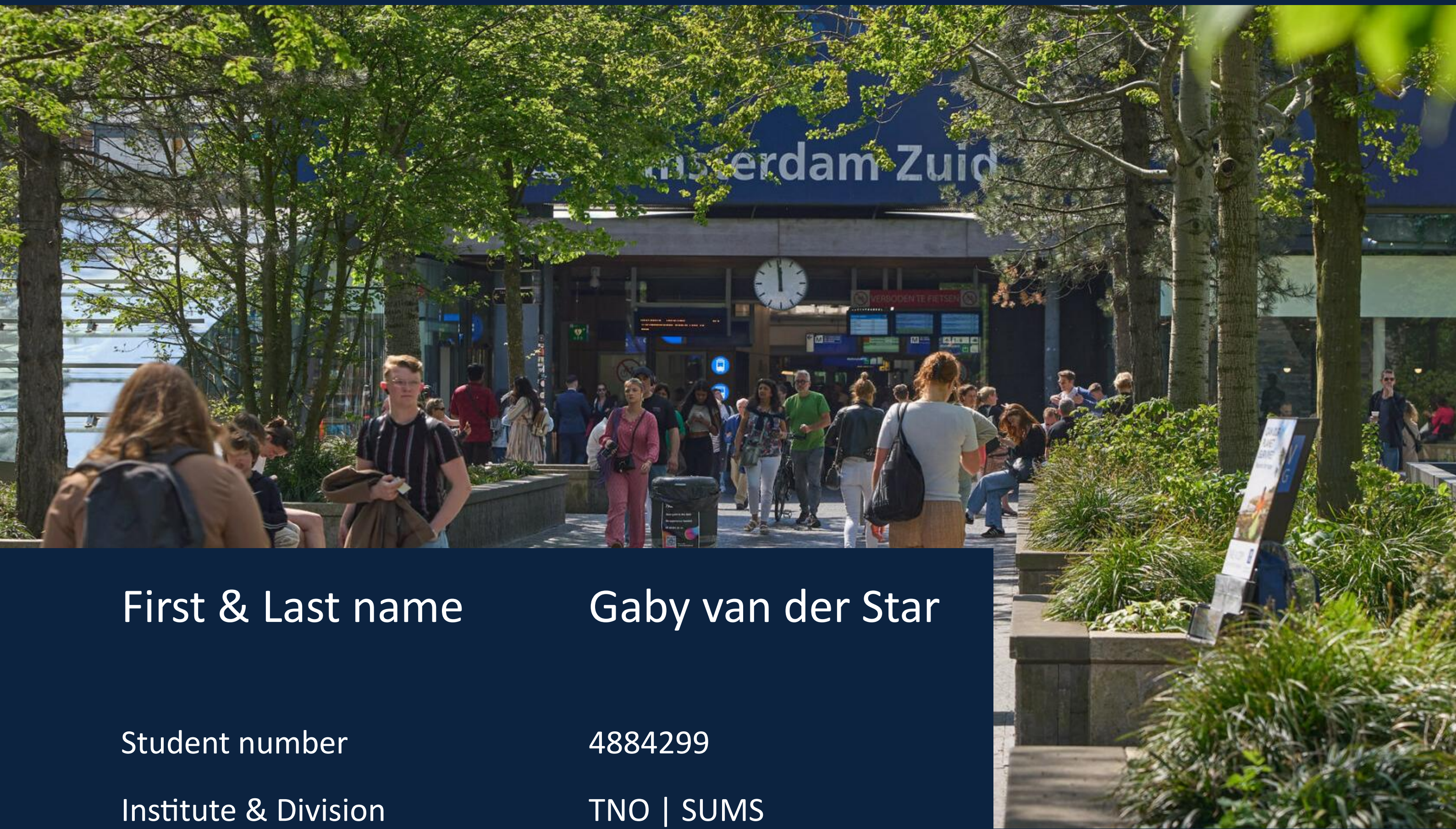
Building-level Pedestrian trip generation for Dutch urban areas using open data

Master Thesis

Date: 24/11/2025



First & Last name: Gaby van der Star

Student number: 4884299

Institute & Division: TNO | SUMS

Daily supervisor: Ali Nadi Najafabadi (TNO), Dorine Duives (TU)

Committee: Maaike Snelder, Winnie Daamen

TU Delft
Delft University of Technology

TNO innovation for life

xcarcity

Preface

This thesis is submitted as the final requirement for obtaining my Master's degree in Civil Engineering, track Transport and Traffic Engineering, at Delft University of Technology. I was fortunate to carry out my graduation project at TNO. This opportunity arose after visiting TNO during one of my courses, where I became inspired by their work on digital twins—an area that connects well with my interests in modelling and sustainability. After reaching out to my lecturer, she kindly introduced me to TNO, which led to this thesis project.

I would first like to express my sincere gratitude to my chair, **Maaike Snelder**, for facilitating this connection and making the internship possible. I am also deeply thankful to my daily supervisor at TNO, Ali Nadi Najafabadi (, whose weekly guidance on data and technical aspects of modelling was invaluable. At TU Delft, my daily supervisor **Dorine Duives** provided essential support on theoretical matters, while **Winnie Daamen** contributed greatly by helping me improve the quality of my writing and ensuring the overall clarity of this report. I truly appreciate the time and effort my committee dedicated to meetings, discussions, and providing valuable feedback throughout the process.

Furthermore, I am grateful to the TNO colleagues who made sure interns felt connected and involved, as well as to the other interns and fellow TU Delft students with whom I shared many study sessions. Defining a manageable scope within the given timeframe was challenging, and I learned that coding often takes longer than expected—especially when applying new methodologies. At the same time, I gained valuable insights into modelling, particularly in understanding pedestrian behaviour and validating patterns in data and outputs.

Overall, I had a very rewarding experience at TNO, both through the guidance of my daily supervisor and the collaborative atmosphere with the other interns.

Delft, November 2025

Abstract

Cities are increasingly pressured by rapid urbanisation, climate change, and congested transport networks, amplifying the need for sustainable mobility solutions that prioritise walking and cycling. Designing effective pedestrian-friendly interventions requires reliable, city-scale models of pedestrian activity. However, most existing transport models remain car-oriented or focus on microscopic pedestrian dynamics. Macroscopic pedestrian models that do exist are often highly data-intensive, depend on household surveys or extensive counts, and are not suited to the Dutch context, where such data are limited. This creates a clear gap: the need for a practical, open-data-based, building-level pedestrian trip generation model for Dutch cities.

To address this gap, the objective of this study is to develop a building-level pedestrian trip generation model with fine spatial and temporal resolution, tailored to Dutch conditions and relying exclusively on widely available open data. This leads to the central research question: *To what extent can an existing pedestrian trip-generation model be adapted to reflect pedestrian trip-generation dynamics in the Dutch urban context?*

To answer this question, the study adapts Sevtsuk's Urban Network Analysis (UNA) framework into a Dutch-specific, building-level model referred to as BPT-Gen (building level pedestrian trip generation). This framework was selected because it offers a practical balance between spatial detail and data requirements.

The first major adaptation involves identifying and classifying buildings using the Dutch datasets BAG (for buildings) and OSM (for land use, amenities, and public transport).

The second key modification concerns the derivation of building activity weights, defined as the number of unique daily users per building. hese weights are calculated using detailed study-area data where available; otherwise, Dutch building and occupancy standards or proxy indicators are applied.

Building on the UNA framework, the model then incorporates accessibility-adjusted weights. Each building is connected to the pedestrian network, and destination-specific Reach indices are calculated using a Dutch-relevant walking radius. These indices adjust the baseline activity weights to reflect behavioural tendencies and destination attractiveness. Sensitivity analyses are conducted to examine how choices related to walking radius and normalisation methods influence these adjustments.

A major extension to the original UNA model is the integration of hourly walking trip rates. Trip purposes from ODiN are mapped to building types, and hourly origin–destination trip rates are applied accordingly. Combining these rates with the accessibility-adjusted weights produces hourly, building-level pedestrian trip generation for the study area.

Finally, the model's performance is assessed through face-validation using observed pedestrian counts, providing an initial indication of how well the adapted framework captures real-world patterns of walking activity.

The results from the face validation show that the adapted model successfully reproduces realistic spatial and temporal patterns for building types supported by reliable activity data. Particularly housing, healthcare facilities, and major train stations. Underprediction is observed for offices, leisure facilities, and bus stops, where weights depend on proxy indicators and where walking trips are likely underreported in ODiN. Sensitivity analyses confirm that trip-generation

outcomes depend strongly on proxy selection and accessibility assumptions, revealing important structural uncertainties.

Despite these uncertainties, the model provides clear practical value: it identifies major generators of walking trips, peak hours, and accessibility-driven hotspots, even where some building types rely on imperfect proxies. By linking trip purposes to building types at hourly resolution using only open data, BPT-Gen fills a gap in existing literature and offers a transparent, reproducible foundation for Dutch pedestrian modelling.

The findings imply that open-data pedestrian models can offer meaningful insights without extensive calibration datasets. The BPT-Gen framework enables planners to explore how walking activity responds to land-use changes or network modifications and supports accessibility-oriented planning approaches. At the same time, the study highlights the need for improved activity indicators, better representation of workplace and leisure walking, and validation using street-level pedestrian flow data once trip distribution and assignment are incorporated.

Table of Contents

BPT-Gen Moodel....1

Preface i

Abstract ii

Table of Contents iv

1. Introduction 1

2. Literature Review on Pedestrian Models 3

2.1 Overview of Existing Pedestrian Models....3

2.2 Comparative Analysis of Pedestrian Models 3

2.2.1 Model Performance....4

2.2.2 Behavioural Validity....4

2.2.3 Spatial and Temporal Resolution....4

2.2.4 Applicability to the Dutch Context....5

2.2.5 Data Demands....5

2.3 Reflection on the Comparison of Pedestrian Demand Modelling 7

3. Research Methodology....8

3.1 General BPT-Gen Model Outline....8

3.1.1 Explanation of the UNA Tool....8

3.1.2 Explanation of the Adaptation of the UNA Structure in BPT-Gen 9

3.2 Using Dutch Open Data for Model Development....11

3.2.1 Scraping Building, Network and Household Survey Data....11

3.2.1.1 OSM scraping for Building Information....11

3.2.1.2 BAG Scraping for Building Information 12

3.2.1.3 OSM scraping for pedestrian network 13

3.2.1.4 ODiN scraping for Trip Rate Data 13

3.2.2 Mapping Trip Purpose (ODiN) to Building Type 14

3.2.2.1 Trip Purpose to Building Type....14

3.2.2.2 Main Mode of Transport to Building Type....15

3.2.2.3 Mapping Building Function to Building Type 16

3.2.3 Assigning Building Weights 17

3.2.3.1 Weight computation using study area specific data 17

3.2.3.2 Weights Based on Dutch Building Standards 19

3.2.3.3 Proxy Weights....20

3.2.3.4 Public Transport Weights 20

3.2.3.5 Model Sensitivity: Building Weights 20

3.2.3.5.1 Alternative Office Weights 20
3.2.3.5.2 Alternative Catering Weights...21
3.2.3.5.3 Alternative Public Transport Weights 21
3.2.4 Elastic Building Weights 22
3.2.4.1 Inserting buildings into the pedestrian network 22
3.2.4.2 Reach Calculation 23
3.2.4.3 Model Sensitivity: Reach Normalization and Radius Parameter 24
3.2.4.3.1 Alternatives Reach Normalization 24
3.2.4.3.2 Alternatives Radius in Reach 26
3.2.5 Trip Rate Computations 26
3.2.5.1 Calculating the Trip Rates 26
3.2.5.2 Model Sensitivity: Effect of Regional Filtering 28
3.2.6 Calculating Generated Trips 28
3.3 Model Face Validation 29
4. Case Study Description 32
5. Results 36
5.1 Scraping of the Building Data 36
5.1.1 OSM scraping for Building Information Results 36
5.1.2 BAG scraping for building information results 38
5.1.3 OSM scraping pedestrian network...40
5.1.4 ODiN scraping for trip rate data 40
5.2 Mapping Trip Purpose Odin to Building Type 41
5.3 Assigning Building Weights 41
5.3.1 Weights based on study-area-specific data 41
5.3.2 Weights based on Dutch building standards 43
5.3.3 Proxy-based weights 45
5.3.4 Public transport (excluding the train station)...46
5.3.5 Model Sensitivity Results: Building Weights 46
5.3.5.1 Alternative Office Weight Results 46
5.3.5.2 Alternative Catering Weights Result...47
5.3.5.3 Alternative Public Transport Weight Result...48
5.4 Elastic Building Weights 49
5.4.1 Inserting Buildings into the Pedestrian Network 49
5.4.2 Reach Calculation..50
5.4.3 Model Sensitivity: Reach Normalization and Radius Parameter 52
5.4.3.1 Alternatives Reach Normalization 52

5.4.3.2 Alternatives Radius in Reach 54
5.5 Trip Rates Computations 56
5.5.1 Calculating Trip Rates 56
5.5.2 Model Sensitivity: Effect of Regional Filtering 57
5.6 Calculating Generated Trips 63
5.7 Model Face Validation 66
6. Discussion 70
6.1 Relation to Prior Work 70
6.2 Limitations of the BPT-Gen Model 71
7. Conclusion 73
8. References 74
9. Appendix 80
9.1 Appendix 1: Building Weights 80
9.2 Appendix 2: Trip Rate Patterns 90
9.3 Appendix 3: Generated Trips 104

1. Introduction

Cities today face major societal challenges such as rapid urbanization and climate change, which places increasing pressure on urban centres and strain their accessibility (Sweere, 2022 and TNO, 2023). As a result, many urban areas struggle with traffic congestion, pollution, and an overloaded transport network, highlighting the growing need for sustainable mobility solutions (PTV Group, n.d.).

One such solution is the implementation of low-traffic areas, which aim to reduce motor vehicle presence in residential zones, improve the quality of public space, and enhance conditions for walking and cycling (Sweere, 2022). While reducing car usage is difficult due to its perceived benefits, targeted interventions in high-traffic neighbourhoods or new developments offer more feasible opportunities (TNO, 2023).

In order to effectively implement such low-traffic measures careful planning, resilient policies, and community cooperation are crucial (Sweere, 2022). Traffic modelling and simulation play a key role in this by allowing cities to assess the potential impacts of measures under different scenarios and to communicate these impacts clearly to stakeholders (PTV Group, n.d.).

To effectively support low-traffic cities, such models must incorporate new mobility concepts (TBO, 2023), a central element of which is the representation of pedestrian movements. Knowing where and how many pedestrians are on the streets enables planners to make informed decisions, and the ability to model pedestrian activity at the city scale is essential for designing effective infrastructure in low-traffic areas and public spaces (Kaziyeva et al., 2023). Both hard measures, such as enhancing connectivity within the transport network, and soft measures, such as initiatives aimed at influencing travel behaviour, depend on accurate city-scale pedestrian flow predictions (Kaziyeva et al., 2023). This highlights the fact that macroscopic pedestrian models play a crucial role in addressing the challenges faced by modern cities, as city-scale simulations of pedestrian activity provide the insight needed to plan and design effective infrastructure and public spaces in emerging low-traffic areas.

For this reason, the municipality of Amsterdam and TNO aim to incorporate a macroscopic pedestrian model in their project to develop a new low-traffic neighbourhood, of which this report forms part without having to collect substantially more data. To be useful, the model must be fully applicable to the Dutch urban context and meet specific spatial and temporal resolution requirements. Spatially, it should provide sufficient detail to capture pedestrian volumes on individual street segments and assess the impact of new infrastructure or buildings, enabling informed decisions on, for example, pavement widths and building placements. Temporally, the model should ideally operate on an hourly basis to allow integration with microscopic pedestrian models for dynamic assessments, such as intersection safety analyses. Meeting these requirements ensures that the macroscopic model can effectively support both strategic planning and detailed, safety-focused evaluations of pedestrian infrastructure.

However, traditional transportation models often struggle with these requirements. Many were developed when automobiles were the primary focus, and as a result, current frameworks remain heavily biased toward car-based travel in both data and methods and struggle to represent pedestrian behaviour (Sevtsuk et al., 2024).

As the literature review in chapter 2 will show, most recent research on pedestrian models has concentrated on the microscopic scale, such as flows within specific facilities, leaving a gap in city- or regional-scale pedestrian modelling and the few macroscopic pedestrian models that have been developed tend to be highly data-dependent and often overlook behavioural aspects of walking (Kaziyeva et al., 2023).

Existing macroscopic pedestrian models range from regression-based to agent-based approaches, but none are tailored to the Dutch context. Most depend on detailed household travel surveys or extensive pedestrian counts, neither of which are available at the required resolution in the Netherlands. Consequently, a clear gap remains for macroscopic modelling approaches that are both practical and suited to Dutch data conditions.

This research project aims to tailor the first stage in pedestrian macroscopic modelling, trip generation, in order to establish a robust and reliable foundation for predicting pedestrian activity in the eventual full model which will include trip distribution and trip assignment. The objective of this study is thus to develop a pedestrian trip generation model with fine spatial and temporal resolution, tailored to the Dutch urban context and based on widely available data. This new tailored trip generation model enables planners to assess the impact of individual buildings and streets on pedestrian activity, while remaining feasible despite limited survey and count data.

This objective leads to the main research question:
To what extend can an existing pedestrian trip generation model be adapted to reflect the dynamics of pedestrian trip generation in a Dutch urban context?

To address this overarching question, the research is structured around the following sub questions:

SQ1 How do existing pedestrian trip generation models capture pedestrian behaviour and built environment characteristics?

SQ2. How can an existing pedestrian trip generation model be adapted to accurately reflect the Dutch urban context?

SQ3. How can Dutch open data sources be utilised to parameterise the pedestrian trip generation model?

SQ4. To what extent does the adapted pedestrian trip generation model reproduce observed pedestrian patterns in Dutch urban areas?

These sub questions form the structure of this report. Chapter 2: Literature Study addresses SQ1, which provides an overview of existing macroscopic pedestrian models, comparing their approaches and informing the methodological choices for this study. Chapter 3: Methodology , answers SQ2 concerning the adaptation of an existing model to the Dutch context. In particular, section 3.1: General Model Outline explains the functioning of the existing pedestrian count–calibrated model and its adjustments for Dutch urban conditions and temporal. SQ3 is addressed in section 3.2: Using Dutch Open Data for Model Development, which details how available data sources are leveraged to parameterise the model. The evaluation of the adapted model, including validation and sensitivity analysis under limited data conditions, is described in section 3.3. Chapter 4 introduces the case study to which the model is applied for validation, and the resulting model performance is presented in Chapter 5. Chapter 6: Discussion, reflects on the applicability of the developed model, its limitations, and opportunities for future research. Finally, Chapter 7: *Conclusion* summarises the main findings and provides clear answers to the research questions.

2. Literature Review on Pedestrian Models

This chapter examines how existing pedestrian models represent pedestrian behaviour and incorporate built-environment characteristics. It begins in section 2.1 with a brief overview of current pedestrian modelling approaches, with a particular focus on how pedestrian behaviour and land-use factors are incorporated and data requirements. This is followed by section 2.2, a comparative analysis of the identified models to highlight their respective strengths and limitations. Finally, section 2.3 reflects on the comparative analysis and the insights derived from this review form the foundation for the methodological choices made in this study.

2.1 Overview of Existing Pedestrian Models

A range of modelling approaches have been presented that can predict pedestrian demand. Hankey et al. (2011), for instance, use regression models linking street counts to build environment and sociodemographic, while Clifton et al. (2016) combine household survey data with regression-based walk mode split and destination choice models. Tian et al. (2017) focus narrowly on household-level trip generation using built environment and sociodemographic factors. In contrast, Kaziyeva et al. (2023) adopt an agent-based model (ABM) where trips emerge from individual activity schedules at minute-level resolution. Sevtsuk et al. (2024) introduce a building-level framework integrating trip generation, distribution, and assignment, calibrated with street counts. More recently, two studies propose hybrid approaches: Zhang et al. (2024) link Clifton's macroscopic model with an ABM, and Ullrich et al. (2024) combine a regression-based network model with an agent-based particle model.

2.2 Comparative Analysis of Pedestrian Models

Each of the pedestrian demand prediction models has its advantages and disadvantages. Table 1 summarizes the key characteristics of the macroscopic pedestrian models mentioned above. The table provides a quick overview of each model's type, scale, data requirements, behavioural representation, and spatial and temporal resolution. This overview serves to illustrate how existing models represent pedestrian behaviour and built-environment characteristics, while also identifying which modelling approaches are most compatible with the available Dutch data and the requirements of this study. Presenting the models in this format facilitates easier comparison in the following section, where their strengths and limitations will be analysed in greater depth.

The reviewed macroscopic pedestrian models are compared in terms of model performance, behavioural validity, spatial and temporal resolution, applicability to the Dutch context, and data requirements. Performance and behaviour validity indicates which approaches produce the most accurate predictions and ensures the model can represent realistic pedestrian decision-making, which are important for planning and scenario analysis. Spatial and temporal resolution is crucial for assessing pedestrian flows at the correct scale for mixed neighbourhood level and coupling with microscopic models. Applicability to the Dutch context evaluates whether models can be implemented with available local data, while data requirements determine feasibility given limited survey and pedestrian count availability. Together, these criteria guide the selection of a suitable framework for developing a building-level, hourly-resolution pedestrian trip generation model for Dutch urban areas.

2.2.1 Model Performance

Across the reviewed studies, model performances ranges from moderate to high. Early regression-based model of Hankey et al. (2011), captured broad spatial and socio-demographic trends but struggled with variability across time and locations, leading to only moderate predictive accuracy. Similarly, Tian et al. (2017) and Kaziyeva et al. (2023) achieved reasonable performance, with their models outperforming simpler benchmarks but still underestimating daily volumes or failing to fully capture variability in walking patterns. Clifton et al. (2016) demonstrated good explanatory ability for destination choice , however the mode split predictions remained weaker. More recent approaches show stronger predictive power. The UNA framework (Sevtsuk et al., 2024) and hybrid approaches (Zhang et al., 2024; Ullrich et al., 2024) performed better overall, with the UNA framework achieving very high accuracy after calibration (explaining over 90% of variation) and maintaining strong predictive power even out-of-sample. The hybrid MITO–MoPeD and ABM–NM models also improved behavioural realism and spatial distribution of trips, though their performance still depended on local calibration and were constrained by simplified representations of certain trip types (e.g., long-distance walks, transit access).

2.2.2 Behavioural Validity

In terms of behavioural validity, the models vary widely in how comprehensively they represent factors shaping pedestrian behaviour in trip generation. Tian et al. (2017) incorporates sociodemographic and built environment variables but capture only limited temporal variability. Clifton et al. (2016) further integrates household characteristics and trip purposes, though still with modest explanatory power. By contrast, agent-based approaches such as Kaziyeva et al. (2023) achieve higher realism by modelling individual level heterogeneity in socio-demographics, activity schedules, and even tourist flows. Sevtsuk et al. (2024) and Zhang et al. (2024) achieve realism by introducing finer spatial resolution and accessibility-based trip generation, though they omit socio-demographic inputs. Finally, Ullrich et al. (2024) balances macroscopic flow prediction with microscopic route simulation, but behavioural richness is constrained by reliance on built environment predictors and limited temporal differentiation.

2.2.3 Spatial and Temporal Resolution

With respect to spatial and temporal resolution, regression-based frameworks such as Clifton et al. (2016) and Tian et al. (2017) provide only limited temporal differentiation, though their spatial resolution varies from PAZ zones (80×80 m) to the household level. Kaziyeva et al. (2023) achieves the highest granularity, simulating pedestrian movements on street networks with minute-by-minute dynamics. Sevtsuk et al. (2024) also offers strong resolution by modelling at the building level and distinguishing morning, midday, and evening flows. Zhang et al. (2024) improves spatial detail with a 100×100 m raster but may remain too coarse for neighbourhood level analysis. Ullrich et al. (2024) delivers street segment resolution, but its temporal treatment relies on scaling factors rather than explicit modelling of time-of-day patterns.

Clifton et al. (2016) allows coupling with macroscopic multimodal models through spatial aggregation from PAZ to TAZ, while temporal alignment would be needed to integrate with microscopic pedestrian models. Tian et al. (2017) similarly requires spatial aggregation to link with multimodal models but lacks explicit mechanisms for temporal integration. Kaziyeva et al. (2023) naturally supports multimodal integration with multiple modes facilitating macroscopic coupling, and an ABM is already microscopic. Sevtsuk et al. (2024) could be linked to multimodal models if building-level flows are aggregated, while temporal alignment would be required for microscopic integration. Zhang et al. (2024) and Ullrich et al. (2024) already provide strong integration potential, with hybrid or OD-based approaches enabling coupling at both spatial and temporal levels.

2.2.4 Applicability to the Dutch Context

Notably, none of the reviewed models were developed in a Dutch context. Therefore, the primary criterion for evaluation is their transferability to the Dutch urban context, given the availability and limitations of local data sources. In the Netherlands, built environment data are widely accessible through open datasets such as OpenStreetMap and the Basisregistratie Adressen en Gebouwen (BAG). However, pedestrian count data are scarce, typically limited to a few locations near train stations and measured only over short periods. The main behavioural dataset available is the national ODiN survey, which provides socio-demographic characteristics, postal-code level origins and destinations, and departure times. While valuable, it is not household-based, tends to underreport walking trips, and lacks the spatial and temporal detail found in dedicated pedestrian surveys.

When looking at transferability to other contexts, Clifton et al. (2016) achieves fine spatial detail through PAZs but relies on detailed U.S. datasets and is designed for metropolitan areas, making transfer to Dutch neighbourhood-scale contexts moderate. Tian et al. (2017) operates at household and neighbourhood levels, allowing good spatial resolution, and could be transferred to other cities if comparable household travel 18 surveys are available. Kaziyeva et al. (2023) is conceptually transferable but requires extensive local data and full recalibration for new urban contexts. Sevtsuk et al. (2024) has been applied at scales similar to this study, and with calibration using Dutch pedestrian counts, it is transferable to similar-sized urban areas. Zhang et al. (2024) demonstrates transferability between cities, though calibration constants are needed to adapt to local conditions. Ullrich et al. (2024) shows strong transferability, with the NM trained in Stockholm and tested in Gothenburg, and the modular ABM structure allowing adaptation to different urban contexts.

2.2.5 Data Demands

The reviewed models differ significantly in their data demands. Hankey et al. (2011) and Ullrich et al. (2024) are moderately demanding, relying mainly on built environment data and count data for calibration, though the latter are scarce. Clifton et al. (2016), Zhang et al. (2024) and Tian et al. (2017) require detailed household travel surveys combined with parcel- or neighbourhood-level data, which are difficult to replicate in the Dutch context. Kaziyeva et al. (2023) is the most data-intensive, requiring full travel diaries and extensive counts, making application impractical. By contrast, Sevtsuk et al. (2024) is more feasible, as its building-level inputs can be sourced from open datasets, though reliable multi-period count data remain a bottleneck. Overall, while open spatial data provide a strong foundation, the lack of detailed travel surveys and comprehensive pedestrian counts limits the feasibility of data-heavy frameworks. Models with moderate requirements, such as Hankey et al. (2011), Sevtsuk et al. (2024), and Ullrich et al. (2024), are therefore more applicable with the available data in this study.

Table 1: Comparison of Pedestrian Models

Model/ Study	Type	Scale	Data	Explanatory Variables	Spatial Reso	Temp. Reso	Main Conclusion
Hankey et al. (2011) – network model	Regression Model	City	Street Counts	weather, socio-demographics, built environment characteristics	Street Segments	Daily	Incorporates wide range of explanatory variables, but, it struggles with variability across space and time. Its reliance on extensive pedestrian count data for calibration restricts its transferability.
Clifton et al. (2016) – First three steps of four step model	Regression Model	Metropolitan	Travel Survey	built environment and socioeconomic variables	Pedestrian Analysis Zones (80x80m)	Daily	Limited spatial and temporal resolution, restricting capture of fine-grained neighbourhood-level variations. Relies on highly detailed, disaggregate data, reducing transferability to contexts with limited datasets.
Tian et al. (2017) – two step home based walking trip generation	Regression Model	Neighbourhood	Travel Survey	built environment and socioeconomic variables	Building Level (only includes residential buildings as origins)	Daily	Data-intensive, relying on detailed household travel surveys and GIS-based data. Provides good neighbourhood-level spatial resolution, but lacks temporal detail and considers only households as trip origins, limiting integration and transferability.
Kaziyeva et al. (2023) – ABM modelling decision processes	Agent - Based Model	Regional	Travel Survey, Street Counts	built environment, socioeconomic and mode variables, activity type	Street level	Minute	High behavioural realism and multimodal representation with fine spatial and temporal detail, but highly data-intensive and difficult to transfer due to reliance on detailed local surveys and network data.
Sevtsuk et al. (2024) – Urban Network Analysis	Accessibility-based network model	Neighbourhood	Street Counts	Built environment	Building Level, Street Segments	Morning, midday and evening periods	High spatial accuracy with detailed built-environment representation, but limited behavioural and socio-demographic realism and dependent on availability of pedestrian counts for calibration.
Zhang et al. (2024) – Hybrid MITO-MoPeD Framework	Hybrid regression and agent-based model	City	Travel Survey	built environment, socioeconomic variables, trip purpose	Pedestrian Analysis Zones (100x100 m)	Daily	Integrates macroscopic and agent-based models to improve behavioural realism and sensitivity to short walking trips. Requires detailed household survey data and calibration constants for transferability across cities.
Ullrich et al. (2024) – Hybrid network and particle model	Hybrid regression and agent - based (particles) model	City	Street Counts	Built environments, Interaction with environment and other pedestrians	Street Segments	Hourly (adjusted with factors)	High spatial resolution and transferability and good temporal resolution which does rely on context specific factors and is limited in behavioral variables (limited built environment variables and no socioeconomic variables)

2.3 Reflection on the Comparison of Pedestrian Demand Modelling

When evaluating across all criteria, a clear trade-off emerges between behavioural realism, spatial and temporal resolution, data requirements, and applicability to the Dutch urban context.

Models with high behavioural realism, such as regression-based and agent-based approaches, require detailed household travel surveys and extensive pedestrian counts, which are not readily available in the Dutch context. While agent-based and hybrid models offer fine spatial representation, their data demands limit their practical applicability. In contrast, the building-level UNA framework (Sevtsuk et al., 2024) provides suitable spatial detail while relying mainly on built environment data and limited calibration counts, making it the most feasible option for application in Dutch urban neighbourhoods.

Moreover, the building-level spatial resolution of the UNA framework allows for the representation of trips between non-residential buildings, making it particularly suitable for mixed-use neighbourhoods. This level of spatial detail also supports the inclusion of access and egress movements at transit nodes and boundary connections, an aspect highlighted as critical by Zhang et al. (2024) and Tian et al. (2017), the latter demonstrating that such trips can be captured by applying a buffer around the study area reflecting typical walking distances.

Finally, regarding integration potential, Ullrich et al. (2024) emphasize that effective coupling between macroscopic and microscopic models requires compatible temporal resolutions. As the microscopic model used in this project operates at an hourly resolution, the trip generation model must reflect a similar temporal scale. While Sevtsuk et al.’s framework incorporates temporal variation, it does so post-assignment through calibration, limiting its suitability for direct integration. In contrast, agent-based and hybrid models inherently capture fine-grained temporal variation through travel survey or count-based data, but their data demands remain a constraint.

3. Research Methodology

This chapter presents the methodology for developing a pedestrian trip generation model tailored to Dutch urban neighbourhoods. It addresses the central methodological question SQ2:

> *how can an existing pedestrian trip generation framework be adapted to accurately represent the Dutch context?*

Section 3.1 introduces the overall modelling approach, describing which components from existing pedestrian trip generation models are adopted and how these are adapted to form the new framework. Section 3.2 then details the specific adaptations required for application in the Dutch context. Section 3.2.1 explains the use of Dutch building data. Section 3.2.2 discusses how the different categories in ODiN and BAG are linked, while Section 3.2.3 describes how buildings are assigned weights based on study-area data, Dutch datasets, or national building guidelines. Section 3.2.4 outlines how the elastic weights from the base model (UNA model) are adjusted using available Dutch data, and Section 3.2.5 presents how trip rates are derived from the national household survey to achieve finer temporal resolution. Section 3.2.6 then explains the final calculation of generated trips. Finally, Section 3.3 discusses the model face validation strategy and the challenges posed by the limited availability of pedestrian data.

3.1 General BPT-Gen Model Outline

This section provides an overview of the key principles of the UNA model of Sevtsuk who's trip generation, will be used as a base of the new macroscopic pedestrian trip generation model which will be called the BPT-Gen model and why these key principles will be adapted to be for the Dutch context and to have a finer temporal resolution.

3.1.1 Explanation of the UNA Tool

Sevtsuk et al. (2024) present the urban network analysis (UNA) tool, which provides an accessibility-based framework, see figure 1. This UNA framework assesses how built environment shapes pedestrian travel. Before introducing the specific adaptations, a brief overview of the full UNA modelling process is provided to clarify its main components. Applied at the city level, UNA generates walking trips from origin points weighted by user-defined factors, distributes them to destinations within a search radius using a Huff destination choice model. Route assignment is modelled through a customized betweenness algorithm that incorporates a detour ratio to reflect pedestrians' tendency to avoid strictly shortest paths. The model calibrates the outputs with observed pedestrian volumes on selected streets.

The UNA framework defines multiple origin types, such as homes, amenities, parks, schools, places of worship, parking lots, bus stops, institutions, and employment locations. Each origin is assigned a numeric weight representing its trip-generating potential. For example, homes are weighted by the number of residents, employment locations by jobs or university enrolments, food businesses by seat counts, transit stations by the number of lines, parks by area, parking lots by capacity, and tourist attractions by TripAdvisor review counts (Sevtsuk et al., 2021). If a person count is unknown, the weight is set to one.

The next step in the UNA framework is to insert the building as nodes in a pedestrian network since all spatial relationships are analysed along networks. This pedestrian network is not only needed for the route choice model but also for the trip generation since this framework adjust the weights of the origins elastically based on accessibility.

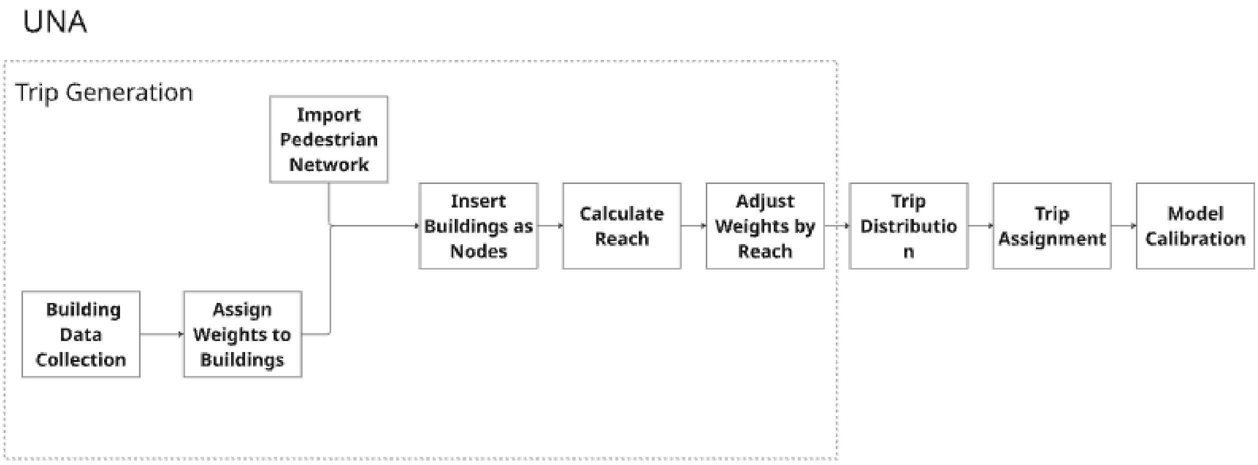


Figure 1: Flow chart describing the structure of UNA

The UNA framework allows the weights to be elastic with respect to the destination accessibility since research has shown that better pedestrian accessibility to destination can lead to more pedestrian trips being generated to such destinations (Sevtsuk et al., 2024). To adjust the weights of the origins users must first measure gravity access from each origin to a given destination type and then adjust trip generation rates at the origins with the gravity accessibility results, saving out a new elastic trip generation weight for each origin (Sevtsuk et al., 2024). The Hansen (1959) gravity-based accessibility index measures how reachable destinations are by combining their attractiveness with the travel cost required to reach them. It includes an empirically derived exponent that determines how strongly additional destinations contribute to accessibility, allowing for diminishing returns as destination availability increases.

A less data-intensive alternative considered by Sevtsuk, is the cumulative opportunities accessibility index, also called the Reach index (Bhat et al., 2001). For a given origin, the Reach index counts how many destinations of a type can be accessed within a specified walking distance. For example, how many retail shops can be reached from each residential building within a 600-m walk (Sevtsuk et al., 2024). Only destinations reachable via the shortest paths in the pedestrian network are included, and their weights are summed to calculate the index. Still, the UNA model applies the gravity accessibility index, as it simultaneously considers both the weights of destinations and the travel costs. The precise effect of this formulation on the results, however, remains uncertain, partly because calibration and validation in the UNA framework occur only after the trip assignment stage.

The elastic weights of the UNA framework serve as a first estimate of the generated trips, and only after the trip assignment temporal variations are included by calibrating the model for the morning, midday, and evening periods.

3.1.2 Explanation of the Adaptation of the UNA Structure in BPT-Gen

The BPT-Gen model focuses exclusively on the trip generation component of UNA, with several adaptations to suit Dutch urban conditions and available data, see figure 2. In the BPT-Gen, the model also starts with defining the origins, only difference here is the data collection , where in the BPT-Gen model Dutch data, since the model will be applied in the Dutch context.

BPT-Gen

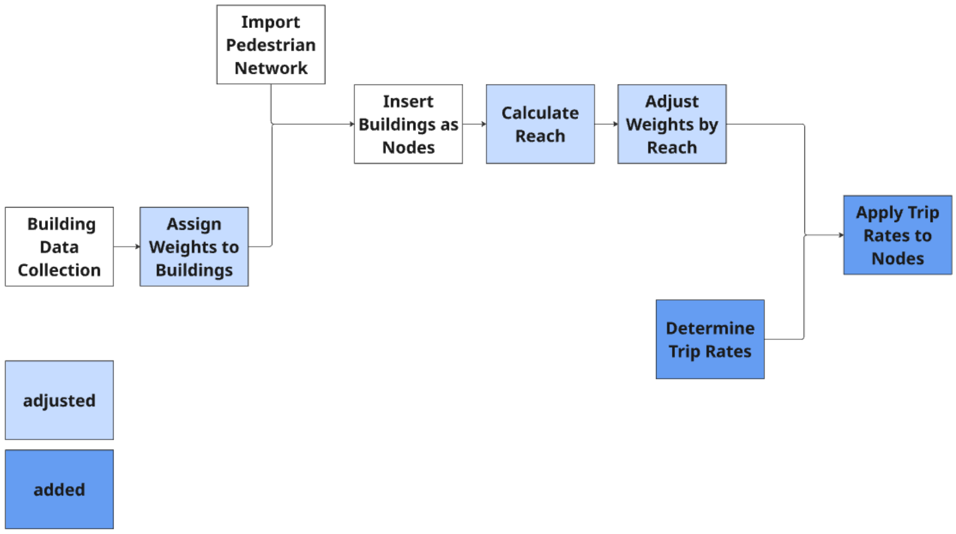


Figure 2: Flow chart describing structure of the BPT-Gen

The first major adaptation of the UNA framework concerns the assignment of building weights. In the BPT-Gen model, this step is modified to achieve a finer temporal resolution and to compensate for the absence of a calibration phase. As a result, the initial weight estimates are designed to more closely reflect real-world activity levels. The revised weighting method is described in detail in Section 3.2.3.

The BPT-Gen model closely follows the method of the UNA model for inserting buildings as nodes described in Section 3.2.4.1. This approach was adopted because it can be easily applied to Dutch data without requiring substantial modification.

Despite the uncertainty of the impact of accessibility calculations, the elastic accessibility approach is still adopted to try and capture the behavioural assumption that pedestrian trip generation is influenced by the accessibility of relevant destinations. However due to the fact that there is not enough data to empirically derive the exponent in the gravity accessibility the reach index will be used instead. How this is applied in the BPT-Gen model specific for the Dutch area is described in section 3.2.4.2 and how to evaluate how this affects model performance, an impact analysis will be conducted see section 3.2.4.3.

To get an hourly temporal resolution for the BPT-Gen model the approach developed combines the building-based trip generation of Sevtsuk et al. (2024) with survey-based trip rate estimation using the Dutch national ODiN travel survey. While ODiN lacks the detail required for full agent-based calibration, it provides sufficient information to estimate hourly trip rates by trip purpose. These rates are used to disaggregate trip generation to an hourly scale, drawing conceptually on ABM frameworks that use survey data to capture fine-grained temporal dynamics.

The outcomes from both components, the accessibility-adjusted building weights and the hourly trip rates, are then combined to calculate the total number of trips generated per node (building) per hour. Since this report focuses only on the trip generation stage, trip distribution and assignment are not yet included. Consequently, the model will not undergo full statistical calibration. Instead, results will be assessed through face validation and plausibility checks, to evaluate whether the generated trip patterns provide a robust foundation for future model development and integration with microscopic pedestrian simulations.

3.2 Using Dutch Open Data for Model Development

This section describes how Dutch open data sources are incorporated into the development of the BPT-Gen model. First**,** section 3.2.1 details how information regarding buildings and the pedestrian is scraped from the Dutch building datasets. Section 3.2.3 . describes how functional weights are assigned to the buildings. And section 3.2.4 presents how buildings are positioned within the pedestrian street network to enable network-based accessibility calculations and how these building weights are made elastic using the reach index.

The use of national household travel survey data to derive trip rates and temporal variation is then described in section 3.2.5, followed by section 3.2.5, which outlines how building-level characteristics and trip rates are combined to generate pedestrian trips. Finally, section 3.3 presents the approach to model face validation under limited pedestrian count availability.

3.2.1 Scraping Building, Network and Household Survey Data

In BPT-Gen, buildings serve as origins for pedestrian trip generation, with their location, type, and size determining both trip weights and temporal activity patterns, reflecting differences in pedestrian behaviour across building types.

Similar to the original UNA framework, buildings are defined as origin and destination nodes in the pedestrian network. For each building, attributes such as location, function, and size are obtained from Dutch open data sources. In particular, Open Street Map and BAG are adopted because they jointly contain relevant information pertaining to the buildings in the case study area.

3.2.1.1 OSM scraping for Building Information

The collecting of these attributes starts with defining the study area boundaries for which OpenStreetMap (OSM) is used.

OSM is a volunteered geographic information project with several advantages: the data is freely available, it covers almost all countries worldwide, and it is updated continuously (Zhou, 2022). For European countries in particular, the data has relatively high accuracy, and its quality is generally sufficient for land-use mapping (Zhou et al., 2022). OSM is therefore selected as a base source because it combines open accessibility with sufficient accuracy for the purposes of this study.

To account for walking trips entering or leaving the study area, a buffer is applied around the case study boundaries. The buffer size is set based on the mean walking distance relevant to the case study, derived from the ODiN survey. This ensures that trips starting outside the study area but continuing on foot within it are included in the analysis, while also capturing trips that begin within the study area and exit on foot. Transport hubs located within the buffer further help represent walking flows originating from outside the study area.

OSM is scraped for trip generating functions amenities (e.g., catering industries), public transport stops, bicycle and car parking facilities, and parks. Catering and parks are treated as leisure destinations that generate walking trips, while public transport stops and parking facilities represent access and egress points, for instance where commuters arrive by train or car and continue their journey on foot within the study area.

OSM amenity types are then aggregated into broader functional categories (for example, grouping all food- and drink-related amenities under ‘catering’). This step reduces complexity in the dataset by consolidating numerous detailed OSM categories, many of which lack information on footprint or GFA and therefore cannot be weighted directly. Aggregation also ensures that OSM features match the functional categories used for applying trip rates, since the trip rates differentiate only at a broad level (e.g., leisure trips) and do not distinguish between more specific activities such as visiting catering establishments, going to a park, or even between different types of catering. At this stage, it is assumed that this aggregation does not substantially affect model outcomes, even though some discrepancy with real-world behaviour is expected. In practice, different leisure activities likely attract pedestrians at varying intensities and times of day, meaning that the simplified representation may not fully capture temporal or behavioural variation. Nonetheless, this level of aggregation is maintained to ensure consistency with the available trip rate data.

For each of these OSM features, geometry, type, and footprint are stored.

3.2.1.2 BAG Scraping for Building Information

Even though OSM is accurate enough, it does lack detailed information regarding buildings. That is why in addition to OSM, building data are also imported from the *Basisregistratie Adressen en Gebouwen* (BAG). In this study, the 2023 BAG dataset was used to ensure temporal consistency with the other data sources. BAG is the official Dutch national registry of all addresses and buildings, widely used by both governmental and private organizations. It contains detailed and authoritative information such as construction year, floor area, usage type(s), and precise geometry (EduGIS, n.d.). The database is openly accessible and updated monthly (EduGIS, n.d.). This dataset was thus selected because it is openly available and provides a reliable, authoritative source of building information for the Netherlands.

BAG is preferred over OSM for building-level trip generation because it provides more disaggregated functional information. For example, within a single physical building (*pand*), multiple *verblijfsobjecten* (VBOs) may be recorded, each with its own designated function. This makes BAG particularly valuable in mixed-use contexts, as it captures the coexistence of different uses such as offices, shops, or meeting spaces within the same structure. This level of detail is assumed to be important because different building functions attract a different amount of people and distinct user groups and exhibit different temporal patterns of activity. This assumption is supported by Kuiper (2021), who notes that BAG enables a finer level of granularity, with precise identifiers for individual units of use (VBOs) that facilitate targeted analysis of trip generation.

In addition, BAG provides gross floor area (GFA) information, whereas OSM typically only records building footprint area. It is assumed that the GFA values in BAG are both correctly measured and properly allocated across multiple VBOs within the same building footprint. The GFA is important for estimating the relative weight of different building functions. Because both the functional detail of individual VBOs and the availability of GFA data are essential for capturing variations in activity levels across building types, the use of BAG is considered most appropriate for this study. This is particularly relevant since trip rates are explicitly incorporated into the BPT-Gen model, requiring accurate information on building use and size to estimate trip generation more reliably. This importance is also highlighted by Kuiper (2021), who explains that the level of detail of BAG, which 21 categorizes buildings based on function and provides specific attributes such as GFA, allows for more accurate trip generation estimates based on the expected trip rates for different land uses. OSM may not provide the same level of granularity, focusing more on roads and geographic features than on building specific data (Kuiper, 2021).

From BAG, two datasets are imported: *verblijfsobject* (VBO) and *pand*. The dataset *pand* is imported in order to visualize the building footprints in plots making it easier to identify the locations but the most valuable information needed for the model are given in the VBO dataset.

The VBO dataset represents individual building units with an official use designation. All VBOs located within the buffered study area are selected. Because BAG does not overwrite records when updates occur but instead adds new entries under the same identification number with a more recent timestamp, the dataset is filtered to retain only the most recent version of each VBO. It is filtered for the most recent one because keeping only the latest version ensures temporal consistency with the 2023 reference year. Further cleaning of the data is done. The VBO dataset is further cleaned by retraining only the attributes relevant for this study: identification number, usage purpose, gross floor area (GFA), and location. These variables provide the essential information needed to assign functions to units, weight them by size, and place them spatially within the study area. Other BAG fields, such as construction year, are excluded because they do not directly influence trip generation and would only add unnecessary complexity to the dataset. This results in a streamlined dataset of building use units that captures both the functional role and the scale of each unit.

The harmonized OSM features and BAG VBOs are subsequently combined into a single dataset that defines the complete set of potential origins and destinations for pedestrian trips.

3.2.1.3 OSM scraping for pedestrian network

For the accessibility calculations, buildings from BAG and OSM are represented as nodes within a pedestrian network. Since the model focuses exclusively on walking trips, the network was extracted from OSM and filtered to include only walkable paths within the boundaries of the study area. Specifically, only streets, footpaths, and other links designated as accessible to pedestrians were retained, see figure 8, while roads intended solely for vehicles or restricted access were excluded. This ensures that the network accurately reflects the paths pedestrians can actually use within the case study area. OSM was chosen because it provides detailed, open, and reproducible data suitable for creating a comprehensive pedestrian network.

3.2.1.4 ODiN scraping for Trip Rate Data

Next to building and network information. Also, insights into the average trip rate per building type is required. The only available dataset that features insights into the average trip rate is the *Onderzoek Verplaatsingen in Nederland* (ODiN), a large-scale national travel survey conducted annually by Statistics Netherlands (CBS). ODiN provides detailed information on individuals' daily mobility behaviour in the Netherlands, collected through survey responses and linked to contextual variables such as demographics, household composition, and place of residence.

In ODiN, each respondent is assigned a unique personal identifier, which allows all trips reported by that individual within a single day to be linked. Trips are recorded as individual entries, each representing a movement from an origin to a destination or a complete tour, and are sequentially numbered to capture the order in which trips are undertaken throughout the day. Each trip includes several variables describing the movement, such as the origin and destination (recorded at the 4-digit postcode and municipality level), the purpose of travel (categorized into 14 types), the main mode of transport used, the start time and activity duration, a tour indicator to identify whether the movement is part of a closed tour or a single trip, which is important for analysing return behaviour, and the trip distance. Some movements may include multiple legs; for example, if a person cycles to the train station, takes the train, and then walks to work, ODiN records this as a single movement with three legs. Each leg corresponds to a specific transport mode, with its associated distance and mode recorded separately.

The most recent available ODiN dataset, from 2023, was imported and prepared for analysis. To derive walking trip rates, the raw survey data must first be filtered and processed to ensure it captures only relevant trips and aligns with the study context.

The initial step involves filtering the ODiN dataset for the correct region. Two options are considered see section 3.2.5.2.

The second step involves splitting tours into individual trips to enable explicit origin–destination modelling and to capture return behaviour. Tours are broken down into single-trip records, and the destination of the previous trip is used to construct the corresponding return trip. If no prior destination is available, the return trip is assumed to lead back home. For these generated return trips, the origin and destination are swapped, and the departure time is adjusted to account for the activity duration at the destination. In cases where activity duration is missing, a default of 30 minutes is imputed, reflecting a reasonable upper bound for walking to a typical destination such as a leisure activity. All return trips are assigned new sequential trip numbers to preserve the correct chronological order.

The third step involves filtering the dataset to focus on regular commuting patterns. Only trips occurring on weekdays are retained, and the analysis is restricted to trips with origins, destinations, and home locations within major urban areas comparable to the study area, ensuring that the results provide a representative reflection of travel behaviour in the context of the study.

3.2.2 Mapping Trip Purpose (ODiN) to Building Type

BPT-Gen requires trip rates at the building level. Unfortunately, in ODiN origins and destinations are reported at the 4-digit postcode level.

3.2.2.1 Trip Purpose to Building Type

To bridge this gap, the recorded purpose of travel in ODiN is used as a proxy to assign trips to corresponding building types (see table 2). For example, a trip purpose coded as *going home* is mapped to the building type *home*. Similarly:

- The purposes *work*, *business visit in a professional context*, and *professional activity* are grouped under the building type *work*.
- The purposes *pick up/drop off a person*, *pick up/drop off goods*, and *other purpose* are grouped as *other*.
- *Shopping/groceries* is mapped to *shopping*.
- *Touring/strolling* and *other leisure activities* are mapped to *leisure*.
- *Sport/hobby* is grouped as *sport*.
- *Services/personal care* is mapped to *healthcare*. Which has been shown that many trip purposes strongly correlate with specific land-use types (Lee et al., 2015)

Please note that this mapping approach makes several simplifications and assumptions. Several ODiN purposes are combined into a single building type, even though in practice the destination could vary. For instance, *Business visit in a professional context* is assigned to *work*, based on the assumption that most such visits occur in workplaces, even though they could occasionally take place at private residences. *Hobby* activities are grouped under *sport* to avoid merging them with hospitality or event-related buildings, though this means they are no longer aligned with other gathering or cultural locations. Furthermore, in the study of Lee et al. (2015) it is stated that not all trip purposes correlate to land use. Trip purposes, such as “work related”, “business related”, “change type of transportation” or “recreation and entertainment” can go to a variety of land-use types (Lee et al., 2015).

However, the grouping of trip purposes ensure comparability and reduces the level of detail to a manageable set of purpose categories in the dataset. And according to Lee et al (2015) most of the trips go to highly predictable land-use types (given their purpose), which means that overall, there is much information in the land-use map that could be exploited to predict the trips, given their purpose. At the same time, they introduce limitations, as certain nuances of travel purpose (e.g., business visits at non-work sites) cannot be distinguished in this classification.

Table 2: Trip Purposes and Building Types Mapping

Trip Purpose	Mapped Building Type
Going home	Home
Work	Work
Business visit	Work
Professional activity	Work
Pick up / drop off a person	Other
Pick up / drop off goods	Other
Education / course	Education
Shopping / groceries	Shopping
Visiting friends or staying overnight	Home
Touring / strolling	Leisure
Sport / Hobby	Sport
Other leisure activities	Leisure
Services / personal care	Healthcare
Leisure	Hotel
Other purpose	Other

3.2.2.2 Main Mode of Transport to Building Type

For walking leg trips a similar mapping was applied. However, instead of using the trip purpose to identify the building type, the main mode of transport were mapped to building type see table 3.

Table 3: Main mode and building type mapping

Main Mode of Transport	Mapped Building Type
Car / touring car	Car Parking
Bicycle / speed pedelec / electric bicycle / other without motor	Bicycle parking
Motor / moped	Motorcycle parking
Train	Train station
Subway	Subway station
Tram	Tram stop
Bus	Bus stop

3.2.2.3 Mapping Building Function to Building Type

To ensure consistency between the land-use dataset and the ODiN-derived trip rates, the detailed BAG and OSM function labels are consolidated into broader origin and destination categories that align with the categories used in the trip rate data, which are only available for coarser functional level. Note that this last mapping is performed after the weights per building are computed (see section 3.2.3), as the detailed building types are required to compute the weights.

The Mapping from Building function to building type is provided in table 4. In particular:

- Offices and industrial buildings are merged into a single *work* category to represent employment-related activity as these are expected to have similar travel behaviour.
- Assembly functions (e.g., conference centres, theatres), catering, and parks are grouped under *leisure* to capture social, cultural, and recreational trips, as it is assumed that all have the same temporal behaviour.
- Other functions are relabelled directly to match the trip rate categories: *home* (residential), *education* (educational buildings), *shopping*, *sport*, *healthcare*, and *other*. Parking facilities are labelled as *car parking*, *bicycle parking*, and *motorcycle parking*.
- Public transport infrastructure is divided by type into *train station*, *subway station*, *tram stop*, and *bus stop*.

This grouping reduces the original level of detail in BAG and OSM to a functional typology consistent with trip rate data. While it enables systematic application of trip rates, it also introduces assumptions. For instance, assembly functions are treated the same as parks and catering under *leisure*, even though travel behaviour may differ. Likewise, combining offices and industrial uses under *work* simplifies employment trips but may obscure variations in activity patterns. Such aggregation is necessary to operationalize the model, but it inevitably smooths out some land-use distinctions.

Table 4: BAG/OSM building function to defined building type

Original Building Function	Mapped Building Type
Residential use	Home
Office	Work
Industry	Work
Educational	Education
Shopping	Shopping
Assembly function	Leisure
Catering	Leisure
Park	Leisure
Sport	Sport
Healthcare facilities	Healthcare
Other	Other
Parking	Car Parking
Bicycle parking	Bicycle parking
Motorcycle parking	Motorcycle parking
Station	Train station
Subway	Subway station
Tram stop	Tram stop
Platform	Bus stop
lodging	Lodging

3.2.3 Assigning Building Weights

Each origin must be assigned a weight that reflects its relative trip-generating potential. In the BPT-Gen framework, building weights represent the estimated intensity of pedestrian activity for each origin type, expressed as the approximate number of unique individuals associated with a building over the course of a typical day. These weights are derived using different approaches depending on the availability and quality of data.

This section is structured accordingly. The first part describes how weights are assigned using study area–specific data, where detailed local information is available. The second part explains the use of Dutch building standards, applied when no area-specific data exist. The third part introduces proxy weights, which are used for origins lacking gross floor area (GFA) or footprint data; in these cases, alternative indicators are employed to approximate trip generation potential. The next section explains the weighting approach for public transport stops. This section concludes with a sensitivity analysis, explaining how some of the categories have a different weighting approach.

3.2.3.1 Weight computation using study area specific data

This weight computation strategy is applied to building types for which study-area-specific data are available, being: residential, educational, and healthcare functions. Focusing on these functions first allows the use of detailed, locally relevant information to assign weights for trip generation, providing a reliable baseline before addressing other building types where only general proxies are available. In this context, ground floor area is treated as the primary driver of trip generation, regardless of actual intensity of use or occupancy patterns.

For **residential buildings**, study-area-specific data on the total number of residents was obtained from the CBS *Kerncijfers wijken en buurten* 2023 (CBS, n.d.), which provides statistics at the level of municipalities, districts, and neighbourhoods. However, this dataset does not provide disaggregated information at the level of individual buildings or six-digit postal codes. Moreover, BAG does not distinguish between different types of residential buildings, such as social housing, owner occupied housing or luxury apartments. Therefore, the number of residents per building is assumed to only depend on the GFA. The total number of residents in the study area was obtained directly from the CBS dataset, where the study area corresponds to a specific entry in the table. Residential weights are then estimated by first calculating the occupancy rate by distributing the total population of the study area proportionally across the gross floor area (GFA) of all residential VBOs in the study area, see formula 1. The weight of a residential building is than calculated by multiplying the occupancy rate with the GFA, see formula 2. This approach assumes that larger buildings accommodate more residents, without differentiating between building types. Although this simplification introduces some uncertainty due to variations in actual population densities, using GFA as a proxy is considered a reasonable method for estimating trip generation in the absence of building-level population data.

$$Occupancy\ rate_{residential} = \frac{Total\ residents\ in\ the\ study\ area}{Total\ residential\ GFA\ in\ study\ area} \qquad [\frac{persons}{m^2}]\ (1)$$

$$Weight_{\ Building} = GFA_{\ Building} \cdot Occupancy\ rate \qquad [persons]\ (2)$$

For **educational buildings**, data on the number of students and staff at institutions within the study area were collected from *allecijfers.nl* and institutional websites. While these figures are available per specific building, for consistency with the residential weighting approach and to simplify the allocation, the total number of students and staff for each institution was summed and then divided by the total educational GFA of the study area, see formula 3. Then the number of staff and student per educational building where determined with formula 2. Comparison of the allocated number of students and staff per educational building with the actual figures shows a close correspondence, confirming that weighting by GFA provides a reasonable proxy for trip generation. An important assumption in this approach is that students and staff are present on a daily basis. Although this does not fully reflect reality, it provides a consistent and transparent method for estimating educational activity.

$$Occupancy\ rate_{education} = \frac{Total\ students\ and\ staff\ in\ the\ area}{Total\ educational\ GFA\ in\ study\ area} \qquad [\frac{persons}{m^2}](3)$$

Healthcare building weights were determined using the total number of hospital beds and staff (ESMO, n.d.). For the study area only aggregate data was available for the entire hospitals, while BAG identifies multiple healthcare buildings per hospital. To allocate activity across the individual hospital buildings, beds and staff were distributed proportionally by GFA, see formula 4 and 2. This approach assumes that the combination of beds and staff provides a reasonable representation of daily activity, that fluctuations due to staff shifts and patient turnover do not substantially affect trip generation. This assumption is supported in the literature, where Al-Masaeid et al. (2021) note that the number of beds serves as a reliable proxy for hospital-related trips. The choice to distribute the number of beds and staff over the different hospital buildings assumes that beds and staff are evenly distributed across all buildings within a hospital.

$$Occupancy\ rate_{healthcare} = \frac{Total\ beds\ and\ staff\ in\ study\ area}{Total\ healthcare\ GFA\ in\ study\ area} \qquad [\frac{persons}{m^2}] (4)$$

The train station, NS Amsterdam Zuid, was weighted using observed daily check-ins and check-outs from the NS dashboard, providing a more direct measure of passenger activity. The total was divided by two (see formula 13) to avoid double-counting, based on the assumption that a passenger who checks out in the morning will likely check in again in the evening (and vice versa). This adjustment ensures that each individual is counted only once when estimating daily pedestrian activity.

$$Weight_{\ Train} = \frac{Daily\ Check-ins\ and\ Check\ outs}{2} \qquad [persons]\ (5)$$

3.2.3.2 Weights Based on Dutch Building Standards

For other building functions, including **offices**, **industrial buildings**, **retail**, **sports facilities**, **assembly functions**, **lodging**, **parks**, **prisons**, and various types of **parking**, no study-area-specific data are available. In the absence of such data, regulatory design standards from the Dutch Bouwbesluit (BRIS, n.d.) provide a consistent and reproducible basis for estimating potential occupancy. These standards specify the average required floor area per person for different building uses, which allows the assignment of weights in a way that reflects likely trip generating activity.

For **office buildings**, an additional adjustment is applied to account for hybrid work patterns, recognizing that not all employees are present onsite every day. This ensures a more realistic estimate of trip generation compared to assuming full occupancy.

Parking weights were calculated based on typical space requirements per vehicle type: multi-level rack capacities for bicycles, standard Dutch parking dimensions for cars, and an estimated half-space requirement for motorcycles. The number of vehicles accommodated was then multiplied by the footprint area, with a small constant added to prevent zero weights for very small parking areas. This approach ensures that even small parking facilities contribute to trip generation and that footprint-based calculations approximate the number of vehicles that can realistically access or depart from the area. Only public-access parking is considered, since residential parking is assumed to correspond to home-based car trips rather than walking trips within the study area. For example, parking near shopping areas generates access/egress trips where pedestrians continue into the study area on foot.

The general occupancy rate and weight formulas are:

$$Occupancy\ rate_x = \frac{1}{Required\ floor\ area\ per\ person\ or\ vehicle_x} \qquad [\frac{persons}{m^2}] (6)$$

$$Weight_{building,x} = GFA_{building} \cdot Occupancy\ rate_x \qquad [persons] (7)$$

For office spaces, the occupancy rate is further adjusted by the average daily presence of workers to account for hybrid working patterns:

$$Occupancy\ rate_x = \frac{1}{Required\ floor\ area\ per\ person_x} \cdot average\ daily\ occupance \qquad [\frac{persons}{m^2}]\ (8)$$

3.2.3.3 Proxy Weights

For **catering functions**, no study-area specific data, GFA or footprint was available from OSM or other sources. To estimate trip generating activity, two proxies are tested. The first proxy is based on regional catering statistics. This approach was chosen because this data is readily available and reproducible for other areas, making it a practical solution in the absence of direct measurements. The proxy calculates the estimated number of customers per day per venue by dividing total provincial revenue for the catering sector by the average national spending per customer, then dividing this by the total number of catering venues and by 300 assumed active days per year. The resulting figure was assigned as the weight for each catering venue in the study area, see formula 8. This method relies on several strong assumptions: catering activity in the study area reflects the regional average, customer activity is evenly distributed across venues, and each venue operates approximately 300 days per year. While these assumptions introduce uncertainty, this approach provides a consistent and transparent estimate of daily catering activity. Alternative methods, such as using TripAdvisor or Google Maps business hours and popularity data, could offer more precise proxies but may be less easily generalizable.

$$Weight_{catering} = \frac{Total\ Revenue_{province}}{Avg.spending\ per\ customer_{national}} \cdot \frac{1}{nr\ of\ catering\ venues_{province} \cdot active\ days\ per\ year} \quad [persons](9)$$

3.2.3.4 Public Transport Weights

Weights for **tram**, **bus**, and **metro** stops were calculated based on both the number of lines and their service frequency, whereas Sevtsuk et al. (2021) used only the number of lines as a proxy. Sevtsuk et al. (2021) was used as a guide to estimate pedestrian activity, but to capture 24 daily activity more accurately, this study also incorporates frequency (the number of times a specific line stops at each stop). The weight for each stop is therefore calculated as the number of lines serving the stop multiplied by their frequency, see formula 9 where *i* is the public transport line serving the stop, *n* is the total number of lines at the stop and $Frequency_i$ is the number of departures per line per day.

$$Weight_{PTstop} = \sum_{i=1}^{n}(Frequency_i) \quad [persons](10)$$

3.2.3.5 Model Sensitivity: Building Weights

To evaluate how assumptions in building-level weights influence pedestrian trip generation, a series of sensitivity tests was carried out for key building categories. These weights play a crucial role in the BPT-Gen model because they determine the relative contribution of each building to trip production and attraction. However, several building types—particularly offices, catering establishments, and public transport facilities—require proxy-based or aggregated assumptions due to limited availability of establishment-specific activity data. The sensitivity analysis therefore examines how alternative weighting formulations affect model behaviour by systematically varying these assumptions.

3.2.3.5.1 Alternative Office Weights

To evaluate the impact of occupancy assumptions on final generated trips from offices two different occupancies where tested. In the baseline approach, weights were based on the average daily occupancy rate of office workers, approximately 30–35% in the Netherlands. Peak days, particularly Tuesdays and Thursdays, reach occupancy rates of 50–60% (Verwoerd, 2024). In the first iteration, the average weekly occupancy was applied, while in the second iteration, weights were recalculated using the maximum observed occupancy of 60% to represent peak activity. This allows analysis of how higher occupancy affects trip generation.

3.2.3.5.2 Alternative Catering Weights

The catering building function is weights with a lot of assumptions, to evaluate the sensitivity to these assumptions an alternative method is tested.

The alternative method, see formula 10 and 11, begins with the total number of hospitality jobs in Amsterdam, which is scaled to Amsterdam Zuid using the ratio of Zuid's population to the total population of Amsterdam. This assumes that employment in the hospitality sector is roughly proportional to population, providing a reasonable estimate in the absence of Zuid-specific employment data. The scaled employment total is then divided by the number of catering businesses in Amsterdam Zuid, yielding the average number of employees per establishment. Averaging across venues is necessary because detailed employment per venue is not available, and this approach ensures all venues are represented systematically.

To translate staffing levels into estimated customer activity, the number of employees per venue is multiplied by the number of customers that a worker can serve and by the average number of seat turnovers per day. For these variables, the mean values for restaurants and cafés are used. This uniform assumption across venues is a simplification, but it provides a consistent, reproducible approach given that all catering types are aggregated into a single category. While this aggregation eliminates distinctions between different types of catering establishments, it allows the method to operate with open data and to generate weights for all venues in a systematic manner.

$$Employees_{venue} = \frac{Total\ Jobs_{Amsterdam} \cdot \frac{population\ zuid}{population\ amsterdam}}{Nr\ of\ catering\ venues\ zuid} \qquad [\text{persons/venue}](11)$$

$$Weight_{catering} = Employees_{venue} \cdot customers\ per\ worker \cdot seat\ turnover\ per\ day \qquad [\text{persons}](12)$$

Finally, while more precise estimates could be obtained using establishment-specific employment data from the BAG–ABR dataset, such access was not available for this study. Overall, this labour-based approach represents an alternative method to the revenue-based weights, using openly available data to systematically approximate catering-related activity within the study area.

3.2.3.5.3 Alternative Public Transport Weights

Unlike other building types, the public transport weight proxy does not directly reflect daily user volumes but assumes that the number of departures is proportional to the volume of passengers using the stop and provides a reasonable estimate of relative pedestrian activity even though the actual number of passengers is not directly observed. To test how sensitive the model is to these assumptions, the model is first run with the baseline weights, and then rerun with revised weight estimates based on more activity-oriented proxies.

To better align with the user-based proxies used for residential, healthcare, and educational buildings, weights are revised to include vehicle capacity and average occupancy, such that the weight is calculated from the service frequency, the number of lines serving the stop, the vehicle capacity, and the average occupancy (formula 13). Incorporating capacity and occupancy provides a closer approximation of the actual number of passengers passing through a stop daily. An older study by Rietveld (2001) suggests that the average daily occupancy rate in Dutch train services was approximately 35%. While this figure relates to trains and predates 2023, it reflects the structural challenge of aligning vehicle capacity with actual demand and the tendency for services to operate well below full utilisation. In the absence of more recent and mode-specific data, this 35% occupancy rate is assumed to provide a sufficiently accurate proxy for the current analysis.

$$Weight_{PTstop} = \sum_{i=1}^{n} (Frequency_i \cdot Capacity_i \cdot Occupancy\ rate) \qquad [\text{persons}](13)$$

3.2.4 Elastic Building Weights

Initial building weights, assigned in the previous section, represent baseline trip-generating potential but do not account for location or accessibility. In this section, weights are adjusted based on each building's accessibility to different destination types, creating elastic weights.

First, buildings are integrated into the pedestrian network to calculate network-based distances to destinations. These distances are used to compute reach, a measure of accessibility, which then adjusts the initial weights for each destination type. Then the method for adjusting building weights using the Reach index is described as well as the sensitivity approach. The actual implementation and results of the tests are presented in section 5.4.3.

3.2.4.1 Inserting buildings into the pedestrian network

BAG *verblijfsobjecten* (VBOs) and OSM features such as public transport stops and catering amenities are already provided as point geometries. Whereas other OSM facilities such as parks and parking areas are represented as polygon geometries rather than points, requiring the centroid of each feature to be calculated to represent its spatial position. These point and centroid locations serve as consistent geometric proxies when no entrance data is available.

Each centroid is then snapped to the nearest edge of the pedestrian network (Sevtsuk et al., 2025). The nearest edge was determined by computing straight-line distances between the centroid and all street segments, selecting the shortest as the connection (Stack Overflow, 2022; Sevtsuk et al., 2025), see figure 3. The precise point on this edge was then calculated using Shapely geometry functions, and stored with the centroid coordinates, building type, and building weight. This nearest point effectively acts as a proxy for the building's exit, that is, the location from which walking trips are assumed to originate.

It is assumed that these proxy locations provide a sufficiently accurate spatial representation for subsequent analyses, such as network-based accessibility calculations, since accessibility primarily depends on relative distances along the network rather than exact entrance placement. However, some uncertainty remains, particularly for large or spatially complex features such as parks or buildings with multiple exits, where the centroid may not fully capture the true pattern of pedestrian access.

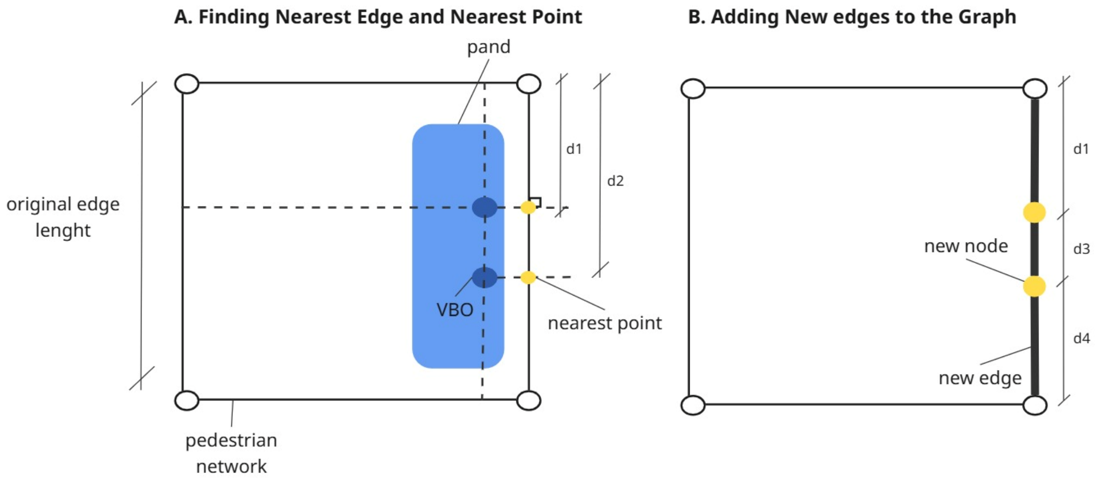


Figure 3: Inserting building point into network

Each nearest point is assigned an unique node identification number. These points are inserted as nodes into the pedestrian graph, see figure xB, with their coordinates and attributes added. The attributes include geometry, weight, and function, as geometry is needed for spatial analysis, while weight and function are essential for trip generation (Stack Overflow, 2022).

To insert the nearest points in the correct order and avoid overlaps or crossing edges, buildings located on the same edge were grouped and connected sequentially. This step follows the methodology of the Medina model (Sevtsuk et al., 2025).

When an edge contained only one new node, it was split into two, with lengths recalculated as straight-line distances between consecutive nodes. If multiple new nodes were located along the same edge, their distances from the edge's starting point were computed, and the nodes were ordered accordingly. Connecting nodes in this way preserves the spatial order of access points, ensuring that shortest path calculations remain accurate and avoiding artificial detours.

3.2.4.2 Reach Calculation

In the BPT-Gen model, the trip-generating potential of each building is made *elastic* with respect to accessibility, reflecting the assumption that locations with better access to destinations generate more walking trips. Building on the behavioural logic of the UNA framework (Sevtsuk et al., 2024), this section details how the initial building weights are adjusted according to accessibility levels within the Dutch pedestrian network.

As explained in Section 3.1, the Reach index was selected as a practical alternative to the more data-intensive gravity-based accessibility measure used in the original UNA model. Here, the methodological implementation of this index is described in detail.

Reach is calculated for each building node individually, ensuring that accessibility adjustments are sensitive to the building's specific position within the pedestrian network. This is necessary to capture spatial heterogeneity, as not all buildings of the same function type are equally connected. Moreover, Reach is computed separately for each destination type, since trip generation in the BPT-Gen model is defined by origin–destination pairs. As a result, each origin receives a set of accessibility-adjusted weights, one for every destination type. These elastic weights are later used to scale the corresponding trip rates for each OD combination, meaning that origins with higher accessibility to a given destination type will generate proportionally more trips of that type.

For example, a residential building with strong access to train stations will generate more train-related trips than one located farther away, while the same building may produce fewer bus trips if it is less connected to bus stops. By structuring the adjustment in this way, the model ensures that trip generation reflects both the functional role of each building and its network accessibility to specific destinations, resulting in more behaviourally realistic estimates.

The formula 14 for the Reach index is given by Sevtsuk et al. (2024) and is computed as the sum of the weights W of all destinations of type t that can be reached along the pedestrian network within a walking radius r, where $d[i,j]$ denotes the shortest-path distance between origin i and destination j. This measure captures the cumulative opportunities available to a building for a given destination type, forming the basis for adjusting its trip-generation weight.

$$Reach_t\,[i]^r = \sum_{j \epsilon G-\{i\}, d[i,j] \leq r} W_t[\,j] \qquad [persons] (14)$$

The Reach indices calculated for each building and destination type are then used to adjust the initial trip-generation weights. Each origin receives a separate adjusted weight for every destination type, which is subsequently used to scale the corresponding trip rates for that origin–destination combination.

3.2.4.3 Model Sensitivity: Reach Normalization and Radius Parameter

This section examines how sensitive the model's adjusted building weights are to two key assumptions: the method used to normalise Reach values and the choice of walking radius in the Reach calculation. Different normalisation approaches are tested to assess how they alter the relationship between accessibility and trip generation, while varying the walking radius evaluates how changes in the assumed catchment size influence Reach outcomes and resulting weights. Together, these tests clarify how methodological choices shape the model's behaviour and robustness.

3.2.4.3.1 Alternatives Reach Normalization

Several normalization formulas are tested to observe how the adjusted weight of the central building changes as the walking radius increases. The goal was to identify which normalization approach best represents the expected pedestrian accessibility behaviour. Meaning that higher perceived attractiveness or utility when more destinations are reachable within a reasonable walking distance. This is done by making a synthetic scenario where buildings of the same weight of a 100.

The first formula observed is a mean-based method where each building's weight relative to the average reach value in the study area. Once the Reach indices has been computed, the original weights are retrieved from the node attributes W_i . To establish a benchmark, the average reach value across all buildings is calculated for each destination type:

$$\overline{Reach}(t) = \frac{1}{N}\sum_{i=1}^{N} Reach(i,t) \qquad (15)$$

where *N* is the number of origin buildings. The adjusted weight of building *i* for destination type *t* is then computed as:

$$W_i^{*(t)} = \begin{cases} W_i \cdot \left(1 + \dfrac{Reach(i,t) - \overline{Reach}(t)}{\overline{Reach}(t)}\right), if\ \overline{Reach} > 0 \\ 0 \qquad\qquad\qquad\qquad , if\ \overline{Reach} = 0 \end{cases} \qquad [personen](16)$$

This ensures that buildings with higher-than-average accessibility to a destination type receive proportionally greater adjusted weights, while those with below-average accessibility are down-weighted. If no destinations of type *t* exist in the study area, the adjusted weight is set to zero, reflecting the fact that no pedestrian trips will be generated toward such destinations. The mean-based method centres values around the average, highlighting relative deviations and making differences between moderately higher or lower accessible buildings more visible ("Normalization, Standardization and Data Transformation", n.d.). This approach is robust to scale differences and does not rely on strict proportionality, making it suitable for diverse data types

The second approach applies a Z-score normalization, which standardizes each building's reach value relative to the mean and standard deviation of all reaches in the study area. Formally, the standardized reach for building *i* and destination type *t* is calculated as:

$$Z(i,t) = \frac{Reach(i,t) - \overline{Reach}(t)}{\sigma(t)} \qquad (17)$$

where σ(t) is the standard deviation of reach values for destination type *t*. The adjusted weight is then defined as:

$$\widehat{W_i^*}(t) = W_i \cdot \left(1 + Z(i,t)\right) \qquad [personen](18)$$

This formulation increases the weight of buildings that are more than one standard deviation above the mean accessibility, while reducing it for those with below-average access.

In the ratio-based normalization, the adjusted weight is scaled directly by the ratio between a building's reach and the mean reach in the study area. The formula can be written as:

$$\widehat{W_i^*}(t) = W_i \cdot \frac{Reach(i,t)}{\overline{Reach}(t)} \qquad [personen](19)$$

The ratio-based normalization scales weights strictly in proportion to reach, producing a linear adjustment that is simple and interpretable. It is most appropriate when the relationship between variables is expected to be strictly proportional, but it can be misleading if this assumption does not hold, potentially creating spurious results or overly smooth gradients (Deng et al., 2019).

In addition to the mean-based, Z-score, and ratio-based approaches, a fourth option is log-based normalisation, commonly used in accessibility modelling to reduce the influence of extreme values (Geurs & van Wee, 2004). Because the synthetic test scenario uses a perfectly homogeneous dataset, a log transformation would have no observable effect.

For the case study dataset, however, theory suggests that the logarithmic function—characterised by rapid growth at low values and flattening at high values—would impose diminishing marginal utility on accessibility (Ewing & Cervero, 2010). This means it would up-weight low-accessibility buildings more strongly and compress differences in highly accessible areas, producing overly smooth gradients.

While behaviourally plausible in some contexts, this smoothing would downplay meaningful distinctions between highly accessible locations, precisely the differences the model aims to highlight. For this reason, log-based normalisation was not considered suitable for application in the case study.

3.2.4.3.2 Alternatives Radius in Reach

As for the parameter walking radius r , the initial run will follow Santana et al. (2022), who suggest that Reach measures can approximate Gravity measures effectively when thresholds align with realistic travel behaviours, the initial value of r is set to correspond to the mean walking distance observed in the Dutch context, derived from the ODiN travel survey to ensure regional relevance.

To assess the sensitivity of the model to this r parameter, the walking radius is later varied to test its influence on accessibility and trip generation outcomes. Specifically, three radii were evaluated: a smaller radius (representing a more localized catchment), the mean walking distance (the baseline based on the observed mean walking distance), and a larger radiusi (representing an extended catchment). For each radius, new Reach values and adjusted building weights are computed, and the model is rerun. The outcomes are compared by examining how the distribution of adjusted building weights changes, how clusters of high-influence origins shifted spatially, and how predicted flows at the counted buildings differed across scenarios.

3.2.5 Trip Rate Computations

Trip rates quantify how many walking trips are generated per person per hour and form a core component of the BPT-Gen model. In this study, trip rates are derived from the ODiN travel survey, with the data preparation and extraction procedure described in section 3.2.1.4 an section 3.2.5. A structured approach is required because walking movements occur both as main-mode trips and as access or egress legs of multimodal trips, meaning that origins and destinations must be inferred differently depending on the trip type.

The section first outlines how origins, destinations, and hourly departures are assigned across five categories of walking trips, ensuring all relevant walking movements are consistently classified. It then presents the formulas used to compute weighted trip rates using ODiN's survey expansion factors. The section concludes with a sensitivity analysis.

3.2.5.1 Calculating the Trip Rates

In order to obtain trip rates for the defined building types from the ODiN dataset, a multi-step procedure is applied that distinguishes between home-based trips, transfer-related walking trips, and other walking trips. This distinction is required because the definition of origins and destinations varies across trip types. For instance, for walking trips where walking is the main mode, origins and destinations can be derived from the purposes of consecutive trips. However, for the first trip of the day, the origin is assumed to be home, as no previous trip exists. For trips where walking occurs only as an access or egress leg of another mode, the main transport mode determines the associated origin or destination. The definition of a trip rate in transportation planning is expressed as the total number of trips divided by the number of "units," where these units can refer to dwellings, households, building area, or other relevant indicators for the study context, as discussed by Takyi (1990) and Pan and Sharifi (n.d.). In formula form:

$$\text{Trip Rate} = \frac{\text{Total Nr of Trips}}{\text{Nr of Units (persons, area, etc)}} \qquad [\text{trips/unit}] (20)$$

Since ODiN is a survey dataset, only a sample of the population is observed. To ensure that the results are representative of the entire population, each respondent is assigned a survey weight. In ODiN, this weight is denoted as FactorP, which indicates how many individuals in the Dutch population the respondent represents (CBS, n.d.).

For the ODiN dataset filtered to areas comparable to the study area, the estimate of how many real-world individuals are represented is calculated as:

$$N = \sum_{i \in \text{unique persons}} \text{FactorP}_i \qquad [\text{persons}] \quad (21)$$

where N denotes the total weighted number of persons and $FactorP_i$ represents the survey weight of respondent i. This value provides the population base against which trip rates per person are derived.

The distinction between home-based trips, transfer-related walking trips, and other walking trips is further refined into five categories to enable precise filtering of the ODiN dataset. Each category requires a different approach for identifying origins and destinations, depending on whether walking occurs as the main mode or as an access or egress leg. For instance, access trips (walking to another transport mode) must be filtered by the *first walking leg* of a trip, whereas egress trips (walking from another mode) require filtering by the *last walking leg*. For trips where walking is the main mode, the *main mode field* is used instead. Defining these five categories ensures that all relevant walking movements are captured and correctly classified.

The first category consists of home-based walking trips, defined as the first trips of the day where the main mode of transport is walking. For these trips, the origin is assumed to be *home*, and the destination is derived from the reported trip purpose, mapped to building types. The corresponding trip rate is given by:

$$R_{t,\text{home}\to d} = \frac{\sum_{i \in \tau_{t,\text{home}\to d}} \text{FactorP}_i}{N} \qquad [\frac{trips}{person \cdot hour}] \quad (22)$$

Where $\tau_{t,home\to d}$ denotes the set of all home-based walking trips departing in hour t with destination type d.

The second category consists of home-based walking legs to transfer points. These are also first trips of the day, and are therefore assumed to originate from *home*. In this case, the main mode of the movement is not walking, but the first leg of the trip is. The destination of the walking leg is defined as the transfer facility associated with the main mode. The trip rate is given by:

$$R_{t,\text{home}\to m} = \frac{\sum_{i \in \tau_{t,\text{home}\to m}} \text{FactorP}_i}{N} \qquad [\frac{trips}{person \cdot hour}] \quad (23)$$

where $\tau_{t,home\to m}$ is the set of all home based first leg walking departing in hour t to transfer facility m.

The third category consists of non-first walking trips, meaning they are not the first trips a person makes during the day, and the main mode of transport is walking. In these cases, the origin is derived from the purpose of the previous trip, and the destination corresponds to the purpose of the current trip. The trip rate is calculated as:

$$R_{t,o\to d} = \frac{\sum_{i \in \tau_{t,o\to d}} \text{FactorP}_i}{N} \qquad [\frac{trips}{person \cdot hour}] \quad (24)$$

where $\tau_{t,o\rightarrow d}$ denotes the set of all walking trips departing in hour t, with origin type o and destination type d.

The fourth category consists of outbound walking legs to transfer points. These trips are not the first movement of the day, and while the main mode of transport is not walking, the first leg of the trip is. Here, the origin is determined by the purpose of the previous trip, and the destination is the transfer facility associated with the main mode. The trip rate is given by:

$$\mathrm{R}_{\mathrm{t,o\rightarrow m}} = \frac{\sum_{\mathrm{i}\in\tau_{\mathrm{t,o\rightarrow m}}} \mathrm{FactorP_i}}{\mathrm{N}} \qquad [\frac{trips}{person \cdot hour}] \tag{25}$$

where $\tau_{t,o\rightarrow m}$ is the set of all outbound walking legs in hour t from origin type o to transfer facility m.

The fifth and final category consists of inbound walking legs from transfer points. In this case, the main mode of the movement is not walking, but the last leg of the trip is. The origin is therefore the transfer facility associated with the main mode, and the destination is the reported trip purpose. The trip rate is calculated as:

$$\mathrm{R}_{\mathrm{t,m\rightarrow d}} = \frac{\sum_{\mathrm{i}\in\tau_{\mathrm{t,m\rightarrow d}}} \mathrm{FactorP_i}}{\mathrm{N}} \qquad [\frac{trips}{person \cdot hour}] \tag{26}$$

where $\tau_{t,m\rightarrow d}$ denotes the set of all inbound walking legs departing in hour t from transfer facility m to destination type d.

All five categories are merged into a single dataset of hourly trip rates, defined as:

$$\mathcal{R} = \{\mathrm{R}_{\mathrm{t,o\rightarrow d}} | \mathrm{t} \in [0,23], \mathrm{o} \in \mathrm{O}, \mathrm{d} \in \mathrm{D}\}$$

where O and D are the sets of possible origin and destination categories, respectively.

3.2.5.2 Model Sensitivity: Effect of Regional Filtering

To examine how spatial filtering influences the estimated walking trip rates, two alternative subsets of the ODiN data are used.
First, a narrow filter selects postcode areas that closely resemble the case study context—major Dutch station-area business districts such as Amsterdam Zuidas, Utrecht Centraal, Rotterdam Kop van Zuid, The Hague Beatrixkwartier/Binckhorst, Amsterdam Sloterdijk, Leiden Stationsgebied, and Eindhoven Stationsgebied. This subset reflects walking behaviour in dense, transit-oriented office districts.

Second, a broader municipality-level filter includes a wide set of large urban municipalities to increase sample size and reduce irregularities observed in the narrow subset. This broader filter is intended to provide more stable estimates of general urban walking behaviour.

For each filter, hourly origin–destination trip rates are calculated and later compared (section 5.4.3) to assess the sensitivity of the model to the choice of spatial selection.

3.2.6 Calculating Generated Trips

In this step, the previously derived trip rates are applied to the adjusted building-level weights in order to estimate the number of trips generated by each node in the study area. Trip generation is performed per node, per hour, and per destination type, ensuring that the temporal and functional dimensions of pedestrian travel are captured.

For each building node of a given origin type, the number of trips generated for a given hour, and destination type is calculated by multiplying the trip rate for that origin–destination pair by the node's accessibility-adjusted weight for the corresponding destination type.

$$G_{n,t,o\to d} = W_{n,d} \cdot R_{t,o\to d} \qquad [trips/hour] \tag{27}$$

Where $G_{n,t,o\to d}$ is the number of generated trips at node *n,* departing in hour *t* from origin type *o* to destination type *d,* and $W_{n,d}$ is the weight of the node for destination type *d*. This approach follows standard trip generation principles, where the number of trips is proportional both to the size or activity level of the origin (its weight) and to the likelihood of trips per unit (the trip rate). A key difference in this study is that trip rates are differentiated by destination type. This distinction is chosen because pedestrian trip patterns vary not only by origin but also by destination: the likelihood, distance, and duration of trips can differ substantially depending on the land use of the destination. For example, trips to a nearby grocery store may be shorter and more frequent than trips to a workplace or transit hub. By making node weights destination-specific, the model preserves these variations, ensuring that pedestrian activity is accurately represented across different origin destination pairs. Additionally, the weights have been adjusted to account for each origin's accessibility to destinations, further refining the spatial realism of trip generation.

The output is a table of generated trips per node, per hour, and per destination type. This dataset provides the spatial and temporal distribution of pedestrian trip generation within the study area, and serves as input for the subsequent trip distribution stage.

As a final step, the total number of trips generated per node is calculated by summing across all destination types. This provides a straightforward consistency check to validate the plausibility of the generated trips.

3.3 Model Face Validation

Because BPT-Gen does not include the trip distribution and assignment components present in the UNA framework, and due to the limited availability of pedestrian count data, the validation and sensitivity analysis must follow an adapted approach. The subsequent section outlines the methods used to assess model performance and robustness under these constraints.

To validate the model, the framework is applied to a case study (see Chapter 4). The resulting hourly generated trips are compared with manual pedestrian counts collected on-site. These manual counts record the number of pedestrians entering or exiting selected buildings during predefined observation periods.

For the selection of count locations, best practice identifies four approaches: (1) representative community-wide sites, (2) sites linked to a specific project or facility, (3) control sites unaffected by changes to be compared with affected sites, and (4) randomly sampled locations within defined categories (Ryus et al., 2022). In this study, count sites were not selected to be representative of the entire neighbourhood, as this would require more extensive data collection than was feasible. Instead, sites were deliberately chosen at buildings expected to generate high pedestrian trip volumes (i.e., buildings with high trip rates or weights) and where flows could be observed from a single vantage point (ensuring all entrances and exits were visible). This prioritizes locations where model uncertainty is highest and where validation observations are most informative.

Guidelines recommend covering four peak periods for a total of eight hours, with no more than two consecutive hours counted at a time. In this framework, these principles are followed by observing the morning peak, the midday/lunch period, and the afternoon peak. The evening peak is excluded due to safety considerations at the study site.

Count sites, see figure 4, are selected to focus on building types with the highest trip rates: home, work, leisure, education, and train stations. Due to the counting being conducted in the summer period, education buildings were replaced with healthcare facilities, and the train station category was substituted with a bus stop because municipal data were already available for stations and ODiN indicated low bus/tram trips. Shopping was excluded due to time constraints. Within each category, high-weight buildings were prioritized, as they are expected to have the greatest influence on overall pedestrian flows and clearer observation points.

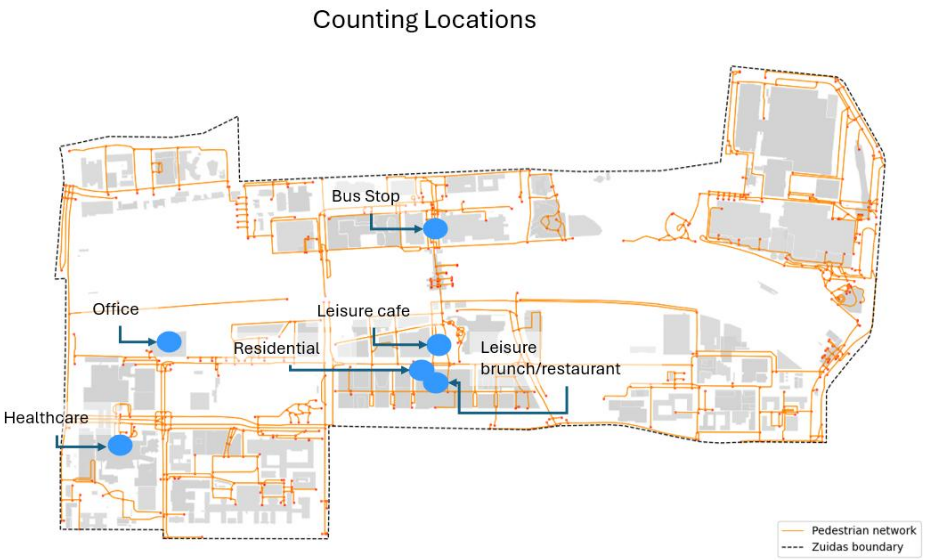


Figure 4:Manual Count Locations

Figure 5 illustrates the locations of pedestrian street counts conducted by the municipality of Amsterdam, with the Amsterdam Zuid train station located between counting points 7 and 8. Counts in the direction away from the station were summed and compared with the model's predicted pedestrian volumes (see table 9), using the mean over twelve weekdays between late September and late October 2024.

Some uncertainties are introduced with this counting method. According to Ryus et al. (2022) manual methods using paper sheets or clickers can underestimate pedestrian volumes by 8–25%. Additionally, the counts for this report are conducted during the summer vacation period, when pedestrian behaviour may differ from other times of the year. Pedestrian flows are also highly variable from day to day, and short-duration counts may not fully capture average patterns. Due to resource limitations, counts were limited to two-hour periods, as longer sessions risked observer fatigue and increased uncertainty. Without continuous counting equipment, the resulting data should therefore be interpreted as indicative rather than precise (Ryus et al., 2022). Consequently the small sample size, limited temporal coverage, and contextual factors such as weather, holidays, and daily variability mean the results are not statistically representative and are intended as a check on the model's performance and not as formal validation. Nevertheless, observing a subset of high-weight nodes provides a useful indication of overall model behaviour and highlights potential underreporting in the ODiN survey data, while illustrating how local conditions influence pedestrian activity.

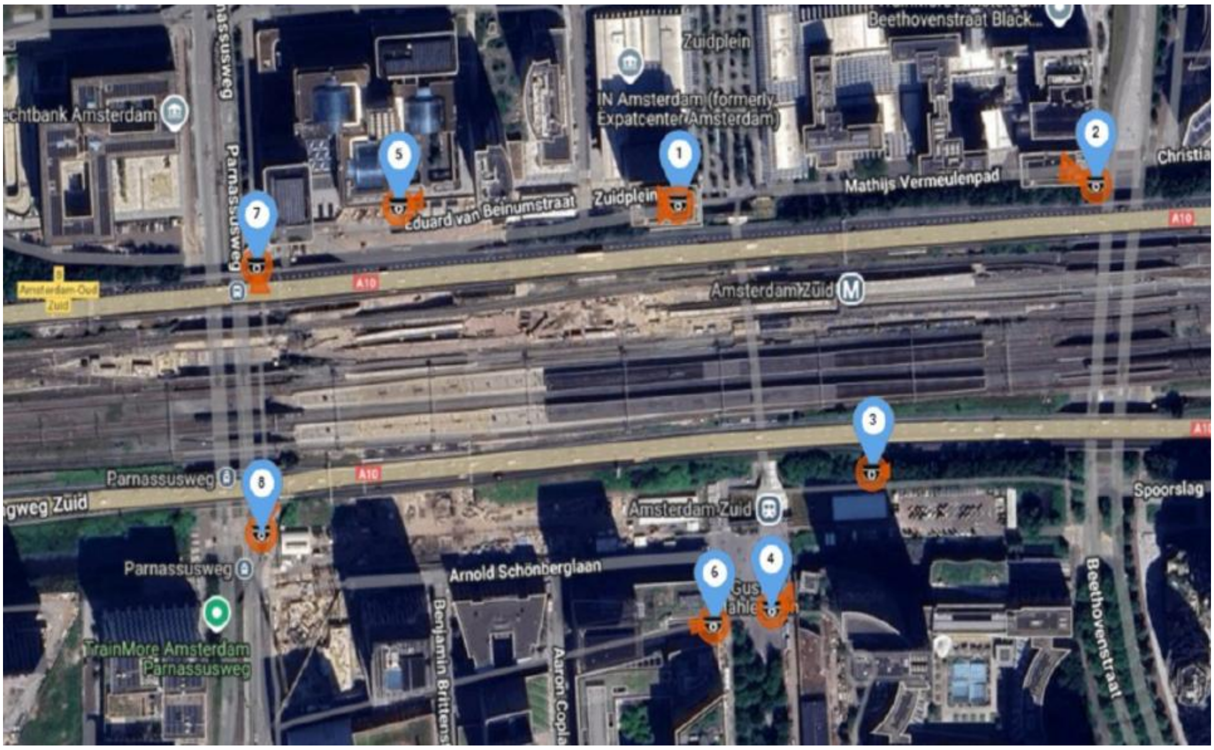


Figure 5: Location Counts Conducted by the Municipality

Since the collected manual data is not enough for a proper validation the validation applied here is a plausibility validation. The objective is to assess whether the model produces trip volumes of the correct order of magnitude, temporal profile, and whether relative differences between origin types in the model align with real-world behaviour. For each counted building, observed pedestrian flows are compared with the model-generated hourly trip estimates for the corresponding time periods.

For each observed location and time period, the modelled number of generated trips was extracted for the corresponding origin type and hour. To compare observed and predicted flows, a simple ratio metric was used:

$$Factor = \frac{observed\ (count)}{predicted\ (model)} \tag{28}$$

Where a factor bigger than one indicates that the model underpredicts pedestrian activity and a factor smaller than one indicates that the model overpredicts pedestrian activity.

This comparison provides a scaled measure that is interpretable across different building types and trip magnitudes. Because the counts cover only a subset of nodes and conditions vary by day and season, the validation is interpreted as a plausibility check, not a formal accuracy assessment.

4. Case Study Description

This chapter introduces the case study area in which the pedestrian trip generation model is applied. The case study serves to ground the methodological framework in a real-world urban environment, demonstrating its practical applicability. Zuidas, located in southern Amsterdam, was selected as the case study area. The choice is motivated by its integration within the TNO project as well as its highly mixed use character and the train station located in the area. These qualities make Zuidas both a relevant and challenging environment for testing a building level pedestrian trip generation model.

Zuidas is a mixed-use district in southern Amsterdam, spanning about 245 hectares. The area features a mix of high-rise office buildings and major institutions, including Vrije Universiteit Amsterdam, Amsterdam UMC, and the RAI Convention Centre. It is bordered by the neighbourhoods of Buitenveldert and Oud-Zuid, with the A10 ring road, the Nieuwe Meer lake, and the Amstel River forming its natural and infrastructural boundaries (Zuidas, n.d.).

In the figure below, the district's layout is illustrated: the main train station is located centrally within the 'Dokzone' (the red zone), the RAI is in the northeast (the brown zone), and the VU University and Amsterdam UMC are in the southwest (the pink zone). Consequently, two major pedestrian flows are expected, one moving from the train station towards the RAI and another towards the university and hospital.

Figure 6: Zonal map of Amsterdam Zuidas, reprinted from Zuidasdok (https://www.planviewer.nl/imro/files/NL.IMRO.0363.K1402BPGST-OW01/t_NL.IMRO.0363.K1402BPGST-OW01.html)

As of 2023, Zuidas was home to over 5,300 residents, 50,000 employees, and 32,000 students, underscoring its role as a dense, multifunctional urban hub (Zuidas, n.d.).
An overview of building types based on development plans and Google Maps shows clear spatial patterns. Residential buildings are primarily concentrated in the orange zone (see figure 6), where most early-morning trips are expected to originate as people commute to work or study. Office buildings are more evenly distributed but are densest around the train station, an area associated with late-afternoon outbound trips as employees return home. Catering facilities are

dispersed throughout the district, contributing to leisure-related movement, along with the large park located in the northern part of the area.

To gain a clearer understanding of how different building types are distributed, figure 7 presents clusters of building types within the study area. Each cluster groups buildings with similar characteristics or spatial proximity, and the accompanying bar charts show the three largest origin types per cluster based on gross floor area (GFA).

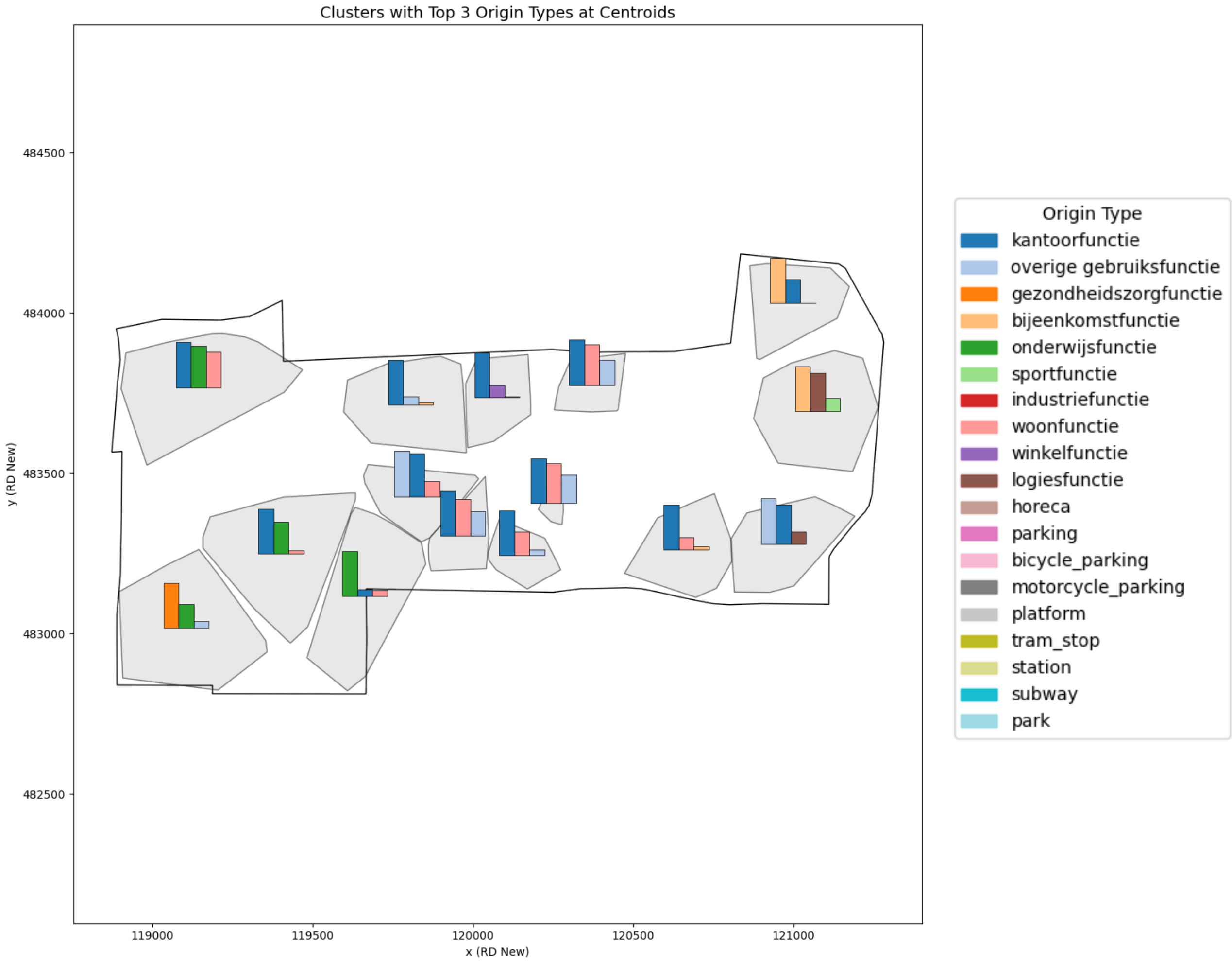


Figure 7: Clusters with most common building types

The results indicate that residential buildings are indeed primarily concentrated in the eastern clusters, with additional residential areas located north of the railway tracks. Office buildings dominate most clusters, while residential functions remain more localized. The spatial pattern also reveals a clear division not only between the northern and southern parts of the area, defined by the A10 ring road and railway tracks, but also between the eastern and western sections. The eastern side is characterized by educational and healthcare institutions, whereas the western side hosts hotels and convention centres.

Each cluster exhibits a degree of mixed use, suggesting that various trip purposes contribute to activity throughout the day. Although catering, sports, and transit-related buildings appear less prominent in the clustering visualization, these functions are distributed across the district, supporting diverse urban activities and contributing to the continuous movement patterns that define Zuidas.

The diversity of land uses within Zuidas, combined with Amsterdam Zuid Station as a major multimodal transit hub, generates a wide range of pedestrian activity associated with commuting, education, leisure, and access to public transport. The extent of this activity, however, also depends on the quality and connectivity of the pedestrian network, which determines how easily destinations can be reached on foot.

The physical structure of the area is strongly influenced by the A10 motorway and the railway corridor, which divide Zuidas into northern and southern sections. These are connected by several viaducts and underpasses that provide pedestrian access across the barrier. Analysis of the pedestrian network, which includes only walkable paths within the study area (figure 8), indicates that Zuidas is generally well connected. Limited connectivity is observed in sections where no viaducts cross the A10; nevertheless, the network ensures that all buildings within the district remain accessible to one another by foot.

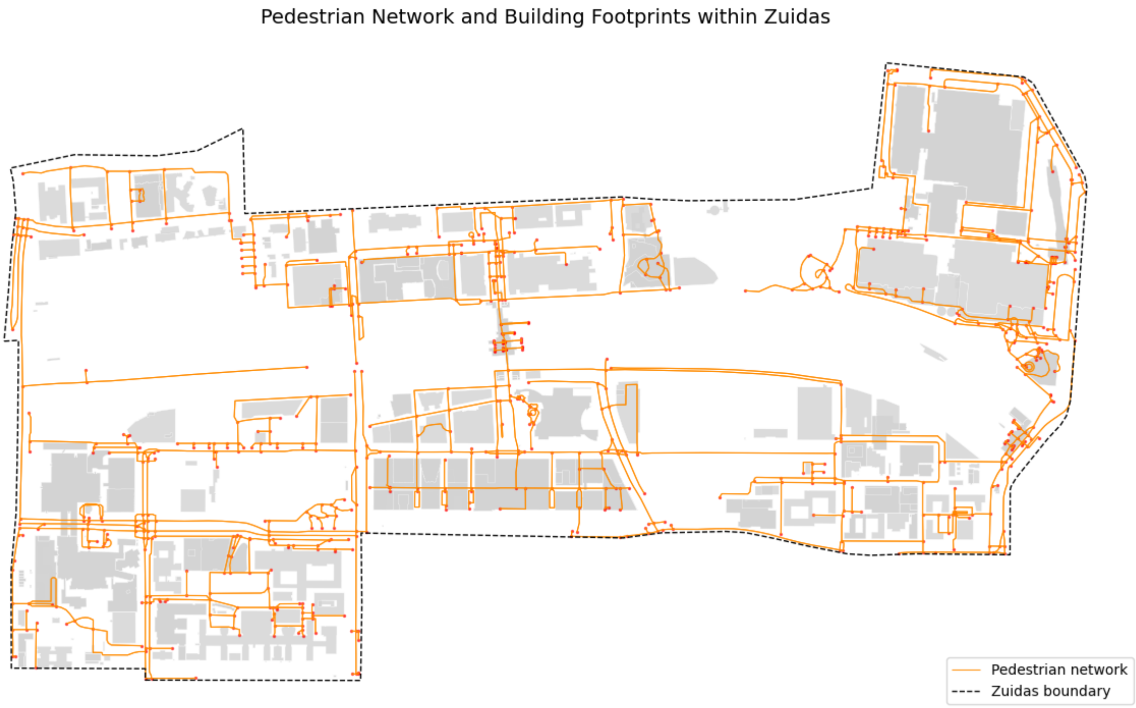


Figure 8: Pedestrian network Zuidas

The level of pedestrian activity in Zuidas is influenced not only by its diverse mix of land uses and strong public transport connections but also by the walking behaviour typically observed in comparable urban contexts. National insights from the Netherlands Institute for Transport Policy Analysis (KiM) reports average walking distances of approximately 1.3 km for work trips, 2.0 km for leisure trips, and less than 1 km for other trip purposes (de Haas et al., 2019).

To better understand how these patterns translate to dense, mixed-use urban districts such as Zuidas, the national travel survey ODiN was examined for areas with similar characteristics, Amsterdam Zuidas, Utrecht Centraal, Rotterdam Kop van Zuid, The Hague Beatrixkwartier/Binckhorst, Amsterdam Sloterdijk, Leiden Stationsgebied, and Eindhoven Stationsgebied. This subset included 325 unique individuals. Across all trips, the mean walking distance was 1.70 km, decreasing to 0.93 km when trips longer than 3 km were excluded. These values are slightly higher than those reported by KiM, possibly due to underreporting of very short walking trips, overestimation in the sample, or contextual differences in the urban environment.

Given Zuidas's mixed-use character, with a high concentration of employment and education functions, and well-connected pedestrian infrastructure, substantial walking demand is anticipated throughout the day. The highest pedestrian volumes are expected during commuting hours, morning and late afternoon, driven primarily by work and education trips. Additional peaks are likely around midday and in the evening, reflecting leisure, shopping, and dining activities. The area's high network connectivity further supports continuous pedestrian movement, reinforcing its function as a walkable urban hub.

The ODiN data largely confirm these expectations to a large extent. When looking at the trip purpose distribution in figure 9, it becomes clear that home- and leisure-related walking trips dominate, followed by shopping and work trips. This hierarchy aligns with national findings from KiM, where leisure trips are also identified as the most common purpose for walking. The inclusion of walking legs within multimodal journeys in ODiN explains slightly higher shares for certain trip purposes compared to KiM, which considers only stand-alone walking trips.

The temporal distribution of walking activity also aligns with expected behaviour. A clear morning peak around 08:00 corresponds to work and school commutes, a midday rise around 12:00 reflects lunch and short leisure activities, and a pronounced afternoon peak between 16:00 and 18:00 marks the end of the working day. A smaller evening peak around 20:00 relates to social and dining activities. These temporal patterns are consistent with trends observed in mobile phone data analyses (Jacobs-Crisioni et al., 2014), reinforcing the plausibility of the modelled pedestrian dynamics in Zuidas.

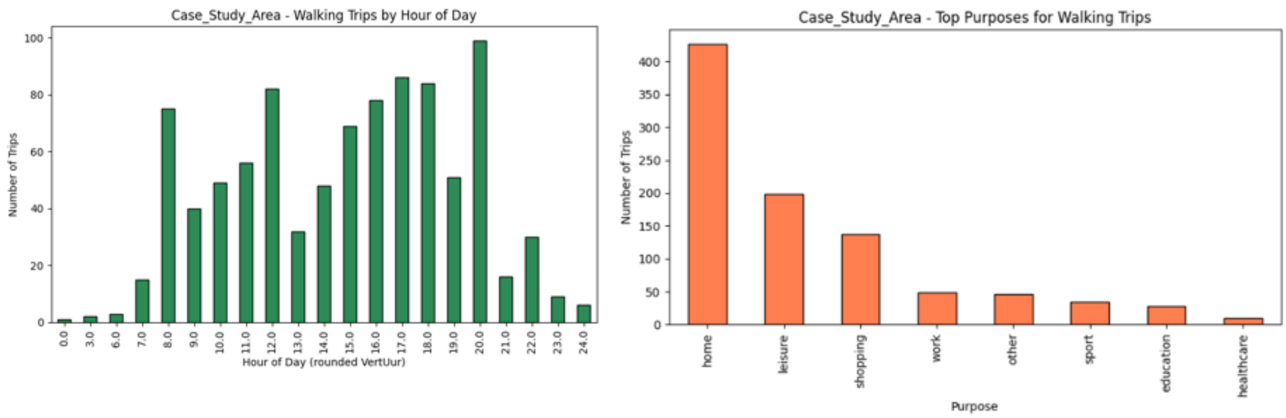


Figure 9: Distributions of departure hour and trip purpose for Zuidas

5. Results

This chapter presents the outputs of the BPT-Gen model described in Chapter 3, following the stepwise structure of the methodology. Section 5.1 examines the scraping of data and the combining of BAG and OSM building data, assessing their accuracy and spatial representation. Section 5.2 looks at the results of mapping trip purpose categories to building types. Section 5.3 presents building weights, showing their distribution by type and comparing different weighting methods to evaluate plausibility. Section 5.4 focuses on elastic building weights, examining how the insertion of building nodes into the network affects reach-based adjusted weights and how varying the walking radius influences accessibility. Section 5.5 discusses trip rates, including the impact of reclassified building types, potential underreporting in ODiN, and differences between study-area-specific and full national datasets. In Section 5.6, building weights and trip rates are combined to generate hourly trips per building, illustrating spatial and temporal patterns of pedestrian activity. Finally, section 5.7 validates the model outputs by comparing generated demand with observed pedestrian counts, assessing whether the results reflect the expected magnitude and temporal dynamics. Together, these analyses provide a comprehensive evaluation of the model's intermediate and final outputs and the influence of methodological choices on pedestrian trip estimates.

5.1 Scraping of the Building Data

This section outlines the scraping and preparation of all datasets used as inputs for BPT-Gen. Building information was collected from OSM and BAG (sections 5.1.1–5.1.2), providing footprints, functions, and floor area for all origins and destinations. The pedestrian network was extracted from OSM to construct the walking graph (section 5.1.3). Finally, walking trips from the ODiN survey were scraped and filtered to derive empirical trip rates for areas comparable to the Zuidas context (section 5.1.4). Together, these datasets form the basis for the model's building weights, accessibility calculations, and trip rate estimation.

5.1.1 OSM scraping for Building Information Results

To evaluate the reliability of the inputs for BPT-Gen, the results first focus on the building data that define all potential origins and destinations. Since the model depends on accurate spatial locations, functional classifications, and size attributes, the OSM datasets was analysed to determine their completeness and consistency within the buffered study area. This section summarises the structure of each dataset, highlights their differences, and identifies how they complement each other in forming the basis for subsequent building weighting.

The case study area boundary was first established using OSM and expanded with a buffer of 1,100 m, corresponding to the mean walking distance calculated from ODiN data. This distance is considered a realistic threshold for capturing walking trips that enter or exit the area, in line with conclusions drawn in the KiM report on walking.

The OSM building dataset contains 16,680 footprints within the buffered study area. Each record includes geometry, a function label, and footprint area, but no gross floor area (GFA). Apartments make up the largest share (≈7,500 units, 45%), a distribution that aligns with municipal statistics and visual checks using Google Maps, see figure 10. Building sizes also vary widely, from small ancillary structures of only a few square metres to very large complexes exceeding 61,000 m^2. This confirms that OSM captures the overall diversity of building forms well, ranging from small residential units to major office complexes. However, functional detail remains limited: only 1.2% of records contain a building name, and OSM does not represent mixed-use buildings. For example, an office and residential function within the same structure cannot be distinguished in OSM, even though this is a defining characteristic of Zuidas.

The OSM amenities dataset contributes 3,466 features across four categories: parking, catering, leisure (parks), and other local facilities, see figure 11. These functions are not represented in BAG VBO, making OSM a valuable complementary source for non-residential trip generators. Parking dominates the dataset with 74% of entries, followed by catering-related amenities. Most features are small in footprint (median 10 m^2), but the inclusion of very large facilities, such as multi-level garages exceeding 18,600 m^2, demonstrates that OSM captures both small-scale and large-scale generators. This is particularly relevant for parking, where even small facilities may produce noticeable pedestrian demand.

Public transport infrastructure is also well represented in OSM, with 193 features identified. The majority consist of bus platforms (≈61%). These facilities highlight OSM's added value in providing access and egress points for pedestrian trips, supplementing the building-focused perspective of BAG.

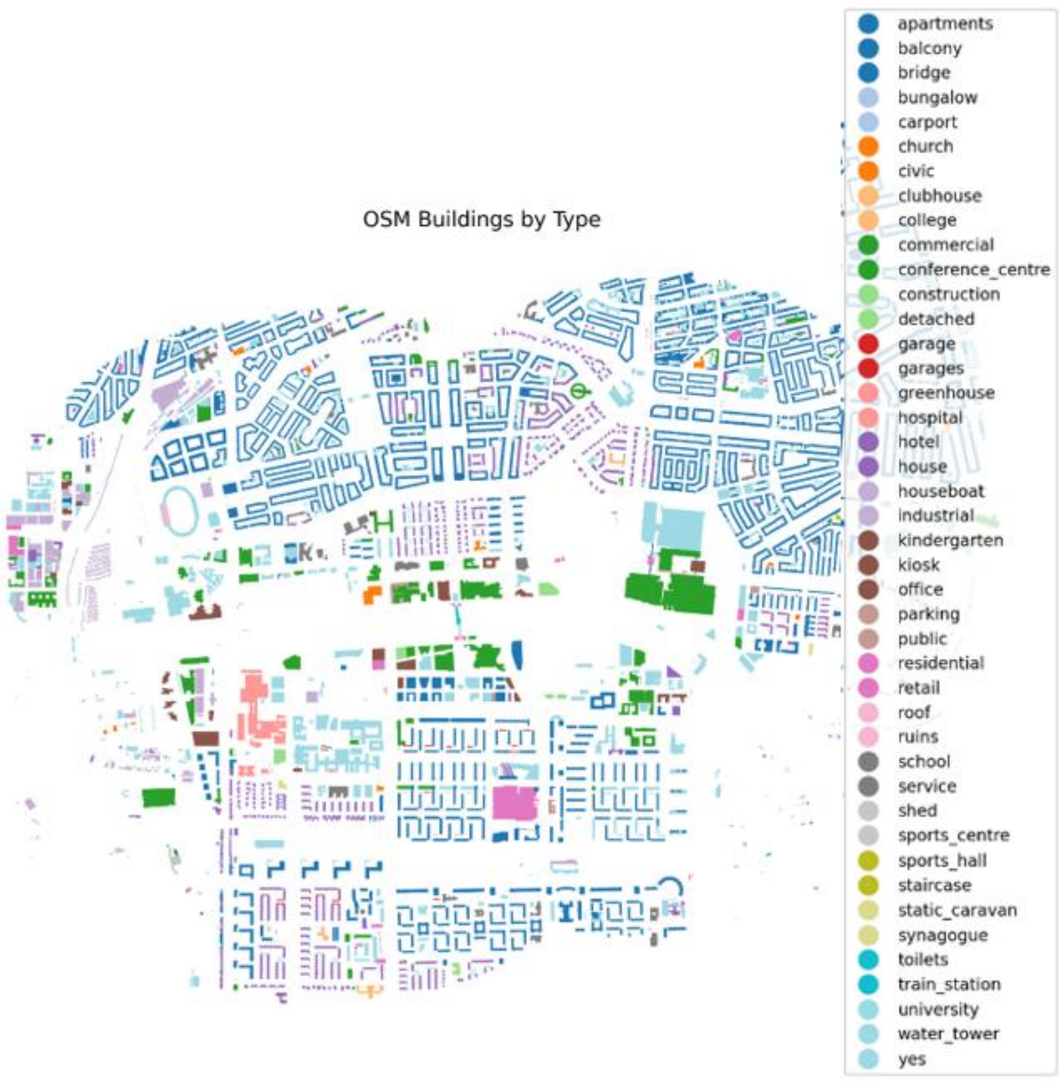


Figure 10: OSM Building Types

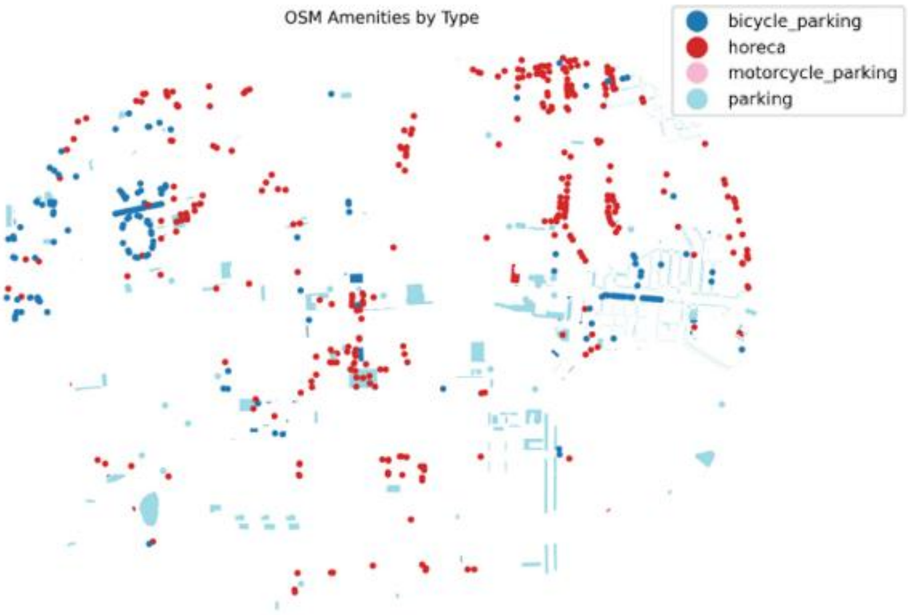


Figure 11: OSM amenities

The OSM datasets generally reflect the character of the study area, as confirmed through comparisons with Google Maps and on-site observations. However, their lack of GFA information and inability to represent mixed-use buildings limit their suitability for trip-generation modelling. For this reason, the BAG dataset is also examined and used as the primary source. OSM then serves as a complementary dataset, mainly providing functions not included in BAG, such as parking facilities, public transport stops, and parks.

5.1.2 BAG scraping for building information results

The BAG VBO dataset provides 59,464 individual building-use units, making it the most detailed and functionally rich source for this study, see figure 12. Forty-eight usage categories are represented, with residential units dominating, while offices, retail, and other functions make up the remaining 13% concentrated in the Zuidas core. This distribution also corresponds with field impressions and visual comparison against Google Maps. Crucially, BAG provides GFA for every record, allowing a direct link between building size and estimated activity. Moreover, BAG captures mixed-use structures by linking multiple use units (VBOs) to a single building footprint (pand), enabling more accurate representation of the co-existence of residential, office, and retail functions within the same building. The dataset includes 11,544 unique building footprints in total, providing a spatial anchor for these functions.

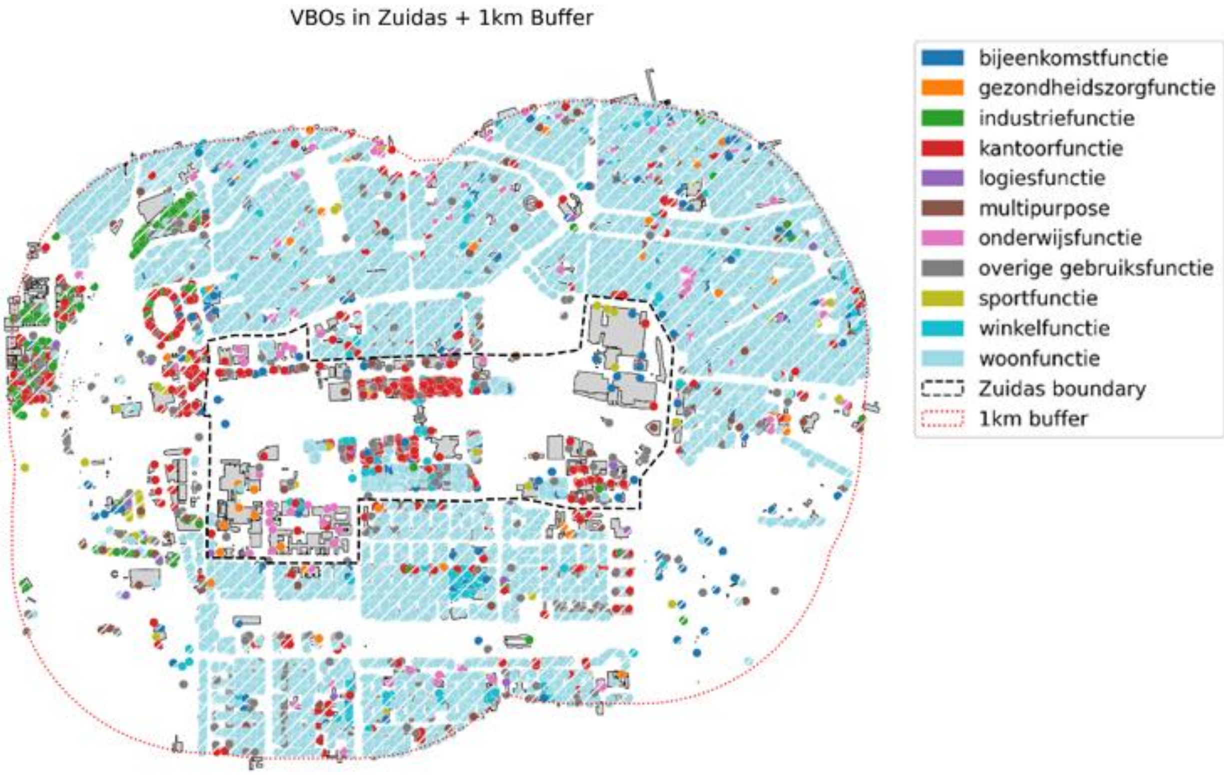


Figure 12: BAG VBO building types

BAG therefore forms the primary basis for building-level analysis, while OSM serves mainly as a complementary source for functions not covered in BAG, such as parking facilities, public transport stops, and parks, see figure 13.

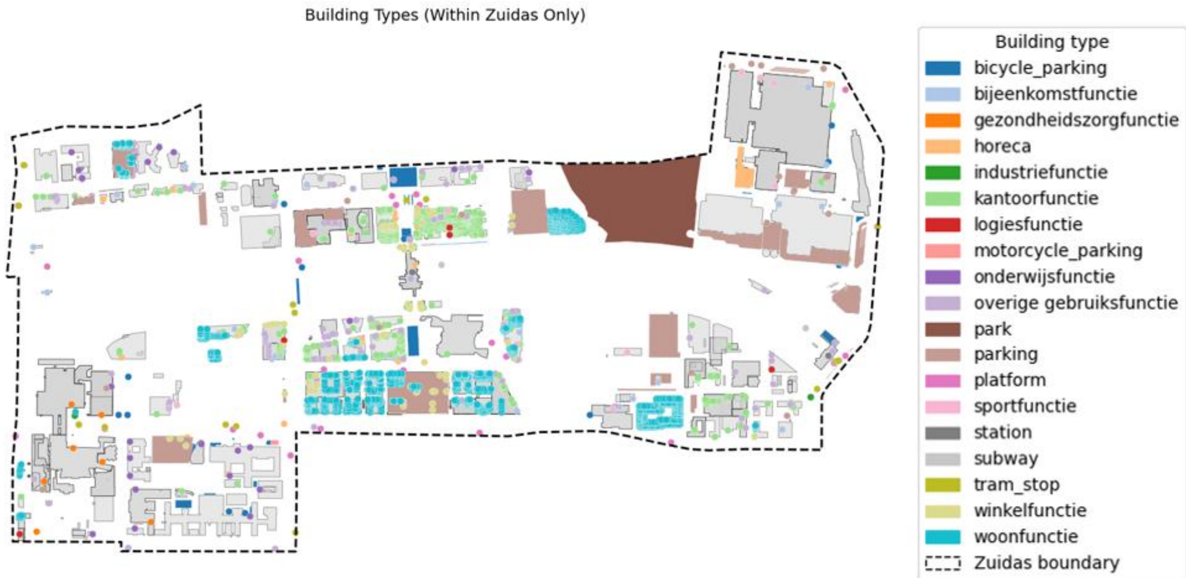


Figure 13: Combined BAG and OSM dataset

The integrated dataset with BAG as the primary source of building functions and floor area with OSM amenities and public transport locations relies on the methodological assumptions that BAG provides correctly measured gross floor area (GFA) and accurate allocation of floor space across multiple building-use units, and that aggregating OSM amenities into broader functional categories does not distort their role as pedestrian trip generators.

Visual comparison with Google Maps, supported by on-site observations, indicates that the resulting map aligns well with the actual land-use structure. This supports the assumption that the integrated dataset is sufficiently complete and accurate for identifying potential trip-generating locations. BAG delivers the necessary functional detail, particularly for mixed-use buildings, while OSM adds complementary coverage of amenities and transport access points. Together, these datasets provide an adequately accurate representation of the study area for subsequent modelling.

5.1.3 OSM scraping pedestrian network

The pedestrian network extracted from OSM provides a dense and continuous representation of all walkable paths within the buffered study area. After filtering the raw OSM data to retain only links accessible to pedestrians (e.g., footpaths, sidewalks, shared streets), the resulting network consists of *814* nodes (representing points where pedestrians can change direction) and *2016* edges (figure 8). Visual inspection confirms that the network aligns well with the observed street and footpath structure from Google Maps and on-site observations. In line with previous studies showing that OSM offers sufficient positional accuracy and completeness for pedestrian network analysis in European cities (Zhou, 2022; Zhou et al., 2022), the extracted network is assumed to be reliable for the reach calculations performed in this study. While small gaps or local inconsistencies cannot be ruled out, common limitations in volunteered geographic information, the available evidence indicates no major missing segments or misclassifications that would significantly affect accessibility outcomes. The resulting network is therefore considered an adequate basis for computing building-level Reach values in the subsequent analysis.

5.1.4 ODiN scraping for trip rate data

The filtered ODiN dataset, for weekdays in areas similar to the case study area, was evaluated in terms of sample size, walking distance distributions, and trip purposes to assess its reliability and alignment with expected pedestrian behaviour, as well as to identify potential underreporting of certain trip types. The dataset was restricted to postal code areas comparable to major Dutch urban centres with characteristics similar to the Zuidas area, including Amsterdam Zuidas, Utrecht Centraal, Rotterdam Kop van Zuid, The Hague Beatrixkwartier/Binckhorst, Amsterdam Sloterdijk, Leiden Stationsgebied, and Eindhoven Stationsgebied. After filtering, the sample consisted of 325 unique individuals and 708 walking trips, providing a modest but focused dataset for analysis. The evaluation of ODiN was framed in the context of prior studies, including KiM reports on ODiN data, and linked to Jacobs-Crisioni et al.'s framework, which associates human activity patterns and mobile phone usage densities with land-use types such as workplaces, retail, leisure facilities, and residential areas.

The walking distance distributions, and trip purposes of Zuidas and similar areas are given in the case study chapter and align reasonably well with expected pedestrian behaviour, trip rates per hour and per origin–destination (OD) type were calculated according to the methodology outlined in chapter 3.

5.2 Mapping Trip Purpose Odin to Building Type

More detailed BAG and OSM labels were consolidated into broader origin–destination categories to match the trip rate data. This aggregation ensures consistency between the building dataset and travel behaviour data (see figure 14). While it is assumed that mapping detailed uses into coarser land-use categories does not critically distort model outcomes, heterogeneity within categories remains, and this assumption cannot be fully verified within the current analysis.

Figure 14: Building categories consolidated to match trip purposes

5.3 Assigning Building Weights

The assigned weights represent a proxy for the relative intensity of human presence and activity associated with each building throughout the day. Visualisation of the resulting patterns shows substantial variation across building types, reflecting differences in function, scale, and the nature of the underlying data sources. For mapping purposes, VBO-level weights for office and residential buildings were aggregated to the *pand* level to improve readability. A complete overview of weights for all building types is provided in Appendix 1.

5.3.1 Weights based on study-area-specific data

The first group of weights, residential, educational, healthcare buildings, and the train station, relies on locally available quantitative data.

Residential buildings, see figure 15, receive comparatively low weights (avg. ≈ 1.4 per VBO), with variation primarily driven by differences in gross floor area, reflecting the assumption that population totals are distributed proportionally across residential GFA. This produces realistic values for a young, housing-dominated district.

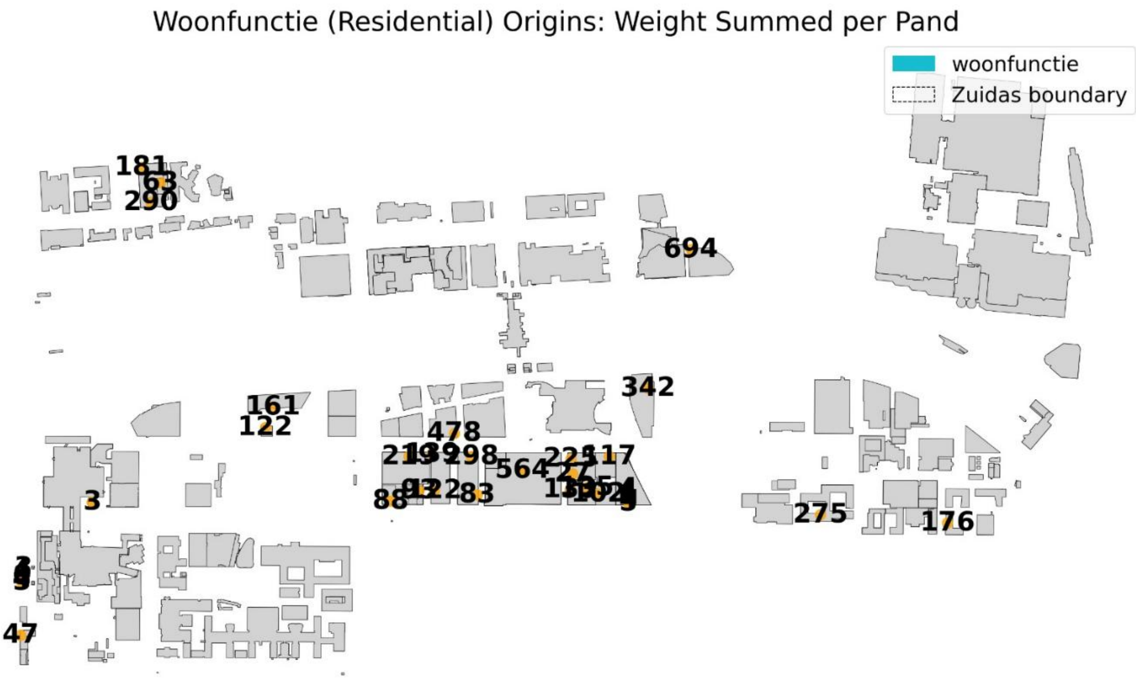


Figure 15: Residential building weights

Educational buildings, see figure 16, show substantially higher weights (max ≈ 13,950), consistent with enrolment figures and staff counts obtained from AlleCijfers.nl and institutional sources. Healthcare buildings follow the same logic, where patients and workers are distributed over the available GFA; however, the method cannot distinguish variation in intensity between facilities.

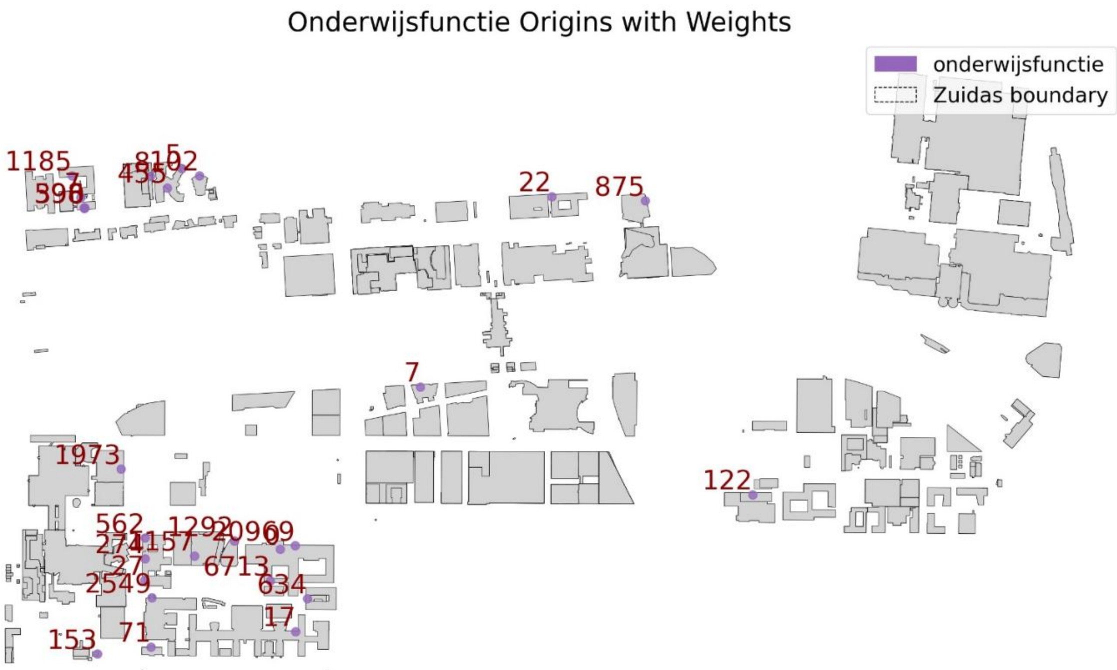


Figure 16: Education building weights

Amsterdam Zuid station stands out with the highest weight (≈ 28,901). This value is directly derived from NS passenger counts and therefore represents the most reliable indicator of daily human activity in the dataset, see figure 17.

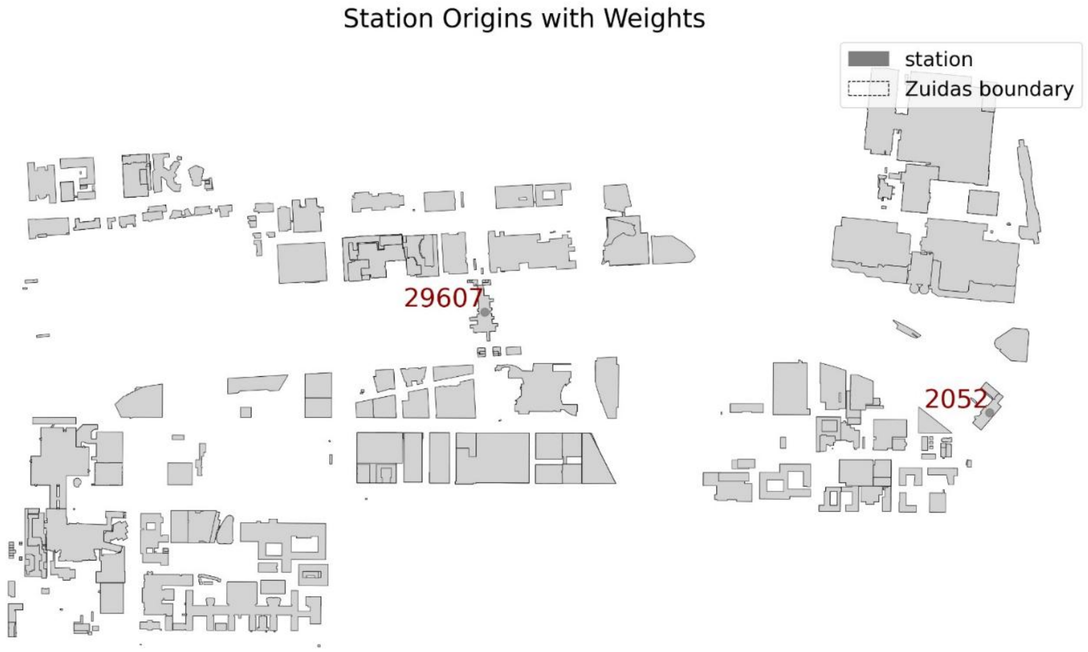


Figure 17: Train station weights

5.3.2 Weights based on Dutch building standards

The second group includes building types for which weights were derived using national building standards combined with GFA: offices, sports facilities, assembly buildings, and parks. For offices two average daily occupancy rates were tested, see section 5.3.5.1, the chosen alternative is presented in figure 18, which is the alternative with the peak daily occupancy rate. Office buildings receive moderate to high weights depending on floor area, reflecting expected worker densities.

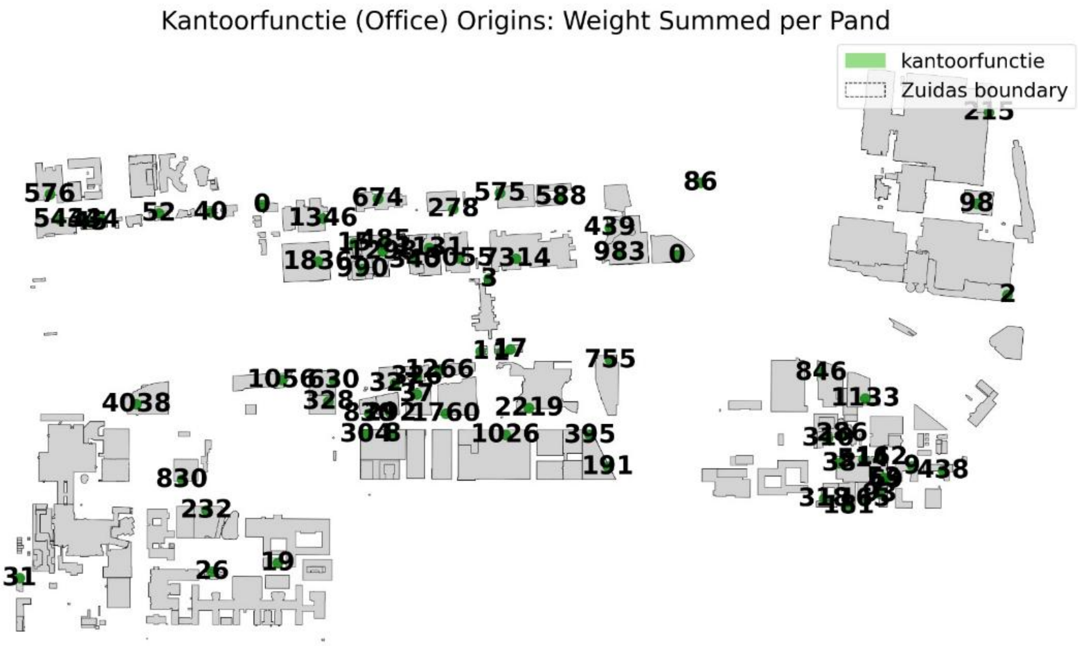


Figure 18: Office building weights with peak daily occupancy rates

Sports and assembly functions, however, display very large weights (sport max ≈ 8,413; assembly ≈ 9,090), see figure 19 and 20. These high values reveal a limitation of the method: applying uniform GFA-based standards to vastly different facility types (e.g., small gyms vs. stadium-like halls) does not capture variation in actual utilisation.
Parks form the extreme upper tail of this category, with weights exceeding 448,000 due to their large surface areas and assumed recreational activity levels.

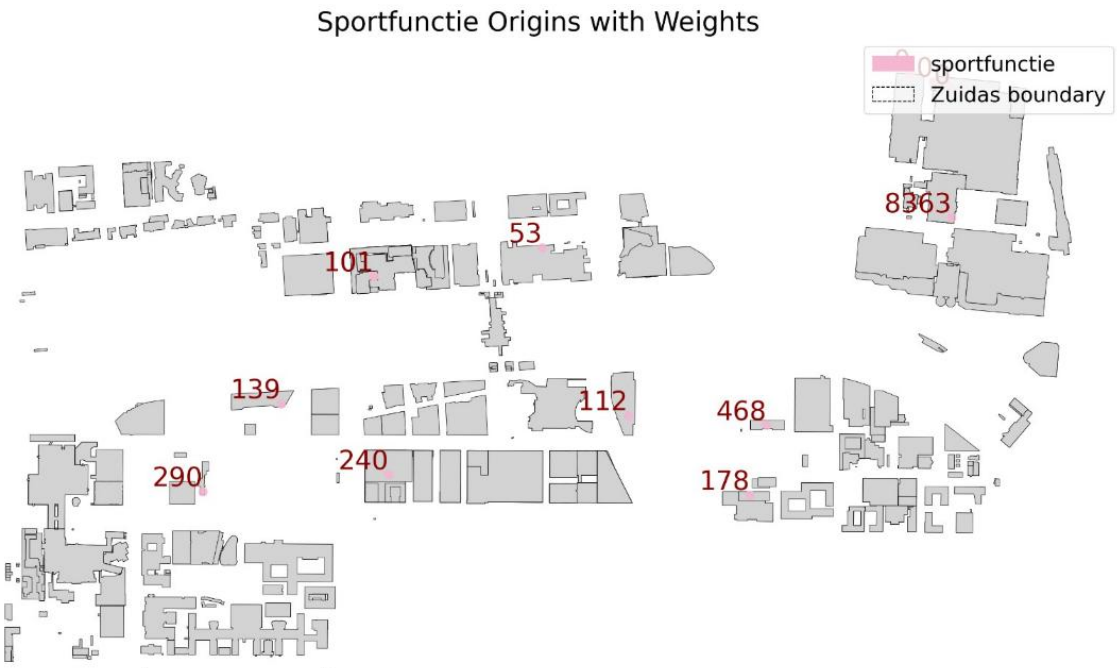


Figure 19: Sport building weights

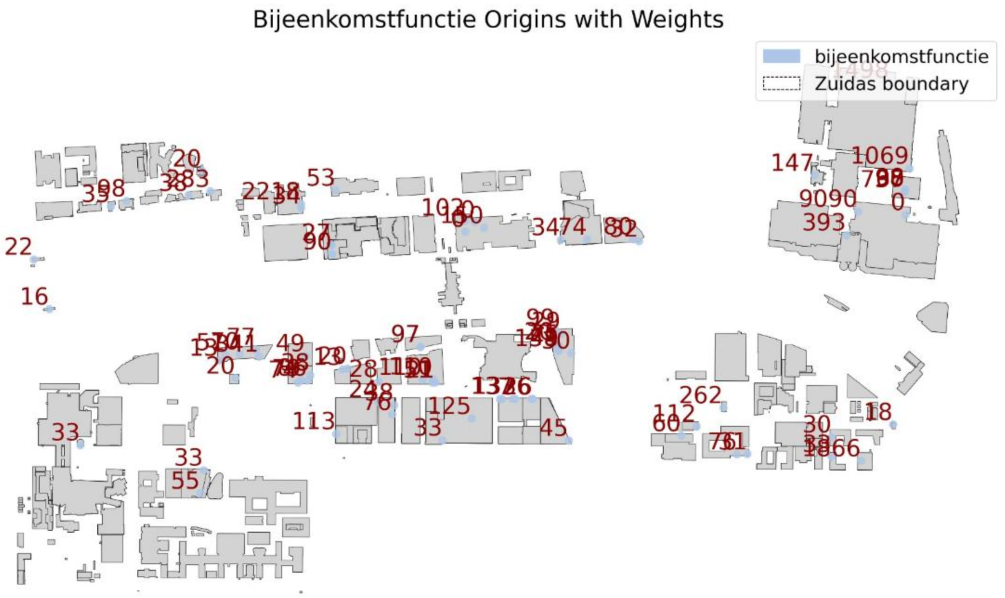


Figure 20: Assembly function building weights

5.3.3 Proxy-based weights

Catering functions were weighted using proxy assumptions because no activity data were available. Based on the comparison with manual counts, see section 5.3.5.2, the labour-based weighting (Alternative 2) was selected as the preferred approach for catering functions, as it produces trip estimates of the correct order of magnitude, unlike the regional-statistics method, which significantly underestimates activity. For results see figure 21. However, this choice remains a coarse proxy: all catering establishments, ranging from cafés and take-away outlets to full-service restaurants, still receive identical weights because no establishment-specific activity data (e.g., visitor numbers, seating capacity, opening hours) is available. As a result, the selected weighting improves overall magnitude but does not capture differences in intensity between individual establishments, leaving substantial uncertainty in spatial detail.

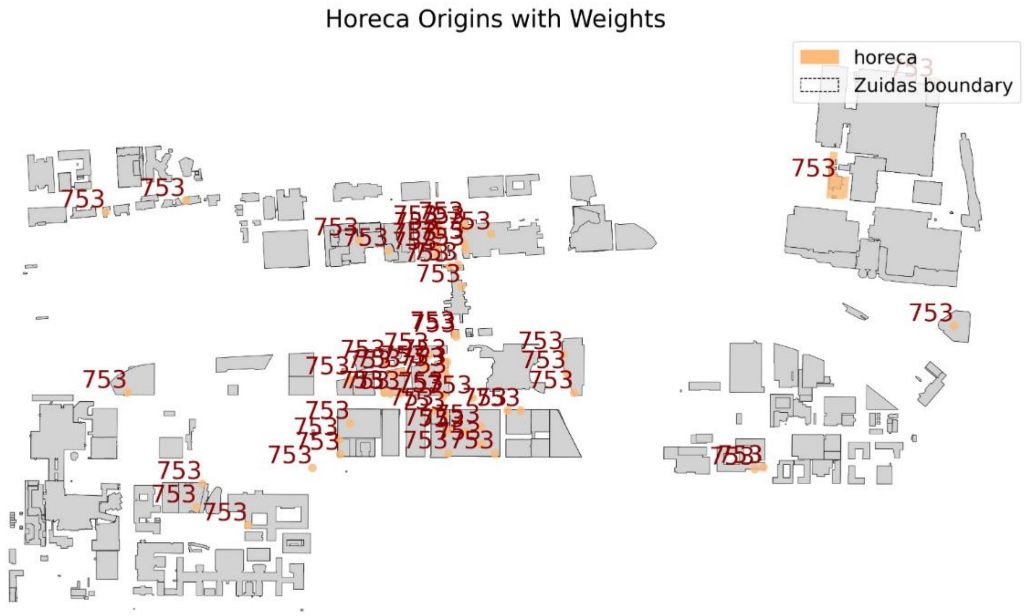


Figure 21: Catering building weights

5.3.4 Public transport (excluding the train station)

For other public transport functions, also two methods where testes, see section 5.3.5.3 , the second alternative. the second alternative, incorporating vehicle capacity and average occupancy, was selected as the preferred weighting method for public transport stops. This approach produces predicted trip volumes that are of the correct order of magnitude. While this provides a behavioural approximation of activity intensity, the lack of stop-level boarding and alighting data means these weights are less precise than those derived from NS ridership at the train station.

5.3.5 Model Sensitivity Results: Building Weights

Even though weights vary across building types, reflecting both functional diversity and differences in size, the method relies on multiple proxies and different weighting strategies depending on building type. For example, some categories are weighted by redistributing totals over GFA, others by the number of individual buildings, and still others, such as public transport, by service frequency and number of lines. These methodological choices introduce potential biases and uncertainty. For three building types, two alternative weighting methods were tested to assess how the assumptions underlying the weight calculations affect the resulting trip generation.

5.3.5.1 Alternative Office Weight Results

The first building type for which alternative weighting methods were tested is offices. In the first approach, weights were based on average occupancy across the week, see figure 22, while the second used maximum daily occupancy. The second method nearly doubles the weights of the first, reflecting the shift from mean to peak activity levels, see figure 18.

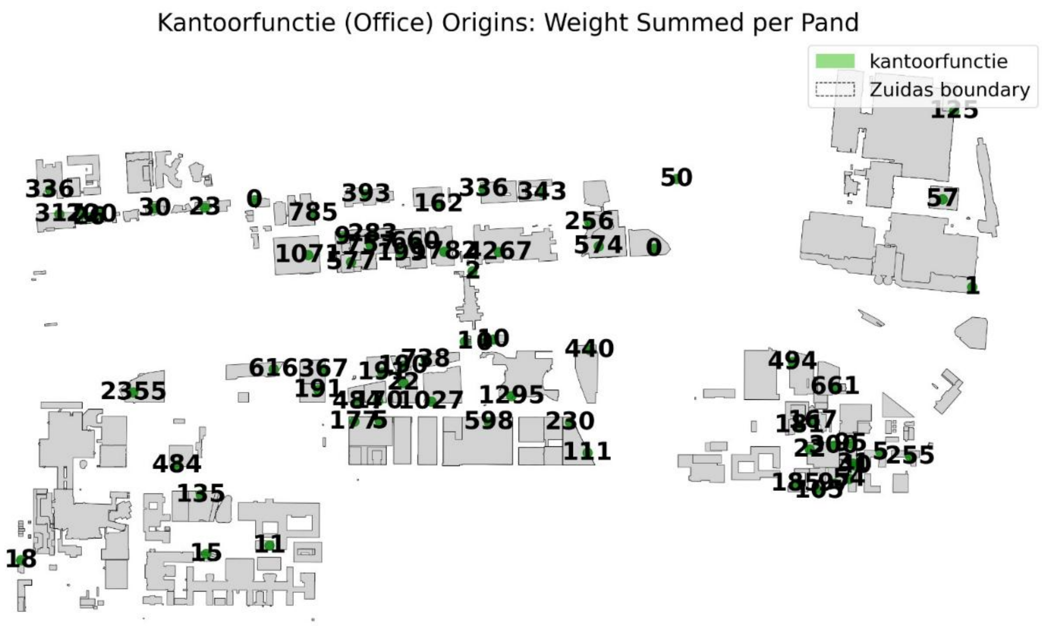


Figure 22: Office weights with average daily occupancy

Using peak occupancy for office buildings produces predicted pedestrian volumes that are closer to the manual counts, though still not within the correct order of magnitude (see table 5). Furthermore, substantial day-to-day variation in office attendance was observed. Since the model applies a single occupancy value for all days, it is unable to represent this variability in actual office use. However, the reliability of the office counts is limited (section 5.7).

Table 5: Predicted vs. Observed Pedestrian Trips for Office Buildings Using Alternative Occupancy Weights

Departure hour	Counted	Predicted Alternative 1	Predicted Alternative 2	Factor Difference A1	Factor Difference A2
11	122	3	5	40.67	24.4
12	197	7	12	28.14	16.42
16	239	16	28	14.93	8.54
17	512	21	36	15.93	9.29

5.3.5.2 Alternative Catering Weights Result

Catering functions remain difficult to weight accurately because large differences in customer activity between establishments, and their dependence on opening hours, are not captured by the current data. Although later accessibility adjustments reduce some of this variation, meaningful calibration will still be required to align these weights with actual usage patterns. Two alternative weighting approaches were tested. The first, based on regional catering statistics, produced relatively low daily visitor estimates for each establishment, each being assigned the weight of 72 people. The second, a labour-based proxy, resulted in higher weights. The weight of 753 people, but still assigned identical values across all establishments and therefore remained a coarse approximation.

The labour-based catering weights are on average 10.46 times higher than the initial weights, resulting in predicted trip totals that are approximately nine times greater. Although relative differences between establishments remain unchanged, the absolute magnitudes increase substantially.

When comparing model outputs with manual counts, the baseline model, alternative 1, significantly underestimated catering-related activity. Using the higher labour-based weights, alternative 2, brought the predicted trips much closer to observed levels, reducing the discrepancy from factors from 33 to 2.75 and from 1.5 to 0.16, depending on the hour (table 6). This indicates that the labour based weights produce results of the correct order of magnitude. However, because the increase is applied uniformly across establishments, local variability is still not fully captured, and fine-grained spatial differences remain subdued.

Table 6: Predicted vs. Observed Pedestrian Trips for Catering Buildings Using Alternative Methods

Departure Hour	Counted	Predicted regional Alternative 1	Predicted labour Alternative 2	Factor Differences A1	Factor Differences A2
8	8	0	3		2.67
9	28	0	5		5.6
11	16	1	9	16	1.77
12	33	1	12	33	2.75
11	0	1	9		
12	17	1	11	17	1.55
16	3	2	19	1.5	0.16
17	9	2	18	4.5	0.5

5.3.5.3 Alternative Public Transport Weight Result

For public transport facilities, two weighting methods were tested. The first alternative accounted only for the number of lines and their service frequency, while the second incorporated vehicle capacity and occupancy to better reflect actual passenger volumes. The resulting differences in assigned weights for bus stops and tram stops are shown in figure 23.

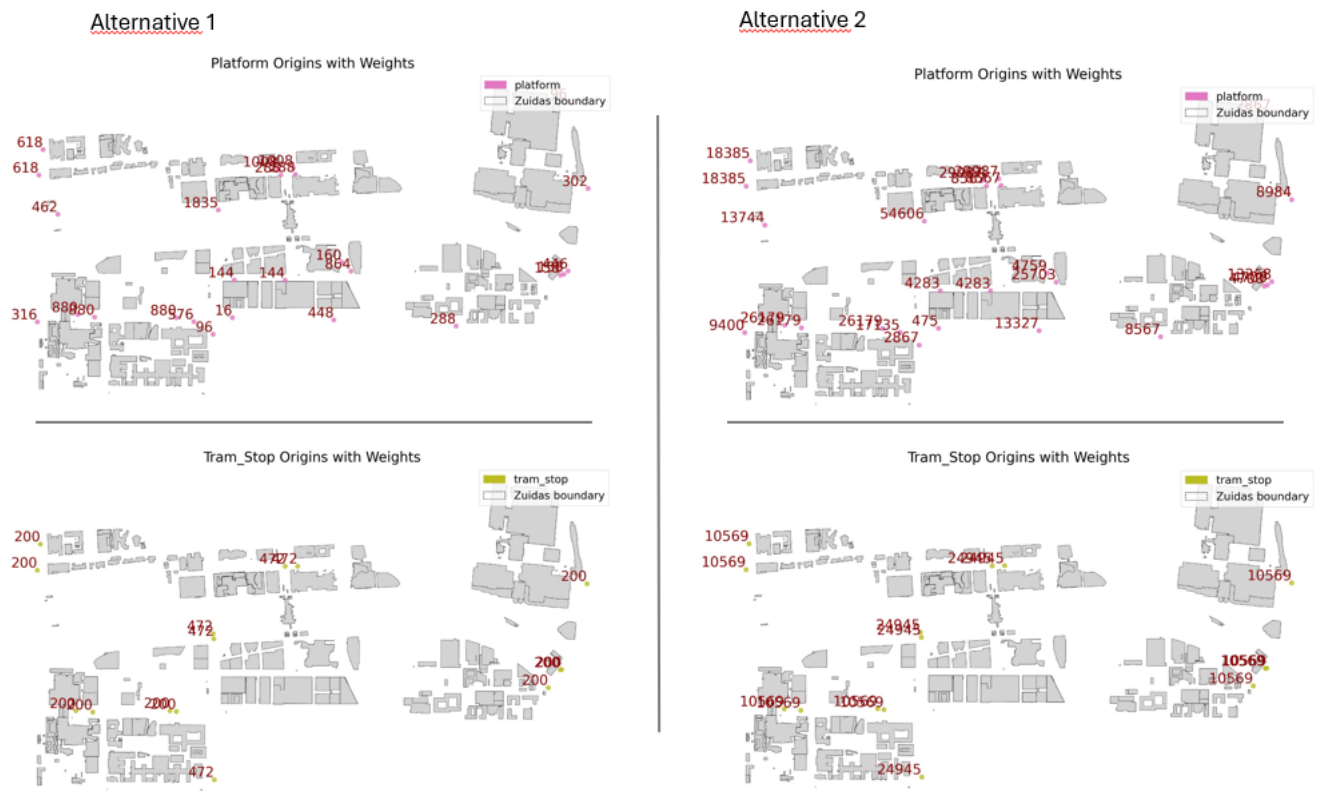


Figure 23: Comparison alternatives public transport weights

The second alternative, results in higher weights with about a factor of 30. This results in higher predicted generated trips at these public transport stops of about a factor of 33. The relative distribution of trips across locations remains unchanged, meaning the hotspots identified in the baseline model continue to dominate pedestrian traffic. The increase in predicted trips is closer to the counted trip going from a factor difference of for example 104 to 2.7, see table 7. Meaning the second alternative is the correct order of magnitude.

Table 7: Predicted vs. Observed Pedestrian Trips for Public Transport Stops Using Alternative Methods

Departure hour	Counted	Predicted Alternative 1	Predicted Alternative 2	Factor Difference A1	Factor Difference A2
8	167	8	261	20.88	0.64
9	178	9	294	19.77	0.61
4	109	4	136	27.25	0.80
3	154	3	77	51.33	2
2	208	2	75	104	2.77
5	235	5	156	47	1.51

5.4 Elastic Building Weights

This section explains how elastic building weights were generated by integrating buildings into the pedestrian network and adjusting their weights based on accessibility. It begins by outlining the procedure used to insert buildings as active nodes in the network and discusses key assumptions and limitations of this integration. The section then describes how reach values were calculated and used to adjust building weights. Finally, it examines two major sources of model sensitivity: alternative normalization methods for reach and the influence of different walking radii, illustrating how these choices shape the resulting spatial distribution of adjusted weights.

5.4.1 Inserting Buildings into the Pedestrian Network

The integration procedure successfully embedded buildings into the pedestrian network, allowing them to function as active nodes for accessibility calculations. Figure 24 shows a section of the original graph alongside the updated version after building nodes were added. In the original OSM-derived network, nodes represent locations where pedestrians can change direction, such as intersections, junctions, or endpoints of paths. After integration, additional nodes were introduced to represent buildings, and these nodes now carry attributes including building type and assigned weight, enabling them to act directly as origins and destinations for trip generation.

This demonstrates that the graph update procedure generally worked as intended, with buildings now represented directly as origins and destinations in the network. One deviation from the procedure outlined by Sevtsuk et al. (2025) was observed: for edges containing multiple neighbouring buildings, the original edge was not removed, which may result in redundant links. At a qualitative level, embedding buildings as nodes improves the realism of the model by allowing shortest-path calculations to start and end at actual building locations rather than arbitrary street intersections. However, this approach rests on several assumptions. First, centroids are treated as proxies for entrances. For simple buildings with a single VBO, comparison of the nearest snapped point with actual entrances suggests this assumption is sufficiently accurate. For large or irregular buildings with many VBOs and multiple exits, however, this representation is less precise; the model therefore assumes that centroids still provide an adequate approximation. Second, the snapping procedure assumes that the nearest network edge reflects the true pedestrian access route, which may not always hold in practice. Finally, the procedure assumes that OpenStreetMap provides sufficiently accurate pedestrian geometries. While checks indicate good agreement between OSM and field conditions, no alternative pedestrian network is available for full verification, and OSM accuracy is therefore assumed to be adequate.

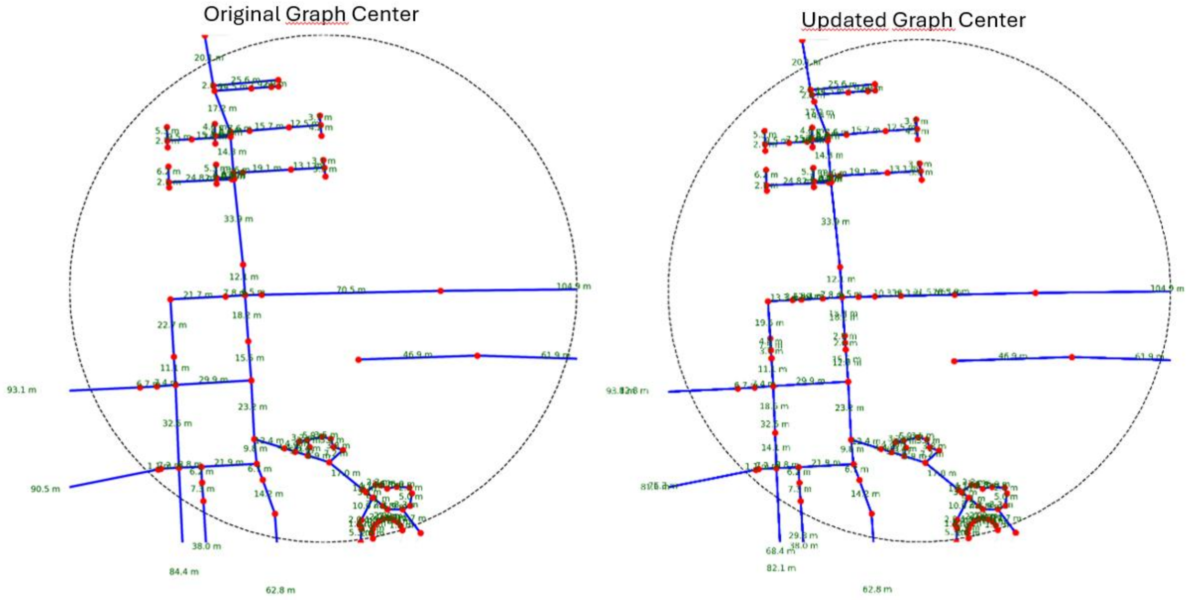


Figure 24: Original and updated graph center node distances within 100m buffer

Overall, the addition of nodes strengthens the model by directly linking land-use points to the pedestrian network, consistent with recent advances in urban network analysis (Sevtsuk & Kalvo, 2024). At the same time, robustness could be improved by addressing limitations such as centroid-based entrances, the handling of edges with multiple buildings, and the reliance on OSM data that may not be fully complete or error-free.

5.4.2 Reach Calculation

The mean-based normalization method and the 1100 m radius (mean walking distance from ODiN) were selected for the final analysis because together they provide a balanced and interpretable adjustment of building weights, preserving meaningful spatial variation without overemphasizing extreme differences. The resulting patterns confirm that accessibility-based weight adjustment can substantially reshape trip potentials. For office buildings, for example, weights can increase by an order of magnitude or be reduced to zero relative to the original values (figure 16).

In the case of office-to-train-station trips (figure 25), most buildings located within the walking radius show weight increases of roughly tenfold, reflecting their strong accessibility to the station. Conversely, all office buildings beyond this radius receive an adjusted weight of zero, indicating that, under the current threshold-based model, no workers are expected to walk to the station from those locations. This reflects a clear spatial logic: proximity to the station strongly amplifies pedestrian trip potential, while distance effectively eliminates it.

A similar but more nuanced pattern appears for office-to-bus-stop trips (figure 26). Here, some buildings experience a doubling of their weights because they lie within reach of several bus stops, while others see their weights halved, perhaps having access to only one stop. Buildings with no bus stop within the walking radius again receive an adjusted weight of zero. These outcomes intuitively reflect that an office with multiple nearby bus stops is likely to generate more pedestrian trips toward bus services, while an office with no nearby stops will generate none.

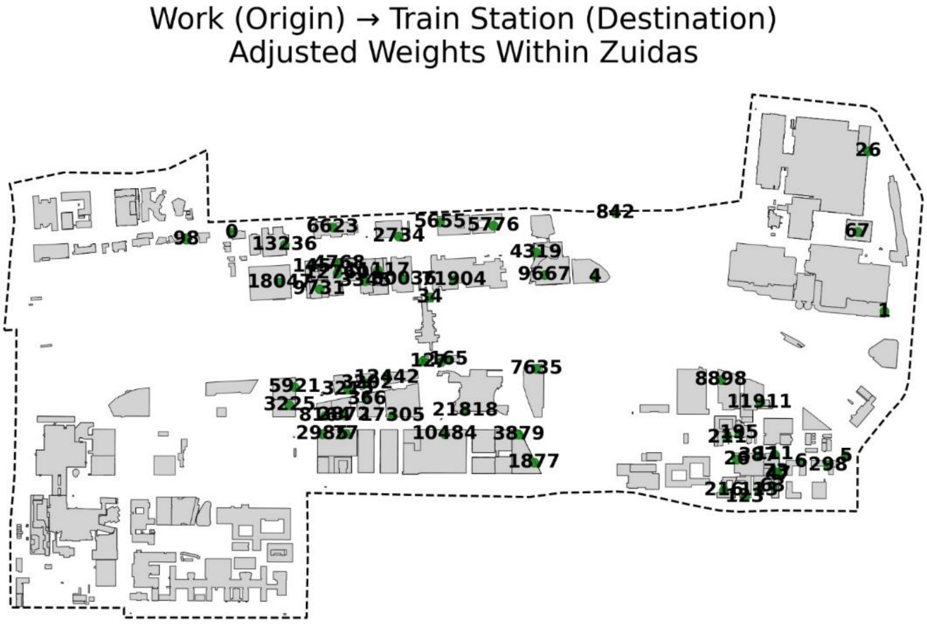


Figure 25: Adjusted weights for offices to destination type train station

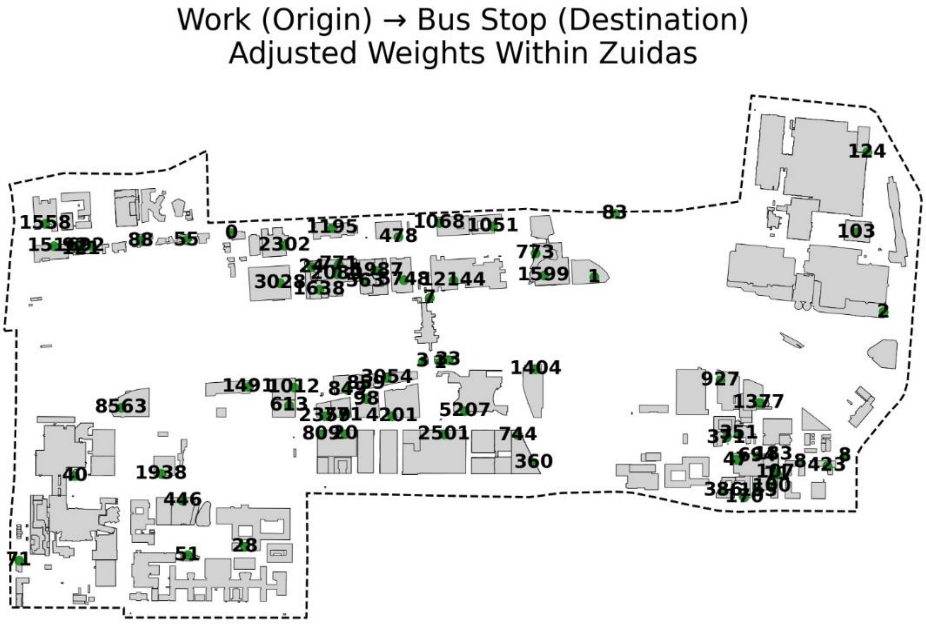


Figure 26: Adjusted weights for offices to destination type bus stops

Overall, the results appear consistent with expected pedestrian behaviour: buildings with better access to relevant destinations generate more walking trips, and those without access generate fewer or none. However, further research is required to validate the behavioural realism of these effects. The current approach applies a strict distance threshold, buildings are either within reach or not, without accounting for distance decay (i.e., gravity effects), which would soften these boundaries in real-world conditions. Additionally, more validation data are needed to assess whether the magnitude of the adjustments, especially cases where weights increase tenfold or decrease to zero, faithfully represents actual pedestrian patterns.

5.4.3 Model Sensitivity: Reach Normalization and Radius Parameter

This subsection examines how sensitive the adjusted building weights are to two key modelling choices: the method used to normalize reach values and the walking radius that defines the scale of accessible destinations. By comparing alternative normalization formulas and varying the walking radius, the analysis shows how these parameters influence the magnitude, spatial distribution, and interpretability of the resulting accessibility-based weights.

5.4.3.1 Alternatives Reach Normalization

The effect of the mean-based adjustment (formula 17) is shown in figure 27. As the walking radius increases, the central building's adjusted weight rises initially, reflecting the fact that its reach expands more rapidly than the study-area mean. This occurs because additional nearby buildings fall within the radius, enhancing its relative accessibility. After a certain threshold, the growth stabilizes and eventually declines slightly, as nearly all buildings become mutually reachable and differences in reach diminish. This produces a realistic nonlinear pattern: accessibility increases with radius, but the marginal benefit declines once the accessible set saturates—consistent with expected pedestrian behaviour.

The Z-score–based adjustment (formula 18) produces the same general shape as the mean-based approach but exaggerates differences. Small deviations from the mean are amplified by division through the standard deviation, leading to inflated contrasts between buildings with moderately different reach values. While this can be analytically useful for highlighting extremes, it also generates weight variations that are behaviourally unrealistic. In particular, buildings only slightly above the mean receive disproportionately high adjusted weights, while those slightly below the mean are heavily down-weighted. Compared to Formula 16, the Z-score method therefore overemphasizes spatial contrasts and is less suitable for representing perceived pedestrian accessibility.

In the synthetic scenario, the ratio-based and mean-based formulas produce identical patterns because all buildings start with equal weights. The primary difference lies in behavioural interpretation: whereas formula 19 applies strict proportional scaling, Formula 16 moderates this scaling by centring values around the mean and accentuating smaller deviations. The ratio-based method assumes a perfectly proportional relationship between reach and perceived attractiveness, which may not reflect actual pedestrian decision-making. For this reason, and because Z-score normalization (Formula 18) produced exaggeration, mean-based normalization (formula 16) was selected for the final analysis. It offers a balanced adjustment that preserves meaningful spatial variation without amplifying differences beyond what accessibility theory would support.

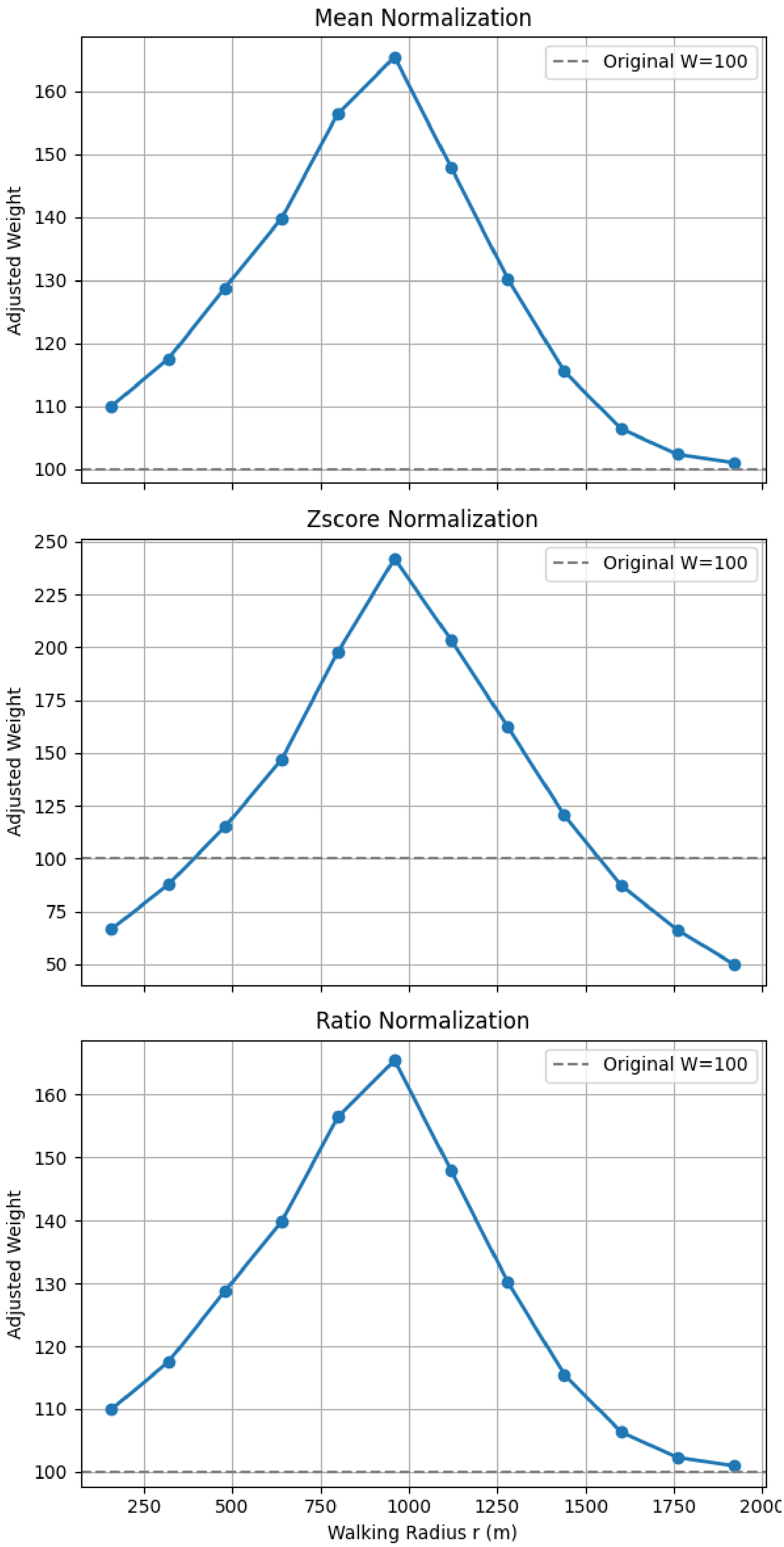


Figure 27: Impact normalization method reach on homogenous synthetic dataset

5.4.3.2 Alternatives Radius in Reach

The influence of the walking radius *r* on the mean-based Reach normalization (Formula 16) is illustrated in figure 28. As expected, increasing the radius expands the set of destinations included in each building's reach, progressively smoothing the spatial pattern of adjusted weights. At the smallest radius (900 m), accessibility is highly localized: only buildings located close to dense clusters of destinations achieve high reach values, producing a pronounced central peak, while peripheral buildings receive much lower adjusted weights due to their limited connectivity.

When the radius increases to 1100 m, the baseline based on the mean walking distance from ODiN, more buildings fall within reach for most origins. This reduces spatial contrast and broadens the pattern of elevated weights. By 1500 m, nearly all buildings in the study area become mutually reachable, resulting in an almost uniform distribution of adjusted weights. At this point, the mean-based normalization differentiates very little between buildings because their reach values converge around the study-area average.

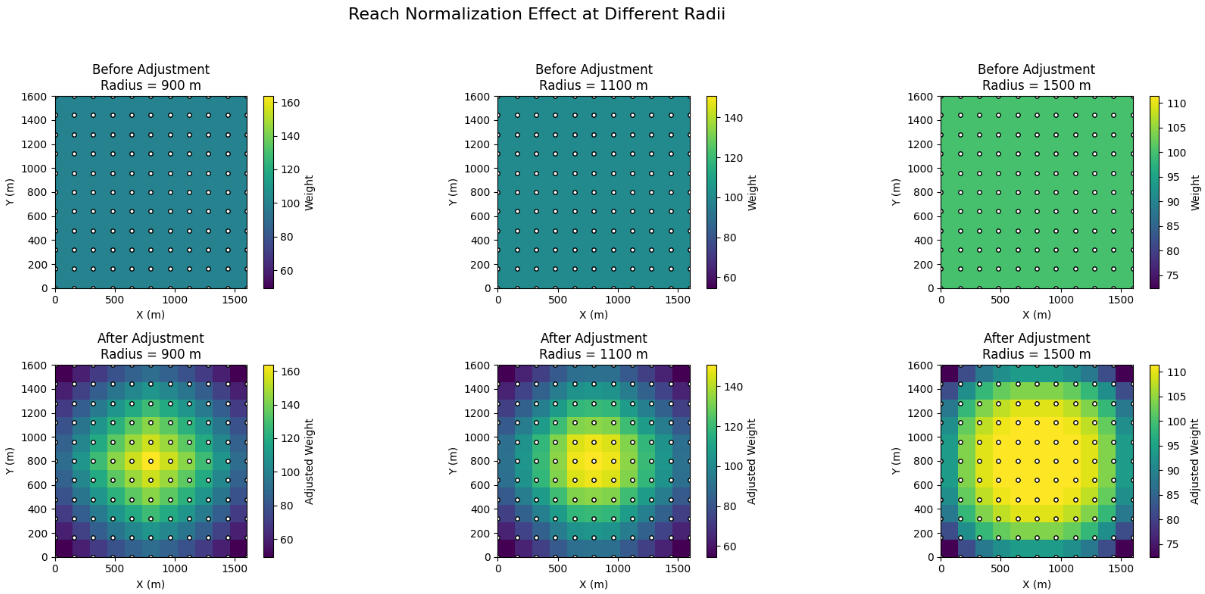


Figure 28: Effect of different radius in reach calculation

These differences are clearly visible when examining the residential-to–train station scenario (figure 29). At 900 m, only buildings near Amsterdam Zuid station lie within reach, resulting in strong amplification: a central building's weight increases from 478 to 5577, while a peripheral building only rises from 275 to 222. At 1100 m, the central weight decreases to 4699 and the peripheral to 245, indicating a more balanced pattern. At 1500 m, the same buildings show adjusted weights of 3305 and 2034 respectively, reflecting near-saturation of the reachable area. Across all radii, several office buildings that originally had small non-zero weights (up to 336) were reduced to zero because no train station fell within the corresponding threshold.

Overall, smaller radii emphasise highly localized pedestrian access, producing steep gradients and strongly favouring buildings close to major destinations. Larger radii diffuse these gradients and flatten meaningful differences. The baseline radius of 1100 m provides the most realistic balance: it captures local accessibility variations without overstating a small number of central nodes or excessively homogenising the pattern. This highlights the importance of radius selection, as it directly shapes the magnitude and spatial distribution of the adjusted building weights.

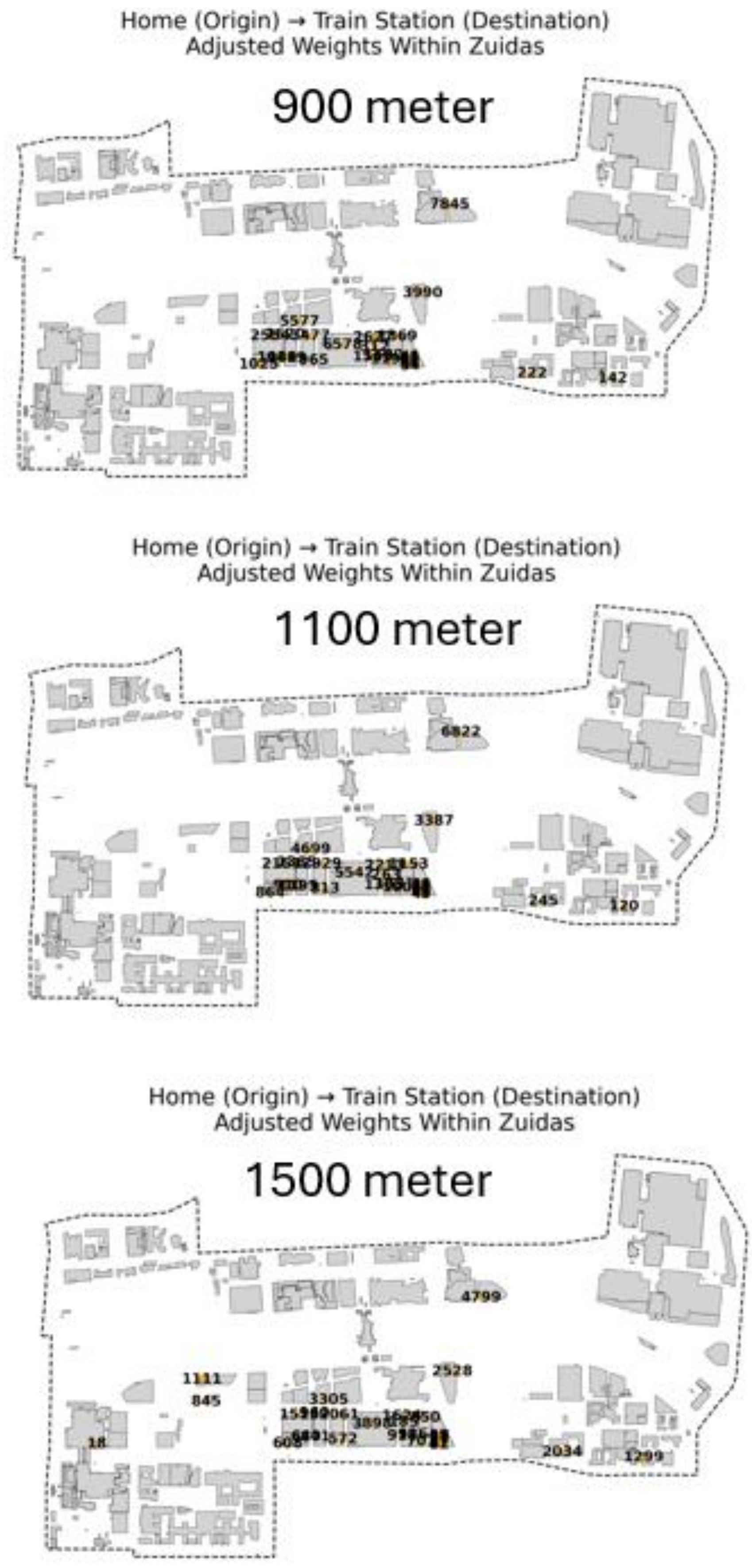


Figure 29: Differences in adjusted weights for home to train station for different radii

5.5 Trip Rates Computations

This section presents the derivation and evaluation of hourly pedestrian trip rates used in the model. The aim is to assess how well the ODiN travel survey represents walking behaviour in urban, multimodal environments and to determine which filtering strategy yields the most reliable inputs for demand estimation.

Section 5.5.1 describes the computation of trip rates using the broader urban-municipality dataset and summarises the resulting temporal patterns. Section 5.5.2 examines model sensitivity by comparing these results with those obtained from a narrowly filtered set of Zuidas-like postcode areas, highlighting the effects of sample size, representativeness, and underreporting. Together, these analyses inform the selection of trip rates used in the final model.

5.5.1 Calculating Trip Rates

Using the broader selection of large urban municipalities to filter the ODiN dataset produces a more stable set of hourly trip rates. When simplified to a heatmap of origin types (figure 30), the resulting patterns show clear peaks at expected times, for example, morning demand from home or train stations, and afternoon returns home. However, for some origins, particularly public transport stops, trip rates remain lower than anticipated, indicating that certain walking legs are still underreported.

Despite these limitations, the broader dataset is assumed to capture daily fluctuations in pedestrian activity sufficiently for estimating baseline trip rates. Underreporting, will be examined further during face validation, and additional calibration data may be required to refine the estimates.

A key assumption underlying this process is the mapping of ODiN trip purposes to building types. For instance, “business visit” is assigned to workplaces, while several detailed purposes are merged into broader categories (e.g. “touring/strolling” as leisure, “sport/hobby” as sport) to reduce noise. These aggregations simplify the analysis but inevitably obscure more nuanced destination patterns, and further data would be needed to confirm the accuracy of these assumptions in real-world contexts.

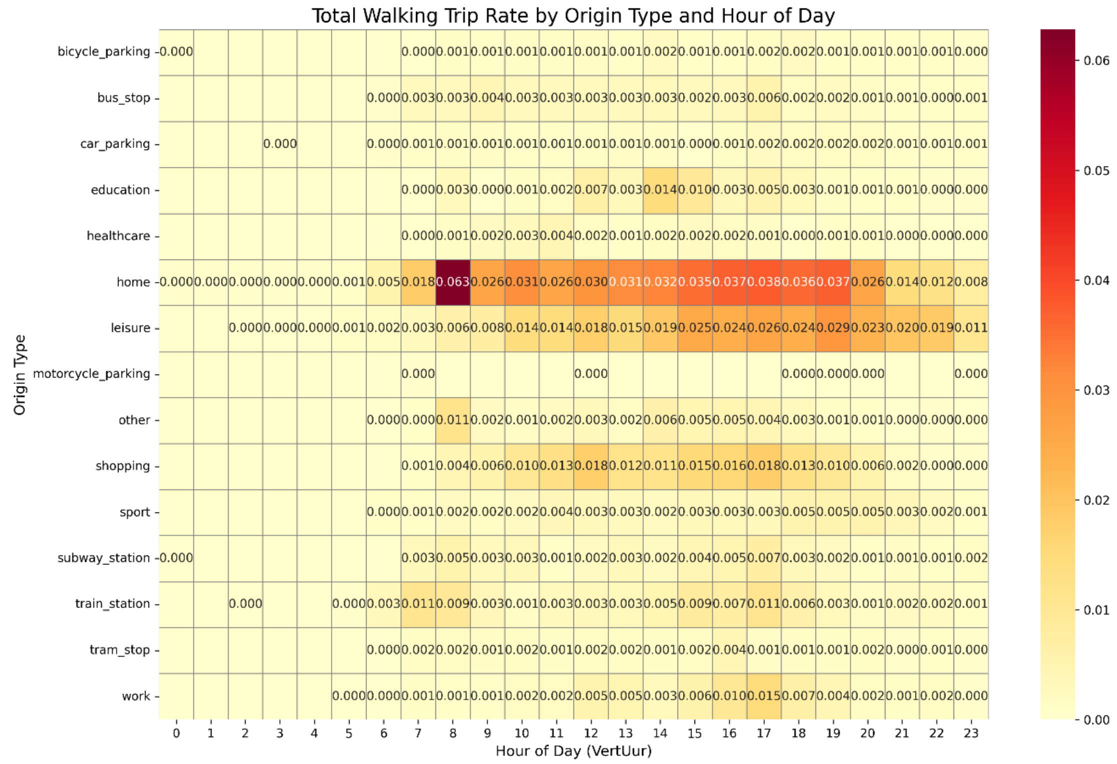


Figure 30: Trip rate heatmap per origin type

5.5.2 Model Sensitivity: Effect of Regional Filtering

The resulting hourly OD trip rates for the narrow selection of postcode areas highly similar to the case study context, are visualized as a heatmap in figure 32. The heatmap reveals that for many OD–hour combinations, values are zero, reflecting the limited occurrence of certain walking flows in the ODiN sample. Nevertheless, several clear temporal and directional patterns emerge.

A pronounced morning peak occurs for walking trips from train stations to work, reaching 0.019 at 8:00. Return flows from work to train stations are weaker and more dispersed, with peaks of 0.006 at 15:00 and 0.002 at 18:00, while values are unexpectedly zero at 16:00 and 17:00. For home–work commuting trips, morning peaks appear at 6:00 (0.006), 8:00 (0.012), 9:00 (0.008), and 10:00 (0.005), while return flows from work to home are concentrated later in the day, with a peak of 0.009 at 17:00, followed by smaller peaks of 0.005 at 18:00 and 19:00, and 0.004 at 20:00. These patterns broadly reflect expected commuting rhythms, although the asymmetry in station-related flows suggests potential under-sampling or misrepresentation of certain walking legs in ODiN. Small peaks from work to leisure at 11:00 and 12:00 (0.002) indicate that some lunch-break strolls are captured, even if their frequency is low.

Among all OD pairs, shopping-to-home trips show the highest trip rates, with a temporal distribution consistent with expected shopping activity throughout the day. Similarly, leisure-to-home trips exhibit relatively high trip rates concentrated in the afternoon and evening hours, aligning with typical leisure patterns.

A notable limitation is that trip rates involving bus and tram stops are consistently zero. Given the mixed-use character of the study areas, with high concentrations of office workers and strong multimodal connectivity, higher values would be expected. This discrepancy highlights a limitation in ODiN for capturing short walking access and egress legs to public transport, and it suggests that hourly trip rates derived from the survey may not fully capture the daily fluctuations in pedestrian activity.

Trips originating from work, education, and transit stations are unexpectedly low. This suggests that many walking trips, particularly walking legs that are part of multimodal journeys as well as return trips, are underreported. Such underreporting has also been documented in previous research; for example, Lugtig et al. (2022) note that respondents often forget to include details such as the mode of transport.

To mitigate the limitations observed in the smaller Zuidas-focused dataset, the filtering approach was adjusted. Rather than restricting the sample to areas comparable to Zuidas, the dataset was filtered at the municipality level. The selected municipalities include Groningen, Almere, Deventer, Enschede, Zwolle, Nijmegen, Wageningen, Amersfoort, Utrecht, Amstelveen, Amsterdam, Haarlem, Alphen aan den Rijn, Capelle aan den IJssel, Delft, Dordrecht, Gouda, Leiden, Rotterdam, Schiedam, Breda, Eindhoven, Tilburg, and Maastricht. This broader filtering results in 4,964 unique individuals and 10,719 walking trips, substantially increasing the sample size. The mean walking distance across all trips is 1.75 km; when capped at 3 km, the mean drops to 1.0 km (figure 31), which is slightly higher than in the smaller-area dataset.
The distribution of trip purposes in the larger dataset is broadly consistent with the smaller dataset, with home-based walking trips dominating overall behaviour.

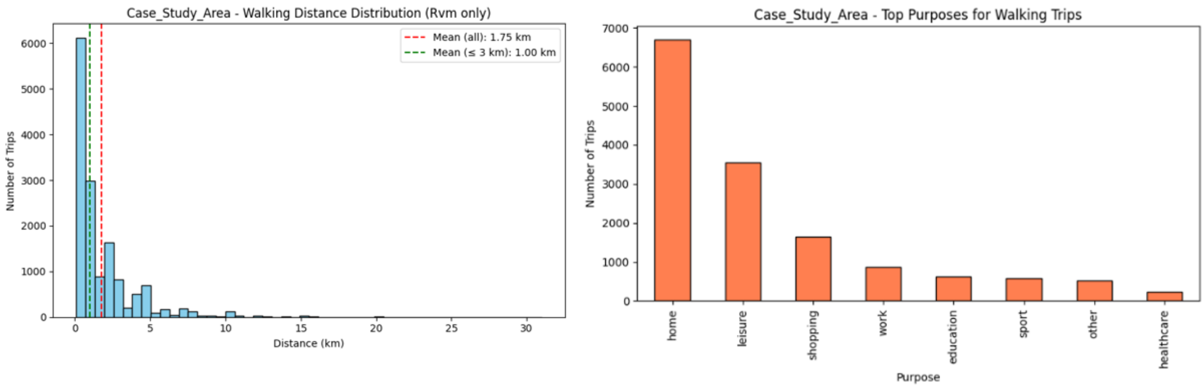


Figure 31: Distributions of walking distance and trip purpose for urban municipalities

Temporal patterns (figure 33) show clear peaks at 08:00, 12:00, and 17:00, corresponding to morning commuting, midday activities such as shopping, leisure, or lunch trips, and evening returns home. Compared to the smaller dataset, the evening peak at 20:00 is less pronounced, suggesting that late-day activities are relatively less concentrated across the broader set of municipalities.

A notable difference is that the midday peak (12:00) is more pronounced and less fragmented, with fewer abrupt drops (e.g., the dip at 13:00 is shallower). This indicates a smoother distribution of lunch and midday errands across municipalities, rather than being concentrated in high-density areas alone. Overall, the larger dataset emphasizes that walking demand is influenced not only by commuting demand but also by shopping and leisure activities, producing smoother temporal patterns compared with the smaller-area sample.

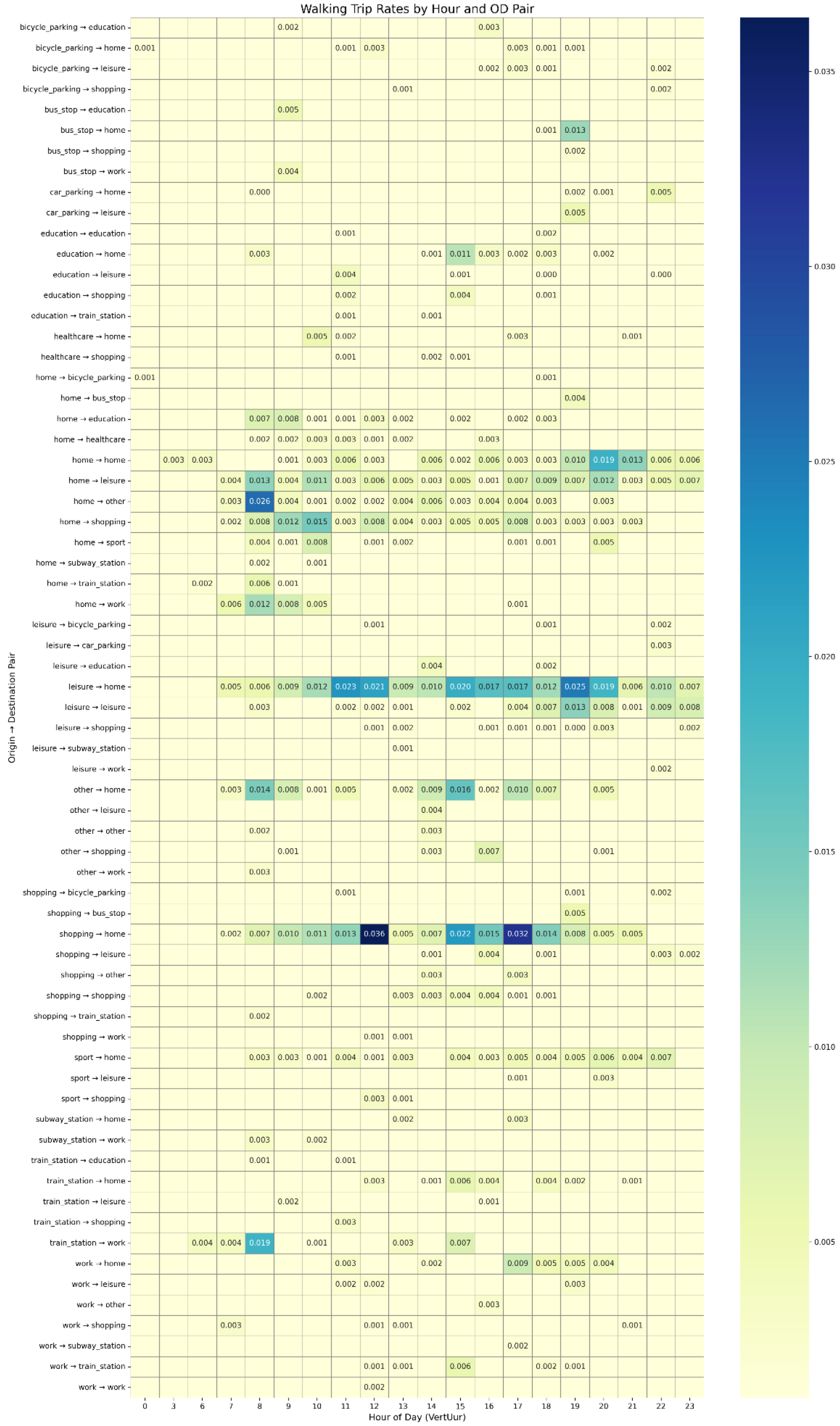


Figure 32: Heat map trip rates narrow filtered ODiN set of zuidas like postcode areas

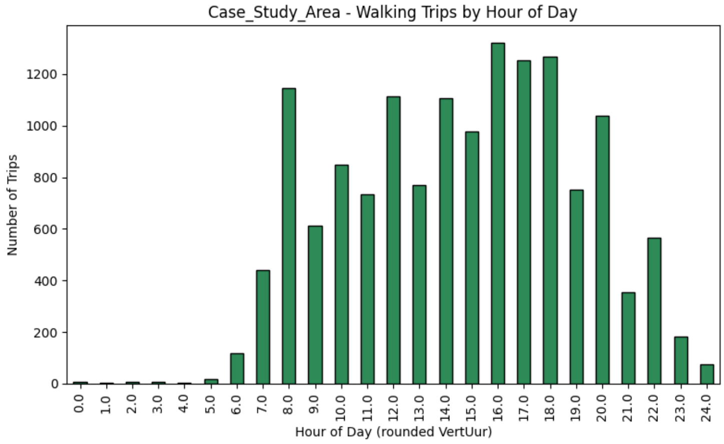


Figure 33: Temporal pattern for urban municipalities

The patterns observed in this broader dataset remain consistent with expected pedestrian behaviour reported in the literature. Despite the wider geographic coverage, the study maintains the assumption that, once weighted, the sample provides a reasonable representation of real-world travel behaviour in the study area. Consequently, this dataset is used to compute the trip rates. The resulting origin–destination trip rates are presented in figure 34. This larger dataset offers a more robust foundation for estimating pedestrian demand, while still acknowledging the inherent limitations of survey-based data.

In the updated dataset, zero-value cells are considerably reduced, though the maximum trip rates are lower overall. The highest observed rate is now for home → education trips at 0.021, while leisure → home and shopping → home remain the OD pairs with the largest overall trip rates. For train station → work trips, the pronounced 8:00 peak decreases from 0.019 to 0.006, but demand become more evenly distributed across the morning: at 07:00 the rate increases from 0.004 to 0.009, and at 09:00 it rises from 0.000 to 0.001. Trips involving bus and tram stops remain low, though fewer zero entries are observed; for instance, bus stop → work trips now show non-zero rates at 07:00 (0.001), 08:00 (0.002), and 09:00 (0.002), compared to the original dataset where values were zero at 07:00 and 08:00 and 0.004 at 09:00.

To better understand the differences between the Zuidas-based and broader municipalities datasets, trip rates per origin type were plotted with separate lines for each destination type.. The broader municipalities dataset produces smoother and more consistent temporal patterns, whereas the Zuidas dataset, despite showing higher peak rates, exhibits more irregularities due to the smaller sample size. This is particularly evident for origins such as bus stops, car parking, healthcare, and tram stations. For example, in the Zuidas dataset, trips from bus stops, see figure 35, occur only at 09:00, 18:00, and 19:00, with the highest peak being bus stop → home at 19:00 (0.013). In the broader dataset, trips from bus stops are observed between 06:00 and 23:00, with the highest peak at 17:00 for bus stop → home (0.005). Despite differences in magnitude, both datasets reflect the expected morning commute to work or education and evening returns home.

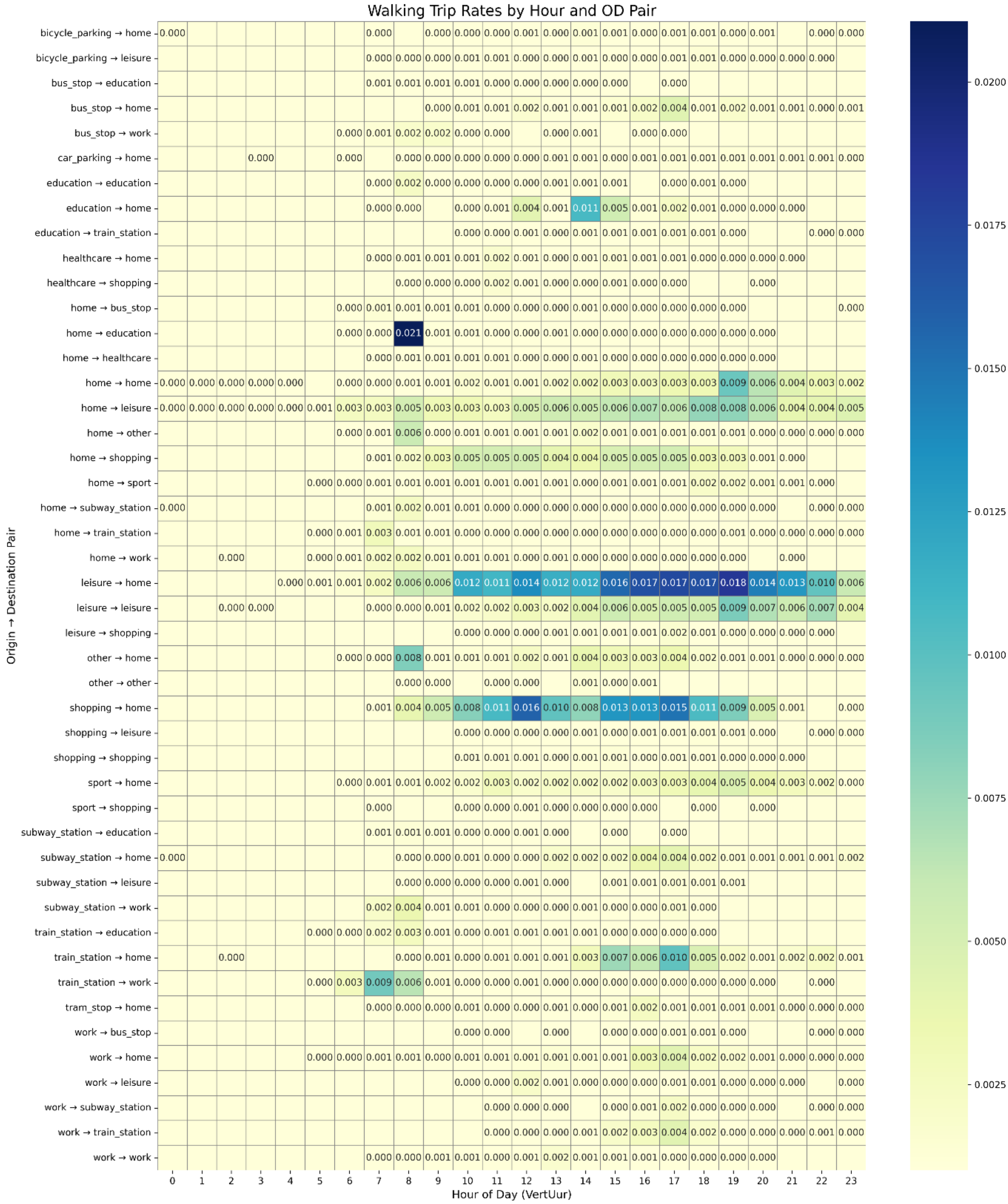


Figure 34: Heat map trip rates broad filtered ODiN set of urban municipalities

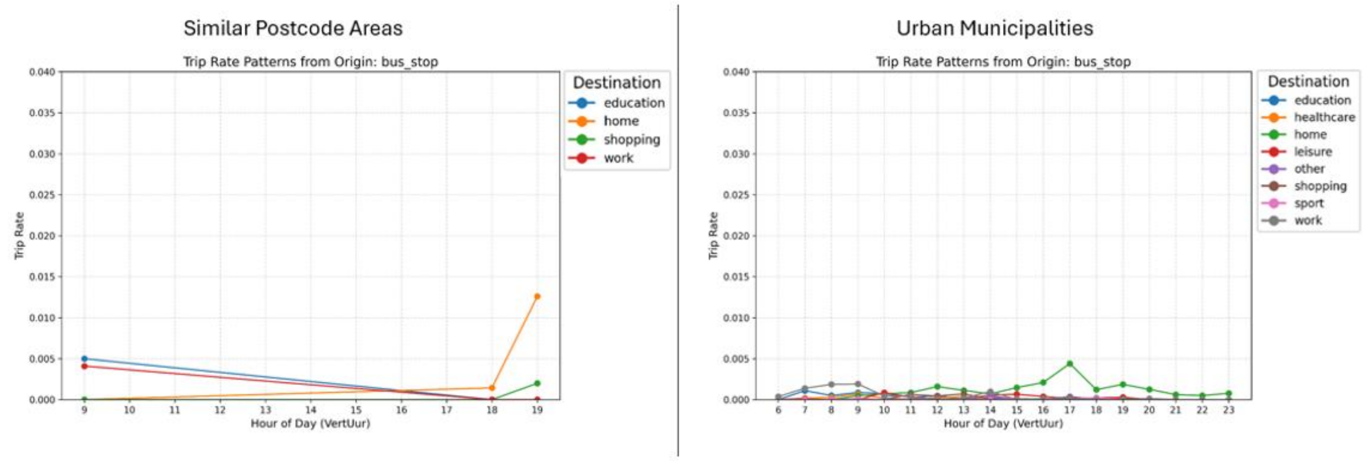


Figure 35: Trip rate patterns of bus stops for narrow and broad dataset

Education-origin trips show broadly similar patterns across datasets, see figure 36, with home-bound demand peaking in the afternoon (15:00 and 17:00). Differences appear in secondary demand: the broader dataset shows a minor peak at 08:00 for education → home and at 17:00 for education → train station, while the Zuidas dataset exhibits a strong 11:00 peak for education → train station and unusually high education → shopping and leisure trips. These deviations may reflect concentrated university activity in Zuidas-like areas or artefacts of the smaller sample size. Overall, the broader dataset produces smoother and more plausible curves for education trips.

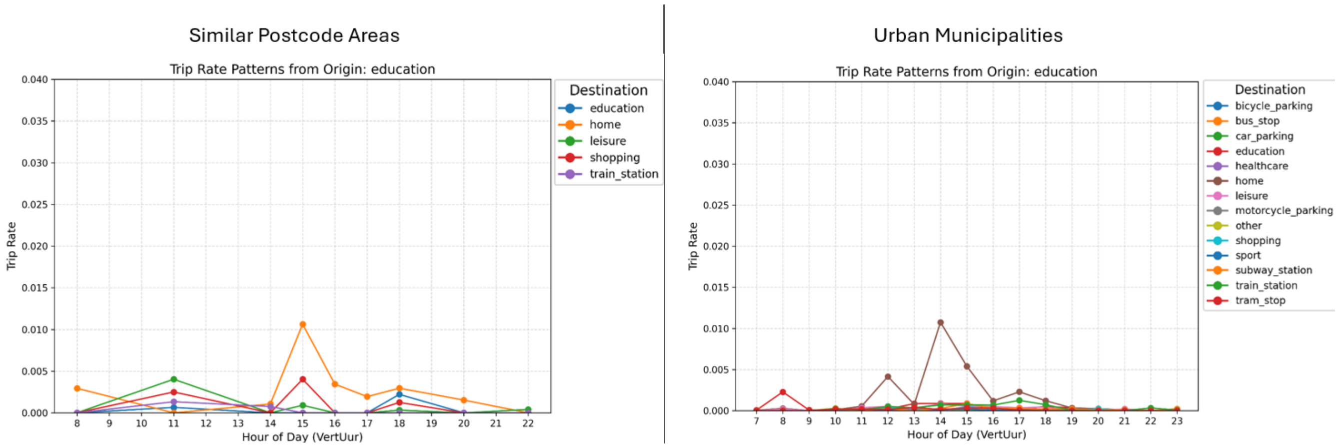


Figure 36 Trip rate patterns of education for narrow and broad dataset

Train station origins further illustrate the contrast, see figure 37. In the Zuidas dataset, trips to work are sharply concentrated at 08:00, reflecting the dense office environment, whereas the broader dataset spreads these trips between 05:00 and 09:00, providing a more realistic temporal distribution but underestimating the intensity of station-related walking in office-dense areas. The “other” category also shows lower rates in the broader dataset, which is beneficial, as it reduces uncertainty by allocating trips to more meaningful categories.

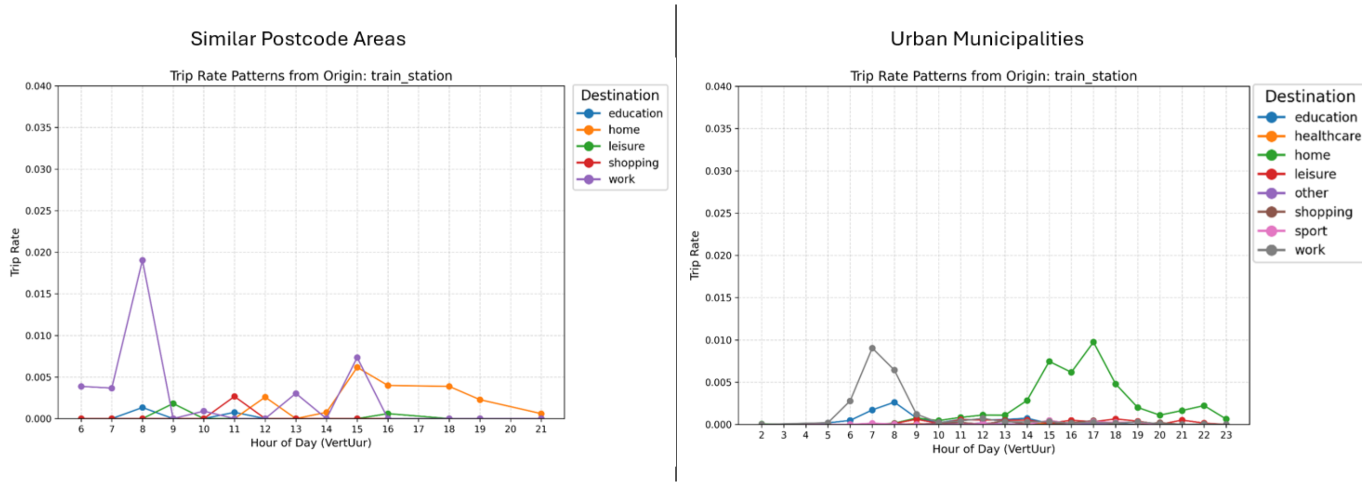


Figure 37: Trip rate patterns of train station for narrow and broad dataset

Overall, for each origin type see appendix 2, the comparison highlights a trade-off between filtering approaches. The Zuidas-based dataset captures higher peaks but suffers from irregularities due to limited sample size. The broader municipalities dataset benefits from a larger number of observations, resulting in smoother and more consistent temporal patterns, though maximum trip rates, particularly for public transport origins, are lower than would be expected in highly urbanized areas. Consequently, the broader dataset is considered more suitable for estimating general urban walking behaviour, while acknowledging that it likely undercounts short access and egress walking legs in multimodal trips, particularly in dense office and transport hubs such as Amsterdam Zuidas.

5.6 Calculating Generated Trips

Applying the trip rates derived in the previous section to the adjusted building weights results in estimates of the number of pedestrian trips generated at each node. These estimates are disaggregated by hour and by origin–destination type, allowing both daily rhythms and functional differences across land uses to be captured. As described in the methodology, trip generation is assumed proportional to building activity (weight) and the likelihood of trips from the ODiN data. Destination-specific rates are used to account for variation in trip frequency, distance, and duration across land uses (e.g., frequent short trips to grocery stores versus less frequent but longer trips to workplaces), providing a basis for more detailed analyses of trip distribution in future research. In addition, node weights are adjusted for accessibility, under the assumption that this sufficiently reflects the spatial realism of trip generation. Finally, hourly trip rates are applied to preserve the temporal distribution of walking activity, even though some short access and egress trips may be underreported. Together, these steps yield a dataset of predicted pedestrian demand per node, per hour, and per destination type, forming the basis for the subsequent distribution and validation analyses.

The generated trips were visualized across the study area by origin type and hour. For example, trips originating from homes clearly show the difference between morning and afternoon peaks, with variations between buildings of different sizes and locations apparent (figure 38 and 39). Visualizations for other building types, for morning and afternoon peak, are provided in Appendix 3. The accuracy of these generated trips is discussed in the validation section.

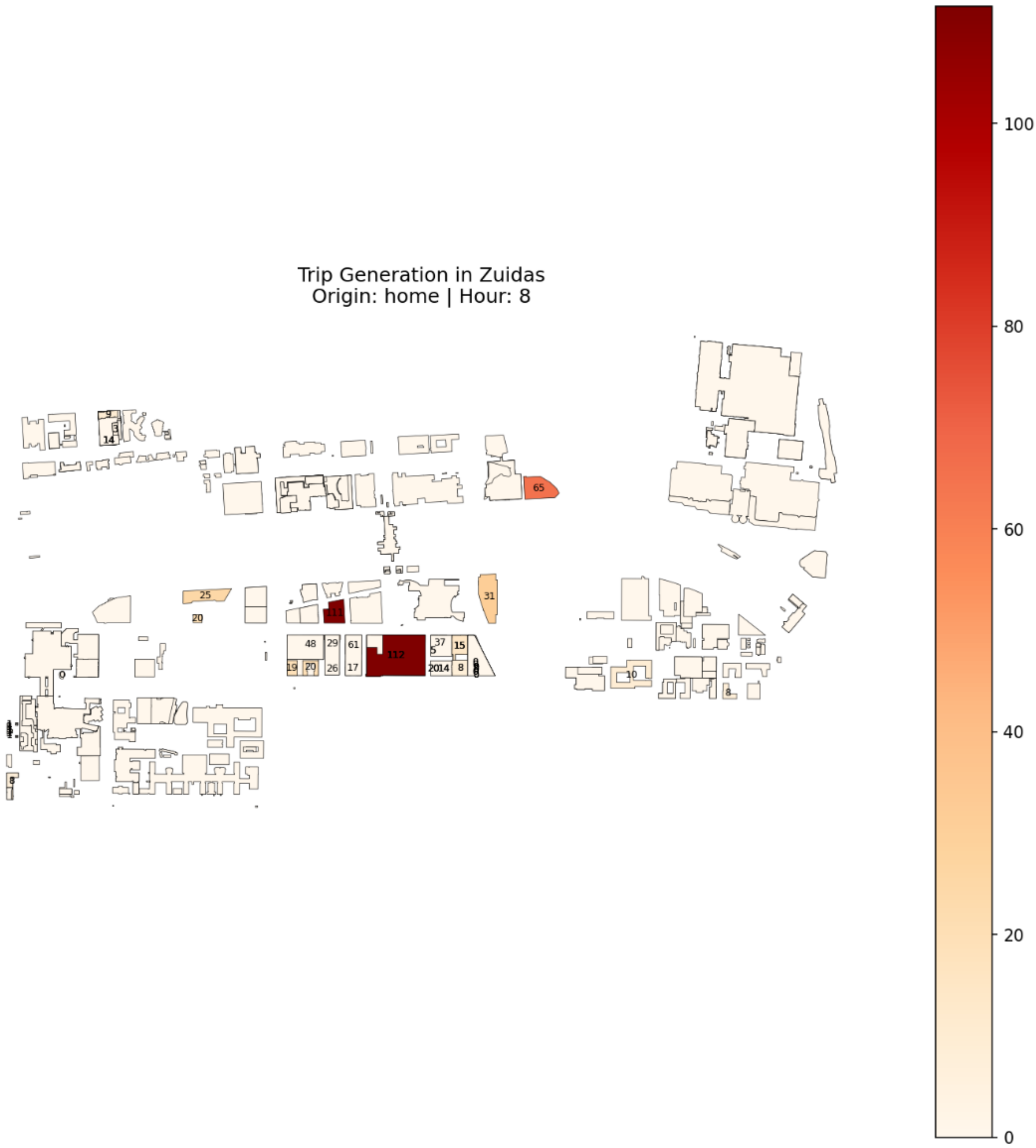


Figure 38: Generated Trips for homes at hour 8

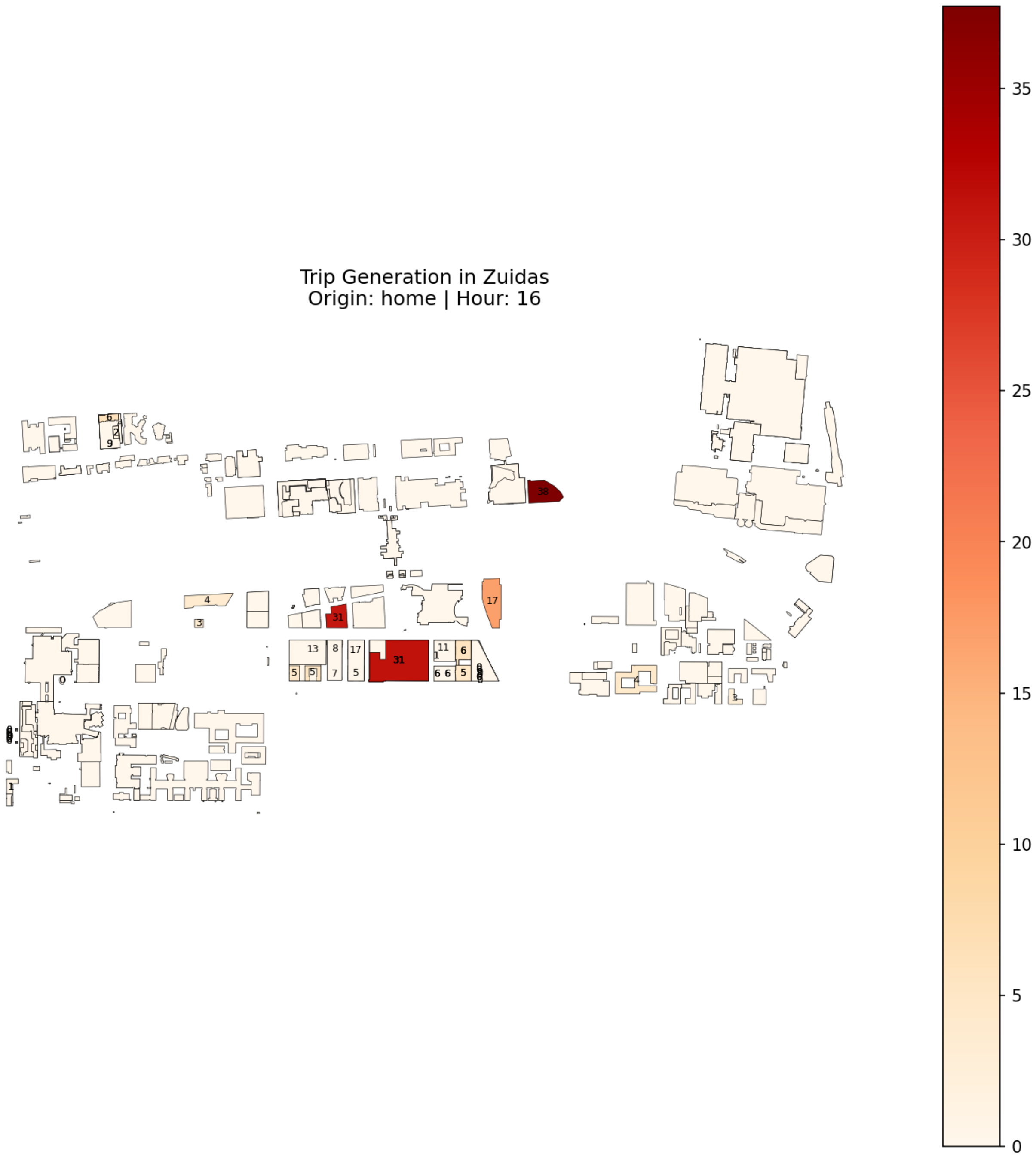


Figure 39: Generated trips for homes at hour 16

5.7 Model Face Validation

Conducting manual counts revealed several practical challenges as well as useful insights. At residential locations, both exits were visible, though occasionally obstructed, making it difficult to capture all movements. Office counts proved more complex due to group departures, brief trips to pick up bicycles or cars, and significant weather differences between observation days. Lunchtime walking appeared noticeably underreported in ODiN, while after-work peaks were strongly shaped by building-specific factors such as internal catering.

Leisure locations showed substantial variation between subtypes; cafés, brunch spots, restaurants, and bars, which the model cannot represent individually because they are aggregated into a single category. At the hospital, staff circulation, short visitor movements, and smokers stepping outside introduced additional uncertainty. Counting at the bus stop was particularly challenging due to group arrivals and its proximity to a train station. Tourist activity, which is not included in the model, added further variability.

Together with the limitations discussed in the methodology, these observations confirm that manual counts are best interpreted as plausibility checks rather than a basis for robust validation.

To assess the plausibility of the model outcomes, predicted walking trip volumes are compared with observed counts using a simple comparison factor (Table 8). In addition, predicted trips originating from Amsterdam Zuid train station are compared with pedestrian counts conducted by the municipality (Table 9). This second comparison provides a more robust benchmark, as the station benefits from a reliable activity-weight proxy and the counts cover multiple weekdays outside the vacation period.

Overall, the comparison between predicted and observed walking trips indicates that the model produces plausible results and captures the correct order of magnitude across locations. The highest accuracy is observed for origin types with study area–specific weighting data, namely residential buildings and the train station. For homes, population counts are used directly, while for the train station, daily check-ins and check-outs provide an accurate representation of activity levels. These data sources substantially improve model consistency and explain the strong performance for these origin types.

In contrast, offices and leisure locations are systematically underpredicted. These categories rely on indirect proxies of varying accuracy, with leisure locations in particular using the least precise proxies, which likely contributes to their underestimation. Bus stops show reasonably good performance despite relying on a different weighting method based on service frequency, number of lines, and vehicle capacity. This suggests that the estimation of activity levels for bus stops is appropriate and results in plausible predictions.

Notably, healthcare facilities, despite using study area–specific data for weight assignment, perform similarly to offices, which rely on Dutch building standards. This may be because the applied healthcare weight, based on the number of staff and beds, does not fully reflect pedestrian activity levels. Most walking trips associated with hospitals are generated by employees and visitors, rather than by patients or staff numbers alone, meaning that the chosen proxy may not adequately represent actual walking demand.

However, discrepancies for leisure locations cannot be attributed to weighting uncertainty alone. They also indicate that a key model assumption, that aggregating building types into broad functional categories does not significantly distort activity patterns, may not fully hold. In practice, cafés, restaurants, theatres, and parks generate distinct spatial and temporal walking patterns, which are not captured when these amenities are combined into a single leisure category.

A related issue arises for healthcare facilities due to limitations in mapping trip purposes to building types in ODiN. In the survey, only the trip purposes *services* and *personal care* are linked to healthcare buildings. Trips made by employees are classified under the *work* purpose, which is only associated with buildings classified as *work*, while *visiting* trips are coupled to residential buildings. As a result, a substantial share of hospital-related walking trips, particularly work-related and visitor trips, is not correctly attributed to the hospital location, leading to structural underestimation.

Further uncertainty stems from known limitations of ODiN, which underreports non-commuting trips and walking segments related to public transport access and egress. These limitations are clearly reflected in the face validation results. Temporal patterns show that the model performs best during the morning peak, indicating that commuting-related walking trips are generally well captured.

In contrast, midday and evening trips, including lunch breaks and return trips, are more frequently underestimated, likely because these trip types are underreported in ODiN. Consequently, the largest factor differences are observed for trip purposes that are known to be less reliably captured in the survey, such as lunchtime movements from offices and return trips to home.

Trips originating from bus stops were expected to be underpredicted due to the known underreporting of short walking legs in ODiN. However, the results show overprediction in the morning and underprediction in the evening. This pattern may partly be explained by the use of summer holiday counts, which likely reduced or shifted commuting activity. Additionally, the proxy-based weighting for bus stops may be higher than actual activity levels, while underreporting of short transfer walks and return trips in ODiN further amplifies discrepancies later in the day.

Table 8: Counts and Results

Origin Type	Time Period	Counted [persons]	Predicted [persons]	Factor Difference [-]
Home	8:00-9:00	(28+20)/2	18	1.33
Home	9:00-10:00	(14+12)/2	6	2.17
Office	11:00-12:00	(207+37)/2	5	24.4
Office	12:00-13:00	(355+39)/2	12	16.42
Office	16:00-17:00	(292+186)/2	28	8.54
Office	17:00-18:00	(354+315)/2	36	9.29
Leisure (coffee)	8:00-9:00	8	3	2.67
Leisure (coffee)	9:00-10:00	28	5	5.6
Leisure (brunch)	11:00-12:00	16	9	1.77
Leisure (brunch)	12:00-13:00	33	12	2.75
Leisure (dinner)	11:00-12:00	0	9	-
Leisure (dinner)	12:00-13:00	17	11	1.55
Leisure (dinner)	16:00-17:00	3	19	0.16
Leisure (dinner)	17:00-18:00	9	18	0.5
Healthcare	8:00-9:00	5	1	5
Healthcare	9:00-10:00	39	4	9.75
Healthcare	11:00-12:00	20	2	10
Healthcare	12:00-13:00	-	2	-
Healthcare	16:00-17:00	82	5	16.4
Healthcare	17:00-18:00	16	1	16
Bus Stop	8:00-9:00	167	261	0.64
Bus Stop	9:00-10:00	178	294	0.61
Bus Stop	11:00-12:00	109	136	0.80
Bus Stop	12:00-13:00	154	77	2
Bus Stop	16:00-17:00	208	75	2.77
Bus Stop	17:00-18:00	235	156	1.51

The train station shows the best overall performance in the model, likely because its weight is based on actual daily check-ins and check-outs, highlighting the importance of accurate building weight information. Nevertheless, substantial discrepancies remain: the model overestimates demand at time period 7:00-8:00, underestimates them at time period 8:00-9:00 and 9:00-10:00, underpredicts trips during lunch hours, and drops to zero at 19:00, likely reflecting underrepresentation in ODiN for certain periods.

Table 9: Counts and Results for train station

Origin Type	Time Period	Counts [persons]	Predicted [persons]	Factor Difference [-]
Train Station	7:00-8:00	431	1393	0.31
Train Station	8:00-9:00	1777	1144	1.55
Train Station	9:00-10:00	967	270	3.58
Train Station	11:00-12:00	358	162	2.21
Train Station	12:00-13:00	645	196	3.29
Train Station	13:00-14:00	640	163	3.93
Train Station	16:00-17:00	302	164	1.84
Train Station	17:00-18:00	442	251	1.76

Train Station	18:00-19:00	368	135	2.73
Train Station	19:00-20:00	211	0	-
Train Station	20:00-21:00	146	45	3.24

Taken together, the results indicate that model accuracy is strongly influenced by the quality of the underlying assumptions, particularly those related to activity weighting and the use of ODiN data. The analysis shows that reliable, area-specific weighting data substantially improves the consistency and accuracy of the predicted walking volumes, as demonstrated for homes and the train station. However, even where such data are unavailable, maintaining the correct order of magnitude proves to also result in plausible results.

Large factor differences occur for trip types that ODiN is known to underreport, such as lunch-break movements, short access and egress legs, and return trips later in the day. As a result, the model performs most reliably for morning peak periods, when commuting trips dominate and the correspondence between trip purposes and building types is more straightforward. In contrast, predictions for offices, leisure locations, and healthcare are more uncertain because these locations often involve aggregated building types for which no direct mapping to ODiN trip purposes exists.

Overall, the discrepancies observed in the face validation do not indicate a failure of the modelling approach, but rather highlight its sensitivities and limitations. The model produces plausible results, preserves the correct order of magnitude, and captures key spatial and temporal variations in walking demand. At the same time, the findings provide clear guidance for future refinement, particularly through improved activity weighting and a more nuanced treatment of trip purposes that are poorly represented in survey data.

6. Discussion

This chapter interprets the results of the building-level pedestrian trip generation model (BPT-Gen) adapted for the Dutch urban context, situating the findings within existing research and evaluating their significance and addresses the limitations of the model.

In section 6.1 results and assumptions in the BPT-Gen model are compared with models from the literature review. This is followed by section 6.2 which names the limitations of the BPT-Gen model.

6.1 Relation to Prior Work

The findings of this study demonstrate that adapting the UNA-based pedestrian trip generation model to the Dutch context is feasible, though model performance depends heavily on the quality of weighting proxies for different building types. These results can be contextualized with respect to prior pedestrian trip generation studies, including Clifton et al. (2016), Tian et al. (2017), and hybrid pedestrian models.

A key extension of the original UNA framework (Sevtsuk et al., 2024) is the incorporation of hourly trip rates, enabling temporal prediction of pedestrian demand. While the use of trip rates in pedestrian modelling is not entirely novel. Clifton et al. (2016) also apply them, the differences in implementation are significant. Clifton's trip rates are linked to households rather than individual buildings, because their framework uses parcel zones as origins. This reduces the model's ability to detect localized hubs of pedestrian activity, such as malls or recreational facilities. On the other hand, small uniform grid cells or parcel zones may introduce unnecessary complexity or aggregation bias, particularly for larger land uses such as parks or campuses. In contrast, the building-level approach used in BPT-Gen is better suited for neighbourhood-scale mixed-use areas, while Clifton's framework is more appropriate for regional-scale analyses.

Tian et al. (2017) also use trip rates linked to households, and while their approach allows single buildings as origins, it is limited to residential buildings. In comparison, the BPT-Gen model can generate trips from a wider variety of origin types, including offices, leisure facilities, and transport hubs. This broader scope makes the model less dependent on socio-demographic inputs, although it does limit the ability to capture demographic influences on walking behaviour, an aspect that Tian et al. emphasize as a strong predictor of pedestrian activity.

Another important distinction is temporal resolution. Clifton et al. use daily trip rates, whereas BPT-Gen employs hourly trip rates, revealing temporal patterns that aggregated daily totals obscure. For example, BPT-Gen identifies underpredictions during midday and evening for offices, leisure locations, and walking legs to other modes, which aligns with travel-demand literature (e.g., Lugtig et al., 2022) indicating that incidental, discretionary, and short access trips are often underreported in household surveys. These temporal discrepancies highlight the value of using higher-resolution trip rates to capture behavioural nuances that daily totals miss.

Accessibility is another under-researched aspect in macroscopic pedestrian modelling. While Clifton et al. (2016) and Tian et al. (2017) show that longer distances reduce the likelihood of walking to a destination, the impact of origin accessibility is rarely considered. In BPT-Gen, accessibility adjustments using network-based reach indices allow building weights to reflect behavioural tendencies, providing a more nuanced view of pedestrian demand.

Hybrid models, such as Ullrich's, demonstrate strong predictive performance but require extensive, heterogeneous datasets, including household surveys, detailed infrastructure data, and street-level counts. These models are typically limited to scenarios similar to the conditions under which they were trained. In contrast, BPT-Gen relies exclusively on openly available datasets (BAG, OSM, ODiN) and can be applied more flexibly to different case studies. The trade-off is that building-level activity must sometimes be inferred from proxies, introducing additional uncertainty.

Overall, comparing BPT-Gen to existing literature highlights several advantages. The building-level, hourly approach reduces aggregation bias, captures temporal variability, and can represent diverse origin types beyond households. At the same time, it introduces new sensitivities, particularly around proxy selection, trip purpose reporting, and accessibility assumptions. Unlike zone-based or socio-demographic models, BPT-Gen primarily relies on land-use structure and observable activity indicators, making it well-suited to fine-grained Dutch urban contexts. In conclusion, this framework complements existing models, offering new insights into pedestrian behaviour at high spatial and temporal resolution, while also emphasizing the importance of careful data selection and methodological choices to ensure reliability.

6.2 Limitations of the BPT-Gen Model

Even though the model performs best for building types supported by robust activity data and well-reported ODiN trip purposes, several sources of uncertainty remain. These limitations affect how confidently results can be interpreted, particularly in areas with mixed land uses or buildings represented through proxies.

A key limitation lies in the aggregation required to map ODiN trip purposes to BAG and OSM building types. Diverse amenities are merged into broad functional categories, workplace-related purposes are combined despite differences in employment types, and several ODiN purposes do not align neatly with specific land uses. This harmonisation is necessary for model implementation but inevitably obscures behavioural variation, masks differences in opening hours, usage patterns, and simplifies temporal activity profiles.

The construction of building weights relies on several structural assumptions. The method employs multiple proxy strategies depending on building type. For example, redistributing residential totals over GFA, or estimating public transport activity from service frequency and number of lines. These methodological choices, and the lack of building-specific data such as residents per building, students per school, or boarding counts per station, introduce potential biases and uncertainty.

The accessibility component depends on assumptions regarding walking distances and catchment sizes. Using the national mean walking distance from ODiN ensures consistency but may not reflect local mobility patterns, particularly in dense or highly mixed areas. Sensitivity analysis confirmed that the choice of walking radius and normalization method substantially alters the distribution of building weights, indicating that accessibility parameters represent an additional source of structural uncertainty.

The model inherits underreporting biases in ODiN, especially for short access/egress movements, incidental leisure, and midday office trips. These missing trip types contribute to systematic underprediction for buildings whose weighting relies heavily on proxies, such as offices, leisure facilities, and bus stops.

Finally, the model could not be formally calibrated or validated due to the limited availability of high-quality pedestrian count data. The manual counts conducted for this study were short in duration, influenced by summer conditions, and subject to undercounting, particularly where visibility was constrained or where it was unclear whether people were walking or switching to another mode. As a result, the counts serve only as indicative reference points rather than statistically reliable benchmarks.

Together, these limitations imply that model outputs should be interpreted with greater confidence for building types grounded in robust, activity-specific data and with more caution where proxy measures, aggregated categories, or underreported trip purposes dominate.

7. Conclusion

This thesis examined to what extent an existing pedestrian trip-generation model can be adapted to the Dutch urban context using only open data and locally relevant behavioural parameters. The Urban Network Analysis (UNA) framework was reconfigured by classifying buildings with BAG and OSM, recalculating activity-based building weights, integrating network-based accessibility, and incorporating hourly walking trip rates from ODiN. The resulting BPT-Gen model produces building-level, hourly pedestrian trip generation for Dutch cities without relying on household surveys or extensive counts.

The main findings show that the adapted model successfully reproduces realistic spatial and temporal patterns for building types supported by reliable activity data; particularly housing and major train stations. Underprediction occurs for offices, leisure facilities, and health care functions, where weights rely heavily on proxy indicators, where ODiN underreports walking trips or where trip purposes do not map well to the building type. Sensitivity analyses confirm that trip generation outcomes depend strongly on proxy choice and accessibility assumptions, highlighting important structural uncertainties.

Despite these uncertainties, the model already provides clear practical value. It identifies the main generators of walking trips, peak hours, and accessibility-driven hotspots, even where some building types rely on imperfect proxies. This makes it useful for urban design, scenario testing, and accessibility analysis, where insight into relative patterns often matters more than perfectly calibrated totals. The model also fills a gap in existing literature by providing the first open-data, building-level pedestrian trip generation model for Dutch cities. By explicitly linking trip purposes to building types at hourly resolution, it strengthens the trip-generation step itself and creates a transparent and reproducible foundation for future pedestrian modelling.

For research and practice, these results imply that open-data pedestrian models can offer meaningful insights even without extensive calibration datasets. The BPT-Gen framework enables planners to explore how walking activity responds to land-use changes or network modifications and supports accessibility-oriented planning approaches. At the same time, the findings underline the importance of more accurate activity indicators, better representation of workplace and leisure walking, and validation using street-level flow data once trip distribution and assignment are incorporated.

Looking ahead, several steps would strengthen the model. First, ambiguity in building-exit counts could be reduced by generating all trips first and then applying a mode-choice model, instead of modelling only walking trips. Second, validation should move beyond building-exit plausibility checks by comparing assigned pedestrian flows against independent street-level counts once the full modelling chain is implemented. Third, deriving trip rates from ODiN could be improved through a more detailed household survey that directly links destinations to building types. Finally, extending this work to include trip distribution and assignment will allow the complete pedestrian modelling framework to be tested, refined, and applied in practice.

8. References

3D BAG (Basisregistratie Adressen en Gebouwen). (n.d.). EduGIS. Retrieved August 21, 2025, from https://www.edugis.nl/project/3d-bag-basisregister/

AlleCijfers.nl. (n.d.). *AlleCijfers.nl*. AlleCijfers.Nl. https://allecijfers.nl/

Al-Masaeid, H. R., Khedaywi, T., Shehadeh, E., & Al-Shafie, R. (2021). *Trip and Parking Generation of Hospitals and Medical Centers in Jordan*. *15*(4). https://www.researchgate.net/publication/353885038_Trip_and_Parking_Generation_of_Hospitals_and_Medical_Centers_in_Jordan

Bhat, C., Handy, S., Kockelman, K., Mahmassani, H., Chen, Q., Srour, I., & Weston, L. (2021). *ASSESSMENT OF ACCESSIBILITY MEASURES* (No. FHWA/TX-01/4938-3). https://www.researchgate.net/profile/Issam-Srour/publication/277295747_ASSESSMENT_OF_ACCESSIBILITY_MEASURES/links/55b4d99e08ae9289a08a5698/ASSESSMENT-OF-ACCESSIBILITY-MEASURES.pdf

BRIS. (n.d.). *Bouwbesluit online*. BRIS. https://www.bouwbesluitonline.nl/docs/wet/bb2012?q=bouwbesluit

CBS. (n.d.-a). *Dutch National Travel survey*. CBS. https://www.cbs.nl/en-gb/our-services/methods/surveys/brief-survey-description/dutch-national-travel-survey

CBS. (n.d.-b). *Kerncijfer wijken en buurten 2023*. Centraal Bureau Voor de Statistiek. https://www.cbs.nl/nl-nl/maatwerk/2025/13/kerncijfers-wijken-en-buurten-2023

Clifton, K. J., Singleton, P. A., Muhs, C. D., & Schneider, R. J. (2016). Representing pedestrian activity in travel demand models: Framework and application. *Journal of Transport Geography*, *52*, 111–122. https://doi.org/10.1016/j.jtrangeo.2016.03.009

de Haas, M., & Hamersma, M. (2019). *Walking Facts*. Ministry of Infrastructure and Water Management. https://english.kimnet.nl/documents/2019/10/3/walking-facts

Deng, Y., Zhu, Y., Wang, H., Khadka, V. S., Hu, L., Ai, J., Dou, Y., Li, Y., Dai, S., Mason, C. E., Wang, Y., Jia, W., Zhang, J., Huang, G., & Jiang, B. (2019). Ratio-Based Method To Identify True Biomarkers by Normalizing Circulating ncRNA Sequencing and Quantitative PCR Data. *Analytical Chemistry*, *91*(10), 6746–6753. https://doi.org/10.1021/acs.analchem.9b00821

ESMO. (n.d.). *ESMO*. VU University Medical Center, Department of Medical Oncology. https://www.esmo.org/designated-centres/vu-university-medical-center-department-of-medical-oncology

Ewing, R., & Cervero, R. (2010). Travel and the Built Environment: A Meta-Analysis. *Journal of the American Planning Association*, *76*(3), 265–294. https://doi.org/10.1080/01944361003766766

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

Hankey, S., Lindsey, G., Wang, X., Borah, J., Hoff, K., Utecht, B., & Xu, Z. (2012). Estimating use of non-motorized infrastructure: Models of bicycle and pedestrian traffic in Minneapolis, MN. *Landscape and Urban Planning*, *107*(3), 307–316. https://doi.org/10.1016/j.landurbplan.2012.06.005

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

Jacobs-Crisioni, C., Rietveld, P., Koomen, E., & Tranos, E. (2014). Evaluating the Impact of Land-Use Density and Mix on Spatiotemporal Urban Activity Patterns: An Exploratory Study Using Mobile Phone Data. *Environment and Planning A: Economy and Space*, *46*(11), 2769–2785. https://doi.org/10.1068/a130309p

Johnstone, D., Krista, N., & Lowry, M. (2017). *Collecting Network-wide Bicycle and Pedestrian Data: A Guidebook for When and Where to Count* (No. WA-RD 875.1). Portland State University. Transportation Research and Education Center. /view/dot/34663

Kaziyeva, D., Loidl, M., & Wallentin, G. (2021). Simulating Spatio-Temporal Patterns of Bicycle Flows with an Agent-Based Model. *ISPRS International Journal of Geo-Information*, *10*(2), 88. https://doi.org/10.3390/ijgi10020088

Kaziyeva, D., Stutz, P., Wallentin, G., & Loidl, M. (2023). Large-scale agent-based simulation model of pedestrian traffic flows. *Computers, Environment and Urban Systems*, *105*, 102021. https://doi.org/10.1016/j.compenvurbsys.2023.102021

Kuiper, J. M. (2021). *Using Dutch Land and Property Data to Improve Trip Generation based on Open Data* [University of Twente]. https://essay.utwente.nl/89024/1/Kuiper_MA_ET.pdf

Lee, M., & Holme, P. (2015). Relating Land Use and Human Intra-City Mobility. *PLOS ONE*, *10*(10), e0140152. https://doi.org/10.1371/journal.pone.0140152

Lugtig, P., Roth, K., & Schouten, B. (2022). Nonresponse analysis in a longitudinal smartphone-based travel study. *Survey Research Methods*, 13-27 Pages. https://doi.org/10.18148/SRM/2022.V16I1.7835

Moeckel, R., Kuehnel, N., Llorca, C., Moreno, A. T., & Rayaprolu, H. (2020). Agent-Based Simulation to Improve Policy Sensitivity of Trip-Based Models. *Journal of Advanced Transportation*, *2020*, 1–13. https://doi.org/10.1155/2020/1902162

Normalization, standardization and data transformation. (n.d.). My Research Lab. https://www.myrelab.com/learn/normalization-standardization-and-data-transformation

NS. (n.d.). *Reizigersgedrag 2024*. NS Dashboards. https://dashboards.nsjaarverslag.nl/reizigersgedrag/amsterdam-zuid?utm_source=chatgpt.com

Pan, Q., & Sharifi, S. (n.d.). First Step of Four Step Modeling (Trip Generation). In *Transportation Land Use Modeling and Policy*. Mavs Open Press. https://uta.pressbooks.pub/oertransportlanduse/chapter/chapter-10-first-step-of-four-step-modeling-trip-generation/

PTV Group. (n.d.). *Why Would You Use a Traffic Simulation?* PTV GROUP Part of Umovity. https://www.ptvgroup.com/en/application-areas/traffic-simulation

Rietveld, piet. (2001). *Environmental Effects of Public Transport*. universiteit Amsterdam, and Tinbergen institute. https://papers.tinbergen.nl/01108.pdf

Ryus, P., Musunuru, A., Bonneson, J., Kothuri, S., Monsere, C., McNeil, N., LaJeunesse, S., Nordback, K., Kumfer, W., & Currin, S. (2022). *Guide to Pedestrian Analysis*. Transportation Research Board.

Santana Palacios, M., & El-geneidy, A. (2022). Cumulative versus Gravity-based Accessibility Measures: Which One to Use? *Findings*. https://doi.org/10.32866/001c.32444

Sevtsuk, A., & Alhassan, A. (2025). Madina Python package: Scalable urban network analysis for modeling pedestrian and bicycle trips in cities. *Journal of Transport Geography*, *123*, 104130. https://doi.org/10.1016/j.jtrangeo.2025.104130

Sevtsuk, A., Basu, R., & Chancey, B. (2021). We shape our buildings, but do they then shape us? A longitudinal analysis of pedestrian flows and development activity in Melbourne. *PLOS ONE*, *16*(9), e0257534. https://doi.org/10.1371/journal.pone.0257534

Sevtsuk, A., & Kalvo, R. (2025). Modeling pedestrian activity in cities with urban network analysis. *Environment and Planning B: Urban Analytics and City Science*, *52*(2), 377–395. https://doi.org/10.1177/23998083241261766

Sevtsuk, A., Kollar, J., Pratama, D., Basu, R., Haddad, J., Alhassan, A., Chancey, B., Halabi, M., Makhlouf, R., & Abou-Zeid, M. (2024). Pedestrian-oriented development in Beirut: A framework for estimating urban design impacts on pedestrian flows through modeling, participatory design, and scenario analysis. *Cities*, *149*, 104927. https://doi.org/10.1016/j.cities.2024.104927

Stack Overflow. (2022). *Add node between existing edge in Networkx Graph generated by OSMnx*. Stack Overflow. https://stackoverflow.com/questions/72523683/add-node-between-existing-edge-in-networkx-graph-generated-by-osmnx?utm_source=chatgpt.com

Sweere, N. (2022). *IMPROVING THE QUALITY OF STAY: A TRANSITION TOWARDS A CAR-FREE CITY* [university of Groningen]. https://frw.studenttheses.ub.rug.nl/3847/1/MasterthesisEIP_NikkiSweere_Finalversion.pdf

Takyi, I. (1990). *TRIP GENERATION ANALYSIS IN A DEVELOPING COUNTRY CONTEXT* (Transportation Research Record). https://www.semanticscholar.org/paper/TRIP-GENERATION-ANALYSIS-IN-A-DEVELOPING-COUNTRY-Takyi/bc0d91a9099599726a2d36722a0a4bf5bbf3c57b

Tian, G., & Ewing, R. (2017). A walk trip generation model for Portland, OR. *Transportation Research Part D: Transport and Environment*, *52*, 340–353. https://doi.org/10.1016/j.trd.2017.03.017

TNO. (2023, June). *XCARCITY: effective digital twins for tomorrow's low-traffic city*. TNO Innovation for Life. https://www.tno.nl/en/newsroom/insights/2023/06/xcarcity-effective-digital-twins/

Ullrich, A., Hunger, F., Stavroulaki, I., Bilock, A., Jareteg, K., Tarakanov, Y., Gösta, A., Quist, J., Berghauser Pont, M., & Edelvik, F. (2024). A hybrid workflow connecting a network and an agent-based model for predictive pedestrian movement modelling. *Frontiers in Built Environment*, *10*, 1447377. https://doi.org/10.3389/fbuil.2024.1447377

Verwoerd, F. (2024). *Office Occupier Benchmark Report 2024*. https://www.cbre.nl/en-gb/insights/reports/cbre-office-occupancy-benchmark-report

VU. (n.d.). *This page contains current numbers of the students and staff of Vrije Universiteit Amsterdam*. Vrije Universiteit Amsterdam. https://vu.nl/en/about-vu/more-about/vu-in-numbers

Zhang, Q., Moeckel, R., & Clifton, K. J. (2024). MoPeD meets MITO: A hybrid modeling framework for pedestrian travel demand. *Transportation*, *51*(4), 1327–1347. https://doi.org/10.1007/s11116-022-10365-x

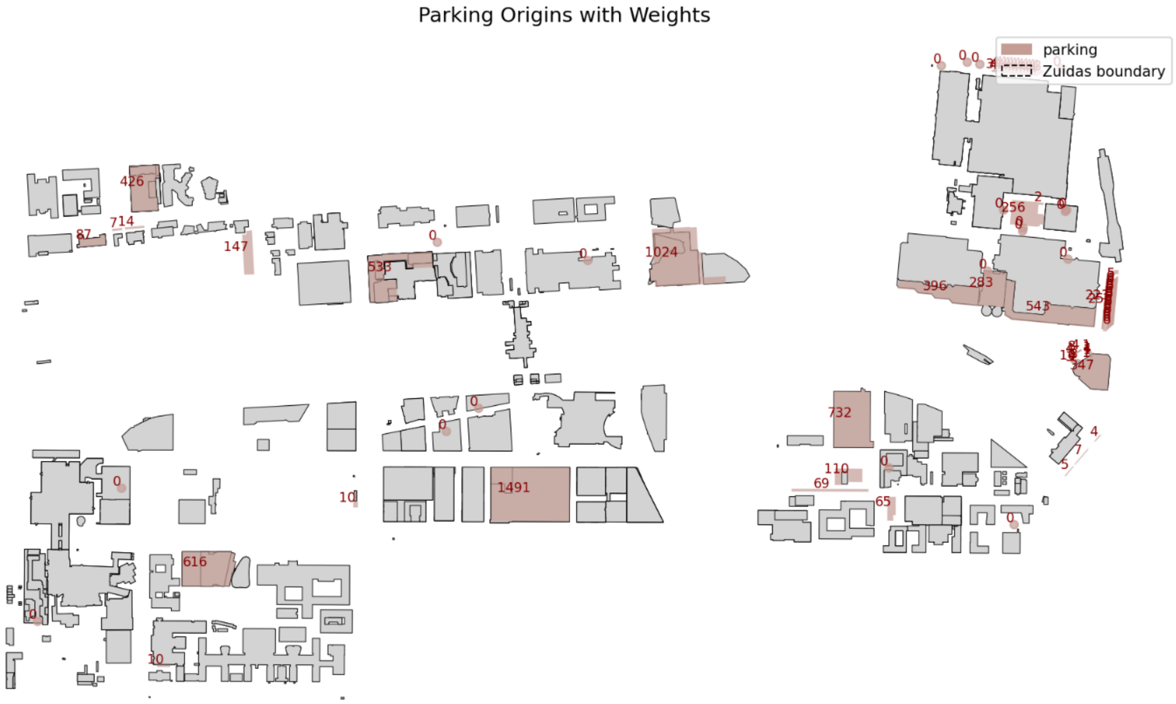
Zhou, Q., Wang, S., & Liu, Y. (2022). Exploring the accuracy and completeness patterns of global land-cover/land-use data in OpenStreetMap. *Applied Geography*, *145*, 102742. https://doi.org/10.1016/j.apgeog.2022.102742

Zuidas. (n.d.). *Zuidas*. Zuidas, Houdt Je Op de Hoogte. https://zuidas.nl/

(N.d.).

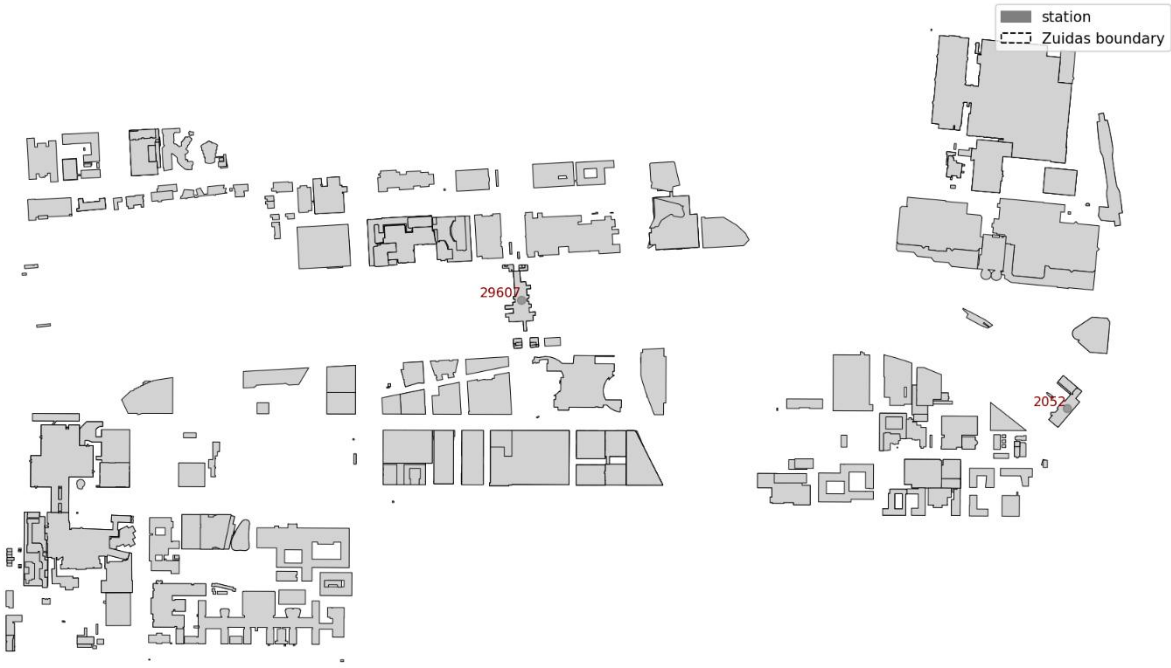
9. Appendix

9.1 Appendix 1: Building Weights

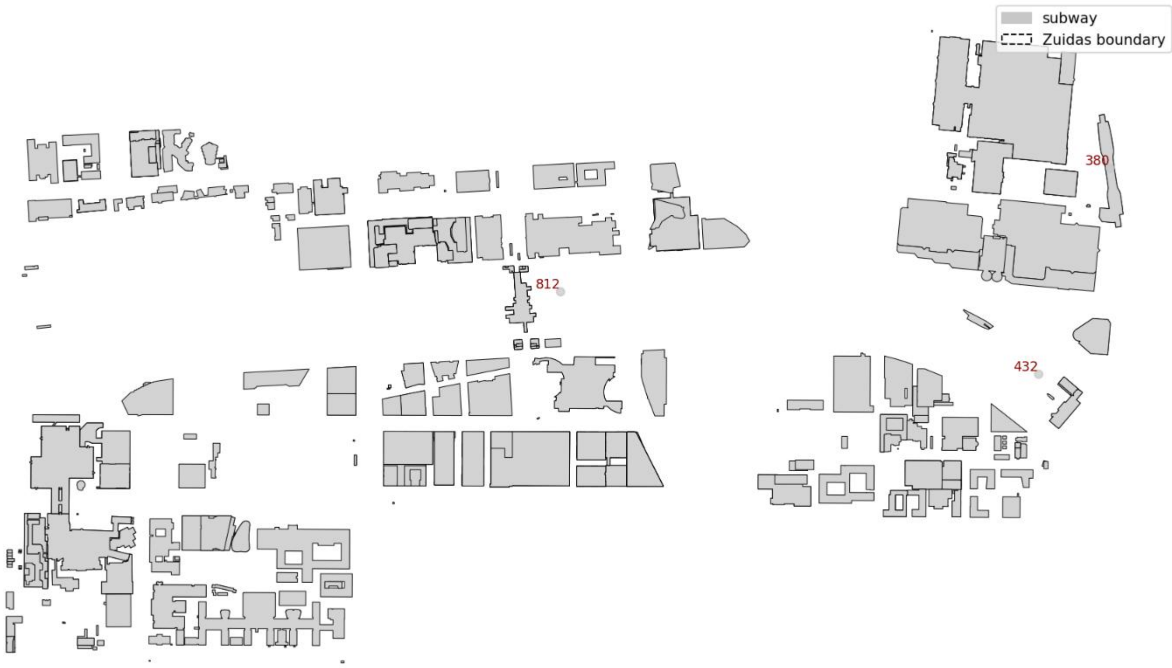


Sportfunctie Origins with Weights

Station Origins with Weights



Subway Origins with Weights



Tram_Stop Origins with Weights

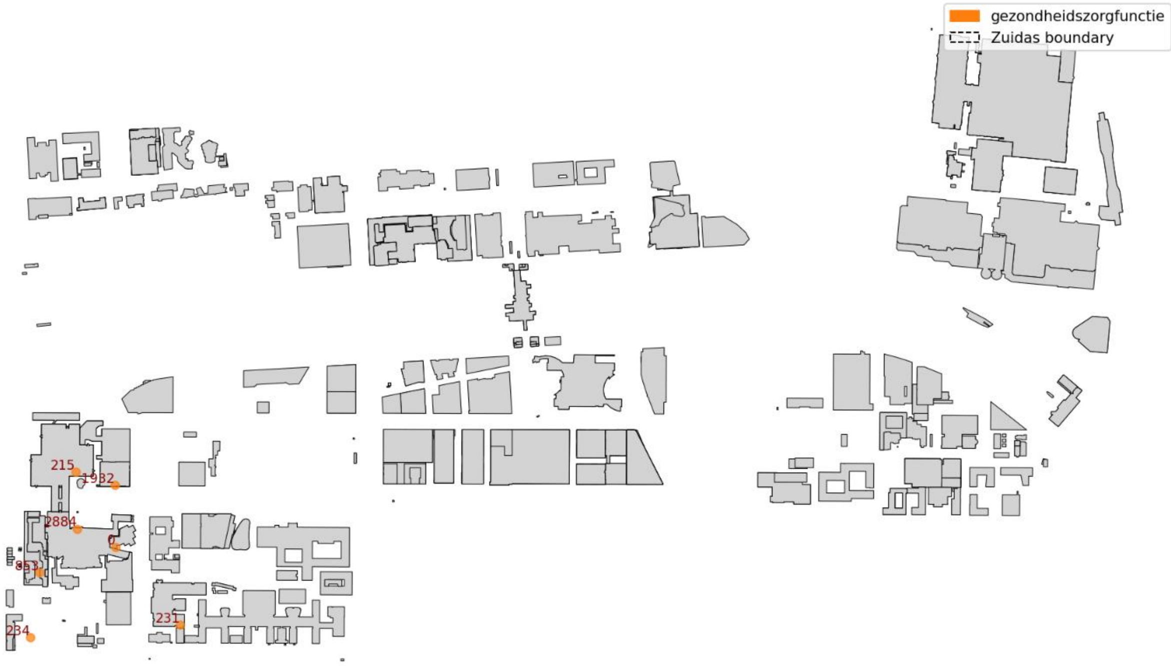
Winkelfunctie Origins with Weights

Woonfunctie (Residential) Origins: Weight Summed per Pand

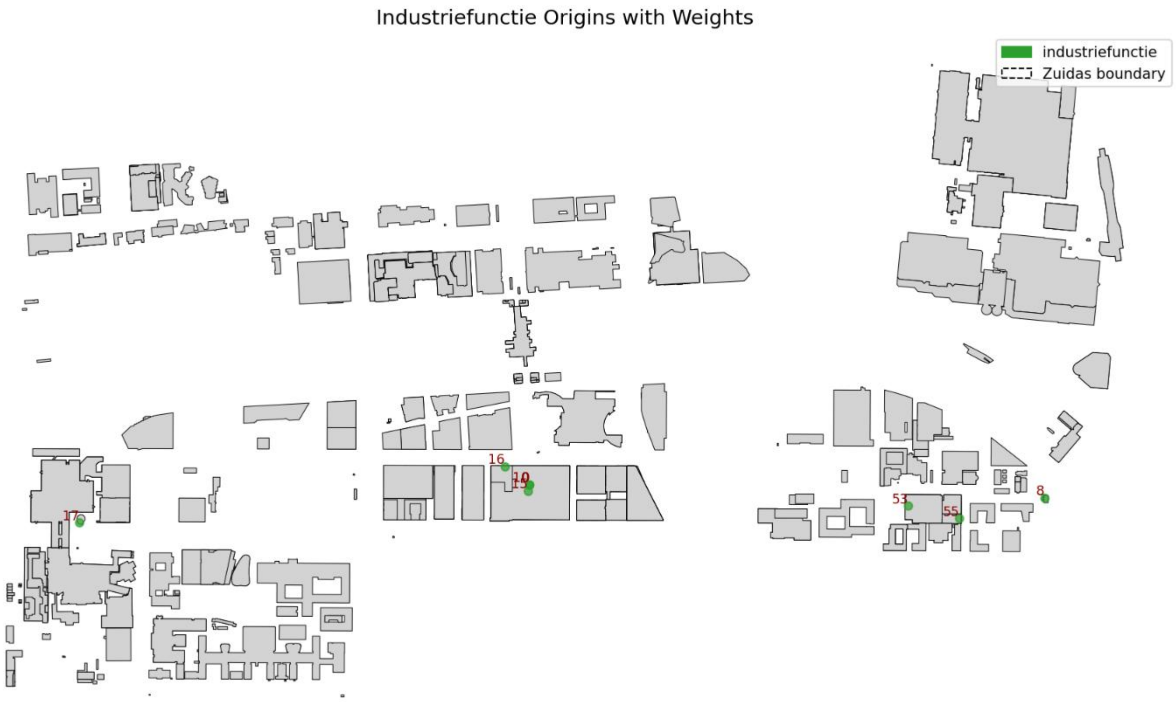
Bicycle_Parking Origins with Weights

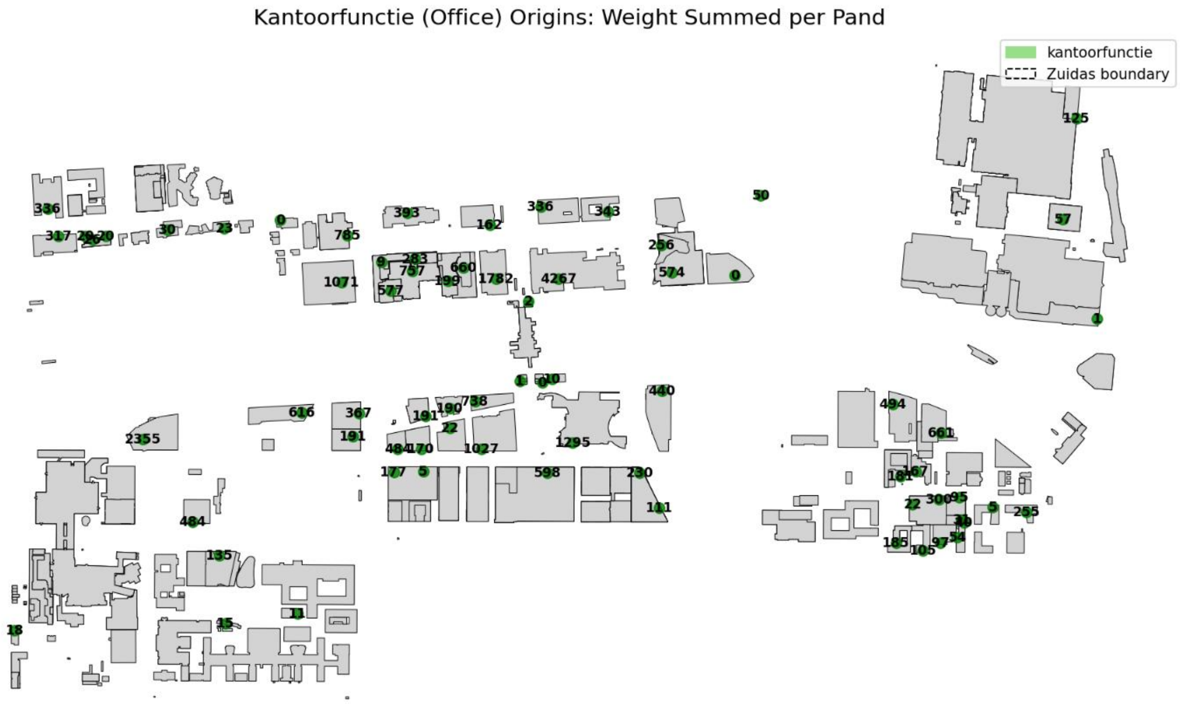
Bijeenkomstfunctie Origins with Weights

Gezondheidszorgfunctie Origins with Weights



Horeca Origins with Weights


Industriefunctie Origins with Weights
industriefunctie
Zuidas boundary

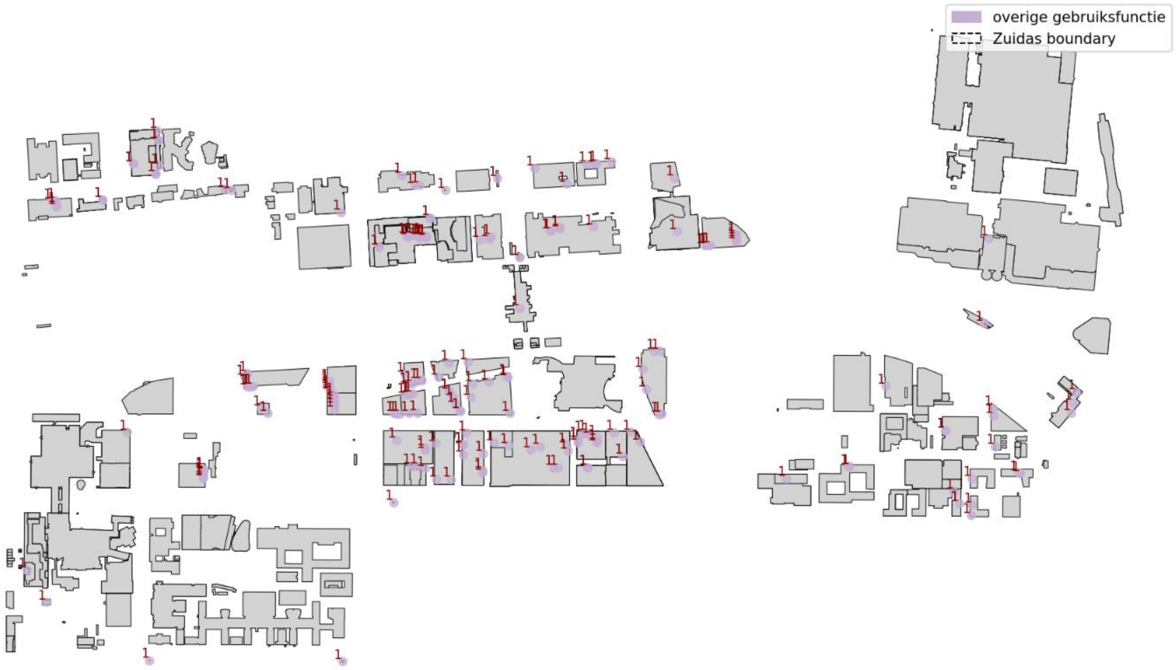

Kantoorfunctie (Office) Origins: Weight Summed per Pand
kantoorfunctie
Zuidas boundary

Logiesfunctie Origins with Weights

Motorcycle_Parking Origins with Weights

Onderwijsfunctie Origins with Weights

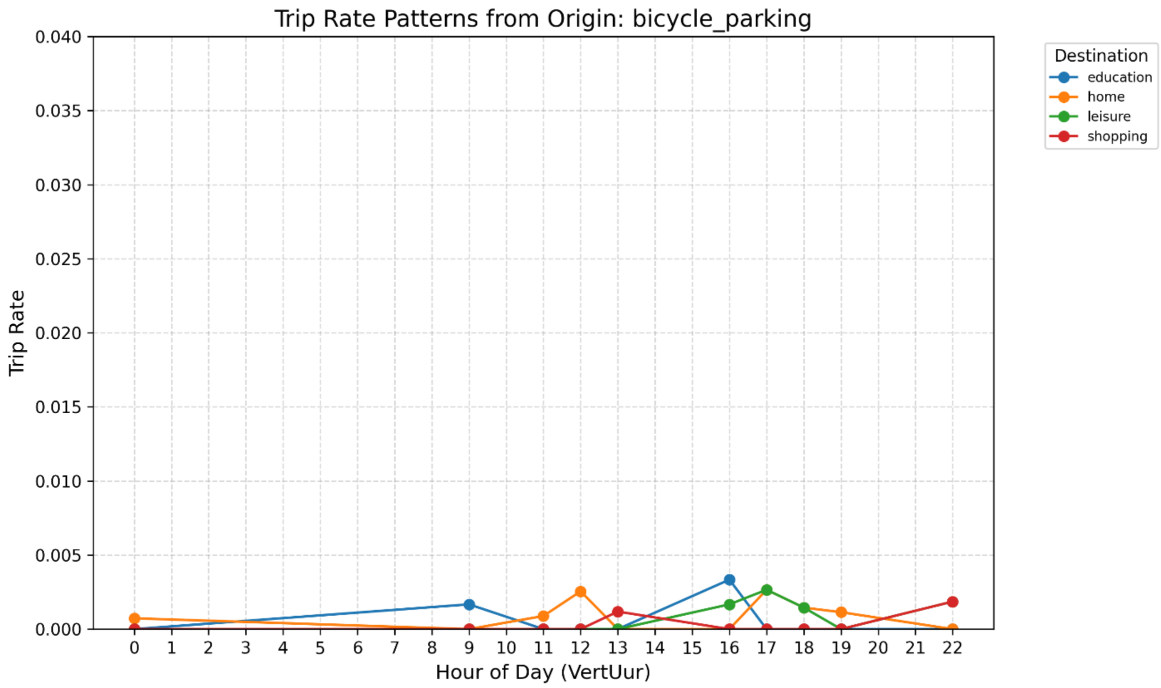
Overige Gebruiksfunctie Origins with Weights



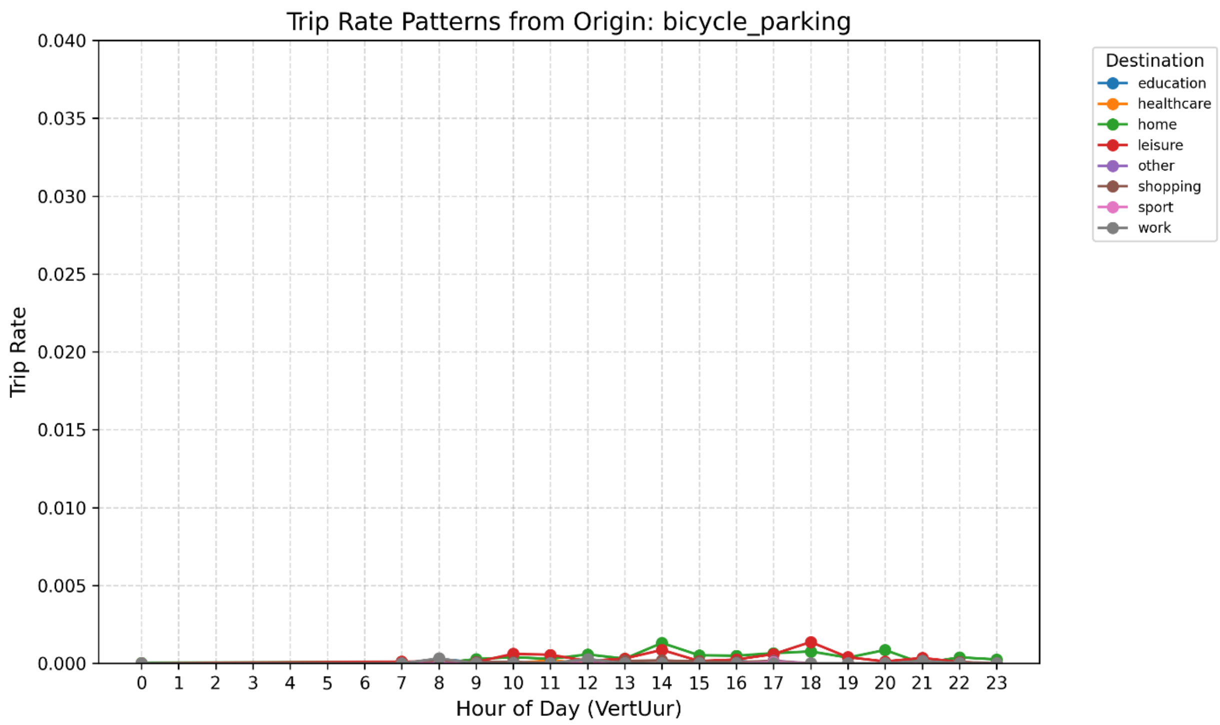
Park Origins with Weights

9.2 Appendix 2: Trip Rate Patterns

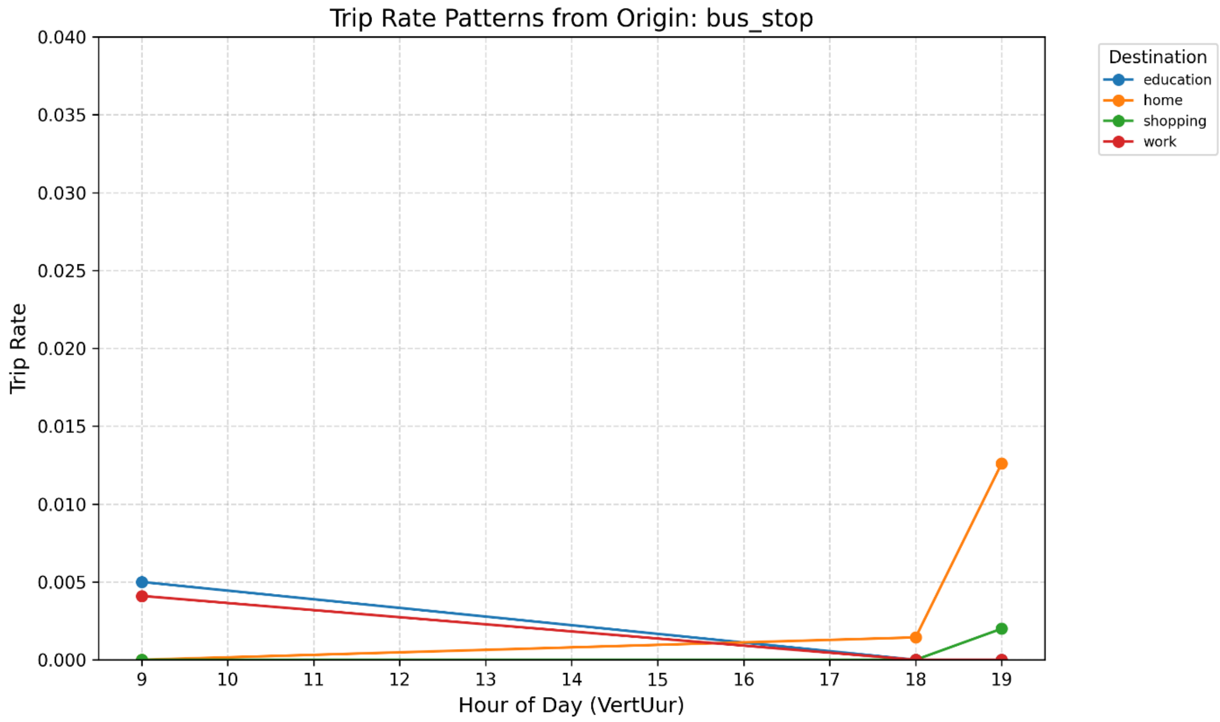
Similar Postcode Areas:



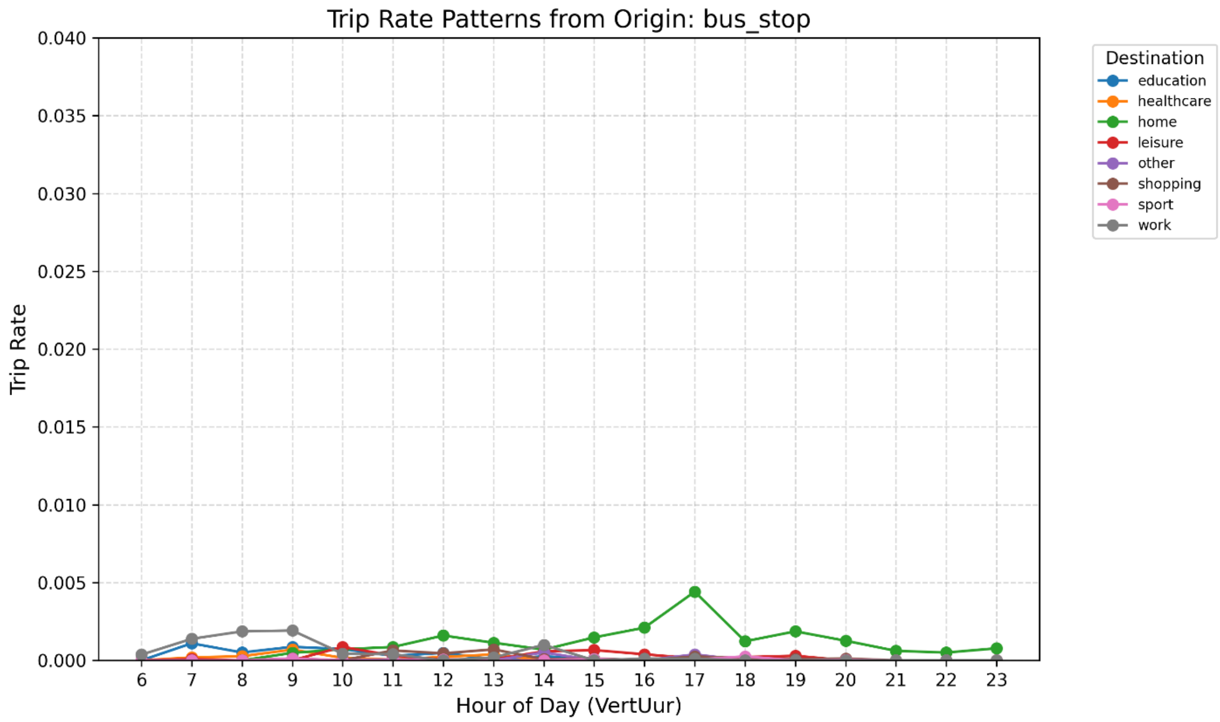
Urban Municipalities:



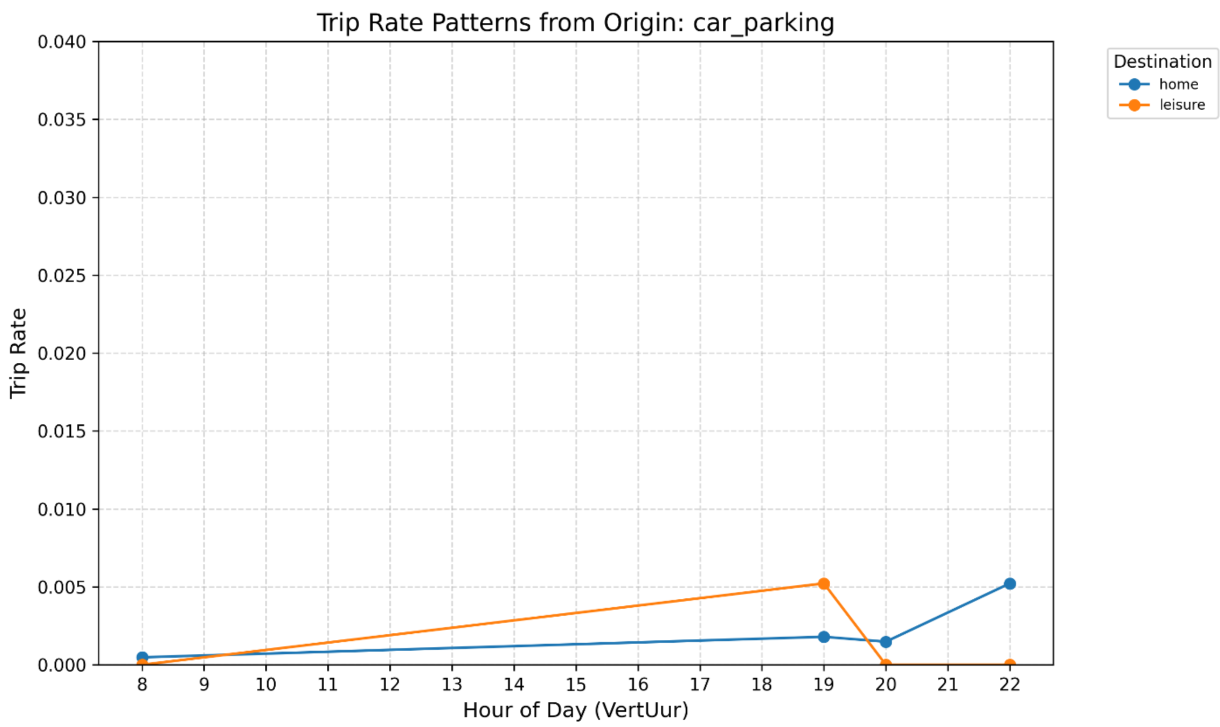
Similar Postcode Areas:



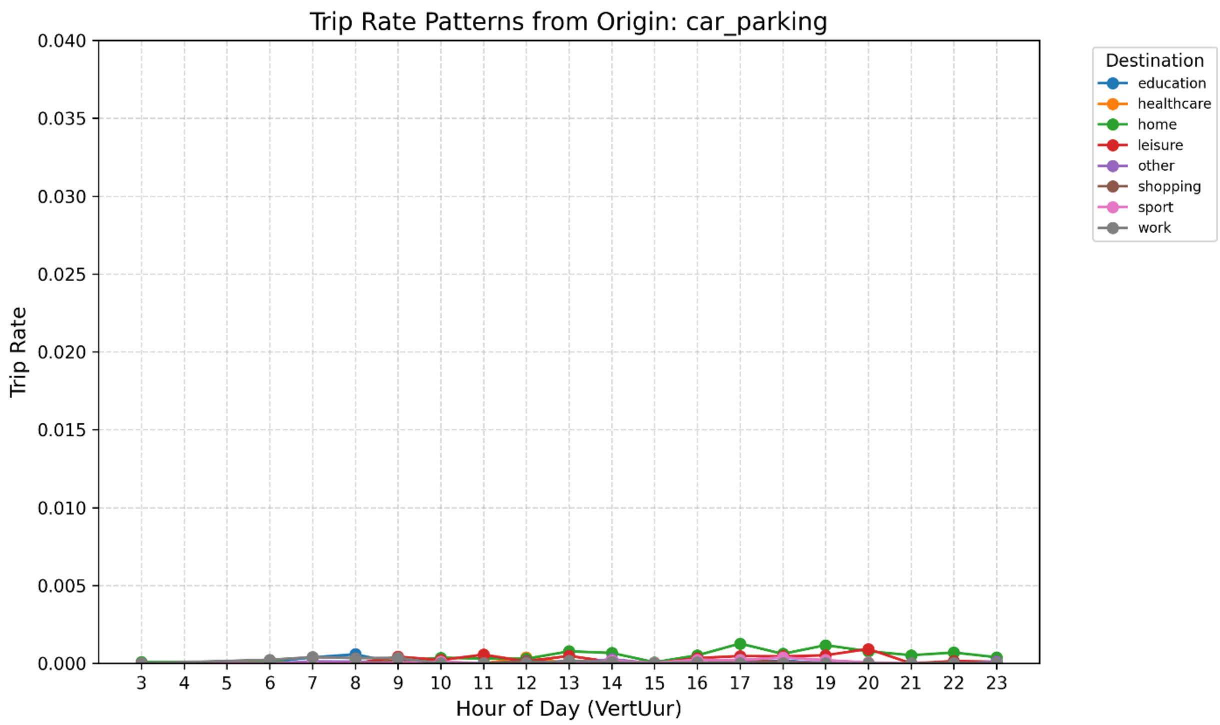
Urban Municipalities:



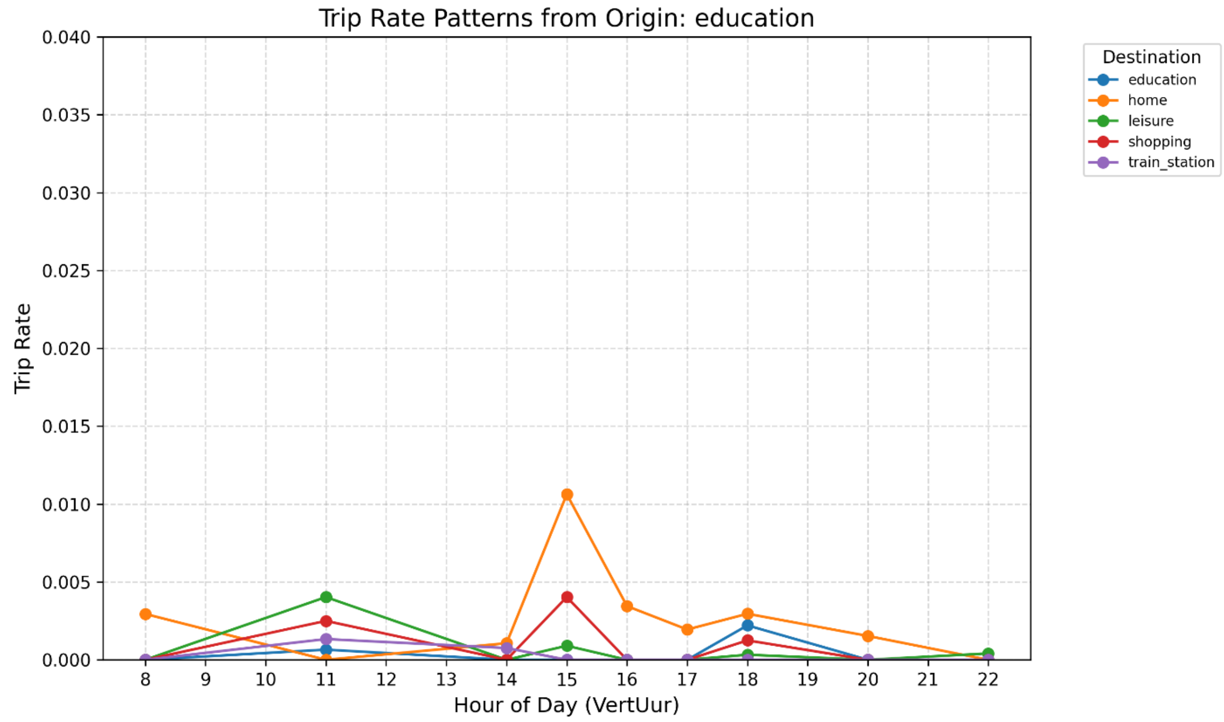
Similar Postcode Areas:



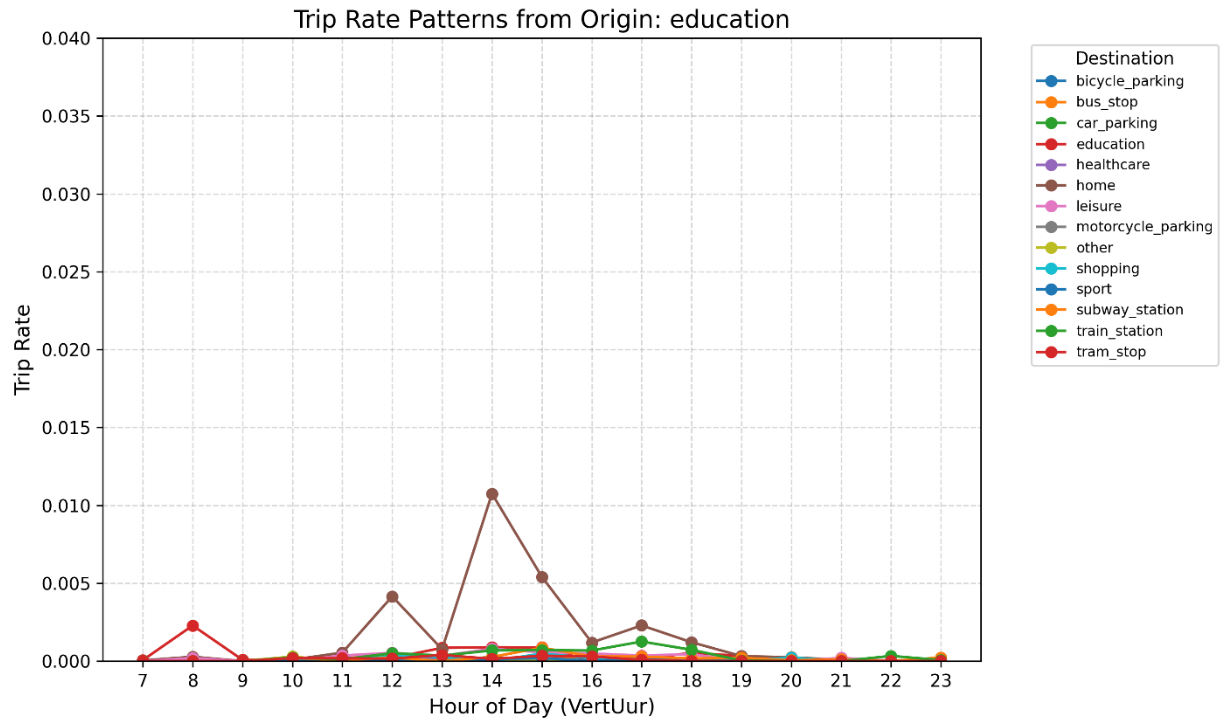
Urban Municipalities:



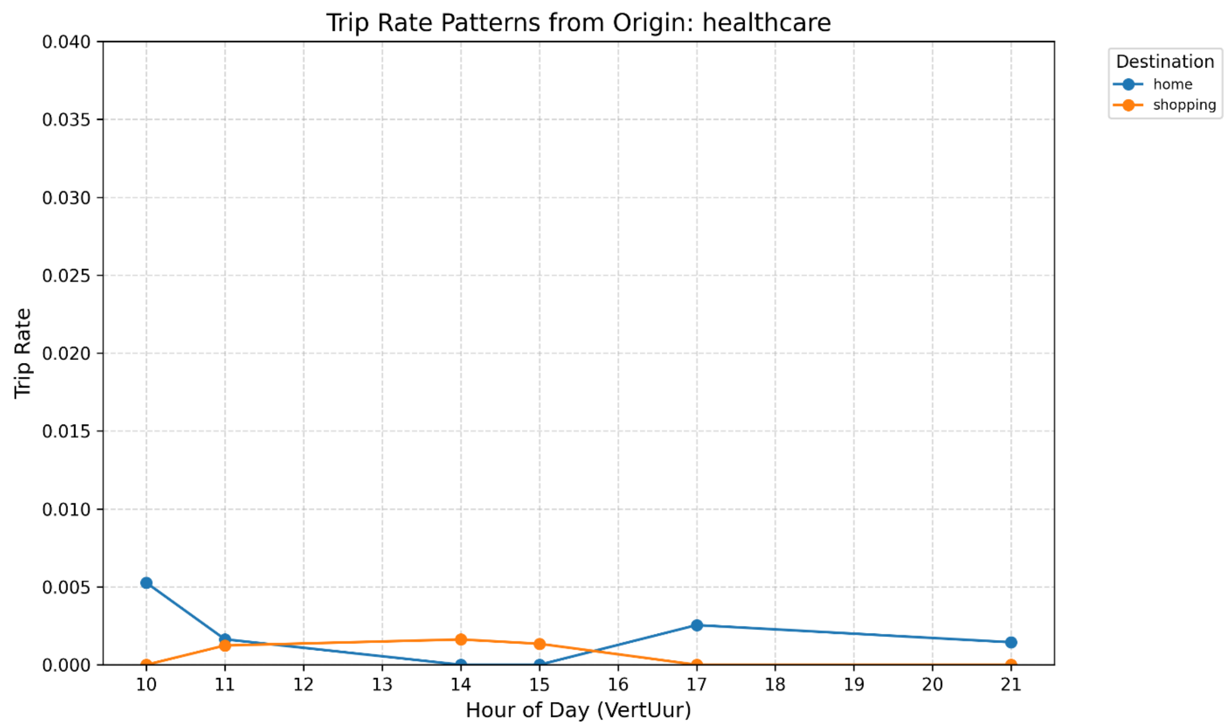
Similar Postcode Areas:



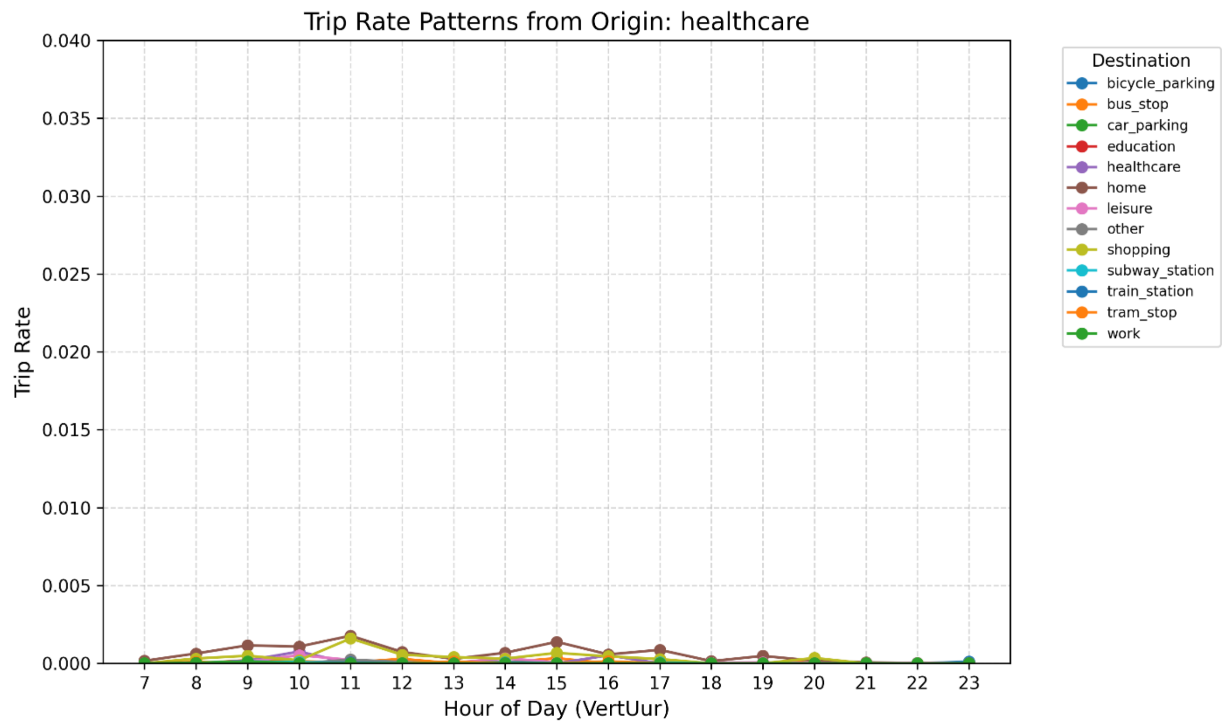
Urban Municipalities:



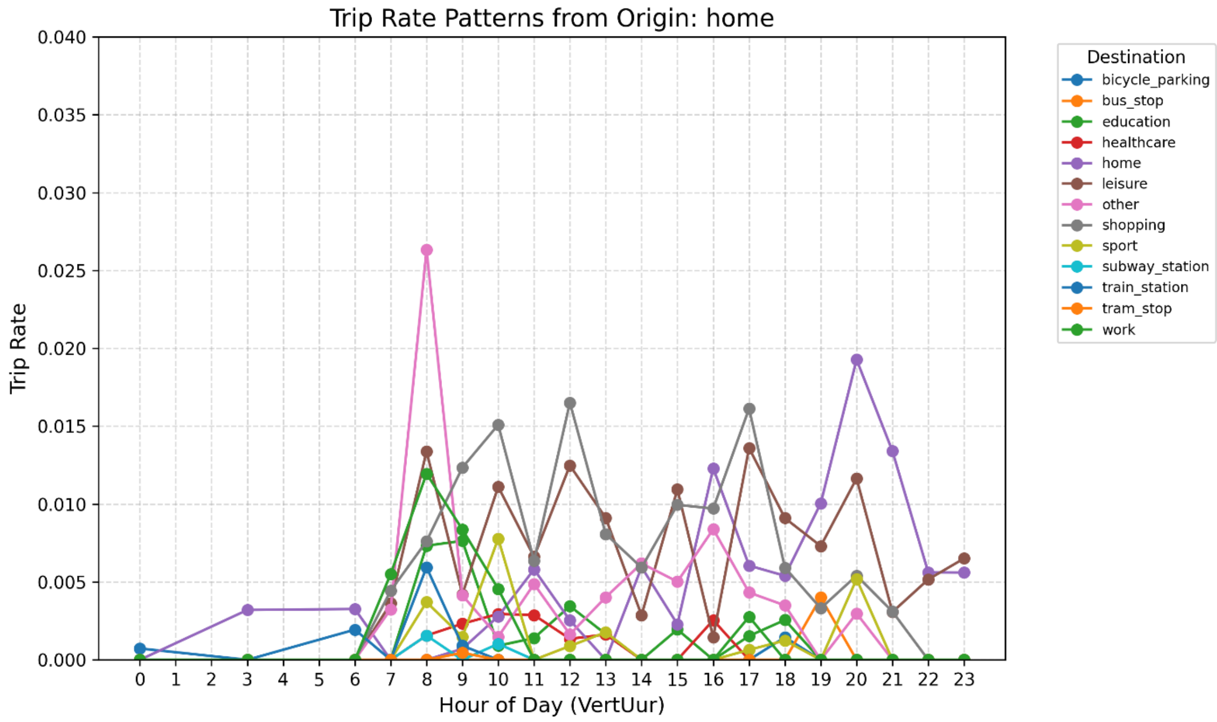
Similar Postcode Areas:



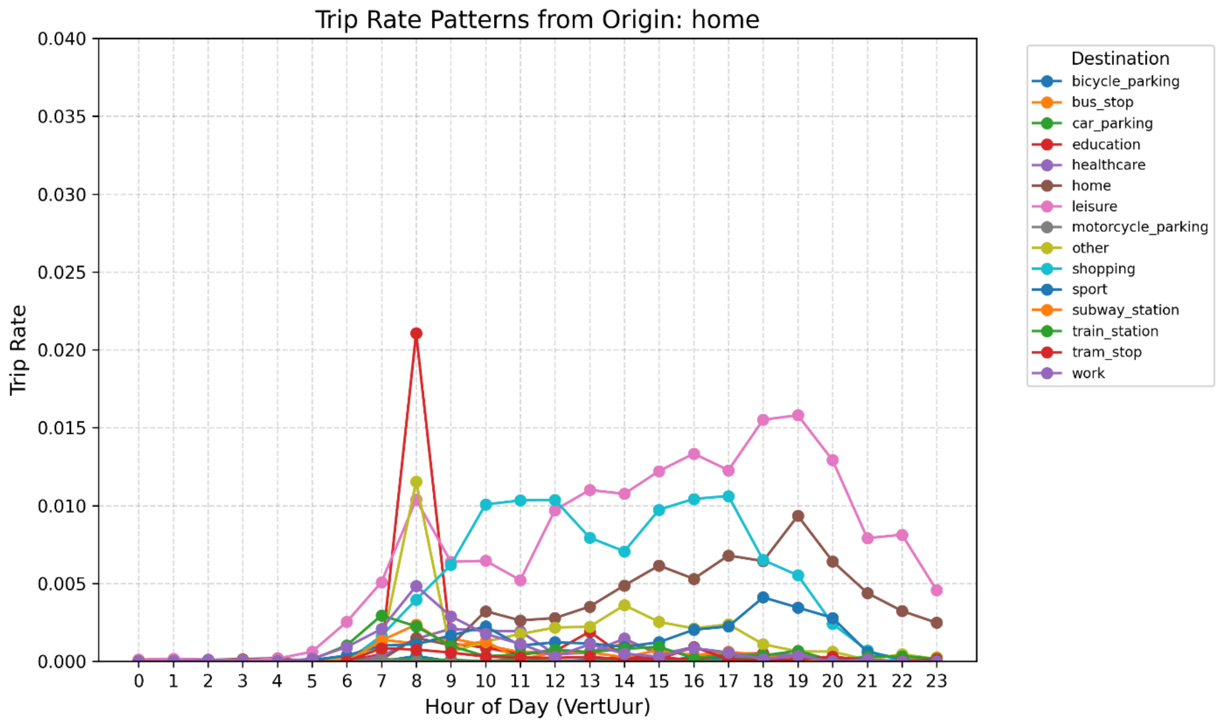
Urban Municipalities:



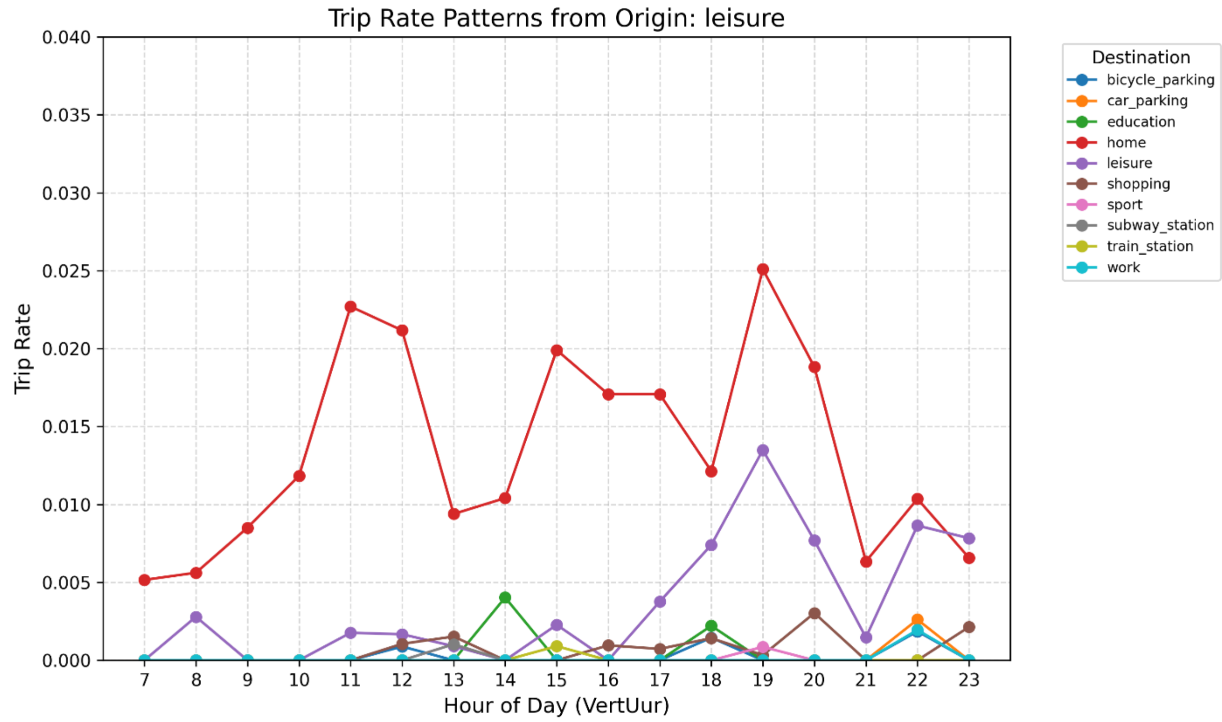
Similar Postcode Areas:



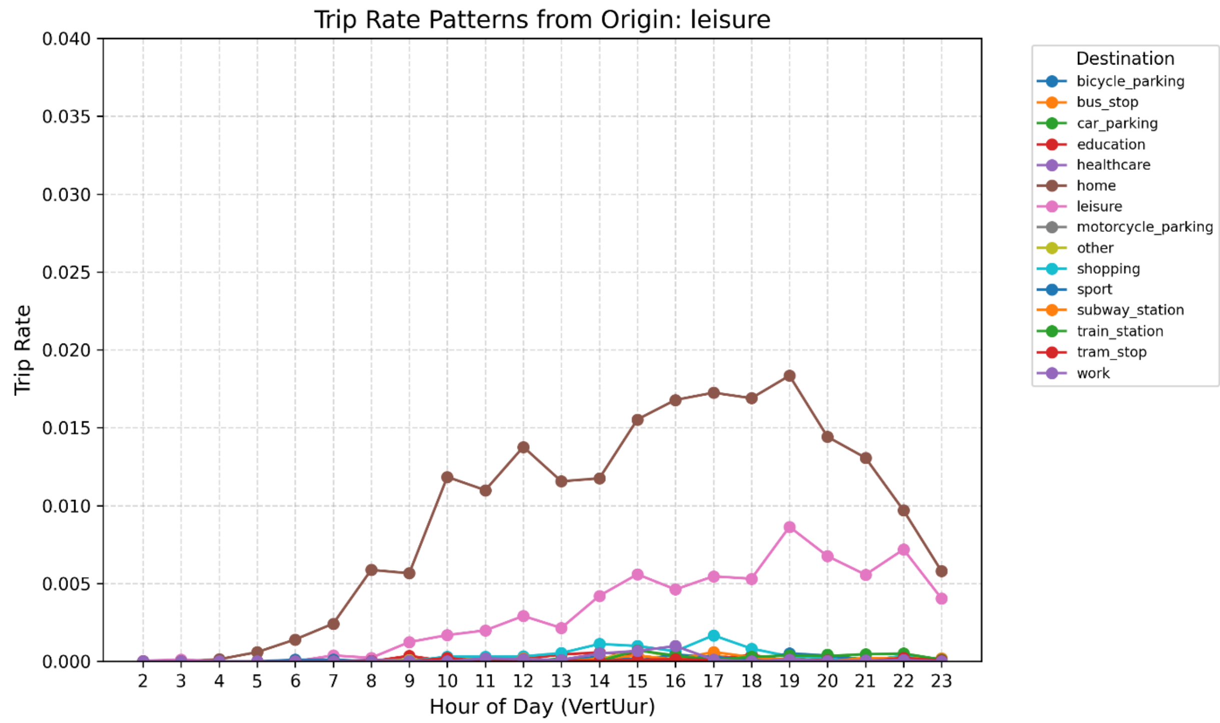
Urban Municipalities:



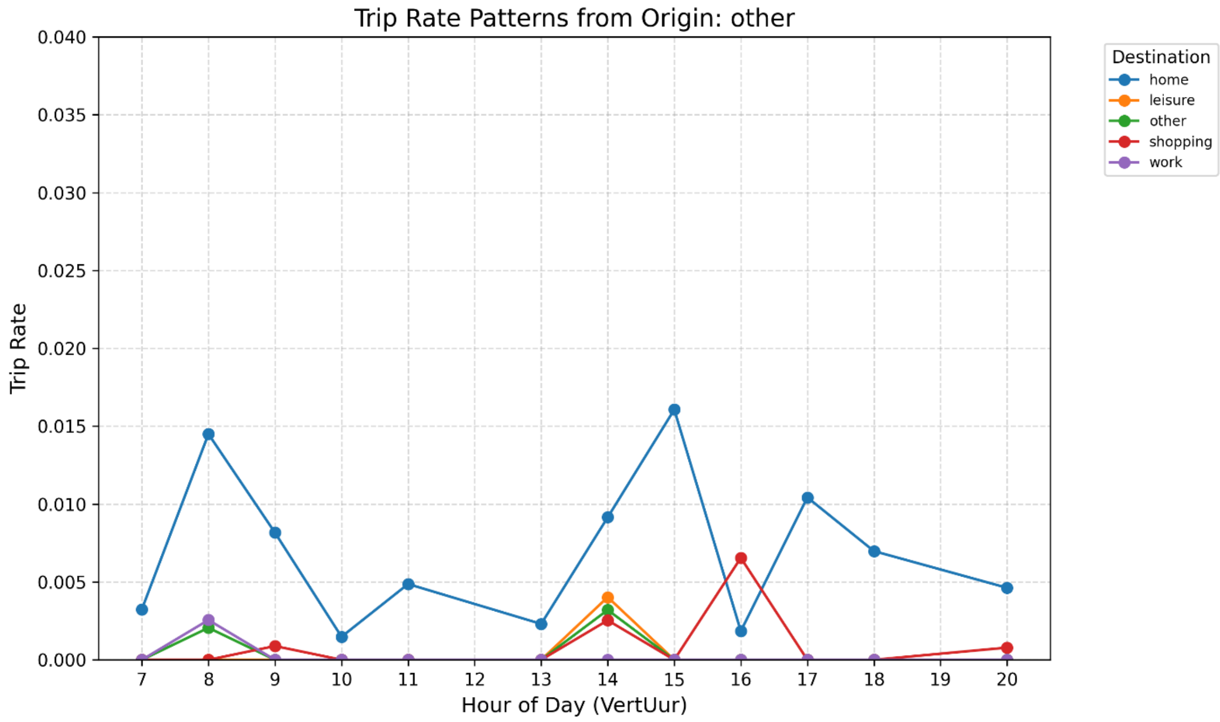
Similar Postcode Areas:



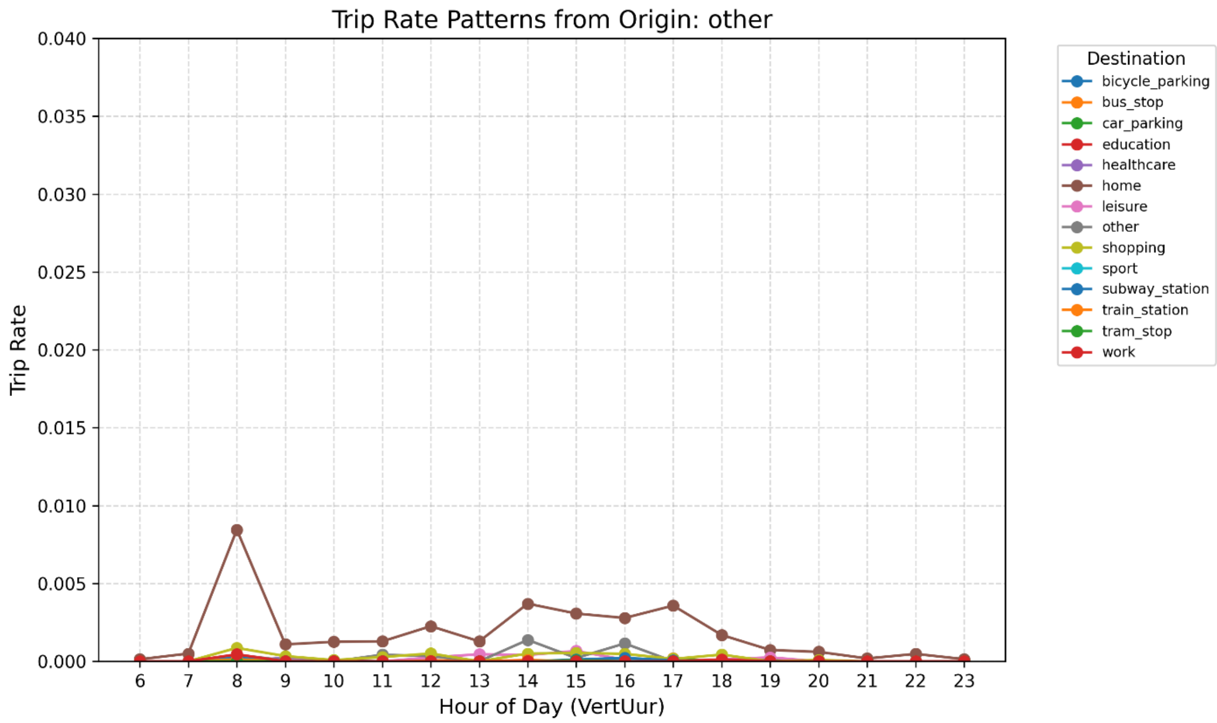
Urban Municipalities:



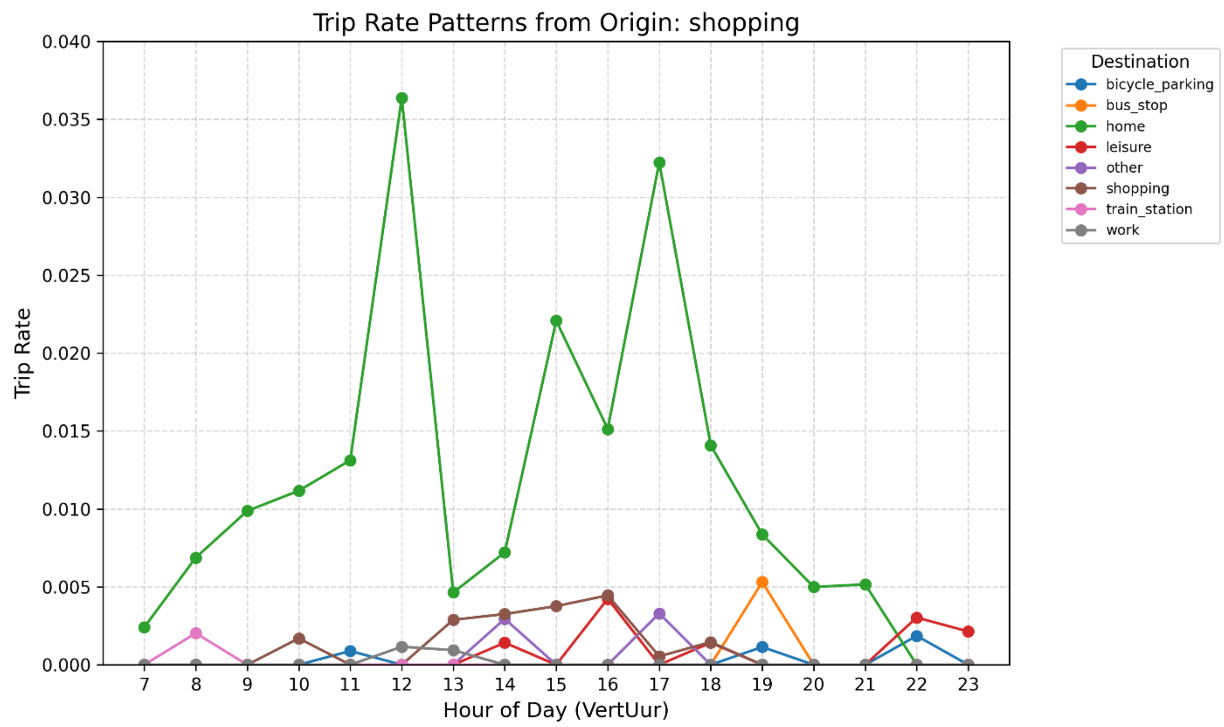
Similar Postcode Areas:



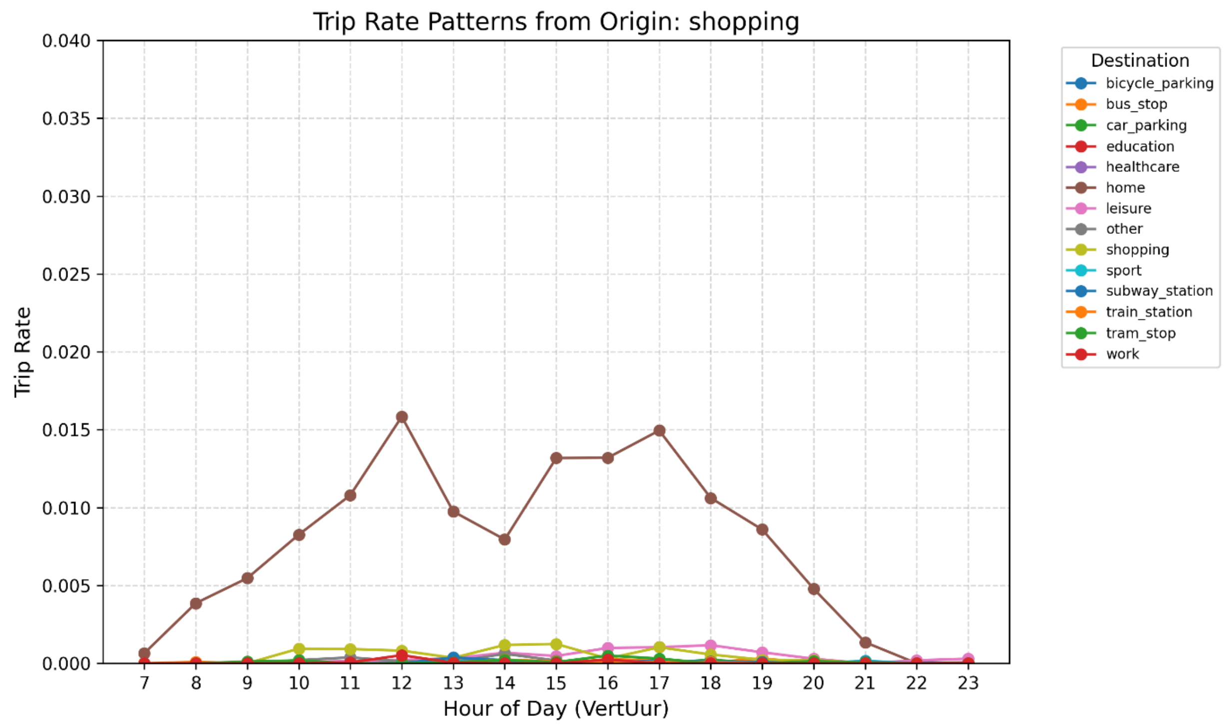
Urban Municipalities:



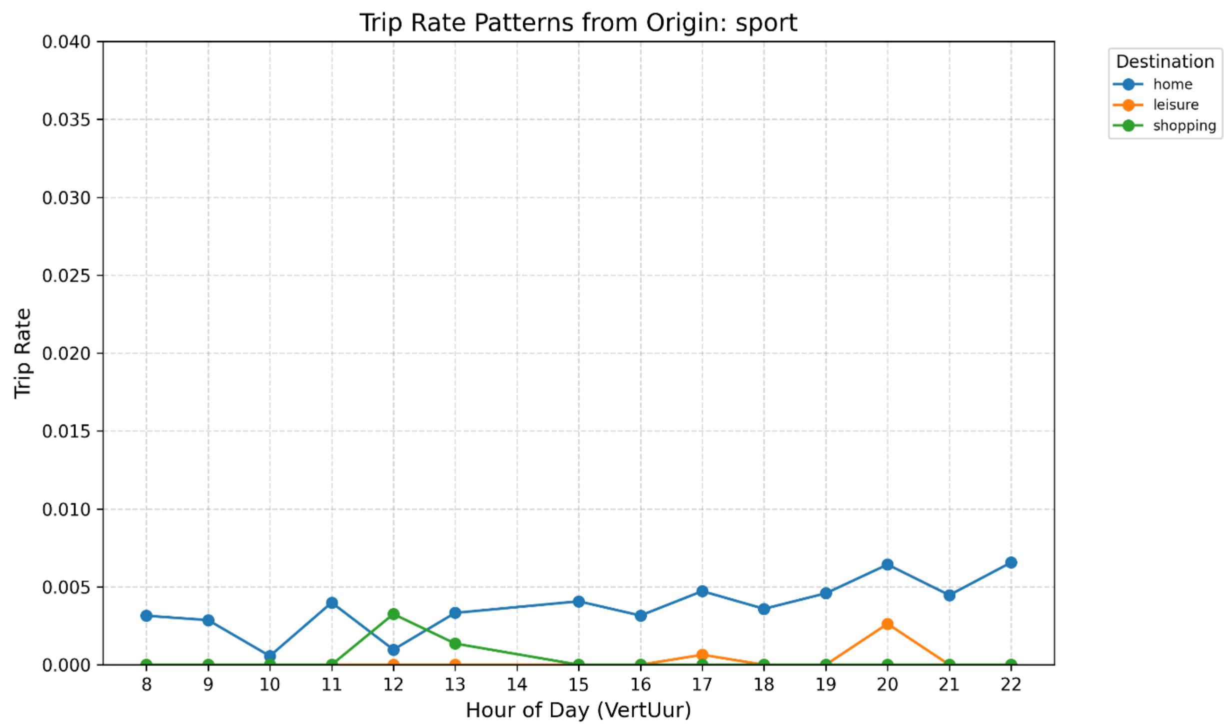
Similar Postcode Areas:



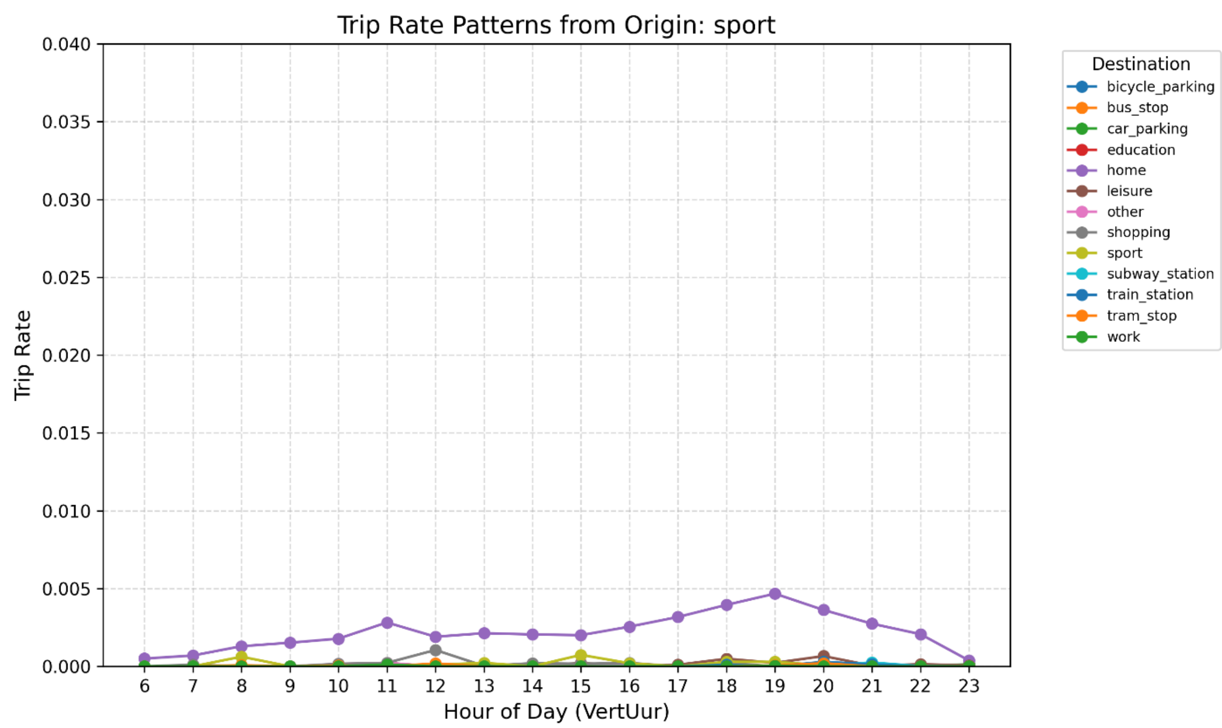
Urban Municipalities:



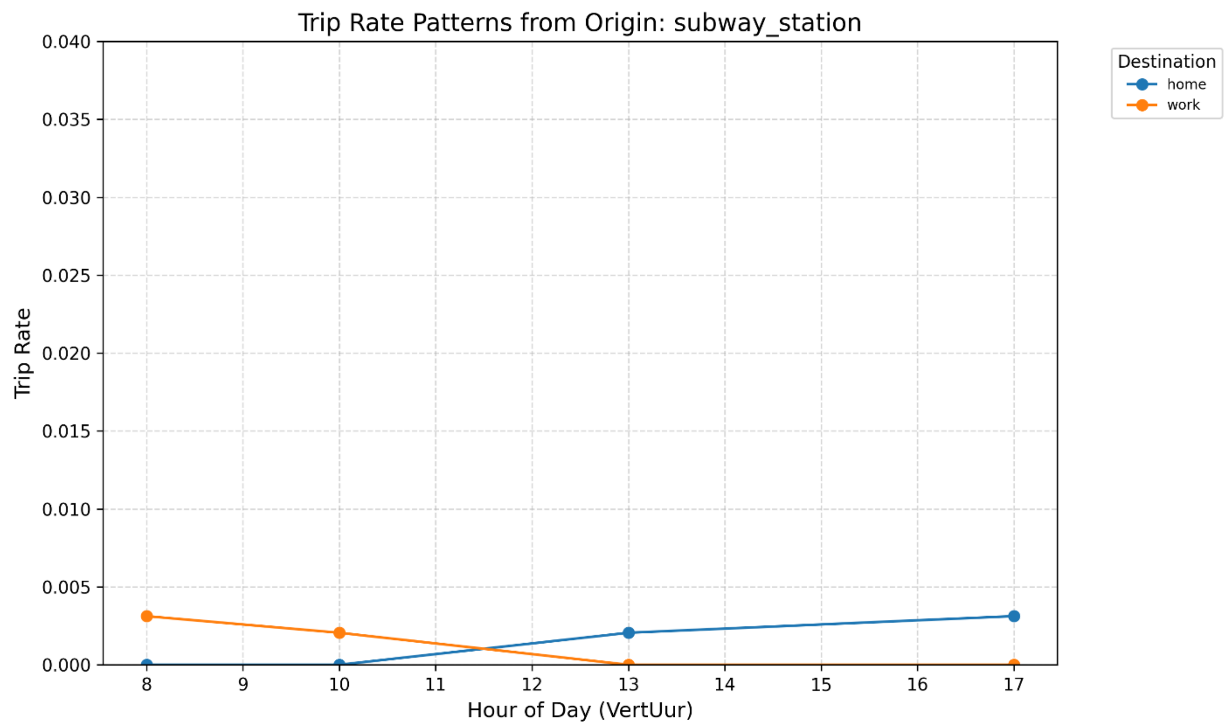
Similar Postcode Areas:



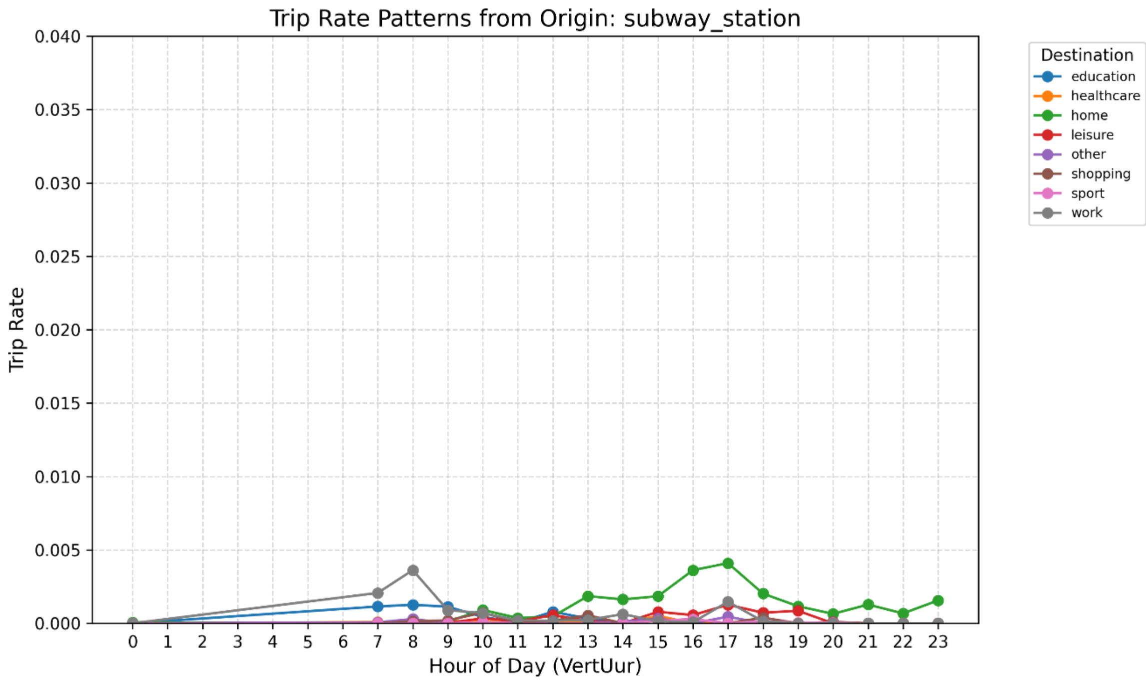
Urban Municipalities:



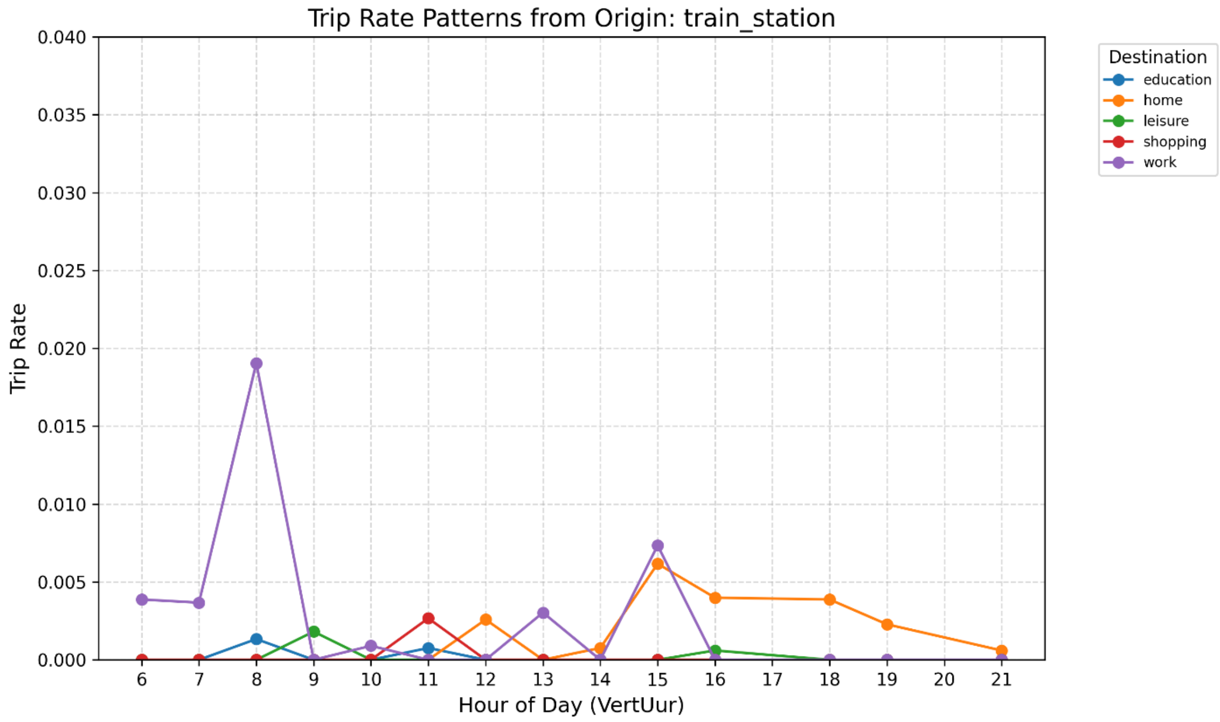
Similar Postcode Areas:



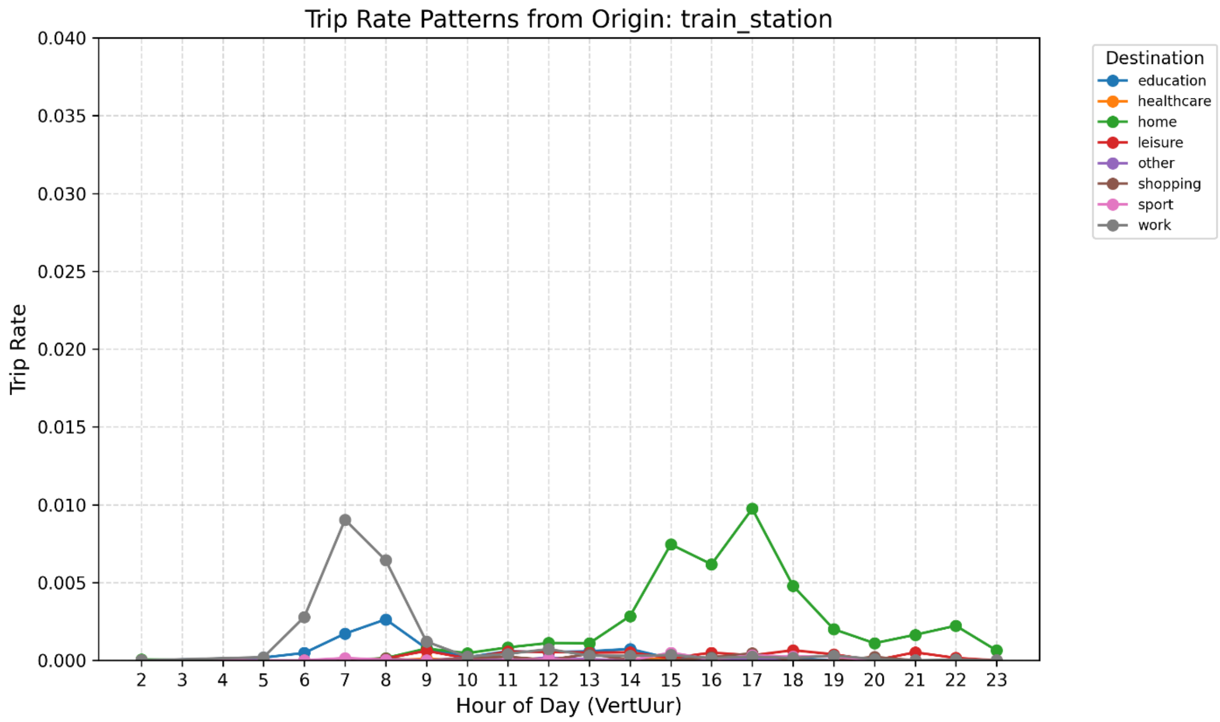
Urban Municipalities:



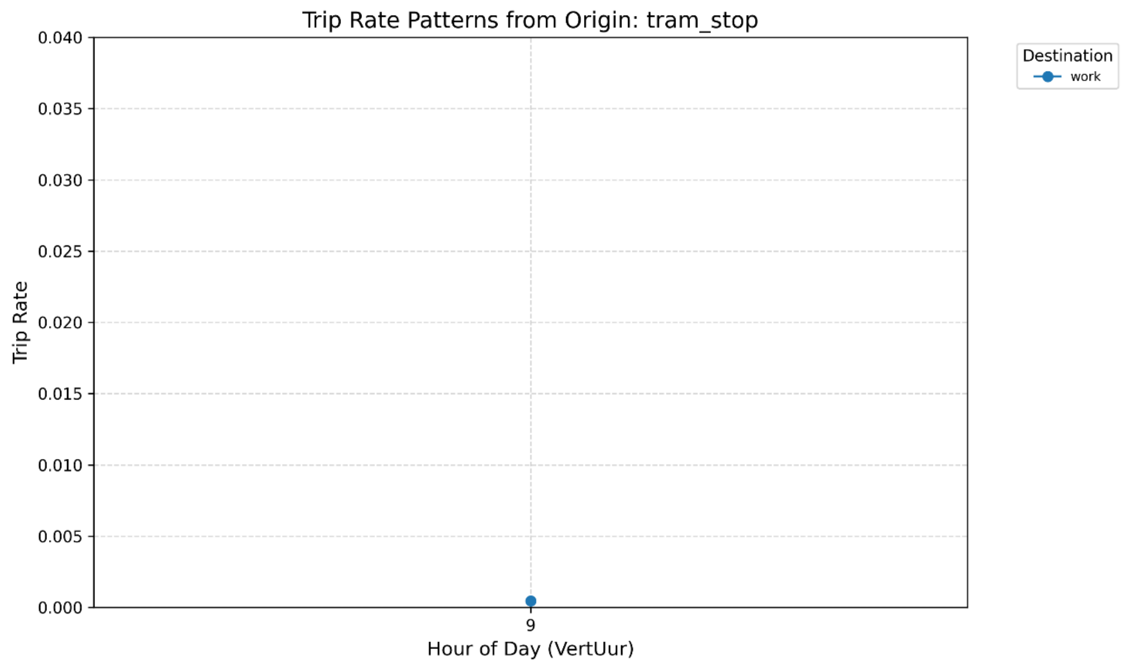
Similar Postcode Areas:



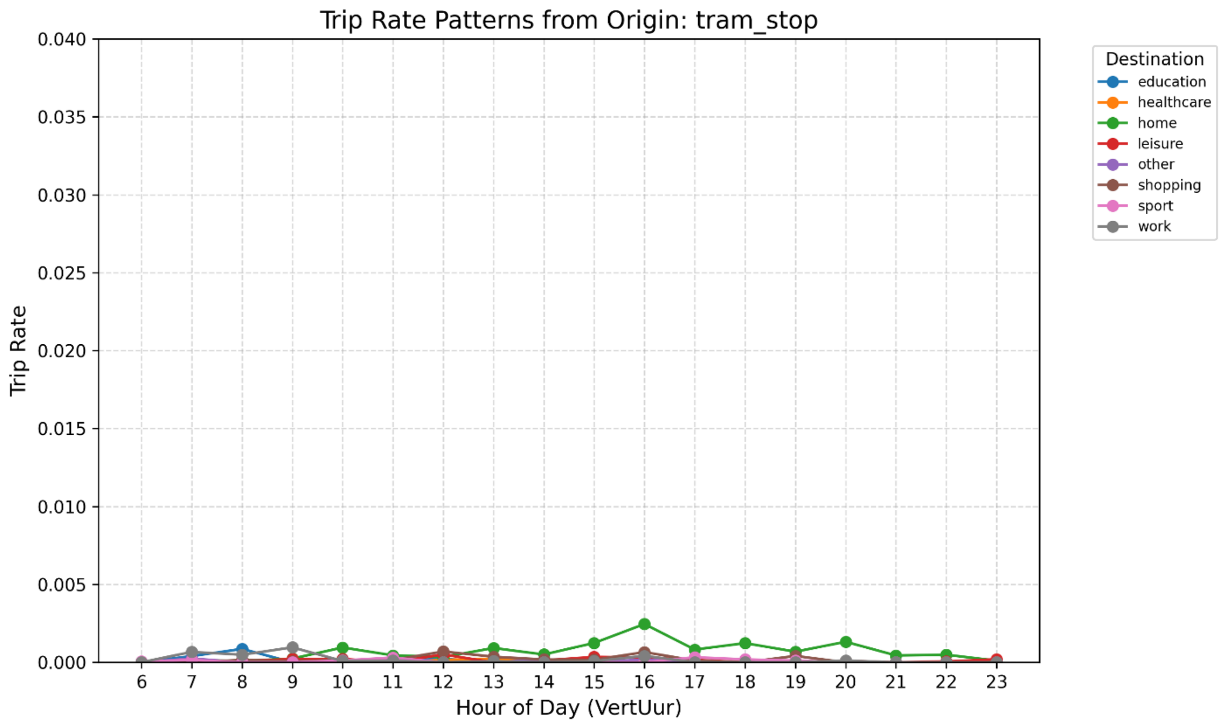
Urban Municipalities:



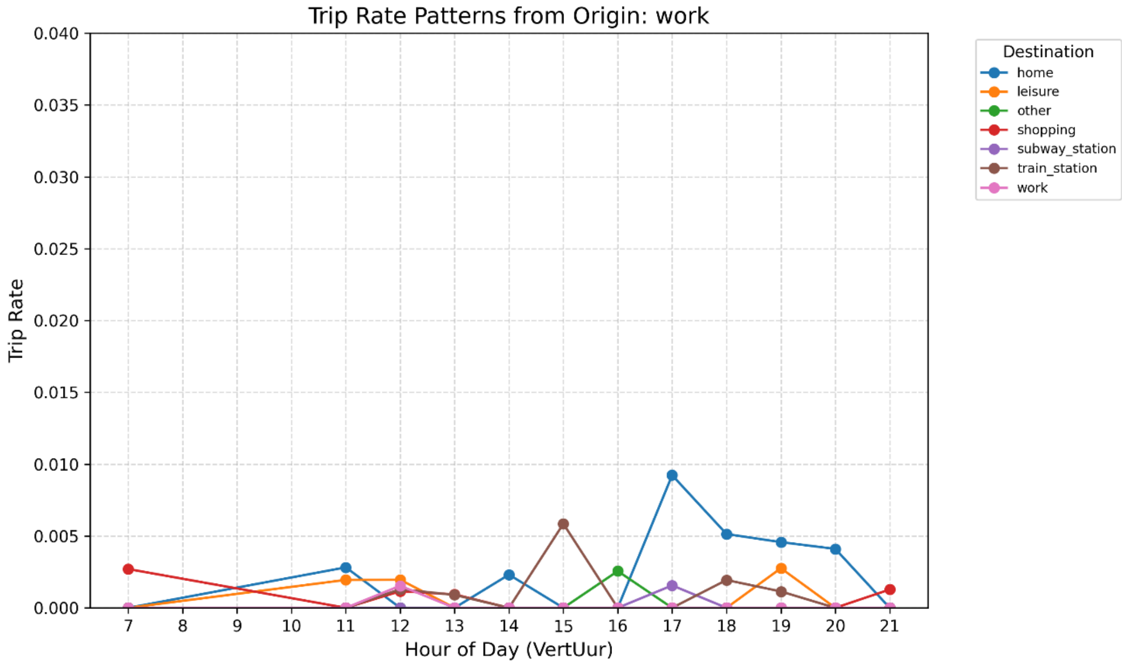
Similar Postcode Areas:



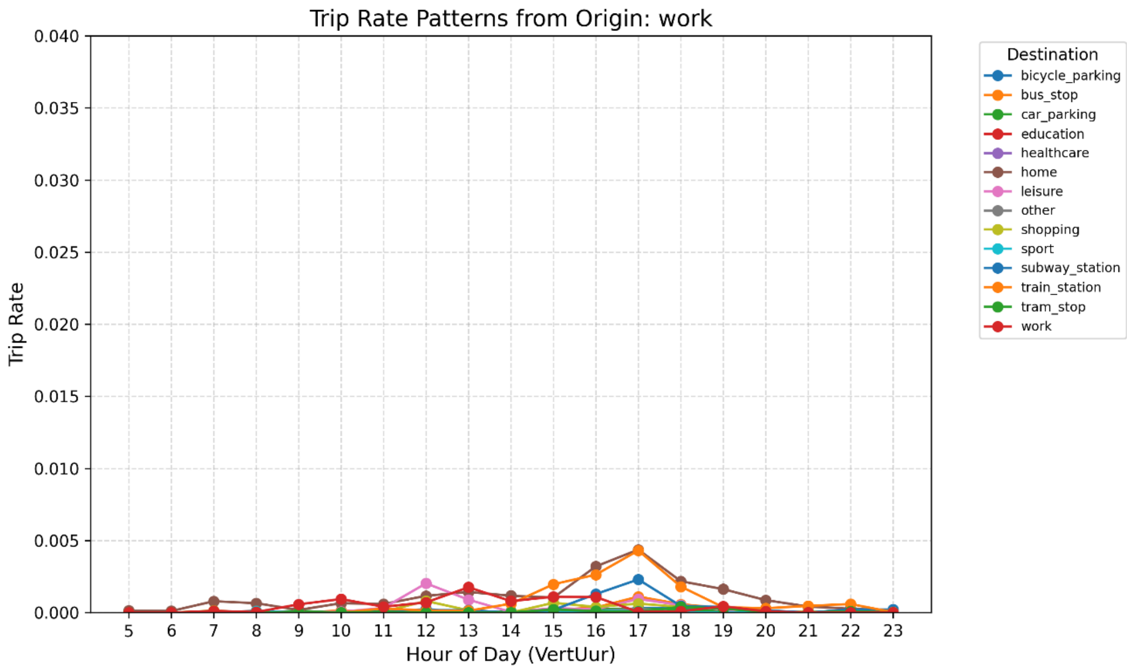
Urban Municipalities:



Similar Postcode Areas:

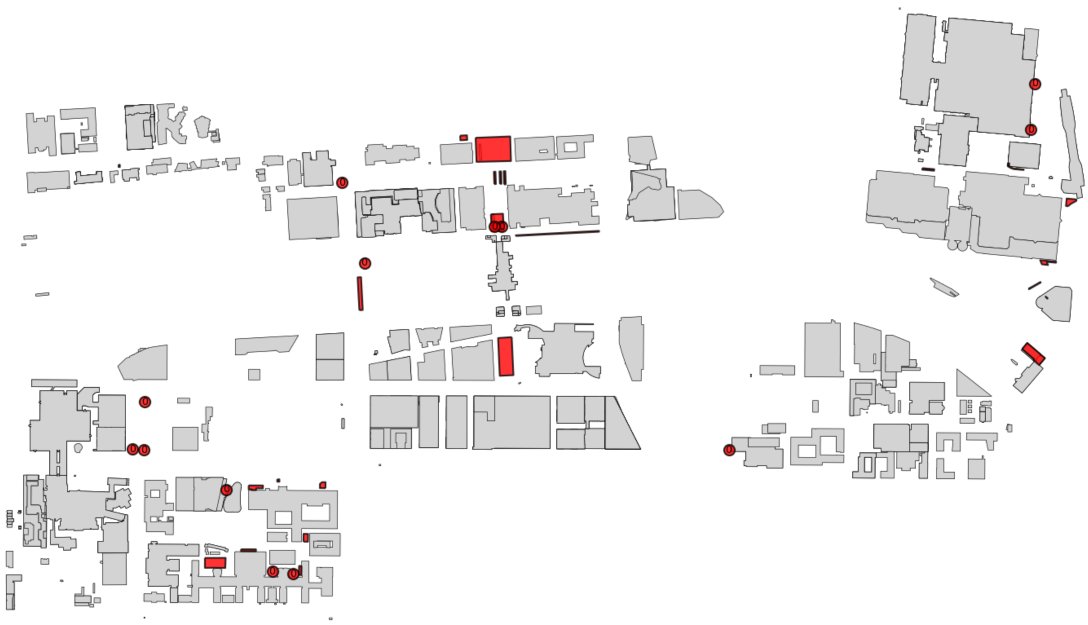


Urban Municipalities:

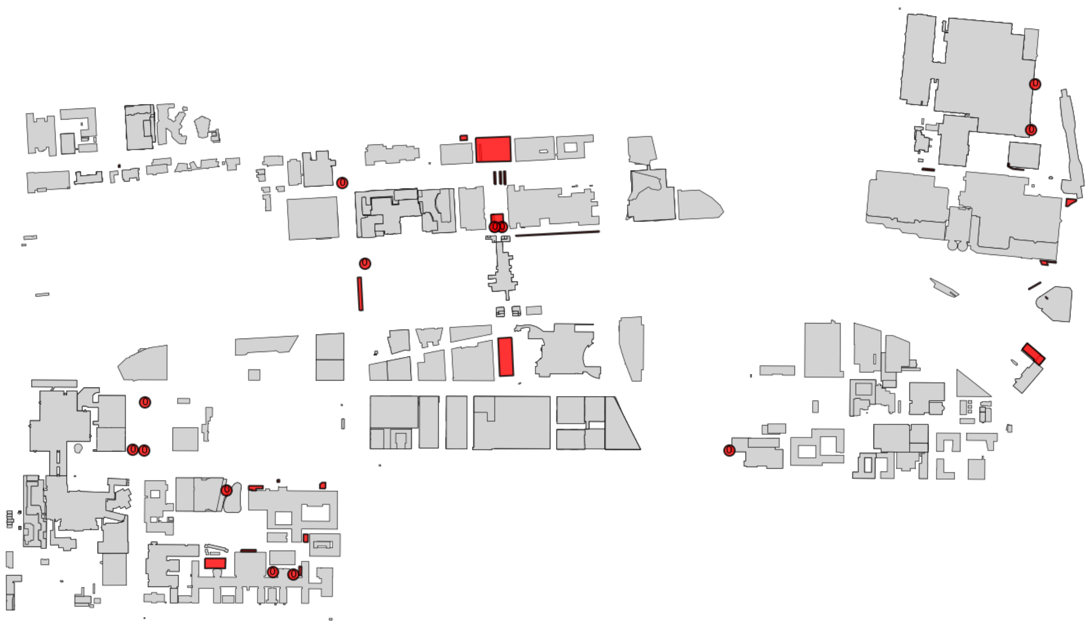


9.3 Appendix 3: Generated Trips

Trips in Zuidas — Origin: bicycle_parking | Hour: 8



Trips in Zuidas — Origin: bicycle_parking | Hour: 16

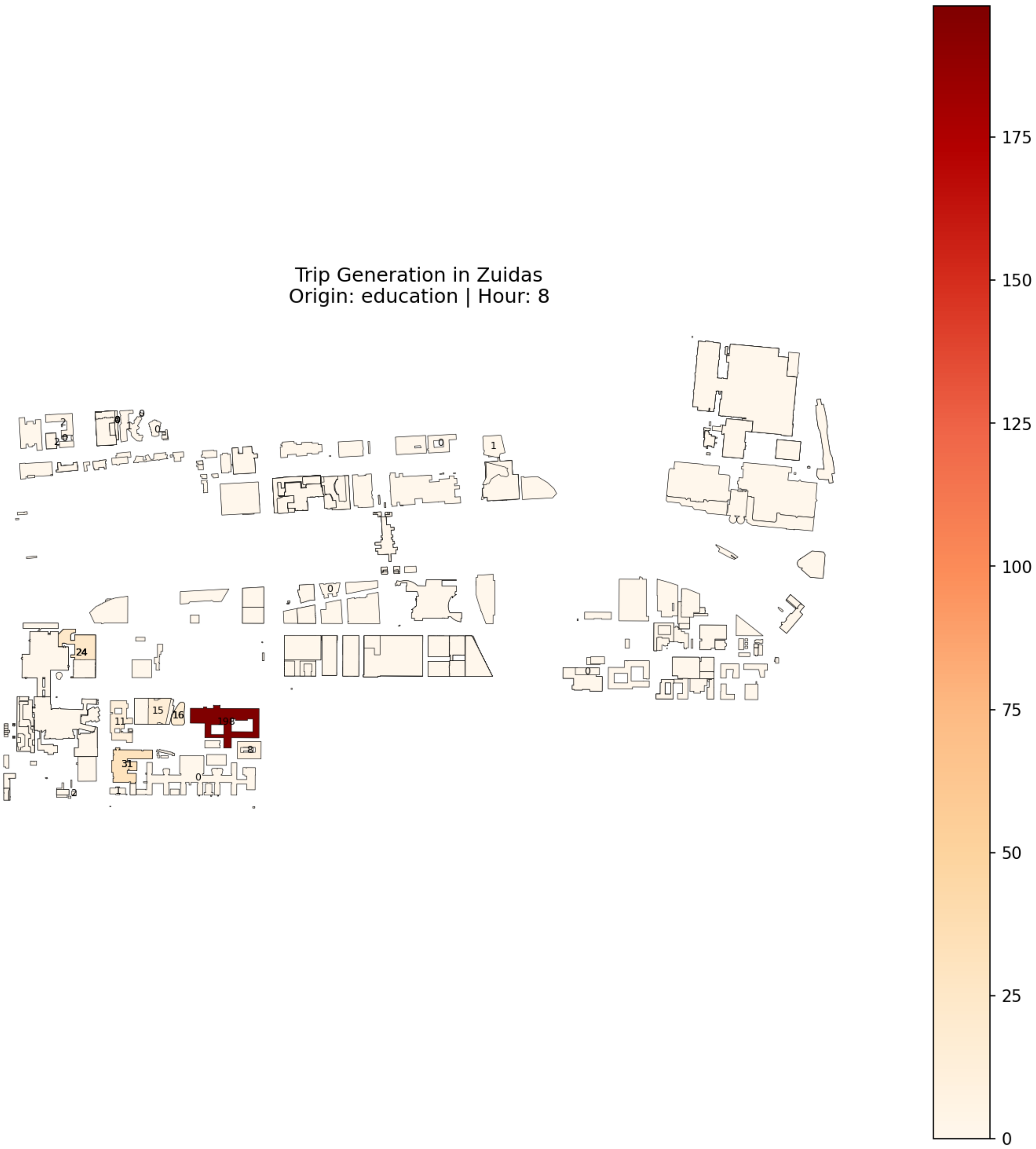


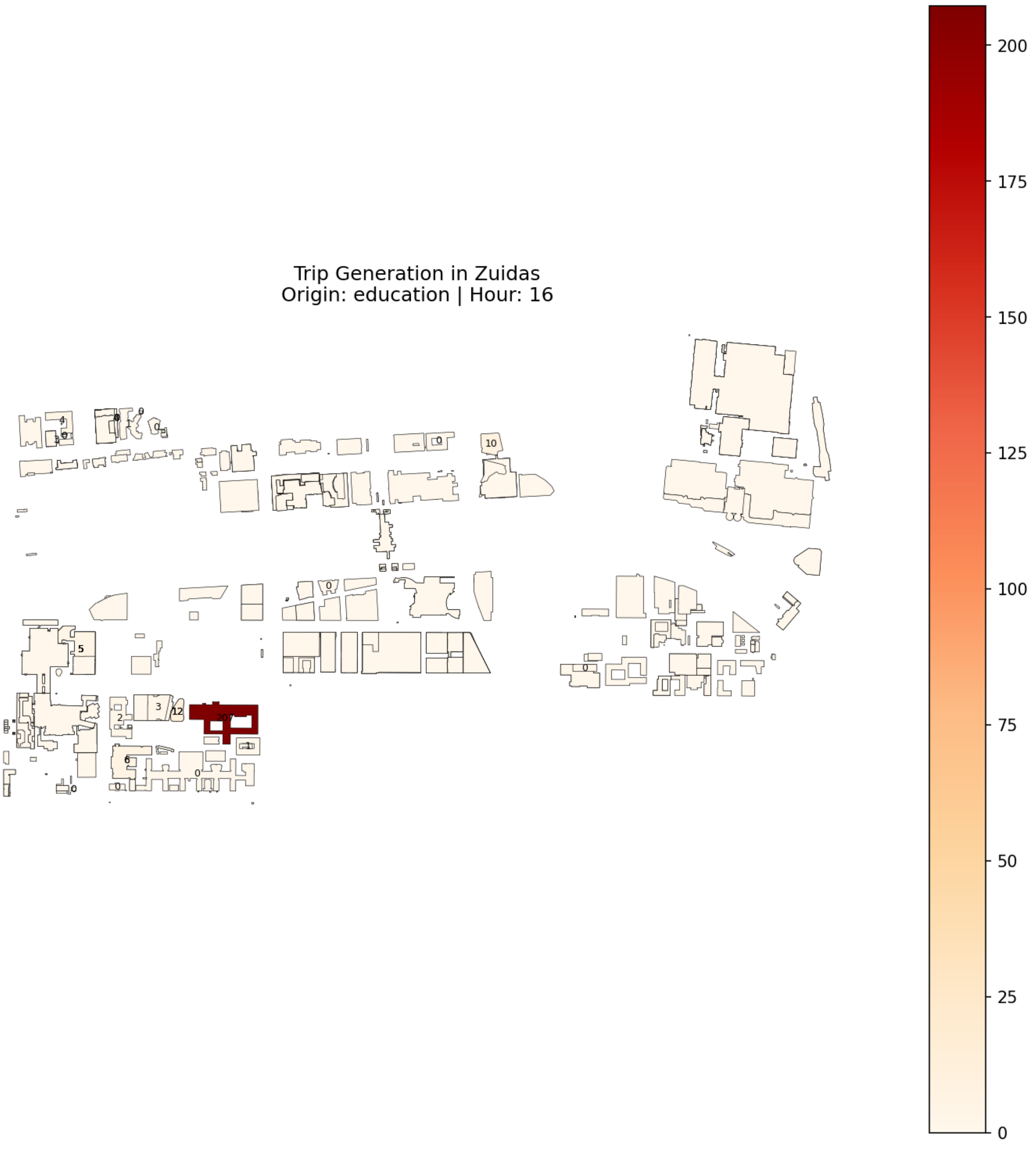
Trips in Zuidas — Origin: bus_stop | Hour: 8

Trips in Zuidas — Origin: bus_stop | Hour: 16

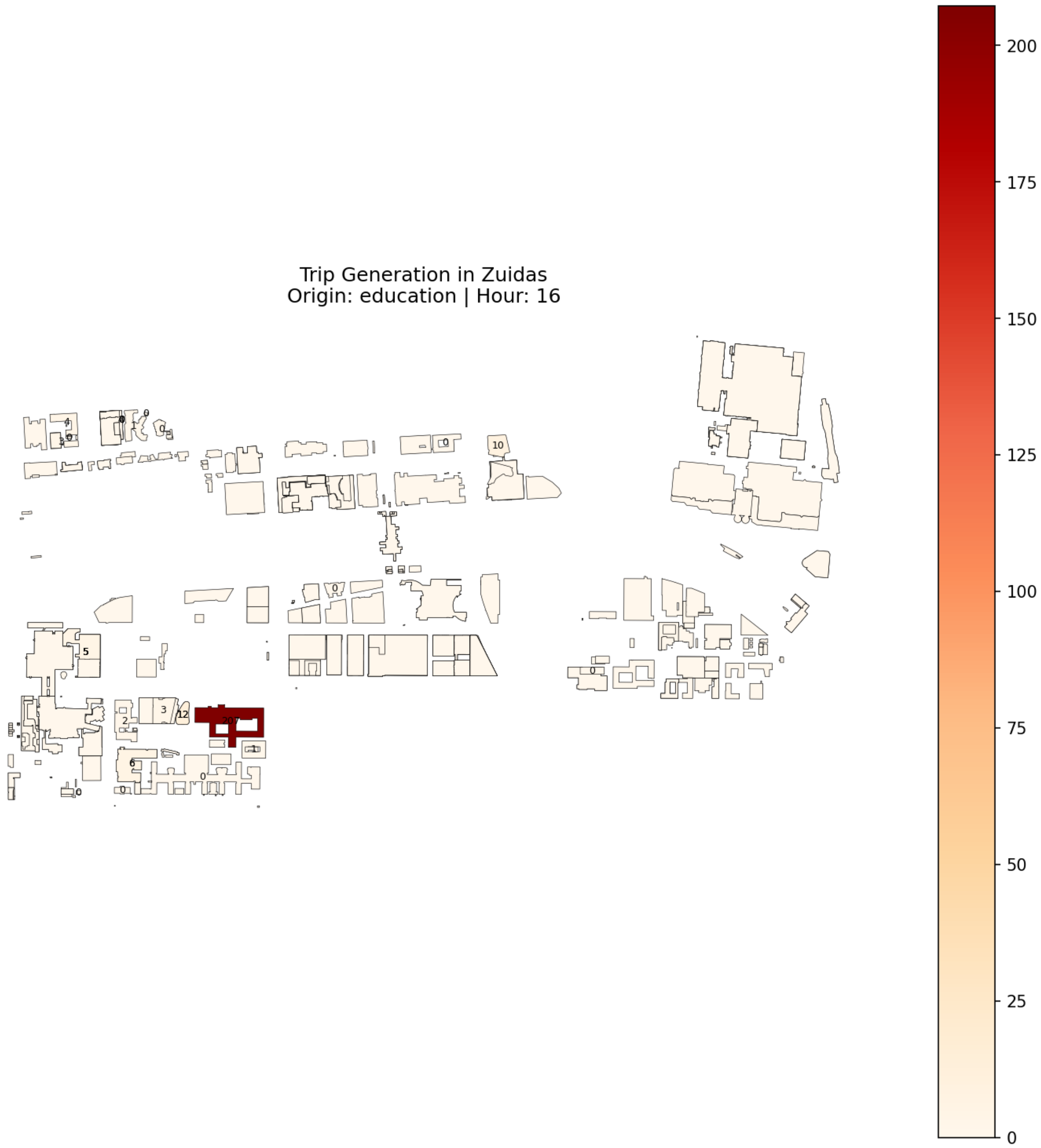
Trips in Zuidas — Origin: car_parking | Hour: 8

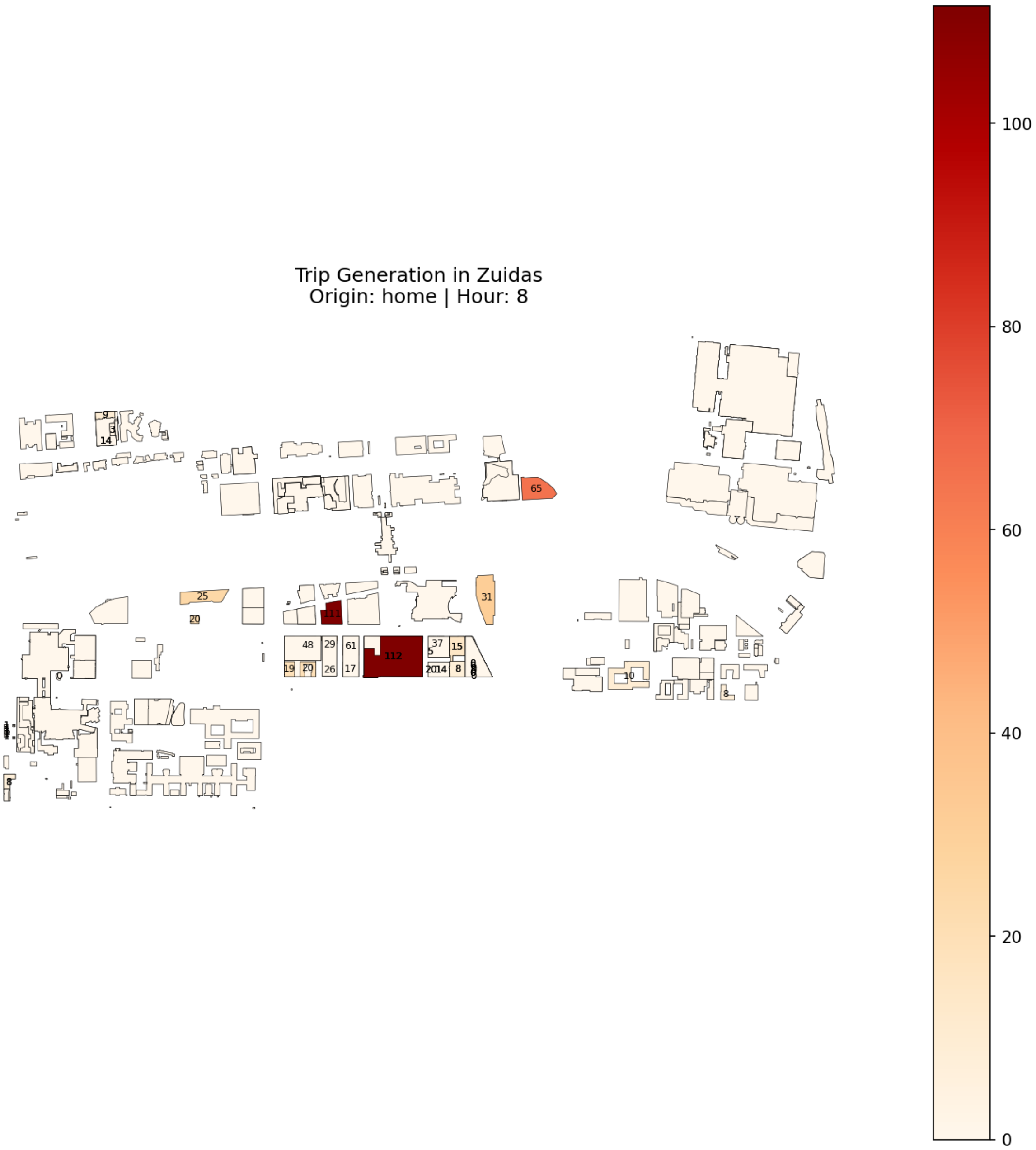
Trips in Zuidas — Origin: car_parking | Hour: 16


Trip Generation in Zuidas
Origin: education | Hour: 8
175
150
125
100
75
50
25
0

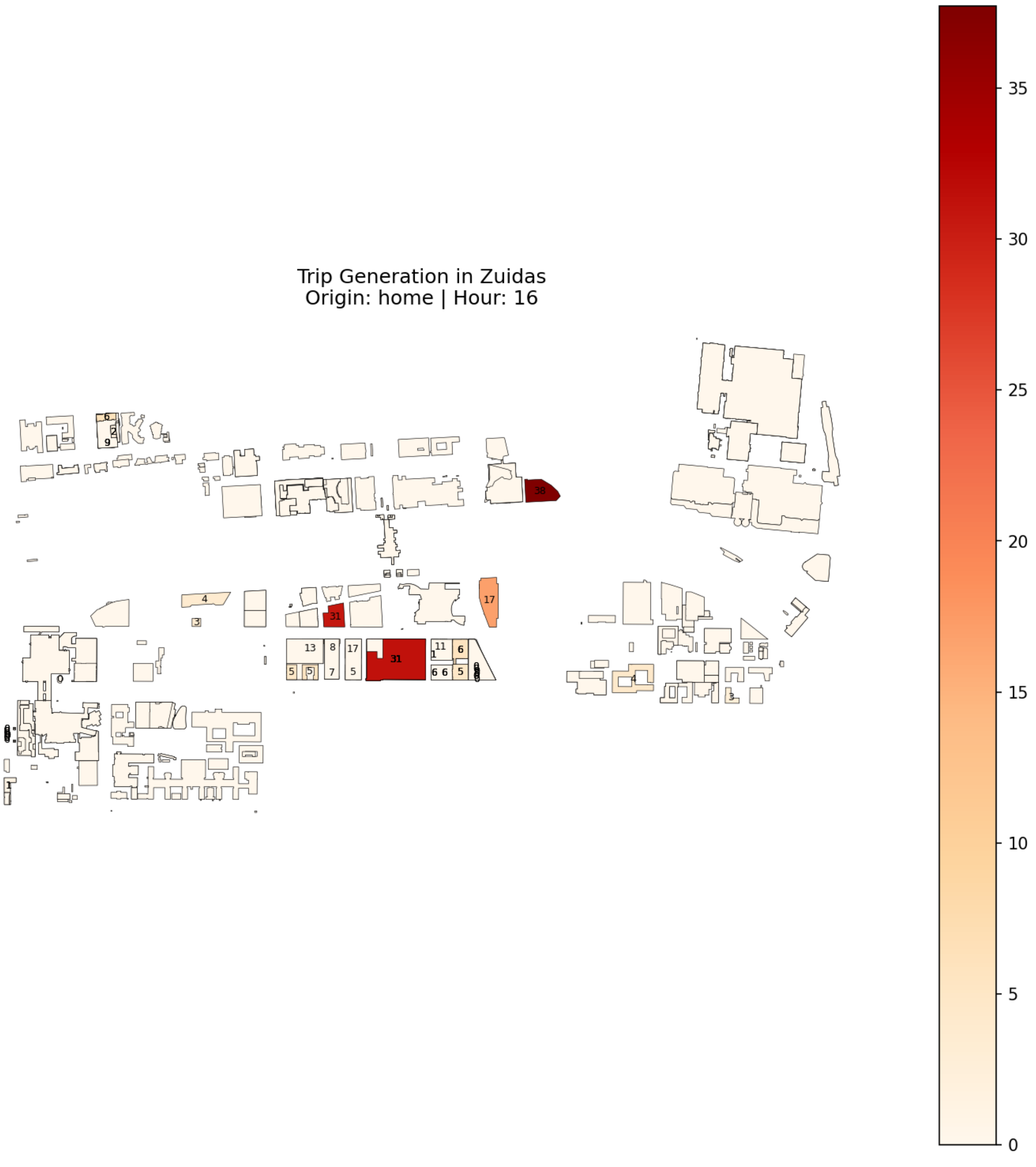

Trip Generation in Zuidas
Origin: education | Hour: 16
200
175
150
125
100
75
50
25
0
207
12
10
6
5
4
3
2
1
0

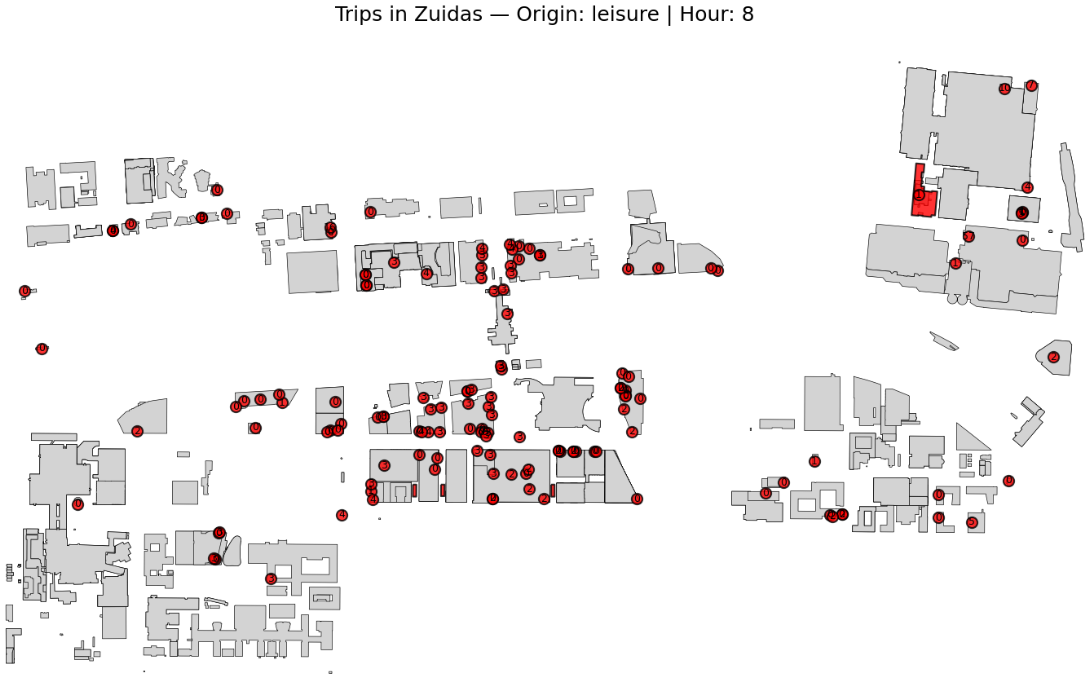
Trip Generation in Zuidas
Origin: education | Hour: 16

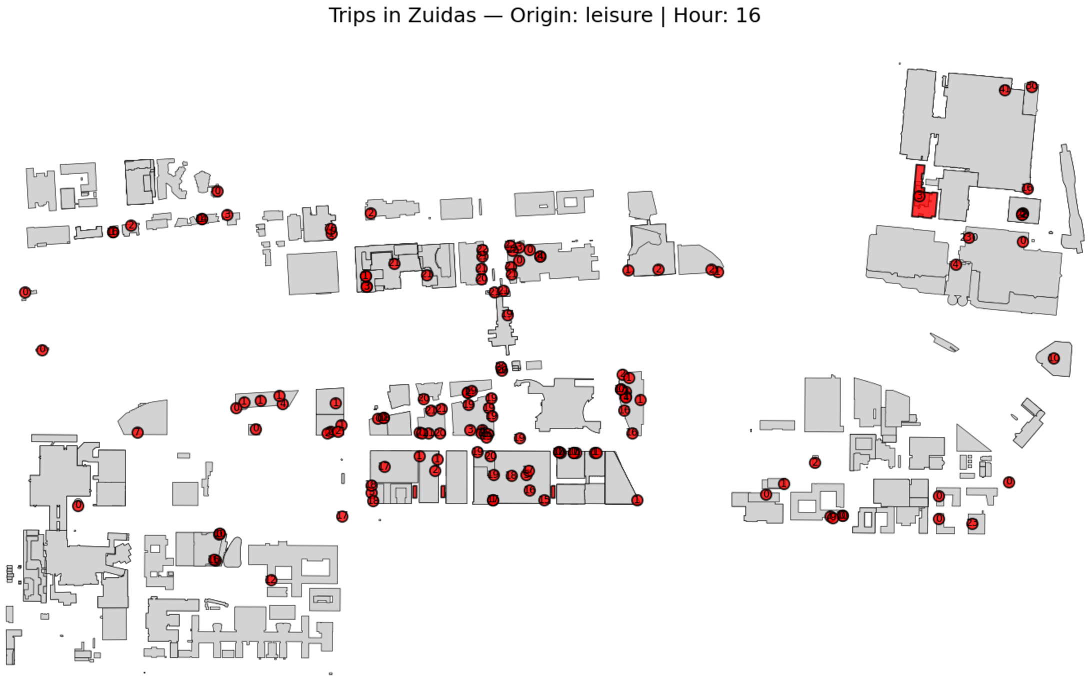


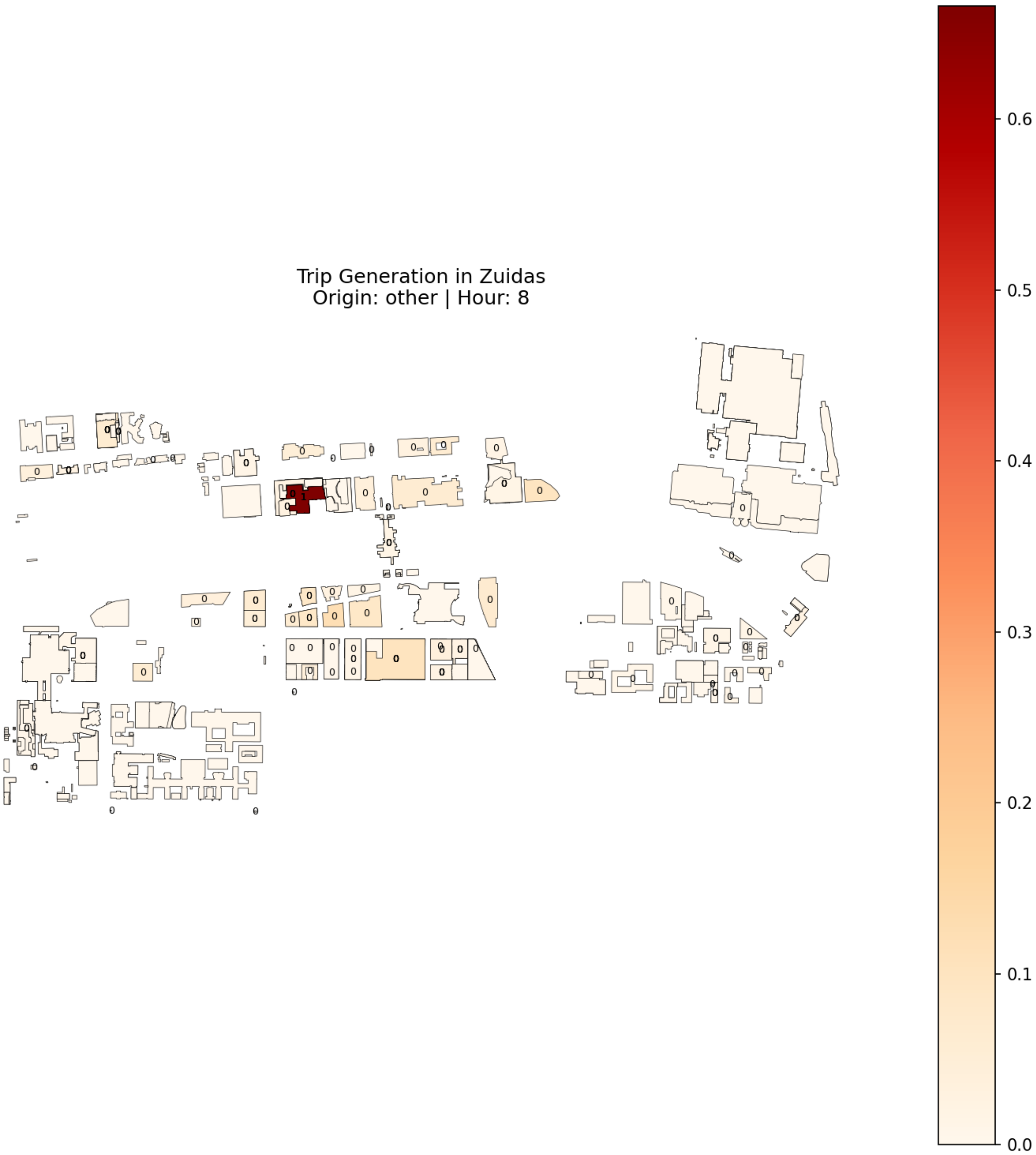

Trip Generation in Zuidas
Origin: home | Hour: 8
65
25
20
31
48
29
61
112
37
15
19
20
26
17
14
8
10
0
20
40
60
80
100

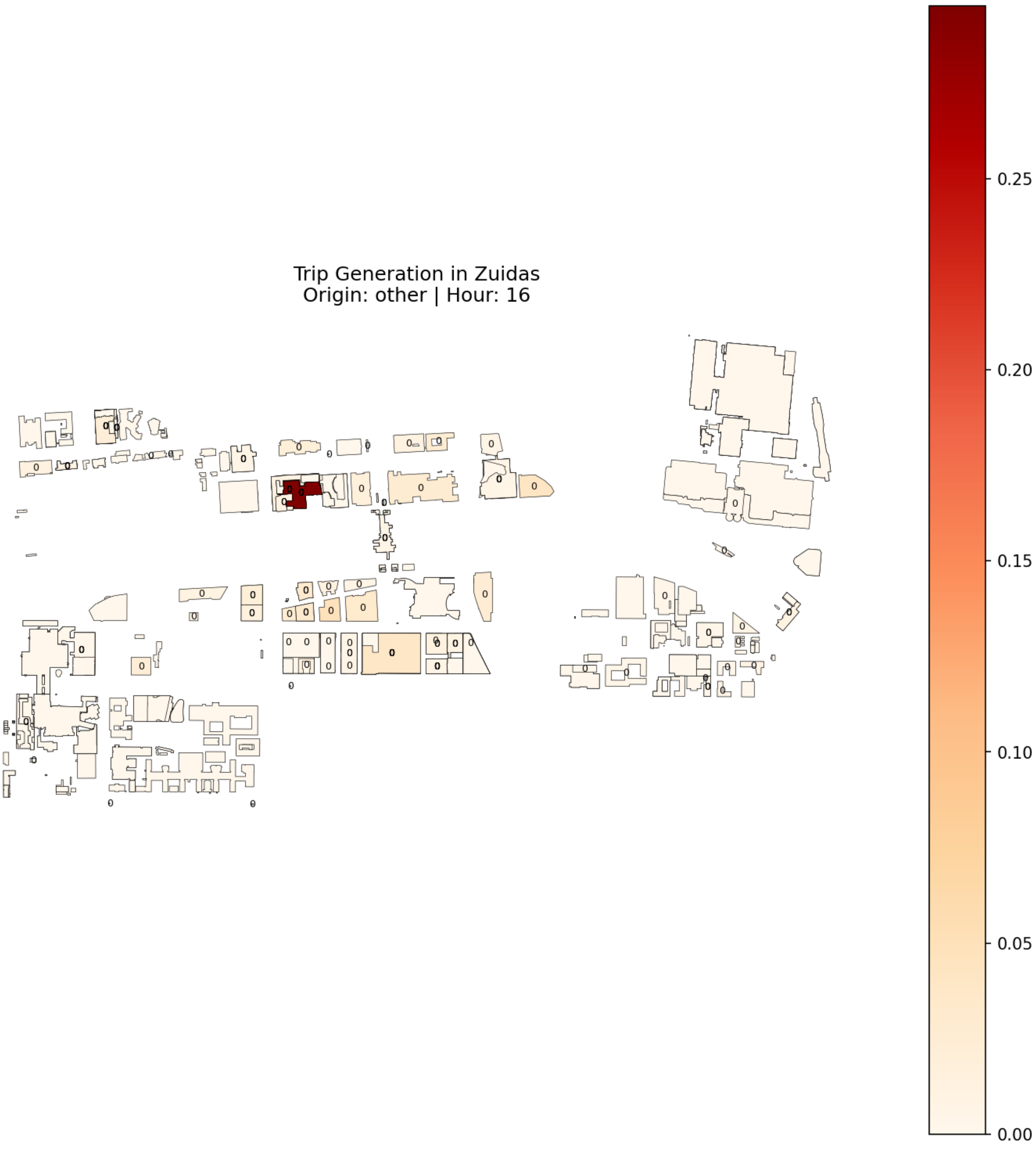
Trip Generation in Zuidas
Origin: home | Hour: 16

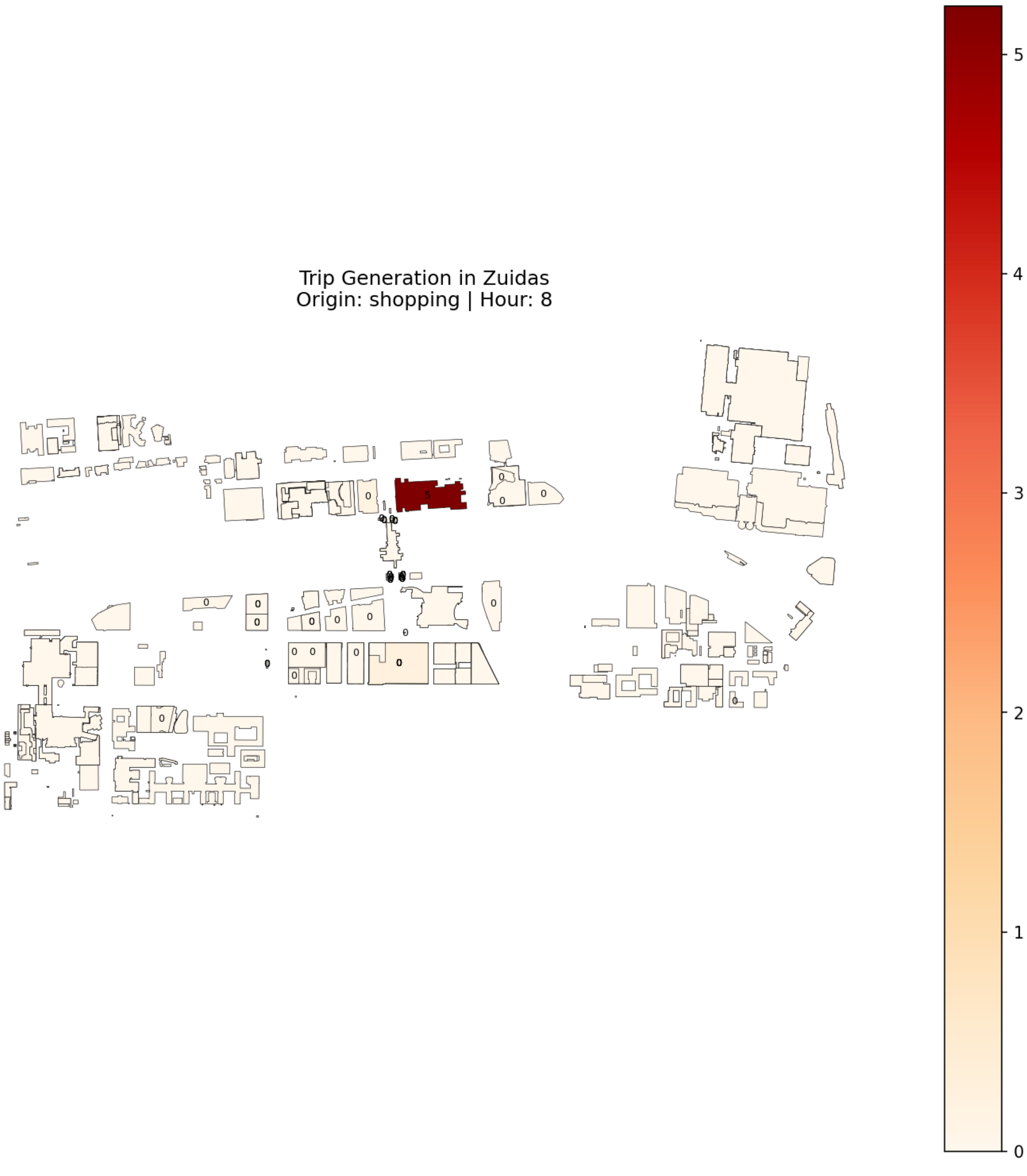



Trips in Zuidas — Origin: leisure | Hour: 8

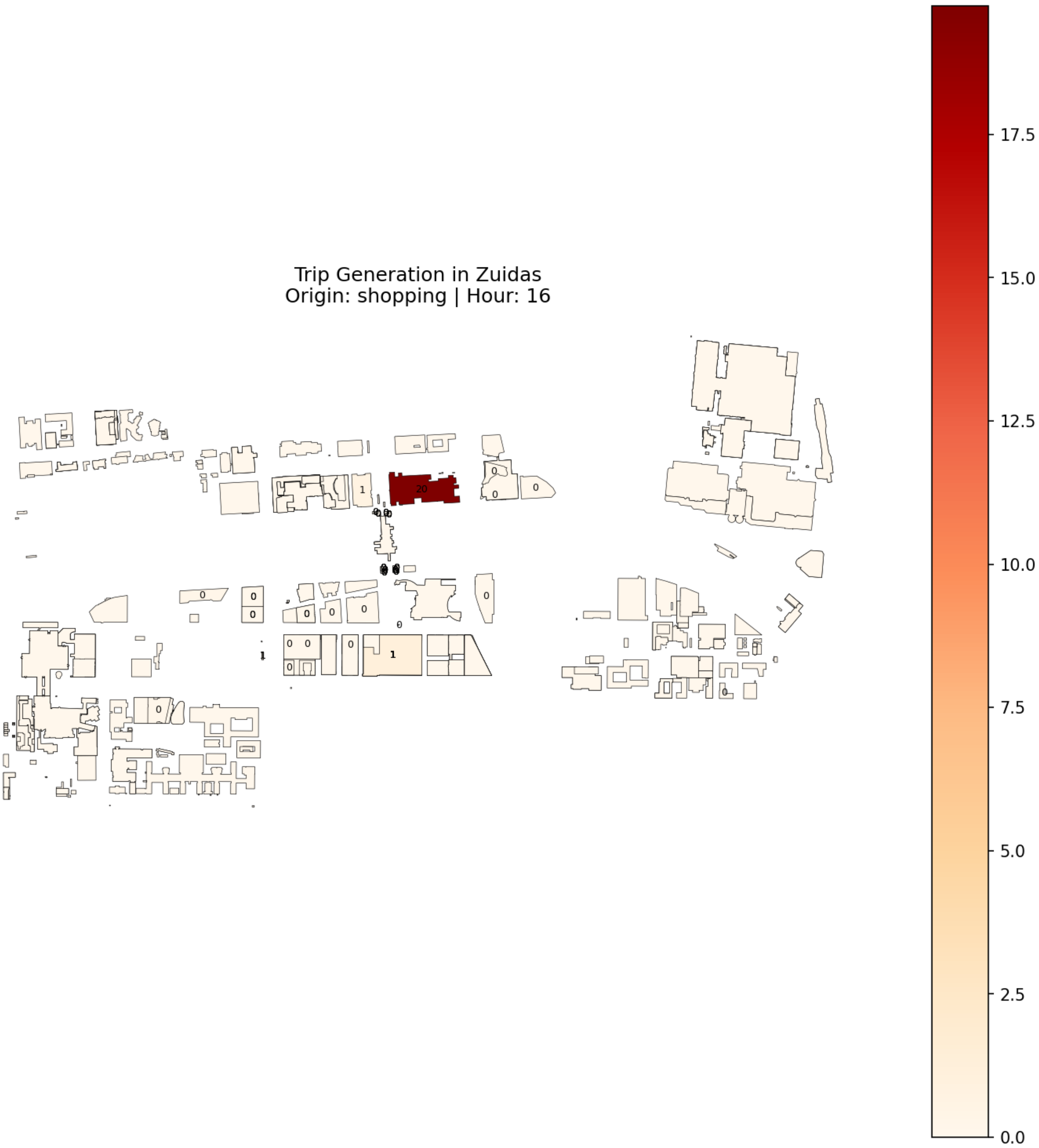

Trips in Zuidas — Origin: leisure | Hour: 16

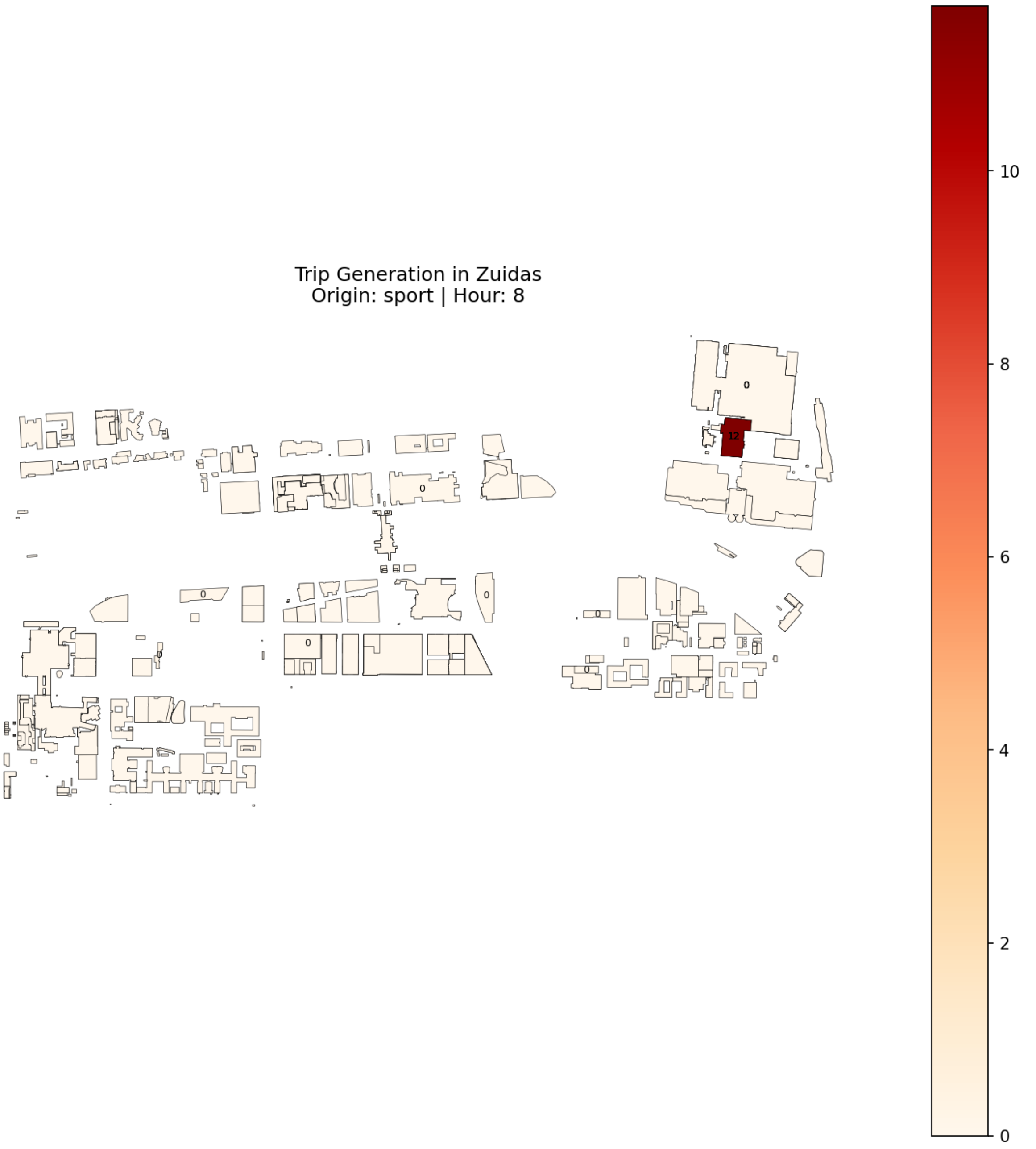

Trip Generation in Zuidas
Origin: other | Hour: 8
0.6
0.5
0.4
0.3
0.2
0.1
0.0

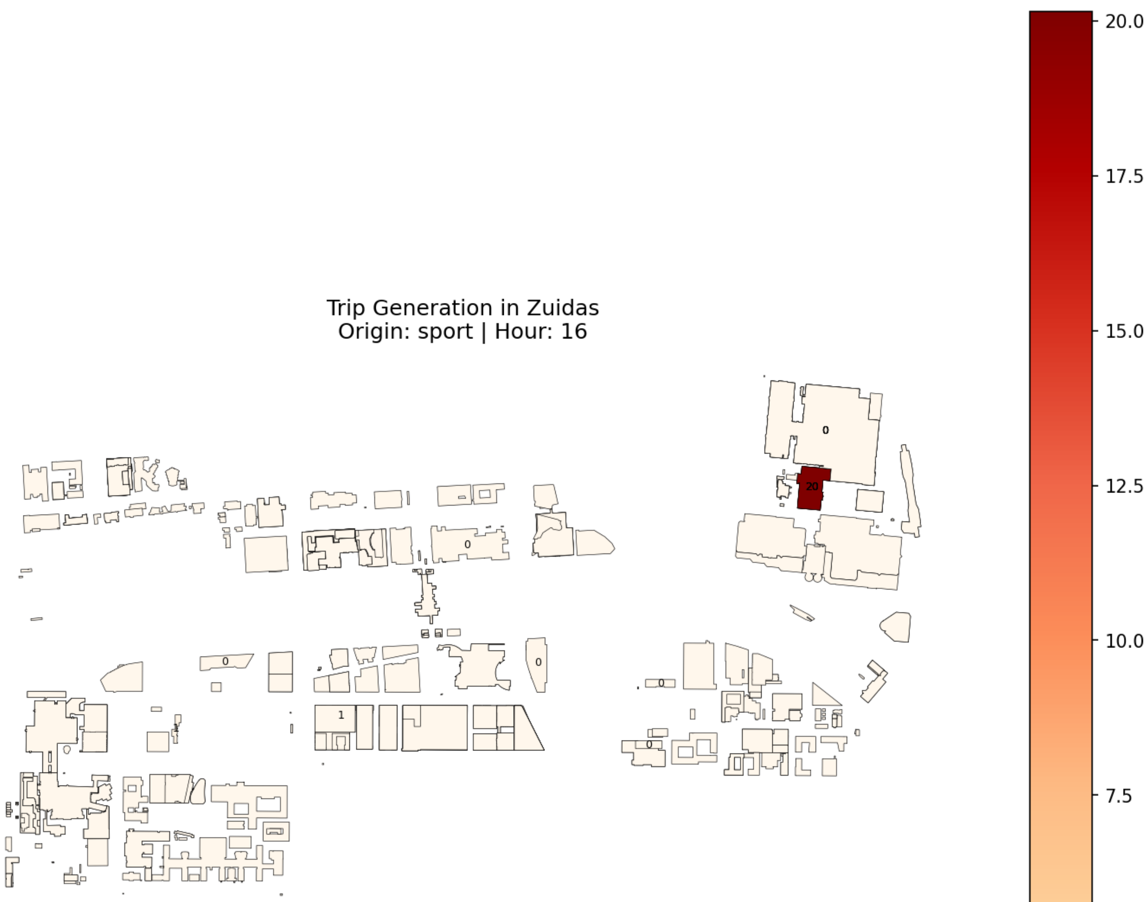

Trip Generation in Zuidas
Origin: other | Hour: 16
0.25
0.20
0.15
0.10
0.05
0.00



Trip Generation in Zuidas
Origin: shopping | Hour: 8
5
4
3
2
1
0


Trip Generation in Zuidas
Origin: shopping | Hour: 16
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0


Trip Generation in Zuidas
Origin: sport | Hour: 8
0
12
0
0
0
0
0
0
0
0
10
8
6
4
2
0

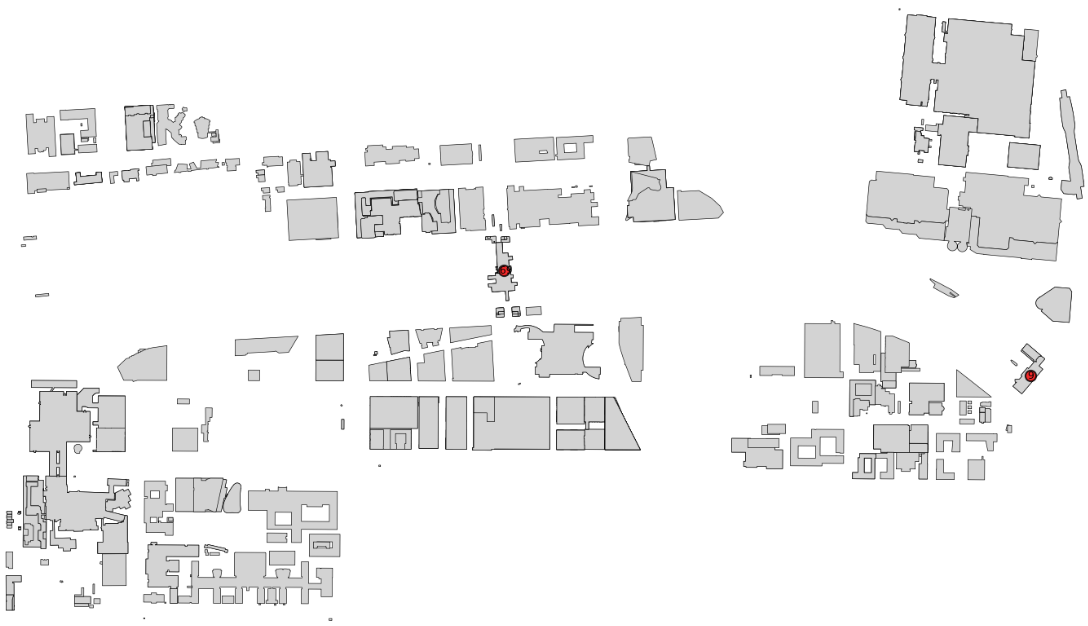

Trip Generation in Zuidas
Origin: sport | Hour: 16
0
20
0
0
0
0
0
1
0
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0

Trips in Zuidas — Origin: subway_station | Hour: 8

Trips in Zuidas — Origin: subway_station | Hour: 16

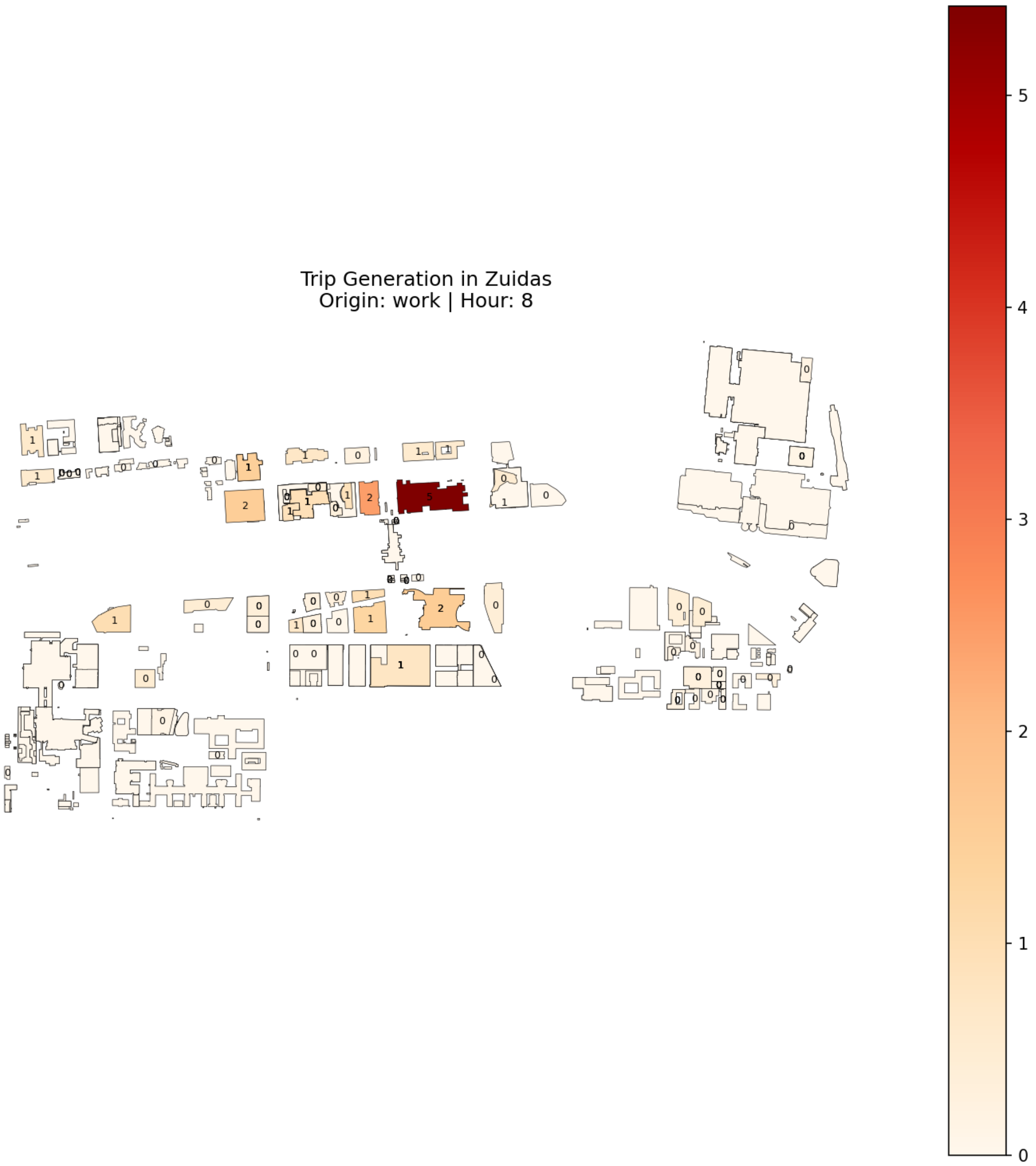
Trips in Zuidas — Origin: subway_station | Hour: 16

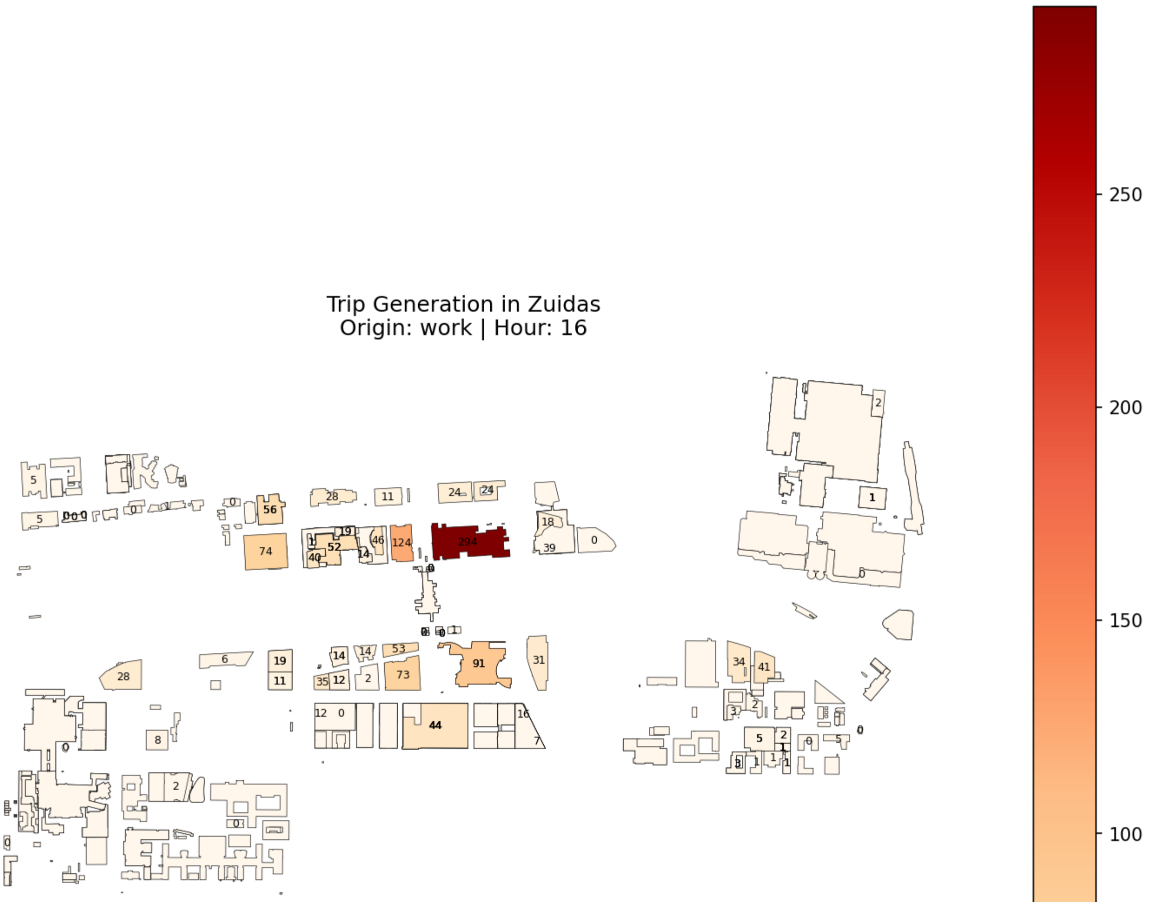
Trips in Zuidas — Origin: train_station | Hour: 16



Trips in Zuidas — Origin: tram_stop | Hour: 8

Trips in Zuidas — Origin: tram_stop | Hour: 16


Trip Generation in Zuidas
Origin: work | Hour: 8
0
1
2
3
4
5


Trip Generation in Zuidas
Origin: work | Hour: 16
250
200
150
100
50
0