UNIVERSIDAD DE CANTABRIA

E.T.S. Ingenieros de Caminos, Canales y Puertos

PhD Thesis:

METHODOLOGICAL APPROACH
FOR ASSESSING THE DEMAND AND QUALITY OF BIKE-SHARING SYSTEMS
USING MANUAL AND AUTOMATIC DATA

Tesis Doctoral:

METODOLOGÍA PARA EL ESTUDIO DE LA DEMANDA Y LA CALIDAD DE LOS SISTEMAS PÚBLICOS DE BICICLETA MEDIANTE DATOS MANUALES Y AUTOMÁTICOS

Author:

María Bordagaray Azpiazu

Supervised by:

José Luis Moura
Luigi dell'Olio

Santander (Spain) - December, 2015

A Amatxu,

a Aitatxu,

y a Iñigo

"Caminar es un peligro y respirar es una hazaña en las grandes ciudades del mundo al revés."

"Walking is risky and breathing a challenge in the great cities of the looking-glass world."

EDUARDO GALEANO (1998)

Patas arriba: la escuela del mundo al revés

Upside Down: A Primer for the Looking-Glass World

Acknowledgements

Agradecimientos personales

No puedo dejar pasar esta oportunidad de agradecer a quienes más debo, a pesar de que, en muchos casos, su aportación va más allá de esta etapa de formación. Son esas personas que dan sentido y felicidad a mi vida: mi familia y mis amigos. Deseo agradecerles su interés por mí durante estos años. En especial, dedico mi agradecimiento a Daniela y Juan, quienes desde la mayor generosidad se ofrecieron a compartir lo que fue una inolvidable experiencia para mí, pero sabiendo que también era la última etapa de esta tesis. ¡Gracias por vuestro apoyo, empatía y por hacerme sentir como en casa!

Mi agradecimiento final, y el más especial, se dirige a mi madre, a mi padre y a mi hermano. Eternamente gracias por vuestro amor y cariño, y por vuestras enseñanzas: sois mis verdaderos maestros.

Abstract

A service is due to its demand and planning and evaluation are both necessary for its success. Indeed, even though numerous bike-sharing systems (BSSs) have been launched in the last years, many have also been closed. The insufficient mass demand has been claimed caused by suboptimal localization and density of the network, deficient operating hours or inefficient marketing among other reasons. Nevertheless, these could either be avoided or overcome with data, modelling, and analyses.

In addition, there is an actual need for new methods to model the demand for cycling since classic models and traditional assumptions do not hold for this transport alternative due to its particularities. On the other hand, it should be considered that the demand for cycling may be either derived from any other activity, or a leisure activity itself, implying a different behaviour. In any case, the fact is that cities experience increasing cycling rates and, consequently, it is necessary to study this evolution to redesign the bike-sharing schemes in order to meet the society's requirements.

This thesis aims to tackle the previous issues by proposing a thorough methodology to assess the quality and demand for bikeshare. Derived from this main objective, such a methodology is particularly intended to determine how the quality of the service is perceived, how the system is actually used and how its demand is influenced by the weather and also time-based variables. These specific research questions have been responded for the case of Santander (Spain), serving as a test to validate the approaches that can analogously be applied to any automatically operated system.

The first of the research questions is dealt with random ordered probit models that estimate the overall quality of the service as a function of the perceived quality of its attributes, and socio-economic and trip variables. For this purpose, the citizen involvement is proposed in the first stage of the research to provide the necessary context from the users' perspective, and to identify the variables that influence the demand for the system

and the perceived service quality. In addition, the heterogeneity in perceptions is also considered in different forms, confirming the influence of the gender, age, purpose of the trip, travel time, accessibility to the system and type of subscription.

After modelling the users' perceived quality, the second question is responded through data mining techniques applied to smartcard transactions and developed thanked to the knowledge obtained in the first stage from manually collected responses. A difference is made between usage and travel behaviour: the usage is described by the trip-chaining contained in the automatic footprints and is directly influenced by the limitations of the BSS as a public renting service, whilst the travel behaviour relates to the spatio-temporal distribution, the travel time and trip purpose. The approach is based on the hypothesis that there are systematic usage types which can be described through a set of conditions that permit classifying the rentals and reduce the heterogeneity in travel patterns. The two sequential data mining processes proposed permit to clean the data set, complete the characterisation of the demand for the system, support the study on the quality perceived by users, and optimise the operator's performance as well as the decision making.

The third specific aim is addressed through regression models directed to determine the impact of the weather and time-based variables on the trip production at terminals predominantly demanded for leisure and those experimenting a predominant demand for the BSS for transport purposes. This last stage of the research is built upon the relation between trip purpose, travel time and spatial distribution of the demand evidenced in the previous two stages. Interesting insights are uncovered that contribute to the knowledge on the influence of weather and time-based trip conditions on the demand for bikeshare. In particular, the models confirm the weather inertia: the rainfall not only affects the trip production in the instant when it is occurring but its impact is also significant in the following hour.

The results and conclusions drawn from the application of the proposed methodology provide the knowledge to design efficient and sustainable strategic, tactical and operational policies. The models and processes yield several indicators that are valuable to develop incentives to control the operator's performance and quality standards as well as to promote the pro-bike culture. Therefore, the approach is beneficial to the operator and to decision makers, but eventually, it is the entire society that would benefit from this tool as a result of the optimisation of the resources dedicated to the bike-sharing scheme.

Resumen

La demanda de un servicio justifica su oferta y su éxito depende de su planificación y evaluación. De hecho, a pesar de que numerosos sistemas de bicicleta pública se han puesto en marcha en los últimos años, también muchos se han cerrado. Las principales causas suelen ser la subóptima localización de terminales y de densidad de la red, insuficientes horas de funcionamiento o marketing ineficiente, entre otras razones. Sin embargo, estos problemas podrían evitarse a partir de información, su análisis y modelado.

Adicionalmente, se hace necesario desarrollar nuevos métodos para modelar la demanda ciclista ya que los modelos e hipótesis clásicas no siempre se ajustan a esta alternativa de transporte debido a sus particularidades. Por otro lado, es importante considerar que la demanda ciclista puede ser tanto derivada de cualquier otra actividad, o una actividad de ocio en sí misma, lo que implica un comportamiento diferente. En cualquier caso, lo cierto es que las ciudades experimentan un aumento de la movilidad ciclista y, en consecuencia, es necesario estudiar esta evolución para rediseñar los sistemas de bicicletas públicas con el fin de satisfacer las necesidades de la sociedad.

Esta tesis tiene como objetivo hacer frente a las cuestiones anteriores, proponiendo una metodología completa para evaluar la calidad y la demanda de bicicletas de préstamo. Derivado de este objetivo principal, dicha metodología se dirige particularmente a determinar cómo se percibe la calidad del servicio, cómo se utiliza el sistema y el impacto del clima y variables temporales en su demanda. Estos objetivos particulares de la investigación han sido desarrollados para el caso de Santander (España), que sirve como caso de estudio para validar la metodología que de forma análoga se puede aplicar a cualquier sistema de préstamo automático.

El primero de los objetivos específicos ha sido abordado por medio de modelos *random ordered probit* que estiman la calidad global del servicio en función de la calidad percibida de sus atributos, y variables socio-económicas y condiciones de viaje. Para

ello se propone la participación ciudadana en la primera etapa de la investigación, para proporcionar el contexto necesario desde la perspectiva de los usuarios, y para identificar las variables que influyen en la demanda del sistema y la calidad percibida del servicio. Además, la heterogeneidad en las percepciones también se ha considerado en diferentes formas, confirmando la influencia del sexo, edad, propósito del viaje, el tiempo de viaje, la accesibilidad al sistema y el tipo de suscripción.

Después de modelar la calidad percibida por los usuarios, para abordar el segundo objetivo se proponen técnicas de minería de datos aplicadas a las transacciones de acceso al sistema mediante tarjeta canceladora. Esto es posible debido a los conocimientos obtenidos en la primera etapa de la tesis a partir de los datos obtenidos de los usuarios mediante encuestas. En esta etapa se establece una diferencia entre el uso del sistema y los patrones de viaje: el uso del servicio se describe a partir de la secuencia de usos del servicio contenida en la información generada automáticamente por el sistema y está directamente influenciado por las limitaciones del sistema como servicio de alquiler público, mientras que los patrones de viaje se relacionan con la distribución espacio-temporal, el tiempo de viaje y el propósito del viaje. El método se basa en la hipótesis de que hay prácticas de uso que pueden describirse a través de un conjunto de condiciones que permiten la clasificación de los alquileres, reduciendo la heterogeneidad en los patrones de viaje. Los dos procesos de minería de datos secuenciales que se proponen permiten depurar la base de datos, completar la caracterización de la demanda del sistema, enriquecer el estudio sobre la calidad percibida por los usuarios, y optimizar el desempeño del operador, así como la toma de decisiones.

El tercer objetivo específico se aborda a través de modelos de regresión que estiman el impacto de la meteorología y de variables temporales en la generación de viajes en las terminales predominantemente demandadas por ocio y aquellas que experimentan una demanda predominante con fines de transporte. Esta última etapa de la investigación se sustenta sobre la relación entre el propósito del viaje, el tiempo de viaje y la distribución espacial de la demanda que se evidencia en las dos etapas anteriores. Los resultados contribuyen al conocimiento sobre la influencia del clima y las condiciones temporales en la demanda de bicicletas públicas. En particular, los modelos confirman la inercia meteorológica: la precipitación no sólo afecta a la generación de viajes en el instante en que se está produciendo, sino que su impacto es también significativo en la hora siguiente.

Los resultados y las conclusiones extraídas de la aplicación de la metodología propuesta proporcionan los conocimientos necesarios para diseñar políticas estratégicas, tácticas y operativas eficientes y sostenibles. Los modelos y procesos producen diversos indicadores que son valiosos para el desarrollo de incentivos destinados a controlar el desempeño del operador y la calidad de la oferta, así como para promover la cultura pro-bicicleta. Por lo tanto, el enfoque es beneficioso para el operador y para los organismos decisores, pero eventualmente, toda la sociedad se verá beneficiada tras la aplicación de esta metodología y como resultado de la optimización de los recursos dedicados al sistema de reparto de bicicletas.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Any service is due to its demand. The dictionary defines the term *demand* as *the desire or need of customers for goods or services which they want to buy or use*. Thus, the demand is both the reason for the provision of a service and the source for revenues. Having said that, there is no doubt that it is of utmost importance that services are carefully designed and evaluated in order to determine to what extent resources are efficiently fulfilling their commitment and the supply is adjusted to the actual demand.

Specifically, public transport services are due to the citizenship and the success of a transport facility or service will necessarily come from an in-depth study of its demand. But in order to accomplish such a task, data is required. Data has historically been an issue in the transportation field of research and practice. However, in the last decades there has been a surprising development of information and communication systems which are the source of enormous amount of detailed objective data. This circumstance is believed to be the key to more efficient models and predictions, although the social dimension of the demand for transit services still requires traditional data collection via personal surveys.

The data permit analysing, modelling, simulating and predicting. In terms of a transport system, this means the service can be optimised so as the gap is minimized between demand and supply. Such an objective should be a priority since transport systems are generally non self-sustained and resources should not be wasted. Therefore, the public administration would benefit from the insights brought from the modelling and analyses of a transit system as long as it guides the decision making towards an efficient service.

As a transit mode, all the above is applicable to a bike-sharing system (BSS). These

schemes have been one of the most implemented actions to encourage everyday cycling for transport. Local governments conceive public bicycles as an alternative form of public transportation and thus, any trip that the bike is chosen instead of any motorized mode, and more precisely, instead of the private vehicle, is a success on the way towards a more liveable and sustainable city. But such benefits could only be ensured when the demand is deeply analysed to understand the cyclists' behaviour and citizens' transport requirements.

The supply of a BSS differs from that of traditional transit modes in an essential characteristic: it is an individual transit mode which lacks of a calendar and timetable and thus, the bikes are usually available to rent at determined pick-up and drop-off points 24 hours a day, 7 days a week. Furthermore, the particularities of this active mode leads to an *a priori* heterogeneous demand due to cycling being both conceived as a mode of transport and as a recreational activity or physical exercise. Therefore, it could be expected that these different conceptions are associated to substantial variability in the demand and travel patterns. On the contrary, there is a similarity of BSSs with conventional transit systems: the concession framework providing the service influence the operator's performance as a result of the dissimilar objectives of the public administration and those of the operator providing the service.

Due to the previous circumstances, there is an actual need to develop methodologies to specifically analyse and model the demand for bike-sharing since traditional methods applied to the conventionally regulated transit modes are not sufficient to reproduce the particularities of bicyclists' behaviour accurately. As a contribution to fill such a gap, this thesis proposes a novel methodology that provides the ingredients for the system to be optimised by assessing the quality and demand for bikeshare schemes. Derived from this main objective, the models and data mining processes evaluate how the quality of the service is perceived, how the system is actually used and how its demand is influenced by the weather and also time-based variables. These specific research questions have been responded for the case of Santander (Spain), serving as a test to validate the approaches that can analogously be applied to any automatically operated system.

The results uncover diverse aspects that need to be improved such as the fare structure, the subscription framework, the rental conditions, or the location of terminals since these are the factors that directly condition the demand. Therefore, all stakeholders will benefit from the application of the methodology: on the one hand, from the operator's per-

spective, the insights derived from the methodology permit to optimise the performance in terms of operational tasks such as bike redistribution while providing higher quality standards; on the other hand, the public administrator can anticipate the impact of any strategic measure towards the promotion of the BSS and can also design efficient incentives to stimulate the operator's commitment. Eventually, it is the society the principal beneficiary and the reason behind the need to evaluate and improve BSSs as a transit mode.

In terms of the scientific contribution, this thesis describes an integral methodology to assess the demand and quality of bike-sharing systems. This research has developed techniques to collect and use traditional survey data and to apply and manage automatically collected data. In addition to the methodological contribution, this thesis is also a step forward in the knowledge on the demand for bike-sharing and the cyclists' requirements and behaviour.

The remainder of this chapter presents the motivation of this thesis, its objectives and its contributions. Subsequently, the case study selected to validate the various methodological approaches is described and justified. Finally, the structure of the thesis is presented.

1.1 Motivation

The demand and supply interact within a transport system: the demand adapts its behaviour to the supply and, at the same time, the reason for providing the supply is actually to respond to travellers' transport needs. According to this statement, suppliers should constantly evaluate to what extent the citizens find the service adequate to their transport requirements. Otherwise, the service will not be attractive enough to justify its provision. Moved by this believe, this thesis aims to establish an integral methodology directed to evaluate the demand for bikeshare by responding to the following questions:

- What do citizens demand from a bike-sharing system?

- How do users perceive the quality of the service?

- How do subscribers actually use the service?

- Is the users' behaviour sustainable in terms of capacity?

- To what extent do diverse factors such as the land use or the weather affect the demand for bikeshare?

The previous questions provide a picture of the scheme. Needless to say, they are interrelated: depending on how the quality of the service meets the travellers' expectations, the use of the system will vary. Furthermore, only by responding to these questions on the quality and demand-supply interaction of a BSS could it be possible to plan long-term strategies and decide on the resources allocation.

The society but also the bike-sharing service provider and the public administrator can benefit from the methodology presented in this thesis. Insights are obtained to guide the decision making towards an optimised bike-sharing scheme, promoting healthier and more sustainable urban lifestyles and reducing congestion, and noise and air pollution. Furthermore, the scientific community and transport practitioners are provided with tools to evaluate the demand and identify the efficiency in the provision of such a service.

Additionally, two aspects are a motivation *per se* in this thesis: the application of this research to bike-sharing schemes and the use of two types of data in terms of their nature, this is, automatically and manually collected information. These two aspects are actual interests nowadays and it is thereby the time to provide contributions to both lines of research and practice.

Bike-sharing schemes

The sharing market is increasing its presence; nowadays, sharing systems provide cars, offices or bikes to rent. However, although the sharing essence is just the same for bikes than it is for books in libraries, the truth is that the first are constantly under debate in the last years.

An interesting review of the literature based on BSSs is provided by García-Palomares *et al.* (2012) and Fishman *et al.* (2013). Other recommended readings are DeMaio (2009) and Midgley (2011), who reviewed the fundamentals of BSS based on worldwide schemes and in relation to the infrastructure and access to the service, as well as the business-management model.

The third generation of BSSs started in the late 90's and has spread the most; it is characterized by automatic 24-hour provision of bikes at specific points across the city, where the access is on a self-service basis requiring subscription and use of a smartcard.

Every time a bike is picked up or returned to the system, the transaction is recorded, thereby providing large amounts of automatic and rich data on bike mobility. This source of information is increasingly being applied for a variety of purposes in the transportation field of research(Froehlich *et al.*, 2008).

In 1995, Copenhagen installed a bike-sharing system and the first automatic program was launched in Rennes (France) in 1998 (Midgley, 2011). From that point on, the initiative has spread all over the world, reaching a great number of cities of all sizes. In Europe, the fourth generation of BSSs (DeMaio, 2009) is currently spreading: Stuttgart, Prague and Donostia-San Sebastián offer electric bikes, and Madrid has launched its fully electric system in 2014.

The majority of programs are open to both registered users and casual demand. In this regard, Bicing, the system in Barcelona is particular since it is merely offered to residents. Its use is widely extended for everyday trips but also known to reach capacity at peak hours and to experience intense asymmetries caused by the steep slopes. Interestingly, Bicing now offers electric bikes to face the later issue.

The international picture shows these initiatives are very differently conceived both by decision makers and by the public. Indeed, there is a trend that supports BSSs as an alternative transit mode, whilst another believes that these systems should facilitate a first cycling experience to be later substituted by a private bike once it is being routinely demanded. In this regard, maybe moved by a wish of more flexibility, maybe the will to avoid coming across 100% full or empty terminals, some people end up buying their own bicycle after testing bike-sharing.

Therefore, it is interesting to determine where the demand for cycling comes from and the conditions under which the uptake is originated. For instance, Efthymiou *et al.* (2013) estimated ordered logit models to assess the willingness to join car and bike-sharing schemes, finding that bike-sharing is attractive to people that currently go on foot. The authors also inform on the increase in the use of public transport and especially bikes in Greece since the economic crisis began. The uptake of bike-sharing was also studied for the case of Hangzhou's (China) BSS by Shaheen *et al.* (2011). A survey was conducted to both members and non-members, finding that the demand for the BSS is taken from a wide range of modes, including private cars and taxis, but also walking and transit. Another interesting insight that should be remarked upon from the system in Hangzhou

relates to the actual members having higher access to private cars than non-members.

A BSS infrastructure successfully integrated within the public transport network implies that even when distances are high, the bike can get the traveller to zones not covered by other conventional public transport (PT) modes, leading to the possibility of multi-stage trips and providing an alternative to the private car. For instance, in Paris, the number of multimodal trips with use of public bike increased after the launch of Vélib' system (DeMaio, 2009) and the synergy bike-train has also been assessed (Martens, 2004; Martens, 2007; Rietveld, 2000a; Rietveld, 2000b).

The occasional studies that directly or indirectly focus on the impact of BSS on the citizens' mobility patterns conclude the difficulty to achieve an actual shift from private car to bike-share and even in such a case, the change is suggested scarce (DeMaio, 2009; Fishman *et al.*, 2013; Midgley, 2011). However, there are always examples that serve to encourage designing bike-friendly policies. Such is the case of Lyon, since Jensen *et al.* (2010) examined the daily flows and routes and found double demand for the BSS when the other public modes are on strike.

In order to promote the modal shift and more sustainable trip patterns, it is necessary to first understand the demand for cycling. Specifically, in this thesis the demand for bike-sharing is assessed.

Automatic versus manual data

Handling and managing data automatically collected by Intelligent Transport Systems (ITS) is currently a major opportunity but also a challenge for transport professionals due to the great amount of objective data generated. With these data, methods of analysis have been developed to understand the demand for public transport. These are mainly based on trip-chaining assumptions imposed to the automatic footprints that ITS record. Some interesting contributions in this regard were authored by Cortés *et al.* (2011), Gordon (2012), Munizaga and Palma (2012), Wang *et al.* (2011) and Wilson *et al.* (2009). The initials of this niche of research focused on deriving origin and destinations from sequential smartcard uses. As resulted in the previously mentioned international research, most of the trips are successfully inferred through data mining techniques and this way, the travel patterns can be studied and the behaviour modelled.

This thesis guides the management of smartcard data from public bikes by providing

criteria to infer the trip production and travel patterns described by the actual behaviour registered through bike pick-ups and drop-offs, some information only available from automatically operated bike-sharing systems. Therefore, this work has been motivated by the need for studying and improving the potential of bike-share smartcard system data for travel demand analysis.

Nevertheless, there is valuable data that could only be directly obtained by personal interviews. This research proposes a survey design and traditional sampling method to collect subjective measurements such as service quality perceptions. In any case, indirect indicators of quality are also obtained in this thesis derived from the data mining applied to smartcard transactions.

1.2 Objectives

The main objective of the present thesis is to provide a thorough methodology to assess and analyse the quality and demand for a bike-sharing system in order to identify the actions to be taken to optimise the system.

Particularly, the present thesis is intended to:

- Determine the service attributes that influence the most in the users' perceived quality. The first model in this thesis identifies the service characteristics that users prioritise when rating the service quality and thus those that should perform the highest quality standards.

- Identify the conditions in the usage of the BSS that introduce systematic heterogeneity in travel patterns. This is achieved by the application of automatically recorded data that inform on the actual behaviour. These data undergo various data mining processes that classify the rentals into groups that show increased homogeneity in terms of travel patterns.

- Determine the deficiencies in the system and the gap between supply and demand so as to be improved by the public administrator and/or the operator.

- Provide new applications for the still under-utilized ITS data, taking advantage of the potential of objective measurements, whilst manually collected data is also applied to assess subjective measurements.

- Evaluate the different use of registered and casual users in terms of perceived quality and travel patterns with the purpose to provide further knowledge on the different behaviour when cycling is a derived demand, this is, a tool for transportation, compared to when it is the an activity itself such as a leisure ride or physical exercise.

- Assess the different impact of the weather and time-based variables on the trip production at terminals predominantly demanded for leisure and those experimenting a predominant demand for the BSS for transport purposes.

- Provide indicators that could be monitored in terms of incentives to the operator and to promote cycling as a transport alternative while ensuring a sustainable demand for the BSS.

1.3   Contributions

Methodological contribution

Methodologically, this thesis contributes to the state of the art with a series of analyses, algorithms and models to assess the quality and demand for a BSS.

Particularly, the research is a step forward in various aspects:

- the heterogeneity explaining quality perceptions and travel behaviour is considered in diverse forms throughout the research;

- a guidance is provided on the management and processing of conventional survey data and automatically collected data sets;

- novel data mining processes have been developed that detect deficiencies in the service and that serve as a clustering technique to take advantage of objective data to infer travel patterns;

- count data models are proposed as an alternative to multiple linear regression to estimate trip production;

- the non-linear effect of weather variables on the demand is tested.

Contributions to the knowledge

The outcomes obtained from the methodology provide knowledge on:

- the attributes of a BSS that users place the most importance on when rating the quality of the service,

- the factors that induce variability in perceptions,

- the conditions on the usage behaviour that introduce heterogeneity in travel patterns,

- the usage framework of automatically operated BSS and how it differs based on the type of subscription,

- the effect of the weather and time-based attributes on bike trip-production

Published work from this thesis

Several articles and conference proceedings have been published that disseminate the contents of this thesis:

- Bordagaray M., dell'Olio L., Ibeas, A., Barreda M.R., & Alonso B. (2013). Modelling the service quality of public bicycle schemes considering user heterogeneity. *International Journal of Sustainable Transport, 9*(8), 580-591.

- Bordagaray M., Dell'Olio L., Fonzone, A., & Ibeas Á. Capturing the conditions that introduce systematic variation in bike-sharing travel behaviour using data mining techniques. Transportation Research Part C: Emerging Technologies. *(Under review)*

- Bordagaray M., Fonzone A., Dell'Olio L., & Ibeas Á. (2014) Considerations about the analysis of ITS data of bicycle sharing systems. *Elsevier Procedia - Social and Behavioral Sciences*, 162, 340 – 349.

- Bordagaray M., Ibeas A., & dell'Olio L. (2012). Modeling user perception of public bicycle services. *Elsevier Procedia - Social and Behavioral Sciences*, 54, 1308-1316.

- Bordagaray M. (2014) Metodología de minería de datos para identificar patrones de demanda de los sistemas públicos de préstamo de bicicletas. *XI Congreso de*

Ingeniería del Transporte. Santander (España). Conference Proceedings. (ISBN 978-84-697-0359-5)

- Bordagaray M., Fonzone A., Dell'Olio L., & Ibeas Á. (2014) Considerations about the analysis of ITS data of bicycle sharing systems. *XIII Congreso Panamericano de Ingeniería de Tránsito, Transporte y Logística.* Santander (España). Conference Proceedings. (ISBN 978-84-617-0085-1)

1.4   Application to Santander (Spain)

Santander is the capital of the Autonomous Community of Cantabria, one of 17 in Spain and located on the north coast. It is a medium-sized city, covering 36 m^2 with a population of about 200,000.

The city is particularly characteristic for its environment, topography and meteorology. It is a port city and its limits give shape to the Bay of Santander. This provides beautiful sights of the naturally generated forms of the coast line, beaches and towns in the surroundings.

Several characteristics of the city make it an ideal case study to test and validate the methodology proposed in this thesis:

- The BSS attracts not only residents but also tourists. Santander is characterised by beautiful views, open areas and an impressive coastline, and these attractions are equally valuable for tourists and residents. Therefore, tourism and local daily activities coexist in the city and this is expected to be reflected on the travel behaviour.

- As stated in the handbook delivered within the OBIS project (2011) funded under the Intelligent Energy Europe Programme (IEE), "*topography and climate are significant for how and when people find it agreeable enough to use the cycling mode.*" This case study permits to provide further knowledge on the conditions that these attributes should perform so that cyclists are inclined to choose the bike-sharing service as a mode of transport or for recreational purposes.

- The city was one of the leading smart-cities in Spain. Indeed, an extensive network of sensors was displayed to gather data on diverse aspects such as air conditions, noise or available street-parking lots. As a result, the city's commitment towards the

development of tools to apply automatically collected data has also been a motivation in this thesis.

Further information is presented of the city to provide the context to understand the demand for the bike-sharing scheme and the users' behaviour that will be analysed and modelled along this thesis.

Topography

North-south mobility is restricted because the steep slopes (greater than 15º) of parallel hills and valleys running north-east to south-west meant few of the important city routes were built in that direction.

Meteorology

The climate in Santander is oceanic. The average yearly temperature is 10ºC, oscillating from the average minimum of around 5ºC in winter to 25ºC in the summertime. The rainy weather is described by an average amount of 1200 mm of water fallen per year within around 130 precipitation days. Humidity is also high, reaching beyond 90%.

Public transport services: bus and bike-share

The public transport has historically been provided by a network of bus lines and bus stops located at less than 300 meters apart. Additionally, Santander is currently served by a public bicycle service provided by a fleet of 200 bicycles distributed between 14 recently installed docking stations and with further plans for expansion.

The BSS, named TusBic, was installed in 2009. The automated bike-sharing system improved the conditions of the previous service, which was not operated automatically and was limited in opening times and number of docking stations. This encouraged the use of the bicycles for leisure purposes but was not seen enough as an alternative for transport. The upgrade to the automatically operated system, which was made via a concession to a private advertising firm, was intended to provide an alternative mode of transport.

As suggested in the literature, the infrastructure dedicated to cycling plays a key role in its promotion (Akar & Clifton, 2010; Dill, 2009;, Dill & Carr, 2003; Pucher *et al.*, 2010). Dell'Olio *et al.* (2011a) also identified the infrastructure to determine the uptake of cycling as a mode of transport in Santander. As determined by Faghih-Imani *et al.*
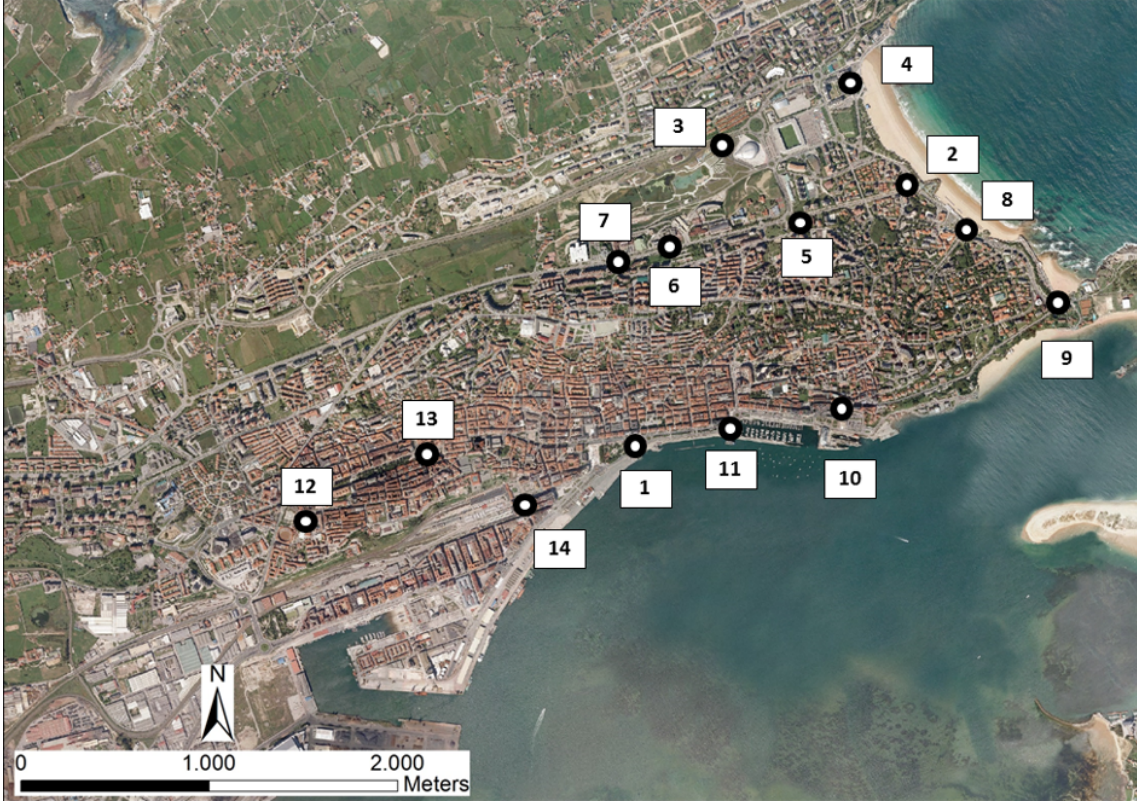
Figure 1.1: Localization of the terminals of TusBic system in Santander

(2014), the land use and facilities in the surroundings of the terminals also influence the demand for the BSS. In the present case, except for the ordinary bike parking facilities, which are spread throughout the entire city, the rest of the infrastructure dedicated to cycling avoids the steep slopes and is therefore mostly located along the perimeter of the city. The city's main attractions are served by the BSS and bike lanes (Figure 1.1 and Figure 1.2): the city center (terminals 1, 11 and 13), the train station and that for regional and national buses (terminal 14), the main touristic attractions such as the Park of La Magdalena (terminal 9), the Casino (terminal 8), the coastline (terminals 1, 2, 4, 8, 9, 10, 11), Las Llamas Park (terminal 3), and the University Campus (terminals 6 and 7). Terminals 5, 6, 7, 10, 12, 13, and 14 are located in mainly residential areas.

In 2011, the system consisted of 14 terminals, about 200 bikes and more than 300 docks. The BSS in Santander is offered to all public. Three types of subscription are available that give annual, a week or a day access to the system. Annual cards are worth €10, whereas week and day subscriptions cost €5 and €1 respectively. Furthermore, it should be mentioned that week and day subscribers (casual users) can register on the go

Terminal	Capacity (docks)	Bike lane	Beach	Panoramic view	Commercial area	City centre	University	Residential area	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				✓	✓		✓	✓

Figure 1.2: Description of the facilities and land use in the surroundings of the terminals

as long as a credit card is provided at a terminal to charge the fee and an amount of €150 is blocked as deposit. On the contrary, annual users should first register themselves online. The subscription covers trips lasting less than an hour. In order to discourage prolonged use, annual subscribers are charged €0.30 for every extra half an hour. Analogously, weekly subscribers are charged €0.50 and daily users €0.60.

The data applied in this thesis is limited to the summer period of 2011 and this is justified by the following facts:

- A previous research conducted in Santander assessed the negative influence of the weather on the mode choice towards the bike (dell'Olio *et al.*, 2011a). This explains that the presence of bikes sharply decreases during the rainiest months of the year, being the demand for the BSS in July and August a 30% of the yearly demand. As a consequence, summer is the time when the service can reach its capacity. In fact, this occurs to the bus service: the system experiences a peak of demand when both residents and tourists require the service.

- In the summer time, the variability in the demand is more accused since all types of trip purposes are expected due to the attraction of tourists and the chilly weather (which can also get rainy), at the same time that residents are still working but may also be on vacation. Therefore, travel patterns are expected to vary substantially across the population of bike rentals.

- Summer is when the weather conditions are the most favourable to use the bike for transport and it is thereby when the utilitarian uses of public bikes can be identified and characterised presenting a more stable and reliable routine.

Therefore, the demand and variability within the BSS are at their highest values during the summer months in Santander and thus, these circumstances make it an ideal period to test and validate the demand, considering all types of subscribers (casual and registered) and uses (recreational and utilitarian). Thereby, the traditional sampling method through surveys was conducted in July and August 2011, to collect the data to model the service quality, whilst the dataset comprising the smartcard transactions includes 26,290 records, recorded from the 1st of July to the 31st of August 2011. For each rental, the system records:

- Bicycle pick-up and drop-off docking station

- Identification number of the stand occupied by the bicycle at the origin station

- Time (date and time – hour, minute, second) of the bike pick-up and return

- User's type of subscription (annual, weekly or daily)

- Subscriber identification number

- Bicycle identification number

1.5 Structure of the thesis

The methodological contents of this thesis are structured in three main chapters ordered according to a sequence of phases designed to guarantee the accuracy of the results and conclusions.

The first stage of this research, and presented in Chapter 2, describes the data collection method and proposes an active involvement of the citizens in focus group sessions and interviews with the users of the bike-sharing scheme. Two main reasons justify the public involvement to be proposed in the beginning of the thesis: to provide the necessary context to the research from the users' perspective, and to identify the attributes, individual characteristics and trip conditions that influence the demand for the system and the perceived service quality.

After obtaining valuable information from the citizens and users, and from the demand and quality indicators obtained in Chapter 2, an alternative source of data is examined in Chapter 3: smartcard transactions automatically recorded with every use of the system. Thanked to the knowledge obtained in the first stage (Chapter 2), methods can be designed to obtain the most from automatic data. This way, and as a complement to the results on the quality perceived by users, a first data mining process has been developed that detects deficient bikes through the description of a usage pattern called "bike trial with substitution". The application of the method not only consists of a satisfaction indicator on the bike performance but also provides interesting data to support operational tasks such as bike redistribution across the system with rebalance and maintenance purposes. A subsequent data mining process is proposed in Chapter 3 to cluster the rentals into less heterogeneous groups of demand, after the bike trials detected in the previously applied algorithm have been removed from the data set. The clustering permits to characterise the demand for bike-sharing based on the objective usage behaviour contained in the automatic footprints. The analyses of the various usage behaviours yield various interesting insights: the characterisation of travel patterns in terms of travel time and spatial distribution across the system. The analyses to the objective automatic data proposed in Chapter 3 provide a detailed picture of the demand that enrich the information collected through traditional manual techniques applied in Chapter 2.

Chapter 4 presents a regression approach to determine the impact of the weather and time-based variables on the trip production at terminals predominantly demanded for leisure and those experimenting a predominant demand for the BSS for transport purposes. This last stage of the research is built upon two assumptions derived from the outcomes in the previous phases: first, that the travel time and the quality perceptions differ based on the trip purpose (Chapter 2), and second, that the demand for a terminal and the travel behaviour are associated to the land use and main activities in the surroundings of the bike station (Chapter 3). Interesting insights are uncovered that complement previous research on the influence of weather and time-based trip conditions on the demand for cycling.

Finally, Chapter 5 summarises the main insights obtained in this thesis and concludes with the future research that the methodological approach and the results obtained leave open.

Chapter 2

A methodology to assess the users' perceived quality based on manual data

2.1 Introduction

Providing quality in addition to a service itself has been a major marketing policy in developed countries in the last decades. As maintained by Parasuraman *et al.* (1988), delivering quality is a business strategy that adds value to the service and helps differentiate from competitors. In the transportation field, this research interest arose in the 90's but it was made official among practitioners at the moment when the Transit Capacity and Quality of Service Manual (TCQSM) was first published in 1999 to "provide guidance on transit capacity and quality of service issues and the factors influencing both", in a similar manner as the traditional Highway Capacity Manual (HCM, High Capacity Manual, 2010) provides assistance in the design of infrastructures.

But what exactly does *quality* refer to? Such a complex term demands to first look up the word in the dictionary for a first approximation of the concept. Two definitions are provided:

- the standard of something when it is compared to other things like it;

- how good or bad something is.

From the first statement, it can be said that quality is not an absolute term but a relative

measure. By the second definition it could be interpreted that quality is a measurement by itself, but in the end, such measurement is inherently related to the evaluator's own experiences or expectations. As a consequence to this, when addressed from the customers' perspective, the service quality may be greatly affected by personal and cultural circumstances due to the subjective nature of the problem.

This chapter of the thesis describes the sequential steps on the data collection and modelling intended to determine how the quality is perceived by users and to identify what aspects should be prioritised to enhance the service.

Although infrastructure availability has been demonstrated fundamental in the promotion of cycling (Akar & Clifton, 2010; Dill, 2009; Dill & Carr, 2003), underutilized BSSs have proven that this is not enough. Hence, the quality of bike-sharing systems should be examined in order to guarantee that the infrastructure meets its purpose and to assist in its improvement and resources reallocation. Nevertheless, although international contributions have been published focused on cycling and bike-sharing systems, the quality of these services has not been examined. This chapter is aimed at addressing this deficiency through the examination of the users' perceptions. For this to be accomplished, an active participation is required from the users, and their contributions guide the subsequent stages of the demand study. For this reason this was the first objective to be addressed in this thesis.

The ordered scale defined to mark the quality of the overall service and its attributes suitably holds the assumptions of the proposed modelling framework: Random Ordered Probit. The modelling outcomes reveal the aspects that should be prioritised to generate the highest impact on the users' satisfaction and, thus, to promote the use of the public bicycle service.

The inherent heterogeneity existent in the subjective valuations of BSS users has been dealt with through a pair of techniques: by allowing the parameters to distribute randomly across the population in order to account for the unobservable variability, and by the introduction of the conditions that cause systematic variations in the perception of quality such as personal, socio-economic and trip circumstances. The trustworthiness of the models resides in the principal role given to the users of the system from the very first initials of the study of the demand for the public bikes in the tasks concerning the data collection and the sampling design.

The chapter is structured as follows. Section 2.2 reviews the international scientific literature addressing the concept and assessment of quality, particularly those concerned on public transport services. Section 2.3 describes the research methodology proposed to examine the service quality of BSSs and presents the sampling method and the analysis of the data collected. In section 2.4, the modelling is explained: first, the theoretical framework is described; second, the estimated models are presented and the results are discussed. Finally, section 2.5 gathers the conclusions drawn from this research.

2.2   Previous research on transit quality

Perspectives on service quality

The literature on market research and particularly in the transportation field has demonstrated the key role that customers play in the evaluation of the service performance. The users' opinions should be taken into account since they are who decide and since they may eventually change their minds whenever the service does not reach the standards they expect. This fact has motivated the design of performance-based contracts in the transport sector (Hensher *et al.*, 2003; Hensher & Prioni, 2002; Mokonyama & Venter, 2013) and it has been a particular niche of research in the last years as a stimulant for quality provision.

The literature provides an extensive spectrum of tools for measuring the service quality whilst also a wide range of perspectives have been adopted with this aim. The specific objective of the study together with the conceptualization of the term *quality* determine the modelling, which eventually requires particular data to be collected through a specific sampling method. The service quality refers to the level of service with regard to the characteristics describing the service, both tangible and intangible. Thus, it is the service performance that is under evaluation when talking about service quality and the performance can be evaluated based on objective measures gathered by the operator or on the customers' perceptions and experience.

On the one hand, techniques have been developed that evaluate the service performance in terms of the efficiency or productivity based on a set of quantities that the firm collects for their analysis and monitoring. For instance, the Data Envelopment Analysis (DEA) is a method that establishes production benchmarks based on measurements of

the resources and performance, a method that is widely applied to logistics in the case of the transportation field (Cullinane & Wang, 2006). It should be noted that such perspective is uniquely aimed at maximizing the firm's benefits lacking the clients' perceptions and, thereby, omitting the customers' repurchase intentions as a result of their satisfaction. There is an alternative stream of practice that evaluates the service quality from the supplier's perspective based on the believe that the quality pertaining the service design, management and performance do have an effect on the customers' satisfaction and loyalty. In the transportation sector, this position is supported by the TCQSM, a manual that provides guidance to transit agencies in terms of the design and delivery of service standards. With this premise, the TCQSM defines intervals of values for a set of parameters in order to assess the level of service quality provided.

On the other hand, Parasuraman *et al.* (1985) propose considering both the service provider's and the customers' viewpoint through the measurement of the discrepancies between the willingness, expectations and objectives of transit agencies and passengers, and their perceptions, experiences and inputs. This combined perspective has also been considered in transport research by various authors. Rietveld (2005) warned on the overestimation of service quality when considering supply-oriented indicators. His research measured the systematic differences in the quality if compared based on objective measurements and on the users' perceptions. The author highlights the errors involved in quality studies that only consider the operator's viewpoint. As well as other authors in the literature (Friman *et al.*, 2001; Pedersen *et al.*, 2011), Rietveld (2005) argues on the important effect of negative incidents on the users' impression. As the author holds, since a greater number of passengers travel in peak periods, the majority of users perceive the quality provision under such circumstances, a point to be taken into account in the design of the data collection processes but also a point that is overlooked when only objective measures are used.

Nathanail (2008) used data delivered from the operator together with survey responses to evaluate the service performance through multicriteria analyses, whereas some years later, Eboli and Mazzulla (2011) proposed a unique index that simultaneously considers both objective and subjective measures of the quality of transit performance. Friman and Fellesson (2009) also focused on the link between service performance (vehicle km/inhabitant, total public transport place km/inhabitant, and average public transport

speed) and customers' perception of it (frequency, seat, travel time and overall satisfaction) in 6 European cities. They concluded it is not always true that more public transport results in more satisfaction, suggesting that the service performance is only a part of their satisfaction. Specifically, their research found no correlation between supply and frequency satisfaction and a negative correlation between average speed and satisfaction with the travel time, this outcome interpreted caused by the fact that commuters making longer journeys tend to mark travel time lower. The main conclusion that the reader gets from their research is the need to accurately identify the objective supply indicators that resemble the satisfaction attributes the best or vice versa. The research by Rietveld (2005) contributes in this direction with the provision of supply-oriented quality measures equivalent to demand-oriented ones.

Concepts involved in the assessment of service quality

As already mentioned in this thesis, there is an ongoing debate on what service quality actually is. Various authors have explicitly defined diverse concepts and have attempted to confirm the nature of such concepts through experiments. What do customers expect from the service? Parasuraman *et al.* (1991) raised the question whether there are different types of expectations, their stability over time and the factors influencing them. They established two levels of expectations: adequate service quality and desired service quality, leaving a "zone of tolerance" in between (Parasuraman *et al.*, 1994). Alternatively, as treated by Zeithaml *et al.* (1993) and dell'Olio *et al.* (2011b), expectations are related to the desired quality.

Satisfaction is usually referred to the level of fulfilment of expectations, this is, a relative measure. Satisfaction is, thus, a subjective measure relative to what it is expected or desired; the user is being asked about his or her feelings as a demander. For this reason, Parasuraman *et al.* (1994) proposed modelling the gap by considering perceptions and expectations. Alternatively, Caruana (2002) argued that the customers' expectations need not be collected; only a score of the level of fulfilment of their expectations is needed for each item.

In contrast to the relativity inherent in the definition of satisfaction, the term *perception* may be interpreted an absolute measurement of the actual quality. In this case, the user is an evaluator of the performance but once again, the individual rates based on

previous experiences, so, eventually, it is a relative value.

The previous definitions imply delicate differences between them which should be internalised before conducting the experiment and, particularly, before the design of the data collection process.

Proposed methodologies to understand the users' perceived quality and satisfaction

SERVQUAL (Parasuraman *et al.*, 1988) is probably the most widely known scale developed for measuring the service quality. Initially proposed to evaluate and manage the quality in bank services, this method groups factors into 5 general dimensions: tangibles, reliability, responsiveness, assurance, and empathy, which are, at the same time, described by several items that are scored by customers. Numerous other scales and quality indexes have been developed based on the principles of SERVQUAL, such as SERVPERF (Cronin Taylor, 1994). The same authors continued developing their scale finding it can be constructed on 3 and not 5 dimensions, merging some of them into the same one, at the same time that the ordered scale was redefined into a nine-point scale (Parasuraman *et al.*, 1994).

In the transportation field, Hu and Jen (2006) applied a similar method, where 20 items were rated to determine the quality perceived in the bus system in Taipei. Similarly, Wen *et al.* (2005) proposed 20 items to be rated in a 5-point scale from *Very unsatisfied* to *Very satisfied*, and uncovered three latent factors: operational performance, on-board amenity, the crew's attitude. Specifically, passengers were willing to pay the most for ride smoothness, safety, on-time performance, ease of purchasing tickets and cleanliness in the bus. Alternatively, the research by Tyrinopoulos and Antoniou (2008) proposes 23 items grouped into 4 dimensions that were evaluated through a semantic five-point scale from *Very dissatisfied* to *Very satisfied*.

The majority of studies define quality in 5, 7, or 9 levels. A 10-point scale and 16 items were proposed by Eboli and Mazzulla (2007) to study a bus system whereas a 11-point scale was defined by Nathanail (2008) to evaluate the quality of 23 factors in order to determine the behaviour and satisfaction of users of the railway services in Greece.

Not only satisfaction ratings and quality marks have been applied in the research on transit quality but also the level of agreement with a series of statements (Friman & Fellesson, 2009), whereas the importance of the various attributes describing the service

were collected by Cantwell *et al.* (2009) and introduced to model the citizens' behaviour in terms of quality perceptions.

Both satisfaction and importance were collected via personal interviewees in the experiment designed by Garrido and Ortúzar (1994). The importance of a set of factors was revealed through rankings on the level of service in the bus system in Santiago (Chile). Twelve aspects were ranked in order of importance, selecting the seven most relevant; the variables were also marked in a 7-point scale going from *Very bad* to *Excellent*. The authors derived the willingness to pay for improving the level of service described by the various attributes. Eboli and Mazzulla (2009) also collected importance and satisfaction ratings that were applied to construct a satisfaction index.

Key factors in transit quality

Before addressing the quality of service of the BSS, and in order to draw a general picture of how transit quality is perceived, it is interesting to gather the main outcomes obtained in previous research.

From an in-depth review of the international literature on transit quality it can be concluded that the most important characteristics of a transport system from the passengers' viewpoint are usually reliability and safety, but also cleanliness and the staff's behaviour. However, the particularities of each experiment lead to different variables identified as priorities; this situation is interpreted caused by the objective of the study, the survey design, sampling method and sample characteristics, the variables selected or the social circumstances and lifestyle, among others.

The TCQSM establishes "quality of service frameworks" described by two general concepts: availability and comfort and convenience. The availability is evaluated in terms of the spatial, temporal, capacity and information availability, but also considering the accessibility to different modes, the bicycle access and the access to passengers with disabilities. The comfort and convenience concern on passenger loading, reliability, travel time, safety and security, cost, appearance and comfort, and the treatment to customers. In the research by Garrido and Ortúzar (1994), the seven most relevant items selected by interviewees were the cost of travel, the variability of waiting time, the waiting time, the possibility of travelling seated, the in-vehicle travel time, the accident risk and the vehicle comfort. The price was found to be the most important factor for students and workers

and such attribute was used to derive willingness-to-pay measures, obtaining that peak and off-peak users are willing to pay the most for comfort and security, as well as for bus driver appearance and driving behaviour.

Transit service coordination, service frequency and accessibility were determined to highly influence the users' satisfaction in the study by Tyrinopoulos and Antoniou (2008). Interestingly, dell'Olio *et al.* (2011b) demonstrated that potential users do not care about the same attributes as current users do: potential users are not concerned about the cleanliness and comfort, as current users are; instead, potential users place importance on the journey time, waiting time and bus occupancy.

Eboli and Mazzulla (2010) found the service frequency to be the attribute showing the highest weight, followed by the reliability, the cleanliness and the bus stop facilities. Lai and Chen (2011) obtained that vehicle safety, cleanliness and complaint management significantly influence behavioural intentions. More recently, de Oña *et al.* (2013) obtained three latent constructs: service, comfort and personnel. The authors found that service had the highest influence on the overall service quality and, specifically, the frequency and the speed. Their model also identified the staff behaviour to be more relevant to passengers than the comfort.

Behavioural implications of quality provision

Even when the priorities are clear, what does this imply in terms of customers' behaviour? Zeithaml *et al.* (1996) concerned on this issue, moved by the highest objective in the business field: predicting profitability of actions. As the authors state, the best business strategy is to maintain the present demand, more than attracting new customers. The TCQSM defends that "Better quality of service is more attractive to potential passengers and generates higher ridership than lower quality of service". The manual dedicates a section to the expected variations in ridership as a result of the improvement on the service standards.

The repurchase intentions were examined by Mittal and Kamakura (2001), who proposed a model to determine the effect of individual characteristics on their purchase behaviour after their satisfaction has been revealed. The authors hold there is an actual latent satisfaction which differs from the observed one but that is more directly connected to the individual's purchase intentions. With this premise, the probability of remaining

loyal evaluates whether the latent and "true" satisfaction is greater than a given threshold that considers that competing alternatives also affect the consumers' expectations. Two five-point semantic scales were defined to rate satisfaction and purchase intentions. These were defined in terms of likelihood: from *Very likely* to *Very unlikely*. Interestingly, their research concludes that although the satisfaction-intention relation yield lower returns, the satisfaction–behaviour yield greater returns. The authors suggested that service suppliers should focus on specific user segments by considering the aspects to those individuals place the most importance on.

Service loyalty is also considered by Caruana (2002) together with satisfaction and service quality in a same model. The author reviews the definitions and studies on loyalty and motivates the research into loyalty and its connection with education, occupation or income. Lai and Chen (2011) identified three dimensions of customer loyalty: word-of-mouth, purchase intentions and price insensitivity. The authors contribute to the knowledge on the role of the individual's involvement together with attitudes on the customer's experience and behavioural intentions. The level of involvement refers to an individual's feelings regarding the relevance of the service and the sense of concern and care; this factor was found to directly and indirectly influence the repurchase intentions.

Friman (2004) found that the satisfaction experienced by the user due to improvements in the quality of service has a limit. The author analysed passenger satisfaction before and after these improvements and confirmed that they do not always lead to an increase in satisfaction. Three reasons were found in such case. First, a negative experience with the service or incident shows to affect satisfaction during long periods of time. Second, the explicit knowledge that the service has been improved increases passenger expectations and results in a higher number of incidents reported when the service does not meet those expectations. Finally, the different departments comprising the operator may show deficiencies in terms of cooperation and communication.

Modelling the service quality

Diverse approaches have been proposed in the literature on transit quality to better understand how the service performance is evaluated and identify the main attributes from the passengers. Satisfaction indexes have been widely developed to understand the passengers' satisfaction, in the line of the SERVQUAL scale. These propose mathematical expressions

that consider the quality evaluations of both the overall quality and the specific attributes comprising the service (del Castillo & Benítez, 2012; Eboli & Mazzulla, 2009).

Another technique that has shown potential in the identification of the main attributes of a service are decision trees. This non-parametric method has been applied by de Oña *et al.* (2012) to satisfaction ratings collected from passengers, obtaining the derived importance of the set of attributes predefined by the authors.

The majority of studies apply discrete choice models (Ben-Akiva & Lerman, 1985), assuming individual choice scenarios where the service quality is introduced. The selection of the precise discrete choice model is determined by the objective of the study, the theoretical framework and the definition of the choice framework. Multinomial models have been applied based on stated preferences (SP). In these studies it is assumed that the choice is actually the alternative providing the highest service quality and is thus proposed by Eboli and Mazzulla (2008a) to assess a service quality index, whereas dell'Olio *et al.* (2011b) define their SP experiment in terms of quality desired by users. A similar approach is proposed by Eboli and Mazzulla (2010); in this case the choice model considers not the levels of the attributes but the rating provided by the user for each level.

More complex discrete choice models have also been proposed that overcome the simplistic assumptions of multinomial logit. A nested structure was proposed by Hensher *et al.* (2003) whereas a mixed logit approach was proposed by Eboli and Mazzulla (2008b) and Cirillo *et al.* (2011). Interestingly, the study by Espino *et al.* (2008) explores and compares the explanatory power of different specifications with the aim of considering the preference heterogeneity in the assessment of the willingness to pay for improving service quality. The authors obtained lower values using mixed logit models than multinomial, concluding the existent variability should be considered in order to accurately measure and predict the effect of quality improvements.

Hybrid models introducing latent variables into multinomial logit models were proposed to improve models on travel behaviour by Ben-Akiva *et al.* (2002) and Morikawa *et al.* (2002). This approach has been applied in several studies concerned on the potential benefit of incorporating quality perceptions and attitudes in discrete choice models (Raveau *et al.*, 2010, 2012; Tam *et al.*, 2010; Wen *et al.*, 2005). In such a case, latent unobservable variables are introduced in a choice model, increasing its explanatory power. Exploratory factor analyses and confirmatory factor analyses are first computed to uncover

the latent unobservable variables explaining a set of predefined items that are ranked or marked by individuals. In a more complex way, Structural Equations Models (SEM) have also been applied to describe the complete conceptual framework around latent factors where these are treated as both exogenous and endogenous variables, leading to simultaneous equations describing the relations within the phenomenon. Golob (2003) focused on the use of this technique to understand travel behaviour whilst several studies opted for it to explore transit quality (De Oña *et al.*, 2013; Eboli & Mazzulla, 2007, 2012; Lathia *et al.*, 2012).

When not actual choices are collected but quality ratings, satisfaction marks or the level of agreement, ordered logit and probit models have been shown ideal specifications for modelling the users' perceptions (dell'Olio *et al.*, 2010; Hensher *et al.*, 2010; Rojo *et al.*, 2013; Tyrinopoulos & Antoniou, 2008). The ordered scales used to define the service quality within a discrete choice framework permit to successfully reproduce the users' behaviour and identify the most important attributes testing for diverse sources of heterogeneity.

The literature suggests various form to consider the variability in quality perceptions. The travel requirements, the socio-economic characteristics and other trip circumstances have been shown to systematically influence on perceptions (Gonzalo-Orden *et al.*, 2011; Mittal & Kamakura, 2001; Rojo *et al.*, 2013; Tyrinopoulos & Antoniou, 2008). For instance, Garrido and Ortúzar (1994) found that peak and off-peak travellers perceived the quality differently, justified by the different level of service in the two time periods. Tyrinopoulos and Antoniou (2008) compared the different importance places on the same attributes but different transit systems. Their results suggest that metro travellers perceive the service reliable, and thus their satisfaction goes beyond the level of service since cleanliness, staff behaviour and ticketing systems are the variables that play a key role. On the contrary, bus and trolley buses were found mostly evaluated in terms of frequency, coverage and punctuality. However, the quality of transport services has not only been demonstrated to vary across systems but also across lines within the same system, as found by Bordagaray *et al.* (2013), suggesting actions should be directed and adapted to specific services and bus lines in order to make the most of the resources applied.

Models require data that can be collected via two sampling methods: customer satisfaction surveys and stated preferences, this is, observations on the behaviour given

hypothetical choices (Ortúzar Willumsen, 2011). As pointed by Parasuraman *et al.* (1988), the choice should be made with caution since the former method should only be directed to past or present users that have an experience on the service, whereas the latter could be applied to any case, users and non-users since the hypothetical choices could be understood as the quality desired or the expected quality.

A review on the demand for cycling

Cycling has been widely analysed and modelled and, in the last years, automatically operated BSSs have attracted the scientific interest. Nevertheless, the users' perceived quality regarding these services has been overlooked. This has motivated the present research, together with a conclusion by Tyrinopoulos and Antoniou (2008), who hold that not the same attributes are important in the evaluation of the service quality in different transit systems. Consequently, the attributes that have been identified important in other transit modes in previous studies on transit quality may not be so in the case of bike-sharing. Furthermore, the particularities of this mode of transport which is also conceived as a tool for recreation make it a rather different problematic and, thus, this study constitutes an original contribution to the knowledge on transit quality.

Notwithstanding, some previous studies on cycling and BSSs have been very insightful at the time of defining the attributes that comprise such a public service and which thus have been considered in the present research.

There are specific profiles of people whose behaviour with respect to the bicycle has been addressed in detail: commuters and women. The commuters represent an interesting study group due to the high number of journeys they make travelling to and from their work places and the particular requirements of this type of journey. This explains why certain research has specifically focused on the influential variables on mobility patterns related to trips to work (Gatersleben & Appleton, 2007; Heinen *et al.*, 2010, 2013; Parkin *et al.*, 2008). On the other hand, the preferences of women with respect to the bicycle have been researched by Bernhoft and Carstensen (2008) and Garrard *et al.* (2008). The double viewpoint proposed by Bonham and Wilson (2012) is interesting: they studied the factors that influence women to both start and stop cycling. Akar *et al.* (2013) determined the different demands of men and women with reference to safety and feasibility.

Safety is indeed a factor which is generally associated with the availability of bicycle

infrastructure. This aspect has shown to have an important effect on women's behaviour (Akar *et al.*, 2013, Garrard *et al.*, 2008), but also in general among the population. In fact, the literature highlights the infrastructure as one of the decisive factors in the promotion of the bicycle as a mode of transport (Akar & Clifton, 2010; Dill & Carr, 2003; Dill, 2009). However, any infrastructure should be at least socially justified to be installed and then the capacity of bikes to be used as a mode of transport needs to be assessed.

The potential of the bicycle in multimodal trips has also been a major interest to the scientific community. Especially railway modes have been demonstrated powerful in combination with the bike (Martens, 2007; Midgley, 2011; Replogle, 1993; Rietveld, 2000b), concluding another attribute that describe the quality of the service: the connection of the BSS with other transit modes.

As found by Heinen *et al.* (2013) and Maldonado-Hinarejos *et al.* (2014), the lifestyle is also an influential factor in the mode choice towards cycling. These authors introduced socio-economic characteristics in the choice models such as age, gender, ethnicity and the residential area, assessing the explanatory value of these characteristics on the mode choice. The car ownership (Rietveld & Daniel, 2004) and the income level (An & Chen, 2007; Parkin *et al.*, 2008) have also been demonstrated influential in the attitude towards cycling.

Other variables that influence the demand for cycling have to do with the topography (Rietveld & Daniel, 2004), the weather conditions (dell'Olio *et al.*, 2011a) or the land use (Faghih-Imani *et al.*, 2014). Furthermore, the trip distance has been recognised a limitation to cycling (Heinen *et al.*, 2013) whilst below a certain threshold, the bike is claimed a competent urban mode for transport.

In terms of bicycle rental systems, the majority of studies analyse journey patterns and flows depending on time of day and day of the week as is the case of Lyon (Borgnat & Abry, 2011), Barcelona (Froehlich *et al.*, 2008) and Paris (Nair *et al.*, 2013). This information is very interesting to better understand the trip conditions under which the bike is demanded as public transport. Precisely motivated by its conception as a transit mode, and given that such systems should respond to public demand and count on a limited number of bicycles and docking stations, the scientific community has also worked on the provision of guidelines to consider in order to introduce new rental schemes (dell'Olio *et al.*, 2011c; DeMaio, 2009; Midgley, 2011; Monzón & Rondinella, 2010). In this regard,

several methodologies have also been proposed to determine the number and location of the stations comprising the system (Romero *et al.*, 2012) in the line that other studies apply optimisation techniques to determine the location and capacity of ordinary bike-parking points, as well as the routes between origin and destination pairs (Lin & Yang, 2011).

The literature review has provided the basis and the starting point of this research. A knowledge gap has been identified and is addressed in the present chapter: a methodology to determine how the quality provided within the BSS is perceived by users. Although these schemes have been a niche of study in recent years, studies have mainly focused on travel patterns and the optimisation of bike redistribution routing. This research seeks to provide knowledge on the existent need to ensure that the service is efficiently managed and that the supply fulfils the citizens' transport demands. With this premise, the ordered discrete choice theory is proposed to understand the users' perceptions and determine the aspects that should be prioritised in the management and improvement of the service. Furthermore, considering the conclusions from the literature review, the heterogeneity is included in the modelling in different forms in order to determine what causes variability in perceptions. Therefore, the proposed methodology considers the previously reviewed scientific knowledge and seeks to contribute to determine the attributes of a BSS that users place importance on.

2.3 Methodological approach

The aim of this research is to characterise the quality perceived by users of a public bicycle hiring service, recently introduced in numerous towns and cities worldwide, as is the case of the town being studied (Santander).

Three points should be highlighted from the methodology followed in this research: the important role of the citizens involved in the first stages through their participation in focus groups and in the data collection process, the consideration of different causes of heterogeneity in preferences, and the application of the modelling to a recently installed public bicycle system. The different phases guarantee the reliability of the results and thereby permit their inference to the overall population. This information is of great value for the efficient design of the improvements to make to a public bicycle system that will result in a positive impact on the public's perception of service quality.
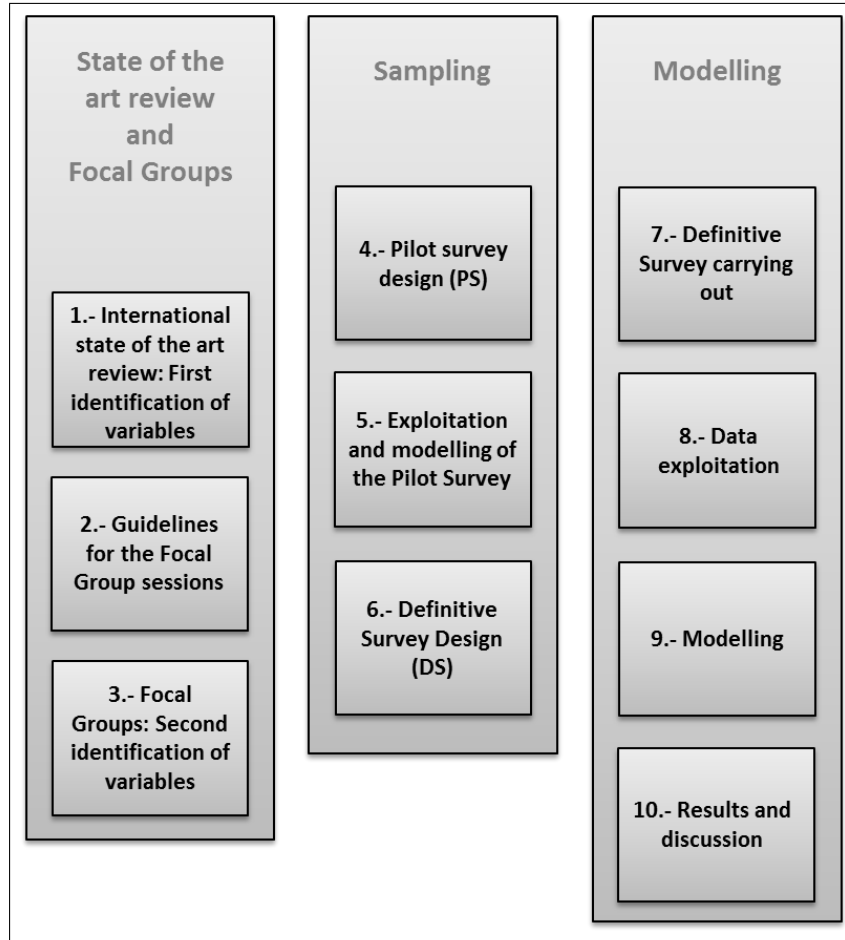
Figure 2.1: Methodological sequence to assess the users' perceived quality

Figure 2.1 shows the methodology followed to assess the quality perceived by the users of the BSS. After the literature review, the citizen involvement is of utmost importance when such an infrastructure as the bike-sharing system has been recently implemented and even more when the demand is not established yet. The results from phases 1, 2 and 3 are considered in the sampling design (phases 4, 5 and 6). The data collected finally undergo the mathematical modelling of the perception of quality in the public bicycle service. Hence, this research proposes a mathematical and scientific methodology that relies on the public perceptions, sensitivity and willingness to use the bike-sharing service.

Phases 2 and 3 (Figure 2.1) represent the initial exploratory research aimed at determining the most relevant variables to be included in the pilot survey. Focus groups (FG) are meetings made up of people with certain characteristics and interests in common; they provide qualitative data and information by participating in a discussion focused

around a determined subject (Ibeas *et al.*, 2011), in this case the demand for a bike-renting service.

Regarding the case study of Santander, the participants in the FG considered that more docking stations should be installed at the entrances to the city, both for private car users and public transport users (park and ride installations at bus and train stations). They also claimed for adequately signposted and segregated bike lanes to enter the city from the outskirts and other municipalities. In this respect, it is essential that more parking facilities are installed throughout the city, such as shopping centres, financial and business districts, public services, education establishments, etcetera. Incentives should be directed to encourage businesses to provide their workers with good infrastructure related to cycling (showers, secure parking areas, etcetera). The members of the focus group affirmed that these changes would encourage modal change in favour of the bicycle with a corresponding reduction in private car use for making everyday trips.

The knowledge acquired from the literature and the results of the focus groups provided the background for phase 4 (Figure 2.1) in which the pilot survey was designed. The design of the survey was defined by the goal of characterising the quality of service perceived by users of public bicycle systems. With this goal in mind a Revealed Preferences (RP) survey was proposed to study the current quality standards provided by the system. The interviews were held at the docking stations and responded by the cyclists returning the bicycle. The user was encouraged to score the standards of diverse aspects of the service they experienced on their just-finished trip. The pilot data was analysed and modelled so that any necessary changes could be made to ensure the accuracy of the definitive survey (phase 6, Figure 2.1).

2.3.1 Survey design

The ideas that arose in the focus groups were introduced into a pilot survey and the results were evaluated, modelled and corrected leading to the design of the final survey on perceived quality in a public bicycle hiring service. The final survey was administered in July-August 2011 to users returning the bicycles to the docking stations since the presence of cyclists was still scarce in Santander but less accused in summer, as confirmed by previous research conducted in this city by Monzon *et al.* (2010) and dell'Olio *et al.* (2011a).

The sampling method consisted in surveying a ratio of users proportional to the demand at each station at different intervals of peak and off peak times. The survey was divided into two parts in order to characterise both the users and their journey patterns as well as to collect their quality perceptions. The first part of the survey provided the individual's socioeconomic characteristics (Ortúzar & Willumsen, 2011) as well as data on origin and destination, journey purpose, frequency of journey and other data. The analysis of this data (phase 8, Figure 2.1) provides useful information about different types of service users and their characteristics which help explain the different perceptions or behaviour of the users. The second and last section of the survey provide the data required in the following stages of the methodology to determine the influential variables on user perception of quality. The first and the last of these questions are actually the same one and refer to the overall quality perceived by the user. Each of these responses will be the dependent variable in two separate models. In between both responses, the interviewees are asked to evaluate the quality of the characteristics of the service, one by one. The scale to mark the quality is semantically defined: *Very bad*; *Bad*; *Not good, not bad*; *Good*; *Very good*, whilst an ordered and numeric scale going from 1 to 5 has been used to analyse the data. These ratings (the quality perceptions of the attributes of the service) were introduced as explanatory variables of the overall quality in each model.

Hence, the questionnaire provided the following information: socio-economic characteristics, origin, destination, journey purpose, frequency of journey, access, journey, and egress time, and ratings to evaluate the quality of the various attributes and the whole service. 195 interviews were complete and used to model the service quality, with an average of 500 daily rentals in the summer of 2011. The sampling format to score the quality standards was first proposed by dell'Olio *et al.* (2010) and applied to determine the most important attributes for users of the bus service. The research concluded that the overall quality of service is perceived differently by the user after having evaluated certain specific aspects than it was at the beginning of the survey. The potential for explaining the choice mechanisms occurring in these different moments during the survey is the reason why the perception of the system as a whole is asked twice: when the user rates the overall quality for the first time, several aspects influence their perception, whereas after the interviewer verbally names the characteristics making up the system and the user scores them individually, their perception changes.

2.3.2 Analysis of survey data

The user and journey characterisations provide some very interesting information about the demand and the main uses of the service. With this knowledge, actions can be designed targeting determined user segments depending on the characteristics of the current and potential demand.

The analysis of the collected data reveals that 51% of users are women and 49% are men. It is noteworthy that more than half of those surveyed (56%) are under 25 years old, 15% are between 35 and 44, 13% are aged 45 to 54, 12% from 55 to 64 and 4% are over 64 years old. While 85% confirmed that they had a driving license, only 67% had access to a private car. In terms of household income, 3% stated under €900 per month, 17% between 900 and €1,500, 30% between 1,500 and €2,500 and 23% replied their income was over €2,500. However, 28% did not know or did not answer that question. 74% of the interviewees were residents in Santander. The journeys were mainly made for leisure (44%), 37% were travelling home, 1% for education, 3% for work, 2% for health purposes, 3% for shopping and 11% were travelling for other reasons. The service is used on a daily basis by 42%, weekly by 31%, monthly by 2% and sometimes by 25% of the interviewees. Finally, regarding the use of the bike lane during the journey, 58% said they had used it, 9% said they had not used it and 33% said they had used it in certain sections (because it did not exist along the entire route or they had taken an alternative one).

The data collected in the second part of the survey deal with the perceived quality of the public bicycle service and are used to calibrate the models. Users were asked to mark the overall quality of the service twice: at the beginning and at the end of the survey. Each of these scores are the data used in the dependent variable of one of the two separate models that will provide insights into the users' perception in those two moments: the initial and the final perceived quality (IPQ and FPQ). The explanatory variables in both models are the ratings regarding the quality of the specific features of the public bike system. The average score turned out to be over 4, meaning that the quality of the service is perceived to be in between *Good* and *Very good*. The highest scores were given to the cost and the travel time and, on the contrary, the bike lane design and the quality of the bicycle received the poorest ratings.

Finally, Table 2.1 reveals interesting information based on the interviewees' responses.

Table 2.1: Average travel time and initial (IPQ) and final (FPQ) perceived quality according to the trip purpose

Trip purpose	Av. travel time (mins)	Av. IPQ	Av. FPQ
Leisure	34	4.27	4.24
Work/study	10	4.17	4.00
Shopping	24	4.20	4.00
Home	21	4.13	4.03
Other purposes	33	4.28	4.08

The demand for the public bike for leisure and to go to work or study yield very different values. Those demanding the BSS for leisure travel 34 minutes in average (the highest, according to Table 2.1) and are the most satisfied with the service. On the contrary, users demanding the service as a mode of transport to go to work or study spend about 10 minutes and are the most critic on the quality of the service.

Therefore, there is evidence that the purpose of the trip is associated with different travel behaviour in terms of travel time and that different quality standards are perceived depending on the aim of the trip made.

Interestingly, in all cases the final perceived quality (FPQ) is lower than the initial perceived quality (IPQ), which means that users are more critic after making them reflect on each of the attributes characterising the bike-sharing service.

2.4 Modelling the user's perceived quality

2.4.1 Theoretical framework

Discrete choice modelling is proposed to estimate the service quality perceived by users. It is a disaggregate approach that models choice processes (Ortúzar & Willumsen, 2011) where each response corresponds to the behaviour of an individual facing a choice from a limited number of alternatives and subject to certain conditions and constraints.

These models are based on the Random Utility Theory, where an individual q chooses the alternative i which provides the greatest benefit or utility U_{iq} following the expression:

$$U_{iq} = V_{iq} + \varepsilon_{iq} \tag{2.1}$$

Where the utility is made up of the terms: V_{iq}, or systematic utility, and ε_{iq}, random

component represents the likes and dislikes of each person which are not explained in the group of attributes, as well as any possible mistakes made in measuring or during data collection, thereby endowing the model with realism. This leads to the conclusion that it is not possible to predict the user's behaviour with absolute certainty and, therefore, probabilities are assessed that a specific individual will choose one of the available alternatives.

The systematic utility V_{iq} is explained by a set of variables which are affected by their corresponding coefficients in the following way:

$$V_{iq} = \sum_k \theta_{ki} X_{kiq} \tag{2.2}$$

Where X_{kiq} represents the value of the variable k of the alternative i for an individual q, and each parameter θ_{ki} represents the weight placed on the variable k of the alternative i. In turn, these parameters rank the variables by importance within the utility, which in this case is the overall perception of quality.

The model in this research is based on the specification proposed by McKelvey and Zavoina (1975), who defined the ordered model as a latent regression adapted to ordinal outcomes. In this case study, the semantic and ordered ratings provided by users were converted into discrete but ordinal numeric values. The score on the overall service was introduced as the dependent variable following the expressions 2.3 and 2.4 below (Hensher *et al.*, 2005):

$$y_i^* = \theta_{ki}x_i + \varepsilon_i \quad \text{where} \quad \varepsilon_i \sim F(E(\varepsilon_i)|Var(\varepsilon_i)), E(\varepsilon_i) = 0, Var(\varepsilon_i) = 1 \tag{2.3}$$

$$y_i = j \quad \text{if} \quad \mu_{j-1} < y_i^* \leq \mu_j \tag{2.4}$$

Where j represents each of the responses comprising the numeric scale. The expressions 2.3 and 2.4 imply that the unobservable, dependent and continuous variable, y_i^* is transformed to an ordinal and discrete variable y that contains the observed responses. In the Ordered Probit model, the error term ε_i in 2.3 is assumed to be normally distributed.

The maximum likelihood method (ML) is applied to estimate the parameters θ_{ki} and μ in discrete choice models. This method produces the most probable answer for each individual according to the responses given to the explanatory variables, the socio-economic characteristics and their mobility constraints. The series of probability functions associated with each outcome j is given by the following expression:

$$Pr(y_i = j|x_i) = F(\mu_j - \theta_{ki}x_i) - F(\mu_{j-1} - \theta_{ki}x_i) > 0, \quad j = 0, 1, ..., J \tag{2.5}$$

The log likelihood is the logarithm of the probability expression in equation 2.5:

$$logL = \sum_{i=1}^{n} \sum_{j=0}^{J} m_{ij} log[F(\mu_j - \theta_{ki}x_i) - F(\mu_{j-1} - \theta_{ki}x_i)] \qquad (2.6)$$

where $m_{ij} = 1$ if $y_i = j$, and 0 in other cases.

In the case of the ordered model, μ is directly related to the probability, within the population, of choosing as a quality rating each specific level j on the ordered scale.

The generalisation of the ordered model assumes that the parameters θ_{ki} are not constant across the population but, instead, they are randomly distributed. This means that the coefficients of the various explanatory variables of the model vary from one individual to another. This generalisation introduces variability in the preferences and permits variables to be correlated.

The random coefficients respond to the following definition:

$$\theta_{ki} = \theta_{ki}^* + \Delta z_i + \Gamma v_{ki} \qquad (2.7)$$

where θ_{ki}^* is the average value of the parameter in the population and the systematic heterogeneity in the mean is induced by the variables z_i which are normally socio-economic in nature or journey restrictions. The variability across the population in the perception of each explanatory variable k of the alternative i is taken into account through the unobservable random component v_{ki}, which will present the distribution which best fits the existing heterogeneity.

The mean of the parameters for a specific individual characterised by z_i is:

$$E[\theta_{ki}|x_i, z_i] = \theta_{ki}^* + \Delta z_i \qquad (2.8)$$

The variance of the parameters is:

$$Var[\theta_{ki}|x_i, z_i] = \Gamma I \Gamma^{'} = \Omega \qquad (2.9)$$

Where Γ is the lower triangular matrix that introduces the correlation between the random parameters. Similarly, the parameter thresholds μ may vary across the population. In the ordered model, the final conclusions about the choice mechanism are derived from the partial effects (Hensher *et al.*, 2005), which report on the impact of a variable's variation on the choice probability of a specific result of y in equation 2.5.

The partial effects report on the effect a variation in a variable has on the probability of a specific result of y following the equation:

$$\delta_j(x_i) = \frac{\partial Pr(y_i = j|x_i)}{\partial x_i} = [f(\mu_j - \theta_{ki}x_i) - f(\mu_{j-1} - \theta_{ki}x_i)]\theta_{ki}^* \qquad (2.10)$$

The values of these effects may be both positive and negative in accordance with an increase or a decrease in the probability of choosing each alternative or scale level y.

The accumulated value of the partial effects of all the variables is also of interest:

$$\frac{\partial Pr(y_i \leq j|x_i)}{\partial x_i} = \sum_{m=0}^{j}[f(\mu_m - \theta_{ki}x_i) - f(\mu_{m-1} - \theta_{ki}x_i)]\theta_{ki}^* = -f(\mu_m - \theta_{ki}x_i)\theta_{ki}^* \quad (2.11)$$

2.4.2 Estimated models and discussion of results

Phase 9 2.1 deals with modelling the data collected in the definitive survey. This data is introduced into the mathematical specification previously described to assess the users' overall quality perceptions. Thus, the survey data has been used to estimate the ordered probit models where the potential explanatory variables of the overall perception of quality are the attribute ratings collected in the survey:

- IPQ: Initial perceived quality

- ATQ: Quality perceived regarding the access time

- JTQ: Quality perceived regarding the journey time

- COSTQ: Quality perceived regarding the cost of the service

- DESTQ: Quality perceived regarding the time to the final destination

- DISQ: Quality perceived regarding the distribution of terminals

- BKRQ: Quality perceived regarding the bike-lane routes

- PSQ: Quality perceived regarding the payment system

- BICQ: Quality perceived regarding the quality of the bicycle

- CONQ: Quality perceived regarding the connection with other transport modes

- JSQ: Quality perceived regarding the journey safety

- THEFTQ: Quality perceived regarding the anti-theft bike system

- INFQ: Quality perceived regarding the information about the service

- FPQ: Final perceived quality

Two methods have been applied to consider the heterogeneity in perceptions: random parameters are estimated to account for the general variability across the population; on the other hand, the interaction of the attribute ratings with socio-economic variables or trip circumstances that may have a systematic effect on the quality perceptions. In the former case, different distributions have been tested (normal, logarithmic, uniform, triangular) to represent the variation around the mean of the parameters affecting the explanatory variables. To consider the latter type of heterogeneity, the systematic variations in preference are introduced by the interaction between the score given to an attribute and a *dummy* variable representing whether or not each individual is characterised by a socio-economic circumstance or trip conditions. The resulting interaction term is a new variable containing the evaluation of the selected attribute, responded only by the group of users categorised by the socio-economic feature. Very diverse causes of interaction have been tested accounting for the profile features of each individual collected in the survey: gender, age, net monthly household income, residence in Santander, frequency of use of the service, trip purpose, use of the bike lane during the journey and whether a desire for more infrastructure has been stated, or for the improvement of the existing facilities.

During the interview, users were also asked about quantitative aspects of their trip: the access time to the docking station, the travel time and the cost. As with any mode of transport making up an urban mobility system, it is helpful, even necessary, to know what influence these variables have on the travellers' behaviour. They were not only introduced into the models just as the quality valuations directly collected in the survey (ATQ, JTQ and COSTQ), but also through the use of dummy variables characterising different time and cost ranges, as presented in Table 2.2. Each variable (access time, journey time and cost) was subdivided into three different categories so that the journey was characterised more specifically according to the purpose of the journey (in the case of the journey time, it was considered that 44% of the interviewees travelled for leisure).

Table 2.2 shows the distribution of intercepted users according to the trip purpose (home, education, work, health, shopping and other purposes) and the categories in which access time (AT) and journey time (JT) have been structured. The cost of the service is defined by the type of subscription acquired by the user: day-ticket, week-ticket or annual ticket.

Table 2.2: Distribution of leisure and other trip purposes according to the access time (AT) and journey time (JT)

Interval	Dummy variable	Leisure trips	Other trip purposes
$AT \leq 5$ mins	AT5	80%	88%
$5 < AT \leq 10$ mins	AT5-10	13%	11%
$AT > 10$ mins	AT10	7%	1%
$JT \leq 25$ mins	JT25	43%	73%
$25 < JT \leq 50$ mins	JT25-50	42%	21%
$JT > 50$ mins	JT50	15%	6%

Each of the two different ratings of the overall quality were introduced in different models as the dependent variable. As a result, the first model (Table 2.3) shows the variables influencing the perception of the quality at the beginning of the interview (IPQ) and the second model (Table 2.4) presents the attributes explaining the overall score at the end of the interview (FPQ), after the user was asked to score the individual attributes making up the system. The same explanatory variables have been introduced into both the initial and final models, yielding the significant parameters in Table 2.3 and 2.4, in each case.

As uncovered by the estimated models presented in Table 2.3 and 2.4, the following dummy variables lead to systematic variations in preference:

- AT5: Access time less or equal to 5 minutes

- JT25: Journey time less or equal to 25 minutes

- G: Gender (female=1; male=0)

- ESPP: Specific trip purpose (work, study, health care, shopping or other reasons)

- JT25-50: Journey time from 25 to 50 minutes

- AT10: Access time longer than 10 minutes

- Y: Young people (younger than 34 years old)

- M: Middle-aged users (34 to 54 years old)

Table 2.3: Initial perceived quality estimation model

Variable	Coefficient	T-test
Non-random parameters		
Constant	-1.19	-1.83
ATQ	0.04	0.23
ATQ·ESPP	0.39	2.50
COSTQ	1.25	4.93
COSTQ·AT5	-0.41	-2.68
COSTQ·G	-0.19	-2.07
BICQ	0.32	2.24
BICQ·JT25	0.56	3.92
PSQ	0.45	3.55
Means for random parameters		
DISQ	0.71	5.17
Scale parameters for random parameters		
DISQ	0.29	4.57
Threshold parameters for probabilities		
μ_1	1.56	3.38
μ_2	4.91	8.35
Observations: 195		
Log likelihood function: -111.09		

For the models to be estimated, every alternative or available response of the ordinal scale has to be represented in the sample so that the estimated model can reproduce a similar share of responses for all the scores and, thereby, reproduce the reality. Thus, given the small representation of the two lowest scores of the dependent variables (initial and final overall perceived quality), the models were estimated by aggregating the *Very bad* and *Bad* responses from the survey into a single category regarding negative perceptions.

The positive sign of the parameters needs to be assured so that the higher the system attributes are marked, the better is the overall quality of the service perceived. It also has to be noted that since all the independent variables are defined on the same scale, all parameters can be directly compared to identify the attributes the users place the most

Table 2.4: Final perceived quality estimation model

Variable	Coefficient	T-test
Non-random parameters		
Constant	-3.03	-3.41
BICQ	1.02	3.60
PSQ	1.00	4.48
INFQ	1.60	4.92
INFQ·M	-1.07	-3.38
INFQ·G	-0.49	-2.58
ATQ·JT25-50	-0.70	-3.79
ATQ·AT10	2.82	4.38
JSQ·Y	-1.14	-3.73
Means for random parameters		
ATQ	1.02	4.56
JSQ	2.38	5.12
DISQ	0.66	3.12
Scale parameters for random parameters		
ATQ	0.41	4.71
JSQ	0.60	4.99
DISQ	0.34	3.52
Threshold parameters for probabilities		
μ_1	3.50	4.22
μ_2	11.41	6.77
Observations: 195		
Log likelihood function: -77.63		

importance on, and thus, the partial effects will yield the same relative importance among the attributes.

The estimated models (Table 2.3 and 2.4) also inform about different sources of heterogeneity in the perception of the quality across the population of users. To begin with, Table 2.3 shows that the initial overall evaluation is heterogeneously explained by the perception of the distribution of the docking stations (DISQ). However, the model

determining the users' perception of quality at the end of the survey (Table 2.4) reports on the presence of three attributes that heterogeneously explain the final valuation of the service: the distribution of the docking stations (DISQ), the perception of the access time (ATQ) and journey safety (JSQ). In all cases, the normal distribution was found to yield the best fit of the parameter.

The comparison of the two models permits to realise the adjustment in the perceptions before and after the characteristics of the service have been explicitly listed and rated: the quality rating given to the cost of the public bicycle service (COSTQ) only influences the overall quality perceived at the beginning of the interview (IPQ) (Table 2.3). On the other hand, there are two variables influencing the perception of service quality at the end of the interview which did not do so in the initial model: the satisfaction on the safety during the journey (JSQ) and the evaluation of the available information (INFQ) (Table 2.4). Hence, although initially the cost of the service is considered in the overall perception of the quality, once the individual service attributes have been thought about, the cost ceases to have any influence and other factors show to be of importance in the final evaluation of the overall service, factors which had not even been considered at the start of the interview.

The model also reveals the existence of systematic variations in preference, which are studied through the interaction variables created by multiplying different factors. The categories of users showing this systematic variation in variable perception are affected by both parameters: that of the variable itself and that of the interaction term. Therefore, when evaluating the overall service at the beginning of the interview (Table 2.3), the perception of those users that have experienced journey times of less than 25 minutes is the sum of the interaction parameter BICQ·JT25 and the parameter BICQ.

Similarly, in the first model, those users whose journeys have been made for specific purposes (ESPP: work, education, health, shopping or other purposes) perceive the access time (ATQ) differently than the other users, who travel for leisure or to get home. Considering the systematic variations in preference that have been found significant in the second model (Table 2.4), it is worth pointing out that, the access time is again not homogeneously perceived across the population: those users who perceive an access time of over 10 minutes (AT10) are not only affected by the parameter ATQ but also by ATQ·AT10. Other systematic variations found in the model regarding the final valuation of the quality

(Table 2.4) indicate the specific perception of the time accessing the station in the case that the trip lasts between 25 and 50 minutes (JT25-50) and, the quality perceived of the quality during the trip by users younger than 34 years old (Y).

In the FPQ model, the parameters show a confidence level of greater than 99%. However, on the contrary, the coefficients of the IPQ model have a lower significance and the loglikelihood value also indicates a poorer fit. The significance of the parameters indicates the possibility of transferring these results onto the overall population. As the quality perceived at the end of the interview (FPQ) is explained by more significant parameters, it is possible to state that the individual's behaviour is known with greater precision in this case. For this reason, the partial effects of the FPQ model are analysed and discussed (Table 2.5). This uncovering of the difference between the overall service quality scores provided at the start and at the end of the interviews is interpreted as the effect of having previously considered the quality of each individual attribute before providing a final score for the overall service quality (dell'Olio *et al.*, 2010).

Table 2.5: Partial effects of the final perceived quality model

Variable	Very bad / Bad	Not good, not bad	Good	Very good
ATQ	0	0	-0.0107	0.0107
ATQ·JT25-50	0	0	0.0073	-0.0073
ATQ·AT10	0	0	-0.0294	0.0294
BICQ	0	0	-0.0106	0.0106
DISQ	0	0	-0.0069	0.0069
PSQ	0	0	-0.0104	0.0104
JSQ	0	0	-0.0249	0.0249
JSQ·Y	0	0	0.0119	-0.0119
INFQ	0	0	-0.0167	0.0167
INFQ·M	0	0	0.0112	-0.0112
INFQ·G	0	0	0.0051	-0.0051

The partial effects quantify the unit increase or decrease (positive or negative sign) in the probability of choosing each response available in the scale (*Very bad/bad*; *Not good, not bad*; *Good*; *Very good*) to score the overall quality as a result of a unit improvement on the satisfaction mark given to each influential variable. The scale defined to score quality,

as well as the amplitude of the intervals determined by the parameters μ of the model (Table 2.4), affect the probability of choosing each scale level to rate the overall service quality.

As the partial effects shown in Table 2.4 reveal, the improvement of the performance of any of the attributes and, thus, the increase in satisfaction regarding that service characteristic will result in an increase in the maximum valuation *Very good*. Since the average score of the overall valuation turned out to be in between *Good* and *Very good*, any improvement accomplished in the system would result in an increased probability of a *Very good* response to rate the overall quality and, therefore, the same reduction in the probability of perceiving the quality as *Good*.

As the coefficients in the model estimating FPQ already suggested (Table 2.4), the greatest impact on the overall final quality is caused by a unit increase in the satisfaction regarding firstly, the safety during the journey (JSQ) and, secondly, the available information about the service (INFQ). Measures designed to achieve these improvements would cause an increased probability of 2.49% and 1.67% respectively, representing an accumulated increased probability of 4.16% of rating the service quality with the maximum score. After these two attributes, the greatest positive impacts on the perception of quality would come from actions directed at service improvements regarding access time (1.07%), bicycle quality (1.06%), payment system (1.04%) and the distribution of docking stations across the city (0.69%).

Likewise, the partial effects accounting for the interactions should be added to the effect of the variable alone (in a similar way to the interpretation of the parameters, when the individual is characterised by the socio-economic variable or trip circumstance associated to the interaction term). In this respect, it is worth pointing out that users with access times over 10 minutes (AT10) place much more importance on this aspect than the general population of users do, as can be confirmed by adding 2.94% (AT10·ATQ) to the 1.07% (ATQ) that for the general population increases the probability of responding with the highest score to rate the overall quality when that particular aspect (ATQ) is improved. Table 2.5 also shows that the user categories involved in the remaining interaction terms experience a lower positive impact compared to the expected effect on the general population's perception. Such is the case of perceived quality in relation to access time for those users whose journeys last between 25 and 50 minutes (-0.73%), in the case

of the valuation of the safety perceived during the journey by consumers aged under 35 (-1.19%), and among females (-0.51%) and users aged between 35 and 54 (-1.12%) regarding the quality relative to the available information about the service. The direction of these effects was previously revealed by the negative signs of their corresponding parameters (Table 2.4), which remove weight from the importance placed on the corresponding variables by the general population. In summary, the interactions indicate several circumstances under which the impact of the quality of some attributes on the overall quality systematically variates. The quality perceived regarding the access time has been found to be less important when the trip lasts between 25 and 50 minutes, whereas it has a higher impact on the overall quality when more than 10 minutes are required to access the nearest station. On the other hand, safety is perceived less important among users younger than 35, and women and users aged between 35 and 54 tend to reduce the importance of the quality of the information provided about the service.

2.5 Conclusions

Aimed at improving the efficiency and management of bike-sharing systems, this chapter describes the integral process for evaluating the quality of service that users perceive they are receiving based on their experiences with the existing service.

The proposed methodology involves the public from the very first phases to guarantee a reliable representation of reality during the modelling process. The collected data is the result of successive phases involved with identifying variables and survey design which have then been introduced into ordered probit models to understand the perception of quality of service as a function of the characteristics which define it.

The scientific methodology described here can be applied to any town or city in a similar way to its application to Santander. The modelling provides interesting results on various aspects. This research has once again showed the difference in the perception of service quality before and after the interviewee has had time to reflect about and score different aspects of it. Variables such as the safety during the journey and the available information go unmentioned in the initial evaluation of overall quality while they turn out to be of greatest weight at the end of the interview. The opposite occurs with the cost, with the highest *a priori* score, but which ceased to be of importance for the users after considering each and every characteristic of the service. This and the scarce presence of

the level-of-service variables (access time, time to destination, journey time and cost) lead to the conclusion that conventional transport attributes have a certain influence on the perception of service quality, but they are not decisive for the BSS in Santander. This could also be a result of the still low modal share of the bicycle within urban mobility as a whole and that its use is still mainly for recreational purposes.

Evidence has been found that the purpose of the trip is associated with different travel behaviour in terms of travel time, and that different quality standards are perceived depending on the aim of the trip made: those demanding the BSS for leisure travel 34 minutes in average (the highest, according to Table 2.1) and are the most satisfied with the service, whilst users demanding the service as a mode of transport to go to work or study spend about 10 minutes and are the most critic on the quality of the service. Interestingly, in all cases the final perceived quality (FPQ) is lower than the initial perceived quality (IPQ), which means that users are more critic after making them reflect on each of the attributes characterising the bike-sharing service.

The proposed models represent a tool with a great potential as an aid in the management of bicycle-sharing services because not only does it report on the different service variables and their degree of importance on quality, but it also provides information on population variability and the systematic differences in perceptions resulting from different socio-economic and journey characteristics. Heterogeneity in the population appears in the perception of the distribution of bicycle docking stations while among the causes of systematic variability are gender, age, purpose of journey, type of ticket purchased and the access and journey times. Specifically, access time is the most important aspect for those users who experience an access time longer than 10 minutes. As a result, as reflected in the accumulated value of the corresponding partial effects, the improvement of this attribute is the one which generates the most noteworthy impact on the overall perception of quality among this group of users, suggesting the accessibility is fundamental for the users' satisfaction and to promote the uptake of bike-sharing, as already concluded in previous research (dell'Olio *et al.*, 2011a, Fishman *et al.*, 2013; García-Palomares *et al.*, 2012; Maldonado-Hinarejos *et al.*, 2014).

Finally, the presented research concludes by stressing the need to encourage users to evaluate each of the characteristics of a service to make their overall perception of quality more complete as, in this way the model is loaded with more realism. It is also

recommendable to report to the citizenship on improvements made to the service in order to guarantee the maximum impact on the overall quality of service perceived by users.

Chapter 3

A methodology to describe the demand for the service based on the usage behaviour recorded through smartcard data

3.1 Introduction

After obtaining valuable information from the citizens and users, and from the demand and quality indicators obtained in Chapter 2, an alternative source of data is examined in this chapter: smartcard transactions automatically recorded with every use of the system. The analyses proposed in this chapter to these objective automatic data are intended to provide a detailed picture of the demand that enrich the information collected through traditional manual techniques applied in the previous chapter.

Indeed, one of the major reasons why bike-sharing systems have spread so rapidly is the straightforward automatic operation and use. Users register themselves into the system and have access to any bike. The Intelligent Transport Systems (ITS) register every use of the service and associate each rental to the individual's subscription number. Undoubtedly, the huge amount of objective information generated by ITS can enhance transport estimations on travel behaviour and predictions. However, making the most of these data is difficult because of several reasons: incomplete knowledge about the

potential use of these records, lack of awareness of the limits to the validity and reliability of ITS data, and complexity of managing raw data and developing efficient algorithms its application.

As Vogel *et al.* (2011) noticed, it is limited the research that had applied data mining algorithms to study the demand for public bike schemes. So far, the main concern among practitioners and academics is the bike imbalances that the systems generally experience due to predominant travel patterns in terms of spatial and temporal distribution, as is the objective of the research by Vogel *et al.* (2011). Additionally, diverse methods have been proposed to infer bike availability predictions and to optimise the system so as to improve the supply as much as possible given a certain demand (Romero *et al.*, 2012). Although the interest in the demand for cycling has increased in the last decades, on the contrary, less numerous are the contributions dealing with bike sharing systems, and more specifically, those applying data from ITS such as smartcard transactions. This stage of the research aims to respond to this need by the development of new methods to obtain the most from automatic data. As a consequence, the research presented in this chapter is motivated by the need to study and take advantage of the potential and validity of bike-share smartcard system data for travel demand analysis.

However, extracting valuable insights from records generated by the ITS requires both data mining skills and some knowledge on the analysed system. In other words, the better the bike-sharing system is known, the more fruitful the data mining would be, and it is thanked to the knowledge obtained in the first stage (Chapter 2), that new methods can be designed to obtain the most from automatic data. This way, two data mining processes are proposed as a complement to the results on the quality perceived by users and on the characterisation of the demand for bike-sharing. These have been developed with an ultimate objective: to obtain the most from such a rich and valuable source of objective data as ITS are and while providing tools for practitioners to be less dependent on traditional manual data, being these limited and pricey. Undoubtedly, questionnaires permit to obtain valuable and direct responses regarding personal trip circumstances, or subjective issues such as the quality perceived by the users of the system, as addressed in Chapter 2. Nevertheless, although it is for certain that automatic records cannot explicitly gather these or even the purpose of the trip, far from being this a limitation, it is the challenge that motivates the present research, which takes advantage of the richness

and quality of the information recorded with every smartcard transaction.

The specific purpose of the research presented in this chapter is twofold. The first phase of the research aims to propose a data mining process to detect the rentals that are actually not trips but trials where the bike is promptly returned and substituted by a new one, suggesting a deficient performance of the first bike. As will be concluded, failing to remove bike trials from the data set to undergo the modelling would lead to inflated intra-zonal demand and reduced average travel time. Alternatively, a second process is described that identifies the BSS usage conditions that introduce systematic heterogeneity in travel patterns so as to infer travel patterns from the objective information of the users' actual behaviour.

In order to validate the methodology and algorithms, real data from the BSS in Santander (Spain) is used. As explained in section 1.4, this case study is of interest due to many reasons related to the city itself. The size of the BSS as well as the presence of tourists are the two main characteristics of the case study to justify its convenience in order to test and validate the methodology presented in this thesis and in this chapter in particular. The latter condition is expected to introduce substantial variability in the travel patterns and, thus, in the usage behaviour recorded within every smartcard transaction since local daily activities and tourism coexist in the city, mostly during summer.

The analysed dataset includes 26,290 records, collected from the 1st of July to the 31st of August 2011. In 2011 the system consisted of 14 terminals, about 200 bikes and more than 300 docks. Each smartcard transaction is described by the following fields that the system automatically records:

- Bicycle pick-up and drop-off docking station

- Identification number of the stand occupied by the bicycle at the origin station

- Time (date and time – hour, minute, second) of collection and return of the bicycle

- User's type of subscription (annual, weekly or daily)

- Subscriber identification number

- Bicycle identification number

The challenge consists of taking advantage of this information and understanding how the service is conceived by analysing the registered usage behaviours. For this to be

accomplished, a difference is made in this research between usage behaviour and travel patterns: the usage is described by the actual trip-chaining gathered with every smartcard transaction and is directly influenced by the limitations of the BSS as a public renting service, whilst the travel behaviour relates to the spatio-temporal distribution, the travel time and trip purpose. Two original data mining process have been developed that consider this difference and take advantage of the potential of automatic data to provide the variables and demand indicators to ensure an efficient design of the system and allocation of resources.

The next section reviews previous applications of automatic data to assess the transit demand. After that, the contents of the chapter follow the sequence of methods and analyses applied to the automatic data. Section 3.3 presents the algorithm to detect a usage pattern called "bike trials" and remove them from the data set for further analyses. In section 3.4 overall demand indicators are assessed that can be compared with other worldwide BSSs. Section 3.5 describes the data mining process to cluster the rentals into less heterogeneous groups of demand, providing a detailed picture of the demand that enrich the information collected through traditional manual techniques applied in the previous chapter. Finally, the conclusions are presented in section 3.6.

3.2 Previous applications of automatic data to estimate transit demand

Academics and practitioners in public transport have been moved by the great amount of rich data provided by the ITS. Valuable analyses and tools have been developed to understand the demand for public transport. These are mainly based on the characteristic trip-chaining in travel behaviour which can be constructed from the footprints that smartcard uses leave behind. Some interesting contributions in this regard were authored by Cortés *et al.* (2011), Gordon (2012), Munizaga and Palma (2012), Wang *et al.* (2011) and Wilson *et al.* (2009). The initials of this niche of research focused on deriving origin and destinations from sequential timestamps linked to the records of the vehicle location (bus/train). Most of the trips are successfully inferred through the data mining techniques proposed in the literature and this way, their travel patterns can be studied, the behaviour modelled and predictions derived.

In the case of automatic bike-sharing systems the timestamps at both origin and destination are recorded. As a consequence, the scientific research has been directed to characterise the particular demand for public bikes. The first studies applying ITS records from bike-sharing were aimed at analysing travel patterns. Some of the first successful BSSs were examined in this regard by Froehlich *et al.* (2008) (Barcelona), Noland and Ishaque (2006) and Lathia *et al.* (2012) (London), and Borgnat and Abry (2011) and Jensen *et al.* (2010) (Lyon). Froehlich *et al.* (2008) analysed the availability of bicycles throughout the day and proposed a clustering method to understand the relationship between the terminals locations and public bike mobility in Barcelona. They found that not all clusters refer to nearby stations, meaning that terminals spread all over the city may experience similar behaviour and that it is not a matter of location but of the land use, a conclusion that was also drawn by Vogel *et al.* (2011). These authors applied 3 clustering techniques to the data on the activity observed at the terminals comprising the system in Vienna so as to determine predominant travel patterns. The clustering is based on land use and facilities in the surroundings of the stations, confirming the effect of such factors on the demand levels.

Analogously, Borgnat and Abry (2011) clustered the data to uncover the flows between stations within Lyon's Vélo'v system and analysed the demand evolution, daily patterns and trip distance distribution. Furthermore, the authors calibrated the demand in order to forecast the hourly fluctuations. Jensen *et al.* (2010), examined the daily flows and routes observed in the same BSS, assessing bikeshare cycling indicators of the demand.

Interestingly, O'Brien *et al.* (2014), compared the bike mobility in 38 systems around the world based on automatic data. Their research is insightful in two ways: firstly, the study found 6 dominant patterns as a synthesis of the existent worldwide variability; secondly, the 38 schemes were clustered, resulting on groups of similar systems based on infrastructure and demand attributes.

Once a BSS is implemented, the level of service is mostly dependent on its operational performance: the bikes' management and their maintenance. For this reason, the other major interest in the recent years has been dealing with the bike imbalances that the asymmetrical demand for BSSs leads to (Chemla *et al.*, 2013). Smartcard data also permit the optimisation of the system based on the detailed information that is gathered

automatically. García-Palomares *et al.* (2012) do not apply smartcard records but GPS data to optimise the location and capacity of terminals. Their research also evaluate the demand for isolated terminals and their contribution within the system's demand. The authors test two objective functions and conclude the maximization of the coverage is more useful than minimizing impedance.

The potential of smartcard transactions within bike-sharing systems is still to be explored and the present research proposes a particularly different perspective. The difference is placed between the usage behaviour, described by the objective trip-chaining recorded with sequential timestamps and the travel behaviour, described by the spatio-temporal distribution, the travel time and trip purpose. The relation between these different perspectives is examined in this chapter and it is proposed to be considered to take advantage of the great amount of valuable objective data for evaluation and predicting purposes.

3.3 Database setup and first analyses: detection of bike trials

Generally, preliminary elaboration of raw datasets is required to prepare the data to undergo automated routines. The case of Santander allows illustrating some of the most common necessary processes.

Data interpretation. A clear understanding of the meaning of the variables that have been measured in each record is essential. This kind of information sometimes is not available or clear, and collaboration with the system operator is needed. Problems may arise when the subject interested in data analysis (e.g. a city council) is not the owner of the data (which often belongs to the system operator).

Database layout. The very first stage is to organize the files supplied by the system so that they can be easily analysed. For instance, the TusBic system in Santander provides a file for each day of operation. All the records have been collected in a single file to make data analysis easier. However, it should be highlighted that this was possible in our study because we have analysed only two months of a small system. In general, datasets can include huge numbers of records, and this can make data management burdensome.

Data cleaning. Before the analysis, the data need to be cleaned and the following

aspects should be checked:

- Repeated entries: The dataset in this research contained 882 duplicated records of two specific days during the period of study

- Missing data. In automated systems, missing data is often identified by special entries. Our of the 26,290 rentals recorded in July and August 2011 in the TusBic system, 46 showed a label "not referenced" in the field of the destination docking station. These records have been omitted in the analysed dataset. When the missing data is not used in the analysis, other information included in the record can still be used.

At the end of the cleaning process, 26,244 rental records, made by 5,615 users are left. Figure 3.1 reports the usage of the system per kind of subscription.
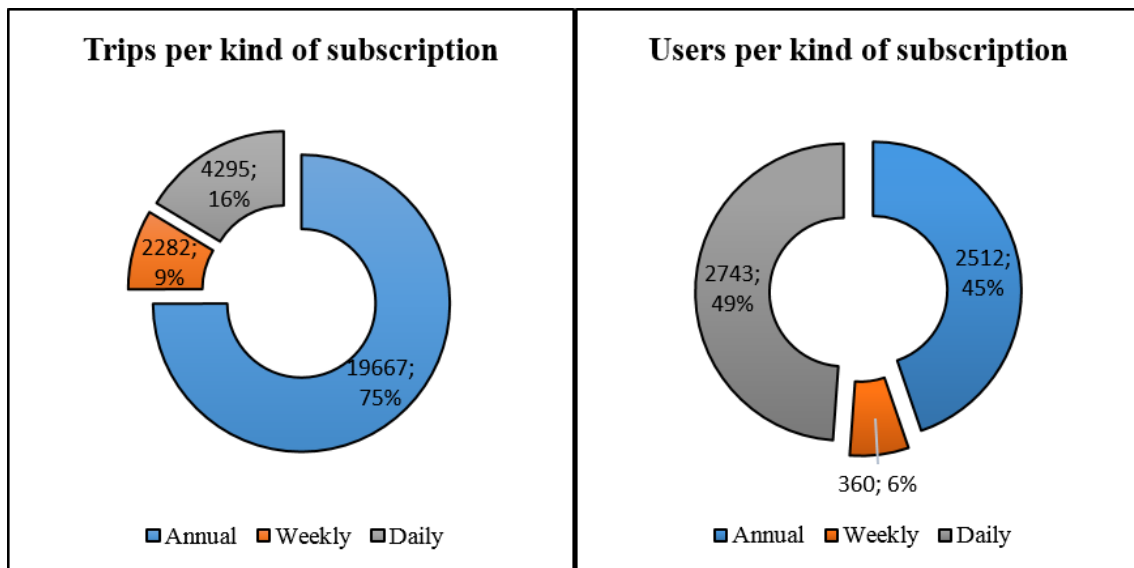
Figure 3.1: Sample description regarding the type of subscription: (a) share of trips; (b) share of users.

Data elaboration. New variables and quantities may be calculated from the automatically registered fields in order to extract the most from the data and to facilitate future analyses. In this case study, the following fields have been derived from the available data:

- the day of the week and the type of day (weekday, Saturday, Sunday; normal day/holiday);

- the rental duration or travel time as the time elapsed between timestamps at origin and destination.

The very first analysis expected to uncover interesting insights is the distribution of travel time (or rental duration) distinguishing the trips in which the origin and the destination coincide (hereafter COD trips) and those with different docking stations at origin and destination (DOD trips). The distribution of rentals turns up to be as follows:

- COD rentals: 6,335 cases (24.2%)

- DOD rentals: 19,909 cases (75.8%)

Figure 3.2 represents the frequency of trips according to the rental duration for these two behaviours.
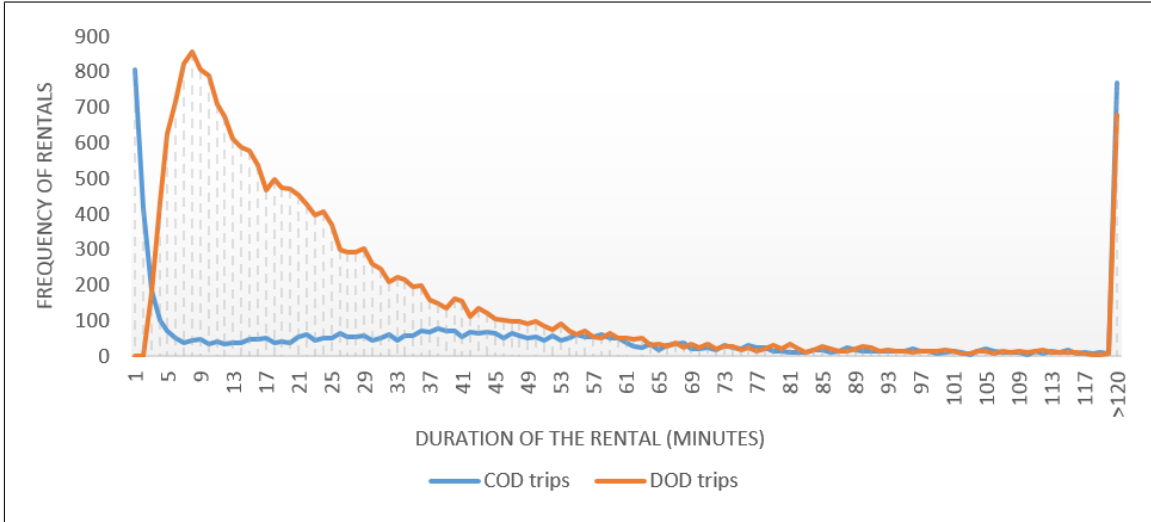
Figure 3.2: Histogram of rentals that the origin and destination are the same ($O = D$) or different ($O \neq D$), according to the duration of the rental

The curve of DOD rentals shows an increasing, quasi-linear trend for durations up to 7-8 minutes, where the frequency reaches a maximum of 856 rentals. The DOD case in Santander results in an interquartile range of 10-33 min (Table3.1). The COD journeys present a surprisingly different distribution with a rather big interquartile range of 5-67 min; the maximum number of rentals happens for a duration of less than a minute and 25% of records show duration shorter than 5 minutes.

The very high number of COD journeys showing reduced travel time is interpreted as the cases in which the bicycle is tested and then returned to the terminal without actually

Table 3.1: Quartiles of the duration of rentals

	COD rentals	DOD rentals
1st quartile	5 mins	10 mins
Median	37 mins	18 mins
3rd quartile	67 mins	33 mins

making the trip, presumably because it is faulty and so the users decide to replace it before heading out for their destination. This behaviour is called "bike trial". If our interpretation is correct, most of the short COD trips should be immediately followed by a new rental generated at the same terminal. To the aim of demand analysis, it could be considered that bike trial rental and the following one, this is, the trip itself, are actually the same "journey with bike substitution". Clearly failing to detect and consider bike rentals would result in biased demand analyses, regarding both the number of trips generated and attracted by each docking station (and so by different city zones) with inflated intra-zonal demand, and the average duration of trips to be reduced. However, to the author's knowledge, this pattern described as bike trials and bike substitutions have not been considered in the literature before. The presence of these particular rentals has been either ignored or directly omitted without even calibrating the pattern. The latter is the case of the research by Vogel *et al.* (2011), who directly remove rentals where the bike has been returned to the same station as origin within the first minute of rental. On the contrary, the interest in this section is actually describing this pattern so as to detect and remove the records from future analyses since, as showed by Figure 3.2, COD trips do not only last 1 minute bus also 2 or 3, meaning those could neither be considered actual trips.

Following the previous premise, a data mining process has been developed that is presented in Figure 3.3 and describes the conditions resembling bike trials and journeys with bike substitution.

The proposed process involves two steps. In the first step, every rental in the database is evaluated (the record considered in this stage is called Record 1), and two characteristics are checked:

1 - whether the docking station at destination (Destination$_1$) coincides with the origin (Origin$_1$);

2 - if the rental duration (Duration$_1$) is lower than X minutes. X needs to be
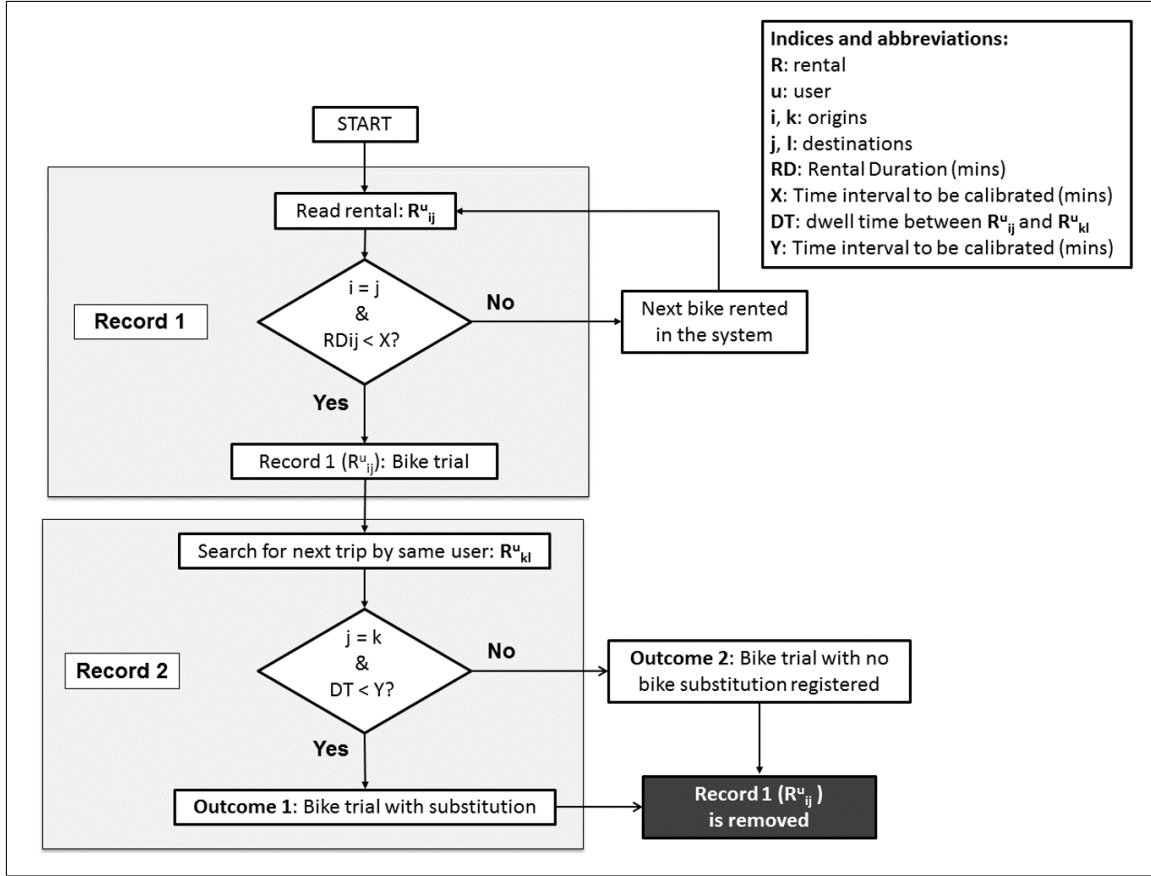
Figure 3.3: Data mining algorithm to detect bike trials

calibrated in each case study. If this condition is not verified, then the rental is interpreted as a journey itself and it remains in the database for demand analysis.

If both conditions hold for Record 1, the rental is considered a bike trial. In the latter case, the following rental made by the same user (Record 2) is evaluated to determine:

1 - whether the same user picks a new bike at the same docking station where the previous one was returned;

2 - if the interval between the return of the first bicycle and the rental of the second is smaller than Y minutes. Y needs to be calibrated in each case study.

When a Record 1 is detected, there are two possible scenarios:

- A Record 2 exists which satisfies both conditions (Outcome 1): Record 1 and Record 2 are considered part of the same journey corresponding to a bike trial with substitution.

- No Record 2 is detected that holds both conditions (Outcome 2): If the second rental

takes place at the same docking station of the first one, depending on the interval between the two, several scenarios may be hypothesized: the user does not intend to go anywhere but just to test a bicycle, or he would like to make a trip but cannot find a suitable bicycle, or he started his planned journey but was then he decided to abort it.

In any of the previous two cases (Outcome 1 or Outcome 2) is Record 1 considered in the dataset that will undergo the demand analysis. Nevertheless, distinguishing the case of Outcome 1 from the of Outcome 2 is interesting, because the former can be linked to malfunctioning of bicycle whereas the latter can either be a simple test of a bike or a fail to find a new suitable bike because the rest are not adequate or because there are no other bikes available in the terminal.

The results of the algorithm depend on the values of X and Y. The former is the minimum rental duration for a record to be considered a real trip and not a bike trial or equivalently the maximum duration of a bike trial. The latter is the maximum interval between returning a bike and collecting a new one from the same terminal in the case of a journey with bike substitution. A possible approach to choosing X and Y is illustrated in the TusBic case. Figure 3.2 shows that COD trips present a peak of frequency for rental durations up to 5 minutes, whereas the distribution is rather uniform for longer rentals. As a consequence, it is assumed that the two cases represent different types of rentals: the former, bike trials; the latter, proper trips starting and ending at the same docking station. Hence, a value of 5 minutes is assumed for X. With this threshold, 1580 rentals are classified as bike trials, where 77.9% correspond to annual subscribers, 8.5% to weekly and 13.6% to daily subscribers. This share is similar to the distribution of trips in the sample (Figure 3.1a), which means that the percentage of bike trials compared to the total number of trips made by each of the three types of subscribers remains constant for all of them and results in the 5-6%.

On the other hand, Figure 3.4 shows the sensitivity of the number of Outcomes 1 (and, hence, of Outcomes 2) against the magnitude Y for $X=5$ min. Y is the time to choose the new bike and to book it using the automatic interface at the terminal. As Figure 3.4 reflects, the number of additional Outcomes 1 (trips with bike substitution) decreases sharply as Y increases so as to detect more than 95% of the cases for $Y=5$. However, as it can be seen in Figure 3.4, it is for Y equal to 14 minutes that the number of Outcomes 1

and 2 establishes, so *Y=13* minutes is considered the threshold for Santander. Although such interval may appear too long for a bike change in the case of commuting trips – when users can be in a hurry – or rentals made by experienced system users, it is not unreasonable for recreational trips and, above all, when users are not familiar with the system, this is, when more time may be needed to choose the new bike and deal with the interface. Furthermore, sometimes the registration of daily or weekly subscribers may take longer than average because the system fails in detecting the credit card where the fee is charged. This may cause queues to access the machine to register and delays in booking bicycles, so the value of *Y* should also consider this possibility to calibrate the pattern and detect these rentals.
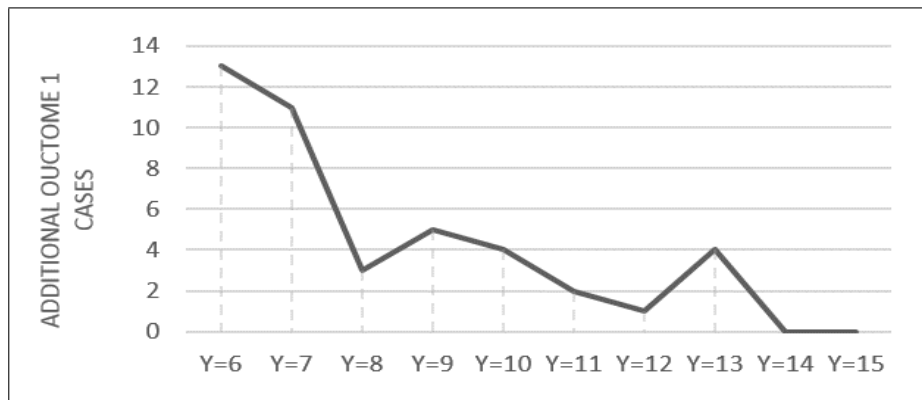
Figure 3.4: Sensitivity of cases detected as a bike trial with substitution

Table 3.2 informs on the distribution of Outcomes 1 and 2 detected according to the subscription.

Table 3.2: Number of Outcomes 1 and 2 cases regarding the type of subscription

	Annual	Weekly	Daily	Total
Outcome 1 (substitution)	994 (80.7%)	118 (88.1%)	185 (85.6%)	1,297 (82.1%)
Outcome 2 (no substitution)	237 (19.3%)	16 (11.9%)	30 (14.4%)	283 (17.9%)
Total bike trials	1,231	134	215	1,580

The first insight relates to the fact that, in the case of Santander, most of the bike trials (at least the 80%) detected with X=5 min can be considered first legs of trips with bike substitution. Furthermore, although the difference is not dramatic, it can be seen that Outcome 1 cases are more frequent for less regular users (weekly and daily subscribers). Such result can be explained by the fact that experienced users can be more able to

distinguish bicycles in good conditions. However, further analysis is needed because the result may also indicate that these users place less importance on the bike itself, or it may also be linked to the purpose of the journey.

In this case study, this pattern corresponds to 5% of the total records. Undoubtedly, considering them in the analysis of the demand leads to an impact in terms of inflated intra-zone demand and travel times. Furthermore, it should be highlighted that the transactions identified as bike trials provide interesting information, such as potentially deficient bikes that are being returned to the system in similar conditions. On the one hand, the agency responsible for the maintenance of the bikes would benefit from such information and the described process for detecting bike trials could be considered to optimise the operator's routing for terminal rebalancing purposes. On the other hand, the number of times that each bicycle has been substituted is proposed an indicator of the bike conditions and performance. Furthermore, this information is indeed an indirect observation of the perceived quality of a bike. Indeed, it was an attributed rated and modelled in the first stage of the research in this thesis, as described in Chapter 2. Consequently, the algorithm provides interesting insights that could be analysed in more depth or even used to support the modelling with survey data.

3.4 Overall demand indicators

In order for a BSS to be compared with other schemes in terms of the demand, a set of indicators are needed to give a general picture of the service. Therefore, some typically used indicators have been assessed for the case study of this research.

193 bicycles and 62 days are represented in the dataset, which means an average of 2 trips per day per bike. Recalling the studies by Fishman *et al.* (2013) and O'Brien *et al.* (2014), this value informs on the low demand for the BSS in Santander in comparison with other worldwide systems. Annual members show greater variation of the number of users per day due to the non-limitation of their access to the service but in average, one bike is rented a day per subscriber and the 90% of them do not exceed 17 rentals within the two summer months.

Relevant information is also obtained from the analysis of the demand profile along an average day, and even more interesting is to compare the demand profiles of the transport modes that coexist, as is proposed in Figure 3.5a.
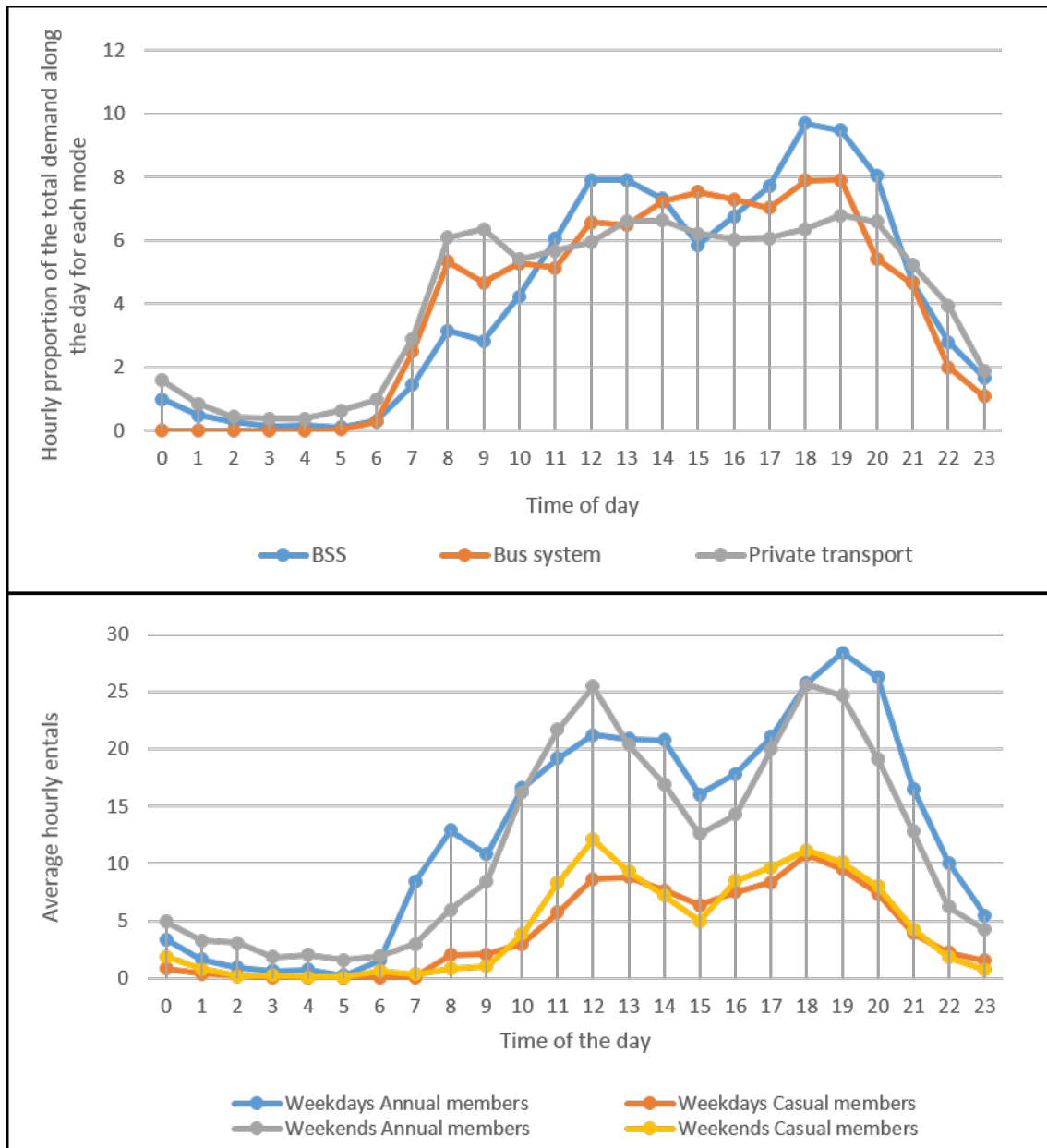
Figure 3.5: Transport demand profiles along the day in Santander: a) hourly proportion of the total demand for the BSS, the bus system and the private car; b) average bike rentals distinguishing the type of day and subscription

Various interesting insights are revealed by Figure 3.5. On the one hand, the comparison of the three modes leads to the conclusion that the peaks are accused in the case of the BSS but smother variations along the day can be found for the bus and the private car. It should be noted that the car experiences a peak at lunch time from 2 to 4 pm, in line with the Spanish customs. Although the BSSs in Barcelona and Zaragoza (Froehlich *et al.*, 2008; Kaltenbrunner *et al.*, 2010; O'Brien *et al.*, 2014) do reflect this same routine, in Santander, the lunch time effect corresponds to a significant fall in bike mobility as observed in Figure 3.5a, thereby indicating that cycling may be consider an activity itself for many rentals, in most cases taking place during the late morning and towards the afternoon-evening. However, an evident peak period in the morning from 7 to 9 am is identified by Figure 3.5, indicating the possible presence of commuting trips since it is simultaneous to the morning commuting peaks in the rest of massive modes in Santander, as previously reported by Alonso (2010) and represented in Figure 3.5a.

In a more detailed analysis of the demand for the BSS regarding the day and type of user, Figure 3.5b shows that the tendency is very similar in all cases with peak demands in the mid-morning and mid-afternoon. The difference comes from the trip rates associated to the type of subscription. However, it should be noted that a relative maximum is uncovered in Figure 3.5b: a morning peak period (7 to 9 am) among annual subscribers in weekdays that does not occur at weekends, suggesting the service may also be demanded for commuting.

3.5 A clustering method to determine travel patterns based on the heterogeneity existent in the usage behaviour

An algorithm is proposed that describes the behavioural framework within a bikeshare scheme through a set of conditions that classify rentals into systematic bike usage behaviours. These are a consequence of the particularities of a BSS for being a public and renting service such as its fare structure, the interval of time to rent under the flat fee, or the distribution of terminals, among other aspects.

The process is presented in Figure 3.6. Various attributes of every rental are examined that are recorded by the system in every smartcard transaction: the user ID, the rented bike ID, the terminals at origin and destination and their timestamps. The usage
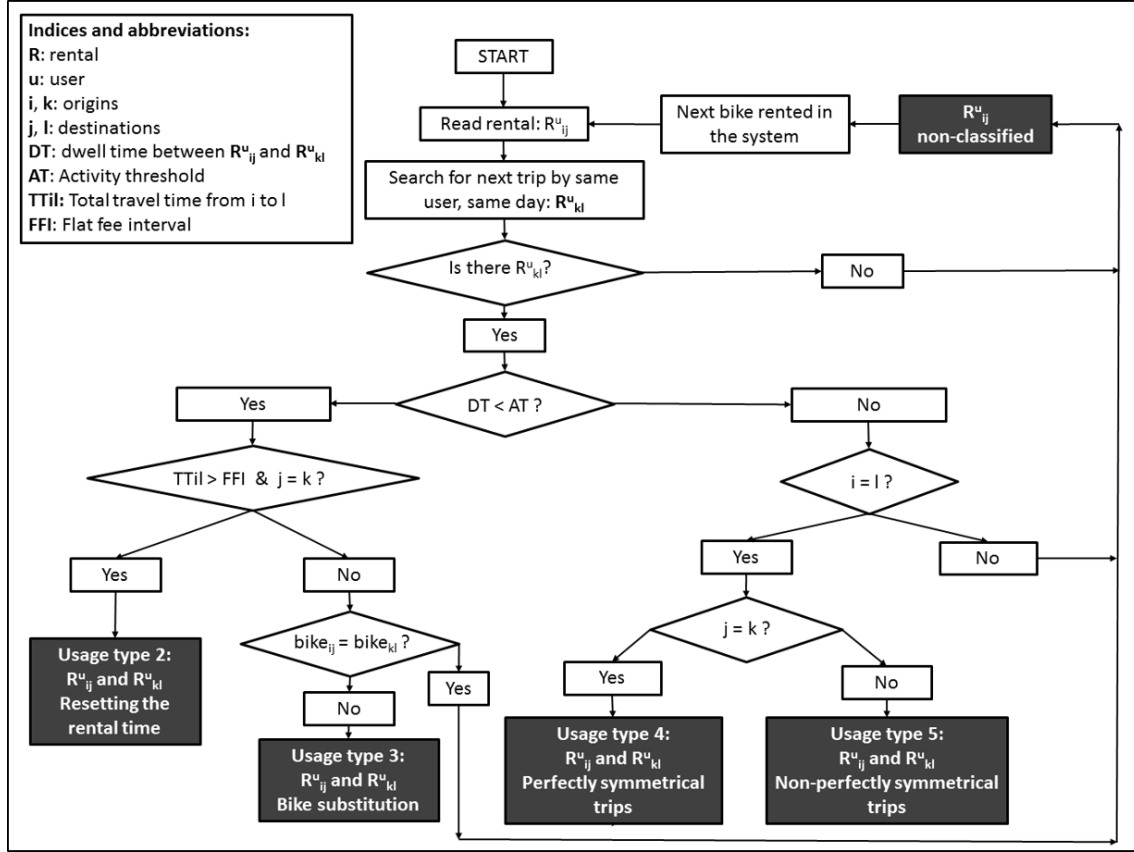
Figure 3.6: Algorithm to describe the BSS usage behavioural framework

behaviour is evaluated by relating each trip with the preceding and following one made by the same user within a same day. Only in the case of round trips and in the rentals that are eventually not classified in any of the defined circumstances in the algorithm, the trips are assumed isolated, not related to the previous or following bike rental. In the rest of cases, the sequence of usage patterns are evaluated considering the trips made within the same day by the same user.

Previously to the application of the method, a first condition evaluates whether the terminals at origin and destination of a trip are the same. This way, the rentals described as round trips are identified and classified into the first usage type.

As can be seen in Figure 3.6, the algorithm evaluates the following conditions in order to classify the rentals into usage patterns:

- The dwell time or elapsed time between the return of a bike and the next bike that is picked up within the same day by the same user.

  A time threshold (AT, activity threshold) is imposed in the data mining process

assumed to discriminate whether or not an activity has taken place in between the two bike rentals.

- The added travel time of two subsequent trips when a rental begins right after the previous bike has been dropped off.

  A travel time threshold is introduced intended to establish the difference between a bike substitution due to problems with the performance of the bike and the will to avoid being charged an extra fee. BSSs generally permit the use of the bike with no extra cost under what it is called a flat fee interval (FFI). Most worldwide BSSs impose a FFI of 30 minutes, and this value could be used as the threshold in the algorithm. In case a longer interval is permitted with no extra cost, then it is advised that the threshold ensures an accurate discrimination between the use of the bike for transport and a recreational ride. The literature has determined an average of 15-20 minutes and 2 to 5 kilometres up to a maximum of 7-8 kilometres depending on the travel patterns in the city (González *et al.*, 2015; Jensen *et al.*, 2010; Midgley, 2011; Pucher & Buehler, 2008), a threshold under which the bike is demanded as a mode for transport and above which the bike is expected to be used for recreation or physical exercise.

- The symmetry between two subsequent trips.

  Symmetric trips are especially linked to commuting and daily routines but can also be linked to any other activity. In any case, no matter the purpose at destination, the importance lays on the fact that bike trips classified as symmetric are expected to consist of a use of the bike for transport to get to where the activity will take place.

The behavioural casuistry considered in the process (Figure 3.6) is explained in detail below. Each usage behaviour identified through the algorithm reflects a different conception of the BSS as a service and thus, the travel patterns are expected to systematically vary between the different groups. Hence, the heterogeneity in travel patterns is expected to be reduced among the rentals classified into the same usage type in the algorithm: the terminals, time of day and travel time are assumed derived from the cycling desire and needs, which is assumed indirectly related with the service usage behaviour. On the contrary, the trips that are not classified through the algorithm are assumed isolated trips, not part

of a sequence and thus the greatest variability is expected to be encountered among the non-classified bike rentals since no extra information is provided by the data that could talk about a systematic travel behaviour.

In order to relate one trip with its subsequent or previous one, the trips made within the same day are evaluated through the algorithm in Figure 3.6. This way, activities may be assumed between subsequent trips occurred in a same day but overnight home stays are ignored. This way the algorithm is not restricted to big BSSs which are mainly used for commuting but also for BSSs open to casual users that may not perform trip-chaining routines.

3.5.1 Description of the behavioural framework

Usage type 1: Round trips

Round trips are the first to be gathered into a specific database. These records do not undergo the algorithm but are previously detected by the evaluation of the coincidence of the terminals at origin and destination.

Attending at their nature, round trips are expected to be related to mostly recreational rides and thereby, these trips would not have been made in any other motorized mode if a bike-share scheme was not available. Under this characterization, cycling for physical exercise is also a possibility. However, the above condition could also be related to the use of public bikes as a mode of transport to make a short activity such as a quick shopping, and the bike be locked somewhere in the street and be returned to the original station. Even users could assume the extra cost to make a longer activity and park the bike in the street where no docking stations are available.

In order to deeper understand this usage behaviour, the analysis of travel times, together with the most demanded terminals is recommended. The spatial dimension may further inform on whether the most demanded terminals are located in recreational zones or the bikes are mostly rented and returned to typical "home terminals" where the residential rates, hotels and accommodations are located. The distribution of travel times would enrich the analysis and determine the variability existent.

The analysis of round trips can be very insightful in decision making on the following aspects:

- the FFI if recreational rides are to be avoided;

- the redesign or extension of the BSS since, in the case of non-recreational round trips, such usage type could be pointing out the need to provide a more dense network to encourage dropping off the bike at a terminal for other users while an activity is taking place.

Usage type 2: Resetting the rental duration

A tricky usage type is identified when users require a bike for a longer period than the FFI and avoid paying an extra fee by resetting the rental time, this is, by returning the bike and immediately renting a new one or even the same. This is a systematic behaviour characterised by the following simultaneous conditions (Figure 3.6):

- the dwell time between two subsequent trips within a same day is shorter than a limited interval called "activity threshold" ($DT \leq AT$);

- the second bike (the same or different as the previous) is rented in the same terminal where the preceding was dropped off ($j = k$);

- the added travel time considering the two rentals is greater than the FFI, or any time interval over which the ride could be assumed a recreational activity itself ($TT_{il} \geq FFI$).

The dwell time to be shorter than the threshold implies that no activity has taken place in between the two trips. The third condition determines that the total travel time considering the two rentals is over the imposed flat rate interval, and thereby identifies a usage pattern that goes against the sharing essence of the BSS. In any case, instead of the FFI, any time could be defined as the threshold under which the bike can be expected to be potentially demanded as a mode for transport and above which the bike is expected to be used for recreation or physical exercise.

Note that the anxiety regarding the FFI could also lead to reset the rental duration even when the ride would not last longer than the flat fee interval; but the total travel time could still be greater than that evidenced in the literature for a utilitarian use of the bike (up to 20-30 minutes or 8 kilometres at the most).

It should also be noted that a first short rental would suggest a bike substitution for a failure more than the will to initialize the rental duration. However, the short length

of the first trip is difficult to define in minutes and a deeper research would be needed to accurately introduce and set that extra condition.

The analysis of the rentals identified under this type could provide interesting information to characterize this specific usage. As a first clarifying description of the subsample of rentals, it is recommended to determine the percentage of "pseudo round trips" where the origin of the first rental coincides with the destination of the second one. A recreational usage could be expected in these cases and the terminals demanded for bike changing could be characterised. The analysis of the spatial patterns would also inform on whether "home terminals" generate these trips or they are alternatively demanded where activities take place.

Usage type 3: Bike substitution

This systematic behaviour is characterised by the following simultaneous conditions (Figure 3.6):

- the dwell time between two subsequent trips within a same day is shorter than a limited interval called "activity threshold" ($DT \leq AT$);

- the second bike is not the same as the previous ($bike_{ij} \neq bike_{kl}$);

- the second bike is rented in the same terminal where the preceding was dropped off ($j = k$);

- the added travel time considering the two subsequent rentals is shorter than the FFI, or any time interval over which the ride could be assumed a recreational activity itself ($TT_{il} \leq FFI$).

Similarly to the usage type 2 (resetting the rental duration), this behaviour is also characterized by a limited dwell time between the first rental and the second, being assumed that no activity takes place in between both trips. The second and fourth conditions identify a bike change that implies the will to substitute the bike for a new one to finish the trip, maybe due to a problem with its performance. The total travel time of the two trips is shorter than the FFI or than any other time threshold that suggests the two trips form a unique trip and the bike is interpreted to be substituted for a better performance. Hence, it is assumed that this usage behaviour was not planned before the trip was generated but simply occurred that the bike needed to be replaced. Then it is expected that

this usage type does not explain a homogeneous travel behaviour since it is derived from an unexpected situation. In other words, in terms of travel patterns, the heterogeneity is expected to still be present across the trips in this group.

Interestingly, since the algorithm identifies the bikes being substituted, these could be revised by the operator and alternatively taken for repair. It would be highly recommended to analyse the number of times that the same bike has been returned in similar circumstances by a different users or in the case the bike has similarly being identified substituted in a bike trial through the data mining proposed in section 3.3, which should be applied right before this algorithm.

Usage type 4: Perfectly symmetrical trips

Symmetries are typical in mobility behaviour, especially in commuting trips. Indeed, this characteristic is the basic condition on which trip-chaining is based in the transport studies based on data collected by the ITS, as pointed in the literature review in section 3.2.

In the case of a BSS, this usage pattern is described by the following conditions simultaneously (Figure 3.6):

- the dwell time between two subsequent trips is longer than a limited interval called "activity threshold" ($DT \geq AT$);

- the origin of the first trip is the same as the destination of the second ($i = l$);

- the destination of the first trip is the same as the origin of the second ($j = k$).

This usage type is assumed to be related to activity-based transportation, so it is important to highlight that even when the activities are recreational such as attending social events, the bike is assumed rented as a mode for transport to get to the place where the activity will take place. This implies that if there was no BSS, other transport mode would be selected.

The analysis of travel times is expected to be in agreement with previous literature on the length and duration for which the bike is attractive as a mode for transport. Additionally, an analysis of the dwell time would be insightful in activity-based transport research if related to the spatial distribution of this usage type.

Usage type 5: Non-perfectly symmetrical trips

Non-perfectly symmetrical trips imply an activity in between the two rentals and the second bike being picked up in a different terminal from that where the first bike was dropped off. Expressed in an algorithmic form, this behaviour is identified with the following simultaneous conditions (Figure 3.6):

- the dwell time between two subsequent trips within a same day is longer than a limited interval called "activity threshold" ($DT \geq AT$);

- the origin of the first trip is the same as the destination of the second ($i = l$);

- the destination of the first trip is not the same as the origin of the second ($j \neq k$).

In comparison with the perfectly symmetrical behaviour, the further analysis to be applied in this case is the distance from the terminal where the first bike was dropped off and that where the second trip began. The user may have moved around on foot or by any other mode. However, the change in terminals may also suggest that it was not possible to return the first bike at a specific terminal due to capacity limitations and was eventually returned to the closest station or, alternatively, that a bike was required in the same previous terminal but was empty and the bike was eventually rented at another station.

3.5.2 Validation of the behavioural framework

The dataset for validation corresponds to the smartcard transactions experienced in the summer period of 2011 (July-August). As reasoned in section 1.4 of the introduction, this period is justified by the following facts:

- A previous research conducted in Santander assessed the negative influence of the weather on the mode choice towards the bike (dell'Olio *et al.*, 2011a). This explains that the presence of bikes sharply decreases during the rainiest months of the year, being the demand for the BSS in July and August a 30% of the yearly demand. As a consequence, summer is the time when the service can reach its capacity. In fact, the bus system experiences a peak of demand when both residents and tourists require the service.

- The summer time is when the existent variability is more accused since all types of trip purposes are expected due to the attraction of tourists and the chilly weather (which can also get rainy), at the same time that residents are still working but may also be on vacation. Therefore, travel patterns are expected to vary substantially across the population of bike rentals.

- Summer is when the weather conditions are the most favourable to use the bike for transport and it is thereby when the utilitarian uses of public bikes can be identified and characterised.

In short, the demand and variability within the BSS are at their highest values during the summer months in Santander and thus, these circumstances make it an ideal period to test and validate the developed data mining process.

As informed in section 1.4 in the introduction, except for the ordinary bike parking facilities, which are spread throughout the entire city, the rest of the infrastructure dedicated to cycling avoids the steep slopes and is therefore mostly located along the perimeter of the city. The city's main attractions are served by the BSS and bike lanes (Figure 3.7 and Figure 3.8): the city center (terminals 1, 11 and 13), the train station and that for regional and national buses (terminal 14), the main 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). Terminals 5 and 12 are located in mainly residential areas, as it also occurs with terminals 13, 14, 10, 6 and 7.

The fare structure and registering options are also key attributes of the system, as maintained by Midgley (2011). The BSS in Santander is offered to all public. Three types of subscription are available that give annual, a week or a day access to the system. Annual cards are worth €10, whereas week and day subscriptions cost €5 and €1 respectively. Furthermore, it should be mentioned that week and day subscribers (casual users) can register on the go as long as a credit card is provided at a terminal to charge the fee and also block an amount of €150 as deposit. On the contrary, annual users should first register themselves online.

Regarding the fare structure, another point that should be considered is the flat fee interval that BSSs impose which is intended to ensure that bikes are actually "shared" and not conceived as one's own bike. Any bike rental does not cost extra money within
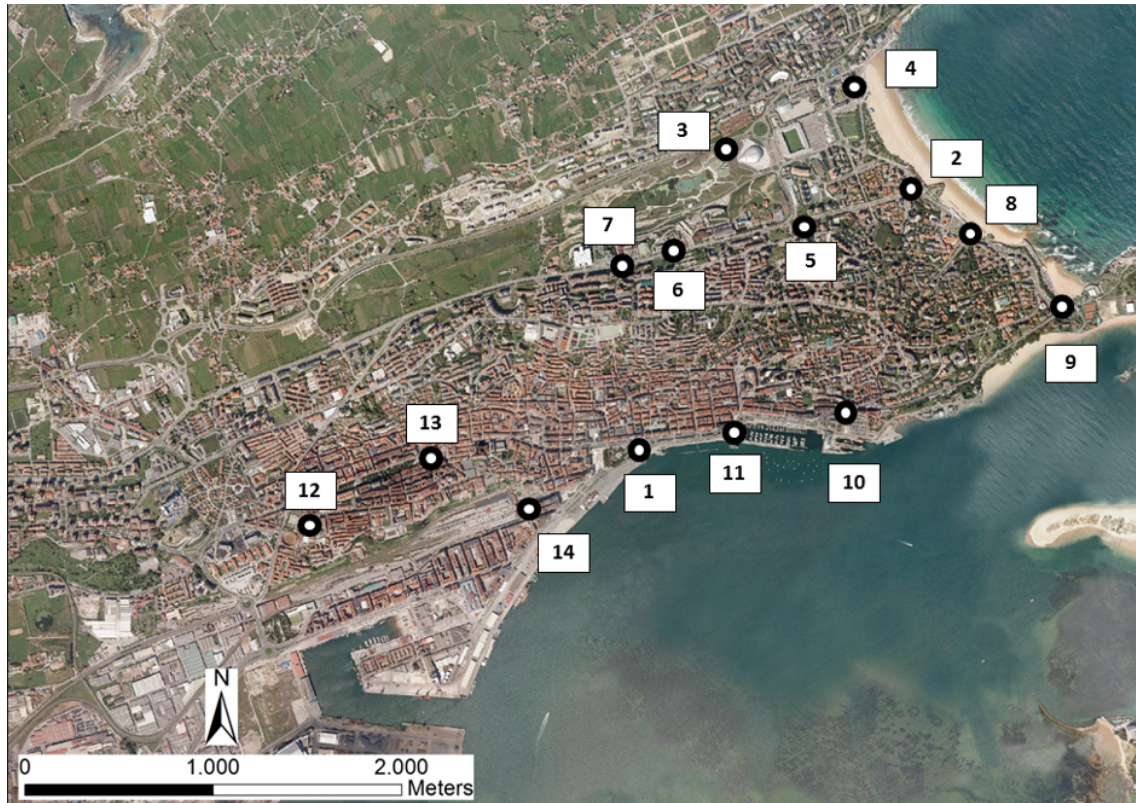
Figure 3.7: Localization of the terminals of TusBic system in Santander

Terminal	Capacity (docks)	Bike lane	Beach	Panoramic view	Commercial area	City centre	University	Residential area	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				✓	✓		✓	✓

Figure 3.8: Description of the facilities and land use in the surroundings of the terminals

that interval of time (in addition to the registration fee), which is generally 30 minutes. In the case of Santander, up to 60 minutes of use are permitted before charging extra money for every half an hour.

Assumptions for the case of Santander

All things considered, the following assumptions have been made in the present case study to apply the data mining process previously described:

- The activity threshold (AT) assumed for two subsequent rentals has been set in 15 minutes. This value is based on the result obtained in section 3.3, where it was found that even when the second bike is generally picked up a very few minutes after the preceding one had been returned, it is up to 15 minutes that the pattern holds for Santander. Hence, in this case study it is considered that if a second bike is rented in the same terminal within the next 15 minutes after a bike has been dropped off, no activity has taken place; then it would be just a bike substitution or a will to reset the rental time in order not to pay an extra fee.

- Instead of the actual flat fee interval, which is established in 60 minutes in this case study, 40 minutes have been considered instead. This value has been established, first assuming that longer bike uses are recreational trips based on the literature (Dill & Gliebe, 2008), and second, recalling the survey data collected in the first stage of the thesis (Chapter 2) and the results in Table 2.1, where the intercepted leisure trips lasted 34 minutes in average. Therefore, 40 minutes is a conservative threshold that thereby guarantees discriminating between a substitution of the bike and the will to initialize the rental time again to avoid being charged extra money in a leisure ride.

It should also be mentioned that in this case study, a day is assumed from 6 am to 6 am of the next day. This assumption is made based on the algorithm, which identifies the sequence of trips within a same day and so it is assumed the very first trips of the day take place after 6 am and, therefore, this makes it possible to identify return trips from a night out event as long as the bike is rented before 6 am.

Considering the previous values, a 47% of rentals have been classified into any of the five usage types described in the algorithm proposed in this research. This value is

obviously characteristic to each BSS and informs on how the system is conceived, and thereby used. As previously mentioned, the non-classified records are isolated bike rentals in the sense that they are not related with any preceding or subsequent trip in any of the forms hypothesized in the data mining described in Figure 3.6. In other BSSs, for instance in cities where the bike is routinely used or greatly demanded for commuting, a greater percentage of rentals could be expected to be classified through the data mining process since commuting trip-chaining is easier to characterise as perfectly and non-perfectly symmetrical behaviour.

Figure 3.9 presents the classification of the rentals that have been identified via the algorithm presented in Figure 3.6, discriminating those made by registered users (Figure 3.9a) from those by casual users (Figure 3.9b). As Figure 3.9 shows, registered and casual users tend to perform different usage types. Almost half of the uses by casual subscribers are round trips whereas annual card-holders show a similar representation of round and perfectly symmetrical trips. Furthermore, registered subscribers are shown to substitute the bike more often. These differences complement the research by Lathia *et al.* (2012), in this case the comparison being made in terms of usage behaviour or types. However, the usage behaviour characterized with the proposed algorithm is assumed to explain the travel patterns, as in this case study.

The operator should place importance in the second and third usage types (resetting the rental time and bike substitution) since that demand is related to the will to reset the rental time or to substitute the bike. The first behaviour 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.

Although the classification presented in Figure 3.9 are reasonable and in line with those that were expected, further validation is presented in the following subsections through binary probit models and based on the rental duration and spatial analysis.

Validation through binary probit models

The validation is required to confirm that the hypotheses and conditions used in the data mining process do reduce the heterogeneity among rentals. With this aim, binary probit models have been estimated in order to confirm the different impact of the variables
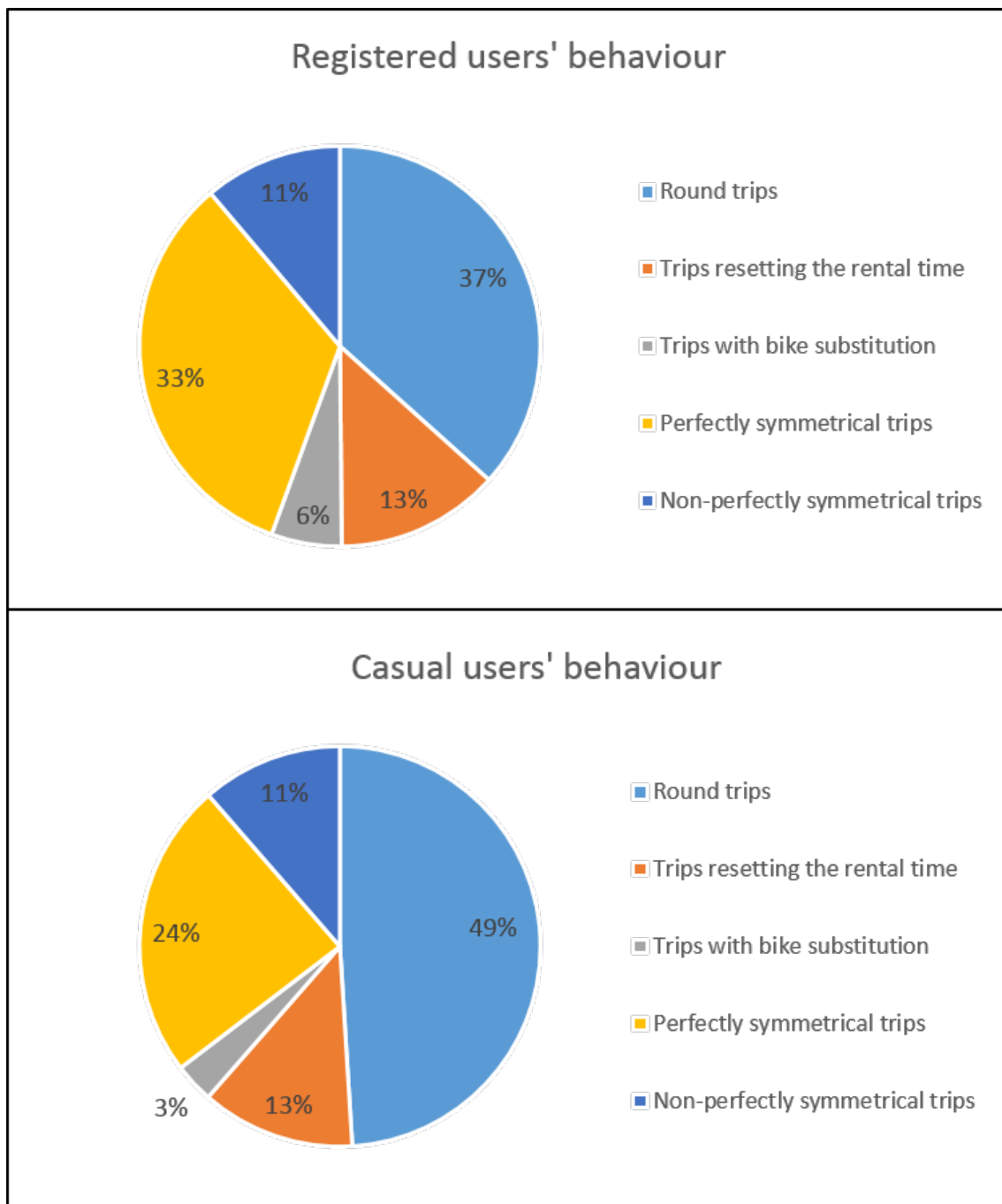
Figure 3.9: Distribution of usage behaviours in Santander: a) registered users; b) casual users

describing the travel patterns (travel time and spatial distribution) on the utility of each usage type. This way, it could be accepted that the travel behaviour is significantly different for each usage behaviour and thus the clusters are justified and the algorithm in figure 3.6 is a useful tool to be applied previously to the demand analysis.

Table 3.3: Binary probit models to validate round trips and resetting rentals

Model 1 - Round trips			
	Coefficient	Standard error	T-test
Annual subscriber	-0.0869	0.0301	-2.89
Rental duration	0.0070	0.0003	24.05
Peak period	-0.5404	0.0717	-7.54
Origin/Destination residential	-0.9936	0.0391	-25.39
Origin/Destination bike lane	0.1759	0.0333	5.29
Origin/Destination capacity	0.0166	0.0018	9.19
Log-likelihood function: -4863.51			
Model 2 - Resetting rentals			
	Coefficient	Standard error	T-test
Annual subscriber	-0.8799	0.2695	-3.27
Rental duration	0.2098	0.0209	10.06
Peak period	-1.4577	0.6137	-2.38
Origin1 residential	-0.4683	0.3212	-1.46
Origin1 bike lane	-1.0325	0.3132	-3.3
Origin1 capacity	-0.0291	0.0157	-1.85
Destination1 residential	-0.1184	0.2998	-0.4
Destination1 bike lane	-2.7148	0.4451	-6.1
Destination1 capacity	-0.0610	0.0175	-3.48
Destination2 residential	0.3373	0.1878	1.8
Destination2 bike lane	0.1841	0.1789	1.03
Destination2 capacity	-0.0781	0.0159	-4.91
Log-likelihood function: -84.68			

Table 3.4: Binary probit models to validate bike substitutions and perfectly symmetrical rentals

Model 3 - Bike substitutions			
	Coefficient	Standard error	T-test
Annual subscriber	1.1700	0.3233	3.62
Rental duration	-0.2613	0.0284	-9.19
Peak period	1.5406	0.6859	2.25
Origin1 residential	0.1274	0.3483	0.37
Origin1 bike lane	1.1129	0.3596	3.09
Origin1 capacity	0.0347	0.0176	1.98
Destination1 residential	0.1974	0.3367	0.59
Destination1 bike lane	2.9386	0.5104	5.76
Destination1 capacity	0.0692	0.0205	3.38
Destination2 residential	1.1915	0.3698	3.22
Destination2 bike lane	1.4509	0.3684	3.94
Destination2 capacity	0.0418	0.0185	2.26
Log-likelihood function: -67.79			
Model 4 - Perfectly symmetrical rentals			
	Coefficient	Standard error	T-test
Annual subscriber	-0.0679	0.0381	-1.78
Rental duration	-0.0193	0.0006	-30.99
Peak period	0.6807	0.0693	9.82
Origin1 residential	0.8303	0.0508	16.34
Origin1 bike lane	-0.1465	0.0470	-3.12
Origin1 capacity	0.0003	0.0024	0.11
Destination1 residential	-0.1475	0.0477	-3.09
Destination1 bike lane	-0.1118	0.0534	-2.09
Destination1 capacity	-0.0136	0.0024	-5.67
Log-likelihood function: -3367.35			

Table 3.5: Binary probit model to validate non-perfectly symmetrical rentals

Model 5 - Non-perfectly symmetrical rentals			
	Coefficient	Standard error	T-test
Annual subscriber	-0.2201	0.0994	-2.21
Rental duration	0.0027	0.0020	1.36
Peak period	0.4633	0.1940	2.39
Origin1 residential	0.2979	0.1190	2.5
Origin1 bike lane	0.0952	0.0946	1.01
Origin1 capacity	0.0121	0.0055	2.21
Destination1 residential	-0.2840	0.1086	-2.61
Destination1 bike lane	0.0813	0.1486	0.55
Destination1 capacity	0.0106	0.0067	1.58
Origin2 residential	0.0141	0.0894	0.16
Origin2 bike lane	0.0063	0.0930	0.07
Origin2 capacity	-0.0103	0.0064	-1.6
Log-likelihood function: -583.32			

The data set for the binary models was built as follows: in each model (for instance model 1 - Table 3.3, describing the choice towards a round trip), the choice variable values 1 if the algorithm has classified the rental as a round trip, 0 otherwise. This way, each model determines how the choice of the usage behaviour (in this case, round trips) is explained by the set of variables describing the travel patterns in terms of rental duration and spatial distribution, as well as other aspects such as the type of subscription and time of the day. The travel time is summed up for the two subsequent rentals in two usage types: when the trip duration is reset (model 2 - Table 3.3) and in bike substitutions (model 3 - Table 3.4) since they are indeed comprised of two subsequent rentals showing a limited time lapse in between them. Indexes 1 and 2 for origin and destination refer to the first and second trip-legs detected in the usage behaviours described in models 2 to 5.

As can be concluded from Table 3.3, 3.4, and 3.5, the BSS usage behaviour is differently explained by the land use and bike facilities at origin and destination of the trip-legs, by the rental duration, the type of subscription and the period of the day.

Interestingly, peak periods (7 to 9 am and 1 to 4 pm in weekdays) increase the

probability of a rental to be either a symmetrical behaviour (model 4 - Table 3.4) or a bike substitution (model 3 - Table 3.4). Furthermore, except in model 5 (in which case the parameter is not even significant - Table 3.5), the impact of the travel time or rental duration is inverse to that of the peak period, confirming the pattern that peak periods are related to reduced travel times and symmetric behaviour. This usage type is also explained to be produced in terminals located at residential areas, whereas round trips and those were the rental time has been reset show lower probability to be generated in residential zones, suggesting a tendency to demand the BSS for recreation.

Hence, the binary probit models presented confirm that the travel patterns differently explain the usage of the BSS, thereby validating the potential of the data mining method developed to cluster the smartcard transactions into homogeneous groups of demand.

Validation based on the spatial analysis and rental duration

As the models in Table 3.3, 3.4, and 3.5 show, the facilities and land use in the surroundings of the terminals at origin and destination of the trip-legs do explain the choice of the use of the BSS. In order to visualize how the usage of the BSS informs on the spatial preferences of a trip, the spatially distributed trip productions are presented in Figure 3.10 for each usage type.

Figure 3.10 confirms that spatial patterns are actually heterogeneous across different usage behaviours. For the sake of simplicity, the bike substitution case is here represented together with the non-classified trips since the two groups are expected to be highly heterogeneous in terms of travel behaviour due to the absence of a planned behaviour, as previously reasoned. In addition, it should be mentioned that only those trips presenting a subsequent rental (this is, the first trip-leg) are represented in Figure 3.10 in the case of resetting the rental time, and perfectly and non-perfectly symmetrical types. The reason for this is that the trip is assumed moved by the activity at destination and the origin of the following trip-leg is obviously dependent on the destination of the previous, so the travel information in the three usage types mentioned is basically provided by the trips with a subsequent one that has been detected with the algorithm.

The non-classified trips together with the trips where the bike has been substituted later on present smoother variations in the demand across terminals, this is, all of them
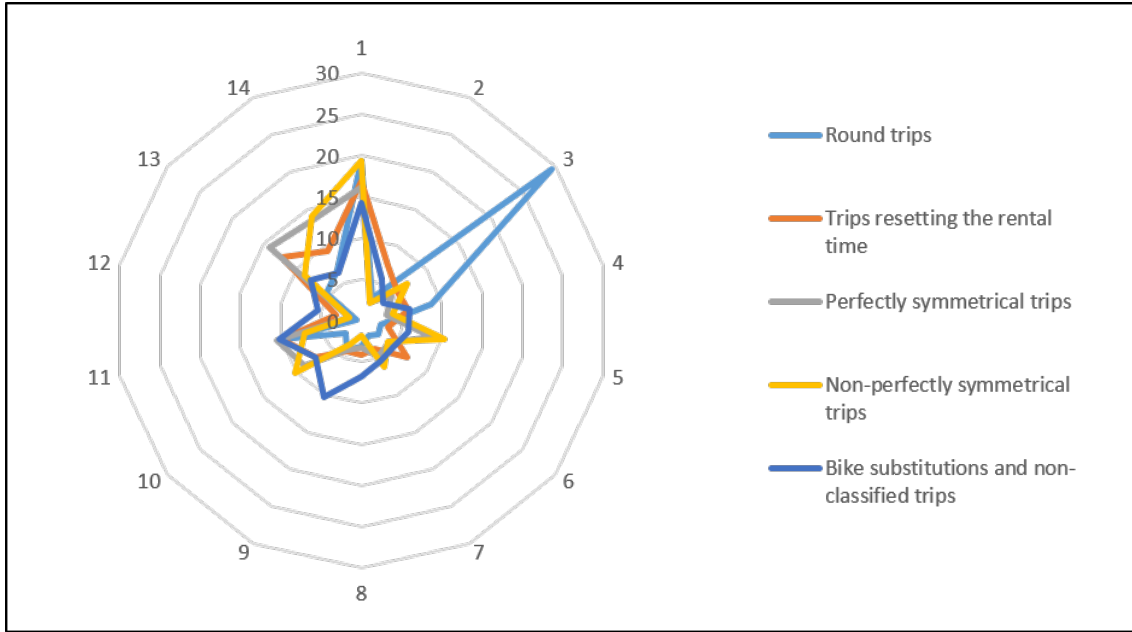
Figure 3.10: Spatial demand at origins according to the usage behaviour (percentage demand within the same usage behaviour)

are almost equally demanded. On the contrary, the rest of the usage types, those described by the predefined conditions in the algorithm do show a clear preference for some of the terminals whereas the demand for others is scarce. These results support the hypothesis that the classified usage types show greater homogeneity in the spatial demand at origins than in the case of isolated rentals that do not follow any of the predefined types of behaviour.

The previous results call for a first attempt to classify the terminals in terms of home-based and activity-based. Symmetric trips (both perfectly and non-perfectly symmetric) will be considered to classify the spatial demand in terms of residential and activity-based. For this, the results in Figure 3.10 are compared with the information in Figure 3.8. The graph identifies the following most demanded terminals for symmetric behaviour: 1, 5, 10, 11, 13 and 14. These terminals are all located in residential areas, as informed by Figure 3.8, which supports this thesis. This finding also suggests that the distance to the household is determinant to the use of public bikes, as previously obtained for Santander in Chapter 2 and by dell'Olio *et al.* (2011a), and in various other case studies (Fishman *et al.*, 2013; García-Palomares *et al.*, 2012), together with the greater impact of a more dense network claimed by Maldonado-Hinarejos *et al.* (2014). Furthermore, the spatial demand

at origins is similar for the types of resetting the rental time and symmetric behaviour, meaning it could be expected that these usage types are mainly home-based generated.

On the contrary, the least demanded terminals at the origin of a symmetric behaviour are 2, 3, 4, 6, 7, 8, 9 and 12. Indeed, as informed by Figure 3.8, stations 2, 3, 4, 8 and 9 are located in recreational open areas with panoramic views and maybe a beach close. Stations 6 and 7 are residential but also serve the university campus, so it appears they may be more demanded as destination zones but not as origins of outward routes. Finally, the low demand for terminal 12 is not explained by its attraction as an activity zone but, recalling the literature (Akar & Clifton, 2010; Maldonado-Hinarejos *et al.*, 2014; Pucher *et al.*, 2010), it is interpreted affected by two circumstances: its isolated situation within the BSS network (Figure 3.7), and the lack of a bike lane (Figure 3.8), which ends kilometres away from its location.

It should also be noted that terminal 1 is the greatest demanded in any case whilst the most popular terminal for round trips is number 3. As indicated in Figure 3.7, and Figure 3.8, terminal 1 is characterized by the greatest capacity for bikes since and its high demand is expected due to its location in the city center, attractive to all possible trip purposes: home, business, shopping, restaurants and night life. It also boasts one of the most famous panoramic views of the Santander Bay. The most surprising result is the demand for terminal 3. However, when the land use is analysed, the finding is not unexpected. Such terminal is located in the extreme of the biggest park in Santander, a spot with no shops nor restaurants but an open and semi-natural area with lakes which also provides firm paths for skating and cycling. Furthermore, there is a free parking lot right next to the docking stations and people usually park their cars and do exercise or just wander around. Since round trips are mostly generated at this spot, it can be concluded that cycling in this park is actually the activity, the purpose of these trips and that, as expected, round trips are mainly related to trips where cycling is conceived an activity itself.

Finally, the variability in the travel time is analysed so as to confirm that this factor indeed causes heterogeneity among the demand for the BSS. Figure 3.11 presents box plots of the rental duration of each behaviour identified through the proposed algorithm.

The travel time is summed up for the two subsequent rentals in the case that the rental duration is reset and in bike substitutions. Trips above 500 minutes have been
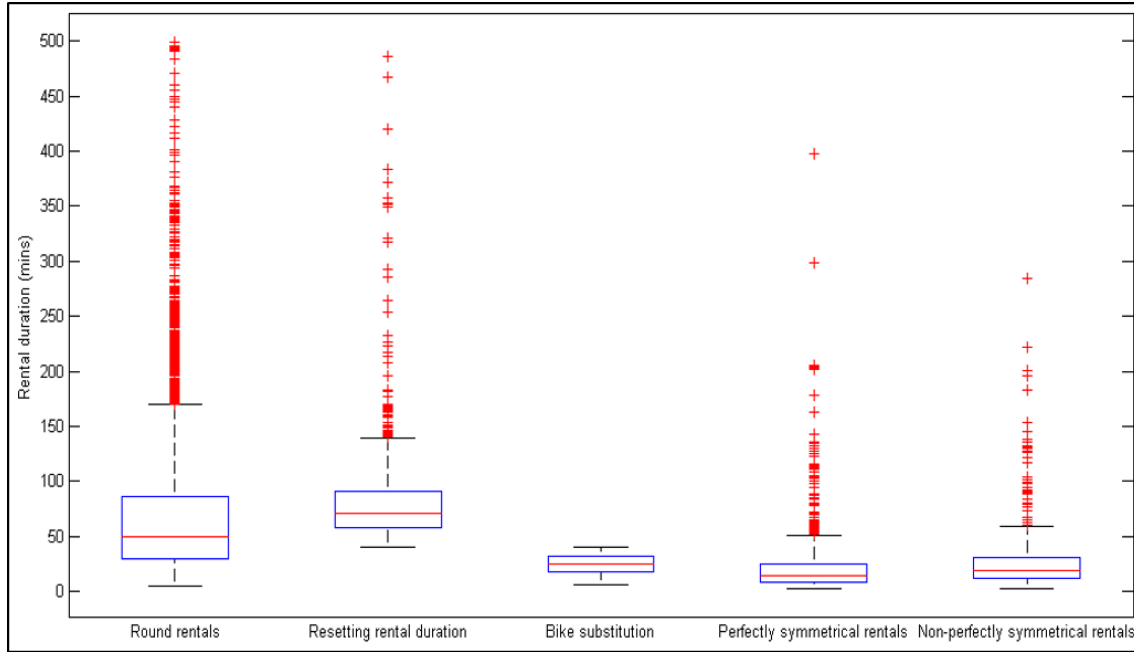
Figure 3.11: Rental duration according to the usage behaviour

omitted and considered outliers.

Figure reveals the travel time varies depending on the usage type. Analogously, Table 2.1 in Chapter 2 uncovered the travel time differs according to the trip purpose, being the greatest in the case of cycling for leisure purposes and the lowest if it is a commuting ride. Considering the two outcomes, round trips and those resetting the rental time could be associated to recreational cycling, whilst the travel time in perfectly symmetrical trips can be compared with bike riding for transport purposes.

As a consequence, the presented models and analyses validate the developed algorithm: the hypotheses assumed in the algorithm are suitable to classify rentals into groups performing homogeneous travel patterns.

Insights into the travel patterns in Santander derived from the observed bike usage

It is interesting to analyse each usage type in detail with the aim to characterise their travel behaviour and to better identify the implications of such a different conception of the BSS in each case. Therefore, once the methodology has been validated, the results for the case study are presented for each usage type based on various analyses as well as the binary models above presented.

Round trips

Approximately 41% of all the classified rentals have been identified under this label for the case of Santander. As can be seen in Figure 3.9, the percentage is substantially deviated among rentals made by casual users (50%) compared to those made by annual card-holders (38%). Together with previously highlighted evidences, this fact supports the hypothesis that the round nature of these trips is likely to relate to recreational rides more than cycling for transport since the share for round trips is substantially greater among casual users than among registered subscribers.

In order to further understand this usage in terms of travel demand, Figure 3.10 presents the travel times registered by users making round trips.

The highest occurrence of round trips show around 40 minutes of travel time. 86% of round trips last longer than 20 minutes, the average travel found in other systems (Dill & Gliebe, 2008; Jensen *et al.*, 2010) characterized by a routine use of public bikes so the fact that a high proportion last longer suggest the use of the bike for leisure rides. It needs to be remarked that 75% of trips last longer than 40 and the 38% even overcome the flat fee interval of 60 minutes. On top of the recreational hypothesis for round trips, the interpretation of results together with the average of 5 km and maximum of 7-8 kilometres up to which the bike can be attractive for transport (Midgley, 2011), indicate that the threshold of 40 minutes for the condition of the added travel time is justified adequate in this case study instead of applying the system's FFI of 60 minutes.

Interestingly, 75% of the round trips intercepted in the first stage of this thesis focused on the quality perceived by the users (Chapter 2) were made for a leisure ride, as responded by interviewees. Furthermore, as informed by Table 2.1 leisure trips are associated with the greatest travel time of all trip purposes: an average of 34 minutes, very close to the aforementioned values. Therefore, round trips can be assumed mostly recreational.

Trips resetting the rental duration

The values specified for the activity threshold (AT=15 minutes) and the added travel time (40 minutes instead of the FFI) lead to a total of 6.2% of rentals to be classified into the type of resetting the rental duration.

As explained in the description of the algorithm and the usage type in section 3,

this use of the BSS is hypothesized to be related to the demand for public bikes as a leisure activity or physical exercise and identifies a tricky use of the service against its sharing nature. In fact, two circumstances directly confirm the type is actually to reset the rental time. Firstly, the 40% of the cases show the same bike is actually rented in the two subsequent rentals, meaning it is for sure that the bike was not substituted and therefore confirming the purpose of initialize the rental time again. Secondly, the 34% of all the identified cases are perfectly symmetrical trips, which means the second bike was rented right after the previous one was returned and the user has ended up at the same terminal where the preceding trip started. This pattern describes a pseudo round trip where the travel time was long enough to be cautious to reset the rental by changing the bike or even renting the same one immediately after it was returned. Finally, the 12% of all identified cases simultaneously show the previously mentioned conditions: the same bike being rented in the two trips and these to be perfectly symmetrical. Even when these two circumstances are not true, it does not mean the will to reset the rental time does not hold true in those cases; it means it is not possible to be 100% confirmed due to data limitations.

In addition, the travel patterns describing this usage type are analysed. The average total travel time considering subsequent rentals is 90 minutes with a standard deviation of 70, so substantially long rides are mostly expected, at least more than 20 minutes that is found in average in other studies for cycling for transport (Dill & Gliebe, 2008; Jensen *et al.*, 2010; Pucher & Buehler, 2008).

In terms of spatial distribution, Figure 3.12 shows the different preferences for terminals at the origin of the first trip and those selected for resetting the rental time.

The most demanded origins in the first rental were already represented in Figure 3.10 and are similar to those enumerated for symmetrical trips, this is, terminals located in mainly 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 Figure 3.8 and Figure 3.7). This result also support the hypothesis that these trips are mainly leisure rides and the bike is returned but rented again only to avoid paying extra fee.
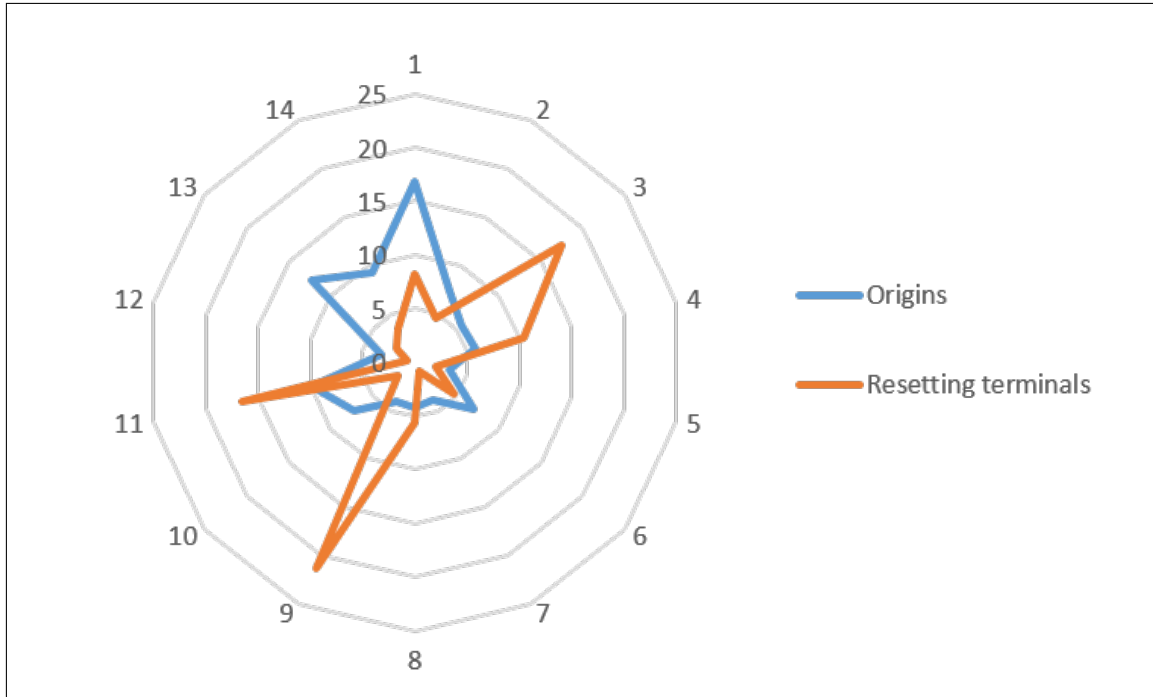
Figure 3.12: Spatial distribution of trips resetting the rental duration (percentage demand within this usage type)

Bike substitution

As pointed in the description of the algorithm, the interest in the analysis of this usage type lays on the identification of the bikes that are not performing correctly and may need to be repaired. The quality perception of the user is of utmost importance. A simple data representation provides interesting information. For instance, Figure 3.13 shows the number of bikes that have been replaced for a new one according to the times they have been substituted.

Obviously, the first bikes to be checked should be those that have been returned to the system the most. The operator may localize these bikes and optimise the redistribution route by considering this information.

The rentals interrupted by a bike substitution are considered as one trip in the following analysis. As already reasoned, this usage type is expected to lack of clear travel patterns. The spatial distribution of origins and destinations are presented in Figure 3.14.

Peak demands are shown for some terminals in Figure 3.12 but it should be noted that except for number 10 and 12, all the rest are substantially represented either as origins or as destinations. In addition, it should be noted that terminal 12 was the least

Figure 3.13: Number of bikes according to the number of times that they have been replaced by another bike

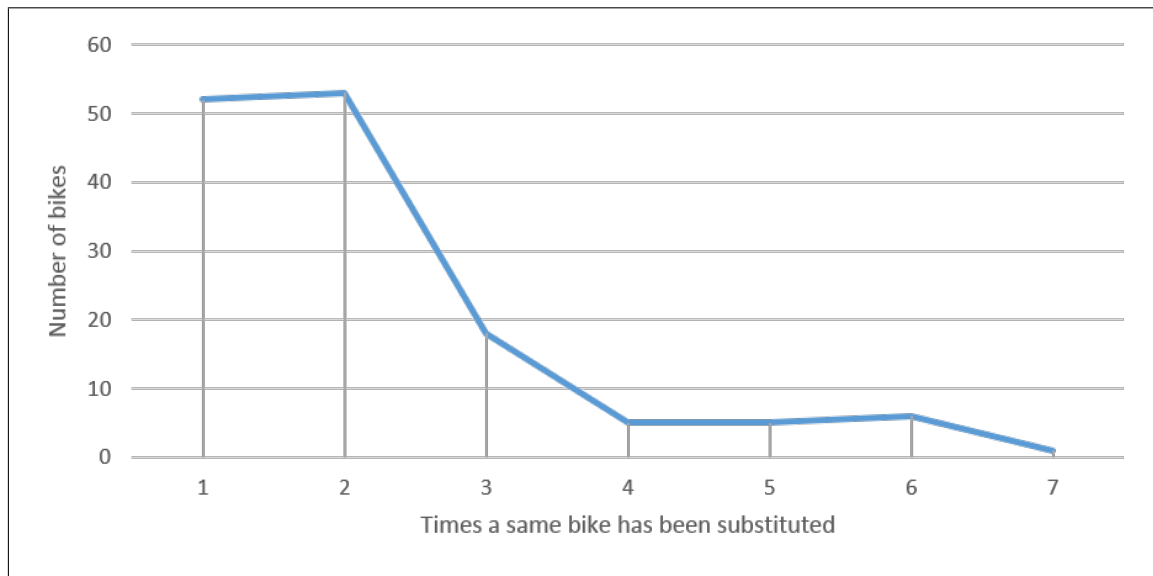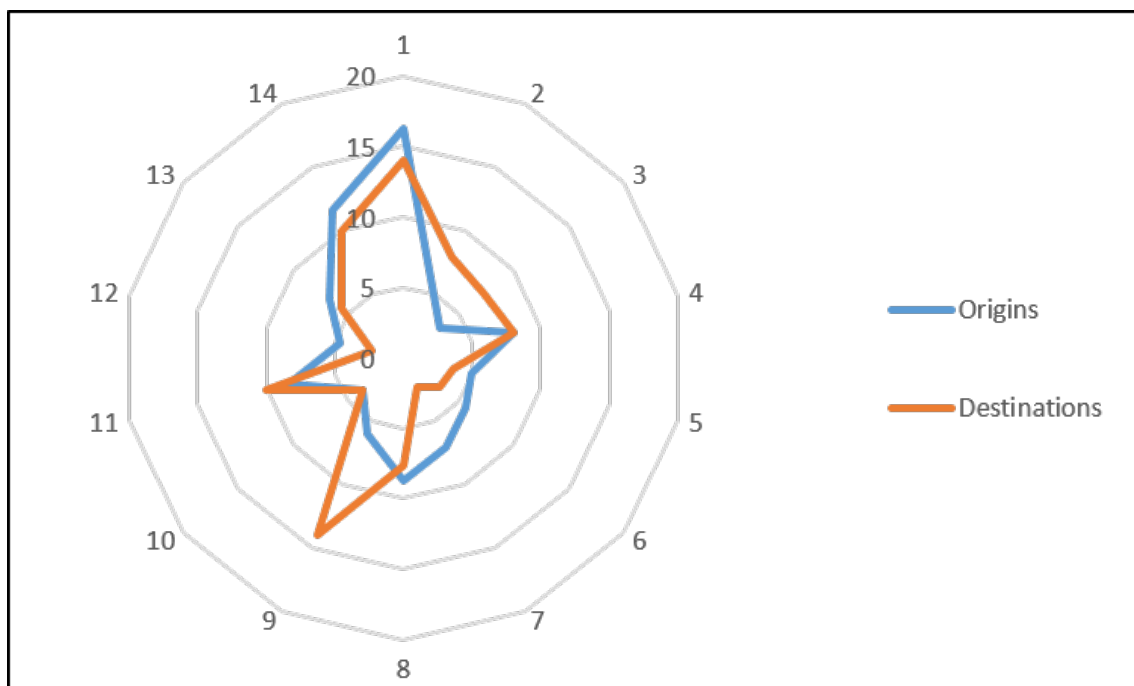Figure 3.14: Spatial distribution of bike substitutions (percentage demand within this usage type)

demanded in all usage types presented in Figure 3.10 so this result is not unexpected.

The average travel time of these trips is 24.7 minutes, presenting a standard deviation of 9.8 minutes. Therefore, most of these trips could be considered made with an actual need for transportation since the average travel time of cycling for transport is said to be 15-20 minutes or 2.5 kilometres (Jensen *et al.*, 2010; Pucher & Buehler, 2008) with a maximum range between 5 to 8 kilometres. Indeed, this result is supported by that in Chapter 2: users on a journey shorter than 25 minutes place more importance on the quality of the bike, a result that suggest that users demanding the service for transport (limited travel times) are more critic with the quality of the bike. The interviewers in such research on users experiencing a bad performance of their bike and changing for a new one at the first terminal they come across on their way. When reported, problems were related with the seats of the bikes failing to stay at a determined height.

Perfectly symmetrical trips

Perfectly symmetrical trips are mainly represented among registered users as informed by Figure 3.9, a result suggesting that annual subscribers demand the service for activity-based transportation more than casual users do.

The average travel time considering each leg of perfectly symmetric rentals is 20 minutes with a standard deviation of 27 minutes with positive asymmetric coefficient. The average is in agreement with the travel times obtained in other cities such as Lyon (Jensen *et al.*, 2010) or Santiago of Chile (González *et al.*, 2015).

As above analysed, the demand at origin of a symmetric behaviour is suggested majorly home-based in Figure 3.10. Figure 3.15 compares that with the spatial demand at destination, this is, where the activity takes place.

As expected, the destinations substantially differ from the origins. Activities mostly take place within the area of influence of terminals 9, 1, 8, 4, 13, 2, 11 and 14 and these zones are consistent with the localization of activities in the city (Figure 3.8 and Figure 3.7). Furthermore, symmetric patterns show considerably greater demand for terminal 14 than other usage types (Figure 3.10). This docking station is located next to the train station, a result that contributes to the literature addressing the potential of BSS to enhance the transit network (González *et al.*, 2015; Jäppinen *et al.*, 2013; Martens, 2004, 2007; Pucher & Buehler, 2008, 2009). It is also a densely populated area in the
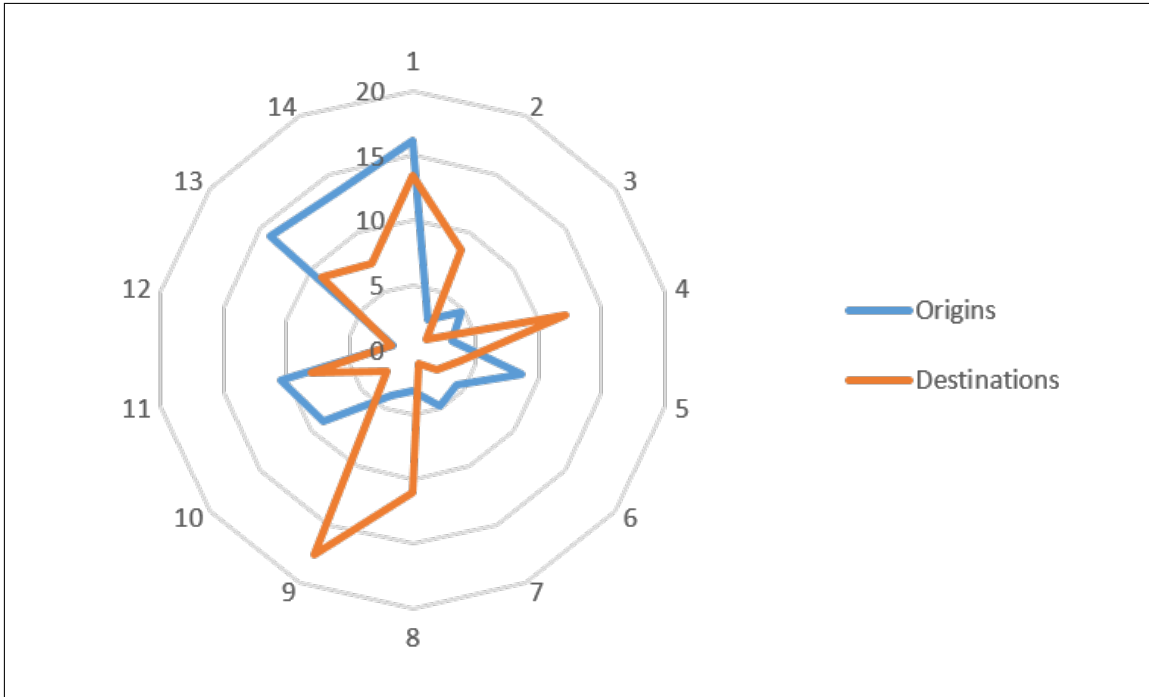
Figure 3.15: Spatial distribution of perfectly symmetric behaviour (percentage demand within this usage type)

city center that is not generally the interest of visitors and is not connected with any bike lane, as informed by Figure 3.7) and Figure 3.8. The lack of bike lane seems not to influence the production of trips, thereby presuming the bikes are more being rented as a mode for transportation than a tool for leisure, if considered the literature evidences on the smaller influence of pro-bike infrastructure among experienced cyclists choosing the bike for transport (Akar & Clifton, 2010; Dill, 2009; Dill & Gliebe, 2008).

The time in between two subsequent and perfectly symmetrical rentals, which is assumed the duration of the activity, goes from 16 minutes to 13 hours with an average of around 3 hours.

Non-perfectly symmetrical trips

Analogously as in the case of perfectly symmetrical trips, non-perfectly rentals are spatially represented in Figure 3.16. The profiles are similar to those in Figure 3.15 and although in this case the destination of the outward trips does not coincide with the origin of the return trips, their very similar shapes indicate that the same set of terminals can be interpreted to be activity-oriented: 1, 2, 4, 8, 9 and 11. From the land use around these terminals
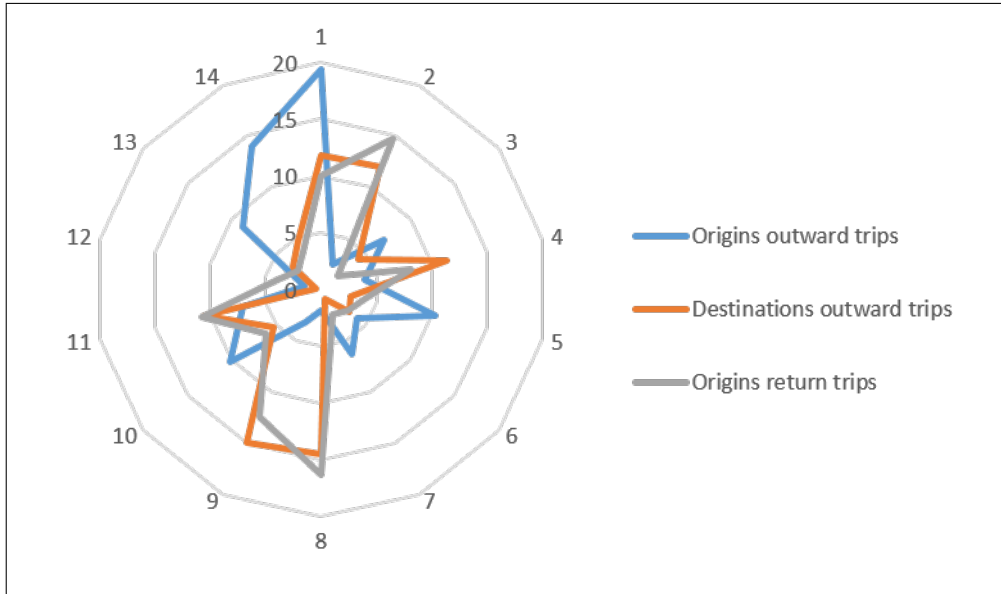
Figure 3.16: Spatial distribution of non-perfectly symmetric behaviour (percentage demand within this usage type)

(Figure 3.8 and Figure 3.7), it can be concluded that, in general, leisure activities are those demanding the BSS as a transport mode in this case study.

Also in line with the perfectly symmetric behaviour, the range of time between two subsequent rentals goes from 16 minutes to 10 hours with an average of around 3 hours. In terms of distances, the average distance between the destination of a trip and the origin of the subsequent is 990 meters and a standard deviation of 580 meters. Hence, it could be expected that these short distances are overcome on foot but it also raises the question whether the users in some of the identified non- perfectly symmetrical cases wanted to behave symmetrically but could not due to capacity circumstances. Indeed, some surveyors in the research in Chapter 2 on the quality perceived by users reported users approaching full stations and then forced to search for a place to drop the bike off in the closest terminal. These situations were encountered at midday in sunny days and terminals located along the coastline.

3.6   Conclusions

Extracting valuable insights from records generated by the ITS requires both data mining skills and some knowledge on the analysed system. The possibilities to analyse and model

the data directly depend on the nature and quality of the recorded information; but even if the automatically recorded variables are limited, it should be highlighted that it is the entire population that is gathered by the ITS, thereby avoiding the sampling bias inherent to traditional manually collected data.

From the first analysis of raw data, a non-ignorable pattern has been uncovered named "bike trial with substitution", which is made up of two trip-legs: an initial short rental in which a first bike is tested and then returned to the docking station being probably faulty, followed by the rental of a new bicycle which allows carrying out the initially planned journey. Failing to remove the first trip-leg (this is, the bike trials) from the data set to analyse the demand of the system would lead to inflated intra-zonal demand and also reduces the average travel time. Thus, it is important to apply the data mining to assess the presence of this kind of rentals before further use of the database and posterior modelling.

The detection of these specific rentals with bike substitution can provide valuable information regarding:

- The service quality and user's satisfaction: the return of a bike to be substituted by a new one indicates the user does not find the quality of the bike as expected to make the trip. Thus, the detection of these rentals provides interesting information to support the study of the service's perceived quality approached in Chapter 2.

- The management and maintenance of the fleet of bikes: since the ITS record the bike of each rental, it is possible to identify the bikes that have been returned more frequently and so to arrange for physical checks and reparation more effectively. Therefore, the application of the data mining process to the automatic records provide useful insights to the operator to improve the performance and to optimise the routing to rebalance bikes at terminals.

The second data mining process developed permits to deeply understand travel patterns based on objective usage behaviour. The time-lapse between consecutive rentals, together with the spatial trip-chaining are key conditions to classify rentals into less heterogeneous demands. As a result, the proposed algorithm could be considered a clustering technique that permits characterizing the predominant travel patterns. The developed data mining process has been validated, implying that the hypothesis assumed do identify

usage patterns that show systematic heterogeneity in travel patterns. Hence, the proposed methodology permits to deeply understand travel patterns based on automatically collected data.

The method aimed to facilitate the analysis of the demand for public bikes both in the case of large and small systems. As the population of rentals is represented by the applied data, sample size limitations are overcome, thereby permitting the identification of less usual mobility patterns such as bike substitutions that would not have been obtained through traditional collection techniques. Nevertheless, it is for certain that the main shortcoming of this type of data is failing to know the trip purpose. Therefore, results should be read with caution and be based on ingenious analyses together with related scientific literature; these ingredients have permitted an insightful discussion of the results obtained in the case study adopted, assisting in the understanding of the usage and travel behaviour.

In particular, this approach detects tricky practices: when a user does not behave as expected in a sharing service but his behaviour is still permitted under the normative of the BSS. Indeed, two specific usage types identified in the proposed data mining process are particularly insightful for operators: bike substitutions and the will to reset the rental time. A bike substitution may be informing of a bad performance of the bike, a circumstance which influences the users' overall perception of quality, as concluded in Chapter 2. Since the bike IDs are recorded, they can be spotted and considered in the bike redistribution route to check them. On the other hand, the second behaviour might compromise the capacity of the system as it goes against the sharing nature of the service. The times and conditions under which the flat fee interval is overcome or when subsequent rentals are made to avoid paying extra money are identified. Their analysis is the key for an efficient management and very valuable information to improve and redesign the service. In conclusion, even though these systems limit the use of the bikes, it is possible to violate such a sharing conception without being charged for it. The proposed methodology provides the knowledge to develop incentives that could avoid such circumstances.

The two possible conceptions of cycling (for transport and as an activity itself) can also be identified through the proposed algorithm. As Heinen *et al.* (2010) state, the behaviour is necessarily different in the two cases. The type of subscription has also been shown to yield a different picture of usage behaviour, which eventually influences the

travel patterns. These differences complement the research by Lathia *et al.* (2012), but in this case the comparison being made in terms of usage types. However, this research has evidenced the usage behaviour can be used to characterize the demand for bike-sharing in terms of travel patterns. Additionally, the discussion on the results obtained with the application have indirectly inferred the results that would be expected in any other type of BSS considered in the classification proposed by O'Brien *et al.* (2014). For instance other BSSs showing a routinely use of bikes being greatly demanded for commuting should expect a greater presence of symmetric trips. Hence, 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.

All things considered, insightful conclusions are drawn from the application of the two data mining processes proposed that have been demonstrated beneficial to all stakeholders: operators, decision makers, and eventually citizens, once the system is optimised.

Chapter 4

A weather and time-based approach to understand the trip production

4.1 Introduction

Once the demand for a bike-sharing system has been characterised, models can be estimated and outcomes interpreted based on the previously uncovered insights. The two sources of information have been analysed (manually collected data - Chapter 2 -, and automatically collected data - Chapter 3), providing very valuable information to the modeller. After applying the two algorithms presented in the previous chapter, the data set is qualified to undergo modelling procedures. As a result, this chapter presents a regression approach to determine the impact of the weather and time-based variables on the trip production at terminals predominantly demanded for leisure and those experimenting a predominant demand for the BSS for transport purposes. This last stage of the research is built upon two assumptions derived from the outcomes in the previous phases: first, that the travel time and the quality perceptions differ based on the trip purpose (Chapter 2), and second, that the demand for a terminal and the travel behaviour are associated to the land use and main activities in the surroundings of the bike station (Chapter 3).

It is said that one of the major barriers to cycling is unfavourable weather. Such a negative effect has been noticed even in cultures where cycling is extended as is the case of the Netherlands (Rietveld & Daniel, 2004) and Sweden (Bergström & Magnusson, 2003)

for instance. The literature on the variables that stimulate cycling is extensive but the particular impact of the meteorological conditions is scarce yet. The present research aims to address what Böcker *et al.* (2013) and Nankervis (1999) suggest efforts are needed: the knowledge on the weather impact on everyday transport decisions and more specifically, on the demand for bike-sharing. Consequently, the present chapter takes advantage of the automated data from a bike-sharing system and from meteorological sensors in order to determine such an influence on the demand for recreational and utilitarian bike-sharing. Apart from further understanding of what drives cycling as a mode of transport and for a leisure ride, the modelling will allow predicting the bikeshare demand and, thereby opti-mising the system and bike relocation tasks. Additionally, not only bike-sharing managers can benefit from the results in this last stage of this research but also managers of the alternative modes when the weather is adverse.

The methodology is applied to the bike-sharing system in Santander in order to complete the demand analysis of the system and specifically determine the effect of weather and time-based trip conditions in the users' behaviour. As an extension to the study by dell'Olio *et al.* (2011a), the present research, focuses on the effect that meteorological conditions and temporal trip circumstances have on the bikeshare demand. Dell'Olio *et al.* (2011a) identified the rainy and humid weather typical in Santander to affect the mode choice towards cycling together with the available infrastructure. This previous research was conducted when the bikeshare scheme was still not implemented and so the interest is now placed on assessing the weather effect assumed a bike-sharing infrastructure is provided. Regression models are proposed to explain the bikeshare trip production experienced in terminals mainly demanded in leisure rides and recreational trip purposes, separately from the rental activity registered by the stations that are mainly linked to non-recreational trip purposes.

Particularly, the modelling is aimed at understanding the different impact of the predefined conditions on the bikeshare demand for recreational and utilitarian purposes. In terms of data, such a distinction is made on the classification of terminals based on the conclusions drawn in Chapter 3. The spatial distribution of the terminals across the city and the land use in their surroundings have been evidenced to be related to bikeshare trip patterns and, eventually, purposes since the spatial distribution of activities is also associated to the land use and services. Therefore, the approach attempts to measure the

different extent to which the weather impacts on the trip production in those terminals which are mainly used for recreation cycling and those which are related to non-recreational trip trips, as concluded in 3. The weather and time-based effect on the demand is thus separately modelled as the trip production of each type of terminals. With this premise, the demand for bikeshare riding is regressed on factors such as temperature, humidity and rain as well as time-based trip conditions: the morning and afternoon peak periods and the day being Sunday or holiday.

The following points introduce the main contributions of the present research:

- The weather effect on bikeshare demand is separately studied based on two differentiated purposes: recreational and utilitarian.

- Count regression models are proposed: Poisson and negative binomial; these are hardly represented in the literature to estimate trip production.

- Previous conclusions from the literature have been considered in this study as indicated in the description of the modelling setup, ensuring that the research is complete and up-to-date.

The present chapter is structured as follows: the literature is reviewed on the first place to provide the state of knowledge on the field; the methodology is presented in section 4.3, containing the modelling fundamentals followed by the description of the application to the case study, the bike-sharing system in Santander. Section 4.4 presents the estimated models and a discussion is provided on the results. Finally, section 4.5 highlights the conclusions drawn from this research, suggesting further work in this field.

4.2 Previous research on the meteorological and time influence on cycling

Various approaches and case studies can be found in the literature that have measured the impact of weather conditions on travel behaviour. It is of great interest the review by Böcker *et al.* (2013) , who inform that the majority of studies linking weather variables with the transport sector are concerned on service performance and accidents whereas the literature on individual behaviour is less extensive. They also point out that although the interest is generally placed on the impact of extreme weather conditions, the everyday

conditions and short-term effects on travel decisions still needs to be studied since the literature is scarce in this regard.

Among the existent transport alternatives, active modes are the ones that experience the greatest impact caused by weather; especially cycling has been found much influenced by the meteorology being unsheltered against wind or rain. From the literature concerned on this individual mode it can be concluded that temperature and rain are the main drivers of cycling in terms of meteorology, although the wind speed and humidity have also been identified to influence the use of bikes. Precipitation is generally a barrier whereas temperature has been found to stimulate cycling except for the case of extreme temperatures, that can also act as a deterrent Böcker *et al.* (2013). Obviously, the specific research methodology together with cultural and lifestyle circumstances lead to slight differences in the behaviour towards cycling and the influence of weather conditions. For instance, Dill and Carr (2003) found that the rain was non-significant on commuting trips although the sign was negative in their study, which analysed 35 large cities in the USA. Similarly, Cervero and Duncan (2003) found no significant effect of rainfall on cycling rates in the San Francisco Bay Area. On the other side, temperature tend to positively increase the utility of cycling except under extreme values.

Böcker *et al.* (2013) remark that previous studies suggest precipitation affects departure times. Somehow in line is the lagged effect that raining cause on the decision to ride a bike, as uncovered by Miranda-Moreno and Nosal (2011), which can have important implications in terms of mode choice, being cycling rates influenced by the weather conditions in the previous hours.

In any case, the literature suggests that weather conditions may be perceived differently depending on the trip purpose. However, little is known in this field. The majority of studies focus on a specific type of cycling. Commuting and recreational cycling have been examined but also university campuses have been the interest of some research.

Winters *et al.* (2007) differentiate cycling among students from the general population using multilevel regression models. They found that precipitation and very cold temperatures were linked to lower cycling activity in general whereas students were less determined by climatic conditions, mostly by cold temperatures but not by rain. Students were also analysed by Nankervis (1999), who directed efforts towards understanding the effect of seasonal climate and weather conditions such as extreme temperature (daily

highest expected value), rain and wind on cycling. To this aim, daily counts of parked bikes were examined based on the daily weather. The author's main conclusion is that the climatological effect on students is smoother than expected and that they do not give up cycling easily although adverse conditions do impact student cycling negatively.

Among the literature on commuting cycling, the main interest has been placed on determining the modal shift caused by weather conditions. The research by Saneinejad *et al.* (2012) contributes to understand the relative importance that non-captive commuters place on temperature, rain, wind speed and the sky conditions (clear or cloudy) in the mode choice. Their research also determines the weather influence according to age and gender, surprisingly finding that young people are the most affected by temperature. Their models suggest that commuters shift from cycling to walking under unfavourable weather. In terms of trip generation, their results suggest that people tend to postpone trips when the weather is bad.

In a more specific research, Bergström and Magnusson (2003) focused on the potential shift from car to bike during winter in Sweden. The authors determined an important seasonal mode change caused by temperature and precipitation and identified the snow clearance to be essential to promote cycling in winter. A tendency to change mode from bikeshare to metro under adverse weather conditions was evidenced in the research by Gebhart and Noland (2014). They applied their approach to the bike-sharing system in Washington DC and obtained that the sensitivity to rain was greater in those bike terminals closer to metro stations.

Brandenburg *et al.* (2007) did propose the comparison of cycling purposes as an original contribution and found greater sensitivity of recreational cycling to weather. The authors intercepted cyclists in the main access points of a recreation area in Vienna and modelled daily manual data from interviews through simple and multiple linear regression specifications in order to understand the affection of climatological variables. The authors based their study on air temperature, precipitation and a scale of human weather comfort called the "physiologically equivalent temperature" (PET) and developed by Matzarakis *et al.* (1999). Thomas *et al.* (2012) also directed their research to determine the differences between recreational and utilitarian cycling and found that the amount of sunshine was the weather attribute that better distinguishes both cases. In line with Brandenburg *et al.* (2007), they also obtained higher sensitivity for recreational cycling. More

recently, Faghih-Imani *et al.* (2014) also applied automated data from a bike-sharing system, Montreal in this case. They modelled the minutely station activity in the 410 terminals (arrivals and departures) through linear regressions in order to test the effect of weather, temporal characteristics, bicycle infrastructure, land use and built environment attributes. The authors determined that the weather conditions affect bike pick-ups more than returns, an interesting whilst expected result.

This study proposes an updated approach that extends the research by dell'Olio *et al.* (2011a), which was applied to the same city, Santander (Spain) when the bike-sharing system was not implemented yet. The authors identified the rainy and humid weather, typical in the North of Spain, to affect the mode choice towards cycling. The research was conducted based on data collected through travel diaries and stated preferences and evaluated the increase on the share for cycling as a result of both weather conditions and infrastructure provision. The study revealed that the weather is the major driver for potential bikers, more than the provision of bike facilities. The interest of the present research is thus to determine the impact of meteorological conditions on bikeshare trip production, once the BSS infrastructure has been provided.

This chapter of the thesis aims to contribute to the scientific knowledge on this field both in terms of methodology and results. For this purpose, the study takes into account the previously reviewed literature and addresses the weather effect on the trip production discriminating for recreational and non-recreational trip purposes.

The conclusions and suggestions in the existent literature have been considered in the present research. Furthermore, some of the identified limitations from previous studies are overcome. For instance, the research proposed by Brandenburg *et al.* (2007), also concerned the different effect of weather on recreational and utilitarian cycling, but daily manual data from interviews were modelled through simple and multiple linear regression specifications whilst smartcard and hourly transactions are applied in this case, consisting of an important step further. Hence, the research ensures an up-to-date and accurate methodology with the original contributions enumerated in the previous section.

4.3 Methodology

4.3.1 Data collection and analysis

The aim of this research is to evaluate the impact that the weather conditions and temporal circumstances separately have on the trip production the terminals comprising the BSS, discriminating those mainly demanded for recreational purposes and those for non-recreational trip purposes. In order to achieve this aim, the modelling uses hourly data since modelling at a day level would not permit to evaluate the direct effect of weather conditions on the decision to rent a bike; it would only inform on general effects and tendencies regarding the trip production. In fact, the difficulty in measuring the effect on behaviour is even more accused in the case of casual activities, this is, recreational rides and shopping activities which are usually spontaneously decided. Hence, the hourly trips registered from 7 am to 10 pm along the months of July and August, 2011 have been used to determine how various meteorological and temporal factors impact on the demand for bike-sharing in Santander. The selected time window seeks to delimit the research on bike renting while other transport services are provided and when activities take place throughout the city.

It should be mentioned that this chapter considers the results and conclusions in the previous one as a starting point, but now the emphasis is placed on understanding the different impact that the weather and time-based trip conditions have on the demand for bike-sharing for leisure purposes compared to bike-sharing as a transport mode. Such distinction is based on the terminals registered at origin and destination of the rentals. Chapter 3 concluded that the demand for the BSS is explained according to the facilities and land use in the surroundings of the terminals. For instance, residential areas are negatively associated with longer bike rentals suggesting recreational rides, whereas symmetrical trip patterns are associated with shorter trips and residential origins, suggesting a predominant use of the public bike for transport. Therefore, the assessment of the effect of the weather and time-based variables on the recreational and non-recreational demand for the BSS is proposed in terms of the trip production registered in each group of terminals.

Figure 4.1 reflects the classification of terminals assumed based on the results in the previous chapter.

It should be noted that terminal 1 has been considered both a recreational and
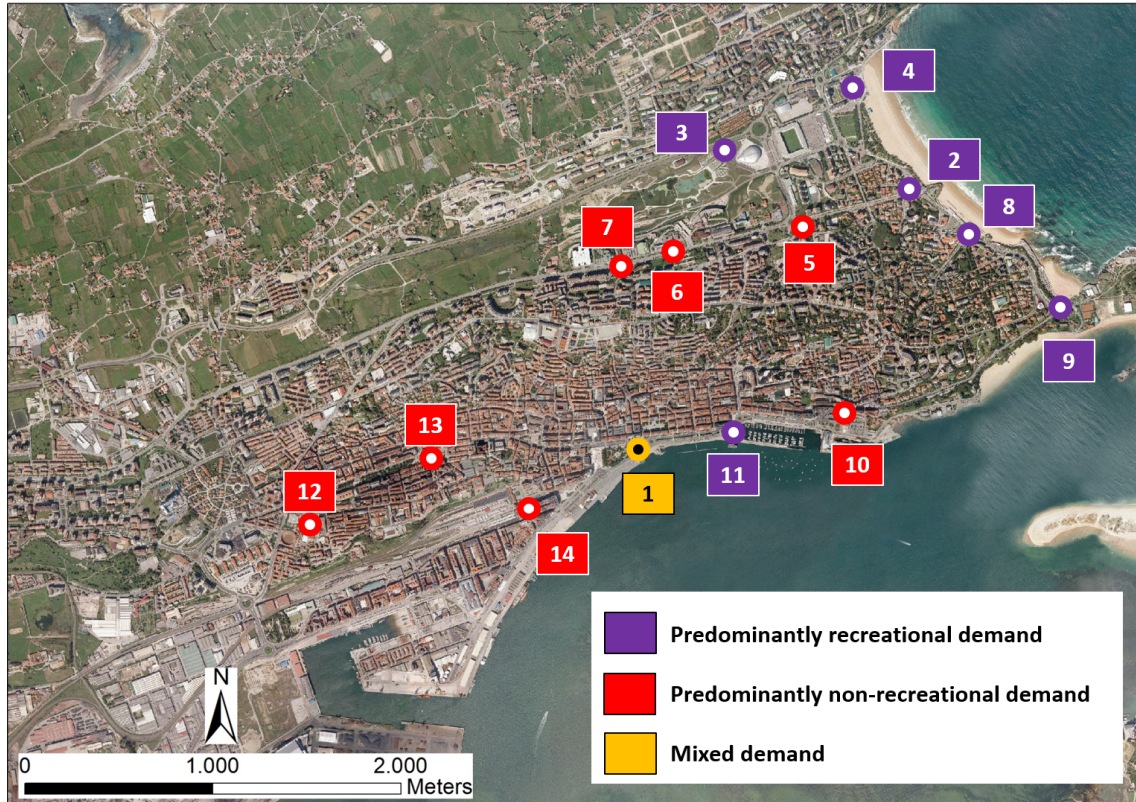
Figure 4.1: Classification of terminals according to their predominantly recreational and non-recreational demand

non-recreational station since all usage types demand this station, as revealed in Chapter 3.

The recorded bike rentals have been gathered into the following groups:

- Recreational trips: the bike has been rented in and returned to a recreational terminal.

- Non-recreational trips: the bike has been rented in and returned to a non-recreational terminal.

- Mixed trips: a recreational and a non-recreational terminal have been demanded either at origin or destination.

In order to avoid ambiguity, only the first two sets of trips are modelled to determine the factors affecting the demand for the two groups of terminals separately. The two endogenous variables that will be modelled are statistically described in Table 4.1: the

trip production (total hourly rentals) in terminals experiencing predominantly recreational and non-recreational activity.

Table 4.1: Descriptive statistics of the endogenous variables (hourly rentals)

Variable	Mean	Std. Dev.	Min.	Max.	Skewness	Kurtosis
Recreational rentals	11.70	8.58	0	54	0.73	3.49
Non-recreational rentals	4.98	3.22	0	16	0.62	3.08

The following sensor data on weather variables are automatically collected and provided by the Regional Government of Cantabria:

- Wind speed (m/s)

- Temperature (º C)

- Relative humidity (%)

- Rainfall (l/$m²$)

The collected data is presented in Table 4.2.

Table 4.2: Descriptive statistics of weather variables (hourly data)

Variable	Mean	Std. Dev.	Min.	Max.	Skewness.	Kurtosis
Wind speed (m/s)	1.86	0.99	0.30	6.70	1.41	5.85
Temperature (ºC)	22.11	2.99	14.40	32.50	0.28	2.73
Relative humidity (%)	73.15	13.37	35.20	100	-0.10	2.35
Rainfall (l/$m²$)	0.14	0.79	0	9.20	8.07	75.42

In addition, the "weather inertia" is tested in the models through a binary variable that values 1 in the case that any rainfall has been registered in a predefined interval of time and 0 if not. As mentioned earlier, the lagged effect of precipitation on cycling was confirmed by Miranda-Moreno and Nosal (2011) and in this case, the purpose is to determine whether rainfall affects the demand for bike-sharing in the forthcoming hours and if the impact differs depending on the trip purpose. The explanation of the weather impact on the demand for bike-sharing is expected to gain reality and to be improved considering this circumstance, although most of the times in the literature the precipitation is introduced merely as the measurement within the time interval when a trip occurs.

The meteorological variables have been introduced in their continuous definition but also in the form of binary variables.Temperatures above 20, 25, 28 and 30 °C and relative humidity levels over 70%, 80% and 90% have been tested. The binary variables represent whether or not these thresholds have been overcome and allow to consider the likely non-linear influence of weather factors remarked by Böcker *et al.* (2013). The summer meteorology in Santander already limits the values in the lower bound as can be seen in Table 4.2, therefore binary variables were defined for representing extreme low weather conditions for temperature and humidity.

Chapter 3 also identified particular temporal patterns when the bike is a tool for leisure compared to when it is rented for transport. Such patterns were described in terms of peak and off-peak periods and these conditions are considered explanatory factors of the demand through the following binary variables that represent whether the bike has been rented within any of the defined time windows:

- MPP: morning peak period, from 7 to 9 am of weekdays

- APP: afternoon peak period, from 13 to 16 h of weekdays

- OPP: rest of weekdays and weekends

In other words, in the modelling, the previous periods are hypothesized as trip conditions that may explain the growth or decrease of the demand within the same time intervals. It should be mentioned that the presented periods, which are used throughout the whole thesis, are based on the city's global mobility, whereas the demand for bike-sharing shows an inverse profile along the day, as shown in Figure 4.2. This means the MPP and APP periods are not maximums in the demand for bike-sharing

An average of 365 rentals per day were registered in the period of study, distributed as presented in the graphic in Figure 4.2. As revealed by the graphic, the peak periods for bike-sharing occur precisely during the off-peak periods of the general mobility in the city (OPP), this is, at midday and evening. Nevertheless, as revealed by the demand profile in Figure 4.2, the MPP period accounts not only for the mobility in the usual modes but also for the bike-sharing since a relative maximum is identified.

Considering that the mobility and city life is particularly different on Sundays and holidays as a result of the non provision of services, this effect is also introduced in the models through a binary regressor.
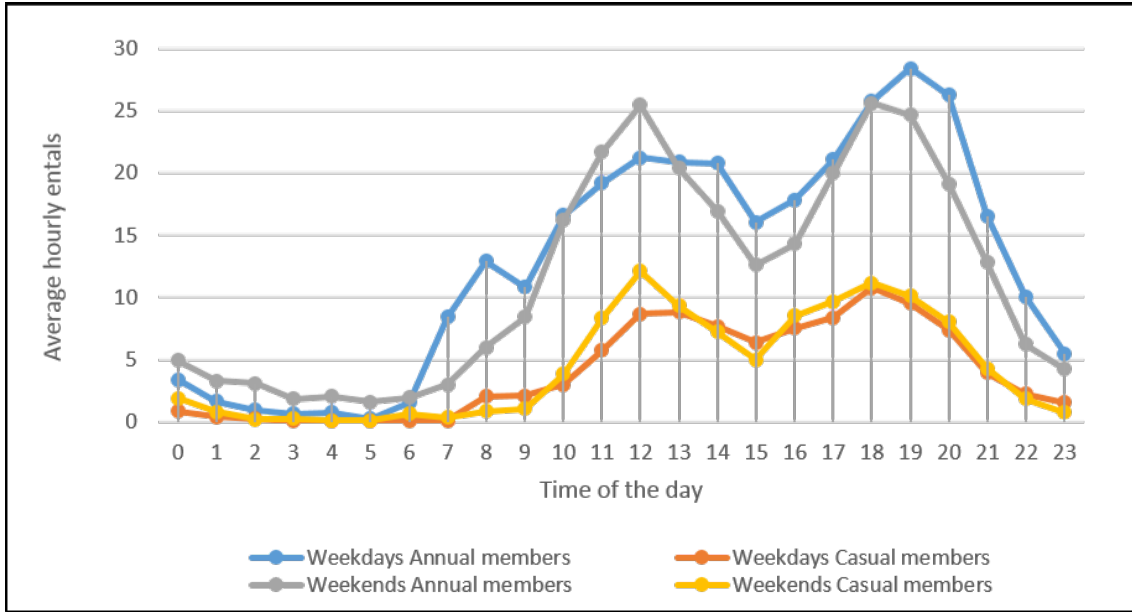
Figure 4.2: Demand profile along an average day (weekdays versus weekends)

4.3.2 Modelling fundamentals

The trip production is generally modelled through regressions in transport studies (Ortúzar & Willumsen, 2011). Regressions permit to estimate the trip rates as a function of a group of variables explaining the phenomenon. In this case, the demand for bike-sharing is modelled using the information recorded through the smartcard transactions as well as weather measurements obtained through sensors. Although multiple linear regression was the first technique to be applied to this research, it was not consistent with the assumptions underlying this method. Specifically, the Breusch-Pagan test yielded a high value for the Chi-square, rejecting the null hypothesis of homocedasticity in all the estimated models. As a result, the linear regression was not accurate enough since the data did not perform as assumed.

In fact, the low trip rates, especially for the trips generated in non-recreational terminals (Table 4.1), suggested an alternative approach was required. Recreational cycling shows greater activity, although still low. Hence, count regression models are suitable to explain and estimate the trip production of the two endogenous variables. As defined by Cameron and Trivedi (2003), "*an event count refers to the number of times an event occurs*". Counts are non-negative and discrete variables generally presenting heteroskedasticity and right skewness, and their variability commonly increases with the mean of the

distribution Hilbe (2011); precisely these characteristics are in contrast with the requirements for the estimation of a standard regression.

The Poisson and Negative Binomial (NB) models are the most popularly applied regression specifications for the estimation of count data. It should be mentioned that both Poisson and NB distribution consider the existence of zero values, this is, the no occurrence of an event, as is the case of this research.

Poisson is a discrete probability distribution with density function Greene (2003):

$$f(y_i|x_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}, \quad y_i = 0, 1, 2, ... \tag{4.1}$$

where y_i is the number of events occurred in the time interval i, x_i corresponds to the values that take the set of explanatory variables in the time interval i, and μ_i is the only parameter involved in the Poisson function which represents the expected value of y_i. The μ parameter is the log-linear function of the explanatory variables denoted by x.

$$\mu_i = exp(x_i^{'}\beta) \tag{4.2}$$

where β represents the parameters that account for the effect of each regressor on the dependent variable y. The following relation defines the Poisson regression:

$$V(y_i|x_i) = E(y_i|x_i) \tag{4.3}$$

Equation 4.3 holds for the equidispersion restriction imposed in Poisson that assumes that the conditional variance is equal to the expected value of y. This property can be unrealistic; overdispersion has been evidenced in numerous case studies and should thereby be tested (Cameron & Trivedi, 1986; Greene, 2003). Poisson's restriction is relaxed in the NB specification through a parameter that measures the overdispersion:

$$V(y_i|x_i) = E(y_i|x_i)(1 + \alpha E(y_i|x_i)) \tag{4.4}$$

where α is the parameter introducing the dispersion. The data is overdispersed when the variance is greater than the mean, which implies α is positive; a negative outcome of α informs that the variance is lower than the mean, the case of underdispersion.

The estimation of the parameters involved in the previously mentioned structures is based on the maximum likelihood estimator (Hilbe, 2011). The likelihood function assumes the values are independent on the period of time i; it is thereby calculated as the

joint probability of reproducing the sample:

$$L(\beta) = \prod_{i=1}^{n} f(y_i|x_i, \beta) \tag{4.5}$$

$$\mathcal{L}(\beta) = lnL(\beta) = \sum_{i=1}^{n} lnf(y_i|x_i, \beta) \tag{4.6}$$

The regression and model parameters are obtained from the derivative of the log-likelihood, whereas the second derivative of equation 4.6 informs on the variance-covariance matrix and the parameter standard errors.

One of the main interest resides on the partial effects that can be derived from these models. These inform on the expected impact on the dependent variable caused by the variation in a explanatory regressor. Considering equation 4.2, the partial effects are computed through the following equation:

$$\frac{\partial E(y_i|x_i)}{\partial x_i} = \mu_i \beta \tag{4.7}$$

4.4 Estimated models and discussion of results

This section presents the parameters obtained from the count models for each type of use of the public bikes considered: recreational and non-recreational. Firstly, the results are discussed concerning the various weather and temporal trip conditions whilst the final subsection evaluates the impacts comparing the trip production of recreational and non-recreational terminals.

The resulting models are presented in Table 4.3.

In the first place, it can be concluded that the restriction of equidispersion assumed in Poisson does not hold true in any of the two types of demand under study. The dispersion parameter, α, has been found statistically significant in both cases, as well as the two hypothesis test computed in the Poisson estimations and proposed by Cameron and Trivedi (1990). It should also be noted that the standard errors of the parameters in the Poisson estimation are smaller than in the NB. This is induced by the relaxation of the dispersion parameter in the NB model, thereby suggesting that the Poisson restriction is unrealistic for the two trip purposes in this research.

Regarding the model fit, the log-likelihood function is improved in the NB specifications as it occurs with the Information Criterion AIC. Equation 4.4 implied the Poisson model is the restricted version of the NB for alpha equal to zero, then the likelihood ratio

Table 4.3: Estimated regression models for recreational and non-recreational demand
(Index 1 refers to binary variables)

Trip production at terminals considered mainly recreational				
	Poisson		Negative Binomial	
	Coefficient	T-test	Coefficient	T-test
Constant	1.458	17.65	1.213	5.36
Morning peak period: 7 - 9 am [1]	-1.740	-20.15	-1.712	-14.37
Afternoon peak period: 1 - 4 pm [1]	-0.232	-7.82	- 0.250	-2.43
Sunday/holiday [1]	-0.119	-4.75	-0.105	-1.96
Temperature (^{o} C)	0.053	14.89	0.063	6.14
Rainfall [1]	-0.190	-3.33	-0.207	-2.38
Rainfall in previous interval [1]	-0.491	-8.03	-0.464	-5.77
Relative humidity > 90% [1]	-0.279	-5.81	-0.268	-3.87
Dispersion parameter (α)			0.379	17.83
Log-likelihood function	-3980.20		-3024.52	
Overdispersion tests	$g=\mu_i$	10.64		
	$g=\mu_i^2$	9.44		
Inf. Cr. AIC/N	8.577		6.524	
Trip production at terminals considered mainly non-recreational				
	Poisson		Negative Binomial	
	Coefficient	T-test	Coefficient	T-test
Constant	2.321	25.27	2.344	17.09
Morning peak period: 7 - 9 am [1]	0.343	6.69	0.342	5.1
Afternoon peak period: 1 - 4 pm [1]	0.151	3.61	0.145	2.15
Sunday/holiday [1]	-0.408	-8.89	-0.411	-7.36
Temperature > 28 ^{o} C [1]	-0.228	-2.48	-0.243	- 1.81
Rainfall (l/m^2)	-0.063	-2.19	-0.062	-2.07
Rainfall in previous interval [1]	-0.475	-6.23	-0.480	-6.23
Relative humidity (%)	-0.009	-7.13	-0.009	-5.09
Dispersion parameter (α)			0.159	8.6
Log-likelihood function	-2343.41		-2258.81	
Overdispersion tests	$g=\mu_i$	9.47		
	$g=\mu_i^2$	8.76		
Inf. Cr. AIC/N	5.057		4.877	

test can be used to compare the different explanatory power of the two models. The null hypothesis assumed in the likelihood ratio test performed for the two models are rejected, thereby indicating that the extra parameter that introduces the overdispersion in the NB model does improve the replication of the data significantly. Consequently, the NB specification is preferred in the two circumstances: the total trip production in pro-recreational and non pro-recreational terminals in the BSS in Santander. It should be mentioned that all models are reasonable in terms of tendencies since the estimated parameters confirm the initial hypotheses and go in line with previous studies reviewed in the literature.

Once the most accurate model is selected for each case, this is, the NB models, it is time to discuss on the results. From a general perspective, as can be concluded from Table 4.3, not the same factors explain bikeshare for transport and for leisure, as also obtained by Brandenburg *et al.* (2007). It should also be noted that the wind speed has not resulted a significant influence in any case, as also reported by Winters *et al.* (2007), Nankervis (1999), and Rietveld and Daniel (2004), but in contrast to Saneinejad *et al.* (2012), who determined that the wind speed negatively affects cycling, and doubly than walking.

4.4.1 Temperature

As Table 4.3 informs, the temperature takes part in different forms. On the one hand, those terminals that are mostly demanded for trip purposes other than leisure experiment a decrease in their activity when the temperature is over 28 ºC. This result is in agreement with other case studies which also found extreme heat to decrease cycling (Gebhart & Noland, 2014; Miranda-Moreno & Nosal, 2011) and also supports the point made by Böcker *et al.* (2013) about the sometimes non-linear impact of weather on travel behaviour. On the contrary, Saneinejad *et al.* (2012) obtained non sensitiveness of commuting cycling to temperatures above 15ºC, in other words, only cold weather was influential in the choice of the bike as a commuting mode. As above mentioned, the present case study applies summer data, and as informed in Table 4.3, cold temperatures are not represented in the models, so their influence cannot be measured.

On the other hand, temperature positively affects bike rides for leisure, as informed by Table 4.3. No significance has been obtained for specific temperature thresholds so the temperature and recreational bikeshare rides are linearly related. This trend was

previously obtained by Bergström and Magnusson (2003), Pucher and Buehler (2006), and Faghih-Imani *et al.* (2014) among others.

4.4.2 Rainfall

As expected and confirmed in the majority of the reviewed studies in the literature, the parameters that account for the raining conditions all exhibit a negative sign. However, it should be noted that in the case of cycling for a purpose other than leisure, the more rainfall the less bikes are rented, whereas rainfall is conceived as a dummy variable (yes/no) for recreational cyclists. These results imply that free-time cycling is significantly determined by any rain that could fall, whereas non- recreational trip production is linearly affected by the amount of rainfall.

Additionally, the models contribute to confirm the impact of the weather inertia in terms of precipitation: the hourly demand for bikes decreases when it has rained the previous hour both in the case of recreational and utilitarian cycling. Hence, as the literature suggests, it may be expected that activities are cancelled or postponed (Faghih-Imani *et al.*, 2014), especially recreational ones, whereas in the case of obliged trips, people may eventually shift from bike to other modes under the effect of rainfall, as evidenced by Gebhart and Noland (2014).

4.4.3 Humidity

As is revealed in Table 4.3, the relative humidity has a negative influence on the number of bikes rented for a transport purpose. This factor has been pointed out as a negative influence on cycling (Faghih-Imani *et al.*, 2014; Gebhart & Noland, 2014), possibly because it stimulates sweating, a major inconvenient in trips heading to work or study. Miranda-Moreno and Nosal (2011) noticed that over the 60% of relative humidity the bike rentals tend to decrease. In the case of the recreational demand in Santander it is over the value of 90% that this factor decreases the rental generation.

4.4.4 Time-based trip production

The estimated models (4.3) also inform on the influence of time-based variables. In this regard, it is remarkable that the morning and afternoon peak periods have opposed influence on the demand for the two trip purposes: from 7 to 9 am and from 1 to 4 pm

non-recreational terminals increase their demand whilst the terminals considered related to leisure activities present lower activity. Such is a valuable result in terms of understanding the demand for bike sharing. It implies that the likelihood of bikes being rented as a mode for transport increases at the city's peak hours in weekdays, meaning that the terminals considered non-recreational (Figure 4.1) present a greater activity at such periods.

It is surprising the result of the negative effect of Sundays and holidays on the demand, which occurs even in the case of recreational mobility. It should be remarked that the research conducted by Miranda-Moreno and Nosal (2011) and Faghih-Imani *et al.* (2014) also found a lower bikeshare activity at weekends. In the case of Santander, the outcome is interpreted in line with the accused decrease on the general mobility in the city on Sundays.

4.4.5 Evaluation of impacts

In addition to the positive or negative influence of the explanatory variables on the demand for bike-sharing, their actual impact can be assessed from the estimated parameters. The marginal effects are presented in Table 4.4 and inform on the expected change on the trip production as a result of a unitary change of each variable (equation 4.7), *ceteris paribus.* It should be noted that the average hourly trip production is 11.70 and 4.98 rentals for the case of terminals demanded for recreational cycling and those demanded for transport, respectively(Table 4.1); as a result, the impacts are also greater in the former than in the latter case.

In terms of weather, the rainfall impacts the most on the demand for both types of terminals: those mainly used for recreational bikeshare and those for non-recreational cycling. In the first case, the highest impact (an average change of 4.4 rentals an hour) is expected the consequence of rainfall occurring in the preceding hour. Therefore, although it does not hold true for the case of the terminals mainly demanded for utilitarian bikeshare, recreational ones show a great weather inertia. In the second place, in terms of weather, the trip production is influenced by a relative humidity above the 90%, with an expected decrease of 2.809 rentals in an hour under this condition. On the other hand, if it is raining, a decrease of 2.2 rentals is expected in an hour, whereas a unit change on temperature does not impact as much.

Table 4.4: Estimated partial effects (Index 1 refers to binary variables)

Trip production at terminals considered mainly recreational	
Morning peak period: 7 - 9 am [1]	-10.177
Afternoon peak period: 1 - 4 pm [1]	-2.687
Sunday/holiday [1]	-1.199
Temperature (º C)	0.743
Rainfall [1]	-2.214
Rainfall in previous interval [1]	-4.444
Relative humidity > 90% [1]	-2.809
Trip production at terminals considered mainly non-recreational	
Morning peak period: 7 - 9 am [1]	1.969
Afternoon peak period: 1 - 4 pm [1]	0.762
Sunday/holiday [1]	-1.789
Temperature > 28 º C [1]	-0.047
Rainfall (l/m^2)	-1.952
Rainfall in previous interval [1]	-0.309
Relative humidity (%)	-1.085

The demand for terminals associated to non-recreational cycling is mainly influenced by the rainfall: an extra l/m^2 of precipitation is expected to generate a decrease of about 2 rentals the total hour demand. The relative humidity is the second effect in terms of weather on utilitarian trip production.

Considering the temporal circumstances, utilitarian cycling increases during the morning and afternoon peak periods, but mostly in the morning, when an increase of almost 2 rentals is expected compared to the reference case, this is, the off-peak period. Sundays and holidays, on the contrary, impact negatively on bike renting for transport, expecting an average of 1.79 less trips than in the off-peak period, which also considers Saturdays.

The demand for the system as a tool for recreation is also affected the greatest by the precipitation registered in the previous hour and secondly, by relative humidity levels above 90%, leading to an average of 4.44 and 2.81 less rentals respectively. As expected, the temperature stimulates cycling for leisure whereas rain is predicted to act as a deterrent to cycling translated into an average of 2.21 less bikes rented. Contrary to utilitarian cycling, the morning and afternoon city's peak period are associated to an

important decrease in the demand for leisure with approximately 10 and 2.6 bikes less than in the off-peak period, respectively. As has already been mentioned, the lower rates of recreational cycling on Sundays and holidays were not expected but are interpreted in line with the also calm traffic that is experienced in such days.

The two models inform on the different impact of weather and temporal circumstances on recreational and utilitarian bike renting. The recreational demand is negatively impacted by all weather variables except for the temperature whereas the demand for the bike as a mode of transport is negatively impacted by all weather aspects found significant. It should be noted that both demands are similarly (and substantially) affected by the precipitation in the previous hour, being the rainfall the major barrier to cycle either for leisure and for transport. Finally, as resulting in previous research, a greater sensitivity is revealed in the case of cycling for leisure. This result supports the conclusions from the literature review, particularly that from Nankervis (1999), who stated that "*those who ride, ride despite the conditions*". In any case, the previously mentioned results should be referred to Table 4.1 for an overall perspective of the phenomenon under study.

4.5 Conclusions

This last stage of the thesis has been built upon the results obtained in the previous analyses of Chapters 2 and 3 and is intended to understand the different impact of weather and time-based trip production at a station level and according to the main character of the bikeshare demand: for transport or for leisure. For this purpose, Poisson and negative binomial regressions have been performed, finding substantial differences between the two trip purposes.

This research uncovers the factors affecting the demand for bike-sharing in Santander. As was expected, the general conclusions support previous research concerning on the climatological influence on cycling such as the positive effect of the temperature on recreational cycling and the negative impact of humidity and high temperatures on cycling for transport purposes. In this regard, it should be remarked that above all, the highest impact is derived from precipitation.

Further insights have resulted from the estimated models that contribute to the scientific knowledge. First, the users' behaviour presents an evident inertia when it comes to rainfall, finding that the hourly bike-sharing demand decreases when it has rained in

the previous hour. Second, as some authors suggest, the effect of weather on cycling is not always linear; this study has yielded both linear and non-linear influences of temperature and relative humidity. Third, it can be concluded that recreational demand is more elastic to weather conditions, as is generally concluded in the literature. Fourth, the impact of the morning and afternoon peak periods in weekdays is opposite for recreational cycling and utilitarian, confirming the suitability of the spatio-temporal approach proposed in this thesis and previously presented in the previous chapter: on the one hand, the classification of rentals into recreational and non-recreational is justified with the travel patterns analysed in the previous chapter and, indirectly, the trip purpose is assumed linked to the land uses at origin and destination; on the other hand, the time of the day should be considered when studying cycling for leisure and for transport separately.

The interest of this research is threefold:

- Further knowledge is provided on what drives bike-sharing demand;

- New uses of automatic data are proposed that overcome traditional manual data limitations: rentals are classified into recreational and non-recreational according to the terminals at origin and destination and are separately regressed on a predefined set of variables. These models could not have been estimated with manual data due to limited observations.

- The different response that recreational and non-recreational bike mobility exhibit to meteorological conditions and temporal circumstances has been measured.

From a methodological point of view, the conclusions drawn in previous international research have been considered so that the present research contributes to gain further knowledge on the topic. The present approach overcomes the likely violation of the regression assumptions when the phenomenon exhibits a natural interpretation as counts, showing the potential of alternative specifications to fit the data more accurately. In any case, cautions should be taken in the selection of the method for prediction and new specifications should be tested and evaluated. Alternative methodologies could be applied to understand how the demand responds and new factors should be tested. For instance, the tourism rates may have an influence on recreational bikeshare rentals; however only monthly measurements were accessible for this case study, implying insufficient variability for modelling the two-months demand. Seasonal bikeshare demand could address this

issue and contribute in this line.

Finally, the study strengthens the potential of automatic data to deeply understand the demand for transit services, particularly that of bike-sharing systems. The outcomes and conclusions drawn from the models are useful insights for the efficient management of this type of services and the design of actions to promote the bike mobility as a complementary strategy towards sustainable cities. It should be highlighted that this research is also of potential benefit for the management of alternative transit services since it is expected that the demand for cycling will potentially shift to the rest of modes under unfavourable weather.

Chapter 5

Conclusions and further research

This thesis proposes a methodological approach for assessing the demand and quality of bike-sharing systems using automatic and manual data. This research contributes to the scientific knowledge and practice with new methods to analyse the particular demand of BSSs, but also with insights and valuable ingredients to optimise the system.

The application of the methodologies presented in this thesis permits to respond to the set of questions raised in the introduction:

- What do citizens demand from a bike-sharing system?

- How do users perceive the quality of the service?

- How do subscribers actually use the service?

- Is the users' behaviour sustainable in terms of capacity?

- To what extent do diverse factors such as the land use or the weather affect the demand for bikeshare?

The answers are indeed the knowledge derived from the results obtained from the models and data mining techniques. The answers are particular to the BSS and city to which the methodology is applied although the approach is applicable to any case study. In the case of the present thesis, interesting insights have been revealed on the TusBic system.

The thesis follows a sequence of methods starting with information and data collection, followed by data analysis and concluding with modelling techniques intended to assess the demand for a bike-sharing system. Along the proposed approach, the service

strengths and weaknesses are identified, uncovering demand indicators to support the decision making. The sequence is justified by the need to get a picture of the service: the actual demand for it, the user's needs and expectations and the use of the system as a result of the operator's performance and the system's limitations. Hence, one of the most important conclusions from this research is the need to adequately design the sequential stages so that the models are specified and their outcomes interpreted after the modeller knows the service in depth.

As is the case of this thesis, the public involvement provides insightful information and a very necessary perspective that would enrich subsequent stages of the research or study. The debate that is generated in focus group sessions give context to the study and is therefore an ideal starting point of any research or study, and essential when it is aimed at understanding the demand of a service. In particular, focus groups permit to identify the variables that influence the user's behaviour so that all important attributes, users' characteristics and travel needs are considered in the study. Indeed, important variables may be explanatory of the demand but may also be the source of heterogeneity, as occurs in the presented case study and is considered along the thesis. Various forms of heterogeneity have been determined and the variability has been demonstrated in quality perceptions, the service usage, travel patterns and trip production.

The methodology described in this thesis is ultimately intended to provide indicators to assess the demand and this way, establish a range of performance in which the concession should operate. The incentive-disincentive framework should consider the business-management model, and the responsibilities of both the operator and the public administrator. In order to ensure the social benefit and the efficiency, the supplier's economical benefit should depend on the percentage of fulfilment of the conditions in the contract with regard to the level of service and quality standards. However, the benefits from the application of the presented approach are diverse since so are the implications of the results obtained. Actions should be directed to improve the aspects that are the most important to users whilst bike trials should be overcome with bike relocation routes that consider the bikes that have been returned and substituted by an alternative one.

The decisions could be strategic, tactical or operational but in any case, the benefit will be shared by different stakeholders. The redesign of the system will have a direct effect on the users' satisfaction and fidelity and, luckily, it will promote the uptake of

bike-sharing. It should be noted that also depending on how the strategy is traduced into supply standards, the BSS may be conceived differently by the citizens and/or tourists. In other words, as it has been demonstrated in this thesis, the characteristics of the supply (fare structure, coverage or spatial distribution of terminals) do have a direct impact on the usage of the system and thus is the key to promote certain types of demand: casual users versus registered card-holders, recreational rides versus utilitarian trips, and eventually tourists versus residents.

The presented methodology is a contribution itself since it provides original techniques to understand the demand and detect the deficiencies of the system. In this regard, the main contribution lies in the combination of both automatically and manually collected data, concluding their potential depends on the objective of a method or model. However, as it has been demonstrated in this thesis, both sources of information present limitations but also interesting advantages. It has been shown that questionnaires permit to obtain valuable and direct responses regarding personal trip circumstances, or subjective issues such as the quality perceived by the users of the system, as addressed in Chapter 2. Nevertheless, although it is for certain that automatic records cannot explicitly gather these or even the purpose of the trip, far from being this a limitation, it should be faced as a challenge. Extracting information from data provided by ITS requires both data mining skills and knowledge of the analysed system. Undoubtedly, the analyses and models that could be performed directly depend on the nature and quality of the recorded data; but even if the automatically recorded variables are limited, it is the entire population that is recorded by the ITS, thereby avoiding the sampling bias inherent to traditional manually collected data. The challenge is actually to take the most of each source of information being rigorous in the data collection and management since the outcomes and conclusions will be drawn from them.

Conclusions from the study of the service quality perceived by users (Chapter 2)

The proposed methodology to model the quality perceived by users involves the public from the very first phases to guarantee a reliable representation of reality during the modelling process. The collected data is the result of successive phases aimed at designing the questionnaire and identifying the variables that influence the users' behaviour. The

ordered nature of quality ratings justifies the estimation of random ordered probit models to understand the perception of the quality of the service as a function of the characteristics which define it. The proposed models represent a tool which has great potential as an aid in the management of bicycle hiring services because not only does it report on the different service variables and their degree of importance on quality, but it also provides information on population variability and the systematic differences in perceptions resulting from different socioeconomic and journey characteristics. Heterogeneity in the population appears in the scores given for the distribution of the docking stations while among the causes of systematic variation in the perceptions of different attributes are gender, age, purpose of journey, type of ticket purchased and the access and journey times.

Regarding the outcomes obtained for the case study, variables such as the safety during the journey and the available information go unmentioned in the initial evaluation of overall quality while they turn out to be of greatest weight at the end of the interview. The opposite occurs with the cost, with the highest *a priori* score, but which ceased to be of importance for the users after considering each and every characteristic of the service. Additionally, in line with the literature, it has been determined that the accessibility to terminals greatly influences the perceived quality when this attribute presents low standards, concluding the desire of a dense network that would reduce access time to the system.

Conclusions from the data mining processes and the BSS usage characterisation (Chapter 3)

Two data mining processes have been proposed. The first one identifies a non-ignorable pattern that has been calibrated and named "bike trial with substitution". This behaviour is described by two trip-legs: an initial short rental in which a first bike is tested and then returned to the docking station, probably deficient or not working properly, followed by the rental of a new bicycle which allows carrying out the initially planned journey. Failing to remove the first trip-leg, this is, the bike trials, would lead to inflated intra-zonal demand and a reduced average travel time. Thus, it is important to apply the data mining to assess the presence of this kind of rentals before further use of the database and posterior modelling.

The detection of these specific rentals with bike substitution can provide valuable

information regarding:

- User satisfaction: in fact, the return of a bike to be substituted by a new one indicates the user does not find the quality of the bike as expected to make the trip. Thus, the detection of these rentals provides interesting information that would also enrich the study of the service's perceived quality approached in Chapter 2 with data from the TusBic bike-sharing service in Santander.

- Management and maintenance of the fleet of bikes: since the ITS system records the bike of each rental, it is possible to identify the bikes that have been returned more frequently and so to arrange for physical checks and reparation more effectively.

Alternatively, a second process has been proposed that identifies the BSS usage conditions that introduce systematic heterogeneity in travel patterns so as to infer travel patterns from the objective information of the users' actual behaviour. In this case, a difference is made between usage behaviour and travel patterns: the usage is described by the actual trip-chaining gathered with every smartcard transaction and is directly influenced by the limitations of the BSS as a public renting service, whilst the travel behaviour relates to the spatio-temporal distribution, the travel time and trip purpose.

It is concluded that the time-lapse between consecutive rentals or trip-legs, together with the spatial trip-chaining are key conditions to classify rentals into less heterogeneous demands. As a result, the proposed algorithm could be considered a clustering technique that permits characterizing the predominant travel patterns. The developed data mining process has been validated, implying that the hypothesis assumed do identify usage patterns that show systematic heterogeneity in travel patterns.

Sample size limitations are overcome in the algorithms where smartcard data has been applied, thereby permitting the identification of less usual mobility patterns such as bike substitutions that would not have been obtained through traditional collection techniques. Nevertheless, it is for certain that the main shortcoming of this type of data is failing to know the trip purpose. Therefore, results should be read with caution and be based on ingenious analyses together with related scientific literature; these ingredients have permitted an insightful discussion of the results obtained in the case study adopted, assisting in the understanding of the usage and travel behaviour.

In particular, as has been shown with this approach, tricky behaviours can be detected for instance when a user does not behave as expected in a sharing service but still

being permitted. Indeed, two specific usage types identified in the proposed data mining process are particularly insightful for operators: the will to reset the rental time and the bike substitutions. The first behaviour might compromise the capacity of the system as it goes against the sharing nature of the service. The times and conditions under which the flat fee interval is overcome or when subsequent rentals are made to avoid paying extra money are identified. Their analysis is the key for an efficient management and very valuable information to improve and redesign the service. On the other hand, a bike substitution may be informing of a bad performance of the bike, a circumstance which influences the users' overall perception of quality, as concluded in Chapter 2. Since the bike IDs are recorded, they can be spotted and considered in the bike redistribution route to check them. In conclusion, even though these systems limit the use of the bikes, it is possible to violate such sharing conception without being charged for it. The proposed methodology provides the knowledge to develop incentives that could avoid such circumstances.

This research has evidenced the usage behaviour can be used to characterize the demand for bike-sharing in terms of travel patterns. Additionally, the discussion on the results obtained with the application have indirectly inferred the results that would be expected in any other type of BSS considered in the classification proposed by O'Brien *et al.* (2014).For instance other BSSs showing a routinely use of bikes being greatly demanded for commuting should expect a greater presence of symmetric trips. Hence, the data mining techniques proposed 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.

Conclusions from the estimation of the impact of weather and time-based variables on the trip production (Chapter 4)

Chapter 4 presented a regression approach to determine the impact of the weather and time-based variables on the trip production at terminals predominantly demanded for leisure and those experimenting a predominant demand for the BSS for transport purposes. This last stage of the research is built upon two assumptions derived from the outcomes in the previous phases: first, that the travel time and the quality perceptions differ based on the trip purpose (Chapter 2), and second, that the demand for a terminal and the travel

behaviour are associated to the land use and main activities in the surroundings of the bike station (Chapter 3).

Interesting insights have been uncovered that complement previous research on the influence of weather and time-based trip conditions on the demand for cycling. In terms of trip production, this thesis provides new evidences that the weather differently affects the demand for terminals depending on the land use in the surroundings of the terminals and, ultimately, on the activities developed in the area of influence. The temperature has a direct impact on the trip production in terminals located in recreational areas whereas high temperatures above 28 º C reduce the demand for terminals located in residential areas where the bike has been shown to be mainly demanded as a mode of transport as concluded from previous stages of the thesis. Another interesting conclusion drawn from this thesis is the weather inertia: the rain has shown a significant effect on the trip production of the next hourly period, proving not only does it affects the instant when it is raining but it has a longer impact.

5.1 Further research

The following lines of study are left open to further research.

First, the application of the proposed methodology to other case studies would yield interesting information regarding the demand for bike-sharing. The comparison of different systems would allow to provide a solid characterisation of the demand in general terms, but also to identify new sources of variability among BSSs, this is, the particular behaviour and bikeshare needs depending on the system (size, fare structure, subscription possibilities, schedule), the type of city (size, activities, topography, meteorology), or the cycling culture.

Second, the system used to validate the sequential methods is ready to be optimised considering the results regarding the quality perceived by users, the bike trials and the rest of the usage behaviours uncovered with the methodology developed in this thesis as well as the impact of the time period, type of day and weather on the trip production.

Third, as a natural extension of this thesis, the variables that have been determined influential in the users' behaviour need to be included in the utility function of the bikeshare alternative in mode choice model. Due to the particular nature of quality attributes, these are considered latent variables in a modal choice, thereby requiring specific

techniques to be introduced in the utility function.

Fourth, the study of the service quality presented in this thesis is only a starting point; new approaches in terms of theoretical fundamentals, specifications, attributes or data collection should be considered. Specifically, panel data collection on quality percep-tions and the definition of various forms to determine the attribute importance are two interesting lines of research derived from this thesis.

Fifth, new data mining processes need to be developed to take full advantage of the information generated by automatic systems and sensors. The algorithms presented in this thesis could be improved and further analysis and models could be applied to each of the usage patterns identified. In addition, the trip-chaining assumptions could gain complexity and multi-modal trip-chaining could be considered in a case study performing this behaviour.

Finally, new approaches could be developed to assess the demand for bike-sharing. Regarding the knowledge obtained from the modelling outcomes, it is interesting to com-pare the impact of the long-term climate and seasonal tendencies on the demand, with the weather impact in the short term obtained in this thesis.

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, 2140*.

Akar, G., Fischer, N., & Namgung, M. (2013). Bicycling Choice and Gender Case Study: The Ohio State University. *International Journal of Sustainable Transportation, 7*(5), 347–365.

Alonso, B. (2010). *Bus stop location model considering the operation and incidents on the network*. Ph.D. thesis, University of Cantabria.

An, M., & Chen, M. (2007). Estimating Nonmotorized Travel Demand. *Transportation Research Record: Journal of the Transportation Research Board, 2002*.

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

Ben-Akiva, M., Walker, J., Bernardino, A., Gopinath, D. A., Morikawa, T., & Polydoropoulou, A. (2002). Integration of choice and latent variable models. In H. S. Mahmassani (Ed.), *Perpetual Motion: Travel Behaviour Research Opportunities and Challenges* (pp. 431–470). Amsterdam: Pergamon.

Bergström, A., & Magnusson, R. (2003). Potential of transferring car trips to bicycle during winter. *Transportation Research Part A: Policy and Practice, 37*(8), 649–666.

Bernhoft, I. M., & Carstensen, G. (2008). Preferences and behaviour of pedestrians and cyclists by age and gender. *Transportation Research Part F: Traffic Psychology and Behaviour, 11*(2), 83–95.

Böcker, L., Dijst, M., & Prillwitz, J. (2013). Impact of Everyday Weather on Individual

Daily Travel Behaviours in Perspective: A Literature Review. *Transport Reviews, 33*(1), 71–91.

Bonham, J., & Wilson, A. (2012). Bicycling and the Life Course: The Start-Stop-Start Experiences of Women Cycling. *International Journal of Sustainable Transportation, 6*(4), 195–213.

Bordagaray, M., dell'Olio, L., Ibeas, A., & Cecín, P. (2013). Modelling user perception of bus transit quality considering user and service heterogeneity. *Transportmetrica A: Transport Science, 10*(8), 705–721.

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.

Brandenburg, C., Matzarakis, A., & Arnberger, A. (2007). Weather and cycling—a first approach to the effects of weather conditions on cycling. *Meteorological Applications, 14*, 61–67.

Cameron, A. C., & Trivedi, P. K. (1986). Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests. *Journal of Applied Econometrics, 1*(1), 29–53.

Cameron, A. C., & Trivedi, P. K. (1990). Regression-based tests for overdispersion in the Poisson model. *Journal of Econometrics, 46*(3), 347–364.

Cameron, A. C., & Trivedi, P. K. (2003). Essentials of Count Data Regression. In *A Companion to Theoretical Econometrics* (pp. 331–348).

Cantwell, M., Caulfield, B., & O'Mahony, M. (2009). Examining the factors that impact public transport commuting satisfaction. *Journal of Public Transportation, 12*(2), 1–21.

Caruana, A. (2002). Service loyalty: The effects of service quality and the mediating role of customer satisfaction. *European Journal of Marketing, 36*(7), 811–828.

Cervero, R., & Duncan, M. (2003). Walking, Bicycling, and Urban Landscapes: Evidence From the San Francisco Bay Area. *American Journal of Public Health, 93*(9), 1478–1483.

Chemla, D., Meunier, F., & Wolfler Calvo, R. (2013). Bike sharing systems: Solving the

static rebalancing problem. *Discrete Optimization, 10*(2), 120–146.

Cirillo, C., Eboli, L., & Mazzulla, G. (2011). On the Asymmetric User Perception of Transit Service Quality. *International Journal of Sustainable Transportation, 5*(4), 216–232.

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.

Cronin, J. J., & Taylor, S. A. (1994). SERVPERF versus SERVQUAL: and Reconciling of Service Measurement Perceptions Quality. *Journal of Marketing, 58*(1), 125–131.

Cullinane, K., & Wang, T. F. (2006). Data Envelopment Analysis (DEA) and Improving Container Port Efficiency. *Research in Transportation Economics, 17*, 517–566.

De Oña, J., De Oña, R., & Calvo, F. J. (2012). A classification tree approach to identify key factors of transit service quality. *Expert Systems with Applications, 39*(12), 11164–11171.

De Oña, J., de Oña, R., Eboli, L., & Mazzulla, G. (2013). Perceived service quality in bus transit service: A structural equation approach. *Transport Policy, 29*, 219–226.

Del Castillo, J. M., & Benitez, F. G. (2012). Determining a public transport satisfaction index from user surveys. *Transportmetrica A: Transport Science, 9*(8), 1–29.

Dell'Olio, L., Ibeas, A., Bordagaray, M., & Ortúzar, J. D. D. (2011a). 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., & Cecín, P. (2010). Modelling user perception of bus transit quality. *Transport Policy, 17*(6), 388–397.

Dell'Olio, L., Ibeas, A., & Cecin, P. (2011b). The quality of service desired by public transport users. *Transport Policy, 18*(1), 217–227.

dell'Olio, L., Ibeas, A., & Moura, J. L. (2011c). 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. *Jour-

nal 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.

Dill, J., & Gliebe, J. (2008). *Understanding and measuring bicycling behavior: A focus on travel time and route choice.* Oregon Transportation Research and Education Consortium, Portland, OR.

Eboli, L., & Mazzulla, G. (2007). Service Quality Attributes Affecting Customer Satisfaction for Bus Transit. *Journal of Public Transportation, 10*(3), 21–34.

Eboli, L., & Mazzulla, G. (2008a). A Stated Preference Experiment for Measuring Service Quality in Public Transport. *Transportation Planning and Technology, 31*(5), 509–523.

Eboli, L., & Mazzulla, G. (2008b). Willingness-to-pay of public transport users for improvement in service quality. *European Transport / Trasporti Europei, XIV, 38*, 107–118.

Eboli, L., & Mazzulla, G. (2009). A New Customer Satisfaction Index for Evaluating Transit Service Quality. *Journal of Public Transportation, 12*(3), 21–37.

Eboli, L., & Mazzulla, G. (2010). How to Capture the Passengers' Point of View on a Transit Service through Rating and Choice Options. *Transport Reviews, 30*(4), 435–450.

Eboli, L., & Mazzulla, G. (2011). A methodology for evaluating transit service quality based on subjective and objective measures from the passenger's point of view. *Transport Policy, 18*(1), 172–181.

Eboli, L., & Mazzulla, G. (2012). Structural Equation Modelling for Analysing Passengers' Perceptions about Railway Services. *Procedia - Social and Behavioral Sciences, 54*, 96–106.

Efthymiou, D., Antoniou, C., & Waddell, P. (2013). Factors affecting the adoption of vehicle sharing systems by young drivers. *Transport Policy, 29*, 64–73.

Espino, R., Martín, J. C., & Román, C. (2008). Analyzing the effect of preference heterogeneity on willingness to pay for improving service quality in an airline choice context.

Transportation Research Part E: Logistics and Transportation Review, 44(4), 593–606.

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.

Friman, M. (2004). Implementing Quality Improvements in Public Transport. *Journal of Public Transportation, 7*(4), 49–65.

Friman, M., Edvardsson, B., & Gärling, T. (2001). Frequency of negative critical incidents and satisfaction with public transport services. I. *Journal of Retailing and Consumer Services, 8*(2), 95–104.

Friman, M., & Fellesson, M. (2009). Service Supply and Customer Satisfaction in Public Transportation: The Quality Paradox. *Journal of Public Transportation, 12*(4), 57–69.

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.

Garrard, J., Rose, G., & Lo, S. K. (2008). Promoting transportation cycling for women: The role of bicycle infrastructure. *Preventive Medicine, 46*(1), 55–59.

Garrido, R. a., & Ortúzar, J. D. D. (1994). Deriving Public Transport Level of Service Weights from a Multiple Comparison of Latent and Observable Variables. *Journal of the Operational Research Society, 45*(10), 1099–1107.

Gatersleben, B., & Appleton, K. M. (2007). Contemplating cycling to work: Attitudes and perceptions in different stages of change. *Transportation Research Part A: Policy and Practice, 41*(4), 302–312.

Gebhart, K., & Noland, R. B. (2014). The impact of weather conditions on bikeshare trips

in Washington, DC. *Transportation, 41*(6), 1205–1225.

González, F., Melo-Riquelme, C., & Grange, L. De. (2015). A combined destination and route choice model for a bicycle sharing system. *Transportation.* DOI: 10.1007/s11116-015-9581-6

Gonzalo-Orden, H., Dell'Olio, L., Ibeas, A., & Rojo, M. (2011). Modelling gender perception of quality in interurban bus services. *Proceedings of the ICE - Transport, 164*(TR1), 43–53.

Gordon, J. B. (2012). *Intermodal Passenger Flows on London's Public Transport Network.* Ph.D. thesis, Massachusetts Institute of Technology.

Greene, W. H. (2003). Discrete choices and event counts. In *Econometric Analysis* (6th ed., pp. 760–832). Prentice Hall.

Heinen, E., Maat, K., & van Wee, B. (2013). The effect of work-related factors on the bicycle commute mode choice in the Netherlands. *Transportation, 40*(1), 23–43.

Heinen, E., van Wee, B., & Maat, K. (2010). Commuting by Bicycle: An Overview of the Literature. *Transport Reviews, 30*(1), 59–96.

Hensher, D. a., Mulley, C., & Yahya, N. (2010). Passenger experience with quality-enhanced bus service: The tyne and wear "superoute" services. *Transportation, 37*(2), 239–256.

Hensher, D. A., & Prioni, P. (2002). A service quality index for area-wide contract performance assessment. *Journal of Transport Economics and Policy, 36*(1), 93–113.

Hensher, D. a., Rose, J. M., & Greene, W. H. (2005). *Applied choice analysis: A primer.* Cambridge University Press.

Hensher, D. a., Stopher, P., & Bullock, P. (2003). Service quality - developing a service quality index in the provision of commercial bus contracts. *Transportation Research Part A: Policy and Practice, 37*(6), 499–517.

Highway capacity manual 2010. (2010). *Transportation Research Board of the National Academies.* Washington, D.C.

Hilbe, J. M. (2011). *Negative Binomial Regression* (2nd ed.). Cambridge University Press.

Hu, K.-C., & Jen, W. (2006). Passengers' Perceived Service Quality of City Buses in Taipei: Scale Development and Measurement. *Transport Reviews, 26*(5), 645–662.

Ibeas, A., Dell'Olio, L., & Montequín, R. B. (2011). Citizen involvement in promoting sustainable mobility. *Journal of Transport Geography, 19*(4), 475–487.

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.

Transit Capacity and Quality of Service Manual (2013). Third Edition, TCRP Report 165. Washington, D.C.

Lai, W.-T., & Chen, C.-F. (2011). Behavioral intentions of public transit passengers—The roles of service quality, perceived value, satisfaction and involvement. *Transport Policy, 18*(2), 318–325.

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.

Lin, J.-R., & Yang, T.-H. (2011). Strategic design of public bicycle sharing systems with service level constraints. *Transportation Research Part E: Logistics and Transportation Review, 47*(2), 284–294.

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 feedering 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.

Matzarakis, A., Mayer, H., & Iziomon, M. G. (1999). Applications of a universal thermal index: physiological equivalent temperature. *International Journal of Biometeorology, 43*, 76–84.

McKelvey, R. D., & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *The Journal of Mathematical Sociology, 4*(1), 103–120.

Midgley, P. (2011). *Bicycle-sharing schemes: Enhancing sustainable mobility in urban areas.* New York.

Miranda-Moreno, L. F., & Nosal, T. (2011). Weather or not to cycle; whether or not cyclist ridership has grown: a look at weather's impact on cycling facilities and temporal trends in an urban environment. *Transportation Research Record: Journal of the Transportation Research Board, 2247*, 42–52.

Mittal, V., & Kamakura, W. a. (2001). Satisfaction, Repurchase Intent, and Repurchase Behavior: Investigating the Moderating Effect of Customer Characteristics. *Journal of Marketing Research, 38*(1), 131–142.

Mokonyama, M., & Venter, C. (2013). Incorporation of customer satisfaction in public transport contracts - A preliminary analysis. *Research in Transportation Economics, 39*(1), 58–66.

Monzón, A., & Rondinella, G. (2010). *PROBICI. Guía de la Movilidad Ciclista. Métodos y técnicas para el fomento de la bicicleta en áreas urbanas.* (I. para la diversificación y ahorro de la energía (IDAE), CICCyP, & Transyt, Eds.). Madrid.

Morikawa, T., Ben-Akiva, M. E., & McFadden, D. (2002). Discrete choice models incorporating revealed preferences and psychometric data. *Econometric Models in Marketing,*

16, 29–55.

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.

Nair, R., Miller-Hooks, E., Hampshire, R. C., & Bušić, A. (2013). Large-Scale Vehicle Sharing Systems: Analysis of Vélib'. *International Journal of Sustainable Transportation, 7*(1), 85–106.

Nankervis, M. (1999). The effect of weather and climate on bicycle commuting. *Transportation Research Part A: Policy and Practice, 33*(6), 417–431.

Nathanail, E. (2008). Measuring the quality of service for passengers on the hellenic railways. *Transportation Research Part A: Policy and Practice, 42*(1), 48–66.

Noland, R. B., & Ishaque, M. M. (2006). Smart Bicycles in an Urban Area: Evaluation of a Pilot Scheme in London. *Journal of Public Transportation, 9*(5), 71–95.

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.

Ortúzar, J. D. D., & Willumsen, L. G. (2011). *Modelling Transport.* Fourth edition. Chichester, United Kingdom: John Wiley & Sons.

Parasuraman, a, Zeithaml, V. a, & Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing, 49*, 41–50.

Parasuraman, A., Berry, L. L., & Zeithaml, V. A. (1991). Understanding Customer Expectations of Service. *Sloan Management Review, 32*(3), 39–49.

Parasuraman, A., Zeithaml, V. a, & Berry, L. L. (1988). SERQUAL: A Multiple-Item scale for Measuring Consumer Perceptions of Service Quality. *Journal of Retailing, 64*(1), 28.

Parasuraman, A., Zeithaml, V. a, & Berry, L. L. (1994). Alternative scales for measuring service quality: A comparative assessment based on psychometric and diagnostic criteria. *Journal of Retailing, 70*(3), 201–230.

Parkin, J., Wardman, M., & Page, M. (2008). Estimation of the determinants of bicycle

mode share for the journey to work using census data. *Transportation, 35*(1), 93–109.

Pedersen, T., Kristensson, P., & Friman, M. (2011). Effects of critical incidents on car users' predicted satisfaction with public transport. *Transportation Research Part F: Traffic Psychology and Behaviour, 14*(2), 138–146.

OBIS Project: Optimising Bike Sharing in European Cities. A Handbook. (2011). Funded under the Intelligent Energy Europe (IEE)

Pucher, J., & Buehler, R. (2006). Why Canadians cycle more than Americans: A comparative analysis of bicycling trends and policies. *Transport Policy, 13*(3), 265–279.

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.

Raveau, S., Álvarez-Daziano, R., Yáñez, M. F., Bolduc, D., & de Dios Ortúzar, J. (2010). Sequential and Simultaneous Estimation of Hybrid Discrete Choice Models. *Transportation Research Record: Journal of the Transportation Research Board, 2156*, 131–139.

Raveau, S., Yáñez, M. F., & Ortúzar, J. D. D. (2012). Practical and empirical identifiability of hybrid discrete choice models. *Transportation Research Part B: Methodological, 46*(10), 1374–1383.

Replogle, M. (1993). Bicycle access to public transportation: learning from abroad. *Transportation Research Record: Journal of the Transportation Research Board, 1396*, 75–80.

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.

Rietveld, P. (2005). Six reasons why supply oriented indicators systematically overestimate

service quality in public transport. *Transport Reviews, 25*, 319–328.

Rietveld, P., & Daniel, V. (2004). Determinants of bicycle use: do municipal policies matter? *Transportation Research Part A: Policy and Practice, 38*(7), 531–550.

Rojo, M., Dell'Olio, L., Gonzalo-Orden, H., & Ibeas, Á. (2013). Interurban bus service quality from the users' viewpoint. *Transportation Planning and Technology, 36*(7), 599–616.

Romero, J. P., Ibeas, A., Moura, J. L., Benavente, J., & Alonso, B. (2012). A Simulation-optimization Approach to Design Efficient Systems of Bike-sharing. *Procedia - Social and Behavioral Sciences, 54*, 646–655.

Saneinejad, S., Roorda, M. J., & Kennedy, C. (2012). Modelling the impact of weather conditions on active transportation travel behaviour. *Transportation Research Part D: Transport and Environment, 17*(2), 129–137.

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.

Tam, M. L., Lam, W. H. K., & Lo, H. P. (2010). Incorporating passenger perceived service quality in airport ground access mode choice model. *Transportmetrica, 6*(1), 3–17.

Thomas, T., Jaarsma, R., & Tutert, B. (2012). Exploring temporal fluctuations of daily cycling demand on Dutch cycle paths: the influence of weather on cycling. *Transportation, 40*(1), 1–22.

Transit Capacity and Quality of Service Manual. Third edition. Transportation Research Board.

Tyrinopoulos, Y., & Antoniou, C. (2008). Public transit user satisfaction: Variability and policy implications. *Transport Policy, 15*(4), 260–272.

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 Analyses Using Automated Data Collection Systems. *Journal of Public Transportation, 14*(4), 131–150.

Wen, C.-H., Lan, L. W., & Cheng, H.-L. (2005). Structural equation modeling to determine passenger loyalty toward intercity bus services. *Transportation Research Record: Journal of the Transportation Research Board, 1927*, 249–255.

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.

Winters, M., Friesen, M. C., Koehoorn, M., & Teschke, K. (2007). Utilitarian bicycling: a multilevel analysis of climate and personal influences. *American Journal of Preventive Medicine, 32*(1), 52–8.

Zeithaml, a, Berry, L., & Parasuraman, A. (1996). The behavioral consequences of service quality. *Journal of Marketing, 60*(2), 31–46.

Zeithaml, V. a, Berry, L. L., & Parasuraman, A. (1993). The nature and determinants of customer expectations of service. *Journal of the Academy of Marketing Science, 21*(1), 1–12.