Journey through Crowds!

MODELING PASSENGER DISTRIBUTION ON RAILWAY PLATFORMS IN THE NETHERLANDS: A DISCRETE CHOICE APPROACH

THESIS

submitted in partial fulfillment of the

requirements for the degree of

MASTER OF SCIENCE

in

CIVIL ENGINEERING AND GEOSCIENCES

by

Muhammed Aldarawsheh

born in Damascus, Syria

Civil Engineering Master Thesis

Department of Transport & Planning

Faculty CITG, Delft University of Technology

Delft, the Netherlands

www.ewi.tudelft.nl

www.tudelft.nl/en/ceg

ProRail

Moreelsepark 3, 3511 EP Utrecht

Utrecht, the Netherlands

www.prorail.nl

Journey through Crowds!

MODELING PASSENGER DISTRIBUTION ON RAILWAY PLATFORMS IN THE NETHERLANDS: A DISCRETE CHOICE APPROACH

Author: Muhammed Aldarawsheh

Student id: 5209803

Email: M.Aldarawsheh@student.tudelft.nl

Thesis Committee:

Chair: Dr.ir. D.C. Duives, Faculty CITG, TU Delft

University supervisor: Dr. Y. Yuan, Faculty CITG, TU Delft

Company supervisor: Drs. L. Verhoeff, Beleidsadviseur, ProRail

Committee Member: Dr. C.N. van der Wal, Faculty TPM, TU Delft

Abstract

This study investigates the factors influencing passenger distribution on railway platforms in the Netherlands, using Eindhoven Central Station as a case study. The objective is to quantify how spatial characteristics, environmental conditions, and human behavioral factors jointly shape passenger distribution and to derive actionable recommendations for platform design and management.

A combination of descriptive statistics, spatial analysis, and discrete choice modeling was applied to a high-resolution dataset comprising 142,256 sensor-based observations collected under 49 situational scenarios. The platform was discretized into spatial cells characterized by distance to entrances, information boards, and the track edge, proximity to seating, leaning areas and kiosks, passenger density, lighting conditions, and weather.

Descriptive analyses reveal systematic clustering near entrances and comfort-related facilities, confirming the central role of accessibility and physical support in waiting behavior. Passengers generally avoid track-adjacent areas, while moderate social clustering occurs at intermediate densities. Environmental conditions further influence spatial patterns, with adverse weather and poor lighting reinforcing concentration in sheltered zones.

A multinomial logit model identifies nine statistically significant determinants of waiting location choice. Seating exhibits the strongest positive effect, followed by entrance proximity and leaning facilities, highlighting comfort and accessibility as primary drivers. Safety considerations reduce the attractiveness of areas near the track, although this effect weakens under favorable lighting and weather conditions. Areas near kiosks and information boards are avoided, indicating the disutility associated with congestion and circulation conflicts. The model demonstrates strong predictive performance and reproduces observed passenger distributions with high accuracy.

The findings show that platform waiting behavior reflects structured trade-offs between comfort, safety, accessibility, congestion avoidance, social context, and environmental conditions. Based on these results, design recommendations are proposed, including redistributing seating and leaning facilities, dispersing entrance flows, relocating kiosks and information boards toward circulation corridors, applying adaptive lighting strategies, and maintaining clear safety buffers near the track edge. The study provides a behavioral and empirical foundation for improving comfort, safety, and operational efficiency at Dutch railway stations.

Preface

This thesis presents the findings of my research on passenger distribution on railway platforms, using Eindhoven Central Station as a case study. The study investigates how spatial characteristics, environmental conditions, and human behavior influence where passengers choose to wait, with the aim of providing actionable and evidence-based recommendations for platform design and management.

In conducting this research, digital tools and artificial intelligence-based systems were used in a supportive manner, primarily to assist with language refinement, code debugging, and the organization of technical content. All conceptual development, methodological design, data analysis, modeling decisions, and interpretation of results remain entirely the author's own responsibility. The use of such tools followed the principles of academic integrity, transparency, and responsible research practice.

I would like to sincerely thank my supervisors for their continuous guidance, constructive feedback, and academic support throughout this research. I am also grateful to the ProRail advisers for their expert advice and assistance, which were essential for accessing data and contextualizing the empirical analysis. Special thanks go to my family and friends for their encouragement and patience, which have been a constant source of motivation throughout this research journey. This work reflects not only individual academic effort, but also the collective support and inspiration of everyone who contributed directly or indirectly to its completion.

Table of Contents

List of Figures

List of Tables

1. Introduction

Railway platforms are critical nodes within urban transit systems, acting as interfaces between passengers and trains where safety, efficiency, and comfort must be balanced. The effective management of passenger distribution on platforms plays a key role in minimizing dwell times, preventing overcrowding, and ensuring smooth train operations. Uneven crowding not only heightens safety risks, such as pushing incidents near platform edges, but also leads to train delays and network-wide inefficiencies (Daamen, Bovy, & Hoogendoorn, 2005, de Ana Rodríguez, Seriani, & Holloway, 2016)[1][2]. As urbanization intensifies and travel demand grows, optimizing passenger flows on platforms has become an increasingly urgent challenge for transport planners and operators.

Passenger distribution on platforms is shaped by a combination of infrastructural, environmental, and behavioral factors. Elements such as entrance and exit locations, seating distribution, visibility, weather conditions, and lighting quality all influence where passengers choose to stand and wait (Daamen & Hoogendoorn, 2003, Van Hagen, 2011)[3][4]. Moreover, platforms function as both transit corridors and waiting spaces, producing dynamic and often uneven spatial usage patterns, particularly during peak hours (Løvås, 1994)[5]. These imbalances can result in localized congestion zones near preferred boarding areas, while other regions of the platform remain underused. Such conditions complicate both passenger experience and train operation efficiency.

Technological and design interventions, such as platform edge doors, adaptive signage, and real-time crowd information, have shown potential to improve distribution uniformity and safety (Zhang, Han, & Li, 2008, Qi et al., 2014)[6][7]. Yet, these solutions often overlook the complex interplay between environmental context and human behavior. For instance, passengers tend to cluster under shelters during adverse weather or avoid poorly lit zones due to perceived insecurity (Páez & Scott, 2005, Fruin, 2012)[8][9]. Thus, a data-driven and behaviorally grounded understanding of passenger distribution is essential for designing more resilient and user-centered railway platforms.

1.1 Background and Research Problem

Railway stations play a central role in Dutch daily mobility, accommodating growing passenger volumes that challenge operational efficiency, safety, and comfort (ProRail, 2022, KiM, 2021) [16][17]. Platforms, as the primary interface between passengers and trains, are dynamic spaces where infrastructure, environmental conditions, and human behavior intersect. Understanding how passengers select waiting locations is crucial, as spatial clustering near train doors, sheltered areas, or well-lit sections can create local congestion, underutilized spaces, and safety hazards, particularly during peak hours (Daamen, Bovy, & Hoogendoorn, 2005, Páez & Scott, 2005, Van Hagen, 2011) [1][4][8].

Previous research has provided valuable insights into pedestrian flow and boarding behavior using macroscopic flow models and microscopic agent-based approaches (Fruin, 1971, Helbing et al., 2000, Løvås, 1994)[10][12][5], and more recent studies employ data-driven and discrete choice models to capture individual behavior (Ben-Akiva & Lerman, 1985, Xu et al., 2020)[13][14]. However, passenger distribution along platforms is often treated as a secondary outcome rather than as a behavior influenced by infrastructure, environmental factors, and perceived utility. Factors such as distance to exits or seats, lighting, visibility, and weather conditions shape waiting location choices, with passengers tending to occupy sheltered or well-lit areas during adverse conditions (Seriani & Fujiyama, 2019, Fruin, 2012) [15][9]. Despite advances in real-time monitoring and crowd management technologies, predictive modeling of static spatial distribution remains limited, leaving a knowledge gap in empirically grounded, behaviorally realistic representations of platform occupancy.

This study aims to fill this gap by developing a model to estimate static passenger distribution on railway platforms in the Netherlands, integrating spatial, environmental, and behavioral factors. Using data from operational monitoring systems and empirical modeling techniques, this research quantifies the relative importance of variables affecting waiting location choice, allowing for a better understanding of how passengers distribute themselves along platforms.

By providing an empirically validated framework for platform occupancy, the study contributes actionable insights for improving safety, efficiency, and passenger comfort, supporting rail operators such as ProRail and NS in managing growing demand and optimizing platform design and management strategies.

1.2 Research Objective

The primary objective of this research is to develop and validate a behavioral model that explains and predicts passenger distribution patterns on railway platforms based on a combination of spatial, environmental, and behavioral factors. Through a data-driven and empirically grounded approach, the study aims to quantify how elements such as distance to entrances and exits, proximity to seating, visibility, weather conditions, and lighting collectively shape passengers' standing preferences and spatial dispersion.

To achieve this, the research employs a discrete choice modeling framework, allowing for the estimation of individual-level decision behavior under varying contextual conditions (Ben-Akiva & Lerman, 1985) [13]. This approach provides a probabilistic representation of passenger choices, integrating both observable variables (e.g., physical distances, lighting levels) and unobserved preferences. The model will be trained and tested using empirical data collected from real platform environments, ensuring statistical robustness and behavioral realism (Daamen, Bovy, & Hoogendoorn, 2005, Xu et al., 2020) [1][14].

1.3 Societal Relevance

Ultimately, the objective extends beyond theoretical modeling. By providing an interpretable and empirically validated framework, this research seeks to support railway planners, designers, and operators, such as ProRail and Nederlandse Spoorwegen (NS), in developing evidence-based strategies to enhance safety, efficiency, and passenger comfort on platforms. This aligns with broader policy goals emphasizing sustainable and user-centered mobility solutions in the Netherlands (KiM, 2021, ProRail, 2022) [17][16].

1.4 Research Scope

This research focuses on the modeling and validation of passenger distribution behavior on railway platforms within the context of Dutch railway stations, using empirical data and behavioral modeling techniques. The study's scope is defined by both its spatial boundaries and analytical framework, ensuring methodological focus while maintaining relevance to real-world operational settings.

The empirical scope of the study is limited to train station platforms characterized by open-air, multi-access configurations, such as those managed by ProRail and Nederlandse Spoorwegen (NS). These environments typically exhibit heterogeneous spatial and environmental conditions, including variations in platform geometry, seating distribution, weather exposure, and lighting quality (ProRail, 2022)[16]. By focusing on such settings, the study aims to capture the behavioral diversity that arises from interactions between infrastructure and passenger preferences (Daamen & Hoogendoorn, 2003, Van Hagen, 2011)[3][4].

Analytically, the research employs a discrete choice modeling approach to represent passenger spatial decisions probabilistically. The model uses individual-level or aggregated spatial data to estimate the likelihood of passengers choosing specific platform zones based on explanatory variables such as distance to access points, proximity to seating, visibility, weather, and lighting (Ben-Akiva & Lerman, 1985, Xu et al., 2020)[13][14]. These factors were selected for their behavioral relevance and measurability in real environments, allowing for replicable and interpretable analysis.

The validation of the developed model constitutes a core part of the study's scope. Both training and testing datasets are used to assess generalization and predictive robustness through performance metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and rank correlation measures (Train, 2009, Ortúzar & Willumsen, 2011)[18][19]. Additionally, spatial validation indicators, including mean spatial deviation and Top-N Hit Rate, are applied to evaluate how accurately the model reproduces observed spatial distributions (Horni et al., 2016)[20].

However, certain limitations are acknowledged within this scope. The analysis does not explicitly account for temporal dynamics such as crowd evolution over time or train dwell interactions, which could be addressed in future extensions using dynamic or agent-based models (Helbing et al., 2000)[12]. Similarly, the model abstracts from individual socio-demographic attributes due to data availability constraints, focusing instead on spatial-environmental determinants of behavior.

1.5 Research Questions

Based on the background, problem statement, objectives, and scope, this study is guided by a central research question supported by a set of sub-questions. Together, they structure the investigation and ensure a comprehensive understanding of passenger distribution on railway platforms in the Netherlands, linking theoretical insights, empirical data, and practical applications.

Central Research Question:

How do passengers distribute themselves over railway platforms in the Netherlands in response to infrastructure, environmental and crowding conditions?

To address this broad question, the study is structured around the following **sub-questions**, which are logically connected to ensure a systematic investigation:

1. *What elements significantly influence passenger choices on train platforms, and how do these factors interact to shape spatial distribution patterns?*
2. *What theoretical frameworks contribute to understanding passenger distribution on train station platforms, and which methodologies can effectively model the waiting location distributions?*
3. *What types of data are available regarding passenger waiting behavior on train station platforms, and which methods will be used to collect, process, and analyze these data to support robust modeling?*
4. *How do environmental, contextual, and crowding factors influence passenger location choices and contribute to uneven platform distribution?*
5. *How can insights from the modeling outcomes inform policy to improve platform safety, efficiency, and user experience?*

These sub-questions break the central research question into actionable components, guiding the study in a structured way. The first identifies key spatial, environmental, and behavioral factors that shape passenger choices, providing the foundation for understanding spatial patterns. The second examines relevant theoretical frameworks and modeling approaches to ensure the analysis is grounded in appropriate methodologies. The third focuses on the availability and processing of empirical data on waiting behavior, enabling robust, evidence-based models. The fourth explores how contextual factors, such as weather, lighting, and crowding, influence location choices and uneven platform utilization. Finally, the fifth translates modeling outcomes into practical insights for policy, linking empirical and theoretical findings to improvements in platform safety, efficiency, and user experience. Together, these sub-questions allow the study to systematically explain how passengers distribute themselves across platforms, providing a comprehensive answer to the main research question.

1.6 Thesis Outline

This thesis is structured to systematically investigate passenger distribution on train platforms. Chapter 1 introduces the research background, problem, objectives, scope, and research questions. Chapter 2 reviews the literature, which focuses on key factors affecting passenger behavior, identifies the research gaps, and presents the conceptual framework and key models. Chapter 3 describes the research methodology. In particular, this chapter presents the data requirements, data collection strategy, data sampling strategy, attribute definition, choice set development, descriptive analysis, statistical testing procedure, discrete choice modeling approach, and validation procedures. Chapter 4 presents the case study at Eindhoven Train Station, detailing data collection, choice sets, descriptive statistics, model specification, parameter estimation, and validation results. Chapter 5 discusses findings in relation to the research objectives, highlighting behavioral patterns, model performance, and practical insights for platform design and management. Chapter 6 provides the conclusions, limitations, and recommendations. This structure ensures a coherent flow from theoretical foundations and empirical analysis to practical applications, addressing the central research question and sub-questions.

2. Literature Review

This comprehensive literature review explores the factors that influence waiting location distribution and models that simulate passenger distribution on platforms, using insights from multiple disciplines to unravel complex spatial dynamics. Passenger distribution is usually affected by factors that lead to uneven distribution.

2.1 Key Factors

The literature on waiting behavior at railway platforms identifies six key factors that consistently influence where passengers choose to wait: (1) station location, position, and overall design (Baron, 2021) [21], (Fruin, 1987) [23], (van der Werf, 2019) [22], (Zhang, 2021) [26], (2) wayfinding and safety systems (Ahn et al., 2017) [25], (Xu et al., 2020) [27], (Bandara & Hewawasam, 2020) [30], (Romero et al., 2023) [33], (3) train operations and train-car design (Li & Zhu, 2016) [39], (Xu et al., 2022) [38], (Oliveira et al., 2019) [42], (Szplett & Wirasinghe, 1984) [43][44], (4) platform layout and physical configuration (Seriani et al., 2016) [46], (Bosina et al., 2015) [47], (5) pedestrian movement behavior (Bosina et al., 2015) [47], (Zhang et al., 2020) [48], and (6) human factors and environmental conditions (Ferri & Popp, 2022) [49], (Van Hagen, 2011) [4], (Kuipers et al., 2021) [52], (Seriani & Fujiyama, 2019) [15]. Each of these factor's shapes waiting location distribution through different mechanisms, ranging from physical infrastructure constraints to operational conditions, behavioral tendencies, and external environmental influences.

The following six subsections (Sections 2.1.1-2.1.6) discuss each factor in turn, summarizing existing knowledge, methodological approaches, and empirical findings from previous studies. By examining these factors one-by-one, the review clarifies how platform design, passenger behavior, operational elements, and environmental conditions jointly contribute to uneven or patterned distribution along platforms. Section 2.1.7 synthesizes the findings from the preceding subsections to present the overarching factors that shape passenger distribution on platforms, bringing together insights on infrastructure, information systems, operational conditions, and human behavior. This integrative conclusion highlights key themes and remaining gaps in the literature, providing a foundation for the conceptual and methodological framework developed in the following chapters.

2.1.1 Station Location, Position, and Design

The location and position of a station within the rail network and the city, as well as the platform's capacity to handle large traffic flows during peak hours, significantly impact user experience and safety at the station (Baron, N., 2021) [21]. Platform design also contributes to overall passenger experience. Scholars such as van der Werf (2019) [22] emphasize the importance of user-centered design principles in creating platforms that meet diverse passenger needs.

Major factors, such as the location and position of a station, inherently influence other aspects like platform safety. Fruin (1987) [23] shows that the physical placement of a station affects pedestrian flow and safety protocols. The platform layout, designed with passengers' preferences and ease of guidance in mind, significantly impacts wayfinding systems (Passini, 1992) [24]. Ahn et al. (2017) [25] further highlights how well-designed platforms enhance safety and overall passenger experience by improving wayfinding efficiency, reducing confusion, and ensuring that high-priority areas are effectively managed during peak hours. These interconnected factors illustrate the complexity of station design and its influence on passenger behavior and safety.

Zhang (2021) [26] used an agent-based model to simulate passenger distribution at key interchange stations, revealing higher congestion at strategically important stations. Their model, based on detailed passenger movement data and network connectivity, provided insights into congestion patterns and the impact of station location.

Agent-based models in these studies rely on field observations, video analytics, and network connectivity metrics. They are effective in simulating passenger behavior under various station layouts but may not fully capture individual preferences or the impact of environmental variables.

2.1.2 Wayfinding and Safety systems in station areas

Effective wayfinding systems have proven to be an important factor in improving passenger navigation, reducing congestion by guiding passengers, and influencing uniform distribution on the platform (Ahn et al., 2017) [25]. The visibility and accessibility of emergency features such as exits and evacuation routes influence the management of passenger distribution, especially in emergency situations (Xu, H et al., 2020) [27]. Ergonomic considerations, clear signage, and seating arrangements contribute to passenger satisfaction. (Ibrahim, A.N. H. et al., 2020) [28] and (Rüger, B. 2018) [29]. Lighting also plays an important role from a safety point of view on the platform. A study by Bandara & Hewawasam, (2020) [30] identified convenience, safety, and comfort as important factors influencing the design of pedestrian facilities at railway stations. Several studies have investigated various factors related to platform security, user experience, and operational efficiency. A study by T. Mizuno and K. Tokuda (2023) [31] highlights the important role of physical infrastructure such as platform fences and Braille blocks in preventing accidents and ensuring passenger safety. Research shows that passenger numbers on trains are likely to increase by 10.5% due to passengers feeling safe on trains and at stations (Crime Concern 2002) [32].

However, the literature emphasizes the integration of cutting-edge technologies to improve the capabilities of the platform such as Romero et al., (2023) [33] discussed the integration of digital technologies holds promise for improving wayfinding and the provision of real-time information. Another study by Virgona, A et al. (2015) [34] describes the use of sensor-based systems to monitor passenger flow and optimize platform occupancy. Additionally, various techniques such as video surveillance are used to measure passenger distribution on the platform. Real-time video analysis allows you to collect data on passenger movement and distribution patterns. Zheng, Z et al. (2023) [35] used video surveillance data to analyze passenger distribution and identify areas at risk of congestion. Moreover, sensors and probes strategically placed on the platform detect passenger movement and provide valuable insight into distribution patterns. Hänseler, F et al. (2020) [36] used infrared sensors to detect passenger density and distribution.

Existing research underscores the critical role of effective wayfinding systems, emergency feature visibility, ergonomic considerations, and clear signage in enhancing passenger navigation, reducing congestion, and ensuring satisfaction. Overall, while current research has made significant strides in understanding and addressing these key factors influencing platform safety, operational efficiency, and passenger distribution, the integration of innovative technologies continues to be an area of focus for future advancements in railway station design and management.

Xu et al. (2020) [27] employed a field experiment and video analysis approach to assess the impact of emergency exits, seating arrangements, and signage on passenger behavior. This study used real-time video data to track passenger movement and identified key areas where wayfinding systems significantly influenced distribution patterns, especially in emergency situations.

Bandara & Hewawasam (2020) [30] utilized sensor-based data collection methods combined with statistical modeling to evaluate the impact of lighting and signage on pedestrian movement in stations. Their study demonstrated that improved lighting and clear signage reduces pedestrian congestion and enhances safety. The use of real-time sensor data allowed for a dynamic analysis of how these ergonomic factors influence passenger behavior.

The use of field experiments, video analysis, and sensor-based data collection provides robust insights into how wayfinding systems and safety features influence passenger distribution. However, these models often lack the ability to predict individual decision-making processes in varying environmental conditions.

2.1.3 Train Operations and Design of Train Cars

A study by Zhang et al. (2020) [37] emphasized that the platform is of great importance in ensuring the safety and efficiency of train operations within the station, especially during unexpected train delays. This highlights the importance of efficient platform management to reduce security risks during operational disruptions. Additionally, a study by Xu et al. (2022) [38] addressed the optimization of train operations at high-speed stations and emphasized the need for accurate platform allocation to ensure operational efficiency and safety. This highlights the importance of efficient platforms in maintaining safe and smooth train operations. Train operations, including train frequency, schedule, and delays, have a significant impact on passenger distribution, as rush hours lead to increased concentration on platforms (Li, W., & Zhu, W. (2016) [39], Hartleb, J., & Schmidt, M. (2022) [40]). Li, W., & Zhu, W. (2016) [39] found that, in contrast to normal train operations, disruptions such as train delays lead to changes in both passenger route preferences and the distribution of passenger flows on timetable-based networks. Timetables can influence passengers' decisions about when and where to wait on the platform (Ingvardson et al. 2018) [41]. However, train design, including door placement and class configuration, plays a role in determining passenger movement, selection, and boarding locations, thereby influencing the spatial dynamics on the platform (Oliveira et al., 2019) [42]. Analyzing this problem may require deeper knowledge of passenger behavior on the platform in terms of the distribution of passengers across the train doors, along with station design factors. Only a very limited number of researchers have been working on this problem. One of the rare studies that addressed this issue was by Szplett and Wirasinghe (1984 i, 1984 ii) [43][44], who investigated train dwell time and passenger boarding and alighting times at an LRT station in Calgary, Canada. Their study included a comprehensive analysis of passenger distribution between train doors depending on platform design. The analysis revealed that the location of platform entrances and exits has a significant impact on passenger distribution. At stations with only one entrance/exit at the end of the platform, passengers concentrate near the entrance/exit. It turns out that this passenger distribution follows a negative exponential distribution. Stations with multiple platform entrances have a more even distribution of passengers and are modeled using a normal distribution. Another study in this area was conducted by Krstanoski, N. (2014) [45]. In this study, they proposed a potential energy model to describe the passenger distribution between train doors before the train arrives at the station, based on data collected at a metro station in Toronto, Canada. The data used to build the model was collected by manual counting and video recording. The existing research, as reflected in the cited studies, underscores the critical role of efficient platform management in ensuring the safety and operational efficiency of train operations, especially during unexpected delays. While these studies shed light on platform dynamics, the reliance on manual counting and video recording for data collection suggests opportunities for more advanced data collection methods. The current literature provides valuable insights into the complex relationship between train operations, platform design

"2.1.4 Platform Layout and Design", and passenger distribution, highlighting the need for further research with advanced methodologies to deepen our understanding of these intricate dynamics.

Li & Zhu (2016) [39] used a Discrete Event Simulation (DES) to model how delays and disruptions impact passenger movement and waiting patterns. Their simulation incorporated real-world operational data, such as train schedules and passenger arrival rates, to analyze how disruptions affect platform congestion and waiting times.

Krstanoski, N. (2014) [45] applied a potential energy model to study how passengers distribute themselves near train doors. The model, based on manual passenger counts and boarding time observations, provided insights into the natural clustering of passengers around doors and highlighted the need for better design of train car entrances.

Discrete event simulation and potential energy models offer valuable insights into the operational dynamics of passenger distribution but are limited by their dependence on static data collection methods. These models often overlook the influence of environmental variables and individual preferences, leading to gaps in understanding real-time passenger behavior.

2.1.4 Platform Layout and Design

While Section 2.1.1 addressed station-level factors such as location, position, and overall design, this section focuses specifically on platform layout and design and their influence on passenger distribution.

The physical layout and design of the platform strongly affect how passengers position themselves. Seriani et al. (2016) [46] show that platform length and width influence distribution patterns: long platforms promote more even passenger distribution, whereas narrow platforms lead to higher concentration. Bosina et al. (2015) [47] highlight that obstacles and the size of queuing zones near platform entrances are important for pedestrian distribution. Passengers generally prefer waiting areas where they are undisturbed, allowing for more efficient system use. Only when traffic density is high and train arrival is imminent do pedestrians begin to occupy walking areas as temporary waiting space. Manual mapping of pedestrian locations has been used effectively to collect such data. Research underscores the pivotal role of platform layout in shaping passenger distribution. Seriani et al. (2016) [46] used cellular automata models to simulate pedestrian flow on platforms of varying lengths and widths, validated with video-tracking data, confirming that longer platforms encourage even distribution while narrow platforms cause crowding. Bosina et al. (2015) [47] applied space syntax analysis to study how obstacles like columns and queuing zones affect passenger movement, showing that poorly placed obstacles increase congestion and reduce efficiency. Cellular automata and space syntax models effectively simulate platform layout impacts on pedestrian flow but may not fully capture dynamic factors such as crowd density or individual preferences.

2.1.5 Pedestrian Movement Behavior

Passenger density, flow rate, and walking speed significantly affect passenger distribution on platforms. Bosina et al. (2015) [47] show that higher density leads to uneven distribution and congestion near entrances and exits. Similarly, slow walking speeds cause accumulation in certain platform areas (Zhang et al., 2020) [48].

Existing studies emphasize the importance of these factors in shaping platform distribution and provide valuable insights into passenger behavior and platform dynamics. However, future research could further explore mitigation strategies for managing density and flow to improve crowd management, safety, and efficiency on railway platforms.

Zhang et al. (2020) [48] employed macroscopic flow models using video footage and automated counting data to analyze the relationship between density and flow rates. Their results show that high-density conditions lead to uneven distribution and reduced walking speeds, contributing to congestion. While these models provide a broad understanding of crowd dynamics, they lack detailed representation of individual decision-making.

2.1.6 Human Factors and Environmental Influences

Passenger distribution on platforms is influenced by both human factors and environmental conditions and is commonly analyzed using three main types of models. Behavioral models focus on passenger decision-making and preferences, such as comfort, proximity to entrances or exits, and avoidance of crowded or obstructed areas, using observational data or surveys to explain individual behavior and platform utilization (Ferri & Popp, 2022) [49], Van Hagen, 2011 [4].

Environmental impact models examine how weather, lighting, and shelter affect location choices. Adverse weather encourages passengers to gather in covered areas, while favorable conditions result in more even distribution along the platform (Kuipers et al., 2021) [52]. These models typically combine weather sensor data with passenger observations to quantify environmental effects on waiting behavior (Seriani & Fujiyama, 2019 [15]).

Integrated simulation models combine behavioral and environmental factors to simulate platform occupancy under different conditions. Agent-based and pedestrian flow simulations show how infrastructure, passenger preferences, and external conditions interact to shape distribution patterns (Zhang, 2021 [26], Seriani et al., 2016 [46], Bosina et al., 2015 [47]).

Together, these approaches identify key influences on waiting location distribution, including passenger preferences, crowding tendencies, and environmental conditions, and provide a basis for improving platform design, wayfinding, and operational strategies to enhance safety, efficiency, and user experience (Daamen et al., 2005 [1], Páez & Scott, 2005 [8]).

2.1.7 Conclusion

The literature indicates that passenger distribution on train platforms is shaped by several interacting factors (Daamen et al., 2005) [1], (Páez & Scott, 2005) [8]. Station and platform configuration, including location within the station, access points, circulation space, and geometric layout, strongly influence where passengers position themselves (Baron, 2021) [21], (Fruin, 1987) [23], (Seriani et al., 2016) [46], (Bosina et al., 2015) [47]. Platforms with well-distributed entrances, wider standing areas, and unobstructed sightlines support smoother flows and reduced clustering (Seriani et al., 2016) [46], (Bosina et al., 2015) [47].

Wayfinding and information systems also play a central role in guiding waiting passengers (Ahn et al., 2017) [25]. Clear signage, real-time information, and visible platform markers help travelers select appropriate waiting locations and reduce uncertainty, particularly during busy periods or service disruptions (Xu et al., 2020) [27], (Romero et al., 2023) [33]. Train operations, including timetable structure, dwell times, and door or carriage positions, directly affect waiting locations (Li & Zhu, 2016) [39], (Ingvardson et al., 2018) [41], (Oliveira et al., 2019) [42]. Irregular operations, such as delays or platform changes, can rapidly alter distribution patterns as passengers adjust their expectations (Li & Zhu, 2016) [39]. The physical characteristics of platforms, such as obstacles, furniture, shelters, and available standing width, determine how efficiently space is used (Bosina et al., 2015) [47], (Seriani et al., 2016) [46]. Local bottlenecks or poorly placed objects can create crowding even when overall density is moderate (Bosina et al., 2015) [47].

Much of the literature relies on static or limited-scope data collection methods, including manual observations, passenger counts, surveys, and fixed-camera analytics (Szplett & Wirasinghe, 1984) [43][44], (Krstanoski, 2014) [45], (Zhang, 2021) [26], (Xu et al., 2020) [27]. While these provide useful snapshots, they often cover short time windows or specific locations, limiting insight into variability across operational scenarios, weather conditions, and passenger groups (Seriani & Fujiyama, 2019) [15], (Kuipers et al., 2021) [52]. This highlights the need for more continuous, high-resolution, and data sources capturing behavioral variation (Zheng et al., 2023) [35], (Hänseler et al., 2020) [36].

Finally, human behavioral factors, including personal preferences, crowd avoidance, and responses to environmental conditions, significantly shape waiting patterns (Ferri & Popp, 2022) [49], (Van Hagen, 2011) [4]. Weather, perceived comfort, and anticipation of train arrival further influence passenger positioning (Kuipers et al., 2021) [52], (Seriani & Fujiyama, 2019) [15].

Overall, platform occupancy emerges from the combined effects of infrastructure, information, operations, and human behavior (Daamen et al., 2005) [1], (Páez & Scott, 2005) [8]. Understanding these relationships is essential for designing safer and more efficient platforms and for improving passenger distribution during both regular and disrupted operations (Ahn et al., 2017) [25], (Zhang, 2021) [26].

2.2 Key models to simulate waiting location choice

Research on passenger distribution has evolved through several modelling approaches, each offering different insights into how people choose where to wait on train platforms. Broadly, these approaches can be grouped into three main modeling approaches: (1) movement-based simulation models, (2) spatial-structural and flow-based models, and (3) behavioral choice models. Each paradigm captures different aspects of platform dynamics and waiting behavior.

2.2.1. Movement-based simulation models

Early and widely used tools mainly replicate physical movement. Agent-based models (ABM) simulate passengers as autonomous agents navigating station layouts using predefined movement rules (Zhang, 2021) [26]. They are valuable for analyzing how geometry, crowding, and access points affect flow patterns, but they underrepresent the behavioral motivations behind preferences for specific waiting locations.

A similar limitation applies to cellular automata models, which discretize platforms into grids and simulate movement step by step (Seriani et al., 2016) [46]. These models effectively capture congestion formation and the effects of platform dimensions and obstacles, yet they prioritize movement efficiency over individual preferences, limiting their ability to explain detailed waiting location choices.

2.2.2 Spatial-structural and flow-based models

To understand how the physical environment channels passenger movement, researchers have applied space syntax analysis (Bosina et al., 2015) [47], which quantifies how structural elements such as columns, barriers, and queuing zones affect visibility, accessibility, and spatial integration. Although useful for identifying design-induced bottlenecks, this method does not capture behavioral or environmental factors influencing voluntary waiting locations.

Other studies focus on operational dynamics. Discrete Event Simulation (DES) models how delays and timetable disruptions affect passenger accumulation on platforms (Li & Zhu, 2016) [39]. While these simulations capture sudden changes in crowd levels, they treat passengers as a collective flow rather than heterogeneous decision-makers.

Macroscopic flow models analyze density, walking speed, and flow rate using aggregate data (Zhang et al., 2020) [48]. These models identify when and where congestion occurs but cannot explain the underlying choice mechanisms behind waiting location selection.

Some studies incorporate location-specific attraction effects. Potential-based models, such as those by Szplett & Wirasinghe (1984i, 1984ii) [43][44] and Krstanoski (2014) [45], assign attraction strengths to platform areas to represent how passengers gravitate toward certain locations. However, these models rely on manual counts or static video data and approximate behavioral reasoning without explicitly modelling trade-offs between competing attributes.

2.2.3 Behavioral and choice-oriented models

More recent research has incorporated behavioral surveys, observational data, and environmental measurements to create behavioral, environmental, and integrated simulation models (Ferri & Popp, 2022 [49], Kuipers et al., 2021 [52], Seriani & Fujiyama, 2019 [15]). These approaches reveal how comfort, shelter, lighting, weather, and crowd avoidance influence waiting patterns. While they provide rich descriptive insights, they do not formally quantify individual utilities or choice probabilities, limiting their ability to predict how passengers would respond to new platform designs or operational scenarios.

Across these three modelling approaches, one recurring limitation becomes clear: most approaches simulate movement, space usage, or crowd dynamics well, but they struggle to represent the decision-making process behind waiting location choice in a formal and predictive manner.

2.2.4 Discrete Choice Models as the preferred approach

For this reason, Discrete Choice Models (DCMs) have become increasingly important in platform research. Unlike movement-based or flow-based models, DCMs explicitly model individual decision-making by assuming that passengers choose the waiting location that maximizes their perceived utility (Train, 2009) [18]. Utilities can be expressed as functions of observable platform attributes such as proximity to exits, expected train door positions, shelter availability, crowding level, lighting conditions, or walking distance, as well as unobserved factors.

DCMs offer several advantages over the alternative approaches discussed above:

- They explicitly represent behavioral trade-offs, rather than implicit attraction or movement rules.
- They allow the estimation of attribute importance (e.g., how much crowding or distance passengers are willing to tolerate).
- They support preference heterogeneity through advanced structures such as Nested Logit or Mixed Logit models.
- They enable policy and design evaluation, making it possible to simulate how changes in platform layout, information systems, or shelter provision would alter waiting location choices.

In contrast, ABMs and cellular automata are strong in reproducing physical interactions but weak in behavioral realism, space syntax and flow models explain spatial constraints but not voluntary choices, and potential-based models lack a formal decision-theoretic foundation.

Given its ability to directly capture individual preferences, quantify attribute effects, and predict behavioral responses under alternative scenarios, the Discrete Choice Modelling framework provides the most suitable methodological foundation for analyzing waiting location choice in this study and is therefore adopted as the core modelling approach.

2.3 Research Gap identification

Although existing studies provide valuable insights into platform design, passenger movement, and waiting behavior, several important gaps remain. First, environmental influences on waiting behavior, such as weather, lighting, and shelter availability, are only partially incorporated in existing waiting behavior models. While some studies highlight how adverse weather conditions lead to clustering in covered areas (Kuipers et al., 2021) [52], most modeling frameworks either treat these factors superficially or omit them entirely. This results in behavioral models that insufficiently reflect the situational conditions under which waiting decisions are actually made.

Second, the empirical basis of many studies relies on static or limited-scope data collection methods, including manual observations, periodic counts, surveys, and fixed-camera video analytics (e.g., Daamen & Hoogendoorn, 2003)[3]. These methods provide valuable snapshots of behavior but do not fully capture the context-dependent nature of waiting location choice, particularly under varying environmental conditions. While Section 2.1 discussed the strengths and limitations of these methods, the literature shows a clear need for more integrated, multi-source datasets capable of capturing real-time behavioral variation.

Third, and most critically, there is a lack of comprehensive modeling approaches that jointly represent spatial, and environmental determinants of waiting location choice. Existing models often focus on isolated components: spatial layout (e.g., obstacles, platform geometry), or environmental influences (e.g., weather effects), but rarely consider their interactions within a unified predictive framework. As Sections 2.1.1-2.1.6 showed, waiting behavior emerges from the interplay of passenger preferences, platform design, crowding tendencies, and environmental conditions, yet existing models typically capture only parts of this complexity.

To address these gaps, this study develops an integrated waiting location choice model that combines spatial, environmental, and behavioral factors into a single predictive framework. By drawing on Discrete Choice Modeling and incorporating empirically grounded spatial, behavioral and environmental attributes, the model aims to more accurately reflect the complex decision-making processes shaping passenger distribution on train platforms.

2.4 Conceptual framework

2.4.1 Comprehensive Conceptual Framework

The comprehensive conceptual framework Figure 1_2.4.1.1 is developed directly from the literature synthesis presented in Section 2.1. These sections demonstrate that passenger distribution on railway platforms is a multidimensional phenomenon shaped by infrastructure design, operational processes, behavioral mechanisms, and environmental context. To reflect this complexity, the comprehensive framework brings together all major categories of influence identified in previous studies into a single integrated structure, with passenger distribution over the platform positioned as the central dependent variable.

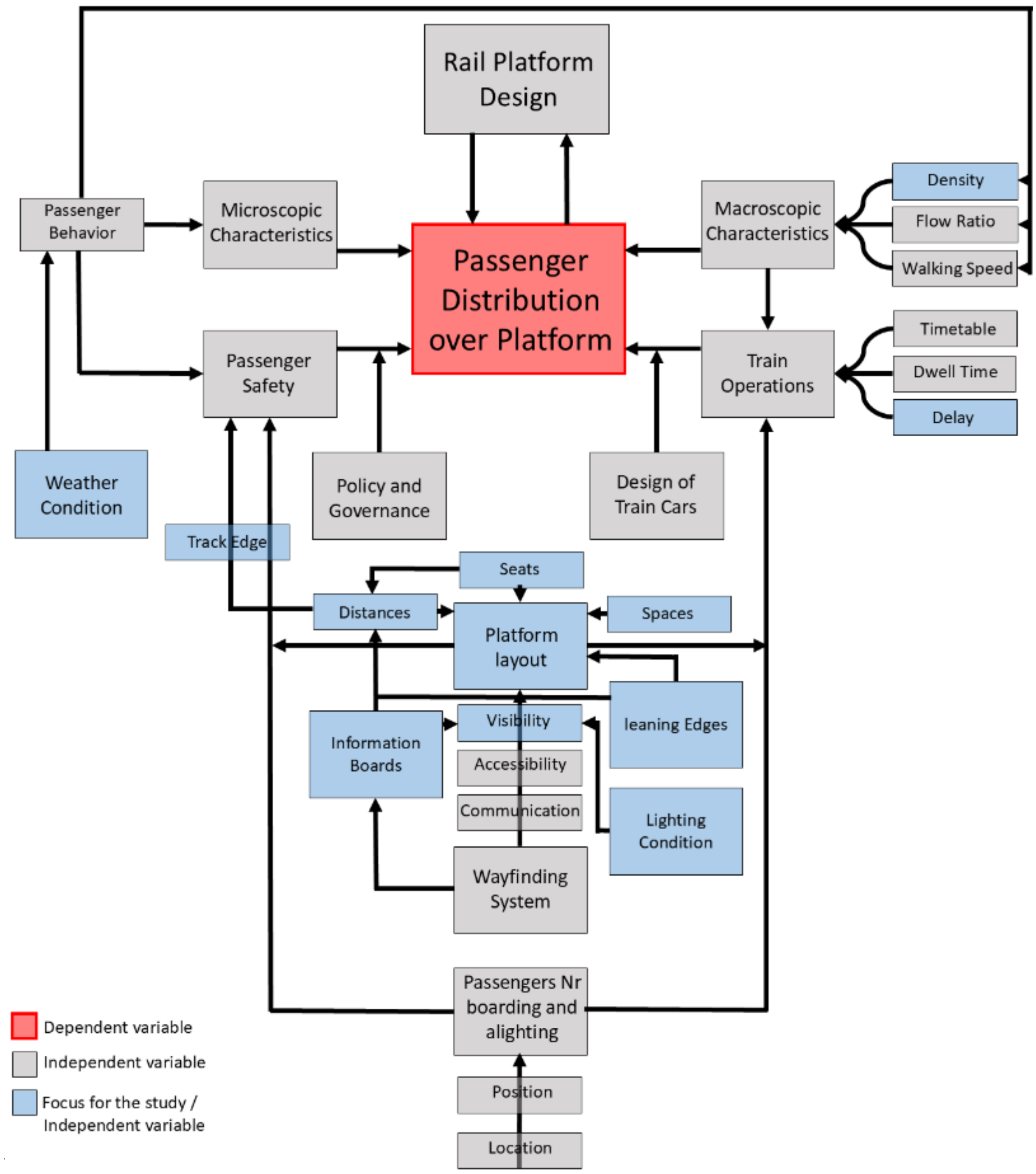

Figure 1_2.4.1.1 Comprehensive interaction conceptual framework for passenger distribution on railway platforms

***Figure 1_2.4.1.1:* Conceptual framework synthesizing findings from the literature to illustrate how infrastructure design, operational characteristics, policy and governance, train car design, passenger behavior, and environmental conditions interact to determine passengers' spatial distribution and waiting location choices on railway platforms.**

At the highest level, the framework distinguishes four primary groups of independent variables: rail platform design, train operations, policy and governance, and design of train cars. Rail platform design captures the physical configuration of the station environment, including platform width and length, circulation space, access points, and the placement of facilities. These elements directly shape pedestrian movement opportunities and spatial constraints, thereby influencing where passengers can and prefer to wait.

Train operations represent the temporal and service-related dimension of platform use. Timetable structure, dwell time, and delays determine the rhythm of passenger arrivals and departures and strongly affect short-term crowd accumulation and redistribution processes. These operational variables give rise to macroscopic characteristics of the system, such as overall passenger density, flow ratios, and walking speeds, which describe the collective dynamics of movement and congestion on the platform.

Policy and governance are incorporated to account for institutional and regulatory influences, including safety regulations, platform usage rules, and crowd management strategies. Although these factors do not always manifest in directly observable spatial patterns, they shape both operational practices and individual behavior by defining permissible actions, guiding infrastructure provision, and structuring passenger flows.

The design of train cars forms a further structural component of the framework. Characteristics such as the number and placement of doors, seating layout, and carriage configuration affect boarding and alighting processes and therefore influence where passengers position themselves along the platform prior to train arrival.

At the microscopic level, passenger behavior is explicitly represented as a key driver of distribution. Individual movement patterns, waiting preferences, and avoidance strategies determine how passengers respond to the spatial and operational conditions they encounter. These behavioral processes interact closely with passenger safety considerations and weather conditions, such as rain, cold, or heat, which can alter perceived comfort and risk and thereby shift preferred waiting locations.

The framework further specifies platform layout as a central mediating layer between infrastructure design and observed passenger distribution. Platform layout is decomposed into concrete spatial attributes, including the availability and location of seats, open spaces, distances to functional elements, visibility conditions, information boards, leaning edges, track edges, and lighting. Together, these attributes define the set of feasible and attractive waiting locations. A dedicated wayfinding system, encompassing signage, accessibility features, and communication mechanisms, supports this layout by guiding passenger movement and shaping spatial awareness.

Finally, the framework incorporates the number of passengers boarding and alighting, as well as their typical positions along the platform, to account for demand-related pressures that directly modify local crowding levels and clustering tendencies.

By integrating these micro- and macro-level components, the comprehensive conceptual framework offers a holistic representation of the mechanisms that generate passenger distribution patterns on railway platforms. Its purpose is not to prescribe a single causal pathway, but to map the full network of interactions identified in the literature and to ensure that all theoretically relevant influences are acknowledged before analytical simplification.

2.4.2 Focused Conceptual Framework for the Empirical Study

The discussion in Section 2.1 has shown that passenger distribution on train platforms is not the result of a single dominant factor, but rather emerges from the interaction between infrastructure design, operational conditions, environmental context, and individual decision-making processes. The literature review identified six recurring groups of determinants, station and platform design, wayfinding and safety systems, train operations and train design, platform layout, pedestrian movement behavior, and human and environmental factors, and demonstrated how each contributes to shaping where passengers choose to wait.

Building on this structured synthesis of previous research, a conceptual representation is required to translate the theoretical insights from the literature into an analytical framework that can guide empirical modelling. The comprehensive conceptual framework therefore serves as the first step in this process. It integrates the full range of factors discussed in the preceding sections into a single system, explicitly linking operational variables, spatial characteristics of stations and platforms, passenger behavior attributes, and environmental conditions to the central outcome of interest: passenger distribution over the platform. In doing so, it mirrors the breadth of the literature and ensures that the complexity of real-world platform dynamics is adequately acknowledged.

However, while the comprehensive framework is valuable for conceptual completeness, not all of its components can be directly operationalized in an empirical model. Some variables identified in the literature, such as policy and governance structures, long-term institutional rules, or network-level planning decisions, affect passenger distribution only indirectly and cannot be reliably observed or quantified at the level of individual platforms or waiting locations. Moreover, the modelling approach adopted in this study, namely discrete choice modelling, requires that explanatory variables can be expressed as measurable attributes associated with specific waiting location alternatives.

For this reason, the comprehensive framework is systematically refined into a focused conceptual framework Figure 2_2.4.2.1. This focused version does not replace the broader framework but is derived from it by selecting the subset of factors that (i) are consistently highlighted in the literature as influential for waiting location choice, (ii) can be observed or constructed from the available data sources, and (iii) can be meaningfully incorporated into a behavioral choice model.

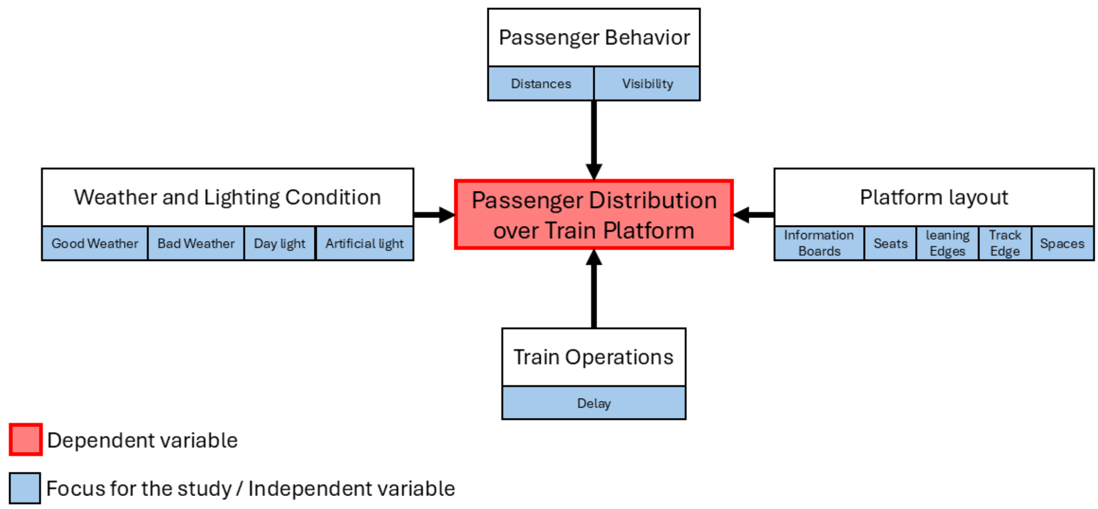

Figure 2_2.4.2.1 Focused conceptual framework for empirical modelling of passenger waiting location choice

Figure 2_2.4.2.1 presents the focused conceptual framework, which is derived from the comprehensive framework and highlights the variables that can be operationalized in the empirical analysis, as well as their direct relationships with passenger waiting location choice on the platform.

In the focused framework, passenger distribution over the train platform remains the dependent variable. The independent variables are narrowed to four interrelated groups that directly correspond to the mechanisms identified in the literature review:

- Platform layout attributes including distances to relevant elements (such as access points and train doors), visibility conditions, the presence and location of seats, leaning edges and track edges, spatial availability, and wayfinding and information boards. These variables operationalize the design-related influences discussed in Sections 2.1.1 and 2.1.4.
- Train operations, represented by delays, reflecting the operational effects on passenger concentration and repositioning described in Section 2.1.3.
- Train design characteristics, limited to aspects that influence boarding location choice, particularly the design of train cars and door configurations, as highlighted in the review of train–platform interaction effects.
- Environmental conditions, restricted to weather and lighting conditions, representing the most consistently documented external influences on waiting behavior in the literature.

Other elements present in the comprehensive framework, most notably policy and governance factors, are retained as contextual background but excluded from the empirical specification because they cannot be directly linked to individual waiting location choices within the available observational setting.

This stepwise development from literature review to comprehensive framework and finally to a focused analytical model establishes a clear connection between the running text and the conceptual structure of the study. The focused framework therefore acts as the bridge between theory and methodology: it translates the qualitative insights from previous research into a coherent set of quantitative variables suitable for discrete choice modelling.

By explicitly documenting this progression, the conceptual framework is not introduced as an isolated diagram but as the logical outcome of the preceding theoretical and methodological discussion. It defines the analytical boundaries of the study, clarifies which mechanisms are examined empirically, and provides the structural foundation for the model formulation and estimation presented in the subsequent chapters.

3 Methodology

Understanding passenger distribution on train platforms requires a comprehensive approach combining data collection and behavioral modeling. This methodology identifies the required datasets, gathers precise sensor data, and applies the Discrete Choice Model (DCM) to analyze passenger behavior and assess how different factors influence distribution patterns.

3.1 Data Requirements

To proceed with the analysis and develop the Discrete Choice Model (DCM) for studying passenger distribution on train platforms, several datasets are required, being:

- Sensor Data to capture passengers' movement trajectories (time, stamped data to distinguish between walking and waiting passengers). This data is used to identify their entry and exit points, as well as their spatial distribution across the platform.
- Platform Layout Data to detail the spatial configuration, such as the dimensions of the platform, seating areas, information boards, and physical features like track edges.
- Environmental factors to capture weather and lighting conditions.

3.2 Data Collection Plan

The data collection plan emphasizes the utilization of sensor data as a primary source to capture detailed passenger behavior on train platforms. While this is the only available data, it is crucial to recognize that sensor data alone has limitations.

The use of sensors in monitoring passenger movement on train platforms has inherent limitations, primarily due to privacy concerns and the capabilities of the technology employed. Currently, these sensors are designed to capture only the x and y coordinates of passenger movements, which restricts the depth of behavioral insights that can be gleaned from the data. While they can effectively track foot traffic patterns, they fail to capture nuanced behaviors such as waiting times, group dynamics, and individual interactions with the environment. This simplification may overlook critical factors influencing passenger decisions and experiences.

The choice of sensor placement, strategically located above escalators and along the platform, was likely dictated by practical considerations. These include the existing station layout and the need to monitor passengers without disturbing them, and cost-effectiveness, as installing more advanced sensors could be prohibitively expensive. This approach ensures a basic level of operational efficiency but may compromise the richness of the data collected. As a result, while the current system provides valuable information on movement patterns, it is limited in its ability to inform comprehensive analyses of passenger behavior, necessitating further research and potentially more advanced technology in the future.

If we could define and install sensor types based on the conceptual framework, a broader array of sensors could be utilized to capture a more comprehensive range of data. For instance, infrared sensors or thermal imaging could help differentiate between different passenger activities, such as sitting, standing, or walking, which the current sensors might not fully capture.

To distinguish walking passengers from those waiting, the sensor data is analyzed for movement speed and direction. Walking passengers typically display continuous movement with a consistent trajectory, whereas waiting passengers have minimal movement and remain in a confined area. This distinction is vital for understanding platform utilization, whether passengers are in transit or waiting for their train. The detailed insights provided by this data include specific traffic variables such as entry and exit points, density distribution, flow rates, and waiting area occupancy.

3.3 Dataset description

The data collected in this study is carefully structured to analyze passenger choice behavior with high temporal granularity. The choice of a 0.1 second interval for sensor data collection is based on the need to capture real-time variations in passenger movements. This level of detail ensures that subtle behavioral shifts, such as walking speeds and positional changes, are captured, providing a more accurate understanding of the decision-making process. Additionally, the dataset is synchronized with train arrival times, focusing on the critical two-minute window before a train's arrival. This window was chosen based on observed passenger behavior, where decisions about waiting positions are typically made during this period of heightened attentiveness. A shorter window would miss key moments of decision-making, while a longer one might dilute the focus on final movements. Passengers make choices based on proximity to their train's boarding point, visibility, and the availability of seating or shelter. These assumptions are grounded in previous literature on pedestrian behavior but will be tested through this data. Key environmental variables, such as weather and lighting conditions, are included to understand how external factors influence decisions. The study spans four months in 2024, covering both good weather (June, September) and bad weather (February, November), defined quantitatively by temperature and rainfall data and by avoiding holiday periods when the flow is relatively limited. The distinction between daylight and artificial lighting periods is aligned with meteorological data for each season, ensuring accurate representation of environmental conditions.

To provide clarity, the dataset is organized to represent each passenger's decision context. At every snapshot in time, all passengers are treated as decision-makers, and the data is duplicated with each passenger removed in turn, reflecting their choice within that moment. This approach, while offering high-resolution insights, introduces some limitations. For instance, it does not account for sequential passenger arrivals or interactions, potentially oversimplifying dynamic interactions on the platform. Moreover, duplicating snapshots could lead to data redundancy and increase the risk of overfitting, where the model becomes overly sensitive to the specific context of each snapshot rather than generalizable trends.

Despite these limitations, the high-frequency data provides a robust basis for analyzing individual and collective behavior on the platform. By isolating individual decisions and focusing on key contextual factors, the dataset offers deep insights into passenger preferences and decision, making processes. However, to fully understand the limitations of this method, further research needs to consider dynamic interactions, particularly how passengers influence each other's positioning on the platform. The final dataset features 142,256 rows, Table A (Appendix) a snapshot of the final dataset.

3.4 Sampling

The sampling process in this study was designed to create a representative subset of passengers while preserving the diversity of environmental, temporal, and operational conditions present in the full dataset. The original dataset contained 142,256 passenger records, each representing an individual detected on the platform together with its associated timestamp, weather condition, lighting condition, and platform location. Before sampling, all records were enriched with environmental and platform attributes by joining the timestamp of each passenger with external weather and lighting datasets and mapping the passenger's coordinates to platform sections. This ensured that every pedestrian observation contained both behavioral and contextual information.

The first step was defining how many records were needed for statistical validity. Based on a 95% confidence level, a 5% margin of error, and a proportion $p = 0.5$, the standard sample size formula Equation 3.4.1 for a finite population Israel, G. D. (1992) [53] was applied:

$$n = \frac{Z^2 \cdot p \cdot (1-p)}{E^2} \qquad n_{adjusted} = \frac{n}{1 + \frac{n-1}{population\_size}}$$

Equation 3.4.2: sample size formula

With $N = 142{,}256$, this resulted in an adjusted minimum sample size of 384 observations. However, initial inspection of the dataset revealed that passenger presence and temporal patterns varied strongly across different environmental and lighting conditions. To visualize these differences and assess the balance of observations across scenarios, Figures 3.4.1a–3.4.1c present heatmaps of passenger record frequencies by day of week and time period for three representative cases: good weather & good lighting, bad weather & good lighting, and bad weather & bad lighting.

Figure 3.4.1a shows that under favorable conditions, passenger counts concentrate strongly during weekday morning and evening peak periods, particularly on Mondays and Fridays. Figure 3.4.1b indicates that under bad weather but adequate lighting, overall presence decreases, especially during weekends, while peak-hour flows remain clearly visible. In contrast, Figure 3.4.1c demonstrates that the combination of bad weather and poor lighting leads to a sharp reduction in recorded passengers, with several time periods (notably Sunday evenings) exhibiting very sparse observations.

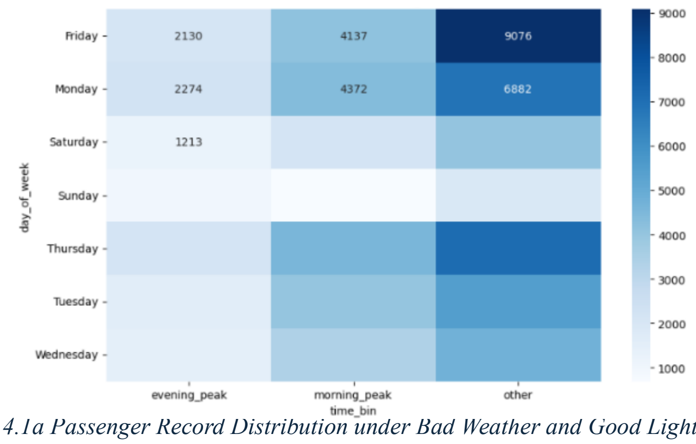

Figure 4_3.4.1a Passenger Record Distribution under Bad Weather and Good Lighting Conditions

***Figure 3_3.4.2a:* Heatmap illustrating the frequency of passenger records by day of week and time period under bad weather but good lighting conditions. Compared with fair-weather conditions, overall passenger presence is lower, especially during weekends, while morning and evening peak periods remain clearly identifiable.**

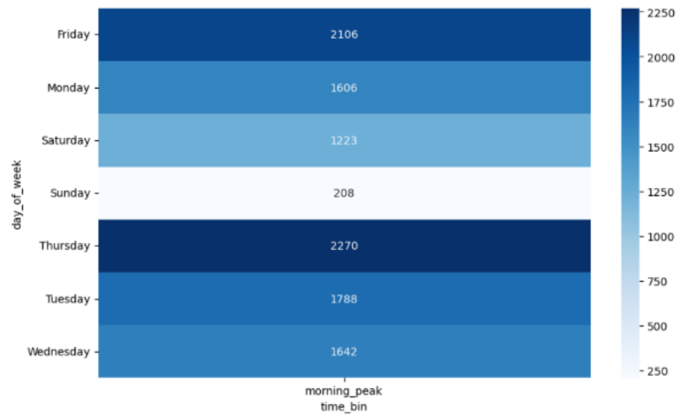

Figure 3_3.4.2b Passenger Record Distribution under Bad Weather and Good Lighting Conditions

***Figure 4_3.4.2b:* Heatmap illustrating the frequency of passenger records by day of week and time period under bad weather but good lighting conditions. Compared with fair-weather conditions, overall passenger presence is lower, especially during weekends, while morning and evening peak periods remain clearly identifiable.**

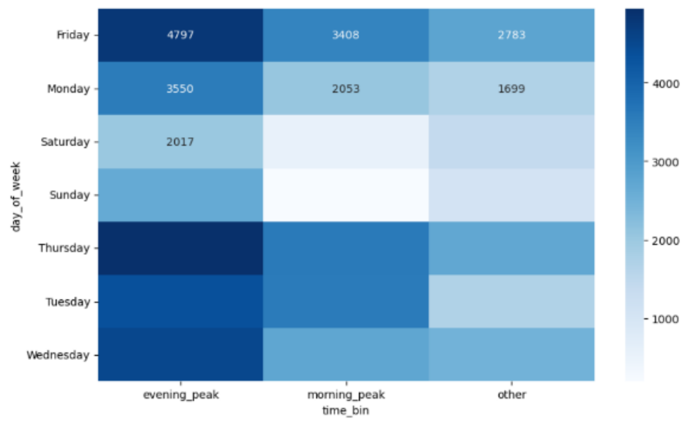

Figure 5_3.4.3c Passenger Record Distribution under Bad Weather and Bad Lighting Conditions

***Figure 5_3.4.3c:* Heatmap depicting the frequency of passenger records by day of week and time period under combined bad weather and poor lighting conditions. The figure shows a substantial reduction in observations across most periods, with particularly sparse records during off-peak hours and weekend evenings, reflecting the strong suppressing effect of adverse environmental conditions on platform usage.**

These visual insights revealed an important challenge: the dataset was highly imbalanced, meaning some situations were heavily represented while others had very few observations. Because the purpose of the study is to model waiting location choice across a broad range of realistic platform conditions, relying solely on a random sample of 384 records would severely underrepresent many valid combinations of weather, lighting, time of day, and day of the week.

To overcome this, the sampling strategy was expanded into a structured, stratified sampling approach. The full dataset was divided into 84 distinct situations, derived from all possible combinations of:

- **Weather:** Good (June, September) / Bad (February, November)
- **Lighting:** Good (daylight) / Bad (artificial light)
- **Time of day:** Morning Peak (07:00–09:00), Evening Peak (16:00–18:00), Other Times
- **Day of the week:** Monday–Sunday

This creates

$$2 \times 2 \times 3 \times 7 = 84 \text{ unique situations.}$$

However, when mapping the dataset into these 84 bins, 35 situations contained too few observations to be usable. For example, combinations of good weather with bad lighting rarely occur because good weather generally coincides with sufficient natural light. Similarly, bad weather with good lighting occurred mainly during morning peaks, where overall pedestrian counts are lower. To maintain reliability, these sparse situations were excluded from sampling, leaving 49 well-represented situations.

The next step involved selecting samples within each of these 49 situations. Instead of sampling the same number of passengers from each bin, which would distort their true distribution, a proportional sampling method was used. For every bin, a number of passengers was drawn proportional to its relative frequency in the full dataset, while ensuring that every situation remained represented. This process resulted in a final sample of 2,499 pedestrian observations. This enlarged sample ensured adequate coverage of all relevant combinations of environmental conditions and platform attributes, while still keeping the computational running time of the model reasonable.

By structuring the data in this way, each final pedestrian record in the sample was fully connected to its environmental attributes (weather and lighting at the time of observation) and platform attributes (precise platform section, distance to entrances, and other spatial characteristics). As a result, the sampled dataset is not only statistically sound but also behaviorally and contextually rich. This prepares a solid foundation for modeling waiting location choice in the subsequent analysis.

3.5 Attributes

Understanding the attributes influencing passenger decision-making on the train platform is essential for optimizing both platform design and operational management. Each attribute captures a specific aspect of the platform environment, infrastructure, or user behavior that affects where passengers choose to stand or wait. These variables were identified through descriptive analysis, literature findings, and behavioral reasoning derived from the conceptual framework.

The attributes included in the systematic component of the utility function represent observable characteristics of the alternatives (cells) and interaction effects that capture how environmental and spatial factors jointly influence passenger preferences.

The selected attributes are described as follows:

- Passenger Density (B_persons_per_m2)

  Represents the number of passengers per square meter within a cell, reflecting perceived crowding and comfort. Higher density typically reduces a cell's utility due to discomfort, lower accessibility, and safety concerns.

- Weather Conditions (B_weather)

  Capture the impact of environmental conditions (e.g., dry vs. rainy) on passenger spatial preferences. Under unfavorable weather, passengers tend to cluster in sheltered or central areas, indicating behavioral adaptation to environmental stressors.

- Lighting Conditions (B_lighting)

  Represent the level of illumination across the platform. Adequate lighting improves perceived safety and visibility, while poor lighting conditions discourage passengers from occupying darker or peripheral zones.

- Distance to Entrance (B_dist_to_entrance)

  Measures the cell's distance from the main platform entrance, representing accessibility and convenience. Shorter distances increase attractiveness, especially for passengers boarding or alighting frequently.

- Distance to Information Board (B_dist_to_information_board)

  Indicates how close a cell is to information boards. Proximity facilitates access to real-time travel information and enhances situational awareness, though excessive clustering near these areas may cause local congestion.

- Proximity to Leaning Areas (B_is_next_to_leaning)

  Denotes whether a cell is adjacent to a leaning rail. Leaning facilities provide comfort and mild physical support, making nearby areas more attractive, especially during long waiting periods.

- Proximity to Seats (B_is_next_to_seats)

  Indicates whether a cell contains or is adjacent to seating areas. Seating availability is a critical comfort factor and often drives passengers to cluster around these amenities.

- Proximity to Information Boards (B_is_next_to_information_board)

  Captures whether a cell is near information displays, which are essential for navigation and trip planning. Cells near these boards are typically preferred during uncertain or dynamic travel conditions.

- Proximity to Kiosks (B_is_next_to_kiosk)

  Measures whether a cell is near retail or refreshment kiosks. Proximity enhances convenience and is associated with higher waiting satisfaction and dwell time.

- Proximity to Track Edge (B_is_next_to_track)

  Indicates whether a cell is directly adjacent to the track. This factor involves a trade-off between visibility (desirable for train anticipation) and perceived safety risk (undesirable for comfort).

Table 1_3.5.Aa Attributes Included in the Utility Function for Passenger Waiting Location Choice

Table 1_3.5.1a presents all attributes, their descriptions, and corresponding variable codes used in the model estimation.

Attribute	Description	Type	Value
persons_per_m2	Represents the density of passengers in a particular cell location, influencing passenger choices and indicating crowded conditions.	Ordinal	
weather	Reflects weather conditions that impact passenger behavior and preferences on the platform, guiding decisions based on environmental factors.	Dummy	Good weather 1 Bad weather 0
lighting	Indicates the quality of lighting on the platform, essential for safety and visibility, particularly during nighttime or low-light conditions.	Dummy	Day light 1 Artificial light 0
dist_to_entrance	Measures the proximity of cell location to the platform entrance, influencing convenience and accessibility for boarding and disembarking passengers.	Ordinal	
dist_to_information_board_below	Distance to the right information board from the entrance	Ordinal	
dist_to_information_board_upper	Distance to the left information board from the entrance	Ordinal	
dist_to_kiosk_upper	Distance to the upper (left) side of the kiosk (seating area)	Ordinal	
dist_to_kiosk_below	Distance to the lower (right) side of the kiosk (seating area)	Ordinal	
dist_to_kiosk_entrance	Distance to the entrance side of the kiosk	Ordinal	
dist_to_kiosk_non_entrance	Distance to the back side of the kiosk (leaning edge)	Ordinal	
dist_to_seats_upper	Distance to the upper (left) side of the seats at the back of the platform	Ordinal	
dist_to_seats_below	Distance to the lower (right) side of the seats at the back of the platform	Ordinal	
dist_to_seats_right	Distance to the right side of the seats	Ordinal	
dist_to_seats_left	Distance to the left side of the seats	Ordinal	
dist_to_seats	Overall distance to the seats (to the centroid of the seats cells)	Ordinal	
dist_to_information_board	Overall distance to the closest information board	Ordinal	
dist_to_lean_edge_upper	Distance to the upper (left) side of the leaning edge at the front side of the platform	Ordinal	
dist_to_lean_edge_bottom	Distance to the lower (right) side of the leaning edge	Ordinal	
dist_to_track_edge_upper	Distance to the upper(left) side of the track edge	Ordinal	
dist_to_track_edge_bottom	Distance to the lower (right) side of the track edge	Ordinal	
dist_to_track_edge	Overall distance to the closest track edge	Ordinal	
dist_to_lean_edge	Overall distance to the closest leaning edge	Ordinal	
is_visible_from_entrance	Is the cell visibility from the entrance	Dummy	Visible 1 O otherwise
is_next_to_leaning	Consider the proximity to leaning areas, offering passengers a comfortable spot to rest or lean against while waiting.	Dummy	Next to leaning edge 1 O otherwise
is_next_to_seats	Assesses the availability of nearby seating options, influencing passenger choices for added comfort and convenience.	Dummy	Next to seating area 1 O otherwise
is_next_to_information_board	Examine the proximity to information boards providing essential updates and announcements, guiding decisions based on.	Dummy	Next to Information Boards 1 O otherwise
is_next_to_kiosk	Is the location next to a kiosk	Dummy	Next to Kiosk 1 O otherwise
is_next_to_track	Is the location next to the track (yes/no)	Dummy	Next to track 1 O otherwise

3.5.1 Interaction Terms

To capture more realistic behavioral dynamics, several interaction terms were incorporated into the utility function, representing the combined influence of environmental and spatial factors on waiting location choice. In discrete choice modelling, interaction terms are commonly used to relax the assumption that attributes affect utility independently and to account for context-dependent preferences (Train, 2009)[18].

The interaction terms were developed through a three-step process:

1. Theoretical motivation from the literature: Previous studies on platform behavior show that weather and lighting conditions significantly affect perceived comfort, safety, crowd tolerance, and the use of platform facilities such as seats, information boards, and sheltered areas (Seriani & Fujiyama, 2019, Ferri & Popp, 2022, Kuipers et al., 2021)[15, 49, 52]. These findings suggest that spatial attributes (e.g., distance to seats or platform edges) do not influence passengers uniformly but rather depend on environmental conditions.
2. Initial model specification with main effects and pairwise interactions: The modelling process started with a baseline specification including only main effects for spatial, behavioral, and environmental attributes. Subsequently, pairwise interaction terms (e.g., Weather × Lighting, Weather × Distance to Seats) were tested to explore whether simple dependency structures improved model performance.
3. Empirical testing and refinement: Alternative specifications were estimated and compared using standard goodness-of-fit indicators (log-likelihood, adjusted ρ^2) and coefficient stability. Pairwise interactions generally showed limited statistical significance and weak improvements in explanatory power. In contrast, interaction terms combining environmental conditions with spatial attributes produced more consistent coefficient signs, stronger behavioral interpretability, and measurable improvements in model fit. Only interaction terms that were theoretically meaningful and empirically robust were retained in the final specification.

Although interactions between two attributes are common in discrete choice models, waiting behavior on platforms is shaped by *joint environmental conditions* rather than by weather or lighting alone. Weather and lighting together determine overall environmental comfort and perceived safety, which in turn condition how passengers evaluate specific waiting locations. For example, proximity to seats may be attractive under dry and well-lit conditions but less relevant during rain or under poor lighting. Modelling only pairwise effects would implicitly assume that weather and lighting influence preferences independently, which contradicts both observational evidence and behavioral findings in platform studies (Seriani & Fujiyama, 2019, Ferri & Popp, 2022, Kuipers et al., 2021)[15, 49, 52]. Therefore, three-factor interaction terms of the form, Weather × Lighting × Spatial Attribute, were adopted to represent this conditional valuation process more accurately.

Interactions involving four or more factors were deliberately avoided, despite being theoretically possible, due to: increased risk of multicollinearity, reduced parameter stability, lower statistical power given the available sample size, and diminished behavioral interpretability.

This modelling strategy balances behavioral realism with statistical robustness and model transparency, in line with best practice in discrete choice analysis (Train, 2009)[18].

The final model includes the following interaction terms:

- Weather × Lighting × Proximity to Seats (interaction_weather_lighting_seats)
  Captures how the combined effect of favorable weather and bright lighting enhances the attractiveness of areas near seating. Under good environmental conditions, passengers are more likely to occupy comfort-oriented zones close to seats.
- Weather × Lighting × Visibility (interaction_weather_lighting_visibility)
  Reflects how passengers' visual comfort and spatial safety perceptions are jointly influenced by lighting quality, weather conditions, and the visibility of the cell from the platform entrance.
- Weather × Lighting × Density (interaction_weather_lighting_persons_per_m2)
  Represents how environmental conditions influence passengers' tolerance for crowded areas. Under poor lighting or unfavorable weather, densely occupied zones tend to be avoided, while favorable conditions increase crowd acceptance.
- Weather × Lighting × Distance to Entrance (interaction_weather_lighting_dist_to_entrance)
  Shows how lighting and weather jointly affect passengers' preference for proximity to platform entrances. Pleasant conditions reduce the need to remain close to entrances, whereas adverse conditions increase the value of quick access.
- Weather × Lighting × Distance to Information Boards (interaction_weather_lighting_dist_to_info_board)
  Captures how environmental comfort modifies passengers' willingness to wait near information boards for improved awareness and reassurance.
- Weather × Lighting × Distance to Track Edge (interaction_weather_lighting_dist_to_track_edge)
  Describes how environmental conditions influence safety-related positioning near the track edge, with poor lighting or weather discouraging proximity.
- Weather × Lighting × Distance to Leaning Edge (interaction_weather_lighting_dist_to_lean_edge)
  Represents how environmental comfort affects the attractiveness of leaning edges as informal waiting and resting locations.

Together, these interaction terms allow the model to represent context-dependent and nonlinear preference structures that cannot be captured by main effects alone. Their inclusion is grounded in behavioral theory (Train, 2009)[18], supported by previous empirical findings on platform environments (Seriani & Fujiyama, 2019, Ferri & Popp, 2022, Kuipers et al., 2021)[15, 49, 52], and justified through systematic model testing.

In combination with the main attributes, the final specification provides a behaviorally realistic and statistically robust representation of passenger waiting location choice, enabling the Discrete Choice Model to quantify both individual and combined effects of environmental conditions and spatial platform characteristics on passenger distribution.

Table 2_3.5.2 Interaction Terms Included in the Utility Function

Table 2_3.5.2: Overview of the interaction terms incorporated in the discrete choice model to represent the combined effects of weather, lighting, and spatial platform attributes on passengers' waiting location choices.

Attribute	***Description***	***Type***
interaction_weather_lighting_seats	Represents how the combined effect of weather and lighting influences passengers' preference for proximity to seating areas. Under good weather and sufficient lighting, seats become more attractive locations for waiting.	Interaction
interaction_weather_lighting_visibility	Captures how weather and lighting together influence the visibility of cells from the entrance, affecting perceived safety and comfort.	Interaction
interaction_weather_lighting_persons_per_m2	Reflects how environmental conditions (weather and lighting) modify passengers' tolerance to crowding levels on the platform.	Interaction
interaction_weather_lighting_dist_to_entrance	Indicates how weather and lighting affect passengers' preference for proximity to the entrance, where accessibility and exposure vary with environmental comfort.	Interaction
interaction_weather_lighting_dist_to_info_board	Describes how favorable environmental conditions influence the attractiveness of areas near information boards, where visibility and clarity are essential.	Interaction
interaction_weather_lighting_dist_to_track_edge	Represents how lighting and weather together shape passengers' willingness to stand near the track edge, balancing safety and convenience.	Interaction
interaction_weather_lighting_dist_to_lean_edge	Models how weather and lighting jointly impact the appeal of leaning edges as waiting locations, reflecting comfort preferences under different environmental conditions.	Interaction

3.6 Choice Set and Alternatives

In this study, the concept of the choice set is fundamental to analyzing passenger behavior on train platforms. A choice set refers to the range of alternative options passengers have when selecting where to stand or wait. These options are shaped by various factors such as proximity to the train, access to seating, visibility of information boards, and environmental considerations like weather and crowd density. The alternatives in the choice set typically represent different spatial cells on the platform, each offering distinct advantages based on passenger needs.

The platform is divided into multiple cells, with each cell representing a possible location a passenger may choose to occupy. These cells differ in terms of their location relative to platform amenities, such as the distance to entrances, seats, leaning edges, and track edges. Passengers are assumed to weigh these factors based on their preferences, mobility, and real-time conditions (e.g., crowding, lighting). For example, a passenger with limited mobility might prefer a cell closer to the platform entrance or near available seating, while others might choose a spot near the train tracks to ensure timely boarding.

The variety of alternatives in the choice set allows for an analysis of the complex decision-making process involved in passenger distribution. This approach ensures that all critical factors influencing passenger behavior are accounted for, enabling a robust understanding of how platform design and environmental conditions affect where passengers choose to wait.

The concept of the choice set is fundamental to this study, as it defines the range of possible spatial alternatives available to passengers when deciding where to stand or wait on the platform. In line with the conceptual framework, the choice set operationalizes how platform layout, environmental conditions, operational characteristics, and human factors interact to shape passenger distribution. Each cell within the platform represents a distinct alternative, characterized by attributes such as distance to entrances, distance to amenities (seating, leaning rails, kiosks, information boards), lighting, weather conditions, and local density. These factors correspond directly to the determinants identified in the literature review: spatial design (Section 2.1.1), ergonomic and safety considerations (Section 2.1.2), environmental influences (Section 2.1.5), and behavioral patterns (Section 2.1.6).

The total platform area at Eindhoven Central Station (Platform 1-2) measures approximately 475 meters in length and 8.5 meters in width, yielding a total surface area of roughly 4,037 m^2. This area was divided into 264 discrete cells, each measuring 1.8 × 2.0 meters (3.6 m^2). The chosen cell size corresponds approximately to the personal space footprint of standing passengers under moderate crowding conditions, as reported in studies of pedestrian flow and platform occupancy (Daamen & Hoogendoorn, 2003, Fruin, 1971). This scale allows for capturing meaningful behavioral variations, such as proximity to amenities or crowding differences, while maintaining computational efficiency in model estimation.

The decision to allocate up to four passengers per cell is a behavioral modeling assumption rather than a physical restriction. It reflects typical comfort and safety thresholds under moderate crowding (≈1.1 passengers/m²), consistent with comfort standards for standing density on platforms (Transportation Research Board, 2013)[66]. The passengers themselves are not aware of these virtual cell boundaries, rather, the segmentation serves as an analytical tool to translate continuous passenger positions into discrete alternatives required for the Multinomial Logit (MNL) model estimation. Thus, this structure facilitates model computation without constraining real passenger movement. Within each cell, passengers are assumed to evaluate multiple factors when choosing their location. Cells nearer to entrances offer accessibility advantages, especially for passengers with mobility constraints or luggage, linking to the platform layout and human factors dimensions of the framework. Cells adjacent to seating or leaning rails relate to comfort and ergonomic considerations, encouraging longer dwell times and aligning with findings on waiting behavior (Daamen & Hoogendoorn, 2003)[3]. Proximity to information boards and kiosks supports convenience and situational awareness, connecting to both wayfinding and operational characteristics. Conversely, track-edge distance reflects the safety component of the framework, where passengers balance convenience against perceived risk. Environmental factors, such as weather and lighting, modify these preferences dynamically, and passengers cluster in sheltered, well-lit cells under poor conditions but distribute more evenly under favorable conditions. Overall, defining the choice set in this structured way translates the conceptual framework into a spatially operational model. Each factor, layout, safety, environment, operations, and human behavior, is represented through measurable cell attributes, allowing the discrete choice model to statistically validate which aspects most strongly influence passenger decisions. This approach ensures behavioral realism while maintaining methodological rigor in analyzing passenger distribution patterns on the platform. Figure 6_3.6.1 illustrates the spatial grid used to construct the choice set and its alignment with platform features.

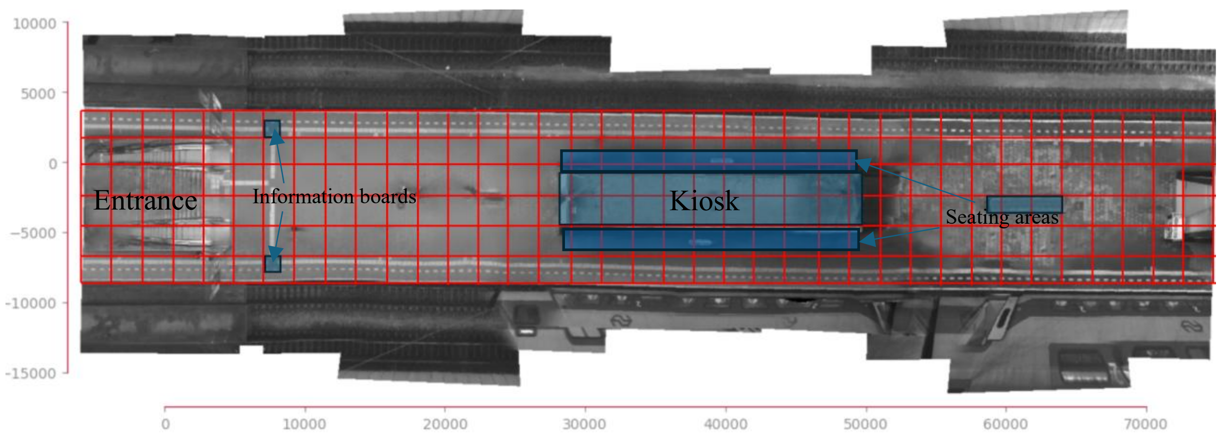

Figure 6_3.6.1 Spatial Discretization of the Platform into Choice Set Cells

Figure 6_3.6.1:* *Representation of the platform at Eindhoven Central Station divided into discrete spatial cells used to define the choice set in the discrete choice model, showing the alignment of cells with key platform features and amenities.

3.7 Descriptive Knowledge

Descriptive analysis is essential in this research as it provides foundational insights into passenger behavior and platform usage. By creating heat maps, the study visualizes passenger density, highlighting areas of congestion and underutilization. These visualizations are necessary to inform platform design and operational strategies. Key travel variables considered include density, and positional patterns, which are derived from sensor data. The heat maps will specifically show the average number of persons per square meter under different scenarios. By analyzing these results, the study identifies critical zones for targeted improvements in platform design, such as optimizing space usage, enhancing wayfinding, and improving overall passenger flow.

3.8 Correlation analysis of attributes

In this study, various statistical tests are applied to validate relationships between variables and interpret passenger distribution patterns on the platform. A correlation matrix assesses linear relationships between variables such as distances to amenities, lighting, and crowd density, identifying potential multicollinearity. Descriptive statistics, including histograms, are used to understand the skewness and density variations in passenger distribution across cells. These statistical tests help ensure the data is well-understood and aligned with the assumptions required for the Discrete Choice Model (DCM), enabling more accurate predictions of passenger behavior.

3.9 Modeling Approach

The Discrete Choice Model (DCM) is a statistical framework that predicts individual decision-making when faced with multiple alternatives. It operates on the assumption that each individual selects the option with the highest perceived utility, where utility is a function of various attributes and characteristics associated with each alternative. In this research, the focus is on using DCM to predict the probability that a passenger will choose a specific location on the train platform, influenced by factors such as platform layout, environmental conditions, and interactions with other passengers.

3.9.1 Model Specification

In this study, passenger waiting location choices on the platform are modeled using a Discrete Choice Model (DCM), specifically the Multinomial Logit (MNL) model. The MNL model is well suited for contexts where individuals choose among multiple discrete spatial alternatives. The model is grounded in Random Utility Theory (RUT), which assumes that each passenger selects the location that yields the highest perceived utility.

Under RUT, the utility of each alternative consists of two components:

- A deterministic (systematic) part, which captures the observable characteristics of each location, such as distance to entrances, benches, signage, shelter, density of surrounding passengers, lighting conditions, and weather.
- A random error component, capturing unobserved influences and individual-specific preferences.

The general utility function for each alternative i is expressed as:

$$U_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + \varepsilon_i$$

Where:

- U_i= total utility associated with alternative i
- $X_{1i}, X_{2i}, \ldots, X_{ki}$= observable attributes of alternative i
- $\beta_1, \beta_2, \ldots, \beta_k$= parameters to be estimated, representing the sensitivity of passengers to each attribute
- ε_i= random error term, assumed to be independently and identically distributed (IID) following a Gumbel distribution, consistent with the MNL assumptions

The probability that a passenger selects location i is determined by comparing the systematic utility of i to the utilities of all other available alternatives. The MNL model assumes rational decision-making, meaning passengers choose the alternative with the highest expected utility based on the attributes they perceive as important.

Thus, this specification provides a robust framework for capturing how multiple spatial, and environmental factors jointly influence passenger waiting location behavior on the platform.

3.9.2 Factors Included in the Model

The key factors included in the model are directly tied to the conceptual framework:

- Platform Infrastructure: Variables like distance to seating areas, proximity to track edge, visibility from entrances, and distance to information boards.
- Environmental Conditions: Weather and lighting conditions, as well as their interactions with the platform infrastructure, influencing passenger preferences.
- Passenger Interactions: crowd density which affects individual choices.

3.9.3 Data Description and Related Variables

Data for the model comes from sensor-based observations, capturing precise x and y coordinates, walking speeds, density, and spatial distribution on the platform. This data allows for a detailed analysis of how passengers move and where they choose to wait.

3.9.4 Estimation Procedure

The Multinomial Logit (MNL) model will be estimated using the Maximum Likelihood Estimation (MLE) method, which determines the set of parameter values (β coefficients) that maximize the likelihood of observing the given data (Ben,Akiva & Lerman, 1985)[13]. In other words, MLE identifies the parameters that make the observed passenger choices most probable under the model's assumptions. This approach is widely applied in discrete choice modeling because it ensures statistical efficiency and provides unbiased and consistent parameter estimates when the model is correctly specified (Train, 2009)[18].

In this study, 70% of the available dataset will be used for model estimation, while the remaining 30% will be reserved for model validation and performance testing. The rationale behind this split is to maintain a balance between estimation accuracy and the ability to generalize the model's predictive power. Using a sufficiently large portion of data for estimation allows the model to capture the underlying behavioral patterns and relationships between explanatory variables and passenger choices. Meanwhile, holding out a separate testing set prevents overfitting and enables a robust evaluation of how well the model performs on unseen data (Hastie, Tibshirani, & Friedman, 2009)[54].

3.9.5 Estimation Strategy

To identify the attributes that significantly influence passenger cell platform selection behavior on a train platform, a comprehensive estimation strategy was employed. This strategy began with an exploratory analysis of the dataset's descriptive statistics, providing insights into the distribution, variability, and relevance of various attributes associated with both the alternatives (platform cells) and the decision makers (passengers). Attributes were initially selected based on their theoretical importance and observed influence on platform choice behavior.

To refine the model and ensure its robustness, a stepwise forward selection procedure was adopted. The process started with a null model containing no attributes, and variables were sequentially added based on their theoretical relevance and contribution to the model's predictive performance. At each step, the statistical significance of each coefficient was assessed using its p-value, with a threshold of 0.05. Variables with p-values greater than or equal to 0.05 were considered non-significant and excluded from the model.

The forward stepwise method was guided by both theoretical expectations and empirical performance metrics. Initially, each variable was evaluated independently to determine its impact on model fit, using the Rho-square (init) as a measure of explanatory power. The variable with the highest Rho-square (init) was selected as the starting point. Subsequently, each remaining variable was added iteratively, and its contribution to improving model fit was evaluated. This process continued until no further improvement was achieved, ensuring that all included variables meaningfully enhanced the model's predictive ability.

In addition to statistical significance, correlation analysis was conducted to examine relationships between marginally significant variables ($0.05 \leq p < 0.1$) and other attributes. Correlation coefficients provided insights into the direction and strength of associations among variables. Positive correlations indicated that two variables tended to increase together, while negative correlations suggested inverse relationships. Variables exhibiting very weak correlations (close to zero) and high p-values were considered redundant and excluded, whereas weak to moderately correlated variables with statistically significant p-values were retained for further analysis, as they still provided meaningful explanatory information.

Finally, the Likelihood Ratio Test (LRT) was employed to guide model selection. The LRT compares the fit of the current model with a simpler model excluding the candidate variable, providing a statistical basis for determining whether the inclusion of a variable significantly improves model performance. By combining stepwise forward selection, p-value thresholds, correlation analysis, and LRT, this estimation strategy ensures a systematic, rigorous, and parsimonious approach. The resulting model captures the most relevant attributes influencing passenger distribution on train platforms, offering a robust foundation for understanding and predicting platform choice behavior.

3.9.6 Evaluation Criteria and Model Validation

A rigorous evaluation and validation process was implemented to ensure that the developed model not only fits the observed data but also generalizes well to unseen conditions. This section outlines the methodological framework used to assess the model's performance, including data splitting, probabilistic evaluation, spatial validation, and generalization testing. The overall goal of this process is to verify the model's predictive reliability and behavioral realism across different datasets (Ben, Akiva & Lerman, 1985, Train, 2009)[13][18].

The validation process was designed to evaluate the model's performance at two complementary levels:

1. Individual validation, assessing how well the model reproduces the empirical probabilities and spatial patterns within each dataset (training and testing separately),
2. Comparative validation, examining the consistency and stability of results between the two datasets to identify potential signs of overfitting or generalization errors.

This dual perspective allows both absolute and relative evaluation of the model's predictive quality.

3.9.6.1. Data Splitting and Validation Framework

To evaluate the model's generalization ability, the available dataset was divided into two independent subsets: a training set and a testing set. The training set was used to estimate the model parameters, while the testing set was reserved for post-estimation validation.

This approach follows standard model validation procedures in data-driven and discrete choice modeling to avoid overfitting and to test predictive robustness (Arlot & Celisse, 2010, Hastie, Tibshirani, & Friedman, 2009)[55][54]. A random split was employed to ensure statistical representativeness, with approximately 70% of the data used for model training and 30% for testing. Both subsets maintained similar distributions across the main explanatory variables (e.g., spatial coordinates, probabilities, and contextual attributes) to guarantee fair comparison.

After estimation, model outputs from both datasets were used to compute a set of comparable evaluation metrics, enabling a direct quantitative assessment of model performance consistency between training and testing phases.

3.9.6.2. Probabilistic Evaluation

The probabilistic evaluation focuses on how accurately the model reproduces the observed empirical probability distribution of choices.

Each predicted probability represents the likelihood that a given alternative (e.g., spatial cell or platform section) is chosen, and these predictions are compared with the empirical probabilities computed from observed frequencies.

The following statistical indicators were used to measure the model's predictive accuracy and ranking behavior:

- Mean Absolute Error (MAE): Quantifies the average magnitude of errors between predicted and observed probabilities, independent of direction. It provides a clear measure of overall predictive deviation (Chai & Draxler, 2014)[56].
- Root Mean Squared Error (RMSE): Similar to MAE but penalizes larger deviations more heavily, making it sensitive to high-magnitude prediction errors.
- Pearson Correlation Coefficient (r): Measures the linear association between predicted and empirical probabilities, indicating how well the model reproduces the overall probability distribution (Ben,Akiva & Bierlaire, 1999, Train, 2009)[57][18].
- Spearman's Rank Correlation (ρ) and Kendall's Tau (τ): Assess the degree of correspondence between the predicted ranking of probabilities and the empirical ranking. High rank correlations indicate that the model correctly reproduces the ordering of preferences or likelihoods across spatial alternatives (Kendall, 1938, Spearman, 1904)[58][59].

These probabilistic indicators were computed separately for the training and testing datasets. The comparison between both provides insights into how well the model generalizes probabilistic accuracy beyond the data it was calibrated on.

3.9.6.3. Spatial Validation

Beyond statistical accuracy, spatial consistency is crucial in transport and behavior modeling, where the physical dimension of choices, such as spatial location, is fundamental (Horni et al., 2016, Páez & Scott, 2005)[20][8]. Therefore, a spatial validation procedure was applied to assess how closely the model's predicted spatial distribution aligns with the observed empirical distribution.

This process involved computing spatial error and ranking measures as follows:

- Euclidean Distance: The geometric distance between predicted and observed coordinates for each spatial cell or location, capturing the spatial displacement between model predictions and real-world observations.
- Mean and Median Distance: Summarize the central tendency of spatial deviations, providing interpretable indicators of overall spatial accuracy.
- Mean Rank Deviation: Measures how much the predicted rank of each spatial cell differs from its observed empirical rank, offering insight into the accuracy of spatial preference ordering.
- Weighted Distance: A probability-weighted version of the Euclidean distance that gives higher importance to frequently chosen or more relevant locations.

These spatial metrics, particularly the weighted distance and rank deviation, jointly capture not only geometric proximity but also behavioral relevance, ensuring that the most probable locations are predicted with higher spatial precision. This approach reflects best practices in spatial choice and accessibility modeling (Páez & Scott, 2005, Horni et al., 2016)[8][20].

3.9.6.4. Top-N Hit Rate and Cumulative Coverage

To further assess the model's spatial ranking performance, a Top-N Hit Rate analysis was conducted. This metric evaluates how many of the top-N predicted locations (those with the highest predicted probabilities) match the top-N empirically observed locations. For example, a Top,10 Hit Rate measures the percentage of overlap between the ten most likely predicted and observed locations. Additionally, the cumulative empirical probability coverage was calculated to quantify how much of the total observed probability mass is captured by the top predicted locations (Zhu et al., 2013, Choudhury et al., 2018)[60][61].

Together, these indicators provide an intuitive understanding of the model's ability to correctly identify the most probable locations and replicate the concentration of observed choices. This is particularly relevant for behavioral models that operate over a discrete spatial grid, where ranking and location precision are both essential (Horni et al., 2016)[20].

3.9.6.5. Comparative Validation Between Training and Testing Sets

All probabilistic and spatial metrics were computed for both datasets and then compared systematically. This comparison aimed to identify any discrepancies between the training and testing performance across all criteria. Such cross-validation helps determine whether the model's predictive behavior remains consistent when applied to new, unseen data.

The comparative evaluation included:

- Statistical accuracy (MAE, RMSE)
- Rank-based consistency (Spearman's ρ, Kendall's τ, Pearson's r)
- Spatial proximity (Mean Distance, Weighted Distance)
- Behavioral ranking (Top-N Hit Rate and Cumulative Coverage)

By comparing these metrics between the training and testing datasets, the analysis identifies whether the model maintains consistent predictive patterns across different data subsets. Small deviations across metrics (e.g., less than 5% in MAE or less than 10% in Top-N hit rates) indicate strong generalization and low overfitting, whereas larger differences may highlight specific behavioral or spatial biases requiring refinement.

This comparison supports the individual evaluations and checks whether the model is stable and works across datasets. (Arlot & Celisse, 2010, Hastie et al., 2009)[55][54].

4. Case Study: Eindhoven Train Station

This chapter aims to address research sub-questions related to how platform-specific factors influence passenger distribution. It focuses on investigating the dynamics of passenger choices on train platforms, with Eindhoven Station serving as the primary study site.

4.1. Case Study Area Description

The case study for this research is Eindhoven Central, focusing on waiting-location behavior at platform 1-2. Eindhoven Central is the main railway station in Eindhoven, North Brabant (Netherlands), serving as a major interchange for intercity and regional train services, which makes it a suitable context for analyzing real-world passenger distribution under varying conditions (Eindhoven Centraal, n.d., Ruland Architecten, n.d.) [62][63].

Figure 7_4.1.1 shows the main entrance for Eindhoven Central railway station

Figure 7_4.1.1 main entrance for Eindhoven Central railway station

4.1.1 Station Classification and Role

As one of the busiest Dutch railway stations outside the Randstad, Eindhoven Central handles a high volume of daily passengers, recent data place the station's daily passenger count at approximately 65,500 travelers (ExpatINFO Holland, n.d.) [64]. This high demand situates Eindhoven among the top ten busiest rail hubs in the Netherlands, ensuring that platform 1-2 experiences substantial passenger flows and diverse crowding conditions (ExpatINFO Holland, n.d., Eindhoven Centraal, n.d.) [64][62].

4.1.2 Platform Environment

Platform 1-2 at Eindhoven Central exhibits typical characteristics of a major intercity station platform: it is covered by a roof structure, and features platform sitting areas and a kiosk. These platform buildings, together with the sheltered roof and covered waiting facilities, provide passengers with amenities and sheltered spaces to wait, features that are relevant when studying waiting-location choice under varying conditions of weather and lighting. The renovation works carried out between 2013 and 2016 expanded station capacity and updated station passageways, enabling the station to accommodate growing passenger demand and improve accessibility to platforms through widened tunnels, escalators, and lifts (Heijmans, 2019) [65].

These physical characteristics, roof cover, waiting-room availability, shelter, and accessible platform layout shape the waiting environment and offer different waiting-location options (e.g., near shelters, along open platform edges, close to access points), which makes platform 1-2 especially suited for analyzing distribution patterns.

Figure 8_4.1.2.1 shows Eindhoven station

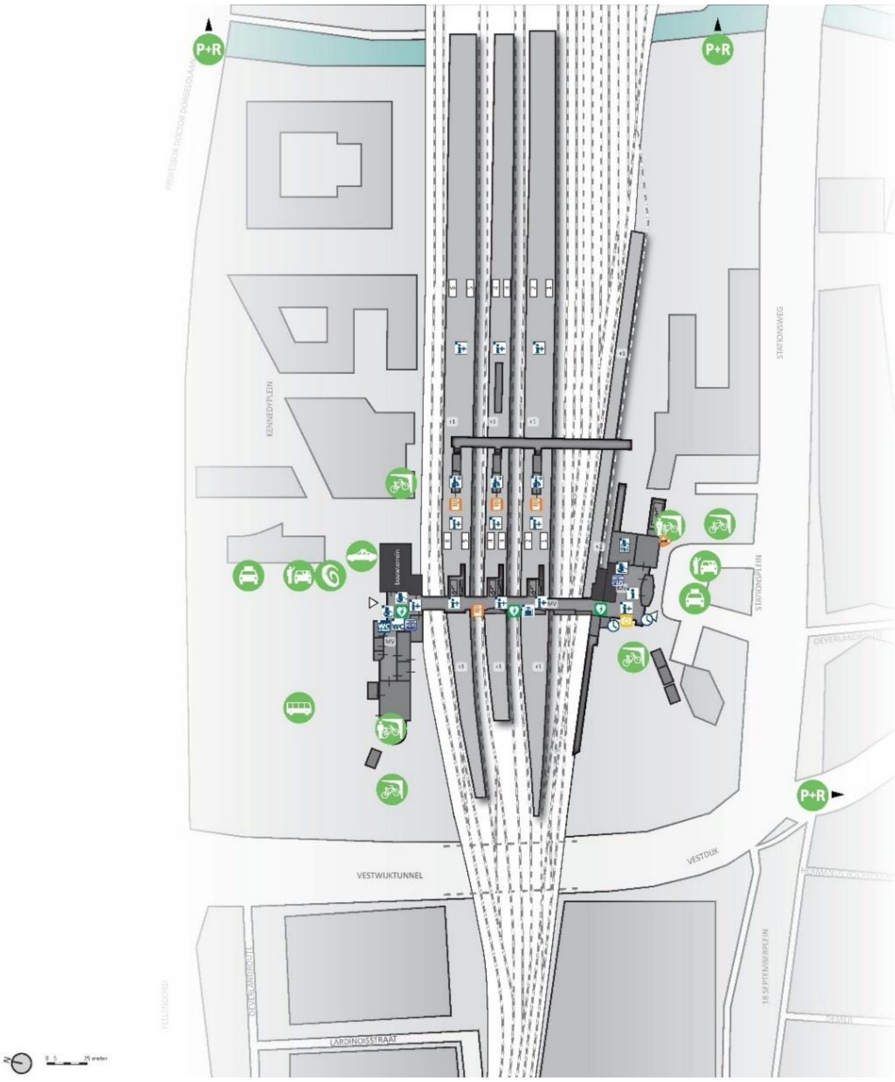

Figure 8_4.1.2.1 The Eindhoven station map

4.1.3 Passenger Demand and Travel Flows

With an estimated 65,500 daily passengers, Eindhoven Central sees substantial annual throughput, confirming that the station routinely operates under high-demand conditions (ExpatINFO Holland, n.d.) [64]. Given the station's role as an interchange with both intercity and regional services (Eindhoven Centraal, n.d.) [62], the variety of services and frequent train schedules imply that platform 1-2 supports a mix of boarding and alighting events throughout the day.

4.1.4 Types of Travelers Using the Station and Platform

Eindhoven Central caters to a heterogeneous mix of travelers: daily commuters using intercity and local trains, regional passengers connecting to smaller towns, transfer passengers changing trains, and occasional long-distance or intercity travelers. This diversity of user types leads to a range of waiting behaviors, from short-term waiting (commuters catching sprinters) to longer waiting times for intercity passengers or connecting travelers. The repurposed platform buildings (waiting rooms, kiosks) and the platform's shelter and layout make it likely that some passengers choose waiting-locations based on comfort, shelter, and convenience, while others may prefer quick boarding proximity, depending on their time constraints and travel purpose (Ruland Architecten, n.d., Heijmans, 2019) [63][65].

4.1.5 Suitability of Eindhoven Central

Because of its high passenger volumes, varied services, and a platform that combines sheltered waiting areas, physical layout, and multiple access points, platform 1-2 at Eindhoven Central offers a realistic and rich environment for studying waiting-location choice. The combination of structural amenities, environmental exposure (e.g., exposure to weather if parts of the platform are open), and diverse traveler types allows the analysis to capture a range of behaviors under varying conditions. Furthermore, the entire platform roof is equipped with sensors that cover a large part of the platform. Using this station as a case study helps ensure that findings are relevant to real operational contexts and may be generalizable to similar large stations in the Netherlands or Europe.

4.2 Descriptive statistics

The purpose of the descriptive statistics in this study is to establish a clear, empirical understanding of how passengers distribute themselves across the platform under different environmental conditions, and to identify whether spatial features, such as the leaning edges, seating areas, and information bords, appear to influence waiting-location choices. This descriptive step is essential because it provides the baseline evidence needed before developing and estimating the behavioral model introduced later in the thesis.

All results presented in this section are based on measurements of passenger density across the platform, recorded two minutes before train arrival. This timestamp is significant because it reflects a moment when most passengers have already selected and settled into their preferred waiting location. Therefore, the data represents actual choice behavior rather than transitional movement.

Distances to platform features were originally measured in centimeters and were converted to meters to improve interpretability. Table 3_4.2.1 summarizes the descriptive statistics for the three distance attributes used in the analysis.

Table 3_4.2.1 Descriptive Statistics of Distance Attributes (in meters)

Variable	***Mean***	***Std. Dev.***	***Min***	***Max***
Distance to leaning edge (m)	6.85	6.16	0.10	24.97
Distance to seats (m)	26.71	19.56	0.01	72.98
Distance to information boards (m)	29.18	20.39	0.02	80.14

The histogram of distances between observation points and leaning edges (Figure 9_4.2.1) shows a strong concentration of observations close to the leaning edge, with a long right tail extending beyond 20 meters. This suggests that many high-density areas are located near the track edge, possibly because passengers tend to position themselves closer to expected train door locations or prefer visual contact with the arriving train.

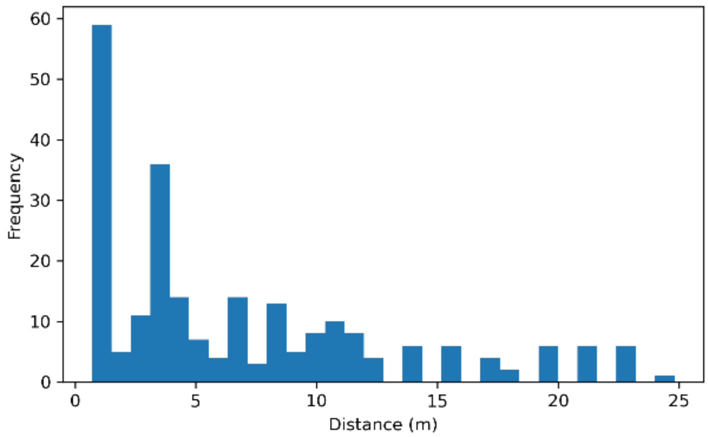

Figure 9_4.2.1 Distribution of Distances to Leaning Edges

***Figure 9_4.2.1:** Histogram showing the distances between observation points and leaning edges on the platform. Most passengers tend to stay close to the leaning edge, with a long tail indicating some passengers stand farther than 20 meters from the edge.*

Distances to seating show a very different pattern (Figure 10_4.2.2). The multimodal distribution reflects the discrete placement of benches along the platform: clusters of observations fall close to benches, while many others remain far away from any seating area. The relatively high mean distance (26.71 m) and large standard deviation indicate that seating is sparse and unevenly distributed.

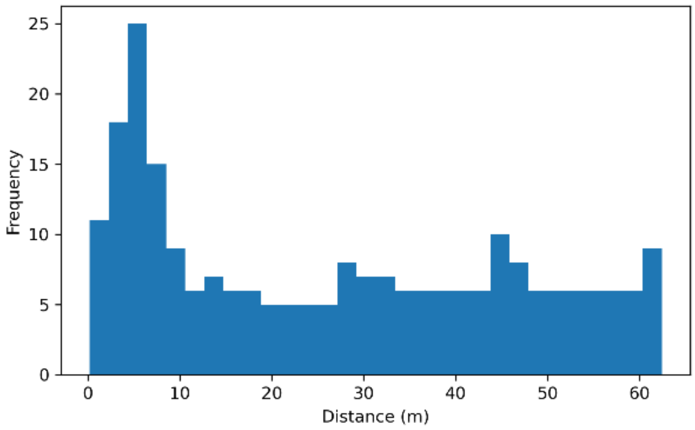

Figure 10_4.2.2 Distribution of Distances to Seating Areas

***Figure 10_4.2.2:* Histogram illustrating the distances between observation points and seating. The multimodal pattern reflects clusters near benches and larger gaps elsewhere, indicating unevenly spaced seating along the platform (mean distance: 26.71 m).**

A similar pattern is observed for signage (Figure 11_4.2.3). Information boards are typically placed at fixed, widely spaced intervals, which produces large gaps between signs. The average distance of 29.18 m reflects this sparse distribution. The tail of the distribution shows that some platform segments are more than 70 meters from the nearest sign.

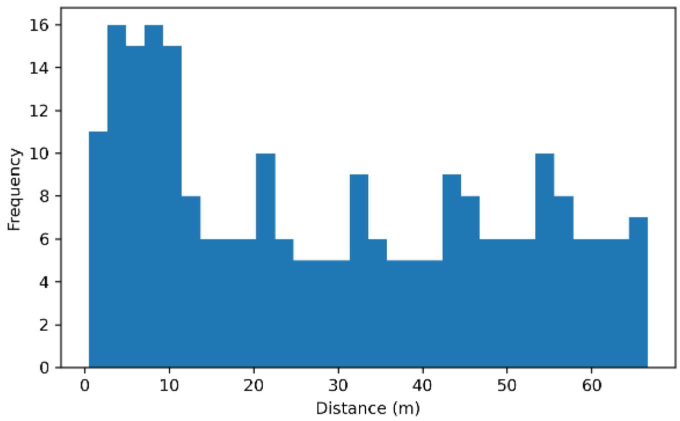

Figure 11_4.2.3 Distribution of Distances to Signage

***Figure 11_4.2.3:* Histogram of distances from observation points to signage on the platform. Signs are sparsely and regularly placed, resulting in an average distance of 29.18 m and long tails extending beyond 70 m in some segments.**

In order to explore how environmental conditions shape waiting location choices, passenger density Figure 12_4.2.4(a, b, c) is visualized using heat maps generated for three distinct environmental situations: (a)Good weather & daylight, (b)Bad weather & daylight and (c)Bad weather & artificial lighting

Weather conditions were operationalized quantitatively using hourly meteorological data, specifically air temperature and rainfall. Periods were classified as good weather when temperatures were within a comfortable range and no rainfall was recorded, representing conditions under which thermal discomfort and shelter-seeking behavior are minimal. Conversely, bad weather was defined as periods with low temperatures and/or measurable rainfall, reflecting conditions associated with increased discomfort and stronger incentives to seek shelter or safety-related platform features.

Lighting conditions were classified based on sunrise and sunset times obtained from meteorological records. Observations conducted during natural daylight hours were labelled as daylight, while observations after sunset and before sunrise were classified as artificial lighting, corresponding to periods when platform illumination is provided primarily by installed lighting systems.

The combination of weather and lighting theoretically yields four possible states. However, only three were retained for analysis for both empirical and behavioral reasons.

First, the combination of good weather & artificial lighting rarely occurs in practice within the study context. Good weather conditions are predominantly observed during summer and early autumn months, when daylight extends into the evening peak hours. Consequently, periods with comfortable temperatures and rainfall-free conditions typically coincide with sufficient natural light. Instances of good weather combined with artificial lighting were extremely limited and did not provide enough observations to support reliable density estimation.

Second, the category bad weather & daylight captures transitional and peak-demand periods, particularly during daytime rain or cold conditions, when passenger volumes remain high and behavioral responses to discomfort are most pronounced.

Third, bad weather & artificial lighting represents the most challenging environmental setting, combining reduced visibility with the discomfort associated with heat and rainfall. This category is particularly relevant for studying safety perception, shelter-seeking behavior, and avoidance of exposed platform areas.

Together, these three categories capture the dominant and behaviorally meaningful environmental regimes observed during the four-month study period, while maintaining sufficient sample sizes for robust spatial analysis.

These categories allow the analysis to identify whether passengers systematically adjust their waiting locations by moving toward sheltered areas, clustering near seating zones, positioning themselves closer to information boards, or avoiding exposed locations such as platform edges under deteriorating environmental conditions. This distinction is essential because the core assumption of the research is that waiting location choice is systematically related to environmental exposure and spatial platform attributes.

The heat maps provide a visual pre-analysis of these relationships by highlighting high-density clusters around key spatial features under different environmental regimes. Rather than examining long-term temporal variations, the focus is deliberately placed on instantaneous density distributions, as these directly reflect passengers' final waiting decisions prior to boarding. Incorporating broader temporal patterns, such as daily or weekly demand cycles, would not contribute to this objective and would obscure the spatial mechanisms of interest.

The outcome of this descriptive analysis is therefore a platform-level behavioral portrait: where passengers tend to position themselves, how this positioning varies with weather and lighting, and whether observed patterns align with the spatial features hypothesized to influence waiting behavior.

These descriptive findings form the empirical foundation for the subsequent modelling stage, in which the identified relationships are formally quantified using the discrete choice framework. Figure 12_4.2.4 presents the heat maps used for this analysis.

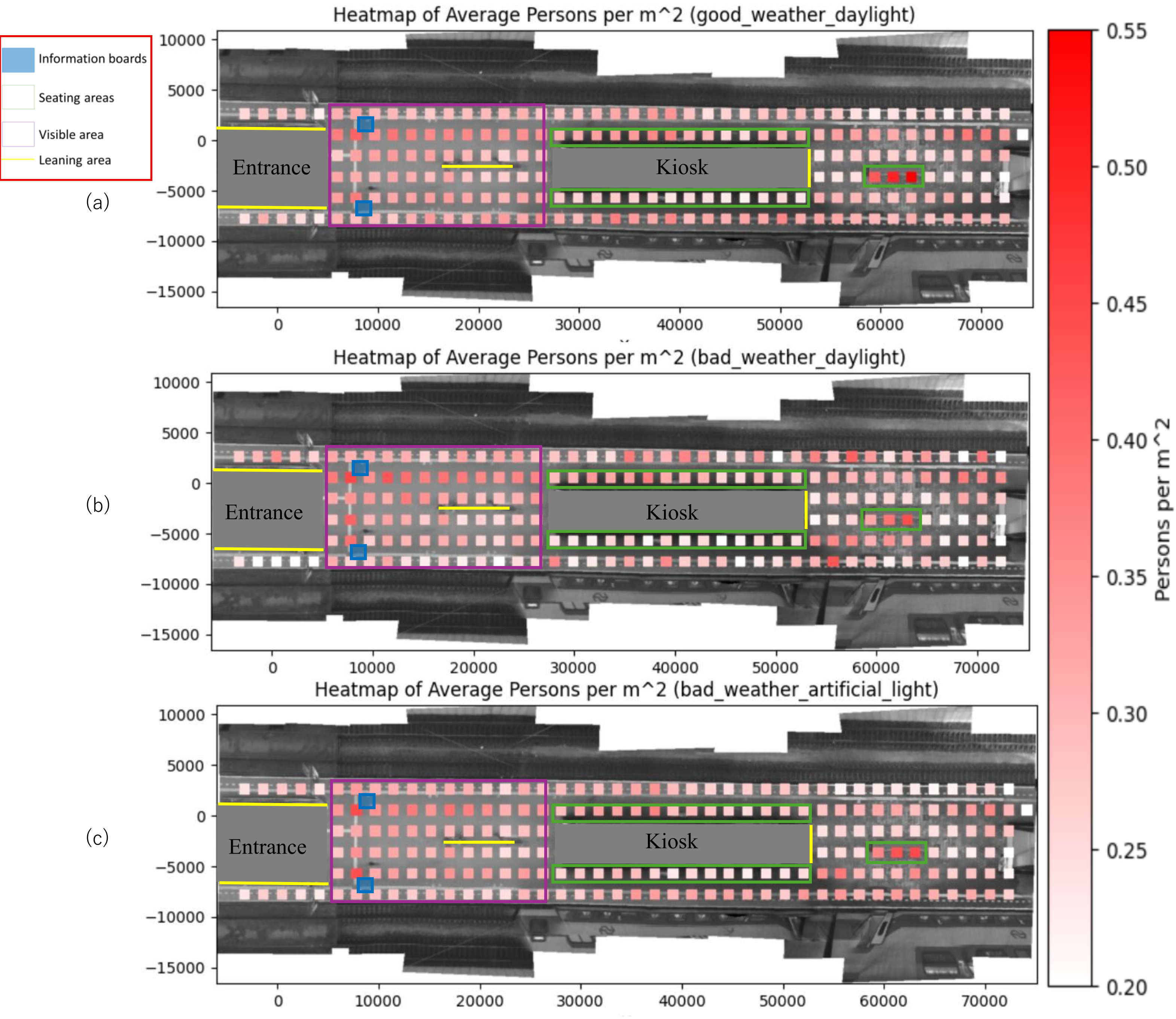

Figure 12_4.2.4 Passenger Density Heatmaps under Different Conditions

<u>Figure 12_4.2.4</u>: Heatmaps showing passenger distribution on the platform under different weather and lighting conditions. (a) Good weather, daylight, high densities spread across the central platform and toward track edges. (b) Bad weather, daylight, passengers cluster near entrances and seating, avoiding edges. (c) Bad weather, artificial light, overall densities are low, with most passengers near entrances and illuminated areas.

4.2.1. Pedestrian distribution on the platform

Figure 12_4.2.4 illustrates heatmaps of average passenger density per square meter under three environmental conditions: (a) good weather with daylight (top), (b) bad weather with daylight (middle), and (c) bad weather with artificial lighting (bottom). Each heatmap overlays passenger density on the same platform layout, which includes the entrance area, kiosk, seating areas, and information boards. The color intensity represents the concentration of passengers, with darker red tones indicating higher densities.

Across all three subfigures, a consistent pattern emerges: passengers tend to cluster around the entrance zone (highlighted in grey) and seating areas (green outline). This indicates that, regardless of weather or lighting, passengers prefer locations that offer accessibility and comfort. The entrance provides convenient access to trains and exits, while the seating areas allow passengers to wait more comfortably, especially during longer intervals between trains.

4.2.2 Impact of weather conditions

However, differences become noticeable when comparing weather conditions:

- Figure 12_4.2.4 (a) (good weather, daylight) shows the highest density level overall, with heat spreading across the central platform and extending closer to the track edges. This suggests that during pleasant conditions and higher passenger volumes, individuals disperse more widely across the platform, even towards less safe or exposed areas. The wider spread indicates that maintaining personal space becomes a stronger priority than safety when crowding increases.
- In contrast, Figure 12_4.2.4 (b) (bad weather, daylight) shows a more concentrated cluster near the entrance and central seating area, with fewer passengers near the edges. Poor weather appears to encourage passengers to remain in more sheltered or familiar zones, prioritizing comfort and safety over personal space.
- Figure 12_4.2.4 (c) (bad weather, artificial light) reveals the lowest overall densities. Passenger concentration is again strongest near the entrance and kiosk, while peripheral areas show minimal activity. This may reflect both the reduced number of passengers during off-peak or darker periods and a tendency to stay near accessible or illuminated sections of the platform.

Taken together, these three subfigures reveal a clear behavioral pattern: passengers consistently cluster near entrances and seating areas (comfort and accessibility), but their spatial spread towards the edges increases under good weather or crowded conditions (personal space). In contrast, under adverse weather or low visibility, safety and shelter become stronger influencing factors, keeping passengers closer to central and covered zones.

4.3 Correlation analysis of attributes

The purpose of this section is to identify whether any strong linear relationships exist between the explanatory variables, and to evaluate whether multicollinearity may pose a problem for the later Discrete Choice Modelling (DCM) stage. This step is essential because DCM assumes that each attribute contributes independently to the utility function. If attributes are strongly correlated, it becomes difficult to disentangle their individual effects. Therefore, the aim here is not to explain passenger behavior, but rather to diagnose the structure of the dataset, determine which variables may overlap, and confirm that DCM is an appropriate modelling approach.

Figure 13_4.3.1 presents the correlation matrix for all continuous variables included in the study. Overall, the correlations between variables are weak, indicating that no major multicollinearity issues are present. This is important because it suggests that the attributes can be included concurrently in the choice model without violating key assumptions.

Some expected patterns, however, do appear in the matrix. A strong positive correlation is observed between dist_to_kiosk_upper and dist_to_kiosk_below, and similarly between dist_to_kiosk_entrance and dist_to_kiosk_non_entrance. These correlations are a direct reflection of the physical layout of the platform, where the kiosk and its surrounding areas create spatial clusters. In other words, cells that are close to one kiosk-relevant location are also likely to be close to the others simply due to geometry. These relations do not pose a methodological issue, as long as only one representative variable from each spatial cluster is used in the model.

A modest negative correlation between lighting and persons_per_m2 indicates that denser areas tend to occur in zones with poorer lighting. However, the correlation is not strong enough to imply a direct causal mechanism, instead, it is likely reflecting indirect spatial effects, such as crowded conditions around sections of the platform that rely on artificial lighting during peak hours.

Another notable pattern is the positive correlation between the distance to information boards and the distance to platform entrance. This is consistent with the platform layout: information boards are positioned close to the main access points, meaning that moving away from one naturally increases distance from the other. Conversely, a weak negative correlation appears between the distance to information boards and distance to seating. This likely reflects behavioral tendencies in addition to spatial structure: passengers who prefer sitting may pay less attention to information boards, relying instead on visual cues such as approaching trains.

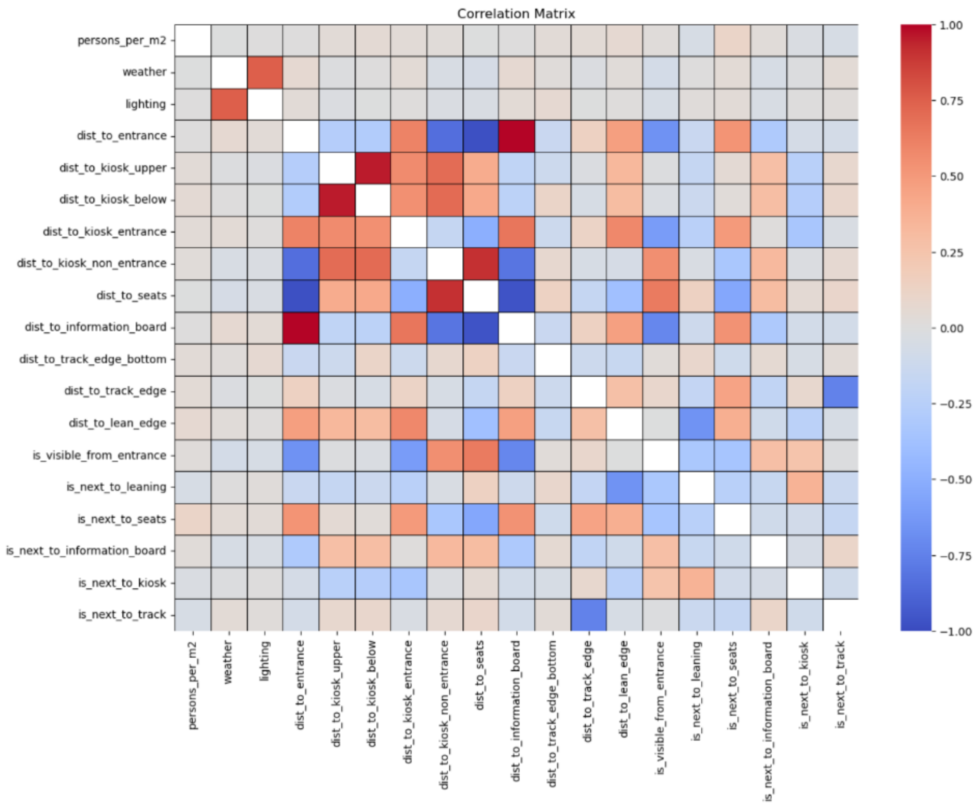

Figure 13_4.3.1 Correlation Matrix of Continuous Variables

***Figure 13_4.3.1:** Shows correlations between all continuous variables. Overall correlations are weak, indicating no major multicollinearity, with some expected spatial patterns between kiosk, entrances, and information boards.*

Figure 14_4.3.2 shows a histogram of the number of passengers per cell, revealing a highly right-skewed distribution. Most cells contain only one passenger, while only a few cells contain three or more. This pattern confirms that the platform is generally sparsely occupied, with small, localized areas of higher density. Such a distribution reinforces the need for analytical methods capable of handling diverse behavioral data.

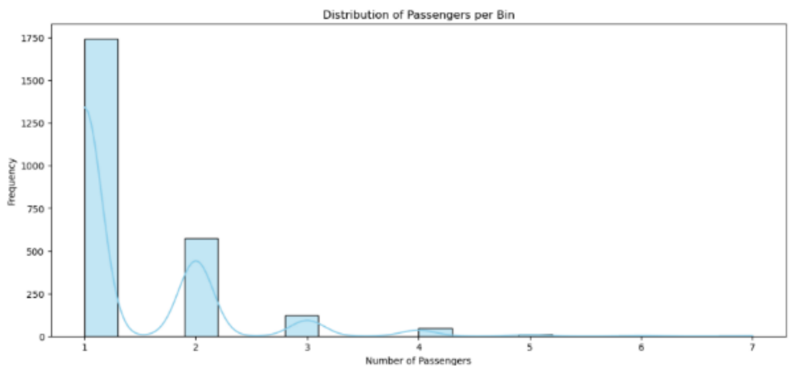

Figure 14_4.3.2 Number of passengers per cell

In summary, the correlation analysis shows that the dataset contains mostly independent variables with only layout-related correlations. These findings support the use of a DCM, which is particularly suited to modelling situations where multiple attributes influence behavior simultaneously and non-linearly. The absence of strong correlations ensures that the model can more clearly disentangle the effects of each attribute on waiting-location choice.

5 Choice Models

To model passenger platform cell selection behavior, this study employs a Multinomial Logit (MNL) framework, which assumes that each passenger chooses the cell that provides the highest perceived utility among all available alternatives. Each cell on the platform represents a discrete alternative, characterized by a unique combination of spatial and environmental attributes. The decision-making process is formalized through a utility function that quantifies the satisfaction a passenger derives from selecting a specific cell.

Interaction terms, such as the combination of weather, lighting, and proximity to seating, were also included to capture the combined effects of environmental and spatial conditions on passenger preferences. These interactions provide a more realistic representation of behavior, acknowledging that passengers' choices are shaped not only by single factors but also by their contextual interplay.

The model was estimated using 70% of the available dataset, ensuring sufficient data to accurately capture the relationships between variables and passenger choices. The remaining 30% of the data was reserved for model validation, allowing assessment of predictive performance and robustness on unseen observations. This data split ensures that the estimated coefficients are not overly tuned to the training data (i.e., avoiding overfitting) while preserving the model's generalizability to real-world conditions.

Before estimation, an availability dictionary (av) was constructed to represent which platform cells were accessible for passenger selection. Each cell (from 1 to 264) was iteratively checked for its availability status, where a value of 1 indicated an available cell and 0 denoted an unavailable one (e.g., due to structural obstructions or physical constraints). Only available cells were included in the estimation process, ensuring behavioral realism in the observed choice set.

After estimation, many variables were not significant in explaining passenger waiting location choice. Eliminating these insignificant factors led to a reduced utility function that captures the most influential determinants of passenger behavior:

$$
\begin{aligned}
U_i = {} & \beta_{\text{persons\_per\_m2}} \cdot \text{persons\_per\_m2}_i + \beta_{\text{dist\_to\_entrance}} \cdot \text{dist\_to\_entrance}_i + \beta_{\text{dist\_to\_information\_board\_upper}} \\
& \cdot \text{dist\_to\_information\_board\_upper}_i + \beta_{\text{dist\_to\_track\_edge\_upper}} \\
& \cdot \text{dist\_to\_track\_edge\_upper}_i + \beta_{\text{is\_next\_to\_leaning}} \cdot \text{is\_next\_to\_leaning}_i + \beta_{\text{is\_next\_to\_seats}} \\
& \cdot \text{is\_next\_to\_seats}_i + \beta_{\text{is\_next\_to\_kiosk}} \cdot \text{is\_next\_to\_kiosk}_i + \beta_{\text{is\_next\_to\_track}} \\
& \cdot \text{is\_next\_to\_track}_i + \beta_{\text{interaction\_weather\_lighting\_dist\_to\_track\_edge}} \cdot (\text{dist\_to\_track\_edge}_i \\
& \cdot \text{weather}_i \cdot \text{lighting}_i) + \varepsilon_i
\end{aligned}
$$

The MNL model quantifies how the identified attributes influence passengers' spatial choices. Each estimated coefficient (β_k) reflects the marginal utility contribution of its associated factor, showing whether it increases or decreases the probability of a passenger choosing a given cell. For instance, a negative coefficient for distance to entrance implies a preference for cells closer to the entrance, whereas a positive coefficient for lighting level indicates a tendency to select well-lit areas.

This modeling approach enables a detailed understanding of passenger behavior at the micro-level of platform space. It provides a foundation for predicting how passengers redistribute under varying environmental and infrastructural conditions, thereby supporting design and policy decisions aimed at improving platform efficiency, comfort, and safety.

Table 4_5.1 presents the estimated coefficients and associated statistical measures for the model.

Table 4_5.1 Estimated Model Coefficients

Log,Likelihood : -9005.751

Variable	***Estimated Beta***	***Robust Std. Error***	***Robust t,test***	***Robust p,value***
β _dist_to_entrance	0.061	0.023	2.635	0.008
β _dist_to_information_board_upper	-0.098	0.012	-8.325	0.000
β _dist_to_track_edge_upper	-0.032	0.011	-2.850	0.004
β _interaction_weather_lighting_dist_to_track_edge	0.026	0.012	2.132	0.033
β _is_next_to_kiosk	-0.430	0.132	-3.252	0.001
β _is_next_to_leaning	0.200	0.061	3.299	0.001
β _is_next_to_seats	1.222	0.088	13.900	0.000
β _is_next_to_track	-1.020	0.074	-13.708	0.000
β _persons_per_m2	0.329	0.092	3.592	0.000

5.1 Model Parameter

The final model retained nine statistically significant attributes that meaningfully influence passenger cell selection decisions on the platform: distance to entrance, distance to information boards, distance to track edge, proximity to seats, leaning areas, kiosks, proximity to the track, passenger density, and the interaction of weather, lighting, and distance to the track edge (see Table 2). All these variables had p-values below 0.05, providing strong statistical evidence of their influence on passenger positioning behavior.

The model achieved a log-likelihood of –9005.75, indicating a robust fit for this behavioral model. The log-likelihood quantifies how well the model predicts the observed passenger choices: values closer to zero (less negative) indicate better predictive accuracy. Considering the behavioral complexity of passengers on busy platforms, this strong log-likelihood confirms that the model effectively captures the key determinants of location selection.

The behavioral interpretation of the significant attributes, in relation to existing literature, is as follows:

- Distance to entrance (positive effect): Passengers prefer cells closer to the entrance for convenience, minimizing walking distance and enabling rapid access. This aligns with studies by Bosina et al. (2015) [47], which highlight the importance of accessible entry points in reducing congestion and improving platform efficiency.
- Distance to information boards (negative effect): Passengers tend to avoid areas immediately adjacent to information boards, likely due to congestion or obstructed movement. This is consistent with Ahn et al. (2017) [25] and Xu et al. (2020)[14], who indicate that high-traffic zones near infrastructure can limit comfort and mobility.
- Distance to track edge (negative effect): Safety considerations drive passengers to maintain distance from the track edge. Similar findings are reported by Seriani et al. (2016)[46] and Bosina et al. (2015)[47], highlighting edge avoidance as a common risk mitigation behavior.
- Proximity to seats (strong positive effect): The largest coefficient indicates a strong preference for seating, reflecting comfort and ergonomic needs. Bandara & Hewawasam (2020)[30] and Mizuno & Tokuda (2023)[31] emphasize the critical role of seating in influencing passenger micro-location choices.
- Proximity to leaning areas (positive effect): Leaning areas attract passengers, suggesting that opportunities for rest and support guide positioning decisions, as noted by Kuipers et al. (2021)[52].
- Proximity to kiosks (negative effect): Avoidance of kiosks likely reflects congestion or obstruction, consistent with Zhang et al. (2020) [48] and Ahn et al. (2017)[25], who link obstacles and high-traffic areas with reduced accessibility.
- Proximity to track (negative effect): Confirms that passengers strategically avoid risky areas, prioritizing safety while waiting, aligning with Seriani et al. (2016)[46] and Bosina et al. (2015)[47].
- Passenger density (positive effect): Passengers appear to prefer staying near others, reflecting social behavior, safety-in-numbers perception, or psychological comfort in moderately crowded areas. This is consistent with the literature by Zhang et al. (2020)[48] and Kim et al. (2018)[50], which indicates that human factors and perceived crowd safety strongly influence platform location choices.
- Environmental interaction (weather × lighting × track edge distance, positive effect): Favorable weather and lighting conditions encourage passengers to position themselves closer to the track edge. This suggests that comfortable environmental conditions reduce perceived safety risks, allowing passengers to spread more evenly and occupy desirable areas near the track. These findings are supported by Xu et al. (2020)[14], emphasizing that environmental comfort significantly shapes platform behavior.

Overall, the significant coefficients validate insights from literature. The strong statistical significance of the attributes, together with the robust log-likelihood, confirms the model's effectiveness in capturing the complex behavioral patterns of platform users.

5.2 Model Validation

Model validation was conducted to assess the robustness, predictive power, and behavioral realism of the developed spatial choice model.

The goal was to determine whether the model accurately reproduces observed patterns of choice in both the training (in-sample) and testing (out-of-sample) datasets. Validation was performed using complementary probabilistic, spatial, and rank-based indicators.

5.2.1 Data Split and Validation Procedure

The complete dataset was randomly divided into 70% for model estimation (training) and 30% for validation (testing). The estimated model parameters (β coefficients) from the training data were then applied to both subsets to compute predicted probabilities of spatial choices.

The validation was designed to evaluate three main aspects:

- Probabilistic accuracy: how closely predicted probabilities match the empirical (observed) ones.
- Spatial accuracy: how close predicted spatial locations are to the observed ones.
- Behavioral consistency: how stable ranking and preference patterns remain across both datasets.

This multi-criteria validation ensures that the model is not only statistically accurate but also spatially and behaviorally consistent, avoiding overfitting and ensuring realistic representation of travel and choice behavior.

5.2.2 Probabilistic Validation

The first part of the validation focused on the alignment between predicted and empirical probabilities.

Table 5_5.2.2.1 presents the main probabilistic indicators for both training and testing datasets.

Table 5_5.2.2.1 Probabilistic Validation Metrics

Metric	Training Set	Testing Set
Mean Predicted Probability	0.000454	0.001052
Mean Empirical Probability	0.004739	0.005102
Mean Absolute Error (MAE)	0.004286	0.004059
Mean Squared Error (MSE)	0.000033	0.000033
Correlation (Predicted vs Empirical)	0.797	0.706

The model's predictive performance was first evaluated using the training dataset, where it demonstrated a strong ability to replicate observed passenger behavior. Predicted and empirical choice probabilities are closely aligned, with a correlation coefficient of 0.80, indicating that the model effectively captures the underlying structure of passengers' decision-making. The very low error metrics, an MAE of 0.0043 and an MSE of 0.000033, confirm that deviations between predicted probabilities and observed choices are minimal. These results collectively show that the model is not only statistically robust but also internally coherent. In practical terms, the estimated utility structure successfully reflects the actual distribution of waiting location choices on the platform, giving confidence that the model is correctly specified.

When applied to the testing dataset, the model continues to perform well. The correlation between predicted and observed choice probabilities remains high at 0.71, showing that the relationships captured during estimation persist with unseen data. Additionally, MAE and MSE values in the testing phase remain almost identical to those in the training set, indicating that prediction errors do not increase outside the estimation sample. This consistency provides strong evidence that the model is not overfitted and maintains high probabilistic reliability. Overall, these results demonstrate that the model possesses both internal and external validity, confirming its effectiveness for predicting passenger waiting location choices across varying environmental conditions and platform characteristics.

5.2.3 Spatial and Rank-Based Validation

Spatial validation was used to evaluate how closely the model predicts the physical location of empirical choices.

For each matched predicted–observed pair, the Euclidean distance was computed based on the x–y coordinates of cells.

In addition, rank-based indicators (Spearman's ρ and Kendall's τ) were used to assess the consistency of the preference structure between predicted and empirical probabilities.

Table 6_5.2.3.1 shows the Spatial and Rank-Based Validation Metrics

Table 6_5.2.3.1 Spatial and Rank-Based Validation Metrics

METRIC	TRAINING SET	TESTING SET	DIFFERENCE (TEST – TRAIN)
MEAN DISTANCE (M)	25.3396	26.1299	+0.7903
MEDIAN DISTANCE (M)	21.2037	21.6539	+0.4501
MAE (PROB)	0.0043	0.0040	−0.0002
RMSE (PROB)	0.0057	0.0056	−0.0001
SPEARMAN'S Ρ	0.9975	0.9836	−0.0139
KENDALL'S Τ	0.9698	0.9208	−0.0491
MEAN RANK DEVIATION	3.3507	7.6582	+4.3075
WEIGHTED DISTANCE	0.1089	0.1215	+0.0127

In the training dataset, the model demonstrates a high level of spatial precision. The mean distance between predicted and observed choices is 25.3 meters, and the median distance is 21.2 meters, indicating that the predicted locations align closely with where passengers actually chose to stand. The strength of this alignment is further reflected in the rank correlations: Spearman's ρ reaches

0.998 and Kendall's τ reaches 0.97, both of which signal near-perfect agreement between the predicted ordering of preferred locations and the empirical ordering derived from the data. Together, these measures confirm that within the estimation sample, the model not only replicates the spatial pattern of passenger behavior but also captures the underlying preference structure with remarkable accuracy.

When the model is evaluated on the testing dataset, its performance remains highly stable. The mean (26.1 m) and median (21.6 m) distances between predicted and observed choices are almost identical to those observed in training, demonstrating that the model retains spatial consistency even when confronted with new, unseen data. Although the rank correlations decrease slightly, Spearman's ρ to 0.984 and Kendall's τ to 0.921, they remain exceptionally strong, indicating that the model continues to reproduce the behavioral ordering of platform locations with high accuracy. The increase in mean rank deviation, from 3.35 in training to 7.66 in testing, reflects the natural variability expected when predictions are made outside the estimation sample. Nevertheless, the overall ranking structure remains robust.

Taken together, these results show that the model is both spatially precise and behaviorally reliable across datasets. It captures not only where passengers prefer to position themselves on the platform but also the relative structure of these preferences, and it maintains this performance when applied to new observations, demonstrating strong generalization ability and confirming the validity of the model's behavioral representation.

5.2.4 Top-N Predictive Performance

To further assess predictive realism, a Top-N hit rate analysis was conducted. This evaluates how frequently the empirically chosen locations fall within the top N highest-probability predictions, as well as how much cumulative empirical probability those top predictions capture (coverage). Table 7_5.2.4.1 shows the Top-N Hit Rates and Cumulative Coverage

Table 7_5.2.4.1 Top-N Hit Rates and Cumulative Coverage

Top-N	***Hit Rate (Train)***	***Coverage (Train)***	***Hit Rate (Test)***	***Coverage (Test)***	***Δ Hit Rate***	***Δ Coverage***
5	80.0%	8.98%	60.0%	8.80%	−20.0%	−0.18%
10	50.0%	13.44%	50.0%	14.67%	0.00%	+1.23%
20	55.0%	23.33%	45.0%	24.13%	−10.0%	+0.81%
50	78.0%	48.83%	68.0%	49.73%	−10.0%	+0.91%

The model's predictive performance was further assessed using the Top-N hit rate and cumulative coverage metrics, which evaluate its ability to identify the most probable locations chosen by passengers. In the training dataset, the model demonstrates a Top,5 hit rate of 80%, indicating that in four out of five cases, the empirically observed cell falls among the five highest predicted alternatives. This high hit rate illustrates that the model reliably captures the key areas where passengers tend to wait. As the number of alternatives considered increases, cumulative coverage rises steadily, reaching nearly 49% at the Top,50 level. This demonstrates that the model successfully identifies a majority of high-probability spatial regions, highlighting its effectiveness in mapping passenger behavior across the platform.

When applied to the testing dataset, the Top,5 hit rate decreases moderately to 60%, which is expected when evaluating unseen data. Despite this reduction, the cumulative coverage remains stable or even slightly improves, showing that the model continues to capture the correct spatial zones of passenger activity. This stability suggests that, while the exact ranking of specific cells may vary, the model consistently identifies the areas of highest passenger presence. Overall, these results confirm that the model generalizes well beyond the estimation sample, providing a robust and realistic representation of passenger waiting behavior on the platform.

5.2.5 Spatial Visualization of Predicted and Empirical Probabilities

To complement the quantitative validation measures presented earlier, the figure 15_5.2.5.1 below presents four spatial heat maps, two for the training dataset and two for the testing dataset. These visualizations provide an essential qualitative assessment of the model's behavior by illustrating how predicted and empirical probabilities are distributed across the platform space. While numerical indicators summarize overall accuracy, spatial heat maps reveal where the model performs well and whether it reproduces the behavioral logic implied by the environment.

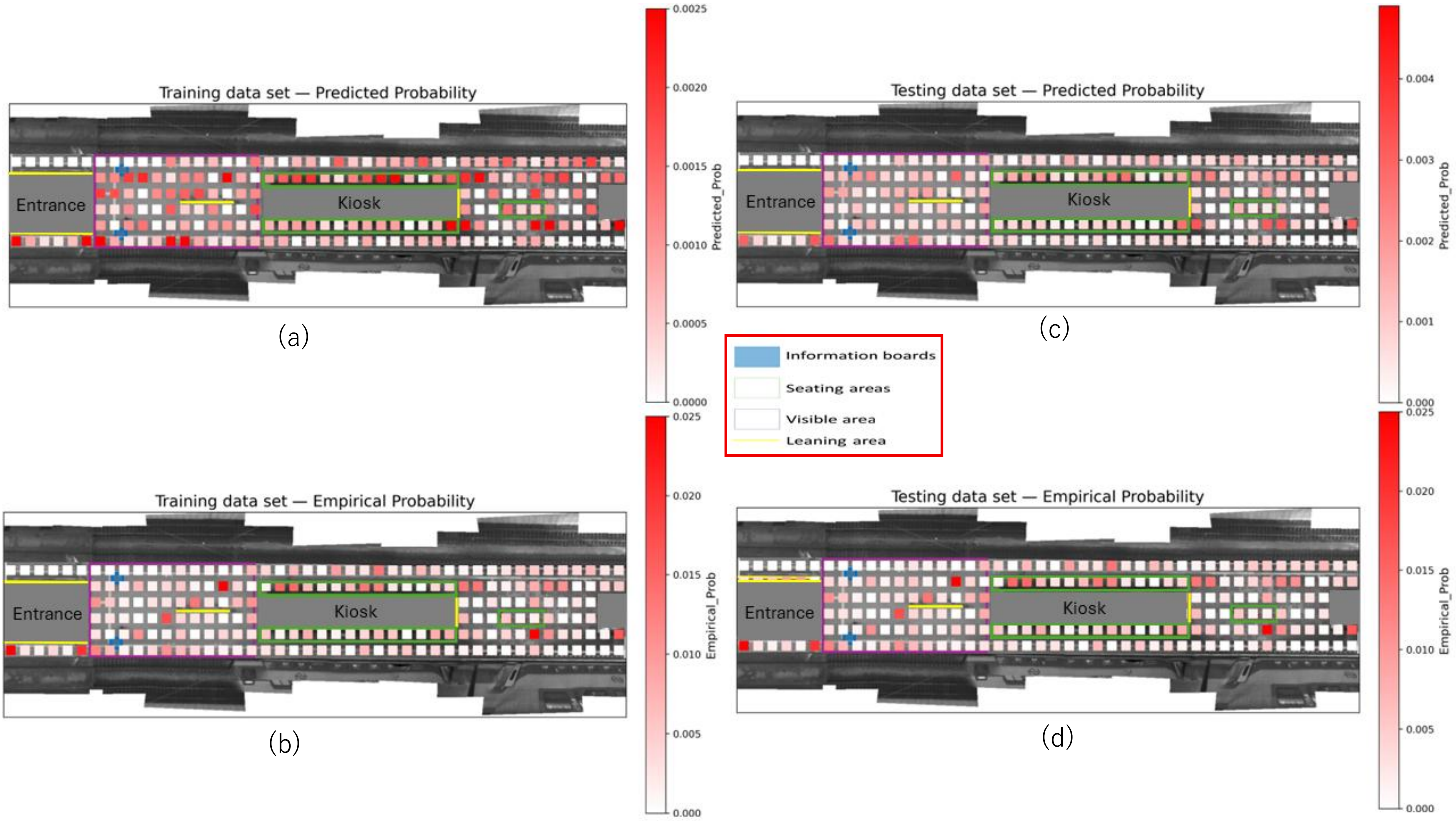

Figure 15_5.2.5.1 Predicted and Empirical Probabilities spatial heat maps

***Figure 15_5.2.5.1:** Shows predicted and observed passenger probabilities for training (a, b) and testing (c, d) datasets. Patterns highlight key platform areas such as entrances, leaning zones, and seating, demonstrating that the model captures spatial behavior consistently across datasets.*

The top-left map (a) in the figure shows the predicted probability distribution for the training dataset. This map reflects how the model allocates choice likelihoods across space when applied to the data used for estimation. Higher predicted values cluster near behaviorally important elements such as the entrance, the leaning areas, and the surrounding seating zones. These patterns are consistent with expected passenger behavior, where convenience, accessibility, and comfort typically drive spatial decisions. The gradient of colors, ranging from darker reds to lighter tones, illustrates how the model differentiates between more and less attractive areas based on the estimated parameters.

The bottom-left (b) map displays the empirical (observed) probability distribution for the same training dataset. This representation reflects actual passenger movement and standing preferences. High-activity areas correspond to key features, including major access points and waiting areas. When compared with the predicted training map above it, the spatial patterns show strong alignment: areas of concentrated empirical activity largely coincide with regions the model identifies as highly likely. This similarity indicates that the model effectively captures the spatial logic present in the observed data.

Turning to the testing dataset, the top-right map (c) presents the predicted probability distribution for observations not used during model estimation. Despite being applied to new data, the predicted spatial patterns remain consistent with those seen in the training predictions. High-probability zones again appear near prominent features such as the entrance, the leaning areas, and the surrounding seating zones. This repeated pattern is an important indicator of generalization: the model is not merely memorizing the training sample but is instead capturing underlying behavioral tendencies that persist across datasets.

The bottom-right map (d) shows the empirical probability distribution for the testing dataset. Although some natural differences appear compared to the training data, as expected in any out-of-sample setting, the overall structure remains comparable. Regions with higher observed activity align closely with those emphasized in the predicted testing map directly above. This spatial coherence suggests that passengers exhibit similar behavioral patterns across samples, and that the model is able to anticipate these patterns with reasonable accuracy.

Together, the four maps confirm the conclusions drawn from the probabilistic and rank-based analyses. The model replicates realistic spatial behavior within the training sample, maintains consistent spatial logic when exposed to new data, and avoids signs of overfitting. By visually demonstrating how predicted and observed patterns overlap across both datasets, the heat maps provide a robust qualitative validation that supports the reliability and behavioral credibility of the developed spatial choice model.

5.2.6 Overall Assessment

Taken together, the probabilistic, spatial, and rank-based validation results provide a comprehensive and coherent assessment of the model's performance. Across all three dimensions, the evidence consistently indicates that the developed spatial choice model behaves reliably, generalizes well, and reflects plausible passenger decision-making mechanisms.

First, the probabilistic indicators show that the model reproduces observed choice probabilities with a high degree of accuracy and minimal deviation. The close alignment between predicted and empirical probabilities in both the training and testing datasets demonstrates that the estimated parameters capture meaningful behavioral patterns rather than noise or dataset-specific effects.

Second, the spatial validation confirms that the model accurately reconstructs the physical locations where passengers tend to position themselves. Predicted choices are spatially close to observed choices, and the spatial structure present in the empirical data, particularly the concentration of activity near key functional zones such as entrances, leaning areas, and seating regions, is consistently mirrored by the model. The similarity in spatial distributions across training and testing sets further highlights the robustness of this behavior.

Third, the rank-based metrics reveal that the ordering of alternatives implied by the model aligns closely with the ranking derived from empirical observations. This consistency persists even out-of-sample, indicating that the model maintains stable preference structures when applied to new observations.

Finally, the Top-N predictive performance reinforces these findings by demonstrating that the empirically chosen locations frequently fall within the highest-ranked model predictions. The stability of coverage across datasets suggests that the model not only identifies the correct spatial zones of activity but also maintains predictive relevance beyond the estimation sample.

Overall, the small differences between training and testing results across all validation dimensions confirm that the model avoids overfitting and exhibits strong generalization. The combined evidence therefore supports the conclusion that the model provides reliable, interpretable, and behaviorally meaningful predictions of spatial choice behavior on the platform.

6. Conclusions, Discussion, and Implications

This chapter synthesizes the main findings of the study and situates them within the existing literature on passenger behavior and platform design. It begins by answering the main research question and summarizing the principal empirical results. It then discusses these results in relation to previous studies, followed by an evaluation of the strengths and limitations of the research. The chapter subsequently outlines the theoretical and practical implications of the findings, with particular attention to platform design and management in the Dutch railway context.

6.1 Conclusions: main research question and key findings

This study aimed to understand how passengers choose where to wait on railway platforms and to develop a behavioral model capable of predicting these location choices. Using empirical observations from a Dutch railway platform, a multinomial logit model was estimated incorporating spatial, behavioral, and environmental attributes.

The results demonstrate that passenger waiting location choice is systematic rather than random and is driven by a consistent set of behavioral mechanisms. Nine variables were found to be statistically significant: distance to entrances, distance to information boards, distance to the track edge, proximity to seats, proximity to leaning areas, proximity to kiosks, proximity to the track, passenger density, and an interaction between weather, lighting, and distance to the track edge.

Among these, proximity to seating emerged as the strongest determinant, followed by proximity to entrances and leaning areas, confirming the central role of comfort and accessibility in waiting behavior. Safety-related factors, particularly distance from the track edge, also exerted a strong influence, while congestion-related features such as kiosks and information boards reduced the attractiveness of nearby waiting locations. Passenger density showed a positive effect, indicating moderate social clustering behavior. Finally, environmental conditions were found to moderate safety perceptions, with passengers positioning themselves closer to the track under favorable lighting and weather conditions.

The model exhibited strong predictive performance and reproduced observed spatial distributions with high accuracy, demonstrating the suitability of discrete choice modeling for analyzing passenger waiting behavior on platforms.

6.2 Discussion of results in relation to literature

6.2.1 Accessibility and entrance proximity

The strong positive effect of proximity to entrances corroborates earlier findings by Bosina et al. (2015) and Seriani et al. (2016) [47][46], who observed clustering near access points due to reduced walking effort and improved boarding convenience. However, this study extends existing knowledge by quantifying the relative importance of entrance proximity within a multivariate behavioral framework. Even after controlling for density and environmental conditions, entrance distance remained one of the dominant predictors of waiting location, highlighting that accessibility is a primary driver of spatial choice rather than merely a secondary consequence of crowding.

6.2.2 Comfort infrastructure: seating and leaning areas

Consistent with Bandara & Hewawasam (2020) and Mizuno & Tokuda (2023) [30][31], seating was found to exert the strongest positive influence on waiting location choice. A novel contribution of this study is the explicit identification of leaning areas as nearly equivalent in attractiveness to seating. While previous ergonomic studies emphasized seating facilities, the present results indicate that passengers value a broader category of supportive micro-infrastructure, suggesting that comfort is not limited to formal seating but also includes opportunities for physical support during waiting.

6.2.3 Safety and track-edge avoidance under environmental conditions

Track-edge avoidance has been widely documented as fundamental safety behavior (Seriani et al., 2016, Bosina et al., 2015) [46][47]. This study confirms this tendency but further demonstrates that it is context-dependent rather than fixed. The significant interaction between weather, lighting, and distance to the track edge reveals that favorable environmental conditions reduce perceived risk, leading passengers to occupy areas closer to the track.

This finding extends environmental behavior studies such as Kuipers et al. (2021) and Seriani & Fujiyama (2019) [52][15], which identified weather and lighting as influential factors, by quantitatively integrating them into a location choice model. The result challenges the assumption in earlier models that safety margins remain constant across conditions.

6.2.4 Avoidance of kiosks and information boards

Contrary to some studies that describe information facilities as spatial attractors, the present analysis indicates that proximity to kiosks and information boards reduces location utility. This aligns with Ahn et al. (2017) and Zhang et al. (2020) [25][48], who noted congestion and circulation conflicts near infrastructure elements. The divergence from studies that report attraction effects suggests that contextual layout characteristics, particularly in compact Dutch platforms, may transform functional facilities into local disamenities for waiting passengers.

6.2.5 Social behavior and density

The positive coefficient for passenger density supports findings by Zhang et al. (2020) and Kim et al. (2018) [48][50], indicating a tendency toward moderate social clustering. This behavior may reflect perceived safety, social comfort, or information-seeking motives and confirms that passenger distribution is partially governed by psychological and social mechanisms in addition to physical constraints.

6.2.6 Methodological contribution

While previous studies relied heavily on agent-based simulations, cellular automata, or macroscopic flow models (Seriani et al., 2016, Zhang et al., 2020) [46][48], this study demonstrates that a statistically estimated discrete choice model can capture spatial, environmental and behavioral determinants with high predictive accuracy. The favorable log-likelihood value indicates competitive performance compared to other pedestrian choice applications and supports the integration of behavioral modeling into station design analysis.

6.3 Main observations on waiting behavior

The results collectively highlight several core behavioral principles:

- Passengers prioritize comfort first, particularly seating and physical support.
- Accessibility strongly shapes spatial concentration, especially near entrances.
- Safety considerations dominate near hazardous areas but are moderated by environmental context.
- Waiting locations are influenced by congestion avoidance, not merely functional proximity to services.
- Passengers exhibit moderate social clustering rather than dispersion or isolation.

These patterns confirm that waiting behavior reflects a structured decision-making process combining physical constraints, psychological preferences, and contextual adaptation.

6.4 Contribution and strengths of the Study

This study contributes to platform research by integrating behavioral theory with detailed spatial and environmental variables in a single empirical modeling framework. It shows that waiting location choice is shaped by the combined influence of comfort, safety, accessibility, congestion, and social behavior rather than by isolated factors.

By using high-resolution observational data for both estimation and validation, the analysis reflects real passenger responses to actual platform conditions, moving beyond purely conceptual or simulation-based approaches. The explicit inclusion of weather and lighting, together with interaction effects, further advances existing work by demonstrating that safety perception and comfort are context-dependent and vary across environmental situations.

Finally, the application to a major Dutch railway station grounds these methodological contributions in a realistic operational setting, strengthening the relevance of the findings for both academic research and station design practice.

6.5 Limitations

Despite its contributions, this study has several limitations:

- Temporal and spatial scope: The dataset represents a limited temporal window, focusing on a two-minute period before train arrivals, and may not fully capture peak-hour dynamics or short-term fluctuations. Only one station layout was analyzed, limiting the generalizability of the results.
- Passenger heterogeneity and interactions: Variations among passengers (e.g., age, mobility constraints, trip purpose, group behavior) were not explicitly modeled. The dataset treats each passenger independently and does not account for sequential arrivals or dynamic interactions, potentially overlooking social influences that affect decision-making.
- Sensor and data limitations: The study relies on high-frequency sensor data capturing only x and y coordinates, which restricts the depth of behavioral insights. Nuanced behaviors, such as waiting patterns, sitting versus standing, or individual interactions with the environment, are not fully captured. Sensor placement was dictated by practical and cost considerations, which may reduce data richness. High-frequency data also introduces some redundancy and increases the risk of overfitting in the models.
- Operational and environmental variables: Real-time operational factors, including delays, train composition, and platform announcements, were not dynamically incorporated. Some environmental and contextual factors, such as extreme weather conditions or holiday periods, were excluded.

Future research should employ more advanced sensors (e.g., infrared or thermal imaging), multi-station datasets, and passenger segmentation to improve the accuracy and generalizability of behavioral predictions, while also incorporating sequential passenger arrivals, interactions, and operational variables for a more comprehensive understanding of passenger behavior.

6.6 Theoretical implications

The findings contribute to transport and pedestrian behavior theory in three ways:

- They empirically validate that waiting location choice can be modeled as a utility-maximizing decision process.
- They demonstrate that comfort infrastructure plays a dominant role, exceeding traditional accessibility factors.
- They provide quantitative evidence that safety perception is environmentally adaptive, supporting integrated behavioral–environmental modeling frameworks (Ferri & Popp, 2022, Van Hagen, 2011) [49][4].

This supports a shift from purely flow-based representations toward hybrid behavioral-spatial models of platform use.

6.7 Practical implications and design recommendations

The results translate directly into actionable guidance for station design and management:

- **Seating and leaning infrastructure**
  Seating and leaning supports should be evenly distributed along platforms to prevent localized congestion. Incorporating leaning areas can provide cost-effective comfort benefits where seating capacity is limited.
- **Entrance placement and flow guidance**
  Multiple access points and visual guidance systems should be used to disperse passengers deeper into platforms, mitigating clustering near entrances.
- **Kiosks and information boards**
  Retail facilities and information boards should be placed along circulation corridors rather than in primary waiting zones to avoid obstruction and crowding.
- **Track-edge safety design**
  Clear buffer zones near the track should be maintained, supplemented by adaptive lighting strategies that consider weather-related behavioral shifts.
- **Environmental quality**
  Improving lighting and shelter enhances perceived safety and spatial balance, supporting more uniform distribution during adverse conditions.

6.8 Concluding remarks

By combining empirical observation with behavioral modeling, this study advances understanding of how passengers interact with platform environments at a spatial level. It demonstrates that waiting behavior reflects rational adaptation to comfort, safety, accessibility, social context, and environmental conditions. The proposed framework provides both theoretical insight and practical tools for designing safer, more comfortable, and more efficient railway platforms in the Netherlands and beyond.

References

[1] Daamen, W., Bovy, P. H., & Hoogendoorn, S. P. (2005). Influence of changes in level on passenger route choice in railway stations. Transportation Research Record, 1930(1), 12–20.

[2] de Ana Rodríguez, G., Seriani, S., & Holloway, C. (2016). Impact of platform edge doors on passengers' boarding and alighting time and platform behavior. Transportation Research Record, 2540(1), 102–110.

[3] Daamen, W., & Hoogendoorn, S. P. (2003). Experimental research of pedestrian walking behavior. Transportation Research Record, 1828(1), 20–30.

[4] Van Hagen, M. (2011). Waiting Experience at Train Stations. Eburon Uitgeverij BV.

[5] Løvås, G. G. (1994). Modeling and simulation of pedestrian traffic flow. Transportation Research Part B: Methodological, 28(6), 429–443.

[6] Zhang, Q., Han, B., & Li, D. (2008). Modeling and simulation of passenger alighting and boarding movement in Beijing metro stations. Transportation Research Part C: Emerging Technologies, 16(5), 635–649.

[7] Qi, X., Mao, B., Li, M., & Feng, X. (2014). Simulation of passenger flows on urban rail transit platforms based on adaptive agents. Journal of Transportation Systems Engineering and Information Technology, 14(1), 28–33.

[8] Páez, A., & Scott, D. M. (2005). Spatial statistics for urban analysis: A review of techniques with examples. GeoJournal, 61(1), 53–67.

[9] Fruin, J. J. (2012). Pedestrian Planning and Design. Elevator World, Inc.

[10] Fruin, J. J. (1971). Pedestrian Planning and Design. Metropolitan Association of Urban Designers and Environmental Planners.

[11] Henderson, L. F. (1974). On the fluid mechanics of human crowd motion. Transportation Research, 8(6), 509–515.

[12] Helbing, D., Farkas, I., & Vicsek, T. (2000). Simulating dynamical features of escape panic. Nature, 407(6803), 487–490.

[13] Ben-Akiva, M., & Lerman, S. R. (1985). Discrete Choice Analysis: Theory and Application to Travel Demand. MIT Press.

[14] Xu, H., Wang, Z., & Zhou, M. (2020). Predicting passenger flow distribution on platforms using machine learning and behavioral models. Transportation Research Part C, 118, 102732.

[15] Seriani, S., & Fujiyama, T. (2019). Modelling passengers' distribution on railway platforms: A review of methods and applications. Transport Reviews, 39(4), 514–536.

[16] ProRail. (2022). Annual Report 2022. Utrecht: ProRail.

[17] KiM Netherlands Institute for Transport Policy Analysis. (2021). Mobiliteitsbeeld 2021. Ministry of Infrastructure and Water Management.

[18] Train, K. (2009). Discrete Choice Methods with Simulation (2nd ed.). Cambridge University Press.

[19] Ortúzar, J. de D., & Willumsen, L. G. (2011). Modelling Transport (4th ed.). Wiley.

[20] Horni, A., Nagel, K., & Axhausen, K. W. (2016). The Multi-Agent Transport Simulation MATSim. Ubiquity Press.

[21] Baron, N. (2021). Designing Paris Gare du Nord for pedestrians or for clients? New retail patterns as flow optimization strategies. In Spatial Implications and Planning Criteria for High-Speed Rail Cities and Regions (pp. 181-200). Routledge.

[22] van der Werf, I. (2019). Changing the mindset: how public transport can become more user centered. Towards User-Centric Transport in Europe: Challenges, Solutions and Collaborations, 105-117.

[23] Fruin, J. J. (1987). Pedestrian Planning and Design, rev. ed. Elevator World.

[24] Passini, R. (1992). Wayfinding in architecture. (No Title).

[25] Ahn, S., Hutchison, D., Jones, L., Kundu, R., & Lam, W. N. (2017, November). Evaluating the Impact of Improvement of Wayfinding at a Train Station: Agent-based Simulation Study. In Proceedings of the 39th Australasian Transport Research Forum, Auckland, New Zealand (pp. 27-29).

[26] Zhang, J. (2021). Agent-based optimizing match between passenger demand and service supply for urban rail transit network with NetLogo. Ieee Access, 9, 32064-32080.

[27] Xu, H., Tian, C., & Li, Y. (2020). Emergency evacuation simulation and optimization for a complex rail transit station: a perspective of promoting transportation safety. Journal of advanced transportation, 2020.

[28] Ibrahim, A. N. H., Borhan, M. N., Md Yusoff, N. I., & Ismail, A. (2020). Rail-based public transport service quality and user satisfaction–a literature review. Promet-Traffic&Transportation, 32(3), 423-435.

[29] Rüger, B. (2018). Influence of passenger behaviour on railway-station infrastructure. In Sustainable Rail Transport: Proceedings of RailNewcastle 2017 (pp. 127-160). Cham: Springer International Publishing.

[30] Bandara, D., & Hewawasam, C. (2020). A comparative study on effectiveness of underpass and overpass among pedestrians in different urban contexts in Sri Lanka. Journal of Service Science and Management, 13(5), 729-744.

[31] Mizuno, T., & Tokuda, K. (2023). Reducing falls among visually impaired individuals on railway platforms: Field research on environmental challenges and solutions. Heliyon, 9(3).

[32] Concern, C. (2002). People's Perceptions of Personal Security and Their Concerns about Crime on Public Transport: Literature Review. The Department.

[33] Romero, C., Zamorano, C., & Monzón, A. (2023). Exploring the role of public transport information sources on perceived service quality in suburban rail. Travel Behaviour and Society, 33, 100642.

[34] Virgona, A., Kirchner, N., & Alempijevic, A. (2015, January). Sensing and perception technology to enable real time monitoring of passenger movement behaviours through congested rail stations. In ATRF 2015-Australasian Transport Research Forum 2015, Proceedings.

[35] Zheng, Z., Wang, H., Liu, W., & Peng, L. (2023). Toward real-time congestion measurement of passenger flow on platform screen doors based on surveillance videos analysis. Physica A: Statistical Mechanics and its Applications, 612, 128474.

[36] Hänseler, F. S., van den Heuvel, J. P., Cats, O., Daamen, W., & Hoogendoorn, S. P. (2020). A passenger-pedestrian model to assess platform and train usage from automated data. Transportation Research Part A: Policy and Practice, 132, 948-968.

[37] Zhang, Y., Umar, A. M., & An, M. (2020). Overview of optimization models and algorithms for train platforming problem. In Proceedings of the 4th International Conference on Electrical and Information Technologies for Rail Transportation (EITRT) 2019: Rail Transportation System Safety and Maintenance Technologies (pp. 707-716). Springer Singapore.

[38] Xu, J., Wang, Z., Yao, S., & Xue, J. (2022). Train operations organization in high-speed railway station considering variable configuration. Sustainability, 14(4), 2365.

[39] Li, W., & Zhu, W. (2016). A dynamic simulation model of passenger flow distribution on schedule-based rail transit networks with train delays. Journal of Traffic and Transportation Engineering (English Edition), 3(4), 364-373.

[40] Hartleb, J., & Schmidt, M. (2022). Railway timetabling with integrated passenger distribution. European Journal of Operational Research, 298(3), 953-966.

[41] Ingvardson, J. B., Nielsen, O. A., Raveau, S., & Nielsen, B. F. (2018). Passenger arrival and waiting time distributions dependent on train service frequency and station characteristics: A smart card data analysis. Transportation Research Part C: Emerging Technologies, 90, 292-306.

[42] Oliveira, L. C., Fox, C., Birrell, S., & Cain, R. (2019). Analysing passengers' behaviours when boarding trains to improve rail infrastructure and technology. Robotics and Computer-Integrated Manufacturing, 57, 282-291.

[43] Szplett, D., & Wirasinghe, S. C. (1984). An investigation of passenger interchange and train standing time at LRT stations:(i) Alighting, boarding and platform distribution of passengers. Journal of advanced transportation, 18(1), 1-12.

[44] Wirasinghe, S. C., & Szplett, D. (1984). An investigation of passenger interchange and train standing time at LRT stations:(ii) estimation of standing time. Journal of advanced transportation, 18(1), 13-24.

[45] Krstanoski, N. (2014). MODELLING PASSENGER DISTRIBUTION ON METRO STATION PLATFORM. International Journal for Traffic & Transport Engineering, 4(4).

[46] Seriani, S., Fujiyama, T., & Holloway, C. (2016, September). Modelling passenger distribution and interaction on platform train interfaces. PANAM 2016.

[47] Bosina, E., Britschgi, S., Meeder, M., & Weidmann, U. (2015, April). Distribution of passengers on railway platforms. In Proceedings of the 15th Swiss Transport Research Conference.

[48] Zhang, N., Chen, F., Zhu, Y., Peng, H., Wang, J., & Li, Y. (2020). A study on the calculation of platform sizes of urban rail hub stations based on passenger behavior characteristics. Mathematical Problems in Engineering, 2020, 1-14.

[49] Ferri, A., & Popp, M. (2022). 'Wayfeeling': Navigating through emotional and sensorial responses in public transit. Wellbeing, Space and Society, 3, 100104.

[50] Kim, J., Prestona, J., & Revella, K. (2018). Exploring the effect of occupancy information on passengers' behaviour change to relieve train crowding. In World Conference on Transport Research Society 00, 000–00 (2018) (pp. 1-19).

[51] Lyon, A. R., & Koerner, K. (2016). User-centered design for psychosocial intervention development and implementation. Clinical Psychology: Science and Practice, 23(2), 180.

[52] Kuipers, R. A., Palmqvist, C. W., Olsson, N. O., & Winslott Hiselius, L. (2021). The passenger's influence on dwell times at station platforms: a literature review. Transport Reviews, 41(6), 721-741.

[53] Israel, G. D. (1992). Determining sample size.

[54] Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758). New York: springer.

[55] Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. Statistics Surveys, 4, 40–79.

[56] Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. Geoscientific Model Development, 7, 1247–1250.

[57] Ben-Akiva, M., & Bierlaire, M. (1999). Discrete choice methods and their applications to short-term travel decisions. In R. Hall (Ed.), Handbook of Transportation Science. Springer.

[58] Kendall, M. G. (1938). A new measure of rank correlation. Biometrika, 30(1/2), 81–93.

[59] Spearman, C. (1904). The proof and measurement of association between two things. American Journal of Psychology, 15(1), 72–101.

[60] Zhu, S., Zhang, L., & Wang, Q. (2013). Modeling and analysis of pedestrian behavior in public transport facilities. Transportation Research Part C, 31, 1–12.

[61] Choudhury, C., et al. (2018). Passenger distribution modeling in transport hubs using sensor data. Journal of Transportation Engineering, 144(6), 04018028.

[62] Eindhoven Centraal. (n.d.). In Somda RailWiki. https://railwiki.nl/index.php/Eindhoven

[63] Ruland Architecten. (n.d.). Platform buildings Eindhoven railway station. https://ruland.nl/en/portfolio-item/20-12181_perrongebouwen-en/

[64] ExpatINFO Holland. (n.d.). 15 busiest Dutch train stations. https://expatinfoholland.nl/top-10/netherlands-15-busiest-train-stations/

[65] Heijmans. (2019). Station Eindhoven – reconstructie en uitbreiding van de stationspassage. Heijmans Civiel. https://www.heijmans.nl/nl/projecten/Station-Eindhoven/

[66] National Academies of Sciences, Engineering, and Medicine. (2013). Transit capacity and quality of service manual.

Appendix

Table A: Snapshot of the Dataset

Table A:* *Shows a sample of the final dataset (142,256 rows) capturing passenger decisions, environmental conditions, and contextual factors at 0.1 second intervals within the two minutes before train arrivals.

datetime	bin_id	persons_per_bin	area	persons_per_m2	weather	lighting
2023,06,03 09:00:00+02:00	211	1	3.83689	0.26062802	1	1
2023,06,03 09:00:00+02:00	214	1	3.83689	0.26062802	1	1
2023,06,03 09:00:00+02:00	219	2	3.83689	0.521256039	1	1
2023,06,03 09:00:00+02:00	220	1	3.83689	0.26062802	1	1
2023,06,03 09:00:00+02:00	225	1	3.83689	0.26062802	1	1
2023,06,03 09:00:00.100000+02:00	211	1	3.83689	0.26062802	1	1
2023,06,03 09:00:00.100000+02:00	214	1	3.83689	0.26062802	1	1
2023,06,03 09:00:00.100000+02:00	219	2	3.83689	0.521256039	1	1
2023,06,03 09:00:00.100000+02:00	220	1	3.83689	0.26062802	1	1
2023,06,03 09:00:00.100000+02:00	225	1	3.83689	0.26062802	1	1

Appendix

dist_to_kiosk_upper	dist_to_information_board_upper	dist_to_information_board_below	dist_to_entrance	centroid
12.41925305	52.8090899	51.85684021	56.03203399	POINT (59352 ,7708.5)
10.10612113	51.98811157	52.14716717	56.019	POINT (59352 ,1459.5)
12.29831258	54.01092891	53.80520367	57.861	POINT (61194 ,3542.5)
11.93977455	53.8254538	53.97909573	57.861	POINT (61194 ,1459.5)
14.087546	55.84247579	55.64352255	59.703	POINT (63036 ,3542.5)
12.41925305	52.8090899	51.85684021	56.03203399	POINT (59352 ,7708.5)
10.10612113	51.98811157	52.14716717	56.019	POINT (59352 ,1459.5)
12.29831258	54.01092891	53.80520367	57.861	POINT (61194 ,3542.5)
11.93977455	53.8254538	53.97909573	57.861	POINT (61194 ,1459.5)
14.087546	55.84247579	55.64352255	59.703	POINT (63036 ,3542.5)

dist_to_seats_right	dist_to_seats_below	dist_to_seats_upper	dist_to_kiosk_non_entrance	dist_to_kiosk_entrance	dist_to_kiosk_below
6.745252126	4.3585	5.3585	10.52610831	32.00478333	10.52610831
5.224451574	1.8905	0.8905	10.052	31.852	10.52670196
3.311599651	0.1925	1.1925	11.894	33.694	11.93959975
3.4238321	1.8905	0.8905	11.894	33.694	12.29780412
1.476601588	0.1925	1.1925	13.736	35.536	13.7755037
6.745252126	4.3585	5.3585	10.52610831	32.00478333	10.52610831
5.224451574	1.8905	0.8905	10.052	31.852	10.52670196
3.311599651	0.1925	1.1925	11.894	33.694	11.93959975
3.4238321	1.8905	0.8905	11.894	33.694	12.29780412
1.476601588	0.1925	1.1925	13.736	35.536	13.7755037

dist_to_track_edge_upper	dist_to_lean_edge_bottom	dist_to_lean_edge_upper	dist_to_information_board	dist_to_seats	dist_to_seats_left
11.4585	56.03203399	56.85405292	51.85684021	4.3585	4.376999686
5.2095	56.24531093	56.1257205	51.98811157	0.8905	0.977033392
7.2925	57.93653534	58.12585163	53.80520367	0.1925	2.252241606
5.2095	58.08013396	57.96432921	53.8254538	0.8905	2.414234092
7.2925	59.77620777	59.95971577	55.64352255	0.1925	4.090532025
11.4585	56.03203399	56.85405292	51.85684021	4.3585	4.376999686
5.2095	56.24531093	56.1257205	51.98811157	0.8905	0.977033392
7.2925	57.93653534	58.12585163	53.80520367	0.1925	2.252241606
5.2095	58.08013396	57.96432921	53.8254538	0.8905	2.414234092
7.2925	59.77620777	59.95971577	55.64352255	0.1925	4.090532025

is_next_to_seats	is_next_to_leaning	is_visible_from_entrance	dist_to_lean_edge	dist_to_track_edge	dist_to_track_edge_bottom
0	0	0	10.52610831	1.0415	1.0415
1	0	0	10.052	5.2095	7.2905
1	0	0	11.894	5.2075	5.2075
1	0	0	11.894	5.2095	7.2905
1	0	0	13.736	5.2075	5.2075
0	0	0	10.52610831	1.0415	1.0415
1	0	0	10.052	5.2095	7.2905
1	0	0	11.894	5.2075	5.2075
1	0	0	11.894	5.2095	7.2905
1	0	0	13.736	5.2075	5.2075

pt_y	pt_x	is_next_to_track	is_next_to_kiosk	is_next_to_information_board
,7709	59352	1	0	0
,1460	59352	0	0	0
,3543	61194	0	0	0
,1460	61194	0	0	0
,3543	63036	0	0	0
,7709	59352	1	0	0
,1460	59352	0	0	0
,3543	61194	0	0	0
,1460	61194	0	0	0
,3543	63036	0	0	0