USING SMART CARD DATA FOR MODE SHIFT ESTIMATION: ENHANCING THE EFFICIENCY OF İZMİR'S TRANSIT SYSTEM

A Thesis Submitted to
the Graduate School of
İzmir Institute of Technology
in Partial Fulfillment of the Requirements for Degree of

MASTER OF SCIENCE

in Civil Engineering

by

Fırat Enver KESMEZ

June 2023
İZMİR

We approve the thesis of **Fırat Enver KESMEZ**

Examining Committee Members:

Prof.Dr. Serhan TANYEL
Civil Engineering, Dokuz Eylül University

Assoc. Prof. Volkan Emre UZ
Civil Engineering, İzmir Institute of Technology

Assist. Prof. Tolga ERCAN
Civil Engineering, İzmir Institute of Technology

22 June 2023

Assoc. Prof. Volkan Emre UZ
Supervisor, Civil Engineering, İzmir Institute of Technology

Prof.Dr. Cemalettin DÖNMEZ
Head of the Department of Civil Engineering

Prof.Dr. Mehtap EANES
Dean of the Graduate School

ACKNOWLEDGEMENTS

First of all, I would like to express my gratitude to my advisor, Assoc. Prof. Volkan Emre UZ, for his guidance and support throughout my M.Sc. study. This thesis would not have been possible without his kind, supportive, and patient mentorship. I would also like to extend my thanks to Assist. Prof. Tolga ERCAN for his valuable guidance and expertise whenever I needed it.

I want to express my gratitude to the authorities in ESHOT for their invaluable contribution to this study by sharing knowledge and experience, which made this research possible.

Lastly, I would like to express my heartfelt gratitude to my family, whom I am indebted to forever. Their unconditional love, support, sharing, and patience have been instrumental in my journey. I am immensely thankful to them for giving me every opportunity to become a better person.

ABSTRACT

USING SMART CARD DATA FOR MODE SHIFT ESTIMATION: ENHANCING THE EFFICIENCY OF İZMİR’S TRANSIT SYSTEM

The burden of highly subsidized public transportation services necessitates careful planning of operations and resource management. Traditionally, planning involves utilizing costly and cumbersome methods to collect historical data on passenger behavior and travel demand. This study aims to contribute to the use of passive data in the public transportation planning processes by utilizing smart card data from a one-weekday to estimate the potential mode shifts from bus transits following the commencement of the extended metro line in Narlıdere, İzmir. For this purpose, first, the trip chaining algorithm, widely used in literature, is used to estimate the alighting location of passengers. Then, the competitive bus routes are determined, and this process is accompanied by other algorithms developed to create alternative scenarios where the passengers repeat their trips by utilizing the metro. Finally, the mode shift behavior of the passengers is estimated by employing travel time saving, which is calculated both deterministically and considering the passenger’s travel convenience.

As a result, 39 bus routes are identified operating inside the service area of the metro extension. Sixteen of them are selected for further analysis based on their competitiveness, which is higher than 55%. The results showed that almost 30 to 55% of the passengers on the competitive bus routes favor the metro opening by lowering travel time and/or increasing travel convenience. Additionally, out of 30,779 passengers using competitive bus routes, it is either not possible or feasible for 5,517 of them to switch to the metro because 76% start and end their journeys outside the metro's service area. Furthermore, the results of the spatial analysis on travel time savings revealed that investments benefit not only those who live nearby but also those who live further away. This emphasizes the significance of enhancing public transportation services, which promotes convenience and accessibility in mobility. Lastly, it is essential to note that the outputs of this study are contingent on assumptions and that varying assumptions will alter the outcomes.

ÖZET

AKILLI KART VERİLERİNİ KULLANARAK MOD DEĞİŞİMİ TAHMİNİ: İZMİR ULAŞIM SİSTEMİNİN VERİMLİLİĞİNİN ARTTIRILMASI

Toplu taşıma sistemlerinin ihtiyaç duyduğu ağır sübvansiyon miktarları iyi bir planlama ve etkin kaynak yönetimi süreçlerini zorunlu kılmaktadır. Planlama süreçleri genellikle yolcu davranışı ve yolculuk talebini belirlemek için maliyetli ve zahmetli geleneksel veri toplama yöntemlerini kullanmayı gerektirir. Bu çalışma bir günlük akıllı kart verisini kullanarak, İzmir'in Narlıdere ilçesine uzatılacak olan metro hattının hizmete açılmasından sonra otobüs hatlarından metroya olası mod değişimini tahmin etmek ve böylece planlama süreçlerinde pasif veri kullanımına katkıda bulunmayı amaçlamaktadır. İlk olarak, literatürde yaygın olarak kullanılan seyahat zinciri algoritması, yolcuların iniş noktalarını tahmin etmek için kullanılmıştır. Ardından, rekabetçi otobüs hatları belirlenmiş ve bu hatlardaki yolcuların seyahatlerini metroyu kullanarak tekrarladıkları alternatif senaryolar oluşturulmuştur. Çalışmanın son adımında, yolcuların seyahat sürelerindeki tasarrufları deterministik olarak ve yolcuk elverişliliği dikkate alınarak iki farklı şekilde hesaplanmıştır.

Sonuç olarak hali hazırda uzatılan metro hattının servis alanında çalışan 39 otobüs hattı tespit edilmiştir. Bunlardan 16 tanesi rekabetçilik yüzdelerini %55'ten fazla olduğu ve bu otobüs hatlarını kullanan yolcuların %30 ila %55' inin yolculuk sürelerini azalttığı ve/veya yolculuk elverişliliğini arttığı için metro hattına geçiş yapabilecekleri tahmin edilmiştir. Ayrıca, bu otobüs hatlarını kullanan yolculardan yaklaşık %18'inin uzatılan metroyu kullanmayı tercih etmelerinin makul olmayacağı tespit edilmiştir. Ayrıca, bu yolcuların büyük bir çoğunluğunun (%76) yolculuklarına metro servis alanı dışında başlayıp bitirdiği tespit edilmiştir. İlaveten, seyahat süresi kazanımlarının mekansal analizi, yatırımların yalnızca yakındaki değil, daha uzak bölgelerde yaşayan bireylere de hareketlilik elverişliliği/erişilebilirliğini iyileştirerek katkı sağladığını göstermektedir. Son olarak, farklı tahmin yaklaşımlarının sonuçları bu çalışmanın çıktılarının varsayımlara bağlı olduğunu ve farklı varsayımların sonuçları değiştireceğini göstermektedir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Public transit is viewed as a low-cost solution to traffic congestion in central business districts and as a more environmentally friendly alternative to private transit due to its ability to reduce carbon emissions (Gao et al., 2022a; Mohammed & Oke, 2022). Although there seem to be many benefits that come with public transit, they all depend tightly on the efficiency of planning and operational processes (Miller et al., 2016). Cautious planning processes are necessary to ensure the efficiency of public transportation infrastructures because they are expensive, particularly rail systems. Mass transit, e.g., bus and metro transit, plays a vital role in densely populated urban areas and has often been seen as an effective way of relieving traffic congestion and improving mobility (Gao et al., 2022a; Mohammed & Oke, 2022; Shao et al., 2022). However, transit agencies must periodically measure the ridership and assess the service quality to ensure efficient, effective, and reliable public transportation system service.

In general, the operation of public transportation services must be subsidized to compete with the rise in private vehicle ownership and use, as well as labor costs (Van Goeverden et al., 2006). In most countries, governments are motivated to cover the deficits by some kind of subsidy scheme to prevent the social exclusion of the handicapped people and to address the urban transit problems (Van Goeverden et al., 2006). However, the increase in the required subsidies, particularly due to increase in the fuel prices, compels the transit operators to allocate resources as efficiently as possible. For example, starting from January 2, 2023, the new fare scheme has been implemented for public transit boardings in İzmir and, under this scheme, the full fare for a boarding is set at 8.78 ₺ (approximately 0.47 $)[1], while the discounted fare is 3.00 ₺ (approximately 0.16 $). Additionally, it has been stated that the average subsidy amount for one boarding is 16.79 ₺ (approximately 0.90 $). Taking into account the distribution in public transit usage based on fare types in İzmir, we can roughly say that the subsidy per boarding covers about 70% of the total cost of a boarding.

[1] On January 2, 2023, 1 USD was equivalent to 18.7180 Turkish Lira.

In this study, we aim to contribute to the operational efficiency of İzmir's bus transit system. We aim to establish a framework in determining the competition between bus routes and rail systems. In addition, we aim to provide results in several aspects, such as travel time savings, that can be used in the operational adjustments. As a case study, Narlıdere metro, which is planned to operate late in 2023, is considered. Our goal is to give the essential inputs for more efficiently aligning competitive bus routes according to the metro line, allowing us to remove unnecessary competition, improve the customer experience, and maximize resource usage. In turn, this will contribute to a more sustainable and accessible transportation network that serves the changing needs of İzmir's expanding population. The significance of this research resides in its potential to improve the quality of public transportation services in İzmir, which would benefit both residents and the environment.

The current metro will be extended in an east-west direction, called Narlıdere metro line in this study, and start to operate in 2023. Several bus routes are currently operating in the service area of the new metro line. The goal is to estimate the impact of the metro extension in terms of passenger ridership change on these bus routes. So that these bus routes can be adjusted to be executed most efficiently, therefore, it is necessary to establish the change in the passengers' route choice to estimate the mode shift from buses to the metro.

This study uses one-day (weekday) multimodal Automatic Fare Collection (AFC) data, i.e., smart card (SC) data. It employs the trip chaining method to estimate the alighting location for each trip. After estimating the alighting stop for each trip, stop-based origin, transfer, and destination matrixes are created for each passenger. Then, these outputs are used in further analyses to improve the public transportation planning process. To fulfill the purposes of this study, we created several algorithms using Python language. It is important to emphasize that all the data manipulation processes are constructed for general daily use, so constructed algorithms can be practically employed and used in strategic transit planning phases. All of the data manipulation work is performed on Spyder 5.4.1 software, a scientific Python development environment. In addition, ArcGIS ArcMap 10.3 is used mainly for map visualization for data analyses, such as calculating the route length of bus lines. We use Microsoft 365 Excel and Power BI Desktop software to create graphs and tables.

For the sake of avoiding confusion, all of the concepts that are discussed in this research are defined in the following way:

- The term “stop” is only used for bus transit.
- The term “station” is used for metro, tram, commuter rail, and ferry transit.
- When the mode is unknown, "location" is sometimes used to refer stops and stations.
- The term “run” defines each bus operation in a direction.
- The concise term “transit” refers to public transit, public transportation, or public transport; these terms are reciprocal and commonly used in literature.

Finally, “trip” is used to refer to each boarding transaction, and “journey” is used to define one or group of trips made to reach from one activity (origin) to another (destination). In other words, a journey is a collection of single or multiple one-way trips and transfers.

1.1. Research Objectives

Mainly, there are three main objectives in this thesis and their brief explanation can be found in the paragraphs below.

Inferring the trips’ alighting locations, identifying transfer and destination locations, and creating origin-destination matrices. Most of the public transit systems use only-entry automatic fare collection systems. Thus, alighting stop information is not available in the smart card data. The trip chaining method is relatively easy (compared to probability and machine learning methods) and capable of estimating the majority of alighting locations. A trip chaining algorithm applicable to our smart card data is developed by adopting similar logic presented by (Trépanier et al., 2007). Details about the developed trip chaining algorithm are presented in section 4.1. After the alighting estimation, we can identify whether the alighting location is a transfer or destination location by implementing the activity time threshold (section 4.1.1.5). After that, the origin-destination matrices can be constructed, as explained in section 4.1.3.

Establishing inter-route relationships. Inter-route relationships can be simplified into competition and cooperation. This process is vital to determine the passenger who may shift from bus to metro after the metro extension. Thus, the bus routes, which carry the passengers who might benefit from the metro extension, must be identified first. A service buffer is created around the metro stations to determine the targeted bus routes.

Then competition and cooperation indices are utilized to characterize these bus routes. This process can be found in section 0.

Estimating the mode shifters to see the ridership change on buses. The main effort is to estimate the change in ridership on buses operating within the service area of the metro extension. To analyze this change, we utilize passenger load profiles. First, we determined the bus runs in section 4.3.1, passenger in-vehicle, and alighting times in section 4.3.4. Then, passenger flow groups are created, and an alternative trips scenario, where the passengers are forced to use the metro to repeat their actualized trip, for each passenger group is identified in section 4.4. Finally, the passenger's decision on shifting to the metro or staying on buses is estimated in section 4.6 employing two different approaches: deterministic and travel convenience.

The results of these objectives can be directly or indirectly used for several purposes. For example, passenger load profiles at both route and run levels can be used in transit performance analysis.

It is important to emphasize that all the analyses in this thesis are performed only using passive data, without using any kind of surveys or any external data gathered from participants or respondents.

1.2. Public Transportation in İzmir

İzmir is the third-largest city located on the western coast of Turkey. According to the Turkish Statistical Institute (TÜİK), the population of İzmir in 2022 is 4,462,056 (TÜİK, 2023). A large proportion of the population (78.6%) lives in densely populated areas. In İzmir, there are 30 districts, and they are colored regarding the district's population in Figure 1.1. The map in the upper right shows the PT trip generation obtained based on the smart card data. The population is densely located in and around the central business district (CBD). Central business districts are designated as Konak, Karabağlar, Karşıyaka, Çiğli, Bayraklı, Bornova, Buca, Gaziemir, Balçova, Narlıdere and Güzelbahçe (İzBB, 2022). As of 2022, the Buca district has the highest population, with 517,963 people.

Public transit services are managed by İzmir Metropolitan Municipality. In İzmir, there are metro, bus, tram, commuter rail, and ferry transit systems, and each allows boarding via smart card. These modes are integrated via transfer centers. The primary transfer centers are located at high-capacity intersections of public transportation routes. It may be in the city's central business district (CBD) or in areas such as airports, stations, and transit terminals outside the city center, where intercity or international transportation connections are available. The majority of public transportation routes, as well as access and transfer for pedestrians and cyclists, are linked to the primary transfer centers. Transfer centers are generally restricted or prohibited private vehicle access zones, typically in city centers. There are eight prominent transfer locations in Izmir. Besides, in İzmir Transportation Master Plan 2030 (UPİ2030), it is expected to have 21 minor transfer locations and 23 transfer points planned for 2030. This indicates the preference towards transfer-oriented development in transportation planning in İzmir. So, the integration, convenience, and quality of these transfer locations will be essential to implementing this development policy in public transportation.

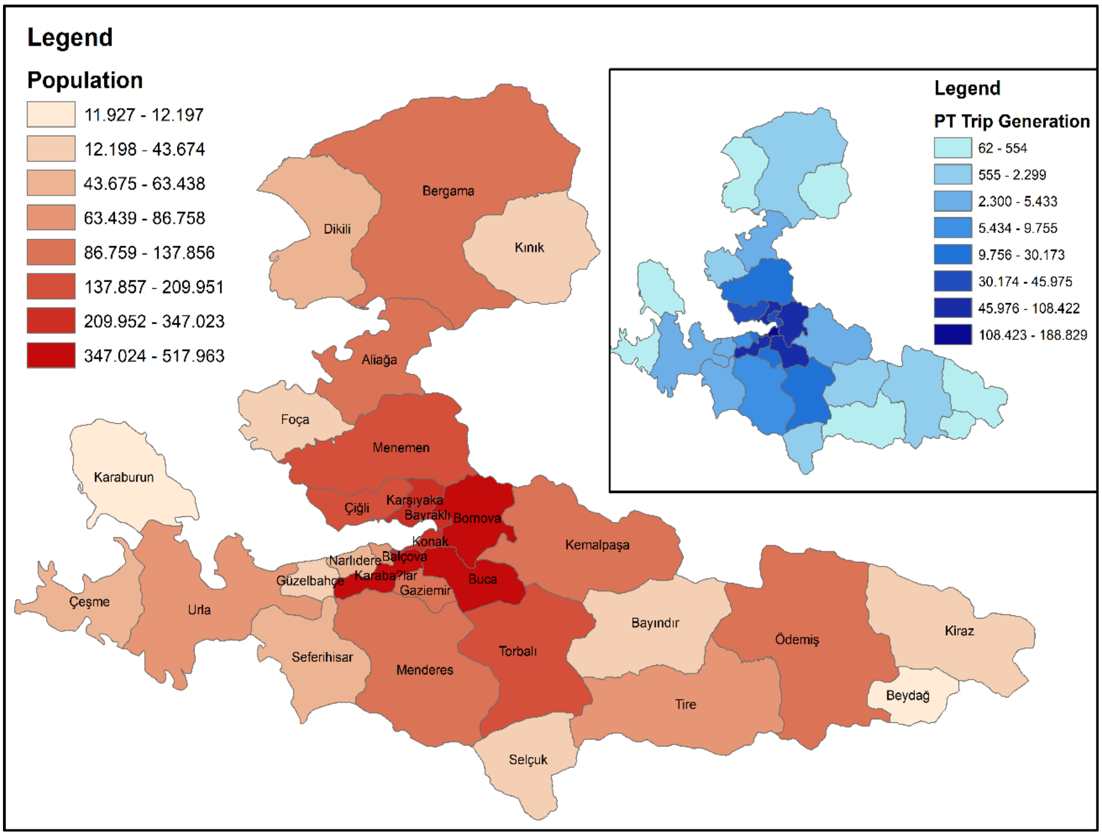


Figure 1.1. İzmir’s population and public transportation trip generation maps.

İzmir's transit system uses several fare policies for several passenger groups. For instance, older people (above 60 years old) can use the transit service charge free. Also, the fare is discounted for students and teachers. Besides, public transportation is free for disabled people, national athletes, relatives of martyries and war veterans, etc. Also, transfers made within 120 minutes are free for students, teachers, and older people. For the full-fare passengers, first and second transfers are about half and one-third of the full fare, respectively. Also, all fares are halved to encourage the use of public transport during rush hour.

Bus transportation in İzmir is administered by the General Directorate of Electricity, Water, Coal Gas, and Public Transport (ESHOT), established in 1943. At the time, ESHOT is responsible for operating 364 bus routes and 12,073 bus stops (6,042 and 6,031 stops for outbound and inbound, respectively), utilizing 1,780 buses of different sizes. On the 29th of December 2019, ESHOT incorporated the minibus operators working as private-public transportation providers. These incorporated minibus operators serve remote areas of İzmir, such as Seferihisar and Kiraz. In Figure 1.2, the bus routes in the outbound direction can be seen. Bus routes densely operate within the central business district (CBD). Besides, there are bus routes that connect remote districts to the CBD. In general, these bus routes operate on very long routes. Table 1.1 presents general information about the longest, shortest, and most-used bus routes. Besides, most long bus routes are operated with very low frequencies, e.g., the headway is one hour for route 987, and route 806 only has two daily runs.

A commuter rail system (İzban) operates on a 136 km route with 41 stations and 219 carriages (see Figure 1.3). It connects several remote districts to CBD. The target is to serve 550 thousand passengers daily (İzBB, 2023). The fare policy adopted on commuter rail service requires passengers to tap both while entering and exiting the system, i.e., the "pay as you go" policy. The passengers are charged with the full fare when boarding and some of the payment is returned, based on trip distance, if the passenger alights before the terminus. This process allows recording true alighting stop information.

There is one metro line (İzmir Metro) serving along the shore at the northeast-southwest axis. It has 17 subway stations; 5 out of 17 are open stations (Hilal, Halkapınar, Stadyum, Sanayi, and Bölge), one station is called a "half-open station" (open but under the ground level), and the rest are underground.

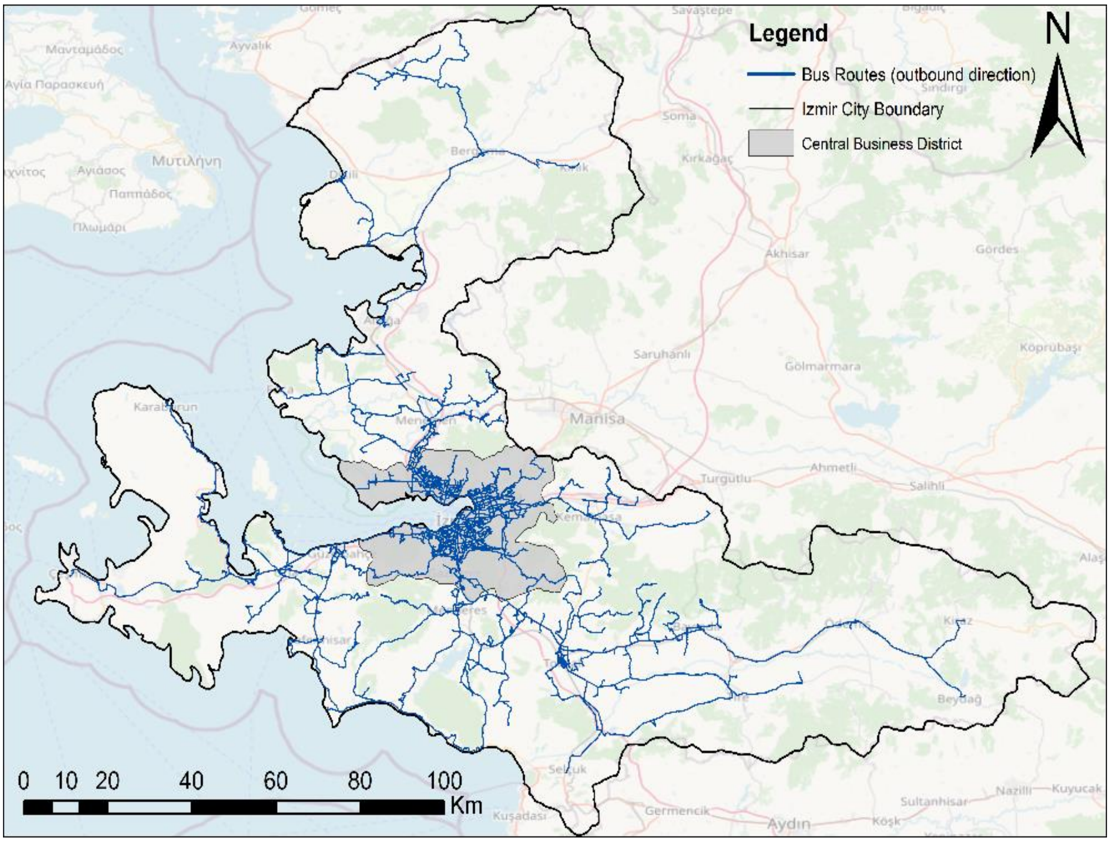


Figure 1.2. Bus routes in the outbound direction in İzmir.

In addition, there are two tram lines (see Figure 1.6): T1 operates between Alaybey and Ataşehir, and T2 operates between Fahrettin Altay and Konak. In public, T1 and T2 are called Karşıyaka and Konak trams, respectively. The Karşıyaka tram commenced its service on the 11th of April 2017; the Konak tram was introduced to the public on the 24th of March 2018, respectively.

Table 1.2 shows total boarding transactions on metro, bus, tram, and commuter rail in respective years. Also, a percent change in the ridership between successive years is given. In İzmir, there were two tram lines (T1 and T2): T1, which operated between Alaybey and Ataşehir, and T2, which operated between Fahrettin Altay and Konak. T1 and T2 were put into service on the 11th of April 2017 and on the 24th of March 2018, respectively. Thus, 1) there were no tram boardings in 2016, 2) In 2017, the number of boardings was low as it served only seven months, and 3) there was an extreme increase (451%) in tram boardings between 2017 and 2018 because the second tram was put into service.

Table 1.1. An overview of the characteristics of bus routes in İzmir.

	Bus Routes	Outbound		Inbound	
		# of Stops	**Length (km)**	**# of Stops**	**Length (km)**
Longest 5	**806**	182	145.72	179	142.34
	795	114	72.71	111	72.63
	761	101	72.37	100	72.33
	987	135	71.59	128	71.00
	837	69	66.23	69	65.03
Shortest 5	**596**	6	2.60	6	4.28
	412	10	3.08	10	2.99
	114	12	3.61	12	3.69
	29	10	3.67	10	3.43
	89	14	3.84	15	3.83
Most used 5	**912**	23	18.57	24	18.52
	304	21	9.37	18	9.30
	800	51	33.64	53	33.15
	691	38	16.33	39	16.53
	680	23	9.18	24	9.31

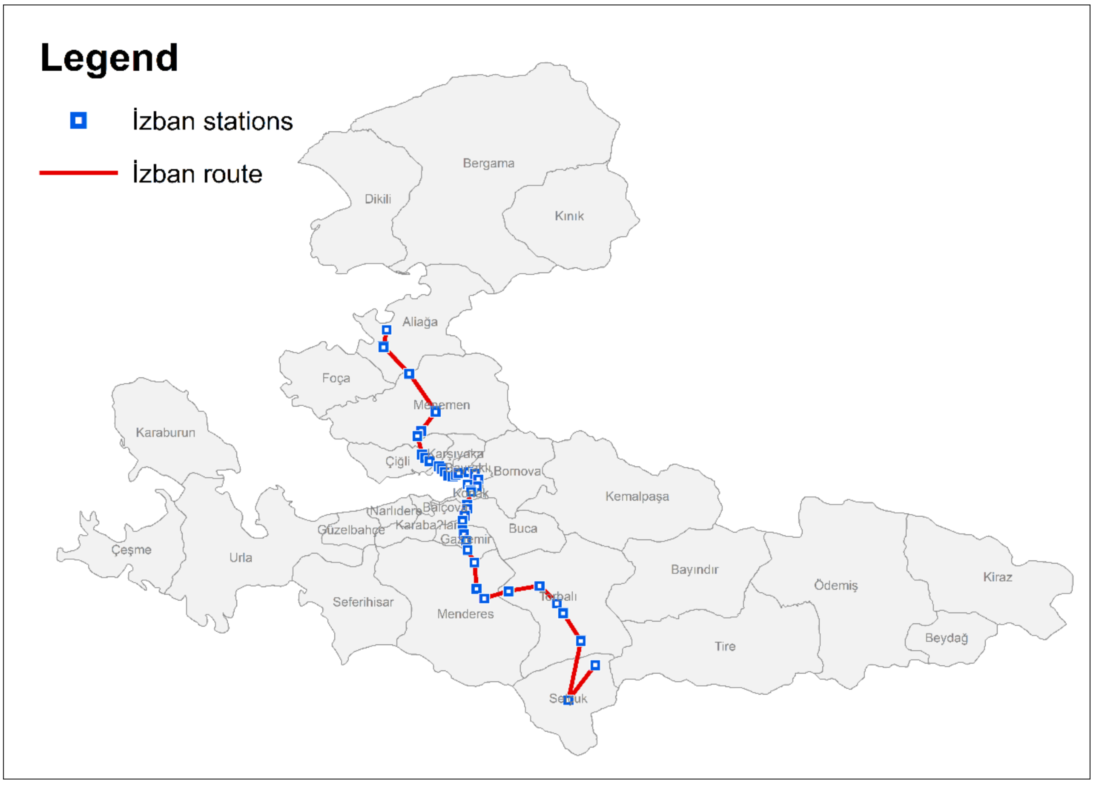


Figure 1.3. İzban (commuter rail) route and station locations.

Important to mention that İzmir also has a docked bicycle sharing system that allows locals to rent a bike using a smart card and credit card. Although the number of stations and bicycles has increased, its primary use is recreational and spatial integration with the transit system, which has several aspects that need improvement.

Table 1.2. Annual transit boarding counts by mode between 2016 – 2021.

	Metro		**Bus**		**Tram**		**Com. Rail**	
Year	Boardings (millions)	Change %	Boardings (millions)	Change %	Boardings (millions)	Change %	Boardings (millions)	Change %
2016	93.85		315.06		0.00		83.83	
2017	96.69	*3%*	307.63	*-2%*	3.39	*0%*	93.42	*11%*
2018	95.95	*-1%*	302.29	*-2%*	18.67	*451%*	80.11	*-14%*
2019	96.20	*0%*	307.37	*2%*	34.14	*83%*	75.82	*-5%*
2020	48.62	*-49%*	174.41	*-43%*	16.69	*-51%*	42.96	*-43%*
2021	57.99	*19%* *(40%)*	201.71	*16%* *(-34%)*	19.80	*19%* *(-42%)*	53.44	*24%* *(-30%)*

Note: Percentages in parentheses are the percent change between pre-pandemic (2019) and 2021, which is a pandemic year.

The model splits are presented in Figure 1.4 based on the smart card data. The first graph of the figure shows the public transportation mode split for a weekday, the 5th of April 2022. The second graph on the figure, which shows the modal split of 2022, proves that smart card data is a good representative of average public transportation use in İzmir. Also, modal split percentages of smart card data align with the yearly modal splits presented in Figure 1.4. There is a clear dominance of bus transit because İzmir's rail systems' service coverage is very limited. Thus, passengers are very dependent on bus transit, for example, the residents in Buca district. This is also an environmental issue since the number of electric transit buses is very small. Besides, bus transit dependency means more road congestion, especially during rush hours.

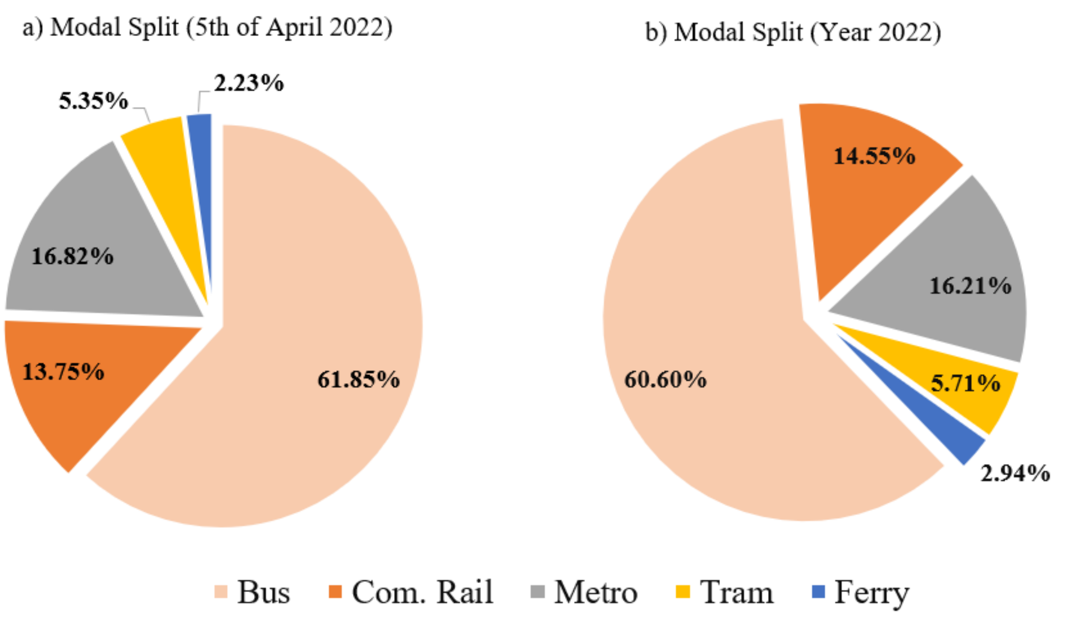


Figure 1.4. a) Modal split of trips b) Modal split of cardholders

Analyses should be done considering the peak hours (rush hours) when the demand reaches its peak on the surrounding road network (Meyer, 2016) or in public transportation. In Figure 1.5, the transit demand by the hours of the day is visualized using smart card data on a weekday, 05.04.2022. The grey bars show the total demand tracked by smart card data and indicate a narrow and clear morning peak between 7 a.m. and 9 a.m. On the other hand, at the evening peak, the demand is spread over a wider range of hours between 4 p.m. and 7 p.m. It is worth noting that a consistent demand for transportation services exists between morning and evening peaks. This demand is particularly evident when considering bus transit, represented by the graph's red line. The weight of bus transit on the overall transit demand becomes apparent from this representation, and it is clear that it shapes the general picture of the transit demand. On the other hand, although the figure does not provide a clear insight into the variation in tram usage, there seems to be a relatively low fluctuation in tram demand throughout the day. This observation suggests that trams also serve leisure trips, likely due to their routes that closely follow the shoreline.

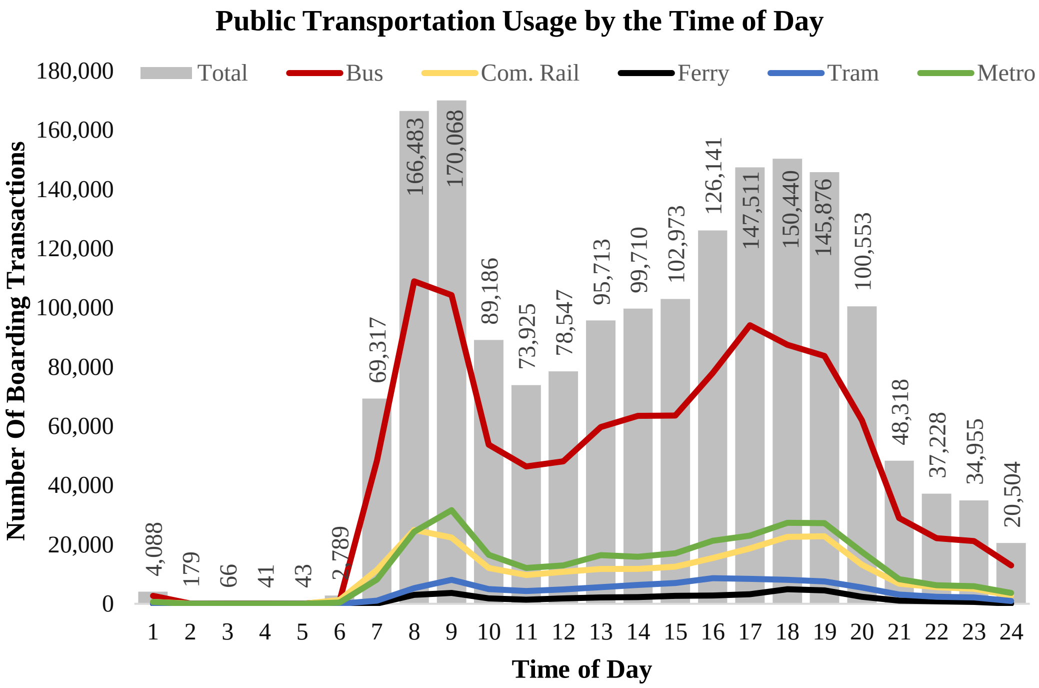


Figure 1.5. Number of boarding transactions by the time of the day

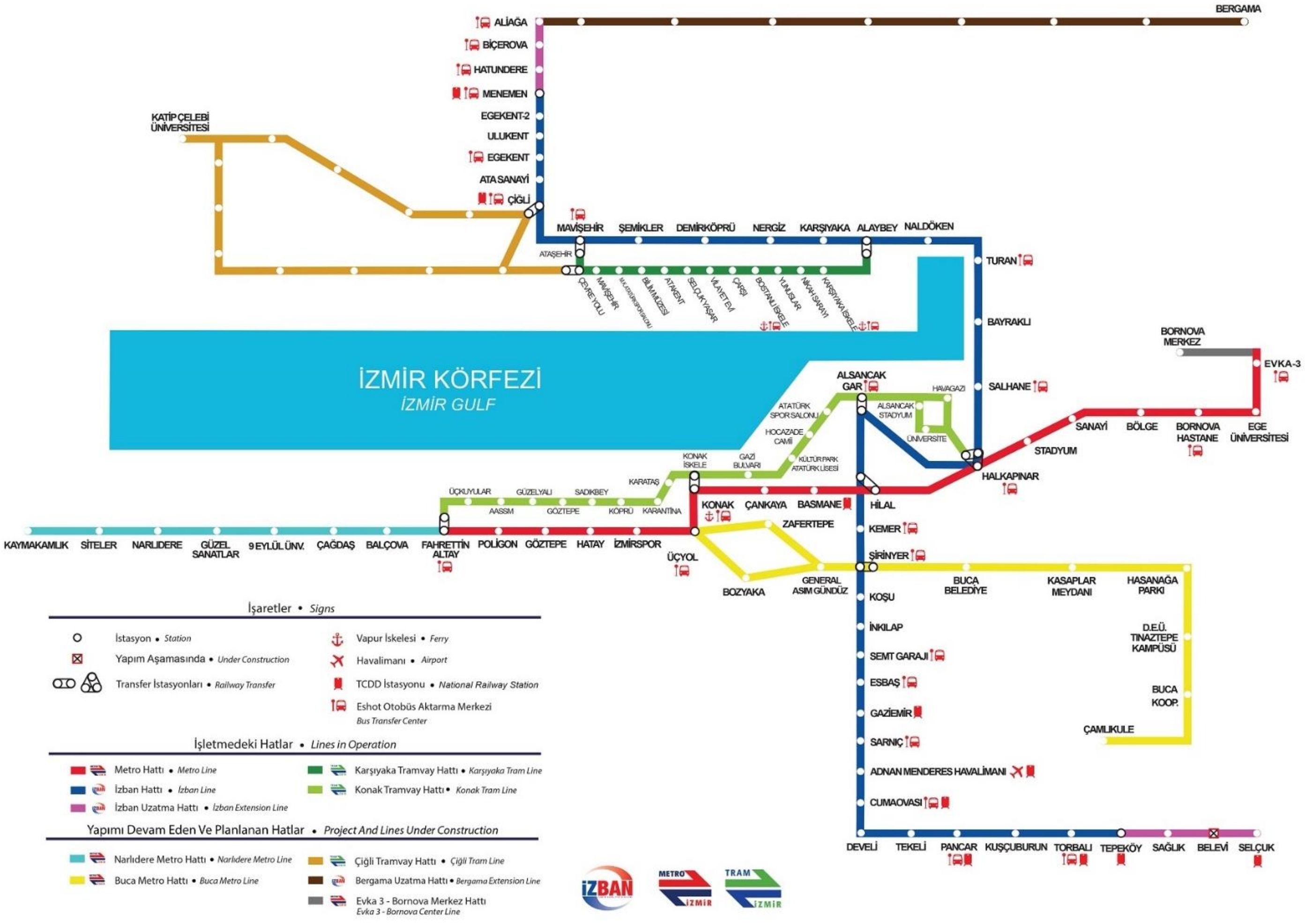


Figure 1.6. The display map of the İzmir's rail systems (İzBB, 2019).

CHAPTER 2

LITERATURE REVIEW

As stated by Meyer (2016), travel can be defined as a means to destinations where activities occur because of economic and personal needs. One of the important concepts in travel demand modeling is viewing travel as a derived demand because it is a result/outcome of activities occurring at destinations. Additionally, passengers are customers of travel, as they are purchasing one of the available modes of travel over another. Since passengers are considered as customers, market segmentation is used to estimate travel patterns more precisely, assuming segments with similar characteristics have similar travel needs. Several attributes are used to simulate the choices of these segments: household income, auto ownership, gender, age, and household structure. As there is a choosing process, one mode over another, it is considered that each available travel mode has its utility, and this utility, which is often the cost of the travel, or the travel time required, differs from one market segment to another (Meyer, 2016).

Understanding and modeling this choosing process is one of the main subjects in transportation planning. Especially for public transport operators, estimating travel demand and responses of the users in case of any service changes, e.g., fare, schedule, or network level changes, is very important because of being dependent on subsidies and accountability mechanisms. As in the definition presented above, passengers' willingness to choose public transportation modes is highly relevant to the utility of the mode presented to the passenger. So, public transportation system operators continuously analyze and evaluate their operations to increase their service quality to be a more appealing alternative for transportation, with the responsibility of utilizing their sources more cost-effectively and efficiently. For this purpose, smart card data can be used as one of many sources.

The literature review section presents the smart card data and summarizes its use in public transportation planning. In this study, we have employed smart card data consisting of tap-in and tap-out data collected by the automated fare collection (AFC) system; thus, studies using smart card data to infer alighting locations are summarized. Later, load profile estimation and its use for determining the bottlenecks in bus transportation are introduced. Finally, the effect of a major network change, e.g., metro line extension, on the current transportation network, i.e., bus routes, is presented based on the literature.

2.1. Smart Card Use in Public Transport

Smart cards (SC) are an application of the AFC system and are used to store and process data, particularly for identification, authorization, and payment (Pelletier et al., 2011). Accordingly, smart cards are used in public transportation systems to validate the user and automatically collect the fare. Depending on the public transit agencies' policy, the AFC system may require passengers to indicate their boarding and alighting activity by tapping the smart card (Mohammed & Oke, 2022). Then, public transit operator stores these individual passenger data collected by the AFC system (A. Cui, 2006).

Despite the major disadvantage of implementing an Automated Fare Collection system based on smart card technology in public transit vehicles being its capital cost, the system also offers notable benefits for both public transportation users and operators. Many transit agencies are adopted smart card based Automated Fare Collection systems around the world (António et al., 2016; Bagchi & White, 2005; Hussain et al., 2021), as its name implied, to automate the face collection, easing public transport use for passenger and increasing the efficiency of revenue collection (António et al., 2016). According to Hao (2007), these advantages can be summarized as follows:

- Smart cards can reduce the cost compared to cash fare payment by improving staff utilization (Hao, 2007). Furthermore, PT drivers can carry out their duties without interruptions caused by payment procedures. This reduced workload increases their performance as they have fewer interactions with passengers (Chira-Chavala & Coifman, 1996; Hao, 2007).
- The capacity and built-in flexibility of the smart card make it possible to create complex and adjustable pricing systems if required (Hao, 2007). This way, fare packages can be arranged specifically for the market segments, such as students, elderly, disabled users, etc.
- With the existence of AFC systems, flexible fare schemes can be implemented more efficiently (Hao, 2007). At the time, in İzmir, transferring between public

transportation modes within 120 minutes is charged approximately half of the full payment. It is free of charge for students, teachers, and passengers older than 60 years old. In addition, to increase public transportation use and decrease the tendency towards private vehicles, fares are half the price on peak hours, between 06:00-07:00 and 19:00-20:00.

- Smart cards can be used on several public transportation systems, thus, increasing the adherence between different PT modes and easing integration (Bagchi & White, 2005).
- Adopting the AFC system based on smart cards decreases the dwelling time on stops (Chira-Chavala & Coifman, 1996; Deri, 2018; Hao, 2007), increasing the reliability (on-time performance) and the level of service.

Finally, smart card data presents an opportunity for collecting individual travel data, which can be used to create passenger flows and better understand travel behavior. The quality of the data can be further increased by combining it with additional data collected from other Automated Data Collection (ADC) systems, such as Automated Vehicle Location (AVL) and Automated Passenger Count (APC) (Jinhua, 2004). Important to note that AFC data lack socio-demographic attributes, so traditional data collection methods, e.g., household surveys, can enrich the data to compensate for this need (Bagchi & White, 2005; Trépanier et al., 2009). Overall quality and reliability of the studies that employed automatically collected data depend highly on the quality of the system that collects and processes the data. However, AFC data often have problems during data collection, possibly caused by hardware, software, or the user (Robinson et al., 2014). For this reason, a data cleaning process is necessary before using the data in an analysis. The following section will discuss using smart card data in subjects related to public transit planning.

2.2. Public Transport Planning

Public transit agencies are generally dependent on the information gathered by manual data collection methods, such as household or OD surveys, which are costly and unreliable, to utilize in planning, managing, and evaluation processes (Zhao et al., 2007). Although AFC systems are used to manage revenue collection in PT systems (Trépanier et al., 2007), they record and provide access to massive amounts of continuous data (available for long periods), which can be linked to individual card and used in PT planning (Bagchi & White, 2005).

There are two kinds of AFC systems: one is the entry-only (open) system that only requires tapping just for boarding as in many cities such as İstanbul, Madrid, New York, Santiago de Chile, Porto, and London; the second is entry-exit (closed) system that requires tapping both while boarding and alighting as in Seoul in Korea and SEQ in Australia (Cengiz, 2022). As explained in SECTİON…, both entry-only and entry-exit AFC systems are used in İzmir. Metro, two tram lines, and most bus routes have entry-only AFC systems, whereas closed AFC systems are used on only 13 bus routes and the commuter rail system (İzban).

Generally, AFC systems work with geographic positioning systems (GPS) and record the following information: card ID, public transportation mode, vehicle ID, driver ID, stop/station ID, route no, route direction, fare segment (student, full-payment, older people, etc.), transfer status and transaction time (Deri, 2018). The availability of this data provide an opportunity for researchers/planners to use the data for public transit planning, including the analysis of PT users' travel patterns and travel behaviors, performance assessment, and planning of the PT systems (T. Li et al., 2018; Pronello et al., 2018), service adjustment and determination of routes' passenger load profile (Pelletier et al., 2011).

In a comprehensive review, Pelletier et al. (2011) used three categories to classify the research areas of the studies on the use of smart card data in public transport. The first category includes strategic-level studies related to long-term planning, such as understanding user behavior, demand management, fare policy analysis, anticipating network extensions, and modeling the loyalty of transit users. The second covers tactical-level studies, such as service adjustment, determination of load profiles, destination, and transfer inference, and determination of origin-destination matrix. The third is operational-level studies, mostly related to performance evaluation, such as schedule adherence, vehicle kilometers, and person-kilometers, and providing real-time information (Pelletier et al., 2011).

2.2.1. Origin Destination Matrix Estimation

The origin-destination (OD) matrix is the fundamental input in transit planning (Hussain et al., 2021). It shows the travel demand between points or areas in consideration, and it may include multiple trips and transfers but does not involve the components of access and egress to public transit (Alsger et al., 2015). The conventional approach to derive OD matrices is surveying transit users, asking about the location of starting and ending points of their daily trips. However, generally, surveying is cumbersome and comes with biases. One of the automated fare collection (AFC) data applications in transit planning is estimating the OD matrix of transit users (Hussain et al., 2021).

As Mohammed and Oke (2022) indicated, two types of OD matrices can be considered: route-level OD matrices and network-level OD matrices. Route-level OD matrices show the passenger flow from a stop/station on a route to another stop/station on the same route. Route level OD matrices are primarily utilized in planning decisions, such as re-scheduling the vehicles, building new routes, or extending the existing ones. Network-level OD matrices show the true origin and destinations of trips, including transfers. Once the OD matrices are developed, the passenger flow between traffic zones can be obtained and used in planning decisions, such as constructing a new transit alternative to supply the demand (Mohammed & Oke, 2022).

The studies in the literature that used the trip chaining method to create origin-destination matrices are presented in Table 2.1.

2.3. Trip Chaining Method

The method was first presented by Barry et al. (2002) for inferring the destinations of metro users using entry-only AFC system data of New York City Transit (Barry et al., 2002). In the last two decades, the trip chaining method has been used in many studies (Table 2.1) to infer the alighting stop of passengers' trips and is further used to establish route or network-level OD matrixes.

Table 2.1. Studies used trip chaining algorithm, sample information, matching, and validation rates. Validation rates are in parentheses.

Reference	Location	Data Source	Sample length	Sample Year	Sample Size	AFC system	% Matching / (Validation)	Validation Source	Mode
(Gordon et al., 2013)	London, England	Oyter and iBus	10-weekday	6-10 and 13-17 June 2011	3-3.1million per day	Entry only	74.50	/	Bus, Rail
(Jánošíkova et al., 2014)	Slovak Republic	DPMŽ	1 week	12-18 Oct. 2009	115,007	Entry only	80.72	/	Bus, Trolleybus
(Trépanier et al., 2015)	Ontario, Canada	STO	1-month	October 2009	908,303	Entry only	80.64 [1]	/	Bus
(Huang et al., 2020)	China	/	/	/	/	/	80 (90)	Survey	Bus
(Jafari Kang et al., 2021)	Iran	7 BRT lines	2 months	Sep. Oct. 2016	716,000	/	52.00	/	BRT
(Kumar, 2019)	Twin cities, America	Metro Transit U-pass	4 days	7-10 March 2016	85,456	Pay-exit	85.00	/	Bus, Rail
(Deri, 2012)	İzmir, Turkey	/	1 week each	March 2009 June 2010	378,260 771,239	Entry only	80.77 83.01	/	Bus, Rail, Ferry
(Palamutçuolu & Gerşil, 2020)	İzmir, Turkey	/	15 days 30 days	2015 2017	1,142,515	Entry only	66.05	/	Bus, Rail, Ferry
(Lee et al., 2022)	Sejong City, South Korea	/	42 weekdays	1 Apr. - 31 May 2018	116,194	Entry-exit	60 - 74.9 (65.6) [2]	Tap-out data	Bus
(Cengiz, 2022)	Madrid, Spain	CRTM	1 weekday 1 weekend	4 Feb. 2020 1 Feb. 2020	709,170 441,086	Entry only	/	/	Bus, Train
(Kim et al., 2017)	Seoul-Gwangju, Korea		1 day 1 week	20-Oct-2015 18-24 April 2016	4,614,149 959,578	Entry-exit	78.2 (93.6) 81.6 (94)	Tap-out data	Bus
(Fidanoğlu & Gökaşar, 2022)	Bursa, Turkey	8 bus routes	3 months	/	1,861,791	Entry only	70%	/	Bus
(Yan et al., 2019)	Beijing, China	Downtown bus data	1 week	8-14 October	/	Entry-exit	87.23-62.66 [3] (72.98-71.88)	/	/
(Mosallanejad et al., 2019)	Adelaide, Australia	DPTI	1 day	May 2017	1,177	Entry only	80.2 (98%)	OD Survey	Bus
(Farzin, 2008)	São Paulo, Brazil	/	/	2006	658,000	Entry only	76.7 [4]	Household survey	Bus

Notes: 1) In the study, unlinked trip destinations were estimated using the Kernel Density method, and the matching rate was increased by 10.9%. **2)** The clustering method improved the matching rate of the base algorithm from 60% to 74.9%. The exact validation rate was 48.2%, validation rate was 65.6% within one stop difference. **3)** Matching and validation rates were for regular and irregular users. **4)** Destinations were inferred at the zone level.

Table 2.2. Studies used trip chaining algorithm, sample information, matching, and validation rates. Validation rates are in parentheses.

Reference	Location	Data Source	Sample length	Sample Year	Sample Size	AFC system	% Matching / (Validation)	Validation Source	Mode
(Assemi et al., 2020)	SEQ - Australia	GoCard - TransLink	1 weekday	20 Mar. 2013	159,815 (138,122)	Entry-exit	72.2 (79.5)	Tap-out data	Bus, Train, Ferry
(Barry et al., 2002)	America	NYCT	1 weekday	Wednesday, September	6,000,000	Entry only	83 (90)	Travel diary, metro exit counts	Metro
(Jinhua, 2004)	America	CTA	1 week	/	2,602,819	Entry only	65.5	/	Train
(A. Alsger et al., 2016)	SEQ - Australia	GoCard - TransLink	1 weekday	20 Mar. 2013	161,446	Entry-exit	86.5 (72.6)	Tap-out data	Bus, Train, Ferry
(Nunes et al., 2016)	Porto, Portugal	STCP	1 month	April 2010	~3000000	Entry only	81.60	/	Bus
(M. Munizaga et al., 2014)	Santiago, Chile	Transantiago	/	/	715 [5]	Entry only	84.2 (90)	Metro OD survey	Bus, Metro
(Munizaga et al., 2012)	Santiago, Chile	Transantiago	2 weeks	March 2009 June 2010	36,000,000 38,000,000	Entry only	80.77 83.01	/	Bus, Metro
(D. Li et al., 2011)	Jinan, China	Bus company	/	/	10,000 per day	Entry only	75-85 [6]	/	Bus Route 115
(W. Wang et al., 2011)	London, England	TfL	/	/	7,386 [5]	Entry only	66 65	BODS	Route 185 NB and SB
(Nassir et al., 2011)	Twin cities, America	Metro Transit	1 weekday	10 Nov. 2008	84,413 (10,886)	Entry only	60.74% (95.4%)	APC and AVL	Bus
(Barry et al., 2009)	America	NYCT	2-week	April 19-May 2, 2004	>7,000,000 for a weekday	Entry only	(90%)	Metro exit counts, bus ride checks	Subway, bus, ferry, tram
(Zhao et al., 2007)	Chicago, America	CTA	6 days	January 2004	>2,500,000	Entry only	71.20%	/	Bus, Train
(Trépanier et al., 2006)	Gatineau, Canada	STO	/	July 2003 October 2003	378,260 771,239	Entry only	66%	/	Bus
(A. Cui, 2006)	America	NYCT	5 weekdays	12-16 Sep. 2005	2,736,454	Entry only	79%	/	Bus, Rail
(Cheng et al., 2019)	Guangzhou, China		3 months	1 Jul. - 30 Sep. 2017	200,670	Entry-exit	85.27 (60.32-87.4)	Tap-out data	Metro

Notes: 5) Validated sample sizes were used in the model. 6) Off-peak and peak hour matching rates, respectively.

Table 2.1 summarizes 30 studies that used the trip chaining method in their analysis. The majority of the studies were from America and China, 8 and 4, respectively. As can be seen, the sample size ranges from 715 to 38 million records depending on the purpose of the study. Similarly, the sample lengths range from 1 day to 3 months, depending on the assumptions in the trip chaining algorithm and the historic data requirement of additional methods to improve the matching performance. 21 out of 29 studies employed entry-only AFC data (one was pay-exit, which provides the alighting stop instead of the boarding stop), whereas six studies had information on both boarding and alighting stops of each record (entry-exit system). Since it is important to validate the inferred stops and obtain the effect of trip chaining assumption on the estimation accuracy, entry-exit AFC data provides a valuable opportunity for endogenous validation.

On the other hand, the validation process for the studies that used entry-only AFC data requires exogenous sources, such as travel diaries, OD surveys, automated passenger count (ADC) data, tally counts on stations, etc. For example, Farzin (2008) utilized 1997 OD matrices developed based on the household survey to validate only the 5% of 2006 OD matrices estimated in their study (Farzin, 2008). Barry et al. (2002) used a metro travel diary containing the origin and destination information of 300 residents to validate the accuracy of their main assumptions on metro trips (Barry et al., 2002). However, in the case of facilitating such exogenous sources, the sample sizes are generally much smaller, thus creating a bias in validation. More reliable data sources, which include real alighting information (location and time) as in entry-exit systems, can give insightful information on the validity of trip chaining assumptions (A. Alsger et al., 2016).

2.3.1. Method Assumptions

The method was first presented by Barry et al. (2002) for inferring the destinations of metro users using entry-only AFC system data of New York City Transit. Their trip chaining method was built on two main assumptions: 1) the majority of passengers alight at the station where they start their next trip, and 2) the majority of passengers return to their origin (boarding station of their very first trip) at the end of their last trip (Barry et al., 2002). These two baseline assumptions have been improved, and new assumptions have been added by researchers to make the algorithm compatible with different data types collected from PT systems worldwide and to increase either the matching or validation rates.

Hora et al. (2017) and Cerqueira et al. (2022) summarized all the trip-chaining assumptions established so far. Based on their compilation, once again, the main assumptions used in trip-chaining studies are listed chronologically as follows (Cerqueira et al., 2022; Hora et al., 2017):

1. The majority of passengers alight at the station where they start their next trip (Barry et al., 2002);
2. The majority of passengers return to their origin (boarding station of their very first trip) at the end of their last trip (Barry et al., 2002);
3. The alighting stop of a trip cannot be inferred if it is the only trip of the cardholder for the day (Barry et al., 2002). This type of trip is termed as a unit (Trépanier & Chapleau, 2006) or a single trip, as in this study;
4. The alighting stop cannot be inferred if consecutive transactions (boardings) occur at the same station (Barry et al., 2002);
5. Passengers can only alight at a subsequent stop on the given direction of the route (Trépanier & Chapleau, 2006);
6. The distance between the boarding stop of the next trip and the alighting stop of the previous trip must be less than the allowable walking distance (Trépanier & Chapleau, 2006);
7. The alighting stop of the last trip of the day must be within the allowable walking distance of the boarding stop of the first trip of the day (Trépanier & Chapleau, 2006);
8. In case of failed alighting stop estimation for the last trip on a day, the model can use the first boarding stop of the following day to infer the alighting stop of the last trip (Trépanier & Chapleau, 2006);
9. For a trip, which the alighting stop estimation is failed, all the other trips of the month for the same user are compared to find a similar trip where the destination would have been found (Trépanier & Chapleau, 2006), i.e., historical data is used to find a similar trip pattern of a passenger;
10. Passengers do not use any other transportation modes, i.e., shared mobility, private car, etc., but walk between any consecutive trip segments (Zhao et al., 2007);

11. As a temporal check, estimated alighting action (alighting timestamp) must occur before the boarding time of the next trip (Nassir et al., 2011);
12. A passenger is required a certain time to perform an activity, activity threshold (Nassir et al., 2011);
13. A passenger must transfer in a given time, transfer threshold (Nassir et al., 2011);
14. If the model is not able to infer the alighting stop of the trip in the given direction, the model can be relaxed by searching a possible alighting stop in the opposite direction (Nassir et al., 2011);
15. The hour with the lowest activity in the day can be used as the virtual midnight, for example, 3 a.m. (M. Munizaga et al., 2014).

Some of the assumptions above are the relaxed form of their primer versions; for example, the first assumption is relaxed by using the sixth assumption, which allows the algorithm to search for an alighting stop within the allowable walking distance of the previous trip's boarding stop. Similarly, the seventh assumption can be preferred instead of the second assumption for relaxing the algorithm. Here, the matching rate and accuracy of the trip chaining algorithm became a function of the selected allowable walking distance, discussed in 4.1.1.3. In Table 2.3, the maximum (acceptable) walking distances used in various studies are presented. Obviously, different values have been considered for allowable walking distance in trip-chaining studies because of several logical reasons, such as public transit network and the architecture of the city in analysis, demographics, terrain, transfer conditions, and PT mode involved in the analysis (Hussain et al., 2021). The trip chaining method is capable of estimating the destination of trips (more specifically, linked trips) with a matching rate of around 60% to 85% depending on the rules applied in the algorithm (Cheng et al., 2019).

2.3.2. Method Deficiencies

There are several weak spots in the trip chaining method. First, the algorithm is not cable of including the non-integrated modes of transport, such as bike-sharing systems, taxis, car-sharing, or minibuses, and this is one of the main sources of error in estimating the alighting stop (M. Munizaga et al., 2014). Second, the method cannot estimate the alighting location of an unlinked trip. As illustrated in Figure 2.1, the trip chaining algorithm cannot infer an alighting stop for the second and third trips. Finally, the effect of the built environment on the passengers' decisions is difficult to implement.

Table 2.3. Allowable walking distances used in trip chaining studies.

Study	Location	Walking Distance (meters)
Trepanier 2006	Gatineau, Canada	2000
He 2015	Canada	2000
Palamutçuoğlu 2020	İzmir, Türkiye	1500
L'udmila 2014	Slovak Republic	1250 [1]
Cui 2004	NYC, America	1100
Munizaga 2012, 2014	Santiago, Chile	1000
Wei Wang 2011	London, England	1000
Deri 2012,2018	İzmir, Türkiye	1000
Yang 2019	Beijing, China	1000
Assemi et al. (2020)	SEQ - Australia	800
Alsger 2015, 2016 [5]	Australia	800
Nassir 2011	Twin cities, America	800
Kumar 2019	Twin cities, America	800
Gordon 2013	London, England	750
Cengiz 2022	Madrid, Spain	650
Nunes 2016	Porto, Portugal	640
Lee 2022	Sejong City, South Korea	500
Kim 2017	Seoul - Gwangju, South Korea	500
Jinhua Zhao 2004, 2007	Chicago, America	402
Da (2021)	Nanjing, China	400
Mona 2019	Adelaide, Australia	400
Yap 2017	Hague, Netherland	400 [6]
Fidanoğlu	Bursa, Türkiye	300-500 [4]

Notes: 1) 15 minutes of walking distance at 5 km/h walking speed. 2) 500 meters is for transfer walking distance; 300 meters is for last trip. 3) Various walking distance values ranging between 200 m to 1,600 m were used in the study.

Alternative methods, machine learning, and probability models often perform better than trip chaining. In terms of the matching rate of the mentioned methods: probability models show the best performance with matching rates close to 90%, and a deep learning model employed by Jung and Sohn (2017) has 60% and 87% matching rates for tight and relax criterion, respectively, trip chain models (see Table 2.1) have matching rates ranging between 65 and 95.4% (T. Li et al., 2018). A combination of the models also exists; for example, Assemi et al. (2020) used ANNs and improved the matching rate from 72.2% to 79.5%

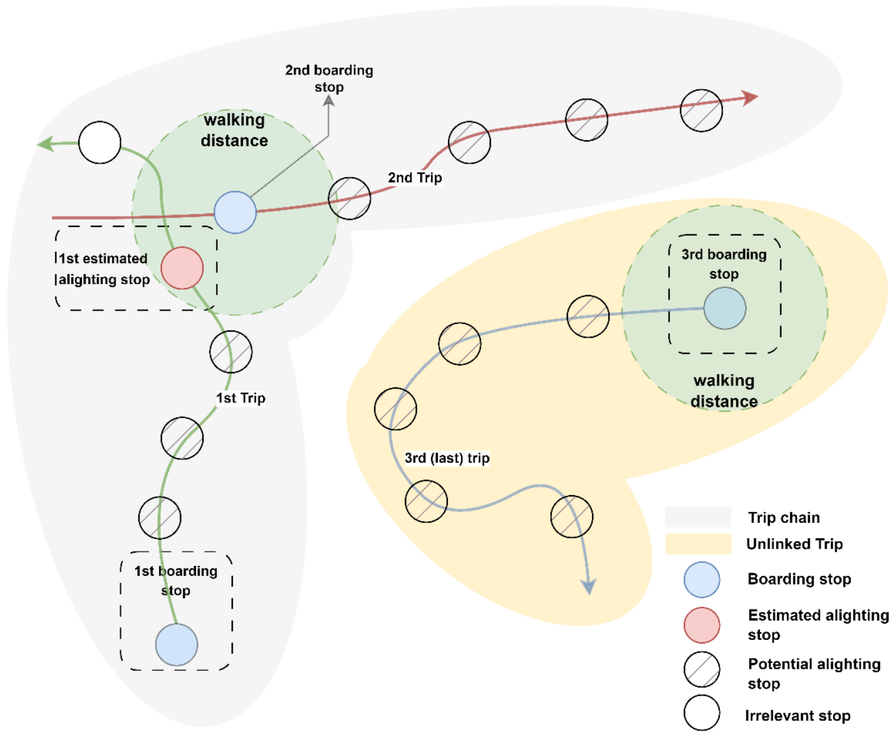


Figure 2.1. Illustration of unlinked and chained trips (Z. Cui et al., 2021).

2.4. Public Transport Service Area

In most cases, a public transport user must walk, cycle, drive or share a ride to reach the closest or preferred public transport mode. Because public transportation systems typically serve major destinations and stops or stations are generally present at centralized locations, making it difficult for passengers to travel from their homes to stops or stations (access) and vice versa (egress). A public transport (PT) system's overall performance greatly depends on access and egress connections. Further, access and egress distance or time are the most significant factors considered in public transport mode choice (Rahman et al., 2022). Hence, a public transport system's service (or influence) area is derived from the distance passengers are willing to walk to access or egress to a station, and it is a key performance measure because it is directly related to the percentage of the population being served (El-Geneidy et al., 2014).

In transit-oriented development (TOD), there is a general acceptance of public transport (PT) planning for walkable access (or egress) distance, which is a quarter mile (400 m) for the bus stops and half a mile (800 m) for the rail stations (Canepa, 2007; El-

Geneidy et al., 2014). Nevertheless, these de facto values may depend on various factors, such as trip characteristics; type of PT service, transfers, waiting time, or personal characteristics; household, gender, age, or environmental factors (El-Geneidy et al., 2014; van Soest et al., 2020). The recent increases in the usage of shared micro-mobility vehicles, such as bicycles, e-bikes, and e-scooter, may have increased the distance of a passenger willing to travel to access or egress, hence, the influence area of public transport modes.

2.4.1. Inter-route Relationships

In public transportation planning, the inter-route relationship is related to the physical connection between two or more routes. These relationships can be established by considering the overlapping service areas of the routes.

Several studies in the literature have defined and established a set of factors for identifying inter-relationships between public transportation modes. According to Peng (1994), three inter-route relationships may be defined: independent, competing, and complementary. These routes are independent if no overlapping route buffers (service areas) exist between the routes. If the route buffers overlap, the relationship can develop in two ways; competing or complementary. The routes are complementary if their service areas intersect at one end, such as a terminus, or an intermediate location, such as a transfer location, and at least one end of the routes is different. Bus routes integrated with metro lines have this kind of inter-route relation and generally operate as a feeder mode. On the other hand, if route buffers overlap linearly and have at least one common end, they are in a competing relationship. For example, bus routes running on the same or parallel streets or a bus route running on the same corridor with a metro line is in a competing relationship (Peng, 1994).

Consider two bus routes operating within the service area of a metro line with a similar amount of overlapping service areas, as illustrated in Figure 2.2. Bus route A collects passengers from the urban outskirts and runs mainly perpendicular to the metro line. Consequently, this type of bus route undertakes a feeding role for the metro line, gathering and evacuating passengers (F. Wang et al., 2022), and shares a complementary relationship. On the other hand, bus route B runs primarily parallel to the metro line. In this case, the bus route and metro line have a competitive relationship. As it can be seen from Figure 2.2, it will be misleading to say whether the route has a competing or complementary relationship by only considering the amount of overlapping area. Here, collinearity can be utilized to distinguish this difference.

In literature, there are different collinearity definitions based on a similar logic (J. Cui et al., 2020; Du et al., 2018; F. Wang et al., 2022; Wei et al., 2020). It is basically obtained by calculating the linearly overlapped service areas of the routes under consideration. The result is further used to determine the degree of competition. The assumption is that when the collinearity between routes increases, the level of competition between routes also increases (Peng, 1994), thus the likelihood of passenger flow exchange from one route to another (F. Wang et al., 2022). These relationships can also be viewed from the perspective of the passengers, who have the option to choose one of the competing routes to reach their destination, on the other hand, they must have a transfer from one complementary route to another to reach their destinations (S. J. Zhang et al., 2018). Since high collinearity between public transport systems, particularly between bus routes and rail system, increases the likelihood of passenger exchange, collinearity is considered a significant factor in bus route optimization studies.

Du et al. (2018) used three categories, which are collinear, feeding, and intersecting routes, to classify bus routes based on their relationship with the rail transit stations. Their main focus was the route optimization of competing and intersecting bus routes. In their study, 500 m buffers were used on the train stations to define the direct attraction area of a station. As for the collinear relationship, they regarded a bus route with stops within the direct attraction area of two or more successive train stations. On the other hand, a bus route was considered to have a feeding relationship if its stops were within one train station's direct attraction area. If the rail transit route and the conventional bus route intersect without the direct attraction area of rail transit stations, they have an intersecting relationship (Du et al., 2018).

Wang et al. (2022) considered the collinearity between bus routes and a rail track as an indicator in their bus route optimization study, as well as other factors such as travel time and service area. The collinearity minimization effort is due to the correlation between collinearity and competition, and minimizing collinearity is expected to solve the competing routes problem. They simply defined the bus routes' collinearity as the ratio of collinear route length with train track to total route length (F. Wang et al., 2022).

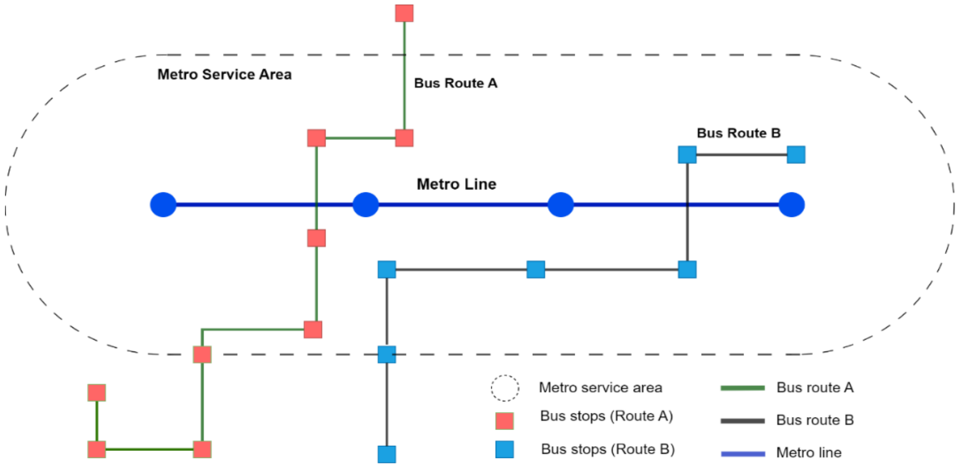


Figure 2.2. Illustration of competing (bus route B) and complementary (bus route A) bus routes within a metro service area.

In another study on bus route optimization, Wei et al. (2020) developed cooperation and competition indices for bus routes to identify those that require optimization. They used 500 and 800 meters of buffers to define the service areas for bus stops and metro stations, respectively. They defined the cooperation index (coo) as the ratio of the total number of cooperation stations (T_{coo}) to the total number of bus stops and metro stations (T_{BM}) (Wei et al., 2020).

$$\mathrm{coo} = \frac{\mathrm{T_{COO}}}{\mathrm{T_{BM}}} = \frac{(\mathrm{T_M} - \mathrm{N})(\mathrm{T_B} - \mathrm{N})}{(\mathrm{T_M} - 1)(\mathrm{T_B} - 1)} \tag{2.1}$$

$$\mathrm{T_{BM}} = (\mathrm{T_M} - 1)(\mathrm{T_B} - 1) \tag{2.2}$$

$$\mathrm{T_{coo}} = (\mathrm{T_M} - \mathrm{N})(\mathrm{T_B} - \mathrm{N}) \tag{2.3}$$

Furthermore, Wei et al. (2020) defined the competition index (com) as the ratio of competition stations (T_{com}) to the total number of bus and metro stations (T_{BM}). As can be seen, they incorporated spatial variables, such as the overlapping area of the bus stations and the length between bus and metro stations. In the formula, N is the total number of overlapping stations, s_b is the buffer area of a single bus station, s_i is the overlapping area of bus station number i, and s_{i+j} is the overlapping area of bus station number $(i + j)$, $L_b^{i,i+j}$ is the travel distance between bus station number i and $(i + j)$, and $L_m^{i,i+j}$ is the travel distance from metro station number i to $(i + j)$ (Wei et al., 2020).

$$\text{com} = \frac{T_{com}}{T_{BM}} \tag{2.4}$$

$$T_{com} = 2 \times \sum_{i=1}^{N} \sum_{j=1}^{N-1} \left[\frac{\Delta s_i}{s_b} \times \frac{\Delta s_{i+j}}{s_b} \times \frac{\min\left(L_b^{i,i+j}, L_m^{i,i+j}\right)}{\max\left(L_b^{i,i+j}, L_m^{i,i+j}\right)} \right] \tag{2.5}$$

$$T_{BM} = (T_M - 1)(T_B - 1) \tag{2.6}$$

Overall, they investigated 245 bus routes and 12 of them were determined as competitive, while 157 bus routes marked as cooperative to the Changsha Metro Line 2. Since the remaining ones (76) showed both properties, they were identified as competitive and cooperative. In addition, authors combined cooperation and competition indices to develop a co-opetition coefficient for their optimization (Wei et al., 2020).

Cui et al. (2020) aimed to determine the bus routes that would need to be adjusted prior to the launch of the rail transit. First, they started by identifying the bus routes collinear with the rail transit line within the 750 m service buffer of the rail line. They considered a bus route to be collinear if its direction was parallel and within the service area of the rail line. In their study, three types of collinear lines (see Figure 2.3) were defined: fully collinear line, endpoint collinear line, and intermediate collinear line. Further, if the collinear part of a bus route was longer than 6 km, it was selected as the bus route that required adjustment. Then, they further analyzed the passengers' travel patterns by employing a generalized cost function to determine the flows that would have benefited if the passengers were used the train in the collinear section instead of bus (J. Cui et al., 2020).

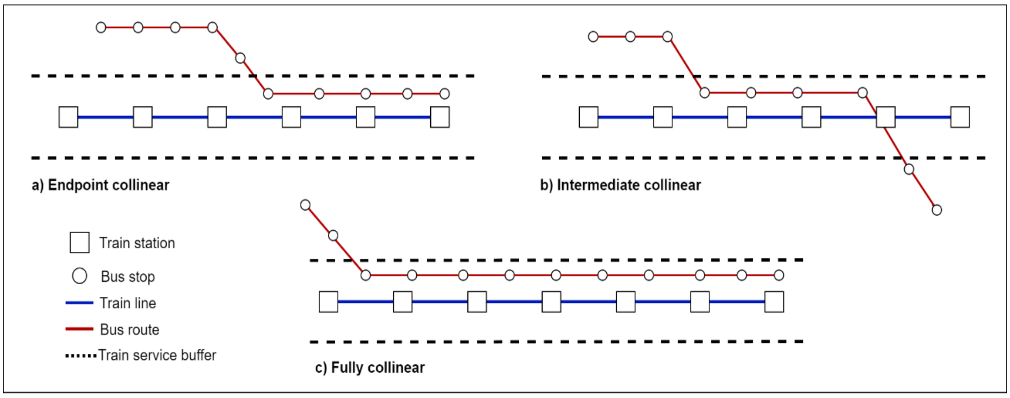


Figure 2.3. Types of collinearity: a) endpoint collinear, b) intermediate collinear, c) fully collinear. (J. Cui et al., 2020)

Gadepalli et al. (2022) used 400 and 2000 meters of network buffers on metro stations to determine the candidate bus stops which can be used to transfer to the metro. To determine the overlapping bus routes, they select the bus routes having at least two stops in the 400 m metro influence area. Then, they further characterized the overlapping bus routes as parallel-competing, parallel-non-competing, and feeder based on the number of stops of a bus route within the 400 m metro influence area as a proxy to collinear distance. Feeder routes were identified as having less than 25% of their stops within 400 m of the metro. On the other hand, bus routes, which have more than 25% of their stops within 400 m of the metro, were identified as parallel routes. Additionally, parallel routes were grouped into two; if the route length within 400 m influence area is higher or equal to 2 km, it was labeled as competing, if not, marked as non-competing route. Then, in the trip level analysis, they categorized the bus trips into two categories, short overlap trips (less than 2 km) and long overlap trips (equal or more than 2 km). Their assumption was that long overlapping trips will likely shift from bus to metro.

In Table 2.4, we summarize the terms and definitions found in the studies that aim to determine competitive bus routes within the rail transit service area. In conclusion, the collinearity of a bus route can be identified by determining the number of stops or the length of the route within the rail transit service area. And, whether the bus route has a competitive or complementary relationship with the rail transit line depends on the amount of its collinearity. There are several approaches employed in literature to establish this distinction, for example Luo et al. (2010) combined ecological theories with transit characteristics to assess the competition between rail and bus based on Lotka-Volterra model (Q. Y. Luo et al., 2010).

In this study, a slightly altered version of the methodology developed by Gadepalli et al. (2022) is used to identify the bus routes that operate within the service area of a rail system (Gadepalli et al., 2022). In addition, we adopt a simplified version of the method presented by Wei et al. (2020) to determine the nature of the relationship (Wei et al., 2020).

2.5. Valuating Travel Convenience

After the opening of the metro extension, passengers traveling within the metro service area will have an alternative mode to get to their destinations. In this study, one of the objectives is to find an answer to the question of which passengers would prefer to use the metro service instead of the current mode of transportation, which is mainly limited to bus transit. In the way of answering this question, we implement the approach where we benefit from the valuation of convenience (or inconvenience) of public transport. For this purpose, perceived travel time studies, where the objective is to reveal the difference between real and perceived travel time components, are reviewed. The main gain from the review is the time multipliers of the travel stages to calculate the inconvenience of the two alternatives: bus, or metro. Since we work with time multipliers instead of monetary values, some degree of transferability is conveniently available (Wardman, 2014). In addition, travel time perception studies are also reviewed to establish general understanding on the importance of the perception as well as the travel convenience. Before going further, the time components considered in a passenger's journey are presented with their definitions.

Table 2.4. Inter-route relationship categories and definitions in the literature.

Study	Terms	Definitions
(Peng, 1994)	Complementary	intersects at one end or at an intermediate location, and at least one end is different
	Competing	one common end and linear overlapping
	Independent	no overlapping
(Du et al., 2018)	Collinear	having stops within the direct attraction area of two or more consecutive train stations
	Feeding	having stops within the direct attraction area of one train station
	Intersecting	intersects without being in a direct attraction area of a train station
(Gadepalli et al., 2022)	Feeder	routes having less than 25% of their stops in 400 m influence area
	Parallel	routes having higher than 25% of their stops in 400 m influence area
	Competing	parallel routes having 2 km or more route length in 400 m influence area
	Non-competing	parallel routes less than 2 km route length in 400 m influence area
(Wei et al., 2020)	Cooperation index	ratio of total number of cooperation stations to the total number of bus and metro stations
	Competition index	the ratio of competition stations to the total number of bus and metro stations
(F. Wang et al., 2022)	Collinearity	ratio of collinear route length with train track to total route length
(J. Cui et al., 2020)	Collinear route	a bus route is collinear if its direction is parallel to and within the service area of the train line, three types: endpoint, intermediate and fully collinear

2.5.1. Travel Time Components in a Passenger's Journey

There are several stages observed in the passengers' journey made on public transit. The complete travel time of a passenger consists of in-vehicle and out-of-vehicle components. In some cases, especially in a multimodal public transport trip, out-of-vehicle time may be very close to the in-vehicle time, for instance, Meng et al. (2018) presented that out-of-vehicle time accounts for 40.3% in average actual travel time (Meng et al., 2018). In- and out-of-vehicle time stages are access/egress, waiting, in-vehicle and walking times presented in Figure 2.4. Out-of-vehicle time components capture the inconvenience of public transportation not being instantaneously accessible and available, whereas in-vehicle time components capture the inconvenience of crowding (and having to stand) and transfer time components capture the inconvenience of having to transfer, walk, and wait (Wardman, 2014).

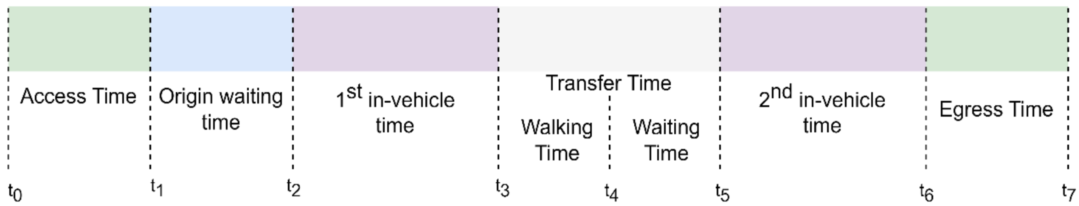


Figure 2.4. Travel time components in a journey (Brands et al., 2022).

Access/egress time is the time spent, most likely by walking, to reach their first origin of their journey, such as workplace or home (Brands et al., 2022), i.e., the time passed between t_0 and t_1 in Figure 2.4. Access and egress times are mainly related to the catchment area of transit mode and the performance of the transit service highly correlated with them (Rahman et al., 2022). According to Van der Waard (1988), access time is affected by the anxiety of attempting not to miss the vehicle, therefore, it is perceived as longer than egress time (Van der Waard, 1988).

Waiting time is the total time spent at stops/stations while waiting for the transit, i.e., the sum of times spent between $t_1 - t_2$ and $t_4 - t_5$ in Figure 2.4. Passengers tend to reduce the time by planning their arrival and departure times. Thus, *origin waiting time* $(t_1 - t_2)$ may be lower than *transfer waiting time* $(t_4 - t_5)$ due to reliability and integration problems between transit modes. In addition, some studies in literature used *hidden waiting time* to capture the potential waiting time due to service quality problems along with platform waiting time (Furth & Muller, 2006; Nielsen, 2000; Nielsen et al., 2021). Hidden waiting time may be defined as half of the headway of the least frequent transit services used in a journey (M. K. A. Anderson, 2013). On the other hand, this issue was also put in the literature by the terms of *headway* and *displacement time*, which were used to represent not being able to travel at the desired time (Wardman, 2014).

In-vehicle time includes the time passed in the transit vehicle. In other words, total of times between $t_2 - t_3$ and $t_5 - t_6$ in Figure 2.4. It can be approximated by using the length of the trip and the commercial speed of the transit vehicle (km traveled per unit of time, including the time spent on stops/stations) (Amirgholy et al., 2017). Here, *crowding level* or *penalty for not seating* components can be employed to distinguish the difference in perception of in-vehicle time, for example between a passenger who travels seating and standing.

Transfer time includes the *walking time* and *waiting time* components at each transfer point (Brands et al., 2022). Besides, the phenomenon of transferring in public transportation may include *transfer penalty* along with walking time and waiting time (Jara-Diaz et al., 2022). Transfer penalty refers to a purely psychological aspect of the transfer process that is influenced by the environment in which the transfer occurs (Guo & Wilson, 2011).

2.5.2. Time Multipliers for Valuing Travel Convenience

Time multipliers obtained from the relative valuations of components based on the people's travel choices collected using stated preference or revealed preference techniques (Wardman, 2014). SP approaches in public transportation basically mimic the travel conditions and put into respondents evaluations by offering multiple choices that require them to trade-off one relevant variable against another (Wardman, 2014). On the other hand, RP method depend on the actualized choices where travelers assumed to be familiar with all the transit alternatives (Wardman, 2014). To gain insights about the relative valuations of travel time components, mode and route choice models may exploit. Generally, these studies assess travel experience of passengers in terms of travel time components: access, egress, waiting, in-vehicle and transfer times and number of transfers (M. K. Anderson et al., 2017; Rui, 2016). Those results can be used in the policy evaluation and predicting the impact of a change in the service on the route choice (Nielsen et al., 2021). Keeping in mind that these valuations, i.e., time multipliers, may vary across countries depending on several factors: cultural differences, different standards and expectations, operating practices, sampled public, socioeconomic composition, travel conditions (Wardman, 2014) and weather (Nielsen et al., 2021). Nonetheless, they may be considered inherently more transferrable than monetary based valuations (Wardman, 2014).

In the following sections, time multipliers (relative to in-vehicle time) under several headings will be presented mainly based on a comprehensive review by Wardman (2014). In addition, we reviewed the recent public transport route choice and perceived travel time studies to present the current state of work by presenting rates of substitutions alongside the time multipliers.

2.5.3. Walk and Wait Time Multipliers

These two attributes form the out-of-vehicle experience of a passenger, where passenger may be affected by the inconvenience, effort and frustration, and common practice was to apply weight of two (2.0) to walk and wait time (Wardman, 2014). In Table 2.5, time multipliers for out-of-vehicle components are presented based on a study by Wardman (2014) where large scale reviews on value of time bus with insights into time multipliers were collected.

Here, one distinction should be made that access/egress walk and first waiting time components may be evaluated separately. This is important because access time multipliers may be lower than walk time because access may involve modes that require less effort than walking, and first waiting may be avoided by planning arrival times, thus can be lower than the transfer waiting multipliers (Wardman, 2014). Also, one other reason for the low access time multipliers, especially for rail, is that some studies in literature worked with public transportation networks, named as hub-and-spoke, where the alternative routes are very limited (Bovy & Hoogendoorn-Lanser, 2005; Raveau et al., 2011; Vrtic & Axhausen, 2002) and mainly focused on metro and train corridors where bus transit was considered as access and egress modes (M. K. Anderson et al., 2017). Thus, modes of access may have a lesser disutility than walking (Wardman, 2014).

Table 2.5. Time multipliers for out-of-vehicle time components. (Wardman, 2014)

Study	**Wardman et al. (2013)**						**Abrantes and Wardman (2011)**	
Attribute	*All*	*Bus*	*Rail*	*UK*	*Non-UK*	*RP*	*UK*	*RP*
Walk Time	1.68	1.64	1.65	1.62	1.93	2.01	1.65	1.84
Access Time	1.68	1.62	1.29	1.57	1.93	1.88	-	-
Wait Time	1.8	1.74	1.49	1.68	1.93	2.22	1.7	2.32
Transfer Wait	1.84	1.92	1.83	1.72	1.93	2.03	-	-

Note: In the table RP is the abbreviation for revealed preference survey. This distinction is made because time multipliers found using RP surveys found to be higher than the SP values.

Further, time multipliers obtained from revealed preference surveys are larger and consistent with the common practice of assuming two (2) as time multiplier for out-of-vehicle components (Wardman, 2014). Also, there might be some external factors, e.g., crowding at stop/station etc., may have influence on walk and wait time multipliers. Hence, crowding factor may be implemented to gain precision to these multipliers (Wardman, 2014).

2.5.4. Transfer Penalty

Transfer walking and waiting times are covered in the previous section. In this section, the fixed cost, which is named transfer penalty, used to reflect the deterrence of transferring is discussed. The transfer penalty is significant due to the inconvenience and risks involved, independent of the time spent walking and waiting, and there is a chance that the next service will be missed, and any ongoing activities will be interrupted (Wardman, 2004).

One issue here is whether the transfer penalty is obtained alongside the transfer time components or solely obtained without considering any separate time component. Wardman (2001) reviewed studies in this aspect and distinguished studies which presented a pure transfer penalty independent of time effects (Wardman, 2001). When the time effects are not excluded, transfer penalty averaged around 30 minutes, whereas it was averaged around 17.6 minutes in case of considering pure transfer penalties

(Wardman, 2001). Furthermore, Vrtic and Axhausen (2002) stated that when the walking and waiting time at transfer locations were omitted, the transfer penalty ratio increased from 5.7 min/transfer to 13.1 min/transfer (Vrtic & Axhausen, 2002). Other than time effects, Nielsen et al. (2021) stated several other contributing factors, e.g., ease of wayfinding, shopping availability, having escalators and level changes, on the substitution rate are also presented. Their results for transfer penalties were ranging from 5.4 to 12.1 minutes of bus in-vehicle time for the best and worst transfer attribute scenarios (Nielsen et al., 2021). Guo et al (2011) studied the variability of transfer penalty based on transfer stations of London Underground (LUL) system and the results showed that the minimum (for the best transfer station) and maximum (for the worst transfer station) penalties were 0.5 minutes and 9 minutes in IVT units, respectively (Guo & Wilson, 2011).

On the other hand, Anderson et al. (2017) presented the penalties for transfer considering modes as 18.8, 14.2 and 7.2 min of bus in-vehicle times for bus-to- bus, bus-to-train, and train-to-train, respectively. This finding indicates that transferring to rail transit from bus is felt less burdensome by the transit passengers. This may be due to rail transit is more reliable and frequent. In addition, transfer penalties for longer trips were found to be higher than for shorter trips because of the tendency of passengers to remain on board and prefer in-vehicle activities such as reading and sleeping (Anderson et al., 2017). Pursula & Weurlander (1999) found that transfer penalty equal to 10 min of door-to-door travel time. In addition, passengers were found to prefer travelling 15 minutes longer to get a seat for the trip (Pursula & Weurlander, 1999). Besides, relatively low transfer penalty of 5 minutes was also found in literature (M. Yap et al., 2020).

Further, because of the very high frequency of the metro, metro in-vehicle time substitution rate was found to be lower than both bus and train. The variation in transfer penalty ratios may be explained in several ways, one reason for finding smaller ratios may be the successful coordination of arrival and departure times of PT modes in those locations, hence, having to transfer from one PT mode to another is perceived less deterrent (Jánošíkova et al., 2014). The transfer penalties form the literature can be seen in Table 2.6. These values are mainly gathers from the study (Wardman, 2014). Some of the more recent transfer penalty values are added to the table, others are mentioned above.

Table 2.6. In-vehicle time equivalencies for transfer penalties in literature.

Study	Condition	Transfer Penalties					
Wallis et al. (2013)		*Peak*	*Off-peak*				
	Same mode	4	12.5				
	Between modes	9	17				
Wardman (2001)		*5 km*	*25 km*	*100 km*	*250 km*		
	Commute	6.9	5.8	5.0	4.5		
	Other	13.7	11.5	9.9	9		
(Douglas and Jones, 2013		*Bus Short*	*Bus Med.*	*Bus all*	*Rail Short*	*Rail Med.*	*Rail All*
	Bus to Bus	14.8	14.6	14.5	18.1	28.6	23.3
	Bus to Rail	11.1	16.6	15.1	15.8	19.3	17.5
Wardman and Shires, 2000		*1st Transfer*	*2nd Transfer*				
	30 min	9.1	5.8				
	60 min	11.1	7				
	120 min	15.1	9.4				
Raveau et al., 2013		*London*	*Santiago*				
	AM Peak Com.	5.24-7.70	5.84-10.30				
	Off Peak Non-Com.	10.65-15.66	9.05-15.95				
Anderson et al. (2017)		*Work*	*Leisure*				
	Bus to Bus	18.8	26.2				
	Bus to Train	14.2	21.8				
	Train to Bus	17.3	26.5				
	Train to Train	7.2	13.4				
Nielsen et al. (2021)		*Work*	*Leisure*				
	With Tr. Attr.	7.92	9.13				
	w/o Transfer Attr.	9.44	12.76				

Note: The values should be seen by considering both conditions. For example, transfer values gather from Nielsen et al. (2021) are presented for work and leisure trips in two ways: considering transfer attributes (With Tr. Attr.) and not considering transfer attributes (w/o Transfer Attr.).

2.5.5. Crowding Convenience

Crowding multipliers are applied to the in-vehicle times and their magnitudes vary according to crowding level. Crowding level may be indicated in terms of passenger per meter square (pass./m²) or the load factor in percentages (see Table 2.7). The load factor is the ratio of the number of passengers to the seating capacity. Hence, the increase in the loading factor results in a reduction in service quality, while simultaneously enhancing cost efficiency.

The inconvenience of standing is readily apparent and can be anticipated to significantly increase the value of travel time, thus used in transit route choice studies

(Nielsen et al., 2000; Raveau et al., 2014) overcrowding will also impact those who are seated to the extent that they will experience an increase in inconvenience and discomfort, leading to a rise in time values (Wardman, 2014). In addition, the burden of travelling on passengers in crowd conditions depends on its severity It is important to quantify the crowding inconvenience to reveal its influence on passengers' travel choices (Shao et al., 2022; M. Yap et al., 2020). For example, Shao et al., (2022) studied the bus crowding influence on bus in-vehicle times and concluded that increase in the crowding degree leads to longer perceived travel time. They classified crowding level using standing density lower limits of 4.2 pass./m² and 7.5 pass./m² for crowded and very crowded conditions. The crowding level time multipliers ranged between 1.11 (4 pass./m²) and 2.51 (10 pass./m²).

In Table 2.7, the crowding time multipliers found in relevant literature are presented. The time multipliers are presented for seating and standing passengers. The negative effect of the in-vehicle crowding on the seating passengers is also present and the multipliers are not very low. For instance, when a passenger with a seat on a bus with 35 seating capacity travels with 70 passengers, the perceived travel time is 31 minutes for a 20-minute trip.

Table 2.7. Crowding multipliers from literature (Wardman et al., 2011; Kroes et al., 2014)

	Seat					**Stand**				
	Wardman et al. (2011)		**Kroes et al. (2014)**			**Wardman et al. (2011)**		**Kroes et al. (2014)**		
Load Factor	*Com.*	*Leis.*	*All*	*Metro*	*Bus + Tram*	*Com.*	*Leis.*	*All*	*Metro*	*Bus + Tram*
50%	0.86	1.04	1.0	1.0	1.0	-	-	-	-	-
75%	0.95	1.14	1.0	1.0	1.0	-	-	-	-	-
100%	1.05	1.26	1.083	1.077	1.102	1.62	1.94	-	-	-
125%	1.16	1.39	1.165	1.155	1.204	1.79	2.15	1.289	1.27	1.342
150%	1.27	1.53	1.248	1.232	1.307	1.99	2.39	1.394	1.362	1.467
175%	1.4	1.69	1.248	1.232	1.307	2.2	2.64	1.394	1.453	1.593
200%	1.55	1.86	1.330	1.309	1.409	2.44	2.93	1.499	1.453	1.593
250%	-	-	1.413	1.386	1.511	-	-	1.604	1.545	1.718

Note: "Com." and "Leis." are the abbreviations for commuter and leisure trips.

2.5.6. Summary

It can be foreseen that, over time with changes in expectations, multipliers may change. Also, multipliers might be tempered by conditions, so for example, commuters in Tokyo might have somewhat greater tolerance to crowding than commuters in many other metro systems (Wardman, 2014). The recommended indicative multipliers by Wardman (2014) are presented in Table 2.8.

Table 2.8. Recommended time multipleries for time attirubutes by Wardman (2014).

Convenience Term	Indicative Multiplier
Walking with more than normal effort	4.0
Waiting in crowded conditions	2.5 - 4.0
Walking in crowded conditions	2.0 - 3.5
Walking and waiting in normal conditions	1.75 - 2.0
Standing	1.50 - 2.0
Headway	0.5 - 0.8
Displacement time	0.4 - 0.6
Transfer penalty	5 - 15 minutes

2.6. Travel Time Perceptions

Public transportation users perceive time differently depending on their preferences and quality of service provided by transit mode (Nielsen et al., 2021). The satisfaction and preferences of transit passengers are affected by their perception of travel time thus it is essential to consider in demand forecasting, modeling, planning and operating public transportation systems (Currie, 2005; Meng et al., 2018). Using objective performance data in planning models is insufficient to reflect the subjective images of passengers (O'Farrell & Markham, 1974). The difference in perception affects the route choice behavior of passengers, thus creating a necessity for considering the perceived travel time in route and mode choice analysis. Perceived travel time can be defined as the duration felt by passengers while traveling (Meng et al., 2018) and include the sum of waiting, in-vehicle and transfer (walking etc.) time across all trips (Jenelius, 2018).

There are several factors influence the passengers' travel time perception, such as travel cost, level of comfort, and reliability of the service (Dewulf et al., 2012; Hess et al., 2004; Meng et al., 2018; Padhye et al., 2020; Psarros et al., 2011). Brands et al. (2022) concluded that, on average, transit users perceive their travel times 11% longer than the actual travel time. In addition, Meng et al. (2018) found that socioeconomic characteristics may contribute to a 5% increase in the perceived travel time. Furthermore, they found that passengers perceive travel times at each stage greater than the actual and women tend to overestimate their walking, waiting and in-vehicle time than men (Meng et al., 2018). Besides, transit users generally prefer the fastest way of travel from their origin to their destination. Because of that, in general, transit passengers have more positive attitude toward rail transit modes than buses because rail transit modes are usually more regular, reliable and comfortable (M. K. Anderson et al., 2017; Nielsen, 2000). However, attitudes can change depending on the circumstances; for instance, passengers generally do not prefer metro transit for short trips; thus, the perception of passengers with short travel times on metro transit may be more negative than that of passengers on buses (Brands et al., 2022).

O'Farrell and Markham (1974) studied on the perception of car-owning commuters on the to or from the work journey variables, namely waiting time, in-vehicle time and cost. They found that regular train users overestimate their in-vehicle train time by 7.7%, on the other hand car users underestimate by 5.1%. On the other hand, they found that bus in-vehicle times were 26% and 32% for morning and evening trips, respectively. Importantly, bus users stated that they perceived bus waiting times 76% and 205% higher than the actual waiting times in the morning and evening, respectively (O'Farrell & Markham, 1974). Brands et al. (2022) compared the actual travel times from AVL data with the perceived travel times reported in a survey. Survey results showed that around 40% of the respondents reported their travel time higher than the actual. On average, in-vehicle travel times are perceived 11% higher than the actual travel times. Their results indicated that, short trips are likely to denigrate more than the long trips (Brands et al., 2022). Meng et al. (2018) stated that passengers perceive more while waiting than walking. In addition, transfer walking time has been established to be more onerous than the access and egress walking times. Also, it was stated that the time spent on the preceding stage has little impact on the perception. They concluded that, on average, passengers perceive 6 to 10%, 11 to 12%, and 3 to 5% higher than the actual walking, waiting and in-vehicle times, respectively (Meng et al., 2018). Dewulf et al. (2012) studied walking time perceptions of people with different destination locations. It can be interpreted from their results that walking time perceptions mostly (86.1%)

reported equal or higher than the actual times and the overestimation was 23% (Dewulf et al., 2012). Watkins et al. (2011) studied the effect of real time information systems on bus waiting times. Their results showed that passengers using traditional information perceive waiting times approximately 16% longer than the actual. Further, real time information systems decrease the perceived travel time by 31%. Interestingly, they also found that real time information systems users wait almost 2 min less than those using traditional information (Watkins et al., 2011).

2.7. The Impact of the Network Change

To determine the impact of any network change, such as opening a new transit line, before-and-after (ex-post) analysis is used. However, the opening of a new transit line is typically accompanied by changes to the rest of the transportation network, land use, and transit systems (Brands et al., 2020), such as the adjustment (or termination) of bus routes collinear to the new metro line (Gao et al., 2022b). Most of the impact studies focus on a new rail service that replaces an existing bus service, as opposed to a case in which the two modes compete. Thus, the observed situation does not involve a choice between rail and bus, but rather a forced conversion and attraction to a new service that provides a higher level of service (Ben-Akiva & Morikawa, 2002). Hence, observed impacts, such as total ridership, travel time, reliability, or modal split, cannot be directly attributed to the new transit line (Brands et al., 2020).

For example, extending an existing metro line requires replanning existing bus routes to minimize the competition between modes. This is mostly the case in ex-post analysis. Thus, making a before and after comparison between bus and metro systems does not represent the true mode choice behavior. This fact constraints to make a meaningful ridership attraction comparison between bus and metro (Ben-Akiva & Morikawa, 2002). External factors and corresponding service changes in the network need to be isolated to make a meaningful comparison between before and after conditions.

Brands et al. (2020), used two longitudinal smart card data sets, which include 5-6 weeks of multimodal tap-in and tap-out transactions before and after opening periods of the new metro line in Amsterdam, and corresponding AVL data are employed for the ex-post analysis. For travelers to get used to the network change, the later data set starts 7 weeks after the new metro line is opened. To be able to make before and after journeys comparable, they clustered the stops using the agglomerative hierarchical clustering (AHC) method and performed their analysis at the cluster level instead of the stop level. Although some clusters near the new metro line showed a decrease in ridership, total ridership increased by 4% indicating that passengers changed their route choice, including the origin and destination of their journeys or access/egress modes. Regarding travel time, their results do not show a significant change in average. This is interesting because the metro tends to travel in shorter running times compared to bus and tram systems. Thus, a decrease in travel time is likely to be expected, however, it seems to be that arranging bus and tram service to complement the new metro line and forcing travelers to change their flow pattern did not enhance their travel time experience. Clusters along the new metro line showed the highest gain in travel times, contrary to that, the outskirts of the city experienced a negative impact. Due to the induced demand (newly attracted passengers), their results showed a slight decrease in the number of transfers. But there was a slight increase in the number of transfers when only travelers in the before case were included.

Gadepalli et al. (2022) studied the impact of a newly introduced metro rail on existing urban bus routes in terms of overall ridership change and time savings experienced by transport users. As was stated in the study, the bus service network was not adjusted according to a new metro line. That is generally never the case because opening a new metro line, and adjusting the transport network, e.g., bus routes, are planned and performed simultaneously (Brands, Dixit, and van Oort 2020). Therefore, this type of a study is important for understanding trip characteristics and user preferences on collinear public transport routes, for example which type of trips are performed, e.g., long, or short distance, or for what purpose the transport system is used, e.g., as feeder or main mode. The results of the study showed that travelers prefer the metro for long distance (more than 5 km) trips when they have bus and metro options. On the other hand, 66% of the bus trips within the metro influence area were shorter than 5 km, whereas the overall share of short bus trips was 42-52%. Trip level analysis results showed that 25% of the passengers of only 27 of 1,474* collinear bus routes travelled parallel to metro.

The findings indicate that public transport users prefer the bus for short-distance trips, and the metro for long-distance trips. In addition, it seems that some fixed bus demand is present even if there is a metro alternative. This may be due to the fact that bus stops are generally more accessible than metro stations; hence, elderly people tend to prefer the bus over the metro. Therefore, this need should not be overlooked when adjusting the bus system.

CHAPTER 3

DATA AND DATA PROCESSING

3.1. Data Collection

İzmir municipality provides General Transit Feed Specification (GTFS) data for the all the public transportation modes available which are bus, metro, commuter rail, tram, and ferry. Most of the required information can be obtained from these GTFS data, such as operative bus routes, bus stop identification numbers and locations, names and locations of metro, commuter rail, tram, and ferry stations. However, stops are not assigned to bus routes in the bus GTFS data. To create lists that show the stops that are on specific bus routes, for example the list of stops that are on the going direction of bus route 10, are provided through Application Programming Interface (API) by İzmir Municipality. Python Spider IDE interface is used to gather all the updated information related to public transportation modes, such as operative bus lines, bus stop identification numbers and locations, rail (metro, commuter rail and tram) station names and locations, and finally ferry station names and locations.

3.1.1. Bus Stop and Rail Station Data

The geographical coordinates of the stops and stations are obtained from the open source GTFS data (İzBB, 2022). This data also includes the name and ID of the stops and station. The information of which stop is on the route of which bus is gathered from the open data source using API service (ESHOT, 2021b). To create the paths of the bus routes that are followed on the road network, the GPS data collected for the bus routes is retrieved from (ESHOT, 2021a). This route data is then created on the geographical information system environment using ArcGIS ArcMap 10.3.

3.1.2. Smart Card Data

ESHOT General Directorate provided one weekday (05.06.2022) smart card data which includes 1,969,185 both tap-in and tap-out transactions. In the smart card data, 201,511 (10.23%) of the total transactions are tap-out (alighting) transactions on commuter rail (İzban) and thirteen bus routes which abide “pay as you go” policy. The remaining is tap-in (boarding) records account 1,767,674 (89.77%) transactions.

Since the boarding transactions are only required, alighting transactions are separated from the smart card data. This “alighting data” is further used to validate the estimations of developed trip chaining algorithm in section 5.1.3. More importantly, since the data may include some errors, data cleaning and error correction processes are applied to the remaining “boarding data”. These processes are explained in the following sections.

3.2. Data Cleaning

Data cleaning (or cleansing) process is often required in studies working with big data. As explained by Hussain et al. (2021), data cleaning process often requires prior knowledge about the sources (human or equipment) of the errors to be able to refine or correct these errors. Human or equipment failures may lead to several problems in smart card data. For example, in entry-exit systems, there may be some alighting time records earlier than or equal to boarding times. In addition, there may be some record with missing information on boarding time/location or card ID, or duplication of tap-in or tap-out events (Hussain et al., 2021).

Before feeding the smart card data to the trip chain algorithm, the data cleaning process is required to eliminate errors. Some errors occur during the fare collection, such as *the one-stop boarding problem* (section 3.2.2.2). Some occur due to the public transit service changes, such as *removed bus routes* (section 3.2.1). The transactions on these bus routes are extracted from the data set. There might also be some issues due to *changed or removed bus stops* (section 3.2.2.1). This is due to the difference between the dates of smart card data and the analysis. In addition to the errors, some passenger behavior causes the trip chain algorithm to fail, such as *group boarding problem (section 3.2.2.3)*; these are the boarding transactions done by the same cardholder at the same stop or station with a 5-10 second time gap. An algorithm, which completes the cleaning process with an acceptable running time, should be created to eliminate these problems. All of the errors discovered in the smart card data, as well as the strategies used to eliminate or fix these errors in order to improve the effectiveness of the trip chaining method are explained in the following sections.

It is necessary to state that, before constructing the data cleaning algorithm, prior information is gathered from transit authorities for the possible errors (systematic or personal) in smart card data. In addition, preliminary analyses are also performed to gain experience with the data. We handle the errors in the SC data in two ways, elimination, and correction. The summary of the data cleaning and other manipulation processes are presented in

3.2.1. Error Elimination

Error elimination is applied to 1,767,674 (89.77% of the total) bus, metro, commuter rail, tram, and ferry transactions. Prior to data correction phase, errors that cannot be corrected are eliminated. Such as, 154 (0.01%) transactions with missing card ID information, 1,154 (0.065%) transactions assigned to one card ID, 736 (0.042%) transactions with no boarding time/location information are deleted.

In addition to this, since there is inconsistency between the dates of the two data sets: the smart card data is from 05.06.2022 and the bus stop and route data is from 09.02.2023. Thus, the smart card data is checked for whether there are any bus routes newly added or shortly removed from operation. 6 (six) bus routes are determined to be removed and corresponding 976 (0.055%) boarding transactions are deleted. As a result, 3,020 (0.17%) transactions are deleted in error elimination and 1,764,654 (89.60%) transactions are remained.

3.2.2. Error Correction

As it was mentioned earlier, in İzmir, commuter rail system (İzban) and 13 bus routes connecting remote districts to CBD require tapping out to get a refund from the full payment charge. This means that smart card data includes alighting information of the passengers who used commuter rail and the mentioned bus routes. Error correction steps are only applied to 1,764,654 boarding transactions. Before feeding the boarding data to the trip-chain algorithm for estimating the alighting locations, 3 (three) steps of error correction process are performed. Error correction steps are summarized below.

First, 17,623 *boardings at stops that were not on the bus route* are corrected by replacing the bus stop with the closest bus stop on the route. Second, 29,632 *boardings at the last stops* are determined and corrected by reversing the trip direction. After reversing the direction of some boarding transactions in the 3rd step, the 2nd step needs to be reapplied, so, 3,543 *boardings at stops that were not on the bus route* are corrected in a similar way. The cleaned and corrected 1,091,356 bus boarding transactions are merged with the rest of the data (673,298 boarding transactions on the other PT modes), resulting in a complete boarding data set of 1,764,654 transactions. Third, *group boardings* are handled by assigning a unique card ID to each transaction in the group.

3.2.2.1. Changed or Non-operative Bus Stops

This problem is only observed on the bus boarding transactions. There are transactions that the boarding stop is not on the given bus route and direction due to the nearly 8 months of time gap between the date of smart card and public transit stop data. During this period, there might be changes applied to the route or the location of stops. In addition, some stops might be taken out of order due to operational changes. This issue is handled by changing the problem stop with the closest stop on the given bus route and direction.

3.2.2.2. One-stop Boarding Problem

This is a major problem that both decreases the accuracy and matching rate of our trip chaining algorithm. Izmir's Automatic Fare Collection (AFC) system requires drivers to change direction at the end of each bus trip. The system denotes the outbound direction as 1 and inbound direction as 2. Bus drivers need to set the direction before starting a new trip to enable the system to correctly record the boarding stops. If the driver fails to change direction before starting a new trip, AFC system assigns all the boardings of the new trip to the last stop of the previous trip. This issue underlies the first type of one-stop boarding problem in three scenarios represented in the Figure 3.1. The 1st scenario may occur because bus drivers often let passengers board at the terminus before setting the direction for the next trip. It is assumed that all the one-stop boarding problems where the boardings assign the last stop are in the majority as in the 1st scenario, hence, correction algorithm is built accordingly. However, as the figure indicates, our correction is not capable of correcting the problem if the problem is realized as in the 2nd and 3rd scenarios.

In the preliminary analysis, it is observed that 29,632 boardings (1.68% of total boardings) are on the last stops of the bus routes. In our algorithm, the following steps are applied to correct the issue represented in scenario one. First, the last stop boardings are identified by matching the boarding stops (in the SC data) with the last stops of the bus routes (obtained from API) by two directions. Second, the direction of the identified last boardings is reversed. For some of bus routes, the same stop is used as the last stop in the inbound direction and the first stop in the outbound direction, or vice versa. Hence stop IDs are the same. Thus, changing only the direction of last stop boardings is sufficient to correct the issue. However, for most bus routes, these stop IDs are different. This problem is solved by applying the previous data cleaning step in section 3.2.2.1.

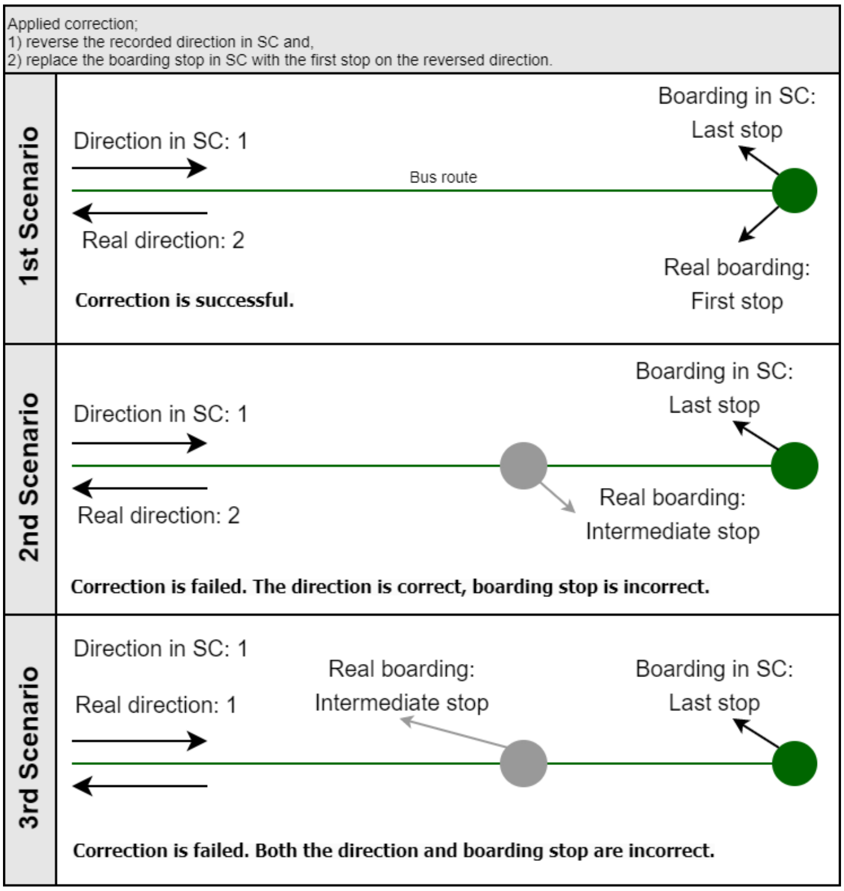


Figure 3.1. Possible scenarios of the issue where all the boardings are assigned to last stops.

On the other hand, independent of the failure of changing the run's direction, there are bus runs that all the boarding transactions are assigned to one stop. Even though these problems may be considered as rare when the size of the whole data is considered, they negatively impact the estimation of the alighting locations of complete journeys.

3.2.2.3. Group Boarding

Repetitive transactions are considered as duplicate records, and discarded from the data (Hora et al., 2017). In İzmir, it is common to see several boarding transactions with small time gaps, e.g., 5-10 seconds, at the same stop/station from the same cardholder, as shown in Table 3.1. The situation may have happened when someone offered to pay (tap) for another, or when a group of people travelled using a single card. In this study, this situation is called group boarding. Since the trip chain algorithm looks for a relation between successive trips, it fails to estimate when two successive boarding transactions occur at the same stop/station, as in the group boarding.

Diker et al. (2016) solved group boarding problem by keeping the first boarding transaction in the passenger's trip chain and treating the other boarding transactions as separate single trips (Diker et al., 2016). In this study, as it is illustrated in the figure, we assigned a new card ID for each successive boarding transaction within a trip.

Table 3.1. An example of a group boarding and assigned new card IDs.

#	Card ID	Boarding Time	Stop ID	New Card ID
1	KART 10111	08:15:25	50318	**KART 10111**
2	KART 10111	08:15:33	50318	**KART 10111 - 1**
3	KART 10111	08:15:41	50318	**KART 10111 - 2**
4	KART 10111	08:15:46	50318	**KART 10111 - 3**

Correction for group boarding problem was applied for 1,764,654 boarding transactions from 633,083 unique card IDs. With the correction, the number of unique card IDs increased by 4,936, i.e., correction created an additional 4,936 passengers. This should be kept in mind that there were one way group boarding trips, i.e., two passengers used the same card for a one-way trip. So, assigning a unique card ID for the two passengers resulted in two single boardings. If the group boarding correction is not applied here, the record in the given example is to be treated as multiple boarding and fed to the trip chain algorithm, consequently, resulting in a failed estimation. This way, the number of single trips may have increased, however, matching rate of the algorithm may also have increased.

Table 3.2. Summary of data cleaning and data manipulation processes.

Phase	Process	# of Transactions	% of Total	% of Boarding Trans.	% in Multiple Boardings	Action
-	**Start**	**1,969,185**	100%	-	-	-
Separation	Alighting Trans.	201,511	10.23%	-	-	Seperation
	Boarding Trans.	**1,767,674**	89.77%	100%	-	Seperation
Preliminary Error Elimination	Missing card ID	154	0.01%	0.01%	-	Elimination
	Over-boarding	1,154	0.06%	0.07%	-	Elimination
	Missing info.	736	0.04%	0.04%	-	Elimination
	Non-operative bus routes	976	0.05%	0.06%	-	Elimination
	Total	3020	0.15%	0.17%	-	
After Elimination	-	**1,764,654**	89.61%	99.83%	-	
Data Cleaning	Boardings at last stops	29,632	1.50%	1.68%	-	Correction
	Boardings at stops not on the bus route	21,166	1.07%	1.20%	-	Correction
	Boardings Remaining from Yesterday	3,240	0.16%	0.18%	-	Correction
	Single trips	**145,690**	7.40%	8.26%	-	Elimination
	Multi boardings	**1,618,964**	82.21%	91.74%	100%	Trip chain input
Trip chaining result (1,250 m walking threshold)	Estimated	1,378,120	69.98%	78.10%	**85.12%**	Trip chain output
	Not estimated	240,844	12.23%	13.65%	**14.88%**	Trip chain output

CHAPTER 4

METHODOLOGY

The data set contains 1,764,654 transactions as a result of the data cleaning processes in section 3. The methodology for this thesis is based on several elements, all of which include data processing approaches that serve to estimate the potential passenger exchange from bus to the metro service after the completion of the metro extension. We used passenger load profiles to illustrate how the ridership on the bus routes, which are currently operating within the metro extension's service area, changed after introducing the metro alternative. This will assist us in understanding the impact of the new metro service and making informed future transportation planning decisions.

First, the assumptions, which are necessary in trip chaining method, are established in section 4.1.1. Then, based on these assumptions, trip chaining algorithm is constructed to estimate the alighting locations in section 4.1.2. Second, a series of algorithms are developed in section 4.3 to create the passenger load profile of bus routes as realistically as possible. Since the primary objective is to simulate the passenger shifts from bus routes to the metro, it is necessary to identify the bus routes that presently operate within the service area of the metro extension. Thus, in section 4.2.1, we aim to determine the bus routes that operate within the service area of the metro extension. Then, in section 0, the competitive or cooperative relationship of these bus routes with the metro is established. Thereafter, the passenger flows on the targeted bus routes are categorized in section 4.4. In this section, for each passenger flow category, we also develop an alternative scenario where the passengers are forced to use the metro, and bus, if necessary, to repeat their actualized trips. In the final part, section 4.5, two travel time cost functions are developed: one for the actualized trip and one for the trip(s) in the alternative scenario. In this section, we explicitly explain our methodology for constructing the travel time cost functions and the underlying assumptions used to configure alternative scenarios. Then, we estimate the mode shifts based on travel time savings.

4.1. Trip Chaining Method

The trip chaining method is used in many studies (see Table 2.1) mostly to create origin destination matrices to further use in transportation planning. In this study, trip chaining method is employed to estimate the alighting location of the trips conducted on public transportation modes. The trip chaining method is relatively easier (compared to probabilistic and machine learning methods) to construct and has a reasonable performance in estimating alighting locations. However, the method has two deficiencies: first, by its nature, it is not capable of estimating alighting locations of single trips (singular boarding transactions); second, alternative modes other than walking are not considered, such riding bicycles or e-scooters.

Even though the algorithm's performance appears to be above average compared to those presented in the literature, we acknowledge that our algorithm lacks the most recent developments presented to enhance the trip chaining method's capabilities. In the following section, the adopted assumptions, which form the foundation of the trip chaining algorithm, are presented.

4.1.1. Assumptions in Trip Chaining Algorithm

The trip chaining algorithm is essentially a set of assumptions and principles that are defined based on real-world conditions in order to estimate a passenger's alighting location. The assumptions and rules that are adopted in the developed trip chaining algorithm are listed below.

1. The majority of passengers alight at the station where they start their next trip (Barry et al., 2002).
2. The majority of passengers return to their origin (boarding station of their very first trip) at the end of their last trip (Barry et al., 2002).
3. The alighting stop of a trip cannot be inferred, if it is the only trip of the cardholder for the day (Barry et al., 2002). This type of trip is termed as a unit (Trépanier & Chapleau, 2006) or a single trip as in this study.
4. The alighting stop cannot be inferred if consecutive transactions (boardings) occur at the same station (Barry et al., 2002).

5. Passengers can only alight at a subsequent stop on the given direction of the route (Trépanier & Chapleau, 2006). This rule is only applicable to bus boarding transactions since the boarding transactions on other public transportation modes do not have trip direction information.
6. The distance between the boarding stop of the next trip and the alighting stop of the previous trip must be less than the *allowable walking distance* (Trépanier & Chapleau, 2006).
7. The alighting stop of the last trip of the day must be within the allowable walking distance of the boarding stop of the first trip of the day (Trépanier & Chapleau, 2006).
8. Passengers do not use any other transportation modes, i.e., shared mobility, private car etc., but walk between any consecutive trip segments (Zhao et al., 2007).
9. A passenger is required a certain time to perform an activity, *activity threshold* (Nassir et al., 2011).
10. A passenger must transfer in a given time, *transfer threshold* (Nassir et al., 2011).
11. The hour with the lowest activity in the day can be used as the *virtual midnight* (M. Munizaga et al., 2014).

4.1.1.1. Single Trips

As in stated in the 3rd assumption, the alighting location cannot be inferred if it is the only trip (single trip) of the card holder (Barry et al., 2002). Due to the inability of the trip chaining method to determine the destinations of single trips, these trips should be excluded from the analysis.

As a result of the data cleaning process, a data set of 1,764,654 transactions from 646,757 cardholders is obtained. The first graph on Figure 4.1 indicates that there are 145,690 (8.26%) single trips that cannot be used in the trip chaining algorithm. These transactions are excluded from the analysis. Besides, the second graph on Figure 4.1 shows the number of occurrences of card IDs in the complete data set. The percentage of cardholders who have only one trip record on the 5th of April 2022 is 22.53%.

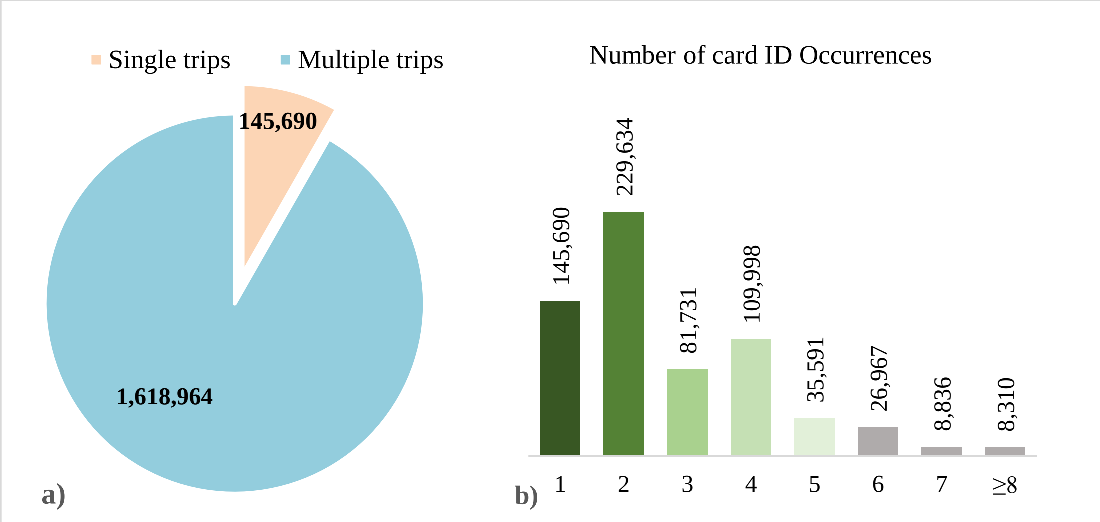


Figure 4.1. a) Number of single b) Number of card ID occurrences.

Table 4.1 shows the data set, in which each cardholder has two or more trip records. This data set contains 1,618,964 (91.74%) transactions of 500,876 unique card IDs.

Table 4.1. Number of occurrences of the unique card IDs in the smart card data.

Occurrence	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**>8**	**Total**
Counts	229,634	81,731	109,998	35,591	26,967	8,836	4,415	3,895	500,876
Percentages	*45.85%*	*16.32%*	*21.96%*	*7.11%*	*5.38%*	*1.76%*	*0.88%*	*0.78%*	*100%*

Note: The occurrence value can be viewed as a representation of the number of trips made by a cardholder. The mean occurrence (trip) for the smart card data is 3.94. 229,634 cardholders use the public transportation twice.

4.1.1.2. Same Boarding and Alighting Location

For several reasons, the trip chaining algorithm estimates the alighting location of some trips at the same location where the boarding happened. In regard to 4^{th} trip chaining assumption, these issues are label as failure. Due to this issue, estimations for 54,878 transactions failed (see Figure 5.1).

4.1.1.3. Walking Distance

In Table 2.3, various walking distance values in the literature can be seen. As the table indicates, different values have been preferred for allowable walking distance because this threshold is affected by several factors, including city architecture, demographics, terrain, transfer conditions, and transit mode in the analysis (Hussain et al., 2021).

To implement the walking distance in the algorithms, haversine distance function can be utilized (Assemi et al., 2020). Haversine distance is relatively easier to calculate than the network walking distance which often requires additional service, for example Google Maps distance matrix service. Using the latitude and longitude information of two points (stop/station) and the radius of the earth (6,371 km), haversine distance between these points can be calculated (Assemi et al., 2020). For this purpose, the built in “haversine” function of python library is used in our algorithms.

In our study, a sensitivity analysis is conducted to select the proper walking threshold, i.e., allowable walking distance. To perform sensitivity analysis, the haversine distance between the boarding stop/station and the closest stop/station on the route of the next trip is computed for all trips except for the last trip in a passenger's journey. For the last trip, the distance between the first trip's boarding stop and the nearest stop on the last trip's route is calculated. Since 2,000 meters is the maximum allowable walking distance used in trip chaining studies according to Table 2.3, higher distance thresholds are not considered.

The distribution of the calculated distances is given in Figure 4.2. As the figure indicates, the majority of the calculated distances (90.77%) are equal or lower than 2,000 meters. For instance, the cumulative percentage line on the figure indicates that alighting locations of 88.51% of the transactions can be estimated if the walking distance threshold is selected as 1,250 meters. Besides, it can be seen from the figure that increasing the walking threshold beyond 1,250 meters does not increase the number of alighting location estimation significantly. In light of the sensitivity analysis and after considering the preferences of local passengers and the characteristics of the public transportation infrastructure, we have determined the allowable walking distance threshold as 1,250 meters.

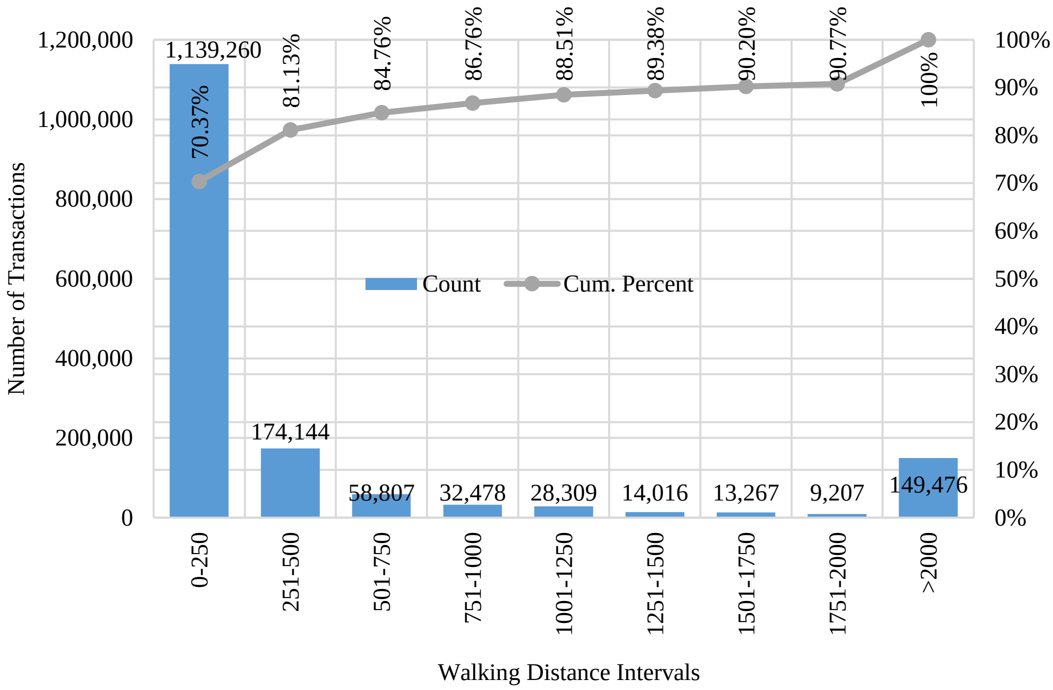


Figure 4.2. The number of transactions whose alighting location can be estimated based on walking distance intervals.

This interpretation also suggests that increasing the walking distance threshold in the trip chain algorithm over 1,250 has no significant influence on estimation performance. This is similar to another walking distance sensitivity analysis performed by António A. Nunes and Dias (2016), the failed estimation was 10.5, 7.7 and 5.8% for 400, 640 and 1,000 m thresholds, respectively (A. Nunes & Dias, 2016). In addition, Alsger et al. (2015) also found that transfer walking distance was less than 600 m for about 85% of passengers and increasing the walking distance beyond 800 m had no significant effect on the results (Alsger et al., 2015). However, in some cases, e.g., if there is a stop that provides higher level of service than the closest stops, passengers may be willing to walk longer distances to benefit from the high level of service (Assemi et al., 2020).

4.1.1.4. Transition Hour

In our smart card data, the day starts at 00:00 and ends at 23:59 by default. It means that trips that are the continuation of the previous day's trip chain may be present in the data. Since the algorithm sorts the data by the time to define the first and the last trips of an individual, this results in incorrect sorting. This is important because the chain algorithm infers the alighting stops of intermediate trips based on boarding location of the next trip, and infers the last alighting stop in reference to first boarding stop. To reduce the error and number of ignored trips (Cengiz, 2022), time internal definition of a day must be made adequately by determining the transition hour (virtual midnight). In the literature, several transition hours have been used; 4 a.m. (Cengiz, 2022; M. Munizaga et al., 2014), 5 a.m. (António A. Nunes & Dias, 2016), and 3 a.m. (Barry et al., 2009; Hora et al., 2017; Nassir et al., 2011). Meaning that, it may vary by location due to different PT policies adopted by the local authorities. Generally, the lowest activity observed in an hour on the day of analysis is used as the virtual midnight (António A. Nunes & Dias, 2016; Cengiz, 2022; M. Munizaga et al., 2014). We adopted the methodology presented by Munizaga et al. (2014) to determine the transition hour of a working day. Table 4.2 shows the number of transactions occurred after midnight. After 4 a.m. there are no transactions in the smart card data, thus these hours are not presented in the table.

In our case, the lowest number of activities (27) occurs between 3:00 and 4:00 a.m., so 4:00 a.m. is chosen as the transition hour. We presumed that public transportation operations for all modes begin at 4 a.m. and end at 3:59 a.m. the following morning. In addition, all transactions before 4 a.m. are moved to the following day under the assumption that passengers exhibit similar trip behavior.

Table 4.2. The number of transactions (activities) occurred during the time intervals.

Time Interval	**00:00 - 00:59**	**01:00 - 01:59**	**02:00 - 02:59**	**03:00 - 03:59**	**Total**
# of Activities	*3036*	*130*	*47*	*27*	*3240*

Note: The least number of trips are seen between 03:00 - 03:59. Thus, 4 a.m. is selected as transition hour. In total, 3,240 transactions are transferred to the next day.

4.1.1.5. Activity and Transfer Detection

Passengers may need to make one or more trips to reach their destinations. To determine whether a passenger performs a transfer trip, spatial and/or temporal thresholds must be defined. In general, the amount of time between consecutive journeys and spatial thresholds are used to determine whether an alighting location is a transfer point or the journey's final destination. Alsger et al. (2015) presented insightful findings about the effects of different transfer times and distance thresholds on the estimated O-D matrices using South East Queensland smart card data. They showed that passengers spend 85% of their transfer time for waiting or short activities not for walking, further, change in transfer time from 15 to 90 minutes had a slight effect, and the last alighting location (destination) for most passengers was within 800 m of their first boarding location (origin) (Alsger et al., 2015).

In addition, Nassir et al. (2011) considered all the trips made within 2.5-h interval as transfer trips based on the current Metro Transit fare policy, which states that all the trips made in the time span are free of charge (Nassir et al., 2011). It is critical to set proper time thresholds that reflect actual circumstances to determine the transfer trips correctly. Besides, it was found that passenger do not wait more than 30 minutes in the areas serves by high frequency bus and rail transit network (Ali et al., 2016).

Since the alighting time information is not available, the time gap between boarding times of consecutive trips is controlled for the time threshold, which is set as 120 minutes. This is the time span currently adopted in the fare policy that allows passengers to transfer free or with discount.

4.1.2. Alighting Estimation using Trip Chaining Algorithm

After establishing the necessary assumptions, the trip chaining algorithm can be constructed. The developed trip chaining algorithm follows similar rules that are presented by Trepanier (2007), but with the limitation that since the SC data is only for one day; it is not possible to establish a link with the historic data in order to estimate an alighting stop for unlinked trips.

The output data set of data cleaning process is the input of the trip chaining algorithm. The primary objective of performing both elimination and correction operations before implementing the trip chaining algorithm is to reduce its execution time. The alighting estimation process, which processes İzmir’s transit data (containing 1,618,964 transactions in total) in less than 18 minutes on an eight-core 2.6 GHz laptop PC with 32 GB of RAM. This execution time makes the process feasible, so it can be utilized multiple times per day.

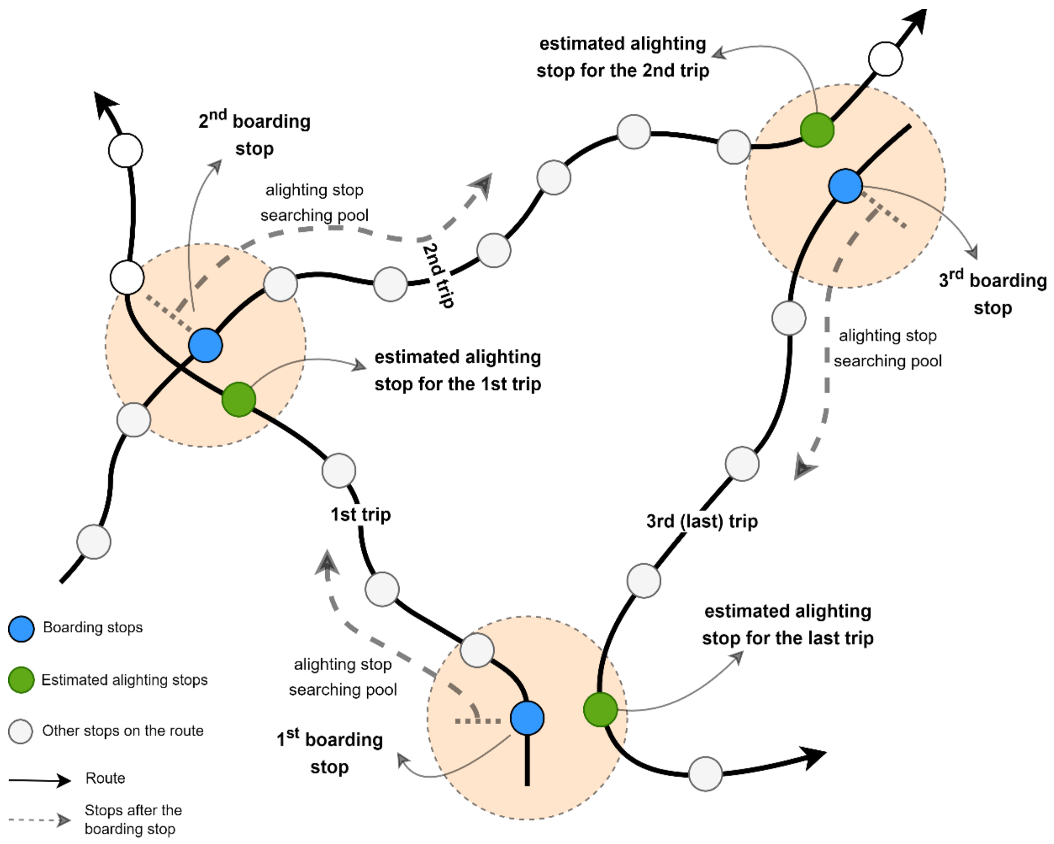


Figure 4.3. Alighting estimation process in the developed trip chaining algorithm.

The alighting estimation process is illustrated in Figure 4.3. The figure illustrates the procedure as if all the three trips are made on bus routes, but almost the same logic applies to metro, tram, commuter rail, and ferry. We have information about the direction of bus trips, allowing us to narrow our search for the alighting stop by applying the 5th assumption. This is necessary for the bus routes since the bus may travel a different route in each direction. However, direction information is unavailable for the other modes. Consequently, the alighting station is searched considering all the stations on the mode's route. Since the routes of the other modes of public transportation are identical in both directions, it is not anticipated that failure to implement the fifth assumption will compromise the results' reliability.

As illustrated, all trips except the final trip (first and second trips in Figure 4.3) are associated with the subsequent trip to estimate the alighting stop/station. In contrast, the alighting stop/station of the final trip (third trip in Figure 4.3) is estimated based on the boarding stop/station of the initial trip. To estimate the alighting stop or station, a buffer is evaluated around the boarding stop/station. Here, the allowable walking distance threshold is used as the radius of the buffer. In addition, an alighting stop search pool is established for trips. If a stop/station in the alighting stop searching pool is within the buffer zone of the subsequent or initial trip's boarding stop/station, this stop/station is selected as the alighting stop. If multiple stops/stations are present within the buffer zone, the closest stop/station is selected. The primary steps in the trip chaining algorithm are as follows:

1) Sort the input data by transaction time.
2) Create a list of unique card IDs. (The following steps will be applied sequentially until all the elements in this list are processed.)
3) Select the card ID in the sequence.
4) Filter the data by card ID. The result of this step is a data set of all the trips done by the cardholder with the selected card ID.
5) Select the trip in the sequence.
6) For the trips but the last trip of the cardholder, the “next trip” is the trip which is right after the current trip in the sequence. If the trip is the last trip, consider the first trip of the card holder as the “next trip”.
7) Create the alighting stop/station searching pool, which is the stops/stations that the cardholder may have alight, based on the current mode. If the mode is bus, the alighting stop searching pool consists of only the stops in the direction that are after the boarding stop.
8) Calculate the distances between the stops/stations in the searching pool and the boarding stop/station of the next trip. Select the closest stop/station.
9) If the distance is higher than the allowable walking distance threshold, alighting stop/station is not estimated. Return to step 2 if the current trip is the last trip. If else, return to step (4).
10) If the distance is smaller than the allowable walking distance threshold, select the closest stop/station the searching pool as the alighting stop/station.
11) If the alighting stop/station is the same as the boarding stop/station, alighting stop/station is not estimated. Otherwise, the closest stop/station is selected as the alighting stop/station. Return to step 2 if the current trip is the last trip. If else, return to step (4).

4.1.3. OD Matrix

Although trip level alighting estimation is sufficient for the main objective of this study, an algorithm is implemented to the trip chaining method that creates the origin and destination pairs for every journey of a cardholder. This implementation increased the running time of the trip chaining algorithm by 8 minutes (26 min in total).

The process is that if a cardholder does not take a trip in 120 minutes (activity threshold) after the previous trip's boarding time, the alighting location is considered as destination of the passenger's journey, as explained in section 4.1.1.5. Accordingly, an additional Excel sheet is produced containing the number of journeys, origin, and destination of the journeys (stop/station level) and also transit modes used in the journey are presented, as in Table 4.3.

In total, 909,402 journeys from 500,876 cardholders are identified based on the activity threshold. The destination information for 181,645 (20%) of the journeys is not available due to trip chaining algorithm could not find an eligible alighting location regarding walking threshold. Also, based on the applied thresholds, it is obtained that 105,891 cardholders only have one journey in the day. 58,305 (55.06%) single journeys do not have an estimated alighting location.

Table 4.3. An example of the output of the OD algorithm.

Card ID	Journey Order	Origin	Destination	Transit Modes
KART 459410	1	50517	Halkapınar	Bus - Metro - Com. Rail
KART 459410	2	Halkapınar	50517	Metro - Bus
KART 440125	1	31364	31272	Bus
KART 124355	1	40955	40088	Bus
KART 124355	2	40088	40018	Bus - Bus
KART 124355	3	40019	40953	Bus - Bus
KART 481905	1	Demirköprü	Naldöken	Com. Rail
KART 481905	2	Naldöken	Demirköprü	Com. Rail
KART 401058	1	31318	Tepeköy	Bus - Com. Rail
KART 401058	2	Tepeköy	31319	Com. Rail - Metro - Bus

Further, using ArcGIS ArcMap 10.3, the stops and stations within the districts' boundaries are identified. There are 31 districts in İzmir and 11 of them constitute the central business district as shown in Figure 4.4. These districts are considered as traffic zones to be able to produce zonal passenger flows.

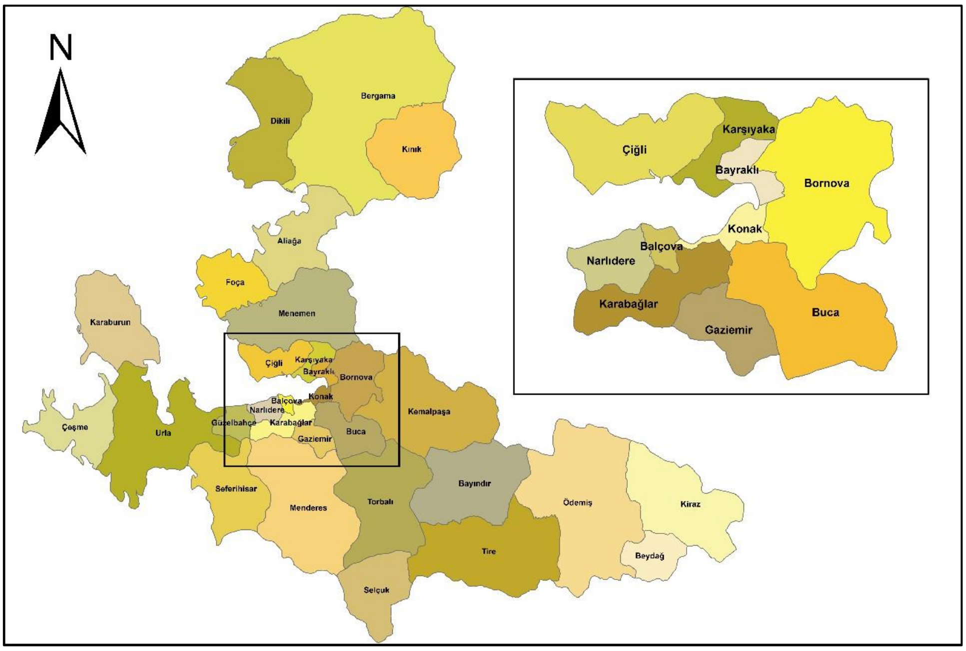


Figure 4.4. The districts (left) and CBD (top-right) of the İzmir city.

4.2. Inter-route Relationships

Here, the primary objective is to identify the bus routes which are expected to be affected by the extension of the metro line, as well as their relationship with the metro. For this purpose, in section 4.2.1, we determine the bus routes operating within the metro extension's service area. Then, in section 0, the relationship between these bus routes and the metro are established by calculating the competitiveness and cooperativeness of the bus routes. In order to do that, two indices are developed: cooperation index and competition index. Finally, the bus routes with the highest degree of competition are deemed to be the most affected from the extension of the metro line due to the presumption that the likelihood of passenger exchange is high between routes in competition (F. Wang et al., 2022; J. Zhang & Li, 2014). These bus routes are further analyzed to simulate the passenger's mode shift.

Important to note that, instead of using haversine distance, we prefer to use network walking distance to define the service area of a station. This process is explained in section 4.2.1. The primary steps in the procedure are outlined below.

1) Define the direction of the analysis, inbound or outbound.
2) Filter the data set of bus stops based on the analyzed direction. This data set consists of all the unique bus stops on the bus routes in the given direction.
3) For each unique bus stop, calculate the network walking distance to all the stations.
4) Determine the minimum distance and the related station.
5) If the minimum distance is below the threshold, i.e., if the stop is within the service area of any station, record the following: stop name, related station and distance. Otherwise, pass.
6) Determine the bus routes that have any of the recorded stops in their routes.
7) For each identified bus route:
    a) obtain the number of stops within the service area of the rail system on its route,
    b) obtain the number of stations whose service area contains at least one stop,
    c) calculate the competition and cooperation indices.

The result of the process contains the targeted bus routes as well as their competition and cooperation degrees with the rail systems. Details can be found in the following sections.

4.2.1. Targeted Bus Routes

To determine the bus routes that can be alternative to the metro, service are of the metro extension should defined (Soza-Parra et al., 2022). This service area is defined as a buffer zone which the area set around the metro stations that covers any possible alighting locations of transit passengers (Alsger et al., 2015). Based on the literature and considering the topography around the metro line, the service radius is selected as 600 meters. This value is lower than 800 meters which is the commonly accepted service radius for rail system’s station. 800 meters may be acceptable for access and egress activities; however, passengers tend to prefer walking shorter distances for transferring. Additionally, the extension of the metro line is expected to serve residents of the Narlıdere and Balçova districts. These districts have residential areas located on hilly terrain, and there are bus routes serving these areas. A large service radius for a metro station may cover portions of these areas that are not easily accessible by walking due to severe street slopes. Thus, 600 meters is selected as the radius of the service buffer around a metro station. Besides, we make use of Google Maps API Distance Matrix service to calculate the network walking distance between two points. The network walking distance is calculated to determine if the stop is within the service area of that station. The bus stop is considered to be within the station's service area if the distance to the station is equal to or less than 600 meters.

As illustrated in Figure 4.5, haversine distance results in much longer distances, thus may cover some streets that are not feasible to walk. In the case of using the haversine distance method, 72 unique bus stops are within the service area. On the other hand, if the network walking distance is used, 54 unique bus stops are withing the service area. By using network walking distance, we believe that the accuracy of the algorithms is enhanced. This is also important because of the factors, which are the number of stops and related stations, used to establish the relationship between bus routes and the metro line. After establishing the service radius around the metro stations, the stops within the service areas are determined. Then, the targeted bus routes are identified according to the stops within the service area of the stations.

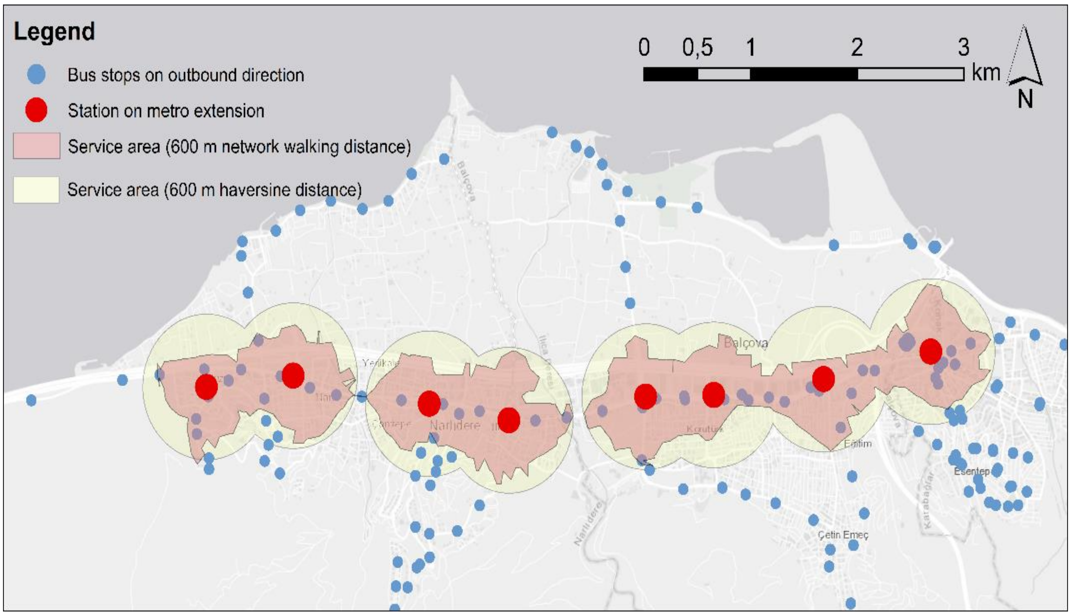


Figure 4.5. The illustration of the difference using network walking and haversine distance to define the service area of a station.

4.2.2. Competitive and Cooperative Bus Routes

As explained in the literature review, a relationship can be developed in two ways; competitive or cooperative. In addition, a contradictory relationship, which is called co-opetition relationship, may also be observed for some bus routes (Wei et al., 2020). The assumption is that when the collinearity between a bus route and a rail system increases, the level of competition also increases (Peng, 1994). The degree of competition and cooperation of a bus route is determined based on its collinearity with the rail system. To define the collinearity mathematically, we use two factors: 1) the number of bus stops within the rail system's service area, 2) the number of metro stations with at least one bus stop within their service area. Furthermore, competition and cooperation indices are developed to calculate the competitiveness and cooperativeness of a bus route. The above-mentioned collinearity factors used in these indices are explained in the sections 4.2.2.1 and 4.2.2.2.

The developed algorithm for this process is so adaptable that it can determine bus routes operating within a service area of any rail system. Since our goal is to determine the bus routes that operate within the service area of the metro extension, the algorithm is fed with the coordinates of the stations on the metro extension. Then, two factors are used to calculate competition (I_{com}) and cooperation (I_{coo}) indices.

4.2.2.1. Competition Index

As explained in the previous section, two factors are used to define the collinearity and competition index (I_{com}) is constructed as below. The first factor (α) is the ratio of the number of bus stops within the metro service area (b_w) to the total number of bus stops (b_t).

$$\alpha_{com} = \frac{b_w}{b_t} \tag{4.1}$$

$$\beta_{com} = \begin{cases} 0 \quad , & m_w = 1 \\ \dfrac{m_w}{m_t} \quad , & m_w > 1 \end{cases} \tag{4.2}$$

$$I_{com} = \frac{\alpha_{com} + \beta_{com}}{2} \tag{4.3}$$

where,

b_w is the number of bus stops within the defined service area,

b_t is the total number of bus stops on the route,

m_w is the number of collinear stations, i.e., number of stations which has a stop within its service area,

m_t is the total number of the stations,

α_{com} is the ratio of number of bus stops withing the defined service area,

β_{com} is the collinearity ratio,

I_{com} is the competition index which gives the competitiveness of a bus route.

Instead of simply using b_w, α_{com} is used to consider the length of the bus route outside the rail system's service area. For example, a bus route can be fully collinear to the rail system, but it may also operate outside of the rail system's service area for a reasonable amount of length. In this case, assigning a high competitiveness to this bus route will be misleading. Because this route also shows cooperative features by collecting passengers outside of the rail system's service area and transferring them to the rail system.

To calculate the value of the second factor (β_{com}), we defined a conditional function with respect to definitions in the literature (see Table 2.4). According to definitions in the literature, if b_w equals 1 (one), the competitiveness of the bus route is deemed to be zero, thus β_{com} is zero. Because this circumstance, in which a bus route serves within the service area of only one station of the rail system, is a characteristic of complementary (feeder) bus routes. Here, we adopt an approach similar to that of Gadepalli et al. (2022) and use this ratio as an approximation for the collinear length of the bus route to the metro (Gadepalli et al., 2022). Thus, if b_w is higher than 1 (one), β_{com} is the proportion of the number of metro stations with at least one bus stop within their service area (m_w) to the total number of metro stations (m_t).

For example, bus route 5 has 28 stops on its route in outbound direction where 24 of them are within the metro service area. Consequently, s_t and s_w are equal to 28 and 24, respectively. In this case, α_{com} is calculated as 0.857. In addition, bus route 5 is related to all the stations of metro extension, thus m_w, which is 8 (eight), is equal to m_t. Then, β_{com} is equal to 1 (one). Finally, the competitiveness of bus route 5 is calculated as 0.929 by using competitiveness index.

4.2.2.2. Cooperation Index

The cooperation index (I_{coo}) is defined as the exact opposite of the competition index. It means that the summation of cooperation and competition index results of a bus route is equal to 1 (one). Once again, two factors are used to calculate the cooperation index (I_{coo}) and is defined as below. The definition of the variables is already presented in the previous section; thus, they are not repeated here.

Since the cooperation between a bus route and a rail system decreases with an increase in the collinearity between them, α_{coo} is defined as (4.4. To calculate the second factor (β_{coo}), a conditional equation is used (equation (4.5). We defined a conditional function and when the m_w equals to 1 (one) β_{coo} is also equal to 1 (one). As explained in the previous section, the reason is that being related to only one station is a feature of complementary (feeder) bus routes.

$$\alpha_{coo} = 1 - \frac{b_w}{b_t} \tag{4.4}$$

$$\beta_{com} = \begin{cases} 1 \quad , & m_w = 1 \\ 1 - \dfrac{m_w}{m_t} \quad , & m_w > 1 \end{cases} \tag{4.5}$$

$$I_{coo} = \frac{\alpha_{coo} + \beta_{coo}}{2} \tag{4.6}$$

4.2.2.3. An Example for Competitive and Cooperative Bus Routes

In Figure 4.6, three bus routes in relationship with the metro line are visualized to discuss competing and complementary relationships on real examples. These bus routes are currently in operation. The metro line, represented with red, and its stations' service area are indicated by hatched areas. We use 600 meters network walking distance to define the service areas. As it is seen, Route 10 (green line) runs parallel to the metro line on the collinear corridor between F. Altay and Konak stations. One might consider this bus route competitive by only examining the relationship heuristically because it runs fully collinear to the metro. However, this bus route is only related to three metro stations with 4 stops, and it is complementary rather than competitive. Further, Route 167 (blue line) is an endpoint collinear bus route primarily operating as a feeder bus having complementary relationship with the metro line. On the other hand, Route 971 (pink line) also has a complementary relationship with the metro line, thereafter, serving the passengers traveling within Narlıdere and Balçova. In addition, this bus route will be an intermediate collinear bus route after the extension of the metro line from F. Altay to Narlıdere. Although characteristics of bus routes can be roughly identified heuristically by looking at their spatial representation, this is not a practical and an accurate approach. Hence, a systematic approach should be adopted. In this study, we have developed a systematic framework based on the abovementioned definitions to determine the competing and complementary bus routes within the service area of any given group of stops/stations in section 4.2.1.

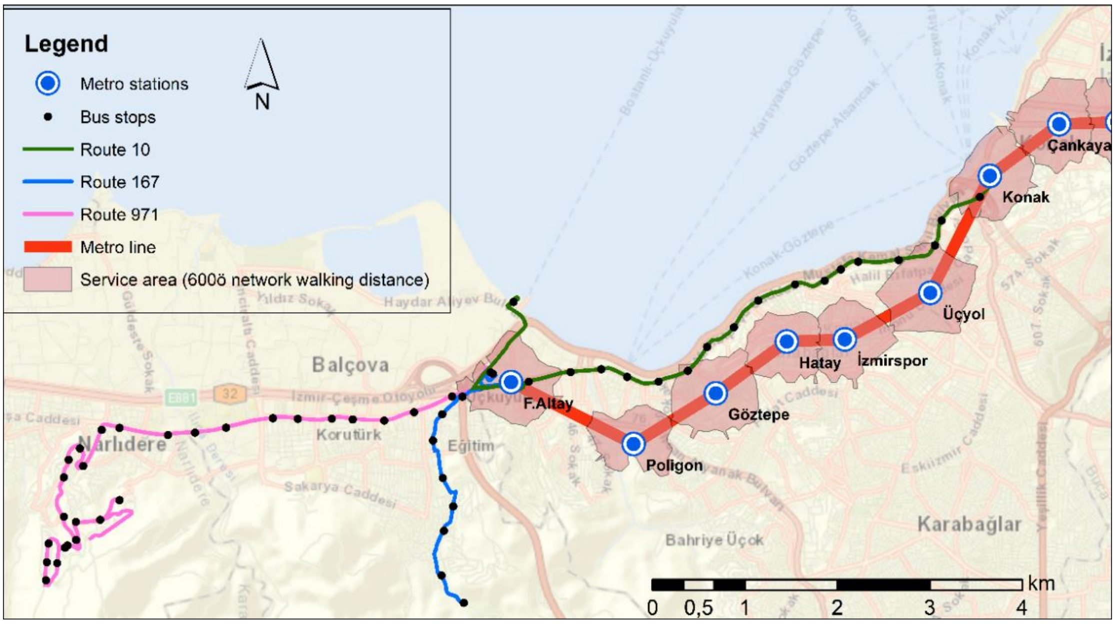


Figure 4.6. Example bus routes for cooperation and competition.

4.3. Passenger Load Profile

Passenger load profiles are used in the operational level transit planning especially for service adjustments (Pelletier et al., 2011). Although manual surveys have frequently been used to estimate onboard passenger loads, such surveys are too expensive to be completed daily across all available services, and they are also prone to mistakes and biases (D. Luo et al., 2018). For this reason, utilizing passive-data to obtain such transit planning inputs is valuable. For example, passenger loads can be determined by combining AFC, AVL and GTFS data which are collected from the urban transit systems (D. Luo et al., 2018).

Passenger load profile or occupancy levels may be determined using the boarding and alighting counts on stops/stations. In this thesis, an algorithm is developed to determine the passenger load profile of any bus route. Passenger load profiles provide a rough insight into the bottlenecks in bus operation in terms of service quality. We have utilized one-day of smart card data containing boarding and corresponding alighting locations, which are estimated using trip-chaining algorithm in the previous section.

Since both boarding and estimated alighting locations are available, load profile of any bus route can be obtained. However, there are trips with a failed alighting stop estimation due to reasons explained in section 2.3.2, such as unlinked trips. Barry et al. (2009) assigned an alighting stop to passengers who have no estimated alighting stop information by uniform sampling on the basis of observed distribution from passengers at the same stop/station with estimated alighting stop information (Barry et al., 2009). This issue also creates unbalanced trip generation and attraction pairs, thus requires a correction to create OD matrices. To estimate system-wide OD matrix, A. Cui (2006) was scaled the number of the trips with inferred alighting stops to match the control totals (entry-exit counts) for the rail system (A. Cui, 2006). In the purpose of obtaining load profile of the bus routes, the last stop on the direction is assigned as an alighting stop for failed estimations. Although this assumption has a negative impact on the accuracy, it has an insignificant influence in our case because of high matching rates (see Table 5.4) found for the bus routes operating within the metro extension’s service area. In other words, the number of passengers whose alighting stop is not estimated is very low for the targeted bus routes.

Besides, in the smart card data, there is not any information about the bus runs of a bus route. Thus, the runs of a bus route must be determined priori for obtaining the passenger load profiles. In addition, the determination of bus running times, headways and frequencies of a bus route also requires the identification of bus run numbers since this information cannot be directly obtained from the smart card data.

4.3.1. Bus Runs

Similar to the method one presented by Deri (2018), vehicle plates and bus stop orders are used to calculate the number of bus runs performed in a day. The time gap between the earliest and latest boarding transactions is primarily utilized to prevent incorrect run assignment, such as that caused by the one-stop-boarding problem. As with all other procedures, the data on the smart card is sorted by boarding time, from earliest to latest. The steps followed in the *bus frequency determination* algorithm are presented below.

First, smart card data is filtered based on the bus route number and the direction, for example bus route 551 and direction 1 (outgoing). Then for every boarding transaction on the bus route, vehicle plate, assigned bus run number and boarding stop order are recorded in separate lists named as *plate list, run number list,* and *stop order list*, respectively. The main objective for generating these lists is to be capable of retrieving the previous run number and stop order for the current transaction's vehicle plate. Previous run number and prior stop order refer to previously recorded run number and stop order for the present bus plate. For example, the bus plate, run number and boarding stop order for a transaction are AB123, 4 and 15, respectively. And assume that, for the next several transactions, a different vehicle plate is present. When vehicle plate AB123 is observed again, the algorithm requires to reach the boarding stop order and assigned run number of the last time this vehicle plate was seen. Hence, these lists record the history of the inputs and outputs, so they can be reached any time.

For the first transaction, the vehicle plate and the transaction are assigned to the first bus run. Then, the vehicle plate and the run number are stored in *vehicle plate list* and *run number list*, respectively. Also, the corresponding boarding stop order is stored in the *stop order list*.

For the next boarding transaction, if the vehicle plate is encountered before, i.e., if it is in the *vehicle plate list*, the current bus stop order is compared with the previous bus stop order. If the current bus stop order is equal to the previous stop order, it is assumed that the bus run continues, and the current transaction is assigned to the previous run number. Here, a control parameter must be added in case of boardings happening at the same stop on different bus runs. For example, the last boarding on the previous bus run and the first boarding on the current bus run may have happened at the same stop. In this case, the time difference between these consecutive transactions can be used to distinguish bus runs. Even in extremely high demand situations, the time difference between two boarding transactions at the same stop, i.e., dwelling time, is usually not expected to be greater than 2-3 minutes. In our case, there are two exceptions for this assumption: 1) drivers let passengers board buses at terminus on layover time, thus, there might be transactions happened in longer time gaps, 2) one-stop-boarding problem. In one-stop-boarding problem, all the boarding transactions are assigned to one stop on the bus route, however, boarding times are correct. In case of one-stop boarding problem and low demand situations, the time difference between boardings might be nearly high as bus running time. Thus, the time difference threshold here is set as 20 minutes. If the time difference between the corresponding boarding time of the previous transaction and the boarding time of the current transaction on the current vehicle is higher than 20 minutes, the vehicle plate and the transaction is assigned to the next bus run.

If the current stop order is higher than the previous stop order, it is assumed that the latest bus run continues, and the current transaction is assigned to the previous run number. We also need to consider the possibility that the run may have changed even though the current stop order is higher than the previous stop order. So, a time threshold, which is 30 minutes in our algorithm, must be used to distinguish these bus runs. To determine the value for this threshold, the worst scenario is considered. For example, consider a bus run having a boarding transaction at its first stop. The next boarding on the same run may have happened at a stop near the terminus, i.e., after some time that is close to route's running time. Thus, the time threshold needs to be higher than the typical running time. If the time difference between consecutive boardings on the same vehicle is higher than 30 minutes, the vehicle plate and the transaction is assigned to the next bus run. Finally, if the current bus stop order is lower than the previous one, it is assumed that the bus has started a new bus run; the vehicle plate and the transaction are assigned to the next bus run.

If the vehicle plate is not encountered before, i.e., if it is not in the *plate list*, the vehicle plate and the transaction are assigned to the next bus run. The bus runs of a bus route can be determined using this algorithm as well as the bus frequency.

Here, due to failing to change the direction, there are bus runs that seem to last from the beginning to the end of the operation. This causes to have some bus run with unusually high passenger count. This problem is tried to eliminate in section 4.3.6.

4.3.2. Bus Running Speeds

Before explaining the algorithm, the following should be stated. The stop data set has the stop-to-stop haversine distances calculated using the latitudes and longitudes of consecutive stops for each bus route. Furthermore, to obtain the real bus route lengths, all the bus routes are created on ArcGIS ArcMap software using open-source GPS data of bus routes provided by İzmir municipality. Then, the real route length of a bus route is determined. Generally, haversine distances are scaled by $\sqrt{2}$ to obtain realistic distances (M. D. Yap et al., 2017). Instead of this approach, stop-to-stop distances are scaled to equalize the total haversine route length and real route length of a bus route and added to the bus stop data as *corrected distance*.

After the determination of the runs of a bus route, the bus running speed can be calculated. To determine the running speeds for all the runs of a bus route, we simply utilized the boarding times at first and last stops where the boarding transactions occurred on a run. For the sake of clarity, the capital letter F is used to denote the first stop which has a boarding transaction on a bus run. Likewise, L is used to represent the last stop. Then, using the latest transaction time at F and the earliest transaction time at L, the travel time is calculated. And then, the distance between stop F and L and travel time is used to calculate the running speed of the run. Finally, all the calculated speed values are recorded in a list based on the relevant run number so that they can be accessed at any time.

However, there are several complications that must be considered. First, there might be only one boarding (low demand situation) on the run. In this case, there is not enough data to calculate the running speed. The second issue is the one-stop-boarding problem which is the case where all the boarding transactions are assigned to one stop. Third, some several reasons, the results of running speed calculations might be inconsistent, i.e., as too high or low speed values could be calculated. In our algorithm, the upper and lower limits for a bus speed are set as 15 and 55 km/h, respectively. In any of the three cases, the algorithm first looks at the list where the running speeds are recorded.

4.3.3. Bus Running Times

The total travel time, i.e., running time, of a bus route run can be calculated by dividing the route's total length to its running speed. A route's total length is obtained by determining the distance between its first and last stops. When the running speed between the farthest stops with boarding transactions is determined, this speed value can be presumed to be constant for that run. The following equation is used to calculate the run times of a bus route:

$$\mathrm{RT} = \frac{\mathrm{distance}(\mathrm{s}_1, \mathrm{s}_\mathrm{n})}{\mathrm{RS}} \tag{4.7}$$

where,

RT is running time for the run of a bus route,

$distance(s_i, s_j)$ is the function that returns the distance between given two stops on a bus route,

RS is the calculated speed for the run of a bus route,

s_1 is the first stop on a bus route,

s_n is the last stop on a bus route.

Bus running times are utilized in section 4.3.6 to eliminate the unrealistic results due to not properly setting the bus direction.

4.3.4. Travel Times (Bus in-vehicle Times) and Alighting Times

In the previous section, the running speeds of bus routes, which are specific to each run, are determined. By using the running speed information, bus in-vehicle times and the alighting time of each passenger can be obtained.

Bus in-vehicle time is the time that is spent in the bus to travel from the passenger's boarding stop (s_b) to estimated alighting stop (s_a). There are two approaches used to calculate a passenger's bus in-vehicle time. First, we look for a record of a boarding transaction at the estimated alighting stop. If there is a boarding transaction at the estimated alighting stop, it means that we have the information about when the bus arrived at the stop. In this case, the passenger's alighting time is assumed to be equal to the boarding transaction's time stamp. So, the travel time (bus in-vehicle time) of a passenger is calculated using the following equation:

$$\mathrm{IVT_b} = \mathrm{TS_b} - \mathrm{Tstamp(s_a)} \tag{4.8}$$

where,

TS_b is the time stamp of the passenger's boarding transaction,

$Tstamp(s_i)$ is the function that returns the time of the earliest boarding transaction at the given stop s_i,

s_a is the estimated alighting stop of the passenger.

If there is not a boarding transaction record at the estimated alighting stop, we use the running speed and the distance between the boarding (s_b) and alighting (s_a) stops of the passenger to calculate the bus in-vehicle time, IVT_b. In this case, bus in-vehicle time of a passenger is calculated using the following equation:

$$\mathrm{IVT_b} = \frac{\mathrm{distance(s_b, s_a)}}{\mathrm{RS}} \tag{4.9}$$

where,

RS is the running speed calculated for the bus run that the passenger boards,

$distance\left(s_i, s_j\right)$ is the function that returns the distance between bus stops s_i and s_j.

With the calculation of passengers' travel time (bus in-vehicle time), the alighting time of passengers can be determined by simply adding the value of IVT_b to their boarding time. Furthermore, a control check is employed for the determined alighting times if the calculation is done based on the running speed. Because there is an approximation in using the running speed for calculating the travel times since the speed of a bus is not likely to be constant. Thus, there might be some alighting time estimations that are later or earlier than the actualized alighting times. For the early estimations, there is not any information to use to correct the estimation. However, for the late alighting time estimations, we utilize the boarding time of the passengers' next trip. If the alighting time estimation is later than the next trip's boarding time, the boarding time of the next trip is considered as the alighting time of the previous trip.

4.3.5. Trip Purpose: Transfer or Activity

In section 4.1.1.5, we roughly determine the origin and destination of passengers by considering walking distance and time thresholds. However, as previously mentioned, the time threshold is controlled between boarding times since the alighting time information is not available. Since the alighting time of a passenger is calculated in section 4.3.4, it can be used to determine whether the trip is a transfer or not as illustrated in Figure 4.7. In this way, the transfer identification is enhanced. Important to note that, surely the alighting time estimation process could be done simultaneously in the trip changing algorithm, however, this would dramatically increase the running time of the algorithm. On the other hand, in this section, we handle the process for bus route level, thus the data size is relatively very small. Hence, sacrifice can be made here on the running time performance in order to enhance the results.

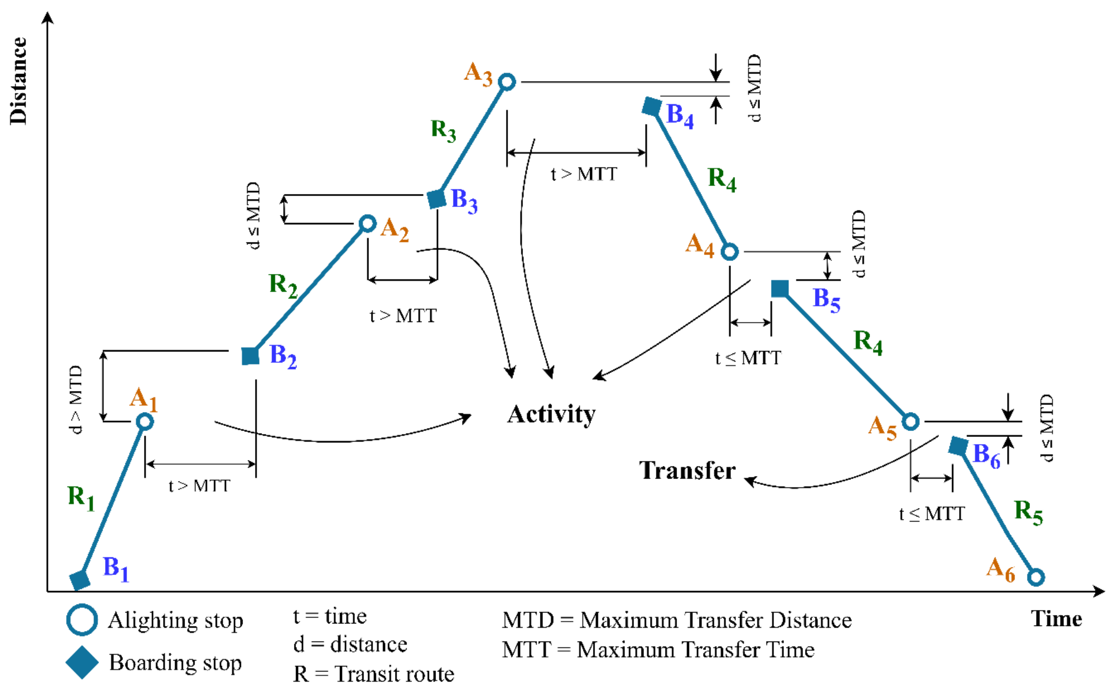


Figure 4.7. Spatial and temporal visualization of transfer and destination detection procedure.

Transfer time is simply calculated as the difference between the boardings time of the subsequent trip and the estimated alighting time of the previous trip (Wang et al., 2011). In the light of these definitions, a trip is defined as transfer if the distance between related boarding and estimated alighting locations is not higher than 1,250 meters (MTD) and the time difference between estimated alighting and next boarding time is not higher than 20 minutes (MTT). In this case, the alighting location is labeled as transfer point. On the other hand, if the time difference is higher than 20 minutes, the trip is defined as activity and the alighting location is considered as destination point.

4.3.6. Error Correction

Although determination of bus runs is sufficient to produce the passenger load profile of a bus route, due to one-direction problem, a correction is necessary to be applied. In summary, the one-direction problem is the case that all the runs of a bus route seem to operate in only one direction, outbound (direction 1) or inbound (direction 2). This problem also results in having one-stop boarding problem because smart card system is not able to relate the position of the bus with nearest bus stop on the bus route in the correct direction, it assigns all the boardings to the last stop in the incorrect direction.

In these cases, the time thresholds used in the bus run determination algorithm become insufficient because these thresholds consider the time difference between successive transactions. Thus, when the direction is always the same, the time difference between successive transactions is most likely to be under 20 minutes which is set to solve the issues due to one-stop boarding problem in bus run determinations algorithm.

For this reason, an additional step is applied. This correction is only applied to the runs that are labeled as "one-stop boarding problem" by the bus run determination algorithm. In this case, the summation of the time difference between consecutive transactions is considered. When this summation is equal to the bus running time, which is the mean running times of all the bus runs determined in section 4.3.3, the current direction of the run is changed, and summation is started over. This process is continued for all the runs that have one-stop boarding problem. Then, the transactions with the changed direction are excluded from the passenger profile analysis.

One deficiency here is assuming that the initial direction is correct. For example, the first bus run may be in the inbound direction but in the smart card data, it can be assigned to outbound direction. The algorithm cannot distinguish this situation.

This correction provides more reliable occupancy analysis. On the other hand, due to the one-direction problem, the number of runs performed for a bus route may have been estimated incorrectly.

4.4. Passenger Flow Groups and Alternative Scenarios

As stated in the earlier sections, our objective is to establish which bus passengers are likely to shift to the metro based on their perceived travel time savings. For this purpose, alternative scenarios are simulated in which the bus passenger is forced to use the metro. Important to note that alternative scenarios are constructed to assess the travel time gain or loss for only considering the travels within collinear section. We consider sufficient for estimating the passengers' decision by solely assessing the effect of metro within the collinear section on the travel time saving without considering the travel history of passengers prior or later (J. Cui et al., 2020). However, as suggested, analysis must be done at passengers' journey level instead of trip level to be able to explain the passengers behavior (Soza-Parra et al., 2022). Hence, we enhance our trip level analysis by looking at the next trip's purpose (transfer or destination) and mode (metro or others) while defining passenger flow groups. Then a set of rules are defined accordingly to construct the most plausible alternative scenarios. For this purpose, passenger trips performed within the service are of the metro extension, i.e., trips performed using targeted bus routes identified in section 4.2.1, are categorized in a accordance with (Gao et al., 2022b).

In total, eight different kinds of trip behavior are defined and summarized in Figure 4.8. Then, we construct alternative scenarios of the actualized trips by considering what the passenger would have done if she had used the metro. For example, consider a passenger who boards and alights at stops within the metro service area. For this passenger, performing a metro trip is sufficient since the origin (boarding stop) and the destination (alighting stop) of the passenger's trip are within the service area of the metro alternative. Hence, only a metro trip is required to be simulated in the alternative (simulated) scenario. On the other hand, if both the boarding and the alighting stop of a passenger are not within the metro alternative, this passenger needs to perform transfer activities as well as the metro trip in the simulated scenario. In the case of the latter example, the passenger 1) makes a bus trip to transfer to metro, then 2) performs a metro trip, finally 3) performs another bus trip to reach the destination of the original trip. There are several different cases which require different configurations in the simulated scenarios. Hence, the configuration of the alternative (simulated) scenarios is constructed according to the passenger flow groups observed in the buses operating within the service area of the metro extension.

The passenger flows on bus routes operating within the metro extension's service area are categorized into two main groups. The main groups are created based on whether the boarding stop of a passenger is within the service area or not. We use the capital letter "A" if the boarding stop of a passenger is within the service area. If the boarding stop is not within the service area, passenger flow is denoted as "B".

Furthermore, these two main groups are divided into subgroups based on the location of the estimated alighting stop. Here, we also utilize the information of whether the trip is transfer to metro or has another purpose. Because if the bus trip has been made to transfer to the metro in the first place, the transfer time costs in the alternative scenario can be neglected. Thus, the main groups are further divided into three and denoted by using the numbers 1, 2 and 3, respectively.

Furthermore, there are some trips on the targeted bus routes that are not feasible to replace with a metro trip. For example, there are passengers having trips taken outside the metro extension's service area. These passenger flows (P2) cannot be replaced by the metro. On the other hand, there are very short bus trips taken within the service area, however only related to one metro station. In other words, there are trips starting and ending in the service area of only one metro station. In addition, a trip might start outside of the service area however ended in the service area of the first station on the trip's route. These trips are also impractical to be replaced by the metro, thus the latter two trip types are grouped as P1. As a result, groups P1 and P2 indicate the bus trips that are not possible or impractical to be repeated by the metro.

In summary, the passenger trips are categorized according to the locations of boarding and alighting stops. This process is applicable for every bus route operating within the metro service area. The algorithm begins by filtering all the transactions on a bus route based on predetermined bus runs and then applies the steps explained in the following sections to each passenger trip.

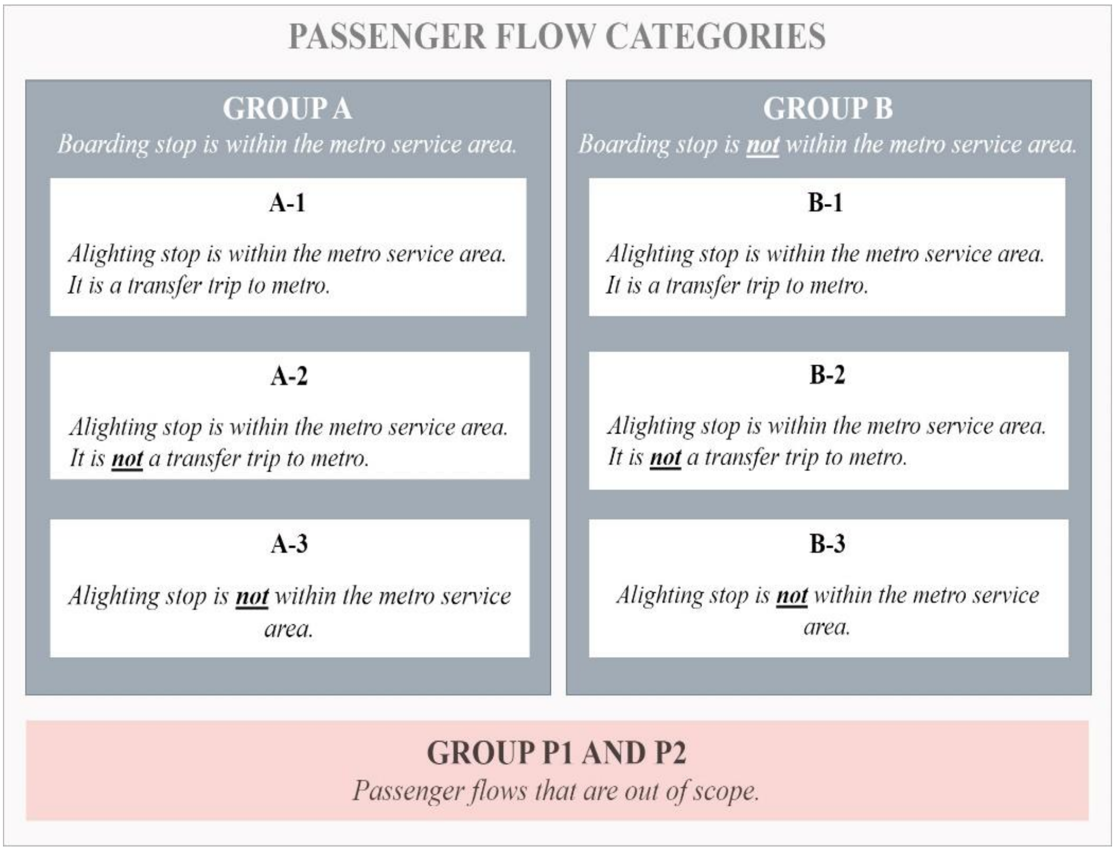


Figure 4.8. Passenger flow categories used in this study.

In the following sections, the alternative scenarios for each passenger group are explained. Since in this study the mode shift estimation of the passengers is determined according to the time savings, the time components are also presented. The explanation and calculation of mentioned time components in the following sections are presented in section 4.6.1.

4.4.1. Boarding Stop within the Metro Service Area (Group A)

If the boarding stop of passengers is within the service area of the metro extension, their trips are collected under the main group A. Group A is further divided into three subgroups according to the location of alighting stops and trip purpose. These subgroups and configured alternative scenarios for these trips are explained below.

1. *If the alighting stop (s_a) is within the service area of the metro extension and the trip is to transfer to metro.*

In the Figure 4.9, an example of a trip in group A-1 can be seen. In the actualized bus trip, passenger boards (s_b) and alights (s_a) at stops withing the service area of the metro extension. In the alternative (simulated) scenario in which the passengers are forced to use the metro, the following travel activities are assumed:

1) Passengers take the metro instead of taking the bus,
2) Passengers access to the metro at station M_b that is the closest to their boarding stop (s_b) in the actualized bus trip,
3) Passengers take a metro trip to reach the destination of their <u>subsequent trip</u>.

In the actualized trips, passengers in this group access the bus stop and wait for the bus. The first assumption states that the passengers do not require to perform any additional activity. In the same way, they access the metro station instead of accessing the bus stop.

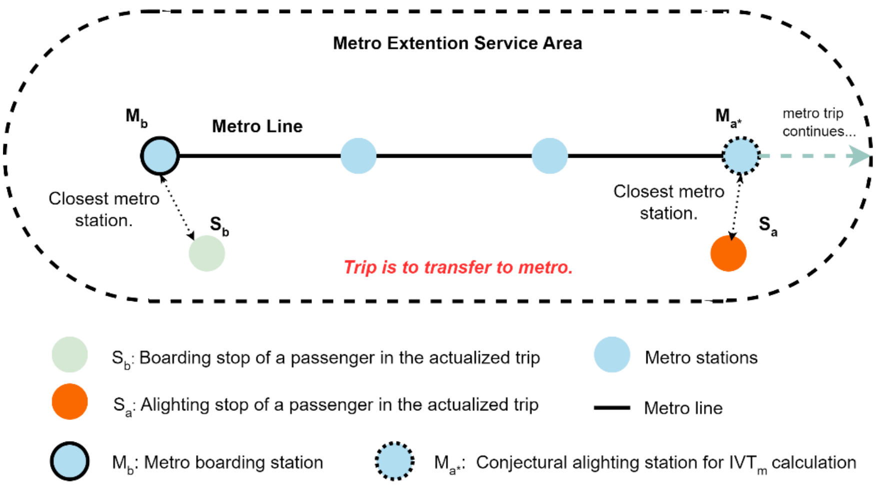


Figure 4.9. Illustration of the alternative scenario configured for group A-1.

Important to note that, to be able to compare the bus in-vehicle time of the actualized trip with the metro in-vehicle time of the alternative scenario, the travel time (IVT_m) between the boarding station (M_b) and the closest station (M_{a^*}) to the alighting stop of their actualized trips is calculated for the trips in this group.

2. *If the alighting stop (s_a) is within the service area of the metro extension and the trip is* ***<u>not</u>*** *to transfer to metro.*

The alternative scenario configured for the A-2 trips is nearly identical to the A-1 and illustrated in Figure 4.10. There are only in-station walking activities, for access and egress, additionally. The reason for grouping these trips, which are not to transfer to the metro, is to use this information to characterize the bus routes in further analysis. Also, since these passengers do not transfer to the metro, they alight at the station (M_a) which is the closest to their alighting bus stop (s_a).

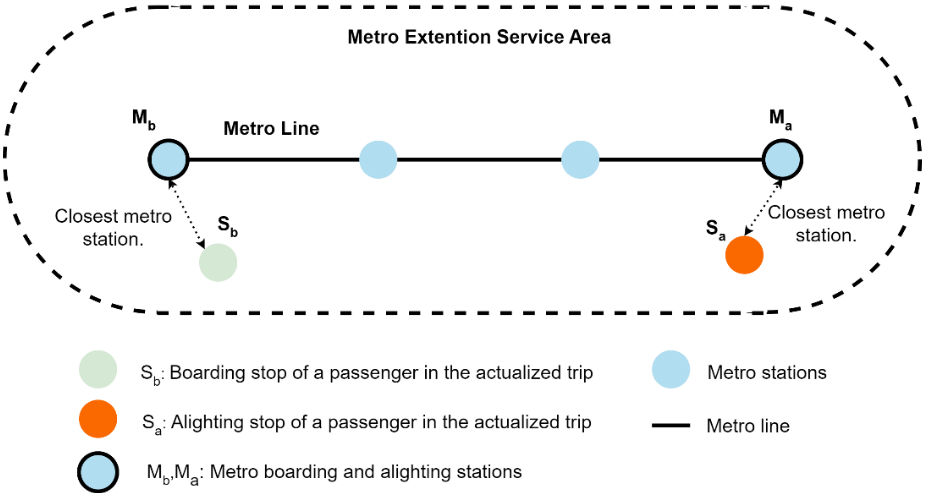


Figure 4.10. Illustration of the alternative scenario configured for group A-2.

3. *If the alighting stop (s_a) is **not** within the service area of the metro extension.*

In this group (Figure 4.11), passengers end their trips at a bus stop which is not within the service area of the metro extension. In this case, if the passengers prefer to use the metro, they also need to transfer to a bus route from the metro to reach their destination. For this reason, the following travel activities are assumed in the alternative scenario:

1) Passengers access the metro at station M_b, which is the closest to their boarding stop (s_b),
2) Passengers walk at the station M_b (in-station walking time),
3) Passengers take a metro trip (metro in-vehicle time),
4) Passengers alight at station M_a, which is the closest to any stop (s_{tr}) on the bus route,
5) Passengers walk at the station M_a (in-station walking time),

6) Passengers walk (transfer walking time) from M_a to s_{tr} to transfer to any bus route,
7) Passengers wait for a bus route (transfer waiting time) to board,
8) Passengers make a bus trip (bus in-vehicle time) to reach their destination (s_a).

As can be seen from the configuration of trips, the alternative scenario includes additional activities compared to A-1 and A-2.

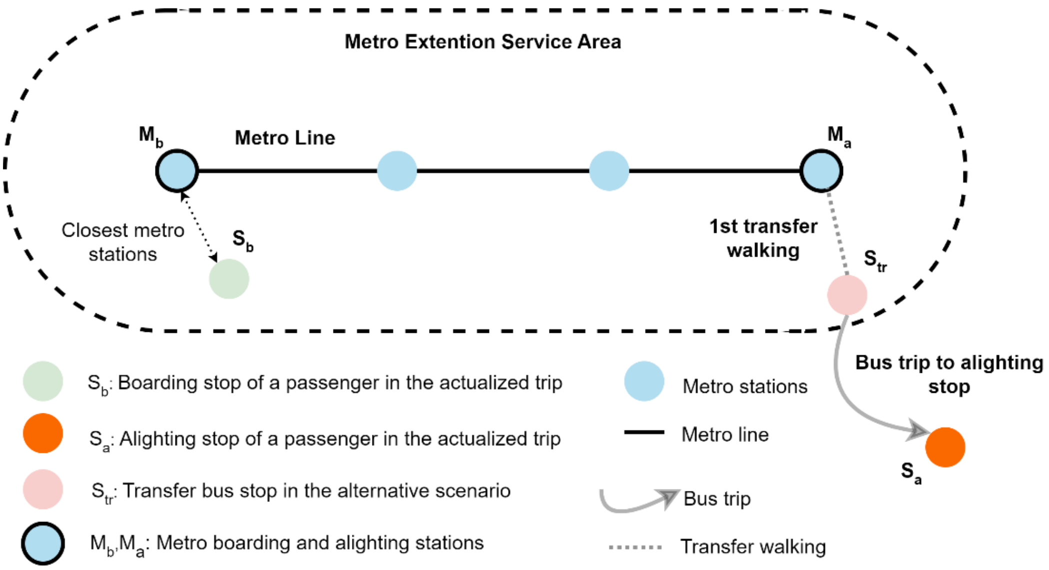


Figure 4.11. Illustration of the alternative scenario configured for group A-3.

4.4.2. Boarding Stop is not within the Metro Service Area (Group B)

If the boarding stop of the passengers is not within the service area of the metro extension, the trips of these passengers are grouped under main group B. In the same manner, the main group B is further divided into three subgroups according to the location of alighting stops and trip purpose. These subgroups are explained below.

1. *If the alighting stop (s_a) is within the service area of the metro extension and the trip is to transfer to metro.*

Passengers in this group start their trip at a bus station (s_b), not within the metro extension's service area. Besides, the purpose of the actualized trip is to transfer to the metro at a station that is also not within the service area. In other words, the only difference between actualized trip and alternative scenario is that passengers transfer to the metro at a closer station in the alternative scenario by taking a shorter bus trip. The activities in the alternative scenario can be listed as follows:

1) Passengers access the bus route at stop s_b,
2) Passengers take a bus trip (bus in-vehicle time),
3) Passengers alight at stop s_{tr} on the bus route, which is determined by the rules presented in section 4.4.3,
4) Passengers walk (transfer walking time) from s_{tr} to M_b to transfer to the metro,
5) Passengers walk at the station M_b (in-station walking time),
6) Passengers wait for the metro (transfer waiting time) to board,
7) Passengers have a metro trip (metro in-vehicle time) to reach the destination of their subsequent trip.

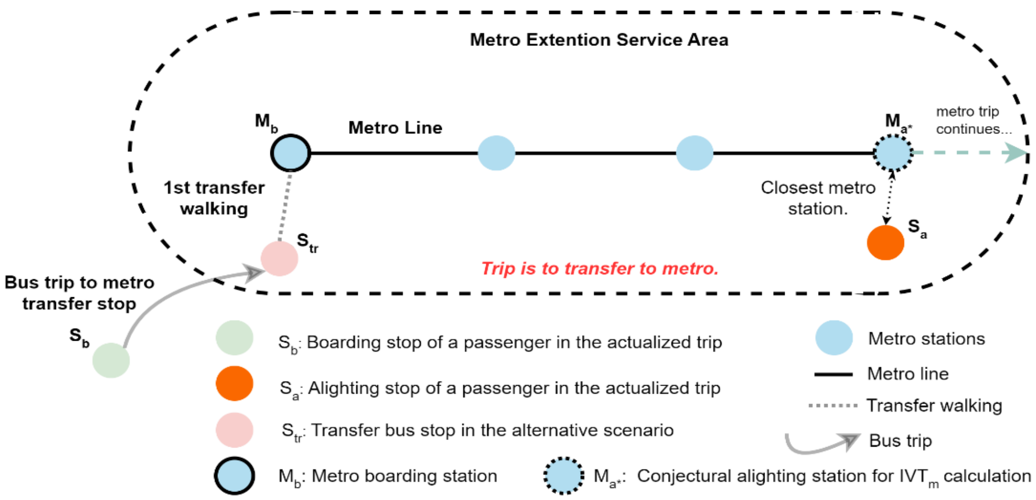


Figure 4.12. Illustration of the alternative scenario configured for group B-1.

As for the group A-1, the metro travel time (IVT_m) between the boarding station (M_b) and the closest station (M_{a^*}) to the alighting stop of their actualized trips is calculated to be able to calculate the travel time cost of the alternative scenario.

Importantly, it is assumed that all transfer activities prior to the metro trip in the alternative scenario, namely transfer walking (step 4) and transfer waiting (step 5), have no time cost. Because, if we consider the passengers' journey, these passengers will anyhow transfer to the metro at the end of their actualized trip.

2. *If the alighting stop (s_a) is within the service area of the metro extension and the trip is* ***not*** *to transfer to metro.*

As in group B-1, the boarding stop of the passengers is not within the service area of the metro extension. Unlike group B-1, passengers do not perform their actualized bus trip to transfer to the metro. Thus, they alight at a station (M_a), which is the closest to their alighting bus stop (s_a).

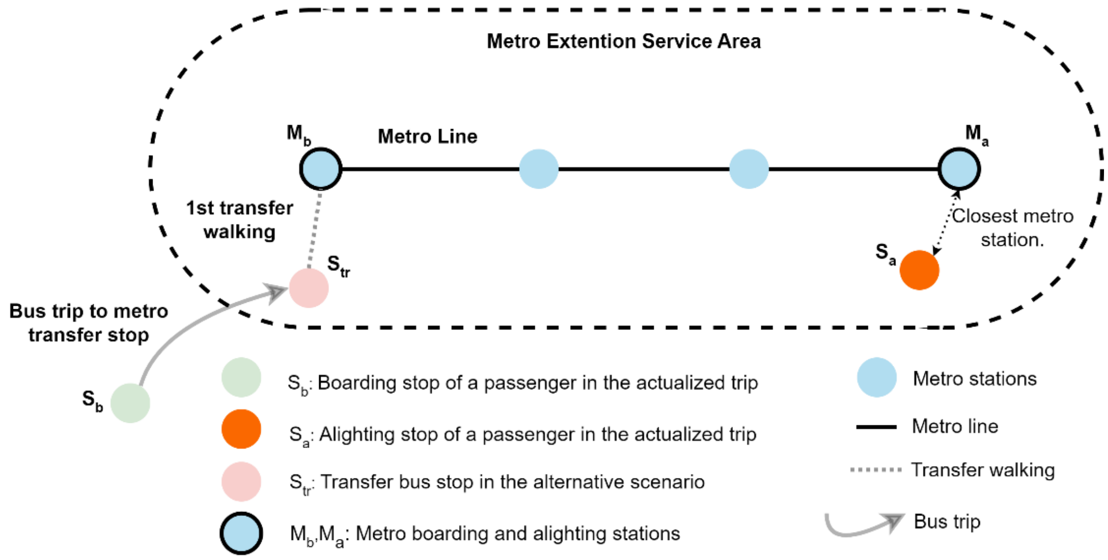


Figure 4.13. Illustration of the alternative scenario configured for group B-2.

For this reason, there are transfer costs to consider. The alternative scenario for this group is constructed as follows:

1) Passengers board the bus route at the stop s_b,
2) Passengers have a bus trip (bus in-vehicle time),
3) Passengers alight at a bus stop (s_{tr}) on the bus route, which is determined by the rules presented in 4.4.3,

4) Passengers walk (transfer walking time) from s_{tr} to M_b to transfer to the metro,
5) Passengers walk at the station M_b (in-station walking time),
6) Passengers wait for the metro (transfer waiting time) to board,
7) Passengers take a metro trip (metro in-vehicle time),
9) Passengers alight at station M_a, which is the closest to stop s_a on the bus route.
10) Passengers walk at the station M_a (in-station walking time),

Differently, these passengers do not transfer to the metro in their journey. Thus, there are transfer costs, namely transfer walking (step 4) and waiting (step 5), and in station walking times (steps 5 and 10) to consider. The calculations of these transfer costs are explained in 4.6.1.

*3. If the alighting stop (s_a) is **<u>not</u>** within the service area of the metro extension.*

In this group, both boarding and alighting stops of the passengers are outside the service area of the metro extension. Therefore, these passengers are the most disadvantaged in the alternative scenario. The configuration of the alternative scenario is as follows:

1) Passengers board the bus route at the stop s_b,
2) Passengers have a bus trip (bus in-vehicle time),
3) Passenger alight at a bus stop (s_{tr}^1) on the bus route, which is determined by the rules presented in 4.4.3,
4) Passengers walk (transfer walking time) from s_{tr}^1 to M_b to transfer to the metro,
5) Passengers walk at the station M_b (in-station walking time),
6) Passengers wait for the metro (transfer waiting time) to board,
7) Passengers have a metro trip (metro in-vehicle time),
11) Passengers alight at a station (M_a) that is the closest to any bus stop (s_{tr}^2) on the bus route,
12) Passengers walk (transfer walking time) from M_a to s_{tr}^2 to transfer to any bus route,
13) Passengers walk at the station M_a (in-station walking time),
14) Passengers wait for a bus route (transfer waiting time) to board,

15) Passengers take a bus trip (bus in-vehicle time) to reach their destination (s_a).

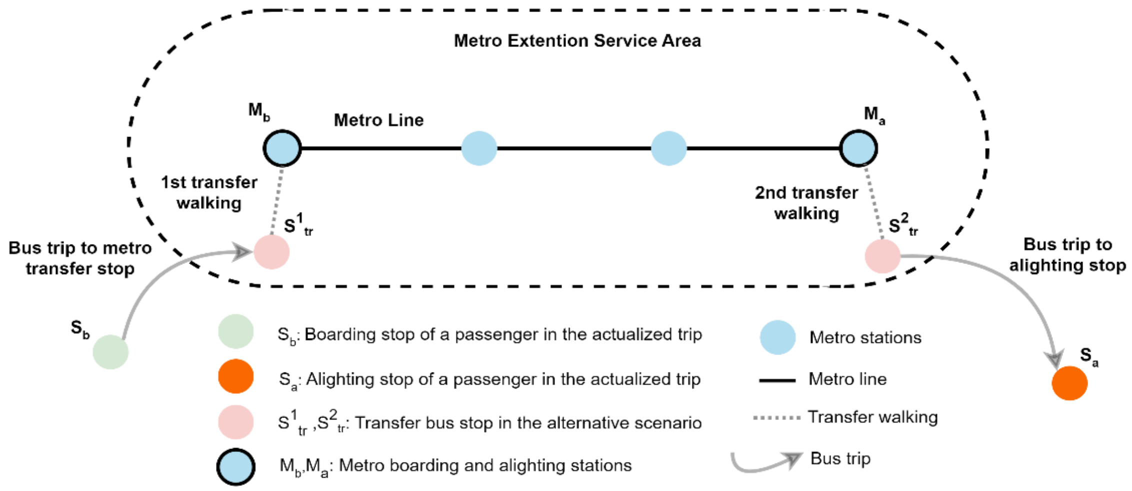


Figure 4.14: Illustration of the alternative scenario for group B-3.

It is clear that these passengers need to take three trips (2 bus trips and 1 metro trip) in the alternative scenario just to use the metro in their journey. The expected behavior from these passengers is to continue to use the bus routes after opening the metro extension.

4.4.3. Transfer Locations in the Alternative Scenarios

As explained in the previous section, some passengers must have transfer trip(s) if they want to use the metro in their journey. For this reason, the most plausible transfer locations are required to be determined. We define two different ways to determine plausible transfer locations: one for the transfers to the metro and one for the transfers from the metro.

The process of determining the transfer stop and station for the transfer trips to metro is illustrated in the Figure 4.15. The transfer stop is determined by considering transfer walking and bus in-vehicle times for each bus stop in the service area. In this process, first, the possible transfer stops, the closest stations to these stops, and the walking distances are determined. The walking distances are calculated using Google Maps API Distance Matrix service. This service requires the geographical coordinates of two points and calculates the walking distance in between. In this case, these two points are the possible transfer stop and the metro station. By using walking distance rather than haversine distance, we can determine the transfer locations with greater accuracy. After the walking distances are obtained, the transfer walking time (TWT) is calculated walking speed.

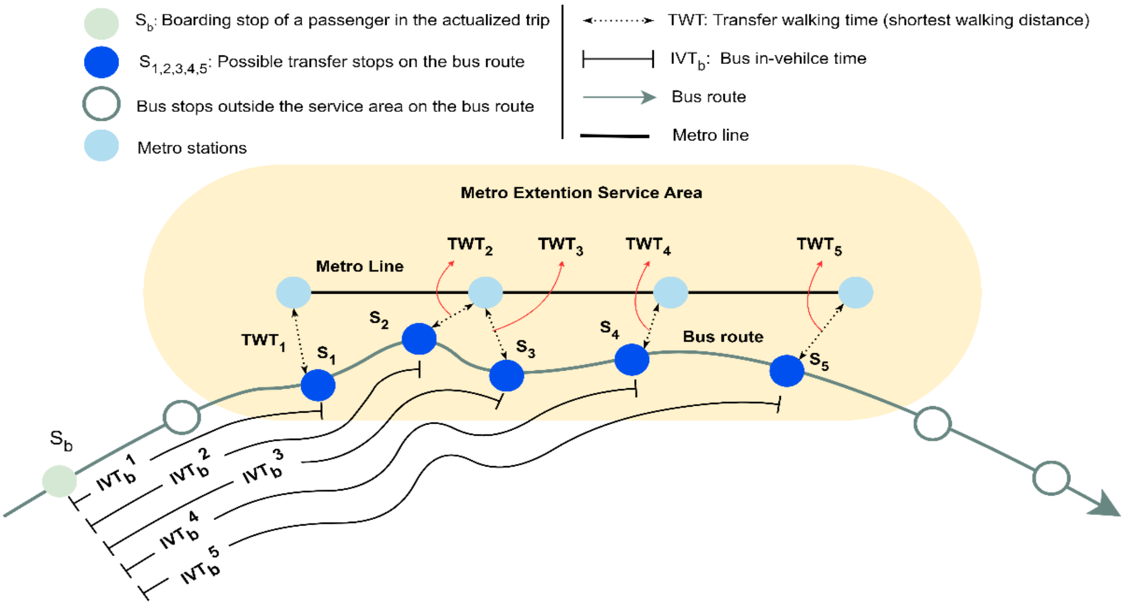


Figure 4.15. Presentation of transfer stop and station selection rules.

The walking speed of passengers is generated based on whether they are student, elderly, or others, as explained in section 4.6.1.1. Calculation of the transfer walking time is explained in section 4.6.1.5. Then, the bus in-vehicle time is required to determine the transfer stop. Bus in-vehicle time is calculated for all the stops within the service area of the metro extension as if they are the transfer stop. Also transfer walking times for these stops are calculated. The summation of bus in-vehicle and transfer walking times for each stop is calculated. The stop, which gives the minimal summation is selected as the transfer stop, and the related metro station is the transfer station.

For the transfers from the metro, the transfer station is simply determined by choosing the furthermost bus stop on the bus route, which is within the metro service area. We assume that passengers will prefer to travel by metro as long as possible to exploit its benefits. The reason for using these two different methods is that the distances between bus stops are relatively shorter than the distances between metro stations. Thus, passengers have the flexibility to choose a stop that minimizes the transfer walking time as well as the bus in-vehicle time. For these reasons, the transfer stop for bus passengers is selected based on the total of transfer walking and bus in-vehicle times.

4.5. Travel Time Calculations

Passengers considers several factors in Since the travel time has been a significant indicator of passengers' mode choice (Fan et al., 2016; Gao et al., 2022b), travel time is evaluated regarding two different approaches: 1) considers the convenience of passengers, 2) considers passengers' travel time saving calculated deterministically. The time cost of actualized trips and trips in the alternative scenarios are calculated based on both travel convenience and deterministic approach using the functions developed in section 4.5.1 and section 4.5.2, respectively. These functions are evaluated and compared based on the time savings to predict the passengers who most likely will shift from bus to metro. Factors used in these functions are in units of time (minutes) and mainly are in-vehicle time, transfer penalty, penalty if no seat and, transfer walking and waiting times. The definitions of these terms are presented in section 0. The monetary cost is not included in the functions because fare prices of bus and metro are equal and transfers within 120 minutes are free of charge for all the fare classes but full fare class. Even though it is neglected in this study, there will be a minor additional monetary expense for the full fare passengers due to the necessity to transfer in the alternative scenarios. In the travel convenience approach, we utilize the time multipliers to implement the perceived inconvenience of passengers while traveling.

The monetary costs are not included and variables are scaled using time multipliers, so the functions developed for the travel convenience approach can be called generalized time functions (Wardman, 2014). On the other hand, the functions constructed for the deterministic approach are called time functions.

It is important to note that these functions are evaluated for each passenger trip instead of passenger flow groups. This approach results in a longer computational time but provides the flexibility to use disaggregated factors, such as bus running speed and seating capacity of the bus, as well as to define rules in the algorithm in micro levels, such as transfer location selection in section 4.4.3. One benefit of using bus running speeds is that the variation of running speeds due to the variation in traffic congestion can be considered. Hence, the variation in the service quality of a bus route is reflected in the calculation of the passenger's bus in-vehicle time. Furthermore, the seating capacity of a bus can be utilized to reflect the discomfort of travel standing on a bus.

4.5.1. Travel Convenience Approach

One of the earliest convenience definitions made by Claffey (1964) in a travel context. The definition can be put as that the convenience is greatest for passengers when they put least effort to adjust their plans and habits to use the transit, and when challenges to getting to transit stations and boarding transit vehicles are minimized (Claffey, 1964). We have used convenience evaluation for identifying the passenger most likely to prefer metro over bus after the opening of the metro extension, i.e., Narlıdere metro. It is a widely acknowledged fact that making public transportation more convenient increases its likelihood of being chosen over other modes and the demand for it (Wardman, 2014). There are several justifications to implementing convenience (or inconvenience) evaluation to this process. 1) It is an important component of the overall attractiveness of public transportation directly affecting the wellbeing of travelers, 2) Poor performance significantly discourages to use the transit, 3) Important to understand the trade-offs between public transportation convenience in the practice of transit planning, and 4) Convenience may capture the sensitivity of broader mobility objects (Wardman, 2014).

Generally, the relative importance of the variables, which have an influence on the passenger's mode selection, is obtained by conducting SP or RP surveys. Unfortunately, since such data is not available, the range time multiplier of the variables found in literature is considered in the evaluation of the time functions. We believe that time multipliers presented in section 2.5.2 are transferrable to some degree and can be used in the evaluation of time functions. For this reason, we collect the time multipliers presented in literature to gain insight into their variation. Nevertheless, comparing the time multipliers across studies is difficult due to the specific values may depended on the context of the case and other conditions, such as weather, in different countries (Nielsen et al., 2021). We prioritize simulating the conditions where shifting to metro is the most and least favorable for the bus passengers. Thus, the minimum and maximum values of the time multipliers for each time component are considered.

Based on the values presented in section 4.5.1.1, min and max values of time multipliers for each variable are defined. Since the time multipliers are used in the time functions, i.e., results give an impression about the inconvenience, low and high values represent the best and worst conditions. Thus, we can see the possible mode shift conditions by evaluating the functions using different relative importance attached to the variables.

4.5.1.1. Selected Min and Max Values for The Time Multipliers

Here, we determine the min and max time multipliers for transfer walking and waiting, transfer penalty and crowding inconvenience based on literature review and especially considering the values in Table 2.8.

First, for the transfer walking multiplier, the min and the max values are selected as 1.68 and 2.01, respectively. The time multipliers for transfer waiting are selected as 1.72 and 2.03 for the min and the max values, respectively. The max values were obtained from the analysis using RP surveys in which waiting and walking time multipliers generally found higher than SP values (Wardman, 2014). In addition, the recommended maximum value for out-of-vehicle time multipliers for calculations is two (2) (Wardman, 2014).

Second, we determine the min and max values for transfer penalty according to the review in section 2.5.4. The variation in the transfer penalties in the literature is very high, ranging between 5 to 30 minutes of in-vehicle time. Review shows that transfer penalties vary regarding the mode of transit and the purpose of trips. The min and max values are selected as 5 and 12 minutes for the transfer penalty, respectively. The upper bound of the transfer penalty is constraint by considering the transit policy, travel habits and expert view from public transit authorities.

Third, crowding time multipliers are aimed to determine. The important decision is made here that crowding effect is solely considered for the bus travels and disregarded for the metro travel. This decision is motivated by the lack of capacity and the passenger demand information since the extension has not yet commenced operations. However, it is worth noticing that this assumption, while having some influence, does not deviate the evaluation due to the postulated inherent advantages of the metro service, such as comfort, reliability, and time efficiency compared to the bus (Ben-Akiva & Morikawa, 2002). In addition, we believe that the effect of favoring metro in-vehicle time is compensated by using max values for other components. As a result, we make use of the load factor and defined set of crowding multipliers for min and max conditions based on Paris crowding multipliers (Kroes et al., 2014). These values are presented in Table 4.4.

Table 4.4. Crowding time multipliers for bus trips based on load factors.

Load Factor	**75%**	**100%**	**125%**	**150%**	**200%**	**250%**
Min	1.0	1.083	1.289	1.394	1.499	1.604
Max	1.0	1.102	1.342	1.467	1.593	1.718

Note: These values are taken from the study by Kroes et al., (2014).

4.5.1.2. Generalized Time Function for Actualized Trips

In order to reflect the convenience, crowding time multiplier is utilized. The calculated bus in-vehicle times (actual travel time) are multiplied with the defined crowding time multipliers. Since the analysis is conducted on trip level, actualized trips are singular trips and there are no transfer activities, thus the cost function only consists of bus in-vehicle time. In addition, a time multiplier, which reflects the effect of comfort on perceived travel time, is utilized. The cost function for the actualized trips is calculated using the following equations:

$$\mathrm{G_{ac}} = \alpha_{\mathrm{cr}}^{\mathrm{l}} \times \mathrm{IVT_b} \tag{4.10}$$

$$\mathrm{l} = \frac{\mathrm{V_i}}{\mathrm{V_{seat}}} \times 100 \tag{4.11}$$

where,

G_{ac} is the cost function for the actualized trips,

IVT_b is the travel time of a passenger on the bus, i.e., bus in-vehicle time,

l is the load factor,

α_{cr}^{l} is the crowding time multiplier for the calculated load factor at the boarding stop,

V_i is the passenger load, i.e., occupancy level, on the bus at the boarding bus stop of the passenger,

V_{seat} is the seating capacity of the bus (specific to each bus plate).

By using a crowding time multiplier, the convenience (or comfort) of a passenger is assured to be taken into account. To establish whether the magnitude of the crowding time multiplier, the loading factor is utilized.

4.5.1.3. Generalized Time Function for Alternative Scenarios

The second function is constructed for alternative scenarios where the passengers are forced to use the metro in their journey. In other words, this is the scenario where the passengers have to use the metro to repeat their actualized. This generalized time function consists of in-vehicle and out-of-vehicle time components. Here, in-vehicle time components refer to bus and metro in-vehicle times. Transfer walking and waiting times, in-station walking time, and transfer penalty components constitute the out-of-vehicle time function. The calculation of these time components is explained in section 4.6. Generalized time function for alternative trips is calculated using the following equations:

$$\mathrm{G_{alt}} = \alpha^{\mathrm{l}}_{\mathrm{cr}} \times \mathrm{IVT_b} + \mathrm{IVT_m} + \mathrm{T_{ovt}} \tag{4.12}$$

$$\mathrm{G_{ovt}} = \delta \times \left(\sum_{\mathrm{i}=1}^{2} \mathrm{TWT_i} + \sum_{\mathrm{j}=1}^{2} \mathrm{IWT_j} \right) + \theta \times (\mathrm{WT_b} + \mathrm{WT_m}) + \sum_{\mathrm{k}=1}^{2} \mathrm{TP_k} \tag{4.13}$$

$$\mathrm{WT_b} = \frac{\mathrm{H_b}}{2} \tag{4.14}$$

$$\mathrm{WT_m} = \frac{\mathrm{H_m}}{2} \tag{4.15}$$

where,

G_{alt} is the generalized time function for alternative trips,

G_{ovt} is the generalized time function for out-of-vehicle time components,

IVT_b is the bus in-vehicle time,

α^{l}_{cr} is the crowding time multiplier for the calculated load factor at the boarding stop,

IVT_m is the metro in-vehicle time,

TWT_i is the time for transfer walking, which may be required up to two times in our configuration,

WT_b is the transfer waiting time at a bus stop,

WT_m is the transfer waiting time at a metro station,

IWT_j is the in station walking time; up to two times,

H_b is the headway of a bus route,

H_m is the headway of the metro,

TP_k is the penalty for a transfer activity, up to two times,

δ and θ are the time multipliers attached to related attributes.

The transfer walking and waiting time multipliers, and the transfer penalty component are used to establish the passenger's travel convenient of the time in the alternative scenarios. Additionally, a constant in-station walking time is considered. The time multipliers defined for each attribute can be seen in 4.5.1.1.

4.5.2. Deterministic Approach

In this approach, we disregard the convenience or perception factors in the process of estimating the passengers who are likely to shift to metro after its introduction. Time components are computed for actualized trips and alternative trips using the developed time functions presented in the following sections.

4.5.2.1. Time Function for Actualized Trips

In contrast to generalized time functions, there are no time multipliers to scale the impedance of the time components. Thus, there is only bus in-vehicle time in the time function since access walking and waiting times are neglected. In summary, the time function for the actualized trips (T_{ac}) is equal to the bus in-vehicle time (IVT_b) calculated in section 4.3.4.

4.5.2.2. Time Function for Alternative Scenarios

In the alternative scenarios passengers might have transfer walking and waiting, in-station walking as well as bus trips to access or egress from the metro. Thus, the time function is evaluated for the alternative scenarios using following equations:

$$\mathrm{T_{alt}} = \mathrm{IVT_b} + \mathrm{IVT_m} + \mathrm{T_{ovt}} \tag{4.16}$$

$$\mathrm{T_{ovt}} = \sum_{i=1}^{2} \mathrm{TWT_i} + \mathrm{WT_b} + \mathrm{WT_m} + \sum_{j=1}^{2} \mathrm{IWT_j} \tag{4.17}$$

where,

T_{alt} is the time function for alternative scenarios,

T_{ovt} is the function for the out-of-vehicle time components.

Other attributes are already defined in the previous section. Main differences from the convenience approach are that time multipliers and transfer penalty are not used in the calculations.

4.6. Passenger's Decision: Shift or Stay

In the case of travel convenience approach, the decision of passengers is presumed to favor the more convenient travel option. Thus, we compare G_{alt} and G_{ac} values for each passenger. If G_{alt} is found lower than G_{ac}, the passenger is assumed to shift to metro. If otherwise, the passenger is assumed to stay, i.e., continue using the bus.

In the deterministic approach, we define two criteria: 1) the result of the T_{alt} must be lower than T_{ac}, 2) there must be at least 10% time saving in the alternative scenario. If these two conditions are satisfied, the passenger is assumed to shift to metro.

4.6.1. Calculation of Travel Time Components

In-vehicle time term refers to the amount of time a passenger spent in any public transportation mode while traveling. There are bus and metro in-vehicle times that are to be calculated for the actualized trips and the trips in the alternative scenarios. Transfer walking time and transfer waiting time are the components of transfer activities. In addition to these two, the transfer penalty is utilized to reflect the deterrence of transferring on the passengers' travel time perception. The alternative scenarios configured for the groups A-3, B-2 and B-3 include transfer components.

4.6.1.1. Walking Speed

Walking activity is a mode of transport (Wigan, 1995) which is regarded as the most efficient one (Gore et al., 2020). Since most of the trips on public transit services start and end with walking, it is important to understand what influences pedestrian walking. In particular, the influence of access, egress and transfer walking times are investigated in public transportation route choice studies and they are stated as some of the main descriptors of the passengers' route choices (Bovy & Hoogendoorn-Lanser, 2005; Eluru et al., 2012; Nielsen, 2000; Nielsen et al., 2021). For that reason, it is critical to calculate these values as accurately as possible. Since we use the walking speed in the transfer walking time calculations, it becomes important to be defined properly.

The walking speed of a pedestrian may vary depending on several factors, such as age, gender and walking purpose of the pedestrian as well as traffic and surrounding environmental characteristics (Knoblauch et al., 1996). For example, presence of parked vehicles found to have influence on walking speed (Gore et al., 2020) and using carriageways instead of sidewalks found to decrease the walking speed of passengers (Gore et al., 2020; Rastogi et al., 2011). Also, using cellphone during walking found to reduce the walking speed in general (Rastogi et al., 2011).

Rastogi et al. (2011) indicated that walking speed in the educational areas (85.27 m/min; 5.12 km/h) found to be 26% higher than the mean walking speed (67.87 m/min; 4.07 km/h). They also presented the walking speeds for different age groups and high variation was observed. The walking speeds for older, middle-aged and young pedestrians were 55.17 m/min (3.31 km/h), 68.73 m/min (4.12 km/h) and 77.23 m/min (4.63 km/h), respectively (Rastogi et al., 2011). On the other hand, Knoblauch et al. (1996) presented mean walking speeds as 1.51 m/sec (5.44 km/h) and 1.25 m/sec (4.50 km/h) for pedestrians younger and older than 65 years old, respectively (Knoblauch et al., 1996). Wigan (1995) presented walking speeds for men and women in different age groups based on the data collected in 1986, Australia. For men under 60 years old, the walking speed varied between 3.2 and 5.9 km/h with a mean of 4.7 km/h. For women, walking speeds varied between 3.3 and 5.8 km/h with a mean of 4.0 km/h. For pedestrians aged above 60 years old, the walking speeds were ranged between 3.5 - 4.5 km/h and 2.8 – 3.3 km/h for men and women, respectively (Wigan, 1995). Tanaboriboon et al. (1986) found mean walking speeds 54 m/min (3.24 km/h) and 76 m/min (4.56 km/h) for elderly and young pedestrians in Singapore, respectively. They stated that mean walking values were higher in U.S. and Britain, 81 m/min (4.86 km/h) and 78 m/min (4.68 km/h), respectively (Tanaboriboon et al., 1986). Dewulf et al. (2012) presented walking speed values for various age groups by genders as 4.71, 4.95 and 4.59 km/ hour for young, senior and elderly passenger where these values were 4.77, 4.79 and 4.28 for females (Dewulf et al., 2012). Cengiz (2022) considered 50 m/min (3 km/h) walking speed of senior people in Madrid to calculate their transfer times. These findings indicate that walking speeds vary regarding age and gender etc., thus walking should also be defined accordingly.

Table 4.5. The walking speed values from studies in literature.

Studies	Rastogi et al. (2011)	Knoblauch et al. (1996)	Wigan (1995)		Tanaboriboon et al. (1986)		
Segment			*Man*	*Women*	*Singapore*	*U.S.*	*Britain*
Mean	4.07	-	-	-	-	4.86	4.68
Educational area	5.12	-	-	-	-	-	-
Older	3.31	4.50	3.5 - 4.5	2.8 - 3.3	3.24	-	-
Middle-aged	4.12	-	-	-	-	-	-
Young	4.63	5.44	3.2 - 5.9	3.3 - 5.8	4.56	-	-

There is an obvious variation in walking speeds for pedestrian groups. Thus, we used fare type information, which is available in the smart card data, to classify passengers as students, elderly, and others. The latter group is considered to consist of a mix of passengers, thus represented with mean values. Based on the findings above, a lower and upper limit as well as the mode of walking speeds are defined for the three groups as in Table 4.6.

Table 4.6. Selected walking speed limits in our study.

Limits and Mode	Student	Older People	Others
Lower limit (km/h) *Left*	5.0	3.4	4.5
Peak (km/h) *Mode*	5.5	4.0	4.8
Upper limit (km/h) *Right*	5.8	4.4	5.0

To select a random walking speed, we use triangular distribution function in “NumPy” library in Python. Triangular distribution is a continuous probability distribution with the lower limit to the left, the peak at the mode, and the upper limit to the right. It is mostly used in simulations to represent the randomness in the choices. Instead of using explicit values for the walking speed, we prefer adding some randomness.

Unlike other distributions, these parameters define the probability distribution function's shape explicitly.

4.6.1.2. Bus In-vehicle Times

The calculation of bus in vehicle time of passengers is established in 4.3.4. Besides, the alternative scenarios may also include bus trips to and/or from the metro. Specifically, trips where the origin (boarding stop) and/or destination (alighting stop) are not within the metro service area, i.e., trips grouped under A-3, B-1, B-2, and B-3, includes bus trip(s). For those bus trips, bus in-vehicle times are required to be calculated. The transfer stops for those bus trips are determined in 4.4.3. So, bus in-vehicle times can be calculated using (4. presented in section 4.3.4.

4.6.1.3. Metro In-vehicle Times

Metro in-vehicle times are only required to be calculated for the alternative scenarios. The calculation procedure is similar to the latter approach presented in bus in-vehicle time calculation, thus (4. can be altered for this purpose. For metro operating speed, we use 40 km/h in the calculations. Similarly, $distance(s_i, s_j)$ function is used to calculate the distance between boarding (u_b) and alighting (u_a) stations determined in section. The determination of metro boarding and alighting stations of a passenger in the alternative scenarios is explained in 4.4.3.

4.6.1.4. Transfer Penalty

As the literature review indicates, passengers do not prefer the transfer activities and a constant disutility term (transfer penalty) is employed along with other transfer components to reflect this attitude. The alternative scenarios configured for the groups A-3, B-2 and B-3 include 5 minutes of transfer penalty for each transfer. Transfer penalty is not applied to A-1 and A-2 since they do not include transfer activities in their alternative scenarios. B-1 is also excluded even though its alternative scenario includes a transfer. The reason is that for the actualized trips are performed to transfer to metro in group B-1, thus, the transfer cost can be ignored.

4.6.1.5. Transfer Walking Times

Transfer walking time refers to the time that passengers spent walking to the boarding location of their next trip. In the alternative scenarios, passengers may require walking from a stop to station or vice versa. As explained in section 4.4.3, the transfer walking distances, i.e., network walking distances, are obtained using the Google Maps API. Then transfer walking time is calculated using passenger's walking speed, which is determined as explained in section 4.6.1.1.

The alternative scenarios of groups A-3, B-2 and B-3 include transfer walking. As explained above, the transfer walking time is disregarded for group B-1 because passengers in this group are having the bus trip to transfer the metro in the first place. This component is not required to be calculated for the other groups since they do not include a transfer activity.

4.6.1.6. In-station Walking Time

To represent the time passengers spent walking at stations, a fixed value of time is considered for the three groups: student, elderly, and others. This in-station walking time can be estimated by considering the walking time at the entry/exit of the station, walking time on the passage, stairs or escalator, and the walking time on the platform; requiring information about the stations (J. Cui et al., 2020). Since such data is not available, instead, we used fixed values. For example, Cengiz (2022) considered 5 minutes of in-station time in the transfer time calculations for Madrid (Cengiz, 2022).

Since in-station walking time is dependent on the walking speed of a passenger, we defined a constant in-station walking distance of 100 meters and in-station walking time is calculated based on given walking speeds.

4.6.1.7. Transfer Waiting Times

Since the smart card data only includes the boarding times, it is not possible to estimate the passengers waiting time at the stops/stations. In order to quantify the passengers transfer waiting times, we set the waiting time as equal to half of the scheduled headway of the boarding line in the defined peak hours (Yap et al. 2020a). The determined headways for the bus and metro are explained below.

For the metro, we defined 5 and 3 minutes of headways for the operating hour between 6 am to 7 am and between 7 am to 8 pm, respectively. For the hours outside of these time durations, the headway is defined as 7.5 minutes. These headways are based on the metro schedule declared for a weekday on the official website.

For the bus headways, morning and evening peak hours are determined as the hours in which the buses operate at the highest frequencies. Hence, AM peak and PM peak hours are obtained as 7 am to 9 am and 4 pm to 7 pm based on planned schedules, respectively. For the AM and PM peak hours, bus headway is taken as 7.5 minutes. For the exogenous hours, the headway is 15 minutes. Here, one explanation must be made. According to our configuration for alternative scenarios, metro-to-bus transfers are allocated where the station is closest to any bus stop that is on the boarding bus route. It means that there will be several bus routes operating at the transfer location. Thus, considering the individual headways will be misleading since there are other bus alternatives with different frequencies. For these reasons, we used common headway values.

4.6.2. Other Time Components

In this study, *access and egress times* are neglected. One reason is that the original location of access and egress activities are not available. Thus, whether the mode is bus or metro, the distance that the passenger travelled (by walking etc.) to and from the stop/station is unknown. We preferred to neglect these components for the sake of simplicity and to reduce having unrealistic results by implementing another assumption which is not supported by a local analysis.

CHAPTER 5

RESULTS

Explanatory analysis is conducted on the data which is obtained in section 3.2.1. In the following section, this dataset is referred to as the boarding dataset. This data contains 1,764,654 (89.60%) boarding, and 201,511 (10.23%) alighting transactions. As is stated, 3,020 (0.17%) transactions are removed in the error elimination process, since their proportion is very small, their quantity is neglected while presenting the results.

All the analyses after the trip chaining results are only performed for the 16 bus routes found competitive to the metro extension. Also, all the analyses are only done for the outbound direction.

5.1. Trip Chaining Algorithm Results

The trip chaining algorithm is applied to the data set in which each cardholder has multiple trips. This data consists of 1,618,964 trip records and in a suitable condition to be fed to the trip chaining algorithm.

5.1.1. Matching Rate

The matching rate is the proportion of trips with an estimated alighting location to the total number of trips. Our trip chain algorithm is able to estimate an alighting location of 1,378,120 (85.12%) trips. Note that, there are no single trips in the input dataset. If the single trips are also considered, the matching rate decreases to 78.10%.

In Figure 5.1, the performance of the developed trip chaining algorithm can be interpreted. Due to the nature of the trip chaining algorithm, 8.26% of the trips cannot be used in the process. Besides, alighting estimation of 3.11% of the trips are failed. There might be several reasons of this issue, such as passengers might have used another mode or transportation that cannot be tracked by smartcard data, e.g., minibuses.

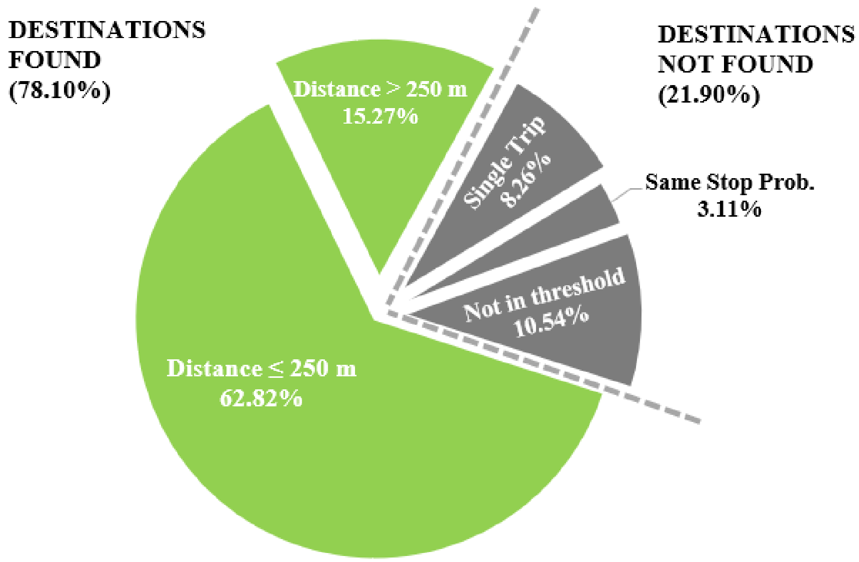


Figure 5.1. Trip chaining algorithm matching results. Distance is the calculated haversine distance between estimated alighting and next trip's boarding stops.

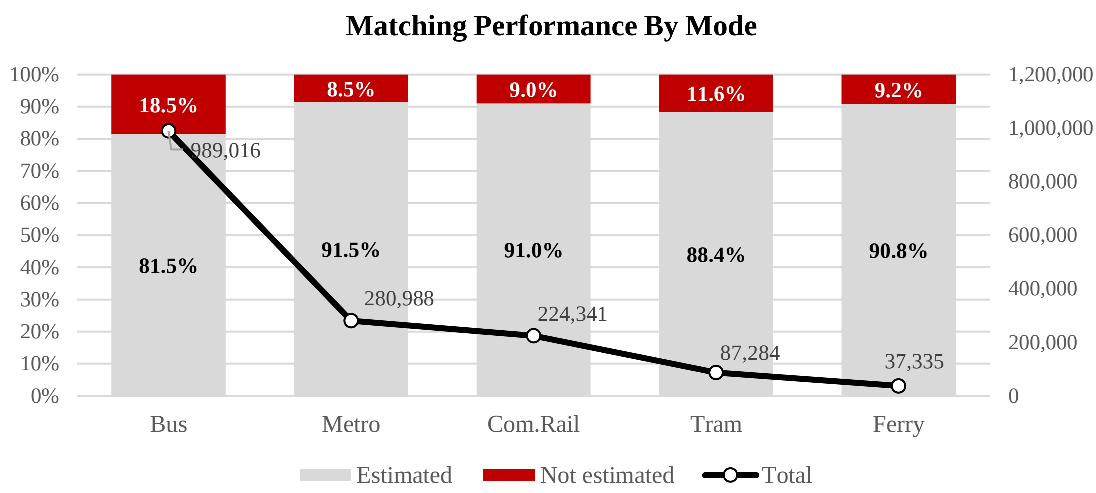


Figure 5.2. Trip chaining matching performance by transit mode.

In Figure 5.3, trip chaining results by the hour of the day can be seen. According to the figure, the percentage of single trips are high on the hours after midnight. This indicates the importance of using at least two days of data. Also, unlinked trips are more likely to be observed due to midnight bus services. Furthermore, the decrease in matching rates during these hours may be attributed to the fact that rail transit services are predominantly unavailable after midnight, and there is a relatively high reliance on bus transit during these hours. If we look at the hours in a day, the proportion of the single trips became constant independent to the number of trips.

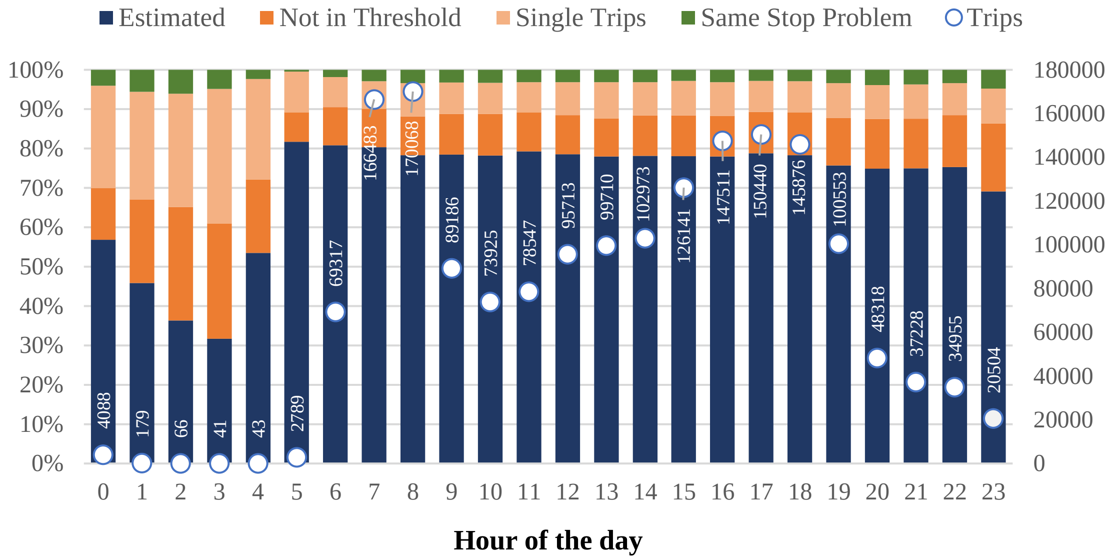


Figure 5.3. Results in percentages by hour of the day.

The matching performance by transit modes can be seen in Figure 5.2. The highest matching rate is established for the metro. Bus transit has the least matching rate with 81.5% estimation performance. In Table 5.1, the percentages are given on the total of the transactions. The numbers clearly indicate that by improving the trip chaining algorithm, there is a potential to increase the mathing rate by 13.65%. However, single trips requires to implement different methods, such as probability and machine learning models, thus historic data.

Table 5.1. Matching rates regarding complete dataset by transit modes.

Mode	**Bus**	**Com. Rail**	**Metro**	**Tram**	**Ferry**	**Total**
Estimated	805,829	204,139	257,064	77,199	33,889	**1,378,120**
% in grand total	*45.66%*	*11.57%*	*14.57%*	*4.37%*	*1.92%*	***78.10%***
Estimation Failed	183,187	20,202	23,924	10,085	3,446	**240,844**
% in grand total	*10.38%*	*1.14%*	*1.36%*	*0.57%*	*0.20%*	***13.65%***
Single Trips	102,340	18,286	15,835	7,156	2,073	**145,690**
% in grand total	*5.80%*	*1.04%*	*0.90%*	*0.41%*	*0.12%*	***8.26%***
Modal Totals	**1,091,356**	**242,627**	**296,823**	**94,440**	**39,408**	**1,764,654**
% in grand total	***61.85%***	***13.75%***	***16.82%***	***5.35%***	***2.23%***	***100%***

5.1.2. OD Matrix

In total, 909,402 journeys from 500,876 cardholders are identified based on the activity threshold. The destination information for 181,645 (20%) of the journeys is not available due to trip chaining algorithm could not find an eligible alighting location regarding walking threshold. Also, based on the applied thresholds, it is obtained that 105,891 cardholders only have one journey in the day. In other words, it is most likely that these passengers have taken one way trip or round trip in a short time. To be able to distinguish that, we can look at the number of one-journey cardholders those alighting location is not estimated. In this case 58,305 (55.06%) of one-journey situations have no estimated alighting location, thus it can be said that these are the one-way journeys where passengers do not return to their first origin of the day.

5.1.3. Validation

In İzmir, all commuter rail system and thirteen bus routes that run through rural areas are operated according to "pay as you go" policy. This means that when boarding these bus routes and commuter rail, passengers are charged with the full fare, and then when they get off the bus, they must tap out in order to receive a refund. Therefore, as stated in section 3.1.2, the tap-out location information is available for 201,511 trips. 192,632 (95.6%) of those are from the commuter rail, the rest 8,879 (4.4%) are the records from the buses. By using the card ID and transaction time information, a dataset consisting of both tap-out and corresponding tap-in transactions is obtained. And then, from this dataset, a joint data set which consists of both estimated and true alighting locations of 168,661 trips is created.

The objective here is to assess the accuracy of the developed trip chaining algorithm. He et al. (2015) defined the accuracy of the algorithm as the ability to find a destination within an acceptable distance from the true destination. They determined the accuracy by calculating the distance between estimated destinations and the true destinations. Their results indicated that the accuracy in the case of perfect matching (0 distance) was 65.8%, and it reached to 85% at 1,000 m tolerance distance (He et al., 2015). In accordance, distribution of the distances between estimated and true alighting locations are obtained and presented in Figure 5.4

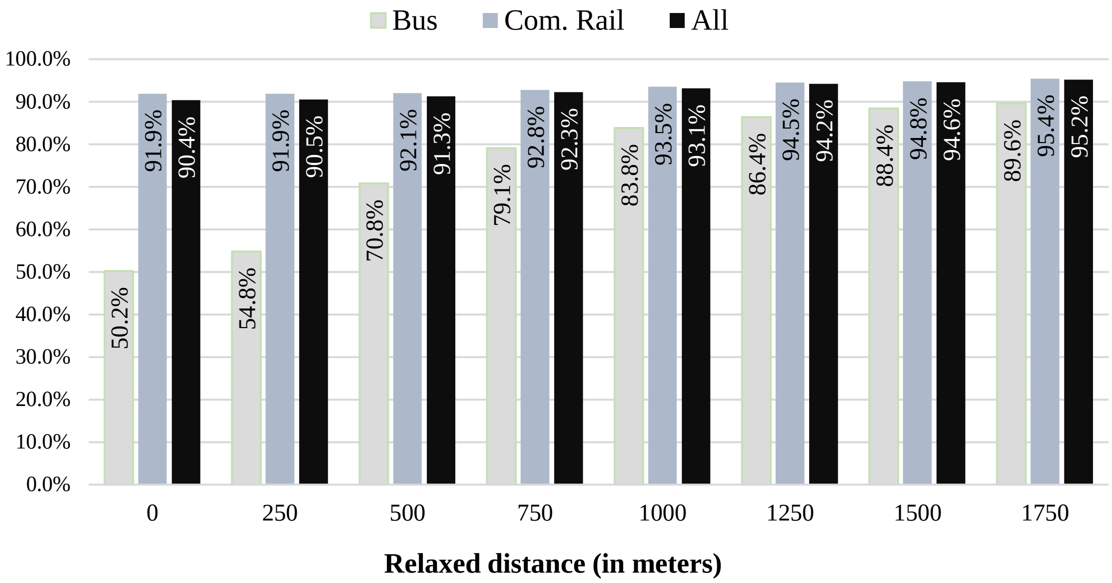


Figure 5.4. The matching percentages based on the amount of relaxed distance between estimated and true alighting locations.

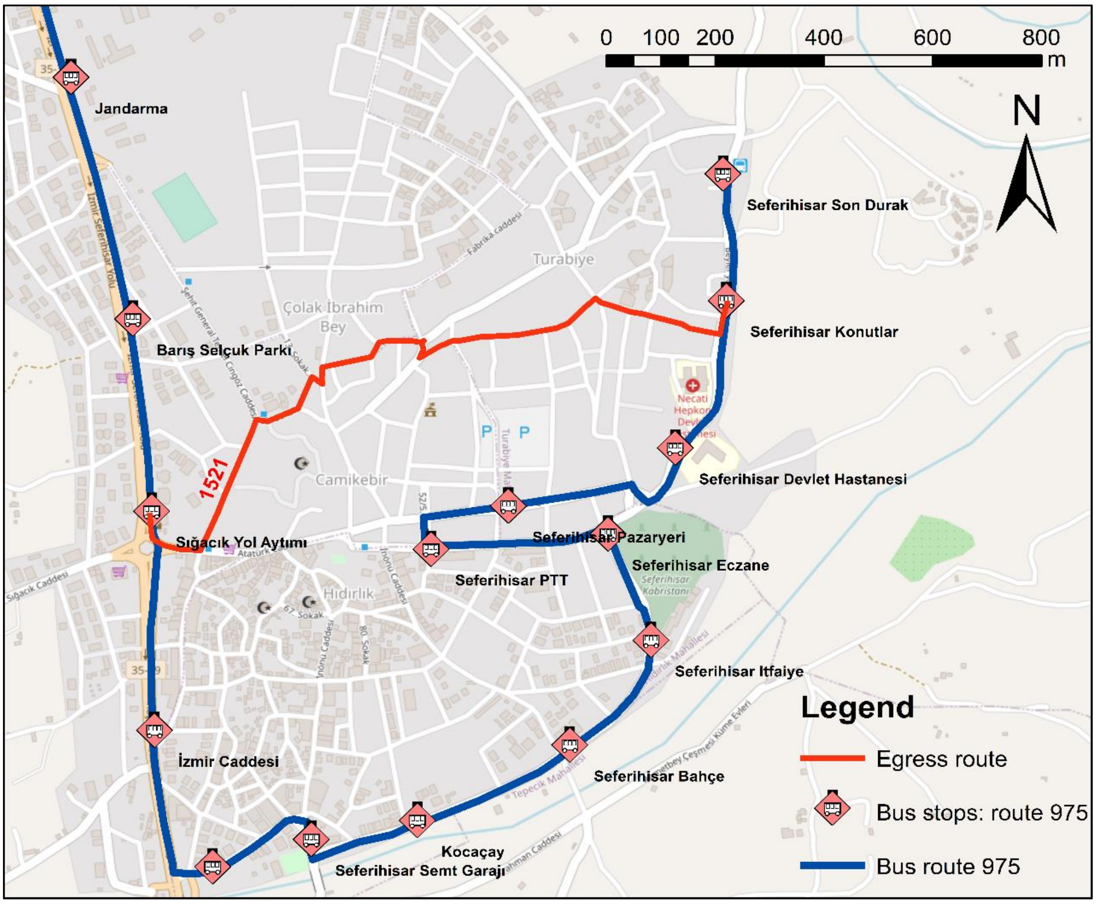


Figure 5.5. A real example for an egress behavior that results in incorrect alighting location estimation.

There might be several arguments to support to consider the relaxed matching instead of perfect matching:

- Passengers may prefer to avoid spending time in congestion and alight several stops earlier.
- Passengers may prefer to walk further than usual and alight earlier or later than the closest stop, i.e., they might not prefer to get off from the PT vehicle at the stop closest to next boarding stop for walking or running errands (see Figure 5.5) especially when they are returning their home or if there are stops close to shopping/entertainment centers (W. Wang et al., 2011).

5.2. Inter-route Relationship Results

As stated in section 4.2.1, our algorithm used to establish inter-route relationships is applicable to any set of station for a given direction; inbound or outbound. As an example, we present the inter-route relationship results obtained for current metro line, Konak tram and planned metro extension in the following paragraphs. The service areas for the analyses are created for each station and their coverage is defined by the 600 meters network walking distance.

As mentioned in section 1.2, there is currently one metro line with 17 stations operating from Fahrettin Altay to Evka-3 station in İzmir. 178 bus stops are found within the defined service areas for outbound direction (see Figure 5.6). Furthermore, 176 bus routes are identified as being in relationship with the metro. However, the competition index results for all the bus routes are found to be lower than 55%. The highest belongs to the bus route 681 with 52.2% competition value, which has 15 stops within the service area of the 7 metro stations, namely Fahrettin Altay, Poligon, Göztepe, Hatay, İzmirspor, Üçyol and Çankaya. In the meantime, Fahrettin Altay, Bornova and Konak stations are found to be most related stations to the bus routes with having 26, 24 and 24 stops within their service areas, respectively.

Similarly, this process is repeated for Konak tram line for outbound direction. There are 19 stations on its route. The rail system’s service area includes 124 bus stops that serve 122 bus routes. Three bus routes are found to have competition index values higher than 55%. These are bus route 10, 253 and 811 having competition index values 76.3%, 63.4% and 57.9% respectively. Remember that bus route 10 is given as an example for the metro competition in section 4.2.2.3. For the Konak tram, this bus route shows competitive features because all of its stops are within the service area of the tram and being related to 10 stations. The most related stations are found to be Fahrettin Altay, Konak İskele and Halkapınar stations.

Finally, considering our objective, the process is also utilized for the determination of the interrelating bus routes with the metro extension for outbound direction. There will be an additional 7 stations with the opening of the metro extension. The analysis is performed on these stations including Fahrettin Altay station. As a result, the algorithm detects 54 bus stops within the service area of the metro extension. Then after, from those stops, 39 bus routes, which are related to the metro extension in some way, are determined. According to analysis, if the current conditions are preserved, there will be several bus routes with relatively high competition (above 70%); bus routes 5, 551, 6 and 305 with competition values of 83.9%, 76.8%, 71.5%, and 70.0%, respectively.

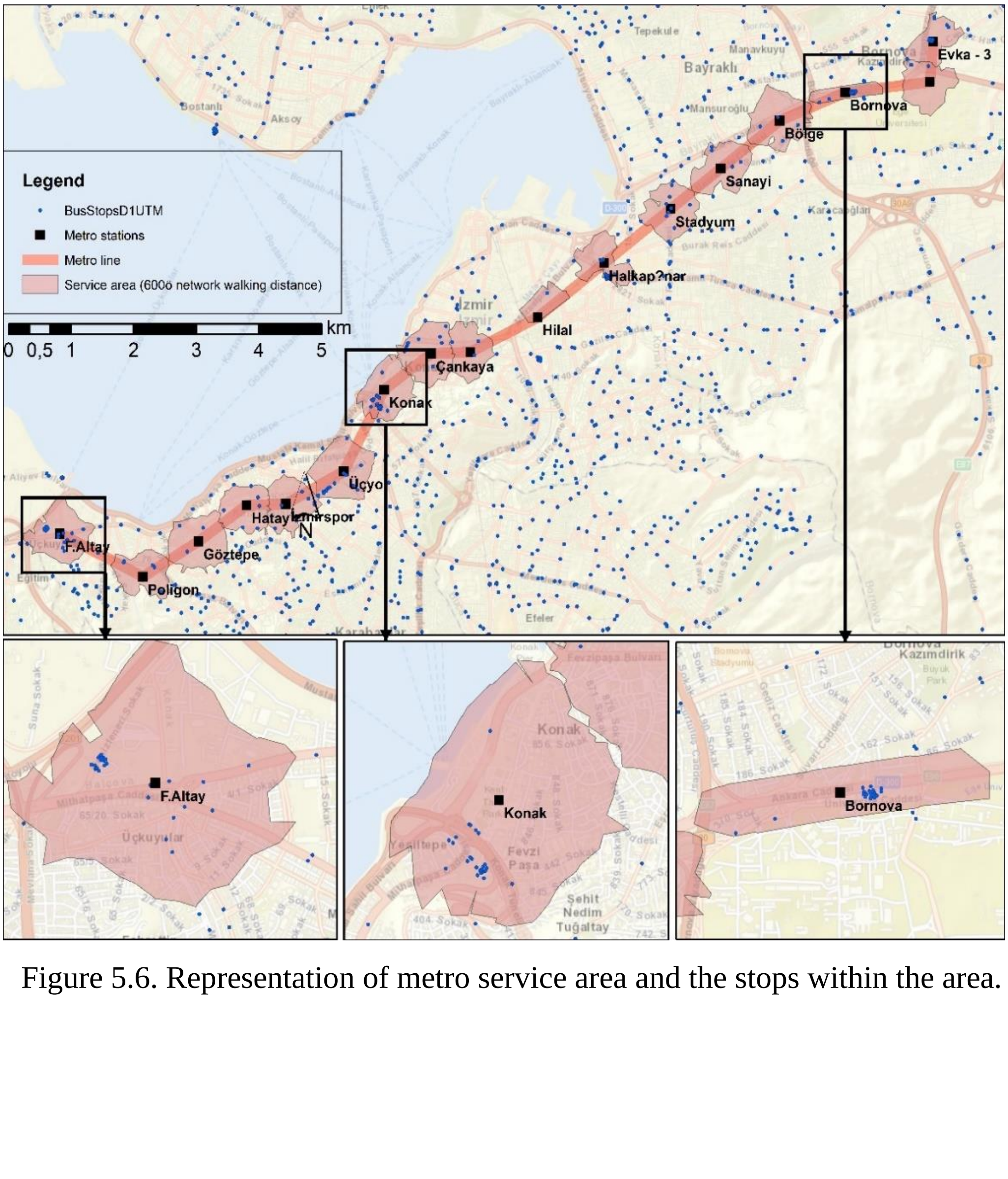


Figure 5.6. Representation of metro service area and the stops within the area.

For a representation of the results obtained from these processes, some of the results are presented in Table 5.2. For the metro and Konak tram lines, there are not any bus routes with a high competitiveness except bus route 10 for the Konak tram. This bus route operates parallel to both the metro and Konak tram line knowingly due to the high elderly demand. In Figure 5.7, the passenger profile of bus route 10 is compared to all bus routes. To do that, we utilize the fare types explained in section 1.2. If we look at the passenger profile of bus route 10 there is a distinct difference where the percentages of elderly and free-of-charge passengers are higher than the average of bus transit, and student usage is very low. This figure also supports the necessity of operating bus route 10 in the current conditions.

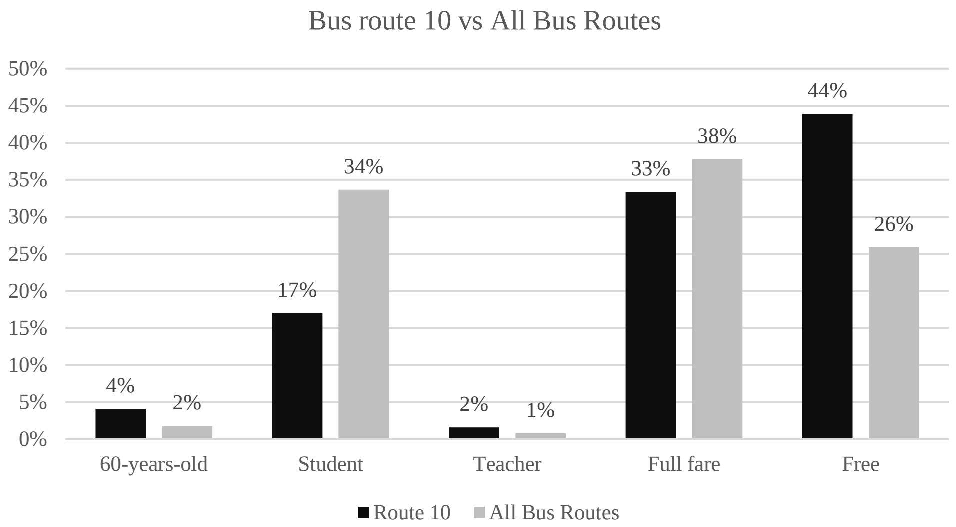


Figure 5.7. Passenger profile of bus route 10 according to fare types.

Table 5.2 also shows that there will be several bus routes highly competitive to the metro extension. Other than the first three competitive bus routes to the metro extension, there are several bus routes similar to 984 which is related to all the stations on the metro extension however they run outside the service area in majority. These bus routes both show competitive and cooperative features. Even though the competitiveness of bus route 984 is found relatively low, some adjustments may be utilized, such as shortening its route.

In conclusion, determination of inter-route relationships may ensure efficient allocation of transit resources by preventing to operate services in competition. In addition, prior knowledge about the competitive bus routes to a rail system may be favorable in case of rail system's disruption. These bus routes can be boosted to satisfy the unmet demand. Also, to pull the sustainable alternative to push the less favored, more complex fare policies can be adopted for the competitive modes.

Table 5.2. Some of the inter-route relationship results for metro, Konak tram and planned metro extension for outbound direction.

System	Route No	# of Stops	# of Stops Within	Related Stations	Order	Competition	Cooperation
Metro Extension	**5**	28	19	8	1	*83.9%*	*16.1%*
	551	28	15	8	2	*76.8%*	*23.2%*
	6	27	15	7	3	*71.5%*	*28.5%*
	984	87	16	8	12	*59.2%*	*40.8%*
	311	26	7	4	20	*38.5%*	*61.5%*
Konak Tram	**10**	19	19	10	1	*76.3%*	*23.7%*
	253	13	11	8	2	*63.4%*	*36.6%*
	811	5	5	3	3	*57.9%*	*42.1%*
	486	16	5	3	10	*23.5%*	*76.5%*
	480	26	4	3	20	*15.6%*	*84.4%*
Current Metro	**681**	23	15	7	1	*53%*	*47%*
	15	10	8	4	2	*52%*	*48%*
	21	9	7	3	3	*48%*	*52%*
	35	20	5	3	10	*21.3%*	*78.7%*
	520	32	5	3	23	*16.6%*	*83.4%*

Note: Order represents the order of the bus route's competitiveness order among other bus routes.

5.3. Passenger Load Profile Results

Passenger load profile is an important input used in the service adjustments in which maximum loading point/segment is required (Pelletier et al., 2011). One of the major compliances in determining transit service is the selection of the most efficient headway, or frequency (vehicle/hour), for each route in the system considering the time of the day and day of the week (Ceder, 2007). These service adjustments are the essence of maintaining adequate service quality and minimizing number of required vehicle runs (Ceder, 2007). Generally, ride check surveys, which are not impractical to conduct frequently on every system, are performed to obtain passenger load profiles. One of the important outcomes of our study is the passenger load profiles that can be created for any bus route. Since the stop level alighting locations are estimated using trip chaining algorithm, we are able to calculate the passenger load between consecutive stops of bus routes for the day of analysis. In addition, since we also determine the bus runs (section 4.3.1), the passenger load of bus routes can be produced at the run level. For each run, the passenger load between stops can be calculated. Then, this information can be used to identify bottlenecks in the service as well as to distinguish the effect of a service change on the ridership as it is done in section 0. Before going into analysis results, it is important to note that there are some passengers whose alighting stop is not estimated by the trip chaining algorithm. To have a complete view and consistency in the results, for those, the alighting location is assumed to be the last stop.

The passenger load profiles are presented at the route and the run level. While presenting the passenger load profiles of bus routes at the run level, we also indicate the seating and theoretical capacity of the vehicle, and the average passenger load on that run. Average passenger load is used to interpret the service quality of the run, and the bus route, in terms of crowding and seating possibilities. In addition, on the horizontal axis, the stop names are indicated in the order of the operation for a given direction. Besides, the vertical axis shows the passenger load, i.e., occupancy, and the load at the corresponding stop is visualized with a stepped line. All the analysis is done considering outbound direction.

The largest population in İzmir resides in Buca district. However, it lacks the advantages of rail transit systems due to poor connectivity. Thus, the transit demand is mainly satisfied by bus transit. Consequently, the ridership on the bus routes serving the Buca district is high. In Figure 5.8, the passenger loads on bus routes, 171, 268 and 878, which are selected giving credits to their high ridership, are visualized. In the figure, passenger load on the buses at each segment, i.e., between consecutive stops, are shown. By looking at the load changes for bus route 171 (green), most of its passengers (about 75%) use this bus to reach Konak, which is a major transfer hub. In addition, there is a stop where the passenger can transfer to commuter rail at Koşu station.

In purpose of presenting the outputs obtained the process described in section 4.3, the bus route 171 is analyzed. In total, 97 bus runs were detected by our algorithm. Based on our calculations the mean, minimum, and maximum running speeds are found 19.46, 26.65 and 15.13 kilometers per hour, respectively. The corresponding running times are calculated as 41.40, 29.95 and 52.75 minutes. The minimum and maximum running times are observed at 10 p.m. and 4 p.m. This indicates a soft spot that the current route of 171 is very prone to traffic congestion. Figure 5.9 gives a clear view about the performance and service quality of the bus route. The x and y axes indicate the stops orders and the run numbers, respectively. For instance, the passenger load in AM peaks is very high after the 14th stop, hence, the service quality is low and most likely there are denied boardings between 14th and 26th stops.

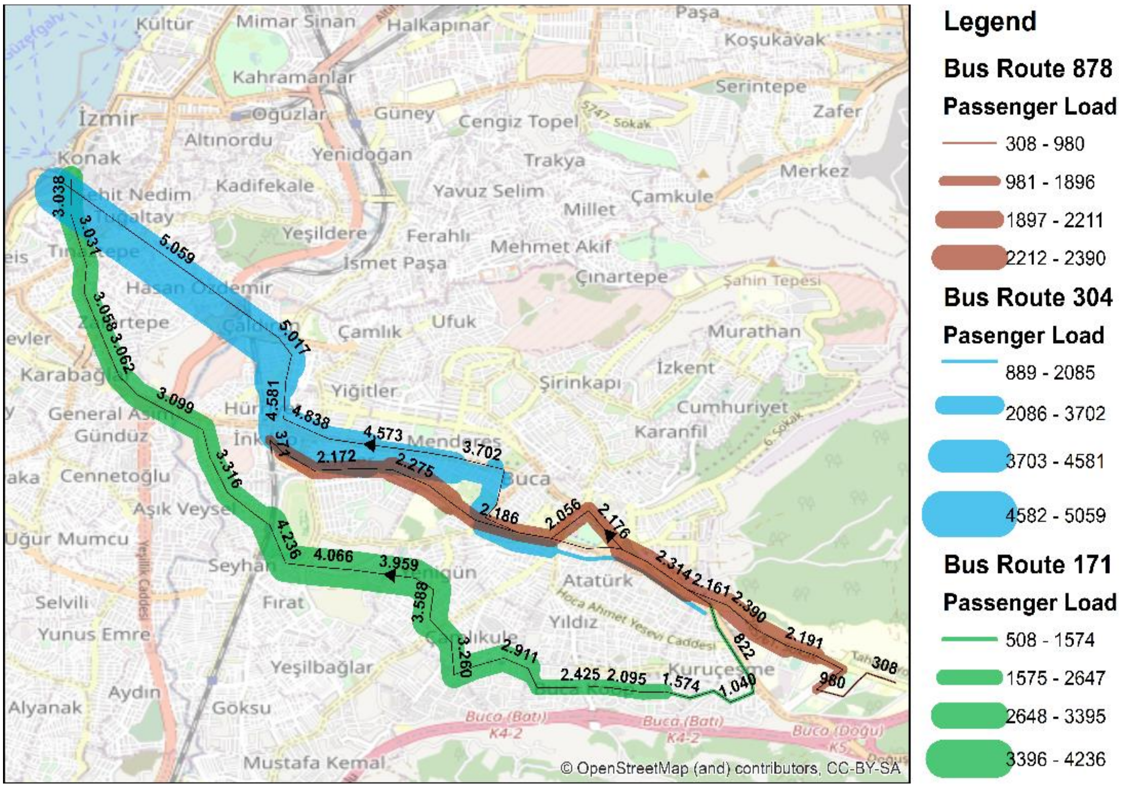


Figure 5.8. Passenger load of bus routes 868, 307 and 171 for the day of analysis on the outbound direction.

Although the passenger load for a day provides an insight into the operational characteristics of the bus route, several issues cannot be captured, such as crowding, denied boardings etc. In Figure 5.10, the passenger load profiles of 14th , 41st and 74th runs are visualized to represent the conditions for morning, afternoon, and evening peaks, respectively. The red and blue dashed lines show the average seating capacity and theoretical capacities. Average seating capacity is the average of the vehicles' seating capacities used to operate the bus route. Theoretical capacity is calculated by multiplying the total capacity (seating plus standing capacities) with 0.75, which is the ratio used in transit planning. The following interpretations may be done:

1. There are small differences in the passenger loads throughout the day for the part of the route which is before 19th stop,
2. For morning and evening peaks, it is most likely that passengers are not able to board between 20th and 26th stops,
3. The 26th stop is a transfer stop, and around 15% of passengers alight at this stop,
4. For the majority of the runs, passengers travel standing after the 14th stop,
5. There is a significant difference between morning and evening runs,
6. There are several stops where the demand is relatively high, for example the 8th and 21st stops.

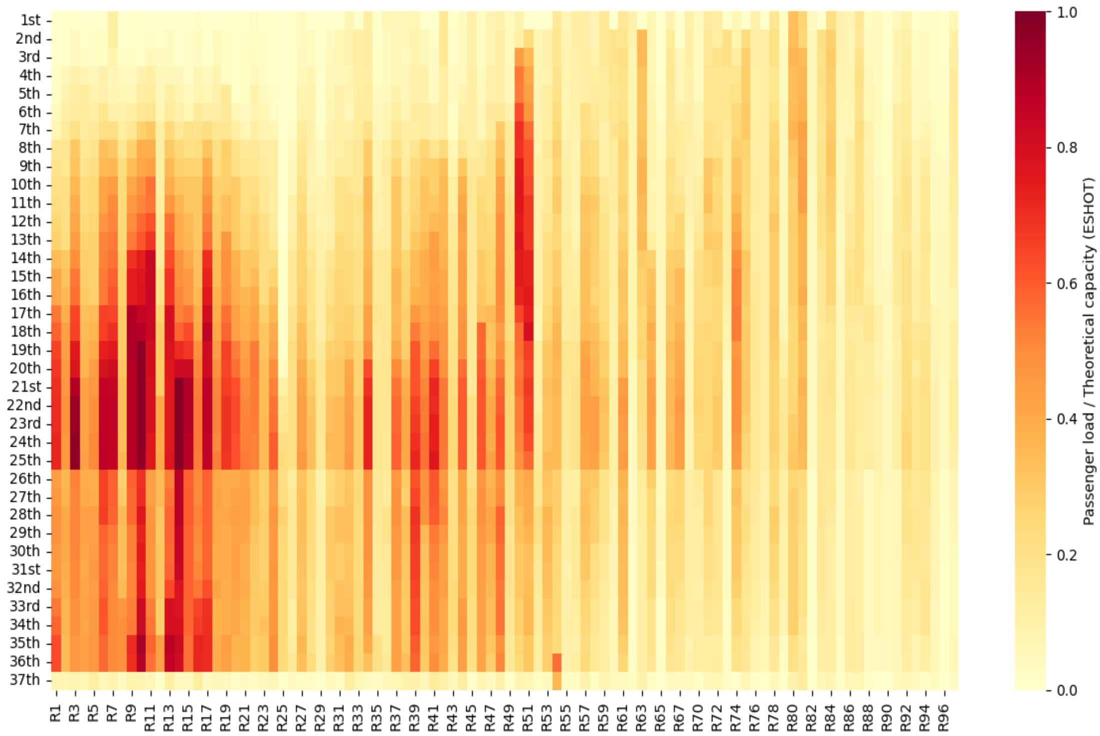


Figure 5.9. Occupancy heatmap of route 171.

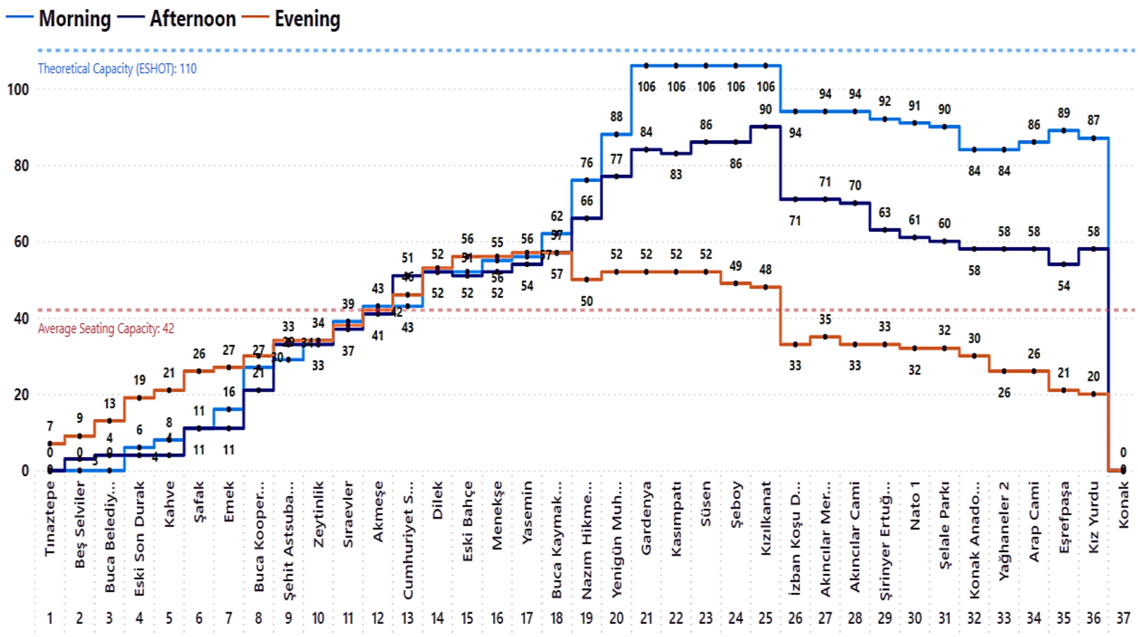


Figure 5.10. Passenger load profile of selected runs for morning, afternoon, and evening peaks for bus route 171.

5.4. Results of Mode Shift Estimation Processes

In this section, the results of the process for estimating the ridership exchange from bus routes to metro extension are presented. The process starts with the identification of targeted bus routes. Then categorization of the passenger flows is performed. After that, travel time components are calculated based on two approaches presented in the previous sections. Finally, the decision of a passenger is estimated, and analyses are presented. Then, on the selected bus routes, the ridership change is further analyzed in the micro level by utilizing the passenger load profiles. This selection process is necessary since it is impractical to present all the results obtained for all the bus routes. Important to note that all the analysis is done for the outbound direction (denoted as 1). First, the results, which are obtained for the bus routes in relationship with the metro extension, are presented. In advance, the process has resulted in 39 bus routes which are found related to metro extension and the results are presented in Table 5.3.

5.4.1. Inter-route Relationships

The bus routes found to have a relationship with the metro extension were identified in section 5.2. The relationship characteristics of all the related bus routes are presented in Table 5.3.

Table 5.3. Competition and cooperition index values of target bus routes.

#	Bus Route	Comp. %	Coop. %	#	Bus Route	Comp. %	Coop. %	#	Bus Route	Comp. %	Coop. %
1	**5**	83.9	16.1	14	**985**	57.9	42.1	27	**883**	8.3	91.7
2	**551**	76.8	23.2	15	**981**	57.2	42.8	28	**945**	7.7	92.3
3	**6**	71.5	28.5	16	**987**	55.9	44.1	29	**873**	6.5	93.5
4	**305**	70.0	30.0	17	**650**	53.6	46.4	30	**24**	6.3	93.8
5	**7**	67.2	32.8	18	**950**	43.8	56.3	31	**969**	6.3	93.8
6	**8**	64.3	35.7	19	**510**	38.8	61.2	32	**486**	6.3	93.8
7	**82**	61.4	38.6	20	**311**	38.5	61.5	33	**10**	5.3	94.7
8	**321**	60.7	39.3	21	**480**	20.2	79.8	34	**202**	4.5	95.5
9	**971**	59.9	40.1	22	**167**	10.0	90.0	35	**671**	4.4	95.6
10	**983**	59.8	40.2	23	**17**	9.4	90.6	36	**681**	4.3	95.7
11	**982**	59.4	40.6	24	**690**	8.8	91.2	37	**946**	3.3	96.7
12	**984**	59.2	40.8	25	**25**	8.3	91.7	38	**517**	2.3	97.7
13	**975**	58.2	41.8	26	**977**	8.3	91.7	39	**879**	2.2	97.8

Evidently, there are 18 bus routes with cooperation index value higher than 90%. Mostly, the current transfer hub Fahrettin Altay station is the only station where these bus routes link to the metro extension. Important to note that the cooperation index does not vary too much, thus the distinction between bus routes is not clear for cooperative bus routes. This indicates the necessity of utilizing additional factors in the cooperation and competition indices.

The objective of identifying the relationship type is based on the conclusion derived from a review of the relevant literature, that is, the exchange of ridership between competing routes is more likely to occur. Therefore, we have decided to focus on the bus routes that have a competitive relationship. The bus routes having a competition index value higher than 55% are considered competitive. According to Table 5.3, the first 16 bus routes are matching with this condition. These bus routes are considered as the main focus group, and further analysis is performed for these bus routes. The general information about the sixteen bus routes is presented in Table 5.4.

Table 5.4. General information about the determined bus routes.

No	**Route No**	**# of Bus Runs**	**# of Transactions**	**Alighting stop is estimated**		**Alighting stop is not estimated**	
1	551	93	4,770	4,476	93.8%	294	6.2%
2	8	71	4,636	4,202	90.6%	434	9.4%
3	975	100	3,710	3,467	93.5%	243	6.5%
4	984	60	3,372	3,146	93.3%	226	6.7%
5	971	69	3,038	2,807	92.4%	231	7.6%
6	82	46	2,157	1,988	92.2%	169	7.8%
7	6	42	1,729	1,616	93.5%	113	6.5%
8	5	42	1,715	1,608	93.8%	107	6.2%
9	7	33	1,511	1,412	93.4%	99	6.6%
10	321	27	1,329	1,241	93.4%	88	6.6%
11	985	18	632	587	92.9%	45	7.1%
12	305	17	590	548	92.9%	42	7.1%
13	983	20	542	514	94.8%	28	5.2%
14	987	12	498	458	92.0%	40	8.0%
15	982	10	385	349	90.6%	36	9.4%
16	981	4	165	159	96.4%	6	3.6%
Total	-	-	30,779	28,578	-	2,201	-

Note: These bus routes are determined based on the competition degrees with the metro extension as explained in section 4.2.1.

There are 30,779 boarding transactions recorded on the 16 targeted bus routes. The alighting location of 28,578 (92.8%) the transactions are estimated using the trip chaining algorithm. This matching rate is above the average of the trip chaining average matching rate (85.12%) when single trips are excluded. This is promising because the negative effect of assuming the passengers, whose alighting location cannot be estimated, alight at the last stop will be minimal on the results.

Figure 5.11 visualizes the inputs used in the inter-route relationship determination process and highlights the characteristics of the competitive bus routes. For example, there are eight stations on the metro extension including the Fahrettin Altay station, which is currently operational, and if a bus route is related to all the eight stations, the green line shows the value of 100%. Thus, even though the percentage of route withing the service area is low for some bus routes, their competitiveness is higher than 50% because they are found to be related to all the stations. The bus routes having short route lengths (lower than 12 km) and high competitiveness values are clearly the ones requiring detailed analysis. A simple classification is made here, and the bus routes, 305, 5, 6, 551, 7 and 971, are classified as urban bus routes. The rest (8, 82, 321, 983, 984, 975, 985, 982, 981 and 987), which have longer route lengths, are categorized as suburban bus routes. The results are presented in detail for the suburban and urban bus routes.

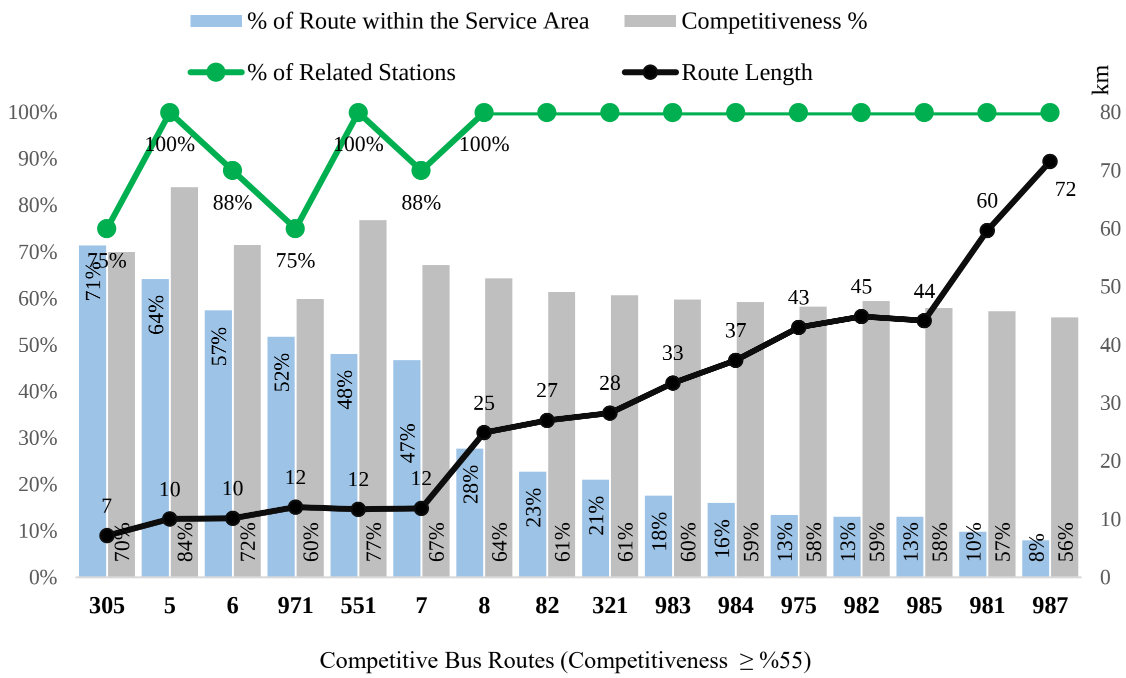
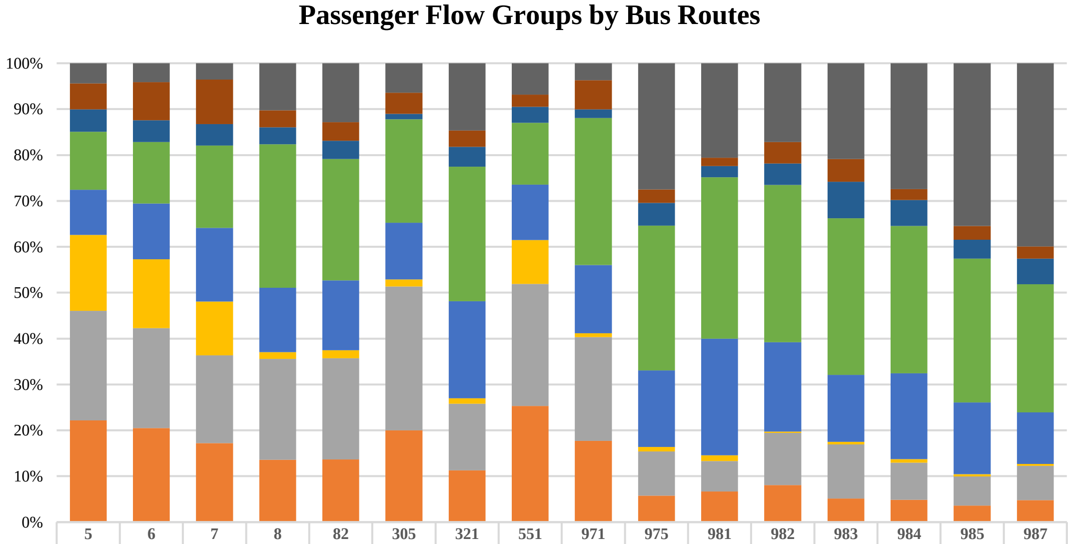


Figure 5.11. Graphical inter-route relationship evaluation of competitive bus routes.

5.4.2. Passenger Flow Group Counts

The mode shift estimation process relies on calculating the travel time savings for passengers in different scenarios tailored to specific passenger flow groups. Therefore, the number of passengers falling into these groups provides insights into the potential magnitude of mode shift that could take place. For this reason, the percentages of the passenger flow groups on the selected bus routes are presented in Figure 5.12. It is clear that the passenger flow groups that will benefit the most from the opening of the metro extension are A-1 and B-1, since these passengers use the bus transit to transfer to the metro in the first place. The groups A-2 and B-2 follow the previously mentioned groups. Important to note that, since there is a transfer stage to or from the metro in groups A-2 and B-2, the proportion of shifting passengers in these groups is greatly dependent on the transfer conditions in the area. Thus, in the convenience approach, the effect of these transfer stages is expected to be captured.



	5	6	7	8	82	305	321	551	971	975	981	982	983	984	985	987
P2	4%	4%	4%	10%	13%	6%	15%	7%	4%	28%	21%	17%	21%	27%	35%	40%
P1	6%	8%	10%	4%	4%	5%	4%	3%	6%	3%	2%	5%	5%	2%	3%	3%
B-3	5%	5%	5%	4%	4%	1%	4%	3%	2%	5%	2%	5%	8%	6%	4%	6%
B-2	13%	13%	18%	31%	26%	23%	29%	13%	32%	32%	35%	34%	34%	32%	31%	28%
B-1	10%	12%	16%	14%	15%	12%	21%	12%	15%	17%	25%	19%	15%	19%	16%	11%
A-3	17%	15%	12%	1%	2%	2%	1%	10%	1%	1%	1%	0%	1%	1%	0%	0%
A-2	24%	22%	19%	22%	22%	31%	15%	27%	23%	10%	7%	11%	12%	8%	6%	7%
A-1	22%	21%	17%	14%	14%	20%	11%	25%	18%	6%	7%	8%	5%	5%	4%	5%

Figure 5.12. Passenger flow group percentages on the bus routes.

5.4.3. Mode Shift

After the determination of the bus routes that are expected to be most affected with the opening of the metro extension, the mode shift estimation process can be performed. Mode shift estimation for the bus transit users is established using two methods. One method, namely deterministic approach (DA), determines whether passengers will shift or not based on the pure time saving which must be at least 10% of their travel time. The second method adds a stochasticity to the estimation process by considering travel convenience, namely convenience approach and denoted as CAMin and CAMax. In this method the minimum, and maximum time multipliers are used for CAMin and CAMax, respectively, to observe the range in the number of passengers shifting to metro. The mode shift estimation results for bus routes considering different estimation methods are given in Figure 5.13.

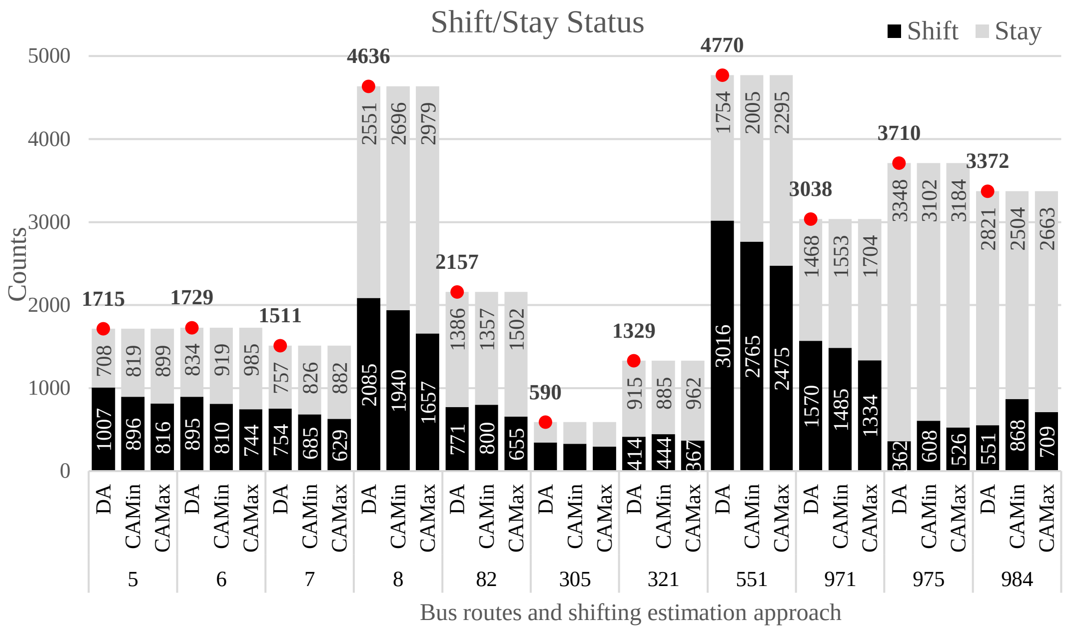
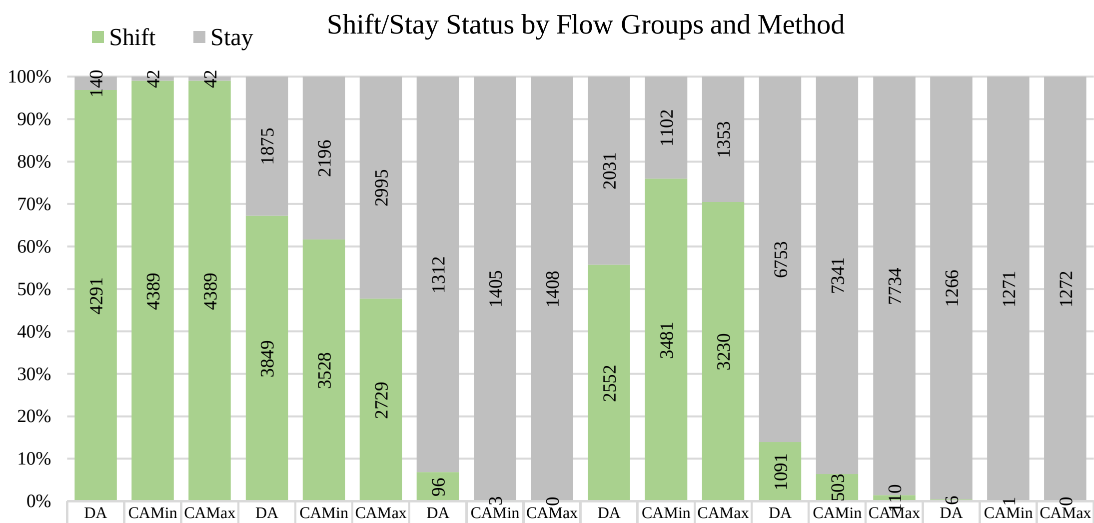


Figure 5.13. Shift and stay status of bus routes in numbers by applied methods.

According to Figure 5.13, about half of the passengers that used the bus routes 5, 6, 7 and 8 are found shifting to the metro after the opening of the metro extension. The deterministic approach represents the evaluation condition where most passengers shift to the metro for the majority of the bus routes. The CAMin results in higher number of shifts on the bus routes where passenger travel long distances in a crowd condition, such as 984 and 975.

If we look at the number of passengers shifting to the metro regarding the passenger flow groups and mode shift estimation method (see Figure 5.14), almost all the passengers in group A-1 are shifting to the metro. It appears that a very small number of passengers in group A-1 benefit the most from continuing to use the bus. The reason is that the median travel distance of the passengers, who are found to be not shifting, are 2.9, 2.0 and 2.2 km for the methods DA, CAMin and CAMax, respectively. So, the advantages of using the metro cannot aggregate on these passengers travelling such short distances. As expected, A-3 shows the least amount of shifting to the metro since the passengers who fall in this group are required to transfer twice in the alternative scenario. On the other hand, the convenience approaches result in lesser number of shifting passengers in groups except A-1 and B-1. In the case of A-1, the crowding conditions are in play, while the determinant is the transfer conditions for B-1. Besides, a classification based on the effort required to convince a passenger to shift to the metro considering the flow groups can be made by looking at the numbers in Figure 5.14. For the passengers grouped under P1, P2, A-3, B-2, and B-3, using the metro is either not possible or not feasible, hence they are labeled as “impractical” as in Table 5.5 .



	A-1			A-2			A-3			B-1			B-2			B-3		
	DA	CAMin	CAMax	DA	CAMin	CAMax	DA	CAMin	CAMax	DA	CAMin	CAMax	DA	CAMin	CAMax	DA	CAMin	CAMax
Stay	140	42	42	1875	2196	2995	1312	1405	1408	2031	1102	1353	6753	7341	7734	1266	1271	1272
Shift	4291	4389	4389	3849	3528	2729	96	3	0	2552	3481	3230	1091	503	110	6	1	0

Figure 5.14. Shift and stay proportions by passenger flow groups and mode shift estimation methods.

The aggregated results are presented in Table 5.5 for the 16 bus routes. As is evident, the demand for these bus routes is expected to decrease 30 to 55 percent. This indicates the importance of preparation on the service adjustments prior commencing the metro extension. In addition, based on our analysis, consideration of travel convenience to estimate the passengers’ decision has an effect on the results about 5-10%. This points out the importance of valuing passengers’ perception in transportation planning.

Table 5.5. The mode shift estimation results by methods for the bus routes.

Result Boundary	**Method**	**Shift**		**Stay**				
		Total	%	Total	Impractical (% in total)		Potential (% in total)	
All the 16 bus routes	DA	11,885	*38.6%*	18,894	8,095	*26.3%*	10,799	*35.1%* ↑
	CAMin	11,905	*38.7%*	18,874	8,193	*26.6%*	10,681	*34.7%* ↑
	CAMax	10,458	*34.0%*	20,321	8,197	*26.6%*	12,124	*39.4%* ↑
Urban bus routes *(5, 6, 7, 305, 971 and 551)*	DA	7,587	*56.8%*	5,766	2,990	*22.4%*	2,776	*20.8%* ↑
	CAMin	6,970	*52.2%*	6,383	3,088	*23.1%*	3,295	*24.7%* ↑
	CAMax	6,292	*47.1%*	7,061	3,091	*23.1%*	3,970	*29.7%* ↑
Suburban bus routes *(8, 82, 321, 983, 984, 975, 985, 982, 981 and 987)*	DA	4,298	*24.7%*	13,128	5,105	*29.3%*	8,023	*46.0%* ↑
	CAMin	4,935	*28.3%*	12,491	5,105	*29.3%*	7,386	*42.4%* ↑
	CAMax	4,166	*23.9%*	13,260	5,106	*29.3%*	8,154	*46.8%* ↑

Note: We consider the passengers found not shifting to the metro and fall into groups A-1, A-2, B-1, and B-2 as potential non-shifters. On the other hand, passengers in groups P1, P2, A-3, and B-3 are considered impractical non-shifters. This classification is based on the average travel time savings required for the non-shifting passengers in the groups (Figure 6.8).

5.4.4. Passenger Load Profiles

Passenger load profiles are used to visualize the effect of opening the metro on the competitive bus routes. This output is valuable in that it gives deeper insight on the operational perspective at micro level. In the next paragraphs the passenger load profiles of bus routes 5 and 984 are analyzed.

Since the number of runs for a bus route can go high as 100, it is also necessary to select representative bus runs. To do that, the maximum and average passenger load, i.e., average occupancy, values at each run are considered for selecting the representative bus runs for morning (AM) and evening (PM) peaks, and off-peak hours. The AM and PM peaks are considered as the hours between 7 and 9 a.m. and 4 and 7 p.m., respectively. The rest of the hours are considered off-peak hours. For each time segment, the bus run with the highest mean and maximum occupancies is selected between representative runs. In Figure 5.15, maximum and average occupancy levels are presented for bus route 5. Red bars represent the runs having the one-stop boarding problem. Runs with green colors selected as representatives. In the majority, vehicles with 26 seating capacity are used at the runs of bus route 5. The corresponding theoretical capacity is 66 passengers. The mean occupancy for the day of analysis is found 15 passengers considering all runs except the ones with one-stop boarding problem. The figure also indicates there is no clear peak and off-peak working conditions, but the evening usage seems high. In the end, the runs numbered 11, 20, 30, and 31 are selected as representatives. Two bus runs (30^{th} and 31^{st} runs) are selected for PM peak because the flow characteristics are found to be worth analyzing. This process is applied for the other two bus routes to be analyzed.

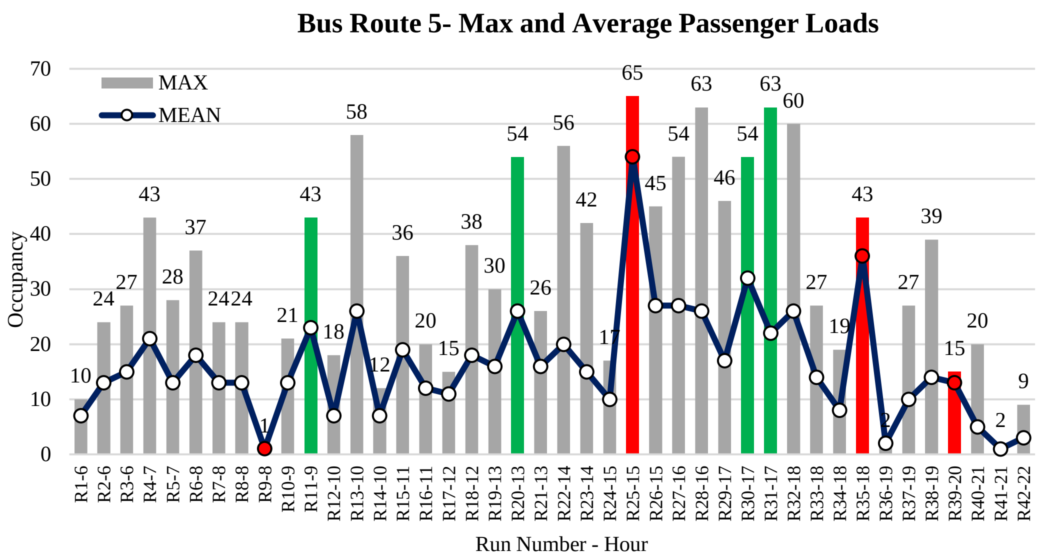


Figure 5.15. Maximum and average passenger load values of bus route 5.

The passenger load profiles of bus route 5's 11^{th} run for current, deterministic (DA), minimum (CAMin), and maximum (CAMax) convenience conditions are presented in Figure 5.16. Valuable interpretations can be made by analyzing the load profiles:

1. In the best scenario, nearly 60% of the passengers are shifting to the metro, as a result, decreasing the average passenger load of 11^{th} run from 22 to 9 passengers,
2. Passengers do not transfer to the metro at the earliest station due to long transfer walks, instead they transfer at the 13^{th} stop where the transfer distance is 150 meters,
3. Travel convenience factors, such as transfer penalties and in-station walking, are affective and reduce the number of passenger shifts. For passengers who do not switch routes after considering the minimum travel convenience components, the average time saved is 4 minutes for the deterministic

approach (DA), while there is an average time decrease of -3.6 minutes for the minimum convenience approach (CAMin).

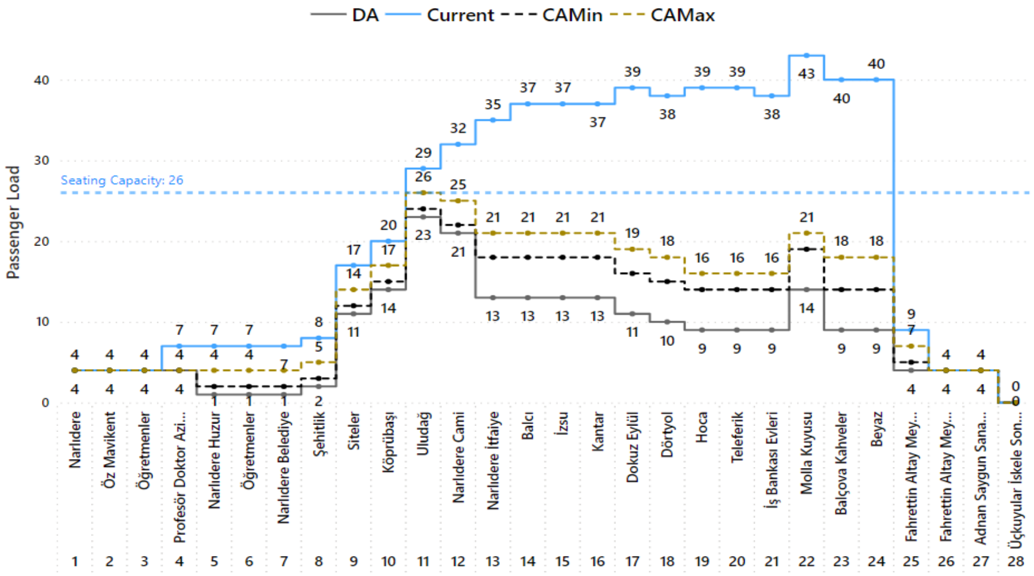


Figure 5.16. Passenger load profile comparison of the 11th run of bus route 5.

The load profiles are very similar for the 20th and 30th runs (Figure 5.17 and Figure 5.18), indicating that the passenger flow characteristics do not vary to much between morning and afternoon hours. However, a transient demand peak is occurred on 31st run due to students' mobility at the 17th stop which is close to a university (Figure 5.19).

For bus route 984, the 17th , 32nd and 41st runs are presented. The length of the bus route is 72 km, and it has 87 stops on its route. For this reason, only the last 33 stops are shown in the load figures. Also, the mode shift may only occur after the 67th, which is the first stop falls within the service area of the metro extension. It is clear in Figure 5.20 that the bus runs in a crowding condition for many of stops. After the metro opens, nearly 40% of its passengers will shift to the metro. There are some passengers who won't shift to the metro even though they are saving time. This is due to their high travel times, so the amount of time saved should also be high. Besides, max and min convenience coefficients are only significant if transfer activities are present. Otherwise, mode shift behavior overlaps regardless of using max or min time multipliers.

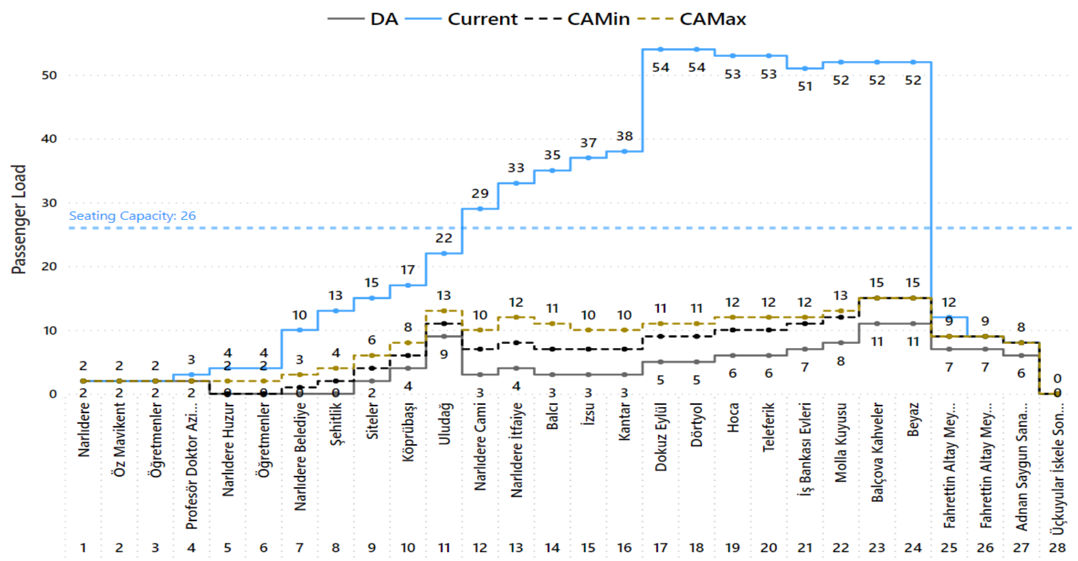


Figure 5.17. Passenger load profile comparison of the 20th run of bus route 5.

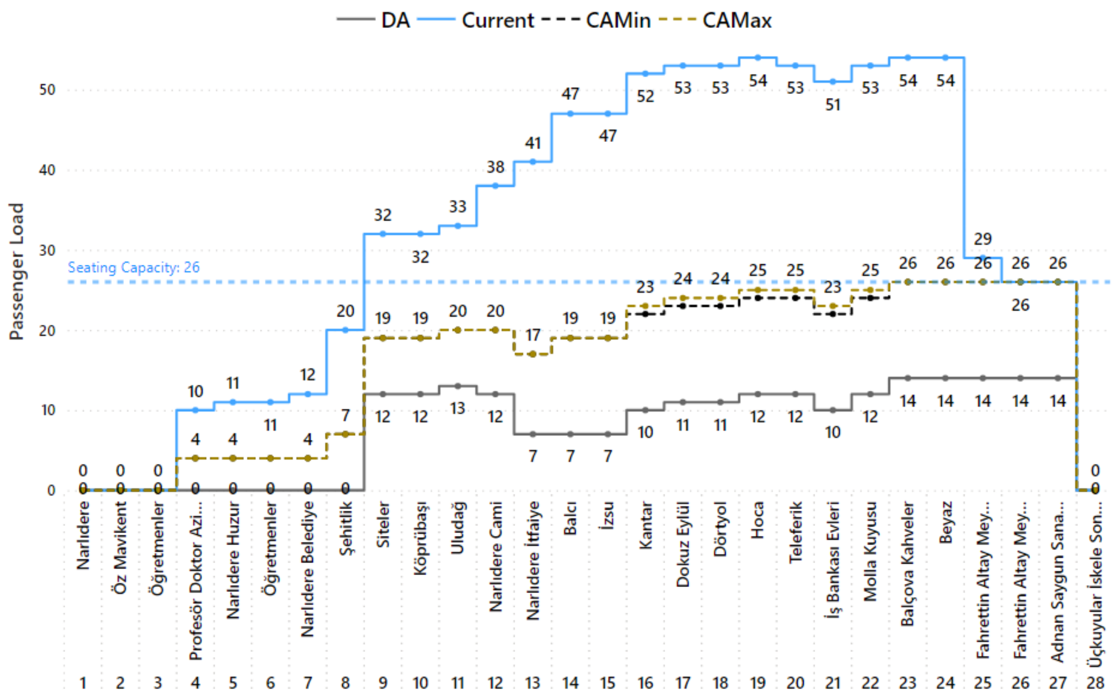


Figure 5.18. Passenger load profile comparison of the 30th run of bus route 5.

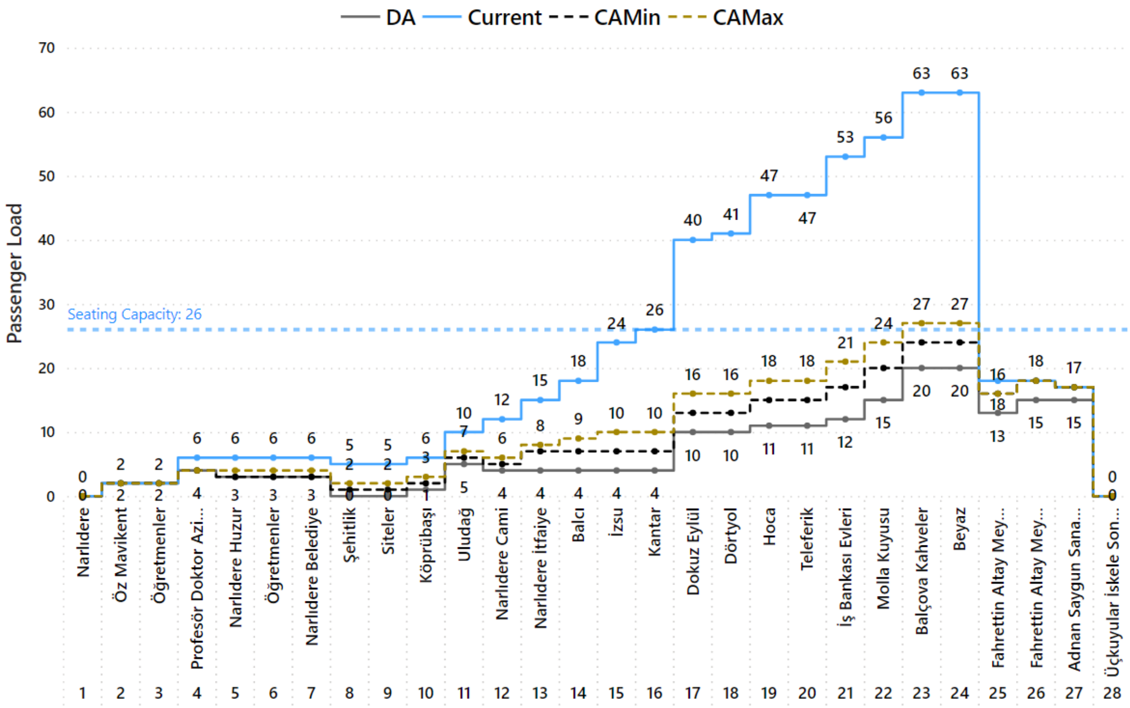


Figure 5.19. Passenger load profile comparison of the 31st run of bus route 5.

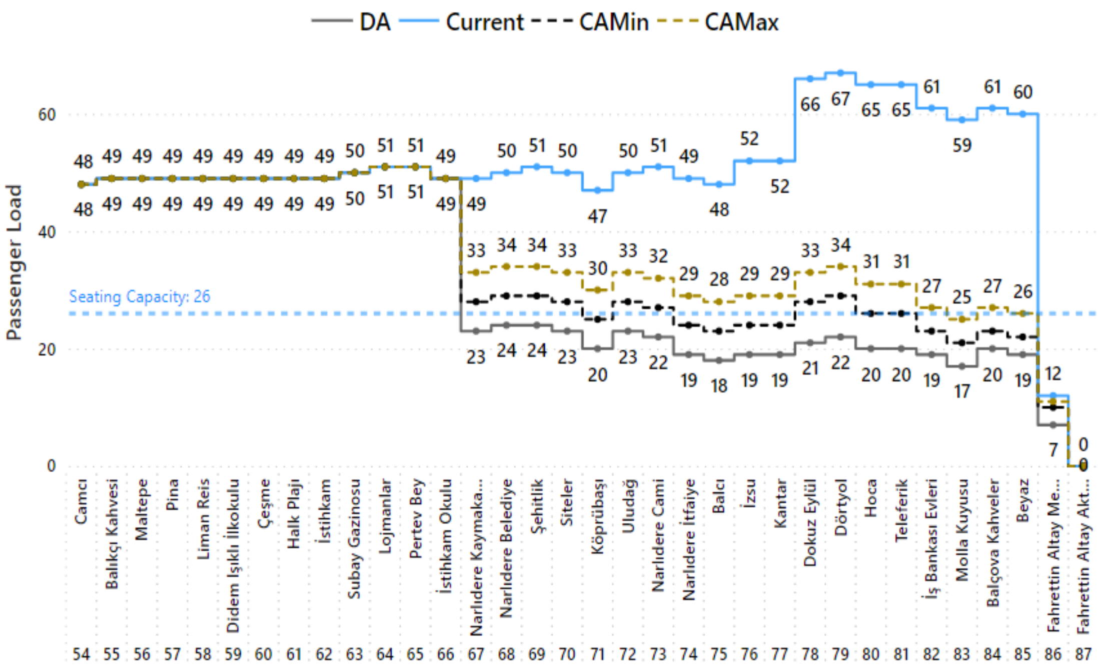


Figure 5.20. Passenger load profile comparison of the 17th run of bus route 984.

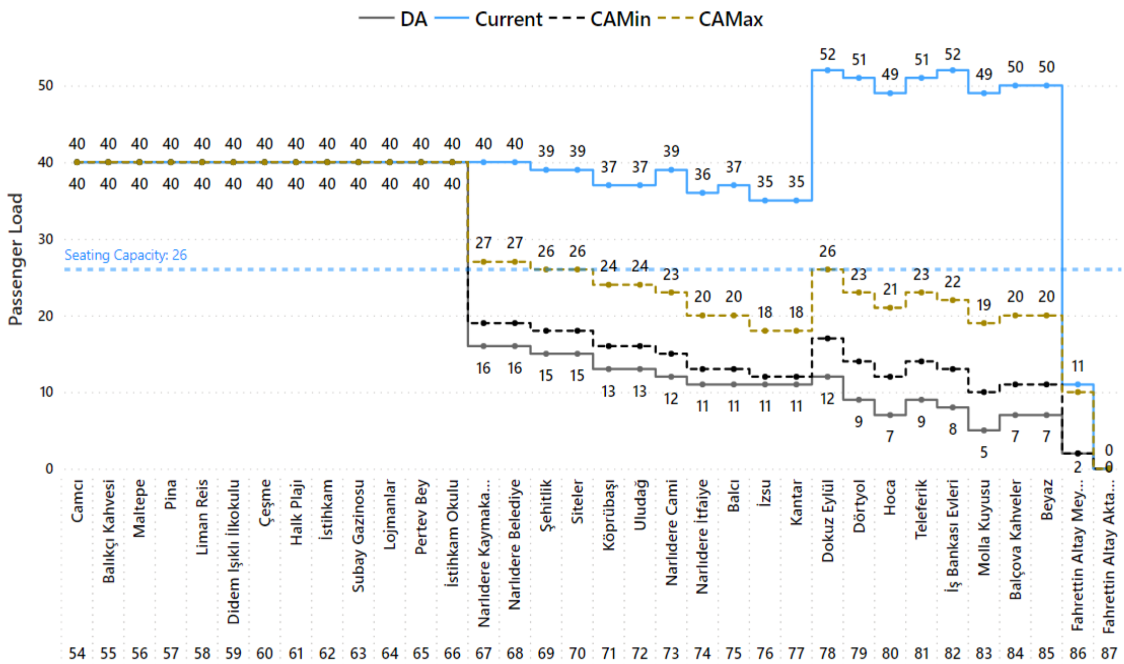


Figure 5.21. Passenger load profile comparison of the 32nd run of bus route 984.

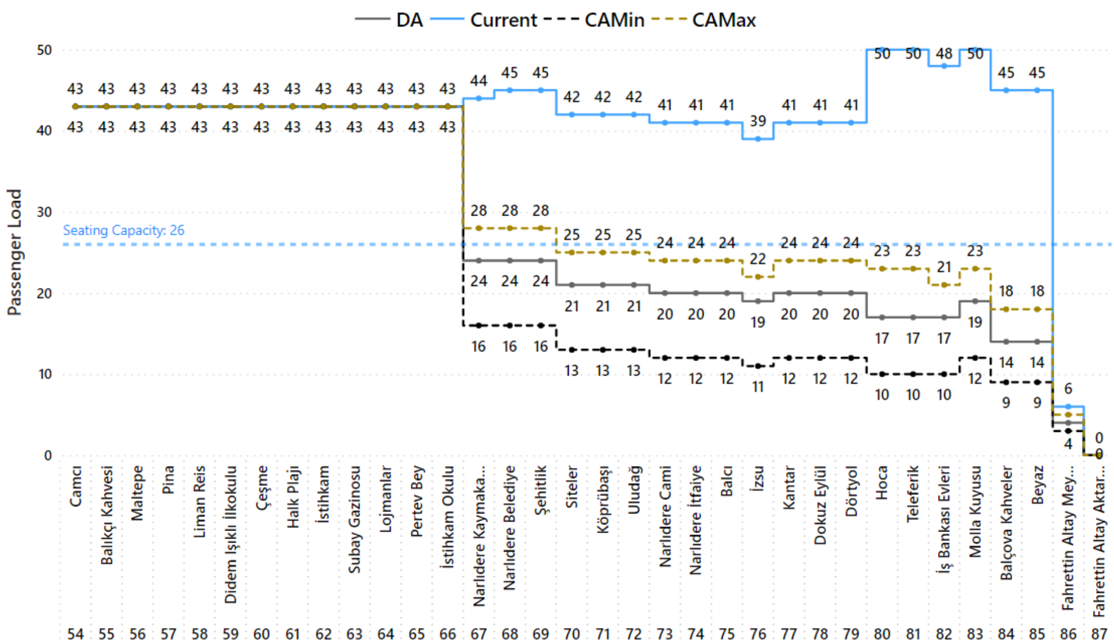


Figure 5.22. Passenger load profile comparison of the 41st run of bus route 984.

5.4.5. Urban and Suburban Bus Routes

Selected sixteen bus routes are classified as urban and suburban considering their route lengths based on Figure 5.11. We present the results based on this classification in order to have further insights.

We considered the bus routes 8, 82, 321, 975, 981, 982, 983, 984, 985 and 987 as suburban bus routes. These bus routes have a significant amount of route length outside of the service area of the metro extension and expected to have P1 and P2 passenger groups in higher proportions (see Figure 5.12). There are 17,426 (out of 30,779) passengers on these bus routes. The number of passengers fall into groups P1 and P2 is 4,104. Meaning that it is either impractical or not feasible for these passengers to use the metro. The majority of the non-shifting passengers are the ones in group B-2 (28.52%) (see Figure 5.23). Their average bus travel distance (actual trip distance) is 22.7 km and the average metro in-vehicle time in the alternative scenarios of these passengers is 7.6 minutes (approximately 5km). In addition, the average travel time savings are -3.36 and -11.72 minutes considering DA and CAMin methods. This means that these passengers start their trips far from the closest metro station on their route and forcing them to transfer to the metro causes some amount of travel time loss. The passengers in groups B-2 and B-1 can be avertable by enhancing transfer conditions since the required travel time improvements, which are -3.36 and -0.96 minutes on average for DA method, are relatively low.

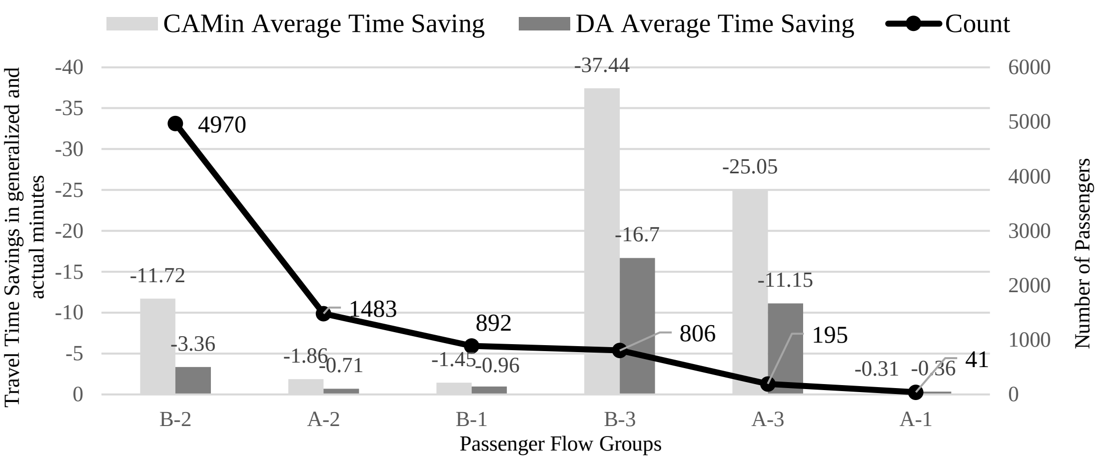


Figure 5.23: Suburban bus routes; travel time savings of non-shifting passengers by groups.

For the urban bus routes, the number of non-shifting passengers falling into group A-3 is relatively higher according to Figure 5.24. These passengers are ending their trips outside the metro service area, for example the passengers alighting at a stop near the ferry station, Üçkuyular İskele. The average travel time savings

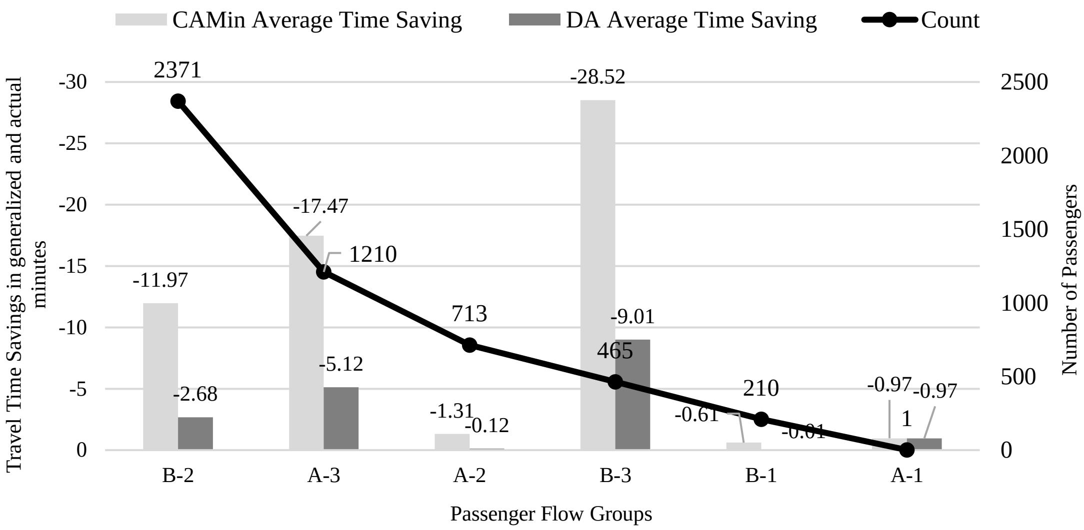


Figure 5.24: Urban bus routes; travel time savings of non-shifting passengers by groups.

CHAPTER 6

DISCUSSION & CONCLUSION

The present study aims to shed light on the effect of a major service change (extending a metro line) on the current bus transit system. In the previous sections, we have examined the influence of travel time calculated using the deterministic and travel convenience approaches to estimate the passenger's mode shifting patterns between competitive bus routes and the metro. In the literature, it is commonly acknowledged that following the implementation of a rail system, transit authorities typically prioritize the adjustment of bus transit within the area to align with the goal of optimizing travel demand and promoting the use of the rail system (Brands et al., 2022; Gao et al., 2022b). However, it is crucial to ensure that this adjustment process does not lead to passenger dissatisfaction. Thus, the focus of this section is to identify potential areas within the transit network where issues may arise and prioritize improvements accordingly. First, the shifting behavior is discussed considering the travel distance of passengers. Then, the effect of the transfer walking on the passengers' estimated decisions is established. Then, the spatial evaluation of the candidate transfer bus stops and the change in the travel time of passengers after commencing the metro extension is discussed. Following that, we comment on the possible adjustment strategies considering the presented findings for competitive urban and suburban bus routes. The discussion section is finalized presenting the comments on the applied travel time calculation approaches and the sensitivity of the results to the assumptions.

6.1. Shifting Behavior by Travel Distance

Some of the advantages of the two alternative modes may change depending on the passengers' travel distance. Passengers tend to prefer metro for long distances and utilize buses for short travels within the metro influence area (Gadepalli et al., 2022). In Figure 6.1, the travel distances of the passengers on the targeted bus routes are visualized along with the proportion of their decisions. In order to gain insights into the mode shift behavior of bus passengers along the parallel section with the extended metro line, we have excluded passengers whose origin locations are located far from the metro stations (beyond a distance of 3-4 km). So, the origin of the passengers using these bus routes are either within or very close to the metro stations. Thus, we will be able to see which mode is preferred for which travel distance intervals based on our approaches. Figure 6.1 clearly shows that bus transit is mostly preferred for travel distances shorter than 2.5 km. For the travel distances above this value, shifting behavior towards the metro increases. It seems the benefits of using metro will be more enhanced after 3-4 km which is in line with the findings in literature (Gadepalli et al., 2022). For travel distances higher than the length of the metro extension, buses are becoming more preferred. This is due to the deterrence of transfer activities coming into play. These results are similar to findings in literature, meaning that our assumptions are capable of representing the expected behavior of passengers.

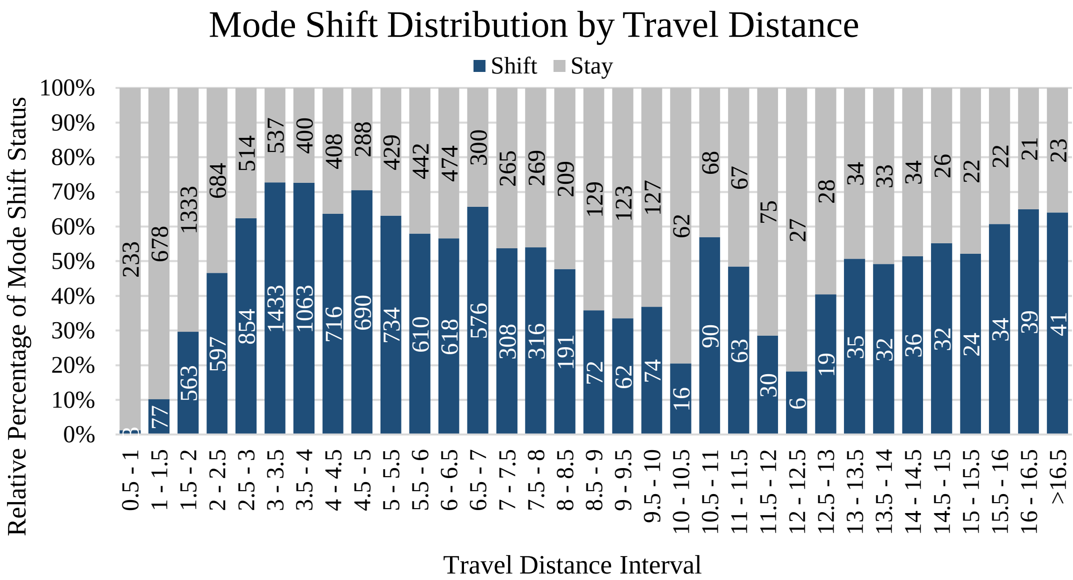


Figure 6.1. Passengers’ mode shift pattern regarding their travel distances.

6.2. Transfer Walking Distances in the Area

We have determined the transfer walking distances, which are the network walking distances between stops to stations that are the closest, in section 4.6.1.5. Hence, these transfer walking distances can be used in purpose of understanding the transfer conditions from bus to metro considering the current location of bus stops. This is important due their closeness directly affects the integration efficiency between modes through the effect on travel time convenience. The distribution of transfer walking distances and average walking distances by stations are visualized in Figure 6.2.

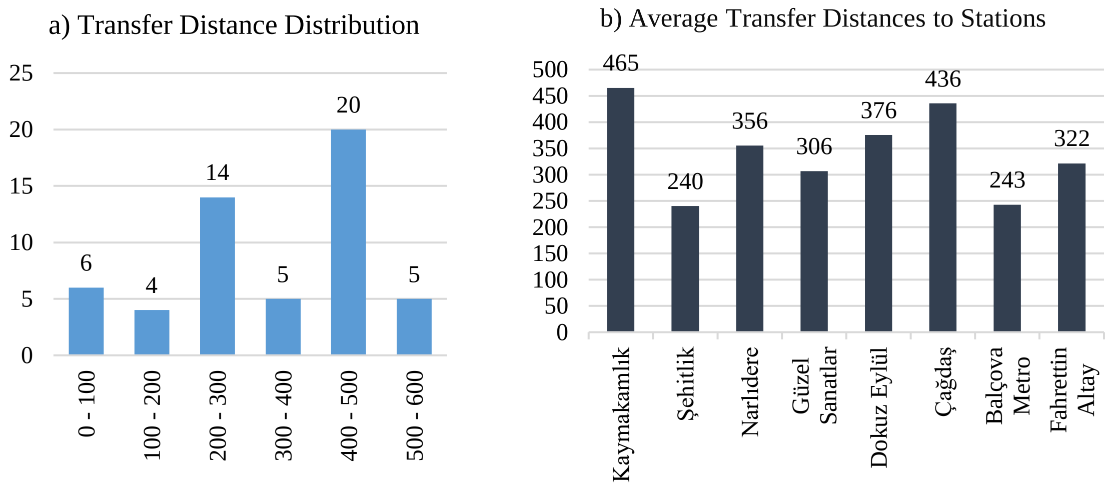


Figure 6.2. a) Shows the number of stops regarding the transfer walking distance (to closest metro station) intervals b) Average transfer walking distances by station.

To provide further insights, we delve into the question of the potential number of passengers who might consider shifting their mode of transportation if the transfer walking distances in the area were reduced. For that, the passengers estimated not shifting the metro based on CAMin method (13,357 passengers) are collected. Note that passengers fall into P1 and P2 flow groups are excluded. The distribution these non-shifting passengers' travel time savings in generalized time minutes according to CAMin method is presented Figure 6.3. A total of 6,030 passengers, who require at most 10 minutes of improvement in terms of generalized time, have the potential to exhibit a preference for transitioning to the metro system if transfer conditions are adequately enhanced. This can be achieved through measures such as enhancing the convenience of in-station walking times or optimizing transfer stops to reduce the distance passengers need to walk during transfers. This finding holds significance as it suggests that approximately 45% of passengers who initially do not consider shifting modes could be attracted by solely implementing strategic adjustments.

Furthermore, a sensitivity analysis is performed in terms reduction in minutes of transfer and in-station walking times in Table 6.1. Since the transfer walking time multiplier is 1.68 in CAMin method, 1 minute of reduction in transfer or in-station walking time decreases 1.68 generalized time minutes in the generalized travel time. For passengers who have transfer walking activity (10,000 of 13,357), the average transfer walking time is 3.81 minutes. For instance, by adjusting the location of transfer stops to decrease the walking distance about 10% of the non-shifters can be attracted to use the metro. On the other hand, improving the passengers in-station walking time perception by increasing the comfort and attractiveness of metro stations, such as providing escalators and shops etc., up to 32% of non-shifting passengers may prefer to shift to the metro.

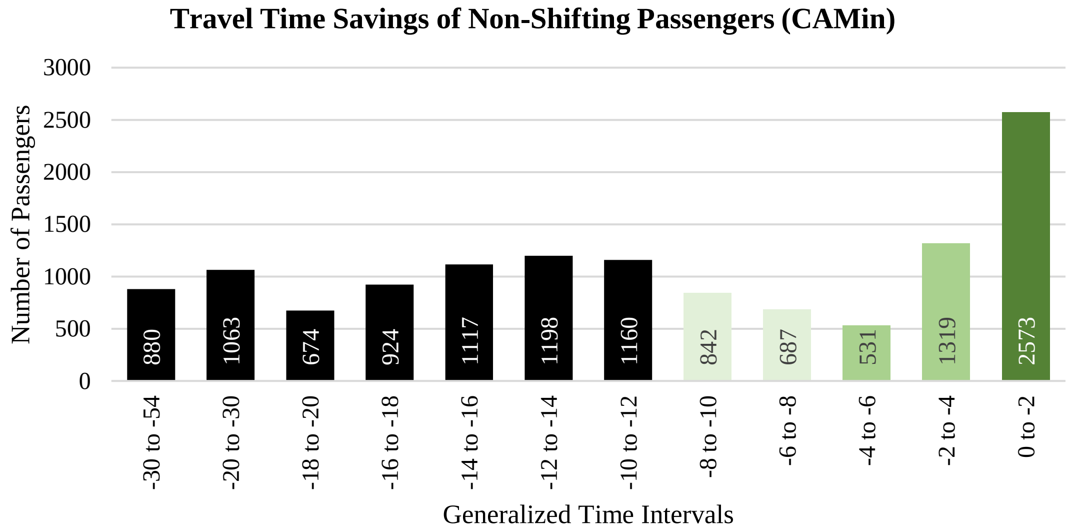


Figure 6.3. Number of non-shifting passengers by generalized time savings under CAMin method.

6.3. Transfer Stops

In Figure 6.4, the pie charts are the locations of bus stops, which are either the transfer stop in the alternative scenarios (for groups B-1, B-2, and B-3) or the boarding stop of the actual trip (for groups A-1, A-2, and A-3). The yellow color represents the amount of passengers who use the particular bus stop to transfer to the metro in their alternative scenarios. On the other hand, blue color is used to represent the amount of passengers who do not need to have a transfer to shift to the metro. As is evident from Figure 6.4, certain bus stops will serve as crucial transfer points after commencing the metro extension. There seems to be several stations prioritized by the passengers of the competitive bus routes to transfer to the metro. As expected, the first station of the extension is highly utilized to transfer to the metro by the passengers of suburban bus routes. The intermediate stations are mainly used by the passengers of the bus routes serving the outskirts of Narlıdere and Balçova districts, such as bus route 5, 6, 7, 305 and 971. Indicating that there will be several minor transfer stations.

Table 6.1. The change in number of non-shifting passengers by decrease in the transfer and in-station walking times (CAMin method).

Time Component	**Reduction (minutes)**	**# of new shifters**	**% of non-shifting passengers (13,357)**
TWT	0.5	98	0.73%
	1.0	208	1.56%
	1.5	333	2.49%
	2.0	496	3.71%
	3.0	878	6.57%
	4.0	1,304	9.76%
	5.0	1,932	14.46%
IWT	0.5	1,099	8.23%
	1.0	2,203	16.49%
	1.5	3,064	22.94%
	2.0	3,652	27.34%
	3.0	4,212	31.53%
	4.0	4,641	34.75%

Ensuring the safety of pedestrians and implementing traffic regulations around the area hold paramount importance to increase accessibility of the metro stations. This is essential to mitigate possible hazards, enhance travel convenience and consequently increase metro ridership. Besides, the implementation of traffic regulations also promotes sustainable mobility by reducing traffic congestion and promoting sustainable ways of transport, such as cycling, walking etc. For instance, providing parking facilities for micromobility vehicles around the stops where the most transfer activities occur may also provide seamless and efficient transportation system. These measures contribute to enhancing the overall experience for passengers, encouraging multimodal travel, and promoting sustainable urban mobility.

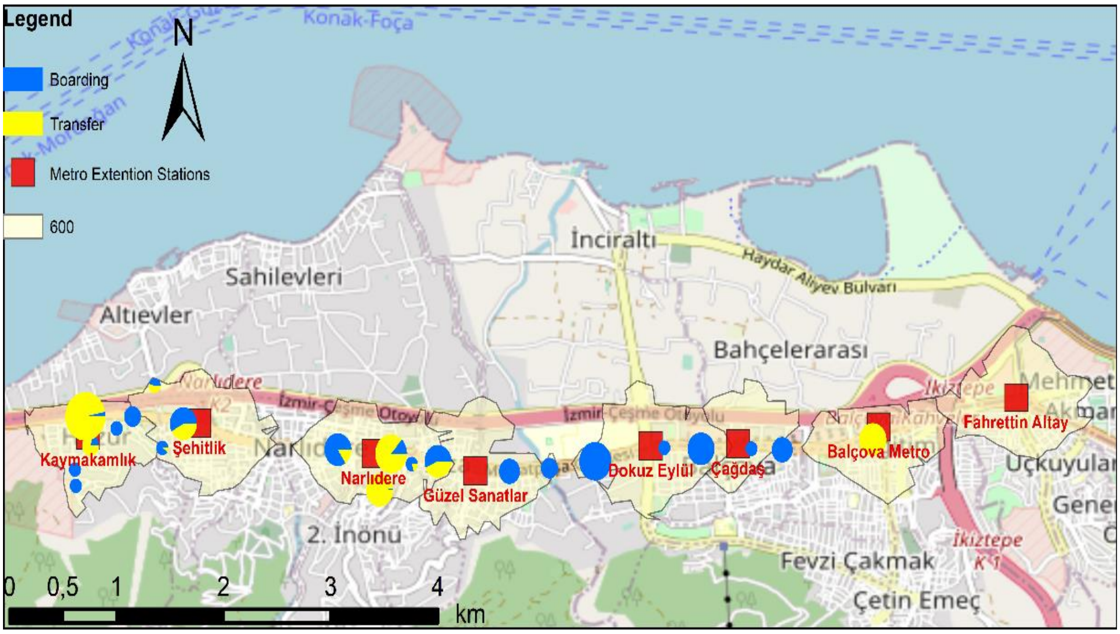


Figure 6.4. Possible transfer stations.

6.4. Travel Time Savings

Travel time saving is one of the main influencing factors in the passengers' mode choice. Also, it is a powerful indicator to establish the efficiency of service changes. In Figure 6.5 and Figure 6.6, the geographical distribution of travel time savings by origin stops of competitive bus passengers considering CAMin method are shown. At first glance, in the current conditions, the passengers starting their trips within or close to the service area of the metro extension are benefiting from the advantages of the metro. Additionally, there are passengers in distant districts, such as Güzelbahçe and Gülbahçe, who are experiencing significant travel time reductions, typically ranging from 1 to 8 generalized time minutes. It is highly likely that these passengers are the ones transferring to the metro at Fahrettin Altay station.

In Figure 6.6, spatial distribution of travel time savings based on the origin stops of bus passengers within the service area of the metro extension is shown. It is clear the passengers who require transfer trip to use the metro are suffering from the deterrence of transferring on the travel convenience even though their origin is relatively close to the metro. This could also be attributed to the relatively short travel distances, resulting in minimal disparities for passengers when choosing between the metro and buses. It is also important to not there is a rarity in the high travel time savings around the metro stations.

This is due to the consideration of travel convenience instead of looking at pure travel time savings. This indicates that such expensive public transportation investments are also beneficial to people living distant from its direct service area. This may be enhanced by integrating the insufficient transit services in the remote districts with efficient, comfortable, and fast alternatives. Consequently, reducing the travel times and increasing the travel convenience of the people living remote districts and ensuring the equity and accessibility by making the mobility easier.

In Figure 6.7, the average time savings by bus routes considering the shifting, non-shifting, and all passengers separately based on CAMin method. The average time saving for shifting passengers is around 6 minutes of generalized time minutes. However, there is a sharp distinction between shifting and non-shifting passengers. If we consider the average travel time saving of all the passengers, it is negative for all bus routes in consideration. For the bus routes having longer route length, bus route 984 and so on, this value is relatively lower indicating that the number of passengers who benefit from utilizing the Narlıdere metro line is lower on these bus routes. This could also be interpreted as a requirement for the bus routes to continue operating in some capacity.

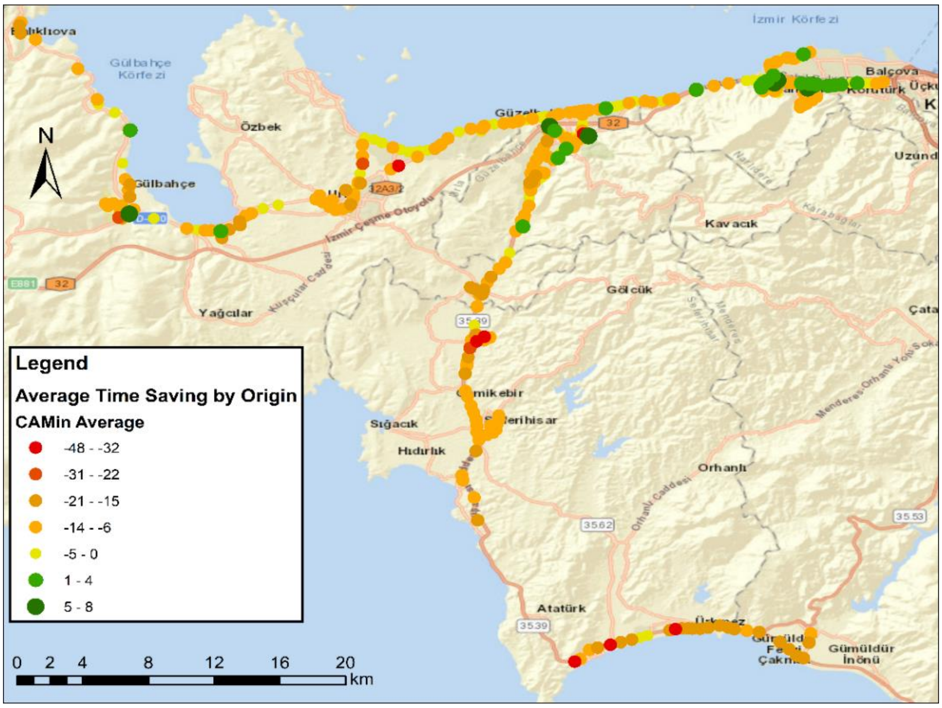


Figure 6.5. Geographical distribution of travel time savings by origin (stop) for the 16 bus routes in the analysis for the day.

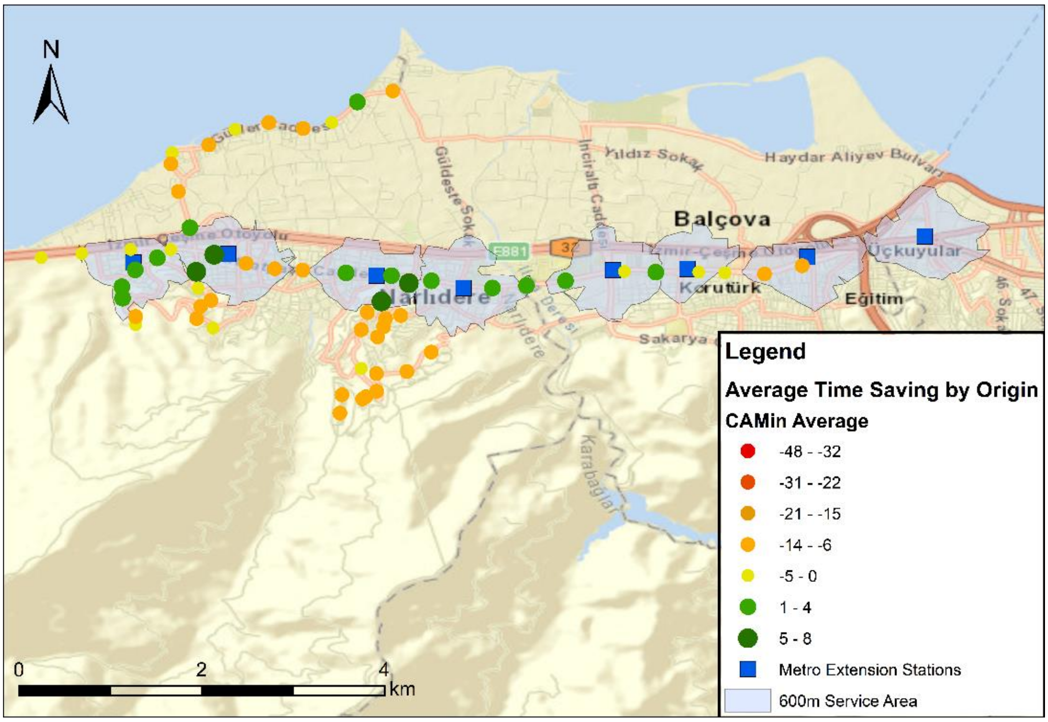


Figure 6.6. Geographical distribution of travel time savings by origin (stop) close to the metro extension for the day of analysis.

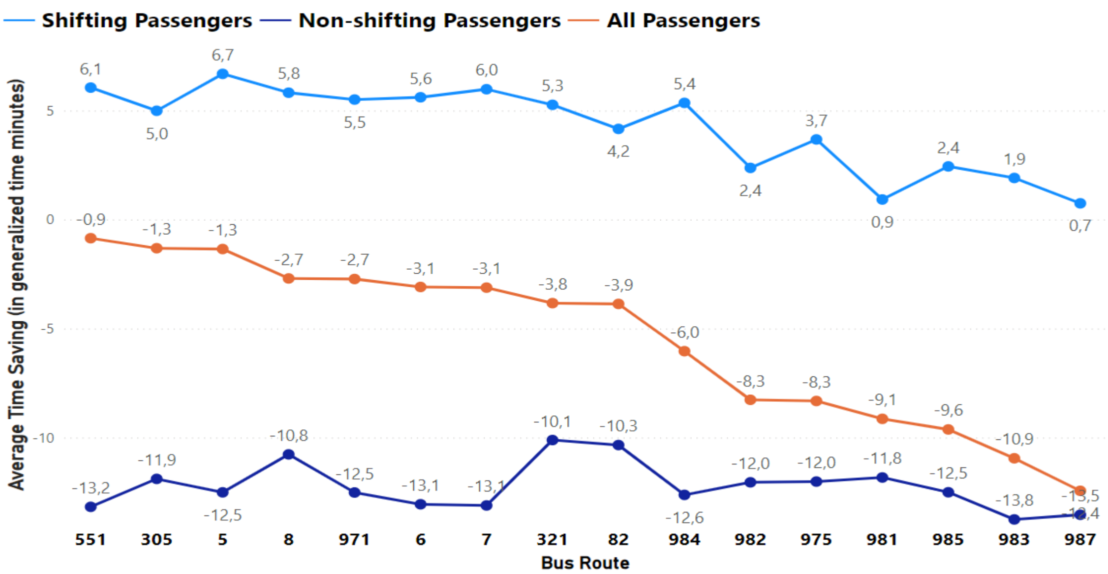


Figure 6.7. Average generalized time saving of the passengers on the competitive bus routes based on CAMin method.

The Figure 6.8 shows the average travel time savings for non-shifting passengers on the competitive bus routes. This figure can be interpreted as showing the required average travel time improvement for non-shifting passengers in the groups. If we analyze this figure with Figure 5.14, we can see that, non-shifting passengers in group A-2 are the potential shifters after a slight improvement in the travel times (2.2 generalized time minutes for CAMax method). Also, passengers in group B-2 can be considered as potential shifters since 4 minutes (actual time) of improvements can capture the majority of these passengers. Thus, as in Table 5.5, we can consider the passengers found not shifting to the metro and fall into groups A-1, A-2, B-1 and B-2 as potential shifters. On the other hand, passengers in groups P1, P2, A-3, and B-3 can be considered impractical shifters. This classification is based on the average travel time savings required for the non-shifting passengers in the groups.

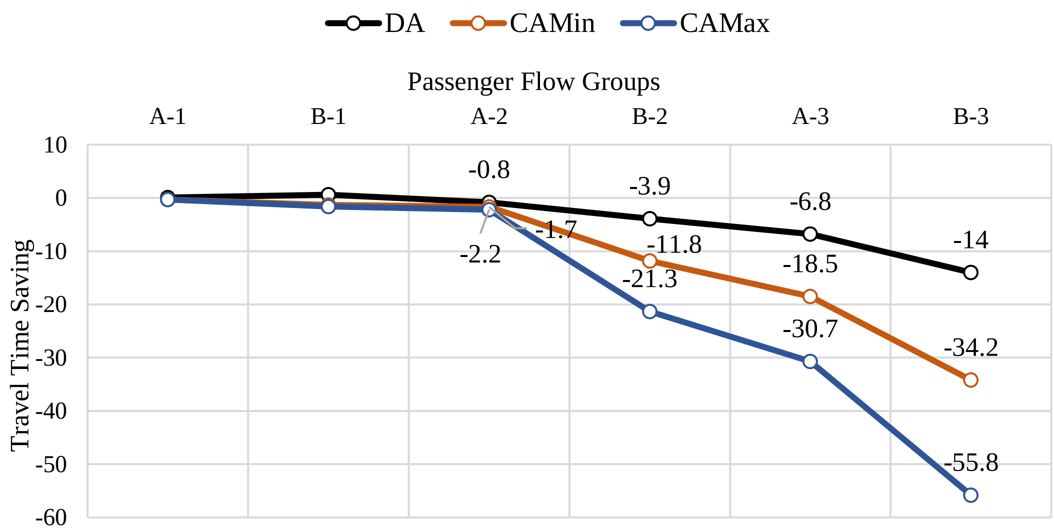


Figure 6.8: Average travel time savings of non-shifting passengers on the competitive bus routes by groups and method

6.5. Urban and Suburban Bus Routes

The result presented in section 5.4.5 gives important insights about the ways of efficiently adjusting the bus routes found competitive to the metro extension. The results indicate the possibility of shortening suburban bus routes by ending that the closest station of the metro extension. There are 17,426 passengers on these bus routes and 12,491 found not shifting to the metro. If we exclude the 3,529 passengers in group P2, since the will not be affected by the shortening of the bus routes, there are 8,962 passengers to be considered. The passengers in groups B-1 and B-2 (6,453 in total, 72%) are potential passengers that can be forced to use the metro and their travel time losses can be prevented by planning actions. However, the number of passengers in groups A-3 and B-3 (1,001 in total) indicates a need to run an alternative mode parallel to the metro line which also serves to stops outside of the metro extension's service area. These passengers along with the ones in urban areas can be served by bus routes operating parallel to the metro. On the other hand, the number of passengers who start or end their trip outside the metro service area can be seen as the need of continuing the operation of urban bus routes in the current conditions.

6.6. Sensitivity Analysis on Deterministic Approach Threshold

In the deterministic approach (DA), the estimation of mode shift is based on the assumption that passengers will only shift to the metro if they save more than 10% of their travel time. Notably, the selection of this 10% threshold is subjective and can alter depending on a variety of factors. While this threshold is commonly used in literature, it is essential to note that the estimation process is sensitive to the threshold value chosen. Therefore, the favored threshold value can influence the estimation's results and outcomes. Figure 6.9 shows the sensitivity of the shifting behavior on the selected threshold. If the threshold is set to zero, the percentage of shifting passenger increases about 24% considering the 10% travel time saving threshold.

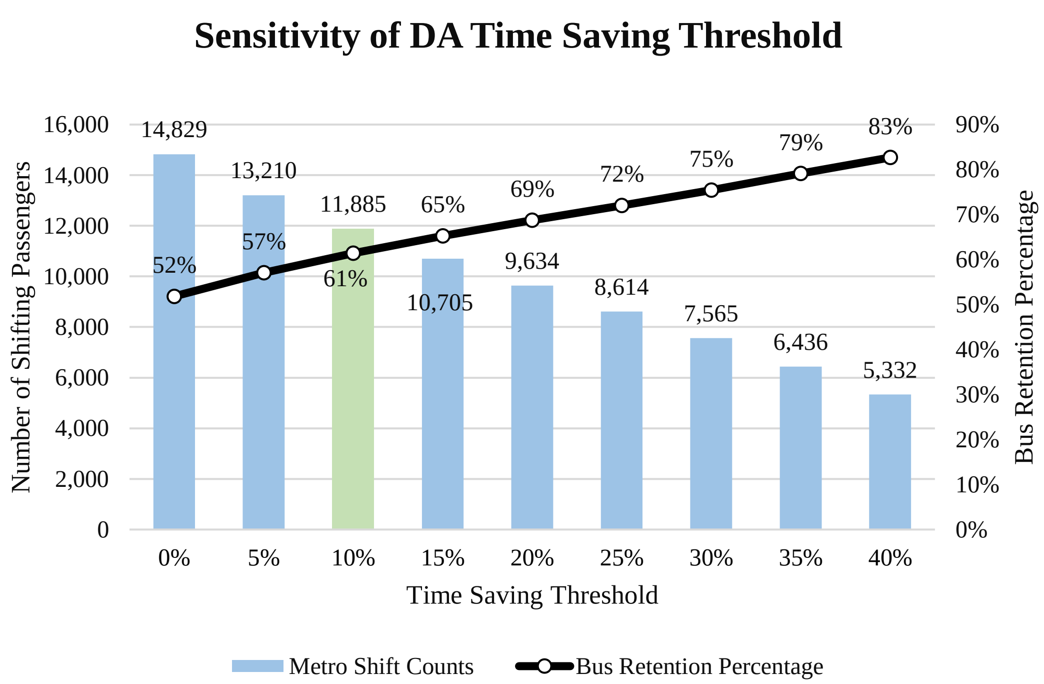


Figure 6.9. The number of shifting passengers by different decision thresholds in DA method.

6.7. Conclusion

First of all, we would like to emphasize that all the processes in our study are developed in such a way that they can be implemented practically and adoptable to alternative scenarios. There are several sections where improvements are necessary. For example, we have constructed a trip chaining algorithm and established a successful matching rate. However, our algorithm lacks the recent improvements presented in literature. Probably the most important deficiency is that our algorithm does not improve its matching rate by utilizing multiple days of smart card data. In further analysis, we first aim to adopt our algorithm for multiple days of data, such as one week of one month.

Furthermore, we believe that the process of establishing competitive and cooperative services with another transit service can be used in several aspects. For example, determining competitive services to a rail system is important to determine the transit systems that must be strengthened in case of a failure of the rail system. Also, to promote sustainable transportation modes, their alternatives can be made less attractive.

Using different method in the travel time calculation while showing that there are variability in approaches and the importance of considering different aspects of traveling also states that the number do not fluctuate significantly and even the simplistic approaches may provide sufficient information for the aggregate level planning processes.

Results show that the number of passengers who may shift to the metro is significant, and some adjustments are required. This will clearly help to decrease the congestion level on the streets by decreasing the number of buses in the traffic. On the other hand, the number of passengers that can be seen as resistant to mode shift even if the conditions are improved is also high (see Table 5.5). Similar to the findings presented by Gadepalli et al., (2022), some passengers on these parallel bus routes will not likely shift to the metro. About 40-50% of passengers are found to be not likely to prefer their bus trip to a metro trip. Thus, some parallel bus transit services are required.

One other important result obtained by using network walking distance is the essentiality of regulatory precautions around the transfer bus stops. Because network walking distance also considers the location of pedestrian crossings. There are stops very close to a station in bird’s-eye view, however they lack the necessary regulations, such as pedestrian crossings, etc. This shows that making stops solely close to the stations will not improve the transfer conditions, thereby encouraging passengers to shift to the metro.

The mode preference of passengers is generally understood by analyzing stated or revealed preference surveys. There is a body of literature on passengers' perception, mode choice, and route choice in public transportation that can be reviewed in order to understand the passenger behavior. In this study, instead of putting effort into conducting a survey, we focused on collecting the findings on the issue and directly implementing them in the analysis. However, we acknowledge that behavioral indications are very local and not easily transferable.

Furthermore, the results of the spatial analysis of travel time savings showed that investments benefit not only residents within proximity but also individuals living in more distant areas. This emphasizes the significance of enhancing public transportation services, which promotes equity in mobility and accessibility. Successful integration strategies can leverage these enhancements to increase the benefits for all individuals. The results of this study can be used to prevent inefficiency in transit services by allocating resources and making specific adjustments. However, it is important to note that our results are sensitive to the assumptions we made. Additionally, this study can be improved by utilizing inputs, which are acquired from local data, that better characterized passengers' behaviors and preferences.

REFERENCES

Abrantes, P. A. L., & Wardman, M. R. (2011). Meta-analysis of UK values of travel time: An update. *Transportation Research Part A: Policy and Practice, 45*(1), 1–17. https://doi.org/10.1016/j.tra.2010.08.003

Ali, A., Kim, J., & Lee, S. (2016). Travel behavior analysis using smart card data. *KSCE Journal of Civil Engineering, 20*(4), 1532–1539. https://doi.org/10.1007/s12205-015-1694-0

Alsger, A. A., Mesbah, M., Ferreira, L., & Safi, H. (2015). Use of smart card fare data to estimate public transport origin-destination matrix. *Transportation Research Record, 2535*, 88–96. https://doi.org/10.3141/2535-10

Alsger, A., Assemi, B., Mesbah, M., & Ferreira, L. (2016). Validating and improving public transport origin-destination estimation algorithm using smart card fare data. *Transportation Research Part C: Emerging Technologies*, *68*, 490–506. https://doi.org/10.1016/j.trc.2016.05.004

Amirgholy, M., Shahabi, M., & Gao, H. O. (2017). Optimal design of sustainable transit systems in congested urban networks: A macroscopic approach. *Transportation Research Part E: Logistics and Transportation Review, 103*, 261–285. https://doi.org/10.1016/j.tre.2017.03.006

Anderson, M. K. A. (2013). *Behavioural Models for Route Choice of Passengers in Multimodal Public*. PhD Thesis, Department of Transport, Technical University of Denmark.

Anderson, M. K., Nielsen, O. A., & Prato, C. G. (2017). Multimodal route choice models of public transport passengers in the Greater Copenhagen Area. *EURO Journal on Transportation and Logistics*, *6*(3), 221–245. https://doi.org/10.1007/s13676-014-0063-3

António A. Nunes, & Dias, T. G. (2016). Passenger Journey Destination Estimation from Automated Fare Collection System Data Using Spatial Validation. *IEE Transactions on Intelligent Transportation Systems*, *17*(1), 133–142.

Assemi, B., Alsger, A., Moghaddam, M., Hickman, M., & Mesbah, M. (2020). Improving alighting stop inference accuracy in the trip chaining method using neural networks. In *Public Transport* (Vol. 12, Issue 1). https://doi.org/10.1007/s12469-019-00218-9

Bagchi, M., & White, P. R. (2005). The potential of public transport smart card data. *Transport Policy*, *12*(5), 464–474. https://doi.org/10.1016/j.tranpol.2005.06.008

Barry, J. J., Freimer, R., & Slavin, H. (2009). Use of entry-only automatic fare collection data to estimate linked transit trips in New York city. *Transportation Research Record*, *2112*, 53–61. https://doi.org/10.3141/2112-07

Barry, J. J., Newhouser, R., Rahbee, A., & Sayeda, S. (2002). Origin and destination estimation in New York City with automated fare system data. *Transportation Research Record*, *1817*, 183–187. https://doi.org/10.3141/1817-24

Ben-Akiva, M., & Morikawa, T. (2002). Comparing ridership attraction of rail and bus. *Transport Policy*, *9*(2), 107–116. https://doi.org/10.1016/S0967-070X(02)00009-4

Bovy, P. H. L., & Hoogendoorn-Lanser, S. (2005). Modelling route choice behaviour in multi-modal transport networks. *Transportation*, *32*(4), 341–368. https://doi.org/10.1007/s11116-004-7963-2

Brands, T., Dixit, M., & van Oort, N. (2020). Impact of a new metro line in amsterdam on ridership, travel times, reliability and societal costs and benefits. *European Journal of Transport and Infrastructure Research*, *20*(4), 335–353. https://doi.org/10.18757/ejtir.2020.20.4.4084

Brands, T., Dixit, M., Zúñiga, E., & van Oort, N. (2022). Perceived and actual travel times in a multi-modal urban public transport network: comparing survey and AVL data. *Public Transport*, *14*(1), 85–103. https://doi.org/10.1007/s12469-022-00298-0

Canepa, B. (2007). Bursting the bubble: Determining the transit-oriented development's walkable limits. *Transportation Research Record*, *1992*(1), 28–34. https://doi.org/10.3141/1992-04

Ceder, A. (2007). Public Transit Planning and Operation. In *CRC press*. https://doi.org/10.1201/b12853

Cengiz, D. (2022). *Estimation of Origin-Destination Matrix from Smart Card Data of Public Transportation of the Community of Madrid*. Universidad Politécnica de Madrid.

Cerqueira, S., Arsenio, E., & Henriques, R. (2022). *Is there any best practice principles to estimate bus alighting passengers from incomplete smart card transactions ? 00*, 1–8.

Cheng, Z., Trépanier, M., & Sun, L. (2019). *Inferring Trip Destinations in Transit Smart Card Data Using a Probabilistic Topic Model*. *October*. https://exo.quebec/en

Chira-Chavala, T., & Coifman, B. (1996). Effects of smart cards on transit operators. *Transportation Research Record*, *1521*, 84–90. https://doi.org/10.3141/1521-12

Claffey, P. J. (1964). User criteria for rapid transit planning. *Journal of the Urban Planning and Development Division*, *90*(1), 5–14.

Cui, A. (2006). *Bus Passenger Origin-Destination Matrix Estimation Using Automated Data Collection Systems by*. Massachusetts Institute of Technology.

Cui, J., Gao, Y., Cheng, J., & Shi, L. (2020). Study on the Selection Model of Staying Adjustment Bus Lines along Rail Transit. *Journal of Advanced Transportation*, *2020*. https://doi.org/10.1155/2020/6385359

Cui, Z., Wang, C., Gao, Y., Yang, D., Wei, W., Chen, J., & He, T. (2021). Alighting Stop Determination of Unlinked Trips Based on a Two-Layer Stacking Framework. *Mathematical Problems in Engineering*, *2021*. https://doi.org/10.1155/2021/6464980

Currie, G. (2005). The Demand Performance of Bus Rapid Transit. *Journal of Public Transportation*, *8*(1), 41–55. https://doi.org/10.5038/2375-0901.8.1.3

Deri, A. (2012). *Akıllı Kart Verileri Kullanılarak Toplu Ulaşım Yolculuk Talebinin Belirlenmesi ve Sefer Çizelgeleme Optimizasyonu*. *25*(2), 122–128. http://acikerisim.deu.edu.tr/xmlui/handle/12345/7791

Deri, A. (2018). *The Use of Smart Card Data in Public Transport Assignment Models*. Dokuz Eylül Üniversity.

Dewulf, B., Neutens, T., Van Dyck, D., de Bourdeaudhuij, I., & Van de Weghe, N. (2012). Correspondence between objective and perceived walking times to urban destinations: Influence of physical activity, neighbourhood walkability, and socio-demographics. *International Journal of Health Geographics*, *11*, 1–10. https://doi.org/10.1186/1476-072X-11-43

Diker, A. C., Nasiboğlu, E., & Nasibov, E. (2016). Toplu Ulaşımda Akıllı Kart Verilerinin Analizi ile Yolcu Yoğunluğunun Tahmin Edilmesi. *Big Data: Imkanları, Multidissiplinar Problemləri və Perspektivləri" I Respublika Elmi-Praktiki Konfransı*, 19–21.

Du, M., Cheng, L., Huang, L., & Tu, Q. (2018). Network Optimization of Conventional Public Transit Based on Urban Rail Transit. *CICTP 2018: Intelligence, Connectivity, and Mobility - Proceedings of the 18th COTA International Conference of Transportation Professionals, March*, 766–774.

El-Geneidy, A., Grimsrud, M., Wasfi, R., Tétreault, P., & Surprenant-Legault, J. (2014). New evidence on walking distances to transit stops: Identifying redundancies and gaps using variable service areas. *Transportation*, *41*(1), 193–210. https://doi.org/10.1007/s11116-013-9508-z

Eluru, N., Chakour, V., & El-Geneidy, A. M. (2012). Travel mode choice and transit route choice behavior in Montreal: Insights from McGill University members commute patterns. *Public Transport*, *4*(2), 129–149. https://doi.org/10.1007/s12469-012-0056-2

ESHOT, Y. Ş. M. (2021a). *İzmir Open Data Portal - Bus Line Routes*. İzmir Büyükşehir Belediyesi. https://acikveri.bizizmir.com/dataset/eshot-otobus-hat-listesi

ESHOT, Y. Ş. M. (2021b). *İzmir Open Data Portal - Stops on bus routes*. İzmir Büyükşehir Belediyesi. https://acikveri.bizizmir.com/tr/dataset/otobus-gidis-yonune-gore-hatta-ait-duraklar

Fan, Y., Guthrie, A., & Levinson, D. (2016). Waiting time perceptions at transit stops and stations: Effects of basic amenities, gender, and security. *Transportation Research Part A: Policy and Practice*, *88*, 251–264. https://doi.org/10.1016/j.tra.2016.04.012

Farzin, J. M. (2008). Constructing an automated bus origin-destination matrix using farecard and global positioning system data in São Paulo, Brazil. *Transportation Research Record*, *2072*, 30–37. https://doi.org/10.3141/2072-04

Fidanoğlu, A., & Gökaşar, I. (2022). A Method for Public Transit OD Estimation Using Transit Smart Card Data. *Journal of Transportation and Logistics*, *7*(2), 479–488. https://doi.org/10.26650/jtl.2022.1192128

Furth, P. G., & Muller, T. H. J. (2006). Service reliability and hidden waiting time insights from automatic vehicle location data. *Transportation Research Record*, *1955*, 79–87. https://doi.org/10.1177/0361198106195500110

Gadepalli, R., Gumireddy, S., Bhamidipati, S., & Cats, O. (2022). Impact of introducing a metro line on urban bus services. *Case Studies on Transport Policy*, *10*(2), 940–947. https://doi.org/10.1016/j.cstp.2022.03.007

Gao, Y., Guo, Z., Long, Y., Cui, Z., & Li, X. (2022a). Passengers' travel behavior before and after the adjustment of regular bus collinear sections: A case study in the incipient phase of metro operation in Xiamen. *Travel Behaviour and Society*, *26*(October 2021), 221–230. https://doi.org/10.1016/j.tbs.2021.10.006

Gao, Y., Guo, Z., Long, Y., Cui, Z., & Li, X. (2022b). Passengers' travel behavior before and after the adjustment of regular bus collinear sections: A case study in the incipient phase of metro operation in Xiamen. *Travel Behaviour and Society*, *26*(January 2021), 221–230. https://doi.org/10.1016/j.tbs.2021.10.006

Gordon, J. B., Koutsopoulos, H. N., Wilson, N. H. M., & Attanucci, J. P. (2013). Automated inference of linked transit journeys in london using fare-transaction and vehicle location data. *Transportation Research Record*, *2343*, 17–24. https://doi.org/10.3141/2343-03

Gore, N., Dave, S., Shah, J., Jain, M., Rathva, D., & Garg, V. (2020). Comparative analysis of pedestrian walking speed on sidewalk and carriageway. *Transportation Research: Proceedings of CTRG 2017*, 65–76.

Guo, Z., & Wilson, N. H. M. (2011). Assessing the cost of transfer inconvenience in public transport systems: A case study of the London Underground. *Transportation Research Part A: Policy and Practice*, *45*(2), 91–104. https://doi.org/10.1016/j.tra.2010.11.002

Hao, X. U. (2007). *Evaluation of Benefits and Effectiveness of Smart Cards for Public Transport*. University of Leeds.

He, L., Nassir, N., & Hickman, M. (2015). *Validating and Calibrating a Destination Estimation Algorithm for Public Transport Smart Card Fare Collection Systems Transport Smart Card Fare Collection Systems*. *October*.

He, L., & Trépanier, M. (2015). Estimating the destination of unlinked trips in transit smart card fare data. *Transportation Research Record*, *2535*, 97–104. https://doi.org/10.3141/2535-11

Hess, D. B., Brown, J., & Shoup, D. (2004). Waiting for the bus. *Journal of Public Transportation*, *7*(4), 67–84. https://doi.org/10.5038/2375-0901.7.4.4

Hora, J., Dias, T. G., Camanho, A., & Sobral, T. (2017). Estimation of Origin-Destination matrices under Automatic Fare Collection: The case study of Porto transportation system. *Transportation Research Procedia*, *27*, 664–671. https://doi.org/10.1016/j.trpro.2017.12.103

Huang, D., Yu, J., Shen, S., Li, Z., Zhao, L., & Gong, C. (2020). A Method for Bus OD Matrix Estimation Using Multisource Data. *Journal of Advanced Transportation, 2020*. https://doi.org/10.1155/2020/5740521

Hussain, E., Bhaskar, A., & Chung, E. (2021). Transit OD matrix estimation using smartcard data: Recent developments and future research challenges. *Transportation Research Part C: Emerging Technologies*, *125*(February), 103044. https://doi.org/10.1016/j.trc.2021.103044

İzBB. (2019). *İzmir Rail Sytem Map*. İzmir Metropolitan Municipality. https://www.izmir.bel.tr/tr/ulasim-haritalari-/507/2168

İzBB. (2023). *İzmir's Commuter Rail System*. İzban. https://www.izban.com.tr/Sayfalar/Single.aspx?MenuId=15

İzBB, T. U. H. Ş. M. (2022). *İzmir Open Data Portal - GTFS Data*. İzmir Büyükşehir Belediyesi. https://acikveri.bizizmir.com/dataset/toplu-ulasim-gtfs-verileri

Jafari Kang, M., Ataeian, S., & Amiripour, S. M. M. (2021). A procedure for public transit OD matrix generation using smart card transaction data. *Public Transport*, *13*(1), 81–100. https://doi.org/10.1007/s12469-020-00257-7

Jánošíkova, L., Slavík, J., & Koháni, M. (2014). Estimation of a route choice model for urban public transport using smart card data. *Transportation Planning and Technology*, *37*(7), 638–648. https://doi.org/10.1080/03081060.2014.935570

Jara-Diaz, S., Monzon, A., Cascajo, R., & Garcia-Martinez, A. (2022). An international time equivalency of the pure transfer penalty in urban transit trips: Closing the gap. *Transport Policy*, *125*(April), 48–55. https://doi.org/10.1016/j.tranpol.2022.04.019

Jenelius, E. (2018). Public transport experienced service reliability: Integrating travel time and travel conditions. *Transportation Research Part A: Policy and Practice*, *117*(December 2017), 275–291. https://doi.org/10.1016/j.tra.2018.08.026

Jinhua, Z. (2004). The planning and analysis implications of automatic data collection systems: Rail transit OD matrix inference and path choice modelling examples. In *Thesis of Master, M.S. thesis, MIT, USA.*

Jung, J., & Sohn, K. (2017). Deep-learning architecture to forecast destinations of bus passengers from entry-only smart-card data. *IET Intelligent Transport Systems*, *11*(6), 334–339.

Kim, K., & Lee, I. (2017). Public transportation alighting estimation method using smart card data. *Journal of the Korean Society for Railway*, *20*(5), 692–702. https://doi.org/10.7782/JKSR.2017.20.5.692

Knoblauch, R. L., Pietrucha, M. T., & Nitzburg, M. (1996). Field studies of pedestrian walking speed and start-up time. *Transportation Research Record*, *1538*, 27–38. https://doi.org/10.3141/1538-04

Kroes, E., Kouwenhoven, M., Debrincat, L., & Pauget, N. (2014). Value of crowding on public transport in Île-de-France, France. *Transportation Research Record*, *2417*(2417), 37–45. https://doi.org/10.3141/2417-05

Kumar, P. (2019). Transit Origin Destination Estimation using Automated Data. *ProQuest Dissertations and Theses*, 73. https://www.proquest.com/dissertations-theses/transit-origin-destination-estimation-using/docview/2307163412/se-2?accountid=15198%0Ahttp://www.yidu.edu.cn/educhina/educhina.do?artifact=&svalue=Transit+Origin+Destination+Estimation+using+Automated+Data&sty

Lee, I., Cho, S. H., Kim, K., Kho, S. Y., & Kim, D. K. (2022). Travel pattern-based bus trip origin-destination estimation using smart card data. *PLoS ONE*, *17*(6 June), 1–20. https://doi.org/10.1371/journal.pone.0270346

Li, D., Lin, Y., Zhao, X., Song, H., & Zou, N. (2011). *Estimating a Transit Passenger Trip Origin-Destination Matrix Using Automatic Fare Collection System*. 502–513. https://link-springer-com.ezproxy.uniandes.edu.co:8443/content/pdf/10.1007%2F978-3-642-20244-5_48.pdf

Li, T., Sun, D., Jing, P., & Yang, K. (2018). Smart card data mining of public transport destination: A literature review. *Information (Switzerland)*, *9*(1), 28–30. https://doi.org/10.3390/info9010018

Luo, D., Bonnetain, L., Cats, O., & van Lint, H. (2018). Constructing Spatiotemporal Load Profiles of Transit Vehicles with Multiple Data Sources. *Transportation Research Record*, *2672*(8), 175–186. https://doi.org/10.1177/0361198118781166

Luo, Q. Y., Sun, B. F., & Zhong, Y. P. (2010). Collinear competition between rail transit and bus transit based on population competition model. *Proceedings - 2010 WASE International Conference on Information Engineering, ICIE 2010*, *3*, 313–315. https://doi.org/10.1109/ICIE.2010.251

Meng, M., Rau, A., & Mahardhika, H. (2018). Public transport travel time perception: Effects of socioeconomic characteristics, trip characteristics and facility usage. *Transportation Research Part A: Policy and Practice*, *114*(January), 24–37. https://doi.org/10.1016/j.tra.2018.01.015

Meyer, M. D. (2016). Transportation Planning Handbook. In *Transportation Planning Handbook*. https://doi.org/10.1002/9781119174660

Miller, P., de Barros, A. G., Kattan, L., & Wirasinghe, S. C. (2016). Public transportation and sustainability: A review. *KSCE Journal of Civil Engineering*, *20*(3), 1076–1083. https://doi.org/10.1007/s12205-016-0705-0

Mohammed, M., & Oke, J. (2022). Origin-destination inference in public transportation systems: A comprehensive review. *International Journal of Transportation Science and Technology*. https://doi.org/10.1016/j.ijtst.2022.03.002

Mosallanejad, M., Somenahalli, S., & Mills, D. (2019). Origin-destination estimation of bus users by smart card data. *Lecture Notes in Geoinformation and Cartography*, 305–320. https://doi.org/10.1007/978-3-030-19424-6_17

Munizaga, M. A., & Palma, C. (2012). Estimation of a disaggregate multimodal public transport Origin-Destination matrix from passive smartcard data from Santiago, Chile. *Transportation Research Part C: Emerging Technologies*, *24*, 9–18. https://doi.org/10.1016/j.trc.2012.01.007

Munizaga, M., Devillaine, F., Navarrete, C., & Silva, D. (2014). Validating travel behavior estimated from smartcard data. *Transportation Research Part C: Emerging Technologies*, *44*, 70–79. https://doi.org/10.1016/j.trc.2014.03.008

Nassir, N., Khani, A., Lee, S. G., Noh, H., & Hickman, M. (2011). Transit stop-level origin-destination estimation through use of transit schedule and automated data collection system. *Transportation Research Record*, *2263*, 140–150. https://doi.org/10.3141/2263-16

Nielsen, O. A. (2000). A stochastic transit assignment model considering differences in passengers utility functions. *Transportation Research Part B: Methodological*, *34*(5), 377–402. https://doi.org/10.1016/S0191-2615(99)00029-6

Nielsen, O. A., Eltved, M., Anderson, M. K., & Prato, C. G. (2021). Relevance of detailed transfer attributes in large-scale multimodal route choice models for metropolitan public transport passengers. *Transportation Research Part A: Policy and Practice*, *147*(January), 76–92. https://doi.org/10.1016/j.tra.2021.02.010

Nielsen, O. A., Hansen, C. O., & Daly, A. (2000). A Large-scale model system for the Copenhagen-Ringsted railway project. *International Conference on Travel Behaviour Research, 9th, 2000, Gold Coast, Queensland, Australia, Vol12*.

O'Farrell, P. N., & Markham, J. (1974). Commuter Perceptions of Public Transport Work Journeys. *Environment and Planning A: Economy and Space*, *6*(1), 79–100. https://doi.org/10.1068/a060079

Padhye, P., Nagakumar, M. S., Sunil, S., & Reddy, A. H. M. (2020). Activity-based travel demand models to evaluate transport policies. In *Lecture Notes in Civil Engineering* (Vol. 45). https://doi.org/10.1007/978-981-32-9042-6_20

Palamutçuolu, B. T., & Gerşil, M. (2020). *Şehir İçi Toplu Taşıma Planlamasında Akıllı Kart Veri Madenciliği ile Sefer Sıklığı Optimizasyonu*.

Pelletier, M. P., Trépanier, M., & Morency, C. (2011). Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies*, *19*(4), 557–568. https://doi.org/10.1016/j.trc.2010.12.003

Peng, Z. (1994). A Simultaneous Route-level Transit Patronage Model: Demand, Supply, and Inter-route Relationship. In *Dissertations and Theses*. https://doi.org/https://doi.org/10.15760/etd.1158

Pronello, C., Longhi, D., & Gaborieau, J. B. (2018). Smart card data mining to analyze mobility patterns in suburban areas. *Sustainability (Switzerland)*, *10*(10), 1–21. https://doi.org/10.3390/su10103489

Psarros, I., Kepaptsoglou, K., & Karlaftis, M. G. (2011). An empirical investigation of passenger wait time perceptions using hazard-based duration models. *Journal of Public Transportation*, *14*(3), 109–122. https://doi.org/10.5038/2375-0901.14.3.6

Pursula, M., & Weurlander, M. (1999). Modeling Level-of-Service Factors in Public Transportation Route Choice. *Transportation Research Record: Journal of the Transportation Research*, *1669*(1), 30–37. https://doi.org/doi:10.3141/1669-04

Rahman, M., Akther, M. S., & Recker, W. (2022). The first-and-last-mile of public transportation: A study of access and egress travel characteristics of Dhaka's suburban commuters. *Journal of Public Transportation*, *24*(April), 100025. https://doi.org/10.1016/j.jpubtr.2022.100025

Rastogi, R., Thaniarasu, I., & Chandra, S. (2011). Design Implications of Walking Speed for Pedestrian Facilities. *Journal of Transportation Engineering*, *137*(10), 687–696. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000251

Raveau, S., Guo, Z., Muñoz, J. C., & Wilson, N. H. M. (2014). A behavioural comparison of route choice on metro networks: Time, transfers, crowding, topology and socio-demographics. *Transportation Research Part A: Policy and Practice, 66*(1), 185–195. https://doi.org/10.1016/j.tra.2014.05.010

Raveau, S., Muñoz, J. C., & de Grange, L. (2011). A topological route choice model for metro. *Transportation Research Part A: Policy and Practice*, *45*(2), 138–147. https://doi.org/10.1016/j.tra.2010.12.004

Robinson, S., Narayanan, B., Toh, N., & Pereira, F. (2014). Methods for pre-processing smartcard data to improve data quality. *Transportation Research Part C: Emerging Technologies*, *49*, 43–58. https://doi.org/10.1016/j.trc.2014.10.006

Rui, T. (2016). *Modeling Route Choice Behaviour in Public Transport Network*. https://core.ac.uk/download/pdf/83107872.pdf

Shao, M., Xie, C., Li, T., & Sun, L. (2022). Influence of in-vehicle crowding on passenger travel time value: Insights from bus transit in Shanghai, China. *International Journal of Transportation Science and Technology*, *11*(4), 665–677. https://doi.org/10.1016/j.ijtst.2021.09.001

Soza-Parra, J., Raveau, S., & Muñoz, J. C. (2022). Public transport reliability across preferences, modes, and space. *Transportation*, *49*(2), 621–640. https://doi.org/10.1007/s11116-021-10188-2

Tanaboriboon, Y., Hwa, S. S., & Chor, C. H. (1986). Pedestrian characteristics study in Singapore. *Journal of Transportation Engineering*, *112*(3), 229–235. https://doi.org/10.1061/(ASCE)0733-947X(1986)112:3(229)

Trépanier, M., & Chapleau, R. (2006). Destination estimation from public transport smartcard data. *IFAC Proceedings Volumes (IFAC-PapersOnline)*, *12*(PART 1). https://doi.org/10.3182/20060517-3-fr-2903.00209

Trépanier, M., Morency, C., & Agard, B. (2009). Calculation of Transit Performance Measures Using Smartcard Data. *Journal of Public Transportation*, *12*(1), 79–96. https://doi.org/10.5038/2375-0901.12.1.5

Trépanier, M., Tranchant, N., & Chapleau, R. (2007). Individual trip destination estimation in a transit smart card automated fare collection system. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, *11*(1), 1–14. https://doi.org/10.1080/15472450601122256

TÜİK, T. İ. K. (2023). *Kent-Kır Nüfus İstatistikleri,2022*. Nüfus ve Vatandaşlık İşleri Genel Müdürlüğü.

Van der Waard, J. (1988). The relative importance of public transport trip time attributes in route choice. *PTRC Summer Annual Meeting, 16th, 1988, Bath, United Kingdom*.

Van Goeverden, C., Rietveld, P., Koelemeijer, J., & Peeters, P. (2006). *Subsidies in public transport*.

Van Soest, D., Tight, M. R., & Rogers, C. D. F. (2020). Exploring the distances people walk to access public transport. *Transport Reviews*, *40*(2), 160–182. https://doi.org/10.1080/01441647.2019.1575491

Vrtic, M. ;, & Axhausen, K. W. (2002). The impact of tilting trains in Switzerland: A route choice model of regional-and long distance public transport trips. *Proceedings of the 82nd Annual Meeting of the Transportation Research Board*, 1–25. https://doi.org/10.3929/ethz-a-004403563

Wang, F., Ye, M., Zhu, H., & Gu, D. (2022). Optimization Method for Conventional Bus Stop Placement and the Bus Line Network Based on the Voronoi Diagram. *Sustainability (Switzerland)*, *14*(13). https://doi.org/10.3390/su14137918

Wang, W., Attanucci, J. P., & Wilson, N. H. M. (2011). Bus passenger origin-destination estimation and related analyses using Automated Data Collection Systems. *Journal of Public Transportation*, *14*(4), 131–150. https://doi.org/10.5038/2375-0901.14.4.7

Wardman, M. (2001). A review of British evidence on time and service quality valuations. *Transportation Research Part E: Logistics and Transportation Review*, *37*(2–3), 107–128. https://doi.org/10.1016/S1366-5545(00)00012-0

Wardman, M. (2004). Public transport values of time. *Transport Policy*, *11*(4), 363–377. https://doi.org/10.1016/j.tranpol.2004.05.001

Wardman, M. (2014). Valuing Convenience in Public Transport. *Valuing Convenience in Public Transport*, 1–70. http://www.oecd-ilibrary.org/transport/valuing-convenience-in-public-transport_9789282107683-en

Watkins, K. E., Ferris, B., Borning, A., Rutherford, G. S., & 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*, *45*(8), 839–848. https://doi.org/10.1016/j.tra.2011.06.010

Wei, J., Long, K., Gu, J., Ju, Q., & Zhu, P. (2020). Optimizing bus line based on metro-bus integration. *Sustainability (Switzerland)*, *12*(4), 1–14. https://doi.org/10.3390/su12041493

Wigan, M. (1995). Treatment of walking as a mode of transportation. *Transportation Research Record*, *1487*, 7–13.

Yan, F., Yang, C., & Ukkusuri, S. V. (2019). Alighting stop determination using two-step algorithms in bus transit systems. *Transportmetrica A: Transport Science*, *15*(2), 1522–1542. https://doi.org/10.1080/23249935.2019.1615578

Yap, M., Cats, O., & van Arem, B. (2020). Crowding valuation in urban tram and bus transportation based on smart card data. *Transportmetrica A: Transport Science*, *16*(1), 23–42. https://doi.org/10.1080/23249935.2018.1537319

Yap, M. D., Cats, O., Van Oort, N., & Hoogendoorn, S. P. (2017). A robust transfer inference algorithm for public transport journeys during disruptions. *Transportation Research Procedia*, *27*, 1042–1049. https://doi.org/10.1016/j.trpro.2017.12.099

Zhang, J., & Li, T. (2014). Urban rail transit and bus collinear analysis based on competition model. *J Transport Info Safety*, *32*(4), 108–112.

Zhang, S. J., Jia, S. P., Bai, Y., Mao, B. H., Ma, C. R., & Zhang, T. (2018). Optimal Adjustment Schemes on the Long Through-Type Bus Lines considering the Urban Rail Transit Network. *Discrete Dynamics in Nature and Society*, *2018*. https://doi.org/10.1155/2018/8029137

Zhao, J., Rahbee, A., & Wilson, N. H. M. (2007). Estimating a rail passenger trip origin-destination matrix using automatic data collection systems. *Computer-Aided Civil and Infrastructure Engineering*, *22*(5), 376–387. https://doi.org/10.1111/j.1467-8667.2007.00494.x