universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master‘s Thesis

“Data Exploitation for Public Transit Planning“

verfasst von / submitted by

Beauty Burman

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2024 / Vienna 2024

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 914
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Internationale Betriebswirtschaft
Betreut von / Supervisor:	Univ.- Prof. Dr. Jan Fabian Ehmke

Table of Contents

Acknowledgement

Firstly, I would like to express my gratitude to my supervisor, Univ. - Prof. Dr. Jan Fabian Ehmke, Vice Dean – Faculty of Business, Economics & Statistics, who gave me the opportunity to write my master's thesis under his supervision. His unlimited support, encouragement and advice guided me throughout the whole process and beyond. He provided me with valuable feedback and suggestions during all stages of my work and his quick responses kept me pushing towards my goal. Also, truly thankful for understanding my situation and giving me as much time i needed to complete my thesis. Thank you for motivating me from the beginning till the end.

Secondly, I am grateful to my family, my husband, daughter, mama, papa, aunt for helping me directly and indirectly during these years of Master studies, and supported me, encouraged me every single day. You all are my strength and inspiration. Thank you, and I love you all.

Last but not least, my appreciation goes to my colleagues and friends, Iman, Olga, Julia. You gave me the idea - how to write a thesis proposal by sharing your thesis proposal with me. You also gave me some valuable tips regarding how to proceed with thesis work. Likewise, both Julia and Ms. Gislind Stefan helped me further by sharing their thesis work with me , which gave me idea about what to expect while writing my thesis, and took their time respnding to my queries. Not to forget Sandra, who helped me with German translation of thesis abstract.

To all of you, my deepest thanks!

Abstract

As we are living in a technologically advanced era, getting real-time information takes barely a few seconds. Advancements in communication and information technology have made our lives very easy and more comfortable. This advancement in technology has spread in every sector including public transport. With the help of smartphones and Apps, accessing real-time data has never been so easy. New information sources from public transport providers and wearable devices provide opportunities for novel applications in urban transit planning, which improve citizens‘ quality of life.

The aim of this thesis is to obtain an overview of how far technology has been exploited in order to increase the efficiency and reliability of public transit. This is accomplished by posing the following research questions: "Where does data come from and in which area of transit planning has it been used extensively? Up to what level does data has been exploited?"

A systematic literature review was performed in order to answer these questions. The latest scientifc papers on data driven public transit planning, operation, optimization, disturbance management, and transit policy were reviewed.

Ultimately, this thesis helps in leading the stakeholders such as researchers, transit planners and policy makers towards finding a common goal which will be beneficial for the overall improvement of the trasnport sector.

Keywords: Real-time data, transit planning, public transport, reliability, efficiency, exploitation, operation, optimization, disturbance management, transit policy, transport sector.

Kurzfassung

Da wir in einer technologisch fortgeschrittenen Ära leben, dauert der Erhalt von Echtzeitinformationen nur wenige Sekunden. Der Fortschritt in der Kommunikation und Informationstechnologie hat unser Leben sehr einfach und komfortabler gemacht. Dieser technologische Fortschritt hat sich in allen Sektoren, einschließlich des öffentlichen Verkehrs, ausgebreitet. Mit Hilfe von Smartphones und Apps war der Zugriff auf Echtzeitdaten noch nie so einfach. Neue Informationsquellen von öffentlichen Verkehrsmitteln und mit tragbaren Geräten , werden Möglichkeiten für neuartige Anwendungen in der urbanistischen Verkehrsplanung geboten, die die Lebensqualität der Bürger verbessern.

Ziel dieser Arbeit ist es, einen Überblick darüber zu geben, inwieweit Technologie genutzt wurde, um die Effizienz und Zuverlässigkeit des öffentlichen Nahverkehrs zu erhöhen. Dies wird durch das Stellen folgender Forschungsfragen unterstützt: "Woher stammen die Daten und in welchem Bereich der Verkehrsplanung wurden sie umfassend genutzt? Bis zu welchem Niveau wurden Daten genutzt?“

Um diese Fragen zu beantworten, wurde eine systematische Literaturrecherche durchgeführt. Neueste wissenschaftliche Arbeiten zu datengesteuerter Planung, Betrieb, Optimierung, Störungsmanagement und Transitpolitik wurden überprüft.

Letztendlich trägt diese Arbeit dazu bei, die Interessengruppen bzw. Stakeholder wir Forscher, Verkehrsplaner und politische Entscheidungsträger, dazu zu bringen, ein gemeinsames Ziel zu finden, das für die allgemeine Verbesserung des Verkehrssektors von Vorteil sein wird.

List of Figures

List of Tables

List of Abbreviations

AFC – Automated Fare Collection

APC – Automated Passenger Counter

AVL – Automatic Vehicle Location

ITS – Intelligent Transportation System

GPS – Global Positioning System

GTFS – General Transit Feed Specification

O-D Matrix - Origin-Destination Matrix

RTI – Real-Time Information

GMTT – Global Marine Technology Trends

MNO – Mobile Network Data

GNSS – Global Navigation Satellite System

A-GPS – Assisted Global Positioning System

GPRS – General Packet Radio Service

UHF – Ultra High Frequency

RBS – Real Bus Stop

CBM – Condition Based Maintenance

MTR – Mass Transit Railway

RCP – Robustness Critical Points

ATIS – Advanced Traveler Information Sysrem

ESTRI – Efficient Spatiotemporal data retrieval method based on Impala

PPDP – Privacy Preserving Data Publishing

PPTR – Privacy Preserving Task Recommendation

GIS – Geographic Information System

VSP – Vehicle Scheduling Problem

IVCSP – Integrated Vehicle Crew Scheduling Problem

POI – Point of Interest

SAP – Spatial Affinity Propagation

ADC – Automated Data Collection

SP – Schedule Plan

IPTS – Intelligent Public Transit System

DTA – Drive-Time Areas

1. Introduction

In 2024, everyone is talking about data. Data is everywhere. Apps are everywhere to access this data which will be further used for providing assistance, control and planning. Main reason behind data collection by many transport service providers is to make the travel planning reliable by increasing the transparency of collected data. This can happen only if the data is fully exploited for providing “future traveler information systems“ by virtue of having “huge potential for forming an important piece of future information“ (Redmond et al., 2019). Numerous studies have been conducted so far to increase the efficiency and reliability of public transit, hence improving the State-of-art. In my thesis, I will focus mostly on reliability aspect and I will investigate further the extent of data utilization to increase the reliability of public transit.

Empirical studies conducted in recent years by Hossan et al. (2016); Koutsopoulos et al. (2017); Devarasetty et al. (2012); Ehreke et al. (2015) have shown that travelers actually value the reliability of travel time along with short travel times. On the contrary, when it comes to developing multi-modal itineraries, travel websites mainly focus on cost and time factors without considering reliability. Therefore, the existence of “gap between information available to the travelers“ and what these travelers actually value while making itinerary related decision, has been seen. This type of situation creates uncertainty in travelers’ minds which can further lead to distrust in the mode of transportation. Therefore, travelers’ decision-making process is highly affected by transparent reliability information for the whole itinerary, in addition to price (Redmond et al., 2019).

So far, automated data collection systems are helping transit sectors with transformation of planning, monitoring, scheduling and operations control of transit systems. Automated data collection is done through Automated fare-collection(AFC), Automated passenger counting systems(APC) and Automated vehicle location system (AVL). Real time AVL data is commonly available at the control center, while, real time availability of APC and AFC data is absent and the data is uploaded only overnight. This reason leads to “real time applications of AFC data” which has been emerging recently (Koutsopoulos et al., 2019).

Availability of advance technology is not sufficient, appropriate use of these technologies are important for real-time information dessimination. According to Cats and Loutos (2013), informations related to disruptions in service, crowding conditions, the remaining time until the arrival of the next vehicle or otherwise known as “prescriptive journey planners” can be obtained from Real-time information (RTI). Zhang et al., (2008) have conducted a study to “analyze the impact of RTI provision” on various aspects of travelers’ experience, including the satisfaction level of passengers. Whereas, Mishalani et al., (2006) and Dziekan and Kottenhof (2007) have conducted a study to see the impact of RTI on “perceived waiting time”, and Watkins et al., (2011) have studied the impact of RTI on actual waiting time. There are several benefits of installing RTI systems, such as; uncertainty reduction and to choose departure time, time saving through path choice change facilitation (Nökel and Wekeck, 2009; Cats et al., 2011). Unfortunately, RTI dissemination systems often lacks storing historical data (Cats and Loutos, 2013).

In modern cities, data can also improve citizens’ quality of life through the availability of new information sources from public transport providers. There are also wearable devices such as; smart watch, which provides the opportunity for novel application of new technologies and contribute towards improving citizens' quality of life. As the buses and vehicles are equipped with sensors and GPS technologies, they record and store their own paths, delays and other issues while moving around a city. Sensors installed in the streets also capture the volume of traffic. The information obtained from these sources helps in the development of applications that benefit the citizens' quality of life by opening multiple opportunities (Katakis, 2015). However, it is not an easy task to manage the public daily commute network, especially in recent times with accelerated urbanization, accompanied by increase in the population, seen mostly in developing nations. Hence, appropriate tools must be adopted by public transportation systems for the purpose of addressing such challenges by exploiting data (Ge et al., 2021).

As the service demand for passengers is increasing, the requirement for new management is also increasing, by making the evaluation of bus service an important aspect for the healthy development of urban traffic. (Shi et al., 2021). For building a smart city concept, data-driven transport policy act as a key pillar. Therefore, to increase the competitiveness and effectiveness in all sectors of the economy, especially transport sector, the capability to analyze big data provided by information and communication technologies and the capability to use this information in the decision-making process, serves as an important element for the management of urban mobility (Urbanek, 2019). Lock et al., (2021) have stated that, "present societal and governmental trends, emerging technologies, and new tech-enabled transport businesses suggest that the current system can change dramatically. As we continue to invest heavily in public transport infrastructure, big data analytics can further monitor the medium-and long-term performance outcomes and make better-revised decisions".

Apart from this, The Global Marine Technology Trends (GMTT) 2030 reports the need to exploit data from a wide range of sources varying and constantly changing structure and architectures by integrating a range of emerging technologies for developing robust, reliable and efficient solutions (Clarke, 2015). This raises the concern of level of exploitation of data till date and what are the probable gaps which still needed to be filled in order harness all the advantages of latest technology to make our life more comfortable. Therefore, the thesis will dive deep into answering the research question:

Where does data come from and in which areas of transit planning has it been used extensively? Up to what level does data has been exploited?

Research questions will be answered using literature review on the most recently published scientific papers that uses data/ big data in public transit planning. The second chapter is the most important one as it will serve as the foundation for answering research question based solely on review of literature. Therefore, chapter two will be subdivided into three mini topics where the first topic will cover the research question about "where does data come from" and other important factors related to data-driven public transit planning. Second mini topic will cover the research question about "in which areas of transit planning data has been used extensively". Third mini topic will cover last part of the research question on "up to what level data has been exploited?". Third chapter will be comprised of two parts. First part will highlight some challenges and opportunities while working with data. And second part will cover the perspectives of transit operators, passengers and policy makers. Fourth chapter will analyze the thesis based on the findings in the previous chapters which will then be followed by a discussion. Discussion part presents the point of view of author related to the thesis topic and its corresponding research questions. Chapter five concludes the thesis by summarizing all the findings of literature review.

2. Literature Review

A detailed literature review was preceded strategically for finding relevant papers. The goal was to find the most recent published scientific papers where data has been used in aiding citizen's life through improved public transit planning. Subsequently, this strategy will lead us towards finding the most appropriate answers to research questions. The results of the search were analyzed and categorized according to the points included in the five mini topics of second and third chapter. The research was initiated with a search process which is roughly divided into four parts:

1. General search: it was carried out by inserting key words in different combinations on several well known platforms including the u:search of Vienna University Library, Directory of Open Access Journals (DOA), IEEE Xplore Digital Library, Science Direct, MDPI, Research Gate, ACM digital library. Backward search was used to ensure complete search results.
2. Pre-selection: this step was carried out by analyzing the search result based on title and abstract of the papers.
3. Selection: full-text analysis of selected papers was done in this step which further limits the number of papers that are relevant for the literature review of data exploitation in public transit planning.
4. Analysis: it is the final step, where selected papers were analyzed extensively according to various criteria such as; year of publication, area of application, relevant to research question, availability and accessibility of full text article online, and contribution towards improving overall public transit planning.

This thesis have tried to maintain year of publication as latest as possible- in some cases, it was not possible to find the latest literature. The majority of the literature reviewed in this thesis are from well known and reputed journals such as; Transportation Research, Transportation Research Record, transport Policy, Elsevier, IEEE Transactions on Intelligent Transportation Systems, Public Transport etc.,

More than 300 references were cited in this thesis, where 126 miscellaneous consists of project reports, proceedings, thesis work, books, conference paper and single time cited journals. Among other journals, the ones, that have been cited for five times or more were presented in the form of chart in figure 1.

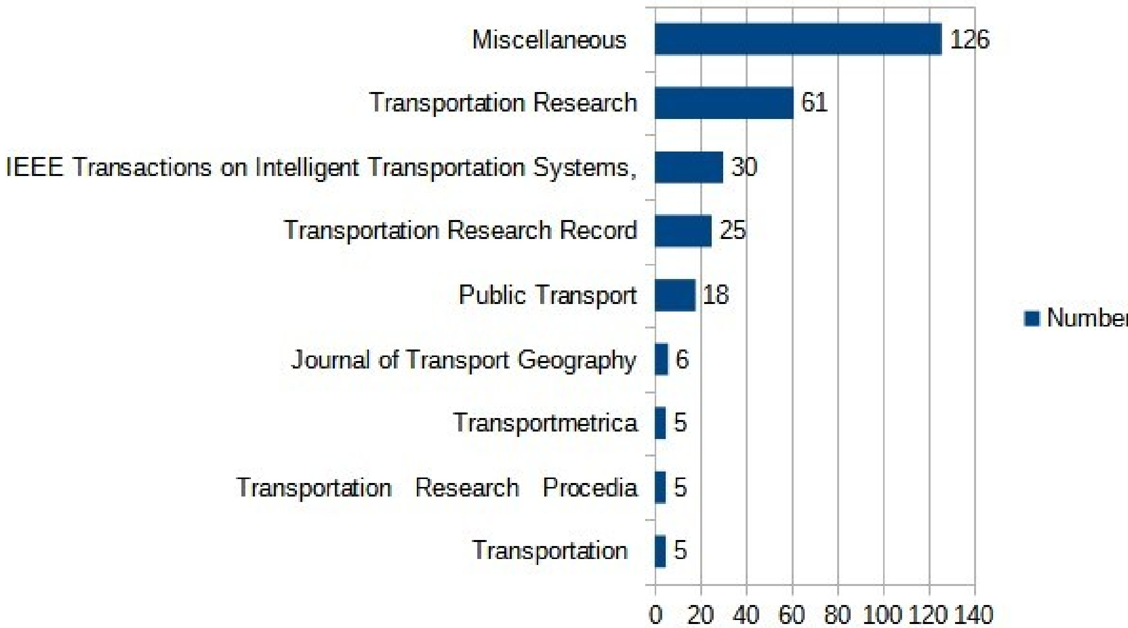

Figure 1: Most frequently cited journals in this thesis

Source: Own illustration

General search was initiated using different combinations of search terms in each of the databases. Few such combinations includes:

- “data” AND “public transport” AND “exploitation”
- “data-driven” AND “planning” AND “public transport”
- “data exploitation” AND “public transit” AND “improve reliability”
- “data-driven” AND “reliability” AND “public transport”
- “data” AND “latest technology” AND “exploitation” AND “transport”

This way, for each topic, sub-topic, and all the important points included in this thesis, similar search process was carried out. At the end of the search, more than 450 articles were returned from the databases. At pre-selection stage, review of abstract was done for the determination of relevancy of content to the thesis and to its research questions. After pre-selection, number of articles was reduced to 330.

Articles to be were included or excluded was not only based on the relevancy to the research questions, but also based on relevancy to many other topic and sub-topic and important factors mentioned in this thesis. This thesis is covering not one single aspect of data in public transport, but the whole environment involved in the smooth functioning of transport sector with the help of data. Which is why there is a need for literature review in almost every areas of public transport and large number of articles were cited in this thesis. Further, the inspiration for the construction of thesis structure was derived from transportation data hub, presented in the form of graphics in figure 2.

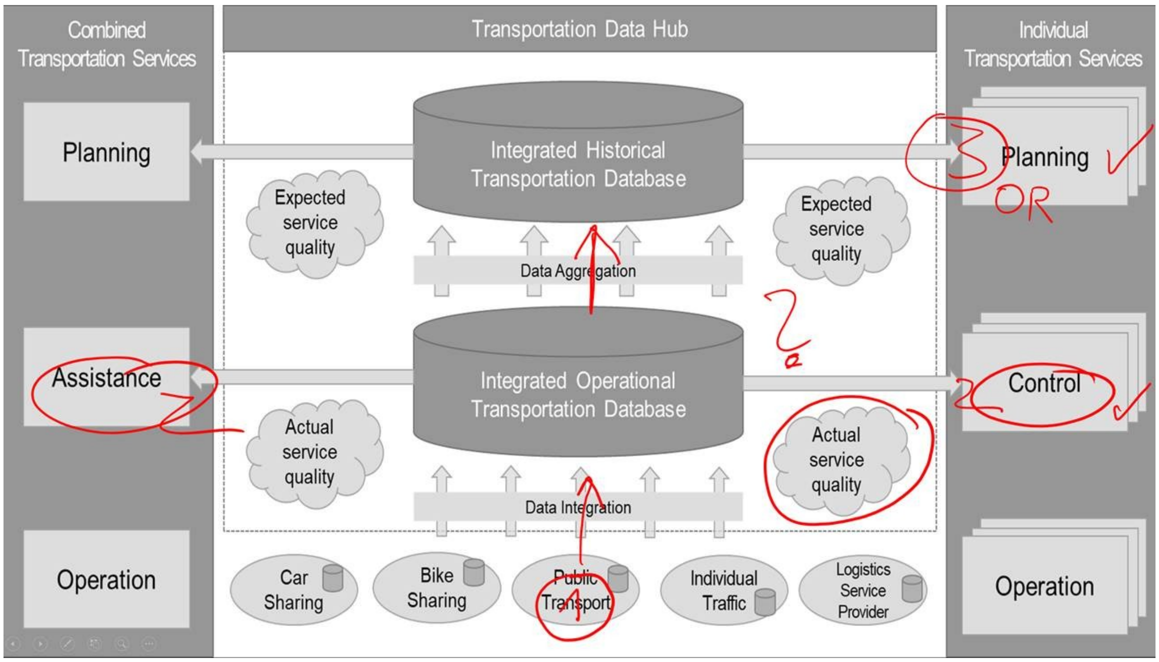

Figure 2: Transportation Data Hub

Source: From the repository of Univ. Prof. Dr. Jan Fabian Ehmke

2.1 literature review on – "*Where does data come from*", and factors related to data-driven transit planning and operation

Data types

Different types of data has been found throughout the literature which were used in transport sector. These are; AVL(automatic vehicle location)(Pi et al., 2018), AFC(automated fare collection)(Sanches-Martinez and Munizaga, 2016), Smart card data (Zhao et al., 2018), APC(automatic passenger counters)(Jung and Casello, 2019), GIS(geographic information system) (Liu et al., 2023; Droj et al., 2021), GPS(global positioning system)(Valdes et al., 2017), GNSS(global navigation satellite system) (Willumsen, 2021), GTFS(general transit feed specification)(Bast et al., 2014), ITS(intelligent transportation systems) (Iliopoulou and Kepaptsoglou, 2019), ATIS(advanced traffic information system) (Williams, 2019), Mobile network data (Willumsen, 2021).

Data sources

Bast et al., (2014) introduced a framework for public transit based on freely available General Transit Feed Specification (GTFS) timetable data. With the help of this framework a world-wide live map was created, which can capture the movement of all buses, trains, subways and ferries.

Shafique and Hato, (2014) have predicted the mode of transportation using acceleration data. In an attempt to detect a device carrier with level of accuracy, acceleration data was used by transport service providers. As smartphones comes with an in-built accelerometer and other sensors, recorded data can be used for the transportation mode determination of customers. These recorded data can also act as “an alternative“ to traditional common “travel surveys“, with the advantages of performing customer-centric advertisements.

Sanches-Martinez and Munizaga, (2016) reports that, there are various sources of data being used in public transportation. These sources includes; automatic vehicle location (AVL), automated fare collection (AFC), and automatic passenger counting (APC), sensors installed on vehicles and stations, Wifi, video surveillance, crowdsourcing, Bluetooth, and cellphone networks, and various forms of published urban data. In order to track bus, GPS combined with odometers and inertial navigation system has been seen to be used. Whereas, train tracking systems are often done with the help of track circuit occupancy along with radio-frequency identification (RFID) and satellite-based location. Data can also be crowd sourced while travelling through smartphone applications and web. These applications are namely; Google Maps, Waze, Moovit, Twitter, and Facebook.

Zhao et al., (2016) have acquired data for their study from single sensor. The computer vision-based traffic surveillance system can collect traffic data such as; speed, density and volume of data for uninterrupted flow corridors in the form of big data to monitor traffic parameters with a coverage of 25 square miles.

Ferreira et al.,(2016) have used various sources of big data, among which “a data mining approach based on a Naïve Bayes (NB)” was adopted to correctly analyse and pull beneficial knowledge from different data sources. Another approach used by them was tracking down citizen’s mobility and to make them aware of their footprint through mobility invoice system. Application can be developed using mobile device data.

A project carried out by UN Global Pulse, (2017) on ‘Using big data analytics for improved public transport’, uses data such as GPS bus data to map problematic locations, and using passenger tap-in data to gain insights on passenger behaviour through origin-destination (OD) statistics and waiting times.

Anda et al., (2017) have found some advance big data sources and by harnessing theses sources can aid in understanding travel behaviour of the passengers. These data can also allow the transport planners to compute what-if scenario. Examples of these data sources includes; mobile phone call data records, geo-coded social media records, and smart card data, which allows the transit operators to observe and understand the mobility behaviour and other detail.

Valdes et al., (2017) did a project on ‘Development of transit performance measures using big data’ in San Juan Metropolitan Area of Puerto Rico. They needed data “generated by GPS-based Automatic Vehicle Location systems installed in public transit vehicles” to help improve the service of public transit. This was done by merging the collected data “with transportation demand related data available from other sources” in the form of Big Data management program to estimate system performance measures and proposing new metrics accordingly.

Nagy and Simon, (2018) referred to some wide variety of sensor technologies as data sources which were used for traffic flow prediction. These sensors are: inductive loop sensors, magnetic sensors, video image processors, infrared sensors, microwave radar sensors, audio sensors, laser radar sensors, along with Global Positioning System sensors (GPS) equipped in smartphones and vehicles.

Pi et al., (2018) uses AVL and APC data in order to understand the transit system. They have introduced a transit data analytics platform based on APC and AVL for assessing the service quality of public transit systems systems. With the help of this platform, users and decision makers can systematically examine many aspects of the system performance such as; service quality, bus travel time, bus bunching level, passenger waiting time, on-time performance, stop-skipping frequency and bus fullness. Their work was based on archived AVL and APC data. Planned future work of authors is to incorporate real-time AVL & APC data to detect and resolve bus bunching issues.

Zhao et al., (2018) have used transit smart card data from London, UK in order to propose a model for individual mobility prediction.

Williams, (2019) have carried out a study where the author explored the use of ATIS (advanced traffic information system) "to manage traffic congestion in developing countries". The author have stated that, "ATIS is extensively used in the developed countries and is regarded as an important component for road traffic control and management".

Willumsen, (2021) refers to two main sources of data for transport modeling: 'mobile network data' or MNO data and 'smartphone data'. Smartphones uses the data generated by integrated microchip Global Navigation Satellite System (GNSS), in the form of GPS. GPS alone cannot increase location accuracy, in order to do so, GPS in smartphones is combined with cellular data and Wi-Fi network signals in the form of Assisted-GPS (A-GPS).

Ait-Ali and Eliason, (2021), in an attempt to study the passenger origin-destination for transport planning, have found the existance of various sources for data collection. But, it involves an additional cost for "additional data collection". Therefore, while "estimating time-dependent" origin–destination matrices for their study, the authors have also studied the value of additional data "using a case study from the London Piccadilly underground line". The sources of additional data were APC; vehicle weighing system and travel surveys.

Ge et al., (2021) reviewed the potentials, challenges and complements of transit data sources. For their study, they refers to AVL, APC and AFC as endogenous sources of data and weather, social media, smartphone, traffic and survey data as exogenous sources of data.

Collection/acquisition of data

Basic concept

According to Torre-Bestida et al., (2018) "collection of transportation and mobility data refers to set of different tasks aimed at (i) acquiring the data captured and sampled from signals, (ii) converting the samples in digital values that can be manipulated by a computer, and (iii) bringing such digital data into the processing flow of a given data system". Therefore, "in transportation and mobility, data collection can be understood as a task of compiling data captured by devices and sensors and generated by citizens and vehicles through different channels".

AVL data collection technique

According to Tilocca et al., (2017), there are three different ways of collecting AVL data: (1) with infrastructure based sensors, (2) with mobile devices such as GPS, and (3) with mixed sensors such as dead reckoning or odometric location. These collected AVL data are then transmitted through AVL system based on two technologies (radio signal and GPRS). Not real-time bus location data is shared with AVL centre at regular interval and real-time bus location data is shared in case of exceptions.

APC-AFC data collection techniques

According to Antrim (accessed on 7.7.2022), APC-AFC comes with components or sensors, which automatically collects passengers data. APC data is collected with the help of sensors "along with GPS locations and time stamps onboard each vehicle" and communicated through "a wireless connection to a central server". Whereas, AFC data is communicated through validation of electronic fares, where validation id done either by automated agency-owned device-based or by rider-owned device-based validation.

GTFS data acquisition techniques

Wu et al., (2023) have reported that, GTFS is a data service providers, which acts as "an intermediate layers between users and transport operators", where "it offers a data service bringing together all the modes of existing transportation". With the help of interface, GTFS service providers connects "with a multimodal transport system and cloud services" for data collection, where "real-time information is shared by cloud computing services". These collected data underwent pre-processing steps of data validation, data de-duplication. Processed data of real-time and static data are then integrated in the form of data fusion. These processed data are then "transferred to the Transport Performance and Analytics (TPA) database" for further use.

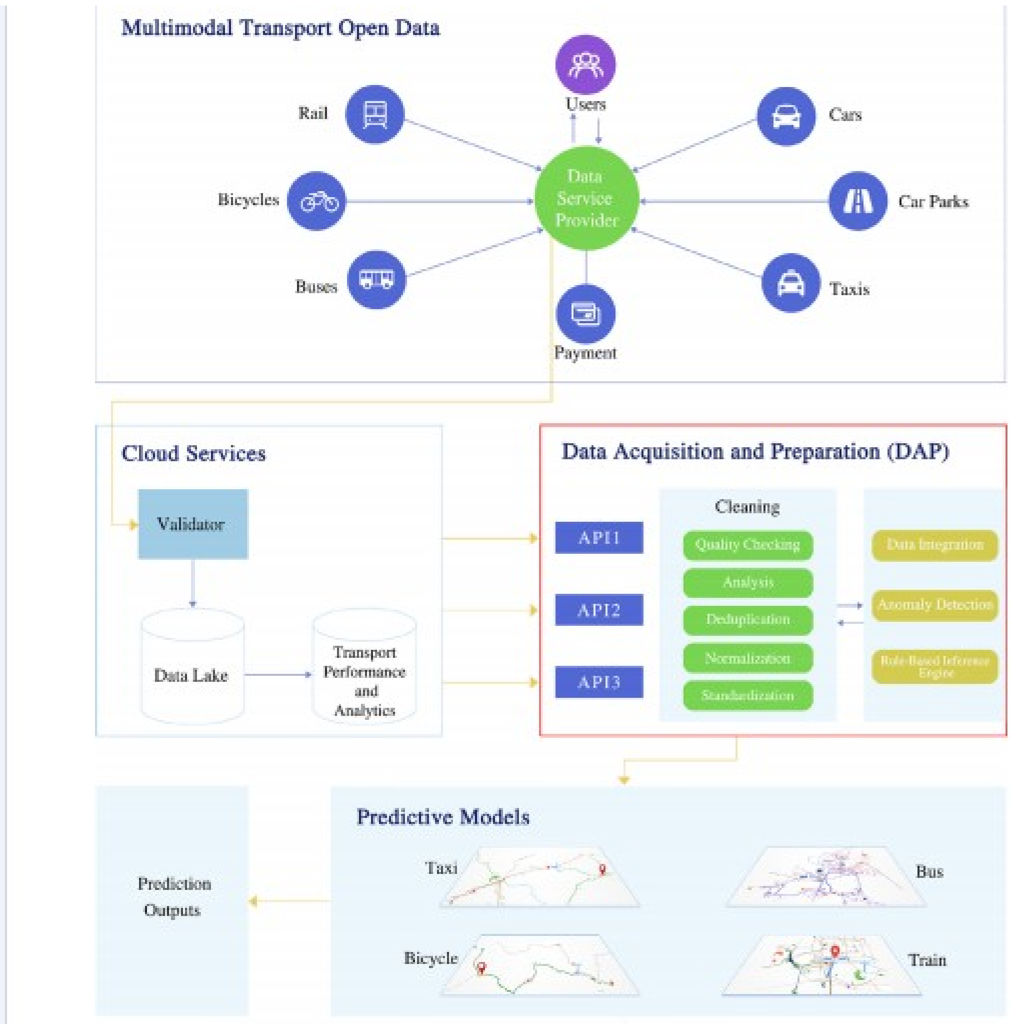

Figure 3: “A GTFS data acquisition and processing frameworks”
Source: Wu et al., (2023) A GTFS data acquisition and processing framework and its application to train delay prediction

ATIS data collection and acquisition technique -

Williams, (2019) have presented the architechture of ATIS composed of “flawless data collection practice, processing, and transmission”, which is done step by step of data acquisition, followed by data fusion then data transmission. ATIS providers collects data in two different ways: (1) with stationary inductive loop detectors, and (2) by gathering road traffic data. Data collection with stationary inductive loop detectors is costly and owned by government. Road traffic data are gathered through floating car data (FCD) and floating phone data (FPD). There is a three different ways of collecting FCD: (a) through GPS, (b) with personal navigation devices, and (c) with short range communication system. In case of floating phone data (FPD), data is collected through the detection of mobile phone position of the person inside the vehicle. Data fusion is the next step after data collection. Data fusion is done for the purpose of making the data complete by closing the gaps within data. Data fusion is then followed by data transmission. Data is transmitted either through FM radio broadcast , or through digital audio broadcasting (DAB).

Infrastructure

Ge et al., (2021) revealed the importance of infrastructure in data driven decision making. Through appropriate infrastructure, the capacity for data acquisition from external (weather, traffic) and internal sources of data(AVL, APC, AFC) can be increased and the process of handling data can be improved. Infrastructure such as GPS and smart cards allows to extract information concerning vehicle location “and the number of passengers“ from AVL, APC and AFC systems. Mobile phones are the example of other type of infrastructure which allows to collect data about passenger density in public transport means. Infrastructure is also needed to allow for large-scale video-based approaches to passenger counting. Therefore, Ge et al., (2021) concluded that: to harness the advantages of data-driven decision making, there is an urgent need for “investment in infrastructure” with varried capacities based on the “budget and expertise of the agencies”.

Integration

Ge et al., (2021) emphasized on data integration as it is one of the important aspects which involve “merging different data sources”. According to the authors, simply acquiring data from various sources, which are of “heterogeneous and inconsistent” in nature is not enough. There is a need of combining these acquired data of different forms into homogeneous ones with the help of “pre-processing” tools for the purpose of easy use by all stakeholders of transport sector. Integration is done through “standardization, validation and matching” of data.

Standardization

According to Ge et al., (2021), “standardization is a functional requirement to avoid an additional computational task for data pre-processing”. Presentation of data sources such as AFC, differs from agency to agency, which creates the “problem of standardization“. This problem can be solved by exploiting data and extending to a common formats, such as GFTS and NeTEx. Although adoption of data standardization is an emerging trend nowadays and the agencies are increasingly embracing the concepts of open data, but there is a need for incorporation of non standard data into data standardization process for improved analysis. However, the authors have noted that, public transport agencies must take the issues of standardization while publishing data that are of “frequently provided from statistical services“. A work done by Tibaut et al., (2012) was found in this regard who proposes a sustainable European Passenger Information System(EPIS) as a solution to this problem.

An example of standardization using NeTEx was presented by Network Timetable Exchange (accessed on 06.07.2022). “NeTEx is a technical standard of the European Committee for Standardization (CEN)”, which serves as a means for data exchange and provides information about schedules, stops, fares, timetables and related data of public transport to its passengers. Several parts of NeTEx with “specific functional subsets” allows for proper information to the passengers. Public transport network topology provides information about scheduled timetables, fare information along with general passenger information

But works on a part of “technical specifications” are still in progress. The main purpose of this standard is to create a general XML format which will allow “an efficient exchange of transport data among distributed systems” and which is currently active in Germany (https://www.vdv.de/netex.aspx, accessed on 06.07.2022). The core application and concepts of German VDV includes: obtaining seamless passenger information by linking various modes with regions through interoperable “passenger information systems and related data”. German/European DELFI (“Durchgängige elektronische Fahrgastinformation”) system is another popular example that can be found among “various systems and projects” which has been seen to develope over the period of time (accessed on 06.07.2022).

Validation

Ge et al., (2021) states the importance of data validation in order to integrate different data sources. Integration of different data sources is done to get similar information from those different data sources. By integrating AFC data, smartphone data and survey data, their mutual information could be validated for further use by various agencies. The authors have stated that, the reason behind consideration of data validation is, massive increase in data availability along with its resulting challenges arising with transit data. In order to make the most of these available data, there is an urgent need for data validation. Same information can be retrieved from different data sources and transit agencies can combine these information according to their needs. The authors have also presented an instance of encountering an erroneous APC system on a bus by using smart card data and an AFC system which was caused due to the lack of validation of data.

Munizaga et al., (2014) have worked on validation of public transport origin-destination (OD) matrices by combining AFC and AVL data. The result of their study was used for planning purposes, both by the operators and by the transport authority.

Nunes et al., (2016) have presented anther example of data validation. The spatial validation features was proposed to increase the accuracy of destination inference results and to verify key assumptions present in previous origin-destination estimation literature. Additionally, “a methodology for estimating the destination of passenger journeys from” automated fare collection (AFC) system data was also described by the authors.

Liu et al., (2017) gave “insights on the validation of both AVL and AFC data”. They also proposed a data driven methodologies to increase data validity using pattern recognition and statistics inference.

Matching

Ge et al., (2021) have encountered a problem related to data matching that was emerging during the integration of “different data sources on linked information”. According to the authors, there is a chance for arising such situation while “combining APC bus data with static stop data”, when the buses are not stopping at the intended stop or stopping at “an area slightly different from their estimated” stop. The authors have also claimed that, while matching APC and AVL data, APC data could be more noisy and unstructured than the AVL data. An example to solve this data matching problem was seen in the work of Barabino et al., (2014). Barabino et al., (2014) have matched the data with real bus stop (RBS) data before analyzing the APC data, and any ambiguity deriving from possible overlapping areas (OAs) were also removed.

Data Processing

Data processing is the immediate step after collection and integration of data. According to Ilyas and Chu, (2019), problem arising during data management was related to maintaining the data quality, as poor data quality often “leads to inaccurate” analysis, hence, inaccurate decision making. The authors have reported that, “poor data across businesses and the U.S. government are reported to cost trillions of dollars a year”. From the findings of other survey studies, the authors have revealed that, scientists were also reported having theoretical and practical problem to work with poor data quality, and finding it difficult to arrive at “effective and efficient data cleaning solutions”.

Massobrio et al., (2016) have “addressed the problem of processing large volumes of historical GPS data from buses”. Their case study was done in Montevideo, Uruguay. According to the authors, “the main goal of the data processing is to compute relevant metrics to assess the efficiency of the public transport system”, as the main problem with collected GPS data from buses was to compute “statistical values to assess the quality of the public transportation system”. They have proposed a solution to solve this problem by adopting map reduce approach. Before implementation of this approach, they have processed the collected data by diving into pre-processing stage and statistical generation stage. Pre-processing stage delivered the filtered data which to be used as input for further analysis. The further analysis involved an experiment, where the experimental results showed that proposed map reduce “solution scales properly when processing large volumes of input data”, and achieved a speed upto 22.16 through the usage of “24 computing resources”.

Wu et al., (2022) in their quest for designing a ready-to-use GTFS data acquisition and processing format, did a detail study on GTFS data processing. They studied static and real-time GTFS data to achieve their goal which was neccessary step for the prediction of multivariate multistep train delay in Sydney, Australlia. Additionally, they have also proposed a comprehensive data cleaning workflow which further aided in delivering less noisy GTFS data. Their proposed format was further validated through an experiental test method which successfully predicted the regular and irregula train delays.

Ou et al., (2022) have also shown the importance of data processing in their study. The authors were working towards quantifying bus delay which was caused by COVID-19 from the period of February 2020 to March 2020 across the Sydney metropolitan region.

As the data collected for their study was heteregenously sourced, there was an urgent need for data processing. Through data processing a unified input data was provided which was being used by the authors to study and solve the bus delay due to the pandemic.

Privacy and security

Sarkar et al., (2015) used number of applications to address the privacy and security issue. One such example of preserving privacy in smart phone data is by encrypting the data.

Yang et al., (2020) have acknowledged the privacy issue while working with transit data. In order to reduce the privacy breach issues while dealing with transit smart card data, they have proposed “privacy preserving data publishing algorithm”. The proposed algorithm was suitable to publish “differential private” sanitized data set which was found efficient than previously mentioned algorithms, reviewed by the authors. The authors have claimed that, the proposed algorithm is not only suitable to solve the privacy issue while using transit data, but can also efficiently deal data set collected from social media, navigation maps and ride sharing.

According to Ge et al., (2021), “data privacy and security is an essential issue that has to be taken into consideration when storing and processing data sources, especially AFC and smartphones”. The authors also further states that, both privacy and security issues are interrelated with one another.

Shu et al., (2021) proposed an approach to encrypt the data before being outsourced to the crowd sourcing platform. They proposed “a privacy-preserving task recommendation scheme (PPTR) for” crowd sourcing, with the ability of achieving “the task-worker matching” by maintaining privacy preservation of “both task privacy and worker privacy” and establishes a security goal. The proposed scheme was further tested for its efficiency against other three schemes using data collected from third-party platform.

Stockburger et al., (2021) found a way to manage identity in public transport by developing self-sovereign identity(SSI) - a user-controlled resilient identity management systems that are enabled using block-chain based ledger technology. The authors have claimed that ITS being at its peak and easy connectivity through internet also pose a threat for data breach and compromises confidential data. The authors found block-chain as a solution to this problem and conceptualized bloch-chain technology in transportation sector. For this purpose, the authors have also classified and defined different types of identities, which was followed by comparision of four block-chain based SSI technology such as; Sovrin, uPort, Civic, and Namecoin. The authors have further identified different stakeholders and their requirements for a seamless travel experience by managing de-centralized identity for a pan-European transportation ticketing system.

Storage

In general, data are categorized as structured and unstructured data then storage is done accordingly. “Structured data is comprised of clearly defined data types with patterns that make them easily searchable. Unstructured data is comprised of data that is usually not as easily searchable, such as; audio, video and social media postings” (https://www.datamation.com/big-data/structured-vs-unstructured-data/ accessed on 28.02.2022).

Wong, (2009) and Khwaja et al., (2010) have defined unstructured Condition-Based Maintenance data as, “technical documentation (TD) of tools and component, photos taken by Thermo graphic cameras, frequency spectrums collected by vibration spectrum”.

Whereas, Maktoubian, (2017), have defined semi-structured data as, tags or other types of markers, text-based, are being employed for the depiction of some elements. Such data fitting in this category are, “emails, XML files, and log files, etc”.

Li et al., (2011) in their attempt to study the traffic control and management system in the evolving computing paradigm, came across to the agent-based traffic management system, which dealt with the dynamic traffic environment. They have proposed cloud computing in the form of intelligent transportation clouds to cope up with large amounts of storage arising from mass transportation data.

Ge et al., (2021) raised an important issue of data storage in transportation with respect to the budget. Real-life traffic data such as AVL is growing in developong countries leading to less storage capacity of data compared to the data growth. This was a matter of concern as the transportation systems are producing huge amount of data from variors sensors and cameras to run the services efficiently and smoothly. The authors suggested two options as a solution to this problem; use of cloud computing and use of available open data tool. They have stated that, “cloud computing technologies help agencies manage large amounts of storage“. GTFS as a open data tool is also very useful to manage data but there is a need for involvement from the transportation agency with an additional requiremement for investment in underlying technologies in order to facilitate the use of this tool.

Barnes et al., (2020), have also confirmed the use of GFTS (General transit feed specification), to solve the storage issue as the GTFS specification is extended by a real-time transit data feed since August 2011. In their quest for predicting real times bus travel times using machine learned model, GTFS data allowed stripping down real-time transit data to training data format for their study. The authors preferred working with GTFS-real-time data as it was easy to work with and provides several information about trip updates, bus delays, cancellations and route changes..

Information management: Endogenous data vs. Exogenous data

Voss and Gutenschwager, (2001) have defined information management (IM) as "purpose-oriented provision, processing and distribution of the resource information for decision support, as well as the provision of respective infrastructure".

An interview with Axelsson, (2020) at the Swedish Transportation Administration served as an example of information management in transport sector. As a "Building Information Management (BIM) strategist", the aim was to ensure management of information efficiently and management of infrastructure sector with the help of a common digital facility. BIM is a tool to manage asset and lifecycle of asset information, which is used widely for design, building and maintenance purpose. The "BIM standardisation issues of rails, roads and bridges at the Swedish Transport Administration" was solved by focusing on (1) establishment of information structure and hierarchy based on category system, (2) well defined information provision (3) standardised way of information delivery between various stakeholders, leading to a overall management of asset with efficiency.

Puusaag and Palmi, (2021) have also presented us with the neccessity of information management in transport sector through the implementation of building information management (BIM) in Estonian Transport Adminstration. The BIM was implemented in Estonian Transport Adminstration with "five main goals; ease design and construction processes, ease data collection and usage, quality BIM models, common information exchange, transparent design, construction and maintenance processes". The authors have stated that, efficient digitalization and BIM in transport sectore is possible by, BIM working in close cooperation with “Estonian Construction Cluster and also with Estonian Asphalt Pavement Association“.

Ge et al., (2021) have perceived information management as “an instrument for making information distribution operable”. However, the latest advancement in information and communication technologies are failing to make distribution of information operable up to certain extend. The reason was for this was, “lack of putting data into perspective in the sense of this definition”. Hence, the authors came up with an idea of developing a three layer model of information management to aid in distribution of information, which will further aid in putting data into perspectives, thus aiding in better decision making by different stakeholders. In the three layer model, first layer consists of layer of information usage/deployment, followed by layer of information system, which is followed by layer of information infrastructure. Then, the authors have categorised the information management into groups; information management of external data, information management of internal data. They further categorised the external data into static and dynamic. These categories of information management then put to basic three layer framework model for better understanding, where layers are shown in vertical arrangement and categories of information management in horizontal arrangement (Figure 4). By making the division of information management into external and internal data, and, also, by categorising the external data into static vs. Dynamic (Figure 5), it became easier to address different types of issues by their corresponding stakeholders (transit operators, policy makers and passengers).

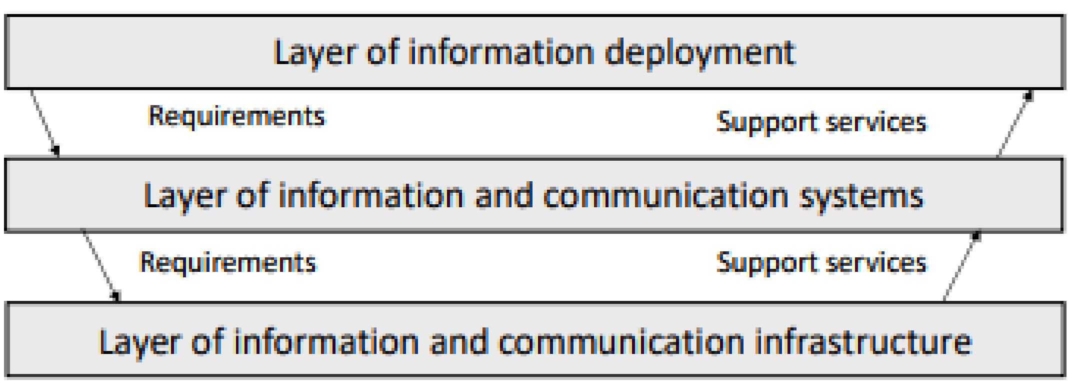

Figure 4. Information Management framework based on three-layer model

Source: Ge et al., (2021) Review of Transit Data Sources: Potentials, Challenges and Complementarity

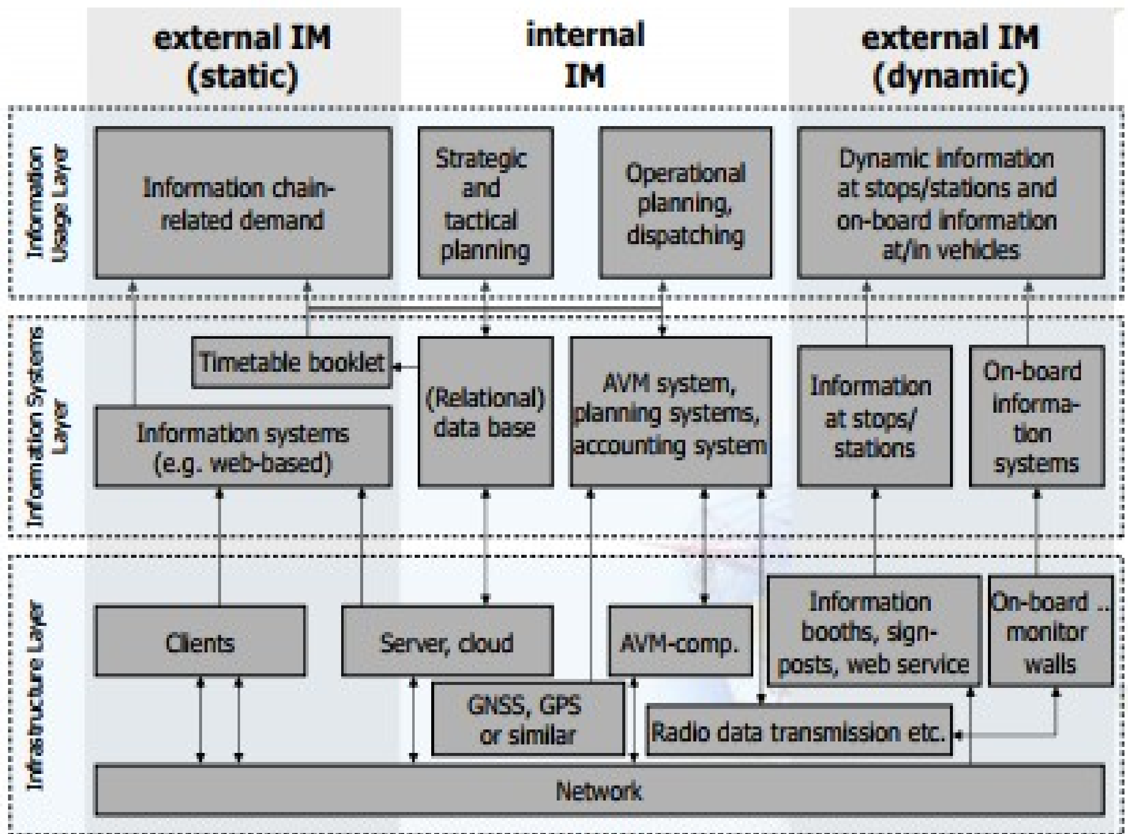

Figure 5: Management of internal and external information at different layers
Source: Ge et al., (2021) Review of Transit Data Sources: Potentials, Challenges and Complementarity

The systematic implementation of transit big data

According to Lu et al., (2021), the main purpose behind implementating transit data systematically, is for deriving “valuable insights for decision support and automation from huge volumes of heterogeneous sensor data”. Due to this reason, several enterprises and organisations were found investing on big data technologies for storage (“Hadoop, Data Lakes, NoSQL Databases“), processing (“ Spark, Hadoop, data governance“) and analysis (“ Spark, cloud computing, edge computing, artificial intelligence, modeling and optimization“) of data. The insights obtained through implemetation of big data are later served as a foundation for decision support for transit planning, operation and policy making.

Liu et al., (2021) have presented us with their foundational work, where one can see, how systematically implemented data can be further used for solving various transit related problems. One such case was seen through the work of the authors, where the goal was to predict passenger flow from both ‘macroscopic level’ and ‘microscopic level’. The passenger flow was predicted with decision tree based model and this model was implemented in china, using intelligent IC card data. Decision tree based model was later complemented with deep neural network model for accurate passenger flow prediction. The authors concluded that, their proposed model had performed better in the real world scenario than the existing models.

The work of Lu et al., (2018) provided us with some insights which can be helpful in sound decision and policy making. The result of their integration of interdisciplinary work showed that, there is a “gap between the idealized service and the real service”.

However, the work of Iliopoulou and Kepaptsoglou, (2019) was all about finding out the extent to which data driven implementation has been successful so far in public transit planning and optimization and what can be done by the researchers and transit operators in coming future. The authors have concluded that, there is a need for modification of existing models and algorithms for achieving “the transition to data-driven planning“. So far, the authors have found no straight path for such transition. They have also confirmed that, one single independent data is not sufficient for efficient application into “transportation-related optimization models and algorithms“. Therefore, future research work should be focused on “automated validation procedures“, where ambiguities in data set can be detected and rectified automatically with the help of latest techniques of artificial intelligence.

Digital twin

Kaewunruen and Xu, (2018) have implemented the concept of digital twin for their study involving railway building construction. The authors have applied digital twin in the form of Building information Management (BIM) for renovation of old railway station of King’s Cross station, London. According to the authors, digital twin or BIM can provide wide range of integrated information which is easily comprehensible, which has the potential to change the traditional way of working construction industry. BIM application using

Revit-based simulation helped in adoption and transformation of 3D building model to 6D model for renovation work. The authors were certain that, the result of their study will be beneficial to provide guidance to the construction works involving BIM application for planning, designing and operating sustainable and efficient construction projects.

Ge et al., (2021) stated that, digital twin represents “a physical process, person, place, system, or device“ digitally. a digital twin consists of a digital representation of a physical process, person, place, system, or device. It is a favorable technology “for digital transformation“ with the advantages of merging data from various sources. The authors have emphasized the need to take advantage of latest data acquisition technologies to bring improvement in data collection. Digital twin facilitates this process by providing necessary information, collected from different data sources. One such case was presented for exploiting digital twin concept was through data integration from AVL vehicles data, AFC passengers data and in APC traffic data.

White et al., (2021) have associated the concept of digital twin with smart cities. In their attempt to create smart city based on digital twin where citizen can interrect and give their feedback, also revealed other advantages of using digital twin concept. Digital twins can pinpoint strengths and weaknesses of each plan before being unfold in to real world, through simulation . The authors have conducted a simulation study by taking the smart city data from Dublin, Ireland. With the help of Unity 3D Software, they perfoermed simulation on skyline, green space of the city, user tagging simulation, flooding simulation and crowd simulation. Their aim was to make the simulation study as much accurate as possible so that it can help in better policy making. Indeed, the result of their simulation study was successful in proving that digital twin can be used for range of urban planning decisions and policy including mobility.

Exploitation

Basic concept

Ge et al., (2021) have defined exploitation as a way to make use of collected information which “is beneficial to the transit service and to the passengers”. According to them, exploitation of data is done through judging the passengers information from various perspectives which includes “visualization and service optimization”.

Exploitation of public transit data through visualization

Kwoczek et al., (2014) have developed a prediction and visualization technique which is capable of predicting non-recurrent traffic congestion a way ahead of time. In order to predict traffic congestion, which is occuring due to planned special events, a special visualization tool was developed by the authors. The study was carried out in Cologne, Germany, using the traffic information generated from various sources such as; floating car data, GSM probe data, data from stationary sensors. The result of the study was successful in identifying the influence of planned special events on traffic, and performed up to 35% better than the existing state of art solution.

Tao et al., (2014) have observed that, despite the growing implementation of BRT (bus rapid transit) system worldwide, there was a lack of in-depth studies about their spatio-temporal dynamics. Therefore, in order to fill this gap, the authors have carried out a study which will explore further potentials of BRT. Their study used smart card data from the city of Brisbane, Australia. They have combined smart card data along with GTFS data to study the spatio-temporal dynamics of the passengers. They used GTFS data because, GTFS data provides geographic coordinate and alighting information which smart card does not provide. The result of their study was successful in providing important insights on BRT in sustainable urbon transit planning.

Corsar et al., (2017) have provided an example of information exploitation in their work. They have acknowleded that, rural areas are lacking neccessary infrastructure to provide real time passenger information and involves huge cost in order to deploy and maintain latest technological infrastructure outside urban areas. As a solution to this problem, the authors proposed a GetThere smartphone app for the rural passengers to share their location while travelling with the buses. This app can access the bus information stored in the ecosystem. With the help of semantic web technologies, the whole services can be exploited. “Semantic Web technologies are required to support the integration of heterogeneous mobility data that is annotated with meta-data providing schema and provenance information“.

Huang et al., (2017) adopted a visualization approach to study the relationship between destination choice and business cluster. They performed travel analysis with the help of social media data such as; Twitter data. Due to the increasing popularity of social media data, at the same time with advancement in data collection techniques which “has evolved from traditianal dairy household survey, GPS studies, smartphone-based surveys, to social media data“, the authors have found huge opportunity to collect travel behavioral data at microscopic level. Through visualization technique, the authors have found that, short “distance can attract more common visitors“, long distance and “better connections of roads and its association with more check-ins“.

Another best example of data exploitation was seen in the work of Jevinger A. and Persson, (2019), where collected information was used for improving public transport disturbance management. In their quest for „exploring the potential of real time traveler data in“ disturbance management, the authors have discovered that, AFC system can be successfully used to manage unplanned disturbances. AFC provide real-time informations which is very helpful in supporting various actions and decisions in case of unplanned scenarios. It is because, AFC system detects passengers while on board, at the entrance of the vehicle, at the vehicle exit, on platform, which provides better opportunities for acquiring better information.

Prommaharaja et al., (2020) have further explored the untapped potentials of GTFS data through their study where they used GTFS data for visualizing the operation of public transit system. Therefore, the authors aimed at describing the recent development of “GTFS data visualization tool“ to draw qualitative visions and information about spatio-temporal transit service patterns. Their study was carried out using GTFS data from Calgiry transit. With the help of GTFS data, they further aimed to develop “a new public transit system operation visualization tool (called PubtraVis)“ composed of “six visualization modules“ which has the ability to “reflect on different transit system operational characteristics; mobility, speed, flow, density, headway, and analysis“. The result of their study proved that, GTFS visualization tool can assist “communication between transit operators, city authorities and the general public regarding public transit planning and operation“.

Lock et al., (2021) have acknowledged that, there are tremendous amount of static and real- time transport data has been generated and being archived everyday. These data were generated ans stored in different formats in different places. In order to make use of these valuable data for better planning and operation in transport sector, the authors have suggested for visualization techniques. For their study, a general framework for data visualization was demonstrated followed by exploring different data visualization methods. They have presented “a series of visualisation techniques applied in a real-world scenario to create a framework which enables the exploration of these large, continuously broadcast public transport performance datasets". Real-time GTFS feed was the primary source of collected data for their study. The result of the study was successful in describing effective collection and use of public transport data which is continuously updating. Through visualization, updated public transport data can be easily generalized and adopted globally, by establishing a framework, where, all interested stakeholders "can view and integrate this data at a granular scale, and, implements this framework in a real-world case study". Through their study, the authors have also recommended to the transport agencies about the deployment of "large-scale visual analytics of open-source generalised global transportation feeds".

Pi et al., (2021) in their attempt to identify cause of traffic congestion in everyday life, propsed "a technique" which has derived from traffic flow theory for the purpose of analyzing reason for traffic congestion, by extracting vehicle flow data from "GPS trajectory and vehicle detector data". Then, they proposed a visual analytic system in order to analyse the cause and effect of traffic congestion. The proposed analytic system was implemented by undertaking a case study in Beijing, China. The result of their was successful in analysing causes and effects of traffic congestion in Beijing. Moreover, they were also successful in "classifying abnormal traffic congestion patterns, such as traffic accidents and bad weather, by combining external data".

Fazari et al., (2021) have presented an example of exploitation of technologies through visualization by developing a visualization tool called 'Mobisuite'. Smart card data was used for their study because this data have been used by public transportation due to its meaningfulness for mobility trend analysis, which further can assist "public authorities and transportation organizations" in the process of decision making. Mobisuite can successfully estimate some of the passenger indicators such as; "passenger kilometers, maximum load, number of on-boarding passengers, average load" through virtual check out algorithm. This visualization tool was applied and studied in the Piedmont region of Italy. The authors have concluded that, "Mobisuite" has the advantages of making the communication simple between different stakeholders. It can also display the whole transport network of public transport system at a glance".

Exploitation through service optimization

Pelletier et al (2011) also aimed to explore the potentials of smart card data apart from being used for revenue collection. Their intention was to fill the gap between technical element of smart card system and the data collected from this system. By doing so, an example can be created for interested parties, which will support in backing up the existed values of smart card system. For this purpose, the authors have carried out this research work based on literature review. Their ressearch work was indeed successful in exploring several other potential of smart card system which are; “providing data to researchers and planners, to enhance the strategic, tactical, and operational performance of transit authorities“.

Ma and Wang, (2014), have confirmed the several opportunities of latest data collection technologies which is yet to be explored and exploited. The authors tried to take the initiative of exploring the potentiality of these technologies by developing an online platform for monitoring transit performance using AFC and AVL data. Their proposed framework was implemented in Beijing where they have estimated individual rider’s origin and destination with the help of data mining technique. The result of their study proved that, this online eplatform can successfully monitor the performance of transit network with an intention “to take advantage of e-science initiative for data-driven transportation research and applications“.

Anda et al., (2017) in their study, aimed to present the reader with an understanding about latest advances in Big data source and how it is shaping the mobility flow. To accomplish their goal, the authors have carried out a literature based study, which revealed that smart card data can be exploited in many ways for transport planning. Among them, the most frequent ways to exploit smart card data are for; individual trip reconstruction, estimation of origin and destination matrices and inclusion of smart card data to agent based simulations.

Brakewood and Watkins, (2019) have conducted a study with an objective “to compile a literature review of studies“ for the purpose of assessing the benefits of real-time information provision to the passengers. Studies on real-time information is more concentrated towards providing and improving transit services, because of the development of GTFS data and which further complemented by passengers carrying smartphones, informations are easily available in mobile formats. The result of their study showed that, real-time data sources played an important role in optimizing overall transit services.

Another study done by Iliopoulou and Kepaptsoglou, (2019) gave detailed information on exploiting Intelligent Transport Systems (ITS) data for public transit planning. They revealed that, exploitation of ITS data in the area of strategic planning has only recently attracted the attention from public transport planners. By enabling "information collection" from various fronts, ITS data was exploited for measuring the "performance of public transport", helped to learn about "ridership and demand pattern". The authors have taken AVL, AFC and APC data in the form of ITS data to study the exploitation of latest technology where they aimed to systematically review literature in the field of optimization of public transport system and the services it provides to its users. When it comes to operational level planning, the authors have also revealed that, there are limitted study in this topic so far, which covers the exploitation of AVL data with an objective of generating optimal vehicle schedules.

According to Ge et al., (2021), AVL and AFC data offers exploitation through service optimization by measuring the reliability and punctuality of the service, which also takes passenger behavior into consideration. According to the authors, information about service optimization can also be available in real time. They have also stated that, passengers upon "informed about delays" in advance, can make changes in their schedules and choose other options. The authors have also noted that, transit services can be dynamically optimized and adapted with the help of these data sources.

Lyu et al., (2021) were interested in improving the service by proposing a framework for matching bus trajectories by identifying bus routes. They wanted to address the issues arising due to "malfunctioning of bus positioning devices", "delayed update of database", "erroneous information about route stored in positioning devices". Therefore, the authors have designed a bus route identification framework pertaining to "the metric of partial Fréchet distance", which does not require any base maps for identification. In order to implement and evaluate the proposed framework, the authors have carried out an experiment in a small city Fuyang and in large city Shenzhen, China. The result of the experiment showed that, the proposed framework not only performed better than the existing method but also provides better interpretable information than the existing method.

Topics yet to be exploited fully:

- Iliopoulou and Kepaptsoglou, (2019) have revealed through their literature review that, there is a research gap at strategic and tactical level of planning with huge scope for exploitation. At strategic level, the authors have reported from the work of Lovrić et al., (2013) that, exploitation of AFC data are yet to be explored in the aarea of passenger preferences reproduction which encompasses strategic decision on “temporal flexibility and sensitivity to fare and service changes“, as well as fiscal policy related decisions. From the work done by Ali et al., (2016), the authors have found that, there is still room for improving “route choice and passenger behavior modeling accuracy“ with the help of “AVL/GPS data“. In general, “strategic analysis using ITS data” can be strengthen by factual representation of demand and supply, which can depict its potential to make long-term decisions more significant and impressive. Therefore, restructuring of public transport networks and definitions of “appropriate problem formulations“ is only possible by further researching and exploring the ways of ITS data exploitation. At tactical level of planning, it was revealed from the work of Ordóñez Medina and Erath, (2013) , that, there is a lack of use of data in “optimization model, to define more appropriate design objectives for passenger-oriented service adjustments”. At operational level of planning, in general, there was also limitted studies “on operational planning decisions using AVL/AFC data”. In particular, there is still room for AVL data driven vehicle scheduling to be exploited. From the work of Desaulniers and Hickman, (2007), the authors revealed that, operational planning problem, related to “vehicle parking and dispatching need to be addressed daily”. Overall, at operational planning stage, “ITS supported optimization method” is still yet to be fully explored. In case of real-time operations, topics hold the potentiality of data driven exploitation are; “Travel time prediction algorithms” for estimation of the effect of “schedule deviations” on running times (Okunieff,1997), “combining optimal control models and prediction methodologies” (Yu and Yang, 2009), “availability of real-time passenger demand data” to manage overcrowding by “improving the performance of control models” (Sánchez-Martínez et al., 2016), favouring simulation environment for real-time implementation of holding strategies (Berrebi et al., 2018). Williumsen (2021) have revealed some topics where uses of Mobile network data and smartphone data are still limited and yet to be exploited. These topics includes: optimization of combined new data and conventional data for the purpose of transport model development and improvement in policy making, developing new forms of model which will stimulate the introduction of new mobility technologies along with it implications on “users, operators, citizens and the environment“, big data in understanding day to day travel demand variability for over a period of time, travel demand monitor for the purpose of knowing changing preferences, to explore the role of big data in understanding uncertainty for better travel demand forecast and better decision making.

Other topics yet to be exploited are:

- “Bus route generation and schedules“ (Yan et al., 2006).
- “Optimal stop spacing“ (Li and Bertini, 2008).
- Reproducing the preferences of diverse passenger “based on AFC, including temporal flexibility and sensitivity to fare and service changes, thus a series of strategic decisions , including fiscal policy“ (Lovrić et al., 2013).
- Studies on design related, studies on problems related to operational-level (Moreira-Matias et al., 2015).
- AVL data based bus route networks adjustment to “improve performance“ (Moreira-Matias et al., 2015).
- To infer trip patterns and “bus network design“ (Liu et al., 2017).
- “Devising robust vehicle schedules based on observed trip times“ (Ceder, 2007; Shen et al., 2016; Shen et al., 2016; Shen et al., 2017).

Transit Evaluation

Lu et al., (2018) have defined “transit evaluation“ as “way to obtain the service feedback from the passengers and operators“. Through transit evaluation, transit operators knows areas needed to improve for better services. These areas of evaluation includes; network reliability, transit accessibility, timetable robustness, and energy-consuming evaluation“. Their aim was to review state-of-art models and approaches to highlight problems arising in a smart urban transit system and the possible solution to these problems. While doing so, they came across transit plan and transit network planning problem as important elements to smart urban transit system. By combining transit plan and transit network planning problem in order to solve the arising problem from smart urban areas, the authors have revealed that, urban transit can be classified into five parts which are connecetd and interconnected with each other. These are; “strategic planning, tactical planning, operational planning, transit evaluation, marketing and policy“. By evaluating transit system based on “service reliability, service accessibility, timetable robustness, and energy consuming“, the authors have found that, in practice, the real service actually differs from idealized service.

Evaluation of public transit network and services based on accessibility

De Oña et al., (2013) have proposed an approach called Structural Equation Model (SEM) to “determine the influence of a series of characteristics describing the quality of the bus transit service on the Overall Service Quality (OSQ)“. This approach will help the transit operators to successfully attract more passenger.

Xu et al., (2017) have also done a study on transit accessibility with an aim to increase the transit efficiency, reducing tranffic congestion by making the public transit attractive and easilly accessible to all income group of population. The authors are stressing the fact that, when poeple can reach their job locations easilly on time through a good transit connection between home and job location, there is no point of bringing/driving the private cars to the central business district area. Therefore, the authors have proposed a measure called Expected Locational Accessibility Index (ELAI) to measure Expected Locational Accessibility(ELA) with the help of expected normalized total reachable job locations (NTRL) to evaluate transit accessibility of commuters between home and job locations. The proposed methods was implemented in an emperical study using empirical transit network of Xiamen City, China. The study result showed that, “the expected locational accessibility concept and measures would considerably assist in evaluating the configuration of transit network, as well as the advancing service quality of transit stations“.

Xu and Yang, (2019) have carried out another transit evaluation study based on transit accessibility while evaluating urban land use plan. The authors have aimed at identifying the performance of urban land use plan, due to the fact that many urban cities are facing with traffic congestion, energy related, health related and environment related issues. In order to solve these problems by providing an environment friendly sustainable solutions, evaluation of transit accessibility becomes crucial. Therefore, the authors assumed transit accessibility as a dependent variable of urban land use in a geographically weighted regression function. The study result showed that, “transit accessibilty could be used as a critical index to determine if each category of urban land parcels has been deployed well all over the city".

Liu et al., (2022) have addressed the transit evaluation issue by measurring "accessibility using high resolution real-time data" such as GTFS. The have used "realizable real-time accessibility based on space–time prisms" (STP) to address 'accessibility unreliability', which is arising inspite of the availability of real-time data and advancement in the technologies.

Ravensbergen et al., (2022) have done literature review based accessibility study with the focus on older adults. In the context of global aging population, the authors focused on "older adults' ability to access the "destinations using public transport". In their review, the authors have categoried their review into three groups; defining the age of older adults, the destinations used by the older adults, how accessibility is measured. The authors have concluded that, literatures on public transit accessibility by older adults are just emerging, with great variation in "study design", the way older adults conceptualize the whole theme of accessibility, different ways of meassuring accesibility, types of destination used by the older adults. Therefore, these situations create difficulty in drawing conclusions "on the status of accessibility by public transport for older adults in a general review".

Evaluation of public transit based on reliability

Chen et al., (2009) have attempted to evaluate the public transit by measuring the transit reliability, in order to bridge the gap between transit agencies perspective and passengers perspective. From the perspectives of passengers, the authors have found their sensitivity towards "reliability at stop level". At the same time, inability of transit agencies in precise determination of bus running at en route stop was also identified by the authors, which led them to propose a reliability measures for buses. To be specific, these measures were proposed with regard to "variability of journey time at routes and headway at stops", which will help the transit agencies "to identify potential service problems at the stop, route, or network level, which may ultimately lead to improved bus services". The assessment of bus reliability was conducted in Beijing, China with the help of three proposed parameters namely; "punctuality index based on routes (PIR)", "evenness index based on stops (EIS)" and "deviation index based on stops (DIS) " to estimate service reliability in route, at the stop, and for all network levels. The analysis was done using AVL and IC card data. The authors have concluded that, these proposed performance parameters was successful in measuring the reliability of public transit in Beijing as the parameters can be adopted from “macro-scale to the micro-scale under a framework with similar metrics“.

Van Oort, (2014) have undertaken an international survey in order “to gain insights into how public transport authorities deal with (improving) service reliability and the design instruments“. The author have observed that, traditional way of transit reliability measurement was done by considering only vehicle’s performance. In other words, transit reliability measurement was done by considering the perspective of transit operators solely, and very less attention was paid to passengers perspective on reliability, which led to wrong outcome of reliability measurement. The survey result showed that, transit operators were concerned only about the vehicle performance by measuring service variability and no focus was given to the impacts of this service variability on passengers. The result of the survey also revealed that, service reliability was not clearly defined by the transit operators, which might create other concers regarding the quality indicators. Therefore, through this study, the author explored the area of the transit evaluation by including other indicator such as; ‘additional travel time’, a sustainable way to represent the service reliability level, which was ignored by traditional indicators.

Diab et al., (2015) through literature review, have also acknowledged that, many studies on reliability are reffering to either from passengers point of view or from transit operators point of view. But there was no work done on reliability by integrating the these two seperate perspectives. Therefore, their study aimed at undertaking public transit evaluation by combining both passengers and transit operators perspective on service reliability. By considering and combining two separate perspectives on transit reliability can actually increase the attractiveness and “effective and valuable policy-relevant information“ can be provided to the “transit planners and decision makers“ for efficient decision making.

Tran et al., (2015) have addressed the transit reliability issue by dealing uncertainty. In their study, the authors were considering factors such as; passenger waiting time, overcrowding are the main reasons for uncertainty especially on high frequency bus routes. The authors were interested to provide for solution against transit unreliability by considering passengers perspective. Therefore, the authors have proposed an approach called dynamic Bayesian networks for real-time multi-criteria decision-making and future states prediction along “with multi-objective optimization“ as a means to finding “appropriate strategies“ for the maintenance of “bus service reliability”. This proposed approach was implemented in the city of Wollongong, Australia, using AVL and AFC data from Gwynneville-Keiraville bus route. The study result showed that, proposed algorithm was facilitating “bus service reliability“ prediction and handling of “multi-criteria decision making for“ real-time information control.

Cats and Loutos, (2016) have undertaken a transit evaluation study to “improve overall level-of-service“ by investigating “the performance of a commonly deployed real-time information generation scheme”. Through this study, the authors were interested in addressing the transit uncertainty by using “real time information” services with a focus to reduce waiting. Moreover, with the help of “real time information”, performance of real-world operational system can be easily evaluated. Therefore, the authors developed a real time information generation scheme and implemented it in Stockholm, Sweden. The study was undertaken by considering the perspectives of passengers as well of transit operators. The result of their study showed that, “real time information has the potential to reduce waiting time uncertainty and facilitate proactive control and operations strategies”.

Redmond et al., (2019) have addressed reliability issue arising from to multi-modal-itinerary planning in their study. In author’s own words, “multi-modal travel itineraries are based on traversing multiple legs using more than one mode of transportation“. Therefore, traveler’s challenge while choosing an itinerary depends on the number of “combinations of legs and modes“. As a solution, the authors have opted for developing a stochastic optimization model by combining both schedule based and un-scheduled based travel modes such as, airline network and taxi transportation respectively. Therefore, the authors have presented with “stochastic network search algorithm“ using historical data from five aiports to find out the flight performances along with the determination of on drive times to and from airport. The airports choosen by the authors was of three different sizes; small, medium, large for the comparison of reliability. The result of their study showed that, although “state-of-the-art traveler information systems can create multi-modal travel itineraries“, itinerary creation was independent of reliability factor “of the chosen transportation services“. Therefore, keeping this information in mind, can be helpful in better decision making by transit operators and travelers.

Kathuria et al., (2020) realized the importance of time value of travelling which is increasing as the economics activities are increasing. At this circumstances, reliability of public transit plays an important role for passengers commuting with public transit. In the process of revewing reliability by the authors, they were focusing mainly on service reliability measures of public transportation, came across to different perceptions about reliability from passengers point of view and of transit operators point of view. The authors have found that, parameters of reliability according to passengers perception are; boarding time, waiting time, alighting time, in vehicle time, seat availability, total travel time, adherence to pre-trip information time, missed connections, transfer time. Whereas, parameters of reliability from transit operators point of view are; headway regularity., total ridership, frequency of break downs, interference of transits units with other vehicle. Due to the differences in perception about transit reliability for different stakeholders, the authors found it very challenging to evaluate public transit reliability. The authors have divided the reliability measures into four quadrants such as; headway regularity, travel time measures, transfer time and waiting time, for reliability evaluation. By taking this theoretical basis, the authors have conducted a case study in Ahmedabad, India using ITS data. The result of their study showed that, “headway regularity and travel time measures used GPS data to understand the travel time variation. Transfer time measure were dependent on the waiting time data at the transfer point and hence could be collected from the smart card data“.

Anderson and Daganyo, (2020) performed a transit reliability study by evaluating transit signal priority. To be specific, the authors designed a special form of transit signal priority, known as; ‘conditional signal priority(CSP)‘ to alleviate bus retardation occuring due to bus bunching and which will eventually increase the transit reliability of both scheduled as well as un-scheduled systems. The study result showed that, “CSP can be used“ for acceleration of buses as well as for reliability improvement in all types of scenarios.

Robustness evaluation of transit services

De-Los-Santos et al., (2012) have measured the robustness of railway transit by proposing an index in case of link failure. In their view, breakdowns in railway links cannot be easily repaired and the “affected trains cannot be overtaken“. In this study, the authors have assumed two types of disruptions which is likely to appear such as, “without-bridging disruptions and with-bridging disruptions“, and based on this assumptiontotal system travel time was measured. The indexes were also defined and divided into two groups for measuring planned and unplanned failures. The result of their study proved that, proposed index was “useful for descriptive measurements“, and increases robustness by assessing and comparing upgraded networks.

Andersson et al., (2013), in their attempt to solve congestion problem in railway transit have proposed a measure based on “Robustness Critical Points (RCP)” to increase the robustness of railway line. The proposed measure was considering train timetable as a main factor for increasing robustness. The proposed measure was implemented to the Swedish railway line and compared against the existing robustness measure. The result of the experimental study showed that, “critical points can easily be used in the timetabling process to identify weaknesses“. Moreover, RCP measure has the advantages of providing meaningful suggestions to the timetable constructors about where and which service needs to be corrected and revised.

Goerigk et al., (2013) have also done a similar study of increasing robustness in railway transit by comparing different line planning approaches and their impact on timetables. By doing so, the authors wanted to evaluate the line planning with an ultimate aim to increase the transit robustness by reducing vulnerability to delays. Therefore, the authors have proposed a framework called LinTim framework, which was used to analyze the existing single line planning steps. This framework was implemented in Liebchen, Athens metro system and in Göttingen. The authors concluded that, “the described framework can be used to measure the apriori robustness of solutions”.

Corman et al., (2014) have also done a railway disturbance evaluation study to measure the robustness of railway schedule using Intelligent Transportation Systems. For their study, the authors have experimented with different timetables for railway schedule to measure the robustness during disturbances. They used an optimization framework to asses the evaluation of different timetables. The authors have concluded that, robustness during disturbances can be increased by extending the railway service with shuttle service for high priority needs for specific origin-destination. However, this extended services requires extra resources and extra management cost.

Marketing and Policy

Brown et al., (2003) have found some special scheme for university students such as, “Un-limited Access Program” for the purpose of providing “fare-free transit service” to faculty members and students. The un-limitted access called “BruinGo”, where the students and staffs does not have to pay for the ride, instead, the university pay on behalf of them to the transit providers. This scheme is a way to motivate the students and faculty staff to commute with public transit and avoid private cars, thereby reducing the demand for car parking zones. Their work served as a best example of good policy and its contribution towards attaining important planning goals. Also, by reforming transportation planning, huge benefits can be produced at lower cost and enhance the growth potential of transportation sector.

The work of Ivaldi and Vibes, (2008) in policy matter deals with market competition where the authors have implemented their framework by “considering competition for passenger traffic on the Cologn-Berlin link, Germany“. Through this study, the authors have provided valuable tools for studying market competition in transport industry and many other valuable instruments for measuring and evaluating cost and benefits ratio arising from new policy implementation which will effect the socio-economic welfare of the popullation.

The work of Li et al., (2009) dealt with the investigation of transit fare under different market structure by proposing a network-based model. Through this work, the authors wanted to address the reliability issue of transit fare considering the perspectives of both transit operators and passengers to better design and plan the transit services. Therefore, the authors investigated and compared the fare fluctuation in the monopoly and oligopoly market by incorporating unreliability factors. According to the authors, “the proposed model provides a useful tool for modeling a competitive transit market and evaluating transit policies at the strategic level“.

Tirachini, (2013) have done an analysis which can be applied to transport policy making for more informed decision making through upgraded fare-collection policy which will reduce travel time of the passengers. The authors showed the superiority of this upgraded version of fare policy over existing other policies which will be helpful in reducing travel time for high demand areas.

Basso and Silva, (2014) have proposed three urban congestion management policies such as; “transit subsidization, car congestion pricing, and dedicated bus lanes” and an analysis was carried out to check “the efficiency of and the substitutability between“ the proposed policies. With the help of this study, the authors wanted to set an example of policy implementation about “reducing transit deficit without affceting the consumer welfare“, which can be achieved through careful planning.

Eliasson and Börjesson, (2014) have worked on the importance of using verifiable principle for railway timetable construction for judicious use of railway investments. With the help of this principle, a meaningful cost benefit analysis can direct the transport economists to make good transport policy by favoring of welfare of the society.

Thomas and Bertolini, (2015) have done their study on transit oriented development with regards to transfer of policy ideas by describing and assessing the process of policy transfer “to Dutch land-use and transportation planners“. Their study stressed in transfering policy and therefore, examined 16 critical success factors with an intention to implement them in Netherlands for transit-oriented development. In particular, the authors “aimed to determine whether these lessons from other cities, if removed from their political, geographical and social contexts, could be seen as learning opportunities during a policy transfer exercise“. Different types of policy transfers and involvement of different stakeholders was studied by the authors too. The authors have realized the importance of involvement from different groups including academics, policy practitioners, politicians, historians for better outcome.

The work of De Palma et al., 2015) can help the policy makers to make certain policies while dealing with crowd, congestion and discomfort in public transit and how does these factors influencing scheduling and pricing. The authors discovered that, there was not yet any crowding measuring consensus developed so far. Therefore, the authors developed a discomfort cost function by considering several discomfort scenarios for the analysis. The authors have concluded that “passengers make decisions on the basis of a given anticipated trip time and congestion level“, which is an important parameter to be considered while making new policies with regards to scheduling and pricing.

The work of Wu et al., (2016) was about providing a tool to the decision makers for quantification of smooth transfer of the passengers by adjusting timetables. This work was accomplished through the investigation of “multi-objective re-synchronizing of bus timetable”.

Lu et al., (2018) have stressed the importance of good marketing strategies and attractive policies to attract more investors and ultimately increasing the performance of transporatation sector and making it more efficient. According to the authors, through implementation of better marketing strategies “such as fare policy, cooperation with other transportation modes, investment strategies and trade-offs”, society and residents can gain higher return.

Lu et al., (2021) have worked on reviewing big data and its applications in urban transit system. Their collected review work on policy applications mainly focused on “land use, financial applications and data privacy with transit big data“. The auhtors have found a direct correlation between several factors which will affect the transit policy planning. These factors includes; home and business locations, travel distances, transit accessibility, residential density, land use planning in transit corridors, retail development, bus rapid transit, mass rail transit, transit investment, population displacement etc.,. And, the research topics involving varying degree of big data application has been seen by the authors for “network accessibility, route choice, activity purpose analysis, and timetable optimization and rescheduling“.

2.2. Literature review on- “*In which areas of transit planning data has been used extensively?“*

Literature review will cover the data utilization in the area of - (A) Planning (B) Assistance (C) Control

2.2.A. Planning

According to Van Oort and Cats, (2015), innovative planning and operation of public transport services can be promoted with big data. This big data is instrumental throughout public transit planning process, where “planning process consists of strategic, tactical, operational and control“. The authors have presented us with cases of Sweden and Netherlands, where offline data were used for strategic and tactical planning in Netherlands, and real-time data were used for operational planning and control in Stockholm, Sweden. By big data, the authors were mainly referring to and using AVL, APC and AFC data. In case of Netherlands, the authors were using big data such as; smart card data for “passenger behaviour estimation and ridership prediction“.

Smart card data enables the authors to construct a predicting model for estimating passenger behaviour. For real-time information, AVL data, complemented with APC and AFC data were used for origin-destination estimation in Stockholm. This study has served as a proof of successful use of data throughout the planning and operational processes in public transit.

Pelletier et al., (2011) have done a literature review on using smart card data in public transit planning. Their focus of data uses in transit was mostly on strategic level, tactical level and operational level of planning. From the review of several literature, the authors has seen how data can assist the transit operators in various stages of planning ranging from long term network planning to calculating demand and supply indicators. The authors have concluded that, "in addition to the fare collection function, smart card systems can be useful for providing data to both planners and researchers".

Shen et al., (2016) have carried out AVL data based "public transit planning and scheduling" work in China in order "to raise the service level and utilization of resources" by improving transit operation in terms of timetable setting and scheduling. Three cities were chosen for their study, where AVL data played an crucial role in successfully crraying out the planned projects. AVL system and "the integrated circuit (IC) card system, a kind of automatic fare collection (AFC) system, were installed before the planning project started". It was less time consuming because of using AVL data, as compared to the previous situations without using data. Installing AVL system in the bus also allowed for vehicle scheduling where prominent "scheduling parameters such as headways (frequencies) and TTs are set automatically based on AVL data". Therefore, with the help of stored AVL data, the authors were able "to measure service reliability" and an elaborated timetable was compiled. In particular, the authors have conducted AVL based vehicle scheduling by handling AVL data, followed by identification of homogeneous running time (HRT) bands, then service reliability measurement, setting the scheduling parameters, ultimately followed by compilation of schedule. in a successive order. Their case study proved the potentials of using data in transit planning by showing the improvement in service level and management level and"resources use in public transport" in those chosen cities.

Stages / parts of transit planning

Desaulniers and Hickman, (2007) and Iliopoulouand Kepaptsoglou, (2019) have classified public transport planning into four stages: *strategic level planning, tactical level planning, operational level planning* and *real-time operations*.

Whereas, Lu et al., (2018) have stated that, “urban transit system is an important part of city transportation” and smart transit system can be only build by considering certain set of sub-problems such as: transit evaluation, traffic design problem, marketing and policy model. Moreover, the authors have also emphasized the need of attractive marketing policy implementation for smart urban transit system which will draw more residents. An adequate evaluating methods also contributes to improve the transit system through evaluation. Therefore, the authors have classified “the urban transit system into these five parts which are connected and interacted with each other, such as: *strategic planning, tactical planning, operational planning, transit evaluation, marketing and policy*“.

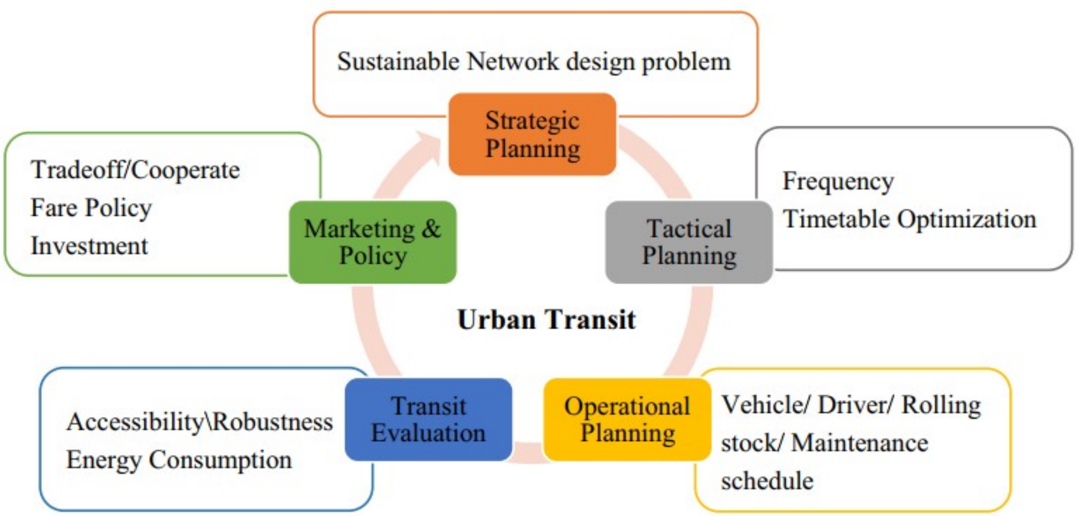

Figure 6: Five parts of the transit system

Source: Lu et al., (2018) Smart Urban Transit Systems: From Integrated Framework to Interdisciplinary Perspectives

Strategic Planning

Basic concepts

According to Ibarra-Rojas et al., (2015), in transit network planning process, strategic planning decisions are long term decisions, that focuses on transit network design by determining types of vehicles, the lines and stop spacing, in order to meet the needs of population's movement.

Whereas, Desaulniers and Hickman, (2007) have stated that, at the strategic level of transit planning, transit routes and networks were designed.

Crainic, (2001) have revealed that, the goal of network design problem is to accomplish the "economic and customer service goal" by finding an optimal resource allocation and its utilization.

Whereas, according to Lu et al., (2018), network design problem aims for total travel time minimization and generalization of travel cost.

Strategic planning with data

Furth et al., (2006) have shown the importance and uses of data in transit planning. Their work deals with strategic planning in general with special emphasis on route and schedule matching in particular, where AVL-APC data played an important role in providing real-time information to the passengers. Not only this, but also the details of other current and potential uses of data in transit sector was presented by the authors. The authors used the archived data and have showed that, for investigating certain situations, archived data can be used as "play-back mode". They have also stated that, transit management and performamce can be improved using "archived AVL-APC" with the highest possibility.

Pelletier et al., (2011) have done a literature based study on uses of smart card data in transit planning and presented us with a scenario which showed the neccessity and the potentials of using data in transit planning. In strategic planning, the authors have found the use of smart card data for better understanding the behavior of the passengers. Few examples of long term planning or strategic planning with smart card data includes, turnover analysis, marketing (Bagchi and White, 2005), itinerary construction (Chu and Chapleau, 2008), creation of future demand matrix (Park and Kim, 2008), fare policy analysis (Utsunomiya et al., 2006), lifespan calculation of smart card user through loyalty and retention improvement (Trépanier and Morency, 2010).

Schmöcker et al., (2013) have also worked on understanding the behavior of passengers by predicting their travel patterns and travel demand with the help of smart card data. By obtaining time series route choice data from smart card of a local bus service provider from Japan, the researchers can see the “actual lines boarded by passengers over longer time periods”. Hence, this historical smart card data played a key role in long term strategic planning by analysing choice behavior of the passengers. Based on this, the transit operators can make better future decisions on setting route, frequency based on the choice behavior of passenger groups.

Moreira-Matias et al., (2015) have worked on “improving the mass transit operations by using AVL-based systems”. Among three main approaches considered by the authors for improving transit operations, one approach was to improve the network design with the help of AVL data. The authors have referred to historical AVL data in order to define the network by “changing the location of bus stops”. They have drawn conclusion on how important role AVL data plays in “building simple frameworks capable of learning from location-based data streams and of providing predictions with low uncertainty”.

Liu et al., (2017) have presented us with another success story of using data in public transit for human mobility pattern mining and bus route network optimization. Their goal was to attract more likely bus riders by detecting and re-planning the faulty and fewer efficient bus routes without seeking for an optimization plan which is globally recognized for feasible applications. The authors have used smart card and GPS data for their study. The proposed method for bus route optimization and transportation mode choice modeling was further evaluated by Beijing Bus Company as the proposed method was implemented using real-world data of this particular bus company. The proposed bus route optimization was successful in attracting more bus riders from taxi by making the “public transportation more attractive to riders”.

Schmöcker et al., (2017) in their book, have emphasized the use of smart card data in strategic planning. According to them, by aggregating smart card data “one can get knowledge and create graphs to illustrate details of travelers demand for strategic planning”. Apart from this, smart card data also has an advantage of tracking individual passenger’s behavior. Moreover, smart card data enables the detection of probabilistic distribution of route choice and can determine the group of traveler based on their route choice preferences and variation in demand.

Qi et al., (2018) have worked on strategic planning which concerned about analyzing and predicting the mobility of the bus passengers based on smart card data and point of interest data. Their work contributed to prove the importance of data by successfully achieving the goal of the study where it has been seen the role and contribution of smart card data in analyzing the mobility pattern of the commuters, and contribution of point of interest data in predicting the mobility of the commuters. Obtaining the knowledge about the regional mobility pattern of the passengers from smart card and point of interest data enables the transit planners, urban planners and managers to better understand and predict the regional mobility pattern which “effectively support evidence-based and forward-looking planning and intelligent transportation management to achieve the healthy development and effective operation of urban cities“.

Iliopoulou and Kepaptsoglou, (2019) have carried out a systematic literature review on optimization of public transport systems and services”, by harnessing AVL, APC and AFC data with a goal to fill the gap between existing work and potential future research work, which will help “relevant implementation and methodological issues” to receive the much needed attention for the accomplishment of their goal. The literature review in the area of strategic planning involving the use the AVL, APC, AFC data was seen for transit assignment and network design. Among the existing work, the authors have found that, there is a lack of studies which investigates design and operational level related problems. But, harnessing ITS data has a huge potential for future research work especially at strategic level of transit planning by adopting data-driven service design and a demand reactive public transport. Concepts like; transit network re-design, inclusion of equity and accessibility in the design process, passengers response to fiscal policy may strengthen the strategic level of planning with the help of ITS data. The authors have concluded that, there is a need to modify “algorithms and models including transit assignment, route design and shortest path” for a successful transition to data driven planning. In terms of strategic planning and demand- oriented improvements, to attain a sustainable mobility model, an immense emphasis must be given “to incorporate ITS data into optimization frameworks” through the adoption of “an appropriate data manipulation strategy”, which should be meaningful and computationally feasible.

Tactical Planning

Basic concept

Lu et al., (2018) defined tactical planning as a “way of transferring the transit design to the transportation service”, where it connects passengers with operators. Therefore, at tactical level of planning, it deals with frequency setting and timetable optimization.

According to Ge et al., (2022), tactical level planning refers “to setting up a timetable with all its specific considerations”. These specific considerations includes; “coordination of different lines” for schedules synchronization, which will facilitate the “possible transfers of passengers”. Only when the timetable has been fulfilled, appropriate plans can be made possible “for transport vehicles (buses, trains), related personnel (e.g. crew, drivers, conductors)”.

Whereas, Ibarra-Rojas et al., (2014) have defined tactical planning as “a stage where, the bus lines are designed and then the timetables (or departure times of all trips) are set. This stage focuses on offering a high level of service for clients: frequency of the lines, waiting times, and short transfers, among others”.

Ibarra-Rojas et al., (2015) have stated that, “tactical planning related decisions” are associated to the Transit Network Problem (TNP). Problems related to transit network deals with “improving the level of service and reducing the operational costs”. Decisions taken at this level of planning are relates to “frequency setting, transit network timetabling, and design of operational strategies”.

Desaulniers and Hickman, (2007) were interested in constructing “frequencies of route” and determination of service schedule, which are the “intermediate steps in the tactical planning process”. The orientation of structuring and improving services towards passengers are the main reason behind inclusion of “these two parts together in the tactical level” planning by the authors.

Tactical planning with data

Pelletier et al., (2011) have presented us with some example of successful uses of smart card data for tactical planning. From literature review, the authors have found a solution to the problem related to service adjustments during weekdays(Utsunomiya et al., 2006), which occurs due to the variations arising in ridership on those days. Smart card data enables to solve this problem route by route by deriving load profiles at each run(Trépanier et al., 2007). With the help of smart card historical data, individual trips can also be planned by deriving load profile of route data on boarding and alighting point from each transaction. Smart card data also provides information about transfer habits

(Hoffman et al., 2009) as the transit operators can reorder “their network geometry and schedules” through reconciliation in order to best match the requirements of traveler”. Not only this, but also the “effects of the application of new policies on ridership (White, 2010)” can be measured with smart card data.

Hadas and Shnaiderman, (2012) have worked on setting optimal frequency by “considering stochastic demand and travel time”, along with “the costs associated with running empty-seats and passenger overload”. This work was successfully carried out using AVL and APC data as AVL enables en route vehicle location tracking, and APC enables collection of passengers’ number “alighted and boarded at each stop”. The authors have concluded that, AVL and APC data are fundamental to the setting of optimal frequency model. Developing models for the estimation of statistically distributed demand and travel times can only be justified depending on whether the “AVL and APC data” are available throughout “transit agencies and operators”.

The work of Sun et al., (2014) has presented us with another case of importance of data and its uses for transit planning. The authors worked on designing a timetable for metro services based on demand using smart card based AFC data. With the help of smart card data, the authors were able to “conduct an in-depth analysis of daily transit demand pattern variation and obtain detailed spatial–temporal service loading profiles”. The authors have also stated the advantages of using “continuous stream of smart card data” in transit planning. These advantages are; “maintaining demand homogeneity by frequent updates”, and “exploring passenger response to a new timetable”.

The work of Nassir et al., (2015) have presented us with the behavioral aspects of the passengers in finding the best route choice using fare card data. The authors have mentioned about the difficulty in obtaining travel strategy choice of the passenger with traditional data collection methods. With fare card data this problem was solved and real-world strategy choice were easily modeled. Fare card data along with the proposed strategy-based transit assignment model and boarding strategies of the passengers helped the authors to successfully analyze the “behavioral aspects of passengers’ path choice”.

The work of Han and Sohn, (2016) was all about promoting the use of smart card in a manner, where the authors are providing us with a situation of successfully obtaining trip purpose, trip chains with the help of smart card data. In their study, an unsupervised model was adopted for the purpose of recognition of hidden activities “behind a smart-card holder's trip chain”. In order to impute the trip purposes from smart card data, the authors applied continuous hidden Markov model in the Seoul metropolitan area. The study result showed that, the proposed continuous hidden Markov model was helpful in inferring the latent activities of smart card data, which further shed light in planning by estimating “the most probable activity sequence for a trip chain, given that the observed features were known”.

Yap et al., (2017) have worked on transfer inference algorithm using smart card data. The authors have observed that, some algorithms have flaws of illogical inferring the transfers and cannot suggest an appropriate option for route choice to the passengers during disruption. However, the proposed robust transfer inference algorithm with the help of smart card data was successful in helping the passengers and the transit planners to plan the route in both disrupted and undisrupted operations. The proposed algorithm was also compared with the state-of-the-practice, and state-of-the-art in Netherlands, using individual AFC transaction dataset and seen to “improve transfer inference” without “compromising on general inference quality”.

Gkiotsalitis and Cats, (2018) have worked on determining the frequencies of bus by setting optimal dispatching headways by considering variability in “demand, travel time and passenger waiting times”. This type of tactical planning related work was crucial for transport operators to address in order to bring changes in passengers demand and traffic conditions such as; traffic jams, accidents, road works, traffic flow etc.,. This proposed method was implemented using GTFS data from Stockholm, where the authors used detailed AVL and APC data from central bus lines 1 and 3. from their work. It has been seen how important historical AVL and APC data were in examining “optimal bus dispatching headways based on” aspects such as “passenger demand coverage, waiting time variability at stop level, operational costs, cost of utilizing extra buses”, through the investigation of the above mentioned aspects and impact of theses aspects on its corresponding objective function.

Kieu et al., (2018) have attempted to study the spatial-behavioral market segmentation with an aim to provide the transit operators with information on “condensed spatial segments” to develop better tactical and operational strategies. Therefore, the authors studied spatial-behavioral passenger segmentation, by proposing “new algorithm named Spatial Affinity Propagation (SAP)” for large-scale spatial-behavioral transit market segmentation" with the help of AFC data from Sydney, Australia. The authors showed the importance of using smart card based AFC data as the aspects such as behavioral features of passengers, relating to frequency of use of smart card data, travel behavior randomness, train usage frequency and "frequency of transfers are mined from Smart card AFC data". The authors have also noted that, in spite of having many advantages, smart card data also have dis-advantages of not able to provide actual home location of the customers.

The work of Zou et al., (2018) was all about "detecting home location and trip purposes of card holders" which was done with the help of smart card data. In Beijing, trip purpose was identified using two types of AFC data, where one type of AFC data is flat fare system, and the other one is distance-based fare system. By using "travel information, such as departure time, activity time duration, and land use type of destination", purpose of a trip was indirectly infered from AFC data. This was done by sorting the departure time from AFC trip records. The authors have acknowledged the potential of using smart card data in the absence of socioeconomic travel data. Detecting trip purposes with the help of smart card data can be later used for developing sound price policies based on types of travellers.

Chen et al., (2019) have dealt with "examining regional mobility patterns of public transit and automobile users based on the smart card and mobile internet data". The authors wanted to study the relationship between travel patterns and land use characteristics, as it can aid urban and transit planners in designing, planning and managing varried "transportation systems and infrastructures". Therefore, the authors have conducted a case study in Chengdu, China by focusing on analysing the mobility pattern on" stop and route levels" and made a comparison of mobility between regions and between the different modes.

Huang et al., (2019) have used GPS and smartphone mobile sensor data to study the transport mode detection of the passengers for the purpose of understanding travel behavior. Transit planners can use this information for later use of scheduling or schedule planning. The author were also interested in learning the characteristics of mobile phone data and data pre-processing "to detect and remove noise/outliers, without mistakenly removing the actual trip points". Through this work, the authors were successful in showing how data is playing a “key role“ in attaining objective of the study, in other words, this type of study is impossible without the uses of data.

The work of Gan et al., (2020) was all about AFC based mobility pattern understanding from spatiotemporal perspectives. The objective of their study was to extract spatiotemporal ridership data to deepen the understanding of urban mobility pattern, to explore “the linkage between ridership patterns and local LCLU“. This study will help the authors to understand the diversity in “ridership patterns“ for the purpose of highlighting the relation between “urban structure“ and its corresponding “urban mobility during a diurnal cycle“. The study was undertaken in Nanjing metropolitan area, China using smart card data from Nanjing Metro Corporation. Collected data was used for cluster analysis to catagorize the Nanjing Metro stations based on diverse diurnal travel patterns. Then, “multinomial logit model“ was used for the estimation of “relationship between local Land Cover/Land Use(LCLU) characteristics and clusters“. The result of their study showed that, there is a difference in mobility pattern between employment-oriented stations and residential-oriented stations. The multinomial logit model showed that, “spatial distribution of different urban functions (LCLU)“ is closely linked to “the spatial distribution of different types of metro stations“.

Tang et al., (2021) have used smart card and GPS data for the purpose of studying “a multi-objective timetable optimization model, to minimize the passenger waiting time and number of departure times“. With the help of smart card and GPS data, the authors have calculated the input variable such as, passenger volume and number of passengers on the bus during boarding and alighting, time-dependent travel time between two adjoining stations, and bus dwell time of each bus stop. The authors have fused smart card and GPS data with the bi-objective optimization model to minimize the passenger waiting time and to reduce the bus departure times. This experiment was carried out in Beijing, China, where the experimental result was successful in reducing passenger waiting time and number of departure times with the proposed optimization model.

Causa et al., (2023) have dealt with urban air mobility through strategic and tactical path planning using GNSS and GIS. With the help of unmanned aerial vehicle (UAV), the authors aimed “to define the best route from a start to a goal point“. In tactical path planning, the authors aimed “to tackle with unpredicted events during flight“ with the help of GNSS data. The result of their study proved the effectiveness of UAV to ensure the “safety and effectiveness of the path“ through tactical path planning. Data along with “multi-objective and multi-constraint planner algorithms“ also helps in reducing computational burden of tactical planner.

Operational Planning

Basic concepts and terminologies

According to Lu et al., (2018), the focus of “operational planning problems” is on the provision of “proposed service at minimum cost”. Problems related to service provision at minimum cost includes “vehicle scheduling, driver scheduling, rolling stock schedule, and maintenance schedule”.

Ibarra-Rojas et al., (2015) have described that, short-term decisions of operational planning focuses on the minimization of cost through judicious utilization of resources such as; vehicles, fuel and drivers. Operational planning problems of resource utilization is then further splitted into “(i) Vehicle Scheduling Problem (VSP), (ii) Driver Scheduling Problem (DSP), and (iii) Driver Rostering Problem (DRP)”. According ot the authors, strategical and tactical planning decisions are used as input by the transit agencies for the purpose of best resource use determination to save the cost.

Ceder, (2007) in his book, have decomposed the operational planning process into “four basic activities such as; (1) network route design, (2) timetable development, (3) vehicle scheduling, and (4) crew scheduling”. These activities are usually perform in an arranged manner.

According to Desaulniers and Hickman, (2007), in a given situation where timetabled trips are operating, resources set of buses and drivers are available and distributed evenly “among the transit agency depots”, in such situations, the aim of “operational planning phase” is the construction of “vehicle and crew schedules” for the purpose of total cost minimization by considering all the “operational constraints and work regulations”.

Operational planning - vehicle scheduling

Basic concepts and terminologies

According to Shen et al., (2016), with the given previously aggregated timetables, “the Vehicle Scheduling Problem (VSP)” covered fleet allocation of vehicles with an aim to efficiently carry out all the timetabled trips. With the help of an efficient schedule, transport operators can save a significant amount of operating cost along with saving other property.

Trip, Trip time, Pull-out, Pull-in. Idle time, Deadhead, Depot-return

Shen et al., (2016) have defined “a trip” as a “task unit” that is being carried out by a vehicle. It is also called as “service trip”. Every trip has a start time and end time at its departure and arrival points respectively and the duration of a trip is called trip time. With pull-out from a depot, the work of a vehicle starts everyday and at a pull-in to the depots, it ends. The waiting time of the vehicle is called “idle time”. If there is an “empty movement between any two consecutive trips”, it is called deadhead. Temporary return of vehicle to a depot upon imposition due to the large gaps between consecutive vehicles is called “depot-return”.

Huisman et al., (2004) have defined the objectives of vehicle scheduling problem as minimization of total cost which includes fixed cost and variable cost. Variable costs are the operating costs which includes idle times, trip times and deadheads under certain constraints such as: (1) precisely one vehicle is assigned with each trip (2) first and last trip of every vehicle has to start at a pull-out from a depot and ends at a pull-in to the same depot respectively (3) there must be a time feasibility between the connection of any two consecutive trips in a vehicle.

Operational planning of vehicle scheduling with data

The work of Shen et al., (2016) concerned about enhancing the robustness of vehicle schedule by considering delay propagation. Therefore, they studied static vehicle scheduling problem along with stochastic trip times and scheduled departure times. The trip times used in this study was taken from historical AVL data. The authors have proposed and implemented probabilistic vehicle scheduling model, using the trip time distribution of AVL data, collected from Route 4 of Haikou Bus (HKB4), China.

Shen et al., (2017) have proposed a new vehicle schedule model with the objective “to compile schedules being able to achieve the expected on-time performance with minimal total cost (including the fleet size and operating cost)”. The proposed model uses the values of trip which varies within certain time range instead of taking fixed time value and thereby making it easier for the schedulers to set suitable scheduled trip times. With the help of AVL system, time range of each trip time distribution was obtained by the authors for the study. Their proposed model was implemented in the real world scenario of China, “using the AVL data derived from Route 4 of Haikou Bus”. The result of the study showed that, vehicle schedule problem with variable time was not only successful in obtaining the objectives but also “produce better result than the deterministic vehicle schedule problem”.

Gkiotsalitis, (2019) have also focused on reducing operational cost and passenger waiting time at stops by generating and integrating alternative lining options into resource allocation and frequency setting. This work was undertaken with the help of GTFS data and AFC data from the Hague city. The proposed approach was superior in allocating buses by offering larger benefits, which can help the transit operators and bus service planners in making better decisions with certain preferences.

The work of Shang et al., (2019) dealt with addressing “a multidepot vehicle scheduling problem considering the passenger waiting cost”. For this purpose, the authors have conducted a case study in Beijing, China, by proposing “an extended vehicle scheduling model together with the deficit function (DF)” on three important bus lines, using bus smart card data along with direct observation and video tape recording. The timetable synchronization optimization along with vehicle scheduling was successful in reducing “feet size of bus operators and the waiting cost of passengers“.

Lai et al., (2022) have used historical GPS data for the purpose of flexible vehicle scheduling and route optimization, in order to improve the “flexibility and convenience in urban public transport systems”. Their work served as an important example in showing the crucial role of data in executing certain objectives of transit planning. Not only the use of data but the authors have also emphasized on perfectly integrating vehicle and passenger data for “overall optimization of transit system”. The authors have conducted an experimental study in China using Xiamen Taxi Dataset for the simulation. The result of the study showed that, the proposed flexible transit system was successful in reducing passenger waiting time in comparison to existing fixed transit system.

Operational planning of crew scheduling

Basic concepts

According to Shen et al., (2013), “The crew scheduling problem in public transport (e.g. train, bus and tram) involves finding the most efficient way of assigning notional crews to the daily operations of a fleet of vehicles”.
Beasley and Cao, (1996) have defined crew scheduling as basic problem of assignment schedule where each crew is assigned with a fixed task along with fixed start and finish times so that none of the crew “does not exceed a limit on the total time it can spend working”.
Cendales, (2019) have defined crew scheduling which is “also called DSP (Driver Scheduling Problem) or duty scheduling”, where drivers’ duties are defined by cost minimization and by “considering all labor regulations“. The output of duty scheduling are starting working time and end working times of drivers by considering meal breaks and pauses. According to the author, duty scheduling is based on vehicle scheduling problem. Optimization of duty scheduling is the key to increase the proft of transit operators as trips cannot be performed without drivers and drivers “are the main cost for operators“. The author have performed the optimization process by taking into account “total driving time per day, continuous driving time, pauses and break meals” as constraints.
According to Heil et al., (2020) “at strategic and tactical level with a time horizon longer than one year”, the task of crew planners is to deal “with the long-term availability and capabilities of crew members” in addition to manage “the location and the capacities of crew depots (crew management)”. Whereas, at operational level, the task of crew planners is to build “work schedules for crew members to operate a planned timetable”.

Operational planning of crew scheduling with data

Cendales, (2019)’s thesis work was about measuring the robustness of bus crew schedule, where the author have used AVL and GPS data along with historical data from Frihamnen depot, Stockholm. “Robustness and planning parameters (slack%, average schedule deviation and passengers) of crew bus schedules” was considered for simulation created by Excel tool. Result of the study showed that, collected data along with the simulation tool is desired by the transit operators for the analysis of crew robustness.

Operational planning with data – Integrated vehicle and crew scheduling

After searching in all the database such as, ScienceDirect, Google scholar, Elsvier, Research Gate, IEEE, no literature has been found in integrated vehicle and crew scheduling problem, where the uses of data such as; AVL, APC, AFC, GPS, GFTS, ITS has been mentioned. In other words, data has not been used in operational planning of integrated vehicle and crew scheduling.

Operational planning with data - Maintenance scheduling

The work of Xie et al., (2020) was favoring the importance of data-driven study for predictive maintenance of railway tracks. From the literature review based study, the authors have found that data from cameras and sensors upon combining with machine learning techniques can perfectly avert needless replacement of track components, improve the safety, save costs, and making the railway service available and efficient. Through review of literature, the authors have found that, several “maintenance strategies such as; corrective, preventive, condition-based, and predictive maintenance” has been used so far for saving maintenance cost, among which predictive maintenance along with machine learning method found most popular. The authors have found three main issues in the literature where data-driven model was applied in order to find the solution. These issues were: “rail geometry irregularity, rail head defect, and missing rail component detection”. The authors have also concluded that, model selection criteria was solely depending upon types of data being collected.

Mohammadi et al., (2021) have used geospatial data along with other data such as; track geometry measurements, rail defects, service failure, maintenance frequency, to study “the problem of rail renewal and maintenance planning” in US. Federal Railroad Administration of US was responsible for providing the data for this study. These geopatial data provides exact location and time in the network, which was helpful for carying out the maintenance planning and scheduling. The authors have carried out this work in order “to determine the type and time of maintenance tasks needed to be performed in each segment of a network while meeting quality criteria, resources and time limitations”. They have proposed a mixed integer linear programming model to acquire better understanding for “rail maintenance and renewal planning”. This data-driven study result was successful in providing robust strategies for railway maintenance during uncertain situations and can help the decision makers “in selecting optimal strategies with respect to expected rail quality and budget and resources availability”.

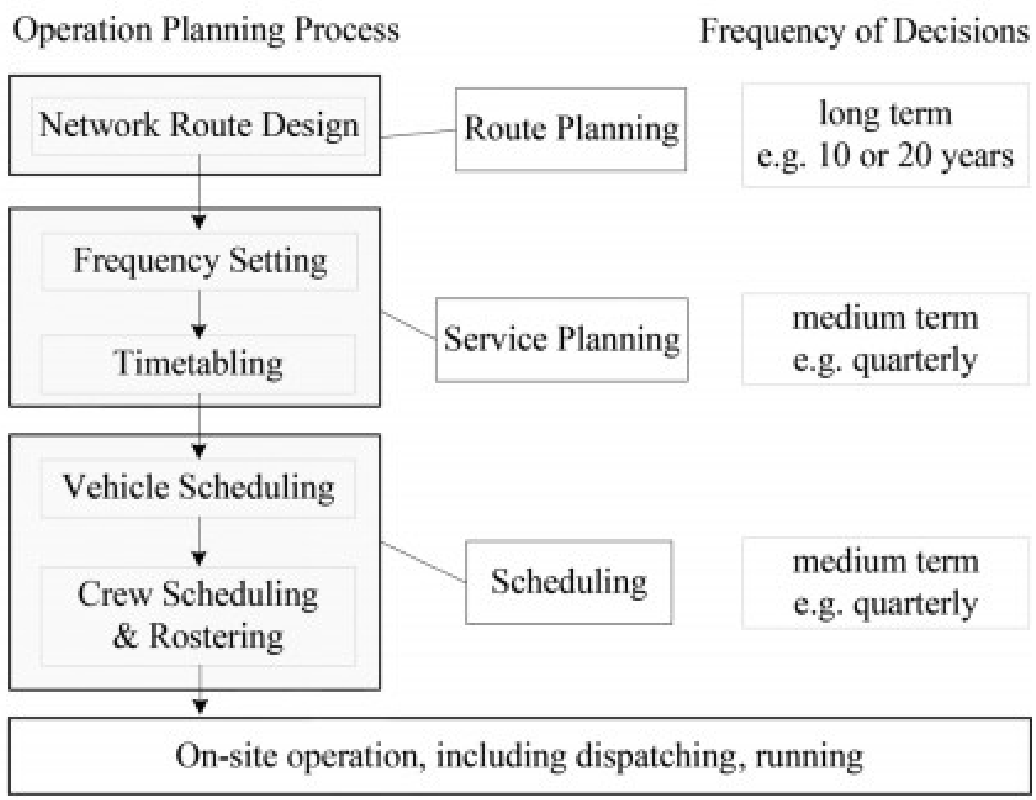

Figure 7: Public transit operation planning process

Source: Ceder, (2007) Public transit planning and operation: Theory, modeling, and practice

Operation optimization

Basic concept

Lu et al., (2021) have stated that, "operation optimization contains operation plan, schedule and the fare scheme".

According to Gkiotsalitis, (2023), operation optimization "of public transport systems" comprises several levels, including "strategic planning and moving to tactical planning and near real-time (operational) control ".

When it comes to using algorithm for operation optimization, the most popular ones are: Genetic Algorithm(GA) (Yang et al., 2017; Shi et al., 2016; Zhen and Jing, 2016), Branch-and-Bound GA (Albrecht, 2009), Simulation based GA (Yang et al., 2016), nondominated sorting genetic based algorithm (NSGA-II) and hybrid artificial bee colony algorithm (Chen et al., 2015) and Tabu search (Cao et al., 2016).

Operation optimization with data

Basak et al., (2019) have used GTFS and sensor data for optimization of public transit schedule. However, the authors have emphasized that, not only the GTFS and sensor data, but their study can also be carried out by using AVL and APC data as AVL data set worked in the form of “GTFS realtime feed“ in the form of “automatic vehicle locator (AVL)“ , which can report the bus location in real-time, was installed. Whereas, datasets about “passenger counts and departure/arrival times at stops” was recorded by APC system. But great emphasis was given to data cleaning process before their uses as “raw transit dataset“ were full of inaccurate samples with some dupplicate and missing contents. Cleaned processed data were later used for timetable optimization. The authors have used empirical cumulative distribution function (CDF), a greedy algorithm for timetable optimization as well as other heuristic algorithms such as; “Genetic algorithm(GA)“ and “Particle swarm optimization (PSO)“. The study result showed that, implementation of PSO have strengthen the bus on-time performance along with less execution time in comparison to heuristic methods.

Cyril et al.,(2019) have emphasized the need of using data through their study. The authors have worked on “performance optimization of public transport” by proposing “integrating the analytical hierarchy process (AHP) and goal programming (GP)“, which considers the perceptions of both operators and users. The case study was carried out in Kerala, India, using audit data from Kerala State Road Transport Corporation and Economic Review of kerala. This study did not use data such as: ITS, GFTS, AVL, APC, AFC, GPS, GIS etc., for performance optimization. However, the result of optimal solution and sensitivity analysis suggests that, through the introduction of technological advancements namely, intelligent transport system (ITS), electronic ticketing machines (ETMs), global positioning system (GPS)-based ETM, automated ticketing systems, passenger information system and computerized crew scheduling, the performance of public transport can be improved from the operators’ point of view.

Shatnawi et al., (2020) have combined GIS data along with “Particle Swarm Optimization (PSO) and Genetic Algorithm (GA)“ for the purpose of modelling “the location of bus stops in Amman city in Jordan“ by finding serviceability of stops and optimal travel time with an aim to promote the use of public transit to private car users. GIS data along with produced Geo-database was used as input data to “Particle Swarm Optimization (PSO) and Genetic Algorithm (GA)“ for the purpose of optimization “to generate two optimized bus stops locations maps“. The authors have concluded and recommended that, “GIS modelling“ has the advantages of reducing travel time for the streets with the unnecessary “bus stops located at irregular intervals“. Travel time reduction occured due to the “increase of service areas of the bus stop and the increase for public transportation demand“. Therefore, the authors have recommended that, the decision-makers should use the developed models for the enhancement of “the existing public transportation network and design of the expansion of the network“.

Lv et al., (2021) have carried out a bus operation optimization work using Smart card AFC data in Beijing, China. This study was initiated from the authors personal daily experience by taking line number 87, which was the combination of regular buses and shuttle buses in the morning peak time. The authors have “proposed a bus operation plan model considering the shuttle bus based on predicted data (PSBO)”, which dealt with solving the prediction of passenger flow, prediction of bus travel time, and the optimization of bus operation. Smart card data along with bus station data helps in providing passenger number from each station. Then data on bus travel time was obtained from “different bus stations using long short term memory (LSTM)”. Then passenger number and bus travel data were combined and used as input to work out bus operation optimization. Through this, authors have showed the importance of data in solving several relevant operation optimization problems. The total dataset used for this study “contains around 150 million transactions over 7000 bus stops from December1, 2018 to December 31, 2018”. The fields of extracted dataset contains: passengers’ smart card data, time related to passengers got on and off the bus, bus line taken by the passengers, the station where passengers got on and off the bus.

Shi et al., (2021), with the help of “multi−source data including bus smart card transaction data, bus location data and static attribute data of bus network“, worked on bus route optimization by proposing an evaluation indicator system for bus service. The authors have found that, earlier studies on optimization were focusing mostly on certain indicators and there was lack of “post optimization evaluation method“ that deals with all kinds of bus adaptation. Therefore, the authors have presented an extensive evaluation method which covers “the bus route adjustment, station adjustment and dispatching optimization“, along with “post evaluation method of bus route optimization“ with an aim to mirror the effect of multi−dimensional optimization method on bus routes adjustment scheme. Smart card data helps to achieve this goal as, it “includes the information of passengers’ smart card number, card type, transaction time, transaction type, route number, vehicle number, boarding and alighting station number“. The information about boarding and alighting of passengers can be obtained from ground trnsactin data. Information regarding vehicle number, route number, speed, longitude and latitude coordinates can be obtained derived from bus location data. This bus location data helped in analysing information regarding arrival time, speed, mileage and operation time of the bus. Travel chain data helps in analysing “travel time related information“ by recording each travel stage of passengers. The authors have concluded that, bus operation data and data on individual passenger’s travel attributes in the form of big data helps in the development of “a weight coefficient calculation model“ which can evaluate the bus route optimization effects systematically.

2.2.B. Assistance (during plan execution)

Pangilinan et al., (2008) have carried out a pilot project on bus supervision deployment strateggy in Chicago. Their aim was to improve bus service reliability using AVL data. Archived AVL data used to develop a simulation model for route which assists in predicting “the effects on service reliability when real-time AVL information used in bus supervision“.

Cats et al., (2011) have used data from Stockholm metro network for their study and showed how data can assist in the form of real-time information in transit path-choices and saving potential times, and improves transfer coordination.

Nielsen et al., (2012) conducted a study on disruption management of railway rolling stock “within a certain time horizon from the time of rescheduling“, using real time data collected from Netherlands Railways (NS). The result of their study showed that, real time data assists in real time decision making as “real-time rescheduling needs to deal with uncertain information as the environment evolves (Yu and Qui, 2004)“.

Borole et al., (2013) have presented through their case study, where, the authors have proved the advantages of GPS data for the purpose of assistance in transit trip planning through the incorporation of “the delays into the transit network“. This is done in “real-time to minimize the gap“ between trips. Therefore, the authors have combined real time GPS data with “K- shortest path algorithm“ for the purpose of computation of “multi-criterion optimal plans“ by considering “delays in all routes“, where “arterial ones and frequently missing GPS signals“ was included. With the help of GPS data, the authors were successful in producing a feasible, accurate and time responsive plans for better trip planning.

Other studies in assistance category also cover uses of big data that can help agencies evaluate their services and identify potential improvements. A study conducted by Faizrahnemoon et al., (2015) have shown how data can assists in developing a model for the purpose of “proactively predict the effect of some control actions (e.g., removing a bus stop, increasing the frequency of a train, adding an extra bus line) in an existing public transportation network”. Therefore, the authors have stated the importance of constructing a model with directly obtained real time transportation data in near real time. By using real time data such as; “GPS, mobile phone data, loop detectors, cameras, etc”, the authors were motivated in developing “mathematical models” for various modes of transport such as; buses, trains, taxis, bikes, walking and private cars.

The work of Bruglieri et al., (2015) concerned about designing and developing “real time mobility information system“, where the author have used real time GTFS data, which assists in managing “unexpected events, delays and service disruptions concerning public transportation in the city of Milan“. Their study also showed the advantages of using data by “exploiting the information on the status of urban mobility and on the location of citizens, commuters and tourists“, which can assist in rescheduling of their movements in real-times. Through this study, the authors have promoted MOTUS((MObility and Tourism in Urban Scenarios) travel planner against other famous travel planners such as Google Transit and Moovit, and have found that, MOTUS has the advantages of proposing alternative paths for the purpose of bypassing disruption as compared to other travel planners.

The work of Li et al., (2016) was concerned about using real traffic GPS data to predict traffic congestion. “Floating car data are the real time traffic GPS data” used by the authors for their study. According to the authors, by predicting congestion, an adequate traffic management measure can be easily adopted by traffic departments. Congestion prediction also helps the users to select wiser trip strategies for “route and departure time selection“. Therefore, the authors have paired real time GPS traffic data with adaptive k-means cluster algorithm, which was successfully predicting the congestion with high accuracy and robustness.

Zhao et al., (2016) have proposed a tool called “efficient spatiotemporal data retrieval method based on Impala (ESTRI)” for the enhancement of efficient sharing of mass data. Here, the data/using data has been seen to execute the plans of transit planners through easy access and retrieval of data itself because of finite “computational capabilities” and fragmental “data organization mechanisms”. The proposed method was implemented in Taiyuan city, where ESTRI was compared against MongoDB. The experimental result showed that, effective retrieval for large volume data was obtained through ESTRI as compared to MongoDB. In fact, ESTRI performed around seven times greater than MongoDB. Therefore, the experimental result proved the effectiveness of ESTRI in the retrieval of “mass spatio-temporal trajectory data“.

Liebig et al., (2017) have proposed a trip planner, where both historical and real-time data assists in “dynamic route planning with real-time“ prediction. The proposed trip planner was successfully implemented in the city of Dublin, Irland, where the incorporation of “ future traffic hazards in routing“ was carried out through the computation of “future traffic conditions“ by a sensor readings based Spatio-Temporal Random Field.

The work of de Winter et al., (2019) puts emphasis on sharing and using data from sensors, cameras and telematics applications as remedial precaution against road traffic accidents. During the execution of plan, if events occurred such as; traffic congestion, accidents, delays or pandemic, strategies are adopted or supports are provided in order to smoothen the service. The authors were favoring the “development and deployment of automated cars“ by supporting the sharing of data and its uses to reduce traffic accidents.

Williams, (2019) have presented another example where data has been seen to assists the transit planners in reducing traffic congestions as the author have explored the possibility of managing traffic congestion in developing countries with the help of “advanced traveler information system (ATIS)”. The traffic information presented by ATIS allowed the drivers and commuters to choose their destination, mode of transport, route and departure time. Thus, “an optimal use of the existing road infrastructure” was made possible through the implementation of ATIS.

Guo et al., (2020) have paired “real time taxi GPS vehicle trajectory data” with machine learning technique such as; “3-Dimensional Convolutional Networks (C3D) with Convolutional Neuron Networks (CNNs) and Recurrent Neuron Networks (RNNs), called CRC3D“ with an aim to assists forecast of traffic in urban areas. The proposed mapping technique was found successful in yielding satisfactory results.

When it comes to provide better assistance during pandemic such as COVID-19, the work of Pozo et al., (2020) have served its purpose successfully. In their work, it has been seen that, smart card can assists the transit planners to response in an emergency situation like COVID-19. It was because, with smart card, it was easy analyze percentage of ridership during peak pandemic, along with types of transport mode used by the travelers. The authors have used smart card data from Madrid, Spain for their study which covered all stages of COVID-19 pandemic. The authors have also concluded that, the findings of their study can contribute “to extend the global knowledge about COVID-19 impact on transport and comparing results with other cities worldwide”.

Sostaric et al., (2021) through their study, have showed that how data can assist various stakeholders in transit planning and sustainable urban mobility policy creation. They used GPS data which were collected from mobile telecommunication networks “in the form of big data“ and was applied to a pilot project conducted in Croatia. The data-driven study was aimed to assist “sustainable urban mobility assessment and improvement“.

2.2.C. Control

Zolfaghari et al., (2004) have developed a holding strategy for real-time operation-control characterized by high frequency, where real-time information was obtained from GPS based AVL system. This GPS based AVL system provides the “locations of buses along a specified route“. The control strategy was developed with an objective to reduce “the waiting time of passengers at all stops on that route“ with an option to adopt situation based on high bus occupancy or low bus occupancy. This control strategy was successfully carried out by combining real-time data with a mathematical model such as; “heuristic algorithm based on simulated annealing“, which was suitable to solve problems “for all buses on the route“ simultaneously. Data plays an important role in public transit, and the authors have also stated that, “design of computer-based real-time decision support systems for public transit“ was accelerated only due to the advent of automatic vehicle location(AVL) and global positioning systems(GPS) like technologies.

The work of Yu and Yang, (2009) revealed the importance of using new information technologies such as; AVL, APC to enhance the real-time control operation. The authors have developed a “dynamic holding strategy” with the help of support vector machine (SVM) in order to forecast the departure times of early bus from the following stop onwards, where SVM uses historical data for time series forecasting. Although the authors have successfully carried out the dynamic holding strategy based on historical data, however, they emphasized on need and importance of using real-time data for real-time control strategies for Advanced Public Transport System (APTS) as APTS is “designed to enhance the systems ability to remedy unexpected problems as they occur”. Additionally, the authors have concluded that, in extension to growing applications of emerging information technologies of AVL, AVI and APC, the control strategy can be further extended to on-time performance consideration of early buses at several stops concurrently in the form of “dynamic multi-stop holding strategy”.

Chen et al., (2013), investigates the control strategy where “a group of buses“ was hold either “at a single or multiple control point(s)“ with the help of real-time AVL and APC data. The authors undertook this study as a remedy against solving bus bunching problem by proposing holding strategy which considered the passenger boarding during holding. The proposed strategy was successfully implemented at Chicago. When real-time data is combined with an online algorithm with fast computational skill, where AVL and APC data were continuously updating, enables the continuous service monitoring and assessment of transit due to the availability of real-time bus data. This investigative study revealed the intention of holding strategy which was through the insertion of “slack into bus schedule“ without helping a bus which is “falling behind schedule“. Consequently, if the goal of transit providers is to obtain “better performance with an integrated model“, recommenedation was made to combine holding strategy with other strategies such as; stop skipping, adjusting bus cruising speed.

The work of Moreira-Matias et al., (2015) was about “improving both planning and control on public road transportation companies using automatic vehicle location (AVL) data”. The main aim behind their study among others was, to reduce the manpower amidst the AVL based “automatic control strategies” revision, where, the focus of traditional planning will be shifted “to an autonomous data-driven real-time control”. The authors have also stated that, AVL data facilitates the automatic operation and control strategies allows to prevent the “instability and unreliability in the network while the buses are” in operation.

Nesheli and Ceder, (2015) have worked on “implementing proper control action“ to prevent the missed transfer with the help of real-time fare card data. This case study was conducted in Auckland, New Zealand, with an objective to improve reliability through “performance measurements of a“ public transport system by “offering direct transfers on multi-legged trips“. The authors were successful in implementing control action with the help of “Auckland Integrated Fares System, OD pairs and“ public transit “demand data“ by minimizing “passenger travel time“.

The work of Andres and Nair, (2017) are about creating “data-driven predictive-control approach to reduce bus bunching“ with the help of GPS based AVL data. The authors have implemented their work in Dublin, to reduce bus bunching problem by combining “bunching prediction and corrective control“. The author were successfully applying the proposed predictive control to convince the transit service providers and customers over other “classical approaches“.

Luo et al., (2017) have used AVL data and IC smart card data along with proposed “dynamic control framework“ to manage “the bus bunching problem with capacity constraints“. The authors have claimed that, the proposed control framework can simultaneously improve headway adherence and maintain passenger load stability within the capacity of bus including similar and varried situations. The proposed framework was implemented at Chengdu, China. The result of the study showed that, the proposed control framework was successfull in reducing bus bunching by considering varried requirements “at bus stations and the bus capacity constraints“ which led to cost minimization of total passengers at the same time keeping the buses inside “their capacity constraints“.

Asgharzadeh and Shafahi, (2017) have proposed “real-time bus holding control strategy“ to “minimizes passenger waiting time“. The proposed strategy was implemented and evaluated with the help of real-time AVL data and static APC data. The proposed strategy was tested on a simulated route for bus in the city of Mashad, Iran. The proposed strategy was composed of “a mathematical model aimed to minimize the passenger waiting time“, where the mathematical model was solved by engaging “a heuristic solving method based on the gradient descent algorithm“. The authors have found that, the proposed strategy was able to reduce the “total passenger waiting time by 8.65% during a 150-min simulation run“.

Berrebi et al., (2018) have carried out performance comparison study of three different control strategies in order to reduce bus bunching. With the help of AVL data, APC data and loop detector data along with historical data were collected from “Portland Oregon Regional Transportation Archive Listing“ for the comparison of “performance of closed-form bus holding methods“. For their study, the authors have compared methods such as; scheduled-based, headway-based, prediction-based, on the basis of “headway instability and mean holding time“. The authors have found that, the prediction based holding methods was superior than two other holding methods by achieving “the best compromise between headway regularity and holding time on a wide range of desired trade-offs“.

Gavrilidou and Cats, (2018) with the help of AFC and APC data, have developed real-time transfer synchronization controllers in order to study “the performance of transfer synchronization control“. Use of real-time data allowed the implementation of “real-time control strategies“ especially in “high frequency services“. The result of the study showed that, maximum “time saving per transferring passenger“ was obtained only when the level of demand was low, allowing the controller to maintain synchronization more periodically.

The work of Iliopoulou and Kepaptsoglou, (2019) was a literature review based study, where the authors have found that, AVL, APC, and AFC data are in the core of ITS technology for the execution of “real-time control strategies“.

Malucelli and Tresoldi, (2019) have used “automated vehicle monitoring (AVM) systems and mobile telecommunication devices“ to manage disruption in public transit in Milan. A real-time vehicle and driver scheduling based Control strategy was used by the authors to solve the problem regarding “delays and small disruptions“. Therefore, the authors have proposed “a simulation based optimization system“ which helps in coping with delays and small disruptions“, and was successful in reducing irregularity in the services and operational costs minimization.

Wang et al., (2021), in their work, have proposed “proactive control strategy“ with an aim to increase service reliability with the help of real-time IC card data and GPS data. Through this strategy, the authors were intending future probable disturbance prediction through “monitoring and inferring the system operation“, and prevention of service irregularity by taking measures ahead. The proposed control strategy was implemented and “tested with the data of bus route No. 52 in Beijing city of China“, and have successfully yielded satisfactory result than other control strategy.

Xue, et al., (2022) have combined “ITS(Intelligent Transportation System)“ and “single-line real-time control strategy“ with an aim “to ensure the balanced operation of buses on the route“, through the avoidance of bus bunching incident and thereby reduction in bus delays.

In order to acheive their goal, three aspects of bus bunching were considered by the authors. These aspects are; “bus cruising time, dwell time, and bus load rate“ were incorporated into the proposed control strategy. The proposed strategy was implemented at “No. 245 bus line in Nanchang City, China“ to solve the frequent occurence of bus bunching problem. The result of the study showed that, the proposed real-time control strategy can successfully avoid the occurence of bus bunching and a balanced operation of bus line can be effectively acheived.

Yap et al., (2022) have adopted optimization based control strategy using AFC data in an attempt to quantify disruption impact propagation “due to the multi-level nature of public transport“ network. Their proposed framework was successfully implemented to a “multi-level case study network in the Netherlands“. The authors have recommened that, consideration about effects of control strategies on train disruption management should be made by “train network managers“, otherwise impacts of disruption is likely to be combined with urban PT network. It was because, in practice, monitor and control of trips and passengers are not done by operators in whole part of the network, but only to a assigned part of the network. Therefore, quantification and control of disruption impacts on overall public transport network is essential than to focus on particular level where disruption happened.

2.2.D. Maintenance

According to UITP (Union Internationale des Transports Publics), (26.02.2020) “The purpose of maintenance is to keep assets performing their prescribed functions at the optimum cost”. Maintenance is the main element of broad asset management strategy which beside “design, acquisition, operation and de-commissioning”. Traditional way of executing public transport asset management was through combined actions of "corrective and preventive" measures. But nowadays, with the help of data "smarter asset management is possible" as the system is self learning, allowing self diagnosis and prediction of failures followed by precise maintenance activation.

From UITP perspectives, "assets here can mean any tool or public transport system such as a ticket machine, bus, tram, rail tracks etc". UITP have classified maintenance into four categories; interval-based preventive maintenance, conditional preventive maintenance, condition-based maintenance (CBM), predictive maintenance. According to UITP, zero failure public transport system can be achieved with "CBM and predictive maintenance" as it can predict the physical status of assets and plan the better maintenance duty accordingly. Data played an integral part to make predictive maintenance easier, where data was handled "through four steps: data measurement, data transmission, data analysis and based on that, prognosis with the use of algorithms and statistical data". The use of processed data by different stakeholders for the purpose of decision making is only possible once the data has been "captured and transmitted to the cloud" for analysis.

The work of Maktoubian, (2017) was to solve transportation problem related to "rising maintenance cost" in real-time through "Condition-Based Maintenance (CBM)". The authors have found that, managing asset traditionally with planned maintenance(PM) is not possible as different sources continuously generating high volume of maintenance data everyday. Therefore, the author have proposed "a knowledge-based approach of CBM using streaming big data analysis (SBDA)" for the purpose of solving challenges arising from "real-time big data management, storage and computation". The proposed knowledge based approach is also capable of solving problems arising during "predictive data analysis in CBM systems". The CBM also has the advantages of managing on the spot system failure, when it (CBM) is combined with technologies such as; "Apache Kafka, Mesos, Spark streaming and HDFS".

Davari et al., (2021) did a survey based predictive maintenance study in railway industry. This predictive maintenance study was based on machine learning and deep learning, where "events like temporal behavior and fault events, anomaly detection in time-series" can be monitored and industrial equipment can be logged in with the help of "sensors installed in different parts of an industrial plant". The result of the survey revealed that, anomaly detection can be enabled by sensors and by detecting anomaly, predictive maintenance for downtime can be reduced.

ITF, (2021), reported the “potential of data-driven approaches” in improving the maintenance of transport infrastructure. According to the International Transport Forum (ITF), data driven transport infrastructure maintenance gained momentum subsequently due to “four simultaneous technological innovations“. These are; “development of digital technologies”, “computing technologies”, “internet of things” and “artificial intelligent”. The transport forum mentioned about “two forms of preventive maintenance”, namely; pre-determind maintenance and condition-based maintenance (CBM). Sensor technologies, being a part of “data-related technologies” played a major role in “increasing global use of condition-based maintenance policies for transport infrastructure”. A successful data-driven transport maintenance can be carried out using three main elements; information logistics, analytics, context awareness. The forum also presented some examples of data driven transport maintenance implementation which are functioning actively; railways, roads, bridges, tunnels, and in airports.

2.3. Literature review on– “*Up to what level data has been exploited*?“

2.3. A. Review of studies on improving service quality and reliability

Turnquist, (1981) have carried out an analysis with an aim to improve transit service reliability of buses involving “four major classes of strategies“. These strategies are: vehicle-holding, signal preemption, number of stops reduction, and exclusive right-of-way provision. The result of the study showed that, schedule-based holding strategy worked best for less frequent services where less than 10 buses run in every hour. Signal preemption works well for mid frequent services where10-30 buses run every hour. While, in the presence of suitable control point, headway-based holding strategy can perform well too. Strategy such as; an exclusive lane in combination with signal preemption must be considered in situation with high frequency where more than 30 buses are running every hour.

De Oña et al., (2013) with an aim to attract more passenger for the successful public transit system, undertook a transit service quality evaluation study. For this purpose, the authors have proposed an approach called Structural Equation Model (SEM) to “determine the influence of a series of characteristics describing the quality of the bus transit service on the Overall Service Quality (OSQ)“. The authors were also interested in revealing unnoticed aspects of service serving as “the main characteristics of the services“ quality, and described by its traits.. The proposed approach was implemented in the city of Granada, Spain. The result of the study revealed that, a powerful tool such as; SEM methodology has a potential to find out latent aspects of services “that are hidden under a series of attributes describing the quality of the service“, when used as an alternative technique. The latent variable found in the study which defines the overall service quality were; service, personnel and comfort.

Diab et al., (2015) have reported few strategies, used by transit agencies to improve service reliability, “by decreasing frequency of appearance order“, among which, bus rapid transit (BRT) and BRT-like systems (rapid transit system or networks) are the most attractive ones than the traditional approach as BRT and BRT-like system combines more than one approach, and can effectively increase the ridership, efficiency and service reliability. The authors have integrated the perspective of passengers’ and transit operators’ in order to improve transit service reliability. The authors have further studied the impact of strategies on services. Regarding passengers perception of service reliability, the authors have concluded that, there were many studies on how passengers “value their time during a transit trip“, but limited studies were found on how passengers value their “time savings“ during transit trip. The authors have observed that, there were studies reporting on how “positive impact of service improvement strategies“ can change user’s perception after the recent “implementation of a new strategy“, but rare investigation has been done in finding out, why service improvement startegies impact the perception of passengers and why and “how these perceptions change over time“. After analyzing the reliability perception of transit agencies, the authors concluded that, report on “users’ satisfaction regarding service reliability“ was not disclosed by the transit agencies as often in spite of its recognized importance. Although transit agencies employ combination of strategies for the enhancement of service, limited studies have concentrated on the exploration of “ impact of a set of strategies on the service variation“.

Oort, et al., (2015) carried out an AVL data based study, which will help the transit service providers to find out their “inefficiencies, bottlenecks and potentials“. The authors have emphasized the importance of using data in public transit as the recorded data allowed the transit operators to improve continually. Therefore, the authors took an example of GOVI data “(Borderless Public Transport Information initiative [Grenzeloze Openbaar Vervoer Informatie in Dutch, GOVI)“ and performed a “line-based analysis“, “network-based analysis“ and “inter-operator transfers“ which helped in finding out “inefficiencies, bottlenecks and potentials“ of transit service providers. From the case study, it was revealed that, shorter and reliable travel times are the main factors for efficient public transport. The authors also found that, by reducing cost, service quality can also be improved, which ultimately leading to “increasing ridership and revenues“. Their research showed and proved, how data can be used in increasing the overall service quality and reliability of public transit.

Chen et. al., (2016) have conducted a study on “reliability-based transit trip planning“ using static and real-time GFTS data by considering schedule adherence and travel time uncertainty. The aim of their study was “to encourage the public to take transit“. For their study, measures such as; transit service reliability and stop-to-stop travel time were estimated exploiting GTFS data. The result of their study showed, the importance of data in advanced public transit system “such as transit trip planners“ by helping the passengers in accessing and utilizing transit information to plan their trips which ultimately increases the transit reliability.

Firew, (2016) did a research on public transport service reliability in order to attract private car users and keeping existing passengers. The thesis work was conducted in Helsinki using Bus Line 550. With the help of AVL and APC data, “five different service reliability measures were“ analyzed by the author. Measures such as; “on-time performance, headway adherence, vehicle trip-time variability“ were agency oriented for measuring reliability. Whereas, waiting time and travel time of passengers were the passenger oriented reliability measures. The result of the study revealed that, for short headway services, regular headway keeping on the route is highly significant than irregular schedule. Improving waiting time and travel time aspects of services can increase the satisfaction of passengers as the reliability percieved by the passengers were mainly in terms of waiting and travel time.

Stelzer et al., (2016) have undertook their work on “improving service quality in public transportation systems“ where the authors have considered the importance of integrating the feedback from transit service user i.e., customers through an “automated customer feedback“ approch. The authors have revealed that, there is a lack of traveler information transmission to the transport service providers, followed by lack of dispatchers for the purpose of permitting automatic information processing. The authors have concluded that, integration of this approach will facilitate taking advantages of technological advances to its maximum level. Recommendation was made by authors to integrate traveler aspects for the purpose of optimising planning and dispatching process which will result in the enhancement of “service quality of public transport in general“.

Dixit et al., (2019) have addressed the reliability issue of using public transit by extending “the existing reliability buffer time” with the help of “smart card and automatic vehicle location data“. Therefore, “a new metric for travel time reliability measurement“ was proposed, which considered “multimodal transit journeys“ for all legs besides, waiting time and transfer times. The proposed metric was implemented in “Amsterdam transit network“ for all the public transit mode such as; bus, tram, train. The authors have found that, by developing reliability metric at disaggregrated level, it allowed for freedom of flexibility subject to the goal. From the real-world implementation, the authors have observed that, least reliable source of public transit were trams in and near city center areas in terms of single-leg journey. “Bus to bus transfer“ was discovered to be least reliable source of public transit in terms of multiple legs, resulting from missed connections leading to “longer headways“ and “longer transfer times“. Issues of specific unreliability arising at “particular route/OD pair“ can be counterreacted by flexibility aggregation at small sccale by acknowledging reliability variation among “different OD pairs and routes“.

Li and Tan,(2020) have conducted reliability evaluation study based on stop-level, route-level and network- level using “bus AVL and smart card data“. AVL data helped to capture the bus operation process, with which systematic establishment of “headway-based reliability evaluation“ was carried out. Whereas, with the help of smart card data, aggregation of “weighted stop reliability“ was performed “for route and network reliability“. The proposed reliability evaluation method was implemented at “Ningbo, Zhejiang Province, China“. The result of the study showed that, at stop-level, there was a significant reduction in reliability as soon as bus departs from stop. At route-level, with the help of graded matrix diagram, “routes with both serious unreliability and high ridership“ are easily located which helped in developing a overall understanding of complete reliability. Whereas, in case of “network-level reliability“, demonstration of temporal changes in “ridership on weekdays and the weekend“ was performed. As the reliability worsen in the peak period with larger ridership, the authors have stressed upon reduction in “bus delay at stops in the peak hours on weekdays“ and enhancement of on road “bus operation environment“.

Liu and Miller, (2020) have dived deeper to investigate the reliability of public transport by comparing “RTI apps users and non-RTI user“. The authors have carried out this study at “the Central Ohio Transit Authority (COTA)“ using GTFS and APC data, where, GTFS data represents passenger data in real-time and APC data characterize real behaviour of transit system in terms of on-time performance. Four strategies “for both RTI and non-RTI users“ were compared to evaluate the transit reliability. The authors have found that, real-time information(RTI) strategy, uses “a prudent tactic (PT)“ which has been insured with optimized buffer, and performed almost same “as a simple, follow-the-schedule tactic (ST) that does not use RTI“.

Liu et al., (2022) have addressed the transit evaluation issue by measurring accessibility by adopting “high resolution real-time data“ such as GTFS. This study is one of its kind which has actually identified a gap in terms of accessibility and realibility of public transit inspite the availability of real-time data and latest advances in technology. The authors have introduced a concept called “realizable real-time accessibility based on space–time prisms (STP)“ a measure more of conservative and realistic in nature which can solve problem regarding accessibility overestimation. The authors have defined few concepts for the measurement of schedule based overestimation and retrospective accessibility measurement. These concepts are; “accessibility unreliability“ and “potential path area“. This proposed method was implemented in Central Ohio Transit Authority bus system, Columbus, Ohio, USA by conducting a case study using high resolution GTFS data. For the evaluation purpose, the authors have divided the space time prism (STP) into three group; scheduled – STP, retrospective real-time STP, relializable real-time STP. The study result revealed that, the novel realizable STP is “user-centric and conservative measure for future transit planning and operation“ with asymmetric transit planning pattern, where high standard for transit users and operators were set by the transit system. They have also highlighted the fact that, using only schedule data is not enough for planning related measurment, as it cannot address unreliability issue because of differences in perspectives. Afterall, it is the commuters/users who are completing the trips in real world.

Alkubati et al., (2023) performed a bibliometric study on public transport reliability where they reviewed all the publications made in the last 15 years starting from 2005 to 2021 with an objective to fulfill their concerns and provide answer to the questions regarding public transport reliability. The aim of their study was, illustration of transportation reliability concept, along with providing "an overview" by listing the names of predominant "countries, journals, institutions" which have undertaken research and publication in this particular topic of "transportation reliability" with a focus "to support future studies and cooperation". The findings of their study revealed that, there is a "dramatic annual growth of public transport reliability publications". Moreover, some fascinating points related to reliability some important indicators was also revealed through this study. The authors have concluded that, a consideration must be made about transport reliability related aspects while studying this subject, which includes; "Travel Time (TT), Arrival Time (AT), Waiting Time (WT), Departure Time (DT), Boarding Time (BT), and Alighting Time (ATT), Seat Availability (SA), Missed Connections (MC), Commitment to the scheduled Time before the trip (CT)". Between 2005 to 2021, "reliability-related research" conducted in 5 top most countries includes; China, Netherlands, England, Australia and Poland respectively. During that same time period (2005 to 2021), top 5 journals in the same fields ("reliability-related research") were; "Transportation research part a-policy and practice, Transportation research record, Transportation policy, Transportation research part c-emerging technologies, Transportation".

2.3. B. Review of studies on – "Real-time information- how accurate is it?"

Dziekan and Kottenhoff, (2007) have presented their study as a means of proof that real-time information can reduce perceived waiting time of the passengers by 20% along with other "positive psychological effects of at-stop real-time information displays". The authors have constructed a framework by considering seven main effects of real-time information displays such as: : “(A) reduced wait time, (B) positive psychological factors, such as reduced uncertainty, increased ease-of-use and a greater feeling of security, (C) increased willingness-to-pay, (D) adjusted travel behaviour such as better use of wait time or more efficient travelling, (E) mode choice effects, (F) higher customer satisfaction and finally (G) better image“, which can be used as a base for evaluation studies. The authors have undertaken two studies, where study 1 proved the effect of real-time information in reducing waiting time, and study 2 proved the other positive psychological effects due to the “at-stop real-time display“. This study can be used as a base for evaluation studies, despite its incompleteness and many related effects.

Cebon and Samson, (2011) have undertaken a study with an aim to improve transit effectiveness using real-time information. Through this work, the authors have claimed that, real-time information have the potentials in solving “cities congestion“, “transporting the people in the cities effectively“, “informed uses of public transport“ and “effective ridesharing“. Survey result undertaken for their study revealed that, a combination of technology and real-time hub software can increase the systems utility to its passengers by publishing real-time information, resulting in time saving.

Watkins et al., (2011) have presented their survey report, where the authors wanted to see if there is a difference of “impact of mobile real-time information“ between the perceived waiting time and actual waiting time of transit riders. The findings of their study revealed that, passengers have experienced decrease in actual waiting time using real-time information in comparision to traditional information system. The reported reduction in waiting was nearly 2 minutes less while using real-time information.

Brakewood et al., (2015) have undertaken a survey based analysis of real-time information on the commuters of Boston city. The analysis of wait times revealed that, commuters, who consulted rail real-time information have reported about “a decrease in self-reported usual wait times“. But the authors have found that, on the days of survey, there was no “statistically significant difference“ in perceived waiting times between real-time information users and non-real-time information users.

Bruglieri et al., (2015) have worked on displying the role of real-time information to counter the situations of delays and disruptions. Therefore, the authors aimed at designing and development of a real time mobility information system for the management of unexpected events, delays and service disruptions concerning public transportation in the city of Milan. The authors were able to tackle such situation with the development of of MOTUS travel planner. Through their work, the authors have came across users perception on the service provided by this particular travel planner, that is; reliability perception and perception about robustness can be raised depending on how quick and automatically disruptions are handled, which in a way can reduce uncertainty, and increases the attractiveness of public transit.

The work of Harmonx and Gayah, (2017) presented another aspect of real-time information provision by the transit authorities to the passengers. According to the authors, it was neccessary to know, which information in real-time passengers were interested to receive (information demand) and transit authorities ability to supply this particular information. However, in this process, the authors have encountered that, various stakeholders were “involved in the development, dissemination, and operation of real-time transit information systems“, and some constraints related to the implementation of real-time transit information systems. Funding and staffing were the main reasons, behind keeping the latest tehnology from being fully exploited to best serve the transit reliability.

Brakewood and Watkins, (2019) have carried out a literature based study on the benefits of real-time information to the passengers. Their work was focused on; reducing passenger waiting time, decreasing total travel time, information. “increased transit use“, raising of transit satisfaction, raising the awareness of “personal security“ with the help of real-time information. The authors have found strong evidence in the literature that, real-time information was successfully reducing “perceived and actual wait times of passengers“. Moreover, real-time information was also seen benefiting the passengers by increasing the “perception of personal security“ and overall satisfaction with transit service.

Liu and Miller, (2020) have conducted an empirical analysis on whether real-time information (RTI) apps reduces waiting time of passengers or not. According to the authors, “popular RTI apps aim to diminish waiting time to zero”. But, this attempt made by RTI app might be risky, as the actual bus arrival time may change as soon as the person have left home, and RTI app does not update vehicle location and bus arrival time on a continuous basis. This distinction “between RTI app and reality“ might lead the passenger to miss the bus with an acquired longer waiting time. In order to tackle this kind of situation, the authors have examined the impacts of RTI on waiting time of public transit users' subjected to "the empirical performance of a public transit system" by comparing "two RTI-based strategies" with non-RTI-based arbitrary arrival and following schedule strategy. Two RTI-based strategy used for the comparison were: greedy strategy and prudent strategy. This comparison study was further carried out using real-time GTFS data and APC data from "Central Ohio Transit Authority (COTA) in Columbus, Ohio, USA". The result of the study showed that, the so called "best RTI strategy, a prudent tactic (PT)" performed no differently than non-RTI based "follow-the-schedule tactic (ST)". Moreover, the waiting time of passengers is longer in RTI based prudent tactic, with higher chances of missing the bus than passengers using non-RTI based schedule tactic (ST). Whereas, RTI-based "greedy tactic (GT)" for the purpose of achieving zero waiting time have appeared to be the worst RTI-strategy than any non-RTI-based strategy. Therefore, the authors have concluded that, the so called best RTI-strategy despite its advantages of saving time at certain stops, hours of the day for certain users, was unable to outperform globally upon "simply following the published schedules". According to the authors, accuracy and reliability of RTI apps can be improved through the addition of "pre-calculated insurance buffers" subjected to certain "risk attitudes of users".

3. Challenges and opportunities of data utilization, and Perspectives of different stakeholders

3.A. Literature review on challenges and opportunities of using data

Challenges faced with data utilization

Prarthana and Jayakumer, (2017) have mentioned that, security issue was the biggest challenges while dealing with big data. In their views, collecting data is not a problem but to manage it and make sense out of it while maintaining data privacy was the main challenge.

Torre-Bastida, (2018) have pointed out few challenges while working with big data in transport sector. The first problem was related to “real-time data analytics“, where the authors have revealed that, ITS infrastructure and deployment of its platform does not meet safety standard for real-time transport data processing and analysis. The only solution in future to this problem is “real-time service automation and optimisation“. Second problem was related to “security and privacy“ which is a long-established concern. In terms of security and privacy, transportation industry faced difficulty in collecting private data. Third problem was arising from “new data sources“, due to its large volume and amount of data coverage, variety, value and quality of data. Based on these problems encountered by the authors in terms of transportation and mobility, the only ways for successful continuation of big data technologies through the incorporation of further ingestion and data processing to tackle the “increasingly challenging properties in terms of volume, heterogeneity and dynamism“.

Neilson et al., (2019) have reported four main challenges found throughout their literature review based study while using big data for transportation. These challenges were; challenges arising during data collection, challenges while storing data, maintaining data quality and data security. A graphical representation of these challenges were given by the authors in figure 8 below.

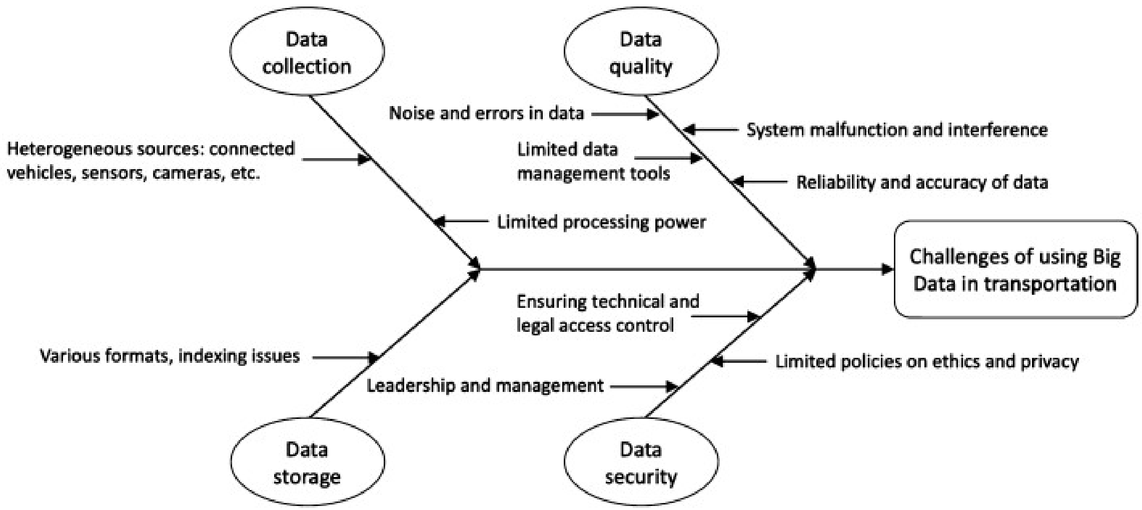

Figure 8: “Key challenges of utilizing Big Data in transportation“
Source: Neilson et al., (2019) Systematic review of the literature on big data in the transportation domain: concepts and applications

Ge et al., (2021) concluded in their study that the main challenge in using transit data is processing of data. Findings of their study showed that, most often data needed “to be processed to derive meaningful information“. Apart from this, the authors also mentioned other problems during data utilization such as; significant overlap between different data sources, the appropriate functionality requirement of data was not defined without which advantages of these data cannot be taken, the concept of “multi-modality in public transport“ must be included for (Redmond et al., 2019) comprehensive “analysis of public transport data“, which influence the decision making process of passengers in using various forms of transport, “including cars and bicycle“.

Opportunities of data utilization

UITP (2021) have reported that, through data sharing, “bringing net positive value for mobility for life” was possible. The ecosystem for sustainable data sharing can be achieved only through bringing “net positive value across all stakeholders”, where, the shareholders are composed of five capitals, namely; financial, human, social, physical and natural. Data is efficiently benefiting citizens by enabling the pubic transport sector to lead towards innovation and digitalization.

Ewoldsen, (2022) have compiled the report presented by Transit Research Board's (TRB's) Transit Cooperative Research Program (TCRP), where, the report declared that, data sharing has the potentials to improve "customer satisfaction, revenue and service performance". Another report presented by National Academies of Sciences, Engineering, and Medicine, on The Role of Transit, Shared Modes, and Public Policy in the New Mobility Landscape, have showed that, among all the benefits of data sharing and collaboration, "multimodal transportation options" can also "reduce carbon and other emissions and improve equity". Another report presented by TRB's Airport Cooperative Research Program (ACRP) notes the importance of airport operators investing in data sharing for passengers. According to the report, data sharing allowed the passengers to purchase discounted public transit ground transportation tickets to access the two area airports.

Ortelio (2018) have identified some opportunities that if taken into consideration can significantly improve the way transport organisations adopt and utilize Big Data technologies. Theses are; business opportunities, opportunity for collaboration, transfer ability of data, adopting standards for data, opportunities for practitioners and "big data application developers" into becoming the skilled personnel. SMEs, large corporations, transport authorities can gain "competitive advantage" by developing new innovative and sustainable business models offering better services to the transportation sector (Grover et al., 2018; Hashem et al., 2016). Business opportunity could also be the development of a mobile application that could gather transport related datasets, which could be either sold to a transport authority or could be provided as open and lead to new better transport related services. The scope for "collaboration between transport providers, businesses, research centres and software developers" was highlighted by Kim et al., (2014). According to the authors, transport organisations have the opportunity to work with businesses and academia for developing intellectual capital so that academics and businesses "focus their research on transportation related problems". The potentials of transferability of "Big Data" from context to context of application has been acknowledged by Lycett, (2013); Tallon et al., (2014). The need for adopting data standards by agencies has been proposed by Kujala, (2018) so that one single data can be used for various agencies for studying various issues. One such example was, using timetable data for the purpose of studying accessibility, level of service and the

“structure and organization“ of public transport. Data also served a huge potentials for the practitioners and application developers where, staff can be educated on the topic of uses of data, creation of business strategies by stressing on data-driven analysis.
According to Ge et al., (2021) every data collected from their respective sources has distinct applications and can deliver uncaptured information that are yet to captured from other sources. The authors have emphasized on the possibility where, each data has the ability of providing distinct information, where each data is capable of impacting the behavior of either passengers’ or the transit service providers or both.
Jacobs, (Accessed on, 05.09.2023) Director Mobility and Transport, Lufthansa industry solution, revealed the “potentials uses of big data in publuc transportation“. According to him, most common challenges faced by companies in public transport sector are related to operation optimization, costs cutting, revenue increase. Companies must exploit the potentials of new technologies in the form of big data, if they desire to overcome these challenges and assure success in future. In a circumstance with less financial freedom where municipalities having budget constraints, governments are cutting the subsidy (in case of Germany), provision for “with high-quality services” to the passengers is possible only through “digital transformation“, accompanied by exploitation of “big data technologies“ as these technologies delivers the opportunity to transportation companies for cutting cost and increase in the number of customers with the help historical data analysis.

3.B. Literature review on perspectives of public transit operators, passengers, policy makers and researchers

Perspectives of public transport service providers

Walker, (2008) have revealed that, public transport have two main goals to achieve in general; patronage goal and coverage goal, and perspectives of public transport service providers vary accordingly the type of goal they are aiming to achieve. Because, patronage goals include: “goals related to financial return”, “goals related to vehicle trip reduction”. Whereas, coverage goals focused on service availability of socially disadvantaged population, covering as much of geographic area as possible. Based on these two distinct goals, “both short-term service design decisions and long-range network planning” were carried out by the transport service providers.

Oort et al., (2015); Cats et al., (2015) have added that public transport operators are concern about “quality improvements”, through increasing “robustness and service reliability”. Data plays an instrumental role in addressing these challenges which was shown and proved by the authors through their work. One such case was, calculating “additional travel time and travel time variation” with the help of smart card data.

Schmöcker et al., (2017) have showed the Transit operators effort in reducing the money handling fees through smart card data. Both the passengers and transit operators can exploit the available data to best suit their needs. Transit operators have targeted the use of smart card data in such a way that, spliting the revenue become “easier to integrate the fare systems of several operators within a city”.

According to Noursalehi et al., (2018), transit operators are aiming to develop and apply predictive analytics to find the solutions to transportation problems which cannot be seen before, until recently. With the help of AFC data this type of predictive analysis study was successfully carried out, which can help the transit operators with timely prevention of “unwanted deterioration in service quality” in due time.

Koutsopoulos et al., (2019) have revealed the main goals of transit operators were, to use the collected data for measurement, planning, policy, control, communications, proactive decision support, information and service customization. According to the authors, these goals can be achieved through fusing data from the various sources.

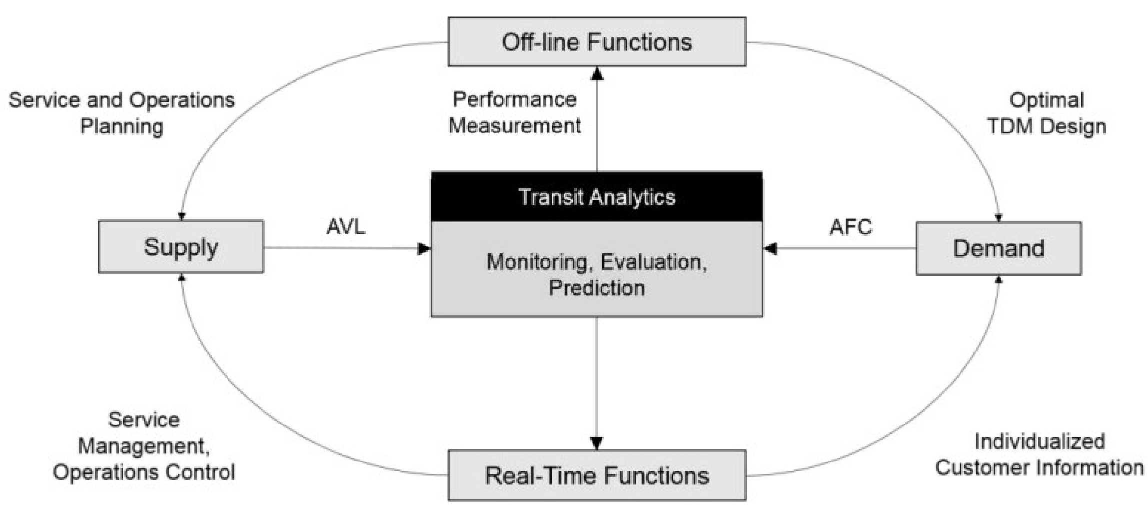

Figure 9. “Automated data collection and key functions”
Source: Koutsopoulos et al., (2017) Automated data in transit: Recent development and applications

Khan et al., (2020) have revealed that, public transit operators and stakeholders are mostly concerned about increasing the "patronage in terms of planning principles and measures". In order for the planning strategies to convert into reality, public transit service providers also need good governance, political and public support, and "well-functioning collaboration between the involved organisations".

Leng and Corman, (2020) have found Transit operators keen interest in improving services especially during transport disruptions through offering information to passengers. According to the authors, the passengers can use the services offered by the public transport networks if they have knowledge about them.

Redmond, (2020) have reported that, transportation providers are interested in making travel planning more reliable through transparent data collection.

UITP, (2020) reported that, "public transport is a key accelerator of the economy, job creation, social inclusion and sustainability". In Europe, public transport companies were aiming to recover the European economy by saving costs which was arising due to traffic congestion, through investment on networks and creating jobs, promoting health and sustainability by encouraging "active lifestyles" by involving walking, cycling and taking public transport in to their journeys.

Keller et al., (2022) have found that, along with providing services to the customers, Public transport companies are aiming to provide clean environmental friendly sustainable means of transport.

Perspectives of passengers

According to Casello et al., (2009) passengers perception of transit reliability was all about on time departure, reasonable in-vehicle time limits, on time arrival at the designated station where traveler can "reach his or her final destination without being late".

Li and Hensher, (2013) have reviewed "public transport crowding valuation research, using a number of primary studies conducted in the UK, USA, Australia and Israel", and have revealed the perception about in-vehicle crowding. The authors have found that, where public transit took "crowding" as an "important component" with an opportunity to improve the quality of public transit services, the passengers were more concerned about service reliability and have "willingness to pay(WTP) for reduced crowding".

Increase in crowd level was equally proportional to the increase in passengers perception about “value of travel, walking or waiting time”.

Drevs et al., (2014) have done “an empirical analysis on the effect of subsidies on individual willingness to pay for public transportation”, and revealed that, each individual passenger have major “effects of personal attitudes and beliefs on crowding effects, in line with behavioral theories”. From the perspectives of passengers, the authors have also found out that, when public transport was used as an option, some choose to pay “higher fare“ as they are mindful about the fact that “public transport is financed by subsidies”. The authors also suggests that, when passengers are aware about subsidies, they will give more value to the services offered by the public transport.

Ehreke et al., (2015) have studied the cost-benefit analysis of “the value of reliability (VOR) and the value of travel time (VOT)“ and compared personal travel with business travel. The authors have found that, different groups of passengers perceived the interpretation of reliability differently. The authors have seen that, passengers making trips with cars gave more value to travel time saving, in comparison with variability reduction. Whereas, in case of passengers making trips with the public transport, less value was given to reliability and more value was given to travel time saving as well.

Kou et al., (2017) were interested in analyzing “commuters’ willingness to choose the travel schemes with the different reliabilities”, for which the authors have carried out a survey on stated preferences. Their survey was aimed to find out the commuters mode of choice by taking into account income and time constraints. The survey result confirmed that, commuters are more concerned about reliable means of public transport than travel time in general. “Commuters with higher income levels generally prefer a travel scheme with less travel time and higher reliability“.

Haywood et al., (2017) have surveyed 1000 respondents from Paris on crowding in order to find out the passengers perspective on the matter. From other research works on crowding, the authors have found that, most works focuses on the “trade-offs between crowding and time or money“ as a means to investigate the crowding effects. However, through this work, the authors have proved that, with in-vehicle density (IVD),“crowding costs grow linearly“. When it came to wealthy passengers, “users’ satisfaction decreases more quickly with IVD“. They have also revealed that, passengers were feeling discomfort due to “not being seated“, limitted convinience to utilize the time during journey, and being physically close to “other travelers *per se*“.

Hörcher et al., (2017) in an attempt to estimate the crowding cost of passengers with "large scale smart card and vehicle location data", have revealed that, passengers willingness to pay is inversely proportional to public transport services. The more the "attractiveness of the service drops and customers become less willing to use it or pay the same fare". Therefore, the authors have concluded that, by measuring "crowding discomfort" of the passengers, an "optimal supply-side decision" can be derived.

Koutsopoulos et al., (2019) have found reliability and crowding to be given importance by the passengers. With these two points, passengers are measuring system performance. According to the authors, passengers experiences reliability between specific origin-destination at each level of journey. When in comes to high frequency services, reliability of public transit services are resolved through the prediction of journey times, without clinging to published schedule. In fact, high frequency services are so reliable that, passengers hardly confer "a schedule before starting their journeys".

Chocholac et al., (2020) were interested to know the "customers' perception and satisfaction regarding urban public transport companies" as a means to promote the use of public transit for "sustainable city logistics". In their study, the authors have found that, different groups of passengers have different perceptions about "individual service quality". Not only this, but also perceptions about "customer care and comfort" was seen to be varied according to "different age group of respondents". Older respondents with 55 -90 age group were satisfied with the quality of service. Whereas, younger respondents from 35-44 "age group were satisfied with comfort, information support and customer care services, but quite unhappy with "price accessibility and availability". Therefore, the authors have concluded that, there is a requirement for public transport operators/ agencies in order to pay endless "attention to the communication with individual groups of passengers" through the adoption of pertinent communication means for each group of passengers which will help to increase the number of public transport users.

Luke and Heyns, (2020) have done an emperical analysis and found that, customers are having very low perceptions about public transport service quality "that most public transport users aim at buying a private motor vehicle, as soon as they can afford it". This study was based on Johannesburg city, where the aim was to promote environment "friendly modes of transport" through the achievement of 80:20 modal split corresponding to "public to private transport use". In order for to acheive the stated modal split, the customers expected more attractive and viable form of transport alternative with higher service quality, only then “the travel patterns in the city“ can be changed.

Wang and Zacharias, (2020) have conducted a survey to find out the most relevent factor behind percieved crowding in public transit. By doing so, the authors were determined to improve the travel comfort for effective “switch to public transport“. The survey result revealed the passengers perception on comfort through the “evaluation of crowding“, which was related to real passenger density, noise and smell in a subway car. Additionally, travellers concern about overcrowding could be addressed through noise reduction and improving the existing environmental condition and ultimately upgrading the comfort of passengers.

Aghabayk et al., (2021) have also conducted a case study to find out the perception of passengers on crowding during pandemic such as Covid-19. Their case study was undertaken in Tehran metro rail, where maintainig social distance was allmost impossible. The study result showed that, “passengers perceived more disutility for crowding during a pandemic outbreak compared to a time before the pandemic“. The authors have also found that, passengers percieved the negative effects of crowding when they had to stand, all seats were taken and unable to maintain social social distancing during pandemic. The authors concluded that, by “understanding passengers’ perceptions of crowding“ can help transport operators and planners in proper management of risk involved in mass transportation not only during pandemic but in further situations too.

Perspectives of researchers on public transport

Ceder, (2021) have worked on visualizing the future of urban transport based on “dramatic change on urban mobility“. From the author’s point of view, there are some excessive confusion arising from “the evolution of automated/autonomous vehicles (Avs)“ which is threatening the future research scope of researchers on this particular subject for the improvement of urban mobility in future. Thats is why, the author found it neccessary to emphasize on profound thinking of the future possibilities instead of their prediction. The author have revealed “four important themes of studies“ such as; behavioural trends, decisions, automated transit networks, and modelling and technologies, with requirement for future research, can clear the prevalent confusion.

Through the proposal of “only conceptually – a Visionary, Feasibility-related and Realist (VFR) perspective-based approach“ possible future visions can be explored which can create endurable and impressive public transit systems.

Tennoy, (2022) have presented a work from a researchers point of view by highlighting few important tips which can be incorporated by transit planners in order to make the public transport more attractive. The author presented the result of his report from the study conducted in Norwegian cities. First tip; “knowledge acquired through studies on larger cities might be relevant for smaller cities”. Based on the priority, patronage coverage in smaller cities can be increased. The tip also helped in understanding, how the reorganization of system can increase the attractiveness of public transit system. Second tip; a goal oriented increase in public transport competitiveness while reducing car usage, coupling with intensified knowledge and competencies among public transport planners, and organisational changes, can pass down more power to the professionals. By following these tips, public transit could be made more attractive and the patronage of “small and medium-sized cities” could be increased.

Perspectives of public transport policy makers

Chowdhury et al., (2018) with an aim to develop an “integrated public transport systems“ conducted an detailed analysis on comparing the perceptions between transit users and policy makers. The authors believed that, without an alignment in perceptions between the public and transit policy makers, an integrated public transport systems cannot be attained. The users and policy makers of Auckland, New Zealand, were asked to give weights to few important attributes and sub-attributes in order to create an integrated system. The findings of the study revealed that, policy makers gave weighted value to “network integration“ as one of the most important attribute for integrated public transport systems. From the perspectives of policy makers and passengers, “fare and ticketing integration” should be considered as an important attribute. From the perspectives of frequent travelers making regular transfers, “integrated timed-transfer“ should be considered as one of the key attribute, which was not taken into consideration as one of the important feature by none of the policy makers that were being interviewed. Apart from these, there were also differences in attributes valued by policy makers and the passengers. Therefore, it was advised that, policy makers should take note of these differences before making any transit policy related to integrated public transport systems in order to retain more patronage.

Tang, (2018) have worked on a case study, where the authors have adopted an operational strategy which helped in reducing “operating cost“ along with reduction in “undesirable emissions“. The author have considered operational strategies such as: full route operation, short turn, limited stop, and a mix of limited stop and short turn which was implemented in Dalian of China. Through this strategy, the author could achieve 12.5% reduction in the number of vehicles and “emissions per pollutant (CO2, HC, CO, NOx, PM2.5)“ was reduced to 13% in comparision to “full route operation scenario“. Additionally, there was 20% reduction in emissions due to contactless card payment and off-board payment. Therefore, the author believed that, this particular work can serve as best example in better policy making for environment friendly commuting.

Paulsson, (2018) have worked on finding out the means of translating a new policy “in the political–administrative nexus“ by carying out a case study in “Regional Public Transport Authority (RPTA) in Stockholm, Sweden“. The author have encountered conflicts of interest among various stakeholders while translating the policy regarding “sustainable public transport“. The author have highlighted that, conflicts arising between different shareholders was mainly due to the lack of basis for comparision and different laws and political parties are having their own objectives. From author’s perspective, public transport is a tool for politicians, and there is a need for finding a common ground by strengthening the relationship between politicians, administrative local configuration and stakeholders, as this reliationship can literally shape the policy translation and its related outcomes.

ITF, (2021) based on their findings, have recommended few points which could be helpful in making better transport policy. These recommendations are: data collection should be done for described goals for the sake of data protection; big data can be used in transport model by developing guidelines with periodical revision of the guidelines; user consent based and secure apps promotion in smartphone for location data collection; privacy protection through numerous solutions; household travel survey with defined roadmaps; smartphone based household travel survey; smartphone based data mining with AI; promotion of data steward function for secure exchange of public and private data; investment in public- sector workforce for data-related training and data specialist recruitment for precise goal.

UITP, (2021) have presented a policy brief about “sustainable data sharing“, where it was concluded that, for the mindset to shift towards “data as an opportunity“, there is a need of focus by all stakeholders on value creation for end users and the objectives alignment of stakeholders with the end users. Additionally, there is a need for collaboration between different sectors and stakeholders for the purpose of providing “strong, consistent and sustainable mobility services to citizens“.

Skartland, (2022) conducted a case study in two Norwegian cities with an aim to see the affect of unfavorable transit planning such as; “lack of knowledge, lack of collaboration or political conflicts“ on increasing transit competitiveness. The study was based upon two Norwegian cities, where qualitative case study was undertaken because of the initiated projects aiming to increase “transit competitiveness versus the private car”. The author have interviewed different shareholders and other important personalities involved with the project, several documents were studied and interpreted by the author with the help of “existing theories and case studies” which were further essential in determining probable causes for unfavorable decisions regarding transit competitiveness. The case study revealed that, “conflicting politics” was the main reason behind “unfavorable decisions in transit planning”. Therefore, the author have suggested that, there is a need for urban planner to inform and awaken the politicians and decision makers in order to avoid conflicting situations through knowledge production, leading to the prevention of unfavorable decision making within transit planning and urban planning.

4. Analysis

4.A. Findings

4.A.1. Findings – Sources of data and factors related to data-driven transit planning and operation

- ➔ According to the literature review, types of data used in public transport were; customer data produced by passeners, operational data produced by transit operators, mobility data produced by both users and transit operators, exogenous data produced by third parties.
- ➔ The review of literature has revealed that, there were different data sources that were being used in public transport such as; AVL, APC, AFC, GIS, GPS, GTFS, GNSS, ITS, ATIS, mobile network data, smart card data, API and big data. Data has also been seen to categorized into endogenous and exogenous sources. Data included in endogenous category were; AVL, APC, AFC, GPS, GNSS, GTFS, Smart card data etc., due to the nature of data which were; easy accessibility, updated, reliabile, correct and consistent. Whereas, exogenous sources of data included smartphone data, social media data, traffic data, weather and survey data.
- ➔ Data collection is "a task of compiling data captured by devices and sensors and generated by citizens and vehicles through different channels" (Torre-Bestida et al., 2018). This master's thesis has reviewed data collection technique of AVL, AFC, APC, GTFS and ATIS systems. Each system has specific ways of collecting and transmitting data for further use. AVL, AFC, APC, GTFS and ATIS systems has being providing us with both real-time and not-real-time data, which is being continuously used by transport service providers for numerous purposes.
- ➔ After literature review, it has been revealed that, there are some important factors affecting the uses of data in public transport. First of all, infrastructure played the most important part in acquiring data from various sources. In order to collect data and make the data available in real-time, needed an appropriate installation of devices such as; GPS, smart card, videos and cameras. A better infrastructure performed a better role in handling different kinds of transit data according to its various needs. The factors such as; integration, standardization, validation, matching are pre-processing steps which played an important role

once the data have been collected. Integration of heterogeneous and inconsistent data from various sources were combined to deliver structured and homegeneous data through standardization, validation and matching of data. NeTEx, DELFI are some popular examples technical standards for allowing efficient exchange of transport data among distributed systems by bridging different modes, regions etc. Data validation allowed ingetration of data by providing reliable information. Once the collecetd data from various sources has been validated, it could be used further by various agencies for different purposes. Data matching allowed for unifying data by eliminating duplicate. AVL and APC data matching, APC bus data matching with static bus stop data are some example found in literaure. Processing of data is the next step after completing pre-processing steps. Processing of large data volumes allowed "to compute statistical values to assess the quality of the public transportation system" (Massobrio et al., 2016). Problems related to processing of GPS and GTFS data has been encountered in the literature, where map reduce solution was adopted to process the large volume of GPS data and data cleaning workflow was adopted in case of GTFS data for less noisy data delivery (Wu et al., 2022).

➔ Once the data processing step has been taken care of, next comes the concern relating to storage of data and protection of privacy and security. According to the literature, storage of data is done depending on its categorization into; structured data, semi-structured data and unstructured data. But, it was reported in the literature about storage issues which is arising due to the increasing volume of transportation data. Solution to this problem was, cloud computing and use of open data tool have been found in the literature. GTFS was used as open data tool to solve storage issues of transportation data. Additionally, it was also recommended in the literature that, investment for underlying technologies is needed if we want to use the this tool. When it comes to privacy and security issues, many authors have showed the concern regarding data protection and ways to achieve it. These options are; encrypting the data, "privacy preserving data publishing algorithm" to reduce the privacy breach of smart card data, " privacy-preserving task recommendation scheme" before data "being outsourced to the crowd sourcing platform“ and use of block-chain based self-sovereign identity technology.

➔ Digital twin is another important factor needed to be considered in the era of digitilization. This concept has been used in the area of construction, in the process of data acquisition and integration of data from various sources, and in simulation study of smart cities. Through digital twin, many option can be simulated digitally before being implemented in real-world, which will be helpful in identifying faults and qualities of each plan.

➔ Once the data have been collected and stored, the next step is to take the advantages of these data for the overall improvement of transport sector. This was done through visualization and through service optimization. Examples of exploitation through visualization found in literature review are; “predicting non-recurrent traffic congestion“ by developing “prediction and visualization technique“ with the help of floating car data, GSM probe data, data from stationary sensors (Kwoczek et al., 2014), to get important insights on the potentials of bus rapid transit by studying the spatio-temporal dynamics of the passengers using smart card and GTFS data (Tao et al., 2014), to study the relationship between destination choice and buseness cluster by adopting a visualization approach with the help of social media data (Huang et al., 2017), improving “public transport disturbance management“ by exploring the potential of real time traveler data in disturbance management with the help of AFC data (Jevinger and Persson, 2019), visualizing the operation of public transit system with the help of GTFS data in order to gain “qualitative information“ and useful insights for better “transit system design and planning“ (Prommaharaja et al., 2020), creation of a framework through visualization technique in order to explore “continuously broadcast public transport performance datasets“ for better planning and operation in transport sector with the help of real-time GTFS data (Lock et al., 2021), use of visual analytic system in order to analyse the cause and effect of traffic congestion with the help of GPS vehicle flow data (Pi et al., 2021), development of a visualization tool called ‘Mobisuite’ for mobility trends analysis and providing assistance to the transport service providers in making decision with the help of smart card data (Fazari et al., 2021). Exploitation through service optimization instances has been found in few literatures, where the study was specifically aimed towards service optimization. These cases were; enhancing “the strategic, tactical, and operational performance of transit authorities“ by exploring the potentials of smart card system (Pelletier et al., 2011); development of online platform for transit performance monitoring using AFC and AVL data for the purpose of exploring latest data collection techniques (Ma and Wang, 2014); exploring cases related to frequent exploitation of smart card data for shaping readers understanding on the latest advances in big data sources with its role in mobility flow (Anda et al., 2017); assessing “passenger benefits of real-time information provision“ for optimizing overall transit services which was complemented by the development of GTFS data and smartphones carried by passengers (Brakewood and Watkins, 2019); measuring the performance of public transport using AVL, AFC, APC data in the form of ITS data to learn about ridership and demand pattern, where “generation of optimal vehicle schedules“ was the only problem addressed with the help of AVL data exploitation (Iliopoulou and Kepaptsoglou, (2019); measuring the reliability and punctuality of the services with the help of AVL AFC data in the form of exploitation through service optimization (Ge et al., 2021); service improvement by developing a bus route identification framework which addresses few important issues on bus position and route (Lyu et al., 2021).

Table 1: An overview of data exploitation in public transportation

Author	**Data**	**Exploited for**
Iliopoulou and Kepaptsoglou, (2019)	AVL	“Generation of optimal vehicle schedules”
Gordon et al., (2013)	AFC, AVL	“To infer boarding and alighting times”, to derive the location of individual bus passengers and to derive the transfer of passenger trips between “various public modes”, and origin–destination matrices construction
Ji et al., (2015)	AFC, AVL	“Estimation of Origin-Destination flow metrices“
Gordon et al., (2018)	AFC, AVL	“Estimating scaled passenger flows while preserving the intermediate interchange information”
Trépanier et al., (2007)	AFC	To justify “similarities between trips over successive days” and identification of “transit alighting points”
Moreira-Matias et al., (2015)	AVL	“To adjust bus route networks”

Pelletier et al (2011)	AFC	“To understand travelers behavior”
Moreira-Matias et al., (2015)	Survey data	“Long time transport planning”
Anda et al., (2017)	Smart cards	“Individuel trip reconstruction, estimation of OD matrices”
Lyu et al., (2021)	GPS	“To identify bus routes”
Ge et al., (2021)	AVL, AFC	“Service optimization”
Jevinger and Persson , (2019)	AFC	“To improve public transport disturbance management”
Tao et al., (2014)	AVL, AFC, GTFS	“To reconstruct travel trajectories of bus passengers”

Source: Own illustration

➔ In the literature, authors have listed few topics which are yet to be fully exploited. A tabular presentation of unexploited or partially exploited topic are given below.

Table 2: Lists of unexploited or partially exploited topics

Author	Topic
Yan et al., (2006)	“Bus route generation and schedules“
Li and Bertini, (2008)	“Optimal stop spacing“
Lovrić et al., (2013)	AFC based reproduction of passenger preferences
Moreira-Matias et al., (2015)	“Studies on design related, studies on problems related to operational-level“
Moreira-Matias et al., (2015)	AVL data based bus route network adjustment
Liu et al., (2017)	“Inferring trip patterns along with bus network design“
Shen et al., 2017	“Devising robust vehicle schedules based on observed trip times“
Iliopoulou and Kepaptsoglou, (2019)	Strategic level decision, fiscal policy
Ali et al., (2016)	To improve AVL/GPS based “route choice and passenger behavior modeling”
Ordóñez Medina and Erath, (2013)	AVL/AFC based “operational planning decisions, AVL data based vehicle scheduling
Desaulniers and Hickman, (2007)	“ITS supported optimization method in the operational planning stage”
Sánchez-Martínez et al., (2016)	Improving “the performance of control models” during overcrowding
Berrebi et al., (2018)	“Favoring simulation environment for real-time implementation of holding strategies”

Williumsen, (2021)	Use of mobile networking data for transport model development and policy making, development of new mobility technologies, travel demand variation, to study changing travel preferences, better travel demand forecast by understanding uncertainty

Source: Own illustration

➔ Next important factor is about evaluation of transit services. In the literature, transit evaluation has been done based on accessibility, reliability and robustness. It has been seen from the literature review, how crucial role accessibility is playing for the evaluation of transit services. Transit accessibility was one of the important factor to increase the public transit attractiveness, reducing congestion (Xu et al., 2017), commuter friendly urban land use plan (Xu and Yang, 2019), in addition to addressing 'accessibility unreliability' (Liu et al., 2022). But, when the accessibility issue was focusd on older adults (Ravensbergen et al., 2022), there is a need to define certain accessibility criteria for older adults, before drawing any conclusion. There were many studies on evaluation of reliability found through the literature, where measures used for evaluation were: estimation of service reliability at stop, route and network levels by bridging the gap between transit agencies perspective and passengers perspective (Chen et al., 2009); through international survey for gaining insights about "how public transport authorities deal with (improving) service reliability" (Van Oort, 2014); public transit evaluation by combining both passengers and transit operators perspective on service reliability (Diab et al., 2015); providing solution against transit unreality by considering the perspective of the passengers (Tran et al., 2015); by implementing real-time information generation scheme to "improve overall level-of-service" (Cats and Loutos, 2016); addressing the reliability issue arising from to multi-modal-itinerary planning by developing a stochastic optimization model (Redmond et al., 2019); comparing the parameters of reliability according to passengers perception against the perception of transit service providers (Kathuria et al., 2020); through the evaluation of transit signal priority (Anderson and Daganyo C, 2020). Robustness of transit services was evaluated: by adopting an index in case of link failure in order to measure total travel time (De-Los-Santos et al., 2012); by introducing a measure called "Robustness Critical Points (RCP)" to solve the railway congestion problem (Andersson et al., 2013); comparing different line planning approaches and their impact on railway timetables (Goerigk et al., 2013); measuring the robustness of railway schedule using Intelligent Transportation Systems (Corman et al., 2014).

- ➔ Marketing and policy were another important factors found in literature, which played a crucial role in implementing any policy and to make the transportation sector more attractive. Many studies on marketing and policy have been found throughout the literature review, where findings are about need to make transit related investment (Lu et al., 2018), increasing market competition (Ivaldi and Vibes, 2008; Li et al. 2009), implementing upgraded version of fair policy (Tirachini, 2013), urban congestion management policies (Basso and Silva, 2014), scheduling and pricing policies (De Palma et al., 2015), transport policy favoring welfare of the society (Eliasson and Börjesson, 2014) and factors affecting transit policy planning (Lu et al., 2021). Whereas, the work of Brown et al., (2003) was about presenting a real life example of scheme called “BruinGo” where the students and staffs does not have to pay for the ride, instead, the university pay on behalf of them to the transit providers. This type of scheme can motivate transit service providers and policy makers in producing sound “transportation planning reforms“. The work done by Thomas and Bertolini, (2015) was about the process of policy transfer “on transit-oriented development (TOD)“ to land owners and “transportation planners“. This kind of work helps to understand the role each stakeholders are playing while implementing a new policy for better outcome.
- ➔ Next, comes the importance of implementing transit big data systematically. It helps the researchers and transit operators in deriving insights for decision support for transit planning, operation and policy making (Lu et al., 2021; Lu et al., 2018). It can also help in solving various transit related problems (Liu et al., 2021). It is also helpful in identifying “efficient strategies“ by undertaking comparative assessment “between different approaches“ and finding out research gaps (Iliopoulou and Kepaptsoglou, 2019).
- ➔ The basic concept of information management (IM) was provided by Voss and Gutenschwager, (2001) and Ge et al., (2021). An example of IM in transport sector was provided by Axelsson, (2020) where, Building Information Management (BIM) was used to ensure “a common digital facility for efficient

information management in the building and infrastructure sector“ with an ultimate goal to manage asset efficiently. A case of Swedish transport Administration was presented for better understanding. The need for implementing IM was also shown in the form of maintaining cooperation wiith other third party along with efficient digitalization in transport sector (Puusaag and Palmi, 2021). Better IM also aid in information distribution for putting data into perspectives for better decision making (Ge et al., 2021).

4.A.2. Findings – Data utilization for Planning

➔ First of all, throughout the literature, evidence showed the successful use of data in planning of public transport. Then, use of data or big data in public transit was indicated by using data such as; AVL, APC, AFC (Van Oort and Cats, 2015).

➔ Planning of public trasport was done at various levels. In the lierature, it has been seen that, some researchers/academicians have categorized these levels into: strategic level planning, tactical level planning, operational level planning and real-time operations (Desaulniers and Hickman, 2007; Pelletier et al., (2011). Other researchers have categorized stages of planning from the perspectives of policy makers or economists into “five parts which are connected and interacted with each other, such as: strategic planning, tactical planning, operational planning, transit evaluation, marketing and policy“ (Lu et al., 2018).

➔ Literature review revealed the data driven strategic transit planning in many areas. Few of such findings have been presented in the tabular format:

Table 3 : Overview of studies using data at strategic level of planning

Authors	**Date**	**Data**	**Research purpose**
Furth et al.,	2006	AVL, APC	“Transit management”, “transit performance”
Pelletier et al.,	2011	Smart card data	“Passenger behavior analysis”
Schmöcker et al.,	2013	Smart card data	“Passenger behavior analysis”
Moreira-Matias et al.,	2015	AVL	“Network design”
Liu et al.,	2017	GPS	“Network optimization”, “mobility pattern analysis”
Schmöcker et al.,	2017	Smart card	“Route choice”, “travel behavior”

Qi et al.,	2018	Smart card	“Mobility pattern analysis”
Iliopoulou and Kepaptsoglou	2019	AVL, APC, AFC	“Transit assignment”, “Network design”

Source: Own illustration

- ➔ Through the review of literature, it was also revealed that, there are untouched areas for data-driven strategic planning with huge potentials which needed to be studied further. These un-explored areas are: use of “archived AVL-APC for improving transit management and performance“ (Furth et al., 2006); transit network re-design, inclusion of equity and accessibility in the design process, passengers response to fiscal policy with a need to modify algorithms and models including transit assignment, route design and shortest path for a successful transition to data driven planning (Iliopoulou and Kepaptsoglou, 2019).
- ➔ In tactical level of transit planning where data have been successfully used to obtain certain objectives in the literature have been presented in the tabular format:

Table 4: Overview of studies using data at tactical level of planning

Authors	**Date**	**Data**	**Research purpose**
Gkiotsalitis and Cats	2018	GTFS, AFC, AVL	“Frequency setting”
Kieu et al.,	2018	AFC	“Spatial-behavioral passenger segmentation”
Zou et al.,	2018	Smart card	“Trip purpose analysis”, “Home location detection”
Chen et al.,	2019	Smart card, mobile data	“Mobility pattern”
Huang et al.,	2019	GPS, Sensors	“Travel behavior analysis”
Gan et al.,	2020	AFC	“Mobility pattern analysis”
Tang et al.,	2021	GPS	“Timetable optimization”
Causa et al.,	2023	GNSS, GIS	“Route optimization”

Source: Own illustration

- ➔ The concept of operational level of transit planning found in the literature was presented differently by different authors with the similar main focus in this level of planning was to reduce/minimize total cost. From the perspective of transit service providers, cost minimization can be done by considering “the usage of vehicles, fuel consumption, and drivers’ wages“. Therefore, by considering “ strategical and tactical planning decisions as input“ operational planning problem was divided into: “(i) Vehicle Scheduling Problem (VSP), (ii) Driver Scheduling Problem (DSP), and (iii) Driver Rostering Problem (DRP)“ (Ibarra-Rojas et al., 2015). Whereas, from the researchers perspective, operational planning problem includes “vehicle scheduling, driver scheduling, rolling stock schedule, and maintenance schedule”(Lu et al., 2018). Based on the activities involving at this level, operational planning process has been decomposed into “(1) network route design, (2) timetable development, (3) vehicle scheduling, and (4) crew scheduling” which was performed sequencially (Ceder, 2007).

➔ Throughout the liretature review, uses of data has been seen in vehicle scheduling, crew scheduling, maintenance scheduling. No literature related to data-driven driver rostering, integrated vehicle and crew scheduling has been found. Literature comprising data-driven operational planning has been presented in a tabular format.

Table 5: Overview of studies using data at various operational level of planning

Authors	**Date**	**Data**	**Research purpose**
Shen et al.,	2017	AVL	“Vehicle scheduling”
Gkiotsalitis	2019	GTFS, AFC	“Frequency setting”
Shang et al.,	2019	Smart card data, Camera	“Vehicle scheduling”
Cendales	2019	AVL, GPS	“Crew scheduling”
Xie et al.,	2020	Camera, Sensors	“Maintenance planning”
Mohammadi et al.,	2021	Geospatial data	“Maintenance planning and scheduling”
Lai et al.,	2022	GPS	“Vehicle scheduling”, “Route optimization”

Source : Own illustration

➔ Literature on data-driven maintenance scheduling has been found only for railways. No literature on data-driven bus maintenance scheduling has been found so far.

➔ Literature on data-driven operation optimization reviewed are presented in the tabular format.

Table 6: Overview of studies on data-driven operation optimization

Authors	Data used	Optimization algorithm	Research purpose
Basak et al., (2019)	GTFS, Sensor data	Genetic algorithm(GA) and Particle swarm optimization (PSO)	“To improve bus on-time performance”
Cyril et al.,(2019)		“Analytical hierarchy process”, “goal programming”	“Performance optimization”
Shatnawi et al., (2020)	GIS; Geo-database	Genetic algorithm(GA) and Particle swarm optimization (PSO)	“to generate two optimized bus stops locations maps“
Lv et al., (2021)	Smart card AFC data	“Deep learning method”	Prediction of passenger flow and travel time, and the “bus operation optimization”.
Shi et al., (2021)	Smart card, AVL,	“Analytic hierarchy process”	“Bus route optimization”

Source: Own illustration

- ➔ Apart from the findings presented here in the tabular format, there were also some interesting points which worth mentioning such as: while working with GTFS real-time data, data cleaning process must carried out for raw dataset for better results (Basak et al., 2019); performance optimization work carried out by Cyril et al.,(2019) was without the use of data such as:ITS, GFTS, AVL, APC, AFC, GPS, GIS etc., but the result of sensitivity analysis suggests that, there is a strong need for introducing ITS “passenger information system, electronic ticketing machines (ETMs), global positioning system (GPS)-based ETM, and automated ticketing systems” for better improvements of the public transport performance; use of big data not only helps to achieve the goal of bus route optimization, it also helps in developing “a weight coefficient calculation model“ for comprehensive evaluation of “bus route optimization“ effect(Shi et al., 2021).

4.A.3. Findings – Data utilization for Assistance

- ➔ Literature review has provided the thesis with the evidence on uses of data for the purpose of assistance in public transit. Such of these findings includes: use of real-time data in “real-time rescheduling” in the form of real-time decision making (Nielsen et al., 2012); real-time GPS data “to minimize the gap“ in transit trip planning (Borole et al., 2013); real time data such as; “GPS, mobile phone data, loop detectors, cameras, etc” are assisting in proactive prediction of bus stop removal, train frequency increase prediction, prediction of extra bus line addition (Faizrahnemoon et al., 2015); real-time GTFS data assists transit operators in managing “unexpected events, delays and service disruptions concerning public transportation in the city of Milan“ (Bruglieri et al., 2015); real-time traffic GPS data can assists the traffic departments in congestion prediction, adoption of “smarter trip strategies, including route and departure time selection” (Li et al., 2016); data assists in obtaining and extracting “mass spatio-temporal trajectory data“ for the enhancement of efficient mass data sharing (Zhao et al., 2016); both historical and real-time data assists in “dynamic route planning with real-time“ prediction through a trip planner (Liebig et al., 2017); data in the form of “advanced traveler information system (ATIS)” assists the transit planners in reducing traffic congestions (Williams, 2019); use of “real time taxi GPS vehicle trajectory data” to forecast traffic in urban areas (Guo et al., 2020); smart card data can assists the transit planners to response better in an emergency situation like COVID-19 (Pozo et al., 2020); GPS data in the form of big data can assist various stakeholders in transit planning and sustainable urban mobility policy creation (Sostaric et al., 2021).

➔ Therefore, it has been clear that, data was used extensively to assist transit operators, users and policy makers in various instances to attain certain objectives along with the capacity to tackle emergency situation like COVID-19.

4.A.4. Findings – Data utilization for Control

➔ Literature review have revealed numerous works on data-driven real-time control strategies to be adopted for smoothe functioning of transit operation. These studies have been presented in a tabular format below:

Table 7: Overview of data-driven control strategies found in the literature

Authors	**Data used**	**Control strategies**	**Research purpose**
Zolfaghari et al.,	GPS, AVL	“High frequency based	“To minimize the waiting time of

(2004)		holding strategy”	passengers at all stops on that route“
Yu and Yang, (2009)	Historical data	“dynamic holding strategy”	To forecast ”the early bus departure time from the next stop”
Chen et al., (2013)	real-time AVL and APC data	“Holding strategy”	“To solve bus bunching”
Moreira-Matias et al., (2015)	AVL		To reduce the manpower by shifting the “traditional focus on planning to an autonomous data-driven real-time control”
Nesheli and Ceder, (2015)	Real-time fare card data		“To implement proper control action”, “to improve reliability”
Andres and Nair, (2017)	GPS, AVL	“bunching prediction and corrective control“	“To reduce bus bunching”
Luo et al., (2017)	AVL data and IC smart card data	“dynamic control framework“	To manage “the bus bunching problem with capacity constraints“
Asgharzadeh and Shafahi, (2017)	real-time AVL and static APC data	“real-time bus holding control strategy“	To “minimizes passenger waiting time“
Berrebi et al., (2018)	AVL, APC, Loop detector	“Schedule-based, headway-based and prediction-based holding methods”	“To reduce bus bunching”
Gavrilidou and Cats, (2018)	AFC, APC	“real-time transfer synchronization controllers“	To study “the performance of transfer synchronization control“
Malucelli and Tresoldi, (2019)	AVM and mobile phone data	“real-time vehicle and driver scheduling based control strategy“	To solve “delays and small disruptions”
Wang et al., (2021)	real-time IC card data and GPS data	“proactive control strategy“	“To increase service reliability“
Xue et al., (2022)	ITS	“single-line real-time control strategy“	To prevent bus bunching, reduce bus delays
Yap et al., (2022)	AFC	“optimization based control strategy“	“To quantify disruption impact propagation“

Source: Own illustration

- ➔ It has been seen from table 6, how data has helped to ease out inconvenient situations arising during transit operation by adopting control strategies based on its respective situations.

4.A.5. Findings – Data utilization for Maintenance

- ➔ It has been also found from the review of literature the importance of transit asset maintenance with the help of data. Requirements for public transit asset management was needed for the smooth functioning of assets such as; "ticket machine, bus, tram, rail tracks etc" at "optimum cost". Data played a key role for condition-based maintenance and predictive maintenance (UITP, 26.02.2020). Other authors have addressed the problems related to rising real-time big data maintenance cost by proposing CBM, a "knowledge-based approach", where "streaming big data analysis (SBDA)" was used for the purpose of solving problems in real-time such as; big data management, challenges arising during storage and computation, „and predictive data analytics in CBM systems" (Maktoubian, 2017). Use of sensor has been found for anomaly detection in railway industry as a part of predictive maintenace plan (Davari et al., 2021).

4.A.6. Findings- Data utilization to improve service quality and reliability

- ➔ In the literature, there were four major classes of strategies have been found, which has been used for the improvement of the transit service reliability: "vehicle-holding strategies, reduction of the number of stops made by each bus, signal preemption, and provision of exclusive right-of-way" (Turnquist, 1981). Other strategies, based on decreasing frequency of appearance order includes: transit signal priority (TSP), new buses (low-floor buses and articulated buses), bus rapid transit (BRT) or BRT-like systems (rapid transit system or networks), intelligent transportation system (ITS), reserved bus lanes, limited-stop services (express buses), AVL/APC systems in buses, and smart cards. Also, before embarking on data utilization journey to improve service quality and reliability, there was a need "to find out which are the latent aspects that are hidden under a series of attributes describing the quality of the service". With the help of Structural Equation Model, "the influence of a series of characteristics describing the quality of the bus transit service on the Overall Service Quality (OSQ)" can be determined (De Oña et al., 2013). Moreover, an emphasis was also given to integrate “traveller information in the decision process” (Diab et al., 2015) by adopting automated exchange of information between travelers and transportation company (Stelzer et al., 2016) for the improvement of service quality provided by public transport. Uses of data to improve service quality and reliability have been presented in a tabular format below.

Table 8: Overview of studies on service reliability improvement using data

Authors	Data used	Research purpose	Methods used
Oort et al., (2015)	AVL	To increase ridership and revenues through finding out “inefficiencies, bottlenecks and potentials“	“Line-based analysis“, “network-based analysis“ and “inter-operator transfers
Chen et al., (2016)	Static and real-time GTFS	“To encourage the public to take transit”	
Firew, (2016)	AVL, APC	“To attract private car users and keeping existing passengers“	Measurement of “on-time performance headway adherence“, “vehicle trip-time variability“, “passenger wait time“ and “passenger travel time“
Dixit et al., (2019)	Smart card data, AVL	“To address specific unreliability issues for a particular route/OD pair”	Extending the existing “reliability buffer time”, “developement of reliability metric at disaggregated level“
Li and Tan , (2020)	AVL, smart card data	“Reliability evaluation”	Evaluation based on stop-level, route-level and network-level using “bus AVL and smart card data“
Liu and Miller, (2020)	GTFS and AFC data	“Reliability investigation”	Comparison of strategies between “RTI apps users and non-RTI user“.
Liu et al., (2022)	GTFS	“Reliability evaluation”	Accessibility measurement by introducing “realizable real-time accessibility based on space–time prisms (STP)“
Alkubati et al., (2023)		Illustration of transportation reliability concept, highlighting the need of uncovered studies	“Bibliometric analysis”

Source: Own illustration

- ➔ The "top 5 countries in reliability-related research from 2005 to 2021" includes; China, Netherlands, England, Australia and Poland respectively. "Top 5 leading journals in reliability-related research from 2005 to 2021" were; "Transportation research part a-policy and practice, Transportation research record, Transportation policy, Transportation research part c-emerging technologies, Transportation (Alkubati et al., 2023).

4.A.7. Findings – Real-time information delivery to the passengers

- ➔ Literature review on how accurately real-time information was delivered to the passengers, revealed that, all the literaures were supporting the fact that, real-time information reduces passenger waiting time (Dziekan and Kottenhoff, 2007; Cebon and Samson, 2011; Watkins et al., 2011; Brakewood et al., 2015; Brakewood and Watkins, 2019), contributed towards passenger's positive psychological effects and overall satisfaction(Dziekan and Kottenhoff, 2007; Bruglieri et al., 2015; Brakewood and Watkins, 2019). Other authors have stressed upon finding out 'types' of information that passengers wanted in real-time in the form of 'information demand' and transit authorities should be in a position to supply the particular information to the passengers which will increase the passengers utility and overall satisfaction with the services provided by the public transport.
- ➔ Despite all the advantages of Real-time information reported in literature, two literature stand out differently. In one literature, the authors have found that, there was no "statistically significant difference" in perceived waiting times between real-time information users and non-real-time information users (Brakewood et al., 2015). In another literature, the author did an actual enquiry about the effectiveness of real-time information by questioning "whether real-time information (RTI) apps reduces waiting time of passengers or not". In order to answer this question, the author first looked into the strategy used for RTI apps. Among the two strategies; greedy tactic and prudent tactic, greedy tactic of RTI aimed to achieve zero waiting time, was the worst strategy. Moreover, when appropriate trip planning strategy is missing in the app, the waiting time of the passengers could be significantly longer. Whereas, prudent tactic strategy used in RTI was the so called "best strategy", was found performing same as " follow-

the-schedule tactic (ST) that does not use RTI”. Therefore, based on the findings, the authors have concluded that, inspite of the advantages of RTI best strategy with actual time saving capacity only for specific users at specific hours for specific stops, it cannot “globally outperform simply following the published schedule” strategy (Liu and Miller, (2020).

4.A.8. Findings – Perspectives of public transport service providers

➔ The review of literature revealed that, based on the goal persued by the transit operators, the perspectives might vary. These goals are: patronage goal and coverage goal. Based on these two distinct goals, “both short-term service design decisions and long-range network planning” were carried out by the transport service providers (Walker, 2008). Other concerns were related to: quality improvement by increasing robustness (Oort et al., 2015; Cats et al., 2015); reducing the money handling fees with easy revenue split (Schmöcker et al., 2017); development of data-driven predictive analysis “prevent unwanted deterioration in service quality” in due time (Noursalehi et al., 2018); fusing data from various sources and applying it efficiently for the purpose of planning, control, communication, proactive decision support, measurement, information and service customization (Koutsopoulos et al., 2019); increasing “patronage in terms of planning principles and measures” (Khan et al., 2020); service improvement during disruptions by offering information to the passengers (Leng and Corman, 2020); increasing transparency for reliable travel planning (Redmond, 2020); saving cost and recover the economy by creating jobs (UITP, 2020); providing clean environmental friendly sustainable means of transport (Keller et al., 2022).

4.A.9. Findings – Perspectives of passengers

➔ Perspectives of passengers found in the literature were mostly concerned about comfort and time while taking bus, trams or trains. These findings were: on time departure from origin station, reasonable in-vehicle time limits, timely arrival at the destination enabling the passengers to reach his or her final destination without delay (Casello et al., 2009); “willingness to pay(WTP) for reduced crowding” with increasing “value of travel, walking or waiting time“ (Li and

Hensher, 2013); for personal travel,"the value of reliability (VOR) and the value of travel time (VOT)" varried greatly than during business travel, depending on the means for travel used (Ehreke et al., 2015); commuters are more concerned about reliable means of public transport than travel time in general (Kou et al., 2017); in-vehicle density and users satisfaction are inversely proportional, discomfort due to "not being seated", had less opportunities for making the journey time useful and being physically close to "other travelers *per se*" (Haywood et al., 2017); passengers willingness to pay is inversely proportional to public transport services (Hörcher et al., 2017); passengers' perception of reliability for high frequency services are broadly decided upon how predictabile their journey times are, instead of complying with published schedule (Koutsopoulos et al., 2019); different age group of individuals have different perception and different level of satisfaction (Chocholac et al., 2020); crowding, discomfort arising due to noise and smell of fellow passengers (Wang and Zacharias, 2020); crowding level and maintaining social distance inside public transport during the outbreak of pandemic (Aghabayk et al., 2021).

➔ Luke and Heyns, (2020) have revealed that, before measuring passengers perception especially, in terms of reliability, there was a need to spread awareness about "friendly modes of transport" and public transport service quality among the users, which will help to attract more users to commute with the public transport and discourage the users from buying private cars. Not only this, passengers must also be aware about the public subsidies while commuting with public transport as it can "lead to a higher valuation of the services offered" (Drevs et al., 2014).

4.A.10. Findings- Public transport from researchers perspective

➔ The perspective of researchers on creating sustainable public transport is in the form of exploration of visions but only conceptually a feasible and realist version which will reduce the confusion arising due to "the evolution of automated/autonomous vehicles (Avs)". "The evolution of automated/autonomous vehicles (Avs)" might reduce the future research scope for the researchers in the context of future urban mobility improvement (Ceder, 2021). Another perspective found was regarding the increasing attractiveness of public transport. By following few important tips, patronage of “small and medium-sized cities” could be increased. First tip; to acquire knowledge about patronage coverage in larger cities and applying this same knowledge to the smaller cities for the same purpose (patronage coverage) by reorganising the transit system into providing high frequency services with “fewer, straighter, faster and simpler routes“. Second tip; to have a sharp focus on increasing public transport competitiveness with the reduction in car usage, coupling with knowledge and competency enhancement amidst the planners of public transport. Other tips such as; to bring changes in organisational power construct where professionals have more leeway to plan and decide (Tennoy, 2022).

4.A.11. Findings- Public transport from policy makers perspective

- ➔ Literature review revealed that, perspectives of policy makers are sometimes similar with the users and sometimes it differs. Therefore, there is a need to take note of these differences by policy makers before implementing transit policy related integration (Chowdhury et al., 2018). There is also need for collaboration “between the politicians and the civil servants, and their relationships to their respective group of stakeholders“ while translating transit policy in order to avoid conflicts (Paulsson, 2018). Collaboration among different stakeholders and sectors is also necessary “to provide strong, consistent and sustainable mobility services to citizens“ (UITP, 2021). Political conflicts must also be resolved in order to avoid “unfavorable decisions in transit planning” by educating policy makers, decision makers in various transit and urban planning related topics. Urban planner can best serve the role in this regard to inform the politicians and decision makers and awaken them to a need for more knowledge production among them(Skartland, 2022). While working with data, policy makers should consider few important points for better transport policy making are: minimum data collection for defined purposes only, using big data in transport models through advanced guidelines, enabling smartphone apps to collect location data, privacy protection, taking advantage of artificial intelligence to mine data, data steward function creation and its promotion in both public and private sectors, investment for public sector workforce in data-related training (ITF, 2021).

4.A.12. Findings- Challenges faced during data utilization

➔ The challenges encountered in the literature while working with data in public transit planning and operation has been presented in tabular format.

Table 9: Lists of challenges related to data utilization

Authors	Challenges
Prarthana and Jayakumer, (2017)	Security issue, management of data and making sense of data
Torre-Bastida, (2018)	“real-time data analytics“, security and privacy issue, new data sources, lack of infrastructure,
Ge et al., (2021)	processing of data, overlap between different data sources, functionality requirement of data
Redmond et al., (2019)	Exclusion of multi-modality notion in public transport

Source: Own illustration

4.A.13. Findings - Opportunities of data in public transportation

➔ Although few challenges have been reported in the literature with respect to data utilization in public transport, but the scope and opportunities of data has found to outperform its challenges. Theses findings are presented in the tabular format below.

Table 10: Opportunities of data utilization in public transport

Authors	Opportunities
UITP (2021)	“bringing net positive value for mobility for life” through data sharing
Ewoldsen, (2022)	to improve the satisfaction of customers, to increase revenue and service performance, carbon and other emissions reduction and equity improvement through multimodal transportation, to purchase discounted public transit ground transportation tickets to access the two area airports by passengers
Ortelio (2018)	business opportunities, opportunity for collaboration, transferability of data, adopting standards for data, opportunities for practitioners and “big data application developers” into becoming the skilled personnel
Ge et al., (2021)	“each data source has specific applications and could provide information not captured from other sources“

Jacobs, (Accessed on, 05.09.2023)	Operations optimization, costs reduction and increase revenues, high-quality service provision to customers despite “less financial leeway in the public transportation sector”

Source: Own illustration

4.B. Discussion

The main purpose of this review based thesis work is to dive deep and explore the research works done on the uses of data in different areas of public transport. The aim was to find out the current practices on the uses of data in public transit with an aim to improve reliability, efficiency and sustainability of public transport system. Throughout analysis of the literature, it was observed that, despite the latest advances in the technology, its implementation in transportation sphere is still at its inception. Also, most of the research works were done by research institutions and universities as a part of special projects, with limited real-life implementations. Those projects, that have been implemented in real world were merely in the form of test samples without large scale deployment of proposed methods and designs. Hence, it is quite difficult to measure the benefits of the latest technology in public transport. Transport authorities must implement all the proposed methods in a large scale and keep a record of it. When this record is shared among different stakeholders, general knowledge about exploitation of latest technologies is gained by all stakeholders of public transport.

The thesis begin with reviewing different sources and types of data in transportation domain, followed by the data collection techniques. During the review of literature, it has been noticed that, there are certain important factors playing direct and indirect role in the fucntioning of transport sector for data-driven planning and operation. Therefore, these important factors were highlighted and reviewed briefly for better understanding the whole system for facilitating the data exploitation in public transport. Hence, it is partly answering the research question “up to what level data has been exploited?“

To be specific, the interest of this thesis work is to have a general idea about extensive uses of data in three important domains of transport. That is why, second part of literature review was done in the area of planning, assistance and control, in order to find the answer on “in which areas of transit planning data has been used extensively?“. There were numerous literature on data-driven transit planning where historical or offline data were used for strategic and tactical planning and real-time data used for operation and control (Van Oort and Cats, 2015). Review of literature was performed separately for each stages of planning namely; strategic, tactical and operational to have an idea about main topics investigated by researchers and where the research work is lacking. It has been seen that, no article was found using data on the topic of crew rostering , integrated vehicle and crew scheduling. In other words, literatures are lacking on data driven crew rostering, integrated vehicle and crew scheduling.

Similarly, literature review was carried out with an aim to discover in which types of situation, data-driven control or assistance was administered. By finding out different situations, will also help to find out the answer to the research question. Meanwhile, area of data-driven maintenance was also explored, as data-driven transport planning uses infrastructure for its functioning (Ge et al., 2021), which needed to be maintained.

In order to answer the last part of research question, “up to what level data has been exploited“, literature review was carried out on the topic related to service quality, reliability and accuracy of real-time information. It is because, previous works on this topic were also focusing on the improvement of customers/passengers overall satisfaction with the service provided by the public transport. By looking at the findings, it has been clear that, many authors were keen interested in finding out current reliability status of public transport. As the countries like; China, Netherlands, England, Australia and Poland were among top five to engage “in reliability-related research from 2005 to 2021“, indicating, these nations are yet to be satisfied with public transit in terms of reliability. The findings of “real-time information delivery to the passengers“, showed that, in general, real-time information is successful in overall reduction of passenger waiting time along with its other advantages. However, two studies have reported otherwise. Detail review into this topic revealed by the authors concludes that, there was no difference between real-time information apps and non real-time follow-the-schedule tactic, and, real-time information does not necessarily reduces waiting time of the passenger. “Although the best RTI strategy can indeed save time“ for specific users in particular stops and during specific hours, it is impossible for this RTI strategy to globally outperform by barely pursuing the “published schedule” (Liu and Miller, (2020). This finding have indirectly answered the last part of research question. There is still huge scope for improvements for real-time information delivery to the passengers and data is yet to be fully exploited.

Once, the areas of data utilization in public transport has been thoroughly explored, there is also a need to address the challenges faced while working with data. After the challenges has been addressed, the unexplored opportunities of data utilization has also been addressed , which will act as a way shower to future researchers for taking up new projects in this unexplored avenues of data utilization.

To improve the overall satisfaction with the transit services, it is also therefore necessary to understand the perspectives of transit operators, passengers and policy makers. Whereas, perspectives of researchers, helps to bridge the gap (if any) between all other stakeholders of public transport as it is the researchers/academicians, who, through their work, are always in a quest to find out what has been done so far, what is the situation at the present moment and what needed to be done in the future.

4.C. Limitations to the study

This master thesis is literature based, and not to test any findings presented by other authors.

There is no standard set for the “extent of data utilization“ in public transit planning. Therefore, it was difficult to answer the reserach question with precision.

This thesis covers many areas of data-driven transit planning in general and limited to provide detail/specific information about topics included.

There might be chances of bias while typing “key words“ due to the large number of topics included, and possibility of finding/including all the relevant literature is beyond the scope of this thesis.

5. Conclusion

Public transit plays an essential role in making urban living easy and comfortable. Technological advancement in the form of real-time information provision for the passengers has contributed to improving the quality of life of urban dwellers. With rapid urban expansion and its corresponding increase in population (Ge et al., 2021), management of the daily commute network seemed impossible without taking the advantages of data-driven technologies. However, how far has the transport sector has been able to take advantages of these latest technologies? In other words, to what extent has data been exploited to improve efficiency, service reliability and overall transit performance?

In order to find an answer to this question, one must know which types of data are used in the transport sector and what its sources are. The most popular data types found throughout the literature were, AVL, AFC, APC, smart card, GIS, GPS, GNSS, GTFS and Mobile network data. These data can be sourced from GTFS, ATIS, ITS, AVL system, AFC, APC; smartphone applications and the web, mobile phone call data records, sensors, weather data, traffic data and survey data. In brief, the data acquisition process of AVL, AFC-APC, GTFS and ATIS was presented in this thesis for better understanding.

Data acquisition capacity depends on the presence of appropriate infrastructure which further allows better data handling. Therefore, investment in infrastructure is apparent for data-driven decision making. After acquiring data, it is important to know and follow a few steps before data can be used to serve a particular purpose. These steps include; data integration, standardization, validation, matching and data processing. Meantime, problems arising due to storage, privacy and security issue also need to be addressed. Information management allowed data distribution and helped in better decision-making by different stakeholders by putting data into perspectives.

Through systematic implementation of transit big data, various transit related problems can be solved. Valuable insights gained from systematic implementation of data serve as a foundation which further aids in sound decision and policy making. Data in the form of digital twin enables digital transformation of “physical process, person, place, system, or device“ (Ge et al., 2021) by integrating data from various sources. Digital twin is a form of data exploitation where simulation can be carried out before taking physical action in the real world. Therefore, digital twin can be used for better urban planning and mobility related decisions. To be specific, transit data exploitation has been found in the area of visualization and prediction, “public transport disturbance management“, “transit system design and planning“, planning and operation, service optimization, development of visualization tools, performance measurement among many. However, despite the availability of data alongside the knowledge of the latest technological advancement, data is yet to be fully exploited at strategic, tactical and operational level of planning.

Through evaluation of transit services, transit agencies can identify their strengths and weaknesses. This thesis underwent literature review in the area of network and service accessibility, reliability and robustness evaluation of public transit. Transit accessibility related studies helped with information provision regarding urban land use plan, distance between passengers' home and business/ working areas and how people are ariving at work, “accessibility unreliability“ and a means to solve it with real-time data. From the latest study on accessibility focusing on 'older adult', it was revealed that, there were not enough studies conducted on this topic. More study is needed in this regard in order to draw a sound conclusion on how older adults are conceptualizing accessibility, types of destinations used by them and how they are measuring transit accessibility. In terms of evaluating public transit based on reliability, four types of strategy were found in the literature; reliability evaluation through performance measurement, reliability evaluation by considering transit agencies' point of view, reliability evaluation by considering passengers point of view, and integrated passenger and transit operators perspectives based reliability evaluation. The integrated passengers transit operators' perspectives based strategy seemed out performed all other strategies due to the fact that, it can bridge the gap (if any) between these groups. Attention must also be paid towards multi-modal-itinerary planning which involves “using more than one mode of transportation“ during multiple legs traverse. This is because, “state-of-the-art traveler information systems“ can only create multi-modal travel itineraries, without considering the reliability of the chosen transportation services during the itinerary creation. Literature on robustness evaluation of transit services was found only in the case of railway transit. Robustness evaluation study was carried out for situations like; link failure, to solve congestion, to reduce delays, and disturbance management.

After pointing out all other important factors of data-driven transit planning and operation, it is also important to include an attractive marketing strategy and a sound transit policy.

After implementation of every new policy, a cost-benefit analysis must be carried out to determine the welfare of the popullation. This is because, with the right transit policy, better land use planning can be administered, an attractive fare policy can be exercised, and transit big data can be protected through the maintenance of data privacy and security.

The findings of this thesis revealed that, data is utilized for planning, assistance, control and maintenance. At strategic level of planning, AVL, APC, AFC, GPS and smart card data was used to study route choice, travel behavior, to analyze passenger behavior, to analyze mobility patterns, network design, transit management, performance evaluation, network optimization. However, there are some unexplored areas of strategic planning, where exploitation of data is yet to take place; use of "archived AVL-APC for improving transit management and performance" (Furth et al., 2006), network re-design, route design, shortest path for a successful transition to data driven planning (Iliopoulou and Kepaptsoglou, 2019). At tactical level of planning, GTFS, AFC, AVL, AFC, GPS, GNSS, GIS, Sensors, smart card data and mobile data were used to study frequency setting, trip purpose, travel behavior, mobility pattern, timetable optimization, route optimization, home location detection, spatial-behavioral passenger segmentation. At the operational level of planning, data has been utilized for vehicle scheduling, frequency setting, crew scheduling, maintenance planning. However, this thesis did not come across a single literature on data-driven integrated vehicle and crew scheduling and maintenance scheduling. Data utilization has also been found in the area of public transit operation optimization. For this, a combination of data and optimization algorithm was used for the purposes such as; to improve bus on-time performance, to generate two optimized bus stops locations, to predict passenger flow, to predict bus travel time, to optimize bus operation, bus route optimization.

Instances of data-driven assistance found include; "real-time re-scheduling", real-time proactive prediction, real-time management of unexpected events, real-time congestion prediction, adoption of smarter trip strategies, "accessing and retrieving mass spatio-temporal trajectory data", real-time dynamic route planning, in reducing traffic congestion, to response in a emergency situation like Covid-19, assisting various stakeholders in transit planning and sustainable urban mobility policy creation.

Numerous literature reveals the use of data for public transit control operations. Both real-time and static data have been used for control. Among the specific situations found in the literature where control operations are carried out were; to reduce bus bunching, to solve delays and disruptions, to improve reliability, to quantify disruption impact propagation, to minimize waiting time, "to forecast early bus departure time from the next stop", to reduce manpower, to synchronize transfer. It has been clear that data has helped in the adoption and functioning of appropriate control strategies to ease out inconvenient situations arising during transit operation.

Application of data was also seen for the maintenance of public transit assets. The requirements for the maintenance of public transit assets such as "ticket machine, bus, tram, rail tracks etc" at "optimum cost" are crucial for the smooth functioning of the whole transport system. Most popular maintenance strategies used in transport sector are condition-based maintenance and predictive maintenance, where data plays a key role in monitoring the performance of assets.

After accumulating all the information surrounding data and its utilization in public transport from literature, the next task is to critically analyze its real world successful implementation. This was done by reviewing literature in the area of 'service quality and reliability improvement' and 'accuracy of real-time information'. There were four major classes of strategies that have been found to improve the transit service reliability: "vehicle-holding strategies, reduction of the number of stops made by each bus, signal preemption, and provision of exclusive right-of-way". Based on decreasing frequency of appearance order, other strategies for improving bus service reliability includes: bus rapid transit (BRT) or BRT-like systems (rapid transit system or networks), transit signal priority (TSP), reserved bus lanes, new buses (low-floor buses and articulated buses), limited-stop services (express buses), AVL/APC systems, smart cards and intelligent transportation system (ITS). Numerous studies have showed that, data has been used for the purposes like: increasing ridership and revenues, encouraging the public to take transit, attracting private car users to take public transit, "to address specific unreliability issues for a particular route/OD pair", reliability evaluation, reliability investigation. Efficiency of real-time information delivery to passengers has been reported in several literature through reduction in waiting time, increases in overall satisfaction, increasing the utility of passengers and contribution towards passengers positive psychological effects. However, the conclusion found in two literatures were reporting otherwise. In one literature, the authors found no "statistically significant difference" in perceived waiting times between real-time information users and non-real-time information users

(Brakewood et al., 2015). In another literature, the authors have concluded that, the so called best RTI-strategy despite its advantages of saving time at certain stops, hours of the day for certain users, “cannot globally outperform simply following the published schedules“ (Liu and Miller, (2020). These findings indicate that, there is still room for improvements and data is yet to be fully exploited for public transit planning.

There were some challenges faced while using data in transit planning and operation. These challenges must be addressed by finding right solutions to them. Huge scope and opportunities of data utilization reported in the literature were: “bringing net positive value for mobility for life” through data sharing, to improve “customer satisfaction, revenue and service performance”, “reduce carbon and other emissions and improve equity” through multimodal transportation, to purchase discounted public transit ground transportation tickets to access the two area airports by passengers, business opportunities, unique information provision by each data, operation optimization, cutting cost.

The perspective of public transport stakeholders played a big role in the overall functioning of the whole transit system. However, the perspectives of transit service providers, passengers and policymakers do not necessarily coincide due to the fact that each stakeholder has different concerns and priorities. Seldom overlapping in the perspectives of different stakeholders are found in the literature. Therefore, researchers and academicians have huge potential here to bridge the gap between them through research work. A common platform is needed to be established where transport planners, academic researchers and policymakers can meet and easily share information and ideas which can ultimately increase the overall performance of the transport sector.

Bibliography

Aghabayk, K., Esmailpour, J., and Shiwakoti, N., 2021. Effects of COVID-19 on rail passengers' crowding perceptions. Transportation Research. Part A, Policy and Practice, vol. 154, pp. 186 - 202.

Ait-Ali, A. and Eliasson, J., 2022. The value of additional data for public transport origin–destination matrix estimation. Public Transport, 14, pp. 419–439.

Aklilu, A. and Necha, T., 2018. Analysis of the spatial accessibility of addis Ababa's light rail transit: The case of East–West corridor. Urban Rail Transit, vol. 4, no. 1, pp. 35–48.

Albrecht, T., 2009. Automated timetable design for demand-oriented service on suburban railways. Public Transport vol. 1, no. 1, pp. 5–20.

Ali, A., Kim, J., and Lee, S., 2016. Travel behavior analysis using smart card data. KSCE Journal of Civil Engineering, vol. 20, no. 4, pp. 1532–1539.

Anda, C., Erath, A. and Fourie, P. J., 2017. Transport modelling in the age of big data. International Journal of Urban Sciences, vol. 21, no.1, pp. 19-42.

Andersson, E. V., Peterson, A. and Krasemann, J. T., 2013. Quantifying railway timetable robustness in critical points,Journal of Rail Transport Planning & Management, vol. 3, no. 3, pp. 95-110.

Andres, M., and Nair, R., 2017. A predictive-control framework to address bus bunching. Transportation Research Part B: Methodological, 104, 123–148.

Antrim A. (accessed on 07.07.2022). Automatic passenger counting (APC) and automatic fare collection (AFC) technology, Benefits, technologies, components, data standards, experiences and recommendations, Developed for ODOT by Trillium, Portland, Oregon.

Antrim A., (accessed on 07.07.2022). Automatic passenger counting (APC) and automatic fare collection (AFC) technology. Benefits, technologies, components, data standards, experiences, and recommendations, Trillium Solutions, 610 SW Broadway Portland, OR 97212.

Asakura, Y., Iryo. T., Nakajima, Y. and Kusakabe, T., 2012. Estimation of behavioural change of railway passengers using smart card data. Public Transp 4(1):1–16.

Asgharzadeh, M., and Shafahi, Y., 2017. Real-time bus-holding control strategy to reduce passenger waiting time. Transportation Research Record, 2647,9–16.

Axelsson, P., 2020. Building Information Management with the Swedish Transport Administration. ISA², European Commission.

Bagchi, M., and White, P. R., 2005. The potential of public transport smart card data. Transport Policy, vol. 12, no. 5, pp. 464–474.
Baptista, M., Sankararaman, S., de Medeiros, I. P., Nascimento, C., Prendinger, H., Henriques, E. M. P., 2018. Forecasting fault events for predictive maintenance using data-driven techniques and ARMA modeling. Computers & Industrial Engineering, vol. 115.

Barabino, B., Lai, C., Casari, C., Demontis, R. and Mozzoni, S., 2017. Rethinking Transit Time Reliability by Integrating Automated Vehicle Location Data, Passenger Patterns, and Web Tools. IEEE Trancastions on Intelligent Transportation Systems, vol. 18, no. 4.

Barabino, B., Di Francesco, M. and Mozzoni, S., 2016. An Offline Framework for the Diagnosis of Time Reliability by Automatic Vehicle Location Data. IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 3, pp. 583-594.

Barabino, B., Di Francesco, M. and Mozzoni, S., 2014. An Offline Framework for Handling Automatic Passenger Counting Raw Data. IEEE Transactions on Intelligent Transportation Systems, vol. 15, no. 6, pp. 2443-2456.

Barabino, B., Di Francesco, M. and Mozzoni, S., 2014. An Offline Framework for Handling Automatic Passenger Counting Raw Data. IEEE Transactions on Intelligent Transportation Systems, vol. 15, no. 6, pp. 2443-2456.

Barabino, B., Di Francesco, M. and Mozzoni, S., 2015. Rethinking bus punctuality by integrating Automatic Vehicle Location data and passenger patterns. Transportation Research Part A: Policy and Practice. vol. 75, pp. 84–95.

Barnes, R., Buthpitiya, S., Cook, J., Fabrikant, A. Tomkins, A., and Xu, F., 2020. BusTr: Predicting Bus Travel Times from Real-Time Traffic. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Virtual Event.

Barnes, R., Buthpitiya, S., Cook, J., Fabrikant, A., Tomkins, A. and Xu, F., 2020. BusTr: Predicting Bus Travel Times from Real-Time Traffic. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Virtual Event.

Basso, L. J., and Silva, H. E., 2014. Efficiency and Substitutability of Transit Subsidies and Other Urban Transport Policies. American Economic Journal: Economic Policy, vol. 6, no. 4, pp. 1-33.

Bast, H., Patrick, B., Sabine S., 2014. Real-time movement visualization of public transit data. The 22nd ACM SIGSPATIAL International Conference (SIGSPATIAL '14). Dallas, Texas, Nov 4, 2014 - Nov 7, 2014. In: HUANG, Yan, ed.. Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. New York, NY:ACM Press, pp. 331-340.

Beasley, J. E. and Cao, B., 1996.A tree search algorithm for the crew scheduling problem. European Journal of Operational Research, vol. 94, no. 3, pp. 517-526.

Berkcan, O., and Aksoy, T., 2022. Assessing the effect of proactive maintenance scheduling on maintenance costs and airline profitability: The case of Turkish Airlines Technic .International Journal of Research in Business and Social Science (2147- 4478), vol.11, no. 4, pp. 213–238.

Berrebi, S. J., Hans, E., Chiabaut, N., Laval, J. A., Leclercq, L., and Watkins, K. E., 2018. Comparing bus holding methods with and without real-time predictions. Transportation Research Part C: Emerging Technologies, 87, 197–211.

Berrebi, S. J., Crudden, S. Ó., and Watkins, K. E., 2018. Translating research to practice: Implementing real-time control on high-frequency transit routes. Transportation Research Part A: Policy and Practice, vol. 111, pp. 213–226.

Bertini, R. L., and Tantiyanugulchai, S., 2004. Transit Buses as Traffic Probes: Use of Geolocation Data for Empirical Evaluation. Transportation Research Record, vol. 1870, no. 1, pp. 35–45.

Binder, S., Maknoon, Y. and Bierlaire, M., 2017. The multi-objective railway timetable rescheduling problem. Transportation Research Part C: Emerging Technologies 78:78–94.

Borgi, T., Zoghlami, N., & Abed, M., 2017. Big data for transport and logistics: A review. In Advanced Systems and Electric Technologies (IC_ASET), International Conference on (pp. 44-49). IEEE.

Borole, N., Rout, D., Goel, N., Vedagiri, P. and Mathew, T. M., 2013. Multimodal Public Transit Trip Planner with Real-time Transit Data. Procedia - Social and Behavioral Sciences, vol. 104, pp. 775-784.

Brakewood, C., Rojas, F., Zegras, P. C., Watkins, K. and Robin, J., 2015. An Analysis of Commuter Rail Real-Time Information in Boston. Journal of Public Transportation, vol. 18, no. 1, pp. 1-20.

Brakewood, C. Watkins, K. , 2019. A literature review of the passenger benefits of real-time transit information. Transportation Reviews, 39, 327–356.

Briand, A. S., Côme, E., Mahrsi, M. K. El and Oukhellou, L., 2016. A mixture model clustering approach for temporal passenger pattern characterization in public transport. International Journal of Data Science and Analytics, vol. 1, no. 1, pp. 37–50.

Brown, J., Daniel, H. and Donald, S., 2003. Fare-Free Public Transit at Universities: An Evaluation. Journal of Planning Education and Research, vol. 23, no. 1, pp. 69-82.

Bruglieri, M., Bruschi, F., Colorni, A., Luè, A., Nocerino, R. and Rana, V., 2015. A Real-time Information System for Public Transport in Case of Delays and Service Disruptions. Transportation Research Procedia, vol. 10.

Camus, R., Longo, G. and Macorini, C., 2005. Estimation of transit reliability level-of-service based on automatic vehicle location data. Transportation Research Record, Journal of the Transportation Research Board, vol. 1927, pp. 277–286.

Casello, M. J., Nour, A. and Hellinga, B., 2009. Quantifying Impacts of Transit Reliability on User Costs. Transportation Research Record: Journal of the Transportation Research Board. No. 2112, pp. 136-141.

Cats O., Koutsopoulos H. N., Burghout W. and Toledo T., 2011. Effect of real-time transit information on dynamic path choice of passengers. Transportation Research Record, 2011; 2217(1): 46-54.

Cats, O. and Loutos, G., 2013. Real-time bus arrival information system-an empirical evaluation. 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013), pp. 1310-1315.

Cats, O. and Loutos, G., 2016. Real-time bus arrival information system: an emperical evaluation. Journal of Intelligent Transportation SystemsTechnology, Planning and Operations, vol. 20, no. 2.

Causa, F., Franzone, A. and Fasano, G., 2023. Strategic and Tactical Path Planning for Urban Air Mobility: Overview and Application to Real-World Use Cases. Drones, vol. 7, no. 1.

Cebon, P. and Samson, D., 2011. Using real time information for transport effectiveness in cities. City, Culture and Society, vol. 2, no. 4, pp. 201-210.

Ceder, A., 2007. Public transit planning and operation: Modeling, practice and behavior. Boca Raton: CRC press.

Ceder, A., 2011. Optimal Multi-Vehicle Type Transit Timetabling and Vehicle Scheduling. Procedia - Social and Behavioral Sciences, vol. 20, pp.19-30.

Ceder, A., 2021. Urban mobility and public transport: future perspectives and review. International Journal of Urban Sciences, vol. 25, no. 4, pp. 455-479.

Cendalles K. L., 2019. Robustness simulation of bus crew schedules. Master thesis, Degree Project in the Built Environment, Second Cycle, 30 Credits Stockholm, Sweden.

Cevallos, F., Wang, X., Chen, Z. and Gan, A., 2011. Using AVL data to improve transit on-time performance. Journal of Public Transportation, vol. 2011, no. 14, pp. 21–40.

Chakirov, A. and Erath, A., 2012. Activity identification and primary location modelling based on smart card payment data for public transport. In Proc. 13th International Conference on Travel Behaviour Research, pp. 1–24.

Chapleau, R., Trépanier, M., and Chu, K. K., 2008. The ultimate survey for transit planning: Complete information with smart card data and GIS. In Proceedings of the 8th international conference on survey methods, Annecy, France.

Chen, Y., Yang S., Hu M., & Wu Y. J. 2016. A reliability-based transit trip planning model under transit network uncertainty. Public Transport, 8(3), 477–496.

Chen, J., Liu, Z., Zhu, S. and Wang, W., 2015. Design of limited-stop bus service with capacity constraint and stochastic travel time. Transportation Research Part: E Logistics Transportation Review, vol. 83, pp. 1–15.

Chen, P.W., and Nie,Y.M., 2015. Optimal transit routing with partial online information. Transportation Research Part B: Methodological, 72, 40–58.

Chen, Q., Adida, E., and Lin, J., 2013. Implementation of an iterative headway-based bus holding strategy with real-time information. Public Transport, vol. 4, no. 3, pp. 165–186.

Chen, S., Claramunt, C. and Ray, C., 2014. A spatio-temporal modelling approach for the study of the connectivity and accessibility of the Guangzhou metropolitan network. Journal of Transport Geography, vol. 36, pp. 12–23.

Chen, X., Yu, L., Zhang, Y. and Guo, J., 2009. Analyzing urban bus service reliability at the stop, route, and network levels. Transportation Research Part A: Policy Practice, vol. 43, no. 8, pp. 722–734.

Chen, X., Wang, Y., Tang, J., Dai, Z. and Ma, X., 2020. Examining regional mobility patterns of public transit and automobile users based on smart card and mobile Internet data: A case study of Chengdu, China. IET Intelligent Transport Systems, vol. 14, no. 1, pp. 45 – 55.

Cheng, T., Tanaksaranond, G., Brunsdon, C. and Haworth, J., 2013. Exploratory visualisation of congestion evolutions on urban transport networks. Transportation Research Part C: Emerging Technologies, vol. 36, pp. 296–306.

Cheon, S. H., Lee, C. and Shin, S. 2019. Data-driven stochastic transit assignment modeling using an automatic fare collection system. Transportation Research Part C: Emerging Technologies, vol. 98, pp. 239-254.

Chidlovskii, B., 2017. Mining Smart Card Data for Travellers' Mini Activities. IEEE Transactions on Intelligent Transportation Systems, vol. 19, pp. 3676-3685.

Chocholáč, J., Sommerauerová, D., Hyrslová, J., Kučera, T., Hruška, R., and Machalík, S., 2020. Service quality of the urban public transport companies and sustainable city logistics. Open Engineering, vol. 10, pp. 86 - 97.

Chowdhury, S. Hadas, Y., Gonzalez, V. and Schot, B., 2018. Public transport users' and policy makers' perceptions of integrated public transport systems. Transport Policy, vol. 61, pp. 75-83.

Chu, K.K., Chapleau, R., 2008. Enriching Archived Smart Card Transaction Data for Transit Demand Modeling. Transportation Research Record: Journal of the Transportation Research Board, No. 2063, Transportation Research Board of the National Academies, Washington, DC, pp. 63–72.

Corman, F., D'Ariano, A., and Hansen, I. A., 2014. Evaluating disturbance robustness of railway schedules. Journal of Intelligent Transportation Systems: technology, planning, and operations, vol. 18, no. 1, pp.106-120.

Corsar, D., Edwards, P., Nelson, J., Baillie, C., Papangelis, K. and Velaga, N., 2017. Linking open data and the crowd for real-time passenger information. Journal of Web Semantics, vol. 43, pp 18-24.

Covic, F., and Voß, S., 2019. Interoperable smart card data management in public mass transit. Public Transport, vol. 11, pp. 523–548.

Crainic, T. G., 2000. Service network design in freight transportation. European Journal of Operational Research, vol. 122, no. 2, pp. 272–288.

Cyril, A., Mulangi, R.H. and George, V., 2019. Performance Optimization of Public Transport Using Integrated AHP–GP Methodology. Urban Rail Transit, vol. 5, pp. 133–144.

Daganzo, C. F. and Pilachowski, J., 2011. Reducing bunching with bus-to-bus cooperation. Transportation Research Part B: Methodological, vol. 45, no. 1, pp. 267–277.

Daganzo, C. F., 2009. A headway-based approach to eliminate bus bunching: Systematic analysis and comparisons. Transportation Research Part B: Methodological, vol. 43, no. 10, pp. 913–921.

Dalla, C., B., Barabino, B., Corona, G., 2013. Sistemi di localizzazione automatica dei veicoli e delle flotte. In: ITS nei trasporti stradali-Tecnologie, metodi ed applicazioni. In Dalla Chiara, B., Bifulco, G.N., Fusco G., Barabino, B., Corona, G., Rossi R., Studer L. (Ed.). EGAF, Forlì, pp. 1-477.

Davari, N., Veloso, B., Costa, G. d. A., Pereira, P.M. Ribeiro, R.P. and Gama, J. A., 2021. Survey on Data-Driven Predictive Maintenance for the Railway Industry. Sensors, vol. 21, no. 17.

De Palma, A., Kilani, M. and Proost, S., 2015. Discomfort in mass transit and its implication for scheduling and pricing. Transportation Research Part B: Methodological, vol.71, pp.1–18.

de Winter, J. C. F., Dodou, D., Happee, R. and Eisma, Y. B., 2019. Will vehicle data be shared to address the how, where, and who of traffic accidents? European Journal of Futures Research, vol. 7, no. 2.

DELFI, (accessed on 06.07.2022). “Durchgängige elektronische Fahrgastinformation”, seamless electronic passenger information; https://www.delfi.de/;

Desaulniers, G. and Hickman, M., 2007. Public transit. In: Laporte G, Barnhart C (eds) Transportation. Handbooks in operations research and management science, vol. 14, pp. 69–127. Elsevier, Amsterdam.

Dessouky, M., Hall, R., ZHANG, L, Singh, A., 2003. Real-time control of buses for schedule coordination at a terminal. Transportation Research Part A: Policy and Practice, vol. 37, no. 2, p. 145-164.

Devarasetty, P. C., Burris, W. and Douglass Shaw, W., 2012. The value of travel time and reliability-evidence from a stated preference survey and actual usage. Transportation Research Part A: Policy and Practice, vol. 46, no. 8, pp. 1227-1240.

Devillaine, F., Munizaga, M. and Trépanier, M., 2012. Detection of activities of public transport users by analyzing smart card data. Transportation Research Record, vol. 2276, no. 1, pp. 48–55.

Diab, E. I., Badami, M. G. and El-Geneidy, A. M., 2015. Bus Transit Service Reliability and Improvement Strategies: Integrating the Perspectives of Passengers and Transit Agencies in North America. Transport Reviews, vol. 35, no.3, pp. 292-328.

Dixit, M., Brands, T., Oort, N. V., Cats, O. and Hoogendoorn, S., 2019. Passenger travel time reliability for multimodal public transport journeys. Transportation Research Record, vol. 2673, no. 2, pp. 149-160.

Dong, H. And Wang, Y., 2018. Bus passenger flow and running status analyzation system based on MAC address. International Conference on Transportation and Development 2018: Traffic and Freight Operations and Rail and Public Transit.

Drevs, F., Tscheulin, D. Lindenmeier, J. and Renner, S., 2014. Crowding-in or crowding out: An empirical analysis on the effect of subsidies on individual willingness-to-pay for public transportation. Transportation Research Part A: Policy and Practice, vol. 59, pp. 250–261.

Droj, G., Droj, L., and Badea, A., 2021. GIS-Based Survey over the Public Transport Strategy: An Instrument for Economic and Sustainable Urban Traffic Planning. ISPRS International Journal of Geo Information, vol. 11, no. 16.

Dziekan, K. and Kottenhoff, K., 2007. Dynamic at-stop real-time information displays for public transport: effects on customers. Transportation Research Part A: Policy and Practice, vol. 41, no. 6, pp. 489-501.

Eberlein, X. J., Wilson, N. H. and Bernstein, D., 2001. The holding problem with real–time information available. Transportation Science, vol. 35, no. 1, pp. 1–18.
Ehreke I., Hess, S., Weis, C. and Axhausen, K., 2015. Reliability in the German value of time study. Transportation Research Record: Journal of the Transportation Research Board, vol. 2495, pp.14-22.

El-Geneidy, A. M., Horning, J. and Krizek, K. J., 2011. Analyzing transit service reliability using detailed data from automatic vehicular locator systems. Journal of Advanced Transportation, vol. 45, no. 1, pp. 66–79.

Eliasson, J., Bo¨rjesson, M., 2014. On timetable assumptions in railway investment appraisal. Transportation Policy ,vol. 36, pp. 118–126.

Eom, J, K., Choi, M. H. andLee, J., 2012. Evaluation of metro service quality using transit smart card data. In: Transportation Research Board 91st annual meeting, No. 12-1314.

Eriksson, k. J., 2020. Visualizing Public Transport with Heat-Maps: Comparing the Scalability of SVG and Canvas for Heat-Maps. Bachelor 's Thesis, Mid Sweden University, Östersund, Sweden.

Espinoza, C., Munizaga, M, Bustos, B. and Trépanier, M., 2018. Assessing the public transport travel behavior consistency from smart card data. Transportation Research Procedia, vol. 32, pp. 44-53.

Ewoldsen, A., 2022. Data sharing presents major opportunities for transportation. National Academics, Sciences Engineering Medicine.

Faizrahnemoon, M., Arieh, S., Lorenzo, M., Emanuele, C., and Robert, S., 2015. A big-data model for multi-modal public transportation with application to macroscopic control and optimisation. International Journal of Control, vol. 88, pp. 1-28.

Faouzi, N. E. E., Leung, H. and Kurian, A., 2011. Data fusion in intelligent transportation systems: Progress and challenges – A survey. Information Fusion, vol. 12, no. 1, pp. 4-10.

Fazari, A., Arnone, M., Botta, C., Caroleo, B., Pensa, S., 2022. Mobisuite: A User-Friendly Tool to Exploit E-Ticketing Data and Support Public Transport Planning. In: Martins, A.L., Ferreira, J.C., Kocian, A. (eds) Intelligent Transport Systems. INTSYS 2021.

Ferreira, J. C., Monteiro, V., Afonso, J. A. and Afonso, J.L., 2016. Methodology for Knowledge Extraction from Mobility Big DataDistributed Computing and Artificial Intelligence. 13th International Conference, vol. 474.

Firew, T., 2016. Analysis of Service Reliability of Public Transportation in the Helsinki Capital Region: The Case of Bus Line 550. Master thesis, Alto University School of Engineering.

Fransen, K., Neutens, T., Farber, S., De Maeyer, P., Deruyter, G. and Witlox, F., 2015. Identifying public transport gaps using time-dependent accessibility levels. Journal of Transport Geography, vol. 48, pp. 176–187.

Frumin, M., and Zhao, J., 2012. Analyzing Passenger Incidence Behavior in Heterogeneous Transit Services Using Smartcard Data and Schedule-Based Assignment. Transportation Research Record, vol. 2274, no. 1, pp. 52–60.

Fu, L., and Yang, X., 2002. Design and implementation of bus–holding control strategies with real-time information. Transportation Research Record, Journal of the Transportation Research Board, vol. 1791, no. 1, pp. 6–12.

Furth, P.G., Hemily, B., Muller, T., and Strathman, J., 2006. Using Archived AVL-APC Data to Improve Transit Performance and Management, Transportation Research Board, Washington, DC, USA, Tech. Rep. TRCP 113.

Gavriilidou, A. and Cats, O., 2018. Reconciling transfer synchronization and service regularity: Real-time control strategies using passenger data. Transportmetrica A: Transport Science, vol. 15, no. 2, pp. 215–243.

Ge, L., Kliewer, N., Nourmohammadzadeh, A., Voss, S.and Xie, L. 2022. Revisiting the richness of integrated vehicle and crew scheduling. Public Transport.

Ge, L., Sarhani, M., Voß, S. and Xie, L., 2021. Review of Transit Data Sources: Potentials, Challenges and Complementarity. Sustainability, vol. 13, 11450.

Gill, B., Coffee-Borden, B., & Hallgren, K., 2014. A Conceptual Framework for Data-Driven Decision Making. Mathematica Policy Research Reports.

Gkiotsalitis, K. And Cats, O., 2018. Reliable frequency determination: Incorporating information on service uncertainty when setting dispatching headways. Transportation Research Part C: Emerging Technologies, vol. 88, pp. 187–207.

Gkiotsalitis, K. (2022). Strategic Planning of Public Transport Services. In: Public Transport Optimization. Springer, Cham.

Goerigk, M., Schachtebeck, M. and Schöbel, A., 2013. Evaluating line concepts using travel times and robustness. Public Transport, vol. 5, pp. 267–284.

Gordon, J. B., Koutsopoulos, H. N., & Wilson, N. H., 2018. Estimation of population origin–interchange–destination flows on multimodal transit networks. Transportation Research Part C: Emerging Technologies, vol. 90, pp. 350–365.

Gordon, J., Koutsopoulos, H., Wilson, N., & Attanucci, J., 2013. Automated inference of linked transit journeys in London using faretransaction and vehicle location data. Transportation Research Record, vol. 2343,pp. 17–24.

Grover, V., Chiang, R. H., Liang, T. P., & Zhang, D., 2018. Creating Strategic Business Value from Big Data Analytics: A Research Framework. Journal of Management Information Systems, vol. 35, no. 2, pp. 388-423.

Guo, J., Liu, Y., Yang, Q., Wang, Y. and Fang, S., 2020. GPS-based citywide traffic congestion forecasting using CNN-RNN and C3D hybrid model. TRANSPORTMETRICA A: TRANSPORT SCIENCE; vol. 17, no. 2, pp. 190-211.

Hadas, Y. and Shnaiderman, M., 2012. Public-transit frequency setting using minimum-cost approach with stochastic demand and travel time. Transportation Research Part B: Methodological, vol. 46, no. 8, pp. 1068– 1084.

Hadas, Y., 2013. Assessing public transport systems connectivity based on Google transit data. Journal of Transport Geography, vol. 33, pp. 105–116.

Hamidi, S., and Hamidi, I., 2021. Subway Ridership, Crowding, or Population Density: Determinants of COVID-19 Infection Rates in New York City. American journal of preventive medicine, vol. 60, no.5, pp. 614-620.

Han, G., and Sohn, K., 2016. Activity imputation for trip-chains elicited from smart-card data using a continuous hidden Markov model. Transportation Research Part B: Methodological, vol. 83, pp. 121–135.

Harmony, X. J. and Gayah, V. V., 2017. Evaluation of Real-Time Transit Information Systems: An information demand and supply approach. International Journal of Transportation Science and Technology, vol. 6, no. 1, pp. 86-98.

Hashem, I. A. T., Chang, V., Anuar, N. B., Adewole, K., Yaqoob, I., Gani, A., ... & Chiroma, H., 2016. The role of big data in smart city. International Journal of Information Management, vol. 36, no. 5, pp. 748-758.

Hassan, M et al., 2016. Modeling transit users stop choice behavior: Do travelers strategize? Journal of Public Transportation, vol. 19, no. 3, pp. 98–116.

Hassan, M. et al., 2016. Modeling transit users stop choice behavior: Do travelers strategize? Journal of Public Transportation, vol. 19, no. 3, pp. 98–116.

Haywood, L., and Koning, M., 2015. The distribution of crowding costs in public transport: New evidence from Paris. Transportation Research Part A: Policy and Practice, vol. 77, pp. 182- 201.

Haywood, L., Koning, M. and Monchambert, G., 2017. Crowding in public transport: Who cares and why? Transportation Research Part A: Policy and Practice, vol. 100, pp. 215-227.

Heil, J., Hoffmann, K. and Buscher, U., 2020. Railway crew scheduling: Models, methods and applications. European Journal of Operational Research, vol. 283, no. 2, pp. 405-425.
Hellinga, B., Yang, F. and Hart-Bishop, J., 2011. Estimating Signalized Intersection Delays to Transit Vehicles: Using Archived Data from Automatic Vehicle Location and Passenger Counting Systems . Transportation Research Record, Journal of Transportation Research Board, vol. 2259, pp. 158–167.

Hernández, D., Muñoz, J. C., Giesen, R. and Delgado F., 2014. Analysis of real-time control strategies in a corridor with multiple bus services. Transportation Research Part B: Methodological, vol. 78, pp. 83-105.

Hickman, M., 2003. Robust passenger itinerary planning using transit AVL data. In: The IEEE 5th International Conference on Intelligent Transportation Systems, Singapore, pp. 840–845. IEEE. https://doi.org/10.1109/ITSC.2002. 1041329.

Hoffman, M., Wilson, S., White, P., 2009. Automated Identification of Linked Trips at Trip Level Using Electronic Fare Collection Data. 88th Annual Meeting of the Transportation Research Board, Washington, 18 p. (CD-ROM).

Hörcher, D., Graham, D. J. and Anderson, R. J., 2017. Crowding cost estimation with large scale smart card and vehicle location data. Transportation Research Part B: Methodological, vol. 95, pp. 105-125.

Hu, S., and Chen, P., 2021. Who left riding transit? Examining socioeconomic disparities in the impact of COVID-19 on ridership. Transportation Research Part D: Transport and Environment, vol. 90, no.102654.

Hua, W., and Ong, G. P., 2018. Effect of information contagion during train service disruption for an integrated rail-bus transit system. Public Transport, vol. 10, pp. 571–594.

Huang, A., Gallegos, L., Lerman, K., 2017. Travel analytics: Understanding how destination choice and business clusters are connected based on social media data. Transportation Research Part C: Emerging Technologies vol. 77, pp. 245–256.

Huang, H., Bucher, D., Kissling, J., Weibel, R. and Raubal, M., 2019. Multimodal Route Planning with Public Transport and Carpooling. IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 9, pp. 3513-3525.

Huisman, D., Freling, R. and Wagelmans, A. P. M., 2004. A Robust Solution Approach to the Dynamic Vehicle Scheduling Problem. Transportation Science, vol. 38, no. 4, pp. 447-458.

Hunter, T., Abbeel, P. and Bayen, A., 2014. The Path Inference Filter: Model-Based Low-Latency Map Matching of Probe Vehicle Data. IEEE Transactions on Intelligent Transportation Systems, vol. 15, no. 2, pp. 507-529.

Ibarra-Rojas, O. J., Giesen, R., and Rios-Solis, Y. A., 2014. “An integrated approach for timetabling and vehicle scheduling problems to analyze the trade-off between level of service and operating costs of transit networks”. In: Transportation Research Part B: Methodological, vol. 70, pp. 35-46. DOI: 10.1016/j.trb.2014.08.010.

Ibarra-Rojas, O. J., Delgado, F., Giesen, R. and Muñoz, J. C., 2015. Planning, operation, and control of bus transport systems: A literature review. Transportation Research Part B: Methodological, vol. 77, pp. 38-75.

Illyas I. F. and Chu X., 2019. Data cleaning. ACM, New York, 2019.

Iliopoulou, C., and Kepaptsoglou, K., 2019. Combining ITS and optimization in public transportation planning: state of the art and future research paths. European Transportation Research Review, vol. 11, no. 27.

Ingvardson, J. B., Nielson, O. A., Ravau, S. and Nielson B. F., 2018. Passenger arrival and waiting time distributions dependent on train service frequency and station characteristics: A smart card data analysis. Transportation Research Part C: Emerging Technologies, vol. 90 , pp. 292-306.

ITF, 2021. “Data-driven Transport Infrastructure Maintenance”. International Transport Forum Policy Papers, no. 95, OECD Publishing, Paris.

Ivaldi, M. and Vibes, C., 2008. Price competition in the intercity passenger transport market: a simulation model. Journal of Transportation Economics and Policy, vol. 42, no. 2, pp.225–254.

Jacobs A., (Accessed on, 05.09.2023). Potential uses of big data in public transportation. Lufthansa industry solutions.

Jevinger, Å.; Persson, J.A., 2019. Exploring the potential of using real-time traveler data in public transport disturbance management. Public Transport, vol.11, pp. 413–441.

Ji, Y., Mishalani, R. G. and McCord, M. R., 2014. Estimating transit route OD flow matrices from APC data on multiple bus trips using the IPF method with an iteratively improved base: Method and empirical evaluation. Journal of Transportation Engineering, vol. 140, no. 5.

Ji, Y., Mishalani, R. G., and McCord, M. R., 2015.Transitpassenger origin–destination flow estimation: Efficiently combining onboard survey and large automatic passenger count datasets. Transportation Research Part C: Emerging Technologies, vol. 58, pp. 178–192.

Jiang, S., Ferreira, J. and Gonzalez, M. C., 2017. Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore. IEEE Transactions on Big Data, vol. 3, no. 2, pp. 208–219.

Jung, YJ. And Casello, J.M., 2020. Assessment of the transit ridership prediction errors using AVL/APC data. Transportation, vol. 47, pp. 2731–2755.

Katakis, I., 2015. Mining urban data, (Part A). Information Systems, vol. 54, pp. 113 – 114.

Keller, C., Glück, F., Gerlach, C. F. and Schlegel, T., 2022. Investigating the potential of data science methods for sustainable public transport. Sustainability, vol. 14, no. 7. https://doi.org/10.3390/ su14074211

Kennedy, H., Moss, G., 2015. Known or knowing publics? Social media data mining and the queation public agency. Big Data Soc , vol. 2, no. 2, pp. 1–11.

Khaleghi, B., Khamis, A., Karray, F. O. and Razavi, S. N., 2013. Multisensor data fusion: A review of the state-of-the-art. Information Fusion, vol. 14, no. 1, pp. 28-44.

Khan, J., Hrelja, R. and Pettersson-Löfstedt, F., 2021. Increasing public transport patronage – An analysis of planning principles and public transport governance in Swedish regions with the highest growth in ridership. Case Studies on Transport Policy, vol.9, Issue 1, pp 260-270.

Khwaja, H. A., Gupta, S. and Kumar, V., 2010. A statistical approach for fault diagnosis in electrical machines. IETE Journal of Research, vol. 56, no. 3, pp. 146-155.

Kieu, L. M., Ou, Y.. and Cai, C., 2018. Large-scale transit market segmentation with spatial-behavioural features. Transportation Research Part C: Emerging Technologies, vol. 90, pp. 97-113.

Kim, K. M., Hong, S. P. Ko, S. J. and Kim, D., 2015 Does crowding affect the path choice of metro passengers? Transportation Research Part A: Policy and Practice, vol. 77, pp. 292–304.

Kim, J., Corcoran, J. and Papamanolis, M., 2017. Route choice stickiness of public transport passengers: Measuring habitual bus ridership behaviour using smart card data. Transportation Research Part C Emerging Technologies, vol. 83, pp. 146–164.

Kon, F., Ferreira, É. C., de Souza, H. A. et al., 2021. Abstracting mobility flows from bike-sharing systems. Public Transport, vol. 14, pp. 545–581.

Kou, W., Chen, X., Yu, L., Qi, Y. and Wang, Y., 2017. Urban commuters' valuation of travel time reliability based on stated preference survey: A case study of Beijing. Transportation Research Part A: Policy and Practice, vol. 95, pp. 372-380.

Koutsopoulos, H. N., Ma, Z., Noursalehi, P. and Zhu, Y., 2019. Transit data analytics for planning, monitoring, control, and information. In Mobility patterns, big data and transport analytics, pp. 229-261.

Koutsopoulos, H. N., Noursalehi, P., Zhu, Y., and Wilson, N. H. M., 2017. Automated data in transit: Recent developments and applications. In Proceedings of the 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT- ITS2017), pp. 604-609.

Kujala, R., Weckström, C., Darst, R. K., Mladenović, M. N., and Saramäki, J., 2018. A collection of public transport network data sets for 25 cities. Scientific data,vol. 5, 180089.

Kumar, P., Khani, A., and He, Q., 2018. A robust method for estimating transit passenger trajectories using automated data. Transportation Research Part C: Emerging Technologies, vol. 95, pp. 731-747.
Kusakabe, T. and Asakura, Y., 2014. Behavioural data mining of transit smart card data: a data fusion approach. Transportation Research Part C: Emerging Technologies, vol. 46, pp.179–191.

Kwoczek, S., Martino, di S., Nejdl, W., 2014. Predicting and visualizing traffic congestion in the presence of planned special events. Journal of Visual Languages & Computing, Vol. 25, no. 6.

Lai, D. S. W. and Leung, J. M. Y., 2018. Real-time rescheduling and disruption management for public transit. Transportmetrica B: Transport Dynamics, vol. 6, no.1, pp.17-33.

Lai, Y., Yang, F., Meng, G. and Lu, W. 2022. Data-Driven Flexible Vehicle Scheduling and Route Optimization. IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 12, pp. 1-15.

Langlois, G. G., Koutsopoulos, H. N. and Zhao, J., 2016. Inferring patterns in the multi-week activity sequences of public transport users. Transportation Research Part C: Emerging Technologies, vol. 64, pp. 1–16.

Lee, D. H., Sun, L. and Erath, A., 2012. Study of bus service reliability in Singapore using fare card data. In: 12th Asia-Pacific Intelligent Transportation Forum.

Leng, N. and Corman, F., 2020. The role of information availabiöity to passengers in public transport disruptions: An agent-based simulation approach. Transportation Research Part A 133(2020), pp. 214-236.

Li, F., Gong, J., Liang, Y. and Zhou, J., 2016. Real-time congestion prediction for urban arterials using adaptive data-driven methods. Multimed Tools Appl, vol. 75, pp. 17573–17592.

Li, H. and Bertini, R. L., 2008.Optimal bus stop spacing for minimizing transit operation cost. In Traffic and transportation studies , pp. 553– 564.

Li, H., and Bertini, R. L., 2008. Optimal bus stop spacing for minimizing transit operation cost. In Traffic and transportation studies, pp. 553– 564.
Li, Y. and Tan, H., 2020. En-route headway-based bus reliability with real-time data at network scale. Transportation Safety and Environment, vol. 2, no. 3, pp. 236-245.

Li, Z., Chen, C. and Wang, K., 2011. Cloud Computing for Agent-Based Urban Transportation Systems. IEEE Intelligent Systems , vol. 26, pp. 73–79.

Li, Z. and Hensher, D. A., 2011. Crowding and public transport: A review of willingness to pay evidence and its relevance in project appraisal. Transport Policy, vol. 18, no. 6, pp. 880-887.

Li, Z.C., Lam, W.H. and Wong, S.C., 2009. The optimal transit fare structure under different market regimes with uncertainty in the network. Netw Spat Econ, vol. 9, no. 2, pp.191–216.

Liebig, T., Piatkowski, N., Bockermann, C. and Morik, K. 2017. Dynamic route planning with real-time traffic predictions, Information Systems. vol. 64, pp. 258-265.

Lin, H., Tang, C., 2021. Analysis and optimization of urban public transport lines based on multiobjective adaptive particle swarm optimization. IEEE Transactions on Intelligent Transportation Systems, pp. 1–13.

Lin, T., et al., 2014. Spatial analysis of access to and accessibility surrounding train stations: A case study of accessibility for the elderly in Perth, Western Australia. Journal of Transport Geography, vol. 39, pp. 111–120.

Liu, G., Yin, Z., Jia, Y., and Xie, Y., 2017. Passenger flow estimation based on convolutional neural network in public transportation system. Knowledge Based System, vol. 123, pp. 102-115.

Liu, L. and Miller, H. J., 2020. Does real-time transit information reduce waiting time? An empirical analysis,Transportation Research Part A: Policy and Practice, vol. 141, pp. 167-179.

Liu, X., Payakkamas, P., Dijk, M. and de Kraker, J., 2023. GIS Models for Sustainable Urban Mobility Planning: Current Use, Future Needs and Potentials. Future Transportation, vol. 3, pp. 384-402.

Liu, Y., Liu, C., Yuan, N. J., Duan, L., Fu, Y., Xiong, H., Xu, S., and Wu, J., 2017. Intelligent bus routing with heterogeneous human mobility patterns. Knowledge and Information Systems, vol. 50, no. 2, pp. 383–415.

Liu, Y., Liu, C., Yuan, N. J., Duan, L., Fu, Y., Xiong, H., Xu, S., and Wu, J., 2017.

Intelligent bus routing with heterogeneous human mobility patterns. Knowledge and Information Systems, vol. 50, no. 2, pp. 383–415.

Lock, O., Bednarz, T., & Pettit, C. ,2021. The visual analytics of big, open public transport data – a framework and pipeline for monitoring system performance in Greater Sydney. Big Earth Data, vol. 5, no. 1, pp. 134-159. doi:10.1080/20964471.2020.1758537

Lovrić, M., Li, T. and Vervest, P., 2013. Sustainable revenue management: A smart card enabled agent-based modeling approach. Decision Support Systems, vol. 54, no. 4, pp. 1587-1601.

Lu, K., Han, B. and Zhou, X., 2018. Smart Urban Transit Systems: From Integrated Framework to Interdisciplinary Perspective. Urban Rail Transit, vol. 4, pp. 49–67.

Lu, K., Liu, J., Zhou, X. and Han, B., 2021. A Review of Big Data Applications in Urban Transit Systems. IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 5, pp. 1–18.

Luke, R. and Heyns, G. J., 2020. An analysis of the quality of public transport in Johannesburg, South Africa using an adapted SERVQUAL model. Transportation Research Procedia, vol. 48, pp. 3562-3576.

Luo, X., Liu, S., Jin, P. J., Jiang, X., and Ding, H., 2017. A connected-vehicle-based dynamic control model for managing the bus bunching problem with capacity constraints. Transportation Planning and Technology, vol. 40, no. 6, pp. 722–740.

Lv, Y., Lv, W., Ren, Y. and Ouyang, Q., 2021. Optimizing the bus operation plan Based on Deep Learning. Microprocessors and Microsystems.

Lycett, M., 2013. "Datafication: making sense of (big) data in a complex world" European Journal of Information Systems, 22 (4) (2013), pp. 381-386, 10.1057/ejis.2013.10

Lyu, C., Wu, X., Liu, Y. and Liu, Z., 2021. A Partial-Fréchet-Distance-Based Framework for Bus Route Identification. IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 7, pp. 9275-9280.

Ma, X., and Wang, Y.. 2014. Development of a data-driven platform for transit performance measures using smart card and GPS data. Journal of Transportation Engineering, vol. 140, no. 12, 04014063.

Ma, X., Wu, Y. J., Wang, Y., Chen, F. And Liu, J., 2013. Mining smart card data for transit riders' travel patterns. Transportation Research Part C: Emerging Technologies, vol. 36, pp. 1–12.

Maktoubian, J., 2017. Proposing a streaming big data analytics (SBDA) platform for condition based maintenance (CBM) and monitoring transportation systems. EAI Endorsed Transactions on Scalable Information Systems, vol. 4, no. 13.

Maktoubian, J., Noori, M., Ghasempour-Mouziraji, M., Amini, M., 2017 Analyzing large-scale smart card data to investi- gate public transport travel behaviour using big data analytics. Journal of Inormation Technology and Software Engineering, vol. 7, no. 4, pp. 1–3.

Malucelli, F., and Tresoldi, E., 2019. Delay and disruption management in local public transportation via real-time vehicle and crew re-scheduling: a case study. Public Transportation, vol. 11, pp. 1–25.

Mandelzys, M. and Hellinga, B., 2010. Identifying Causes of Performance Issues in Bus Schedule Adherence with Automatic Vehicle Location and Passenger Count Data. Transportation Research Record, Journal of the Transportation Research Board, vol. 2143, pp. 9-15.

Manolakis, D. E., 1996. Efficient solution and performance analysis of 3-D position estimation by trilateration. IEEE Transactions on Aerospace and Electronic Systems, vol. 32, no. 4, pp. 1239-1248.

Mesquita, M., Paias, A. and Respício, A., 2009. Branching approaches for integrated vehicle and crew scheduling. Public Transpport, vol. 1, pp. 21–37.

Mishalani, R. G., McCord M. M. and Wirtz J., 2006. Passenger Wait Time Perceptions at Bus Stops: Empirical Results and Impact on Evaluating Real-Time Bus Arrival Information. Journal of Public Transportation, vol. 9, no. 2, pp. 89-106.

Mohammadi, R., He, Q., and Karwan, M., 2021. Data-driven robust strategies for joint optimization of rail renewal and maintenance planning. Omega, vol. 103.

Moreira-Matias, L., Mendes-Moreira, J., Sousa, J. F., and de Gama, J ., 2015. Improving mass transit operations by using AVL-based systems: A survey. IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 4, pp. 1636–1653.

Morency, C., Trépanier, M., and Agard, B., 2007. Measuring transit use variability with smart-card data. Transportation Policy, vol. 14, no. 3, pp. 193–203.

Munizaga, M., Devillaine, F., Navarrete, C. and Silva, D., 2014. Validating travel behavior estimated from smartcard data. Transportation Research Part C: Emerging Technologies, vol. 44, no. 4, pp. 70–79.

Nagy, A. M. and Simon, V., 2018. Survey on traffic prediction in smart cities. Pervasive and Mobile Computing, vol. 50, pp. 148-163.

Nassir, N., Hickman, M. and Ma, Z., 2015. Behavioral findings from observed transit route choice strategies in the farecard data of Brisbane. Proc. 37th ATRF, Sydney, NSW, Australia, 2015, pp. 1–11.

Nassir, N., Khani, A., Lee, S. G., Noh, H. and Hickman, M., 2011. Transit stop-level origin–destination estimation through use of transit schedule and automated data collection system. Transportation Research Record, vol. 2263, no. 1, pp. 140–150.

Nesheli, M.M. and Ceder, A. A. 2015. Improved reliability of public transportation using real-time transfer synchronization. Transportation Research Part C: Emerging Technologies, vol. 60, pp. 525–539.
Network Timetable Exchange, (accessed on 06.07.2022), NeTEx. Available online: http://netex-cen.eu/

Neilson, A., Ben Daniel, I. and Tjandra, S., 2019. Systematic Review of the Literature on Big Data in the Transportation Domain: Concepts and Applications. Big Data Research, vol. 17, pp. 35-44.

Nielsen, B. F., Frølich, L., Nielsen, O. A., and Filges, D., 2013. Estimating passenger numbers in trains using existing weighing capabilities. Transportmetrica A: Transport Science, vol. 10, no. 6, pp. 502–517.

Nielsen, L. K., Kroon, L. and Maróti, G., 2012. A rolling horizon approach for disruption management of railway rolling stock. European Journal of Operational Research, vol. 220, no. 2.

Nishiuchi, H., King, J., and Todoroki, T., 2013. Spatial-temporal daily frequent trip pattern of public transport passengers using smart card data. International Journal of Intelligent Transportation Systems Research, vol. 11, no. 1, pp. 1–10.

Niu, H,, Zhou, X. and Gao, R., 2015. Train scheduling for minimizing passenger waiting time with time-dependent demand and skip-stop patterns: Nonlinear integer programming models with linear constraints. Transportation Research Part B, Methodologies, vol. 76, pp. 117–135.

Nökel, K., and Wekeck, S., 2009. Boarding and Alighting in Frequency-Based Transit Assignment. Transportation Research Record, vol. 2111, no. 1, pp. 60–67.

Noursalehi, P., Koutsopoulos, H.N., and Zhao, J., 2018. Real time transit demand prediction capturing station interactions and impact of special events. Transportation Research Part C: Emerging Technologies, vol. 97, pp. 277-300.

Nunes, A.A., Dias, T.G., e Cunha, J.F., 2016. Passenger Journey Destination Estimation from Automated Fare Collection System Data Using Spatial Validation. IEEE Transactions on Intelligent Transportation Systems, vol. 17, pp.133–142.

Nuzzolo, A., Crisalli, U., Comi, A., and Rosati, L., 2016. A mesoscopic transit assignment model including real-time predictive information on crowding. Journal of Intelligent Transportation Systems, vol. 20, pp. 316 - 333.

Okunieff, P. E., 1997. AVL systems for bus transit: A synthesis of transit practice.

On˜a, J.de, On˜a, R., Eboli, L. and Mazzulla, G., 2013. Perceived service quality in bus transit service: a structural equation approach. Transportation Policy, vol. 29, pp. 219–226.

Oort, N. V., 2014. Incorporating service reliability in public transport design and performance requirements: International survey results and recommendations. Research in Transportation Economics., vol. 48, pp. 92–100.
Oort, N.V., Sparing, D., Brands, T., and Goverde, R.M., 2015. Data driven improvements in public transport: the Dutch example. Public Transport, vol. 7, pp. 369-389.

Ordóñez Medina, S. A. and Erath, A., 2013. Estimating dynamic workplace capacities by means of public transport smart card data and household travel survey in Singapore. Transportation Research Record, vol. 2344, no. 1, pp. 20–30.

Ortelio, 2018. Project Title: Novel Decision Support tool for Evaluating Strategic Big Data investments in Transport and Intelligent Mobility Services, Topic: Horizon 2020 Research and Innovation Programme, MG-8-2-2017 - Big data in Transport: Research opportunities, challenges and limitationsMG-8-2-2017: Big data in Transport: Research opportunities, challenges and limitations.

Osang, G., Cook, J., Fabrikant, A., & Gruteser, M., 2019. LiveTraVeL: Real-time matching of transit vehicle trajectories to transit routes at scale. IEEE Intelligent Transportation Systems Conference, ITSC 2019, pp. 2244-2251.

Ou, Y., Mihăiţă, A-S., and Chen F., 2022. 14 - Big data processing and analysis on the impact of COVID-19 on public transport delay. Data Science for COVID-19, Academic Press, vol. 2, ISBN 9780323907699,

Pangilinan, C., Wilson, N., and Moore, A. 2008. Bus Supervision Deployment Strategies and Use of Real-Time Automatic Vehicle Location for Improved Bus Service Reliability. Transportation Research Record, Journal of the Transportation Research Board, vol. 2063, no. 1, 28–33.

Park, J.Y., Kim, D.J., 2008. The Potential of Using the Smart Card Data to Define the Use of Public Transit in Seoul. Transportation Research Record: Journal of the Transportation Research Board, No. 2063, Transportation Research Board of the National Academies, Washington, DC, pp. 3–9

Park, J., Chowdhury, S. and Wilson, D. J., 2020. Gap between Policymakers' Priorities and Users' Needs in Planning for Accessible Public Transit System. Journal of Transportation Engineering, Part A: Systems. vol. 146, no. 4.

Paulsson, A., 2018. Making the sustainable more sustainable: public transport and the collaborative spaces of policy translation. Journal of Environmental Policy & Planning, vol. 20, no. 4, pp. 419-433.

Pelletier, M. P., Trépanier, M. and Morency, C., 2011. Smart card data use in public transit: A literature review. Transportation Research Part C: Emerging Technologies, vol. 19, no. 4, pp. 557–568.

Pendyala, R. M., Yamamoto, T. and Kitamura, R., 2002. On the formulation of time-space prisms to model constraints on personal activity-travel engagement. Transportation, vol. 29, no. 1, pp. 73–94.

Pi, M., Yeon, H., Son, H. and Jang Y., 2021. Visual Cause Analytics for Traffic Congestion. IEEE Transactions on Visualization and Computer Graphics, vol. 27, no. 3, pp. 2186-2201.

Pi, X. Egge, M., Whitmore, J. and Qian, Z., 2018. Understanding Transit System Performance Using AVL-APC Data: An Analytics Platform with Case Studies for the Pittsburgh Region. Journal of Public Transportation, vol. 21, no. 2, pp.19-40.

Pozo, R. F., Wilby, M. R., Díaz, J. J. V. and Rodriguez, A. B., 2022. Data-driven analysis of the impact of COVID-19 on Madrid's public transport during each phase of the pandemic. Cities, vol. 127.

Prarthana, T. and Jayakumar, J., 2017. Challenges and Opportunities with Big Data. International Journal of Engineering Research and Technology, vol. 6, no. 7.

Prommaharaj, P.; Phithakkitnukoon, S.; Demissie, M.G.; Kattan, L.; Ratti, C., 2020. Visualizing public transit system operation with GTFS data: A case study of Calgary, Canada. Heliyon 2020, vol. 6, no. 4.

Puusaag, E. and Palmi, A., 2021. Implementing building information management (BIM) in Estonian Transport Administration. IOP Conference Series: Materials Science and Engineering, IBRC 2021, vol. 1202.

Qi, G., Huang, A., Guan, W., Fan, L., 2018. Analysis and prediction of regional mobility patterns of bus Travellers using smart card data and points of interest data. IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 4, pp. 1197–1214.

Qiu, G., Song, R., He, S., Xu, W. Jiang, M., 2019. Clustering Passenger Trip Data for the Potential Passenger Investigation and Line Design of Customized Commuter Bus. IEEE Transactions on Intelligent Transportation Systems, vol. 20 pp. 3351-3360.

Rajbhandari, R., Chien, S. and Daniel, J., 2003. Estimation of bus dwell times with automatic passenger counter information. Transportation Research Record, vol. 1841, pp. 120-127.

Ravensbergen, L., Liefferinge, M. V., Isabella, J., Merrina, Z. and El-Geneidy, A., 2022. Accessibility by public transport for older adults: A systematic review, Journal of Transport Geography, vol. 103.

Redmond, M. A., 2020. Reliable itinerary planning in transportation networks. Ph.D. thesis, University of Iowa, United States.

Redmond, M., Campbell, A. M. and Ehmke, J. F., 2020. Data-driven planning of reliable itineraries in multi-modal transit networks. Public Transport, vol. 12, pp. 171 – 205.

Riasetiawan, M., Harwanto, I. M., Falakh, A. I., Harryajie, A. K., Munjazi, J., and Adji, T. B., 2017. Profiling and clustering methods for transaction profiling in BRT transaction. 9th International Conference on Information Technology and Electrical Engineering (ICITEE), pp. 1–6.

Saavedra, M., Hellinga, B. and Casello, J. M., 2011. Automated Quality Assurance Methodology for Archived Transit Data from Automatic Vehicle Location and Passenger Counting Systems. Transportation Research Record, Journal of the Transportation Research Board, vol. 2256, pp. 130-141.

Sáez, D., Cortés, C. E., Milla, F., Núñez, A., Tirachini, A., and Riquelme, M. , 2012. Hybrid predictive control strategy for a public transport system with uncertain demand. Transportmetrica, vol. 8, no. 1, pp. 61–86.

Sánchez-Martínez, G. E., 2016. Inference of public transportation trip destinations by using fare transaction and vehicle location data: Dynamic programming approach. Transportation Research Record, vol. 2652, pp. 1–7.

Sánchez-Martínez, G. E. and Munizaga, M., 2016. Workshop 5 report: Harnessing big data. Research in Transportation Economics, vol. 59, pp. 236 -241.

Sarkar, C., Treurniet, J. J., Narayana, S., Prasad, R. V. and de Boer, W., 2015. SEAT: Secure Energy-Efficient Automated Public Transport Ticketing System. Journal of Latex Class Files, vol. 14, no. 8.

Schade, J. and Schlag, B., 2003. Acceptability of urban transport pricing strategies. Transportation Reseach Part F: Traffic Psychology and Behaviour, vol. 6, no. 1, pp. 45–61.

Schmöcker, J. D., Shimamoto, H., and Kurauchi, F., 2013. Generation and calibration of transit hyperpaths. Procedia - Social and Behavioral Sciences, vol. 80, pp. 211–230.

Schmöcker, J. D., Krauchi, F. and Shimamoto, H., 2017. Public transport planning with smart card data. CRC Press, Tyalor and Francis Group, 6000 Broken Sound Parkway, NW, Suite 300, Boca Raton, FL 334872742.

Shafique M. M. and Hato, E., 2014. Use of acceleration data for transportation mode prediction. Transportation, vol. 42, pp. 163–188.

Shang, H., Liu, Y., Huang, H. and Guo, R.-Y., 2019. Vehicle Scheduling Optimization considering the Passenger Waiting Cost. Journal of Advanced Transportation, vol. 2019, pp. 1-13.

Shatnawi, N., Al-Omari, A. and Al-Qudah, H., 2020. Optimization of Bus Stops Locations Using GIS Techniques and Artificial Intelligence. Procedia Manufacturing, vol. 44, pp. 52-59.

Shaukat, S., Katscher, M., Wu, C. L., Delgado, F. and Larrain, H., 2020. Aircraft line maintenance scheduling and optimisation. Journal of Air Transport Management, vol. 89.
Shen, Y., Xu, J., and Wu, X., 2017. Vehicle scheduling based on variable trip times with expected on-time performance. International Transactions in Operational Research, vol. 24, no. (1–2), pp. 99–113.

Shen, Y., Xu, J., and Li, J., 2016. A probabilistic model for vehicle scheduling based on stochastic trip times. Transportation Research Part B: Methodological, vol. 85, pp. 19–31.

Shen, Y., Xu, J. and Wu, X., 2017. Vehicle scheduling based on variable trip times with expected on-time performance. International Transactions in Operational Research, vol. 24, no. (1–2), pp. 99–113.

Shen, Y., Xu, J. and Zeng, Z., 2016. Public transit planning and scheduling based on AVL data in China. International Transactions in Operational Research, vol. 23, no. 6, pp. 1089–1111.

Shen,Y., Xu,J., and Zeng, Z., 2016. Public transit planning and scheduling based on AVL data in China. International Transactions in Operational Research, vol. 23, no. 6, pp. 1089–1111.

Shi, Q., Zhang, K., Weng, J., Dong, Y., Ma, S. and Zhang, M., 2021. Evaluation model of bus routes optimization scheme based on multi-source bus data. Transportation Research Interdisciplinary Perspectives, vol. 10.

Shi, Q., Zhang, K., Weng, J., Dong, Y., Ma, S., & Zhang, M., 2021. Evaluation model of bus routes optimization scheme based on multi-source bus data. Transportation Research Interdisciplinary Perspectives, vol. 10, 100342.

Shi, R.J., Mao, B.H., Ding, Y., Bai, Y. and Chen, Y., 2016. Timetable optimization of rail transit loop line with transfer coordination. Discrete Dynamic in Nature and Society, vol. 2016, pp. 1–11.
Shu, J., Jia, X., Yang, K. and Wang, H. 2021. Privacy-Preserving Task Recommendation Services for Crowdsourcing. IEEE Transactions on Services Computing, vol.14, pp. 235–247.

Siebert, M. and Ellenberger, D., 2020. Validation of automatic passenger counting: introducing the t-test-induced equivalence test. Transportation, vol. 47, pp. 3031–3045.

Skartland, E. G., 2022. Unfavorable transit planning: Lack of knowledge, lack of collaboration, or political conflicts? A case study of two Norwegian cities aiming to increase transit competitiveness. Progress in Planning, 100656.

Šoštarić, M., Vidović, K., Jakovljević, M. and Lale, O. 2021. Data-Driven Methodology for Sustainable Urban Mobility Assessment and Improvement. Sustainability, vol. 13, no. 7162.

Stelzer, A., Englert, F., Hörold, S. and Mayas, C., 2016. Improving service quality in public transportation systems using automated customer feedback. Transportation Research Part E: Logistics and Transportation Review, vol. 89, pp. 259-271.

Sun, L., Jin, J. G., Lee, D. H., Axhausen, K. W. and Erath, A., 2014. Demand-driven timetable design for metro services. Transportation Research Part C: Emerging Technologies, vol. 46, pp. 284–299.

Sun, L., Jin, J. G., Lee, D. H., Axhausen, K. W., and Erath, A., 2013. Demand- driven timetable design for metro services. Transportation Research Part C: Emerging Technologies, vol. 6, pp. 284–299.

Sun, L.,Axhausen, K. W., Lee, D. H. and Huang, X., 2012. Understanding metropolitan patterns of daily encounters. Proc. Natl. Acad. Sci. USA, vol. 110, no. 34, pp. 13774–13779.

Sun, S., Akhtar, N., Song, H., Zhang, C., Li, J., and Mian, A.S., 2018. Benchmark Data and Method for Real-Time People Counting in Cluttered Scenes Using Depth Sensors. IEEE Transactions on Intelligent Transportation Systems, vol. 20, pp. 3599-3612.

Sun, Y. and Xu, R., 2012. Rail transit travel time reliability and estimation of passenger route choice behavior: Analysis using automatic fare collection data. Transportation Research Record, vol. 2275, no. 1, pp. 58–67.

Sun, Y. and Schonfeld, P. M., 2016. Schedule-based rail transit path-choice estimation using automatic fare collection data. Journal of Transportation Engineering, vol. 142, no. 1.

Sun, Y., Shi, J. and Schonfeld, P. M., 2016. Identifying passenger flow characteristics and evaluating travel time reliability by visualizing AFC data: a case study of Shanghai Metro. Public Transport, vol. 8, no. 3, pp. 341–363.

Tallon, P.P., Ramirez, R.V. and Short J.E., 2014. "The information artifact in IT governance: toward a theory of information governance" Journal of Management Information Systems, vol. 30, no. 3, pp. 141-177, 10.2753/MIS0742-1222300306

Tang, J., Yang, Y., Hao, W., Liu, F. and Wang, Y. 2020. A Data-Driven Timetable Optimization of Urban Bus Line Based on Multi-Objective Genetic Algorithm. IEEE Transactions on Intelligent Transportation Systems, vol. 22, pp. 2417-2429.

Tao, S., Corcoran, J., Mateo-Babiano, I. and Rohde, D., 2014. Exploring bus rapid transit passenger travel behaviour using big data. Applied Geography, vol. 53, pp. 90–104.

Tavassoli, A., Mesbah, M. Hickman, M., 2018. Application of smart card data in validating a large-scale multi-modal transit assignment model. Public Transport, vol. 10, pp. 1-21.

Tennøy, A., 2022. Patronage effects of changes to local public transport services in smaller cities. Transportation Research Part D: Transport and Environment, vol. 106.

Tian, Q. and Wang, H., 2022. Optimization of preventive maintenance schedule of subway train components based on a game model from the perspective of failure risk. Sustainable Cities and Society, vol. 81.

Tibaut, A., Kaučič, B. and Rebolj, D., 2012. A standardised approach for sustainable interoperability between public transport passenger information systems. Computers in Industry, vol. 63, no.8, pp, 788-798.

Tilocca, P., Farris, S., Angius, S., Argiolas, R., Obino, A., Secchi, S., Mozzoni, S. and Barabino, B., 2017. Managing Data and Rethinking Applications in an Innovative Mid-sized Bus Fleet.Transportation Research Procedia, vol. 25, pp. 1899-1919.

Tirachini, A., 2013. Estimation of travel time and the benefits of upgrading the fare payment technology in urban bus services. Transportation Research Part C: Emerging Technologies, vol. 30, pp. 239–256.

Torre-Bastida, A. I.,Javier, D. S., Nekane, B., Maitena, I., Sergio, C. and Ibai, L., 2018. Big Data for Transportation and Mobility: Recent Advances, Trends and Challenges. IET Intelligent Transport Systems, vol. 12, no. 8.

transpor t: Harmonisation and data comparability pp. 25–31.

Tran, V.T., Eklund, P. and Cook, C., 2015. Toward Real-Time Multi-criteria Decision Making for Bus Service Reliability Optimization. Foundations of Intelligent Systems. ISMIS 2015, vol 9384, pp. 371-378.

Trépanier, M., Morency, C., 2010. Assessing transit loyalty with smart card data. In: Presented at the 12th World Conference on Transport Research, Lisbon, No. 2341.

Trépanier, M., Tranchant, N. and Chapleau, R., 2007. Individual trip destination estimation in a transit smart card automated fare collection system. Journal of Intelligent Transportation Systems, vol. 11, no. 1, pp. 1–14.

Turnquist, M. A., 1981. Strategies for improving reliability of bus transit service. Transportation Research Record, vol. 818.

UITP, 2020. Digitalisation in public transport: implementing predictive assets maintenance. Knowledge brief, Rail unit of the UITP secretariat, Rue Sainte-Marie 6, B-1080 Brussels, Belgium.

UITP, 2021. A framework for sustainable data sharing in public transport. Policy brief, Rue Sainte-Marie 6, B-1080 Brussels, Belgium.

UN Global Pulse, 2017. Using Big Data Analytics for Improved Public Transport, Project Series, no. 25.
Urbanek, A., 2019. Data-Driven Transport Policy in Cities: A Literature Review and Implications for Future Developments.

Utsunomiya, M., Attanucci, J., Wilson, N., 2006. Potential Uses of Transit Smart Card Registration and Transaction Data to Improve Transit Planning, Transportation Research Record: Journal of the Transportation Research Board, No. 1971, Transportation Research Board of the National Academies, Washington, DC, pp. 119–126.

Valdes, D., Cruzado, I., Martinez, J. and Taveras, Y., 2017. Development of Transit Performance Measures Using Big Data. Transportation Informatics University Transprtation Center.

Van. Oort , N. and Cats, O., 2015. Improving Public Transport Decision Making, Planning and Operations by Using Big Data: Cases from Sweden and the Netherlands. 2015 IEEE 18th International Conference on Intelligent Transportation Systems, Gran Canaria, Spain, pp. 19-24.

vdv.de/ netex.aspx (accessed on 06.07.2022), VDV. Soll-Daten-Schnittstellen: Europäische Norm NeTEx (CEN), 2022. (In German). Available online: https://www.vdv.de/ netex.aspx

Voß, S., 2001. Gutenschwager, K. Informationsmanagement; Springer: Berlin, Germany.

Wang, W., Zong F. and Yao, B., 2021. A Proactive Real-Time Control Strategy Based on Data-Driven Transit Demand Prediction. IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 4, pp. 2404-2416.

Wang, Y., Zhu, L., Lin, Q. And Zhang, L., 2018. Leveraging big data analytics for train schedule optimization in urban rail transit systems. 21st International Conference on Intelligent Transportation Systems (ITSC) pp. 1928–1932.

Watkins, K. E., Ferris, B., Borning, A., Rutherford, G. S. and Layton, D., 2011. Where Is My Bus? Impact of mobile real-time information on the perceived and actual wait time of transit riders. Transportation Research Part A: Policy and Practice, vol. 45, no. 8, pp. 839-848.

Welch, T. F. and Widita, A., 2019. Big data in public transportation: a review of sources and methods. Transport Reviews, vol. 39, pp. 795-818.

White, G., Zink, A., Codecá, L., Clarke, S., 2021. A digital twin smart city for citizen feedback. Cities, vol. 110.

Williams A., 2019. Exploring the use of advanced traffic information system to manage traffic congestion. Scientific African, vol. 4.

Willumsen, L., 2021. Use of Big Data in Transport Modelling. International Transport Forum Discussion Papers, No. 2021/05, OECD Publishing, Paris.
Wong, W. K., Tan, P. N., Loo, C. K. and Lim, W. S., 2009. An Effective Surveillance System Using Thermal Camera. 2009 International Conference on Signal Acquisition and Processing, pp. 13-17.

Wu, J., Du B., Gong, Z., Wu, Q., Shen, Z., Zhou, L., Cai C., 2023. A GTFS data acquisition and processing framework and its application to train delay prediction. International Journal of Transportation Science and Technology, vol. 12, no. 1.

Wu, Y., Yang, H., Tang, J. and Yu, Y., 2016. Multi-objective re-syn- chronizing of bus timetable: model, complexity and solution. Transp Res Part C Emerg Technol, vol. 67, pp.149–168.

Walker, J., 2008. Purpose-driven public transport: creating a clear conversation about public transport goals. Journal of Transport Geography, vol. 16, no. 6, pp. 436-442.

Wang, B. and Zacharias, J., 2020. Noise, odor and passenger density in perceived crowding in public transport. Transportation Research Part A: Policy and Practice, vol. 135, pp. 215-223.

Xiao-Feng, X., 2014. Coping with real-world challenges in real-time urban traffic control.

Xie, J., Huang, J., Zeng, C., Jiang, S. H. and Podlich, N., 2020. Systematic Literature Review on Data-Driven Models for Predictive Maintenance of Railway Track: Implications in Geotechnical Engineering. Geosciences, vol. 10, no. 425.

Xu, W., Zhang, W. and Li, L., 2017. Measuring the expected locational accessibility of urban transit network for commuting trips. Transportation Research Part D: Transportation Environment, vol. 51, pp. 62–81.

Xu, W. and Yang, L., 2019.Evaluating the urban land use plan with transit accessibility. Sustainable Cities and Society, vol. 45, pp. 474-485.

Xu, X., Xie, L., Li, H. and Qin, L., 2018. Learning the route choice behavior of subway passengers from AFC data. Expert Systems with Applications, vol. 95, pp. 324–332.

Xue, M., Wu, H., Chen, W., Ng, W. S., and Goh, G. H., 2014. Identifying tourists from public transport commuters. In Proc. ACM KDD, New York, NY, USA, pp. 1779–1788.

Xue, Y., Zhong, M., Xue, L., Tu, H., Tan, C., Kong, Q. and Guan, H., 2022. A Real-Time Control Strategy for Bus Operation to Alleviate Bus Bunching. Sustainability, vol. 14, no. 7870.

Yan, S., Chi, C.-J., and Tang, C.H., 2006. Inter-city bus routing and timetable setting under stochastic demands. Transportation Research Part A: Policy and Practice, vol. 40, no. 7, pp. 572–586.

Yan, S., Chi, C.J., and Tang, C. H., 2006. Inter-city bus routing and timetable setting under stochastic demands. Transportation Research Part A: Policy and Practice, vol. 40, no. 7, pp. 572–586.

Yan, Y., Meng, Q., Wang, S. and Guo, X., 2012.Robustoptimization model of schedule design for a fixed bus route. Transportation Research Part C: Emerging Technologies, vol. 25, pp. 113–121.

Yang, L., Dasen, Y. and Xianbiao, H., 2020. A Differential Privacy-Based Privacy-Preserving Data Publishing Algorithm for Transit Smart Card Data. Transportation Research Part: C Emerging Technologies, vol.115.

Yang, X., Chen, A., Ning, B. And Tang, T., 2016. A stochastic model for the integrated optimization on metro timetable and speed profile with uncertain train mass. Transportation Research Part B: Methodological, vol. 91, 2016, pp. 424-445.

Yang, X., Chen, A., Ning, B. and Tang, T., 2017. Bi-objective programming approach for solving the metro timetable optimization problem with dwell time uncertainty. Transportation Research Part E: Logistics Transportation Review, vol. 97, pp. 22–37.

Yap, M. D., Cats, O., van Oort, N., and Hoogendoorn, S. P., 2017. A robust transfer inference algorithm for public transport journeys during disruptions. Transportation Research Procedia, vol. 27, pp. 1042–1049.

Yap, M., Cats, O., Krasemann, J. T., Oort, N. V. and Hoogendoorn, S., 2022. Quantification and control of disruption propagation in multi-level public transport networks. International Journal of Transportation Science and Technology, vol. 11, no.1, pp. 83-106.

Yu, B. and Yang, Z., 2009. A dynamic holding strategy in public transit systems with real-time information. Applied Intelligence, vol. 31, no. 1, pp. 69–80.

Yu, G. and Qi, X., 2004. Disruption Management: Framework, Models and Applications. World Scientific.

Zeng, X., Zhang, Y., Jiao, J. and Yin, K., 2020. Route-Based Transit Signal Priority Using Connected Vehicle Technology to Promote Bus Schedule Adherence. IEEE Trancastions on Intelligent Transportation Systems, vol. 22, pp. 1174–1184.

Zeng, G., 2015. Application of big data in intelligent traffic system. IOSR Journal of Computer Engineering, vol. 17, no. 1, pp. 1–4.

Zhang, F., Shen, Q., and Clifton, K. J., 2008. Examination of Traveler Responses to

Real-Time Information about Bus Arrivals using Panel Data. Transportation Research Record, Journal of the Transportation Research Board, vol. 2082, no. 1, pp. 107–115.

Zhang, Y. S. and Yao, E. J., 2015. Splitting travel time based on AFC data: Estimating walking, waiting, transfer, and in-vehicle travel times in metro system. Discrete Dynamics in Nature and Society., vol. 2015, pp. 1–11.
Zhao, J. et al., 2016. Estimation of passenger route choice pattern using smart card data for complex metro systems. IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 4, pp. 790–801.

Zhao, X., Dawson, D., Sarasua, W. and Birchfield, S., 2016. Automated Traffic Surveillance System with Aerial Camera Arrays Imagery: Macroscopic Data Collection with Vehicle Tracking. Journal of Computing in Civil Engineering. 31.

Zhen, Q. and Jing, S., 2014. Train rescheduling model with train delay and passenger impatience time in urban subway network. Journal of Advanced Transportation, vol. 50, no. 8, pp. 1990–2014.

Zhou, L., Chen, N., Yuan, S. and Chen, Z., 2016. An Efficient Method of Sharing Mass Spatio - Temporal Trajectory Data Based on Cloudera Impala for Traffic Distribution Mapping in an Urban City. Sensors, vol. 16, no. 11, 1813.

Zhu, L., Yu, F. R., Wang, Y., Ning, B., and Tang, T., 2019. Big data analytics in intelligent transportation systems: A survey. IEEE Transactions on Intelligent Transport System, vol. 20, no. 1, pp. 383–398.

Zolfaghari, S., Azizi, N. and Jaber, M. Y., 2004. A model for holding strategy in public transit systems with real-time information. International Journal of Transport Management, vol. 2, no. 2, pp. 99–110.

Zou, Q., Yao, X., Zhao, P., Wei, H., and Ren, H., 2018. Detecting home location and trip purposes for cardholders by mining smart card transaction data in Beijing subway. Transportation, vol. 45, no. 3, pp. 919–944. [Original source: https://studycrumb.com/alphabetizer]