



MASTERARBEIT | MASTER'S THESIS

Titel | Title

Measuring hierarchical depth in selected industries

verfasst von | submitted by

Kim Melissa Bär B.Sc.

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of
Master of Science (MSc)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt |
Degree programme code as it appears on
the student record sheet:

UA 066 915

Studienrichtung lt. Studienblatt | Degree
programme as it appears on the student
record sheet:

Masterstudium Betriebswirtschaft

Betreut von | Supervisor:

Univ.-Prof. Dipl.-Chem. Dr. Markus Georg Reitzig
MBR

Abstract

The interplay and design of organizational design factors significantly influence the efficiency and effectiveness of organizations. This thesis provides a basis for future research to measure hierarchical depth in the financial services and software industries. For the assessment, a rule-based coding approach was designed in Python, which allows job titles to be placed in a hierarchical vertical order. The implementation is based on Lee's (2022) framework but represents a nuanced and extended method. In addition to the vertical hierarchy, a distinction can also be drawn between horizontal and geographical hierarchy depth and seniority. Expert interviews were conducted, and salary data was analyzed for validation purposes. The code allows hierarchical depth to be compared within or between industries. Initial results show similar structures regarding the frequency of job title use and the average hierarchical depth, which lies between 5.5 (financial services) and 5.82 (software). The findings indicate that an increasing number of employees highly influence the deepening of hierarchical structures. This knowledge is important in order to draw conclusions for a successful organizational design.

Abstrakt

Das Zusammenspiel und die Gestaltung der organisatorischen Design-Faktoren beeinflussten die Effizienz und Effektivität von Organisationen maßgeblich. Diese Arbeit bietet eine Grundlage für spätere Forschungen, Hierarchietiefe für die Finanzdienstleistungs- und Software-Industrie zu messen. Für die Messung wurde ein regelbasierter Coding-Ansatz in Python gestaltet, welcher es ermöglicht, Jobtitel in eine hierarchisch vertikale Reihenfolge zu bringen. Die Umsetzung beruht auf Lee's (2022) Framework stellt aber eine nuancierte und erweiterte Methode dar. Neben der vertikalen Hierarchie kann zusätzlich zwischen horizontaler und geografischer Hierarchietiefe, sowie Seniorität unterschieden werden. Zur Validierung wurden Experteninterviews durchgeführt und Gehaltsdaten ausgewertet. Der Code ermöglicht es, Hierarchietiefe innerhalb einer oder zwischen Industrien zu vergleichen. Erste Ergebnisse zeigen ähnliche Strukturen bezüglich der Häufigkeit der Nutzung von Jobtiteln und der durchschnittlichen Hierarchietiefe, die zwischen 5,5 (Finanzdienstleistungen) und 5,82 (Software) liegt. Die Ergebnisse deuten darauf hin, dass eine steigende Zahl der Mitarbeitenden einen großen Einfluss auf die zunehmende Hierarchievertiefung hat. Diese Erkenntnisse sind wichtig, um Rückschlüsse für eine erfolgreiche Organisationsgestaltung zu ziehen.

Table of Contents

Figures	I
Tables.....	III
List of Abbreviations	IV
1. Introduction	1
2. Organizational Design Theory.....	3
2.1 Multi Contingency Theory.....	4
2.2 Information Processing	5
2.3 Universals of Organizing	6
2.4. Organizational Configuration	8
2.4.1 Organizational Complexity	9
2.4.2 Coordination and Control.....	10
3. Research Gap and Research Question	15
4. Rule-based Categorization.....	17
4.1 Data	17
4.2 Data pre-processing	18
4.3 Categorization	19
4.3.1 Categorization Differentiation.....	20
4.3.2 Categories.....	24
4.3.3.1 Categories FS.....	26
4.3.4.2 Categories SW	30
4.4 Validation.....	40
4.4.1 Interviews	40
4.4.2 Salary data	43

4.4.2.1 Salary FS industry	44
4.4.2.2 Salary SW industry	49
4.5 Results	51
4.5.1 Results for the Financial Services Industry	52
4.5.2 Results for the Software Industry	54
5. Discussion	56
6. Conclusion	58
Bibliography	IV
Appendix	XI
Appendix 1: Evolution of the Contingency Perspective	XI
Appendix 2: Definition Universals of Organizing	XIII
Appendix 3: External Job Titles	XIV
Appendix 4: SOC Categories FS Industry	XV
Appendix 5: SOC Categories SW Industry	XVII
Appendix 6: Rule-based Matching – FS Code	XVIII
Appendix 7: Rule-based Matching – SW Code	XXVII
Appendix 8: Interview Guide – English	XXXV
Appendix 9: Interview Guide – German	XXXVII
Appendix 10: Interview – Powerpoint Presentation	XXXIX
Appendix 11: Interview Results – Organizational Charts	XLIV
Appendix 12: Interview Transcripts	LII
Interview 1 – FS industry	LII
Interview 2 – FS industry	LX
Interview 3 – FS industry	LXIX
Interview 4 – FS industry	LXXXII
Interview 5 – SW industry	XCVII

Interview 6 – SW industry	CIX
Interview 7 – SW industry	CXIX
Interview 8 – SW industry	CXXXV
Appendix 13: Salary Data Scraping – Code	CXLV
Appendix 16: Merge number of employees file with salary datafile – Code	CLXI
Appendix 17: Evaluate Salary Data – Code	CLXII
Appendix 17.1: Data-preprocessing and category size calculation for salary data-file – Code	CLXII
Appendix 17.2: Statistics calculation for salary data-file – Code	CLXIII
Appendix 17.3: Salary data including ‘number of employees’ variable – Code	CLXVII
Appendix 18: Results – Code	CLXIX
Appendix 18.1: Differentiation features – Code	CLXIX
Appendix 18.2: Category sizes – Code	CLXX
Appendix 18.4: Include variable “number of employees” – Code	CLXXIII
Appendix 18.5: Number of organizations per ‘num_employees’ group.....	CLXXV

Figures

Figure 1 information processing model (cf. Burton, Obel and Håkonsson, 2020).....	6
Figure 2 Overview of the occupation coding approaches (cf. Nahoomi, 2018).....	15
Figure 3 Hierarchical Levels by Lee (2022).....	20
Figure 4 SOC 2018 hierarchy for computer occupations (cf. Soc manual, 2018)	22
Figure 5 Hierarchical depth in the FS industry.....	26
Figure 6 Horizontal differentiation staff, FS industry	30
Figure 7 Hierarchical depth in the SW industry	33
Figure 8 Horizontal differentiation specialists, SW industry	34
Figure 9 Partial ratio matching FS industry.....	38
Figure 10 Salary distribution FS industry.....	45
Figure 11 Boxplot mean salary per category SW industry.....	49
Figure 12 FS category sizes in percent.....	52
Figure 13 SW category sizes in percent	54
Figure 14 Interview pptx slide 1	XXXIX
Figure 15 Interview pptx slide 2.....	XXXIX
Figure 16 Interview pptx slide 3.....	XL
Figure 17 Interview pptx slide 4.....	XL
Figure 18 Interview pptx slide 5.....	XLI
Figure 19 Interview pptx slide 6.....	XLII
Figure 20 Interview pptx slide 7	XLIII
Figure 21 Organization chart interview 1	XLIV
Figure 22 Organization chart interview 2	XLV
Figure 23 Organization chart interview 3.....	XLVI
Figure 24 Organization chart interview 4.....	XLVII

Figure 25 Organization chart interview 5.....XLVIII

Figure 26 Organization chart interview 6.....XLIX

Figure 27 Organization chart interview 7L

Figure 28 Organization chart interview 8..... LI

Figure 29 Code merge number of employees.....CLXI

Figure 30 Code data cleaning number of employeesCLXI

Tables

Table 1 Support function identifier 24

Table 2 Mean salary FS grouped by number of employees 48

Table 3 Mean Salary SW grouped by number of employees 51

Table 4 Hierarchical depth by number of employees FS industry 53

Table 5 Hierarchical depth by number of employees SW industry 55

Table 6 Number of organizations per number of employees – FS industry CLXXV

Table 7 Number of organization per number of employees – SW industry CLXXV

List of Abbreviations

CEO	Chief Operation Officer
cf.	compare
FS	Financial Services
IC	Individual Contributor
i.e.	Id est; that is
inaud.	inaudible
O*NET	Occupational Information Network
SOC	Standard Occupational Classification
SW	Software
VP	Vice President

1. Introduction

Understanding the organizational structure within industries is pivotal for recognizing how organizations operate and achieve their strategic goals. It also plays a critical role in determining the efficiency and effectiveness of an organization. This thesis delves into the concept of hierarchical depth—a measure of the levels of hierarchy within an organization. Hierarchical depth is crucial because it influences various aspects of organizational performance, including decision-making efficiency, communication effectiveness, and the ability to innovate and adapt to market changes (Burton et al., 2020). Within this thesis a rule-based coding framework is developed to specifically measure the hierarchical depth in the financial services and software industries.

The literature foundation is built upon organizational design theory. Key contributions from scholars like Lawrence and Lorsch (1967), Galbraith (1973), and Burton and Obel (2004) have emphasized the importance of aligning organizational structures with environmental factors and information processing needs. Existing literature has explored various methods of analyzing organizational structures, yet there is a gap in quantitative approaches that systematically measure hierarchical depth across different industries. This research extends and validates the work of Lee (2022), who developed a hierarchical framework for start-ups in the game development industry, by applying and refining this concept for larger, more established industries like financial services and software. This study not only broadens the application of hierarchical depth measurement but also integrates new dimensions such as horizontal, geographical, and seniority differentiation, providing a more comprehensive understanding of organizational complexity. The primary research question guiding this thesis is:

How to measure hierarchical depth in the financial services and software industries effectively?

To close this research gap, the objective of this thesis is the development of a rule-based coding system that categorizes job titles using LinkedIn data across hierarchical levels for the financial services and software industries. Providing a rule-based approach offers a more efficient alternative than existing manual or computer-assisted methods. In today's business landscape, organizations and their environments are increasingly unpredictable, rapidly

changing, and require real-time creative design (Van de Ven et al., 2013). This study's use of current, real-world data, i.e. LinkedIn profiles, ensures that the findings are relevant and applicable to contemporary organizational challenges. Furthermore, organizational design research currently lacks rich, large-scale datasets (Puranam, 2012); using LinkedIn data has the potential to close this gap.

Within six chapters, the approach to measure hierarchical depth is conceptualized, explained, validated, and evaluated. The second chapter starts with delving into existing studies on organizational design and positions hierarchical depth in the broader academic discourse. This review provides a foundation for understanding the interplay of hierarchical depth and other organizational design components. In the third chapter, the development of the rule-based coding framework is described. It details the data sources used and the processes for data pre-processing and rule creation. The chapter also outlines the validation methods, which involve expert interviews and salary data analysis, to ensure the accuracy and applicability of the coding system. This section is crucial for understanding the technical and analytical steps taken to achieve the research objectives. The results section presents the findings from the application of the coding system to the financial services and software industries and highlights key insights into the hierarchical structures within these industries. The fifth chapter explores the potential benefits and limitations of the coding system, followed by the conclusion in chapter six.

2. Organizational Design Theory

One of the main questions concerning organizational research is how to design organizations that operate efficiently and effectively. Starting in the early 20th century, researchers focused their research on developing theories and frameworks to guide organizations to be highly successful. One theory that shaped the research with an overarching view on task design and formal organizational structure is the stream of organizational design (Joseph, Baumann et al. 2018). The following chapter will explain the principles and theoretical underpinnings of organizational design theory from a contingency perspective, providing a comprehensive overview of its relevance and complexity. The multi-contingency theory and its application in analyzing organizational configurations are setting the stage for a deeper exploration of hierarchical structures within the financial services and software industries.

Organizational design aims to outline the ideal structure for an organization to operate both effectively and efficiently. It involves a methodical process of coordinating structures, processes, leadership styles, cultural aspects, personnel, and performance indicators to empower organizations to fulfill their goals and strategic vision. The theoretical underpinnings for organizational design theory are provided by organizational theory. As a positive science, it serves to explain and understand the structure, behavior, and effectiveness of organizations. Organizational design theory is a normative science used to develop recommendations for effective and efficient organizations (Burton & Obel, 2018). What is meant by effective and efficient? According to Burton and Obel (2004) “an organization is effective if it realizes its purposes and accomplishes its goals” (p. 4). Efficiency refers to the utilization of the least number of necessary resources to provide services and products. Overall, an organization should be viable to operate in a market.

Organizational design will be addressed from a contingency perspective, integrated into an information processing framework. The fundamental logic is that one organizational model does not fit all scenarios, thus not being equally effective (Galbraith, 1973). The better an organization fits its environment, the more likely it is to successfully operate (Burton et al., 2002). Organizational design theory has come a long way, which has enabled different research streams to thrive; therefore, it is explicitly stated that this viewpoint reflects only one option for looking at organizational design. The evolution of the contingency perspective can be comprehended in Appendix 1. This thesis uses organizational design as a foundation to dive deeper into the specific characteristics of organizing, thereupon explaining the

contextual significance of hierarchical depth. To do so, the multi contingency view of Burton and Obel (2004), which is based on the assumptions of the information processing framework, will be explained. To illustrate the ongoing relevance, the principles of organizing are outlined until finally the structural configurations, including hierarchical depth, are explained.

2.1 Multi Contingency Theory

The multi contingency theory from Burton and Obel (2004) is widely recognized. By introducing a multi-contingency model, they encountered the difficulty of synthesizing and integrating various approaches and established a starting point for further research. The multi contingency model was built as an integration and extension of many single contingency models (Cornelis et al., 2019). It combines theoretical aspects of Leavitt's model (Leavitt, 1965), the Organizational strategy, structure, and process model by Miles et al. (1978), and the Star model by Galbraith (2002). Doty et al. (1993) set the base proposition by establishing that contextual, structural, and strategic factors should align to avoid a misfit. In the multi-contingency model Burton and Obel (2004) defined them as fit criteria, strategic factors, and organizational configurations, whereby all dimensions are interrelated. The fit criteria represent the efficiency, effectiveness, and viability of an organization. Strategic factors allow to analyze a given organization; they include the leadership and management style, the organizational climate, the size and skill capabilities, the environment, the technology, and the strategy (Pugh et al., 1969). Organizational configurations provide design recommendations; they deal with organizational complexity, formalization, centralization, incentives, coordination and control (Chapter 2.4). The multi-contingency model connects the three dimensions by using the information-processing view (cf. Galbraith 1973; Chapter 2.2).

In their book "Organizational Design: A Step-by-Step Approach" (2020), the authors introduced an expanded version of the multi-contingency model, which particularly takes greater account of digital technologies. However, they also changed the structure of the model, whereby internal and external factors are now presented more strongly integrated with each other. This depicts the complex interplay of all factors more accurately but would also lead to a more complicated representation for the purpose of this master's thesis. It was

therefore deliberately decided to describe the model from 2004. It enables the most important structural configurations to be placed in context and to be clearly differentiated from external factors. At relevant points, further and more recent literature is included.

2.2 Information Processing

To facilitate a multi-contingency approach, Burton and Obel (2004) used the concepts of uncertainty and information processing to integrate the organizational design and structure literature. This approach suggests that organizations should balance their structure to align information-processing capacities with the demands they face. Specifically, Tushman and Nadler (1978), along with Burton and Obel (2004), propose using this paradigm to develop a contingency approach for creating feasible structural alternatives for organizations.

Central to this perspective is the idea that higher uncertainty and interdependency among tasks increase the need for robust information handling within an organization, as noted by Galbraith (1973, 1974). Issues are referred to higher hierarchies with sufficient knowledge and skills if uncertainties evolve at the assigned hierarchy. This procedure reflects the traditional exception-based hierarchical decision-making process. However, when the uncertainty exceeds the hierarchy's capacity, it necessitates a shift towards more decentralized task management to alleviate decision-making bottlenecks. Furthermore, Galbraith (1974) articulates strategies to achieve what he terms a fit between demand and capacity. Organizations might structure into semi-independent units to reduce the continuous demand for information or enhance their communication capabilities, both hierarchical and lateral, to increase their overall information-processing capacity.

Figure 1 illustrates how the approach contributes to the context of multi-contingency theory. The strategic factors create a need for information processing. Whereas the organizational configurations represent the ability to process information. Ultimately, a fit of internal and external factors can be achieved.

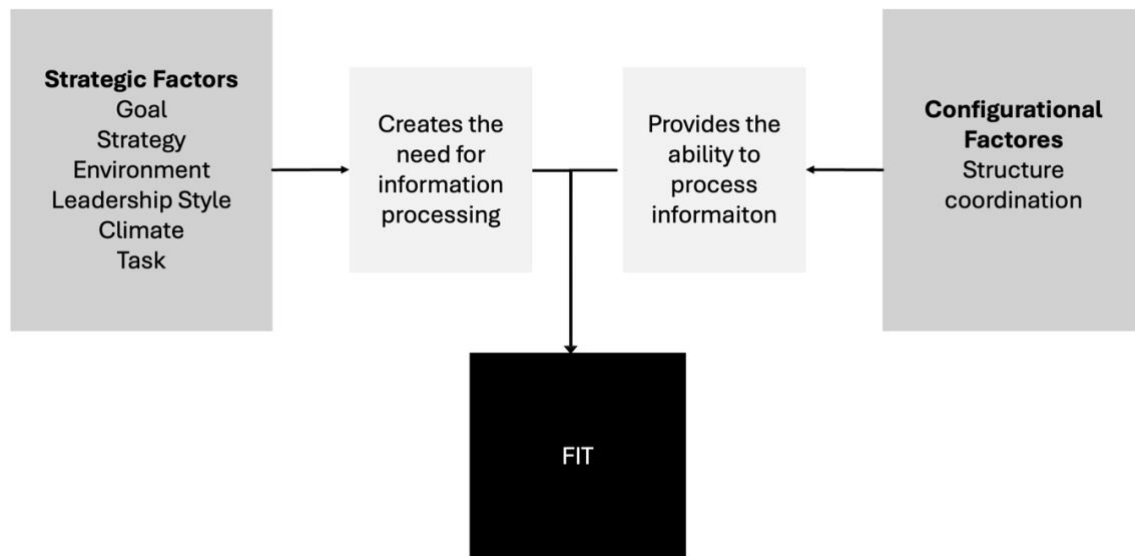


Figure 1 information processing model (cf. Burton, Obel and Håkansson, 2020)

By leveraging this integrated information-processing framework, organizations can effectively design structures that are responsive to both current and future uncertainties. This holistic approach not only provides a strategic pathway for organizational design but also emphasizes the importance of adaptability in the dynamic business environment (Burton et al., 2020).

2.3 Universals of Organizing

The micro-structural approach is a promising view to better adapt organizational design to new market conditions. In the microstructural approach to organizational design, large and complex organizations are conceptualized as aggregations of smaller, simpler, and recurring patterns known as micro-organizations. This perspective focuses on two fundamental processes that are universal to all organizations: the division of labor and the integration of effort (Puranam, 2018). Burton et al. (2020) refer to these universal problems as structure and coordination. These processes facilitate the disaggregation and reaggregation of tasks and responsibilities, effectively linking individual contributions to the broader organizational objectives (Kretschmer & Khashabi, 2020). In order to facilitate this universalizability, it is necessary to clarify the definition of an organization. Puranam et al. (2014) examined common definitions in the field and identified four characteristics that an organization must fulfill: An organization is “a multiagent system with identifiable

boundaries and system level goals toward which the constituent agent's efforts are expected to make a contribution.” (p. 163). This definition enables a broader view of organizations, allowing, for example, individual teams, departments, alliances, or corporations to become the focus object. Assuming that new theories that are embedded in a rapidly changing context lose their validity prematurely, they set up an approach to identify new forms of organizing rather than defining a new theory. By doing so, they specified four problems along the division of labor and integration of effort that must be solved. If at least one of these problems is solved in a novel way, compared to other organizations pursuing similar objectives, a new organizational form can be considered to exist. Furthermore, if all problems are solved, the conditions of the organizational definition should be fulfilled in equal measure. Resulting in the logical inference that the process of solving those four problems, or, to put it differently, organizing these four dimensions, is both individually necessary and collectively sufficient for the existence of an organization and can therefore be representative of "universals of organizing" (p. 166).

The four universals of organizing are, according to Puranam et al. (2014) task division and task allocation, as well as the distribution of rewards and the provision of information. Task division refers to “what needs to be done” and task allocation specifies “who does what”, combined resulting in the definition of the division of labor. Integration of effort specifies the provision of rewards (who gets what) and provision of information (who talks to whom). A more detailed description of the four universals can be read in Appendix 2.

Since these four dimensions do not operate in isolation but are deeply interwoven, they influence and reinforce each other to create a dynamic and responsive organizational structure. But rather than framing new organizational structures as entirely novel, they present them as unique arrangements of solutions to these four problems (Puranam et al., 2014). This viewpoint allows applying existing organizational theories to understand and analyze new forms of organizing, such as those enabled by technology and global connectivity. Furthermore, they laid the foundation to highlight the relevance of organizational design in today's environment.

Recent work by Burton et al. (2020) establishes a similar perspective to the problem-solving approach to organizational design by Puranam et al. (2014). As mentioned above, Burton et al. (2020) refer to the two dimensions as structure and coordination. Structure pertains to how tasks are allocated to individuals or subunits, how resources are distributed among these

units, and how specific markets and customers are assigned to various segments, essentially dividing broader challenges into manageable parts. Coordination is the process of synthesizing these diverse elements, units, and tasks to form a unified entity that fulfills the organization's primary objectives. This synthesis is achieved through multiple mechanisms such as effective communication, information technology, leadership initiatives, cultural norms, incentive programs, established routines, and procedural guidelines, all of which are components of what is broadly termed management (Van de Ven et al. 1976; Burton et al. 2015). In their reasoning, organizational design can be broken down to the simple question of "who does what when" (Burton et al., 2020).

2.4. Organizational Configuration

The conceptualization of organizational configurations helps to understand the self-reinforcement and dynamics of organizing and the embedding of hierarchical depth into it. It is assumed that the definitions will help to understand this thesis in a sufficient way, although organizational design is highly complex, and its implications can differ greatly.

Organizational configuration reflects how work is distributed, how tasks are divided, and how activities are coordinated. It is often depicted by an organizational chart, which shows how the tasks are divided (Burton & Obel, 2004). Based on the organizational chart, it is in most cases quickly apparent whether the tasks are principally organized on a functional or product/market level, which allows conclusions to be drawn about the formal communication patterns. In fundamentally applying task division and allocation, work is coordinated; a common supervision system is introduced, mutual resources are closely allocated and shared, and due to the proximity of work processes, mutual alignment around positions and departments evolves (Mintzberg, 1989). In addition, formal authority can often be read from the chart (Dobrajska et al., 2015). The chosen configuration is a key factor in achieving the desired goals and serves as a basis for further decisions at the strategy level. A mismatched configuration jeopardizes the efficiency and effectiveness of an organization in the short term, which can have a long-term impact on viability (Burton et al., 2020).

The configurations most widely recognized in the past are the simple, functional, divisional, and matrix structures. The most basic structure, the simple structure, has only one head for decision-making, coordination, and control. The functional structure organizes units based

on specialized functions, facilitating efficiency and expertise development within specific domains. Contrastingly, in divisional structures, units are organized into largely autonomous groups overseen by central headquarters. These groups are typically formed around product lines, customer types, or geographical regions. The matrix structure is unique for its dual reporting lines, where specialists from functional departments work on interdisciplinary teams led by a project leader (Burton & Obel, 2004). Organizations nowadays are structured in various forms, from strict hierarchies to flexible network models. Guidelines for decision-making are essential, whether these rules are explicitly formalized or informally understood (Burton et al., 2020). There is a necessity for a mechanism to facilitate information distribution, even in settings where the distribution of information is unrestricted (Puranam et al., 2014).

2.4.1 Organizational Complexity

Organizational complexity reflects the shape of the organization's configuration. It can be evaluated through the lens of horizontal, vertical, and spatial differentiation. They reflect the breadth, depth, and dispersion of complexity, which symbolizes hierarchy.

Horizontal differentiation refers to the breadth of job titles and the extent to which employees possess specialized training and academic qualifications. A higher level of horizontal complexity is evident in organizations that feature a multitude of narrowly defined experiences, educational backgrounds, and specific training. Thus, horizontal differentiation is associated with *specialization* (Burton & Obel, 2004). Moreover, horizontal differentiation also refers to employees' values, motivation, and willingness to perform, which also determine the need for communication and leadership (Fallgatter, 2020).

Vertical differentiation refers to the layers of hierarchical levels that span from top to bottom in the organization. This concept may also be characterized as *hierarchical depth*, which reflects the extent of vertical differentiation. It can be assessed by examining the deepest vertical reach within any given unit and the average hierarchical depth across the organization (Burton & Obel, 2004). This thesis places significant emphasis on the analysis of vertical differentiation.

However, it is essential to acknowledge that horizontal and vertical differentiation are intrinsically linked, with both dimensions contributing to an organization's **span of control**. The span of control refers to the number of subordinates effectively managed by a superior

(Mintzberg, 1989). A broader span of control facilitates increased horizontal differentiation, whereas a narrower span normally enhances vertical differentiation by requiring a downward chain of command to accomplish the task volume. Among the factors that influence the span of control are the complexity of tasks, the cognitive abilities of managers, their management style, and the extent to which decision-making processes are supported by the organization's information and knowledge systems. Thus, fluctuations in one of these measures are likely to necessitate corresponding adjustments in the others (Burton et al., 2020). In organizations where tasks and processes are highly standardized, coordination reduces significantly, thus allowing for a greater span of control. Underlying factors are the reduced time for managerial interventions and direct supervision. The trend of an economy that tilts to be knowledge-based results in tasks that are likely to be rather complex. Therefore, information sharing, and face-to-face communication are highly relevant for the coordination of tasks, which should reduce the span of control.

Spatial complexity is a measure of an organization's *geographical differentiation*. Greater spatial complexity is characterized by a broad geographic distribution of facilities and personnel. For example, the total number of operating locations, the average distance between these locations and the headquarters, and the worker distribution ratio among these sites are among the metrics used to measure spatial complexity (Burton & Obel, 2004).

2.4.2 Coordination and Control

Coordination and control are highly impacted by organizational complexity. Coordination reflects how the management brings units together to achieve an overall purpose (Burton & Obel, 2018; Van de Ven et al., 1976). Given that they mirror managerial responsibilities, it appears evident that their function is to guarantee the availability of pertinent information when needed and to allow that the right decisions can be reached (Burton & Obel, 2004).

Information and knowledge systems supply critical data required for the functioning of coordination and control mechanisms, which lead them to form the third element of essential infrastructure for organizations. When structuring these systems, organizations must decide on the level of formalization and the degree of centralization to adopt (Burton et al., 2020).

Centralization is a reference to the hierarchical level that has *formal authority* over decisions. It refers to the extent to which coordination and control are concentrated at a central point (Liao et al., 2011). Centralization is measured by examining how closely

managers are involved in key decision-making processes, such as budget allocations, oversight of evaluations and rewards, personnel decisions, and procurement activities. Through high centralization, direct supervision and, thus, tight coordination are ensured. On the other end of the dimension is **decentralization**. Congruently, it refers to the degree to which decision-making authority is distributed to lower levels within the organization, such as individual managers or sub-units, away from centralized corporate headquarters or a singular hierarchical level. Decentralization considers more than formal authority. Similar to differentiation, it can be separated into two categories: vertical decentralization, the delegation of formal authority to line managers lower in the hierarchy; and horizontal decentralization, the dispersal of formal or informal authority away from the line hierarchy to non-managerial roles (Burton & Obel, 2004; Mintzberg, 1989).

Many contemporary organizations are shifting towards more decentralized models for managing coordination and control. This shift acknowledges the value of decentralization in addressing the varied needs of a geographically dispersed organization and in empowering local units with greater responsibility for organizational actions. Reitzig (2022) argues that decentralized decision making can enhance to adaptive capabilities of organizations and has the potential to enhance the working attitudes among motivated personnel. The centralization-versus-decentralization distinction is particularly relevant to operational decisions. When strategic decisions are reserved for the upper echelons of an organization while operational decisions are delegated to lower levels or sub-units, the organization leans towards decentralization. This suggests that coordination and control systems are more focused on designing work processes than on formulating strategy or policy (Alonso et al., 2008; Burton et al., 2020). Hart and Moore (2005) noted in their work that the benefits of centralization and coordination must be considered against each other. The organization should be centralized if the benefits of centralization outweigh the additional coordination effort; if coordination is moderate, the organization should be decentralized. Moreover, they stated that an organization should split up if the synergetic benefits compared to coordination are small. Centralization and decentralization are, as expected, connected to other design decisions, and the best option is situational.

One concept that is closely linked to centralization and decentralization is an organization's **information processing capacity**. Grounded in the principle of *bounded rationality*, it is posited that individuals possess a finite capability for absorbing, retaining, and interpreting

data. The extent to which an individual can manage information is contingent upon their competencies and level of education. One way of arguing is that decentralization typically increases when information processing demand increases, since a broader distribution of decision-making authority is necessary. Especially when organizations grow, decentralization occurs since the magnitude of the workload exceeds the capabilities of one person. On the other hand, decentralization often is assumed to decrease when the information processing demand increases, which leads to a more centralized decision-making structure to effectively manage the increased amount of work.

Formalization is the codification of organizational regulations. The degree of formalization reflects the amount of rules that regulate how work is done, who does it, and which constraints apply under which circumstances (Burton & Obel, 2004). Organizations that mainly depended on formalization to ensure coordination were typically called bureaucracies (Mintzberg, 1989). Formalization is high when organizational rules are detailed, codified, and uniformly conveyed to the workforce. The introduction of training programs, demonstrating a specific behavior, and verbal guidelines can equally lead to high formalization. In contemporary practice, rules are frequently embedded discreetly within IT systems and user interfaces. In contrast, when formalization is low, there are neither explicit nor recognized rules or codes, leading to greater variability and flexibility in the operational practices of the organization. Typically, organizations find themselves operating on a continuum that ranges from high to low formalization (Burton et al., 2020). The implementation of formalization serves as an instrumental approach to achieve standardized behavior, which can result in lower costs, higher product quality, and more efficient operations. Problems resolving from increased organizational complexity can be counteracted with increased formalization. Typically, in scenarios where both attributes are high, the level of formalization may not be as extensive as it otherwise would be. Empirical research suggests that the impact of these factors is considerably influenced by organizational size. Moreover, the modification of employee behavior can be achieved through the strategic deployment of incentives (Burton & Obel, 2004).

Incentives are tools employed to motivate and evaluate the performance and contributions of organizational members. They are designed to promote specific behaviors or outcomes among employees or groups within the organization. Beyond monetary compensation such as salaries and benefits, incentives also include non-financial rewards like praise,

recognition, and social inclusion. Non-material incentives such as commendation can also be significant motivators. Ultimately, for incentives to be effective, they must be perceived by employees as equitable and motivational, leading them to commit to and excel in their roles. An incentive is generally considered sufficient when it is seen as proportionate with the effort required and consistent with rewards given to others in similar roles or situations (Burton et al., 2020). Promotions often reflect an efficient incentive to motivate employees and enhance their job satisfaction. Furthermore, they offer a clear career path, which can increase the commitment and reduce the turnover (Jiang et al., 2012). The design of incentive systems is crucial and should align with either behavioral norms and procedures or with the expected outputs of tasks. Incentives focused on behavior encourage adherence to and efficiency in organizational processes, while those based on outputs steer efforts towards fulfilling organizational goals and maximizing the impact of those efforts. A key factor in implementing incentive systems is the ability to monitor and evaluate whether the focus is on processes or outcomes. This distinction between monitoring behaviors and outcomes is essential, as it influences the management approach—whether it involves overseeing the process of work or assessing the results (Burton & Obel, 2004).

In general, it can be summarized that an increase in organizational complexity, thus also in hierarchical depth, is expected to be followed by intensifying coordination, control, and information processing challenges (Burton et al., 2020). More complex organizations will likely have a higher amount of information exchange and thus, needing higher information processing capacity and robust control mechanisms (Tushman & Nadler, 1978). In turn, formalization and incentives can help to reduce the need for higher information processing and can function as control mechanisms. Effective coordination mechanisms can be highly beneficial for specialized organizations. Although specialization might facilitate overall efficiency, by enabling the personnel to do what they can do best and thus minimizing coordination effort, it might promote individual development of terminologies, tools, and processes in specialized subgroups, which increases the coordination amount (Perrow, 1977). Well-established information systems can further enhance effectiveness by reducing the hurdles higher vertical, horizontal, and especially geographical differentiation introduces (Burton et al., 2020).

This chapter aimed to define hierarchical depth and contextualize the conceptualization. It highlighted the complexity of the interplay between organizational configurations, with the

potential for further research based on hierarchical depth to provide important implications. Additionally, the work of Puranam et al. (2014) should have demonstrated that principles of organizing remain relevant for new organizational forms and that such studies are essential for academic research. Building on these insights, the next chapter shifts focus towards a more quantifiable aspect of organizational structure—hierarchical depth. This measurement is pivotal for understanding how traditional and emerging business models differ not only in their management and operational approaches but also in their structural depth across various industries. For this purpose, a code was developed that systematically measures the depth of hierarchy in the financial services (FS) and software (SW) industries and thus, provides a concrete tool for analyzing and comparing the structural complexity of organizations. This approach will allow to examine the implications of hierarchical structures on organizational efficiency and adaptability, crucial for navigating the complexities of today's dynamic business environments. By quantifying hierarchical depth, the extent to which organizations adhere to or deviate from traditional hierarchical configurations can be better assessed, and how these structural choices impact their operational success and innovation capacity can be explored.

3. Research Gap and Research Question

To facilitate the named advantages, hierarchical depth needs to be quantifiable. The approach introduced in this chapter will allow a measurement of hierarchical depth.

Explicitly classifying job titles into a sphere of hierarchical depth to visualize organizational structure is very limited in existing literature. More often, for statistical purposes, occupations are classified, commonly based upon a variety of factors such as educational background, skills, and work performed (Nahoomi, 2018). Occupational coding is defined as a system to classify job titles into standardized categories. It is essential for social science research, enables international comparisons, and is the starting point for a variety of scale measurements, especially for social inequality and social stratification (Ganzeboom & Treiman, 2003). Occupational classification methodologies are primarily categorized into three types: manual, computer-assisted, and fully automated systems (Nahoomi, 2018), as delineated in Figure 2, which provides a schematic representation.

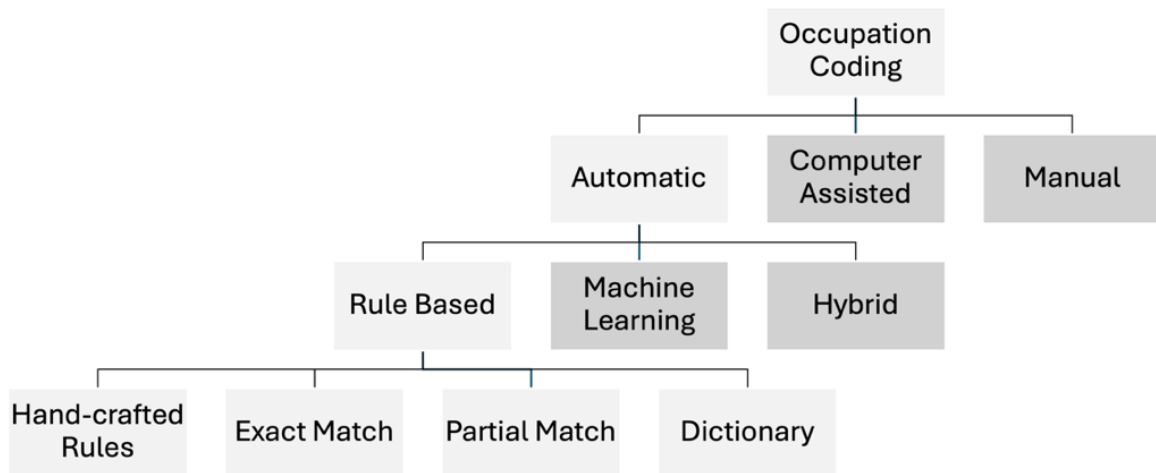


Figure 2 Overview of the occupation coding approaches (cf. Nahoomi, 2018)

Traditionally, occupational titles were categorized by human coders, who employed detailed guidelines to perform their classifications manually. This method, while thorough, is notably labor-intensive and financially demanding. Moreover, the limited availability of skilled coders presents a significant constraint (Patel et al., 2012). Manually coding occupational titles into standardized classifications demands considerable time and resources. Burstyn et al. (2014) stated that a single trained coder needs between four and eight years to categorize

a database containing 200,000 job titles. For reference the proposed code classified over two million job titles within hours.

The automation of occupational classification has the potential to address many of these challenges. In the realm of computer-assisted classification, specialized software is employed to facilitate the coding process. This software proposes several potential categories, from which coders then select the most appropriate one (Schierholz, 2014). This approach is more efficient in terms of both time and costs compared to manual classification, although it still relies on human judgment for final category selection.

Fully automated classification systems directly assign categories to occupational titles using algorithms, enhancing both time efficiency and cost effectiveness while reducing human error. Nonetheless, perfecting these automated systems to ensure high-quality output consistently remains an ongoing challenge (Gweon et al., 2017). In rule-based classification, a predefined set of rules is applied to categorize job titles. This method is more efficient than manual classification, but the setup requires substantial datasets and meticulously crafted rules to ensure dependable categorization (Nahoomi, 2018).

Lee (2022) used rule-based coding to measure hierarchical depth in the game development industry, focusing on high-growth start-ups. The work showed significant differences in various success measurements based on the hierarchical depth. Leading to further implications for the organizational design to create an organizational fit. Structural configurations determine how successfully an organization can adapt to external changes and thus stay highly relevant when designing (new forms of) organizations. Lee's approach has not been validated so far and is limited to one industry. This thesis aims to expand Lee's framework and propose a more detailed hierarchical depth measurement for the FS and SW industries. Which leads to the research question:

How to measure hierarchical depth in the financial services and software industries effectively?

4. Rule-based Categorization

This chapter introduces the rule-based occupational coding framework. It evolves from describing the data and the data pre-processing to introducing rule-based categorization and finishes with the validation of the proposed hierarchical categories and the measured depth in the industries.

4.1 Data

The dataset used in this thesis covers almost 40 million LinkedIn profiles from the United States to draw information about job titles and organizations in the FS and SW industries. The variety of LinkedIn job titles is extremely high, since users are free to specify their own titles. Thus, data cleaning and a sufficient analysis process are highly important to reduce noise (Bekkerman & Gavish, 2011).

The industry classification was evaluated using Crunchbase data. Crunchbase is a leading platform that provides private-company data to identify valuable insight and get verified contact details (Crunchbase, 2024). The LinkedIn and Crunchbase data for the selected industries were provided by the chair of Strategic Management at Vienna University. As the quantity of data is immense, the majority of the coding and testing were carried out with smaller samples. Fractions were then taken from various files to improve validity (i.e. Appendix 6).

The FS and SW industries were selected based on available interview partners, who provided significant insights into their organizational structure with the aim of verifying the proposed hierarchical structure for validation purposes. Another decision criterion was the complexity of the industries. Whereas the SW industry is recently thriving and evolves around new organizational forms (Donohue & Power, 2012), the FS industry comprises a variety of different types of institutions (Van den Berghe & Verweire, 2001). For further research, a high variety of the structure of organizations and the size of organizations within the industry is optimal; both selected industries fulfill this condition.

4.2 Data pre-processing

Data pre-processing is a critical step in the analytical process that entails the modification and manipulation of raw data to derive significant insights. Particularly, data cleaning is an essential preliminary activity that focuses on the correction, elimination, and enhancement of data that may be inaccurate, extraneous, duplicated, or improperly formatted (Batini & Scannapieco, 2006).

Out of 26 columns, twelve were collected—*name*, *id*, *experience title*, *experience company*, *legal name*, *experience start date*, *experience end date*, *experience duration short*, *city*, *type*, *Crunchbase industries*, *industry dummy*:

- *Id* works as a unique identifier for a person,
- *experience title* contains the job title specified by the individual on LinkedIn for the current and previous jobs,
- *experience company* expresses the organizations name, whereas *legal name* was kept as well because of a greater accuracy of the organization name but due to many missing entries it solely would not be useful,
- *type* mainly has two different entries “company” and “investor”, whereas investors are being excluded from the dataset,
- *Crunchbase industries* and *industry dummy* were used as an identifier of the selected industries. Organizations that failed to be classified with a relevant industry tag in Crunchbase were subsequently excluded from the work. Resulting in 77.5% of the entries in the FS dataset being dropped and 74.4% in the SW dataset.
- The remaining columns were kept for further analysis, remaining unused in this study.

In order to facilitate working with the data, all missing entries have been overwritten with the string "no_info". Using a placeholder ensures that operations such as visualizations and calculations can proceed without errors due to unexpectedly missing values. Placeholders are a non-technical method to ensure data consistency and facilitate various stages of data processing (VanderPlas, 2016). The *legal name* column displayed additional quotation marks, which hindered the processing of the data. The text contents were converted to strings. Furthermore, to avoid case sensitivity, the contents of the columns were changed to lowercase. Compared to the basis of other text-based analysis, job titles are considerably

short, with an average of three words per job title. Thus, most studies do not recommend using stemming, some propose to remove stop words, and others recommend using neither (Nahoomi, 2018). This study went along with the last recommendation. The attempt to use methods to autocorrect spelling mistakes (e.g., TextBlob) resulted in many miscorrections, especially for abbreviations. Thus, the job title spelling was left unchanged. Instead, common misspellings were included in the categorization base (rules); for example, in addition to *director*, the notations *directo*, *diirecto*, and *diirector* were added.

It was noticed after an analysis of the test dataset that company names such as "self at home", "self", "owner at home", "retired", and "www" did not represent organizations, therefore they were excluded. Under *experience title* job titles were identified that are not related to the work of the selected industries and consequently could not be categorized in any associated hierarchy, such as waitress, police officer, and nurse. A detailed overview is provided in Appendix 3. In addition, five organizations were identified in the FS dataset whose core activity is to enable other people to start their own businesses in various sectors. Customers of these organizations entered one of these five company names under *legal name* and "owner" as the *experience title*. Consequently, entries that contained both specifications were excluded; this procedure allows to remove the clients but also retain employees of these organizations in the dataset.

After the execution of all data preprocessing steps, 19.9% of all data, resembling 3.8 million observations in the FS dataset and 23.9%, 4.9 million observations in the SW dataset, were the starting point for further analysis.

4.3 Categorization

The categorization of the job titles represents the main work of this thesis. The work is based on Lee's (2022) 11+1 level structure, which he used to measure the hierarchical depth of start-ups in the game development industry.

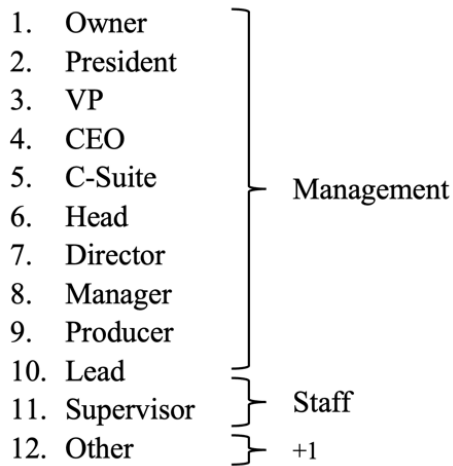


Figure 3 Hierarchical Levels by Lee (2022)

This work extends Lee’s structure to 16+2 (FS) and 20+2 (SW) general hierarchical levels, whereas +2 represents support functions and the *Other* category. Several levels have a more detailed second-level structure; the logic of these second-level structures will be explained in a general manner in the following, as well as the matching algorithm. The final categorization for each industry will be explained afterwards.

4.3.1 Categorization Differentiation

The hierarchy levels are divided into three organizational groups: management, staff, and support functions.

Regarding the management group, it has proven to be useful to additionally differentiate between **seniority** on four levels. Seniority denotes the duration for which an employee has held a specific role. It can be attached to the length of employment, the duration in a specific function, or a blend of both factors. Seniority serves as a criterion for assigning benefits, duties, and pay scales across the organizational hierarchy (Dessler, 2020). Promotions usually represent a leap in seniority, and, depending on the size and structure of the organization, the management structure may also change as a result (Xu et al., 2018). Interviewee 1, who works for ZEB stated:

“It [the promotion] doesn't really change who you report to, as long as the person who you reported to before is also in the same role and is still higher in hierarchy than you, so if you would get promoted to the same hierarchy level then it might change.” (00:02:40)

This leads to the assumption that a seniority in job titles does not necessarily reflect a deepening of the hierarchical structure. The additional consideration of seniority at the second level provides an opportunity to group all persons with the same management job title but to examine a more precise distinction if required. A distinction is made between:

- x.1 executive,
- x.2 senior,
- x.3 title without seniority indicator, and
- x.4 junior or assistant.

The category *Manager* shows further detailed differentiation. Additionally, this category is divided between:

- x.01 program manager,
- x.02 project manager, and
- x.03 department manager in the industry.

The job requirements for a manager can differ significantly; for example, someone in project management may have a much broader scope of responsibility and span of control compared to a "junior information technology manager," where the title "manager" serves more as a prestige enhancer. An inflation of manager titles has occurred in recent years, with smaller organizations in particular often opting for titles that are associated with higher professional qualifications and greater prestige (Martinez et al., 2008). Interviewee 2, who works in a startup providing cash systems with approximately 85 employees, explained her promotion as follows:

“In our team, it's a bit different because everyone has the title manager, and so I got the title [...]. But nothing changed, and none of my tasks changed. So, it was just the title.”

Within the staff group, **horizontal differentiations** of job titles were implemented. As the scope of the data covers an entire industry, individual job descriptions, and therefore responsibilities and required qualifications, might vary greatly depending on factors such as strategic focus, organizational size, and organizational culture (Cascio, 2015). The analysis

of job descriptions and, accordingly, the categorization is a highly complex task (Ikudo et al., 2019). The *Standard Occupational Classification (SOC) System* offers the possibility of implementing a reliable system to differentiate job titles on a horizontal sphere. It is published by the United States Bureau of Labor Statistics and serves the purpose of classifying employees and jobs into occupational categorizations with the aim of gathering, computing, analyzing, or distributing data. All jobs in the national economy are covered by the SOC. The SOC 2018 is a hierarchical, four-level system, starting with 23 categories, which are then organized into 98 subgroups. These subgroups are further specified into 459 general occupations; the fourth level comprises 867 specific occupations. Every occupation in the SOC can be identified using a six-digit code, which gets more specific at every level (U.S. Bureau of Labor Statistics, 2018c). This work focuses on the categories for Business and Financial Operations (13-0000 and 43-0000) and Computer and Mathematical Occupations (15-0000). Figure 4 represents the hierarchical differentiation of the computer occupations (15-1200) on the first three levels. The categories used for the SW data follow this differentiation for the *Specialists* category.

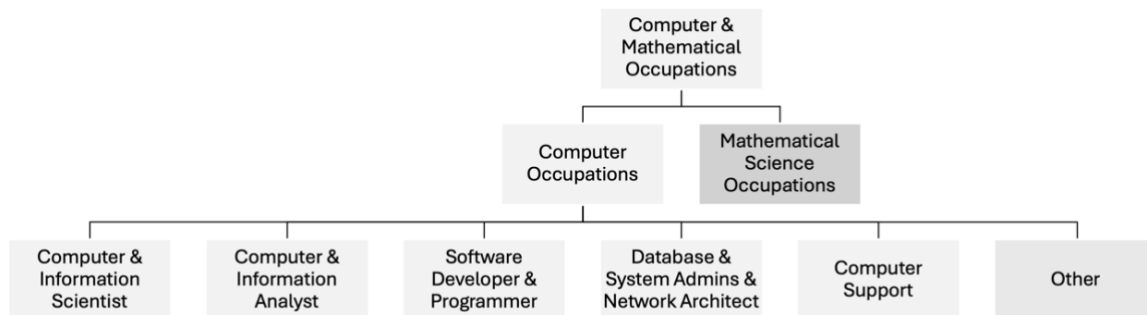


Figure 4 SOC 2018 hierarchy for computer occupations (cf. Soc manual, 2018)

In general, the SOC manual provided the starting point for the structure of the *Specialists* and *Workers* categories. It includes an explanation of the structure, a short job description, and up to four job title examples. More job title examples were extracted from the direct match title file (U.S. Bureau of Labor Statistics, 2018a). To expand the existing categories, the data provided by the Occupational Information Network (O*NET) was manually implemented. The O*NET Resource Center (2024a) provides a wide online database that links the occupational codes to skills, knowledge, and abilities. More importantly, it offers data on alternate job titles (O*NET Resource Center, 2024b) and frequently reported job titles reported by experts (O*NET Resource Center, 2024c), which were used to enrich the categories. Horizontal differentiation is again indicated within the rules by x.01 – x.09.

Additionally, the *Specialist and Expert* groups summarize all senior positions at the level indicated by the category number with the addition x.2, e.g. senior analyst. Following the same logic as for the manager group, that a differentiation in seniority can indicate a hierarchical deepening but does not have to.

Summarized, the *Specialist and Expert* groups show detailed horizontal differentiation, whereas the management group only divides the *Managers* category horizontally. This logic is based on the diversity of job titles and the associated tasks. In the management group, the levels are usually sufficiently represented with fewer than five job title variations, whereas in the specialist groups, the titles and title variations exceed this number. Contrarily, the management group has a more detailed structure regarding seniority, as the data showed.

Furthermore, **geographical differentiation** is implemented, adding a new hierarchical dimension. Organizational geographical differentiation is more prominent than ever in today's globalized world. Job titles can indicate in which regional dimension an individual works, thus implying different skill sets and expertise about different economic conditions and cultural expectations. A matching skill and knowledge set can provide a more nuanced understanding of regional labor markets and the specific demands of various locals (Aufiero et al., 2023; Lazarova et al., 2023). This hierarchical indicator is based on alphabetical values. This allows a sufficient combination with the other indicators. The classification is as follows:

- a. Global,
- b. Global Regions,
- c. Countries,
- d. Regional Areas, and
- e. Branch.

Besides job titles that evolve around core functions, organizations usually need employees involved in support tasks. **Support functions** are essential in organizations because they facilitate core business activities that directly generate revenue by providing necessary secondary services like human resources, marketing, and sales. These functions are pivotal for smoothing operations and enhancing overall business efficiency and effectiveness (Smet et al., 2022). Lee's (2022) framework does not sufficiently distinguish between core and support functions, in contrast, this work integrates an additional filter step. Thus, a support

function categorization for sales, marketing, human resources, business development, legal, and either finance or information technology is included. In contrast to the explicit categorization of the core job titles, keywords (e.g., sales) were used for the support divisions. Leading to a more general identification of job titles belonging to the support functions rather than an explicit identification of the job title hierarchy.

Support Function Type	Category
Finance or Information Technology	30.01
Sales	30.02
Marketing	30.03
Human Resources	30.04
Business Development	30.05
Legal	30.06

Table 1 Support function identifier

In addition, job titles in the C-Suite that belong to the support functions are named separately. This peculiarity is taken into account because of the common practice of abbreviating these titles, e.g. CMO for Chief Marketing Officer. Including these titles in the categorization level (30.01- 30.06) would make it considerably more difficult to assign them to the correct level. The logic becomes clear when the first step of job title matching is addressed. In general, this procedure enables a simplified form of matching; a precise categorization of the five other areas is beyond the scope of this thesis.

4.3.2 Categories

The final levels for categorization, reflecting vertical differentiation, should complete the picture of the drawn explanations of the previous chapters. Firstly, the categories regarding the FS industry are explained, any further additions in the SW industry will be explained afterwards and will be marked blue in Figure 7. As already mentioned, Lee's (2022) levels provided the base for the hierarchical depth development. Except for the expansion, further changes compared to Lee were undertaken. The level *Producer* is excluded, and the levels *CEO* and *C-Suite*, also called the Chief-Suite, were pulled to the fourth and fifth highest positions, respectively. Several sources suggest that a director has to be above the head position since the span of control is most likely higher for directors (Nagpal, 2024). For the FS industry, the level *Supervisor* is positioned above the level *Lead*. Since a lead in the FS

industry more likely represents a team lead, which has less authority compared to a supervisor or a technical lead (SW industry). In the code the defined levels are reflected by rules, called tuples in technical terms. Each rule contains several strings, which indicate the job titles; in total, the rules in the FS industry contain 543 strings and in the SW industry, 613. Thus, the rules will be explained in a comprehensive manner, more details can be extracted from the code in Appendix 6 and 7.

The identified categories and the hierarchical order of the categories are based on the following aspects: The identified job titles in the LinkedIn dataset, published data about organizational charts, published articles, conducted interviews, and fetched salary data. The LinkedIn job titles were checked several times manually in an iterative process. The starting point were Lee's (2022) categories, which allowed for an initial categorization of those levels. The remaining (uncategorized) job titles were then checked regarding frequency and possible grouping. Additional levels were drawn from websites providing information on organizational charts, e.g. The Org (theorg.com), Organimi (www.organimi.com), The Official Board (www.theofficialboard.com), and levels.fyi (www.levels.fyi). The academic literature on hierarchical depth based on job titles is very limited. Therefore, much information on how to categorize the levels was drawn from business websites or job portals providing guiding articles, most often for job seekers. Validation is built upon expert interviews and salary data analysis.

4.3.3.1 Categories FS

Figure 5 shows all the hierarchical levels identified in the FS industry.

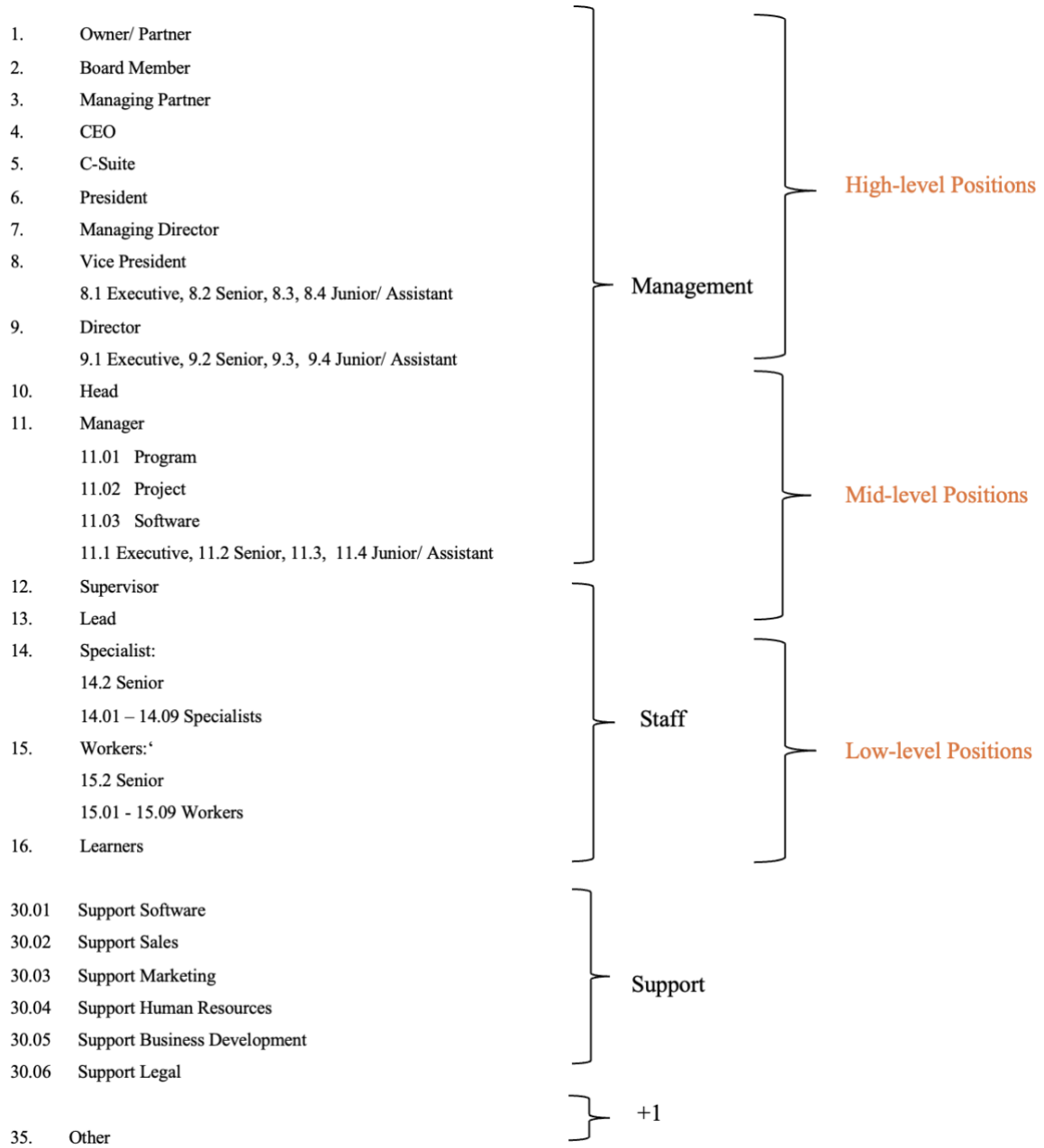


Figure 5 Hierarchical depth in the FS industry

The level *Owner* (1) was expanded with the job title *partner*. A partnership reflects shared ownership, including legal and profit-related regards (Kopp, 2024).

The second level *Board Member* was added. Board members take a governance position and often oversee the organization’s activities. Their responsibilities can be divided into legal duties, strategic oversight, and management support. Board members can be internal and

external personnel. For large public organizations it is common that they are external, thus they do not have a day-to-day role in organizational operations. Due to their duties, a comprehensive knowledge about the organizational activities is required. The CEO reports to the board members, which gives them a special governance position at the top (Saab, 2023). Interviewees 4 and 8 confirmed the hierarchical chain of CEOs and board members (Appendix 12, Interview 4 – FS industry, Interview 8 – SW industry).

The third position, *Managing Partner*, is also newly added and compromises managing and leading partners. A partner is usually not involved in the management operations; comparatively, a managing or leading partner is involved but still has ownership interest (Zippia Team, 2022). Depending on the organizational type, a leading partner can hold the highest position in an organization (Appendix 12, Interview 1 – FS industry).

The Chief Executive Officer (*CEO 4*) is widely known as the top executive in an organization; five out of eight interviews verified this assumption. The other interviews led to the categorization above. The CEO sets the strategic goals, vision and mission, and decides which strategies the organization implements. The focus of a CEO must comprise all aspects of the organization, contrarily the C-Suites focus on a single aspect or division. Therefore, below the CEO typically evolves the *C-Suite (5)*. The titles here are ordered horizontally, mostly around the support and core functions of an organization. They are responsible for strategic decision-making (Saab, 2023).

The sixth hierarchical level is the *President*. Compared to the chiefs, the president is more involved in the day-to-day business functions. They often determine how the vision and mission are achieved with interim objectives or set strategic intervals to reach those goals. Next to the president, the *Managing Director (7)* is often considered the highest-ranking position outside of the C-Suite. A managing director also oversees the day-to-day business strategy and implements strategies (Bachman, 2023). The term managing director is not as frequently used as CEO or president, but sometimes they are used interchangeably. The term managing director is most often used in British organizations instead of CEO. In smaller organizations, the managing director often resembles the president, which is more frequently used in the USA. It is unlikely that one organization uses both job titles; however, since the utilization of job titles is not static, it cannot be ruled out (Bean-Mellinger, 2019).

A plan for reaching the strategic goals will be most likely established by the *Vice Presidents* (8). Vice presidents are essential in crafting strategies and policies for their respective areas, ensuring these align with the broader goals of the organization.

Directors (9) typically oversee entire departments or functions, playing a decisive role in realizing the strategic goals established by the C-Suite. They usually report to vice presidents or directly to C-Suite executives, depending on the organizational hierarchy (Nagpal, 2024). Their function evolves around overseeing, mentoring, and teaching employees.

Heads (10) usually manage departments, e.g., head of marketing, or small working groups. The title head of is more commonly used in young or smaller organizations, whereas the title director is often found in more established and bigger organizations. Their areas of responsibility can have considerable similarities, with the biggest difference being the scale of the teams. A head usually oversees smaller teams, which allows for more direct contact with the employees (Cook, 2022). As suggested by Interview 5, the *Director* level is thus positioned above the *Head*, different from Lee's (2022) structure (Appendix 12, Interview 5 – SW Industry).

The last level in the management group are the *Managers* (11), noteworthy is the differentiation between horizontal and seniority titles. In general, managers ensure that their teams meet operational targets and project deadlines. Horizontally, program, project, and department managers are categorized. A program manager coordinates and supervises multiple related projects. They are responsible for setting program controls, managing budgets, overseeing interdependencies, and ensuring that milestones are achieved (Joubert, 2020). Project managers oversee the life cycle of specific projects and ensure they meet time constraints within budget and the required quality standards. Their duties include defining project goals, developing detailed plans, and monitoring progress (Ahmed et al., 2021). In general, department managers oversee departmental operations, ensuring activities align with the organizational strategy and goals. They implement strategies, manage budgets, and oversee team performance. Often, they are closer to the employees than the upper management, thus having better influence and control over the (productive) work environment, performance targets, and evaluations to improve efficiency and effectiveness (Indeed Editorial Team, 2023).

In the management group, the levels of *Vice President*, *Director*, and *Management* have additionally a seniority differentiation.

The next group resembles the Staff, covering the *Supervisors*, *Lead*, *Specialists*, *Workers*, and *Learners*. *Supervisors* (12) generally possess work experience and engage in tasks similar to those performed by the employees in the *Specialist* category. Thus, they are experts with additional supervisory responsibilities (U.S. Bureau of Labor Statistics, 2018b).

Different from supervisors, a *Lead* (13) in the FS industry often reflects a team leader. Team leaders work with the team but are responsible for guiding and managing it. To motivate, communicate, and resolve conflicts are key tasks for those roles. Leads in general often provide technical or subject-matter expertise (Mazzetti & Schaufeli, 2022).

Specialists (14) are well trained and crucial for the operative tasks. *Workers* (15) resemble office administrative and support occupations with less educational training or experience (U.S. Bureau of Labor Statistics, 2018b).

The last official level *Learners* (16) inherits all staff in training positions, likely individuals who are at the beginning of their careers.

The categories *Specialists* and *Workers* differentiate between senior and non-senior roles. Executive roles are more likely to be found in the management group, and junior roles resemble the differentiation between specialists and workers. As already explained, both categories are horizontally differentiated, following the SOC 2018. Figure 6 shows the detailed differentiation of the categories, and Appendix 4 includes the matching SOC levels for the categories. As described in Chapter 4.3.1, the job titles within the categories are allocated based on O*Net and SOC data (O*NET Resource Center, 2024b, 2024c; U.S. Bureau of Labor Statistics, 2018c). The title identification process was a manual multi-iteration cycle until a sufficient default rate was reached.

14.2 Senior Specialists

- 14.01 Buyers
- 14.02 Coordinators and Consultants
- 14.03 Accountants and Auditors
- 14.04 Credit Analyst
- 14.05 Financial Analyst, Advisors, and Underwriters
- 14.06 Financial Examiner
- 14.07 Credit Counselors and Loan Officers
- 14.08 Tax Examiners, Collectors, and Revenue Agents
- 14.09 Others

15.2 Senior Workers

- 15.01 Bill and Account Collectors
- 15.02 Billing and Accounting Clerks
- 15.03 Bookkeeping, Accounting, and Auditing Clerk
- 15.04 Tellers
- 15.05 Credit Authorizers, Checkers and Clerks
- 15.06 Loan and New Account Clerks
- 15.07 Personal Banker
- 15.08 Personal Assistants
- 15.09 Others

Figure 6 Horizontal differentiation staff, FS industry

4.3.4.2 Categories SW

For a better comprehension of the SW industry, it is useful to briefly outline the different career paths. Essentially, one can distinguish between three career paths: the management, individual contribution, and project management career paths. A managerial career path involves leading teams and departments to achieve organizational objectives. Managerial staff is responsible for setting goals, developing strategies, and ensuring their team's alignment with the organization's vision. Key responsibilities include performance management, team development, conflict resolution, resource allocation, and communication (Nagpal, 2024).

Project managers focus on planning, executing, and closing projects on time and within budget. They are responsible for defining the project scope, setting timelines, and managing

budgets. Project managers coordinate cross-functional teams, mitigate risks, and ensure project deliverables meet quality standards. Their role requires a blend of technical understanding, organizational skills, and leadership capabilities to manage project life cycles effectively (Ahmed et al., 2021).

The third path resembles the individual contributor (IC). They use their expertise to contribute directly to the tasks and goals of the organization without assuming managerial responsibilities. The role is characterized by specialization and depth of knowledge in a specific domain. ICs are critical for their technical skills, problem-solving abilities, and innovative contributions. They are often responsible for producing high-quality work, driving technical projects, and providing insights that influence strategic decisions.

Based on the chosen career path, the career progression goals might differ greatly. Individuals following the management path might strive to climb to the very top of the hierarchy; contrarily, the IC progression normally involves becoming a subject matter expert, leading complex technical projects, and possibly moving into senior technical roles, like principal engineer. Nevertheless, a relatively new literature stream suggests that those narrowly portrait groups cannot withstand the rapidly changing business environment and that nowadays a technical career progression often leads to a point where leadership competencies gain relevance (Joseph et al., 2012; Maurer & London, 2018). The hierarchical levels added for the SW industry present a more nuanced IC career path, with hierarchical levels that reach the middle management classification. Nevertheless, this categorization does not necessarily indicate management responsibilities. The code does not allow to show the progression of the two paths next to each other; rather, the levels resemble different hierarchical levels.

The career progression in the software industry often includes a promotional ladder with up to eight levels. While some organizations choose to represent those promotions as numbered levels (e.g., I – VI), others use titles. When focusing on those job titles, a progression might look like this:

- Junior Engineer, Tier I
- Engineer, Tier II
- Senior Engineer, Tier III
- Staff Engineer, Tier IV

- Lead Engineer, Tier V
- Principal Engineer, Tier VI
- Distinguished
- Fellow

To identify the categorical hierarchy for the engineering titles, the webpage [levels.fyi](http://www.levels.fyi) (www.levels.fyi) was especially helpful to compare the title structure of several organizations in the industry. The hierarchical levels in the SW industry are visualized in Figure 7.

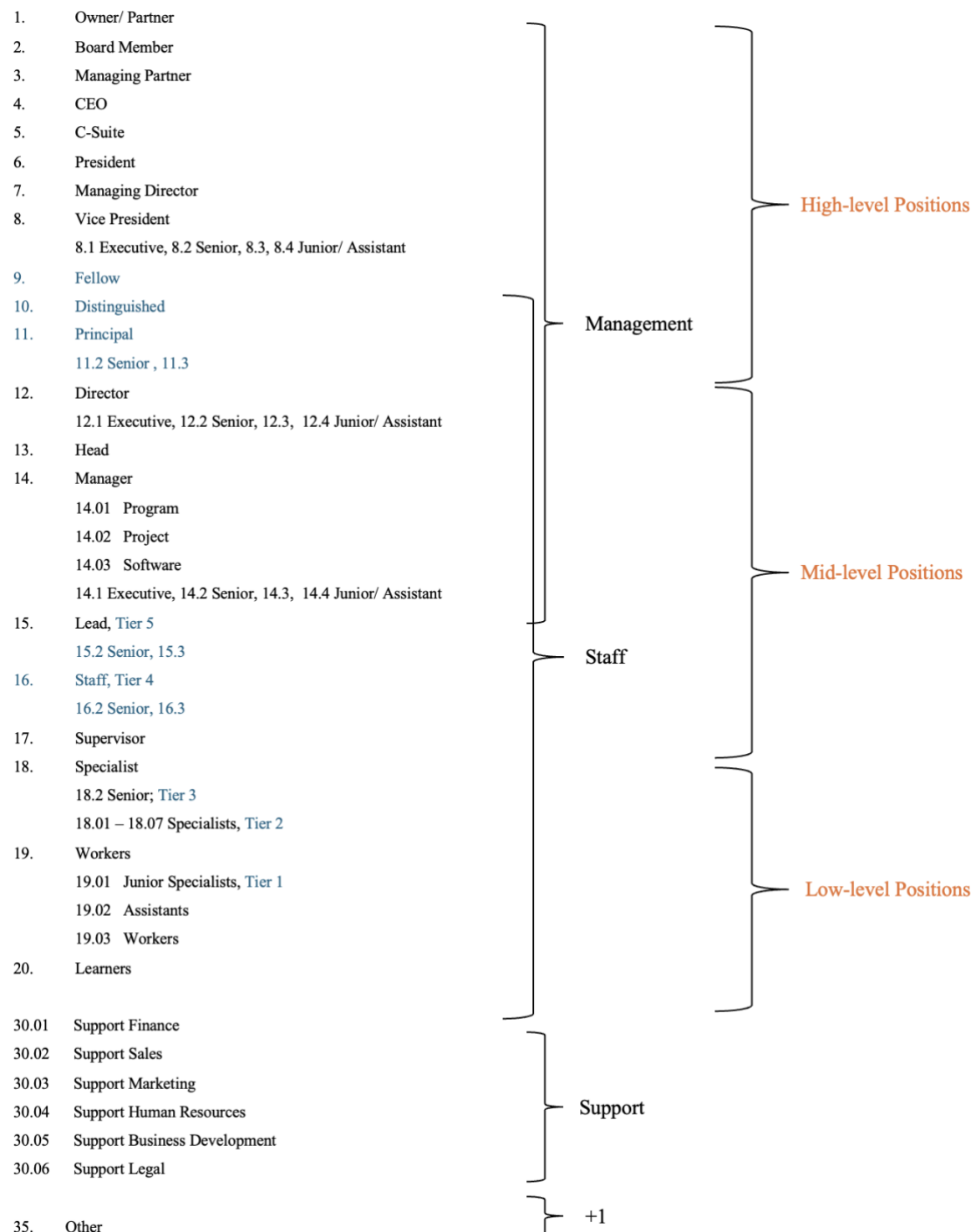


Figure 7 Hierarchical depth in the SW industry

Starting at the bottom of the proposed hierarchy, the first category different from the FS industry is the *Workers* (19) category. The horizontal differentiation is not as detailed as for the FS industry. This is explained by the fact that the SOC manual has only made significant changes and additions to the main group (here: *Specialists*) with the current edition from 2018, but the information on the workers is still limited. The *Workers* category is therefore

limited to the following classifications: *Assistants (19.03)*, unspecified worker titles (*Workers 19.02*) and *Junior Specialists (19.01)*. Junior specialists describe entry positions and are often called “Tier 1”.

Congruently the next upper category are the *Specialists (18)*, individuals in this category often use the additional title Tier 2. The horizontal differentiation follows basically the same structure as the third level of the SOC manual, except the category *18.05 Software and Web Developers, Programmers and Testers*. This category is more detailed and structured down to the fourth level. The explanation lies within the number of identified job titles and the seen relevance of those occupations. Comparably, the whole category *Specialists* contains 272 strings with 106 strings just in *18.05*. Similar to the FS industry, the senior specialists are extracted from all horizontal categories and summed in the category *18.2*, which also resembles the Tier 3 level. The differentiation can be seen in Figure 8, the matching SOC titles can be drawn from Appendix 5.

18.2 Senior Specialist, Tier III

- 18.01 Computer and Information Analysts
- 18.02 Computer and Information Research Scientists
- 18.03 Computer Support Specialists
- 18.04 Database and Network Administrators and Architects
- 18.05 Software and Web Developers, Programmers, and Testers
 - 18.051 Computer Programmers
 - 18.052 Software Developers
 - 18.053 Software Quality Assurance Analysts and Testers
 - 18.054 Web Developers
 - 18.055 Web and Digital Interface Designers
- 18.06 Consultants
- 18.07 Others, Tier II

Figure 8 Horizontal differentiation specialists, SW industry

The category *Staff (16)* is the initial step on the ladder of an IC, positioned just above *Senior Specialists*. In addition to strong technical abilities, essential leadership skills such as critical thinking, judgment, listening, empathy, and communication are vital at this level of seniority. They support the organization by providing technical context and direction, defining technical specifications, and documenting processes. Their reporting structure may differ

across organizations, but they generally report to a manager while maintaining control over their day-to-day activities (Valette, 2021).

The category *Lead* (15), also Tier V, was already included in the FS industry, but here the term often describes a more technical role, for example, a lead engineer. This position often includes maintaining and optimizing technical aspects of projects to enable positive project completion (Indeed Editorial Team, 2024b).

Principal engineers (*Principal 11*) possess a strategic understanding of organizational priorities and make technical decisions to address business challenges. They serve as the link between the engineering work on the ground and senior executives. In general, they operate with a high degree of autonomy but must collaborate closely with management to ensure alignment with business objectives. Although they do not hold formal managerial authority, they lead through influence and strategic advice.

The title distinguished engineer (*Distinguished 10*) is a prestigious designation awarded to individuals with exceptional technical achievements, often just below the rank of fellow. This role is typically found in larger organizations and involves shaping the strategic direction of the organization, creating strategic roadmaps, and achieving business goals. Distinguished engineers have the autonomy to develop their own areas of expertise, which strengthens the organization. They work closely with other stakeholders and executive teams to maintain alignment and avoid friction. Their role also involves significant mentorship and influence, necessitating excellent interpersonal skills (Valette, 2021).

The title *Fellow* (9) is rarely awarded. Fellows play a crucial role in leading organizational strategy. They must possess a comprehensive understanding of the organization's vision and long-term goals, ensuring alignment across technical teams. In addition to their technical acumen, fellows often represent the organization externally, acting as ambassadors. They facilitate communication between technical teams and non-technical executives, requiring a blend of exceptional technical expertise and managerial capabilities (Chapman, 2023). The categories *Principal*, *Lead*, and *Staff* differentiate additionally between senior and non-senior positions.

4.3.2 matching algorithms

The previous chapter explains the general setup of the categories, leaving the question open on how exactly the core job titles are matched with the categories. Following the matching

order in the code, the first matching logic that is explained regards the support function. To identify the job titles belonging to the support functions, all words from rules 30.1 to 30.6 and the abbreviations of the C-Suite are summed up in one comprehensive list that serves as an identifier. A function now screens all data in the column *experience title* for matches with the words contained in the list; this is the essential part of the matching. Each keyword in the list is compared with those in the *experience title* column using the *Fuzz Partial Ratio*-method from the *fuzzywuzzy* library in order to calculate a match score. This method assesses the similarity between two strings, providing a percentage score based on partial matches. The calculation is based on the Levenshtein distance, which means it identifies the closest substring that maximizes the match score between part of one string and the entire other string (SeatGeek). The function used here specifically looks for a perfect match score (100), indicating that the keyword is a substring of the entry in *experience title*. When a perfect match is identified and the length of the keyword is less than or equal to the length of the string in *experience title*, the function iterates through the rules specifying the hierarchical category. If the keyword is found in any set of the predefined keywords within the rules, the corresponding category is returned and saved to a new column called *support functions*. If no match is found, the function default is returned; in this case, the default entry is equal to 0. Creating a new column allows the rows with entries different from zero to work as a dummy to identify the support functions. Especially the job titles in the management group are overly generic throughout the core and support functions; thus, to get a more detailed picture of the industries, all observations, including those identified in the previous step, are scanned in the next steps. The categorization will be different in some cases because of the comprehensive list of support function indicators used in this step. If the job titles in the *experience title* column are not matched to the specified words, other matching opportunities arise.

The job title matching regarding the core functions was done in four steps: 1) exact matching, 2) partial ratio matching, 3) geographical matching, and 4) adjustments. The code for all steps can be retraced in Appendix 6 (FS) and 7 (SW). As already mentioned, job titles are considerably short, and abbreviations often consist of two or three letters. In turn, what might be obvious in an occupational context for humans, for example, that COO is equal to chief operations officer but not equal to coordinator, is not always as easy to implement while coding. To avoid that kind of miscategorization, word boundaries were used. A word boundary delineates the start and end of words within strings. In regular expressions, they

are denoted by the “\b” metacharacter. This metacharacter functions as an anchor, matching positions where a word character is immediately followed or preceded by a non-word character. A word character typically includes alphanumeric characters (letters and digits) and underscores, whereas non-word characters are any characters that do not fit these categories, such as spaces or punctuation (Python Software Foundation, 2024). LinkedIn does not specify a structure or set a framework for how job titles are supposed to be referenced, so the variety is significantly high. Bekkerman and Gavish (2011), for example, identified approximately 40.000 different versions of the job title “software engineer” when analyzing LinkedIn data. Moreover, to avoid listing all possible word combinations, it can be easier to use word character representations. One example is the “senior manager” title. A specification about the department would lead the rule-based algorithm introduced here to not match titles like “senior financial manager” or “senior customer manager”. To avoid listing a great variety of word combinations the expressions “\w+” and “\s+” were used. The expression “\w+” matches one or more word characters of undefined length, and “\s+” ensures that there is at least one whitespace character separating the words in the pattern (Python Software Foundation, 2024). The introduction of word boundaries led to the first matching step:

1) The exact matching

As explained above, word boundaries lead to an avoidance of matching abbreviations or short job titles with words where those terms are just a component. Nevertheless, fuzzy partial ratio matching, as explained for the support functions, calculates the similarity of strings, which might lead to the recognition of a match that is not supposed to be one. On the other hand, more than one match can be found, due to a typical order of letters, like “md”. To reduce the chance of abbreviated or single-word job titles being miscategorized, the rule *rules_exact*, which follows the same hierarchical pattern but only contains abbreviations or stand-alone terms that were recognized as miscategorized in the test runs, is introduced.

The matching steps follow the same logic as with the support functions, but the matching algorithm is different: 1) the identified words are compromised in a list; 2) the words in the list are then checked for exact matches with words in the column *experience title*. This time the strings must be the exact same (length, letters, order); 3) if an exact match is found, the function iterates through *rules_exact* and categorizes the found match; 4) observations where

no match is found level 35 (*default*) is returned. All results are saved to the new column *exact level*.

2) Partial ratio matching

The partial ratio matching accounts for most of the matches being found. Figure 9 helps to understand the steps to reach a match.

```
new_df = category_35[category_35['exact_level'] == 35].copy()

def categorize_with_partial_ratio(experience_title, rules):
    best_match = (35, 0)
    # Iterate through the rules except the default
    for level, keywords in reversed(rules[:-1]): #reversed filter - double titles cat to highest
        for keyword in keywords:
            if len(experience_title) >= len(keyword):
                #Calculate the match score using partial_ratio
                match_score = fuzz.partial_ratio(keyword.lower(), experience_title.lower())
                # If a better match is found, update the best_match tuple
                if match_score == 100:
                    best_match = (level, match_score)
    return best_match[0]
```

Figure 9 Partial ratio matching FS industry

Before starting the partial ratio matching, a new data frame, *new_df* is created, which just includes the observations that could not be categorized before, thus being categorized to the default value, *level 35*. The *categorize_with_partial_ratio* function efficiently categorizes job titles by evaluating the similarity between an observation in *experience_title* and the predefined categories in *rules*, which consist of a set of predefined keywords associated with the hierarchical levels. The function initializes a *best_match* tuple to (35, 0), with 35 as a default level in the absence of matches, and 0 indicating no match yet detected. The core process involves iterating over the *rules* list in reverse order, excluding the *default category 35*. This reverse iteration ensures that in cases where multiple matches yield the same score, the highest level in the hierarchy is selected, for example, “senior manager” instead of “manager”. This also applies to individuals who state two job titles, like “vp, banking manager”, in this case, the individual will be categorized under *Vice President*. Within this loop, the function only considers keywords if the length of *experience_title* equals or exceeds (“>=”) that of the keyword. This guarantees that the keywords set a minimum of given information but allows individual job titles, like “banking manager at worldbank”, to be identified. It then computes the match score using *fuzz.partial_ratio*, which compares the

keyword and the job title after converting both to lowercase, thereby assessing how closely the keyword matches a substring of the job title. A more detailed explanation of the partial ratio method was explained before in the support function context. If this fuzzy matching process results with a perfect score of 100, indicating an exact substring match, the *best_match* tuple is updated with the current level. Achieving a 100 percent match score with fuzzy matching essentially means that the job title contains the keyword exactly as it appears in the list, embedded within the larger string of the title. Thus, it can be seen as a second reassurance of accurate categorization next to the “>=” condition. Finally, the function returns the hierarchical level, as a component of *best_match*.

3) Geographical matching

The *find_geo_division* function iterates through a list of tuples, *rules_g*, where each tuple contains a geographical classification (denoted as letters “a” to “e”) and a list of associated keywords. To verify that a keyword appears within the title, the function examines each observation in the *experience_title* column. If a match is found, the function returns the corresponding letter, representing the geographical classification of the title. In cases where no match is found, the function returns *none*. The results are stored in a new column called *geo_division*, which represents the geographical classification of each observation based on the defined rules.

4) Adjustments

During the test runs, it was noted that despite the steps undertaken, there are occasional outliers in the vice president (VP) titles. The outlier rate is very low but easy to adjust and was therefore taken into account. They are mostly found in LinkedIn titles that have two or more job titles in one field, which may be due to the length of the VP title. Additionally, in the matching process, a recategorization to the higher-level *President* might happen, as there is also a 100 percent match here and the filter scans bottom up. To correct possible mismatches, a pattern was defined with all possible VP title forms. This pattern is then used to rescan all relevant observations that have already been categorized—relevant are all categories below and the category *President*. In the event of a match, the observations are consequently re-categorized to the category *Vice President*.

In the management group, the assistant and junior categories (x.4) can be recategorized to the category above (x.3) because of the bottom-up approach of the partial ratio filter. The

correcting procedure is identical to the one described above. This also applies to the categories *Principal*, *Lead*, and *Staff* in the SW industry.

In the SW industry, hierarchical job progression often follows a numerical logic, especially for technical job titles. The level *Senior Specialist (18.2)* is equal to e.g., software engineer III; in addition to the senior titles, level three indicators are scanned for and included in the category. Similar to the process of filtering for junior titles within the management group, a filter that screens level one (tier 1) indicators is applied to the *Specialist* category and recategorizes possible miscategorizations to the *Junior Specialist* level (19.01).

As already indicated, the *Specialist* category has a combined senior level, this also applies to the *Expert* category. Summarized means that all horizontal subdivisions are displayed at one senior level. The procedure is again equivalent to the previous one: a senior pattern is created, all horizontal groups are examined, and matches are placed in the appropriate category, e.g., in the FS industry, the horizontally differentiated senior specialists categorized within 14.01 to 14.09 will be recategorized to 14.2.

4.4 Validation

The presented structures in the previous chapter reflect the final categories for both industries; the process to validate this categorization included expert interviews and the analysis of salary data. Both validation steps were crucial for some fundamental changes and corrections regarding the hierarchical order.

4.4.1 Interviews

A total of eight interviews were conducted, four for each industry, with employees from the respective industry. It should be noted that although data from the United States was used, the organizations interviewed are based in Austria and Germany. For a better flow of conversation, the respondents were given the option of conducting the interviews in German or English, with the indication that English would be preferred. Nonetheless, four of the eight participants opted for German. All transcripts in English can be found in Appendix 12. The interviews helped to clarify whether the assumptions could withstand in practice.

The interviews were based on an interview guideline (Appendix 8 (EN) and 9 (GER)). The first part comprises a brief introduction, a characterization of the organization (sector, name, size), and questions about the interviewee's personal career. The second part was supported by a PowerPoint presentation, which served as an additional guideline (Appendix 10). The aim of the second part was to depict the own organizational structure as accurately as possible in a “work-together” process. The presentations included exemplary organizational charts, which illustrated the aim of the interview; a portrayal of the overlapping career paths, which served to question possible parallel structures; and finally, an illustration of the categorization of this thesis at that time. Alongside questions about the respondents' organization, questions related to whether they were familiar with such structures from other organizations or had already gained experience in other organizations. The information content of the interviews varied greatly based on the data available to the employees. If the employees had access to an organizational chart, more information could be obtained, and the validity of the statements increased. Interviewee 7, for example, a software engineer working at SAP, was able to share the organizational chart but only for his track, which limited the amount of information received (Appendix 12, Interview 7 – SW industry). Some interviewees looked in their email histories for signatures to extract specific titles. If no access to an organizational chart or other indicating tools were given, the information was limited. This led to two implications: Firstly, understanding organizational hierarchy is highly complex, and knowledge of employees is limited. Interviewee 3, who works at American Express Austria, stated:

“These are of course internal company notations that may not be so clear to outsiders.” (00:02:04)

Interviewee 3, was also elaborating on the difference between hierarchical differentiation and promotions, he explained:

“We have these band levels. So, every person or every position has a band level, and in Austria, you normally start at band level 25, and then it goes up and mostly, as far as I know, in increments of five [...]. And what I can't quite say yet is whether the title, whether it will change.” (Appendix 12, Interview 3 – FS industry, 00:05:23).

Secondly, the initial planned number of conducted interviews had to be increased to gain more information. After conducting all the interviews, valuable knowledge could be

extracted. Implemented alterations are explained in the next section, the initial categorization can be seen in presentation slides five and six (Appendix 10). Starting at the top of the hierarchy, the first interview with ZEB led to a restructure of the second category (former) *Associate Partner* (Appendix 12, Interview 1 – FS industry). The new structure focuses on managing and leading partners (3); they have similar tasks as a CEO but commonly own shares.

The category *Managing Director* raises doubts on the right placement, while N26 (Interview 5 – SW industry) handles their C-Suite and engineering director on the same hierarchical level, they also assign the titles managing director and chief risk officer to the same individual. American Express contrarily places the managing director right under the CEO and the C-Suite beneath the managing director–information was drawn from the internal organizational chart (Appendix 12, Interview 3 – FS industry). Double checking the organization’s website, no information about a managing director is indicated; instead, the title president is identified (American Express, 2024). Additional sources were therefore analyzed in order to draw a clear conclusion. As indicated in the previous chapter, the final order is *CEO, C-Suite, President, Managing Partner*.

Interviewees 4 (Kreditgeschäft Kreissparkassen), 5 (N26), and 6 (Walter Bösch GMBH & Co. KG) indicated that the director is hierarchically above the head; congruent with other sources, the alteration was implemented. Interviewee 5 helped further identify the level of *Principal* in the SW industry.

Additionally, the interviews helped to determine an overarching geographical differentiation for all categories, especially respondents 3 (American Express) and 4. Before the interviews, this was only intended for the *Manager* category.

The interview with interviewee 7 (SAP) indicated the implementation of a category for data architects above the *Manager* category (Appendix 12). The title architect is deeply rooted in the *Specialist SW* category, which is based on the SOC (U.S. Bureau of Labor Statistics, 2018b). Additional web research suggests that a data architect is on the same hierarchical level as a data engineer but fulfills different tasks (Indeed Editorial Team, 2024a). Therefore, a new category was not introduced. The interviews did not verify the usage of the technical levels in the SW industry, i.e., Fellow, Distinguished, and Staff. Nevertheless, evidence for the usage and order of those titles is explicit.

The interviews showed that the depth of hierarchy in smaller organizations is rather straightforward and more easily identifiable. They also showed that some titles, such as manager, head of, director, and CEO tend to be clearer and more frequent. In essence, the categories presented were recognized by the interviewees, and no further changes were suggested. In fact, the proposed hierarchical arrangement was largely accepted and found to be coherent. Further changes are accordingly attributed to the salary data and further research.

4.4.2 Salary data

As a second validation measurement, salary data from the page H1B was scraped. H1B provides almost 5 million indexed records containing yearly salary data from 2013 to 2023 (h1bdata.info, 2024). Research indicates that there is a proportional difference in compensation levels between adjacent hierarchical positions within an organization. The relationship between hierarchical levels and corresponding salaries is widely recognized and commonly expected in formal organizations, as both elements signify status and value to some extent (Mahoney, 1979). Based on this assumption, a downward curve along the categories is expected.

For a better comparison, the salary data was standardized. Before standardizing the data, observations with a lower yearly salary than \$10,000 were dropped. In addition, observations that were identified as support functions in the first categorization step are also excluded from the calculations. To standardize the salary data, the following methodological steps were applied: Firstly, the mean salary was calculated for the dataset. This involved computing the average of all the salary values present in the *salary* column. Subsequently, the standard deviation of the salary data was determined, providing a measure of the dispersion or variability of the salary values around the mean. Once the mean and standard deviation were obtained, the standardization of each salary value was calculated with:

$$z = \frac{x - \mu}{\sigma}$$

Where z represents the standardized value, x denotes the original salary value, μ is the mean salary, and σ is the standard deviation of the salary data. The standardized salary values were stored in a new column, *standardized_salary*. Standardization techniques ensure that the salary data is normalized, which simplifies meaningful comparisons and further statistical analyses.

It should be enhanced that the number of observations per category differs greatly, which might lead to potential impacts on the results, due to observation bias. Categories with larger sample sizes, such as category 18 (*Specialists*) in the SW industry with over 400,000 observations, exhibit higher statistical power, due to the law of large numbers, which states that as a sample size grows, its mean gets closer to the average of the whole population. In contrast, categories with smaller sample sizes, such as category 6 (*Director, SW*) with 78 observations, have less statistical validity, since they are more sensitive to the effects of outliers or random variations. The presence of these major differences can result in larger categories dominating the analysis. For example, if the large category tends to exhibit higher or lower salaries, this could disproportionately affect the standardized salary data. Conversely, categories with fewer observations may not adequately represent the actual salary distribution in that category. This lack of representativeness can introduce bias and reduce the generalizability of the results (Wolf et al., 2013).

4.4.2.1 Salary FS industry

The data for the FS industry is distributed as follows, most observations, 34%, belong to the *Specialist (14)* category, 29% belong to the *Support (30)* category and will not be analyzed, 20% belong to the *Manager (11)* category, and 8% of the data belong to uncategorized personnel, category 35, thus being also excluded from the analysis. The support category is excluded because of a less detailed categorization system. Furthermore, it is expected that other specialists field show varying salaries which lead to potential misleading of the results. The categories *Owner/Partner (1)*, *Head (10)*, *C-Suite (5)*, *Learners (16)*, *Managing Director (7)*, *CEO (4)*, and *Managing Partner (3)* are underrepresented with less than 500 observations. Conclusions drawn from those categories may lack robustness due to the limited sample size, which can lead to less reliable estimates. Data for category 2 (*Board Members*) is completely missing. In total, 195,821 observations qualified to be included in the analysis (cf. Appendix 17.1).

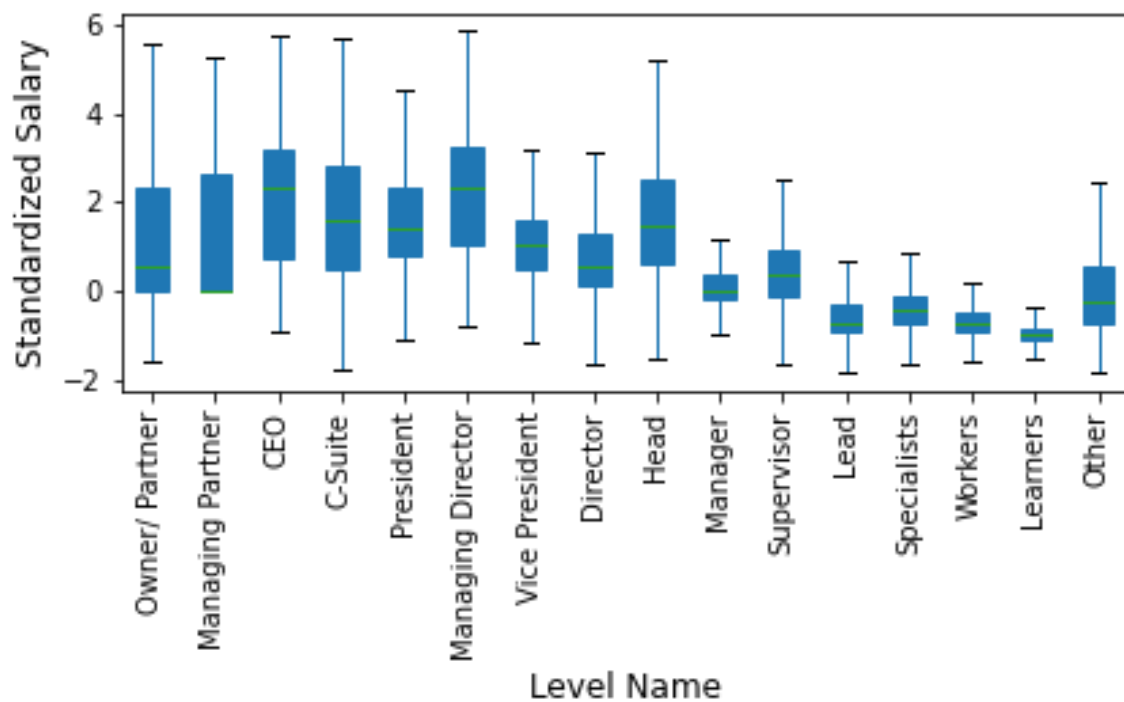


Figure 10 Salary distribution FS industry

The provided box plot visualizes how salary distributions vary across the hierarchical levels, reflecting differences in compensation structures within the organizations. The code for the calculation can be extracted from Appendix 17.2.

High-level positions, including *Owner/Partner* (1), *Managing Partner* (3), *CEO* (4), *C-Suite* (5), *President* (6), *Managing Director* (7), and *Vice President* (8), exhibit significantly higher standardized salaries. These positions show substantial variability in compensation, which might reflect diverse responsibilities, strategic importance, and varying organizational sizes. Notably, *Managing Partner* (3) stands out with lower median salaries compared to some other top executive roles. This might be due to equity ownership or other compensation agreements. *Managing Director* (7) on the other hand, shows a higher mean salary, which might be due to the interchangeable usage of managing director and CEO in some organizations (Bean-Mellinger, 2019). Overall, high-level positions are characterized by their critical roles in leadership and decision-making, commanding higher pay to reflect their influence and accountability within the organization (Mahoney, 1979).

Mid-level positions, such as *Director* (9), *Head* (10), *Manager* (11), and *Supervisor* (12), demonstrate moderate to high standardized salaries. *Directors* (9) and *Heads* (10) have

relatively higher median salaries compared to *Managers (11)* and *Supervisors (12)*, indicating their higher managerial and strategic roles. However, the category *Heads (10)* show higher median salaries than *Directors (9)* and *Vice Presidents (8)*, which may be due to different scopes of responsibility or organizational priorities. Nevertheless, following the conducted interviews, task descriptions, and due to the limited number of observations for this category, namely 429, the hierarchical order is kept unchanged. Managers and supervisors have more consistent but lower salaries, reflecting their operational focus and narrower range of responsibilities compared to higher-level positions. *Managers (11)* have a median standardized salary around 0.5, indicating moderate compensation, with less variability compared to higher-level positions. *Supervisors (12)* show more variability than the neighboring categories. The small variability in these categories, suggest more uniform salaries. Mid-level roles play a crucial part in bridging the strategic directives of high-level positions with the operational execution of lower-level roles (Colombo & Grilli, 2013).

Lower-level and support positions, including *Lead (13)*, *Specialist (14)*, *Workers (15)*, and *Learners (16)*, generally have lower standardized salaries with more consistency. *Specialists (14)* show slightly higher median salaries compared to *Workers (15)* and *Learners (16)*, reflecting their specialized skills and experience. The comparatively low salary of *Leads (13)* might be due to a more frequent use in other industries. *Learners (16)*, as expected, have the lowest standardized salaries, significantly below 0, reflecting their entry-level status and lower compensation. These positions are essential for executing day-to-day operations and supporting the broader organizational goals set by higher and mid-level positions. The relatively uniform compensation within these roles highlights the more standardized nature of pay structures for operational and support tasks (U.S. Bureau of Labor Statistics, 2018b).

If the seniority differentiation is included, a clear difference between the salaries depending on the seniority title can be seen. These differences in seniority might explain the variability in the salary data for the respective categories. The category *Director (9)*, for example, includes almost 10,000 observations with information for all four forms:

Executive Director	193,365 \$/year
Senior Director	187,910 \$/year

Director	137,255 \$/year
Junior/ Assistant Director	101,468 \$/year

Secondly, the salary structure is tested to follow an increasing pattern for larger organizations. Empirical research repeatedly shows that larger organizations, i.e. those with more than 500 employees, pay higher wages than smaller ones, a phenomenon known as the size wage premium, whereby wage disparities are sometimes as high as 35%. This salary discrepancy can be attributed to the fact that larger companies seek higher quality staff, employ efficient pay practices and take advantage of economies of scale that lead to higher productivity and allow them to offer better compensation and job security (Oi & Idson, 1999). To sort the salary data by organizational size, the data was merged with another Crunchbase dataset. The data cleaning steps, and the merging steps conducted can be comprehended in Appendix 17.3. Summarized, the strings in the column *company* were cleaned to enhance the merging process, and observations missing data about the number of employees were dropped. The merged dataset resulted in a final data frame with 12,797 relevant observations. The significantly reduced number of observations reduces the statistical power of the interpretation of the results and therefore should be treated with caution.

Table 2 illustrates the standardized mean salary data across various hierarchical levels, segmented by the number of employees in different size categories. The standardized salaries are categorized into intervals based on the legend provided, ranging from high positive values (>2) to low negative values (<-1).

main categories	1 - 10	11 - 050	51 - 100	101 - 250	251 - 500	501 - 1000	1001 - 5000	5001 - 10000	10001 - max
1		-0,37	0,26	0,07	0,48	-0,04	0,12		0,26
4		0,91	1,73			-0,65			2,13
5	1,27	0,76		2,06	2,06	-0,62			-0,20
6	2,22	0,72		0,53	0,56	1,02	1,85	2,17	0,76
7					2,32		1,11		4,84
8	0,47	0,16	0,55	0,28	0,20	0,17	1,18	1,98	0,65
9	0,30	0,54	0,91	0,29	0,78	0,60	1,53	0,75	0,57
10	0,60		0,50	0,97	0,97				0,26
11	0,06	-0,20	-0,21	-0,16	0,23	-0,23	0,26	0,21	0,28
12	-0,03	-0,12	0,39	0,13	1,07	0,20	0,56	-0,15	0,30
13		-0,62	-1,09	0,14	-0,70	-0,03	-0,09	-0,62	-0,42
14	-0,47	-0,56	-0,50	-0,14	-0,26	0,19	-0,06	-0,30	-0,39
15	-0,33	-1,04	-1,14	-0,84	-0,54	-0,13	-0,70	-0,84	-0,49
16									-0,79
Color Codes	Standardized salary intervals								
	> 2								
	1 bis 2								
	0,5 bis 1								
	-0,5 bis 0,5								
	-1 bis -0,5								
	< -1								

Table 2 Mean salary FS grouped by number of employees

Higher-level positions generally exhibit higher standardized salaries with significant peaks in certain employee size categories, particularly for *CEO* (4: 2.13 in 5,001–10,000 employees) and *C-Suite* (5: 2.06 in 101–250 and 251–500). The salary for *Managing Director* (7) reflects the highest deviation in the biggest organizational group, but all calculations for this category are based on eight observations, thus the interpretative validity is low (4.84 in 10001–max employees). This reflects the high compensation expected for strategic and executive roles in larger organizations. Nevertheless, the number of observations might play a crucial role in evaluating the robustness of the data for each hierarchical level and is especially low for the upper-management categories (<100). The mid-level positions show mixed results, with some categories like *Director* (9) exhibiting moderately positive salaries in certain ranges, but generally lower than high-level positions. The category *Manager* (11) shows predominantly negative standardized salaries. Lower-level positions exhibit lower standardized salaries, especially in smaller organizations. The *Workers* (15) category shows the lowest paying observations in small organizations.

There is a visible trend that larger organizations tend to pay better, especially for high-level positions such as *CEO* (4), *C-Suite* (5), *President* (6), and *Managing Director* (7) as evidenced by the higher standardized salaries in larger employee size categories. For mid-level positions, the trend is less consistent but still present in certain roles like *Specialists* (14) and *Supervisors* (13). The trend is less evident for lower-level positions: although there are some positive salary trends in larger companies, overall compensation is lower.

4.4.2.2 Salary SW industry

H1B provided 520,340 observations for 20,187 organizations in the SW industry. Looking at the distribution of the observations for the SW industry, 80.23% of all observations are in category 18 (*Specialists*). Further 14% are in the categories *Manager* (14), *Lead* (15), and *Workers* (19), which leads to the conclusion that those categories are well represented. In descending order, the categories *Principal* (11), *Owner/ Partner* (1), *C-Suite* (5), *Head* (13), *CEO* (4), *Managing Director* (7), *President* (6), *Distinguished* (10), *Learners* (20), and *Managing Partner* (3) include all less than 1000 observations, with category 3 (*Managing Partner*) only including 5 observations. No observations for categories 2 (*Board Members*) and 9 (*Fellow*) could be captured.

The provided boxplot in Figure 11 visualizes the standardized salaries across the main categories (without horizontal, geographical, and seniority differentiation).

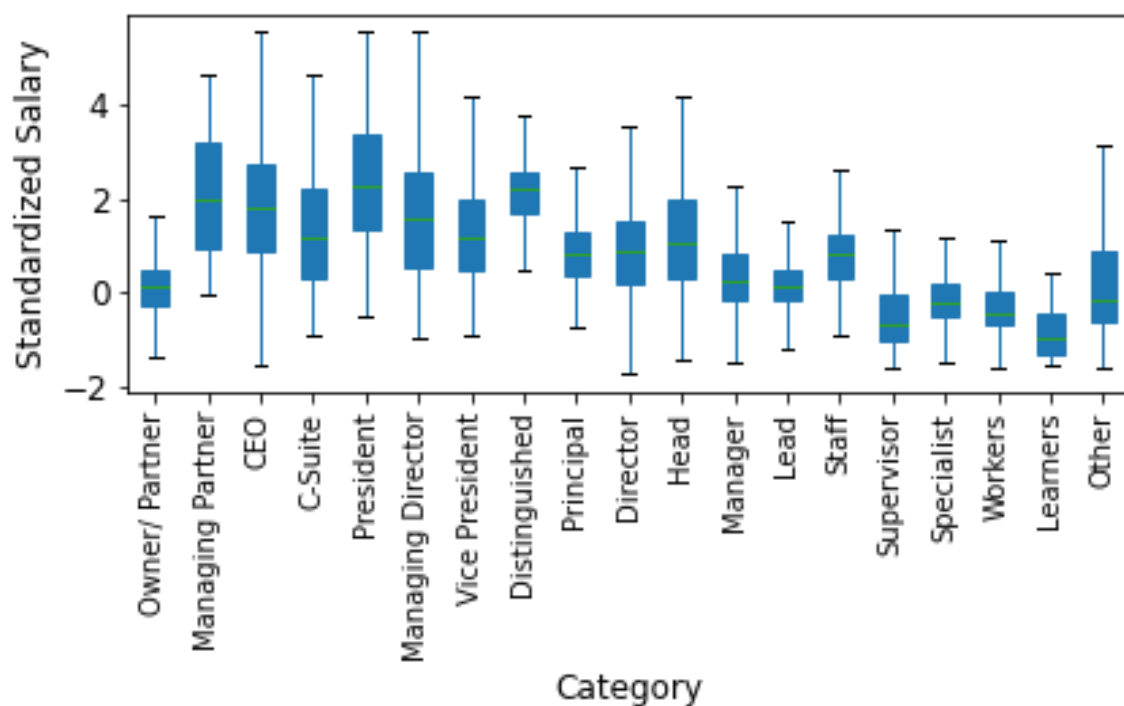


Figure 11 Boxplot mean salary per category SW industry

Similar to the FS industry, high-level positions, such as, *Managing Partner* (3), *CEO* (4), *C-Suite* (5), and *President* (6), demonstrate higher median standardized salaries with significant variability. The *Owner/Partner* (1) category shows a smaller range with some very high outliers, indicating substantial differences in compensation within this group. This might be

due to different organizational sizes or profit-sharing-based compensation (Kopp, 2024; Oi & Idson, 1999). Similarly, *Managing Director* (7) and *Vice President* (8) positions also have relatively high median salaries, although with less variability compared to the top executive roles. It should be noted that all categories until *President* (6) belong to those with a low number of observations. The category *Distinguished* (10) shows high median salaries, representing seniority.

Mid-level positions, including *Principal* (11), *Director* (12), and *Head* (13) roles, have median standardized salaries closer to the overall mean, with moderate variability. This suggests more uniformity in compensation compared to higher executive positions. *Manager* (14) and *Lead* (15) positions show narrower salary distributions with lower medians than the higher executive roles, reflecting their place in the organizational hierarchy. The category *Staff* (16) has a surprisingly high mean salary, despite a respective number of observations. Nevertheless, the hierarchical order is confirmed by several sources mentioned before. Contrarily, the *Supervisors* (17) category shows low salaries but high variability.

Lower-level and support positions generally have lower median standardized salaries and less variability, aligning with the expected remuneration for these roles. *Specialists* (18), *Workers* (19) and *Learners* (20) exhibit lower median salaries, fitting within the anticipated structure.

Similar to the FS industry, a clear difference between the seniority titles can be detected. For instance, the mean salary for the *Vice President* (8) category is as follows:

Executive VP	350,000 \$/year
Senior VP	232,716 \$/year
VP	192,125 \$/year
Junior/ Assistant VP	122,341 \$/year

To analyze the salary distribution based on organizational size, the merged Crunchbase dataset was used. After the merge, based on the *company title* columns and the drop of all rows lacking information on the number of employees, 38,121 observations remained. Resulting in a sufficient number of observations (>1.000) only for the categories *Lead* (15)

and *Specialists* (18). The standardized mean salary per category, grouped by the number of employees in an organization, can be seen in Table 3.

main_categories	1 - 10	11 - 050	51 - 100	101 - 250	251 - 500	501 - 1000	1001 - 5000	5001 - 10000	10001 - max
1	6,79	0,15	0,70	0,20	0,99	0,47	0,80		0,34
4	3,77	6,03	3,29	6,14		-0,90			
5	2,52	2,90	1,64	5,19			2,30		-0,36
6			-0,12				3,53	10,06	
7						-0,12			
8	-0,20	2,05	1,10	2,42	2,88	3,29	2,84	0,80	4,42
10					10,80				2,71
11	0,75	2,76	4,34	1,41	0,58		2,64		1,46
12	1,30	1,80	1,94	1,16	1,22	1,02	2,10	1,46	2,04
13	2,33	-0,50	1,89	1,04	1,70	0,00	0,64		2,61
14	-0,18	0,11	0,25	0,27	0,47	0,37	0,83	1,08	0,96
15	0,35	0,54	0,64	0,30	1,31	0,74	0,35	0,92	0,71
16		-0,76	-1,05	0,07	0,54	1,07	0,95		
17	-1,64	-0,35		-0,72		-0,70	-0,48		-0,38
18	-0,28	-0,24	-0,09	-0,05	0,09	-0,11	-0,13	0,08	0,34
19	-0,64	-0,52	-0,34	-0,46	0,00	-0,14	-0,42	0,47	-0,35
20			-1,38						-1,38

Color Codes	Standardized salary intervals
	> 2
	1 bis 2
	0,5 bis 1
	-0,5 bis 0,5
	-1 bis -0,5
	< -1

Table 3 Mean Salary SW grouped by number of employees

As the number of observations was significantly reduced, the explanatory power was further minimized. An interpretation of individual results, which can be taken from Table 3, should therefore be avoided. Nevertheless, the table highlights that salary data for higher management categories (categories 1 to 5) is rarely available for larger organizations (<250 employees). The categories *President* (6) and *Vice President* (8) tend to show an increase in salary with increasing organizational size; a similar trend can be seen for the category *Manager* (14). The *Supervisor* (17) category shows extremely low salaries in small organizations and among the lowest salaries for mid-size organizations. Lower categories show an inconsistent salary progression with a growing number of employees.

4.5 Results

Since this thesis works as a base to analyze the relationship between hierarchical depth and organizational performance, the results will provide an overview of possible evaluation variations, the distribution of the data, and the identified hierarchical levels in the industries. The code needed to extract the results can be found in Appendix 18.

The results can be viewed in four different ways. The first option includes all possible variations, including vertical, horizontal, geographical, and seniority differentiation. The second option only includes seniority differentiation; the third option adds the geographical component. The fourth option is similar to the base for the salary analysis and only represents the main categories (vertical differentiation). Further modifications of the code to vary the differentiation options are relatively easy to implement. The distribution of the data will be shown for the main categories, thus neglecting horizontal, geographical, and seniority differentiation.

4.5.1 Results for the Financial Services Industry

The final dataset for the FS industry encompassed 3,830,806 observations. This dataset included 55,937 unique organizations and 970,847 unique job titles.

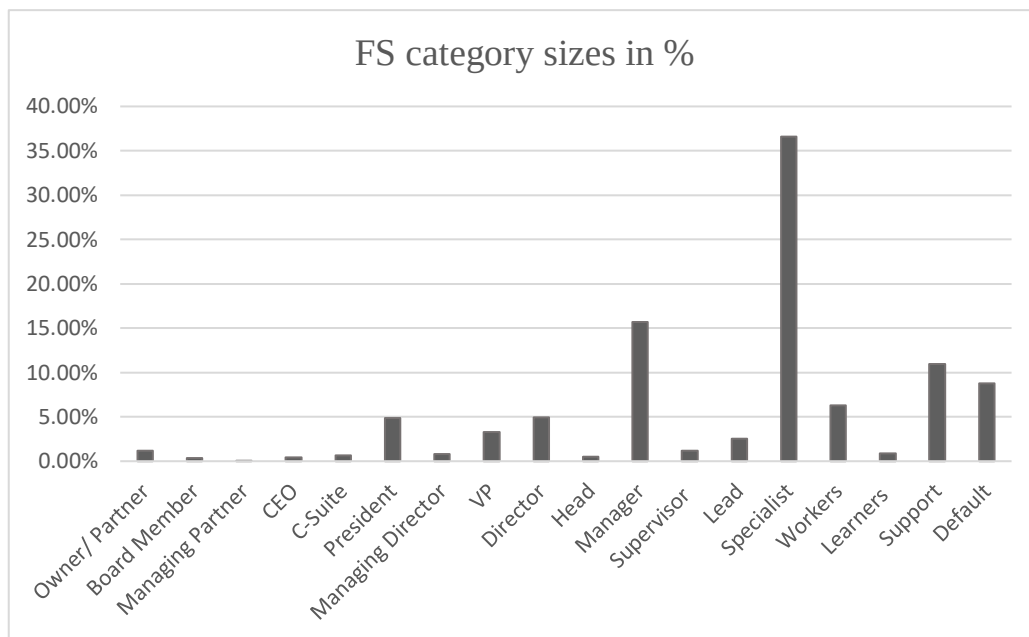


Figure 12 FS category sizes in percent

The distribution in Figure 12 reflects a pyramidal structure typical of many organizations, where the number of positions increases as one moves down the hierarchy. The heavy reliance on specialized skills and knowledge in the financial services industry is reflected in the high percentage of *Specialists* (36.61%). The second largest category reflects managers (15.73%) and the importance of middle management in coordinating and supervising the work of specialists and operational staff (Colombo & Grilli, 2013). Higher level positions, with each category contributing less than 1%, collectively represent a very small percentage of the overall distribution. The considerable percentage of *Support* roles reflects their vital

contribution to organizational efficiency. In total, the dataset contained 847,328 entries (22.12%) specifically related to support functions.

An analysis of the hierarchical depth reveals that the mean number of hierarchical levels within organizations is 5.5. The median number of levels is four, with a distribution in the range of the first and third quartiles from three to eight levels. A significant finding from the analysis is the identification of 466 outliers, each with a minimum of 16 hierarchical levels. These outliers indicate a high degree of organizational complexity. The presence of such outliers suggests that while the majority of organizations maintain a relatively flat structure, a minority exhibit deeply hierarchical configurations. Among the organizations with 18 identified hierarchical levels are American Express, Cisco, Deloitte, Epsilon, and Oliver Wyman; in total, 35 organizations encounter 18 levels. Nevertheless, as an increase in hierarchical depth is expected with a growing number of employees, a corresponding analysis was conducted. Consistently low hierarchical depth while the number of employees increases would, among other assumptions, indicate a centralized organizational structure. One finding of the analysis, which is important for the context of the results, indicates that the number of organizations decreases with an increasing number of employees; for example, the dataset is limited to 196 organizations with more than 10,000 employees. Comparably, the two smallest employee groups (up to 50 employees) inhibit more than 16,000 organizations (c.f. Appendix 18.5). To include the variable “number of employees”, the second dataset from Crunchbase was merged with the result files; the code can be seen in Appendix 18.4. The merging and cleaning of the data resulted in a total of 36,631 unique organizations being included in the analysis. Table 4 shows the hierarchical depth of the organizations grouped based on the number of employees. The x axis reflects the number of employees divided into groups; the y axis shows the first quartile, the median, the third quartile, and the mean.

num_employees	1 - 10	11 - 50	51 - 100	101 - 250	251 - 500	501 - 1000	1001 - 5000	5001 - 10000	10001 - max
q1	2	3	4	5	7	6,25	7	8	9
median	3	5	7	8	10	11	12	12	13
q3	4	7	10	11	13	14	15	16	17
mean	3,74	5,21	7,27	8,11	9,61	10,16	10,64	11,40	12,27

Table 4 Hierarchical depth by number of employees FS industry

The increase in hierarchical depth follows a perfectly increasing trend. Whereas small organizations with 1-10 employees have a median of 3 hierarchical levels, larger organizations with more than 10,000 employees have a median of 13 hierarchical levels.

This finding indicates that a growing number of employees lead organizations to increase their hierarchical depth. The flatter hierarchical structures in smaller organizations and their higher incidence might explain the low representation of higher management positions identified in Figure 12. Furthermore, this trend is not only evident in the median, but also in the quartiles, which reinforces its general validity. Only the lower quartile in the 501-1000 employee group deviates slightly from the surrounding values by 0.75.

4.5.2 Results for the Software Industry

The final dataset for the SW industry contains 4,898,656 observations, with 1,266,646 unique job titles and 59,465 unique organizations.

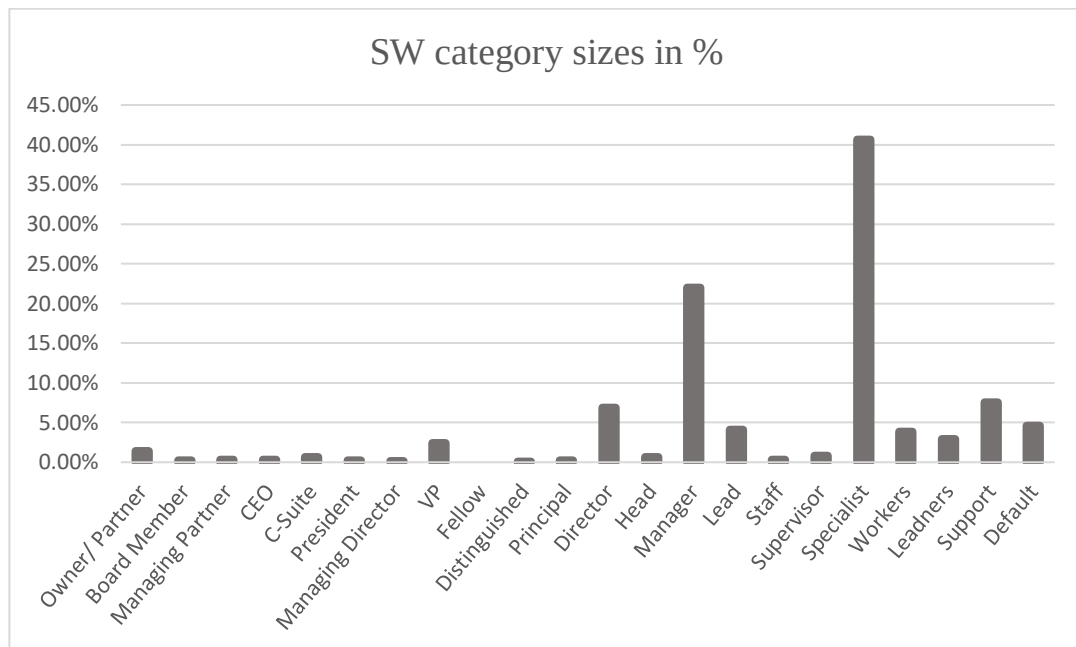


Figure 13 SW category sizes in percent

The distribution shown in Figure 13 suggests that some titles are awarded more frequently and that these categories therefore increase in size. A similar distribution could be seen in the FS industry, also resembled by the particularly low filling rates of the high-level positions. Also striking is the low filling of the technical positions, i.e., those positions that have been added in the SW industry categorization. The *Specialist* category is with 40,66% clearly the most densely represented, which is not surprising given the large number of job titles identified. The *Support* category only shows the uncategorized but identified observations; in total, 20.05% of the titles (982,181 entries) belong to the support functions. 4.46% of the data could not be categorized by the code.

To investigate the hierarchical depth of the organizations, those with observations only for one hierarchical level have been excluded. Regarding the main categories, 50% of the organizations show five hierarchical levels, with a mean of 5.82 levels. Organizations between the first and third quartile show between three and eight hierarchical levels. The eight biggest companies that have job titles categorized to 21 levels are CSC, Google, IBM, Microsoft, Motorola, Salesforce, Teradata, and Unisys. In total, 689 outliers with at least 16 hierarchical levels were identified. Since all categories, including support and default, would sum up to 22 levels, it is indicated that even the biggest organizations are not using all titles. Especially the category *Fellow* contains less than 0.05% of the observations, which might explain the limitation to 21 categories as a maximum. Nevertheless, this low representation fits the description of industry insiders about the distinction in this title (Chapman, 2023).

For a conclusive analysis, the second Crunchbase dataset, which contains the number of employees, was used again. Organizations with information about only one hierarchical level got excluded, leading to a drop of around 13,000 organizations. The interpretation in this section is based upon 43,445 observations, which represent the number of unique organizations. Table 5 shows the boxplot data for the hierarchical depth of the main categories classified by the number of employees.

num_employees	1 - 10	11 - 050	51 - 100	101 - 250	251 - 500	501 - 1000	1001 - 5000	5001 - 10000	10001 - max
q1	2	3	4	5	6	7	8	7	7
median	3	5	7	8	10	11	13	12	11
q3	5	7	10	11	13	14	16	16	17
mean	4,07	5,52	7,44	8,18	9,29	10,49	11,78	11,35	11,42

Table 5 Hierarchical depth by number of employees SW industry

The average hierarchical depth increases with the number of employees. The two groups reflecting the highest number of employees deviates slightly from this trend with an average lower hierarchical depth than the organizations with 1,001, to 5,000 employees. Although the 21 possible categories (including support functions) could have been assigned, the industry trend shows flatter hierarchies, with a maximum mean value of 11,76.

5. Discussion

The thesis successfully extended Lee's (2022) framework and adapted it to the FS and SW industries. The validation of the results revealed weaknesses in Lee's hierarchical order and provided important insights into the industries. As expected, the size of the organization had a significant influence on its hierarchical depth. Summarizing the results presented above, the analysis showed that both industries exhibit a pyramidal structure, where the number of positions increases as one moves down the hierarchy, with more operational staff members than top-level executives. Specialists form a sizable share of the labor force in both industries. Specialists account for 36.61% of the workforce in the FS industry and 40.66% in the SW industry, respectively, demonstrating an extensive dependence on specialized knowledge and abilities.

The number of middle management positions in both sectors indicates the importance of managers in organizing and supervising the work of operational personnel. In both industries, higher-level executive roles comprise a relatively modest portion of the total distribution; less than 1% falls toward most categories. Support roles are essential in both industries, contributing significantly to organizational efficiency. They account for approximately 20% of the workforce.

Regarding the hierarchical depth, the mean number of hierarchical levels is similar in both industries: The mean for the FS industry is 5.5 levels, while the mean for the SW industry is 5.82 levels. In the FS and SW industries, the median number of levels is four and five, respectively. Both industries demonstrate a clear trend where hierarchical depth increases with the size of the organization. Smaller organizations have fewer hierarchical levels, facilitating faster decision-making and more agile operations. Larger companies, on the other hand, add more levels, perhaps in order to handle growing complexity.

In both industries, organizations with exceptionally high hierarchical complexity were identified. The FS industry had 466 outliers with at least 16 hierarchical levels, while the SW industry had 689 outliers. This indicates that a minority of organizations in both sectors exhibit deeply hierarchical configurations.

Besides these findings, this study has several limitations that may affect the interpretation of the results. Firstly, the individual use of job titles within organizations is often arbitrary, and

relying solely on job titles might be insufficient (Donohue & Power, 2012). This was also reflected in the enormous variety of job titles identified; around one million per industry. On the other hand, the code enables around 95% of these titles to be categorized.

Secondly, occupational coding often incorporates factors such as education and experience, which job-titles may not fully capture. Since LinkedIn users often specify their educational background and professional experience, the same dataset can be facilitated to evaluate the influence of it. Nevertheless, the reliance on LinkedIn data poses another limitation, as this data was not formally collected and may contain potential biases. The combination of LinkedIn and Crunchbase data showed that around 80% of the data, which was classified by LinkedIn users to belong to the respective industries, could not be verified by Crunchbase. And even after applying the Crunchbase industry filter, the exclusion of external titles was necessary.

On the one hand, a machine learning approach might have been more forward-looking instead of rule-based coding. On the other hand, the coding of machine learning tools is more complex and requires an enormous amount of data, time, and expertise. Also, compared to computer-assisted or manual occupational coding this thesis provides an efficient alternative.

Furthermore, the categorization would benefit from representing all industries comprehensively, including support functions. A more inclusive categorization would provide a complete picture of the organizational hierarchy, reflecting the diversity of roles across different sectors. This could already be shown by the interdependencies of the FS and SW industries.

Although the number of interviews conducted was increased, additional interviews would provide more robust verification. Since the inferences were limited and insufficient to verify all categories comprehensively. Consequently, some categories were confirmed by only a few interviews, restricting the generalizability of the findings. Additionally, the interviews showed that observed differences between small and large organizations reflect industry norms but may introduce biases in data interpretation.

Lastly, the salary data used in this study could be further expanded to allow for more precise classification and interpretation of (all) hierarchical levels. More comprehensive salary data would enhance the robustness of the findings, providing deeper insights into the correlation between job titles and compensation.

6. Conclusion

This thesis provides an effective approach to measuring hierarchical depth in the financial services and software industries by developing a rule-based occupational coding framework. The use of LinkedIn data enabled large amounts of data to be categorized based on job titles and to make conclusions about the distribution in those industries. The coding was done with Python Pandas and fuzzywuzzy functions were used in several categorization steps to enable precise matching.

One of the key findings of this thesis is the extension of Lee's (2022) hierarchical framework from 11+1 levels to 16+2 levels for the financial services industry and 20+2 levels for the software industry. It was possible to identify a more nuanced differentiation of technical positions in the software industry, which led to more hierarchical levels. In addition, a more detailed differentiation between seniority titles at certain levels and a horizontal differentiation for the *Managers*, *Specialist* and *Worker* categories were included in the framework. To complete the representation of organizational complexity, a measure of geographical differentiation was added for all categories. The separate identification process of support function employees, which reflect around 20% of the workforce, enhances the options to evaluate organizational complexity further. Summarized, provides the inclusion of additional differentiation measures a more robust and detailed framework than suggested by Lee (2022), which can be used as a foundation for further analysis of organizational design.

The proposed structures were further enhanced through expert interviews. In addition, salary data that were categorized with the same code as the LinkedIn data served to determine the categorical order. The analysis of the standardized salary data also highlighted the expected trend towards higher salaries at higher organizational levels, with notable peaks in larger organizations. In comparison, mid- and lower-level positions tend to have more consistent and lower standardized salaries.

The results show several patterns that apply to both industries. The high proportion of specialists in the workforce indicates that both industries rely heavily on specialized skills and knowledge. Middle managers also account for a significant proportion of the workforce.

The use of managers underscores the importance of bridging the gap between top management and operational staff.

In addition, the hierarchical depth was found to increase steadily as the number of employees increased. In general, companies in both industries had a mean value between 5.5 (FS) and 5.82 (SW) hierarchical levels.

To reduce the limitations, future research might especially want to increase the validation measurements and extend the categorization to all industries. Nevertheless, the included validation measurements proved the proposed categorization, with a code allowing for a low rate of uncategorized job titles. To the best of my knowledge, no published framework offers the opportunity to analyze hierarchical depth and differentiate between vertical, horizontal, geographical, and seniority job titles. This work can offer a nuanced foundation to further analyze the interplay of hierarchical depth and other organizational configuration factors. Furthermore, the exploration of which parameters (strategic factors) influence hierarchical depth and what impact (fit criteria) these have on the organization's success can be another topic of interest. Research regarding organizational design is highly valuable to draw implications for long-term organizational success. By measuring and analyzing hierarchical depth, researchers can better understand the impact of structural choices on organizational efficiency, adaptability, and innovation capacity.

Bibliography

- Ahmed, R., Philbin, S. P., & Cheema, F.-e.-A. (2021). Systematic literature review of project manager's leadership competencies. *Engineering, Construction and Architectural Management*, 28(1), 1-30.
- Alonso, R., Dessein, W., & Matouschek, N. (2008). When does coordination require centralization? *American Economic Review*, 98(1), 145-179.
- American Express. (2024). *Executive Committee & Directors*. Retrieved 23.05.2024, from <https://ir.americanexpress.com/governance-and-corporate-responsibility/executive-committee-and-directors/default.aspx>
- Aufiero, S., De Marzo, G., Sbardella, A., & Zaccaria, A. (2023). Mapping job complexity and skills into wages. *arXiv preprint arXiv:2304.05251*.
- Bachman, K. (2023). *Managing Director vs. CEO: What's the Difference?* Retrived 17.05.2024, from <https://www.huntclub.com/blog/managing-director-vs.-ceo>
- Batini, C., & Scannapieco, M. (2006). *Data Quality: Concepts, Methodologies and Techniques*. Springer.
- Bean-Mellinger, B. (2018). *The Difference Between CEO, President & Managing Director*. Chron. Retrieved 18.05.2024, from <https://smallbusiness.chron.com/corporate-level-rank-structure-73315.html>
- Bekkerman, R., & Gavish, M. (2011). High-precision phrase-based document classification on a modern scale. Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining,
- Burstyn, I., Slutsky, A., Lee, D. G., Singer, A. B., An, Y., & Michael, Y. L. (2014). Beyond crosswalks: reliability of exposure assessment following automated coding of free-text job descriptions for occupational epidemiology. *Annals of occupational hygiene*, 58(4), 482-492.
- Burton, R. M., Lauridsen, J., & Obel, B. (2002). Return on assets loss from situational and contingency misfits. *Management science*, 48(11), 1461-1485.
- Burton, R. M., & Obel, B. (2004). *Strategic organizational diagnosis and design: The dynamics of fit* (Vol. 4). Springer Science & Business Media.
- Burton, R. M., & Obel, B. (2018). The science of organizational design: fit between structure and coordination. *Journal of Organization Design*, 7(1), 1-13.

- Burton, R. M., Obel, B., & Håkonsson, D. D. (2020). *Organizational Design: A Step-by-Step Approach* (4 ed.). Cambridge: Cambridge University Press. <https://doi.org/https://doi-org.uaccess.univie.ac.at/10.1017/9781108681162>
- Cascio, W. F. (2015). *Managing human resources*. McGraw-Hill New York, NY, USA:.
- Chapman, L. (2023). *The reality of being an engineering fellow*. LeadDev. Retrieved 22.05.2024, from <https://leaddev.com/personal-development/reality-being-engineering-fellow>
- Colombo, M. G., & Grilli, L. (2013). The creation of a middle-management level by entrepreneurial ventures: Testing economic theories of organizational design. *Journal of Economics & Management Strategy*, 22(2), 390-422.
- Cook, B. (2022, 18.05.2022). *The Differences Between Department Head and Director*. Retrived 15.05.2024, from <https://fellow.app/blog/management/the-differences-between-department-head-and-director/>
- Cornelis, T., Dubois, P., Omhover, J.-F., & Fercoq, A. (2019). Organisation design seen through systematic design. Proceedings of the Design Society: International Conference on Engineering Design,
- Crunchbase. (2024). *What Can I Do with Crunchbase Pro?* Retrieved 02.05.2024 from <https://www.crunchbase.com>
- Dessler, G. (2020). *Human resource management* (Sixteenth edition, global edition ed.). PearsonProQuest Ebook Central. <https://ubdata.univie.ac.at/AC16334847>
- Dobrajska, M., Billinger, S., & Karim, S. (2015). Delegation within hierarchies: How information processing and knowledge characteristics influence the allocation of formal and real decision authority. *Organization science*, 26(3), 687-704.
- Donohue, P., & Power, N. (2012). Legacy job titles in IT: the search for clarity. Proceedings of the 50th annual conference on Computers and People Research,
- Doty, D. H., Glick, W. H., & Huber, G. P. (1993). Fit, equifinality, and organizational effectiveness: A test of two configurational theories. *Academy of Management Journal*, 36(6), 1196-1250.
- Drazin, R., & Van de Ven, A. H. (1985). Alternative forms of fit in contingency theory. *Administrative science quarterly*, 514-539.
- Fallgatter, M. J. (2020). *Management und Managementenerfolg: Analyse, Prognose und Gestaltung von Wertschöpfung*. Springer.
- Galbraith, J. (1973). Designing complex organizations. *Reading, Mass.*

- Galbraith, J. R. (1974). Organization design: An information processing view. *Interfaces*, 4(3), 28-36.
- Galbraith, J. R. (2002). *Designing Organizations: An Executive Guide to Strategy, Structure, and Process*. Wiley.
- Ganzeboom, H. B., & Treiman, D. J. (2003). Three internationally standardised measures for comparative research on occupational status. In *Advances in cross-national comparison: A European working book for demographic and socio-economic variables* (pp. 159-193). Springer.
- Gweon, H., Schonlau, M., Kaczmirek, L., Blohm, M., & Steiner, S. (2017). Improving the Accuracy of Automated Occupation Coding at Any Production Rate. Gweon, H., Schonlau, M., Kaczmirek L., Blohm, M., Steiner, S. *Three Methods for Occupation Coding Based on Statistical Learning. Journal of Official Statistics*, 33(1), 101-122.
- h1bdata.info. (2024). *H1B Salary Database 2024*. Retrieved 26.05.2024, from h1bdata.info
- Hart, O., & Moore, J. (2005). On the design of hierarchies: coordination versus specialization. *Journal of political Economy*, 113(4), 675-702.
- Ikudo, A., Lane, J. I., Staudt, J., & Weinberg, B. A. (2019). Occupational classifications: A machine learning approach. *Journal of Economic and Social Measurement*, 44(2-3), 57-87.
- Indeed Editorial Team. (2023). *What Does a Department Manager Do? (And How to Become One)*. Indeed. Retrieved 18.05.2024, from <https://ca.indeed.com/career-advice/finding-a-job/what-does-a-department-manager-do>
- Indeed Editorial Team. (2024a). *Principal Engineer vs. Architect: Compare Duties and More*. Indeed. Retrieved 23.05.2024, from <https://www.indeed.com/career-advice/finding-a-job/principal-engineer-vs-architect>
- Indeed Editorial Team. (2024b). *What Does a Lead Engineer Do? (With Salary and Skills)*. Indeed. Retrieved 22.05.2024, from <https://sg.indeed.com/career-advice/finding-a-job/what-does-lead-engineer-do>
- Jiang, K., Lepak, D. P., Hu, J., & Baer, J. C. (2012). How does human resource management influence organizational outcomes? A meta-analytic investigation of mediating mechanisms. *Academy of Management Journal*, 55(6), 1264-1294.
- Joseph, D., Boh, W. F., Ang, S., & Slaughter, S. A. (2012). The career paths less (or more) traveled: A sequence analysis of IT career histories, mobility patterns, and career success. *Mis Quarterly*, 427-452.

- Joseph, J., Baumann, O., Burton, R., & Srikanth, K. (2018). Reviewing, revisiting, and renewing the foundations of organization design. In *organization design* (pp. 1-23). Emerald Publishing Limited.
- Joubert, S. (2020). *What's the Difference Between Project, Portfolio, and Program Management?*. Northeastern University. Retrieved 18.05.2024, from <https://graduate.northeastern.edu/resources/project-management-vs-portfolio-management-vs-program-management/>
- Kopp, C. M. (2024). *Partnership: Definition, How It Works, Taxation, and Types*. Investopia. Retrieved 17.05.2024, from <https://www.investopedia.com/terms/p/partnership.asp>
- Kretschmer, T., & Khashabi, P. (2020). Digital transformation and organization design: An integrated approach. *California Management Review*, 62(4), 86-104.
- Lazarova, M., Caligiuri, P., Collings, D. G., & De Cieri, H. (2023). Global work in a rapidly changing world: Implications for MNEs and individuals. *Journal of World Business*, 58(1), 101365.
- Leavitt, H. J. (1965). Applied organizational change in industry: Structural, technological and humanistic approaches. In *Handbook of Organizations (RLE: Organizations)* (pp. 1144-1170). Routledge.
- Lee, S. (2022). The myth of the flat start-up: Reconsidering the organizational structure of start-ups. *Strategic Management Journal*, 43(1), 58-92.
- Liao, C., Chuang, S.-H., & To, P.-L. (2011). How knowledge management mediates the relationship between environment and organizational structure. *Journal of business research*, 64(7), 728-736.
- Mahoney, T. A. (1979). Organizational hierarchy and position worth. *Academy of Management Journal*, 22(4), 726-737.
- Martinez, A. D., Laird, M. D., Martin, J. A., & Ferris, G. R. (2008). Job title inflation. *Human resource management review*, 18(1), 19-27.
- Maurer, T. J., & London, M. (2018). From individual contributor to leader: A role identity shift framework for leader development within innovative organizations. *Journal of Management*, 44(4), 1426-1452.
- Mazzetti, G., & Schaufeli, W. B. (2022). The impact of engaging leadership on employee engagement and team effectiveness: A longitudinal, multi-level study on the mediating role of personal-and team resources. *Plos one*, 17(6), e0269433.

- Miles, R. E., Snow, C. C., Meyer, A. D., & Coleman Jr, H. J. (1978). Organizational strategy, structure, and process. *Academy of management review*, 3(3), 546-562.
- Mintzberg, H. (1979). An Emerging Strategy of "Direct" Research. *Administrative science quarterly*, 24(4), 582-589. <https://doi.org/10.2307/2392364>
- Mintzberg, H. (1989). The Structuring of Organizations. In D. Asch & C. Bowman (Eds.), *Readings in Strategic Management* (pp. 322-352). Macmillan Education UK. https://doi.org/10.1007/978-1-349-20317-8_23
- Nagpal, A. (2024, 17.05.2024). *C Suite | V Suite | D Suite*. Retrived 08.05.2024 from <https://www.linkedin.com/pulse/c-suite-v-d-amit-nagpal-phd-jca4c/>
- Nahoomi, N. (2018). *Automatically Coding Occupation Titles to a Standard Occupation Classification* University of Guelph.
- O*NET Resource Center. (2024a). *About O*NET*. Retrieved 08.05.2024, from <https://www.onetcenter.org>
- O*NET Resource Center. (2024b). *Alternate Titles* Retrieved 08.05.2024, from https://www.onetcenter.org/dictionary/28.0/excel/alternate_titles.html
- O*NET Resource Center. (2024c). *Sample of Reported Titles*. Retrieved 08.05.2024, from https://www.onetcenter.org/dictionary/28.0/excel/sample_of_reported_titles.html
- Oi, W. Y., & Idson, T. L. (1999). Firm size and wages. *Handbook of labor economics*, 3, 2165-2214.
- Patel, M. D., Rose, K. M., Owens, C. R., Bang, H., & Kaufman, J. S. (2012). Performance of automated and manual coding systems for occupational data: a case study of historical records. In: Wiley Online Library.
- Perrow, C. (1977). The bureaucratic paradox: The efficient organization centralizes in order to decentralize. *Organizational Dynamics*, 5(4), 3-14.
- Pugh, D. S., Hickson, D. J., Hinings, C. R., & Turner, C. (1969). The context of organization structures. *Administrative science quarterly*, 91-114.
- Puranam, P. (2012). A future for the science of organization design. *Journal of Organization Design*, 1(1).
- Puranam, P. (2018). *The microstructure of organizations*. Oxford University Press.
- Puranam, P., Alexy, O., & Reitzig, M. (2014). What's "new" about new forms of organizing? *Academy of management review*, 39(2), 162-180.
- Python Software Foundation. (2024). *Regular expression operations* <https://docs.python.org/3/library/re.html#>
- Reitzig, M. (2022). *Get Better at Flatter*. Springer.

- Saab, H. (2023). *What Is A CEO?*. Wall Street Oasis. Retrieved 17.05.2024, from <https://www.wallstreetoasis.com/resources/careers/jobs/CEO-chief-executive-officer>
- Schierholz, M. (2014). *Automating survey coding for occupation*
- SeatGeek. *FuzzyWuzzy: Fuzzy String Matching in Python*
<https://github.com/seatgeek/fuzzywuzzy>
- Smet, A. D., Hewes, C., Mygatt, E., & Rieckhoff, K. (2022). *Redefining corporate functions to better support strategy and growth*. McKinsey. Retrieved 08.05.2024, from <https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/redefining-corporate-functions-to-better-support-strategy-and-growth>
- Tushman, M. L., & Nadler, D. A. (1978). Information processing as an integrating concept in organizational design. *Academy of management review*, 3(3), 613-624.
- U.S. Bureau of Labor Statistics. (2018a). *Direct Match Title File, 2018 SOC*. Retrieved from https://www.bls.gov/soc/2018/soc_2018_direct_match_title_file.pdf
- U.S. Bureau of Labor Statistics. (2018b). *The Standard Occupational Classification*. Retrieved from <https://www.bls.gov/soc/2018/home.htm>
- U.S. Bureau of Labor Statistics. (2018c). *Standard Occupational Classification (SOC) Manual: 2018 Edition*. Retrieved from https://www.bls.gov/soc/2018/soc_2018_manual.pdf
- Valette, M. (2021). *Who are staff, principal, and distinguished engineers?*. LeadDev. Retrieved 22.05.2022, from <https://leaddev.com/career-ladders/who-are-staff-principal-and-distinguished-engineers>
- Van de Ven, A. H., Delbecq, A. L., & Koenig Jr, R. (1976). Determinants of coordination modes within organizations. *American sociological review*, 322-338.
- Van de Ven, A. H., Ganco, M., & Hinings, C. R. (2013). Returning to the frontier of contingency theory of organizational and institutional designs. *Academy of Management Annals*, 7(1), 393-440.
- Van den Berghe, L., & Verweire, K. (2001). Convergence in the financial services industry. *The Geneva Papers on Risk and Insurance. Issues and Practice*, 26(2), 173-183.
- VanderPlas, J. (2016). *Python data science handbook: Essential tools for working with data*. " O'Reilly Media, Inc."

- Weber, M. (1947). The theory of economic and social organization. *Trans. AM Henderson and Talcott Parsons. New York: Oxford University Press.*
- Weber, M. (1987). The economy and society.
- Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and psychological measurement*, 73(6), 913-934.
- Xu, H., Yu, Z., Guo, B., Teng, M., & Xiong, H. (2018). Extracting Job Title Hierarchy from Career Trajectories: A Bayesian Perspective. *IJCAI*
- Zippia Team. (2022, 17.05.2024). *What is the difference between a partner and a managing partner?*. Zippia. Retrieved 10.05.2024, from <https://www.zippia.com/answers/what-is-the-difference-between-a-partner-and-a-managing-partner/#>

Appendix

Appendix 1: Evolution of the Contingency Perspective

Following the contingency perspective, the evolution of organizational design theories has significantly shifted from seeking a universal "best way" of management to embracing the variability of approaches based on specific contextual needs (Van de Ven et al., 2013). Early theorists initially focused on defining task design and formal structures as essential aspects of organizational management. For instance, Weber (1987) depicted the properties of a bureaucratic price, spotlighting crucial elements such as specialization, coordination, hierarchical authority, and formal rules. The analysis of bureaucracy, especially the underpinnings of rational-legal authority, defined by Weber (1947), was also a driving factor leading to the contingency theory. During industrialization, authority patterns changed significantly, resulting in the rise of rational-legal authority, which emphasized tasks designed based on expertise and supported by formal policies (Joseph et al., 2018).

The movement away from a one-size-fits-all approach was further advanced by Lawrence and Lorsch (1967) and Thompson (1967) by suggesting that the ideal structure varies according to the specific conditions of the organization.

Thompson (1967) made it prominent that there is not one single template for good organizations, he highlighted the interplay of the external environment and the firm's own interdependencies. Lawrence and Lorsch (1967) argued that the most effective organizational structure depends on the specific external and internal conditions it encounters. They specifically identified differentiation and integration as crucial elements of organizational design. They posited that environments with high uncertainty require greater internal differentiation to cope effectively, which in turn requires a higher level of integration to maintain unity and coordination within the organization.

This foundational work on contingency theory significantly influenced subsequent strategic research, which shifted focus from hierarchical structures to more flexible design configurations that facilitate effective information processing and align structural characteristics with the organization's environment, technology, and scale (Joseph et al., 2018). Notable contributions in this area include those by Galbraith (1974) and Tushman and Nadler (1978), who emphasized the role of information processing in organizational design, and Drazin and Van de Ven (1985),

as well as Mintzberg (1979), who further explored how organizational structures can be adapted to match environmental conditions.

In synthesizing these developments, it becomes clear that the modern contingency perspective emphasizes a tailored approach, where the design and management strategies are contingent upon the specific circumstances faced by the organization, marking a significant evolution from earlier management theories that sought a universal solution to organizational design.

Appendix 2: Definition Universals of Organizing

Task division is the foundational process where organizational goals are broken down into specific, manageable tasks, setting the stage for effective delegation and execution. This structuring determines how tasks are defined and organized, directly influencing the complexity and manageability of operations. Subsequently, task allocation becomes crucial, as it involves distributing these defined tasks among organizational members based on their skills, experience, and capacities. This allocation is pivotal for optimizing resource utilization and ensuring that tasks are executed by the most qualified individuals, thereby enhancing productivity and operational efficiency.

Concurrently, reward provision plays a significant role in motivating and incentivizing employees to perform these tasks effectively. This dimension encompasses not only monetary rewards such as salaries and bonuses but also non-monetary incentives like recognition and growth opportunities, which help align individual motivations with the broader organizational goals. The effectiveness of reward systems is inherently tied to the clarity and appropriateness of task division and allocation, as these define the benchmarks for performance evaluation. Lastly, information provision binds all these dimensions by ensuring that all members have access to the necessary information to execute their tasks efficiently. This involves establishing robust communication channels and information systems that support transparency, facilitate knowledge sharing, and aid in decision-making processes across the organization. Effective information provision is essential for maintaining coordination and coherence across the various layers of the organization, enabling it to respond agilely to both internal and external changes (Puranam et al., 2014).

Appendix 3: External Job Titles

- nurse, nursing, rn, surgeon, pharmacist, medical, therapist, physician, dental, health, pharmaceutical, paramedic, clinical,
- navy sea, air forcek-9, pilot, police, police officer, detective, investigator, lieutenant, captain, fbi agent, sergeant, guard, crime, firefighter, fire, fescue, guard
- Technician, technical, electrician, electrical, technical support, mechanic, automotive, auto, parts, car, driver, manufacturing, manufacture, production, logistics, assembly, warehouse, plant, operator, energy, telecommunication, network, wireless,
- farmer, agriculture, crop, environmental, environment, biologist, geologist, climate, paymaster,
- realtor, real estate, property, title officer,
- historian, teacher, teaching, phd, professor, postdoc, postgraduate, researcher, undergraduate, student, lawyer, law clerk, paralegal,
- chef, stylist, fashion, wardrobe, store, cashier, photographer, graphic designer, event planner, janitor, plumber, actor, actress, artist, singer, editor, author, writer, proofreader, bookseller, waitress, waiter, barista, bartender, pcb, resident assistant, pastor, fundraising, fundraiser, personal trainer, coach
- freelance, no_info, seeking employment, stay at home, dad, mom, mother, wife, family leave, retired, retirement, unemployed, not working, retiree, sabbatical, summer, various, homemaker

Appendix 4: SOC Categories FS Industry

Specialists			
14.01	Buyers	13 – 1020	Buyers and Purchasing Agents
14.02	Coordinators and Consultants	13 – 1100	Management Analysts
14.03	Accountants and Auditors	13 – 2010	Accountants and Auditors
14.04	Credit Analyst	13 – 2040	Credit Analysts
14.05	Financial Analyst, Advisors, and Underwriters	13 – 2050	Financial Analysts, Advisors, and Underwriters
14.06	Financial Examiner	13 – 2060	Financial Examiners
14.07	Credit Counselors and Loan Officers	13 – 2070	Credit Counselors and Loan Officers
14.08	Tax Examiners, Collectors, and Revenue Agents	13 – 2080	Tax Examiners, Collectors, Prepares, Revenue Agents
14.09	Others	13 – 2090	Others
Workers			
15.01	Bill and Account Collectors	43 – 3010	Bill and Account Collectors
15.02	Billing and Accounting Clerks	43 – 3020	Billing and Posting Clerks
15.03	Bookkeeping, Accounting, and Auditing Clerk	43 – 3030	Bookkeeping, Accounting and Auditing Clerks
15.04	Tellers	43 – 3070	Tellers

15.05	Credit Authorizers, Checkers and Clerks	43 – 4040	Credit Authorizers, Checkers, Clerks
15.06	Loan and New Account Clerks	4130 + 4140	Loan Interviewers + New Accounts Clerks
15.07	FS Sales Agents	41 – 3030	Securities, Commodities, and Financial Services Sales Agents
15.08	Personal Assistants	43 – 6000	Secretaries and Administrative Assistants
15.09	Others	43 – 3090	Others

Appendix 5: SOC Categories SW Industry

Specialists	
17.01 Computer and Information Analysts	15 – 1210 Computer and Information Analysts
17.02 Computer and Information Research Scientists	15 – 1220 Computer and Information Research Scientists
17.03 Computer Support Specialists	15 – 1230 Computer Support Specialists
17.04 Database and Network Administrators and Architects	15 – 1240 Database and Network Administrators and Architects
17.05 Software and Web Developers, Programmers, and Testers	15 – 1250 Software and Web Developers, Programmers, and Testers
– 17.051 Computer Programmers	– 1251 Computer Programmers
– 17.052 Software Developers	– 1252 Software Developers
– 17.053 Software Quality Assurance Analysts and Testers	– 1253 Software Quality Assurance Analysts and Testers
– 17.054 Web Developers	– 1254 Web Developers
– 17.055 Web and Digital Interface Designers	– 1255 Web and Digital Interface Designers
17.06 Consultants	13 – 1110 Consultants
17.07 Others	15 – 1290 Others

Appendix 6: Rule-based Matching – FS Code

Create smaller files to test the code

```
from unittest import case
import pandas as pd

# Read the large CSV file
large_df = pd.read_csv('/Users/kimmelissabar/Documents/GitHub/Masterarbeit/datasets/transfe
r_121697_files_a38c0739/combined_0.csv')

# Create a smaller subset (e.g., 200MB)
sample_df = large_df.sample(frac=0.12) # Adjust fraction to get around 200MB

# Save the sample as a new CSV file
sample_df.to_csv('/Users/kimmelissabar/Documents/GitHub/Masterarbeit/CSV files created/try/
try.csv', index=False)
```

Create a loop to execute all files at once

```
import numpy as np
import pandas as pd
import os
from fuzzywuzzy import fuzz

# Specify the directory where your CSV files are located
directory_path = "/Users/kimmelissabar/Documents/Masterarbeit Schreiben/Code/financial/outp
uts"

# Get the list of file names in the specified directory
file_names = os.listdir(directory_path)

# Create an empty DataFrame to store the final processed data
empty_list = []

#define a function to exclude misleading companies and job titles that do not belong to the
industry; outside the loop
def filter_conditions(columns_used_exp):
    cond1 = columns_used_exp['legal_name'].str.contains('self financial, inc.', case=False)
    & columns_used_exp[
        'experience_title'].str.contains('owner', case=False)
    cond2 = columns_used_exp['legal_name'].str.contains('primerica, inc.', case=False) & co
lumn_s_used_exp[
        'experience_title'].str.contains('owner', case=False)
    cond3 = columns_used_exp['legal_name'].str.contains('hoom inc.', case=False) & column
s_used_exp[
        'experience_title'].str.contains('owner', case=False)
    cond3_1 = columns_used_exp['legal_name'].str.contains('hoom inc.', case=False) & column
s_used_exp['experience_company'].str.contains('home', case=False)
    cond4 = columns_used_exp['legal_name'].str.contains('cisco', case=False) & column_s_
used_exp[
        'experience_title'].str.contains('owner', case=False)
    cond5 = columns_used_exp['legal_name'].str.contains('haus services inc', case=False) &
column_s_used_exp[
        'experience_title'].str.contains('owner', case=False)
    cond_company = columns_used_exp['experience_company'].str.strip().str.contains(
        'self at home|owner at home|retired|www.|purdue university', case=False)

    cond_external1 = columns_used_exp['experience_title'].str.strip().str.contains('nurse|n
ursing|surgeon|internist|pharmacist|pharmacy|medical|therapist|physician|dental|health|phar
maceutical|paramedic|clinical|navy seal|air force|k-9|pilot|police|police officer|detective
|investigator|lieutenant|captain|fbi agent|sergeant|guard|crime|firefighter|fire|rescue|tec
hnical|electrician|mechanic|automotive|auto|parts|car|driver|manufacturing|manufacture|prod
```

```

uction|logistics|assembly|warehouse|electrical|plant|operator|energy|telecommunication|network|wireless|farmer|agriculture|crop|environmental|environment|biologist|geologist|climate|paymaster|city|diplomat|fundraising|fundraiser|receptionist|realtor|real estate|property|historian|teacher|lecturer|teaching|phd|professor|postdoc|postgraduate|researcher|undergraduate|student|chef|stylist|fashion|wardrobe|store|cashier|photographer|graphic designer|event planner|janitor|actor|actress|artist|singer|musician|editor|author|writer|proofreader|bookseller|waitress|waiter|barista|bartender|pcb|resident assistant|pastor|freelance|no_info|seeing employment|stay at home|homemaker|dad|mom|mother|wife|family leave|retired|retirement|unemployed|not working|retiree|sabbatical|summer|plumber|various|personal trainer|coach|nanny|mover|library|librarian|linguist|self|self-employed|self employed|master', case=False)

```

```

all_conditions = cond1 | cond2 | cond3 | cond3_1| cond4 | cond5 | cond_company | cond_external1

```

```

filtered_chunk = columns_used_exp[~all_conditions]

```

```

return filtered_chunk
pass

```

```

#predefine support functions title matching outside the loop

```

```

def categorize_support_words(experience_title):

```

```

    if pd.notna(experience_title): # Check if the value is not NaN
        words_to_match_support = ['software', 'developer', 'software development', 'engineer', 'ux designer', 'programmer', 'programming', 'solution architect', 'solutions architect', 'system architect', 's/w', 'web designer', 'systems', r'\bqa\b', r'\bqu\b', 'qa tester', 'solutions', 'performance', 'principal developer', 'principal engineer', 'sql', 'sap', 'systems', 'test', 'data scientist', 'information security', 'database', 'system', 'system administrator', 'quality assurance', r'\bqa\b', 'sales', 'sales', 'registered representative', 'reg. rep', 'representative', 'registered rep', 'sales training', 'salesforce', 'procurement', 'marketing', 'social media', 'advertising', 'promotions', 'communication', 'brand ambassador', 'content creator', r'\bhr\b', r'\bhrm\b', 'human resource', 'recruiter', 'recruiting', 'recruitment', 'staffing', 'payroll', 'timekeeping', 'employee benefits', 'compensation', 'wage', 'salary', 'talent acquisition', 'workforce', 'hrm', 'workforce management', 'human resource manager', 'title officer', 'sourcer', 'people operations', 'personnel', 'position descriptor', 'talent', 'acquisition', 'pension', 'legal counsel', 'notary public', 'legal advisor', 'legal administrator', 'legal researcher', 'legal liaison', 'legal ninja', 'legal operations', 'legal secretary', 'legal services', 'legal translator', 'legislative aide', 'patent', 'legal', 'notary', 'licensing', 'lawyer', 'law', 'notary', 'licensing', 'law clerk', 'paralegal', 'attorney', 'counsel', 'csr', 'diversity', 'inclusion', 'csr', 'corporate responsibility', 'policy', 'compliance', 'business development', 'bus. dev', 'research and development', 'r&d', 'governance', 'client service', 'client', 'client support', 'support', 'support representative', 'customer service', r'\bcco\b', r'\bcdo\b', r'\bcho\b', r'\bci\b', r'\bclo\b', r'\bcmo\b', r'\bcpo\b', r'\bcso\b', r'\bcto\b', 'ccao', r'\bcos\b', r'\bro\b', 'lmts', 'lead member of technical staff', 'technical staff']

```

```

    for word in words_to_match_support:
        match_score_1 = fuzz.partial_ratio(word.lower(), experience_title.lower())
        # If a better match is found, update the best_match tuple
        if match_score_1 == 100 and len(word) <= len(experience_title):
            for category, keywords in rules:
                if word in keywords:
                    return category

```

```

    return 0

```

```

#filtered_data = filter_conditions(columns_used_exp)

```

```

# Use a loop to process each CSV file and append the filtered results

```

```

for file_name in file_names:

```

```

    if file_name.endswith(".csv") and not file_name.startswith("."):

```

```

        # Read the CSV file

```

```

        df = pd.read_csv(os.path.join(directory_path, file_name), encoding='ISO-8859-1', on_bad_lines='warn')

```

```

        print("Loop at file:", file_name)

```

```

        columns_used_exp = df.loc[:, ['name', 'id', 'experience_title', 'experience_company',
        'legal_name', 'experience_start_date', 'experience_end_date', 'experience_duration_short',
        'city', 'type', 'crunchbase_industries', 'industry_dummy']]

        #Replace "\N" with NaN (Not-a-Number)
        columns_used_exp.replace('\N', np.nan, inplace=True)

        #fill the nan rows with the string "no info"
        columns_used_exp = columns_used_exp.fillna('no_info')

        columns_used_exp = columns_used_exp.loc[~columns_used_exp['type'].str.contains('investor')]
        # Drop the entire 'type' column
        columns_used_exp.drop(columns=['type'], inplace=True)

        #delete double quotation marks
        columns_used_exp['legal_name'] = columns_used_exp['legal_name'].str.replace('"', '')
    )

    #avoid case sensitivity
    columns_used_exp['legal_name'] = columns_used_exp['legal_name'].str.lower()
    columns_used_exp['name'] = columns_used_exp['name'].str.lower()
    columns_used_exp['experience_title'] = columns_used_exp['experience_title'].str.lower()
er()
    columns_used_exp['experience_company'] = columns_used_exp['experience_company'].str.lower()
    columns_used_exp['crunchbase_industries'] = columns_used_exp['crunchbase_industries'].str.lower()

    #exclude all entries without "financial services"-tag
    columns_used_exp = columns_used_exp.loc[columns_used_exp['industry_dummy'].astype(str).str.contains('1.0')]
    #delete the column
    columns_used_exp.drop(columns=['industry_dummy'], inplace=True)

    #adding the conditions function that is pre-defined
    filter_conditions(columns_used_exp)
    combined_conditions_df = filter_conditions(columns_used_exp)

    #check number of rows
    num_rows = len(df)
    num_rows_industry = len(columns_used_exp)
    num_rows_external = len(combined_conditions_df)
    # Printing the number of rows
    print(f"The number of rows in the DataFrame started with: {num_rows}; dropped to after crunchbase filter: {num_rows_industry}; is now: {num_rows_external}")

    #Define the rules for categorization
    rules = [
        (1, ['owner', 'founder', r'\bcreator\b', 'created', 'made', 'ownere', r'\bpartner\b', 'co-owner', 'business partner', 'founding member']),
        (2, ['board member', 'board of directions', 'board of directors', 'chairman']),
        (3, ['leading partner', 'managing partner']),
        (4, ['ceo', 'chief executive', 'cheif executive']),
        (5, [r'\bcco\b', r'\bcdo\b', r'\bcfo\b', r'\bcho\b', r'\bcio\b', r'\bclo\b', r'\bcmo\b', r'\bcoo\b', r'\bcpo\b', r'\bcso\b', r'\bcto\b', 'ccao', 'ceo', r'\bcos\b', r'\bro\b', 'chief officer', 'chief\s+\w+\s+officer', 'chief technology officer', 'chief operating officer', 'chief people officer', 'chief marketing officer', 'chief staff officer', 'chief financial officer', 'chief revenue officer', 'chief product officer', 'chief information officer', 'chief sales officer', 'chief of staff', 'chief investment officer', 'chief strategy officer']),
        (6, ['president', 'presidente']),
        (7, [r'md\b', 'managing director']),
        (8.1, [r'evp\b', 'executive vp', 'executive vice president', ' executive vice-president', 'exec vice pres']),
    ]

```

(8.2, ['senior vice president', 'sr vice president', 'snrvp', 'svp', 'senior v. p.', 'sr. vice president', 'r\'senior v\bp', 'sr vp', 'r\'sr. \bvp\b', 'r\'bsenior\s+\w+\s+vp\b']),

(8.3, [r'\bvp\b', r'\bv\.\?p\b', 'vice president', 'vice-president', 'vice presidente', 'vice pres', r'\bv\b\.\?s?p\b', r'\bvp,\b']),

(8.4, ['avp', 'assistant vice president', 'junior vice president', 'r\'bjr. vp\b', 'r\'bjn\b\s+\vp\b']),

(9.1, ['executive director', 'exec. dir', 'exec director']),

(9.2, ['senior director', 'sr. director', 'r\'bsenior\s+\w+\s+director\b']),

(9.3, ['director', 'directo', 'diercto', 'r\'dir\b', 'dierctor']),

(9.4, ['junior director', 'r\'bjun dir\b', 'assistant director', 'r\'bjunior\s+\w+\s+director\b', 'r\'bassistant\s+\w+\s+director\b', 'jr. director', 'r\'bjr dir', 'r\'basst d ir\b', 'r\'asst\s+dir']),

(10, [r'\bhead\s+\w+', 'r\'bheads\b', 'r\'bhead\s+\of\w+', 'r\w+\s+\head\s+\w+', 'r\w+\s+\head', 'head of']),

(11.01, ['program manager']),

(11.02, ['project manager', 'r\'bpmo\b', 'project management officer', 'r\'senior \bpm\b', 'r\'pm\b', 'r\'pm\b manager']),

(11.03, ['banking center manager', 'banking manager', 'city comptroller', 'treasurer', 'comptroller', 'fiscal manager', 'international bank manager', 'controller', 'bank manager', 'account manager']), *#manager finance*

(11.1, ['executive manager']),

(11.2, ['senior manager', 'senior management team', 'sn manager', 'sr. manager', 'sr mng', 'sr manager', 'r\'bsenior\s+\w+\s+\manager\b', 'r\'bsr.\s+\w+\s+\manager\b', 'r\'bsenior\s+\w+\s+\w+\s+\manager\b']),

(11.3, ['manager', 'manager', 'manger', 'mananger', 'mgr', 'r\'gm\b', 'department manager', 'division manager']),

(11.4, ['junior manager', 'assistant manager', 'r\'bassistant\s+\w+\s+\manager\b', 'management assistant']),

(12, ['supervisor', 'instructor', 'instructer', 'mentor']),

(13, ['lead', 'leader', 'lead member of', 'team leader', 'lead of']),

#14.2 senior specialists

(14.01, ['buyer', 'client relationship manager', 'buying agent', 'broker', 'dealer', 'trader', 'purchasing agent', 'purchaser', 'importer', 'purchasing', 'tradesman']),

(14.02, ['coordinator', 'consultant', 'financial consultant', 'personal finance coordinator', 'pfc']),

(14.03, ['accountant', 'auditor', 'audit', 'assurance', 'cash management']),

(14.04, ['credit analyst', 'portfolio credit manager', 'pcm', 'claims analyst', 'credit risk analyst', 'credit administrator', 'r\'bcredit\s+\w+\s+\analyst\b', 'credit specialist', 'factorer', 'loan review analyst']),

(14.05, ['wealth analyst', 'investor', 'wealth strategist', 'underwriter', 'underwriting', 'analyst', 'associate', 'financial planner', 'financial planning', 'adviser', 'advisor', 'generalist', 'pcii', 'personal client investment specialist', 'private client as sociate', 'pca', 'risk control analyst', 'rca', 'pcis', 'personal finance specialist', 'pfs', 'private banking specialist', 'pbs', 'private banking officer', 'portfolio management officer', 'portfolio officer', 'pmo', 'portfolio development manager', 'pdm', 'stock analyst', 'trust officer', 'loan funder', 'mortgage officer', 'revenue officer', 'financial risk officer', 'finance officer', 'financial officer', 'portfolio manager', 'r\'bpm\b', 'portfolio administrator', 'bank officer', 'financial counselor', 'individual pension consultant', 'insurance writer', 'pricing evaluator', 'asset management', 'financial services officer', 'financial service officer', 'financial services strategist', 'finance service advisor', 'account executive', 'insurance producer', 'client relationship manager']),

(14.06, ['loan examiner', 'loan reviewer', 'bank compliance officer', 'bank examiner', 'financial compliance examiner', 'home mortgage disclosure act specialist', 'payroll examiner', 'pension examiner', 'examiner', 'claims investigator', 'appraiser']),

(14.07, ['credit counselor', 'debt counselor', 'case manager', 'loan validator', 'validator', 'credit adjuster', 'adjuster', 'credit officer', 'loan originator', 'originator', 'loan officer', 'loan administrator', 'credit counselor', 'counselor', 'loan concierge', 'financial respondent', 'lender', 'lending officer', 'debt management counselor', 'financial aid counselor', 'loan counselor', 'lending officer', 'commercial lender', 'loan analyst', 'loan reviewer', 'credit manager', 'financial aid', 'loan advisor', 'mortgage loan', 'personal loan']),

(14.08, ['tax examiner', 'preparer', 'revenue agent', 'licensed tax consultant', 'income tax preparer', 'income tax advisor', 'tax', 'city collector', 'customs appraiser', 'tax adjuster', 'revenue agent', 'revenue collector', 'revenue enforcement agent', 'tax compliance officer', 'tax compliance representative', 'tax investigator', 'corporate tax preparer', 'tax consultant', 'tax specialist', 'adjustment examiner', 'block advisor', 'city c

collector', 'city tax', 'customs', 'income tax', 'internal revenue', 'tax']],
 (14.09, ['financial specialist', 'financial expert', 'practitioner', 'specialist', 'advocate', 'financial administrator', 'trust administrator', 'actuary', 'financial services professional', 'bondsman', 'superintendent', 'financial services dealer', 'financial services professional', 'financial service concierge', 'financial professional', 'personal banker', 'r'\bpb', 'personal banking officer', 'pbo', 'banking officer', 'banker', 'financial services administrator']),
 #15.2 senior workers
 (15.01, ['repossessioner', 'debt collector', 'collecting agent', 'telex operator', 'accounts collector', 'bill collector', 'billing representative', 'collection agent', 'collections clerk', 'collections representative', 'installment agent', 'loan collector', 'insurance collector', 'payment collector']),
 (15.02, ['wholesale processor', 'agent', 'billing clerk', 'invoice clerk', 'invoice control clerk', 'posting clerk', 'statement clerk', 'statement processor']),
 (15.03, ['bookkeeper', 'bookkeeping', 'book keeper', 'accounts receivable assistant', 'accounts receivable clerk', 'auditing clerk', 'fixed capital clerk', 'foreign exchange position clerk', 'mortgage accounting clerk']),
 (15.04, ['teller', 'foreign exchange clerk', 'money order clerk', 'vault teller']),
 (15.05, ['authorizer', 'credit reviewer', 'checker', 'credit rating checker', 'credit investigator', 'credit processor', 'credit reference clerk', 'finance writer']),
 (15.06, ['loan opener', 'mortgage processor', 'mortgage closing clerk', 'mortgage loan closer', 'mortgage loan processing clerk', 'banking services clerk', 'new accounts banking representative', 'new client banking', 'services clerk', 'loan servicing', 'loan service']),
 (15.07, ['foreign exchange trader', 'investment banker', 'securities trader', 'stock broker']),
 (15.08, ['personal assistant', 'assistant', 'assistan', 'assitant', 'r'\bpa\b', 'alumni secretary', 'office secretary', 'personal secretary', 'asst', 'secretary']),
 (15.09, ['processor', 'financial representative', 'claims representative', 'investment service', 'r'\bstaff\b', 'r'\bstaf\b', 'clerk', 'employee', 'administrative clerk', 'office assistant', 'office clerk', 'office administrator', 'financial services operator', 'financial services rep', 'financial services support', 'bank vault', 'financial operations clerk', 'financial reserve clerk', 'vault teller']),
 (16, ['trainee', 'trainie', 'research student', 'r'\bintern\b', 'volunteer', 'training program', 'mentee', 'apprentice']),
 (30.1, ['developer', 'software development', 'software', 'engineer', 'ux designer', 'programmer', 'programming', 'solution architect', 'solutions architect', 'system architect', 'software architect', 'r'\bsw\b', 'web designer', 'it consultant', 'programmer analyst', 'systems analyst', 'test analyst', 'qa analyst', 'qa tester', 'qu tester', 'solutions analyst', 'performance analyst', 'programmer analyst', 'principal developer', 'principal engineer', 'data specialist', 'unix system', 'r'\bsql\b', 'r'\bui\b designer', 'sql', 'sap', 'system administrator', 'systems administrator', 'system administrator', 'security officer', 'systems administrator', 'data scientist', 'application architect', 'database administrator', 'enterprise architect', 'lmts', 'lead software', 'lead technical']),
 (30.2, ['sales', 'sales representative', 'registered representative', 'reg. rep', 'registered rep', 'sales manager', 'sales associate manger', 'sales training manager', 'sales consultant', 'salesforce', 'sales associate', 'sales agent', 'sales assistant', 'procurement', 'client service', 'client advocate', 'client support', 'support representative', 'customer service']),
 (30.3, ['marketing', 'advertising', 'promotions', 'marketing manager', 'brand ambassador', 'social media', 'communication', 'content creator']),
 (30.4, ['r'\hr\b', 'human resources', 'recruiter', 'recruiting', 'staffing', 'payroll', 'timekeeping', 'employee benefits', 'compensation', 'wage', 'salary', 'talent acquisition', 'talent', 'talent professional', 'workforce management', 'human resource manager', 'hrm', 'recruitment manager', 'payroll manager', 'personnel', 'job analysis', 'position descriptor', 'talent acquisition coordinator', 'payroll analyst', 'payroll assistant', 'pension administrator', 'title officer', 'technical sourcer', 'sourcer', 'talent management', 'people operations']),
 (30.5, ['csr', 'diversity', 'inclusion', 'csr', 'corporate responsibility', 'policy', 'compliance', 'business development', 'bus. dev', 'research and development', 'r&d', 'governance', 'eoo']),
 (30.6, ['notary public', 'legal liaison', 'legal operations', 'legal services', 'legal translator', 'legislative aide', 'law', 'lawyer', 'attorney', 'general counsel', 'legal counsel', 'legal advisor', 'legal administrator', 'legal researcher', 'legal ninja', 'legal secretary', 'legal translator', 'patent', 'legal', 'notary', 'licensing', 'lawyer', 'law clerk', 'paralegal', 'attorney', 'counsel', 'internal audit']),

```

    (35, []) # Default category
]

#funcitons to categorize job titles

combined_conditions_df = combined_conditions_df.copy() # Create a copy to avoid the SettingWithCopyWarning

combined_conditions_df['support_functions'] = combined_conditions_df['experience_title'].apply(categorize_support_words)

#extract 100% exact matches - the abbreviations have partially distorted the results because they are word components. -> solvable by \bvp\b

rules_exact = [
    (1, ['partner', 'creator']),
    (4, ['ceo']),
    (5, ['cco', 'cdo', 'cfo', 'cho', 'cio', 'clo', 'cmo', 'coo', 'cpo', 'cso', 'cto', 'ccao', 'cceo', 'cos', 'cro']),
    (6, ['president']),
    (7, ['md']),
    (8.1, ['evp']),
    (8.2, ['snrvp', 'svp', 'sr vp', 'sr. vp']),
    (8.3, ['vp', r'v\.?p', 'vice president', 'vp,']),
    (8.4, ['avp', 'jr. vp', 'jr vp']),
    (9.3, ['director', 'dir']),
    (9.4, ['jun dir', 'jr dir', 'asst dir']),
    (10, ['head']),
    (11.02, ['pm', 'pmo']),
    (11.3, ['manager', 'manger', 'mananger', 'mgr', 'gm']),
    (13, ['lead', 'leader']),
    (14.02, ['pfc']),
    (16, ['intern'])
]

def categorize_exact_words(experience_title):
    if pd.notna(experience_title): # Check if the value is not NaN
        words_to_match = ['partner', 'principal', 'manager', 'president', 'vice president', 'svp', 'vp', 'v.p', 'rvp', 'vp', 'avp', 'ceo', 'cco', 'cdo', 'cfo', 'cho', 'cio', 'clo', 'cmo', 'coo', 'cpo', 'cso', 'cto', 'ccao', 'cceo', 'cos', 'cro', 'sr mng', 'pm', 'gm', 'md', 'director', 'dir', 'vp,', 'head', 'lead', 'leader', 'intern']

        for word in words_to_match:
            if experience_title.lower() == word: # Check for exact match
                for level, keywords in rules_exact:
                    if word in keywords:
                        return level # Return the level based on the matched word

        return 35 # If no match found

#new df to just work with cat 35, thus a re-categorization of the exact matches is avoided
category_35 = combined_conditions_df

category_35 = category_35.copy()
# Apply categorization based on exact words only for rows categorized as 35 by external matching
category_35['exact_level'] = category_35.apply(lambda row: categorize_exact_words(row['experience_title']), axis=1)

# Update the original DataFrame with the re-categorized values
combined_conditions_df.loc[category_35.index, 'supervision_level'] = category_35['exact_level']

#filter by partial ratio, now it will be matched if the term from the rule appears in the title and the title must be longer than the rule

```



```

#including elif with less than 100% results in many mismatches --> because there is no AI behind the matching logic but pure letter analysis
new_df = category_35[category_35['exact_level'] == 35].copy()
new_df['experience_title'] = new_df['experience_title'].astype(str)

def categorize_with_partial_ratio(experience_title, rules):
    best_match = (35, 0)
    # Iterate through the rules except the default
    for level, keywords in reversed(rules[:-1]): #reversed filter - double titles c
at to highest
        for keyword in keywords:
            if len(experience_title) >= len(keyword):
                #Calculate the match score using partial_ratio
                match_score = fuzz.partial_ratio(keyword.lower(), experience_title.
lower())
                # If a better match is found, update the best_match tuple
                if match_score == 100:
                    best_match = (level, match_score)

        return best_match[0]

new_df['level_three'] = new_df['experience_title'].apply(lambda x: categorize_with_
partial_ratio(x, rules))

# Update the original DataFrame with the re-categorized values
combined_conditions_df.loc[new_df.index, 'supervision_level'] = new_df['level_three
']

#check column counts
spalte = 'supervision_level'
wert_counts = combined_conditions_df[spalte].value_counts()
print(wert_counts)

#Avoid setting with copy warning
combined_conditions_df['senior_supervision_level'] = combined_conditions_df['superv
ision_level']

# Regular expression for 'vice president'
pattern_vp = r'vice president|vp\.|\bvp\b|\bsvp\b|v\.p\.|vice president|vice presid
ent,|vp,'

# Use str.contains with regular expression and case sensitivity=False
mask = combined_conditions_df['experience_title'].str.contains(pattern_vp, case=False, regex=True)

all_below_vp = combined_conditions_df['senior_supervision_level'].astype(float).isin([6, 35, *range(9, 16)])

# Update based on the mask
combined_conditions_df.loc[mask & all_below_vp, 'senior_supervision_level'] = 8.3

#do the same for assistant columns

pattern_jun = r'\bassistent\b|\basst\b| assistant\s+\w+\s+\w+\s+\w+\b|assistant\s+\w+\s+\w+\b|asst.\s+\w+\s+\w+\b|\bjunior\s+\b|\bjun\b|\bjr\.\b|assistant|junior'
mask_2 = combined_conditions_df['experience_title'].str.contains(pattern_jun, case=False, regex=True)

#vp
all_junvp = combined_conditions_df['senior_supervision_level'].astype(float).between(6, 8.3)
# Update based on the mask
combined_conditions_df.loc[mask_2 & all_junvp, 'supervision_level'] = 8.4
#Transfer results to supervision level column

```

```

#director
all_jundirector = combined_conditions_df['senior_supervision_level'] == 9.3
# Update based on the mask
combined_conditions_df.loc[mask_2 & all_jundirector, 'supervision_level'] = 9.4

#create senior specialist and expert category
condition_spec = combined_conditions_df['senior_supervision_level'].astype(float).between(14.01, 14.09)
condition_work = combined_conditions_df['senior_supervision_level'].astype(float).between(15.01, 15.09)

combined_conditions_df.loc[condition_spec & combined_conditions_df['experience_title'].str.contains(r'\bsenior\b|\bsr\.\|bsn\.\|bsr\b', case=False), 'supervision_level'] = 14.2

combined_conditions_df.loc[condition_work & combined_conditions_df['experience_title'].str.contains(r'\bsenior\b|\bsr\.\|bsn\.\|bsr\b', case=False), 'supervision_level'] = 15.2

# categorize geographical differentiation function
rule_goe = [
    ('a', ['global']),
    ('b', ['emea', 'usa', 'europe', 'apac', 'asia-pacific', 'asia pacific', 'north america', 'mena', 'middle east', 'north africa', 'africa', 'asean', 'cis', 'soviet republic', 'north asia', 'south asia', 'central america', 'the caribbean', 'national']), #global regions
    ('c', ['usa', 'canada', 'mexico', 'brazil', 'argentina', 'united kingdom', 'germany', 'france', 'italy', 'spain', 'china', 'india', 'japan', 'south korea', 'australia', 'south africa', 'russia', 'turkey', 'saudi arabia', 'united arab emirates', 'nigeria', 'egypt', 'indonesia', 'malaysia', 'singapore', 'thailand', 'vietnam', 'philippines', 'austria']), #countries
    ('d', ['northwestern', 'northwest', 'north', 'northern', 'northeastern', 'northeast', 'easter', 'east', 'central', 'midwest', 'pacific', 'atlantic', 'mountain', 'plains', 'gulf', 'regional']), #regional areas
    ('e', ['branch']),
]

# Function to apply the rules and find the matching letter for each experience title
def find_geo_division(title, rule_geo):
    for letter, keywords in rule_geo:
        if any(keyword.lower() in title.lower() for keyword in keywords):
            return letter
    return None # Return None if no keywords match

# Apply the function to the 'experience_title' column and create the 'geo_division' column
combined_conditions_df['geo_division'] = combined_conditions_df['experience_title'].apply(lambda title: find_geo_division(title, rule_goe))

empty_list.append(combined_conditions_df)

combined_conditions_df = pd.concat(empty_list, ignore_index=True)
combined_conditions_df.to_csv('/Users/kimmelissabar/Documents/Masterarbeit Schreiben/Code/financial/outputs/financial_all_new3105.csv', index=False)

```

Check default category

```

df_all = pd.read_csv('/Volumes/X9 Pro/Masterarbeit-extern/CSV files created/financial/financial_all_new1805.csv')

# Filter entries where supervision_level is 35
check_35_df = df_all[df_all['supervision_level'] == 35]

```

```

# Count the most appearing strings in experience_title
most_appearing_titles = check_35_df['experience_title'].value_counts()

# Check occurrences of specific words 'finance' and 'bank' in experience_title
finance_count = check_35_df['experience_title'].str.contains(r'\bfinanc\b', case=False, regex=True).sum()
bank_count = check_35_df['experience_title'].str.contains(r'\bbank\b', case=False, regex=True).sum()

# Display results
print("Most appearing titles in 'experience_title':")
print(most_appearing_titles.head(40)) # Display top 40 most appearing titles

print("\nOccurrences of 'financ' in 'experience_title':", finance_count)
print("Occurrences of 'bank' in 'experience_title':", bank_count)

```

Appendix 7: Rule-based Matching – SW Code

Create a loop to execute all files at once

```
import numpy as np
import pandas as pd
import os
from tqdm import tqdm
tqdm.pandas()
from fuzzywuzzy import fuzz

# Specify the directory where your CSV files are located
directory_path = "/Volumes/X9 Pro/Masterarbeit-extern/CSV files created/software/software output"

# Get the list of file names in the specified directory
file_names = os.listdir(directory_path)

# Create an empty DataFrame to store the final processed data
empty_list = []

#predefine filter conditions

def filter_conditions(columns_used_exp):
    cond_company = columns_used_exp['experience_company'].str.strip().str.contains('self at home|self|owner at home|retired|www', case=False)
    cond_position = columns_used_exp['experience_title'].str.strip().str.contains('retired|self|self-employed|self employed|Master', case=False)
    cond_external = columns_used_exp['experience_title'].str.strip().str.contains('nurse|nursing|surgeon|internist|pharmacist|pharmacy|medical|therapist|physician|dental|health|pharmaceutical|paramedic|clinical|patient|navy seal|air force|k-9|pilot|police|police officer|detective|investigator|lieutenant|captain|fbi agent|sergeant|guard|crime|firefighter|fire|rescue|electrician|mechanic|automotive|auto|parts|car|driver|manufacturing|manufacture|logistics|assembly|warehouse|electrical|plant|operator|energy|telecommunication|wireless|farmer|agriculture|crop|environmental|environment|biologist|geologist|climate|paymaster|city|diplomat|fundraising|fundraiser|receptionist|realtor|real estate|property|historian|teacher|lecturer|teaching|phd|professor|postdoc|postgraduate|undergraduate|student|chef|stylist|fashion|wardrobe|store|cashier|photographer|event planner|janitor|actor|actress|artist|singer|musician|editor|author|writer|proofreader|bookseller|waitress|waiter|barista|bartender|pcb|resident assistant|pastor|freelance|no_info|seeking employment|stay at home|dad|mom|mother|wife|family leave|retired|retirement|unemployed|not working|retiree|sabbatical|summer|plumber|various|personal trainer|coach|nanny|mover|library|librarian|linguist', case=False)

    all_conditions = cond_company | cond_position | cond_external
    filtered_chunk = columns_used_exp[~all_conditions]

    return filtered_chunk
pass

#predefine support functions title matching outside the loop
def categorize_support_words(experience_title):
    if pd.notna(experience_title): # Check if the value is not NaN
        words_to_match_support = ['financ', 'banker', 'finanical services', 'buyer', 'teller', 'adjuster', 'examiner', 'accountant', 'account executive', 'credit', 'claims', 'portfolio', 'loan', 'lender', 'tax', 'accounts', 'exchange', 'invoice', 'mortgage', 'processor', 'sales', 'registered representative', 'reg. rep', 'representative', 'registered rep', 'salesforce', 'procurement', 'marketing', 'advertising', 'promotions', 'brand ambassador', 'social media', 'communication', 'content creator', r'\bhr\b', 'human resources', 'recruiter', 'recruiting', 'staffing', 'payroll', 'timekeeping', 'employee benefits', 'compensation', 'wage', 'salary', 'talent acquisition', 'talent', 'workforce', 'human resource', 'hrm', 'recruitment', 'personnel', 'job analysis', 'position descriptor', 'talent acquisition coordinator', 'pension', 'title officer', 'technical sourcer', 'sourcer', 'people operations', 'csr', 'diversity', 'inclusion', 'corporate responsibility', 'policy', 'compliance', 'business development', 'bus. dev', 'research and development', 'r&d', 'governance', 'legal', 'notary public', 'legislative aide', 'patent', 'notary', 'licensing', 'lawyer', 'law clerk', 'paralegal']
```

```

, 'law', 'attorney', 'counsel', r'\bcco\b', r'\bcfo\b', r'\bcho\b', r'\bclo\b', r'\bcmo\b',
r'\bcoo\b', r'\bcpo\b', r'\bcro\b']

    for word in words_to_match_support:
        match_score_1 = fuzz.partial_ratio(word.lower(), experience_title.lower())
        # If a better match is found, update the best_match tuple
        if match_score_1 == 100 and len(word) <= len(experience_title):
            for category, keywords in rules:
                if word in keywords:
                    return category

    return 0

#predefine exact match function
def categorize_exact_words(experience_title):
    if pd.notna(experience_title):
        words_to_match = ['partner', 'managing partner', 'leading partner', 'principal', 'ma
nager', 'mgr', 'president', 'vice president', 'svp', 'evp', 'v.p', r'\v.p\b', 'rvp', 'vp', '
avp', 'ceo', 'cco', 'cdo', 'cfo', 'cho', 'cio', 'clo', 'cmo', 'coo', 'cpo', 'cso', 'cto', '
ccao', 'cceo', 'cos', 'cro', 'sr mng', 'pmo', 'pm', 'gm', 'md', 'lmts', 'head', 'director',
'dir', '6', '5', '4', '3', '2', '1']

        for word in words_to_match:
            if experience_title.lower() == word: # Check for exact match
                for level, keywords in rules_exact:
                    if word in keywords:
                        return level # Return the level based on the matched word

    return 35 # If no match found

#predefine partial ratio function
def categorize_with_partial_ratio(experience_title, rules):
    best_match = (35, 0)
    # Iterate through the rules except the default
    for level, keywords in reversed(rules[:-1]): #reversed filter - double titles cat to hi
ghest
        for keyword in keywords:
            if len(experience_title) >= len(keyword):
                #Calculate the match score using partial_ratio
                match_score = fuzz.partial_ratio(keyword.lower(), experience_title.lower())
                # If a better match is found, update the best_match tuple
                if match_score == 100:
                    best_match = (level, match_score)

    return best_match[0]

#start of the loop
for file_name in file_names:
    if file_name.endswith(".csv") and not file_name.startswith("."):
        # Read the CSV file
        df = pd.read_csv(os.path.join(directory_path, file_name), encoding='ISO-8859-1', on
_bad_lines='warn')
        print("Loop at file:", file_name)

        #columns for new files
        columns_used_exp = df.loc[:, ['name', 'id', 'experience_title', 'experience_company
', 'legal_name', 'experience_start_date', 'experience_end_date', 'experience_duration_short
', 'city', 'type', 'crunchbase_industries', 'industry_dummy']]

# Replace "\N" with NaN (Not-a-Number)
        columns_used_exp.replace('\N', np.nan, inplace=True)
#fill the nan rows with the string "no info"
        columns_used_exp = columns_used_exp.fillna('no_info')
#filter the column type and exclude all investors
        columns_used_exp = columns_used_exp.loc[~columns_used_exp['type'].str.contains('inv

```

```

estor'))]
# Dropping the entire 'type' column
columns_used_exp.drop(columns=['type'], inplace=True)
#exclude double quotation marks
columns_used_exp['legal_name'] = columns_used_exp['legal_name'].str.replace('\"', '')
)

#avoid case sensitivity
columns_used_exp['legal_name'] = columns_used_exp['legal_name'].str.lower()
columns_used_exp['name'] = columns_used_exp['name'].str.lower()
columns_used_exp['experience_title'] = columns_used_exp['experience_title'].str.lower()

columns_used_exp['experience_company'] = columns_used_exp['experience_company'].str.lower()
columns_used_exp['crunchbase_industries'] = columns_used_exp['crunchbase_industries'].str.lower()
columns_used_exp['experience_title'] = columns_used_exp['experience_title'].astype(str)

#exclude all entries without "financial services"-tag
columns_used_exp = columns_used_exp.loc[columns_used_exp['industry_dummy'].astype(str).str.contains('1.0')]
#delete the column
columns_used_exp.drop(columns=['industry_dummy'], inplace=True)

#adding the conditions function that is pre-defined
filter_conditions(columns_used_exp)
combined_conditions_df = filter_conditions(columns_used_exp)

#control print statement
num_rows = len(df)
num_rows_industry = len(columns_used_exp)
num_rows_external = len(combined_conditions_df)
# Printing the number of rows
print(f"The number of rows in the DataFrame started with: {num_rows}; dropped to after crunchbase filter: {num_rows_industry}; is now: {num_rows_external}")

rules = [
(1, ['owner', 'founder', r'\bcreator\b', 'created', 'made', 'ownere', r'\bpartner\b', 'co-owner', 'business partner', 'founding member', 'member of the founding team']),
(2, ['board member', 'board of directions', 'chairman', 'board chair', 'board person']),
(3, ['leading partner', 'managing partner']),
(4, ['ceo', 'chief executive', 'chief executive']),
(5, [r'\bcco\b', r'\bcdo\b', r'\bcfo\b', r'\bcho\b', r'\bcio\b', r'\bclo\b', r'\bcmo\b', r'\bcoo\b', r'\bcpo\b', r'\bcso\b', r'\bcto\b', 'ccao', 'cceo', r'\bcos\b', r'\bro\b', 'chief officer', 'chief\s+\w+\s+officer', 'chief technology officer', 'chief operating officer', 'chief people officer', 'chief marketing officer', 'chief staff officer', 'chief financial officer', 'chief revenue officer', 'chief product officer', 'chief information officer', 'chief sales officer', 'chief of staff', 'chief investment officer', 'chief strategy officer']),
(6, ['president', 'presidente']),
(7, [r'md\b', 'managing director']),
(8.1, [r'evp\b', 'executive vp', 'executive vice president', ' executive vice-president', 'exec vice pres']),
(8.2, ['senior vice president', 'sr vice president', 'snrvp', 'svp', 'senior v.p.', 'sr. vice president', r'senior v\b', 'sr vp', r'sr. \bvp\b', r'\bsenior\s+\w+\s+vp\b']),
(8.3, [r'\bvp\b', r'\bv\.?p\b', 'vice president', 'vice-president', 'vice presidente', 'vice pres', r'\bv\b\.?s?p\b', r'\bvp\b']),
(8.4, ['avp', 'assistant vice president', 'junior vice president', r'\bjr. vp\b', r'\bjn\b\s+vp\b']),
(9, [r'fellow\s+\w+', r'\bfellow\b']), #fellow
(10, ['distinguished', 'tier 7', r'7\b', r'\bvii\b']),
(11.2, ['senior principal engineer', 'senior principal\s+\w+']),
(11.3, ['principal engineer', r'\bprincipal\s+\w+', r'6\b', 'tier 6', r'\bvi\b'])

```

```

),
    (12.1, ['executive director', 'exec. dir', 'exec director']),
    (12.2, ['senior director', 'sr. director', r'\bsenior\s+\w+\s+director\b']),
    (12.3, ['director', 'directo', 'diercto', r'dir\b', 'dierctor']),
    (12.4, ['junior director', r'\bjun dir\b', 'assistant director', r'\bjunior\s+\w+\s+director\b', r'\bassistent\s+\w+\s+director\b', 'jr. director', r'\bjr dir', r'\basst dir\b', r'asst\s+\dir' ]),
    (13, [r'\bhead\s+\w+', r'\bheads\b', r'\bhead\s+\of\w+', 'r\w+\s+\head\s+\w+', 'r\w+\s+\head', 'head of' ]),
    (14.01, ['program manager']),
    (14.02, ['project manager', r'\bpmo\b', 'project management officer', r'senior \bpm\b', r'pm\b', r'pm\b manager']), #project managers all
    (14.03, ['system administrator', r'qa pm\b', 'pm/se', 'pm windows', 'product manager', 'technical account manager', 'sw manager', 'software manager', 'quality assurance manager', 'qa manager', 'software developer manager', 'software engineering manager', 'software development manager']), #manager software (department managers)
    (14.1, ['executive manager']),
    (14.2, ['senior manager', 'senior management team', 'sn manager', 'sr. manager', 'sr mng', 'sr manager', r'\bsenior\s+\w+\s+\manager\b', r'\bsr.\s+\w+\s+\manager\b', r'\bsenior\s+\w+\s+\w+\s+\manager\b']),
    (14.3, ['manager', 'manager', 'manger', 'mananger', 'mgr', r'gm\b', 'department manager', 'division manager']),
    (14.4, ['junior manager', 'assistant manager', r'\bassistent\s+\w+\s+\manager\b', 'management assistant', 'coordinator']),
    (15.2, ['senior lead', 'senior leader', 'senior lead member of', 'team manager' ]),
    (15.3, ['lead', 'leader', 'lmts', 'lead member of', 'team leader', r'\bv\b', r'b5\b', 'tier 5']),
    (16.2, ['senior staff\s+\w+', 'smts', 'senior member of technical staff' ]),
    (16.3, [r'staff\s+\w+', r'staff\s+\w+\s+\w+', 'mts', 'member of technical staff', 'member technical staff', r'\biv\b', r'4\b', 'tier 4']),
    (17, ['supervisor', 'instructor', 'instructer', 'mentor']),
    #18.2 senior specialists, 3, iii
    (18.01, ['analyst', 'computer system analyst', 'information security analyst', 'system specialist', 'vendor', 'application package specialist', 'system integrator', 'software analyst', 'applications analyst', 'computer systems engineer', 'system engineer', 'systems analyst', 'systems planner', 'programmer analyst', 'systems architect', 'security specialist', 'systems security analyst', 'risk specialist', 'network security analyst', 'sap', 'security officer', 'six sigma', 'security advisor', 'enumerator']), #analyst
    (18.02, ['research scientist', 'computational theory scientist', 'computer scientist', 'programming methodology', 'languages researcher', 'data scientist', 'user experience researcher', 'ux researcher', 'researcher', 'scientist']),
    (18.03, ['support', 'helpdesk', 'help desk', 'support specialist', 'network support', 'user support', 'client support', 'system installer', 'systems installer', 'services advisor', 'service advisor', 'game advisor', 'product advisor', 'support technician', 'network technician', 'help desk representative', 'help desk specialist', 'help desk analyst', 'help desk technician', 'tech support', 'field engineer', 'pfe', 'pc tech', 'guest advisor', 'site reliability engineer', 'sre', 'support analyst', 'technical advisor', 'client executive', 'client service', 'customer service rep']), #support
    (18.04, ['administrator', 'network', 'database', 'computer network architect', 'database architect', 'database administrator', 'system administrator', 'technology architect', 'sql dba', r'dba\b', 'peoplesoft dba', 'solution architect', 'platform engineer', 'platform eng', 'platform specialist', 'platform architect', 'solutions architect', 'computer network engineer', 'network designer', 'network developer', 'network engineer', 'processing planner', 'database administration manager', 'database coordinator', 'database programmer', 'database security administrator', 'data architect', 'data integration specialist', 'data warehousing specialist', 'database developer', 'lan administrator', 'lan systems administrator', 'network administrator', 'network analyst', 'network coordinator', 'network support coordinator', 'network systems administrator', 'network systems coordinator', 'wan systems administrator', 'peoplesoft administrator', 'oracle dba', 'systems administrator', 'architect', 'linux', 'middleware administrator', 'oracle app', 'unix admin', 'lotus notes administrator' ]), #admin & architects
    (18.051, ['computer programmer', 'applications programmer', 'computer language coder', 'coder', 'it programmer', 'mainframe programmer', 'systems programmer', 'unix system', 'programming']), #programmer
    (18.052, ['software developer', 'software development', 'sql processor', 'software engineer', 'software eng', 'software engineer', 'pfe', 'solution specialist professional' ]),

```



```

, 'application package designer', 'software solutions specialist', 'pde', 'product development engineer', 'php developer', 'software architect', 'integration engineer', 'applications developer', 'application developer', 'applications engineer', 'systems engineer', 'systems engineering', 'applications developer', 'applications architect', 'applications designer', 'software designer', 'systems software specialist', r's/w\b', 'sql', 'sde', 'software design', 'pts', 'software professional', 'ios developer', 'mobile developer', 'solutions engineer', 'oracle app dev', 'sft. eng', 'software eng', r'se\b', 'unix', 'software administrator']], #software

(18.053, ['quality assurance', 'performance', 'software quality assurance', 'quality assurance analyst', 'tester', 'performance quality', 'product safety', 'psr', 'product safety representative', 'security professional', 'qa analyst', r'qa\b', 'performance analyst', 'qa engineer', 'performance engineer', 'software quality control specialist', 'software quality engineer', 'software test engineer', 'qu tester', 'performance specialist', 'quality assurance engineer', 'security engineer', 'quality engineer', 'test engineer', 'product safety rep', r'psr\b', 'sqa', r'qa\b', 'quality control', 'ste', 'sqe', 'qa advisor', 'system test']), #tester/ quality assurance

(18.054, ['web developer', 'intranet developer', 'web applications developer', 'web architect', 'web content developer', 'ux engineer', 'web dev']), #web

(18.055, ['web designer', 'interface designer', 'digital designer', 'piping designer', 'painter', 'product designer', 'graphic web designer', 'web content specialist', 'ux designer', r'ui\b designer', 'designer', 'design eng', 'user interface', 'ux design', 'ui/ux', 'hardware design', 'graphic designer']), #design

(18.06, ['consultant', 'it consultant', 'systems consultant', 'software consultant', 'solutions consultant', 'itc']),

(18.07, ['product administrator', 'project specialist', 'business planner', 'print specialist', 'computer specialist', 'strategist', 'storage solutions specialist', 'processing specialist', 'specialist', 'performer', 'technologist', 'planner', 'product specialist', 'producer', 'process engineer', 'computer console operator', 'computer laboratory technician', 'data center operator', 'paycom specialist', 'process expert', 'thinkpad specialist', r'it\b specialist', 'engineer', 'developer', 'programmer', 'pc strategist', 'technician', 'pc tech', 'technician', 'tech info spec', 'devops', 'sme', 'product owner', r'dev\b', r'dev\b', 'adjuster', 'examiner', 'accountant', 'processor', 'lawyer', 'legal counsel', 'advisor', 'title officer', 'coordinator', r'2\b', r'bii\b', 'tier 2']), #others

(19.01, ['associate', 'engineer in test', 'sdet', 'machinist', 'in test', r'1\b', r'bi\b', 'tier 1', 'junior', 'jr.', r'bjn\b']),

(19.02, ['personal assistant', 'executive assistant', 'assistant', 'assistan', 'assitant', 'bpa', 'alumni secretary', 'office secretary', 'personal secretary', 'asst', 'office admin', 'secretary', 'administrative assistant']),

(19.03, ['clerk', r'brep\b', 'employee', 'worker', 'agent', 'paraprofessional', 'mac genius', 'machinest', 'maintenance', 'technical sourcer', 'technical aide']),

(20, ['trainee', 'trainie', 'research student', r'bintern\b', 'volunteer', 'training program', 'mentee', 'apprentice']),

(30.01, ['financ', 'financial services', 'credit', 'claims', 'portfolio', 'loan', 'lender', 'tax', 'accounts', 'exchange', 'invoice', 'mortgage', 'banker', 'buyer', 'teller']),

(30.02, ['sales', 'sales representative', 'registered representative', 'reg. rep', 'representative', r'brep\b', 'registered rep', 'sales manager', 'sales associate manager', 'sales training manager', 'sales consultant', 'salesforce', 'sales associate', 'sales agent', 'sales assistant', 'procurement', 'account executive']),

(30.03, ['marketing', 'advertising', 'promotions', 'social media', 'communication', 'brand ambassador', 'content creator']),

(30.04, [r'hr\b', 'human resource', 'recruiter', 'recruiting', 'staffing', 'payroll', 'timekeeping', 'employee benefits', 'compensation', 'wage', 'salary', 'talent acquisition', 'talent', 'talent professional', 'workforce management', 'human resource manager', 'hrm', 'recruitment manager', 'payroll manager', 'personnel', 'job analysis', 'position descriptor', 'talent acquisition coordinator', 'payroll analyst', 'payroll assistant', 'pension administrator', 'title officer', 'technical sourcer', 'sourcer', 'talent management', 'people operations']),

(30.05, ['csr', 'diversity', 'inclusion', 'csr', 'corporate responsibility', 'policy', 'compliance', 'business development', 'bus. dev', 'research and development', 'r&d', 'governance']),

(30.06, ['notary public', 'legal liaison', 'legal operations', 'legal services', 'legal translator', 'legislative aide', 'law', 'lawyer', 'attorney', 'general counsel', 'legal counsel', 'legal advisor', 'legal administrator', 'legal researcher', 'legal ninja', 'legal secretary', 'legal translator', 'patent', 'legal', 'notary', 'licensing', 'lawyer', 'law clerk', 'paralegal', 'attorney', 'counsel', 'internal audit']),

(35, []) # Default category

```



```

]

#functions to categorize job titles/ pre-defined
combined_conditions_df = combined_conditions_df.copy() # Create a copy to avoid the
SettingWithCopyWarning
#call predefined function
combined_conditions_df['support_functions'] = combined_conditions_df['experience_title'].apply(categorize_support_words)

category_35 = combined_conditions_df

# use of rules exact, bc word boundary usage in the rules lead to a no match in the == word
case. No use of word boundaries in the fuzz matching leads to a match with random word like
cooperation = coo
rules_exact = [
    (1, ['partner', 'creator']),
    (4, ['ceo']),
    (5, ['cco', 'cdo', 'cfo', 'cho', 'cio', 'clo', 'cmo', 'coo', 'cpo', 'cso', 'cto', 'ccao', 'cceo', 'cos', 'cro']),
    (6, ['president']),
    (7, ['md']),
    (8.1, ['evp']),
    (8.2, ['snrvp', 'svp', 'sr vp', 'sr. vp']),
    (8.3, ['vp', r'v\.\?p', 'vice president', 'vp']),
    (8.4, ['avp', 'jr. vp']),
    (9, ['fellow']),
    (10, ['distinguished', '7', 'tier 7']),
    (11.3, ['principal', '6', 'tier 6']),
    (12.3, ['director', 'dir']),
    (12.4, ['jun dir', 'jr dir', 'asst dir' ]),
    (13, ['head']),
    (14.01, ['program manager']),
    (14.02, ['pmo', 'pm']),
    (14.03, ['qa pm', 'pm/se']),
    (14.3, ['manager', 'manager', 'manger', 'mananger', 'mgr', 'gm']),
    (15.3, ['lead', 'leader', 'lmts', '5', 'tier 5']),
    (16.2, ['smts']),
    (16.3, ['mts', 'staff', '4', 'tier 4']),
    (18.07, ['2', 'tier 2']),
    (19.01, ['1', 'tier 1']),
    (30.01, ['rep']),
    (30.04, ['hr']),
    (30.05, ['csr', 'r&d', 'bus dev']),
]

category_35 = category_35.copy()
# Apply categorization based on exact words only for rows categorized as 35 (default)
by external matching
#categorize_exact_words(experience_title)
category_35['exact_level'] = category_35.apply(lambda row: categorize_exact_words(row['experience_title']), axis=1)

combined_conditions_df.loc[category_35.index, 'supervision_level'] = category_35['exact_level']

new_df = category_35[category_35['exact_level'] == 35].copy()

#include progress bar: progress_apply
new_df['level_three'] = new_df['experience_title'].progress_apply(lambda x: categorize_with_partial_ratio(x, rules))

# Update the original DataFrame with the re-categorized values
combined_conditions_df.loc[new_df.index, 'supervision_level'] = new_df['level_three']

```

```

        #extra filter of vp
        combined_conditions_df['senior_supervision_level'] = combined_conditions_df['supervision_level']

        # regular pattern for 'vice president'
        pattern_vp = r'vice president\s+\w+|vp\.|vp\b|v\.p\.|vice president|vice president,|vp,'

        mask = combined_conditions_df['experience_title'].str.contains(pattern_vp, case=False, regex=True)

        all_below_vp = combined_conditions_df['senior_supervision_level'].astype(float).isin([6, 35, *range(9, 20)])

        # update based on mask
        combined_conditions_df.loc[mask & all_below_vp, 'senior_supervision_level'] = 8.3

        combined_conditions_df.loc[combined_conditions_df.index, 'supervision_level'] = combined_conditions_df['senior_supervision_level']

        #due to bottom to top filter lowest cat. might be not fully filled
        pattern_jun = r'\bassistant\b|\basst\b| assistant\s+\w+\s+\w+\s+\w+\s+\w+\b|assistant\s+\w+\s+\w+\b|asst.\s+\w+\s+\w+\b|\bjunior\s+\b|\bjun\b|\bjr\.\b|assistant|junior'

        mask_2 = combined_conditions_df['experience_title'].str.contains(pattern_jun, case=False, regex=True)

        #vp
        all_junvp = combined_conditions_df['senior_supervision_level'].astype(float).isin([6, 8.3])
        combined_conditions_df.loc[mask_2 & all_junvp, 'supervision_level'] = 8.4

        #director
        all_jundirector = combined_conditions_df['senior_supervision_level'] == 12.3
        combined_conditions_df.loc[mask_2 & all_jundirector, 'supervision_level'] = 12.4

#junior specialists
        condition_15 = combined_conditions_df['senior_supervision_level'].astype(float).between(18.01, 18.07)

        combined_conditions_df.loc[condition_15 & combined_conditions_df['experience_title'].str.contains(r'\bsenior\b|\bsr\.\b|\bsn\.\b|\bsr\b|\b3\b|\biii\b|\btier 3\b', case=False), 'supervision_level'] = 18.2

#junior dev level i und 1
        condition_i = combined_conditions_df['senior_supervision_level'].astype(float).isin([18.2]) | combined_conditions_df['senior_supervision_level'].astype(float).between(18.01, 18.07)

        combined_conditions_df.loc[condition_i & combined_conditions_df['experience_title'].str.contains(r'\bi\b|\b1\b|\bi\.\b|\bjunior\b|\bassociate\b|\bjr\b|\bjn\.\b|\bjn\b', case=False), 'supervision_level'] = 19.01

        # Define the rules as provided in the prompt
        rule_goe = [
            ('a', ['global']),
            ('b', ['emea', 'usa', 'europe', 'apac', 'asia-pacific', 'asia pacific', 'north america', 'mena', 'middle east', 'north africa', 'africa', 'asean', 'cis', 'soviet republic', 'north asia', 'south asia', 'central america', 'the caribbean', 'national', 'international']), #global regions
            ('c', ['usa', 'canada', 'mexico', 'brazil', 'argentina', 'united kingdom', 'germany', 'france', 'italy', 'spain', 'china', 'india', 'japan', 'south korea', 'australia', 'south africa', 'russia', 'turkey', 'saudi arabia', 'united arab emirates', 'nigeria', 'egypt

```

```

', 'indonesia', 'malaysia', 'singapore', 'thailand', 'vietnam', 'philippines', 'austria']],
#countries
    ('d', ['northwestern', 'northwest', 'north', 'northern', 'northeastern', 'north
east', 'easter', 'east', 'central', 'midwest', 'pacific', 'atlantic', 'mountain', 'plains',
'gulf', 'regional']), #areas
    ('e', ['branch']),
]

# Function to apply the rules and find the matching letter for each experience title
def find_geo_division(title, rules_g):
    for letter, keywords in rules_g:
        if any(keyword.lower() in title.lower() for keyword in keywords):
            return letter
    return None # Return None if no keywords match
# Apply the function to the 'experience_title' column and create the 'geo_division' column
combined_conditions_df['geo_division'] = combined_conditions_df['experience_title']
.apply(lambda title: find_geo_division(title, rule_goe))

empty_list.append(combined_conditions_df)

combined_conditions_df = pd.concat(empty_list, ignore_index=True)
combined_conditions_df.to_csv('/Volumes/X9 Pro/Masterarbeit-extern/CSV files created/software
re/software_all_new_220524.csv', index=False)

#read output file
import pandas as pd
df_all = pd.read_csv('/Volumes/X9 Pro/Masterarbeit-extern/CSV files created/software/software
re_all_new_220524.csv')

```

Check default category

```

check_35_df = df_all[df_all['supervision_level'] == 35]

# Count the most appearing strings in experience_title
most_appearing_titles = check_35_df['experience_title'].value_counts()

# Check occurrences of specific words 'finance' and 'bank' in experience_title
finance_count = check_35_df['experience_title'].str.contains(r'\bsoftware\b', case=False, r
egex=True).sum()
bank_count = check_35_df['experience_title'].str.contains(r'\bsystem\b', case=False, regex=
True).sum()

# Display results
print("Most appearing titles in 'experience_title':")
print(most_appearing_titles.head(40)) # Display top 40 most appearing titles

print("\nOccurrences of 'software' in 'experience_title':", finance_count)
print("Occurrences of 'system' in 'experience_title':", bank_count)

#count unique companies in output file
df_all.info()
experience_company_list = df_all['experience_company'].unique().tolist()
len(experience_company_list)

#count unique titles in output file
experience_title_list = df_all['experience_title'].unique().tolist()
len(experience_title_list)

experience_company_list

```

Appendix 8: Interview Guide – English

Introduction:

- Goal of the interview:

The aim of this master's thesis is to develop and validate a method for determining hierarchical depth in companies in the Financial Services/ Software industry. Over the past few months, I have developed this method using data, primarily based on job titles, and have now reached the validation phase. This interview is intended to confirm or refine the categorization or to identify and improve potential misconceptions.

The interview will not be used to specifically analyze your company, the insights gained will only be used to gain a better overview of the industry and to validate theoretical assumptions with practical knowledge.

Company:

- Control question: Are you working in the financial services/software industry?
- For which company?
- How many employees does the company employ (approx.)?

Background/personal questions:

- In which department do you work?

Can you explain the different departments or divisions of the company?
(here especially core/support functions)

- Which position do you currently hold in the company (job title)?
- Do you have management responsibility in your current position?
- Promotions (which ones, when/after which seniority)?
- Which career path applies best to you?
 1. Specialist career path
 2. Project management
 3. Management career path

Questions about the hierarchy:

- Who do you report to? And who reports to you? (only job titles, no names)
-

reminder of the actual goal

From here on, work together to create a diagram of the organizational hierarchy

Possibly helpful:

- How are the career paths in your department/ in your professional path/ in this company usually structured?
- Keep in mind: Hierarchy steps in the case of a change of supervisor
- What exchange exists between the career paths (e.g. manager with senior engineer?)
- Which departments or teams work closely with your department/other departments/your line manager?
- In my assumption position X is inferior/ superior to position Y, does this correspond to reality?

Industry wide:

- Have you had experience of this hierarchy being found in other companies?
- Do you know whether you would join a (competitor) company with a different title/level?

Conclusion

Appendix 9: Interview Guide – German

Einleitung:

Ziel des Interviews:

Ziel dieser Masterarbeit ist es, für die Financial Services/ Software-Industrie eine Methode zur Bestimmung der Hierarchietiefe in Unternehmen zu entwickeln und diese zu validieren. In den letzten Monaten habe ich diese Methode mit Hilfe von Daten, primär basierend auf Jobtiteln entwickelt und befinde mich nun in der Phase der Validierung. Dieses Interview soll dazu dienen die Kategorisierung zu bestätigen, zu verfeinern oder auch potenzielle Fehlannahmen zu erkennen und zu verbessern.

Das Interview dient nicht zur spezifischen Analyse Ihres Unternehmens, die gewonnen Erkenntnisse werden ausschließlich dazu verwendet, eine bessere Übersicht der Industrie zu erlangen und um theoretische Annahmen durch praktisches Wissen zu validieren.

Erklärung Ablauf:

Teil 1: Im ersten Teil können und sollen die Fragen mit einzelnen Wörtern/ in 1-2 Sätzen beantwortet werden.

Teil 2: Im zweiten Teil geht es um das gemeinsame Erarbeiten der Unternehmenshierarchie, ein lockerer Gespräch Fluss ist hier gewünscht. Zuvor abgefragte individuelle Informationen sollen in die Unternehmensstruktur eingearbeitet werden.

Unternehmen:

- Kontrollfrage: Sind Sie im Financial Services/ Software Sektor tätig?
- Bei welchem Unternehmen?
- Wie viele Mitarbeitende hat das Unternehmen (ca.)?

Hintergrundfragen/ Persönliches:

- Wie heißt die Abteilung, für die Sie tätig sind?
- Welche Position haben Sie aktuell im Unternehmen (Jobbezeichnung)?
- Besitzen Sie innerhalb der aktuellen Position Führungsverantwortung?
- Beförderungen:
wie hat sich der eigene Titel geändert?
hat sich die Führungsperson geändert – wenn ja, von Titel zu?

- Welcher Karriereweg trifft am ehesten auf Sie zu:

Fachlaufbahn

- Projektmanagement
- Führungslaufbahn

Fragen zur Hierarchie:

Teil 2

Hier nochmal Erinnerung an das eigentliche Ziel

Zeigen von 1) Organisationsorganigrammen, 2) typische Machtstrukturen, 3) Hierarchische Gliederung der Struktur

→ Ab hier gemeinsames erarbeiten einer Abbildung der Unternehmenshierarchie

Eventuell hilfreich:

- Können Sie die verschiedenen Abteilungen oder Bereiche des Unternehmens erläutern?
(hier v.a. core/support functions)
- An wen berichten Sie? Und wer berichtet an Sie? (lediglich Jobbezeichnungen, keine Namen)
An welche Position/ Job Titel berichtet ihr*e Vorgesetzte*r, und deren?
- Wie sind üblicherweise die Karrierewege in ihrer Abteilung/ in ihrem Karriereweg/ in diesem Unternehmen gestaltet?
- *Im Hinterkopf: Hierarchiesprung bei wechselnder Führungsperson*
- Welcher Austausch besteht zwischen den Karrierewegen (z.B. Manager mit Senior Eng.?)
- Welche Abteilungen oder Teams arbeiten eng mit Ihrer Abteilung/anderen Abteilungen/ ihrem/r Vorgesetzten zusammen?
- In meiner Annahme ist Position X Position Y unter-/ übergestellt, entspricht dies der Realität?

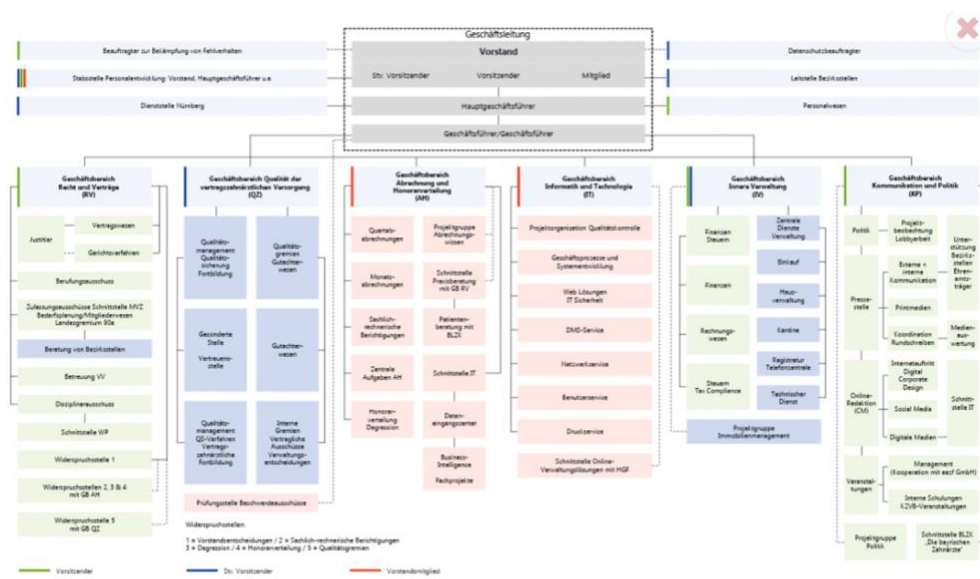
Industrieweit:

- Haben Sie die Erfahrung gemacht, dass diese Hierarchie auch in anderen Unternehmen zu finden ist?
- Wissen Sie, ob Sie bei einem (Konkurrenz-)Unternehmen mit einem anderen Titel/Level einsteigen würden?

Schluss

- ➔ Bewusst kein Erfragen von persönlichen Daten, wie Name, Alter, Geschlecht, Ethnie
- ➔ Ggf. anbieten, dass Unternehmensname nicht genannt wird.

Appendix 10: Interview – Powerpoint Presentation



Source: Business Wissen, 2024, extracted from: <https://www.business-wissen.de/hb/organigramm-erstellen-beispiel-definition-und-vorlagen/>, date 05.01.2024

Figure 14 Interview pptx slide 1



Figure 15 Interview pptx slide 2



Figure 16 Interview pptx slide 3

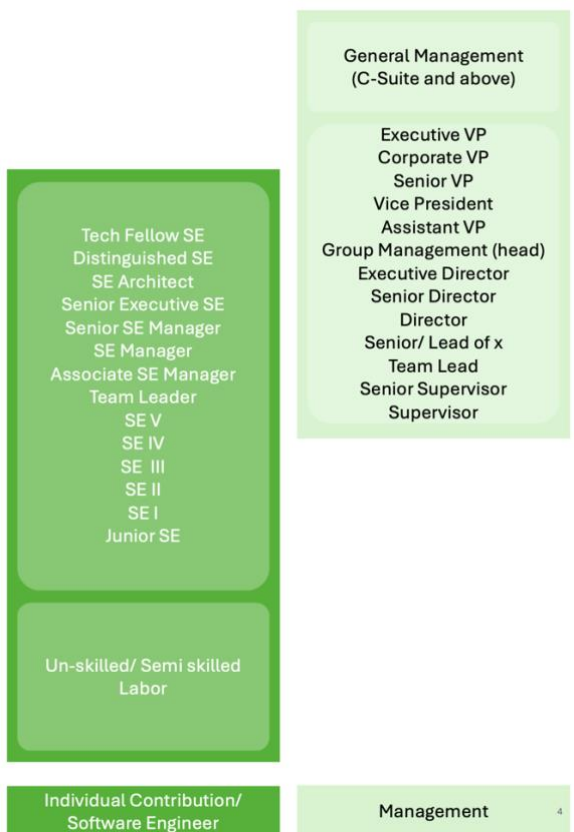


Figure 17 Interview pptx slide 4

Categories Financial Services

- 1. Owner ; Partner
 - 2. Associate partner; Principal
 - 3. Board member
 - 4. CEO
 - 5. Managing director
 - 6. C-Suite
 - 7. President
 - 8.1 Executive vp
 - 8.2 Senior vp
 - 8.3 VP
 - 8.4 Assistant vp; junior vp
 - 9. Head
 - 10.1 Executive director
 - 10.2 Senior director
 - 10.3 Director
 - 10.4 Junior director
 - 11.x Manager
 - 12 Lead
 - 13. Senior specialists
 - 13.01-13.10 Specialist
 - 14. Supervisor
 - 15. Senior worker
 - 15.01-13.10 Workers
 - 16. Trainee
-
- 30.1-30.6 support functions (software, sales, marketing, HR, law, bus. Dev.)
 - 34 client service (unsure → partly software; partly only support functions)
 - 35 default

5

Figure 18 Interview pptx slide 5

Categories Software

- 1. Owner ; Partner
- 2. Associate partner; Principal
- 3. Board member
- 4. CEO
- 5. Managing director
- 6. C-Suite
- 7. President
- 8.1 Executive vp
- 8.2 Senior vp
- 8.3 VP
- 8.4 Assistant vp; junior vp
- 9. Head
- 10.1 Executive director
- 10.2 Senior director
- 10.3 Director
- 10.4 Junior director
- 11.1 Senior fellow
- 11.2 Fellow
- 12. Distinguished
- 13. x Manager
- 14. Lead , V
- 15.1 Senior staff
- 15.2 Staff, IV
- 16. Supervisor
- 17. Senior specialists, III
- 17.1-17.7 Specialists , II
- 18. Junior specialists , I
- 19. Trainee

- 30.1-30.6 support functions (finance, sales, marketing, HR, law, bus. Dev.)
- 35 default

(maybe additional personal assistant + secretary)

6

Figure 19 Interview pptx slide 6

Manager Categories

Area	Seniority	Skills
• 11.001 global • 11.002 national • 11.003 market • 11.004 regional • 11.005 division • 11.006 office	• 11.01 executive • 11.02 senior • 11.03 junior	• 11.1 program • 11.2 project • 11.3 support functions e.g. sales manager • 11.4 general manager

7

Figure 20 Interview pptx slide 7

Appendix 11: Interview Results – Organizational Charts

Interview 1–ZEB (DE)

Management Career Path
Senior Consultant

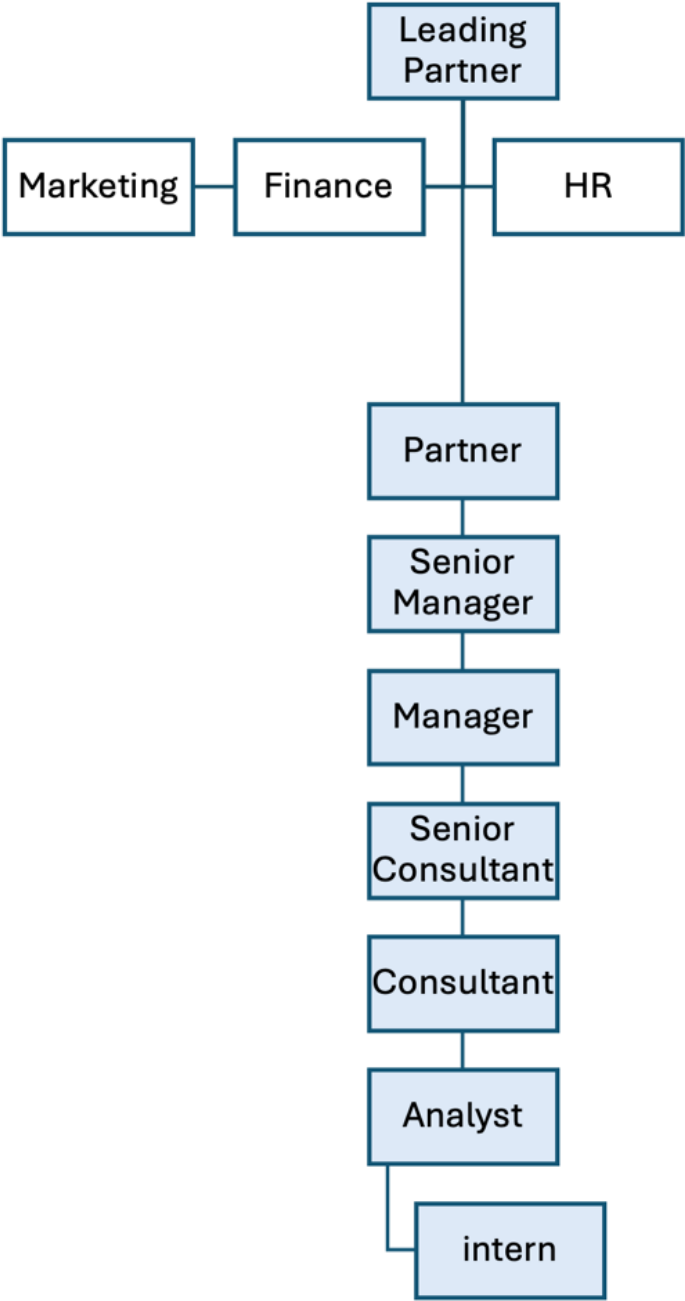


Figure 21 Organization chart interview 1

Interview 2—ready to order

Management Career Path
Marketing Manager

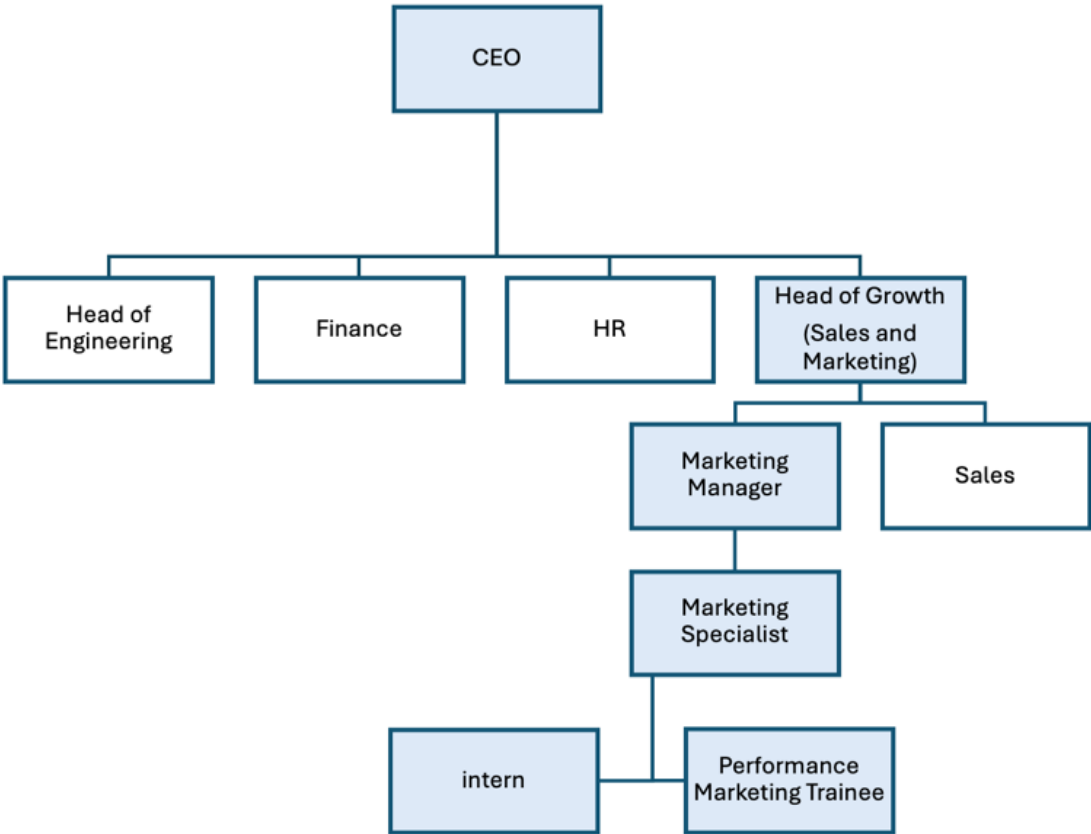


Figure 22 Organization chart interview 2

Interview 3–American Express (AT)

Individual Contribution Path Digital Aquisition Coordinator

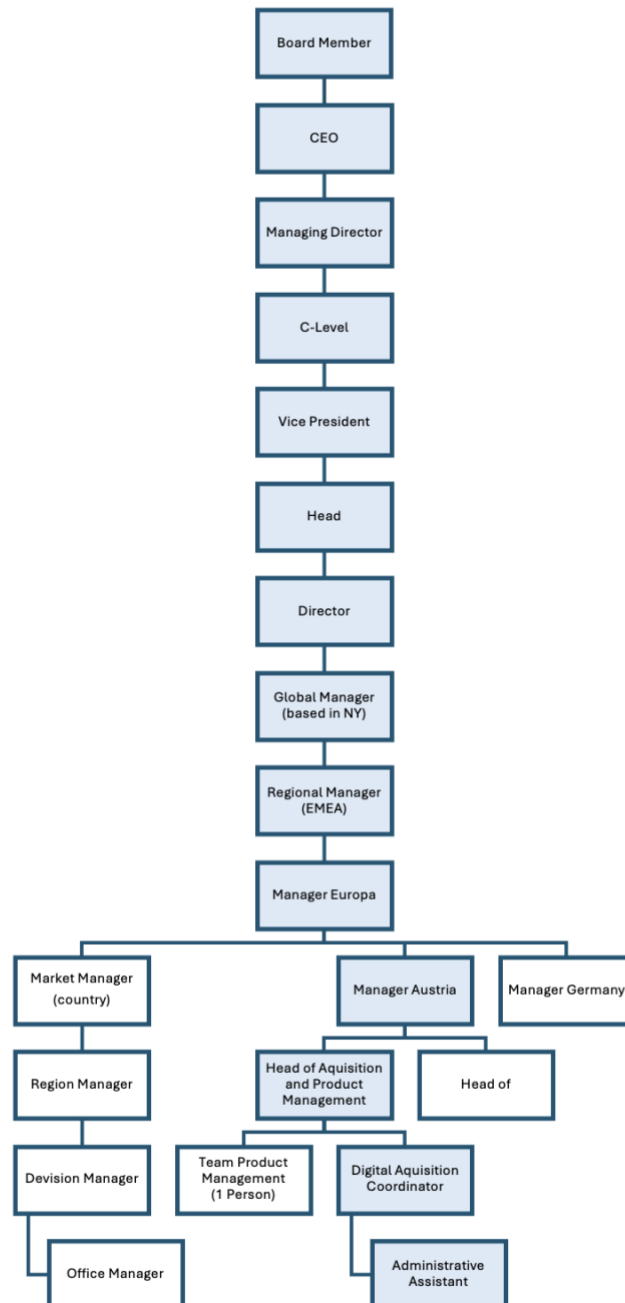


Figure 23 Organization chart interview 3

Interview 4–Kreditgeschäft Kreissparkassen (DE)

Management Path

Landesdirektion West

*the job titles are not translated, because the organization decided consciously not to use english titles

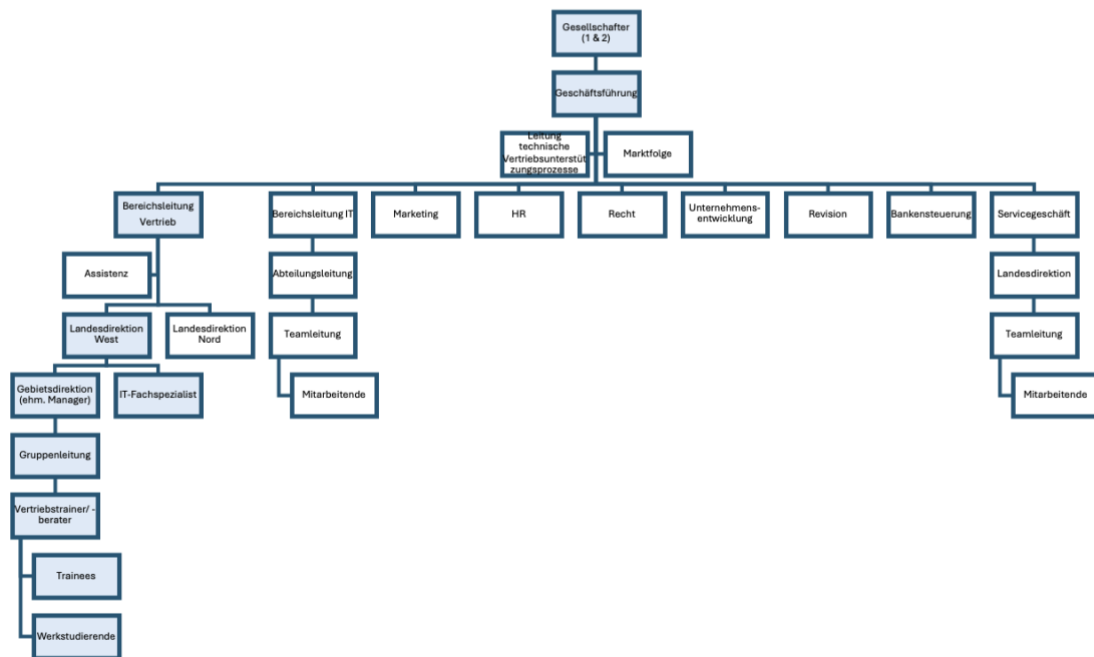


Figure 24 Organization chart interview 4

Interview 5–N26 (AT)

Individual Contribution Path Lead Engineer

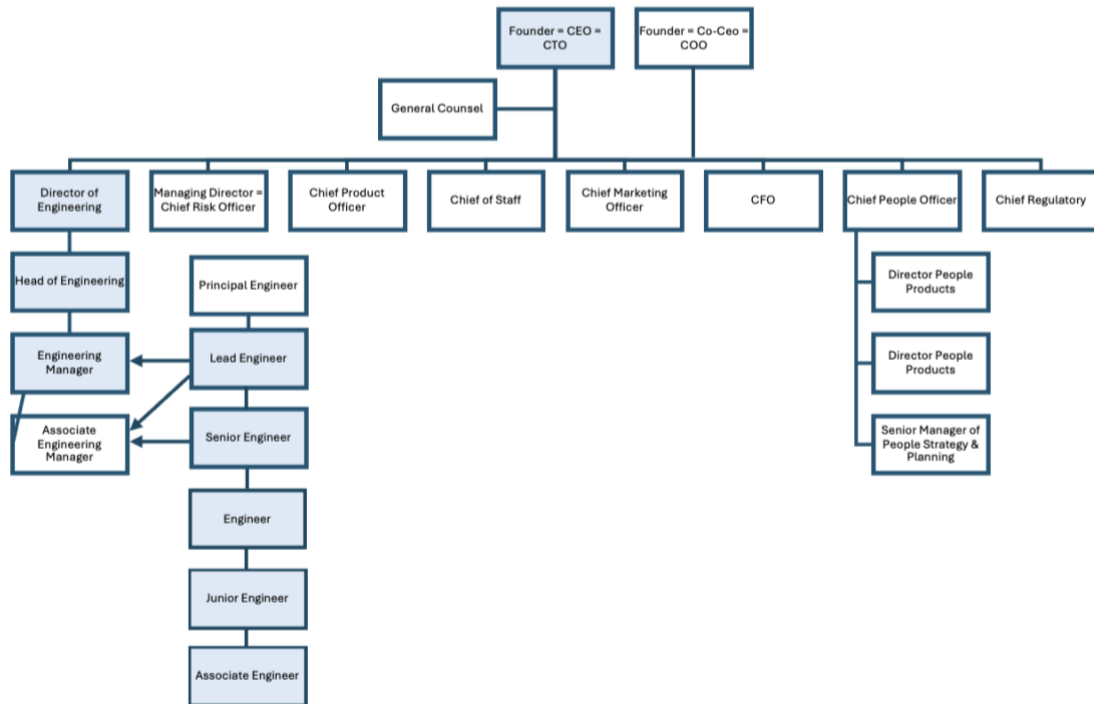
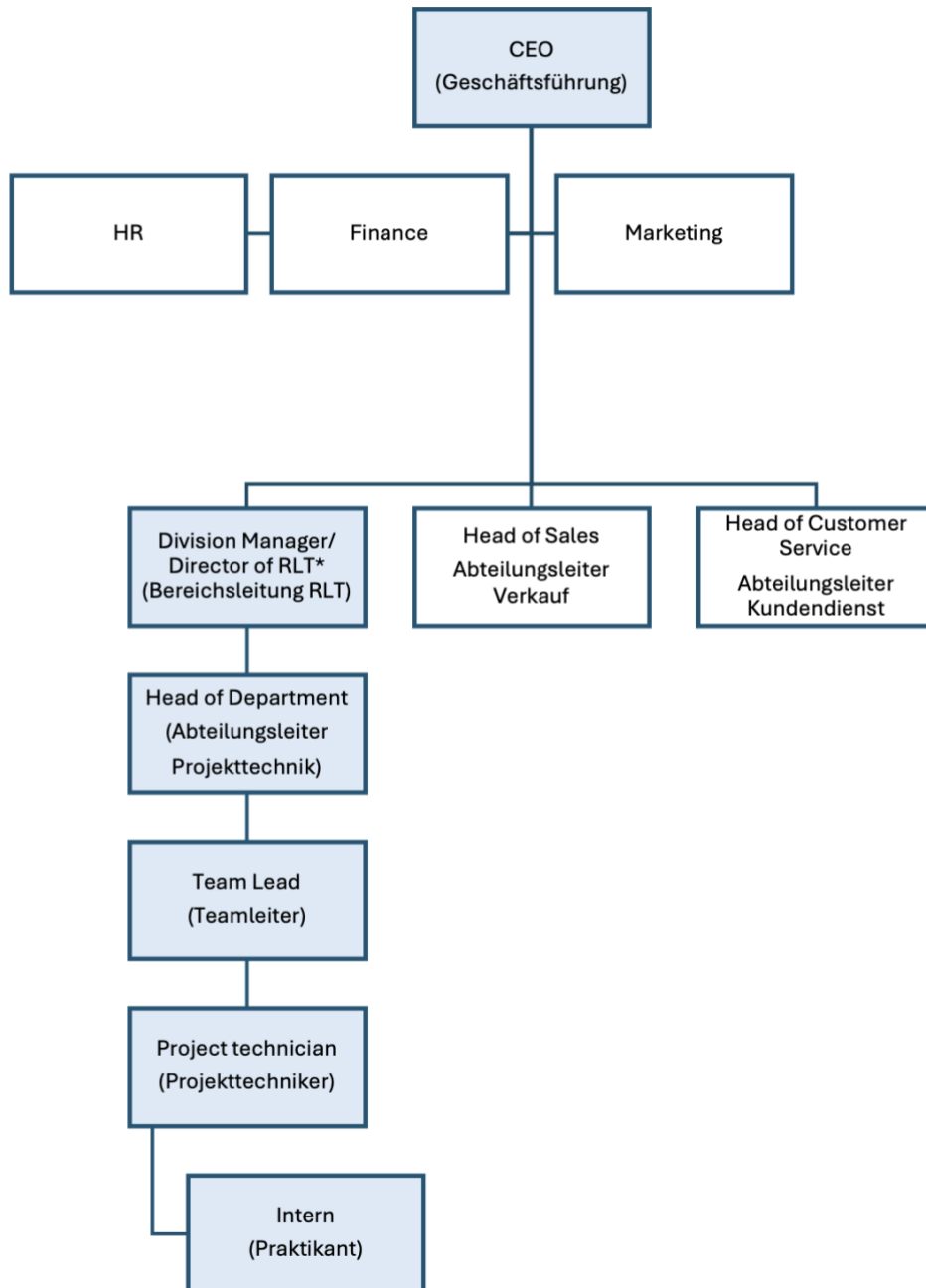


Figure 25 Organization chart interview 5

Interview 6–Walter Bösch GMBH & Co. KG (AT)

Individual Contribution Path

Technician



*Raum und Lufttechnik

Figure 26 Organization chart interview 6

Interview 7–SAP (DE)

Management Path Associate Developer

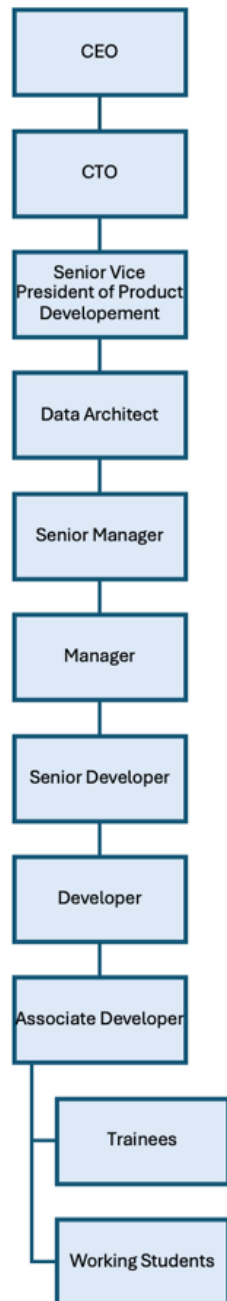


Figure 27 Organization chart interview 7

Interview 8–Platform X GmbH & Co. KG (DE)

Individual Contribution Path
Platform Engineer

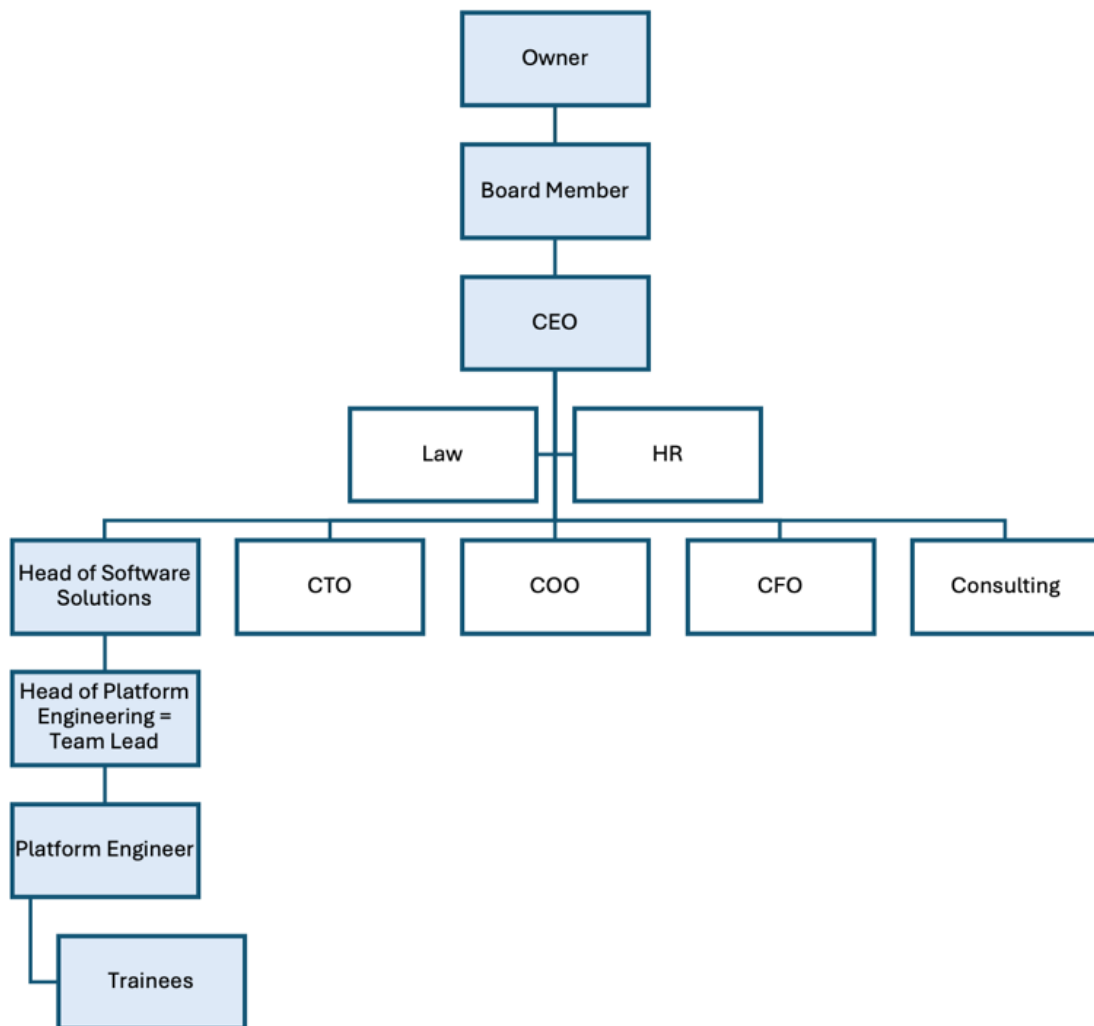


Figure 28 Organization chart interview 8

Appendix 12: Interview Transcripts

Interview 1 – FS industry

00:00:00 Interviewer

Are you working in the financial service industry?

00:00:05 Interviewee

Yes, I'm working currently as a consultant in the financial services industry with a focus on large banks.

00:00:12 Interviewer

Perfect. And for which company?

00:00:16 Interviewee

For ZEB, which is a medium-sized German consultancy company, but I'm working in Vienna, and the company has multiple offices around Europe.

00:00:27 Interviewer

How many employees does the company have? Just approximately like, is it a startup or a bigger company?

00:00:34 Interviewee

Around 1000 employees.

00:00:38 Interviewer

Perfect. Can you explain a little bit further in which department do you work?

00:00:45 Interviewee

Yeah, sure. I work as a senior consultant. So as the second hierarchy, after you start and finish university. And I'm mainly doing projects in the area of corporate and institutional banking, which serves larger institutions with financing or with other financial products.

And the products, the projects that we do are used for these strategies, for example, for market entries or for product strategies, also for growth or also for business reductions in certain areas. And also, we do a lot of market analysis, for example, with a revenue potential tool that we developed in-house, which we sell and then implement in clients and institutions. We said that we also do risk.

00:01:50 Interviewer

OK, perfect. That sounds good. So, you already mentioned that you work as a senior consultant. So, this is your current job title, right?

Have you had a job title before or is this your first job title? The one which you started with in the company.

00:02:08 Interviewee

No, that's the second job title. The first title was consultant; after finishing university and now the second title is senior consultant.

00:02:17 Interviewer

So, you got a promotion, right?

00:02:19 Interviewee

Yeah, exactly.

00:02:21 Interviewer

Congratulations on that.

00:02:24 Interviewee

Thank you.

00:02:26 Interviewer

Do you still work there?

00:02:30 Interviewee

Yeah, I do. I mean the promotion just happened, and it doesn't really change who you report to, as long as the person who you reported to before is also in the same role and is still higher in hierarchy than you, so if you would get promoted to the same hierarchy level then it might change. But as long as the person is still higher in hierarchy than you, then it's probably not going to change.

00:03:03 Interviewer

And what's the job title of the person you report to?

00:03:10 Interviewee

Usually, manager or senior manager.

00:03:14 Interviewer

OK.

00:03:14 Interviewee

Right now, it's mainly managers.

00:03:18 Interviewer

OK, so there are no hierarchical titles in between. Do I understand that right?

00:03:23 Interviewee

Yeah. So, the manager title is the one that comes after senior consultant. So, it's the hierarchy directly after. After that comes a senior manager and after that comes a partner. So, the number of titles that you can get is limited. But of course, you can be multiple years within one hierarchy.

00:03:43 Interviewer

OK. So, would it be common that you are the person someone reports to, so a lower-level hierarchy reports to you and just to you, or would it then always be like a two steps report system?

00:04:02 Interviewee

Yeah, it depends on the type of project. Usually also on the size of the project. If you have a bigger project then it's also possible that a lower hierarchy like a consultant or also an intern, for example, reports to me. But if it's a smaller project then it is also possible that the consultant does mostly independent work and reports directly to the manager.

00:04:31 Interviewer [slide 2]

If you had to describe a career path, would it be the specialized career path, the project management career path or the management career path?

00:04:42 Interviewee

In my role and my company, it will be the management career path.

00:04:46 Interviewer

OK, so no project management, although you do project management.

00:04:52 Interviewee

In our company, yes, it can also be project management, but not in my role because that depends mostly on the size of the project. But it's a really big project that needs its own project management capacity then also that can occur. But I usually do smaller scale projects. So, if you have a team of five people maximum, then usually you don't need your own project management.

00:05:19 Interviewer

OK.

00:05:20 Interviewee

For our type of projects.

00:05:22 Interviewer

Really helpful. Thanks. You already mentioned shortly that there are some interns, but are the levels in between so intern, consultant, and then senior consultant?

00:05:34 Interviewee

Yeah, there's another position in between that is analyst. So, intern is the lowest position that you can have, which is students who are still university than analysts, analysts in our company are kind of in a special position because you can only become an analyst after you finish your bachelor's and you didn't start your masters, but you want to start. And then you become an analyst. After you have done your master's, you come back to the company as a consultant. So, it's more of an educational role, but it's still paid, so you still count as an employee, that is below consultant. So, the first real position with the finished degree role is consultant.

00:06:17 Interviewer

OK. And after, so you get (promoted to) consultant, then you are a manager and after being a manager, which roles are coming afterwards?

00:06:26 Interviewee

After consultant comes senior consultant, manager, senior manager, and partner.

Of course, the pyramid becomes more and more narrow to the top, so there are a few partners and more hierarchies below.

00:06:45 Interviewer

But do you still have, like, a CEO? Of course, right or?

00:06:50 Interviewee

I mean the legal form of our consultancy and of many consultancies is a partnership. So, we do not have a CEO like a listed company, for example. We have a managing Director of the whole company. So, he's one of the partners. All the partners are shareholders of the company and he's the one who was elected to be the leading partner, so to speak.

00:07:18 Interviewer

OK. Right. Nice. Umm. So, you said you have like smaller projects, but are there like other projects or other departments which have bigger projects and where it gets more technical or where you also need more software engineers, for example, or just employees that don't really have the consulting like hierarchy?

00:07:51 Interviewee

Yeah, definitely. In other parts of the company, you have bigger projects, and you have more variety of roles also. So, my projects are more focused and are more strategy based. We usually work in smaller teams. But for example, in IT projects, which we also offer, you also have IT developers. And you have in general way bigger teams and the volumes of the project are bigger. For example, if the bank wants to implement a new software tool and yeah, it can become a project of one year for example, and revenue is in the 1,000,000. So of course, you have way bigger teams, and you have a lot of roles. So, you have Project management, normal consultants, but you have also other skilled labor workers in terms of IT developers for example. So yeah, in that kind of project you can also have roles beyond the consulting level.

00:08:58 Interviewer

OK, so the structure I'm showing you right now, those this is kind works for your company as well, or just maybe more like the right sides, the management side. So also, being with supervisor and senior supervisor or is it just called completely differently.

00:09:21 Interviewee

I mean it works in terms of the general logic, but of course it has more layers. So, our company is pretty lean when it comes to the number of layers. So, there are not too many

hierarchy depths that you can take. But the principle is the same I would say, but it's not really next to each. It's more like a pyramid, so you have the top and it goes down and the skill sets are more and less comparable. The only thing that is maybe left like it's from here are the technical roles like IT developers. And what I see there is also unskilled, semi-skilled labor. Yeah, unskilled is maybe not the right word, but of course we have also professional roles that don't really work in consulting, like marketing, like HR, like Finance internally. So, all the internal roles that are not client-facing, they of course also are next to the pyramid organization next to the consulting roles.

00:10:39 Interviewer

OK. And do you happen to know how they are structured, like do they have similar titles or is it completely differently structured?

00:10:51 Interviewee

It is definitely structured. Definitely. I cannot tell you all the names, but for example the role titles will be more like professional, senior professional, or for example head of finance, head of travel accounting, so it is more kind of a different structure, and the career paths are not that much preset like in consulting roles. They are slower development paths and of course. Also, fewer people in leading roles and more people in normal working roles.

00:11:29 Interviewer

OK, so you have 2 screenshots here. They are not too good in quality, but you can kind of tell that the support functions and the core functions work next to each other. Would you say it's similar in your company? [shows slide 1]

00:11:43 Interviewee

Yeah, yeah, you can do that. We have the real - the content part of the company which is sold to the clients, and you have the internal part which is working to enable the work that is sold to the client, that is correct.

00:12:00 Interviewer

OK, perfect. I think that's it for now. If I have more questions, I will come back to you if that's alright and if you have any other questions, please feel free to ask them now.

00:12:16 Interviewee

Yeah, yeah, not right now. Yeah. Thank you for the interview.

00:12:20 Interviewer

Thank you too, Jakob.

Interview 2 – FS industry

00:00:02 Interviewer

Thanks for your time. And just to let you know, so the interview will go as follows. In the beginning, we will have some short questions that you can answer with just one sentence. And in the following we have like a more work together part, but I will tell you when the point is reached and then I will also show you a PowerPoint which helps us a little bit to work together. Is this all right with you?

00:00:29 Interviewee

Sure, perfectly fine everything.

00:00:31 Interviewer

OK, perfect. So, are you working in the financial service industry?

00:00:37 Interviewee

I'm working in a company where they're doing cash systems, so I would consider it as financial services.

00:00:45 Interviewer

Very good. What's the name of the company?

00:00:48 Interviewee

The name is ready to order.

00:00:51 Interviewer

Can you tell me approximately how many employees do they have?

00:00:55 Interviewee

Umm I think approximately like 80 or 90.

00:01:00 Interviewer

Quite a lot. And in which department, do you work?

00:01:08 Interviewee

I work in the marketing department.

00:01:10 Interviewer

OK. Can you explain a little further which other departments the company have?

00:01:15 Interviewee

So, we have the marketing department maybe starting with that we were like 15 people that do all the spectrum of marketing things. Another big part is the finance part, yeah, where all the financial things from our company are like calculated and observed. Another smaller part is HR, where we obviously have, I think two or three persons that are responsible for - who's going to be hired or everything around that. And like company meetings, stuff like that. And a very big part for the engineers because yeah, the main software has to be written, programmed. It has to be contained, and since that's a big part of the product, we have a lot of engineers. I think half of our team are engineers.

00:02:11 Interviewer

OK, that's good to know because actually I am working with the software industry and with the financial service industry. So maybe your answers will help both of my industries. OK. So that being said, we come back to the information later and now I have some more personal questions. So, in which position do you currently work?

00:02:33 Interviewee

The name of my position is marketing specialist.

00:02:38 Interviewer

And what is your/ Do you have management responsibilities in this position?

00:02:47 Interviewee

Mh No.

00:02:50 Interviewer

Did you get any promotions earlier or was this your starting point in the company?

00:02:57 Interviewee

I started as part time employee during my master's, so I worked like 15 to 20 hours depending on like what time of the year. And when I finished the master, I got. Yeah, maybe it was a promotion. My job switched a bit since I have more hours. I got more tasks, and my role name switched a bit from performance marketing trainee to marketing specialist.

00:03:28 Interviewer

OK, so you didn't get any further promotions or job title changes, to be specific.

00:03:42 Interviewee

Actually, I got a my/ yeah, but in our team it's a bit different because everyone has the title manager and so I got the title, like I think a few months ago, I got the title marketing manager. But nothing changes and none of my tasks changed. So, it was just the title.

00:03:48 Interviewer

OK.

00:04:02 Interviewee

Because I said everyone in our team is a manager, so I want to have this in my role as well, but yeah.

00:04:08 Interviewer

OK. That's really good to know since I just worked with the titles and sometimes it's really hard to find the real purpose of a title.

So, in general I saw doing my research that there are three different career paths. I will tell them now and you can tell me which one suits you best. So, the first career path is called specialist career path. The second one project management and the third one management career path. Which one describes your position the best or your career path?

00:04:46 Interviewee

What is/ Can you maybe give me one sentence for each that I understand you?

00:04:49 Interviewer

Yeah. So, a specialist career path is probably most likely to be very/ has a lot of knowledge in a specific field, for example, a software engineer. And the highest position you get in a specialist career path, is probably without management responsibility.

00:05:07 Interviewee

OK.

00:05:07 Interviewer

The project management is probably pretty self-explaining. You work in projects and the projects follow a basic project structure. And the third career path is the one I think most known where you just start as an employee and then your kind of goal, or not your goal, might be that you go to the management areas and maybe end up as a CEO, or similar.

00:05:35 Interviewee

I think in my work and ready to order, it's most likely the management path because although we work on projects we have management structures. I mean I started as a part-time person and now projects and responsibilities grow but I still have the same bosses. That works well. So yeah, the role is a bit growing in itself. But yeah, no plan on actually being like the team leader of 16 people since I think our company is not too big to hire someone from the team to be the head of the team.

00:06:15 Interviewer

OK. And now a little bit more to a simpler question. So, what is the job title of your manager or from the person next higher to you and next lower to you, or next lower to you probably the trainee position, right you already said before?

00:06:33 Interviewee

Yeah, trainee or like, internships is lower, and above is only/ or his name is Head of Growth.

00:06:45 Interviewer

Head of Growth, is he the company lead or?

00:06:49 Interviewee

No, he is the lead of sales. OK, which I by/ No, I realized I forgot that, as part of the structure in the first question. I think because we have a big sales team—Since we're working with businesses, and they always have asked questions [inaud. 00:07:02]. He's leading sales and marketing, so he's Head of Growth.

00:07:12 Interviewer

OK. Head of Growth and above him?

00:07:16 Interviewee

Just the CEO.

00:07:18 Interviewer

Just the CEO, OK.

OK, so now these were the basic questions, and you answer already quite a lot of the questions that will follow now. But I will show you a PowerPoint that will help us to see and make it a little more clear. Just let me show you my screen here. Can you already see it?

00:07:51 Interviewee

Mhm. [affirmative]

00:07:53 Interviewer [slide 1]

OK, so what you see here is like kind of a basic structure of bigger companies. So, you see like the Heads like in the middle, the grayish one and to the left and right are the support functions like marketing. And then we have the core functions with a little bit more levels. Are we clear on that or should I specify a little bit more?

00:08:19 Interviewee

It's fine, I can't read it too much.

00:08:21 Interviewer

Ohh yeah, it's really small.

00:08:25 Interviewee

It's Fine, I think.

00:08:29 Interviewer

Yeah, it's kind of a basic structure just to make clear what kind of my job is/ to see just the hierarchy. So, and here we have/ Like it just makes it a little bit easier to explain. So, on the right side you have the management path and on the left side you have the individual contribution career path. And as you can see, both paths are kind of crossing each other. So, the experts on the left side are kind of on the same level as the top management. So maybe like a lead or also a vice president, depending on the company. So, both paths can cross, and it doesn't mean that you're like lower level than a team leadership just because you're an expert. That's my theory and the interviews are like supposed to serve my theory as wrong or right [slide 2]. Here we have some examples, how it can be – But I think I'm going to skip this a little bit right now, just if it's OK with you [slide 3 and 4].

00:09:36 [slide 5]

And here we have the actual structure. I give you a second so you can read through and maybe you identify some of the positions and can tell me if you have them as well, maybe especially in the software department.

00:09:53 Interviewee

I mean, we have a CEO. We have, yeah, a head, head of Software, I think. A Manager, as I said, since everyone is a manager. Trainee, of course, as I said. And yeah.

00:10:18 Interviewer

So, the managers in your company are more likely to be specialists, right?

00:10:24 Interviewee

Yeah. I think the term specialist manager is basically the same.

00:10:29 Interviewer

OK yeah. So, for me would be important to know if there's like some more steps between like the manager or like if they're more specific names, job titles being used in the company.

00:10:46 Interviewee

I think in most teams we're just, I mean, we have to have the CEO, underneath, we have all the heads and underneath we have like, yeah, all the team members and it's not, I think it's maybe in some parts it's one person between the head and the other managers, but in some not, and especially in our software team our head of engineering, I think he's basically a really good manager. I mean, he's rather a good engineer than a real manager because I think he's really good at all the engineering parts. And of course he's the team lead, so he knows everything. But yeah, he's just an expert with some managing things and not a manager in a classical sense like my boss.

00:11:46 Interviewer

OK. Do you know his job title?

00:11:52 Interviewee

He is Head of Engineering.

00:11:54 Interviewer

OK, so he's a Head then?

00:11:54 Interviewee

Yeah.

00:11:57 Interviewer

OK, so the head of engineering is above the manager that is your superior, right?

00:12:06 Interviewee

No, they're on the same level.

00:12:08 Interviewer [slide 7]

OK, nice. Good to know. Actually, I came to the manager problem and what I realized is that there are a lot of manager titles and it's quite hard to say which is higher, which is lower. So, I came up with three different areas of managers. So, on the left side we have like the real area. So, if you are like a global manager, which you probably just have when you're a really big company. Then you have the seniority which is kind of a hierarchical level thing where you going to jump up. And then on the right side you have the more individual career path managers. Do you identify any of them?

00:12:54 Interviewee

I would say skills, maybe because we have project manager. I am the marketing manager. So, it seems like the support function manager, we have a lot of sales managers. So, I would say maybe in the direction of skills.

00:13:11 Interviewer

And the project manager is above the support function manager.

00:13:15 Interviewee

Not really. They're more like, I mean. To be fair or marketing team really has changed since I started there, so the structure we have I think is not a structure that was initially planned,

but rather like came up. So, the project manager maybe should develop into a function that's a bit above. But right now, it's rather on the same level.

00:13:49 Interviewer

OK. So, basically that were all my questions. So, we're good to go.

00:13:55 Interviewee

OK, that was everything?

00:13:56 Interviewer

Yeah, Thank you. I will stop sharing my screen and stop recording.

Interview 3 – FS industry

The interview was conducted in german and translated afterwards.

00:00:01 Interviewer

Let me briefly explain what the aim of my Master thesis is, so that you understand what the aim of the interview is. In my master's thesis, I tried to illustrate the hierarchy levels for an entire industry based on job titles, using LinkedIn data from the USA. For the study, I examined the financial services industry and the software industry. The software industry is not relevant in our case. And now the interview is about looking at how many levels I have identified, i.e. jobs within my job titles, can be mapped in the company and, above all, are they identified in the right order, i.e. in the right hierarchical structure.

And to do this, we would first go through a few basic questions, which can also be answered in one sentence, and then we will move on to the second part, where I will also show a PowerPoint to specify or simplify some things. We can then try to create a hierarchy overview of your company together. Does that work for you?

00:01:18 Interviewee

That works, yes.

00:01:19 Interviewer

Great, let's start with the first question.

Do you work in a financial services company?

00:01:28 Interviewee

Yes, I do.

00:01:29 Interviewer

And which company do you work for?

00:01:32 Interviewee

Well, I work for American Express. But in Austria, right.

00:01:38 Interviewer

Great, can you tell me approximately how many people work at American Express?

00:01:47 Interviewee

I can't say how many people work there worldwide. I know that there are about 120 of us in Austria.

00:01:54 Interviewer

Okay. Now a few personal questions, what department do you work in?

00:02:04 Interviewee

My department is/ These are of course internal company notations that may not be so clear to outsiders. My department is the external acquisition department. So, you could actually say I'm in the marketing department.

00:02:25 Interviewer

Okay. And what's your current job title?

00:02:33 Interviewee

Digital Acquisition Coordinator.

00:02:37 Interviewer

Long title. And do you have management responsibility within your position?

00:02:46 Interviewee

I don't have any management responsibility, although we - we have Steffi. I don't know how many hours. I think 18 hours, roughly, and she's supportive. She also gets work assignments from my colleague and from me, but that's not really management responsibility, I would say.

00:03:22 Interviewer

Do you have her job title?

00:03:30 Interviewee [looks for job title in e-mail signature]

I think she/ I can have a quick look.

I don't think she has a job title. [...]

00:04:24

Ah, she has "Administrative Assistant" in her e-mail signature.

00:04:34 Interviewer

Assistant ok, we can work with that.

00:04:42 Interviewer

Did you already have a promotion or did your job title change during your time at AMEX?

00:04:49 Interviewer

It hasn't changed. Right. And I've been with the company and in the position for just over 2 years. Exactly 2 years and 5 months approximately and it hasn't changed. It may change in the near future, but nothing has changed at the moment or so far.

00:05:12 Interviewer

OK, do you know how that would change when the promotion comes? That just sounded like a good outlook.

00:05:23 Interviewee

Exactly that/ I don't know exactly yet. It's all so complicated with us because it's not quite/ So many things in our company are very transparent, which is great, and some things are not transparent, I think, or are a bit contrived. We have these band levels. So, every person or every position has a band level and in Austria you normally start at band level 25 and then

it goes up and mostly as far as I know in increments of five, which means that there might be a jump from band level 25 to band level 30 for me or a colleague of mine in the near future. And what I can't quite say yet is whether the title, whether it will change, yes.

00:06:32 Interviewer

OK, well, maybe we'll come back to that later. In general, just for your information in the master's thesis/ the interview is not necessarily about eliciting details from your CV or internal company information, but really about being able to confirm this industry hierarchy or not. Accordingly, specific company information is of course necessary to practically confirm this information. But this is not being analyzed now - in other words, the company American Express is not being analyzed now. Or trying to find out company-specific or company secrets that you wouldn't get otherwise, but simply to verify this work and find out whether my structure makes sense.

00:07:37 Interviewee

Yes, so please, please stop me or if I'm talking and it's not helping at all or if it's going too much in a different direction.

00:07:45 Interviewer

Mhm no, so far what you've said has been helpful anyway, I just don't want you to feel like you have to give me information that you might not want to give. So let me say that again at this point. The next question would be in terms of career direction. So, there are three different career paths. I don't know if that rings a bell, but that would be the specialist career path, then the project management career path and then the management career path. Have you heard that before?

00:08:16 Interviewee

No, I haven't.

00:08:18 Interviewee

OK, then I would discuss that in the second part, I also have a PowerPoint, so that might be a bit helpful. One more question would be what the job title of the person above you is. Which I think is still Klemens, isn't it?

00:08:34 Interviewee

What is the job title of the person above me?

00:08:36 Interviewer

Exactly, so the next highest, so from the person you're reporting to.

00:08:42 Interviewee [looking for signature]

Exactly. That's right, it's still Klemens, and Klemens has a longer title. He often doesn't send his signature. Where you see that, I could have picked it all out already, I'm sorry.

00:09:21 Interviewer

All good, we have time. I definitely have time; I hope you do too.

00:09:48 Interviewee [looking for signature in email history]

He never sends the signature. I do it every time, why he doesn't.

Well, the profile only says Manager Marketing. That's a bit, just a bit. Yes, exactly, you can say Manager Marketing or, right there I have it now, you can also say "Head of Acquisition and Product Management" in his signature.

00:10:30 Interviewer

OK, that's good that you looked it up, because according to my logic, the Head of/ So Head of is definitely higher up than Manager.

So, it's good that you looked it up. We're actually already on the second part, so I'd like to share my screen now. Let me know if you can see it.

00:10:55 Interviewee

Yes, I can see it. But I can't see the individual parts, it's quite small.

00:10:59 Interviewer [slide 1, upper diagram]

Exactly, it's relatively poor also in terms of quality. It's actually a classic company organization chart, in grey you can see the board of directors and the management and from there the individual positions here now "responsible for combating misconduct". Those who report directly to the Executive Board and those who report from the Executive Board are different business units. So, what exactly is written there is not really relevant, but simply to call an organizational structure into your head. This could be, for example, the core business of American Express in the center and then Marketing or Sales on the side, where there are individual hierarchical structures within these lines. Is that clear so far?

00:11:56 Interviewer

Yes.

00:11:56 Interviewer [slide 3]

Right. And now we come back to these career paths. On the one hand, we have the individual contribution, which actually means that you're basically pursuing a career within a field of expertise, but you're not necessarily looking to take on management responsibility. What it also implies, however, is that you can be positioned at a similar level to top management when you reach the top. But you are simply so specialized in your expertise, for example as a software engineer, that it doesn't really make sense for the company to give you management responsibility and then put you in charge of organizational matters, which takes time away from the important knowledge you have in software engineering, for example. And then we have the other career path, which is actually the classic one, the management career path, where you start at the bottom. For example, like me as a student alongside work, and then you work your way up and yes, maybe end up at the top as CEO at some point. Those are two different paths and then there's the third, the project management career. This is actually rather irrelevant for my work. Exactly, the question now is, would you see yourself in one of the two paths?

00:13:39 Interviewer

Ahhhm

00:13:41 Interviewer

Was that even coherent?

00:13:44 Interviewee

Yes, that was understandable. But it's not that easy to say where I see myself and of course in practice the boundaries become a bit blurred.

00:13:58 Interviewer

Mhm, totally. So that's exactly what's difficult to discuss. Where exactly/ also, that the expert might be at the same level as the team leader.

00:14:15 Interviewee

Yes, yes, but then perhaps more from the left-hand side, where the focus is more on specialist knowledge and expertise, without this management responsibility.

00:14:33 Interviewer

OK, yes.

00:14:36 Interviewer [slide 3 and 4]

Yes, those would be examples. I'd prefer to skip them now and show you the structure that I've developed. And then we can see if you recognize any of it. Does that work?

00:14:54 Interviewee

Ja.

00:14:56 Interviewer [slide 5]

Perfect, alright. That's actually the structure, that you see here, which I've developed based on the LinkedIn titles and further research. And yes, otherwise maybe just take a minute or

a second, as long as you need to read through it. And you can also tell me in between if you recognize any titles from your company.

00:15:25 Interviewee

That sounds pretty good. Yes, that sounds logical, it looks good, yes. [...]

Sorry, what did you want to know from me or me?

00:15:56 Interviewer

Whether you/ whether you recognize individual titles in your company, i.e. whether the titles exist in your company.

00:16:02 Interviewee

Oh, ah, whether I recognize individual titles. Or yes/

00:16:05 Interviewer

Exactly or, well, the C-suite is of course not specified, but it means Chief of Marketing, for example. So, whether these titles are assigned in your company and if so, whether you can also tell me if they are assigned in this order.

00:16:35 Interviewee

Yes, some of the titles are assigned in this order, but some of them are also called something else because there are simply internal titles for them, which then/ we have our own acronyms or designations. Companies – what do you call them, not slang, but yes. But of course, there are CEOs, but then we also have mixed roles in some cases. For example, the man I report to, Klemens, who is now Head of Acquisition and Product Management, has also been the second Manager (Geschäftsführer) since last year, which means there are also dual roles. For example, he has two departments, Product Management, which he does on his own, and then Acquisition, where he has a team, yes.

Ah, and we don't have the number 12, for example. So, there are no intermediate levels sometimes. You've already listed it very granularly and it's not that granular with us, I'm sure it depends on the size of the company or the size of the department.

00:17:59 Interviewer

Exactly, so in the/ what you see here that would also be the maximum that exists. In other words, what the largest company in the entire market would have, but perhaps it doesn't even exist. So, it's actually about mapping whether you could map all the companies that operate in this industry within this structure. In other words, if you say, yes, we have 1, 4, 7, 9 and so on and so forth, then it would shrink to this hierarchy level for your company. Right. So, it would be great if you could be a bit more specific, even if they are acronyms. So basically, the titles that you see there are actually groups in my code. For example, 1 doesn't just say Owner and Partner, but also CO owner, Leading Partner - category 1 would be a simple level, but Senior Specialist, for example, includes around 300 terms, if not more. I think it even has around 500 terms.

00:19:30 Interviewee

Mhm, yes.

00:19:34 Interviewer

In other words, the levels, the terms there, try to keep it simple and understandable for everyone.

00:19:36 Interviewee

So, we definitely have 3, 4, 5, 8, 9, 10, 11, 12, 13.

00:19:57 Interviewer

That's a lot. Does the 8 also exist in all forms, including Executive, Senior, Vice President and then Assistant? So, in these four leaps?

00:20:13 Interviewee

No, it's not that granular. For example, I know a bit about the situation in Germany, with my German colleagues, and there it is of course - there are several levels or several gradations and subdivisions, exactly, yes.

00:20:29 Interviewer

Okay. And then probably the same for 10, right?

00:20:39 Interviewee

Yes, exactly the same at 10, there's actually only one level with us, yes.

00:20:46 Interviewer

OK, so the term senior is generally used in the company, but not in Austria, and does it indicate that you are one level higher? And it definitely indicates that you've made a small leap.

00:21:02 Interviewee

Exactly, yes

00:21:04 Interviewer [slide 7]

Yes, then I have the last slide, I realized that the manager category is partly, I wouldn't call it misused, but I call it alienated now, and that there are many types of managers and that it is sometimes very difficult to tell these managers apart. Because they range from a global manager of an entire company to a junior manager of area X. Do you have any insights for me?

Please take a moment and have a look at what I've created.

00:21:59 Interviewee

Yes – So your question to me is whether there are these gradations or these categories – whether I know them like that or whether I can confirm them like that?

00:22:08 Interviewer

Exactly, so global would be more or less above the national manager and so on and how exactly then. And also, for example, whether there would be a global senior manager, i.e. whether they would be used in a mixed way or whether they would be used in general.

00:22:22 Interviewee

Yes, we/ can I understand it like this, for example: The global manager is really the most senior person, who is based in New York, for example. And then there's an international, I don't know, then there's someone in between who's responsible for Europe. Then there's someone who's responsible for the market and the country, so there's definitely that.

Then there are further subdivisions, i.e. by region or division or office - we don't have any, because we only have one location and haven't subdivided it again.

According to seniority, there is - I would have to think about it now. But yes, there is another situation. And by skills and by areas of responsibility, yes.

00:23:31 Interviewer

Okay, quite a big company.

00:23:39 Interviewee

In principle, we are organized in a kind of matrix organization. In other words, we have here/ we have our superiors, or we report to people who are in our own market or in our own country, for example. And they now/ for example, there is someone/ so I report to my line manager (Vorgesetzten), and he reports to his line manager and it's now about marketing or new customer acquisition, but then there is another reporting line, or an exchange that goes beyond this line, or goes alongside it. Where there are then supervisors for these respective areas, or Century Teams.

00:24:43 Interviewer

Okay.

00:24:44 Interviewee

Yes, well, I don't know whether that made sense, but there are also separate central teams for each specialist area, so whether it's a specific marketing area, i.e. new customer acquisition, there are also separate central teams that deal with areas and think about things and provide reports and are in dialogue with the markets.

00:25:11 Interviewer

OK, and then above these specialist areas would be the entire marketing department, so to speak, and then above that would be the entire marketing department for country group X and then probably above that would be continent and then global. Can I imagine it like that?

00:25:34 Interviewee

Yes, exactly, so for us it's organized in the USA and then it's International and then I think it's subdivided into continents, for example EMEA.

00:25:55 Interviewer

Yes, OK. Very good.

That would be the main questions, right. So, it's really primarily about this depth/ whether that's the case, for example, whether the President is above the Vice President, that's perhaps a simple example, or whether the Lead is below the Director.

Exactly, obviously the director has to be present for that, but I think you said 10, if I remember.

00:26:34 Interviewee

Yes exactly, so director in any case, yes. So exactly, the way you're showing the structure, it's definitely there with us too. I wouldn't have noticed anything that is organized differently, that I would organize differently. It's just that it's definitely much more compact or compressed here because of the size of the site.

00:27:00 Interviewer

Okay. Well, do you know if there's a chief here at the site? So is there already a vice president at the Austrian location, for example, or is that first introduced in the global company, so only for America?

00:27:24 Interviewee

Exactly, he will only be introduced for larger markets or for the USA. We don't have one right now. There's no Chief Operations Officer or CFO sitting in our company (in Austria) right now or a CTO. That doesn't exist, yes.

00:27:43 Interviewer

Yes, OK. Well, I think those were all my questions then.

00:27:52 Interviewee

Yes, great.

00:27:54 Interviewer

Great. I'll stop the recording.

Interview 4 – FS industry

00:00:00 Interviewer

Very good. Yes, I'm pleased that we've come to meet here today. Right, very briefly about the aim of my master's thesis, I've already told you a bit, but I've just used the last few months to determine a hierarchy depth for an industry, and in this case that's the financial services industry. And I did this using LinkedIn data and actually only using the job title. The interview is actually about validating whether this hierarchy sequence that I have established can be confirmed in practice. The interviews are not necessarily about mapping, confirming or validating this complete depth, but simply to see how individual companies are set up. But again, it's not really about analyzing an individual company or passing on company secrets; as I said, it's a master's thesis and therefore remains at a scientific level. The interview is also anonymized.

Yes, the interview would take place in such a way that we actually have two parts. The first part is a few more personal questions, or questions about your employment, and you are welcome to answer these in one or two sentences, and in the second part we would then try to work out your company hierarchy together. Yeah, does that work for you?

00:01:23 Interviewee

Yes, wonderful.

00:01:24 Interviewer

Very good. Right, do you work in the financial services sector?

00:01:29 Interviewee

Yes, I actually do. In fact, yes, we do practically the entire lending business of all the Sparkassen. I'm actually responsible for major Sparkassen (Großsparkassen).

00:01:43 Interviewer

OK, that fits perfectly. You just answered the next question. The next question would have been at which company, but you've already answered that. Can you say approximately how many employees the company has?

00:01:56 Interviewee

Yes, we are about 700 without the whole service area. Otherwise, of course, we are over 1000.

00:02:03 Interviewer

Okay, and what's the name of the department you work for?

00:02:07 Interviewee

My department is the regional head office (Landesdirektion) for the network businesses two (Verbundgeschäft 2). So that's basically the sales department for the western region if you want to narrow it down a bit.

00:02:22 Interviewer

OK, and what is your job title in the company?

00:02:26 Interviewee

My job title is actually Regional Director (Landesdirektion) West, i.e. Sparkassen West. And the Sparkassen are divided into networks (Verbünden) accordingly.

00:02:34

Mhm. And do you have management responsibility?

00:02:45 Interviewee

Yes.

00:02:47 Interviewer

Okay, then we'll talk later about what the position under you are called. If that works for you.

00:02:52 Interviewee

Yeah, sure.

00:02:53 Interviewer

And have you had any promotions? And if so, how did your job title change during the promotion?

00:03:00 Interviewee

Yes, I've had various promotions, and my job title has changed every time. Would you like to know exactly what my name was before?

00:03:13 Interviewer

That would be great.

00:03:14 Interviewee

Yes, okay, great. So below that is now/ Before it was the area director (Gebietsdirektor), below that it's a sales trainer (Vertriebs-Trainer). Oh yes, sorry, between these two is the group management (Gruppenleitung).

00:03:34 Interviewer

Okay.

00:03:35 Interviewee

Then it's always oriented towards sales in a certain region, yeah.

00:03:46 Interviewer

OK, right, and that were already the questions for the first part and now we'll move on to the second part. I'm going to try to share my screen now, but since we're doing it via smartphone, it's possible that it won't help at all. Do you see my screen?

00:04:07 Interviewee

No, not yet.

00:04:14 Interviewer

Still not, right?

00:04:16 Interviewee

No, otherwise just read it out to me.

00:04:20 Interviewer

OK, wait, I think maybe now if I make it full screen.

00:04:24 Interviewee

Ah god, yes, so it's really/ very tiny/

00:04:28 Interviewer [Slide 1]

OK, I think we can skip this slide. It's really just a reminder of what a company organizational chart can look like. But I think you're aware of that. So, what exactly it says is not relevant.

00:04:46 Interviewee

Yes.

00:04:50 [Slide 2]

Exactly. And then there are actually two career paths. I could actually hear a bit from your description that there are/ Well, there are two career paths and that you actually took the management career path.

00:05:00 Interviewee

Correct.

00:05:02 Interviewer

What is relatively interesting for my work and important in this context is that there are these two paths and that they intersect at a certain point. So, in this case, professionals and team leadership, and it's not entirely clear from the job titles where these paths intersect. So, if

you have any information for me, for example, that you are reporting to a technical manager, who is perhaps involved in programming and not really in management, and the job title or general information about that, I would definitely be pleased.

00:05:37 Interviewee

Yes, I actually do. I'll have a look in the IT area now, that would be something like that, they don't necessarily/ IT here.

00:05:49 Interviewer

Yes.

00:05:50 Interviewee

There's a so-called, that's the top manager, yes, that's the division manager (Bereichsleitung). There is of course a suitable assistant there. Directly below that is the head of department (Abteilungsleiter) with the respective officers. Then there are the management and control systems. So, it's actually always broken down to the respective management/team management, as division management (Bereichs-), department management (Abteilungsleitung), team management and then the employees.

00:06:25 Interviewer

Okay.

00:06:26 Interviewee

We only have these titles. In the other cases, it's more like state directorates (Landesdirektion), regional directorates (Gebietsdirektion). Then there is actually group management (Gruppenleitung), team management, broken down to the person who is then also active on site, i.e. the sales trainers. The only connection that I see is in the third hierarchical level from the bottom, that we always talk about group management here, with us and with the others it's just department management.

00:07:01 Interviewer

OK, so the group management is at the same level as the department management. Do I understand that correctly?

00:07:06 Interviewee

Yes, correct.

00:07:09 Interviewer [Slide 5]

Very good, very helpful.

Exactly. And that's already the structure I've made. Can you see that on your screen?

00:07:20 Interviewee

Yes, that's really (very small), I'll make it big now. You're also standing sideways (inaud. 00:07:26) you can't turn it, can you? I could probably read it if you turned it.

00:07:32 Interviewer

Wait. Can I turn it (...) I can zoom. It's difficult to turn it, but I can zoom.

00:07:44 Interviewee

Yes, just like that, so what does it say here - Owner, Partner, Board Member, CEO, Managing Director. OK, yes, that's actually the case with us / With us, if you take the owner, that's actually our shareholders, clearly. It's a shareholder structure, two shareholders. Because we are actually a GmbH, but otherwise that would also fit.

00:08:16 Interviewer

Exactly. So, these are all the English terms now, simply because I was working with data from the USA. I hope that's not too distracting, but a lot of English terms are already used widely.

00:08:26 Interviewee

Yes, we're not allowed to do that, because we've actually written ourselves, the Sparkassen, on the flags, that's why we're there/ So we try to stay quite German on purpose, because that's what's desired in these areas.

00:08:40 Interviewer

Yes, but that shouldn't really bother because, in theory, the translation still exists.

00:08:46 Interviewee

Exactly, yes, yes, it's exactly the same with the other banks.

00:08:50 Interviewer

Yes.

00:08:54 Interviewee

Senior, yes. (inaud. [goes through list]) But it actually fits in with what I said. In terms of the structure, the organization chart, right?

00:08:59 Interviewer

Exactly. Shall I go further down? Or have you already been to Head?

00:09:03 Interviewee

Yes, down. Yes, so here/ Yes exactly, so here with all the directors, that's also a bit typical, when I look around on some platforms myself and see who I know and what they call themselves or what their real title is, that's interesting. So that wouldn't fit in with our hierarchy, so to speak, but that's what happens / Well, that's an insane hierarchy tree you've set up there. So, we're a bit leaner on the board, although we actually used to have more hierarchies.

00:09:45 Interviewer

Ok, so what would you not depict?

00:09:54 Interviewee

Yes, all the individual directors in between, the junior and senior and that sort of thing. Of course, we also have, for example, I'm the regional director now, so of course that's above the regional director, you could compare that a bit with junior and senior, yes. But we don't have anything in between. At least not in this area now, but even if I open the other areas here, IT or Treasury or something like that, I don't see that either.

00:10:32 [reads out in fragments what she sees in the internal company organization chart]

I'll take another look. Distribution organization/ divisional management/ Do I have anything else here, no, that's all. We simply differentiate between/

00:10:46

So really on the non-sales side it's always called management (Leitung): head of department, group management and so on. Only with us it's called director, that's just the difference.

00:10:55 Interviewer [refers to slide 5]

OK, but would you then theoretically/ Because Lead is actually 12 here and Director is 10, so two levels above that, would that fit?

00:11:05 Interviewee

Yes, with us, so with me there's only one boss and then the management.

00:11:13 Interviewer

Ah okay.

00:11:14 Interviewee

Well, it's called management (Geschäftsführung), then the top manager, the top divisional management (oberste Bereichsleitung), that's really what we call divisional management. It's even higher than the regional management (Landesdirektion). That's basically one person who then manages many, many Sparkassen, and then below that we come from the state directorate, which is then Hesse, then North Rhine-Westphalia and so on, and below that come the regional directors (Gebietsdirektoren), where these individual areas are divided up again, because it's huge. So North Rhine-Westphalia is still small and then there are the

group managers, who are responsible for looking after the individual employees on site, so to speak, and then report accordingly, both downwards and upwards.

00:12:06 Interviewer

OK, I'll stop sharing my screen for a moment, because then I'll see you again.

00:12:14 Interviewee [adjusts smartphone]

Yeah, I'm a little oblique here.

00:12:20 Interviewer

So, now I see you again.

What positions come under you then, so what job titles?

00:12:30 Interviewee

Exactly. Yes, we have the area directors (Gebietsdirektoren).

00:12:34 Interviewer

Mhm [affirmative].

00:12:36 Interviewee

The group management (Gruppen Leitung) and depending on, well not in my case now, but depending on how big this area is, there's even a team management underneath and then there are the sales trainers who then also accompany the actual sales, so to speak.

00:12:57 Interviewer

OK, but then you don't use the term manager either, do you?

00:13:01 Interviewee

We actually used to use it, but because the Sparkasse world is very German speaking, we've now dropped it. So, the manager would have basically been the area director (Gebietsdirektor) for us.

00:13:20 Interviewer

Ah ok, very good and even further down? We start very high up with you.

00:13:27 Interviewee

Yes, as I said, we only have / Well, we have the management (Geschäftsführung), then the divisional management (Bereichsleitung) of the entire association (Verbund). Then comes the directorate/ regional directorate (Landesdirektion), i.e. director, what I am, for the individual region, i.e. North Hesse, North Rhine-Westphalia and so on. Then there are area directors (Gebietsdirektoren). Then there are group leaders (Gruppenleitung), then team leaders and then there is the employee who is on site, who is called the sales trainer (Vertriebstrainer). But we've also had sales consultants. So, it changes from time to time.

00:14:04 Interviewer

Okay, good. And do you know whether you also have subdivisions for the employees? Or whether they can move up again? So, you start as an intern, for example, and then maybe it's something like junior/senior?

00:14:18 Interviewee

What we actually have are trainees, i.e. working students, and we don't have any interns in this area.

00:14:28 Interviewer

Okay.

00:14:29 Interviewee

What we still have, of course, is the more unusual ones, for example, are specialists, I think that's what you were just talking about. These are really more the areas that go in certain

directions, that are really like, IT or what else do they do? Everything that really has nothing to do with the actual business.

00:14:57 Interviewer

Mhm, probably then also /

00:15:03 Interviewee

But they are also embedded, actually, under us. So, they're one hierarchy below me, they're part of my team.

00:15:15 Interviewer

Yes.

00:15:16 Interviewee

But she's a specialist, so she has nothing to do with the actual work in sales, because she just analyzes the figures, evaluates them, product research or something like that. But that's actually the case for each area.

00:15:30 Interviewer

Okay, and how is the team then structured, how does it go from there?

00:15:34 Interviewee

This team not at all. We only ever have one or two people and they have equal rights, they're divided up between the regions, so they're not assigned to a specific other area.

00:15:48 Interviewer

OK.

00:15:48 Interviewee

So, we have our own mini specialist, so to speak. Without marketing or IT or I don't know what, so we have an extra (inaud. 00:16:01).

00:16:03 Interviewer

Control question, sort of: In any case, you probably also have the marketing area, the HR area.

00:16:07 Interviewee

Yes, yes.

00:16:11 Interviewer

You probably also have a law department, so law.

00:16:12 Interviewee

Yes, we have law too.

00:16:14 Interviewer

Very important for banks.

00:16:17 Interviewee

Yes, for example, we have / We also have (the) Corporate Development and Law department, the Audit department, of course. Then we have the back office, which is, as I said, we are set up as a bank in such a way that you have the market, i.e. sales and the back office, those are the ones who control us, that we do everything well and do it right. Then we have - yes, what else do we have - we already had Legal, we already had Audit, then we have the IT division, and we have the Bank Management division, which is of course something special for us that probably doesn't exist in normal companies. And then we also have the service business area, so there are the colleagues who work in telephony, provide customer support when you call the office and so on. There's that too, of course.

00:17:09 Interviewer

OK, yes, very good and they come under you, so under your department?

00:17:16 Interviewee

Yes, yes, that's parallel. Yes, of course they come / we are not responsible at all. It's always like that, I just read out: we now have areas, service business, there's one person who's responsible for that, that would be exactly the same position I have, because there are really only the employees, then underneath, there's no - except for a team leader, there's no great hierarchy anymore. That's why it's like this, so sideways from us.

00:17:53 Interviewer

And then they report to them?

00:17:54 Interviewee

Management directly. Yes, because that's very important for us, actually because of customer satisfaction, so of course that's due to the system. Then we also have the management of technical sales support processes, but these are actually things that are not necessarily next to the management in terms of hierarchy, but everything that is immediately reported to the management, which is particularly important for us, is immediately reporting to the management and therefore they are actually next to it in the organization chart.

00:18:29 Interviewer

OK, that's very good to know. My head is already rattling with how I need to map that in my code.

00:18:43 Interviewee

Yes, I have the organization chart here, but unfortunately, I'm not allowed to give it to you. I can't do that, there are names there.

00:18:50 Interviewer

No, I really don't need the names either. So, nothing, no information that is not given to me in a scientific sense.

Yes, exactly, but I think you already have a lot of information anyway/ I'm just thinking about what else I could use, but actually, you said who's above you, you said who's below you with the title, you said which departments are next to you, you said where career paths

overlap, you gave me as much insight as possible. Which is also totally complex with a company the size of yours. Yes, and that's actually all the information I need, I think.

00:19:34 Interviewee

Yes, that's nice. Oh well, our parent company is a bit bigger, it has over 2000 employees and that's the way it is - I worked there before - but even there you often find these hierarchies in English areas. In fact, that's the difference to a public limited company - there's a bit more structure at the top, but ultimately there's not much difference. So, the designations here are also a bit different/ So actually only that you're not in the German area, but in English, but otherwise it's actually the same.

00:20:16 Interviewer

Okay. If I could share my screen again now, could you show me the numbers/ can you see my screen yet?

00:20:28 Interviewee

Yes, but wait, I have to turn you upside down again, so (...).

00:20:47 Interviewer

Can you just tell me the numbers that you know in the English context and if it/

00:20:55 Interviewee

All of them, actually.

00:20:57 Interviewer

All of them?

00:20:59 Interviewee

Yes, if I look at it that way, I mean, if I look at our parent company. We don't have that here, from the board and so on. But if I look at our parent company, we do have that, yes, we have that too, we actually have everything.

00:21:15 Interviewer

OK, yes, I like to hear that.

00:21:18 Interviewee

Yes.

00:21:19 Interviewer

And then I just had what you couldn't say now/ I had then subdivided the employees more into Senior Specialist and Specialist. Then they would be supervised again and then there would be the workers, i.e. senior workers, workers and trainees, so there would be another five levels.

00:21:37 Interviewee

Exactly, we actually have that in the parent company too. That's not really the case with us, so we don't have that. Maybe that will come, I don't know, but at our current size, we were more or less a small startup back in 2011 and now we already have over 11 billion [invested?]. So, we're already turning this thing around a lot, but I think we'll stay in this structure, which is brand new, for the next while.

00:22:16 Interviewer

OK, very good. Yes, I think those are all the questions then.

00:22:24 Interviewee

Yes, fine.

00:22:24 Interviewer

Yes, great, thank you very much. I'll stop the recording now.

Interview 5 – SW industry

00:00:00 Interviewer

I already told you what the interview is going to be about and that it will have two parts. So in the first part are some more basic questions, where you can answer with really short sentences or just a few words. And then in the second part, it's more about working together and trying to find the hierarchical order of your company. Well, you already told me that you can show me the order of the company. So maybe it's an easy part.

Can we start with the first question?

00:00:30 Interviewer

Sure, go for it.

00:00:32 Interviewer

Perfect. So, are you working in the financial service industry or in the software industry?

00:00:31 Interviewee

Practically in both. It's a financial company, and I'm working as a software engineer.

00:00:44 Interviewer

Perfect. So, it's really perfect for my purpose of studies. And for which company are you working?

00:00:52 Interviewee

N26, a financial start-up.

00:00:55 Interviewer

Cool. Do you know approximately how many employees the company has?

00:01:02 Interviewee

I don't have the latest number, but I would guess around 1600, 1700, 1800. Not a small but also not a very big company, something around that.

00:01:08 Interviewer

OK, yeah, that's more than enough. Just an approximate number. And in which department do you work?

00:01:18 Interviewee

I'm working in tech and product, and we have different segments, and right now, I am in customer risk.

00:01:29 Interviewer

OK. And which position do you currently hold? So, what's the title of your position?

00:01:39 Interviewee

Lead Engineer.

00:01:40 Interviewer

Lead Engineer. Thank you. And do you have management responsibility?

00:01:46 Interviewee

Management? So, I don't have any direct reports, but I have more or less technical responsibilities that ensure the solution is working, scalable, and maintainable. So, from the technical perspective but from the management perspective, I would say no.

00:02:04 Interviewer

OK. And did you have any promotions in your current company, and - you already shook your head; if so, the question would have been which other titles did you have, but you already answered it with a head shake.

00:02:25 Interviewee

Exactly, yeah.

00:02:26 Interviewee

OK, so I think you also already answered the questions I will ask you now. So, there are three different career paths. I don't know if you know them: So, there's the specialist career path, the project management career path, and the management career path. Do you see yourself in any of these three career paths?

00:02:45 Interviewee

Yeah. So, in the technical/ specialist path.

00:02:49 Interviewer

Perfect. That helps a lot. And what's the title of the person you report to?

00:02:57 Interviewee

It's an Engineering Manager.

00:02:59 Interviewer

Engineering manager, do/

00:03:00 Interviewee

And he's assisting the management.

00:03:02 Interviewer

Do you also know the title of whom he reports to?

00:03:06 Interviewee

Yeah. So, my manager reports to the head of engineering.

00:03:10 Interviewer

OK. So, this was already the first part, so the short question part and now I can show you my screen, and maybe we can dive a little bit deeper into the topic. I have a PowerPoint; hopefully, it helps a little bit.

00:03:35 Interviewee

Actually, one correction. So, the Manager is true; he reports to the Head of Engineering. But right now, because the position is not filled, he is reporting to the Director of Engineering. So, there's one person missing more or less.

00:03:44 Interviewer [slide 1]

OK, okay. Let me start with the first slide. Can you see my screen?

00:03:52 Interviewee

Yeah, I can see it, but it's too small to read anything.

00:03:54 Interviewer

Yeah, I know. Actually, it doesn't really matter what stands in the fields; it just to show you like a normal organigram of a company. Just to remind you kind of what my job was and what's the purpose of the study. So, in greyish, you can see the heads, and then on the right and left side, you can see that there are some people directly reporting to the board, and under the box are the units, which have also a hierarchical order. So just to clarify what we're talking about and if you know if something like that also exists in your company?

00:04:38 Interviewee

So not for the whole company, but just for product and tech, we have it.

00:04:44 Interviewer [slide 2]

Yeah. Perfect. Wait. So, we already talked a little bit about the career paths. So, on the left side, you can see an individual career path, which also called individual contribution, which is your career path, kind of, and on the right side, you can see the management career path. And what you also can see here is that both career paths kind of like overlapping each other. So, an expert can be on the same level as the top management. And this was also kind of

tricky for me to find out like how high can experts be and how are they comparable to each other, so it would be nice if we just have a look at it later and maybe find out - or maybe you can give me some insights how it is in your company.

Those are some examples; I will just skip them now. Also, if we will need them later, we can just come back to them. But yeah, I will just skip them [slide 3 and 4] if it's OK for you.

00:05:51 Interviewee

Sure.

00:05:52 Interviewer [slide 6]

Perfect. And these are the categories that I came up with – based on research, of course. And just take a minute or as long as you need to read through them. And if you recognize any of them, just tell me – this would be great.

00:06:21 Interviewee

I think overall, I mean I have not worked for companies with these titles, but I know Vice president. Actually, in my former company, we had a vice president, so I am aware of it. So yes. And also, I don't know, but friends told me, in Cisco, there have been senior vice presidents and so on. So also, no fellow/ For me, I don't really know these titles. Senior staff, we have it – no staff engineer, we had it at one point in N26, but we got rid of the role because it's too related to some other roles we have.

Yeah. Supervisor, I have no idea about this one. And yeah, senior specialist, specialist and junior specialist yes, I know them, but we just call them engineer specialist, but it's the same thing.

00:07:24 Interviewer

OK, so do you have the Engineer 1, 2, 3 kind of scheme that you're working with?

00:07:31 Interviewee

Sorry, what again?

00:07:31 Interviewer

Do you have the engineer one to five scheme. So, engineer one, engineer two, engineer/

00:07:41 Interviewee

So, we have up to 6 levels on the specialist side.

00:07:46 Interviewer

Oh 6 levels, that's good.

00:07:47 Interviewer

Up to six levels, yes.

00:07:49 Interviewer

Do you know how they are comparable, for example, like what I found was that a lead is comparable to an engineer on Stage 5.

00:08:04 Interviewee

If you want, I can show you my screen.

00:08:09 Interviewer

Yeah. Give me a second. Can you share it now?

00:08:19 Interviewee

Can you see it?

00:08:21 Interviewer

Now, I can see it. Yeah.

00:08:23 Interviewee

Is it big enough?

00:00:24 Interviewee

Yes, of course.

00:08:26 Interviewee

We start as associate engineer, so I think in your slides it starts with trainee.

And then we're a junior, engineer and senior engineer. I think this is also more or less like yours.

Then on the technical track side, we have the lead engineer and the principal engineer, so principal engineer or principal developers are also kind of common, next to senior engineer.

And your question is, how does it translate to the management and then you see a senior engineer is on the same level like an engineering manager. Lead engineer is on the same level like engineering manager so I am on the same level like my boss, just on two different tracks. And theoretically/ no, I think all principal engineers are either reporting to director of engineering processing, but I'm not sure how this works.

Yeah, but I'm not sure about/ and yeah after management, Head of Engineering and Director and after that, I think is already the C- Level, so CTO in my case. Also, other companies might have VPs or junior VPs or other management positions.

00:09:57 Interviewer

Ok. Do you mind if I take a screenshot?

00:10:01 Interviewee

No, I don't think it's sensitive data there.

00:10:04 Interviewer

OK. Thanks. Perfect. This is really helpful, and it was way easier to just see it like that, then trying to come up with it on our own. So, for your company, I understand it right that there just comes the C-level, like the C- Suit and then probably also the owner.

00:10:26 Interviewee

Yeah.

00:10:26 Interviewer

And/ but you had the V levels in different companies, right?

00:10:33 Interviewee

Exact, with another company we had a more (putting? #00:10:38#) tracks, so there was a team lead. So, there were not two tracks, just one track with a team lead. So, I have directly reported to the manager, and was also contributing, and also technical responsibility and I was reporting to my director and my director was reporting to a VP. And the VP was reporting to the CT/ founder.

00:11:06 Interviewer

OK, so the founder was the same as the CTO?

00:11:10 Interviewee

So, because the company was quite small, we were around 60 people. And also, in N26 both founders have C positions. So, the one is CTO and the other one – I have to look it up – is COO

00:11:35 Interviewer

Do you know if there comes anything beneath the associate engineer?

00:11:43 Interviewee

So, I can just talk for product and tech, again no. But we also have customer support and I have no idea how they're leveled but sure they could have.

00:11:55 Interviewer

But do you have trainees in general?

00:11:59 Interviewee

Not that I know. But like I said, maybe in other departments, in other, maybe we have them. So, I know we have also senior investigators and maybe they have trainees, but I do not know.

00:12:12 Interviewer

OK. Yeah, especially for a company like, if it's the bigger than 100 people, it's getting a little bit too much to know everything.

00:12:24 Interviewee

Exactly. We just have the tech and product overview, and I don't know if there is any full overall company overview. Where we have everything in one place.

00:12:39 Interviewer

OK, so basically/

00:12:41 Interviewee

But the other departments have for sure different types and different structures.

00:12:45 Interviewer

OK. Do you know if your company have the finance, sales, marketing and HR department?

00:12:55 Interviewee

Yes, we have for sure. But how they structured, again, I don't know really. But for sure we have a Head of HR, Head of Marketing, maybe a C-Level for this. Yeah, actually I could look this up.

00:13:11 Interviewer

Thank you.

00:13:27 Interviewee

So, Valentin is one of the founders since he has no specific title. Then OK, here we have VP, Chief Financial Officer, CTO, Chief Marketing Officer, Chief Product Officer, Chief of Staff, Senior Manager, Office Management and Chief People Officer.

00:13:59 Interviewer

OK. That helps a lot. Yeah. I wish every company has something like that publicly available.

00:14:12 Interviewee

I don't think this is not public. This is an internal tool.

00:14:16 Interviewer

No, no.

00:14:17 Interviewee

Yeah, I guess, for sure we have also on the main page some information about the whole company structure but/ Yeah, so here you have a bit information for the public: the founders, then manager of Chief Risk Officer as managing director, Chief regulatory. Yeah, so here you have even more people, that are somehow involved with the management level. Then also the board is/

00:15:09 Interviewer

Ah, so you have a supervisory board.

00:15:11 Interviewee

I think so, yeah, I could not [inaud. 00:15:17]

00:15:19 Interviewer

OK. Yeah, I just have one more question relating manager titles.

00:15:39 Interviewer [internet problems]

Can you hear me?

00:15:41 Interviewee

Yes.

00:15:42 Interviewer

OK, I think you were gone for a second. Yeah, I just have one more question relating the manager titles. I share my screen again shortly. In my research I found that the manager title is a little bit overused, kind of. And it was quite hard to structure the manager titles. Do you recognize any of these manager titles in your company?

00:16:27 Interviewee

Not really.

00:16:27 Interviewer

Oh.

00:16:33 Interviewee [thinks again]

Senior, junior, yes, for sure, sure. Senior, we have. Executive not that I am aware of, maybe in other departments we have them. Yeah, maybe it is a typical title, but I am not aware that N26 is using any of those. 00:16:56 Interviewer

OK, but in general does that kind of like differentiation makes sense to you. So that there are area managers, that you get a manager promotion title based on your seniority, and also on your skills so that you have like the department managers for example sales manager and then you also have the department in Vienna and then you have the department more globally spoken in or internationally spoken.

00:17:33 Interviewee

Probably, but this is mhh/ So product and tech is more or less is in three locations, but we don't have any department or location, yeah we have one for Vienna and one for Barcelona, because we all work together. For sure for the banking level, it's different, because if you

have banking accounts, for Spain, for example, then of course someone has to be responsible for that, so there maybe are area titles but at least I am not aware of.

00:18:08 Interviewer

So, I think that's it. I will stop the recording now, or do you have any more questions?

00:18:19 Interviewee

No.

00:18:21 Interviewer

Okay thank you, the recording is stopped.

Interview 6 – SW industry

The interview was conducted in german and translated to english afterwards

00:00:00 Interviewer

I'm pleased that we're meeting here today. Right, I would now briefly explain the aim of my master's thesis and the interview and then briefly explain the structure of the interview, would that be, okay?

00:00:13 Interviewer

Yes.

00:00:14 Interviewer

Very good. So, the aim of the master's thesis is to develop a method for determining the depth of hierarchy in the financial services sector and then also for the software industry and subsequently to validate this method. And these interviews are being used for validation. Over the last few months, I have already developed these hierarchy levels based on research and LinkedIn data from the USA. Our goal today would be to see whether it can be confirmed in practice, of people who work in the industry, what I have developed. The first part of the interview will be shorter questions, so you're welcome to answer them in one or two sentences or even with a few words, and in the second part we'll try to identify the depth of your company's hierarchy.

00:01:06 Interviewee

OK.

00:01:08 Interviewer

OK. So, the first question would be whether you really work in the software industry.

00:01:15 Interviewee

Yes, I have worked in the software sector for 3 years.

00:01:21 Interviewer

Very good and at which company?

00:01:24 Interviewee

The company was called Walter Bösch GMBH und Co KG.

00:01:29 Interviewer

Do you know approximately how many employees the company has?

00:01:32 Interviewee

Yes, the company has about 700 employees throughout Austria.

00:01:37 Interviewer

OK, and what was the name of the department you worked in?

00:01:43 Interviewee

I was in the project technology department. The larger department, it's simply called RLT Technik, or room ventilation technology, and more precisely I worked in project technology.

00:01:55 Interviewer

Okay. And what position did you have? So, what was your job title?

00:02:05 Interviewer

My job title was simply project technician in RLT, so RLT project technician.

00:02:12 Interviewer

Okay, and did you have management responsibility in that position?

00:02:17 Interviewee

No, no management responsibility.

00:02:20 Interviewer

Were you promoted during your time at RLT?

00:02:27 Interviewee

You can't really call it a promotion. I just started as an intern and then after half a year I switched to a permanent position.

00:02:40 Interviewer

Did your title change at the same time?

00:02:40 Interviewer

Just from trainee to technician, project technician.

00:02:50 Interviewer

Okay, and are you familiar with career paths? There are 3 different ones, that would be the specialist career, project management and the management career, have you heard that before?

00:03:06 Interviewer

Yes.

00:03:07 Interviewer

Would you identify with one of these specialist careers?

00:03:11 Interviewee

Yes, in my company it was more along the lines of a specialist career.

00:03:15 Interviewer

Mhm. [affirmative]

00:03:17 Interviewee

Because the RLT, the department is only a large part of what the company does. And that's more in the specialist area.

00:03:29 Interviewer

OK, and what was the name of your/ so who did you report to and what job title did the person have?

00:03:38 Interviewee

Yes, I wrote that down. I'll have to have a quick look. So, my boss, the name is not important, but that was Jonas Hxxx and he was the team leader in project technology, so project technology team leaders, in projects yes, team leader simply.

00:03:57 Interviewer

Team leader. Did you have other people in the team who were also under the team leader?

00:04:04 Interviewee

Well, that was pretty new for him. So, he was/ he was basically only promoted for me, so that he could manage me. And the goal was actually to get more team members, but I was the first. And at the time I resigned, no one else was there yet.

00:04:28 Interviewer

Okay, do you know who your team leader reported to?

00:04:33 Interviewee

Yes, that was the head of the entire project engineering department (Chef), that was Klaus Sailer, and then he was simply the head of the entire project engineering department (Abteilungsleiter).

00:04:43 Interviewer

Head of department OK, so that were actually the basic questions, part one. I'll now move on to part two and share my screen.

00:04:53

Mhm.

00:04:54 Interviewer

Right wait, I'll go to full screen just a moment. Can you see the screen?

00:05:01 Interviewer

Yes.

00:05:02 Interviewer

Great. Here we see the management career and the specialist path next to each other, so management career on the right, specialist career on the left and what we also see here is that these specialist careers overlap and that the experts at the very top can be at the same level as top management, so to speak. And one of the aims of my master's thesis is to crystallize the job titles where these people intersect [slide 2]. Right, I'll show you the overview I've created and give you a minute or as long as you need to look at it. If anything looks familiar to you, just tell me the number. However, if you have any questions, just let me know [slide 6].

00:05:51 Interviewee

Can you explain that to me again? I didn't quite understand what is going to happen.

00:05:59 Interviewer

I'm actually just illustrating that there are these two specialist careers and that they overlap. Not everyone is necessarily aware of that. And the actual goal was or would be to figure out in the interview at which positions they overlap and on the next slide we actually see the categorization that I developed in the master's thesis with all the levels that the largest company in this industry can have. In other words, it is very unlikely that your company or many companies in general will cover all the levels shown here. Exactly, but it is also possible, for example, that you are now telling me that your Senior Specialists do not actually

have to be at category level 17, but rather at the same level as the Head of. Simply because they would be two different career paths.

00:07:01 Interviewee

Mhm. [affirmative]

00:07:02 Interviewer

Exactly, but basically it would be interesting to know whether you have any of these titles, although this is of course the English version, because I have worked with English data, and these are only generic terms for groups of titles. But this overarching term reflects that. Right, I hope it works anyway, the rethinking.

00:07:20 Interviewee

Yes, uhm.

00:07:24 Interviewer

Otherwise, we can also work with the German terms.

00:07:28 Interviewee

Yes, I'm not sure how all of them, all the people above me, how that can be translated into English.

00:07:37 Interviewer

Then we'll work with the German terms, and I'll do the research.

00:07:41 Interviewee

OK. Exactly, it's also possible that nobody gives a 1 to 1 translation, but that's not a problem.

00:07:52 Interviewee

Yes, so I should just start with what's overlapping. What was it again? Management and specialization earlier, right?

00:08:02

Exactly.

00:08:07 Interviewee

I was basically the lowest level in our department and at the beginning I was more of a trainee, of course, but then at the end I was more of a specialist in the RLT area. And yes, then came the team leader. He basically just gave me the tasks. Of course, he knew his stuff very well. Then came the department manager, who was also more generally knowledgeable, but was still involved in everything from a technical point of view, I don't even know how to explain it - but then came the division manager and he was then less responsible for the technical aspects and more simply for the organization.

00:09:23 Interviewer

More personnel responsibility probably too.

00:09:25 Interviewee

Exactly, yes, in the RLT department there were also several different areas such as planning and project technology and something else, but I can't remember the names. And then there was the management level.

00:09:43 Interviewer

OK, yes.

00:09:44 Interviewee

Yes.

00:09:46 Interviewer

And at the management level, you know roughly how it was set up in the hierarchical order.

00:09:51 Interviewee

No, unfortunately, unfortunately not, I don't remember.

00:09:56 Interviewer [shows slide 7]

Okay. You just talked about the leader and the leader/ I would translate leader (Leiter) in English as manager. Precisely because manager is actually a very frequently used and widely used term and that would also partly correspond here with your [description] - so in the right-hand column you see skills. Exactly, so would recognize that there is, for example, the program manager or the project manager and then the function manager?

00:10:34 Interviewee

Yes, what do you mean by program? [refers to program manager]

00:10:42 Interviewer

Program/

00:10:42 Interviewee

So, here's another quick question, is this about the managers or about the division manager function (Bereichsleiterfunktion) or what?

00:10:54 Interviewer

Right, so it's also possible that this doesn't apply to an Austrian company, but in American companies the term manager is used a lot. But some have different levels. For example, the global manager is definitely above the office manager because he or she manages the whole thing and a global company globally and is therefore actually very high up in the chain. And then there are - traditionally, junior senior and executive titles are always awarded in relatively large companies, or in any case the senior title is very popular and simply indicates a certain level of maturity. Well, that you are above a certain level, i.e. above another position, and then there are also skills, which would then partly correspond to the specialist career path or partly to the project career path, which I have not shown here, that you initially only manage one project, but above this project there are usually other projects, or next to this project there are other projects and this is then considered a program, for example, if you now had/ would you then have all the projects that work for this RLT – is that the right term? Right, that would be the manager who manages all the projects/ everyone who works under it, that would be the program manager.

00:12:16 Interviewee

OK, yes.

00:12:19 Interviewer

Exactly. Would that be transferable?

00:12:23 Interviewee

Yes, absolutely yes. So yes.

00:12:24 Interviewer

Okay. Is there anything else you can confirm or not confirm? Well, recognize, rather.

00:12:50 Interviewee

Well, with us the specialists were all in one, so there weren't different levels for specialists, but rather that everyone was specialized in something else. Yes, then maybe the supervisor was also like the head of department (Abteilungsleiter), right?

00:13:17 Interviewer

Yes.

00:13:20 Interviewee

Yes.

00:13:33 Interviewer

But the way they're shown here now, you can't recognize the upper titles, so to speak.

00:13:40 Interviewee

Well, not really, I've had too little to do with that.

00:13:44 Interviewer

So OK, then that's it. OK. Then I'll stop the recording. Thank you.

00:13:56 Interviewee

Mhm. [affirmative]

Interview 7 – SW industry

00:00:01 Interviewer

Thanks that we came together today, and we already talked a little bit about it, but just to let you know so the interview will have two parts, kind of. And in the first part you can answer with really short sentences or just a few words. And there are more basic questions. And in the second part we will have a like more work together part where we're trying to identify the structure of your company, so hierarchical structure. Is this all right?

00:00:34 Interviewee

Yeah, yeah, sure. Go ahead.

00:00:36 Interviewer

OK, starting with the first question, so are you working in the software industry?

00:00:42 Interviewee

Yes.

00:00:43 Interviewer

And for which company?

00:00:46 Interviewee

SAP.

00:00:48 Interviewer

Do you know approximately how many employees they employ?

00:00:56 Interviewee

Worldwide or just Germany based?

00:01:00 Interviewer

Whatever number you have.

00:01:03 Interviewee

All right. The number that comes up to my mind is I think 105,000 employees worldwide.

00:01:13 Interviewer

OK. I think it's the biggest company I've interviewed so far. And in which department do you work?

00:01:24 Interviewee

The department is called Technology and innovation. So explained by job titles. I'm in the department led by the CTO, so the technical department.

00:01:38 Interviewer

OK. And in which position do you currently work in the company, so what's your job title?

00:01:47 Interviewee

I think the job title is associate developer.

00:01:52 Interviewer

And do you know which/ like do you have any job titles that come beneath you?

00:02:04 Interviewee

Yeah, I wish I could say yes, but no.

00:02:07 Interviewer

Trainees would be alright as well.

00:02:11 Interviewee

Yeah, yeah, trainees and working students, but otherwise no, for full time.

00:02:19 Interviewer

And to which job position do you report?

00:02:27 Interviewee

I think it's just plainly called manager.

00:02:31 Interviewer

Plainly called manager.

00:02:32 Interviewee

Of software developer - development manager. I guess it's manager.

00:02:36 Interviewer

Ok. But definitely manager that's interesting.

So yeah, there are three kind of career paths, the specialist career path, the project management, career path and the management career path. Do you know them or that you heard before of them?

00:03:00 Interviewee

I do know them but concerning my job as a developer, there's only two tracks. That I can embark on.

00:03:16 Interviewer

OK, which 2 tracks would that be?

00:03:24 Interviewee

Hmm. I never really asked what the name is, but one track is definitely called the management track and the other track is something like – I think it's called the expert track.

00:03:39 Interviewer

Oh yeah, so this is like probably the same as the specialist career path, but I can show you a slide a little bit later. So maybe an example makes it a little more clear. OK. And this was actually the first part of the interview, and I will go on with sharing my screen. I hope the internet connection is good enough. Just tell me when you can see my screen. Oh, now you see the whole screen, right?

00:04:10

Yep, Yep.

00:04:11 Interviewer

OK. So, this [overview] is really small, and it doesn't really matter what's specifically in the fields. Just to remind you what a company structure can look like. So, in gray, you see the board and then on the left and right side you see people who directly report to the board and then you see the different units. Is this clear for you?

00:04:41 Interviewer

Mhm mhm. [affirmative]

00:04:41 Interviewer

OK. Perfect. Yeah, and here we come back to the 2 tracks. So, on the left side, we have the individual contribution track, which is kind of the expert track. I guess that's what you were talking about. And on the right side is the management track. So, as you can see, you kind of just start an individual contribution track and at one point you can decide for the management track, but you also can stay in the individual contribution track, and if you're really up high in the individual contribution track, you can be as high as the top management. Although like they work a little bit differently, so you probably don't have control over employees because your individual knowledge is so valuable for the company that they would just like bury you with, like organizational stuff and they wouldn't do that. Yeah, like does it make it a little more clear.

00:05:47 Interviewee

Yeah, yeah. Totally makes sense.

00:05:50 Interviewee

So, would you describe your path as the expert path, or like the path you're aiming for, or as a management path?

00:06:00 Interviewee

Well, as you correctly portrayed like the start of the path always is the individual contribution path, so that's where I'm at right now, but I do aim to pursue the management path at some point.

00:06:19 Interviewer

OK, that's interesting. So here we have basically the categories that I kind of developed during my master thesis. Maybe take a minute and just read through them. Also, when you recognize any of the categories you can tell me or just read them all and then we talk about it together.

00:07:18 Interviewee

OK, I've read through them.

00:07:19 Interviewer

OK. Do you recognize any of them in your company?

00:07:26 Interviewee

Yes, starting from the top, the first would be CEO. Then I think terms like president. Forms of director are not used in our company or not to my knowledge.

00:07:55 Interviewer

But do you have President and Vice president, do you use both of them?

00:08:03 Interviewee

As I said, I wouldn't know who the president of my company is, so I only know -

00:08:07 Interviewer

Oh, you don't? You don't have to know the name. Just if there is a president.

00:08:09 Interviewee

No, no what I wanted to say is I don't know if there's a position called President, just the CEO is the Head of the Executive board.

00:08:16 Interviewer

Ah, OK. OK. Sorry.

00:08:26 Interviewee

Yeah. And then the next one I guess would be manager. So that's the terms that we use. So, we have the board and below the board there's some layers of management.

00:08:46 Interviewer

Mhm.

00:08:47 Interviewee

And then it trickles down like you have/ So for a little bit of background because I work in a German company, we have a collective wage agreement, so there are certain positions where there's a spectrum of wage that you can get and there are certain job titles and no matter what specific area you work in you get round about the same money and you are a part of this collective wage agreement and at some point you can like break through the ceiling of this collective wage agreement. And that's the top management positions, so I don't know if they have different job titles.

00:09:45 Interviewer

Mhm. [affirmative]

00:09:48 Interviewee

But in the collective wage agreement the highest would be senior manager. And below would be manager and I guess a manager can have other managers below him. And then at some point the manager just has a team of junior to senior specialists.

00:10:14 Interviewer

OK, so you just work with junior, and senior add ons to the titles?

Right? I also have another slide that might be helpful regarding the manager categories. I came up with/ like there are a lot of different managers and a manager is not equal to a manager.

00:10:36 Interviewee

Yeah, yeah.

00:10:37 Interviewer

I thought that it might be useful to have managers in three different categories. So, area based, seniority based, and skills based. Yeah, and you already talk about the seniority, so you have senior engineers. Do you also know if you have executive managers?

00:10:58 Interviewee

No, I don't think so, at least I haven't heard the term in the company.

00:11:10 Interviewer

OK. Do you recognize any of the other manager titles?

00:11:20 Interviewee

So, concerning the area, not really. Concerning the seniority, we don't have junior managers, we just have managers and senior managers.

00:11:32 Interviewer

Mhm. [affirmative]

00:11:33 Interviewee

And as I said, everything above that is outside of the collective wage agreement. So, they probably do not have a defined job title.

00:11:46 Interviewer

Ah, OK.

00:11:52 Interviewee

So as soon as you break out of this ceiling you probably can talk with the HR department and then agree on some job title, but it's not written down in a collective agreement.

00:12:15 Interviewer

OK.

00:12:17 Interviewee

And skill based, we do have project managers. They are called product managers. And I think here the terms mix up, so, like managers in the development area are managers because they manage people and product managers are product managers because they manage products, and they can also manage people, but they don't have to.

00:12:55 Interviewer

Yeah. OK. Yeah, managers to categorize is quite hard.

00:13:03 Interviewee

Yeah, yeah.

00:13:06 Interviewer

OK, so you definitely work with senior and with the normal title just without any additional words. OK, I just go back to the other slide. Do you maybe know if you have like a supervisor or so. Which would be category "Supervisor", 16.

00:13:37 Interviewee

Nope, not to my knowledge.

00:13:40 Interviewer

OK.

00:13:40 Interviewee

Like below the manager, it's just senior developer, developer, or associate developer.

00:13:47 Interviewer

OK, so you also don't work with developer 1, 2, 3, 4, 5.

00:13:59 Interviewee

Yeah, there are three grades of developers and they're called associate developer, which then if he or she gets promoted, gets to be a software developer and then with the next promotion, that's to be a senior software developer. That's the job titles.

00:14:17 Interviewer

OK, got it. Have you ever heard of the word staff? Because that's what I run over. But like I couldn't get it confirmed by now. Like staff developer, for example. Not staff as a general term for employees.

00:14:46 Interviewee

No. Nope, never heard of it.

00:14:50 Interviewer

OK, so the exact job title of the next person above you is manager, right?

00:15:03 Interviewee

Yeah.

00:15:05 Interviewer

And do you know the job title above him or her?

00:15:10 Interviewee

Yeah. Yeah, that's senior manager.

00:15:12 Interviewer

OK. And if you go further up the chain, do you know any more of those titles? So, manager senior manager. And then?

00:15:30 Interviewee

As I said above that there is no like collective agreed upon title for the managers.

00:15:39 Interviewer

Yeah, totally.

00:15:42 Interviewee

I think it would be like a specific title. Like when you talk about it with other employees, you just say like top management.

00:15:53 Interviewer

OK.

00:15:53 Interviewee

And that's that, but no, I don't know the job title.

00:16:00 Interviewer

OK, I thought maybe you know like the specific one, by being in contact with the senior manager and he is like "oh, you know I have to talk with x."

00:16:12 Interviewee

I can have a look into our company structure.

00:16:20 Interviewer

Oh yeah, this would be amazing.

00:16:21 Interviewee

There might be some titles flying around there.

00:16:28 Interviewer

Yeah, it would be amazing because most of the time those data are not publicly available. Like, I don't think they're too sensitive, but still most of the time it's just don't available for normal people in the internet.

00:16:48 Interviewee

Yeah.

00:16:49 Interviewer

Can you look it up now or should I just come back to you at another point?

00:16:54 Interviewee

Yeah, I'm looking at it right now. But what I'm seeing is/ The only thing that's visible for me is the title manager.

00:17:07 Interviewer

OK. So, there's not like very much information available. I think all that I, as an employee, looking into the org chart should get from it is that if the person is a manager, then that means that he or she has a team of people that he or she manages and that's that. But I cannot see the job title.

00:17:36 Interviewer

OK. That's really interesting. I thought like for a big company as SAP, it would be like public.

You know when you want to get this information as an employee you can see how much you have to climb up the ladder to be like in the really, really top management.

00:18:00 Interviewee

Oh, I do see how many people are above me. I don't see the job title, yeah.

00:18:08 Interviewer

Ah, OK.

00:18:11 Interviewee

I drove down and now I do see some more information. So, it is divided into Manager, managing people and Manager, managing managers.

00:18:34 Interviewer

And they are all called managers. Wow. This doesn't make it complicated at all.

00:18:44 Interviewee

And/ Ah, OK, I see the external titles. If you're interested in that, because what I see is that all people who are inside the collective wage agreement don't have an external title because the internal title is the same as the external title, senior manager, manager. But for the people above that, there's an external title.

00:19:18 Interviewer

Oh yeah, this would be great.

00:19:20 Interviewee

If you're interested so/

00:19:23 Interviewer

Please go on.

00:19:28 Interviewee

Manager. So above me is the manager. Above him is the senior manager above that is someone with the external title data architect.

00:19:47 Interviewer

I have this title actually in my code.

00:19:51 Interviewee

Ok, and now we have proof that I'm a liar. The external title of the person above that is senior vice president of product development.

00:20:06 Interviewer

But then you are already at vice president?

00:20:10 Interviewee

Senior Vice President of product development. Yeah.

00:20:14 Interviewer

OK. And above that, probably just the C-Suit

00:20:20 Interviewee

And above that is CTO.

00:20:22 Interviewer

CTO, and then?

00:20:26 Interviewee

Yeah, that's it. That's the end of the org chart.

00:20:29 Interviewer

OK, this was helpful.

00:20:33 Interviewee

All right.

00:20:35 Interviewer

I think, I mean, that's probably all the information I need. I mean you can't see anything else right?

00:20:42 Interviewee

Yeah, sorry, I could have given it to you directly. Since I have access to the org chart.

00:20:50 Interviewer

No, it's alright. Just one more question because I also work with the support functions. So, you have the finance, sales, marketing, HR and law support functions in your company?

00:21:13 Interviewee

What exactly do you mean by support function?

00:21:18 Interviewer

So I mean SAP is a software company, so you sell software. But for a bigger company, I think it's probably normal that you also need people who work in HR. People who work in a marketing team, and I mean, if you have those support units, so HR, marketing, finance.

00:21:43 Interviewee

OK, everything that is outside the core area of the company?

00:21:46 Interviewer

Right.

00:21:48 Interviewee

Yeah.

00:21:52 Interviewer

But you don't know how they are structured by accident, right?

00:21:58 Interviewee

It's a good question. If I look into the org chart. I can only see the org chart of the area that I am in, so everything that's below the CTO.

00:22:24 Interviewer

Mhm. [affirmative]

00:22:31 Interviewee

I search a little bit. Yeah, seems like I can only look above and below me.

00:22:47 Interviewer

Below, is there anything else?

00:22:50 Interviewee

Yeah, I had to mention it, but no, there's just – there is nothing, no.

00:22:58 Interviewer

OK. Just to be sure, not that I miss anything there, but this aligns actually with my code, so this is good.

00:23:08 Interviewee

OK.

00:23:08 Interviewer

OK. This is like basically my main work, and it was nice to talk about it with you. So, I would stop the recording if this is all right, or do you have any questions for me at this point?

00:23:25 Interviewee

No.

Interview 8 – SW industry

The interview was conducted in german and translated to english afterwards.

00:00:00 Interviewer

Yes, actually, now (I'm recording). Right, so the topic of the dissertation is to develop a method for the software industry to determine the depth of hierarchy. And that's what I've been doing over the last few months, using LinkedIn data from the USA. And the interview is now intended to support the validation of this data and the hierarchy that I have defined.

Exactly, the interview will be conducted in two parts. In the first part, you can answer very briefly, in one or two sentences or even words, and in the second part we will then work together on your company organization chart or the depth of your company's hierarchy. Does that sound alright to you?

00:00:41 Interviewee

Mhm, yes.

00:00:44 Interviewer

Do you work in the software industry?

00:00:51 Interviewee

In the broadest sense, yes. Well, I tend to work for a publisher, but in IT or rather in the software department.

00:01:00 Interviewer

Okay. What's the name of the company?

00:01:06 Interviewee

The company used to be called VNR Verlag, but it has changed its name to Platform X GmbH & Co. KG.

00:01:13 Interviewer

Do you know approximately how many employees the company has?

00:01:18 Interviewee

Well, the whole publisher has around 400 to 500 employees. I think Platform X has about 200.

00:01:26 Interviewer

OK, and what's the name of the department you work in?

00:01:32 Interviewee

Team Platform Engineering.

00:01:36 Interviewer

OK, and what's your position?

00:01:44 Interviewee

Officially I'm a working student, but the job title is platform engineer.

00:01:52 Interviewer

Platform Engineer, OK. Do you have management responsibility?

00:02:03 Interviewee

Not directly. Well, if we get an intern or someone like that or a trainee, then I'm supposed to show them things, but I'm not authorized to give instructions, no.

00:02:13 Interviewer

Okay. Have you already had a promotion in the company and if so, has your job title changed?

00:02:25 Interviewee

Promotion would now mean something like from Junior to Senior or from Junior to /

00:02:34 Interviewer

Engineer, yes.

00:02:36 Interviewee

Exactly, no.

00:02:40 Interviewer

There are three career paths in total, does that ring a bell? There's the specialist career path, the project management career path and the management career path, have you ever heard of those?

00:02:54 Interviewee

I can guess, but maybe you can explain it to me again.

00:02:56 Interviewer

OK, so with the specialist career path, you actually have expert knowledge. And as a software engineer, for example, your expert knowledge is so valuable that they don't want to involve you in administrative tasks. And that would be the specialist career path. But, as I will also show in a moment, someone who is very high up in their specialist career can be just as high as top management. In other words, in the hierarchy.

The management career path would then be the classic top management path. And then there's the project career path, where you have larger projects, and there's a bit of a hierarchical structure of its own, which is partly a mix of the two. Right, where would you see yourself in this? But I can also show you my screen.

00:03:55 Interviewee

So, management career / leadership career is also really when you are specifically in the management position, so when you start working in small or middle management and then in high management.

00:04:08 Interviewer

Exactly! Can you see my screen.

00:04:12 Interviewee

Yes.

00:04:13 Interviewer [slide 2]

So, you can actually see the management career path on the right, which doesn't start at the bottom. And on the left, you see this Individual Contribution Path, i.e. the specialist career path, which starts at the bottom, but at some point these two paths overlap.

00:04:32 Interviewee

Okay, so according to this model, if you're an expert then you're at least in middle management.

00:04:40 Interviewer

Actually, yes

00:04:41 Interviewee

Okay.

00:04:46 Interviewer

And you are?

00:04:48 Interviewee

I see. I would see myself as somewhere between Skilled and Professional in my field.

00:05:01 Interviewer

Okay. Well, that was actually the first part. Another important thing to say here is that it was relatively difficult for my work to determine with which job title you are actually in the Individual Contribution Path and with which job title you are in this Management Path, or

for which job titles these two paths overlap. So, if you have any input on this later, please feel free to share.

Exactly, and the [Slide 6] is actually the work of the last few months. It looks very simple. Or rather, the overview only shows you an overview of categories, in this case for English companies and in English. I hope that's not a problem to rethink here. Take a minute and see if you recognize any titles in your company.

00:06:33 Interviewee

Mhm. [affirmative]

00:06:35 Interviewer

And?

00:06:42 Interviewee

Well, the company where I work is officially a public limited company, but it's not public, but there are "owners", "board members" probably too, there's definitely a CEO. We don't have vice presidents. Exactly and yes –

00:07:13 Interviewer

Do you have CTOs, or –

00:07:18 Interviewee

So yes, we have CTO, CEO, COO, CFO, I think. I don't think we have CIO anymore.

00:07:30 Interviewer

OK, but they come directly below the CEO? So, there's no level in between?

00:07:38 Interviewee

I'm just wondering whether they should be placed underneath or at the same level, but I think that they also report to the CEO. So, there is no Managing Director between CEO and C-

Level.

00:07:57 Interviewer

Okay, you don't have the VPs, the Vice Presidents. Do you have Head Ofs?

00:08:04 Interviewee

Exactly. For us it continues with Head of xxx. So, my boss's boss is the "Head of Software Solutions", and my boss is also the "Head of Platform Engineering", so of the team I'm in.

00:08:30 Interviewer

So, another subdivision in the Head Ofs actually, in terms of management responsibility.

00:08:37 Interviewee

Yes.

00:08:41 Interviewer

And what are the names of your colleagues? Well, you've already said your title. But did you

00:08:47 Interviewer

Right –

00:08:49 Interviewer

Yes, sorry.

00:08:51 Interviewee

We're platform engineers, me and two colleagues, and our boss is the team leader and head of platform engineering, because we're the platform engineering team and he - if he's the team leader, he's also the head of platform engineering in some sense.

Exactly. And then all the teams have sub-teams - no, actually all the teams have a team leader and he's also a Head of and they all report to the Head of Software Solutions, who's in charge of everything. And although he's not a C-level, he's at the same level as the C-level. It's a bit complicated, so he's also the company director (Geschäftsführer), so to speak.

00:09:47 Interviewer

Okay.

00:09:50 Interviewee

Because that's such an important sector for us.

00:09:54 Interviewer

Ah, which brings us to the management and individual career path. Because of the importance your department has in the company, it makes sense to place your managers at the very top so that they can report directly—not via several levels.

00:10:13 Interviewee

Exactly. So, I think you could also call him a CTO—so I don't think there's much more to it.

00:10:24 Interviewer

Ok, do you still have classifications within the company like senior or junior?

00:10:37 Interviewee

Not officially, so you don't see them, but I think you start internally as a junior and that's how you start. But if I look somewhere in Teams or in Outlook, it doesn't say Senior.

00:10:54 Interviewer

But you don't work with a different system either, so the software industry often classifies its employees as Developer 1, Developer 2, Developer 3, for example. So not that either, right?

00:11:07 Interviewee

No.

00:11:10 Interviewer

And you said at the beginning that there were still trainees below you?

00:11:18 Interviewer

Exactly. And next to your –

00:11:21 Interviewee

But it's not as if I exist among my colleagues. I'm more or less on the same level as my colleagues, under the leadership of my boss.

00:11:28 Interviewer

Okay, so company affiliation, for example, is not really "relevant" for the job title here. Are there any other units besides your unit that you know of?

00:11:49 Interviewee

Yes. Should I mention them to you, the name, right?

00:11:54 Interviewer

Actually. Do they still report to the same person?

00:11:59 Interviewee

Yes.

00:12:01 Interviewer

How many are there? Do you know or can you tell me the name?

00:12:07 Interviewee

Phew. Speaking off the top of my head, I would say there are seven to eight teams reporting to one person, and because we are a publishing house, we also have these individual specialist publishers (Fachverlage). So, I think there are seven or eight divisions (Fachverlage) specialist publishers and they also have sub-teams and each specialist

publisher has a specialist publishing manager, I think. But that's no longer in the IT sector, right? So, then we just leave our IT sector.

00:12:54 Interviewer

Okay, so then you're a support function in the company. And other functions would be - I've just categorized them as Finance, Sales, Marketing, HR and Law, would that fit for you as well?

00:13:13 Interviewee

Finance, yes. Law, I think we only have one or two lawyers, so it's not that big. HR of course, yes. What else we have - Consulting, I don't know. Consulting but not externally, they do a lot of project management, for example "We want to develop a new product now" and then they take over the project management or rather the product owner role and then we develop.

00:13:54 Interviewer

Would they then be at the level of a manager, so to speak, i.e. product manager? So, above developers in theory, or would you report to them.

00:14:14 Interviewee

If you want to see it that way, probably yes.

00:14:17 Interviewer

OK, in general I noticed that you don't work with the title Manager at all. Is that right?

00:14:26 Interviewee

Yes.

00:14:27 Interviewer [slide 6]

That's interesting. Very interesting. Okay. I think that covers all the questions, though. Or do you think you can expand on the diagram, or that we have discussed everything that would be relevant for this illustration? Or do you see mistakes in it that you know differently in other companies or from acquaintances/friends, that it is different?

00:14:57 Interviewee

Well, I know about the Vice President, that they have them between C-Level and Head of, but otherwise I would say that looks plausible.

00:15:14 Interviewer

So, the VPS are, as you know, also below C-level or are they above C-level?

00:15:22 Interviewee

They're above C-Lev/ no, they're above C-Level!

00:15:26 Interviewer

Okay. Thank you.

Appendix 13: Salary Data Scraping – Code

scrap salary data

```
#create a list with unique company names
df_all.info()
experience_company_list = df_all['experience_company'].unique().tolist()
#check number of unique companies
len(experience_company_list)

#set up salary data scraping from H1B
import requests
from bs4 import BeautifulSoup
import pandas as pd
import time
import os
import datetime

session = requests.Session()
session.max_redirects = 100
headers = {
    "User-Agent": "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/100.0.4896.60 Safari/537.36"
}
session.headers.update(headers)

def scrape_data(experience_company_list, output_file, start_position):
    base_url = "https://h1bdata.info/index.php?em="
    all_rows = []

    for index, company in enumerate(experience_company_list[start_position:], start=start_position):
        print(f"Scraping data for {company} at position {index}...")
        try:
            response = session.get(base_url + str(company), timeout=10)
            if response.status_code == 200:
                soup = BeautifulSoup(response.content, "html.parser")
                table = soup.find("table", {"class": "tablesorter tablesorter-blue hasStickyHeaders"})
                if table:
                    rows = table.find_all("tr")[1:] # Skip header row
                    for row in rows:
                        data = [td.text.strip() for td in row.find_all("td")]
                        if len(data) == 6:
                            all_rows.append(data)
                    time.sleep(1)
                else:
                    print(f"Failed to fetch data for {company}. Status code: {response.status_code}")
            except requests.exceptions.RequestException as e:
                print(f"Error fetching data for {company}: {e}")

        # Save the current position after each company
        with open("last_saved_position.txt", "w") as file:
            file.write(str(index + 1)) # +1 to start from the next company on restart

    # Create DataFrame from collected rows and save to CSV
    columns = ["company", "job_title", "salary", "location", "submit_date", "start_date"]
    result_df = pd.DataFrame(all_rows, columns=columns)
    result_df.to_csv(output_file, index=False)
    print(f"Data saved to {output_file}")

directory_path = '/Volumes/X9 Pro/Masterarbeit-extern/CSV files created/financial/financ
```

```

ial_all.csv'
df_all = pd.read_csv(directory_path)
experience_company_list = df_all['experience_company'].unique().tolist()

#if scrapping stops unexpectedly start at last saved position
last_saved_position_file = "last_saved_position.txt"
start_position = 0

# Load the last saved position if available
if os.path.exists(last_saved_position_file):
    with open(last_saved_position_file, "r") as file:
        start_position = int(file.read())

output_directory = "/Volumes/X9 Pro/Masterarbeit-extern/CSV files created/financial/financial_salary/"
output_file = f"{output_directory}output_{datetime.datetime.now().strftime('%Y%m%d_%H%M%S')}.csv"

scrape_data(experience_company_list, output_file, start_position)

```

Rename file and count companies

```

import os

# Current file path
current_file = '/Volumes/X9 Pro/Masterarbeit-extern/CSV files created/financial/financial_salary/output_20240413_174853.csv'

# New file name
new_file = '/Volumes/X9 Pro/Masterarbeit-extern/CSV files created/financial/financial_salary/financial_salary_all.csv'

# Renaming the file
os.rename(current_file, new_file)

import pandas as pd

df_count = pd.read_csv('/Volumes/X9 Pro/Masterarbeit-extern/CSV files created/financial/financial_salary/financial_salary_categorized_1805.csv')
# Count unique values in the 'company' column

unique_count_companies = df_count['company'].nunique()
unique_count_titles = df_count['job_title'].nunique()

# Print the count
print("Unique count in 'company' column:", unique_count_companies)
print("Unique count in 'job titles' column:", unique_count_titles)

```

Appendix 14: Categorize salary data – Code FS

Apply categorization code to the salary data

```
import pandas as pd
import os
from fuzzywuzzy import fuzz

df_salary = pd.read_csv('/Volumes/X9 Pro/Masterarbeit-extern/CSV files created/financial/financial_salary/Archiv 1805/financial_salary_all.csv')

#data cleaning
def convert_to_lowercase(dflow):
    # Loop through each column in the DataFrame
    for column in dflow.columns:
        # Check if the column's data type is 'object', typically used for strings
        if dflow[column].dtype == 'object':
            dflow[column] = dflow[column].str.lower()
    return dflow

# Assuming df_salary is already defined and contains string data
df_salary = convert_to_lowercase(df_salary)

df_salary['job_title'] = df_salary['job_title'].astype(str)

cond_external = df_salary['job_title'].str.strip().str.contains('nurse|nursing|surgeon|internist|pharmacist|pharmacy|medical|therapist|physician|dental|health|pharmaceutical|paramedic|clinical|navy seal|air force|k-9|pilot|police|police officer|detective|investigator|lieutenant|captain|fbi agent|sergeant|guard|crime|firefighter|fire|rescue|technical|electrician|mechanic|automotive|auto|parts|car|driver|manufacturing|manufacture|production|logistics|assembly|warehouse|electrical|plant|operator|energy|telecommunication|network|wireless|farmer|agriculture|crop|environmental|environment|biologist|geologist|climate|paymaster|city|diplomat|fundraising|fundraiser|receptionist|realtor|real estate|property|historian|teacher|lecturer|teaching|phd|professor|postdoc|postgraduate|researcher|undergraduate|student|chef|stylist|fashion|wardrobe|store|cashier|photographer|graphic designer|event planner|janitor|actor|actress|artist|singer|musician|editor|author|writer|proofreader|bookseller|waitress|waiter|barista|bartender|pcb|resident assistant|pastor|freelance|no_info|seeking employment|stay at home|homemaker|dad|mom|mother|wife|family leave|retired|retirement|unemployed|not working|retiree|sabbatical|summer|plumber|various|personal trainer|coach|nanny|mover|library|librarian|linguist', case=False)

combined_conditions_df = df_salary[~(cond_external)]

#Define the rules for categorization
rules = [
    (1, ['owner', 'founder', 'chairman', 'creator', 'created', 'made', 'ownere', r'\bpartner\b', 'co-owner', 'business partner', 'founding member']),
    (2, ['board member', 'board of directions', 'board of directors']),
    (3, ['leading partner', 'managing partner']),
    (4, ['ceo', 'chief executive', 'cheif executive']),
    (5, [r'\bcco\b', r'\bcdo\b', r'\bcfo\b', r'\bcho\b', r'\bcio\b', r'\bclo\b', r'\bcmo\b', r'\bcoo\b', r'\bcpo\b', r'\bcso\b', r'\bcto\b', 'ccao', 'cceo', r'\bcos\b', r'\bcro\b', 'chief officer', 'chief\s+\w+\s+officer', 'chief technology officer', 'chief operating officer', 'chief people officer', 'chief marketing officer', 'chief staff officer', 'chief financial officer', 'chief revenue officer', 'chief product officer', 'chief information officer', 'chief sales officer', 'chief of staff', 'chief investment officer', 'chief strategy officer']),
    (6, ['president', 'presidente']),
    (7, [r'\bmd\b', 'managing director']),
    (8.1, [r'\bvp\b', 'executive vp', 'executive vice president', 'executive vice-president', 'exec vice pres']),
    (8.2, ['senior vice president', 'sr vice president', 'snrvp', 'svp', 'senior v.p.', 'sr. vice president', r'\bsenior\b', 'sr vp', r'\bsr.\b', r'\bsenior\s+\w+\s+vp\b'])
]
```

,
 (8.3, [r'\bvp\b', r'\bv\.?p\b', 'vice president', 'vice-president', 'vice presidente', 'vice pres', r'\bv\.?s?p\b', r'\bvp\b']),
 (8.4, ['avp', 'assistant vice president', 'junior vice president', r'\bjr. vp\b', r'\bjn\b\s+\vp\b']),
 (9.1, ['executive director', 'exec. dir', 'exec director']),
 (9.2, ['senior director', 'sr. director', r'\bsenior\s+\w+\s+director\b']),
 (9.3, ['director', 'directo', 'diercto', r'dir\b', 'dierctor']),
 (9.4, ['junior director', r'\bjun dir\b', 'assistant director', r'\bjunior\s+\w+\s+director\b', r'\bassitant\s+\w+\s+director\b', 'jr. director', r'\bjr dir', r'\basst dir /b', r'asst\s+\dir']),
 (10, [r'\bhead\s+\w+', r'\bheads\b', r'\bhead\s+\of\w+', 'r\w+\s+\head\s+\w+', 'r\w+\s+\head', 'head of']),
 (11.01, ['program manager']),
 (11.02, ['project manager', r'\bpmo\b', 'project management officer', r'senior \bpm\b', r'pm\b', r'pm\b manager']), *#project managers all*
 (11.03, ['banking center manager', 'banking manager', 'city comptroller', 'treasurer', 'comptroller', 'fiscal manager', 'international bank manager', 'controller', 'bank manager']), *#manager finance (department managers)*
 (11.1, ['executive manager']),
 (11.2, ['senior manager', 'senior management team', 'sn manager', 'sr. manager', 'sr mng', 'sr manager', r'\bsenior\s+\w+\s+\manager\b', r'\bsr.\s+\w+\s+\manager\b', r'\bsenior\s+\w+\s+\w+\s+\manager\b']),
 (11.3, ['manager', ' manager', 'manger', 'manager', 'mgr', r'gm\b', 'department manager', 'division manager']),
 (11.4, ['junior manager', 'assistant manager', r'\bassitant\s+\w+\s+manager\b', 'management assistant', 'coordinator']),
 (12, ['lead', 'leader', 'lmts', 'lead member of']),
 (13, ['supervisor', 'instructor', 'instructor', 'mentor']),
#14 senior specialists
 (14.01, ['buyer', 'client relationship manager', 'buying agent', 'broker', 'dealer', 'trader', 'purchasing agent', 'purchaser', 'importer', 'purchasing', 'tradesman']),
 (14.02, ['coordinator', 'consultant', 'financial consultant', 'personal finance coordinator', 'pfc']),
 (14.03, ['accountant', 'auditor', 'audit', 'assurance', 'cash management']),
 (14.04, ['credit analyst', 'portfolio credit manager', 'pcm', 'claims analyst', 'credit risk analyst']),
 (14.05, ['wealth analyst', 'investor', 'wealth strategist', 'underwriter', 'underwriting', 'analyst', 'associate', 'financial planner', 'financial planning', 'adviser', 'advisor', 'generalist', 'pcii', 'personal client investment specialist', 'private client associate', 'pca', 'risk control analyst', 'rca', 'pcis', 'personal finance specialist', 'pfs', 'private banking specialist', 'pbs', 'private banking officer', 'portfolio management officer', 'portfolio officer', 'pmo', 'portfolio development manager', 'pdm', 'stock analyst', 'trust officer', 'loan funder', 'mortgage officer', 'revenue officer', 'financial risk officer', 'finance officer', 'financial officer', 'portfolio manager', r'\bpm\b', 'portfolio administrator', 'bank officer', 'financial counselor', 'individual pension consultant', 'insurance writer', 'pricing evaluator', 'asset management', 'financial services officer', 'financial service officer', 'financial services strategist', 'finance service advisor', 'account executive', 'insurance producer', 'client relationship manager'])
 ,
 (14.06, ['loan examiner', 'loan reviewer', 'bank compliance officer', 'bank examiner', 'financial compliance examiner', 'home mortgage disclosure act specialist', 'payroll examiner', 'pension examiner', 'examiner', 'claims investigator', 'appraiser']),
 (14.07, ['case manager', 'loan validator', 'validator', 'credit adjuster', 'adjuster', 'credit officer', 'loan originator', 'originator', 'loan officer', 'loan administrator', 'credit counselor', 'counselor', 'loan concierge', 'financial respondent', 'lender', 'lending officer', 'debt management counselor', 'financial aid counselor', 'loan counselor', 'lending officer', 'commercial lender', 'loan analyst', 'loan reviewer', 'credit manager']),
 (14.08, ['tax examiner', 'preparer', 'revenue agent', 'licensed tax consultant', 'income tax preparer', 'income tax advisor', 'tax', 'city collector', 'customs appraiser', 'tax adjuster', 'revenue agent', 'revenue collector', 'revenue enforcement agent', 'tax compliance officer', 'tax compliance representative', 'tax investigator', 'corporate tax preparer', 'tax consultant', 'tax specialist']),
 (14.09, ['financial specialist', 'financial expert', 'practitioner', 'specialist', 'advocate', 'financial administrator', 'trust administrator', 'actuary', 'financial servi

```

ces professional', 'bondsman', 'superintendent', 'financial services dealer', 'financial
services professional', 'financial service concierge', 'financial professional', 'financ
ial representative', 'claims representative', 'financial services rep' ]),
#15 senior workers
(15.01, ['reposessor', 'debt collector', 'collecting agent', 'telex operator', 'acc
ounts collector', 'bill collector', 'billing representative', 'collection agent', 'colle
ctions clerk', 'collections representative', 'installment agent', 'loan collector', 'ins
urance collector', 'payment collector'])),
(15.02, ['wholesale processor', 'agent', 'billing clerk', 'invoice clerk', 'invoice
control clerk', 'posting clerk', 'statement clerk', 'statement processor'])),
(15.03, ['broker', 'investment banker', 'bookkeeper', 'bookkeeping', 'book keeper',
'accounts receivable assistant', 'accounts receivable clerk', 'auditing clerk', 'fixed c
apital clerk', 'foreign exchange position clerk', 'mortgage accounting clerk', 'financia
l services administrator'])),
(15.04, ['teller', 'foreign exchange clerk', 'money order clerk', 'vault teller'])),
(15.05, ['authorizer clerk', 'credit reviewer', 'checker', 'credit rating checker',
'credit investigator', 'credit processor', 'credit reference clerk', 'finance writer'])),
(15.06, ['loan opener', 'mortgage processor', 'mortgage closing clerk', 'mortgage lo
an closer', 'mortgage loan processing clerk', 'banking services clerk', 'new accounts ba
nking representative', 'new client banking', 'services clerk', 'loan interviewers', 'loa
n servicing', 'loan service', ])),
(15.07, [ 'personal banker', r'\bpb', 'personal banking officer', 'pbo', 'banking of
ficer', 'banker', 'bank vault', 'financial operations clerk', 'financial reserve clerk']
),
(15.08, ['personal assistant', 'assistant', 'assistan', 'assitant', r'\bpa\b', 'alum
ni secretary', 'office secretary', 'personal secretary', 'asst', 'secretary'])),
(15.09, ['processor', 'investment service', r'\bstaff\b', r'\bstaf\b', 'clerk', 'empl
oyee', 'administrative clerk', 'office assistant', 'office clerk', 'office administrator
', 'financial services operator', 'financial services support', 'research assistant', 'r
esearchassistant'])),
(16, ['trainee', 'trainie', 'research student', r'\bintern\b', 'volunteer', 'trainin
g program', 'mentee', 'apprentice'])),
(30.01, ['developer', 'software development', 'software', 'engineer', 'ux designer',
'programmer', 'programming', 'solution architect', 'solutions architect', 'system archit
ect', 'software architect', r'\bsw\b', 'web designer', 'it consultant', 'programmer ana
lyst', 'systems analyst', 'test analyst', 'qa analyst', 'qa tester', 'qu tester', 'soluti
ons analyst', 'performance analyst', 'programmer analyst', 'principal developer', 'princ
ipal engineer', 'data specialist', 'unix system', r'\bsql\b', r'\bui\b designer', 'sql', '
sap', 'system administrator', 'systems administrator', 'system administrator', 'security
officer', 'systems administrator', 'data scientist', 'application architect', 'database
administrator', 'enterprise architect', 'client service', 'client advocate', 'client', '
client support', 'support representative', 'customer service'])),
(30.02, ['sales', 'sales representative', 'registered representative', 'reg. rep', '
representative', 'registered rep', 'sales manager', 'sales associate manger', 'sales tra
ining manager', 'sales consultant', 'salesforce', 'sales associate', 'sales agent', 'sal
es assistant', 'procurement'])),
(30.03, ['marketing', 'advertising', 'promotions', 'marketing manager', 'brand ambas
sador', 'social media', 'communication', 'content creator'])),
(30.04, [r'\hr\b', 'human resources', 'recruiter', 'recruiting', 'staffing', 'payroll
', 'timekeeping', 'employee benefits', 'compensation', 'wage', 'salary', 'talent acquisi
tion', 'talent', 'talent professional', 'workforce management', 'human resource manager'
, 'hrm', 'recruitment manager', 'payroll manager', 'personnel', 'job analysis', 'positio
n descriptor', 'talent acquisition coordinator', 'payroll analyst', 'payroll assistant',
'pension administrator', 'title officer', 'technical sourcer', 'sourcer', 'talent manage
ment', 'people operations'])),
(30.05, ['csr', 'diversity', 'inclusion', 'csr', 'corporate responsibility', 'policy
', 'compliance', 'business development', 'bus. dev', 'research and development', 'r&d',
'governance'])),
(30.06, ['notary public', 'legal liaison', 'legal operations', 'legal services', 'le
gal translator', 'legislative aide', 'law', 'lawyer', 'attorney', 'general counsel', 'le
gal counsel', 'legal advisor', 'legal administrator', 'legal researcher', 'legal ninja',
'legal secretary', 'legal translator', 'patent', 'legal', 'notary', 'licensing', 'lawyer
', 'law clerk', 'paralegal', 'attorney', 'counsel'])),
(35, []) # Default category
]

```

```

def categorize_support_words(job_title):
    if pd.isna(job_title): # Check if the value is not NaN
        words_to_match_external = ['software', 'developer', 'software development', 'engineer', 'ux designer', 'programmer', 'programming', 'solution architect', 'solutions architect', 'system architect', 's/w', 'web designer', 'systems', r'\bqa\b', r'\bqa\b', 'qa tester', 'solutions', 'performance', 'principal developer', 'principal engineer', 'sql', 'sap', 'systems', 'test', 'system', 'system administrator', 'quality assurance', r'\bqa\b', 'sales', 'sales', 'registered representative', 'reg. rep', 'representative', 'registered rep', 'sales training', 'salesforce', 'procurement', 'marketing', 'social media', 'advertising', 'promotions', 'communication', 'brand ambassador', 'content creator', r'\bhr\b', r'\bhrm\b', 'human resource', 'recruiter', 'recruiting', 'recruitment', 'staffing', 'payroll', 'timekeeping', 'employee benefits', 'compensation', 'wage', 'salary', 'talent acquisition', 'workforce', 'hrm', 'workforce management', 'human resource manager', 'title officer', 'sourcer', 'people operations', 'personnel', 'position descriptor', 'talent', 'acquisition', 'pension', 'legal counsel', 'notary public', 'legal advisor', 'legal administrator', 'legal researcher', 'legal liaison', 'legal ninja', 'legal operations', 'legal secretary', 'legal services', 'legal translator', 'legislative aide', 'patent', 'legal', 'notary', 'licensing', 'lawyer', 'law', 'notary', 'licensing', 'law clerk', 'internal audit', 'paralegal', 'attorney', 'counsel', 'csr', 'diversity', 'inclusion', 'csr', 'corporate responsibility', 'policy', 'compliance', 'business development', 'bus. dev', 'research and development', 'r&d', 'governance', 'client service', 'client', 'client support', 'support', 'support representative', 'customer service']

    for word in words_to_match_external:
        match_score_1 = fuzz.partial_ratio(word.lower(), job_title.lower())
        # If a better match is found, update the best_match tuple
        if match_score_1 == 100 and len(word) <= len(job_title):
            for category, keywords in rules:
                if word in keywords:
                    return category

    return 0

#functions to categorize job titles

combined_conditions_df = combined_conditions_df.copy() # Create a copy to avoid the SettingWithCopyWarning

combined_conditions_df['support_functions'] = combined_conditions_df['job_title'].apply(categorize_support_words)

#exact matching rules and filter
#new df do just work with cat 35
category_35 = combined_conditions_df

rules_exact = [
    (1, ['partner', 'creator']),
    (4, ['ceo']),
    (5, ['cco', 'cdo', 'cfo', 'cho', 'cio', 'clo', 'cmo', 'coo', 'cpo', 'cso', 'cto', 'cao', 'cceo', 'cos', 'cro']),
    (6, ['president']),
    (7, ['md']),
    (8.1, ['evp']),
    (8.2, ['snrvp', 'svp', 'sr vp', 'sr. vp']),
    (8.3, ['vp', r'\.?p', 'vice president', 'vp,']),
    (8.4, ['avp', 'jr. vp', 'jr vp']),
    (9.3, ['director', 'dir']),
    (9.4, ['jun dir', 'jr dir', 'asst dir' ]),
    (10, ['head', 'lead', 'leader']),
    (11.02, ['pm', 'pmo']),
    (11.3, ['manager', 'manager', 'manger', 'mananger', 'mgr', 'gm']),
    (13, ['lead', 'leader']),

```

```

(14.02, ['pfc']),
(16, ['intern'])
]

def categorize_exact_words(job_title):
    if pd.isna(job_title): # Check if the value is not NaN
        words_to_match = ['partner', 'principal', 'manager', 'president', 'vice president',
                           'svp', 'evp', 'v.p', 'rvp', 'vp', 'avp', 'ceo', 'cco', 'cdo', 'cfo', 'cho', 'cio', 'clo',
                           'cmo', 'coo', 'cpo', 'cso', 'cto', 'ccao', 'cceo', 'cos', 'cro', 'sr mng', 'pm', 'gm', 'md',
                           'director', 'dir', 'vp',]

        for word in words_to_match:
            if job_title.lower() == word: # Check for exact match
                for level, keywords in rules_exact[::-1]:
                    if word in keywords:
                        return level # Return the level based on the matched word

        return 35 # If no match found

category_35 = category_35.copy()
# Apply categorization based on exact words only for rows categorized as 35 by external matching
category_35['exact_level'] = category_35.apply(lambda row: categorize_exact_words(row['job_title']), axis=1)

# Update the original DataFrame with the re-categorized values
combined_conditions_df.loc[category_35.index, 'supervision_level'] = category_35['exact_level']

#partial ratio filter; now matching takes place if the term from the rule also appears in the title + title must be longer than rule
new_df = category_35[category_35['exact_level'] == 35].copy()

def categorize_with_partial_ratio(job_title, rules_match):
    best_match = (35, 0)
    # Iterate through the rules except the default
    for level, keywords in reversed(rules_match[::-1]): #reversed filter - double titles cat to highest
        for keyword in keywords:
            if len(job_title) >= len(keyword):
                #Calculate the match score using partial_ratio
                match_score = fuzz.partial_ratio(keyword.lower(), job_title.lower())
                # If a better match is found, update the best_match tuple
                if match_score == 100:
                    best_match = (level, match_score)

    return best_match[0]

new_df['level_three'] = new_df['job_title'].apply(lambda x: categorize_with_partial_ratio(x, rules))

# Update the original DataFrame with the re-categorized values
combined_conditions_df.loc[new_df.index, 'supervision_level'] = new_df['level_three']

spaltea = 'supervision_level' # Change to your column name
wert_counts = combined_conditions_df[spaltea].value_counts()
print(wert_counts)

combined_conditions_df['senior_supervision_level'] = combined_conditions_df['supervision_level']

# Regular expression pattern for 'vice president'

```



```

pattern_vp = r'vice president|vp\.|\bvp\b|\bsvp\b|v\.p\.|vice president|vice president,|
vp,'

# Use str.contains with regular expression and case sensitivity=False
mask = combined_conditions_df['job_title'].str.contains(pattern_vp, case=False, regex=True)

all_below_vp = combined_conditions_df['senior_supervision_level'].astype(float).isin([6,
35, *range(9, 16)])

# Update based on the mask
combined_conditions_df.loc[mask & all_below_vp, 'senior_supervision_level'] = 8.3

# regular expression pattern for assistant and junior
pattern_jun = r'\bassitant\b|\basst\b| assistant\s+\w+\s+\w+\s+\w+\b|assistant\s+\w+\s+
\w+\b|asst.\s+\w+\s+\w+\b|\bjunior\s+\b|\bjun\b|\bjr\.\b|assistant|junior'
mask_2 = combined_conditions_df['job_title'].str.contains(pattern_jun, case=False, regex
=True)

#vp
all_junvp = combined_conditions_df['senior_supervision_level'].astype(float).between(6,
8.3)
# Update based on the mask
combined_conditions_df.loc[mask_2 & all_junvp, 'supervision_level'] = 8.4 #Transfer resu
lts to supervision level column

all_jundirector = combined_conditions_df['senior_supervision_level'] == 9.3
# Update based on the mask
combined_conditions_df.loc[mask_2 & all_jundirector, 'supervision_level'] = 9.4

condition_spec = combined_conditions_df['senior_supervision_level'].astype(float).betwee
n(14.01, 14.09)
condition_work = combined_conditions_df['senior_supervision_level'].astype(float).betwee
n(15.01, 15.09)

combined_conditions_df.loc[condition_spec & combined_conditions_df['job_title'].str.cont
ains(r'\bsenior\b|\bsr\.\|bsn\.\|bsr\b', case=False), 'supervision_level'] = 14.2

combined_conditions_df.loc[condition_work & combined_conditions_df['job_title'].str.cont
ains(r'\bsenior\b|\bsr\.\|bsn\.\|bsr\b', case=False), 'supervision_level'] = 15.2
# define the rules and filter for geographical differentiation
rule_goe = [
    ('a', ['global']),
    ('b', ['emea', 'usa', 'europe', 'apac', 'asia-pacific', 'asia pacific', 'north ameri
ca', 'mena', 'middle east', 'north africa', 'africa', 'asean', 'cis', 'soviet republic',
'north asia', 'south asia', 'central america', 'the caribbean', 'national']), # global re
gions
    ('c', ['usa', 'canada', 'mexico', 'brazil', 'argentina', 'united kingdom', 'germany'
, 'france', 'italy', 'spain', 'china', 'india', 'japan', 'south korea', 'australia', 'so
uth africa', 'russia', 'turkey', 'saudi arabia', 'united arab emirates', 'nigeria', 'egypt',
'indonesia', 'malaysia', 'singapore', 'thailand', 'vietnam', 'philippines', 'austria
']), #countries
    ('d', ['northwestern', 'northwest', 'north', 'northern', 'northeastern', 'northeast'
, 'easter', 'east', 'central', 'midwest', 'pacific', 'atlantic', 'mountain', 'plains', 'g
ulf', 'regional']),
    ('e', ['branch']),
]

# Function to apply the rules and find the matching letter for each experience title
def find_geo_division(title, rules_g):
    for letter, keywords in rules_g:
        if any(keyword.lower() in title.lower() for keyword in keywords):
            return letter
    return None # Return None if no keywords match

```

```
# Apply the function to the 'job_title' column and create the 'geo_division' column
combined_conditions_df['geo_division'] = combined_conditions_df['job_title'].apply(lambda
a title: find_geo_division(title, rule_goe))

combined_conditions_df.to_csv('/Volumes/X9 Pro/Masterarbeit-extern/CSV files created/financial/financial_salary/financial_salary_categorized_1805.csv', index=False)
```

Appendix 15: Categorize salary data – Code SW

Apply categorization code to the salary data

```
import pandas as pd
import os
from fuzzywuzzy import fuzz
from tqdm import tqdm
tqdm.pandas()

df_salary = pd.read_csv('/Volumes/X9 Pro/Masterarbeit-extern/CSV files created/software/
software_salary/software_salary_all.csv')

#data cleaning
def convert_to_lowercase(dflow):
    # Loop through each column in the DataFrame
    for column in dflow.columns:
        # Check if the column's data type is 'object', typically used for strings
        if dflow[column].dtype == 'object':
            dflow[column] = dflow[column].str.lower()
    return dflow

df_salary = convert_to_lowercase(df_salary)
df_salary['job_title'] = df_salary['job_title'].astype(str)
cond_external = df_salary['job_title'].str.strip().str.contains('nurse|nursing|surgeon|i
nternist|pharmacist|pharmacy|medical|therapist|physician|dental|health|pharmaceutical|pa
ramedic|clinical|patient|navy seal|air force|k-9|pilot|police|police officer|detective|i
nvestigator|lieutenant|captain|fbi agent|sergeant|guard|crime|firefighter|fire|rescue|el
ectrician|mechanic|automotive|auto|parts|car|driver|manufacturing|manufacture|logistics|
assembly|warehouse|electrical|plant|operator|energy|telecommunication|wireless|farmer|ag
riculture|crop|environmental|environment|biologist|geologist|climate|paymaster|city|dipl
omat|fundraising|fundraiser|receptionist|realtor|real estate|property|historian|teacher|
lecturer|teaching|phd|professor|postdoc|postgraduate|undergraduate|student|chef|stylist|
fashion|wardrobe|store|cashier|photographer|event planner|janitor|actor|actress|artist|s
inger|musician|editor|author|writer|proofreader|bookseller|waitress|waiter|barista|barte
nder|pcb|resident|assistant|pastor|freelance|no_info|seeking employment|stay at home|dad
|mom|mother|wife|family leave|retired|retirement|unemployed|not working|retiree|sabbatic
al|summer|plumber|various|personal trainer|coach|nanny|mover|library|librarian|linguist'
, case=False)

combined_conditions_df =df_salary[~(cond_external)]

rules = [
    (1, ['owner', 'founder', r'\bcreator\b', 'created', 'made', 'ownere', r'\bpartner\b'
, 'co-owner', 'business partner', 'founding member']),
    (2, ['board member', 'board of directions', 'chairman']),
    (3, ['leading partner', 'managing partner']),
    (4, ['ceo', 'chief executive', 'cheif executive']),
    (5, [r'\bcco\b', r'\bcdo\b', r'\bcfo\b', r'\bcho\b', r'\bcio\b', r'\bclo\b', r'\bcmo
\b', r'\bccoo\b', r'\bcpo\b', r'\bcso\b', r'\bcto\b', 'ccao', 'cceo', r'\bcos\b', r'\bcro
\b', 'chief officer', 'chief\s+\w+\s+officer', 'chief technology officer', 'chief operat
ing officer', 'chief people officer', 'chief marketing officer', 'chief staff officer',
'chief financial officer', 'chief revenue officer', 'chief product officer', 'chief info
rmation officer', 'chief sales officer', 'chief of staff', 'chief investment officer', '
chief strategy officer']),
    (6, ['president', 'presidente']),
    (7, [r'\bmd\b', 'managing director']),
    (8.1, [r'evp\b', 'executive vp', 'executive vice president', ' executive vice-presid
ent', 'exec vice pres', 'extended board member']),
    (8.2, ['senior vice president', 'sr vice president', 'snrvp', 'svp', 'senior v.p.',
'sr. vice president', r'senior v\bvp', 'sr vp', r'sr. \bvp\b', r'\bsenior\s+\w+\s+vp\b']),
    (8.3, [r'\bvp\b', r'\bv\.?p\b', 'vice president', 'vice-president', 'vice presidente
```

```

', 'vice pres', r'\bv\b\.\?s?p\b', r'\bvp\b']],
(8.4, ['avp', 'assistant vice president', 'junior vice president', r'\bjr. vp\b', r'\bjn\b\s+\vp\b']),
(9, [r'fellow\s+\w+', r'\bfellow\b']),
(10, ['distinguished', 'tier 7', r'7\b', r'\bvii\b']),
(11.2, ['senior principal engineer', 'senior principal\s+\w+']),
(11.3, ['principal engineer', r'\bprincipal\s+\w+', r'6\b', 'tier 6', r'\bvi\b']),
(12.1, ['executive director', 'exec. dir', 'exec director']),
(12.2, ['senior director', 'sr. director', r'\bsenior\s+\w+\s+director\b']),
(12.3, ['director', 'directo', 'diercto', r'dir\b', 'dierctor']),
(12.4, ['junior director', r'\bjun dir\b', 'assistant director', r'\bjunior\s+\w+\s+director\b', r'\bassitant\s+\w+\s+director\b', 'jr. director', r'\bjr dir', r'\basst di r/b', r'asst\s+dir']),
(13, [r'\bhead\s+\w+', r'\bheads\b', r'\bhead\s+\of\w+', 'r\w+\s+\head\s+\w+', 'r\w+\s+\head', 'head of']),
(14.01, ['program manager']),
(14.02, ['project manager', r'\bpmo\b', 'project management officer', r'senior \bpm\b', r'pm\b', r'pm\b manager']), #project managers all
(14.03, ['system administrator', r'qa pm\b', 'pm/se', 'pm windows', 'product manager', 'technical account manager', 'sw manager', 'software manager', 'quality assurance manager', 'qa manager', 'software developer manager', 'software engineering manager', 'software development manager']), #manager software (department managers)
(14.1, ['executive manager']),
(14.2, ['senior manager', 'senior management team', 'sn manager', 'sr. manager', 'sr mng', 'sr manager', r'\bsenior\s+\w+\s+manager\b', r'\bsr.\s+\w+\s+manager\b', r'\bsenior\s+\w+\s+\w+\s+manager\b']),
(14.3, ['manager', 'manager', 'manger', 'mananger', 'mgr', r'gm\b', 'department manager', 'division manager']),
(14.4, ['junior manager', 'assistant manager', r'\bassitant\s+\w+\s+manager\b', 'management assistant', 'coordinator']),
(15.2, ['senior lead', 'senior leader', 'lmts', 'senior lead member of', 'team manager']),
(15.3, ['lead', 'leader', 'lmts', 'lead member of', 'team leader', r'\bv\b', r'\b5\b', 'tier 5']),
(16.2, ['senior staff\s+\w+', 'smts', 'senior member of technical staff']),
(16.3, ['staff\s+\w+', 'mts', 'member of technical staff', 'member technical staff', r'\biv\b', '4', 'tier 4']),
(17, ['supervisor', 'instructor', 'instructor', 'mentor']),
#18.2 senior specialists, 3, iii
(18.01, ['analyst', 'computer system analyst', 'information security analyst', 'system specialist', 'vendor', 'application package specialist', 'system integrator', 'software analyst', 'applications analyst', 'computer systems engineer', 'system engineer', 'systems analyst', 'systems planner', 'programmer analyst', 'systems architect', 'security specialist', 'systems security analyst', 'risk specialist', 'network security analyst', 'sap', 'security officer', 'six sigma', 'security advisor', 'enumerator']), #analyst
(18.02, ['research scientist', 'computational theory scientist', 'computer scientist', 'programming methodology', 'languages researcher', 'data scientist', 'user experience researcher', 'ux researcher', 'researcher', 'scientist']),
(18.03, ['support', 'helpdesk', 'help desk', 'support specialist', 'network support', 'user support', 'client support', 'system installer', 'systems installer', 'services advisor', 'service advisor', 'game advisor', 'product advisor', 'support technician', 'network technician', 'help desk representative', 'help desk specialist', 'help desk analyst', 'help desk technician', 'tech support', 'field engineer', 'pfe', 'pc tech', 'guest advisor', 'site reliability engineer', 'sre', 'support analyst', 'technical advisor', 'client executive', 'client service', 'customer service rep']), #support
(18.04, ['administrator', 'network', 'database', 'computer network architect', 'database architect', 'database administrator', 'system administrator', 'technology architect', 'sql dba', r'dba\b', 'peoplesoft dba', 'solution architect', 'platform engineer', 'platform eng', 'platform specialist', 'platform architect', 'solutions architect', 'computer network engineer', 'network designer', 'network developer', 'network engineer', 'processing planner', 'database administration manager', 'database coordinator', 'database programmer', 'database security administrator', 'data architect', 'data integration specialist', 'data warehousing specialist', 'database developer', 'lan administrator', 'lan systems administrator', 'network administrator', 'network analyst', 'network coordinator', 'network support coordinator', 'network systems administrator', 'network systems coordinator', 'wan systems administrator', 'peoplesoft administrator', 'oracle dba', 'systems adminis

```

trator', 'architect', 'linux', 'middleware administrator', 'oracle app', 'unix admin', 'lotus notes administrator'])), #admin & architects

(18.051, ['computer programmer', 'applications programmer', 'computer language coder', 'coder', 'it programmer', 'mainframe programmer', 'systems programmer', 'unix system', 'programming'])), #programmer

(18.052, ['software developer', 'software development', 'sql processor', 'software engineer', 'software eng', 'software enginer', 'pfe', 'solution specialist professional', 'application package designer', 'software solutions specialist', 'pde', 'product development engineer', 'php developer', 'software architect', 'integration engineer', 'applications developer', 'application developer', 'applications engineer', 'systems engineer', 'systems eng', 'applications developer', 'applications architect', 'applications designer', 'software designer', 'systems software specialist', 'r's/w\b', 'sql', 'sde', 'software design', 'pts', 'software professional', 'ios developer', 'mobile developer', 'solutions engineer', 'oracle app dev', 'sft. eng', 'software eng', 'r'se\b', 'unix', 'software administrator'])), #software

(18.053, ['quality assurance', 'performance', 'software quality assurance', 'quality assurance analyst', 'tester', 'performance quality', 'product safety', 'psr', 'product safety representative', 'security professional', 'qa analyst', 'r'qa\b', 'performance analyst', 'qa engineer', 'performance engineer', 'software quality control specialist', 'software quality engineer', 'software test engineer', 'qu tester', 'performance specialist', 'quality assurance engineer', 'security engineer', 'quality engineer', 'test engineer', 'product safety rep', 'r'psr\b', 'sqa', 'r'qa\b', 'quality control', 'ste', 'sqe', 'qa advisor', 'system test'])), #tester/ quality assurance

(18.054, ['web developer', 'intranet developer', 'web applications developer', 'web architect', 'web content developer', 'ux engineer', 'web dev'])), #web

(18.055, ['web designer', 'interface designer', 'digital designer', 'piping designer', 'painter', 'product designer', 'graphic web designer', 'web content specialist', 'ux designer', 'r'ui\b designer', 'designer', 'design eng', 'user interface', 'ux design', 'ui/ux', 'hardware design', 'graphic designer'])), #design

(18.06, ['consultant', 'it consultant', 'systems consultant', 'software consultant', 'solutions consultant', 'itc'])),

(18.07, ['product administrator', 'project specialist', 'business planner', 'print specialist', 'computer specialist', 'strategist', 'storage solutions specialist', 'processing specialist', 'specialist', 'performer', 'technologist', 'planner', 'product specialist', 'producer', 'process engineer', 'computer console operator', 'computer laboratory technician', 'data center operator', 'paycom specialist', 'process expert', 'thinkpad specialist', 'r'it\b specialist', 'engineer', 'developer', 'programmer', 'pc strategist', 'technician', 'pc tech', 'technician', 'tech info spec', 'devops', 'sme', 'product owner', 'r'dev\b', 'r'dev\b', 'adjuster', 'examiner', 'accountant', 'processor', 'lawyer', 'legal counsel', 'advisor', 'title officer', 'coordinator', '2', 'r'\bii\b', 'tier 2'])), #others

(19.01, ['associate', 'engineer in test', 'sdet', 'machinist', 'in test', '1', 'r'\bi\b', 'tier 1', 'junior', 'jr.', 'r'\bjn\b'])),

(19.02, ['personal assistant', 'executive assistant', 'assistant', 'assistan', 'assistant', 'bpa', 'alumni secretary', 'office secretary', 'personal secretary', 'asst', 'office admin', 'secretary', 'administrative assistant'])),

(19.03, ['clerk', 'employee', 'worker', 'agent', 'paraprofessional', 'mac genius', 'machinest', 'maintenance', 'technical sourcer', 'technical aide'])),

(20, ['trainee', 'trainie', 'research student', 'r'\bintern\b', 'volunteer', 'training program', 'mentee', 'apprentice'])),

(30.01, ['financ', 'financial services', 'credit', 'claims', 'portfolio', 'loan', 'lender', 'tax', 'accounts', 'exchange', 'invoice', 'mortgage', 'banker', 'buyer', 'teller'])),

(30.02, ['sales', 'sales representative', 'registered representative', 'reg. rep', 'representative', 'r'\brep\b', 'registered rep', 'sales manager', 'sales associate manger', 'sales training manager', 'sales consultant', 'salesforce', 'sales associate', 'sales agent', 'sales assistant', 'procurement', 'account executive'])),

(30.03, ['marketing', 'advertising', 'promotions', 'social media', 'communication', 'brand ambassador', 'content creator'])),

(30.04, ['r'hr\b', 'human resource', 'recruiter', 'recruiting', 'staffing', 'payroll', 'timekeeping', 'employee benefits', 'compensation', 'wage', 'salary', 'talent acquisition', 'talent', 'talent professional', 'workforce management', 'human resource manager', 'hrm', 'recruitment manager', 'payroll manager', 'personnel', 'job analysis', 'position descriptor', 'talent acquisition coordinator', 'payroll analyst', 'payroll assistant', 'pension administrator', 'title officer', 'technical sourcer', 'sourcer', 'talent management', 'people operations'])),

(30.05, ['csr', 'diversity', 'inclusion', 'csr', 'corporate responsibility', 'policy

```

', 'compliance', 'business development', 'bus. dev', 'research and development', 'r&d',
'governance' ]),
(30.06, ['notary public', 'legal liaison', 'legal operations', 'legal services', 'le
gal translator', 'legislative aide', 'law', 'lawyer', 'attorney', 'general counsel', 'le
gal counsel', 'legal advisor', 'legal administrator', 'legal researcher', 'legal ninja',
'legal secretary', 'legal translator', 'patent', 'legal', 'notary', 'licensing', 'lawyer
', 'law clerk', 'paralegal', 'attorney', 'counsel', 'internal audit'])),
(35, []) # Default category (here eg pa (personal assistant))
]

def categorize_support_words(job_title):
    if pd.isna(job_title): # Check if the value is not NaN
        words_to_match_support = ['financ', 'banker', 'financial services', 'buyer', 'te
ller', 'adjuster', 'examiner', 'accountant', 'account executive', 'credit', 'claims', 'po
rtfolio', 'loan', 'lender', 'tax', 'accounts', 'exchange', 'invoice', 'mortgage', 'proces
sor', 'sales', 'registered representative', 'reg. rep', 'representative', 'registered re
p', 'salesforce', 'procurement', 'marketing', 'advertising', 'promotions', 'brand ambass
ador', 'social media', 'communication', 'content creator', r'\bhr\b', 'human resources',
'recruiter', 'recruiting', 'staffing', 'payroll', 'timekeeping', 'employee benefits', 'c
ompensation', 'wage', 'salary', 'talent acquisition', 'talent', 'workforce', 'human reso
urce', 'hrm', 'recruitment', 'personnel', 'job analysis', 'position descriptor', 'talent
acquisition coordinator', 'pension', 'title officer', 'technical sourcer', 'sourcer', 'p
eople operations', 'csr', 'diversity', 'inclusion', 'corporate responsibility', 'policy',
'compliance', 'business development', 'bus. dev', 'research and development', 'r&d', '
governance', 'legal', 'notary public', 'legislative aide', 'patent', 'notary', 'licensin
g', 'lawyer', 'law clerk', 'paralegal', 'law', 'attorney', 'counsel', r'\bcco\b', r'\bcf
o\b', r'\bcho\b', r'\bclo\b', r'\bcmo\b', r'\bcoo\b', r'\bcpo\b', r'\bcro\b']

        for word in words_to_match_support:
            match_score_1 = fuzz.partial_ratio(word.lower(), job_title.lower())
            # If a better match is found, update the best match tuple
            if match_score_1 == 100 and len(word) <= len(job_title):
                for category, keywords in rules:
                    if word in keywords:
                        return category

    return 0

combined_conditions_df = combined_conditions_df.copy() # Create a copy to avoid the Set
tingWithCopyWarning

combined_conditions_df['support_functions'] = combined_conditions_df['job_title'].apply(
categorize_support_words)

category_35 = combined_conditions_df#

rules_exact = [
(1, ['partner', 'creator']),
(4, ['ceo']),
(5, ['cco', 'cdo', 'cfo', 'cho', 'cio', 'clo', 'cmo', 'coo', 'cpo', 'cso', 'cto', 'c
cao', 'cceo', 'cos', 'cro']),
(6, ['president']),
(7, ['md']),
(8.1, ['evp']),
(8.2, ['snrvp', 'svp', 'sr vp', 'sr. vp']),
(8.3, ['vp', r'\.?p', 'vice president', 'vp,']),
(8.4, ['avp', 'jr. vp']),
(9, ['fellow']),
(10, ['distinguished', '7', 'tier 7']),
(11.3, ['principal', '6', 'tier 6']),
(12.3, ['director', 'dir']),
(12.4, ['jun dir', 'jr dir', 'asst dir' ]),
(13, ['head']),
(14.01, ['program manager']),

```

```

(14.02, ['pmo', 'pm']),
(14.03, ['qa pm', 'pm/se']),
(14.3, ['manager', 'manager', 'manger', 'mananger', 'mgr', 'gm']),
(15.3, ['lead', 'leader', 'lmts', '5', 'tier 5']),
(16.2, ['smts']),
(16.3, ['mts', 'staff', '4', 'tier 4']),
(18.07, ['2', 'tier 2']),
(19.01, ['1', 'tier 1']),
(20, ['intern']),
(30.01, ['rep']),
(30.04, ['hr']),
(30.05, ['csr', 'r&d', 'bus dev']),
]

#predefine exact match function
def categorize_exact_words(experience_title):
    if pd.isna(experience_title):
        words_to_match = ['partner', 'principal', 'manager', 'mgr', 'president', 'vice pr
esident', 'svp', 'evp', 'lead', 'v.p', 'r.v.p\b', 'rvp', 'vp', 'avp', 'ceo', 'cco', 'cdo',
'cfo', 'cho', 'cio', 'clo', 'cmo', 'coo', 'cpo', 'cso', 'cto', 'ccao', 'cceo', 'cos', 'c
ro', 'head', 'sr mng', 'pmo', 'pm', 'gm', 'md', 'lmts', 'director', 'dir', '7', '6', '5',
'4', '3', '2', '1']

        for word in words_to_match:
            if experience_title.lower() == word: # Check for exact match
                for level, keywords in rules_exact:
                    if word in keywords:
                        return level # Return the Level based on the matched word

        return 35 # If no match found

category_35 = category_35.copy()
# Apply categorization based on exact words only for rows categorized as 35 by external
matching
category_35['exact_level'] = category_35.apply(lambda row: categorize_exact_words(row['j
ob_title']), axis=1)

# Update the original DataFrame with the re-categorized values
combined_conditions_df.loc[category_35.index, 'supervision_level'] = category_35['exact_
level']

new_df = category_35[category_35['exact_level'] == 35].copy()

def categorize_with_partial_ratio(job_title, rules_match):
    best_match = (35, 0)
    # Iterate through the rules except the default
    for level, keywords in reversed(rules_match[:-1]): #reversed filter - categorize str
ings with two titles to highest category
        for keyword in keywords:
            if len(job_title) >= len(keyword):
                #Calculate the match score using partial_ratio
                match_score = fuzz.partial_ratio(keyword.lower(), job_title.lower())
                # If a better match is found, update the best_match tuple
                if match_score == 100:
                    best_match = (level, match_score)

    return best_match[0]

new_df['level_three'] = new_df['job_title'].progress_apply(lambda x: categorize_with_par
tial_ratio(x, rules))

# Update the original DataFrame with the re-categorized values
combined_conditions_df.loc[new_df.index, 'supervision_level'] = new_df['level_three']

```

```

combined_conditions_df['senior_supervision_level'] = combined_conditions_df['supervision_level']

# regular pattern for 'vice president'
pattern_vp = r'vice president\s+\w+|vp\.\vp\b|v\.\p\.|vice president|vice president,|vp,'

mask = combined_conditions_df['job_title'].str.contains(pattern_vp, case=False, regex=True)

all_below_vp = combined_conditions_df['senior_supervision_level'].astype(float).isin([6, 35, *range(9, 19)])

# update based on mask
combined_conditions_df.loc[mask & all_below_vp, 'senior_supervision_level'] = 8.3

combined_conditions_df.loc[combined_conditions_df.index, 'supervision_level'] = combined_conditions_df['senior_supervision_level']

#due to bottom to top filter lowest cat. might be not fully filled
pattern_jun = r'\bassitant\b|\basst\b| assistant\s+\w+\s+\w+\s+\w+\b|assistant\s+\w+\s+\w+\b|asst.\s+\w+\s+\w+\b|\bjunior\s+\b|\bjun\b|\bjr\.\b|assistant|junior' #assistant ja eigentlich bei Level 17 = pa #\bregional\b|\breg\.\b/regional
mask_2 = combined_conditions_df['job_title'].str.contains(pattern_jun, case=False, regex=True)

#vp
all_junvp = combined_conditions_df['senior_supervision_level'].astype(float).isin([6, 8.3])
combined_conditions_df.loc[mask_2 & all_junvp, 'supervision_level'] = 8.4

#director
all_jundirector = combined_conditions_df['senior_supervision_level'] == 12.3
combined_conditions_df.loc[mask_2 & all_jundirector, 'supervision_level'] = 12.4

#senior specialists
condition_15 = combined_conditions_df['senior_supervision_level'].astype(float).between(18.01, 18.07)

combined_conditions_df.loc[condition_15 & combined_conditions_df['job_title'].str.contains(r'\bsenior\b|\bsr\.\b|\bsn\.\b|\bsr\b|\b3\b|\biii\b|\btier 3\b', case=False), 'supervision_level'] = 18.2

#junior dev level i und 1
# check specialist cat
condition_i = combined_conditions_df['senior_supervision_level'].astype(float).isin([18.2]) | combined_conditions_df['senior_supervision_level'].astype(float).between(18.01, 18.07)

combined_conditions_df.loc[condition_i & combined_conditions_df['job_title'].str.contains(r'\bi\b|\b1\b|\bi\.\b|\bjunior\b|\bassociate\b|\bjr\b|\bjn\.\b|\bjn\b', case=False), 'supervision_level'] = 19.01

# Define the rules as provided in the prompt
rule_goe = [
    ('a', ['global']),
    ('b', ['emea', 'usa', 'europe', 'apac', 'asia-pacific', 'asia pacific', 'north america', 'mena', 'middle east', 'north africa', 'africa', 'asean', 'cis', 'soviet republic', 'north asia', 'south asia', 'central america', 'the caribbean', 'national']),
    ('c', ['usa', 'canada', 'mexico', 'brazil', 'argentina', 'united kingdom', 'germany', 'france', 'italy', 'spain', 'china', 'india', 'japan', 'south korea', 'australia', 'south africa', 'russia', 'turkey', 'saudi arabia', 'united arab emirates', 'nigeria', 'egypt', 'indonesia', 'malaysia', 'singapore', 'thailand', 'vietnam', 'philippines', 'austria'])

```



```

    ']), #'market not included bc of misuse
    ('d', ['northwestern', 'northwest', 'north', 'northern', 'northeastern', 'northeast',
    'easter', 'east', 'central', 'midwest', 'pacific', 'atlantic', 'mountain', 'plains', 'gulf', 'regional']),
    ('e', ['branch']),
]

```

Function to apply the rules and find the matching letter for each experience title

```

def find_geo_division(title, rules_g):
    for letter, keywords in rules_g:
        if any(keyword.lower() in title.lower() for keyword in keywords):
            return letter
    return None # Return None if no keywords match

```

Apply the function to the 'job_title' column and create the 'geo_division' column

```

combined_conditions_df['geo_division'] = combined_conditions_df['job_title'].apply(lambda
a title: find_geo_division(title, rule_goe))

```

```

combined_conditions_df.to_csv('/Volumes/X9 Pro/Masterarbeit-extern/CSV files created/software/software_salary/software_salary_categorized_220524.csv', index=False)

```

Appendix 16: Merge number of employees file with salary datafile – Code

Data cleaning steps needed to use the Crunchbase dataset with the number of employees.

The explanations of the steps are indicated by the # symbol.

```
#Load the CrunchBase csv file for the numbers of employees and company names
import pandas as pd

# Define the path to the CSV file
file_path = '/Volumes/X9 Pro/Masterarbeit-extern/20230214_CrunchbaseUS_v3.csv'

# Define the columns of interest
cols_to_use = ['company', 'num_employees']

# Read the CSV file, specify the columns to use, set the correct delimiter and encoding
df_crunchbase_numemp = pd.read_csv(file_path, usecols=cols_to_use, delimiter='*', encoding='iso-8859-1')

# Replace double quotation marks in the 'num_employees' and 'company' columns
df_crunchbase_numemp['num_employees'] = df_crunchbase_numemp['num_employees'].str.replace('"', '')
df_crunchbase_numemp['company'] = df_crunchbase_numemp['company'].str.replace('"', '')

#save the modified DataFrame back to a CSV
df_crunchbase_numemp.to_csv('/Volumes/X9 Pro/Masterarbeit-extern/num_employees_crunchbase.csv', index=False)
```

Figure 29 Code merge number of employees

Screenshot two shows the data cleaning steps to improve the match making before merging the two datasets based on the *company* column.

```
#match the companies in the categorized salary file with those in the num_employees file.

#data cleaning to enhance match finding: replace full stops with spaces and transform everything to lower case
df_salary_cat['company'] = df_salary_cat['company'].str.strip().str.lower().replace(['.'], ' ', regex=True)
df_num_emp['company'] = df_num_emp['company'].str.strip().str.lower().replace(['.'], ' ', regex=True)

# Remove common company suffixes
common_suffixes = ['inc', 'corp', 'corporation', 'co', 'ltd', 'llc', 'limited']
pattern = r'\b(?:{})\b.?'.format('|'.join(common_suffixes))
df_salary_cat['company'] = df_salary_cat['company'].str.replace(pattern, ' ', regex=True)
df_num_emp['company'] = df_num_emp['company'].str.replace(pattern, ' ', regex=True)

#Normalize other potential discrepancies
df_salary_cat['company'] = df_salary_cat['company'].str.replace('international', 'intl', regex=False)
df_num_emp['company'] = df_num_emp['company'].str.replace('international', 'intl', regex=False)

merged_df = pd.merge(df_salary_cat, df_num_emp[['company', 'num_employees']], on='company', how='left')

#drop rows with no information about num employees
merged_df.dropna(subset=['num_employees'], inplace=True)
```

Figure 30 Code data cleaning number of employees

Appendix 17: Evaluate Salary Data – Code

Appendix 17.1: Data-preprocessing and category size calculation for salary data-file – Code

Data preprocessing

```
import pandas as pd
import matplotlib.pyplot as plt

df_salary_cat = pd.read_csv('/Users/kimmelissabar/Documents/Masterabreit Schreiben/Code/financial/salary_fs/financial_salary_categorized_1805.csv')

# Convert 'salary' to a numeric type, coercing errors and removing unwanted characters
df_salary_cat['salary'] = pd.to_numeric(df_salary_cat['salary'].replace(['\$',], '', regex=True), errors='coerce')

# Drop rows where salary could not be converted to numeric
df_salary_cat = df_salary_cat.dropna(subset=['salary'])

# drop rows where yearly salary is below 10.000$
df_salary_cat = df_salary_cat[df_salary_cat['salary'] >= 10000]

# Define the mapping
level_mapping = {
    1: "Owner/ Partner",
    2: "Board Member",
    3: "Managing Partner",
    4: "CEO",
    5: "C-Suite",
    6: "President",
    7: "Managing Director",
    8: "Vice President",
    8.1: "Executive VP",
    8.2: "Senior VP",
    8.3: "VP",
    8.4: "Junior/ Assistant VP",
    9: "Director",
    9.1: "Executive Director",
    9.2: "Senior Director",
    9.3: "Director",
    9.4: "Junior/ Assistant Director",
    10: "Head",
    11: "Manager",
    11.01: "Program Manager",
    11.02: "Project Manager",
    11.03: "Software Manager",
    11.1: "Executive Manager",
    11.2: "Senior Manager",
    11.3: "Manager",
    11.4: "Junior/ Assistant Manager",
    12: "Supervisor",
    13: "Lead",
    14.2: "Senior Specialist",
    14: "Specialist",
    15.2: "Senior Workers",
    15: "Workers",
    16: "Learners",
    30.01: "Support Software",
    30.02: "Support Sales",
    30.03: "Support Marketing",
    30.04: "Support Human Resources",
    30.05: "Support Business Development",
```

```

    30.06: "Support Legal",
    35: "Other"
}

df_salary_cat['level_name'] = df_salary_cat['supervision_level'].map(level_mapping)

#transform floats to integer to have the option for a calculation without seniority and horizontal differentiation

# Convert 'supervision_level' from float to integer and create a new column 'main_categories'
df_salary_cat['main_categories'] = df_salary_cat['supervision_level'].astype('float').astype('int')

unique_companies_count = df_salary_cat['company'].nunique()
print("Number of unique companies:", unique_companies_count)

```

Calculate the sizes of the categories and observations

```

#count the size of the main cat (r + %)
#one way to confirm the validity of the code by checking the (similar) distribution of a different database

pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)

if 'main_categories' in df_salary_cat.columns:
    # Count the number of entries for each unique number in the 'supervision_level' column
    count_per_number = df_salary_cat['main_categories'].value_counts()
    #replace 'main_categories' with 'supervision_level' calculate the category sizes respectively

    # Calculate the percentage overview
    percentage_overview = count_per_number / len(df_salary_cat) * 100

    # Print the count and percentage overview
    print("Count of each unique number:")
    print(count_per_number)
    print("\nPercentage overview:")
    print(percentage_overview)
else:
    print("Error: 'main_categories' column not found in the dataframe.")

# Properly filter the DataFrame where 'support_functions' is 0.00 and 'column_name' is not 35
df_filtered = df_salary_cat[(df_salary_cat['support_functions'] == 0.00) & (df_salary_cat['main_categories'] != 35)]

# Count the number of observations in the filtered DataFrame
count_observations = len(df_filtered)

print(f"Number of observations where 'support_functions' is 0.00 and 'column_name' is not 35: {count_observations}")

Number of observations where 'support_functions' is 0.00 and 'column_name' is not 35: 195821

```

Appendix 17.2: Statistics calculation for salary data-file – Code

Statistics calculation

Calculate statistics and plot

```

# Group by 'supervision level' and calculate the min, max, and mean salaries
stats = df_salary_cat.groupby('supervision_level')['salary'].agg(['min', 'max', 'mean'])

# Rename the columns for clarity
stats.columns = ['Minimum Salary', 'Maximum Salary', 'Average Salary']

# Save the statistics to a CSV file
stats.to_csv('/Users/kimmelissabar/Documents/Masterabreit Schreiben/Code/financial/salary_fsy_statistics.csv')

# Display the statistics in a table format
print("Salary Statistics by Supervision Level:")
print(stats)

# Plotting
ax = stats.plot(kind='bar', figsize=(10, 5), title='Salary Statistics by Supervision Level')
plt.ylabel('Salary')
plt.xlabel('Supervision Level')
plt.xticks(rotation=45)
plt.tight_layout()
plt.legend(title='Statistics')

# Save the plot as a PNG file
plt.savefig('/Users/kimmelissabar/Documents/Masterabreit Schreiben/Code/financial/salary_fs/salary_statistics_chart.png')

# Show the plot
plt.show()

```

Calculate boxplot for main categories

```

#standardized
#without support functions

# Filter the DataFrame to include only rows where 'support_function' is 0.00
df_salary_cat = df_salary_cat[df_salary_cat['support_functions'] == 0.00]

# Standardize the salary data
mean_salary = df_salary_cat['salary'].mean()
std_salary = df_salary_cat['salary'].std()
df_salary_cat['standardized_salary'] = (df_salary_cat['salary'] - mean_salary) / std_salary

# Group by 'main categories' to prepare for statistics
grouped = df_salary_cat.groupby('main_categories')['standardized_salary']

# Prepare a DataFrame to hold the statistics
stats_df = pd.DataFrame({
    'Min': grouped.min(),
    'Q1': grouped.quantile(0.25),
    'Median': grouped.median(),
    'Q3': grouped.quantile(0.75),
    'Max': grouped.max(),
    'Count': grouped.count() # Count of observations per main categories
})

# Calculate Interquartile Range (IQR)
stats_df['IQR'] = stats_df['Q3'] - stats_df['Q1']

# Display the statistics in a table format
print("Detailed Boxplot Statistics by Main Categories (Support Function = 0.00):")
print(stats_df)

```

```

# Save the statistics to a CSV file
stats_df.to_csv('/Users/kimmelissabar/Documents/Masterabreit Schreiben/Code/software/salary_sw/detailed_boxplot_statistics_standardized_Main_no_support.csv')

# Filtering extreme values for better visualization (if necessary)
filtered_df_salary_cat = df_salary_cat[(df_salary_cat['standardized_salary'] >= -3) & (df_salary_cat['standardized_salary'] <= 3)]

# Plotting a boxplot
plt.figure(figsize=(20, 10)) # Increase figure size for better readability
boxplot = filtered_df_salary_cat.boxplot(column='standardized_salary', by='main_categories', grid=False, fontsize=10, patch_artist=True, showfliers=False)
plt.title('Boxplot of Standardized Salaries by Main Categories (Support Function = 0.00)', fontsize=16)
plt.suptitle('') # Suppress the default title to clean up the plot
plt.ylabel('Standardized Salary', fontsize=14)
plt.xlabel('Main Categories', fontsize=14)
plt.xticks(rotation=90, fontsize=10) # Rotate to 90 degrees to better manage space

# Adjust Layout
plt.tight_layout()
plt.subplots_adjust(bottom=0.3) # Adjust bottom to give more space to x-axis labels

# Save the boxplot as a PNG file
plt.savefig('/Users/kimmelissabar/Documents/Masterabreit Schreiben/Code/software/salary_sw/salary_boxplot_std_main_no_support.png')#
# Show the plot
plt.show()

```

Calculate boxplot with seniority and horizontal diff

```

#w/o support
#standardized
#include x-axis label with cat. names instead of numbers

# Define the Level mapping e.g. SW industry
level_mapping = {
    1: "Owner/ Partner",
    2: "Board Member",
    3: "Managing Partner",
    4: "CEO",
    5: "C-Suite",
    6: "President",
    7: "Managing Director",
    8: "Vice President",
    8.1: "Executive VP",
    8.2: "Senior VP",
    8.3: "VP",
    8.4: "Junior/ Assistant VP",
    9: "Fellow",
    10: "Distinguished",
    11: "Principal",
    11.2: "Senior Principal",
    11.3: "Principal",
    12: "Director",
    12.1: "Executive Director",
    12.2: "Senior Director",
    12.3: "Director",
    12.4: "Junior/ Assistant Director",
    13: "Head",
    14: "Manager",
    14.01: "Program Manager",

```

```

14.02: "Project Manager",
14.03: "Software Manager",
14.1: "Executive Manager",
14.2: "Senior Manager",
14.3: "Manager",
14.4: "Junior/ Assistant Manager",
15: "Lead",
15.2: "Senior Lead",
15.3: "Lead",
16: "Staff",
16.2: "Senior Staff",
16.3: "Staff",
17: "Supervisor",
18: "Specialist",
#18.2: "Senior Specialist",
#18.01: "Specialist",
19: "Workers",
#19.2: "Senior Workers",
#19.01: "Workers",
20: "Learners",
30: "Support",
30.01: "Support Finance",
30.02: "Support Sales",
30.03: "Support Marketing",
30.04: "Support Human Resources",
30.05: "Support Business Development",
30.06: "Support Legal",
35: "Other"
}

# Filter the DataFrame to include only rows where 'support_functions' is 0.00
df_salary_cat = df_salary_cat[df_salary_cat['support_functions'] == 0.00]

# Standardize the salary data
mean_salary = df_salary_cat['salary'].mean()
std_salary = df_salary_cat['salary'].std()
df_salary_cat['standardized_salary'] = (df_salary_cat['salary'] - mean_salary) / std_salary

# Group by 'main_categories' to prepare for statistics
grouped = df_salary_cat.groupby('main_categories')['standardized_salary']

# Prepare a DataFrame to hold the statistics
stats_df = pd.DataFrame({
    'Min': grouped.min(),
    'Q1': grouped.quantile(0.25),
    'Median': grouped.median(),
    'Q3': grouped.quantile(0.75),
    'Max': grouped.max(),
    'Count': grouped.count() # Count of observations per main_categories
})

# Calculate Interquartile Range (IQR)
stats_df['IQR'] = stats_df['Q3'] - stats_df['Q1']

# Display the statistics in a table format
print("Boxplot by Main Categories (Support Function = 0.00):")
print(stats_df)

# Save the statistics to a CSV file
stats_df.to_csv('/Users/kimmelissabar/Documents/Masterarbeit Schreiben/Code/software/salary_sw/detailed_boxplot_statistics_std_nosupport_name.csv')

# Filtering extreme values for better visualization (if necessary)

```

```

filtered_df_salary_cat = df_salary_cat[(df_salary_cat['standardized_salary'] >= -3) & (df_salary_cat['standardized_salary'] <= 6)]

# Plotting a boxplot
plt.figure(figsize=(60, 30)) # Increase figure size for better readability
boxplot = filtered_df_salary_cat.boxplot(column='standardized_salary', by='main_categories', grid=False, fontsize=12, patch_artist=True, showfliers=False)

# Update x-axis labels to use level_name from level_mapping
plt.xticks(ticks=range(1, len(boxplot.get_xticklabels()) + 1), labels=[level_mapping.get(float(cat.get_text()), cat.get_text()) for cat in boxplot.get_xticklabels()], rotation=90, fontsize=10)

plt.title('')
plt.suptitle('') # Suppress the default title to clean up the plot
plt.ylabel('Standardized Salary', fontsize=12)
plt.xlabel('Category', fontsize=12)

# Adjust Layout
plt.tight_layout()

# Save the plot to a file
plt.savefig('/Users/kimmelissabar/Documents/Masterabreit Schreiben/Code/software/salary_sw/boxplot_std_salaries_nosupport_name.png')

# Show the plot
plt.show()

```

Appendix 17.3: Salary data including ‘number of employees’ variable – Code

```

#Load the modified crunchbase file with the number of employees
df_num_emp = pd.read_csv('/Users/kimmelissabar/Documents/Masterabreit Schreiben/Code/num_employees_crunchbase.csv')

#exclude all rows that do not have an explicit entry in the column num_employees
df_num_emp = df_num_emp[df_num_emp['num_employees'] != '\\N']

```

Match the companies in the categorized salary file with those in the num_employees file

```

#data cleaning to enhance match finding: replace full stops with spaces and transform everything to lower case
df_salary_cat['company'] = df_salary_cat['company'].str.strip().str.lower().replace(['.,'], '', regex=True)
df_num_emp['company'] = df_num_emp['company'].str.strip().str.lower().replace(['.,'], '', regex=True)

# Remove common company suffixes
common_suffixes = ['inc', 'corp', 'corporation', 'co', 'ltd', 'llc', 'limited']
pattern = r'\b(?:{})\b\.'.format('|'.join(common_suffixes))
df_salary_cat['company'] = df_salary_cat['company'].str.replace(pattern, '', regex=True)
df_num_emp['company'] = df_num_emp['company'].str.replace(pattern, '', regex=True)

#Normalize other potential discrepancies
df_salary_cat['company'] = df_salary_cat['company'].str.replace('international', 'intl', regex=False)
df_num_emp['company'] = df_num_emp['company'].str.replace('international', 'intl', regex=False)

merged_df = pd.merge(df_salary_cat, df_num_emp[['company', 'num_employees']], on='company', how='left')

```



```
#drop rows with no information about num employees
merged_df.dropna(subset=['num_employees'], inplace=True)
```

Calculations with number of employees

Creating a pivot table with the mean salary per employee number group and cat

```
merged_df['num_employees'] = merged_df['num_employees'].str.replace(r'["\\N]', '', regex=True)
merged_df = merged_df[merged_df['num_employees'] != ''] # Excludes all entries that became empty after replacement
```

```
# Drop rows where 'salary' or 'num_employees' could not be cleaned or converted
merged_df.dropna(subset=['salary', 'num_employees'], inplace=True)
```

```
# Create the pivot table for the main categories
pivot_table = merged_df.pivot_table(
    values='salary',
    index='main_categories', #exchange with 'supervision_level' for other option
    columns='num_employees',
    aggfunc='mean'
)
```

```
# Save the pivot table to an Excel file
output_excel_path = '/Users/kimmelissabar/Documents/Masterarbeit Schreiben/Code/software/meansalary_numemp_main.xlsx' #adapt name if index is changes
pivot_table.to_excel(output_excel_path, sheet_name='Mean Salaries')
```

```
# Print a message to confirm the file has been saved
print(f"Pivot table saved to {output_excel_path}")
```

```
# Create the pivot table for the main categories
#standardized
```

```
#import pandas as pd
from scipy.stats import zscore
```

```
# Standardize the 'salary' data
merged_df['salary_z'] = zscore(merged_df['salary'])
```

```
# Create the pivot table using the standardized 'salary_z' instead of 'salary'
pivot_table = merged_df.pivot_table(
    values='salary_z', # use the standardized salary values
    index='main_categories',
    columns='num_employees',
    aggfunc='mean'
)
```

```
# Save the pivot table to an Excel file
output_excel_path = '/Users/kimmelissabar/Documents/Masterarbeit Schreiben/Code/software/meansalary_numemp_main_stand.xlsx'
pivot_table.to_excel(output_excel_path, sheet_name='Mean Salaries')
```

```
# Print a message to confirm the file has been saved
print(f"Pivot table saved to {output_excel_path}")
```

Appendix 18: Results – Code

Read file

```
import pandas as pd
import re

df = pd.read_csv("/Users/kimmelissabar/Documents/Masterabreit Schreiben/Code/financial/o
utputs/financial_all_new3105.csv") #define path to right file for the selected industry

#drop the column supervision level
df = df.drop(columns='senior_supervision_level')

#count the number of unique companies and job titles
unique_count_companies = df['experience_company'].nunique()
unique_count_titles = df['experience_title'].nunique()

# Print the count
print("Unique count in 'company' column:", unique_count_companies)
print("Unique count in 'job titles' column:", unique_count_titles)
```

Appendix 18.1: Differentiation features – Code

Transformations to look at various differentiation features

```
#transform floats to integer to have the option for a calculation without seniority and
horizontal differentiation
# = create main categories

# Convert 'supervision_level' from float to integer and create a new column 'main_catego
ries'
df['main_categories'] = df['supervision_level'].astype('float').astype('int')

#transform the categories to have an option to ignore horizontal differentiation

#Function to check if the value matches the pattern number.number (e.g. 15.06)
def should_transform(x):
    return bool(re.match(r'^\d+\.\d$', str(x)))

# Apply transformation conditionally
df['seniority_diff'] = df['supervision_level'].apply(
    lambda x: int(x) if should_transform(x) else x
)

# Function to check if the value matches the pattern number.number (e.g. 15.2)
def should_transform(x):
    return bool(re.match(r'^\d+\.\d(h1bdata.info)$', str(x)))

# Apply transformation conditionally
df['horizontal differentiation'] = df['supervision_level'].apply(
    lambda x: int(x) if should_transform(x) else x
)

df
```

Appendix 18.2: Category sizes – Code

Calculate category sizes

```
#count the identified support functions observations
count_non_zero_support = df[df['support_functions'] != 0.00].shape[0]

print("Number of non-zero entries in 'support' column:", count_non_zero_support)

#calculate size of ALL categories + % value
# Set display options to show all rows and columns
pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)

if 'supervision_level' in df.columns:
    # Count the number of entries for each unique number in the 'supervision_level' column
    count_per_number = df['supervision_level'].value_counts()
    #replace supervision_level with 'main_categories', 'seniority_diff' or 'horizontal differentiation' to calculate the category sizes respectively

    # Calculate the percentage overview
    percentage_overview = count_per_number / len(df) * 100

    # Print the count and percentage overview
    print("Count of each unique number:")
    print(count_per_number)
    print("\nPercentage overview:")
    print(percentage_overview)
else:
    print("Error: 'supervision_level' column not found in the dataframe.")
```

Appendix 18.3: Measuring hierarchical depth – Code

Calculate hierarchical depth

count levels per company with seniority and hierarchical differentiation

```
unique_counts = df.groupby('experience_company')['supervision_level'].apply(pd.Series.nunique)

#create a boxplot
import matplotlib.pyplot as plt
import seaborn as sns
# Reset index if 'unique_counts' is a Series; this step ensures it's a DataFrame
unique_counts_df = unique_counts.reset_index(name='Unique Counts')

# Create a boxplot
sns.boxplot(x='Unique Counts', data=unique_counts_df)

# Enhance the plot with titles and labels
plt.title('Distribution of Unique Supervision Levels by Company')
plt.xlabel('Number of Unique Supervision Levels')
plt.ylabel('Companies') # This label is somewhat generic as the y-axis actually doesn't have a meaningful label in a boxplot

# Save the plot to a PNG file
plt.savefig('unique_supervision_levels.png')
```

```
# Show the plot
plt.show()
```

count levels per company main: without seniority and without hierarchical differentiation

exclude companies with only one hierarchical level

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import os

# Ensure the output directory exists
output_dir = '/Users/kimmelissabar/Documents/Masterabreit Schreiben/Code/financial' #adapt directory
os.makedirs(output_dir, exist_ok=True)

#count levels per company without seniority and without hierarchical differentiation
unique_counts_main = df.groupby('experience_company')['main_categories'].apply(pd.Series
.nunique)
#to examine other variables that indicate differentiation exchange var 'main_categories'
#attention: also rename the other 'main_categories' indicators

# Reset index if 'unique_counts_seniority' is a Series; this step ensures it's a DataFrame
unique_counts_df_2 = unique_counts_main.reset_index(name='Unique Counts Main')

# Exclude companies with only one level
filtered_unique_counts_df_2 = unique_counts_df_2[unique_counts_df_2['Unique Counts Main']
> 1]

# Calculate descriptive statistics
statistics_df = filtered_unique_counts_df_2['Unique Counts Main'].describe(percentiles=[
.25, .5, .75])
statistics_df = statistics_df.to_frame(name='Value')
statistics_df['iqr'] = statistics_df.loc['75%', 'Value'] - statistics_df.loc['25%', 'Value'] # Calculate IQR
lower_bound = statistics_df.loc['25%', 'Value'] - 1.5 * statistics_df['iqr'].values[0]
# Ensure scalar by accessing the first element
upper_bound = statistics_df.loc['75%', 'Value'] + 1.5 * statistics_df['iqr'].values[0]
# Ensure scalar by accessing the first element

# Identify outliers
outliers = filtered_unique_counts_df_2[
    (filtered_unique_counts_df_2['Unique Counts Main'] < lower_bound) |
    (filtered_unique_counts_df_2['Unique Counts Main'] > upper_bound)]

# Create a boxplot
sns.boxplot(x='Unique Counts Main', data=filtered_unique_counts_df_2)
plt.title('') #Distribution of Unique Supervision Levels by Company (Excluding Single Level)
plt.xlabel('Number of Unique Supervision Levels')
plt.ylabel('Companies')

# Save the plot
plot_path = os.path.join(output_dir, 'filtered_unique_supervision_levels_main_fs.png') #
rename when looking at other differentiation aspects
plt.savefig(plot_path)
plt.close() # Close the plot explicitly after saving

# Save the original counts and statistics to an Excel file
excel_path = os.path.join(output_dir, 'filtered_unique_counts_main_analysis_fs.xlsx') #adapt the file name, to not override the result when looking at other differentiation variables
try:
```

```

        with pd.ExcelWriter(excel_path) as writer:
            filtered_unique_counts_df_2.to_excel(writer, index=False, sheet_name='Unique Counts')
            statistics_df.to_excel(writer, index=True, sheet_name='Descriptive Statistics')
            outliers.to_excel(writer, index=False, sheet_name='Outliers')
            print(f"Data and statistics saved to Excel at {excel_path}.")
    except Exception as e:
        print(f"Failed to save Excel file: {e}")

print(f"Plot saved to {plot_path}.")

```

include geographical differentiation

```

# Set display options to show all rows and columns
pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)

if 'supervision_level' in df.columns:
    # Count the number of entries for each unique number in the 'supervision_level' column
    count_per_number = df['supervision_level'].value_counts()

    # Calculate the percentage overview
    percentage_overview = count_per_number / len(df) * 100

    # Print the count and percentage overview
    print("Count of each unique number:")
    print(count_per_number)
    print("\nPercentage overview:")
    print(percentage_overview)
else:
    print("Error: 'supervision_level' column not found in the dataframe.")

# Calculate the number of unique combinations of 'supervision_level' and 'geo_division'
unique_combinations = df.drop_duplicates(subset=['supervision_level', 'geo_division'])

# Count the unique combinations
num_unique_combinations = len(unique_combinations)

# Print the number of unique combinations
print("\nNumber of unique combinations of 'supervision_level' and 'geo_division':", num_unique_combinations)

# Calculate the size of all categories and their percentage
category_counts = df.groupby(['supervision_level', 'geo_division']).size().reset_index(name='count')
category_counts['percentage'] = category_counts['count'] / len(df) * 100

# Print the category counts and their percentage
print("\nSize of all categories and their percentage:")
print(category_counts)

# Sort the entire DataFrame by 'geo_division' and 'supervision_level'
df_sorted = df.sort_values(by=['geo_division', 'supervision_level'], ascending=[True, True])

# Print the sorted DataFrame
#print("\nSorted DataFrame by 'geo_division' and 'supervision_level':")
#print(df_sorted)

```

look at a specific company with titles sorted numerical in combination with alphabetical (geo)

```

# Replace 'company' with the actual string you're searching for in the 'experience_company' column.
company_geoinfo = 'company'

# Filter the DataFrame for non-NaN entries in 'experience_company' that contain the specific string
df_filtered = df[df['experience_company'].notna() & df['experience_company'].str.contains(company_geoinfo, na=False)]

# The DataFrame will be firstly sorted by 'geo_division' and then by 'supervision_level' within each 'geo_division' group
df_sorted = df_filtered.sort_values(by=['geo_division', 'supervision_level'], ascending=[True, True])

# This sorted DataFrame can then be used for further processing or saved to a file

```

Appendix 18.4: Include variable “number of employees” – Code

Include number of employees as variable

```

#merge df num_employees and output file result to create a pivot table with the num employees and num hierarchical levels
#main categories
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import os

# Define file paths for input CSVs
sw_cat_file = '/Users/kimmelissabar/Documents/Masterabreit Schreiben/Code/financial/outputs/financial_all_new3105.csv'
num_emp_file = '/Users/kimmelissabar/Documents/Masterabreit Schreiben/Code/num_employees_crunchbase.csv'

# Load CSV files
df_fs_cat = pd.read_csv(sw_cat_file)
df_num_emp = pd.read_csv(num_emp_file)
#data cleaning to enhance match finding: replace full stops with spaces and transform everything to lower case
df_fs_cat['company'] = df_fs_cat['experience_company'].str.strip().str.lower().replace(['.'], ' ', regex=True)
df_num_emp['company'] = df_num_emp['company'].str.strip().str.lower().replace(['.'], ' ', regex=True)

df_num_emp.dropna(subset=['num_employees'], inplace=True)

# Merge dataframes
merged_df = pd.merge(df_fs_cat, df_num_emp[['company', 'num_employees']], left_on='experience_company', right_on='company', how='left')
print(f"Number of rows before merge: {df_fs_cat.shape[0]}")
print(f"Number of rows after merge: {merged_df.shape[0]}")

# Identify duplicated rows based on all columns
duplicates = merged_df[merged_df.duplicated()]

# Drop duplicated rows based on all columns
merged_df = merged_df.drop_duplicates()

# Verify the number of rows after dropping duplicates

```

```

print(f"Number of rows after dropping duplicates: {merged_df.shape[0]}")

merged_df.dropna(subset=['num_employees'], inplace=True)

create main category

#single observations
merged_df['main_categories'] = merged_df['supervision_level'].astype('float').astype('int')
print('count 0, Filtered unique df hat ',merged_df.shape[0], 'Werte')

# Count Levels per company main: without seniority and without hierarchical differentiation
#from here for companies
unique_counts_main = merged_df.groupby('experience_company')['main_categories'].apply(pd.Series.nunique)
print('count 1',unique_counts_main.shape[0])

# Reset index to ensure it's a DataFrame
unique_counts_df_2 = unique_counts_main.reset_index(name='Unique Counts Main')
print('count', unique_counts_df_2.shape[0])

# Exclude companies with only one Level
filtered_unique_counts_df_2 = unique_counts_df_2[unique_counts_df_2['Unique Counts Main'] > 1]

# Merge the filtered unique counts with the num_employees data -> to ensure tht every organization is included (only) once
filtered_merged_df = pd.merge(filtered_unique_counts_df_2, merged_df[['experience_company', 'num_employees']], on='experience_company', how='left').drop_duplicates()

hierarchical depth calculation

# Verify the size of the data frame
print(f"The original number of observations is: {filtered_merged_df.shape[0]}")

# Calculate descriptive statistics per num_employees category
pivot_table = filtered_merged_df.groupby('num_employees')['Unique Counts Main'].agg(
    q1=lambda x: x.quantile(0.25),
    median='median',
    q3=lambda x: x.quantile(0.75),
    mean='mean'
).reset_index()

# Ensure the output directory exists
output_dir = '/Users/kimmelissabar/Documents/Masterarbeit Schreiben/Code/financial'
os.makedirs(output_dir, exist_ok=True)

# Save the pivot table to Excel
pivot_table_excel_path = os.path.join(output_dir, 'pivot_table_analysis_numemp_main_fs.xlsx')
try:
    with pd.ExcelWriter(pivot_table_excel_path) as writer:
        pivot_table.to_excel(writer, sheet_name='Pivot Table', index=False)
        print(f"Pivot table saved to Excel at {pivot_table_excel_path}.")
except Exception as e:
    print(f"Failed to save pivot table Excel file: {e}")

# Create a boxplot for Unique Counts Main by num_employees categories
plt.figure(figsize=(10, 6))
sns.boxplot(x='num_employees', y='Unique Counts Main', data=filtered_merged_df)
plt.title('Distribution of Unique Supervision Levels by Number of Employees')
plt.xlabel('Number of Employees Categories')
plt.ylabel('Number of Unique Supervision Levels')

```

```
plt.xticks(rotation=90)

# Save the plot
plot_path = os.path.join(output_dir, 'boxplot_supervision_levels_by_num_employees_fs.png')
plt.savefig(plot_path)
plt.close() # Close the plot explicitly after saving

print(f"Boxplot saved to {plot_path}.")

# Count the number of observations included in the analysis
num_observations = filtered_merged_df.shape[0]
print(f"The number of observations included in the analysis is: {num_observations}")
```

Appendix 18.5: Number of organizations per ‘num_employees’ group

```
# Display how many organizations belong the every category based on number of employees
num_employees_counts = filtered_merged_df['num_employees'].value_counts().reset_index()
num_employees_counts.columns = ['num_employees', 'count']

print("Distinct values in the 'num_employees' column and their counts:")
print(num_employees_counts)

# Ensure all values are printed
with pd.option_context('display.max_rows', None, 'display.max_columns', None):
    print(num_employees_counts)
```

The results showed that the number of organizations in the dataset matched with the number of employees are as follows:

num employees	num organization
1 - 10	12140
11 - 050	14782
51 - 100	2593
101 - 250	3905
251 - 500	1422
501 - 1000	798
1001 - 5000	645
5001 - 10000	151
10001 - max	196

Table 6 Number of organizations per number of employees – FS industry

num employees	num organizations
1 - 10	12196
11 - 50	17306
51 - 100	4381
101 - 250	3963
251 - 500	1514
501 - 01000	957
1001 - 5000	654
5001 - 10000	220
10001_max	262

Table 7 Number of organization per number of employees – SW industry