Gutiérrez, A.; Domènech, A.; Zaragozí, B. & Miravet, D. (2020). Profiling tourists' use of public transport through smart travel card data. *Journal of Transport Geography*, 88. https://doi.org/10.1016/j.jtrangeo.2020.102820

Profiling tourists' use of public transport through smart travel card data

Aaron Gutiérrez[a], Antoni Domènech[a], Benito Zaragozí[a], Daniel Miravet [bc]

[a] Universitat Rovira i Virgili, Department of Geography, C/ Joanot Martorell, 15 43480 Vila-seca, Spain
[b] Consortium of Public Transport of Camp de Tarragona, C/ Anselm Clavé, 1 43004 Tarragona, Spain
[c] Universitat Rovira i Virgili, Department of Economics, Research Center on Economics and Sustainability (ECO-SOS), Av. Universitat, s/n 43204, Reus, Spain

Highlights

- Smart card data are used to define public transport passenger profiles.
- The study is based on a coastal tourist destination during peak season.
- The analysis focuses on travellers using a multipersonal transport fare.
- Model-based clustering is used to identify passenger profiles.
- The method provides useful information for public transport policy and management.

Abstract

Data collected through smart travel cards in public transport networks have become a valuable source of information for transport geography studies. During the last two decades, a growing body of literature has used this sort of data source to study the behaviour of public transport users in cities and regions around the world. However, its use has been scarce in contexts where public transport demand is highly influenced by the activities of the tourist sector. Therefore, it remains to be seen whether these data can be leveraged to optimize the supply of public transport. In this article, data drawn from the Camp de Tarragona automated fare collection system extracted during 2018 are used to study tourists' use of public transport in Costa Daurada (Catalonia, Spain). This is a popular coastal destination with a high concentration of visitors during the summer period. The analysis focuses on the use of the T-10, a multipersonal transport fare with no time limitations on its use which makes it appealing for tourists. Model-based clustering has been applied to identify different clusters of passengers according to their activity and spatial profiles. Differences between profiles are significant and, as a result, this study allowed the validation of a method that could be replicated in other contexts, as it provides highly useful information for public transport policy and mobility management.

Keywords: smart card data; public transport; tourism destination; travel behaviour; tourist profiles; model-based clustering analysis

1. Introduction

1.1. Tourism and public transport networks

The importance of transport networks and infrastructure is essential in tourism development (Duval, 2007; Prideaux, 2000). In recent years, a growing body of literature has been focusing on the role of local transport networks and, more specifically, on public transport systems, to enhance a more sustainable tourism development (Le-Klähn & Hall, 2015). The synergies are bidirectional. Public transport is important for tourism, but tourism can also play a role in supporting the provision of public transport services (Hall et al., 2017). These synergies also feed evident challenges that need to be addressed by local policy-makers. The continuous growth of the number of visitors in tourist cities and regions could imply extra resources for public transport networks but also additional pressure for the transport system (Albalate & Bel, 2010). These and other research topics related to tourist mobility have benefited from the advances experienced during recent years in the domain of location technologies. Increasingly, researchers can take advantage of large volumes of automatically generated geodata to examine these questions. Indeed, big data coming from tracking devices, sensor-based smartphone applications, mobile phone traces and smart travel cards, among others (Kwan & Schwanen, 2017), provide high-resolution spatial and temporal information with no subjective bias related to human collection.

1.2. Smart travel cards and transport studies

Automatic fare collection systems through smart travel cards became popular during the 1990s, but a proliferation in managing payments in public transport networks around the world took place especially during the 2000s. Octopus card in Hong Kong, Navigo in Paris, Oyster in London, Bip! in Santiago, and OV-Chip in the Netherlands are well-known examples of the implementation of these systems. Each fare collection system has its own data collection method, but what is common in all of them is that two kinds of information are automatically collected: the monetary transactions when an individual acquires a card or adds credit, and journey transactions when passengers board buses or enter the metro/train stations (El Mahrsi et al., 2017). The transaction data collected are valuable for transport geography research, since they provide comprehensive and continuous spatial and temporal information of all public transport users in a city or region. Consequently, the monitoring of travel behaviour allows transport authorities to detect the weaknesses of transport networks, to analyse the typology and mobility patterns of users, and to adopt specific measures for the optimization of the transport system. In this regard, a growing body of literature has developed providing approaches and methods to mine this data source (Bagchi & White, 2005; Kurauchi & Schmöcker, 2017; Pelletier et al., 2011). However, despite this methodological richness, a notorious lack of research exists reviewing the potentialities that this data source offers for public transport management in tourist contexts. To the best of the authors' knowledge, only Ortega-Tong (2013) for London and Lu et al. (2019) and Xue et al. (2014) for Singapore have been able to identify visitors after analysing smart card data. While most studies put the focus solely on the identification of recurrent users (Ghaemi et al., 2016).

The lack of research exploring the benefits of smart travel card data for public transport studies within tourist cities and regions contrasts with the increasing complexity of these contexts. In tourist areas public transport is used by both residents and a growing number of visitors. However, the volume of users is not constant in time but rather is highly influenced by the arrival of visitors. Therefore, the demand is uncertain (Domènech & Gutiérrez, 2017; Hall et al., 2017) and could evolve to where it could even fully reshape the mobility patterns of a whole area (Gutiérrez & Miravet, 2016a). As a consequence, the achievement of an optimal fit between supply and demand becomes a challenge in itself.

In contexts where seasonality is an idiosyncratic element of the tourism activity, the availability of data capturing the dynamics of public transport flows is critical to properly forecast demand at different times (Domènech et al., 2020). Public transport authorities should be able to manage the combined flows of the locals, which tend to be quite constant throughout the year, with the visitor flows, which alternatively tend to be irregular. If supply design does not optimally suit these combined demands, capacity problems could arise, which could seriously damage the users' perceived service quality or even the reputation of the tourist destination (Albalate & Bel, 2010).

However, the greater degree of complexity of the management of public transport within tourist cities and regions is not only caused by seasonal oscillations of demand; it is also connected to different patterns of use between residents and non-residents. These divergences have been identified by works using smart travel card data and mainly embrace the geographical patterns of use (Gutiérrez & Miravet, 2016a; Wu et al., 2015), the distribution of demand throughout the day (Domènech et al., 2020) as well as fare preferences (Wu et al., 2015). As a result, public transport authorities must properly meet the needs of groups that tend to show clearly divergent demands. This is a demanding challenge for public transport providers, since visitors tend to enjoy a higher degree of choice between transport modes rather than locals (Gronau and Kagermeier, 2007). Therefore, an optimal coverture of visitors' catchment area along with a high-quality service that meets their needs are crucial to attract them to public transport.

In this sense, the use of public transport by visitors is also conditioned by the transport fares offered. For instance, multipersonal tickets are good value for money, as they allow the use of public transport by groups with just one card. Usually, the physical form of these fare types are cards allowing ten transactions with a time validity that clearly exceeds the visitors' length of stay (i.e. the end of the year). It is not a rare system in Europe, although many cities or metropolitan areas (e.g. London, Rome, Vienna) have chosen to offer unipersonal tickets, which allow their owners unlimited use of the local public system during a limited period of time. Other cities let their visitors to choose between the unipersonal and multipersonal options. For instance, the Paris fare system allows the acquisition of the "carnet de 10", ten single tickets with a price reduction, and at the same time the "Paris Visite", a unipersonal card that ranges from one to five days. In Madrid, the "Metrobus" multipersonal ten-journey card also coexists with the unipersonal day tickets (1,2,3,4,5 or 7 days of validity). Barcelona has discarded its traditional T-10 (multipersonal ticket with ten journeys), replacing it with the "T-familiar" (only 8 journeys that can be used by groups and within a 30-day time span) and the "T-casual" (a unipersonal 10 journey transport ticket with no time restrictions). Berlin offers both a unipersonal day pass and a multipersonal day pass (limited to five people). However, not only urban areas embrace specific public transport fares oriented to visitors; it also occurs in non-urban tourist destinations. Cote d'Azur offers the multipersonal fare “MULTI 10”, allowing ten journeys with no time restrictions. In the island of Mallorca the smart card can be recharged with the T-20, a multipersonal pass that allows 20 transactions. In Malta, “Explora” (unipersonal fare allowing unlimited journeys over 7 days) coexists with a multipersonal fare with 12 journeys with just one day of validity). In the metropolitan area of Grenoble, a touristic mountain environment, Transisère replaced in 2012 the “la carte 10 trajets” (10 journeys) by the “la carte 6 trajets” (6 journeys). Other popular tourist destinations do not offer fares that are attractive for visitors. This is the case of Tenerife and Gran Canaria (Spain), Antalya (Turkey) or Rimini Riviera (Italy).

1.3. Objectives

One of the fundamental challenges of tourist cities and regions is the management of mobility flows of seasonal populations to minimize their externalities. Therefore, it is key to acquire knowledge about the mobility patterns of tourists and attract/redirect them towards more sustainable modes of transport. Accordingly, this article contributes to this demanding need of tourist regions through the provision of tools that will help researchers and planners to untangle visitors using public transport. Considering the need for accurate information, especially in highly dynamic scenarios affected by tourism, and the potential interest of smart travel card data, the main objective of this article is to identify profiles of tourists who use public transport during their stay at a destination and characterise their distinctive use of it. To this end, data referring to 2018 and collected by the Camp de Tarragona (South Catalonia, Spain) automated fare collection system has been used. More concretely, smart cards corresponding to the T-10 multipersonal transport fare that were active only during the summer season were analysed. The novelty and interest of this data extraction lies not only in this fare being the most-used transport fare in the region but also in it being the most attractive for seasonal users such as tourists. The region under analysis hosts the Costa Daurada tourism brand, one of the most important Spanish Mediterranean coastal destinations. Therefore, it is characterised by the seasonal nature of its tourism activity.

In order to detect public transport users' profiles, this study is based on latent profile analysis (LPA), a model-based clustering technique using the expectation–maximization algorithm. Two different cluster analyses have been implemented considering, first, the activity and, second, the spatial variables of the smart cards under study. Then the relations between both clusters have been explored to define the spatiotemporal behaviour of the different profiles identified. Overall, clustering techniques have proven to be effective in identifying patterns in smart card datasets (Liu et al., 2020; Liang et al., 2018; Manley et al., 2018; Faroqi et al., 2017; Briand et al. 2017; Trépanier al., 2012). However, none of the published studies has explicitly mentioned the use of the model-based technique used in the present study (LPA). It is for this reason that a secondary and more-instrumental objective of this research is to validate a method that could be replicated elsewhere.

From this introduction, the article is organised as follows: the second section introduces the area of study and explains how the data have been obtained and processed. In the third section, the methods are explained and justified. The results of the study are shown in the fourth section. The discussion of our findings is presented in the fifth section, while the last section includes the concluding remarks of the article.

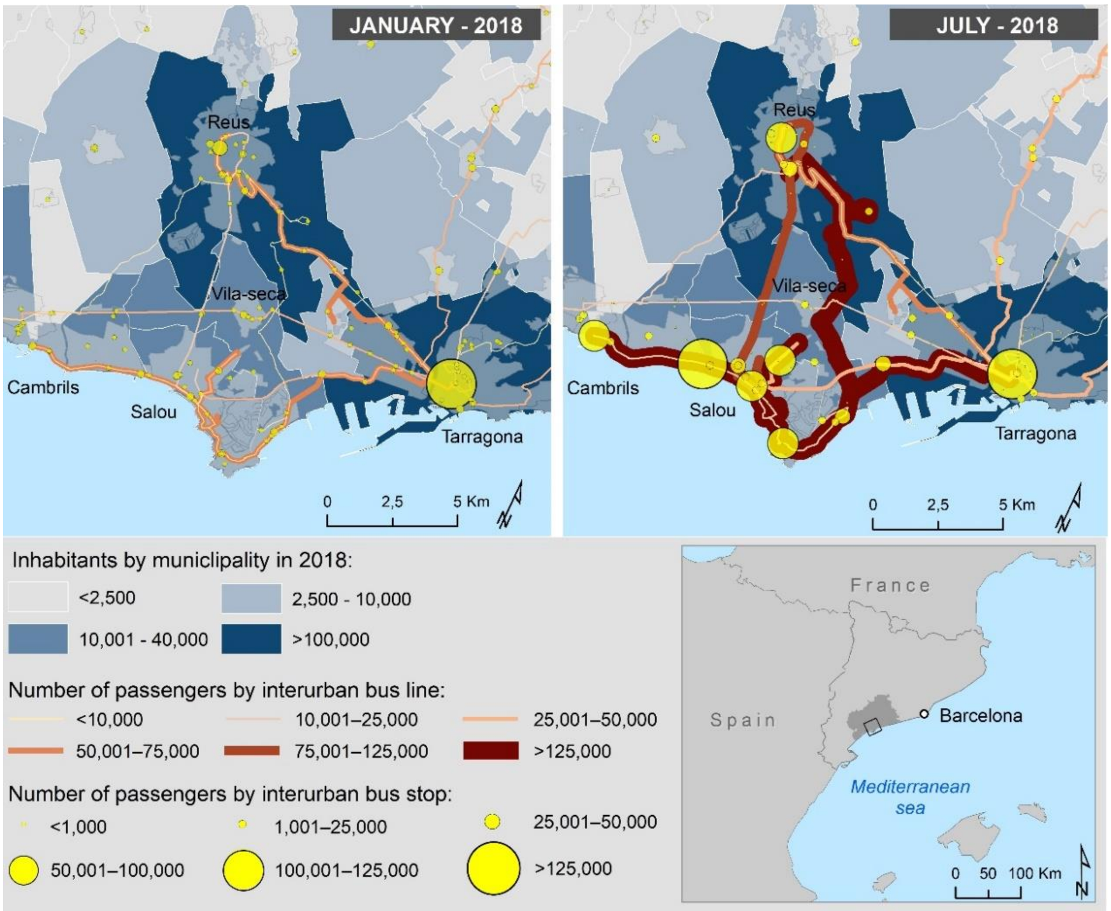
2. Data

2.1. Study area

The chosen study area is Camp de Tarragona, a region in the northeast of Spain belonging to Catalonia, with a population of 617,504 inhabitants (INE, 2018) and where a public transport fare integration system has been implemented through the use of transport smart cards. For more than a decade, this region has collected all the necessary data that could allow the identification of tourist profiles and study how they use public transport. The Territorial Mobility Authority of Camp de Tarragona (henceforth ATM Camp de Tarragona) is the consortium responsible for the management of public transport concessions in this area and manages the information system providing data for this research.

The Camp de Tarragona region has an uneven spatial distribution of economic activity and population, most of which is concentrated along the coast. Tarragona and Reus (with populations of over 100,000 each) and three coastal towns of Cambrils, Salou and Vila-seca (whose populations exceed 20,000 each) constitute the main regional functional centrality. These three coastal municipalities constitute the core of one of the most important Spanish tourism brands, the Costa Daurada, and account for 77% of the total hotel capacity in Camp de Tarragona. According to data provided by the Tourism Observatory of Catalonia, these three municipalities hosted more than 20.2 M overnight stays in regulated accommodation in 2018. The tourism activity within the area is markedly seasonal. Two-thirds of total overnight stays take place during the summer months. Furthermore, since 1995, the presence of PortAventura (one of the largest theme parks in Europe) in the municipalities of Vila-seca and Salou also boosts tourism in the region and shapes the territorial dynamics with an attraction of more than 3.5 M visitors per year (Rubin, 2015). Considering the characteristics of the region and the concentration of tourism activity during summer, the most significant challenge for public transport provision in the area is the management of the unbalanced seasonal demand (see Figure 1).The seasonal behaviour of the public transport demand in the region is described in detail in Domènech et al. (2020).

Figure 1. Public transport transactions in Costa Daurada (January and July 2018).



Source: Own research with data provided by ATM Camp de Tarragona.

2.2. Data pre-processing and enrichment

The automated fare collection system of the ATM Camp de Tarragona continuously collects smart card transactions done on board interurban bus services, which altogether represent the main mode of public transport used to interconnect the cities and towns in the region. The fare-managing system of the integrated transport network, SIGIT (from the acronym in Catalan: “Sistema de Gestió de la Integració Tarifaria”), stores information regarding the time at which the passenger boarded the vehicle, the location of the bus stop where the user got on board, the bus line identification and the type of transport fare.

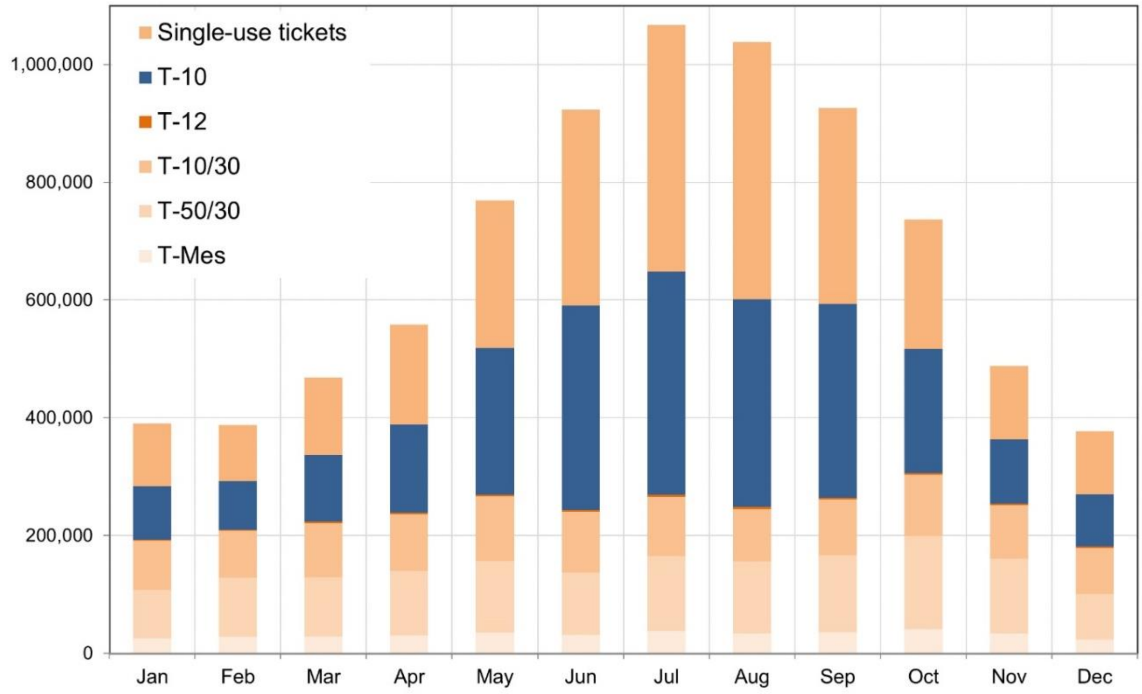
The files generated by SIGIT lack a well-structured format and not all the data collected by the system was relevant to this study. For these reasons, a pre-processing phase was carried out to create an adequate database to carry out different types of spatio-temporal aggregations (Zaragozí et al. 2020). The logs generated by SIGIT (9.1 GB in CSV) were imported into an SQL database which is a convenient framework for analysing medium to large smart transport card data sets (Li et al., 2018).

In this study, all the records collected during 2018 were analysed. This implies 5,414,028 interurban travel transactions of 135,365 different smart cards, of which 98.3% were used more than once, and 2,725,852 transactions of single-use tickets. The records were ordered by date and time of the transaction, the code of the card that made the transaction and the type of transaction carried out. Only essential attributes of the smart cards were selected. The resulting database organises the records into different tables following a simple relational structure with a main transactions table linked to other reference tables with details of the main entities (agencies, fares, stops, municipalities and routes). In addition to the data provided by the ATM Camp de Tarragona, the database was enriched with layers of geographic information: (a) municipalities, roads and coastline, downloaded from the Institut Cartogràfic i Geològic de Catalunya (www.icgc.cat) and adapted to the needs of this research; (b) stops and routes were manually digitised; municipal population data for 2018 (https://www.idescat.cat/); (c) and the zoning of the ATM Camp de Tarragona in service areas.

2.3. Knowledge-based sampling

As explained in the introduction, the existence of multipersonal fares is a key element that shapes the demand of public transport in tourist destinations. T-10 is the unique multipersonal fare in the ATM Camp de Tarragona system. It allows travelling in a group (consecutive transactions when boarding). The standard T-10 card covers ten transactions. It can be recharged as many times as wanted, but a maximum accumulation of 30 transactions at a time is allowed. The other ATM fares are unipersonal: T-10/30 (10 transactions maximum in 30 consecutive days), T-50/30 (50 transactions maximum in 30 consecutive days), T-MES (unlimited transactions in 30 consecutive days), T-12 (special fare for children up to 12 years old), and T-70/90 (large families). Figure 2 shows the distribution of transactions by months and fares in 2018 in the region. From the figure, it is possible to observe that the increase of transactions during summer is concentrated in single-use tickets and T-10 fares.

Figure 2. Interurban bus transactions by months and fare (2018).



Note: T-70/90 is not represented in the figure due to the low number of monthly transactions.

Source: Own research with data provided by ATM Camp de Tarragona.

As already shown, three factors shape the use of public transport in the study area: the concentration of tourist arrivals during summer; the increasing volume of travellers on interurban public transport during that period (Figure 2); and their spatial concentration in the tourist municipalities (Figure 1). Taking these factors into account, it makes sense to analyse the types of transport fares that are likely to be used mainly by visitors during the high season.

Despite their interest in describing the global situation, it is obvious that single-use tickets cannot be tracked, and their analysis would not provide any information to study the profiles of those users. Among the smart cards that registered 100% of their transactions in summer (n=37,393), 92.6% were T-10 (n=34,641), which represents 90% of the transactions performed with cards used only in summer (Table 1). Transactions from T-10 cards used only in summer represent 11% of all yearly transactions and deserve to be analysed separately from the rest of unipersonal fares used only in summer (7.4% of the cards and 10% of the transactions). These unipersonal fares could correspond to other profiles, such as local residents travelling for leisure or holiday purposes in summer and temporary workers in the tourism sector, among others.

Table 1. Number of cards exclusively used in summer and their volume of transactions, showing the preponderance of T-10 over the remaining fees.

Cards used only in summer	Cards	Transactions
All fares	37,393	665,562
T-10	34,641 (92.6%)	599,028 (90%)

Source: Own research with data provided by ATM Camp de Tarragona.

The preponderant role of T-10 fares during the summer period should not mislead one to assume that the use of T-10 is only important during the peak tourist season. As Figure 2 reveals, its use is also widespread during the rest of the year. Its characteristics are attractive for visitors who visit this area outside the summer months and for some segments of the resident population as well. Indeed, its fees are appealing for those who are not frequent public transport users, and for those who do not travel alone. Nevertheless, as this study is focused on profiling tourists' using public transport, we decided to concentrate the analysis on the peak season.

3. Method

3.1. Extraction and selection of indicators

Specific SQL queries were developed to extract 14 simple but meaningful indicators for each card. These were related to the level of activity of each user and its spatial distribution (Table 2). These indicators are aggregation queries that could be calculated virtually for any other dataset collected through a smart card system. These indicators separate between activity and spatial variables, depending on their intrinsic value for describing temporal or spatial activity patterns at different granularity levels. Carrying out these analyses separately makes the behaviour of each variable clearer and thus obtains more useful results. This approach using separate groups between different types of variables was also used in previous studies employing data mining and clustering techniques to identify traveller profiles from smart travel card data (Manley et al., 2018; El Mahrsi et al., 2017; Briand et al., 2017).

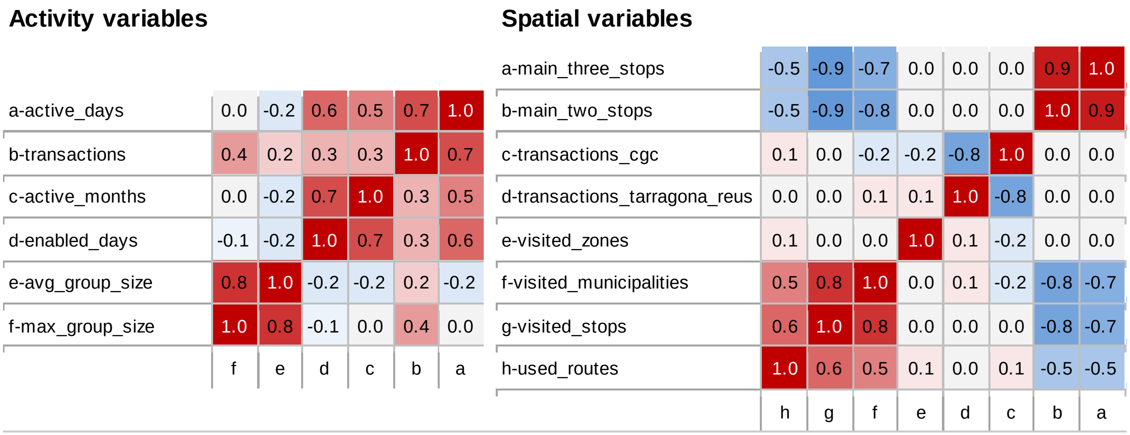
Table 2. Descriptive statistics of activity and spatial-related variables derived from the processed dataset (T-10 cards exclusively used in summer).

Type	Name	Description (range)	Mean (SD)	Median [min. max.]
Target	card	Grouping variable (n=34,641)	-	-
Activity	transactions	Total number of transactions (n=599,028)	17.3 (14.4)	11 [1, 288]
	enabled_days	Number of days between the first and the last day the card was used	9.0 (13.0)	5 [1, 93]
	active_days	Number of days the card was used	3.6 (3.0)	3 [1, 51]
	active_months	Number of Months the card was used	1.2 (0.5)	1 [1, 4]
	avg_group_size	Average number of consecutive transactions in any stop	2.6 (1.5)	2 [1, 30]
	max_group_size	Maximum number of consecutive transactions in any stop	3.3 (1.9)	3 [1, 30]
Spatial	visited_stops	Number of stops visited during the entire period	3.7 (1.7)	4 [1, 14]
	visited_municipalities	Number of municipalities visited during the entire period	2.9 (1.2)	3 [1, 8]
	visited_zones	Number of zones visited during the entire period	1.0 (0.1)	1 [1, 3]
	used_routes	Number of routes used during the entire period	1.9 (0.7)	2 [1, 9]
	main_two_stops	Percentage of transactions concentrated in the two most visited stops	75.0% (19.8)	75% [20,100]]
	main_three_stops	Percentage of transactions concentrated in the three most visited stops	88.0% (13.9)	94% [30, 100]
	transactions_tarragona_reus	Percentage of transactions concentrated in the main cities —Tarragona and Reus— over 50,000 inhabitants	19.9% (22.5)	15% [0, 100]
	transactions_cgc	Percentage of transactions concentrated in the main touristic cities —Cambrils, Salou and Vila-seca—, between 20,000 and 50,000 inhabitants.	72.0% (25.7)	75% [0,100]

Source: Own research with data provided by ATM Camp de Tarragona.

Before carrying out a clustering algorithm, the redundancy between the extracted indicators has to be checked. In this regard, correlation matrices were calculated, and the results are presented in Table 3. The corrplot package (Wei & Simko, 2017) was used to create two correlograms highlighting the most redundant variables in each set of extracted features. Some variables are highly correlated (>±0.75). Combining this criterion with the utility for furthering our understanding of patterns in public transport, these variables were excluded: active_months, max_group_size, visited_stops, visited_zones, main_two_stops and transactions_tarragona_reus.

Table 3. Correlograms of the calculated variables, reordered using hierarchical clustering (all coefficients are statistically significant at p<0.01).



Activity variables

	f	e	d	c	b	a
a-active_days	0.0	-0.2	0.6	0.5	0.7	1.0
b-transactions	0.4	0.2	0.3	0.3	1.0	0.7
c-active_months	0.0	-0.2	0.7	1.0	0.3	0.5
d-enabled_days	-0.1	-0.2	1.0	0.7	0.3	0.6
e-avg_group_size	0.8	1.0	-0.2	-0.2	0.2	-0.2
f-max_group_size	1.0	0.8	-0.1	0.0	0.4	0.0

Spatial variables

	h	g	f	e	d	c	b	a
a-main_three_stops	-0.5	-0.9	-0.7	0.0	0.0	0.0	0.9	1.0
b-main_two_stops	-0.5	-0.9	-0.8	0.0	0.0	0.0	1.0	0.9
c-transactions_cgc	0.1	0.0	-0.2	-0.2	-0.8	1.0	0.0	0.0
d-transactions_tarragona_reus	0.0	0.0	0.1	0.1	1.0	-0.8	0.0	0.0
e-visited_zones	0.1	0.0	0.0	1.0	0.1	-0.2	0.0	0.0
f-visited_municipalities	0.5	0.8	1.0	0.0	0.1	-0.2	-0.8	-0.7
g-visited_stops	0.6	1.0	0.8	0.0	0.0	0.0	-0.8	-0.7
h-used_routes	1.0	0.6	0.5	0.1	0.0	0.1	-0.5	-0.5

Source: Own research with data provided by ATM Camp de Tarragona.

3.2. Clustering technique selection

As a result of the data preparation and the selection stages of the indicators, the dataset to be analysed included processed records of the 34,641 selected T-10 cards, with four activity indicators and four spatial indicators for each card (Table 2). For each set of indicators, a clustering model has been implemented to identify traveller profiles. In general terms, clustering techniques are algorithms for grouping a set of physical or abstract objects into classes of similar objects (smart travel cards in this case).

The range of clustering techniques that could be applied is wide. According to He et al., (2020), three main groups of clustering techniques can be distinguished: (1) partitioning clustering algorithms, that require the user to specify in advance the number of clusters to be created (e.g. k-means, k-medoids, CLARA); (2) hierarchical clustering, that does not require the user to specify in advance the number of clusters and has the great advantage of creating a dendrogram for visualising the data structure (e.g. agglomerative clustering, divisive clustering); and (3) a heterogeneous group of methods that combine or modify the above mentioned techniques (e.g. hierarchical K-means, fuzzy clustering, model-based clustering or density-based clustering).

Fonseca (2013) presented some key arguments in favour of model-based clustering in social sciences and, more recently, some authors achieved good results using model-based clustering for characterising traveller activity profiles (Liu et al., 2020; El Mahrsi et al., 2017; Briand et al., 2015). Model-based clustering considers that the observations come from a distribution that is a mixture of more than one component or cluster. In principle, each cluster can be described by any density function, but it is usually assumed that they follow a normal or Gaussian distribution, which is characterised by the parameters: mean vector, covariance matrix, associated probability (each point has a probability of belonging to each cluster). To estimate the parameters that define the distribution function of each cluster, the expectation-maximization (EM) algorithm is used. This resolves different models in which the volume, shape and orientation of the distributions can be considered the same or different between clusters (e.g., one possible model is constant volume, variable shape, variable orientation). The key advantage of the model-based approach, compared to the standard clustering methods (k-means, hierarchical clustering, etc.), is the suggestion of the number of clusters and an appropriate model. One common model-based approach is the use of finite mixture models, which provides a flexible modelling framework for the analysis of the probability distribution. Finite mixture models are a linearly weighted sum of component probability distribution.

Considering all this information and that all the variables of the present study are continuous, a latent profile analysis (LPA) technique, also known as the “Gaussian (finite) mixture model” or “binomial (finite) mixture model”, has been used. As a person-centred approach —in contrast to variable-centred— LPA has been recurrently used in social sciences (Stanley et al., 2017; Jones et al., 2014, Fonseca 2013) and health studies (Machimbarrena et al., 2019; Sullivan et al., 2019; Imaginario et al., 2018), and, there are a few examples where LPA was used in mobility studies (Ton et al., 2019, McBride et al., 2018). However, the authors were not able to find public transportation smart card data studies that made explicit mention of LPA.

Among other R libraries, the mclust package is a very powerful and widely used package for Gaussian mixture modelling. It allows us to calculate many model configurations, and it can be used for different purposes (Scrucca et al., 2016). Although LPA can be performed directly using the mclust R library, the tidyLPA package (Rosenberg et al., 2018) has been used, since it provides a selected set of models that are common to LPA and that can be calculated using a “tidy” approach (Wickham et al., 2014). In order to select the best models, a series of LPA were performed. The best model and profile number was chosen by exploring fit statistics on the four different models available through tidyLPA, ranging from two to ten profiles.

Finally, once the clusters for the temporal activity variables and those for the spatial activity variables were obtained, the relationship between them was assessed by means of an alluvial diagram. This diagram is structured by two columns (strata) showing the sizes of activity profiles on the left side and spatial profiles on the right side, with lines linking both columns (alluvium) representing individual smart cards and showing how they were classified by the LPAs. These lines are coloured to show the number of times the card was recharged —being said that ten is the minimum buy and 30 the maximum— and ordered to plot similar lines together.

4. Results

In this section, the different clusters obtained and their structure are presented. First, the results of the clusters for the temporal activity variables and those for the spatial activity variables are described. Then the relationships between activity and spatial clusters are detailed.

4.1. Clustering profiles from activity variables

LPA based on activity variables was performed following the above mentioned parameters (4 different models and 2:10 profiles). Table 4 presents the values for the different models and the number of cards for each profile. All solutions returned a significant bootstrapped likelihood ratio test (BLRT) value, indicating that models with a greater number of profiles were the preferred solutions. In fact, an analytic hierarchy process, based on the fit indices AIC, AWE, BIC, CLC and KIC, suggests the best solution is the model with seven profiles (Akogul & Erisoglu, 2017).

Although the seven-profile solution had the lowest fit indices, a five-profile solution was chosen since it is the model that provides the most meaningful distribution and internal structure of groups, avoiding groups with less than 1% of the total cards. Furthermore, the mean posterior probabilities across the five profiles are high, ranging from 0.88 to 0.99. It indicates that individuals are classified into their respective profiles more accurately than the six-profile and seven-profile models. In addition, there is a significant difference in terms of entropy between the five-profile and the six-profile solutions.

Table 4. Model/clustering comparison – activity variables

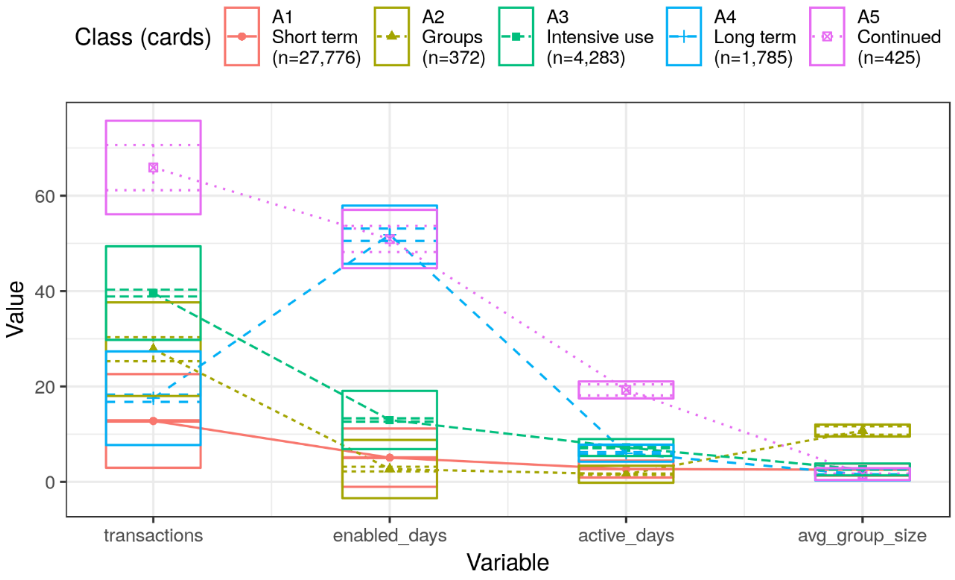
n profiles	AIC	BIC	Prob.		Entropy	*n per profile*						
			min	max		1	2	3	4	5	6	7
2	819,426.36	819,536.25	0.98	1	0.99	32,252	2,389					
3	809,434.36	809,586.51	0.93	1	0.99	379	2,384	31,878				
4	793,562.75	793,757.17	0.85	0.99	0.96	364	2,918	29,387	1,972			
5	**775,788.47**	**776,025.15**	**0.88**	**0.99**	**0.95**	**1,785**	**27,752**	**4,307**	**372**	**425**		
6	771,937.57	772,216.51	0.61	0.98	0.90	4,089	2740	221	423	1,773	25,395	
7	761,351.97	761,673.17	0.62	0.98	0.91	368	3,857	901	2,806	24,806	1,682	221

Note: AIC Aikake Information Criterion; BIC Bayesian Information Criterion; BLRT Bootstrapped Likelihood Ratio Test with $p < 0.01$

Source: Own research with data provided by ATM Camp de Tarragona.

The descriptive statistics of the activity variables for each of the five profiles obtained by means of LPA are shown in Figure 3. Moreover, further statistical details of the profiles obtained are presented in Table 5. As can be observed, each profile has received a taxonomy according to its characteristics and the previous knowledge of the study area.

Figure 3. Descriptive statistics of the profiles obtained by means of LPA with activity variables.



Source: Own research with data provided by ATM Camp de Tarragona.

Overall, the mean card lifespan (enabled_days) is relatively short (9.03), while the average number of transactions is 17.3, and it is evident that T-10 is attractive for groups, as the average group size is 2.6 people. Differences between groups, which have allowed the suggested taxonomy, are nevertheless substantial. Profile A1 (short term; n=27,776) is clearly distinguished in terms of the number of cards that compose it. In fact, it concentrates 80.6% of all the T-10 cards. The taxonomy assigned to this profile is short term, since it statistically resembles a pattern that could clearly fit the behaviour of tourists spending a short stay at the destination (less than one week). In general, the values of the four activity variables are below the mean. But specifically, the users (cards) within this profile show a low-activity pattern concentrated in a short time period, and they tend to travel individually or in small groups. This profile is associated with the average tourist of the Costa Daurada that stays overnight in tourist accommodation for an average length of around 5 nights. This group could be dominated by the presence of tourists traveling with family and couples, which are the main types of tourism in this destination (Gutiérrez & Miravet, 2016b). Profile A2 (groups; n=372) stands out for its very high value in the size of the group (avg_group_size) and its moderate or low values in the rest of the variables. This profile is composed of potential excursionists travelling in a group using single or various cards during their stay at the destination. Thus, their geographical behaviour might be more dispersed, in terms of number of areas visited. Profile A3 (intensive use; n=4,283) mainly presents high values in the number of transactions, but distributed in a moderate number of days. This profile seems to include tourists doing a longer stay than the first group and travelling more intensively. The intensive use of the card can either be concentrated along the coast due to a combination of sun and sand tourism with other activities linked to leisure and sightseeing, or due to a more sprawled behaviour that integrates a diversity of activities. Profile A4 (long term; n=1,785) comprises the highest scores in the card lifespan (enabled_days) and relatively high values on active days, but a lower number of transactions. It might be representative of individuals that spent long or multiple stays at the destination, such as people owning a second residence in the area, or local residents travelling sporadically for leisure purposes. Finally, cards in profile A5 (continued; n=425) are mainly characterised by the highest values of transactions and active days, high values of card lifespan, and small group size. This profile could concentrate seasonal workers that, for any reason, decided to use the T-10 fare instead of others more convenient for them (T-MES, T-10/30, T-50/30).

Table 5. Models results – Profiles according to activity variables

	Profiles					
	A1 **Short term** **(n=27,776)**	**A2** **Groups** **(n=372)**	**A3** **Intensive use** **(n=4,283)**	**A4** **Long term** **(n=1,785)**	**A5** **Continued** **(n=425)**	**Total** **(n=34,641)**
transactions						
Mean (SD)	12.8 (7.40)	27.6 (18.4)	40.5 (14.9)	17.6 (11.5)	66.2 (31.3)	17.3 (14.4)
Median [Min, Max]	10.0 [1.00, 45.0]	20.0 [8.00, 146]	39.0 [9.00, 150]	15.0 [2.00, 82.0]	61.0 [19.0, 288]	11.0 [1.00, 288]
enabled_days						
Mean (SD)	5.09 (4.50)	2.64 (4.42)	13.2 (6.75)	51.8 (14.8)	50.8 (17.3)	9.03 (13.0)
Median [Min, Max]	4.00 [1.00, 33.0]	1.00 [1.00, 43.0]	11.0 [3.00, 42.0]	49.0 [29.0, 93.0]	49.0 [14.0, 92.0]	5.00 [1.00, 93.0]
active_days						
Mean (SD)	2.70 (1.36)	1.57 (0.942)	7.35 (2.14)	6.03 (3.05)	19.4 (6.81)	3.64 (2.98)
Median [Min, Max]	3.00 [1.00, 9.00]	1.00 [1.00, 5.00]	7.00 [3.00, 18.0]	5.00 [2.00, 15.0]	18.0 [11.0, 51.0]	3.00 [1.00, 51.0]
avg_group_size						
Mean (SD)	2.54 (1.19)	10.9 (4.47)	2.62 (1.19)	1.55 (0.785)	1.64 (0.762)	2.58 (1.54)
Median [Min, Max]	2.00 [1.00, 7.00]	10.0 [7.00, 30.0]	2.00 [1.00, 10.0]	1.00 [1.00, 8.00]	1.00 [1.00, 4.00]	2.00 [1.00, 30.0]

Source: Own research with data provided by ATM Camp de Tarragona.

4.2. Clustering profiles from spatial variables

In this subsection, the approach to discovering groups of cards that exhibit similar patterns from a purely spatial standpoint is presented. Table 6 presents the values for the different models obtained by means of LPA, and the number of cards for each profile. According to the analysis of several fit indices (Akogul & Erisoglu, 2017), the solution of six profiles was the most appropriate. In this case, models with more than six profiles could not be calculated due to model convergence issues. As in the clustering from activity variables, all solutions returned a significant bootstrapped likelihood ratio test (BLTR) value. Although the six-profile solution has the lowest BIC, the four-profile solution has been chosen. The selected model provides more meaningful profiles, presents higher probabilities ranging from 0.92 to 0.96, and shows the highest value in terms of entropy. Therefore, the four-profile provides a better classification than the five- and six-profile models.

Table 6. Model/clustering comparison – spatial variables.

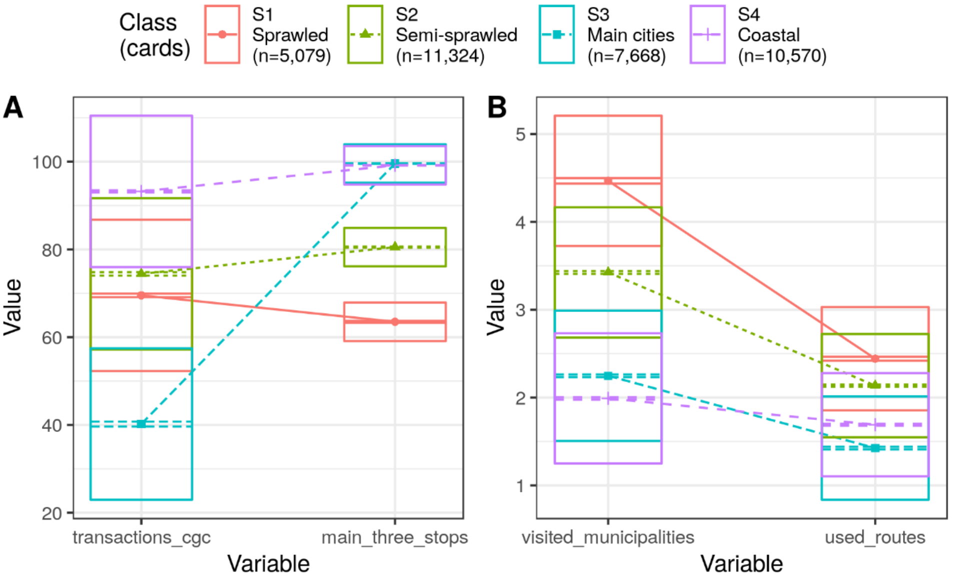
n profiles	AIC	BIC	Prob.		Entropy	*n* per profile					
			min	max		1	2	3	4	5	6
2	745,241.88	745,351.76	0.96	0.97	0.88	14,267	20,374				
3	735,277.63	735,429.78	0.91	0.96	0.86	14,101	7,782	12,758			
4	**717,908.56**	**718,102.98**	**0.92**	**0.96**	**0.89**	**5,082**	**11,321**	**10,570**	**7,668**		
5	714,856.74	715,093.42	0.70	0.95	0.86	4,958	5,988	10,849	3,743	9,103	
6	709,979.29	710,258.23	0.64	0.95	0.84	4,962	11,062	7,722	5,305	1,569	4,021

Note: AIC Aikake Information Criterion; BIC Bayesian Information Criterion; BLRT Bootstrapped Likelihood Ratio Test with $p < 0.01$.

Source: Own research with data provided by ATM Camp de Tarragona.

In Figure 4, the descriptive statistics of the spatial variables for each of the five profiles obtained are presented. Table 7 provides more statistical details of each profile. In the same way as in the first clustering from activity variables, each profile has received a taxonomy according to its characteristics (see Table 7).

Figure 4. LPA with spatial variables. Graph divided into sub-figures due to the different scaling of the variables.



Source: Own research with data provided by ATM Camp de Tarragona.

The general spatial patterns of the whole cards are clear: there is an intense concentration of transactions in the coastal municipalities (72.1%) and in a limited number of stops (88% of transactions in the top-three stops of each card). Nevertheless, differences between the four spatial profiles can also be identified.

In this regard, profile S1 (sprawled; n=5,082) is characterised by cards with a higher number of visited_municipalities and used_routes, and with the lowest percentage of transactions concentrated in three stops (main_three_stops). This means that this profile of activity is the most dispersed in the study area and, therefore, it has received the taxonomy of “sprawled”. Cards in profile S4 (coastal; n=10,570) generally show a concentrated spatial pattern around the most touristic municipalities located on the coast (Cambrils, Salou and Vilaseca concentrate 94.4% of the transactions). Since this profile is concentrating its activities along beach towns, there is a high probability of being composed of tourists interested in sun and sand tourism. Inversely, profile S3 (Main cities; n = 7,668) is characterised by cards with higher spatial concentration, in terms of number of stops used, and a lower presence in the touristic municipalities. This implies that most of the transactions were done in the two most-important cities, Tarragona and Reus. This group could be, therefore, mainly associated with tourists interested in cultural tourism in urban environments. The last profile, S2 (semi-sprawled; n=11,324), generally presents on-average values in all variables, but it shows relatively high values of activity concentrated in the tourist municipalities. In this regard, this group could involve tourists mainly inclined for beach tourism but also with certain interest in urban tourism.

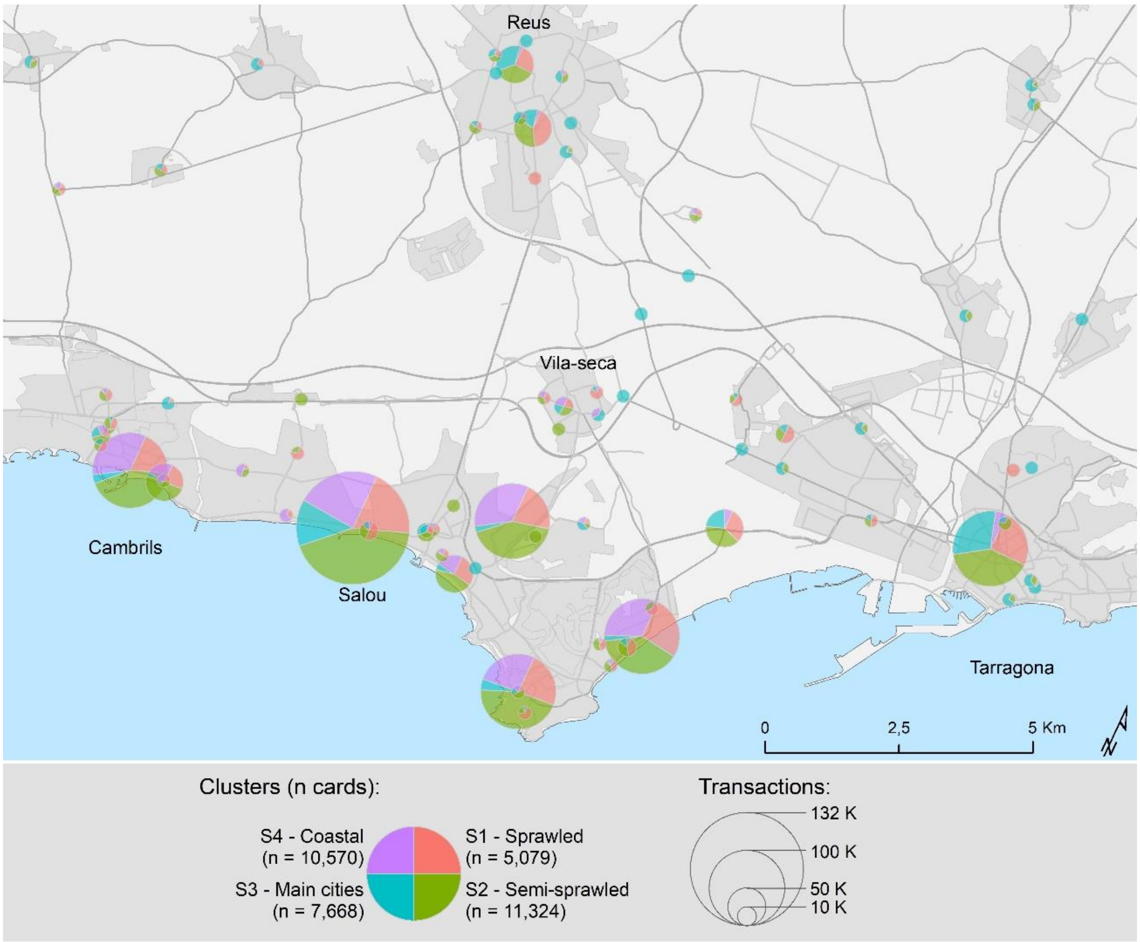
Table 7. Models' results – Profiles according to spatial variables.

	Profiles				
	S1 Sprawled (n=5,079)	**S2 Semi-sprawled (n=11,324)**	**S3 Main cities (n=7,668)**	**S4 Coastal (n=10,570)**	**Total (n=34,641)**
transactions_cgc					
Mean (SD)	69.5 (14.0)	74.4 (18.3)	39.5 (21.1)	94.4 (10.9)	72.1 (25.7)
Median [Min, Max]	70.0 [0.00, 100]	78.0 [0.00, 100]	50.0 [0.00, 69.0]	100 [58.0, 100]	75.0 [0.00, 100]
main_three_stops					
Mean (SD)	63.4 (6.36)	80.7 (5.47)	99.6 (1.70)	99.3 (2.40)	88.0 (13.9)
Median [Min, Max]	64.0 [30.0, 79.0]	80.0 [67.0, 95.0]	100 [82.0, 100]	100 [86.0, 100]	94.0 [30.0, 100]
visited_municipalities					
Mean (SD)	4.48 (0.900)	3.41 (0.771)	2.24 (0.628)	1.98 (0.681)	2.87 (1.16)
Median [Min, Max]	4.00 [2.00, 8.00]	3.00 [2.00, 8.00]	2.00 [1.00, 5.00]	2.00 [1.00, 4.00]	3.00 [1.00, 8.00]
used_routes					
Mean (SD)	2.44 (0.759)	2.13 (0.538)	1.41 (0.614)	1.70 (0.513)	1.89 (0.685)
Median [Min, Max]	2.00 [1.00, 9.00]	2.00 [1.00, 6.00]	1.00 [1.00, 5.00]	2.00 [1.00, 5.00]	2.00 [1.00, 9.00]

Source: Own research with data provided by ATM Camp de Tarragona.

These spatial profiles can be easily visualised using a pie chart map. Figure 5 shows the volume of transactions made per bus stop in the core of the study area, together with the proportion of smart card transactions made by each profile. This map evidences three mixed patterns. First, a very high spatial concentration of activity around the most touristic areas (Cambrils, Salou and Vila-seca) is detected. Second, a moderate volume of transactions in the main cities can be observed. Third, a very marked difference, in terms of the proportion of profiles, is appreciated between the stops near the sea and those further away (especially Tarragona and Reus). Regarding the last pattern, multiple observations emerge from the observation of the map. S1-profile (sprawled) has a significant presence in all bus stops, but it is slightly more important in stops with lower activity. Moreover, the presence of the S2-profile (semi-sprawled) in the two main cities is practically non-existent. On the contrary, this profile tends to be more predominant on coastal stops, but concentrated in fewer routes and visiting fewer different municipalities than the S1-profile. The S3-profile (main cities) is clearly more important in the two main cities of the region. Finally, the S4-profile (coastal) is clearly defined by its highest percentage in all the main stops near the sea. This characteristic is not possible to be appreciated directly from the LPA results, but it needs the spatial component to become evident.

Figure 5. Transaction by bus stops and spatial activity profiles.



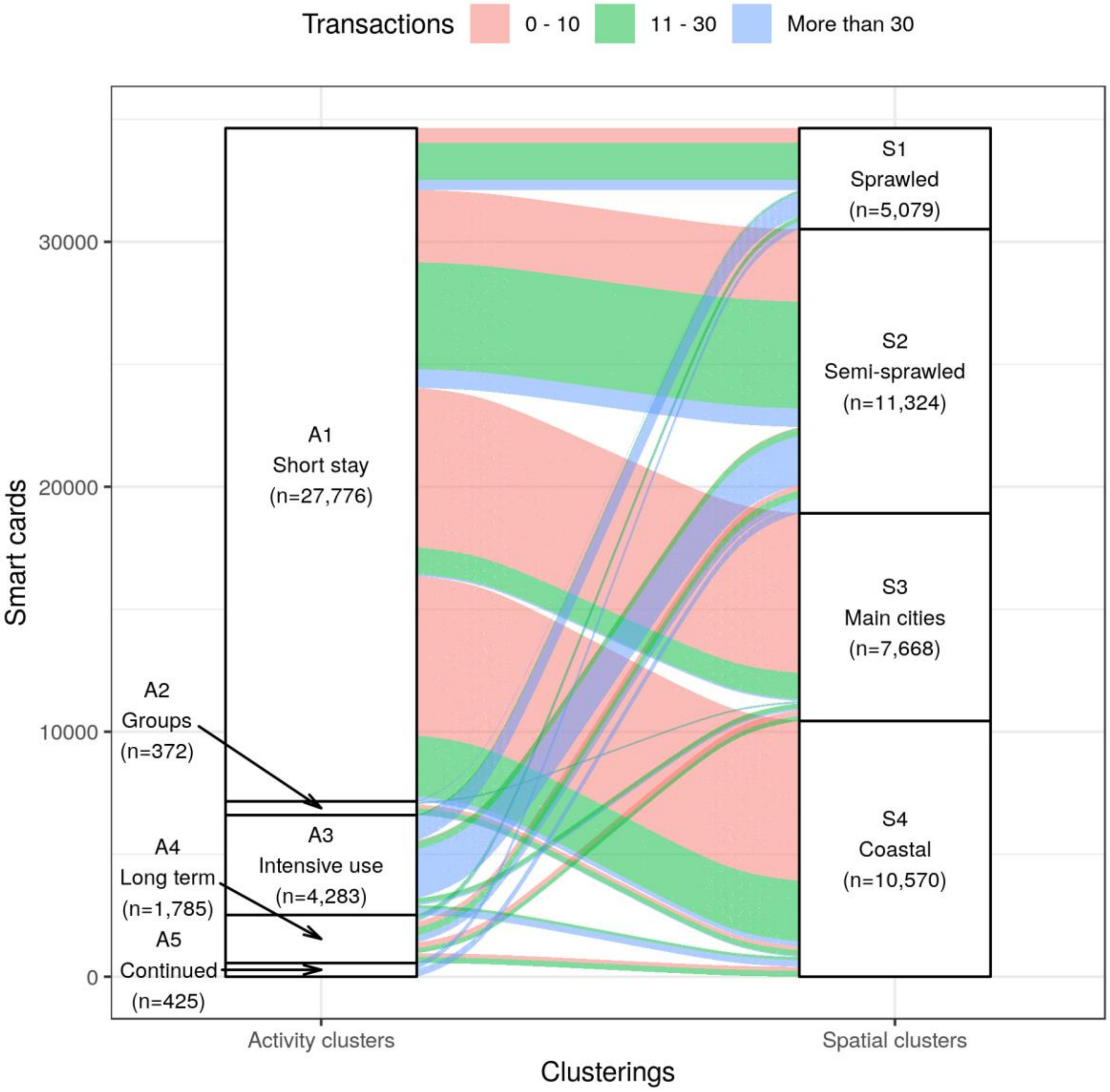
Source: Own research with data provided by ATM Camp de Tarragona.

4.3. Relations between activity and spatial clusters

Beyond the above-described activity and spatial patterns, the approach also makes it possible to explore the relationships between clusters of different nature. This approach allows the development of more parsimonious and easily interpretable models. Therefore, in Figure 6, an alluvial diagram is used to show the relationships between activity and spatial clusters. The two columns (strata) show the sizes of profiles, with the one from the left belonging to the activity clusters, and the one from the right to the spatial clusters. The spatial profiles are clearly more equally distributed in terms of the volume of cards assigned to each cluster.

The lines linking both columns (alluvium) represent individual smart cards and show how they have been classified by the LPA. These lines are coloured to show the number of times the card was recharged —being said that ten is the minimum buy and 30 the maximum— and ordered to plot similar lines together.

Figure 6. Alluvial diagram showing relationships between the activity and spatial clusters.



Source: Own research with data provided by ATM Camp de Tarragona.

On the one hand, from Figure 6, it is possible to visualise that cards in profiles A2 (groups), A3 (intensive use) and A5 (continued) have important activity levels, since most of them registered more than 30 transactions. However, cards in the A2-profile are more irregular in time; meanwhile, cards belonging to the A3-profile tend to perform many transactions in short periods of time, and cards from the A5-profile have a much longer lifespan. On the other hand, cards in profiles A1 and A4 present lower activity levels for short and longer periods of time, respectively. It is very interesting to note that most of the cards in profile A1 did not even recharge once during the summer period.

The relationships between activity and spatial profiles emerge in the visualisation in Figure 6. It can be seen that cards in profile S1 (sprawled) —which visited more municipalities, stops and used more routes than cards in other profiles— are mainly high-activity cards. Cards in S4-profile (coastal) are characterised by a slightly lower activity level compared to the first. The two remaining profiles, S3 (main cities) and S2 (semi-sprawled), mostly include cards that made fewer transactions.

The alluvial diagram also allows us to understand the direct relations between activity and spatial clusters. Obviously, the A1 profile (short term) has the highest proportion of cards in every spatial cluster. Nevertheless, important transfers between clusters can be identified. The S3 (main cities) together with the S4 (coastal) are the clusters with the highest presence of A1 (short term) and low-transaction cards. These are, therefore, representative of short-stay tourists who either have a greater interest in urban areas or a major interest in coastal destinations. Those cards from the A1 profile included in the S1 (sprawled) and S2 (semi-sprawled) profiles tend to have a medium amount of transactions (11– 30), in contrast to the usual A1 cards. They might be families and relatives moving together, with a more generic interest in widely exploring the territory, combining urban visits with visits to the Mediterranean coast. Similar to the previous group, the A3 (intensive) users have the analogous motivations in terms of spaces visited, but they travel more intensively. The A4 (long term) profile shows clear patterns that could be associated with those of long-stay temporary residents, such as second-home owners. They tend to have a semi-sprawled behaviour, combining visits to coastal and urban areas. Finally, profile A5, which could correspond mainly to workers, has high transaction values and is distributed almost equally among the sprawled and semi-sprawled spatial profiles. This indicates that they are users with both their residence and workplace locations distributed in the territory, rather than concentrated in the coastal municipalities or in the main cities.

5. Discussion

The present work has analysed the spatial and temporal behaviour of visitors of the Costa Daurada, a very popular Spanish coastal tourist destination, through data collected via smart travel cards in public transport networks.

The mining of this type of data source is not a novel issue in transport research. In fact, Bagchi & White (2005) have already highlighted their potential use in the field of travel behaviour analysis. However, the analysis of smart travel card data in tourist cities and regions has been incorporated on very few occasions and just to separate residents' public transport demand from visitors' (Ortega-Tong, 2013; Lu et al., 2019; Xue et al., 2014). But there is still a lack of studies focused on profiling visitors' use of public transport. Therefore, the present study addressed this research gap and demonstrated the potential of mining smart travel cards to tackle the specific challenges of public transport management in tourist contexts. The characteristics of this source of data have demonstrated themselves to be valuable tools with which to approach the dynamic evolution of public transport demand in contexts of seasonality, which can completely reshape the mobility patterns of a whole territory (Domènech et al., 2020).

Previous studies with smart travel card data stressed the usefulness of the identification of profiles at different levels of management of public transport authorities (strategic, tactical and operational) (Pelletier et al., 2011), as well as to forecast the evolution of demand (Pelletier et al., 2011), to identify bottlenecks (Makimura et al., 2017), or to develop tailor-made performance indicators (Trépanier & Morency, 2017). The present study extends these contributions to contexts where seasonality poses additional pressure on the management and adjustment of the provision of optimised public transport services. Moreover, this understanding of visitors' mobility is also useful for tourism stakeholders, since deeper knowledge of how the different places of the destination are visited and integrated into the network is provided.

The empirical results obtained in this study point out that the data registered by the T-10 multipersonal smart cards have made it possible to identify different profiles of visitors and locals travelling by public transport in the Camp de Tarragona region. The diverse use of the public transport services amongst the groups detected reveal the existence of users with different needs, thereby raising policy implications. The results clearly reveal that visitors use the T-10 multipersonal fare due to the flexibility advantages it offers. Similarly, the study shows how a reduced number of local users (i.e. temporary workers in the tourist sector) and long-term tourists also make use of the fare, although in some cases other transport fares might be better adapted to their needs. In this sense, the results emanate the importance of the existence and promotion of multipersonal transport fares to encourage public transport use among visitors and temporary users. This is increasingly important in those destinations, such as Costa Daurada, where familiar tourism is the cornerstone of their demand. Therefore, efforts should be made to intensify information campaigns, which could even be personalized among users who acquire a card, in order to promote public transport among tourists and visitors as well as to familiarize and generate loyalty amongst local users and long-term tourists (e.g. second-home owners). This consideration contrasts with other decisions made by authorities managing public transport systems in many popular destinations based on the existence of multipersonal fares. On the one hand, cities such as London, Rome or Vienna, and coastal destinations such as the Canary Islands (Spain) or Antalya (Turkey) do not offer any multipersonal transport fares. On the other hand, other coastal tourist destinations such as Malta have given particular characteristics to the multipersonal fares that make them less appealing to visitors. Therefore, the results of this study suggest reinforcing the attractive qualities of multipersonal fares for temporary users, given the fact that extending the use of public transport among visitors is central to ensuring sustainable mobility in tourist destinations and very useful in increasing the competitiveness of a tourist destination (Mandeno, 2012; Hall et al., 2017)..

Apart from contributions at the management level of the public transport system, this article has allowed the testing and validation of a method that could be useful for researchers and public authorities aiming to take advantage of smart travel card data to understand tourists' mobilities and visitors' public transport use. Previous to this study, other researchers have applied clustering techniques to a diverse range of purposes, such as the loyalty of public transport users (Trépanier et al., 2012), the analysis of each passenger's recurrent travel patterns and behaviours (Liu et al., 2020; Liang et al., 2018), the separation of frequent and infrequent travellers (Manley et al., 2018), and the understanding of the spatial and temporal patterns of traveller profiles (Faroqi et al., 2017; Briand et al., 2015; Briand et al., 2017). However, none of the published studies that have analysed public transport smart card data have explicitly mentioned the use of the model-based technique used in the present study (LPA).

Nevertheless, the lack of available sociodemographic information of users, travel purposes and origin-destination matrices (Kurauchi & Schmöcker, 2017) are some of the limitations that all kinds of studies using smart card data have to face. Furthermore, there is another implicit limitation: all the journeys using single-use tickets have been excluded from the analysis, as they do not allow the tracking of the public transport users. This is a relevant issue because an important volume of tourists tends to use single tickets. As a result, an important volume of tourist mobilities by public transport is excluded from the analysis. In addition to these limitations, the replicability of this study is subject to some conditions. First, each public transport network has its own diversity of transport fare categories. Therefore, the subsequent selection and use of transport fares might be very different depending on the territorial context and the network structure. Second, the data availability among the diverse automated fare collection systems is far from homogenous, as each one has its own design structure. Finally, public transport networks in the largest cities generate such a huge amount of data that in some cases, the databases may be periodically refreshed and ancient data deleted.

6. Conclusions

This study has demonstrated that the multipersonal fare T-10 is mainly used by visitors in the Costa Daurada tourist region. Due to its characteristics (value-for-money and multipersonality), it is undoubtedly interpreted as a tool that guarantees tourists mobility in the region. In brief, three main findings have been detected after the mining of smart travel card data. First, what really makes T-10 attractive among visitors is the fact that it can be used by groups. Second, common temporal and spatial patterns among those passengers have been found: reduced activity time and concentration of transactions in the coastal stops. Third, it has been possible to identify the spatiotemporal logics of visitors by means of the interaction of the four profiles obtained according to the temporal activity and the other five profiles based on the spatial behaviour. According to these results, the first objective of this work, which was to portray the profiles of public transport users in a seasonal tourist destination, has been accomplished. In fact, the implications of the study go far beyond the research interest, as the study provides valuable information for the proper design and adjustment of transport policies. Besides, the results obtained validate a methodology to analyse the public transport patterns of visitors, which was our second goal.

In any case, this study and its derived results open up a diversity of new and interesting lines of research in the fields of transport geography and travel behaviour analysis. First, the impact of seasonality on the efficiency of public transport services could be analysed. In this sense, to assess to what extent the seasonal increase of the number of users has an impact on the quality of the services (i.e. time required for boarding, time required for the whole journey, delays, number of buses overcrowded, among other potential issues) is of particular interest. A second topic would be the measure of the impact of these circumstances on the behaviour of residents. Additional pressure on public transport services could lead to the development of subjectivities and perceptions that are detrimental to the well-being of residents and, more worryingly, to possible modal changes to other less sustainable modes of transport. A third line of research involves the combination of smart card data with other georeferenced information, such as the location of hotels or points/areas of tourist interest. It would help to detect what role these potential generators of mobilities play, and, at the same time, it would provide ideas on how to better adapt services to visitors' needs. It would also be interesting to implement this type of methodology to depict visitors' profiles in contexts characterised by inverted seasonality, such as winter mountain destinations or urban environments with more stable flows of visitors. Beyond the use of public transport by tourists, given the very particular characteristics of this transport fare which are attractive to infrequent users and people travelling in a group, a similar profiling of the T-10 by the resident population would be of high interest. One of the key potentialities of the smart card data is the availability to develop interannual comparative analyses (Briand et al., 2017). This would allow researchers to study the evolution of traveller profiles and their spatial behaviour, the overall redefinition of public transport fluxes in the region, and longitudinal analysis of changes of behaviour amongst individual and/or specific groups. This possibility would be of special interest for future studies analysing the multiple disruptive effects of the COVID-19 pandemic in the usage of public transport.

References

Akogul, S. & Erisoglu, M. (2017). An approach for determining the number of clusters in a model-based cluster analysis. *Entropy*, 19(9), 452. https://doi.org/10.3390/e19090452

Albalate, D., & Bel, G. (2010). Tourism and urban public transport: Holding demand pressure under supply constraints. *Tourism Management*, 31(3), 425-433. https://doi.org/10.1016/j.tourman.2009.04.011

Bagchi, M., & White, P. R. (2005). The potential of public transport smart card data. *Transport Policy*, 12(5), 464-474. https://doi.org/10.1016/j.tranpol.2005.06.008

Briand, A. S., Côme, E., El Mahrsi, M. K., & Oukhellou, L. (2015). A mixture model clustering approach for temporal passenger pattern characterization in public transport. *Proceedings of the 2015 IEEE International Conference on Data Science and Advanced Analytics, DSAA 2015*, 1(1), 37–50. https://doi.org/10.1109/DSAA.2015.7344847

Briand, A. S., Côme, E., Trépanier, M., & Oukhellou, L. (2017). Analyzing year-to-year changes in public transport passenger behaviour using smart card data. *Transportation Research Part C: Emerging Technologies*, 79, 274–289. https://doi.org/10.1016/j.trc.2017.03.021

Domènech, A. & Gutiérrez, A. (2017). A GIS-based evaluation of the effectiveness and spatial coverage of public transport networks in tourist destinations. *ISPRS International Journal of Geo-Inormation*, 6 (3). http://dx.doi.org/10.3390/ijgi6030083

Domènech, A., Miravet, D., & Gutiérrez, A. (2020). Mining bus travel card data for analysing mobilities in tourist regions. *Journal of Maps*, 16(1), 40-49. https://doi.org/10.1080/17445647.2019.1709578

Duval, D.T. (2007). Tourism and Transport. Modes, Networks and Flows. Bristol: Channel View.

El Mahrsi, M. K., Come, E., Oukhellou, L., & Verleysen, M. (2017). Clustering Smart Card Data for Urban Mobility Analysis. *IEEE Transactions on Intelligent Transportation Systems*, 18(3), 712–728. https://doi.org/10.1109/TITS.2016.2600515

Faroqi, H., Mesbah, M., & Kim, J. (2017). Spatial-temporal similarity correlation between public transit passengers using smart card data. *Journal of Advanced Transportation*, 2017. https://doi.org/10.1155/2017/1318945

Fonseca, J. R. S. (2013). Clustering in the field of social sciences: that is your choice. *International Journal of Social Research Methodology,* 16(5), 403–428. https://doi.org/10.1080/13645579.2012.716973

Ghaemi, M.S., Agard, B., Trépanier, M., & Partovi Nia, V. (2016). A visual segmentation method for temporal smart card data. *Transportmetrica A: Transport Science*, 13(5), 381-404. https://doi.org/10.1080/23249935.2016.1273273

Gronau, W., & Kagermeier, A. (2007). Key factors for successful leisure and tourism public transport provision. *Journal of Transport Geography*, 15(2), 127-135. https://doi.org/10.1016/j.jtrangeo.2006.12.008

Gutiérrez, A., & Miravet, D. (2016a). Estacionalidad turística y dinámicas metropolitanas: un análisis a partir de la movilidad en transporte público en el Camp de Tarragona. *Revista de Geografía Norte Grande*, (65), 65-89. http://dx.doi.org/10.4067/S0718-34022016000300004

Gutiérrez, A., & Miravet, D. (2016b). The determinants of tourist use of public transport at the destination. *Sustainability* 8 (9), 908. https://doi.org/10.3390/su8090908

Hall, C.M., Le-Klähn, D.T., & Ram, Y. (2017). Tourism, public transport and sustainable Mobility. Birstol: Channel View.

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. *Transportmetrica A: Transport Science*, 16(1), 56–75. https://doi.org/10.1080/23249935.2018.1479722

Imaginário, C., Rocha, M., Machado, P., Antunes, C., & Martins, T. (2018). Self-care profiles of the elderly institutionalized in Elderly Care Centres. *Archives of Gerontology and Geriatrics*, 78, 89–95. https://doi.org/10.1016/j.archger.2018.05.012

Jones, M., & Huh, J. (2014). Toward a multidimensional understanding of residential neighborhood: A latent profile analysis of Los Angeles neighborhoods and longitudinal adult excess weight. Health and Place, 27, 134–141. https://doi.org/10.1016/j.healthplace.2014.01.011

Kurauchi, F., & Schmöcker, J.D. (Eds). (2017). Public transport planning with smart card Data. London: CRC Press.

Kwan, M-P. & Schwanen, T. (2017). Geographies of mobility: recent advances in theory and method. Routledge.

Le-Klähn, D.-T., & Hall, C. M. (2015). Tourist use of public transport at destinations – a review. *Current Issues in Tourism*, 18(8), 785–803. https://doi.org/10.1080/13683500.2014.948812

Li, T., Sun, D., Jing, P., & Yang, K. (2018). Smart card data mining of public transport destination: A literature review. *Information* 9(1), 28–30. https://doi.org/10.3390/info9010018

Liang, Q., Weng, J., Zhou, W., Santamaria, S. B., Ma, J., & Rong, J. (2018). Individual Travel Behavior Modeling of Public Transport Passenger Based on Graph Construction. *Journal of Advanced Transportation*, 2018. https://doi.org/10.1155/2018/3859830

Liu, Y., & Cheng, T. (2020). Understanding public transit patterns with open geodemographics to facilitate public transport planning. *Transportmetrica A: Transport Science*, 16(1), 76–103. https://doi.org/10.1080/23249935.2018.1493549

Lu, Y., Mateo-Babiano, I., & Sorupia, E. (2019). Who uses smart card? Understanding public transport payment preference in developing contexts, a case study of Manila's LRT-1. *IATSS Research*, 43(1), 60–68. https://doi.org/10.1016/j.iatssr.2018.09.001

Lu, Y., Wu, H., Xin, L., Chen, P., & Zhang, J. (2019). TourSense: A Framework for Tourist Identification and Analytics Using Transport Data. *IEEE Transactions on Knowledge and Data Engineering,* 31(2), 2407–2422. https://doi.org/10.1109/TKDE.2019.2894131

Makimura, K., Nakamura, T., Ishigami, T & Imai, R. (2017). Assessment of traffic bottlenecks at bus stops. Kurauchi, F., & Schmöcker, J.D. (Eds). Public Transport Planning with Smart Card Data. (pp. 225–244) London: CRC Press.

Mandeno, T. G. (2012). *Is Tourism a Driver for Public Transport Investment?* (Doctoral dissertation, University of Otago).

Manley, E., Zhong, C., & Batty, M. (2018). Spatiotemporal variation in travel regularity through transit user profiling. *Transportation*, 45(3), 703–732. https://doi.org/10.1007/s11116-016-9747-x

McBride, E. C., Davis, A. W., & Goulias, K. G. (2018). A Spatial Latent Profile Analysis to Classify Land Uses for Population Synthesis Methods in Travel Demand Forecasting. *Transportation Research Record*, 2672(49), 158–170. https://doi.org/10.1177/0361198118799168

Ortega-Tong, M.A. (2013). Classification of London's public transport users using smart card data. Master Thesis. Massachusetts Institute of Technology. http://hdl.handle.net/1721.1/82844

Pelletier, M. P., Trépanier, M., & Morency, C. (2011). Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies*, 19(4), 557–568. https://doi.org/10.1016/j.trc.2010.12.003

Prideaux, B. (2000). The role of the transport system in destination development. *Tourism Management*, 21, 53-63.

Rosenberg, J. M., N. Beymer, P., J. Anderson, D., & A. Schmidt, J. (2018). tidyLPA: An R Package to Easily Carry Out Latent Profile Analysis (LPA) Using Open-Source or Commercial Software. *Journal of Open Source Software*, 3(30), 978. https://doi.org/10.21105/joss.00978

Rubin, J. (Ed.) (2015). TEA/AECOM 2014 Theme Index and Museum Index: the global attractions attendance report. Burbank (CA): Themed Entertainment Association (TEA).

Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. The R Journal, 8(1), 289–317. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/27818791

Stanley, L., Kellermanns, F. W., & Zellweger, T. M. (2017). Latent Profile Analysis: Understanding Family Firm Profiles. *Family Business Review*, 30(1), 84–102. https://doi.org/10.1177/0894486516677426

Sullivan, M. O., Gallagher, L., & Heron, E. A. (2019). Gaining Insights into Aggressive Behaviour in Autism Spectrum Disorder Using Latent Profile Analysis. *Journal of Autism and Developmental Disorders*, 49(10), 4209–4218. https://doi.org/10.1007/s10803-019-04129-3

Wei, T., & Simko, V. (2017). R package "corrplot": Visualization of a Correlation Matrix. Retrieved from https://github.com/taiyun/corrplot

Ton, D., Zomer, L. B., Schneider, F., Hoogendoorn-Lanser, S., Duives, D., Cats, O., & Hoogendoorn, S. (2019). Latent classes of daily mobility patterns: the relationship with attitudes towards modes. *Transportation*. https://doi.org/10.1007/s11116-019-09975-9

Trépanier, M., Habib, K. M. N., & Morency, C. (2012). Are transit users loyal? revelations from a hazard model based on smart card data. *Canadian Journal of Civil Engineering*, 39(6), 610–618. https://doi.org/10.1139/L2012-048

Trépanier, M., & Morency, C. (2017). Evaluation of Bus Service Key Performance Indicators Using Smartcard Data. F. Kurauchi & J.. Schmoker (Eds.), *Public Transport Planning with Smart Card Data*, 181-196

Wickham, H. (2014). Tidy Data. *Journal of Statistical Software*, 59(10). https://doi.org/10.18637/jss.v059.i10

Wu, H., Tan, J. A., Ng, W. S., Xue, M., & Chen, W. (2015, May). FTT: A system for finding and tracking tourists in public transport services. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data* (pp. 1093-1098). ACM. https://doi.org/10.1145/2723372.2735367

Xue, M., Wu, H., Chen, W., Ng, W. S., & Goh, G. H. (2014). Identifying tourists from public transport commuters. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1779–1788. https://doi.org/10.1145/2623330.2623352

Zaragozí, B., Gutiérrez, A., & Trilles, S. (2020). Towards an affordable GIS for analysing public transport mobility data: A preliminary file naming convention for avoiding duplication of efforts. In *Proceedings of the 6th International Conference on Geographical Information Systems Theory, Applications and Management* (pp. 302–309). SCITEPRESS - Science and Technology Publications. https://doi.org/10.5220/0009766303020309