Identifying the Transit Needs of Socioeconomic Groups by Evaluating the Relationship between a Network's Supply and Demand

A Case Study of the City of Amsterdam

Benjamin Drybrough

Technische Universiteit Delft

TU Delft

Cover photo: https://unsplash.com/photos/XrQD9OqkjYQ

Identifying the Transit Needs of Socioeconomic Groups by Evaluating the Relationship between a Network's Supply and Demand

A Case Study of the City of Amsterdam

by

Benjamin Drybrough

This research is done for the partial fulfilment of requirements for the Master of Science degree at the TU Delft, the Netherlands

To be defended publicly on November 25th, 2020.

Department of Transport and Planning, Delft University of Technology

Student number:	4775090	
Project duration:	February 10, 2020 – November 25, 2020	
Thesis committee:	Dr. Oded Cats,	TU Delft, Chair
	Dr. ir. Niels van Oort,	TU Delft, Daily Supervisor
	Dr. ir. Maarten Kroesen,	TU Delft
	Machiel Kouwenberg,	Vervoerregio Amsterdam

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

Acknowledgements

I would like to take this opportunity to thank everyone who helped me over the entire thesis period. For myself, it was a thrilling opportunity to work on a project in the city that peaked my interest of transport and motivated me to complete my studies at the TU Delft. However, without the help of many people the completion of the research would not have been possible.

I would like to first say thank you to my committee at the TU Delft consisting of Oded, Niels, Maarten, and Malvika. To Oded, thank you for introducing myself to the topic at the Vervoerregio, providing guidance throughout each stage of the process, and recently starting the refreshing Monday morning masters kick-offs. To Niels, thank you for your supervision and support during the tough Covid-19 months, providing opportunities to share my thoughts on my research, and suggesting almost a year and a half ago how I should go upon searching for practical experience within the Dutch public transport industry. To Maarten, thank you for challenging me in the later phases of my research and pushing me to a point where I am proud of the finished product. And to Malvika, thank you for your thoughtful answers surrounding transport accessibility research. You all played an important role in my thesis, so thank you very much.

I would also like to extend a thank you to all the colleagues I met at the Vervoerregio Amsterdam but specifically to Machiel and Mark of team Kennis & Onderzoek. Machiel, thank you for your calming guidance throughout the entire process and connecting me with the people and information required to complete my project. Mark, thank you for having utmost enthusiasm for my project and pushing me to get over the finish line. While I would have wished to spend more time at the Vervoerregio office I will remember my time at the Vervoerregio fondly.

Special thanks goes out to all my family and friends who I hold very close to my heart. Specifically to my parents, Dan and Marjolein, for their unconditional love and steadfast support that allowed me to fulfill this idea of completing my masters in the Netherlands. Lastly, I would like to thank my girlfriend Marieke, who has shown more love, support, and patience during this thesis period than I could have ever asked for. Thanks to everyone for making this thesis possible.

Ben Drybrough. November 2020. Amsterdam

Executive Summary

An inclusive and sustainable transport network allows people the opportunity to access their daily needs. A transportation planner attempts to facilitate and support these needs by providing a transport network that reaches into neighbourhoods of a transport region. When this is not completed in an effective manner a discrepancy between someones transport need and their transport network arises. This is termed as *transport poverty* which has negative consequences such as making employment or their social circles inaccessible.

It is argued that people of lower socioeconomic and demographic status are most at risk of experiencing transport poverty. This occurs as their transport needs are higher relative to other groups of people with more transport options. This is exacerbated by the fact that current transport networks are often designed for an idealised group of citizens and fails to cater to the specific needs of all individuals. A recent study commissioned by the European Commission explained that transport needs are not constant within a city and outlined eight guiding principles that should be considered when building a transport network for all individuals (Tovaas, 2020). This is an important step in defining specific transport needs for individuals rather than assuming levels of transport need based on their socioeconomic and demographic data.

This study attempts to move further away from these assumptions by building on a framework of latent demand proposed by Clifton and Moura (2017). In this framework, transport needs can be identified by observing the transition between latent demand into effective demand when presented with an improved transport network. By observing the transition for different socioeconomic and demographic groups, further insights into people's transport needs are identified. This study will aim to build on this framework and for which the following research question is developed:

How does observing the relationship between the transport network supply and demand provide insights into the transport needs of neighbourhoods defined by their socioeconomic and demographic characteristics in the city of Amsterdam, The Netherlands?

When answering this question, it is expected that the transition is quicker for groups of lower socioeconomic and demographic characteristics as they are historically assumed to have the most transport need. This is because these groups do not have the same options available in comparison with groups of higher socioeconomic groups. The methodology of this research allows for policy makers to identify needs in a city and make network improvements which attempt to solve these problems.

Research Methodologies and Case Study

The methodology for this research is developed and implemented in the form of a case study for the transit network of city of Amsterdam. This is completed as the transport authority of the Amsterdam region (Vervoerregio Amsterdam) identifies the necessity of supporting individuals transport needs in the region. There are four main components of the methodology developed in order to answer the research question.

First, is the clustering of Amsterdam neighbourhoods using income, car ownership, and family composition data. This step is completed using a Latent Class Clustering Analysis (LCCA) which defines clusters of neighbourhoods that maximizes the similarities of characteristics within and the differences between clusters. The distance to the centre as well as urbanity of the neighbourhood are added as covariates to the model in order to refine the allocation of neighbourhoods to a certain cluster. The result is a characteristic profile that defines each cluster based on socioeconomic and demographic

characteristics which are relevant for the identification of variable transport needs. The clustering sets the basis on which the relationship between supply and demand for differing socioeconomic and demographic groups is analyzed.

The second step of the research defines three supply indicators - *walking coverage*, *supply frequency* and cumulative opportunity accessibility (termed *accessibility* in this thesis) - for the GVB public transport network in Amsterdam (Carleton and Porter, 2018; Deboosere and El-Geneidy, 2018; Wang et al., 2017). General Transit Feed Specification (GTFS) data is used in a GIS network analysis environment in order to calculate the indicators. These indicators are built to reflect the tangible needs within a public transport network as defined by the Inclusion Project (Tovaas, 2020). The *walking coverage* is calculated for neighbourhoods to represent the convenience of a potential traveller reaching a transit stop. The *supply frequency* is calculated to describe the efficiency of the network a potential traveller experiences in their neighbourhood. The last indicator, *accessibility*, reflects the number of people that a traveller may access within 30 minutes by a public transport journey. This last metric is a common metric in accessibility research which incorporates the convenience of moving around the city, the ability for multiple routes and vehicles to complete a trip, and providing ample number of locations and destinations that a person can access. Each of these indicators represent a potential transit need for an individual and could be of conscious or unconscious consideration before a trip is completed.

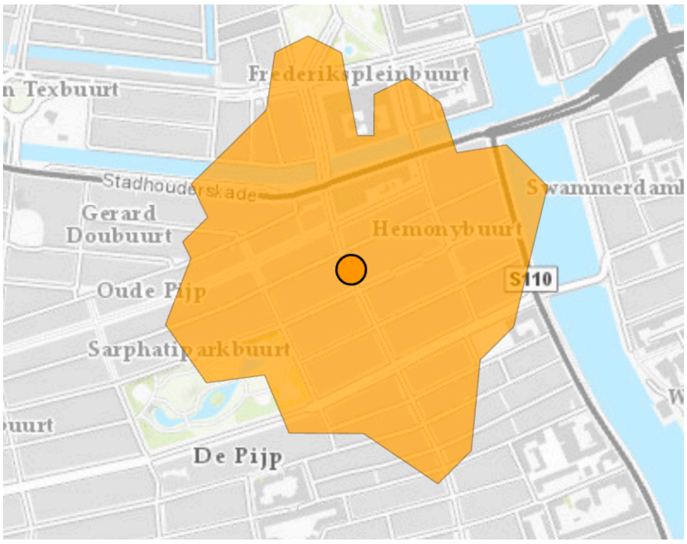

Figure 1: Representation of the walking buffer surrounding transit stop in the city of Amsterdam used within each of the supply and demand indicators.

The third component of the methodology looks at the number of public transport trips made from neighbourhoods in Amsterdam during morning rush hour of June 2019. The demand indicator is a proxy for the use of the transit network and reflects the trip generation of individuals under different quantity of supply indicators of the transit network. This study utilizes public transport smart card data (OV-Chipkaart in the Netherlands) to record the number of trips started at each transit stop across the GVB network. Transit trips are allocated and attributed to the surrounding neighbourhoods using a percentage of the walking buffer from a transit stop that is present within that neighbourhood. The walking buffer represents the area a person can reach after walking 400m. The pedestrian network of Amsterdam is gathered from OpenStreetMaps and the walking buffers are calculated using network analysis in the GIS environment (Figure 1).

These first three steps of the methodology allow for the socioeconomic clusters to be defined, the constraints/opportunities of the transit network to be calculated, and the use of the network to be found. The last part of the methodology finds the relationship between each of these three components. A simple linear is used in order to determine the correlation between each of the supply indicators and the demand of the GVB network. This regression is first completed for all neighbourhoods in Amsterdam and does not yet consider the socioeconomic and demographic clusters for the neighbourhood. This is later incorporated through interaction effects in the linear regression where it is determined whether the correlations of each defined cluster are statistically different from each other. This provides the main crux of the thesis and allows the main research question to be answered and discussed.

Results

Six socioeconomic and demographic clusters are defined upon completion of the LCCA. The different clusters are defined in order to maximize the similarity of characteristics within clusters and differences between clusters. Clusters consisting of one through nine clusters were evaluated through the analyses BIC values (calculating the statistical fit for the model) and the bivariate values, which reflect the statistical dependency between pairs of socioeconomic and demographic characteristics. The six clusters are chosen as this analysis minimizes the BIC value and limits the dependency between the characteristics in each cluster.

The six neighbourhood clusters within Amsterdam each have a defining profile that allows for a comparison between clusters and a platform to test for differing transport needs. Cluster 1 is defined by high income city centre dwellers who have low car ownership. Cluster 2 are low-income families that live on the peripheries of the city. Cluster 3 are high income homeowners who are located within the Amsterdam ring. Cluster 4 is family focused on high density neighbourhoods to the east of the city centre. Cluster 5 are middle income families with a high proportion of young children. Lastly, Cluster 6 is defined by neighbourhoods that, despite living in the city centre, are more car dependent than any other. A visual for the various clusters is presented in Figure 2.

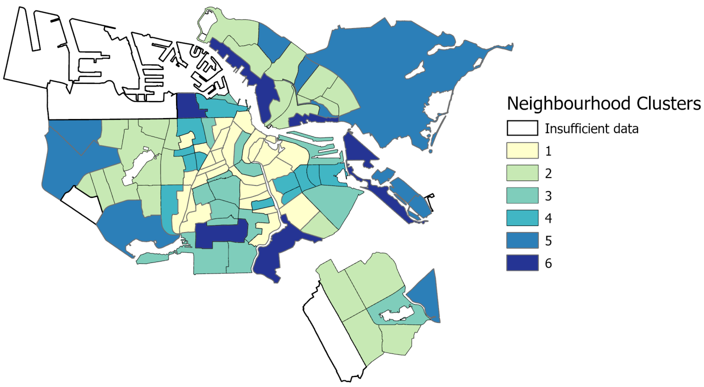

Figure 2: Socioeconomic and demographic neighbourhood clusters of Amsterdam defined through the completion of Latent Class Clustering Analysis

These clusters are evaluated in terms of their supply indicators, demand of the GVB network, and the relationship between these variables. Throughout Amsterdam the majority of neighbourhoods experiences a walking coverage about 75% which represent that most of the population is within 400m of a public transport stop. The neighbourhoods with the highest percentages are located largely within city centre while the neighbourhoods on the outskirts experience the lowest. The supply frequency has a different distribution across the city with only a few neighbourhoods experiencing the largest number of departures transport hubs while the majority of neighbourhoods have a supply frequency that are closer to the minimum value than the maximum value. The distribution in accessibility for Amsterdam neighbourhoods is recognized to have two peaks showing neighbourhoods of higher accessibility (around 300,000 people) and lower accessibility (around 150,000 people). This shows that the accessibility to other people is not homogenous across the city. Certain neighbourhoods, largely located in the city centre, experience higher access to resdients than other neighbourhoods in the city.

The calculation of these indicators demonstrated that GTFS data and the pedestrian network allowed for an overview of how the transport network is reaching the neighbourhoods of Amsterdam. The subsequent use of the network is calculated through the OV-Chipkaart data which demonstrate the most trips starting near the Amsterdam Centraal Station but also neighbourhoods which have metro stations such as Zuidas, Amsterdam Noord, Frankendael, and neighbourhoods located along the Noord-Zuidlijn.

To test the assumptions of the research, these indicators are incorporated into the simple linear regression and interaction effects regression. A significant positive relationship is found between the walking coverage, supply frequency, and accessibility indicator throughout Amsterdam and the demand indicator. This represents a correlation between an improved transit network, in terms of the defined indicators, and an increase in ridership. The simple linear regression model which accounted for the most variance is the supply frequency with an R^2 value of 0.691 while walking coverage and the accessibility indicator have values of 0.146 and 0.197 respectively. This demonstrates that the supply indicators account for some of the variance in the use of the network (69.1%. 14.6%, and 19.7%), but other factors still need to be considered in the overall evaluation of the network.

The interaction effects linear regression, finds that the correlations between the supply indicators and demand indicators are not statistically different when comparing the socioeconomic and demographic clusters of Amsterdam. Only two significant differences are found between Cluster 1 and Cluster 6 in terms of the accessibility and walking coverage indicator with Cluster 6 representing the smallest number of neighbourhoods. These findings were not the expected result and suggests that the convenience and efficiency needs in a transit network, representing two guiding principles of an inclusive transit network, are the same for all of the defined clusters of Amsterdam.

Discussion and Recommendations

By observing these relationships, improved walking coverage, supply frequency, and accessibility are recognized as transit needs for the city of Amsterdam rather than the specific socioeconomic and demographic clusters. The argument is made that a cluster of Amsterdam Zuid-Oost needs the supply frequency, walking coverage, and accessibility similarly to a higher income cluster in the city centre. The positive correlation found between these indicators is expected as an improved transit system intuitively results in an increase of use. However, it is not expected that the clusters of different socioeconomic and demographic characteristics would have a statistically similar relationship based on previous transport poverty research. This suggests that the clusters defined in this study do not show any differences in transport need as they are not at high risk of experiencing transport poverty. It may also suggest the current state of the network is already adequate enough to sufficiently satisfy the convenience and efficiency needs for Amsterdam neighbourhoods. This is important when considering new infrastructure projects which improve any of these indicators. Changes can be made where necessary from a network perspective and not take into consideration *needy* clusters of the city not receiving extra service they require.

Despite these valuable findings there are limitations of the research method that should be recognized. These limitations largely arise from the assumptions regarding the aggregation of socioeconomic data and OV-Chipkaart demand data at a geographic level. The socioeconomic data is tied to the neighbourhood and not attributed to the individual making a transit trip. The same issue surrounding the aggregation of demand data is recognized where trips are allocated to the surrounding neighbourhoods with no understanding of how far a transit stop is accessed and by which mode. A causation between the indicators cannot start to be discussed without truly determining the transport behavior and perception of an individual traveller. These assumptions place a heavy emphasis on the neighbourhood and its socioeconomic attributes and may not accurately portray the complex dynamic nature of a cities transport network and may mask any transit need differences in the meantime.

Following this discussion, a series of recommendations are made for the scientific community in terms of future research and the Vervoerregio Amsterdam in terms of potential policy actions. The method built in the research study allows for significant correlations to be found between the supply indicators and the demand indicator for the city of Amsterdam. However, in order to refine our understanding of transport needs across the city a couple of scientific recommendations are the following.

First, is to focus more on the individual traveller rather than socioeconomic and demographic characteristics largely tied to a particular geographic area. One method to do so is to utilize data capture techniques such as household travel surveys which record the specific movements of individuals. Moving to this method, instead of smart card data allocated to a transit stop, allows a researcher to be certain about the makeup of the traveller and limit assumptions based on the average socioeconomic and demographic characteristics of surrounding neighbourhoods. Travel surveys would also allow for the collection of information about the perception of the transit network by the potential transit user. In this study the indicators are built as a proxy for convenience and efficiency of the transit network. Focusing rather on the actual perception, increases the validity of the results in this study. Such techniques also allows for intangible insights such safety and empowerment to be considered. Therefore, more than two guiding principles are assessed when determining the transit needs of individuals within a geographic area. Future studies should also build on the LCCA method incorporated in order to classify the socioeconomic and demographic clusters. Further work can be completed on other case studies, compared against more heuristic methods, and whether these two options can support each other in building an accurate picture of the differences of transit users in a study area.

The thesis was conducted in order to provide insights for the Vervoerregio Amsterdam. Based on the findings a few recommendations are made for policy makers as they continue to develop the transit network of the Vervoerregio. It is recommended to include the supply indicators in monitoring of the region as well as potential forecasting within the network. Through the positive correlation between the supply indicators and demand indicator it can be argued that improvement projects which improve the walking coverage, supply frequency, and accessibility results in increase ridership on the network. The indicators should be monitored over time to determine whether the relationship remains consistent for socioeconomic clusters or whether certain *needs* arise as the network continues to develop. Further definition of which socioeconomic and demographics need to be monitored may also improve the effectiveness in monitoring the city and implementing changes that truly improve the transit needs of all groups of people. The last recommendation states that the Vervoerregio should analyze the distances transit users travel in order to reach a particular transit stop in the network. The access and egress distances is a noted limitation of this study and understanding this further may provide insights into the distance and time people consume to access a *quality* public transport stop.

To summarize, the research provides a novel methodology to investigate the transit needs of different socioeconomic and demographic characteristics. This is completed through observing the relationship between the network's supply and demand for clusters within Amsterdam. Transit *needs* are determined to be equal in terms of the indicators chosen which has implication for further scientific research as well as implications for policy.

Contents

List of Figures

List of Tables

1

Introduction

1.1. Motivation and Background

An inclusive and sustainable transport network allows people the opportunity to access their daily needs. People on average spend an hour each day travelling to their desired destinations in the Netherlands. These trips may involve travelling to work, completing groceries, or visiting their social circles. There are a number of reasons why a person wants to travel, and a transport network should be built to facilitate these desires. However, despite the best efforts of transportation researchers and planners, an individual's intended destination cannot always be facilitated by their available transport options. This discrepancy between one's transport desires and their accessible transport possibilities is termed as transport poverty. It is suggested that the current transport system is built for an idealised group of citizens and fails to cater to the specific needs of all individuals (Tovaas, 2020). If true, a certain segment of the population can experience the negative effects of transport poverty greater than the idealised individual. These negative effects include prolonged travel times, lack of access to job opportunities and groceries, and feeling excluded from the rest of society.

A recent systematic analysis supported by European Horizon 2020 called the *Inclusion Project* recognized guiding principles in order to build an inclusive transport network that works to reduce transport poverty (Tovaas, 2020). These principles are both tangible and intangible principles and focus on the needs of the traveller in mind. The tangible principles include the physical accessibility of the transport network such as step-free access, affordability for all users, convenience via the distance to a stop, as well as the efficiency of the network once in a vehicle. The intangible principles focus on the empowerment of taking a trip, the necessity of offering a helping hand, gender equity, and safety while using the transport network. Each principle plays an important role in improving current networks and developing new transit systems into inclusive transport systems. What remains a challenge is the principles do not have equal importance for all people using the transport system. Therefore, any future initiative is at risk of causing or perpetuating transport poverty for those desiring to use a transport system, especially of concern for people that fall outside the idealised group of citizens. (Tovaas, 2020)

The key in these guiding principles is the focus on the human traveller and their needs of a transport network. This is a move away from classical transportation engineering which focused on automobile networks and the speed and time to travel from point A to point B (Clifton and Moura, 2017). Accessibility research is the branch of transportation research which shifts away from this paradigm and attempts to highlight the desire behind every human traveller. The efforts are instead focused on a transport networks ability to support the needs of the different socioeconomic groups instead of focusing on the idealised user group. Accessibility researchers have conducted extensive research in order to define methods that identify areas within a population of potential transport poverty. A common methodology compares the supply of the transport network with the need for transport within a neighbourhood. The defined need for public transport is largely assumed from socioeconomic and demographic characteristics such as low-income neighbourhoods, lack of car ownership, and areas of young and/or old people. A recent analysis completed by the Central Bureau for Statistics Netherlands (CBS) dove into the socioeconomic and demographic variables that were most likely associated with transport poverty in the country (Central Bureau for Statistics, 2017). This analysis was completed as transport poverty is a growing area of concern for governmental agencies. If exacerbated, transport poverty results in social exclusion where an individual does not fully participate in, contribute to, and reap the benefits that a society has to offer. The CBS methodology developed risk factors involving the aforementioned socioeconomic and demographic risk groups and identified neighbourhoods in the cities of Utrecht and Heerlen which fell under the high risk category. They provided a further call to action for other cities in the Netherlands to analyze whether there are pockets of transport poverty in a similar fashion. This provides an opportunity for the nations transport authorities to follow the guiding principles and develop infrastructure which provide solutions for transport poverty in these areas.

Finding the disparity between a transport networks supply and the transport need is at the crux of the transport poverty definition. Results from past accessibility studies frequently recommend transport operators to increase the transport service where this disparity is the largest (Carleton and Porter, 2018; Currie, 2004; Fransen et al., 2015). It remains unclear whether this solution is effective in supporting the needs of a potential traveller and facilitate their use of the transport network. Most of the uncertainty arises from the misunderstanding of the travellers needs in respective to the inclusive transport guiding principles presented by Tovaas (2020) and not just the idealised group of citizens that it is built for. A recent paper by Clifton and Moura (2017) attempts to build on the concept of need for a transport system through the idea of latent demand. Their work provides a methodological framework for researching and understanding unmet needs that people experience through their transportation network. The term *need* is loosely defined as travel that is desired but unrealized because of bestowed constraints (Clifton and Moura, 2017). If these constraints are reduced, latent demand can be realized and manifest as effective demand (use) of the system. The latent demand for the network needs to be considered in order to build a sustainable transport system that reduces transport poverty (Clifton and Moura, 2017). Latent demand is broken down into two categories: redistributed demand and generative demand. The former concerns relative changes to travel constraints and the rearrangement of the use of transport throughout the city. Generative demand takes into consideration exogenous factors on top of travel constraints such as social, economic, cultural, and technological factors (Clifton and Moura, 2017). Building the understanding of generative demand is the least understood but most significant aspect of latent demand in order to reduce the potential for transport poverty. Through evaluating the manifestation of effective demand from latent demand, it is possible to uncover the travelers needs of a transport system, instead of making assumptions as found in previous research.

A recent case study by Curtis et al. (2019) in Gothenburg, Sweden attempted to identify areas of latent demand as proposed by Clifton and Moura (2017). The work utilizes an accessibility indicator as a tool to uncover areas of latent demand for socioeconomic and demographic clusters within the city. The indicator uses the number of jobs and people reachable within 30 minutes of transport as a proxy for the ability to participate in society through the transport network. Curtis et al. (2019) findings discovered a strong relationship between the accessibility indicator and the public transit modal share in Gothenburg. More importantly, they recognized that the relationship between transport supply and transport demand is not equal for all the clusters within the city. Their findings highlight multiple components of generative demand which has significant implications for the understanding of transport poverty. First, is the observation that the city consists of multiple clusters of people rather than one idealised group of people. Second, is the observation of more effective demand (public transport modal share) when able to reach more jobs and people in the city. Third, and most importantly, is recognizing that the needs varied across clusters of the city in terms of the accessibility metric. This finding can highlight the varying requirements for the guiding principles of an inclusive transport system. (Curtis et al., 2019)

If operationalized, the work completed by Curtis et al. (2019) is an important first steps in using latent demand as a means of capturing the needs of the individual traveller. Transportation providers and authorities may use these findings to be able to focus their efforts and investments in areas that require the most attention. A key component of this is looking at the needs of different socioeconomic groups amongst a population rather than serving an idealized user group. In doing so, the efforts can focus on supporting the needs of the human traveller and work on reducing transport poverty throughout a transport region.

1.2. Research Gap and Objective

1.2.1. Research Gap

Accessibility research has striven to capture areas at risk of transport poverty through measuring the disparity between a transport networks supply and the transport need of the socioeconomic and demographic groups of a neighbourhood. Past work provides a strong basis, however, the definition of need is largely based on assumptions. The main assumption insinuates that groups of lower socioeconomic status benefit the most from improvements to the transport network in their built environment. However, there is less focus on what is actually needed by these various socioeconomic and demographic groups from their transport network. The guiding principles presented by Tovaas (2020) in the Inclusion Project, provides insights into what constitutes these needs. However, operationalized methods to capture these needs are sparse in accessibility research.

As presented by Clifton and Moura (2017), a main component of combating transport poverty is the transition between latent demand and effective demand of the system when presented with different supply of the system. Clifton and Moura (2017) provides a framework of latent transport demand in order to capture the needs and desires of the individual traveller. Their belief states that the role of the transport planner is to uncover latent transport demand in communities and allow it to manifest as the effective demand for the transport system. This is an interesting development as there is a disconnect within past accessibility research in observing whether the use of a transport network changes when the constraints to the network are altered. This is especially true in relation to exogenous factors of a transport system such as socioeconomic and demographic characteristics. The case study by Curtis et al. (2019) in Gothenburg, Sweden attempted to fill this gap by observing the relationship between an accessibility indicator and public transport modal share for different socioeconomic clusters in the city. However, more work must be completed on analyzing the relationship between the principles of an inclusive transport network, strengthening methods to capture the effective demand of the network, and whether this relationship differs for various socioeconomic and demographic groups. If analyzed, people's needs of a transport network are better defined and improvements to the network can be made accordingly.

In this research, it is hypothesized that the needs of all transport network users are not equal but vary based on their socioeconomic and demographic status. This is why it is important to analyze multiple clusters of people within a city instead of viewing the city as one group of idealised citizens. Clustering techniques differentiate groups across a city but allow for the assessment of responses to transport constraints within groups of similar people. This is valuable when testing the relationship between the transport network and the use of the network in order to better define the needs of different socioeconomic and demographic groups. The case study by Curtis et al. (2019) used an ad hoc method to cluster neighbourhoods of Gothenburg into similar socioeconomic and demographic groups. However, defining and operationalizing a statistical clustering technique is not presented, which is required for the application of future case studies.

These research gaps are summarized as follows:

1. Further definition of the need of a transport network for various socioeconomic and demographic groups to reduce the potential of transport poverty in neighbourhoods.

2. Testing the relationship between the transport supply and the use of the network for various supply indicators which describe the guiding principles of an inclusive transport network.

3. The use of statistical clustering techniques that allow a city to be viewed as multiple socioeconomic and demographic clusters instead of one idealised group of people.

These research gaps motivate and support the objective of this research study which is outlined in the following section.

1.2.2. Research Objective

There are two main components for the objectives of this research study. The first component surrounds the domain of accessibility research and strengthening the definition of *need* within a transport system. A framework is provided that helps define the need of a transport system as the latent transport demand where travel is desired but cannot be realized because of transport constraints. The objective of this research is to move away from assuming transport need and instead defining need by observing how latent demand manifests as effective demand for a transport network. This is completed by analyzing how different socioeconomic and demographic groups use the transport network when presented with varying qualities of the transport network. Once completed, recommendations can be made that minimize the discrepancy between the supplied transport network and what is needed from the transport network. Minimizing this discrepancy is essential in reducing the potential for transport poverty in areas within a transport network

The second component of this research concerns the case study that the research is conducted on. The study completed by Uitbeijerse et al. (2019), and subsequent political pressure, has made the topic of transport poverty a priority for municipalities and transport authorities of the Netherlands. The Vervoerregio Amsterdam, one of the nation's transport authorities, is requiring methods to recognize areas of potential transport poverty and provide solutions which limits the negative effects. The goal of this research is to build a methodology and case study to support the goals of the Vervoerregio Amsterdam to reduce transport poverty in the Amsterdam region. This is supported by a host of data available for the socioeconomic and demographic characteristics throughout Amsterdam as well as data on the public transport network supply and demand. It is the goal to use this data and support the requirements of reducing transport poverty within Amsterdam. The analysis is completed by understanding the different socioeconomic and demographic clusters within the city of Amsterdam and focuses on understanding the needs of these groups in terms of their transport network. A better definition and quantification of this need can aid the policy makers of Vervoerregio Amsterdam in predicting the effects of future transportation investments in the city of Amsterdam. Once implemented, this methodology can provide a future framework for other cities and regions in their analysis of transport *needs* amongst its residents.

1.3. Research Questions

The research objectives provide the aim for this research study. The research questions are defined in order to complete these objectives. They are defined as follows:

How does observing the relationship between the transport network supply and demand provide insights into the transport needs of neighbourhoods defined by their socioeconomic and demographic characteristics in the city of Amsterdam, The Netherlands?

The main research question is supported by four sub-questions:

1. What are the socioeconomic and demographic clusters of Amsterdam neighbourhoods based on available census data using a statistical clustering analysis technique?
2. What is the quality of the transit network, reflected through relevant indicators, for neighbourhoods within the city of Amsterdam?
3. What is the recorded demand of the public transport network for each neighbourhood within the city of Amsterdam?
4. What is the relationship between the transport network supply and network demand for Amsterdam neighbourhoods in respect to their defined socioeconomic and demographic cluster?

1.3.1. Research Expectations

These research questions and objectives are built with expectations in mind. These expectations are based on the definition of transport need frequently reiterated in literature. The researcher's interpretation of the assumptions surrounding need is that certain socioeconomic groups have more to gain than other socioeconomic groups from improvements to their neighbourhood's transport network. The groups to benefit the most are those at greatest risk of transport poverty, defined by the CBS as groups that are low-income, do not have a car at their disposal, and are either of young or of old age. (Central Bureau for Statistics, 2017)

This study centres the transport need instead within latent transport demand and the transition to effective demand of a transport network. The argument is made that a groups transport needs are reflected in this transition. For example, if a segment of the population truly needs improvements to the accessibility of jobs, they are more likely to use the transport network if presented with improved accessibility. Further, transport needs are assumed to vary between socioeconomic and demographic groups, meaning that the utilization of the network depends on the affiliated socioeconomic and demographic group of the traveler. Therefore, the relationship between supply and demand is expected to be different and stronger in lower socioeconomic and demographic groups. If found, this has significant policy implications as authorities and operators can focus system improvements for neighbourhoods of cities with greater *need*. These transport improvements would support neighbourhoods in *need*, reducing the chance of further transport poverty, and ensure a significant return on investment as ridership is expected to quickly increase. Identifying the relationship between Amsterdam's public transport supply and demand, relative to neighbourhoods socioeconomic and demographic characteristics, is at the core of the research study.

1.4. Thesis Structure

The research study is divided into the following chapters. Chapter 2 consists of the literature review where the conceptual model for the research study is built in an effort to answer the series of research questions. Chapter 3 presents the methodology that arises from the developed conceptual model. The methodology is presented for a general scenario so that it can be applied to other cities in the future. Chapter 4 contains the description of the Amsterdam case study. It outlines the case study and how this specific scenario is implemented within the research methodology. Chapter 5 is where the results of the methodology implemented within the case study are presented. Chapter 6 is where the results are discussed, the answers to the research questions are presented, as well as limitations of the assumptions that were experienced during the research process. The study is completed with Chapter 7 where the research is summarized and recommendations for future research as well as policy decisions are provided.

2
Literature Review

Classical transport engineering analyzes the transport networks through travel times and the manner of getting from point A to point B. However, this approach fails to include the user's needs which may be reflected in the design of the networks. A more modern approach attempts to highlight the needs behind every human traveller. This section analyzes and evaluates past research surrounding the definition of transport needs and the identification of transport poverty within neighbourhoods. Common techniques involving socioeconomic groups, transport supply, and transport use are highlighted and critically reviewed in this chapter. By showing the limitations and benefits of past research, this aids the development of a conceptual model and the methodology for the remainder of this research study.

2.1. Identification of Socioeconomic and Demographic Groups

A key component in transport accessibility research, when analyzing the potential of transport poverty, is defining the socioeconomic and demographic characteristics. These characteristics are used within literature to define the potential need of public transport for residents of a region. Information on these residents and their different socioeconomic groups are usually attributed to the geographic area of their residence. The definition of the geographic area depends on the case study and the method in which information is gathered in the different ways. These geographic areas are commonly termed neighbourhoods, census tracts, as well as traffic analysis zones (TAZs). Multiple census tracts are incorporated into a neighbourhood, while the same is true for TAZs within a census tract. Therefore, TAZs are the smallest granularity and are the most common unit of geographic area for traffic models (Nazari Adli and Donovan, 2018). Neighbourhoods provide an aggregation of the smaller units of area and are a frequent geographic used in accessibility research. (He et al., 2018; Jones and Lucas, 2012; Lucas, 2019)

Information on these geographic areas are usually gathered through the latest completed census where a detailed questionnaire is provided to an area's residents (Behrisch et al., 2017). However, due to poor survey design and unresponsiveness of residence, these questionnaires often fail to provide a detailed overview of the demographic. Another employed option is household travel surveys for a segment of the population that may be collected via a transportation authority, statistics bureau, or similar type of governmental agency. While extensive in information, detailed census and household travel surveys are not completed every year. For example, a study completed by Currie (2004) in Hobart, Australia utilized detailed census data from eight years prior to the completion of the study. This fact may limit the accuracy of a study which analyzes the present day transportation network with socioeconomic and demographic groups defined eight years prior. On-board transit surveys limit these problems by acquiring information that is relevant to the traveller once they enter a transit vehicle. This technique also allows for the direct fusing of passenger data and the transport network data at a particular geographic location. In a study by Agrawal et al. (2017) it is found that this method has its limits as it only represents people that are currently travelling with public transport and it is impossible to capture all passengers through such a method. Based on these cases, it is evident that there are a few methods and considerations made into the identification of socioeconomic and demographic characteristics within a region (Agrawal et al., 2017; Behrisch et al., 2017; Currie, 2004; Gadepalli et al., 2018). Regardless, this step is imperative to identify different socioeconomic groups in the assessment of potential transport poverty and identifying the cities needs in a transport system.

Through the chosen technique, a number of different socioeconomic and demographic characteristics can be incorporated into a study. The selection may involve one characteristic or multiple characteristics depending on the purpose of the study and data accessibility. Currie (2004) chose characteristics including adult car ownership, physical disabilities, vulnerable age groups (young and old), adults with low income, unemployed adults, and students. It is a common technique to take these multiple socioeconomic and demographic characteristics and develop a composite needs index for a case study (Carleton and Porter, 2018; Currie, 2004; Foth et al., 2013; Jaramillo et al., 2012). A composite index is a statistical method that groups together the different socioeconomic characteristics and makes one representative index. A composite needs index does not specify the transport needs of an individual but instead explains where potentially disadvantaged people are located in a city. The underlying assumption in this method is that disadvantaged or vulnerable people are in the most need of a nearby transit network.

Currie (2004) utilized a technique that identifies the population of disadvantaged groups in each neighbourhood and applied a weighting to each of the groups. The result was a score between 1 and 100 that was applied to each neighbourhood in the city. A composite technique is also utilized by Fayyaz et al. (2017) on dynamic accessibility in Wasatch, Utah as well as Carleton and Porter (2018) which describes the challenges in calculating transit equity in a region. A study by Jaramillo et al. (2012) in Cali, Colombia used factors of transport disadvantage from their census statistics such as vehicle ownership, age, disabilities, employment, education level, number of children, and literacy. These characteristics are similar to the chosen characteristics by Currie (2004), but highlight components important and specific for the city Cali with the inclusion of average education level as well as literacy. These inclusions highlight the importance of choosing characteristics relevant to the specific case study. Apart from this, differing to Currie's (2004) approach, Jaramillo et al. (2012) built their composite index through unequal weighting. Each characteristic had a weighting based on their deemed significance to the analysis of transport poverty. In some cases, these weightings had been defined by governmental agencies in the region ahead of time indicating that some characteristics are thought to be more/less in need than other characteristics. However, this weighting is not universal between studies, so no consensus is determined. Regardless, the choices regarding the characteristics and weighting, the composite needs index is frequent technique that is employed to identify areas of potential transport poverty in a transport network.

It is evident, that the exact justification of including certain characteristics is understated in the vast majority of studies involving a composite needs index. A common reasoning for including certain characteristics is their association with transport poverty/disadvantage. The underlying assumption is that people with a physical disability, lower income, etc. are in need of public transport as their alternative transport options are limited. The exact association between these characteristics and transport poverty is provided as a given rather than a supported claim. Another factor may be that the number of meaningful and quantifiable differences between people are limited within the scope of a census or household travel survey. Similar to the inclusion of certain socioeconomic and demographic characteristics, this claim is not supported by a research methodology, but is a possible explanation for this noted inconsistency in common transport poverty research.

A study by Fransen et al. (2015), attempted to highlight the characteristics that differentiate one neighbourhood from another. Their specific selection of characteristics for a composite needs index is completed through a factor analysis in order to determine the most relevant indicators. Each characteristic is not chosen for its relevance to transport poverty but in terms of identifying differences between neighbourhoods. This technique is valuable as it moves past including all available census data that is recognized in past studies. Instead, it attempts to make a composite index thereafter which uses differentiating characteristics throughout the transit network.

On the other hand, there are a series of studies which do not attempt to define a combined need indicator and instead look at one socioeconomic characteristic at a time (Albacete et al., 2017; Deboosere and El-Geneidy, 2018; Ferguson et al., 2012). In the case of the study by Albacete et al. (2017), multiple characteristics were involved, but their relationship to the transport network was assessed one at a time. This technique is usually utilized to determine whether a transport network is more and/or less equitable for one characteristic in comparison with another. These characteristics, similar to when a composite index is established, are largely centred around income, car ownership, family composition, etc. In this manner, the socioeconomic and demographic characteristics do not change but the manner in which they are used in the study is altered.

Another approach represented in the literature uses only one socioeconomic or demographic status for the duration of the entire study (He et al., 2018; Legrain et al., 2016; Liu and Kwan, 2020). Taking one socioeconomic or demographic characteristic for the duration of the entire study means that the effects and relationship to the supply of the network are emphasized and recommendations are made only for this one demographic. This approach is commonly seen when the network is assessed for varying levels of income across a city. Studies involving income are often in North America seemingly highlighting the priority of transport planners to connect its cities purely on an economic basis. In this technique, the varying levels of income groups are often tied to the type of jobs that they are able to access in the city. For example, low-income neighbourhoods are tied to the number of low-income jobs while higher income neighbourhoods are tied to higher income business districts. The studies by Wang et al. (2017) and Deboosere and El-Geneidy (2018) seemingly take the more economic sense when observing potential for transport poverty in their cities.

It is evident that an essential component of transport poverty studies is evaluating neighbourhoods in terms of socioeconomic and demographic factors. However, the exact choice and number of characteristics is not consistent. It occurs, that they must be tied in some way to the established belief that these certain statistics are most likely to be involved within potential transport poverty. Another approach, where studies identified one factor and completed an analysis on this specific group, was able to identify which socioeconomic and demographic characteristics are most valuable in distinguishing one neighbourhood from another. This technique seems best suited to differentiate neighbourhoods rather than just applying a score where differences between the neighbourhoods may actually be negligible. These differing methods are considered when building the methodology in this study in order to evaluate the socioeconomic and demographic characteristics of the neighbourhoods.

2.2. Supply of the System

The socioeconomic and demographic status of the people potentially using the network is the first component. The second component, and the focus of this section, is defining the supply of the transport network within the defined areas of a transport region. Through defining supply of the transit network, researchers are able to suggest whether a certain geographic area provides enough service relative to travelers transport network needs. This is frequently referred to as the transport gap in literature and is operationalized and properly explained by Currie (2004). However, there does not seem to be a consensus on which indicator to use as every study is specific to their study area, purpose, and available data sources. There also does not seem to be a consensus on the justification on which indicators are used in the first place. An overview of the justifications as well as which supply frequencies are frequently utilized are explained in the following paragraphs.

The most common transport supply indicator is a form of cumulative opportunity accessibility indicator which is explained and utilized by Deboosere and El-Geneidy (2018), Foth et al. (2013), and Paez et al. (2009). As explained by Geurs and Van Wee (2004), this indicator counts the number of opportunities that are reachable within a certain time frame or cost of the trip. The rationale behind this indicator is that, when presented with more opportunities, a person is more likely to participate and contribute to the society that offers the opportunities. This indicator incorporates information on the transport network as well as land-use information for the region within the transport network. Therefore, a synthesis of data sources is required in order to complete the analysis.

Transport poverty research also varies in defining what constitutes an opportunity in various case studies (Legrain et al., 2016; Páez et al., 2010). The most common opportunity used in cumulative opportunity accessibility indicators is the number of jobs that are available for its citizens. This type of methodology places a large emphasis on the economic effects of a transport system. As mentioned in an earlier section, this is especially true for American studies where a large emphasis is placed on capturing the jobs that are accessible to low-income households within cities, especially the inner city (Karner and Niemeier, 2013; Legrain et al., 2016). The main goal in these studies is to match residents with the types of jobs that they are most likely to acquire. In these studies, the assumption is made that people within a certain income bracket are only looking to access jobs within their income bracket. As Páez et al. (2010) stated, this poses the risk of having *transport islands* where people of lower socioeconomic status do not have access to the locations they desire.

Besides the number of jobs, amenities within a studied geographic area is another opportunity that is often researched within accessibility studies (Farber et al., 2016). These amenities are locations such as hospitals, grocery stores, retail stores, cafes, and places of leisure. The decision can be either to identify one type of amenity, such as grocery stores for low-income people, or a multitude of trip purposes as done by Currie (2004). What is evident is that these amenities are chosen to represent actions in someone's life outside the obvious work commute. This potentially reflects another perspective surrounding transport poverty in which the transport network should provide opportunities in all aspects of life such as purchasing food, accessing healthcare, or spending leisure time at a café or retail store. A case study by Curtis et al. (2019) incorporated the number of people accessible from the transport network in order to emphasize the social aspect and desires of an individual's social life. This provides another approach and a further example that the definition of the opportunities in a case study varies from city to city.

The analysis of the quality of the transport network is not limited to cumulative opportunity indicators. Indicators may involve other proxies for *quality* such as the number of stops, cost of an average trip in time and money, frequency and route diversity throughout the region, as well as the overall coverage of the network (Bertolaccini et al., 2018; Carleton and Porter, 2018; Wang et al., 2017). All of these indicators are mentioned as important components of a quality transit network. The various metrics are frequently presented as common knowledge instead of being directly relevant to the needs of the individual. A study by Farber et al. (2018) is the only study that discussed how different components of the transport network were perceived by the potential travellers. The study focused on the stated needs and desires of Syrian refugees in Durham, Ontario. These perceptions were gathered through surveys as well as focus groups which describes a much more time expensive method than other studies completed in this domain (Agrawal et al., 2017). This is an interesting change of method which may signify a change in how the needs of people are acquired.

The previously mentioned supply indicators are regularly combined in order to define a composite supply indicator, as done by Ricciardi et al. (2015), Bertolaccini et al. (2018), and Carleton and Porter (2018). However, it is key that the indicator is easily understandable to public transit operators and contain fundamental information about the system and the community it serves. A technique that is commonly observed is the fusing of the data sources surrounding a cities pedestrian network and elements of the transport network. An example is seen in a paper by Carleton and Porter (2018), where the area surrounding a transit stop influences the quantification of the frequency at the stop. This attempts to highlight the combination of elements in a transit network and is recognized as a frequently utilized technique in order to quantify the quality of the transit system.

No matter what supply indicator is chosen the underlying intent is that the quality of the transport network is evaluated. The quality of the network is important in order to evaluate how it supports the needs of its potential users. The manner that certain supply indicators manifest in public transit ridership as well as activity levels in an area are highlighted in the next section.

2.3. Activity and Transit Use

Limited studies have looked into the actual use of the transport network in comparison to its transit supply indicators and socioeconomic and demographic groups. For the few studies that incorporated this component (Allen and Farber, 2020; Curtis et al., 2019; El-Geneidy et al., 2016), the assumption is that a low level of use may reflect present barriers within the transportation networks. In identifying the use of the network, two common methods are utilized. The first method evaluates the public transit modal share within a geographic area. This information is gathered from the travel surveys and were gathered at the city level by El-Geneidy et al. (2016) and at neighbourhood level by Curtis et al. (2019). The public transport share is subsequently compared with the accessibility indicators of the geographic areas, in order to determine whether a significant relationship is found.

The second method looks at the activity levels within a cities neighbourhoods in relationship to the provided transit accessibility, income, and car ownership. Activities may involve grocery trips or meeting a friend across the city. While activities are not inherently tied to public transport, they do reflect people living their life and participating in society. This participation is a key component of reducing the chances of social exclusion. El-Geneidy et al. (2016) found that areas of low activity participation tend to concentrate in the automobile oriented inner-suburbs; poorer, where existing levels of transit service are not meeting the needs of residents. This study highlights the assumption that people with a lower income have the most need of a transit service and should be built on further.

2.4. Temporal Aspects

The temporal aspect of the analysis is important to consider in order to properly capture the state of the network. It is also important to choose a time period, together with the correct spatial aspect, to answer the posed research questions.

In terms of the accessibility indicator, an assumed average travel time is usually within a 30 to 60 minute time frame. The exact chosen time frame depends on the city size of the study. For example, Chen et al. (2018) chose a 60 minute time frame as the metric, as the study was took place in Edmonton, a larger North American city. However, in the smaller European city of Gothenburg, Sweden, Curtis et al. (2019) chose a time frame of 30 minutes.

When measuring the service frequency, the temporal aspect also needs to be considered. A study by Chen et al. (2018) investigated the frequency of busses within an entire week, while a study by Fayyaz et al. (2017) focused on the number of buses within one peak hour during a weekday. These decisions are dependent on whether a broader view of the network is required or whether a micro-approach is needed. Since Fayyaz et al. (2017) looked at travel patterns of morning peak hour commuters, a micro approach is chosen. Additionally, this approach is chosen when home-based trips and the socioeconomic and demographic characteristics attached to the homes are important.

In a study completed by Fayyaz et al. (2017), it became evident that the transit accessibility is not consistent overtime. In their analysis, it was demonstrated that there are temporal variations of service across the day and that usual accessibility measurement tools ignore these fluctuation. If it is a key component of the analysis, it is important to take these fluctuations into consideration. Fayyaz et al. (2017) was able to conduct a dynamic transit supply and need gap analysis over the course of the day. However, if the dynamic accessibility is not a key aspect of a research study, taking an average over a predetermined time window can be utilized to take the variation into account (Fayyaz et al., 2017).

Most of the variation of the temporal aspect refers to the time within a particular day. In the literature, less emphasis has been placed on the time in a year that these analysis have been completed, which suggests that the time within the year has little affect on the analysis. Rather, it should be chosen based on data availability as well as the research question that is investigated. For example, using the summer months may not be effective in building an argument for transit usage and scheduling in order to reach schools as this is usually vacation time when students are not commuting to schools.

2.5. Conceptual Model

The literature surrounding transport poverty and transit networks is evaluated in order to understand past research and how it may support the methodology of this research. When answering the research question it is expected that people of lower socioeconomic and demographic status have the most need in their transport network. The argument is made that a groups transport needs are reflected in the transition between latent transport demand and the effective demand of the network. Therefore, a group with higher *need* will transition quicker from latent demand to the effective demand. In order to evaluate this transition there are three requirements in terms of information and data. First, are the socioeconomic characteristics of the travellers or the neighbourhoods that they live in. Second, are indicators which represent the transport network describe its quality across a study area. Lastly, is the use of the network (demand) by individuals in a study area and in reference to the quality of transport network that they are experiencing. Past research is analyzed in this literature review to show the limitations and benefits of methods utilized by researchers to fulfill these three components. By doing so, the below conceptual model is developed in Figure 2.1 and the methodology may be built to build on these past methods for this research study. The conceptual model is visualized to represent the order of steps taken and mimic the various data layers required to complete such as study. First are the neighbourhood attributes in terms of transit supply as well as socioeconomic and demographic characteristics. Next are the definition of indicators in which these characteristics are represented. Next travel behaviour and transit use may be included in the analysis. Lastly is the completion of the analysis where potential risk for transport poverty is realized or identification of the relationship between the supply and use of the system.

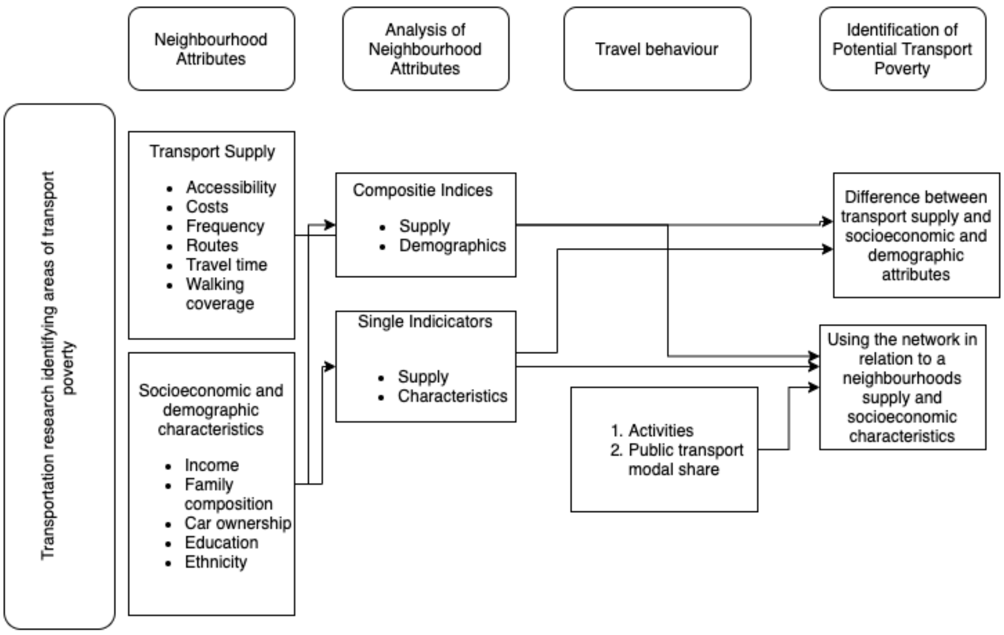

Figure 2.1: Conceptual model for literature surrounding transport poverty analysis

3 Methodology

The goal of this research is to provide insights into the relationship between supply and demand of a public transport network. The methodology is built to identify the relationship between supply indicators and transit ridership for different socioeconomic groups within a city. In the first section of the methodology, the definition of neighbourhoods and different socioeconomic and demographic clusters within a geographic area are described. This is followed by the method in which transit supply indicators are calculated. Third, the methodology to allocate demand (ridership) to the neighbourhoods of the study area is presented. To finalize, the methods in which the relationship between supply and demand for different socioeconomic and demographic groups is presented.

3.1. Geographic study area

The methodology aims to provide a framework that can be applied to any city or chosen geographic study area, in order to provide insights into the relationship between supply and demand of a public transport network. The hypothesis of this research states that neighbourhoods with varying socioeconomic characteristics interact with the built transit network differently in terms of ridership. This assumption is supported by findings that people of lower socioeconomic status, old age, and reduced car ownership are the most to benefit from an increased quality in the public transport network (Currie, 2004; Jaramillo et al., 2012). It is the goal of this research to provide an accurate representation of the neighbourhoods in the study area in terms of these socioeconomic and demographic characteristics. The following section describes the methodology in which the neighbourhoods are clustered and which socioeconomic and demographic data are used for the analysis.

3.1.1. Clustering of Neighbourhoods

A clustering technique is utilized in order to create a profiles for different neighbourhoods in the study area. This is in place of composite need indices frequently observed in transport poverty literature. This decision is made to guarantee the city consists of clusters with multiple like-wise neighbourhoods. This allows the clusters, and their corresponding socioeconomic and demographic characteristics, to have varying levels of transit supply applied to the clusters, and their subsequent ridership to be evaluated. The steps are fully explained in Sections 3.2, 3.3, and 3.4 of this methodology. However, clustering guarantees that there are groups of similar neighbourhoods being evaluated. Cluster analysis is a multivariate statistical technique used to organize observations into groups (clusters) so that the observations within clusters have a high degree of similarity while the clusters themselves are distinct from each other (Pedigo et al., 2011). The clustering of neighbourhoods in the study area allows for the assessment of each cluster of neighbourhoods individually. A Latent Class Cluster Analysis (LCCA) is used as the clustering technique of choice for this research. The goal of LCCA is to allow a discrete latent variable to account for observed differences between a set of observations, so that any associations between the observations become independent (Vermunt and Magidson, 2003). The final goal of the model is to find the smallest amount of classes which can describe the associations between the indicators.

A study completed by Curtis et al. (2019) used ad hoc clustering techniques (mostly from local knowledge) in order to create a profile of neighbourhoods in the city and provide a basis for deliberation as to where to prioritise future investments in public transport. Instead of using the ad hoc technique, LCCA utilizes a probabilistic model which calculates the ideal number of classes and allocates each neighbourhood to the cluster that it has the highest probability of belonging. LCCA is also beneficial as it allows for continuous variables which is not the case for other clustering methods such as fuzzy or k-means clustering. This technique attempts to remove bias and provide an objective view of the different types of neighbourhoods that the study area is able to provide. The mathematical formulation of the model takes on the following form which represents the probability of the observations for a particular cluster.

$$f(y_i) = \sum_{x=1}^{K} P(x|z) \cdot \prod_{m=1}^{M} f(y_{im}|z) \tag{3.1}$$

Each neighbourhood, based off of its own characteristics, has a certain probability of belong to a particular cluster. Where x is the latent variable with K categories, y_{im} are a neighbourhoods characteristics of socioeconomic variable m (M being the number of indicators) and z are the covariates for each neighbourhood. The model consists of two probabilities: one that provides the probability of belong to a certain latent class given the covariate values as well as the probability of observing the neighbourhood characteristics given the latent class membership (Molin et al., 2016). Both probabilities are calculated through a multinomial logit model. LCCA is also beneficial as it allows for continuous variables which is not the case for other clustering methods such as fuzzy or k-means clustering.

3.1.2. Socioeconomic and demographic variables

The LCCA is completed on the socioeconomic, demographic, and geographic studies variables that are attributed to each of the neighbourhoods. As mentioned previously, multiple studies have assessed single variables and their relation to the public transport network. A study completed by Chen et al. (2018) in Edmonton, Canada assessed the majority of elderly people in each of the neighbourhoods in the study area. Another study by Fayyaz et al. (2017) in Toronto, Canada looked at neighbourhoods in the study area in terms of the percentage of single parents. In a study completed by Allen and Farber (2020) which clustered neighbourhoods in Hamilton, Canada socioeconomic and demographic variables described the majority of variation of neighbourhoods characteristics. The aforementioned study by Curtis et al. (2019) took into account multiple variables but clustered the neighbourhoods in an ad hoc basis.

For this study, the clustering of the study area is based on socioeconomic, demographic, and geographic data. Each type of data can be found in Table 3.1 and motivated by the conceptual model presented in Chapter 2. The socioeconomic and demographic data are the observations and indicators for the LCCA. The geographic data of each neighbourhood will be used as covariates and helps predict the class for each neighbourhood.

3.1.3. Selection of clusters – Model Estimation

The socioeconomic and demographic variables presented in Table 3.1 are the input for observations in the LCCA where clusters of neighbourhoods will be the output. It is important to realize that the model will not unilaterally state which number of clusters is ideal for the input observations. Instead the goodness-of-fit indexes and bivariate residuals of each model need are utilized to choose which model best fits the data.

Latent class models with 1 to 9 classes are fitted to the neighbourhood characteristics. Variables are entered into the models as percentages (low income households which is defined differently per

Socioeconomic Variables	Demographic Variables	Geographic Variables
Cars per household	Percentage of people $\geq$ 65 years	People density
Salary per person	Percentage of people $\leq$ 15 years	Urbanity
Percentage of low income households		
Average home value		

Table 3.1: Socioeconomic and demographic variables used in the LCCA for neighbourhoods in the research study area. Geographic variables are added to the model as covariates.

government, people $\geq$ 65 years old, people $\leq$ 15 years old) or as a ratio (cars per household, salary per person (1000 euros, average home value 1000s euros, kilometers to Centraal Station , and Urbanity as the addresses per km^2). The LCCA is completed in LatentGold version 5.1. The criteria to determine the model of choice is the percentage change in the Bayesian Information criterion (BIC) as well as the value of the bivariate values. BIC is a measure of model fit with penalization for additional classes where models with lower values are considered a better model fit. The percentage change in the BIC was compared for each model, selecting models where there is a difference by adding another class (Fairley et al., 2014). The bivariates are a measure of the dependence between observations and provides information on whether the assumption of local independence is met. Bivariate values are estimates of the improvement to the model if direct effects are added to the model and the local independence assumption is relaxed. Both of these factors are considered when determining the number of clusters as the results show in section 5.1.1.

3.2. Supply Indicators

The analysis of the public transport supply is composed of three parts. The three indicators calculated are the *walking coverage* indicator, the *supply frequency* indicator, and the *cumulative opportunity* indicator (termed *accessibility* for the remainder of the thesis). These indicators are calculated at the neighbourhood level of the study area, instead of census tract or TAZ level. This is chosen as the latter two options are determined too small a unit for the purpose of this study. The walking coverage indicator aims to represent the first stage of a public transport journey. This first stage requires the individual to reach the public transport stop that allows the vehicle to be taken. The walking coverage indicator attempts to explain the ease and simplicity of reaching the public transport stop. The second supply indicator, the supply frequency indicator, attempts to explain the second stage of the public transport journey which is the arrival of the bus, tram, or metro. The third stage of the public transport journey is delivering the individual to their final destination. The number of destinations that a person can reach is referred to as the accessibility indicator in the literature.

Each of these supply indicators are chosen to represent the constraint a rider experiences before completing a transit trip. Recognizing and quantifying these constraints are essential in identifying latent transit demand within neighbourhoods. To reiterate, latent demand is defined as the demand that is desired but cannot be facilitated through the public transport network. These supply indicators quantify a subset of the constraints a potential transit rider may experience. The walking coverage describes how far an individual walks before reaching a public transport stop. This value falls under the *Convenient* section outlined by Tovaas (2020) in the Inclusion paper on building an inclusive transport network. Second, the supply frequency describes the waiting time a potential transit rider experiences before the arrival/departure of a public transit vehicle. This indicator falls under the *Efficient* category described within the Inclusion paper (Tovaas, 2020). The last indicator, accessibility, approaches this analysis through the classical cumulative opportunity approach lens. Calculating the number of people reachable, this metric attempts to capture the human desire in accessing people within their social environment. Providing access to as many people as possible is integral within the transit system which is why it is calculated within this study.

3.2.1. Walking Coverage

The walking coverage is described as the percentage of a neighbourhood which is within 400m of a public transport stop. Distances between 300m and 600 m are historically used as an adequate distance to a transit stop (Gutiérrez and García-Palomares, 2008). 400 m is also chosen for relevance to the Vervoerregio which is the region for which the methodology is applied on. The area that is within 400m of a public transport stop is calculated by taking into consideration the street network of the analyzed city. Using ARCGIS, the area around the public transport stop (alternatively called the walking buffer) is calculated of the pedestrian accessible areas. A representation of the transit stop buffer can be seen in 3.1. The walking coverage of each neighbourhood is calculated using the following equation:

$$SC_i = \frac{\sum_{j=1}^{n} A_j}{A_I} \tag{3.2}$$

Where SC_I refers to he walking coverage of neighbourhood (i), A_J refers to the area of the buffer from stop (j) which is located within neighbourhood (i), and A_I refers to the total area of neighbourhood (i).

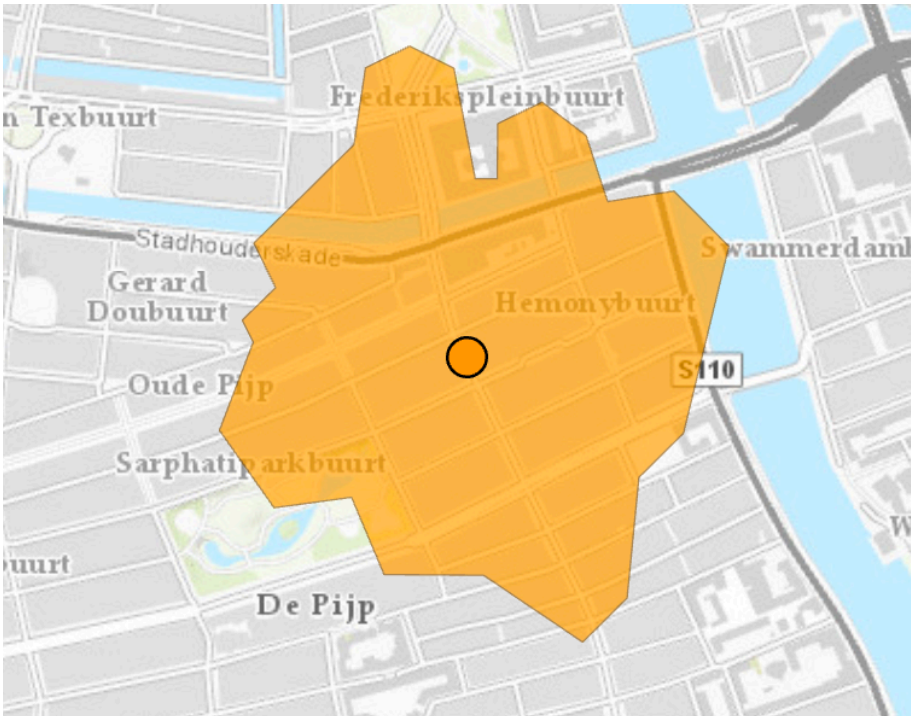

Figure 3.1: A walking coverage buffer area 400m away from a transit stop in Amsterdam, The Netherlands. The buffer is used in the development of supply and demand indicators through GIS analysis

It is important to note that buffers may overlap one stop or more. In this methodology the area of these overlapping buggers is only taken into account once. This removes the chance of a walking coverage being recorded as greater than the area of the neighbourhood. Therefore, the buffer denoted as A_j may be a part of and tagged as belonging to more than one stop. This allows for an accurate portrayal of the entire walking buffer for each transit stop.

3.2.2. Supply Frequency

The supply frequency of public transport is described as the number of transit vehicles that depart from a certain stop per hour and attributed to the surrounding neighbourhoods. It has been shown that an increase of transit vehicles past a stop will increase the use and perception of quality of the public transport network by the public transport users (Fayyaz et al., 2017). In this study, supply frequency is the second supply indicator and provides information on the transit vehicle that the transit user will use. The BetterBusBuffers toolbox of ARCGIS allows the number of transit departures from each stop to be calculated over a certain time frame. In order to reflect the number of vehicle departures from each neighbourhood, instead of the stop level, a method is used which is adapted from Jung and Casello (2019) and Currie (2010). The supply frequency indicator takes into account the number of vehicles as well as the walking buffer that is created 400m around each transit stop in the neighbourhood. The equation in order to compute the number of departures from each neighbourhood is as follows:

$$SL_{NBH} = \frac{\sum_{i=1}^{n} \frac{A_{i_n} \cdot SL_i}{A_i}}{A_{NBH}} \tag{3.3}$$

Where, SL_{NBH} refers to the supply frequency indicator of each neighbourhood, $A_{i_{NBH}}$ refers to the area of the buffer from stop (i) that is located in the neighbourhood, A_i is the total area of the transit stop buffer (i), A_{NBH} refers to the area of the neighbourhood and SL_{B_n} refers to the number of departures from the transit stop.

There are three components of the equation that come to the final supply frequency of the neighbourhood. The first part of the summation incorporates the percentage of the entire transit buffer in each neighbourhood. The second are the number of vehicles from the transit stop. The percentage of the buffer in the neighbourhood allows for the transit stops to be allocated to each of the neighbourhoods it is a part of. The final component of the equation is the area of the neighbourhood and taking the total size of the area into account. The assumption is made that people and land-use are uniform across the entire area of the neighbourhood. It is an oversimplification of the methodology but necessary in order to allocate the transit departures to the neighbourhood the transit stop is in but also to the surrounding neighbourhoods. This is with the understanding that the 400m buffer from a transit stop is not a part of one neighbourhood but could be over a couple parcels of neighbourhood. Buffer space belonging to more than one transit stop is calculated twice which is unlike the approach taken in the walking coverage indicator where each buffer is only calculated once. This approach is chosen as the buffer may be spatially the same from two or more stops but can vary with the number of departures leaving the transit stop over the morning peak hour.

3.2.3. Accessibility

The final supply indicator is named the accessibility indicator. Accessibility is a common metric that is used by researchers to describe the number of places and/or opportunities reachable within a given time frame (Curtis et al., 2019; Farber et al., 2016; Legrain et al., 2016). The number of people was chosen as the proxy for accessibility in this study. This decision was made for two main reasons:

1. The aim of identifying transport poverty is reducing social exclusion within the study area. Someone experiencing social exclusion is not able to reach their jobs but also are not able to reach friends and family to the extent that is desired. The decision is made that choosing both jobs and people as a proxy requires weighting jobs and people as they cannot be identified as the same. For the sake of simplicity, the number of people is chosen as the proxy to highlight the *social* element of the analysis. The validity of this approach is also tested as it is not recognized greatly in the literature where most of the focus is upon the economic aspect of jobs.

2. The study area (described in Chapter 4) is recognized as having a land-use design where jobs and residential buildings are largely intertwined. There are industrial and commercial zones but less than other cities with a large central business district. The mixed-use neighbourhoods, coupled with the ease of access of population data, aids in the decision to use resident data as the proxy for accessibility opportunities.

The time frame that was measured was 30 minutes from the centroid of each neighbourhood that was measured. The 30 minutes includes both the access and egress time from the centroid to the first transit stop and from the final transit stop to the end destination respectively. The access and egress walking times are calculated using the street network of the study area with an assumed walking speed of 1.4 m/s. A representation of the buffer area is seen in Figure 3.2 The number of people is calculated using population statistics of the study area from Central Bureau for Statistics (2017). The total number of people that live in each neighbourhood is known. It is assumed that the distribution of people living within the neighbourhood is equal across the entire neighbourhood and does not take into account the land-use of the neighbourhood. The equation of the number of people accessible within 30 minutes of the centroid is as follows:

$$Access_i = \sum_{i=1}^{n} \frac{A_{iso}}{A_i} \cdot P_i \tag{3.4}$$

where $Access_i$ refers to the accessibility (number of people reached) of neighbourhood (i), A_{iso} is the total area of the isochrome within each neighbourhood of the study area, A_i is the total area of the neighbourhood (i), and P_i refers to the population of each neighbourhood (i) in the study area. This equation is calculated for every neighbourhood (i) for the entire set of neighbourhoods (n).

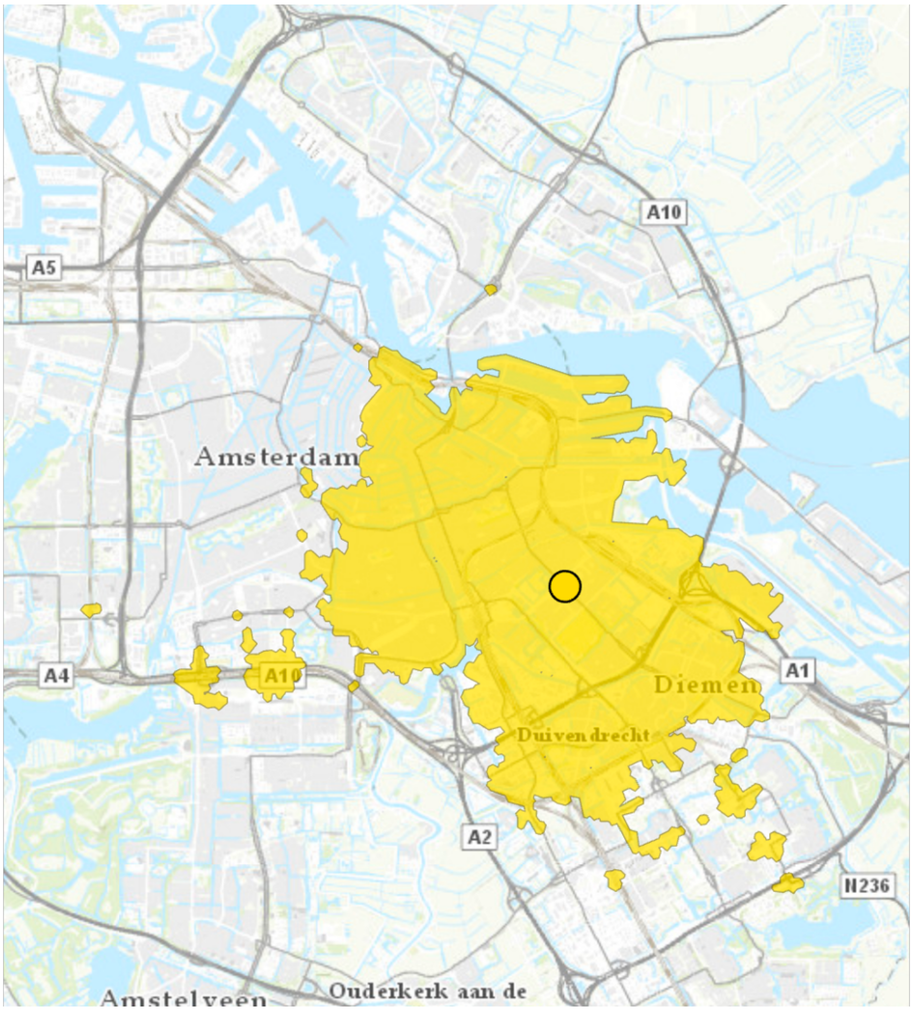

Figure 3.2: Reachable area within 30 minutes of a transit stop in Amsterdam, The Netherlands

3.3. Demand Indicator

The next component of the analysis is identifying the demand for transit in the neighbourhoods of Amsterdam. The transit demand is calculated as the ridership of the transit network from different stops throughout the study area. The precise specifications on the ticketing technology for the study area is described in the case study in chapter 4. The purpose of capturing the transit demand is to identify the effective demand of the public transport network. The effective demand is the result of latent demand manifesting itself once constraints to the transit system are removed. Section 3.4 of the methodology explains how the transition from latent to effective demand is analyzed. However, first the effective demand technique needs is explained within this methodological section.

Similarly to the supply indicators the demand of the public transport network needs to be allocated to the neighbourhoods of the city. Information and demand on the network in this study is provided by check-in data that is collected upon entry of the public transport vehicle. The data is aggregated at the stop level and is provided as the number of check-ins over a given time frame. However, the data that is collected is not attributed to the neighbourhoods that these stops are located. Allocating these trips to the neighbourhood that the stop belongs to, but also for the surrounding neighbourhoods, is an important step in this analysis but also for using the information in regional travel forecasting models. A method used to allocate check-ins to neighbourhoods is provided by a study completed by Jung and Casello (2019). These two studies both use a GIS approach to allocate stop-level boarding and alighting trips into trips from the neighbourhood level. Similarly, to the supply frequency equation seen in equation 3.3, the methodology will use the areas of the 400m buffer around a transit stop. The method can be mathematically represented in the following equation:

$$Demand_{NBH} = \frac{\sum_{i=1}^{n} \frac{A_{i_{NBH}}}{A_i} \cdot D_i}{A_{NBH}} \tag{3.5}$$

where $Demand_{NBH}$ refers to the travel demand of each neighbourhood (i), $A_{i_{NBH}}$ refers to the area of the buffer from stop (i) that is located in the neighbourhood, A_i is the total area of the transit stop buffer (i), A_{NBH} is the total area of the neighbourhood for which the demand is calculated (i), and D_i refers to the transit demand (number of check-ins) from transit stop (i).

A transit stop buffer area may belong to two or more stops as found in supply frequency and walking coverage indicators. The approach is taken the same as for the supply frequency where a buffer area is calculated twice in the calculation. This choice is made as the area is the same spatially but differs in the number of demand allocated towards the varying neighbourhoods.

3.4. Regression Analysis

The previous two sections outline the different constraints/opportunities of a transit network as well as the demand of the network through GVB ridership data. This methodology sets up the constraints and effective component of latent demand framework. The first section of the methodology, the clustering of neighbourhoods, groups like-wise neighbourhoods together and allows for the analysis of similar neighbourhoods. The goal is to identify whether the manifestation of effective demand, relative to the varying levels of constraints, is dependent only on the constraint itself or is also determined by exogenous socioeconomic and demographic variables. It is this final component which is identified by the clusters of neighbourhoods. In identifying the variables behind the effective demand, assumptions may be made on the underlying latent transit demand of these clusters.

In order to identify this relationship two types of regression analyses are completed. First is a simple linear regression which looks at the relationship between the transit supply indicators and the transit demand between all neighbourhoods. Next, interaction effects are added to the regression analysis to take into the exogenous socioeconomic and demographic variables that are described through the neighbourhood clusters. Both of these analyses are described in the following two subsections.

3.4.1. General City-Wide Regression

The research methodology is built in order to assess the relationship between the supply indicators (walking coverage, supply frequency, and accessibility) and the demand indicator. This relationship will be assessed for the city in general as well as determining whether the relationship varies for the different neighbourhood clusters. The implementation of these clusters is explained in Section 3.1.3 with the results provided in Section 5.1.1.

The assessment of the city as a whole is completed using a linear regression analysis. A regression analysis is used to support future decision making in predicting future transit demand provided the set of transit constraints/benefits. The prediction is based on the past correlation between the two variables. The stronger the correlation between the transit constraints and transit demand, the better our ability to predict a value with only one of the pieces of information. This is beneficial to a transit operator to expect the demand of their network given the network that they supply to a city. Each supply indicator is calculated and considered individually rather than a multiple variables in this analysis. The equation is as follows:

$$y_i = b_0 + b_1 \cdot x_i \tag{3.6}$$

Where y_i refers to the demand indicator for neighbourhood (i), b_0 is the intercept of the regression b_1 refers to the calculated slope between the supply and demand indicator, and x_i equals the supply indicator for that given regression. The strength of the model and the general trends between the supply and demand indicators are represented by the R^2 and r value of the model. The linear regression is conducted within SPSS.

It is reiterated that the first linear regression analysis is completed for all neighbourhoods of the study area. This is in place of taking into account the socioeconomic and demographic clusters that are identified in Section 3.1. The neighbourhood clusters are taken into account in the following section through interaction effects.

3.4.2. Interaction Effect Regression

The following analysis utilizes interaction effects to calculate the effects of neighbourhood clustering on the relationship between supply indicators and transit demand in the study area. Interaction effects indicate whether a third variable influences the relationship between the independent supply indicator variable and the dependent transit demand variable. The third variable in our study is the socioeconomic and demographic group for a neighbourhood. This methodology is built to describe the last component of identifying the generative component of latent demand. As discussed, generative demand occurs when previously suppressed trips are realized through a change of transit constraints. These trips are calculated through the transit ridership while the transit constraints are described through the transit supply indicators. The last component is that this the generative demand may be because of exogenous socioeconomic and demographic variables. It is through the interaction effect regression that the exogenous variables and their effect on the regression are explained. If found, insights into the *need* of public transit are realized. These findings may support policy decisions in guarding against transport poverty.

In order to complete the interaction effect regression analysis, each neighbourhood is allocated to a cluster based on the cluster they were allocated to in the LCCA. The analysis is completed in SPSS using the built-in regression analysis. The different clusters are categorized as dummy variables in the analysis where the number of dummy variables is equal to k-1 (k = number of clusters). The first cluster is used as the reference group in the regression and therefore is not included in the analysis. The interaction between dependent variable and clusters is included in the regression by multiplying the independent supply variable and the cluster dummy variable. This is recognized in equation 3.7.

$$y_i = b_0 + b_1 \cdot x_i + b_2 \cdot x_j + b_3 \cdot x_i \cdot x_j \tag{3.7}$$

Where y_i refers to the demand indicator for a particular neighbourhood (i).x_i refers to the supply indicator for that given regression and x_j is equal to the cluster membership of a particular neighbourhood. b_0 refers to the intercept of the regression, b_1 represents the slope and effect of the supply indicator on the demand indicator (similar to the simple linear regression). b_2 represents for the slope and the effect of the class membership on demand when supply is equal to zero. As there is almost always a transit service b_2 most likely is insignificant. However the interaction between the class and supply indicator is calculated in b_3 it is determined whether the slopes on the graph are statistically different from each other.

3.5. Chapter Summary

To summarize, this chapter sets up the methodological approach for the remainder of the research study. There are four main components of the methodology which are built to help identify levels of latent demand through a particular region. These steps include the clustering of neighbourhoods based on socioeconomic and demographic data, calculating transit supply indicators to identify the opportunities and strengths provided by public transit, as well as allocating check-in data in order to determine the levels of transit use in neighbourhoods. Within the methodology of the neighbourhood clustering, the LCCA and the method to choose the most accurate number of clusters is explained. The last component of the methodology utilizes linear regression and interaction effects to determine the factors involved in the transfer between latent demand and effective demand. These factors involve changes to the transit supply as well as exogenous factors through the socioeconomic and demographic factors. This methodological approach in the following chapter is applied to the City of Amsterdam. This choice is made to facilitate the policy needs of the Vervoerregio Amsterdam to monitor for transport poverty in their transport region. The specifications of the City of Amsterdam and how the methodology is applied to the case study is presented in the following chapter.

4

Case Study

The methodology presented in Chapter 3 is structured to be applied to any entity and geographic area attempting to identify areas of latent demand in order to combat transport poverty. However, past studies surrounding accessibility research have largely focused on case studies where one city or area governed by a particular transport authority are analyzed. These decisions are made as cities are unique in their transport networks, unique in the residents using their networks, and unique in the capital available for future investment into the system.

This chapter outlines the case study that is assessed within this thesis. It is completed within the city of Amsterdam and supported by the Vervoerregio Amsterdam. Therefore, the city of Amsterdam is used as the geographic study area. The decision is supported by the extensive data necessary in regards to the transportation network, ridership, as well as socioeconomic statistics available for the city of Amsterdam. In this chapter these various data sources are explained. All these data sources are placed in the context of the GIS software which is used for completing this analysis, ARCGIS.

4.1. The Geographic Study Area

The analysis is conducted as a case study for the City of Amsterdam for the Vervoerregio Amsterdam. Amsterdam is the most populous city of the Netherlands with a population of 844,947 in the 2017 census (Central Bureau for Statistics, 2017). Located within the province Noord Holland, Amsterdam is the main economic driver within the entire Amsterdam Metropolitan Region which homes 2,410,060 inhabitants. Amsterdam is one of the most multicultural cities in the world with high international influence through its recognition as a financial centre of Europe, budding technology industry, and high volume of tourists. This is supported by large transport hubs such as Schiphol Airport. The Vervoerregio Amsterdam is the transport authority which governs the transportation goals and investments for the city of Amsterdam as well as fourteen surrounding municipalities.

The city of Amsterdam is broken down into ninety-nine neighbourhoods. Census statistics are made available by Central Bureau for Statistics (2017) and provide information on the citizens living throughout these areas. The neighbourhoods have an average of 8740 inhabitants. These neighbourhoods are heterogeneous in population with the Bijlmer Oost neighbourhood being home to 28,495 inhabitants. Alternatively, the Bedrijventerrein Sloterdijk, is recognized as the least populated neighbourhood with only 145 inhabitants registered in this area. See Figure 4.2 for an overview of the different neighbourhoods.

The city is recognized as a highly urbanized living space with an average population density of 5,111 inhabitants per km^2. CBS recognizes neighbourhoods in the Netherlands with more than 2,500 inhabitants per km^2 as highly urbanized with a majority of Amsterdam neighbourhoods falling into this category. In order to facilitate the movement of citizens through the city, a vast and highly sophisticated transport system has been developed. Governed by the Vervoerregio Amsterdam, the system is integrated into the neighbourhoods and communities of Amsterdam to bring the residents towards their intended destinations. As discussed in the introduction, it is the goal of the Vervoerregio to gain a better understanding of traveler's interactions and needs with the public transport system.

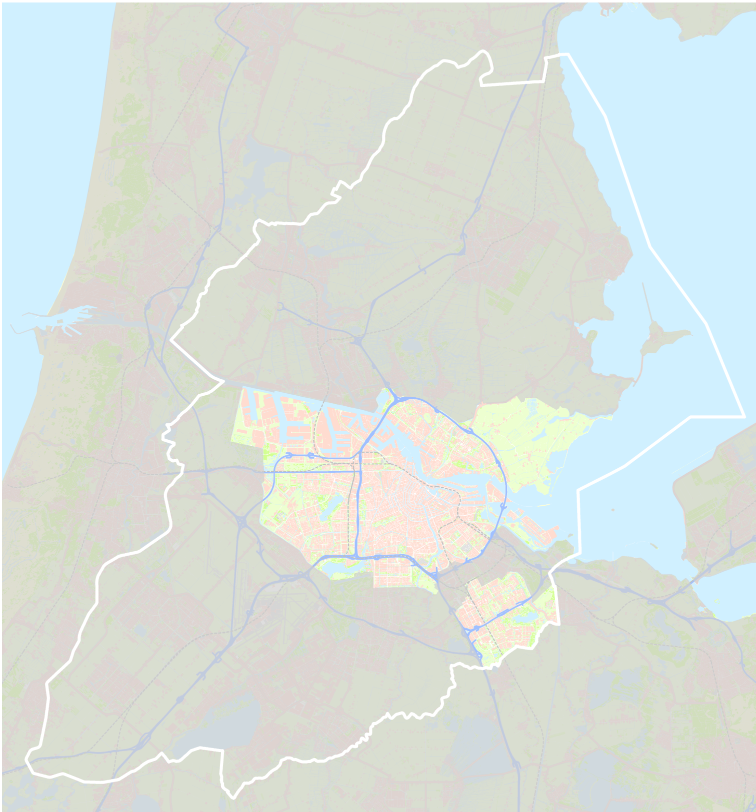

Figure 4.1: An outline of the geographic area the Vervoerregio governs with the highlights city of Amsterdam

4.1.1. Public Transport in Amsterdam

The Vervoerregio Amsterdam governs all modes of transport for the city of Amsterdam and the surrounding municipalities. The Vervoerregio Amsterdam translates the desires of the municipal governments to policies and transport policies. For example, they work to improve "Safe Streets" where the roads are integrated to facilitate pedestrians, cyclists, cars, as well as public transport. The public transport mode is the chosen mode for analysis in this case study. This decision is made as the methodology is built through the public transport accessibility research lens. However, this case study can not be seen in isolation but an integral component of a complex transport system.

The public transport network within Amsterdam provides access to local, regional, national, as well as international connections. International trains arrive and depart from Amsterdam Centraal Station, which is the main transportation hub in the city. The stations daily passengers arrive and depart through the four main transit operators that provide service from this station. Nederlandse Spoorwegen (NS), the main train operator in the Netherlands, provides the national rail service through intercity and sprinter trains throughout the vast majority of the country. Regionally, the Connexxion and EBS bus lines provide transit stops within the city limits (6% of all trips) but also to the surrounding municipalities such as Purmerend, Waterland, and Haarlemmermeer. The main local operator is the GVB, which first began service in 1900 and provides transit service in the city of Amsterdam and also has stops in

Ouder-Amstel, Diemen, Amstelveen, and Haarlemmermeer. The GVB network consists of bus lines, tram routes, metro lines, as well as ferry services which provide travel options across the IJ river. The 46 bus lines and 15 tram routes are on street and provide a mix between separated lanes and integrated with cars. The 5 metro lines are on segregated tunnels as well as elevated track through the core of the city, outer ring, as well as towards Amsterdam Zuid Oost. The transit network has recently seen a shift in structure as the new Noord-Zuid metro line (Line 52) was introduced into some of the most densely populated sections of the Amsterdam City Centre. Completed in summer 2018, the line resulted in a redistribution of the transit network to maximize the potential of the new metro investment by focusing tram and bus lines towards the new metro.

The GVB services provided in the city of Amsterdam, excluding Ouder-Amstel, Diemen, Amstelveen, and Haarlemmermeer, are the transit stops and services that are investigated for the purpose of this case study. The GVB network is at the core of the Amsterdam transit network and consists of a majority of the daily trips of Amsterdam public transit users. The decision to focus on this area was made due to the limited scope of this research as well as access to data and time constraints. To facilitate the research, the GVB provided data on transit ridership. However, the desire is that the methodologies used in the thesis can be extended and applied to the entire transport region at a later time. This is completed with the awareness that the city of Amsterdam presents a different urban density and landscape than the rest of the transport region. The analysis developed strives to explain both high and low density locations that cover the entire area of the Vervoerregio Amsterdam.

4.2. Time horizon

The study is completed for the state of the network in June 2019. The decision is made based on a number of criteria. A month needed to be chosen which represented a *normal* month of the Amsterdam public transport network. Based on recommendations from the Vervoerregio Amsterdam, June was chosen as a month with limited disruptions from school and/or public holidays. Further, June represents a month relatively normal in terms of weather which limits days of inclement weather having an effect on the networks demand. Within the month, considerations are made on when the supply frequency, accessibility, and public transport demand are measured. Walking coverage is not considered as the street network is constant over time and the average walking speed remains the same. The considered indicators are outlined in the following few paragraphs.

The accessibility indicator for the city of Amsterdam is calculated as shown in equation 3.4. The accessibility indicator is calculated for the morning peak hours (7:00 to 10:00) of a weekday in June 2019. The morning peak is chosen as a proxy for the entire day as it is understood that the service did not drastically change across the entire day. The calculations are completed for June 4, 2019 as the posted schedule for June 4 is the same as all other weekdays within the month. In order to account for any variation of accessibility, the analysis began at 07:00 and is calculated at 15 minute intervals until 10:00. The resulting calculations are the average of these 12 time intervals.

The supply frequency indicator is calculated using a similar time frame as for the accessibility indicator. June 4, 2019 is chosen as the day to calculate the number of transit vehicles departing from neighbourhoods in the city of Amsterdam. Measurements are taken for each 15 minute interval between 07:00 and 10:00. A weekday service was chosen as the proxy instead of the weekend schedule. The frequency is calculate for one day instead of the average of multiple days as the posted schedule is the same for each day. The manner in which trips are allocated to neighbourhoods is outlined in equation 3.3 and results presented in Section 5.2.2.

The demand of the network is calculated over a time frame of four weeks within June 2019. These are weeks 23 through 26 of the calendar year. Data on check-ins (entering the vehicle or accessing a metro station) are collected at the stop level in one hour aggregates. The travel demand, as outlined in equation 3.5, is calculated at the neighbourhood level as an aggregate of weekday check-ins in between the hours of 07:00 and 10:00. Upon completion of the case study it became evident that neighbourhoods with train stations of the national rail service (NS) received a disproportionately large number of check-ins during this period of time. Presumably this is a result of travellers egressing from the NS network and subsequently accessing the GVB network. This fact disturbs the assumption of this thesis that trips taken from a transit stop are from inhabitants that reside in the neighbourhoods within 400 m. To rectify this issue the average number of egressing NS passengers from 07:00 to 10:00 are subtracted from the total calculated through the methodology in equation 3.5 (Nederlandse Spoorwegen (NS), 2019). The final demand indicator values are presented as the average of the morning peak hour for weeks 23 through 26. The results are presented in Section 5.3.

4.3. Implementation in GIS

The Amsterdam case study and the designed methodology requires multiple data sources where the geospatial information is retained. Geographic Information Systems (GIS) have a diverse set of tools that allow to store, manipulate, and analyze the spatial and geographic data that is required for this study. The ArcGIS Pro software, as developed by Esri, is the software chosen for this study. The software layers data sources relating to its geography, which allows for a comparative analysis. The different data sources are outlined in the following sections. A GIS system is able to process each data source needed in this study and visualize the findings in an easily communicable manner.

The GIS helps with each step as follows:

1. Dividing the case study area into neighbourhoods as defined by the Gemeente Amsterdam
2. Calculating transport supply indicators based on the methodology described in Section 3.2 and the GTFS data explained in a following Section 4.4
3. Allocating demand data that is collected at the stop level to surrounding neighbourhoods through a 400 m walking buffer area
4. Visualizing socioeconomic and demographic clusters of Amsterdam as calculated by Section 5.1.1 in the methodology.

4.3.1. Neighbourhoods and their Characteristics

A major component of the methodology is analyzing neighbourhoods based on their transit supply and ridership statistics. Data sources available for the transit data is available at the stop level which is described in a subsequent subsection. However, in order to allocate this data and analysis to the neighbourhoods of Amsterdam a strict geographic definition is made for the neighbourhoods within GIS. This neighbourhood information is the first base layer of the GIS analysis in which other data layers are layered upon. Geographic data is available from the city of Amsterdam via their maps interface. Neighbourhood geographic are available as shapefiles which become a polygon data type within the GIS environment (Figure 4.2.

For the city of Amsterdam, ninety-nine neighbourhoods are defined. The spatial coordinates of each neighbourhood are provided by the City of Amsterdam. These neighbourhoods are enriched with socioeconomic and demographic data provided by Central Bureau for Statistics (2017) and are the characteristics outlined in the Methodology. The socioeconomic and demographic data is important as the supply and demand of the transit network in Amsterdam are analyzed in relation it.

4.4. Supply Indicator Data

The following layer of the GIS analysis is regarding the transit supply indicators outlined in chapter 3. Information on the pedestrian network, transit network, and reachable opportunities is necessary to calculate the walking coverage, service frequency, and service accessibility metrics.

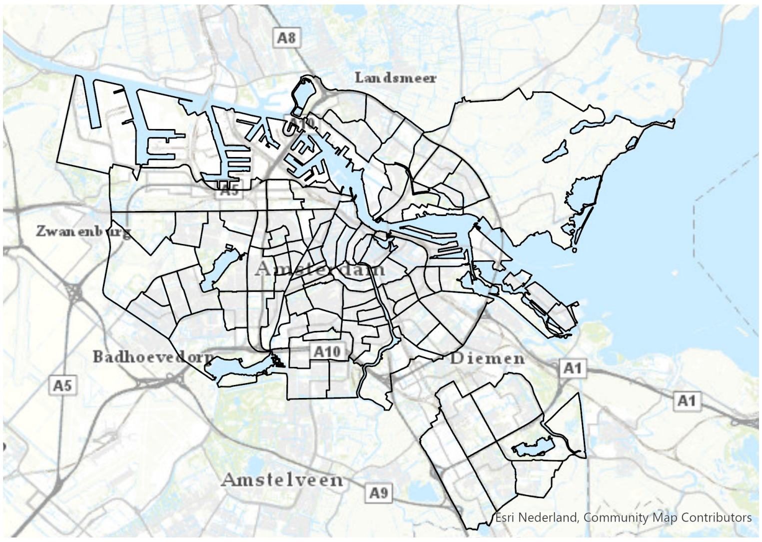

Figure 4.2: Neighbourhood division within the city of Amsterdam.

The first data source processed by the ArcGIS system is the General Transit Feed Specification (GTFS) data of the GVB network. GTFS is a world wide accepted format for transit schedules and its associated geographical attributes (Google, 2020). It is through these geographical attributes that the data can be inputted into ArcGIS for the analysis of the networks supply indicators. In the standardized feed, transit operators provide information such as stops, timetable, transit lines. For this study, the GTFS feed of the GVB was downloaded from ovapi.nl. This GTFS feed included the following files:

1. Calendar_dates.txt - Any services that are not repeated on a certain day are highlighted in this section.
2. Calendar.txt - Provides the service patterns that occur recurrently in a certain day (e.g. weekdays)
3. routes.txt - Identifies distinct routes through the public transport network
4. shapes.txt - The rules in order to draw routes on maps of the transit lines
5. stoptimes.txt - The arrival and departure times of transit services at every stop
6. stops.txt - Identifies the location (coordinates) and stop title of transit stops throughout the netw

Together, these files provide a detailed description of the public transport network for the city of Amsterdam. The public transport stops in Amsterdam are highlighted in Figure 4.3. These represent the nodes on the network at which transit vehicles arrive and depart. The frequency and speed of travelling between nodes is provided by the posted schedule and calculated through the network analysis functionality of GIS for this study.

The network provided by the GTFS data is comprehensive, however, it does not describe the routes people walk to access/egress from the public transport stop. For this reason a pedestrian network of Amsterdam is created in order to recreate the way travellers reach the public transport network. The manner in which this is completed is outlined in the next section, Section 4.4.1.

4.4.1. Pedestrian Network

Public transport is served by the modes in which travellers can access the network. Travellers need to walk to the bus stop from their origin and from the transit stop at their destination. To run the analysis on our transit network the underlying pedestrian network is built. The main source for the pedestrian network is OpenStreetMaps in which the roads of a city are extracted using a QGIS plugin. The shapefile that is made by QGIS can be input into ArcGIS along the public transport network. All roads are assumed to have a pedestrian friendly sidewalk. The roads assumed to have no pedestrian capability are the ring roads around Amsterdam as well as the large roads entering radially into the city. The final pedestrian network can be seen in Figure 4.4.

4.5. Public Transport Ridership Data

Ridership data is used in this study to measure demand throughout the network. In the Netherlands, ticket sales are completed via contactless electronic ticketing via smartcards provided by Translink or ticketing completed at stations. The Dutch smart card, *OV-Chipkaart*, was introduced fully in the Netherlands in 2012. The smart cards used on the GVB network are used throughout the entire country which facilitates a smooth transition between different modes of transportation as well as operators. The *OV-Chipkaarts* uses near-field communication technology to record an active ticket upon entering the network and an inactive ticket upon exiting. A transit rider must *tap in* and *tap out* in order to board and exit a transit vehicle. By tapping, a rider checks-in and/or checks-out of the public transportation network and a trip is recorded in the database. These moments physically occur inside the vehicle for bus and tram and at gates prior to the platform for the GVB metro network. Therefore, there is extensive data on the number of check-ins and check-outs at every station in the Amsterdam transit network.

The check-in and check-out data is collected over four different stages in order to provide a travel history for the individual user. First, the user checks-in or out as previously stated. The second step is the storage of this information at a temporary local location within the public transport operator. Third, this information is stored at a central location of the public transport operator, before the fourth step of sending the information to the national database operated by Translink. It is in this fourth step that transactions are verified and attributed to an individuals account (van Oort et al., 2015).

This travel data allows to perform transportation research at an individuals level. However, keeping user information confidential, the data is masked to ensure privacy. For example, the individual data is stored through an anonymous card ID which only the user has access to through their account. Further, Dutch law only allows data to be kept no longer than 18 months (van Oort et al., 2015). Another fact is that the data is owned by the private transport companies which makes it difficult for the sharing of information between companies. This makes it almost impossible to look at the entire journeys of individuals transfer information between transit operators. This is a major factor in determining that the GVB network is the only network analyzed for the purposes of this study. Once obtained from the GVB, the data is presented in an aggregate number of check-ins and check-outs per station over an hour interval.

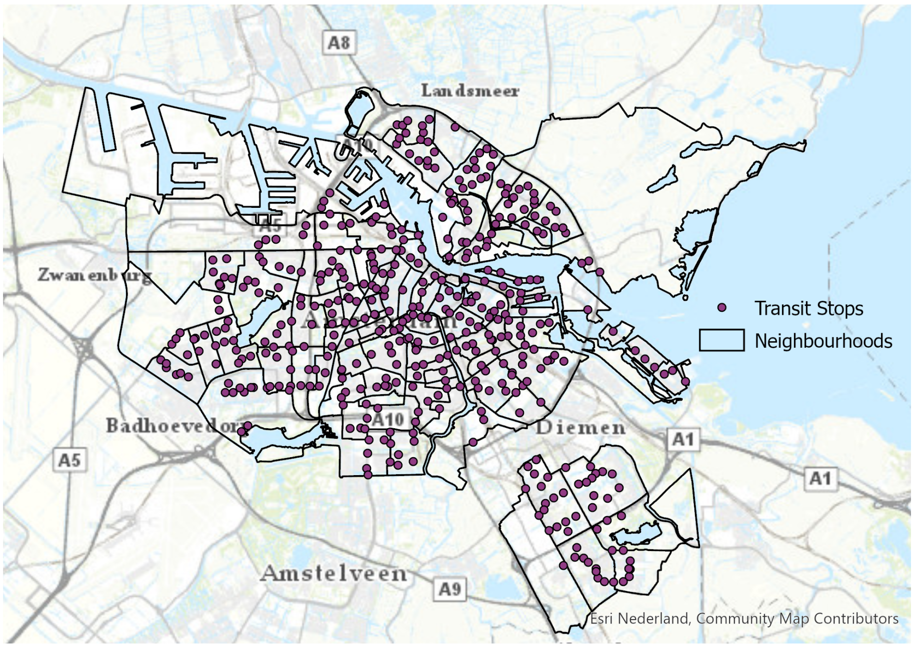

Figure 4.3: Public transport stops for the GVB bus/tram/metro for the city of Amsterdam acquired through GTFS data

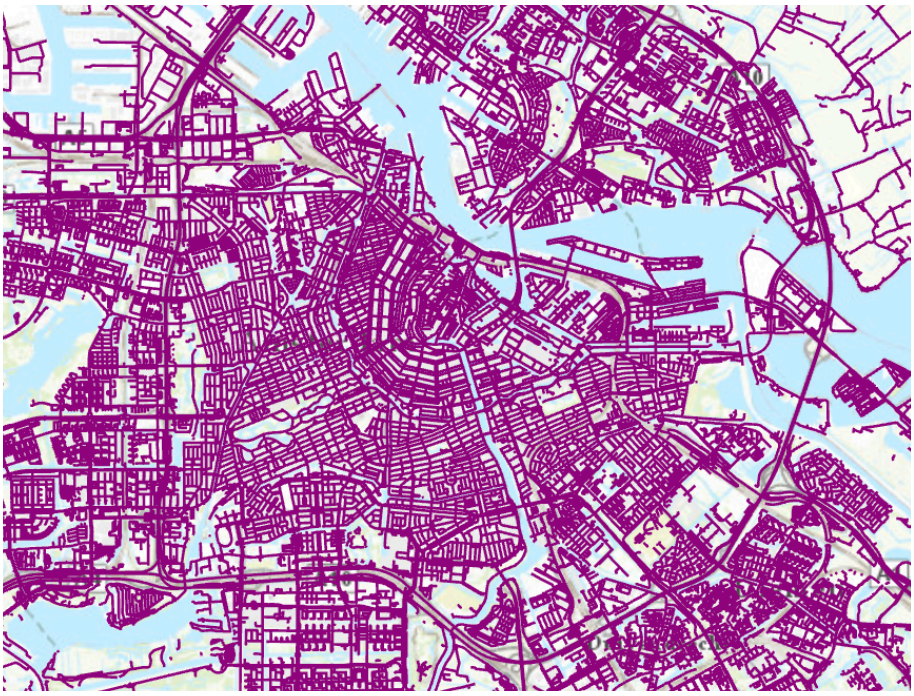

Figure 4.4: The pedestrian walking network of the city of Amsterdam as acquired from OpenStreetMaps

5
Results

In this section of the study the results from the methodology in the context of the case study are presented. First, the neighbourhoods of Amsterdam are placed within their socioeconomic and demographic cluster. The varying clusters are identified and the reasoning behind the total number of clusters through the LCCA is presented. This is conducted first in order to understand first the potential exogenous characteristics of the relation between supply and ridership in the city of Amsterdam. Next, the observations and general trends of the data are presented for the supply indicators as well as the smart card ridership data. The findings are presented per neighbourhood as well as their defined socioeconomic and demographic cluster. The final stage of the results finds the relationship between the supply and demand indicators for the city of Amsterdam and clusters of neighbourhoods. This relationship is provided through the interaction effect regression that was outlined in Section 3.4.

5.1. Clustering

In the first part of the results section the neighbourhoods were described by their three supply indicators: *walking coverage*, *supply frequency*, and *accessibility* (Section 5.2). As well as the demand of the designed network (Section 5.3). The following section of the methodology describes how each of the neighbourhoods is allocated into a cluster based on their socioeconomic and economic characteristics. The result of these clustering techniques is outlined in this following section.

5.1.1. Determination of The Number of Clusters

The LCCA was ran using the Latent Gold version 5.1 software. The model began with no covariates initially in order to determine the number of clusters. This was to eliminate the covariates being a factor in the decision-making process. As stated by the methodology model was ran for models consisting of one to nine clusters. The outputs, BIC values, and corresponding percent change of the BIC are found in Table 5.1.

The BIC value of each model decreases as the number of clusters increases from one through six. The BIC drops initially with a percent change of 9.3% and 10.9% as the number of clusters increase to two and three. As the model increases from five to six clusters the last negative change of BIC (-2.71%) that is noticeably different from 0.0% occurs. The bivariate values for the models have to be minimized as well when choosing the ideal number of clusters. Preferably the value should be below 3.84 in order to maintain the assumption of local independence. As the number of clusters increases the bivariate values generally decrease, as seen by Table A.3. When comparing the models of four through seven clusters bivariate variables no longer changing from double digits to single digits. In the model with six clusters, five out of the sixteen residuals are below the benchmark of 3.84. However, these values are recognized for the pairings of Cars per Household – Low Income Households, Percentage Elderly People – Cars per household, Salary per person – Low Income Households, Salary per person – House

Number of Clusters	BIC	Change in BIC (%)
1	2,064.27	0
2	1,872.26	-9.30
3	1,667.81	-10.92
4	1,635.04	-1.96
5	1,623.31	-0.72
6	1,579.28	-2.71
7	1,579.17	-0.01
8	1,616.35	2.35
9	1,615.64	-0.04

Table 5.1: BIC values and the percentage change during the formation and analysis of Amsterdam clusters

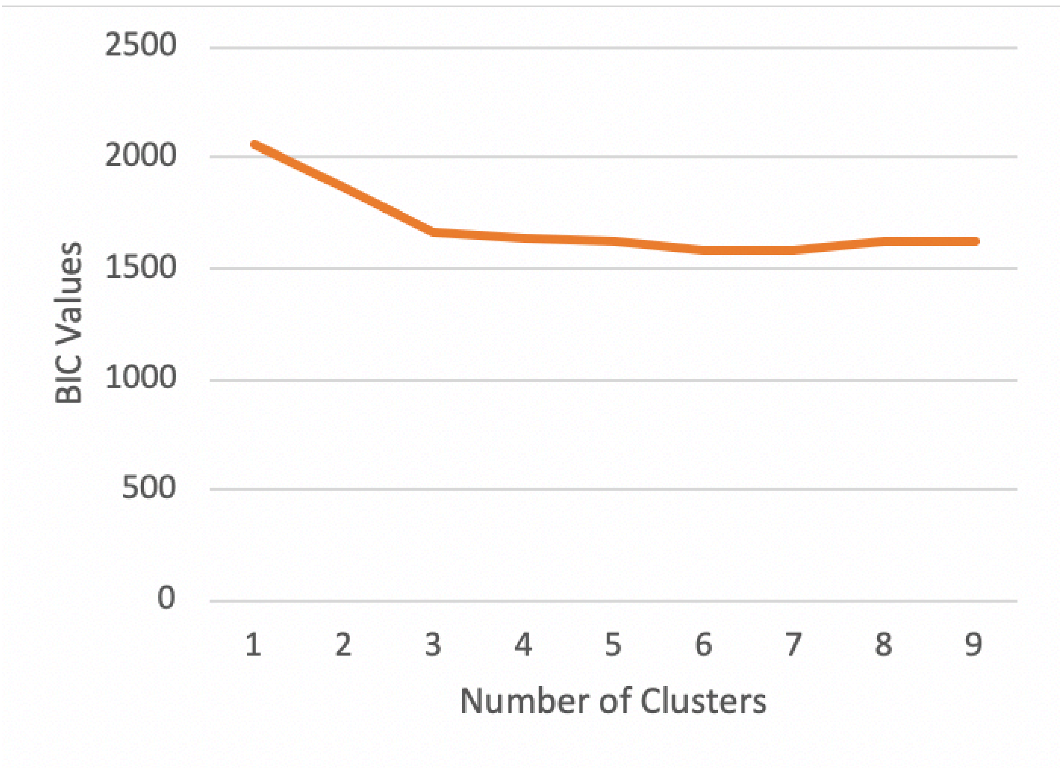

Figure 5.1: Visualization of the change to BIC values during LCCA

Value, and Size of Household – Percentage of Young People. The significant bivariate values for these pairings are deemed acceptable as these relationships clearly have an objective relationship with each other such as salary per person and low income households.

The decision is made to choose the model with six clusters as that is the last model in which a positive percentage (>0.05) change to the BIC is noticed. Table 5.1 reflects that the changes are -0.007. -2.35, and 0.04 respectively for the models with seven, eight, and nine clusters. It is intended that the bivariate residuals are as low as possible (preferably below 3.84) in order to maintain the assumption of local independence between observations. However, there are bivariate values larger than this benchmark for 5 out of the 16 residuals (Table A.4). High residuals are found for the pairings Cars per Household – Low Income Households, Percentage Elderly People – Cars per household, Salary per person – Low Income Households, Salary per person – House Value, and Size of Household – Percentage of Young People. It is clear that these socioeconomic and demographic variables have dependencies between each other. Therefore, it is evident why the local independence assumptions is broken for 5 out of the 15 residual pairings.

Based on the reasoning outlined above, the six cluster model is chosen to represent the ninety-nine neighbourhoods of Amsterdam. Only ninety-five out of the ninety-nine neighbourhoods are allocated to these different clusters. The neighbourhoods of Westelijk Haven, Bedrijventerrein Sloterdijk, Ijburg Oost, and Amstel III/Bullewijk are not clustered as data is not made available for the cars per household. However, in total these four neighbourhoods account for 1000 individuals which is approximately 0.125% of the population. The visual allocation of the neighbourhoods and their clusters is seen in Figure 5.2.

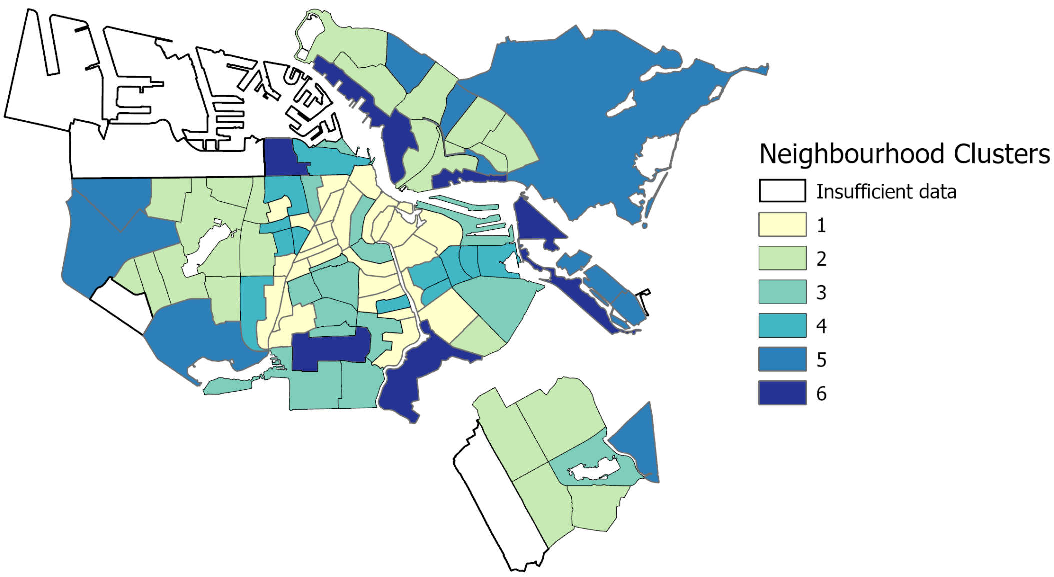

Figure 5.2: Socioeconomic and sociodemographic neighbourhood clusters of Amsterdam

Covariates to the model are next introduced to further explain the clustering of the neighbourhoods and improve the accuracy of the LCCA. Urbanity and distance to Amsterdam Centraal Station were chosen as potential covariates to the six cluster model. Proximity to the city centre as well as the degree of urbanity are highlighted as potential reasons there is an unequal distribution of characteristics of people throughout an urban landscape. They were added as these neighbourhood characteristics can help predict the cluster membership of each neighbourhood more accurately. Future policy recommendations may prove more accurate when presented with information such as degree of urbanity as well as the distance to the city centre. The inclusion of these covariates lowered the BIC value of the six cluster model from 1579 to 1513 demonstrating an improvement to the model and validity in adding these covariates. Further, the inter-dependency of factors within the model did not increase as only the bivariate residual for Degree of Urbanity and Cars per Household is significant. Amongst the observations the residuals remained the same except for the inter-dependency between Size of Household and the Percentage of Children which increased to a value of 10.35. (Table A.4)

5.1.2. Cluster Profiles

The cluster of neighbourhoods are compared in relation to the averages found across the city of Amsterdam. The summary of all cluster characteristics and the Amsterdam average is found in Table 5.2.

	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Cluster 6**	**Average Amsterdam**
Cluster Size (%)	28.35	25	15.79	12.64	11.8	6.42	
Cars per Household	0.36	0.56	0.59	0.36	0.95	1.27	0.50
Salary per Person (1000€)	33.72	20.00	47.37	25.02	29.14	33.28	28.90
Low Income Household (%)	12.59	16.20	7.61	17.79	7.31	11.15	13.70
House Value (1000€)	330.20	186.28	508.47	258.66	341.51	244.33	290.00
Size of Household (# People)	1.57	2.01	1.79	1.76	2.42	1.43	1.80
Young People (%)	11	18	14	14	21	8	15
Old People (%)	11	14	17	10	13	05	12
Distance to Centraal Station (m)	2725.4	5458.7	3944.0	3299.7	6293.3	3986.3	4341.4
Degree of Urbanity (addresses per km^2)	8836.5	3194.8	6087.2	8130.5	1687.3	2244.4	5454.7

Table 5.2: Calculated cluster characteristics the Amsterdam clusters and average defined through the LCCA

- *Cluster 1*. The first cluster is the largest cluster found with 28% of the neighbourhoods in Amsterdam. This cluster is located a majority of the time in the city centre as seen in Figure 5.2. This cluster highlighted by an income that is higher than the Amsterdam average which can also be reflected in the average home value being forty thousand euros. This higher average home value is recognized despite the percentage of families classified as *low-income* being fairly similar to the Amsterdam average. The percentage of elderly people ($\geq$ 65 years) is approximately average while there is a smaller than average portion of the population being $\leq$ 15 years of age (10% versus 15% respectively). The cars per household is also the smallest (along with Cluster 4) of any of the neighbourhood clusters.

- *Cluster 2*. The second cluster is the next cluster of neighbourhoods with 25% percent of the observations falling into this category. These neighbourhoods are relatively further away from the Amsterdam city in Amsterdam Noord, Amsterdam Zuidoost, and Amsterdam Nieuw-West centre coupled with low urbanity. These neighbourhoods are classified by an average salary that is approximately 7 thousand euros a year less than the city average. This is reflected as well in a house value which is 100 thousand euros less than the city average and a percentage of low income families that are higher. The cars per household follow the city average of approximately 0.5 cars per household. Lastly, there are a greater number of young people and old people than the city average as well as a higher size of household than normal.
- *Cluster 3*. The third cluster has the largest mean house value of 500 thousand euros. These neighbourhoods are generally located amongst the inner canals of Amsterdam and spread south towards the Zuidas. The salary per inhabitant is also higher than the city-wide average with half of the number of low income households. The car ownership is also approximate to the city wide average and is similar to Cluster 2 neighbourhoods on the periphery of the city. The size of the household is equal to the Amsterdam average of 1.8 which can be seen in the percentage of young people also being close to the Amsterdam average of 15%. The size of the households can seemingly be influenced the most by having a larger number of inhabitants under the age of 15.
- *Cluster 4*. The fourth cluster has approximately 12 percent of the neighbourhoods with the neighbourhoods being focused in the Amsterdam Oost region. This cluster has the largest percentage of low-income households at 17.79%. Car ownership, salary per person, as well as average house value are below the city average at 0.35%, 25,000 euros, and 258 euros respectively. The average size of the household is only a couple percentage points lower than Cluster 3 and the city average at 1.75 people per household.
- *Cluster 5*. The fifth cluster is the second smallest cluster of 11% of the neighbourhoods but can be categorized by higher than average car ownership, household value, as well as household size. The house value is larger than the Amsterdam average while the salary per person remains similar to the Amsterdam averages. The same as average household size, the percentage of young people to live in these neighbourhoods is higher at 20% than any of the other clusters. As seen in Figure 5.2 the majority of Cluster 5 neighbourhoods are found in the outer parts of the city such as Waterland, far far Nieuwest as well as Driemond.
- *Cluster 6*. The sixth and final cluster is located at different spots throughout the city. These neighbourhoods have the highest car ownership, second to highest salary and percentage of low income families that is closer to the Amsterdam average. Demographically this cluster has below average populations of both young and old people.

Adding the covariates to the model improved the accuracy of the model as stated in the previous section on the determination of the number of clusters. Cluster 1 is the closest to the city centre on average at 2725 metres. Both Cluster 2 and Cluster 5 are significantly further away from the Amsterdam Centraal Station with distances of 5458m and 6293 metres, respectively. The final three clusters (Cluster 3, 4, 6) have scores of 3944m, 3299m, and 3986m. Interestingly these clusters are closer on average to Cluster 1 than Cluster 5 meaning the middle distance clusters are closer to the Amsterdam Centraal Station than they are the furthest neighbourhoods.

Cluster 1, which is closest to Centraal Station is also the highest degree of urbanity at 8836 surrounding addresses km^2. Cluster 5, the furthest away from Amsterdam CS, also records the lowest urbanity score (1687 addresses per km^2) suggesting distance to Amsterdam CS and the urbanity of a cluster are inversely proportional to each other. Clusters 3, 4, and 6, while similarly spaced from to the Centraal Station have a wider range of urbanity scores with values of 6087, 8130, and 2244 addresses per km^2, respectively. This finding suggests that the land-uses and developments are different for each of these three clusters. The final cluster to consider, Cluster 2, has an urbanity that is closest to Cluster 6 with a value of 3194 addresses per km^2. This is despite Cluster 2 being further away from the city centre than cluster 6. These last two statements go against the thinking that urbanity and the distance to the Centraal station are inversely proportional to each other.

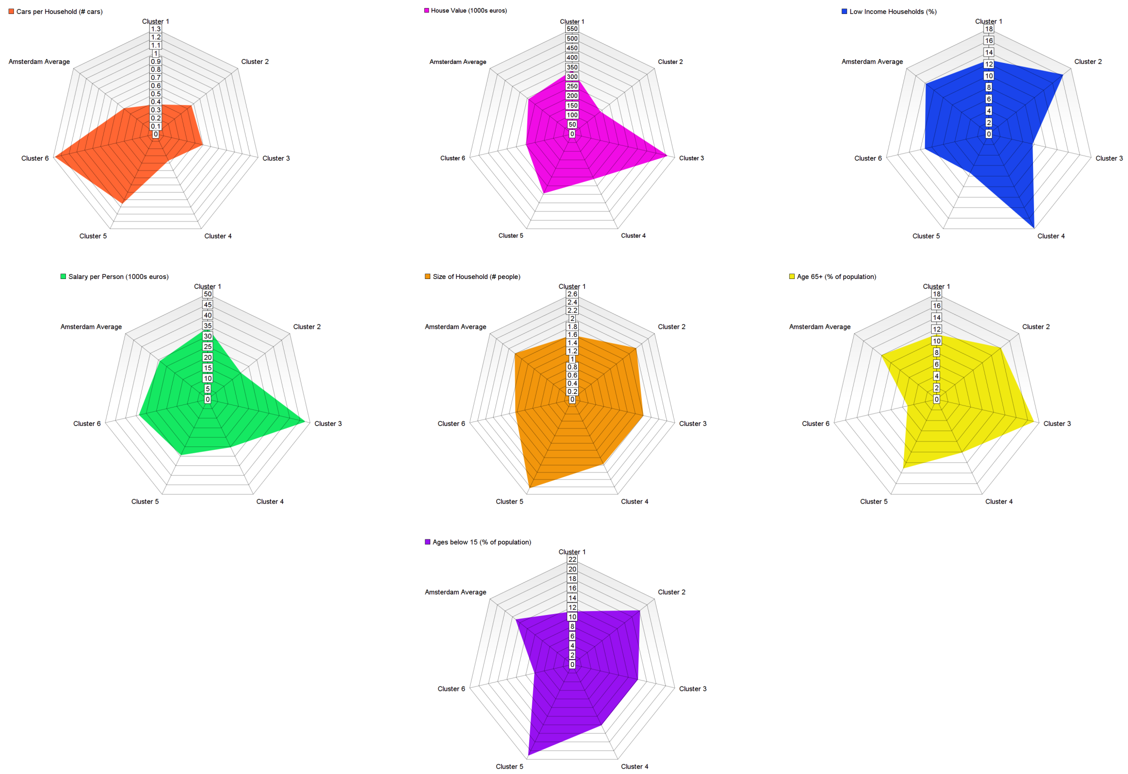

Figure 5.3: Radar graphs showing the relative values for each socioeconomic variable. Cluster 1 is at the top of each graph with Clusters 2 through 6 following clockwise. The Amsterdam average is the last value provided in the each graph. Starting top row to bottom row and moving left to right the characteristics are Cars per Household, House Value, Low income households, Salary per person, Size per Household, Percentage of people older than 65, and percentage of people below 15.

Each neighbourhood of Amsterdam are allocated towards these clusters based on the neighbourhoods socioeconomic and demographic characteristics. The defined cluster of a neighbourhood is the cluster which has the highest probability of matching the characteristics of the neighbourhood.Each of the six clusters has a profile as outlined in the previous paragraphs. Each cluster has a few distinguishing features as a result and are outlined below.

- Cluster 1 (*High Income City Centre Dwellers*): Recognizable by higher than average income in Amsterdam coupled with households that are less dependent on cars. Additionally, these clusters are located more or less in the city centre with a high density of addresses within the neighbourhoods.
- Cluster 2 (*Low-income peripheral neighbourhoods prioritizing family*): Characterized by low home values and family compositions that have more young and old people than the city average. Neighbourhoods are located on the peripheries of the city.
- Cluster 3 (*High house value with low levels of low income households*): The largest house value of any cluster in Amsterdam coupled with higher than average income.
- Cluster 4 (*East ward focussed - High density family neighbourhoods*): Lower income families largely located to the East of the city centre.
- Cluster 5 (*Kid-focussed middle income families*): Highlighted by a high proportion of young people and located in neighbourhoods further from the city centre.
- Cluster 6 (*Car dependent inner city neighbourhoods*): Well off population that is more car dependent than other clusters.

These summaries of the neighbourhood clusters provide an overview of residents living in these clusters. Each neighbourhood has a defining feature, however, it needs to be stated that there is much overlap between the characteristics of each cluster. The next portion of the results is describing the public transit supply that these various types of residents have access to and can utilize to access their daily needs. The supply indicators is the first component necessary to complete the regression analysis outlined in Section 3.2 of the methodology.

5.2. Supply Indicators

The supply indicators are calculated using the methodology outlined in Chapter 3 and applied to the Amsterdam Case Study. The indicators are designed to capture the efficiency and convenience that is deemed important by the Inclusion paper on Transport Poverty. These effects are captured by the Walking Coverage, Supply Frequency, and Accessibility indicators of each neighbourhood. The main results from these indicators are now outlined in the following section.

5.2.1. Walking Coverage

The first supply indicator calculated for the study, represented by equation 3.2, is the walking coverage of a neighbourhood. The walking coverage is the percentage of the neighbourhood within 400m of a transit stop. The neighbourhoods with the most walking coverage are Kinkerbuurt, Geuzenbuurt, Da Costabuurt, Frederik Hendrikbuurt, Staatsliendenbuurt, Chassébuurt, and Hoofdweg (all in the city centre) e.o. which all have scores of 100% (Table A.1). This means that using the GIS methodology used in this study, that the entire neighbourhood is within 400m of a public transport stop. These neighbourhoods are located on the western part of the canal ring radiating outwards towards the highway ring as well as the southern neighbourhoods of the canal ring. In regarding to the clusters described in Section 3.1 the neighbourhoods of highest walking coverage are located in Clusters 1 and Clusters 3.

There is only one neighbourhood, out of ninety-nine Amsterdam neighbourhoods where the population is not within 400m of a GVB transit stop (Figure 5.4). This finding describes that no GVB transit stops are located in the neighbourhood. Additionally, no walking buffer area from neighbouring neighbourhoods fall within this neighbourhood. This neighbourhood is Driemond and is followed by Eendracht, Waterland, Westelijk Havengebied, and Amstel III as the neighbourhood with the lowest walking coverage. Upon conferral with the Vervoerregio, there is a transit stop near Driemond however this is not recognized in the GIS analysis. This most likely occurred during the fusion of data sources which brings to light possible missing data points. Nonetheless, Driemond may still likely fall in the bottom 25% of each of the supply indicators through its location and single nearby bus stop. These neighbourhoods, as well as Lutkemeer/Ookmeer and Bedrijventerrein Sloterdijk, are those in which less than 10% of the neighbourhood is within 400m of a transit stop. These neighbourhoods are located mostly on the peripheries of the city within Clusters 2 and Cluster 5, similar to the minimum neighbourhoods of the accessibility and level of service indicators.

What is striking about the findings is that not all neighbourhoods with high walking coverage are centered in the city center. On average, two-thirds of a neighbourhood are within 400m of a transit stop and the top 50% of walking coverage ranking neighbourhoods have a walking coverage of 73% or greater. These neighbourhoods are found in the low-income family neighbourhoods of Cluster 4 as well as the aforementioned Clusters 1 and 3. The fact that the median of 73% is higher than the mean of 66% demonstrates that on average neighbourhoods are closer to having a maximum walking coverage of 100% than the minimum. This may reflect the Vervoerregios goals of ensuring adequate walking coverage for citizens of Amsterdam. The results are visualized in Figure 5.4.

5.2.2. Supply Frequency

The second supply indicator calculated for Amsterdam was the supply frequency of the neighbourhood. The supply frequency is calculated using equation 3.3 and the entire list of the results can be found in Table A.1 in Appendix A.

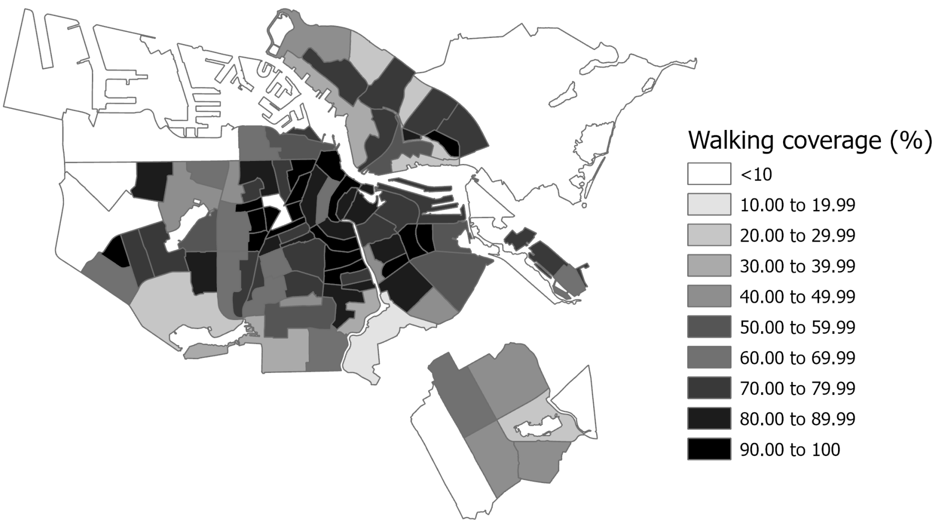

Figure 5.4: Walking coverage for neighbourhoods of Amsterdam measured as the percentage of the neighbourhoods total area.

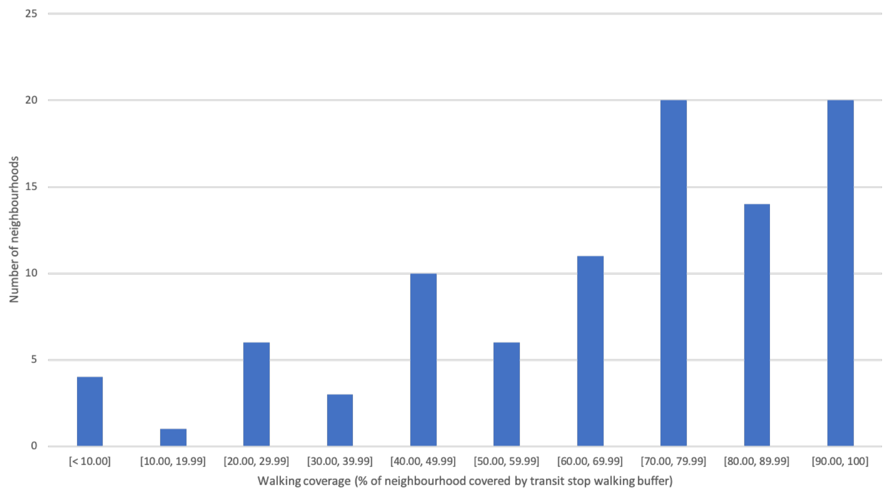

Figure 5.5: Distribution of walking coverage percentages per neighbourhood in Amsterdam

On average there are 123 public vehicle departures per km^2 per hour. The maximum supply frequency is found in the Burgwallen-Nieuwe Zijde with 613 public transport vehicles per km^2 per hour (ptv/km^2/hr) in June 2019. This area is highlighted by the fact that it is close to Amsterdam Centraal Station which is a hub location for GVB bus, tram, and metro as well as other companies such as NS, Connexxion,EBS and international trains which are not taken into consideration in this study. This hub is also close to the Burgwallen-Oude Zijde neighbourhood which is the next neighbourhood in terms of supply frequency with 464 (ptv/km^2/hr). The following three neighbourhoods are De Weteringschans, Indische Buurt West, and the Geuzenbuurt with a calculated supply frequency of 452, 361, and 306 ptv/km^2/hr respectively each falling within the top 25 percentile. De Weteringschans is similar to the Burgwallen neighbourhoods in that there are options for bus, tram, and metro in the neighbourhood. However, this is not the case with Indische Buurt West as well as the Geuzenbuurt in which only bus and trams are available in these neighbourhoods.

The neighbourhoods with the lowest level of supply frequency are those on the outskirts of the city (Clusters 2 and 5) as seen in Figure 5.6. Specifically, these are the neighbourhoods of Kadoelen, Waterland, and Eendracht which all have approximately 1 departure per km^2 per hour in these neighbourhoods. The very low values designates that there is either a very infrequent transit stop within the neighbourhood. However, this also recognizes that some residents of these neighbourhoods may be within 400m of a transit stop in another neighbourhood. This is the most likely reasoning behind the supply frequency value of 1 ptv/km^2/hr. It is also noticed that these are some of the largest neighbourhoods Figure 5.6 respectively. The large neighbourhood areas reduces the significance of a neighbourhing transit stop as very few residents are asssumed to see these stops as a viable transit option. Further, these neighbourhoods either have no or a small number (<2) of transit stops. In the case of the neighbourhoods with no transit stops the recorded transit departure comes from the buffer area that was created by a stop in a nearby neighbourhood.

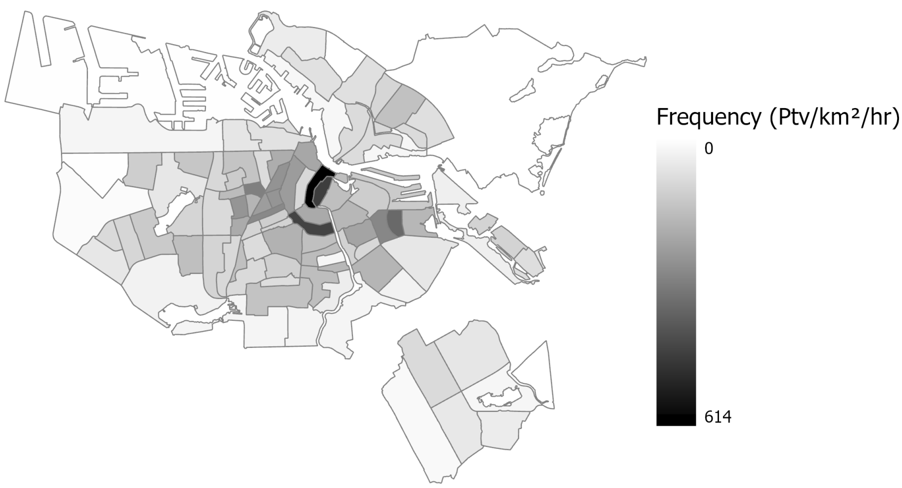

Figure 5.6: Frequency of departures per neighbourhood in Amsterdam

The neighbourhoods with the lowest supply frequency are closer to the mean value of 123 ptv/km^2/hr than the neighbourhoods with the highest supply frequency. This is recognized through the maximum supply frequency values being three standard deviations (103 ptv/km^2/hr) away from the mean neighbourhoods. However, the minimum neighbourhoods are only within one standard deviation of the mean. These results suggest that the very low waiting times associated with the neighbourhoods of the high supply frequency is not the norm. Neighbourhoods which are approximate to the mean are neighbourhoods such as Erasmus Park, Oostelijk Havengebied, as well as Osdorp-Oost are not found within the inner canal ring but closer to the outer highway ring in Clusters 2 and 5. It is recognized that the radiating outward trend of the supply frequency level is present. The supply frequency is highly concentrated at the city center neighbourhoods in Cluster 1 and 3. However, the supply frequency is more concentrated from 100-200 ptv/km^2/hr as the neighbourhoods are assessed going outwards. The level of services are skewed to the higher values as shown by the mean of 123 ptv/km^2/hr and the median value of 99 ptv/km^2/hr. The fact that the Skewness variable is 2.02 also demonstrates that the more neighbourhoods are closer to the minimum value than the maximum value for the supply frequency. Reasons for why this may be the case is outlined in Chapter 6.

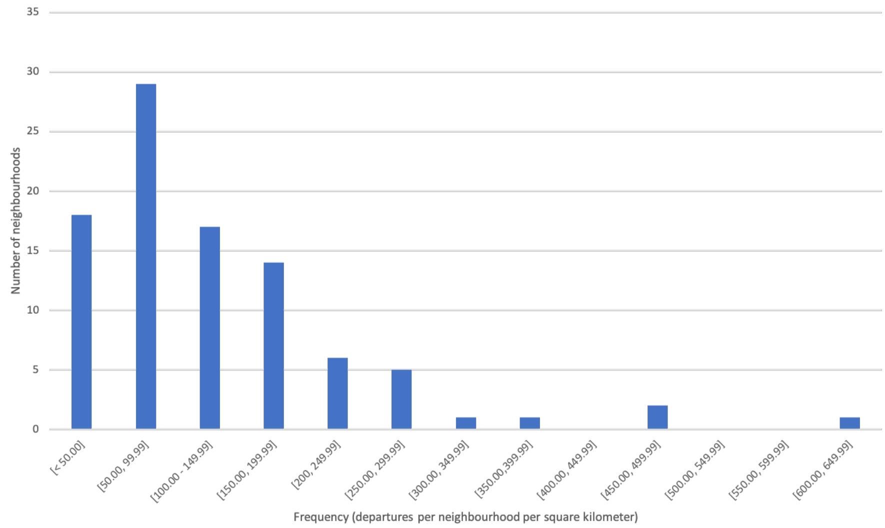

Figure 5.7: Distribution of departures per neighbourhood per km^2 in Amsterdam

5.2.3. Accessibility

The third and final supply indicator was the accessibility of neighbourhoods for the city of Amsterdam. The accessibility indicator is calculated using a cumulative opportunity to other residents measuring from the centroids of neighbourhoods as described by equation 3.4. This is chosen to highlight the social aspect of reducing transport poverty. The entire list of accessibility values can be found in Table A.1 in appendix A. The maximum accessibility value for a neighbourhood in Amsterdam was found in the Weesperzijde neighbourhood with a total reachable people at 484,119, which means that 57.3% of the population of Amsterdam is reachable within 30 minutes of public transport. To conclude the top five neighbourhoods in terms of accessibility is the Vondelbuurt, De Weteringschans, Nieuwmarkt/Lastage, and Grachtengordel-Zuid with accessibility scores of 466,482, 458,229, 427,175, and 424,426 people per 30 minutes of transit travel (ppl/30min), respectively. All of these neighbourhoods are able to reach at least 50% of the population within 30 minutes of a public transport trip.

The neighbourhood with the least accessibility in Amsterdam is the Westelijk Havengebied with an accessibility of 42 ppl/30min. This value is attributed to the lack of public transport services entering this neighbourhood and city of Amsterdam residents living too far away via walking distance. This is similar to the neighbourhoods of Driemond and Waterland where there are no calculated GVB services in this area. It needs to be recognized again that this indicator is only calculated using the GVB network as well as city of Amsterdam residents. There is the potential of more residents being accessibility regionally via nearby municipalities as well as other public transport services. Neighbourhoods which have access to the GVB network, but are still close to the minimum cumulative opportunity levels are Ijburg West and Ijburg Zuid which have accessibility scores of 15,803 ppl/30min and 17,471 ppl/30min which means that 1.8% and 2.1% population are reachable within 30 minutes of public transport respectively.

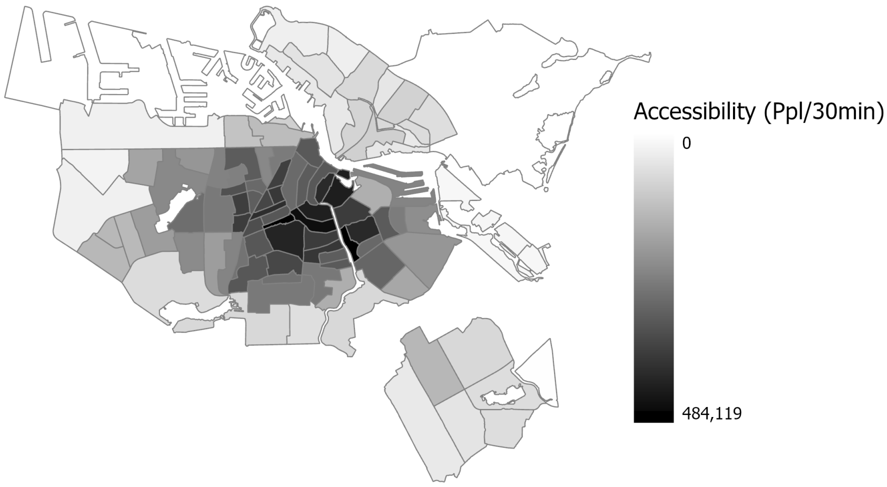

Figure 5.8: Measured accessibility for neighbourhoods in Amsterdam

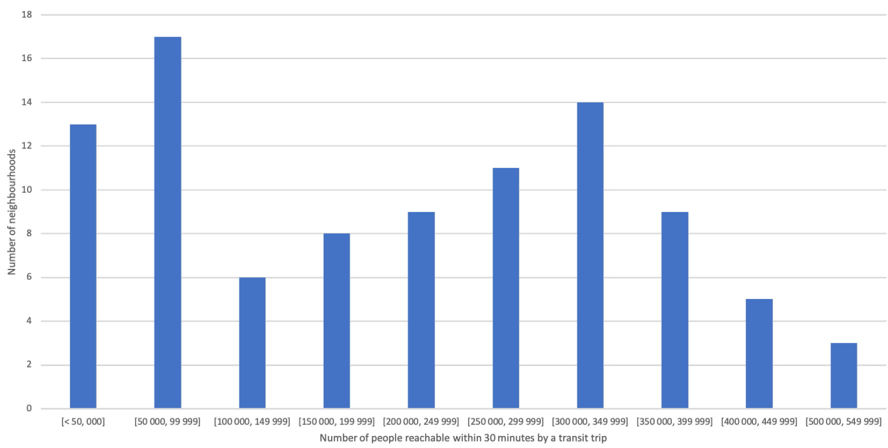

Figure 5.9: Distribution of accessibility per neighbourhood in Amsterdam

The ninety-nine neighbourhoods of Amsterdam have an average accessibility score of 203,683 citizens, or approximately a quarter of the population is reachable by 30 minutes of public transport. However, given the range of 484,076 ppl/30min between the most accessible and least accessible neighbourhoods, with a standard deviation of 136,560 ppl/30min, it demonstrates that the accessibility is variable across the city.

While there is variation of the accessibility indicator throughout Amsterdam there are geographic trends that are evident when looking at Figure 5.8. The neighbourhoods of highest accessibility per km^2 begin in the neighbourhood of Amsterdam Centraal Station and radiate outward. Moving outwards towards the canal ring neighbourhoods are neighbourhoods of accessibility scores of greater than 300,000 people. Visually, the neighbourhoods with the lowest accessibility are at the city limits such as Waterland, Amstel III as well as the Westelijk Havengebied. These trends are clearly general in nature as there are neighbourhoods closer to the city centre which are less accessible than neighbourhoods further away as seen by the Vondelbuurt and Weesperzijde neighbourhoods. This is also evident when observing the neighbourhoods of Landlust, Centrale Markt, Staatsliendenbuurt in Figure 5.8. It is interesting to note though that the metro line of Amsterdam does follow the outer highway ring of the city which may account for the higher accessibility scores such as near Sloterdijk and the Zuidas, as well as Frankendael as the metro line swings back into the city towards the city centre down Wibautstraat.

5.3. Demand Indicator

Calculating the supply indicators for the neighbourhoods of Amsterdam was the first set of indicators that are necessary for the study. Next, the demand indicator is calculated using equation 3.5 of the methodology which signifies how often the network is utilized by people in these neighbourhoods. In order to allocate check-ins at a particular stop, a method is used that is also utilized by Jung and Casello (2019).

The distribution of the levels of transit use can be seen in Figure 5.10. The neighbourhoods with the largest number of check-ins per km^2 are Burgwallen-Nieuwe and Oude Zijde, De Weteringschans, and Geuzenbuurt with 12,810, 7381, 4781, and 4096 travellers respectively. It is noted that the Burgwallen neighbourhoods are the neighbourhoods which are closest to the Amsterdam Centraal Station transport hub. Alternatively, the neighbourhoods with the lowest number of travellers are Driemond, Waterland, Westelijk Havengebied, Kadoelen, and Eendracht. It is noteworthy that these stations are in the bottom five neighbourhoods of the level of service as well as there are either no or minimal amount of stations in these areas.

Looking at the entire city, the average number of travellers from each neighbourhood is 1211. It is also recognized that the median score is at 802 which signifies that the use of the network is more skewed to the minimum scores rather than the very high use areas around the Amsterdam Centraal Station. This congregation of trips can be seen in Figure 5.10. Other points of high demand that are noteworthy is in Amsterdam Noord, the Zuidas as well as other neighbourhoods on the outer ring. These aforementioned locations also are locations of new metro stations from the Noord-Zuidlijn.

5.4. General Regression Analysis

In the literature it was recognized that the walking coverage, supply frequency, and accessibility are important indicators in order to monitor a public transport network. The methodology outlined in the previous explained how the supply indicators are compared with the demand of the transit network. This methodology is conducted under the assumption that an improved public transport network will see an increase of the system. The results from testing these assumptions are outlined in the following section and found in Table 5.3.

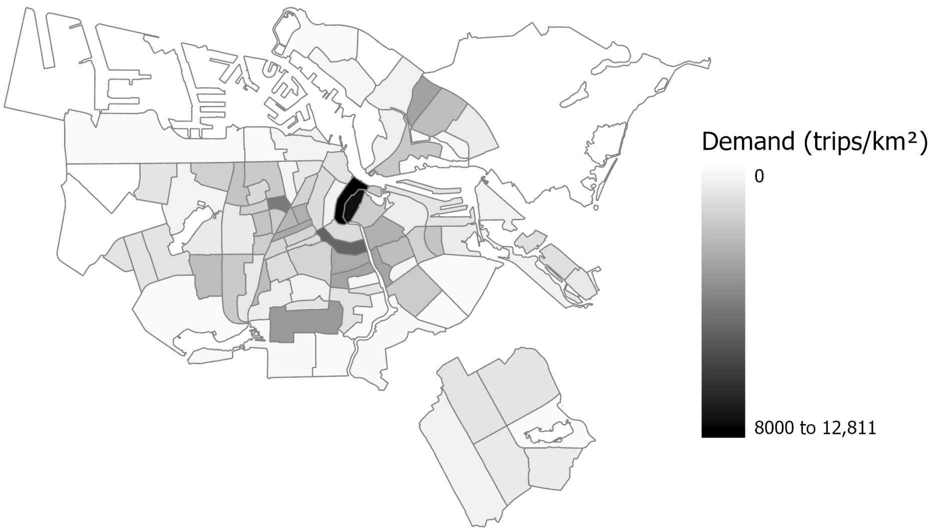

Figure 5.10: Number of check-ins per neighbourhood of Amsterdam

5.4.1. Walking Coverage versus Demand

The first relationship tested against the use of the network is the walking coverage of the study area. This relationship is visualized in Figure 5.11. A significant and positive relationship is found between these two indicators (R^2 = 0.146, r = 0.382). The R^2 value denotes that only 14.6% of the variation in the observations is attributed to the walking coverage indicator with 85.4% resulting from other factors. The calculated beta coefficient of the linear regression model is 2538. The significant positive correlation between the walking coverage and the transit demand suggests that increasing the walking coverage in a neighbourhood increases the demand of the transit network. Increasing the walking coverage can be facilitated by and increase of transit stops as well as improving the pedestrian network connectivity in a neighbourhood. These are the two major inputs and contributors for the walking coverage calculation as described in Chapters 3 and 4. It is reiterated that the walking coverage indicator only accounts for 14.6% of the variation of the demand observations. Other factors affecting transit demand still need to be considered when making policy recommendations to increase transit demand.

5.4.2. Supply Frequency versus Demand

The second relationship tests is between the supply frequency and the use of the network. This relationship is visualized in Figure 5.12. A significant and positive correlation is found between the two indicators. The R^2 value of 69.1% demonstrates that a majority of the variation of the observations is explained by the linear regression model. Looking into the regression model the slope coefficient is 13.332 and significant at a p-value of 0.000 (Table 5.3). The findings suggest that the increase of transit supply frequency has a large responsibility in the increase of transit demand in a neighbourhood. However, a heteroscedasticity after 100 departures per square kilometer is noticed which may suggest less certainty in changes to demand after a certain threshold of supply frequency. Nonetheless, this increase in demand may result from the improved quality of the transit network through the decrease of waiting time. The findings can also suggest the importance of the increase of routes through a neighbourhood providing more options to its residents. These options are discussed further in Chapter 6 through the discussion.

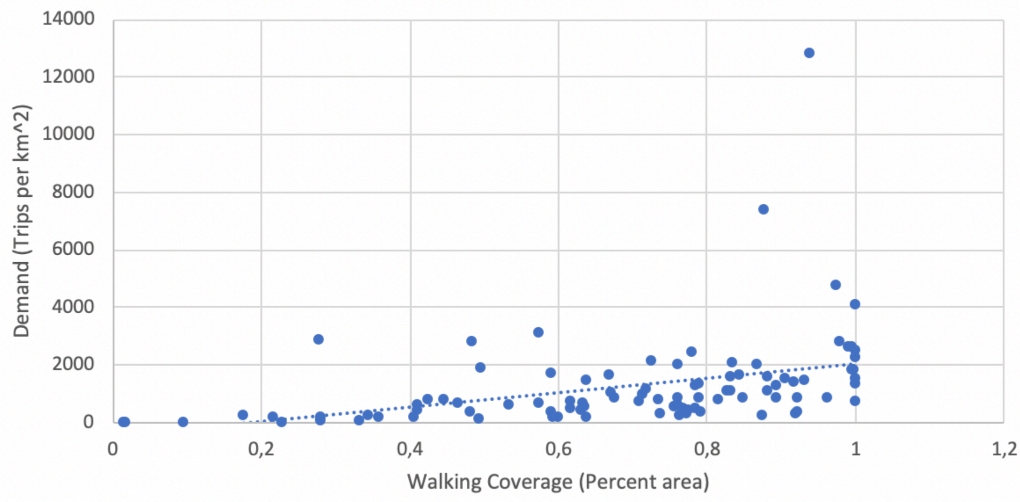

Figure 5.11: Relationship between walking coverage (% of neighbourhood area) and the demand of the Amsterdam public transport network

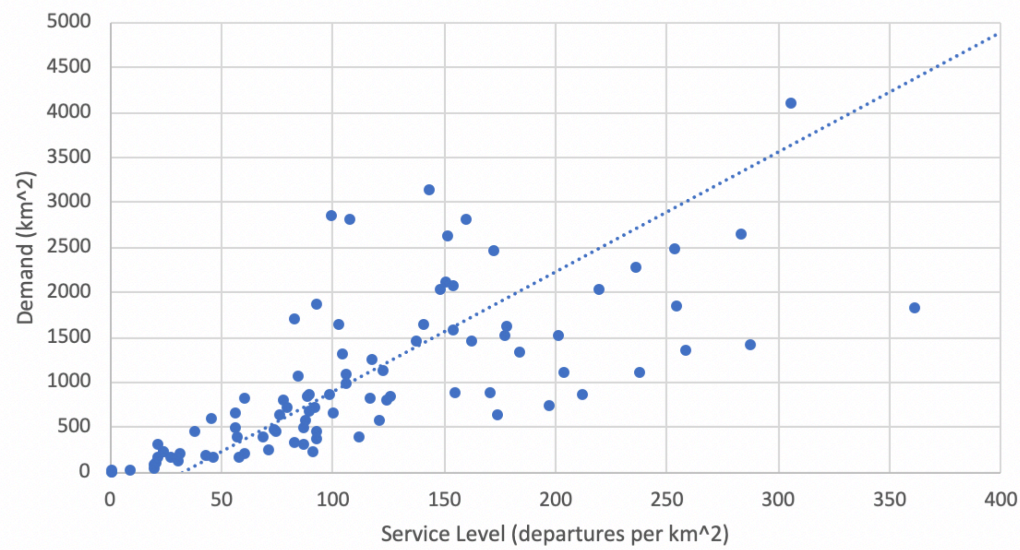

Figure 5.12: Relationship between supply frequency (ptv/km^2/hr) and demand of the network

5.4.3. Accessibility versus Demand

The final relationship between the supply indicators and demand indicator is between accessibility and demand. As shown in the results by Curtis et al. (2019), it is expected that a significant positive relationship is found between the two indicators. In the case study for Amsterdam this relationship is also found as visualized in Figure 5.13. This is statistically supported by Pearson correlation coefficient as 0.444 with a R^2 value of 0.197. This signifies that 19.7% of the variation of the observations is explained by the linear model while the much greater 80.3% is explained by other factors. The beta coefficient for the linear regression is 0.006 and significant at a p - value of 0.000. These findings represent that when a resident is able to access more people throughout a city through transit that their transit ridership will increase. This has important policy implications for both public transit networks as well as land-use policies. These are discussed in the following Discussion chapter.

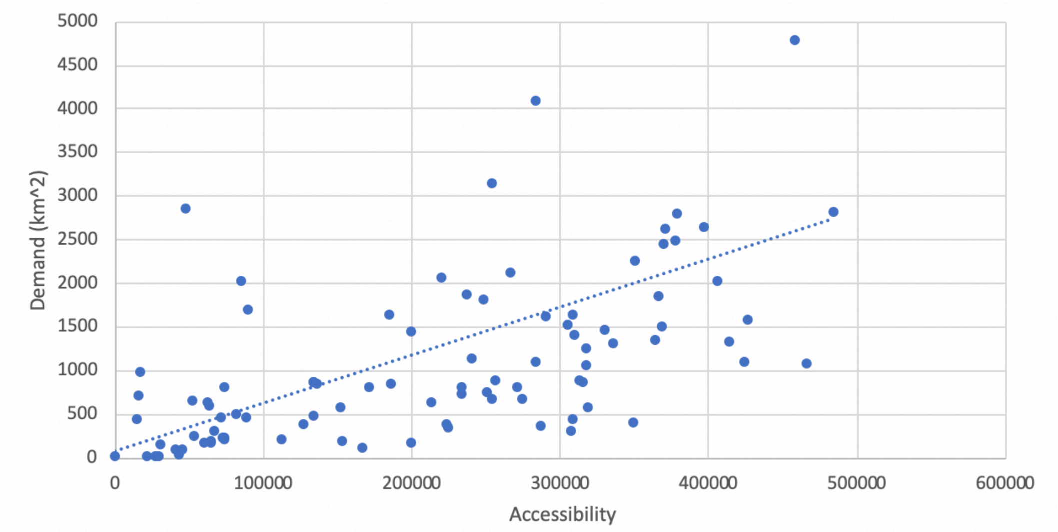

Figure 5.13: Relationship between calculated accessibility (meausured as people reached in 30 minutes via transit) and demand of the network

The results of the regression analysis are summarized in Table 5.3.

Tested Relationship	Beta coefficient	Beta coefficient p-value	R^2 values	Pearson Coefficient
Walking Coverage vs. Demand	2538	0	0.146	0.382
Supply Frequency vs. Demand	13.332	0	0.691	0.831
Accessibility vs. Demand	0.006	0	0.197	0.444

Table 5.3: Statistical values from the simple regression analysis

5.5. Regression Analysis for Interaction Effects

The clusters defined in the following section are created in order to differentiate neighbourhoods across the city. The methodology described in Section 3.4 was conducted in order to see whether the different clusters of the city had an effect on the relationship between the supply and demand indicators. The results of these regression analyses are presented in this section.

	Walking Coverage			**Supply Frequency**			**Accessibility**		
	r	**R^2**		**r**	**R^2**		**r**	**R^2**	
	0.588	0.346		0.763	0.582		0.658	0.433	
	B	**t-stat**	**P-value**	**B**	**t-stat**	**P-value**	**B**	**t-stat**	**P-value**
Intercept	-583.5	-0.831	0.409	213.1	0.779	0.438	98.5	0.274	0.785
Indicator	2178.5	2.429	0.017	-277.0	-0.744	0.459	247.9	0.556	0.579
Class 2	860.4	1.020	0.311	-214.3	-0.503	0.616	305.3	0.627	0.533
Class 3	681.7	0.729	0.468	49.0	0.111	0.912	-220.3	-0.347	0.730
Class 4	319	0.343	0.732	-244.5	-0.557	0.579	55.2	0.108	0.914
Class 5	509.9	0.578	0.565	-272.4	-0.535	0.594	-539.7	-0.702	0.485
Class 6	-2933.8	-1.734	0.087	7.4	3.585	0.001	0.0	0.005	0.996
Interaction Effect Class 2	-1075.4	-0.964	0.338	2.7	0.929	0.356	0.001	0.310	0.758
Interaction Effect Class 3	-467.7	-0.382	0.704	2.0	0.688	0.493	0.001	0.532	0.596
Interaction Effect Class 4	-85.3	-0.069	0.945	-0.9	-0.324	0.747	-0.002	-0.834	0.407
Interaction Effect Class 5	-1099.0	-0.844	0.401	0.3	0.064	0.949	0.004	1.537	0.128
Interaction Effect Class 6	4609.9	2.116	0.037	1.2	0.456	0.650	0.004	2.999	0.004

Table 5.4: The model summary of the regression model which incorporates interaction effects for all supply indicators.

The first supply indicator tested for interaction effects with the demand was the walking coverage. The fit of the regression increased to a R^2 of 0.346 and r = 0.588. The beta value of the walking coverage variable is still significant at a p value of 0.017 with a total value of 2,178 (Table 5.4). The averages for Clusters 2 through 6 are insignificantly different from Cluster 1 as per the p-values. These findings suggest that all socioeconomic clusters in Amsterdam, as defined by the LCCA clustering, do not statistically differ in their average walking coverage. This is despite the calculated walking coverage of 100% in the centre of Amsterdam and <10% in the peripheral neighbourhoods. While discrepancies are recognized on a geographic level, these discrepancies disappear statistically when placing these neighbourhoods into the LCCA determined clusters. The interaction effects for Clusters 2 through 5 are also insignificant in relation to the reference Cluster 1. These findings suggest that the transition from latent demand to effective demand for these clusters of neighbourhoods are not significantly different. Therefore, it may be assumed that the *needs* for walking coverage in these clusters are the same. However, this significance is found for interaction between Cluster 6 and Cluster 1 as the walking coverage indicator is significant at a p value of 0.037 and a beta value of 4609. In comparison with Cluster 1, the transition of latent demand to effective demand is different for Cluster 6. This suggests that the exogenous socioeconomic and demographic characteristics of Cluster 6 play a role in the walking coverage indicator. This also makes the average demand for neighbourhoods in Cluster 6 equal to 3271 (Table 5.4).

After the walking coverage, the interaction effects for the supply frequency and the demand of the network was measured. Cluster 1 was used as the reference group in which the other clusters were compared. The R^2 value was calculated at 0.582 showing a minimal decrease in the percentage of the variance of the observations described by this linear model. The beta value for the supply frequency was calculated as significant at a value of 7.393, a decrease from 13.3 in the base regression. However, the conclusions remain that an increase of the supply frequency in a neighbourhood results in an increase in transit demand. P-values ranging from 0.45 to 0.95, as seen in Table 5.4, also conclude that the average supply frequency in each neighbourhood cluster is not statistically significant from each other. Despite clear differences of supply frequencies throughout the city, these differences are not recognized when the effected neighbourhoods are clustered as per the LCCA. Similar results occurred when the reference group was changed to Cluster 2 to test for any demonstrate the insignificance of the selected reference group. It is noticed that the same R^2 value of 0.582 is calculated after this change. All clusters are still insignificant in relation to the new reference group. These findings suggest that neighbourhoods throughout Amsterdam react similarly to changes in the transport network frequency, regardless their socioeconomic cluster. This is not expected as it is assumed that varying socioeconomic and demographic clusters have different *needs* surrounding the efficiency of their transit network.

To finish the regression analyses was completed to test the interaction effects between the accessibility indicator and the demand indicator. Cluster 1 was used as the reference group with all other clusters tested against it. The overall fit of the linear model increased to a R^2 of 0.433 and r = 0.658. However, the strong positive correlation, as found also in the general regression, is maintained along with the corresponding assumption that an increase in the accessibility results in an increase in transit ridership. The accessibility variable was still significant (p = 0.004) and the beta (slope) metric remained at 0.004. None of the clusters in relation to Cluster 1 had an average (beta value) that was significant in this model. This finding coincides with the findings from the other two supply indicators. There is statistically no difference between the average accessibility's for the clusters based on socioeconomic and economic data. This is despite the clear distribution of accessibility scores across the neighbourhoods of Amsterdam. This insignificance is also found when looking at the interaction effects between the clusters and the accessibility metric in comparison with the reference group. This finding concludes that the relationship between the accessibility indicator and the demand in neighbourhoods is not dependent on the neighbourhoods socioeconomic and demographic cluster (Table 5.4). Instead, the all defined clusters in Amsterdam have the same relationship between the residents they may access and their ridership. This is not expected as it is shown in past literature that public transport modal share, as well as activity level, has a different relationship with accessibility metrics depending on the socioeconomic and demographic group (Allen and Farber, 2020; Curtis et al., 2019).

This chapter presents the results after the application of the methodology on the City of Amsterdam case study. This is completed in order to identify the levels of latent demand throughout clusters of neighbourhoods throughout the city. Six different clusters of neighbourhoods were identified throughout the city. These clusters are identified through the socioeconomic variables presented through the CBS. Upon completion, the transit supply indicators and transit ridership are attributed to individual neighbourhoods under the umbrella of their identified cluster. Significant trends and insights were presented from these data. Lastly, the relationship between the transit supply and transit ridership is calculated, with the lack of interaction effect recognized within the regression analysis. The calculations of this process are presented in this chapter. The implications of these results and what it means for the methodology and policy recommendations are the next goal for this research. These are outlined in the following Discussion chapter.

6 Discussion

In this chapter, the results of this thesis and their implications are discussed. Furthermore, limitations are presented and suggestions for future research are made.

6.1. Identifying Clusters in Amsterdam

The methodology and subsequent analysis of this study is based on transport poverty assumptions that are first introduced in Chapter 1 and thoroughly analysed in Chapter 2. These assumptions state that groups of low socioeconomic and demographic status have the most transport *need*. This study aims to test these assumptions on the city of Amsterdam case study by identifying socioeconomic and demographic clusters within the city. This is completed by using the LCCA to identify clusters of Amsterdam neighbourhoods which maximizes the socioeconomic and demographic similarity within the clusters but maximizes the differences between clusters. Maximizing the differences between neighbourhood clusters allows the transport *needs* to best be evaluated.

Six different socioeconomic and demographic clusters are found within Amsterdam. Each cluster has distinguishing features which separates one cluster from another. For example, Cluster 3 has the largest average home value of any other cluster while Cluster 6 contains the population with the highest car ownership per capita. As the differences are recognized, it may be assumed that they have different transport *needs* falling in line with the studies expectations. If differences are found, it should be understood that the findings are only relevant in terms of the income, car ownership, and family composition characteristics. Past studies have also included characteristics such as education and literacy, but this is not incorporated in the study. Any conclusions from this research needs to understand this context that the results are provided in. There is also a concern (in terms of validity of the methodology) that the clusters are too affluent in order to recognize true differences of transport *need* as they may have all options available. For example, Cluster 1 and 6 are defined as having high average income, and through methods found in literature, these clusters most likely would not be flagged as a risk group. Any differences in transport *need* may not be recognized for the six clusters of Amsterdam for this reason.

Another aspect of the findings is the uncertainty surrounding the definition of six clusters for the city of Amsterdam. The LCCA was chosen in order to operationalize the clustering of neighbourhoods and remove any bias from the analysis as found in prior literature (Curtis et al., 2019). In building the methodology, the classification/clustering of neighbourhoods was intended in a way that could be transferred to another study area. The techniques used by Curtis et al. (2019) are reliant on local knowledge and cannot be standardized for this reason. Further, studies completed in health care showed that clustering techniques were useful in allocating people into subsets of groups (Pedigo et al., 2011). This technique can be applied to other study areas and proves to be more accurate than other clustering techniques such as K-means. However, the results of the LCCA provide viable options from three to seven defined clusters. Six clusters are chosen based on the minimization of the BIC values and bivariates, however, another research may have selected a different model. Local knowledge for the city of Amsterdam is also used when interpreting the clusters and their validity. Cluster 3 located in the city centre and Amsterdam Zuid, as well as Cluster 2 in Nieuw West and Zuid Oost are recognized as *correct* based on this local knowledge. However, this reintroduces bias into the analysis which goes against the original intention behind using LCCA. Therefore, a combination of operationalizing techniques and local knowledge fine-tuning may be the best method for such a clustering study.

Referring back to the research question, the intent of this clustering process is to find the socioeconomic and demographic clusters of Amsterdam. Based on the level of uncertainty surrounding the number of clusters, the defined clusters are not the absolute clusters of Amsterdam. Instead they are a tool, supported by local knowledge, that can support the remainder of the study. This is important to acknowledge when interpreting the transport *needs* of these clusters in further parts of this chapter.

6.2. Transit Network Quality in Amsterdam Neighbourhoods

Once the neighbourhoods of Amsterdam are allocated its socioeconomic and demographic cluster, the quality of the transit network is identified. The intent is to identify the quality of the transit network in line with the guiding principles of an inclusive transit network, as stated by Tovaas (2020). The focus for this study is on tangible components of the network such as the convenience of accessing a stop as well as the efficiency of vehicles in the network. The latter is represented by the walking coverage metric while the efficiency of the vehicles is represented in the supply frequency as well as the accessibility metric. These indicators are also realizable via the chosen GTFS data source which allowed for a seamless integration, using GIS, into the research methodology.

It is recognized that indicators assume how people experience the network rather than being validated by stated preferences. The perception is an important component when assessing the human aspect of the transport system. Building the indicators in line with the guiding principles of an inclusive transport network attempts to include this human element. However, the direct correlation between the assumed and perceived convenience for the potential transit user is missing. This element is also absent in most reviewed literature except for a study completed by Farber et al. (2018) on the transport needs of Syrian refugees. This thesis is no different so readers should realize the empirically derived nature of the results.

The supply indicators do not share a similar distribution across the city (Figure 5.5, 5.7, 5.9). It is therefore difficult to define one definite measure for quality of the transport network. Past studies frequently defined a composite indicator for a city's transit network (Deboosere and El-Geneidy, 2018; Foth et al., 2013). However, a composite indicator of *high quality* may consist of one weak metric and one very strong high quality metric that masks any issues in the network. This scenario provides a simplified description of the state of the current transit network and hide areas where transport needs are not met for certain socioeconomic and demographic groups.

This thesis instead allows for a comparison and general trends for each individual indicator and analyzes transport needs as a result. For example, the left skew of the walking coverage indicator distribution indicates that it is easier for a transport network to offer coverage to the city (Figure 5.5). This is in comparison with the supply frequency of the network, where the distribution of the neighbourhoods is highly skewed to the right (Figure 5.7) where only a few neighbourhoods have a value close to the maximum in Amsterdam. This reflect the costly nature of running high frequency services in comparison to providing a network with a high density of transit stops. Neighbourhoods surrounding Centraal Station are *quality* in all aspects. However, neighbourhoods in the west experience high walking coverage, nonetheless, have supply frequency and accessibility indicators falling below the Amsterdam average. The residents in these neighbourhoods may experience a convenient network, nonetheless, the efficiency of the network may not be sufficient. It is thus apparent that the *quality* of the network in a neighbourhood is relative to each metric and researchers and planners need to take this into consideration for future studies and network planning.

By finding these supply indicator distributions, the constraints/opportunities that residents experience across the Amsterdam transit network are realized. Applying these constraints and opportunities to different socioeconomic and demographic groups is a necessary component in analyzing the transition from the latent demand to effective demand (Clifton and Moura, 2017). Based on this transition, insights into the *needs* of residents are gathered. Once identified, the definition of latent demand/need for public transport may be strengthened. The results in calculating the demand of transit throughout Amsterdam are discussed in the following section.

6.3. Transit Demand for Amsterdam Neighbourhoods

The third research question strives to identify the demand for public transport through the city of Amsterdam. The methodology is defined as the actual use of the network rather than the potential desire to take public transit. The use of the network is recorded as check-ins at all GVB stations during morning rush hour. This decision is contrary to previous public transportation accessibility research as studies have either focused on the desire and *need* of public transport, defined by peoples socioeconomic and demographic backgrounds, or the use of the network, largely defined as the modal share or activities completed. This unfamiliarity provides an opportunity to demonstrate how the smart card ticket technology can be utilized to provide insights to the use of the network at a neighbourhood level.

The check-in data was provided at the stop level and allocated to the surrounding neighbourhoods using the methodology described by Jung and Casello (2019). The distribution of trips in these neighbourhoods is visualized in Figure 5.10. It is evident that there are few large transport hubs with >4000 trips, which is much greater than the average trips of 1211. This is despite check-ins being eliminated from the total by subtracting the average number of passengers who access the GVB network from the NS network. It was the intention of the study to capture riders that were leaving their own neighbourhoods and tie the trips back to the socioeconomic and demographic data of the neighbourhood. However, these large transport hubs and their check-in information, are not as closely tied to their neighbourhoods characteristics as other neighbourhoods. Instead, they have influence from commuters who access the GVB network from other networks (Connexxion and EBS account for 6% of Amsterdam travellers) or active modes such as walking or biking. A larger use of the network is also observed in neighbourhoods with metro stations such as Amsterdam Noord, Amsterdam Zuid, etc. This supports the claim that these neighbourhoods have high access from other neighbourhoods. It should remain clear that these are assumed conclusions rather than supported by access data of each recorded trip.

There are questions asked on the the distance taken to access these trips, however, the methodology is still effective in allocating a trip to a nearby neighbourhood. For this reason a neighbourhoods relative proxy for transit use, in relation to the network's constraints/opportunities, is realized. In terms of the methods validity, no trips are gained or lost between the raw GVB data and the total trips counted after allocation to the neighbourhoods. This demonstrates the effectiveness in defining the geographic location of a transit stop, defining the surrounding neighbourhoods, calculating the buffer area of the transit stop, and calculating the total trips. The method is also adapted from previous literature in which Jung and Casello (2019) demonstrates that the technique is validated by on board surveys asking a passengers home in relation to their first stop. While not conducted in this research, these findings support the techniques of using check-in data for the Amsterdam case study.

6.4. Observing the Relationship between Supply and Demand

The transport needs are defined in the last component of the analysis where the relationship between the networks supply indicators and the demand indicator is examined. The previous three components are the building blocks in order to test the hypothesize, while in the fourth section the hypothesis is tested. To reiterate, the research evaluates the transition between latent demand and effective demand when socioeconomic and demographic groups are presented with varying constraints to the network. It is hypothesized that the needs of groups from a transport network are reflected in this transition where lower socioeconomic and demographic groups are assumed to transition quicker between latent and effective demand.

The transition between latent demand and effective demand is reflected in the simple regression analysis of the relationship between supply and demand indicators. Positive and significant correlations are found in Section 6.1 which validates that the population places value in improvements to the transit network. This is critical as network improvements are currently the main tool for transit operators provide better services to potential riders. The definition of the indicators suggest that improved convenience and efficiency of the network is valued as more trips are taken when these indicators are improved. This is reflected in terms of the increase of ridership rather than stated perceptions from transit users within Amsterdam. Contrary to prior studies this relationship did not vary between socioeconomic and demographic clusters of Amsterdam (Allen and Farber, 2020; Curtis et al., 2019). The insignificance between clusters is recognized in the interaction effects regression and is unanticipated based on expectations communicated in Chapter 1. These expectations are motivated by studies by Curtis et al. (2019) and El-Geneidy et al. (2016) who found that more *need* for public transport may be in neighbourhoods of lower socioeconomic and demographic groups. This is not recognized in this thesis and therefore have consequences in the definition of transport *need* for socioeconomic and demographic groups within Amsterdam.

The findings suggest that the transport *needs* of potential transit users in Amsterdam are the same in terms of convenience and efficiency. It is important to note that this is only when neighbourhoods are clustered through income, car ownership, and family composition, and does not take into consideration other factors. These findings may be reflective on both the transit network as well as the demographics of the neighbourhoods the network is reaching into. A possible explanation is the Amsterdam transit network already sufficiently supports the needs of potential users through the city. Therefore, any network improvements are not drastically *needed* more by one cluster than another cluster. This may reflect and validate the stated goals of the Vervoerregio to provide an equal and quality network to the residents of Amsterdam.

It is also suggested that the defined clusters in Amsterdam are above a *transit need* threshold and are too affluent to experience any significant variances. There is a possibility that methods used in past literature would not define any of the defined clusters as being at risk of transport poverty (Currie, 2004; Jaramillo et al., 2012). Therefore, the added benefits that transit provides to lower socioeconomic and demographic groups are not fully experienced by the defined clusters within Amsterdam. This may be because transit does not have a significant added value in comparison with other modes such as walking and cycling through the city. Extra transit needs that are prevalent in other cities, may be satisfied by these modes within the city limits.

If the equal transit needs hold true for neighbourhoods of Amsterdam it has implications on how the network is designed in the future. This means that when proposed, transit improvement projects may be evaluated without considering varying socioeconomic and demographic needs in the defined geographic space. Instead improvements can be made where it is deemed necessary and the return on investment, in terms of additional ridership, is equal. The necessary improvements may involve evening out the unequal distribution of accessibility seen in Figure 5.9 or skewing the walking coverage distribution further left (Figure 5.5) to ensure everyone has equal convenience of reaching a station.

Another possible explanation for the findings is that the methodology did not adequately capture individuals who's transit *needs* are greater than the rest of the population. It is recognized that a great deal of homogeneity is present in the defined clusters. This may result from the uncertainty in the chosen clusters, characteristics used, or the fact that statistics are aggregated at a neighbourhood level. This last component places a significant weight on a *needy* individual to be recognized only by living in a *needy* neighbourhood. Past research on transport poverty frequently completed analyses at the neighbourhood level in order to identify potential areas of transport poverty and/or higher transport need. It is the intention of this study to maintain this level of aggregation and move from the identification of potential risk to the definition of tangible needs of these neighbourhoods. The lack of noticeable differences may imply that a more microscopic and individual approach is required in order to identify variation of transport needs. This fact highlights that measuring potential transport need takes a different approach than building a concrete definition of the need.

A solution is to improve tying the transit trip exactly to the characteristics of the individual. As discussed previously, the demand data is collected at the stop level and does not reflect the individual making a transit trip. Instead travellers may access a *high quality* transit stop rather than the stop in their neighbourhood. This places a higher relevance to the supply at a transit stop rather than the socioeconomic and demographic characteristics of the neighbourhood. This explanation supports the decisions in past research to observe the use of the transit network through modal share and activities completed through household travel surveys (Allen and Farber, 2020; Curtis et al., 2019). It further challenges the methods surrounding smart card data to be developed and improve the connection between a trip and the travellers individual characteristics.

Another fact of the analysis is that supply indicators targeted only a couple of principles of an inclusive transport network. The indicators attempt to highlight the convenience and efficiency of the network and are only two out of the eight guiding principles that are discussed in the Inclusion paper on inclusive and sustainable transport networks by Tovaas (2020). These indicators are exclusive to the tangible components of the transport network and do not highlight the intangible aspects such as empowerment, empathetic, gender equality, and safe. There is the potential that varying needs between clusters are found in intangible factors such as safety, empowerment, and empathy. There are efforts within the Vervoerregio to increase empowerment and empathy in the transit networks though the introduction of programs designed to help people with lesser ability. These programs aim to support people in accessing and using the public transport network. Continuing this emphasis in transport authorities and incorporating the intangible principles into future accessibility research is an important follow-up from this thesis.

In an ideal world, the needs of every individual person would be met in a transit network. Significant needs of convenience and efficiency are recognized for the entire city of Amsterdam which transport planners can utilize in future projects. However, in order to define differences between different socioeconomic and demographic groups described in past research further work at a more individual level should be completed. In order to properly define recommendations for policy makers and academics, further assumptions made in this research and their limitations are first outlined in this following section.

6.5. Limitations and Assumptions

The results were calculated with the understanding that there are limitations when it comes to the methodology as well as its implication to the case study. In this section the limitations and the implications are explored.

- In the study, the use of public transport between different neighbourhoods was compared. It is assumed that the use of public transport is determined only by the quality of the supplied public transport network as well as socioeconomic factors that may have an impact on ridership. In a normal transport model, the number of public transport users is calculated in comparison with those that use cars/bike/walking etc. The modal share is determined by the ease (travel time) of the other modes. In this study it is assumed that the quality of the car and bike network does not have an impact on the use of public transport. As stated previously, only the quality of the public transport network and socioeconomic factors have an influence.
- The study does not take into account the access/egress aspect of a public transport trip. It is assumed that if a person lives in an area in the city that they will take a public transport trip starting from their neighbourhood. The technique was used by Jung and Casello (2019), and accounts for a 400m buffer walking area around the transit stops. This assumes that the maximum distance a person accesses or egresses from a transit stop is 400m. The assumption sets aside that a person may walk towards a larger transport hub outside of their area as well as bike to a public transport station that provides a more direct connection. In order to combat this limitation it may be necessary to consult mobility surveys in order to understand the entire

- The public transport metrics were calculated for a weekday morning peak hour (07:00 - 10:00) and extrapolated for the entire month. These methods were chosen to simplify the methodology and attempt to capture only riders that were leaving their area of residence in order to tie the ridership and supply indicators to the socioeconomic and demographic variables of the neighbourhood. These reasons are valid, however, it does not reflect the dynamic nature of accessibility and ridership throughout the entire day. Future studies on the relationship between supply and demand indicators may want to identify ways in order to combat this issue.
- Correlations have been identified between the supply indicators and the ridership data across the city of Amsterdam. This is a starting point in identifying the relationship between these two aspects. However, this study was unable to identify why these correlations are present and whether causation can be assumed. Further insights into why these relationships were found, whether it is through surveys, focus groups, etc. would be of value to future research.
- In this study, there was a large amount of aggregation that occurred with the various data sources. First, was the supply and demand data as previously mentioned. Second, was the choice to aggregate neighbourhoods in terms of their socioeconomic and demographics backgrounds. This was chosen in order to incorporate the vast characteristics from a neighbourhood into one cluster. While successful, it results in people and aspects of neighbourhoods being missed.
- A clustering technique was chosen instead of accounting for each individual socioeconomic and demographic characteristic. Many previous studies have only highlighted one characteristic such as age, income, or car ownership. As mentioned previously, this decision was made as a neighbourhood cannot be defined by just individual characteristics but rather a conglomeration of these characteristics. While the clustering decision was made, certain findings regarding the individual characteristics could have been missed, such as the transit use in relation to areas of high and/or low car ownership.
- There is no reliability of the transit service but only the posted schedule. There could be variations in the transit reliability across the city that alter the transit experience for its residents. For example, certain areas may experience busses that come too late or too early more frequently than other areas. While assessing these differences was not in the scope of this study, it may be worthwhile to consider in future iterations of this work.
- The accessibility is calculated without taking into consideration the entire route to a particular opportunity, especially how many transfers are required. It has been shown that transfers add additional perceived minutes onto a transit trip. Certain areas of the city may require more transfers for the same amount of accessibility opportunities than another section of the city. Taking the transfers into consideration may reduce or increase the “perceived accessibility” from different neighbourhoods within 30 minutes.
- Metro, tram, and bus were all considered as one mode for the purpose of this study instead of taking them into account individually. This may be of significance as literature has shown that travellers see rail-bound transit options as more attractive than bus options. Therefore, a neighbourhood with 15 bus services an hour may be less attractive to its residents than a neighbourhood with 15 metro services in the neighbourhood per hour. When observing the supply indicators throughout Amsterdam, and the implications, this fact needs to be taken into account.
- The density of each neighbourhood is considered as equal in the allocation of transit departures as well as ridership. Therefore, a buffer area with 50% in one neighbourhood and 50% in another neighbourhood have half of the departures and transit trips allocated to each. The literature that this was adapted from by Jung and Casello (2019) emphasized the population density of each of these neighbourhoods. This approach was not taken into account for this analysis in order to simplify the calculations as well with the understanding that the city of Amsterdam is homogeneous in nature. Taking the population density into account may have resulted in different results.
- Looking at the geographic case area, only the city of Amsterdam was considered when counting the number of opportunities neighbourhoods experience. However, there are sections of Amstelveen and Duivendrecht which can be reached through 30 minutes of public transport. For

example, the Zuidas may see a reduction in recorded opportunities, as the services south of it are not taken into consideration. This fact needs to be discussed when using this tool for potential policy decisions.

To summarize, this research is based on several assumptions and has its limitations. These should be considered when conducting future research in the field of accessibility research. Overall, the findings of this study allow for recommendations and policy implications as discussed in Chapter 7.

7 Conclusion and Recommendations

7.1. Conclusion

The aim of this thesis was to move away from assuming transport *need* of a transit network. Instead the *need* is defined through quantifying the relationship between the supply and demand of the network for different socioeconomic groups. It is argued that transport networks are built in the past for an idealized group of residents and fail to cater to all the individual needs of its residents. If successful, this research allows these different needs throughout the Amsterdam transit network to be captured. This is important work that falls under the Vervoerregio Amsterdam's goals of limiting the potential for transport poverty in the city.

This approach was applied in a case study of the city of Amsterdam, The Netherlands and its current public transport network of the GVB. First, neighbourhood socioeconomic and demographic data was provided by Central Bureau for Statistics (2017). These neighbourhoods were grouped into like-wise clusters through a Latent Class Clustering Analysis (LCCA) of their characteristics. These clusters describe the city as more than one idealized group of people. Next, a GIS analysis combined GTFS data and the Amsterdam pedestrian network in order to calculate walking coverage, supply frequency, and accessibility indicators. This was completed in order to define the quality of the GVB network within the city of Amsterdam. Further, smart card ridership data from the GVB was used to measure the current demand for neighbourhoods throughout the network. Lastly, a regression analysis was completed, with interaction effects between the neighbourhood clusters calculated, in order to see the relationship between supply and demand and whether there are significant differences between these demographic clusters. These methodological steps allow us to answer the following main research question:

How does observing the relationship between the transport network supply and demand provide insights into the transport needs of neighbourhoods defined by their socioeconomic and demographic characteristics in the city of Amsterdam, The Netherlands?

The findings of this research show that there is a strong positive correlation between each of the supply indicators and the demand of the network for all socioeconomic and demographic clusters of Amsterdam. Further, as identified through the interaction effects, the extent of the relationship does not significantly vary between the six clusters of Amsterdam. By observing these relationships certain transport *needs* are recognized for the whole of Amsterdam rather than for specific socioeconomic and demographic groups. These findings suggest that the convenience and efficiency of the transport network is valued the same for all clusters within Amsterdam. For example, the lower-income families of Amsterdam Zuid-Oost *need* the supply frequency the same as higher income household in the city centre. This is opposite to the expectation that varying levels of socioeconomic and demographic groups react differently to their provided transit network.

These findings suggest that clusters of Amsterdam neighbourhoods have similar convenience and efficiency transport *needs*. This may arise from the fact that Amsterdam neighbourhoods, when defined through a combination of income, car ownership, and family composition, are not different enough to recognize varying *needs*. An alternative explanation is that any difference in *needs* that are present at the neighbourhood level, are already satisfied by other components of the transport network such as bike or walking. This is not empirically supported but a possible explanation for the similar interaction effects in the linear regression analysis. A final explanation is that the transport *needs* of the different clusters do not vary in terms of convenience and efficiency but may centre around intangible desires such as safety and gender equality. Important lessons are made through the use of GIS systems in order to fuse multiple data sources to complete this study. The LCCA technique is also used for the first time in such study and showed promise in being utilized for future research studies. In further studies, the limitations in assumptions should also be taken in how data is aggregated, which indicators are chosen, and how to truly emphasize the individual in a study on transport poverty. All of these considerations are discussed further in section, Section 7.2.

7.2. Recommendations

7.2.1. Future Research

The research objective in this thesis is to define the transport need for different socioeconomic and demographic groups of Amsterdam. Upon completion of the analysis and reflection of the results there are several recommendations for future research.

In this study it is determined that there is a significant positive correlation between the supply indicators (walking coverage, supply frequency, and accessibility metric) and the demand of the network. It is assumed that this increase in ridership is noticed because of improved convenience and efficiency of the network. However, further work needs to be completed to test how improvements of these supply indicators are actually perceived by individuals. Household or on-board travel surveys may be method to capture people's perceptions easier. If collected, this could provide more information on the needs of individuals and further reduce the assumptions made for different socioeconomic and demographic groups.

Continuing the focus on the individual traveller, there is a concern that this research does not properly capture the socioeconomic and demographic characteristics for every recorded trip at a transit stop. This is because trips are directly attributed to the surrounding neighbourhoods and the socioeconomic and demographic characteristics of the neighbourhood. This eliminates the potential for travellers' characteristics to be properly captured if they access a transit stop from further than 400m away. Curtis et al. (2019) and Allen and Farber (2020) use household travel surveys to look at the relationship between activities and accessibility and public transport share and accessibility, respectively. Using these forms of data within this thesis' methodology may capture needs that are not realizable through analyzing the smart card transit data.

Once the demand is matched to the socioeconomic and demographic characteristics, the methodology in this study should be performed on other inclusive transport network guiding principles. Indicators should be developed in order to capture further tangible components such as physical accessibility, as well as intangible components such as safety while using the network. Differences in the neighbourhood clusters of Amsterdam regarding these principles may be recognized which is not present within the analysis of this thesis.

Keeping the socioeconomic and demographic clusters as the focus, further research should be completed on testing the differences between the LCCA and ad hoc/heuristic clustering approaches. It is recognized in this thesis that the clusters determined through the LCCA were unofficially validated by local knowledge of the area. It would be interesting to determine which clusters are identified only through local knowledge and without the aid of a statistical clustering technique. Completing a sensitivity analysis would provide further insights into the validity of the LCCA for future research.

An added dimension to this research would evaluate the different clusters over time. This dimension could provide insights into the changes of the transport network over time, changes to the neighbourhoods socioeconomic and demographics, and whether the transport needs of these clusters of neighbourhoods evolve as well. Research into determining whether changes of the transport network play any role into the evolution of a neighbourhood's socioeconomic and demographic status as well as transport needs would provide insights into the impact a transport network poses.

The last suggestion for future research emphasizes the trip purpose behind a transit trip. Right now, it is assumed that the destination of every single transit trip has equal weighting. However, people may require public transit more and/or less for different trip purposes. For example, an individual commuting far distances to work without a car may highly value adequate public transport. However, their need for public transport may be much lower when accessing a grocery store within their neighbourhood of residence. Completing further research that highlights the need behind different trip purposes is valuable in the development of communities and public transport networks.

The above suggestions highlight how this thesis' methodology may be improved as well as built on for future research. By working through these recommendations, it is the hope that the definition of transport *need* may be strengthened and applied in order to build more sustainable and inclusive transport networks.

7.2.2. Policy Implications

Several policy recommendations can be made from the results of this research. Policy makers should include the supply indicators in monitoring of the region as well as potential forecasting within the network. The positive correlation between the supply indicators and demand supports the assumption, that an increase in ridership occurs if improvements are made to the system. This positive correlation occurs for walking coverage, supply frequency, and accessibility so projects which improve these indicators may result in an increase in ridership.

These different supply indicators should also be monitored over time for the city of Amsterdam. It is recognized that the residents of Amsterdam value these components in the current state of the network. By evaluating the supply indicators, as well as ridership, over time it may be determined whether the GVB network is developed along side the *needs* of Amsterdam residents. This would be a move away from classical transportation engineering and instead monitor how the network is facilitating peoples needs.

A noticeable difficulty in the research process is making assumptions surrounding the distances people travel in order to reach their public transport stop. It is recommended that the Vervoerregio evaluates how far and with which mode people are accessing their public transit stop. Understanding the methods of access for transit trips to be tied better to the socioeconomic and demographic characteristic. It could also provide information into identifying which types of people are willing to go further towards their public transport. This is another potential method to define the transport *need* of people within Amsterdam. Completing house hold surveys with questions involving the access of the transit network may prove valuable.

It is advised, that the Vervoerregio Amsterdam continues to implement policies to serve all socioeconomic and demographic groups equally. The research shows the current policies have served the needs of the defined clusters of Amsterdam in terms of the supply indicators in this study. The Vervoerregio can continue to implement improvements to the network in order to support these homegenous convenience and efficiency transit *needs* in the city. They should look further into developing the policies further to include all of the eight guiding principles of a sustainable and inclusive transport network as outlined by Tovaas (2020). By building this understanding, the network can continue to improve the facilitation of the transit needs throughout the entire city.

To summarize, this research has presented a sound methodology that investigates the quality of the public transport network in the study area. For the purpose of this study, the methodology was applied to the city of Amsterdam. By calculating three supply indicators, *walking coverage*, *supply frequency*, and *accessibility*, it was possible to quantify different measures of network quality. Furthermore, the ridership, *demand*, was calculated. The supply indicators were compared to the demand indicator for different socioeconomic and demographic clusters within the city. This shed light into the needs these different socioeconomic and demographic groups have towards a public transport system. Based on these results, recommendations for future research as well as recommendations for policy adjustments were made.

A
Appendix

Neighbourhood	Walking Coverage	Frequency (per hour per km^2)	Accessibility	Demand (km^2)	Cluster
Driemond	0	0.00	1861.38	0.00	2
Burgwallen-Nieuwe Zijde	0.94	613.58	316656.16	12810.86	1
Burgwallen-Oude Zijde	0.88	464.87	404421.75	7381.57	1
De Weteringschans	0.98	452.24	458229.51	4781.42	6
Indische Buurt West	0.99	361.39	248611.90	1815.81	6
Geuzenbuurt	1.00	306.51	284390.90	4096.08	3
Dapperbuurt	0.92	287.58	310509.96	1407.19	4
Van Lennepbuurt	0.99	283.77	396831.81	2641.52	3
Frederik Hendrikbuurt	1.00	258.94	364396.61	1350.58	1
Hoofdweg e.o.	1.00	254.94	366470.23	1838.97	4
Da Costabuurt	1.00	254.23	377511.59	2475.70	2
Jordaan	0.88	238.15	283675.19	1091.22	1
Kinkerbuurt	1.00	236.84	350721.56	2260.12	2
Oosterparkbuurt	0.87	219.97	406771.01	2014.62	1
Haarlemmerbuurt	0.96	212,74	315592.49	863.84	2
Grachtengordel-Zuid	0.83	203.88	424426.47	1099.94	4
Westindische Buurt	0.90	201.73	369329.11	1501.97	1
Staatsliedenbuurt	1.00	197.96	250853.74	736.53	3

Museumkwartier	0.79	184.65	414838.17	1324.31	3
Frankendael	0.83	178.91	290284.46	1610.14	1
Chassébuurt	1.00	177.38	304972.60	1519.89	4
Indische Buurt Oost	0.53	174.25	213886.17	624.98	3
Weesperbuurt / Plantage	0.78	172.40	369929.07	2443.07	4
IJselbuurt	0.89	170.71	313569.34	878.65	3
Van Galenbuurt	0.93	162.35	330093.84	1452.05	5
Nieuwe Pijp	0.98	160.14	379268.50	2801.02	1
Scheldebuurt	0.85	155.26	256704.63	876.20	4
Slotervaart Zuid	0.83	154.44	220356.00	2062.07	2
Nieuwmarkt / Lastage	0.88	154.38	427175.28	1569.01	2
Oude Pijp	0.99	152.29	370866.67	2618.67	1
Schinkelbuurt	0.73	151.55	266428.34	2110.48	4
Buikslotermeer	0.76	148.87	85002.51	2024.55	3
Zuidas	0.57	143.78	254082.62	3136.83	2
Landlust	0.84	141.35	308829.10	1637.21	1
Slotermeer-Noordoost	0.64	137.88	199829.22	1442.13	2
Osdorp-Oost	0.79	126.61	186331.25	835.23	2
Oostelijk Havengebied	0.74	124.43	234030.48	794.37	2
Erasmuspark	0.72	123.34	240427.37	1130.33	1
Oostelijke Eilanden / Kadijken	0.76	121.05	152711.44	565.73	3
Helmersbuurt	0.78	118.24	318208.30	1253.84	1
Geuzenveld	0.82	117.28	171919.34	807.05	1
Apollobuurt	0.79	112.53	349322.91	391.34	5
Weesperzijde	0.48	108.31	484119.23	2807.33	3
Vondelbuurt	0.83	106.54	466483.54	1071.42	1
IJburg West	0.71	106.33	17471.71	968.33	2
Overtoomse Sluis	0.89	104.66	336465.45	1302.41	1
Westlandgracht	0.67	102.68	185161.23	1640.65	4

Slotervaart Noord	0.58	100.20	274434.48	658.54	6
Elzenhagen	0.28	99.71	47738.11	2850.58	2
Osdorp-Midden	0.76	98.66	133784.49	854.86	2
Grachtengordel-West	0.63	93.34	309041.76	442.78	5
Transvaalbuurt	0.92	92.87	287206.27	353.83	1
De Kolenkit	0.49	92.78	237730.53	1857.62	1
Hoofddorppleinbuurt	0.71	91.91	234107.95	716.00	6
Tuindorp Buiksloot	0.88	91.11	72939.87	219.69	3
Overtoomse Veld	0.63	90.20	254056.84	665.12	1
De Punt	0.92	89.65	134062.33	852.13	1
Bijlmer Centrum	0.68	88.97	136396.57	839.00	2
Stadionbuurt	0.64	88.33	318867.80	571.72	2
Volewijck	0.78	87.27	81916.52	494.03	1
Tuindorp Nieuwendam	0.92	87.18	66866.33	295.32	2
Willemspark	0.67	84.90	318094.81	1059.44	6
IJplein/Vogelbuurt	0.59	83.13	89514.13	1693.15	3
Centrale Markt	0.74	83.09	225480.96	330.06	2
IJburg Zuid	0.62	79.57	15803.69	702.71	6
Prinses Irenebuurt e.o.	0.43	78.59	271606.62	798.90	5
Waterlandpleinbuurt	0.76	76.61	62370.20	629.87	1
Banne Buiksloot	0.77	74.85	72355.56	447.01	3
Houthavens	0.77	74.07	88344.26	463.92	5
Tuindorp Oostzaan	0.76	71.56	53408.74	248.25	2
Spaarndammer- en Zeeheldenbuurt	0.59	69.19	127328.65	379.58	1
Sloterdijk	0.64	61.21	113090.83	195.29	2
Bijlmer Oost	0.45	61.11	74304.89	805.33	5
Middenmeer	0.59	58.70	199448.69	158.67	1
Slotermeer-Zuidwest	0.48	57.57	223866.13	385.42	1
Middelveldsche Akerpolder	0.62	57.07	134049.63	476.63	5

Holendrecht / Reigersbos	0.47	56.32	52192.93	648.75	3
Oostzanerwerf	0.41	46.50	31078.87	149.43	2
Gein	0.41	46.08	64143.10	580.73	3
Rijnbuurt	0.36	43.30	152965.89	177.64	4
Zeeburgereiland / Nieuwe Diep	0.41	38.63	15449.53	441.50	1
Sloter-/Riekerpolder	0.28	31.58	65595.34	192.22	5
Betondorp	0.49	31.28	166915.80	115.52	4
Buitenveldert-Oost	0.60	27.30	61026.42	158.71	2
Omval/Overamstel	0.18	24.73	74432.77	213.34	4
Buitenveldert-West	0.35	24.15	73774.99	217.15	1
Zuid Pijp	0.77	21.64	307259.82	296.48	1
Nellestein	0.22	21.64	65408.25	164.26	4
Nieuwendammerdijk / Buiksloterdijk	0.28	21.05	41187.08	94.62	5
Noordelijke IJ-oevers Oost	0.28	20.60	45263.24	80.44	2
Noordelijke IJ-oevers West	0.33	20.27	43158.45	28.27	5
Lutkemeer / Ookmeer	0.10	9.02	28051.25	11.60	3
Eendracht	0.02	1.27	22076.70	8.26	2
Waterland	0.02	1.09	62.99	1.53	5
Kadoelen	0.23	1.02	29842.81	4.01	2

Table A.1: A summary of the walking coverage, frequency, accessibility, and demand indicator for all neighbourhoods in Amsterdam. Furthermore, the associated cluster allocation

	Walking Coverage	Supply Frequency	Accessibility
Minimum	0.015	1.02	62.99
Maximum	1	613	484119.23
Mean	0.689	127.58	213577.37
Standard Deviation	0.29	103.17	132961.85
25th Percentile	0.52	61.18	74172.41
50th Percentile	0.76	99.96	229755.72
75th Percentile	0.88	164.44	317015.82

Table A.2: Summary of the descriptive statistics for the walking coverage, supply frequency and accessibility indicator

Bivariate Residuals **Indicators**	Cars per household	Salary per person	Low Income households	House Value	Size of household	Percent young people	Percent old people
Cars per household	.						
Salary per person	2.7690	.					
Low Income households	5.9860	6.8706	.				
House Value	1.1679	4.5978	1.4387	.			
Size of household	3.1556	0.8958	1.8866	1.1664	.		
Percent young people	2.7275	1.8769	2.7752	0.6125	9.5341	.	
Percent old people	5.0744	1.5978	4.2365	1.2860	1.0482	3.5189	.

Table A.3: Bivariate scores in the LCCA clustering of neighbourhoods - without covariates

Bivariate Residuals Indicators	Cars per household	Salary per person	Low Income households	House Value	Size of household	Percent young people	Percent old people
Cars per household	.						
Salary per person	1.3067	.					
Low Income households	3.9828	4.2375	.				
House Value	1.4325	7.9255	1.4151	.			
Size of household	3.6847	0.9090	1.4731	0.9255	.		
Percent young people	4.0497	1.7037	2.3587	1.2932	10.3450	.	
Percent old people	3.8484	1.0682	2.5745	1.5575	2.0075	3.8086	.
Covariates	**Cars per household**	**Salary per person**	**Low Income households**	**House Value**	**Size of household**	**Percent young people**	**Percent old people**
Distance to Amsterdam Centraal Station	2.1879	0.0883	0.7668	7.0367	0.3579	1.6412	0.8293
Degree of Urbanity	4.9504	0.0155	0.2201	0.3609	0.1953	0.2175	1.1461

Table A.4: Bivariate scores in the LCCA clustering of neighbourhoods - with covariates

Bibliography

Agrawal, A. W., Granger-Bevan, S., Newmark, G. L., and Nixon, H. (2017). Comparing data quality and cost from three modes of on-board transit surveys. *Transport Policy*, 54(June 2016):70–79.

Albacete, X., Olaru, D., Paül, V., and Biermann, S. (2017). Measuring the Accessibility of Public Transport: A Critical Comparison Between Methods in Helsinki. *Applied Spatial Analysis and Policy*, 10(2):161–188.

Allen, J. and Farber, S. (2020). Planning transport for social inclusion: An accessibility-activity participation approach. *Transportation Research Part D: Transport and Environment*, 78(January):102212.

Behrisch, M., Senaratne, H., Bustos-Jimenez, J., Schreck, T., Mueller, M., Schneidewind, J., Lalanne, F., and Keim, D. (2017). Urban Mobility Analysis With Mobile Network Data: A Visual Analytics Approach. *IEEE Transactions on Intelligent Transportation Systems*, 19(5):1537–1546.

Bertolaccini, K., Lownes, N. E., and Mamun, S. A. (2018). Measuring and mapping transit opportunity: An expansion and application of the transit opportunity index. *Journal of Transport Geography*, 71:150–160.

Carleton, P. R. and Porter, J. D. (2018). A comparative analysis of the challenges in measuring transit equity: definitions, interpretations, and limitations. *Journal of Transport Geography*, 72:64–75.

Central Bureau for Statistics (2017). Kerncijfers wijken en buurten 2017 [dataset]. *Retrieved from https://opendata.cbs.nl/#/CBS/nl/dataset/83765NED/table?ts=1582034278586*. [Online; accessed 27-October-2020].

Chen, Y., Bouferguene, A., Li, H. X., Liu, H., Shen, Y., and Al-Hussein, M. (2018). Spatial gaps in urban public transport supply and demand from the perspective of sustainability. *Journal of Cleaner Production*, 195:1237–1248.

Clifton, K. J. and Moura, F. (2017). Conceptual framework for understanding latent demand: accounting for unrealized activities and travel. *Transportation Research Record: Journal of the Transportation Research*, 2668(1):78–83.

Currie, G. (2004). Gap analysis of public transport needs: measuring spatial distribution of public transport needs and identifying gaps in the quality of public transport provision. *Transportation Research Record*, 1895(1):137–146.

Currie, G. (2010). Quantifying spatial gaps in public transport supply based on social needs. *Journal of Transport Geography*, 18(1):31–41.

Curtis, C., Ellder, E., and Scheurer, J. (2019). Public transport accessibility tools matter: A case study of gothenburg, sweden. *Case Studies on Transport Policy*, 7(1):96–107.

Deboosere, R. and El-Geneidy, A. (2018). Evaluating equity and accessibility to jobs by public transport across canada. *Journal of Transport Geography*, 73:54–63.

El-Geneidy, A., Levinson, D., Diab, E., Boisjoly, G., Verbich, D., and Loong, C. (2016). The cost of equity: Assessing transit accessibility and social disparity using total travel cost. *Transportation Research Part A: Policy and Practice*, 91:302–316.

Fairley, L., Cabieses, B., Small, N., Petherick, E. S., Lawlor, D. A., Pickett, K. E., and Wright, J. (2014). Using latent class analysis to develop a model of the relationship between socioeconomic position and ethnicity: Cross-sectional analyses from a multi-ethnic birth cohort study. *BMC Public Health*, 14(1).

Farber, S., Mifsud, A., Allen, J., Widener, M. J., Newbold, K. B., and Moniruzzaman, M. (2018). Transportation barriers to Syrian newcomer participation and settlement in Durham region. *Journal of Transport Geography*, 68(September 2017):181–192.

Farber, S., Ritter, B., and Fu, L. (2016). Space-time mismatch between transit service and observed travel patterns in the wasatch front, utah: A social equity perspective. *Travel Behaviour and Society*, 4:40–48.

Fayyaz, S. K., Liu, X. C., and Porter, R. J. (2017). Dynamic transit accessibility and transit gap causality analysis. *Journal of Transport Geography*, 59:27–39.

Ferguson, E. M., Duthie, J., Unnikrishnan, A., and Waller, S. T. (2012). Incorporating equity into the transit frequency-setting problem. *Transportation Research Part A: Policy and Practice*, 46(1):190–199.

Foth, N., Manaugh, K., and El-Geneidy, A. M. (2013). Towards equitable transit: Examining transit accessibility and social need in toronto, canada, 1996-2006. *Journal of Transport Geography*, 29:1–10.

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

Gadepalli, R., Tiwari, G., and Bolia, N. (2018). Role of user's socio-economic and travel characteristics in mode choice between city bus and informal transit services: Lessons from household surveys in Visakhapatnam, India. *Journal of Transport Geography*, 88(August 2018):102307.

Geurs, K. T. and Van Wee, B. (2004). Accessibility evaluation of land-use and transport strategies: review and research directions. *Journal of Transport geography*, 12(2):127–140.

Google (2020). General transit feed specification. *Retrieved from https://developers.google.com/transit/gtfs/reference?csw=1*. [Online; accessed 27-October-2020].

Gutiérrez, J. and García-Palomares, J. C. (2008). Distance-measure impacts on the calculation of transport service areas using GIS. *Environment and Planning B: Planning and Design*, 35(3):480–503.

He, S. Y., Cheung, Y. H., and Tao, S. (2018). Travel mobility and social participation among older people in a transit metropolis: A socio-spatial-temporal perspective. *Transportation research part A: policy and practice*, 118:608–626.

Jaramillo, C., Lizàrraga, C., and Grindlay, A. L. (2012). Spatial disparity in transport social needs and public transport provision in santiago de cali (colombia). *Journal of Transport Geography*, 24:340–357.

Jones, P. and Lucas, K. (2012). The social consequences of transport decision-making: Clarifying concepts, synthesising knowledge and assessing implications. *Journal of Transport Geography*, 21:4–16.

Jung, Y. J. and Casello, J. M. (2019). GIS-based transit trip allocation methods converting stop-level boarding and alighting trips into TAZ trips. *Transportation Planning and Technology*, 42(8):848–867.

Karner, A. and Niemeier, D. (2013). Civil rights guidance and equity analysis methods for regional transportation plans: A critical review of literature and practice. *Journal of Transport Geography*, 33:126–134.

Legrain, A., Buliung, R., and El-Geneidy, A. M. (2016). Travelling fair: Targeting equitable transit by understanding job location, sectorial concentration, and transit use among low-wage workers. *Journal of Transport Geography*, 53:1–11.

Liu, D. and Kwan, M. P. (2020). Measuring spatial mismatch and job access inequity based on transit-based job accessibility for poor job seekers. *Travel Behaviour and Society*, 19:184–193.

Lucas, K. (2019). A new evolution for transport-related social exclusion research? *Journal of Transport Geography*, 81(October):102529.

Molin, E., Mokhtarian, P., and Kroesen, M. (2016). Multimodal travel groups and attitudes: A latent class cluster analysis of Dutch travelers. *Transportation Research Part A: Policy and Practice*, 83:14–29.

Nazari Adli, S. and Donovan, S. (2018). Right to the city: Applying justice tests to public transport investments. *Transport Policy*, 66:56–65.

Nederlandse Spoorwegen (NS) (2019). Reizigersgedrag 2019 [dashboard]. *Retrieved from https://dashboards.nsjaarverslag.nl/reizigersgedrag/*. [Online; accessed 27-October-2020].

Páez, A., Mercado, R. G., Farber, S., Morency, C., and Roorda, M. (2010). Relative accessibility deprivation indicators for urban settings: Definitions and application to food deserts in montreal. *Urban Studies*, 47(7):1415–1438.

Paez, A., Ruben, M., and Faber, S. (2009). Mobility and social exclusion in canadian communities: An empirical investigation of opportunity access and deprivation from the perspective of vulnerable groups. *Policy Research Directorate Strategic Policy and Research, Toronto, Canada*.

Pedigo, A., Seaver, W., and Odoi, A. (2011). Identifying unique neighborhood characteristics to guide health planning for stroke and heart attack: Fuzzy cluster and discriminant analyses approaches. *PLOS ONE*, 6(7):1–11.

Ricciardi, A. M., Xia, J. C., and Currie, G. (2015). Exploring public transport equity between separate disadvantaged cohorts: A case study in perth, australia. *Journal of Transport Geography*, 43:111–122.

Tovaas, K. (2020). Typology and description of underlying principles and generalisable lessons. *INCLUSION Project*, D3.4.

Uitbeijerse, G., Kampert, A., and Nijenhuis, J. (2019). Risico op vervoersarmoede onder huishoudens in utrecht en heerlen. *Bijdrage aan het Colloquium Vervoersplanologisch Speurwerk*.

van Oort, N., Brands, T., and de Romph, E. (2015). Short-Term Prediction of Ridership on Public Transport with Smart Card Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2535(1):105–111.

Vermunt, J. K. and Magidson, J. (2003). Latent class models for classifcation. *Computational Statistics & Data Analysis*, 41(December):531–537.

Wang, Y., Huang, Y., Xu, J., and Barclay, N. (2017). Optimal recharging scheduling for urban electric buses: A case study in Davis. *Transportation Research Part E*, 100:115–132.