Journal of Transport Geography 136 (2026) 104808

Contents lists available at ScienceDirect

Journal of Transport Geography

journal homepage: www.elsevier.com/locate/jtrangeo

Following the footprints of visitors: Spatiotemporal public transportation profiles using smart card data

Anastasios Skoufas [a,*], Matej Cebecauer [a], Wilco Burghout [a], Erik Jenelius [a], Oded Cats [a,b]

[a] Department of Civil and Architectural Engineering, KTH Royal Institute of Technology, Stockholm, SE-10044, Sweden
[b] Department of Transport & Planning, Delft University of Technology, Delft, 2628 CN, The Netherlands

ARTICLE INFO

Keywords:
Visitors
Smart card data
Public transportation
Visitor profiles
Clustering analysis
Tourism

ABSTRACT

Visitors' induced seasonal demand spikes make supply dimensioning a challenging task. Smart card data have been widely used for exploring segmented passenger travel behavior; however, their potential in the tourism context remains underexplored. In this study, we portray distinct temporal and spatial visitor profiles in the Stockholm region by means of smart card data analytics. Empirical understanding of visitors' travel patterns can inform public transportation planning in relation to tourist destinations and contribute to the strategic assessment and refinement of policies such as fare structures. In our analysis, visitors are defined using a combination of criteria, including: i) one-time users with a travel span of less than a week (activity criterion), and ii) users utilizing dedicated visitor fare products (fare product criterion). We apply k-means clustering and identify four activity clusters, namely *Multi-travel short-stay visitors*, *Long-stay and frequent travellers*, *Day trippers*, and *Evening visitors*, as well as three spatial visitor clusters denominated as *City center visitors*, *Outside the city center visitors*, and *Mixed visitors*. The clustering reveals that *Day trippers* dominate the activity segmentation, accounting for 38.5% of visitors, while the spatial segmentation results in three more evenly distributed clusters. Notably, 37.3% of visitors primarily utilize the public transportation system outside the city center, highlighting the need for policy attention beyond central areas – for example, accessibility improvements for less accessible visitor destinations. Last, our framework is fully reproducible, demonstrating the potential of smart card data in exploring visitors' travel behavior.

1. Introduction

1.1. Tourism, mobility, and societal impacts

Tourism and transportation are characterized by an interdependent relationship, given that the former cannot occur without the latter. In some cases, transportation is an essential element of visitors' experiences (e.g., by providing accessibility to local destinations), while in other cases it can form the tourist activity per se (e.g., train or cruise journeys) (Page and Ge, 2009). The term *visitor* usually refers to a *"traveller taking a trip to a main destination outside his/her usual environment, for less than a year, for any main purpose (business, leisure, or other personal purpose) other than to be employed by a resident entity in the country or place visited. These trips taken by visitors qualify as tourism trips"* (Eurostat, 0000). The term *tourist* usually is encompassed within this definition, referring to overnight visitors only. In the remainder of this manuscript, we opt for using the more inclusive term *visitor*.

In general, transportation accessibility not only defines visitors' ability to reach destinations but also the use of transportation services at destinations during their stay. In this context, PT plays a crucial role in facilitating visitors' mobility needs once they arrive at their destination. Notably, only a limited research effort has been devoted to understanding visitors' PT travel behavior, despite the many PT-related challenges associated with tourism. To begin with, PT systems cater to the mobility needs of both locals and visitors. Especially in touristic areas, total passenger demand is greatly affected by the demand induced by visitors, which is characterized by high levels of uncertainty (Hall et al., 2017). Furthermore, seasonality, an important characteristic of tourism, affects the correct estimation of visitor PT demand (Domènech et al., 2020) which is critical for accurate PT supply dimensioning. In cases where the PT supply is not able to cater for the combined needs of locals and visitors, this can impact visitors' overall experience, as well as the perceived attractiveness of the destinations (Albalate and Bel, 2010). This evidence is in line with

* Corresponding author.
E-mail addresses: skoufas@kth.se (A. Skoufas), matejc@kth.se (M. Cebecauer), wilco@kth.se (W. Burghout), jenelius@kth.se (E. Jenelius), oded.cats@kth.se, O.Cats@tudelft.nl (O. Cats).

https://doi.org/10.1016/j.jtrangeo.2026.104808
Received 3 October 2025; Received in revised form 25 June 2026; Accepted 12 August 2026
Available online 22 August 2026

the fact that the overall quality of the provided transportation services is crucial in determining the quality of the tourist experience, as well as tourists' travel decisions, including the decision whether to visit a destination altogether (Pagliara et al., 2025; Le-Klähn and Hall, 2015).

Another important characteristic of visitors' mobility is that their travel patterns can be very different from the locals' (Gutiérrez et al., 2020). Further understanding visitors' spatiotemporal travel patterns can assist in strategic urban spatial planning, while helping local authorities appropriately respond to current and future tourism-related challenges (Klepej and Marot, 2024).

The current challenges in understanding visitors' mobility cannot be fully addressed without referring to the broader societal context of tourism. Overall, tourism has a two-fold impact on local societies. On the one hand, it is a catalyst for economic development. In 2024, the tourism sector contributed 10% to the global economy or $10.9 trillion to the global Gross Domestic Product (GDP). The contribution of tourism to the global labor market is crucial as well, supporting 357 million jobs globally (or one in every ten jobs). Last, visitors' spending also spiked to $1.9 trillion (The World Bank, 2025). In line with the aforementioned global trends, tourism currently represents more than 10% of the European Union (EU)'s economy (or €1.8 trillion) and projections show that the contribution to the economy will reach almost €2.3 trillion by 2035, highlighting the vital role of tourism in shaping the European economic landscape (World Travel & Tourism Council, 2025).

Alongside the positive economic effect, tourism also comes with negative impacts on the local societies. In particular, during the last years, the concept of *overtourism* has been widely used when referring to the negative impacts of tourism (Koens et al., 2018). Overtourism describes how the impact of tourism is expanding beyond control, especially due to the widespread use of accommodation platforms (such as Airbnb) (Nilsson, 2020), resulting in significantly exceeding the physical carrying capacity of a destination. This high social density (or overcrowding) can impact locals' quality of life as well as visitors' experience (Papadopoulou et al., 2023). Recently, demonstrations with key messages such as *"Tourism kills the city"* and *"Tourists go home"* are becoming more common in destinations where tourist arrivals are breaking records (Jover and Díaz-Parra, 2022). These slogans express a polarized view on the impact of tourism on the cities, one that usually neglects the heterogeneous nature of visitor types in tourist cities, which is characterized by complexity (Benito and Zerva, 2025). The understanding of visitors' mobility and its distinct types is of utmost importance for relevant policy-making targeting balanced touristic development while preserving locals' quality of life, thereby supporting a better coexistence between visitors and the local population.

1.2. Modeling visitors' travel behavior

Stated Preference (SP) and Revealed Preference (RP) data have been used to capture visitors' actual behavior as well as their behavior in hypothetical scenarios, respectively (Kemperman, 2021). Past studies reveal that visitors' travel behavior at their home location strongly influences their travel behavior at their destinations (Zientara et al., 2024). In addition, visitors' length of stay is positively associated with PT use (Le-Klähn et al., 2014) since daily visitors show a greater desire for more flexible transportation services (Dominik et al., 2017; Romão and Bi, 2021). Survey-based studies have also explored the use of micromobility services by visitors, indicating that the distance from their accommodation to the city center is a highly influential factor for choosing (or not) micromobility services (Jażdżewska-Gutta et al., 2023).

The increased availability of anonymized mobile phone data has been proven crucial in overcoming the limitations of surveys related to participation rates as well as financial-related constraints. In particular, mobile positioning data have been used for distinguishing destinations based on visitor flows (Raun et al., 2016). Furthermore, it is not uncommon to combine mobile position data with survey as well as census data for studying visitors' travel behavior. Specifically, Benito and Zerva (2025) identified different visitor profiles in the city of Barcelona, Spain, accounting for their spatiotemporal distribution. Interestingly, their approach involving several data sources led to the identification of a significantly higher number of visitors compared to the official statistics. This finding indicates the potential of automated data sources in identifying unobserved tourism, and therefore, quantifying its real impact on the rhythm of the city (De Cantis et al., 2015).

While survey and mobile position data have been useful in modeling visitors' overall travel behavior, they lack the precision needed to model visitors' PT travel behavior. Automated Fare Collection (AFC) systems with the use of smart cards (smart card data) have become very popular in modeling PT passenger behavior due to their rich spatiotemporal resolution (Cats, 2023). Despite the use of smart card data in modeling the travel behavior of several PT passenger groups, such as recurring users (Ghaemi et al., 2017), school students (Skoufas et al., 2024), and residents of new urban developments (Skoufas et al., 2025), there is a noteworthy research gap in exploring visitors' PT mobility using smart card data, and therefore extending their potential in the context of tourism.

There is a limited body of research on understanding visitors' travel behavior using smart card data. In particular, Gutiérrez et al. (2020) identify visitors in Catalonia, Spain, based on a fare product targeted for visitor use, and define spatiotemporal visitor profiles using model-based clustering. Moreover, Ortega-Tong (2013) classifies PT users in London, United Kingdom (UK), including visitors, based on their spatiotemporal attributes using k-means clustering. Apart from European studies, there are studies focusing on identifying visitors in Singapore. Specifically, Lu et al. (2019) develop *TourSense*, which identifies tourists using smart card data based on fare product criterion. In a similar direction, Xue et al. (2014) identify visitors based on a fare product criterion using a machine learning approach. A key insight across the above-mentioned studies is that visitors' PT travel behavior can be classified into similar categories based on the length of stay (e.g., short-stay, long-stay) and the degree of PT utilization (e.g., intensive use, light use) across different urban and tourism contexts. This indicates the existence of recurring behavioral patterns among short-term PT users as visitors. All studies have identified visitors relying solely on the fare product. However, this may be unreliable, given that (i) visitors may use different product types, and (ii) locals with occasional PT usage may use similar fare products as visitors.

1.3. Objectives

In this study, we present a framework for identifying visitors entirely based on RP data, especially smart card data, and defining distinct temporal and spatial visitor profiles. Our method is reproducible, given its reliance on individual mobility traces, thereby enhancing its added value. In particular, we utilize PT travel diaries across Region Stockholm, Sweden. Our work significantly contributes to the existing limited literature body on modeling visitors' PT travel patterns using smart card data. First, we define visitors using a combination of criteria accounting for (i) users' temporal activity in the PT system, and (ii) fare product type, overcoming relevant limitations from definitions involving only the fare product type. Then, we analyze visitors' activity (temporal) and spatial attributes and identify representative groups of visitors using clustering. Last, we assess the relations between the estimated activity and spatial clusters by means of an alluvial diagram.

Further understanding of visitors' mobility is useful for multiple stakeholders, such as tourism stakeholders, by gaining insight into how places are (not) visited. Insights from our study can be useful for authorities and operators for integrating touristic attractions into the PT network by (i) adjusting the PT supply based on accurate visitor demand predictions and (ii) (re-) designing the PT network to connect

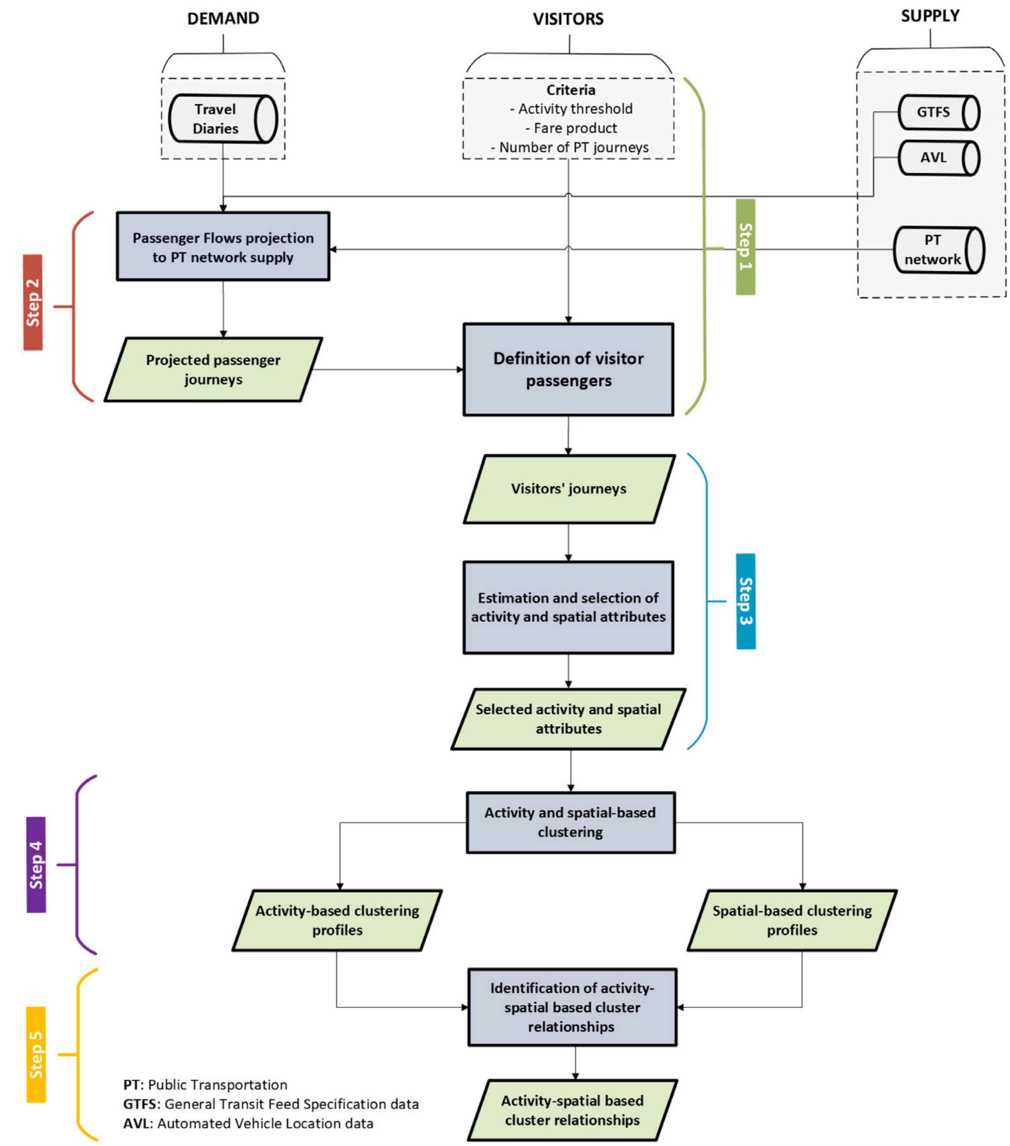

Fig. 1. Methodological framework (elliptic tubes: datasets, rectangles: modules, parallelogram: intermediate databases).

touristic attractions with the rest of the existing network. Furthermore, they can inform the design of tailored fare products for visitors, aligning with the need for more visitor-friendly PT policies towards increasing the attractiveness of tourist destinations (Graham Mandeno, 2012).

The remainder of the paper is structured as follows. Section 2 details the methodological framework. Section 3 presents the case study area, including the local tourism context. Section 4 presents the results of the study, Section 5 discusses the implications of the study, and presents the main conclusions.

2. Methodology

In the following, we present the methodological framework developed in this study. Fig. 1 presents our methodological framework, which has in its core the definition of PT visitors using a combination of criteria (time window, fare product, number of PT journeys). In the next paragraphs, we detail all the methodological steps.

2.1. Visitors' definition using smart card data

Identifying visitors using smart card data is a crucial step in our methodology. It is important to highlight that smart card data do not label whether a user is a resident or a visitor (Egu and Bonnel, 2020). Therefore, the use of a set of criteria is deemed appropriate for defining the visitors' group.

Overall, in the tourism field, visitors can be divided into two categories concerning their destination loyalty: *one-time visitors* and *repeat visitors* (Oppermann, 1999). The latter category concerns visitors who have been to a destination once or multiple times, respectively, within a certain (sufficiently long) time period (e.g., a year). The PT usage patterns of repeat visitors may be similar to the usage patterns of some residents (e.g., residents who use PT occasionally) (Lin et al., 2023). Therefore, there is a risk of failing to distinguish repeat visitors from residents. Given this risk, we choose to consider only one-time visitors in this study. Specifically, we define one-time visitors as those who use PT at least once during the investigated time period with no more than 7 days between their first and last recorded journeys during the investigated period (e.g., 1-year). The selection of the 7-day period over a significantly longer time period (e.g., 1-year) naturally eliminates a significant portion of 'pseudo-visitors' including weekend leisure travellers, and locals attending a Planned Special Events (PSE) (e.g., sports game) more than once. With the term journey, we define the entire travel chain between the origin and the final destination

Table 1
Travel activity and spatial-related variables per visitor obtained for the processed dataset.

Type	Name	Description (unit of measurement)
Target	*CardKey*	Unique smart card identifier
Temporal	*number_of_days*	Total number of active days in the PT system (–)
	number_of_journeys	Total number of PT journeys (–)
	number_of_transfers	Average number of transfers/journey (–)
	journeys_per_day	Average number of journeys/day across the active period (–)
	travel_time	Average travel time/journey (minutes)
	first_journey_time	Average earliest journey start time across the active period (HH:MM:SS)
	last_journey_time	Average latest journey start time across the active period (HH:MM:SS)
Spatial	*distance*	Average road network distance traveled/journey (kilometers)
	total_origin_stops	Total number of unique origin PT stops during the active period (–)
	total_destination_stops	Total number of unique destination PT stops during the active period (–)
	total_origin_zones	Total number of unique origin zones during the active period (–)
	total_destination_zones	Total number of unique destination zones during the active period (–)
	main_OD_%journeys	Share of PT journeys concentrated in the most visited OD (%)
	intra_stockholm_%journeys	Share of PT journeys within Stockholm's inner city (%)

points. We ground the selection of the 7-day threshold in the local tourism context as well as the local fare products, which are particularly designed for visitors. In the Stockholm region, the average visitor stay is estimated to be ca. 2 days in 2023 (The Agency for Economic Growth (Tillväxtverket), 2025). In addition, the regional Public Transportation Authority (PTA) offers the 1, 3, and 7-day visitor fare products (Stockholms Lokaltrafik (SL), 0000). Combining the average visitor stay with the visitor fare products, we opt for an inclusive threshold accounting for visitors having a longer stay, therefore choosing the 7-day upper threshold.

In addition to the travel activity threshold, we use a fare product type criterion, which has been typically used for defining visitors using smart card data (Gutiérrez et al., 2020). Specifically, we also denote users who exclusively use the visitor fare products (1, 3, and 7-day fare products) as visitors.

Last, we apply a minimum travel activity criterion when defining the visitor passenger group. We only consider visitors who have used the PT system at least twice, meaning that we exclude those with a single PT journey. This criterion improves statistical robustness by excluding users who made a single incidental PT journey, therefore reducing the bias induced by very infrequent users.

The combination of the three described criteria (travel activity threshold, fare product, number of PT journeys) leads to the definition of the visitor passenger group that is analyzed and clustered in the next methodological steps (see Fig. 1, Step 1).

2.2. Passenger flows projection to public transportation network

Our methodological framework utilizes passengers' travel records. Using passengers' full travel diaries signifies that the PT mode, line, vehicle departure number, arrival and departure times at the stop level, as well as travel time, are known across each passenger journey i by joining Automated Vehicle Location (AVL) and AFC data. In cases where passengers tap in at a platform or a gate (e.g., metro and commuter rail stations), they are assigned to the earliest feasible departure found in AVL data in relation to their tap-in original timestamp. Last, General Transit Feed Specification (GTFS) data, containing the complete set of departures for all PT lines, can replace AVL data in case of missing or incomplete vehicle departures (see Fig. 1, Step 2).

2.3. Estimation and selection of activity and spatial attributes

Using the projected visitors' PT journeys across the selected time period, we are able to obtain relevant attributes for their temporal and spatial activity. Distinguishing temporal and spatial attributes can highlight the effect of each selected variable during the clustering, therefore obtaining more informative temporal and spatial visitor profiles. Table 1 lists the temporal and spatial attributes included in our analysis for each visitor.

Prior to the clustering analysis, we explore redundancy among all variables by constructing correlation matrices. The highly correlated variables are not jointly used in the clustering process, thereby avoiding redundancy (see Fig. 1, Step 3).

Last, we tested different combinations of the remaining temporal and spatial variable combinations, targeting both conceptually meaningful and distinctive temporal and spatial clusters.

2.4. Activity and spatial clustering

There are multiple clustering techniques that could be applied to identify spatiotemporal profiles of visitors using smart card data, such as (i) centroid-based clustering algorithms (e.g., k-means, k-medoids, clustering large applications (CLARA)), (ii) connectivity-based or also known as hierarchical clustering (e.g., agglomerative and divisive clustering), (iii) density-based (density-based spatial clustering of applications with noise (DBSCAN), ordering points to identify the clustering structure (OPTICS)), (iv) distribution-based (gaussian mixture model (GMM)) and, (v) a group of methods that modify or combine the above-mentioned clustering techniques (e.g., fuzzy clustering, hierarchical k-means, etc.) (He et al., 2020).

For this study, we choose the k-means clustering algorithm. The goal of the k-means algorithm is to define non-overlapping groups of data – mutually exclusive and collectively exhaustive – characterized by a high degree of similarity within each group and a high degree of dissimilarity between the groups. K-means is one of the most popular unsupervised machine learning techniques; it is computationally efficient and therefore suitable for large datasets compared to other clustering techniques (Hartigan and Wong, 1979). K-means has been employed in several PT related studies due to its demonstrated strong scalability in managing multidimensional data, as well as identifying natural clusters for practical applications (e.g., in Kadir et al. (2018) and Liu et al. (2025); see Cats (2023) for a review).

Accounting for the special characteristics of our dataset (large size, multidimensionality), we choose the k-means clustering algorithm and we apply it for grouping visitors based on the selected (i) temporal and (ii) spatial attributes (see Fig. 1, Step 4). In order to identify the number of representative travel activity and spatial clusters, we use internal evaluation indices such as: (i) the total variance within the clusters (Inertia (WCSS) (see Eq. (1)), (ii) the average similarity between each cluster and its most similar one (Davies–Bouldin Index) (see Eq. (2)), and (iii) the intra-cluster cohesion – inter-cluster separation (Silhouette Score) (see Eq. (3)).

$$\mathrm{WCSS} = \sum_{k=1}^{K} \sum_{x_i \in C_k} \|x_i - \mu_k\|^2 \tag{1}$$

$$\mathrm{DB} = \frac{1}{n} \sum_{i=1}^{n} \max_{j \neq i} \left(\frac{S_i + S_j}{M_{i,j}} \right) \tag{2}$$

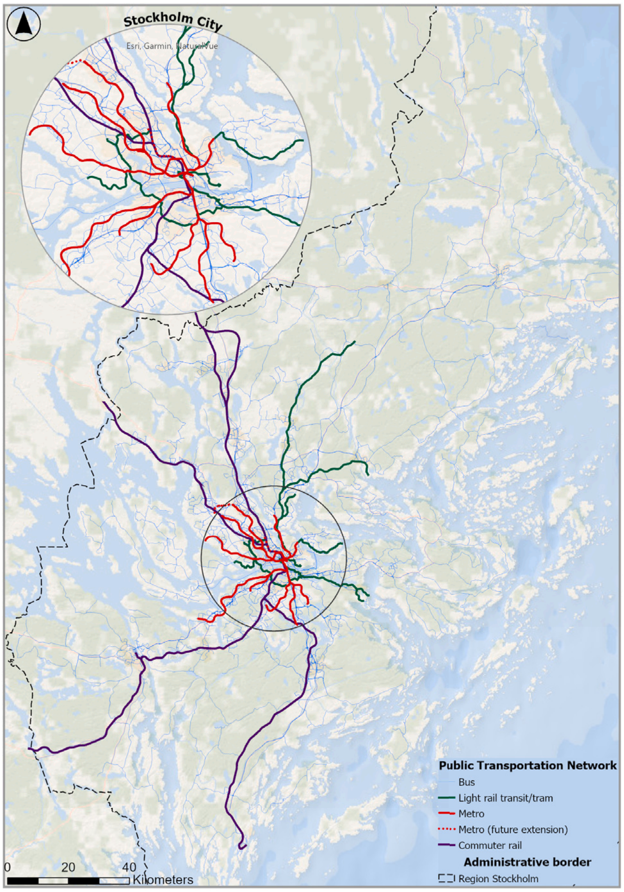

Fig. 2. The multimodal PT system of Stockholm region (basemap credits: Esri, Garmin, GEBCO, NOAA NGDC, and other contributors).

$$\text{Silhouette} = \frac{1}{N}\sum_{i=1}^{N} s_i \tag{3}$$

where K denotes the number of clusters, C_k the set of observations in cluster k, μ_k the centroid of cluster k, and $\| \cdot \|$ the Euclidean distance. Furthermore, S_i represents the average intra-cluster distance of cluster i, $M_{i,j}$ the distance between centroids of clusters i and j. Last, s_i is the silhouette score of observation i, and N is the total number of observations.

Finally, we explore the relationships between the resulting travel activity and spatial clustering profiles (see Fig. 1, Step 5).

3. Case study: Stockholm region

In this section, we present important aspects of the case study area, the Stockholm region, with emphasis on the description of the regional PT system, the data used, and the local tourism context.

3.1. Public transportation system and data overview

Stockholm region has the largest multimodal PT system in Sweden, facilitating the needs of ca. 2.3 million residents (20% of the Swedish population). The PT network comprises 523 PT lines – including bus, metro, bus, commuter rail, light rail/tram, and ferries – and 5549 stations. Fig. 2 presents the multimodal PT system of Stockholm region.

Stockholm region has a tap-in system only, implying that passengers' tap-out locations need to be inferred. The boarding (tap-in) relevant attributes (location, timestamp) of a passenger trip *j* are recorded in the AFC data. The tap-out location of the trip *j* is inferred based on the tap-in location of the next trip *j*+1. Specifically, the inference algorithm is searching for candidate tap-out locations within a defined radius of the tap-in location of the trip *j*+1. The success of the inference is based on i) the time window between the two successive trips *j*, *j*+1 (<8 days), and (ii) the facilitation of the trip between the origin (tap-in location) and the inferred destination by a matching PT line (Cats et al., 2019). A passenger who takes a single daily PT trip on a specific day can have an inferred tap-out location if another trip occurs within the next 8 days. Transfers among the different PT modes are identified by selecting different time thresholds, allowing two consecutive trips to constitute a single journey (and not an intermediate activity) (for more details see Cats et al. (2019), Section 2.5). In addition, statistics on passengers' travel behavior can be aggregated into spatial units (e.g., socioeconomic zones) and more specifically to the Swedish DeSo zones, which are the counterparts of census tracts in the United States (US).

For this study, we analyze visitor demand data from 2023 (January 1–November 23, 05:00–22:00), accounting for 327 days in total. Specifically, we analyze 5,039,516 PT journeys from 1,162,117 identified visitors, corresponding to 21.5% of all cardholders but only 1.7% of the total 291,050,393 PT journeys recorded during the study period. Given, the inability to infer the destination of the absolute final journey of each visitor (1,162,117 out of 5,039,516 journeys, or 23%), we include 77% of the total journeys (3,877,399 journeys) in the clustering analysis. Last, the recognition rate for tap-out locations in 2023 is estimated at 80% of total passenger journeys.

3.2. Swedish tourism context

Tourism in Sweden has seen continuous growth during the last decade (excluding the pandemic period), with almost 33 million visitors (arrivals) spending at least one night in tourist accommodation in 2023 (Statista, 2024a). Importantly, Sweden has been established as the second most popular destination in the Nordics, due to its wide geographical coverage, varying climate, and the numerous cultural events that take place across the country (Statista, 2025). Two-thirds of the total consumption in businesses related to tourism is associated with domestic tourism (Nordregio, 2023). Another element characterizing Swedish tourism is the low seasonality. Specifically, for most Swedish regions, the two busiest months July and August account for up to 25% of the annual total nights spent (Eurostat, 2026).

Travel and tourism significantly contribute to the national economy and labor market. The contribution of travel and tourism to the Swedish economy is estimated to be SEK465.8 billion during 2024 (ca. 2.5% of Sweden's GDP), ca. SEK14 billion more compared to 2023 (Statista, 2024c). Concerning the contribution of travel and tourism to the national labour market, there were ca. 277,000 employees in the travel and tourism field in Sweden in 2023 (Statista, 2024b).

Stockholm region is the 13th most visited destination in Europe, and the most visited region in Sweden. The city of Stockholm, in particular, attracts a significant share of the total visitor arrivals. In 2023, 8 million and 4.7 million visitors arrived in the Stockholm region and the city of Stockholm, respectively, with ca. 70% of them being domestic visitors at both levels (Swedish Agency for Economic and Regional Growth, 2025). The estimated average length of stay in the Stockholm region during 2023 was ca. 2 days, calculated as the ratio of the number of guest nights to the number of arrivals. Activities such as *Walking around the city*, *Sightseeing*, and *Shopping* are the most popular ones. Furthermore, in terms of reasons for visiting Stockholm, the option *"Been here before/wish to return"* is the most popular, indicating the prevalence of repeat visitors (Visit Stockholm, 2023). Regarding visitors' means of travel during their stay, 38% make use of local and regional PT services (Swedish Agency for Economic and Regional Growth, 2020).

Locals have a positive-passive attitude towards tourism, believing that attracting visitors is, in general, a good idea. However, they

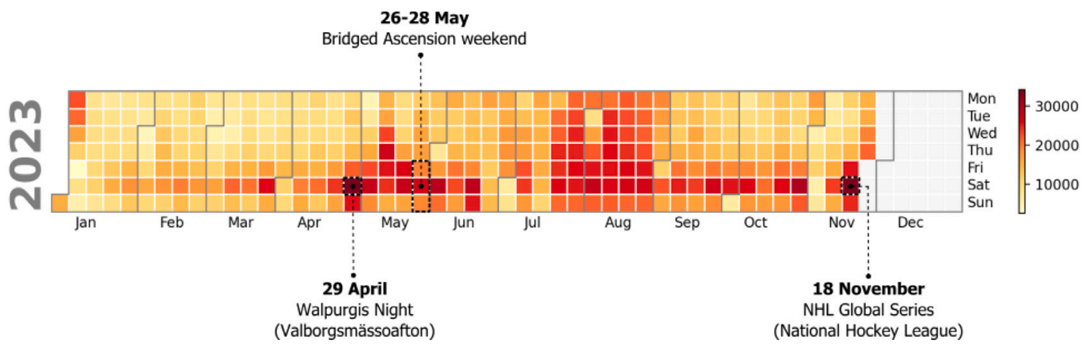

Fig. 3. Number of PT journeys performed by visitors on each day across the investigated time period (January 1–November 23, 2023).

report that *congestion* and *traffic problems* are the most problematic aspects related to tourism, highlighting the need for action by the relevant authorities dealing with the above-mentioned challenges (Visit Stockholm, 2023).

4. Results and analysis

In this section, we present the estimated number of PT journeys performed by visitors during 2023, their travel activity profiles, and spatial visitor clusters, as well as the relations among the defined travel activity and spatial clusters.

4.1. Estimated number of visitors

Fig. 3 presents the number of daily visitor PT journeys identified during the time period (January 1–November 23, 2023, 05:00–22:00).

In total, 5,039,516 PT journeys by 1,162,117 uniquely identified visitors are performed across the investigated time period. Overall, we observe peak periods of visitor activity during mid-May, the second half of July, and August. These seasonal trends are in agreement with the reported temporal trend in the official statistics about tourism, with visitor spikes during summer (Swedish Agency for Economic and Regional Growth, 2025). Notably, the share of visitors PT journeys during July and August alone accounts for 25% of the total sample. In contrast, January, February, and March are characterized by low activity, contributing just 20% of the visitors' journeys.

In terms of day-of-the-week seasonality, we notice higher visitor volumes on Fridays and weekends, consistent with leisure travel patterns. The share of visitor PT journeys on these days collectively reaches 51.2% of the total sample. Interestingly, the total number of visitors on Saturdays is 75% higher compared to Mondays.

The local social-cultural context can explain unusually high visitor volumes on specific days across the year. Specifically, the 29th of April is the day with the highest number of visitor journeys in 2023 (34,209 visitor journeys). On this day, locals celebrate Walpurgis Night (Valborgsmässoafton), a traditional celebration in Sweden involving a significant number of trips to student cities such as Stockholm. Another example is the day after Ascension (May 26), usually bridging a 4-day weekend (19,239 visitor journeys). The existence of local PSEs can also significantly affect visitor volumes. In particular, on the 18th of November, the NHL Global Series – Stockholm, a set of games of the National Hockey League, attracted domestic and global fans of hockey to Stockholm, resulting in 34,081 visitor PT journeys.

In order to assess the reliability of our method in identifying visitors using smart card data, we extract the monthly number of visitors and compare them with official statistics as reported by the Swedish Agency for Economic and Regional Growth (2025). Such a comparison has been employed in the study of Lin et al. (2023) for validating the estimation of visitors using smart card data. We estimate and compare our results for (i) the Stockholm Municipality and (ii) the Region Stockholm level. For this estimation, we identify the Stockholm Municipality visitors as those whose most frequently visited OD pair lies within its geographical boundaries. Fig. 4 presents the estimated monthly number of visitors versus the official statistics at the (a) Stockholm Municipality and (b) Region Stockholm levels, alongside the estimated Pearson correlation coefficient.

The strong correlation between the estimated results and the official statistics confirms the effectiveness of our method (Pearson's correlation coefficient of 0.85 for both comparison levels). The constantly lower number of estimated visitors compared to the official statistics is in line with the fact that official statistics serve as an upper limit (i.e., including also visitors who do not make use of PT).

4.2. Travel activity clustering profiles

Prior to the clustering analysis, we explore the redundancy among all travel activity-related variables by constructing a correlation matrix. Among the temporal variables, *number_of_days* and *number_of_journeys* are found to be highly correlated (>0.80), and therefore are not jointly used in the clustering process. To define activity visitor profiles using the k-means clustering technique, we employ four temporal variables: *number_of_days*, *journeys_per_day*, *first_journey_time*, and *last_journey_time*. The consideration of these variables aims at both conceptually meaningful and distinctive clusters based on visitors' temporal travel patterns. Fig. 5 presents the distributions of the selected activity-related variables.

The distribution of the *number_of_days* variable is right-skewed, indicating visitors' presence in the PT system for a limited number of days. Specifically, the average and the median of the total number of active days is 2 days, while 75% of visitors are active on up to 3 days. The *journeys_per_day* variable also has a right-skewed distribution, reflecting low PT usage by visitors. In particular, the average and the median are estimated to 2 PT journeys per day (interquartile range (IQR): 1.6–2.3 PT journeys). Concerning the *first_journey_time*, the median value is almost at 13:00 (12:48) with 75% of visitors having started their first PT journey by 14:45. Last, regarding the *last_journey_time*, the median time of visitors' last PT journey is around 16:30 (16:26) with 75% of visitors starting their journeys in early evening (18:12).

We choose the number of activity clusters by examining the Inertia (WCSS), the Davies–Bouldin Index, and the Silhouette Score. We obtain those metrics for various numbers of clusters, ranging from 2 to 10, in Fig. 6. We adopt the elbow rule, therefore selecting to proceed with four activity clusters.

The four selected clusters are then interpreted based on differences in visitors' activity characteristics. We label the clusters based on the PT utilization level during the stay, the length of the stay per-se, as well as the time of the last PT journey. The resulting clusters are labeled as follows:

- ***A1 - Multi-travel short-stay visitors***: Visitors with a short stay who make several daily journeys. They are characterized by high activity during the afternoon hours.

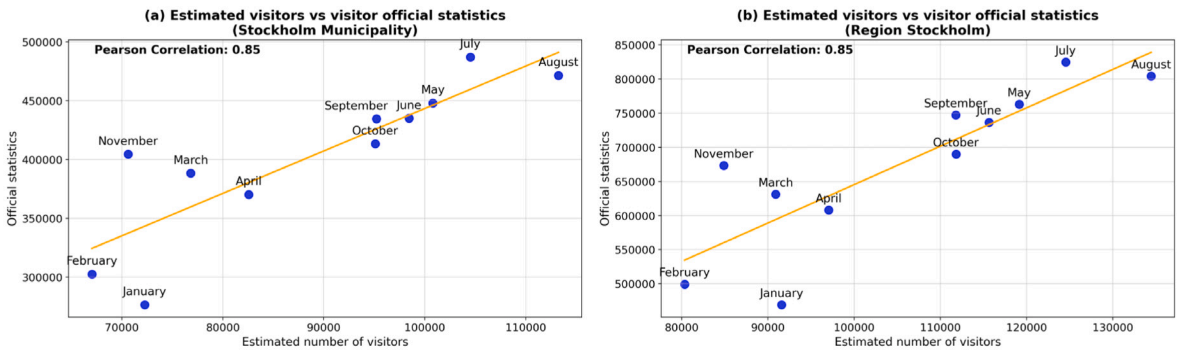

Fig. 4. Monthly number of visitors as estimated by smart card data vs. official statistics for (a) Stockholm Municipality and (b) Region Stockholm.

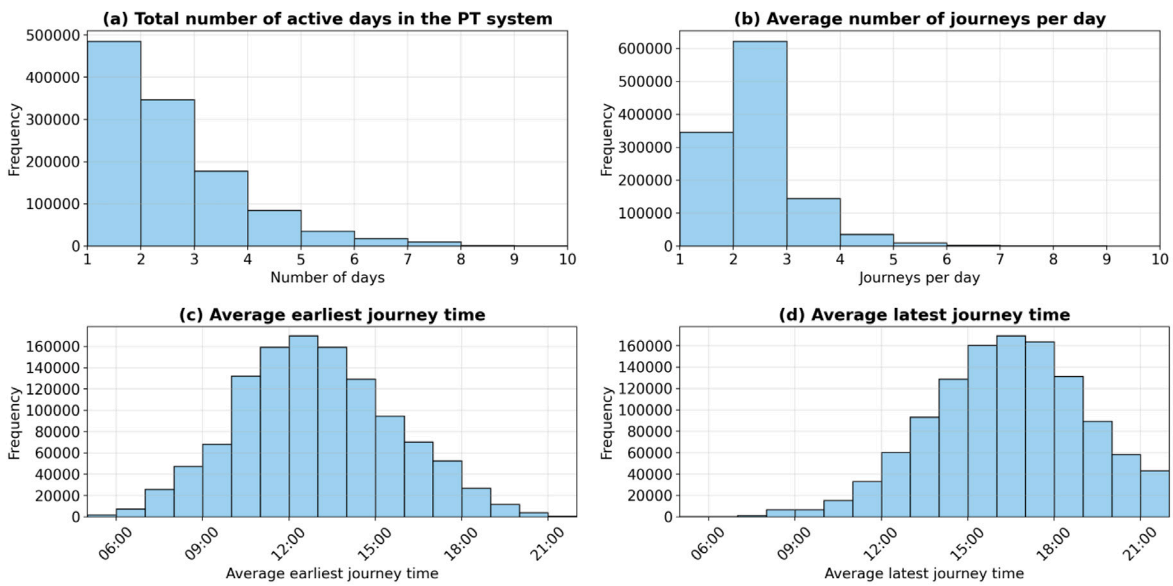

Fig. 5. Distributions of the (a) Total number of active days in the PT system, (b) Average number of journeys per day, (c) Average earliest journey time, and (d) Average latest journey time.

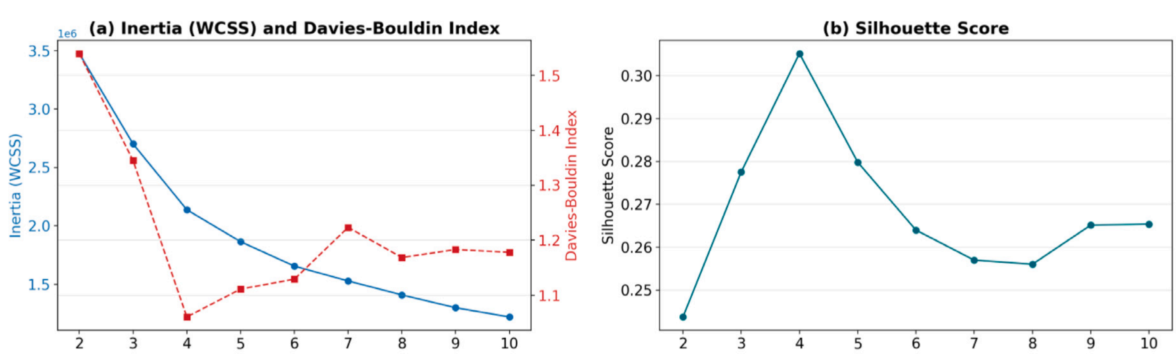

Fig. 6. K-means (a) Inertia (WCSS), Davies–Bouldin Index scores, and (b) Silhouette scores for the explored range of activity cluster numbers.

- ***A2 - Long-stay and frequent travellers***: Visitors who stay for multiple days and make more journeys/day. They are also characterized by high activity during the afternoon.
- ***A3 - Day trippers***: Visitors whose first and last journeys occur early in the day. They are characterized by short stays (a single day) with moderate daily activity (ca. 2 daily journeys).
- ***A4 - Evening visitors***: Visitors whose first and last journeys occur late in the day. They are characterized by short stays with a moderate travel activity per day (ca. 2 daily journeys).

Concerning the activity cluster sizes, cluster A3 represents 38.5% of the total 1,162,117 visitors. Cluster A4 accounts for 29.5% of visitors, and clusters A1 and A2 constitute 14.7% and 17.3% of total visitors,

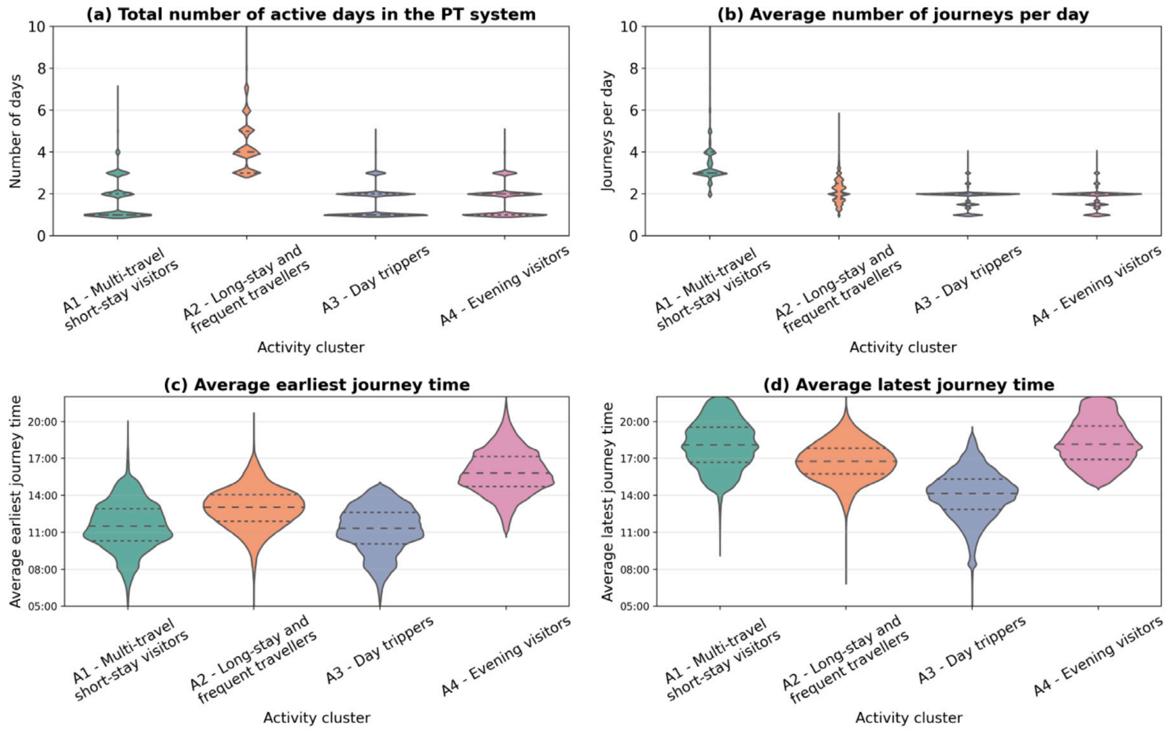

Fig. 7. Distributions of the (a) Total number of active days in the PT system, (b) Average number of journeys per day, (c) Average earliest journey time, and (d) Average latest journey time for each travel activity cluster.

respectively. Fig. 7 presents the distributions of the selected variables for the resulting clusters (A1–A4) (data point: cardholder (passenger)). For visualizing the distributions, we use violin plots, showing the data density at different values, where wider sections indicate higher density (more data points), making them appropriate for comparing distributions across different visitor clusters. The horizontal dashed lines show the 25th, 50th, and 75th percentiles of the corresponding distributions.

Concerning the total number of active days in the PT system, cluster *A2 - Long-stay and frequent travellers* stands out with the highest median (4.3 days) compared to the rest of the clusters with a median of ca. 1.5 days. Regarding the average number of journeys per day, cluster *A1 - Multi-travel short-stay visitors* has the highest median value (3 daily journeys) compared to the rest of the clusters with a median of 2 daily journeys. Notably, 75% of visitors in cluster A2 take up to 4 daily PT journeys, and the maximum value reaches 14 daily PT journeys, indicating an intense PT use. Regarding the average earliest journey times, cluster A3 has the lowest median compared to the rest (11:20), with 75% of day trippers starting their first journey early afternoon (12:36), confirming their activity earlier in the day compared to the rest of the clusters. Cluster A4 has the highest median value (ca. 16:00), reflecting that evening visitors tend to make their first PT journey significantly later compared to other visitors. Last, about the average latest journey time, cluster A4 has the highest median value (18:10) (IQR = 16:55–19:36), indicating evening visitors' PT utilization later during the day compared to visitors from the other clusters. Cluster A3 has the lowest median value (14:09), strengthening its description as visitors who are mostly active during the day. Last, clusters A1 and A2 reflect a mixed visitors' behavior during both the day and the evening.

Enriching further the travel activity visitors' profiles, we extend our analysis to understand the temporal variations of visitors' volumes throughout the year. Fig. 8 presents the monthly visitor volumes per cluster, highlighting seasonal trends and activity peaks.

For clusters A1, A2, and A4, we observe a pattern which is consistent with the overall trend for the number of visitors, reaching a peak during the summer months. In particular, for clusters A1 and A2, this peak is observed in August, while for cluster A4, the peak is reported earlier in July. Contrastingly, cluster A3, the largest amongst the four clusters, has two local peaks in May and October, and a local minimum in July.

Fare product usage per activity cluster can shed more light on visitors' preferences, complementing the analysis while contributing to our understanding of the distinctive characteristics of the identified clusters. Fig. 9 illustrates the differences in fare product composition among the clusters, focusing on the use of the single ticket as well as on the visitor fare products (1, 3, 7-day fare products).

Multi-travel short-stay visitors (cluster A1) have the most diverse fare product use, with substantial shares across all fare product types. Given their combined characteristic of short stay and intense activity, they mainly prefer the 24-hour (1-day), and the 72-hour (3-day) fare product, accounting for 27.5%, and 43.5% of their PT journeys, respectively. On the other hand, long-stay and frequent travellers (cluster A2) mainly utilize the 7-day fare product for half of their corresponding PT journeys (51.8%). The 72-hour ticket is the second most popular choice, accounting for 32.7% of their journeys. Last, for day trippers and evening visitors (clusters A3, A4), the use of single tickets is dominant, given their occasional PT utilization accounting for the 63.4% and 52.1% of their PT journeys, respectively.

4.3. Spatial clustering profiles

In this subsection, we focus on defining visitors' distinct behavior from a purely spatial standpoint. Similarly, when defining travel-activity related visitor profiles, we explore the redundancy among all the spatial-related variables by constructing a correlation matrix. The variables, *total_origin_stops*, *total_destination_stops*, *total_origin_zones*, and *total_destination_zones* are found to be highly correlated (>0.80), and consequently are not jointly used in the clustering process. To reduce skewness and improve the clustering performance, we calculate the natural logarithms (base e) of the *distance* and *total_destination_stops*.

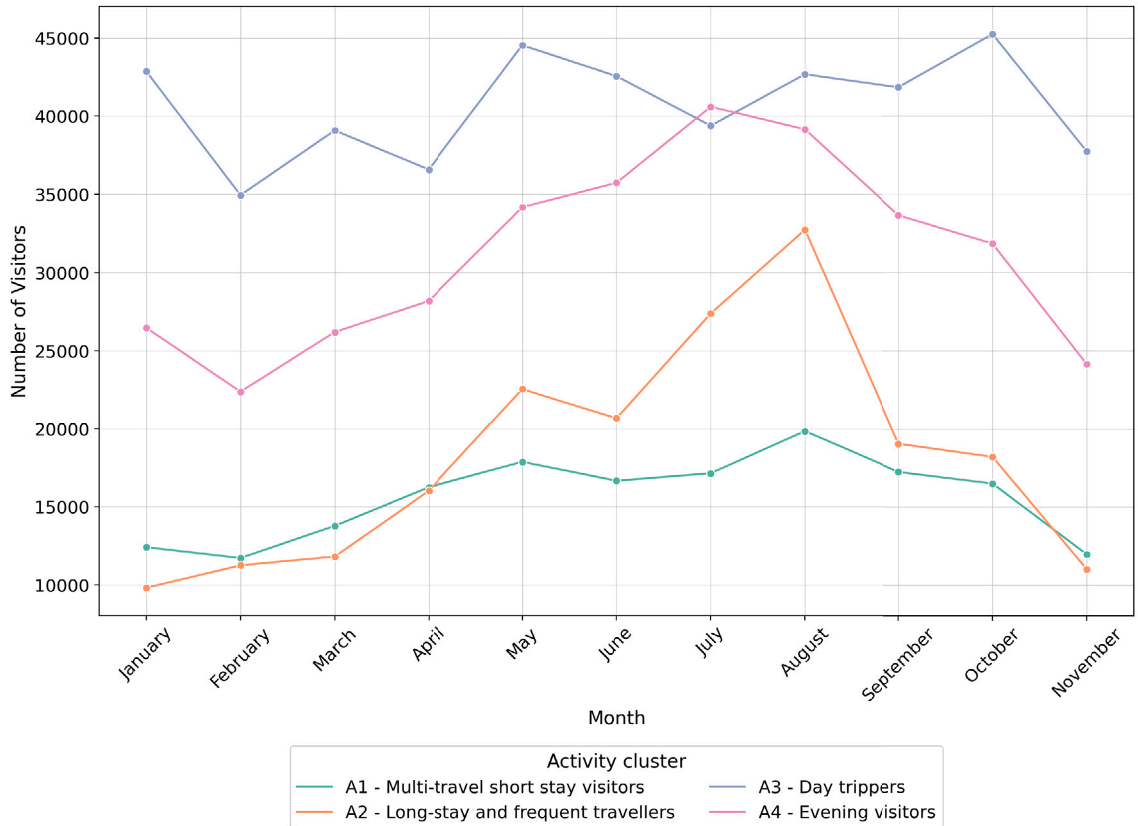

Fig. 8. Monthly visitor volumes per activity cluster.

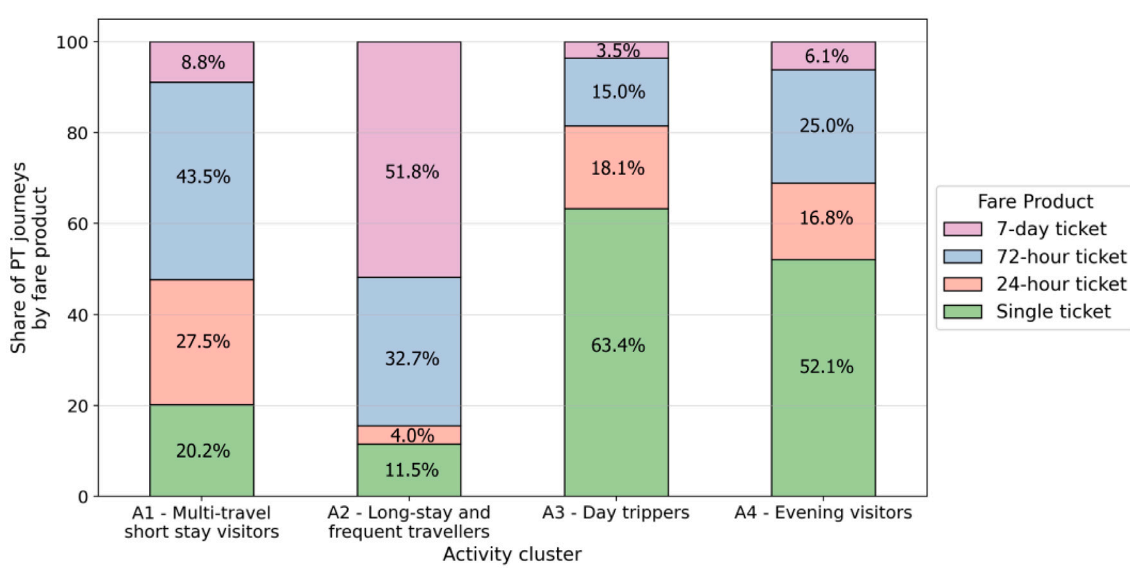

Fig. 9. Fare product composition per travel activity cluster.

The resulting variables *log_distance* and *log_total_destination_stops* are subsequently used in the clustering.

We deploy the k-means clustering technique to define spatial clusters. The variables of *distance*, *total_destination_stops*, *main_OD_%journeys*,
and *intra_stockholm_%journeys* are chosen for defining spatial profiles, aiming at conceptually meaningful and distinctive clusters. Fig. 10 presents the CDF of the selected spatial variables for visitors as well as for non-visitors (data point: cardholder (passenger)).

In the following, we discuss our main observations from contrasting the results obtained for visitors and non-visitors. Concerning traveled distance (see Fig. 10a), the CDFs show that visitors generally travel shorter distances than non-visitors. The median value for visitors is 4.3 km (IQR = 2.8–6.5 km), compared to non-visitors with a median of 6.1 km (IQR = 4.2–8.9 km). Concerning the unique visited destinations

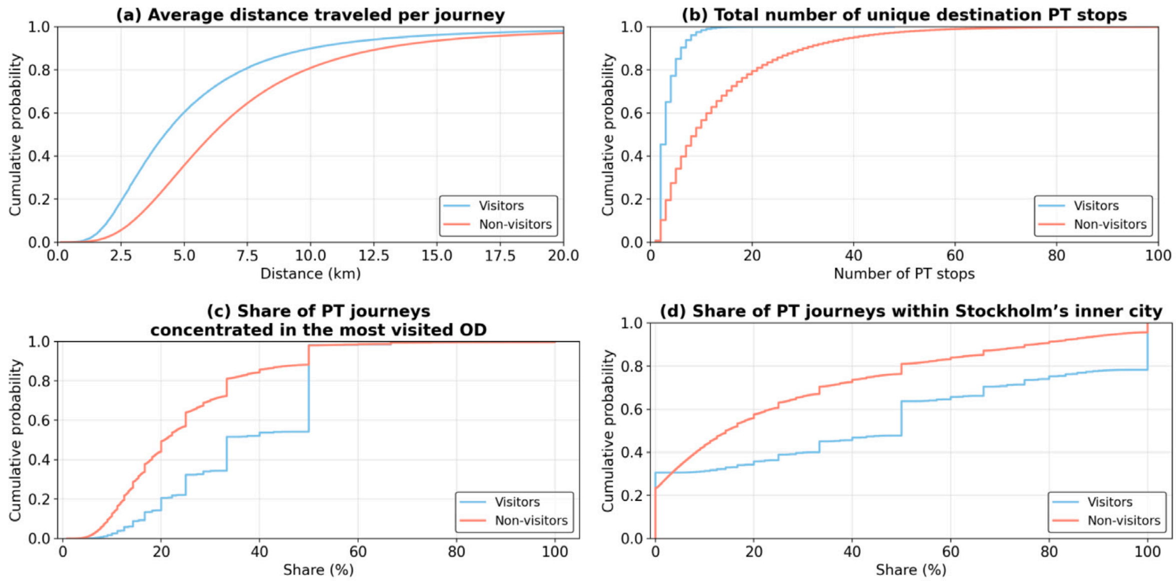

Fig. 10. CDFs of the (a) Average distance traveled per journey, (b) Total number of unique destination PT stops, (c) Share of PT journeys concentrated in the most visited OD, and (d) Share of PT journeys within Stockholm's inner city for visitors (blue) and non-visitors (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

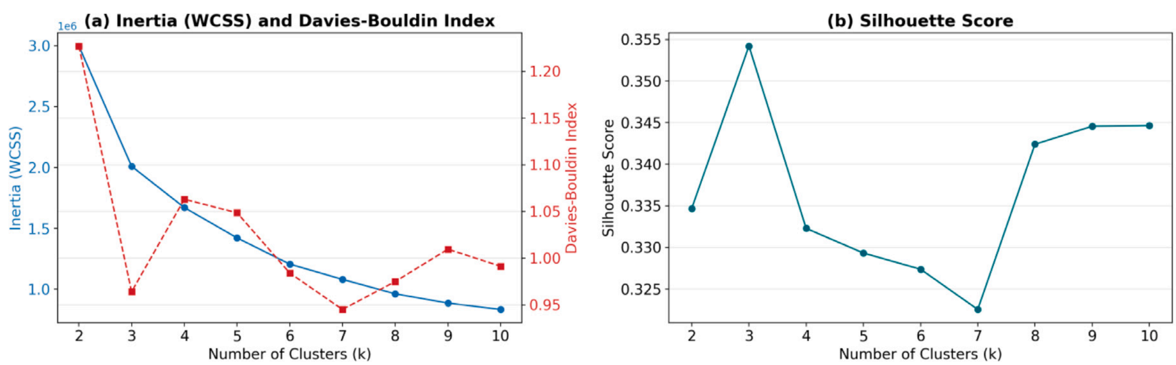

Fig. 11. K-means (a) Inertia (WCSS), Davies–Bouldin Index scores, and (b) Silhouette scores for the explored range of spatial cluster numbers.

(see Fig. 10b), visitors tend to visit considerably fewer PT stops, suggesting a limited spatial variation in visited destinations. The median value for visitors is three PT stops (IQR = 2–4) compared to non-visitors who visit a significantly higher number of PT stops (median = 9, IQR = 4–18)

Regarding the concentration of PT journeys in the most visited OD (see Fig. 10c), visitors' present higher concentration shares. Specifically, the median share of journeys is 33% (IQR = 25%–50%) compared to 21% (IQR = 13%–33%) for non-visitors. Last, regarding the share of PT journeys within Stockholm's inner city (see Fig. 10d), visitors tend to have a much higher share of journeys within the inner city compared to non-visitors. The median share for visitors is 50% (IQR = 0%–80%), compared to non-visitors, with a median share of 15% (IQR = 1%–44%), i.e. one in two journeys versus one in seven or eight journeys.

We choose the number of spatial clusters by examining the Inertia (WCSS), the Davies–Bouldin Index, and the Silhouette Score. We calculate those metrics for various numbers of clusters, ranging from 2 to 10 (Fig. 11). We decide to proceed with three clusters, adopting the elbow rule as well as considering a realistic number of clusters capturing distinct spatial travel behavior.

The three obtained spatial clusters are then labeled based on visitors' place of activity (urban, suburban). The resulting spatial clusters are labeled as follows:

- ***S1 - City center visitors***: Visitors mainly traveling within the city center. They make short-distance trips to a limited number of destinations.
- ***S2 - Outside the city center visitors***: Visitors mainly traveling outside the city center for longer distances to a limited number of destinations.
- ***S3 - Mixed visitors***: Visitors traveling both to central and non-central areas. They demonstrate exploratory behavior due to their tendency to visit multiple destinations over longer distances.

In terms of cluster sizes, visitors are approximately evenly distributed among the three clusters. In particular, cluster S2 accounts for 37.3% of total visitors. Clusters S1 and S3 represent 31.3% and 31.4% of the total visitors, respectively. Fig. 12 presents the distributions of the selected variables for the resulting spatial clusters (S1-S3).

Concerning the average distance traveled per PT journey, visitors outside the city center (cluster S2) have the highest median (6.5 km) and a large IQR (4.8 km–9.4 km), suggesting that visitors from this cluster tend to travel longer with PT. Mixed visitors (cluster S3) have

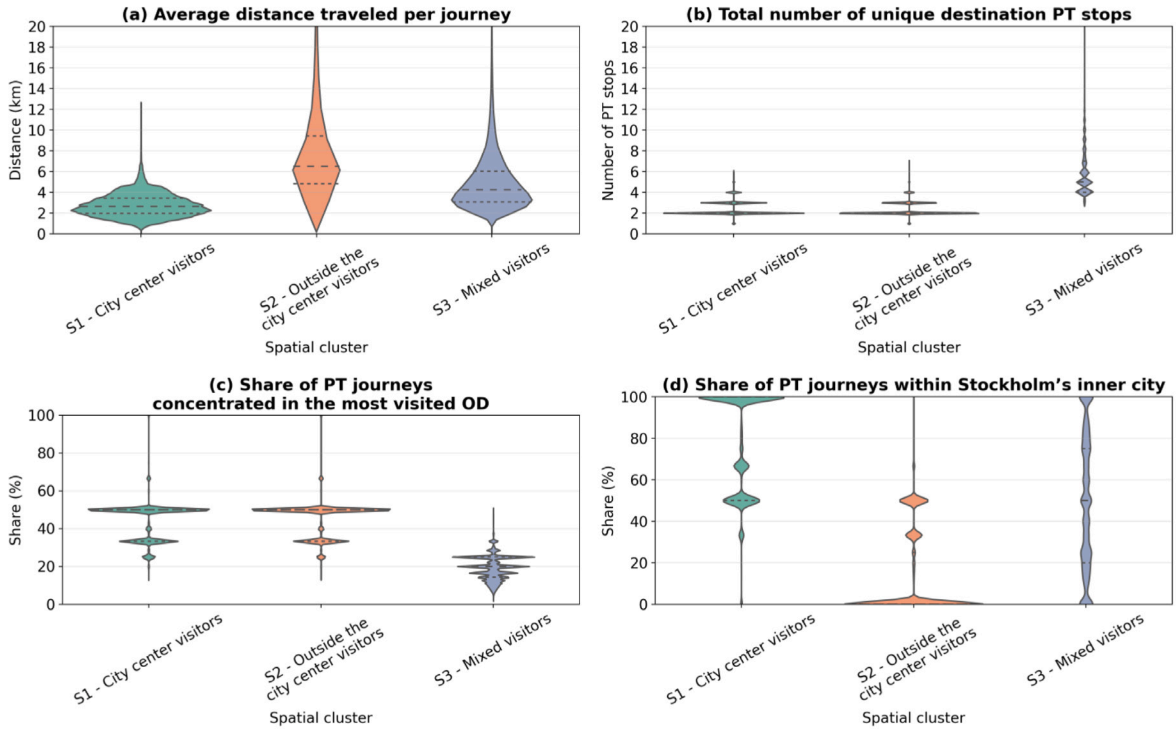

Fig. 12. Distributions of the (a) Average distance traveled per journey, (b) Total number of unique destination PT stops, (c) Share of PT journeys concentrated in the most visited OD, and (d) Share of PT journeys within Stockholm's inner city for each spatial cluster.

the second largest median of 4.3 km (IQR = 3.1 km–6.0 km), while city center visitors (cluster S1) are those traveling the shortest distances with a median average distance of 2.6 km (IQR = 2.0 km–3.4 km), given the geographical limitations of the city center. Concerning the total number of unique destinations, mixed visitors have the highest median of 5 unique destinations (IQR = 4–7 destinations) compared to city center and outside the city center visitors with a median of 2 unique destinations (IQR = 2–3 destinations).

Regarding the share (%) of PT journeys for the most visited OD, city-center and outside the city center visitors exhibit a similar behavior, with half of their journeys being between their most visited OD (median = 50%, IQR = 33.3%–50.0%). Mixed visitors have a more exploratory visitor profile in terms of visiting diverse destinations, with the median of the cluster being just 20% (IQR = 14.2%–50.0%). Last, regarding the share (%) of PT journeys within Stockholm's inner city, city-center visitors have the highest median of 100% (IQR = 50%–100%), while the lowest median of 0% is observed in the cluster S2, as anticipated. Mixed visitors have a more balanced behavior with their median being 50% (IQR = 20%–75%), indicating their presence in the city center as well as in suburban/peripheral areas.

For a better understanding of the geographical presence of the three defined spatial clusters, we display in Fig. 13 the number of generated visitors' PT journeys in each of the spatial DeSo units.

Our results confirm the existence of different spatial travel patterns for visitors. In general, the city of Stockholm and its immediate surroundings, as well as the Stockholm Arlanda Airport zone, play a crucial role in generating PT journeys among all spatial visitor clusters. For city center visitors (S1), these zones act as primary generators, reflecting more compact spatial travel patterns with Stockholm being the epicenter. Contrastingly, visitors outside of the city center (S2) have a broader spatial footprint, including zones further away from Stockholm's urban core (e.g., Ekerö, Vaxholm). Last, mixed visitors exhibit the most spatially dispersed travel behavior, combining urban zones (e.g., T-Centralen, Djurgården), the airport zone, as well as several zones in Stockholm's archipelago as important origin points (e.g., Tyresta Nationalpark).

In order to interpret the spatial dispersion of each spatial cluster further, we are exploring its connection with actual accommodation points of interest (POIs). Specifically, we are identifying the most common origin station per visitor with at least two (2) journeys during the morning (05:00–12:00), assuming that this location is associated with the actual accommodation location during their stay. Finally, 176,965 visitors (15.2% of the total sample) fulfill this criterion. Fig. 14 visualizes the density of common-origin PT stations during the morning for each visitor within each spatial cluster and associates them with the actual accommodation point of interest (POI) retrieved from OpenStreetMap.

For the subset of visitors for whom accommodation anchors could be identified (15.2% of the total sample), results reveal that outside the city center (S2) and mixed (S3) visitors have more spatially distributed accommodation anchors than city center visitors (S1), whose common PT origins are mainly concentrated in Stockholm's city center. Within this subset, this result suggests a relationship between the accommodation location and travel behavior in terms of traveled distance/PT journey. Our results (see Fig. 12a) indicate that clusters S2 and S3 have a higher average distance traveled per PT journey compared to cluster S1, given that centrally located visitors benefit from proximity to main activity areas from their actual accommodation locations.

Last, we are leveraging smart card data to further understand the sequential nature of visitors' PT journeys through a travel trajectory chain analysis. For visitors with an inferred PT accommodation station (15.2% of the total sample), we estimate the *share (%) of PT journeys associated with the inferred accommodation station* (either as origin or destination). Fig. 15 presents the CDF of the estimated share across the identified spatial clusters (data point: cardholder (passenger)).

For 15.2% of visitors, results indicate that the city center and outside the city center visitors (clusters S1, S2) involve their inferred accommodation PT station in a large share of their journeys. In particular, 50% of visitors in both clusters include their inferred accommodation location in at least 85% of their total journeys, suggesting a predominance of *hub-and-spoke* tours. However, mixed visitors (S3) belonging

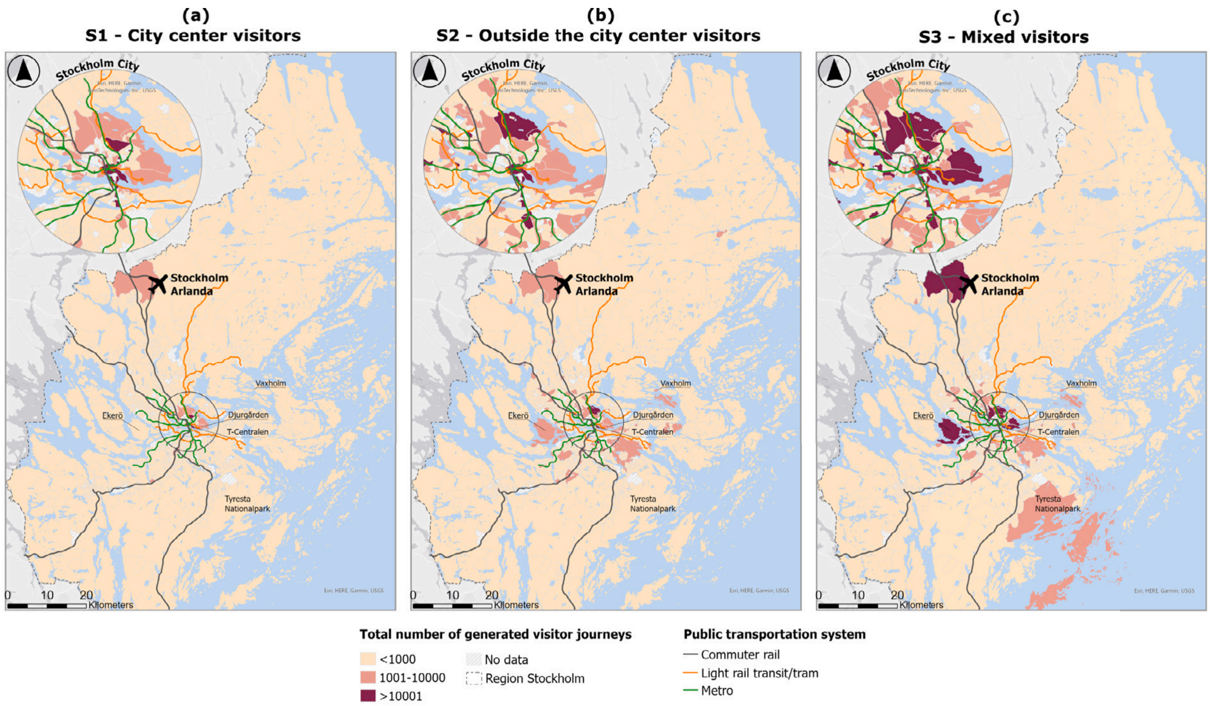

Fig. 13. Total number of generated visitor journeys/DeSo zone for (a) S1 - City center visitors, (b) S2 - Outside the city center visitors, and (c) S3 - Mixed visitors (basemap credits: Esri, Garmin, GEBCO, NOAA NGDC, and other contributors).

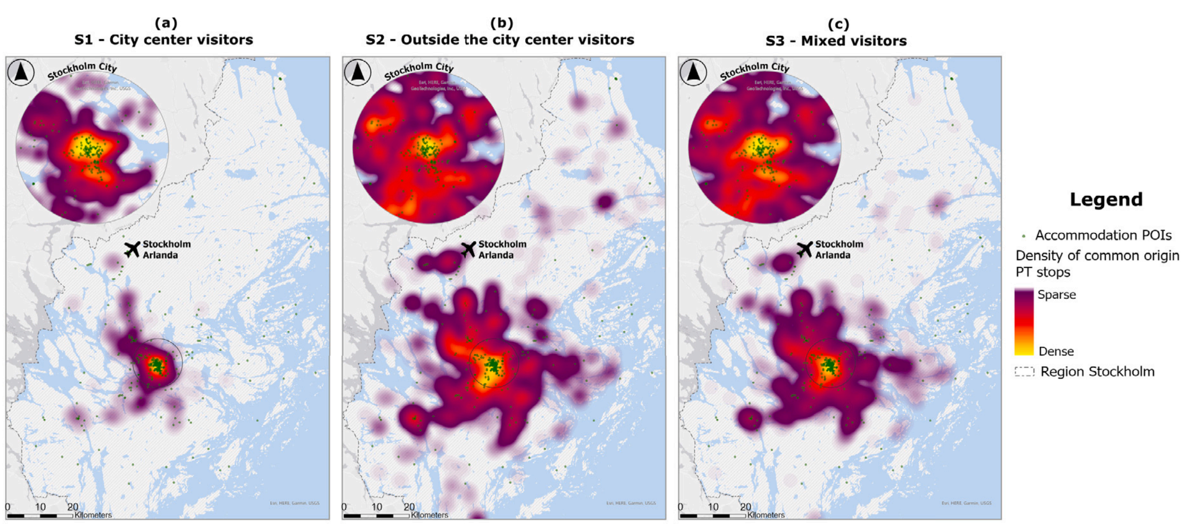

Fig. 14. Density of common origin PT stops and accommodation POIs for (a) S1 - City center visitors, (b) S2 – Outside the city center visitors, and (c) S3 - Mixed visitors.
POIs source: OpenStreetMap contributors (2026), basemap credits: Esri, Garmin, GEBCO, NOAA NGDC, and other contributors.

to the analyzed visitors subset (15.2%) seem more involved in *chain-based* tours, with 50% including their accommodation location in at least 60% of their total journeys. This finding indicates that visitors traveling both urban and suburban areas exhibit a lower reliance on their accommodation location during their visit.

4.4. Activity and spatial clusters relations

Once the temporal and spatial clusters are defined, we assess their correspondence by means of an alluvial diagram (see Fig. 16). This type of diagram is structured by two columns (strata), indicating the relative sizes of the activity profiles (top side) and the spatial profiles (bottom side). The width of the lines linking the two columns (alluvium) represents the total number of visitors, showing how they are classified temporally and spatially by the k-means algorithm.

City center visitors (S1) are primarily associated with day trippers (A3) and evening visitors (A4), highlighting the importance of Stockholm's urban core for diverse activities throughout the day. At the same time, sizeable shares of day trippers (A3) (51.5%) and evening visitors (A4) (42.6%) are associated with visitors' presence out of the city center (cluster S2), indicating the attractiveness of Stockholm's peripheral areas for different types of activities. Last, mixed visitors (S3) are mainly related to multi-travel short stay visitors (A1) (54% of the latter cluster's visitors) and long stay and frequent travellers (A2)

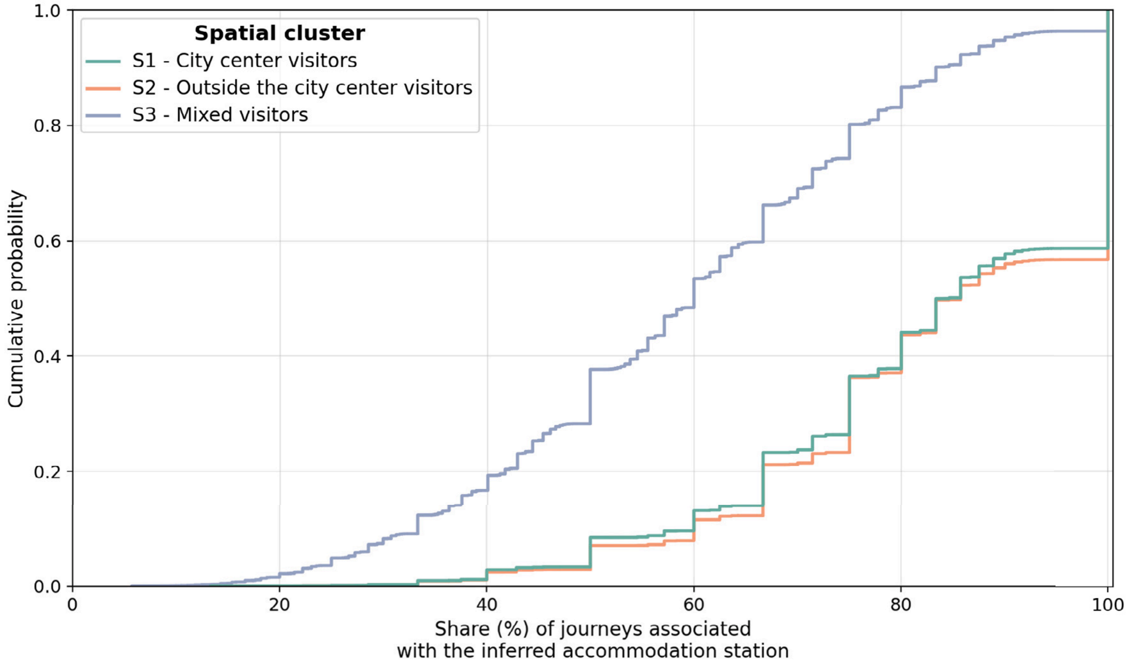

Fig. 15. CDF of the share (%) of PT journeys associated with the inferred accommodation station per visitor and spatial cluster.

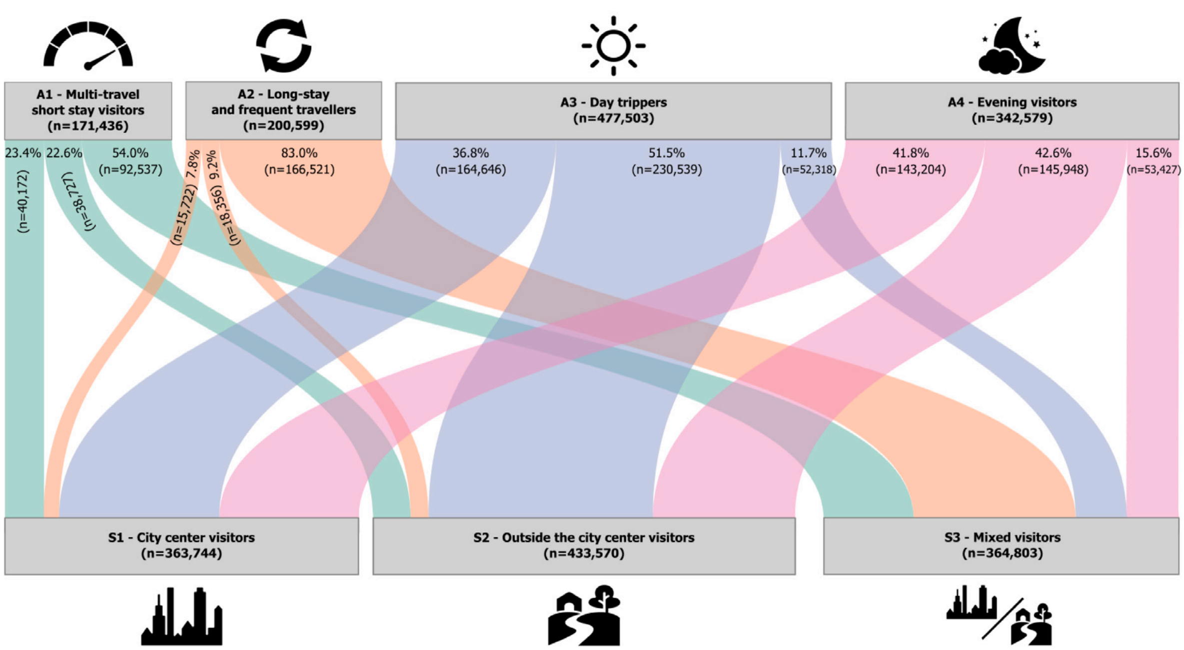

Fig. 16. Alluvial diagram illustrating the relationships between visitors' activity and spatial clusters.

(83% of the cluster's visitors), suggesting a more spatially balanced profile for those visitor categories visiting central and non-central areas.

The observed relationships between travel activity and spatial clusters suggest that visitors with different travel activity patterns have heterogeneous spatial footprints, with shorter-visit visitors having a more limited footprint than longer-visit visitors. These associations highlight the significance of jointly concerning travel activity and spatial presence for relevant policy design. In the tourism context, such insights can inform policies ranging from tailored fare products, and multi-attraction bundles including PT access with admission to tourist attractions, to strategic tourist investments (e.g., investments for improving PT accessibility for less accessible visitor destinations).

5. Discussion and conclusions

We develop a framework for identifying visitors as well as uncovering distinct temporal and spatial profiles based exclusively on using smart card data. We apply our framework to Stockholm region, Sweden, a destination characterized by domestic tourism and low seasonality.

Smart card data have been widely used for exploring the travel behavior of various passenger groups. However, very few studies have incorporated smart card data in identifying and exploring visitors' travel patterns in tourist cities and regions. To this end, our study contributes to the relevant limited literature body in two ways. Firstly, using a joint set of criteria for identifying visitors (activity threshold, fare product, number of PT journeys) overcomes the potential limitations in identifying visitors solely using a fare product criterion given that visitors may use different fare products (e.g., single ticket), especially in contactless-pay-as-you-go PT systems. Correct estimations of the number of visitors is crucial for cost-efficient dimensioning of PT supply, especially in destinations with high visiting seasonality (Kashfi et al., 2015) or destinations with unobserved tourism (De Cantis et al., 2015). Furthermore, correct estimations can assist authorities in assessing visitors' modal split and thereby design policies aimed at modal shifts.

Furthermore, our study enriches the relevant literature focusing on profiling visitors based on their spatiotemporal travel patterns, a current research gap. We identify four temporal profiles which best describe visitors' travel patterns: *Multi-travel short-stay visitors*, *Long-stay and frequent travellers*, *Day trippers*, and *Evening visitors*. Our findings are in line with findings reported in the literature. In particular, Gutiérrez et al. (2020) define *Short term*, *Intensive use*, *Long term*, and *Continued* activity visitor profiles in Barcelona, among others. Similarly, Benito and Zerva (2025) classify visitors as *Tourists*, *Same-day visitors*, *Overnight same-day visitors*, and *Home-return same-day visitors* also in the city of Barcelona. Last, Ortega-Tong (2013) identifies activity profiles distinguishing *regular* and *occasional* visitors for London, UK.

Understanding and segmenting visitors' travel behavior in terms of activity can help authorities better manage the combined passenger flows of visitors and locals. In the case of Stockholm, results reveal that visitor passenger volumes are higher during the summer (see Fig. 8), especially for *evening visitors* and *long-stay and frequent travellers*. These results can guide relevant policy-making in extending service hours for PT services with high visitor volumes. At the same time, attention should be given to the vulnerability of certain PT services, as during summer, there is reduced supply in response to the reduced local demand, potentially exacerbating crowding induced by visitors. Moreover, visitors' activity identified clusters can facilitate the assessment of the use of existing visitors' fare products. In particular, visitors' classification can inform the design of tailored fare products such as evening passes, and multi-attraction bundles or city cards/visitor passes bundling PT access with admission to attractions. The *Vienna City Card*, the *London City Pass*, and the *Tokyo City Pass* serve as good examples for offering tailored bundles based on visitors' level of preferred activity, in terms of PT use and the number of attractions they want to visit during their stay.

In terms of spatial features of visitors' travel patterns, we identify three distinctive profiles: *City center visitors*, *Outside the city center visitors*, and *Mixed visitors*. This aligns with the results reported in Gutiérrez et al. (2020) who defined *sprawled*, *semi-sprawled*, *main cities*, and *coastal* visitors, highlighting that despite differences in the urban form of different cities (e.g., Barcelona and Stockholm), and the geographical context, similar visitor categorizations can emerge. However, it is important to note that despite these similarities, the results of this study may not be directly applicable to other tourism contexts. The definition of visitors, which underpins this analysis, is tailored to the tourism context of Region Stockholm. Practitioners and researchers from other tourism contexts can adjust the reproducible framework proposed in this study to account for different tourism dynamics and contextual factors (e.g., seasonality).

Identifying distinct visitor spatial profiles can offer valuable insights to: (i) relevant authorities in transportation planning for tourist destinations (including PT), and (ii) tourism stakeholders (e.g., service providers) in enhancing marketing campaign effectiveness while preventing overcrowding induced by visitors (Li et al., 2020). For example, among the subset of visitors with an identifiable accommodation anchor (15.2% of total visitors), concerning city center (urban) visitors, who benefit from the proximity to main attractions from their accommodation locations as our results suggest (see Fig. 14a), targeted investments in improving pedestrian facilities or in traffic calming measures can enhance visitors' and locals' satisfaction during their visit (Das and Maitra, 2025). For the case of outside the city center (suburban/peripheral/rural) visitors, our results can enable PTAs to make informed policy interventions regarding strategic PT and tourism investments (e.g., investments for improving PT accessibility for less accessible visitor destinations) (Coppola et al., 2020). Regarding mixed visitors who visit urban and suburban destinations, our analysis that concerns visitors with identifiable accommodation anchor (15.2% of total visitors) shows that they are more involved in *chain-based* tours than *hub-and-spoke* tours compared to other visitor types (see Fig. 15). In this regard, synergies among different transport modes can assist in catering to their complex mobility needs. Mobility-as-a-service (MaaS) can address challenges in relation to visitors' transportation in combined urban and rural areas by integrating different modes such as PT, shared e-scooters and bicycles, as well as ride-hailing services. Such services can provide seamless transportation services among tourist attractions that are not served by PT directly. However, challenges related to intermodal integration, data sharing, and stakeholder collaboration still remain, complicating the implementation of tourism MaaS (Leung et al., 2023).

In terms of limitations for this study, socioeconomic data for visitors and PT trip purposes were not available. Therefore, such attributes were not included in the clustering process. The inclusion thereof could further enrich the defined temporal and spatial clusters. In addition, in this study, visitors with a single PT journey were excluded from the analysis, given the inherent inability to infer their final destinations and therefore to track those visitors in the PT system, using the adopted inference framework. This exclusion may introduce a sample selection bias since a portion of genuine visitors (single-journey visitors) is overlooked. Consequently, visitors' locally induced passenger demand may be underestimated, especially in areas with tourist activity. Furthermore, given the lack of a unique way to identify visitors unambiguously while completely eliminating the inclusion of "pseudo-tourists", our analysis may include a small portion of pseudo-tourists. However, the Swedish public transport barometer on traveller categories reports the existence of a share of infrequent travellers for Region Stockholm (Svenskkollektivtrafik, 2025), supporting the notion that some local users exhibit travel patterns similar to those of visitors. Lastly, regarding the analysis identifying visitors' accommodation anchors, only a minority of the sample (15.2%) is included. Therefore, the relevant findings should be interpreted with caution and viewed as indicative rather than generalizable, given the sample limitation.

We identify the following research directions as potential contributions to the transport geography and travel behavior research fields that are worth exploring. First, research on visitors' mobility combining smart card data with survey data can offer additional insight related to visitors' socioeconomic background and trip purpose, overcoming our study's limitation. Such insight can facilitate tailored policy-making in relation to new fare structures (e.g., dedicated fare products for youth visitors), therefore increasing PT attractiveness among specific demographics within the visitors market. Besides survey data, the use of mobile phone data has opened new opportunities for visitor mobility research. In particular, mobile phone data can enrich visitor profiling by uncovering complete visitors' trajectories, which are difficult to capture with other data sources (Ke et al., 2026). Furthermore, seasonality and its impact on the PT system's performance are important aspects given their significant impact on the quality of PT services. Last, analyzing level of service indicators such as the waiting time, travel time, and on-board crowding during periods of high and low visitor volumes can reveal the real impact of tourism on PT operations.

CRediT authorship contribution statement

Anastasios Skoufas: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Matej Cebecauer:** Writing – review & editing, Methodology, Conceptualization. **Wilco Burghout:** Writing – review & editing. **Erik Jenelius:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Oded Cats:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to express their gratitude to the Swedish Transport Administration (Trafikverket) for funding this study through the "Demand management for reduced congestion in urban traffic from groups of travellers" research project under Grant no TRV 2024/114947. Lastly, the authors extend their appreciation to the Transport Administration in Region Stockholm for providing the necessary data.

Data availability

The authors do not have permission to share data.

References

Albalate, D., Bel, G., 2010. Tourism and urban public transport: Holding demand pressure under supply constraints. Tour. Manag. 31 (3), 425–433. http://dx.doi.org/10.1016/j.tourman.2009.04.011.

Benito, J.A.D., Zerva, K., 2025. From overtourism to overall-mobility. Ann. Tour. Res. Empir. Insights 6 (1), 100161. http://dx.doi.org/10.1016/j.annale.2024.100161.

Cats, O., 2023. Identifying human mobility patterns using smart card data. Transp. Rev. 44 (1), 213–243. http://dx.doi.org/10.1080/01441647.2023.2251688.

Cats, O., Rubensson, I., Cebecauer, M., Kholodov, Y., Vermeulen, A., Jenelius, E., Yusak, S., 2019. FairAccess - How Fair Is the Fare? Estimating Travel Patterns and the Impacts of Fare Schemes for Different User Groups in Stockholm Based on Smartcard Data. tech. rep., https://kth.diva-portal.org/smash/get/diva2:1784265/FULLTEXT01.pdf.

Coppola, P., Carbone, A., Aveta, C., Stangherlin, P., 2020. Assessing transport policies for tourist mobility based on accessibility indicators. Eur. Transp. Res. Rev. 12 (1), 56. http://dx.doi.org/10.1186/s12544-020-00444-4.

Das, P., Maitra, S., 2025. Improvement of pedestrian facilities at tourist destinations based on tourists' satisfaction in the context of walk. Case Stud. Transp. Policy 21, 101505. http://dx.doi.org/10.1016/j.cstp.2025.101505.

De Cantis, S., Parroco, A.M., Ferrante, M., Vaccina, F., 2015. Unobserved tourism. Ann. Tour. Res. 50, 1–18. http://dx.doi.org/10.1016/j.annals.2014.10.002.

Domènech, A., Miravet, D., Gutiérrez, A., 2020. Mining bus travel card data for analysing mobilities in tourist regions. J. Maps 16 (1), 40–49. http://dx.doi.org/10.1080/17445647.2019.1709578.

Dominik, B., Carsten, S., Claudia, W., 2017. Uncommon leisure traffic – Analyses of travel behaviour of visitors. Transp. Res. Procedia 25, 3971–3984. http://dx.doi.org/10.1016/j.trpro.2017.05.236.

Egu, O., Bonnel, P., 2020. Investigating day-to-day variability of transit usage on a multimonth scale with smart card data. A case study in Lyon. Travel Behav. Soc. 19, 112–123. http://dx.doi.org/10.1016/j.tbs.2019.12.003.

Eurostat, Glossary:Tourist. https://ec.europa.eu/eurostat/statistics-explained/SEPDF/cache/1755.pdf?v=6261139684356968.

Eurostat, 2026. Tourism statistics - seasonality at regional level. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Tourism_statistics_-_seasonality_at_regional_level.

Ghaemi, M.S., Agard, B., Trépanier, M., Partovi Nia, V., 2017. A visual segmentation method for temporal smart card data. Transp. A: Transp. Sci. 13 (5), 381–404. http://dx.doi.org/10.1080/23249935.2016.1273273.

Graham Mandeno, T., 2012. Is Tourism a Driver for Public Transport Investment? (Doctoral dissertation). University of Otago, https://ourarchive.otago.ac.nz/esploro/outputs/graduate/Is-Tourism-a-Driver-for-Public/9926479664001891.

Gutiérrez, A., Domènech, A., Zaragozí, B., Miravet, D., 2020. Profiling tourists' use of public transport through smart travel card data. J. Transp. Geogr. 88, 102820. http://dx.doi.org/10.1016/j.jtrangeo.2020.102820.

Hall, C.M., Le-Klähn, D.-T., Ram, Y., 2017. Tourism, Public Transport and Sustainable Mobility, vol. 4, Channel View Publications, https://books.google.se/books?hl=en&lr=&id=GJfzDQAAQBAJ&oi=fnd&pg=PT9&ots=zJX4UczkNQ&sig=64qORpY8vWaKapYe8WsYJTBbvBY&redir_esc=y#v=onepage&q&f=false.

Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136: A K-means clustering algorithm. J. R. Stat. Soc. Ser. C (Appl. Stat.) 28 (1), 100–108. http://dx.doi.org/10.2307/2346830.

He, L., Agard, B., Trépanier, M., 2020. A classification of public transit users with smart card data based on time series distance metrics and a hierarchical clustering method. Transp. A: Transp. Sci. 16 (1), 56–75. http://dx.doi.org/10.1080/23249935.2018.1479722.

Jażdżewska-Gutta, M., Szmelter-Jarosz, A., Borkowski, P., 2023. Micromobility in tourist single- and multimodal travels at destination. Res. Transp. Bus. Manag. 48, 100956. http://dx.doi.org/10.1016/j.rtbm.2023.100956.

Jover, J., Díaz-Parra, I., 2022. Who is the city for? Overtourism, lifestyle migration and social sustainability. Tour. Geogr. 24 (1), 9–32. http://dx.doi.org/10.1080/14616688.2020.1713878.

Kadir, R.A., Shima, Y., Sulaiman, R., Ali, F., 2018. Clustering of public transport operation using K-means. In: 2018 IEEE 3rd International Conference on Big Data Analysis. ICBDA, pp. 427–432. http://dx.doi.org/10.1109/ICBDA.2018.8367721.

Kashfi, S.A., Bunker, J.M., Yigitcanlar, T., 2015. Understanding the effects of complex seasonality on suburban daily transit ridership. J. Transp. Geogr. 46, 67–80. http://dx.doi.org/10.1016/j.jtrangeo.2015.05.008.

Ke, R., Wang, K., Liu, Q., Liu, L., Wang, P., 2026. Uncovering urban tourist mobility patterns based on large-scale mobile phone data. J. Transp. Geogr. 130, 104496. http://dx.doi.org/10.1016/j.jtrangeo.2025.104496.

Kemperman, A., 2021. A review of research into discrete choice experiments in tourism: Launching the annals of tourism research curated collection on discrete choice experiments in tourism. Ann. Tour. Res. 87, 103137. http://dx.doi.org/10.1016/j.annals.2020.103137.

Klepej, D., Marot, N., 2024. Considering urban tourism in strategic spatial planning. Ann. Tour. Res. Empir. Insights 5 (2), 100136. http://dx.doi.org/10.1016/j.annale.2024.100136.

Koens, K., Postma, A., Papp, B., 2018. Is overtourism overused? Understanding the impact of tourism in a city context. Sustainability 10 (12), http://dx.doi.org/10.3390/su10124384.

Le-Klähn, D.-T., Gerike, R., Michael Hall, C., 2014. Visitor users vs. non-users of public transport: The case of Munich, Germany. J. Destin. Mark. Manag. 3 (3), 152–161. http://dx.doi.org/10.1016/j.jdmm.2013.12.005.

Le-Klähn, D.-T., Hall, C.M., 2015. Tourist use of public transport at destinations – a review. Curr. Issues Tour. 18 (8), 785–803. http://dx.doi.org/10.1080/13683500.2014.948812.

Leung, A., Burke, M., Scott, P., 2023. Tourism MaaS – The case for regional cities. Res. Transp. Bus. Manag. 49, 101017. http://dx.doi.org/10.1016/j.rtbm.2023.101017.

Li, C., Ge, P., Liu, Z., Zheng, W., 2020. Forecasting tourist arrivals using denoising and potential factors. Ann. Tour. Res. 83, 102943. http://dx.doi.org/10.1016/j.annals.2020.102943.

Lin, Y., Xu, Y., Zhao, Z., Park, S., Su, S., Ren, M., 2023. Understanding changing public transit travel patterns of urban visitors during COVID-19: A multi-stage study. Travel Behav. Soc. 32, 100587. http://dx.doi.org/10.1016/j.tbs.2023.100587.

Liu, X., Huang, Z., Jian, W., 2025. Nonlinear effects of multilevel factors on public transport commuting in China's cities. Transp. Res. D 143, 104724. http://dx.doi.org/10.1016/j.trd.2025.104724.

Lu, Y., Wu, H., Liu, X., Chen, P., 2019. TourSense: A framework for tourist identification and analytics using transport data. IEEE Trans. Knowl. Data Eng. 31 (12), 2407–2422. http://dx.doi.org/10.1109/TKDE.2019.2894131.

Nilsson, J.H., 2020. Conceptualizing and contextualizing overtourism: the dynamics of accelerating urban tourism. Int. J. Tour. Cities 6 (4), 657–671. http://dx.doi.org/10.1108/IJTC-08-2019-0117.

Nordregio, 2023. Domestic tourism in Sweden. https://pub.norden.org/temanord2023-519/annex-6-country-report-sweden.html.

OpenStreetMap contributors, 2026. OpenStreetMap data: Points of interest for Stockholm extracted via overpass turbo. https://overpass-turbo.eu/.

Oppermann, M., 1999. Predicting destination choice — A discussion of destination loyalty. J. Vacat. Mark. 5 (1), 51–65. http://dx.doi.org/10.1177/135676679900500105.

Ortega-Tong, M.A., 2013. Classification of London's Public Transport Users Using Smart Card Data (Doctoral dissertation). p. 163, https://scholar.google.com/scholar_lookup?title=ClassificationofLondonsPublicTransportUsersUsingSmartCardData&publication_year=2013&author=M.A.Ortega-Tong.

Page, S., Ge, Y., 2009. In: Jamal, T., Robinson, M. (Eds.), Transportation and Tourism: A Symbiotic Relationship? BT - The SAGE Handbook of Tourism Studies. SAGE Publications Ltd, http://dx.doi.org/10.4135/9780857021076.n21.

Pagliara, F., Aria, M., Mauriello, F., 2025. Chapter 2 - Transportation choice in tourism and travel: A bibliometrics literature review. In: Pagliara, F., Aria, M., Mauriello, F. (Eds.), Models and Applications of Tourists' Travel Behavior. Elsevier, pp. 21–46. http://dx.doi.org/10.1016/B978-0-443-26593-8.00002-0.

Papadopoulou, N.M., Ribeiro, M.A., Prayag, G., 2023. Psychological determinants of tourist satisfaction and destination loyalty: The influence of perceived overcrowding and overtourism. J. Travel Res. 62 (3), 644–662. http://dx.doi.org/10.1177/00472875221089049.

Raun, J., Ahas, R., Tiru, M., 2016. Measuring tourism destinations using mobile tracking data. Tour. Manag. 57, 202–212. http://dx.doi.org/10.1016/j.tourman.2016.06.006.

Romão, J., Bi, Y., 2021. Determinants of collective transport mode choice and its impacts on trip satisfaction in urban tourism. J. Transp. Geogr. 94, 103094. http://dx.doi.org/10.1016/j.jtrangeo.2021.103094.

Skoufas, A., Cebecauer, M., Burghout, W., Jenelius, E., Cats, O., 2024. Assessing contributions of passenger groups to public transportation crowding. J. Public Transp. 26, 100110. http://dx.doi.org/10.1016/j.jpubtr.2024.100110.

Skoufas, A., Cebecauer, M., Burghout, W., Jenelius, E., Cats, O., 2025. Ex-post assessment of public transportation on-board crowding induced by new urban developments. Cities 165, 106093. http://dx.doi.org/10.1016/j.cities.2025.106093.

Statista, 2024a. Number of arrivals in tourist accommodation in Sweden from 2010 to 2023. https://www.statista.com/statistics/413261/number-of-arrivals-spent-in-short-stay-accommodation-in-sweden/.

Statista, 2024b. Number of employed persons in the travel and tourism industry in Sweden from 2021 to 2023, by sector. https://www.statista.com/statistics/1563117/travel-and-tourism-employment-by-sector-in-sweden/.

Statista, 2024c. Total contribution of travel and tourism to gross domestic product in Sweden from 2019 to 2024. https://www.statista.com/statistics/790670/travel-tourism-total-gdp-contribution-sweden/.

Statista, 2025. Travel and tourism in Sweden - statistics & facts. https://www.statista.com/topics/6742/tourism-in-sweden/#topicOverview.

Stockholms Lokaltrafik (SL), Visitor tickets - Travelcards. https://sl.se/en/fares-and-tickets/visitor-tickets/travelcards.

Svenskkollektivtrafik, 2025. Kollektivtrafikbarometern (Public Transport Barometer). tech. rep., https://svenskkollektivtrafik.se/app/uploads/2025/02/Arsrapport-Kollektivtrafikbarometern-2024.pdf.

Swedish Agency for Economic and Regional Growth, 2020. Facts about Swedish Tourism. tech. rep., Swedish Agency for Economic and Regional Growth, Stockholm, Sweden, https://tillvaxtverket.se/download/18.6855bfcf184896002ffbcc/1668765860965/FaktaomsvenskturismTA_Slutleverans.pdf.

Swedish Agency for Economic and Regional Growth, 2025. Tourism statistics. https://tillvaxtdata.tillvaxtverket.se/statistikportal#page=72b01aa0-1d4a-425c-8684-dbce0319b39e.

The Agency for Economic Growth (Tillväxtverket), 2025. Accommodation statistics. https://tillvaxtdata.tillvaxtverket.se/statistikportal#page=72b01aa0-1d4a-425c-8684-dbce0319b39e.

The World Bank, 2025. Tourism and competitiveness. https://www.worldbank.org/en/topic/competitiveness/brief/tourism-and-competitiveness#:~:text=In2024%2Ctourismaccountedfor,surged%2CreachingUS%241.9trillion.

Visit Stockholm, 2023. Stockholm and Our Visitors. tech. rep., Visit Stockholm AB, Stockholm, Sweden, https://professionals.visitstockholm.com/documents/223/Stockholm_and_our_visitors_2023_ENG.pdf.

World Travel & Tourism Council, 2025. Travel & tourism to create 4.5MN new jobs across the EU by 2035. https://wttc.org/news/travel-and-tourism-to-create-4-5mn-new-jobs-across-the-eu-by-2035.

Xue, M., Wu, H., Chen, W., Ng, W.S., Goh, G.H., 2014. Identifying tourists from public transport commuters. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '14, Association for Computing Machinery, New York, NY, USA, pp. 1779–1788. http://dx.doi.org/10.1145/2623330.2623352.

Zientara, P., Jażdżewska-Gutta, M., Bąk, M., Zamojska, A., 2024. What drives tourists' sustainable mobility at city destinations? Insights from ten European capital cities. J. Destin. Mark. Manag. 33, 100931. http://dx.doi.org/10.1016/j.jdmm.2024.100931.