

Published as: Bordagaray M., dell'Olio L., Fonzone A., Ibeas A, 2016, Capturing the conditions that introduce systematic variation in bike-sharing travel behavior using data mining techniques, *Transportation Research Part C*, DOI: 10.1016/j.trc.2016.07.009

Capturing the conditions that introduce systematic variation in bike-sharing travel behavior using data mining techniques

Abstract

The potential of smart-card transactions within bike-sharing systems (BSS) is still to be explored. This research proposes an original offline data mining procedure that takes advantage of the quality of these data to analyze the bike usage casuistry within a sharing scheme. A difference is made between usage and travel behavior: the usage is described by the actual trip-chaining gathered with every smart-card transaction and is directly influenced by the limitations of the BSS as a public renting service, whilst the travel behavior relates to the spatio-temporal distribution, the travel time and trip purpose. The proposed approach is based on the hypothesis that there are systematic usage types which can be described through a set of conditions that permit to classify the rentals and reduce the heterogeneity in travel patterns. Hence, the proposed algorithm is a powerful tool to characterize the actual demand for bike-sharing systems. Furthermore, the results show that its potential goes well beyond that since service deficiencies rapidly arise and their impacts can be measured in terms of demand. Consequently, this research contributes to the state of knowledge on cycling behavior within public systems and it is also a key instrument beneficial to both decision makers and operators assisting the demand analysis, the service redesign and its optimization.

Keywords: bike-sharing systems; data mining; smart-card data; demand analysis; cycling; trip-chaining

1. Introduction

Nowadays, promoting cycling is one of the most popular policies implemented by local governments to encourage greener mobility and make cities more liveable. In this context, bike-sharing systems (BSS) have been launched in numerous urban areas as an alternative mode of transport and evidences exist that confirm their potential to encourage cycling (García-Palomares, Gutiérrez, & Latorre, 2012; Pucher, Dill, & Handy, 2010).

Mobility demand is generally considered a derived demand. However, this may not always be the case for cycling. The analysis and comparison of the demand of various worldwide systems provided by O'Brien, Cheshire, and Batty (2014) shows that BSS are not only demanded as a mode for transport but also as a recreational tool. Riding a bike for transport and cycling as an activity itself lead to different mobility patterns. In addition, Lathia, Ahmed, and Capra (2012) find that also occasional and registered BSS users behave differently.

Recent studies on BSS rely on the great availability of data concerning smart-card transactions. Data made available by Intelligent Transport Systems (ITS) in general and by smart-card systems in particular have several advantages over data collected through traditional methods:

- Data collection is not affected by the observer;
- Large samples and often even entire populations are recorded and so the sampling bias is avoided;
- It is an economical source of data compared to traditional sampling methods;

However, a major limitation of ITS data is that it provides information on “how” and “when” but not on “why”, as explained by Lathia et al. (2012).

This research proposes a new methodology to understand the demand for bike-sharing by analysing the actual use of the service as recorded by the ITS. An algorithm is presented which mines the data offline, with the aim of identifying and characterising specific mobility patterns. Knowledge about such patterns can assist in designing and managing BSS so that they can achieve their intended purpose. Consequently, the above-mentioned shortcoming of ITS data is a challenge and an additional motivation to our research rather than a limitation. In fact, our data mining technique has been developed precisely to overcome the lack of insights about the transportation needs to be satisfied with bike-sharing services and so to derive the greatest benefits from raw ITS datasets. At the same time, our research makes use of all the advantages of ITS data since the quantity and the quality of the automatic records allow to identify less usual and scarcely represented mobility patterns that may not have been captured from a limited sample of manually collected data.

The contribution of the present research is derived from the novelty and potential of the methodology. The behavioral framework that is presented brings knowledge on the different usage types of a BSS and how its use is related to the travel patterns: most demanded terminals, travel time and time of day. The proposed algorithm identifies the usage conditions that introduce systematic variability in the travel behavior and that permit the classification of rentals. The analysis of each usage type confirms that they respond to different transportation needs. Furthermore, the results show that the users do not always behave as expected in a sharing service, providing insights on the aspects that provoke bike imbalance across terminals. Indirectly, the algorithm also identifies the conditions of the service supply that promote its use for transport and as a tool for recreation as well as its potential as a touristic attraction.

The implementation of our methodology can bring about many benefits. Firstly, detailed understanding on mobility patterns like that provided by our data mining technique can inform

on regulations of BSS to promote a responsible use. In fact, BSS bicycles are meant to be shared among users and so the time the same person can consecutively use the same bicycle is limited. However, sometimes it is possible to overcome the regulations without being charged for hampering sharing. The application of the data mining technique proposed in this research offers vital information to optimize the system in terms of the subscription policy and the location of terminals for the existing demand features. For instance, it can assist in deciding the location of new terminals depending on the type of demand to encourage, or in redesigning system attributes like the fare structure or the flat fee period. Finally, better understanding of mobility patterns (trip purpose, location of activities, time distribution) derived from our approach can be beneficial to urban transport and land use planning and management in general.

We evaluate our data mining technique and illustrate its potential uses by applying the methodology to the TusBic bike-sharing system in Santander (Spain).

The rest of the paper is organised as follows. The international literature is reviewed in section 2 to provide an overview of the state of the art in the field of the analysis of the demand for bikeshare and based on ITS data. Our approach is presented in section 3, where we introduce the algorithm for data mining and illustrate its potential and benefits. The case study and the results are presented in section 4. Finally, limitations, conclusions and further work are discussed in section 5.

2. State of the art in the field of BSS demand analysis

The great amount of rich data provided by ITS has given momentum to the research on public transport. In particular, academics and practitioners have developed analyses, techniques and tools that can provide detailed information on the demand for transit systems. In general, these approaches apply trip-chaining concepts to the “footprints” left by the smartcards (and more recently by the bank cards) used to access the services to infer trip origin and destination, to observe travel patterns and to model traveler behavior. In the field of public transport demand analyses using ITS data, noteworthy are the contributions of Cortés, Gibson, Gschwender, Munizaga, and Zúñiga (2011), Gordon (2012), Munizaga and Palma (2012), Wang, Attanucci, and Wilson (2011), and Wilson, Zhao, and Rahbee (2009).

Public and politicians progressively consider BSS as new mode of public transport. Useful reviews of the literature on BSSs are provided by García-Palomares et al. (2012) and Fishman, Washington, and Haworth (2013). DeMaio (2009) and Midgley (2011) review the bike-sharing schemes in relation to infrastructure and access to the service, as well as to business-management model. The authors show that the most common type of bike-sharing scheme has emerged in the late 90’s and it is characterized by automatic, 24-hour provision of bikes at specific points across the city. The access is on a self-service basis requiring subscription and use of smartcards. Every time a bike is picked up or returned, the details of the transaction (e.g. user ID, bike ID, station location, time) are recorded, generating large amounts of automatic and rich bike mobility datasets. This source of information is applied for a variety of purposes in the transport research field (Froehlich, Neumann, & Oliver, 2008), and the number and the quality of algorithms and tools to analyze fully-automated BSS data are steadily increasing.

O’Brien, Cheshire and Batty (2014) compare the bike movements in 38 BSS systems around the world, finding 6 dominant patterns. The study suggests a classification of the schemes considering the attributes of the demand and the features of the scheme infrastructure. The level of service of a BSS depends on the management and maintenance of the bikes. Therefore, the bike imbalance generated by asymmetrical demand is one of the most common problems in literature. The imbalance is linked both to the location of the docking stations in relation to the

geography of the city served by the BSS (people tend to cycle downhill), and to the urban land use and daily activity patterns. Froehlich et al. (2008) study the availability of bicycles throughout the day, using a clustering method to understand the relationship between docking station locations and public bike mobility in Barcelona. They prove that the availability of bicycles at different times of the day is linked mainly to land use. The conclusion is confirmed by Vogel, Greiser, and Mattfeld (2011), who apply 3 clustering techniques to the case of Vienna BSS. Borgnat and Abry (2011) implement the same clustering technique to study the flows between the stations of the Vélo'v system in Lyon. Demand evolution, daily patterns and trip distance distributions are analyzed. The clustering method permits the authors to describe the mobility dynamics: trip purposes such as commuting, university and leisure are assumed from the distribution of time and space among each group of rentals. The same BSS is analyzed by Jensen, Rouquier, Ovtracht and Robardet (2010), who find that the demand for bikes doubles when the other public transport modes are on strike.

The occasional studies that directly or indirectly focus on the impact of BSS on citizens' mobility choices lead to diverse conclusions, possibly caused by different cycling culture and bike penetration. The uptake of bike-sharing is studied in the case of Hangzhou (China) by Shaheen, Zhang, Martin, and Guzman (2011). A survey including both members and non-members shows that the demand for the BSS comes from a range of modes, including private cars and taxis, but also walking and transit. Interestingly, in Hangzhou BSS members have higher access to private cars than non-members. However, most studies generally conclude that it is difficult to induce a modal shift from private cars to shared bikes and that, when this happens, the change is small anyway (DeMaio, 2009; Fishman et al., 2013; Midgley, 2011). In line with this, Efthymiou, Antoniou, and Waddell (2013) estimate ordered logit models to assess the willingness to combine the use of cars and BSS. They find that bike-sharing is more attractive to people who currently walk.

When the BSS infrastructure is successfully integrated within the public transport network, bicycles can be used to reach zones not covered by other mass public transport modes, thus fostering the possibility of multi-stage trips and providing an alternative to private cars. For instance, in Paris, the number of multimodal trips including the use of bicycles increased after the launch of the Vélib' system (DeMaio, 2009). A positive synergy has also been proved between BSS and rail systems (Martens, 2004, 2007; Rietveld, 2000a, 2000b). Jäppinen, Toivonen, and Salonen (2013) offer an interesting contribution on the potential of BSS to enhance the overall public transport supply, showing that the provision of bike-share infrastructure has reduced travel time by more than 10% in Greater Helsinki.

The rapidly expanding body of literature using transaction data to study BSS takes for granted that this kind of data cannot provide insights on trip purpose or user satisfaction. Investigations on these aspects are normally based on revealed or stated preference experiments. For instance, Bordagaray, Dell'Olio, Ibeas, Barreda, and Alonso (2014) study the perceived quality of service of the BSS in Santander by means of questionnaires. The possibility of using transactions data to infer the trip purpose is the challenge that motivates the approach presented in this paper.

3. Types of use of bike-sharing systems

We propose an algorithm to classify some of the trips of BSS users in well-defined usage types by mining datasets of automatically collected transaction information. Evaluating the incidence of each kind of usage type and the related mobility patterns (origin and destination, time and duration of the trips) can shed light on the actual use of the BSS and drive the operator in (re-) designing and managing it. The procedure considers different attributes of the smart-card transaction regarding a bike rental (in the following rental and trip are considered synonymous): user ID, bike ID, origin and destination stations, time-stamps of the collection and of the return. When the users carry out more than one trip per day, their behavior is evaluated by relating each trip to the previous and the following one made by the same user in the same day.

The algorithm shown in Figure 1 evaluates the following conditions to classify the current rental:

- The coincidence of the stations at origin and destination. If this is the case, the rental is classified as “round trip”, the first BSS usage type.
- The existence of multiple rentals from the same user in the same day. If the user has rented only one bicycle during the day, his behavior cannot be classified and a new rental is considered.
- The time elapsed between the return of a bike and the next rental of the day by the same user. If this dwell time (DT) exceeds a predetermined threshold (AT, activity threshold), it is assumed that the user has performed an activity different from travelling between the two bike rentals.
- The sum of the durations of two subsequent trips (TT), when a rental begins immediately after the previous bike has been dropped off by the same user. BSSs generally permit the use of a bike with no extra cost within what it is called “flat fee interval” (FFI) that may not be enough, depending on the trip conditions and purpose. As a result, it is advised that the threshold ensures an accurate discrimination between the use of the bike for transport and a recreational ride. In literature, the maximum duration of a rental for transport purpose is assumed equal to 15-20 minutes, corresponding to a distance range of 2-5 to 7-8 kilometres (González, Melo-Riquelme, & Grange, 2015; Jensen et al., 2010; Midgley, 2011; Pucher & Buehler, 2008). Above this threshold, bikes are considered rented for leisure or physical exercise. Consistently, most BSS have a FFI of 30 minutes; however, Santander, the city that constitutes our case study permits a 60-minutes use without an extra charge, and so since it more than doubles the range assumed for transport, we assume a total travel time of 40 minutes in the illustrative application of the algorithm to distinguish a recreational demand for bikeshare from a rental for transport.
- The symmetry of two subsequent trips. Symmetric trips are especially linked to commuting and other recurrent daily activities. In any case, whatever the activity at destination may be, it is important to detect symmetric trips because they correspond to a use of the bike as a transport means to get to where the main activity will take place.

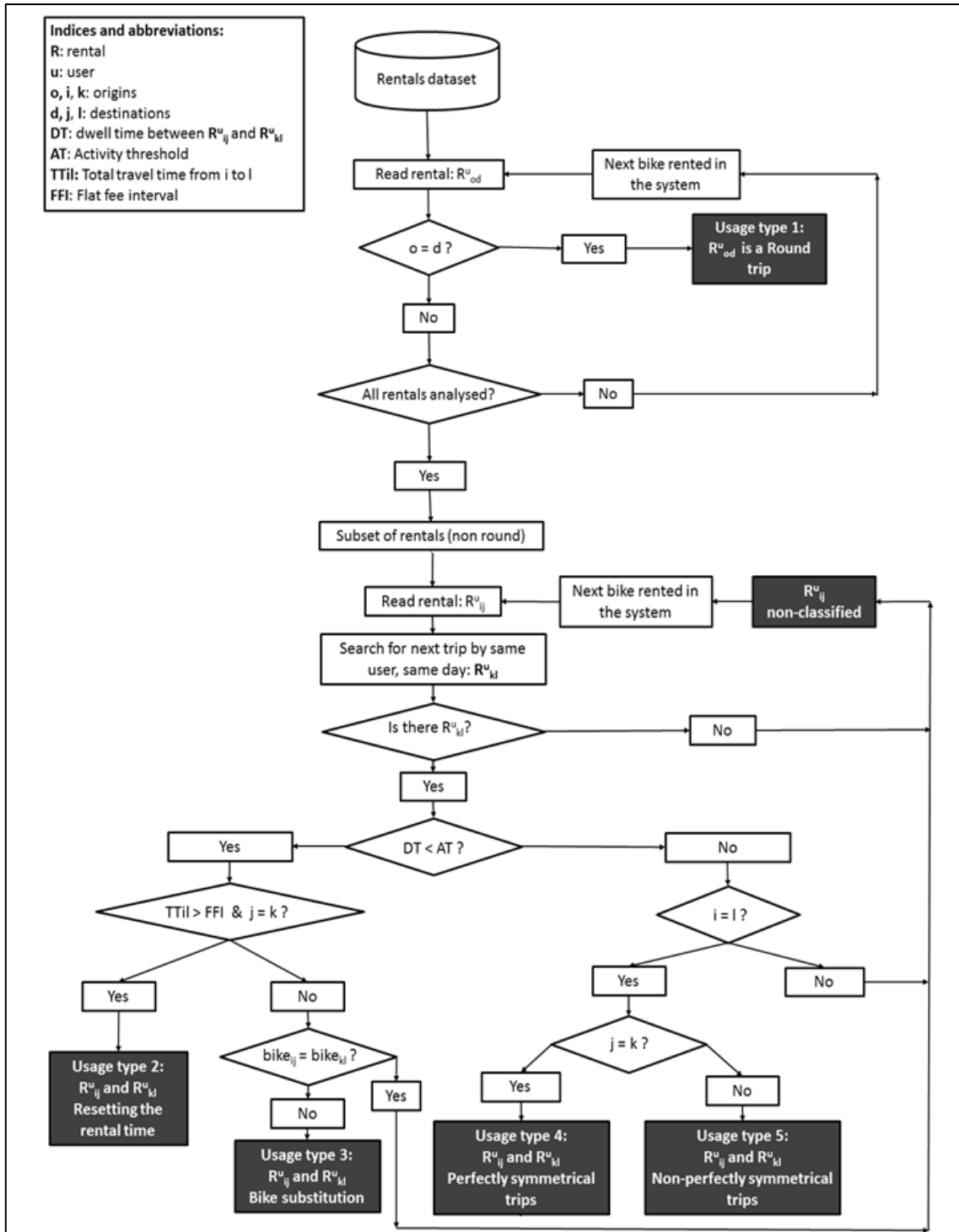


Figure 1. Algorithm to describe the usage behavioral framework within a BSS

The implementation of the algorithm classifies bike rentals in 5 usage types. Our hypothesis is that each of these types is characterised by sufficiently homogeneous mobility patterns, whereas the trips that cannot be classified by the algorithm (isolated trips, not part of a sequence) do not follow any systematic pattern. A description of the 5 rental categories and a short discussion of how the study of their features can give useful insights on the system performance is provided below.

Usage type 1: Round trips

Round trips are those in which the origin coincides with the destination ($o=d$). These are the first identified in the algorithm (Figure 1). Round trips are expected to be associated mainly to recreational rides or to rides made for physical exercise. The corresponding trip may not have taken place at all if the BSS had not been available. However, some round trips may also be made to perform short activities (e.g. quick shopping), during which the bike is locked somewhere in the street. In some cases, the user may decide to pay an extra fee to carry out longer activities and park the bike in the street where no docking station is available.

A precise characterisation of the round trips in terms of rental length and of land use of the zones served by the most demanded terminals can support decisions regarding the duration of the FFI, which should be decreased to discourage the recreational use of the shared bikes. The number and the spatial distribution of the round trips exceeding FFI may inform on the need for the relocation and/or the increase of the number of docking stations, so that bikes can be returned to a station during long activities, avoiding additional fees for the user and reducing bike idle time.

Usage type 2: Rental time reset

Some users requiring a bike for a longer period than the FFI may try to avoid extra fees by resetting the rental time, that is, by returning the bike and immediately renting a new one or even the same. This usage type is identified by the following conditions:

- [1] The dwell time between two subsequent trips in the same day is shorter than the activity threshold ($DT < AT$), i.e. no activity takes place between the two rentals.
- [2] The second bike (which may or may not be the same of the previous transaction) is rented at the same station where the one used in the previous trip is dropped off ($j=k$).
- [3] The overall travel time (considering the two rentals) is greater than the FFI ($TT_{il} > FFI$).

Condition [3] characterises a usage pattern not completely compatible with the sharing essence of the BSS, which is enforced (to a certain level) by levying extra fees for rental longer than the flat fee period. Note that the anxiety caused by the willingness to avoid extra fees could induce to reset the rental duration even when the ride would not be longer than the flat fee interval. A short first rental could also be due to need to change a faulty bike for a failure and not to the will of resetting the rental duration. However, the definition of a threshold of “shortness” requires further research.

If the incidence of the rental time reset usage is deemed unacceptable, the BSS operator may decide to introduce a minimum interval between two rentals by the same user. An interesting sub-group of rentals are the “pseudo round trips”, i.e. rental time reset trips with the origin of the first rental coinciding with the destination of the second one. Pseudo round trips are expected to be typical of the recreational usage of the bikes. The analysis of the spatial patterns would provide further insights on the type of terminals that stimulate this type of use of the BSS: residential or recreational areas.

Usage type 3: Bike substitution

This is the case in which immediately after picking up a bike, the user realises that it is defective and so he returns it and rents a new one. Such a usage type is characterised by the following conditions:

- [1] The dwell time between two subsequent trips in the same day is shorter than the activity threshold ($DT < AT$).

- [2] The destination of the first trip is the origin of the second one ($j=k$);
- [3] The bike in the second transaction is different from that of the previous one ($\text{bike}_{ij} \neq \text{bike}_{ki}$);
- [4] Different from condition [3] in the previous usage type, the cumulative duration of the two rentals is shorter than the flat fee interval ($TT_{i1} < \text{FFI}$).

The first two conditions are the same characterising the usage type 2, the third is a sign that the user does not like the bicycle rented in the first trip and so it chooses a different one, whereas the fourth suggests that the intermediate stop is not linked to the attempt of avoiding the additional fees. Since the behavior is triggered by a random event like the rental of a faulty bike, the travel patterns characteristics of this usage type are expected to be less homogeneous than those of the other kinds of use. In any case, as long as the total travel time is not longer than the FFI, the data only permit to guarantee that a bike substitution has taken place under these circumstances.

Bikes that are frequently involved in the first leg of this type of rentals should be checked as potentially defective. Trips with bike substitution can be identified by the methodology brought forward by Bordagaray, Fonzone, dell'Olio, and Ibeas (2014), which should be applied right before this algorithm. In the previous research, however, the assumed faulty bikes are returned to the system within the first 5 minutes of rental, right before the user rents a new bike at the same terminal. The first short rental is not assumed a trip itself and is thereby removed from the dataset. After the data undergo the previous detection, the algorithm in this research, detects the case when the bike has turned faulty on the way, while using it. The user experiences a bad performance of the bike and decides to change it for a different one (as established in condition [3] above).

Usage type 4: Perfectly symmetrical mobility trips

Symmetric trips (in which the destination of the first leg is the origin of the second one and vice versa) are typical of commuting mobility. Indeed, this characteristic is the key concept of the trip chaining approach widely used in the transport studies based on ITS data, as pointed out in the literature review in section 2. We make use of this feature to identify trips in which the bikes are presumably used as transport means. In this case, the following conditions hold:

- [1] The dwell time between two subsequent trips in the same day is longer than the activity threshold ($DT > AT$).
- [2] The origin of the first trip is the destination of the second ($i=1$).
- [3] The destination of the first trip is the origin of the second ($j=k$).

In this usage type, the use of the bike is ancillary to the activity carried out between the two rentals, including recreational activities. This implies that, in general, if there were no BSS, another transport mode (or a private bike) would be used. For this kind of use, existing literature finds travel times of 20 minutes on average.

An analysis of the dwell time and the type of terminals demanded for this usage type would be insightful on the characterisation of the trip conditions that stimulate cycling mobility.

Usage type 5: Non-perfectly symmetrical mobility trips

Like the previous case, these trips imply an activity between the two rentals, but the second bike is picked up in a station different from that where the first bike is dropped off. This behavior is identified by the following conditions:

- [1] The dwell time between two subsequent trips in the same day is longer than the activity threshold ($DT > AT$).
- [2] The origin of the first trip is the destination of the second ($i=l$).
- [3] The destination of the first trip is different from the origin of the second ($j \neq k$).

In some cases, this usage pattern may be linked to the availability of spaces and bicycles. In fact, the asymmetry may suggest that it has not been possible to return the first bike at the desired terminal because of a lack of available spaces, or, alternatively, that the user has looked for a bike at the station where he left the bike previously but he could not find any. Of course, the asymmetry may be also due to the use of different transport means between the two rentals. The analysis of the spatial distribution (location, distance) of the return and pick-up terminals may help understand the dynamics underpinning this usage pattern, and providing for more capacity if needed.

The set of the 5 usage patterns have been identified through different research steps. On the one hand, bike deficiencies and capacity problems were reported by users when asked in personal interviews carried out by Bordagaray et al. (2015). On the other hand, the occurrence of “tricky” behaviors and so the need to detect them have been identified in a simple analysis of the bike rental duration (Bordagaray et al., 2014). The algorithm has been implemented in Python.

4. Application to TusBic BSS

The smart-card transactions of the TusBic system in Santander have been used to validate the above described data mining methodology. The application to a real case study also permits to confirm our hypothesis that the heterogeneity of travel patterns is reduced within the rentals of the same usage type.

4.1 Santander and TusBic

Citizens face transport choices every day. Their decisions depend on the available alternatives and their relative utility. Demographics, urban shape, available modes and meteorology are recognized key issues in travel behavior, especially in the decisions about cycling. Therefore the characteristics of the city, of its mobility and of the BSS service must be analyzed before implementing the algorithm. This step is essential both to set the thresholds in the conditions evaluated in the data mining procedure, and to support the interpretation of the analysis of the various usage types.

Urban context

Santander is a medium-sized coastal city with 180,000 inhabitants in the north of Spain. It is characterised by a series of parallel hills and valleys running from northeast to southwest. The northern coast of Spain experiences rainy weather and mild temperatures typically ranging from 5 to 25-30°C throughout the year.

The hilly topography together with the fact that the city lays in a peninsula generates a mainly linear transport network (Figure 2) since the steep slopes make transport cross connections

difficult. The bus network covers the entire area of the city, with a stop within 300 meters from any estate. All neighbourhoods are served by an average of 3-4 buses an hour in weekdays, 2-3 at weekends. All the lines pass through the city center.

BSS and bike lanes

As suggested in the literature, the infrastructure dedicated to cycling plays a key role in promoting the use of bicycles (Akar & Clifton, 2010; Dill & Carr, 2003; Dill, 2009; Pucher et al., 2010). As shown by Faghih-Imani, Eluru, El-Geneidy, Rabbat and Haq (2014), land use and facilities around the BSS terminals influence the demand for the BSS. Dell’Olio, Ibeas, Bordagaray and Ortúzar (2011) identify the infrastructure to foster cycling as a mode of transport in Santander.

At the time of the data collection (2011), TusBic was comprised of around 200 bikes and 14 terminals. Figure 2 shows the spatial distribution of the docking stations and in Santander and Table 1 describes the main features of the areas surrounding the terminals.

Apart from the ordinary bike parking facilities, which are spread all over city, the infrastructure dedicated to cycling avoids the steep slopes and is mostly located along the perimeter of the city. The main attractions in Santander are served by the BSS (Table 1): the city center (terminals 1, 11 and 13); the rail station and the inter-urban bus station (terminal 14); the major touristic attractions such as the Park of La Magdalena (terminal 9), the Casino (terminal 8), the coastline (terminals 2, 4, 8, 9, 10), Las Llamas Park (terminal 3); and the University Campus (terminals 6 and 7).

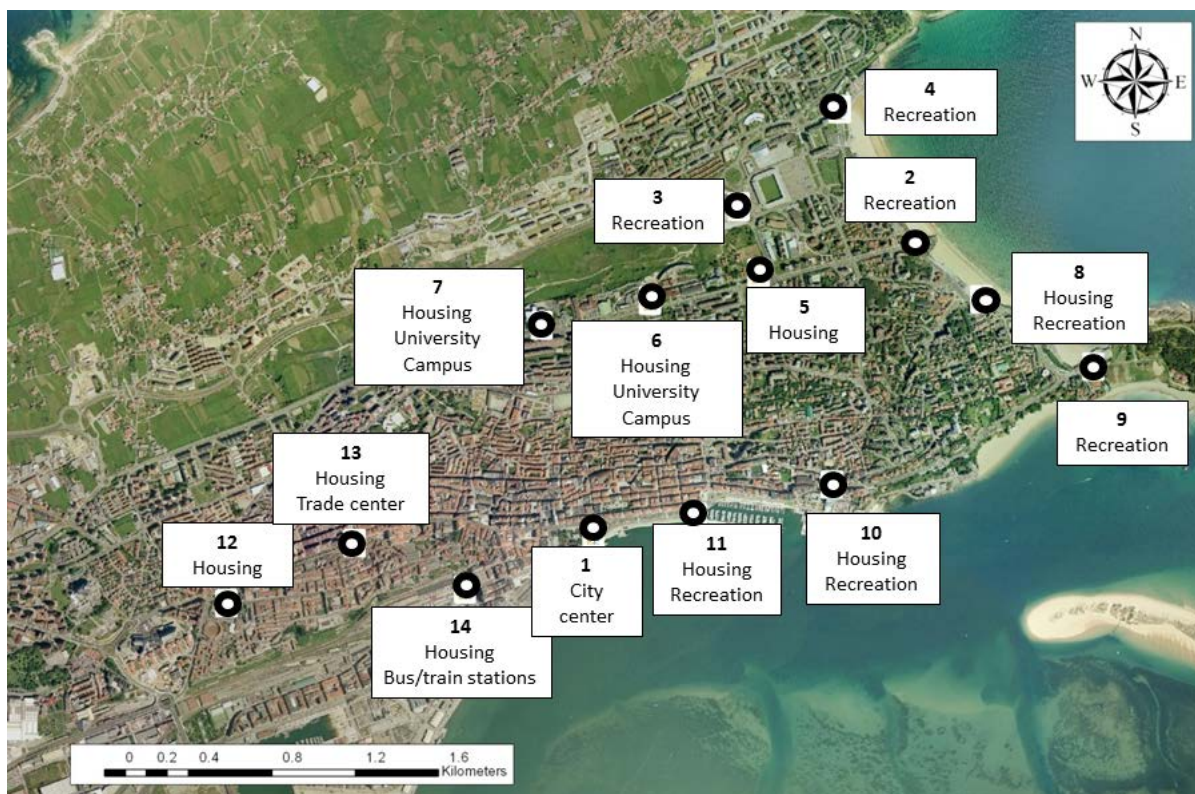


Figure 2. Localization of the TusBic terminals

| Terminal | Number of docks | Bike lane | Beach | Panoramic view | Trade center | City center | University | Housing | Long distance trains / buses |
|----------|-----------------|-----------|-------|----------------|--------------|-------------|------------|---------|------------------------------|
| 1 | 40 | ✓ | | ✓ | ✓ | ✓ | | ✓ | |
| 2 | 20 | ✓ | ✓ | ✓ | | | | | |
| 3 | 20 | ✓ | | ✓ | | | | | |
| 4 | 20 | ✓ | ✓ | ✓ | | | | | |
| 5 | 20 | ✓ | | | | | | ✓ | |
| 6 | 30 | ✓ | | | | | ✓ | ✓ | |
| 7 | 30 | ✓ | | | | | ✓ | ✓ | |
| 8 | 15 | ✓ | ✓ | ✓ | | | | ✓ | |
| 9 | 15 | ✓ | ✓ | ✓ | | | | | |
| 10 | 15 | ✓ | | ✓ | ✓ | | | ✓ | |
| 11 | 25 | ✓ | | ✓ | ✓ | ✓ | | ✓ | |
| 12 | 15 | | | | | ✓ | | ✓ | |
| 13 | 25 | | | | ✓ | ✓ | | ✓ | |
| 14 | 30 | | | | ✓ | ✓ | | ✓ | ✓ |

Table 1. Characteristics of the TusBic terminals

BSS fare structure

The fare structure and the types of subscription are key attributes of a BSS, as maintained by Midgley (2011). TusBic is open to general public. Three kinds of subscription are available: annual (10€), weekly (5€) and daily (1€). Weekly and daily subscribers (casual users) can register directly at the terminals by presenting a credit card. The card is charged with the fee and with a deposit of 150 €. Annual users must register online in advance. In Santander, the flat rate period during which the same bike can be used without additional charge, lasts 60 minutes.

4.2. Data description

Our initial dataset (24,664 observations) includes the data collected in the summer 2011 (62 days in July and August). The TusBic demand and its variability in terms of travel patterns peak during the summer months because of several reasons:

- The summer weather is the most favourable to use the bikes. Previous research in Santander proves the negative influence of the bad weather on the choice of bicycles as mode of transport (Dell’Olio et al., 2011). The use of bikes sharply decreases during the rainiest months of the year, and July and August account for 30% of the yearly TusBic demand.
- In the summer months, the bus system also experiences peak demand in the sunny days, due to demand of residents and tourists heading towards the seaside or the city center. This contributes to increase the attractiveness of the BSS.

- During summer, the variability of bike travel patterns peaks because all types of trip purposes can be expected given the simultaneous presence of tourists and residents, and the unstable weather.

Transactions corresponding to “bike trials”, in which a bike is rented and then immediately substituted with a different one, are removed from the analysis. Vogel et al. (2011) systematically remove the rentals lasting less than 60 seconds whereas a more accurate approach is brought forward by (Bordagaray et al., 2014). By applying the latter methodology, we detect 5% of the total number of rentals to be removed for not being a trip but bike trials; 75% of them are made by registered users, this is, annual subscribers.

The data concerns 193 bicycles, with an average of 2 trips per day per bike. Compared to the statistics published by Fishman et al. (2013) and O’Brien et al. (2014), the demand for the Santander BSS appears quite low. On average, 1 bike a day is rented per subscriber. The annual members show the greatest variability of the number of rentals per day, however 90% of them do not exceed 17 rentals in the two months.

4.3. General demand patterns

Figure 3 shows the average hourly BSS demand compared to the demand for bus and private car in the city. The figure provides interesting insight on the mobility patterns in Santander. In Figure 3a it can be seen that the demand for buses and cars varies in a smoother way than that for shared bikes. The car use experiences a peak at lunch time, from 2 to 4 pm, in line with the Spanish custom of going back home for lunch. Different from the BSS in Barcelona and Zaragoza (Froehlich et al., 2008; Kaltenbrunner, Meza, Grivolla, Codina, & Banchs, 2010; O’Brien et al., 2014) that follow the same trend of motorised modes, in Santander the lunch time corresponds to a significant fall in demand for shared bicycles. This may indicate that cycling is generally considered an activity in itself, which in most cases takes place in the late morning and after the late afternoon. Nevertheless, the minor peak from 7 to 9 am may be ascribed to commuting bike trips since it occurs at the same time of the peaks of the other transport modes (Alonso, 2010). In Figure 3b the demand for bikes is disaggregated per type of user and day of the week. Major differences emerge between the rental rates of annual subscribers and more occasional users, but the general trend is common to all cases, with peaks taking place at mid-morning and late afternoon. The figure highlights the occurrence of a morning peak period (7 to 9 am) of the annual subscriber demand in weekdays that does not occur at weekends and for casual users. The finding suggests the existence of a group of regular TusBic users, who commute by shared bikes.

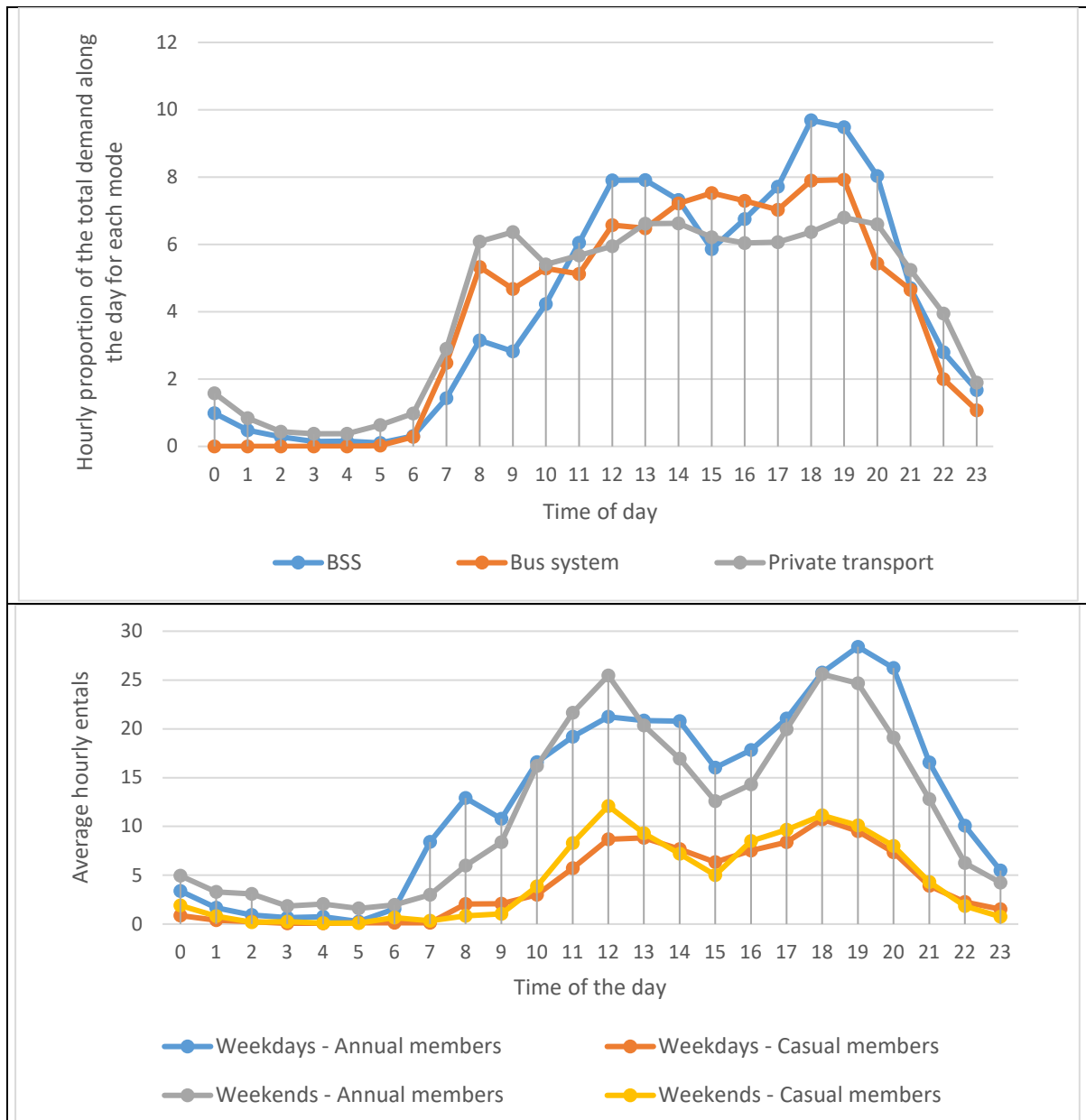


Figure 3. Daily profiles of transport demand: a) Demand for BSS, buses and private cars; b) Average number of TusBic rentals

4.3. Implications of the usage types in the travel patterns

The conditions evaluated in the algorithm have been specified as follows for the present case study:

- The activity threshold (AT) has been assumed equal to 15 minutes. Bordagaray et al. (2014) find that trips in which users change bicycle and the new rental starts no later than 15 minutes after the return of the first bike can be classified as bike trials. Noting that bike trials are a specific case of usage type in which no activity takes place between the two rentals, we think it is reasonable to extend the result of Bordagaray et al. (2014) and to assume no activity has taken place in any case in which the interval between two consecutive rentals of the same user is shorter than 15 minutes. Considering (a) a walking speed of 6km/h and a walking distance of 300m between the docking station and the location of the activity, and therefore a walking time of 6min; and (b) 2 minutes

overall to return the first bike and pick up the second; taking AT equal to 15min means that we assume that no activity can last less than 7min.

- As said, the flat fee interval of the Santander BSS (60min) is longer than the commonly adopted value (30min). Therefore, instead of the FFI itself, we take 40 minutes to distinguish between a bike substitution and the will to avoid being charged extra fee assuming, on the basis of the literature (Dill & Gliebe, 2008), and also considering the size of the city, that longer bike uses are recreational trips. Indeed, Dill and Gliebe (2008) concluded a rather different behavior between utilitarian cycling and other purposes. Cycling as a mode of transport was characterised by an average of 25 minutes of travel time in their study, whereas cycling for exercise resulted in around 45-50 minutes. Consequently, an interval longer than 40 minutes is sufficient to assume purposes other than utilitarian.
- We consider two rentals as being made in the same day if they take place in the 24h interval starting at 6am of a calendar day. This allows characterising return trips from night events as potential legs of journeys started before the midnight of the previous calendar day.

With the described assumptions, we can classify 47% of rentals in one of the usage types defined by our algorithm. As previously mentioned, non-classified records are isolated bike rentals, not related to other trips in the same day. The percentage of classified trips can be expected to be characteristic of each BSS. It can provide initial insights on how the system is conceived and thereby used.

Figure 4 shows the classification of the rentals obtained with the algorithm in Figure 1, distinguished by registered (Figure 4a) and casual (Figure 4b) users. Most of casual users' rentals are round trips (almost half of the transactions) whereas round and perfectly symmetrical trips are almost equally represented in the annual card-holders' rentals and overall the two categories collect 70% of trips. Annual subscribers substitute bikes more often. These differences complement the research by Lathia et al. (2012), in this case the comparison being made in terms of usage types. Lathia et al. (2012) found a higher demand for the BSS in London during weekends and a pattern change at some terminals after the policy that allow casual users was implemented, suggesting diverse travel behavior between registered and occasional users. From an alternative perspective, this research supports the conclusion drawn by Lathia et al. (2012) about the different cycling dynamics of each type of user.

Note the operator should place importance in the second and third usage types presented in the graphs in Figure 4 since that demand is related to the will to reset the rental time or to substitute the bike. The first behavior might compromise the capacity of the system as it goes against the sharing nature of the service, whereas the second one is interpreted to be likely caused by a bad performance of the bike, which will influence in the overall perception of quality that the user may have, as concluded by Bordagaray et al. (2015).

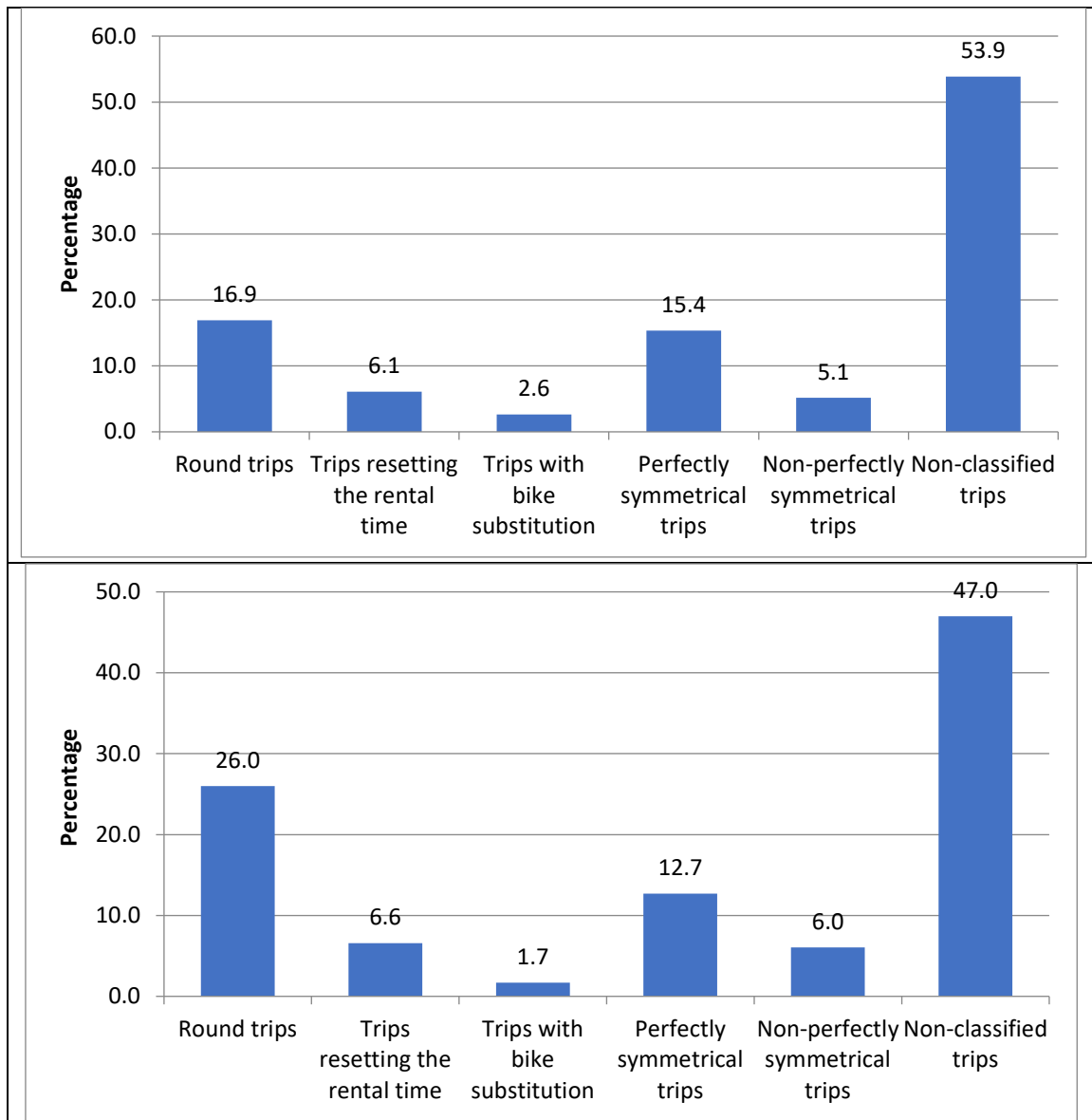


Figure 4. Usage types in TusBic: a) Annual subscribers; b) Weekly and daily subscribers

4.4. Algorithm validation

To validate our interpretation of the reasons underpinning the different usage types, we have estimated binary probit models explaining the probability of having a specific kind of trip with the type of subscription, the rental features (duration and time). Both logistic and normal distributions may be assumed in order to conceptualize the generation of a trip as a binary regression. The hypothesis of the errors distributing normal was accepted in this case due to the fit and the results supporting the common sense.

The models are shown in Table 2. In estimating binary models, the interest is not in measuring how much each independent variable changes the probability of having one behavior with respect to one other behaviour, as it would be in the multinomial case, but compared to all other behaviors. In other words, we assume that the dependent variables of the binary models are not elements of the same multiple-choice exercise for the users, who may not even be aware of the choice set. Rather we consider that each particular type of rental (e.g., a bike trial) takes place in consequence of given situations occurring during the trip (for instance, hiring a faulty bike)

and the aim is to examine how much the independent variables explain each of the behavior strategies separately.

The rental duration and the characteristics of the origin and destination terminal of each trip (described in Table 1) are the potential independent factors validating the heterogeneous travel demand across the defined usage behaviors. In the case of bike substitution and rental time reset types of usage, the pairs of subsequent rentals have been introduced as one unique journey, with the rental duration equal to the value of the total travel time (TT_{ij}) in the algorithm, this is, the sum of the durations of the two rentals. The peak period is related to the time-stamp at the rental origin; in the case considering the sum of the pair of subsequent rentals a unique one, the peak period refers to the origin of the first trip leg.

As shown by Table 2, the occurrence of rentals in peak periods (7-9am and 1-4pm in weekdays) increases the probability of symmetrical trips and bike substitutions. This seems to confirm that symmetrical trips are associated to commuting, which normally takes place in peak times. The influence on the bike substitution usage type may be due to the shortage of well-functioning bicycles in relation to the demand in peak times. In all models but the one concerning the non-perfectly symmetrical trips (in which the variable is not significant), the impact of the travel time and the peak period have opposite effects. This further corroborates the interpretation of symmetrical trips (more frequent in peak times) as trips with mobility purposes, normally shorter than trips for leisure or physical exercises that we think associated to rental time reset trips (more frequent in non-peak times). Finally our hypothesis that symmetrical trips are typical of mobility-related usage, whereas round and rental time reset trips are more common for recreational purposes is supported by the findings concerning the origin of each usage type: symmetrical trips are more frequently generated in residential areas, whereas the other trips tend to begin in non-residential zones. Finally, it should also be highlighted the influence of bike lanes on the probability to rent a bike. As shown by the models, such an infrastructure stimulates the generation of trips except in the case of the terminal where the bike is assumed to be changed to reset the rental time and in the case of symmetric pattern. In fact, together with the positive effect of the peak period on the symmetric behavior, the negative effect of bike lanes in this case means the terminals demanded by the users in this case cannot be reached by a bike lane, so there should be other attributes affecting this demand positively, such as the terminals being located in residential areas, as informed by the Model 4 (Table 2). On the contrary, those behaviors mainly associated with recreational cycling are shown to be positively affected by bike lanes as is the case of round trips and rental time reset.

| | Model 1 - Round trips | | Model 2 - Rental time reset | | Model 3 - Bike substitution | | Model 4 - Perfectly symmetrical journeys | | Model 5 - Non-perfectly symmetrical journeys | |
|--------------------------------|-----------------------|---------|-----------------------------|---------|-----------------------------|---------|--|---------|--|---------|
| | Coefficient t | p-value | Coefficient t | p-value | Coefficient t | p-value | Coefficient t | p-value | Coefficient t | p-value |
| Annual subscriber | -0.087 | 0.004 | -0.880 | 0.001 | 1.170 | 0.000 | -0.068 | 0.075 | -0.220 | 0.027 |
| Rental duration | 0.007 | 0.000 | 0.210 | 0.000 | -0.261 | 0.000 | -0.019 | 0.000 | 0.003 | 0.174 |
| Peak period | -0.540 | 0.000 | -1.458 | 0.018 | 1.541 | 0.025 | 0.681 | 0.000 | 0.463 | 0.017 |
| Origin/Destination_residential | -0.994 | 0.000 | - | - | - | - | - | - | - | - |
| Origin/Destination_bike lane | 0.176 | 0.000 | - | - | - | - | - | - | - | - |
| Origin/Destination_capacity | 0.017 | 0.000 | - | - | - | - | - | - | - | - |
| Origin1_residential | - | - | -0.468 | 0.145 | 0.127 | 0.715 | 0.830 | 0.000 | 0.298 | 0.012 |
| Origin1_bike lane | - | - | -1.033 | 0.001 | 1.113 | 0.002 | -0.146 | 0.002 | 0.095 | 0.314 |
| Origin1_capacity | - | - | -0.029 | 0.065 | 0.035 | 0.048 | 0.000 | 0.913 | 0.012 | 0.027 |
| Destination1_residential | - | - | -0.118 | 0.693 | 0.197 | 0.558 | -0.148 | 0.002 | -0.284 | 0.009 |
| Destination1_bike lane | - | - | -2.715 | 0.000 | 2.939 | 0.000 | -0.112 | 0.036 | 0.081 | 0.584 |
| Destination1_capacity | - | - | -0.061 | 0.001 | 0.069 | 0.001 | -0.014 | 0.000 | 0.011 | 0.114 |
| Origin2_residential | - | - | - | - | - | - | - | - | 0.014 | 0.875 |
| Origin2_bike lane | - | - | - | - | - | - | - | - | 0.006 | 0.946 |
| Origin2_capacity | - | - | - | - | - | - | - | - | -0.010 | 0.110 |
| Destination2_residential | - | - | 0.337 | 0.073 | 1.191 | 0.001 | - | - | - | - |
| Destination2_bike lane | - | - | 0.184 | 0.303 | 1.451 | 0.000 | - | - | - | - |
| Destination2_capacity | - | - | -0.078 | 0.000 | 0.042 | 0.024 | - | - | - | - |
| Log-likelihood function | -4863.510 | | -84.681 | | -67.787 | | -3367.348 | | -583.319 | |

Table 2. Binary probit models for each usage type (Indexes 1 and 2 for origin and destination refer to the first and following rental in the usage behaviors described in models 2 to 5)

As the models in Table 2 show, the characteristics of the origin and destination of a trip explain the usage types. To further investigate this issue, Figure 5 visualizes the spatial distribution of the origin of each usage type. For the sake of simplicity, bike substitutions are represented together with non-classified trips since the two groups are expected to be highly heterogeneous in terms of travel behavior, as previously argued. In the case of rental time reset, perfectly and non-perfectly symmetrical types, only first-leg trips are represented. This is because in trip chains the origin of the second trip is obviously dependent on the destination of the previous one. The figure confirms that spatial patterns of the different usage behaviors actually vary. Non-classified and bike substitution trips present small variations of the demand across terminals. On the contrary, the other usage types do show a clear preference for some of the terminals. These results support our hypothesis that the classified usage types are more homogeneous than isolated rentals.

The distribution of symmetric trips provides insights on the features of the demand. The most demanded terminals for symmetric usage types are 1, 5, 10, 11, 13 and 14. All these docking stations are located in residential areas (Table 1), which supports our thesis that symmetrical usage is typical of trips starting at home, like the commuting ones, as resulted from Models 4 and 5 in Table 2, where the peak period is confirmed to stimulate the generation of symmetric trips. The finding highlights the importance of the distance between terminals and housing estates and of network density household in determining the use of public bikes, as suggested by several existing studies (dell'Olio, Ibeas, & Moura, 2011; Fishman et al., 2013; García-Palomares et al., 2012; Maldonado-Hinarejos, Sivakumar, & Polak, 2014). The spatial distribution of demand at the origins of rental time reset trips is similar to that of symmetric behavior trips, meaning that also the former usage type is mainly home originated. The least demanded terminals at the origin of a symmetric usage are 2, 3, 4, 6, 7, 8, 9 and 12. Indeed, Stations 2, 3, 4, 8 and 9 are located in recreational open areas with panoramic views and close to the seashore. Stations 6 and 7 are residential but they also serve the university campus, so it can be concluded that they are more used as destination zones than as origins of outward routes.

Finally, recalling related literature (Akar & Clifton, 2010; Maldonado-Hinarejos et al., 2014; Pucher et al., 2010), the low demand for terminal 12 can be explained by two circumstances: it is a rather isolated terminal within the BSS network (Figure 2), and it is not served by a bike lane (Table 1). Terminal 1 is the most popular in any case. This can be expected, since it is the docking station with the largest capacity (Table 1) and it is located in the city center (Figure 2), attractive to all possible trip purposes. The most surprising result is the demand for terminal 3. As reported in Table 1, the location of this terminal permits great views of the skyline of the city, and also a bike lane serves and connects the park where it is located. However, it is not a touristic place since the coastline is not visible from the park and neither does the terminal serve the housing in the surroundings. However, the finding is explained by the land use: such terminal is located on the edge of the biggest park in Santander, a spot with no shops nor restaurants but an open and semi-natural area with lakes and paths for skating and cycling. Furthermore, there is a free parking lot next to the docking station, where people usually park to go for physical exercise or just to wander around. This suggests that trips originated at terminal 3 are primarily for leisure/exercise and not for mobility purposes, and therefore it confirms that the round trips are carried out as an activity per se.

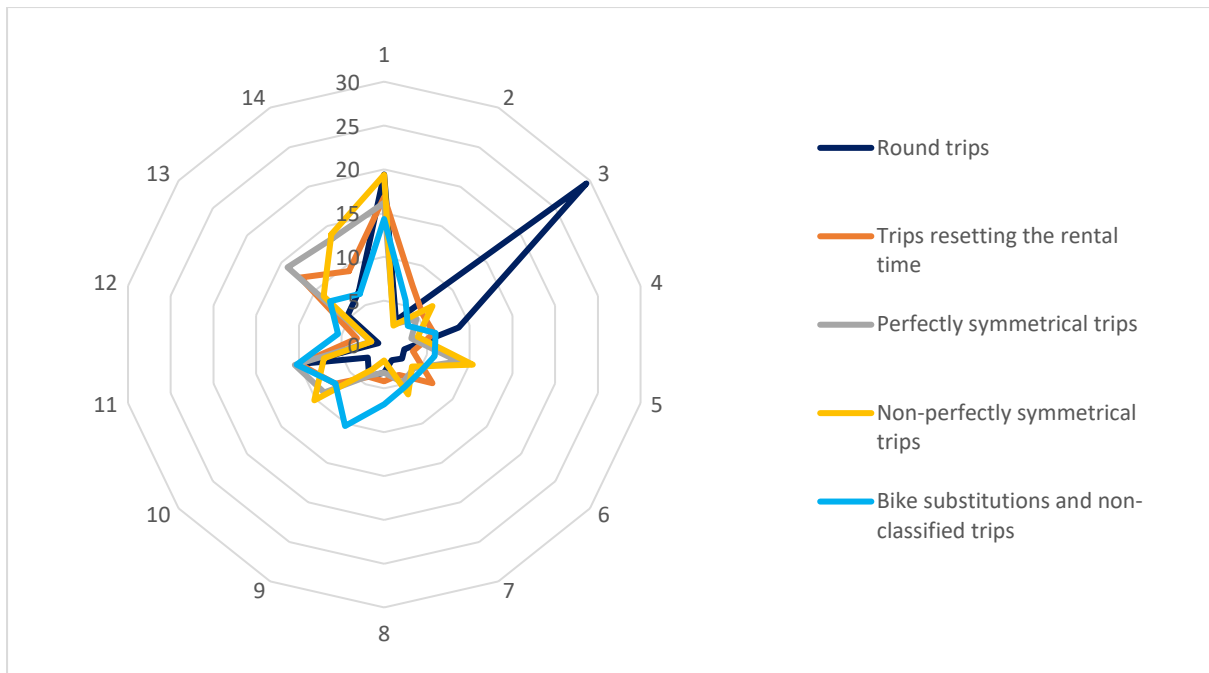


Figure 5. Spatial distribution of the origin of each usage type (percentage demand within the same usage type)

Travel times are also analyzed to characterise the differences between the various types of BSS usage. In Figure 6 the box plots describe the rental duration of each usage type identified by the proposed algorithm. The duration of the two subsequent rentals is considered in the case of rental time reset and in bike substitution. Trips above 500 minutes have been omitted and considered outliers.

As can be seen in Figure 6, round trips and those where the rental time has been reset show the longest uses of the public bikes, whereas perfectly symmetrical trips are the shortest, similar to non-perfectly symmetrical behavior and bike substitutions. Regarding the research by Dill and Gliebe (2008), the results confirm the hypothesis that the first two types are related to leisure or exercise cycling and symmetrical patterns with utilitarian cycling, in line with their findings of an average of 45-50 minutes for the first purpose and 25 minutes for the second.

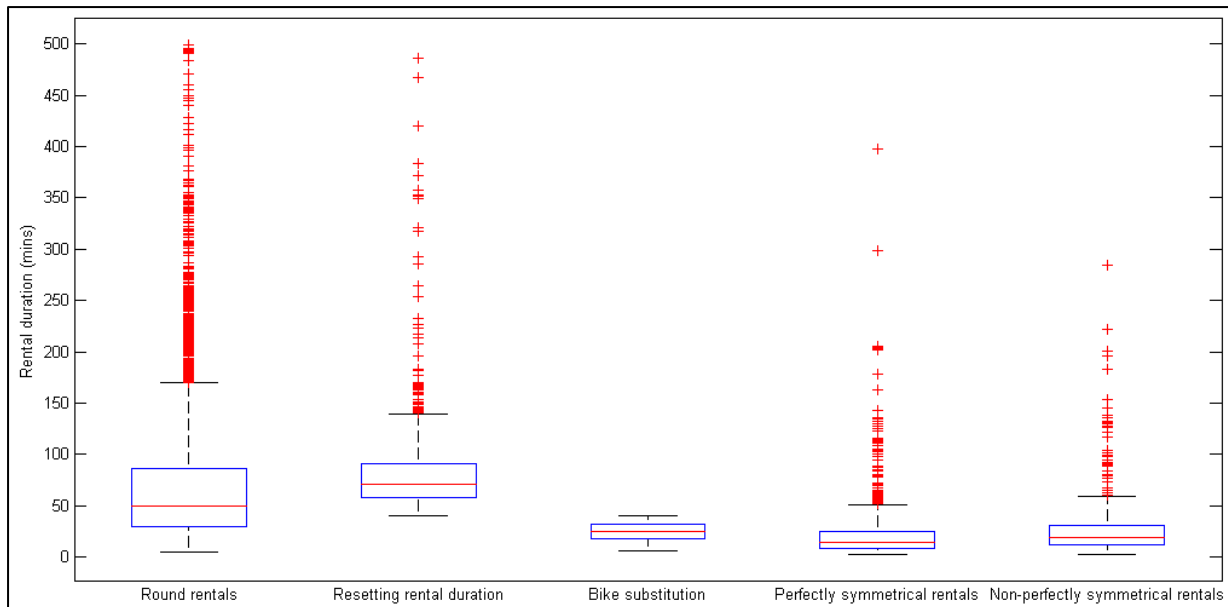


Figure 6. Rental duration of each usage type

The analyses in this section validate the developed algorithm because they show that it allows identifying different and easy-to-characterise usage patterns. In the following section, the usage types in the Santander case are analyzed in detail, to further illustrate the potential of the algorithm.

4.5. Characterisation of the travel patterns of each usage type

Round trips

More than 40% of all the classified rentals in Santander are round trips. As it can be seen in Figure 4, this usage type is more common among casual users (50%) than among annual cardholders (38%). 86% of round trips are longer than 20 minutes, 75% last more than 40 minutes, and 38% even overcome the flat fee interval of 60 minutes. Considering that 20 minutes is the average rental time in systems promoting a mobility-oriented use of public bikes (Jensen et al., 2010), and the duration of about 45-50 minutes ride for exercise informed by Dill and Gliebe (2008), the fact that a high proportion of round trips last longer corroborates the idea that they are mainly recreational rides. This is in line with the survey of Bordagaray et al. (2015), according to which 75% of TusBic round trips are carried out for leisure. In addition, the very probable recreational purpose of the ride in round trips was previously suggested with the results in Model 1 in Table 2: a negative effect of an annual subscription, peak period and residential terminals, at the same time that the rental duration and bike lane show a positive impact on the probability to make a round trip.

Rental time reset

With the proposed thresholds for the duration of an activity ($AT=15$ minutes) and the overall travel time (40 minutes instead of the FFI itself), 6.2% of rentals are classified as rental time reset. As explained earlier, we assume that this use of the BSS is related to the demand for public bikes for leisure or physical exercise. Resetting the rental time may aim to get round the fee imposed to ensure the sharing nature of the system. Two circumstances suggest that the goal of this category of rentals is to reset the time. Firstly, the same bike is actually rented in the two subsequent rentals in 40% of cases. This means that the bike was not returned because

it was faulty. Secondly, 34% of all cases are perfectly symmetrical trips, i.e. the second bike is rented right after the first one is returned and the user terminates their journey where it started. The pattern seems to be typical of a pseudo round trip, whose return travel time is long enough to induce the users to reset the time rental to avoid incurring the fee. 12% of cases simultaneously show the above-mentioned conditions: the same bike in the two consecutive rentals and perfectly symmetrical journeys. Note that when neither condition is true, the aim of this kind of renting behavior may still be resetting the rental time although the evidence provided by data is less strong.

Further indications on the nature of this usage type come from the analysis of the duration and of the geographical distribution. The average of the overall duration of the two rentals is 90 minutes with a standard deviation of 70 minutes. In addition, as shown by Model 2 in Table 2, the rental duration positively affects the probability to reset the rental time. Therefore, the overall journey tends to last longer than the 20 minutes that is considered the average duration of cycling for transport (Dill & Gliebe, 2008; Jensen et al., 2010; Pucher & Buehler, 2008); the average found in this research is even higher than the median found by Dill and Gliebe (2008) in the case of cycling for exercise, 8.5 miles and around 10 miles per hour, which means a 50-minutes ride approximately.

Figure 7 shows the distribution of the journeys corresponding to this renting behavior according to the origin of both the first and the second leg. The distribution of the origin of the first rental (already represented in Figure 5) shows the most popular terminals are located in residential areas. On the contrary, there is a clear tendency to reset the rental duration at touristic and recreational areas: terminals 3, 4, 9 and 11 (see Table 1 and Figure 2). This seems to support further the hypothesis that these trips are mainly leisure rides and the bike is returned only to avoid paying the additional fee.

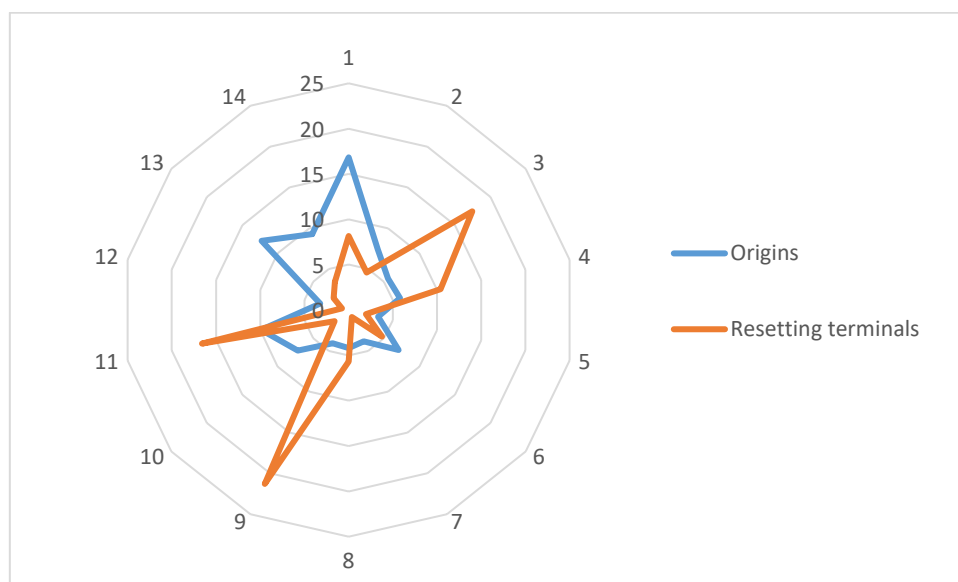


Figure 7. Spatial distribution of journeys with rental time reset (percentage of demand within the usage type)

Bike substitution

As pointed out in section 3, the analysis of this usage type allows identifying the bikes that do not work well and may need maintenance. Figure 8 shows the distribution of the frequency of

substitutions. While it is quite common for a bike to be substituted 1 or 2 times, bikes replaced more than 2 times are much less common. Therefore bikes with more than 2 replacements should be the first to be checked by the operator, which may optimize the redistribution route by considering this information.

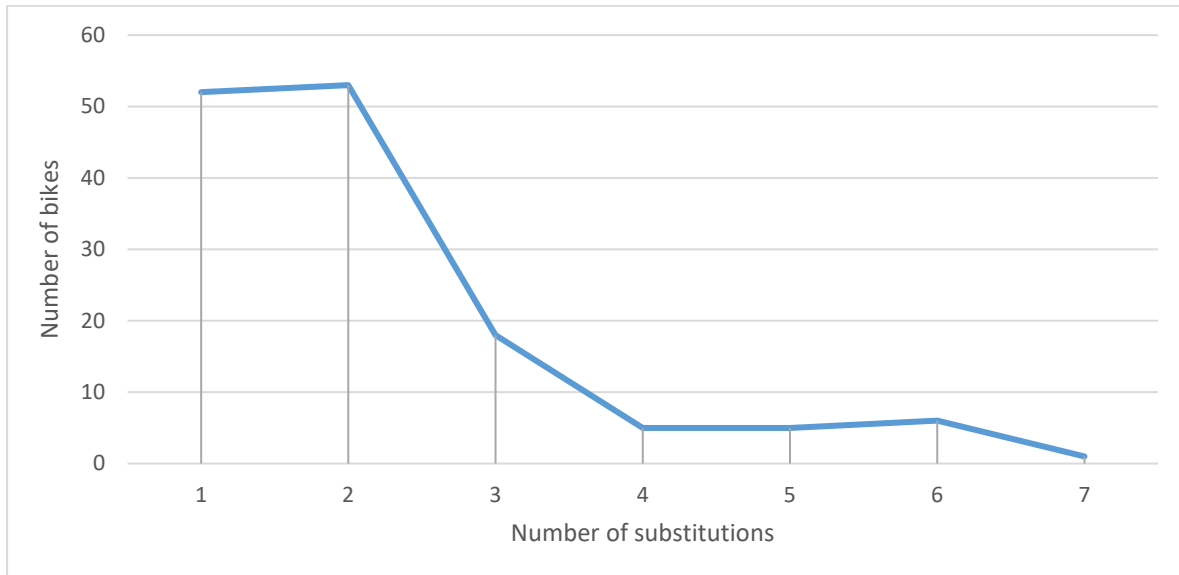


Figure 8. Number of bikes according to the number of times that they have been replaced by another bike

The average total travel time of these trips is 25 minutes (considering the two trip legs), and the standard deviation 10 minutes. Therefore, many of these trips could be considered made for mobility purposes, an argument that is also supported by the positive effect of an annual subscription and the peak period, as well as the negative impact of the rental duration, presented in Model 3 of Table 2. The average travel time also confirms the finding of the survey of Bordagaray et al. (2015), according to which users place more importance on the quality of the bike when their journey is shorter than 25 minutes. This may point to the fact that cyclists using the service for transport (limited travel times) are more “picky” in terms of bike quality.

As explained above, random travel patterns are expected for this usage type. The spatial distribution of origins and destinations represented in Figure 9 seems to confirm the expectation (the rentals ending with a bike substitution are considered single trips in the following analysis). In fact, although demand peaks at some terminals, it should be noted that all the docking stations are origin or destination of this kind of usage type, except for number 10 and 12. The latter is the least demanded in all usage types (Figure 5) so the concerning result is not surprising.

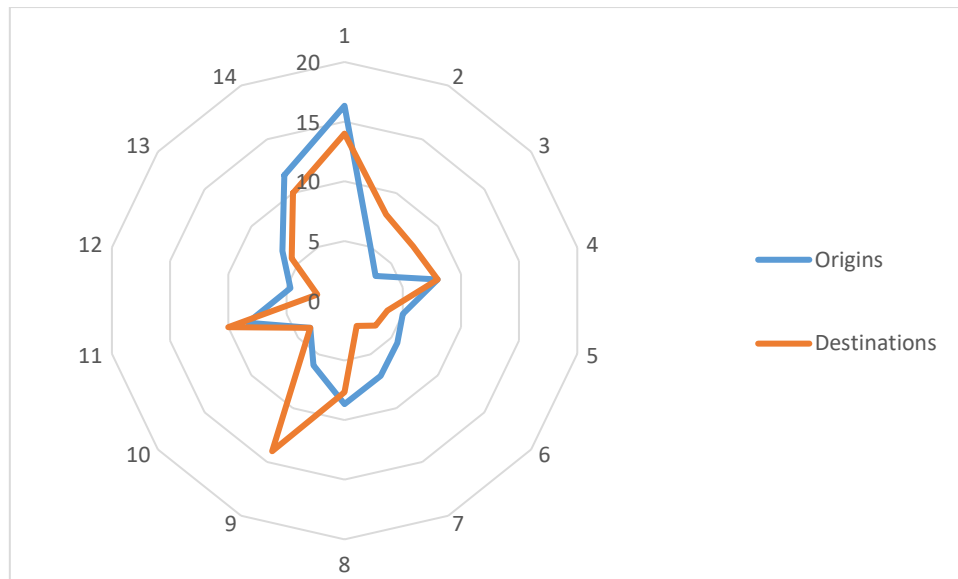


Figure 9. Spatial distribution of journeys with bike substitutions (percentage demand within the usage type)

Perfectly symmetrical trips

Perfectly symmetrical trips are more common among registered users (Figure 4). In addition, recalling the significant and positive impact of the peak period and the negative effect of the rental duration in Model 4 of Table 2, the results suggest that this bikeshare usage resembles a typical ride for mobility purposes.

The average travel time of each leg of a symmetrical journey is 20 minutes with standard deviation of 27 minutes and positive skewness. The average is in agreement with cycling travel times among users of the public bikes in Lyon (Jensen et al., 2010) or Santiago of Chile (González et al., 2015). The time between two perfectly symmetrical rentals, i.e. the duration of the activity that is supposed to be the purpose of travelling, ranges between 16 minutes to 13 hours with an average of around 3 hours.

Figure 10 compares the spatial distribution of the origins of perfectly symmetrical trips with that of the destinations, where the activity takes place. The information revealed by Figure 10 is in line with the results of Model 4 (Table 2). As expected, the destinations substantially differ from the origins. Destinations are localised mainly in the area of influence of terminals 9, 1, 8, 4, 13, 2, 11 and 14, consistently with activity distribution within the city (Table 1 and Figure 2). Furthermore, the demand concerning symmetric patterns at terminal 14 is considerably greater than other usage types (Figure 5). Since this docking station is close to the train station, the result supports the literature highlighting potential synergies between BSS and transit services (González et al., 2015; Jäppinen et al., 2013; Martens, 2004, 2007; Pucher & Buehler, 2008, 2009). The area, densely populated and far from touristic attractions, is not served by any bike lane. Literature provides evidences that experienced cyclists choosing the bike for transport are less influenced by the presence bike lane (Akar & Clifton, 2010; Dill & Gliebe, 2008; Dill, 2009). Therefore the fact that the lack of a dedicated lane does not discourage the demand in the case of the terminal 14 can be interpreted as a proof that the bikes are rented more for mobility purposes than for leisure.

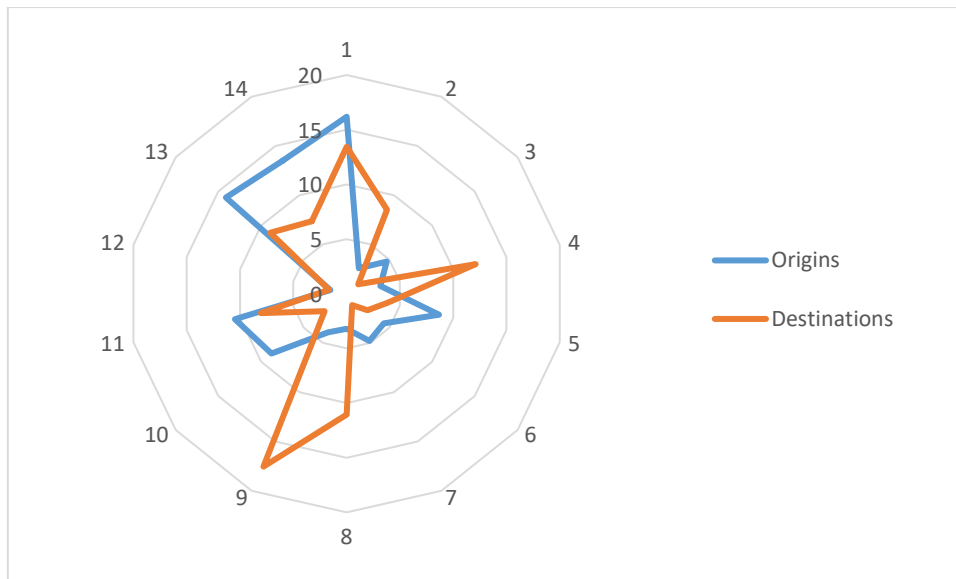


Figure 10. Spatial distribution of perfectly symmetric trips (percentage demand within the usage type)

Non-perfectly symmetrical trips

Similar to the perfectly symmetric renting behavior, the range of time between the two rentals on non-perfectly symmetrical trips ranges between 16 minutes to 10 hours with an average of around 3 hours. The average distance between origin and destination of this kind of trips is 990 meters, the standard deviation 580 meters, a distance that can be easily covered on foot. In some cases, the non-perfect symmetry might be due to an insufficient capacity of the destination terminals.

The spatial distribution of non-perfectly symmetrical rentals is represented in Figure 11. The profiles are similar to the analogous ones in Figure 10, although in this case the destination of the outward trips does not coincide with the origin of the return trips. The similarity of the distribution suggests that also these trips can be interpreted as activity-oriented. The land use (Table 1 and Figure 2) around the most demanded terminals (1, 2, 4, 8, 9 and 11) shows that in this case TusBic is used to reach leisure activities. Furthermore, this hypothesis was previously suggested by Model 5 presented in Table 2, where a residential origin and a non-residential destination of the first leg increase the probability for a non-symmetrical behavior.

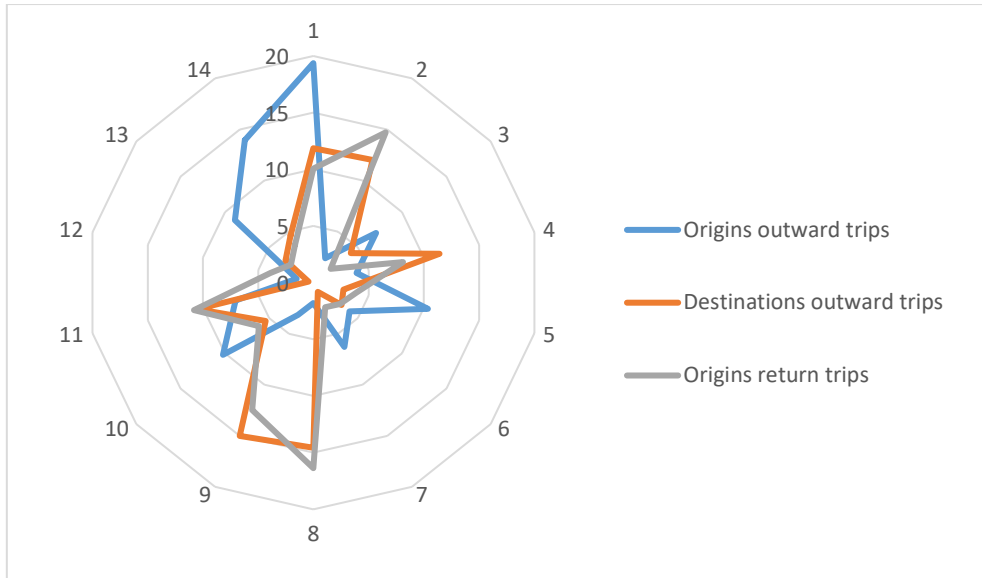


Figure 11. Spatial distribution of non-perfectly symmetric trips (percentage demand within the usage type)

5. Conclusions

The implementation of Intelligent Transport Systems within bike-sharing schemes makes big and rich datasets available to operators. The use of such data to understand and optimise system operations has been proposed and relatively widely explored in literature. Instead, our paper has put forward a method to take advantage of the passively collected data to derive information on the purpose of the bike rental (a tool for transport versus for recreation or physical exercise) from the features of the recorded rentals. We have developed a data mining algorithm aiming to classify rentals in usage types (round rentals, interval duration reset, bike substitution, perfectly symmetrical and non-perfectly symmetrical journeys) corresponding to very well defined user behaviors described in terms of symmetries, origin and destination coincidence or quick bike substitutions.

We have tested the algorithm by an application to a dataset concerning the bike sharing system of Santander, Spain. The limited size of the TusBic system, and our familiarity with it and the city have allowed a thorough interpretation of the results provided by the algorithm. We conclude that the groups of rentals built by the algorithms actually show the features that we were expecting in designing the data mining procedure. Therefore the algorithm can be considered validated. We note that the size of the samples made available by the current technology allows the identification of less usual mobility patterns such as bike substitutions, which would have been difficult to obtain by traditional survey techniques.

Our algorithm can tell the rentals having mobility as their primary scope from those in which cycling itself is the activity (for leisure and/or exercise). The spatial and temporal analysis confirms that the two groups of trips have different features, as found also by Heinen, van Wee, and Maat (2010). Significant differences exist also between the travel patterns followed by users with different types of subscription. This complements the research of Lathia et al. (2012). In general, the discussion of our findings confirms what is expected in the different types of BSS classified by O'Brien et al. (2014). For instance, that the incidence of symmetric trips is higher when the bicycles are used for commuting. Symmetric patterns are typical of daily commuting trips. Wherever the scheme regulations encourage people using the bikes to

commute to register, such symmetrical trips are expected to be wider represented among registered users than among casual ones. In contrast, casual users' demand can be expected to show greater heterogeneity and fewer symmetries. Therefore, casual users' demand (including the touristic demand) is expected to be mostly represented by rental time resets and bike substitutions, and to show an important proportion of unclassified rentals.

The implementation of the algorithm may provide useful insights to manage a BSS. Two specific usage types identified by the proposed data mining procedure are particularly relevant for operators: rental time reset and bike substitution. The former behavior might jeopardize the regular operations of the system, as it is against the sharing nature of the service. Once rentals that belong to this usage type have been identified, their time and spatial characteristics can be determined, and this can cast light on the conditions under which users benefit of a bike for a period longer than the flat fee interval without paying extra money. This is a very valuable information to improve and redesign the service (for instance, the FFI could be changed to half an hour in the case of Santander in order to motivate the bike rotation; also some terminals may need extra capacity such as number 3 at off-peak periods, and also an application could be developed that permits the user to indicate that the bike is faulty when returning the bike at the terminal and so that the bike is then blocked until it is collected for repair). The occurrence of a substitution may be a sign of bad performance of a bike, a circumstance which has a large negative influence on the users' overall perception of the quality of the system, as concluded by Bordagaray et al. (2015). Since the bike IDs are recorded, they can be easily localised and a check can be included in their redistribution route.

It is recommended that the operator applies this tool on a daily or weekly basis in order to optimize the bikes redistribution by the identification of the bikes that have been substituted the most. In any case, the tool should be implemented whenever the demand is expected to undergo a significant change, for instance in the summer or during particular events. Our algorithm would provide interesting insights for before-after evaluation of transport policies, scheme changes (e.g. the installation of a new bike terminal), or land use variations like, for instance, new developments or the opening/closure of schools.

The tool is able to detect a set of travel behaviors whose characterization is of course influenced by both the available data and the general understanding of travel behavior and cycling. However, we think that the accurateness of the classification provided by the algorithm permits to design both strategic and operational measures for the sharing scheme. Other source of information such as GPS data could be used to gain deeper understanding of the described usage behaviors.

Finally, the data mining technique could be further applied to different and larger schemes with two purposes: to validate the power of the methodology in all types of BSSs and to compare the results with the implications derived from the classification of BSSs provided by O'Brien et al. (2014) in terms of the type of use: commuting, leisure or touristic. Nevertheless, further algorithms could be developed with the same smartcard transactions in order to classify users and enrich the characterisation of the bikeshare demand while becoming less dependent on traditional survey methods.

Acknowledgements

The authors would like to acknowledge the financial support provided by the Spanish Ministry of Economía y Competitividad in the projects TRA2010-18068 and TRA2012-39466-C02-02,

and the grant BES-2013-066347. Furthermore, it is the authors' desire to thank the City Council of Santander and JCDecaux for providing the data that has allowed validating this research.

References

- Akar, G., & Clifton, K. J. (2010). Influence of Individual Perceptions and Bicycle Infrastructure on Decision to Bike. *Transportation Research Record: Journal of the Transportation Research Board*.
- Alonso, B. (2010). *Bus stop location model considering the operation and incidents on the network*. Ph.D. thesis, University of Cantabria.
- Bordagaray, M., Dell'Olio, L., Ibeas, Á., Barreda, R., & Alonso, B. (2015). Modelling the Service Quality of Public Bicycle Schemes Considering User Heterogeneity. *International Journal of Sustainable Transportation*, 9(8), 580–591.
- Bordagaray, M., Fonzone, A., dell'Olio, L., & Ibeas, A. (2014). Considerations about the analysis of ITS data of bicycle sharing systems. *Procedia - Social and Behavioral Sciences*, 162(Panam), 340–349.
- Borgnat, P., & Abry, P. (2011). Shared bicycles in a city: A signal processing and data analysis perspective. *Advances in Complex Systems*, 14(3), 415–438.
- Cortés, C. E., Gibson, J., Gschwender, A., Munizaga, M., & Zúñiga, M. (2011). Commercial bus speed diagnosis based on GPS-monitored data. *Transportation Research Part C: Emerging Technologies*, 19(4), 695–707.
- Dell'Olio, L., Ibeas, A., Bordagaray, M., & Ortúzar, J. D. D. (2011). Modeling the Effects of Pro Bicycle Infrastructure and Policies Toward Sustainable Urban Mobility. *Journal of Urban Planning and Development*, 140(2), 1–8.
- dell'Olio, L., Ibeas, A., & Moura, J. L. (2011). Implementing bike-sharing systems. *Proceedings of the ICE - Municipal Engineer*, 164(2), 89–101.
- DeMaio, P. (2009). Bike-sharing: History, impacts, models of provision, and future. *Journal of Public Transportation*, 12(4), 41–56.
- Dill, J. (2009). Bicycling for transportation and health: the role of infrastructure. *Journal of Public Health Policy*, 30, S95–S110.
- Dill, J., & Carr, T. (2003). Bicycle Commuting and Facilities in Major U.S. Cities: If You Build Them, Commuters Will Use Them. *Transportation Research Record*, 1828, 116–123. <http://doi.org/10.3141/1828-14>
- Dill, J., & Gliebe, J. (2008). *Understanding and measuring bicycling behavior: A focus on travel time and route choice*. Portland, OR.
- Efthymiou, D., Antoniou, C., & Waddell, P. (2013). Factors affecting the adoption of vehicle sharing systems by young drivers. *Transport Policy*, 29, 64–73.
- Faghih-Imani, A., Eluru, N., El-Geneidy, A. M., Rabbat, M., & Haq, U. (2014). How land-use and urban form impact bicycle flows: evidence from the bicycle-sharing system (BIXI) in Montreal. *Journal of Transport Geography*, 41, 306–314.
- Fishman, E., Washington, S., & Haworth, N. (2013). Bike Share: A Synthesis of the Literature. *Transport Reviews*, 33(2), 148–165.
- Froehlich, J., Neumann, J., & Oliver, N. (2008). Measuring the pulse of the city through shared bicycle programs. In *Proceedings of the International Workshop on Urban, Community,*

and Social Applications of Networked Sensing Systems, UrbanSense08.

- García-Palomares, J. C., Gutiérrez, J., & Latorre, M. (2012). Optimizing the location of stations in bike-sharing programs: A GIS approach. *Applied Geography, 35*(1-2), 235–246.
- González, F., Melo-Riquelme, C., & Grange, L. De. (2015). A combined destination and route choice model for a bicycle sharing system. *Transportation*. <http://doi.org/10.1007/s11116-015-9581-6>
- Gordon, J. B. (2012). *Intermodal Passenger Flows on London's Public Transport Network*. Massachusetts Institute of Technology.
- Heinen, E., van Wee, B., & Maat, K. (2010). Commuting by Bicycle: An Overview of the Literature. *Transport Reviews, 30*(1), 59–96.
- Jäppinen, S., Toivonen, T., & Salonen, M. (2013). Modelling the potential effect of shared bicycles on public transport travel times in Greater Helsinki: An open data approach. *Applied Geography, 43*, 13–24.
- Jensen, P., Rouquier, J.-B., Ovtracht, N., & Robardet, C. (2010). Characterizing the speed and paths of shared bicycle use in Lyon. *Transportation Research Part D: Transport and Environment, 15*(8), 522–524.
- Kaltenbrunner, A., Meza, R., Grivolla, J., Codina, J., & Banchs, R. (2010). Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system. *Pervasive and Mobile Computing, 6*(4), 455–466.
- Lathia, N., Ahmed, S., & Capra, L. (2012). Measuring the impact of opening the London shared bicycle scheme to casual users. *Transportation Research Part C: Emerging Technologies, 22*, 88–102.
- Maldonado-Hinarejos, R., Sivakumar, A., & Polak, J. W. (2014). Exploring the role of individual attitudes and perceptions in predicting the demand for cycling: a hybrid choice modelling approach. *Transportation, 41*(6), 1287–1304.
- Martens, K. (2004). The bicycle as a feeding mode: Experiences from three European countries. *Transportation Research Part D: Transport and Environment, 9*(4), 281–294.
- Martens, K. (2007). Promoting bike-and-ride: The Dutch experience. *Transportation Research Part A: Policy and Practice, 41*(4), 326–338.
- Midgley, P. (2011). *Bicycle-sharing schemes: Enhancing sustainable mobility in urban areas*. New York.
- Munizaga, M. a., & Palma, C. (2012). Estimation of a disaggregate multimodal public transport Origin–Destination matrix from passive smartcard data from Santiago, Chile. *Transportation Research Part C: Emerging Technologies, 24*, 9–18.
- O'Brien, O., Cheshire, J., & Batty, M. (2014). Mining bicycle sharing data for generating insights into sustainable transport systems. *Journal of Transport Geography, 34*, 262–273.
- Pucher, J., & Buehler, R. (2008). Making Cycling Irresistible : Lessons from the Netherlands, Denmark and Germany. *Transport Reviews, 28*(4), 495–528.
- Pucher, J., & Buehler, R. (2009). Integrating Bicycling and Public Transport in North America. *Journal of Public Transportation, 12*(3), 79–104.
- Pucher, J., Dill, J., & Handy, S. (2010). Infrastructure, programs, and policies to increase bicycling: an international review. *Preventive Medicine, 50*(Supplement), S106–25.
- Rietveld, P. (2000a). Non-motorised modes in transport systems: a multimodal chain

- perspective for The Netherlands. *Transportation Research Part D*, 5(1), 31–36.
- Rietveld, P. (2000b). The accessibility of railway stations : the role of the bicycle in The Netherlands. *Transportation Research Part D*, 5(1), 71–75.
- Shaheen, S. a., Zhang, H., Martin, E., & Guzman, S. (2011). China's Hangzhou Public Bicycle. *Transportation Research Record: Journal of the Transportation Research Board*, 2247, 33–41.
- Vogel, P., Greiser, T., & Mattfeld, D. C. (2011). Understanding Bike-Sharing Systems using Data Mining: Exploring Activity Patterns. *Procedia - Social and Behavioral Sciences*, 20, 514–523.
- Wang, W., Attanucci, J. P., & Wilson, N. H. M. (2011). Bus Passenger Origin-Destination Estimation and Related Analyzes Using Automated Data Collection Systems. *Journal of Public Transportation*, 14(4), 131–150.
- Wilson, N. H. M., Zhao, J., & Rahbee, A. (2009). The potential impact of automated data collection systems on urban public transport planning. In N. H. M. Wilson & A. Nuzzolo (Eds.), *Schedule-Based Modeling of Transportation Networks* (Vol. 46, pp. 75–99). New York: Springer.