Analyzing the impact of fare-free public transport policies on crowding patterns at stations using crowdsensing data

Qing-Long Lu[a], Vishal Mahajan[a], Cheng Lyu[a], Constantinos Antoniou[a]

[a]*Chair of Transportation Systems Engineering, Technical University of Munich, Munich, Germany*

ARTICLE INFO

Keywords:
Fare-free public transport
Crowding pattern
Crowdsensing data
Station heterogeneity
9-EUR ticket

ABSTRACT

Fully or partially fare-free public transport (FFPT) is a measure to make public transport (PT) more attractive and affordable. Cities worldwide are experimenting with a variety of fare discount policies, which will lead to spatiotemporal changes in mobility patterns, including crowding in PT stations. However, the response of PT stations to these policies can vary due to the heterogeneity in their surrounding built environment and other factors. Non-conventional data sources could be used to better model and understand these spatiotemporal dynamics. In this study, we propose a three-step methodological framework to understand the impact of FFPT or extreme fare discount interventions on PT demand patterns, specifically focusing on crowding patterns in PT stations, using crowdsensing data. First, we design a busyness-based similarity measure that leverages the histogram method to capture changes in crowding patterns. Then, we employ a Gaussian Mixture Model (GMM) to cluster PT stations based on their crowding pattern deviations at different stages of policy implementation. This clustering step enables the identification of distinct station types based on their response to the policy. Finally, we train a LightGBM model to learn the relationship between crowding pattern changes and the spatial-temporal characteristics of PT stations, using the busyness-based station types identified by GMM as labels. We apply our methodology to a public transport experiment in Germany during the summer of 2022 when the country introduced a monthly "9-EUR" ticket valid on local and regional PT nationwide. The clustering results show three station types: unaffected, mildly stimulated, and intensely stimulated stations. Furthermore, the classifier indicates that the station's location, activity options near the station, and population within the adjacent area of the station, and the crowding patterns under normal operations (before policy implementation) play a significant role in the heterogeneity of the 9-EUR ticket's impact. The insights gained from our study can help planners better understand and manage the crowding at PT stations during fare interventions.

1. Introduction

Public transport (PT) plays an important role in sustainable urban development. Population growth and urban expansion have led to increased PT ridership, which can result in crowding in the PT system if the transit supply is not augmented concurrently, and therefore reduce PT system performance and rider satisfaction (Li and Hensher, 2011; Halvorsen et al., 2020). Accordingly, PT demand management (PTDM) towards crowding reduction is crucial for improving service quality (Hensher et al., 2011). Fully or partially fare-free PT (FFPT) policies are becoming popular as a PT incentive to increase PT mode share. FFPT allows individuals to utilize public transportation services free of charge or with reduced financial charges, thereby fostering equitable public transport access. These policies can also lead to increased transit demand and ridership (Thøgersen, 2009; Cools et al., 2016). Similarly, if the PT supply is not adjusted correspondingly, such policies might render deteriorating crowding situations (Börjesson et al., 2015). The resulting crowding could persist either temporarily or permanently, depending on the duration and area in which the policy is in effect. The ridership increase resulting from the long-term changes in population demographics and land use requires strategic measures, e.g., investment in infrastructure. On the contrary, short-term or immediate interventions, such as FFPT policies, require tactical and operational responses. PTDM becomes even more critical in the latter case as expanding the PT network and station capacity in the short term is not easily achievable. Understanding the potential impact of FFPT on crowding patterns is imperative to develop effective operational strategies for maintaining satisfactory levels of PT service after its implementation.

qinglong.lu@tum.de (Q. Lu); vishal.mahajan@tum.de (V. Mahajan); cheng.lyu@tum.de (C. Lyu); c.antoniou@tum.de (C. Antoniou)
ORCID(s): 0000-0002-6087-8670 (Q. Lu); 0000-0003-2131-2831 (V. Mahajan); 0000-0002-6356-6947 (C. Lyu); 0000-0003-0203-9542 (C. Antoniou)

Naturally, analyzing the impacts of fare interventions requires data specific to the focus of the investigation. Automated Fare Collection (AFC) and Automated Passenger Counting (APC) systems are the most common transit data sources. These data provide precise and fine-grained PT ridership data; however, they are not available for open transit systems. Also, these data are owned by PT operators and may not be shared with the modeling and research community (Vitello et al., 2023). Even if these data are shared, there can be variability in how the data are collected and shared across different transit agencies. Alternatively, studies can be conducted by using user surveys or location data (e.g., Cools et al., 2016; Börjesson et al., 2015), especially in the cities without AFC and APC systems, like most cities in Germany. However, these data collection methods are time-consuming, labor-intensive, and costly, which limits their coverage area and collection frequency. Due to the need for data with uniformity in the collection, finer temporal resolution, and extensive spatial coverage, empirical analyses are often challenging when analyzing the impact of PT pricing policies on a large scale. As a result, it is imperative to develop a methodology that utilizes easily accessible opportunistic data sources to analyze the impact of PT fare interventions on PT demand patterns.

The proliferation of smartphones equipped with positioning technologies, such as GPS, has provided a novel means of collecting mobility data, opening up new avenues for PT policy evaluation in a more cost-effective manner (Tse et al., 2018). These devices generate vast volumes of real-time data on individuals' activities and mobility through location-based services, social networks, and other mobile applications. Certain smartphone datasets have been leveraged for PT analysis for purposes such as PT network design (e.g., Pinelli et al., 2016, using mobile phone trajectories) and PT accessibility analysis (e.g., Cai et al., 2017, using mobile phone signaling data). Unlike those normally unavailable sensitive personal data, crowdsourced data, such as geotagged tweets (Chaniotakis et al., 2017), usually provide new opportunities to gain more insights into urban mobility, including PT demand patterns (Capponi et al., 2019; Vitello et al., 2023). However, despite the rapid growth of crowdsourced data, their potential for analyzing PT demand patterns, particularly in the context of pricing policies, has not been fully explored (Niu and Silva, 2020).

Crowdsensed check-in rates or busyness data at points of interest (POIs) is a typical kind of crowdsourced data. The raw location data from mobile devices near POIs are collected and transmitted via an internet connection to a central server. These raw data from numerous personal devices are processed, anonymized, and aggregated to provide the busyness data at the POIs. Among others, the real-time busyness data of PT stations, representing the crowding patterns, contains valuable information about PT demand patterns. One can evaluate the impact of PT policies by comparing the crowding patterns extracted from PT station busyness data at different phases of policy implementation. These emerging crowdsensing data have wide coverage and fine resolution in spatial and temporal dimensions, respectively, enabling a detailed empirical analysis of the impact of PT policies on PT demand patterns at a large scale.

When analyzing demand patterns at finer resolution, such as at the level of POI, their heterogeneity in terms of spatiotemporal factors of land use and transportation must be accounted for (Fan et al., 2022). For instance, the complex relationship among surrounding land use, transport network, population, temporal features, and historical demand patterns can lead to different crowding patterns. Therefore, the methodology to analyze the impacts of PT pricing policies should address such heterogeneity and uncover its relationship with those station characteristics. This will contribute to the knowledge of urban mobility and activity patterns and assist policymakers and practitioners in formulating more rational auxiliary measures in sectors beyond transportation for PTDM (Vongvanich et al., 2023).

In view of the above, we see an opportunity to use the opportunistic busyness data to address the aforementioned limitations of existing studies in analyzing PT demand patterns. In this study, we present a three-step busyness-based framework to empirically evaluate the impact of FFPT on crowding patterns at PT stations using a wide-coverage crowdsensing dataset. Specifically, we first devise a similarity measure based on the changes in crowding patterns at different policy implementation stages. We then integrate a clustering model to identify different types of PT stations based on their crowding patterns and a classification model to infer the relationship between the policy impact and station characteristics. Our study offers the following contributions:

(1) We develop a methodological framework leveraging fine-resolution and wide-coverage PT station busyness data for demand pattern analysis.

(2) We apply our framework to the case of the 9-Euro ticket experiment in Germany using opportunistic data and provide empirical findings.

The remainder of this paper is structured as follows. We review the studies investigating the impact of FFPT in Section 2. We introduce the three-step evaluation framework in Section 3. The study area, policy of investigation, and crowdsensing data used are described in Section 4. Section 5 presents the results and model performance. Section 6 provides a discussion of the design and implementation of FFPT based on the results. Finally, major findings, limitations, and future research direction are concluded in Section 7.

2. Literature on free-fare public transport analysis

FFPT emerges as an effective approach to alleviate traffic congestion within urban areas by encouraging the modal shift from private cars to PT (Fujii and Kitamura, 2003). It has been implemented in many cities despite distinctive implementation details (see Kębłowski, 2020, for a review). In contrast to push measures, such as congestion pricing, pull measures like FFPT receive greater public support (Eliasson et al., 2009). Empirical evidence suggests that FFPT enjoys considerable endorsement from the public, even when confronted with its relatively high implementation costs (Börjesson et al., 2015). The survey data collected from Stockholm, Helsinki, and Lyon revealed that around 56% of participants favored FFPT over alternatives like road extension and congestion pricing (Börjesson et al., 2015). The strong public backing for FFPT is associated with environmental considerations and equity concerns, suggesting the need to highlight its benefits in these aspects to encourage modal shifts from private cars towards public transport (Börjesson et al., 2015). The low-cost "environmental protection" ticket in Freiburg, Germany, in 1984, lends support to such findings. This initiative, allowing passengers to use PT services across Freiburg's entire tariff network at a very low price, led to a remarkable annual growth rate of 7.5% in public transport demand between 1983 and 1995 (FitzRoy and Smith, 1998). Notably, this increase outweighed the effects of other traffic restraint measures such as pedestrianization and low-speed zones, thereby reflecting the potency of well-designed fare reduction policies.

Other than the method that the government fully subsidizes the free PT services (or substantial fare discounts), some cities also proposed a so-called "third payer" system, enabling PT operators to collaborate with private companies so that each entity would be responsible for a portion of the commuting expenses for the company's employees. The free PT launched in Brussels, Belgium, in 2005, is one kind of such policy, under which private companies and the government co-fund 80% and 20%, respectively, of the commuting public transport trips (De Witte et al., 2008). In De Witte et al. (2008), the mobility scheme devised by Kaufmann (2017) was employed to explore the factors influencing the commuters' mode choice in the context of such FFPT. This analysis unveiled that aside from lowering the PT price and improving PT quality (frequency, capacity, connections, etc.), company car ownership also plays an essential role in the employee's mode choice. It also suggested that companies' mobility policies could be tailored to facilitate PT usage, such as compensating PT usage rather than providing company cars.

Moreover, FFPT's influence even extends beyond benefiting recipients. An FFPT initiative targeted at students from Flemish-speaking universities and the STIB company in Brussels in 2003 demonstrated that FFPT led to increased public transport usage among both benefiting and non-benefiting students, along with greater usage of non-free public transport services compared to the previous year (De Witte et al., 2006). Similar insights were also obtained in another region of Belgium — Flanders, as pointed out in Cools et al. (2016). In addition, it revealed that the current use of modes poses the most significant effect on mode choice behavior. Namely, under the effect of habitual behavior, individuals are very likely to make the same choice in a stable context. This is also in line with the study conducted in De Witte et al. (2006) and De Witte et al. (2008), in which habitual behaviors were counted in the appropriation factor of the Kaufmann mobility scheme.

Furthermore, interventions like a free month travel card in Copenhagen (Thøgersen, 2009) and a 30-day FFPT ticket in Värmland, Sweden (Friman et al., 2019), yielded similar insights. The former found that the promotion intervention almost doubled the PT usage in the experiment group, and a positive effect remained even six months afterward. Analogously, the latter found that a temporary FFPT intervention could motivate the shift to PT in the long-term perspective, while it seemed less influential in the short term.

In addition to the aforementioned studies that are mostly based on survey data, some have provided empirical analysis by incorporating data from AFC or APC systems. Cats et al. (2014) empirically evaluated the passenger demand changes after the introduction of FFPT in Tallinn, Estonia. It explicitly distinguished the effects caused by supply variables (e.g., increased service frequency) from those solely due to FFPT policy, concluding that only 1.2% out of 3% of the increased passenger demand can be credited to the FFPT policy. Later on, a follow-up investigation was conducted to analyze travel pattern changes based on individual travel habit surveys (Cats et al., 2017). A year after the introduction of FFPT, PT usage increased by 14%, while patronage increased by 24%. Additional studies have substantiated FFPT's potential, illustrating that fare reductions of varying magnitudes can yield notable demand increases over time (Wallimann et al., 2023). A recent FFPT campaign conducted in Germany during the spring of

2022 showed that 3% of participants seem to systematically shift from private cars to PT, as deduced from the analysis of travel diary records and survey data of 1,200 people (Loder et al., 2023). Furthermore, an obvious increase in the frequency of PT utilization was also found among 25% of new PT passengers (Loder et al., 2023). We also note that AFC and APC data have been used to investigate the impact of FFPT from other perspectives, such as demand shift across the time of the day (e.g., Pluntke and Prabhakar, 2013 in Singapore, Halvorsen et al., 2016 in Hong Kong; Loder et al., 2023 in Germany), social and environmental benefits (e.g., Fearnley, 2013 in Europe), user group identification (e.g., Halvorsen et al., 2020 in Hong Kong). However, none have empirically evaluated the impact on PT demand patterns on a large scale, e.g., covering multiple cities, due to the limitation of these data in spatial coverage.

The potential of busyness data for POI's activity and demand pattern analyses has received significant attention in recent years attributed to their wide coverage and fine resolution in spatial and temporal dimensions, respectively. Examples of such data sources include Foursquare check-ins (D'Silva et al., 2018) and popularity trends (Capponi et al., 2019; Timokhin et al., 2020). Previous studies have successfully extracted spatiotemporal demand patterns and activities from these data (Timokhin et al., 2020; MacKenzie and Cho, 2020; Capponi et al., 2019; Möhring et al., 2021). Live POI check-in rates have also been used to analyze the impact of special events or interventions, such as lockdowns during COVID-19, on POI demand patterns (Mahajan et al., 2021), the impact of natural disasters, such as heavy snowfalls, on people daily activities within the urban area (Santiago-Iglesias et al., 2023). Analogously, the busyness data for PT stations also enables detailed empirical analyses of the impact of PT policies on PT demand patterns at a large scale.

To summarize, we note that most existing analyses of FFPT were based on stated preference surveys. Although some empirical studies have been undertaken, they are limited to the scope of macro-level evaluations. A noteworthy gap lies in the absence of empirical analyses that delve into the effects of FFPT at the finer scale of individual PT stations. This station-level investigation is crucial for capturing variations in travel behavior and preferences across different station contexts. On the other hand, there is no nationwide empirical evidence for the impact of FFPT regarding the varying preferences among citizens from diverse cities. To this end, the present study attempts to answer the following questions: (i) How does FFPT influence busyness patterns at PT stations with different characteristics? (ii) What correlations exist between the extent of changes in busyness patterns and the specific characteristics of these PT stations?

3. Methodology

It is expected that the introduction of fare-free policies to PT services will render spatial-temporal demand changes across the network and therefore alter the crowding patterns at PT stations. Due to the diversity of station contexts, the changes in crowding patterns, however, may differ among stations. Understanding how different stations respond to such policies can help the development of effective responses for demand and crowding management. In this section, we propose a framework driven by crowdsensing check-in data (i.e., busyness data) that can be used to evaluate the impact of FFPT policies on crowding patterns in PT stations and estimate the relationship of the impact with station characteristics.

Specifically, we develop a three-step busyness-based evaluation framework, as shown in Figure 1. In addition to the collection and preprocessing of crowdsensing data and station attributes, the framework integrates the following three components:

Step (1) A histogram-based feature engineering approach is adopted to encode the crowding pattern changes in PT stations. The histogram method is a mean for estimating the probability density distribution of a certain random variable. Hence, it can provide an overview of the impact of the policy of interest on the concerned stations. Besides, factor analysis is applied to extract significant patterns from raw station features and reduce noise and feature dimensions in the feature engineering step.

Step (2) A clustering model is used to identify and label different types of PT stations based on the changes in their crowding patterns due to the implementation of the policy. This step is to mine and characterize the changing patterns of PT station crowding patterns. We define those clusters as busyness-based station types. These busyness-based station types (i.e., Y_c in Figure 1) will serve as class labels inputted to the classification model in the following classification step.

Step (3) Finally, a classification model is trained on the spatiotemporal station attributes and the busyness-based station types identified by the clustering model in Step (2) so as to infer the association between them.

Furthermore, it is worth mentioning that in most platforms providing crowdsensing services, the busyness value of a POI may be unavailable if there is a limited number of users opting for location services. In order to estimate the crowding pattern changes, it is necessary to impute the missing data contained in the raw busyness dataset beforehand. We describe our methodology in detail in the following text of this section.

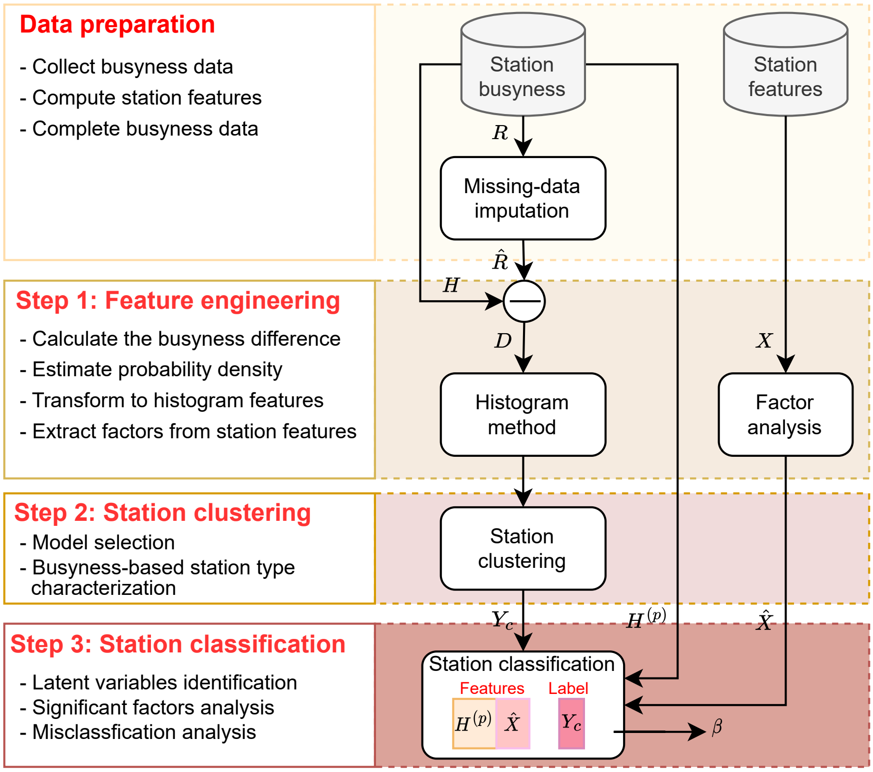


Figure 1: Three-step busyness-based evaluation framework for public transport policies.

3.1. Histogram-based feature engineering from busyness data

Generally, for a specific time stamp, crowdsensing services provide two busyness values for POIs, namely, historical and live busyness. The deviations of live busyness from the historical ones can be used as a proxy for PT demand pattern changes. Here, based on the histogram method, we use the busyness deviations to construct a measurement to describe the similarity among PT stations in terms of the impact of fare-free policies. Figure 2 provides a graphical illustration of the proposed histogram method.

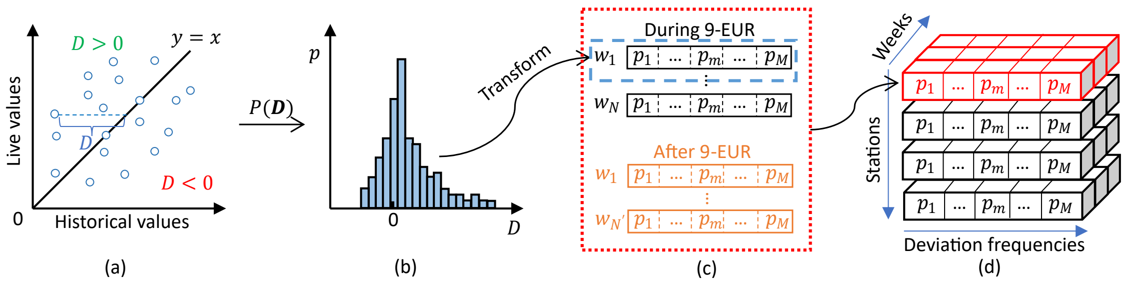


Figure 2: Feature engineering based on the histogram method.

Let $\boldsymbol{H} \in \mathbb{R}^{n_s \times n_t}$, $\hat{\boldsymbol{R}} \in \mathbb{R}^{n_s \times n_t}$ denote the historical and imputed live busyness matrices, where n_s and n_t denote the number of stations and time intervals respectively. We denote H_{ij} and $\hat{R}_{ij}$ the historical and imputed live busyness of station i at time interval j respectively, which can be absolute (number of visitors) or relative values on scale $[0, 100]$ depending on the data source. For instance, Foursquare provides the number of visitors, while Google Maps only provides the normalized value in the range $[0, 100]$. The busyness deviation matrix can then be calculated as $\boldsymbol{D} = \hat{\boldsymbol{R}} - \boldsymbol{H}$ (Figure 2a). The histogram method can then be applied to approximate the probability distribution of the respective busyness deviations for each station and each week separately. As shown in Figure 2b, for one week of data of a station, we can plot the busyness deviations using a histogram; there are 168 deviation values within a week for a platform at which data are updated hourly. In the histogram method, the sample space is divided into M non-overlapping bins, each with width Δ_m. We then count the deviation values located in each bin K_m and approximate the density function by

$$p_m = \frac{K_m l_t}{168\Delta_m}, \quad \forall 1 \leq m \leq M \tag{1}$$

where l_t (h) is the length of each time interval, and 168 is the number of hours each week. Repeating the calculation for every station and every week (Figure 2c), we can obtain a probability density tensor of busyness deviations with dimension $n_s \times M \times n_w$ where n_w is the number of weeks (Figure 2d). Note that bins incapable of distinguishing stations should be discarded. This can be tested by visualizing the distribution of the probability density values per bin. We would show in Section 5.1 that this treatment can improve models' robustness to the hyper-parameters of clustering models. This feature engineering method extracts representative features that can reflect the crowding pattern changes from the PT station busyness data. It is worth mentioning that using deviation values directly for clustering may mislead the model to group together stations with similar crowding patterns but not with similar changes in crowding patterns.

3.2. Clustering for labeling busyness pattern changes

We develop a clustering model to categorize and label PT stations based on their crowding pattern changes due to the implementation of FFPT policies. Denote $\boldsymbol{Q} \in \mathbb{R}^{n_s \times n_p}$ (n_p is the total number of histogram bins of all weeks under consideration after removing insignificant bins) by the probability density matrix formed by stacking probability density vectors of different time intervals and stations along two dimensions. The Gaussian Mixture Model (GMM) is employed to perform the station clustering task. Given the training dataset and a GMM configuration (w.r.t. number of Gaussian components, type of covariance matrices), model parameters (i.e., mean vectors, covariance matrices, and mixture weights) are estimated to maximize the GMM likelihood. Note that each component density is a n_p-variate Gaussian function in this model. The problem can be addressed via the expectation-maximization algorithm proposed in Reynolds et al. (2009). In terms of hyper-parameters, while they are often determined by the amount of data, one can also make a decision based on the Bayesian information criterion (BIC) or the Akaike Information Criterion (AIC). In statistics, these two criteria are used for model selection, particularly to avoid the overfitting issue caused by adding too many parameters into the model, which is achieved by introducing a penalty term for the number of parameters in the model. The present clustering model reveals the heterogeneous impacts of fare-free policies among PT stations and provides labels to the classification model introduced in the following.

3.3. Station classification using spatiotemporal attributes

Many factors can lead to the difference in the response of PT stations to PT policies, including static features such as the network structure, and dynamic features such as activity patterns in the station vicinity. Table 1 summarizes some typical influencing factors together with their representative features. While the clustering model can identify different changing patterns of crowding patterns at PT stations due to the implementation of fare-free policies, we move one step forward by training a classification model to identify the significantly correlated factors with the changing patterns. In this study, we primarily consider the four categories of features listed in Table 1.

Among others, "Location & Context" features are able to capture the local economic situation, communities' life culture, and development pattern to some extent (Fan et al., 2022). U-Bahn and S-Bahn in Germany refer to urban subway and city rapid railway, respectively. Other categorizations in the context of other countries are also applicable. Similarly, "Nearby activity" features represent the land use, demand production and attraction of the adjacent area around the PT station. Apart from the statistics of POIs, the busyness data of POIs are also valuable for modeling neighborhood activity patterns. As per Vongvanich et al. (2023), the popularity trends of the neighboring POIs enable a satisfactory prediction of PT station demand in real time even through simple linear regression models. However, due to a lack of related data, they are not included in our model. "Crowding pattern" features provide information

Table 1
Features considered in the classification model.

Feature categories	Features
Location & Context	Dummy variable for cities and station types (U-Bahn, S-Bahn)
Nearby activity	Counts of different types of POIs around the station, population
Crowding pattern	The crowding pattern of the station for one week (before policy implementation)
Network structure	Features used in complex network analysis: # edges, # nodes, average node degree, etc.

related to the demand pattern of the respective PT stations under normal operation conditions. They are extracted from the historical busyness data collected before implementing new pricing methods. "Network structure" features can, to some extent, capture the traffic conditions of local private transport, which are very important in PT-related assessments given the strong inter-correlation between private and public transport.

Further, note that "Nearby activity" features might be highly correlated. It follows that dimension reduction techniques can be applied to reduce the number of features and preserve most information contained in the original features at the same time. This is also applicable to the "Crowding pattern" features. In this study, we apply factor analysis to achieve this. Factor analysis also serves as an exploratory analytical instrument, and its application is deemed justifiable to the extent that the resulting factors can be meaningfully interpreted.

We apply the LightGBM to perform the classification task. LightGBM is an improved version of the gradient boosting decision tree model that has gained popularity in recent years due to its exceptional performance. Traditional decision tree models classify data into different categories using "if-else" decision rules that partition the feature space into subspaces. However, a single decision tree can overfit easily and doesn't generalize well to new data. Gradient boosting is an ensemble technique that improves decision tree performance by training multiple trees sequentially, with each tree trained to correct the errors of previous trees (Bishop, 2006). LightGBM further improves gradient boosting by introducing gradient-based one-side sampling and exclusive feature bundling techniques (Ke et al., 2017). The former technique selects data points with the largest gradients during training, focusing on the most difficult samples. Exclusive feature bundling groups relate features together to reduce dimensionality and, thus, improve model efficiency. These unique techniques have made LightGBM one of the most competitive machine learning models for a wide range of applications in both academia and industry (Shwartz-Ziv and Armon, 2022; Bojer and Meldgaard, 2021).

A sequential feature selection is used to find the best combination of features for classification. A 10-fold cross-validation is also conducted to evaluate the performance of the LightGBM classifier on the busyness-based station types classification task.

4. Case-study and data

4.1. Study area

In 2022, the German federal government introduced the so-called "9-EUR Ticket" in response to the escalating fuel and energy costs resulting from the geopolitical crisis in Ukraine (Loder et al., 2023). The primary objective of the measure was to encourage commuters to shift from private vehicles to PT by providing them with a monthly flat rate of 9 EUR, which enabled them to travel on all regional, local, and urban PT services, except for long-distance passenger services such as Intercity Express, from June 1 to August 31, 2022. Other than as a pricing intervention, it was also a nationwide real-world experiment in terms of travel behavior and transport policy (Loder et al., 2023). This also distinguishes it from other similar interventions implemented within a relatively small region or for a specific user group, such as the experiment presented in De Witte et al. (2006) in which free PT was only introduced to the students of Flemish universities in Brussels. It provides an unprecedented opportunity for understanding the impact of ticket schemes with substantial discounts on PT operations.

To provide a clearer picture of the extent of PT price reduction during the 9-EUR period, it is beneficial to introduce the pre-existing PT pricing strategy in Germany. Taking the city of Munich, the largest city in southern Germany, as an example, the entire city is divided into 13 tariff zones, denoted as Zone M, Zone 1 to Zone 12. Zone M covers the entire Munich city area, while Zone 1 to Zone 12 cover the districts belonging to the München Verkehrs Verbund (MVV, the PT operator in Munich) area and are numbered to drain outward from the city area. Monthly PT tickets are distinguished among zones, enabling travel within designated areas. The previous monthly Zone tickets in Munich cost from about 50 EUR (Zone M) to about 300 EUR (Zone M - 12)[1] for the public except some specific population groups (e.g., students, children, aging people). The Zone M ticket, being the cheapest, covers travel within the city of Munich. The price increases with the Zone index. For instance, Zone M - 5 is more expensive than Zone M - 1. Note, the PT tickets in Germany are typically city-specific, and the area of validity is confined to the respective cities. In contrast, the 9-EUR ticket is valid in all cities of Germany and costs much less. Furthermore, it is also valid for regional trains, independent of the system of city PT operators, and would otherwise require a separate payment of around tens of EUR for a single trip. This significant reduction in cost and expanded coverage enhances the attractiveness of the 9-EUR ticket.

Figure 3 depicts the PT mobility trends during the 9-EUR implementation period extracted from the COVID-19 Community Mobility Reports[2] collated by Google. The reports plotted movement trends (based on visits to places and length of stay) over time by geographic regions across different categories of places, including PT stations. Mobility trends are calculated by comparing with a baseline (indicating the zero value in Figure 3). The baseline is calculated as the median value, for the corresponding day of the week, during 3 Jan – 6 Feb 2020 (before the COVID-19 pandemic). Therefore, comparing the mobility trends of different regions/cities in Germany pre, during, and post 9-EUR can provide some insights into the heterogeneity of the influence of 9-EUR on these regions/cities. Figure 3 shows that during 9-EUR policy, indeed stimulated PT use in the entire country on average was observed; the PT mobility trends increased during the 9-EUR period and decreased afterward. Besides the 9-Euro policy, seasonal factors can also play an important role in leading to these changes. However, when focusing on specific cities/regions, e.g., no increase was observed in Berlin. In contrast, Hessen, a state of Germany, saw a rising trend similar to the one for Germany, but the former's trend was maintained after the policy as compared to the latter's noticeable fall in the trend. These facts emphasize the necessity of evaluating FFPT region-wise. Moreover, given the distinction in PT station characteristics within a region/city, it is also important to evaluate the policy's influence at a finer spatial resolution, e.g., at the station level. This corresponds with the research questions we proposed in Section 2.

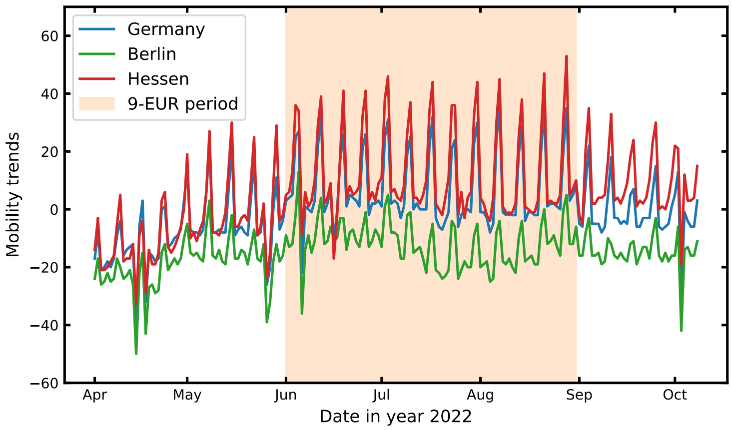


Figure 3: Google community mobility report of Germany.

4.2. Data collection and description

Google's Popular time (GPT) graph (Google, 2023) displays the level of busyness of a Point of Interest (POI) at different times of the day relative to its busiest hour of the week. The historical busyness of a POI is measured on a relative scale of 0 to 100, with 100 indicating the busiest hour. The live busyness is computed based on the relative busyness compared to the corresponding historical period. The popular time graph for POIs is publicly accessible on Google Maps (Google Maps, 2023). Figure 4 provides an example of the GPT graph for the Munich Central Station on April 7, 2023. The blue bars show the historical busyness levels hourly. The red bar represents the live busyness level when it is retrieved. In this example, the Munich Central Station is more crowding at the retrieving moment compared to the same period in the past few weeks.

[1] https://www.bahn.de/
[2] See https://www.google.com/covid19/mobility/ for more details. Access on: 19.08.2023

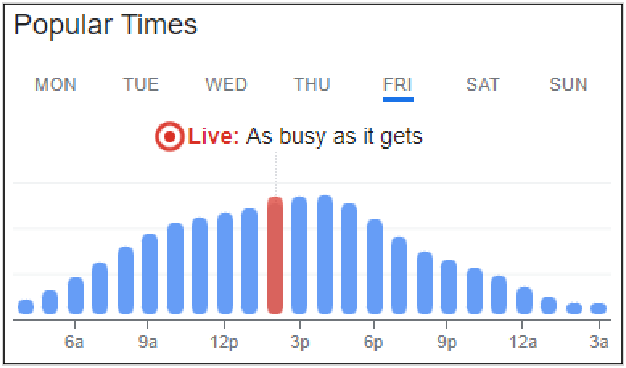


Figure 4: Example of GPT graph showing historical popularity (blue bars) and live popularity (red bar)

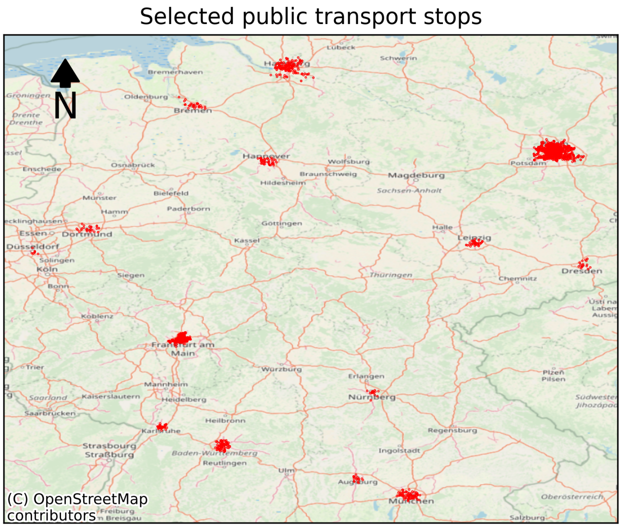


Figure 5: Location of selected PT stations.

The use of GPT graph as a crowdsensing data source for measuring the busyness of POIs, such as PT stations, has been recognized as a valuable approach for exploring urban dynamics (e.g., patterns of urban activity) (Vitello et al., 2023; Niu and Silva, 2020). Given the high penetration of Google services, this publicly accessible proxy for crowding is available at a sheer scale in any city (Vitello et al., 2023; Möhring et al., 2021). GPT data provide information on the busyness of locations based on user check-ins, offering significant potential for opportunistic applications benefiting from their scale, nature of information, and real-time availability.

Therefore, in this study, GPT data is collected and used to measure crowding/demand/busyness patterns in PT stations in Germany at different phases of the 9-EUR ticket policy. Specifically, the GPT data of 2,134 railway stations, including U-Bahn (urban subway) and S-Bahn (city rapid railway) stations, were collected in two-hour intervals before (May 26 – May 31), during (Jun 1 – Aug 31) and after (Oct 9 – Dec 9) the 9-EUR implementation. Figure 5 shows the location distribution of these stations across Germany. The data collected before 9-EUR are used to extract the "Crowding pattern" features for classification, while the data collected during and after the 9-EUR period are used in the histogram method for clustering to identify changing patterns.

Regarding the other features adopted in the classification model, the population is estimated using the population density map of Germany in 2019 (HDX, 2023). For the nearest 5 and 10 stations, we calculate the mean and standard deviation of the population densities to capture the population characteristics. Network neighbors of a given station are defined as the closest stations to it in the entire PT network. These features capture the connectivity characteristics regarding the busyness level within the adjacent area of the neighboring stations. POI statistics are collected from

OpenStreetMaps (84 types are reserved after a POI counts filter). Both the population and POI statistics are estimated and counted within a square are with a side length of 1 km. "Network structure" features are computed using the Python package OSMnx (Boeing, 2017).

4.3. Data analysis and preprocessing

Considering that some stations may have very sparse entries, the collected GPT dataset is cleaned through three filters: (i) Historical busyness cannot be zero at all time intervals for an entire week; (ii) Historical busyness should be updated at least once a week; (iii) Live data of at least one time interval should be non-null in all periods before, during, and after the implementation of 9-EUR. Finally, 293 stations were preserved. Figure 6 shows the average crowding pattern of all stations on weekdays, Saturdays, and Sundays extracted from historical and live datasets during 9-EUR. They are plotted with a two-hour interval which is the data collection frequency. The x-axis is the time of the day and the y-axis is the busyness level. In addition to the average values, the standard deviations across different stations are also provided.

In terms of historical average patterns, two demand peaks are observed on weekdays, including a morning peak at 8 AM and an afternoon peak at 3 PM, while only one peak shows up on Saturdays and Sundays at around 3 PM. It means the daily crowding patterns are different between weekdays and weekends. Furthermore, we can see that the busyness of stations is higher on weekdays than on weekends during the daytime. However, there is a higher busyness level after 8 PM on Saturday, possibly due to more nightlife activities on Saturday evenings. Regarding live average patterns, an apparent lift shows up in Saturday's pattern and Sunday's pattern in all time intervals across the day, compared to the historical ones. It means the implementation of 9-EUR significantly stimulates travel on the weekends. The possible reason might be more people making leisure trips (e.g., exploring the natural landscapes near their places on the weekends, visiting friends) since they can use the regional trains for free, benefiting from the 9-EUR policy. On the contrary, on the weekdays, while the live pattern keeps at a similar level to the historical pattern during the daytime, it presents a higher popularity in the evening. More people opt for PT to perform evening trips and activities. Another interesting point is the standard deviations of live patterns across different stations are much greater than those of historical patterns, implying that the influence of 9-EUR has a nature of inter-station variability. This lends support to the development of the three-step framework and the necessity of integrating station clustering to accurately uncover the policy's influence.

Further, we notice the common missing value issue of live busyness in GPT data; when there are insufficient visitors, the live busyness may not be available. Therefore, it is necessary to apply a data imputation method to complete the live busyness dataset. For ease of reading, its detailed implementation is omitted here and can be found in Appendix A.

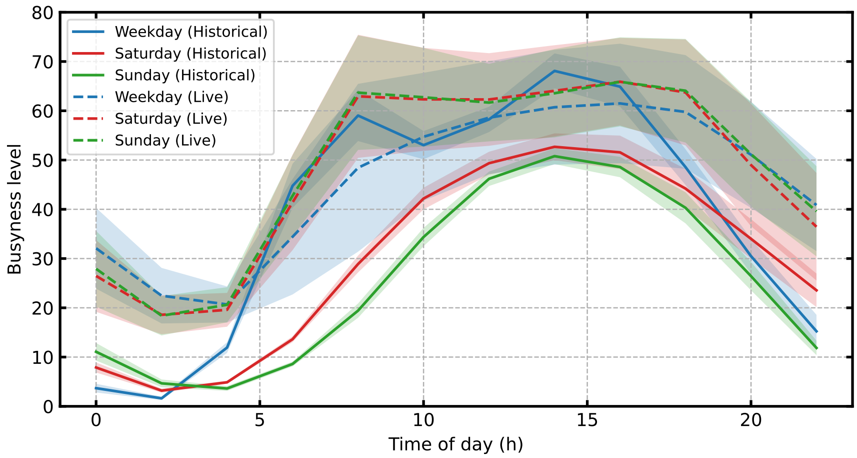


Figure 6: Average patterns on weekdays, Saturdays, and Sundays (during 9-EUR).

5. Results

Applying the proposed three-step busyness-based evaluation framework, in this section, we analyze the impact of the 9-EUR ticket by using the GPT data of PT stations. First, we present the results of station clustering and characterize the identified busyness-based station types. Then, factors extracted from “Crowding pattern” features and other highly correlated features are expounded on. Finally, significant features influencing the response of stations to the 9-EUR ticket are explained.

5.1. PT station clusters characterization and analysis

In the histogram method, we set all bins to the same width $\Delta_m = 10, \forall m$. Since the busyness deviation will be in the range of $[-100, 100]$ (the GPT busyness is in $[0, 100]$), $M = 20$ bins are specified in histograms. Recall that the probability densities of deviation values within these bins constitute the set of features for station clustering. According to Equation 1, after specifying the same l_t and Δ_m for all bins, the probability density is proportional to the counts of deviation values, i.e., $p_m \propto K_m, \forall m$. Thus, we use K_m rather than p_m in the following experiments for a more intuitive understanding. Additionally, the variables representing the range of $[-100, -60)$ and $[80, 100]$ are removed due to their limited capability of distinguishing station samples, i.e., almost all stations have the same value. The rest variables more or less follow a bell-shaped distribution, thereby supporting the application of GMM to the station clustering task. Figure 7 compares the performance of GMM models with different hyper-parameters with respect to the number of components and the type of covariance matrices. A grid search is performed to find the best hyper-parameters based on BIC. A lower BIC is preferable. The top subplot presents the results of considering all histogram variables, while the bottom subplot presents that of after removing the variables representing $[-100, -60)$ and $[80, 100]$. Recall that BIC is used to avoid overfitting by incorporating a penalty term for the number of parameters in the model, which interprets why the GMM models with full or tied covariance matrices (non-empty values in all elements) for each component perform worse on it than the other two types of GMMs. Furthermore, we can see that removing the variables incapable of distinguishing stations can improve the robustness of the GMM with diagonal covariance to the number of components. Finally, focusing on the bottom case, although the GMM models with spherical and diagonal covariance matrices show similar performance regardless of the number of components, the GMM with 3 components and diagonal covariance matrices leading to the smallest BIC is selected for interpretation purposes.

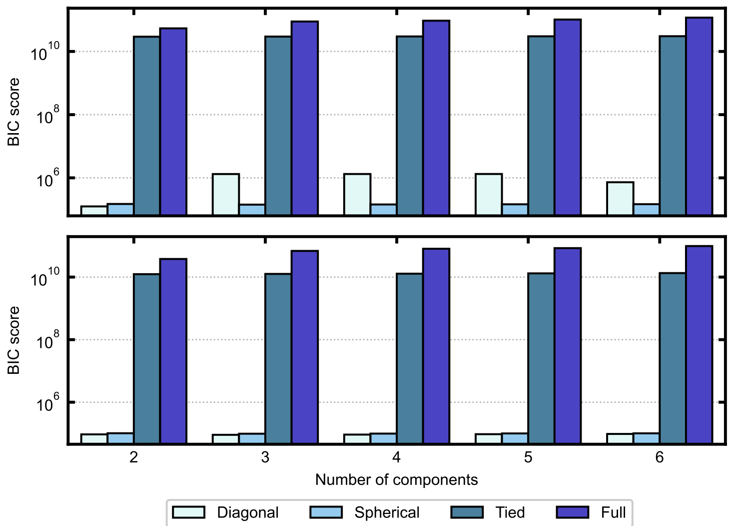


Figure 7: GMM models comparison based on BIC. (top) Considering all variables resulting from the histogram method; (bottom) After removing the variables representing [-100,-60) ∪ [80,100].

After reordering, the three categories of stations are indicated in Figure 8. The x-axis is the bins indices, with every 14 bins constituting a histogram representing one week of data because we have removed the six low-impact intervals mentioned above, i.e., $[-100, -60)$ and $[80, 100]$. We have 14 weeks for the 9-EUR period and 9 weeks for the post 9-EUR period, which are indicated with different colors in the figure. The y-axis represents different PT

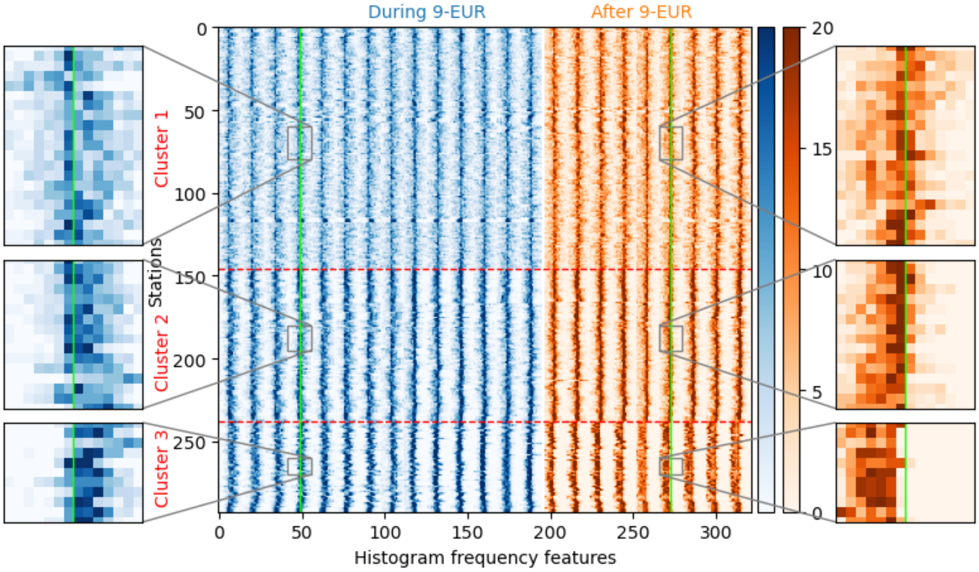
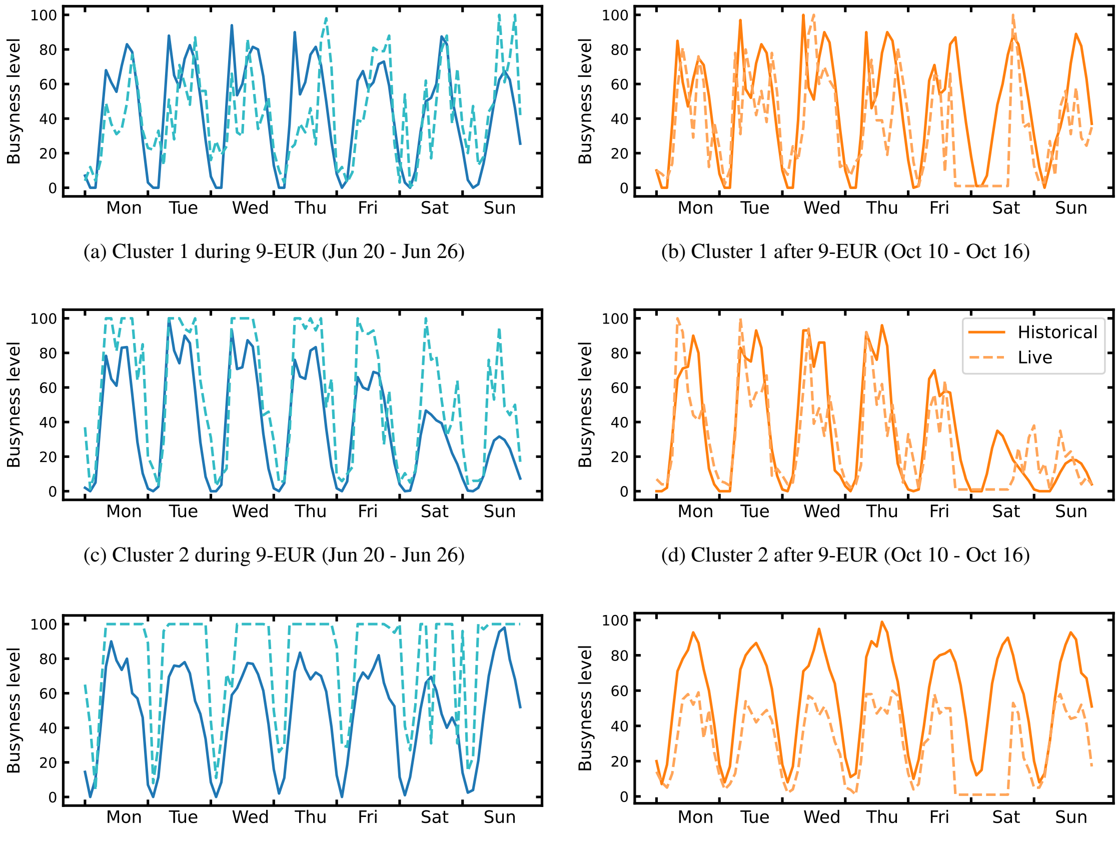


Figure 8: Station clusters based on busyness deviations. (blue) During 9-EUR; (orange) After 9-EUR. Green vertical lines indicate the centers of the respective histograms.



(a) Cluster 1 during 9-EUR (Jun 20 - Jun 26)

(b) Cluster 1 after 9-EUR (Oct 10 - Oct 16)

(c) Cluster 2 during 9-EUR (Jun 20 - Jun 26)

(d) Cluster 2 after 9-EUR (Oct 10 - Oct 16)

(e) Cluster 3 during 9-EUR (Jun 20 - Jun 26)

(f) Cluster 3 after 9-EUR (Oct 10 - Oct 16)

Figure 9: GPT time series samples of the identified busyness-based station types.

stations. Red horizontal dashed lines separate different categories of stations. Zoom-in views of some stations of each category are also provided to better illustrate the difference in their crowding pattern changes. A clear distinction can be observed between every two categories. To be specific, we characterize stations based on the heterogeneous impacts of 9-EUR on their crowding patterns as (i) Cluster 1 – unaffected stations (146 stations). Their demand patterns are almost not affected by the implementation of the 9-EUR ticket policy. Moreover, their crowding pattern deviations are more stochastic than the other types of stations; (ii) Cluster 2 – mildly stimulated stations (92 stations). These stations' demand increase during 9-EUR and recover to the original demand level slowly or keeps a similar level afterward; (iii) Cluster 3 – intensely stimulated stations (55 stations). These stations' demand increases significantly during 9-EUR and reduces immediately afterward. This explains why contradictory results were obtained in Thøgersen (2009) and Friman et al. (2019) regarding the short-term effect of temporary FFPT, i.e., temporary FFPT doubled the PT usage against temporary FFPT showed less influence in the short term. One of the possible reasons could be that the study area of Thøgersen (2009) is dominated by PT stations belonging to cluster 2 and cluster 3, while in the study area that of Friman et al. (2019), stations are relatively unaffected. Figure 9 provides the GPT time series samples during and after 9-EUR for each station type to better illustrate the heterogeneity among different types of stations. These time series demonstrate similar phenomena to those described above based on histogram-based variables in terms of the changes in crowding patterns, which, on the other hand, reflect the correctness of the characterization of stations.

Figure 10 provides the composition of stations in five major cities in Germany that have more than 20 stations in the cleaned set of stations. The cities are ordered based on the number of "unaffected" stations (cluster 1). We can see that the composition is very different from city to city. Specifically, most stations in Stuttgart and Berlin are "unaffected", while more than half of stations in Hamburg and Frankfurt are "mildly stimulated" and "intensely stimulated" respectively. This is consistent with the mobility trends in Figure 3. Namely, the PT use in Berlin did not change much due to 9-EUR, while that in Frankfurt (a city of the Hessen state) increased significantly. On the other hand, Munich has nearly balanced numbers of "unaffected" and "mildly stimulated" stations. Moreover, 87.3% of "intensely stimulated" stations (55 in total) are from Frankfurt.

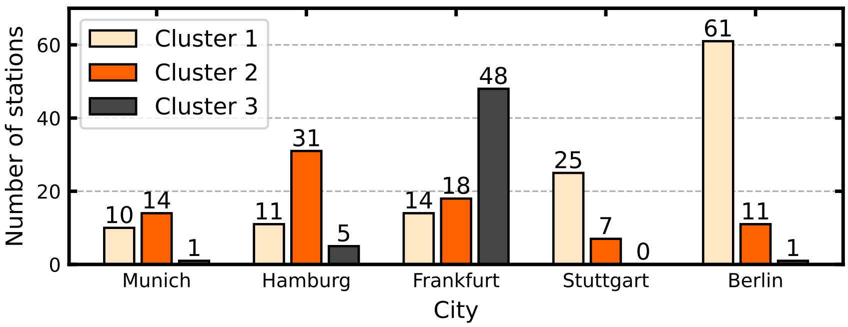


Figure 10: Station types distribution in major cities ($n_s \geq 20$).

5.2. Latent variables extraction based on factor analysis

Factor analysis is conducted on crowding pattern features to extract significant factors that can describe the demand patterns of PT stations. The VARIMAX (Kaiser, 1958) rotation method is used. For the purpose of interpretation, only factors that can explain 2% or more of the total variance are considered. Notably, the justification for employing factor analysis rests on the premise of high communality in the features. The communality represents the percentage of the variance explained by all factors jointly, which can be calculated as the sum of squared factor loadings of the given feature across all factors. Consequently, for interpretation, only features with an absolute factor loading greater than a threshold (here, 0.5) are considered for characterizing the factor. In other words, combining these two constraints, only factors explaining more than 2% of the total variance and showing a factor loading of 0.5 or greater in at least one variable are kept. This treatment leads to 11 factors from 84 crowding features.

Figure 11 depicts the factor loadings of features in different factors. The x-axis shows different periods within a week, while the y-axis indicates the factors which satisfy the constraints. Darker colors indicate a large absolute value of the corresponding factor loading. We denote these factors extracted from crowding pattern features by CP0 to CP10.

Clear distinctions do exist among them. In particular, according to the factor loadings, we can interpret these factors as representatives for different times of the day. We characterize these factors as listed in Table 2. From these factors, we can also see a clear distinction between weekdays, Saturdays, and Sundays, consistent with the findings in Figure 6. Therefore, as per the time intervals with significant factor loadings, we define three basic types of factors: weekday-based, Saturday-based, and Sunday-based factors. All factors can be categorized under at least one category of them. Then, the factors belonging to both Saturday-based and Sunday-based Saturday- and Sunday-based factors would be defined as weekend-based factors, while those relating to all days of the week would be daily factors. However, one should pay attention to the difference in loadings between weekdays, Saturday, and Sunday of these composite factors. More specifically, such differences can be observed in factors CP0, CP1, CP2, CP5, and CP8, as indicated in the Remark column of Table 2.

Essentially, these factors represent the demand patterns at different periods of the week and are latent variables contained in the crowding pattern features. For instance, CP0, associated with the period 14:00 - 22:00, represents the demand pattern at PT stations in the afternoons and evenings. The negative loadings in these intervals say that the busyness level at PT stations during this time period is related in the opposite direction from the factor. The variance each factor can explain is also provided in Table 2.

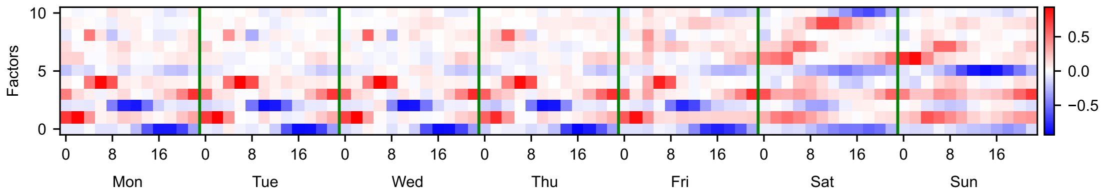


Figure 11: Factors extracted from crowding pattern features.

Table 2
Factors extracted from crowding pattern features.

Factor	Type	Time	Interpretation	Loadings	Var. exp.	Remark
CP0	Daily	14-22	Afternoon and evening	Negative	17.9%	Weekend's pattern is more spread out, and Sunday's pattern is weak
CP1	Daily	0-6	Early morning	Positive	12.3%	Two-hour lag in weekend
CP2	Weekday-based	10-16	Weekday lunchtime	Negative	12.3%	Weekend shows similar patterns but weak
CP3	Daily	22-2	Mid-night	Positive	11%	
CP4	Weekday-based	4-10	Weekday breakfast time	Positive	9.2%	
CP5	Sunday-based	10-22	Sunday after morning	Negative	7%	Saturday shows similar patterns but weak
CP6	Weekend-based	0-6	Weekend early morning	Positive	4%	
CP7	Weekend-based	6-10	Weekend breakfast time	Positive	2.4%	
CP8	Weekday-based	4-6	Weekday before breakfast	Positive	2.2%	Pattern is weak
CP9	Saturday-based	10-14	Saturday lunchtime	Positive	2.1%	
CP10	Saturday-based	16-22	Saturday evening	Negative	2%	

Further, in the set of rest features, factor analysis is also conducted on the highly correlated features, i.e., features with an absolute correlation greater than 0.6, to reduce the feature space. These features all belong to one of the categories of POI statistics, population densities, and network features. Following the same procedure, we obtain 9 factors. The factor loadings are provided in Figure 12. The POI features are sorted in ascending order based on their counts from left to right. We denote these factors by PPN0 to PPN8 with "PPN" representing "POI statistics", "Population" and "Network" respectively.

Following one-to-one correspondence, we can interpret PPN factors as in Table 3. Due to the large amount of POI features and the difficulty of explaining their intrinsic correlation, we do not explicitly categorize different types of POIs and finely interpret the associated latent variables. Instead, we only distinguish the POI-associated factors from the sign of factor loadings, resulting in three POI-based factors with positive loadings (PPN0, PPN3 and PPN5) and two POI-based factors with negative loadings (PPN1 and PPN8). For reference, the POIs with loadings greater than 0.5 and total counts of more than 1,000 are also provided in Table 3. On the contrary, for those factors emphasizing population or network features, we can provide clear interpretations based on the nature of associated features. More specifically, the factor analysis leads to two population-based and two network-based latent variables, respectively. The two population-based factors indicate population and population variance, respectively, given that the former is correlated with the population densities of the vicinity of the station and its network neighbors, while the latter is correlated with the variance in their population densities. The two network-based factors indicate the network summation describing the counts or summations of network elements (nodes, edges, intersections, etc.), and the network average describing the average of network elements. Note that while the factor loadings of the two population-based factors have opposite signs, both network-based factors have positive loadings on the associated features.

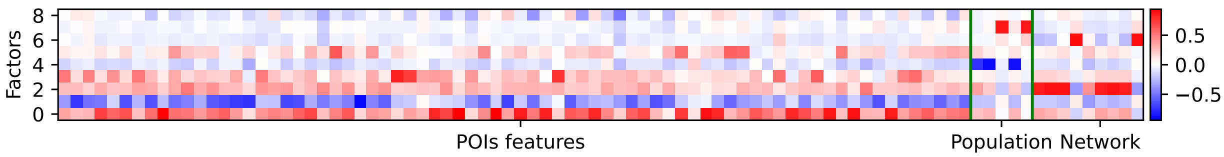


Figure 12: Factors extracted from highly correlated features (i.e., POI features, population features, and network features).

Table 3
Factors extracted from highly correlated features.

Factor	Interpretation	Loadings	Var. exp.	Remark
PPN0	POI positive (1)	Positive	25.5%	Restaurant, vending machine, cafe, fast food, clothes, hairdresser, post box, bakery, bar, ATM, telephone, pharmacy, fountain, bank, jewelry, shoes
PPN1	POI negative (1)	Negative	17.1%	Bicycle parking, waste basket, cafe, fast food, hairdresser, bakery, bar, pub, kindergarten, supermarket, ATM, convenience, bicycle rental, florist
PPN2	Network summation	Positive	11.6%	Significant features are the counts or summations of network elements, including # nodes, # edges, # intersections, # street segments, total edge length, total street length
PPN3	POI positive (2)	Positive	7.4%	bench, clock, toilets
PPN4	Population	Negative	4.2%	Significant features include the population within the one-kilometer circle of the station, the population within the one-kilometer circle of the station's nearest 5 and 10 network neighbors
PPN5	POI positive (3)	Positive	3.9%	Beauty
PPN6	Network average	Positive	2.6%	The two significant features are the average edge length and the average street length
PPN7	Population variance	Positive	2.1%	The two significant features are the standard deviations of population densities of the nearest 5 and 10 station neighbors, respectively
PPN8	POI negative (2)	Negative	2%	Place of worship

5.3. Analysis of features affecting the impact of 9-EUR on stations' crowding patterns

LightGBM is applied to infer the relationship between the selected station features (including the latent variables resulting from factor analysis) and the impact on crowding patterns. Based on the cross-validation, the model achieves a weighted F1 score of 0.704 and a balanced accuracy of 0.701. With the information gain as the importance measurement, the significant features with an information gain greater than 100 are given in Figure 13. As can be seen, the importance of the dummy variable indicating whether the station is located in Frankfurt is more than two times of most of the rest of variables. Referring back to Figure 10, we found that 60% of Frankfurt's stations (48 out of 80) are "intensely stimulated" stations, composing 87.3% of this station type. It is thus understandable that the dummy variable for Frankfurt attains the highest importance. Similarly, the variable indicating whether the station is located in Berlin illustrates a significant importance level since 83.6% of its PT stations belong to "unaffected" stations. Additionally, the second most important to the ninth most important variables (8 out of 13) are latent variables identified using factor analysis introduced in Section 5.2. Four of them are extracted from crowding pattern features, representing the afternoon and evening pattern (CP0), the pattern before breakfast on weekdays (CP8), the early morning pattern (CP1), and the evening pattern on Saturday (CP10), respectively. This means the influence of 9-EUR on a given PT station also has a strong relationship with its daily crowding patterns. The other four factors are PPN factors, representing demand patterns of the POIs near the station (PPN3 and PPN8), population (PPN4) and population variance (PPN7), respectively. The "*Self loop*" variable (i.e., percent of edges that are self-loops) is one "Network structure" variable described in Table 1. In other words, all feature categories listed in Table 1 show up in Figure 13, reflecting the necessity of considering the spatiotemporal attributes of PT stations when assessing the influence of FFPT on the crowding patterns of PT stations. As such, by dedicating measures accurately responding to the crowding pattern changes for each station, one can improve the PTDM towards crowding reduction in the context of FFPT, thereby improving PT service quality and attractiveness. Moreover, 70% of prediction accuracy further validates the existence of a strong association between station characteristics and the heterogeneous impact of the 9-EUR ticket. Therefore, one can expect to improve the demand management in PT stations by implementing appropriate policies in other relevant sectors other than directly acting on the transportation sector.

On the other hand, the confusion matrix tells that most misclassifications happen between "unaffected" and "mildly stimulated" stations. Specifically, the true positive rates of the three classes are 0.713, 0.418, and 0.727, respectively. Notably, 48.2% of misclassifications predict C2 as C1. To further examine the factors behind the misclassification problem, Figure 14 presents the misclassified stations of each class. We can see that the busyness deviations of those stations are more stochastic along different weeks, and it is difficult for clustering models to distinguish their patterns. In other words, those stations are very probably located close to the decision boundaries of the GMM clustering model, especially the boundaries between C1 and C2, from the view of crowding pattern changes.

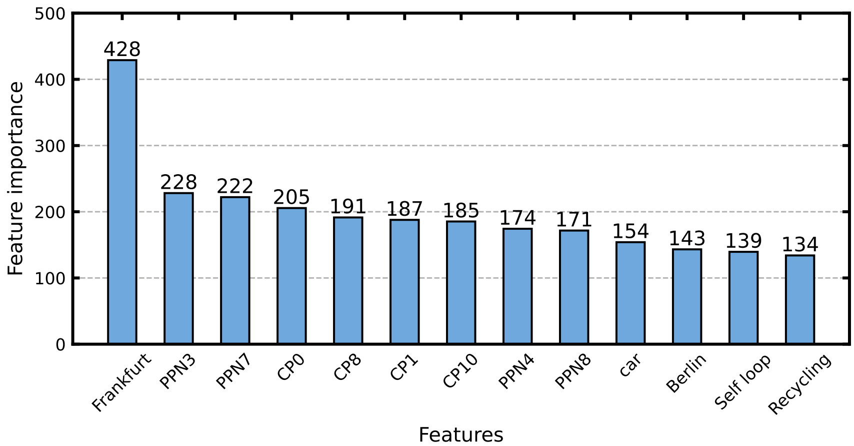


Figure 13: The most important features based on information gain (the information gain is greater than 100).

6. Policy Implications

The findings presented in this study offer valuable insights into the implications of FFPT at the station level and its association with station attributes. These insights bear the potential to enhance the formulation and implementation of nationwide FFPT policies. Due to FFPT-induced changes in demand, transit operators face a challenge to re-orient the supply accordingly. Thus, major policy implications of FFPT are covered under the theme of focused re-balancing of the services in short-term or depending on the duration of the FFPT. We provide examples of a few such aspects below:

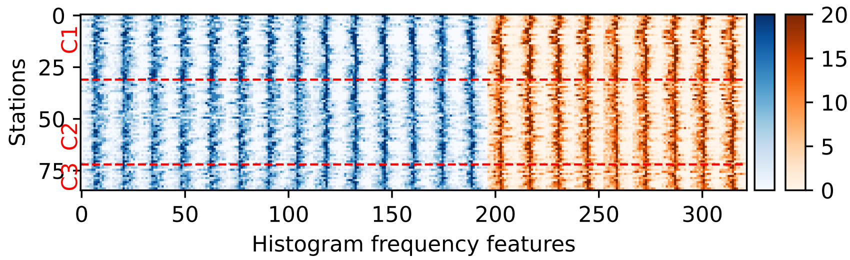


Figure 14: Busyness deviation of misclassified stations.

(1) The study provides empirical evidence of the heterogeneous fluctuations of station busyness stemming from (partially) FFPT policies across multiple PT stations. In other words, the degree of influence varies across different PT stations. Consequently, to improve the efficiency and efficacy of PT services, policymakers and PT operators need to re-balance travel demand and transit supply. For instance, they should explore short-term transit supply measures, such as the integration of skip-stop services into the operation of certain lines after identifying the potential station-level influence of FFPT. Skipping stations can reduce travel times and attract more passengers in general, but such measures can be counterproductive too (Gkiotsalitis and Cats, 2021a,b). These measures can also confuse passengers and impede certain passengers' journeys if not properly planned and implemented, therefore they should be implemented only after rigorous investigation of station selection and sufficiently informing the public such as via announcements, and mobile messages.

(2) An imperative step would involve re-optimizing the schedule plan of the entire PT network to mitigate the effects of crowding at the stations. To be specific, the provision of PT services should be re-balanced by transferring appropriate supply from "unaffected" to "mildly stimulated" and "intensely stimulated" stations. Moreover, the coordination and cooperation of metro and bus have been recognized as a promising approach to reducing passenger delay and improving the resilience of urban PT systems (Saliara, 2014; Jin et al., 2014; Wu et al., 2022). Due to the crowding pattern changes, the metro-bus coordination should also be optimized accordingly to improve the connectivity of the entire PT network.

(3) The knowledge of how PT stations respond to FFPT policies is very useful for the optimization of allocation or deployment of other transport services (such as taxis and shared mobility options) to complement public transport. By discerning the patterns of busyness shifts at PT stations, providers of shared mobility services can adjust the distribution of their fleets to ensure that shared mobility resources are optimally allocated near PT stations. As a complementary alternative for PT, these shared modes also provide a good last-mile solution to PT users. Simultaneously, the optimized allocation of shared modes augments the attractiveness of PT by offering commuters reliable, integrated mobility solutions (Van Kuijk et al., 2022). As the landscape of urban mobility becomes increasingly multifaceted, the symbiotic relationship between PT and shared mobility is poised to flourish. Therefore, policies that could promote the consolidation of this relationship are beneficial to the development of sustainable and green urban transportation systems.

(4) The number of passengers at stimulated PT stations may exceed its typical daily demand. Important implication thus also relates to the crowd management (Drury and Reicher, 2010) at the stimulated stations, especially at the critical or bottleneck entry/exit points to prevent incidents of crowd collapse or crowd crush. The bottleneck capacity not only restricts access to the PT services but also has the potential of controlling crowd densities on critical elements like PT platforms (Seer et al., 2008). The implementation of effective crowd management strategies at stations experiencing surges in demand holds the potential to improve the capacity and safety of PT services. Our methods and the data used can help proactively identify stations with high chances of these incidents, thereby facilitating protective interventions.

(5) The used data and methods provide a transferable way to analyze the effects of other short-term planned or unplanned special events. This holds special potential for locations where PT data is not available or systematically collected. The actual insights can be different depending on the socio-economic context of the study area but the

data and methods can help to accelerate the preliminary crowding investigations for a pro-active management of services at the PT stations. These methods can be further extended and validated with auxiliary and richer data from the public transport authorities or other data owners. For instance, it has been reported that the PT ridership can be precisely estimated by detecting the WiFi or Bluetooth signatures (e.g., Myrvoll et al., 2017). Therefore, combining the proposed methodology with such wireless data can be further utilized to deduce the absolute number of PT passengers at stations before and after policy implementation.

(6) This study underscores the strong relationship between the crowding patterns at a PT station and the activity patterns in the station vicinity. Therefore, POI construction and introduction need to be integrated into the PT network planning if one wants to improve the usage of specific PT stations. It is worth mentioning that different categories of POIs usually exert an impact on different time intervals of the day. For instance, bars will increase the travel demand in the evening. Accordingly, policymakers must align their choice of introduced POIs with the station's role within the network and the anticipated temporal patterns. Additionally, the influence variation across different cities underscores the importance of adapting nationwide FFPT policies to the unique context of each city, enabling its maximum efficacy to be realized.

(7) Stimulated PT stations reflect greater attractiveness under FFPT. This understanding holds substantial implications for businesses operating in proximity to these stations. Informed of the degree of stimulation of specific PT stations, business operators can adjust their operating strategies and schedules accordingly to improve revenue. For instance, consider a PT station that exhibits a noticeable surge in busyness levels during weekends while demonstrating relatively modest changes during weekdays. Restaurants nearby can strategically realign their operational hours, transitioning from weekday-focused schedules to weekend-oriented ones. This adaptive approach enables them to tap into the increased weekend demand to optimize revenue.

Overall, relevant policies can be enacted, with the knowledge of the influence of FFPT, to balance the supply and demand from multiple perspectives, including re-balancing the provision of PT services like (1) and (2), the connectivity to PT services in multi-modal transportation systems like (2) and (3), crowd management at PT stations like (4) and (5), and the business and goods provision in proximity to PT stations like (6) and (7).

7. Conclusions

This study has presented a three-step busyness-based evaluation framework to empirically analyze the impact of fare-free policies on public transport (PT) demand patterns, with a focus on crowding patterns at PT stations. Using the example of the 9-EUR ticket implemented in Germany in 2022, we have showcased the framework's effectiveness in estimating the varying sensitivity of PT stations to fare intervention and inferring the factors influencing the difference.

The clustering model has successfully identified three categories of stations based on their crowding pattern changes: (i) Unaffected stations (demand patterns are not affected by the intervention), (ii) Mildly stimulated stations (demand increases during the implementation period and recovers to the original state slowly or maintains a similar demand level afterward), (iii) Intensely stimulated stations (they are very sensitive to the intervention, as shown by the dramatically increasing demand during the implementation period and the sudden reduction afterward). It has revealed that most stations in Stuttgart and Berlin are "unaffected", while more than half of stations in Hamburg and Frankfurt are "mildly stimulated" and "intensely stimulated" respectively. Munich has balanced numbers of "unaffected" and "mildly stimulated" stations. Further, the classification model suggests that the station location, activity patterns, population variance within the area around the station and around its nearest neighbors in the network, and demand patterns in the afternoon and evening, and weekday early morning play a significant role in the crowding pattern changes of stations.

This study provides insights into the impact of fare-free PT pricing policies on travel behavior and PT demand patterns and adds to the knowledge of factors influencing stations' responses (crowding pattern changes) to these policies. The impact heterogeneity implies that transport operators and planners should review the operation of PT stations to cater to the additional demand. Of course, an increase in public transport mode share needs to be complemented by an increase in PT investments, but our analysis could help to highlight which stations should be prioritized. It could also further support the crowdedness management at stations and schedule optimization of PT lines. Some potential policies are also discussed. More importantly, the proposed methodology uses publicly available data sources and can be applied to other cities or countries, especially those where ridership data is not easily accessible.

However, the framework can still be further improved by integrating other relevant features, such as accessibility to the city center, depending on the cities. Future research could also consider including POI statistics around the areas of the station's network neighbors given the significance of their population densities (and variance) in the classifier. Additionally, in this study, we computed the population and the POI statistics near the station by considering a square area with a side length of 1 km and assigned equal weight to all elements, overlooking the fact that the demand for PT usually responds to a distance decay function. This should also be addressed in order to obtain more accurate estimations of the population and activities in the station vicinity. Another important limitation of this study is that the crowding patterns of other POIs near the station which could capture the activity patterns nearby are not considered and collected. Yet, these activity patterns are expected to be highly correlated to the crowding patterns of the respective stations since people usually travel to perform activities. Our ongoing work on the assessment of the Deutschland-Ticket (D-Ticket)[3] introduced by the German federal government subsequent to the "9-EUR Ticket" attempts to further explore the influence of these factors. Furthermore, while the histogram features extracted from the distribution of crowding pattern deviation can effectively identify PT stations with similar crowding pattern changes, it fails to capture the similarity in the temporal patterns of stations' busyness levels. Future works could explore the development of a method capable of combining both of these aspects.

Acknowledgements

This work was supported by the European Interest Group CONCERT-Japan DARUMA project (Grant Number: 01DR21010) funded by the German Federal Ministry of Education and Research (BMBF). This research was also partially funded by the German Research Foundation DFG (TRAMPA Project, Grant 415208373) and the German Federal Ministry for Economic Affairs and Climate Action (BMWK) (PANAMERA project, Grant Number: 19I21016F).

Appendix A Imputation performance evaluation

The completeness of busyness datasets is a necessity for the calculation of the crowding pattern changes. Due to the potential difference in busyness patterns during the collection periods during and after 9-EUR, live data imputation is performed separately on the dataset of the two periods. The missing rates during and after 9-EUR are 21% and 33%, respectively. Considering the imbalance issue of missing patterns in the dataset, i.e., some stations observe many missing values while observing none otherwise, we apply the Probabilistic Matrix Factorization (PMF) (Mnih and Salakhutdinov, 2007) to address this problem given its good performance on sparse and imbalanced datasets. By PMF, the live busyness matrix will be modeled as a product of two lower-rank matrices, i.e., latent station and time feature matrices. The missing entries in the matrix will be imputed by maximizing (the log of) a posteriori distribution over the station and time features.

To evaluate the performance of the imputation method, we artificially design two random missing scenarios with the same missing rates using the respective historical datasets. In these two synthetic scenarios, PMF leads to mean absolute percentage errors of 2.3% and 1.0%, respectively, indicating a good performance on this task. As an example, Figure 15 shows the busyness time series of the Jaczostr. station, Berlin, in the first eight weeks of 9-EUR, together with their imputed values. We can see that despite live data fluctuating more than historical data, it also follows a clear daily pattern and the imputed values fit quite well with the ground truth. Therefore, we believe that the busyness-based similarity metric computed using the imputed live data is reliable.

References

Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer.

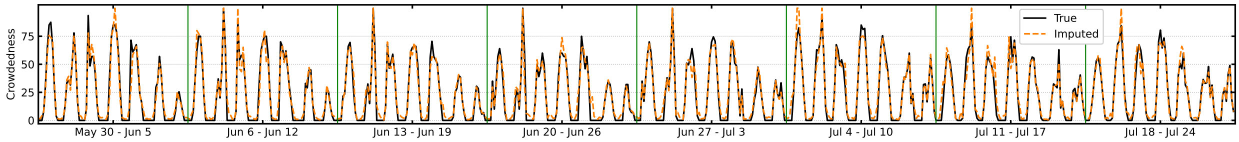
Boeing, G., 2017. Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. Computers, Environment and Urban Systems 65, 126–139.

Bojer, C.S., Meldgaard, J.P., 2021. Kaggle forecasting competitions: An overlooked learning opportunity. International Journal of Forecasting 37, 587–603.

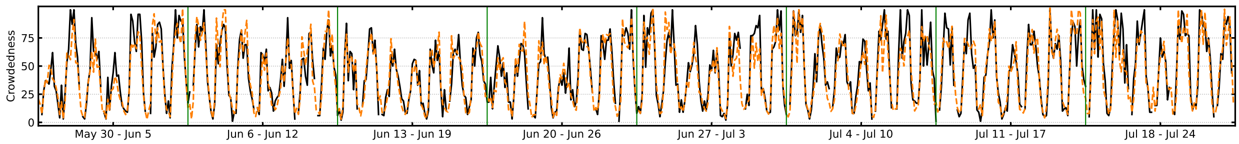
Börjesson, M., Hamilton, C.J., Näsman, P., Papaix, C., 2015. Factors driving public support for road congestion reduction policies: Congestion charging, free public transport and more roads in Stockholm, Helsinki and Lyon. Transportation Research Part A: Policy and Practice 78, 452–462.

Cai, Z., Wang, D., Chen, X.M., et al., 2017. A novel trip coverage index for transit accessibility assessment using mobile phone data. Journal of Advanced Transportation 2017.

[3] https://www.bahn.de/angebot/regio/deutschland-ticket



(a) Imputation result of historical data



(b) Imputation result of live data

Figure 15: Comparison of true and imputed popularity (Jaczostr. station, Berlin).

Capponi, A., Vitello, P., Fiandrino, C., Cantelmo, G., Kliazovich, D., Sorger, U., Bouvry, P., 2019. Crowdsensed Data Learning-Driven Prediction of Local Businesses Attractiveness in Smart Cities, in: 2019 IEEE Symposium on Computers and Communications (ISCC), IEEE. pp. 1–6.

Cats, O., Reimal, T., Susilo, Y., 2014. Public Transport Pricing Policy: Empirical Evidence from a Fare-Free Scheme in Tallinn, Estonia. Transportation Research Record: Journal of the Transportation Research Board 2415, 89–96.

Cats, O., Susilo, Y.O., Reimal, T., 2017. The prospects of fare-free public transport: Evidence from Tallinn. Transportation 44, 1083–1104.

Chaniotakis, E., Antoniou, C., Pereira, F.C., 2017. Enhancing resilience to disasters using social media, in: 2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), pp. 699–703.

Cools, M., Fabbro, Y., Bellemans, T., 2016. Free public transport: A socio-cognitive analysis. Transportation Research Part A: Policy and Practice 86, 96–107.

De Witte, A., Macharis, C., Lannoy, P., Polain, C., Steenberghen, T., Van De Walle, S., 2006. The impact of "free" public transport: The case of Brussels. Transportation Research Part A: Policy and Practice 40, 671–689.

De Witte, A., Macharis, C., Mairesse, O., 2008. How persuasive is 'free' public transport? Transport Policy 15, 216–224.

Drury, J., Reicher, S.D., 2010. Crowd control. Scientific American Mind 21, 58–65. URL: http://www.jstor.org/stable/24943180. full publication date: November/December 2010.

D'Silva, K., Noulas, A., Musolesi, M., Mascolo, C., Sklar, M., 2018. Predicting the temporal activity patterns of new venues. EPJ Data Science 7, 13.

Eliasson, J., Hultkrantz, L., Nerhagen, L., Rosqvist, L.S., 2009. The stockholm congestion–charging trial 2006: Overview of effects. Transportation Research Part A: Policy and Practice 43, 240–250.

Fan, Z., Zhang, F., Loo, B.P.Y., 2022. Rhythm of Transit Stations - Uncovering the Activity-Travel Dynamics of Transit-Oriented Development in the U.S. IEEE Transactions on Intelligent Transportation Systems 23, 12503–12517.

Fearnley, N., 2013. Free Fares Policies: Impact on Public Transport Mode Share and Other Transport Policy Goals. International Journal of Transportation 1, 75–90.

FitzRoy, F., Smith, I., 1998. Public transport demand in Freiburg: Why did patronage double in a decade? Transport Policy 5, 163–173.

Friman, M., Maier, R., Olsson, L.E., 2019. Applying a motivational stage-based approach in order to study a temporary free public transport intervention. Transport Policy 81, 173–183.

Fujii, S., Kitamura, R., 2003. What does a one-month free bus ticket do to habitual drivers? an experimental analysis of habit and attitude change. Transportation 30, 81–95.

Gkiotsalitis, K., Cats, O., 2021a. At-stop control measures in public transport: Literature review and research agenda. Transportation Research Part E: Logistics and Transportation Review 145, 102176.

Gkiotsalitis, K., Cats, O., 2021b. Public transport planning adaption under the COVID-19 pandemic crisis: Literature review of research needs and directions. Transport Reviews 41, 374–392.

Google, 2023. Popular times, wait times, and visit duration. URL: https://support.google.com/business/answer/6263531?hl=en. Accessed on 07.04.2023.

Google Maps, 2023. Google maps. URL: https://www.google.com/maps/. Accessed on 07.04.2023.

Halvorsen, A., Koutsopoulos, H.N., Lau, S., Au, T., Zhao, J., 2016. Reducing Subway Crowding: Analysis of an Off-Peak Discount Experiment in Hong Kong. Transportation Research Record: Journal of the Transportation Research Board 2544, 38–46.

Halvorsen, A., Koutsopoulos, H.N., Ma, Z., Zhao, J., 2020. Demand management of congested public transport systems: a conceptual framework and application using smart card data. Transportation 47, 2337–2365.

HDX, 2023. Germany: High resolution population density maps + demographic estimates. URL: https://data.humdata.org/dataset/germany-high-resolution-population-density-maps-demographic-estimates. Accessed on 20.04.2023.

Hensher, D.A., Rose, J.M., Collins, A.T., 2011. Identifying commuter preferences for existing modes and a proposed Metro in Sydney, Australia with special reference to crowding. Public Transport 3, 109–147.

Jin, J.G., Tang, L.C., Sun, L., Lee, D.H., 2014. Enhancing metro network resilience via localized integration with bus services. Transportation Research Part E: Logistics and Transportation Review 63, 17–30.

Kaiser, H.F., 1958. The varimax criterion for analytic rotation in factor analysis. Psychometrika 23, 187–200.

Kaufmann, V., 2017. Re-thinking mobility: Contemporary sociology. Routledge.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y., 2017. LightGBM: A highly efficient gradient boosting decision tree, in: Advances in Neural Information Processing Systems 30, Curran Associates, Inc.. pp. 3146–3154.

Kębłowski, W., 2020. Why (not) abolish fares? Exploring the global geography of fare-free public transport. Transportation 47, 2807–2835.

Li, Z., Hensher, D.A., 2011. Crowding and public transport: A review of willingness to pay evidence and its relevance in project appraisal. Transport Policy 18, 880–887.

Loder, A., Cantner, F., Adenaw, L., Bogenberger, K., 2023. The 9 EUR-Ticket -A nation-wide experiment: Germany -First findings, in: 102nd Annual Meeting of the Transportation Research Board (TRB 2023).

MacKenzie, D., Cho, H., 2020. Travel demand and emissions from driving dogs to dog parks. Transportation Research Record 2674, 291–296.

Mahajan, V., Cantelmo, G., Antoniou, C., 2021. Explaining demand patterns during covid-19 using opportunistic data: a case study of the city of munich. European Transport Research Review 13, 26.

Mnih, A., Salakhutdinov, R.R., 2007. Probabilistic Matrix Factorization, in: Advances in Neural Information Processing Systems, Curran Associates, Inc.

Myrvoll, T.A., Hakegard, J.E., Matsui, T., Septier, F., 2017. Counting public transport passenger using WiFi signatures of mobile devices, in: 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), IEEE. pp. 1–6.

Möhring, M., Keller, B., Schmidt, R., Dacko, S., 2021. Google Popular Times: towards a better understanding of tourist customer patronage behavior. Tourism Review 76, 533–569.

Niu, H., Silva, E.A., 2020. Crowdsourced Data Mining for Urban Activity: Review of Data Sources, Applications, and Methods. Journal of Urban Planning and Development 146, 04020007.

Pinelli, F., Nair, R., Calabrese, F., Berlingerio, M., Di Lorenzo, G., Sbodio, M.L., 2016. Data-driven transit network design from mobile phone trajectories. IEEE Transactions on Intelligent Transportation Systems 17, 1724–1733.

Pluntke, C., Prabhakar, B., 2013. INSINC: A Platform for Managing Peak Demand in Public Transit. SUSTAINABLE URBAN TRANSPORT .

Reynolds, D.A., et al., 2009. Gaussian mixture models. Encyclopedia of biometrics 741.

Saliara, K., 2014. Public Transport Integration: The Case Study of Thessaloniki, Greece. Transportation Research Procedia 4, 535–552.

Santiago-Iglesias, E., Carpio-Pinedo, J., Sun, W., García-Palomares, J.C., 2023. Frozen city: Analysing the disruption and resilience of urban activities during a heavy snowfall event using Google Popular Times. Urban Climate 51, 101644.

Seer, S., Bauer, D., Brandle, N., Ray, M., 2008. Estimating Pedestrian Movement Characteristics for Crowd Control at Public Transport Facilities, in: 2008 11th International IEEE Conference on Intelligent Transportation Systems, IEEE. pp. 742–747.

Shwartz-Ziv, R., Armon, A., 2022. Tabular data: Deep learning is not all you need. Information Fusion 81, 84–90.

Thøgersen, J., 2009. Promoting public transport as a subscription service: Effects of a free month travel card. Transport Policy 16, 335–343.

Timokhin, S., Sadrani, M., Antoniou, C., 2020. Predicting venue popularity using crowd-sourced and passive sensor data. Smart Cities 3, 818–841.

Tse, R., Zhang, L.F., Lei, P., Pau, G., 2018. Social Network Based Crowd Sensing for Intelligent Transportation and Climate Applications. Mobile Networks and Applications 23, 177–183.

Van Kuijk, R.J., De Almeida Correia, G.H., Van Oort, N., Van Arem, B., 2022. Preferences for first and last mile shared mobility between stops and activity locations: A case study of local public transport users in Utrecht, the Netherlands. Transportation Research Part A: Policy and Practice 166, 285–306.

Vitello, P., Fiandrino, C., Connors, R.D., Viti, F., 2023. TransitCrowd: Estimating Subway Stations Demand with Mobile Crowdsensing Data. preprint. In Review.

Vongvanich, T., Sun, W., Schmöcker, J.D., 2023. Explaining and Predicting Station Demand Patterns Using Google Popular Times Data. Data Science for Transportation 5, 10.

Wallimann, H., Blättler, K., Von Arx, W., 2023. Do price reductions attract customers in urban public transport? A synthetic control approach. Transportation Research Part A: Policy and Practice 173, 103700.

Wu, Z., Zhao, X., Xiao, L., Wang, P., 2022. A coordinated bus bridging and metro short turning model in response to urban metro disruptions. Transportation Safety and Environment 4, tdac003.