Bicycle Travel Demand Estimation Method

TU Delft Campus Case Study

Syakal Hammas KH Hujaemi

Master's Thesis

Bicycle Travel Demand Estimation Method

TU Delft Campus Case Study

by

Syakal Hammas KH Hujaemi

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on May 28 2025

Student number:	5690021	
Thesis committee:	Dr. M. Snelder,	TU Delft
	Dr. ir. A. J. Pel,	TU Delft
	Dr. ir. W. Daamen,	TU Delft

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

Preface

This thesis is the manifestation of my stubborn passion for travel demand modeling, a journey that began during my undergraduate studies in civil engineering in Bandung. The course on transport modeling taught by Prof. Ofyar at that time opened my eyes to how directly civil engineering can connect to people's lives. This realization was further cemented during my Master's in the Traffic and Transport Engineering track at TU Delft, particularly through the course "Sustainable and Resilient Transport Networks and Systems" taught by Dr. Maaike and Dr. Adam.

My perspective on how research should be conducted in the traffic and transport engineering fields was greatly deepened through interactions with Dr. Winnie, who provided invaluable insights into rigorous academic inquiry and methodology.

I would like to take this opportunity to express my profound gratitude to my supervising committee, who have shown immense patience and support despite my many shortcomings. Dr. Maaike, in particular, helped narrow and refine my focus, bringing clarity to my work. Dr. Adam patiently guided me through the complexities of research methods and helped resolve my doubts. Dr. Winnie's guidance was instrumental in shaping my approach to the research process.

Lastly, I must acknowledge my ultimate support system: Allah, my parents, my brother, and my friends—especially Der Derian and Nuraulia—who stood by me during difficult times. Their unwavering belief in me and constant encouragement have been the foundation of my resilience and perseverance.

This thesis is not only a reflection of my academic endeavor but also a testament to the collective effort and support of those mentioned above.

Abstract

Bicycle travel demand analysis remains relatively underexplored, yet it is becoming increasingly important for urban and campus planning. In the Netherlands—particularly within the TU Delft community—cycling is deeply embedded in daily life, creating a strong demand for supportive infrastructure. While congestion may not pose the same threat to bicycle transport as it does to motorized traffic, maintaining the performance and safety of the bicycle network is essential. A busy cycling network may not always be visible at a glance, yet it can present safety risks.

The primary aim of this research is to identify a modeling process and specifications that are compatible with the available data, while laying the groundwork for future improvements to the bicycle network, especially TU Delft Campus. This will help ensure the system remains adaptable and relevant for long-term planning.

To identify an appropriate modeling approach, an exploratory analysis of the data was conducted. A clear pattern emerged in bicycle traffic, characterized by short-interval fluctuations corresponding closely with lecture schedules. An additional notable observation is the occurrence of an average peak in bicycle traffic during midday. These findings support a dynamic analysis approach with a 5-minute interval.

Moreover, the model incorporates specialized variables defined by the study's scope, focusing on trip generation and trip distribution within the established four-step modeling framework, specifically tailored for Origin-Destination (OD) matrix estimation in transportation engineering.

For trip generation, linear regression coupled with backward stepwise elimination via the Ordinary Least Squares (OLS) method was employed to identify significant predictors. For trip distribution, the Iterative Proportional Fitting (IPF) method was utilized. This approach was chosen based on the assumption that impedance is minimal for short-distance travel, a scenario particularly relevant within the TU Delft campus context.

Ultimately, this methodology provides a flexible and responsive framework tailored to the specific transportation dynamics at TU Delft, producing valuable insights for optimizing bicycle network planning.

The developed model is relatively simple but exhibits several shortcomings. One significant limitation is related to data collection, as the available data lack the temporal resolution necessary to fully capture the dynamic travel patterns targeted by the model. Additionally, the linear regression approach used for modeling trip production and attraction yielded unsatisfactory results, with R^2 values below 0.5. Another issue is potential underfitting, as indicated by the improved explanatory power of the model when trained on smaller datasets. Validation using RMSE and comparative plots of modeled versus actual flows further confirms that substantial improvement is needed in the model's reliability and predictive capability.

The trip distribution process, conducted using the Iterative Proportional Fitting (IPF) method, reveals additional areas for improvement. The OD matrix underestimated total production by two bicycles in a 5-minute interval. Although seemingly small, this discrepancy underscores the necessity for more robust input data and methodological refinements. Additionally, direct validation of the OD matrix is crucial to enhance accuracy and reliability in representing actual travel flows.

Despite the shortcomings, the framework provides a balance between interpretability and flexibility, enabling both accurate representation of observed travel behavior and ease of scenario testing—making it a practical tool for supporting data-driven mobility planning and policy evaluation on campus.

Contents

List of Figures

List of Tables

1
Introduction

Bicycles play a significant role in Dutch culture, deeply embedded in the daily lives of citizens who rely on cycling for short-distance travel across various purposes. This phenomenon is particularly evident within university campuses, especially at Delft University of Technology (TU Delft). Despite the diverse backgrounds of the TU Delft community, cycling remains a dominant mode of transport for students, faculty, and staff engaging in campus activities. The widespread use of bicycles in society naturally increases the demand for well-developed cycling infrastructure, regardless of the scale of the community.

The development direction of the TU Delft campus is outlined in the Campus Vision 2040 (TU Delft Campus, 2023), which envisions a more adaptive, inclusive, lively, and sustainable campus. As a global hub for engineering, TU Delft aims to integrate seamlessly with the city, embracing the *"Campus = City"* perspective, which emphasizes connectivity between the university and its surrounding urban environment. This approach fosters an open and accessible campus, welcoming not only students and staff but also visitors and neighboring communities. Additionally, the *"Campus = Meeting Place"* perspective highlights TU Delft's ambition to be a vibrant hub where diverse communities can engage in various activities. Given the anticipated growth of the campus population—driven by historical trends and the increasing demand for engineers—the need for an efficient transport and traffic network is more critical than ever. Ensuring that mobility infrastructure aligns with the university's vision is essential for accommodating this growth while enhancing accessibility and sustainability.

Furthermore, the campus also perceives the need to connect the campus as a whole. It is realized that the *Campus Midden* section is the most crucial part due to the number of education buildings that would attract people. The connectivity of the campus internally becomes important to spread and support the activities across the campus. According to (TU Delft Campus, 2023) the bicycle network is expected to play a crucial role in this case. The expansion of the network is considered.

The bicycle culture, the increase in population, and the vision of the campus in the future give a clear indication for better planning in the transport and traffic network. Regarding this the importance of knowing the demand, to provide the user better, becomes crucial. Understanding the flow could impact the planning to be more efficient and optimum if the network is designed by following the demand of the user. Travel demand is not only the factor that justifies the design of the network. The TU Delft Campus as an academic entity also has several principles that are believed and these principles are interpreted by the authorities into the operations of the campus life.

1.1. Problem Definition

The demand analysis for the motorized mode is very common and has a wide variety of applications. On the other side, the analysis to estimate the travel pattern in the bicycle network exists in research. However, the research of the travel demand analysis for cyclists is still limited in quantity and variety. The practice remains scarce might be because bicycle usage around the world is not as significant as in the Netherlands. Even so, the United States, England, Australia, and some other countries already have research regarding the mapping of their cyclist travel pattern (Bhowmick et al., 2023) for the city or region scale analysis.

It is important to note that the campus scale travel demand analysis is also quite rare. The city-scale travel demand analysis is more popular since the budget and stakes of designing are heavier than the smaller-scale planning. However, it does not make the importance of understanding the travel demand at the campus level becomes annulled. To achieve the vision of the TU Delft campus and to provide growth, the campus scale travel demand analysis should be highlighted. The scarcity of the practice of the campus-scale travel demand analysis for bicycle networks can be handled by contextualization of the existing research.

To ensure a meaningful analysis of travel demand at the TU Delft campus, it is essential to clearly define the model's purpose. The chosen methods must align with this purpose, considering the objectives of the analysis and the constraints, such as data availability and accessible software tools. In line with the campus vision of emphasizing bicycle networks, the travel demand estimation should address the specific needs of the campus environment.

The needs of the model by the campus bicycle network can be determine by identifying the key stakeholders involved in the development of the maintenance of the network. In the case of involved stakeholders, both the university and the municipality are responsible for the bike network. Since TU Delft significantly influences campus traffic, the university needs to understand the infrastructure requirements necessary to support its activities. Simultaneously, collaboration with the municipality is necessary, as it is responsible for the planning, development, and maintenance of transportation infrastructure. Additionally, the Campus Real Estate & Facility Management (CREFM) department should be indirectly involved, as it is responsible for infrastructure and facilities such as buildings and parking areas. The Education and Student Affairs (ESA) office also plays an important role by supporting lecture-related activities, for example in scheduling. Effective coordination between faculties and the education logistics division of ESA could enable a two-way feedback system, ensuring that lecture schedules are optimized to prevent unnecessary busy periods at vulnerable locations that could lead to safety risks. If new infrastructure or redevelopment of the TU Delft bicycle network is necessary, an integrated and cooperative approach between the university and the municipality would be crucial. The university can provide actual insights into mobility needs, while the municipality can facilitate implementation, ensuring a well-planned and sustainable bicycle network.

While issues within the bicycle network are not always evident, certain conditions highlight areas of concern. For instance, the Jaffalaan intersection occasionally experiences high traffic volumes, leading to temporary congestion. Although these high network business events are brief, they still pose risks such as accidents and reduced safety for cyclists. Hence, the analysis that would result the business level of the bicycle way is important to be achieved.

To identify the level of bicycle network utilization that expresses its business, it is essential to estimate link flow volumes. This would provide a clearer representation of network demand and allow for a more detailed analysis of bicycle traffic patterns. To support this, the bike network should be modeled as links, enabling a structured approach to assessing flow distribution and infrastructure needs.

If the bicycleway is modeled as a series of links, the research requires an adequate spatio-temporal resolution. For the TU Delft campus, shorter time intervals may be necessary to effectively capture traffic patterns. However, increasing the resolution also increases the need for detailed data to support the analysis. Current limitations, such as the number of traffic count observations and the availability of socio-demographic and capacity data, must be addressed to enhance the model's accuracy and applicability. While obtaining additional data is an ideal long-term goal, it is essential to establish a modeling pipeline that functions effectively with the available resources and appropriately limits the study area's scope given these circumstances.

Thus, the aim of this research is to understand travel patterns by examining origin-destination relationships within the campus area, represented through an OD matrix. The primary objective is to develop a modeling process and specifications that align with the available data while establishing a foundation for future enhancements to the bicycle network.

1.2. Research Questions

As previously stated, this research aims to identify the modeling process and specifications for trip distribution, represented by the OD matrix, using the data currently available. The research questions are raised to achieve the research aim.

"How can a reliable method for estimating cyclists' travel demand be developed to describe spatiotemporal patterns and support the planning of the bicycle network at the TU Delft Campus?"

To address this main research question, several sub-questions are formulated to structure the analysis:

1. What are the needed data and existing methods for estimating cyclist travel demand at the TU Delft campus?

This question lays the foundation by identifying commonly used data and the state-of-the-art of the OD matrix estimation for the bicycle network. It is discussed in Chapter 2.

2. What model specifications are necessary to apply travel demand modeling effectively for the TU Delft campus bicycle network development?
   This question explores estimation methods compatible with the available data sources within the TU Delft. It is addressed in Chapter 3.
3. How is the origin-destination flow estimated for the TU Delft bicycle network?
   The OD flow, represented by OD matrix, helps analyze cyclist flow and demand. This question includes the assumptions and distribution patterns of OD pairs and is discussed in Chapter 4.

1.3. Research Methodology

This research is structured to answer the research questions systematically. The first sub-research question is addressed through a literature review, which focuses on existing methods for travel demand estimation. In addition to reviewing the methods, the data used in each approach is analyzed to determine its appropriate implementation. The literature is then evaluated for its applicability to the research scope, specifically in estimating a bicycle OD matrix at the campus scale. Through this evaluation, the proposed method is identified.

The second sub-research question is primarily covered in the data description and methodology framework. The available data from the campus environment is evaluated, leading to the formulation of assumptions and hypotheses. Based on this assessment, model specifications, such as the spatio-temporal dimensions, are determined. The estimation method is then defined by correlating the proposed methodology with the available data.

The third research question is addressed through the travel demand estimation process, expressed by the OD matrix estimation for the TU Delft bicycle network. This section implements the defined method and specifications established in the previous section, ensuring consistency in the approach.

1.4. Scientific and Societal Relevance

Existing models and analyses of travel demand often focus on macroscopic trends or specific characteristics, such as riding behaviors, without fully addressing the unique travel dynamics within a university campus. These models tend to either rely on direct travel demand modeling or are built toward more specialized purposes, such as analyzing the physical characteristics of bicycle riding behaviour. However, there is a gap in the literature when it comes to understanding travel demand in smaller, campus-specific settings. This presents an opportunity to develop a travel demand analysis that specifically addresses the campus environment at a macroscopic level with bicycle network as the focus, giving a certain insight in the field.

By creating a travel demand model focused on the campus scale, researchers could gain a better understanding of the unique factors influencing bicycle usage. This model could lead to more informed and effective planning for bicycle facilities and networks on campus, helping to optimize the placement of bicycle facilities and networks. With better insights into travel patterns, the university can make strategic decisions to improve accessibility, reduce congestion, and encourage more sustainable transportation options. Ultimately, this research has the potential to provide a framework for bicycle infrastructure planning, not only benefiting the campus community but also contributing valuable knowledge to the broader field of travel demand modeling. For TU Delft campus context Campus Real Estate & Facility Management (CREFM) is benefited from the existence of such a model since the effect of the development plan can be evaluated in the short moment and the further discussion regarding the bicycle network development could be more dynamic and be processed right away.

1.5. Report Structure

The report of this research is structured as follows. First, the literature review, in which the state-of-the-art practices of bicycle transport model in general with its used data is depicted in Chapter 2. Next, Chapter 3 will provide a data description to grasp a better understanding on the needed specification for the transport model and the implemented model framework based on the description. In Chapter 4, origin-destination matrix model is estimated for the TU Delft campus with the chosen method on Chapter 3. Thereupon, the results are discussed in Chapter 5. At last, Chapter 6 draws up the conclusions of this research.

2

Literature Review

This section explains the state-of-the-art bicycle transport model in general, specifically the travel demand estimation for the campus scale. Furthermore, the literature overview of all the references regarding the demand for bicycle travel is explained in this section. The literature review method is described to give a picture of how the literature is curated and summarized to formulate a methodology used in this research. The data also becomes a consideration in the literature review to evaluate the fitness of the reference method for the research in this thesis.

The perspective of this project still sticks to the exploration of how the travel demand of the bicycle network at TU Delft can be built based on the available data and how to improve it by another source of data that is not available yet. The ability to evaluate the policy and the infrastructure development are also considered in the method. By these chapter's aims, the goal of the literature review is to determine the suitable method to express the travel demand of the TU Delft bicycle network based on the research context, the available data, and the ability to cover necessities for the future direction of the TU Delft Campus development.

2.1. Literature Review Methodology

The literature review is conducted by looking at the existing paper that explains bicycle travel demand retrospectively. It evaluates the applicability of the method and the state of the research in the respective field. In this case, the bicycle travel demand is a part of the transport model of the bicycle network. To be precise, the bicycle travel demand is modeled in the macroscopic transport model that is depicted in the link-volume model. The paper review on the topic is made to evaluate the state of the link volume model of the bicycle network. The review paper discusses the variation in reporting of study characteristics, modeling accuracies, validation methods, their strengths, and limitations of the major modeling approaches (Bhowmick et al., 2023).

Since the paper has been published recently, it has been used as the main reference in the research. Even though it is the main reference for the literature review, the literature reviewed in the paper is sometimes repetitive. Some papers are made by the same authors with similar methods. Even different authors would come up with similar methods. It also proves that the exploration of the methods is still needed to enrich the knowledge in this case. Other than the paper, the literature search is also done by looking at the literature platforms available on the internet and depicted on Table 2.1.

Keyword	#-search results in Web of Science	#-search results in Sciencedirect	#-search results in Google Scholar
bicycle travel demand estimation	45	5,355	about 69,400
campus travel demand estimation	50	4,833	about 108,000
simulation of the bicycle transport model	246	6,888	about 79,300

microsimulation of the bicycle transport model	18	336	about 8,900
campus travel pattern	1,226	23,514	about 613,000
bicycle travel demand estimation AND simulation	3	2,015	about 33,600

Table 2.1: Number of search results in each platform

The literature search on each platform is conducted using the keywords specified in Table 2.1, chosen for their relevance to the research context. The primary keyword, "travel demand estimation," is used to focus the search specifically on methodologies for bicycle travel demand estimation. To refine the results and better align with the research scope, these keywords are combined with specific methods, modes, or scales. This approach helps capture the study's context while filtering out unrelated results. Additionally, backward snowballing is applied to both the main references and their cited sources to expand the literature review and identify relevant studies.

As stated at the beginning of the Chapter 2, obtaining applicable methods for travel demand estimation is implemented by looking at several papers that fit the needed characteristics for the model. The characteristics of the model are broken down into requirements that will meet research needs. The required attributes in the research are to fulfill the research context and also the model needs to be applicable in the future model for developing the TU Delft bicycle network. The requirements for each evaluated characteristic are depicted on Table 2.2.

Characteristics	**Requirements**
Research Context	Methods
	Campus-Scale
	Applicable for campus-scale
	Bicycle
	Is it applicable for bicycle>
Future Application	Is the OD matrix included?
	Is the travel demand prediction included?
	Dynamic/fine temporal resolution
	Policy or infrastructure evaluation included?
	Ground truth data
	Specific Variables
	Validation

Table 2.2: Literature review characteristics and requirements

The research context focuses on the travel demand on a campus scale at the TU Delft bicycle network. In terms of the research scale, the method needs to be fit for the campus scale since the area size and motivation to move in the campus scale are different from another scale. However, the applicability of the method in the case study scale is also important and evaluated. The transport mode in the methods also needs to be applied to the bicycle. Nevertheless, the literature search is not limited to the bicycle mode as long as it could be implemented for the bike network.

The additional model requirements are defined based on its ability to be used in the future. The OD matrix estimation is needed to understand the cyclist's travel pattern. Despite the lack of interaction with the network supply, it could express the pattern comprehensively. It is also interesting to see the evaluated literature on the predictive ability of the bike traffic flow. It is crucial for the planning of the network to provide the future demand. The temporal resolution should be on the finer side for the bicycle network analysis. Even though the aggregated demand on the daily bike is easier to obtain, to capture the fluctuation in the traffic flow, it might be better to analyse the network in the smaller time interval. Smaller time interval allows identifying peak hours and the peak volume accurately. The paper with the policy or infrastructure evaluation is needed as a reference for future planning. If the evaluation is integrated into the model or analysis in one streamline, it could be a positive point to have a comprehensive tool for bike network planning. The usage of ground truth data is essential since the bike count might be done for the whole network in strategic locations. This would increase the accuracy and simplify the modeling process, such as the trip generation of a building can be generated directly from the count of people through the entrance. Once ground truth data exists, the validation and the calibration of the model should be realistic to be implemented and should be included in the planning process. Additionally, the future application of the campus scale might need specific variables that are relevant to the travel pattern or behaviour around the campus area. The short-haul travel around campus has different attributes from the travel within the city, therefore, the specific variables that can be extracted from the campus network might be useful to capture details and representative travel patterns.

The data used in the analysis is also evaluated for all the references. The data that is available in the TU Delft might be limited compared to the data used in all the references. Hence, the travel demand estimation method for this research is also controlled by the data availability. However, the data limitation does not limit the method directly, the method used can be creative based on the specific variable that is more detailed and focused on the behaviour of the people of the campus entities. Furthermore, the data that represent the supply of the network and the demand of the campus are decided and summarized on Table 2.4. The type of data listed in the table is based on the existing paper review on the bicycle link-level volume transport model (Bhowmick et al., 2023).

2.2. Trip-Based Model

The transport model basically could be derived from trips, tours, or activities. A trip refers to a single movement from one origin to one destination. It depends on the purpose and also the mode in the transport model. Furthermore, the chain of trips is called a tour when there is a closed cycle of the trips. A tour typically starts from the house and ends with the house. It also has intermediate locations between the tour's beginning and the last destination. Although the tour is related to the purpose and the mode, the tour is still highly correlated to the location, so the tour is expressed by the location chain (Ortúzar S. et al., 2011). On the other hand, there is the activity-based model, which refers to the origin and destination based on the agent's activity.

The trip-based model heavily focuses on trip distribution based on the production and attraction of the zone or the region. In that case, the four-step model is the classic example of this kind of approach. The other approach with the trip-based perspective is the direct demand model.

2.2.1. Four-Step Model

The classic approach for the macroscopic transport model has been used since the 1950s. The four-step model was used in several USA development plans after World War II, from the Detroit Metropolitan Area Traffic in 1953-1955 to the Chicago Area Transportation Study (CATS) led by Dr. J. Douglas Carroll, Jr. in 1955. Those models were the ones of the first formal application of the four-step model in analysis of the travel demand. In the 1960s, the four-step model was standardized as the primary method for transport modeling in the USA followed by the Federal-Aid Highway Act of 1962 (Weiner, 2008).

The long history of the four-step model gives a comprehensive empirical example of its implementation, including for the bicycle. However, the bicycle transport model is still scarce. The link-level volume models of the bicycle are more familiar as stated before. Generally, the process of the four-step model is captured in Figure 2.1.

The four-step model approach divides the process into sub-models. It starts with trip generation, where all the trips generated are estimated based on the determined zoning and network system. The production and attraction of each zone are estimated to become the controlling values for the trip distribution. The trip distribution starts with initial values for the OD matrix are determined based on the deterrence function that reflects the resistance of the travel willingness depending on the destination. Modal split and trip assignment then follow the process to complete the four-step model. Some practices implement the modal split and trip assignment at the same time or the trip distribution and modal split simultaneously (Ortúzar S. et al., 2011). Nevertheless, the modal split before the trip distribution could be found as a form of simplification. It makes the mode choice not related to the destination, which is a bit inaccurate approach. Hence, the complexity of the four-step model could be related to this point.

In the context of the bicycle network, some practices use the four-step model as the basis of their model, some of them also alter the method to capture specific phenomena or to simply adjust the complexity of the model regarding the data availability and/or level of details. The scale of the model is dominated by the city

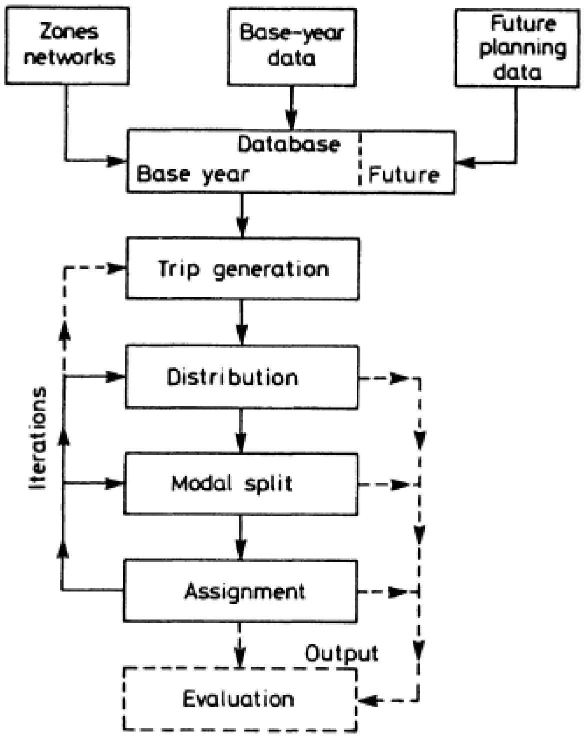


Figure 2.1: Classic transport model: four-step model
Source: Ortúzar S. et al. (2011)

scale since the method itself is suitable for the city scale. The usage of the bike count data is frequent for the four-step model and is supported by the land use data to process the trip generation. Since the commonly found scale is a city, the deterrence function used as the basis of the trip distribution is strongly related to the distance. Hence, the contextualization needs to be made to reflect the campus scale in the analysis later.

The four-step model is quite settled in practice and the usage of commercial software, where some of the steps (trip distribution and trip assignment) or the whole process are being made in the software (Hamad & Obaid, 2022; Jacyna et al., 2017; Oskarbski et al., 2021; van Dulmen & Fellendorf, 2021). The advantage of using commercial software is the ability to analyse the multi-modal network since several layers, scenarios of demand, and controls could be implemented. In this case, the bicycle layer can be determined as one of the layers, and the modal shift can be traced based on the model (Oskarbski et al., 2021; van Dulmen & Fellendorf, 2021).

Since the four-step model consists of sub-models, it could enhance the integration between different platforms. The trip generation could be made based on the existing demand model such as the activity-based model (van Dulmen & Fellendorf, 2021), and integrated with another software that is more relatable for the trip distribution. Hence, the validation needs to be done in every step to avoid error propagation through the succeeding sub-models (Bhowmick et al., 2023).

The four-step model also allows the refinement of the sub-model result. The existing literature shows that the trip distribution can be refined through two stages of the process. The trip distribution can be initially done by using the conventional method such as the gravity model or the growth factor (Ortúzar S. et al., 2011). Furthermore, the result from the gravity model can be the input for the next stage of trip distribution such as the path flow estimator, a logit-based path choice model that is focused on the traffic of the path between origin-destination pairs, rather than the specific link (Ryu, 2020; Ryu et al., 2019). The implementation of the PFE hopefully will decrease route overlapping issues by applying the path-size logit.

Moreover, the transport model on the campus scale has already been attempted to be made by Hamad and Obaid (2022). In the context of Sharjah University City, UAE, the transport model to forecast the travel demand of the campus had been made using the tour-based approach derived from the traditional four-step model. This study also shows how the campus scale transport model generally can be developed. The data and the streamline can be used as the reference in any case of the transport mode.

The data used in the four-step model is the expression of the production and attraction factors for describing the travel demand of the regions to support the trip generation. On the other hand, the network characteristics such as distance to each node and/or zone to support trip distribution reflect the supply of the network. Additionally, the bike count or the traffic flow data is also used as the basis for the validation and calibration process, and for the path-size logit, the bike count is used as the basis of the optimization process by Ryu et al. (2019). The destination is also influenced by the existing of the travel survey to see the preference of the population sample.

The usage of the four-step model is known widely due to its low computational complexity and its easiness of interpretation. The method also unable the possibility of better performance in forecasting by introducing many variables. The model also has the ability to implement the policy or infrastructure changes for the evaluation of its effect. However, since the model consists of several sub-models, the validation is necessary for each process due to the risk of error propagation through the streamline (Bhowmick et al., 2023).

2.2.2. Direct Demand Model

The direct demand models known for a long time, as long as the four-step model perspective. The usage of direct demand was dominated by the transit system, especially in the 1960s. One of the earliest forms of the direct demand model is the SARC by Walter Kraft and Robert Sarc in 1968 (Ortúzar S. et al., 2011) and it is formulated for the urban transit ridership. The SARC estimates the demand through a multiplicative function of activity and socioeconomic variables. Later, the implementation of the direct demands in the non-motorized travel demand estimation started in the early 2000s.

The concept of the direct demand models, which differ from the four-step model, is the simultaneous process of the generation, distribution, and modal split. It could be reflected by only one model (without the building of sub-models), avoiding a sequential process of the model. It could also reduce the efforts because it means the validation and calibration could be conducted on only one representative model (Ortúzar S. et al., 2011).

Direct demand models consist of two types. First, the single estimated equation to connect the travel demand straight to the mode, origin, destination, the traveler characteristics. Second, there is some sort of separation between the modal split and total travel demand (Ortúzar S. et al., 2011). However. those two models are closely related to the socioeconomic variables and relations. One of the examples of the single equation estimation is the SARC model (Ortúzar S. et al., 2011).

$$T_{ijk} = \phi(P_i P_j)^{\theta_{k1}} (I_i I_j)^{\theta_{k2}} \prod_m [(t_{ij}^m)^{\alpha_{km}^1} (c_{ij}^m)^{\alpha_{km}^2}] \tag{2.1}$$

where P is population, I income, t and c travel time and cost of travel between i and j by mode k, and ϕ, θ, and α parameters of the model. The variation of the model then emerges to simplify the complex expression by bringing up composite variables such as Manheim's equation (Ortúzar S. et al., 2011). One of the examples that describes the second type of direct demand model is depicted on Figure 2.2

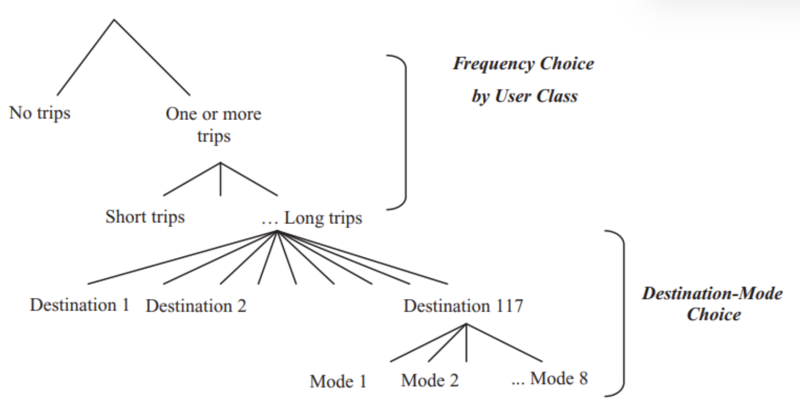


Figure 2.2: Direct demand model with choice model
Source: Ortúzar S. et al. (2011)

The recent direct demand models seem similar to the four-step model since the components of generation, distribution, and mode choice are represented through the choice model paradigm. However, the process is still included through one utility function. The choice model of the direct demand consists of the destination-choice mode in the nested logit model. Still, some models also consider the attractiveness of the destination like the model by Daly (Ortúzar S. et al., 2011).

It needs to be noted that the direct demand models for the non-motorized and active modes were more popular in the late 20th century, aligned with the formal usage of the discrete choice model with random utility maximization theory in transport engineering fields. The advancement in machine learning also allows more complex regression techniques to support equation formulation. Those conditions propose more variety in the purpose and context of the direct demand models. For example, there are practices where the direct demand models are used for traffic flow estimation from the calculated Bikeability Index (BI) (Arellana et al., 2020), link-flow volume estimation (Dadashova et al., 2020; Lin & Fan, 2020; Lu et al., 2018), influential characteristics on bicycle network (Orvin et al., 2021). The direct demand models have become the most used method for the travel demand estimation of the bicycle network.

Arellana et al. (2020) proposes a calculation of the Bikeability Index (BI) and incorporates the direct demand model to calculate the traffic flow of the cyclist based on the index. The direct demand model is introduced to predict the traffic based on the BI as the interaction parameters between two zones. It is very interesting to note since the BI is the parameter in the equation to express the quality of the bicycle way is important to the bike ridership in the city. However, the application is not very compatible with the campus scale since the parameters used in the equation are more familiar to the city scale.

Lin and Fan (2020) and Orvin et al. (2021) implement the direct demand models to determine influential variables for the bicycle network. One of the models uses STRAVA data of Charlotte, USA to execute in the probit model. The cyclist characteristics and the network variables are set as parameters of the utility function in the probit model. The research showed that bicyclists prefer off-street paths, and planners can design more off-street paths to offer a better bike environment for bicyclists in Charlotte (Lin & Fan, 2020). The function is determined by using the backward method.

Another research (Orvin et al., 2021) that inspects the influential variables for the bicycle network uses a latent segmentation-based negative binomial (LSNB) model for Auckland, New Zealand and Kelowna, Canada. The log-likelihood maximization method on the LSNB parameters is used to determine the influential variables of the bicycle network in New Zealand and Canada. The considered variables, such as the road connectivity index, bike index, and AADT (Annual Average Daily Traffic), are included in the LSNB. The research discussed the importance of increasing the BI is important to increase the bike ridership in Kelowna. The model also describes the influence of higher temperature, lower rainfall, increased length of shared paths, lower AADT, and closer distance to bus stops are likely to increase cycling demand in Auckland.

Lu et al. (2018), proposes the direct demand model for bicycles and pedestrians in the form of the equation. Two kinds of models are produced in this research. The first is a spatiotemporal model, which includes time as the model variable. The second model produces 24 sub-models for each time of the day. The used variables are focused on the census, land use, and transportation network data of Blacksburg, Virginia, USA. The research concludes that traditional variables such as land use and transport networks are correlated to bicycle traffic. Likewise, temporal variables also correlated to the demand. However, the variables in the models are highly related to the city/town infrastructure. It is also tough to contextualize on the campus scale.

The direct demand is also used in the AADB (Average Annual Daily Bicycle) counts in several cities in Texas by Dadashova et al. (2020). The AADB is counted based on the variables derived from the transport network such as the OSM functional class and Texas internal inventory. It is found that the functional class of the road is related to bicycle traffic.

It could be seen from all the existing recent models that the traditional variables such as census, land use, and transport network, are highly related to bicycle traffic. It has to be noted that models are estimated for the city or town scale where the distance is a key factor for the cyclist. Another data that can be relatable is temporal variables, either discrete or continuous time. Furthermore, the bicycle count data sourced from the Strava or other GPS-based crowdsourced data are also familiarly used in the direct demand models.

It concludes that the direct demand model offers simplicity since the sub-process could be done in the model. It means the process of validation can be done with less effort and less ground truth data needs to be obtained. The relationship between variables also could be observed more easily than in the four-step model. However, there are several downsides to the model. It is hard to notice the relation of the mode to its network or another mode. Also, the route choice process is not described very well by the model (Bhowmick et al., 2023).

2.3. Activity-Based Model

Travel has always been a "derived demand". The human rarely travels. We rarely travel just for the sake of travelling. We do it in order to satisfy a particular need or requirement at a different location (Ortúzar S. et al., 2011). Hence, human behavior becomes the main driver of the travel demand. Concerning the origin-destination matrix, behavior is the main motivation to choose a destination based on the activity and the mobility of the household. Hence, the activity becomes the main core.

According to Ortúzar S. et al. (2011), Mitchell and Rapkin already established the link between travel and activities. However, the practicality had not happened to be realistic in that era. Still, the idea of 'activity analysis' ran through the researchers. A significant contribution to this kind of approach happened around the 1970s. Hagerstrand, in 1970, proposed a time–geographic approach that implements constraints between location and activity. Peter Jones led a study of activities and travel behaviour in the Transport Studies Unit at Oxford in 1979. After the enhancement of the computational power, the activity-based model reached popularity in the 1990s with TRANSims becoming its pioneer. Nowadays, MATsim might be a commonly used agent-based model tool for transport models.

Activity-based models generally consist of the population's long- and mid-term choices and the personal day simulator (Ortúzar S. et al., 2011). The first long-term choice model decides the long-term characteristics of the population as the fundamental of the activity choices like place of work, car ownership, and season ticket commitments, Medium-term choices describe the household task for the daily activity. The person's daily simulator decides the sequences of activity to fulfill the task. In the context of the bicycle network at TU Delft, the ABM can be built with population synthesis, activity generation, and cyclist simulation.

The ABM (activity-based model) is investigated by many authors for different purposes and model characteristics, thus the example practices of the ABM could become insights for the research. Guo et al. (2013) attempted an agent-based model. The whole model does not represent the OD matrix specifically. There is one part in the model that generates the possibility of the destination based on the generated agent characteristics. The entry point to the campus is determined. It could represent the OD pairing of the campus. The context of this is for the parking and emission analysis. The research is conducted for the campus. Although, the campus dimension is larger than TU Delft. It might be a small influence, but the distance remains considered as the resistance in the model.

Kaziyeva et al. (2021) researched the agent-based model with several choice models. The set of rules is defined based on the travel survey of Salzburg with its adjacent region. The process comes from the agent creation and activity choice with the route and mode choices. The activity choices are defined with a predefined probability by their mode and order of activity. The rule for several specific groups of the population is also restricted like students and university students. The simulation depicts the spatio-temporal dimension of the model well with the peak hours expressed very well.

Another attempt was made by Jafari et al. (2024), which resulted in an activity/agent-based model based on the travel survey in Victoria, Australia. The model focuses more on the multi-modal simulation with the bicycle as one of the modes. Combining the network from OSM, GTFS, and other data, outputting mode share, and road traffic volume with one of the inputs is the hourly bicycle traffic volume sensor data. The lack of inclusion of the impact of the cycling-relevant road infrastructure makes the error on the bicycle volume high.

The basic ABM also attempted by Xu et al. (2018). The research made a framework for an activity-based model based on the household activity pattern problem (HAPP). The framework is comprised of three steps, (i) choice set generation, (ii) choice set individualization, and (iii) model estimation considering several models form. Clustering methods to identify the pattern are introduced.

The common data that is used by the previous research are obtained to support the ABM. However, the travel survey is crucial to determine the characteristics of the population due to the completeness of the data detail. ABM utilizes the preferences indicated by the socio-economics characteristics and the travel behaviour indicated by the origin-destination pairs and travel choices like the mode choice.

The ABM would result in a model with high explanatory ability. The heterogeneity of the rider is captured in the model. The interaction between each person and also the people and their environment could be captured. The future scenarios can be evaluated by this model and the model can be implemented for a long-term analysis. Nonetheless, the model is the most complex of all the approaches and requires enormous and detailed data input.

2.4. Other Methods

The travel demand estimation is also built with another method that implements different concepts of approaches, like simulation, machine learning, or deep learning. For the example of the simulation implementation. Blume et al. (2022) made a Monte Carlo simulation for OD estimation in transit networks with the support of the entry and outflow data count. In the context of the bicycle, it is better to identify the campus entry point and have a traffic count at each bike parking. The OD matrix is not determined specifically but rather expressed by the distribution that represents the probability of the destination chosen by a certain origin. The generation of the OD and test data is crucial and must be based on the validated assumptions. With a similar simulation, Khan and Habib (2023) conducted an MCMC to generate activities. However, there is a simulation based on the probability designed by the MNL for the destination choice.

Cooper (2018) researched an additional problem in applying the four-step model to cycling by including land use accessibility in the model using betweenness. However, betweenness might not be applicable at the campus scale, since the motivation to travel within the campus is not driven by distance. A mixed-methods approach was used between direct demand and the optimization of the GEH.

Besides simulation, machine learning is also used as an approach for travel demand estimation. Baumanis et al. (2023) applied a machine learning method using LASSO (to avoid overfitting) and OLS regression, followed by validation with mean squared error (MSE). Specific variables were used to emphasize the COVID-19 context, capturing phenomena such as the bicycle boom in the US. The scope of the study was the city of Austin (Travis County), considering UT Austin activity (25

Krishnakumari et al. (2020) also conducted research on an unsupervised learning approach using the principal component analysis (PCA) method, aiming to capture a data-driven OD estimation without requiring equilibrium assignment or a network loading model. This was applied to both a small toy network and a large network in Santander City.

2.5. Discussions on Literature Review

The literature review shows important points that need to be included in the research. Overall, the direct demand models are the most common method used in the bicycle network (Bhowmick et al., 2023), as can be seen from the Table 2.3. Four-step model is not used that often, but still has high reliability for the bicycle network if the traditional variables are used in the model. The activity-based model has many examples of usage but the scale is larger than campus and the data variability is high.

The four-step model can be used in the campus context, but it needs modifications due to the data type that is obtained in the campus scale. The ground truth data for this research is not available for every step, which means indirect validation needs to be conducted if the four-step model is used. Scenario testing can be implemented for policy and infrastructure planning. Nevertheless, the process should be repeated from the trip generation since the model needs calibration all over again. Hence the flexibility of the model is quite low if the sensitivity analysis needs to be conducted. Despite that, with modification, the model can still be used for the campus scale with the usage of the specific variable for the generation and trip distribution. It needs to be noted, that the trip distribution needs to consider another factor besides the conventional cost definition in transport like travel time or distance.

The direct demand model used in the literature is mainly used to see the importance of the variables and then the link level volume is estimated based on some sort of simulation process (Dadashova et al., 2020; Lin & Fan, 2020; Lu et al., 2018). It makes the direct demand model need to be contextualized for the current research context since the models are applied to the link in the city scale, not to the OD pairs. Even more, one of the models only gives a latent volume value that does not represent the link flow volume on the real network (Lin & Fan, 2020). However, the model can be contextualized by applying direct demand models to the OD pairs and the Monte Carlo simulation is used to generate the number of people in each OD pair.

The other methods might be intriguing if those are implemented to the model. Machine learning can be implemented in the trip generation process (Baumanis et al., 2023). The principal component analysis is interesting to be implemented. These two machine learning are interesting because there might be factors that could not be explained by the traditional variables. Prevailing, the four-step model and the direct demand model is modified to be used in the research accordingly.

Reflecting on the applied data in existing studies is essential to contextualize the data used in this research. As noted by Bhowmick et al. (2023), understanding the types of data utilized in previous publications helps in identifying relevant variables. In this discussion, the data is categorized into supply-related and demand-related types. This classification serves as the foundation for building hypotheses about how each variable should behave within the model. Such a framework can later support more strategic planning, particularly in the development of bicycle networks. For instance, understanding which variables significantly affect supply or demand allows for data-driven decisions—such as evaluating whether current bike parking facilities are sufficient when student admissions increase, or predicting how expanding bike parking might impact overall traffic patterns.

Based on the formulated hypothesis, supply-related data includes infrastructure and network characteristics. Demand-related data encompasses GPS, Strava, or mobile signal data, as well as weather data, travel surveys, traffic flow data, existing Origin-Destination (OD) data, Points of Interest (POI), bike count data, census data, and land use information, as summarized in Table 2.4.

Based on the formulated hypothesis, supply-related data includes infrastructure and network characteristics. Demand-related data encompasses GPS, Strava, or mobile signal data, as well as weather data, travel surveys, traffic flow data, existing Origin–Destination (OD) data, Points of Interest (POI), bike count data, census data, and land use information, as summarized in Table 2.4.

Supply-related data refers to the physical and structural aspects of the transport system that support actual traffic and mobility. This includes infrastructure and network characteristics, which can be directly observed and are indicative of the system's capacity and functionality.

Demand-related data, on the other hand, captures the activities and needs that drive individuals to travel. This category includes data that reflect human behavior and decision-making, such as GPS traces, Strava activity, mobile signal data, travel surveys, and traffic flow patterns. It also encompasses contextual factors like existing OD data, POIs, bike counts, census information, and land use, all of which help explain where, why, and how people move.

While it may seem counterintuitive to classify weather data as demand-related, it is important to note that weather can significantly influence travel behavior. For instance, weather conditions often affect an individual's mode choice, which in turn is reflected in observed demand. Therefore, weather data plays an important role in modeling demand, particularly in active transport modes such as cycling.

To determine the most suitable method based on the literature, it is crucial to first identify the purpose and goal of the analysis. Since the objective of this research is to develop an OD matrix, the references in Table 2.3 that focus on campus-scale OD matrices or have the potential to capture this aspect are specifically evaluated. Additionally, studies with methodologies that could be applicable in future research—when data collection is expanded, and the scope is enlarged—are also considered for their relevance in supporting long-term model development.

2.5.1. Applicability for OD Matrix Estimation of the Bike Network

These aspects will be evaluated by checking the references listed in Table 2.3. The selected references include Hamad and Obaid (2022), Jacyna et al. (2017), Khan and Habib (2023), Krishnakumari et al. (2020), Jacyna et al. (2017), Lin and Fan (2020), Lu et al. (2018), Orvin et al. (2021), Oskarbski et al. (2021), Ryu (2020), Ryu et al. (2019), and van Dulmen and Fellendorf (2021).

From the review of these references, it is evident that the four-step model aligns well with the research context. This confirms that the four-step model is suitable for OD matrix estimation, with the process consisting of trip generation and trip distribution, leading to the OD matrix as the final product.

It is also evident that the references listed cannot be directly implemented without modifications. Even when a reference fully aligns with the research purpose, additional information is required to adapt the methodology effectively.

2.5.2. Applicability for the Campus Scale

The evaluation of relevant methodologies is conducted by analyzing the references listed in Table 2.3, including Arellana et al. (2020), Baumanis et al. (2023), Guo et al. (2013), Hamad and Obaid (2022), Kaziyeva et al. (2021), Khan and Habib (2023), Krishnakumari et al. (2020), Lin and Fan (2020), Lu et al. (2018), Orvin et al. (2021), Ryu (2020), and van Dulmen and Fellendorf (2021).

While these references demonstrate applicability to campus-scale studies, differences in perspective must be acknowledged. Many of these studies were not originally designed for campus-scale applications, and those that do focus on campuses often incorporate multimodal or car-based models rather than being specifically tailored to bicycle travel. As a result, the variables considered in these models may not align directly with the research objectives of this study.

Additionally, some campus-scale models appear to be extensions of broader city-scale models rather than being dedicated solely to the unique dynamics of university travel, as seen in Hamad and Obaid (2022). However, the work of Baumanis et al. (2023) provides a particularly relevant approach for this research. Although their study does not follow the four-step model used in this research, it aligns well with the challenges posed by data limitations. Their work underscores the importance of identifying specific, data-driven variables that fit the available dataset rather than relying exclusively on conventional variables. This reinforces the need to adapt existing methodologies to suit the specific research context and constraints.

2.5.3. Initial Method Proposal

The selected methodology, based on the previous analysis, is the four-step model, which is used to capture the OD matrix. Additionally, there is potential to extend the model by incorporating trip assignment to enhance its applicability. The four-step model provides flexibility in analyzing the most influential variables affecting travel demand on campus, particularly given the dynamic nature of its socio-demographic characteristics.

Due to the specific conditions of the campus environment, it is necessary to implement variables that accurately capture travel behavior. As seen in the previous section, Baumanis et al. (2023) applied linear regression in a campus-scale context during COVID-19, demonstrating the relevance of such an approach for this study.

For trip generation, Ordinary Least Squares (OLS) regression is employed, as seen in Baumanis et al. (2023), to estimate production and attraction volumes. However, for trip distribution, conventional impedance-based models are not applicable due to the nature of TU Delft's campus travel patterns. Since destination choices are primarily driven by activity rather than proximity, the traditional impedance-based trip distribution method does not align with the study's context. As a result, an alternative method needs to be introduced to better represent the spatial interactions within the campus.

Table 2.3: Literature review: methods characteristics

Author (Year)	Methods	Scale	Applicability (Campus scale)	Mode	Applicability (Bicycle)	OD matrix estimation	Applicability (OD matrix)	Travel demand prediction	Dynamic/fine temporal resolution	Policy or infrastructure evaluation	Ground truth data	Specific variables	Validation / evaluation method
Arellana et al. (2020)	Direct demand	City	✓	Bicycle	✓	×	×	✓	×	✓	✓	✓	GEH
Baumanis et al. (2023)	Machine learning (LASSO & OLS)	County	✓	Bicycle	✓	×	×	✓	×	×	✓	✓	Mean Squared Error (MSE)
Blume et al. (2022)	Bayesian model of the nodal-in and outflow	Test network and City scale	×	Public Transit	×	✓	✓	✓	×	×	✓	×	-
Cooper (2018)	Spatial network analysis	City	×	Bicycle	✓	×	×	×	×	✓	✓	✓	-
Dadashova et al. (2020)	Direct demand	City	×	Bicycle	✓	×	×	×	×	×	✓	✓	comparison with the traffic count
Gosse and Clarens (2014)	Four-step model	City	×	Bicycle	✓	×	×	×	✓	×	✓	×	-
Guo et al. (2013)	Agent-based model	Campus	✓	Car	×	×	✓	×	✓	✓	✓	×	Absolute Error and the Absolute Percent Error (APE)
Hamad and Obaid (2022)	Four-step model	Campus	✓	Multimodal	✓	✓	✓	✓	×	✓	✓	×	GEH
Jacyna et al. (2017)	Four-step model	City	×	Bicycle	✓	×	✓	✓	×	✓	×	×	Comparing with the travel count
Jafari et al. (2024)	Agent-based model	City	×	Multi-modal	✓	×	×	×	✓	×	✓	×	Weighted Absolute Percentage Error (WAPE)
Kaziyeva et al. (2021)	Agent-based mode	City	✓	Bicycle	✓	×	×	✓	✓	✓	✓	×	RMSE & Pearson with the traffic count
Khan and Habib (2023)	Simulation	City	✓	Multi-modal	✓	✓	✓	✓	✓	×	×	×	-
Krishnakumari et al. (2020)	Unsupervised Learning (Principal Component Analysis)	Small network and City	✓	Car	✓	✓	✓	×	✓	×	✓	×	mean absolute percentage error (MAPE)
Lin and Fan (2020)	Direct demand	City	✓	Bicycle	✓	×	✓	×	×	✓	×	✓	MLE (from probit model)
Lu et al. (2018)	Direct demand	University town	✓	Bicycle and pedestrian	✓	×	✓	×	✓	×	✓	✓	R^2

Author (Year)	Methods	Scale	Applicability: Campus scale	Mode	Applicability (Bicycle)	OD matrix estimation	Applicability (OD matrix)	Travel demand prediction	Dynamic/fine temporal resolution	Policy or infrastructure evaluation	Ground truth data	Specific variables	Validation / evaluation method
Orvin et al. (2021)	Direct demand	City	✓	Bicycle	✓	×	✓	✓	✓	✓	✓	×	
Oskarbski et al. (2021)	Four-step model	City	×	Mixed (with bicycle layer)	✓	×	✓	✓	×	✓	✓	×	
Ryu et al. (2019)	Two-stage OD estimation	Campus	✓	Bicycle	✓	✓	✓	×	×	×	×	×	
Ryu (2020)	Two-stage OD estimation	City	×	Bicycle	✓	✓	✓	×	×	×	×	×	
van Dulmen and Fellendorf (2021)	Four-step model	City	✓	Multi-mode	✓	✓	✓	✓	×	✓	×	×	-

Table 2.4: Literature review: data review

Author (Year)	Supply Representing Data		Demand Representing Data								
	Infrastructure data	Network characteristics data	GPS/ Strava/ Signal/ Social Media	Weather Data	Travel survey data	Traffic flow data	Existing OD	POI data	Bike count data	Census data	Landuse data
Arellana et al. (2020)	✓	✓	×	×	✓	✓	×	×	✓	✓	×
Baumanis et al. (2023)	×	×	×	✓	×	×	×	✓	✓	✓	×
Blume et al. (2022)	×	×	×	×	✓	✓	✓	×	×	×	×
Cooper (2018)	×	×	×	×	×	×	×	×	✓	×	✓
Dadashova et al. (2020)	✓	×	✓	×	✓	×	×	×	✓	✓	×
Gosse and Clarens (2014)	✓	✓	×	✓	×	✓	✓	×	✓	×	×
Guo et al. (2013)	✓	✓	×	×	✓	✓	×	×	×	×	×
Hamad and Obaid (2022)	✓	✓	×	×	✓	✓	×	×	×	×	✓
Jacyna et al. (2017)	✓	✓	×	×	×	×	×	×	×	✓	×

Author (Year)	Supply Representing Data		Demand Representing Data								
	Infrastructure data	Network characteristics data	GPS/ Strava/ Signal/ Social Media	Weather Data	Travel Behaviour data	Traffic flow data	Existing OD	POI data	Bike count data	Census data	Landuse data
Jafari et al. (2024)	✓	✓	×	×	✓	✓	×	✓	✓	✓	✓
Kaziyeva et al. (2021)	✓	✓	×	×	✓	✓	×	×	✓	✓	×
Khan and Habib (2023)	×	×	×	×	✓	×	×	✓	×	×	×
Krishnakumari et al. (2020)	✓	✓	×	×	×	✓	×	×	×	×	×
Lin and Fan (2020)	✓	✓	×	×	×	×	×	×	✓	✓	×
Orvin et al. (2021)	✓	✓	×	✓	×	✓	×	×	✓	✓	✓
Oskarbski et al. (2021)	✓	✓	×	×	×	✓	×	×	✓	✓	×
Ryu et al. (2019)	✓	×	×	×	✓	✓	×	×	✓	×	×
Ryu (2020)	✓	✓	×	×	✓	✓	×	×	✓	×	×
van Dulmen and Fellendorf (2021)	×	✓	×	×	✓	✓	×	×	×	✓	×

3

Data Description and Method Framework

3.1. Data Description and Analysis

There is a circumstance in the research that needs to be noted, the data availability must be explored before the methodology can be determined. Adding to that, the study is limited in the execution time, hence, the data collection has to be conducted as efficiently as possible. In this research, the data collection is mainly done by obtaining from another party rather than included in the research scope. This makes the data availability depend on the TU Delft campus activity in data collection around the campus authority and entities.

The data discussed in this research includes both travel supply and demand-related variables, as well as supporting data sources. The supply and demand data are gathered from within the TU Delft campus environment, reflecting characteristics such as population, infrastructure capacity, and land use. In addition, supporting data is incorporated to enrich the analysis and provide broader context. This includes travel behaviour data collected by Statistics Netherlands (CBS), weather data from the Royal Netherlands Meteorological Institute (KNMI), and traffic count data obtained from the Urban Mobility Observatory at TU Delft. Together, these datasets form the foundation for the travel demand modeling conducted in this study.

3.1.1. Supply-related Data

The basic data that needs to be obtained are the bicycle network's supply and demand. In terms of supply, the data from the available open-source network such as Openstreet Map could represent the supply of the network. Furthermore, the network characteristics such as the designed speed, the lane width, and the capacity of each lane can be generalized from the design standard of the cycleway in the Netherlands. The data is described in the following subsections. The data is explored based on the study in Chapter 2. The related variables that might be needed in the travel demand analysis were extracted by reflecting on the TU Delft-related data. The available data was mainly obtained by request but some data is open-sourced and can be obtained freely under specified license. For the trip generation, supply also can be expressed in the model. This section is the explanation of the used supply expressing data.

The data will be described based on its relevance to the analysis, including the rationale for its inclusion, the key travel demand or behavior insights it provides, its implementation within the study, and potential ways to enhance its usability. Additionally, other potential data sources that could further support the analysis may be identified.

3.1.1.1 Lecture and Exam Hall

The demography of the TU Delft campus is certainly dominated by student activities as discussed in Section 3.1.2.1. Hence, the travel demand is mainly motivated by students. The students mostly come to the campus for the sake of attending lectures. This means that determining the pattern of lecture activity is important to estimate the travel demand. The lecture activity itself is also controlled by the spatiotemporal dimension. The travel demand pattern can be calculated if this spatiotemporal factor of the lecture activity can be captured. To support this, the lecture hall capacity could be insightful information since it can describe the demand that can be raised but also the ability of the building to provide service for the lecture activity. For example, a big lecture hall could attract more students to the building since the hall is more flexible in terms of capacity. A specific purpose might be needed such as exams or practical learning with computers and the building with these services is preferable as the activity location. When these capacities are introduced as input into the analysis, it can help to describe the dynamic of the temporal context of the estimation. From a practical perspective, the lecture, exam, and computer capacities can be included as variables with an additional dummy variable of the exam period to emphasize the time context and put importance on the exam capacity. This capacity information can be extracted from the EsViewer website (https://esviewer.tudelft.nl/) provided by the TU Delft campus. These potential variables are depicted on Figure 3.1 and Figure 3.2.

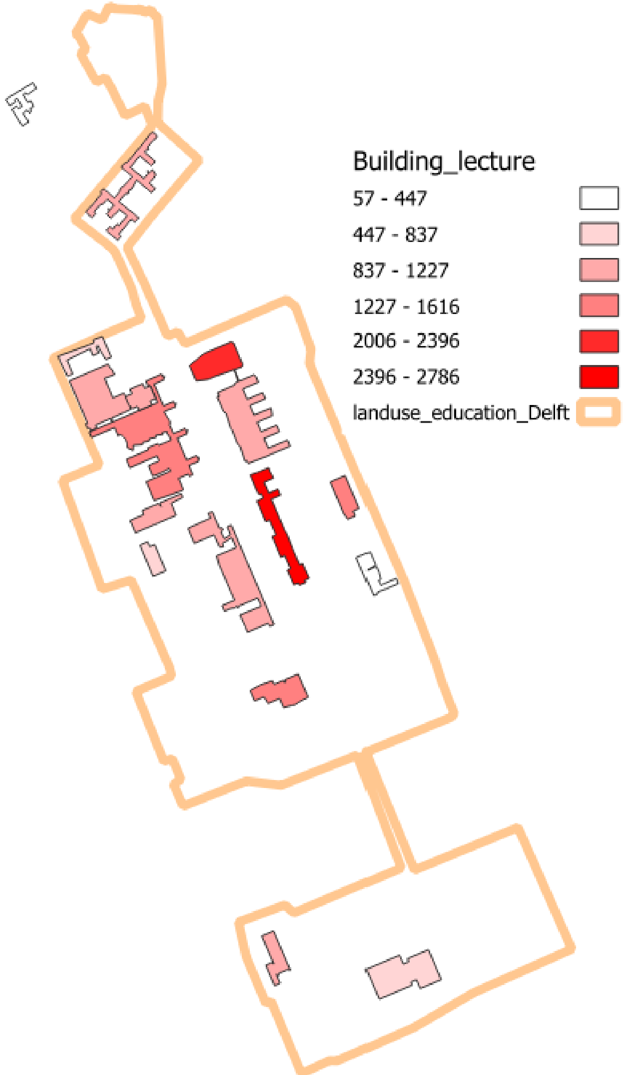


Figure 3.1: TU Delft sampus lecture hall capacity (Seats)

Building Name	Building Number	Seat Capacity	Exam Capacity	Computer Capacity
ABE (Arch)	08	876	0	0
AULA	20	2128	0	0
AP	22	1074	348	71
DUWO	221	100	0	0
CEG	23	2786	571	243
Bouwcampus	26	57	38	0
ECHO	29	1328	144	0
TPM	31	631	278	72
IDE	32	923	354	228
PULSE	33	1020	330	0
ME	34	1302	272	0
DW	35	1136	807	495
EEMCS	36	1112	40	0
X	37	1499	1499	0
FLUX	39	730	365	0
AS (58)	58	531	0	0
AE	62	1071	296	96
FeLS	66	1030	337	156

Figure 3.2: Number of lecture hall

The figure shows that the CEG building contains a large number of lecture seats that can be translated as its capacity for holding the learning activity. Its location also supports its attractiveness, even though for the short-distance haul, it is not relevant. ECHO and PULSE are attractive because of the size of their lecture halls. For exam purposes, Building 35 on Cornelis Drebbelweg and X are often used due to their large exam capacity. Even, X does not allow other activities during the exam period. It positively draws more traffic to the building. It is also worth noting that the Applied Science Building in the TU Zuid does not have exam capacity. It might be because of the data faulty or the number of students, but presumably the students might need to move to another building for the exam and probably induces another origin-destination relationship within the campus area.

The model can incorporate lecture hall capacity as a dependent variable. To capture the influence of lecture and exam periods more accurately, the capacity variable can be differentiated between lecture, exam, and computer lab capacities. Since these variables may show linear relationships and be correlated, the modeling process will eliminate unnecessary distinctions while keeping relevant differentiations to improve the model's accuracy.

The lecture activities can be included in the analysis to provide the time influence on the model/estimation. The lectures can explain the reason behind the peak hours or volumes (if they exist) more clearly. In this case, the data regarding the lecture activities is not recorded in one provider. This is because the timetable service is controlled in a decentralized manner by the faculties. More detailed temporal data can be obtained by occasionally recording the number of people who use the building or lecture hall. The method to estimate the number of people in the building can be the same as the Spacefinder used to give indicator information of the study space occupancy. The method is discussed in Section 3.1.1.2.

3.1.1.2 Study Spaces and Seats

Undoubtedly, the students mainly attend the university for the lectures, but the individual and group studies also influence students' movement to and inside the campus. Furthermore, the study place is surely relevant as an indicator of demand but can also express the capacity of the building to contain the activity of the students. The number of study places expresses the demand because it can describe the attraction of a building for the student. The study place is not related directly to the faculty of the student, thus the origin-destination relationship between each building might be found based on this study place and it justifies the interzonal trip. This pattern can be seen by looking at the buildings not related to the faculty often full of lectures and individual study activities. This building's occupancy and capacities are depicted on the Spacefinder website (www.spacefinder.tudelft.nl).

Based on the Spacefinder, the number of spaces (indicated by rooms) and the number of seats are depicted. The occupancy rate of the buildings is also conceived in the indicator. However, the detail of the indicator is not fine because the method to estimate the occupancy is still not accurate. The Spacefinder team clarified that the team uses the wifi to determine the occupancy rate. Due to technical limitations, the indicator is not detailed to the level at which rooms are more or less occupied and the timestamp of the occupation is not recorded. However, the team still develops the indicator to provide more useful information. The location and the number of seats and spaces for study activities from the Spacefinder are depicted on Figure 3.3 and Figure 3.4.

The Figure 3.3 shows the distribution of the study spaces based on its seats within the TU Delft campus recorded by the Spacefinder. It can be seen that the library is located in the center of the campus and provides a large number of seats to study. It certainly attracts many students to travel to the library for studying activities other than their faculty building. Even though the distance can be assumed to be irrelevant for short-distance travel, the location of the library might affect the attractiveness of the library for the student in this case. The distribution of the study is heavily centered on the TU Midden area. Thus, the bike traffic of ABE buildings and TU Midden might be slightly influenced by the study space capacity. There is a small possibility for the students to come to those buildings solely to study, supposing that the program-related reason gives more influence to motivate the students to come to the TU Zuid rather than study.

It is also interesting to see that there is potential based on the capacity of the IDE building. The study seats of the IDE building are quite high according to Figure 3.4, yet the student number of the IDE faculty is the lowest of all the faculty. There is a possibility that the IDE building could attract another faculty-student solely for studying. It also can be the planning consideration for the stakeholder to distribute the student for individual or group study activities.

The Spacefinder already gives a good overview of the capacities of each building based on the study spaces and seats. In retrospect from prior discussion, the limitation of the occupancy indicator in Spacefinder can be improved since the method to detect the business of the building is rooted to the wifi connection. It means a more detailed record of the building occupancy can be obtained. The time-stamped occupancy will be

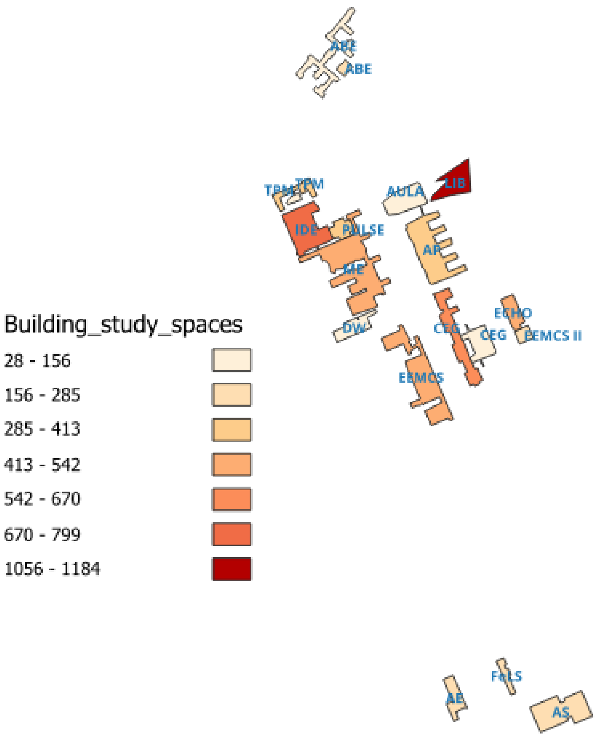


Figure 3.3: TU Delft campus study spaces (Seats)

Buildings	Seats	Spaces
AE	204	5
AP	374	17
AS	215	12
ABE	305	4
AULA	119	4
CEG	646	24
EEMCS II	199	10
DW	58	2
ECHO	422	9
EEMCS	486	11
IDE	689	10
LIB	1184	34
ME	421	24
PULSE	407	8
TPM	465	22
FeLS	228	5

Figure 3.4: Number of study spaces

useful data to be acquired from the transport model perspective since the data would give insight into the travel pattern and can be an input for the dynamic model.

3.1.1.3 Bike Parking Spaces

The bike parking spots are very important to support the cycling culture of the Netherlands and represent the supply of the infrastructure to enhance the motivation to ride bicycles. Hence, bike parking spots can describe the ability of the building to provide conformity for the cyclist, thus indirectly showing the attractiveness of the building to the cyclist. It means the bike parking spot can be included as a variable for the travel demand estimation. This suggestion is strengthened according to Heinen et al. (2011), "bicycle commuting is influenced by attitudes on direct benefits, awareness, and safety". Even though the research is conducted within the context of the mode choice, it can express the importance of the infrastructure that could enhance the direct benefits of bicycle commuting.

The data on the bicycle parking capacity is acquired from the Campus Real Estate and Facility Management Department of TU Delft Campus. The CREFM parking team gives the GIS-based map, then reprocesses to the new shapefile attribute for further analysis and depicted on Figure 3.5. The parking capacities are represented by the number of racks on the figures.

For the context of origin-destination estimation, there are possibilities for the cyclist to choose a location based on its impedance. The parking spot with covers might be more desirable and attractive. Even though,

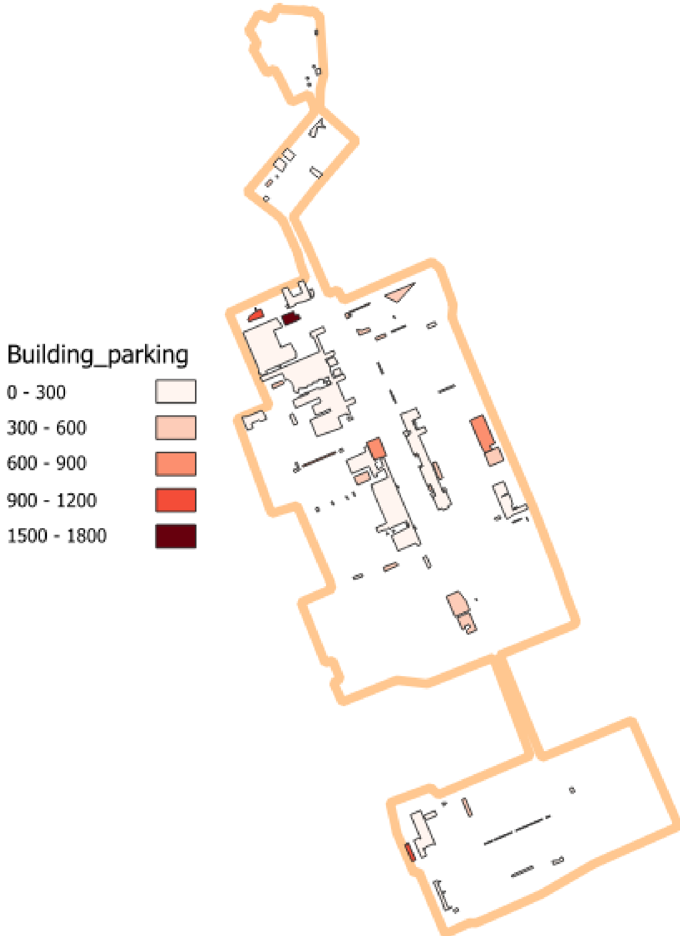


Figure 3.5: TU Delft Campus bike parking spot

the distance to the main activity location is still preferable. Accordingly, the parking spots in TU Zuid and TU Noord areas can be assumed to be provided only for those areas. However, the parking spots in the TU Midden might follow the concept of parking area conformity. For example, people may choose to park their bikes around the CEG building and go to the ECHO building because the ECHO building has a basement parking area that makes people put more effort into parking their bikes. Based on its amount, the parking area close to the IDE and PULSE buildings contains the most parking spot racks. It can be translated into the centrality of the location as mentioned, but this area provides a large number of study spots as discussed in Section 3.1.1.2. It is almost certain that bike parking and study places are important in determining the travel demand. On the other side, it might be those variables are correlated or related and it can be proved in the process of modelling the OD matrix.

The parking spot occupancies with the temporal dimension detail can be insightful information for the travel demand analysis since it shows the rate of activities and the dynamic of cyclist demand. Still, parking spot occupancies are not depicted in the stakeholders' acquired data. It is not also indicated from the GIS-based like the Spacefinder. In terms of recording the exact data of the parking occupancies, additional sensors on each bike can provide an accurate record and can be customized to print the timestamp of the occupancy. Regardless, the amount of effort and equipment translated into the cost will be skyrocketed. Furthermore, to handle this condition, the traffic count sensors can be installed at the nearby entrance to record the number of bikes, although it is not detailed enough to get the occupancy directly.

3.1.2. Demand-related Data

The demand data that would express the possible trip generation or area attractiveness are gathered to express the trip generation of the campus area. The general data of the campus such as the number of students, the number of employees, the number of study spaces, and bike parking capacity, according to the respective buildings or zones might be useful for expressing the demand for the bike network. This subsection describes the data that are related to the demand of bike.

3.1.2.1 TU Delft Population

The number of individuals associated with the TU Delft is analysed since these individuals are more likely to come to the TU Delft than the visitors. The number is assumed to express the basic demand of the TU Delft transport network. The associated individuals that can be recognized in the research are the students and employees and then categorized to see any pattern in travel behaviour. It is known that the number of students who are enrolled in the 2023-2024 academic year and the number of employees of TU Delft is 24,720 and 6,329, respectively. The students and employees have different behaviour and activity. The lecture and academic reasons will control the activity of the students without any consequences if the attendance at the campus is not fulfilled. On the other hand, the schedule of the employees might be more fixed and have consequences if the working hours are not satisfied, It might result in different arrival and departure times. It is also important to note that the difference in quantity between students and staff is quite large, which could allow the students to control the traffic of the bike network. The population of the TU Delft campus is depicted in Table 3.1.

Table 3.1 describes the proportion of the population categorized by the faculty. It can be seen that the proportions for some faculty are quite similar. The Industrial Design Engineering (IDE) is the smallest faculty, yet, it is located in the heart of the middle campus area and connected to the Mechanical Engineering and PULSE building. It made the attraction in the IDE building still the same as the other faculty. It should be assumed that all faculties have quite the same potential in terms of travel demand. The trip distribution between each faculty is assumed to be evenly divided. Nevertheless, this assumption can be ignored if the entrance and exit points to the campus of the cyclists can be determined since the entrance/exit points could be included in the zones and will be the origin point of the cyclist in the analysis scope.

Table 3.1 describes that there are quite dominant characteristics in the population. The scientific and support staff would have a different schedule and behaviors, such as arrival and departure times, vehicle use, movement amount, etc. Regardless, the scientific staff has a bigger proportion in each faculty which means the group might dominate the traffic of the people. It can also be seen on the university scale, that the scientific staff has a larger number than the support and management staff, generally. The method can decide whether to include the staff category as the variable or not. There are some cases in which the analysis needs to include as many variables as possible but can be discarded later if the variable is not significant, like in the backward-step regression. Hence, the students also can be divided to recognize the pattern within this behavior. The MSc and BSc students are supposed to have different schedules since these two programs cannot occupy the exact space and time simultaneously. It is safe to assume that those students would reflect different patterns of activity. Additionally, the ratio between MSc and BSc students in each faculty is quite close. It means there is no dominant group that controls the traffic behaviour of the students. The programme might be meaningful to be included as the variable. Furthermore, the gender proportion is also analysed and depicted on Table 3.2.

Overall, males dominate the population on the TU Delft campus. The male number reached 21,078 making the male 67.9% of the population. Even though, there is a finding that gender gives tendencies to cycling activities depending on the cycling facilities such as protected infrastructure and cycling normalisation (Aldred et al., 2016). However, there is also research by Harms et al. (2014), "gender differences in cycling in the Netherlands are minimal and contrary to findings in other countries where cycling is less popular (and especially women are cycling less often). In the Netherlands women cycle more often than men, especially for trips to and from work and for trips related to shopping". It is shown that gender is not a resistant factor in the choice of mode for the bicycle. It makes the gender irrelevant and most likely does not represent differences in travel behaviour especially on the TU Delft campus.

Since the distance is not relevant, the activity input becomes much more important. Several sources of data might be available but not accessible in the context of this research due to authority restrictions, such as the data on part-time and full-time employees. There are two different patterns between those two and they can be used as the variables. Other data is the employee's attendance record. It can also be simplified to record the arrival time and departure time. Later, it can also be included in the travel survey.

3.1.2.2 Travel Behaviour

For a larger scale, this data would be derived from the travel survey like the survey in the Netherlands. The research, called Onderweg in Nederland (ODiN), has been conducted every year in the Netherlands to de-

Faculty/ corporate office (abbr)	Staff category	Number Employee	Proportion in University (%)	Degree programme type	Number Student	Proportion in University (%)
ABE	Scientific staff	666	10.52	BSC	1348	5.45
	Support and management staff	175	2.77	MSC	1760	7.12
AE	Scientific staff	477	7.54	BSC	1400	5.66
	Support and management staff	116	1.83	MSC	1308	5.29
AS	Scientific staff	661	10.44	BSC	1910	7.73
	Support and management staff	266	4.20	MSC	1044	4.22
CEG	Scientific staff	702	11.09	BSC	1162	4.70
	Support and management staff	202	3.19	MSC	1576	6.38
EEMCS	Scientific staff	942	14.88	BSC	2616	10.58
	Support and management staff	195	3.08	MSC	1854	7.50
IDE	Scientific staff	277	4.38	BSC	992	4.01
	Support and management staff	78	1.23	MSC	937	3.79
ME	Scientific staff	713	11.27	BSC	2621	10.60
	Support and management staff	166	2.62	MSC	2270	9.18
QT	Scientific staff	151	2.39	BSC	-	-
	Support and management staff	72	1.14	MSC	-	-
TPM	Scientific staff	396	6.26	BSC	877	3.55
	Support and management staff	74	1.17	MSC	1045	4.23
	Total	**6329**	**100.00**		**24720**	**100.00**

Table 3.1: Population proportion for each faculty

Faculty/corporate office	Gender	Gender Percentage within Faculty(%) Student	Staff
ABE	Female	56.82	44.83
	Male	43.18	55.17
AE	Female	14.77	26.48
	Male	85.23	73.52
AS	Female	34.70	41.42
	Male	65.30	58.58
CEG	Female	29.07	35.29
	Male	70.93	64.71
EEMCS	Female	17.38	27.70
	Male	82.62	72.30
IDE	Female	56.56	50.14
	Male	43.44	49.86
ME	Female	24.29	26.28
	Male	75.71	73.72
QT	Female	NaN	24.22
	Male	NaN	75.78
TPM	Female	34.29	54.04
	Male	65.71	45.96

Table 3.2: Gender proportion in each faculty

termine the travel behaviour of the Dutch. The research provisions sufficient information on the daily mobility of the Dutch. It comes with origin-destination pairs, time of departure, and the travel purpose (Informatiepunt & Centraal Bureau Voor De Statistiek, 2023), along with the information on the traveler. It makes ODiN perceived as the representing data for the Dutch population's travel behaviour and has adequate information to build models around it. However, this fanciness does not happen in the situation of a small scale. The ODiN data only provides details at the postal code level 4 for the address and the origin-destination pairs. Nevertheless, the ODiN data is still useful in this research. The ODiN data can be used as the basis of the modal split assumption. Before understanding the traffic count data, the data is mixed between pedestrians and cyclists at some parts. There are also traffic count at the entrance of the IDE building. It is believed to be the production trip that is produced by the building but the mode that are used by the people are not defined and will be assumed based on the modal split in the ODiN data.

The production also will be assumed to be related to home. It means the PC4 postal code will be discussed in travel demand estimation as the production factor from external. The mode and departure point of the external zone is depicted in the Chapter 4. The departure time and activity duration is depicted on Figure 3.6 and Figure 3.7. It is shown that the people who come to the PC4 around TU Delft act like student start from the starting departure time matches with the first class time and activity duration that is related to be the trip production. It is shown that the people who come to the PC4 around TU Delft act like student start from

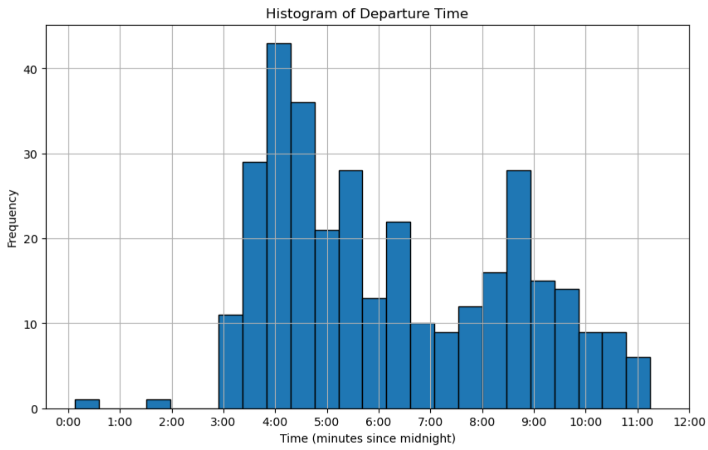


Figure 3.6: Departure time of PC4 related to the TU Delft Campus

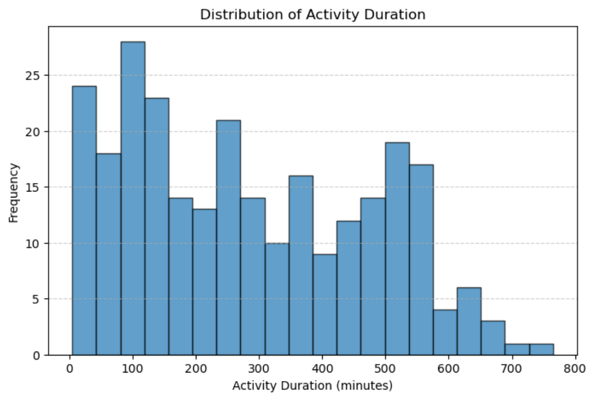


Figure 3.7: Activity Duration

the starting departure time matches with the first class time and activity duration that is related to be the trip production. Since the match between activity and student is drawn, to capture the representative pattern of the TU Delft student, the ODiN data analysis will be focused on the related PC4 that indicates TU Delft as the destination.

The analysis of ODiN data is used to determine modal split factors and the distribution of departure points within the TU Delft area. The analysis is depicted on Figure 3.8 and Figure 3.9. The results indicate that bicycles are the most frequently used transport mode, consistent with the campus's mobility trends. This analysis is conducted based on the PC4 areas that cover the TU Delft campus: 2628 and 2629 (TU Delft Zuid).

Passenger car usage is notably high, which can be attributed to the inclusion of residential areas east of Schoemakerstraat within the PC4 zones. Additionally, the data reveal a significant number of pedestrians,

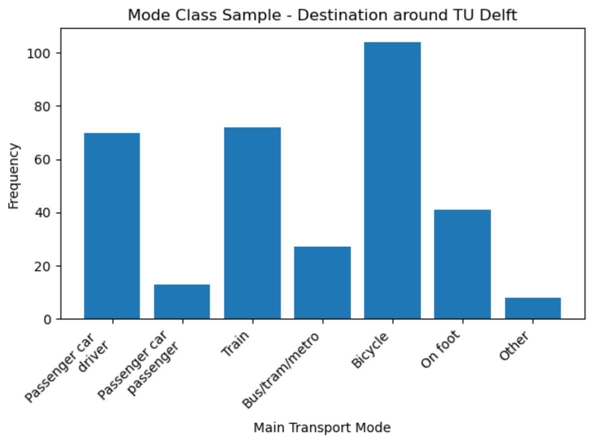
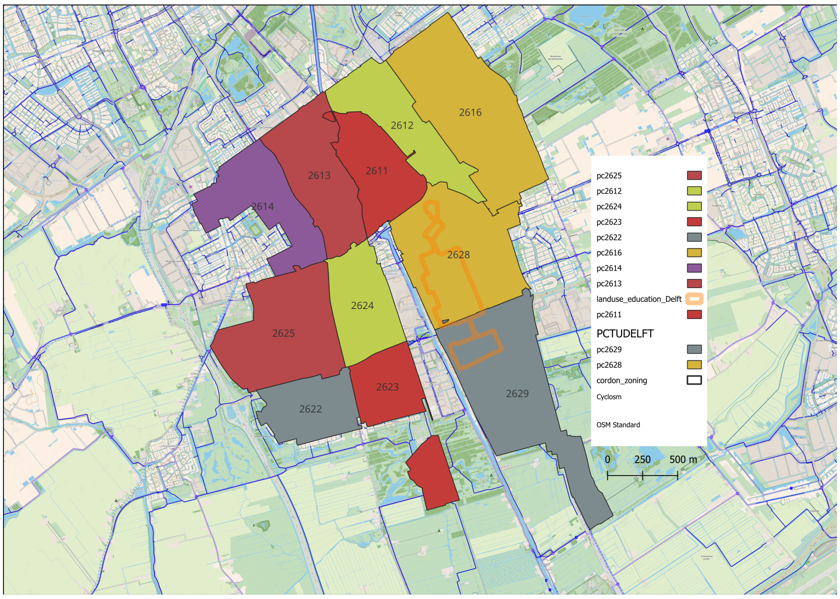
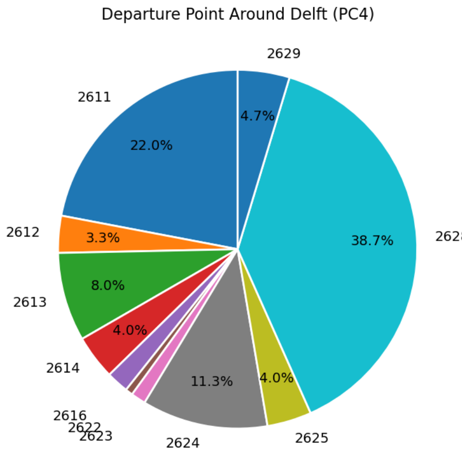


Figure 3.8: Mode Class Proportion



(a) PC4 around TU Delft



(b) Departure Point

Figure 3.9: Mode Split and Location Distribution

reflecting the realistic walking patterns observed in and around the campus. These insights highlight the diverse travel behaviors within the study area.

Train usage accounts for a significant proportion of travel demand due to the wide geographic distribution of TU Delft students. Train users are assumed to primarily rely on bicycles for last-mile travel, and they are included as bicycle users in the modal split analysis. A benchmark of 52.5% bicycle mode share, derived from the ODiN data, is used for trip production and attraction estimations across time categories.

3.1.2.3 Weather Data

The weather is one of the main drivers that motivates or demotivates people from riding the bike. Hence, the weather can indirectly influence the travel demand and it is reasonable to include the weather in the estimation. It is also supported by the existing research. Franco et al. (2014) conducted a set of surveys in Rio de Janeiro to determine the main characteristics of bike trips with the potential demand model based on the behaviour of the cyclist. According to the research, "When asked about the influence of weather conditions on the choice of transport mode, the majority of respondents (64%) stated that when it rains, they do not use the bike, and a small part (4.5%) cited the intense heat as a difficulty, and 31.5% of respondents admitted that no factors affect the use of bicycles for travelling to school or to work. When they cannot use the bike for some reason, they travel, mostly, by bus. The availability of public transport was also questioned and 92.2% of respondents said they have this transportation mode option." Even though, the Netherlands has such an established culture of cycling, sometimes all the hurdles just do not matter. Yet, by referencing this research, the weather is still relatable to the motivation for riding a bike. The weather data is introduced because it can represent variables that change over time except for time itself. Hopefully, by including weather data in the trip production and trip attraction, the dynamic of the bicycle traffic could be captured.

	Daily mean windspeed (in 0.1 m/s)	Daily mean temperature in (0.1 degrees Celsius)	Daily precipitation amount (in 0.1 mm) (-1 for <0.05 mm)	Sunshine duration (in 0.1 hour) calculated from global radiation (-1 for <0.05 hour)
count	244.000000	244.000000	244.000000	244.000000
mean	50.090164	97.057377	36.569672	34.684426
std	22.232777	45.402054	52.724605	32.408574
min	14.000000	-34.000000	-1.000000	0.000000
25%	32.000000	72.000000	0.000000	5.000000
50%	46.500000	96.000000	13.500000	26.500000
75%	63.000000	122.250000	53.250000	57.000000
max	117.000000	201.000000	287.000000	127.000000

Table 3.3: Weather Descriptive Statistics

The weather data is obtained from The Royal Netherlands Meteorological Institute (KNMI). The data provides the average, maximum, and minimum values of the weather measurements on a specified day. The closest weather station to the TU Delft campus is KNMI Station Rotterdam with a distance around 6 km from the campus. It is considered close and the data from the station can be assumed to represent the condition of the TU Delft campus without other supporting station data. The daily mean windspeed, daily mean temperature, daily precipitation, and sunshine duration is used as the variable in the trip generation.

3.1.2.4 Traffic Count of Bike and People

The traffic of the bicycle is important as the basis of the model. Either it is used as a ground truth, or being trained in the model to produce an estimation. The TU Delft campus has the Urban Mobility Observatory (UMO) project that aims to provide new traffic data for the researcher. One of the projects of UMO is the Campus Mobility Dashboard which was built during the COVID pandemic to estimate the travel demand during the lockdown and how it will grow. It needs to be taken that the UMO has a large dataset that contains not only the bicycles but also the people and parking areas in the city center. For the TU Delft campus, smart cameras were installed in the heart of the campus to record the movement and traffic of the campus entities. The locations of the smart cameras are depicted on Figure 3.10. It can be seen that the sensors are mainly located on the Mekelweg where the traffic is centralized.

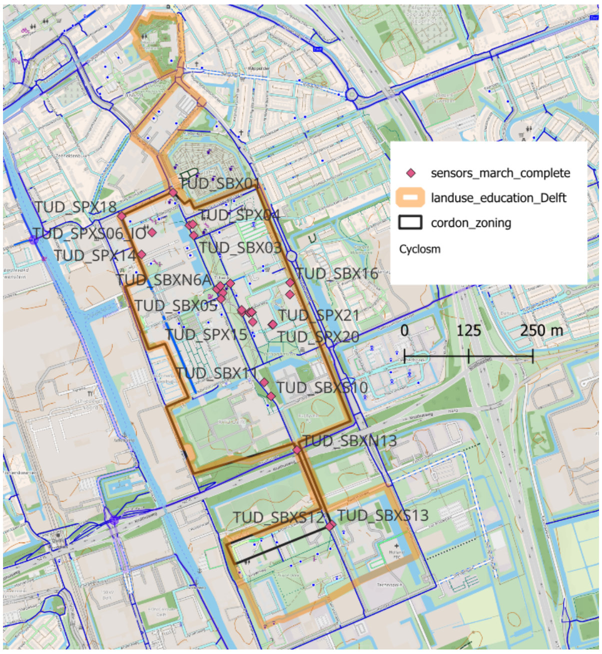


Figure 3.10: Sensor location (blue lines are the bicycle network)

The smart cameras are not only located in the Mekelweg but also in several places that capture the important pattern of the campus activity. The smart cameras with identification "TUD_SBX01" is located at the Jaffalaan intersection, This intersection is very crucial as the presumably source of bottleneck because there are merging flows from two directions and also there is a possible conflict between people from Mekelweg who want to move to Jaffalaaan. The smart cameras were also positioned right before the ECHO parking area at the basement entrance to purely calculate the number of bicycle entrances and exits. However, since the traffic analyzed in this study is exclusively part of the bicycle network, the traffic count data used for model specification is strictly limited to bicycle movements. This ensures that the model accurately reflects cycling patterns without interference from other modes of transport.

The graphs can be seen on the Figure 3.11. The graph illustrates the daily bicycle traffic at the Jaffalaan intersection, highlighting variations in traffic volume across different days within the observed period. It reveals fluctuations in daily bicycle traffic, indicating that travel demand is not constant. Additionally, distinct differences in traffic patterns between lecture periods, exam periods, and holidays are observed, which will be incorporated into the model. Furthermore, the analysis extends to a finer temporal resolution to better capture the detailed patterns of bicycle movement.

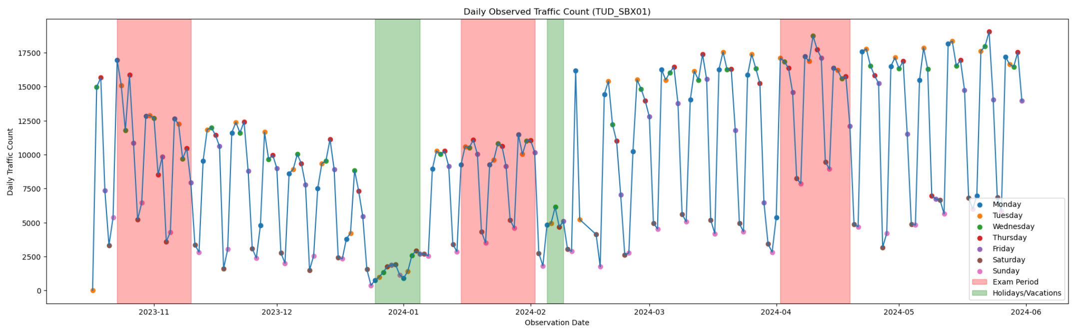


Figure 3.11: Traffic counting analysis TUD_SBX01 sensor: Jaffalaan intersection

Figure 3.12 presents the weekly traffic fluctuations at the Jaffalaan intersection through a boxplot, real-time traffic count data, and the average daily traffic. The boxplot indicates that while there are no extreme fluctuations between days, a significant amount of noise is present in the data, which could impact the model's accuracy. The real-time traffic count reveals high fluctuations, emphasizing the need for a detailed time resolution in the analysis.

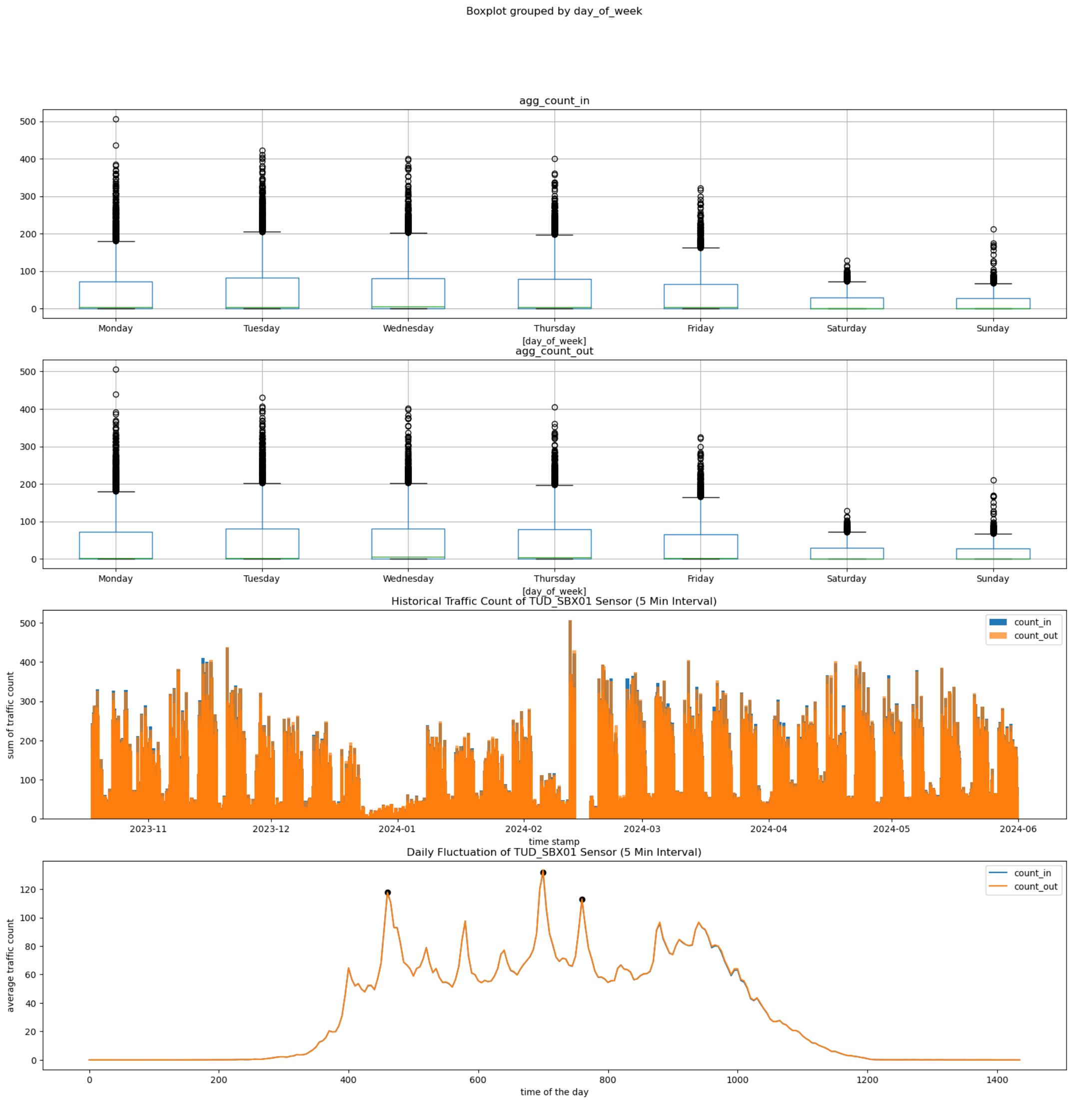
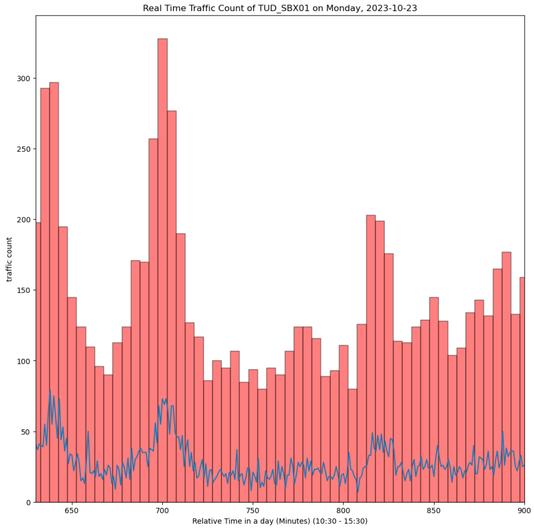


Figure 3.12: Traffic counting analysis TUD_SBX01 sensor: Jaffalaan intersection

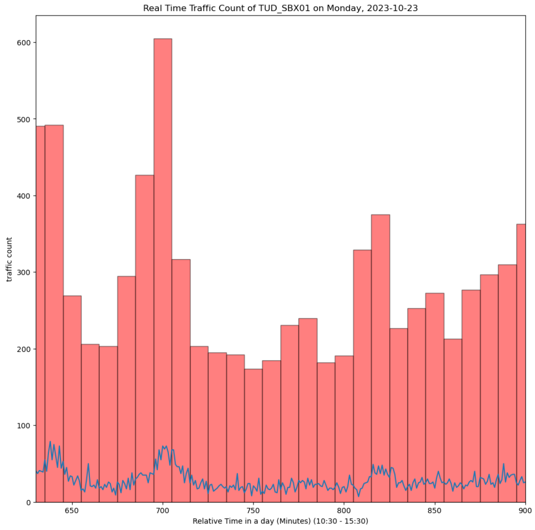
From the average traffic within a day, it is evident that traffic fluctuations align with campus activities, particularly lecture schedules. The start of lecture hours corresponds to increased bicycle traffic, while the highest peak is observed in the middle of the day. This midday peak is likely influenced by break periods when people move for non-academic purposes, such as lunch or informal activities. However, since the midday period consists of mixed travel purposes, it introduces additional uncertainty regarding the factors influencing traffic patterns. Therefore, to ensure a more precise analysis, the focus will be placed on morning and evening peak periods, where travel behavior is more directly related to homebound trip and academic schedules.

Figure 3.13 presents several aggregation plots, illustrating different temporal resolutions of the traffic data. In the visualization, the traffic count is aggregated in the red bars, while the real-time traffic is depicted in the blue plot.

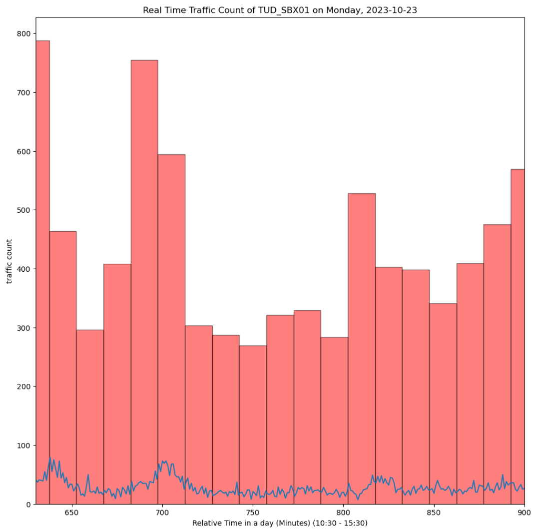
The 5-15 minute aggregation provides a detailed representation of the traffic dynamics, successfully capturing fluctuations and peak periods without excessive data smoothing. Among these, the 5-minute aggregation is selected as the optimal resolution, as it maintains the necessary level of detail while ensuring a manageable computational cost given the volume of data. This resolution effectively balances accuracy and efficiency in the analysis.



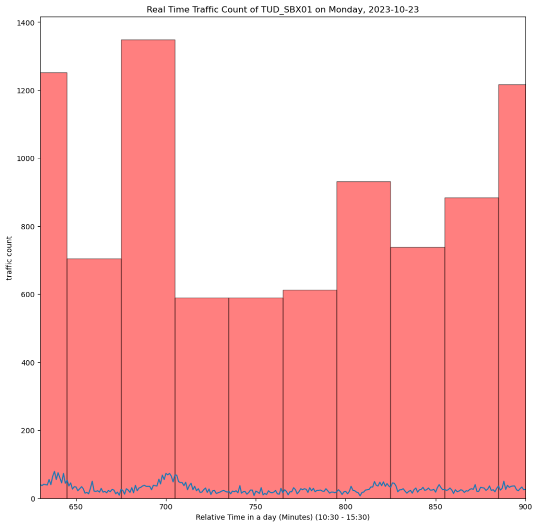
(a) Traffic Count of TUD_SBX01 (5 Minutes Aggregation)



(b) Traffic Count of TUD_SBX01 (10 Minutes Aggregation)



(c) Traffic Count of TUD_SBX01 (15 Minutes Aggregation)



(d) Traffic Count of TUD_SBX01 (30 Minutes Aggregation)

Figure 3.13: 5, 15, 30, and 60 minutes Aggregation Traffic Count Plot

3.1.3. Data Decription Summary

Generally, the data that are gathered do not show any temporal dimension detail. In the case of the weather, the open data from KNMI only provides the daily average, maximum, and minimum values. If the dynamic OD matrix wants to be estimated, the only variable that can be used to represent the dynamic condition is time itself other than the traffic count. The other variables are static and mostly express capacity rather than the activity of the people. From the explanation, it can be concluded that the activity-related data is very important since the literature review indicates that the distance is not heavily related to the trip distribution process on the campus scale. Thus, introducing the variables that can be used in future research and showing their benefits will enrich the knowledge on travel demand estimation with data limitation and small-scale perspective.

This section offers critical insights into the model's specification, confirming the importance of the temporal dimension within the model. The variations noted in bicycle traffic data reinforce the necessity for a dynamic modeling approach that adequately accommodates temporal changes. These data usage, is depicted in the next discussion

3.1.3.1 Data Usage

The dataset includes various variables relevant to modeling bicycle travel demand within the TU Delft campus:

- Lecture and Exam Hall Capacity: Differentiated by lecture, exam, and computer lab capacities, measured in the number of available seats.
- Study Space: Measured in seats to represent available student workspaces.
- Bike Parking Spaces: Categorized into open parking, shared bike parking, and access card-controlled parking, measured in the number of racks.
- TU Delft Population: Differentiated between students and employees. Students are further divided into MSc and BSc groups, as their schedules differ. However, this distinction may later be consolidated into a single category.
- Travel Behaviour Data: Used to inform the modal split assumption. The behaviour of individuals with destinations 2628 and 2629 is analysed to determine the proportion of transport modes, serving as the basis for the modal split.
- Weather Data: Used in nominal values, including daily mean temperature (0.1°C), precipitation (0.1 mm), and wind speed (0.1 m/s) to capture the influence of weather conditions on cycling behavior.
- Traffic Count Data: Used as the target variable for the model. Due to observed fluctuations in traffic counts, data transformation may be required to improve predictability and model performance.

3.2. Methodology Framework

The section explains the methodology of the research based on the current state-of-art review and the framework of the analysis. The previous chapter reveals the indication of the suitable approaches for the research context and the availability of the data and technique.

The four-step model is applied with modifications to better fit the available data. The data used aligns with the proposed method, ensuring compatibility with the OLS regression for trip generation. However, trip distribution must be conducted with limited data that can meaningfully relate to it, as traditional impedance-based methods are not suitable for the campus-scale context.

To address this, the Iterative Proportional Fitting (IPF) method is chosen, as it does not rely on impedance but instead distributes trips based on the interaction between zones. This approach allows for a more accurate representation of campus travel behavior. Additionally, the model specification requires a time interval of 5 minutes, as determined from the time plot analysis, ensuring that the model captures the short-term fluctuations in bicycle traffic effectively.

3.2.1. Data Preprocessing and Assumption Formulation

The data needs to be prepared for the estimation process. In the context of the four steps, it means the zoning is included here. The data is aggregated since it is still scattered in the building level of detail. Other than that, the secondary data is also prepared.

3.2.1.1 Data Preparation

The data needs to be prepared for each timestamp from August 18 to May 31, using a 5-minute interval. The dataset outlined in Section 3.1.3.1 will be used as dependent variables in the regression model, with additional dependent variables including exam and holiday dummy variables.

Time will also be incorporated into the equation, represented as minutes relative to the start of each day. Given that the traffic count data may contain outliers that do not accurately reflect actual travel patterns, an outlier removal process is necessary.

Multicollinearity may arise due to the presence of constant and correlated variables. While these variables are initially included in the model to ensure comprehensive analysis, they may be removed later in the process if they significantly affect the model's performance or interpretability. This approach balances the need for robust modeling with the potential need for simplification.

The traffic count data exhibits high variance, and its distribution does not closely resemble a normal distribution. Therefore, the chosen data cleaning method is based on the distribution shown by the traffic count data. Ensuring that extreme values are filtered while retaining meaningful traffic variations.

Based on the defined zoning system, the cordon model is used to estimate access and egress traffic counts, treating them as the production and attraction of each zone. However, since traffic counts are primarily available for major bicycle corridors, particularly along Mekelweg, certain access and egress points remain unknown. To address this limitation, assumptions are made based on observed traffic count patterns.

The study assumes that traffic along Mekelweg follows a specific pattern: the traffic volume at the Jaffalaan–Mekelweg intersection is expected to be higher than at Stieltjesweg–Mekelweg, with a further decrease at locations farther south, such as Kruithuispad and TU Zuid, as illustrated in Figure 3.14. The figure indicates that sensor `TUD_SBX01` consistently records higher traffic counts compared to `TUD_SBXN6A`, while `TUD_SBXN13` is always less busy during peak hours than the previous sensors. This assumption is derived from a learned comparison of average traffic count data. Based on this, the estimated traffic counts used to represent trip production and attraction are calculated using the following formula:

$$T_C = T_A + \left(\frac{T_B - T_A}{x_B - x_A}\right) \times (x_C - x_A)$$

Where:
T is the traffic count, x is the traffic count observation location relative to `TUD_SBX01`, A and B are known observation points, and C is the estimated observation point.

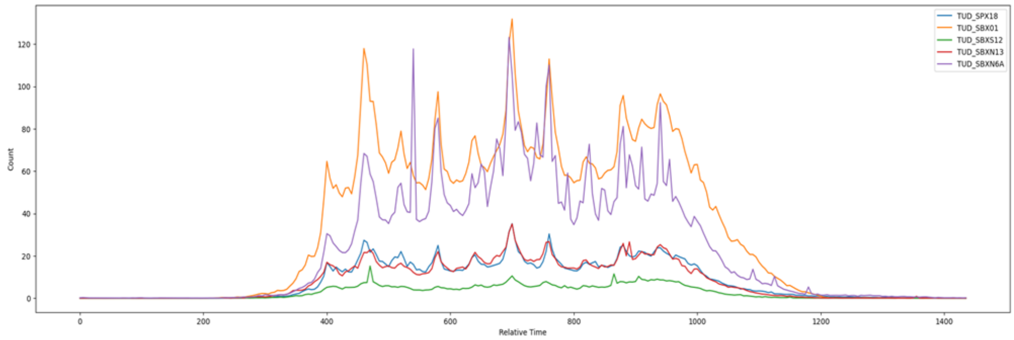


Figure 3.14: Traffic Count Comparison

Additionally, it is assumed that bicycle traffic along the perimeter of the campus decays linearly from Mekelweg, reflecting a gradual reduction in flow as the distance from the central corridor increases. It is

For example, traffic from Leeghwaterstraat is expected to be lower than that on Mekelweg. To validate these assumptions, Strava heatmap is used as a control measure. However, it is not directly integrated into the estimation process, as it may reflect induced traffic from areas beyond the study scope, potentially leading to discrepancies in the traffic count estimation based on the defined principles.

Furthermore, the usage of this assumed extrapolation is explained in Section 3.2.1.2.

3.2.1.2 Zoning System Requirement and Process

The model developed to achieve the research aim should produce a link flow model as the final output, necessitating a high spatial resolution. Since bicycle traffic fluctuates rapidly over short periods, the temporal resolution must also be sufficiently detailed to capture these variations effectively.

The defined 5-minute interval is considered appropriate, as it successfully captures real-time fluctuations in bicycle traffic, as demonstrated in the previous section. This resolution ensures that the model can accurately reflect short-term variations in demand, which is crucial for analyzing the conditions that impact both travel efficiency and safety on the TU Delft campus bicycle network.

To achieve this, the zoning system is defined based on key principles. Given the available data and information, a different approach to determine trip volume is formulate.

It is important to note that the data obtained in this study focuses on the internal zones. As a result, there is a different perspective when determining the traffic volume for internal and external zones, which in turn influences how the zoning system is defined. The cordon line is established based on the outermost traffic count locations, as well as assumed access and egress points to the campus area. For example, the Stieltjesweg–Schoemakerstraat intersection is considered an assumed access/egress point, despite not having a traffic count observation.

Within this cordon, zoning is determined based on the locations of buildings to ensure that the link flow model can be realistically applied. The Traffic Analysis Zones (TAZs) are initially formed based on land use homogeneity, ensuring consistency by using buildings as benchmarks for zone delineation. Additionally, the bicycle network should be modeled as a series of links to the greatest extent possible, allowing for a more accurate representation of movement patterns.

However, due to the absence of data for the external zones, it is assumed that these zones are connected to the traffic count points along the cordon. Therefore, the equation in Section 3.2.1.1 becomes necessary, as many of the assumed access or egress points do not have corresponding traffic count data, like Stieltjesweg–Schoemakerstraat intersection.

To put it simply, the internal zone is determined by the building, but the cordon governs the zoning of external zones.

In this project, it is demonstrated that spatial and temporal details are crucial for accurately modelling travel demand. However, the spatial density of the existing data is very limited, as traffic counts are only available for a few buildings and main roads. As a result, the modelling approach must be adapted to suit both the research context and the limited data availability. The practical approach taken, given the lack of spatial detail, involves not only estimating traffic counts for areas around the campus perimeter but also carefully selecting the data used in the regression analysis. Furthermore, the direct production-attraction relationship can only be observed at three traffic count locations, which will be elaborated in the following section.

3.2.2. Trip Production and Attraction

Trip generation is formulated based on traffic count data, making this information essential for accurately estimating production and attraction volumes. Within the model, the zoning system is structured to distinguish between internal and external zones.

For internal zones, production and attraction volumes can be estimated using identified variables from the available data. However, for external zones, the same approach is not feasible due to the unavailability of comparable data. Since key variables used for internal zones are not available for external zones, modeling their production and attraction directly is not possible.

Given this limitation, achieving a unified model that accurately represents both internal and external zones is challenging. As a result, the estimation will primarily focus on internal zones, where sufficient data is available. For external zones, production and attraction volumes will instead be derived directly from observed traffic counts and estimated traffic flows, ensuring that the model remains as accurate as possible within the constraints of the available data.

It is important to note that the locations of traffic count observations play a critical role, as the production and attraction volumes are modelled based on them. From the existing data, only three locations are considered sufficiently reliable to serve as benchmarks for the trip generation sub-model.

The first is TUD_SPXS06_IO, located at the main entrance of the IDE building. Another sensor, TUD_SPX14, is positioned at the logistics department entrance at the rear of the IDE building. Aside from these, other entrances to the IDE are located within the building, as shown in Figure 3.15. Due to data limitations, it is assumed that all main entrances exhibit the same trip rate, and this assumption is extended to entrances used for service purposes as well. Since the traffic counts at the IDE building record overall (mixed) traffic, a modal split assumption is applied to isolate bicycle traffic from the total.

The second location is TUD_SBXN9A, which captures traffic at the main entrance of the CEG building. A secondary entrance for service or staff purposes is monitored by TUD_SPX21. The same approach used for the IDE building is applied here: it is assumed to represent mixed traffic, and the modal split is used to derive bicycle counts, as illustrated in Figure 3.16.

The third key location is TUD_SBX16, which monitors the entrance to the bicycle parking basement of the ECHO building. This location is considered to capture purely bicycle traffic, eliminating the need for a modal split assumption.

To enhance the interpretability of the model, these three buildings are treated as individual zones, and the regression analysis is conducted using data collected from these specific observation points.

Figure 3.15: IDE entrances (the main entrances are marked by the yellow point

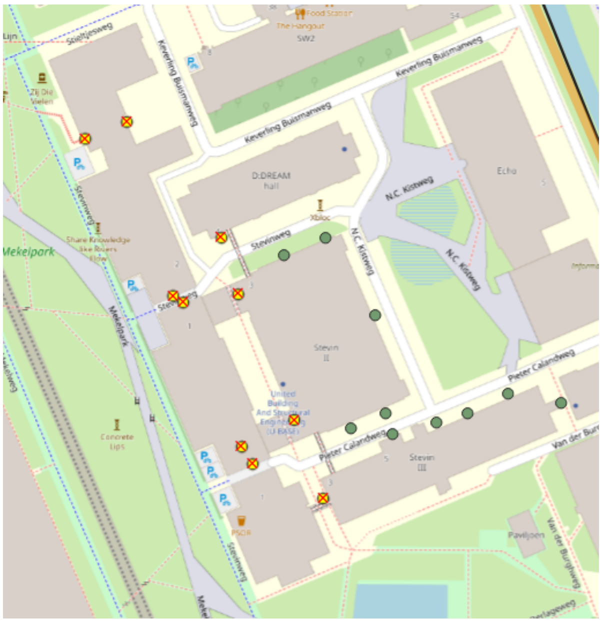


Figure 3.16: CEG entrances (the main entrances are marked by the yellow points

3.2.2.1 Linear Regression

The regression method employed in this study is Ordinary Least Squares (OLS), as utilized in Lu et al. (2018), albeit in a different context. OLS is selected due to its capability to evaluate both the quality of the model and the significance of the variables involved. The equation for OLS is as follows:

$$Y = X\beta + \epsilon$$

There is a potential for multicollinearity due to the implementation of constant variables over time and highly correlated data like lecture halls seats. However, these constant variables are crucial as they capture the capacity and accessibility of buildings, reflecting the activity levels associated with them. To address this, an additional regression model is conducted with the Generalized Least Square method to evaluate whether the impact of multicollinearity can be deemed negligible. The GLS, different from the OLS, does not assume that variables are independent, but the covariance of the variables are considered. In this case, since the real variability of the trip production or trip attraction is unknown, the variables' data covariance is used in the GLS analysis. It is prone to misinterpretation, but this comparative approach ensures the robustness of the regression analysis while retaining meaningful explanatory variables.

The trip production and attraction estimation focuses on the traffic count data located along the cordon of each defined zone. The collected data is divided into two parts: 70% is used for regression analysis (training), while the remaining 30% is reserved for model validation. The validation process employs the Root Mean Squared Error (RMSE) to assess the model's predictive accuracy.

The data split is performed based on the date of observation. Since some buildings lack data on certain dates, the available dates for each building are first identified, and then a randomized selection is used to allocate 70% of those dates to the training dataset. This approach ensures that the training set contains a diverse mix of scenarios, improving the generalizability of the model.

This approach ensures a reliable framework for trip production estimation and model validation. Furthermore, the regression is implemented on the 7.45-9.45 and 16.45-18.45 to really differentiate the peaks and also to describe the purpose of the travel better.

3.2.2.2 Variables Used

The dependent variables are selected based on both data availability and their ability to represent the capacity and trip generation of each zone within the campus. The dataset used aligns with the variables outlined in Section 3.1.3.1, with the inclusion of additional dummy variables to account for specific time-related influences, such as holidays or exam periods.

To further enhance the model's accuracy, an interaction term is introduced to capture the relationship between key variables and time, allowing the model to better reflect temporal variations in bicycle travel demand. This ensures that fluctuations in travel behavior, influenced by both infrastructure and time-dependent factors, are effectively incorporated into the analysis. The interaction term is the multiplication of the variables and the time.

The variables that is incorporated in the regression are:

- Total Employees
- Total Students
- Study space seats
- Lecture hall capacity seats
- Exam dummy
- Holiday dummy
- Day of week
- Morning binary variable
- Daily mean wind speed
- Temperature
- Precipitation
- Time (5 minutes intervals of minutes of hours, for example 7:45, the value is 45-50)
- Total employee interaction term(with holiday dummy)
- Total student interaction term (with holiday dummy)
- Lecture hall interaction term (with exam dummy)

Due to the limited availability of detailed data and the absence of clear indicators distinguishing which variables contribute specifically to trip production or attraction, the basic equation structure for both production and attraction is assumed to be the same. Then, in the form of the equation, it would be shaped as follows:

$$\begin{aligned} T_i = \quad & \beta_{\text{emp}} N_{\text{emp}} + \beta_{\text{stu}} N_{\text{stu}} \\ & + \beta_{\text{study}} N_{\text{study}} + \beta_{\text{lec}} N_{\text{lec}} \\ & + \beta_{\text{exam}} b_{\text{exam}} + \beta_{\text{holiday}} b_{\text{holiday}} \\ & + \textstyle\sum_{j=1}^{7} \beta_{d_j} w_j + \beta_{\text{morning}} b_{\text{morning}} \\ & + \beta_{\text{wind}} V_{\text{wind}} + \beta_{\text{temp}} T + \beta_{\text{precip}} Pr \\ & + \textstyle\sum_{k=1}^{12} \beta_{t_k} d_k \\ & + \beta_{\text{emp-holiday}} (N_{\text{emp}} \cdot b_{\text{holiday}}) \\ & + \beta_{\text{stu-holiday}} (N_{\text{stu}} \cdot b_{\text{holiday}}) \\ & + \beta_{\text{lec-exam}} (N_{\text{lec}} \cdot b_{\text{exam}}) \end{aligned}$$

Where:

N_{stu}	= Total number of students
N_{emp}	= Total number of employees
N_{study}	= Number of study space seats
N_{lec}	= Lecture hall capacity seats
b_{exam}	= 1 if exam period, else 0
b_{holiday}	= 1 if holiday period, else 0
b_{morning}	= 1 if morning peak period, else 0
w_j	= Day-of-week dummy variable:
w_1	= 1 if Monday, else 0
w_2	= 1 if Tuesday, else 0
$\vdots$	$\vdots$
w_7	= 1 if Sunday, else 0
V_{wind}	= Average wind speed
T	= Temperature
Pr	= Precipitation
d_k	= Time dummy variable for each 5-min interval:

$$\begin{array}{ll} \beta_{t_1} \cdot d_1 & \text{if time is within minute 45–50} \\ \beta_{t_2} \cdot d_2 & \text{if time is within minute 50–55} \\ \vdots & \vdots \\ \beta_{t_{12}} \cdot d_{12} & \text{if time is within minute 40-45} \end{array}$$

with each $d_k \in \{0, 1\}$ depending on whether that 5-minute interval applies.

There are several notable decisions made in the formulation of the equation. One of them is the repetition in the use of certain variables, such as the total number of students and the exam dummy. This is done under the hypothesis that nonlinear relationships may exist between traffic volume and these variables. To capture such potential patterns, interaction terms—formed by the multiplication of relevant variables—are introduced. A more practical justification for this approach is that many known variables around the TU Delft campus are static and primarily reflect regular or background traffic levels. These static variables alone may not account for fluctuations in volume caused by time-specific activities or events. Therefore, the interaction terms are expected to capture these fluctuations and provide a more accurate representation of traffic volume in the context of dynamic campus activity.

3.2.3. Trip distribution

Trip distribution is the process of allocating trip productions and attractions to their respective destinations and origins within the study area. Traditionally, distance plays a key role in this step, where travelers are assumed to favor closer destinations due to travel impedance—an assumption that underpins gravity-based models. However, in the context of a compact campus such as TU Delft, distance is relatively uniform and not a significant factor in travel behavior. As a result, this study employs an alternative approach: Iterative Proportional Fitting (IPF). This method distributes trips by adjusting an initial origin-destination (OD) matrix so that its row and column totals match observed or estimated trip productions and attractions. However, the accuracy of the IPF results strongly depends on the quality of the initial skim matrix. Therefore, careful consideration is required in developing this initial matrix to ensure it reflects realistic travel patterns within the study area.

3.2.3.1 Initial Skim Matrix

As previously stated, the initial skim matrix is essential to obtaining an accurate origin-destination (OD) matrix through the IPF method. This initial matrix represents the assumed proportion of trips occurring between each pair of zones, and therefore, plays a critical role in shaping the final distribution outcome. It is important to consider the context of both origin and destination when determining these proportions, as the assumptions must reflect the actual or expected travel behavior. Given the limitations in available data, the relationships between zones in this project are not directly observable and must be carefully assumed. As a result, the treatment of trips must differ depending on the zone types involved—specifically, trips between internal-internal zones, internal-external zones, and external-external zones require separate considerations.

Internal-Internal Zones The internal-internal zones skim value needs to be representative of the relationship between origin and destination zones, and it requires a related variable to achieve this. However, due to data limitations, there are no directly linking variables available between each zone. For example, a suitable linking variable could be the number of lectures held in other faculty buildings, such as when people from CEG need to move to the IDE building for computer labs or to PULSE to access larger classrooms. Unfortunately, such variables are not available for this project. Inspired by the SARC model for direct demand, which connects zones using simple variables derived from production potential and attraction factors, an estimation approach is suggested. This involves introducing the population of each faculty as a proxy for potential trips and bike parking capacity as a proxy for the attraction of each zone. It is interesting to note that the SARC model similarly uses population and income alongside impedance and mode availability as variables. Based on this, the proposed estimation method utilizes these proxies to represent trip production and attraction between zones.

$$S_{ij} = \frac{\big((\text{total\_student}_i + \text{total\_emp}_i) + 1\big) \times \big(\text{total\_parking}_j + 1\big)}{\sum_i \sum_j \big((\text{total\_student}_i + \text{total\_emp}_i) + 1\big) \times \big(\text{total\_parking}_j + 1\big)}$$

$$\begin{aligned} S_{ij} &= \text{Normalized skim value from origin zone } i \text{ to destination zone } j \\ \text{total\_student}_i &= \text{Number of students in origin zone } i \\ \text{total\_emp}_i &= \text{Number of employees in origin zone } i \\ \text{total\_parking}_j &= \text{Bike parking capacity in destination zone } j \end{aligned}$$

Internal-External Zones Relationships Defining the relationship between internal and external zones is more difficult than for internal zones alone, mainly because there are no shared or directly connecting variables available in this project. The only data that links internal and external zones is traffic count data. Therefore, it is assumed that the proportion of trips between internal and external zones can be estimated using this traffic count information. Using traffic counts to relate the skim matrix of internal-external trips to observed traffic volumes is helpful because, unlike internal zone relationships—which are based on fixed variables and don't change over time—traffic counts capture dynamic travel patterns.

For example, consider zone 1 as an internal zone, with its internal-internal skim matrix row already defined. The total trips from this row are adjusted by a scaling factor derived from the ratio of observed traffic between internal buildings and nearby external points. In this project, the scaling factor is calculated using traffic counts from three internal zones (buildings) and three external zones at the campus perimeter. The adjusted total trips are then distributed across the external zones based on their observed traffic volumes. A similar method is used for trips going from external zones to internal zones, applying a different scaling factor calculated from traffic counts.

$$E_{ij} = \left(\sum_{k \in \text{int}} S_{ik}\right) \cdot r_i^{\text{out}} \cdot \frac{C_j^{\text{ext}}}{\sum_{l \in \text{ext}} C_l^{\text{ext}}}$$

$$I_{ji} = \left(\sum_{k \in \text{int}} S_{ki}\right) \cdot r_i^{\text{in}} \cdot \frac{C_j^{\text{ext}}}{\sum_{l \in \text{ext}} C_l^{\text{ext}}}$$

Where:
E_{ij} = Estimated trips from internal zone i to external zone j
I_{ji} = Estimated trips from external zone j to internal zone i
S_{ik} = Skim matrix values between internal zones i and k
r_i^{out} = Scaling factor from internal to external, based on traffic count ratios
r_i^{in} = Scaling factor from external to internal, based on traffic count ratios
C_j^{ext} = Observed traffic count at external zone j

External-External and Intrazonal Skim Matrix Due to data limitations, interactions involving external zones and intrazonal trips must be treated with simplified assumptions. No data has been obtained that directly relates to the interaction between internal and external zones, and the available traffic count data does not account for these external areas. Similarly, there is no specific information reflecting intrazonal movements within individual buildings or zones. As a result, and to maintain consistency with the available dataset, the initial skim matrix values corresponding to both external zone interactions and intrazonal trips are set to zero. This assumption simplifies the model while acknowledging the constraints of the input data.

3.2.3.2 Iterative Proportional Fitting

Iterative Proportional Fitting is used because the gravity model is not quite relevant to the short-distance cycling in the campus area.

$$T_{ij}^{n+1} = T_{ij}^{n} \times \frac{G_i}{\sum_j T_{ij}^{n}} \tag{3.1}$$

$$T_{ij}^{n+1} = T_{ij}^{n} \times \frac{A_j}{\sum_i T_{ij}^{n}} \tag{3.2}$$

Where i is the initial zone, j is the destination zone, and n is the iteration number.

Validating the Origin-Destination (OD) matrix is crucial but challenging in this context. Hereby, several possible methods:

- Traffic assignment is a potential validation method, but it is outside the scope of this research.
- A direct validation through surveys of OD pairs, although these rely on stated preferences .
- An indirect validation method involves summing the rows and columns of the matrix but since the external zones is not modelled, it is hard to implement/

3.3. Method Summary

The framework of the research is depicted on Figure 3.17. The method chosen for this study is guided by data availability, as outlined in Chapter 2 and Chapter 3, ensuring alignment with the study's context and objectives. The four-step model has been selected due to their relevance and applicability to the available data. This approach is supported by Lu et al. (2018), which demonstrated that linear equations can effectively utilize highly specific variables rather than relying solely on traditional inputs such as land use or socio-demographic data. This flexibility enables a detailed modeling approach that fits the dynamic context of the TU Delft campus.

To capture the dynamics of bicycle traffic on the TU Delft campus, a 5-minute interval is selected for the model, with the morning and evening periods are chosen for analysis. This approach reduces computational effort while preserving the essential fluctuations in traffic, avoiding the risk of over-smoothing key patterns.

The methodology involves two primary steps: trip generation and trip distribution. Trip generation is conducted through regression analysis, focusing on the trip attraction and production, which is produced by an equation. Initially, multicollinearity is expected to occur; however, its impact will be evaluated through alternative methods to determine whether it significantly affects the regression model or can be ignored.

Trip distribution is carried out using the Iterative Proportional Fitting (IPF) method. The initial matrix for this process is estimated under the assumption that destination choice is driven primarily by activity-related factors rather than generalized cost minimization. This assumption aligns with the campus environment, where accessibility and activity levels strongly influence travel patterns. Together, these steps provide a comprehensive framework for analyzing bicycle travel demand on the TU Delft campus.

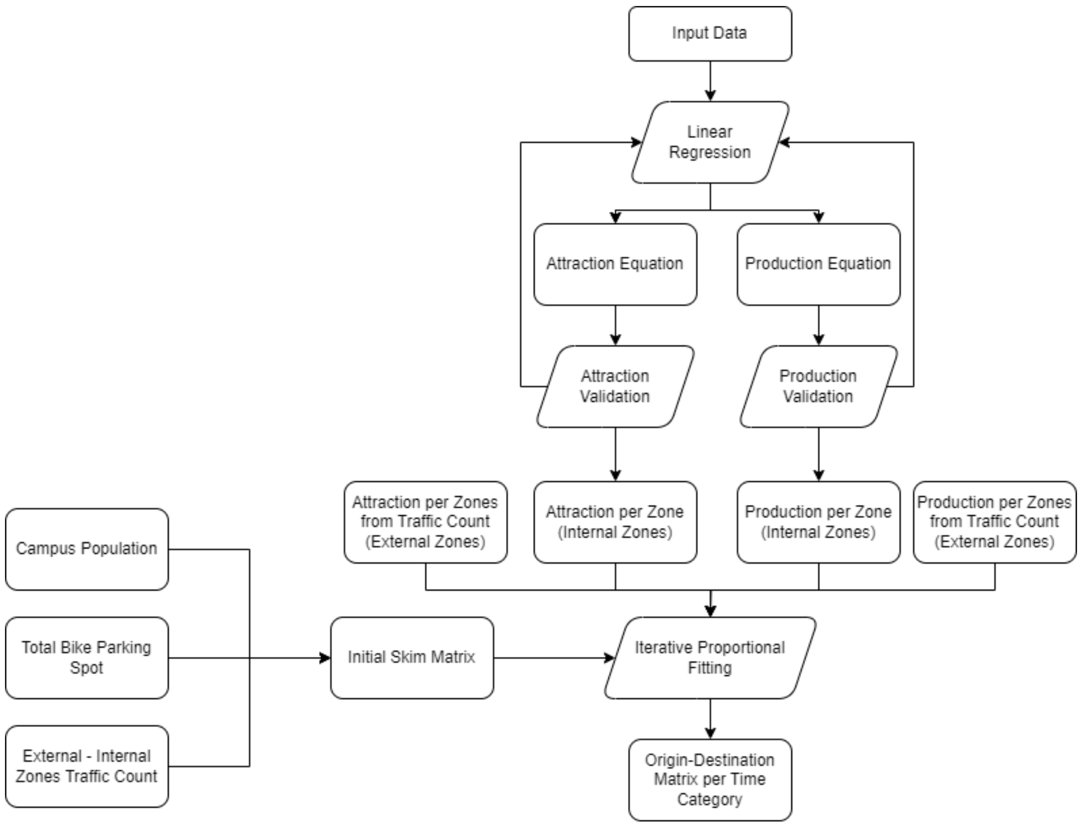


Figure 3.17: Research Methodology

4

Origin-Destination Matrix Estimation

The estimation of travel demand in this study involves two key processes: trip generation and trip distribution. These processes follow the methodology outlined in the previous chapter, making a certain consistency in approach and analysis.

For clarity, the primary output of this chapter and the overall research is the estimation of origin-destination relationships, which are represented by an OD matrix. This matrix serves as a crucial tool for understanding travel patterns within the TU Delft campus.

The results derived from the implementation of these processes will be thoroughly discussed in the following chapter, where their implications and relevance to the study objectives will be analyzed.

4.1. Data Preprocessing and Assumption Formulation

4.1.1. Zoning System

The zoning system follows the requirements outlined in Section 3.2.1.2, which are based on the best-case scenario in terms of data availability from the existing repository—particularly the traffic count data, as the model defined in this project is a cordon-based model. This section begins by describing the ideal zoning configuration that aligns with the objectives of the model and assumes full data availability.

Following this, the section details the practical steps taken to develop the zoning system applied in this research. The final zoning structure is the result of an iterative process of merging and separating the ideal zones, while preserving homogeneity in land use function and maintaining spatial proximity. This method ensures that the resulting zoning system is both data-informed and well-suited to the functional and spatial dynamics of the TU Delft campus.

4.1.1.1 Ideal Zoning System

As explained in Section 3.2.1.2, the model specification must align with the objective of analyzing link-level bicycle volumes on the TU Delft campus. Achieving this requires a fine spatial resolution. Ideally, every building would be represented within the bicycle network model, and all bicycle pathways—whether main routes or shared roads with cars—would be modeled as links. In this structure, each building's bike parking entrance would serve as an access or egress point.

This approach would also minimize unaccounted intrazonal trips, ensuring that most travel is captured and weighted within the network. As illustrated in Figure 4.1, the ideal TAZ configuration enables full coverage of all bicycle links, with direct connectivity between parking access points and the bicycle network. For example, if Mekelweg were found to be overly congested and posed safety risks, the model would allow for alternative planning, such as developing Leeghwaterstraat into a new main bicycle corridor.

However, implementing this ideal zoning model is constrained by the limited availability of traffic count data at access and egress points. Realizing such a detailed model would require extensive data collection, resulting in high costs in terms of time, financial resources, and effort. In the context of a cordon-based model, traffic counts would need to be conducted along the campus perimeter, including locations such as Schoemakerstraat–Stieltjesweg and the area surrounding the BK faculty, which further increases the complexity and resource demands of data collection.

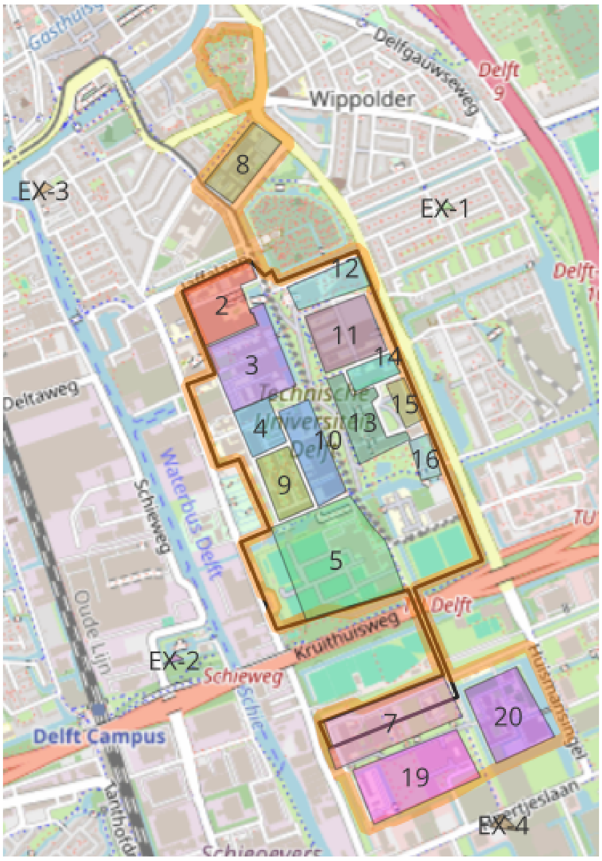


Figure 4.1: Ideal Zoning System

4.1.1.2 Project Zoning System

Since the data used in this project is limited to what is available within the campus environment, certain compromises are necessary in developing the model. As previously stated, these compromises primarily affect the level of spatial detail, while still preserving the model's goal of representing the bicycle network as a system of links as accurately as possible. The attempt to develop the TAZ, for this project is depicted in Figure 4.2.

In this project, the zoning system is designed with a primary focus on maintaining the explainability of the model. Spatial detail is preserved at the building level, as the regression analysis relies on traffic count data collected at individual buildings. This level of granularity aligns with the overarching objective of constructing a detailed link flow model. However, the model's accuracy may be limited due to the use of data from only three buildings, raising concerns about its predictive power for buildings not represented in the dataset.

The decision to retain high spatial detail means that data availability becomes a critical factor in determining which areas are included in the model. For instance, residential zones are excluded, as the selected buildings are primarily academic and do not encompass residential functions. Similarly, areas outside the defined cordon are not considered, due to a lack of data. Instead, a cordon-based modeling approach is applied, using traffic counts at access and egress points around the perimeter to estimate production and attraction volumes of external zones. Nevertheless, this approach introduces further challenges, particularly concerning the accuracy and representativeness of those perimeter-based measurements.

Despite these efforts, the zoning system still has several shortcomings—particularly because there are different approach between zoning inside and outside the cordon. It is happened due to the lack of traffic data at specific access and egress points within the zones and along the campus perimeter. Although the previous chapter introduced a method to estimate traffic volumes in the absence of direct measurements, such assumptions remain limited in their ability to accurately represent real conditions. A notable example is the Stieltjesweg–Schoemakerstraat intersection, a key access point that experiences substantial traffic due to the nearby concentration of student housing. This limitation underscores the need for more comprehensive data collection at critical locations, which would enhance the accuracy of the zoning system and improve the

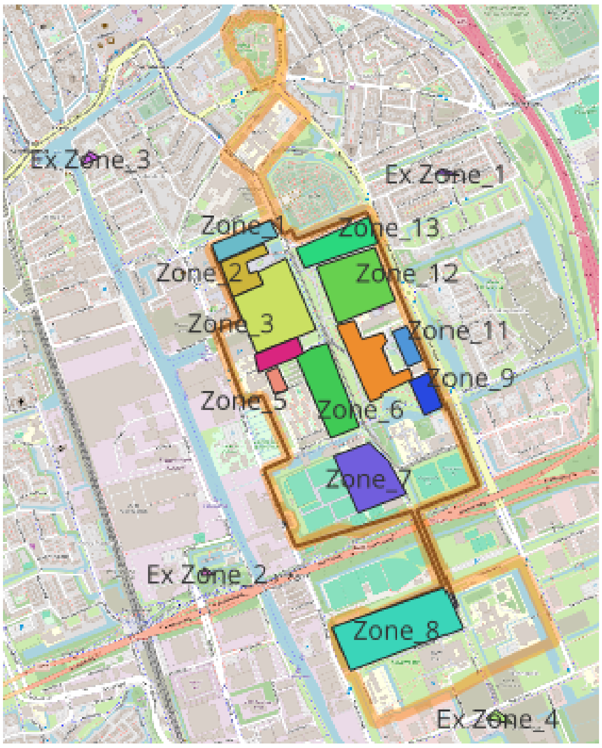


Figure 4.2: Project Zoning System

model's performance in future research.

By introducing this TAZ structure, several key access and egress points are assumed and also recommended for future observation to improve the accuracy and completeness of the model. By introducing these points, the easier implementation of estimating trips produced by or attracted to the external zones is made. These points include:

- Christian Huijgensweg
- Stieltjesweg
- Van der Burghweg
- Cornelis Drebbelweg
- Balthasar van der Polweg
- Kluyverweg

These locations are suggested as critical observation points to enhance the model specification, ensuring it better meets its intended purpose. However, it is important to note that these recommendations are made from the perspective of a cordon-based model. While useful, this modeling approach presents limitations when applied to a campus such as TU Delft, which has a stretched, linear geometry. This shape introduces challenges in capturing internal circulation and peripheral flows accurately, making comprehensive data collection at key access points even more important for refining the model.

4.1.2. Data Preparation

The data preparation process involves consolidating all relevant data into a single dataset for regression analysis, data cleaning for the traffic count, and determining the trip production and attraction for each zone based on traffic counts.

4.1.2.1 Data Cleaning

Data cleaning is essential for traffic count analysis, as the boxplot in the previous section revealed numerous outliers in the data. Further analysis is conducted after translating the traffic count data into observed production and attraction trips within the internal zones. For example, Figure 4.3 shows that the observed production trips follow an exponential distribution with a very long tail, which may not accurately represent the overall data. These outliers introduce noise and can reduce the accuracy of the model. To address this issue, a data cleaning process that trims the tail of the distribution is proposed, using the quantile method. Figure 4.4 illustrates the data after discarding values beyond the 99th percentile.

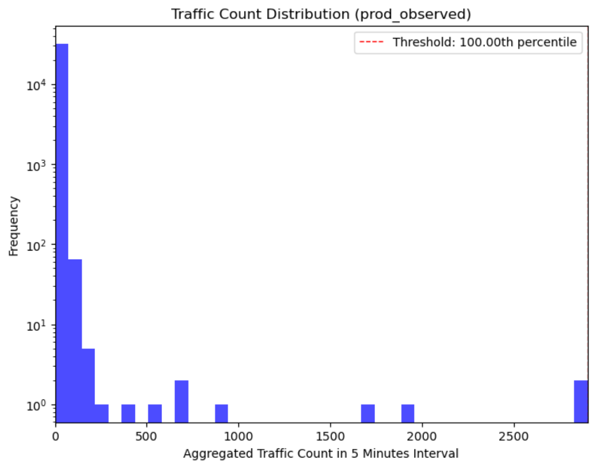


Figure 4.3: Production Observed Before Data Cleaning

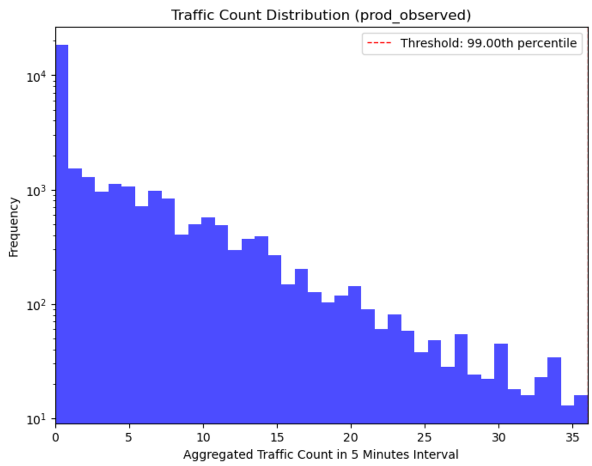


Figure 4.4: Production Observed After Data Cleaning

4.1.2.2 Traffic Count Estimation

In addition to data cleaning, data preparation is essential, especially in relation to the four-step model. According to the zoning system used, each traffic count is linked to the perimeter of specific zones. It is assumed that these counts directly represent the production and attraction values of the zones. However, this approach does not consider the network structure within the zoning process and relies solely on traffic counts. As a result, a single traffic count may correspond to multiple zones. Therefore, accurately locating the traffic count along the zone perimeter is crucial for determining a zone's trip production or attraction.

In this project, the formulation of Traffic Analysis Zones (TAZs) is based on traffic counts. It is assumed that traffic counts between two known observation points can be linearly interpolated based on distance, as shown in Section 3.2.1.1. An example of this approach is the Christian Huijgensweg–Mekelweg intersection, with TUD_SBX01 and TUD_SBXN6A as reference points:

$$T_{intersection} = T_{TUD\_SBX01} + \left(\frac{T_{TUD\_SBXN6A} - T_{TUD\_SBX01}}{x_{TUD\_SBXN6A} - x_{TUD\_SBX01}}\right) \times (x_{intersection} - x_{TUD\_SBX01})$$

x is the location of the intersection relative to the TUD_SBX01. From this intersection, it is assumed that traffic counts decrease with distance from Mekelweg, and by the time it reaches the Christian Huijgensweg–Schoemakerstraat intersection, the reduced traffic count represents trips related to the external zone.

Once the data is properly prepared and organized into a dataframe, it can be used for trip generation analysis using statistical modules available in Python.

4.2. Trip Generation

The zoning system outlined in the model differentiates between external and internal zones. For internal zones, production and attraction volumes can be estimated using variables identified from feasible data obtained within the campus, even though it is limited to the direct volumes from three buildings. However, modeling the production and attraction volumes of external zones is not possible due to the unavailability of comparable data. Consequently, the focus of the model is on understanding the behavior of individuals within the campus. This research aims to examine the circumstances surrounding campus-related travel, particularly during morning and evening peak times. Due to data limitations, the model primarily addresses the travel demand during these peak periods, recognizing that it might not fully explain the purpose of the trips. The morning peak is characterized as a period when the campus acts as a major attraction point due to the start of activities, whereas the evening peak is considered in terms of trip production from the campus. It is important to note that the insights derived from this research might represent only a partial view of the overall travel demand model.

4.2.1. Trip Attraction

The OLS (Ordinary Least Squares) linear regression is conducted for trip generation, as outlined in the proposed method. This research utilizes the statsmodels module in Python, which offers high flexibility for modifying the model and provides comprehensive reports on model explainability, accuracy, and variable significance. The module is also capable of performing analyses for multicollinearity, such as Variance Inflation Factor (VIF) analysis or covariance analysis. However, this research will not address these aspects and will focus solely on the explainability of a linear model. The results of the regression analysis for trip attraction are presented in Table 4.1.

The overall evaluation of the model reveals a low ability to explain the nature of trip attraction, as reflected by an value of only 0.415. This indicates that the model accounts for just 41.5% of the variation in the data. One possible explanation for this limitation is underfitting, as there is a trend showing that R^2 increases when fewer samples are used. The validation plot in Figure 4.5 further supports this, showing that the data points are widely dispersed, making it difficult for the model to accurately predict actual traffic volumes. The figure also shows the bias of the model, since the low actual value has been predicted incorrectly. It could be due to the data that has many zero values as the target. Another possibility is the model itself is not representative, and the model should try capture some non-linear relation. Given these findings, a closer analysis of the variables used in the model is conducted to better understand their impact and potential adjustments. The RMSE shows a 4.85 value for the RMSE means quiet variance is still there.

The overall evaluation of the model reveals a limited ability to explain trip attraction patterns, as indicated by an R^2 value of only 0.415. This means the model explains just 41.5% of the variance in the data. One possible explanation for this low performance is underfitting, supported by a trend showing that R^2 improves when fewer samples are used. The validation plot in Figure 4.5 reinforces this observation, as the data points are widely scattered, indicating the model struggles to predict actual traffic volumes accurately. The plot also highlights model bias, particularly in cases where low actual values are predicted poorly—this may be due to the presence of many zero values in the target data.

Another possibility is that the model itself is not well-suited for the data, potentially missing important non-linear relationships. As a result, a more detailed analysis of the input variables is undertaken to assess their relevance and explore potential improvements. Additionally, the RMSE value of 4.85 suggests a relatively high level of variance remains in the model's predictions.

From the final model, some conclusions can be drawn.

Dep. Variable:	attr_observed	**R-squared:**	0.415
Model:	OLS	**Adj. R-squared:**	0.414
Method:	Least Squares	**F-statistic:**	710.9
No. Observations:	21066	**AIC:**	1.245e+05
Df Residuals:	21044	**BIC:**	1.247e+05
Df Model:	21		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	4.56e-06	8.27e-08	55.129	0.000	4.4e-06	4.72e-06
is_morning	3.1626	0.064	49.273	0.000	3.037	3.288
45-50	0.8052	0.121	6.662	0.000	0.568	1.042
50-55	1.0123	0.120	8.455	0.000	0.778	1.247
55-00	0.7749	0.119	6.492	0.000	0.541	1.009
00-05	0.6364	0.116	5.502	0.000	0.410	0.863
20-25	-0.2374	0.118	-2.010	0.044	-0.469	-0.006
stud_seat	0.0032	5.54e-05	57.061	0.000	0.003	0.003
lecture_cap	-0.0037	5.52e-05	-66.707	0.000	-0.004	-0.004
daily_mean_temp	0.0018	0.001	2.402	0.016	0.000	0.003
daily_mean_prec	-0.0044	0.001	-6.369	0.000	-0.006	-0.003
exam_period	2.4325	0.175	13.910	0.000	2.090	2.775
holiday_vac_period	-4.0592	0.205	-19.796	0.000	-4.461	-3.657
day_1	1.1368	0.079	14.313	0.000	0.981	1.293
day_2	1.8198	0.080	22.647	0.000	1.662	1.977
day_3	1.8102	0.078	23.066	0.000	1.656	1.964
day_4	1.3731	0.079	17.305	0.000	1.218	1.529
day_5	0.2781	0.080	3.483	0.000	0.122	0.435
day_6	-3.2337	0.077	-41.991	0.000	-3.385	-3.083
day_7	-3.1843	0.077	-41.188	0.000	-3.336	-3.033
total_parking	0.0042	7.3e-05	58.204	0.000	0.004	0.004
lecture_cap_int	-0.0010	9.37e-05	-11.104	0.000	-0.001	-0.001
total_student	-0.0006	4.28e-05	-13.444	0.000	-0.001	-0.000
total_emp	-0.0020	3.4e-05	-60.251	0.000	-0.002	-0.002
total_student_int	-0.0027	0.000	-9.369	0.000	-0.003	-0.002
total_emp_int	0.0115	0.001	12.701	0.000	0.010	0.013

Omnibus:	5237.928	**Durbin-Watson:**	0.701
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	13418.923
Skew:	1.354	**Prob(JB):**	0.00
Kurtosis:	5.820	**Cond. No.**	1.07e+19

Table 4.1: Attraction Model - OLS Regression Results

- The variables used in the model generally perform as expected.
- **Morning binary** has a significant coefficient value, reflecting the substantial difference in travel purposes between the morning and evening periods. This variable later demonstrates a stronger influence in the attraction model, as the morning period aligns with the expected influx of individuals arriving at campus, which supports the initial hypothesis.
- **Time binary variables** (indicated by 5-minute intervals, e.g., 7:45 is represented by 45–50):
  - The 45–05 minutes interval is proven to be crucial, as it marks the beginning of lecture hours. The mixture of morning and evening attraction during this window may result in a more generalized coefficient sign, but the significance of this time window is correctly reflected in the model.
  - The 20–25 minutes interval is also considered important, as it aligns with the time when lectures

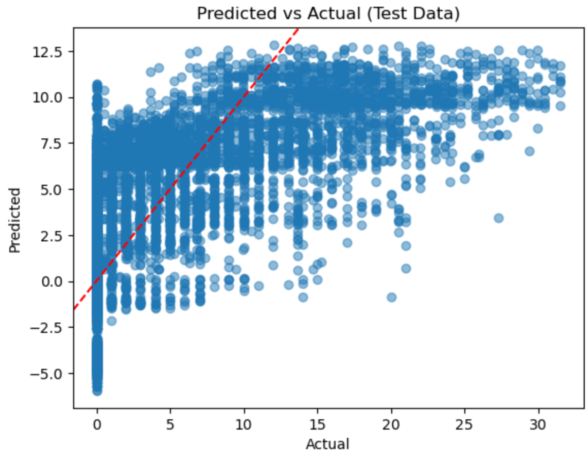


Figure 4.5: Attraction Validation - Predicted vs Actual Test

are about to end, which is visibly represented in the traffic patterns.

- **Capacity-related variables** such as lecture hall seats show negative coefficients, possibly because full spaces may indicate less current inflow. On the other hand, the study space shows a positive coefficient.
- **Weather-related variables** mostly show positive influence, except for precipitation, which has a negative coefficient—reflecting its role as a form of travel impedance.
- **Exam period dummy** contributes positively, consistent with the expectation that students are more likely to be present on campus during exams.
- **Holiday dummy** shows a negative effect, as fewer students and staff are expected on campus.
- **Bike parking capacity** correlates positively with traffic, indicating that well-equipped zones attract more cyclists.
- **The total student and employee variable** shows a negative value, possibly due to overlapping effects with other variables or collinearity. It also might be caused by the attraction is not controlled by the population
- **Day of Week binary variables** show the expected behavior: weekdays have similar coefficients, Friday has a slightly lower influence, and weekends have a negative effect.
- The **interaction term** is significant, as it is found in the final attraction model.

4.2.2. Trip Production

The trip attraction model is developed using the same method as the trip production model, but it focuses on the evening period when home-bound trips are more dominant. This approach allows the model to capture the behavioral shift in travel patterns that typically occurs later in the day, reflecting the return movement from campus to residential areas. The result of the regression is shown in Table 4.2.

Generally, the model performs worse than the attraction model, capturing approximately only 35.7% of the traffic, as indicated by a lower R^2 value. This decreased performance may be attributed to the lack of relatable and significant variables in the model to the evening activity of the campus, increasing the likelihood of underfitting and allowing for a decreased explanation of the observed variation in evening travel behavior. The validation plot also shows similar result to the trip attraction, with a similar RMSE value at 3.92.

On the other hand, the performance of individual variables provides further insights:

- **Morning binary:** This variable shows a lower coefficient value in the production model compared to the attraction model. This indicates that the variable has stronger explanatory power in the attraction context, as it reflects the typical behavior of individuals arriving at campus primarily during the morning hours.

Dep. Variable:	prod_observed	**R-squared:**	0.357
Model:	OLS	**Adj. R-squared:**	0.356
Method:	Least Squares	**F-statistic:**	614.3
No. Observations:	21066	**AIC:**	1.164e+05
Df Residuals:	21046	**BIC:**	1.166e+05
Df Model:	19		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	3.692e-06	6.75e-08	54.669	0.000	3.56e-06	3.82e-06
is_morning	2.1136	0.053	39.879	0.000	2.010	2.217
45-50	0.6923	0.099	6.989	0.000	0.498	0.886
50-55	0.9501	0.098	9.685	0.000	0.758	1.142
55-00	0.6375	0.098	6.517	0.000	0.446	0.829
00-05	0.5020	0.095	5.298	0.000	0.316	0.688
stud_seat	0.0025	4.51e-05	55.577	0.000	0.002	0.003
lecture_cap	-0.0021	4.36e-05	-48.715	0.000	-0.002	-0.002
daily_mean_temp	0.0024	0.001	3.909	0.000	0.001	0.004
daily_mean_prec	-0.0041	0.001	-7.182	0.000	-0.005	-0.003
exam_period	2.2585	0.144	15.704	0.000	1.977	2.540
holiday_vac_period	-3.9825	0.149	-26.737	0.000	-4.274	-3.691
day_1	0.8622	0.066	13.147	0.000	0.734	0.991
day_2	1.4514	0.066	21.877	0.000	1.321	1.581
day_3	1.3596	0.065	20.981	0.000	1.233	1.487
day_4	1.0180	0.066	15.540	0.000	0.890	1.146
day_5	0.1357	0.066	2.058	0.040	0.006	0.265
day_6	-2.4385	0.064	-38.349	0.000	-2.563	-2.314
day_7	-2.3885	0.064	-37.415	0.000	-2.514	-2.263
total_parking	0.0032	5.94e-05	54.522	0.000	0.003	0.003
lecture_cap_int	-0.0009	7.7e-05	-11.888	0.000	-0.001	-0.001
total_student	-0.0011	3.5e-05	-31.807	0.000	-0.001	-0.001
total_emp	-0.0016	2.77e-05	-58.937	0.000	-0.002	-0.002
total_emp_int	0.0034	0.000	13.113	0.000	0.003	0.004

Omnibus:	7109.460	**Durbin-Watson:**	0.862
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	25442.979
Skew:	1.693	**Prob(JB):**	0.00
Kurtosis:	7.186	**Cond. No.**	4.12e+19

Table 4.2: Attraction Model - OLS Regression Results

- **Time binary variables** generally behave as expected for homebound trips:
  - The 45–55 minute interval shows a positive coefficient, reflecting the transition to the next class period—opposite to what is seen in the attraction model.
  - The 55–05 interval also shows a positive coefficient, which aligns with the typical timing of homebound trips.
- **Capacity-related variables** show differing trends. Lecture capacity has a very small coefficient, consistent with the reduction in on-campus activities in the evening. In contrast, study space capacity has a positively influential coefficient, reflecting continued campus presence for individual study sessions.
- The **day of week** variable shows a similar pattern, with the attraction model showing the consistency.
- **Weather-related variables** maintain a similar trend to the attraction model, likely because weather impacts cycling regardless of trip purpose.

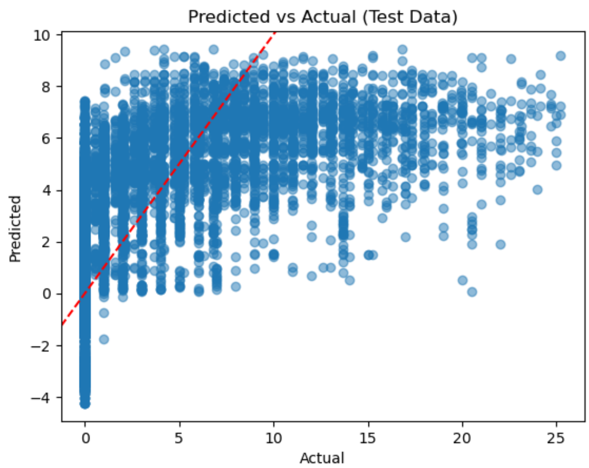


Figure 4.6: Production Validation - Predicted vs Actual Test

- **Exam period dummy** has a positive effect, possibly because students tend to travel a lot around campus during this period.
- **Holiday dummy** also shows a negative effect, as campus activity is reduced with fewer students and staff present.
- **Bike parking spots:** This variable shows a positive coefficient.
- **Interaction term of students and employees:** This variable fits better in the production model and demonstrates greater explanatory power in capturing traffic fluctuations related to time and contextual campus activities.

4.2.3. Trip Generation Result and Discussion

It should be noted that, overall, the model still performs poorly, moderately acceptable at best, with an explainability (as indicated by the R^2 of less than 50%. This low value suggests a high possibility of underfitting, indicating that the model may not fully capture the complexity of the data. A more sophisticated modeling approach or a richer dataset for training could potentially improve performance. Additionally, multicollinearity among the independent variables may also contribute to the model's limitations.

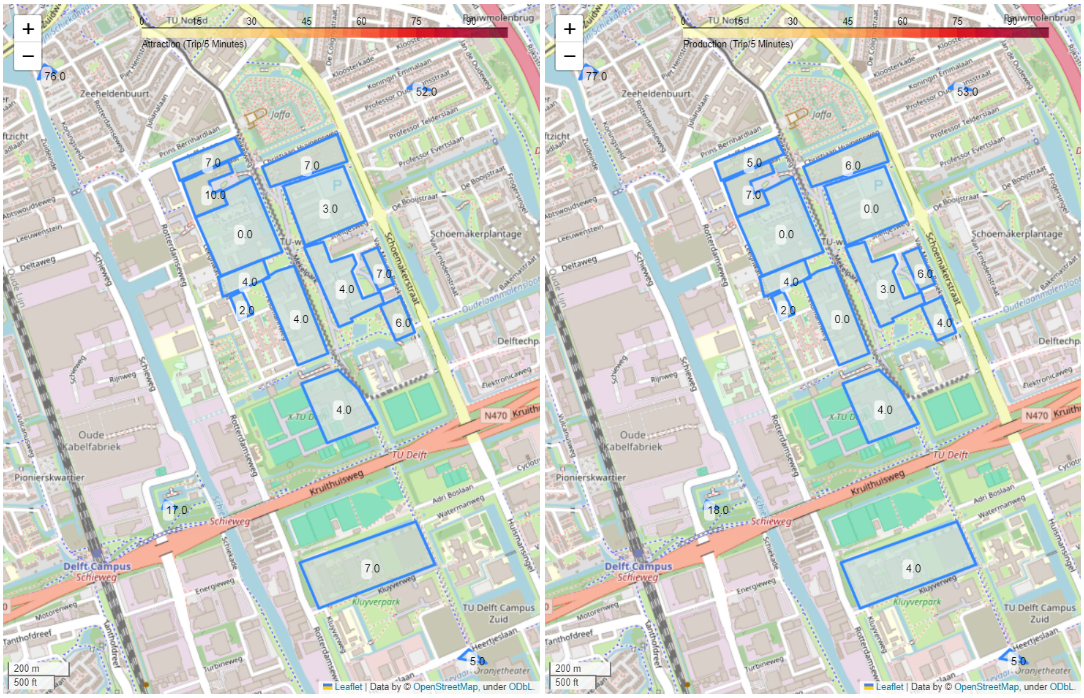
The RMSE values further support the need for improvement, showing that prediction errors remain significant. Moreover, the necessity of transforming the target variable highlights the potential value of exploring alternative modeling techniques—such as time series analysis—which may better capture the temporal dynamics of bicycle travel patterns.

Despite these limitations, it is worth noting that the variables behave in alignment with the expected trip purposes. In the morning, the selected variables effectively represent attraction to the campus, while in the evening, the homebound nature of trips is clearly captured, especially through the time-based binary variables. This indicates that, while the model struggles in overall performance, the variable selection and interpretation remain conceptually sound.

To uncover deeper patterns within the model, it is implemented to estimate production and attraction volumes, with the results visualized in several figures below. These patterns are then compared across different temporal conditions. Figure 4.7 and Figure 4.8 illustrate the contrast between morning and evening demand across the TU Delft campus.

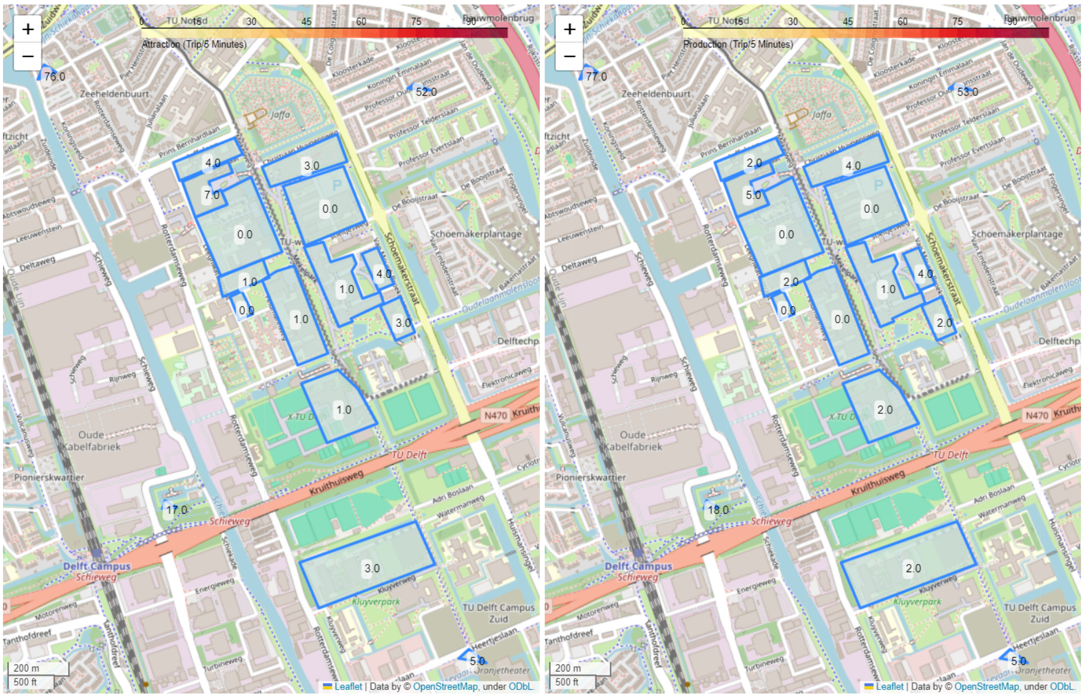
From the model specification, it is evident that capacity- and population-related variables primarily govern the baseline traffic level. Time-related variables, such as 5-minute intervals and the morning binary indicator, are included to capture temporal fluctuations. The morning binary variable, in particular, plays a key role in differentiating the two timeframes shown.

It can be observed that morning traffic generally generates higher volumes across most zones. External zones dominate the total volume within the modeled area during the morning period, reflecting typical home-based trips entering the campus. Additionally, the pattern shows that larger academic buildings both produce and attract significant volumes in the morning, which aligns with their functional role.



(a) Attraction (b) Production

Figure 4.7: Trip Generation Normal Monday at 7:55



(a) Attraction (b) Production

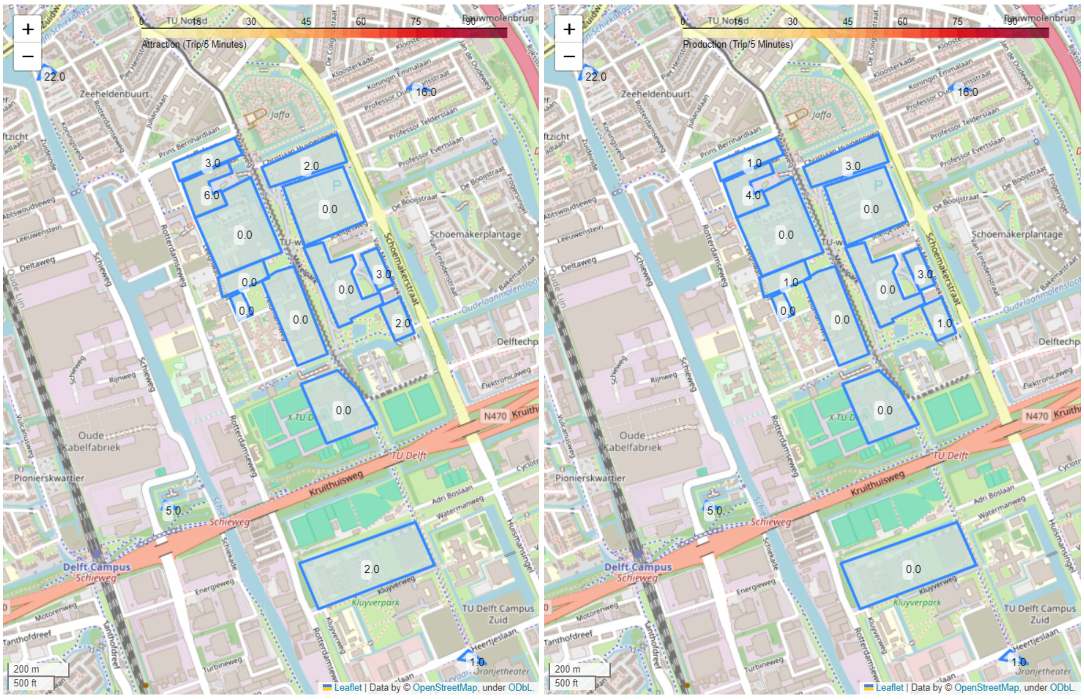
Figure 4.8: Trip Generation Normal Monday at 17:55

However, the presence of several zero-volume predictions for large buildings highlights a limitation of the model—specifically, a tendency to exhibit bias toward either very high or very low actual values.

This outcome, however, may be affected by the spatial placement of the traffic count sensors. Since the cameras are placed in localized positions, they may fail to capture the broader building context. For instance, a camera might register a person exiting a building, counting it as an interzonal trip, while the person remains within the immediate area for activities like buying coffee in a different part of the building and moving from outside.

Such a possibility is highlighted in Figure 3.12, where the historical traffic count shows nearly equal volumes in both directions, possibly indicating intrazonal cycling or minor repositioning. To address this ambiguity, it is recommended to install cameras at both building entrances and corresponding bike parking areas to better differentiate between true interzonal trips and local activities.

Furthermore, travel demand during weekends, holidays, and exam periods is depicted in Figure 4.9, Figure 4.10, and Figure 4.11, respectively. These figures demonstrate the influence of the dummy variables implemented in the model. If the external zone trips appear similar, it is because the method uses the average values from the same scenario.



(a) Attraction (b) Production

Figure 4.9: Trip Generation Saturday at 7:55

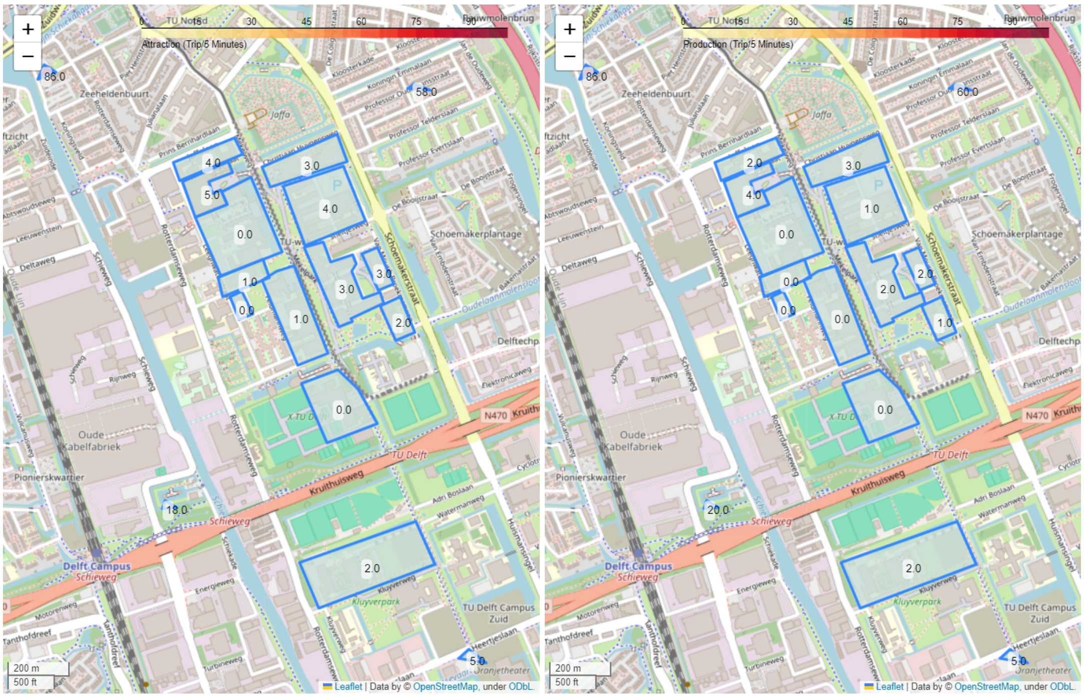
When compared with Figure 4.7, which represents a regular Monday, the figure for Saturday (Figure 4.9) shows significantly lower traffic volumes across all zones. This aligns with the expectation that weekend demand is lower due to fewer campus activities, validating the weekend dummy's effectiveness in capturing this behavioral shift.

Similarly, Figure 4.10, recorded on a holiday that also falls on a Monday, illustrates an even more pronounced drop in trip generation, reflecting the model's ability to represent holiday-induced demand reduction.

During the exam period (Figure 4.11), the model shows a modest increase in volume compared to regular weekdays. While not excessively high, the uptick is consistent with expectations that more people may visit the campus for exams, including those who do not usually attend on regular days.

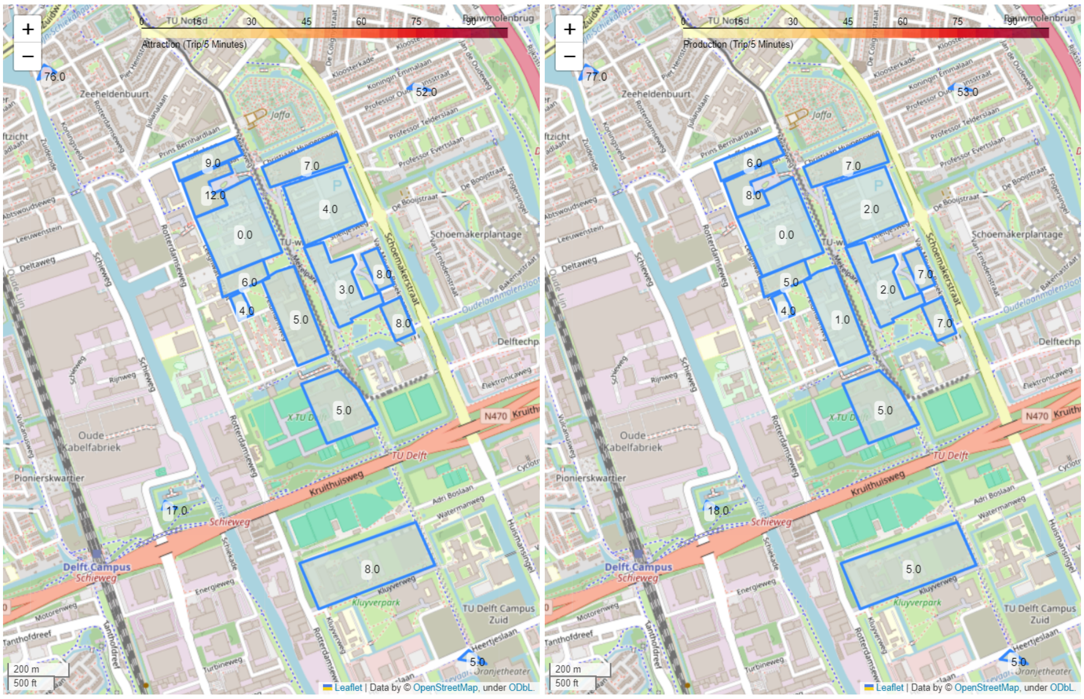
In conclusion, although the current trip generation model may still fall short in predictive accuracy, it successfully captures the hypothesized travel patterns embedded in the model through its variables. For instance, the observed decrease in traffic during weekends and holidays, as well as increased activity during exam periods, are well reflected.

Given this promising pattern recognition, the model's scalability should be enhanced. This can be achieved by incorporating more data points, either by increasing the number of observation locations or by enriching the data collected at existing locations. Additional contextual data—such as localized weather conditions, temporary disturbances, or event schedules—could significantly improve the model's precision and applicability.



(a) Attraction (b) Production

Figure 4.10: Trip Generation Holiday Period at 7:55



(a) Attraction (b) Production

Figure 4.11: Trip Generation Exam Week at 7:55

4.3. Trip Distribution

The trip distribution follows the zoning system that was determined previously. It has to be acknowledged that the production and attraction of the TU Delft campus also depend on home-bound trips. The home-bound trips are addressed by introducing external zones, as discussed in the previous section. While these external zones are not specified in detail, their production and attraction volumes are linked to the traffic counts recorded for flows outside the defined study area. However, it should be emphasized that external zones are not the primary focus of this research, and their effects are only partially represented.

The interpretation of the OD matrix represents an effort to examine the pattern of attraction distribution within a small-scale area like the TU Delft campus. Since distance and impedance are considered negligible in this context, the initial skim matrix for trip distribution is estimated in the same way as explained in the method.

4.3.1. Skim Matrix

The skim matrix is estimated using the same methodology as described in Section 3.2.3.1. Initially, the skim values for intrazonal trips and trips between external zones are assumed to be zero. This assumption is made due to the absence of meaningful impedance within those pairs and the lack of observable data to support alternative estimation.

During the Iterative Proportional Fitting (IPF) process, however, the final Origin-Destination (OD) matrix may still result in non-zero intrazonal trips. This is likely due to inconsistencies between the marginal totals and the row-column totals. While these results may not strictly represent actual movement within zones, they could reflect localized circulation patterns around key points, as previously noted in Section 4.2.3.

Nevertheless, there is a possibility that these non-zero intrazonal values are statistically insignificant—too small to be interpreted as actual trips—and can be considered negligible in the context of physical mobility.

For external-to-internal and internal-to-external zone movements, scaling factors are applied based on sampled traffic counts. For instance, on Monday at 07:55, using external indicators `['TUD_SPX18', 'TUD_SBX01', 'TUD_SBXS12']` and internal indicators `['TUD_SPXS06_IO', 'TUD_SBXN8A', 'TUD_SBX16']`, it is observed that the outbound (internal-to-external) volume is approximately 1.5 times greater than the inbound (external-to-internal) volume. This factor is used to adjust the scaling of external traffic volumes in the skim matrix accordingly.

4.3.2. Trip Distribution Result and Discussion

Based on the previously constructed skim matrix, the Iterative Proportional Fitting (IPF) method is implemented to obtain the final Origin-Destination (OD) matrix.

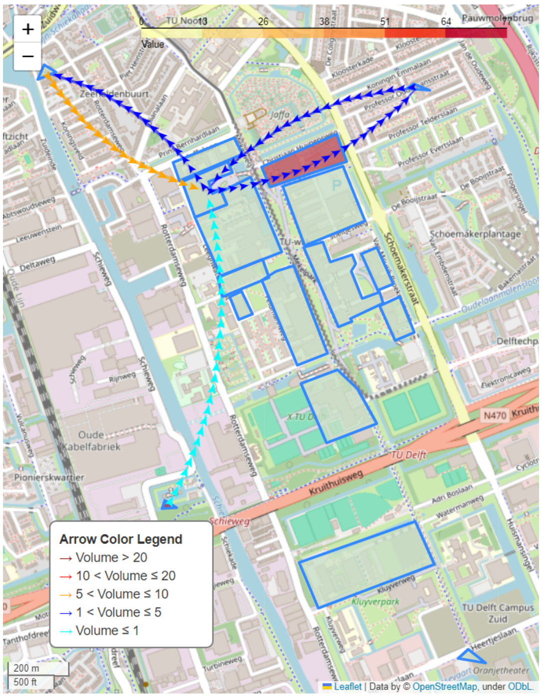
It is important to note that the production and attraction values for the internal zones are estimated using the developed model, while those for the external zones are directly derived from traffic count data. This introduces potential inconsistencies, as the traffic count data reflect movements at the network level rather than the broader macro-scale flows typically considered in trip distribution modeling.

These limitations in data quality and granularity inevitably affect the model's accuracy. Moreover, it must be acknowledged that no formal validation procedure is conducted in this project due to data unavailability, which falls outside the scope of the current study.

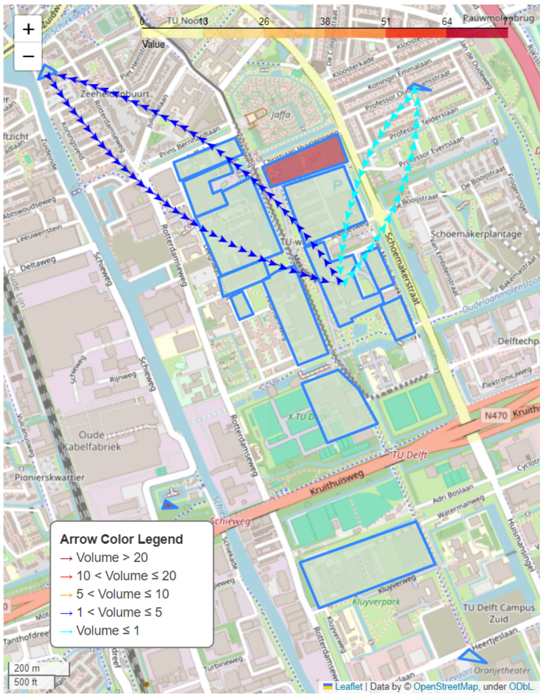
Despite these constraints, the performance of the trip distribution model is evaluated through qualitative analysis—namely, by examining the spatial and temporal patterns generated by the OD matrix and comparing them across different contextual scenarios.

The trip distribution is conducted for all zones; however, to evaluate the model's performance in greater detail and to enable clearer interpretation, the resulting OD matrix is visualized with a focus on selected key zones. For instance, Figure 4.12 displays all OD pairs associated with Zone 2 (IDE), Zone 10 (CEG), and Zone 11 (ECHO)—which were used as the basis for trip generation estimation—as well as Zone 13 (Library and Aula), which functions as a central hub of campus activity. In the figure, different colors indicate the number of bicycle units observed: cyan represents 1 unit, blue represents 2–5 units, and orange represents 5–10 units. Figure 4.12 illustrates that the trips associated with each zone are influenced by the variables used in constructing the skim matrix. Zones with higher values for these variables tend to attract or produce more trips. For example, trips to the IDE building are notably higher, as shown in Figure 4.12d, which aligns since it has the potential and also the research use this building as the main data. Furthermore, the morning period shown corresponds to predominantly home-bound trips originating from external zones. Overall, the patterns depicted are consistent with reasonable expectations based on the model assumptions.

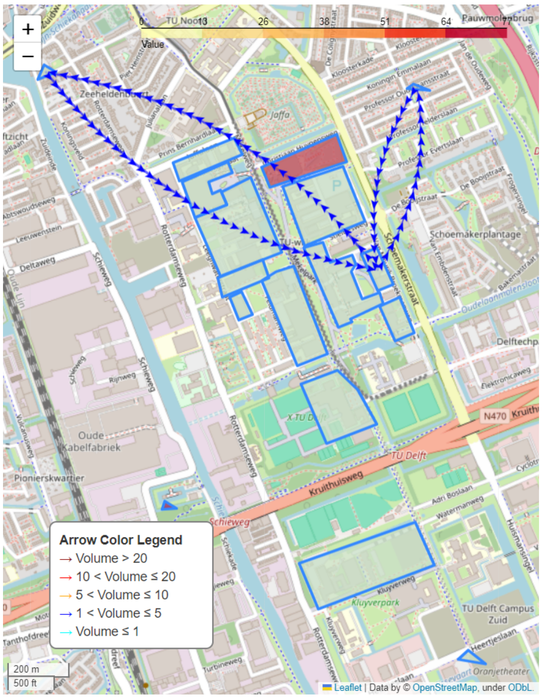
The influence of time on trip distribution is examined in Figure 4.15, with Zone 2 serving as the central reference. Although the interzonal relationships in the skim matrix are static—based on fixed spatial and infrastructural data—the resulting OD patterns change over time due to variations in trip generation and attraction.



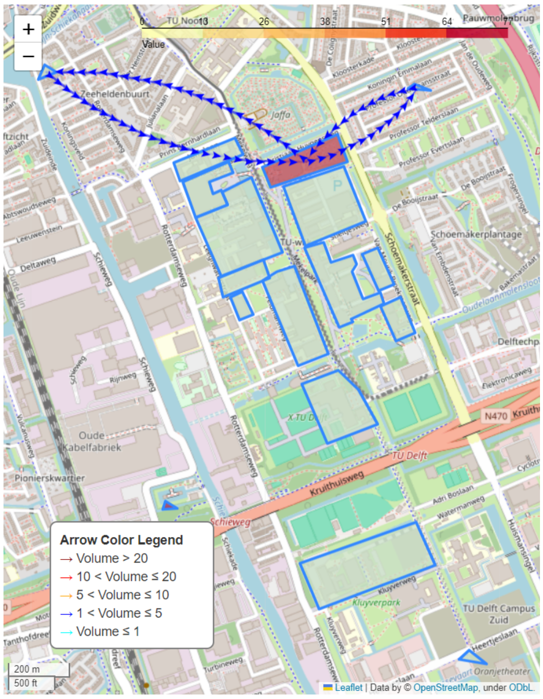
(a) Origin Destination Related to Zone 2 (Normal Monday at 7:55)



(b) Origin Destination Related to Zone 10 (Normal Monday at 7:55)



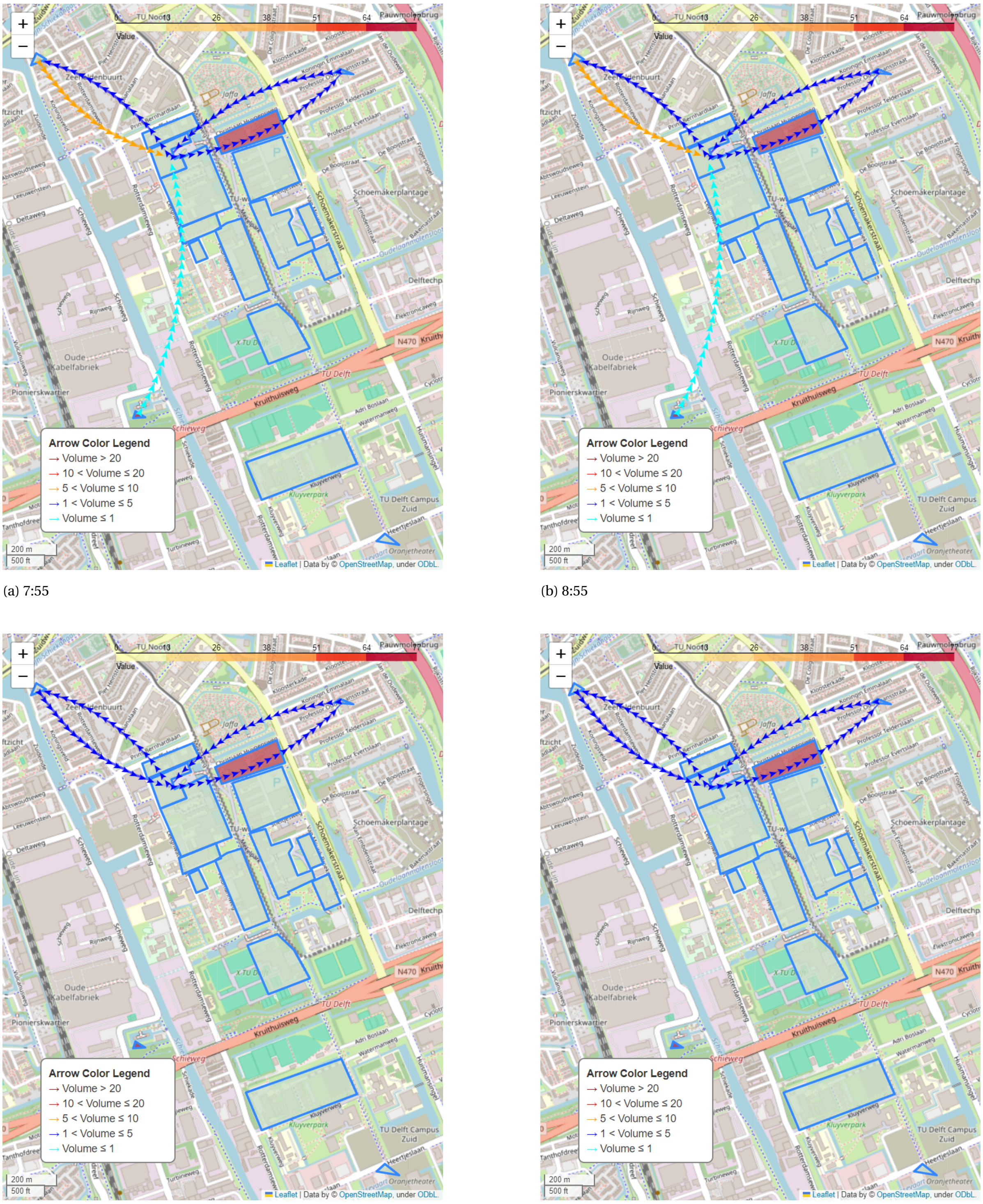
(c) Origin Destination Related to Zone 11 (Normal Monday at 7:55)



(d) Origin Destination Related to Zone 13 (Normal Monday at 7:55)

Figure 4.12: Origin-Destination Volume on Normal Monday at 7:55

It can be observed that traffic volume does not vary significantly in the morning but decreases noticeably in the evening, aligning with the expected patterns of campus activity. In the morning, external zones play a dominant role in trip generation, primarily due to commuting behavior. However, since trips in external zones are determined based on the mean traffic count from their respective sensors, this averaging method also influences the distribution of trips. Nevertheless, it is interesting to note that the morning and evening patterns differ in a way that aligns with the initial hypothesis.



(a) 7:55

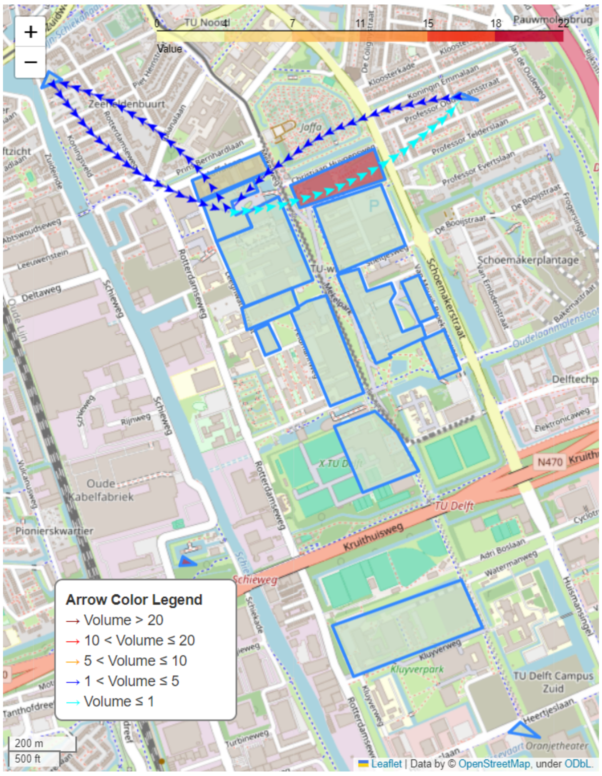
(b) 8:55

(a) 16:55

(b) 17:55

Figure 4.15: Origin-Destination Volume Related to Zone 2 within a Day

While most interzonal trips appear to be bidirectionally balanced—for example, trips from Zone 2 to Ex-Zone 1 are generally similar to those in the opposite direction—some notable exceptions exist. These asymmetries may arise from the initial skim matrix values or from differences in production and attraction estimates based on traffic counts. The assumption of bidirectional symmetry does not always hold, especially in zones with unique access characteristics. Additionally, the similarity in production and attraction volumes across many zones—reflected in their traffic counts, which differ only slightly (typically by a range of 2–5



trips)—may also contribute to this effect.

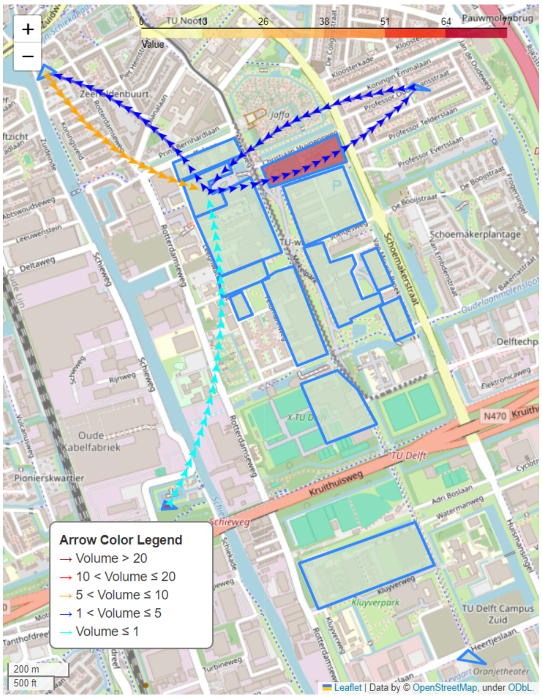
One example is External Zone 3 in the northwest, which is connected to both Delft Station and nearby residential areas. This zone shows high production and attraction volumes in the morning, reinforcing the hypothesis that external zones significantly contribute to inbound flows during typical commute hours.

The influence of different events and scenarios on trip distribution is further explored in Figure 4.16. Figure 4.16a represents a baseline "business-as-usual" case, captured on a Monday during the regular lecture period. The figure clearly shows that IDE is an attractive destination due to the bias of the data, attracting a high number of trips.

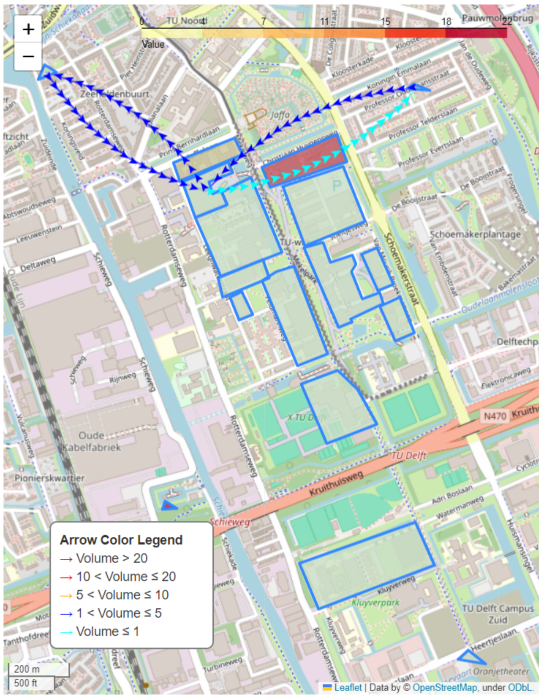
Figure 4.16b illustrates trip patterns on a weekend. As expected, overall traffic decreases, and the remaining trips primarily originate from residential areas or the campus station. This behavior aligns with the model's assumptions and reflects reduced campus activity on non-working days.

A similar trend is observed during the holiday period, as shown in Section 4.3.2. The pattern of travel remains focused on residential and station areas, but with significantly lower volume than on a regular weekday.

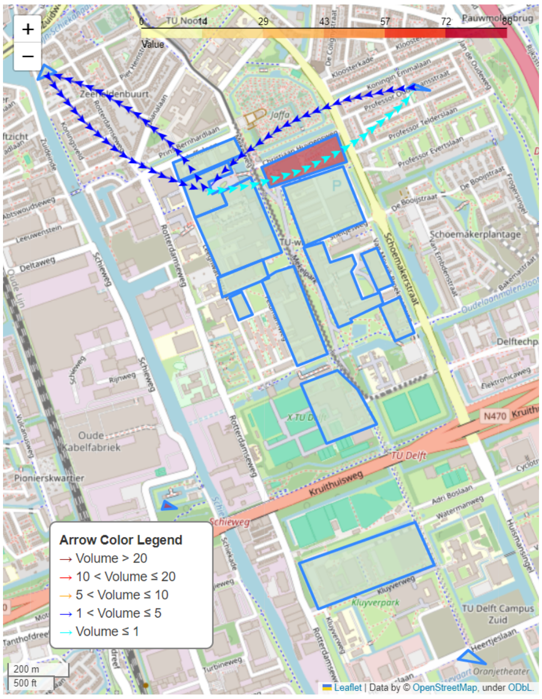
In contrast, a modest increase in traffic is observed during the exam period, which aligns with the hypothesis that campus activity intensifies slightly due to students who do not regularly attend lectures but travel to campus for exams or study purposes. Overall, while the trip distribution still has significant room for improvement, the general patterns are clearly observable. Incorporating new variables—such as lecture schedules—could help generate a more accurate initial skim matrix and enhance the overall model performance.

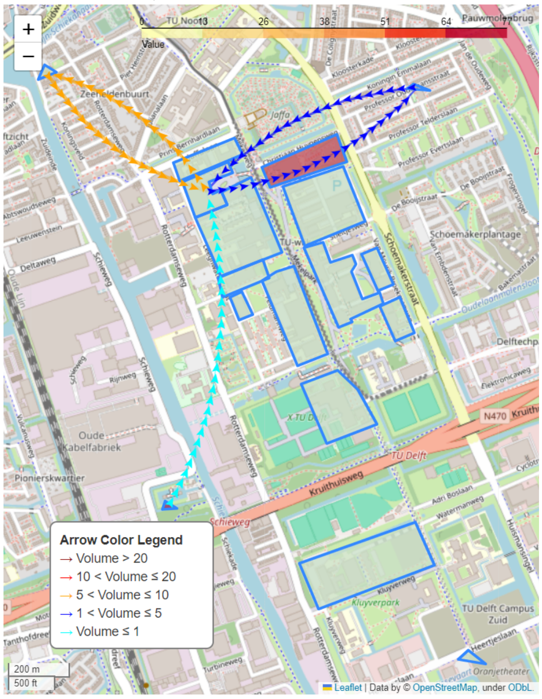


(a) Normal Monday



(b) Normal Saturday





(c) Exam

Figure 4.16: Origin-Destination Volume Related to Zone 2 with Different Occasions

5
Discussions

The previous chapter has already produced some results to be discussed. Chapter 2 explored the required data and existing methods for OD matrix estimation, primarily focusing on transport modeling in broader contexts, later the purpose of the research directs the method specification of the OD matrix estimation. The specification of the model has been depicted and described in Chapter 3, where the available data within the TU Delft campus was translated into a workable methodology for analysis.

The OD estimation has been formed using the defined method, with trip distribution conducted based on the zoning system and assumptions about external zones. The results of this analysis are depicted in Chapter 4.

Each chapter's results provide insights into the research aims. The main aim is to develop a modeling process and specifications that are compatible with the available data while establishing a foundation for future improvements.

5.1. Model Specification

In this section, the data collection and the method framework will be evaluated, and if it is applied, a further suggestion might be raised in the explanation.

5.1.1. Data Collection

The data commonly used in the literature primarily relies on traffic count data. Similarly, this research utilizes traffic count data for the trip generation process, specifically for estimating trip production and attraction volumes. In some cases, models such as 24-hour models (Rob van Nes, 2021) are referenced, where traffic count data is used for validation through the process of trip assignment. However, this research does not conduct the traffic assignment and focuses solely on the development of the OD matrix. This differing perspective from the literature introduces both shortcomings and potential in the research, offering unique insights while acknowledging the limitations in scope and methodology.

The type of data gathered for this research is uncommon for traditional modeling approaches, primarily due to the specificity of the case study. While this should not pose a significant issue, as the method can be contextualized to fit the available data, it does highlight certain limitations. The reliance on less conventional data implies inherent constraints, which may affect the accuracy of the model. Consequently, the expected precision of the outcomes is likely to be moderate rather than high, reflecting the challenges posed by the data limitations.

Furthermore, to emphasize the importance of temporal-spatial analysis in this research, it is crucial to develop data that supports this dimension. The lack of such data introduces challenges in the modeling process, particularly in the trip generation phase, where the absence of detailed temporal-spatial inputs can lead to difficulties and potential misinterpretations in capturing the dynamics of travel demand accurately.

Adding more traffic count observations at the cordon would be highly beneficial for improving the accuracy and granularity of the model. For instance, placing a traffic counter at the Stieltjesweg intersection could provide valuable data on trips originating from the residential area on the eastern side of the campus. This would enhance the model's ability to capture flows from areas currently underrepresented.

Additionally, increasing the number of traffic count points along Mekelweg would help refine the zoning system by enabling the creation of smaller zones. This would provide a more detailed representation of intra-campus traffic flows, allowing the model to capture finer variations and improve its capacity to describe and predict traffic patterns within the campus. Such improvements would strengthen the model's reliability and its application in network development and planning.

In addition to traffic count observations, incorporating data that reflects the relationships between zones is crucial for improving the trip distribution process. Currently, the trip distribution assumes a uniform skim matrix, which may not fully represent the underlying dynamics of travel patterns.

Introducing skim matrices based on more relatable variables, like the lecture schedule, is preferred. This suggests that more directly related data is needed to capture the actual interactions between zones effectively.

For example, incorporating lecture schedules into the model could significantly enhance its accuracy. Lecture schedules provide insights into the movement of students between buildings, reflecting real-world origin-destination relationships. By leveraging such data, the trip distribution could better represent the temporal and spatial dynamics of campus travel, leading to a more reliable and meaningful OD matrix.

5.1.2. Method Framework

The four-step model is widely recognized as a reliable framework for travel demand estimation due to its capacity to produce detailed outputs, such as traffic assignment, while offering macroscopic-level insights through trip distribution. However, for this research's context, improvements to the model's applicability are necessary to address specific requirements.

While the literature indicates that bicycle transport models often rely on direct demand methods, the goal of this research is to develop an OD matrix capable of reflecting changes in OD flows under varying circumstances, such as holidays, events, or high/low peak volumes. The four-step model's structured framework enables the evaluation of variable influences and their relationships with OD flows, making it a suitable choice for this aim.

This research contextualizes existing methodologies by incorporating specific variables not typically found in socio-demographic census data. For example, by referencing (Baumanis et al., 2023), it adapts the four-step model to unique variables relevant to the campus environment, even though it is outside the context of COVID-19 studies. Similarly, it aligns with (Lu et al., 2018) by employing regression methods to estimate link volumes, further tailoring the approach to the study's objectives.

The resulting methodology combines Ordinary Least Squares (OLS) regression for trip generation with Iterative Proportional Fitting (IPF) for trip distribution. However, as with any four-step model, a key limitation is the potential propagation of errors across its stages, which must be carefully managed during implementation to ensure reliable results.

5.2. Implementation

The shortcomings of the trip production and trip generation model are evident, particularly in the zoning system.

To accommodate the location of the smart camera, the zones had to be made significantly large, which compromises the granularity of the analysis. Additionally, limitations are reflected in the linear regression process. Highly correlated variables, such as student data, were initially included as representations of demand (e.g., student and employee data) and supply (e.g., parking and lecture capacity).

It is shown in the OLS model that shows bad performance with only about 30% explainability. The underfitting most likely happened because the model shows a better R^2 when the train sample size gets smaller.

This observation is critical, as the output of the trip generation process serves as an input for the trip distribution stage. Any errors or unexplained variance in the trip generation model may propagate through subsequent stages, potentially impacting the overall accuracy of the OD matrix and the reliability of the final results.

These errors propagate through the trip distribution phase, compounding the effects of assumptions made during the process. While most assumptions are reasonable, instances of over- or undersimplification can occur, further affecting the accuracy and reliability of the model. However, this propagation can be handled if the validation process is done correctly. A refined approach that addresses these issues is necessary to improve the robustness of both the trip generation and distribution processes. The implementation of this method will be fitted into more realistic concept on the next chapter.

The validation of the OD matrix is crucial to minimize errors and inaccuracies. This is particularly important given that the four-step model involves multiple subprocesses, making error propagation inevitable.

However, direct validation of the OD matrix is not possible due to the lack of data that specifies the departure and arrival points of individuals at TU Delft. Validation can only be achieved if OD pair data becomes available. One approach is to conduct a survey, which would allow for direct validation of the OD pairs.

The drawback of this method is that surveys require significant time and effort. Additionally, survey data often involve stated preferences, introducing another layer of potential inconsistencies that could affect the reliability of the validation process.

Since direct validation of the OD matrix is not possible, an alternative approach is to aggregate the matrix and validate its sum against the closest traffic count. This method is partially feasible, especially in locations like TU Delft, where certain buildings, such as ECHO, have smart cameras that directly count bicycle traffic.

However, this approach is not suitable for the specific case study due to the size of the zones used in the analysis. The parking area of ECHO, for instance, is not representative of the broader zone it belongs to, limiting the applicability of traffic count data from such localized sources for validating aggregated matrix flows.

Another indirect method for validating the trip distribution results is through traffic assignment. The validation of the traffic assignment itself can serve as an indirect validation of the trip distribution. However, this approach carries a risk of error propagation, as the traffic assignment process involves several decisions, such as the type of traffic assignment used or whether congestion is considered, which would necessitate using an equilibrium method.

These choices can introduce additional uncertainties, complicating the validation process and potentially amplifying any inaccuracies in the original trip distribution model. While this method provides valuable insights, careful consideration of these factors is essential to minimize the impact of propagated errors.

Thus, implementing validation correctly is highly challenging due to the lack of data. However, the most feasible solution in the current situation is to correlate the OD pairs with the production or attraction volumes of a specific building. While this approach provides a partial validation, it is limited in scope and accuracy.

The ideal solution would be to enhance the data library by adding more comprehensive data sources. This enrichment would not only improve the reliability of the validation process but also open up possibilities for alternative methods or entirely new approaches to address the problem more effectively.

However, there are several apparent advantages to this modeling framework and pipeline:

- **Simplicity and flexibility:** The model is simple and easy to develop due to its foundation in the classic four-step modeling approach. This structure allows for high adaptability, making it straightforward to modify or add parameters as needed. For example, if the CREFM aims to analyze the effect of increasing bike parking spots in specific areas, the impact can be directly assessed by updating the input data or examining the related regression coefficients. This flexibility supports scenario analysis and policy testing in a practical and transparent way.

- **Ability to capture known travel patterns:** The model effectively captures well-established travel behavior patterns, such as variations in traffic volume by time of day, lecture hours, exam periods, and weekends. These patterns are clearly represented in the results, making the model both interpretable and actionable. As a result, it becomes easier to translate insights into campus mobility policies, such as adjusting infrastructure (e.g., increasing parking availability) or managing campus population (e.g., limiting student numbers during peak periods).

6

Conclusion and Recommendation

In summary, this research focused on analyzing bicycle travel demand at a small scale, using the TU Delft campus as a case study. The model was developed under data-limited conditions, contrasting with typical bicycle travel demand analyses, which are usually conducted at a city-wide scale. This study aimed to contextualize bicycle travel demand analysis specifically for a university campus. Ultimately, it provides a step-by-step framework for implementing such an analysis, offering insights into adapting travel demand methodologies to smaller, more localized environments.

The answer of the research question is discussed in Section 6.1. Further recommendations based on previous chapters are also depicted in Section 6.2

6.1. Answering Research Question

This research aims to establish a travel demand estimation process using the available data and methods, with the ultimate goal of supporting the development planning of the TU Delft campus bicycle network. The model's effectiveness is realized when it demonstrates a certain level of sensitivity to changes in key variables, allowing it to detect and reflect their impact on the network.

"How can a reliable method for estimating cyclists' travel demand be developed to describe spatiotemporal patterns and support the planning of the bicycle network at the TU Delft Campus?"

To answer this main research question the formulated sub-research questions are explained to formulate the complete answer of the main research question.

1. What are the needed data and existing methods for estimating cyclist travel demand at the TU Delft campus?
   This sub-research question aims to identify commonly used data and state-of-the-art methods for OD matrix estimation in the context of bicycle networks. The findings are detailed in Chapter 2, where the available data in the TU Delft campus environment is compared to insights from the literature.

   It is evident that traffic count data is a key component in bicycle transport models. While it is often used as validation in traffic assignment, there are cases where it serves as training data, particularly for direct demand models. Beyond traffic counts, data related to demand, such as population, and supply-side data, such as network characteristics or facility capacities, are also commonly employed.

   For this research, the scope has been contextualized by focusing on data that can express both demand and supply within the TU Delft campus. This includes specifically tailored data, such as campus-specific facility capacities. Additionally, weather data is incorporated due to its influence on the motivation to cycle (Franco et al., 2014). By integrating these datasets, the study aligns with existing methodologies while adapting to the unique context of the campus.

   Furthermore, it is observed that in the context of transport modeling, the direct demand approach is widely used, particularly because its output often focuses on link volume models, which are convenient and straightforward to implement. However, insights from the literature reveal aspects that can be contextualized for this research.

For instance, Baumanis et al. (2023) developed a four-step model incorporating machine learning regression to analyze bike traffic during COVID-19, utilizing data specific to the pandemic. This approach inspires the integration of campus-specific data around TU Delft, reflecting the local context and dynamics. Similarly, Lu et al. (2018) created a 24-hour model by incorporating regression techniques to account for traffic variability throughout the day. This approach influences the inclusion of time-related equations in the research, highlighting the importance of temporal dynamics in modeling bicycle travel demand. These contextualized methods provide a foundation for tailoring the OD estimation process to the specific needs of the TU Delft campus.

The objective of developing a model capable of detecting changes in traffic volume based on variations in key variables, while also identifying and explaining the significance of these variables, reinforces the decision to adopt a four-step model with a dynamic temporal dimension. This approach ensures the model can capture temporal variations in travel patterns, providing actionable insights for understanding the factors influencing bicycle demand. Later on, with enhancement, the model can be used as the development planning tools.

2. What model specifications are necessary to apply travel demand modeling effectively for the TU Delft campus bicycle network development?
   This question explores estimation methods compatible with the available data sources within the TU Delft. The answer to this sub-research question is discussed in Chapter 3. While macroscopic-level data may not directly align with the research context, data preprocessing allows it to be combined with specific datasets, making it more applicable. Open data sources play a key role in this research, representing both demand and supply aspects.

   On the supply side, many variables reflect capacity, such as bike parking spots, study places, and lecture halls. These variables are tied to specific locations and act as potential attractors, influencing why individuals travel to certain buildings or zones within the campus.

   Demand, on the other hand, is represented by the population of students and employees on the TU Delft campus. Weather also indirectly represents demand, as it can influence mode choice and the decision to travel or stay at home.

   Given the dynamic approach used in this research, time variables are incorporated into the model through dummy variables. Time is modeled as a binary variable that represents the minutes of an hour, for example (45-50 for 7.45-7.50). This decision is made to observe the influence of the time that is correlated to the activities (lecture) on the bike volume. Additional time-related variables, such as exam periods, break time, and holiday periods, are included because they significantly influence travel patterns, as observed in Chapter 3. This comprehensive approach enables the model to account for both temporal and contextual dynamics in travel demand analysis. The data also shows a pattern of fluctuation in a very short time and analysis determines that a 5-minute aggregation of the bike volume is sufficient to capture the pattern and easier to implement.

   As explained in Section 3.2, this research employs the four-step model to capture the origin-destination relationships within the campus and between campus and external areas. While direct demand models are commonly used for their simplicity, and agent-based models (ABM) are favored for their activity-focused detail, the four-step model was chosen for its ability to provide macroscopic insights, particularly in a campus-scale context.

   The implementation leverages specific variables tailored to the study's scope. The two steps of the four-step model employed are trip generation and trip distribution. Trip generation is conducted using linear regression, achieved through backward stepwise elimination with the Ordinary Least Squares (OLS) technique. For trip distribution, the iterative proportional fitting (IPF) method is applied, with the initial skim matrix assumed from the activity-related data and traffic count due to the negligible impedance to travel within the campus area. This approach ensures that the model remains adaptable and relevant to the specific dynamics of the TU Delft campus.

3. How is the origin-destination flow estimated for the TU Delft bicycle network?
   The OD flow, represented by OD matrix, helps analyze cyclist flow and demand. This question includes the assumptions and distribution patterns of OD pairs and is discussed in Chapter 4.

   The OD flow estimation, detailed in Chapter 4, was executed by forming OD matrices for given the time determined to identify patterns within them. This process involved trip generation and trip distribution, both of which were preceded by data preparation. Data preparation included the establishment of a zoning system based on the traffic count around the cordon.

Trip generation was conducted using linear regression with a backward stepwise approach, employing OLS techniques. The regression was designed to estimate the morning attraction of the zone and evening production of the zone as it can be related to the purpose of the trip, with the resulting equation that has time binary variables based on the minutes of hours. However, due to a lack of data related to the model from external zones, the model's accuracy was limited in this aspect and the model focuses more on the trip volume of the internal zones, though the flow proportions still provided valuable insights.

Trip distribution was performed using the Iterative Proportional Fitting (IPF) method, applying the initial skim matrix -produced based on the lecture hall capacity, bike parking spots, and the traffic count- to distribute flow among zones.

From the research, the OD matrix estimation faced several shortcomings. There is an indication of underfitting with large data but less data as predictors. Multicollinearity was identified in the linear regression due to constant capacity-related variables but ignored.

Additionally, the trip distribution suffered from inaccuracies due to the lack of direct data indicating flow relationships between zones. These limitations highlight the need for improved data and alternative approaches to enhance the reliability of the OD flow estimation process with the addition of a validation method.

However, the framework provides a balance between interpretability and flexibility, enabling both accurate representation of observed travel behavior and ease of scenario testing—making it a practical tool for supporting data-driven mobility planning and policy evaluation on campus.

To address the main research question, the model can be developed through targeted data collection focused on variables directly influencing bicycle travel demand, such as population, infrastructure capacity, and environmental conditions like weather. The dynamic temporal approach is essential, with data aggregated into time intervals that align with the model's purpose. Given the high variability observed in real-time bicycle traffic counts, a finer temporal resolution is preferred. This aligns with practical observations, as bicycle congestion around TU Delft often occurs for brief periods and needs to be captured effectively.

Traffic count data serves as the ground truth for the model, providing a reliable baseline for calibrating and validating outputs. For trip generation, Ordinary Least Squares (OLS) regression is used to estimate production and attraction volumes, establishing relationships between key variables and travel demand. Validation of the model is conducted using Root Mean Square Error (RMSE) and comparison plots between modeled and observed data, ensuring the model accurately reflects real-world traffic patterns.

The trip distribution process employs Iterative Proportional Fitting (IPF) to formulate the OD matrix. Uniform skim matrices are utilized, as impedance can be reasonably ignored for the short trips characteristic of campus travel. This approach integrates data collection, modeling, and validation, creating a robust travel demand estimation model tailored to the unique dynamics of the TU Delft campus and providing valuable insights for bicycle network planning.

6.2. Recommendation

The recommendation for future research is to explore and implement additional approaches in linear regression to achieve better results. Techniques such as Ridge Regression, LASSO, or machine learning-based regression methods could be used to improve model accuracy, address multicollinearity, and enhance the robustness of the trip generation process. A sophisticated model also can be related to the ability to explain the data with less input.

Validation of the trip distribution is another critical area that requires attention, as it was not fully addressed in this research due to time and scope limitations. Establishing a reliable validation framework, potentially through detailed surveys or advanced traffic count data with directional information, would significantly improve the credibility of the results.

Additionally, more comprehensive data collection is essential. This would provide a broader range of variables and enable the development of better-informed assumptions, ultimately resulting in more accurate and insightful travel demand analyses for the TU Delft campus and similar contexts.

Bibliography

Aldred, R., Woodcock, J., & Goodman, A. (2016). Does More Cycling Mean More Diversity in Cycling? *Transport Reviews*, *36*(1), 28–44. https://doi.org/10.1080/01441647.2015.1014451

Arellana, J., Saltarín, M., Larrañaga, A. M., González, V. I., & Henao, C. A. (2020). Developing an urban bikeability index for different types of cyclists as a tool to prioritise bicycle infrastructure investments. *Transportation Research Part A: Policy and Practice*, *139*, 310–334. https://doi.org/10.1016/j.tra.2020.07.010

Baumanis, C., Hall, J., & Machemehl, R. (2023). A machine learning approach to predicting bicycle demand during the COVID-19 pandemic. *Research in Transportation Economics*, *100*, 101276. https://doi.org/10.1016/j.retrec.2023.101276

Bhowmick, D., Saberi, M., Stevenson, M., Thompson, J., Winters, M., Nelson, T., Leao, S. Z., Seneviratne, S., Pettit, C., Vu, H. L., Nice, K., & Beck, B. (2023). A systematic scoping review of methods for estimating link-level bicycling volumes. *Transport Reviews*, *43*(4), 622–651. https://doi.org/10.1080/01441647.2022.2147240

Blume, S. O., Corman, F., & Sansavini, G. (2022). Bayesian origin-destination estimation in networked transit systems using nodal in- and outflow counts. *Transportation Research Part B: Methodological*, *161*, 60–94. https://doi.org/10.1016/j.trb.2022.04.006

Cooper, C. H. V. (2018). Predictive spatial network analysis for high-resolution transport modeling, applied to cyclist flows, mode choice, and targeting investment. *International Journal of Sustainable Transportation*, *12*(10), 714–724. https://doi.org/10.1080/15568318.2018.1432730

Dadashova, B., Griffin, G. P., Das, S., Turner, S., & Sherman, B. (2020). Estimation of Average Annual Daily Bicycle Counts using Crowdsourced Strava Data. *Transportation Research Record: Journal of the Transportation Research Board*, *2674*(11), 390–402. https://doi.org/10.1177/0361198120946016

Franco, L. P. C., Campos, V. B. G., & Monteiro, F. B. (2014). A Characterisation of Commuter Bicycle Trips. *Procedia - Social and Behavioral Sciences*, *111*, 1165–1174. https://doi.org/10.1016/j.sbspro.2014.01.151

Gosse, C. A., & Clarens, A. (2014). Estimating Spatially and Temporally Continuous Bicycle Volumes by Using Sparse Data. *Transportation Research Record*, *2443*(1), 115–122. https://doi.org/10.3141/2443-13

Guo, L., Huang, S., & Sadek, A. W. (2013). A novel agent-based transportation model of a university campus with application to quantifying the environmental cost of parking search. *Transportation Research Part A: Policy and Practice*, *50*, 86–104. https://doi.org/10.1016/j.tra.2013.01.045

Hamad, K., & Obaid, L. (2022). Tour-based travel demand forecasting model for a university campus. *Transport Policy*, *117*, 118–137. https://doi.org/10.1016/j.tranpol.2022.01.001

Harms, L., Bertolini, L., & Te Brömmelstroet, M. (2014). Spatial and social variations in cycling patterns in a mature cycling country exploring differences and trends. *Journal of Transport & Health*, *1*(4), 232–242. https://doi.org/10.1016/j.jth.2014.09.012

Heinen, E., Maat, K., & Wee, B. V. (2011). The role of attitudes toward characteristics of bicycle commuting on the choice to cycle to work over various distances. *Transportation Research Part D: Transport and Environment*, *16*(2), 102–109. https://doi.org/10.1016/j.trd.2010.08.010

Informatiepunt, I., & Centraal Bureau Voor De Statistiek. (2023). ONDERZOEK ONDERWEG IN NEDERLAND - ODiN 2022. https://doi.org/10.17026/SS/BXIK2X

Jacyna, M., Wasiak, M., Kłodawski, M., & Gołębiowski, P. (2017). Modelling of Bicycle Traffic in the Cities Using VISUM. *Procedia Engineering*, *187*, 435–441. https://doi.org/10.1016/j.proeng.2017.04.397

Jafari, A., Singh, D., Both, A., Abdollahyar, M., Gunn, L., Pemberton, S., & Giles-Corti, B. (2024). Activity-based and agent-based transport model of Melbourne: An open multi-modal transport simulation model for Greater Melbourne. *Journal of Intelligent Transportation Systems*, 1–18. https://doi.org/10.1080/15472450.2024.2372894

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

Khan, N. A., & Habib, M. A. (2023). Microsimulation of activity generation, activity scheduling and shared travel choices within an activity-based travel demand modelling system. *Travel Behaviour and Society*, *32*, 100590. https://doi.org/10.1016/j.tbs.2023.100590

Krishnakumari, P., Van Lint, H., Djukic, T., & Cats, O. (2020). A data driven method for OD matrix estimation. *Transportation Research Part C: Emerging Technologies*, *113*, 38–56. https://doi.org/10.1016/j.trc.2019.05.014

Lin, Z., & Fan, W. “. (2020). Bicycle Ridership Using Crowdsourced Data: Ordered Probit Model Approach. *Journal of Transportation Engineering, Part A: Systems*, *146*(8), 04020076. https://doi.org/10.1061/JTEPBS.0000399

Lu, T., Mondschein, A., Buehler, R., & Hankey, S. (2018). Adding temporal information to direct-demand models: Hourly estimation of bicycle and pedestrian traffic in Blacksburg, VA. *Transportation Research Part D: Transport and Environment*, *63*, 244–260. https://doi.org/10.1016/j.trd.2018.05.011

Ortúzar S., J. d. D., Willumsen, L. G., & Ortúzar S., J. d. D. (2011). *Modelling transport* (4. ed). Wiley.

Orvin, M. M., Fatmi, M. R., & Chowdhury, S. (2021). Taking another look at cycling demand modeling: A comparison between two cities in Canada and New Zealand. *Journal of Transport Geography*, *97*, 103220. https://doi.org/10.1016/j.jtrangeo.2021.103220

Oskarbski, J., Birr, K., & Żarski, K. (2021). Bicycle Traffic Model for Sustainable Urban Mobility Planning. *Energies*, *14*(18), 5970. https://doi.org/10.3390/en14185970

Rob van Nes. (2021, September). Transport Modelling - Trip generation [Lecture slides and recording].

Ryu, S. (2020). A Bicycle Origin–Destination Matrix Estimation Based on a Two-Stage Procedure. *Sustainability*, *12*(7), 2951. https://doi.org/10.3390/su12072951

Ryu, S., Su, J., Chen, A., & Choi, K. (2019). Estimating Bicycle Demand of a Small Community. *KSCE Journal of Civil Engineering*, *23*(6), 2690–2701. https://doi.org/10.1007/s12205-019-0415-5

TU Delft Campus. (2023). Campus Vision 2040. https://www.tudelftcampus.nl/wp-content/uploads/2024/03/9278-Campusvisie-2040-ENG-DIGITAAL.pdf

van Dulmen, A., & Fellendorf, M. (2021). Regional Bicycle Network Evaluation and Strategic Planning: A Quantitative Methodological Approach Despite Limited Data Sources for Cycling. *Transportation Research Record*, *2675*(12), 306–316. https://doi.org/10.1177/03611981211028870

Weiner, E. (2008). *Urban Transportation Planning in the United States: History, Policy, and Practice*. Springer New York. https://doi.org/10.1007/978-0-387-77152-6

Xu, Z., Kang, J. E., & Chen, R. (2018). A random utility based estimation framework for the household activity pattern problem. *Transportation Research Part A: Policy and Practice*, *114*, 321–337. https://doi.org/10.1016/j.tra.2018.01.036

A

Appendix: Traffic Count Analysis Graphs

In this appendix part, all the plot and graph that is summarized from the traffic count spots are depicted. It shows a result on assumption and variable of time.

Figure A.1: Traffic count analysis: TUD_SBX01

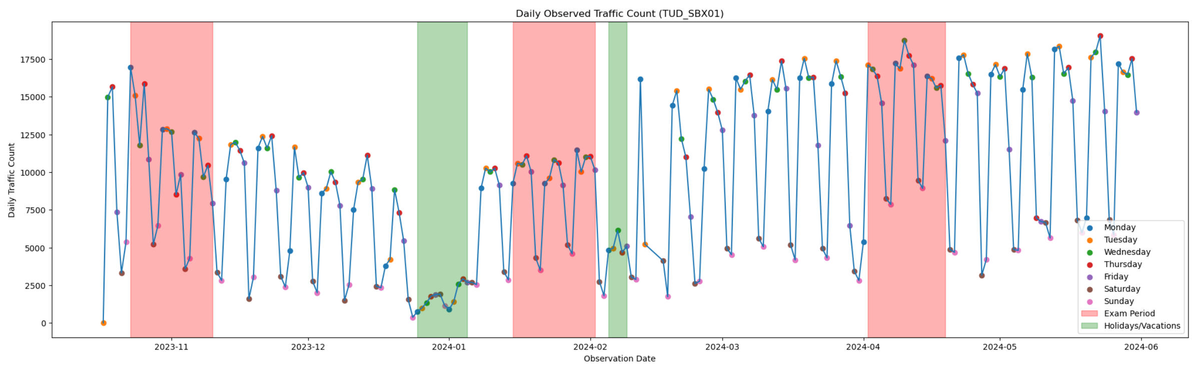


Figure A.2: Daily observed fluctuation in traffic count: TUD_SBX01

Figure A.3: Traffic count analysis: TUD_SBX06B

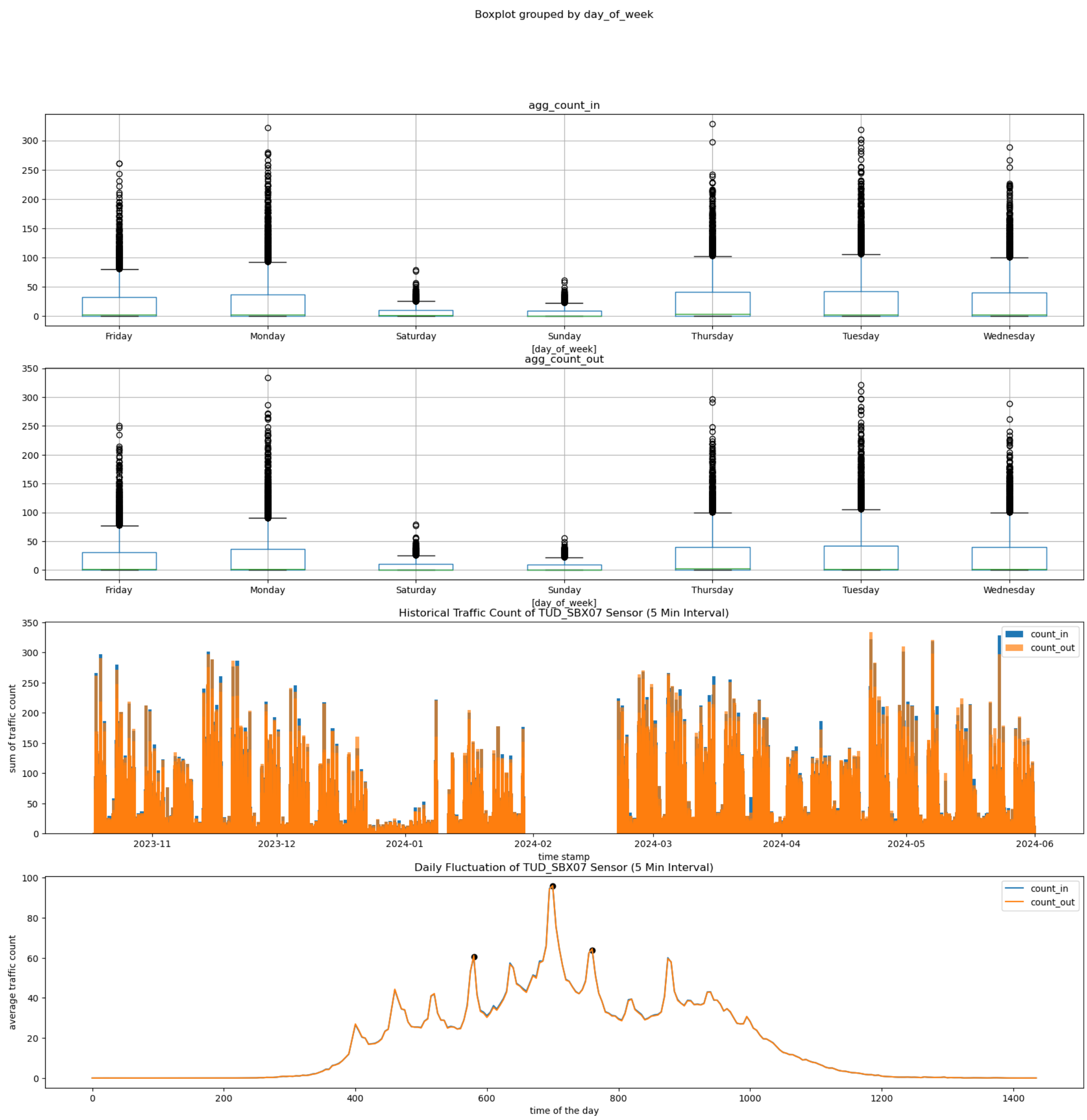


Figure A.4: Traffic count analysis: TUD_SBX07

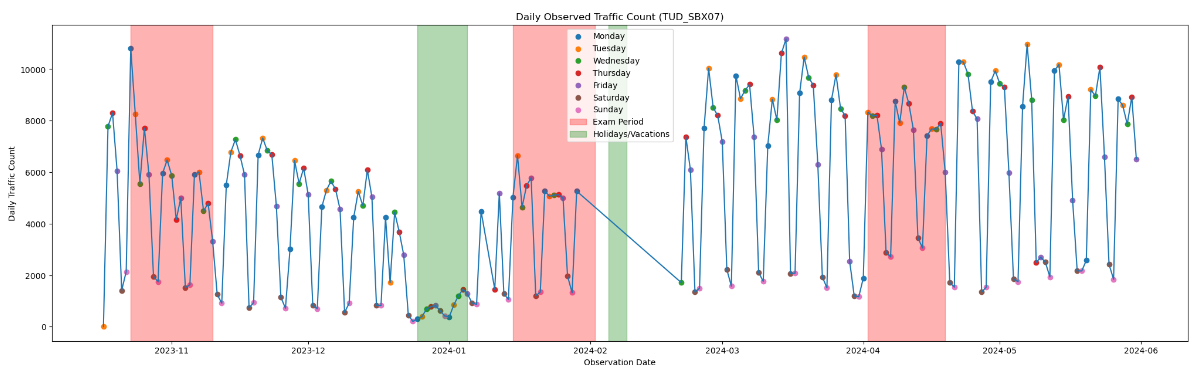


Figure A.5: Daily observed fluctuation in traffic count: TUD_SBX07

Figure A.6: Traffic count analysis: TUD_SBX16

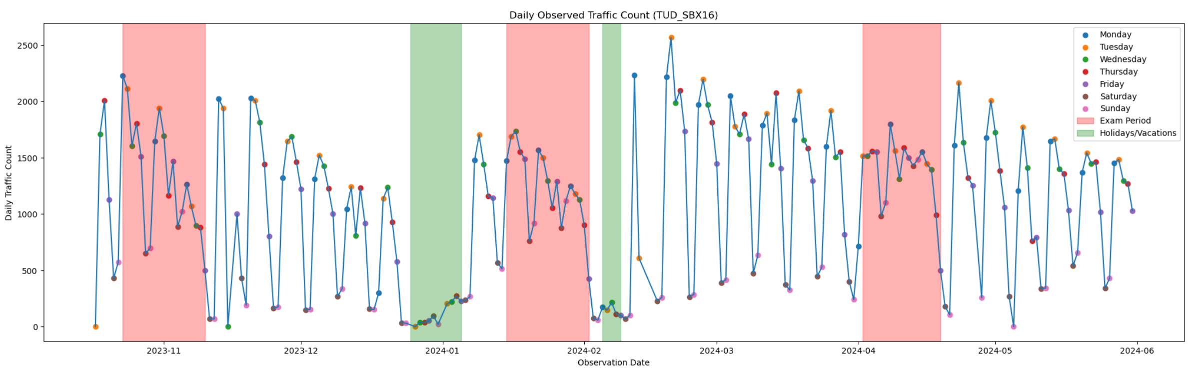


Figure A.7: Daily observed fluctuation in traffic count: TUD_SBX16

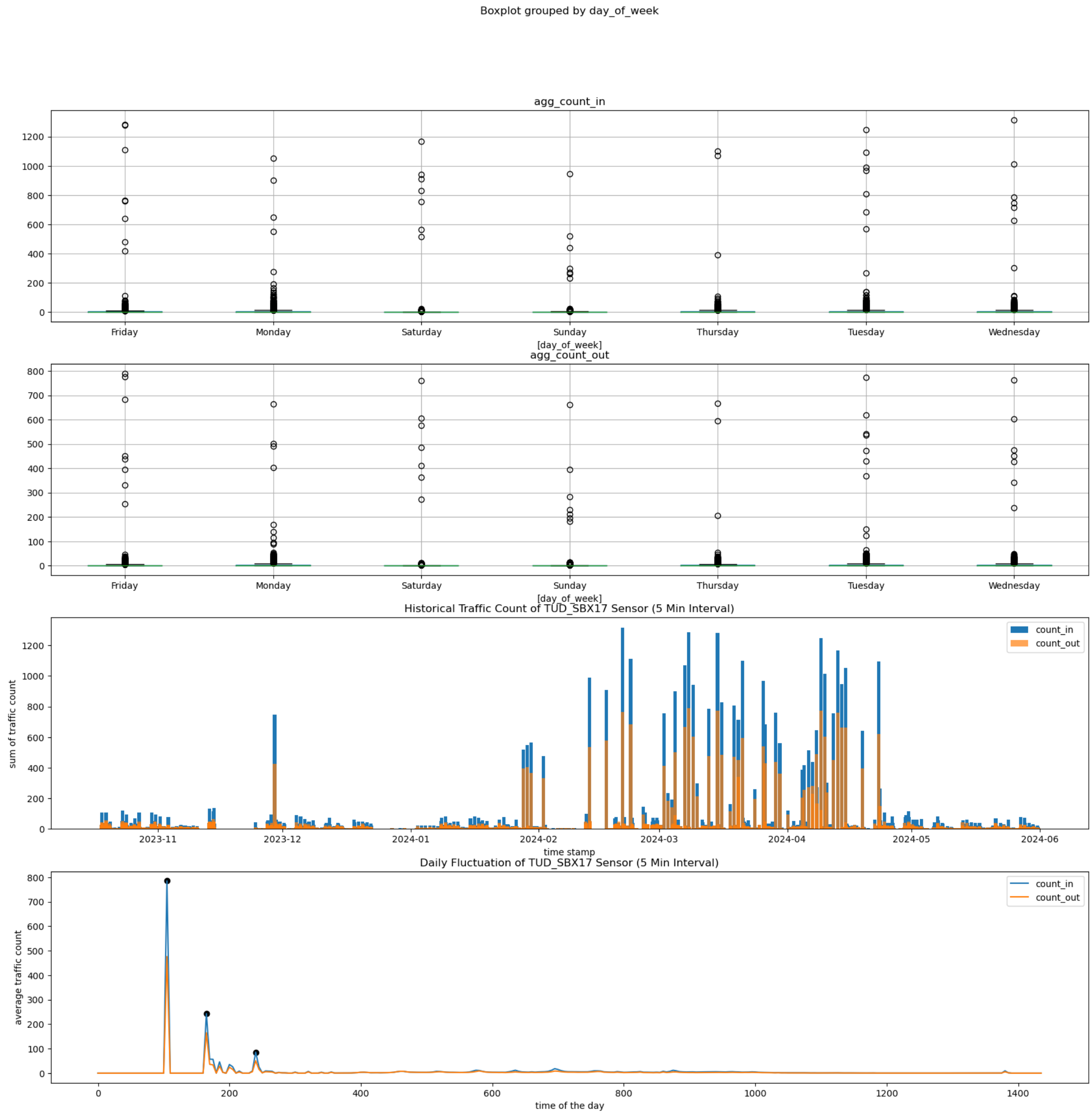


Figure A.8: Traffic count analysis: TUD_SBX17

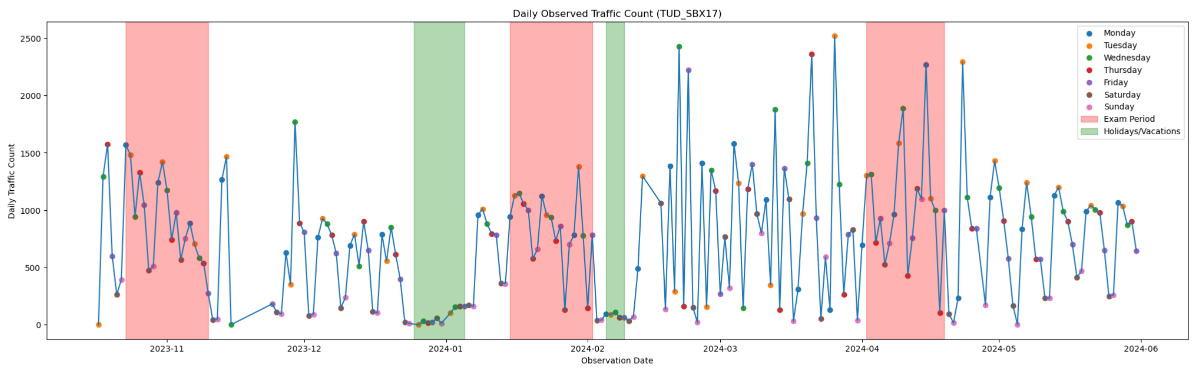


Figure A.9: Daily observed fluctuation in traffic count: TUD_SBX17

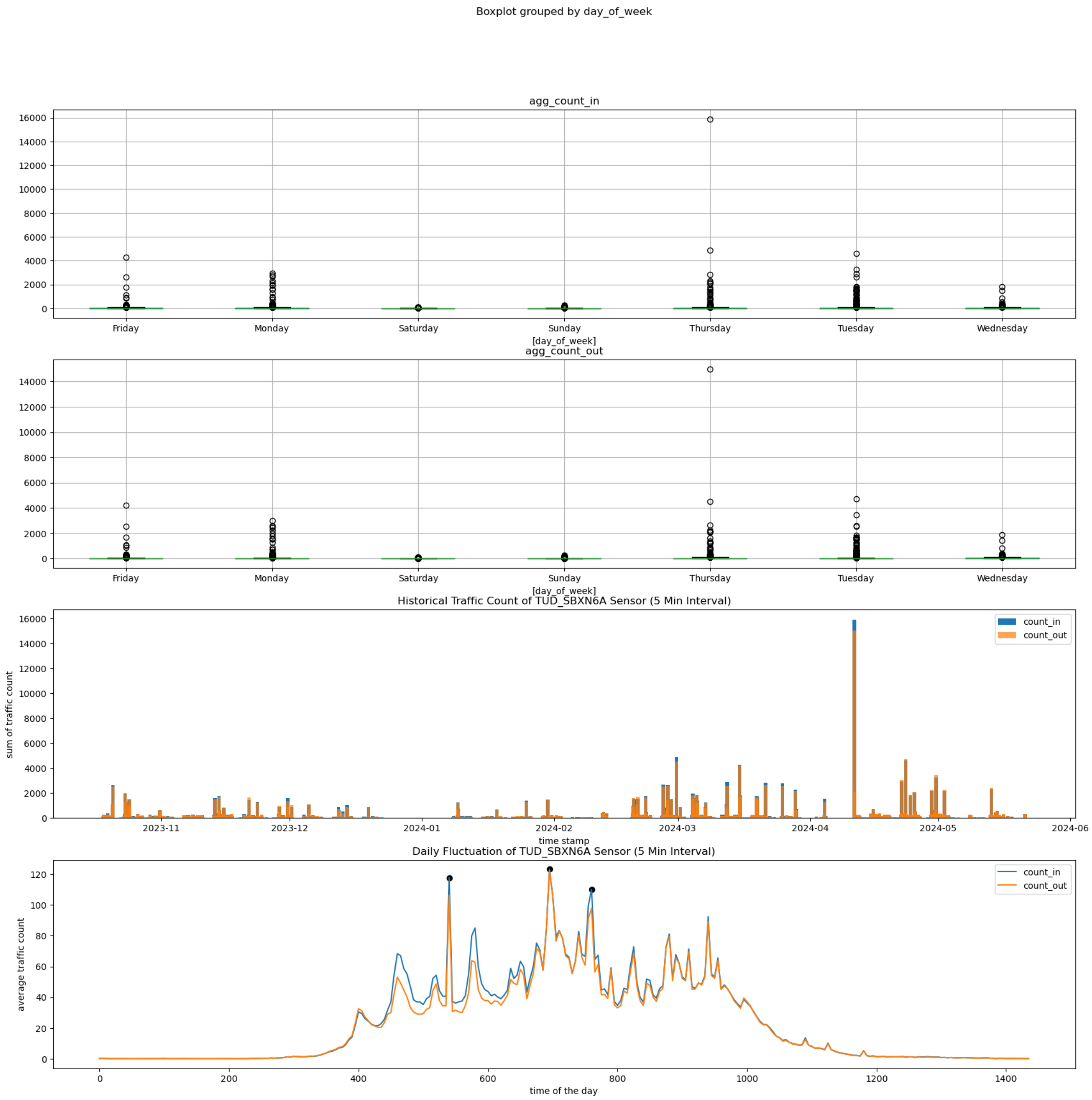


Figure A.10: Traffic count analysis: TUD_SBXN6A

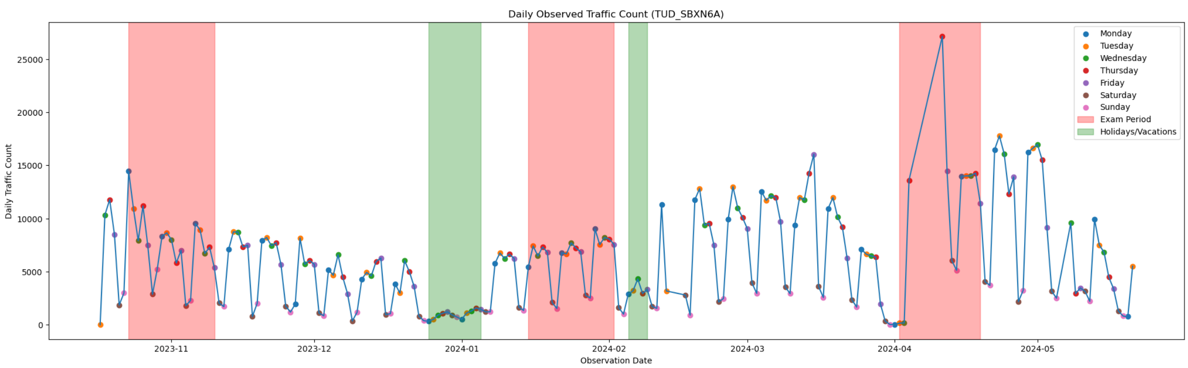


Figure A.11: Daily observed fluctuation in traffic count: TUD_SBXN6A

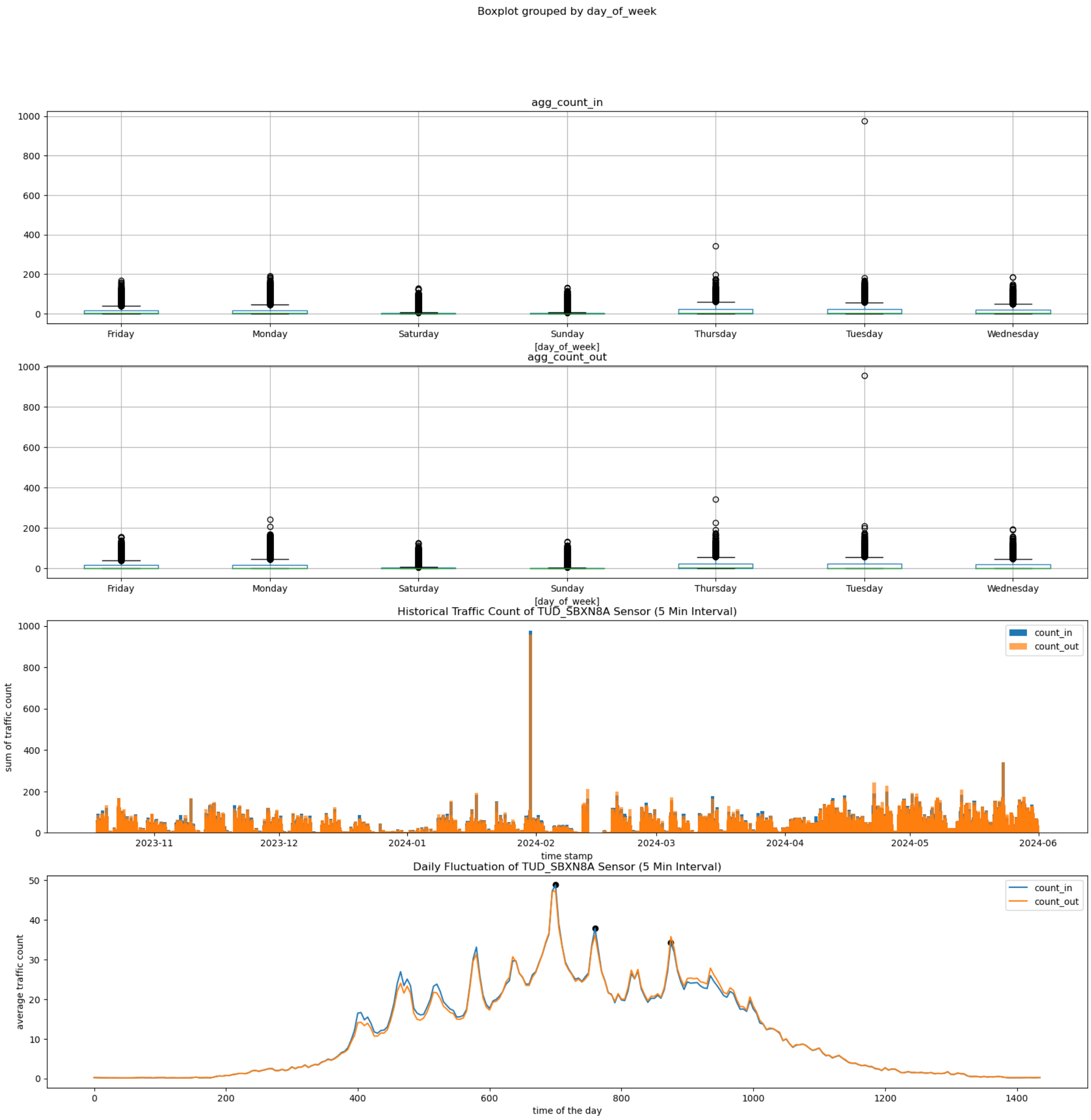


Figure A.12: Traffic count analysis: TUD_SBXN8A

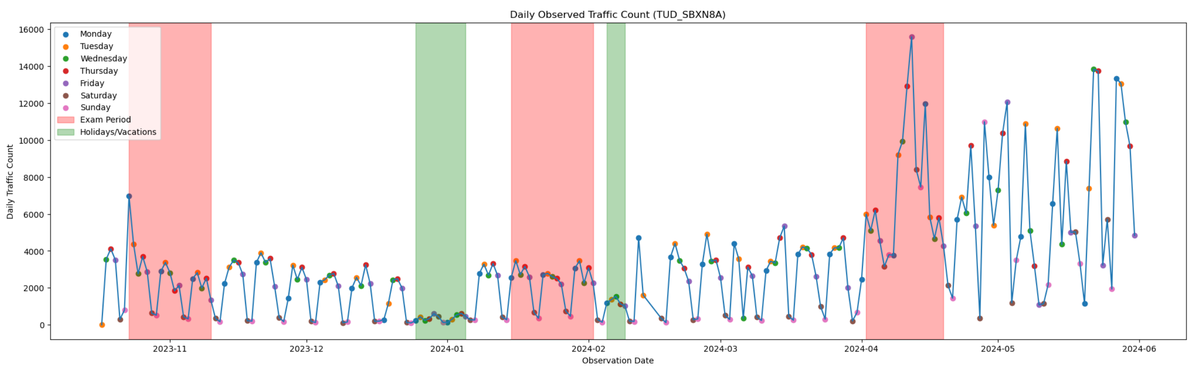


Figure A.13: Daily observed fluctuation in traffic count: TUD_SBXN8A

Figure A.14: Traffic count analysis: TUD_SBXN8B

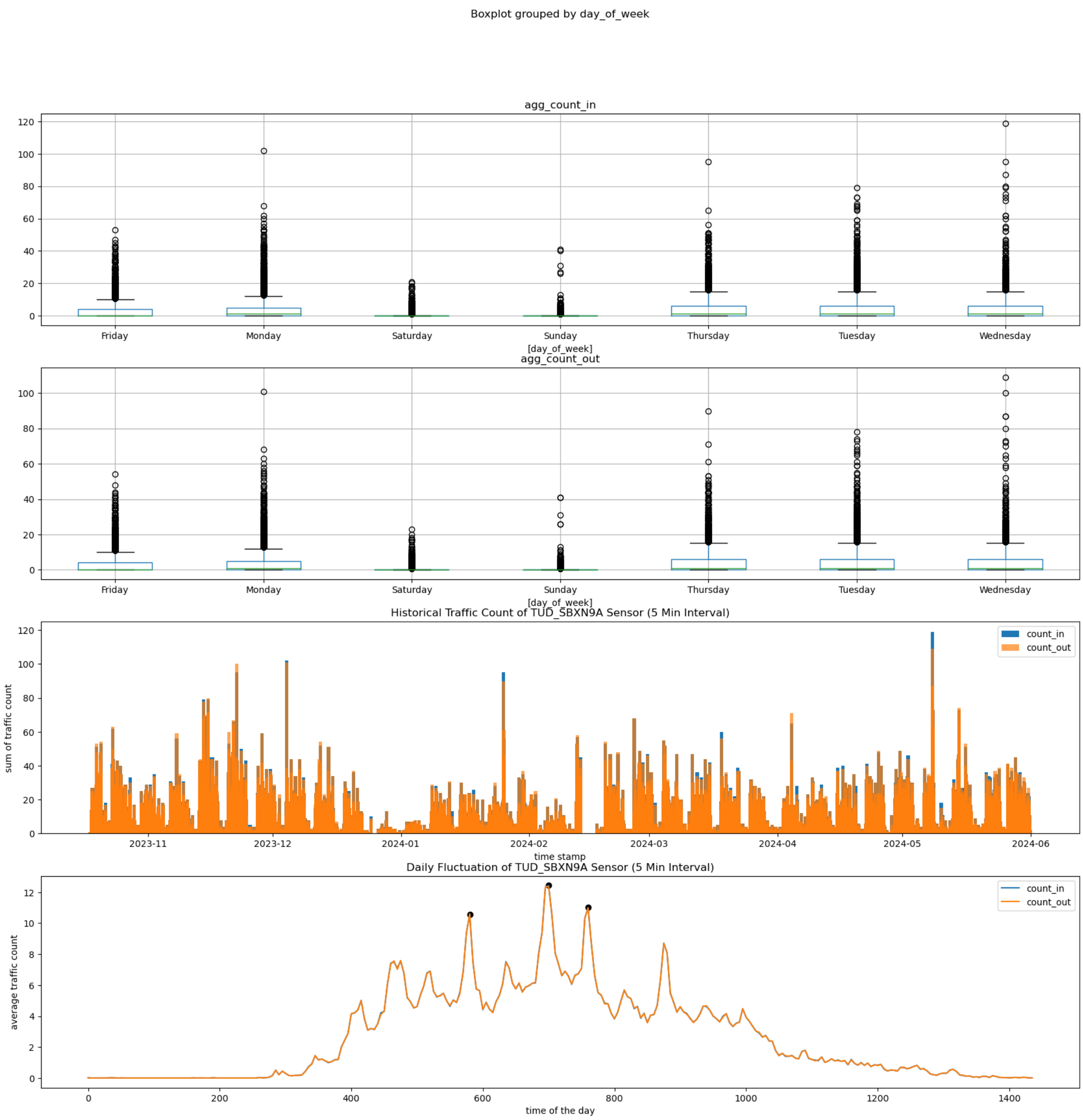


Figure A.15: Traffic count analysis: TUD_SBXN9A

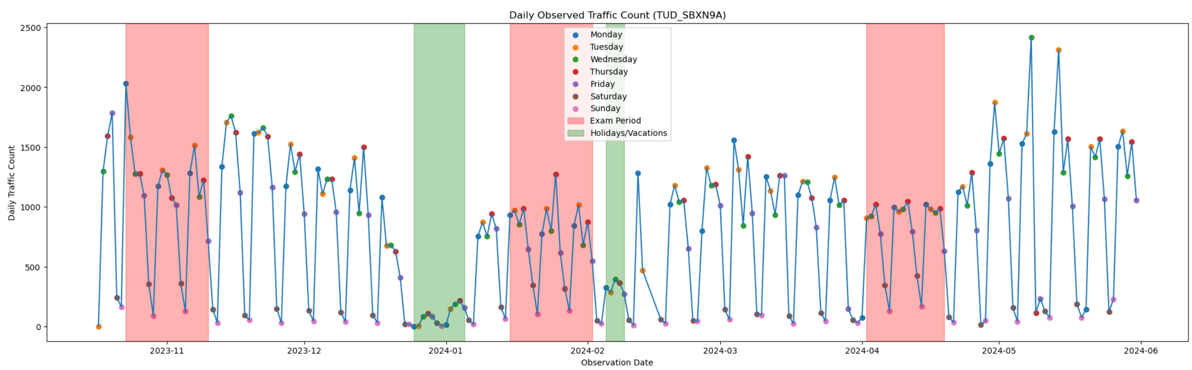


Figure A.16: Daily observed fluctuation in traffic count: TUD_SBXN9A

Figure A.17: Traffic count analysis: TUD_SBXN9B

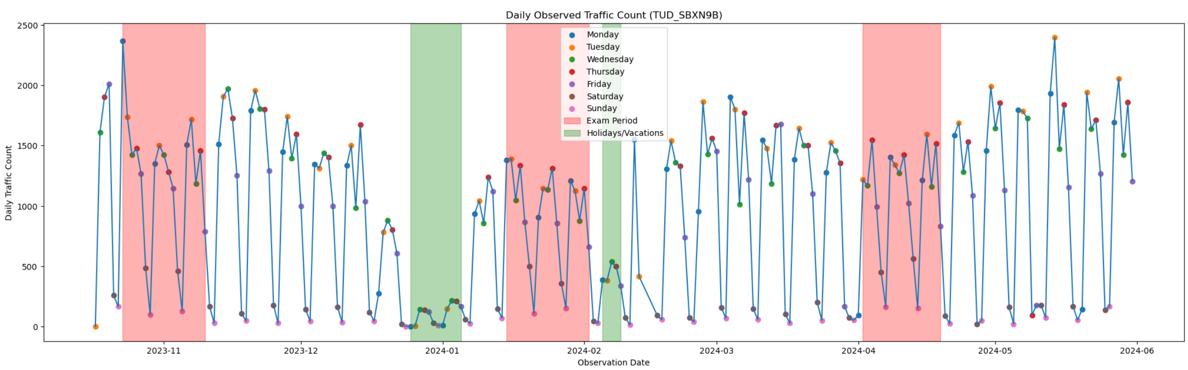


Figure A.18: Daily observed fluctuation in traffic count: TUD_SBXN9B

Figure A.19: Traffic count analysis: TUD_SBXN13

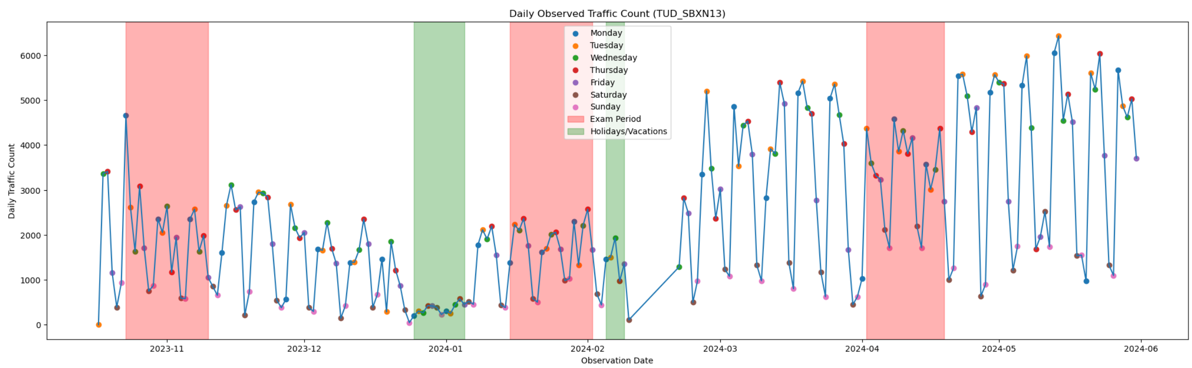


Figure A.20: Daily observed fluctuation in traffic count: TUD_SBXN13

Figure A.21: Traffic count analysis: TUD_SBXS12

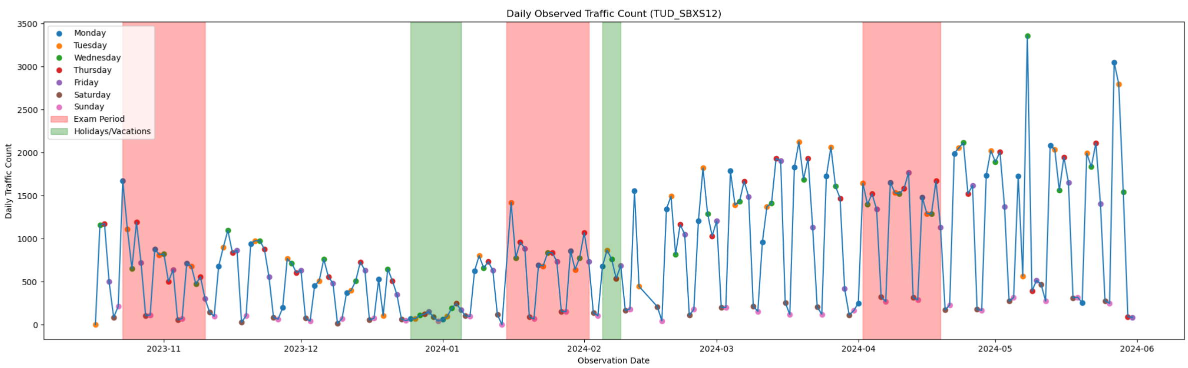


Figure A.22: Daily observed fluctuation in traffic count: TUD_SBXS12

Figure A.23: Traffic count analysis: TUD_SBXN13

Figure A.24: Traffic count analysis: TUD_SPX14

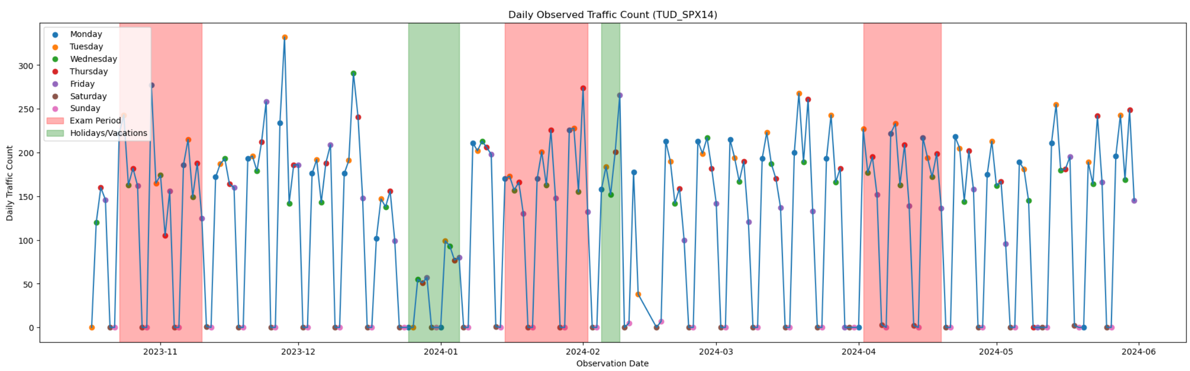


Figure A.25: Daily observed fluctuation in traffic count: TUD_SPX14

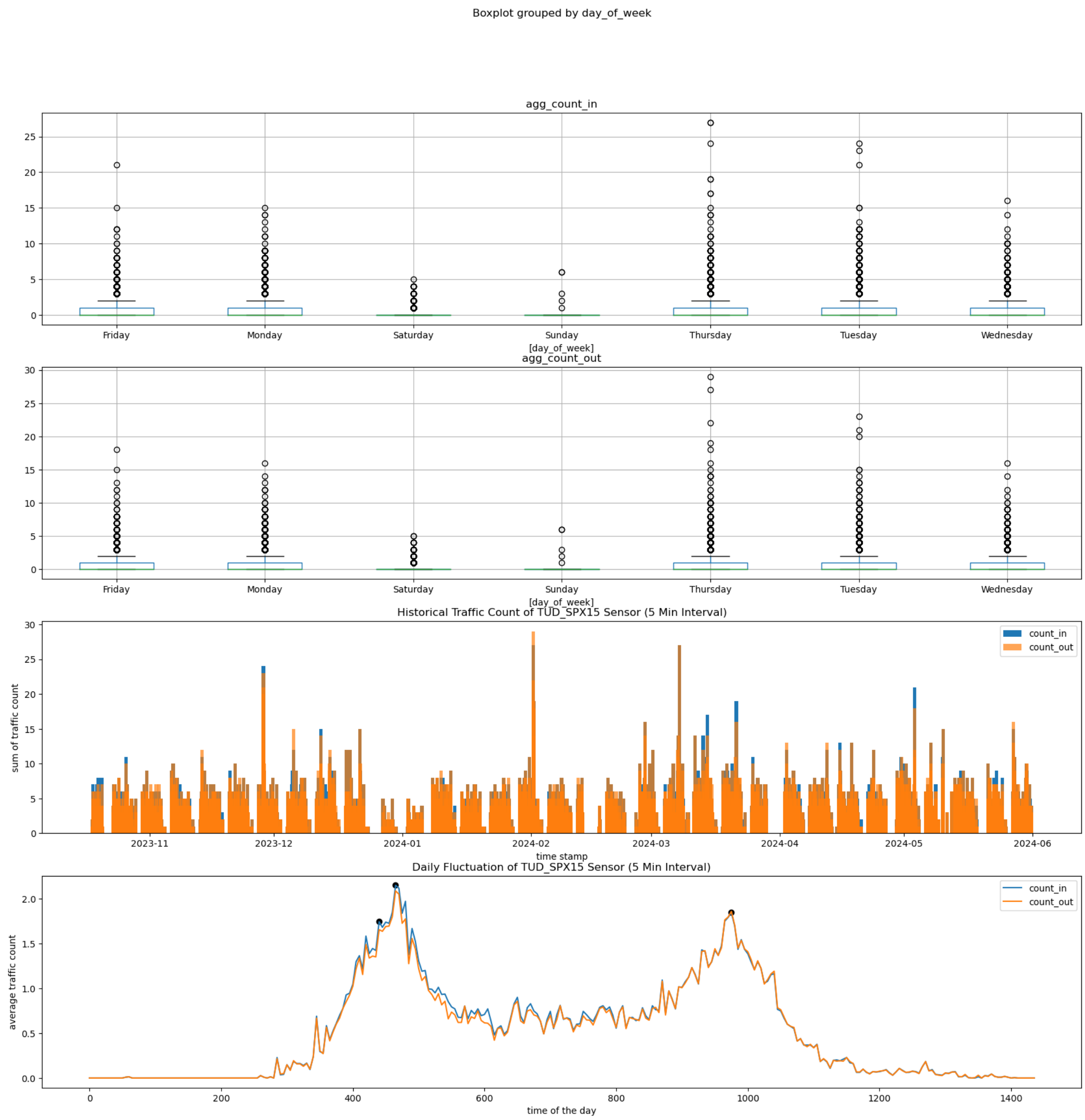


Figure A.26: Traffic count analysis: TUD_SPX15

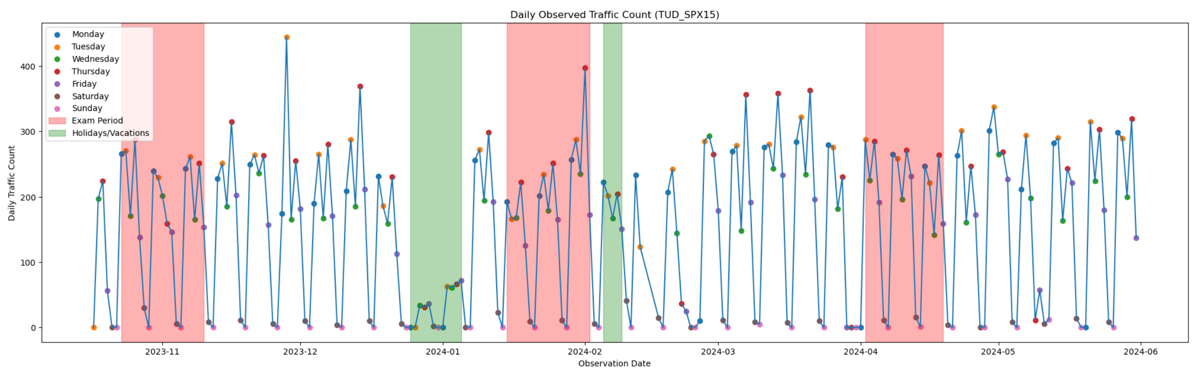


Figure A.27: Daily observed fluctuation in traffic count: TUD_SPX15

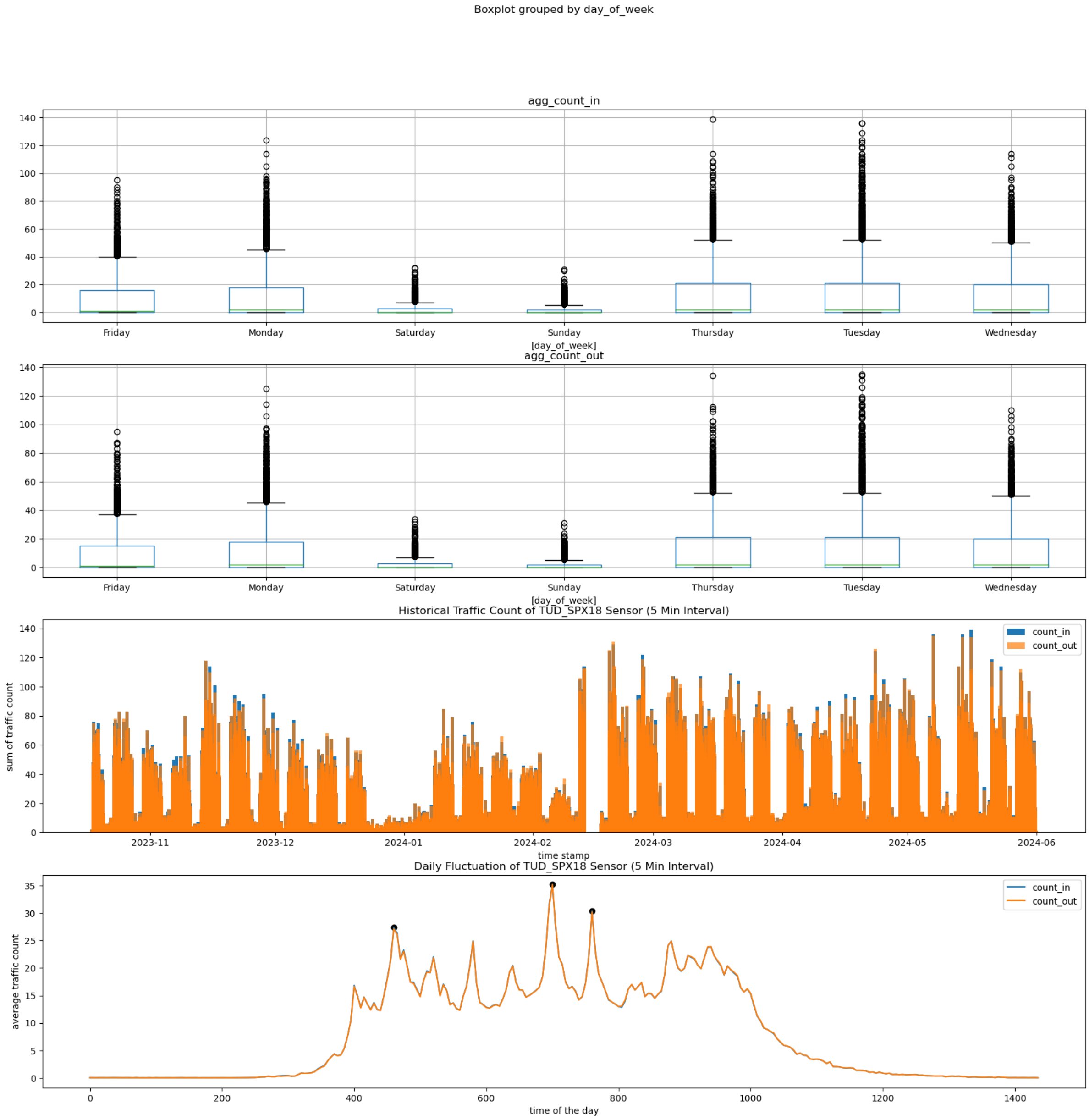


Figure A.28: Traffic count analysis: TUD_SPX18

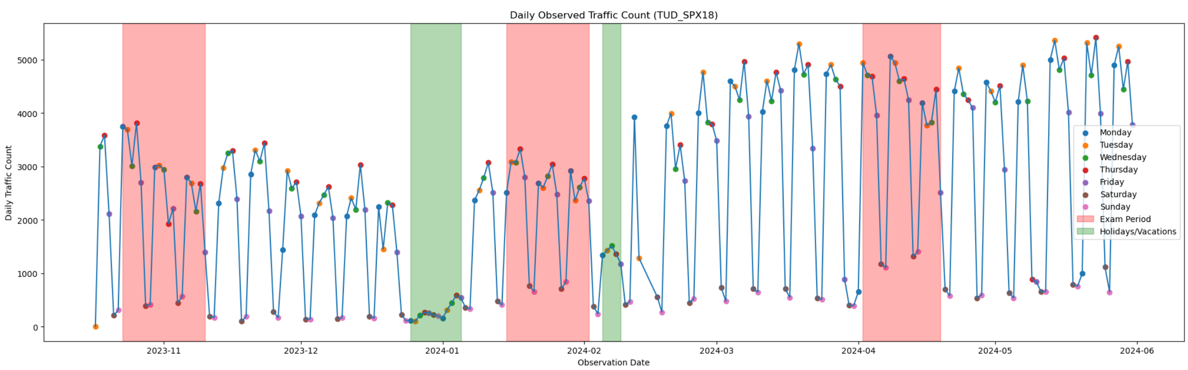


Figure A.29: Daily observed fluctuation in traffic count: TUD_SPX18

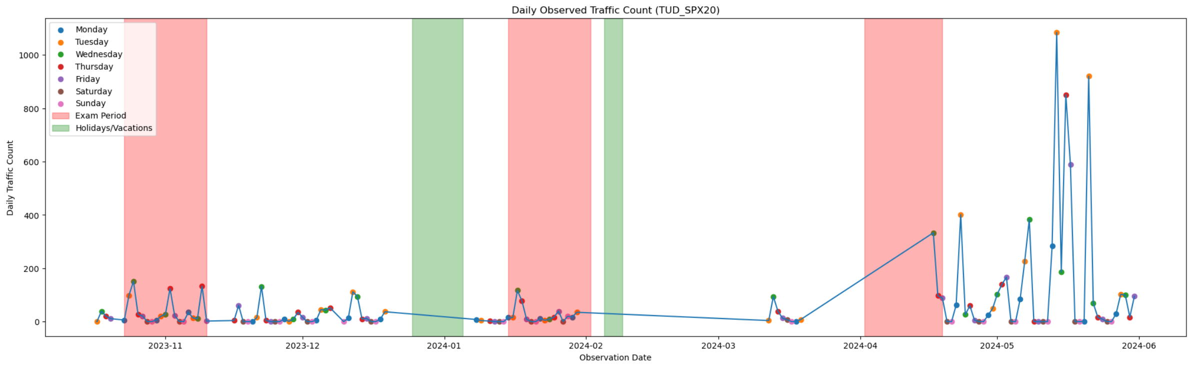


Figure A.30: Daily observed fluctuation in traffic count: TUD_SPX20

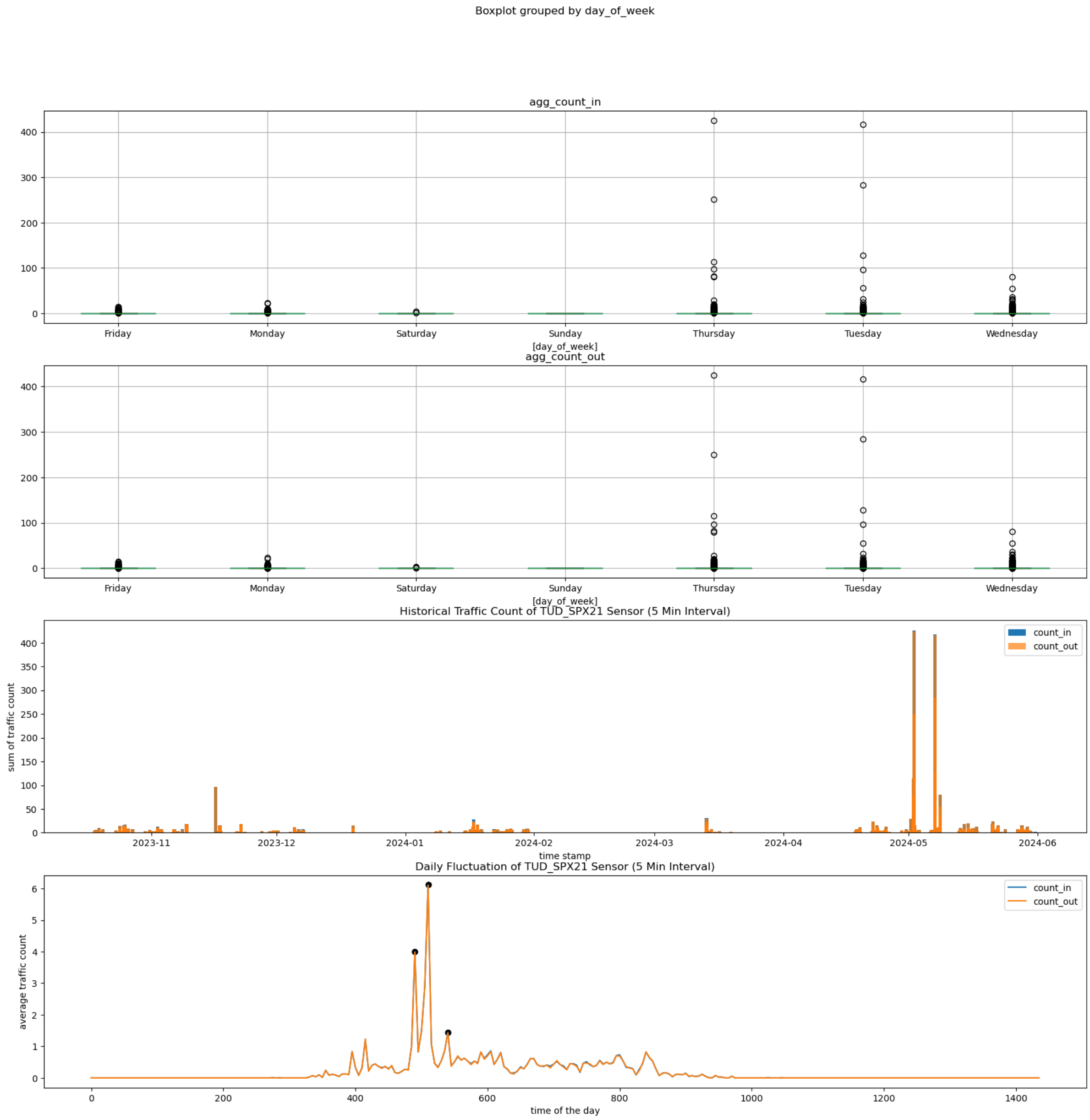


Figure A.31: Traffic count analysis: TUD_SPX21

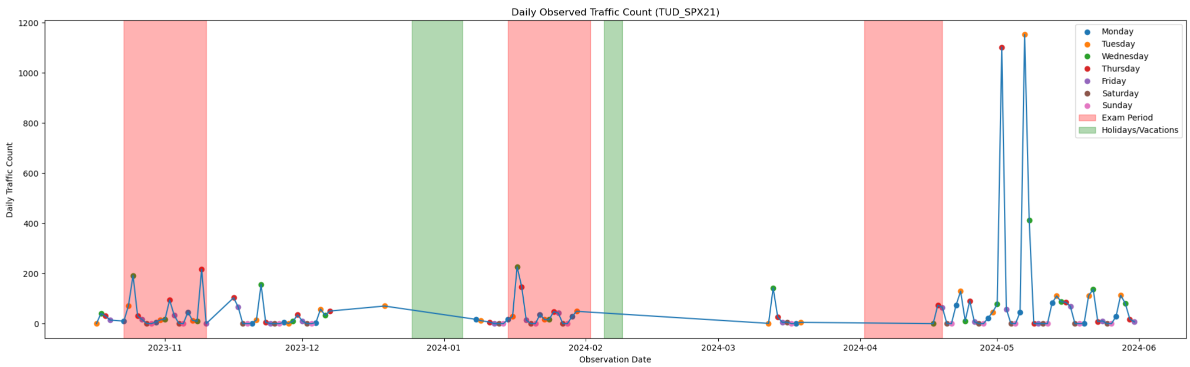


Figure A.32: Daily observed fluctuation in traffic count: TUD_SPX21