Exploring Gender Disparities in Accessibility Levels

A Comprehensive Analysis of Transportation Modes, Activity Types and Personal Characteristics

---

Master thesis submitted to Delft University of Technology

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in **Complex Systems Engineering and Management**

Faculty of Technology, Policy and Management

by

Luisa de La Vega Bayma de Oliveira

Student number: 5584698

To be defended in public on August 25th, 2023

Graduation committee

Chairperson:	Dr.ir. M. Kroesen, Transport and Logistics – TU Delft
First Supervisor:	Dr. T. Verma, Multi Actor Systems – TU Delft
External Supervisor:	Dr. J. Gonçalves, Architecture and the Built Environment - TU Delft

Preface

My journey through the Master's program in Complex Systems Engineering and Management at Delft University of Technology culminates with this thesis. Combining my acquired expertise in Data Science techniques from the program with a personal interest in feminist subjects, I set out to contribute to a more equitable world. The exploration of accessibility, inherently complex in both engineering and management aspects, seamlessly aligns with the essence of the Master's program.

I extend heartfelt gratitude to those who have been instrumental on this path. First, I am immensely thankful to Juliana Gonçalves for not only identifying this compelling topic but also providing unwavering guidance and insightful advice throughout the entire process. I owe a debt of gratitude to Trivik Verma for his invaluable insights and for sparking my passion for data science through his course. I am also deeply appreciative of Maarten Kroesen for his invaluable advice, as well as for agreeing to chair the committee.

Additionally, I extend my gratitude to my parents, Sandra de La Vega and Lucio Bayma, who have always been the strongest advocates for education, inspiring me to pursue my aspirations. To my fiancé, Gabriel Cassano, your ceaseless encouragement and shared journey have been indispensable. My heartfelt thanks to my cherished friends who have brought light and joy to this path. A special acknowledgment to TU Delft and its exceptional staff for imparting education at the highest standard.

It is my aspiration that this research contributes to a more equitable transportation system, and that the methodology I have employed here paves the way for new studies and further scientific exploration.

Luisa de La Vega Bayma de Oliveira

Delft, August 11, 2023

Executive Summary

Accessibility is a fundamental concept concerning urban and transport planning as it is the elementary basis for socio-economic development in cities. It can be described as the potential to reach spatially distributed opportunities. Recent research has identified that excluding person-based features from this analysis might cause inaccurate measurement of accessibility. Age, gender, and income, for instance, are responsible for varying accessibility levels drastically. In addition, studies worldwide have shown that women face different challenges in reaching locations and spatially distributed opportunities. Despite the findings, this research identifies a gap in understanding how a person-based perspective, mainly gender, and other personal characteristics, affect accessibility levels, considering various travel purposes and transport modes.

Thus, this research aims to answer the question, "How do personal characteristics, mainly gender, can impact accessibility levels?". The primary objective is to explore how these characteristics influence accessibility metrics, identify the urban groups most affected by the absence of this perspective, and determine the key personal characteristics that significantly impact accessibility levels. To address these objectives, this study considers a combination of quantitative and case-study research approaches. It investigates the Metropolitan Region Rotterdam-The Hague in The Netherlands by applying surveys to obtain perceived accessibility data. In the same region, spatial analysis is conducted by mapping transport networks and points of interest. Then, this study compares spatially calculated accessibility with self-reported accessibility and the presence of mismatches. In addition, cluster analysis identifies the urban profiles most vulnerable to mismatches and their main characteristics. A Binary logistic regression is conducted to determine the variables' importance in the mismatch occurrence.

From the survey answers, it is identified that women have less access to cars than men. In addition, the comparison between perceived accessibility and spatial accessibility uncovers that women present the most critical mismatches to reach activities by car. In other words, several women perceive the car as an impractical option to access points of interest that are spatially considered reachable by car. It raises the hypothesis that their lack of car access highly impacts their accessibility perception. Furthermore, the clustering analysis reveals that foreign women exhibit a higher prevalence of car-related mismatches when compared to other urban groups. Moreover, this research identifies that fathers of young children also encounter greater disparities across all transportation modes. Additionally, the binary logistic regression underscores the importance of safety as a critical factor influencing women's perception of walking as a viable mode of transportation. This safety importance is also identified from the survey answers.

However, it is essential to acknowledge the limitations of this study, including a small and potentially biased sample size of perceived accessibility data collected from surveys. These limitations can have an impact on the spatial and perceived accessibility comparison, clustering analysis, and logistic regression outcomes. Additionally, the reliability of open-source points of interest descriptions and the sensitivity of threshold definitions impose constraints on spatial analysis.

Several recommendations for further investigation are proposed based on the research findings and limitations. Firstly, an in-depth analysis should be conducted to understand the barriers fathers of young children face when accessing proposed activities. Furthermore, a comprehensive investigation of car access issues among women, particularly foreigners, is recommended. Additionally, evaluating how safety perception of women varies across different spatial contexts can provide valuable insights into factors impacting women's nighttime walkability.

In summary, this study contributes significant value to urban science area by employing a unique combination of techniques to examine the impact of personal characteristics on accessibility levels. The findings raise new hypotheses that warrant investigation in transport engineering, urban planning, and social sciences. Consequently, this research can contribute to developing more inclusive transport policies and establishing a more equitable transport system.

Contents

List of Figures

List of Tables

1. Introduction

Access to people, goods, services, and information forms the foundation of socio-economic development in cities. To comprehend and maintain a certain level of accessibility for citizens, it has become crucial to measure this access, considering each city's distinct spatial structure and transportation system (Rode et al., 2017). Furthermore, the accessibility metric serves as a valuable tool in identifying social inequalities and highlighting the necessity for adjustments in transportation systems or land use patterns. By utilizing accessibility as a key indicator, cities can effectively address disparities, foster sustainable transportation practices, and enhance overall urban well-being (Bhat et al., 2000).

Thus, accessibility is a concept of continuing relevance in urban planning and transportation research. Despite many different definitions, it can be commonly described as the potential for reaching spatially distributed opportunities for employment, recreation, and social interaction (Páez et al., 2012). However, scholars might consider different approaches to obtain this measurement; some studies focus on spatial metrics (Handy & Niemeier, 1997), while others include metrics based on personal characteristics (Páez et al., 2012; Ryan & Pereira, 2021; Tiznado-Aitken et al., 2020).

The physical movement to reach the opportunities, also called mobility, plays an essential role in accessibility. However, accessibility does not just consider people's mobility. Other factors might affect their capability to access opportunities, such as the quality and affordability of transport options and people's ability to use them (Levinson & King, 2020). Thus, access sees movement as means, not an end. The end is the ability to participate in the intended activity of the traveler (Levinson & King, 2020).

Accessibility levels vary dramatically according to individuals' characteristics, such as age, gender, income, or physical and cognitive functioning (Ryan & Pereira, 2021). The distinct ability to participate in intended activities can be identified when analyzing women's main mobility constraints compared to men's. According to studies worldwide, mainly qualitative, many women face frightening situations in their everyday mobility and women feel particularly unsafe in the darkness in car parks, garages, underpasses, and public parks (Stark & Meschik, 2018). Consequently, it might cause women to have restricted mode choices for travel at this time of the day (P. Zhang et al., 2022). Moreover, women choose the nearest alternative due to gendered everyday-life constraints (Gil Solá & Vilhelmson, 2022).

All these factors show that women face different challenges in reaching locations and spatially distributed opportunities. Although cities are becoming more aware of the importance of emphasizing women's needs in urban public spaces, it still needs to be explicitly determined what needs to be done (Priya Uteng et al., 2019). In addition, women's mobility is historically disciplined through patriarchal control (Dulhunty, 2022). Therefore, excluding personal characteristics from the study and definition of transport systems and urban planning analyses might generate overestimated accessibility levels while individual differences are underestimated (Ryan & Pereira, 2021). Consequently, not considering personal characteristics in accessibility analysis can lead to a distorted understanding of accessibility levels, underestimated individual differences, perpetuated gender disparities, and missed opportunities for inclusive urban planning.

The literature review reveals a knowledge gap in investigating the impact of gender and other personal characteristics on accessibility levels, comparing person-based perspective with location-based perspective and including (safety) perception in this analysis. The current body of quantitative (Páez et al., 2012; Ryan & Pereira, 2021) and qualitative (Geurs & Van Wee, 2004)literature on accessibility calculation has explored location-based and person-based perspectives but without a primary focus on gender and other personal characteristics. Similarly, quantitative (Gil Solá & Vilhelmson, 2022; Havet et al., 2021; Lo & Houston, 2018; Tiznado-Aitken et al., 2020) and qualitative (Loukaitou-Sideris, 2020; Porrazzo et al., 2022; Priya Uteng et al., 2019) studies specifically investigating gender and accessibility have been limited to specific contexts such as one transport mode or activity type. In addition, none of these studies compare perceived accessibility (person-based) with spatial accessibility (location-based). Also, most studies do not include participants' (safety) perception, which is a critical component when investigating accessibility disparities among men and women.

Thus, the main research question of this study is "How do personal characteristics, mainly gender, can impact accessibility levels?". This research aims to investigate the difference in accessibility metrics when a person-based perspective is considered. Apart from a primary investigation of men's and women's perceived accessibility distinction, this study analyzes other variables such as economic level, ethnicity, and age. This research investigates The Metropolitan Region Rotterdam The Hague (MRDH) in The Netherlands. However, it aims to become reproducible research in other areas.

This research is a Master's Thesis of the program Complex System Engineering and Management (CoSEM) at Delft University of Technology (TU Delft). This study is linked to this program due to several reasons. Firstly, it consists of the analysis of a socio-technical system. The transport network options describe the technical component, while the transport system users represent the social component. In addition, measuring accessibility is complex since it relies on a non-linear and dynamic network. Therefore, accessibility is modeled and evaluated based on CoSEM methodologies, such as network analysis combined with urban data science techniques and statistical approaches.

The results represent relevant value for society and science since this study aims to provide material for further research and support gender equality in urban planning and transport policies. Moreover, this research represents a multidisciplinary scholarly topic that combines transport engineering, urban planning, and social science. Therefore, the study aims to evaluate relevant distinctions in accessibility perceptions, and the outcomes are potentially valuable to designing interventions in all disciplines.

The structure of this research is as follows: Chapter Two provides an extensive analysis of the theme, focusing on the relationship between gender and accessibility through a comprehensive literature review. In this Chapter, the sub-research questions are described. In Chapter Three, the research approaches of this study are presented. The methodologies employed to address these sub-research questions and ultimately answer the main research question are described in Chapter Four. The findings obtained from the analysis are presented in Chapter Five. Subsequently, Chapter Six delves into a detailed discussion of the results obtained in the previous analysis. Finally, Chapter Seven concludes the project by providing answers to the research questions proposed, thereby offering a comprehensive summary of the research findings.

2. Theoretical Background

This chapter leads to a literature review regarding gender and accessibility. The following sections describe the literature review approach and its main findings. In this sequence, the research gap of this study is presented.

2.1 Literature Review Approach

This literature review is divided into two main parts. First, it investigates accessibility definitions and studies that analyze the heterogeneity of people in accessibility metrics. This part's goal is to model accessibility for this research. Second, this review analyses the main findings about the relationship between gender and accessibility and how the former might impact the other.

This literature review considers the article selection based on three primary sources/techniques. The supervisors indicated some articles, others were searched on Scopus, and finally, one piece was found through citation tracking. Scopus is used because it is a reputable platform covering recent articles, journals, and book publications. Since our primary focus is on the female gender and it was previously stated that mobility is highly related to accessibility, this search process considers a combination of different words such as "Gender," "Women," and "Woman" combined with "Accessibility" and "Mobility."

The selection criteria for the articles available on Scopus is first based on the abstract and its relevance to the problem discussed. In addition, this research prioritizes the most recent articles published, considering that the relationship between women and accessibility/ mobility might have changed in the last few decades. Also, this process finds geographical scope as a criterion. It prioritizes studies concerning countries or regions with similar economic conditions as The Metropolitan Region Rotterdam The Hague (MRDH)-The Netherlands, the chosen geographical area for this research. Considering these criteria, all articles analyzed in this literature review are described in Table 2.1.

Table 2.1: Research techniques and respective articles selected

Research string or source	Author	Research focus	Year	Title	Research Platform	Geographical Scope
Indicated by the supervisor	Páez et al.	Accessibility Definition	2012	Measuring accessibility: positive and normative implementations of various accessibility indicators	-	Canada
Indicated by the supervisor	Ryan & Pereira	Accessibility Definition	2021	What are we missing when we measure accessibility? Comparing calculated and self-reported accounts among older people	-	Sweden
Indicated by the supervisor	Van Wee & Geurs	Accessibility Definition	2004	Accessibility evaluation of land-use and transport strategies: review and research directions	-	None

Indicated by the supervisor	Tiznado-Aitken et al.	Accessibility Definition, Gender and Accessibility	2020	Understanding accessibility through public transport users' experiences: A mixed methods approach.	-	Chile
"Gender", "Mobility"	Havet et al.	Gender and Accessibility	2021	Why do Gender Differences in Daily Mobility Behaviours persist among workers?	Scopus	France
"Accessibility, " "Gender"	Lo & Houston	Gender and Accessibility	2018	How do compact, accessible, and walkable communities promote gender equality in spatial behavior?	Scopus	United States of America
"Gender, "Mobility"	Porrazzo et al.	Gender and Accessibility	2022	Gender and mobility planning: The influence of national culture on planning processes.	Scopus	Denmark
'Gender,' 'Proximity'	Gil Solá & Vilhelmson	Gender and Accessibility	2022	To choose, or not to choose, a nearby activity option: Understanding the gendered role of proximity in urban settings	Scopus	Sweden
Citation Tracking	Loukaitou-Sideris	Gender and Accessibility	2020	Engendering Cities: Designing sustainable space for all: A Gendered View of Mobility and Transport	-	None
'Gender,' 'Mobility'	Priya Uteng et al.	Gender and Accessibility	2021	Chapter Two - Gender gaps in urban mobility and transport planning	Scopus	Both Global North and Global South

2.2 Accessibility definitions and metrics

Despite the vast amount of studies regarding accessibility, its definition considerably varies depending on the analysis purposes. It can be defined as the potential for reaching spatially distributed opportunities for employment, recreation, and social interaction (Páez et al., 2012). Some scholars also include the ease with which this potential can be realized (Ryan & Pereira, 2021; Tiznado-Aitken et al., 2020), the travel costs, and the weight of the activity (Tiznado-Aitken et al., 2020). Depending on the approach, it also includes choice behavior (Wu & Levinson, 2020). The variation of accessibility metrics can be explained by its multi-disciplinary factor. Accessibility metrics come from Topological, Engineering, and Planning; Economics, Computer Science, and Network Science (Wu & Levinson, 2020).

Another relevant accessibility definition provided by Van Wee and Geurs (2004) describes it in four main components: land use, transport, temporal and individual. The land-use component is related to the spatial distribution of activities; the transport component describes the transport system, such as costs and comfort-related; the temporal component reflects temporal constraints, such as the availability of opportunities at different times of the day and finally, the individual component reflects the needs, abilities, and opportunities of individuals. Applied accessibility metrics focus on one or more components of accessibility, depending on the perspective taken.

In addition to the components, four essential perspectives on measuring accessibility can be defined (Geurs & Van Wee, 2004): infrastructure-based metrics, which analysis the transport infrastructure; location-based which analyses the spatial distribution of activities; person-based which consider the activities in which an individual can participate at a given time and utility-based metrics which analysis the economic benefits that people derive from access.

One of the most common accessibility metrics is called Cumulative Opportunities, which considers the location-based perspective (Levinson & King, 2020). It calculates the number of destination opportunities (O) that can be reached from origin i to destination j, constrained by some cost measure (C_{ij}) function (f), as described in Equation 2.1.

$$A_i = \sum_{j=1}^{J} O_j f(C_{ij}) \tag{2.1}$$

The Primal Access metric considers the number of opportunities reached from a fixed cost. Thus, the cost function is constrained by a threshold (t). If the cost is higher than the threshold, the cost function is equal to 0, which annulates the opportunity (O). The function f (C_{ij}) can be described by Equation 2.2.

$$if\ C_{ij}\ \leq t, f\left(C_{ij}\right) = 1, else\ f\left(C_{ij}\right) = 0 \tag{2.2}$$

On the contrary, the Dual Access measure considers the time or cost required to reach a fixed number of opportunities (Wu & Levinson, 2020). Therefore, the primary distinction between the Primal and Dual Access metrics is that the former uses a fixed cost threshold to determine the number of opportunities that can be reached, whereas the dual measure considers the number of opportunities that can be reached to be constant and access is determined by the cost of doing so. The primal and dual metrics are variations of cumulative opportunities metric (Wu & Levinson, 2020).

The function $f\left(C_{ij}\right)$ can be called an impedance factor. Time, distance, money cost, and other travel-related expenses impede travel and reduce access. Some of the most used travel cost approaches are distance or travel time from origin to destination (Wu & Levinson, 2020).

Based on these concepts and equations, scholars have analyzed how the heterogeneity of people impacts the variation of accessibility measurements. In other words, they had included the person-based accessibility perspective on it. Measures of accessibility that consider people's qualities and limitations can be beneficial for social assessments of modifications to land use and transportation (Geurs & Van Wee, 2004). Furthermore, overlooking the heterogeneity in people's perception of their accessibility tend to overestimate accessibility levels and underestimate accessibility inequalities (Ryan & Pereira, 2021).

To analyze accessibility calculation, including the person-based perspective, Páez et al. (2012) distinguish between two accessibility implementations called normative and positive. They describe normative as the desired behavior of travelers or location of services, whereas positive is the actual behavior of travelers or area of services. Páez et al. (2012) rely on the primal access metric and differ normative and positive approaches by calculating thresholds (t) differently.

In their normative approach, the threshold is calculated by the average travel distance of a specific group based on a previous Travel Survey. For instance, a threshold of 3.6 km is the average trip length of women aged 20–35 in the Montreal region (Páez et al., 2012). On the other hand, the positive approach calculates a threshold based on the expansion method. It is a simple tool to generate models with spatially-varying coefficients, which allow the analyst to obtain location- and person-specific estimates of distance traveled. It is a data-intensive method that calculates the threshold considering coefficients based on a person's income, household composition, age, vehicle possession, and other characteristics.

Ryan and Pereira (2021) also focus on the heterogeneity of individual characteristics, mainly older people, and its impact on accessibility measurements. Part of their study also relies on the primal access measurement and calculates the number of critical activities reachable within a timeframe from the study participants' address. For that, it maps the transport network system of public transport, car, cycling, and walking options in a city. It calculates the reachable amount of opportunities for each transport mode type. It categorizes as "less accessible" if the number of critical activities is below the neighborhood's mean and "more accessible" if it is above the mean. In parallel, this study applies surveys to obtain individuals' perceptions about their capabilities to reach an essential activity by a transport mode. Therefore questions were asked regarding the feasibility of using a transport mode to get a key activity from the participant's perspective.

Moreover, Tiznado-Aitken et al. (2020) present a conceptual framework that combines quantitative and qualitative approaches to calculate "Perceived Accessibility." The qualitative approach collects empirical data about user experience and accessibility perception through interviews, and it enables the definition of user-profiles and their main characteristics. For example, one group is more likely to be composed of young people, men, and higher-income users, while the other group is likely composed of women, older people, and low-income users. According to these user profiles, this study defines coefficients such as different walking speeds depending on gender and age categories, different perceptions of waiting time for men and women, and different sensitivity to comfort or crowded conditions. The coefficient values are based on previous studies.

Then, these coefficients are incorporated into the quantitative approach, which calculates the travel time expression of trips considering different transport modes. Table 2.2 summarizes the studies' approach to incorporate person-based analysis on accessibility measurements.

Table 2.2: Research approaches to measure accessibility and include person-based perspective

Author	Accessibility Equation	Location-based perspective	Person-based perspective incorporation	Accessibility Analysis
Páez et al. (2012)	Primal Access Measure Cumulative Opportunities $A_i = \sum_{j=1}^{J} O_j\, I(C_{ij})$ $I(c_{ij} \leqslant \gamma_i) = \begin{cases} 1 & \text{if } c_{ij} \leqslant \gamma_i \text{ (threshold value)} \\ 0 & \text{otherwise} \end{cases}$	This study does not calculate accessibility exclusively by location-based perspective.	Terminology: Normative γ_i = average travel distance of a group. E.g: 3.6 km Terminology: Positive γ_i = expansion method	Comparison between Normative and Positive Cumulative Opportunities Measurements
Ryan & Pereira (2021)	Primal Access Measure Cumulative Opportunities $A_i = \sum_{j=1}^{J} O_j\, I(C_{ij})$ $I(c_{ij} \leqslant \gamma_i) = \begin{cases} 1 & \text{if } c_{ij} \leqslant \gamma_i \text{ (threshold value)} \\ 0 & \text{otherwise} \end{cases}$	Terminology: Calculated Categorized as less or more accessible areas based on the mean of opportunities (O_j) available in a 30-minute timeframe (γ_i) by a transport mode.	Terminology: Self-reported Perceived capability to use a transport mode to reach the opportunities in a 30-minute timeframe (γ_i).	Comparison between Calculated and Self-reported capability of using a transport mode to less and more accessible areas.
Tiznado-Aitken et al. (2020)	Travel time expression $t_{ijk} = \alpha_c \bullet \frac{d_c}{v_k} + \beta_{wk} \bullet t_w + \varepsilon_{ck} \bullet t_t + p_t \bullet n_t$	Terminology: Quantitative Quantitative travel time measurements, a gender-neutral assessment, without differences between age and income.	Terminology: Qualitative User-profile definition from surveys Terminology: Quantitative Calculation of travel time t_{ijk} based on coefficients calculated according to the different user-profile definition	Incorporation of Qualitative data and person-based coefficients on quantitative measurements

The three studies defend that these perspectives can complement one another, leading to better accessibility and understanding and, consequently, better policy outcomes. Páez et al. (2012) describe that positive measures consider the connection between accessibility and social ideals of inclusion and how accessibility varies in disadvantaged populations. According to Ryan and Pereira (2021), traditional accessibility measurements frequently assume that everyone can use a particular mode of transportation to the same degree, which tends to overstate accessibility levels and underestimate accessibility inequities. Tiznado-Aitken et al. (2020) show that it is possible to determine attributes previously overlooked in quantitative studies but are relevant for analyzing the accessibility of population groups. Tiznado-Aitken et al.

(2020) is the only study that presents accessibility analysis distinguished by gender, however, it is not the core of the research.

2.3 Gender and Accessibility correlation

The second part of this literature review analyses how gender influences accessibility perceptions. The main goal of this literature review part is to comprehend the methodologies and outcomes employed by studies that predominantly delve into the intersection of gender and accessibility.

As described previously, mobility is considered a means to access distributed opportunities, and thus, constraints in mobility generate a relevant impact on accessibility results. This literature review identifies gendered distinctions in accessibility aspects, such as mobility barriers, activities preferences, quality, affordability, and travel options. These are seen as relevant aspects that impact the potential to access desired locations and opportunities.

According to Tiznado-Aitken et al. (2020), women are usually more worried about safety and comfort. Fear of sexual harassment and personal security remain significant concerns in negotiating daily mobilities (Priya Uteng, 2021). They change their travel patterns more than men, modifying transport routes, modes, and times and even deciding not to travel. As a consequence, cars or other private alternatives are usually considered. Furthermore, compared to women, men have larger activity spaces and conduct their activities further from home (Lo & Houston, 2018; Porrazzo et al., 2022; Priya Uteng, 2021). Women choose the nearest alternative due to gendered everyday-life constraints and likely environmental concerns (Gil Solá & Vilhelmson, 2022).

In addition, unemployed and employed women shoulder more household responsibilities in childcare and maintenance tasks (Havet et al., 2021; Lo & Houston, 2018); however, this behavior may vary considerably according to the urban space distribution. In the case of compact urban development, couples have greater flexibility to divide household activities outside of the home. On the contrary, the flexibility is lower in the suburbs due to their dispersed land uses. Therefore, close urban development can help alleviate gender inequalities in households' out-of-home responsibilities.

Moreover, in the case of a more distant alternative, women use public transport more than a car, while men use more cars than public transport use. It confirms that women use cars to a lesser extent than men (Priya Uteng, 2021). Gender differences in travel patterns appear to be as much related to differences in access to the car and various factors that strongly influence mobility (e.g., age, number of children, level of education, and income) (Havet et al., 2021).

Furthermore, (Loukaitou-Sideris, 2020) describes four main barriers to mobility among women. Women might face a cultural barrier considering that they are primarily responsible for domestic chores and caregivers for children in many cultures. In addition, there are economic barriers where women lack financial resources for a car, for example. The lack of adequate infrastructure for walking with strolls and kids represents a physical barrier, while the fear of harassment is a psychological barrier.

In brief, this part of the literature review supports that women and men face different constraints to access distributed opportunities spatially. For instance, they might be related to economic or psychological barriers to using a transport mode or their perception of reasonable travel time. Table 2.3 describes the main findings of this literature review part, the correspondent authors, research approach and methodologies. The methodologies are specified according to the four perspectives to measure accessibility described by Van Wee & Geurs (2004): infrastructure, location, person and utility-based. Notably, the majority of the methodologies adopt a person-based perspective, as the focus of the papers is on investigating gender, an individual factor influencing activity participation.

Table 2.3: Main Findings regarding gender and accessibility correlation

Authors	Approach	Methodology	Main Findings
Gil Solá and Vilhelmson (2022)	Quantitative and Case Study	Surveys (Person-based) Urban Densification Analysis (Location-based)	Women choose the nearest alternative, which indicates the presence of gendered everyday-life constraints.
Porrazzo et al. (2022)	Qualitative	Literature Review (Person-based) Experts Interviews (Person-based)	
Priya Uteng (2021)	Qualitative	Literature Review (Person-based)	Women use cars to a lesser extent than men. Women change travel patterns more than men.
Lo and Houston (2018)	Quantitative and Case Study	Survey (Person-based) Built environment indicators (Location-based)	Women shoulder more household responsibilities in childcare.
Havet et al. (2021)	Quantitative and Case Study	Survey (Person-based)	Differences in car access and various factors strongly influence mobility patterns.
Tiznado-Aitken et al. (2020)	Quantitative, Qualitative and Case Study	Surveys (Person-based) Interviews (Person-based) Public Transport Data Analysis (Location and infrastructure-based)	Women are usually more worried about safety and comfort, impacting their transport mode choices.
Loukaitou-Sideris (2020)	Qualitative	Literature review (Person-based)	Women face physical, economical, psychological, and cultural mobility barriers.

2.4 Research Gap

The first part of this literature review has shown that in addition to research regarding the location-based perspective on accessibility calculation, scholars have focused on including a person-based perspective. However equally important, the infrastructure and utility-based perspectives are seen as the last common in this research scope. Páez et al. (2012) include an expansion method to obtain person-based accessibility thresholds. Ryan & Pereira (2021) compares calculated with self-reported accessibility measurements for older adults. Tiznado-Aitken et al. (2021) combine quantitative and qualitative approaches to obtain 'perceived accessibility measurements' in public transport. Van Wee & Geus (2004) does not calculate the accessibility of a specific case but described the fundamental concepts of this research area.

Furthermore, the second part of the literature review describes how gender is incorporated in studies that investigate accessibility levels. Most studies consider the person-based perspective. This research considers that including person-based and location-based analysis enriches the accessibility measure and enables the comparison between both views. Consequently, it is easier to understand the impact of including personal characteristics in accessibility analysis.

In addition, this literature review highlights the significant impact of safety on women's accessibility levels. However, measuring safety is challenging due to its subjective nature, varying according to individuals' perceptions. Consequently, this research argues that integrating perception into the person-based perspective is crucial for exploring the correlation between gender and accessibility. By doing so, a more comprehensive understanding of the relationship between these factors can be achieved.

Figure 2.1 describes each literature review part, and how they are connected. The first part of the literature review focuses on understanding the main perspectives considered when calculating accessibility. The second part investigates the correlation between gender and accessibility. Both literature review parts may overlap when accessibility measures focus on gender.

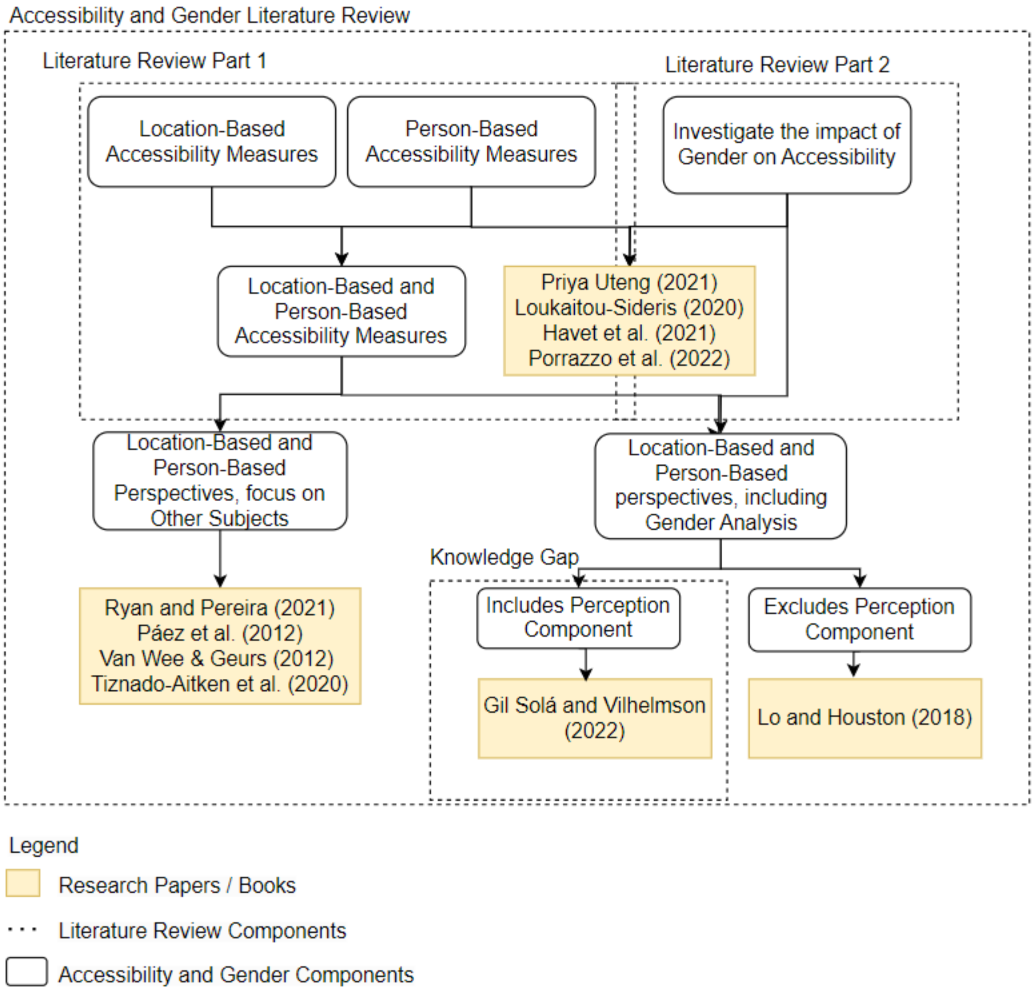


Figure 2.1: Literature Review Diagram

Some studies consider person-based and location-based perspectives on accessibility measures, however, they do not primarily focus on gender (Geurs & Van Wee, 2004; Páez et al., 2012; Ryan & Pereira, 2021; Tiznado-Aitken et al., 2020). Other studies focus primarily on gender and accessibility, however including mainly the person-based perspective and not the spatial one (Havet et al., 2021; Loukaitou-Sideris, 2020; Porrazzo et al., 2022; Priya Uteng, 2021).

Two specific studies that prominently incorporate both person-based and spatial perspectives while primarily focusing on the intersection of gender and accessibility have been identified. (Gil Solá & Vilhelmson, 2022; Lo & Houston, 2018). Among these two studies, only the research conducted by Gil Solá and Vilhelmson (2022) includes a perception component, asking through surveys about people's perception of accessibility options.

Despite the similar research scope, this analysis identifies a knowledge gap in this research scope. First, only this study included all the components mentioned above. However, there are no studies that compare person-based and spatial perspectives. In addition, this study does not investigate safety perception which is described as a critical component of female accessibility. Finally, this study is focused on a specific context and geographical scope, analyzing how men and women choose the nearest alternatives, in Sweden. Further research is essential to explore these disparities, encompassing diverse travel purposes and modes while taking into account individuals' perceptions of their accessibility.

Therefore, the significance of this research stems from its innovative methodological blend, which not only facilitates a distinct comparison between the location-based and person-based perspectives but also incorporates a crucial element of (safety) perception. Furthermore, this study goes beyond the conventional by analyzing a diverse array of transportation modes and points of interest. This comprehensive approach emphasizes the unique importance of this research, as it reveals fresh insights into the differences in accessibility levels when examining factors such as gender and other individual characteristics.

2.5 Sub-Research Questions

Based on the knowledge gap, this research defines the sub-research questions that support the investigation of the main research question, which means, how gender and other personal characteristics impact accessibility levels.

For that, this study gathers two main types of data. One is spatial accessibility data, which means the transport networks and the distribution of amenities in this study's geographical scope. Additionally, this research gathers and analyses perceived accessibility data by survey distribution.

These two data types are chosen because spatial analysis excludes a person-based perspective while surveys do not. Therefore, it enables comparing measurements and the impact of including personal characteristics in accessibility levels. According to the literature review, including a person-based perspective generates a significant variation in accessibility levels. Thus, the first sub-research question of this study is:

1. How can perceived accessibility differ from spatial accessibility?

To answer the first sub-research question, this study investigates the number of mismatches between spatial and perceived accessibility levels (Ryan & Pereira, 2021). It incorporates accessibility self-perception from surveys, enables the analysis of different transport modes, and is based on cumulative opportunity accessibility measurement.

The second part of this research consists of further investigating the groups that present more mismatches when comparing both accessibility perspectives. Thus, the second sub-research question of this study is:

2. What urban groups present an accessibility perception that differs the most from spatially calculated accessibility?

After analyzing the accessibility mismatches and identifying the most impacted urban profiles, this research aims to identify the most relevant personal characteristics that might influence the occurrence of mismatches. As the literature review described the accessibility barriers among women, this part of the analysis focuses mainly on mismatches among women. Thus, the third sub-research question of this analysis is:

3. What are the most influential personal characteristics that impact accessibility perception of women?

3. Research Approach

This chapter discusses the research approaches considered for this study and elucidates the rationale behind their selection. It begins showing insight into the chosen Research Approaches and the underlying motivations. Furthermore, it offers a general overview of how each approach will be implemented. A comprehensive understanding of each methodology and its procedural steps are later presented in Chapter 4: Methodology.

3.1 Research approaches and their motivation

This study considers a Quantitative Research Approach. This approach is chosen since is useful when working to confirm or test a hypothesis. Additionally, it is easier to generalize (Sardana et al., 2023). This research hypothesizes that gender and other personal characteristics highly impact accessibility levels. This analysis focuses on understanding the relationship between Gender and Accessibility and evaluating how the former relates to the other. However, this research analyses not only gender but several other personal characteristics.

Quantitative research, in contrast to qualitative research, deals with data that can be converted into numbers (Sheard, 2018). Thus, the quality of the data gathered, and the decisions and interpretations about the numerical data considerably impact the results, being a limitation of this approach in case of inappropriate decisions conducted by the researcher.

In addition to the Quantitative Research Approach, this research considers a Case-Study Research Approach. This approach is beneficial when there is a need to obtain an in-depth appreciation of an issue, event, or phenomenon of interest in its natural, real-life context (Crowe et al., 2011). It produces a multifaceted knowledge of a complicated problem in its current situation, and it is helpful to explore events or phenomena in everyday contexts, and it is widely applied in a range of fields, especially social sciences (Crowe et al., 2011).

As mentioned previously, the nature of this research theme is a combination of transport engineering, urban planning, and social science. Moreover, it investigates accessibility levels in transport which is a complex subject regarding society's everyday mobility. Thus, including a Case-Study approach suits this research scope and goals.

This study investigates the Rotterdam–The Hague metropolitan area in The Netherlands. This area is chosen considering its diversity. Despite being less evident, the differences in mobility patterns according to gender are described in developing and developed countries (Havet et al., 2021).

Rotterdam, which is the second-largest city in The Netherlands, by the number of inhabitants (Statista, 2022), is diverse in many aspects. It has been historically shaped by migration (Schiller et al., 2023), and it became one of the leading continental port cities at the end of the nineteenth century (van de Laar & van der Schoor, 2019). Consequently, it attracted many low-skilled labor immigrants (Entzinger & Engbersen, 2014). Despite the current attraction of high-skilled immigrants, the predominant population is still low-skilled, and the city presents' a higher share of the immigrant population in the country. Its diversity can be compared to big capitals such as Amsterdam and New York (Entzinger & Engbersen, 2014). Another relevant characteristic for this analysis is that Rotterdam has historically evolved with a car-based 'mobility regime' (Loorbach et al., 2021). Therefore, the diversity of Rotterdam's population and transport mode preferences can contribute to diverse accessibility perceptions.

In addition, The Hague is a fast-growing city and the third largest city in the Netherlands after Amsterdam and Rotterdam. More than half of the city's residents have an immigrant background, and the population composition is very different depending on the neighborhood (Gemeente Den Haag, 2023). Together with other 20 municipalities, it forms the Rotterdam-The Hague metropolitan area with a population of 2,4 million inhabitants across 1200 square kilometers (Gemeente Den Haag, 2023).

3.2 Application of Quantitative and Case Study Approaches

To address all the sub-research questions outlined in section 2.5, this research combines quantitative and case study approaches. It means that it uses quantitative methods such as surveys and statistical analyses (Sardana et al., 2023). In addition, the case study approach allows for an in-depth exploration of the specific context under investigation, for instance, the transport network availability and amenities distribution of the chosen area.

Initially, this research explores the differences between perceived accessibility and spatial accessibility. To achieve this, the study employs both a perceived accessibility investigation through surveys and a spatial accessibility analysis, which involves mapping points of interest within the geographical area under study.

The spatial analysis calculates the Cumulative Opportunity Measure, taking into account various transport modes and their respective transport networks. In the case of Public Transport, the time of the day is also taken into account in this calculation. Additionally, travel time thresholds are established for each transport mode. By conducting both accessibility analyses, the results are compared, and any disparities or mismatches between perceived and spatial accessibility are identified.

Moreover, the study delves into the investigation of mismatches concerning gender, family composition, transport modes, and activity types. This examination aims to understand the factors contributing to these mismatches and gain insights into how different demographic and behavioral aspects impact accessibility perceptions and spatial reality.

Next, the presence of mismatches is examined based on urban groups, aiming to identify the groups most vulnerable to experiencing a mismatch. To accomplish this, Cluster Analysis is employed, an unsupervised machine learning algorithm known for identifying clusters and patterns in unlabeled data (P. Zhang et al., 2022). The primary objective of Cluster Analysis is to determine the most natural grouping within a given dataset, which proves especially valuable in identifying clusters associated with urban vulnerability, based on the specific research context (Garcia-Dias et al., 2020; P. Zhang et al., 2022). Thus, after identifying the urban group's formation, the presence of mismatches in each group is investigated.

Furthermore, this research analyses the influential personal characteristics that impact the occurrence of mismatches. For this purpose, Logistic Regression is utilized, a statistical technique used to model the probability of a discrete outcome based on input variables (Edgar & Manz, 2017). In this analysis, Binary Logistic Regression is chosen, as it is particularly well-suited for analyzing and classifying binary variables (Maalouf, 2011). The mismatches are the dependent variable while the independent variables are the demographic data collected in the surveys, such as age, income, household size, and others. This step aims to gain insights into how individual attributes influence the likelihood of encountering mismatches, providing valuable understanding of the factors at play in the accessibility perceptions and spatial realities of participants.

4. Methodology

This chapter describes the methodology considered for this research. As described in Chapter 3, this project applies a combination of methods to answer the proposed sub-research questions and retrieve results for the main research question. Figure 4.1 presents the Research Framework and its main components which are classified as data sources, data gathering, accessibility calculation, and accessibility analysis.

The literature review serves as a primary data source, providing essential material for crafting the surveys used to collect data on self-reported accessibility. Additionally, OpenStreetMap plays a crucial role as another data source, facilitating the gathering of transport networks and points of interest within the defined geographical scope.

With the data gathered, this research calculates both perceived accessibility and spatial accessibility. By comparing these two sets of data, the analysis focuses on identifying mismatches, vulnerable urban profiles, and the key personal characteristics that influence these disparities. This analytical approach facilitates a thorough exploration of the sub-research inquiries, thereby effectively addressing the overarching research question.

Significantly, while not the primary focus, the calculation of perceived accessibility levels also contributes to comprehending and partially addressing the third sub-research question of this study. This particular inquiry delves into the identification of personal characteristics wielding a significant impact on accessibility levels.

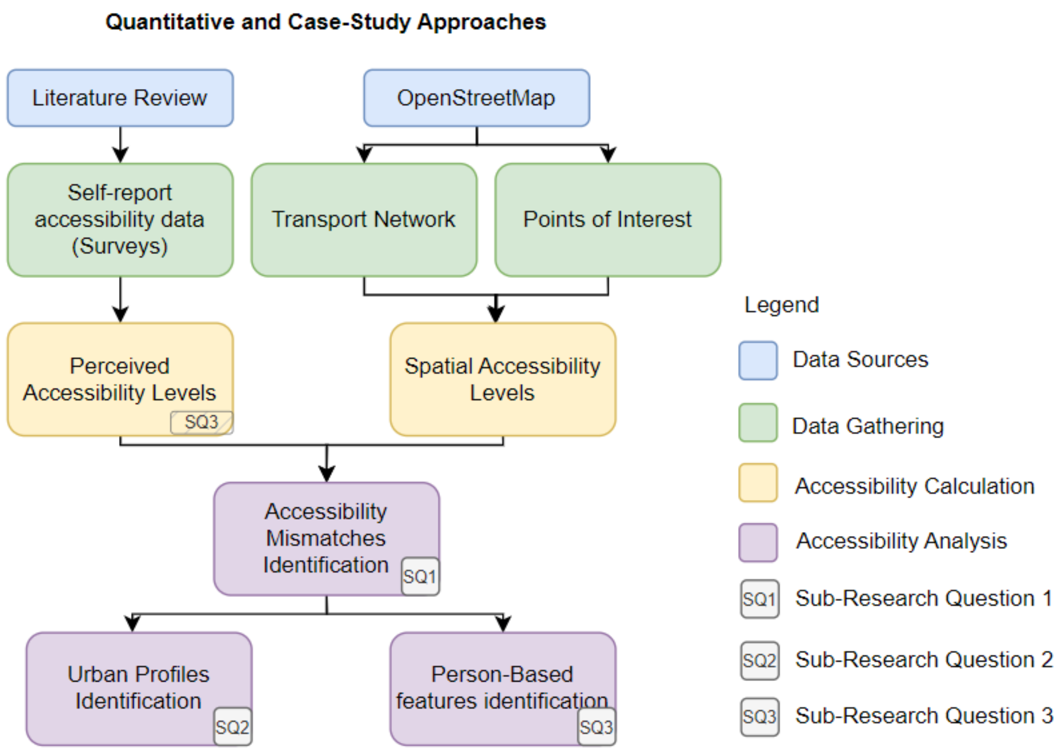


Figure 4.1: Research Framework

The Research Framework employed in this study encompasses the four core components for measuring accessibility, described by Van Wee and Geurs (2004): land use, transport, individual and temporal. The land-use component is incorporated when the amenities distribution is mapped. The transport component is investigated when analyzing the available transport networks in the geographical scope. The individual component is included by incorporating people's perceptions of their accessibility. The temporal aspect is included by investigating public transport availability at specific times of the day. Furthermore, this research defines reasonable travel times as thresholds and analyzes the value of time in choosing a particular transport mode through surveys.

By fully applying Van Wee and Geurs' (2004) framework, which advocates for considering all these components ideally, this study addresses the limitations often encountered in practice, where only a few components are typically included. In focusing on gender and accessibility investigation and applying all these components comprehensively, this research seeks to make a significant scientific contribution to this field of study.

Based on this research framework, the following sections outline the methodologies utilized for calculating Spatial Accessibility, Perceived Accessibility, Accessibility Mismatches Identification, Urban Profiles Identification, and Person-Based Features Identification sequentially.

4.1 Spatial Accessibility Calculation

This research step aims to categorize spatial accessibility levels based on the number of reachable points of interest by a transport mode considering a threshold. In other words, this section describes the procedure for calculating Cumulative opportunity-based measurement and defining accessibility categories. Despite the limitations of this measure, it is proven to give relevant results, and it is vastly used in studies (Ryan & Pereira, 2021). This methodology consists of realizing the Scope definition, Hexagon Grid Application, Threshold Definition, Cumulative Opportunity Measure Calculation, and Accessibility levels Calculation.

4.1.1 Scope Definition

The initial step in calculating spatial accessibility levels involves defining the scope of the analysis, which includes determining the types of points of interest and the geographical area under consideration.

Points of Interest (POIs)

Points of Interest (POIs) can be described into different categories of accessibility: Mobility, Active Living, Entertainment, Food Choices, Community Space, Education, and Health and Well-being (Nicoletti et al., 2022). This study focuses on three types: Food Choices, represented by grocery stores; Education, represented by primary schools or child care and Entertainment, represented by cinemas, nightclubs, bars, pubs, and restaurants. Despite restaurants being part of the Food Choices category, this analysis considers them as points of interest for leisure purposes, as they serve not only as places to satisfy food preferences but also as venues for socializing and enjoying dining experiences. This broader perspective acknowledges the role of restaurants in enhancing individuals' overall leisure and entertainment activities, making them relevant points of interest for this study.

Although these are not explicitly gendered activities, the literature review shows that men and women may shoulder or perceive them differently (Havet et al., 2021; Lo & Houston, 2018). As women shoulder more household and childcare activities, the grocery stores and primary schools / child care are considered. The latter is destinated to participants with children, and the main goal is to investigate the different mobility capabilities to realize these tasks. Finally, this study investigates the mobility capability of traveling to leisure activities at night. This type of point of interest is included since it is a common reason to travel at night, and the time of the day generally impacts the perception of safety, mainly among women (Priya Uteng, 2021; Tiznado-Aitken et al., 2020).

This research uses a combination of Open Street Map (OSM) and the python library osmnx to retrieve the desired Points of Interest. This study uses Python because it is a popular data-driven tool for data and quantitative analysis (Southall, 2021). In addition, OSM offers a free tagging mechanism that includes several number of characteristics describing each landmark (OpenStreetMap Wiki, n.d.). Specifically, this phase uses the osmnx.geometries module. Among several functionalities, it retrieves points of interest from OSM, including their geometries and attribute data, and construct a GeoDataFrame of them (Boeing, 2017). The function used in this research requires a point of origin to retrieve the points of interest, the Euclidian distance from the origin point, and a tag dictionary specifying the point of interest type.

For the most frequently used tags, the community has established particular key and value combinations that serve as unofficial standards (OpenStreetMap Wiki, n.d.). It has two levels of identification: primary features such as amenities, boundaries, airways, and others, and a sub-level with tags for each primary component. According to OSM tags, this study considers the respective tags for each POI group:

- Grocery stores: 'shop' primary group – 'supermarket' and 'greengrocer' tags
- Schools and daycare facilities: 'amenity' primary group – 'school,' 'childcare,' 'kindergarten.'
- Entertainment POIs: 'amenity' primary group – 'nightclub,' 'restaurant,' 'pub,' 'cinema,' and 'theater.'

Despite the efforts to map all points of interest proposed, this approach has a few limitations. Firstly, OSM provides three tags that might represent grocery stores: supermarket; greengrocer, which means a shop focused on selling vegetables and fruits; and convenience, a small local shop carrying a small subset of the items found in a supermarket (OpenStreetMap Wiki, n.d.). The three tags are closely related to grocery stores.

However, a manual check shows that several places named convenience stores are similar to night stores, which sell mainly beverages and snacks. However, the latter category is deliberately omitted from this study for a specific reason. If this research considers fewer amenities than are actually present in a given space, it could lead to instances of false underestimation mismatches. In these situations, users might perceive good accessibility while the spatial analysis suggests limited amenity options. This case is not considered problematic as underestimation mismatches are not the core of this research.

Conversely, if the spatial analysis indicates more amenities than participants actually perceive, it can result in a overestimation mismatch. This occurs when users don't perceive an activity as accessible despite there being numerous amenity options available. As convenience stores typically aren't suitable options for grocery shopping, this research deliberately excludes this category to prevent false overestimation mismatches. The aim here is to ensure the accuracy of the analysis by focusing on relevant amenity types. However, this study recommends further work to analyse the impact of including different amenity tags such as the 'convenience stores' in the results.

The school tags definition faces a similar issue. This research aims to map primary schools and day-care facilities, where adults mostly take children. The tags 'childcare' and 'kindergarten' match the points of interest described in the questionnaire; however, the tag 'school' includes primary, secondary, and high schools. Considering that excluding this tag would also drop all primary schools, the analysis maintains it. However, it is fundamental to be aware of the limitations of this investigation.

The tags related to entertainment activities described in the questionnaire are suitable for this research and did not present any related issues.

Geographical Scope

This research centers around the Rotterdam–The Hague metropolitan area, specifically focusing on six municipalities within the MRDH area: Rotterdam, The Hague ('s-Gravenhage), Delft, Leidschendam-Voorburg, Schiedam, and Rijswijk. The data collected for this study is primarily from these six municipalities. As a result, the spatial analysis is also limited to these areas.

The decision to restrict the spatial analysis to these municipalities is based on two key reasons. Firstly, this approach ensures consistency and enables meaningful comparisons between the perceived and spatial analyses. Secondly, expanding the analysis to include additional municipalities would not yield significant additional insights for this research. It would demand more computational resources without effectively serving the research objectives. The current spatial analysis lacks the inclusion of individual perspectives or person-based data, hence it cannot provide relevant insights to the research goal individually. Since the spatial analysis will be later compared to the perceived analysis, focusing on the surveyed municipalities is a sensible choice, aligning well with the research's main objectives while taking resource constraints into account.

Nonetheless, comparing equally the number of opportunities and transport modes of Rotterdam, the second largest city of the Netherlands, with smaller cities such as Leidschendam-Voorburg may generate inaccurate results. Hence, this research investigates Cumulative Opportunities Accessibility Measurement within each city context.

This research uses a combination of Python and the shape file of The Netherlands (EEA, n.d.) at the municipality level (gemeente in Dutch) and district level (wijk in Dutch) to visualize the geographical scope. First, it filters the five cities mentioned based on their name. Second, this research filters the metropolitan area of Rotterdam, since part of Rotterdam consists of the port or industrial area with few or no residents. Surrounding districts that are residential areas, however, present a considerable distinction from the Rotterdam metropolis were not considered because the comparison may lead to inaccurate results. In brief, the districts considered in Rotterdam are Overschie, Prins Alexander, Noord, Hillegersberg Schiebroek, Kralingen-Crooswijk, Delftshaven, Centrum, Feijenoord, IJsselmonde, Charlois, and Waalhaven. Finally, the areas analyzed in this study are presented in Figure 4.2

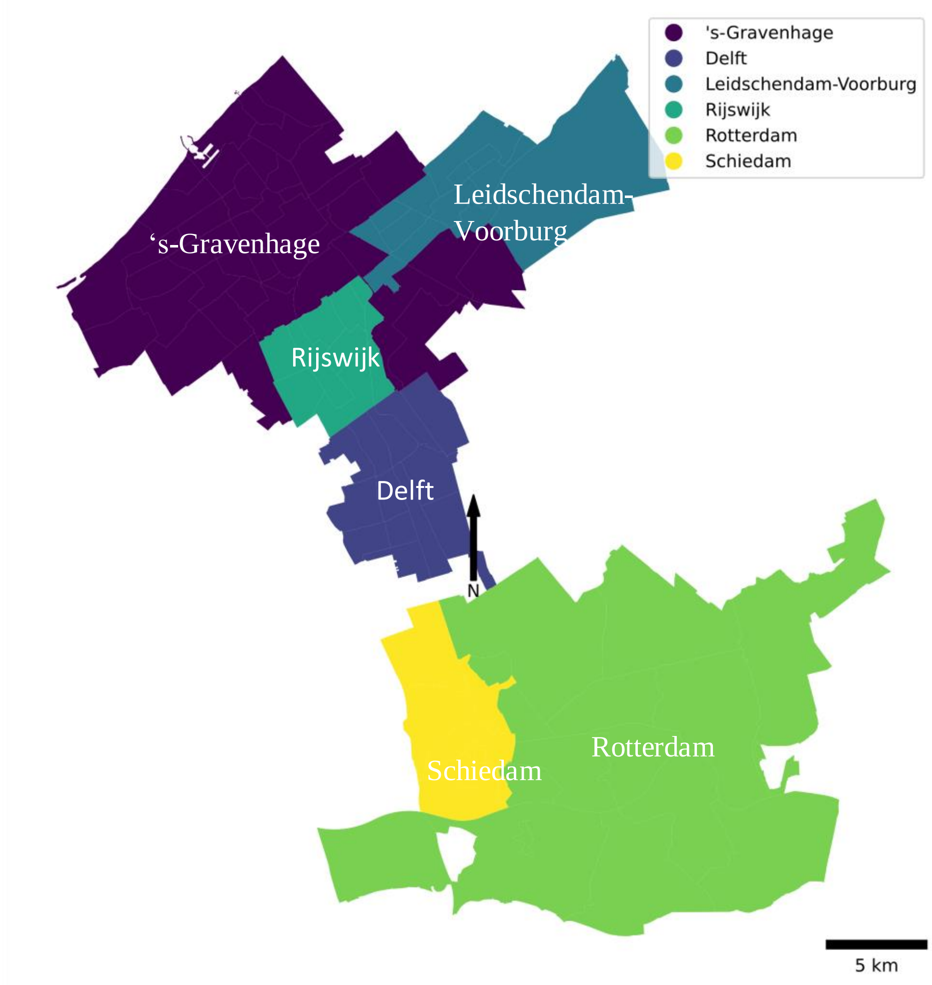


Figure 4.2: Geographical Scope

4.1.2 Modifiable Areal Unit Problem (MAUP) and Hexgrid Application

The goal of this spatial accessibility analysis is to conduct a systematic investigation of the geographical scope. To achieve this, the study divides the area and aggregates data spatially, grouping information into larger areal units instead of examining each individual residential point within the geographical scope. However, it is crucial to be aware that this process of aggregating data into larger units may lead to a phenomenon known as the modifiable areal unit problem (MAUP) (Buzzelli, 2020)

There are two related issues to the MAUP. The ‘scale effect’ concerns the large number of ways of aggregating a set and the influence it causes on the value of statistics derived from spatially aggregated data (Thrift & Kitchin, 2009). In other words, the interpretation of an area and its variables differs according to the scale considered. The second issue is called the ‘zoning effect,’ which considers that the way the geographic boundaries are imposed in grouping the population could introduce bias in the results (Keeler & Emch, 2017). In brief, “Modifiable areal units” of analysis can produce differing analytical results (Thrift & Kitchin, 2009).

Despite the research progress regarding MAUP, a general solution for these issues is still uncovered. The primary action is to be aware of the issue and its impact (Barnes & Forde, 2021). Using the original point data rather than the aggregated ones is one technique to deal with MAUP, although this is typically not possible for privacy-related legal reasons (Su et al., 2011). Additionally, reducing the area units may not totally solve the MAUP but may lessen the likelihood of spatial pattern distortion issues (Su et al., 2011).

Considering the MAUP issues, this research applies a grid, specifically a hexagon grid, to spatially aggregate data. A grid is a pattern of geometrical forms generated using a straightforward mathematical function or formula to divide a surface or territory. A space can be divided into grids to aid with knowledge of the area and placement of things (Apte et al., 2013). There are different grid shapes, such as squares, triangles, and hexagons.

Apte et al. (2013) have proven that hexagon grids are more efficient and accurate than the other grid options. Because hexagons are closest to a polygon with a circular shape that may tessellate to create an equally spaced grid, they lessen sampling bias caused by the edge effects of the grid shape (ArcGIS Pro, n.d.). In addition, it is a common approach used in studies regarding spatial accessibility (Nicoletti et al., 2022; Ryan & Pereira, 2021). Thus, compared to other grid formats, the hexagon grid option reduces the bias caused by the zoning issue.

The fundamental tool for hexagon grid conversion in this research is the library h3py, a Python library of Uber's H3 Hexagonal Hierarchical Geospatial Indexing System in Python. This library is based on H3, a geospatial indexing system that partitions the world into hexagonal cells (H3GEO, n.d.). The H3 specifies a resolution parameter, which defines the size of the hexagons. Table 4.1 describes the average hexagon areas according to each resolution. In addition, the average hexagon length is calculated by the author.

Hexagons with larger areas may hinder different accessibility levels among locations if summarized in one calculation. Moreover, as described previously, smaller areas may reduce spatial bias and mitigate MAUP. On the other hand, smaller sizes might generate several hexagons with the exact accessibility level if the number of amenities does not change within a few meters. This last case may cause an unnecessary computational expensive process.

Table 4.1: Hexagons Resolution of H3 Library

Resolution	Average Hexagon Area (km2)	Average Hexagon Length (km)
0	4357449.416	1295.061026
1	609788.4418	484.4663876
2	86801.7804	182.7840826
3	12393.43466	69.06688948
4	1770.347654	26.10377616
5	252.9038582	9.866242466
6	36.12906216	3.729086015
7	5.16129336	1.409461862
8	0.737327598	0.532726501
9	0.105332513	0.20135169
10	0.015047502	0.076103786

Note: Adapted from (Tables of Cell Statistics Across Resolutions | H3, n.d.)

Based on the process of counting POIs among an area, this research considers that resolution eight (average hexagon length of 533 meters) is the most suitable since resolution seven might hinder accessibility levels, and resolution nine might overcalculate them. Resolution eight in this research context generates hexagons with a length of 557 meters. It means the distance between two hexagon centroids is around 1 kilometer, a reasonable distance between each point.

After transforming the geographical scope into one hexagonal grid, this research identifies the location of each hexagon centroid in each municipality region (*Gemeente* polygons). Figure 4.3 presents the study area divided into hexagons, identified by municipality.

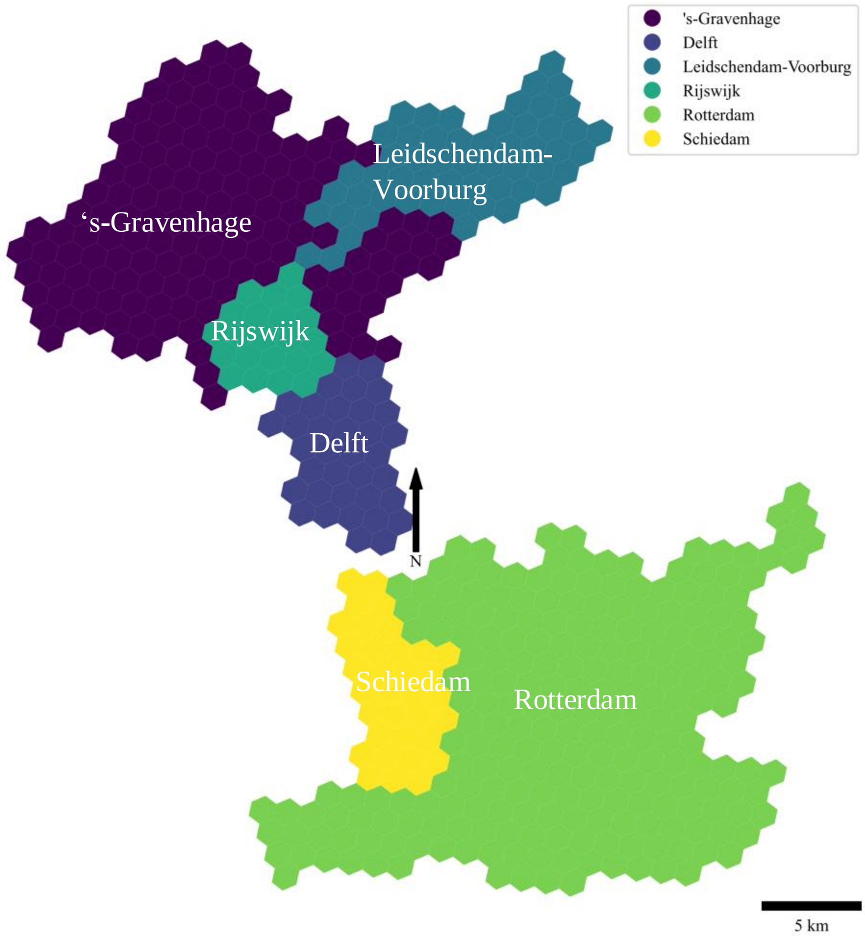


Figure 4.3: Hexagon Grid of Geographical Scope

4.1.3 Cumulative Opportunities Metric Calculation

The calculation of cumulative opportunities for all transport modes follows the same steps, which are described in the subsequent subsections.

Mapping POIs from a geographical location

As described previously in section 4.1.1, the process of mapping POIs uses the osmnx library, specifically the osmnx.geometries.geometries_from_point method. This module requires an origin point, the tags of the desired geometries, and a distance that will be considered from the origin point (Boeing, 2017). The distance is measured in meters and represents the Euclidean distance from the origin point. However, it is important to note that a Euclidean distance of 3km, for instance, may not accurately represent the actual walking distance required to access the same amenity. This initial phase does not consider the transport network, but the subsequent steps calculate the transport network and adjust the reachable POIs accordingly.

Calculate the travel time from the origin to each POI mapped in the previous phase.

This phase relies on r5py, a Python library for rapid realistic routing on multimodal transport networks such as walking, bicycle, car, and public transport (R5py, 2022). This library is used since it is free and allows users to calculate travel time matrices quickly (R5py, 2022). It requires two main files for calculation, a network dataset from OpenStreetMap (OSM) in Protocolbuffer Binary (.pbf) format and a transit schedule dataset in General Transit Feed Specification (GTFS.zip) format, in case of public transport calculations. From here, this research uses the files:

- GTFS data: gifs-nl.zip from OVApi
- OpenStreetMap data in PBF-format: Netherlands-latest.osm.pbf (Geofabrik GmbH, 2018)

These two files generate the transport networks that calculate the travel time between two points. The library r5py uses the module TravelTimeMatrixComputer which requires a transport network, origin point(s), and destination point(s). This step considers the origin point as the hexagon centroid(s) and the destination as the POIs calculated in the first step.

In addition, this r5py module enables the definition of one or a combination of transport modes. The transport modes, car, bicycle, and walk, were defined individually. The public transport travel times were calculated as a combination of public transport and walking.

This combination is selected since this research investigates the accessibility level of participants from their houses. Generally, the access to public transport is combined with walking mode until public transport stops. Other combinations such as cycling or car and public transportation are possible however considered out of the scope of this research.

R5py library also enables the definition of a date and time of departure. Henceforth, this research defines the departure time and day of the week according to the POI type. For instance, in the case of entertainment activities, this phase considers Saturday at 9 pm since it investigates transport mobility at night. On the other hand, when calculating the travel time to schools/daycare, this research considers the departure day as a weekday at 7:40 am, according to the schools’ schedule in The Netherlands (van Mameren, 2023). Ultimately, there isn't a universally fixed time for visiting grocery stores. However, it's generally assumed that a considerable number of people tend to go after their work commitments. Given the usual work schedule, this would typically be during the evenings on weekdays or over the course of the weekend. As a result, this research has chosen a somewhat arbitrary but convenient time frame: a weekday at 6 pm. Table 4.2 summarizes the time departure definition for each activity type.

Table 4.2: Departure Date and Time definition per Activity Type

Activity Type	Departure Date and time
Entertainment Activities	Saturday, 10/06/2023 at 21:00
Schools / Day care	Wednesday, 08/06/2023 at 7:30
Grocery Stores	Wednesday, 08/06/2023 at 18:00

Filter distances within the threshold and count the number of reachable POIs

For each distance calculated in the previous step, this phase checks if it is within the threshold. If it is the case, it counts as a reachable POI. This phase sums the number of reachable POIs for each hexagon grid, per transport mode, per amenity type. The algorithm that represents the complete process is presented in Appendix A.1. The definition of threshold are explained in the following section.

4.1.4 Threshold Definition

As presented in Chapter 2, the cumulative opportunities calculation includes a cost factor. This cost is based on a threshold. The thresholds can be calculated in a normative or positive way (Páez et al., 2012). Normative implementation of accessibility requires only a standard deal or a reasonable supposition of cost, whereas a positive threshold includes person-based behavior. One of this research goals is to compare spatial analysis with person-based analysis. Therefore, normative thresholds are considered for this spatial analysis.

This cumulative opportunities calculation relies on the python library r5py, which calculates travel times between two points by different transport modes (R5py, 2022). For this reason, this phase defines travel duration as the threshold metric. To define an appropriate threshold for each case, this research investigates thresholds based on the average travel time in The Netherlands per transport mode per activity. However, most sources found during this research describe average travel data in kilometers. Thus, this step calculates the average travel time by dividing the travel distance by the average speed of each transport mode.

In situations where specific information about a particular travel purpose is unavailable, general data on average travel behavior per transport mode in the Netherlands is considered. It should be noted that average travel durations may vary across cities. However, this study assumes that the average travel duration in the Netherlands provides a reasonable estimate for comparing spatial and perceived accessibilities. A sensitivity analysis is recommended as future work, considered beyond the scope of this study.

Accordingly, for the walking mode, it considers a general average walking speed of 4.5 km/h and distance of 1.5km in The Netherlands[1], as presented in Table 4.2. Analogously, the average cycling speed of 12.4 km/h is considered for the cycling mode[2]. The average travel distance by car and public transport for each activity type is presented in a national study (CBS, 2022).

The average speed of a car is influenced by the type of road[3]. In this research, it is assumed that trips to grocery stores primarily occur on urban roads, and therefore, the travel time calculation only considers the speed associated with these road types. On the other hand, trips to educational or leisure destinations may involve both urban and motorway road types. As specific information about the average car speed for leisure or education purposes was not found, the average speed is considered to be the speed limit within built-up areas in the Netherlands.

To calculate the travel time for public transport, it is important to consider the different average speeds of various modes such as tram, bus, metro, and train. In this research, the average speed of the tram is used as it is generally the slowest mode within the geographical scope (RET, n.d.). This approach is selected deliberately to estimate the longest travel time threshold. Therefore, this threshold encompasses travel by bus, metro, and train, ensuring a comprehensive analysis of public transport travel durations. Table 4.3 presents the average distance, speed, and calculated duration for each transport mode and activity type.

Table 4.3: Thresholds per each POI type and transport mode

Transport Mode	Grocery Stores			Education Purposes			Leisure Purposes		
	Average Speed (km/h)	Distance (km)	Time (min)	Average Speed (km/h)	Distance (km)	Time (min)	Average Speed (km/h)	Distance (km)	Time (min)
Walking	4.5[a]	1.5[a]	19[g]	4.5[a]	1.5[a]	19[g]	4.5[a]	1.5[a]	19[g]
Cycling	12.4[b]	2[b]	10[g]	12.4[b]	3.29[b]	16[g]	12.4[b]	3.29[b]	16[g]
Car	26[d]	7[c]	17[g]	50[e]	27[c]	33[g]	50[e]	22[c]	27[g]
Public Transport (Tram)	18[f]	7[c]	23[g]	18[f]	10[c]	35[g]	18	9[c]	30[g]

Average Travel values in The Netherlands retrieved from: [a] (de Haas & Hamersma, 2019); [b](Waterstaat, 2017); [c](CBS, 2022); [d](Ligterink, 2016); [e](OlegS, n.d.); [f](RET, n.d.); [g] Calculated by the author

4.1.5 Accessibility Levels Calculation

After computing the number of POIs in each hexagon cell, the next step involves determining the accessibility level for each hexagon based on predefined categories. It is important to emphasize that this research acknowledges the contextual variation in accessibility scales. For example, an area deemed highly accessible in Leidschendam-Voorburg may have a lower number of POIs compared to a highly accessible location in Rotterdam. Consequently, the accessibility levels are defined separately for each municipality. The study examines twelve different scenarios, encompassing three types of POIs and four transport modes. Prior to selecting the most appropriate technique for dataset division, this phase examines the distribution of amenities within each municipality's twelve scenarios.

Figure 4.4 illustrates the distribution of amenities in The Hague. Each graph in this figure represents the number of reachable amenities, which means the number of amenities accessible based on a specific transport type threshold. The x-axis represents this value, while the y-axis shows the frequency of hexagons with each corresponding number of amenities. The distribution of amenities in other cities can be found in Appendix A.2.

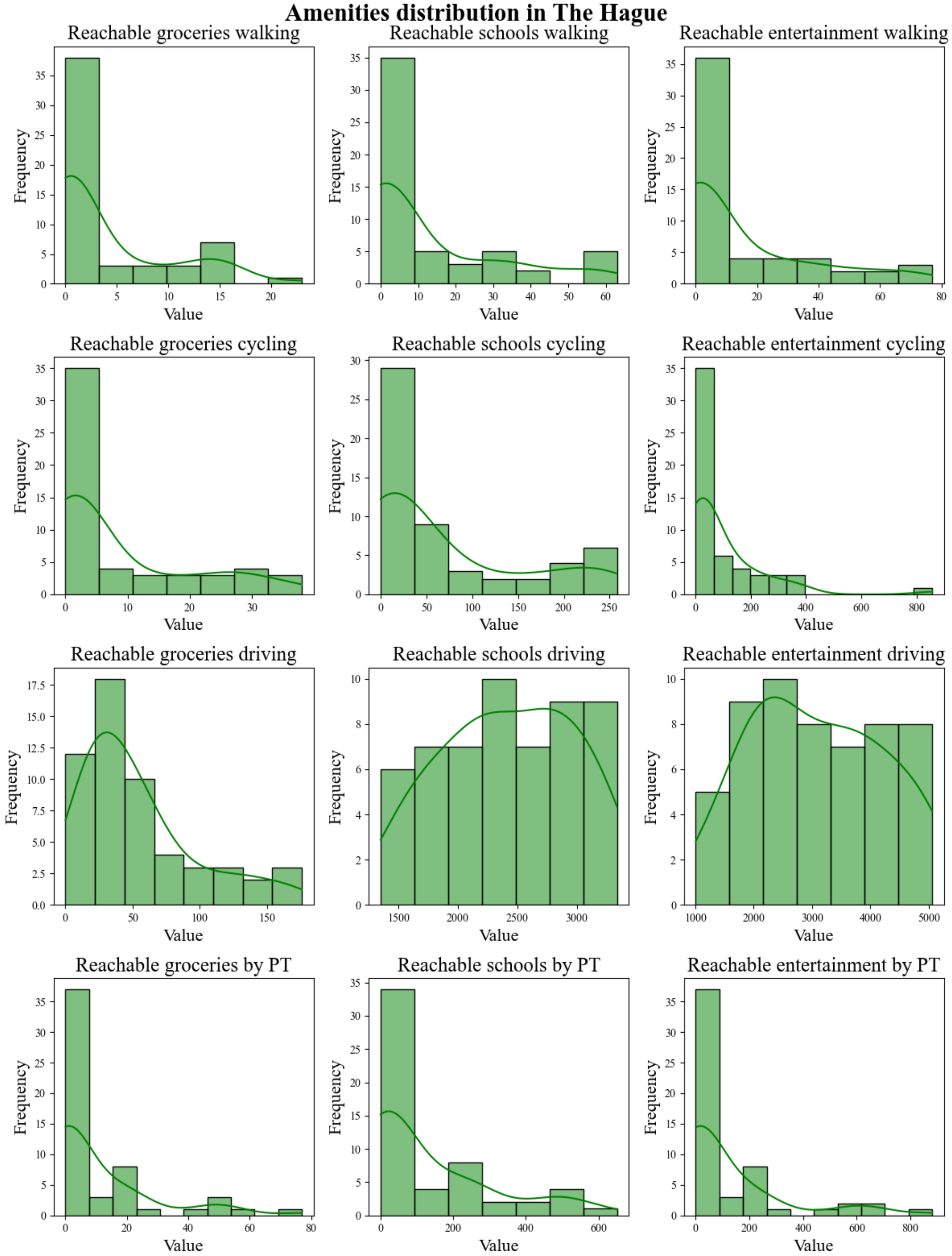


Figure 4.4: Amenities' distribution in The Hague

In Figure 4.4, it is evident that the driving transport mode's distribution for The Hague closely resembles a normal distribution. In contrast, the distributions for walking and cycling options are shifted to the left, indicating that most regions (hexagons) have a relatively low number of amenities accessible by walking or cycling. It is also identified that the data presents outliers of reachable entertainment amenities by walking. It might represent a higher concentration of amenities in city centers.

In addition, this investigation identifies outliers related to shallow values. If the centroid of the hexagon is located in a non-urban area such as a river or a park, it reduces the number of reachable amenities. The r5py algorithm, in some cases, cannot identify a transport network to connect this point.

This phase utilizes three potential techniques to classify the dataset into four accessibility levels. One approach involves using the mean and standard deviation, which measures the extent to which the data is spread around the average (Wilcox, 2012). In this scenario, the first breakpoint for categorization is obtained by subtracting the standard deviation from the mean, the second breakpoint is set at the mean itself, and the third breakpoint is determined by adding the standard deviation to the mean. However, since the mean and standard deviation are highly influenced by outliers, they are not appropriate metrics for this case (Berman, 2016).

Another possible alternative is the division of data based on quartiles. It is a common approach to analyzing measurements of a continuous variable and group subjects into several groups. To create four equal groups, breakpoints are also required to split the data such that 25% of the observations are in each group. The cut-off points are called quartiles, and there are three of them (the middle one also being called the median) (Altman & Bland, 1994). However, this approach is not recommended when the dataset is not generally distributed because it can separate locations with very similar rates and group places with very different rates (Axis Maps, n.d.)

Based on that, this research considers using natural breaks or Fisher Jenks to classify the data. This approach is a standard method for dividing a dataset into a certain number of homogenous classes. The algorithm is commonly used in geographic information systems (GIS) applications (North, 2009). It is an optimal classification scheme that finds class breaks that will minimize within-class variance and maximize between-class differences (Axis Maps, n.d.).

When applying this method, it is noticed that it can be sensitive to outliers. To avoid a significant influence of break definitions based on outlier values, this research realizes the process of Winsorization, which refers to setting severe outliers to a given percentile of the data. In the case of a 90% Winsorization, all observations that fall below the 5th percentile are equal to the value at the 5th percentile, and all observations that fall above the 95th are equal to the value at the 95th percentile. Data is Winsorized when extreme importance in a dataset is converted to less extreme ones (Zach, 2021).

As described previously, outliers were identified in significantly lower and higher amounts of amenities. Hence, this research applies a 90% Winsorization process in the data. After this procedure, this step applies the natural break algorithm to the transformed data. This phase aims to identify four different categories. Therefore, the natural break algorithm identifies four breaks. Figure 4.5 represents the distribution of amenities in The Hague and the breaks identified by the algorithm. This procedure is done for each one of the twelve cases for each municipality studied. The breakpoint for each activity type, per transport mode, for each city is described in Appendix A.3.

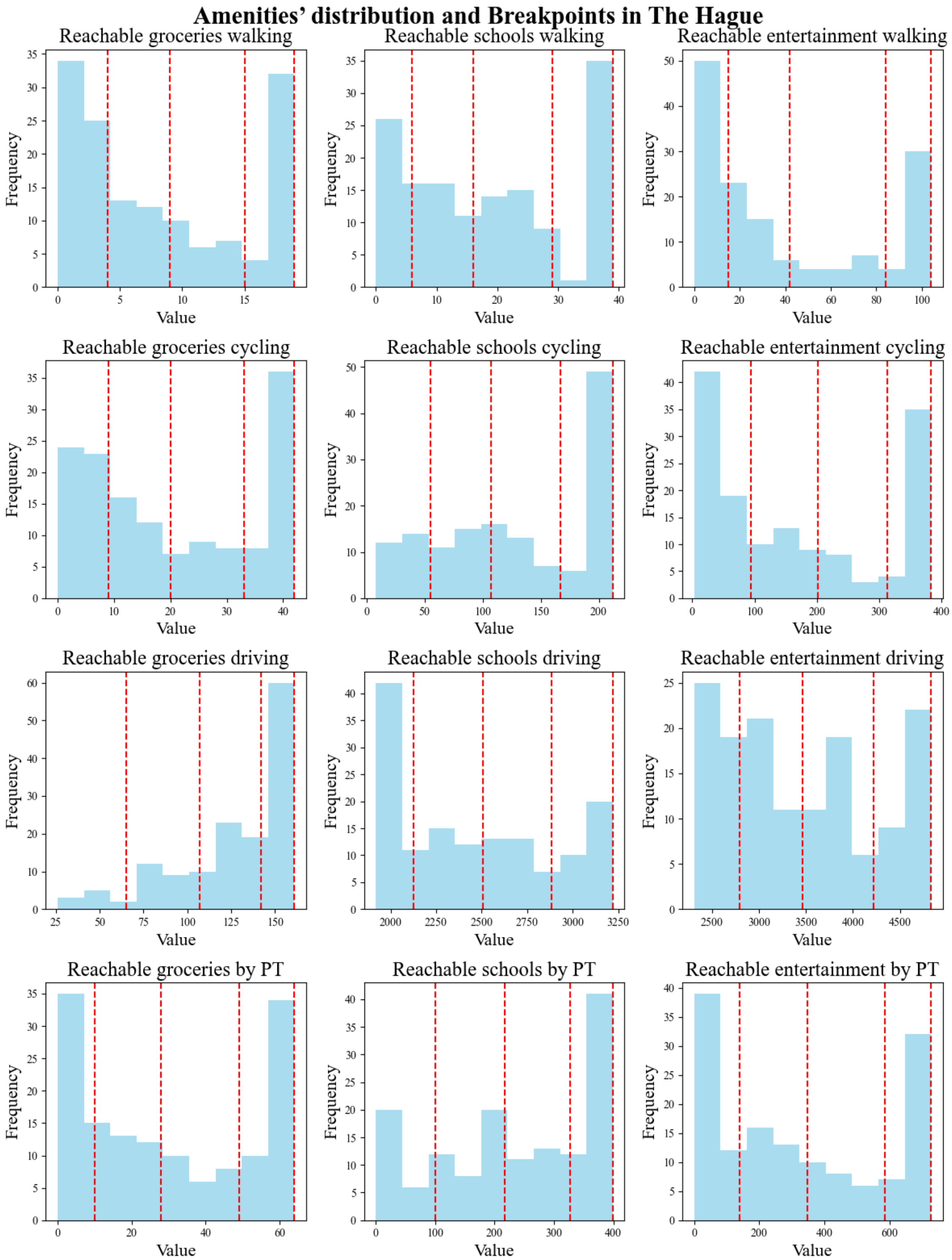


Figure 4.5: Amenities' distribution and BreakPoints in The Hague

Therefore, four main accessibility categories are defined and labeled:

- Limited Accessibility: This category represents areas with significant limitations in terms of access to Points of Interest. It represents the area with the amount of POI(s) below the first breakpoint identified in the city.

- Moderate Accessibility: This category represents areas with a few limitations in terms of access to Points of Interest. It represents the area with the amount of POI(s) between the first and second breakpoints of the city.
- Good Accessibility: This category represents areas with moderate to good access to Points of Interest. It represents the area with the amount of POI(s) between the second and third breakpoints of the city.
- Very Good accessibility: This category represents areas with good to excellent access to Points of Interest. It represents the area with an amount of POI(s) higher than the third breakpoint of the city.

One drawback of the natural breaks approach is each dataset generates a unique classification solution. Thus, it is less interpretable in the case of comparison across maps (Axis Maps, n.d.). Therefore, it is critical reminding that each city presents the four accessibility levels according to its context. Because this research aims to compare spatial and perceived data instead of analyzing the accessibility levels among spatial analysis, this research considers that is not a critical drawback for using this categorization methodology.

4.2 Perceived Accessibility Calculation

Additionally to spatial accessibility data, this research calculates perceived accessibility levels. It applies a survey to gather and analyze perceived data regarding accessibility. This research considers empirical data gathering, such as surveys, due to the highest potential to obtain each individual's perception (The Doctoral Journey, 2014). This phase methodology can be summarized into the survey design and survey implementation steps.

4.2.1 Survey Design

The questionnaire was realized in collaboration with fellow Master's student Iris Roeleven, who has a similar master's thesis theme and the same graduation committee. However, this chapter focuses mainly on the questions and design process related to this research.

The first step of designing the survey is defining the essential information that should be included in the questionnaire. The survey aims to collect data about individual aspects that might influence people's accessibility perception. Thus, several components were included in the questionnaire.

The individual characteristics that are included in the survey and have a strong correlation with accessibility in transport are age, gender, household composition (presence of a partner and number of children), income, and education level (Havet et al., 2021), and culture (Loukaitou-Sideris, 2020). For the latter, this survey asks about the country of origin. Moreover, considering that the individual and household income might represent different financial conditions according to the household composition, both cases are asked in this questionnaire. In addition, car access has been identified as a critical component that impacts accessibility perception (Havet et al., 2021). Thus, the questionnaire asks about driver's license possession (competence resources) and car access (material resources).

Also, the social context can impact accessibility perception (Ryan & Pereira, 2021). For instance, the participants' travel behavior might change if they know someone that could give them car rides. Thus this survey asks the participants about the presence of social support related to transport (social resources). Furthermore, the literature review considers safety as one of the main aspects that differ between mobility behavior of men and women (Priya Uteng, 2021; Tiznado-Aitken et al., 2020). Hence, this survey includes questions regarding safety perception for different transport modes at night, which may potentialize the unsafety perception.

Additionally, Ryan et al. (2019) defend that people's mobility capability facilitates the capability to carry out activities. Therefore, mobility capability is highly related to accessibility perception levels. Thus, this questionnaire includes questions about the perceived capability of the participants to take each transport mode (car, cycling, walking, or public transport) to reach specific key activities (shopping, education, and leisure) in their daily lives. However, this research includes convenience aspects instead of focusing mainly on capability. These questions are critical to compare the perceived and spatial accessibility. From the spatial accessibility analysis, a participant can reach the amenities type by a transport mode if a reasonable number of amenities are spatially available within a threshold. On the other hand, from the perceived accessibility analysis, the participant can reach the amenities type by a transport mode if they are not perceived as inconvenient or not possible. The questionnaire requires the participants' postal codes to compare both cases considering the same locations.

Differently from Ryan et al. (2019) research that categorizes mobility capability into 'possible' and 'not possible,' this study provides a higher amount of categories, assuming that the level of ease to use a transport mode to reach an activity type is more complex than a binary definition. Thus, participants must categorize their use of transport modes according to four labels: convenient, neutral, inconvenient, and not possible. The categories are described as follows:

- Convenient: I find this transport mode easy to use, and it fits well my personal needs;
- Neutral: I find this transport mode acceptable to use;
- Inconvenient: I have a significant restriction(s) to use this transport mode.
- Not possible: I cannot use this transport mode (ex: driving a car without access to a vehicle).

Analogous to the Spatial Accessibility Investigation, this study investigates the mobility capability for three main types of points of interest: grocery stores, primary schools or child care, and places for leisure activities.

Finally, the survey includes questions about travel behavior to understand the audience's preferences. Thus, questions about the primary transport mode used for each POI category are requested. In addition, this questionary asks the participants to rank in order of importance the aspects they value the most when considering a transport mode. This question is included because it may contribute to understanding the main elements considered when classifying a transport mode as convenient, neutral, inconvenient, or not possible.

The survey components, including the link between the perceived accessibility and spatial accessibility analysis, are described in Figure 4.6.

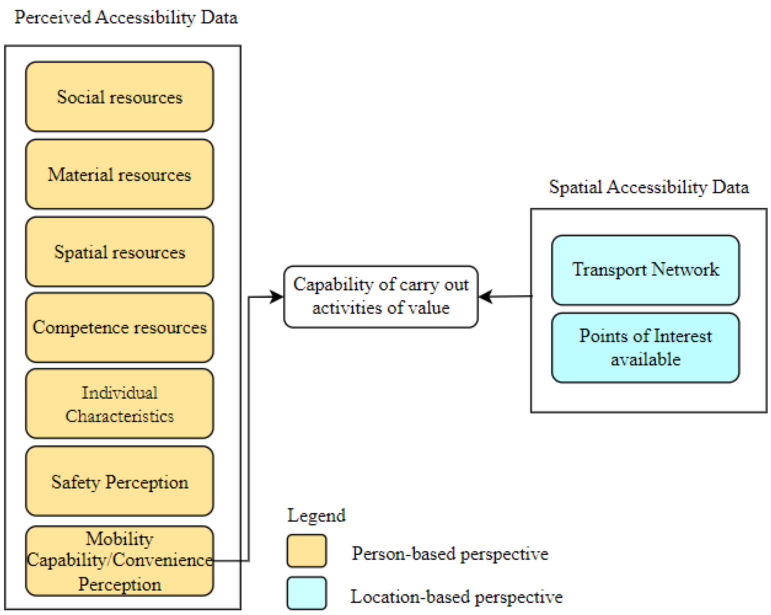


Figure 4.6: Perceived Accessibility Components and connection with Spatial Accessibility Data

Apart from the postal code, all questions are presented in a multiple-choice form. This research relies either on surveys from topic-related studies or the feedback provided by participants during a survey trial to phrase questions and options in the most precise and suitable way possible.

In this survey, the individual characteristics options are presented in ranges. It not only facilitates the data interpretation but also makes the personal characteristics broader, preserving people's privacy and risks of identification. This survey has been analyzed and approved by the Human Research Ethics Committee of TU Delft to guarantee the ethical protection of the participants. The Ethics Application and Forms are presented in Appendix B.1-4.

Table 4.4 describes the components of the perceived data, the respective survey question, and the options. This research includes two questions that depend on previous answers. The question ‘What was approximately the net income of you and your partner together in the last year?’ only appears to participants that previously answered ‘Yes’ to the question ‘Are you living together with a partner?’. In addition, the inquiries related to school and daycare activities, ‘What is the main transport mode you use to go to school, daycare, or a similar place on a daily basis?’ and ‘Please categorize the use of transport modes for going to school, day-care or similar establishments from your house’ are only visible for participants that did not answer ‘No’ or ‘Yes, one child or more older than 12 years old’ to the question ‘Do you have children?’. This research assumes that parents take their children to school or daycare if they have children aged from 0 to 12.

Table 4.4: Description of the questionnaire's components, questions, and options

Component of Perceived Data Gathering	Survey Question	Options
Social resources	If I cannot travel somewhere (important) myself, I think someone in my network (e.g. a friend, a family member) would be available to help me.	Strongly Disagree Disagree Neither disagree nor agree Agree Strongly Agree
Material resources	To what extent do you have access to a car?	I do not have access to a car I can sometimes use a car I sometimes can, and sometimes cannot make use of a car I can usually make use of a car I can always make use of a/my car
Spatial resources	What is your postal code?	Open Question
Competence resources	Do you have a driver's license for a car?	Yes No
Safety resources	How safe do you feel while travelling with the following transport modes during the night (after dark)?	Very Unsafe Unsafe Neither unsafe nor safe Safe Very Safe
Individual characteristics	What gender do you identify as?	Male Female Other
	What is your age?	Between 18 - 25 26-35 36-45 46-55 56-65 66-75 Older than 75
	What is de highest education level you have completed?	Level before middle school Middle school MBO (Secondary vocational education) HBO (higher professional education) bachelor (WO (research-oriented higher education) bachelor HBO (higher professional education) master WO (research-oriented higher education) master PHD

	What is your country of origin?	The Netherlands Another European country Africa North America Central America Caribbean South America Oceania South Asia Central Asia South Eastern Asia East Asia Western Asia
	Are you living together with a partner?	Yes No
	What was approximately your net income in the last year?	Less than €22.000 €22.000 - €43.500 €43.500 - €65.500 €65.500 - €87.500 €87.500 - €109.000 €109.000 - €131.000 More than €131.000 Prefer not to say
	What was approximately the net income of you and your partner together in the last year?	Less than €22.000 €22.000 - €43.500 €43.500 - €87.000 €87.000 - €131.000 €131.000 - €175.000 €175.000 - €218.000 More than €218.000 Prefer not to say
	Do you have children?	No Yes, one child younger than 5 years old Yes, two or more children younger than 5 years old Yes, one child or more between 5 and 12 years old Yes, one child or more older than 12 years old
Mobility Behaviour	What is the main transport mode you use to go to school, daycare, or a similar place on a daily basis?	Car Bus, Tram, Metro or Train Cycling Walking

Mobility Behaviour	What is the main transport mode you use to grocery stores on a daily basis?	Car Bus, Tram or Metro Cycling Walking
Mobility Behaviour	What is the main transport mode you use to go to leisure activities in the evening or at night? It can be restaurants, bars, nightclubs, cinema or similar places.	Car Bus, Tram, Metro or Train Cycling Walking
Mobility Capability	Please categorize the use of transport modes for going to grocery stores from your house.	Convenient Neutral Inconvenient Not Possible
Mobility Capability	Please categorize the use of transport modes for going to school, day-care or similar establishments from your house.	Convenient Neutral Inconvenient Not Possible
Mobility Capability	Please categorize the use of transport modes for going to entertainment activity at night from your house.	Convenient Neutral Inconvenient Not Possible
Mobility critical aspects	Please rank the aspects you value when considering a transport mode convenient or not. Here 1 means the most important and 5 means least important.	Time Safety Comfort Money Sustainability

Moreover, two strategies are considered in the survey design phase to reach a higher number of participants. One of them is a lottery offer of 40 euros for the participants that complete the survey. In this case, participants were forwarded to a new page after finishing the main survey. On this page, participants could fill out their emails if they were willing to participate in the lottery. The independency between the primary survey and the email collection for the lottery is made so the survey answers are not connected to individual emails. Thus, it ensures the participants' data privacy and anonymity.

The second strategy is to provide the questionnaire in five different languages: Dutch, English, Portuguese, French, and Turkish. The questionnaire's authors consider that the higher the number of languages, the higher the chances of a better experience among participants. Based on the research team network, a significant fraction of the participants are not from The Netherlands.

Dutch and English were mandatory languages for this questionnaire since the target group is residents of the Netherlands. The international team involved in this project translated the survey into Portuguese and French. An external collaborator realizes the Turkish translation. Most foreign-born residents were born in Turkey (CBS, n.d.). In Rotterdam, the largest minority groups originate from Surinam (8.7%) and Turkey (7.8%) (COE, n.d.). Thus, it may facilitate the answer to the questionnaire for this significant fraction of the population.

4.2.2 Survey Implementation

The questionnaire was developed and administered using the Qualtrics Platform, a web-based survey tool that offers a range of internal tools and multiple online distribution methods, such as anonymous links, personal links, and QR codes (Qualtrics, 2022). The survey was disseminated through various channels, including the authors' and supervisors' social network posts and interconnected networks. The research team actively shared the survey on platforms such as LinkedIn, Facebook, Whatsapp, and Reddit. Additionally, printed flyers were distributed in key locations, including Rotterdam, Delft, The Hague, and Leidschendam-Voorburg. The association, Vital Cities and Citizens (VCC) from Erasmus University of Rotterdam, also provided support by promoting the survey through their social media channels.

As it takes less time to recruit participants and requires fewer staff members, internet-based research is becoming a more valuable tool for data collecting (Griffin et al., 2022). On the other hand, it also became a weapon for anyone. In online surveys, the term "bot" refers to a script or program designed to repeatedly fill out survey fields with fictitious data and submit the survey, among many reasons, to get the promised payment frequently. Despite this frequency, the presence of Bots in TU Delft research questionnaires is still a new matter.

However, several suspicious activities explain the identification of bots' answers in this survey. In the first two days of survey dissemination, there was an unusual amount of replies (400 responses in a few hours), quick survey replies, several answers computed at the same start time, and inconsistent answers such as postal code different than Dutch format. Most postal codes from this time range presented a North American format, whereas the target group was mainly residents in The Netherlands. Furthermore, there were around 500 fewer participants in the primary survey than the number of email addresses registered in the lottery form. Technically participants would only have access to the lottery form after completing the primary survey, indicating that the lottery might be a reason for the bott's answers. Thus, this project implemented several strategies based on previous literature (Hallberg, 2022; Storozuk et al., 2020) to detect the presence of bot answers. The strategies implemented are described in detail in Appendix B.5.

After cleaning the bot's answer, this phase also considers other criteria to keep relevant answers to this research. First, it maintains only participants that answered valid postal codes within the geographical scope. Postal codes from different regions, empty postal codes, postal codes with the wrong format, or postal codes with the correct format but that do not exist is the Netherlands were dropped from this research. Consequently, the participants were not included in the analysis.

In addition, the mobility capability questions are the link between the perceived and spatial accessibility comparison as presented early in Figure 4.6: Perceived Accessibility Components and connection with Spatial Accessibility Data. Therefore, the analysis did not consider participants who left empty answers to these questions.

4.3 Accessibility Mismatches Identification

This phase describes the methodology for investigating how accessibility calculated by spatial-based perspective differs from person-based accessibility perception. Thus, it compares the spatial accessibility results with perceived accessibility results.

This methodology compares the four spatial accessibility levels (limited accessibility, moderate accessibility, good accessibility, and very good accessibility) described in section 4.1.5 with the four perceived mobility capability categories (not possible, inconvenient, neutral, and convenient) described in section 4.2.1. Hence, there are sixteen possible combinations.

This research considers that if any of the two lower levels of spatial accessibility (limited or moderate) combines if any of the two lower levels of perceived accessibility (not possible or inconvenient), there are no mismatches. The same approach is considered if any of the two higher levels of accessibility of each method combines.

In contrast, if the two lower levels of spatial accessibility combine with the two higher levels of perceived accessibility, it is considered an underestimation mismatch. It means that the analyst underestimates the participant's capability to reach an amenity type by a transport mode. On the other hand, if the two higher levels of spatial accessibility combine with the two lower levels of perceived accessibility, this research defines it as an overestimation mismatch. It means that the spatial analysis overestimates the resident's capability to reach amenities by transport mode. Table 4.5 describes these categories.

Table 4.5: Mismatch types

Spatial Accessibility level	Perceived Accessibility level	Mismatch Type
Very Good or Good	Convenient or Neutral	No Mismatch
Very Good or Good	Inconvenient or Not Possible	Overestimation Mismatch
Moderate or Limited	Convenient or Neutral	Underestimation Mismatch
Moderate or Limited	Inconvenient or Not Possible	No Mismatch

This research investigates the Overestimation Mismatch type explicitly. It assumes that overestimation mismatches may contribute to social inequities, as certain groups may face heightened challenges in accessing amenities despite the seemingly favorable spatial accessibility levels. On the other hand, Underestimation Mismatches may result in an underutilization of potential opportunities. However, it does not indicate possible barriers to accessing amenities.

Therefore, this research further investigates the overestimation mismatch and its intrinsic levels. Within this mismatch type, there are four possible combinations. In the case of Very Good accessibility level combined with Not Possible perceived accessibility, it is categorized as a Strong Mismatch since extreme values are combined. In the case of a Very Good accessibility level combined with Inconvenient perceived accessibility or Good accessibility level combined with Not possible perceived accessibility, it is classified as a Moderate Mismatch. Finally, the combination of Good Accessibility with the Inconvenient category is classified as a Slight Mismatch. Table 4.6 summarizes these Overestimation Mismatch types.

Table 4.6: Mismatch types within Overestimation Mismatch category

Spatial Accessibility level	Perceived Accessibility level	Overestimation Mismatch Type
Very Good	Not Possible	Strong Mismatch
Very Good	Inconvenient	Moderate Mismatch
Good	Not Possible	Moderate Mismatch
Good	Inconvenient	Slight Mismatch

In this phase, the aim is to establish a correspondence between the participant's residential location and the corresponding hexagonal area. This enables the comparison between spatial and perceived analyses. The postal code provided in the survey is utilized to establish this link. The Geopy Python library is employed to extract latitude and longitude coordinates from the postal code. Subsequently, the Shapely library is used to generate a geographical point based on these coordinates. The same library allows for the verification of whether the generated point falls within any of the hexagons within the hexagonal grid. When a participant's point falls within a hexagon, the accessibility classification obtained from the spatial analysis is attributed to that participant. This process ensures that each participant is assigned the appropriate spatial accessibility classification based on their residential location.

4.4 Urban Profiles Identification

This current section outlines the cluster analysis methodology employed to identify urban groups and investigate their potential relationship with one or multiple mismatches. This cluster analysis incorporates demographic data collected from the surveys along with the mismatches identified through the Mismatch Identification phase. By analyzing the key characteristics of each cluster, it becomes possible to assign appropriate labels to the clusters, facilitating their interpretation and understanding.

All data in this research is categorical. Therefore, a categorical cluster analysis is conducted. All cluster analysis types generally entail several fundamental activities—first, feature selection, where the same number of features defines each observation. Second, the analyst chooses similarity metrics, a mathematical function that describes the similarity between observations in the dataset. The third step is the application of the clustering algorithm, where the dataset is grouped based on similarities. Finally, the results are analyzed in a cluster validation step, assessing the methods' performance and the produced groups' accuracy (Garcia-Dias et al., 2020). This section describes the cluster analysis methodology in the same steps, as presented in Figure 4.7. The final step consists of evaluating the model's performance. If it falls short of satisfactory results, it is recommended revisiting previous steps to improve it.

Urban Profiles Identification Approach Diagram

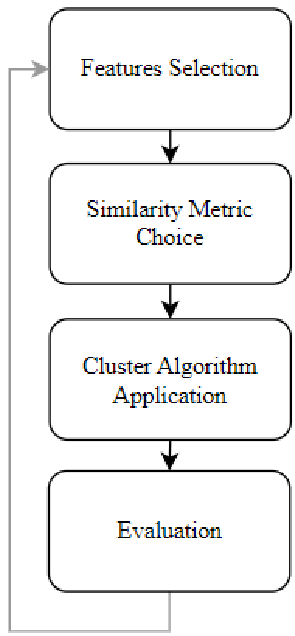


Figure 4.7: Methodology Diagram of Urban Profiles Identification

4.4.1 Feature selection

The initial stage of this analysis involves filtering the key variables to be incorporated in the Clustering Analysis. Considering the investigation of urban profiles and their relationship with mismatch events, this study incorporates demographic data obtained from the survey, coupled with the mismatches identified in the Mismatches Identification section. The analysis concentrates on the most significant mismatches identified in the Mismatches Identification phase, specifically those that exhibit a higher proportion of occurrences within the sample.

In this phase, the selected features are examined to identify and address any missing data. Due to the complexity of replacing missing participants' perceptions within the scope of this study, the decision is made to exclude these missing data points from the analysis. This approach is taken to ensure the integrity and validity of the findings, maintaining the reliability of the results obtained from the available data.

4.4.2 Similarity Metric Choice

This phase involves identifying the most suitable algorithm for this research to identify clusters. The K-means is a popular algorithm in the literature for cluster analysis due to its efficiency and simplicity (Garcia-Dias et al., 2020). It requires as input a matrix of M points (observations) in N dimensions (variables) and a matrix of K initial cluster centers in N dimensions (Hartigan & Wong, 1979). Its procedure divides data into k classes, where each data belongs to the nearest cluster center. During the clustering calculation, the mean of points in the cluster is used as the cluster center for the Euclidean distance calculation and is gradually optimized with iterations. In other words, it minimizes the pairwise deviations of points in the same cluster (Hu et al., 2023).

The simplicity and low computational complexity have given the K-means clustering algorithm wide acceptance in many domains for solving clustering problems (Ikotun et al., 2023). However, K-means requires numerical data, which makes it harder from being used to cluster real-world data containing categorical values (Huang, 1998). Thus, this paper uses the k-modes algorithm, an extension of the k-means algorithm created for categorical domains developed by Huang (1998).

The k-modes approach substitutes the means of clusters with modes. It employs a frequency-based strategy to update modes in the clustering process to minimize the clustering cost function to deal with category items (Huang, 1998). The main difference between k-modes compared to k-means is that the distance between data points X and Y is the number of different observations. It is based on a simple dissimilarity measure according to Equation 4.1:

$$d(X,Y) = \sum_{i=1}^{n} \delta(x_i, y_i) \tag{4.1}$$

where,

$$\delta(x_i, y_i) = \begin{cases} 0, x_i = y_i \\ 1, x_i \neq y_i \end{cases}$$

4.4.3 Cluster Algorithm Application

Before applying the K-modes, the optimal number of K clusters must be identified. This research uses the Elbow and Silhouette Methods for this investigation. The Elbow Method consists of calculating the Within-Cluster-Sum of Squared Error (WSS) for different values of k. The ideal k value is usually identified from a graph (K numbers x WSS) by viewing the elbow location. It means the k location for which WSS becomes first starts to diminish (Mahendru, 2019; Syakur et al., 2018).

However, the elbow method is limited, considering that the discriminant of the number of clusters depends on the manual identification of the elbow points on the visualization curve. This procedure becomes harder when the plotted curve is fairly smooth (Shi et al., 2021). For this reason, this study also considers the Silhouette Method. This approach reveals which items are clearly within their cluster and which ones are just in between clusters by using a silhouette to symbolize each cluster. The data arrangement and the clusters' relative quality can be seen by integrating the silhouettes into a single plot and displaying the full clustering. The average silhouette width may be utilized to choose the right number of clusters (Rousseeuw, 1987). Equation 4.2 calculates the Silhouette score:

$$S\ score = \frac{b - \mathrm{a}}{\max(a, b)} \tag{4.2}$$

Where,

a = cluster width (average intra-cluster distance)

b = average inter-cluster distance.

The optimal k number is the one that presents the higher silhouette score (Bhardwaj, 2020). In case of distinct results between the Elbow and Silhouette method, the latter is prioritized. As described previously, it is less subjective to the analyst interpretation.

4.4.4 Evaluation

There are several evaluation metrics for calculating the degree of accuracy of cluster analysis by comparing predicted values to the cluster labels. Still, unsupervised learning has no ground truth, and it is generally challenging to assess this degree of accuracy (Garcia-Dias et al., 2020; Wong, 2022). Naturally, a satisfying clustering algorithm produces clusters with slight intra-cluster variance and substantial inter-cluster variance (Wong, 2022).

The average silhouette width is used to select an 'appropriate' number of clusters and provides an evaluation of clustering validity (Rousseeuw, 1987). Therefore, this method described in section 4.4.2, it is also the method used to evaluate the performance of this analysis. The silhouette score varies from -1 to 1. The closest the score is to its maximum implies that the intra-cluster distance is much smaller than the smallest inter-cluster distance. It means that the observation is well-clustered, and there is little to no doubt that it has been assigned to an appropriate cluster. When the score is close to 0, it means that the inter and intra-cluster distances are similar, and it is unclear whether the observation should be assigned to the actual cluster or the second-best cluster choice. Thus, it is considered an intermediate case. Finally, the closest score is -1, which means the inter-cluster distance is larger than the intra-cluster distance. Consequently, assigning the observation to the second-best cluster choice would be more natural, and the object might be misclassified (Rousseeuw, 1987).

4.5 Person-Based features identification

Instead of examining general patterns in the formation of urban profiles, this final phase of the methodology employs Binary Logistic Regression to analyze the influence of specific features on the occurrence of mismatches. The objective is to investigate the person-based variables that have the greatest impact on the perception of (women's) accessibility. The methodology for this phase can be summarized into six steps: scope definition, primary data filtering, feature encoding, secondary data filtering, logistic regression modeling, and evaluation metrics. Figure 4.8 illustrates the sequential steps in a diagram. Similar to Cluster Analysis, the final step involves evaluating the model's performance. In the event of unsatisfactory results, revisiting previous steps is advised for improvement. The scope definition and primary data filtering remain unchanged as they align with the research goals. Consequently, features encoding also remains intact. Therefore, if results are insufficient, the focus shifts to reanalyzing the secondary data filtering step.

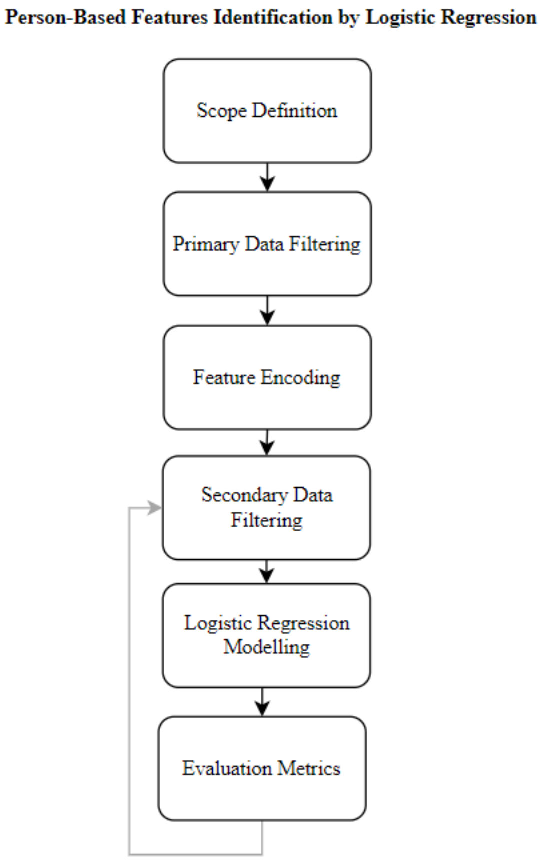


Figure 4.8: Methodology Diagram of Person-Based Features Identification

4.5.1 Scope definition

This section focuses on mismatches among women. It aims to identify the most relevant characteristics that influence the mismatches among women. Even though section 4.3 describes three different mismatch types: strong, moderate, and slight, this analysis focuses on modeling the mismatch occurrence in general. Thus, this section's three mismatch types are generalized to 'mismatches.' This decision is due to some reasons.

First, there may not be enough observations for each mismatch category (strong, moderate, or slight) to create precise logistic regression models on their own. When the sample size for a particular mismatch type is minimal, the logistic regression model's estimations may be unstable and inaccurate. Lower events per variable values during the building of the prediction model have frequently been linked to worse predictive accuracy after validation (van Smeden et al., 2019). Thus, the study gains by merging the mismatch types, enhancing the stability of the model results.

Furthermore, this section prioritizes the mismatches with higher occurrence. Otherwise, the mismatch variable might present significant unbalanced data (i.e., a significantly larger proportion of zeros than ones) (Salas-Eljatib et al., 2018). Some standard classifiers have been observed to perform poorly in machine learning tasks due to disparities in prior class probabilities (Japkowicz & Stephen, 2002).

Finally, the author's main objective is to highlight the features that most influence the likelihood of mismatches among women. The literature review described in Chapter

2 identifies that women face different challenges in reaching locations and spatially distributed opportunities (Gil Solá & Vilhelmson, 2022; Stark & Meschik, 2018; M. Zhang et al., 2022). Therefore, this research aims to give insights into the common elements that lead to general mismatch occurrences across different kinds by concentrating on the total frequency of mismatches, which can be more realistically helpful and applicable.

4.5.2 Primary data filtering

This section investigates the dataset that combines the survey answers with the mismatch cases per participant. A primary feature selection is realized in this dataset, depending on the mismatch transport mode type. Since this section aims to analyze personal characteristics that impact mismatch cases, the questions related to individual elements, social competence, material resources, and safety perception are considered in this analysis. Anyhow, competence (driver's license possession), social (transport support), and material (car access) resources are strictly related to car transport mode. Thus, these features are included only in the case of car mismatch analysis.

Furthermore, it is essential to perform comprehensive data cleaning. Initially, the dataset is filtered to include only women's responses, thereby focusing on this specific target population of interest. In addition, any missing data from other questions are excluded from the analysis, analogously to the cluster analysis procedure presented in section 4.4.

4.5.3 Feature Encoding

Apart from the postal code open question, the survey questions were multiple choice, with either a range or specific options. Thus, all questions in this analysis are categorical, meaning the units of observation differ in type or kind, such as a group membership (Alkharusi, 2012).

However, all variables added to the model for regression analysis must be continuous variables. A continuous variable, like time and height, is one on which participants differ in kind or degree (Alkharusi, 2012). Features can be efficiently coded as integers to solve this requirement. In this project, binary questions such as 'Do you have a driver's license' and 'Do you live with a partner' were encoded into integers where 0 represents No and 1 represents Yes.

Moreover, this phase applies the Dummy Coding method to questions that are not binary. Dummy Generation is a coding technique to express group membership in a mutually exclusive and exhaustive form. Any categorical variable with k categories may often be represented by making (k-1) dummy variables with numerical values. In this procedure, all subjects belonging to one group are given the same numerical value, a code, while all matters belonging to the other groups are granted a distinct number value (Alkharusi, 2012). For instance, this research questionnaire asks the participant's age, giving five range options:18-25, 25-35, 36-45, 46-55, 56-65, 66-75, and More than 75. The Dummy coding method transforms each option into a new feature, with values of 0 and 1 for each feature. In this research, '0' means negative and '1' means positive. Therefore, '1' in the '18-25' feature means the participant has between 18 to 25 years old. The Dummy Generation is applied for all non-binary questions included in this phase.

4.5.4 Second Data Filtering

After the features encoding, the dataset presents 37 to 45 features, depending on the transport mode type. However, assigning significance in high-dimensional regression is challenging. Most computationally efficient selection algorithms cannot guard against the inclusion of noise variables (Meinshausen et al., 2009). Realizing a binary logistic regression with many features can be difficult and problematic for several reasons. In the context of this project, excessive correlation or multicollinearity across characteristics might lead to instability in the estimate of coefficients and make it challenging to comprehend the impact of certain features. Additionally, it could make the model more complicated, making it more challenging to decipher and understand the connections between the attributes and the objective variable (Harrell, 2015).

Thus, this project considers different techniques to identify and select the most relevant input variables to the target variable. One standard methodology is Pearson Correlation (Agarwal, 2022), which investigates the absolute value of Pearson's correlation between the target and other features in the dataset. In Pearson's correlation, the score can vary from -1 to +1, where the closest to -1 means that two objects are highly inversely correlated, the closest to 0 means two objects are uncorrelated and the most relative to the maximum score means two objects are highly similar (Berman, 2016). The Pearson correlation of two objects (x and y) sums the product of their differences from their object means and divides the sum by the product of the squared differences from the object means, as described in Equation 4.3:

$$Pearson's\ correlation = \frac{\sum\ (\mathrm{xi}-\mathrm{x})(yi-y)}{\sqrt{\sum\ (\mathrm{xi}-\mathrm{x})^2}\ \sqrt{(yi-y)^2}} \tag{4.3}$$

Thus, the features that present a higher correlation with the target variable are maintained in the dataset. In addition, the multi-collinearity is checked, which means the correlation among variables excluding the target variable. In the case of multicollinearity between variables, the variable with a higher (absolute) correlation with the target variable is maintained. This methodology analysis the multicollinearity among variables that are higher than 0.7.

Furthermore, this phase tests two different feature selection techniques. The first one is Recursive Feature Elimination (RFE), one of the most popular approaches in the literature in the last two decades (Barbiero et al., 2022). It is a supervised methodology that iteratively removes the worst features based on the performance of the target model (Barbiero et al., 2022). The relevance of each element is first determined by training the estimator on the initial set of characteristics and then by any individual attribute or callable. The least crucial components are then removed from the current list of features. Once the appropriate number of features to pick has been attained, the technique is recursively repeated on the trimmed set (Scikit-learn, n.d.). This research applies the RFE approach by using the sk.learn a method called 'sklearn.feature_selection.RFE'.

The second feature selection technique considered is Chi-squared (X^2) test. It is a nonparametric statistical analyzing method applied only to discrete (categorical) data. The most common use of the test is to assess the probability of association or independence of facts. If properly applied, it answers by rejecting the null hypothesis or failing to reject it (Zibran, 2007). The null hypothesis is one of the two interpretations of a statistical relationship in a sample. It assumes that there is no relationship between the feature and target variable, and the relationship in the sample reflects only sampling error (Chiang et al., 2015). To test this null hypothesis using Chi-test, the X^2 value is calculated by Equation 4.4.

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i} \tag{4.4}$$

Where,

O_i = observed frequencies of a variable

E_i = expected frequencies

i = from i to n

n = the number of observations.

The calculated X^2 value is then compared to the critical value from the X^2 distribution table with degrees of freedom. In this case, the number of degrees of freedom equals the number of columns in the table minus one multiplied by the number of rows in the table minus one, or (r-1)(c-1). The critical X^2 is defined according to the degrees of freedom and chosen confidence level, as presented in Appendix C.1. Finally, we reject the null hypothesis if the calculated X^2 value is bigger than the critical X^2 value. One limitation of the chi-squared test is its sensitivity to sample size (Zibran, 2007).

The feature selection techniques, RFE, and X^2 test were tested in each model individually. The number of features for both metrics is pre-determined by the analyst. Thus, this research tested different features amount for each technique. The most suitable technique is chosen according to the models' performance. The metrics of model performance evaluation are Confusion Matrix, ROC-AUC, and Pseudo R^2, later described in section 4.5.6.

4.5.5 Logistic Regression Modelling

The Logistic Regression function of the Python library sklearn creates the Logistic Regression Model. When learning a dependence from data in machine learning models, it is crucial to divide the data into training and testing sets to avoid overfitting (Gholamy et al., 2018). Empirical studies show that the best results are obtained if we use 20-30% of the data for testing and the remaining 70-80% for training (Gholamy et al., 2018). Therefore, this study considers these ranges to split data into test and training samples. Each model consider a different train/test split within this range aiming to achieve the best model performance.

4.5.6 Evaluation Metrics

This phase considers three main metrics for evaluating models' performance: Confusion matrix, ROC- AUC, and Pseudo R^2. In addition, it considers the Odds Ratio and P-value to assess the variable's influence in the model.

Confusion Matrix

The confusion matrix is a trendy measure used while solving classification problems. It can be applied to binary and multiclass classification problems and represents counts from predicted and actual values (Kulkarni et al., 2020). The confusion matrices present four outputs: True Negatives (TN), which show the number of negative examples classified accurately; True Positive (TP), which indicates the number of positive examples classified correctly; False Positive (FP) describes the number of negative examples classified as positive and False Negatives (FN) represents the number of positive examples classified as negative (Kulkarni et al., 2020).

Based on these concepts, accuracy measurement is calculated, which means the number of true values (TN and TP) divided by all observations (TN, TP, FN, and FP). Accuracy, however, can be deceptive when applied to unbalanced datasets; as a result, alternative measures based on a confusion matrix might help assess performance. Hence, the metrics Precision and Recall are famous metrics for classification. Precision shows how accurate the model is for predicting positive values. Therefore, it is calculated by dividing the TP by the sum of TP and FP. Finally, recall helps measure the strength of a model to predict positive outcomes, and it is also known as the sensitivity of a model. Thus it is calculated by dividing the TP by the sum of TP and FN (Kulkarni et al., 2020).

Receiver operating characteristic curve (ROC) and area under the curve (AUC)

In addition, Kulkarni et al. (2020) present the concept of classifying a model performance using the ROC (receiver operating characteristic) curve. Plotting involves determining the true positive rate (TPR) on the y-axis and the false positive rate (FPR) on the x-axis. A classifier must operate within a predetermined range of thresholds between 0 and 1.

The upper left corner of the ROC curve denotes successful classification, whereas the lower right corner denotes unsuccessful classification. If a classifier reaches the upper left corner, it is considered to be effective. Any classifier with a ROC curve below the diagonal is doing worse than random guessing, which is completely counterproductive. The AUC value is a score that ranges from 0 to 1. Any lower diagonal ROC curve classifier will have an AUC score below 0.5. Similarly, the ROC curve in the top diagonal will receive AUC scores greater than 0.5 (Kulkarni et al., 2020).

Pseudo R^2

In logistic regression, a pseudo-R^2 is a statistical measure that indicates how well the model fits the data. It is referred to as "pseudo" because it is an approximation of the R^2 used in linear regression, which cannot be directly applied to logistic regression due to the different nature of the dependent variable.

McFadden's pseudo-R^2 is a commonly used measure in logistic regression (McFadden, 1974). It compares the likelihood of the entire model to the probability of a null model (intercept-only model) and calculates the proportion of the likelihood ratio. The formula for McFadden's pseudo R^2 is as follows:

R^2 = 1 - (log-likelihood of full model / log-likelihood of the null model)

A higher value of McFadden's pseudo-R^2 indicates a better fit of the model to the data, with a maximum value of 1 indicating a perfect fit (McFadden, 1974).

Odds-Ratio

Odds ratios are used to examine the relative chances of the result of interest, in this case, mismatches, occurring given exposure to the variable of interest, such as individual characteristics (Szumilas, 2010). Thus, the value of OR means:

- OR=1 Exposure has no impact on the likelihood of success.
- OR>1 Exposure is linked to a higher likelihood of success
- OR1 Exposure linked to a reduced likelihood of success

P-value

The P-value is the probability, under the assumption of no effect or difference (null hypothesis), of obtaining a result equal to or more extreme than what was actually observed. The P stands for probability and measures how likely it is that any observed difference between groups is due to chance, and conventionally $P<0.05$ represents the statistically significance of a variable (Meinshausen et al., 2009).

5. Results

This chapter presents the results for each methodology proposed in this research: Spatial Accessibility, Perceived Accessibility, Mismatches Identification, Urban Profiles Identification, and Person-based features Identification.

5.1 Spatial Accessibility Results

This section presents the spatial accessibility levels for accessing different facility types using various modes of transportation. Figure 5.1 provides an overview of accessibility levels for reaching grocery stores through walking, cycling, car, and public transport.

The distribution of reachable amenities is similar for walking and cycling. However, regarding public transport, the distribution shows a slightly more dispersed pattern than the first two modes. Conversely, amenities accessible by car are highly concentrated, with predominantly good and very good accessibility levels.

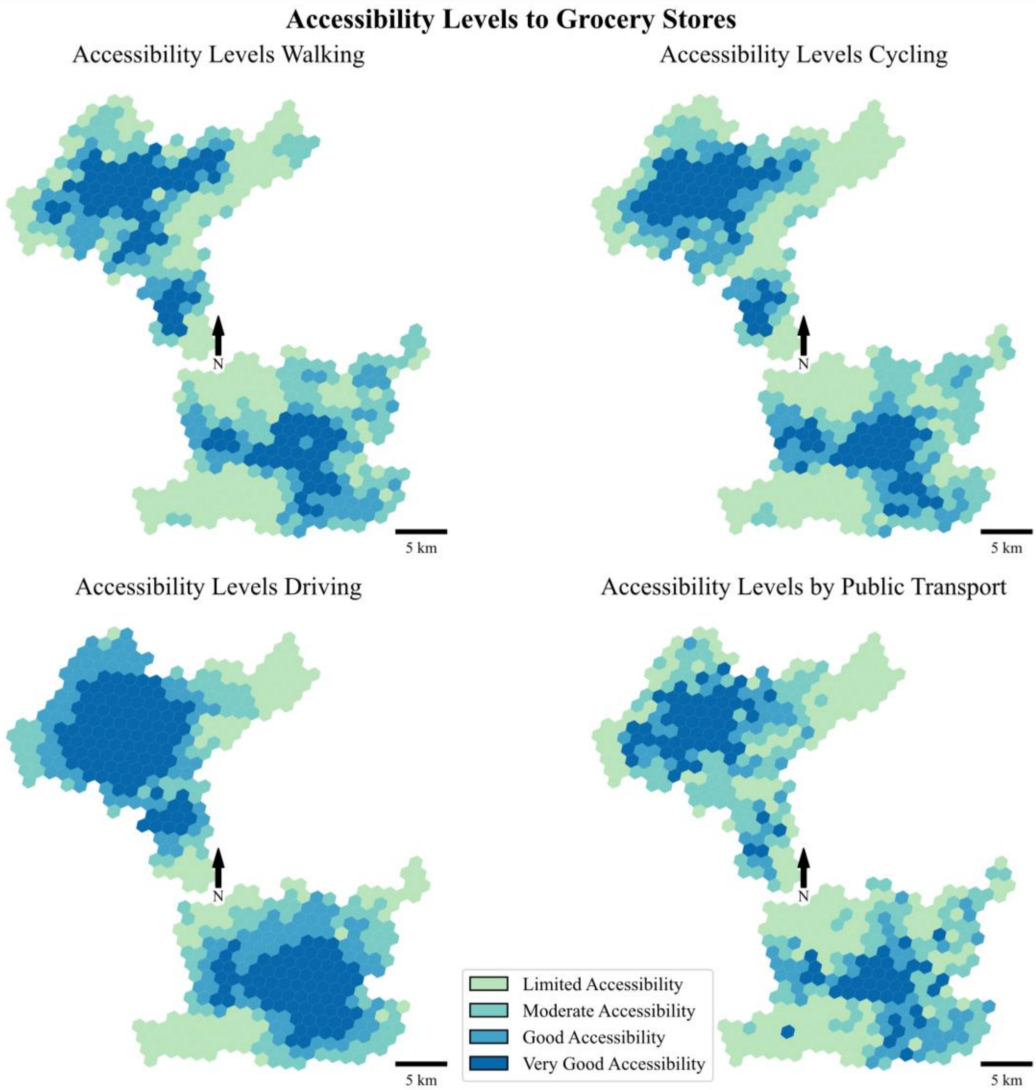


Figure 5.1: Accessibility Levels to grocery stores per transport mode

The concentration of reachable schools by walking, cycling and public transport are located in similar regions. On the other hand, the areas classified as very good or good accessibility by car presents some variances, as presented in Figure 5.2.

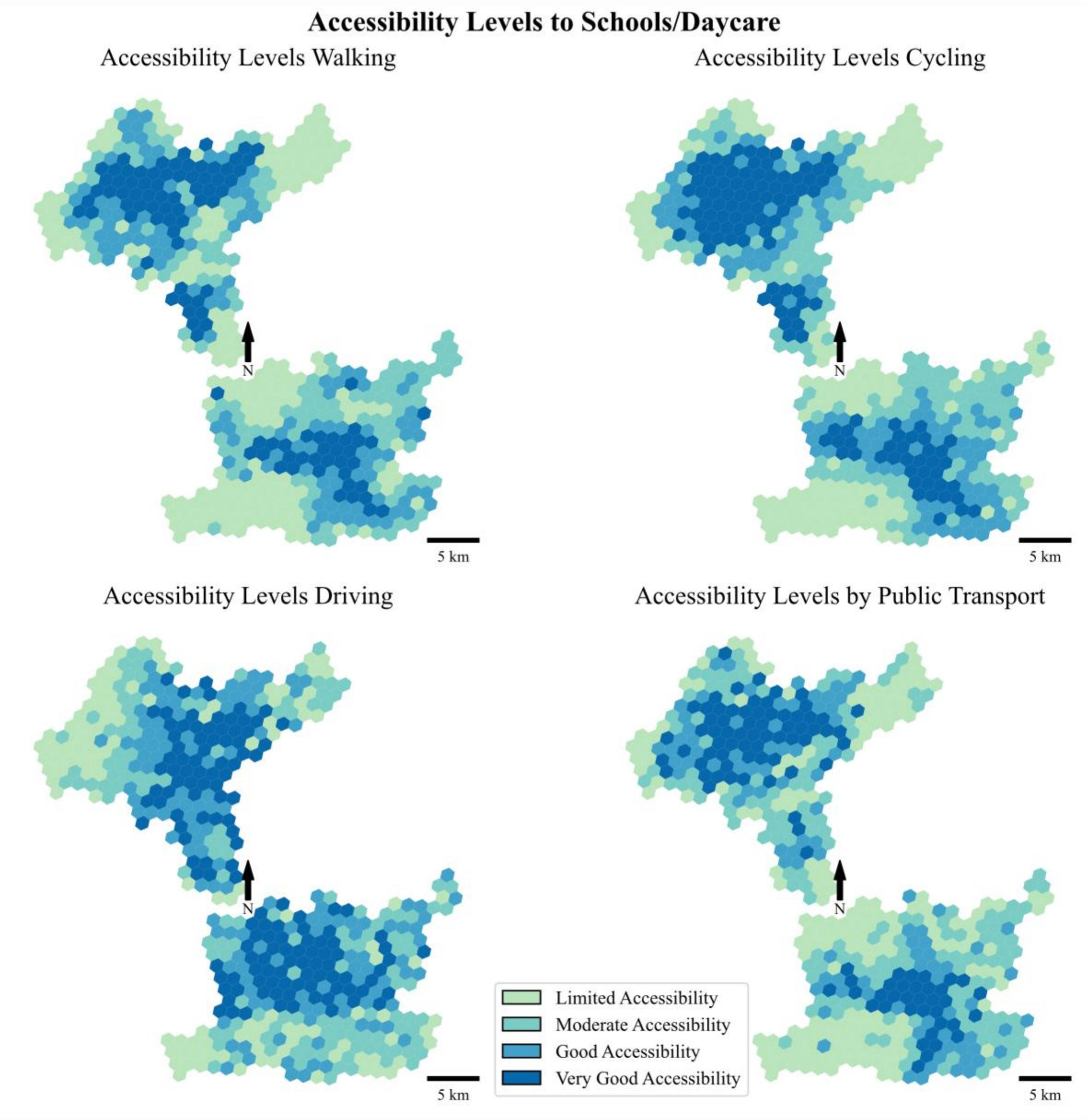


Figure 5.2: Accessibility Levels to Schools/Daycare per transport mode

Furthermore, Figure 5.3 describes the Accessibility levels to Entertainment for each transport mode. The number of regions classified as limited accessibility is higher for walking and cycling modes than other amenity types. The regions considered to have good and very good accessibility by walking or cycling are concentrated in municipality centers. Regarding driving, the accessibility levels are more spread among the geographical scope.

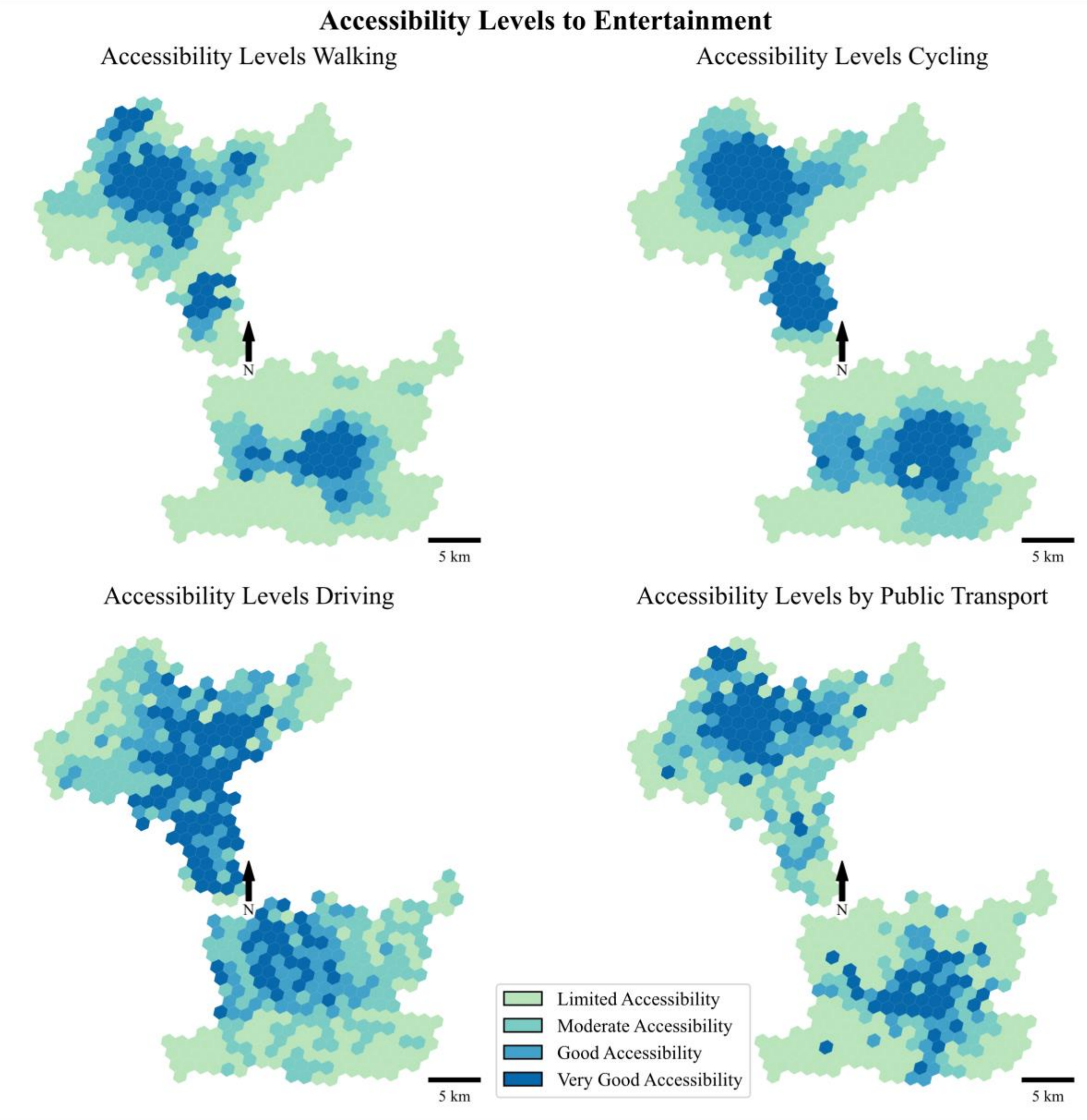


Figure 5.3: Accessibility Levels to Entertainment activities per transport mode

Additionally, when analyzing the cycling accessibility in Delft, most regions are categorized with very good accessibility by bicycle to entertainment spots. On the contrary, in the Leidschendam-Voorburg context, most areas are classified as limited. Figure 5.4 highlights the accessibility levels in both regions.

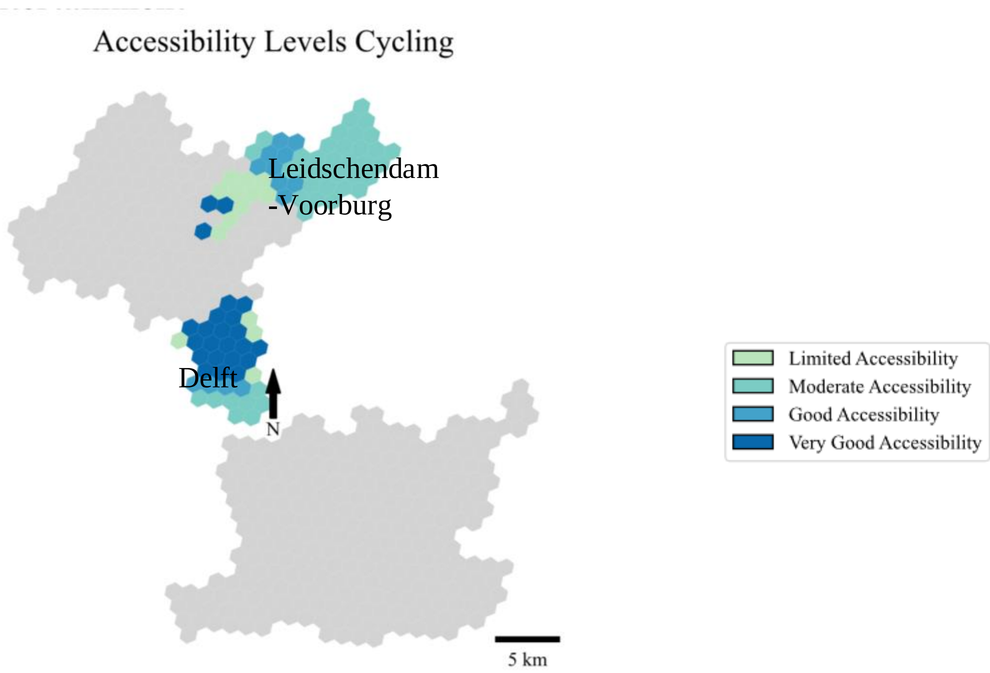


Figure 5.4: Accessibility Levels by cycling to entertainment activities in Leidschendam-Voorburg and Delft

5.2 Perceived Accessibility Results

This section presents the survey application's findings in two sections: descriptive statistics for the sample and a gendered perceived accessibility analysis. The descriptive statistics provide an overview of the participants' characteristics compared to the MRDH population. The gendered perceived accessibility analysis explores how gender influences individuals' perceptions of accessibility.

5.2.1 Descriptive Statistics

The distribution of the questionnaire occurred over eight weeks. This research phase gathered 530 answers, of which 202 were valid after the cleaning process. The answers are distributed in six municipalities of the MRDH: Rotterdam, Schiedam, Den Haag, Delft, Rijswijk, and Leidschendam-Voorburg. This distribution is presented in Table 5.1.

Table 5.1: Survey answers distribution among municipalities

Municipality	Respondents amount	Percentage
Rotterdam	85	42%
Den Haag	41	20%
Leidschendam-Voorburg	36	17.8%
Delft	32	15.8%
Rijswijk	4	2.2%
Schiedam	4	2.2%

Moreover, Table 5.2 shows the sample's descriptive statistics and the Metropolitan Rotterdam-The Hague area population for each demographic variable presented in the questionnaire.

Table 5.2: Descriptive statistics of the sample and MRDH population

Question	Option	Frequency	Percentage (%)	Percentage MRDH (%)
What gender do you identify as?[1]	Female	122	61.6	50.54
	Male	76	38.4	49.46
What is the highest education level you have completed?[2]	Level before middle school (Total)	1	0.5	3.00
	Middle school (Total)	15	7.6	68.00
	HBO (higher professional education) bachelor	53	26.8	-
	HBO (higher professional education) master	20	10.1	-
	MBO (Secondary vocational education)	23	11.6	-
	WO (research-oriented higher education) bachelor	26	13.1	-
	WO (research-oriented higher education) master	55	27.8	-
	PHD	5	2.5	-
	Higher education (Total)	182	91.9	29.00
What is your age?[a]	Between 18 - 25	47	23.7	11.11
	26-35	57	28.8	11.73
	36-45	29	14.6	14.03
	46-55	25	12.6	13.52
	56-65	27	13.6	14.00
	66-75	10	5.1	12.54
	Older than 75	3	1.5	7.83
What was approximately your net income in the last year?[3]	Less than 22.000	43	21.8	-
	109.000 - 131.000	3	1.5	-
	22.000 - 43.500	32	16.2	-
	43.500 - 65.500	55	27.9	-
	65.500 - 87.500	28	14.2	-
	87.500 - 109.000	9	4.6	-
	More than 131.000	3	1.5	-
	Prefer not to say	24	12.2	-
What was approximately the net income of you and your partner together in the last year?[3]	Less than 22.000	7	6.3	-
	22.000 - 43.500	17	15.3	-
	43.500 - 87.000	39	35.1	-
	87.000 - 131.000	22	19.8	-
	131.000 - 175.000	8	7.2	-
	175.000 - 218.000	3	2.7	-
	More than 218.000	3	2.7	-

	Prefer not to say	12	10.8	-
What/where is your country of origin?[2]	Africa	6	3	0.80
	Another European country	37	18.7	5.70
	North America	4	2	-
	Caribbean	1	0.5	-
	Central America	3	1.5	-
	South America	26	13.1	-
	America (Total)	34	17.1	0.50
	Oceania	1	0.5	0.00
	South Asia	3	1.5	-
	South Eastern Asia	7	3.5	-
	Western Asia	4	2	-
	Central Asia	3	1.5	-
	East Asia	5	2.5	-
	Asia (Total)	22	9.5	1.20
	The Netherlands	98	49.5	91.20

Distribution in the population for each demographic variable is obtained from: 1. (UrbiStat S.r.l., n.d.); 2. (OECD, 2016); 3. no relevant source was found for this variable.

Table 5.2 shows that the research sample presents most women (62%); however, the population of MDRH is almost equally divided into women and men. In addition, 92% of the research participants have higher education. It shows that participants are highly educated, compared to 29% of the MDRH population with a higher degree. Furthermore, most participants are young, from 18 to 35 years old, a higher proportion than the population sample.

Another significant discrepancy between distributions is the participant's origin. Approximately half of the participants are foreigners, mostly Europeans (19%) and South Americans (13%). The other foreigner's origins are well distributed among Asia, Africa, and other American Regions. On the other hand, 91% of the MDRH population is Dutch.

Among participants that do not live with a partner, most of them (35.1%) have a salary between 43.500 and 65.500 euros, 21.8% have a salary lower than 22.000 euros, 16% earn between 22.000 and 43.500 euros, 14% between 65.500 and 87.500 euros and 12% prefer not to say. Furthermore, among participants that live with a partner, most couples (35.1%) present a household salary between 43500 and 87000 euros, 20% earn between 87000 and 131000 euros, 15% between 22.000 and 43500 and 11 % prefer not to say. This research did not find relevant sources related to the individual or household income among MDRH population.

This investigation demonstrates, in short, that this sample cannot be regarded as representative of the MDRH community. The sample can be generalized as mainly young foreigners. Still, it is helpful for a preliminary analysis of participant perceptions of accessibility that differ. Care should be taken in generalizing the result to the general MDRH population.

Furthermore, Table 5.3 summarizes the household composition (children and partner presence) and the car-related questions. Most of the participants do not have children (63%) or have one or more children older than 12 years old (25%). It shows that only 12% of the participants are parents of young children. Moreover, 56% live with a partner. Regarding car-related questions, most participants (81%) have a driver's license. The car access is well-distributed among participants: 33% of participants declare always accessing a car, while 30% declare never accessing a car. The other 37% is distributed among the intermediate categories.

Table 5.3: Household composition of the survey's participants

Question	Option	Frequency	Percentage(%)
Do you have children? (You can select multiple options)	No	127	63.2
	Yes, one child or more between 5 and 12 years old	8	4
	Yes, one child or more between 5 and 12 years old,Yes, one child or more older than 12 years old	3	1.5
	Yes, one child or more older than 12 years old	51	25.4
	Yes, one child younger than 5 years old	5	2.5
	Yes, one child younger than 5 years old,Yes, one child or more between 5 and 12 years old	3	1.5
	Yes, two or more children younger than 5 years old	4	2
Are you living together with a partner?	No	87	43.9
	Yes	111	56.1
Do you have a driver's license for a car?	No	38	18.8
	Yes	164	81.2
To what extent do you have access to a car?	I can always make use of a/my car	66	33
	I can sometimes use a car	33	16.5
	I can usually make use of a car	23	11.5
	I do not have access to a car	60	30
	I sometimes can, and sometimes cannot make use of a car	18	9

5.2.2 Gendered Perceived Accessibility Analysis

Since this research aims to identify how accessibility perception differs according to gender, this section presents the following results distinguished by answers of men and women. It's noteworthy that individuals who opted for the 'Others' category in the Gender identification question were excluded from this analysis. This exclusion was made due to the broad range of interpretations encompassed by the 'Others' category, which could potentially dilute the meaningful insights derived from the analysis.

First, the survey results show that the number of men and women who generally agree (options 'agree' or 'strongly agree') have social network support is similar. Figure 5.5 describes these differences. More women 'strongly agree' while more men 'agree.' In addition, a higher fraction of women strongly disagree than men

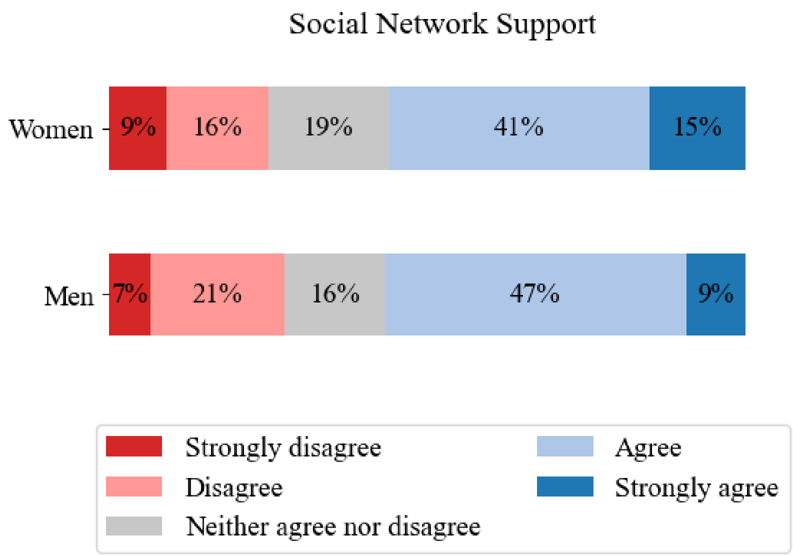


Figure 5.5: Social Network Support Perception by Men and Women

In addition, Figure 5.6 and Figure 5.7 present the car-related questions in a gendered distinction. It shows that despite the majority of driver's license possession in both cases, 12% fewer women have a driver's license. In addition, the 'Car Access Status by Gender' visualization shows that, compared to men, a higher fraction of women do not have access to a car, and a lower fraction usually or always uses a vehicle.

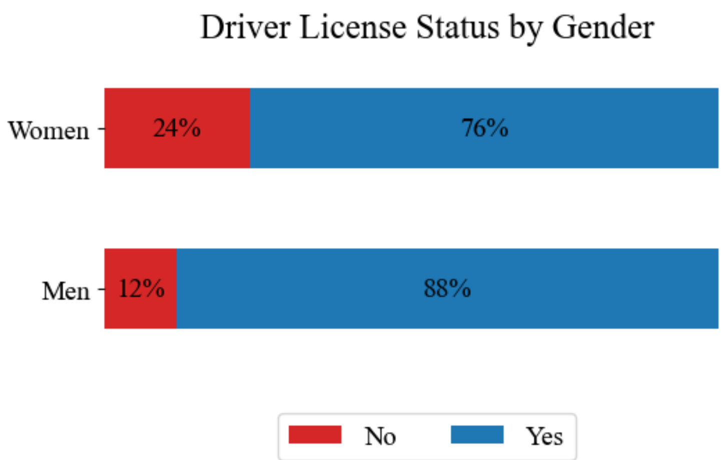


Figure 5.6: Drivers' License Status by Men and Women

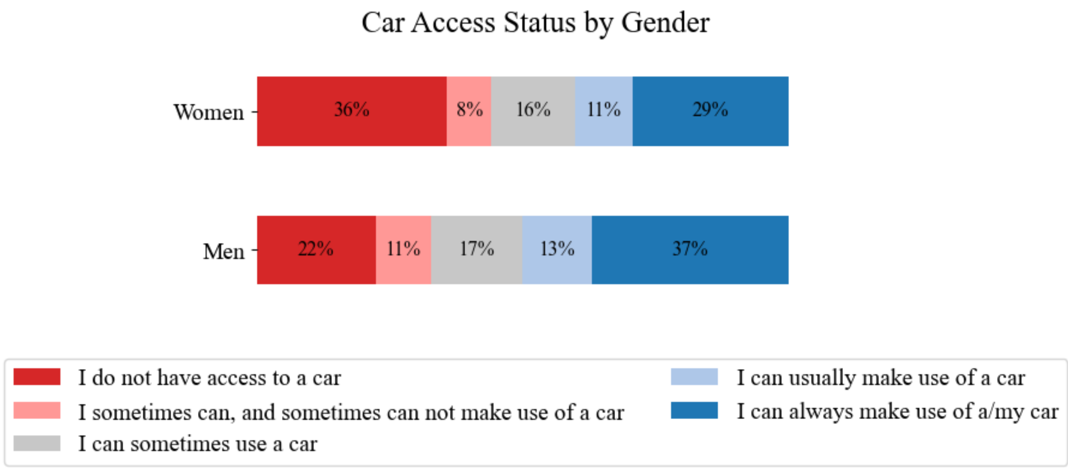


Figure 5.7: Car Access by Men and Women

Moreover, this research asks the participants to rank the aspects according to a ranking order from 1 to 5, where 1 represents the most critical aspect, and 5 represents the last important aspect when they choose a transport mode. Figure 5.8 shows the ranking average for each element, distinguished for women and men. They agree that 'Time' is the most important aspect, 'Money' is the fourth most crucial aspect, and 'Sustainability' is the last important aspect. They disagree on the second and third most important aspects, where women consider 'safety' more critical than 'comfort' while men think the opposite.

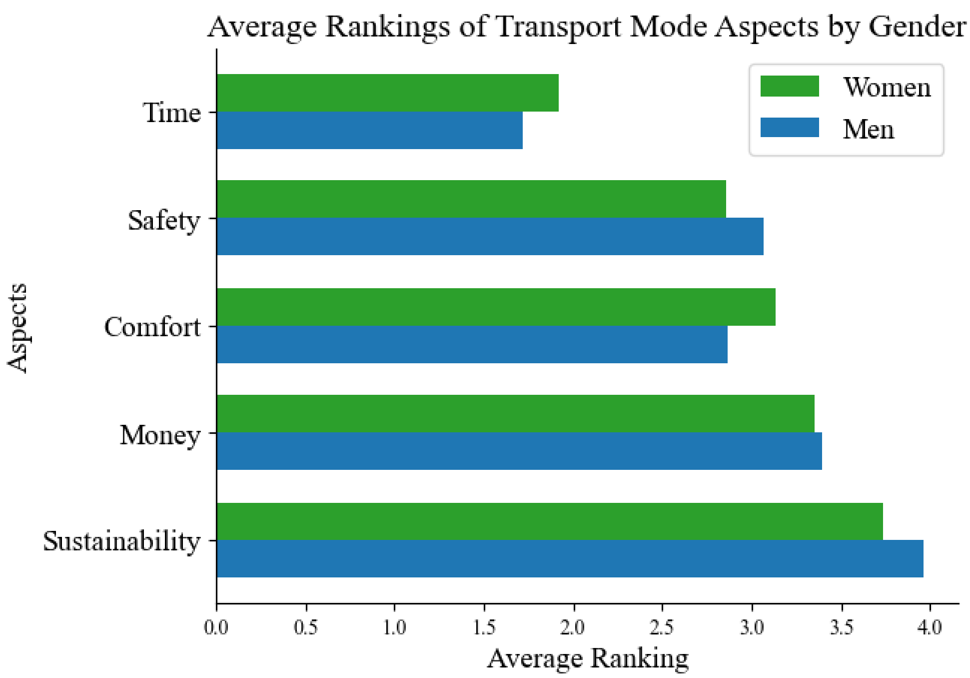


Figure 5.8: Aspects of a Transport Mode Choice ranked by Men and Women

This study also analyzes the safety perception of women and men by using several transport modes at night. Figure 5.9 summarizes the results. The safety perception among women at night by walking, cycling, and public transport presents similar results. The number of women who feel very unsafe or unsafe is higher than men. In addition, the number of women that feel safe or very safe is lower compared to men. Within these three cases, women feel more unsafe or very unsafe walking at night. On the contrary, men feel more unsafe or very unsafe than women by car. The amount of women that feel very safe by car is still lower than men. However, the fraction of women who feel safe by car is 4% higher than men.

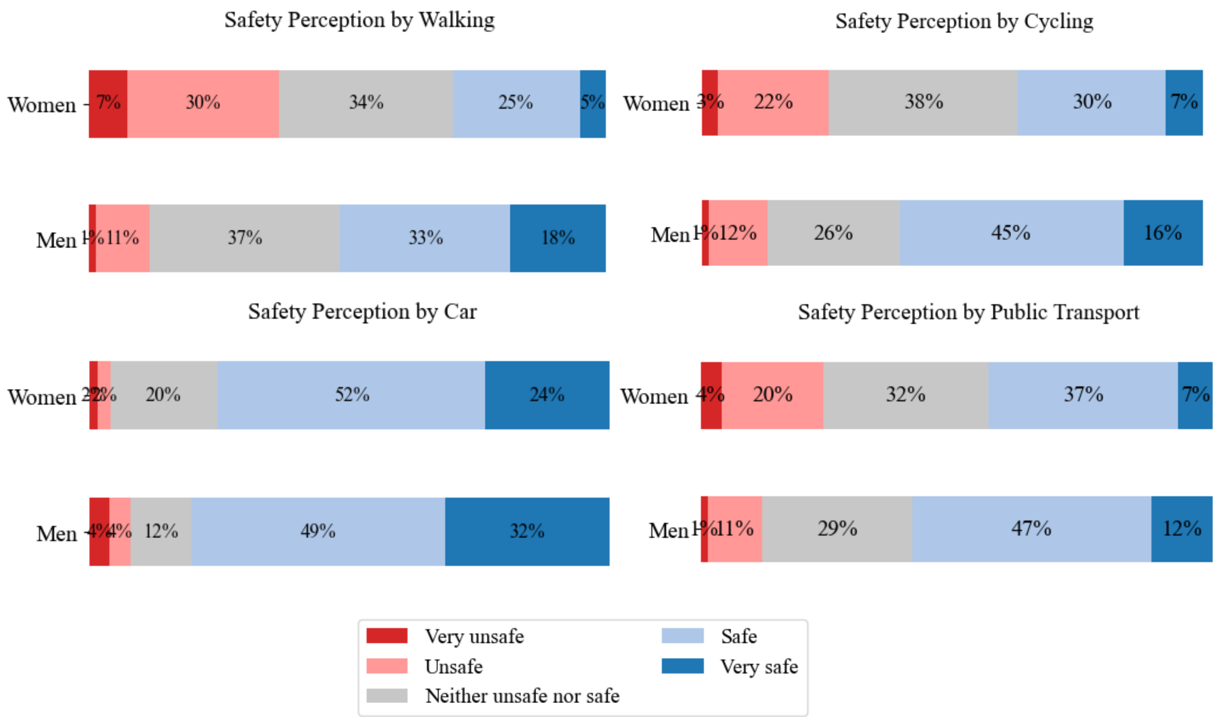


Figure 5.9: Safety Perception of Men and Women at night per Transport Mode

This research contains answers from municipalities that differ in size and population type. Therefore, this chapter briefly investigates the difference between safety perception at night by comparing two distinct areas. This analysis focuses on two municipalities within the MRDH region that differ in terms of population size: Rotterdam, the largest city with 664 thousand habitants (Statista, 2022), and Leidschendam-Voorburg with 77 thousand habitants (Brinkhoff, 2023). While Rijswijk is the smallest city within this geographical scope with 55 thousand habitants (Brinkhoff, 2023), it has a limited number of survey responses, with only four answers. To ensure a robust analysis, we have included Leidschendam-Voorburg, which is smaller than Rotterdam but provides a substantial number of answers (36).

It analyzes walking mode, which presents a higher proportion of 'very unsafe' and 'unsafe' perceptions and public transport due to the difference between public transport options in each municipality. Figure 5.10 shows the safety perception in Rotterdam and Leidschendam-Voorburg by walking. In both cities, the perception of women that feel 'very unsafe' or 'unsafe' is higher than men's. However, 25% of women living in Leidschendam-Voorburg from this sample feel very unsafe walking at night, while only 2% feel the same way in Rotterdam.

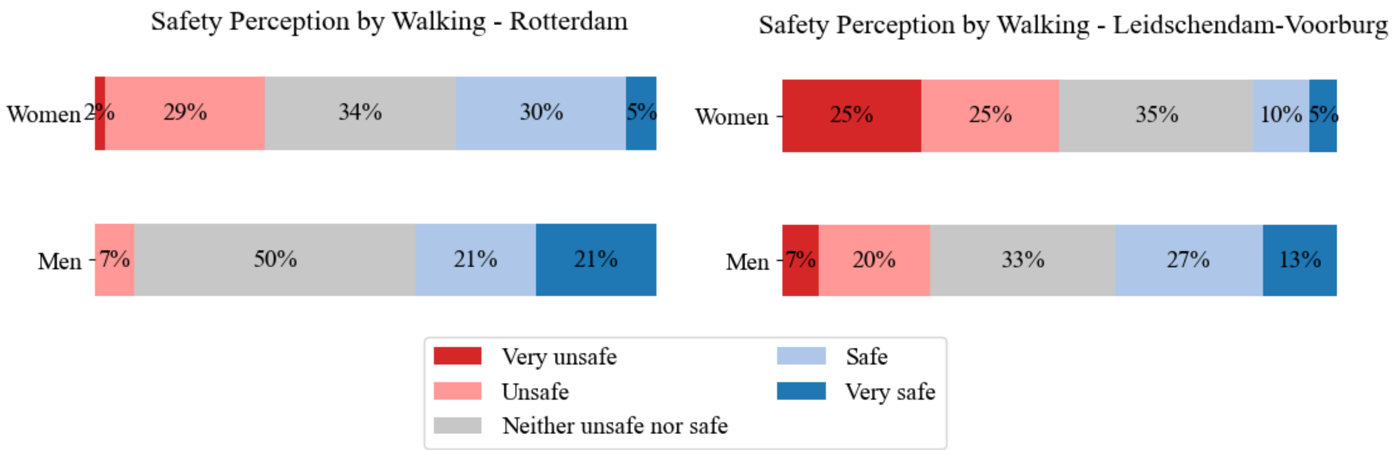


Figure 5.10: Safety Perception of Men and Women at night by Walking in Rotterdam and Leidschendam -Voorburg

Comparing the safety perception of Public Transport by the residents of both municipalities, the proportion of women that feel ‘very unsafe’ or ‘unsafe’ is higher than men. However, as presented in Figure 5.11, the balance of women that feel very unsafe in Leidschendam-Voorburg is 21%, while no women from our sample feel ‘very unsafe’ in Rotterdam. It shows a higher discrepancy regarding safety perception in transport comparing both cities.

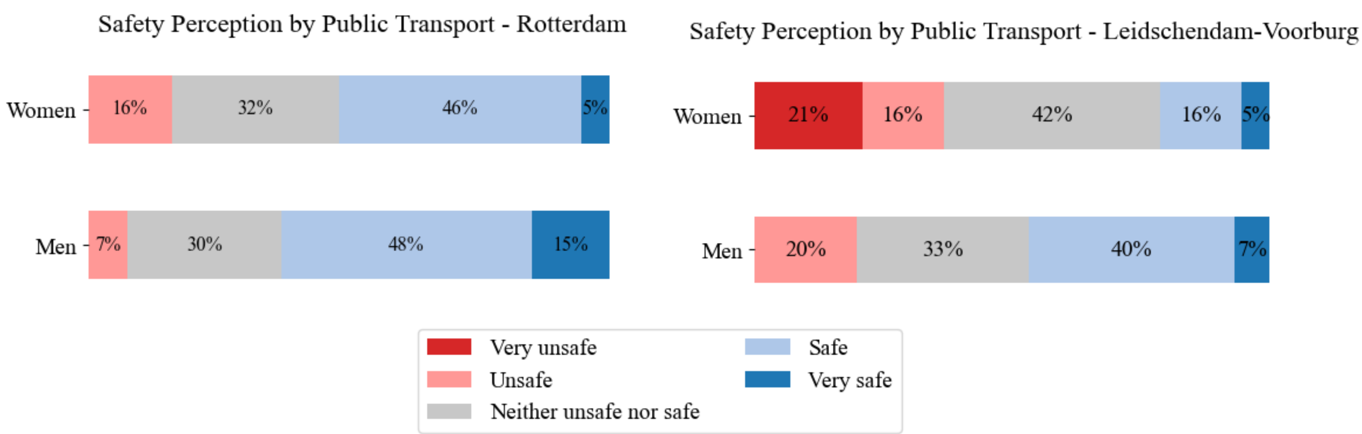


Figure 5.11: Safety Perception of Men and Women at night by Public Transport in Rotterdam and Leidschendam -Voorburg

In addition, the questionnaire includes questions about the primary transport mode used to each of the points of interest investigated in this research. Figure 5.12 describes the primary transport mode men and women use for each activity type.

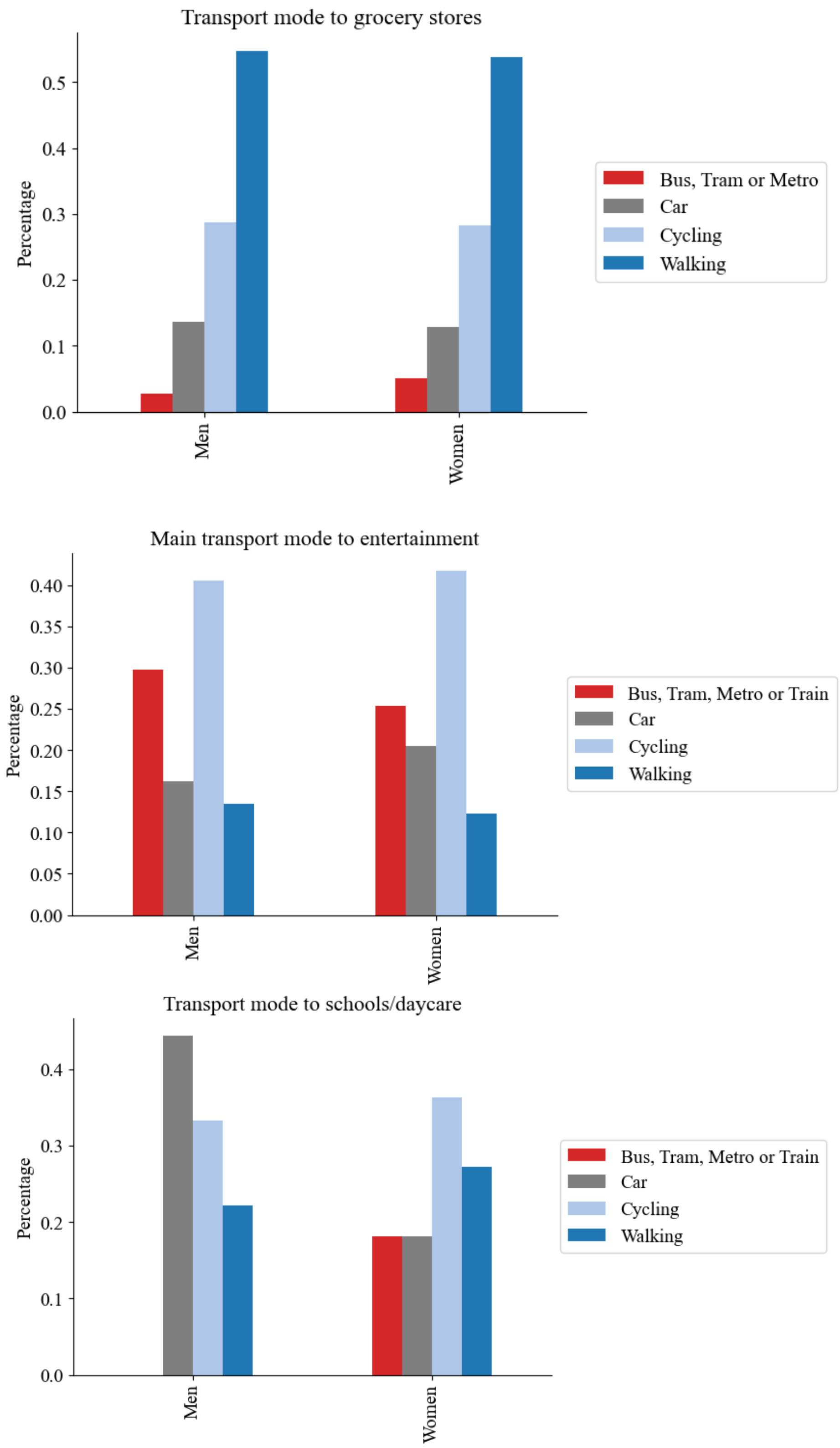


Figure 5.12: Main transport mode of Men and Women for each activity type

The primary transport mode used to grocery stores for men and women is highly similar. Walking is the most popular, followed by bicycle, car and public transport. Women use slightly more public transport than men for this travel purpose. Regarding school or day care, the transport mode preferences differ between men and women. Most women use bicycles as the primary transport mode for this activity type. The second most popular is walking. The exact fraction of women consider Car and Public Transport the primary transport mode, the last popular options for this activity.

On the other hand, men primarily use cars. The second most popular option is cycling, and the third is walking. In this sample, no men use public transport as the primary transport mode to take children to school or day care.

Finally, the proportion of transport modes preferences for entertainment activities is highly similar between men and women. The most popular option is cycling, and the second is public transport. This graph shows that, in this sample, women use public transport as the main transport mode less than men for this activity type. In contrast, women use more cars than men.

The mobility capability answers are not presented in this chapter because they are analyzed in the following section 5.3.

5.3 Accessibility Mismatches Identification Results

This phase consists of comparing spatial accessibility with perceived accessibility. Only aggregated data is presented in this phase to preserve participants' privacy. Thus, this section does not present mismatches per participant but only by groups of men and women. The results are based on cumulative mismatches, transport mode, and mismatches type.

5.3.1 Analysis by Cumulative Amount of Mismatches

This research analyzes a total of twelve combinations resulting from three activity types and four transport modes. However, the study exclusively investigates access to schools and daycares among parents of young children. Therefore, participants without young children are analyzed considering two amenity types and four transport modes, resulting in eight possible mismatches. This phase aims to investigate the cumulative number of mismatches separately for men and women in these two groups.

As mentioned in section 5.5.1, parents of young children represent 12% of the total sample, accounting for 23 participants. Consequently, the investigation of school/daycare amenities and the general analysis of parents of young children are limited to this sample size. Figure 5.13 illustrates the number of mismatches per gender for participants without young children.

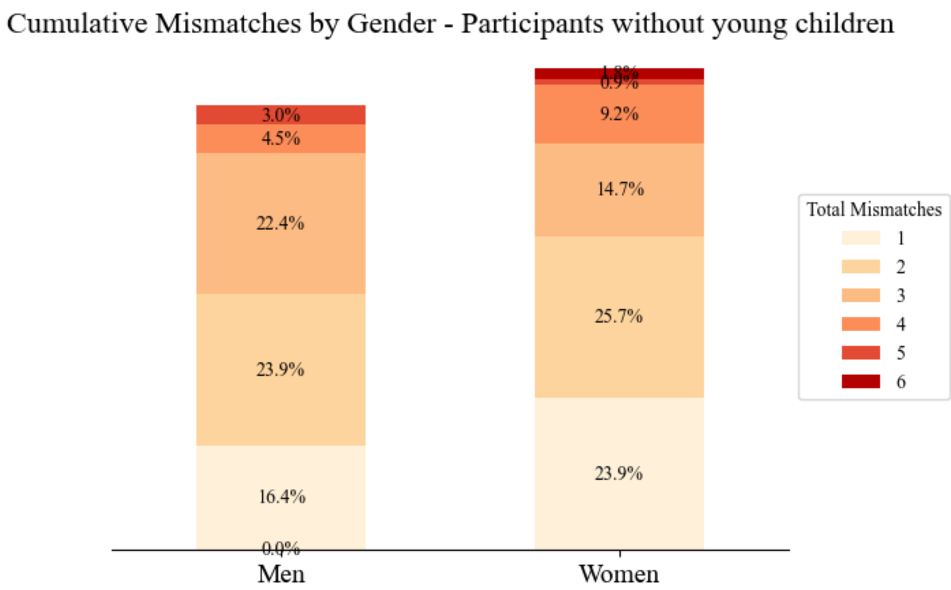


Figure 5.13: Cumulative Mismatches of Participants without young children

The percentage of women with at least one mismatch is higher than men. In addition, the fraction of participants with four or more mismatches is higher between women than men. On the other hand, when visualizing the same analysis for parents with young children, most men present more mismatches compared to women. The results are described in Figure 5.14.

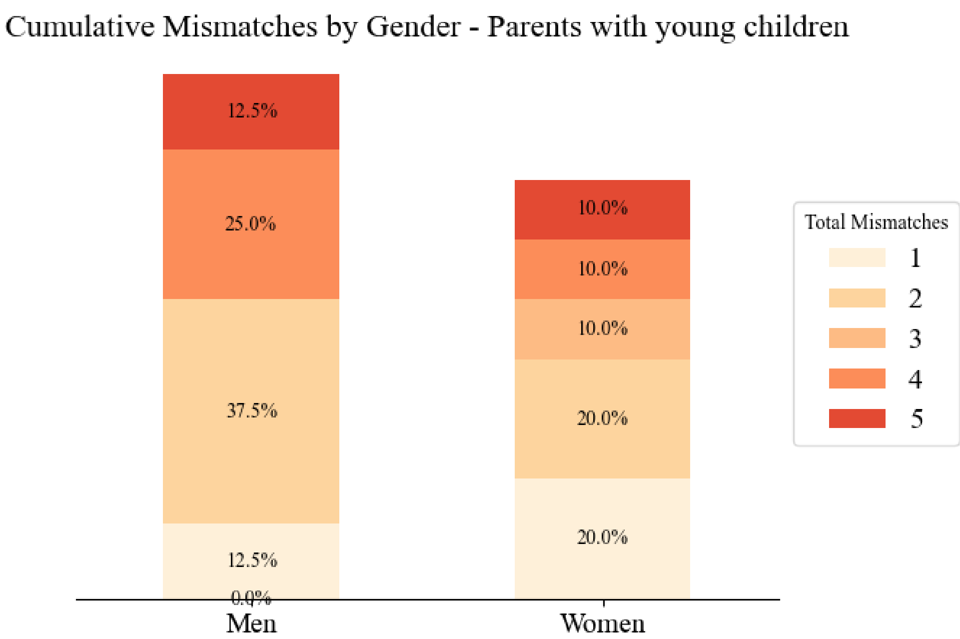


Figure 5.14: Cumulative Mismatches of Participants with young children

5.3.2 Analysis by Transport Mode

Furthermore, this section investigates the fraction of mismatches per transport mode within the two previously described participant groups. The analysis of participants without young children includes grocery stores and entertainment amenities. In contrast, the analysis of participants with young children encompasses primary schools / day care facilities, as well as grocery and entertainment amenities. Figure 5.15 presents the percentage of participants with mismatches per transport mode among participants without young children.

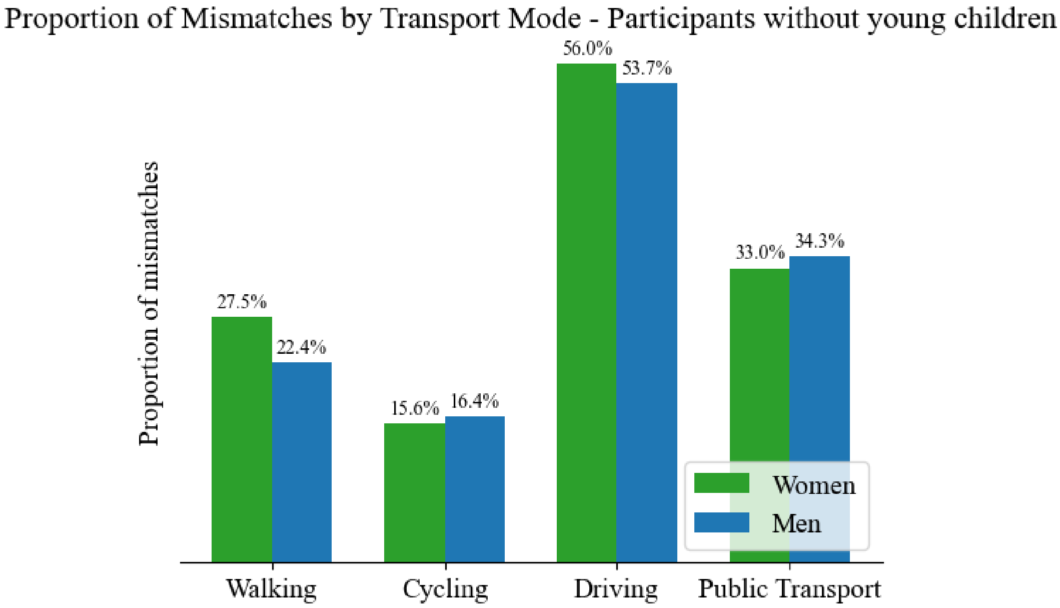


Figure 5.15: Percentage of Mismatches by Transport Mode of Participants without young children

In this analysis, more women perceived accessibility by car or walking as inconvenient or not possible, whereas the area is classified with a good or very good accessibility level. On the other hand, more men perceived not possible or inconvenient cycling or using public transport. It is also noticed that cycling is the transport mode with fewer mismatches while the car presents the higher amount. Among all cases, the walking transport mode presents the higher difference between men and women, which is 5.1%.

When investigating the same analysis among parents of young children, men present more mismatches in all categories, as shown in Figure 5.16. The most gender-equitable transport modes are bicycles and cars, while walking shows a difference of 7.5% between men's and women's mismatches proportion. Furthermore, 22.5% more fathers of young children than mothers see public transport as an inconvenient or not possible transport mode to reach amenities while the areas are considered accessible. It is the most significant disparity between men and women.

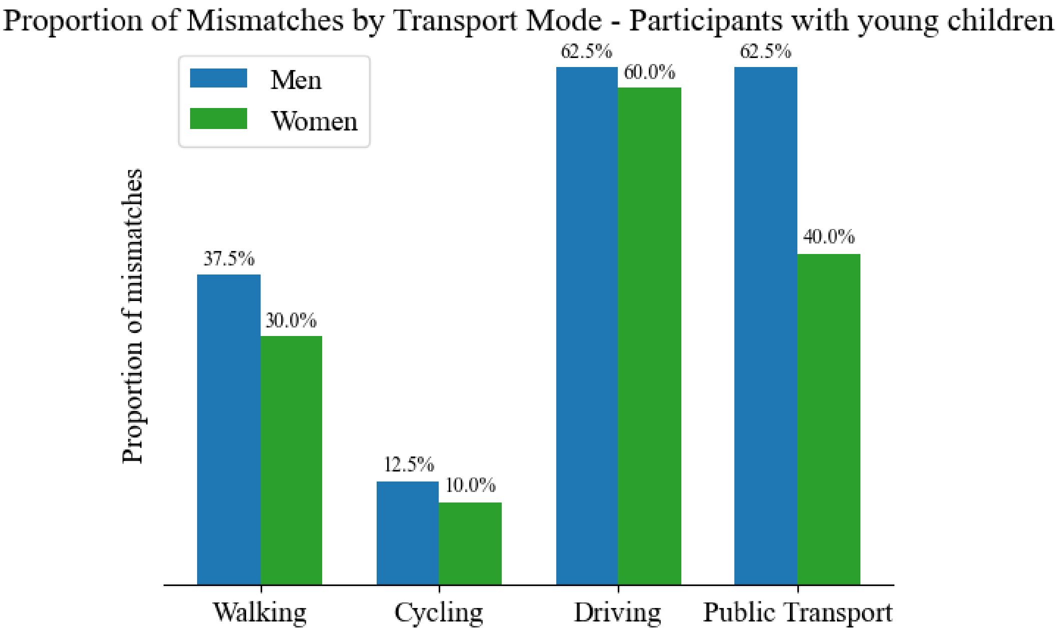


Figure 5.16: Percentage of Mismatches by Transport Mode for Participants with young children

5.3.3 Analysis by Mismatches Type

This study investigates the fraction of mismatch types per activity. This phase analyzes the schools/daycare results only for participants with young children. Figure 5.17 shows the distribution of mismatch types per gender and activity type when walking is the transport mode. Only women perceived accessibility differently than spatial accessibility to reach grocery stores. The mismatches are slight or moderate. When reaching schools or daycare, 37.5% of men perceived it as inconvenient, while the area is categorized with good accessibility. These are all cases of slight mismatches. On the other hand, 10% of encountered a mismatch when reaching schools or daycare; all are moderated mismatches. Finally, the percentage of women that presents mismatches to entertainment activities by walking is slightly higher than men, with more moderate or strong mismatches.

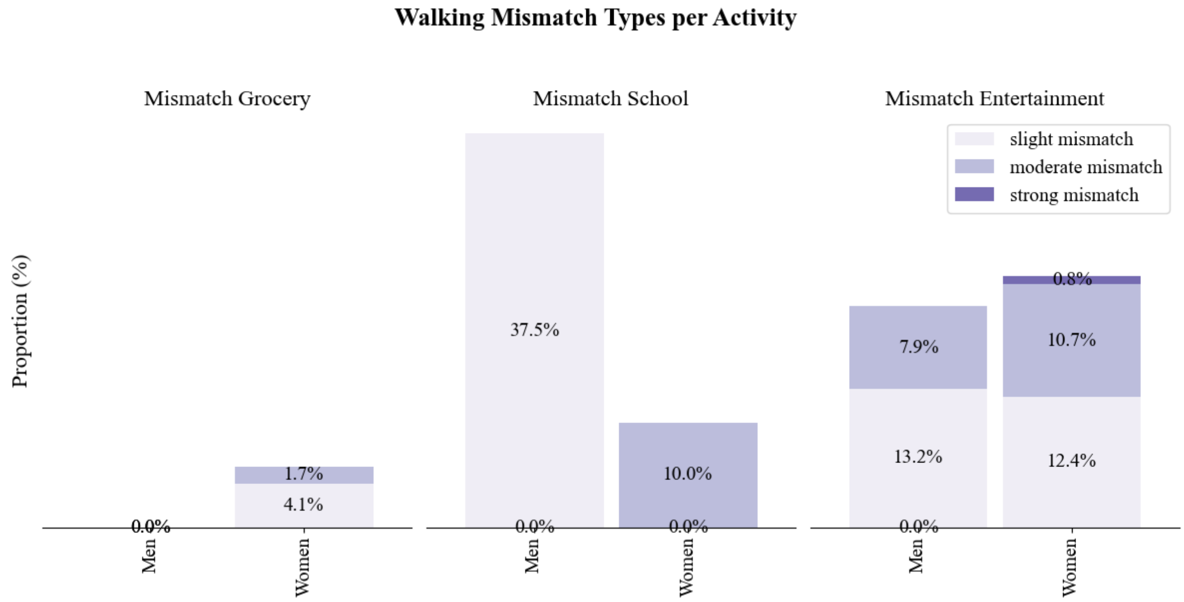


Figure 5.17: Walking Mismatch type per Activity

When analyzing the cycling mismatches, men present a slightly higher proportion in all activity types. Apart from schools/daycare, in general, the differences between gender in this transport mode is relatively low. The results are described in Figure 5.18.

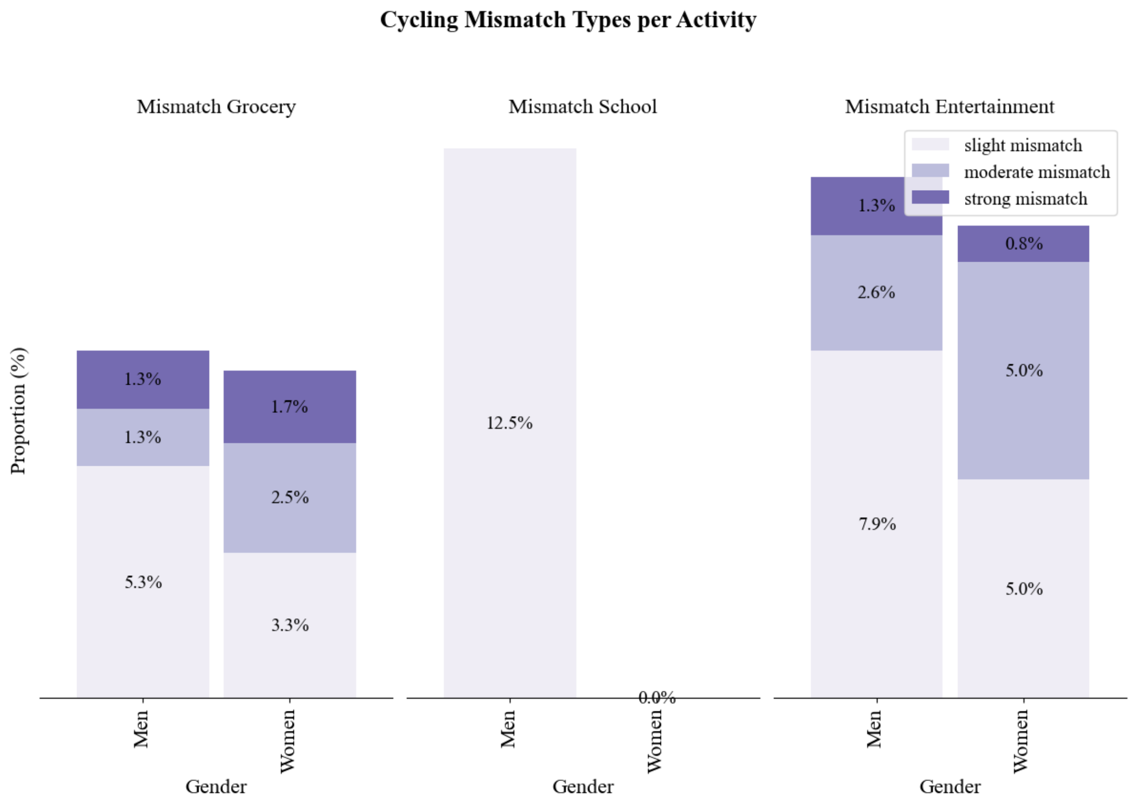


Figure 5.18: Cycling Mismatch type per Activity

Further analyzing the car mismatches, women face higher mismatches when reaching groceries and entertainment, and in both cases, the amount of strong mismatches is higher. On the contrary, men face more mismatches by driving to schools/daycare. These mismatch types are slight or moderate, while the mismatches among women are moderate or strong. The results are presented in Figure 5.19.

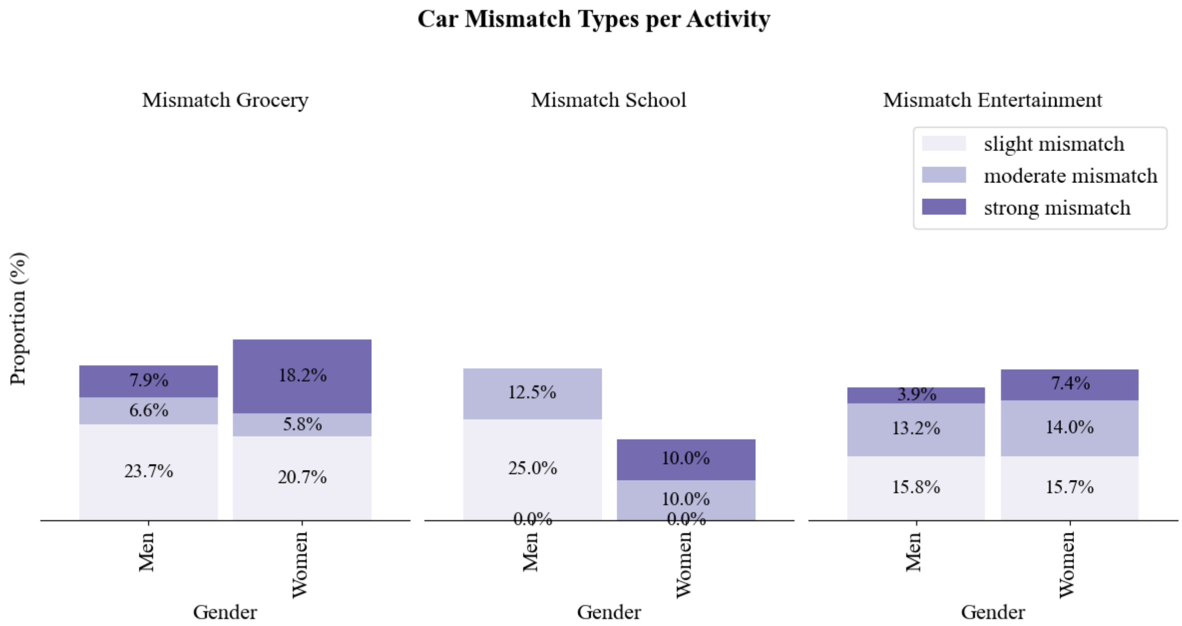


Figure 5.19: Car Mismatch type per Activity

Finally, in the case of public transport, men present a higher mismatch percentage in all activity types. The results are presented in Figure 5.20. The most significant difference between men and women is shown in the schools/daycare analysis.

Regarding access to schools or daycare, most men (62.5%) find it inconvenient or not possible by public transport, where 12.5% of them present strong mismatches. On the other hand, women show only slight or moderate mismatches, representing 20% in total. Finally, men present slightly higher mismatches to entertainment by public transport; however, women present more moderate to strong mismatches among them.

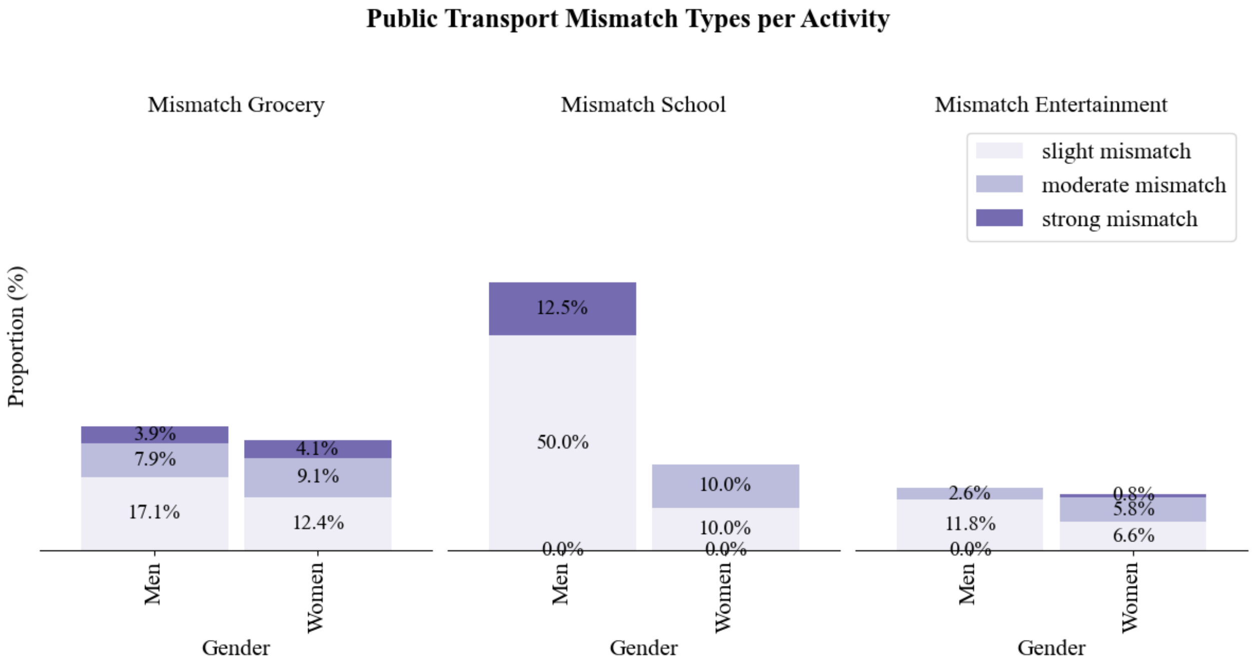


Figure 5.20: Public Transport Mismatch type per Activity

This phase also investigates the proportion of mismatch types for both men and women. Figure 5.21 describes the ratio of mismatch types among all mismatches for two groups, participants without young children and parents of young children.

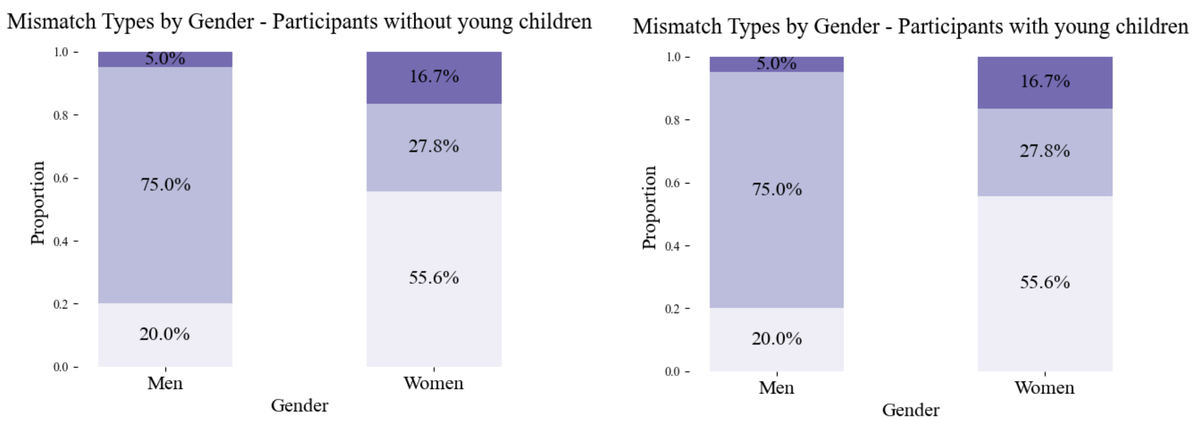


Figure 5.21: Proportion of mismatch types among mismatches

This analysis shows that women in both groups present a higher proportion of strong and slight mismatches than men. In contrast, the proportion of moderate mismatches is higher among men. The discrepancy between mismatch types among men and women is more evident in participants with young children.

5.4 Urban Groups Identification

The identification of urban groups involves conducting two primary clustering analyses: one focusing on clusters that include mismatches related to entertainment activities (named from here on First Cluster Analysis) and another examining clusters that incorporate mismatches associated with grocery stores (named from here on Second Cluster Analysis). The analysis does not consider mismatches concerning travel to schools due to the limited sample size of parents with young children.

This section investigates mainly the individual characteristics, material, and competence components of participants. The perceived social support represented by the question, “If I cannot travel somewhere (important) myself, I think someone in my network (e.g., a friend, a family member) would be available to help me,” is not included in the cluster analysis. Similarly, this phase does not consider the perceived safety question, “How safe do you feel while traveling with the following transport modes during the night (after dark)?”. The main reason behind this decision is that the social support variable did not present differences among clusters, and dropping this variable provided a better cluster performance. The variables related to safety perception were not included in the clustering analysis for the following reasons. Firstly, the safety perception variable consists of four distinct categories based on different modes of transportation. These categories would complicate the clustering analysis and potentially dilute the focus on defining urban groups. Moreover, the influence and impact of safety perception will be thoroughly examined in section 5.5 – Person-based features identification results.

Furthermore, this investigation has chosen not to include the mismatch variable for grocery stores accessed by public transport, as it may result in less reliable findings in the context of the Netherlands. The majority of grocery stores in the area are easily reachable by walking or cycling. Therefore, participants may perceive public transport as inconvenient, not solely due to personal barriers, but also because it is not a common practice considering the distribution of these amenities. Thus, all features included in each clustering analysis are presented in Table 5.4.

Table 5.4: Description of the variables included in each cluster analysis

Variables’ type	1st Cluster Analysis Variables	2nd Cluster Analysis Variables
Individual Components	Gender, Age, Country of Origin, Income, Highest Education Level, Children, Partner presence	Gender, Age, Country of Origin, Income, Highest Education Level, Children, Partner presence
Material Component	Car Access	Car Access
Competence Component	Driver’s License	Driver’s License
Mismatch Analysis	Entertainment activities by walking, cycling, car, and public transport	Grocery stores by walking, cycling, and car.

Both cluster analyses reveal an optimal number of clusters, which is determined to be 3 based on the elbow method and silhouette method. The detailed results for each method can be found in the Appendix D.1 and D.2. The first cluster analysis yields a silhouette score of 0.34, indicating a good score within the range of 0 to 1. Likewise, the second cluster analysis achieves a silhouette score of 0.37. Overall, the results of this section are deemed satisfactory. Further details regarding the results of each cluster analysis are provided in the subsequent sections.

5.4.1 First Cluster Analysis

This analysis identifies three main groups. Table 5.5 describes the most predominant characteristic of each cluster. The first cluster is formed mainly of women from 18 to 25 years old that does not have access to a car but have a driver's license. They are mostly not living with a partner, do not have children, have an HBO Bachelor, and have an income of less than 22.000. Most people in this cluster are from The Netherlands. This cluster does not present any predominant mismatch types for entertainment activities.

The second cluster of this analysis is predominantly composed of women, from 26 to 35 years old and that do not have access to a car. Most of this group have a driver's license, live with a partner, and do not have children. The majority hold a WO master's degree, earn a 43000-65500 salary, and are mainly from other European countries. There is a predominant mismatch in entertainment activities by car.

The third cluster is composed mainly of middle-aged men who always have access to a car and hold a driver's license. The household composition is formed by a partner and one or more children older than 12. Most participants in this cluster have an HBO bachelor, earn 43500 to 65500, and it's from the Netherlands. This cluster does not present any predominant mismatch type.

Table 5.5: Clusters and their predominant values – First Cluster Analysis

Cluster	Mismatch Type				Demographics Data								
	Walk	Bicycle	Car	Public Transport	Gender	Age	Car Access	Driver's License	Living with a partner	Children	Highest Education Level	Net Income	Country of Origin
First Cluster	No	No	Yes	No	Female	26-35	I do not have access to a car	Yes	Yes	No	WO master	43.500 - 65.500	EU
Second Cluster	No	No	No	No	Male	56-65	I can always make use of a/my car	Yes	Yes	Yes, one child or more older than 12 years old	HBO bachelor	43.500 - 65.500	NL
Third Cluster	No	No	No	No	Female	18 - 25	I do not have access to a car	Yes	No	No	HBO bachelor	Less than 22.000	NL

In addition, this analysis investigates the proportion of values in each variable. Figure 5.22 describes the ratio of each mismatch type among the clusters. The proportion of walking and entertainment activity mismatches is similar among all clusters. The first cluster presents slightly less frequency of this mismatch type. When analyzing the car mismatches, most participants in the first cluster present this mismatch type. Around 10% present this mismatch in the second cluster and approximately 30% in the third. The mismatches related to public transport are slightly higher for the second cluster.

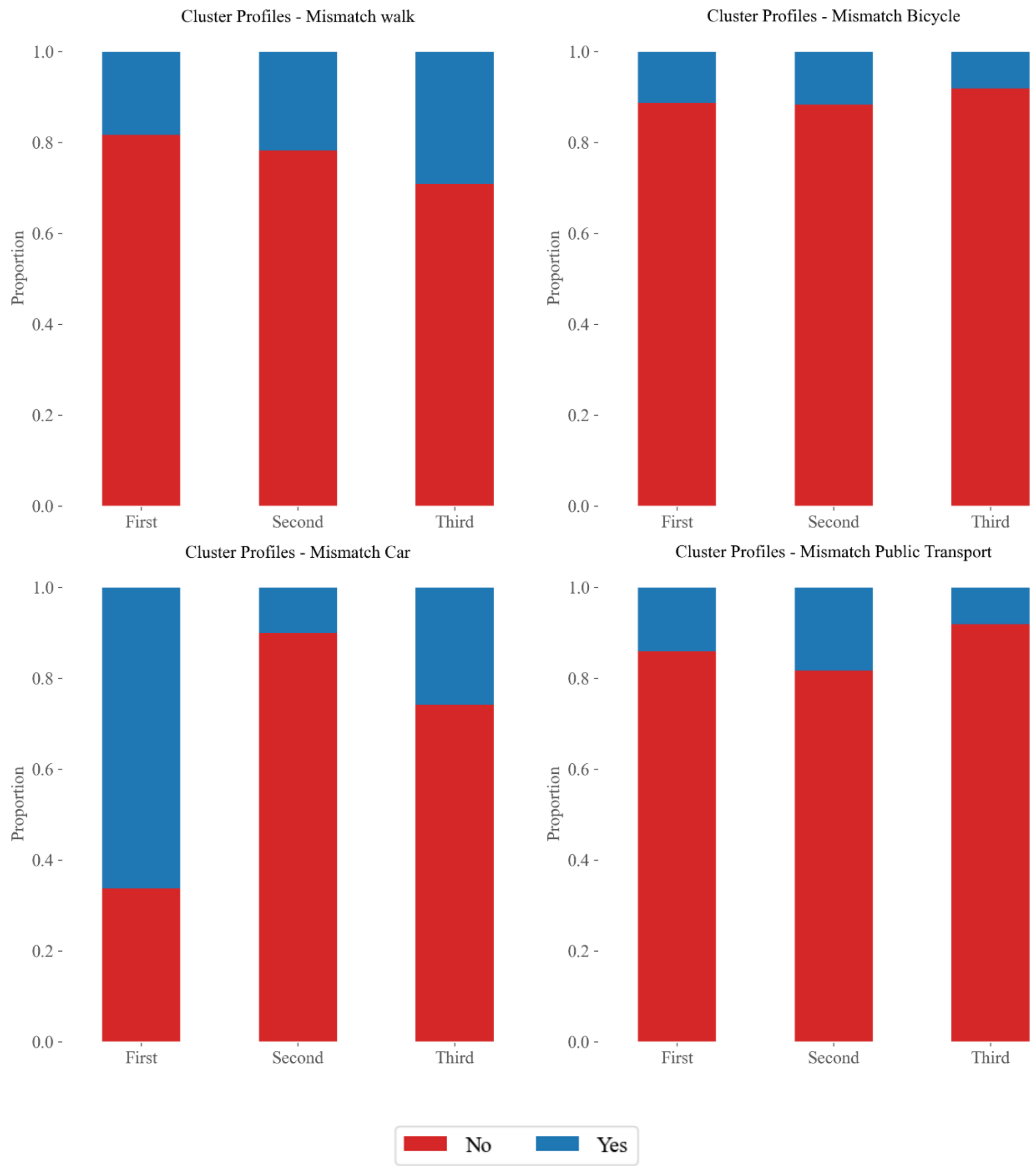
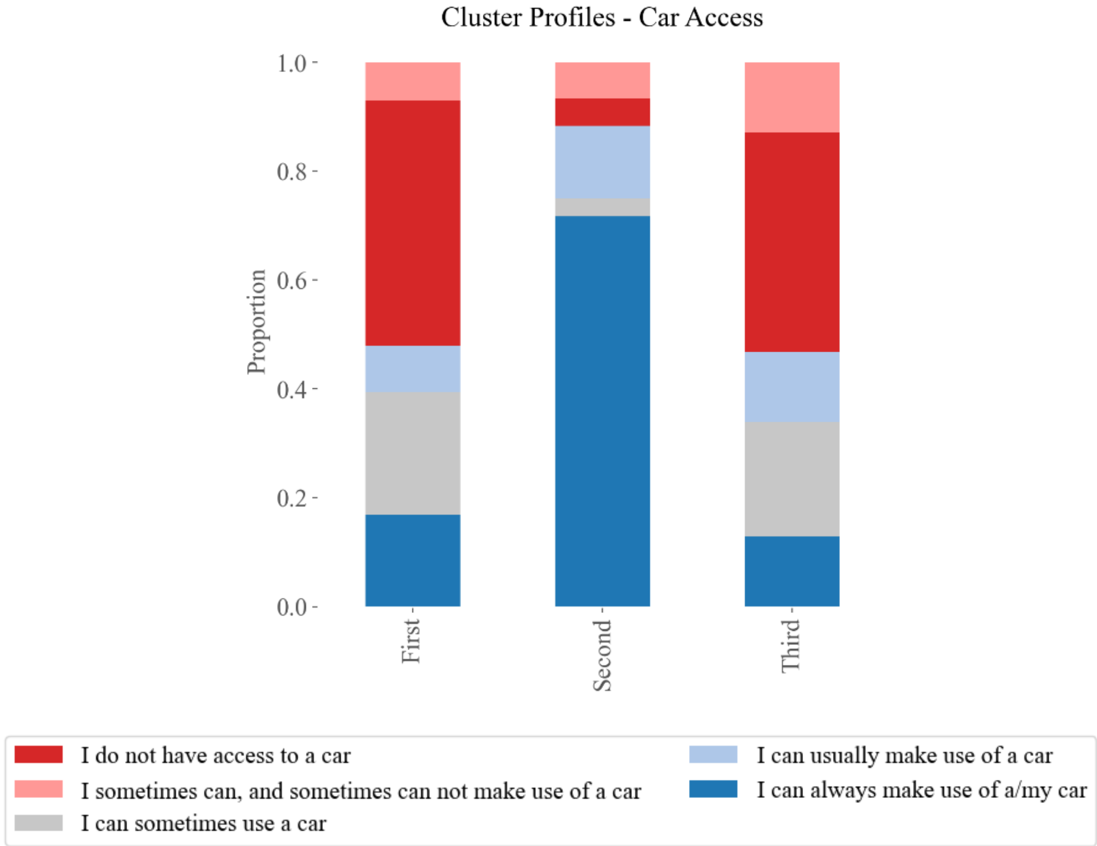


Figure 5.22: The proportion of mismatch types among clusters

In this sequence, Figure 5.23 describes each cluster's proportion of car access and driver's license possession. Most participants in clusters one and three cannot access a car. The majority of participants in the second cluster always have access to a car. The fraction of participants that do not hold a driver's license is similar to clusters 1 and 3, and zero to cluster 2.



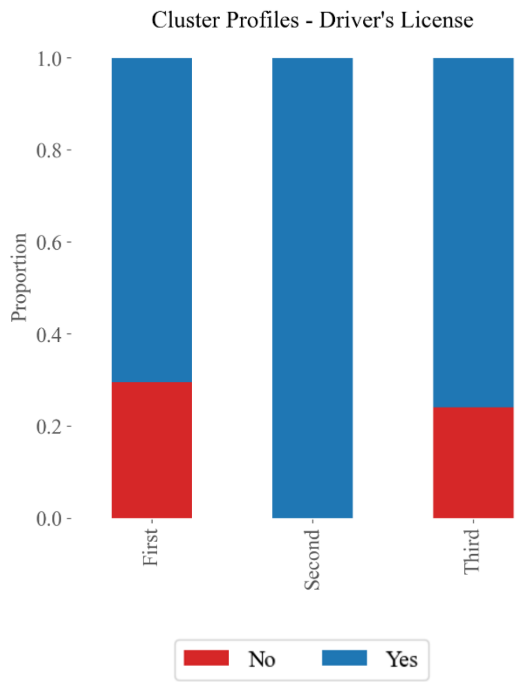


Figure 5.23: Car access and driver's license possession for each cluster

In addition, when analyzing the gender distribution, it is identified that the first cluster is predominantly composed by women. The third cluster is also primarily composed by women, with a higher proportion of men than the first cluster. On the contrary, the second cluster presents mainly men. Regarding age, the first cluster varies mainly from 26 to 45, the second cluster ranges from 46 to 65 and the third cluster is predominantly young, with the most common ages from 18 to 35. Figure 5.24 summarizes these results.

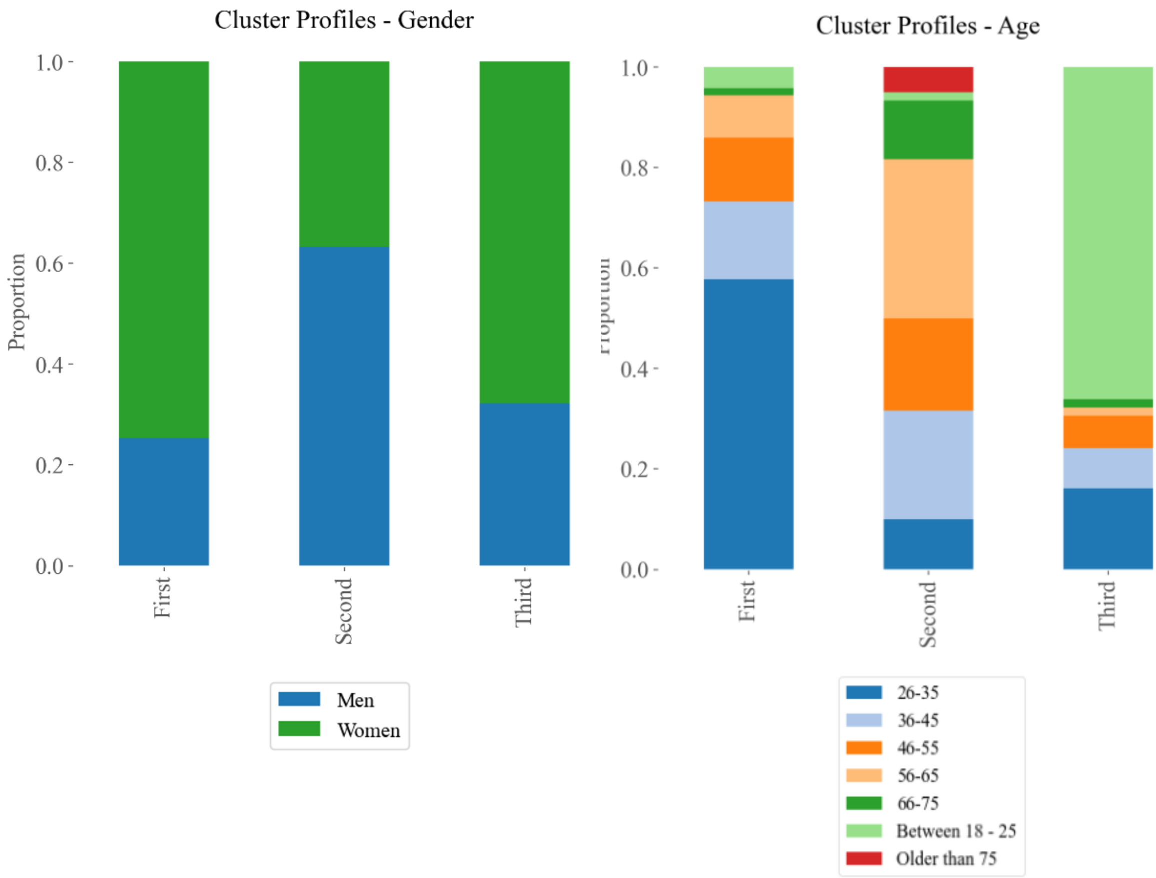


Figure 5.24: Gender and Age distribution among clusters

The household composition per cluster is described in Figure 5.25. The first cluster's majority lives with a partner and does not have children. The second cluster shows mostly people living with a partner and with children and the third one includes mostly participants without children and not living with a partner.

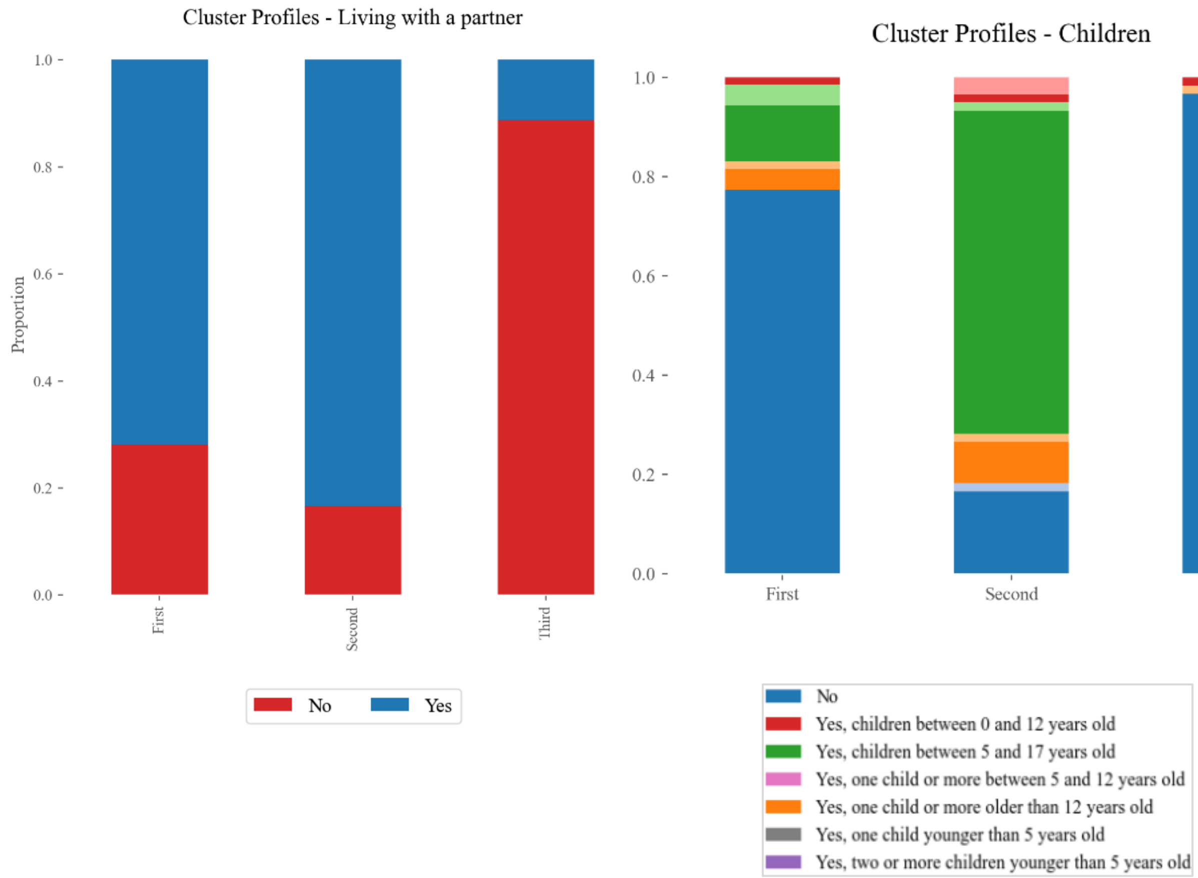


Figure 5.25: Household composition among clusters

Figure 5.26 summarizes each cluster's education level and net income proportions. The education level of the third cluster is well distributed, and the highest ratios are MBO and WO bachelor. The first cluster’s predominant education is also mixed; however, participants with WO master are the majority. The higher education of the second cluster is mostly HBO bachelor and WO master. Regarding income, the income of the first cluster varies mostly from 22.000 to 65.500, the second cluster from 43.500 to 87.500 and the third cluster varies mainly from less than 22.000 to 43.500.

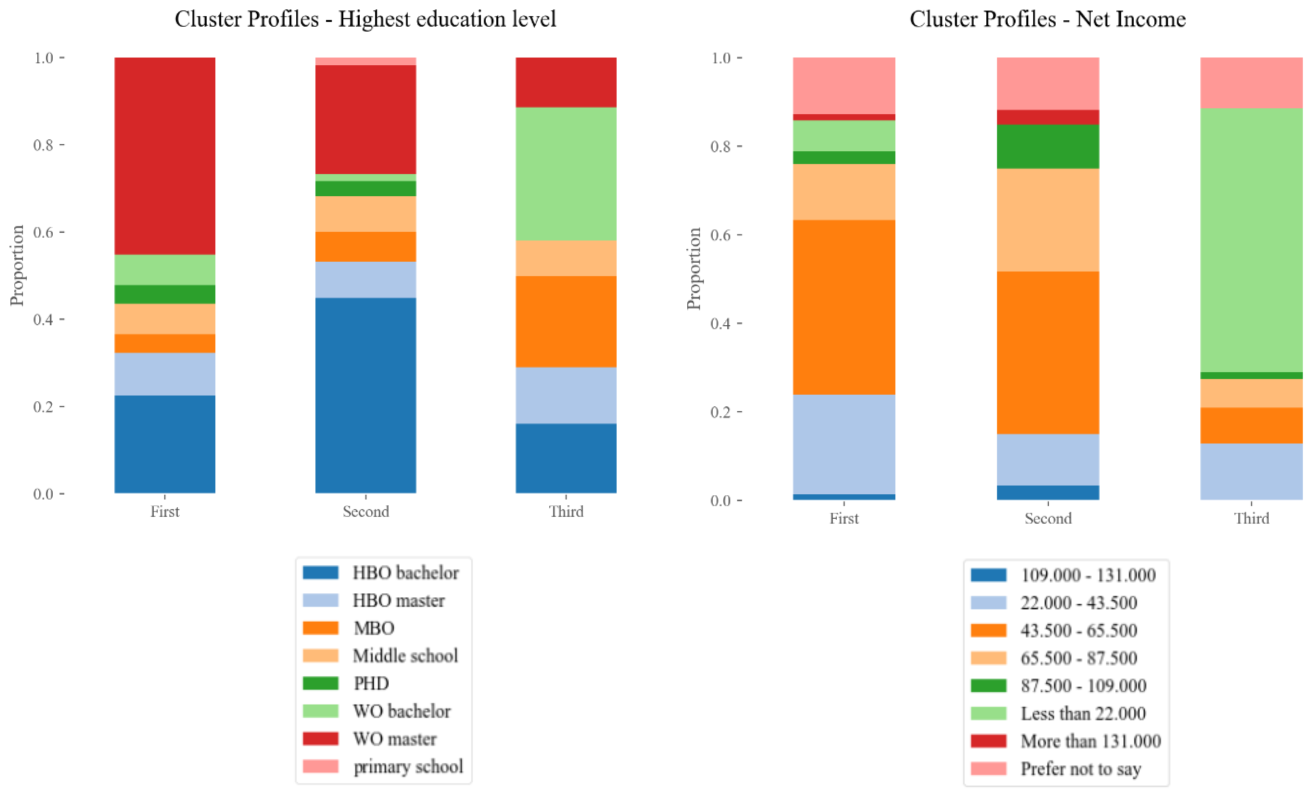


Figure 5.26: Education level and net income among clusters

Finally, the country of origin of each cluster is presented in Figure 5.27. People from the first cluster are predominantly foreign, mainly from other European countries and South America. The second cluster shows the highest amount of Dutch people. Despite the majority of Dutch, the third cluster proportion is almost divided between locals and foreigners.

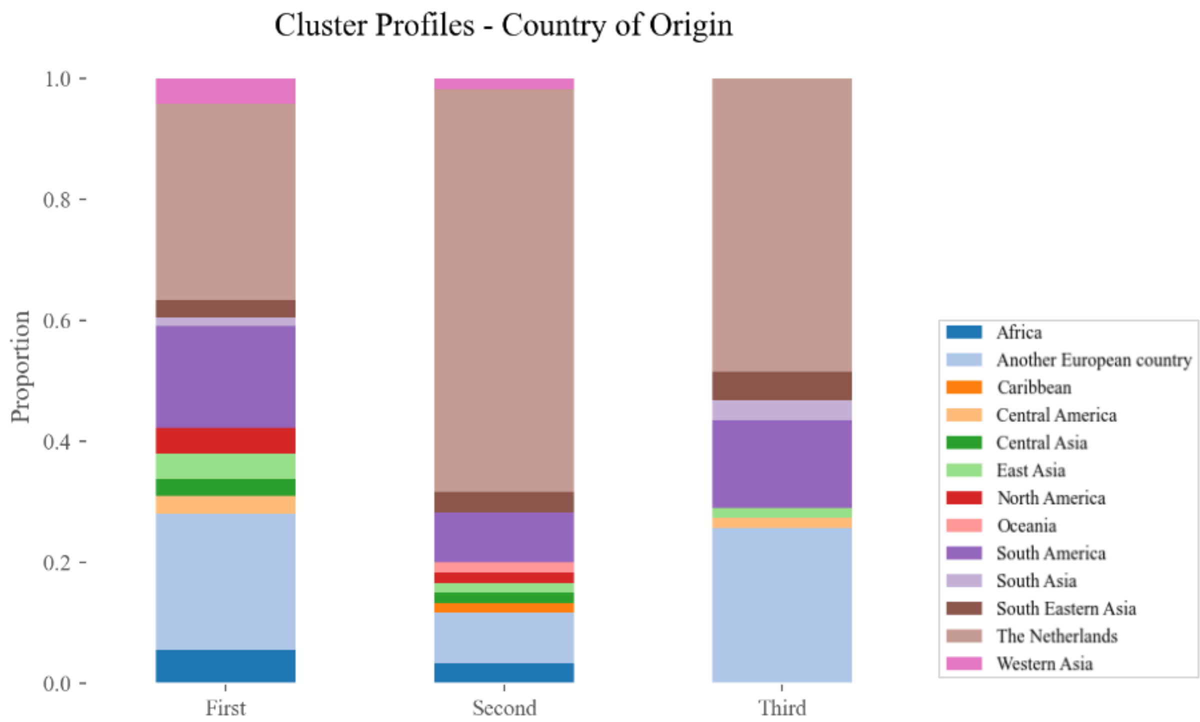


Figure 5.27: Country of Origin among clusters

From this deep investigation, this research describes that the first cluster is mostly of women foreigners around their thirties with an average salary, with partners, and no children. The second cluster comprises senior Dutch men adults with higher salaries, partners, and children. Finally, the third cluster is mainly students that earn a lower salary range, mixed origin with no children or partners.

5.4.2 Second Cluster Analysis

The second cluster analysis investigates the creation of clusters combining demographic data with mismatches related to grocery stores. Table 5.6 presents the clusters created in this investigation and the predominant values for each group.

The first cluster of this analysis shows the same predominant values as the first cluster of the previous examination. It means this cluster is mainly represented by foreign women from 26 to 35 years old, living with a partner, without children, and without access to a car. Similarly, this group presents predominant mismatches to grocery stores by car.

The second group of this analysis is represented mainly by middle-aged women with access to a car, living with a partner, with children, and originally from the Netherlands. This group does not show any predominant mismatch types.

The third cluster is composed of young men from 18 to 25 years old that are not living together with a partner and do not have children. Most participants of this group declared that they could sometimes use a car. They are originally from The Netherlands and do not present predominant mismatch types.

Table 5.6: Clusters and their predominant values – Second Cluster Analysis

	Mismatch Type			Demographics Data								
Cluster	Walk	Bicycle	Car	Gender	Age	Car Access	Driver's License	Living with a partner	Children	Highest Education Level	Net Income	Country of Origin
First Cluster	No	No	Yes	Female	26-35	I do not have access to a car	Yes	Yes	No	WO master	43500-65500	EU
Second Cluster	No	No	No	Female	56-65	I can always make use of my car	Yes	Yes	Yes, one child or more older than 12 years old	HBO Bachelor	43.500 - 65.500	NL
Third Cluster	No	No	No	Male	18-25	I can sometimes use a car	Yes	No	No	WO bachelor	Less than 22000	NL

When further investigating the distribution of mismatch types among clusters, Figure 5.28 presents that the proportion of walking mismatch to grocery stores is very low for all clusters. Still, the fraction is slightly higher for the second cluster, the middle-aged women. Moreover, the mismatches by bicycle to grocery stores are somewhat higher among the first cluster, the foreign women adults. In contrast, the young Dutch men present a lower percentage of mismatches by bicycle. When evaluating the mismatches by car, as previously explained, there is a higher proportion among the first cluster. The second cluster shows the lower ratio of this mismatch type.

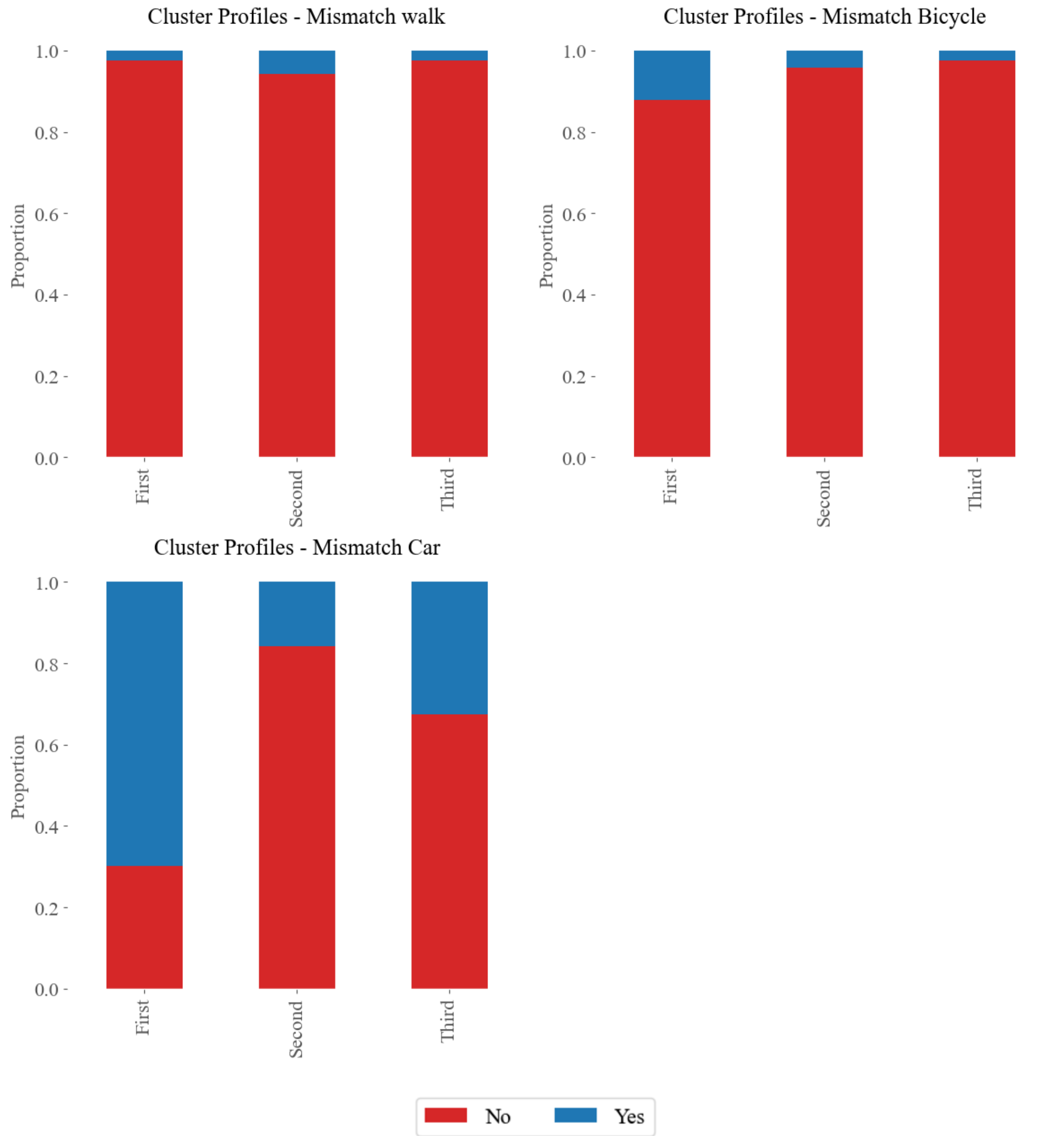


Figure 5.28: The proportion of mismatch types among clusters

The gender and age proportion of these clusters is presented in Figure 5.29. The first cluster is composed mainly of women. The second and third clusters show a more distributed proportion between men and women, with a predominantly group of men in the third cluster. Moreover, the first cluster participants' age is around 26 to 45, from 46 to 65 in the second cluster, and between 18 and 25 in the third one.

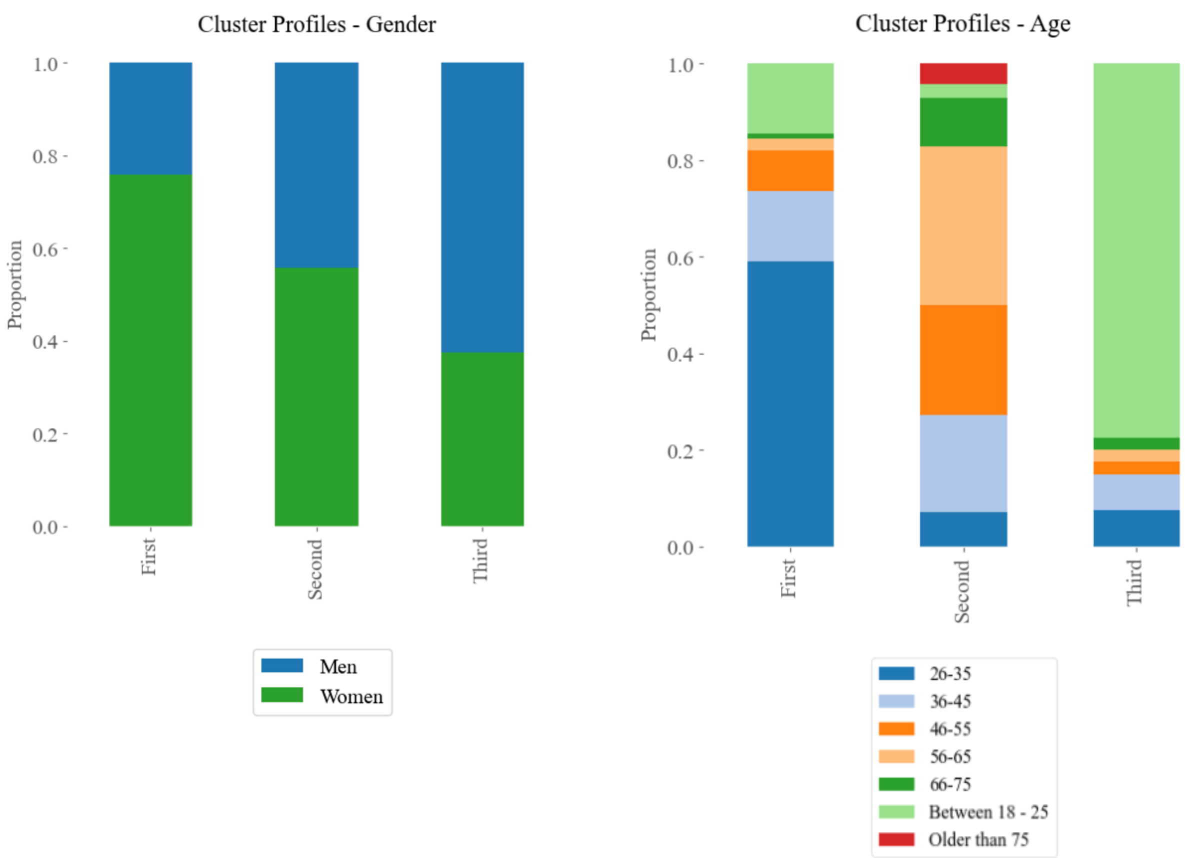


Figure 5.29: Gender and Age distribution among clusters

Furthermore, Figure 5.30 describes each cluster's car access and driver's license possession proportions. The first cluster participants predominantly do not have access to a car, the second can always use a car, and the third cluster presents varied car access. Driver's license possession is around 70% of the participants in the first and third clusters and close to 100% percent in the second cluster.

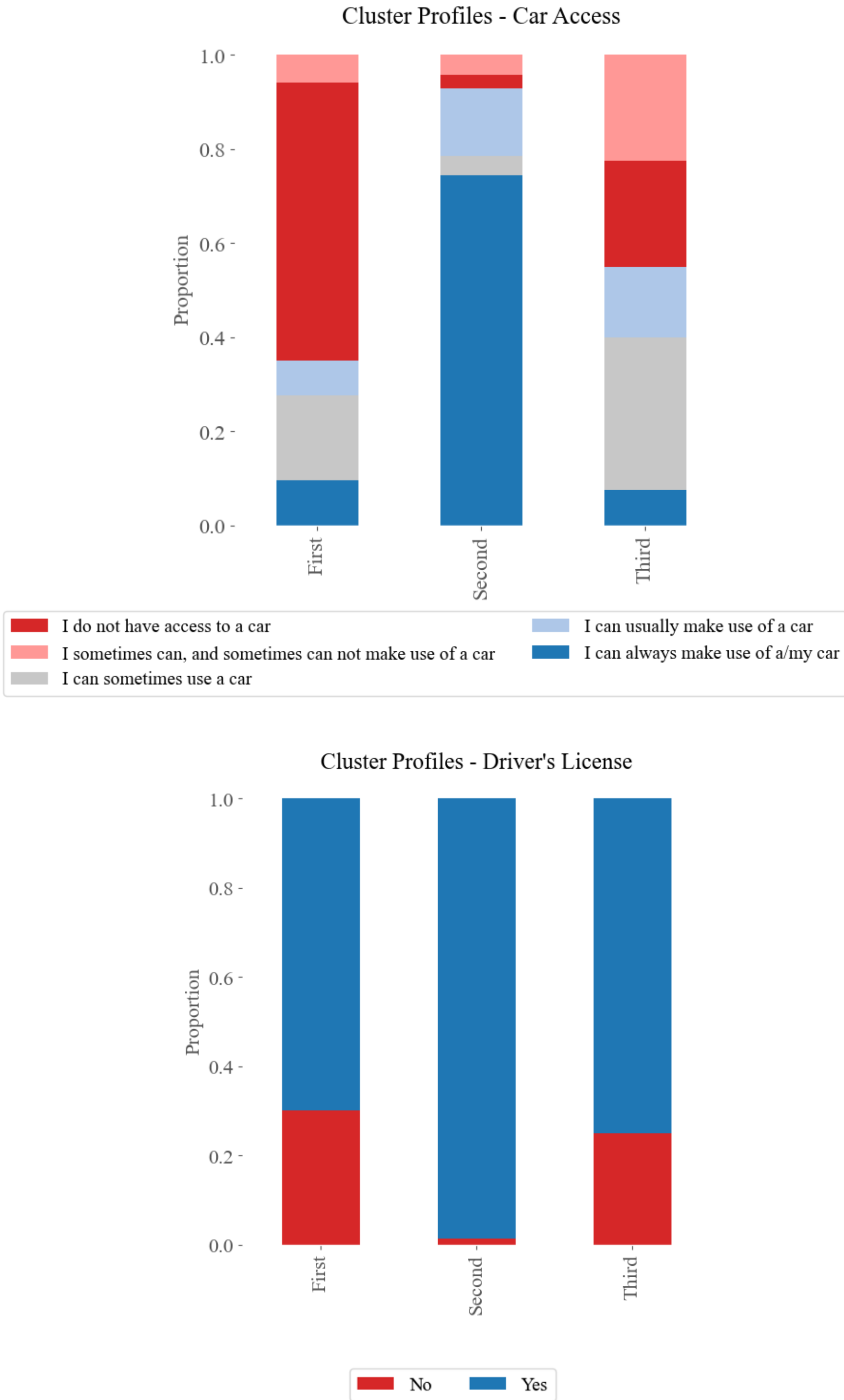


Figure 5.30: Car Access and Driver's License possession among clusters

Regarding the household composition, 60% of the first cluster lives with a partner, while the same variable represents 20% of the second cluster and around 95% of the third cluster. In addition, most participants in the first and second clusters do not have children, the opposite of the second cluster. The results are presented in Figure 5.31.

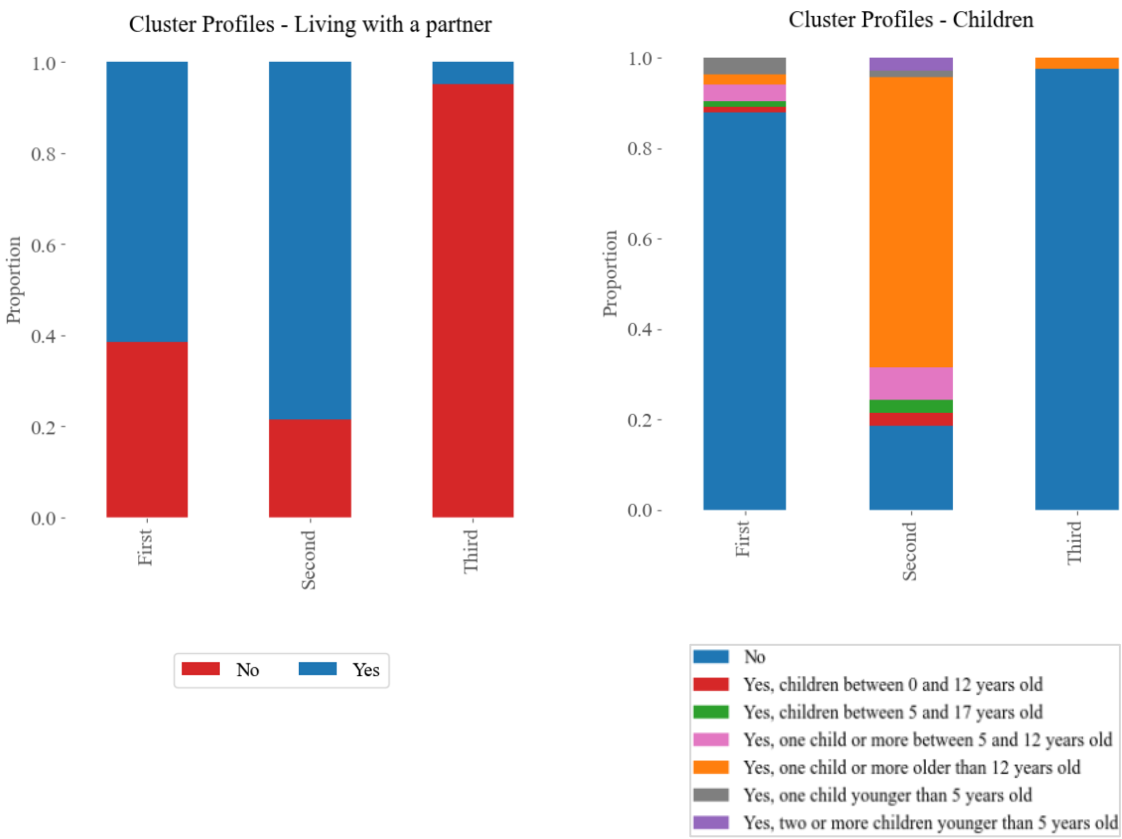


Figure 5.31: Household composition among clusters

The education level among all clusters is quite diverse. Most participants of the first cluster hold a WO master, and most participants of the second cluster have an HBO bachelor and a WO bachelor, in the case of the third cluster. The latter does not include a Ph.D. percentage. The income range of the first cluster is very diverse but predominantly from 43.500 to 65.000. However, the second cluster is similar, with significant proportions of higher salaries. The third cluster presents the lower income range of less than 22.000. The results are shown in Figure 5.32.

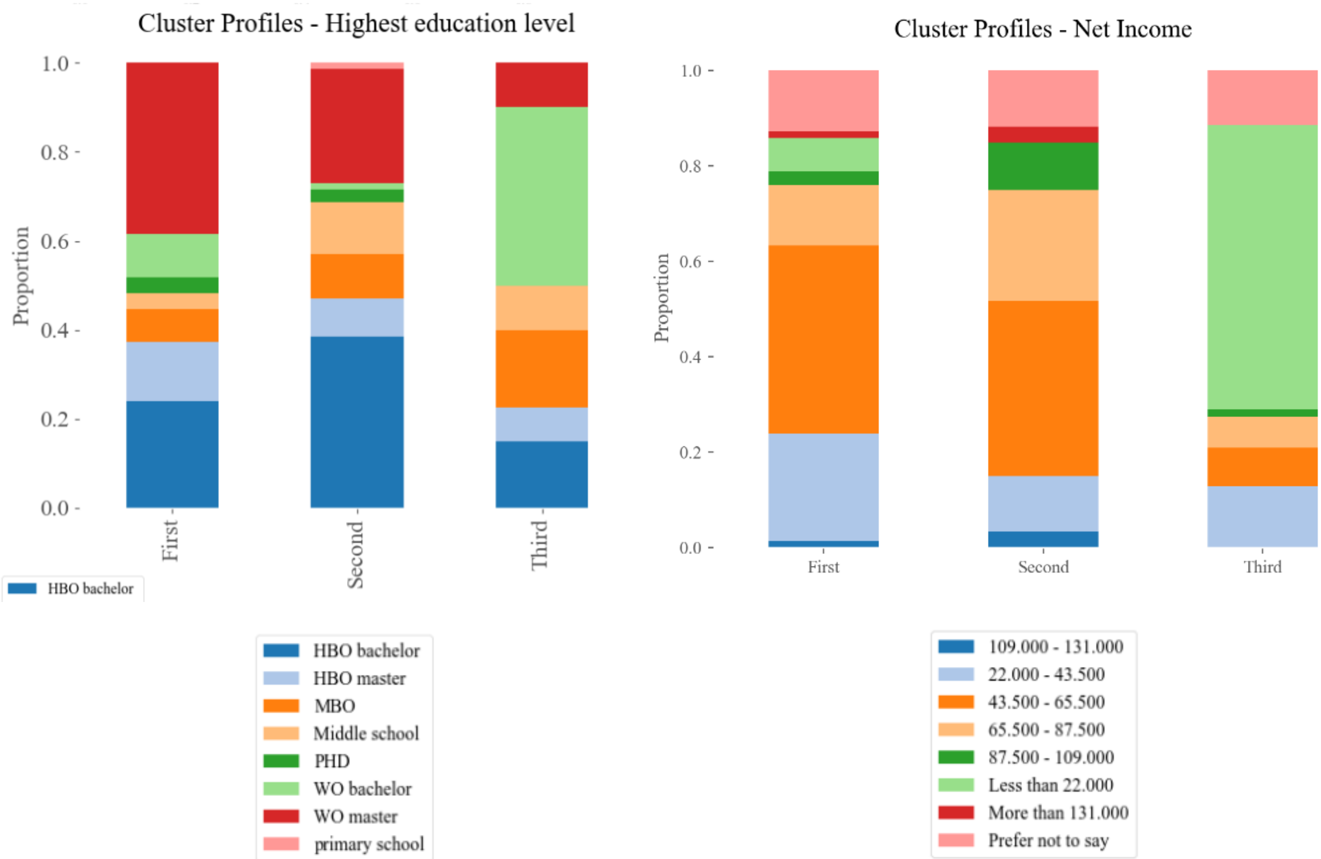


Figure 5.32: Education level and net income among clusters

Finally, Figure 5.33 presents the proportion of the country of origin of participants in each cluster. The first cluster shows mostly foreign participants where most of them from another European country and secondly from South America. The second cluster is 80% represented by Dutch, and the third cluster is also diverse, with around 60% Dutch participants.

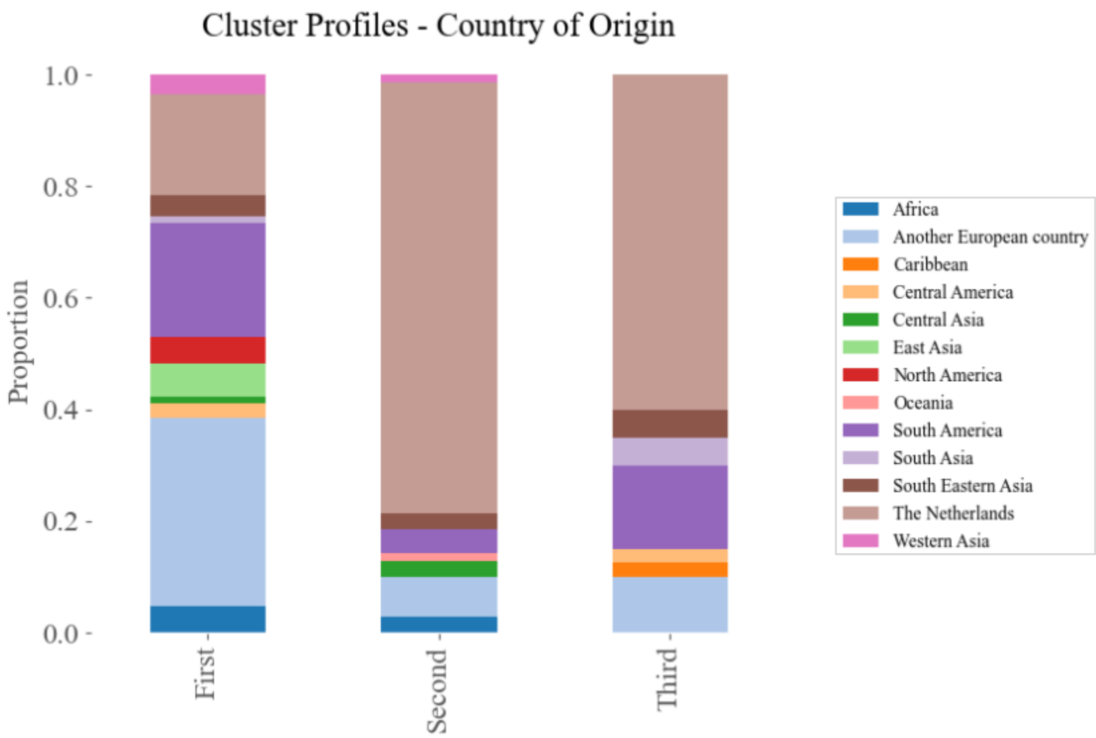


Figure 5.33: Country of Origin among clusters

In brief, the three clusters can be summarized as foreign women adults with an average income, middle-aged Dutch mothers with a higher income, and Dutchmen young adults with a low income. Foreign women adults present a higher proportion of mismatches by car to grocery stores.

5.5 Person-Based Features Identification

This section focuses on investigating the occurrence of mismatches among women and identifying the most influential characteristics associated with these occurrences. The analysis primarily focuses on three prevalent mismatches: accessing entertainment activities by car, by walking and reaching grocery stores by car. These three analyses have been selected due to their highest proportion of mismatches among women. These investigations are referred to as the First, Second and Third Logistic Regression analyses, respectively.

The initial step involves filtering the variables based on the transportation mode under investigation. For example, variables such as car access and driver's license are included only in the first and third analyses. The filtering process is presented in detail in the Appendix E.1.

The next step involves investigating the presence of multicollinearity among variables with a correlation above 0.7. In each analysis, certain variables such as 'No children' and 'One child or more older than 12 years old' exhibit a negative correlation, while variables like 'Income: More than €131,000' and 'Origin: Central America' show a positive correlation. To address multicollinearity, the analysis excludes the variable with the lower absolute correlation with the target variable. The detailed process is provided in the Appendix E.2.

The final feature selection for each analysis, along with the corresponding feature selection technique and the number of features, is presented in Table 5.7. These selections are based on the best model performance achieved.

Table 5.7: Feature Selection analysis for each Logistic Regression

Regression Analysis	Feature Selection Technique	Amount of Features	Features selected
1st Logistic Regression – Mismatch Entertainment by car	RFE	6	- One child or more between 5 and 12 years old - Neither safe nor unsafe - MBO (Secondary vocational education) - I can usually make use of a car - I car always make use of a/my car
2nd Logistic Regression – Mismatch Entertainment walking	X² squared	4	- Age:45-56 - Safe - Very Unsafe - MBO (Secondary vocational education)
3rd Logistic Regression – Mismatch Grocery stores by car	X² squared	7	- Age:56-65 - Neither safe nor unsafe - Very safe - I do not have access to a car - I sometimes can, and sometimes cannot make use of a car - I can usually make use of a car - I can always make use of a/my car

Finally, the model performance and features importance analysis are presented for each logistic regression investigation.

5.5.1 First Logistic Regression – Mismatches to entertainment activities by car

The evaluation metrics of this model's performance are presented in Table 5.8. The confusion matrix shows that out of 36 instances, 19 were correctly predicted as negative, and four were correctly predicted as positive. The evaluation phase considers the test data, therefore 20-30% of the total amount of instances. For this reason, there are 36 instances and not 200, the total amount of answers collected. The evaluation of the other Logistic Regression analyses follow the same logic.

There were nine false negatives (actual positive incorrectly predicted as negative) and four false positives (actual negative incorrectly predicted as positive). The model's accuracy is 0.64, indicating that it correctly predicted the outcome for approximately 64% of the instances. The precision of 0.5 suggests that when the model predicts a positive outcome, it is correct 50% of the time. The recall value of 0.31 indicates that the model correctly identifies only 31% of the actual positive instances. In addition, the pseudo-R-squared value of 0.187 represents the model's goodness of fit or explanatory power, indicating that approximately 16.3% of the variation in the outcome can be explained by the independent variables included in the model.

Overall, the model shows moderate performance with an accuracy of 0.64 and a precision of 0.5. However, the recall is relatively low at 0.31, suggesting that the model may have difficulty identifying positive cases. The pseudo-R-squared value indicates a modest explanatory power, indicating that the included independent variables explain a portion of the outcome variation. An AUC of 0.74 indicates reasonably moderate discrimination ability. The correspondent ROC Curve is presented in the Appendix E.3.

Table 5.8: Model Evaluation Metrics – First Logistic Regression

		Actual Values	
		Negative	Positive
Predicted Values	Negative	19	4
	Positive	9	4

Train/Test: 70/30; Accuracy: 0.64, Precision: 0.5, Recall: 0.31, Pseudo R^2: 0.163, AUC:0.74

When analyzing the model's features, most independent variables exhibit significant associations with the outcome variable, indicating their potential impact on the odds of the outcome occurrence, as presented in Table 5.9. Notably, perceiving oneself as neither safe nor unsafe demonstrates a strong positive association, with a significant p-value of 0.0236 and an odds ratio of 4.1023. Additionally, the variables "MBO (Secondary vocational education)" and "I can always make use of a/my car" exhibit significant associations, although with opposite directions of impact, as reflected by their respective p-values of 0.0202 and 0.0023, and odds ratios of 0.3825 and 0.6433.

It is worth noting that the variable "One child or more between 5 and 12 years old" shows a p-value of 0.1294, suggesting a lack of statistical significance in its association with the outcome variable. It indicates that one or more children in that age range may not significantly impact the odds of the outcome occurring.

Table 5.9: Variables' evaluation metrics – First Logistic Regression

Target Variable	Independent Variables	P-value	Odd-Ratio
Mismatches to entertainment activities by car	One child or more between 5 and 12 years old	0.1394	2.0516
	Neither safe nor unsafe	0.0236	4.1023
	MBO (Secondary vocational education)	0.0202	0.3825
	I can usually make use of a car	0.0345	0.6999
	I can always make use of a/my car	0.0023	0.6433

Therefore, Figure 5.34 presents the Odds ratio of the variables with significant influence on the model.

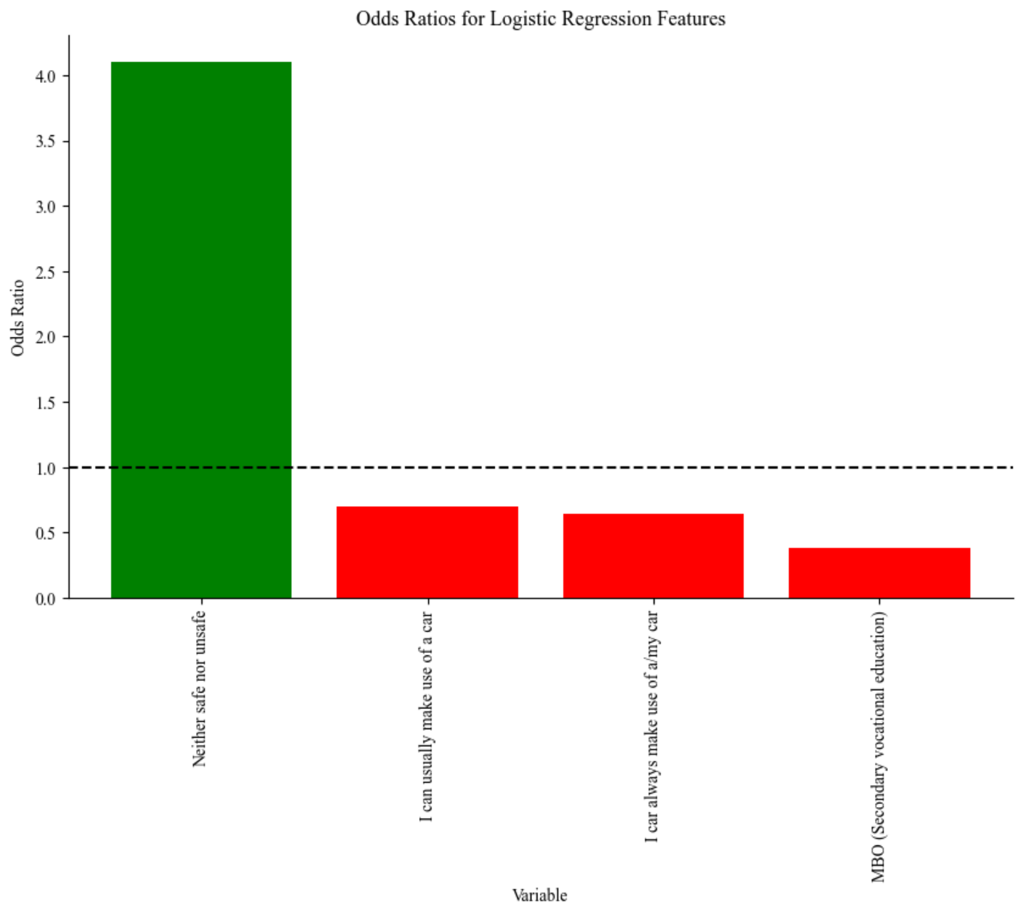


Figure 5.34: Odds Ratios of Significant Features - 1st Regression

5.5.2 Second Logistic Regression– Mismatches to entertainment activities walking

The performance of this model is presented in Table 5.10. It includes Confusion Matrix values, accuracy, precision, recall, Pseudo R^2 and AUC Score. The corresponding ROC Curve is presented in the Appendix E.3.

Table 5.10 - Model Evaluation Metrics – First Logistic Regression

		Actual Values	
		Negative	Positive
Predicted Values	Negative	21	1
	Positive	2	0

Train/Test: 80/20; Accuracy: 0.875, Precision: 0, Recall: 0, Pseudo R^2: -0.018, AUC score: 0.74

The model evaluation results indicate that the model achieved an accuracy of 0.875, meaning that 87.5% of the predictions matched the actual values. However, the precision and recall values are both 0 since the model did not correctly identify any positive instances. The pseudo-R^2 value of -0.018 indicates poor explanatory power, implying that the model may not fit the data well. The AUC of 0.74 contradicts the previous metrics. Overall, the model's performance is characterized by high accuracy but poor precision, recall, and explanatory power, emphasizing the need for further analysis and potential model improvements.

When investigating the variable's influence in the model, the logistic regression analysis results presented in Table 5.11 suggest that several independent variables have varying significance levels concerning the target variable, "Mismatches to entertainment activities walking." Age 46-55 shows a statistically significant relationship, with a 48% decrease in the odds of mismatches for each unit increase in age.

Table 5.11: Variables' evaluation metrics – Second Logistic Regression

Target Variable	Independent Variables	P-value	Odd-Ratio
Mismatches to entertainment activities walking	Age 46-55	0.0314	0.48
	Very Unsafe	0.4199	2.78
	Very safe	0.0009	0.48
	MBO (secondary vocational school)	0.4946	1.59

In addition to the previous summary, it is essential to note that the analysis reveals a clear relationship between the perceived safety levels and the occurrence of mismatches in entertainment activities when walking. The odds ratio of 0.48 for the "Very Safe" category suggests that individuals who perceive the environment as very safe have lower odds of experiencing mismatches.

However, the "MBO (secondary vocational school)" and 'Very Unsafe' variables do not demonstrate statistical significance, suggesting a lack of association with mismatches. Figure 5.35 presents the variables with statistical significance and their Odds Ratios.

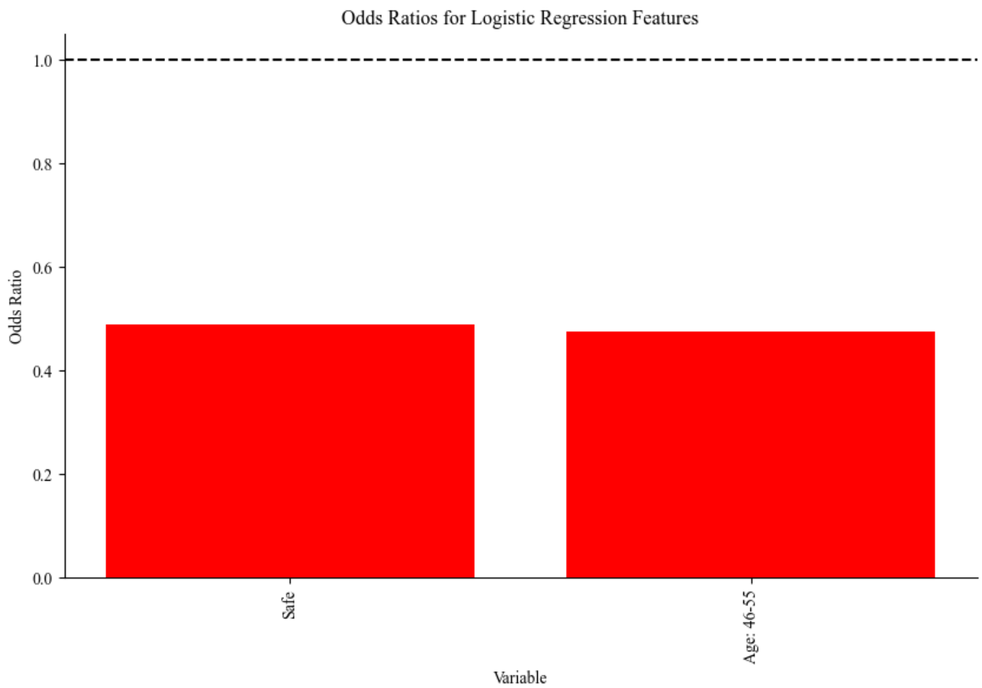


Figure 5.35: Odds Ratios of Significant Features – 2nd Regression

While the model's performance may not be optimal regarding precision, recall, and other metrics, analyzing the logistic regression coefficients can provide insights into the relationships between the independent and target variables.

5.5.3 Third Logistic Regression – Mismatches to grocery stores by car

First, the model performance is presented. Table 5.12 describes the Confusion Matrix values, accuracy, precision, recall, Pseudo R^2 and AUC Score. The correspondent ROC Curve is presented in the Appendix E.3.

Table 5.12: Model Evaluation Metrics – Third Logistic Regression

		Actual Values	
		Negative	Positive
Predicted Values	Negative	9	2
	Positive	6	7

Train/Test: 80/20, Accuracy: 0.67, Precision: 0.72, Recall: 0.54, Pseudo R^2: 0.289, AUC Score: 0.73

The model achieved an accuracy of 0.67, indicating that 67% of the predictions matched the actual values. The precision of 0.72 suggests that 72% of the optimistic predictions were correct. The recall rate 0.54 implies that only 54% of the actual positive instances were identified correctly. The pseudo-R^2 value of 0.289 indicates a moderate level of explanatory power in the model, explaining around 28.9% of the variance in the response variable. The model demonstrates moderate accuracy and precision but struggles to capture all the positive instances, resulting in a relatively lower recall rate. The AUC (Area Under the Curve) value of 0.73 suggests that the model's ability to discriminate between positive and negative instances is relatively good.

Furthermore, the feature's influence on the model is investigated. The variables' P-value and Odd-Ratio are described in Table 5.13.

Table 5.13: Variables' evaluation metrics – Third Logistic Regression

Target Variable	Independent Variables	P-value	Odd-Ratio
Mismatches to grocery stores by car	Age 56-65	0.7154	1.03
	Neither safe nor unsafe	0.0347	3.02
	Very safe	0.4714	0.73
	I do not have access to a car	0.0299	2.85
	I sometimes can, and sometimes cannot make use of a car	0.0643	0.51
	I can usually make use of a car	0.0693	0.61
	I can always make use of a/my car	0.0019	0.21

Among the independent variables, feeling neither safe nor unsafe and not having access to a car show statistically significant associations with mismatches to grocery stores. Individuals who feel neither safe nor unsafe are approximately three times more likely to experience mismatches, while those without car access are around 2.85 times more likely to have mismatches. Additionally, always having car access is strongly associated with a lower likelihood of mismatches, with individuals in this category being significantly less likely to experience mismatches. Other variables, such as age, feeling very safe, and inconsistent car use, do not show significant associations with mismatches to grocery stores by car. Figure 5.36 presents the variables with statistical significance and their Odds Ratios.

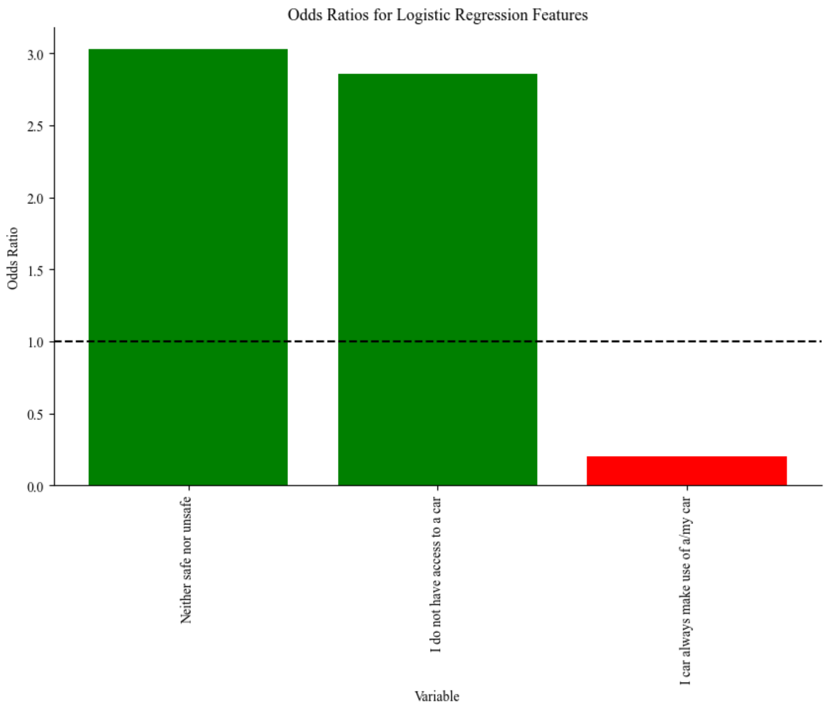


Figure 5.36: Odds Ratios of Significant Features – 3rd Regression

6. Discussion

This chapter provides a comprehensive discussion and interpretation of the research findings while presenting the limitations pertinent to this project. This section interprets the findings in four areas: Perceived Accessibility, Mismatches Identification, Urban Groups Identification, and Person-based Features Identification. The discussion does not include the Spatial Accessibility results, as they are primarily utilized for comparison purposes rather than directly addressing the research questions of this study. Subsequently, based on the findings, the limitation section elucidates the assumptions that can be made and those that should be approached with caution.

6.1 Perceived Accessibility

The main findings from the perceived data analysis show that women perceive safety at night at different levels than men. Regarding the transport modes walking, cycling, and public transport, the safety range among women that includes ‘safe’ and ‘very safe’ perception is lower than for men. In contrast, the unsafety range that includes ‘unsafe’ and ‘very unsafe’ options among women represents a higher proportion than men's perception. A distinct outcome emerges when examining safety perceptions at night concerning car travel, with women exhibiting lower feelings of insecurity compared to men. This noteworthy rise in safety concerns among men, as opposed to women, when opting for cars as a mode of transportation, constitutes a novel discovery not previously discussed in this literature review.

The survey data also shows that women have less access to a car than men. In addition, when ranking valuable aspects of choosing a transport mode, men and women present similar results. Time is considered the most critical aspect when choosing a transport mode, and money and sustainability are the least important. On the contrary, women and men disagree on safety and comfort ranking positions. While women see safety as the second most crucial aspect, men consider it comfort. These results align with the literature review, which identifies safety (Priya Uteng et al., 2019; Tiznado-Aitken et al., 2020) and car access (Havet et al., 2021; Priya Uteng, 2021) as one of the most critical aspects impacting mobility and accessibility of women.

6.2 Mismatches identification

The analysis reveals that regardless of gender, there are fewer mismatches where cycling is the transport mode and more mismatches where cars are the transport mode. Notably, women exhibit a higher proportion of mismatches specifically associated with cars. This result may be related to the lack of car access among women and it raises a hypothesis that this lack of car access highly impacts their accessibility perception by car. This finding is supported by existing literature highlighting the impact of car access on the mobility of women (Havet et al., 2021; Priya Uteng, 2021).

Furthermore, minor gender-related discrepancies are evident in the mismatch ratios for walking, cycling, and public transport use during nighttime entertainment. As described previously, women feel more unsafe in these transport modes at night. Surprisingly, this contrast becomes less pronounced when evaluating the number of mismatches associated with nighttime entertainment activities involving the same transportation options. This pattern gives rise to a hypothesis: although women commonly experience heightened safety concerns, this alone might not be compelling enough to lead them to consider these transport alternatives as unfeasible or inconvenient. This phenomenon could also be labeled as the 'foreign effect'.

Ceccato and Loukaitou-Sideris (2022) underscore the importance of the urban and cultural context in analyzing transit safety perceptions. They suggest that fear among women is more pronounced in the Global South compared to the Global North. In this present study, 50% of international participants are South Americans, representing around 18% of the total sample. Consequently, their outlook could potentially impact these findings.

The 'foreign effect' hypothesis posits that the perception of nighttime walking or cycling as unsafe might be shaped by past experiences and background of growing up in an environment perceived as generally insecure. Paradoxically, this perception contrasts with the Netherlands' reality, where these transportation modes are commonly used without adverse incidents. Thus, the perception of unsafety does not precipitate a decrease in the use of these modes.

Furthermore, when examining participants without young children, women generally experience a higher proportion of mismatches, whereas men with young children encounter more mismatches. The gender disparity becomes particularly pronounced when considering public transport trips to schools/daycare, where men face significantly more mismatches than women. This implies that fathers may have barriers to using public transport for school or daycare transportation. The higher concentration of mismatches among fathers of young children to school and grocery stores might be related to the fact that women shoulder more household responsibilities in childcare and maintenance tasks (Havet et al., 2021; Lo & Houston, 2018), as presented in the literature review. Thus, these activities might be generally perceived as inconvenient among fathers.

Additionally, women consistently demonstrate higher ratios of strong mismatches, indicating their perception of certain modes as impossible to use in areas with accessible amenities. In contrast, men exhibit higher proportions of moderate mismatches. This discrepancy between men and women is particularly evident among parents of young children, suggesting that the presence of young children in a household influences accessibility perception differently for men and women. However, it is crucial to exercise caution when interpreting these findings due to the limited sample size of parents considered in this research, described in more detail in Section 9.2.

6.3 Urban Groups Identification

This analysis aims to identify distinct urban groups based on personal characteristics and the occurrence of mismatch types, explicitly focusing on mismatches related to grocery stores and entertainment activities at night. Notably, the urban group of women adults aged 36-45, predominantly foreigners, with an average salary, living with a partner and no children, and lacking access to a car, exhibits the highest proportion of mismatch types when using a vehicle. This observation leads to a hypothesis suggesting that this particular group is more vulnerable to experiencing mismatches associated with car usage. Interestingly, although the presence of walking mismatches is minimal across all groups, it is slightly less prevalent among this group of women adults. However, it is important to acknowledge that the results may be influenced by the lack of representativeness in the research sample, as discussed in section 5.2.

6.4 Person-based features identification

The findings reveal that the safety perception of 'neither safe nor unsafe' is significant in driving to entertainment facilities and grocery stores. Participants with this neutral safety perception are four times more likely to experience a mismatch in accessing entertainment activities by car and three times more likely to experience a mismatch in accessing grocery stores by car. As this safety perception is neutral, this research assumes that the perception of safety or unsafety does not influence the perception of convenience or the possibility of using a car for grocery stores and entertainment activities.

Furthermore, both mismatches related to driving exhibit strong significance concerning car access variables. In the case of entertainment activities, having constant or regular access to a car decreases the likelihood of experiencing a mismatch. Additionally, participants with an MBO education level have lower chances of encountering this mismatch. The negative impact of education level on this type of mismatch has not been discussed in the existing literature, thus necessitating further investigation to explore possible connections between education level, accessibility levels, and car access. Additionally, it's essential to take into account the potential impact of the data sample's representativeness on these findings. For instance, there's a common correlation between higher education levels and elevated income, and car accessibility. Nonetheless, this correlation might not hold for foreigners, as other barriers, such as possessing a local driver's license and navigating distinct transit regulations, could influence car access despite education levels.

When investigating the mismatches of walking to entertainment activities at night, two variables demonstrate higher significance in the model: safety and age. Specifically, women who feel safe have lower chances of experiencing this mismatch. One intriguing finding is that unsafety feeling does not positively impact the occurrence of mismatches. Thus, it reinforces the hypothesis that women might not perceive this transport mode as not possible or inconvenient even with a feeling of unsafety.

Furthermore, women between the ages of 45 and 56 exhibit a decreased likelihood of experiencing such mismatches. While the literature review underscores age's influence on accessibility perception (Ryan et al., 2019), it remains uncertain whether this age bracket heightens the propensity to view walking as a viable transportation mode. One plausible hypothesis is that women within this age group are less prone to mismatches due to several factors. Primarily, their engagement in nighttime activities is lower in comparison to young adults (students). Additionally, they might reside in residential zones and partake in entertainment outings less frequently.

6.5 Limitations

After presenting the results and their interpretation, it is crucial to address the limitations of this research. These limitations can be categorized into sample size, representativity, methodology, and scope.

6.5.1 Sample Size

A limitation of this research is the small sample size obtained from the survey application. While 200 responses are reasonable, a larger sample size would enhance reliability and potentially yield new insights. A critical limitation is the small number of 23 parents of young children in the sample, representing approximately 12% of participants. Consequently, results for this group are inconclusive. Moreover, small sample size can compromise the performance of Logistic Regression and Cluster analysis.

6.5.2 Sample Representativity

The perceived gathered data lacks representativity compared to the MDHR population. Most participants are highly educated women, with about half of them being foreigners. Therefore, results are limited to this specific demographic and cannot be generalized to the entire geographical population. Additionally, this research shows that the perception of accessibility components, such as safety, varies across spatial contexts. Thus, the analysis results are biased towards the predominant area where the data was collected, specifically Rotterdam.

6.5.3 Methodology

This study employs various methodologies to analyze the impact of gender and personal characteristics on accessibility levels. A limitation of the spatial analysis is the definition of thresholds and their sensitivity to results. Normative thresholds based on average travel times by residents in the Netherlands were used, with assumptions made when information was lacking. The research also calculated average driving speeds for different activity types.

Furthermore, the spatial analysis methodology relies on mapping amenity locations using Python tools and OpenStreetMap. A drawback is its dependence on the accuracy of contributor-provided information, which may lead to incorrect or outdated representation of some amenity types. Moreover, defining tags for amenities like "grocery stores," "primary schools," and "daycare centers" presents challenges, as there are no unique tags for each.

Furthermore, this research assumes that participants commonly engage in nighttime entertainment activities and travel to schools and daycare centers. However, it is essential to acknowledge that this assumption may not accurately represent all participants' experiences. As a result, the study is limited to this specific behavior. Participants who, for instance, prefer individual home care for their children or do not frequent restaurants, clubs, bars, cinemas, or theaters at night may be not represented in this analysis.

6.5.4 Scope

This research evaluates three types of POIs based on the literature review. However, including other amenity types could provide new insights into gender-based accessibility levels. Additionally, this research primarily investigates overestimation mismatches, but exploring the causes and gender distribution of underestimation mismatches could shed light on the differences in accessibility perceptions of women and men.

6.6 Future Research Directions

As discussed earlier, the findings of this research have generated new hypotheses. The lack of access to cars among (foreign) women might negatively impact their accessibility perception by this transport mode. Additionally, although women feel less safe than men while walking, cycling, or using public transport at night, they still perceive it as a possible option due to their previous experiences according to the city context. Furthermore, fathers of young children might present more barriers to accessing different activity types but mainly daycare facilities and primary schools because they shoulder fewer child or household tasks compared to women. Finally, women 45-56 years old presents fewer barriers to going to entertainment activities at night because they go less to these activities compared to younger women. These hypotheses are still inconclusive considering the scope of this research and its limitations. Therefore, future research directions are recommended.

Firstly, it is essential to investigate the reasons behind the lack of car access among women, particularly among foreign individuals, to gain a comprehensive understanding of its implications on accessibility levels. Also, it is valuable to investigate if the unsafety feeling at night and the few mismatches to this activity type are mostly caused by the 'foreign factor' previously described.

Additionally, a more in-depth analysis is recommended to explore mobility barriers faced by fathers of young children, especially concerning their use of public transport. Understanding whether these barriers are related to time constraints, comfort preferences, or other factors would be valuable. Furthermore, conducting a detailed comparison between men and women who are parents of young children will provide insights into whether the disparity in accessibility levels is accentuated within this demographic and identify the underlying reasons.

In addition to the above-mentioned studies, it is recommended to explore other research scopes that were not the primary focus of this study. For instance, evaluating underestimation mismatches, in addition to overestimation mismatches, would be beneficial. Also, understanding how the safety perception varies among cities and what are main causes to it can potentially generate valuable results.

By delving into these suggested areas of investigation, it can be gained a more comprehensive understanding of the complexities surrounding accessibility and safety perceptions. These insights will contribute to the development of more effective policies and interventions to enhance the equity of urban mobility.

For instance, creating policies to facilitate access to cars among foreign women. They might face challenges in accessing private vehicles due to factors such as limited familiarity with local regulations, financial constraints, or lack of documentation. To address this, a policy could be designed to provide tailored support for foreign women to obtain local driver's licenses or access car-sharing programs.

Based on the outcomes of this research, there appears to be a compelling rationale for instituting a policy aimed at bolstering public transport utilization among men. A strategic approach could involve implementing flexible fare structures, strategically tailored to accommodate peak hours or popular routes for male commuters.

7. Conclusion

This research aims to answer the main question ‘How do personal characteristics, mainly gender, can impact accessibility levels?’. Based on a combination of quantitative and case study approaches, it answers this question by investigating the impact of personal characteristics at different levels, represented by sub-research questions. From this analysis, there are several possible conclusions. New questions also present potential directions for further investigation. Thus, this chapter aims to provide summary comments and suggest some potential study issues for the future.

The first sub-research question, ‘How can perceived accessibility differ from spatial accessibility?’ is answered by comparing spatial accessibility analysis and perceived accessibility analysis. From this comparison, this research identifies mismatches from two perspectives: quantity (cumulative analysis) and qualitative (mismatch types). It is determined that the most significant mismatch type, also called strong mismatches, is more evident when using cars by women. It is also identified that fathers of young children present more mismatches by traveling to schools or daycare, regardless of the transport mode. Finally, it is discovered that women generally show a higher proportion of strong and slight mismatches than men. On the contrary, men experience more moderate mismatches than women.

This study's second sub-research question investigates ‘What urban groups present an accessibility perception that differs the most from spatially calculated accessibility?’. Hence, it conducts a clustering analysis to identify the formation of urban groups and the proportion of mismatches for each group. This analysis shows that the most evident mismatches are related to cars as transport modes and are experienced mainly by foreign women adults. In fewer proportions, young adults present low to moderate mismatches to the vehicle, and senior adults present a slightly higher presence of mismatches to public transport compared to other clusters.

Furthermore, the third sub-research question studies ‘What are the most influential personal characteristics that impact accessibility perception of women?’. A binary logistic regression analysis is conducted to answer this question, and the variable's significance and odds ratio is evaluated. This study identifies that car access is a critical aspect that influences the occurrence of mismatches among women by car. In contrast, the perception of safety or unsafety does not impact these mismatches since the most influential variable is a neutral answer of ‘neither safe nor unsafe.’ Additionally, when investigating the event of mismatches among women by walking to entertainment activities, it is identified that safety perception and age 45-56 appear to be the most relevant variables of this study, and both reduce the chances of presenting a mismatch.

In brief, this study contributes significant value to the research area by employing a unique combination of techniques to investigate person-characteristics' impact on accessibility levels. By integrating spatial and perceived accessibility analyses, clustering analysis, and logistic regression, the study identified mismatch levels, urban groups, and factors contributing to differing perceptions of accessibility compared to spatial analysis. This understanding is vital in identifying barriers to using transportation modes to reach points of interest, ultimately supporting the development of new transport policies and more equitable transport systems. These policy measures could encompass efforts to enhance car accessibility rates among foreign women and encourage greater utilization of public transport among men.

To arrive at more distinct conclusions and subsequently develop transport policy initiatives that align, this study proposes several avenues for further investigation.

Primarily, conducting an in-depth analysis of the limitations of car access, especially among foreign women, is essential. Additionally, exploring the potential correlation between women's safety perceptions and their diverse backgrounds and countries of origin is highly recommended.

Moreover, comprehending the disparities in accessibility perception between parents of young children and participants without young children warrants exploration. Further, delving into the frequency of mismatches among fathers of young children, who seem to exhibit a higher incidence of mismatches, is advised.

Furthermore, extending the analysis to encompass not only overestimation mismatches but also underestimation mismatches could reveal novel insights. By addressing these research recommendations in the future, more definitive outcomes can be achieved. These findings can serve as valuable pillars for crafting more inclusive transport policies and establishing a more equitable transportation system.

Resources

Agarwal, R. (2022, March 12). *The 5 Feature Selection Algorithms every Data Scientist should know*. Medium. https://towardsdatascience.com/the-5-feature-selection-algorithms-every-data-scientist-need-to-know-3a6b566efd2

Alkharusi, H. (2012). Categorical Variables in Regression Analysis: A Comparison of Dummy and Effect Coding. *International Journal of Education*, *4*, 202–210. https://doi.org/10.5296/ije.v4i2.1962

Altman, D. G., & Bland, J. M. (1994). Statistics Notes: Quartiles, quintiles, centiles, and other quantiles. *BMJ*, *309*(6960), 996–996. https://doi.org/10.1136/bmj.309.6960.996

Apte, M., Agarwadkar, Y., Azmi, S., & Inamdar, A. (2013). *Understanding Grids and Effectiveness of Hexagonal Grid in Spatial Domain*.

ArcGIS Pro. (n.d.). *Why hexagons?—ArcGIS Pro | Documentation*. Retrieved 26 July 2023, from https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/h-whyhexagons.htm

Axis Maps. (n.d.). *The Basics of Data Classification*. Retrieved 13 June 2023, from https://www.axismaps.com//guide/data-classification

Barbiero, P., Squillero, G., & Tonda, A. (2022). Predictable Features Elimination: An Unsupervised Approach to Feature Selection. In G. Nicosia, V. Ojha, E. La Malfa, G. La Malfa, G. Jansen, P. M. Pardalos, G. Giuffrida, & R. Umeton (Eds.), *Machine Learning, Optimization, and Data Science* (Vol. 13163, pp. 399–412). Springer International Publishing. https://doi.org/10.1007/978-3-030-95467-3_29

Barnes, J. C., & Forde, D. R. (2021). *The Encyclopedia of Research Methods in Criminology and Criminal Justice, 2 Volume Set*. John Wiley & Sons.

Berman, J. J. (2016). Chapter 4—Understanding Your Data. In J. J. Berman (Ed.), *Data Simplification* (pp. 135–187). Morgan Kaufmann. https://doi.org/10.1016/B978-0-12-803781-2.00004-7

Bhardwaj, A. (2020, May 27). *Silhouette Coefficient: Validating clustering techniques*. Medium. https://towardsdatascience.com/silhouette-coefficient-validating-clustering-techniques-e976bb81d10c

Bhat, C., Handy, S., Kockelman, K., Mahmassani, H., Chen, Q., Weston, L., & University of Texas at Austin. Center for Transportation Research. (2000). *Development of an Urban Accessibility Index: Literature Review* (Report No. TX-01/7-4938-1). https://rosap.ntl.bts.gov/view/dot/14882

Boeing, G. (2017). OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems*, *65*, 126–139. https://doi.org/10.1016/j.compenvurbsys.2017.05.004

Brinkhoff, T. (2023). *Netherlands: Administrative Division (Provinces and Municipalities)—Population Statistics, Charts and Map*. https://www.citypopulation.de/en/netherlands/admin/

CBS. (n.d.). *Where were people in the Netherlands born? - The Netherlands in numbers 2022 | CBS* [Webpagina]. Where were people in the Netherlands born? - The Netherlands in numbers 2022 | CBS. Retrieved 24 May 2023, from https://longreads.cbs.nl/the-netherlands-in-numbers-2022/where-were-people-in-the-netherlands-born

CBS. (2022, November 15). *Mobility trend; per stage, modes of travel, purposes of travel, age and sex* [Webpagina]. Statistics Netherlands. https://www.cbs.nl/en-gb/figures/detail/84755ENG?q=grocery

Chiang, I.-C. A., Jhangiani, R. S., & Price, P. C. (2015). *Understanding Null Hypothesis Testing*. https://opentextbc.ca/researchmethods/chapter/understanding-null-hypothesis-testing/

COE. (n.d.). *Rotterdam, Netherlands—Intercultural City—Intercultural cities programme—Www.coe.int*. Intercultural Cities Programme. Retrieved 24 May 2023, from https://www.coe.int/en/web/interculturalcities/rotterdam

Crowe, S., Cresswell, K., Robertson, A., Huby, G., Avery, A., & Sheikh, A. (2011). The case study approach. *BMC Medical Research Methodology*, *11*, 100. https://doi.org/10.1186/1471-2288-11-100

de Haas, M., & Hamersma, M. (2019). *Walking Facts* (p. 16). KiM | Netherlands Institute for Transport Policy Analysis. https://english.kimnet.nl/binaries/kimnet-english/documenten/publications/2019/10/3/walking-facts/KiM+e-book+Walking+Facts.pdf

Dulhunty, A. (2022). Disciplinary mobility and women's empowerment: A complicated connection. *Mobilities*, *0*(0), 1–15. https://doi.org/10.1080/17450101.2022.2088298

EEA. (n.d.). *Netherlands shapefile—European Environment Agency* [Data File]. Retrieved 25 May 2023, from https://www.eea.europa.eu/data-and-maps/data/eea-reference-grids-2/gis-files/netherlands-shapefile

Entzinger, H., & Engbersen, G. (2014). *Rotterdam: A Long-Time Port of Call and Home to Immigrants*.

Garcia-Dias, R., Vieira, S., Lopez Pinaya, W. H., & Mechelli, A. (2020). Clustering analysis. In *Machine Learning* (pp. 227–247). Elsevier. https://doi.org/10.1016/B978-0-12-815739-8.00013-4

Gemeente Den Haag. (2023, May 4). *The Hague in numbers*. https://www.denhaag.nl/en/in-the-city/introducing-the-hague/the-hague-in-numbers.htm

Geofabrik GmbH. (2018). *Geofabrik Download Server*. https://download.geofabrik.de/europe/netherlands.html

Geurs, K. T., & Van Wee, B. (2004). Accessibility evaluation of land-use and transport strategies: Review and research directions. *Journal of Transport Geography*, *12*(2), 127–140. https://doi.org/10.1016/j.jtrangeo.2003.10.005

Gholamy, A., Kreinovich, V., & Kosheleva, O. (2018). Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation. *Departmental Technical Reports (CS)*. https://scholarworks.utep.edu/cs_techrep/1209

Gil Solá, A., & Vilhelmson, B. (2022). To choose, or not to choose, a nearby activity option: Understanding the gendered role of proximity in urban settings. *Journal of Transport Geography*, *99*, 103301. https://doi.org/10.1016/j.jtrangeo.2022.103301

Griffin, M., Martino, R. J., LoSchiavo, C., Comer-Carruthers, C., Krause, K. D., Stults, C. B., & Halkitis, P. N. (2022). Ensuring survey research data integrity in the era of internet bots. *Quality & Quantity*, *56*(4), 2841–2852. https://doi.org/10.1007/s11135-021-01252-1

H3GEO. (n.d.). *Introduction | H3*. Retrieved 26 May 2023, from https://h3geo.org/docs

Hallberg, L. (2022, July 27). *Online Surveys and Data Collection Tools*. https://lifespan.ku.edu[current-page]

Handy, S. L., & Niemeier, D. A. (1997). Measuring Accessibility: An Exploration of Issues and Alternatives. *Environment and Planning A: Economy and Space*, *29*(7), 1175–1194. https://doi.org/10.1068/a291175

Harrell , F. E. (2015). *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. Springer International Publishing. https://doi.org/10.1007/978-3-319-19425-7

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *28*(1), 100–108. https://doi.org/10.2307/2346830

Havet, N., Bayart, C., & Bonnel, P. (2021). Why do Gender Differences in Daily Mobility Behaviours persist among workers? *Transportation Research Part A: Policy and Practice*, *145*, 34–48. https://doi.org/10.1016/j.tra.2020.12.016

Hu, H., Liu, J., Zhang, X., & Fang, M. (2023). An Effective and Adaptable K-means Algorithm for Big Data Cluster Analysis. *Pattern Recognition*, *139*, 109404. https://doi.org/10.1016/j.patcog.2023.109404

Huang, Z. (1998). Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values. *Data Mining and Knowledge Discovery*, *2*(3), 283–304. https://doi.org/10.1023/A:1009769707641

Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, *622*, 178–210. https://doi.org/10.1016/j.ins.2022.11.139

Japkowicz, N., & Stephen, S. (2002). The class imbalance problem: A systematic study. *Intelligent Data Analysis*, *6*(5), 429–449.

Keeler, C., & Emch, M. (2017). Geographic Information Systems (GIS) in Public Health. In S. R. Quah (Ed.), *International Encyclopedia of Public Health (Second Edition)* (pp. 253–255). Academic Press. https://doi.org/10.1016/B978-0-12-803678-5.00174-0

Kulkarni, A., Chong, D., & Batarseh, F. A. (2020). 5—Foundations of data imbalance and solutions for a data democracy. In F. A. Batarseh & R. Yang (Eds.), *Data Democracy* (pp. 83–106). Academic Press. https://doi.org/10.1016/B978-0-12-818366-3.00005-8

Levinson, D., & King, D. (2020). *Transport Access Manual: A Guide for Measuring Connection between People and Places*. https://ses.library.usyd.edu.au/handle/2123/23733

Ligterink, N. E. (2016). *On-road determination of average Dutch driving behaviour for vehicle emission*. TNO.

Lo, A. W.-T., & Houston, D. (2018). How do compact, accessible, and walkable communities promote gender equality in spatial behavior? *Journal of Transport Geography*, *68*, 42–54. https://doi.org/10.1016/j.jtrangeo.2018.02.009

Loorbach, D., Schwanen, T., Doody, B. J., Arnfalk, P., Langeland, O., & Farstad, E. (2021). Transition governance for just, sustainable urban mobility: An experimental approach from Rotterdam, the Netherlands. *Journal of Urban Mobility*, *1*, 100009. https://doi.org/10.1016/j.urbmob.2021.100009

Loukaitou-Sideris, A. (2020). A Gendered View of Mobility and Transport: Next Steps and Future Directions. In *Engendering Cities*. Routledge.

Mahendru, K. (2019, June 17). How to determine the optimal K for K-Means? *Analytics Vidhya*. https://medium.com/analytics-vidhya/how-to-determine-the-optimal-k-for-k-means-708505d204eb

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*.

Meinshausen, N., Meier, L., & Bühlmann, P. (2009). P-Values for High-Dimensional Regression. *Journal of the American Statistical Association*, *104*(488), 1671–1681. https://doi.org/10.1198/jasa.2009.tm08647

Nicoletti, L., Sirenko, M., & Trivik, V. (2022). *Disadvantaged communities have lower access to urban infrastructure*. *50*(3). https://journals.sagepub.com/doi/10.1177/23998083221131044

North, M. A. (2009). A Method for Implementing a Statistically Significant Number of Data Classes in the Jenks Algorithm. *2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery*, *1*, 35–38. https://doi.org/10.1109/FSKD.2009.319

OECD. (2016). *OECD Territorial Reviews: The Metropolitan Region of Rotterdam-The Hague, Netherlands*. OECD. https://doi.org/10.1787/9789264249387-en

OlegS. (n.d.). *Driving in the Netherlands*. Retrieved 26 May 2023, from https://autotraveler.ru/en/netherlands/

OpenStreetMap Wiki. (n.d.). *Map features—OpenStreetMap Wiki*. Retrieved 24 May 2023, from https://wiki.openstreetmap.org/wiki/Map_features#Entertainment,_Arts_&_Culture

Páez, A., Scott, D. M., & Morency, C. (2012). Measuring accessibility: Positive and normative implementations of various accessibility indicators. *Journal of Transport Geography*, *25*, 141–153. https://doi.org/10.1016/j.jtrangeo.2012.03.016

Porrazzo, A., Samson, C., & Freudendal-Pedersen, M. (2022). Gender and mobility planning: The influence of national culture on planning processes. *Case Studies on Transport Policy*, *10*(2), 1102–1109. https://doi.org/10.1016/j.cstp.2022.03.020

Priya Uteng, T. (2021). Chapter Two—Gender gaps in urban mobility and transport planning. In R. H. M. Pereira & G. Boisjoly (Eds.), *Advances in Transport Policy and Planning* (Vol. 8, pp. 33–69). Academic Press. https://doi.org/10.1016/bs.atpp.2021.07.004

Priya Uteng, T., Singh, Y. J., & Lam, T. (2019). 12—Safety and daily mobilities of urban women—Methodolgies to confront the policy of “invisibility”. In K. Lucas, K. Martens, F. Di Ciommo, & A. Dupont-Kieffer (Eds.), *Measuring Transport Equity* (pp. 187–202). Elsevier. https://doi.org/10.1016/B978-0-12-814818-1.00012-3

R5py. (2022). https://r5py.readthedocs.io/en/stable/

RET. (n.d.). *Zakelijke mobiliteitskaart—DeMobiliteitsManager van RET*. Retrieved 12 June 2023, from https://www.ret.nl/zakelijke-mobiliteit/demobiliteitsmanager.html

Rode, P., Floater, G., Thomopoulos, N., Docherty, J., Schwinger, P., Mahendra, A., & Fang, W. (2017). Accessibility in Cities: Transport and Urban Form. In G. Meyer & S. Shaheen (Eds.), *Disrupting Mobility: Impacts of Sharing Economy and Innovative Transportation on Cities* (pp. 239–273). Springer International Publishing. https://doi.org/10.1007/978-3-319-51602-8_15

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Ryan, J., & Pereira, R. H. M. (2021). What are we missing when we measure accessibility? Comparing calculated and self-reported accounts among older people. *Journal of Transport Geography*, *93*, 103086. https://doi.org/10.1016/j.jtrangeo.2021.103086

Ryan, J., Wretstrand, A., & Schmidt, S. M. (2019). Disparities in mobility among older people: Findings from a capability-based travel survey. *Transport Policy*, *79*, 177–192. https://doi.org/10.1016/j.tranpol.2019.04.016

Salas-Eljatib, C., Fuentes-Ramirez, A., Gregoire, T. G., Altamirano, A., & Yaitul, V. (2018). A study on the effects of unbalanced data when fitting logistic regression models in ecology. *Ecological Indicators*, *85*, 502–508. https://doi.org/10.1016/j.ecolind.2017.10.030

Sardana, N., Shekoohi, S., Cornett, E. M., & Kaye, A. D. (2023). Chapter 6—Qualitative and quantitative research methods. In A. D. Kaye, R. D. Urman, E. M. Cornett, & A. N. Edinoff (Eds.), *Substance Use and Addiction Research* (pp. 65–69). Academic Press. https://doi.org/10.1016/B978-0-323-98814-8.00008-1

Schiller, M., Awad, I., Buijse, N., Chantre, M., Huang, Y.-C., Jonitz, E., van den Brink, L., & van Dordrecht, L. (2023). Brokerage in urban networks on diversity and inclusion: The case of Rotterdam. *Cities*, *135*, 104219. https://doi.org/10.1016/j.cities.2023.104219

Scikit-learn. (n.d.). *Sklearn.feature_selection.RFE*. Scikit-Learn. Retrieved 5 June 2023, from https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.RFE.html

Sheard, J. (2018). Chapter 18—Quantitative data analysis. In K. Williamson & G. Johanson (Eds.), *Research Methods (Second Edition)* (pp. 429–452). Chandos Publishing. https://doi.org/10.1016/B978-0-08-102220-7.00018-2

Shi, C., Wei, B., Wei, S., Wang, W., Liu, H., & Liu, J. (2021). A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm. *EURASIP Journal on Wireless Communications and Networking*, *2021*(1), 31. https://doi.org/10.1186/s13638-021-01910-w

Southall, J. (2021). *Oxford LibGuides: Data and Statistics for Social Sciences: Data analysis tools & training*. https://libguides.bodleian.ox.ac.uk/ssdata/analysis

Stark, J., & Meschik, M. (2018). Women's everyday mobility: Frightening situations and their impacts on travel behaviour. *Transportation Research Part F: Traffic Psychology and Behaviour*, *54*, 311–323. https://doi.org/10.1016/j.trf.2018.02.017

Statista. (2022). *Netherlands: Largest cities 2022*. Statista. https://www.statista.com/statistics/993709/largest-cities-in-the-netherlands-by-number-of-inhabitants/

Storozuk, A., Ashley, M., Delage, V., & Maloney, E. A. (2020). Got Bots? Practical Recommendations to Protect Online Survey Data from Bot Attacks. *The Quantitative Methods for Psychology*, *16*(5), 472–481. https://doi.org/10.20982/tqmp.16.5.p472

Su, M. D., Lin, M.-C., & Wen, T. H. (2011). Spatial Mapping and Environmental Risk Identification. In J. O. Nriagu (Ed.), *Encyclopedia of Environmental Health* (pp. 228–235). Elsevier. https://doi.org/10.1016/B978-0-444-52272-6.00636-X

Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018). Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster. *IOP Conference Series: Materials Science and Engineering*, *336*(1), 012017. https://doi.org/10.1088/1757-899X/336/1/012017

Szumilas, M. (2010). Explaining Odds Ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, *19*(3), 227–229.

Tables of Cell Statistics Across Resolutions | H3. (n.d.). Retrieved 26 May 2023, from https://h3geo.org/docs/core-library/restable

The Doctoral Journey (Director). (2014, January 20). *Developing a Quantitative Research Plan: Choosing a Research Design*. https://www.youtube.com/watch?v=UywO2XuFuBA

Thrift, N. J., & Kitchin, R. (2009). *International encyclopedia of human geography*. Elsevier. http://www.sciencedirect.com/science/referenceworks/9780080449104

Tiznado-Aitken, I., Lucas, K., Muñoz, J. C., & Hurtubia, R. (2020). Understanding accessibility through public transport users' experiences: A mixed methods approach. *Journal of Transport Geography*, *88*, 102857. https://doi.org/10.1016/j.jtrangeo.2020.102857

UrbiStat S.r.l. (n.d.). *Demographic statisticsMunicipality of ROTTERDAM, population density, population, average age, families, foreigners*. Retrieved 9 June 2023, from https://ugeo.urbistat.com/AdminStat/en/nl/demografia/dati-sintesi/rotterdam/23055877/4

van de Laar, P., & van der Schoor, A. (2019). Rotterdam's Superdiversity from a Historical Perspective (1600–1980). In P. Scholten, M. Crul, & P. van de Laar (Eds.), *Coming to Terms with Superdiversity: The Case of Rotterdam* (pp. 21–55). Springer International Publishing. https://doi.org/10.1007/978-3-319-96041-8_2

van Mameren, A. (2023, June). *The education system in the Netherlands*. Expatica Netherlands. https://www.expatica.com/nl/education/children-education/dutch-education-system-100816/

van Smeden, M., Moons, K. G., de Groot, J. A., Collins, G. S., Altman, D. G., Eijkemans, M. J., & Reitsma, J. B. (2019). Sample size for binary logistic prediction models: Beyond events per variable criteria. *Statistical Methods in Medical Research*, *28*(8), 2455–2474. https://doi.org/10.1177/0962280218784726

Waterstaat, M. van I. en. (2017, October 23). *Mobility Report 2017 (In Dutch)—Document (research publication)—Netherlands Institute for Transport Policy Analysis* [Rapport]. Ministerie van Infrastructuur en Waterstaat. https://doi.org/10/23/mobility-report-2017

Wilcox, R. (2012). Chapter 4—Confidence Intervals in the One-Sample Case. In R. Wilcox (Ed.), *Introduction to Robust Estimation and Hypothesis Testing (Third Edition)* (pp. 103–136). Academic Press. https://doi.org/10.1016/B978-0-12-386983-8.00004-4

Wong, K. J. (2022, December 9). *7 Evaluation Metrics for Clustering Algorithms*. Medium. https://towardsdatascience.com/7-evaluation-metrics-for-clustering-algorithms-bdc537ff54d2

Wu, H., & Levinson, D. (2020). Unifying access. *Transportation Research Part D: Transport and Environment*, *83*, 102355. https://doi.org/10.1016/j.trd.2020.102355

Zach. (2021, January 22). How to Winsorize Data: Definition & Examples. *Statology*. https://www.statology.org/winsorize/

Zhang, M., Zhao, P., & Tong, X. (2022). Constructing women's immobility: Fear of violence and Women's constricted nocturnal travel behaviour. *Travel Behaviour and Society*, *26*, 178–192. https://doi.org/10.1016/j.tbs.2021.10.002

Zhang, P., Ma, W., & Qian, S. (2022). Cluster analysis of day-to-day traffic data in networks. *Transportation Research Part C: Emerging Technologies*, *144*, 103882. https://doi.org/10.1016/j.trc.2022.103882

Zibran, M. F. (2007). *CHI-Squared Test of Independence*.

Appendix

Appendix A.1
Cumulative Opportunities Metric Algorithm

```
tags_school = {'amenity': ['school','childcare','kindergarten']}
tags_groc = {'shop': ['greengrocer','supermarket']}
tags_ent = {'amenity': ['nightclub','restaurant','pub','cinema','theater']}
```

```
for i, centro in filtered_hexgrid_gdf_calc.iterrows():
    #hexagon centroid
    center_point=(filtered_hexgrid_gdf_calc.loc[i,'geometry'].y,filtered_hexgrid_gdf_calc.loc[i,'geometry'].x)
    #POIs
    df_schools=ox.geometries.geometries_from_point(center_point, tags_school, dist=27000)
    df_schools=df_schools.reset_index()
    #origin definition
    origin=filtered_hexgrid_gdf_calc.loc[i:i+1]
    origin = gpd.GeoDataFrame(origin, crs='epsg:4326')
    ent_amount=0
    if len(df_schools)>0:
        df_schools=df_schools.reset_index()
        df_schools=df_schools.rename(columns={'index':'id'})
        df_schools['geometry']=df_schools['geometry'].centroid

        #travel time by car
        travel_time_matrix_computer = TravelTimeMatrixComputer(transport_network,origins=origin,
                                                               destinations=df_schools,transport_modes=[TransportMode.CAR]
)
        result = travel_time_matrix_computer.compute_travel_times()
        #filtering based on threshold
        result=result[result['travel_time']<33]
        #number of pois available
        ent_amount=len(result)

    filtered_hexgrid_gdf_calc.loc[i,'amount car school']=ent_amount
```

Appendix A.2
Amenities distribution per city

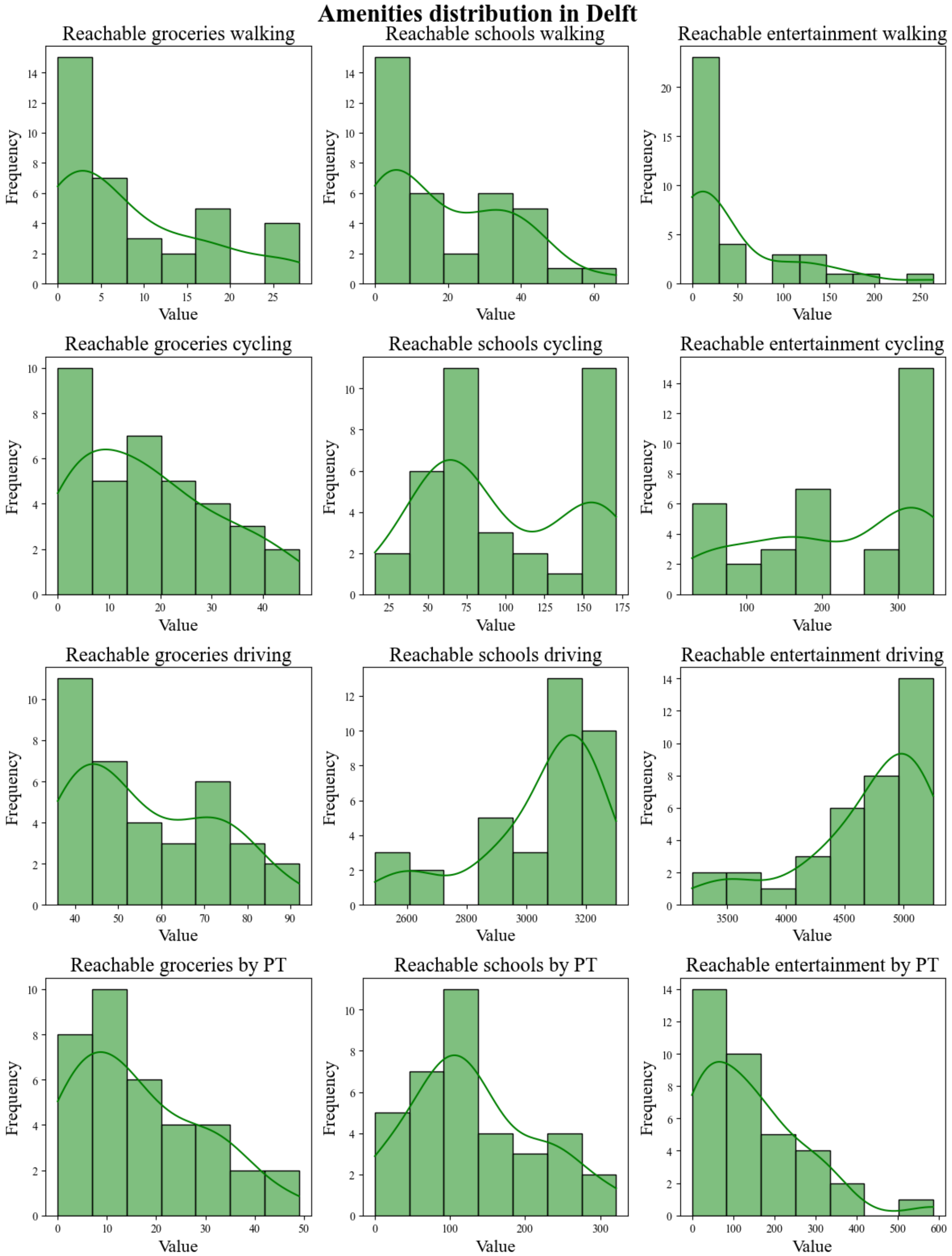


Figure A.2.1: Amenities' distribution in Delft

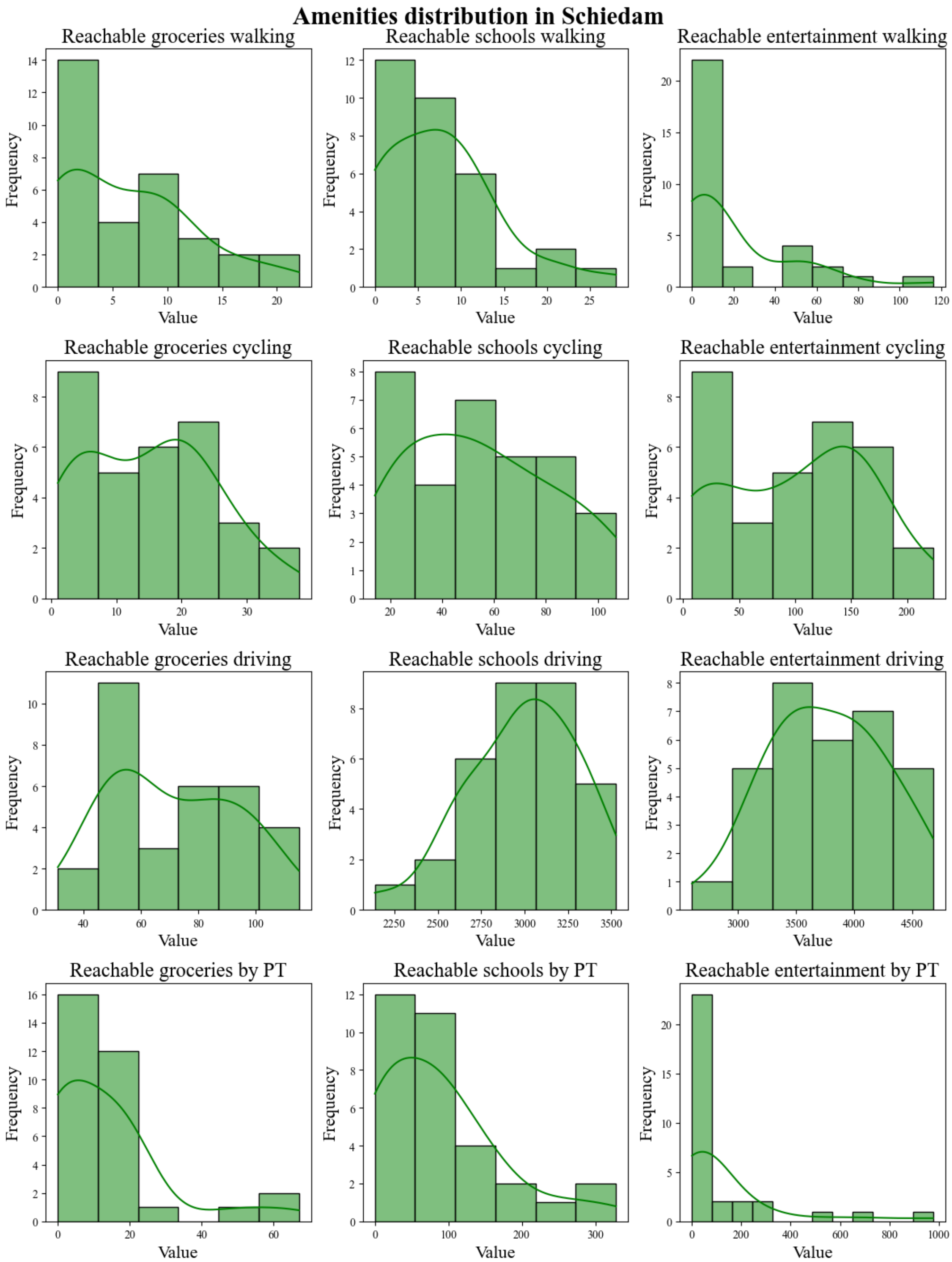


Figure A.2.2: Amenities' distribution in Schiedam

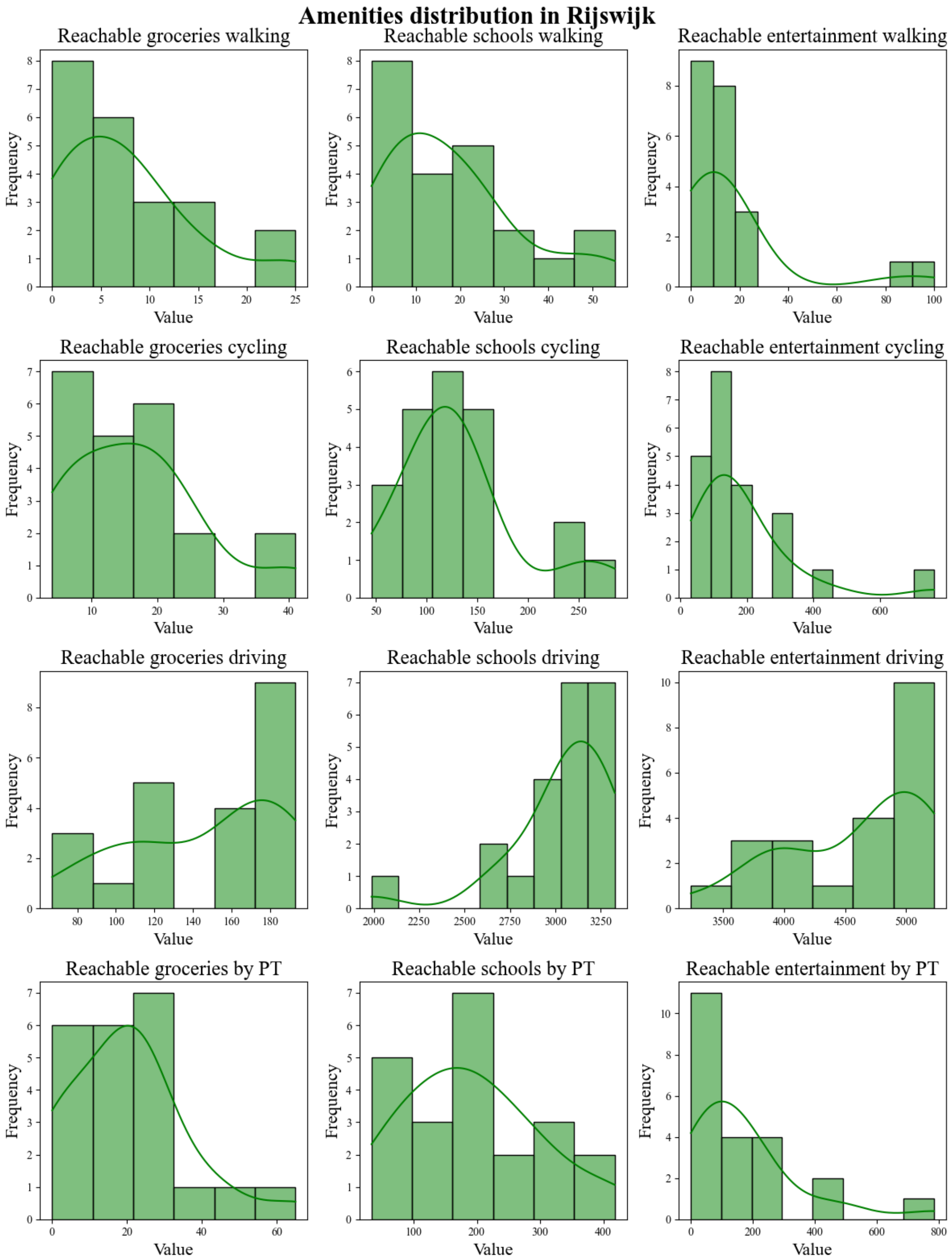


Figure A.2.3: Amenities' distribution in Rijswijk

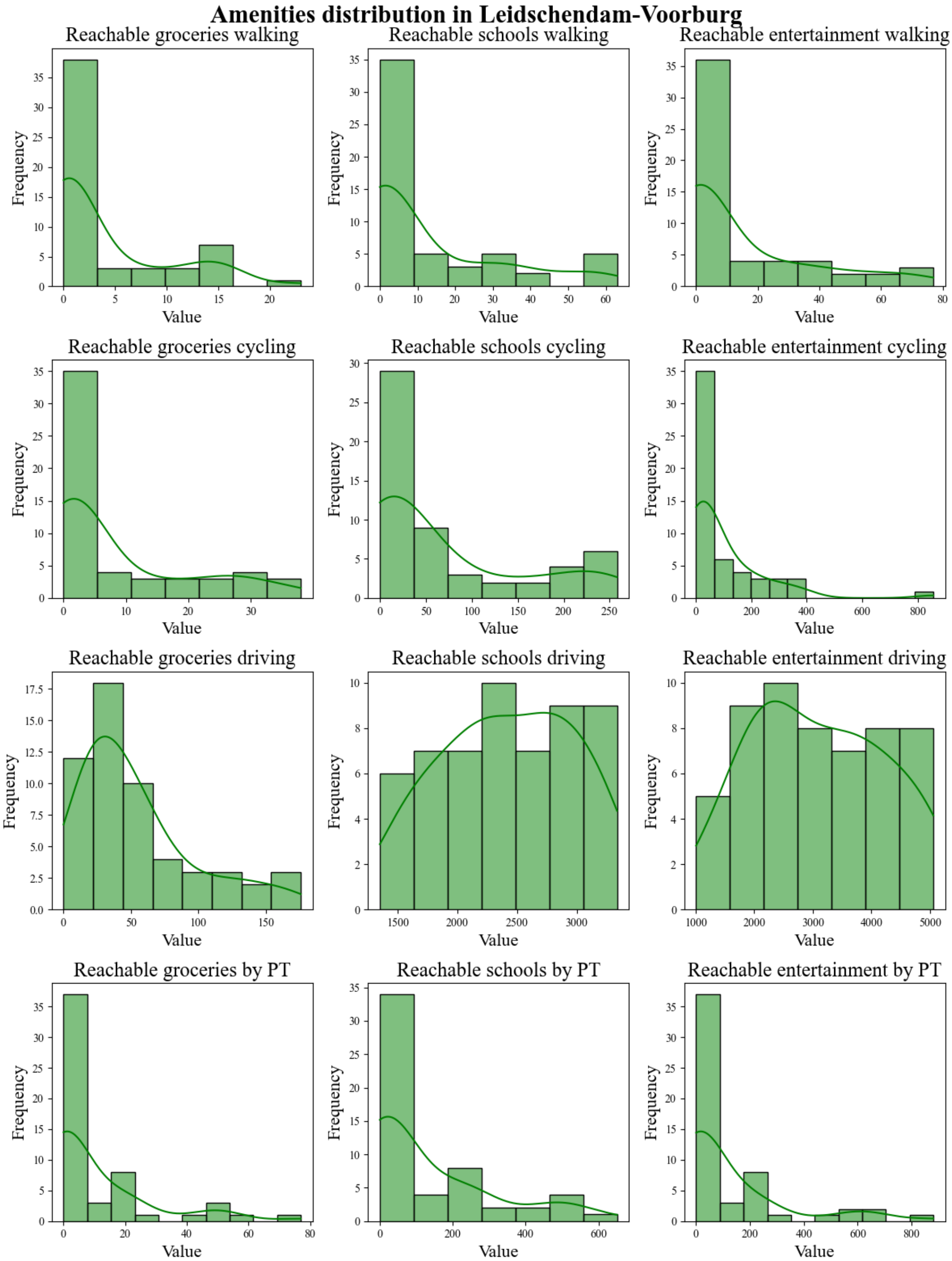


Figure A.2.4: Amenities' distribution in Leidschendam-Voorburg

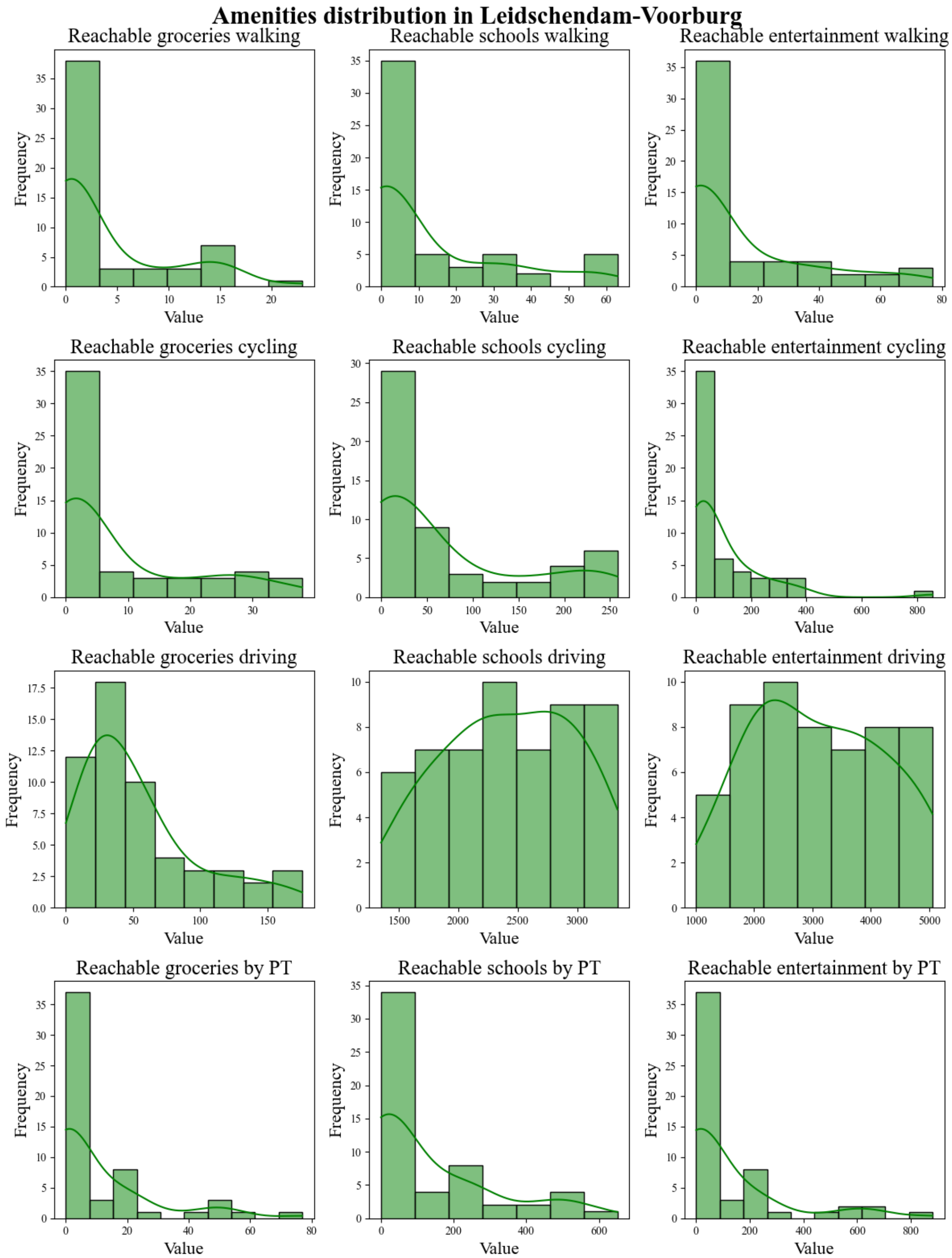


Figure A.2.5: Amenities' distribution in Leidschendam-Voorburg

Appendix A.3

Cities' break points definition through Fisher Jenks algorithm

Amenities' distribution and Breakpoints in Delft

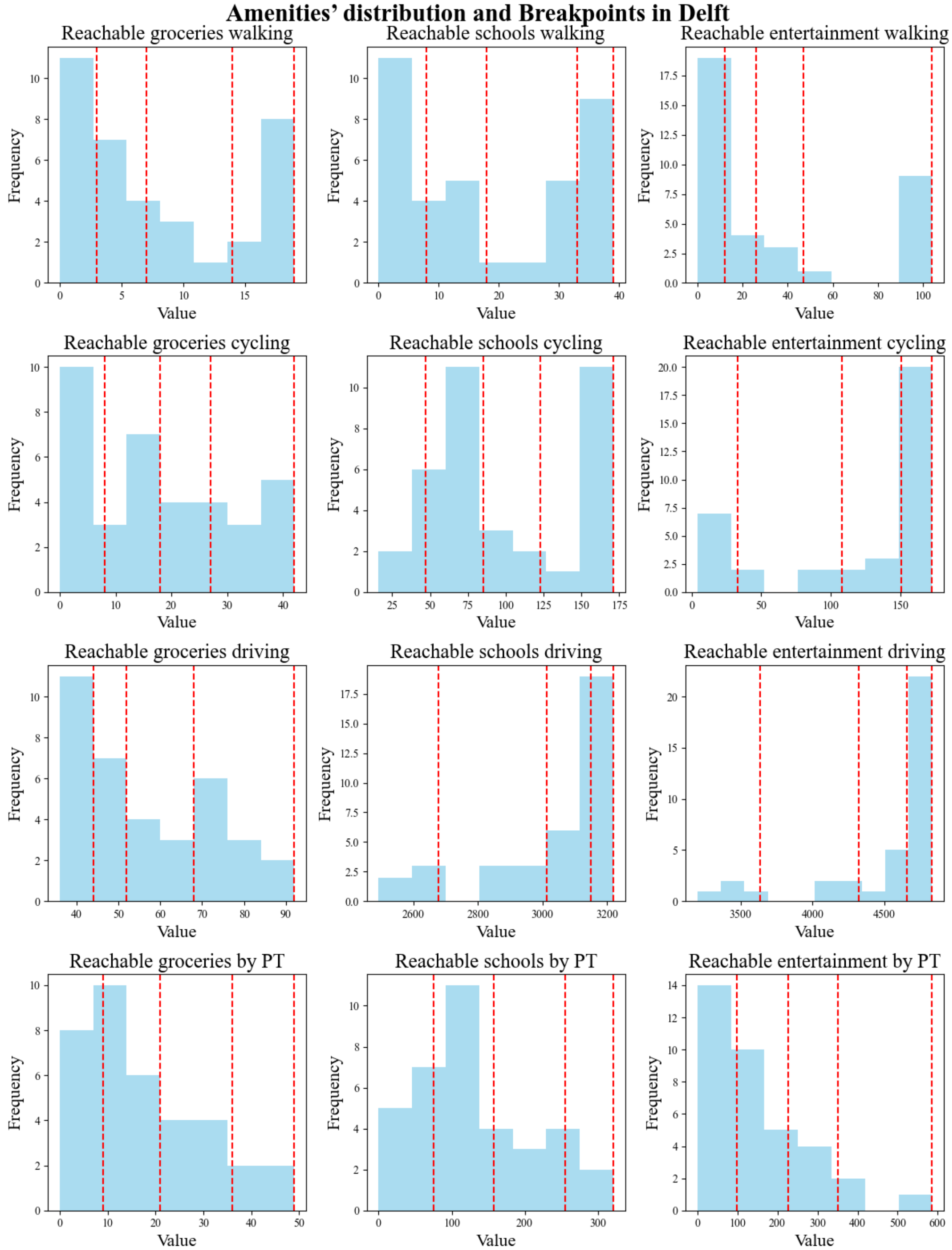


Figure A.3.1: Amenities' distribution and Breakpoints in Delft

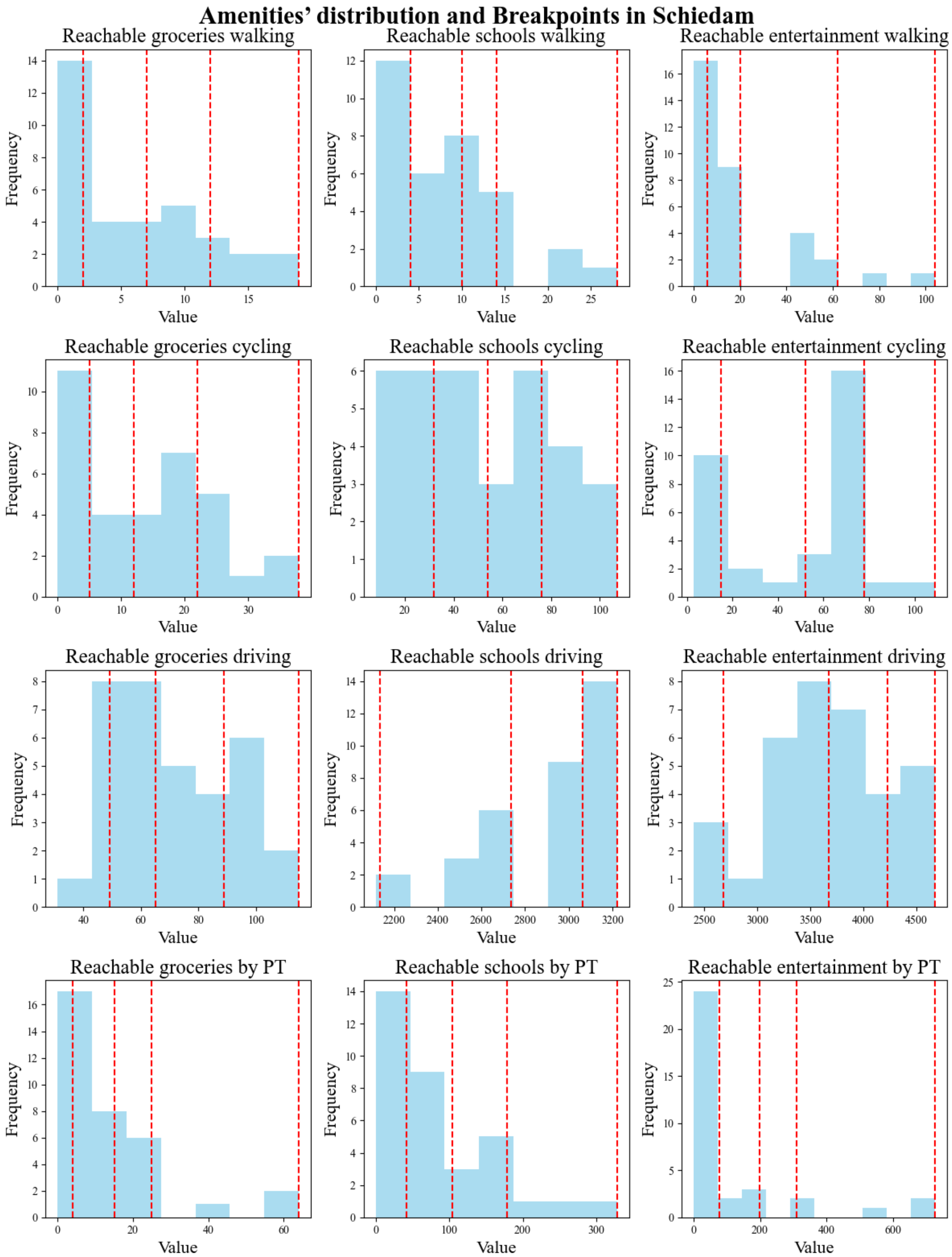


Figure A.3.2: Amenities' distribution and Breakpoints in Schiedam

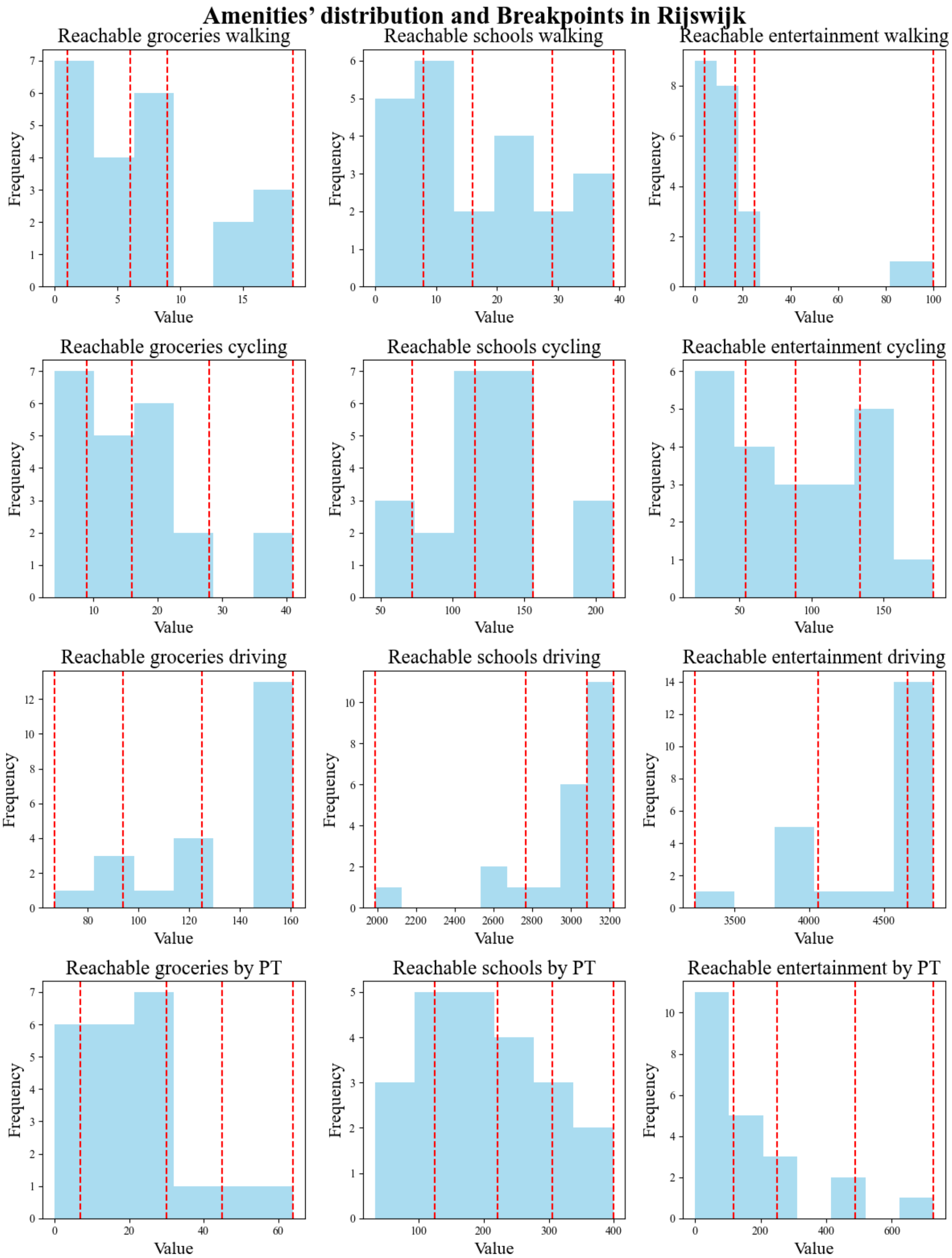


Figure A.3.3: Amenities’ distribution and Breakpoints in Rijswijk

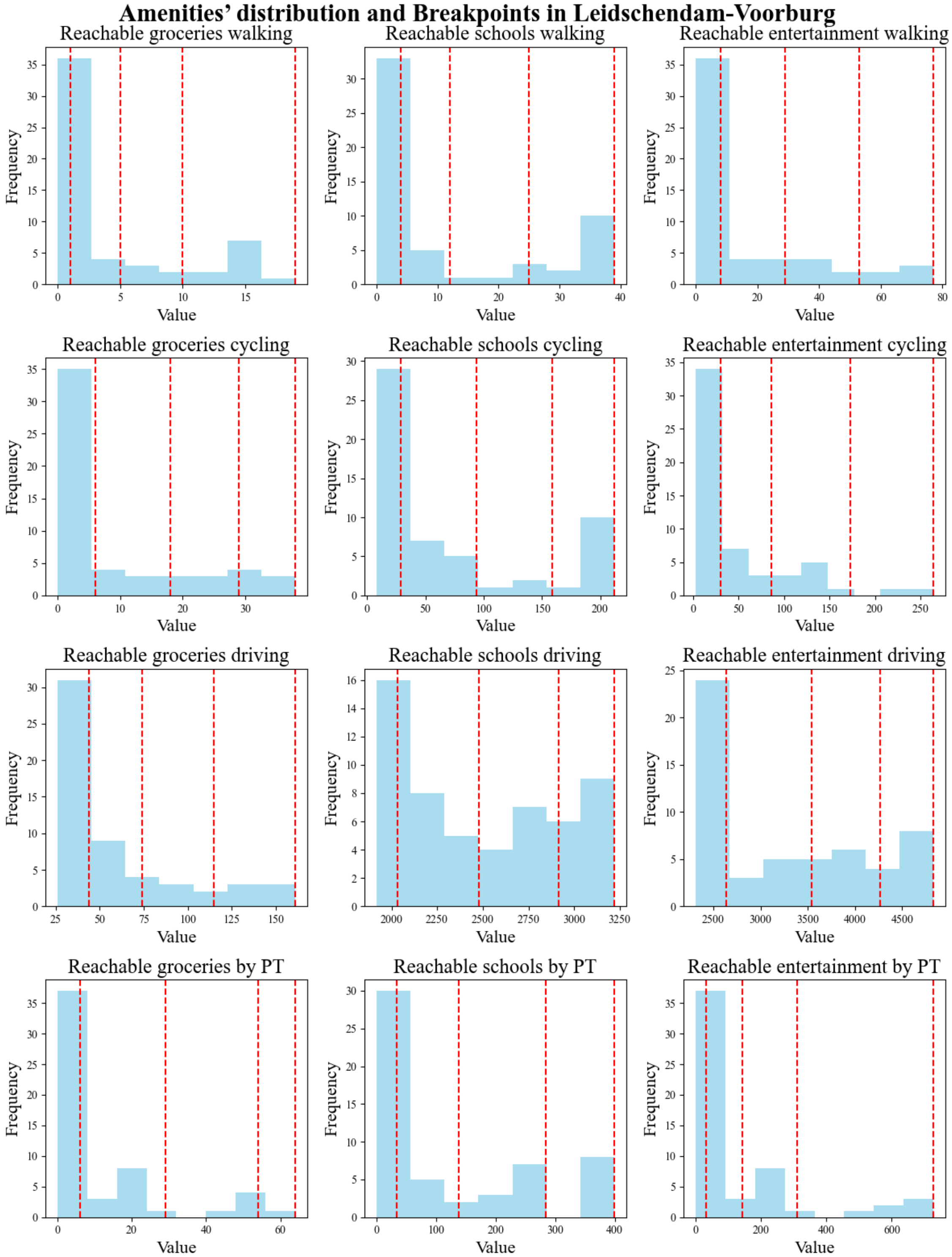


Figure A.3.4: Amenities' distribution and Breakpoints in Leidschendam-Voorburg

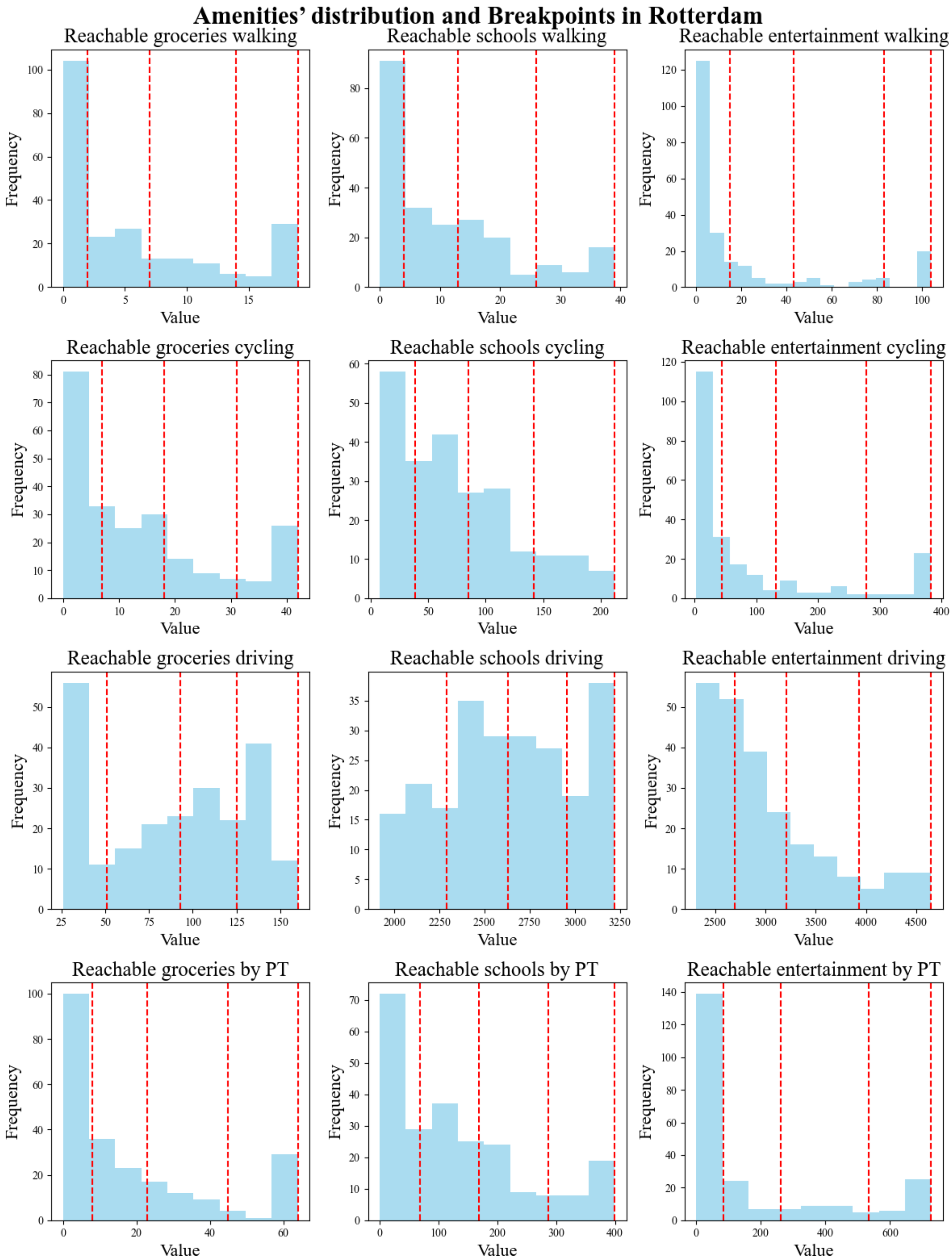


Figure A.3.5: Amenities' distribution and Breakpoints in Rotterdam

Appendix B. 1
Ethics Application Form – Data Management Plan

Master Thesis

0. Administrative questions

1. Name of data management support staff consulted during the preparation of this plan.

My faculty data steward, Nicolas Dintzner, has reviewed this DPM on 28/03/2023.

2. Date of consultation with support staff.

2023-03-28

I. Data description and collection or re-use of existing data

3. Provide a general description of the type of data you will be working with, including any re-used data:

Type of data	File format(s)	How will data be collected (for re-used data: source and terms of use)?	Purpose of processing	Storage location	Who will have access to the data
Gender, age, income, car and driver's license ownership, level of education, Postcode, household composition, and transport modes preferences to go to specific Points of Interest.	.csv files	Online survey	Store the participant's perception of using a transport mode to reach a point of interest.	Project storage drive (TU Delft One Drive)	Project team (author(s) and supervisors)
Spatial data of Points of Interest and Transport Networks in Rotterdam.	Jupyter Notebook to collect the data and .csv file to store the data	GIS data	Calculate the number of points of interest available by different transport modes.	Project storage drive (TU Delft One Drive)	Project team (author(s) and supervisors)

4. How much data storage will you require during the project lifetime?

- < 250 GB

II. Documentation and data quality

5. What documentation will accompany data?

- Data dictionary explaining the variables used
- README file or other documentation explaining how data is organised
- Methodology of data collection

III. Storage and backup during research process

Created using DMPonline. Last modified 31 March 2023 1 of 6

6. Where will the data (and code, if applicable) be stored and backed-up during the project lifetime?

- Another storage system - please explain below, including provided security measures
- OneDrive

GitHub is going to store the jupyter notebook. All should be accessible to my supervisors and co-authors. The questionnaire will be in the Master thesis report, and the other data will be in the One Drive repository.

IV. Legal and ethical requirements, codes of conduct

7. Does your research involve human subjects or 3rd party datasets collected from human participants?

- Yes

8A. Will you work with personal data? (information about an identified or identifiable natural person)

If you are not sure which option to select, ask your Faculty Data Steward for advice. You can also check with the privacy website or contact the privacy team: privacy-tud@tudelft.nl

- Yes

8B. Will you work with any other types of confidential or classified data or code as listed below? (tick all that apply)

If you are not sure which option to select, ask your Faculty Data Steward for advice.

- No, I will not work with any confidential or classified data/code

9. How will ownership of the data and intellectual property rights to the data be managed?

For projects involving commercially-sensitive research or research involving third parties, seek advice of your Faculty Contract Manager when answering this question. If this is not the case, you can use the example below.

This is an internal TU delft master thesis project and no third parties are involved. The data will be stored in the authors' database and a restricted One Drive folder. The access will be open for her and the graduation committee. The author will have the right to control access and be the data owner.

10. Which personal data will you process? Tick all that apply

- Other types of personal data - please explain below
- Gender, date of birth and/or age

Other personal data types will be asked, considering fixed ranges. Therefore, there are no open questions except the postcode.
Other types of personal data:

- level of education:

- Lower than middle school
- Middle school
- Mbo
- HBO bachelor

- WO Bachelor
- HBO master
- WO master
- PHD

- age:

- Between 18 - 25
- Between 26 - 35
- 35 - 45
- 45-55
- 55-65
- 65 - 75
- > 75

- gender:

- Female
- Male
- Other

-annual income range:

- -< 30000 gross
- - 30000 - 40000 gross
- - 40000- 60000 gross
- - 60000- 75000 gross
- - 75000- 90000 gross
- - >90000 gross

- household composition:

- at least one child below 5
- Two or more children below 5
- At least one child between 5 and 12
- Children above 12
- No children

- country of origin:

- Netherlands
- Another European country
- Africa
- Asia
- Oceania
- North America
- South America

- How long have you been living in The Netherlands

- <1 year
- 1 to 5 years
- 5 to 10 years
- >10 years

11. Please list the categories of data subjects

Adult citizens (>= 18 years old) that live in The Netherlands.

12. Will you be sharing personal data with individuals/organisations outside of the EEA (European Economic Area)?

- No

15. What is the legal ground for personal data processing?

- Informed consent

16. Please describe the informed consent procedure you will follow:

All study participants will be asked for their written consent for participating in the study and for data processing before the start of the survey.

17. Where will you store the signed consent forms?

- Same storage solutions as explained in question 6

18. Does the processing of the personal data result in a high risk to the data subjects?

If the processing of the personal data results in a high risk to the data subjects, it is required to perform Data Protection Impact Assessment (DPIA). In order to determine if there is a high risk for the data subjects, please check if any of the options below that are applicable to the processing of the personal data during your research (check all that apply).
If two or more of the options listed below apply, you will have to complete the DPIA. Please get in touch with the privacy team: privacy-tud@tudelft.nl to receive support with DPIA.
If only one of the options listed below applies, your project might need a DPIA. Please get in touch with the privacy team: privacy-tud@tudelft.nl to get advice as to whether DPIA is necessary.
If you have any additional comments, please add them in the box below.

- None of the above applies

22. What will happen with personal research data after the end of the research project?

- Personal research data will be destroyed after the end of the research project
- Anonymised or aggregated data will be shared with others

23. How long will (pseudonymised) personal data be stored for?

- Other - please state the duration and explain the rationale below

The personal data will be stored for two years under the responsibility of the research advisor Juliana Goncalves, email: J.E.Goncalves@tudelft.nl.
The spatial data will be stored in the TU Delft One drive.

24. What is the purpose of sharing personal data?

- Other - please explain below

We will not be sharing personal data.

25. Will your study participants be asked for their consent for data sharing?

- Yes, in consent form - please explain below what you will do with data from participants who did not consent to data sharing

We will not be sharing personal data.

V. Data sharing and long-term preservation

27. Apart from personal data mentioned in question 22, will any other data be publicly shared?

- All other non-personal data (and code) underlying published articles / reports / theses

It will be shared the survey questions, the geographical data gathered, the analysis script and survey aggregated answers.

29. How will you share research data (and code), including the one mentioned in question 22?

- All anonymised or aggregated data, and/or all other non-personal data will be uploaded to 4TU.ResearchData with public access

30. How much of your data will be shared in a research data repository?

- < 100 GB

31. When will the data (or code) be shared?

- At the end of the research project

32. Under what licence will be the data/code released?

- Apache

VI. Data management responsibilities and resources

33. Is TU Delft the lead institution for this project?

- Yes, the only institution involved

34. If you leave TU Delft (or are unavailable), who is going to be responsible for the data resulting from this project?

The Advisor of this master thesis project Juliana Gonçalves, email: J.E.Goncalves@tudelft.nl.

35. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

None.

Appendix B.2

Ethics Application Form – Informed Consent Form

Delft University of Technology
Informed consent form survey

English

You are invited to participate in a research study about accessibility. This study is being done by Iris Roeleven and Luisa de La Vega from the Technical University of Delft (TU Delft).

The purpose of this survey is to understand how accessibility differs from person to person. Accessibility means how easy it is to go to specific places using the existing transport system. For example, we want to understand how convenient it is for you to reach places such as grocery stores using different transport modes and/or during specific times of the day. We are particularly interested in the differences between people who identify as women and men.

The survey will take you approximately 10 minutes to complete. The data from this survey will be used for research purposes as a part of a Master's thesis, which looks into the effect of gender on accessibility to propose improvements to the current transport system. We will ask you for certain information, such as:

- Country of residence and postcode
- Socio-demographic information such as gender, age range, household income range, whether you live together with a partner, and whether you have children
- Your preferences and impressions about the transport modes (safety, quality, cost, availability)
- How convenient it is for you to use different transport modes to reach specific places
- If you wish to participate in a draw to win a voucher from bol.com, your email address (which can be entered in a link to a separate survey shown once you have submitted the main survey)

To the best of our ability, your answers in this study will remain confidential. The data will be used exclusively for research purposes about Accessibility and Gender, aiming to contribute to a more equitable transport system. As with any online activity, the risk of a breach is always possible and there is a risk of re-identification for the participants. We will minimize any risks by separating email addresses from survey answers, only analysing aggregated data, and deleting personal data after two years. Only aggregated survey answers will be published at the end of the study, which means that your answers will not be traced back to you. Content of open questions will not be shared in any way.

Your participation in this study is entirely voluntary and you can withdraw at any time. It will not be possible to remove answers to questions once the survey form has been completed and sent.

You can reach the research team through the following contact information:

- Iris Roeleven (corresponding researcher): I.R.Roeleven@student.tudelft.nl
- Luisa de La Vega (corresponding researcher): l.delavegabaymadeoliveira@student.tudelft.nl
- Maarten Kroesen (responsible researcher): M.Kroesen@tudelft.nl
- Juliana Goncalves (responsible researcher): J.E.Goncalves@tudelft.nl

By clicking through to this online survey and completing all its mandatory questions, you are agreeing to this Opening Statement and providing your informed consent to participate in this study.

Appendix B.3

Ethics Application Form – Human Research Ethics Checklist for Human Research

I. Applicant Information

PROJECT TITLE:	**Gender and Accessibility**
Research period: *Over what period of time will this specific part of the research take place*	**March 2023 – July 2023**
Faculty:	**Technology, Policy and Management**
Department:	**CoSEM**
Type of the research project: *(Bachelor's, Master's, DreamTeam, PhD, PostDoc, Senior Researcher, Organisational etc.)*	**Master's Thesis**
Funder of research: *(EU, NWO, TUD, other – in which case please elaborate)*	**None**
Name of Corresponding Researcher: *(If different from the Responsible Researcher)*	**Luisa de La Vega Bayma de Oliveira**
E-mail Corresponding Researcher: *(If different from the Responsible Researcher)*	**l.delavegabaymadeoliveira@student.tudelft.nl**
Position of Corresponding Researcher: *(Masters, DreamTeam, PhD, PostDoc, Assistant/ Associate/ Full Professor)*	**Masters student**
Name of Responsible Researcher: *Note: all student work must have a named Responsible Researcher to approve, sign and submit this application*	**Juliana Gonçalves**
E-mail of Responsible Researcher: *Please ensure that an institutional email address (no Gmail, Yahoo, etc.) is used for all project documentation/ communications including Informed Consent materials*	**j.e.goncalves@tudelft.nl**
Position of Responsible Researcher : *(PhD, PostDoc, Associate/ Assistant/ Full Professor)*	**Assistant Professor**

II. Research Overview

NOTE: You can find more guidance on completing this checklist here

a) Please summarise your research very briefly (100-200 words)
What are you looking into, who is involved, how many participants there will be, how they will be recruited and what are they expected to do?

Add your text here – (please avoid jargon and abbrevations)
This research aims to analyze the impact of different personal characteristics, mainly gender, on the accessibility perception of citizens in The Netherlands. The target group that will be examined is adults (from eighteen years old) that live in The Netherlands, specifically Rotterdam. This study considers the application of digital surveys for around 500 people. The surveys will be promoted in diverse locations, such as commercial establishments and the university campus in Rotterdam. In addition, the surveys will be applied to the authors' and supervisors' networks. To understand the impact of different personal characteristics on accessibility levels, this study compares accessibility perception collected from surveys with the accessibility calculated by spatial data. The surveys include questions about the participant's perception of using different transport modes to reach points of interest, such as grocery stores from their houses. The spatial accessibility is calculated by mapping the transport mode options and several points of interest available based on the participant's postcode. It should be noted that for the surveys, this research study is combined with the work of fellow CoSEM master student Iris Roeleven, who has the same supervising team. This means one survey will be used to gather data for both studies.

b) **If your application is an additional project** related to an existing approved HREC submission, please provide a brief explanation including the existing relevant HREC submission number/s.

Add your text here – (please avoid jargon and abbrevations)

c) **If your application is a simple extension of, or amendment to,** an existing approved HREC submission, you can simply submit an HREC Amendment Form as a submission through LabServant.

III. Risk Assessment and Mitigation Plan

NOTE: You can find more guidance on completing this checklist ***here***

Please complete the following table in full for all points to which your answer is "yes". Bear in mind that the vast majority of projects involving human participants as Research Subjects also involve the collection of Personally Identifiable Information (PII) and/or Personally Identifiable Research Data (PIRD) which may pose potential risks to participants as detailed in Section G: Data Processing and Privacy below.

To ensure alighment between your risk assessment, data management and what you agree with your Research Subjects you can use the last two columns in the table below to refer to specific points in your Data Management Plan (DMP) and Informed Consent Form (ICF) – **but this is not compulsory**.

It's worth noting that you're much more likely to need to resubmit your application if you neglect to identify potential risks, than if you identify a potential risk and demonstrate how you will mitigate it. If necessary, the HREC will always work with you and colleagues in the Privacy Team and Data Management Services to see how, if at all possible, your research can be conducted.

			If YES please complete the Risk Assessment and Mitigation Plan columns below.		*Please provide the relevant reference #*	
ISSUE	**Yes**	**No**	**RISK ASSESSMENT – what risks could arise?** *Please ensure that you list ALL of the actual risks that could potentially arise – do not simply state whether you consider any such risks are important!*	**MITIGATION PLAN – what mitigating steps will you take?** *Please ensure that you summarise what actual mitigation measures you will take for each potential risk identified – do not simply state that you will e.g. comply with regulations.*	**DMP**	**ICF**
A: Partners and collaboration						
1. Will the research be carried out in collaboration with additional organisational partners such as: • One or more collaborating research and/or commercial organisations • Either a research, or a work experience internship provider[1] *[1] If yes, please include the graduation agreement in this application*		X				
2. Is this research dependent on a Data Transfer or Processing Agreement with a collaborating partner or third party supplier? *If yes please provide a copy of the signed DTA/DPA*		X				
3. Has this research been approved by another (external) research ethics committee (e.g.: HREC and/or MREC/METC)? *If yes, please provide a copy of the approval (if possible) and summarise any key points in your Risk Management section below*						
B: Location						

			If YES please complete the Risk Assessment and Mitigation Plan columns below.		*Please provide the relevant reference #*	
ISSUE	**Yes**	**No**	**RISK ASSESSMENT – what risks could arise?** *Please ensure that you list ALL of the actual risks that could potentially arise – do not simply state whether you consider any such risks are important!*	**MITIGATION PLAN – what mitigating steps will you take?** *Please ensure that you summarise what actual mitigation measures you will take for each potential risk identified – do not simply state that you will e.g. comply with regulations.*	**DMP**	**ICF**
4. Will the research take place in a country or countries, other than the Netherlands, within the EU?		X				
5. Will the research take place in a country or countries outside the EU?		X				
6. Will the research take place in a place/region or of higher risk – including known dangerous locations (in any country) or locations with non-democratic regimes?		X				
C: Participants						
7. Will the study involve participants who may be vulnerable and possibly (legally) unable to give informed consent? (e.g., children below the legal age for giving consent, people with learning difficulties, people living in care or nursing homes,).		X				
8. Will the study involve participants who may be vulnerable under specific circumstances and in specific contexts, such as victims and witnesses of violence, including domestic violence; sex workers; members of minority groups, refugees, irregular migrants or dissidents?	X		As the survey is conducted voluntarily by a sample of Dutch citizens, there may be vulnerable people amongst the people who decide to participate in the survey. These people might feel emotional or mental discomfort while filling out the survey.	No directly identifiable data will be gathered in the survey. Furthermore, the survey will not be designed to collect data about the vulnerabilities of respondents. Lastly, it will be made clear to respondents on the first page of the survey that they can quit at any time. Besides, the survey will be applied only to adults (18 years old or older).		
9. Are the participants, outside the context of the research, in a dependent or subordinate position to the investigator (such as own children, own students or employees of either TU Delft and/or a collaborating partner organisation)? *It is essential that you safeguard against possible adverse consequences of this situation (such as allowing a student's failure to participate to your satisfaction to affect your evaluation of their coursework).*		X				
10. Is there a high possibility of re-identification for your participants? (e.g., do they have a very specialist job of which there are only a small number in a given country, are they members of a small community, or employees from a partner company collaborating in the research? Or are they one of only a handful of (expert) participants in the study?	X		There is a risk of identifying the participants by combining the post code and the demographics data (e.g: age, income, gender).	Individual data will not be published in any way, only aggregated data, so re-identification from published data is impossible. In addition, the survey data will be stored securely in the TU Delft OneDrive and only be privately available in the TU Delft OneDrive for the author, collaborator and supervisors.		
D: Recruiting Participants						

			If YES please complete the Risk Assessment and Mitigation Plan columns below.		*Please provide the relevant reference #*	
ISSUE	**Yes**	**No**	**RISK ASSESSMENT – what risks could arise?** *Please ensure that you list ALL of the actual risks that could potentially arise – do not simply state whether you consider any such risks are important!*	**MITIGATION PLAN – what mitigating steps will you take?** *Please ensure that you summarise what actual mitigation measures you will take for each potential risk identified – do not simply state that you will e.g. comply with regulations.*	**DMP**	**ICF**
11. Will your participants be recruited through your own, professional, channels such as conference attendance lists, or through specific network/s such as self-help groups		X				
12. Will the participants be recruited or accessed in the longer term by a (legal or customary) gatekeeper? (e.g., an adult professional working with children; a community leader or family member who has this customary role – within or outside the EU; the data producer of a long-term cohort study)		X				
13. Will you be recruiting your participants through a crowd-sourcing service and/or involve a third party data-gathering service, such as a survey platform?	X		Qualtrics will be used as a survey platform. As this is a platform from the US, there is a risk of personal data breach, which could lead to (reputational) damage to participants.	Qualtrics has been approved by the TU Delft and therefore is aligned with the regulations. To still minimize the potential damage of a data breach, no directly identifiable personal data will be asked for in the main survey. Additionally, questions regarding personal data will be made as broad as possible, for example, by asking for an age range and income range instead of specific numbers. This way the potential damage of a data breach is lessened. To minimize the risk of data breach itself the guidance from the TU Delft with regard to Qualtrics will be followed, where survey and answers will be downloaded from Qualtrics once sufficient people have filled out the survey, after which the survey and data will be deleted from Qualtrics. To be able to give away a gift card to one or two of the participant, email addresses will need to be collected. However, by linking to a different survey to collect email addresses and turning off the registration of Ip-addresses etc., these email addresses will not be linked to survey answers. This also limits the damage of a data breach.		
14. Will you be offering any financial, or other, remuneration to participants, and might this induce or bias participation?	X		Survey: participants of the survey will be given the opportunity to enter a lottery to win a Bol.com gift voucher. To be able to do this email addresses of those people who want to enter the lottery will need to be collected. This could enlarge the risk of re-identification of participants and their survey	To minimize the risk of re-identification, the email addresses of participants will not be gathered in the same survey as the one in which participants have to answer questions. Instead at the end of this survey, there will be a link to another survey which has the sole purpose of gathering email addresses for a lottery		

			If YES please complete the Risk Assessment and Mitigation Plan columns below.		*Please provide the relevant reference #*	
ISSUE	**Yes**	**No**	**RISK ASSESSMENT – what risks could arise?** *Please ensure that you list ALL of the actual risks that could potentially arise – do not simply state whether you consider any such risks are important!*	**MITIGATION PLAN – what mitigating steps will you take?** *Please ensure that you summarise what actual mitigation measures you will take for each potential risk identified – do not simply state that you will e.g. comply with regulations.*	**DMP**	**ICF**
			answers. Additionally, it could cause a bias for people who are more responsive to a monetary incentive.	draw. By also turning on the anonymise data option in Qualtrics, IP-addresses will not be gathered making the risk of linking the survey answers to the email address minimal and therefore making the survey anonymous By using this method participants will only be able to participate in the lottery draw at the end of the survey. As they have to fill in their email address, this will minimize the risk of people doing the survey multiple times as they would either have to put the same email address multiple times, in which case they will be disqualified, or putting in additional effort by making more email addresses, which would be disproportional for the limited chance of reward.		
E: Subject Matter *Research related to medical questions/health may require special attention. See also the website of the CCMO before contacting the HREC.*						
15. Will your research involve any of the following: • Medical research and/or clinical trials • Invasive sampling and/or medical imaging • Medical and *In Vitro Diagnostic Medical* Devices Research		X				
16. Will drugs, placebos, or other substances (e.g., drinks, foods, food or drink constituents, dietary supplements) be administered to the study participants? *If yes see here to determine whether medical ethical approval is required*		X				
17. Will blood or tissue samples be obtained from participants? *If yes see here to determine whether medical ethical approval is required*		X				
18. Does the study risk causing psychological stress or anxiety beyond that normally encountered by the participants in their life outside research?		X				
19. Will the study involve discussion of personal sensitive data which could put participants at increased legal, financial, reputational, security or other risk? (e.g., financial data, location data, data relating to children or other vulnerable groups) *Definitions of sensitive personal data, and special cases are provided on the TUD Privacy Team website.*		X				

			If YES please complete the Risk Assessment and Mitigation Plan columns below.		*Please provide the relevant reference #*	
ISSUE	**Yes**	**No**	**RISK ASSESSMENT – what risks could arise?** *Please ensure that you list ALL of the actual risks that could potentially arise – do not simply state whether you consider any such risks are important!*	**MITIGATION PLAN – what mitigating steps will you take?** *Please ensure that you summarise what actual mitigation measures you will take for each potential risk identified – do not simply state that you will e.g. comply with regulations.*	**DMP**	**ICF**
20. Will the study involve disclosing commercially or professionally sensitive, or confidential information? (e.g., relating to decision-making processes or business strategies which might, for example, be of interest to competitors)		X				
21. Has your study been identified by the TU Delft Privacy Team as requiring a Data Processing Impact Assessment (DPIA)? *If yes please attach the advice/ approval from the Privacy Team to this application*						
22. Does your research investigate causes or areas of conflict? *If yes please confirm that your fieldwork has been discussed with the appropriate safety/security advisors and approved by your Department/Faculty.*		X				
23. Does your research involve observing illegal activities or data processed or provided by authorities responsible for preventing, investigating, detecting or prosecuting criminal offences *If so please confirm that your work has been discussed with the appropriate legal advisors and approved by your Department/Faculty.*		X				
F: Research Methods						
24. Will it be necessary for participants to take part in the study without their knowledge and consent at the time? (e.g., covert observation of people in non-public places).		X				
25. Will the study involve actively deceiving the participants? (For example, will participants be deliberately falsely informed, will information be withheld from them or will they be misled in such a way that they are likely to object or show unease when debriefed about the study).		X				
26. Is pain or more than mild discomfort likely to result from the study? And/or could your research activity cause an accident involving (non-) participants?		X				
27. Will the experiment involve the use of devices that are not 'CE' certified? *Only, if 'yes': continue with the following questions:*		X				
• Was the device built in-house?		X				
• Was it inspected by a safety expert at TU Delft? *If yes, please provide a signed device report*		X				
• If it was not built in-house and not CE-certified, was it inspected by some other, qualified authority in safety and approved? *If yes, please provide records of the inspection*		X				
28. Will your research involve face-to-face encounters with your participants and if so how will you assess and address Covid considerations?		X				

			If YES please complete the Risk Assessment and Mitigation Plan columns below.		*Please provide the relevant reference #*	
ISSUE	**Yes**	**No**	**RISK ASSESSMENT – what risks could arise?** *Please ensure that you list ALL of the actual risks that could potentially arise – do not simply state whether you consider any such risks are important!*	**MITIGATION PLAN – what mitigating steps will you take?** *Please ensure that you summarise what actual mitigation measures you will take for each potential risk identified – do not simply state that you will e.g. comply with regulations.*	**DMP**	**ICF**
29. Will your research involve either: a) "big data", combined datasets, new data-gathering or new data-merging techniques which might lead to re-identification of your participants and/or b) artificial intelligence or algorithm training where, for example biased datasets could lead to biased outcomes?	X		The research combines different datasets, one originated from mapping transport networks and points of interest in Rotterdam and other from surveys' responses. There is a risk that people can be re-identified. In addition this study applies cluster analysis techniques to identify clusters in the data. There is a risk that the data from surveys is not representative, leading to biased outcomes.	Individual data will not be published, only results from aggregated data and statistical analysis. In addition, the data will be private in TU Delft repository, available only for the author, collaborator and supervisors. To avoid biased data, the surveys will be applied to people from different networks and also they will be advertised in several establishments and public spaces, increasing participants diversity.		
G: Data Processing and Privacy						
30. Will the research involve collecting, processing and/or storing any directly identifiable PII (Personally Identifiable Information) including name or email address that will be used for administrative purposes only? (eg: obtaining Informed Consent or disbursing remuneration)	X		Email addresses of participants who want to participate in the lottery draw will be gathered, which can cause a risk of re-identification.	Only the email addresses of those wanting to participate in the lottery draw will be gathered. To still make the answers anonymous, email addresses will be gathered in a separate survey, to which a link is posted on the final page of the first survey. This way the email addresses will not be linked to survey answers. Additionally, email addresses will be deleted immediately after the gift card winners have been selected.		
31. Will the research involve collecting, processing and/or storing any directly or indirectly identifiable PIRD (Personally Identifiable Research Data) including videos, pictures, IP address, gender, age etc and **what other Personal Research Data** (including personal or professional views) will you be collecting?	X		In the survey data will be collected on participant's gender, age range, income range, highest finished education level, whether they are living together with a partner and whether they have children (adult or minor) and additionally, they will be asked for their postal code. Additionally, people will be asked about their perception of the transport system and how easily it is for them to live a satisfactory life using this transport system. The combination of postal code and certain socio-demographic characteristics, means there is a risk of re-identification of participants, which could mean that participants' answers can be linked back to them. This in turn could cause (reputational) damage to participants.	The risk of re-identification will be minimized as much as possible by keeping questions as broad as possible, while still being useful for the research. By asking for an age range and income range the risk of re-identification is less than when one would ask for precise age or income. Additionally, it is only asked whether people have children and whether those children are still minors, not their exact ages. By keeping answers broad, the risk of re-identification is thus minimized as much as possible. However, because of the question about people's postal codes it is difficult to completely rule out the potential for re-identification and it is therefore especially important to keep the data private. Therefore, only aggregated data will be published. The specific survey answers will be safely stored in TU Delft OneDrive where it is only		

			If YES please complete the Risk Assessment and Mitigation Plan columns below.		*Please provide the relevant reference #*	
ISSUE	**Yes**	**No**	**RISK ASSESSMENT – what risks could arise?** *Please ensure that you list ALL of the actual risks that could potentially arise – do not simply state whether you consider any such risks are important!*	**MITIGATION PLAN – what mitigating steps will you take?** *Please ensure that you summarise what actual mitigation measures you will take for each potential risk identified – do not simply state that you will e.g. comply with regulations.*	**DMP**	**ICF**
				available to the author, collaborator and supervisors. By not sharing the data, the risk of re-identification is thus further minimized. Lastly, no especially sensitive data like race or sexual orientation will be asked for, minimizing the damage of potential re-identification.		
32. Will this research involve collecting data from the internet, social media and/or publicly available datasets which have been originally contributed by human participants		X				
33. Will your research findings be published in one or more forms in the public domain, as e.g., Masters thesis, journal publication, conference presentation or wider public dissemination?	X		The research findings will be published as part of a Master's thesis. This could mean that respondents could be identified based on findings. This could cause further problems for respondents such as reputational issues.	Personal information will not be published in the master thesis, and results will only be shown in an aggregated way. Thus, the exact personal survey answers will not be published in any way. Furthermore, email addresses gathered for the survey lottery draw will also not be published in any way and be deleted once the winner(s) of the gift cards have been selected, and moreover, the email addresses cannot be linked to the survey answers of a respondent		
34. Will your research data be archived for re-use and/or teaching in an open, private or semi-open archive?		X				

H: More on Informed Consent and Data Management

NOTE: You can find guidance and templates for preparing your Informed Consent materials) here

Your research involves human participants as Research Subjects if you are recruiting them or actively involving or influencing, manipulating or directing them in any way in your research activities. This means you must seek informed consent and agree/ implement appropriate safeguards regardless of whether you are collecting any PIRD.

Where you are also collecting PIRD, and using Informed Consent as the legal basis for your research, you need to also make sure that your IC materials are clear on any related risks and the mitigating measures you will take – including through responsible data management.

Got a comment on this checklist or the HREC process? You can leave your comments here

IV. Signature/s

Please note that by signing this checklist list as the sole, or Responsible, researcher you are providing approval of the completeness and quality of the submission, as well as confirming alignment between GDPR, Data Management and Informed Consent requirements.

Name of Corresponding Researcher (if different from the Responsible Researcher) (print)

Luisa de La Vega Bayma de Oliveira
Signature of Corresponding Researcher: Luisa de La Vega B.O.

Date: 29/03/2023

Name of Responsible Researcher (print)

Juliana Gonçalves
Signature (or upload consent by mail) Responsible Researcher:

Date: Juliana Goncalves

V. Completing your HREC application

Please use the following list to check that you have provided all relevant documentation

Required:

- **Always:** This completed HREC checklist
- **Always:** A data management plan (reviewed, where necessary, by a data-steward)
- **Usually:** A complete Informed Consent form (including Participant Information) and/or Opening Statement (for online consent)

Please also attach any of the following, if relevant to your research:

Document or approval	Contact/s
Full Research Ethics Application	After the assessment of your initial application HREC **will let you know if and when you need to submit additional information**
Signed, valid Device Report	Your Faculty HSE advisor
Ethics approval from an external Medical Committee	TU Delft Policy Advisor, Medical (Devices) Research
Ethics approval from an external Research Ethics Committee	Please append, if possible, with your submission
Approved Data Transfer or Data Processing Agreement	Your Faculty Data Steward and/or TU Delft Privacy Team
Approved Graduation Agreement	Your Master's thesis supervisor
Data Processing Impact Assessment (DPIA)	TU Delft Privacy Team
Other specific requirement	Please reference/explain in your checklist and append with your submission

Appendix B.4
Ethics Application Form Approval

Lab Servant <servant_noreply@tudelft.nl>

Fri 07/04, 16:24

Luisa de La Vega Bayma de Oliveira; Trivik Verma

Reply all

> From: HREC
> Date: 31-Mar-2023
>
> Dear Luisa De La Vega Bayma De Oliveira,
>
> Your application titled: *Gender and Accessibility* is **Approved**.
>
> Please click here to view the approval letter and the details of your application (after log-on to the Lab Servant).

Appendix B. 5

Strategies to identify Bots' responses

The survey was launched on 25th April 2023 at 10 am. The first potential bot answer started at 7 pm on the same day. The identification of the suspicious activity, however, was made at 10 am of the following day. Since then, several strategies have been implemented based on previous literature to detect the presence of bot answers:

- Inclusion of attention check questions (Storozuk et al., 2020)

The questionnaire included the question, "This is a brief attention check; please answer 'strongly disagree' to this question instead of other answers like 'agree.'" In addition, the question "Are you 18 or above (years old)?" was included as a double filter. Participants younger than 18 years old were forward to end the survey. This research did not consider the participants who failed at least one of the two attention check questions.

- Monitoring time of survey completion (Storozuk et al., 2020)

A red flag was raised for the answers received in the middle of the night (from 12 am to 6 am). The authors considered that a human's chances of answering the questionnaire at night are considerably lower.

- Monitoring speed of survey completion (Storozuk et al., 2020)

The average time to complete the survey, according to Qualtrics calculation, is 10 minutes. Based on a survey trial with a close network, the survey duration varies from 5 min to 15 min. Thus, answers with a duration time of 4 minutes or less were dropped out.

- Embedding a CAPTCHA into the survey (Storozuk et al., 2020)

Qualtrics Platform provides a Captcha embedding option into the survey. This research arbitrarily dropped answers with Captcha Results lower than 0.7.

- Avoid financial compensation (Hallberg, 2022)

The lottery prize was canceled, the current survey was paused, and a new version was created. In this way, participants that tried the old link would need help to complete the survey.

- Add repeat questions for consistency.

The question "In what Dutch province do you live?" was added to be later compared to the postal code filled in by the participant. In addition, postal codes that presented the correct Dutch format but did not exist in The Netherlands were considered invalid, and inconsistent answers were dropped out.

- Provide Personal Survey Links (Storozuk et al., 2020)

Qualtrics Software enables the creation of different links according to the channel used to distribute the survey. Thus, this research considers using a social media link to be distributed among social media channels and an anonymous link to be shared with a closed network. Moreover, a QR Code option was included in the flyers. Hence, monitoring each communication channel the answers came from became easy.

- Language criteria

As mentioned previously, the potential bot answers present an American Postal Code. Hence, all responses in English that started since the first likely bot at 7 pm, until the following day, in English were dropped out.

- Manual Check

The authors manually checked for other inconsistencies. For instance, participants who answered car as the primary transport mode and said they did not have access to a car in another question were dropped out.

This experience confirms that bots can learn and adapt fast (e.g., accurate answers to attention check questions, slower completion speeds), which makes it harder to spot these fraudulent responses. They improve their skills in mimicking human replies the longer they can access a poll (Storozuk et al., 2020). Thus, the answers from the first questionnaire version were considered valid only if they passed all the above-mentioned criteria. From 1099 answers, only 147 were deemed correct. The later version without the lottery prize received no suspicious bot activity.

Appendix C
Chi-Square Degrees of Freedom

df	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	---	---	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645

28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.87	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

Appendix D.1
Elbow Method Results

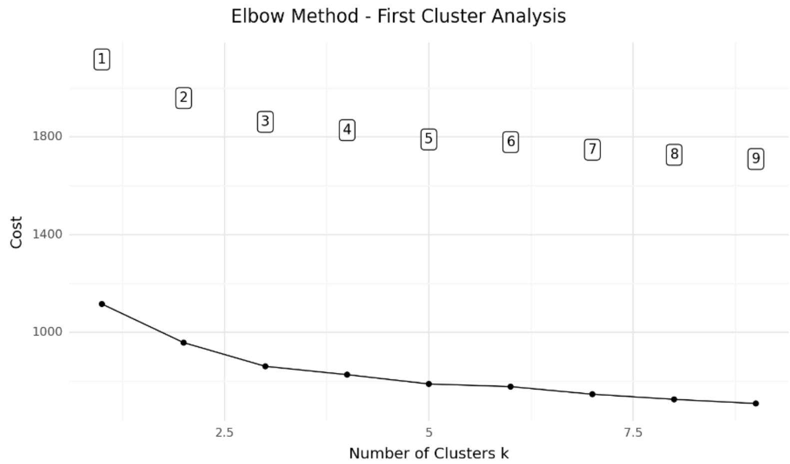


Figure D.1.1: Elbow Method of the 1st Cluster Analysis

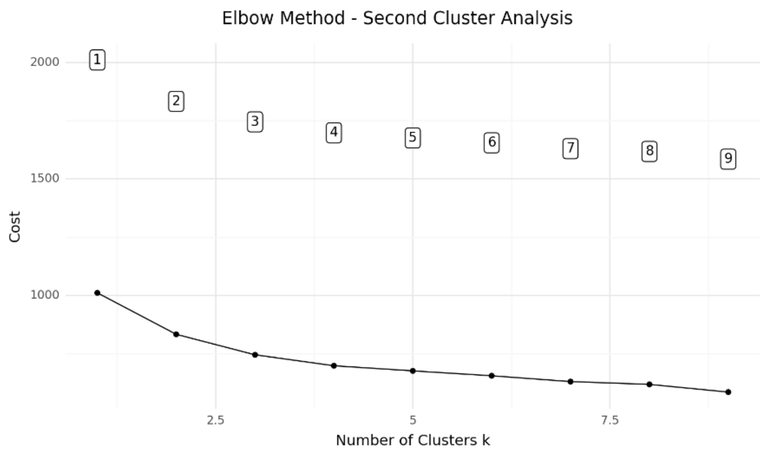


Figure D.1.2: Elbow Method of the 2nd Cluster Analysis

Appendix D.2
Silhouette Method Results

Silhouette score from 2 to 9 clusters for each Cluster Analysis

Analysis number	Silhouette score							
	k=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9
1st Cluster Analysis	0.19	0.34	0.27	0.19	0.12	0.12	0.15	0.11
2nd Cluster Analysis	0.21	0.37	0.27	0.18	0.16	0.17	0.12	0.12

Appendix E.1

Primarily Features selection per Logistic Regression Analysis

Variables' type	1st Logistic Regression – Mismatch Entertainment by car	2nd Logistic Regression – Mismatch Entertainment walking	3rd Logistic Regression – Mismatch Grocery stores by car
Individual Components	Gender, Age, Country of Origin, Income, Highest Education Level, Children, Partner presence	Gender, Age, Country of Origin, Income, Highest Education Level, Children, Partner presence	Gender, Age, Country of Origin, Income, Highest Education Level, Children, Partner presence
Material Component	Car Access	-	Car Access
Competence Component	Driver's License	-	Driver's License
Social resources	Social Support	-	Social Support
Safety Resources	Safety perception by car	Safety perception walking	Safety perception by car
Mismatch type (Target variable)	Entertainment activities by car	Entertainment activities walking	Grocery stores by car

Appendix E.2

Multicollinearity variables analysis for each Regression

Regression Analysis	Multicollinearity between variables (>0.7)	Variables correlation with the target variable	Variable deleted
Logistic Regression n1 – Mismatch Entertainment by car	‘No children’ and ‘One child or more older than 12 years old’ Value: -0.78	‘No children’: 0.006 ‘One child or mode older than 12 years old’: -0.04	‘No children’
	‘Income: More than €131.000’ and ‘Origin: Central America’ Value: 0.70	‘Income: More than €131.000:’ 0.12 ‘Origin: Central America’: 0.033	‘Origin: Central America’
Logistic Regression n2 – Mismatch Entertainment walking	‘No children’ and ‘One child or more older than 12 years old’ Value: -0.79	‘No children’: 0.007 ‘One child or mode older than 12 years old’: -0.025	‘No children’
	‘Income: More than €131.000’ and ‘Origin: Central America’ Value: 0.70	‘Income: More than €131.000’:-0.05 ‘Origin: Central America’: 0.078	‘Income: More than €131.000’
Logistic Regression n3 – Mismatch Grocery stores by car	‘No children’ and ‘One child or more older than 12 years old’ Value: -0.78	‘No children’: 0.22 ‘One child or mode older than 12 years old’: -0.12	‘One child or mode older than 12 years old’
	‘Income: More than €131.000’ and ‘Origin: Central America’ Value: 0.70	‘Income: More than €131.000’: 0,1 ‘Origin: Central America’: 0,012	‘Origin: Central America’

Appendix E.3
ROC Curve and AUC Score for each Regression

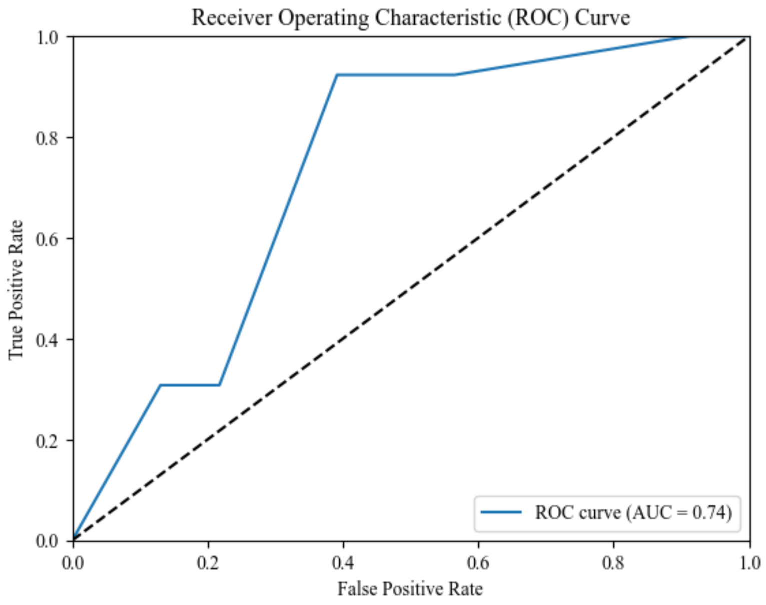


Figure E.3.1: ROC Curve and AUC Score of 1st Logistic Regression Analysis

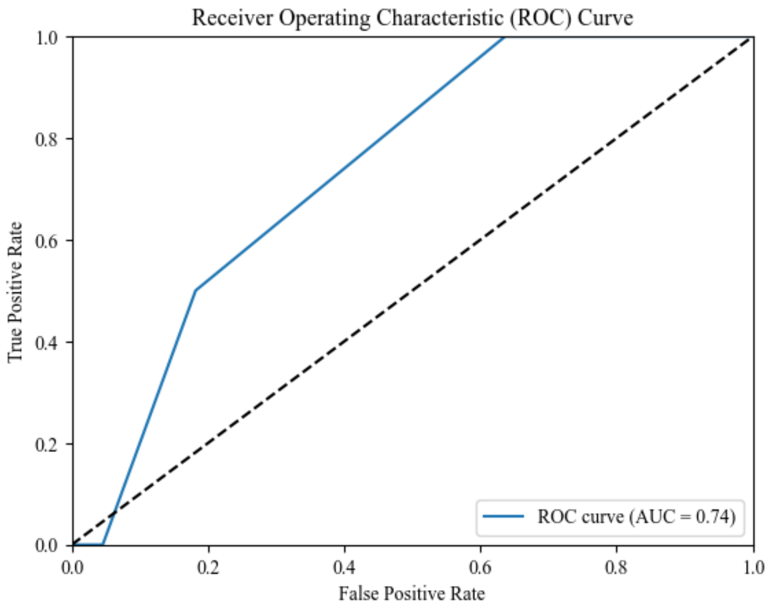


Figure E.3.2: ROC Curve and AUC Score of 2nd Logistic Regression Analysis

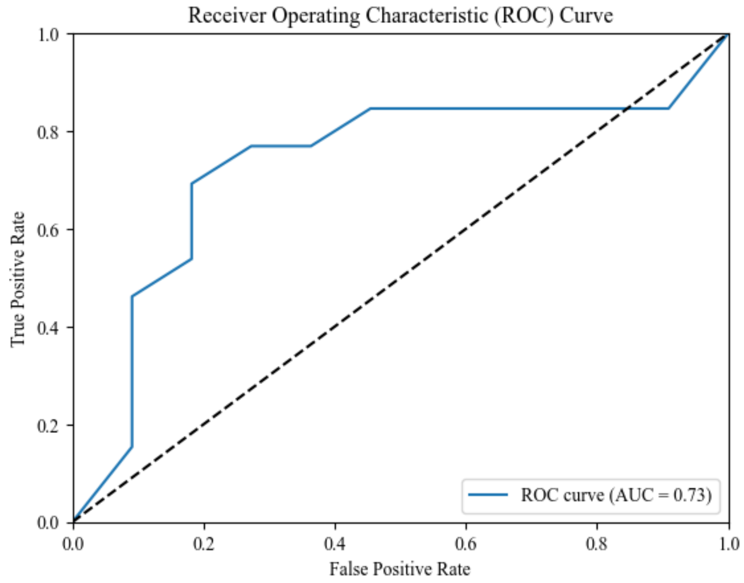


Figure E.3.3: ROC Curve and AUC Score of 3rd Logistic Regression Analysis