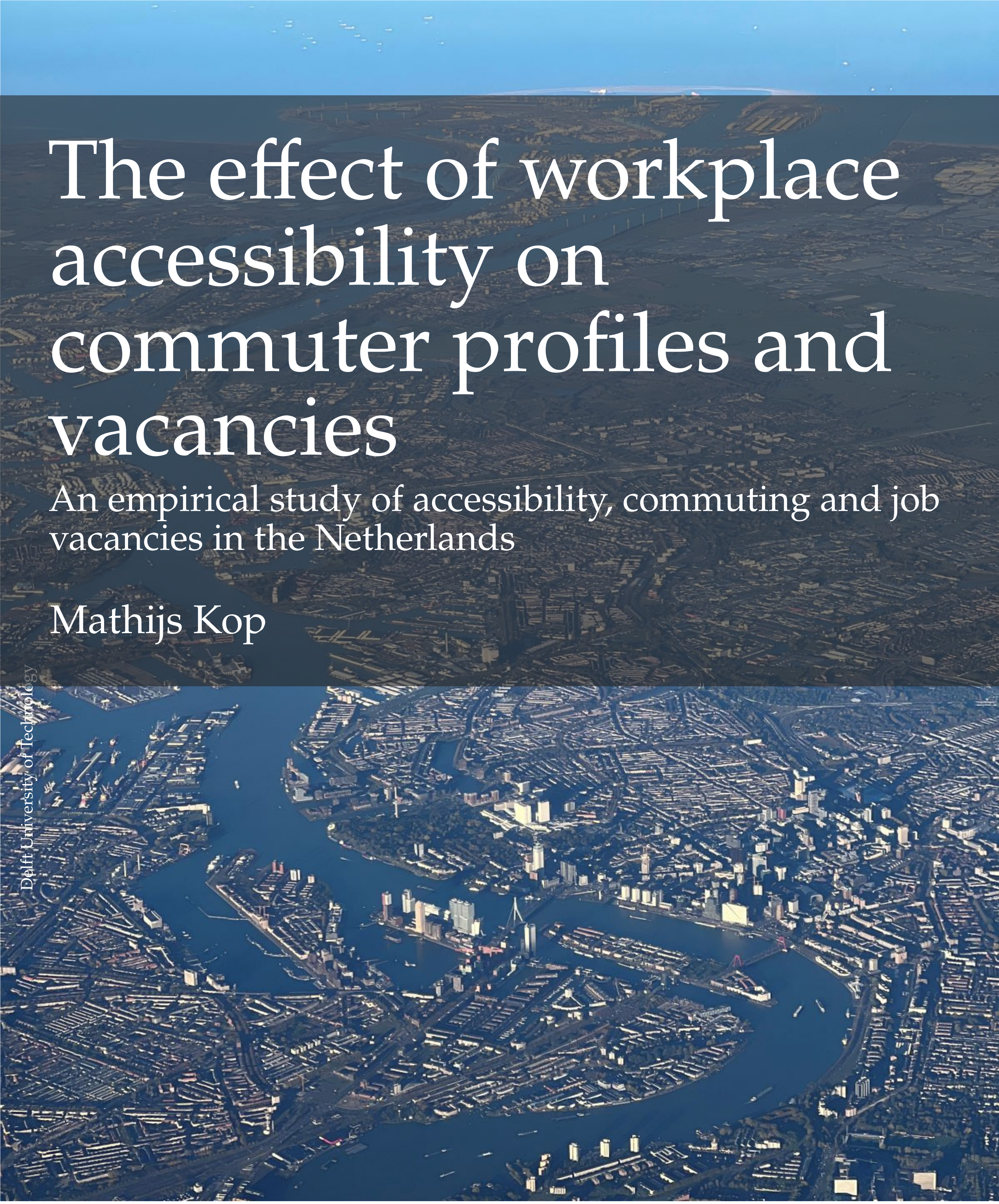


# The effect of workplace accessibility on commuter profiles and vacancies

## An empirical study of accessibility, commuting and job vacancies in the Netherlands

Mathijs Kop

Delft University of Technology

# The effect of workplace accessibility on commuter profiles and vacancies

## An empirical study of accessibility, commuting and job vacancies in the Netherlands


**Thesis committee**

| | |
|---|---|
| 1st Supervisor: | Dr. ir. M. Kroesen |
| 2nd Supervisor: | Dr. ir. C. Maat |
| External supervisor: | Ir. H.A.W. Voskamp |
| | |
| Faculty: | Civil engineering, Delft |



Cover: https://www.industrielinqs.nl/energiebesparing/2024/09/bedrijven-rotterdamse-haven-starten-pilot-grensoverschrijdend-energiebeheer/?gdpr=accept

# Preface

This thesis marks the conclusion of my Master's degree in Transport, Infrastructure & Logistics at the Delft University of Technology. Over the past months, Over the past months, I have done research on the relationship between commuter public transport accessibility and the labour market in collaboration with the Port of Rotterdam.

Researching such a vital economic hub has been a unique and rewarding experience, as well as working for the Port of Rotterdam. Its interesting environment served as a great setting for my research, and I am both proud and delighted that the study has provided new, meaningful insights into the challenges of vacancy fulfilment in the Rijnmond region.

This project would not have been possible without the help and guidance of several people. I would like to thank my graduation committee for their support throughout this period. I am especially grateful to Aral Voskamp from the Port of Rotterdam for providing me with this research topic, and for his weekly guidance and enthusiasm for the project. I would also like to thank my chair, Maarten Kroesen, for our biweekly meetings, his time and advice, and for helping me use all the possible research methods available in travel behaviour research. Additionally, I want to thank Kees Maat for his enthusiasm and his feedback and expertise during our important meetings.

I also want to thank Charlotte Hendrikse for providing me with information on the companies in the port area, and all the other colleagues on the 15th floor who made me feel welcome and appreciated.

Lastly, I would like to thank my family and friends. Thank you for listening to my struggles and celebrating my accomplishments. Your support was the motivation I needed to cross the finish line.

*Mathijs Kop*
*Wednesday 18$^{th}$ February, 2026*

# Abstract

Many employers face persistent difficulties filling their vacancies in the current tight labour markets. While the shift towards sustainability and automation increases the demand for specialized skills, finding qualified candidates is only part of the problem. Physical location and accessibility are just as important. Even if the right people are available, a vacancy might remain hard-to-fill if the workplace is simply too difficult to reach. In economic clusters like the Port of Rotterdam, this challenge is particularly prevalent, with accessibility perceived as a recruitment barrier by 88% of employers. When workplaces are poorly connected, labour shortages may not only reflect a lack of qualified candidates but also a spatial mismatch between job locations and the groups of workers able or willing to reach them. This thesis investigates how public transport workplace accessibility functions as a selective mechanism that shapes commuter profiles and labour market outcomes, exploring whether the recruitment frictions observed in the Rijnmond region are unique to this industrial area or a broader national spatial mismatch. The study approaches accessibility from the perspective of the employer. In contrast to much of the existing literature, which primarily focuses on residential accessibility and individual employment outcomes, this research centres on the accessibility of workplace locations and links it directly to vacancies and vacancy fulfilment. This thesis answers the question:

**To what extent does public transport workplace accessibility shape commuter profiles, and how does this relationship help explain regional differences in vacancy fulfilment in the Rijnmond region?**

The research combines a literature review with quantitative spatial analysis at PC4 level. A national Latent Class Cluster Analysis (LCCA) is used to identify distinct commuter and workplace profiles based on socio-demographics and travel behaviour. This is followed by stepwise regression models in the Rijnmond region, examining how public transport accessibility influences two key labour market dimensions: vacancy duration and vacancy rate.

The results show that workplace accessibility is fundamentally associated ($R^2$ = 8%) with the type of commuter profiles that workplaces attract. Latent Class Cluster Analysis identified six distinct socio-demographic profiles across the Netherlands (table 5.2). Locations with high public transport accessibility are characterized by a higher share of highly educated workers (Clusters 5 and 6), greater public transport use, and lower car dependence. In contrast, poorly accessible workplaces attract more car-oriented and place-bound profiles (cluster 1 and 3), creating a modal mismatch for workers without car access. These patterns are not unique to the Port of Rotterdam as the port functions as an extreme example of a broader national trend where industrial and peripheral areas suffer from a structural connectivity gap.

The regression analysis for Rijnmond demonstrates that public transport accessibility is a significant predictor of labour market efficiency. Potential accessibility (Hansen index) explains 14.8% of the spatial variation in vacancy duration, increasing to 18.2% when accounting for the interaction with urban density (table 5.5). However, the effect is spatially heterogeneous: while better accessibility significantly reduces vacancy duration in peripheral and industrial areas, the relationship reverses in dense urban contexts (figure 5.11). In cities, high accessibility is associated with longer vacancy durations, likely due to increased employer selectivity in larger labour pools. However, accessibility showed minimal explanatory power for vacancy rate, which appears to be driven by structural, sector-specific shortages rather than spatial connectivity.

This thesis concludes that labour shortages stem from spatial, modal, and utility mismatches that filter out urban talent vital for the energy transition. Accessibility acts as a selective filter: industrial car-dependency excludes the highly-educated talent pools concentrated in cities. Solving this requires prioritizing smart, demand-responsive mobility and relocating location flexible functions to urban port hubs rather than expanding traditional urban infrastructure.

# Contents

# 1 Introduction

Across many regions, employers face persistent difficulties in filling vacancies. These recruitment problems are commonly attributed to skills shortages, educational mismatches or demographic change. While these explanations are undoubtedly relevant, they often focus on individual qualifications and labour supply conditions, and less on how spatial conditions shape effective matching between jobs and workers. Jobs and workers are unevenly distributed across space, and whether a vacancy can be filled depends not only on who is qualified, but also on who can realistically reach the workplace. In this sense, workplace accessibility is not merely a matter of commuting convenience, but a structural condition shaping labour market outcomes. Workplace accessibility affects who is willing and able to apply for a job, which travel modes are feasible, and how large the effective labour pool of a firm is. When workplaces are poorly connected to the transport network, potential matches between employers and workers may fail to materialise, not because suitable candidates do not exist, but because reaching the job costs too much time, money, or is not possible due to a modal mismatch. Accessibility thus functions as a selective mechanism: it filters which socio-demographic groups can access certain jobs and which commuter profiles become dominant at specific locations. As a result, labour shortages may partly reflect spatial and modal mismatches rather than an absolute lack of labour supply.

These dynamics are particularly visible in large economic clusters with spatially concentrated employment. The Port of Rotterdam, the largest port in Europe, covers over 6,000 hectares and hosts approximately 3,000 companies. The port directly employs around 100,000 workers and indirectly supports an additional 565,000 jobs. Continuous growth, combined with the ambition to transition towards more sustainable and technologically advanced production processes, has increased the demand for labour. Recent labour market research shows that approximately 6% of all jobs in the port area are vacant, and that 65% of these vacancies are difficult to fill, particularly for highly educated technical occupations (HBO and WO levels) (Dekker & van der Toorn, 2024). These shortages are expected to intensify in the coming years (Joost Boudewijns, 2024). Employers in the port have expressed their concerns and perceive accessibility as a major recruitment barrier, with 88% identifying limited accessibility as a constraint (Dekker & van der Toorn, 2024). Large parts of the port area are only reachable by car, as conventional public transport services cover limited sections of the region (see Appendix B). This reliance on car travel has contributed to congestion, while public transport solutions are complicated by irregular working hours and long distances between terminals and residential areas. Cycling is often not a viable alternative due to these distances. Although innovative initiatives such as the Maasvlaktehopper (an on-demand shuttle service) address parts of the last-mile problem, substantial areas of the port remain poorly accessible by public transport.

While recruitment difficulties in the Port of Rotterdam are well documented, it remains unclear to what extent these problems are unique to this industrial context or reflect a broader, structural mechanism linking workplace accessibility, commuter behaviour and labour market outcomes. Highly educated workers, for example, are known to have a stronger preference for public transport use (Santos et al., 2013). If accessibility systematically filters the socio-demographic profiles of employees, it suggests that transport connectivity shapes the composition of the workforce in addition to traditional commuting patterns. In that case, persistent vacancies may arise even in the presence of sufficient labour supply, due to a mismatch between job locations and the groups of workers willing or able to reach them.

This thesis therefore investigates how public transport workplace accessibility functions as a selective mechanism in the labour market. It examines whether accessibility is systematically associated with distinct commuter and workplace profiles at the national level, and how these patterns help explain regional differences in vacancy fulfilment in the Rijnmond region. By approaching accessibility from the perspective of the workplace rather than the residence, this study aims to clarify whether the challenges observed in the Port of Rotterdam are a local problem or an extreme case of a more general spatial mismatch in the Dutch labour market.

## 1.1. Societal relevance

Workplace accessibility and its implications extend beyond the Port of Rotterdam and affect the labour market in a broader sense. Employers everywhere, particularly in the tech sector, struggle to attract qualified personnel. As vacancies remain unfilled, the productivity of the individual company is hindered and on top of that, the wider economic competitiveness is threatened. This is the case for the Port of Rotterdam competing with other ports and for the Netherlands competing with other countries. Limited knowledge can constrain growth and delay transitions to a more sustainable and digitalised economy. If these shortages are driven by accessibility constraints rather than absolute labour scarcity, traditional labour market or education policies alone will be insufficient to resolve them. Recent work highlights that accessibility is increasingly recognised as a key metric in transport planning, shifting attention away from purely mobility-based approaches (Handy, 2020). As Handy argues, this shift reflects both a methodological and societal evolution, with accessibility offering a more comprehensive framework for linking transport infrastructure to economic and social outcomes. For workplace accessibility, this means that location and connectivity are not just logistical concerns but strategic factors shaping labour market dynamics.

These dynamics are also visible in the Dutch context. The *OECD Review on Local Job Creation for Amsterdam and Dutch Cities* reports persistent difficulties in filling vacancies, pointing to skill mismatches and a very tight labour market as key explanations (Organisation for Economic Co-operation and Development (OECD), 2023). At the same time, the report emphasises that demographic and spatial differences across regions require locally tailored solutions. This raises the question whether part of the observed tightness is structurally reinforced by accessibility constraints that limit who can realistically participate in local labour markets.

Simultaneously, commuting patterns are also directly linked to national discussions regarding sustainability. Climate targets have been set by the Dutch government which require the reduction of car usage, yet not all locations have the necessary public transport, cycling roads and pavements to facilitate change in travel behaviour. Understanding how accessibility changes the modal split can aid in creating more effective mobility policies. Getting sustainable public transport options to align with the needs of the labour market will make the climate targets more achievable, as well as relieve pressure from the already highly saturated road network.

Lastly, accessibility has recently become a topic of social equity. The reliance on cars to reach large economic centres can hinder certain groups such as young workers, low-income households or commuters with physical limitations. People may experience higher barriers or even be excluded from job opportunities. These barriers can translate into reduced job opportunities and reinforce existing socio-economic divides. Unequal access to the labour market not only affects individuals, but also weakens the inclusiveness of society as a whole (Lucas, 2012; Van Wee, 2016).

This study contributes to these societal debates by linking accessibility, commuting behaviour, and vacancy patterns. By identifying who can reach jobs under which conditions, the research provides insights relevant for employers, planners, and policymakers. In this way, this thesis bridges the gap between academic definitions of accessibility and the practical challenges of ensuring sustainable, competitive, and equitable labour markets (Boisjoly & El-Geneidy, 2017; Geurs & van Wee, 2004).

## 1.2. Scientific relevance

While accessibility has been extensively explored in transport geography, the literature offers a wide array of definitions, each tailored to different analytical perspectives. The links between land-use, accessibility, and human activities have been studied for several decades, but research has predominantly focused on the urban scale (Qu et al., 2025). Furthermore, within labour market studies, accessibility is typically analysed from the commuter's perspective, while the standpoint of the company remains less prominent. Additionally, accessibility has rarely been researched as a predictor for open vacancies. Studies have looked into job probabilities (Bastiaanssen et al., 2022, 2025; van Lent et al., 2025) in combination with accessibility, but the employer's perspective in combination with labour market outcomes remains largely unexplored.

Recent directions in accessibility research show that accessibility cannot be generalised over the whole population, as it is heavily dependent on socio-demographic factors such as age, income, education, and car ownership (Boisjoly et al., 2019; Meester et al., 2024). Relating the current challenges in attracting specific workers to the workplace accessibility of these specific commuter profiles can provide new insights into the interplay between accessibility and the labour market. Furthermore, research into persistent vacancies concludes that location is a significant factor in explaining open vacancies, suggesting that a more sophisticated accessibility indicator might yield even greater explanatory power (Dossche et al., 2025).

Finally, the relationship between accessibility and modal split is highly complex and shaped by underlying socio-demographic factors (Nakshi et al., 2025). The spatial heterogeneity inherent in travel behaviour also remains underexplored, particularly regarding how accessibility translates into concrete mode choice outcomes (Cheng et al., 2020; Geurs & van Wee, 2004). To address these interconnected gaps, this research investigates spatial patterns in accessibility, modal split, and labour market outcomes on a national scale, complemented by an in-depth case study in the Rijnmond area.

## 1.3. Research Questions

The aim of this research is to investigate how workplace accessibility shapes commuting patterns and labour market outcomes in the Netherlands, and how spatial differences in accessibility contribute to regional disparities in vacancy fulfilment. This study approaches accessibility as a structural factor that interacts with socio-demographic characteristics, urban context, modal split and labour market outcomes. To that extent, the following research question is formulated:

*To what extent does public transport workplace accessibility shape commuter profiles, and how does this relationship help explain regional differences in vacancy fulfilment in Rijnmond?*

To answer the research question, the following sub-questions are formulated:

1. *What does existing research reveal about workplace accessibility and the relationship with socio-demographics, modal split and labour market outcomes?*
   Evaluating the empirical findings requires a strong theoretical foundation regarding the interactions between accessibility, socio-demographic characteristics, modal split, and the labour market. Previous research provides indications of how factors such as education, income, age and household composition influence commuting behaviour and job location choice. Being aware of these relationships helps to guide the empirical analysis and ensures that observed patterns can be interpreted within an appropriate theoretical and empirical context.
2. *Which socio-demographic commuter profiles can be identified across Dutch postcodes, and how is public transport accessibility associated with these profiles?*
   Commuting behaviour and labour market outcomes are shaped by differences in spatial and socio-demographic conditions. Identifying distinct commuter and workplace profiles allows for a structured exploration of this heterogeneity. Establishing these profiles provides a framework for analysing how accessibility relates to different types of workplaces and commuting environments across the Netherlands.

3. *To what extent do the commuter and public transport accessibility characteristics of the Port of Rotterdam reflect a unique local situation or a broader national pattern?*
   Labour market frictions observed in the Port of Rotterdam are often interpreted as the result of local accessibility constraints. However, similar industrial and peripheral employment locations exist elsewhere in the Netherlands. This question is included to determine whether the patterns observed in the port region are exceptional or indicative of a broader national mechanism. Addressing this distinction is important for assessing the general relevance of the findings.
4. *How does workplace public transport accessibility influence the spatial variation in vacancy duration and market tightness?*
   Spatial differences in vacancy outcomes suggest that the ability of employers to attract workers varies across locations. Understanding whether and how these differences relate to accessibility is necessary for linking transport related factors to labour market outcomes. This question guides the analysis towards identifying spatial regularities in vacancy indicators and assessing whether accessibility significantly contributes to explaining regional variation in vacancy fulfilment.

## 1.4. Research methodology

The research adopts a mixed-method design, combining a literature review with quantitative data analysis. The literature review provides the conceptual and theoretical foundation for the study and serves three purposes. First, it establishes the theoretical framework by clarifying how accessibility has been defined and operationalised in transport and land-use research. Since accessibility is a multi-dimensional concept, reviewing existing approaches supports the selection of a suitable conceptualisation for this study. Second, the review identifies how socio-demographic characteristics such as age, gender, income, and education influence commuting behaviour and modal choice. These insights are essential for understanding differences across groups of workers and for interpreting the interaction between accessibility and actual travel behaviour. Third, the review synthesises methodological contributions on measuring accessibility and analysing spatial variation. This guidance is used to determine how national commuting patterns can be meaningfully compared with local accessibility outcomes in the Port of Rotterdam. Together, the literature provides the conceptual and methodological basis for linking workplace accessibility by public transport to commuting behaviour and labour market outcomes. The full search strategy and synthesis are elaborated in chapter 2.

Quantitative analysis proceeds in three stages. First, descriptive and spatial analyses are carried out to explore accessibility patterns, commuting behaviour, and vacancy distribution across the Netherlands and Rijnmond. Second, a Latent Class Cluster Analysis is conducted to reveal different socio-demographic commuter profiles for different postcodes, using accessibility measures and modal split as covariates. The socio-demographic variables on which the clusters are based are determined by the literature review. Latent Class Cluster Analysis is appropriate for this study because it identifies unobserved commuter groups based on shared patterns in socio-demographics, accessibility conditions and modal choices. Commuting behaviour is multidimensional, and LCCA allows these complex interactions to surface without imposing strong assumptions about group structure (Hagenaars & McCutcheon, 2002). The LCCA is conducted on national scale, and reveals whether the experienced commuter profiles in the port area are unique or part of a larger national mechanism. Lastly, the analysis zooms in on the Rijnmond region, where accessibility is directly linked to vacancy fulfilment through (spatial) regression analysis. Spatial autocorrelation is used to investigate whether accessibility can account for the spatial variance of the model. After that, multiple regression models are tested to understand the relationships between accessibility, address density, vacancy characteristics and the vacancy indicators.

The conceptual model on which this research is based is shown in figure 2.2. The relationship between commuter profiles and vacancies is not directly modelled due to data limitations. However, these two methods are treated as complementary steps in the analysis. While the regression quantifies the impact of accessibility on vacancy duration, the LCCA profiles provide socio-demographic and modal context. It is expected that locations with restricted, car-dependent profiles will correspond with the areas where the regression shows the most significant hiring delays. By linking these findings in the discussion, the research addresses how commuter profiles ultimately influence labour market efficiency. The full description of the used methods are further discussed in chapter 3.

The empirical analysis is conducted at the Postcode 4 (four-digit numerical postcodes, PC4) level and focuses on the accessibility of workplaces rather than residential locations. This spatial focus reflects the employer-oriented perspective of the study, in which recruitment outcomes are linked to the ability of workers to reach job locations. Accessibility heterogeneity is not captured accurately at municipality level or higher, as cities such as Rotterdam would be simplified to one area, even though there are large differences within these cities. While Qu et al. (2025) suggest that higher resolutions generally reduce measurement error, PC4 provides the necessary balance between spatial granularity and the availability of reliable socio-demographic and vacancy data. Accessibility is determined by calculating the public transport travel times between all the postcodes, allowing for the calculation of workplace accessibility by public transport. The various datasets used in this study have different granularity and some need to be aggregated to PC4 level. More information on the data and the preprocessing can be found in chapter 4.

## 1.5. Scope

The geographical scope of this research covers the Netherlands, though it employs varying spatial scales for its two main analytical components. These different scales are a result of data availability and the specific interest in the labour market area in which the Port of Rotterdam resides. The Latent Class Cluster Analysis is conducted at a national level. This ensures a robust analysis by which captures the full picture of Dutch commuter profiles and workplace environments. By analysing all Dutch postcodes, the study can determine whether the profiles found in Rijnmond are unique or reflect broader national trends. The regression analysis is focused on the Rijnmond region. This regional focus allows for a more granular examination of the labour market outcomes within the Rijnmond area, which is most relevant for the Port of Rotterdam. Finally, it should be noted that accessibility at the borders might be lower due to the absence of potential labour supply from Germany and Belgium. However, this is consistent with the research focus on Dutch policy and not significant for the results.

Secondly, the temporal scope differs across datasets. The ODiN data is merged between 2018-2023 in order to have more respondents and less postcodes with too little respondents. A study by KiM on the effects of COVID on commuter traffic revealed that although people travelled less and worked more from home, the modal split remained similarly distributed spatially. Therefore, the modal split is assumed to be constant over time. Only a slight difference in cycling was observed, which is related to the emergence and increased prevalence of the e-bike (KiM Netherlands Institute for Transport Policy Analysis, 2022). However, this assumption needs to be validated with the ODiN data. The travel time data is taken from the year 2025 and the vacancy data is from the years 2023-2025. Travel times change only when changes are made in the infrastructure. While temporary network disruptions, such as roadworks or rail closures, can momentarily alter travel times, these short-term fluctuations do not reflect the underlying structural accessibility that influences long-term commuting choices and vacancy duration. Therefore, utilizing the 2025 baseline provides a stable, robust reflection of the infrastructural reality that governed the labour market dynamics during the 2023–2025 period.

Furthermore, the analysis restricts its modal scope to public transport, assuming walking for both the first and last mile. While this excludes the integration of the bicycle in Dutch commuting, which would significantly expand the catchment area of public transport stops, this restriction deliberately isolates the structural performance of the public transport network itself. Additionally, the travel data is strictly limited to direct home-to-work commuting trips and potential trip chains. Analysing the currently employed population is a logical approach, as their revealed travel behaviour establishes a realistic baseline of the public transport constraints that future job seekers will also face . Because the study adopts the employer's perspective, it evaluates destination accessibility (the ability of a workplace to attract a workforce) rather than individual origin accessibility. Consequently, the research relies on objective indicators, such as calculated travel times and regional socio-demographics, rather than subjective perceptions of safety, comfort, or travel quality. While this means individual psychological barriers to commuting are not captured, this objective approach provides the robust, reproducible framework required to model large-scale spatial relationships and open vacancy outcomes.

For the vacancies, only online vacancies (of which the sector and the required eduction level are known) are considered. Some sectors, such as the agricultural sector, might have low representation, but the most relevant sectors for this study (i.e. logistics, ICT, technical) post the majority of their vacancies online. The data will be aggregated to postcode 4, so it has good spatial coverage, and matches the other data sources. Besides job accessibility, there are many factors that can contribute to filling vacancies. Factors such as salary, labour conditions and level of specialisation required will also influence the open vacancies. The labour market is a complex system which can never be fully explained by job accessibility. This research aims to explore the role of job accessibility in shaping the open vacancies by controlling for as many factors as possible. However, controlling for factors such as salary and labour conditions will not be possible due to data limitations and these influences are therefore considered to be outside of the scope.

Lastly, the methods used in this study are descriptive, exploratory and correlational. The combination of Latent Class Cluster Analysis, spatial analysis and regression analysis allows the study to reveal patterns between accessibility, commuter profiles and vacancies, but it does not establish causal relationships. There is always the possibility of endogenous relationship which cannot be fully ruled out. For example, people who own a car will live in places which are less accessible by public transport. It cannot be clearly determined which variable is causing which. Also, it could be that there are unobserved factors influencing both accessibility and public transport share. Therefore, the relationships examined in this study are theory-driven and grounded in existing literature, allowing the observed associations to be meaningfully interpreted within an established conceptual framework.

## 1.6. Outline

The report starts in chapter 2 with a comprehensive review of the most important relevant literature on accessibility, commuter behaviour, socio-demographics and the connections to the labour market. This is followed by a detailed description of the data analysis techniques used in chapter 3. Chapter 4 describes the data that will be used throughout the research and how it is first prepared for analysis. These data analysis techniques are applied and the results are shown and discussed in chapter 5. Lastly, the research is concluded in chapter 6, as well as their implications and recommendations for further research.

# 2
# Literature review

To evaluate the current state of research on accessibility, travel behaviour, and their interactions with the labour market, a comprehensive literature review was conducted. The primary search was executed using Scopus, and some additional papers were found on Google Scholar. Some keywords that were combined to find these papers were "job", "vacancy", "accessibility", "travel behaviour", "commuter", "modal split", "mode choice", "socio-demographics", "equity" and "employer". Usually, review papers were first selected to limit the amount of papers. Additionally, forward and backward snowballing was used to find more relevant papers.

The remainder of this chapter is structured to narrow down the broad concept of accessibility to the specific research gap addressed in this study. First, the theoretical background of accessibility is discussed to establish the location based framework utilized in this research (2.1). Following this conceptual foundation, the review explores the relationship between accessibility and modal split (2.2), highlighting how spatial configurations and public transport provisions shape actual travel behaviour. The dimension of transport equity is then introduced, emphasizing the socio-demographic barriers that prevent certain groups from fully utilizing the transport network (2.3). Subsequently, these spatial and social dynamics are connected to the labour market (2.4). Here, the focus shifts from the traditional commuter perspective to the underexplored employer perspective, specifically examining spatial mismatch and vacancy durations. Finally, the chapter synthesises these findings to explicitly define the research gap (2.5), resulting in a conceptual model (2.6) that forms the theoretical basis for the methodology applied in the subsequent chapters.

## 2.1. Accessibility as a concept

Building on the themes introduced earlier, it is important to place accessibility to work within a broader picture. While the academic literature offers a wide range of definitions and measures, planners often struggle to translate these into operational indicators. Boisjoly and El-Geneidy (2017) show that although most metropolitan transportation plans formally adopt accessibility as an objective, few clearly define it or use accessibility-based indicators in decision-making. Instead, traditional mobility metrics such as speed or congestion still dominate. Attempt have been made to make accessibility operational such as the MAT model (Patterson et al., 2024). The MAT model provides accessibility based targets to assess whether infrastructure investment and spatial development plans move regions closer to achieving equitable accessibility. However, the application of such models in policy decision making is still limited.

To understand why accessibility is hard to operationalise, the definitions need to be clarified. Accessibility was properly defined for the first time by Hansen (1959) as a measure of potential opportunities. Prior to this new definition, accessibility was merely seen as the costs (in time or distance) of getting to a location. He argued that accessibility was the missing link between transport systems and land use patterns. The paper produced the first reproducible formula based for accessibility based on the amount of opportunities, travel costs and a declining impedance function (3.1). Building on this, Ingram (1971) developed a conceptual framework distinguishing relative and integral accessibility, emphasising the multidimensional nature of accessibility as a function of opportunity, travel impedance, and location attributes. While largely conceptual, Ingram's formulation laid the groundwork for empirical measurement. Morris et al. (1979) operationalised these concepts, developing two widely-used formulas: the cumulative-opportunity measure, which counts the number of opportunities reachable within a given travel cost threshold, and the gravity-based measure, which weights opportunities by a distance-decay function. These contributions established both the theoretical and methodological foundation for modern accessibility research. Shen (1998) introduced a gravity-based model that accounts for job competition among workers commuting by different modes, offering a more nuanced understanding of employment accessibility. Later, Wegener and Fürst (1999) systemised the accessibility relationship in the form of the land use–transport feedback cycle, in which accessibility mediates between transport systems, spatial development, and activity patterns. They explain that accessibility reflects the extent to which locations and activities can be reached, and it is shaped by both the transport system and the spatial distribution of activities. In this framework, land use and accessibility form a dynamic interaction: changes in accessibility influence where activities (including work) are located, while changes in land use in turn alter accessibility patterns. Accessibility does not only depend on the physical distance, but also on the quality and availability of infrastructure. Travel behaviour is shaped by the available opportunities as well as the available transport. These dynamics are shown in figure 2.1.

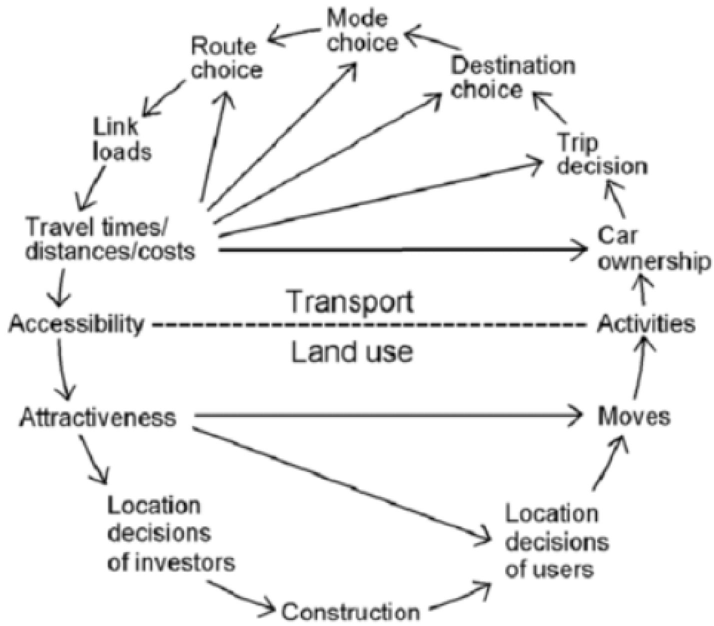



**Figure 2.1:** The land use transport feedback cycle (Wegener & Fürst, 1999)

This dynamic interaction between accessibility and land use has been empirically confirmed in the Dutch context. Zondag et al. (2014) developed a national land-use and transportation interaction model for the Netherlands incorporating a logsu m accessibility component. Their results show that accessibility improvements have a positive influence on the spatial distribution of residences and firms, but that the effects are constrained due to the already dense transport network and compact urban form. This illustrates how accessibility models behave differently in highly connected networks, such as those found in the Netherlands, compared to more car-oriented contexts.

Geurs and Van Wee broadened the definition of accessibility and linked it further to policy-making. They define accessibility as "the extent to which land-use and transport systems enable (groups of) individuals to reach activities or destinations by means of a (combination of) transport mode(s)" (Geurs & van Wee, 2004, p. 128). Accessibility is comprised of multiple components. These are:

- Land use: This component reflects how the land is used and what it has to offer. In the case of the Port of Rotterdam, 3000 companies with their respective location. Also, the demand for these locations from certain origin locations (residences) and lastly the confrontation of supply and demand is incorporated.

- Transportation: The transport system is expressed as "the dis-utility for an individual to cover the distance between an origin and a destination using a specific transport mode" (Geurs & van Wee, 2004, p. 128). Cost, effort and time are all considered in this component. Again, supply and demand are pivotal parts of the component. Supply is the location and characteristics of the infrastructure, whereas the demand relates to the passengers (and freight)
- Temporal component: This component is about the temporal constraints, which considers the availability of opportunities and individuals.
- Individual component: The last component is centred around the individual and its needs, skills and opportunities. In the case of an employee, it could be the education level and the type of job this person want to have.

Geurs and van Wee (2004) argue that effective evaluation require considering all four components simultaneously. Some accessibility measures are easy to interpret, but may overlook certain nuances. Complex measures might capture reality better but are harder to implement. Also, accessibility is not uniform over the population. Socio-demographic characteristics, preferences and constraints matter, which is why the temporal and individual component are important to consider. Following Geurs and van Wee (2004), measures can be grouped into infrastructure-based, location-based, person-based, and utility-based. This research adopts a location-based approach, consistent with Hansen's gravity formulation and subsequent refinements, as these models capture the 4 components of accessibility and are useful for spatial comparisons. The objective is to relate accessibility patterns to modal split and labour market outcomes at the postcode level, this definition is the most applicable.

Although the accessibility definition has been decided, accessibility cannot be determined yet. The accessibility formula contains locations, opportunities, travel costs and an impedance function. Accessibility should be considered mode-specific(Geurs & van Wee, 2004), as each mode has its own infrastructure, travel costs and impedance. Verma and Ukkusuri (2025) demonstrate that travel mode, trip purpose, and socio-demographic characteristics significantly affect accessibility estimates. The paper uses survey data to fit multiple impedance functions to the data to show that the differences mentioned above require different impedance functions and parameters. Most functions are a good fit for the impedance function, so it depends on the context and data which one should be picked. Therefore, it should be empirically tested which function with which parameters is best suited for the data.

Accessibility is often computed as a static measure at certain point in time. However, some studies highlight that transit service and job availability vary across the day. Boisjoly and El-Geneidy (2016) compared constant, static, and dynamic accessibility measures and found them to be highly correlated, which suggests that for many planning purposes the simpler static accessibility may suffice. Pfertner et al. (2023) propose a model for computing place based accessibility using a grid approach. They calculate 30 minutes isochrones for multiple modes and make relative comparisons. This destination based approach was rarely investigated prior to this research and only in the context of jobs-housing-balance. This supports the decision in this research to adopt location-based accessibility measures while taking time fluctuations into account.

## 2.2. Accessibility and modal split

Now that the accessibility as a concept has been defined, the question arises how these measures relate to actual travel behaviour. In transport research, accessibility is often operationalised through its impact on travel choices, in particular in the context of commuting. Understanding this behavioural link is important to see accessibility as something larger than an abstract indicator but as a reason for daily mobility decisions. Van Wee (2011) makes the distinction between the environment and accessibility when explaining travel behaviour. On the one hand, you can argue that the built environment affects how people travel (density, land-use mix, public transport supply). On the other hand, you can argue that people travel differently because accessibility changes the opportunities and costs. This makes it difficult to determine causality. Do individuals drive less because of the density of their neighborhood, the availability of public transport, or due to residential self-selection? While these factors are closely intertwined, Van Wee (2011) argues that accessibility remains the fundamental driver of travel behaviour.

Næss and Lyssand Sandberg (1996) investigated the accessibility of workplaces and how that affects the modal split in urban environments. Workplace location turned out to strongly affect the commuting modal split as peripheral, low-density parts of the urban area tend to have higher car-driving shares. This finding supports the hypothesis that accessibility constraints shape mode choice at the workplace level, not just the residential level. The public transport share is further investigated by Owen and Levinson (2015) by applying a continuous accessibility measure. They use binominal logit models at the aggregate area level to estimate the modal choice probabilities. Results show that accessibility is strongly associated with public transport share. The paper highlights the importance of mode specific accessibility indicators, but also the challenge of comparability between modes. The study also shows that even when controlling for socio-economic and demographic characteristics, accessibility measures remain both statistically and practically significant predictors of spatial and labour market outcomes. This underlines the independent explanatory power of accessibility-based modelling approaches.

In the Dutch context, similar relationships have been observed. Susilo and Maat (2007) analysed commuting behaviour in the Netherlands and found that urban form and built environment characteristics consistently influence modal split, even in a country with a dense transport network and high public transport and cycling quality. Their results show that spatial variation remains a key determinant of commuting mode choice: people living and working in more accessible and denser areas are far more likely to use public transport or cycling. More recent evidence by Hamersma and Roeleven (2024) confirms that people can perceive accessibility in different ways. They found that acceptable travel times differ by mode, purpose, and socio-demographics, with individuals generally accepting longer travel times for public transport than for car. This implies that accessibility affects mode choice not only through the physical transport network, but also through perceived opportunity and individual attitudes towards travel time and convenience.

Recent work highlights that the relationship between accessibility and travel behaviour is not uniform across space. Cheng et al. (2020) analyse commuting in Xiamen, China, using a probit–structural equation modelling framework that distinguishes between urban and suburban origins and different origin–destination types. Their results demonstrate that improvements in public transport accessibility have a substantially stronger effect on transit mode choice in suburban areas. In these contexts, relatively small changes in accessibility can meaningfully alter commuting behaviour, as baseline public transport provision is limited and marginal improvements reduce reliance on private cars. In contrast, in dense urban areas where public transport networks are already well developed, the sensitivity of mode choice to accessibility changes is considerably weaker. Here, accessibility is less of a binding constraint, as most locations already meet a minimum threshold of service. As a result, further improvements yield diminishing behavioural responses. This finding suggests the presence of a non-linear relationship between accessibility and travel behaviour, mediated by urban density and existing transport supply. The study also demonstrates that OD-level characteristics such as travel distance and whether the trip crosses spatial boundaries moderate this relationship. These results emphasise the importance of considering spatial heterogeneity in accessibility–behaviour models, rather than assuming uniform effects across all regions. Santos et al. (2013) dive into the factors that influence the modal split in medium-sized European cities. This research is done for 112 European cities with between 100.000 and 500.000 inhabitants. Results show that car share is positively associated with car ownership. Also economic factors such as GDP contribute to car usage. Higher population and employment densities were positively associated with increased use of public transport. Especially for public transport, quality and reliability are key in promoting usage.

When measuring accessibility, the choice of spatial scale is not merely a technical detail but a fundamental factor that influences the reliability of the results. This challenge is widely recognized in spatial analysis through two related methodological issues: the Modifiable Areal Unit Problem (MAUP) and the ecological fallacy. Recent research by Qu et al. (2025) highlights the MAUP by showing that spatial resolution has a direct impact on the measurement error of public transport accessibility indicators. They argue that:

- The Scale Effect: Lower resolutions (larger areas) often lead to an over- or underestimation of accessibility because they "smooth out" the local variations in transit service, such as the actual walking distance to a bus stop or train station.
- The First- and Last-Mile Nuance: In public transport, the error is particularly sensitive to scale because the first- and last mile stages (walking) occur at a very granular level. Aggregating data to

higher levels can obscure the fact that one side of a postcode might have excellent access while the other is practically disconnected.

Closely related to the MAUP is the risk of the ecological fallacy, which occurs when researchers infer individual-level behaviour or firm-level outcomes solely from aggregated spatial data (Firebaugh, 2001). In the context of accessibility and labour markets, this implies that while a specific postcode may exhibit high average accessibility and short average vacancy durations, one cannot deterministically conclude that every individual firm within that zone experiences identical recruitment success. Firm-level variations, such as specific job requirements or company reputation, are often obscured within the aggregated zonal average. Acknowledging both the MAUP and the ecological fallacy is therefore essential when translating spatial accessibility metrics into labour market expectations.

Taken together, the reviewed studies demonstrate that accessibility significantly shapes commuting mode choices across different spatial and socio-economic contexts. However, most research has primarily focused on individual travel behaviour or variations within urban context, often overlooking how workplace accessibility interacts with mode share on a larger scale when different urbanity levels are involved.

## 2.3. Accessibility and equity

Accessibility and commuting behaviour are not evenly distributed across groups and locations. Differences in socio-demographics between people results in some groups facing more barriers to mobility than others. This perspective looks into the fairness of accessibility and potential social inequalities that it can cause. Van Wee (2016) discusses the challenges in accessibility research and emphasises that there are multiple questions in this area. Lucas (2012) provides a comprehensive review of research on social exclusion in transport-related context. Accessibility inequities are framed as a core dimension of transport equity. The paper highlights that transport disadvantage affects socially excluded groups disproportionately, limiting their access to employment, education, healthcare, and social participation. Lucas (2012) emphasizes that accessibility equity is not only a technical issue but a deeply social one, requiring integrated policy approaches that address spatial, economic, and social barriers. Pereira et al. (2016) have a similar view and argue that equity in transportation should be framed through the lens of distributive justice, where accessibility is understood as a human capability and the real freedom to participate in society. In this view, policies should establish minimum standards of accessibility to essential destinations, ensuring that all individuals have the opportunity to participate fully in society.

In more recent literature, this transport equity is further highlighted. Accessibility constraints can suppress travel even when potential opportunities exist. a research in Sweden shows that limited public transport provision often results in "suppressed travel," particularly among young commuters who face time and transport-related barriers (Ryan et al., 2025). The investigation used a mixed method approach with a survey and a multimodal logistic regression model. However, the paper does question whether a large problem for a few people could or should be compensated and the levels of sufficiency should be different across socio-demographic groups.

Nakshi et al. (2025) investigates transport poverty, which is defined as the combined effect of social and transport disadvantages. They find that transport poverty lowers the feeling of agency over how, when and where you can travel. Also, the link between intention and behaviour becomes weaker. People will plan more on the day-to-day instead of developing routines, depending on availability and affordability of transport. Over time, transport poverty changes how people think about travel. They may lower expectations and accept less favourable modes. While most studies quantify accessibility inequalities using spatial or infrastructural indicators, recent work highlights that disparities also exist at the level of perception. Pot et al. (2021) argue that calculated accessibility (from travel times and opportunity distributions) often deviates from perceived accessibility, which reflects how individuals subjectively experience their ability to reach desired destinations. This mismatch is created by differences in awareness, experience, and evaluation of travel options. Consequently, accessibility inequities can persist even when objective conditions appear equal. Integrating perceived accessibility into transport equity analyses therefore reveals hidden layers of exclusion, complementing the material and attitudinal barriers identified by Nakshi et al. (2025). While this study relies on calculated accessibility metrics, the literature stresses that these capture only part of the accessibility experience. This is important to consider when interpreting the results of the models.

In conclusion, accessibility is also a matter of equity and is not equal for everyone. In light of this research, which explores the relationship between accessibility and vacancies, this perspective is particularly relevant since inequalities in accessibility can help explain spatial differences in labour market outcomes. Certain socio-demographic groups, such as low-income individuals, people without access to a car, or residents of peripheral areas, face structural barriers that limit their ability to reach suitable jobs. When these groups are excluded from parts of the labour market due to limited accessibility, employers in less accessible areas may struggle to attract applicants, leading to longer vacancy durations and higher vacancy rate. Understanding accessibility from an equity perspective has a broader purpose than equal opportunities for everyone. It offers a potential explanation for persistent mismatches between job demand and job supply, highlighting that improving accessibility is not only a social goal but also an economic one, directly linked to how efficiently labour markets function. The connection to the labour market in the literature is further explored in the next section.

## 2.4. Accessibility and the labour market

Concerns regarding accessibility equity highlight that accessibility can determine who can reach which opportunities. Labour market participation and employer recruitment are then likely to be influenced by accessibility. Questions of spatial mismatch and job accessibility have been posed before in literature. Fundamental literature on job accessibility and its effect on the labour market is the Spatial Mismatch Hypothesis (SMH), first introduced by Kain (1968). Kain argued that the physical distance between the residential locations of low-income workers (typically in inner cities) and the shifting locations of employment (increasingly in suburbs) leads to structural labour market imbalances. While this research focused on the US context, the underlying logic of the SMH is still applicable to modern areas like Rijnmond, where the port has expanded over the years to and beyond the coastline, away from the residential concentrations of the required workforce.

The SMH describes that these spatial barriers operate through three main channels: (1) high commuting costs that make distant jobs financially unviable, (2) limited information and social networks regarding vacancies in distant areas, and (3) a lack of adequate transportation options for those without a car. This theory provides a link to the employer's perspective; if a firm is located in an area suffering from spatial mismatch, it essentially faces a restricted labour pool. As a result, the "mismatch" is not only a problem for the job seeker but also for the firm, manifesting in persistent vacancies and higher recruitment frictions.

Another study explores the spatial mismatch in how transportation accessibility in Detroit affects the employment opportunities. Using 2000 census data and a gravity-based model of transportation accessibility, the study investigates differences in access to jobs among various neighbourhoods and demographic groups. The findings suggest that while inner-city neighbourhoods may have greater job accessibility, this advantage depends on car ownership. For individuals without access to a car, employment opportunities are significantly limited, highlighting a "modal mismatch" between job locations and the transportation options available to certain populations (Grengs, 2010).

Accessibility to workplaces strongly influences commute times, transit mode choice, and employment outcomes. Studies show that higher accessibility generally reduces commuting times, particularly for low-income workers, while also expanding the set of accessible job opportunities. Geurs and van Wee (2004) argue that for the case of job accessibility, research is necessary on job accessibility and competition effects in different situations and spatial contexts. Levinson (1998) finds that people living in job-rich residential areas and people working in housing-rich areas tend to have shorter commuting times. Another relevant paper, written by Boisjoly et al. (2019), looks at accessibility equity in commuting. The study uses multilevel mixed-effects statistical models to estimate commute duration as a function of accessibility measures for both car and public transport. The study reveals that low-income individuals are more susceptible to changes in accessibility. low-income people have more to gain and lose when accessibility is low or competition is high. Together, these findings suggest that accessibility functions not only as a spatial facilitator but also as a mechanism that enlarges socio-economic inequalities in the labour market.

Hu and Downs (2019) does not only look into the spatial distribution but also incorporates the time dimension in job accessibility. This is done by modifying Shen's Shen (1998) gravity-based accessibility definition to incorporate temporal fluctuations in the spatial distributions of job opportunities and workers. The proposed method can capture space and time effects but requires high computational power. This approach is not feasible for an analysis on national scale, as the authors indicate themselves, although their proposed solution is larger spatial areas.

Recent studies extend these findings by exploring sectoral and causal dimensions of accessibility and employment. Sharifiasl et al. (2024) investigates the association between transit accessibility and employment density by using gravity-based accessibility, first- and last-mile features and distance-to-infrastructure. In this Harris County case study, local bus accessibility was found to be associated with jobs in all sectors, whereas light rail was linked with employment densities in sectors such as real estate, finance, insurance, food and accommodation services. The proximity to public transport was also found to be a stronger predictor than accessibility itself (Sharifiasl et al., 2024). These patterns suggest that the industry has unique responses to central location proximity. A study by Meester et al. (2024) investigated including socio-demographics in potential job accessibility in low-car and car-free development areas. For this purpose, a stated choice survey was conducted in three Dutch cities and analysed using a Latent Class Logit regression model and Monte Carlo simulations. Notable differences were observed for starters and families in their accessibility levels. The paper concludes that socio-demographics are important to take into account in accessibility assessment as a general accessibility indicator can hide important inequalities, which is in line with research from Verma and Ukkusuri (2025), who emphasise that accessibility measures should reflect behavioural and social heterogeneity.

Another study, conducted in Great Britain, investigates accessibility to help people gain employment. The paper uses a gravity-based model for public transport accessibility and individual level data from a labour force survey. The model also controls for socio-demographic variables such as age, gender, education level, neighbourhood unemployment rate and others. They apply an instrumental variable approach to control for possible reverse causality or omitted variable bias. This is necessary because people who are employed may choose to live in areas with better accessibility, rather than accessibility causing employment. Also, areas with better accessibility might also have stronger economies, better schools or different housing markets, which affect both accessibility and employment. Therefore, simple regression will overestimate the effect of accessibility if these are not controlled for. The results showed that the causal effect between accessibility on employment probability was still significant and positive but smaller than without the instrumental variable. The effect was especially robust for low-income, low-education and low car-ownership groups (Bastiaanssen et al., 2022). These studies show that for different socio-demographic groups, consequences of poor accessibility are different and might limit their chances to get a job.

Bastiaanssen et al. (2025) conducted another similar research in the Netherlands. He explores the connection between accessibility and employment probabilities. Bastiaanssen et al. (2025) combine national administrative employment micro datasets with a novel public transport-bicycle accessibility metric. To address potential endogeneity between job accessibility, vehicle ownership, and employment status, an instrumental variable approach is employed. The study finds that better public transport job accessibility improves individual employment probabilities, particularly in metropolitan areas and among lower-educated groups. Jobs for higher-educated individuals tend to be concentrated in and around city centres, while jobs for lower-educated individuals are more often located outside these prime accessibility areas, reducing job accessibility for low-educated groups. The paper demonstrates that accessibility to suitable jobs increases the probability of employment, particularly for low-educated individuals, and that last-mile connectivity and education-job matching are critical factors (Bastiaanssen et al., 2025). van Lent et al. (2025) is another similar recent study that looks into the impact of teleworking in hybrid job accessibility. The research uses Hansen-based accessibility and combines this with competition effect for jobs (Shen, 1998) and the number of days working from home. The research uses agent based modelling to show that teleworking increases the accessibility of jobs by 40% but competition reduces it by 12%. However, this effect varies across job sectors. Four factors contribute strongly to the accessibility of hybrid jobs: job teleworkability, the use of mode to reach a job opportunity, the size of the pool of competitors and the supply of teleworkable jobs (van Lent et al., 2025). Together, these studies illustrate how accessibility interacts with evolving labour patterns and underline the growing importance of incorporating multimodal and hybrid work factors in accessibility research.

Despite extensive research on accessibility and employment, few studies have examined how accessibility shapes the dynamics of employer recruitment and vacancy duration. However, there are studies which investigate the problems with persistent vacancies. Rouwendal and Rouwendal (2025) investigate whether urban labour markets are inherently more dynamic by analysing how vacancies scale with city size. Their results show that dense urban labour markets have higher vacancy rates alongside higher job mobility, reflecting intensified matching and reallocation processes rather than persistent labour shortages. Thus, vacancies in urban areas do not only show unmet demand but also the dynamic nature of the labour market. The research normalises vacancies by employment to ensure comparability across regions of different size. This supports the use of vacancy rates relative to the number of jobs as an indicator of local labour market pressure, rather than relying on absolute vacancy counts. Their findings further suggest that urban density conditions the interpretation of vacancy outcomes, which is relevant when assessing how accessibility interacts with urban form in shaping spatial labour market performance.

Andrews et al. (2008) look into the determinants of vacancy duration by applying parametric and non-parametric duration models to vacancy micro data. One of their findings is that firms located in town centres (more accessible to potential employees) have a higher likelihood of being filled at any point in time and this outweighs the increased competition (Andrews et al., 2008). Similarly, Mangan and Trendle (2017) investigate the duration of vacancies and use an ordinal accessibility indicator for regions in Queensland, Australia. They find that better accessibility does result in shorter vacancy duration. However, accessibility is a qualitative measure in this research with limited granularity which does not provide much insights in spatial accessibility on a national scale. Lastly, a recent study by Dossche et al. (2025) uses a machine learning framework to analyse hard-to-fill vacancies and determine the most important factors. In line with prior research, the study concludes that the location (region) of the company was one of the important predictors. They used longitude and latitude for location and recognised this as a limitation of their research. They are convinced that using a more sophisticated accessibility measure can strongly improve the predictive performance of the model (Dossche et al., 2025). These literature findings further confirm the research gap that accessibility and open vacancies have not been researched to great extent, and show that this relationship is worth investigating in more detail on a greater scale.

Taken together, these studies indicate that accessibility operates along two interconnected pathways: (1) it affects individual employment probabilities by shaping feasible travel-to-work options, and (2) it influences employer recruitment outcomes by determining the size and composition of the potential labour pool. This dual mechanism links accessibility not only to modal split and socio-demographic exclusion but ultimately to vacancy fulfilment. The evidence suggests that when certain socio-demographic groups (e.g. low-income, low-educated, or car-less individuals) are effectively excluded from reaching job locations due to poor modal accessibility, employers face longer vacancy durations and sectoral shortages. These findings reinforce the idea that accessibility might not operate in isolation, but through its influence on travel behaviour and social inclusion. In subsequent sections, this relationship will be formalised into conceptual models that display two possible mechanisms that explain the relationship between accessibility and vacancies.

## 2.5. Synthesis and research gap

The literature demonstrates that accessibility is a key concept and affects commuting behaviour, spatial equity and employment outcomes. These effects operate through both physical constraints, such as travel time, cost, and modal availability and social mechanisms, including car ownership, income, education, and perceived travel possibilities. As a result, accessibility is not only a transport related concept, but also a key determinant of spatial and economic inclusion.

A substantial body of research demonstrates that higher accessibility is associated with increased use of public transport and active modes (Boisjoly & El-Geneidy, 2016; Cheng et al., 2020), shorter commuting times, and improved employment outcomes. These relationships are observed across different spatial contexts, including highly connected regions such as the Netherlands. At the same time, accessibility effects are spatially heterogeneous: peripheral and suburban areas are generally more sensitive to accessibility constraints than dense urban cores.

Another central insight from the literature is that accessibility is unevenly distributed across population groups. Socio-demographic characteristics strongly influence the extent to which individuals can benefit from transport infrastructure, resulting in accessibility inequalities and differentiated commuting behaviour (Lucas, 2012; Nakshi et al., 2025). These inequalities affect who can realistically reach certain job locations and therefore shape the effective labour supply available to employers.

However, most existing studies examine accessibility from the perspective of the individual. How accessible jobs influence a person's likelihood of employment or mode choice (Bastiaanssen et al., 2022; Levinson, 1998). The labour market from the employer perspective has received much less attention, in particular how a firm's location influences labour market outcomes or persistent vacancies. Research on vacancies is reliant on firm characteristics, job type, or regional market indicators. Even though accessibility is recognised as an influential parameter, it is often treated as a simple locational proxy (Andrews et al., 2008; Dossche et al., 2025; Mangan & Trendle, 2017). If included, it lacks spatial precision or is measured qualitatively. The relationship between vacancy fulfilment and accessibility has not been researched in isolation, which is why this connection remains poorly understood.

Therefore, the gap found in literature lies in the limited understanding of the relationship between accessibility and unfilled vacancies from a spatial and equity perspective. In this research, the gap will be addressed by quantifying accessibility and its effect on commuter profiles and vacancy fulfilment. This connects transport geography and the labour market from a unique perspective.

## 2.6. Conceptual model based on the literature

Based on the information presented in this chapter, a simplified conceptual model can be created to visualise what factors and relationships are highlighted in the literature.

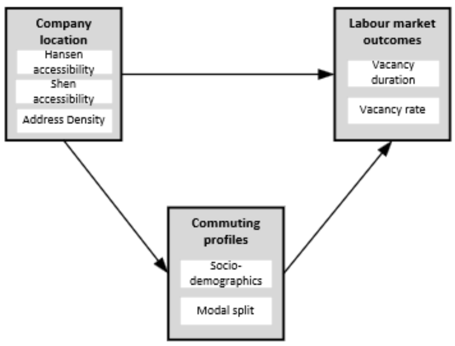



**Figure 2.2:** Conceptual model based on the literature

Figure 2.2 shows the conceptual model based on the literature. Higher accessibility, calculated through the Hansen (1959) or Shen (1998) measure, implies a reduced travel impedance, thus makes the location more attractive for potential employees. This direct link is supported by Andrews et al. (2008) and Mangan and Trendle (2017), who suggest that companies located in dense areas have reduced vacancy duration. Additionally, incorporating competition through the Shen's measure allows the model to account for the market friction caused by other firms competing for the same labour supply.

Beyond physical distance, the model explores an indirect effect through commuter profiles. This represents the individual component of accessibility (Geurs & van Wee, 2004). Literature indicates that accessibility is not a uniform field but is dependent on socio-demographics. Factors such as income, education, and car ownership determine who can realistically reach a location (Bastiaanssen et al., 2025; Lucas, 2012). Moreover, the available infrastructure shapes commuting behaviour. Peripheral areas often lead to car-dependency, whereas dense urban areas support public transport and cycling (Næss & Lyssand Sandberg, 1996; Susilo & Maat, 2007). This indirect path suggests that if a workplace is only accessible to specific groups (e.g., those with cars), it creates a modal mismatch (Grengs, 2010), which ultimately results in higher vacancy rates and recruitment difficulties for firms that cannot access a diverse labour pool.

This conceptualisation guides the analytical approach of this study. The indirect pathway, which links workplace accessibility to labour market outcomes via socio-demographics and commuting behaviour, is examined using Latent Class Cluster Analysis (LCCA). Previous studies highlight that commuter behaviour emerges from the interaction of multiple characteristics rather than from single variables in isolation. LCCA is therefore well suited to identify distinct commuter profiles, capturing how social inequalities, mobility constraints and accessibility conditions jointly shape which groups are able or willing to reach a specific job location. The direct pathway (from accessibility to vacancy fulfilment) will be tested using (spatial) regression to quantify the spatial effect of accessibility on vacancies, independent of socio-demographic or modal choice differences. This approach allows for quantifying the effect of accessibility on vacancies while controlling for vacancy characteristics and urban density. The combination of LCCA and (spatial) regression addresses the following research gaps: research focusing on individual commuting behaviour and social inequality, and studies examining spatial labour market dynamics from an employer perspective. Together, these complementary methods address both the social and spatial dimensions of accessibility as emphasised in the literature. A detailed explanation of these analytical techniques is provided in Chapter 3.

# 3
# Methodology

This chapter outlines the methodological approach used to analyse the relationship between workplace accessibility, commuter profiles and vacancy outcomes. A mixed-method strategy is deployed to analyse the effect of accessibility on commuter profiles and the labour market as shown in figure 3.1. The methodology combines an exploratory clustering approach with confirmatory regression analysis to capture both structural patterns in commuting behaviour and their implications for vacancy fulfilment.

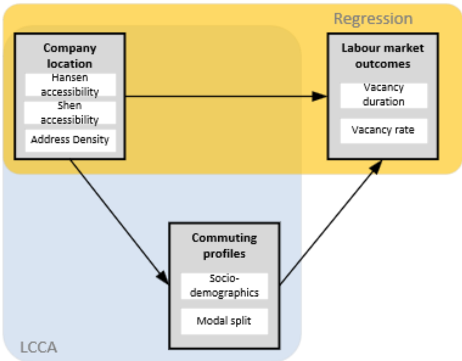



**Figure 3.1:** The conceptual model and the analyses used in this research

The analytical strategy is structured in three complementary parts, each addressing a specific dimension of the research question.

First, a Latent Class Cluster Analysis (LCCA) is applied to identify distinct commuter and workplace profiles across Dutch postcodes. The literature emphasises that accessibility, commuting behaviour and socio-demographic characteristics interact in complex ways and cannot be adequately captured by single indicators or linear relationships. LCCA allows for the identification of unobserved groups that share similar combinations of socio-demographic characteristics, modal split and accessibility conditions. The aim of this step is to uncover structural heterogeneity in commuter profiles and to assess whether workplaces with similar accessibility levels attract systematically different types of workers. By conducting this analysis at the national scale, the study evaluates whether the commuter characteristics observed in the Port of Rotterdam reflect a unique local configuration or a broader national pattern.

Second, spatial dependence is explicitly assessed to determine whether accessibility, vacancy outcomes and their relationships are influenced by spatial clustering or unobserved spatial processes. Both accessibility and labour market indicators are known to exhibit strong spatial structure, as neighbouring areas often share infrastructure, labour pools and economic characteristics. Ignoring spatial dependence may lead to biased inference and overestimation of the explanatory power of regression models.

Assessing spatial autocorrelation therefore serves two purposes: it evaluates whether spatial clustering is present in the outcome variables, and it examines whether accessibility captures part of this spatial structure. This step informs the subsequent regression analysis by clarifying whether spatial effects need to be explicitly modelled.

Third, regression models are estimated to quantify the relationship between public transport workplace accessibility and vacancy indicators, focusing on vacancy duration and vacancy rate. While the LCCA captures indirect mechanisms through socio-demographic composition and commuting behaviour, regression analysis is used to test the direct association between accessibility and labour market outcomes from an employer-oriented perspective. A stepwise modelling approach is adopted to assess how the estimated effect of accessibility changes when controlling for urban density and labour force composition, and to test whether the relationship between accessibility and vacancy fulfilment differs between urban and peripheral contexts. This approach allows the study to evaluate the magnitude, direction and spatial heterogeneity of accessibility effects on vacancy outcomes, directly addressing the last sub-question.

## 3.1. Theoretical framework of accessibility

Building on the four components of accessibility (land use, transport, temporal, and individual), this section formalizes the metrics used in the study. The used metrics are potential accessibility and competitive accessibility.

Potential accessibility, originally formulated by Hansen (1959), remains the most widely used metric because it effectively combines the transport and land-use components. It assumes that the attractiveness of a destination (or in this case, the availability of labour supply) decreases as travel time increases. The measure is suitable as it reflects the "potential" pool of workers. From an employer's perspective, it quantifies the total reachable labour force, which is the primary condition for any recruitment process. Additionally, it uses continuous decay, unlike contour measures, which means that workforce closer is weighted more heavily. In a tight labor market, simply being reachable is insufficient if thousands of other firms are competing for the same worker. To address this, Shen (1998) introduced a correction factor that accounts for demand side competition. By using both metrics, this research can distinguish between geographic isolation (low potential accessibility) and market competition (low competitive accessibility). This distinction is critical for answering the research question: it allows us to see if vacancy duration is driven by a lack of infrastructure (transport component) or by the proximity of competing job clusters (land-use/individual component).

**Accessibility using Hansen (1959)**

Accessibility in the potential definition, measures the extent to which opportunities at all locations can be reached from a given origin, discounted by a travel impedance function. The Hansen accessibility of workplace zone $i$ is defined as the weighted sum of all opportunities at home zones, discounted by their travel impedance:

$$A_i = \sum_j W_j \cdot f(c_{ij}) \tag{3.1}$$

- $A_i$ = Accessibility of location $i$ (company location),
- $W_j$ = Opportunities at location $j$ (number of people that are able to work residing at location j) ,
- $c_{ij}$ = Travel costs (time, distance or generalised costs) between $i$ and $j$ (travel time by public transport),
- $f(c_{ij})$ = the impedance function that gives the declining attractiveness of opportunities $W_j$ as the travel costs $c_{ij}$ increase.

The Hansen accessibility values are then computed via efficient matrix multiplication (dot product), where each row of the matrix $D$ corresponds to a workplace zone $i$, and each column corresponds to an origin $j$. The multiplication evaluates, for each destination, the total amount of reachable population weighted by the decay function. The outcome $A_i$ expresses the potential labour accessibility of workplace zone $i$. Higher values indicate a better location within the regional labour market.

**Accessibility Following Shen (1998)(competition effect)**
Potential accessibility as defined by Hansen looks purely at how many people can be reached in a considerable time. However, locations may also be more favourable if there are no other better options nearby. To illustrate this limitation of the traditional Hansen accessibility measure when applied to labour markets, consider a simplified system inspired by Shen (1998), but interpreted from the perspective of firms rather than households. Each zone contains both jobs and workers, as shown in Figure 3.2. Average intrazonal travel time is assumed to be 10 minutes for all zones. Interzonal travel times are 30 minutes between A and B, and 100 minutes between both A and C and B and C.

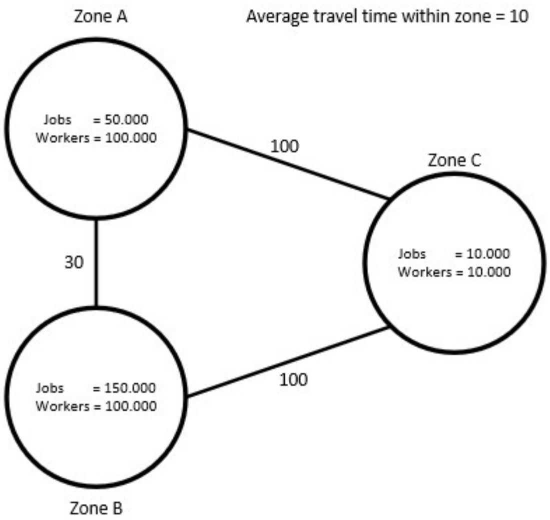



**Figure 3.2:** Example of a three zone system

Let the number of workers in each zone be denoted by $M_j$. Accessibility is measured using the Hansen potential accessibility formulation with $\lambda = 5.207 \times 10^{-5}$ and $\alpha = 2.412$ (4.2.1), where $C$ denotes travel time in minutes. Evaluating this function yields:

$$f(10) \approx 0.987, \tag{3.2}$$
$$f(30) \approx 0.827, \tag{3.3}$$
$$f(100) \approx 0.031. \tag{3.4}$$

The worker distributions are as follows: 100000 workers in zone A, 100000 in zone B, and 10000 in zone C. The Hansen accessibility for firms located in zone A is:

$$\begin{aligned} A_A &= 100{,}000 \cdot f(10) + 100{,}000 \cdot f(30) + 10{,}000 \cdot f(100) \\ &\approx 100{,}000 \cdot 0.987 + 100{,}000 \cdot 0.827 + 10{,}000 \cdot 0.031 \\ &\approx 181{,}647. \end{aligned} \tag{3.5}$$

The Hansen accessibility for firms located in the zone B is:

$$\begin{aligned} A_B &= 100{,}000 \cdot f(30) + 100{,}000 \cdot f(10) + 10{,}000 \cdot f(100) \\ &\approx 100{,}000 \cdot 0.987 + 100{,}000 \cdot 0.827 + 10{,}000 \cdot 0.031 \\ &\approx 181{,}647. \end{aligned} \tag{3.6}$$

Finally, the Hansen accessibility for firms located in zone C is:

$$\begin{aligned} A_C &= 10{,}000 \cdot f(10) + 100{,}000 \cdot f(100) + 100{,}000 \cdot f(100) \\ &\approx 10{,}000 \cdot 0.987 + 100{,}000 \cdot 0.031 + 100{,}000 \cdot 0.031 \\ &\approx 16{,}070. \end{aligned} \tag{3.7}$$

The calaculations show that zone A and zone B have equal accessibility. However, zone A has more workers than jobs, which should be more beneficial from the company perspective as there are still equal amounts of workers available, but less work to compete with.

In addition to the traditional Hansen accessibility indicator, a competitive accessibility measure can be constructed that accounts for the fact that firms compete for the same pool of workers. This approach follows the logic of Shen (1998), but applied from the perspective of firms rather than households. Workers represent the scarce resource, and jobs represent the competing destinations. Let $M_i$ denote the number of workers residing in origin zone $i$, and let $P_j$ denote the number of jobs located in destination zone $j$. The generalised travel impedance between $i$ and $j$ is denoted by $t_{ij}$, and the decay function $f(t_{ij})$ reflects declining interaction with increasing travel time. Workers in zone $j$ have options for jobs in multiple destinations. Thus, the workers in zone $j$ are weighted to the total attraction ($D_j$) of competing firms $k$.

$$D_j = \sum_k P_k \, f(t_{kj}) \tag{3.8}$$

This quantity serves as the normalising factor that distributes workers across all jobs that are reachable from zone $j$. A zone with more competing firms within reasonable distance (higher $D_j$) contributes fewer workers per job, reflecting competition among firms. The competitive accessibility of firm location $i$ is defined as the total effective labour supply that can be drawn from all origin zones:

$$A_i = \sum_j \frac{W_j \, f(t_{ji})}{D_j} \tag{3.9}$$

Therefore, Shen's formula scales the total available workers to other jobs that compete for these workers. In case of Zone A (figure 3.2), other companies are relatively further away (as the jobs within zone A are lower) than for zone B, which results in a slightly higher competitive accessibility for zone A. This measure expresses the labour pool available to firms in a way that accounts simultaneously for distance decay and for the competition created by other firms in the region. This definition might be more suitable for an area like Port of Rotterdam, where densely populated areas with a high potential workforce are close by, but there are also many competing jobs in these areas which might be preferable due to the better travel times, compared to the port area. Therefore, it is interesting to regard both measures in the analysis to investigate which one has more explanatory power.

### 3.1.1. Impedance function

A critical component in the operationalisation of both Hansen (1959) and Shen (1998) metrics is the impedance function, $f(c_{ij})$, which represents the behavioural core of accessibility. While travel time provides a physical measure of connectivity, the impedance function translates this into a measure of perceived distance. This study follows the logic that locations which are reachable within shorter notice are more important than locations which take long to reach. Therefore, it follows a distance-decay pattern where the marginal cost of an additional travel minute increases as the total commute time grows. There are multiple functions that can be used to predict this impedance, which will be fitted to travel times in section 4.2.1 to determine the parameters that reflect Dutch commuting. By applying this function, the accessibility indices move beyond purely spatial measures to reflect the realistic probability of a matching process occurring between a firm in zone $i$ and a worker in zone $j$.

## 3.2. Latent Class Cluster Analysis

In this study, Latent Class Cluster Analysis (LCCA) is employed to identify heterogeneous commuter profiles based on socio-demographic and household characteristics of workers. LCCA is a probabilistic clustering technique that assumes that the population consists of a limited number of latent (unobserved) groups, with each observational unit having a certain probability of belonging to each class. This reduces misclassification biases. Unlike traditional distance-based clustering methods (such as k-means), LCCA does not rely on metric assumptions or linearity, and it can naturally handle categorical and proportional variables. The main concept of LCCA is that a discrete latent variable is able to account for associations between different indicators, so that conditional to the latent class variable, these associations become insignificant (J. K. Vermunt & Magidson, 2002). This is known as the assumption of local independence. The goal is to find a model with an $x$ amount of latent classes that can describe the associations between the variables, without being overly complex. Statistical criteria like the Bayesian Information Criterion (BIC) can be used to decide on the trade-off between explanatory power and model complexity.

### 3.2.1. Conceptual model

The literature revealed that accessibility is not equal for everyone and can determine who can reach what opportunities. The LCCA aims to reveal these equity problems by identifying different commuter profiles which have different need for public transport accessibility. The conceptual model is shown in figure 3.3. The Latent Class Cluster Model is specified using the socio-demographic composition of workers at each destination postcode as the indicators that define the latent classes. The seven included indicators are:

- **Age:** Age is a key determinant of commuting behaviour and accessibility constraints. Younger individuals tend to have lower car ownership rates and are more dependent on public transport and active modes, while older workers often exhibit more stable commuting routines and higher car availability (Hamersma & Roeleven, 2024; Susilo & Maat, 2007). Age is also associated with differences in acceptable travel times and job mobility, which affects both mode choice and labour market participation.
- **Gender:** Gender differences in travel behaviour are well documented in transport literature. Women generally exhibit shorter commuting distances, higher reliance on public transport and active modes, and more complex trip chains related to care responsibilities (Lucas, 2012; Van Wee, 2016).
- **Income:** Income strongly conditions transport mode availability and sensitivity to travel costs. Higher-income individuals are more likely to own a car and tolerate longer commuting distances, while lower-income groups face stronger accessibility constraints (Boisjoly et al., 2019; Lucas, 2012).
- **Education:** Education level is associated with both job location patterns and commuting behaviour. Higher-educated workers are more likely to work in centrally located, highly accessible areas and to use public transport, while lower-educated jobs are often located in peripheral or industrial areas with lower accessibility (Bastiaanssen et al., 2025; Meester et al., 2024).
- **Car ownership:** Car ownership is one of the strongest predictors of commuting mode choice and job accessibility. Individuals without access to a car are significantly more constrained in their feasible job search area, particularly in peripheral locations (Grengs, 2010; Susilo & Maat, 2007).
- **Social participation:** Social participation, defined here as full-time employment, part-time employment or student status, captures structural differences in how often commuting is necessary, income stability and commuting constraints. Full-time workers typically exhibit more stable and longer commuting patterns, whereas part-timers often have other obligations which shorten their accepted travel time (Susilo & Maat, 2007). Students represent a distinct group with high reliance on public transport and cycling, lower car ownership, and greater sensitivity to travel time and cost (Hamersma & Roeleven, 2024).
- **Household configuration:** Household composition influences commuting behaviour through shared resources, care responsibilities and time constraints. Households with children, for example, tend to have shorter and more time-sensitive commutes, while single-person households show greater flexibility in mode choice and travel distance (Susilo & Maat, 2007; Van Wee, 2016).

The selection of socio-demographic indicators included in the analysis is grounded in existing literature on accessibility, commuting behaviour and labour market outcomes. These variables capture structural demographic and household differences that consistently shape travel behaviour and employment access, forming the basis for identifying distinct commuter profiles across company locations. Public transport accessibility is included as an active covariate, to include the theoretical mechanism linking accessibility to the formation of commuter profiles. As an active covariate, accessibility is allowed to influence the probability of class membership, reflecting the evidence that accessibility conditions shape both travel behaviour and the socio-demographic sorting of workers across locations.

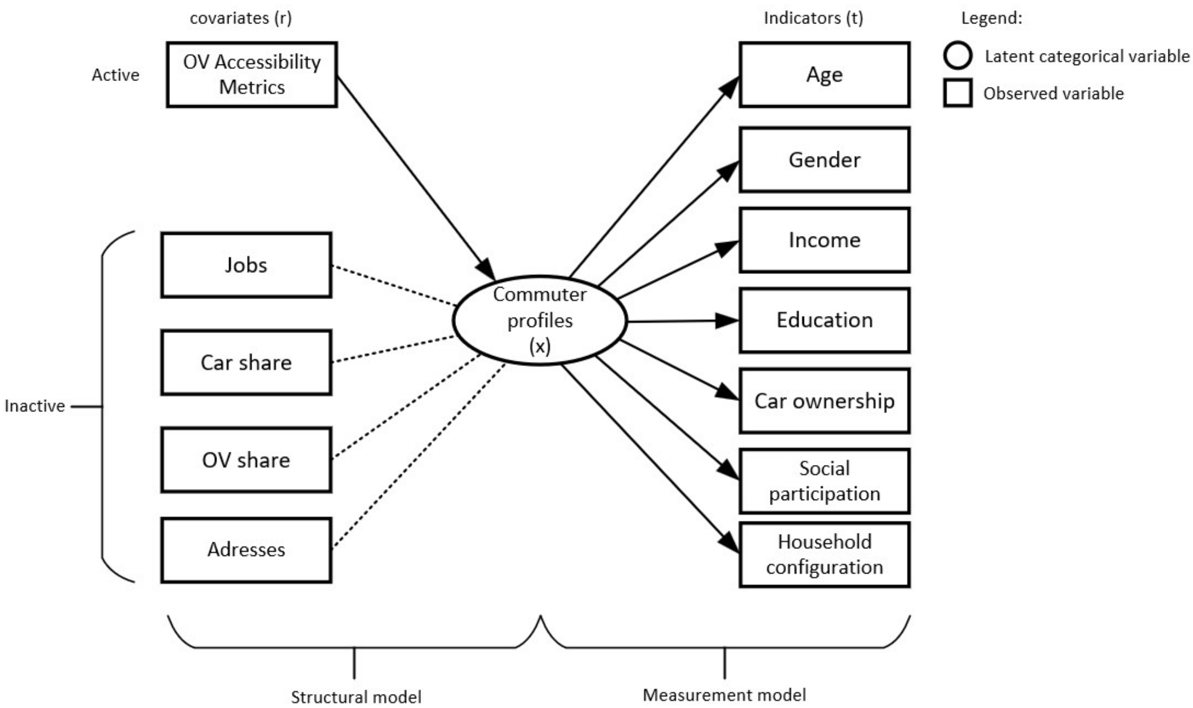



**Figure 3.3:** Graphical representation of the used latent class cluster model including covariates

Four additional contextual variables are used as inactive covariates. address density, commuting modal split (operationalised as area-level public transport share and car share) and the number of jobs at the location are included. These variables do not participate in the formation of the latent classes but are related to them after the clusters have been created. This allows for further assessment on how contextual and behavioural conditions vary across the socio-demographic profiles identified by the model. This specification preserves the conceptual order in which socio-demographic structure defines the classes. Accessibility shapes the likelihood of belonging to those classes, and urban context and modal outcomes are examined as external outcomes. Together, this latent class cluster structure enables the identification of latent worker profiles and their association with accessibility and mobility patterns, which is essential for evaluating potential accessibility equity issues across company locations.

The Latent Class Cluster Model (LCCM) assumes that the associations between $T$ observed indicator variables $\{y_{i1}, \ldots, y_{iT}\}$ for unit $i$ can be explained by a single nominal latent variable $x \in \{1, \ldots, K\}$ representing the latent classes (J. Vermunt & Magidson, 2005). In this study, the indicators are treated as continuous variables and modelled using class-specific normal densities.

Let $y_{it}$ denote the value of indicator $t$ for unit $i$. Conditional on class $x$, each indicator follows a Gaussian distribution with class-specific parameters. The density is:

$$P(y_{it} \mid x) = \frac{1}{\sqrt{2\pi\sigma_{t,x}^2}} \exp\left(-\frac{1}{2}\frac{(y_{it} - \mu_{t,x})^2}{\sigma_{t,x}^2}\right), \tag{3.10}$$

Let $z_{i1}, z_{i2}, \ldots, z_{iR}$ denote the $R$ covariates for unit $i$. In a Latent Class Cluster Model, covariates can be included to predict class membership through a multinomial logit model (J. K. Vermunt & Magidson, 2002). For active covariates, the probability that unit $i$ belongs to class $x$ is:

$$P(x \mid \mathbf{z}_i) = \frac{\exp\left(\gamma_x + \sum_{r=1}^{R} \gamma_{xr} z_{ir}\right)}{\sum_{x'=1}^{K} \exp\left(\gamma_{x'} + \sum_{r=1}^{R} \gamma_{x'r} z_{ir}\right)} \tag{3.11}$$

with one class (typically class 1) serving as the reference category where $\alpha_1 = 0$ and $\beta_1 = 0$ for identification.

Assuming local independence, the likelihood contribution of observation $i$ conditional on latent class $x$ is the product of the indicator densities. The probability of observing a certain pattern of variable outputs is given in equation 3.12.

$$L_i = \sum_{x=1}^{K} P(x \mid \mathbf{z}_i) \prod_{t=1}^{T} P(y_{it} \mid x). \tag{3.12}$$

Essentially, the latent class cluster model is made up of two probabilities: The probability of belonging in a certain class, given the covariate values ($P(x \mid \mathbf{z}_i)$), and the probability of the observed values given the latent class membership ($P(y_{it} \mid x)$).

## 3.3. Regression analysis

In order to relate workplace accessibility directly to labour market performance, this study focuses on two complementary vacancy based indicators: vacancy duration and the vacancy rate.

- Vacancy duration

  Reflects the average time required to fill an open position. In the labour economics literature, vacancy duration is commonly used as an indicator of matching efficiency between labour demand and supply, with longer durations suggesting greater frictions in the matching process as described by Mortensen and Pissarides (1994) and Petrongolo and Pissarides (2001). Spatial variation in vacancy duration can therefore reveal whether accessibility constraints are associated with difficulties in attracting suitable workers, even when vacancies exist.
- Vacancy rate = $\frac{\text{vacancies per year}}{\text{number of jobs}}$

  Captures the relative tightness of the local labour market. This formulation follows the logic used in official labour market statistics from CBS (Centraal Bureau voor de Statistiek, 2024) and the paper by Rouwendal and Rouwendal (2025), where the vacancy rate is interpreted as the number of open vacancies relative to the size of the labour market. By normalising vacancies by the number of jobs, this indicator accounts for differences in labour market size across postcodes and allows for meaningful spatial comparison beyond absolute vacancy counts.

In other labour market literature, vacancy indicators are analysed in combination with unemployment, for instance through the vacancy–unemployment ratio or Beveridge curve frameworks, where vacancies are scaled by the number of unemployed workers (Birinci et al., 2025). However, reliable unemployment data at the postcode level are not publicly available in the Netherlands, making such indicators infeasible in this study. Dividing vacancies by the number of jobs provides a practical and conceptually consistent alternative, as employment levels are directly observed at the same spatial scale and similarly reflect the size of the local labour market.

Together, these two indicators capture distinct but complementary dimensions of labour market outcomes. Vacancy duration reflects the friction in job matching, while the vacancy rate indicates the relative demand for labour. Analysing both allows this study to distinguish between areas where vacancies are numerous but filled efficiently, and areas where vacancies persist over time, potentially due to accessibility constraints or mismatches between workplace location and the available workforce. These indicators will be tested on spatial dependence using the Moran statistics and linear dependence using Ordinary Least Squares (OLS).

### 3.3.1. Spatial autocorrelation

Spatial autocorrelation refers to the extent to which the value of a variable observed in one spatial unit is correlated with values of the same variable in neighbouring spatial units. In spatially referenced data, observations are often not independent, as nearby locations tend to share similar socio-economic, infrastructural, or environmental characteristics. Ignoring spatial autocorrelation may therefore violate the independence assumption underlying classical regression models and can lead to biased inference, most notably through underestimated standard errors (Rey et al., 2023). In the context of this study, spatial autocorrelation is relevant for two main reasons. First, key outcome variables such as vacancy duration and vacancy rate may exhibit spatial clustering due to regional labour market structures or shared accessibility conditions. Second, regression residuals may remain spatially correlated if relevant spatial processes are omitted from the model, indicating potential model misspecification. To assess these issues, spatial autocorrelation is evaluated using Moran's I statistics.

**Global and local spatial autocorrelation**
Moran's I measures the degree of spatial autocorrelation by comparing the similarity of values between neighbouring spatial units to the overall variance of the variable. Two related but conceptually distinct forms are used: Global Moran's I and Local Moran's I.

Global Moran's I provides a single summary measure of spatial autocorrelation across the entire study area. It captures whether similar values tend to cluster together globally, but does not indicate where such clustering occurs. The statistic is defined as (Rey et al., 2023):

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}, \tag{3.13}$$

where:

- $N$ is the number of spatial units;
- $x_i$ and $x_j$ are the values of the variable of interest at locations $i$ and $j$;
- $\bar{x}$ is the mean of the variable;
- $w_{ij}$ denotes the spatial weight between locations $i$ and $j$.

Values of Moran's I range approximately from −1 (perfect dispersion) to +1 (perfect clustering), with values close to zero indicating spatial randomness.

Local Moran's I, also referred to as Local Indicators of Spatial Association (LISA), decomposes the global statistic into location-specific measures. It identifies local clusters and spatial outliers by indicating whether each spatial unit is surrounded by similar or dissimilar values. The Local Moran's I statistic for observation $i$ is defined as (Rey et al., 2023):

$$I_i = \frac{(x_i - \bar{x})}{m_2} \sum_j w_{ij}(x_j - \bar{x}), \tag{3.14}$$

where $m_2 = \frac{1}{N} \sum_i (x_i - \bar{x})^2$ represents the variance of the variable. Positive values of $I_i$ indicate spatial clustering of similar values (high–high or low–low), while negative values indicate spatial outliers (high–low or low–high). While Global Moran's I assesses overall spatial dependence, Local Moran's I is primarily exploratory and descriptive, highlighting where clustering occurs rather than whether it exists overall.

**Spatial weights matrix**
Central to the computation of Moran's I is the spatial weights matrix $W$, which formalises the spatial relationships between observations. The weights matrix defines which spatial units are considered neighbours and how strongly they influence each other. Common specifications include contiguity-based weights, distance-based weights, and $k$-nearest neighbour weights (Rey et al., 2023). In this study, spatial weights are constructed based on queen contiguity, where two polygons are considered neighbours if they share a boundary or a vertex. All neighbouring postcodes are assigned equal weights. The weights are row-standardised such that the weights of each observation sum to one. Row standardisation ensures that spatial influence is comparable across units with different numbers of neighbours and facilitates interpretation of the Moran's I statistic.

**Statistical significance and inference**
The statistical significance of Moran's I is assessed using permutation-based inference. Under the null hypothesis of spatial randomness, the observed values of the variable are randomly permuted across spatial units while keeping the spatial weights matrix fixed. The Moran's I statistic is recalculated for each permutation, generating an empirical reference distribution (Rey et al., 2023). The pseudo $p$-value is computed as the proportion of permuted statistics that are equal to or more extreme than the observed

Moran's I value. In this study, significance is assessed using 999 permutations. Moran's I statistics are computed for both the dependent variables and the residuals of the regression models. Significant spatial autocorrelation in the residuals would indicate remaining spatial dependence, suggesting that relevant spatial processes are not fully captured by the explanatory variables.

### 3.3.2. Non-spatial regression

In order to investigate the effect of workplace accessibility on the labour market outcomes, non-spatial regression analysis is conducted, using different sets of explanatory variables. A total of 5 models will be tested as explained in 3.3.3. The general form of the Ordinary Least Squares (OLS) regression model is given by:

$$y_i = \beta_0 + \sum_{k=1}^{K} \beta_k X_{ik} + \varepsilon_i, \tag{3.15}$$

Where:

- $y_i$ is the dependent variable (e.g., average vacancy duration) for observation $i$;
- $\beta_0$ is the intercept (constant term);
- $\beta_1, \dots, \beta_k$ are the coefficients (slopes) of the independent variables;
- $x_{i1}, \dots, x_{ik}$ are the independent variables (predictors) for observation $i$;
- $\epsilon_i$ is the error term (residual) for observation $i$.

The OLS estimator is obtained by minimising the sum of squared residuals (SSR) and relies on a number of assumptions, including linearity, independence of observations, homoscedasticity and the absence of perfect multicollinearity. Under these assumptions, the OLS estimator is unbiased and efficient. In the context of this study, particular attention is paid to the assumptions of independence and multicollinearity. Since the analysis is conducted at postcode level, spatial dependence between neighbouring observations may violate the independence assumption. This is explicitly assessed using Global Moran's I statistics on both the dependent variables and model residuals (as mentioned in 3.3.1). Multicollinearity can occur when explanatory variables are highly correlated, potentially inflating standard errors and obscuring individual coefficient significance. This is particularly relevant in models that include interaction terms, as the interaction is mechanically correlated with the variables it consists of. To mitigate multicollinearity, accessibility and density variables are mean centred prior to constructing interaction terms. Variance Inflation Factors (VIFs) are used as a diagnostic tool to assess the severity of multicollinearity. VIF values above 5-10 are generally considered to be problematic. While multicollinearity does not bias coefficient estimates, it affects their precision and should therefore be considered when interpreting statistical significance (Kim, 2019).

In this study, the regression models are estimated at the postcode level. The coefficients are interpreted as average partial effects, indicating the expected change in the dependent variable associated with a one-unit change in the explanatory variable, holding other factors constant. To capture heterogeneity in the relationship between accessibility and vacancy outcomes across spatial contexts, interaction terms are included in the regression models. An interaction between accessibility and urban density allows the effect of accessibility to vary depending on the address density. The interaction model can be written as:

$$y_i = \beta_0 + \beta_1 A_i + \beta_2 D_i + \beta_3 (A_i \times D_i) + \varepsilon_i, \tag{3.16}$$

where:

- $y_i$ is the dependent variable for postcode $i$;
- $A_i$ denotes workplace accessibility;
- $D_i$ represents urban density;
- $\beta_3$ captures how the effect of accessibility varies with density;
- $\varepsilon_i$ is the error term.

In the presence of an interaction term, the coefficient $\beta_1$ no longer represents the unconditional effect of accessibility. Instead, the marginal effect of accessibility is given by:

$$\frac{\partial y_i}{\partial A_i} = \beta_1 + \beta_3 D_i. \tag{3.17}$$

This formulation implies that the impact of accessibility depends on the level of urban density. As a result, coefficient significance should be interpreted jointly, and marginal effects are evaluated across the observed range of density values rather than at a single point (Brambor et al., 2006).

### 3.3.3. Regression modelling strategy

To quantify the relationship between workplace accessibility and vacancy outcomes, a series of regression models is estimated at the postcode level. Two dependent variables are analysed: vacancy duration, measuring the median time required to fill vacancies, and vacancy rate, defined as the number of vacancies relative to the number of jobs in a given year. These indicators capture different dimensions of the labour market. Vacancy duration reflects matching efficiency over time, while vacancy rate captures structural labour market pressure(Andrews et al., 2008; Rouwendal & Rouwendal, 2025). The modelling strategy follows a stepwise specification approach, in which complexity is gradually introduced to assess the robustness of the found relationships.

**Model 1: Baseline accessibility models**

The first model compares public transport workplace accessibility from Hansen (1959) or Shen (1998) as the sole explanatory variable. This ensures that the comparison can be made between the two accessibility metrics and explores whether competition effects are necessary to explain labour market dynamics. The overall specification tests the core hypothesis that accessibility of the workplace is associated with vacancy indicators. Similar baseline relationships between location accessibility and recruitment outcomes have been suggested in vacancy duration studies, where firms located in more central or accessible areas tend to fill vacancies faster (Dossche et al., 2025; Mangan & Trendle, 2017). This model establishes whether a direct association exists before accounting for spatial or socio-economic context.

**Model 2: Accessibility + Urban context model**

The second model adds controls for urban context, by including address density. This specification acknowledges that accessibility is strongly correlated with urban form, and that labour markets in dense urban areas differ structurally from those in peripheral regions. Urban labour markets tend to be more lively, have more intense matching processes, and higher vacancy rates, not necessarily due to labour shortages but due to increased activity and competition (Rouwendal & Rouwendal, 2025). Therefore, the vacancy indicators might not be dependent on the accessibility, but on the location and its urban context. Including urban density allows the model to distinguish whether observed vacancy outcomes are attributable to accessibility itself or to broader density effects.

**Model 3: Interaction model: accessibility × density**

The third model addresses that the effect of accessibility does not have to be the same everywhere. In peripheral and low-density areas, limited transport connectivity constrains the effective labour supply, making accessibility a binding factor for vacancy fulfilment. Improvements in accessibility therefore expand the reachable workforce and reduce matching frictions. On the contrary, in highly urbanised areas with dense transport networks and large labour pools, accessibility is less likely to be the primary constraint. Instead, high accessibility may intensify competition among employers for similar workers and increase selectivity in hiring processes, potentially increasing vacancy durations. In these contexts, accessibility no longer relaxes a constraint but interacts with agglomeration effects and labour market competition. The interaction term allows to test if the relevance of accessibility is different for urban or peripheral areas (such as the port areas and other industrial areas). This is in line with literature on spatial mismatch (Kain, 1968) and the call for job accessibility research for different spatial contexts (Geurs & van Wee, 2004).

**Model 4: Accessibility + education**
The fourth model introduces the required education of a vacancy. The vacancy duration is not only determined by accessibility, but also by the qualifications of the available workforce. Vacancies with higher qualification requirements have a smaller labour pool to pick from. High educated people have more options for work and generally more options in dense city areas (Bastiaanssen et al., 2025). Areas with high shares of high educated people tend to have better matching, especially in knowledge-based sectors. This model tests whether the vacancies are explained by spatial mismatch or qualification mismatch

**Model 5: Accessibility + place bound work**
The final model combines accessibility with the need to travel. As van Lent et al. (2025) demonstrated, teleworkability improves the accessibility of an individual. People who can work from home accept longer travel times because they do not have to travel every day. This model estimates if the number of place bound jobs can explain the vacancy indicators.

While the models control for key aspects of labour supply composition, including educational requirements and place-boundedness, they do not aim to fully explain vacancy outcomes. Labour market matching is influenced by a range of additional factors that cannot be observed at the postcode level with sufficient accuracy or with the available data. In particular, variables such as wage levels, working conditions (e.g. shift schedules, physical workload, contract stability), and firm-specific recruitment strategies are known to play an important role in vacancy fulfilment but cannot be included due to data limitations. Although controlling for education level captures part of skill availability, the degree of job-specific specialisation and experience requirements remains unobserved. How these model are created and operationalised is further explained in chapter 4

# 4
# Research data

Following the methodological framework established in chapter 3, this chapter details the datasets and the specific preprocessing steps required to operationalize the variables. The data integration process follows a systematic flow from raw acquisition to the construction of analytical metrics, as visualized in the data flow diagram in figure 4.1. This chapter follows the order of the flowchart from left to right. First, the data is introduced in section 4.1, the accessibility parameters are then defined in section 4.2, followed by the operationalisation of the data in section 4.3.

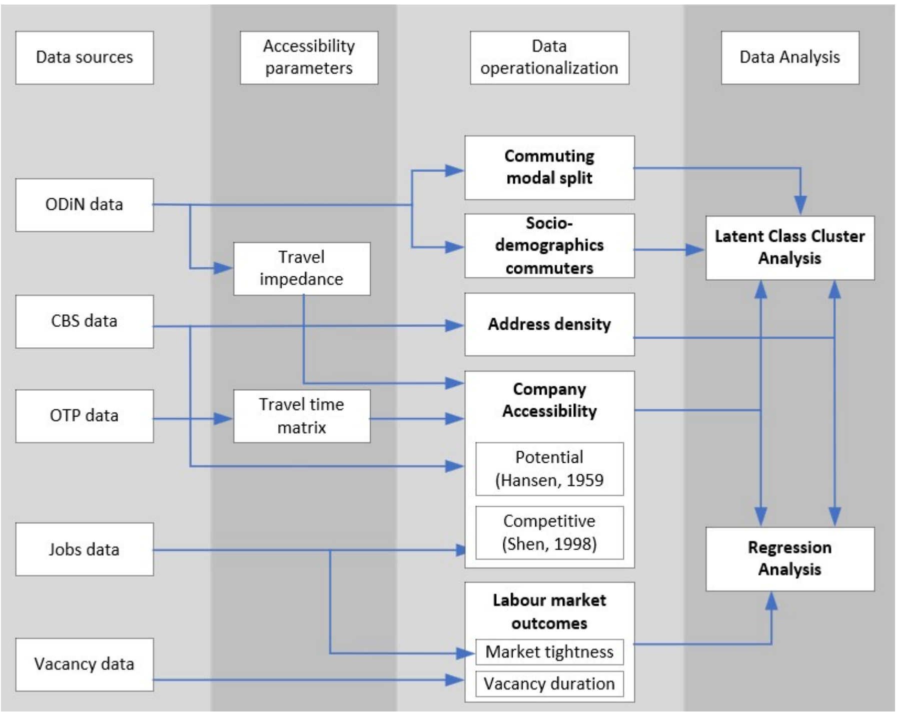



**Figure 4.1:** Flowchart for the data used in this research

## 4.1. Data sources

This research requires multiple types of data, which need to be combined into one dataset. Firstly travel data is required from a large number of people on a personal level. This data is necessary to compute the commuter profiles per postcode, the modal share and to determine the $\beta$ in the impedance function. Additionally, information on postcode 4 areas is also required. This can be found in the data by the CBS (Centraal Bureau voor de Statistiek (CBS), 2025b). This data contains characteristics such as population, urban characteristics and geographical data from each area. Besides this, travel times are necessary in order to calculate the accessibility on postcode 4 level. These public transport travel times are generated for all possible postcode 4 combinations. Lastly, vacancy data is available for usage from Jobdigger (Jobdigger, 2025). The data is further introduced individually in 4.1.1, 4.1.2, 4.1.3, 4.1.4 and 4.1.5.

### 4.1.1. ODiN (individual trip level data)

To analyse commuting behaviour and travel patterns in the Netherlands, this study makes use of the "Onderweg in Nederland" (ODiN) dataset, published annually by Statistics Netherlands(Centraal Bureau voor de Statistiek (CBS), 2023b). The ODiN survey provides nationally representative information on daily travel behaviour of residents aged older than 6 in the Netherlands, covering all modes of transport and trip purposes. Respondents record their trips for a single reference day, including trip origin and destination, purpose, mode, duration, distance, and departure and arrival times. The dataset also includes socio-demographic information such as age, gender, household composition, education level, income, employment status, and car ownership.

The ODiN data is cleaned by the CBS to remove the clearly wrong entries. This includes people not filling in the survey seriously, misinterpreting questions or missing values. The data is enriched with register information (e.g. postcodes and route distances), harmonising trips (merging transit segments, filtering out flights and heavy freight truck movements or purely foreign trips), and deriving new variables such as trip motives and modal classifications (Centraal Bureau voor de Statistiek (CBS), 2024c). In 2023, the initial field responses numbered 64,908, of which 64,459 were retained as “usable” after cleaning procedures (Centraal Bureau voor de Statistiek (CBS), 2024c). These procedures ensure that the ODiN dataset is as accurate and representative as possible, though it remains a cross-sectional snapshot rather than a panel of individual travel behaviour. Similar processing is done in the other years of the survey. The number of respondents of all years is shown in table 4.1 resulting in a total of 367.075 respondents.

**Table 4.1:** Respondents from the included years

| Year | Respondents |
|---|---|
| 2018 | 57,260 |
| 2019 | 53,380 |
| 2020 | 62,940 |
| 2021 | 67,083 |
| 2022 | 61,953 |
| 2023 | 64,459 |

### 4.1.2. CBS data

For additional information on each postcode, CBS data is used. CBS provides extensive population statistics and spatial characteristics of every postcode in the Netherlands (Centraal Bureau voor de Statistiek (CBS), 2025b). This data also includes the number of residents, age groups, distance to facilities and many more elements. The data is used for the year 2023 and includes 4070 postcodes. In 2018, 4061 postcodes existed so there have only been a few new ones over the years, but this might mean that some postcodes are slightly underrepresented in the ODiN data. Additionally, the geometries are provided in ESRI shape formats and include the exact area and contours of the postcode areas. This is used to generate plots to show spatial heterogeneity. These shapes are used in combination with the statistical data of squares from CBS (Centraal Bureau voor de Statistiek (CBS), 2025b), which contain similar data on 100 m x 100 m squares of the Netherlands, to identify appropriate coordinates of each postcode (more on this in 4.2.2). The data has 137 columns with information on the residents, socio-demographics, distances to facilities and also the geometry of the area. Most columns are irrelevant for this research. The used variables are shown in table 4.2.

**Table 4.2:** Selected attributes from CBS data

| Attribute | Description |
|---|---|
| ResN | Number of residents (accessibility calculation) |
| Train | Distance to closest train station (proxy for accessibility) |
| AddrDens | Density of addresses per $km^2$ |

### 4.1.3. Travel time data

The travel time data is generated using OpenTripPlanner (OTP) (OpenTripPlanner Contributors, 2025), which is an open-source family of software projects that provide passenger information and transportation network analysis services. The core server-side Java component finds itineraries combining transit, pedestrian, bicycle, and car segments through networks built from widely available, open standard OpenStreetMap and GTFS data(OpenTripPlanner Contributors, 2025). Given coordinates, time and transport mode, it calculates the best route. Because OTP computes travel times on demand, rather than relying on static matrices, it provides flexible and realistic routing outputs for any origin, destination, date, or departure time. OTP constructs a network graph using OpenStreetMap (OSM) for walkable links, combined with GTFS files for transit schedules, stops and transfers. When a routing request is sent, OTP identifies the nearest network nodes to the origin and destination and computes all feasible walking and transit connections, constrained by timetable. It returns the optimal itinerary based on a specified objective such as the fastest trip. This data is used to calculate the accessibility of each postcode. The process of generating these travel times is explained in 4.2.2.

### 4.1.4. Jobs data

The LISA employment register is a nationwide database that contains all active business and institutional establishments in the Netherlands where paid work is carried out. It is maintained by BIJ12 on behalf of the Dutch provinces and is updated annually through a combination of administrative data from the "Handelsregister" and a large-scale survey among firms. Because LISA is a complete register rather than a sample, it provides highly detailed and reliable information on employment at any spatial or sectoral scale, making it one of the primary data sources for regional labour market analysis(BIJ12, 2024). The dataset contains all the jobs at each postcode in the Netherlands for the years 2023 and 2024. This data can be used to calculate Shen's competitive accessibility. Additionally, it is used as an inactive covariate in the LCCA analysis. Lastly, it is used to arrive at a relative measure for the number of vacancies, as more opportunities generally also lead to more vacancies.

### 4.1.5. Vacancy data

Jobdigger has a comprehensive database of all the vacancies that appear online. Their tool scrapes websites to look for vacancies on every publicly available website. These vacancies are then further processed to find additional information, such as prerequisite skills, education, company, location, etc. For the purpose of this research, all vacancies from 2022 onwards were selected for the Rijnmond region. This region includes Rotterdam and the port area, as well as some other smaller towns and cities (see figure B.1 for a map). The vacancies that were selected all had coordinates of the location where the work would be performed. If only the head office location was provided, the vacancy was excluded. Additionally, intermediaries were excluded since most people do not work at the location of the intermediary itself. The intermediaries would give a false impression of the actual need for work at each location since they are all grouped on one location while they should be spread out. Vacancies are frequently cross-posted across multiple websites. However, Jobdigger identifies and removes these duplicates. Their tool works full time and takes about one week to scrape every website for vacancies. This means that the dates of the vacancies have an uncertainty of 0-7 days. Further preparation of the data is discussed in section 4.4.

## 4.2. Deriving the Accessibility parameters

In this section, the process of deriving the accessibility parameters is discussed. As shown in figure 4.1, the travel impedance is derived from the ODiN data, which contains actual travelled distances and travel times from commuters. The travel time matrix is computed using the OTP graph of the Dutch public transport network. This section summarises the choices made in order to arrive at the necessary variables to compute the company accessibility.

### 4.2.1. Estimating the impedance function and parameters

As mentioned in section 3.1, the accessibility metric relies on a generalized measure of proximity, defined as the potential interaction between an origin and a destination, weighted by the spatial impedance of the travel required to connect them. The functional form of this function can be empirically determined. As this function should resemble the actual behaviour of commuters, the ODiN travel times can be utilised. To this end, the ODiN data has been pre-processed to include only commuting trips to work (as will be explained in more detail in 4.3.1) and filtered for public transport trips.

The distribution of the reported travel times is shown in figure 4.2. This in line with reports from Hamersma and Roeleven (2024), who compute a very similar curve of what commuting times people find acceptable. The ODiN survey data is susceptible to the presence of extreme outlier observations, particularly within the recorded travel time and distance variables. These outliers could be data entry errors by respondents (e.g., mistyping 70 minutes as 700 minutes or misclassification of trip purpose) and in rare cases, they might not even be outliers. Left untreated, a few extreme values can exert a disproportionately large influence on non-linear model estimations, such as the decay functions used in this study, severely compromising the statistical robustness and interpretation of the results. Given the objective of this study, to model the typical commuting behaviour and estimate the general impedance of travel time, it was decided that the benefit of robust model parameter estimation outweighs the risk of excluding a small number of genuinely rare trips. To systematically identify and remove these extreme observations, the Interquartile Range (IQR) method was applied, which is shown in equations (4.1), (4.2) and (4.3).

$$\text{IQR} = Q_3 - Q_1 \tag{4.1}$$

$$\text{Lower bound} = Q_1 - (1.5 \times \text{IQR}) \tag{4.2}$$

$$\text{Upper bound} = Q_3 + (1.5 \times \text{IQR}) \tag{4.3}$$

Applying this equation to the 7528 travel times removes 200 values (2.6%) of the data. As shown in figure 4.2, the maximum travel time is reduced to 122 minutes which is still high but realistic for commuting.

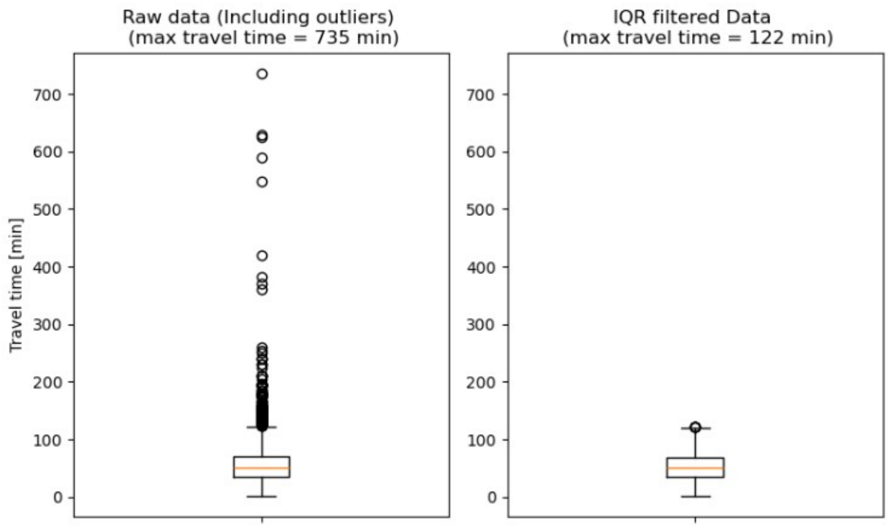



**Figure 4.2:** Boxplots of the public transport trips with and without statistical outliers

The decay function, $f(C)$, models the probability that a trip of cost $C$ (in this case, travel time in minutes) is undertaken. This function is essentially the empirical survival curve $S(C)$, which shows the fraction of trips completed at or beyond a specific travel time. This function is visualised in figure 4.3. The plot can be interpreted as follows: around 80% of the people accepts a travel time longer than 30 minutes, but only 30% accepts a travel longer than 60 minutes. The function decreases mostly stepwise due to people rounding their travel times to 5 minutes. To select the mathematically robust and theoretically sound functional form, seven widely recognized decay functions(Verma & Ukkusuri, 2025) were fitted to the empirical survival curve of the public transport commuting travel times, where $C$ is the travel time in minutes and $\beta_i$ are the estimated parameters:

**Table 4.3:** Decay Functions and Model Fit Statistics

| Function Name | Functional Form | RSS | $R^2$ | $\beta_1$ | $\beta_2$ |
|---|---|---|---|---|---|
| Power | $f(C) = C^{\beta_1}$ | 440.453 | 0.279 | $-1.851 \cdot 10^{-1}$ | |
| Exponential | $f(C) = \exp(\beta_1 C)$ | 141.856 | 0.768 | $-1.454 \cdot 10^{-2}$ | |
| Sheratt–Tanner | $f(C) = \exp(\beta_1 C^2)$ | 10.119 | 0.983 | $-2.700 \cdot 10^{-4}$ | |
| Logistic | $f(C) = \frac{2}{1+\exp(\beta_1(C-\beta_2))}$ | 17.514 | 0.971 | $3.692 \cdot 10^{-2}$ | 19.604 |
| Power–Exponential | $f(C) = \exp(\beta_1 C^{\beta_2})$ | 2.675 | 0.996 | $-5.207 \cdot 10^{-5}$ | 2.412 |
| Tanner | $f(C) = C^{\beta_1} \exp(\beta_2 C)$ | 141.856 | 0.768 | $-5.729 \cdot 10^{-18}$ | -0.015 |
| Quadratic–Exponential | $f(C) = \exp(\beta_1 C + \beta_2 C^2)$ | 10.119 | 0.983 | $-2.050 \cdot 10^{-17}$ | $-2.7 \cdot 10^{-4}$ |

All models were fitted using non-linear least squares regression , minimizing the difference between the observed empirical survival probability, $S_{obs}$, and the predicted probability, $S_{pred}$. The goodness-of-fit was evaluated using two primary criteria(Verma & Ukkusuri, 2025):

- Residual Sum of Squares (RSS): The sum of the squared differences between the observed and predicted values. A lower RSS indicates a better fit of the function to the empirical data.

$$\text{RSS} = \sum_{i=1}^{n} (S_{obs,i} - S_{pred,i})^2 \tag{4.4}$$

- Coefficient of Determination ($R^2$): Represents the proportion of the variance in the dependent variable (the survival probability) that is predictable from the independent variables (travel time). A value closer to 1.00 indicates a superior fit.

$$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}} \tag{4.5}$$

Based on these criteria, the power exponential function provided the best fit with an $R^2$ of 0.996 and an RSS of 2.675. Therefore, the impedance function $f(C)$ used to calculate accessibility for the public transport mode is defined by the following functional form:

$$f(C) = \exp(-5.207 \cdot 10^{-5} \cdot C^{2.412}) \tag{4.6}$$

This function can now be applied to the time travel matrix, to calculate the accessibility of each postcode.

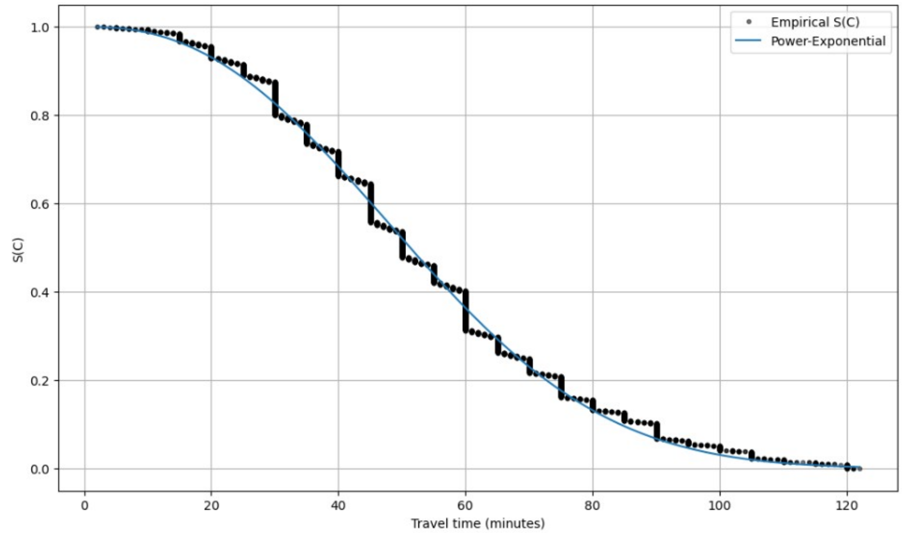



**Figure 4.3:** Survival curve S with the fitted power exponential function

## 4.2.2. Generating the public transport travel time matrix

Now that the impedance function is defined, the travel times need to be computed. In order to calculate the travel times, coordinates need to be defined first. Defining a single representative coordinate for a postcode is challenging, as these areas can be expansive and include uninhabited zones such as water or industrial land. For the calculation of travel time data, accurate coordinates are required as these are direct input and routes cannot be calculated otherwise. Consequently, using the exact geometric centroid often results in an illogical location. Therefore, each postcode was matched with the squares that are 100% inside the contours of the postcode. From these squares, the one with the highest population was chosen and the middle of this square was identified as the coordinates. This is visualised in figure 4.4. For the postcodes in the port, the coordinates were manually chosen, as these areas are even more complex due to the highly industrial nature of the area. A handful of areas still had to be manually changed due to the absence of squares with information. For these areas, the coordinates were picked using google maps at logical locations.

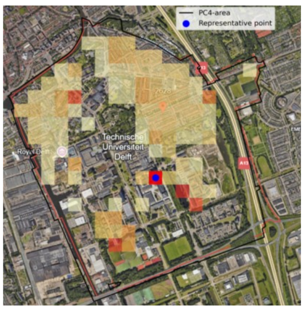



**Figure 4.4:** Example of selecting the representative point for postcode 2628 (Delft)

The routes are calculated using public transport and first and last-mile transport is set to walking. Cycling might have been more logical for first mile transport, but this is not the focus of the analysis. Walking is the minimum viable access mode, which is always available. A maximum walking distance of 5000 meters is used so routes can be computed in most cases, though it should be noted that these last mile distances are probably too far for walking (for first and last mile cycling, this could be a possible distance, but this is not investigated). The departure time is set to 8:00 AM for a realistic commuting time on a representative weekday in September in 2025. Ideally, this would have been computed for 2023 to align with the other data but this produced less reliable results and was harder to verify. The choice was made to run the simulations for one commuting time due to limited computational budget which means that there is some chance involved if the transit schedule aligns well with the departure time. However, Boisjoly and El-Geneidy (2016) concluded from their research that a constant measure at 8:00 AM is representative of the relative accessibility and highly correlated with more sophisticated dynamic measures(Boisjoly & El-Geneidy, 2016).

For collecting the OD travel times, it is important to note that 4069 x 4069 gives over 16 million travel times to compute. Given that this takes a very long time to compute, the process is split up into two parts. In the first part, only pairs (i,j) were calculated which had an Euclidean distance smaller than 150 km. These were 6,169,076 pairs. When this was finished, the pairs (j,i) were calculated which had an euclidean distance smaller than 150 km and of which (i,j) was smaller than 180 minutes. These values were chosen from the boxplots of the commuting travel times and distances in ODiN as they are quite far above what is realistic for commuter behaviour. This can be seen in the boxplots in figure 4.2. This leaves room if the reverse travel time happens to be shorter. The reverse distance left 3,6 million pairs. For the port postcodes, the calculations and coordinates were carefully checked in order to properly implement the Maasvlaktehopper. The Maasvlaktehopper is implemented in OTP as having a transit schedule and making a one way loop, as opposed to being an on demand shuttle service. Since it is supposed to shuttle you directly from the stop closest to you to one of three stops outside the port, the loop would always penalise either arriving or leaving too heavily. Therefore, it was decided that mirroring (i->j) and (j->i) was the most accurate way of solving this. The results of this search tactic is shown in figure C.1.

The travel times are difficult to verify case per case, due to the size of the matrix. Therefore, some random travel times have been plotted for visual inspection and an additional check has been done for the postcodes with the least computed travel times. These analyses can be found in appendix C. It should be noted that these travel times appear to be accurate on average, but some cases are not completely correct. For example, the boat from Friesland to Ameland is not included in this OTP, meaning that Ameland postcodes are only reachable from the island itself. Also, some towns in the bottom of Zeeland had very poor connections even though a bus stop was located there. In these places, the issue was with the transit schedule. Other locations included Alphen (on-demand travel), Veenhuizen (hourly bus with poor connection timing), and America (no public transport). However, all these postcodes had relatively little impact as the population was below average, and travel times would always be low. Additionally, these postcodes are not within the Rijnmond area, or within the postcodes relevant in ODiN. Therefore, further efforts to improve these minor flaws in the data is neglected.

With the impedance function and the travel time matrix defined, the accessibility measures can now be computed as defined in section 3.1. The results are presented in 5.1.1.

## 4.3. Data operationalisation

The accessibility measures are ready to be used for the analysis. However, the ODiN data and the vacancy data still need to be pre-processed as these datasets include individual level data. This section covers the preprocessing of these datasets, and the aggregation to postcode 4 aggregates.

### 4.3.1. Preprocessing ODiN

The collected ODiN data is already filtered and processed by CBS. However, for the specific task of this project, more filtering has to be done. The further preprocessing of the data is discussed in this section.

#### Filtering commuting trips from home to work

The dataset consists of daily trips at random days, thus weekdays often contain the commuting trips of respondents. For the purpose of this research, the non-commuting trips need to be filtered out. There are several indicators which can be used for this filtering. The ones used for this are **Goal**, **CMotive** and **DepLoc**.

The variable **Goal** can take values from 1-14 depending on the goal and/or destination of the movement. The interesting ones are 1 (home), 2 (to work), 3 (Business visit in work environment), 4 (Occupational). Since commuting is defined as "the activity of travelling regularly between work and home, or travelling to different places for work", each of these outcomes are satisfied. For accessibility, the perspective of the workplace is used, so the same should apply for the movements. Therefore, the trips home (1) are excluded. The goal is to find the modal split for commuting which means it would be misleading if all movements on the same day of one person that is related to work would be used. Generally, people who have business visits throughout the day tend to use the car which would result in multiple car entries while the choice for car was likely made based on the information that it would be used multiple times. Similarly, occupational movements are often done using occupational specific vans or cars. Therefore, there is no choice involved and these results are also misleading.

The variable **CMotive** is mainly included as a check, since it is a less specific version of the variable **Goal**. All the motives are filtered out apart from motive 1 (from and to work). Should people have incorrectly filled in one of the variables, this can indicate invalid and inaccurate entries and these people are removed.

Lastly, **DepLoc** is used which has 4 possible values: Own home address (1), other home address(2), work address(3) and other address(4). Clearly, work address has to be filtered out as the person leaving work is not included and business trips were excluded too. (1) and (2) need to stay in as they are a home location. Other addresses can be plenty of things. Looking at the data and checking the movement prior to the movement to work, the pie chart in figure 4.5 can be created. These results show that the majority first picks up or drops off people. Another high value is walking and touring. This could have multiple reasons. People have to walk to the location of their car or public transport and they interpreted the trips and movements wrong, so they did not classify walking as part of their movement to work. Other options include that they actually went for a walk or run before work which is possible given that not everybody starts work early. Important to note that all the work related goals are not represented in this category so it is safe to assume that "other address"(4) can be used for the analysis.

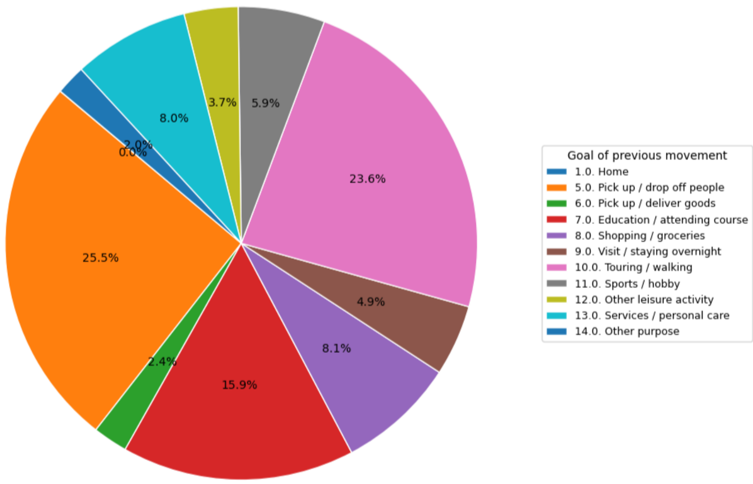



**Figure 4.5:** Distribution of the goals of the previous movement

**Filtering duplicate trips**
After these changes, 82.645 unique respondents remain. However, there are 101.048 rows left. This means that some people have multiple trips and movements from home to work. Movements can consist of multiple trips. For example, a person can cycle from home to the train, take the train, and walk to work. This is considered as 1 movement with 1 goal, cut into 3 trips with 3 different modes. The trip information is not necessary (the variable **Hvm** captures the mode that was used over the longest distance in the movement) and these rows can be removed. However, this still leaves 85.703 movements.

First, the duplicate movements with equal **Vertloc**, **depPC**, **Hvm** and **ArrPC** are selected. These movements are either a wrong entry for going home or someone commuted to work twice with the same vehicle. In either case, the first movement contains the necessary information, and the second movement is double so only the movement with the earliest starting time is kept. This removes 1205 rows. Further inspections of the double movement shows that the first movement is almost always at the time of the morning peak or around lunch, whereas the second movement is around lunch or late in the afternoon (figure 4.6). Sometimes these entries differ in **depPC**, sometimes in **ArrPC**, **Hvm** or a combination. A second check was done by looking into potential sequencing trips by computing the time in between two movements. However, the shortest interval between two movements was 25 minutes, meaning it cannot be concluded that these movements are part of one movement. This appears to be logical from the histograms.

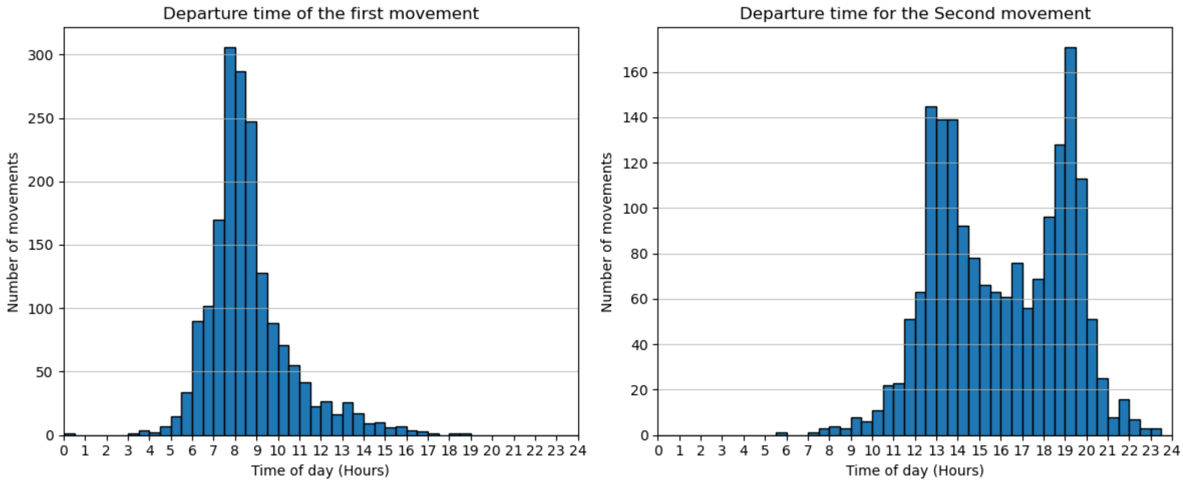



**Figure 4.6:** Departure times of the duplicate movements

The histograms (figure 4.6) show the difference in departure times of the two movements. Departure times in the night (00:00 - 05:00, 15 persons), were manually checked, as night movements could potentially be someone starting with work or ending work. Some cases stood out as people would commute again around the morning peak from a different address, where the night commute would appear to be to a sleeping address. Therefore, the commute around the morning peak was taken in these cases. In the other cases, the assumption that the first commute was the leading one was kept.

One could argue that when a person travels twice to a different location for work, it can still be valid for the modal split, but since most of the data is only one movement per person, and the first movement is most likely to be the movement sought after, it is decided that only the first movement is used. This also takes into account that every persons attitude toward commuting is considered equally and nobody is overrepresented.

### 4.3.2. Defining modal split

In order to calculate the modal split, the modal split first needs to be defined. ODiN has the variable **Hvm** which gives the main transport mode of the movement. This variable will be used to calculate the modal split. The variable is a list of 24 possible transport modes and these all need to be assigned to "car", "public transport", "slow modes" or in some cases "other". The list of transport modes, the category and reasoning can all be found in table E.3. The modal split is then calculated for each postcode by dividing the number of people who used a certain mode by the total of that area:

$$S_{m,i} = \frac{n_{m,i}}{N_i} \quad \text{with} \quad N_i = \sum_m n_{m,i}, \tag{4.7}$$

where:

$S_{m,i}$ = share of mode $m$ in area $i$,
$n_{m,i}$ = number of movements using mode $m$ in area $i$,
$N_i$ = total number of movements in area $i$.

### 4.3.3. Aggregation to postcode 4 level and recoding variables

In order to identify clusters of commuter profiles based on postcode, the data needs to be aggregated to postcode 4 level. This is done using the **ArrPC** variable which is the postcode of arrival of the trip. Most variables are coded as categorical variables, and too detailed for the analysis. Therefore, the variables are recoded to continuous fractions of the population.

**Age**
This variable is a count variable as it can take integer values starting from 6. One option was to take the average age of the postcode, but this method deletes quite some information. An average age of 40 can mean that everyone is around 40 but also that half the population is 20 and half the population is 60. Therefore, new age categories were created and the percentages of people within these categories were determined. For the categories, <25, 25-34, 35-49 and 50-65 were chosen. This was done since these these age categories roughly represent life stages of study, early career, middle career, late career and retirement. This does of course not apply for everyone, but does allow for better interpretation.

**Gender**
The original variable is of binary categorical nature: male or female. This is converted to a percentage of males.

**Education**
The original variable is categorical, with seven categories. This is converted to a percentage of low educated and high educated individuals using the same categories as the CBS applies. The percentage of medium-educated people is intentionally left out to avoid multicollinearity.

**Income**
The original variable is ordinal categorical with ten categories, representing the ten deciles of income. The first 3 deciles are taken together as the low income group and the last 3 deciles are taken together as the high income group. These are then converted to percentages of the population.

**Car ownership**
This original variable is also categorical and indicates how many cars or other vehicles a household has. Here, the definition of car mode (Appendix E.3) is used as it used to calculate the modal split (4.3.2). Again, this value is converted to percentage of the population.

**Social participation**
The original variable is of categorical nature, with eight categories. This is converted to a percentages of people that either work 12-30 hours per week or are a student. The rest category is roughly the people working full-time.

**Household configuration**
The original variable is of categorical nature, with eight categories. This is converted to a percentage of persons with children in their household, and a percentage of single households. The rest category is dual households.

Table 4.4 shows the variables as they are eventually included in the Latent Class Cluster Model. One category per variable is intentionally excluded because the proportional categories sum to one. Including all categories would lead to perfect multicollinearity, as one category can always be inferred from the others. Excluding a reference category avoids this problem while preserving the underlying variation. However, the remaining categories are still inherently correlated: an increase in one share mechanically reduces another. Thus, even with one category omitted, the model continues to reflect the full compositional structure of the data. Since the categories are all percentages of the population, all the indicators are continuous and range from 0 to 1. The covariates are all continuous but they have been scaled in order to match the other variables using min-max normalisation. This is done for better numerical stability and interpretability of the results.

**Table 4.4:** Overview of variables, measurement scales and categories

| Variable | Measurement scale | Categories / Description |
|---|---|---|
| *Indicators* | | |
| Age | | |
| • Age <25 | Continuous (0–1) | Share of commuters younger than 25 |
| • Age 25–35 | Continuous (0–1) | Share of commuters aged 25–35 |
| • Age 35–50 | Continuous (0–1) | Share of commuters aged 35–50 |
| Income | | |
| • Inc_low | Continuous (0–1) | Share of low-income commuters |
| • Inc_high | Continuous (0–1) | Share of high-income commuters |
| Education | | |
| • Edu_low | Continuous (0–1) | Share of low-educated commuters |
| • Edu_high | Continuous (0–1) | Share of high-educated commuters |
| Social participation | | |
| • Students | Continuous (0–1) | Share of student commuters |
| • parttimers | Continuous (0–1) | Share of part-timer commuters |
| Car ownership | Continuous (0–1) | Share of households owning a car |
| Household configuration | | |
| • Single | Continuous (0–1) | Share of single household commuters |
| • Hh_children | Continuous (0–1) | Share of households with children commuters |
| *Active covariate* | | |
| Accessibility | Continuous (0–1) | Hansen-type potential accessibility measure |
| *Inactive covariates* | | |
| Urbanity | Continuous (0–1) | Address density (addresses per km$^2$) |
| Jobs | Continuous (0–1) | number of jobs |
| PT share | Continuous (0–1) | Share of trips by public transport |
| Car share | Continuous (0–1) | Share of trips by car |

Now that the variables have been defined, it is important to check whether these variables are a good representation of the Dutch population, and also whether each postcode has significant number of respondents to get reliable estimates.

### 4.3.4. Representativeness of ODiN data

As mentioned in the previous sections, the ODiN is now filtered on the movements of interest and the variables have been recoded. The analysis is conducted on postcode 4 level, which means that the data needs to have sufficient movements for each postcode in order to perform a valid statistical analysis, For this, a threshold of 30 respondents is the bare minimum as the Central Limit Theorem can be applied. The Central Limit Theorem states that when a large sample size is large enough (30+), the standardized sum (or mean) of independent and identically distributed random variables tends toward a normal distribution, regardless of the original distribution(Feller, 1968). This becomes more valid, the higher the number of respondents. This is a large motivator for using all of the 6 years from the ODiN dataset, even though some differences in travel behaviour were observed during the COVID years. The full motivation for keeping all the years 2018-2023 can be found in appendix E. Even though all years are used, only 829 of the 4070 postcodes had a significant number of movements. This has quite some implications on the representativeness of the data. One can imagine that keeping only the postcodes where many work trips have been made might introduce bias into the data. Table 4.5 shows the representation of urbanity levels in the remaining postcodes, compared to the total dataset. As expected, the postcodes that did not have enough respondents are mostly not urbanised postcodes. One can argue that those postcodes have less work and are thus less relevant, but it does introduce bias in the data.

**Table 4.5:** Representation of urbanity in relevant postcodes in ODiN

| Class | Description | Population (All PC4) | Population % | ODiN (N>30) | ODiN % |
|---|---|---|---|---|---|
| 1 | Extremely urbanised | 443 | 10.9% | 219 | 27.5% |
| 2 | Strongly urbanised | 629 | 15.5% | 220 | 27.6% |
| 3 | Moderately urbanised | 459 | 11.3% | 157 | 19.7% |
| 4 | Hardly urbanised | 542 | 13.3% | 155 | 19.5% |
| 5 | Not urbanised | 1996 | 49.0% | 46 | 5.8% |
| Total | | 4069 | 100% | 797 | 100% |

A plot of the locations of the statistically valid postcodes (figure 4.7) shows that the areas are distributed quite equally over the country, with more valid postcodes in the Randstad, but this is expected, given more people live and work in these areas.


**Figure 4.7:** Spatial distribution of the statistically valid postcodes

The same can be done for the socio-demographics and the Dutch working population. These values are shown in table 4.6. There are a couple of notable differences between the working population of ODiN and the Dutch working population. The lower income and lower education groups are not fairly represented in the data that is left. A reason for this could be that a selection is made of people who have travelled from home to work. This eliminates the working population that does not have a fixed working location (e.g. plumbers, electricians, delivery people) who generally fit the profile that is missing. Additionally, the urban areas attract more educated and wealthy people as the large firms, and universities are located here. Another factor that might influence this is that lower educated people are harder to convince to fill in these kinds of surveys. However, ODiN is collected as such that these differences are partly accounted for, using weights to counteract non-response bias(Centraal Bureau voor de Statistiek (CBS), 2024c). Every socio-demographic group is represented so the population is similar to the actual Dutch population. However, since the trips are recorded for a random day, there is randomness included in which people of the ODiN have work trips, introducing bias in the working population. It is not possible to account for this, and therefore important to keep in mind in the analysis. Although there are some differences, the data is fairly representative for working population. The car ownership percentages for the working population could not be found so they are presented for the total population. The percentage of car ownership is lower in the ODiN data than in the dutch working population. The most likely explanation for this is the over-representation of extremely urbanised postcodes. Car dependency is generally lower for work addresses in extremely urbanised areas.

**Table 4.6:** Representativeness of ODiN data compared to the Dutch working population(Centraal Bureau voor de Statistiek (CBS), 2025a) (note: car ownership is percentage of households of total population)

| | **ODiN data** | | **Dutch working population** | |
|---|---|---|---|---|
| **Category** | **Count** | **Share (%)** | **Count (x1000)** | **Share (%)** |
| **Gender**(Centraal Bureau voor de Statistiek (CBS), 2024a) | | | | |
| Male | 31814 | 57.20 | 5145 | 52.84 |
| Female | 23807 | 42.80 | 4592 | 47.16 |
| **Age**(Centraal Bureau voor de Statistiek (CBS), 2024a) | | | | |
| <25 | 6683 | 12.02 | 1659 | 17.04 |
| 25–35 | 12837 | 23.08 | 2031 | 20.86 |
| 35–45 | 10635 | 19.12 | 1896 | 19.47 |
| 45–55 | 13125 | 23.60 | 1976 | 20.29 |
| 55–65 | 9478 | 17.04 | 1825 | 18.74 |
| >65 | 2863 | 5.15 | 350 | 3.59 |
| **Education**(Centraal Bureau voor de Statistiek (CBS), 2024a) | | | | |
| Low | 7602 | 13.96 | 1971 | 20.32 |
| Medium | 18874 | 34.67 | 3720 | 38.36 |
| High | 27996 | 51.37 | 4007 | 41.32 |
| **Standardised income of household (10% groups)** | | | | |
| First 10% group | 2497 | 4.49 | | 10.00 |
| Second 10% group | 1909 | 3.43 | | 10.00 |
| Third 10% group | 2736 | 4.92 | | 10.00 |
| Fourth 10% group | 3913 | 7.04 | | 10.00 |
| Fifth 10% group | 4728 | 8.50 | | 10.00 |
| Sixth 10% group | 5421 | 9.75 | | 10.00 |
| Seventh 10% group | 6563 | 11.80 | | 10.00 |
| Eighth 10% group | 7778 | 13.98 | | 10.00 |
| Ninth 10% group | 9405 | 16.91 | | 10.00 |
| Tenth 10% group | 9637 | 17.33 | | 10.00 |
| **Car ownership**(Centraal Bureau voor de Statistiek (CBS), 2023a) | | | | |
| Yes | 30318 | 54.51 | 6927 | 68.00 |
| No | 25303 | 45.49 | 1447 | 32.00 |

This bias is not unexpected, as urban regions typically have higher population densities and higher shares of highly educated residents. While this hinders the generalisability of the results to the entire Dutch workforce, it does not pose a substantive problem for the goals of this research, as these groups are of particular interest. Highly educated personnel are exactly the types of individuals that are currently underrepresented in the labour market of the Port of Rotterdam. Understanding their travel behaviour, residential context, and accessibility patterns is therefore crucial for identifying barriers and opportunities to attract them to port related jobs.

Moreover, the filtering of postcodes with insignificant number of respondents ensures a filtering of unrealistic work locations that might impact the accuracy and validity of the result. As a result, the LCCA has more meaning, since the locations reflect actual locations with job opportunities and the mobility patterns and socio-demographics characteristics clusters are significant and comparable to the port of Rotterdam.

## 4.4. Preprocessing vacancy data

For the analysis of vacancies, the vacancy data needs to be preprocessed. The full dataset contains 345,059 initial entries of vacancies with numerous different variables. However, not every vacancy online contains the same information and scrapers cannot always find all the information they are looking for. This results in many incomplete variables such as "contractduration", "experience" and "typeofcontract". Therefore, only the "educationdegreeminimumrequired" and "brc2014beroepsgroep" are selected alongside the "datefound" and "datedeactivated" which are needed for the duration. As mentioned beforehand, the "datefound" variable has an uncertainty of somewhere between 0 and 7 days (see figure 4.8). This means that in some cases, the vacancy duration can even be negative, as the vacancy is found after it has been put online and even filled.

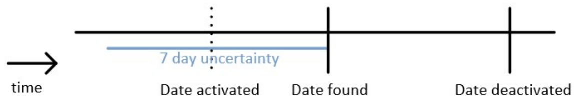



**Figure 4.8:** Uncertainty of the starting date of the vacancy

To account for this, a correction was applied. Since a vacancy can be posted or filled at any point within the 7-day interval, the expected lag is defined as half the scrape interval (7/2 = 3.5 days). This constant was added to the raw duration to reach a more realistic estimation of the actual time a position remained open. Durations that remained negative after this correction (due to immediate deactivation) were truncated to zero to maintain logical consistency. Additionally, 9136 values had no datedeactivated, so these entries were not taken into account.

$$\text{duration} = \text{datedeactivated} - \text{datefound} + 3.5 \tag{4.8}$$

### 4.4.1. Aggregation to PC4 level

In order to run the regression at PC4 level, the data needs to be aggregated. As previously mentioned in section 4.3.4, a minimum of 30 vacancies is needed for statistically significant values. A total of 203 postcodes were present in the data, of which 183 had enough vacancies to be statistically significant. This filtering procedure resulted in a final study sample of $N = 183$ postcodes. This threshold is particularly critical for the vacancy rate indicator. In postcodes with very low vacancy volumes, small absolute differences can lead to disproportionately high relative variance, which can introduce artificial spatial clustering and bias the regression estimates.

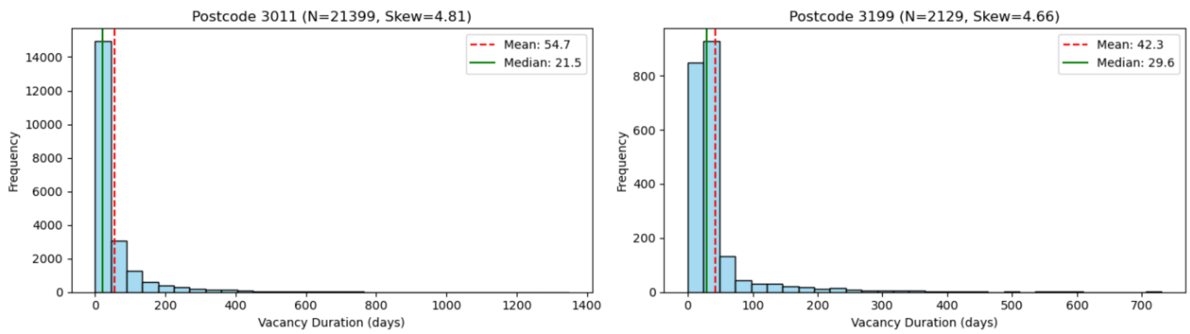



**Figure 4.9:** Comparison of histograms in the city centre of Rotterdam (left) and the Maasvlakte (right)

Figure 4.9 shows histograms for the vacancy duration variable of two statistically significant postcodes. Vacancy duration is typically right-skewed. In a right-skewed distribution, the mean is pulled upward by a small number of extreme outliers (vacancies that remain open for hundreds of days). As shown in the comparison, the Rotterdam City Centre (left) typically has a higher mean than the Maasvlakte (right) due to a higher absolute volume of vacancies and, consequently, more extreme maximum values. However, the median in the centre is often lower, as the majority of vacancies there are filled more efficiently. Therefore, the median vacancy duration is chosen as the primary dependent variable to ensure a robust representation of the vacancy duration. This is consistent with literature as Dossche et al. (2025) also use the median to select the hard-to-fill vacancies.

**Table 4.7:** Overview of Aggregated Variables at PC4 Level for Regression Analysis

| Category | Variable | Description | Transformation |
|---|---|---|---|
| **Outcomes** | Median Duration | Median corrected vacancy duration | |
| | Vacancy Rate | Vacancies per year relative to job count | |
| **Accessibility** | Hansen Acc. | Potential accessibility to labour supply | log / Mean-Centered |
| | Competitive Acc. | Shen's competition-adjusted accessibility | Mean-Centered |
| **Context** | Address Density | Number of addresses per $km^2$ | Mean-Centered |
| | Interaction | $Hansen\_Acc \times Address\_Density$ | Interaction term |
| **Specifications** | Fract. High Ed. | Share of vacancies requiring HBO/WO | Continuous (0–1) |
| | Fract. Placebound | Share of vacancies which are bound to the workplace | Continuous (0–1) |

Table 4.7 shows the variables used in the regression. For the fraction placebound vacancies, the top 20 most frequent professions based on BRC2014 codes (Centraal Bureau voor de Statistiek, 2014) in the Port of Rotterdam were identified. From this top 20 list, professions were classified as "placebound" if the work is inherently location-specific. Examples include construction worker (BRC 721), transport planners (BRC 435), dock workers (BRC 1221), and maintenance technicians (BRC 712). The specific BRC codes identified in the Port area were then used as a filter for the entire Rijnmond dataset. For every postcode, the number of vacancies matching these "placebound" codes was counted and divided by the total number of vacancies in that postcode. For the fraction of highly educated positions, vacancies explicitly requiring HBO or WO degrees were selected.

To ensure the reliability of the regression coefficients, specifically when incorporating interaction terms, a mean-centering procedure was applied to the continuous independent variables Accessibility and Address Density.

$$X_{centered} = X_i - \bar{X} \tag{4.9}$$

Interaction terms (Accessibility × Density) are mechanically correlated with their constituent variables, which often leads to high Variance Inflation Factors (VIF). By subtracting the mean from each value, the correlation between the main effects and the interaction term is significantly reduced. Additionally, this ensures that the coefficients $\beta_1$ and $\beta_2$ can be interpreted as the effect of one variable when the other is at its average value, improving the interpretability of the model.

# 5 Results

This chapter presents the findings of the empirical analyses and interprets them within the theoretical framework of this study. First, section 5.1 explores the general spatial distributions of accessibility and modal split across the Netherlands. Building on these descriptives, section 5.2 details the Latent Class Cluster Analysis, identifying distinct typologies of Dutch commuters and with specific attention to the industrial profile and the Port of Rotterdam. Section 5.3 then shifts the focus to the Rijnmond region, utilizing (spatial) regression to quantify how workplace accessibility predicts local vacancy outcomes. The core findings from both methodological approaches are synthesised in section 5.4 and evaluated against the conceptual model in section 5.5. Finally, section 5.6 provides a discussion on the found results.

## 5.1. Spatial patterns of accessibility and modal split

To establish a foundational understanding of the data before presenting the advanced models, this section explores the spatial distributions of accessibility and modal split across the Netherlands. The socio-demographics were already introduced in section 4.3.4. The variables accessibility and modal split outcomes are shown in this section.

### 5.1.1. Comparative analysis of accessibility measures

Figure 5.2 shows the two accessibility measures side by side. The Hansen measure is simpler as it considers the amount of residents that can reach a certain work location within 2 hours, scaled to the impedance function. As expected, this means that the Randstad scores relatively better than the rest of the Netherlands, due to a higher populations and a much denser public transport network. This also means that the suburban areas in the Randstad still have a relatively high accessibility compared to the rest of the Netherlands as many people can still reach these locations within time. Large cities essentially boost the accessibility of the nearby postcodes. Therefore, Friesland and Zeeland remain largely inaccessible. The port area performs relatively fine in comparison to all of the Netherlands, as it is close to Rotterdam, and it can thus attract quite some people. However, it still performs worse than most other postcodes in the province. The descriptive statistics for both measures is given in table 5.1. Competition is a scaled measure, so the values are lower.

**Table 5.1:** Descriptive statistics for the accessibility measures

| Statistic | Hansen Accessibility | Competitive Accessibility |
|---|---|---|
| Count | 4,070 | 4,070 |
| Mean | 377,526 | 1.28 |
| Std. Deviation | 384,806 | 0.69 |
| Minimum | 0 | 0.00 |
| 25th Percentile | 116,683 | 0.81 |
| Median (50%) | 229,721 | 1.17 |
| 75th Percentile | 513,630 | 1.66 |
| Maximum | 2,852,805 | 5.13 |

Figure 5.1 shows that the distributions of both measures are very different. Hansen measures a total potential, whereas Shen measures a relative potential based on competition. The Hansen accessibility follows a heavily right-skewed distribution with its highest peak near zero, indicating that most locations experience low total potential accessibility. For this reason, the logarithm of the accessibility is used in section 5.3. Conversely, the Shen accessibility shows a more bell-shaped distribution covering a smaller range. Because this measure divides the potential accessibility by the number of other jobs that can be accessed by the same people, it results in a less extreme and more balanced distribution of scores.

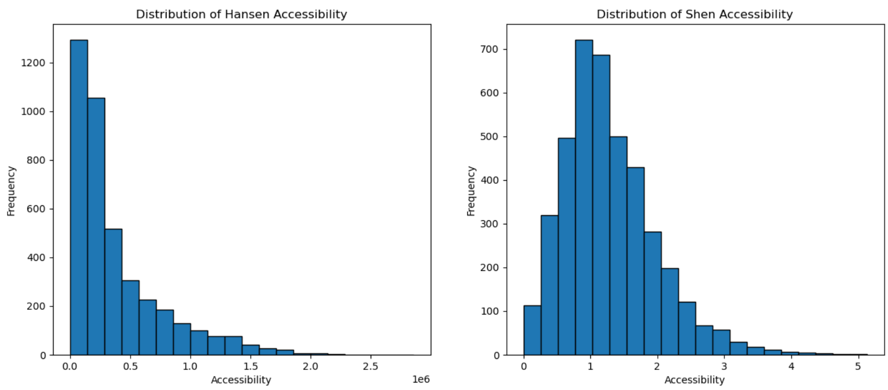
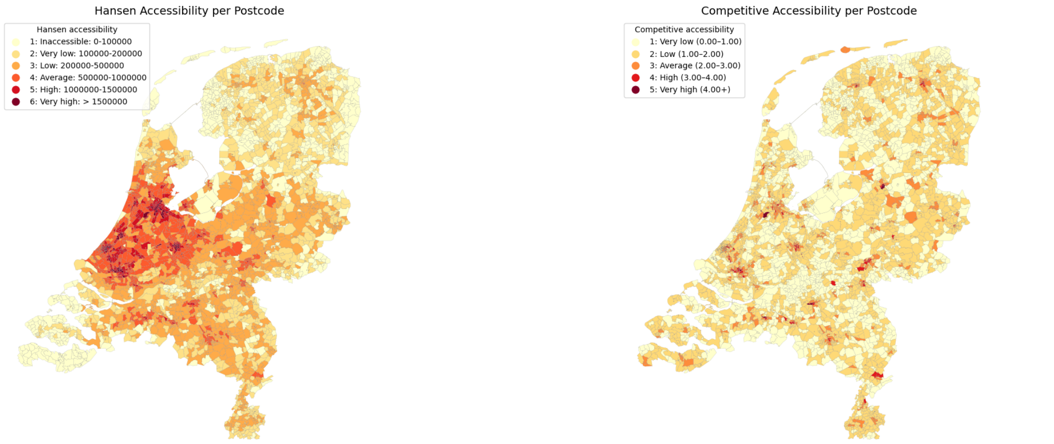



**Figure 5.1:** Distributions of the accessibility measures: Potential (left) and Competitive (right) accessibility

When competition is taken into account, a combination of many nearby residents and low travel times to these residents is needed for a high accessibility. People are divided over the firms where the nearest firms receive the largest shares. Accessibility is lower when there are many competing jobs nearby. Locations with many available jobs and good connectivity are therefore more favoured as they have a strong pull and weigh down the other locations. For example, Schiphol and Rotterdam Centraal are locations where many people work with their own large train station. These locations are always reached faster than nearby postcodes, thus they "steal" residents who can potentially work in surrounding postcodes. In this measure, your position relative to other nearby postcodes is much more important. Therefore, as shown in figure 5.2, the high accessibilities are much more scarce and spread across the country for the competition measure. Note that Zeeland now has quite average accessibility, but their highest accessibility postcodes are still quite far below the maximum. In dense urban centres, the massive supply of reachable residents is divided by an equally massive concentration of competing jobs. This mathematical relationship penalises job-rich areas, bringing their high potential accessibility scores down to a more moderate level. Conversely, in peripheral areas like Zeeland, the absolute number of reachable residents is low, but the competition is also minimal. Therefore, these postcodes maintain a higher relative score than they would in a purely potential model. This results in a distribution where the peaks are lowered and everything is closer together.




**(a)** Accessibility based on Hansen's potential model

**(b)** Competitive accessibility based on Shen's gravity model

**Figure 5.2:** Comparison of accessibility measures

### 5.1.2. Spatial distribution of the modal split

The commuting modal split is plotted in figure 5.3. The commuting modal split for every postcode is visualised using the merging algorithm explained in appendix H. Public transport usage is very dependent on the urbanity of the location. The largest cities in the Netherlands clearly have the highest PT share, especially in the Randstad. The Port of Rotterdam is very car dominated and one of the regions with the highest car share. For the Slow modes, there does not seem to be a very notable pattern, except that the Waddeneilanden seem to have a high share. This is to be expected since everything on these islands is in cycling distance, and commutes to the mainland take a lot of time.

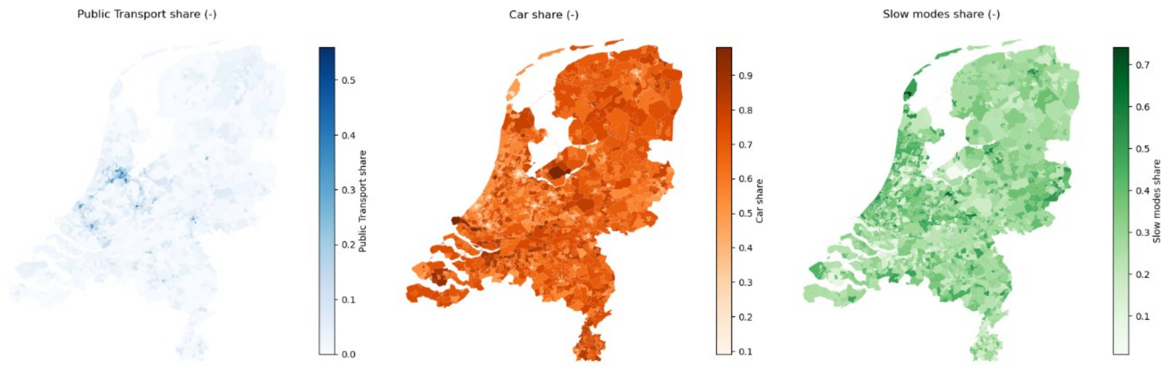



**Figure 5.3:** Commuting modal split in the Netherlands aggregated on PC4 level

## 5.2. Typologies of the Dutch commuter

While the descriptive statistics provide a general overview of accessibility and travel behaviour, they do not reveal the complex ways in which socio-demographics, modal choices, and spatial context interact. To uncover these underlying patterns, this section presents the results of the Latent Class Cluster Analysis (LCCA). By grouping workplaces into distinct typologies, the analysis demonstrates how different segments of the labour market relate to public transport accessibility. This section first outlines the identification and statistical foundation of the commuter profiles (5.2.1). Subsequently, the defining characteristics (5.2.2) and the national spatial distribution (5.2.3) of the six resulting clusters are discussed. Finally, the analysis zooms in on the industrial commuter profile to assess how these national mechanisms work within the specific context of the Port of Rotterdam (5.2.4).

### 5.2.1. Identification of commuter profiles

To move beyond the individual descriptive statistics and uncover the underlying commuter patterns within the dataset, a Latent Class Cluster Analysis (LCCA) was performed. The objective of this analysis is to categorize Dutch postcodes into distinct groups based on shared socio-economic and transport-related characteristics, in order to understand which types of commuters are attracted to which locations, and how these commuters are associated with public transport accessibility. The Latent Class Cluster model was set up using Latentgold (Statistical Innovations Inc., 2016). Firstly, the measurement model was estimated for 1 up until 10 classes. The 6-class model performed best and was expanded with the structural model (see appendix D for a more detailed explanation).

The structural model consisted of an OV accessibility metric, which were all tested to see which had the most explanatory power. The metrics tested were Hansen ($R^2$ = 8.34%), competitive Shen ($R^2$ = 6.65%) and distance to nearest train station ($R^2$ = 2.84%). The standard R-squared measure, which indicates the portion of the variability in the class membership is explained by the added covariate, implies that the potential accessibility measure has the largest impact. The value shows that accessibility does aid in predicting class membership, but it is not a very strong predictor as other covariates are likely to contribute as well. Table D.4 show the parameters and z-values from the final model. The parameters of the accessibility covariate show that clusters 1 and 3 are very negatively associated with accessibility, whereas clusters 5 and 6 are positively associated with accessibility. The inactive covariates do not contribute to the cluster assignment, but can be used to explain differences in the clusters.

**Table 5.2:** The within-cluster distributions of indicators and covariates

| Indicator | Cl1 | Cl2 | Cl3 | Cl4 | Cl5 | Cl6 | Overall |
|---|---|---|---|---|---|---|---|
| *Cluster size* | 31.87% | 18.32% | 15.64% | 14.63% | 13.38% | 6.17% | 100.00% |
| *Gender* | | | | | | | |
| Male | 57.65% | **48.88%** | **75.22%** | **46.38%** | 52.13% | 56.63% | 56.34% |
| *Age groups* | | | | | | | |
| <25 | 13.71% | 8.81% | 7.07% | **22.77%** | 15.06% | 7.55% | 12.90% |
| 25–35 | 18.32% | 21.93% | 21.50% | 16.28% | 26.76% | 32.66% | 21.19% |
| 35–50 | 29.97% | 30.62% | 32.57% | 25.16% | 27.32% | 32.47% | 29.59% |
| 50–65 | 30.66% | 33.22% | 34.62% | 27.61% | 25.48% | 24.38% | 30.22% |
| *Education* | | | | | | | |
| Low educated | 19.89% | 8.66% | 18.05% | **20.50%** | 11.72% | 5.00% | 15.62% |
| High educated | **36.53%** | 59.26% | 37.63% | 37.91% | 57.17% | **70.60**% | 45.93% |
| *Income* | | | | | | | |
| Low income | 11.97% | 11.29% | 9.81% | 14.10% | 19.09% | 12.54% | 12.81% |
| High income | 44.05% | **52.23%** | 47.22% | **43.95%** | 43.39% | **55.13%** | 46.63% |
| *Car ownership* | | | | | | | |
| Car ownership | **57.41%** | 54.68% | **65.02%** | 47.75% | 42.27% | 42.95% | 53.77% |
| *Social participation* | | | | | | | |
| Part-time | 20.45% | 21.62% | **10.67%** | 26.51% | 18.45% | **11.53%** | 19.21% |
| Students | 6.80% | 4.02% | **2.28%** | **12.96%** | **9.11**% | 3.29% | 6.58% |
| *Household* | | | | | | | |
| Single households | 15.18% | 18.51% | 17.75% | 14.18% | 25.02% | 22.75% | 17.83% |
| Households with children | 55.58% | 51.36% | 52.73% | **59.65%** | 45.12% | 45.30% | 52.92% |
| **Covariates** | **Cl1** | **Cl2** | **Cl3** | **Cl4** | **Cl5** | **Cl6** | **Overall** |
| *Active* | | | | | | | |
| Accessibility | 0.1615 | 0.2776 | **0.1478** | 0.2136 | 0.4106 | 0.4837 | 0.2414 |
| *Inactive* | | | | | | | |
| Car usage | 65.19% | 59.70% | **75.63**% | 53.78% | **40.83**% | 40.56% | 59.37% |
| Public transport usage | **3.07%** | 9.19% | 3.22% | 5.34% | 16.88% | **27.51%** | 7.90% |
| Address density | 0.1172 | 0.1956 | **0.0840** | 0.1630 | **0.3940** | 0.3412 | 0.1839 |
| Jobs | 0.0978 | 0.1331 | 0.1287 | **0.0753** | 0.1517 | 0.3186 | 0.1266 |

### 5.2.2. Characteristics of the six clusters

Table 5.2 presents the socio-demographic, mobility, and spatial characteristics of the six identified commuter profiles. Rather than relying on predefined categories, these clusters emerged organically from the latent structure of the data. The defining features of each cluster can be interpreted by comparing their specific indicator shares (such as gender, education, and car usage) against the overall national average. This subsection provides a first characterisation of each cluster.

**Cluster 1: Family car-oriented peripheral commuters**

This is the largest cluster, characterized by low accessibility, a high car share, and a relatively low education level. The cluster is not defined by distinctive socio-demographic factors, indicating a balanced variation of jobs and commuting. This lack of clear differentiation is likely a result of the limited number of distinct suburban or peripheral postcodes, causing them to be aggregated into a single, broader group. Furthermore, address density suggests that this cluster primarily consists of non-urban areas, containing smaller local or family businesses. Public transport accessibility and usage are both quite low, presumably because these villages lack a train station.

**Cluster 2: Older, highly educated multimodal commuters**

This group can be distinguished by relatively older commuters, a high share of women, and highly educated individuals. The workplaces in this cluster are easily accessible by public transport, and public transport usage is accordingly high. These areas are more urban than those in cluster 1 and likely represent the suburbs of major cities or smaller cities. The higher proportions of females and highly educated people could indicate a concentration of jobs in healthcare, education, or retail (Centraal Bureau voor de Statistiek (CBS), 2024b). The group contains relatively few students, suggesting that these urban areas are either not close to universities, or that the local jobs are less attractive to students.

**Cluster 3: Male, car-dependent industry commuters**
This commuter group is highly distinct due to its extremely high share of males, high car ownership, and the lowest level of public transport accessibility. This cluster primarily consists of industrial or agricultural postcodes located in more remote areas. Despite having the lowest public transport accessibility, the total number of jobs rivals that of urban areas. Approximately 61% of the commuters are older than 35, while the number of students and commuters under 25 is the lowest among all groups. Postcodes in the Port of Rotterdam, along with other ports and industrial areas in the Netherlands, all fall within this cluster.

**Cluster 4: Young slow-mode commuters**
This cluster is defined by a large group of young commuters and students who mainly rely on slow modes of transport. The low share of highly educated people and the low job density indicate that this group works primarily for local businesses. Public transport usage is also low in these areas. There are few single-person households, suggesting that these commuters mostly live with their parents or have children themselves. Furthermore, the high share of females indicates that these jobs belong to sectors with strong female representation. These positions could also be part-time or side jobs, which explains the lower average income.

**Cluster 5: Highly educated students and starters**
The fifth cluster is characterized by high shares of commuters younger than 35. The relatively low share of less-educated individuals, combined with a large proportion of highly educated people and students, suggests that this cluster includes individuals studying at the HBO or university level. This is further supported by the low car usage and high reliance on slow transport modes, indicating that people in this cluster live close to their workplaces. The cluster features the highest address density, pointing to mixed-use areas with both residences and workplaces. Interestingly, the proportion of high-income earners is the lowest across all clusters; this can be explained by the large number of students and single households, meaning these commuters are either just starting their careers or have yet to enter the full-time workforce. Public transport accessibility is high and is utilized to a great extent.

**Cluster 6: Career-focused, high-income commuters**
The final cluster is the smallest, yet it represents the wealthiest and most highly educated group. This cluster contains a high proportion of individuals between 25 and 50 years old, and boasts the highest job density of all locations. It likely encompasses postcodes situated in the central business districts of major cities, housing large multinational companies. Additionally, the cluster has the highest average accessibility, suggesting close proximity to central stations or major transit hubs.

### 5.2.3. Spatial distribution of the clusters

The location of the found clusters can reveal more information about what types of jobs are available in these areas and what type of people are attracted. Figure 5.4 shows which postcodes belong to each of the clusters.


**Figure 5.4:** Posterior classification of the postcodes

Clusters 1 and 3 are the most peripheral with the least number of $addresses/km^2$ and this can also be seen in the plot. From the locations of the postcodes of cluster 1, it can be seen that these postcodes are generally isolated and outside densely populated areas. These postcodes generally represent an entire village, which also explains the mixed profile of these postcodes, as a village can facilitate pretty much every type of job. However, high-end jobs that require high education are generally not situated in these locations which explains the low number of high educated commuters. Public transport in these areas is mainly bus traffic which explains why the car is this prominent.

Cluster 2 can be seen in moderately urban areas around the country and in the suburbs of the largest cities. Using google maps, the postcodes of some of the larger hospitals in large cities were compared to the list of postcodes in cluster 2. Almost all these postcodes were assigned to cluster 2, except some research oriented hospitals (cluster 5) or hospitals in dense areas of the city (e.g. Erasmus MC) which were assigned to cluster 6. This also explains why females are strongly represented in this cluster.

As mentioned already, cluster 3 is an industrial cluster which also includes the port of Rotterdam postcodes. All the postcodes from the port of Rotterdam (3196-3199) have a 100% certainty of belonging in this cluster. Some other well known Dutch ports also belong in this cluster. The North Sea Port (Vlissingen, Terneuzen), port of Amsterdam (Westpoort, Zaandam, Ijmuiden) and Groningen Seaport (Delfszijl) are all in this cluster. Also, the Schiphol, Eindhoven, Lelystad and Maastricht airport postcodes belong in this cluster, even though less convincingly. Only Groningen (1, rural location) and Rotterdam airport (2, also contains a hospital) are not part of this cluster. Another clear cluster of industry is the Brainport area. The high-tech companies need to manufacture and distribute their products and this happens mainly in the surrounding area. Lastly, Chemelot is also a large industry cluster in the middle of Limburg which also belongs in this cluster. A more detailed analysis of this cluster follows in section 5.2.4.

The locations for cluster 4 are widely spread across the country. The locations of these areas are in or surrounding moderately urban areas which do not have that many jobs available but generally have (multiple) MBO or HBO schools, which explains the high student density. Also, households with children are highest of all clusters, indicating that these students live at home and cycle to school.

Cluster 5 can be found in all the larger student cities with a university or HBO school. This fits the profile of the cluster. All the postcodes with large campus which is not heavily occupied by companies or start-ups are in this category. The cluster does also have quite high accessibility and a high share of high educated people, but not as homogenous as cluster 6, so this cluster can be found in more places, generally city centres.

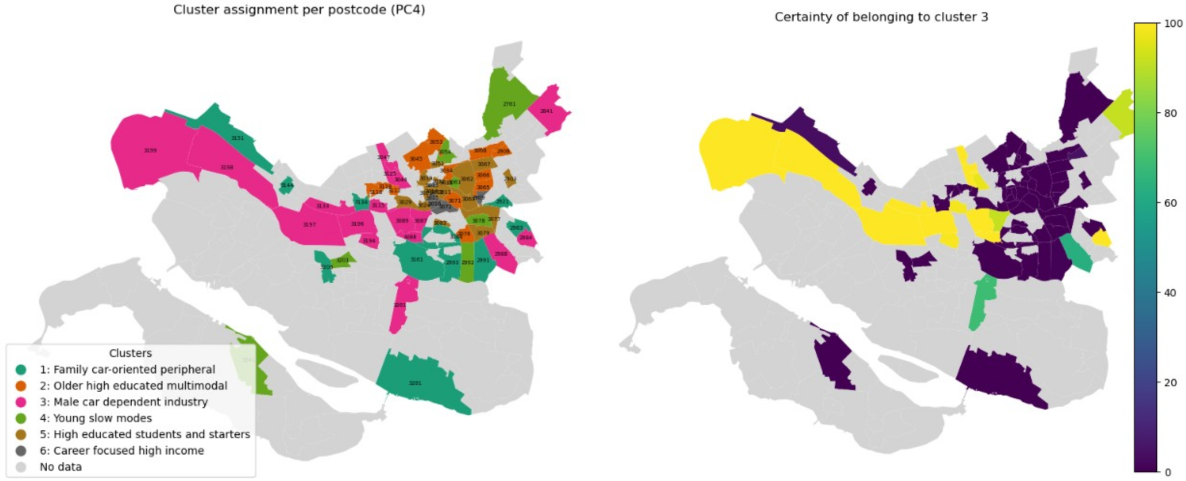
Since cluster 6 is the smallest and wealthiest, the economic knowledge clusters such as the Zuidas in Amsterdam and Kop van Zuid in Rotterdam are logical locations and these are indeed included in this cluster. Other locations are found in Delft (Technopolis), Den Haag (HSD campus), Utrecht (Beurs, Papendorp), Hilversum (Mediapark) and Groningen (Zernike Campus, UMCG). These are all locations that attract very sector specific high educated personnel and are thus more homogenous postcodes. The absence of Eindhoven, which is also a notable high educated tech cluster, can be considered peculiar, but this could be because the ASML headquarters is next to a hospital which is why this postcode is allocated to cluster 2. Additionally, ASML is not as accessible by public transport as the other knowledge clusters so this affected the cluster probabilities as well.

The inherent overrepresentation of wealthy and highly educated individuals in the data distinctly influences the clustering outcomes. Because of this imbalance, the largest, more peripheral clusters lack sharply defined socio-demographic profiles. As a result, the most pronounced differentiation between clusters is found within urban areas, which benefit from a higher number of statistically robust postcodes. Moreover, local commuter profiles are strongly tied to the dominant economic activities in these areas. Cluster 1 is a mixed profile and shows no sign of domination of a particular sector, whereas cluster 2, 3 are clearly dominated by a certain sector, which results in a distinct commuter profile. As for cluster 4 and 5, the education level and urbanity are important, as well as the presence of schooling and or universities

### 5.2.4. Deep dive: The industrial profile and the Port of Rotterdam

To assess whether labour market frictions observed in the Port of Rotterdam are specific to local conditions or reflect a broader national mechanism, this section provides a detailed analysis of cluster 3. As mentioned earlier, the port area falls under cluster 3. The Rijnmond area is shown in figure B.1. All the postcodes with a port function in this area belong to the cluster.

From the plots, it becomes clear that the port area postcodes are easily assigned to the third cluster by the model. This is verified by the shares which can be found in appendix B.1. The postcodes have high




**(a)** Clusters within the Rijnmond area

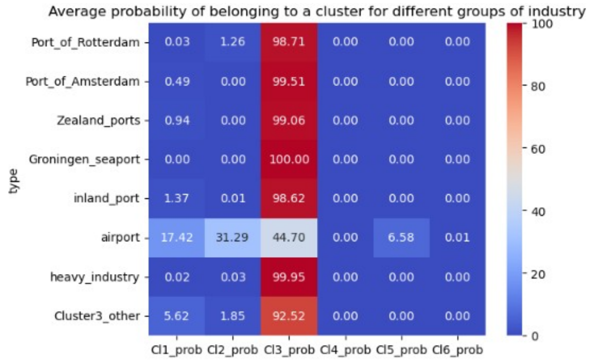
**(b)** Certainty of belonging to cluster 3 [%]

**Figure 5.5:** Clusters in the Rijnmond area and the certainty of belonging in Cluster 3

shares of males, high car ownership, low accessibility and an extremely high car usage. Interestingly, the share of high income is higher than that of cluster 6 even. Even though cluster 6 has a high ceiling of CEOs with extreme salaries, there is also plenty of retail and horeca with relatively low salary. On the other hand, the port area has relatively uniform industrial jobs. There are almost no retail, horeca or other services. On top of that, due to the location, shift work and specialised jobs, the wage floor is quite high. High incomes in Cluster 3 thus reflect wage compensation for shift work, specialisation and limited job mobility rather than flexible or highly accessible employment. Lastly, most of the postcodes are less accessible than the cluster average. Only the postcodes closer to the city centre with almost direct access to the metro have better accessibility than average. In most postcodes, the car is the only used commuting mode.

To assess whether these patterns are specific to the Port of Rotterdam or reflect a broader national mechanism, other industrial regions classified as Cluster 3 are analysed. Some relevant clusters have been grouped and marked in figure 5.6. There are many more areas which could be relevant, but this selection was deemed to be relevant for comparison. The postcodes and their reason for selection is explained in appendix F. All the seaports are undoubtedly part of cluster 3, with limited differences. Both the inland ports, and the other cluster 3 postcodes, have small similarities to cluster 1, which means these postcodes are slightly less dominated by industry.


(a) Cluster 3 postcodes and relevant land use characteristics




**(b)** probability of belonging to cluster 3 for the different groups

**Figure 5.6:** Cluster 3 location subgroups and their probabilities

The notable exception within this cluster is the presence of the airports. This is reflected in the shares, with a notable shares of cluster 1 and 2 also present in these postcodes. This implies a more heterogeneous mix of economic activities within the airport postcodes. Furthermore, it highlights that not all airports serve the same primary function, as their economic activities vary depending on their scale and strategic purpose. The airports have highly different accessibility scores as the Schiphol score is close to 1 (One of the highest scores of all postcodes), whereas Lelystad airport is close to 0, largely a consequence of their difference in purpose. Some of the airports are close to large cities but others are located in much more peripheral areas. Nonetheless, each airport requires jobs such as ground handling, cargo operations, maintenance and other operational jobs that are highly place-bound and car-dependent. Schiphol also has a higher share of high-skilled managerial, engineering and IT related employment. Additionally, The postcode containing Rotterdam the Hague airport also contains a hospital, which explains the presence of cluster 2. This mixed composition highlights that airports combine industrial characteristics with service and knowledge oriented activities. As a result, they show higher average accessibility and greater modal diversity than other industrial locations. This shows that more industrial or place-bound clusters, when well connected by public transport, can attract higher educated and diverse workforce.

**Table 5.3:** Comparison of different location types within Cluster 3

| Indicator | Rotterdam | Amsterdam | Zealand | Groningen | Inland | Airport | Heavy Ind. | Other |
|---|---|---|---|---|---|---|---|---|
| *Age groups* | | | | | | | | |
| <25 | 5.0% | 8.0% | 8.0% | 2.0% | 8.0% | 10.0% | 6.0% | 7.0% |
| 25–35 | 20.0% | 21.0% | 19.0% | 27.0% | 23.0% | 22.0% | 22.0% | 21.0% |
| 35–50 | 35.0% | 29.0% | 33.0% | 12.0% | 33.0% | 36.0% | 27.0% | 33.0% |
| 50–65 | 37.0% | 38.0% | 38.0% | 56.0% | 33.0% | 26.0% | 41.0% | 34.0% |
| 65+ | 3.0% | 3.0% | 3.0% | 2.0% | 3.0% | 6.0% | 5.0% | 4.0% |
| *Gender* | | | | | | | | |
| Male | **82.0%** | **83.0%** | 76.0% | **90.0%** | 75.0% | 69.0% | **82.0%** | 74.0% |
| *Education* | | | | | | | | |
| Low educated | 15.0% | 24.0% | 20.0% | 17.0% | 16.0% | 9.0% | 17.0% | 19.0% |
| High educated | 36.0% | 34.0% | 29.0% | 27.0% | 36.0% | 46.0% | 41.0% | 38.0% |
| *Income* | | | | | | | | |
| Low income | 4.5% | 9.1% | 10.5% | 2.4% | 9.8% | 9.1% | 6.8% | 10.5% |
| High income | **61.0%** | 48.0% | 46.0% | **59.0%** | 48.0% | 52.0% | 54.0% | 45.0% |
| *Car ownership* | | | | | | | | |
| Car ownership | 67.0% | 57.0% | 73.0% | 76.0% | 66.0% | 60.0% | 75.0% | 64.0% |
| *Social participation* | | | | | | | | |
| Part-time | **5.0%** | 7.0% | 13.0% | **2.0%** | 11.0% | 16.0% | **6.0%** | 11.0% |
| Students | **1.0%** | 2.0% | 3.0% | 2.0% | 2.0% | 3.0% | 2.0% | 2.0% |
| *Household* | | | | | | | | |
| Single households | 18.0% | 21.0% | 18.0% | 17.0% | 17.0% | 18.0% | 14.0% | 18.0% |
| Households with children | 52.0% | 54.0% | 48.0% | 46.0% | 55.0% | 51.0% | 53.0% | 53.0% |
| **Covariates** | **Rot.** | **Ams.** | **Zea.** | **Gro.** | **Inl.** | **Air.** | **Hvy.** | **Oth.** |
| *Active* | | | | | | | | |
| Accessibility | **0.12** | 0.23 | **0.02** | **0.04** | 0.08 | 0.27 | **0.10** | 0.16 |
| *Inactive* | | | | | | | | |
| Car usage | **89.0%** | 77.0% | 73.0% | **88.0%** | 78.0% | 79.0% | 73.0% | 74.0% |
| PT usage | **3.0%** | 8.0% | **0.0%** | **0.0%** | 1.0% | 9.0% | 1.0% | 3.0% |
| Address density | 0.05 | 0.01 | 0.04 | 0.05 | 0.06 | 0.03 | 0.07 | 0.09 |
| Jobs | 0.10 | 0.10 | 0.07 | 0.06 | 0.18 | 0.25 | 0.13 | 0.13 |

Table 5.3 demonstrates that although Cluster 3 locations share several structural characteristics, the cluster is internally heterogeneous for in multiple aspects. For instance, accessibility scores vary widely within the cluster, ranging from extremely low values in Zeeland (0.02) and Groningen (0.04) to substantially higher levels at airport locations (0.27). The Port of Rotterdam itself scores 0.12, positioning it between these extremes but still well below the national average.

These differences in accessibility are directly reflected in commuting behaviour. In Rotterdam and Groningen, car usage exceeds 85% and public transport usage is negligible (3% and 0%, respectively), whereas airport locations show lower car dependence (79%) and a notably higher share of public transport use (9%). Similarly, industrial locations in the Amsterdam region display lower car reliance (77%) compared to more peripheral industrial areas such as Zeeland (73%) and Groningen (88%). This can be attributed to the location of the port of Amsterdam which is much closer to the dense public transport network of the city.

Socio-economic composition also differs within Cluster 3. While all subgroups are predominantly male, the share of high-educated workers ranges from 29% in Zeeland to 46% at airport locations. High-income shares are particularly high in Rotterdam (61%) and Groningen (59%), suggesting that in more peripheral industrial contexts, wages partly compensate for limited accessibility and working conditions. In contrast, airport locations combine relatively high accessibility with a more diverse skill profile, reflecting their mixed industrial and service sectors.

These examples indicate that Cluster 3 does not represent a single, homogeneous industrial labour market. Instead, it groups together various industrial areas that share the same practical characteristics: the work is strictly tied to a physical location, resulting in a heavy reliance on cars. The Port of Rotterdam can therefore be understood as an extreme but not unique case within this broader family of industrial locations, where accessibility is generally a limiting factor.

## 5.3. Accessibility as a predictor for vacancies

To assess whether workplace accessibility is systematically associated with vacancies, this section analyses spatial variation in both vacancy duration and vacancy numbers across the Rijnmond area using regression analysis. Before introducing the regression models, some descriptive insights into the spatial distribution of accessibility and vacancy related indicators are shown, to illustrate the heterogeneity within the study area.

### 5.3.1. Spatial patterns in the Rijnmond area

Figure 5.7 illustrates the accessibility distribution in the Rijnmond area, using both Hansen and competitive accessibility measure. In both cases, accessibility is the highest around the Rotterdam city centre and along the larger nodes in the public transport network (figure B.1). The peripheral and industrial areas have a substantially lower accessibility. As mentioned previously, competitive accessibility accentuates the local differences more where otherwise poorly accessible locations receive a higher score because of better accessibility compared to near neighbours.

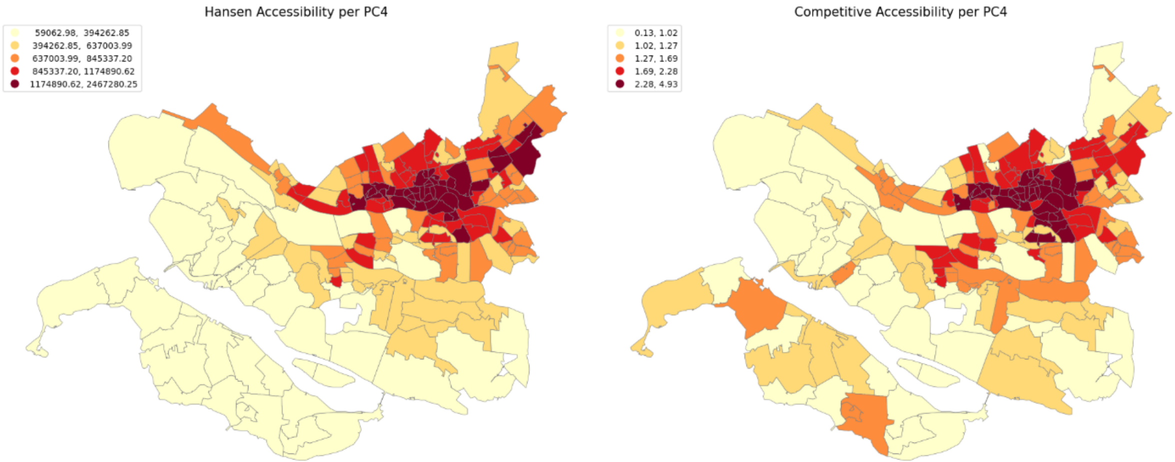



**Figure 5.7:** Hansen accessibility (left) and competitive accessibility (right) descriptives in the Rijnmond area

Figure 5.8 presents quintile plots of two vacancy related indicators. The median duration of open vacancies seems to be longer in the port area and other peripheral areas and durations seems to be more mixed in Rotterdam. The vacancy rate metric seems to be higher in Rotterdam (in line with Rouwendal and Rouwendal (2025)), but also in parts of Zeeland.

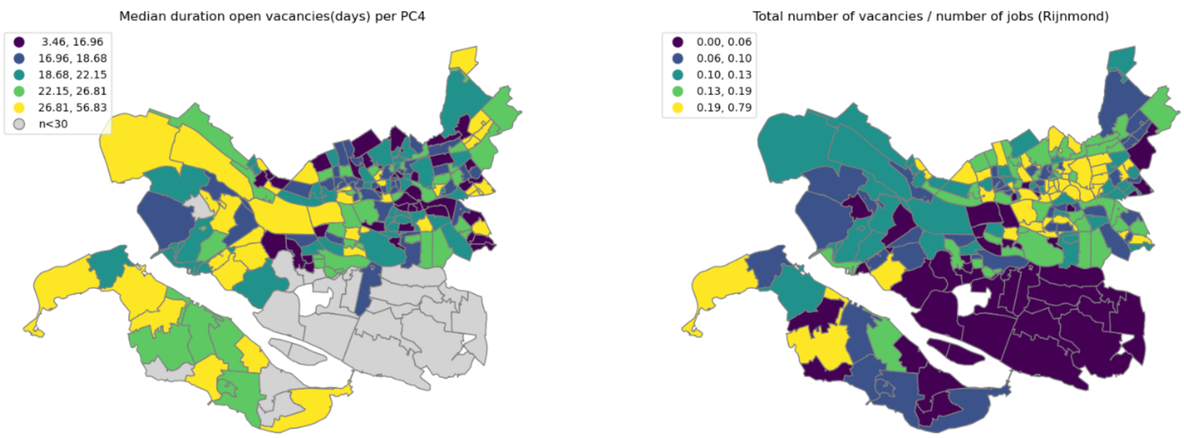



**Figure 5.8:** Vacancy descriptives in the Rijnmond area: Vacancy duration (left) and vacancy rate (right)

At first glance, no straightforward connection can be made between the accessibility metrics and the vacancy outcomes. Areas with high accessibility do not uniformly display shorter vacancy durations, and some low accessibility areas show moderate vacancy performance. This highlights the need for a multivariate analytical approach that controls for confounding factors such as address density, job concentration and workforce composition.

### 5.3.2. Spatial dependence and model validity

To assess the presence of spatial patterns in the vacancy indicators and to validate the independence of the model residuals, a Global Moran's I analysis was performed. The analysis is conducted on the filtered dataset ($N = 183$), ensuring that postcodes with insufficient vacancy observations do not introduce measurement noise. The results, based on a manually adjusted Queen's contiguity matrix that accounts for infrastructural links between the Rijnmond islands, are presented in Table 5.4.

**Table 5.4:** Global Moran's I statistics for vacancy indicators and OLS residuals ($N = 183$)

| Variable | Moran's I | p-value | Result |
|---|---|---|---|
| Median vacancy duration | 0.139 | 0.003 | Significant Clustering |
| OLS residuals (Duration) | 0.050 | 0.107 | Random |
| Vacancy rate | 0.089 | 0.003 | Significant Clustering |
| OLS residuals (Rate) | 0.068 | 0.064 | Random |

Table 5.4 shows that both the raw vacancy duration ($I = 0.139, p = 0.003$) and the vacancy rate ($I = 0.089, p = 0.031$) exhibit significant positive spatial autocorrelation. This confirms that labour market frictions are not randomly distributed but follow a clear regional structure.

After controlling for workplace accessibility and urban density, the spatial dependence in the residuals decreases substantially and becomes statistically non-significant at the $\alpha = 0.05$ level. For the vacancy duration model ($I = 0.050, p = 0.107$), the residuals are randomly distributed. For the vacancy rate model, the p-value (0.064) is also above the significance threshold, although it remains closer to the margin. This indicates that while some localised sectoral effects may still be present, the main spatial structure of vacancy fulfilment is effectively captured by the accessibility and density covariates. The reduction is greater for the vacancy duration, suggesting that the spatial dependence is better explained in this model. For the vacancy rate, the spatial dependence is slightly reduced, but has also become insignificant.

Visual inspection of the OLS residuals (Figure 5.9) supports these statistical findings. The error terms show no distinct geographic clusters of high or low values. Instead, a balanced mix of over- and underestimates is observed across the Rijnmond region, especially for the vacancy duration. Because the OLS specification successfully internalises the primary spatial dependencies, more complex spatial regression techniques are not required for a reliable interpretation of the accessibility effects.

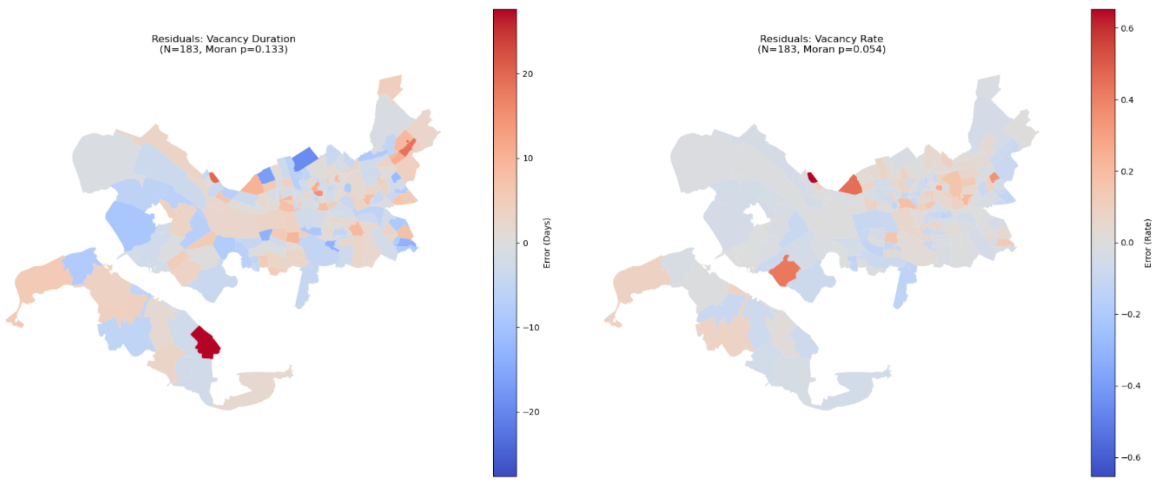



**Figure 5.9:** Map of the residuals for both vacancy indicators: vacancy duration (left) and vacancy rate (right)

This validation justifies the use of the OLS framework to identify the relationship between accessibility and vacancy fulfilment. To further investigate which local labour market characteristics contribute to these differences, the following section introduces a regression analysis linking vacancy outcomes to workplace accessibility and contextual factors.

### 5.3.3. Predicting vacancy outcomes

To add to the understanding of why the found spatial patterns in the vacancies appear, the results of the regression analyses are discussed. Both vacancy duration and vacancy rate are regressed to various sets (section 3.3.3 of explanatory variables in the data.

**Correlations between variables**

Before the results of the regression are discussed, the bivariate correlations are investigated. Vacancy duration shows a moderate negative correlation with both accessibility measures, with potential accessibility (Hansen) exhibiting a stronger association (-0.38) than competitive accessibility (-0.26). This pattern is consistent with the LCCA results and provides initial support for the hypothesis that better accessibility reduces vacancy duration. Additionally, the positive correlation between the share of placebound work and duration (0.24) highlights the specific friction within industrial and logistical sectors. This is likely compounded by a notable spatial mismatch, where roles requiring physical presence are often located in the least accessible postcodes (r = -0.42 between accessibility and placebound share)

Accessibility measures are strongly correlated with address density, indicating that both variables capture aspects of urban structure. However, the correlations are well below 0.9, suggesting that accessibility and density are related but distinct concepts. Overall, the matrix does not show any problematic multicollinearity. However, density and the accessibility metrics should be approached with caution.

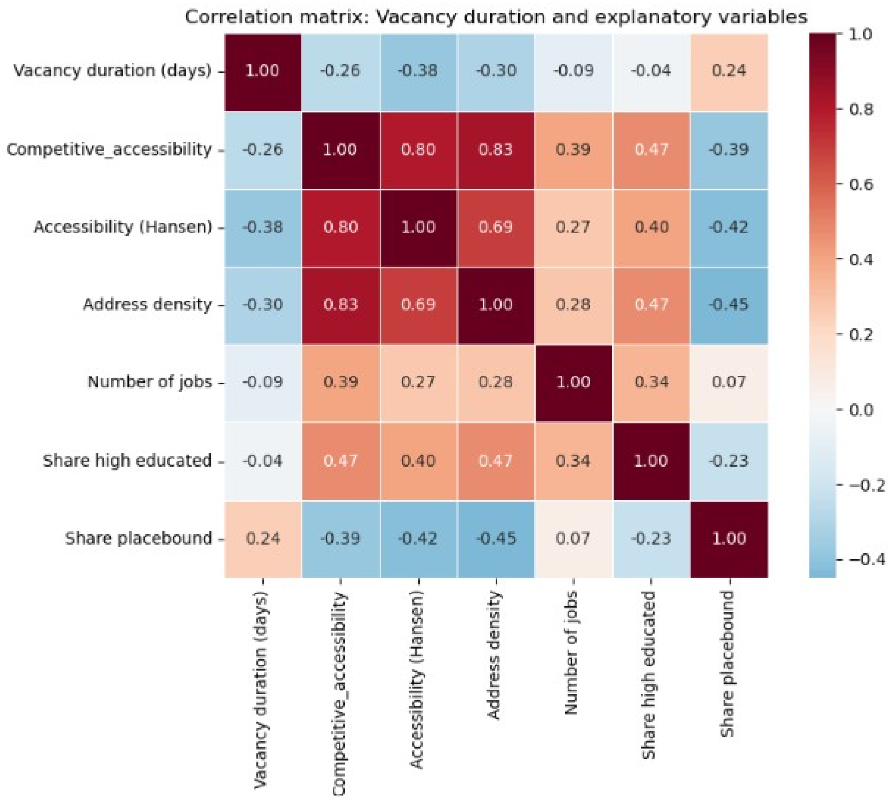



**Figure 5.10:** Bivariate correlations

**Vacancy duration**

The results for the regression are shown in table 5.5. Firstly, models $M1_a$ and $M1_b$ compare the explanatory power of the two accessibility measures. Similar to the LCCA results, the regression results indicate that competitive accessibility ($M1_b$) has lower explanatory power ($R^2$ = 0.070) compared to potential accessibility ($M1_a$, $R^2$ = 0.148). As expected, the sign is negative for both measures, meaning better accessibility results in lower vacancy duration. Accessibility appears to be an important predictor of vacancy duration, even without additional controls. Several factors may explain why a competition metric is less effective in the Rijnmond context: In the Rijnmond region, and specifically the Port of Rotterdam, the friction of distance outweighs the friction of competition. The geographical isolation of the Port area and the physical bounds imposed by the Maas mean that the primary challenge for employers is not necessarily competing for workers, but the fact that workers cannot physically reach the location within a reasonable time. In this context, the absolute potential labour supply (Hansen) is a more direct predictor of recruitment success than a relative competition ratio. Additionally, the performance of competitive accessibility is more sensitive to the spatial unit of analysis. As noted by Qu et al. (2025), lower spatial resolutions can introduce significant measurement errors in public transport accessibility. Because the Shen index requires a ratio of two spatially aggregated variables (labour supply and labour demand), any boundary noise at the PC4 level is compounded. This Modifiable Areal Unit Problem (MAUP) may increase the errors for a more complex measure. With these results in mind, the other models incorporate Hansen as the variable for accessibility.

**Table 5.5:** Regression Results: Determinants of Vacancy Duration in the Rijnmond Region

| Variable | ($M1_a$) Hansen | ($M1_b$) Competitive | ($M2$) Baseline | ($M3$) Interaction | ($M4$) Education | ($M5$) Placebound |
|---|---|---|---|---|---|---|
| Constant | 21.51*** | 25.16*** | 21.40*** | 20.22*** | 18.87*** | 20.89*** |
| | (0.46) | (1.16) | (0.65) | (0.60) | (1.46) | (0.76) |
| log_hansen_c | -3.44*** | | -3.01*** | -0.38 | -3.19*** | -2.81*** |
| | (0.71) | | (1.03) | (1.21) | (1.08) | (1.02) |
| Competitive_acc. | | -2.08*** | | | | |
| | | (0.49) | | | | |
| density_c | | | -0.0003 | -0.0014** | -0.0005 | -0.0002 |
| | | | (0.0003) | (0.0005) | (0.0003) | (0.0003) |
| interaction_c | | | | 0.0013*** | | |
| | | | | (0.0005) | | |
| fract_high_ed | | | | | 6.85* | |
| | | | | | (3.64) | |
| fract_placebound | | | | | | 4.84 |
| | | | | | | (4.12) |
| R-squared | 0.148 | 0.070 | 0.151 | 0.182 | 0.172 | 0.156 |
| Adj. R-squared | 0.143 | 0.065 | 0.136 | 0.168 | 0.154 | 0.137 |
| Max. VIF | 1.00 | 1.00 | 1.91 | 5.90 | 2.09 | 2.02 |
| N | 183 | 183 | 183 | 183 | 183 | 183 |

*Note:* Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. accessibility and density variables are centred.

Model $M2$ tests the effect of accessibility when controlling for address density, as dense areas tend to have a dense public transport network and the variables were highly correlated (0.69). Accessibility remains significant and strong. The addition of address density barely improves the $R^2$ of the model, meaning both variables explain the same variance but accessibility does it better (a model with density as sole determinant of vacancy duration yields $R^2$ = 0.0919).

Model $M3$ extends the analysis by introducing an interaction term between accessibility and address density, which increases the $R^2$ to 0.182. In this specification, the main effect of accessibility loses its significance, while both address density and the interaction term emerge as highly significant predictors.

This shift clearly indicates that accessibility does not operate uniformly across all locations; rather, its effect on vacancy duration depends heavily on the local urban context:

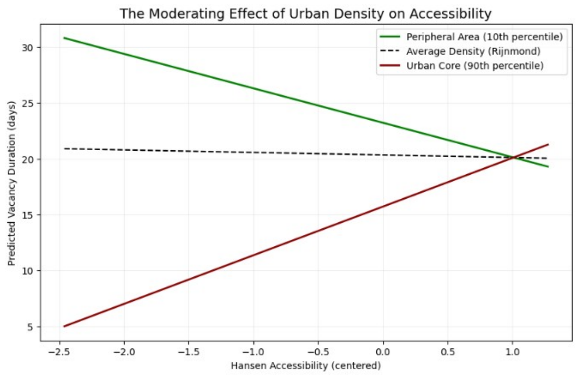
- In low-density areas: Better accessibility strongly reduces vacancy duration.
- In high-density areas: This beneficial effect diminishes and eventually reverses.

This contextual dependency explains why the isolated main effect of accessibility is no longer significant in this model. As visualised in Figure 5.11, the marginal effect of accessibility does not merely flatten out in highly urbanised areas, but actually changes direction. Counterintuitively, better accessibility in the densest urban contexts leads to longer vacancy durations.

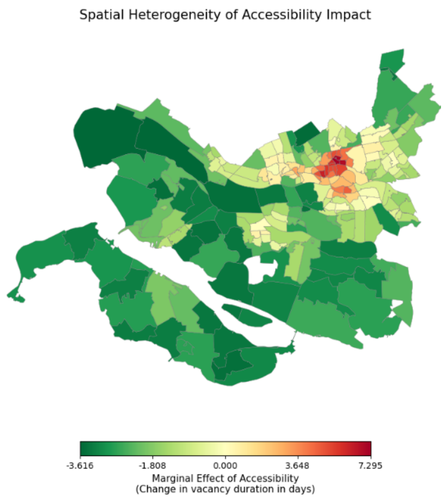
Finally, to ensure the validity of these interaction effects, the Variance Inflation Factors (VIF) were assessed. While introducing an interaction term inherently increases multicollinearity slightly, the maximum VIF in $M3$ reaches only 5.9, remaining well below the conventional threshold of concern (VIF < 10). Furthermore, the coefficient estimates remain stable across specifications, and the sign and magnitude of the baseline accessibility coefficient are comparable to $M2$. This confirms that the interaction effect is robust and not driven by collinearity.

This reversal can be explained through the separation of vacancy duration into an application period and a selection period (Van Ours & Ridder, 1993). In low density environments like the Port of Rotterdam, the primary bottleneck is the application period, thus the time required to attract any suitable candidates. Here, the strong negative correlation between accessibility and duration ($r = -0.38$) suggests that better transport links are essential to fill the pool of applicants. On the contrary, in dense urban areas, accessibility is generally high, ensuring that the application period is completed almost instantaneously. However, as the pool of applicants grows beyond an efficient optimum, the selection period becomes the dominant component of total duration. A larger applicant pool may lead to congestion friction, where the administrative burden outweighs the benefits of increased reach.

Furthermore, in these high access urban markets, employers may raise their employment standards. Instead of hiring the first suitable candidate, the long list of suited candidates allows firms to extend their search in pursuit of a perfect match, paradoxically lengthening the time a vacancy remains open (Puga, 2010). Thus, while accessibility facilitates the search, it can also complicate the selection, leading to the observed reversal of the effect in the most dense postcodes in Rotterdam.




**(a)** Interaction plot




**(b)** Spatial distribution (Map)

**Figure 5.11:** Visualisation of the moderating effect of address density on accessibility impact

Model $M4$ adds the share of vacancies which require high educated personnel. The coefficient is positive and weakly significant, suggesting that areas with a higher concentration of highly educated employment tend to experience longer vacancy durations. This aligns with the idea that higher-skilled positions often require longer matching processes. Importantly, the inclusion of education does not alter the core interaction between accessibility and density, reinforcing the robustness of the spatial mechanism identified earlier.

Model $M5$ introduces the share of place-bound employment. These are vacancies which fall under the 20 most frequent sought after profession groups in the port area, which cannot be performed remotely. According to van Lent et al. (2025), hybrid teleworking increases Hansen accessibility. The positive coefficient suggests that a high share of place bound jobs leads to longer vacancy duration. Although the coefficient is positive, it is not statistically significant once accessibility and density are controlled for. This suggests that sectoral composition alone does not explain vacancy duration beyond what is already captured by spatial accessibility and urban context.

**vacancy rate**
vacancy rate is included to assess whether the accessibility mechanisms identified for vacancy duration translate to an alternative labour market indicator. While vacancy duration captures the employer side difficulty of filling positions, vacancy rate reflects the balance between vacancies and jobs, and may therefore respond differently to spatial and occupational characteristics.

Table G.1 shows that accessibility measures are significantly related to vacancy rate in the simplest model specifications ($M1_a$ and $M1_b$). Both Hansen accessibility and competitive accessibility are positively associated with vacancy rate, indicating that better accessible areas tend to exhibit tighter labour markets. However, the explanatory power of these models is limited, with $R^2$ values below 0.05. Once additional controls are introduced, the accessibility effects lose statistical significance. In contrast to the vacancy duration models, interaction effects between accessibility and address density do not meaningfully improve model fit, and the coefficients remain small and unstable across specifications. The limited explanatory power may partly reflect the sensitivity of the vacancy rate indicator to low vacancy counts, which can dominate spatial patterns despite scaling by employment. vacancy rate is less sensitive to spatial accessibility patterns than vacancy duration, and may be driven more strongly by structural labour market factors. Overall, these results indicate that accessibility plays a secondary role in explaining variation in vacancy rate compared to vacancy duration.

## 5.4. Overall findings

Looking at the results of the Latent Class Cluster analysis as well as the Regression analysis, some final general insights can be derived.

- Latent Class Cluster Analysis shows that Dutch postcodes can be divided into six commuter and workplace profiles, which are strongly related to the dominant work sectors. Accessibility has different but significant impact on these clusters. It explains 8.34% of the variability in the class membership.
- Industrial areas in the Netherlands show similar commuter profiles and are strongly negatively predicated by accessibility. This is not a local port of Rotterdam issue but a broader national mechanism.
- Both methods confirm that Hansen's potential accessibility outperforms Shen's competitive accessibility in predicting commuter profiles and vacancy indicators in Dutch context, suggesting that labour market competition operates primarily through sectoral and skill-based differences rather than through spatial competition alone.
- Vacancy duration can be predicted better by accessibility and other variables than vacancy rate. vacancy rate is more susceptible to noise and outliers, and therefore functions better as a macro- or sector-level indicator rather than a local spatial outcome.
- Accessibility is a significant and strong predictor of vacancy duration ($R^2$ =0.15 as sole predictor). The effect is even stronger when urban context is included as interaction effect ($R^2$ =0.182). In peripheral context, better accessibility reduces the vacancy duration more than in urban context. Additionally, the spatial heterogeneity of the vacancy duration is well captured by this effect.

- Sectoral and education-related variables are not statistically significant in the regression models, suggesting that their influence is largely mediated through spatial context and accessibility, which is in line with the LCCA findings

Together, the results from the Latent Class Cluster Analysis and the regression models provide a coherent picture of how accessibility, sectoral structure and spatial context shape labour market outcomes in the Netherlands. The LCCA reveals persistent, sector-driven commuter profiles that are partially explained by accessibility, while the regression analysis quantifies how accessibility affects vacancy outcomes differently across urban and peripheral contexts. Importantly, the findings indicate that accessibility primarily influences the speed of matching rather than overall vacancy rate, and that its effects are strongest in place-bound and peripheral labour markets. These results highlight the relevance of spatial context in understanding the labour market.

## 5.5. Reflection on the conceptual model

Returning to the conceptual model (section 2.6, figure 2.2) that summarised the relationships identified in the literature, the validity of these pathways can now be assessed based on the empirical results.

The indirect pathway, linking accessibility to recruitment outcomes via commuter profiles, is strongly supported by the LCCA results. This pathway suggests that public transport accessibility acts as a socio-demographic filter, determining the composition of the local labor pool. The sharp contrast between the peripheral industrial clusters (1 and 3) and the dense urban clusters (5 and 6) illustrates this mechanism in detail. Cluster 3, characterised by low PT accessibility(0.15) and low PT usage (3.22%) act as a barrier to the knowledge-oriented and younger demographics found in the city. Instead, Cluster 3 is defined by a car-dependent (65.02%), male-dominated profile (75.22% male). This creates a specific socio-economic profile that relies on wage compensations (61% high income) to overcome spatial barriers. This profile stands in direct opposition to the urban profiles. Clusters 5 and 6 have the highest accessibility scores (0.41 and 0.48 respectively) and successfully attract the high educated, service oriented workforce. This manifests itself in the percentage highly educated personnel in Cluster 6 (70.60%) and the significant presence of students and starters in Cluster 5 (41.82% < 35 years). The high PT usage in these areas (16.88% to 27.51%) confirms that in high density environments, infrastructure facilitates a diverse labour pool that is not reliant on private car ownership. This confirms that the workplace environment acts as a filter, as theorized in the modal mismatch literature (Grengs, 2010).

The direct pathway, which is guided by the assumption that higher accessibility consistently reduces vacancy duration, requires theoretical nuance. While this relationship holds in peripheral industrial zones, the observed 'reversal effect' in dense urban areas suggests that the direct link in Figure 2.2 is mediated by matching frictions that traditional potential accessibility measures like Hansen do not fully capture. The isolated effect of Hansen's potential accessibility shows that this direct pathway is significantly present, but requires some further specification in urban context.

## 5.6. Discussion on the results

There are some considerations that need to be addressed with respect to the general interpretations of the research. Firstly, the Latent Class Cluster Analysis and the regression analysis are based on different spatial scopes. The ODiN data provides personal level data which is aggregated to postcode level in the Netherlands. Requiring a minimum sample size of 30 per postcode for statistical validity restricts the scope of the analysis and introduces a selection bias, disproportionately favouring urban areas with higher employment concentrations. On the other hand, the regression analysis is conducted on the postcodes within the Rijnmond region. There is some overlap in the postcodes, but not enough to integrate the results and derive statistically strong results. However, the two approaches yield complementary insights. The LCCA demonstrates that accessibility is systematically associated with specific commuter and workplace profiles at the national level, while the regression results show how accessibility translates into measurable labour market outcomes at the regional level. The consistency of the accessibility effects across these spatial scales strengthens the interpretation that accessibility is not merely a descriptive characteristic of locations, but a structurally relevant factor shaping labour market functioning.

Interpreting these structural, zonal-level findings requires caution to avoid the ecological fallacy (Firebaugh, 2001). Because the analyses rely on data aggregated to the PC4 level, the observed relationships between accessibility and vacancy duration apply to the postcode as a whole, rather than to any single firm within it. While a postcode with high public transport accessibility may exhibit shorter median vacancy durations on average, individual employers within that zone may still struggle to recruit personnel due to firm-specific characteristics, such as poor reputation, uncompetitive wages, or highly specialized skill requirements. Therefore, the results should be understood as describing spatial labour market conditions and broad matching efficiencies, rather than predicting individual firm recruitment success.

Based on the accessibility and labour market literature, income was expected to play a more prominent role in differentiating commuter profiles and accessibility related outcomes. Previous studies consistently show that low-income individuals are more vulnerable to poor accessibility due to limited car ownership, restricted residential choice and higher sensitivity to commuting costs (Boisjoly et al., 2019; Geurs & van Wee, 2004; Kain, 1968). From this perspective, income is often interpreted as a key socio-economic factor for accessibility equity and labour market participation. However, in the present analysis, income appears to be a relatively weak differentiator between the identified commuter profiles. This deviation from theoretical expectations can be explained by several factors. First, the analysis is conditional on employment: all included individuals are already working and have a fixed workplace. Much of the literature that finds strong income effects focuses on employment probabilities or unemployed, where income constraints are more decisive. Once individuals are employed, income differences appear to play a smaller role in shaping commuting patterns and accessibility sensitivity. Second, income related mechanisms are partly captured indirectly through other variables in the model. Accessibility constraints associated with income often operate via other variables such as car ownership (Grengs, 2010). In this sense, the limited role of income does not contradict the literature, but rather suggests that its effect is mediated through more direct mobility constraints. Finally, income is measured at the household level, which reduces variation at the lower end of the income distribution and limits its ability to distinguish between individual mobility constraints. Previous research has noted that education level and access to transport modes are often more robust predictors of accessibility related inequalities than income alone (Bastiaanssen et al., 2022; Boisjoly et al., 2019). The findings therefore suggest that, within a working population and a place-based accessibility framework, income is a less decisive factor than initially assumed based on the broader accessibility equity literature.

An important and consistent finding across both the Latent Class Cluster Analysis and the regression models is that Hansen's potential accessibility outperforms Shen's competitive accessibility in explaining both commuter profiles and vacancy indicators. At first glance, competitive accessibility sounds promising because it incorporates labour market competition, which should help in predicting labour market outcomes. Literature such as by Geurs and van Wee (2004) and Shen (1998) suggest that potential accessibility greatly overestimates accessibility when competition is not accounted for. However, the literature has not used the competitive measure from the perspective of firm, but rather from the perspective of the individual. Bastiaanssen et al. (2022, 2025) respectively connect potential and competitive accessibility to employment opportunities using individual level micro data. In this context, competition is a relevant mechanism as individuals compete for a limited number of jobs, and accessibility reflects their relative position within that competitive field. In contrast, the present study focuses on firm side outcomes, such as vacancy duration and vacancy rates, where accessibility reflects the size and reach of the potential labour pool rather than individual competition. Moreover, the added complexity in the measure works better on a finer granularity whereas the relatively simple Hansen measure aligns better with the postcode aggregation. In this study, job supply is approximated as the total number of jobs in each postcode and the competing workforce as the total postcode population. This is a strong simplification, which makes the competitive accessibility measure less suitable for explaining specific labour market outcomes. In contrast, the simpler Hansen index appears more robust to the generalisations and aggregation in PC4-level data, as it captures the general potential for interaction. This competition measure might work better when finer sectoral, occupational or skill specific data is available. These findings suggest that while competitive accessibility is well suited to analysing individual employment outcomes at fine spatial scale, its added complexity may be less informative for firm level labour market outcomes in aggregated spatial data.

Another key finding is the difference in predictability of the outcome variables vacancy duration and vacancy rate. This difference can be related to the nature of the variables. Vacancy duration reflects the speed of the matching process between companies and workers, and depends on the amount of applicants for a vacancy. Within the framework by Mortensen and Pissarides (1994) and Petrongolo and Pissarides (2001), the probability that a vacancy is filled depends on the size of the effective search space and the intensity of job search. The applicants depend on who are realistically willing to travel to the location, which seems more intuitively linked to the potential accessibility of the area. The higher the accessibility, the more people can realistically reach that job, thus increasing the number of applicants. This mechanism aligns with earlier findings that accessibility influences commuting behaviour, mode choice and employment opportunities, particularly for groups with constrained mobility (Bastiaanssen et al., 2022, 2025; Geurs & van Wee, 2004). Vacancy rate is determined by the number of vacancies which appeared in a year over the number of jobs in an area. This ratio tells something about local market conditions, but it can mean that there is a strong economic growth or structural shortage in personnel. Especially, aggregated over many companies, the reason for many vacancies remains ambiguous. This interpretation is supported by Rouwendal and Rouwendal (2025), who show that dense urban labour markets tend to exhibit both higher vacancy rates and higher job mobility, reflecting intensified labour market dynamics rather than unfulfilled labour demand. The decision to deploy vacancies is dependent on the company, whereas the time it takes to fill the vacancy is largely dependent on the applicants. Therefore, the link between the duration and accessibility is easier to make. Lastly, the median was taken of the vacancy durations, making the indicator robust to skewed distributions, whereas postcodes with little vacancy appearances introduce bias in the vacancy rate indicator. The decision to only use vacancies of which sector and education level information was known, might have large consequences on the vacancy rate indicator.

Lastly, the observed reversal effect in the interaction also has broader theoretical implications for how accessibility should be interpreted in labour market analyses. While accessibility and density are often conceptualised as an unambiguously positive facilitator of matching (Dossche et al., 2025; Mangan & Trendle, 2017), the results show that its effect depends on which stage of the recruitment process is binding. In areas where where spatial mismatch is prominent, accessibility primarily alleviates search frictions by expanding the applicant pool (Geurs & van Wee, 2004; Kain, 1968). In contrast, in dense urban labour markets, accessibility interacts with agglomeration dynamics, intensified competition and selective hiring behaviour, which can prolong vacancy durations despite a large pool of potential workers (Petrongolo & Pissarides, 2001; Puga, 2010). Consequently, the positive interaction effect serves as empirical evidence that the 'spatial fix' of improving transport infrastructure has inherent limits. While it remains an efficient tool for industrial and peripheral labour markets, its effect diminishes in urban environments where matching complexity takes over. As physical distance ceases to be the primary limiting factor, informational frictions and the administrative burden of sorting through a high volume of applicants become the new constraints on efficiency. This implies that in highly accessible urban areas, vacancy duration is driven more by a company's own hiring expectations than by transport infrastructure.

# 6
# Conclusion & Discussion

This chapter present the answers to the research questions and positions the research in the existing literature. Additionally, implications for policy are discussed along with the limitations of this research. Lastly, some topics for further research are proposed.

## 6.1. Answers to the research questions

Now that the results have been presented, this section answers the research questions formulated in section 1.3. First, the main research question is answered, followed by more elaborate answers to the sub-questions.

**To what extent does public transport workplace accessibility shape commuter profiles, and how does this relationship help explain regional differences in vacancy fulfilment in Rijnmond?**

The results show that public transport workplace accessibility plays a meaningful role in shaping commuter profiles across the Netherlands and contributes to explaining spatial differences in vacancy fulfilment in the Rijnmond region. Accessibility does not operate in isolation, but interacts with urban context and workplace characteristics to structure who is able and willing to work in specific locations. The Latent Class Cluster Analysis reveals 6 distinct clusters (table 5.2), which are all influenced differently by public transport accessibility. Posterior classification shows that the Port of Rotterdam is part of a cluster containing ports, airports and other industrial areas, which is predicted by low accessibility. The results from the spatial autocorrelation reveal that the spatial variation is well captured by accessibility and address density. OLS regression also reveals that Hansen's accessibility can explain 14.8% of the variance of the vacancy duration and interacting with address density, this goes up to 18.2% (table 5.5). In peripheral areas, better accessibility results in a stronger reduction in vacancy duration (figure 5.11). These findings show that accessibility structurally shapes who works in what locations and how hard vacancies are to fill.

*What does existing research reveal about workplace accessibility and the relationship with socio-demographics, modal split and labour market outcomes?*

Existing literature shows that there are many ways to operationalise accessibility, but most literature uses potential accessibility (Hansen, 1959), and more recently competitive accessibility (Shen, 1998). Accessibility is closely intertwined with socio-demographic characteristics and commuting behaviour. High accessibility is associated with high public transport usage. Differences in socio-demographic characteristics, create disparities in accessibility equity, particularly in peripheral areas where the baseline provision of public transport is lower. Especially income, education and car ownership, but also age and household characteristics are influential socio-demographic characteristics. From an individual perspective, higher job accessibility is causally linked to higher employment probability. The limited research into vacancies highlights that location is relevant for the duration of vacancies, especially the difference between urban and peripheral areas. A more profound accessibility metric is suggested to predict vacancy duration even better than location.

*Which socio-demographic commuter profiles can be identified across Dutch postcodes, and how is public transport accessibility associated with these profiles?*

Latent Class Cluster Analysis reveals 6 distinct socio-demographic commuter clusters. High public transport accessibility is strongly associated with Clusters 5 and 6 (high-educated, young, and high-income profiles). In these dense urban cores, the infrastructure facilitates a shift toward public transport and slow modes, effectively matching starters and specialists to high-density job areas. Cluster 3 (Industry/Ports) presents a significant mismatch. While job totals rival urban areas, public transport accessibility is at its lowest. This forces the male-dominated workforce to use the car, regardless of education level. Clusters 1 and 4 (Peripheral and Young Local) face structural barriers: Lower public transport accessibility in these areas restricts commuters to local jobs or forces car ownership, adding to the modal mismatch for those who are less mobile or with lower incomes. Public transport accessibility is not just a service but a dividing factor: It serves as a facilitator for the high-educated, urban workforce but acts as a barrier for industrial and peripheral commuters.

*To what extent do the commuter and public transport accessibility characteristics of the Port of Rotterdam reflect a unique local situation or a broader national pattern?*

The analysis of cluster 3 reveals that the port of Rotterdam is not just a unique local situation, but rather an extreme example of a broader national pattern. While the port of Rotterdam is unique in its size, geographical isolation and reliance on the road network, its socio-demographic and accessibility characteristics reflect a national trend of industrial connectivity gaps. As theorized by the Spatial Mismatch Hypothesis (Kain, 1968), labour market friction occurs when employment hubs shift to peripheral areas while the required workforce remains concentrated in urban centres. While high-educated service sectors (Clusters 5 and 6) benefit from a dense, multimodal urban network, the industry of the region (Cluster 3) is disconnected from the public transport system. Areas with high industrial density seem to always have lower public transport accessibility in all regions in the Netherlands. Consequently, persistent vacancy issues in industrial hubs should be viewed as a structural spatial phenomenon, rather than merely a local shortcoming.

*How does workplace public transport accessibility influence the spatial variation in vacancy duration and vacancy rate?*

Spatial autocorrelation analysis indicates significant positive spatial autocorrelation in vacancy duration, revealing clear clusters of both long and short vacancy durations. However, the residuals of the model display limited to no spatial autocorrelation (table 5.4). Therefore, accessibility succeeds at capturing the spatial effects of the vacancy indicators.

Accessibility determines the efficiency of the labour market. It does not determine whether work is available, but how quickly a match can be found. Table 5.5 shows that vacancy duration is strongly explained by public transport accessibility ($R^2$ = 14.8%), but even stronger when the interaction with address density is taken into account ($R^2$ = 18.2%). In peripheral areas, the mechanism works as expected: An increase in public transport accessibility results in a significant reduction of the vacancy duration, because car and slow modes are often the better or only option. A good public transport connection can make a significant difference and reach a new workforce that were unable to travel here before. In urban areas, higher accessibility actually leads to longer vacancy durations (figure 5.11). The high number of jobs as well as workers leads to higher selectivity amongst employers. Accessibility facilitates the high supply of workers which allows for longer selection procedures.

Contrary to vacancy duration, accessibility hardly has explanatory power for the vacancy rate indicator. vacancy rate is driven by structural shortages in specific sectors (such as technical jobs in the port or healthcare in the city centre). These sectors are clustered in certain locations and the tightness is an indicator of their success. These factors cannot be captured by differences in accessibility.

## 6.2. Implications

The findings of the study indicate that workplace public transport accessibility plays an important role in the labour market matching process, particularly concerning vacancy duration. Accessibility explains a substantial share of the spatial variation in vacancy duration, both directly and through its interaction with address density. This suggests that accessibility functions as a primary condition for efficient labour market matching, influencing how quickly vacancies can be filled rather than whether employment opportunities exist. This shifts part of the explanation away from the individual and firm factors to the organisation of public transport at local and national level. The finding that these areas suffer from the lowest public transport accessibility despite job densities comparable to urban cores indicates a structural modal mismatch. Policy interventions aimed at reducing vacancy durations in the port must consider that current car dependency acts as a barrier for a significant portion of the potential workforce living in the city.

To translate the statistical findings of the interaction model into a practical context, a scenario analysis was conducted for the Maasvlakte (postcode 3199). Figure 6.1 illustrates the predicted effect of improving the area's accessibility. The model estimates that a targeted infrastructure improvement resulting in a 5-minute reduction in total travel time for commuters would decrease the average vacancy duration by approximately one full day ($\Delta y \approx -0.95$). Even though a single day might appear modest, this reduction in vacancy duration can represent a significant economic benefit. This demonstrates that for low-density, remote employment hubs, investments in transport networks remain an effective tool for regional labour market policy.

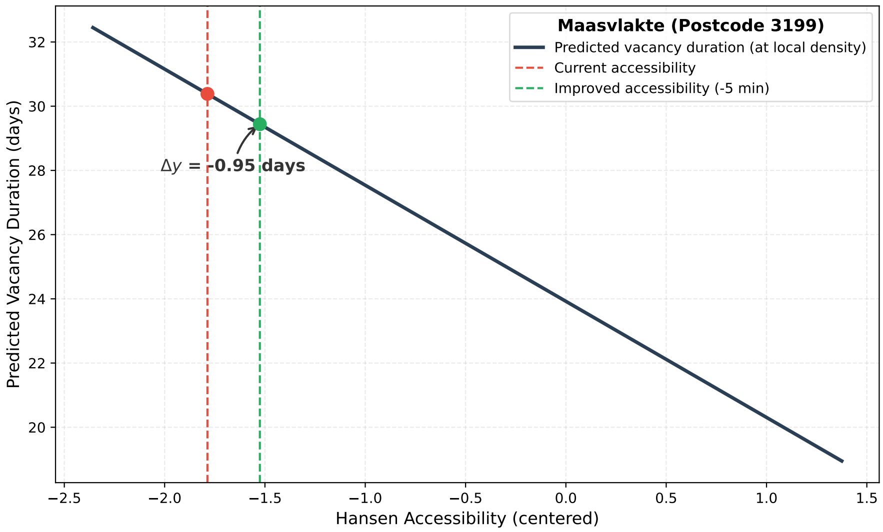



**Figure 6.1:** Predicted impact of a 5-minute travel time reduction on vacancy duration for the Maasvlakte industrial area, based on the local address density.

The simple conclusion would be: Invest in public transport in peripheral industrial areas because the marginal gain is highest in these areas. While this study provides strong arguments for investments in public transport, policymakers must account for low demand density and a culture centred around car commuting. A fundamental tension in transport policy is the debate between efficiency (passenger counts) and accessibility from an equity perspective. In peripheral industrial areas like the Port of Rotterdam (Cluster 3), public transport provision is often viewed through the lens of ridership. Buses which ride without passengers are too expensive and eventually get cut, as there is no rate of return. However, this creates a vicious cycle: poor service forces workers into cars, which justifies further cuts, eventually resulting in an area where car transport is the only viable option. This research suggests that in industrial hubs, public transport should be reframed as an economic facilitator. Even if ridership is low, the existence of a connection reduces the modal mismatch for low-income or young workers, effectively widening the pool of potential applicants for firms struggling with persistent vacancies.

Reframing public transport as an economic facilitator does not, however, imply high-frequency bus lines that largely remain empty. Instead, smart investments should focus on flexibility and the specific needs of an industrial environment. In low-density areas like the Port of Rotterdam, traditional public transport often fails because it cannot compete with the car on speed or convenience. Smart investments prioritise demand-responsive transport (DRT) and first/last-mile solutions that bridge the gap between well-connected hubs and the specific workplace. These investments should be data-driven and made in cooperation with the companies in the area to come to the most efficient solutions.

Another discussion that is relevant is who will benefit from the increased accessibility. As the current workforce in the port especially is male dominated, highly car dependent and relatively old (table B.1, (Dekker & van der Toorn, 2024)), one could argue that improving public transport is futile because the current "Ford Focus driving terminal worker" who has worked in the port for 20 years is unlikely to switch to a bus. This raises the question of self-selection. Is the current workforce car-oriented because the location demands it, or because the current workforce prefers to travel by car. Pot et al. (2023) found that people in peripheral areas generally perceive their accessibility as high and car mobility to be a major contributor to this perception. Contradictory, people who reported low perceived accessibility were more associated with social disadvantages such as being student or unemployed. Therefore, policy should not be designed for the current "Ford Focus driving terminal worker", who has already adapted to spatial constraints of the area. Instead, the focus must shift toward the missing workforce: the urban starters, students, and diverse talent pools identified in Clusters 5 and 6. These groups possess the specialised skills required for the port's energy transition and automation but are currently filtered out by the objective lack of public transport. The link between public transport accessibility and vacancy duration also suggests that the barrier for hiring new people is lowered. Investments in public transport in the port might not result in immediate results, but should be beneficial on the long term.

Given that the focus should be on attracting the future workforce from cluster 5 and 6, the question becomes about how to attract these workers. The spatial mismatch identified in this research can be partly compensated by increasing public transport accessibility and increasing wages. However, there still exists a significant attractiveness gap between urban and peripheral work environments. For the highly-educated urban workforce (Clusters 5 and 6), the decision of where to work is not based solely on salary, but on the total package of the workplace, including its surroundings (social opportunities, safe environment and other quality of life factors) and the ease of the daily commute. According to Glaeser et al. (2001), the success of cities is dependent on the role as centre of consumption, as firms become more mobile. In a tight labour market, these workers have the luxury of choice, and the Port of Rotterdam has difficulties competing as it remains a largely functional environment rather than a social one. This study shows that jobs in the port are often rivalled by urban vacancies that are better accessible from the residential concentrations of the required workforce. This creates a double disadvantage for peripheral employers: they require a longer commute to a less attractive environment. The spatial mismatch is therefore not just a matter of distance, but a utility mismatch. The LCCA results further confirm this utility mismatch through the stark contrast between the industrial cluster and the highly-educated urban workforce clusters. This reality forces a strategic rethink. If the port cannot compete with the liveliness of the city centre, it has two options. First, it can attempt to bring the urban environment to the port area by creating high-quality campus-like environments at transport nodes (such as RDM, the innovation dock in Waalhaven). Second, it can move its knowledge-intensive functions to the city. By relocating engineering and planning offices to areas like M4H (a new development project on the north side of the Maas), firms can bridge the spatial mismatch by placing the work in the exact environment where Clusters 5 and 6 prefer to be.

Additionally, there are also improvements that can be made from the employer side. Even though there are many place bound jobs and shift-workers, it is still important to provide flexibility where possible, such that people can commute outside peak hours and when on-demand transport is available. Furthermore, encouraging other initiatives such as van pooling and carpooling can contribute to better connectivity for those who have a transport disadvantage. Even with improved physical connectivity, the perceived distance to the port remains high for many urban dwellers who view the industrial landscape as a foreign environment. To counter this, employers must invest in recruitment branding that emphasizes the social and innovative nature of the modern industrial workplace (Dekker & van der Toorn, 2024).

Finally, a critical nuance for planners is the finding that in dense urban areas, higher accessibility actually correlates with longer vacancy durations. This implies that in the city, the binding constraint is not accessibility. In high density urban zones, further investment in physical public transport infrastructure may yield diminishing returns for labour market speed. Instead, resources should be shifted toward digital labour market matching and employer information services. The goal here is to help employers navigate the excessive number of applicants that high accessibility provides, reducing the selectivity friction that currently slows down the hiring process in cities. This does not mean that new infrastructure investments are not necessary in urban cores, but it means it does not have benefits for job matching speed.

## 6.3. Scientific contribution

This research provides a significant scientific contribution to the fields of transport geography and urban economics by bridging the gap between spatial connectivity and labour market outcomes. The following text synthesizes these contributions and situates them within the existing body of literature.

In this research, the focus is shifted from the employment probabilities of the individual to the employers perspective. While the majority of existing literature examines how job accessibility influences an individual's probability of gaining employment (Bastiaanssen et al., 2022, 2025), this study investigates how workplace accessibility shapes vacancy duration and vacancy rate. By doing so, it addresses a critical research gap identified in recent predictive modelling of "hard-to-fill" vacancies (Dossche et al., 2025) and provides empirical evidence to earlier findings that suggested a firm's location is a primary determinant of recruitment success (Andrews et al., 2008; Mangan & Trendle, 2017).

Secondly, this study makes a methodological contribution by integrating socio-demographic heterogeneity into spatial accessibility assessments through Latent Class Cluster Analysis (LCCA). By identifying distinct commuter profiles, the research operationalises the concept that accessibility is not equal for everyone. The differences within clusters reveal how transport disadvantage and transport poverty (Nakshi et al., 2025) disproportionately affect specific groups, such as young local workers or industrial commuters, which aids the understanding of distributive justice in transportation (Pereira et al., 2016).

Thirdly, the research provides a modern empirical case of the Spatial Mismatch Hypothesis (Kain, 1968) within a highly developed European context. By highlighting the industrial connectivity gap in areas like the Port of Rotterdam, the study demonstrates a persistent modal mismatch (Grengs, 2010). It shows that even in regions with high job density, the absence of basic public transport provision creates a structural barrier for people without car. This finding extends the theoretical work on transport-related social exclusion (Lucas, 2012) and illustrates how spatial accessibility differences can lead to persistent labour shortages despite high regional accessibility.

Finally, the study refines the understanding of agglomeration economies (Puga, 2010) by exploring the interaction between urban density and accessibility. This research directly addresses the call by Geurs and van Wee (2004) for a more rigorous analysis of job accessibility and competition levels across varying spatial contexts, specifically high-density metropolitan areas versus low-density peripheral zones.

## 6.4. Limitations

There are some limitations to this research, as a result of the data (processing), the used methods, and scoping choices that are important to be aware of. This section highlights these limitations and their effects on the outcomes of the analysis.

- A limitation of this study is that the direct relationship between the commuter profiles (from the LCCA) and the vacancy outcomes (from the regression) was not statistically tested. Due to data constraints, it was not possible to integrate the cluster results as variables directly into the regression models for vacancy duration. This means that the link in the conceptual model between these two components is based on a logical synthesis of results rather than a direct statistical calculation. While the findings suggest that specific profiles, such as the car-dependent cluster 3, correspond with slower hiring processes, the exact causal impact of a commuter profile remains unknown.

- A limitation of this study is the difficulty in isolating the causal effect of public transport accessibility from broader agglomeration economies (Puga, 2010). Highly accessible areas often benefit from unobserved local amenities, economic clustering, and general urban advantages. While the models control for address density, fully isolating the specific benefits of transport connectivity from these spatial advantages remains challenging, especially in urban context. Consequently, the estimated accessibility effects may partly suffer from omitted variable bias, capturing a broader spatial agglomeration effect rather than the direct impact of the transport network alone.
- The use of PC4-level data introduces the Modifiable Areal Unit Problem (MAUP), as postcode boundaries may not align with functional local labour markets and the access points to the public transport network (Qu et al., 2025). Postcodes often contain mixed land uses, meaning that average accessibility may not reflect the actual accessibility of specific workplaces. This spatial aggregation can introduce boundary noise, particularly for competitive accessibility measures (Shen, 1998). As a result, estimated effects and spatial patterns may differ under alternative spatial scales or boundary definitions. Moreover, the reliance on spatially aggregated data introduces the risk of the ecological fallacy (Firebaugh, 2001). The analysis identifies structural relationships between public transport accessibility and vacancy outcomes at the PC4 level. However, drawing conclusions about individual firms or job seekers based on these aggregate spatial patterns must be done with caution. While a highly accessible postcode may exhibit shorter median vacancy durations, individual hiring outcomes within that zone will still vary significantly based on unobserved firm-specific characteristics and individual worker preferences that are obscured by spatial aggregation.
- This research utilised static accessibility measures, which do not account for daily fluctuations in transit frequency or job availability. Although Boisjoly and El-Geneidy (2016) suggest that static measures are often sufficient for general planning purposes, they may overlook the specific constraints faced by shift workers in industrial clusters like the Port of Rotterdam. For these commuters, the temporal availability of public transport is as critical as its spatial presence. Especially in peripheral environments (where buses ride once every 30 minutes or even hourly). In such contexts, schedule misalignment can have substantial consequences, as commuters tend to plan their travel carefully around limited service availability. Additionally, the first- and last-mile transport are modelled as walking, which is not always realistic and can create extreme results for some coordinates which are further from the network.
- The vacancy data used in this study introduces several limitations that primarily affect the interpretation of vacancy rate. Vacancy duration is measured with an uncertainty of up to seven days, which introduces measurement noise, particularly for the vacancies of short duration. This uncertainty is partially mitigated by using median vacancy durations and restricting the analysis to statistically significant postcodes. In addition, only vacancies containing information on sector and education level are included, resulting in a selective sample that may underrepresent certain types of jobs or employers. This primarily affects the vacancy rate as there is no knowing where these vacancies would have been and how they are divided over the postcodes. Furthermore, the use of aggregated micro data obscures variation at the individual vacancy level. Together, these data limitations reduce the reliability and spatial sensitivity of the vacancy rate indicator.
- The empirical analysis relies on the ODiN (Onderweg in Nederland) dataset, which possesses inherent limitations common to surveys (Centraal Bureau voor de Statistiek (CBS), 2023b). Firstly, the data is self-reported, which may introduce reporting biases or inaccuracies regarding travel times and distances as respondents often round figures or suffer from memory recall bias. Secondly, despite the substantial national sample size, the spatial density of respondents at the PC4 level varies, which results in a reduction to roughly 20% statistically significant postcodes at the PC4 level. Lastly, observed travel patterns are based on a single randomly selected day per respondent, which may not be representative of typical behaviour. Even though people were able to flag unusual behaviour, one cannot guarantee that this self-reporting mechanism was used consistently. While this introduces additional uncertainty, the direction and magnitude of the resulting bias cannot be determined.
- Finally, the inclusion of data from the 2020–2022 pandemic period introduces potential behavioural anomalies, as commuting patterns shifted significantly during this time (KiM Netherlands Institute

for Transport Policy Analysis, 2022). While the methodological choice to treat travel behaviour as a passive covariate in the LCCA mitigated the impact on the structural definition of the clusters, the behavioural snapshots may still introduce some bias. A detailed justification and sensitivity analysis of this period are provided in Appendix E.

## 6.5. Recommendations for Further Research

Based on the limitations and findings of this study, several directions for further research are suggested to improve understanding of the relationship between public transport accessibility and labour market outcomes.

- Future research could benefit from the use of longitudinal data or quasi-experimental designs to better identify causal effects between accessibility and labour market outcomes. While this study identifies robust spatial associations, cross-sectional analyses might suffer from endogeneity due to residential and workplace self-selection. Investigating the influence of historical changes in the public transport network, such as the opening of a new rail station or bus corridors, would allow researchers to assess how accessibility improvements affect commuter profiles and vacancy fulfilment over time. This would help to isolate the unique effect of accessibility on labour market outcomes.
- Analyses using smaller spatial units, such as postcode 6 or even company specific accessibility and vacancies, could address Modifiable Areal Unit Problem (MAUP) effects inherent in postcode 4 aggregation. Industrial areas in particular often consist of highly clustered firms located at the edge of administrative units, where small changes in location or infrastructure can substantially affect accessibility. Using smaller spatial units would allow future research to capture local accessibility variation more precisely and assess whether the observed relationships strengthen or weaken at higher spatial resolution.
- This study relies on static public transport accessibility measures, which may not fully capture the commuting constraints faced by workers in industrial and peripheral locations. Incorporating space-time accessibility (Hu & Downs, 2019), with more flexible starting time, first-mile cycling and shift timings, would better capture the actual commuting constraints faced by workers in industrial clusters. Additionally, it would be interesting to incorporate car accessibility (including congestion) in the analyses. Given the strong car dependence observed in several commuter profiles, car accessibility likely plays an important moderating role.
- The vacancy micro data used in this study contains detailed information on contract type, education requirements, sector, skill demands, and working hours, but incomplete coverage limits its current use in multivariate (spatial) analysis. Future research could leverage larger or more complete datasets to investigate how accessibility interacts with specific job characteristics. This would help explain not only where vacancies remain open longer, but which types of jobs are most sensitive to accessibility constraints, strengthening the socio-economic interpretation of accessibility effects.
- The interaction results suggest that the relationship between public transport accessibility and vacancy fulfilment differs fundamentally between peripheral and dense urban contexts. While higher accessibility reduces vacancy duration in less urbanised areas, this effect weakens and even reverses in highly urbanised environments. The linear interaction term assumes a gradual transition. Future research should investigate this specific turning point by employing non-linear modelling techniques to identify the exact level of accessibility or density where the relationship flips from negative (reducing duration) to positive (increasing duration) and the driving mechanisms behind this turning point. This would answer a critical planning question: "How much accessibility is enough?" enabling policymakers to target investments only up to the point of diminishing returns, preventing over-investment in areas where accessibility is no longer the binding constraint.
- An important next step would be to explicitly integrate the results of the Latent Class Cluster Analysis into the regression framework. By interacting accessibility measures with commuter profile membership, future research could test whether different socio-demographic groups structurally experience different labour market outcomes. This would allow for a more direct assessment of group-specific effects of accessibility.

- Finally, examining other regions or countries with different public transport systems, urban forms, and labour market institutions would help assess the external validity of the findings. Comparative analyses could reveal whether the observed dominance of potential accessibility over competitive measures is specific to the Dutch context or reflects a more general mechanism in accessibility–labour market interactions.

# References


Andrews, M., Bradley, S., Stott, D., & Upward, R. (2008). Successful employer search? an empirical analysis of vacancy duration using micro data. *Economica*, *75*(299), 455–480. https://doi.org/10.1111/j.1468-0335.2007.00594.x

Bastiaanssen, J., Johnson, D., & Lucas, K. (2025). The relationship between individual employment probabilities and accessibility to matching jobs: A study of the netherlands. *Transportation Research Part A: Policy and Practice*, *193*, 104398. https://doi.org/10.1016/j.tra.2025.104398

Bastiaanssen, J., Johnson, D., & Lucas, K. (2022). Does better job accessibility help people gain employment? the role of public transport in great britain. *Urban Studies*, *59*(13), 2602–2621. https://doi.org/10.1177/00420980211012635

BIJ12. (2024). Lisa employment register [Accessed 2024].

Birinci, S., Mercan, Y., & See, K. (2025). Mismatch unemployment during covid-19 and the post-pandemic labor shortages. *Journal of Economic Dynamics and Control*, *178*, 105142. https://doi.org/https://doi.org/10.1016/j.jedc.2025.105142

Boisjoly, G., & El-Geneidy, A. M. (2016). Daily fluctuations in transit and job availability: A comparative assessment of time-sensitive accessibility measures. *Journal of Transport Geography*, *52*, 73–81. https://doi.org/https://doi.org/10.1016/j.jtrangeo.2016.03.004

Boisjoly, G., & El-Geneidy, A. M. (2017). How to get there? a critical assessment of accessibility objectives and indicators in metropolitan transportation plans. *Transport Policy*, *55*, 38–50. https://doi.org/10.1016/j.tranpol.2016.12.011

Boisjoly, G., El-Geneidy, A. M., & Levinson, D. (2019). Accessibility and the journey to work through the lens of equity. *Journal of Transport Geography*, *74*, 269–277. https://doi.org/10.1016/j.jtrangeo.2018.12.003

Brambor, T., Clark, W. R., & Golder, M. (2006). Understanding interaction models: Improving empirical analyses. *Political Analysis*, *14*(1), 63–82. https://doi.org/10.1093/pan/mpi014

Centraal Bureau voor de Statistiek. (2014). Beroepenclassificatie 2014 (brc 2014) [Standard Classification of Occupations used in the Netherlands]. https://www.cbs.nl/nl-nl/onze-diensten/methoden/classificaties/beroepen/brc-2014

Centraal Bureau voor de Statistiek. (2024). *De arbeidsmarkt in cijfers 2023* (Jaarcijfers 2023, gepubliceerd april 2024). Centraal Bureau voor de Statistiek (CBS). Den Haag. https://www.cbs.nl/nl-nl/publicatie/2024/18/de-arbeidsmarkt-in-cijfers-2023

Centraal Bureau voor de Statistiek (CBS). (2023a). Duurzame mobiliteit [Accessed: Wednesday 18th February, 2026].

Centraal Bureau voor de Statistiek (CBS). (2023b). Onderweg in nederland (odin): Dutch national travel survey [Accessed: 2025-09-29]. https://www.cbs.nl/en-gb/our-services/methods/surveys/brief-survey-description/dutch-national-travel-survey

Centraal Bureau voor de Statistiek (CBS). (2024a). *85266ned — opleidingsniveau; kerncijfers naar leeftijd en geslacht, 2023*. Retrieved May 15, 2024, from https://opendata.cbs.nl/#/CBS/nl/dataset/85266NED/table

Centraal Bureau voor de Statistiek (CBS). (2024b). Emancipatiemonitor 2024 — werken [Accessed: YYYY-MM-DD].

Centraal Bureau voor de Statistiek (CBS). (2024c). Verwerkingsproces odin 2023 [Geraadpleegd op Wednesday 18th February, 2026]. https://www.cbs.nl/nl-nl/longread/rapportages/2024/onderweg-in-nederland--odin---2023-onderzoeksbeschrijving/6-verwerkingsproces-odin-2023

Centraal Bureau voor de Statistiek (CBS). (2025a). Cbs open data [Accessed: Wednesday 18th February, 2026].

Centraal Bureau voor de Statistiek (CBS). (2025b). Statistische gegevens per vierkant en postcode 2022-2023-2024: Beschrijving cijfers [Geraadpleegd op Wednesday 18th February, 2026]. https://www.cbs.nl/nl-nl/longread/diversen/2025/statistische-gegevens-per-vierkant-en-postcode-2022-2023-2024/4-beschrijving-cijfers

Cheng, Y., Yan, X., & Gao, S. (2020). Exploring spatial heterogeneity in accessibility and transit mode choice. *Journal of Transport and Land Use*, *13*(1), 1–17. https://doi.org/https://doi.org/10.1016/j.trd.2020.102521

Dekker, F., & van der Toorn, A.-J. (2024, November). *Arbeidsmarktonderzoek hic 2024 perspectief in het rotterdamse havengebied* (tech. rep.). SEOR BV. Rotterdam.

Dossche, W., Vansteenkiste, S., Baesens, B., & Lemahieu, W. (2025). Anticipating delays in recruitment: Explainable machine learning for the prediction of hard-to-fill online job vacancies [Published online ahead of print]. *European Journal of Operational Research*. https://doi.org/https://doi.org/10.1016/j.ejor.2025.06.027

Feller, W. (1968). *An introduction to probability theory and its applications, volume i* (3rd). John Wiley & Sons.

Firebaugh, G. (2001). Ecological fallacy, statistics of. In *International encyclopedia of the social & behavioral sciences* (pp. 4023–4026). Pergamon. https://doi.org/10.1016/B0-08-043076-7/00713-3

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

Glaeser, E. L., Kolko, J., & Saiz, A. (2001). Consumer city. *Journal of Economic Geography*, *1*(1), 27–50. https://doi.org/10.1093/jeg/1.1.27

Grengs, J. (2010). Job accessibility and the modal mismatch in detroit. *Journal of Transport Geography*, *18*(1), 42–54. https://doi.org/10.1016/j.jtrangeo.2009.01.012

Hagenaars, J. A., & McCutcheon, A. L. (2002). *Applied latent class analysis*. Cambridge University Press.

Hamersma, M., & Roeleven, I. (2024, September). *Acceptabele bereikbaarheid: Een reizigersperspectief* (tech. rep.) (Geraadpleegd op 10 september 2025). Kennisinstituut voor Mobiliteitsbeleid (KiM). https://www.kimnet.nl/documenten/2024/09/26/acceptabele-bereikbaarheid-een-reizigersperspectief

Handy, S. (2020). Is accessibility an idea whose time has finally come? *Transportation Research Part D: Transport and Environment*, *83*, 102319. https://doi.org/10.1016/j.trd.2020.102319

Hansen, W. G. (1959). How accessibility shapes land use. *Journal of the American Institute of Planners*, *25*(2), 73–76. https://doi.org/10.1080/01944365908978307

Hu, Y., & Downs, J. (2019). Measuring and visualizing place-based space-time job accessibility. *Journal of Transport Geography*, *74*, 278–288. https://doi.org/https://doi.org/10.1016/j.jtrangeo.2018.12.002

Ingram, D. R. (1971). The concept of accessibility: A search for an operational form. *Regional Studies*, *5*(2), 101–107. https://doi.org/10.1080/09595237100185131

Jobdigger. (2025). Dutch vacancy market data 2023–2025 [Data retrieved for the purpose of this research]. https://www.jobdigger.nl/

Joost Boudewijns. (2024, April). *Visie openbaar vervoer, carpoolen en collectief vervoer*. TwynstraGudde. Den Haag.

Kain, J. F. (1968). Housing segregation, negro employment, and metropolitan decentralization. *The Quarterly Journal of Economics*, *82*(2), 175–197.

Kim, J. H. (2019). Multicollinearity and misleading statistical results. *Korean Journal of Anesthesiology*, *72*(6), 558–569. https://doi.org/10.4097/kja.19087

KiM Netherlands Institute for Transport Policy Analysis. (2022, October). *Has covid led to a structural change in travel behaviour?* (Tech. rep.) (Accessed: 2025-09-30). Ministry of Infrastructure and Water Management. https://english.kimnet.nl/documents/2022/10/03/has-covid-led-to-a-structural-change-in-travel-behaviour

Levinson, D. M. (1998). Accessibility and the journey to work. *Journal of Transport Geography*, *6*, 11–21. https://doi.org/10.1016/s0966-6923(97)00036-7

Lucas, K. (2012). Transport and social exclusion: Where are we now? *Transport Policy*, *20*, 105–113. https://doi.org/10.1016/j.tranpol.2012.01.013

Mangan, J., & Trendle, B. (2017). Hard-to-fill vacancies: An analysis of demand side responses in the australian state of queensland. *Economic Analysis and Policy*, *54*, 49–56. https://doi.org/10.1016/j.eap.2016.10.002

Meester, R., Ulak, M. B., & Geurs, K. T. (2024). A framework to include socio-demographic characteristics in potential job accessibility levels in low-car and car-free development areas in the netherlands. *Journal of Transport and Land Use*, *17*(1), 511–536. https://doi.org/10.5198/jtlu.2024.2275

Morris, J., Dumble, P., & Wigan, M. (1979). Accessibility indicators for transport planning. *Transportation Research Part A: General*, *13*(2), 91–109. https://doi.org/10.1016/0191-2607(79)90012-8

Mortensen, D. T., & Pissarides, C. A. (1994). Job creation and job destruction in the theory of unemployment. *The Review of Economic Studies*, *61*(3), 397–415.

Næss, P., & Lyssand Sandberg, S. (1996). Workplace location, modal split and energy use for commuting trips. *Urban Studies*, *33*(3), 557–580. https://doi.org/10.1080/00420989650011915

Nakshi, P., Lucas, K., & Farber, S. (2025). Travelling in the margins: Impacts of transport poverty on travel attitudes and behaviours. *Transportation Research Part A: Policy and Practice*, *196*, 104475. https://doi.org/10.1016/j.tra.2025.104475

OpenTripPlanner Contributors. (2025). Opentripplanner [Accessed: 2025-10-20].

Organisation for Economic Co-operation and Development (OECD). (2023). *Policy options for labour market challenges in amsterdam and other dutch cities: Oecd reviews on local job creation* (OECD Reviews on Local Job Creation, Paris: OECD Publishing). OECD Publishing. Paris. https://doi.org/10.1787/1496859a-en

Owen, A., & Levinson, D. (2015). Modeling the commute mode share of transit using continuous accessibility to jobs. *Transportation Research Part A: Policy and Practice*, *74*, 110–122. https://doi.org/https://doi.org/10.1016/j.tra.2015.02.002

Patterson, Z., Bensmihen, A., & Hermanson, G. (2024). The multimodal accessibility target (mat). *Transportation Research Record*, *2678*(1), 770–780. https://doi.org/10.1177/03611981231172503

Pereira, R. H. M., Schwanen, T., & Banister, D. (2016). Distributive justice and equity in transportation. *Transport Reviews*, *37*(2), 170–191. https://doi.org/10.1080/01441647.2016.1257660

Petrongolo, B., & Pissarides, C. A. (2001). Looking into the black box: A survey of the matching function. *Journal of Economic Literature*, *39*(2), 390–431.

Pfertner, M., Büttner, B., & Wulfhorst, G. (2023). An open-source modelling methodology for multimodal and intermodal accessibility analysis of workplace locations. *Sustainability*, *15*(3), 1947. https://doi.org/10.3390/su15031947

Pot, F., Koster, S., & Tillema, T. (2023). Perceived accessibility in dutch rural areas: Bridging the gap with accessibility based on spatial data. *Transport Policy*, *138*, 170–184. https://doi.org/10.1016/j.tranpol.2023.04.014

Pot, F., van Wee, B., & Tillema, T. (2021). Perceived accessibility: What it is and why it differs from calculated accessibility measures based on spatial data. *Journal of Transport Geography*, *94*, 103090. https://doi.org/10.1016/j.jtrangeo.2021.103090

Puga, D. (2010). The magnitude and causes of agglomeration economies. *Journal of Regional Science*, *50*(1), 203–219. https://doi.org/10.1111/j.1467-9787.2009.00657.x

Qu, Y., Currie, G., & Webb, G. I. (2025). Spatial resolution impact on public transport accessibility measurement error. *Journal of Transport Geography*. https://doi.org/10.1016/j.jtrangeo.2025.104408

Rey, S. J., Arribas-Bel, D., & Wolf, L. J. (2023). *Geographic data science with python*. CRC Press.

Rouwendal, H.-J., & Rouwendal, J. (2025). Are urban labour markets more dynamic? vacancies and urban scaling. *Urban Studies*, *62*(7), 1293–1315. https://doi.org/10.1177/00420980241280887

Ryan, J., Vitrano, C., & Martens, K. (2025). Examining the links between public transport provision, suppressed travel and (in)sufficient accessibility. *Transportation Research Part D: Transport and Environment*, *146*(0), 104883. https://doi.org/10.1016/j.trd.2025.104883

Santos, G., Maoh, H., Potoglou, D., & von Brunn, T. (2013). Factors influencing modal split of commuting journeys in medium-size european cities. *Journal of Transport Geography*, *30*, 127–137. https://doi.org/https://doi.org/10.1016/j.jtrangeo.2013.04.005

Sharifiasl, S., Kharel, S., Pan, Q., & Li, J. (2024). Assessing the impact of transit accessibility on employment density: A spatial analysis of gravity-based accessibility incorporating job matching, transit service types, and first/last mile modes. *Journal of Transport Geography*, *121*, 104053. https://doi.org/10.1016/j.jtrangeo.2024.104053

Shen, Q. (1998). Location characteristics of inner-city neighborhoods and employment accessibility of low-wage workers. *Environment and Planning B: Planning and Design*, *25*(1), 1–17. https://doi.org/10.1068/b2501

Statistical Innovations Inc. (2016). *Latent gold 5.1* [Computer software]. https://www.statisticalinnovations.com

Susilo, Y., & Maat, K. (2007). The influence of built environment to the trends in commuting journeys in the netherlands. *Transportation*, *34*, 589–609. https://doi.org/10.1007/s11116-007-9129-5

van Lent, S. J., Ulak, M. B., Wismans, L. J., & Geurs, K. T. (2025). The impact of teleworking in hybrid job accessibility in the netherlands: A novel agent-based modelling framework. *Transportation Research Part A: Policy and Practice*, *196*, 104496. https://doi.org/10.1016/j.tra.2025.104496

Van Ours, J. C., & Ridder, G. (1993). Vacancy durations: Search or selection? *Oxford Bulletin of Economics and Statistics*, *55*(2), 187–198.

Van Wee, B. (2011). Evaluating the impact of land use on travel behaviour: The environment versus accessibility. *Journal of Transport Geography*, *19*(6), 1530–1533. https://doi.org/10.1016/j.jtrangeo.2011.05.011

Van Wee, B. (2016). Accessible accessibility research challenges. *Journal of Transport Geography*, *51*, 9–16. https://doi.org/10.1016/j.jtrangeo.2015.10.018

Verma, R., & Ukkusuri, S. V. (2025). What determines travel time and distance decay in spatial interaction and accessibility? *Journal of Transport Geography*, *122*, 104061. https://doi.org/10.1016/j.jtrangeo.2024.104061

Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 89–106). Cambridge University Press.

Vermunt, J., & Magidson, J. (2005). Technical guide for latent gold 4.0: Basic and advanced.

Wegener, M., & Fürst, F. (1999). *Land-use transport interaction: State of the art* (tech. rep. No. 47). Institut für Raumplanung, Universität Dortmund. https://doi.org/10.17877/DE290R-240

Zondag, B., de Bok, M., Geurs, K. T., & Molenwijk, E. (2014). Accessibility modeling and evaluation: The tigris xl land-use and transport interaction model for the netherlands. *Computers, Environment and Urban Systems*, *41*, 67–76. https://doi.org/10.1016/j.compenvurbsys.2014.06.001

# A

# Research Paper: Public Transport Accessibility and Labour Market Matching: Evidence from the Netherlands

# Public Transport Accessibility and Labour Market Matching: Evidence from the Netherlands

M.J.Kop[1, 2, 3]
[1]*Delft University of Technology, Faculty of Civil Engineering and Geosciences, The Netherlands*
[2]*Delft University of Technology, Faculty of Technology, Policy and Management, The Netherlands*
[3]*Port of Rotterdam, The Netherlands*

*February, 2026*

**While the Spatial Mismatch Hypothesis has extensively documented how poor accessibility reduces employment probabilities for workers, its impact on the employer's ability to fill vacancies remains understudied. This paper investigates the relationship between public transport workplace accessibility and labour market matching efficiency, operationalised as median vacancy duration. Focusing on the Rijnmond region in the Netherlands, characterized by a stark contrast between the dense Rotterdam city centre and the Port of Rotterdam, we utilise web-scraped vacancy data and computed travel times to estimate stepwise OLS regression models. The results demonstrate that potential (Hansen) accessibility outperforms competitive (Shen) measures in explaining firm-side outcomes at the regional level. Crucially, we identify a significant interaction effect between accessibility and urban density ($R^2 = 0.182$). In peripheral industrial areas, accessibility functions as a binding constraint as poor public transport connectivity creates a modal mismatch that significantly prolongs vacancy durations. Conversely, in dense urban cores, higher accessibility is associated with longer durations. We argue that this reversal effect reflects a shift in matching friction: while accessibility alleviates search constraints in the periphery by expanding the applicant pool, it induces screening congestion and selective hiring behaviour in the city. These findings suggest that public transport investments are a critical economic facilitator for industrial labour markets, whereas urban policy should focus on reducing informational frictions.**

***Keywords*** Job accessibility, Labour Market Matching, vacancies, Regression, Public transport, Modal mismatch, Spatial mismatch

## I. Introduction

Workplace accessibility is a core principle in transport geography and urban planning, representing the ease with which employment opportunities can be reached via a given transport system (Hansen, 1959; Geurs and Van Wee, 2004). While traditionally viewed through the lens of equity (ensuring individuals can access jobs) accessibility is increasingly recognised as a critical factor for regional economic competitiveness (Handy, 2020). In an era characterised by persistent labour shortages and a demographic transition, the ability of firms to attract suitable personnel has become a strategic necessity. This challenge is particularly relevant in industrial clusters undergoing a transition toward sustainable and technologically advanced economies, where the demand for specialised labour is higher than the local supply.

Despite the growing economic importance of recruitment efficiency, the role of transport infrastructure in shaping employer outcomes remains understudied. The dominant theoretical framework, the Spatial Mismatch Hypothesis (SMH), states that the physical separation between job growth and residential locations creates structural unemployment (Kain, 1968). However, this literature has predominantly focused on the consequences for the worker, quantifying reduced employment probabilities and increased transport poverty (Bastiaanssen et al., 2022). Much less is known about the employer perspective. If workers face barriers to reaching specific locations, firms in those areas should theoretically experience greater friction in the matching process. Yet, empirical evidence linking transport accessibility directly to recruitment metrics, such as vacancy duration, remains scarce.

This gap is particularly problematic for port cities and industrial regions. These areas are often characterized by a distinct spatial logic: they are located on the periphery of urban centres, designed for the efficient movement of goods rather than people. As these industries modernize and shift from manual labour to knowledge-intensive roles, they increasingly compete for a workforce that resides in dense urban cores and exhibits different mobility preferences (Susilo and Maat, 2007). This creates a potential modal mismatch: while the industry relies on car-dependent infrastructure, the desired urban workforce increasingly prioritizes public transport and active mobility (Grengs, 2010).

The Rijnmond region offers a compelling case to investigate these dynamics. As home to Europe's largest industrial hub, the Port of Rotterdam is currently undergoing a dual transition toward sustainable energy systems and digitalized logistics (Port of Rotterdam, 2025). This transformation demands a diverse influx of new talent, ranging from IT specialists to process engineers. However, this economic ambition is hindered by a sharp spatial contrast. While the nearby Rotterdam city centre is dense and transit-oriented, the port area remains a car-dependent monoculture. Consequently, 88% of employers identify limited accessibility as a major recruitment constraint (Dekker and van der Toorn, 2024). Conventional public transport covers only a fraction of the vast industrial landscape, and the significant distances between residential areas and terminals often render cycling unviable. Despite the presence of some last-mile solutions, the physical isolation of the workplace suggests that labour shortages in the port reflect a spatial and modal mismatch rather than an absolute lack of labour supply.

This context raises a fundamental question: To what extent is the observed labour market friction a result of spatial isolation? Or, more specifically: *'To what extent does workplace public transport accessibility explain the spatial variation in vacancy duration?'*

In this study, we answer this question by integrating granular transport data with web-scraped vacancy records in a spatial regression framework. We contribute to the literature in three ways. First, we shift the analytical lens of the Spatial Mismatch Hypothesis from the employee to the employer. Second, we operationalise matching efficiency using vacancy duration, a direct measure of recruitment friction. Third, we empirically test whether the impact of accessibility differs between the industrial periphery and the urban core, providing new insights into the heterogeneous effects of transport infrastructure on local labour markets.

The paper proceeds as follows. First, section II reviews the theoretical relationship between accessibility and labour market friction. Building on this foundation, section III introduces the methodological approach and regression strategy. The study area is presented in section IV, after which the details of the data sources and processing are presented (section V). Next, section VI presents the regression results, highlighting the significant interaction effect between accessibility and urban density. The paper concludes in section VII by discussing the theoretical implications and offering policy recommendations to resolve spatial mismatches in transitioning industrial regions.

## II. Literature review

Accessibility is a foundational concept in transport geography, commonly defined as the extent to which land-use and transport systems enable individuals to reach desired activities or destinations (Hansen, 1959; Geurs and Van Wee, 2004; Shi et al., 2020). In the context of labour markets, accessibility reflects not merely physical proximity but the effective reachability of workplaces given transport costs, modal availability, temporal constraints, and individual capabilities (Pfertner et al., 2023; Hu and Downs, 2019). From an employer perspective, accessibility shapes the size and composition of the potential labour pool that can realistically reach a workplace, thereby influencing recruitment processes and outcomes such as vacancy duration.

Early accessibility research conceptualised job access through potential accessibility measures, most notably Hansen's gravity-based formulation, which weights employment opportunities by travel impedance. This approach has remained influential due to its intuitive interpretation and suitability for spatial comparison (Hansen, 1959; Ingram, 1971). Subsequent refinements emphasised that job accessibility is not only a function of travel cost and opportunity distribution but also of competition among workers for the same jobs (Shen, 1998; Geurs and

Ritsema van Eck, 2001). By incorporating labour supply into accessibility calculations, competition-based measures acknowledge that high accessibility does not automatically translate into favourable labour market outcomes when many workers compete for limited positions. Despite methodological advances, most accessibility research has been conducted from a residential perspective, focusing on how easily workers can reach jobs. Workplace accessibility (the extent to which firms can be reached by potential employees) has received considerably less attention (Pfertner et al., 2023; Hu and Downs, 2019). When researching labour market outcomes, this different direction becomes particularly interesting.

The relevance of these accessibility measures is best understood through the Spatial Mismatch Hypothesis (Kain, 1968). Ideally, labour markets function as unified pools, but spatial friction creates segmented markets. In the Dutch context, Bastiaanssen et al. (2022, 2025) have extensively demonstrated this mechanism. They show that transport poverty and low job accessibility significantly reduce employment probabilities, particularly for vulnerable groups. This implies that accessibility is not just a theoretical construct, but a decisive factor in labour market participation. However, the demand-side counterpart of this phenomenon remains underexplored. If workers face barriers to reaching specific locations, firms in those areas should theoretically experience greater friction in the matching process. Recruitment outcomes, such as vacancy duration, reflect this efficiency of matching. While existing studies suggest that firms in well-connected areas fill vacancies faster (Andrews et al., 2008; Mangan and Trendle, 2017), current literature often relies on coarse accessibility proxies, such as simple geographic coordinates or regional dummies (Dossche et al., 2025; Birinci et al., 2025). Consequently, it remains unclear whether transport accessibility affects vacancy duration primarily by expanding the effective labour pool (Hansen), or whether competition effects (Shen) play a decisive role.

Furthermore, the impact of accessibility is likely spatially heterogeneous. Land-use–transport interaction theory suggests that the marginal utility of accessibility improvements depends on the existing spatial context (Wegener and Fürst, 1999; Geurs and Van Wee, 2004). Dense urban areas with extensive public transport networks typically exhibit high baseline accessibility. In these environments, further improvements may yield diminishing returns due to saturation or congestion. In contrast, peripheral or mono-functional industrial areas often face binding accessibility constraints (Qu et al., 2025). Empirical studies confirm that the behavioural relevance of accessibility is highly context-dependent. Susilo and Maat (2007) demonstrate that urban density significantly moderates the relationship between accessibility and commuting behaviour, suggesting that accessibility acts as a variable constraint rather than a universal predictor. This spatial heterogeneity has direct implications for recruitment, particularly in industrial zones outside dense urban areas. In these settings, limited public transport and last-mile issues restrict the effective labour supply, specifically for workers without car access (Boisjoly et al., 2019). Consequently, Geurs and Van Wee (2004) explicitly argue that further research is necessary to understand how job accessibility and competition effects manifest across different spatial situations, a gap this study aims to address.

Taken together, the literature highlights three key gaps. First, while accessibility is widely recognised as a determinant of individual employment outcomes, its role in shaping employer-side recruitment dynamics remains understudied. Second, when accessibility is included in vacancy studies, it often lacks modal specificity or competition effects. Third, few studies explicitly examine how accessibility constraints translate into recruitment difficulties across different spatial contexts. This study addresses these gaps by adopting a workplace-oriented accessibility perspective and linking it directly to vacancy duration. By operationalising both Hansen-type potential accessibility and Shen-type competition accessibility, the analysis assesses whether competition effects are necessary to explain recruitment outcomes across varying urban densities.

## III. Methodology

This section outlines the methodological framework used to analyse the relationship between workplace accessibility and labour market matching efficiency. The research employs a quantitative approach, utilising spatial analysis and regression modelling to determine whether accessibility constraints influence the time required to fill vacancies.

The analytical strategy consists of three phases. First, the accessibility metrics are formalised, distinguishing between potential accessibility (workforce size) and competitive accessibility (accounting for competition). Second, vacancy duration is defined as the primary indicator of matching friction. Third, the relationship between these variables is estimated using stepwise Ordinary Least Squares (OLS) regression models, while explicitly testing for spatial dependence and interaction effects with urban density.

**A. Measuring Accessibility**

This study operationalises accessibility based on the potential of a firm (at location $i$) to reach the available labour supply (at locations $j$). To capture different dimensions of this potential, two metrics are employed: the classic potential accessibility measure by (Hansen, 1959) and the competitive accessibility measure by (Shen, 1998).
For consistency, the following notation is used across both metrics:

- $i$: The workplace location (destination for the commuter, origin of the vacancy).
- $j$: The residential location (origin of the commuter).
- $W_j$: The size of the workforce living in zone $j$ (Labour Supply).
- $P_k$: The number of jobs located in zone $k$ (Labour Demand).
- $t_{ij}$: The travel time between $i$ and $j$.
- $f(t_{ij})$: The impedance function, describing the decay of interaction as travel time increases.

*1. Potential Accessibility (Hansen)*

Potential accessibility measures the absolute size of the workforce that can reach a workplace within a reasonable time. It assumes that the attractiveness of a location depends solely on the proximity to workers, disregarding competition from other firms. Following (Hansen, 1959), the accessibility of workplace zone $i$ is defined as:

$$A_i^{pot} = \sum_j W_j \cdot f(t_{ij}) \tag{1}$$

This metric reflects the *gross* labour pool available to an employer. A high $A_i^{pot}$ indicates that a firm is located close to large concentrations of workers. However, it implicitly assumes that these workers are available to any firm that reaches them, ignoring that these workers may also be within reach of many other employers.

*2. Competitive Accessibility (Shen)*

In dense urban regions, proximity to workers does not guarantee recruitment success if many other firms compete for the same labour pool. To account for this demand-side competition, (Shen, 1998) introduced a competitive accessibility measure. This metric adjusts the gross labour pool by the level of competition.

The calculation follows a two-step process: First, for every residential zone $j$, we calculate the total number of jobs that are reachable for the workers living there. This represents the intensity of demand for their labour:

$$D_j = \sum_k P_k \cdot f(t_{kj}) \tag{2}$$

where $P_k$ represents the number of jobs at any workplace location $k$. A high $D_j$ implies that workers in zone $j$ have many employment options, making them relatively scarce and harder to attract for any single firm. The competitive accessibility of the specific firm location $i$ is then calculated by summing the available workers ($W_j$), but weighing them inversely by their job opportunities ($D_j$):

$$A_i^{comp} = \sum_j \frac{W_j \cdot f(t_{ji})}{D_j} \tag{3}$$

By dividing $W_j$ by $D_j$, this formula calculates the net available workforce. It distinguishes between locations that are simply central (high Hansen accessibility) and locations that hold a strategic advantage (high Shen accessibility) where the supply of labour exceeds the local demand from competing firms.

Both metrics rely on the same impedance function $f(t_{ij})$ to translate physical travel time into perceived cost. This study employs a power-exponential decay function calibrated on Dutch commuting data (as detailed in section V), ensuring that both accessibility measures reflect the realistic commuting tolerance of the Dutch workforce.

*3. Dependent Variable: Vacancy Duration*
To quantify labour market matching efficiency, this study focuses exclusively on Vacancy Duration. This is defined as the median number of days between the publication and filling (or deactivation) of a vacancy within a postcode area. In labour economics, vacancy duration is a widely accepted proxy for matching friction (Mortensen and Pissarides, 1994). Longer durations indicate greater difficulty for employers to find suitable matches, which may be driven by skill mismatches, low wages, or (as hypothesised in this study) poor physical accessibility to the relevant workforce. The median is used to aggregate duration at the postcode level to mitigate the influence of extreme outliers common in survival data.

**B. Regression Analysis Strategy**
The relationship between accessibility and vacancy duration is analysed using regression techniques at the postcode (PC4) level.

*1. Spatial Autocorrelation*
Labour market data are inherently spatial as neighbouring postcodes often share economic characteristics and infrastructure. Ignoring this spatial dependence can lead to biased statistical inference. Therefore, before interpreting regression coefficients, spatial autocorrelation is assessed using Moran's I(Rey et al., 2023):

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \tag{4}$$

A significant Moran's I indicates clustering. If spatial autocorrelation is present in the regression residuals, it suggests that the model fails to capture all spatial processes, which must be considered when interpreting the results.

*2. Regression Models*
A stepwise Ordinary Least Squares (OLS) regression approach is adopted to test the hypotheses. This allows for the isolation of the accessibility effect and the assessment of confounding factors. The general model specification is:

$$y_i = \beta_0 + \beta_1 A_i + \beta_2 \mathbf{X}_i + \varepsilon_i \tag{5}$$

where $y_i$ is the median vacancy duration, $A_i$ is the accessibility metric, and $\mathbf{X}_i$ represents control variables. Five model specifications are estimated:

- Model 1 (Baseline): Tests the bivariate relationship between accessibility (Hansen or Shen) and vacancy duration.
- Model 2 (Urban Context): Adds address density to control for agglomeration effects, distinguishing whether duration dynamics are driven by transport connectivity or broader urban density.
- Model 3 (Interaction): Includes an interaction term ($Accessibility \times Density$). This tests the hypothesis that accessibility acts as a binding constraint in peripheral areas (e.g., the port) but is less critical in dense urban cores.
- Model 4 (Education Mismatch): Controls for the share of high-education vacancies. This tests whether longer durations are caused by spatial mismatch (accessibility) or skills mismatch (education requirements).
- Model 5 (Place-boundedness): Adds the share of place-bound jobs (jobs requiring physical presence) to test if hybrid work potential mitigates accessibility constraints(van Lent et al., 2025).

To mitigate multicollinearity, particularly in the interaction model, continuous variables (accessibility and density) are mean-centred before analysis. While these models control for key spatial and structural labour market factors, it is important to acknowledge that they do not aim to fully explain the variance in vacancy durations. The labour market matching process is influenced by a multitude of unobserved socio-economic and firm-specific variables that are not available at the aggregate postcode level. Critical determinants such as offered wage levels, secondary employment conditions, career advancement opportunities, and firm reputation play a significant role in attracting applicants. Since these variables are absent from the dataset, this study focuses on estimating the specific effect of accessibility on duration, rather than providing a comprehensive predictive model for vacancy filling times.

## IV. Study Area

The empirical analysis focuses on the Rijnmond region (figure 1) in the western Netherlands. This area is interesting because it has two very different sides: the busy, service-oriented city of Rotterdam and the industrial Port of Rotterdam. The port of Rotterdam is the largest port in Europe, which covers over 6,000 hectares and extends approximately 40 kilometres westward into the North Sea. These differences in urban form and land use makes the region particularly suitable for examining how the effects of transport accessibility on labour market outcomes depend on local spatial structure. Besides the city and the port, the region includes several commuter towns like Spijkenisse, Capelle aan den IJssel, and Berkel en Rodenrijs. These are mainly residential areas where many workers live. While these towns have good transport links to Rotterdam's city centre, traveling from these residential areas to the industrial port is often much more difficult and time-consuming by public transport. Finally, the southern part of the region includes more rural areas, such as Middelharnis and Ouddorp. These areas are much less dense and consist mostly of farmland and nature. Unlike the urban cores, these locations have no rail or metro connections. Residents here depend almost entirely on long-distance bus lines or cars to reach the labour markets in the city or the port. The analysis is conducted at the level of 4-digit postal codes (PC4). In the Netherlands, these areas function as spatial proxies for neighborhoods or districts. After filtering for data availability and validity, the final study area comprises $N = 183$ postcodes.

### A. Public Transport Network

Figure 1 visualises the public transport infrastructure within the study area, highlighting a strong monocentric structure. The urban core and immediate suburbs are served by a high-frequency multimodal network consisting of heavy rail, a metro system, and an extensive tram network. This dense grid facilitates rapid transit within the agglomeration and connects residential municipalities such as Schiedam and Capelle aan den IJssel to the Rotterdam city centre. In sharp contrast, the port and industrial areas to the west (Botlek, Europoort, Maasvlakte) exhibit significantly lower connectivity. As shown in Figure 1, the rail infrastructure extending into the port (visualised in orange) is the dedicated *Betuweroute* freight line, which offers no passenger services. The only exception is the *Hoekse Lijn* metro extension (green), which connects the northern bank to the coast but bypasses the major industrial clusters on the southern bank. Consequently, workforce mobility in the port area relies heavily on bus services. Compared to the rail network, these bus connections are characterised by lower frequencies, lower speeds, and the need for transfers, resulting in significantly longer travel times for public transport users compared to car commuters. The port area itself is only served by on-demand transport from Maassluis Centrum, Spijkenisse and Brielle. This discrepancy creates a potential barrier for labour supply in the port region, particularly for workers without access to a car.

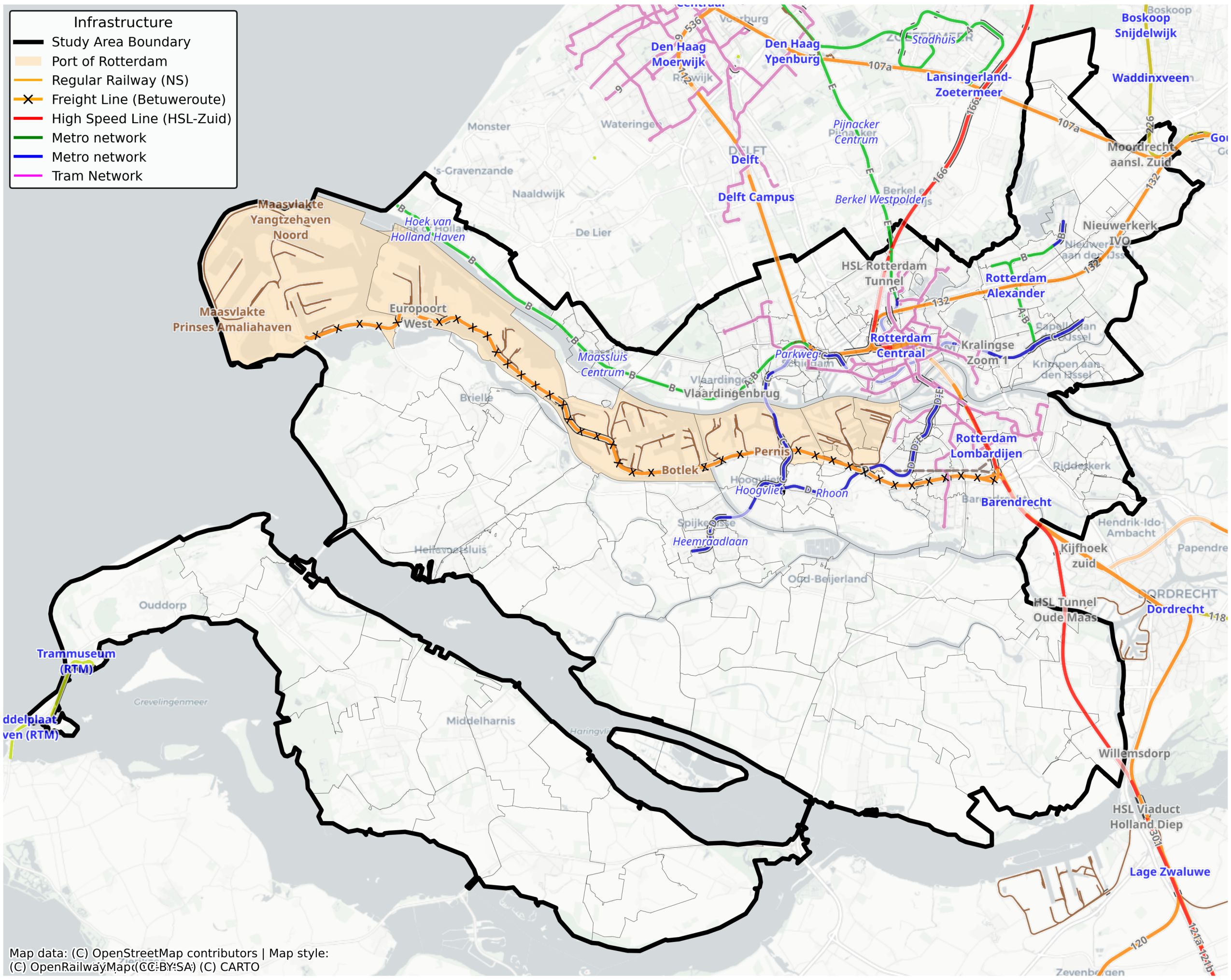



**Figure 1. The Rijnmond area with available public transport infrastructure**

## V. Data and Construction of Variables

To analyse the relationship between accessibility and vacancy duration, this study integrates three primary datasets: online vacancy data (Jobdigger), employment statistics (LISA), and computed travel times (OpenTripPlanner).

### A. Job Vacancies (Dependent Variable)

Vacancy data were obtained from Jobdigger, a comprehensive database of online job postings. The dataset covers the period 2022–2025 and includes information on posting dates, occupational classification (BRC2014), and required education level. To ensure the analysis captures local labour demand rather than administrative flows, the raw data required rigorous preprocessing. Vacancies posted by intermediaries (recruitment agencies) and entries without a fixed, identifiable workplace location were excluded. This filtering step ensures spatial precision, linking the vacancy directly to the physical location where the work is performed.

The primary outcome variable is *vacancy duration*, defined as the number of days a position remains open. This is calculated as the difference between the deactivation date and the initial discovery date. To account for the latency of the web-scraping process, a correction factor of 3.5 days (half the scraping interval) was applied to the raw duration. Given the typically right-skewed distribution of vacancy durations, the median duration per postcode is used to operationalise the vacancy duration(Dossche et al., 2025). To ensure the results are reliable, postcodes with very few vacancies were excluded from the analysis. According to the Central Limit Theorem (CLT), a larger number of observations is needed to get a stable and representative average. Including areas with too little data could lead to unreliable results, as a single outlier could disproportionately affect the median. Therefore, only postcodes with at least 30 vacancies were included to maintain the statistical quality of the study.

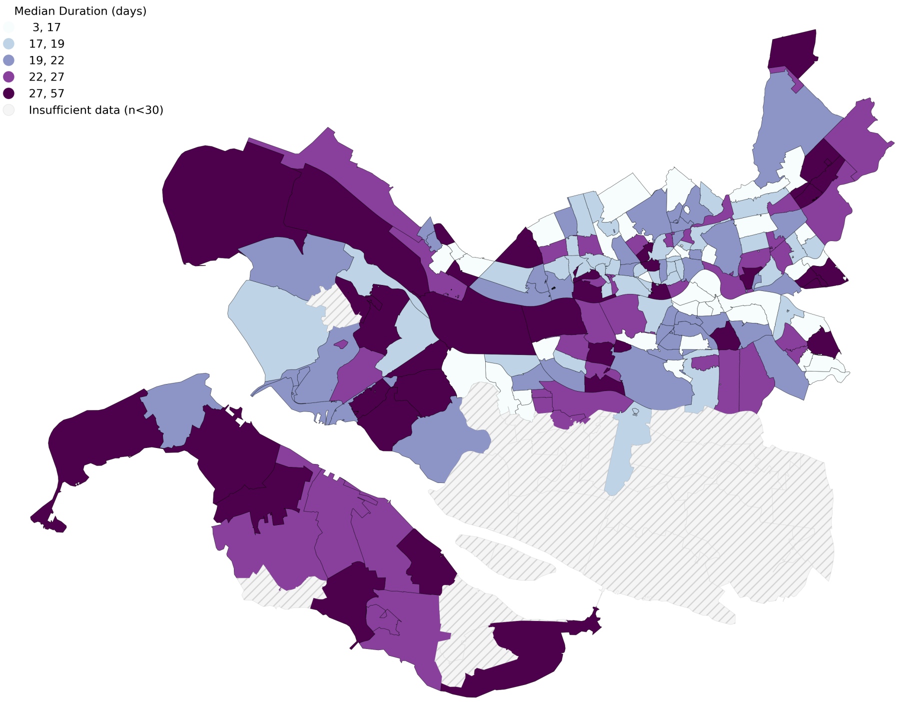



**Figure 2. Spatial distribution of median vacancy duration in the Rijnmond area**

Figure 2 presents the spatial distribution of the median vacancy duration. Visual inspection suggests that vacancies in the port area and other peripheral zones tend to remain open longer compared to the urban core, although the pattern within Rotterdam itself is mixed. This variability highlights the need for a multivariate analytical approach that controls for confounding factors such as address density and job characteristics.

**B. Accessibility Measures (Independent Variable)**

Transport accessibility is modelled as the potential access to the labour supply. While the outcome variable (vacancies) is observed within the Rijnmond region, the accessibility measures account for the potential labour force from the broader Dutch labour market. Travel times ($t_{ij}$) between all 4-digit postcode areas in the Netherlands were computed using OpenTripPlanner (OpenTripPlanner Contributors, 2025). This engine utilises OpenStreetMap (OSM) for walking networks and GTFS data for public transport schedules to generate dynamic travel times. Calculations were performed for a representative weekday at 8:00 AM to capture peak-hour commuting constraints, including access and egress times. Boisjoly and El-Geneidy (2016) concluded from their research that a constant measure at 8:00 AM is representative of the relative accessibility and highly correlated with more sophisticated dynamic measures as presented by Owen and Levinson (2015). To reflect realistic commuter behaviour, the travel times are weighted by an impedance function derived from the Dutch National Travel Survey (ODiN). Based on observed home-to-work commuting trips ($N$ = 7328), a power-exponential decay function provided the best fit ($R^2$ = 0.996) and is applied to the travel time matrix. Finally, the spatial distribution of the potential labour force is derived from the LISA dataset (Landelijk Informatiesysteem Arbeidsplaatsen (BIJ12, 2024)), which contains yearly information on the number of jobs per postcode, allowing for the calculation of competitive accessibility indicators where required.

Figures 3 and 4 visualise the spatial distribution of the two accessibility metrics within the Rijnmond region.

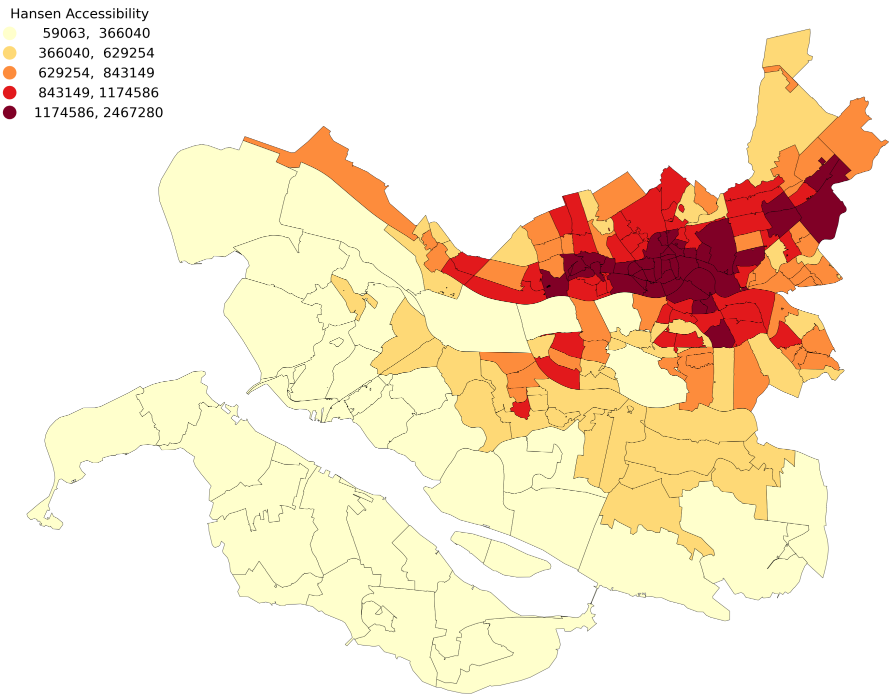



**Figure 3. Hansen accessibility in the Rijnmond area**

Figure 3 displays the Hansen Potential Accessibility(Hansen, 1959), which represents the absolute size of the workforce that can reach a location within a reasonable travel time. As expected, this measure follows a clear mono-centric pattern: accessibility is highest in the dense urban core of Rotterdam and along major public transport axes (such as the metro lines extending to Schiedam and Capelle aan den IJssel). In contrast, the Port of Rotterdam and the peripheral south exhibit significantly lower accessibility due to greater distances and lower public transport frequencies.

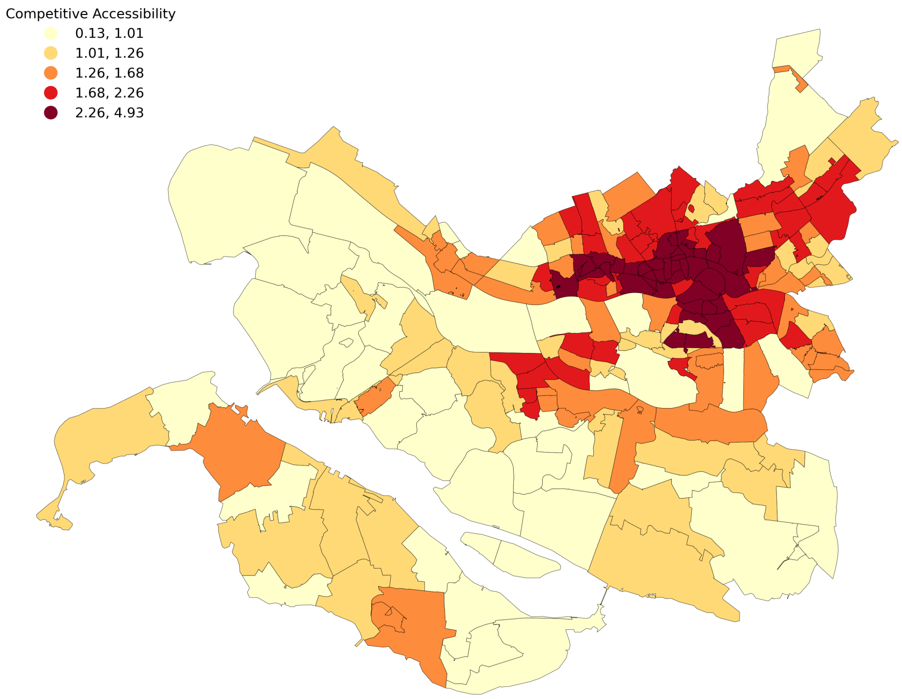



**Figure 4. Competitive accessibility in the Rijnmond area**

Figure 4 illustrates the Shen Competitive Accessibility(Shen, 1998), which adjusts potential accessibility by accounting for competition from other employers. While the general core-periphery structure remains visible, this metric reveals more nuanced local dynamics. Notably, some peripheral locations achieve relatively higher scores compared to the Hansen metric. This occurs because, although these locations have lower absolute connectivity, they face less competition for the local workforce compared to the city centre. Conversely, in the urban core, high connectivity is partially offset by the intense concentration of jobs, meaning that despite the large potential workforce, the effective labour supply per vacancy is weakened by high demand.

### C. Urban Context and Controls

To control for agglomeration effects and local spatial structure, urban context is captured using postcode-level address density, obtained from Statistics Netherlands (Centraal Bureau voor de Statistiek (CBS), 2025). Address density varies substantially within the study area, ranging from very high levels in the inner-city districts of Rotterdam to extremely low densities in the port and industrial zones. In the regression analysis, address density serves as both a control variable and a moderator to test for heterogeneous effects of accessibility across different urban forms(Geurs and Van Wee, 2004).

## VI. Results

This section presents the empirical analysis of the relationship between accessibility and vacancy duration. Building on the descriptive patterns observed in Section V, this section first quantifies the spatial dependence of the variables and then presents the stepwise regression models.

### A. Assessing spatial dependence

To ensure the validity of the Ordinary Least Squares (OLS) assumptions, the presence of spatial autocorrelation in the vacancy duration variable and the model residuals was assessed using Global Moran's I. The analysis is conducted on the filtered dataset ($N$ = 183). The results, based on a Queen's contiguity matrix, are presented in Table 1.

| Variable | Moran's I | p-value | Result |
|---|---|---|---|
| Vacancy duration | 0.139 | 0.003 | Significant Clustering |
| OLS residuals | 0.050 | 0.107 | Random |

**Table 1. Global Moran's I statistics for Median Vacancy Duration** ($N$ = 183)

Table 1 shows that the raw vacancy duration ($I = 0.139$, $p = 0.003$) exhibits significant positive spatial autocorrelation. This confirms that matching frictions are not randomly distributed but follow a clear regional structure. However, after controlling for workplace accessibility and urban density in the OLS specification, the spatial dependence in the residuals decreases substantially ($I = 0.050$) and becomes statistically non-significant ($p = 0.107$). This implies that the included variables, specifically the interaction between accessibility and density, successfully capture the spatial structure of the labour market, leaving no significant unexplained spatial clustering

Visual inspection of the OLS residuals (Figure 5) supports these statistical findings. The error terms show no distinct geographic clusters of high or low values. Instead, a balanced mix of over- and underestimates is observed across the Rijnmond region. Because the regression model successfully internalises the primary spatial dependencies, more complex spatial regression techniques are not required for a reliable interpretation of the accessibility effects.

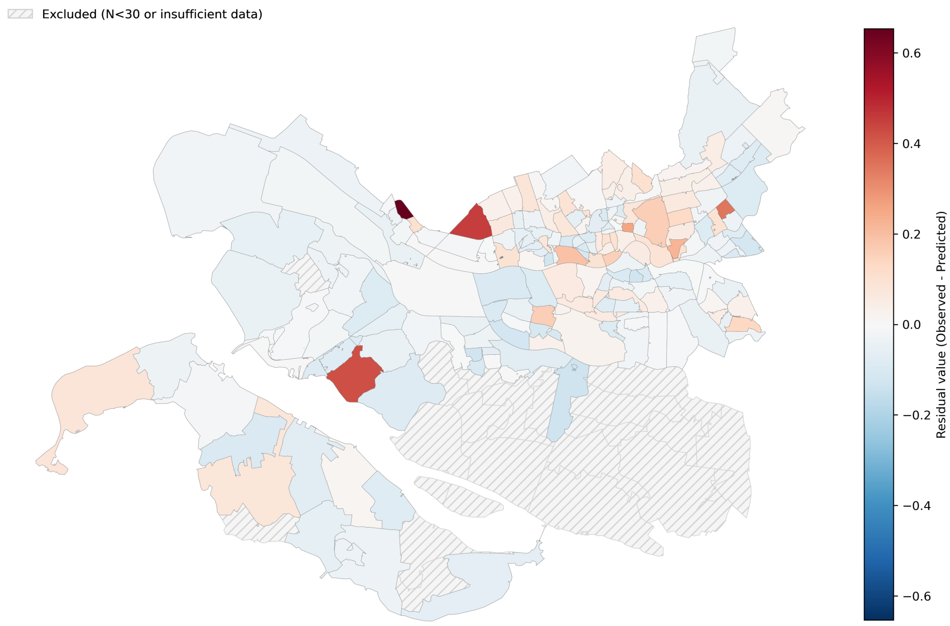



**Figure 5. Map of the OLS residuals for vacancy duration**

**B. Regression Analysis**

To quantify the relationship between accessibility and matching friction, median vacancy duration is regressed against various sets of explanatory variables.

*1. Correlations*

Bivariate correlations indicate a moderate negative correlation between vacancy duration and both accessibility measures, with potential accessibility (Hansen) exhibiting a stronger association ($r = -0.38$) than competitive accessibility ($r = -0.26$). This provides initial support for the hypothesis that better accessibility reduces vacancy duration. Additionally, a positive correlation ($r = 0.24$) is observed between the share of place-bound work and duration, highlighting specific frictions within industrial sectors. Accessibility measures are strongly correlated with address density ($r = 0.69$), but Variance Inflation Factors (VIF) remain below critical thresholds in all models, indicating that multicollinearity does not bias the coefficient estimates.

*2. Regression Results*

**Table 2. Regression Results: Determinants of Vacancy Duration in the Rijnmond Region**

| Variable | ($M1_a$) Hansen | ($M1_b$) Competitive | ($M2$) Baseline | ($M3$) Interaction | ($M4$) Education | ($M5$) Placebound |
|---|---|---|---|---|---|---|
| Constant | 21.51*** | 25.16*** | 21.40*** | 20.22*** | 18.87*** | 20.89*** |
| | (0.46) | (1.16) | (0.65) | (0.60) | (1.46) | (0.76) |
| log_hansen_c | -3.44*** | | -3.01*** | -0.38 | -3.19*** | -2.81*** |
| | (0.71) | | (1.03) | (1.21) | (1.08) | (1.02) |
| Competitive_acc. | | -2.08*** | | | | |
| | | (0.49) | | | | |
| density_c | | | -0.0003 | -0.0014** | -0.0005 | -0.0002 |
| | | | (0.0003) | (0.0005) | (0.0003) | (0.0003) |
| interaction_c | | | | 0.0013*** | | |
| | | | | (0.0005) | | |
| fract_high_ed | | | | | 6.85* | |
| | | | | | (3.64) | |
| fract_placebound | | | | | | 4.84 |
| | | | | | | (4.12) |
| R-squared | 0.148 | 0.070 | 0.151 | 0.182 | 0.172 | 0.156 |
| Adj. R-squared | 0.143 | 0.065 | 0.136 | 0.168 | 0.154 | 0.137 |
| Max. VIF | 1.00 | 1.00 | 1.91 | 5.90 | 2.09 | 2.02 |
| N | 183 | 183 | 183 | 183 | 183 | 183 |

*Note:* Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Accessibility and density variables are centred.

The regression results are presented in Table 2. Models $M1_a$ and $M1_b$ compare the explanatory power of the two accessibility measures. The results indicate that competitive accessibility ($M1_b$) has significantly lower explanatory power ($R^2 = 0.070$) compared to potential accessibility ($M1_a$, $R^2 = 0.148$). This contradicts the general expectation in accessibility literature that competition measures offer superior explanatory power. It suggests that in the Rijnmond region, specifically for the port, the absolute barrier of travel time (Hansen) is a more decisive constraint than the ratio of job seekers (Shen). As expected, the coefficient is negative for both measures, implying that better accessibility is associated with shorter vacancy durations. Model $M2$ introduces address density as a control. Accessibility remains significant and strong, while the addition of density barely improves the model fit. This suggests that in the baseline specification, accessibility is the primary driver of the spatial variation in vacancy duration. The interaction model ($M3$) reveals a critical nuance. The interaction term ($Accessibility \times Density$) is positive and significant ($p < 0.01$), and the model fit improves to $R^2 = 0.182$. The main effect of accessibility becomes insignificant, indicating that the impact of accessibility is conditional on the urban context. The effect is visualised in Figure 6:

- In low-density areas: Better accessibility strongly reduces vacancy duration.
- In high-density areas: This effect flattens out and even reverses.

This suggests that accessibility is the primary barrier in peripheral areas, but in dense urban cores, higher accessibility may correlate with longer durations due to congestion effects or selective hiring processes.

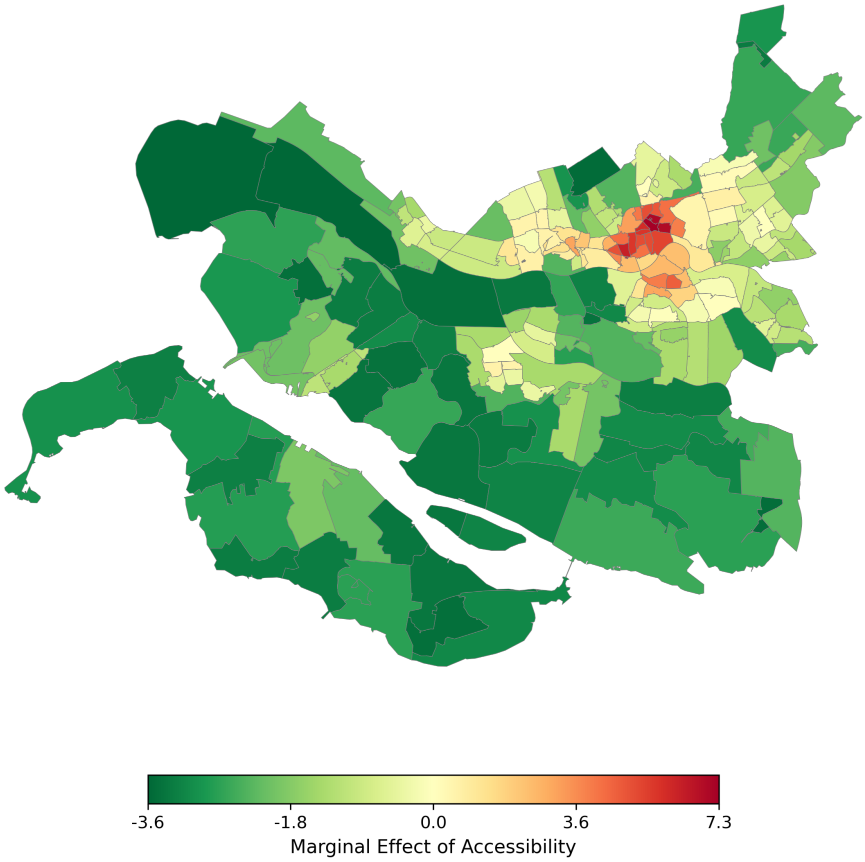



**Figure 6. Spatial heterogeneity of the marginal effect of accessibility on vacancy duration. Negative values (green) indicate that higher accessibility is associated with shorter vacancy durations, while positive values (red) indicate an increase in duration.**

To illustrate the practical implications of this interaction effect, a scenario analysis was performed for the Maasvlakte (westernmost tip of the Port of Rotterdam, directly facing the North Sea), a representative location for the peripheral, industrial cluster. Based on the coefficients of Model $M3$, Figure 7 shows the predicted impact of a 5-minute reduction in public transport travel time. The model estimates that this relatively minor infrastructure improvement would decrease the median vacancy duration by approximately one full day ($\Delta y \approx -1.0$). This confirms that in low-density industrial settings, accessibility is a highly elastic determinant of recruitment efficiency.

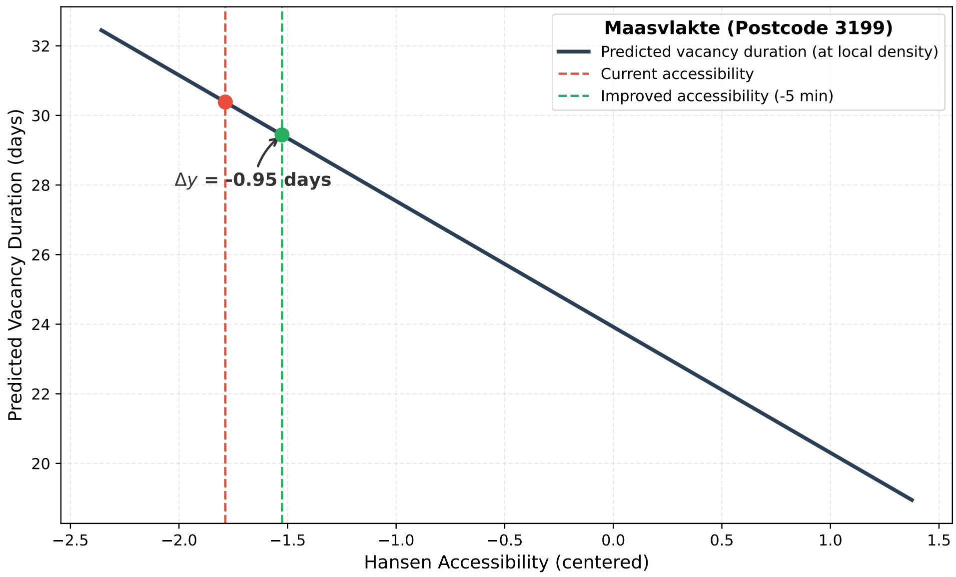



**Figure 7. Predicted impact of a 5-minute travel time reduction on vacancy duration for the Maasvlakte industrial area.**

Models $M4$ and $M5$ test for compositional effects. The share of high-education vacancies ($M4$) shows a weakly significant positive effect, aligning with the theory that specialised matching takes longer. The share of place-bound jobs ($M5$) is positive but not statistically significant. The lack of significance for place-bound work suggests that the spatial location (captured by the interaction term) is a stronger predictor of friction than the sectoral characteristics of the vacancy itself. Importantly, in both specifications, the influence of accessibility remains robust, suggesting that sectoral composition alone does not explain the spatial variation in vacancy duration.

## VII. Conclusion and Discussion

This study set out to quantify the extent to which public transport workplace accessibility explains the spatial variation in vacancy duration in the Rijnmond region. By shifting the analytical lens from the employee's probability of finding work to the employer's efficiency in filling vacancies, this research provides new empirical evidence on the employer-side consequences of accessibility constraints.

### A. Principal Findings

The regression analysis demonstrates that public transport accessibility is a statistically significant determinant of labour market matching efficiency. The Hansen potential accessibility measure, interacting with urban density, explains approximately 18.2% of the spatial variance in median vacancy duration.

The results reveal a distinct spatial paradox. In peripheral and industrial areas, such as the Port of Rotterdam, accessibility functions as a binding constraint. Here, the friction of distance limits the effective labour supply, and improvements in accessibility are associated with a significant reduction in vacancy duration. Conversely, in dense urban cores, the relationship weakens and even reverses. This suggests that while transport infrastructure is a facilitator for matching in peripheral zones, it is not the governing factor in the city centre.

Furthermore, the study finds that the potential accessibility measure (Hansen) outperforms the competitive measure (Shen) in predicting vacancy duration. This indicates that for firm-side outcomes at the regional level, the absolute size of the reachable workforce is a more robust predictor of recruitment speed than complex competition ratios, likely due to the dominant role of physical distance barriers in the port area.

**B. Theoretical Implications: Spatial and Modal Mismatch**

These findings strongly support the Spatial Mismatch Hypothesis (Kain, 1968) from an employer's perspective. The prolonged vacancy durations in the Port of Rotterdam confirm that the physical separation between employment centres and residential concentrations creates structural frictions in the matching process. However, the results suggest that this is not merely a problem of distance, but of Modal Mismatch (Grengs, 2010).

The peripheral industrial clusters are characterised by high car dependence, effectively excluding the segment of the urban workforce that relies on public transport. The strong explanatory power of public transport accessibility in these areas indicates that the matching friction is driven by a lack of modal options. For the non-car-owning workforce, the port is not just far away, it is effectively disconnected.

In contrast, the reversal effect observed in urban areas highlights the limits of the mismatch theory. In high-accessibility zones, the spatial and modal constraints are resolved. Here, the bottleneck shifts from the *application period* (finding candidates) to the *selection period* (screening candidates). The abundance of candidates facilitated by high accessibility may lead to "screening congestion" or encourage firms to act more selectively, paradoxically prolonging the process (Puga, 2010).

**C. Policy Implications**

For policymakers and planners in the Rijnmond region, addressing these mismatches requires a differentiated strategy.

In the Port of Rotterdam, poor public transport provision acts as a structural barrier. Current policy often evaluates public transport based on ridership efficiency. However, in industrial zones, public transport should be reframed as an economic tool to bridge the modal mismatch. Even if ridership is low, the existence of a connection is crucial to unlock the labour potential of non-car-owning demographics (e.g., young starters, urban support staff). Investments here should prioritize coverage and connectivity (e.g., via Demand Responsive Transport) rather than frequency, to effectively widen the applicant pool for "hard-to-fill" vacancies. Additionally, the spatial mismatch can be reduced by bridging the physical divide between the urban core and the industrial port. Strategic land-use interventions should focus on relocating knowledge-intensive and non-place-bound port functions to urban transitional zones with increased connectivity. By bringing these specific employment opportunities closer to the residential concentrations of the workforce, firms can effectively bypass the accessibility constraint, transforming a peripheral vacancy into an accessible urban position.

In the urban core, the observed longer vacancy durations should not be interpreted as a sign of inefficiency, but rather as a symptom of a robust business climate. High accessibility creates a deep and thick labour market, which acts as a major pull factor for firm location (Puga, 2010). In these high-density environments, firms benefit from agglomeration economies, allowing them to be highly selective and wait for the perfect match. Consequently, the "reversal effect" reflects a luxury problem of abundance: accessibility has successfully concentrated such a large talent pool that the challenge shifts from finding an applicant to selecting the best one.

**D. Limitations and Future Research**

*1. limitations*

These findings should be interpreted within the context of several limitations. First, regarding causality, the cross-sectional design cannot fully rule out firm self-selection effects. It remains plausible that firms with hard-to-fill vacancies cluster in the port due to land-use zoning rather than accessibility constraints per se. Second, the reliance on public transport data likely underestimates the accessibility levels in the highly car-dependent port area. Future studies should ideally employ a relative accessibility measure (e.g., the ratio of public transport to car accessibility) to capture the true competitive disadvantage of non-car modes. Finally, the use of aggregate accessibility measures treats the labour market as homogeneous, potentially masking skill mismatches between the available workforce and the specialised technical requirements of the port industry.

*2. Future research*

Future research should verify these findings using longitudinal data to isolate the effect of transport improvements from firm sorting effects. Additionally, integrating space-time accessibility measures that account for shift-work schedules would provide a more realistic picture of the modal mismatch facing the industrial workforce. Additionally, to address the issue of occupational heterogeneity, future research should refine the accessibility measures by segmenting the labour market. By calculating accessibility specifically for technical staff, healthcare workers, or logistics personnel separately, researchers can test whether the spatial mismatch is more severe for specific critical sectors. This would allow for more targeted policy interventions than the current aggregate approach. Lastly, this study identifies a significant interaction between accessibility and density, but the linear interaction term assumes a gradual transition. Future research should investigate the specific turning point, the exact threshold value of accessibility or density where the relationship flips from negative (reducing duration) to positive (increasing duration). This would answer a critical planning question: "How much accessibility is enough?" enabling policymakers to target investments only up to the point of diminishing returns, preventing over-investment in areas where accessibility is no longer the binding constraint.

## References

Andrews, M., Bradley, S., Stott, D., and Upward, R. (2008). Successful employer search? an empirical analysis of vacancy duration using micro data. *Economica*, 75(299):455–480.

Bastiaanssen, J., Johnson, D., and Lucas, K. (2022). Does better job accessibility help people gain employment? the role of public transport in great britain. *Urban Studies*, 59(13):2602–2621.

Bastiaanssen, J., Johnson, D., and Lucas, K. (2025). The relationship between individual employment probabilities and accessibility to matching jobs: A study of the netherlands. *Transportation Research Part A: Policy and Practice*, 193:104398.

BIJ12 (2024). Lisa employment register. Available at: https://www.lisa.nl. Accessed 2024.

Birinci, S., Mercan, Y., and See, K. (2025). Mismatch unemployment during covid-19 and the post-pandemic labor shortages. *Journal of Economic Dynamics and Control*, 178:105142.

Boisjoly, G. and El-Geneidy, A. M. (2016). Daily fluctuations in transit and job availability: A comparative assessment of time-sensitive accessibility measures. *Journal of Transport Geography*, 52:73–81.

Boisjoly, G., El-Geneidy, A. M., and Levinson, D. (2019). Accessibility and the journey to work through the lens of equity. *Journal of Transport Geography*, 74:269–277.

Centraal Bureau voor de Statistiek (CBS) (2025). Cbs open data. `https://opendata.cbs.nl`. Accessed: February 18, 2026.

Dekker, F. and van der Toorn, A.-J. (2024). Arbeidsmarktonderzoek hic 2024 perspectief in het rotterdamse havengebied. Technical report, SEOR BV, Rotterdam.

Dossche, W., Vansteenkiste, S., Baesens, B., and Lemahieu, W. (2025). Anticipating delays in recruitment: Explainable machine learning for the prediction of hard-to-fill online job vacancies. *European Journal of Operational Research*. Published online ahead of print.

Geurs, K. T. and Ritsema van Eck, J. R. (2001). Accessibility indicators: literature review and applications for the netherlands. *RIVM Report*, 408505006:1–57.

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

Grengs, J. (2010). Job accessibility and the modal mismatch in detroit. *Journal of Transport Geography*, 18(1):42–54.

Handy, S. (2020). Is accessibility an idea whose time has finally come? *Transportation Research Part D: Transport and Environment*, 83:102319.

Hansen, W. G. (1959). How accessibility shapes land use. *Journal of the American Institute of Planners*, 25(2):73–76.

Hu, Y. and Downs, J. (2019). Measuring and visualizing place-based space-time job accessibility. *Journal of Transport Geography*, 74:278–288.

Ingram, D. R. (1971). The concept of accessibility: A search for an operational form. *Regional Studies*, 5(2):101–107.

Kain, J. F. (1968). Housing segregation, negro employment, and metropolitan decentralization. *The Quarterly Journal of Economics*, 82(2):175–197.

Mangan, J. and Trendle, B. (2017). Hard-to-fill vacancies: An analysis of demand side responses in the australian state of queensland. *Economic Analysis and Policy*, 54:49–56.

Mortensen, D. T. and Pissarides, C. A. (1994). Job creation and job destruction in the theory of unemployment. *The Review of Economic Studies*, 61(3):397–415.

OpenTripPlanner Contributors (2025). Opentripplanner. `https://github.com/opentripplanner/OpenTripPlanner`.

Accessed: 2025-10-20.

Owen, A. and Levinson, D. (2015). Modeling the commute mode share of transit using continuous accessibility to jobs. *Transportation Research Part A: Policy and Practice*, 74:110–122.

Pfertner, M., Büttner, B., and Wulfhorst, G. (2023). An open-source modelling methodology for multimodal and intermodal accessibility analysis of workplace locations. *Sustainability*, 15(3):1947.

Port of Rotterdam (2025). Rotterdam port vision 2050. Policy document, Port of Rotterdam Authority and Municipality of Rotterdam, Rotterdam.

Puga, D. (2010). The magnitude and causes of agglomeration economies. *Journal of Regional Science*, 50(1):203–219.

Qu, Y., Currie, G., and Webb, G. I. (2025). Spatial resolution impact on public transport accessibility measurement error. *Journal of Transport Geography*.

Rey, S. J., Arribas-Bel, D., and Wolf, L. J. (2023). *Geographic Data Science with Python*. CRC Press, Boca Raton.

Shen, Q. (1998). Location characteristics of inner-city neighborhoods and employment accessibility of low-wage workers. *Environment and Planning B: Planning and Design*, 25(1):1–17.

Shi, Y., Blainey, S., Sun, C., and Jing, P. (2020). A literature review on accessibility using bibliometric analysis techniques. *Journal of Transport Geography*, 87:102810.

Susilo, Y. and Maat, K. (2007). The influence of built environment to the trends in commuting journeys in the netherlands. *Transportation*, 34:589–609.

van Lent, S. J., Ulak, M. B., Wismans, L. J., and Geurs, K. T. (2025). The impact of teleworking in hybrid job accessibility in the netherlands: A novel agent-based modelling framework. *Transportation Research Part A: Policy and Practice*, 196:104496.

Wegener, M. and Fürst, F. (1999). Land-use transport interaction: State of the art. Technical Report 47, Institut für Raumplanung, Universität Dortmund.

# B

# Port area

Some contextual information and specific results from the port area are provided in this chapter.

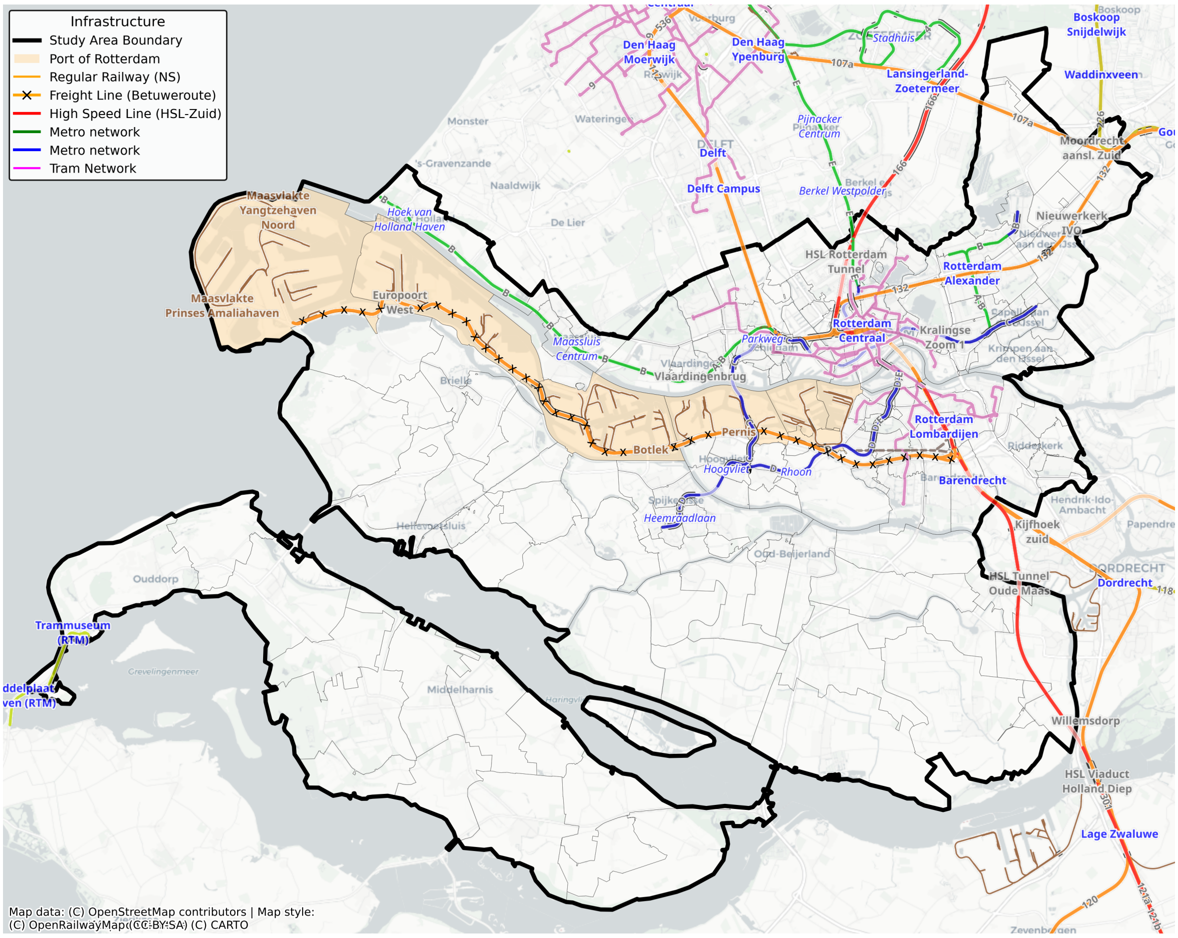



**Figure B.1:** Description of the Rijnmond area with relevant public transport options. The Maasvlaktehopper serves the Europoort and Maasvlakte area from Brielle, Maassluis and Maasland

**Table B.1:** Internal comparison of PC4 areas within the Port of Rotterdam vs. Cluster 3 Reference

| **Indicator** | **3087** | **3088** | **3089** | **3196** | **3197** | **3198** | **3199** | **Cl3 Average** |
|---|---|---|---|---|---|---|---|---|
| *Age groups* | | | | | | | | |
| <25 | 5.00% | 3.17% | 9.23% | 2.67% | 3.43% | 6.38% | 3.54% | 7.07% |
| 25–35 | 22.50% | 19.05% | 15.38% | 18.67% | 18.29% | 15.96% | 30.09% | 21.50% |
| 35–50 | 32.50% | 38.10% | 35.38% | 34.67% | 38.29% | 34.04% | 33.63% | 32.57% |
| 50–65 | 32.50% | 36.51% | 40.00% | 42.67% | 37.71% | 39.36% | 30.09% | 34.62% |
| *Gender* | | | | | | | | |
| Male | 82.50% | 77.78% | 70.77% | 89.33% | 86.29% | 78.72% | 85.84% | 75.22% |
| *Education* | | | | | | | | |
| Low educated | 12.50% | 17.46% | 16.92% | 14.67% | 13.14% | 17.02% | 15.04% | 18.05% |
| High educated | 52.50% | 38.10% | 36.92% | 25.33% | 35.43% | 37.23% | 28.32% | 37.63% |
| *Car ownership* | | | | | | | | |
| Car ownership | 52.50% | 61.90% | 64.62% | 76.00% | 69.14% | 68.09% | 74.34% | 65.02% |
| *Income* | | | | | | | | |
| Low income | 12.50% | 1.59% | 3.08% | 2.67% | 6.29% | 4.26% | 0.88% | 9.81% |
| High income | 52.50% | 52.38% | 60.00% | 68.00% | 65.71% | 69.15% | 60.18% | 47.22% |
| *Social participation* | | | | | | | | |
| parttime | 7.50% | 1.59% | 10.77% | 1.33% | 4.57% | 4.26% | 1.77% | 10.67% |
| student | 2.50% | 0.00% | 3.08% | 0.00% | 1.14% | 1.06% | 0.00% | 2.25% |
| *Household* | | | | | | | | |
| Single households | 12.50% | 23.81% | 20.00% | 14.67% | 18.29% | 19.15% | 20.35% | 17.75% |
| Households with children | 47.50% | 53.97% | 53.85% | 54.67% | 52.00% | 53.19% | 49.56% | 52.73% |

| **Covariates** | **3087** | **3088** | **3089** | **3196** | **3197** | **3198** | **3199** | **Cl3 Ref.** |
|---|---|---|---|---|---|---|---|---|
| *Active* | | | | | | | | |
| Accessibility | 0.27 | 0.18 | 0.12 | 0.09 | 0.11 | 0.02 | 0.04 | 0.1478 |
| *Inactive* | | | | | | | | |
| Car usage | 70.00% | 84.13% | 86.15% | 93.33% | 95.43% | 92.55% | 98.23% | 75.63% |
| Public transport usage | 10.00% | 3.17% | 6.15% | 0.00% | 0.00% | 0.00% | 0.88% | 3.22% |
| Address density | 0.20 | 0.03 | 0.03 | 0.01 | 0.02 | 0.00 | 0.00 | 0.0840 |
| Jobs | 0.10 | 0.09 | 0.10 | 0.05 | 0.18 | 0.08 | 0.11 | 0.1287 |

# C
# Verification of travel times

An assessment was carried out to evaluate whether the calculated travel times were realistic. Figure C.2 presents the travel times to eight randomly selected postcodes in the Netherlands. Several plots clearly display a noticeable boundary around 150 km, indicating that this threshold is both visible in the data and a plausible distance limit. Although certain specific connections, such as the route from Utrecht Centraal to Groningen Centraal, exceed 150 km while remaining just under two hours, such cases occur only for major intercity stations. When walking times and transfers are included, these trips typically surpass the two-hour mark. In addition, these postcodes already exhibit high levels of accessibility, and due to the impedance function approaching zero for long travel times, these exceptional cases contribute very little to the overall accessibility values.

Some travel time plots contain more gaps than others. This can be attributed to the rule that when the travel time from i to j exceeded 180 minutes, the reverse direction (j to i) was not computed. As a result, some degree of incompleteness in the dataset is unavoidable. Figure C.1 illustrates the percentage of travel times that could be calculated for each postcode. The effects of the applied constraints are clearly visible: central areas of the Netherlands have substantially higher coverage compared to regions such as Limburg, Groningen, and Zeeland. A few postcodes appear in red due to having no calculated travel times, although none of these are located within the Rijnmond region.

Furthermore, the ten ODiN postcodes with the lowest coverage were manually examined. Only the postcode with the lowest coverage (8218, Lelystad Airport, 0.5%) appeared to be a significant outlier. Although a bus service operates to Lelystad Airport, valid routes were only identified in a limited number of cases. This is likely due to the surrounding area of the Flevopolder being sparsely populated and the selected grid cell being slightly misplaced. The remaining low-coverage postcodes all had at least 16% coverage and were located in Groningen, Limburg, or Zeeland.

Overall, this analysis indicates that the calculated travel times are generally consistent and reliable.

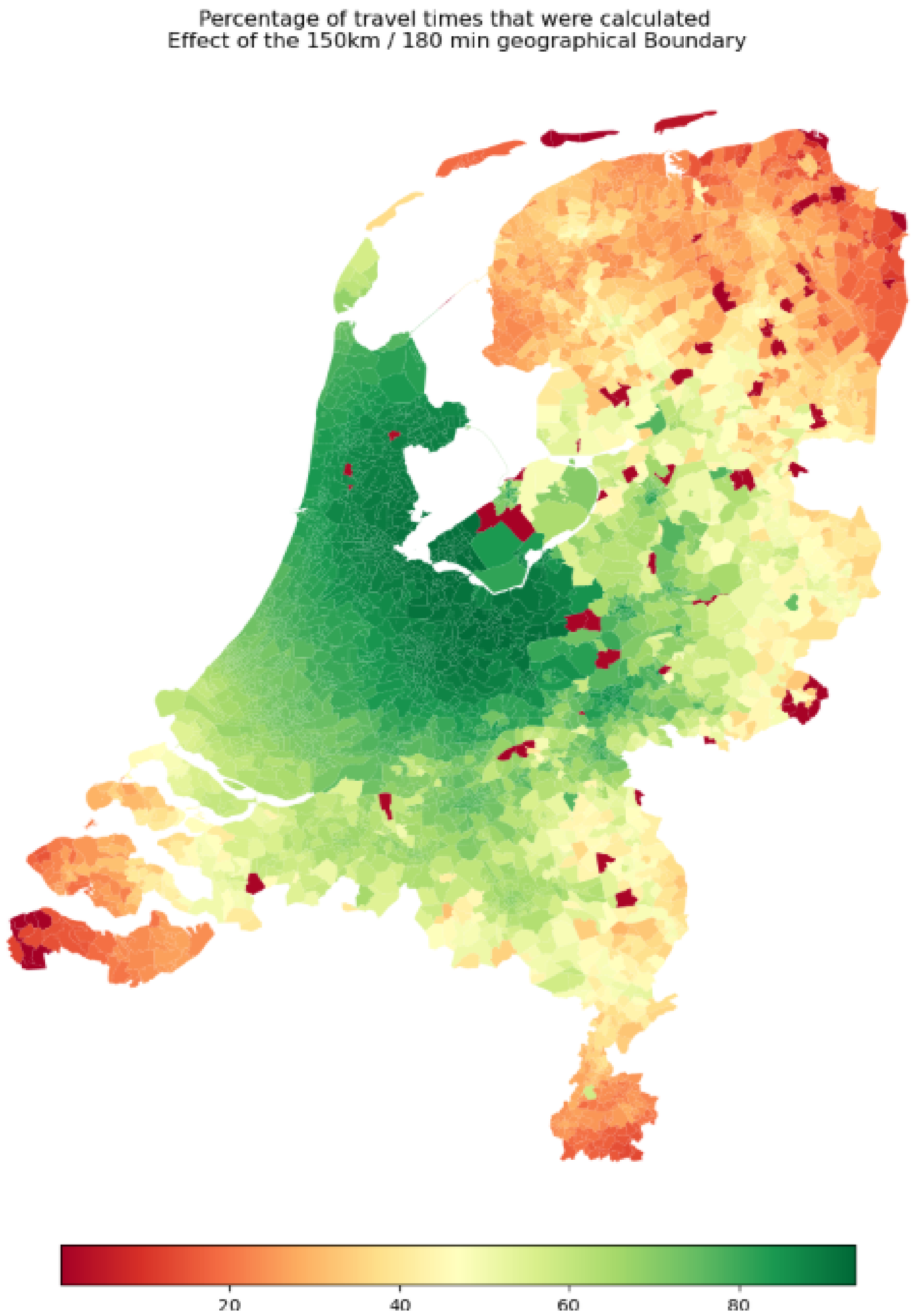



**Figure C.1:** Percentage of travel times that could be calculated for each postcode

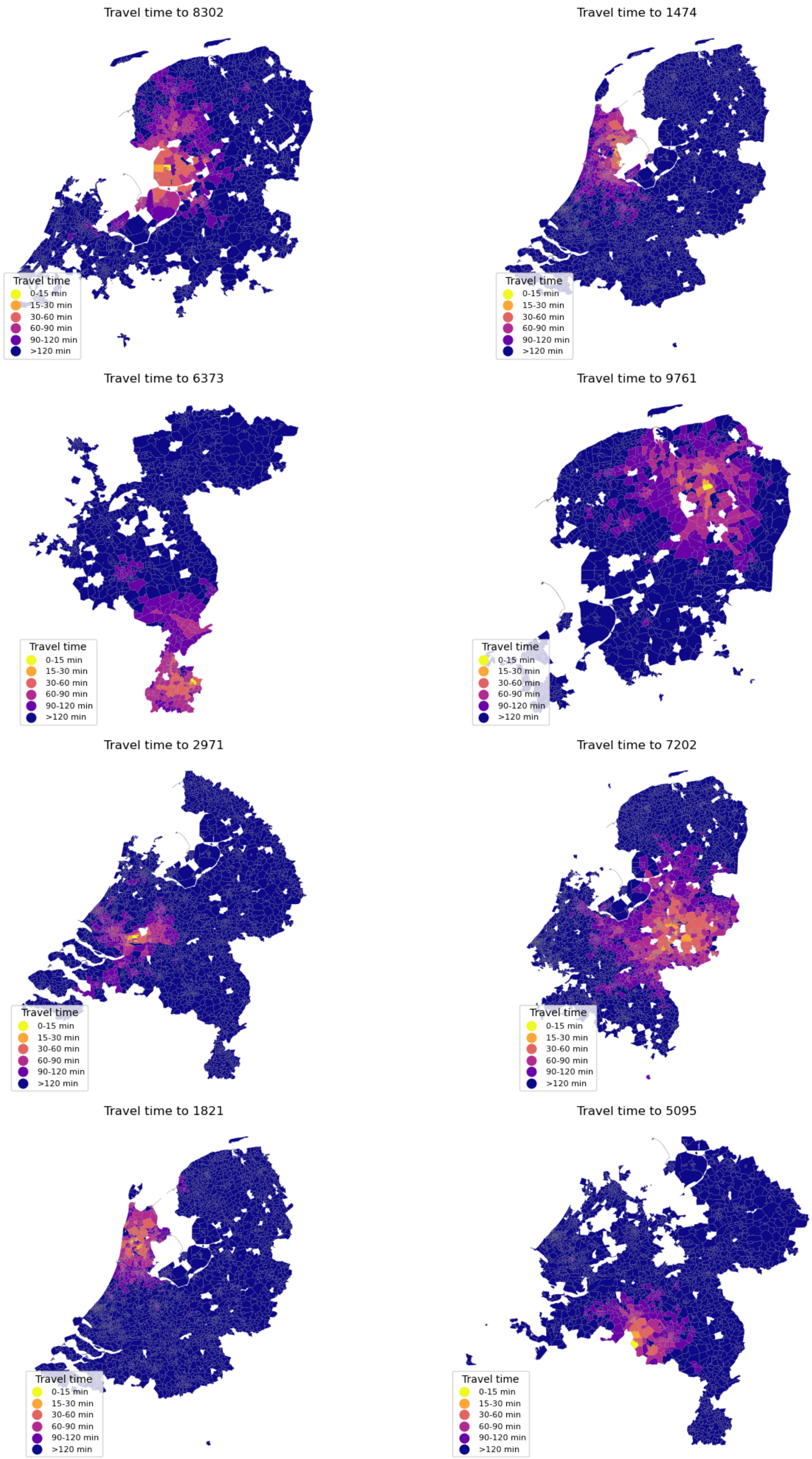



**Figure C.2:** Travel time plot to random postcodes

# D

# LLCA Model estimation, parameters & z-values

In this appendix, estimation process is elaborated upon. The reasoning behind the number of classes chosen is given, as well as the model parameters.

## D.1. Model estimation

In order to determine how many classes are needed to to explain the data, the model is first estimated without the covariates, so the measurement model can be assessed in isolation. The models are estimated using the Latend Gold software(Statistical Innovations Inc., 2016). Table D.1 shows the fit of the model with 1 up until 10 clusters. Log-Likelihood (LL) measures how well the model reproduces the observed data. However, Log-Likelihood increases monotonically as adding more classes allows the model to reproduce the data with more precision. This does not mean that the model is better, because the complexity of the model is not taken into account. Therefore, the Bayesian Information Criterion (BIC) is used most to assess model fit. It uses the LL but penalises complexity. BIC is calculated using formula D.1:

$$\mathrm{BIC} = -2\,\mathrm{LL} + k\ln(n) \tag{D.1}$$

Where $n$ denotes the sample size and $k$ the number of parameters. The lowest BIC generally indicates the best fit. The results from table D.1 show that the model with 7 clusters performs the best. However, the increase in model fit is marginal compared to the 6 cluster model (only -2) for 29 extra parameters and a higher class error. This 7 cluster model is also harder to meaningfully interpret. Therefore, the 6 cluster model was found to be the best model.

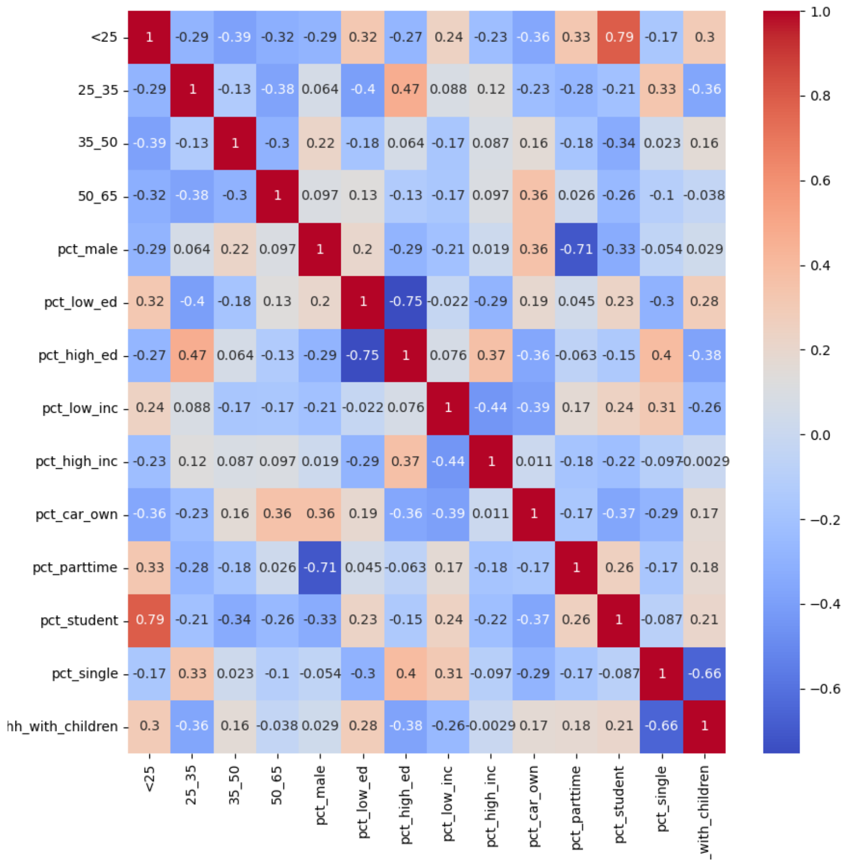
**Table D.1:** Model fit of the latent class cluster analysis for different cluster numbers

| Clusters | LL | BIC(LL) | AIC(LL) | AIC3(LL) | Npar | Max. BVR | Class.Err. | Entr. $R^2$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 12731,5440 | -25274,9218 | -25407,0880 | -25379,0880 | 28 | 521,7628 | 0,0000 | 1,0000 |
| 2 | 13452,8927 | -26522,7329 | -26791,7855 | -26734,7855 | 57 | 311,5581 | 0,0368 | 0,8597 |
| 3 | 14018,8430 | -27459,7470 | -27865,6859 | -27779,6859 | 86 | 182,9918 | 0,0645 | 0,8548 |
| 4 | 14306,8993 | -27840,9733 | -28383,7986 | -28268,7986 | 115 | 120,3556 | 0,0679 | 0,8716 |
| 5 | 14525,9134 | -28084,1152 | -28763,8269 | -28619,8269 | 144 | 91,8875 | 0,0757 | 0,8738 |
| 6 | 14711,6885 | -28260,7790 | -29077,3771 | -28904,3771 | 173 | 73,2746 | 0,0781 | 0,8781 |
| 7 | 14809,9621 | -28262,4396 | -29215,9241 | -29013,9241 | 202 | 63,4783 | 0,1002 | 0,8648 |
| 8 | 14906,2400 | -28260,1090 | -29350,4799 | -29119,4799 | 231 | 40,5976 | 0,0971 | 0,8737 |
| 9 | 14980,5688 | -28213,8804 | -29441,1377 | -29181,1377 | 260 | 38,7402 | 0,1034 | 0,8704 |
| 10 | 15072,8671 | -28203,5905 | -29567,7341 | -29278,7341 | 289 | 32,0534 | 0,1027 | 0,8764 |

One could also look at bivariate residuals. These residuals indicate whether there is a significant improvement in the model fit if the direct effect between two variables is included (if the local independence assumption is relaxed)(J. Vermunt & Magidson, 2005). In this study, however, the high BVRs primarily reflect correlations between the socio=demographic indicators themselves, rather than model misspecification. These variables are highly interdependent in the population, and such dependencies cannot be fully absorbed by adding more classes. Figure D.1 show the correlations of all the variables and table D.2 shows the Bivariate residuals of the six class model. Naturally, the percentage students and the percentage people aged under 25 are highly correlated. These variables also share a high BVR, but this cannot be relaxed by adding more classes. The same can be said for the percentage of parttimers and the percentage males, which have a high negative correlation. Other high BVR values are within indicator groups. Low- and high education share a high negative correlation because they are both percentages from the same sample, where a high percentage in one category by definition results in a low percentage in the other category. Therefore, the BVRs were used only to check whether additional classes would meaningfully improve the model, not as a strict criterion for model rejection. Comparing the 6- and 7-class solutions (tables D.2 and D.3) shows that only the residual between percentage high education and percentage low education decreases substantially (from 45.78 to 27.02). All other high bivariate residuals remain large, indicating that adding a seventh class does not meaningfully reduce the associations. This supports the interpretation that these high residuals largely reflect substantive correlations between indicators, rather than a lack of model fit, thus the 6 class model is selected as optimal.

## D.2. Correlations heatmap of the used indicators

**Figure D.1:** Correlation heatmap of the used indicators



## D.3. Bivariate residuals of the 6 and 7 class model

**Table D.2:** Bivariate Residuals of the 6-Class Model

| Indicators | <25 | 25–35 | 35–50 | 50–65 | pct_male | pct_low_ed | pct_high_ed | pct_low_inc | pct_high_inc | pct_car_own | pct_parttime | pct_student | pct_single | pct_hh_with_children |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **<25** | - | | | | | | | | | | | | | |
| **25–35** | 3.52 | - | | | | | | | | | | | | |
| **35–50** | 9.61 | 14.94 | - | | | | | | | | | | | |
| **50–65** | 9.68 | 25.25 | 38.60 | - | | | | | | | | | | |
| **pct_male** | 2.25 | 0.61 | 3.63 | 2.32 | - | | | | | | | | | |
| **pct_low_ed** | 2.27 | 2.00 | 5.02 | 2.90 | 4.78 | - | | | | | | | | |
| **pct_high_ed** | 2.22 | 3.84 | 2.89 | 1.29 | 7.77 | 45.78 | - | | | | | | | |
| **pct_low_inc** | 1.91 | 2.25 | 2.19 | 1.29 | 1.77 | 2.94 | 1.34 | - | | | | | | |
| **pct_high_inc** | 1.10 | 1.41 | 0.34 | 1.88 | 0.23 | 4.07 | 10.72 | 33.59 | - | | | | | |
| **pct_car_own** | 7.36 | 2.79 | 1.68 | 4.00 | 2.78 | 1.36 | 4.52 | 8.28 | 1.07 | - | | | | |
| **pct_parttime** | 3.05 | 2.20 | 1.11 | 2.33 | 73.27 | 1.96 | 0.77 | 4.45 | 3.95 | 1.15 | - | | | |
| **pct_student** | 46.14 | 0.79 | 4.56 | 1.55 | 2.74 | 0.54 | 0.82 | 3.66 | 3.28 | 3.68 | 4.51 | - | | |
| **pct_single** | 0.66 | 0.89 | 1.94 | 0.66 | 1.48 | 1.19 | 2.65 | 9.06 | 7.13 | 3.83 | 0.84 | 1.62 | - | |
| **pct_hh_with_children** | 3.78 | 3.68 | 13.04 | 5.15 | 1.79 | 1.40 | 2.12 | 8.77 | 3.50 | 1.63 | 0.33 | 1.69 | 46.78 | - |

**Table D.3:** Bivariate Residuals of the 7-Class Model

| Indicators | <25 | 25–35 | 35–50 | 50–65 | pct_male | pct_low_ed | pct_high_ed | pct_low_inc | pct_high_inc | pct_car_own | pct_parttime | pct_student | pct_single | pct_hh_with_children |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **<25** | - | | | | | | | | | | | | | |
| **25–35** | 2.65 | - | | | | | | | | | | | | |
| **35–50** | 9.40 | 13.63 | - | | | | | | | | | | | |
| **50–65** | 8.47 | 22.37 | 28.56 | - | | | | | | | | | | |
| **pct_male** | 1.55 | 0.92 | 3.48 | 2.29 | - | | | | | | | | | |
| **pct_low_ed** | 1.38 | 1.46 | 4.39 | 1.97 | 1.64 | - | | | | | | | | |
| **pct_high_ed** | 1.43 | 3.04 | 2.12 | 1.02 | 3.31 | 27.02 | - | | | | | | | |
| **pct_low_inc** | 1.64 | 1.99 | 3.33 | 1.61 | 2.19 | 2.43 | 1.29 | - | | | | | | |
| **pct_high_inc** | 1.52 | 1.34 | 0.75 | 1.86 | 0.55 | 4.59 | 7.93 | 28.40 | - | | | | | |
| **pct_car_own** | 7.04 | 2.76 | 1.45 | 3.08 | 1.88 | 1.82 | 3.80 | 7.03 | 0.87 | - | | | | |
| **pct_parttime** | 1.84 | 2.58 | 1.04 | 2.89 | 63.48 | 0.89 | 0.98 | 3.69 | 3.92 | 1.22 | - | | | |
| **pct_student** | 43.90 | 0.78 | 4.55 | 1.81 | 2.08 | 0.71 | 1.05 | 2.96 | 2.99 | 4.03 | 3.50 | - | | |
| **pct_single** | 0.98 | 0.66 | 1.38 | 0.91 | 1.19 | 1.16 | 2.01 | 7.75 | 6.61 | 3.44 | 1.12 | 1.20 | - | |
| **pct_hh_with_children** | 3.63 | 3.27 | 10.85 | 4.84 | 1.88 | 1.21 | 2.45 | 7.80 | 3.48 | 1.70 | 0.26 | 1.59 | 40.94 | - |

Now that the measurement part of the model is established, the model can be expanded with the structural part. For the structural part, the model is expanded with an accessibility metric. The results are further interpreted in chapter 5, where the actual distributions of indicators and covariates are also given. No parameters are estimated for inactive covariates, which explains why these are not included in the table.

**Table D.4:** Parameters and z-values of the estimated latent class cluster analysis model with covariate

| **Predictions for the indicators** | **Cl1** | **z-value** | **Cl2** | **z-value** | **Cl3** | **z-value** | **Cl4** | **z-value** | **Cl5** | **z-value** | **Cl6** | **z-value** | **Wald** | **p-value** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Age | | | | | | | | | | | | | | |
| •<25 | 0,0121 | 4,2871 | -0,0368 | -10,0953 | -0,0543 | -18,4172 | 0,1028 | 17,6892 | 0,0256 | 6,4815 | -0,0495 | -14,1175 | 716,6783 | 1,2e-152 |
| •25_35 | -0,0459 | -9,7370 | -0,0098 | -1,8112 | -0,0141 | -2,8707 | -0,0663 | -12,2728 | 0,0385 | 7,5588 | 0,0975 | 12,4438 | 411,8400 | 8,3e-87 |
| •35_50 | 0,0028 | 0,5568 | 0,0093 | 1,4794 | 0,0289 | 5,2034 | -0,0453 | -6,9158 | -0,0237 | -4,3393 | 0,0279 | 4,5600 | 98,7606 | 9,6e-20 |
| •50_65 | 0,0133 | 2,6003 | 0,0389 | 5,6943 | 0,0529 | 8,1930 | -0,0172 | -2,4495 | -0,0384 | -6,8313 | -0,0495 | -6,7275 | 194,7011 | 3,9e-40 |
| Gender | | | | | | | | | | | | | | |
| •pct_male | 0,0150 | 1,7432 | -0,0727 | -5,2648 | 0,1908 | 24,4019 | -0,0977 | -8,4947 | -0,0402 | -4,4294 | 0,0048 | 0,3936 | 635,6083 | 4,1e-135 |
| Education | | | | | | | | | | | | | | |
| •pct_low_ed | 0,0592 | 13,5791 | -0,0531 | -13,4459 | 0,0408 | 7,6694 | 0,0653 | 10,2023 | -0,0225 | -4,4770 | -0,0896 | -20,1879 | 882,2878 | 1,8e-188 |
| •pct_high_ed | -0,1332 | -20,7663 | 0,0941 | 11,1081 | -0,1222 | -13,2018 | -0,1194 | -12,7862 | 0,0732 | 8,3241 | 0,2075 | 21,3344 | 1295,5312 | 5,9e-278 |
| Income | | | | | | | | | | | | | | |
| •pct_low_inc | -0,0116 | -3,1880 | -0,0185 | -4,3529 | -0,0332 | -8,0934 | 0,0097 | 1,6955 | 0,0595 | 9,3061 | -0,0060 | -0,9485 | 135,8351 | 1,4e-27 |
| •pct_high_inc | -0,0361 | -6,2302 | 0,0456 | 6,3503 | -0,0044 | -0,5527 | -0,0371 | -4,6076 | -0,0427 | -5,6319 | 0,0747 | 7,9774 | 200,3380 | 2,4e-41 |
| Car ownership | | | | | | | | | | | | | | |
| •pct_car_own | 0,0573 | 9,9500 | 0,0300 | 4,3271 | 0,1334 | 15,5664 | -0,0393 | -4,4290 | -0,0941 | -11,4135 | -0,0873 | -6,2673 | 447,0447 | 2,1e-94 |
| Social participaction | | | | | | | | | | | | | | |
| •pct_parttime | 0,0224 | 3,6483 | 0,0341 | 4,0402 | -0,0753 | -15,6284 | 0,0831 | 9,9231 | 0,0025 | 0,4278 | -0,0668 | -12,1984 | 489,6907 | 1,3e-103 |
| •pct_student | 0,0039 | 1,7716 | -0,0239 | -8,6012 | -0,0413 | -22,8965 | 0,0655 | 14,3670 | 0,0270 | 7,9581 | -0,0312 | -10,8570 | 705,7453 | 2,8e-150 |
| Household configuration | | | | | | | | | | | | | | |
| •pct_single | -0,0372 | -9,8364 | -0,0039 | -0,8640 | -0,0115 | -2,4300 | -0,0472 | -8,6389 | 0,0613 | 10,9781 | 0,0385 | 5,9272 | 289,9416 | 1,5e-60 |
| •pct_hh_with_children | 0,0395 | 7,6698 | -0,0027 | -0,4539 | 0,0111 | 1,9063 | 0,0803 | 10,2380 | -0,0650 | -10,4805 | -0,0632 | -7,8013 | 297,4229 | 3,6e-62 |

| **Intercepts** | **Overall** | **z-value** | **Wald** | **p-value** |
|---|---|---|---|---|
| Age | | | | |
| •<25 | 0,1249 | 65,3661 | 4272,7324 | 1,9e-930 |
| •25_35 | 0,2291 | 89,1238 | 7943,0454 | 1,4e-1727 |
| •35_50 | 0,2968 | 122,9324 | 15112,3797 | 1,6e-3284 |
| •50_65 | 0,2933 | 106,9193 | 11431,7308 | 3,2e-2485 |
| Gender | | | | |
| •pct_male | 0,5615 | 126,3424 | 15962,3944 | 4,1e-3469 |
| Education | | | | |
| •pct_low_ed | 0,1397 | 65,4411 | 4282,5343 | 1,4e-932 |
| •pct_high_ed | 0,4985 | 127,8863 | 16354,9151 | 2,3e-3554 |
| Income | | | | |
| •pct_low_inc | 0,1313 | 60,0275 | 3603,3009 | 4,7e-785 |
| •pct_high_inc | 0,4766 | 148,0729 | 21925,5735 | 4,5e-4764 |
| Car ownership | | | | |
| •pct_car_own | 0,5168 | 128,3614 | 16476,6479 | 8,6e-3581 |
| Social participaction | | | | |
| •pct_parttime | 0,1821 | 62,6199 | 3921,2536 | 4,1e-854 |
| •pct_student | 0,0641 | 45,0302 | 2027,7211 | 8,6e-443 |
| Household configuration | | | | |
| •pct_single | 0,1890 | 82,6436 | 6829,9683 | 7,5e-1486 |
| •pct_hh_with_children | 0,5162 | 175,2523 | 30713,3726 | 2,2e-6672 |

| **Prediction Latent class membership** | **Cluster1** | **z-value** | **Cluster2** | **z-value** | **Cluster3** | **z-value** | **Cluster4** | **z-value** | **Cluster5** | **z-value** | **Cluster6** | **z-value** | **Wald** | **p-value** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Intercept | 2,3604 | 11,4234 | 0,3210 | 1,2988 | 1,8634 | 8,1570 | 0,8723 | 3,6125 | -1,7339 | -6,3703 | -3,6832 | -5,5347 | 177,4713 | 1,9e-36 |
| **Covariates** | | | | | | | | | | | | | | |
| •Hansen_accessibility | -6,0435 | -7,8339 | 0,9627 | 1,5393 | -7,4346 | -7,1085 | -2,2191 | -2,9124 | 6,0521 | 8,9187 | 8,6825 | 6,7227 | 144,8421 | 1,7e-29 |

# E

# Comparing and joining years 2018-2023

This section explains how the decision was made to include all years in the analysis. The different years need to be compared and evaluated to decide on whether they have enough resemblance to use them together. Table E.1 shows the division of the respondents over the years. Interesting is the shift in total responses compared to table 4.1, which showed that 2021 and 2022 had the most responses. Due to COVID, many people did not travel for workKiM Netherlands Institute for Transport Policy Analysis, 2022 which eliminates relatively more respondents from the dataset.

**Table E.1:** Amount of respondents with commutes over the years

| Year | Respondents |
|---|---|
| 2018 | 15902 |
| 2019 | 14653 |
| 2020 | 11689 |
| 2021 | 12506 |
| 2022 | 13357 |
| 2023 | 14538 |

For the complete dataset, the modal split can now be calculated to see if COVID impacted the results. The categorisation explained in 4.3.2 and allows for a classification of each movement as one of four categories which results in the modal split shown in figure E.1.

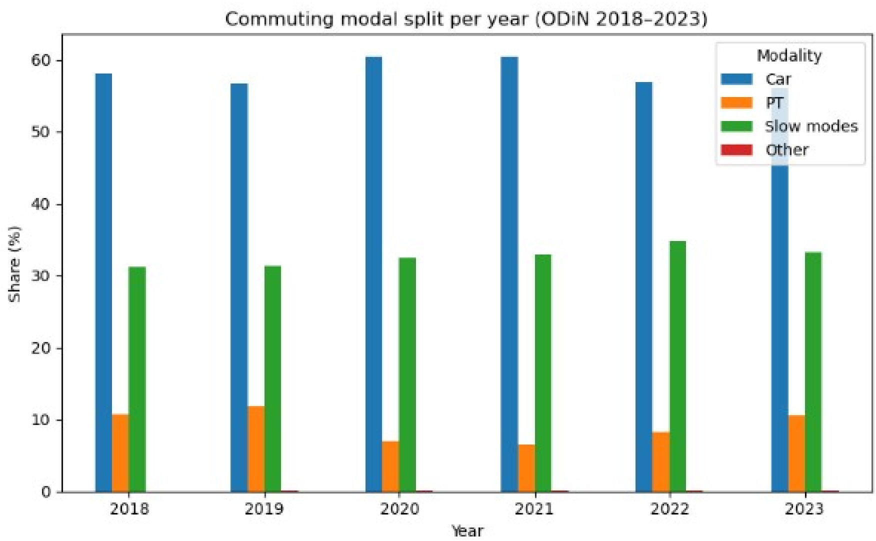



**Figure E.1:** Modal split for each year

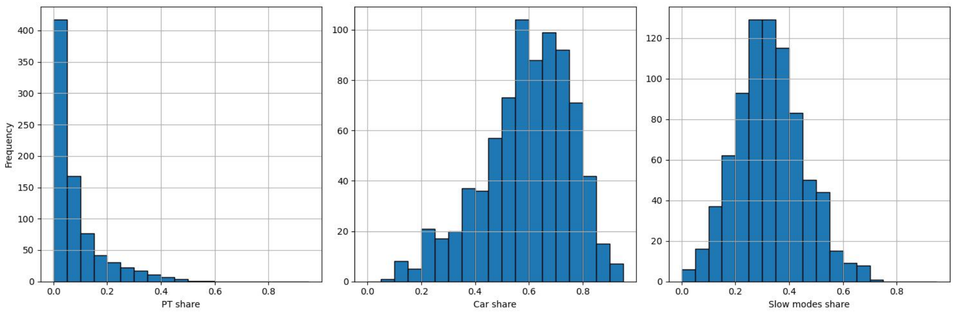
The category "Other" is plotted here, but since it is not useful for the further research and the respondents within this category are limited, these are dropped from the dataset. The COVID years have slightly lower public transport usage and increased car usage. Contrary to the results from the KiMKiM Netherlands Institute for Transport Policy Analysis, 2022, COVID does seem to affect the public transport and car percentages. This gives a couple of options with the usage of the years: The first option is to acknowledge the fact that COVID impacted commuter behaviour and remove the COVID years from the dataset. COVID lasted from 2020 to early 2022, so 2020 and 2021 would have to be removed at least. 2022 shows that the shares are returning to their original values, but are not there yet. However, this cuts severely in the amount of respondents. Option two would be to use all the years despite the slight differences in the modal split, and keep the higher amount of respondents, as it improves spatial accuracy.

**Table E.2:** Comparison of options of using years together

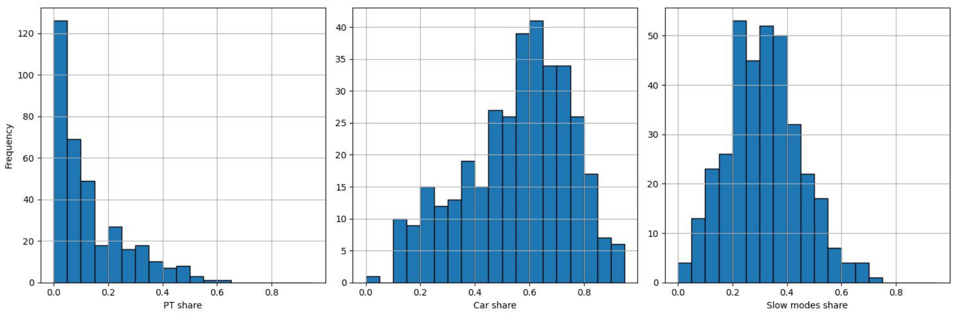
| Options | Use all years | Remove 2020, 2021 | Remove 2020, 2021, 2022 |
|---|---|---|---|
| Mean | | | |
| Car share | 58.15 | 56.96 | 56.99 |
| PT share | 9.16 | 10.36 | 11.04 |
| Slow modes share | 32.70 | 32.68 | 31.97 |
| Standard deviation (mean) | 1.84 | 1.35 | 0.98 |
| Car share | 1.96 | 0.83 | 1.01 |
| PT share | 2.20 | 1.49 | 0.75 |
| Slow modes share | 1.35 | 1.73 | 1.17 |
| Postcodes (of 4070) with more than 30 respondents | 797 | 529 | 311 |

From table E.2, it can be concluded that the modal split does shift a bit by taking all years, but the gain in amount of postcodes is also quite substantial. Therefore, it is more relevant to check how the spatial distribution of the modal split changes. If the public transport share is consistently underestimated, spatial differences will stay intact. Therefore, the national modal split plots of all year option and the 3 year option are compared. These plots are shown in figure E.4. From visual inspection, it can be concluded that the regions with high mode share percentages remain roughly the same in the exception of some specific cases. However, the most notable difference is in the amount of respondents and areas with enough respondents. Additionally, figure E.2 highlights the relative occurrence of certain modal shares within the relevant postcodes. Even though the plots have different magnitudes, the distribution is fairly similar. Based on the info here and the similarities of the plots, the decision is made to continue with the information of all six years since the results become more statistically significant since the mean number of respondents per postcode increases, and more than double the postcodes can be used.

**Figure E.2:** Histogram comparison for each of the shares




**(a)** Distribution of mode share when 6 years of ODiN data is used

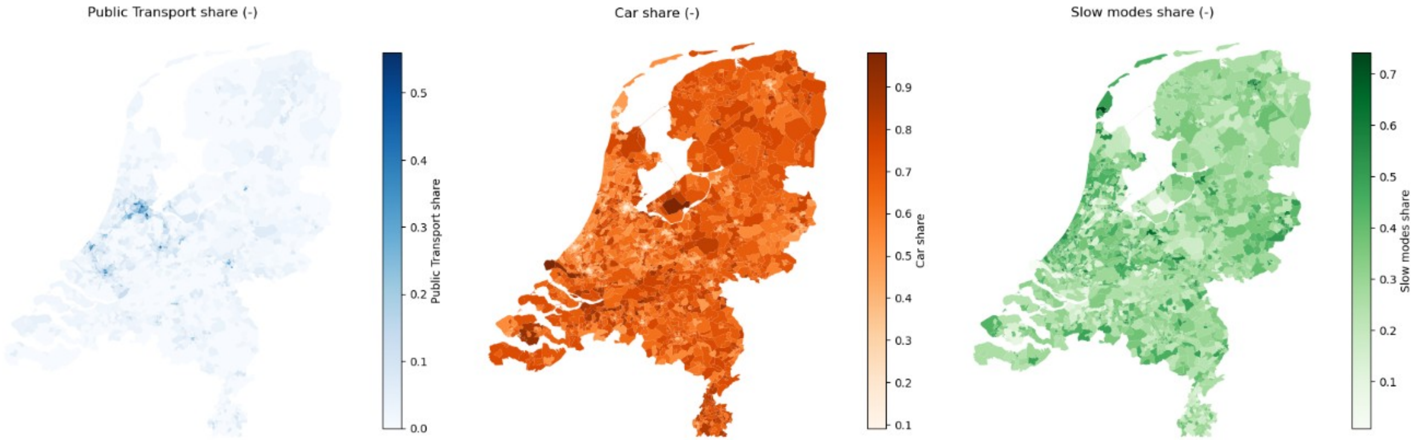



**(b)** Distribution of mode share when 3 years of ODiN data is used

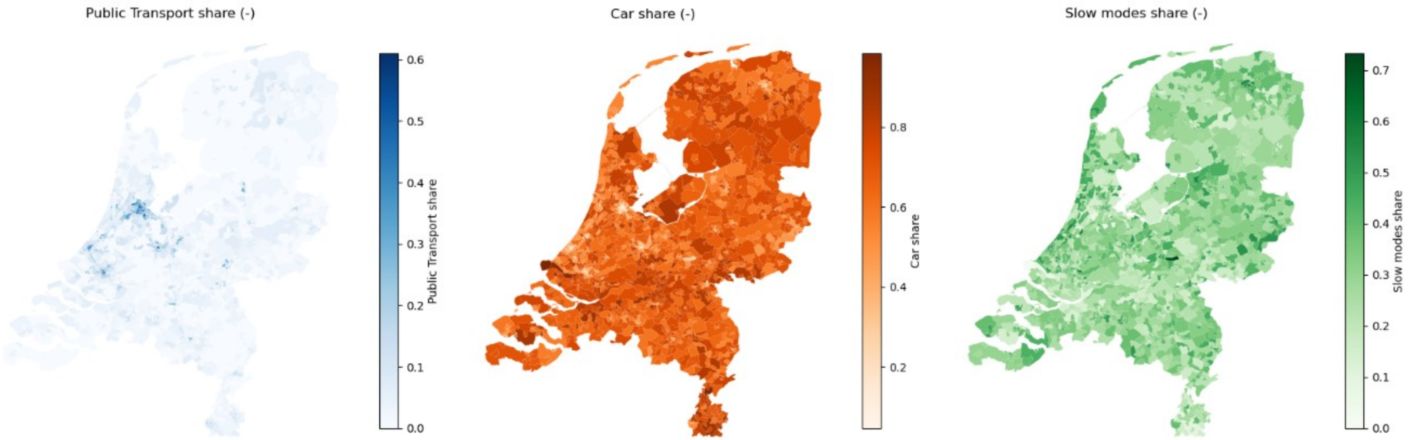
**Table E.3:** types of vehicles and their classification

| ODiN Coding | Transport mode | Classification | Reasoning (special cases) |
|---|---|---|---|
| 1 | Car | Car | |
| 2 | Train | Public Transport | |
| 3 | Bus | Public Transport | |
| 4 | Tram | Public Transport | |
| 5 | Metro | Public Transport | |
| 6 | Speedpedelec | Slow modes | Has a max speed of 25 km/h and needs to use the bike lane |
| 7 | E-bike | Slow modes | |
| 8 | Bike | Slow modes | |
| 9 | Foot | Slow modes | |
| 10 | Touringcar | Public Transport | Collective transport often used for long distance shuttling |
| 11 | Delivery van | Car | |
| 12 | Lorry | Car | |
| 13 | Camper | Car | |
| 14 | Taxi/Taxivan | Public Transport | Partly individual, but collectively organised and often paid and shared transport |
| 15 | Agricultural vehicle | Other | Very specific, not standard for average commuter |
| 16 | Motorbike | Car | Can drive on the road, high speed |
| 17 | Bromfiets | Car | Has a max speed of 45 km/h and more similarities to car then slow modes |
| 18 | Snorfiets | Slow modes | Has a max speed of 25 km/h and needs to use the bike lane |
| 19 | Motorised disability vehicle | Slow modes | Even motorised, uses the cycling roads or pavements and low speed |
| 20 | Un-motorised disability vehicle | Slow modes | |
| 21 | Skates/skeelers/step | Slow modes | |
| 22 | Boot | Other | Not road transport, very specific to certain regions |
| 23 | Other motorised vehicle | Other | Unclear what this can be, so should not be taken into account |
| 24 | Other un-motorised vehicle | Other | Unclear what this can be, so should not be taken into account |

**Figure E.3:** Comparison of six and three year modal split using the merging algorithm

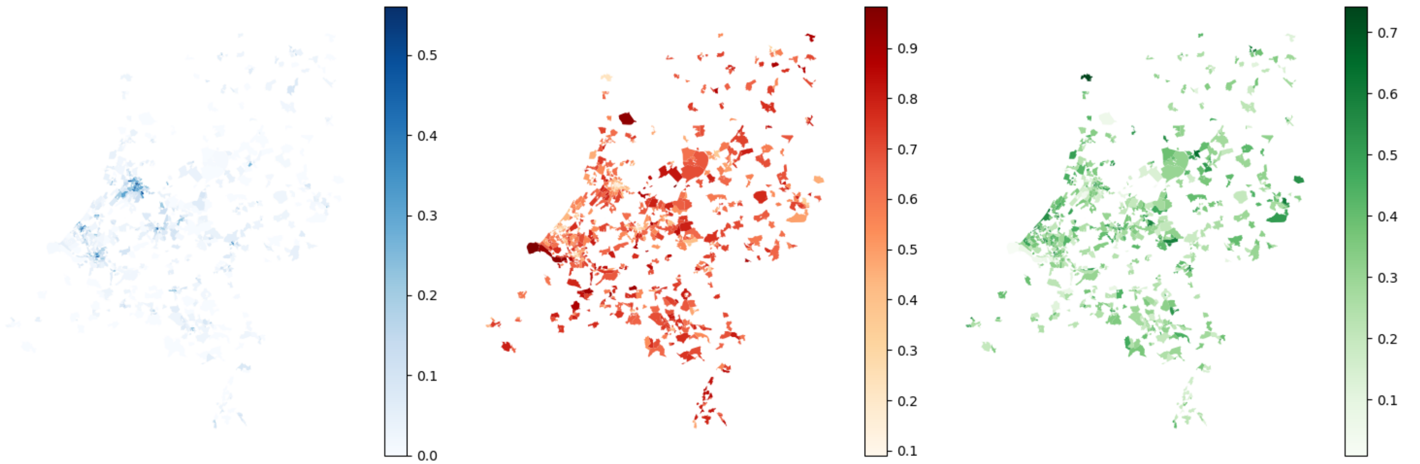



**(a)** modal split based on all six years

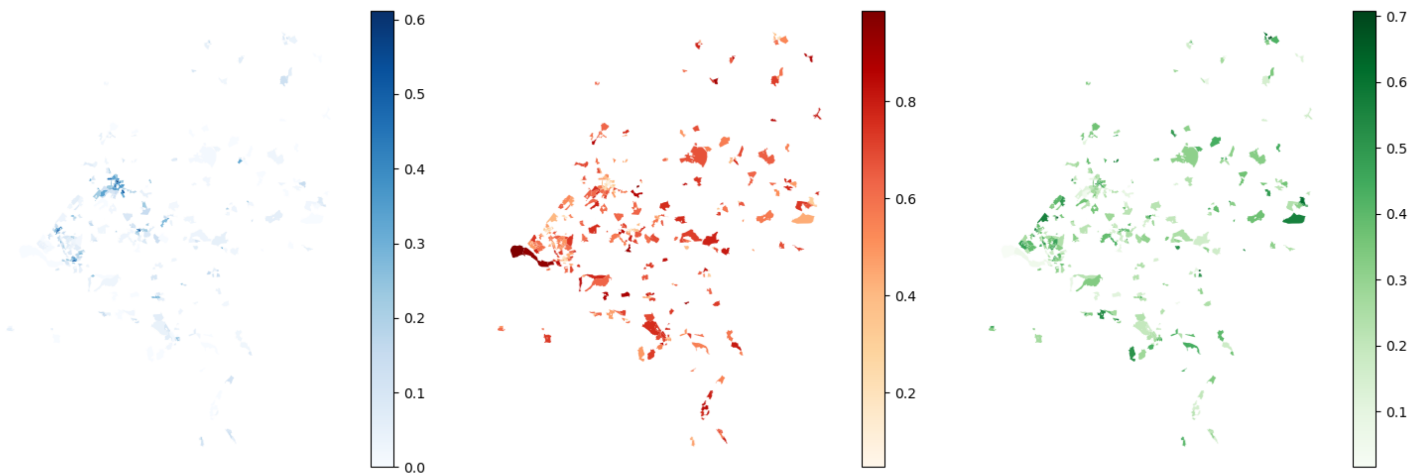



**(b)** Modal split based on 2018, 2019 and 2023

**Figure E.4:** Comparison of six and three year modal split looking at only statistically relevant postcodes




**(a)** modal split based on all six years




**(b)** Modal split based on 2018, 2019 and 2023

# F

# cluster 3 postcode groups

**Table F.1:** Functional classification of Cluster 3 postcode areas

| **Subtype** | **PC4 codes** | **Functional motivation** |
|---|---|---|
| Port of Rotterdam | 3196, 3197, 3198, 3199, 3087, 3088, 3089 | Core port and industrial areas of the Port of Rotterdam, characterised by large-scale logistics, petrochemical industry, and limited residential presence. These areas exhibit strong infrastructural constraints and highly specialised labour demand. |
| Port of Amsterdam (Westpoort) | 1042, 1046, 1047 | Major seaport and industrial zone in the Amsterdam metropolitan area, hosting logistics, energy and bulk activities. Although embedded in an urban region, the area itself is mono-functional and workplace-oriented. |
| Moerdijk Seaport | 4751, 4782 | Industrial seaport with strong logistics and chemical activities, positioned between Rotterdam and Antwerp. Functions as a peripheral but highly specialised employment location. |
| Zealand Seaports | 4389, 4538, 4542 | Seaport and industrial zones in Zeeland, characterised by chemical industry, logistics and port-related employment, with relatively low surrounding population density. |
| Groningen Seaport (Eemshaven) | 9936 | Northern Dutch seaport focused on energy, data centres and heavy industry. Spatially isolated location with limited public transport accessibility. |
| Inland Ports | 3313, 4202, 4906, 5145, 5349, 5928, 5807, 8013, 8028, 7418 | Inland port and multimodal terminal locations along major rivers and canals (e.g. Maas, Waal, IJssel). These areas lack direct sea access but fulfil an important logistics and industrial function, often located outside major urban cores. |
| Airports | 1117, 1118, 5657, 6199, 9761, 3045, 8218 | Airport-related employment zones, including Schiphol, Eindhoven Airport and regional airports. Characterised by large-scale infrastructure, shift-based work, and specific accessibility requirements. |
| Heavy Industry | 6167, 1951, 4612, 4631 | Large industrial complexes such as Chemelot and Tata Steel, dominated by heavy manufacturing and processing industries, often spatially segregated from residential areas. |
| Cluster 3 (Other) | Remaining Cluster 3 PC4s | Postcodes within Cluster 3 that do not fall into one of the predefined functional categories but share similar characteristics, such as low residential density and specialised employment structures. |

# G

# Additional regression results

**Table G.1:** Regression Results: Determinants of Market Tightness in the Rijnmond Region (N=203)

| Variable | $(M1_a)$ Hansen | $(M1_b)$ Competitive | $(M2)$ Baseline | $(M3)$ Interaction | $(M4)$ Education | $(M5)$ Placebound |
|---|---|---|---|---|---|---|
| Constant | 0.1304*** | 0.0957*** | 0.1199*** | 0.1295*** | 0.0730*** | 0.1228*** |
| | (0.0072) | (0.0179) | (0.0106) | (0.0135) | (0.0280) | (0.0127) |
| log_hansen_c | 0.0277*** | | 0.0132 | -0.0079 | 0.0070 | 0.0123 |
| | (0.0082) | | (0.0105) | (0.0174) | (0.0113) | (0.0103) |
| Competitive_acc. | | 0.0208** | | | | |
| | | (0.0083) | | | | |
| density_c | | | 0.0000 | 0.0000* | 0.0000 | 0.0000 |
| | | | (0.0000) | (0.0000) | (0.0000) | (0.0000) |
| banen | | | 0.0000* | 0.0000** | 0.0000 | 0.0000* |
| | | | (0.0000) | (0.0000) | (0.0000) | (0.0000) |
| interaction_c | | | | -0.0000 | | |
| | | | | (0.0000) | | |
| fract_high_ed | | | | | 0.1317 | |
| | | | | | (0.0855) | |
| fract_placebound | | | | | | -0.0263 |
| | | | | | | (0.0550) |
| R-squared | 0.0437 | 0.0290 | 0.0642 | 0.0728 | 0.1051 | 0.0652 |
| Adj. R-squared | 0.0389 | 0.0242 | 0.0501 | 0.0541 | 0.0871 | 0.0464 |
| N | 203 | 203 | 203 | 203 | 203 | 203 |

*Note:* Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

# H
# Alternative merging areas on urbanity

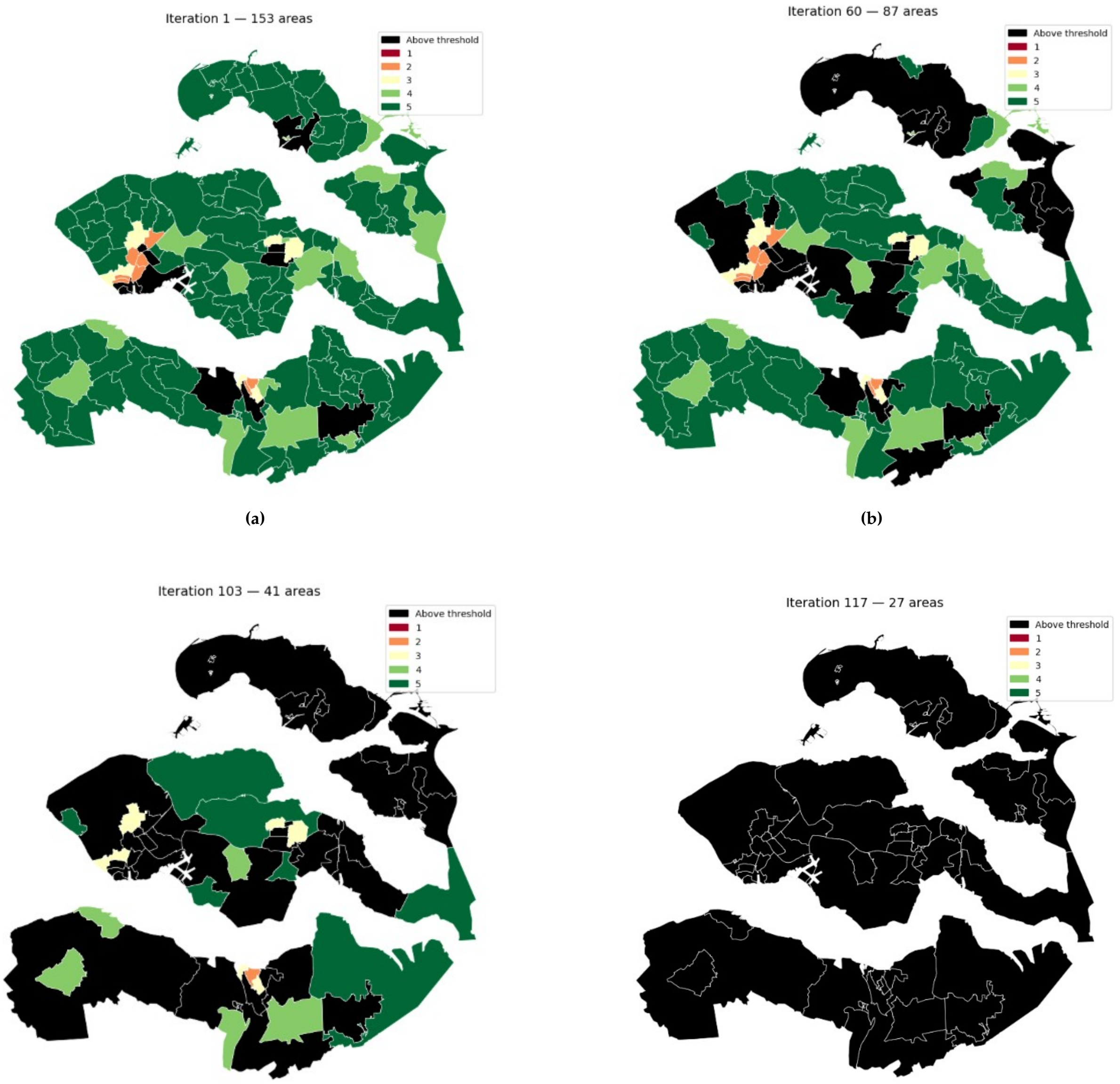
To ensure reliable statistical estimates across spatial units, it is necessary that each area contains a sufficient number of survey respondents. However, at the Dutch PC4 level, many areas have too few respondents for stable analysis. Therefore, an iterative merging algorithm was developed to aggregate postcode areas until all units reach a predefined respondent threshold. The algorithm takes as input a geodataframe containing postcode polygons and other attributes (e.g. urbanity), and the ODiN dataset grouped by **ArrPC**. The output is a new geodataframe with merged polygons and a merged ODiN dataset in which all respondents are assigned to their new area with enough respondents to get statistically significant outcomes. The algorithm iteratively merges areas until every area is above the threshold of 30 respondents.

In an iteration, the following steps are taken:

1. Identify areas under threshold
2. Select the area with lowest n
3. Retrieve key attributes of selected area (postcode, n, town, municipality, degree of urbanisation and geometry)
4. Identify candidate neighbours in the following order: (1) Same town and **urb**, (2) same municipality and **urb**, (3) nearest neighbour
5. Evaluate possible merging combinations and choosing the combination that just exceeds threshold and uses smallest number of areas
6. Merge the areas (geometric data and ODiN data) and validate the shapes
7. Go back to step 1

The algorithm can be constructed with multiple strategies. This merging strategy was chosen since commuting modal split is highly dependent on the urbanity of the area (Næss & Lyssand Sandberg, 1996). The downside to this is that sometimes, areas that are not adjacent or even multiple kilometres apart are used for a merge. However, adding data from a small village to a small city will be less realistic, even though these might be spatially closer together. This is visible in the visualisation of an example merge in Zeeland in figure H.1 and H.2. The algorithm merges the areas with postcodes 4697, 4341 and 4401 (dark purple). These are areas which fall under urbanity class 4 where many surrounding postcodes fall under urbanity class 5. Therefore, these areas are considered together even though they might be relatively far apart. Some areas need many postcodes to be merged, which is to be expected since these areas do not have that many respondents. These areas are generally areas with urbanity class 5 where less people live and work, so less accuracy here cannot be avoided. Additionally, the other datasets have better spatial coverage, so for these analyses areas do not have to be merged.

**Figure H.1:** Visualisation of the merging process with the province Zeeland as an example




**(a)** **(b)**

**(c)** **(d)**

**Figure H.2:** Final result of the merging process for the province Zeeland.