Università degli Studi di Padova

Dipartimento di Scienze Statistiche

Corso di Laurea Magistrale in

Scienze Statistiche

TESI DI LAUREA MAGISTRALE

OPTIMIZATION MODELS FOR BIKE SHARING SYSTEMS

Relatore: prof. Luigi De Giovanni

Dipartimento di Matematica

Padova, Italia

Laureando: Marco Zanetti

Matricola n. 1180248

Anno Accademico 2019/2020

Alla mia famiglia,

e a tutti coloro che mi hanno fatto sentire parte della loro,

tra Este, Taizé, Verona, Rethymno, Padova e Amsterdam.

Abstract

The aim of this thesis is to contribute to make the planet a more sustainable place and to promote a healthy way of moving in the city. In this direction, we consider possible actions to improve Bike Sharing Systems (BSSs) with docking stations. In a BSS, especially during peak hours, in some stations the demand of bikes is higher than others. If no action is taken by the service provider, the stations may face a shortage of bike or docking stations, thus preventing other users from collecting or delivering bikes, respectively. The service provider has a variety of choices to balance the system, permanently or temporarily.

In this thesis, we define and present four optimization models and their application to reduce the disservices for the users. We propose an optimal scale up or scale down of the stations size to meet the users demand of docking stations under limited resources, and an optimal reallocation of the bikes at the beginning of the day to meet the users demand of bikes. Then, we examine the Static Bicycle Re-balancing Problem (SBRP), to re-balance the system during the night by moving bikes along the shortest route, and the Dynamic Bicycle Re-balancing Problem (DBRP), to re-balance the system during the day, taking into account the users demand. The models presented in the thesis are intended to minimize the violations that occur when the station is empty or completely full of bikes, inconveniences that causes a disservice to the users that want to pick-up or deliver a bike, respectively. The models are applied on a case study, the BSS of the city of Padua, in Italy. The main performance indicator is assessed in terms of total number of disservices for the users, to compare scenarios with different permanent or temporary decisions. The resources available, as the number of bikes, docking stations and the usage of a rebalancing vehicle, are limited. The objective is to assess the impact of a more

efficient use of these limited resources on the total number of disservices for the users.

In Chapter 1, we provide a summary of the key events of the bike sharing history, with its possible evolution. Then, we go more in details about basic concepts and metrics to be more familiar with BSSs. We see different terminologies used in a BSS station and we describe the logistic related to service vehicles and users operations. Furthermore, we see how the time, the users demand and various measures of performance are involved in the BSSs.

In Chapter 2, we see the main problems faced in the planning of a BSS at different levels (strategic, tactical, operational) and we recall the state of the art for each problem. We introduce problems at the strategic level, from the choice of the location for the stations to the number of bikes, docking stations and vehicles. Then, we focus on the problems that are faced in this thesis:

- the choice of the stations size;
- the choice of the number of bikes to set in each station at the beginning of the day, called stations state;
- the choice of the night route for the re-balancing vehicle;
- the choice of the daily route for the re-balancing vehicle.

Finally, we introduce the case study of the Padua BSS, in which we face the four problems above, based on the real data kindly provided for this thesis by the service provider Bicincittà S.R.L., with permission from the Municipality of Padua.

In Chapter 3, we introduce the basic concepts of mathematical programming, the modelling tool used to formalize and solve the optimization problems described in this thesis. In particular, we propose two new optimization models, original contribution of this thesis, to suggest a desirable number of parking spaces and bikes per station. Finally, we introduce the static optimization model proposed by *Dell'Amico et al. 2014* for the night re-balancing route and the dynamic optimization model proposed by *Contardo et al. 2012* for the daily re-balancing route.

In Chapter 4, we describe one of the main contribution of the thesis, namely the introduction of a decision-making process to guide in the configuration of a new BSS. The process is composed of various steps, that leads to the creation of scenarios in which the four optimization models introduced in Chapter 3 are applied, or not, to the real data of the BSS of Padua. For each scenario selected for the analysis, different performance indicators are proposed to compare the scenarios between them. We improve the dynamic optimization model proposed by *Contardo et al. 2012*, as to obtain a more parsimonious solution in terms of service vehicle's operations, original contribution of this thesis. We also introduce the rolling horizon methodology to update the route during the day as soon as new information becomes available. Finally, we propose 15 different scenarios that can be evaluated to determine an efficient use of the available resources, as bikes, docking stations and service vehicles.

In Chapter 5, we describe the technological tools used to analyse the data and implement the optimization models. Then, we explore the historical trips done in the Padua BSS for the period 2014-2018 and we describe the process to gather weather data and road distances among station. We provide some measures to compare the current performances of the Padua BSS with respect to the recommended level of performance indicated by the state of the art. We analyse the historical trips, exploring the features that impact the users activity during the day. Finally, we focus on the prediction models using the real data of the BSS of Padua to forecast the users demand of bikes in different time intervals of the day.

In Chapter 6, we apply the optimization models to find different configurations of the stations size, that can be standard, optimal or utopic, and we apply an optimization model to find the stations state at the beginning of the day, that can be standard or optimal, given the stations size. Finally, we define the parameters for the night and the daily route models, to have a complete setting for each of the 15 scenarios.

In Chapter 7, we evaluate the 15 scenarios using different performance metrics and we analyse the impact of the various models on the decrease of the total number of violations, with a detail of the performances per month, weekday, time interval or station. Finally, we analyse the rebalancing vehicle usage, in terms of kilometers per day, number of operations, number of visits in the stations and the time spent to rebalance the BSS in the different scenarios.

In Chapter 8, we conclude the thesis by summarizing contributions and results.

Sommario

L'obiettivo della tesi è di contribuire a soluzioni sostenibili per il nostro pianeta e promuovere un modo salutare di muoversi in città. In questa direzione, consideriamo possibili azioni per migliorare i sistemi di Bike Sharing (*Bike Sharing Systems - BSSs*) con postazioni fisse (*docking stations*). In un BSS, specialmente nelle ore di punta, in alcune stazioni la domanda è più alta che in altre. Se non viene intrapresa alcuna azione dal fornitore del servizio, le stazioni possono riempirsi o svuotarsi velocemente, impedendo ad altri utenti di prelevare o consegnare la propria bici. Il fornitore del servizio ha diverse opzioni a disposizione per bilanciare il sistema, di carattere permanente o temporaneo.

In questa tesi, definiamo e presentiamo quattro modelli di ottimizzazione e la loro applicazione per ridurre i disservizi per gli utenti. Proponiamo un ridimensionamento della capacità delle stazioni per soddisfare la domanda di docking stations, e una riallocazione ottima delle biciclette a inizio giornata per soddisfare la domanda di bici. Esaminiamo quindi lo Static Bicycle Re-balancing Problem (SBRP) per ribilanciare il sistema durante la notte utilizzando il percorso pù corto per spostare le bici, e il Dynamic Bicycle Re-balancing Problem (DBRP) per il ribilanciamento del sistema durante il giorno, tenendo in considerazione la domanda di bici. I modelli di ottimizzazione presentati nella tesi sono improntati a minimizzare le violazioni che si verificano quando una stazione è vuota o satura di bici, inconvenienti che causano un disservizio per gli utenti che vorrebbero prelevare o consegnare la bici, rispettivamente. I modelli sono applicati a un caso reale, il BSS della città di Padova. Le performance sono valutate in termini di numero totale di disservizi per gli utenti, per confrontare tra loro diversi scenari, con effetti permanenti o temporanei. Le risorse disponibili, come il numero di bici, docking stations e l'utilizzo di un veicolo

adibito al ribilanciamento, sono limitate. L'obiettivo è quello di valutare l'impatto di un uso più efficiente di queste risorse sul numero totale di disservizi per gli utenti.

Il Capitolo 1 è dedicato a un riepilogo degli eventi chiave della storia del bike sharing e alla sua possibile evoluzione. Inoltre, dedichiamo spazio ai concetti di base e alle varie metriche utilizzate in un BSS.

Il Capitolo 2 è dedicato a introdurre i principali problemi affrontati nella pianificazione di un BSS a diversi livelli (strategico, tattico, operativo) e a richiamare i principali lavori presenti in letteratura. Introdurremo diversi problemi a livello strategico, dalla scelta della posizione delle stazioni, alla scelta del numero di biciclette, docking stations e veicoli. In particolare, in questa tesi ci focalizziamo sui problemi di livello tattico e operativo, che sono:

- la scelta della dimensione delle stazioni;
- la scelta del numero ideale di bici all'inizio della giornata in ogni stazione;
- la scelta del percorso notturno del veicolo adibito al ribilanciamento;
- la scelta del percorso giornaliero del veicolo adibito al ribilanciamento.

Infine, introduciamo il caso studio del BSS di Padova, in cui vengono applicati dei modelli per i quattro problemi di cui sopra, basandoci sui dati storici gentilmente concessi per questa tesi da Bicincittà S.R.L. su autorizzazione del Comune di Padova.

Nel Capitolo 3, introduciamo i concetti di base della programmazione matematica e due nuovi modelli di ottimizzazione, contributo originale di questa tesi, per trovare un'indicazione sul numero desiderabile di docking stations e biciclette per stazione. Infine, introduciamo il modello di ottimizzazione statico proposto da *Dell'Amico et al. 2014* per il percorso di ribilanciamento notturno, e il modello di ottimizzazione dinamico proposto da *Contardo et al. 2012* per il percorso di ribilanciamento giornaliero.

Nel Capitolo 4, introduciamo un processo decisionale per guidare il fornitore del servizio nella configurazione di un nuovo BSS. Il processo decisionale è composto da diversi passaggi, che conducono alla creazione di uno scenario nel quale i quattro modelli di ottimizzazione introdotti nel Capitolo 3 vengono applicati, o meno, ai dati reali del sistema di Bike Sharing di Padova. Per ogni scenario selezionato per l'analisi, vengono misurati diversi indici di performance per confrontare gli scenari tra loro. Quindi, proponiamo una variazione al modello di ottimizzazione dinamico proposto da *Contardo et al. 2012*, per ottenere una soluzione più parsimoniosa in termini di numero di operazioni effettuate dal veicolo di servizio, contributo originale di questa tesi. Di seguito, introduciamo la metodologia del *rolling horizon* per aggiornare periodicamente il percorso durante il giorno quando saranno disponibili nuove informazioni. Infine, proponiamo 15 diversi scenari che possono essere valutati per determinare un uso efficiente delle risorse disponibili, come biciclette, docking stations e veicoli di servizio.

Nel Capitolo 5, descriviamo gli strumenti informatici utilizzati per analizzare i dati e implementare i modelli di ottimizzazione. Quindi, esploriamo i dati dei viaggi storici effettuati nel BSS di Padova nel periodo 2014-2018 e descriviamo il processo di raccolta dei dati meteorologici e del calcolo della distanza tra le varie stazioni. Di seguito, forniamo alcune misure per confrontare le prestazioni attuali del BSS di Padova rispetto al livello di prestazione ideale indicato nello stato dell'arte. Analizziamo i dati dei viaggi storici e le variabili che influenzano l'attività degli utenti durante il giorno. Infine, ci concentriamo sui modelli di previsione per stimare la domanda di bici utilizzando i dati storici del BSS di Padova.

Nel Capitolo 6, applichiamo i modelli di ottimizzazione al caso di Padova per trovare diverse configurazioni della capacità delle stazioni, che può essere standard, ottimale e utopica, e del numero di bici a inizio giornata, che può essere standard oppure ottimale, date le dimensioni delle stazioni. Infine, definiamo i parametri operativi per il modello di ribilanciamento notturno e giornaliero, per definire i parametri di ciascuno dei 15 scenari selezionati.

Nel Capitolo 7, valutiamo le performance dei 15 scenari nel BSS di Padova, utilizzando diverse metriche e analizzando l'impatto dei vari modelli nella riduzione del numero totale di violazioni. Un'analisi più approfondita viene fatta con un dettaglio per mese, giorno della settimana, intervallo di tempo o stazione. Infine, analizziamo le risorse utilizzate dal veicolo adibito al ribilanciamento, come i chilometri percorsi, il numero di operazioni o viaggi effettuati, e il tempo impiegato per ribilanciare il

BSS nei diversi scenari.

Nel Capitolo 8, concludiamo la tesi con una sintesi dei contributi e dei maggiori risultati ottenuti.

Contents

List of Figures

List of Tables

Chapter 1

Introduction to Bike Sharing Systems

In this chapter, we give a brief history of bike sharing and we introduce some key concepts and definition to better understand the functionality of the Bike Sharing Systems (BSSs) and the notation used throughout the thesis.

1.1 A brief history of Bike Sharing

As urbanization proceeds throughout the world, the demand for efficient and sustainable modes of transportation is constantly growing. Bike sharing fulfils these criteria for short distance travelling within city centres and consequently, bike sharing is getting increasing attention from both governments and the public. A BSS is a mobility service in which public bicycles, located at different stations across an urban area, are available for shared use. These systems contribute towards obtaining a more sustainable mobility and decreasing traffic and pollution caused by car transportation. This service can support the citizens in the city, but it needs to avoid the system unbalance, that occurs when stations are full or empty. Re-balancing operations, like the addition of some bikes to a station with a shortage of them, using a service vehicle, are often applied to support the system to meet the users demand of bikes. In this thesis, we study the functionalities of BSSs with docking stations, focusing on the allocation of docking stations and bikes in the network and on the operations of the service vehicles used to re-balance the system.

1.1.1 First generation

The first bike sharing program began in 1965 in Amsterdam, when a dutchman named Luud Schimmelpennink painted 50 ordinary bikes white and placed them throughout the city for the public use (*DeMaio 2009*). This concept is known as the first generation of bike sharing, and is characterized by bikes painted in a specific color without locks or stations, free of charge, and with no need for membership or registration. A bike could be picked up where found, ridden for as long as needed and then it could be left unlocked for someone else to use. Despite good intentions, the project was unsuccessful. After just a few days all bikes were either stolen, broken, or thrown in the canals (*DeMaio 2009*).

1.1.2 Second generation

In the beginning of the 1990s in Denmark, a few modest scale projects took place in two small cities (*Nielsen et al. 1993*) before the famous Bycyklen service began in Copenhagen in 1995 (*Shaheen et al. 2010*). The second generation of bike sharing was introduced and it had the following features:

- the system included docking stations in which bikes had to be locked, borrowed, and returned;
- customers had to pay a coin deposit when borrowing, which they got back when the bike was safely returned, preventing vandalism and theft.

However, in the second generation of BSSs, there was not the possibility to know the identity of the rider, so thefts were still active.

1.1.3 Third generation

The most important contribution of the third generation of BSS was the introduction of a form of membership, often including a magnetic card used to borrow bikes at the stations. This addition made it possible to keep track of each borrower, and combined with personal information and credit card information, theft became much less attractive. A couple of small membership-based bike sharing programs existed before 2005, but it was in Paris, with the system called *Velib'*, that the third generation got a foothold (*Shaheen et al. 2010*). Velib' was a huge success and got massive attention worldwide as one of the largest bike sharing systems in the world. With over 20,000 bicycles, this system covered the city with 1,800 bicycle stations distributed in an astonishing 300 meters distance between them. One Velib' (bicycle) is rented every second in Paris, coming down to about 86,400 rentals per day. A bicycle can be picked up at these stations by both short-term users and long-term subscribers. Journeys under 30 minutes are always free of charge and for the first additional half-hours you pay an increasing fee. The system is very flexible as users are able to place their bicycle at any of the other bicycle stations and do not have to bring the bicycle back to the station they originated from. When fixed docking stations are used the bicycles can only be locked and thus returned at bicycle stations found throughout the city.

1.1.4 Fourth generation

The fourth generation of bike sharing is related to the application of more advanced technologies. This generation makes use of smartphones and Internet, and may be considered a part of the development towards the so-called "Internet of things". The docking stations, and even each bike, may be connected to Internet, sending real time information to both users and operators using smartphone applications (*Lozano et al. 2018*). This new generation includes a dock-less Bike Sharing solution, in which bikes are equipped with a GPS tracker and they can be found and rented using a mobile application. Dock-less bikes consist of bicycles with a lock that is usually integrated into the frame and does not require a docking station. Due to the fact that this system does not require docking stations and it does not need infrastructures that may require city planning and building permissions, the system has spread rapidly on a global scale. These new systems offer a huge flexibility to the users, but they do not lack of problems of vandalism and massive accumulation of bikes in the most attractive areas of the city.

1.1.5 The evolution

In the last years, the number of BSSs worldwide has increased exponentially, from 17 to 1608 between 2005 and 2018 (*MetroBike 2011*), promoting the use of active modes of transport and decreasing dependency on the car, toward better air quality in the city. Hence, the effects of Bike Sharing Systems indicates a significant reduction in car usage (*Fishman et al. 2014*). Bike sharing has large effects on creating a larger cycling population, increasing transit use, decreasing greenhouse gases, and improving public health. The future of bike sharing is clear: there will be a lot more of it (*DeMaio 2009*). As the price of fuel rises, traffic congestion worsens, populations grow, and a greater world-wide consciousness arises around climate change, it will be necessary for leaders around the world to find new more environmentally sound, efficient, and economically modes of transport.

1.2 Basic definition and notation

1.2.1 Stations

A *station* is a hub in which users can find bicycles and docking stations. We denote by S the set of stations, by $s \in S$ a specific station and by S^N the total number of stations.

A *depot*, also called Station 0, is the store in which the service provider keeps the service vehicles and in which it repairs the damaged bicycles.

A *docking station* is a special bike rack that locks the bike, and only releases it by computer control. The bikes have also their own locker, that can be used for intermediary stops, but the user must conclude the trip locking the bike to a docking station. We denote by C_s^N the number of docking stations of station s, i.e. its capacity, and by C^N the total number of docking stations in the BSS, calculated as

$$C^N = \sum_{s \in S} C_s^N.$$

Finally, let B^N be the total number of bikes in the BSS.

A *docking stations BSS* is a public system for automatic or semi-automatic lending of bicycles (often called city bikes) within a restricted time period and area. A BSS is equipped by S^N stations, so a user can pick-up a bike at one station and deliver it to another station. This is called "A to B service" (but is also possible an "A to A service"). Using the bike may be free of charge, have a fee per hour, or have a monthly or yearly subscription fee.

The *Information Technology* system (IT), in a BSS, is the computer system of the service provider, in which, among the various activities, the users operations and the stations state are recorded. Monitoring the users operations and the stations state is a crucial part of the service provider activities. The control of the stations state is fundamental to understand if there are stations recurrently empty or full. This information will be useful in Section 3.2 and 3.3 to determine if some stations need more bikes or docking stations to meet the user demand. Also, the analysis of the users operation is important because the service provider can have a better understanding of the users demand, for each station, in different time intervals of the day. This information will be useful in Section 3.5 to determine the route of the vehicle during the day, taking into consideration the predictions of the users demand.

1.2.2 Time

In this thesis we consider the time as discrete, so the daily time is split in T^N time intervals $t \in \{1, ..., T^N\}$, as we need time intervals for the user demand analysis of Sections 6.6 and 6.7 and for the optimization models of Chapter 3. Let be T the set of the time intervals and let $t \in T$ be one time interval.

A *time interval* t includes few minutes in which events are considered as simultaneous. For example, if the time interval is a time window of 5 minutes, the day is split in 288 time intervals.

A *time period*, for the scope of thesis, is considered as a set D of days, where $d \in D$ refers to one day and D^N is the total number of days. The set D is composed of a set of working days D_W and a set of holidays D_H, where the set of holidays includes national holidays and weekends:

$$D = D_W \cup D_H.$$

1.2.3 Operations

Let U be the set of users, where $u \in U$ is a user and let R be the set of service vehicle, where $r \in R$ is a service vehicle.

One operation $O_{s,t,d}$ is an action that occurs in station s, at time t of day d.

A user operation

A *user operation* $O^u_{s,t,d}$ is done by a user $u \in U$ that delivers or picks-up a bike. $O^u_{s,t,d}$ values 1 when a user $u \in U$ deliveries or picks-up a bike, 0 otherwise.

A *user pick-up operation* $O^{uP}_{s,t,d}$ is an action that occurs in station s, at time t of day d, when a user u pick-up a bike from the station to start the trip.

A *user delivery operation* $O^{uD}_{s,t,d}$ is an action that occurs in station s, at time t of day d, when a user u delivers a bike to the station to finish the trip.

A vehicle operation

A *re-balancing operation* $O^r_{s,t,d}$ is done by the service vehicle $r \in R$ to pick-up or delivery of one or more bikes to re-balance station s. $O^r_{s,t,d}$ values the number of bikes added or removed from the station by the vehicle r, or 0 if no operation is done.

A *service vehicle pick-up operation* $O^{rP}_{s,t,d}$ is a re-balancing action that occurs in station s, at time t of day d, when a service vehicle r picks-up one or more bikes from the station.

A *service vehicle delivery operation* $O^{rD}_{s,t,d}$ is a re-balancing action that occurs in station s, at time t of day d, when a service vehicle r delivers one or more bikes to the station.

1.2.4 Users demand

Let $U^{P}_{s,t,d}$ be the set of users $u \in U$ that picked-up a bike in station s, at time t of day d. The pick-up demand $D^{P}_{s,t,d}$ refers to the total number of users $u \in U^{P}_{s,t,d}$ that pick-up a bike from station s:

$$D^{P}_{s,t,d} = \sum_{u \in U^{P}_{s,t,d}} O^{uP}_{s,t,d}$$

where $O^{uP}_{s,t,d} \in \{0,1\}$ values 1 if a user u pick-ups a bike, 0 otherwise.

Let $U^{D}_{s,t,d}$ be the set of users that deliver a bike in station s, at time t of day d. The delivery demand $D^{D}_{s,t,d}$ refers to the total number of users $u \in U^{D}_{s,t,d}$ that deliver a bike from a station s:

$$D^{D}_{s,t,d} = \sum_{u \in U^{D}_{s,t,d}} O^{uD}_{s,t,d}$$

where $O^{uD}_{s,t,d} \in \{0,1\}$ values 1 if a user u delivers a bike, 0 otherwise.

The *users activity*, or" pick-up plus delivery" demand $D^{P+D}_{s,t,d}$, is a descriptive statistic useful to understand the activity of the BSS in different time intervals of the day. It can be calculated as:

$$D^{P+D}_{s,t,d} = D^{P}_{s,t,d} + D^{D}_{s,t,d}.$$

The *users demand* refers to the net demand of bikes or docking stations $D_{s,t,d}$ in station s, at time t of day d. The users demand is the difference between the delivery demand and the pick-up demand:

$$D_{s,t,d} = D^D_{s,t,d} - D^P_{s,t,d}.$$

If $D_{s,t,d} > 0$, it means that the delivery demand is higher than the pick-up demand, so more users finish the trip to the station than the ones that begin the trip from that station. If $D_{s,t,d} < 0$, it means that the delivery demand is lower than the pick-up demand, so less users finish the trip to the station than the ones that begin the trip from that station. If $D_{s,t,d} = 0$, it means that, at time t of day d, station s is perfectly balanced between delivery and pick-up demand. Figure 1.1 shows an example of the users demand of a specific station s.

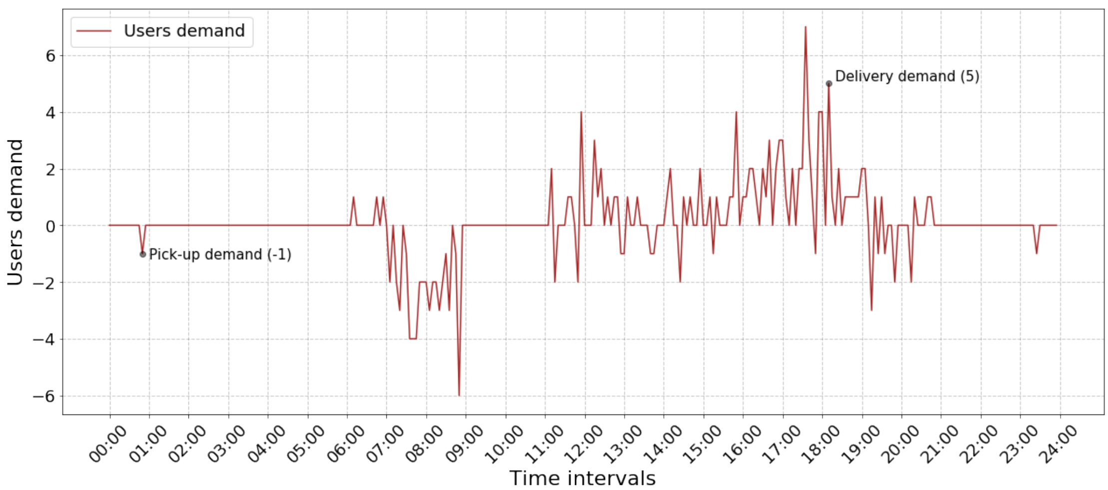

Figure 1.1: An example of the users demand of a specific station s in the course of one day.

The *cumulative users demand* $D^{cum}_{s,t,d}$ is a quantity that can be helpful to see if station s, at time t of day d, tends to be unbalanced. $D^{cum}_{s,t,d}$ can be calculated as the cumulative users demand in the station from the beginning of the day, when t is equal to 1, to the current time interval t:

$$D^{cum}_{s,t,d} = \sum_{t=1}^{t} D_{s,t,d}.$$

The *minimum cumulative users demand* $D^{cum;\min}_{s,d}$ is the minimum cumulative demand in the course of day d for station s:

$$D^{cum;\min}_{s,d} = \min_{t \in T} D^{cum}_{s,t,d}.$$

The *maximum cumulative users demand* $D^{cum;\max}_{s,d}$ is the maximum cumulative demand in the course of day d for station s:

$$D^{cum;\max}_{s,d} = \max_{t \in T} D^{cum}_{s,t,d}.$$

Figure 1.2 shows an example of the maximum and minimum cumulative users demand of a specific station s.

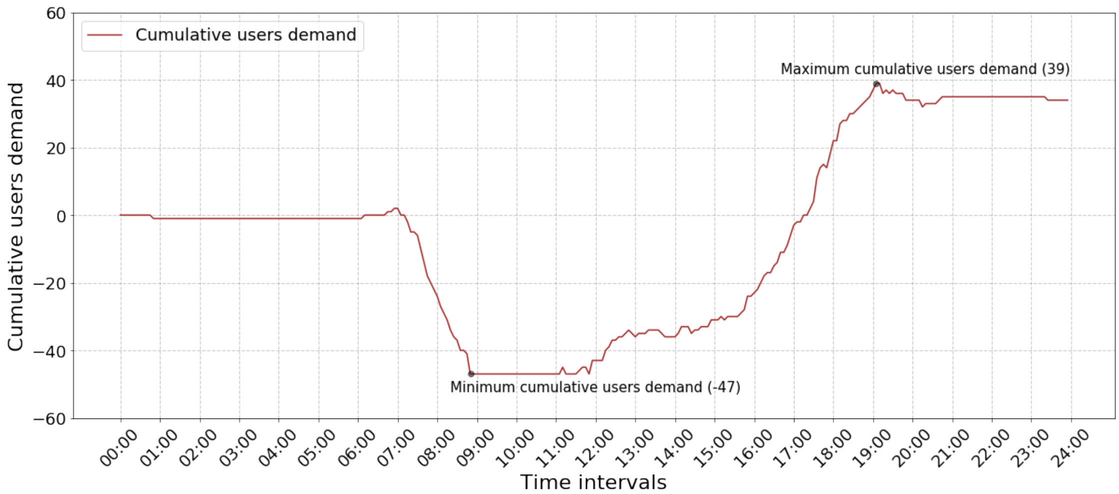

Figure 1.2: An example of the maximum and minimum users demand of a specific station s in the course of one day.

The *users demand forecast* refers to the prediction of the users demand done in the tactical planning, when the BSS is already active in the city. The users demand forecast is based on the historical users demand and can lead to a better understanding of the users behaviour in the different time interval of the day.

The *censured demand* refers to the users demand that the system was not able to register. When a user needs to pick-up a bike but there is a shortage of bikes in the station, the user chooses another way to go to the destination, but the IT system will not record any failed pick-up operation. Likewise, when a user rides to destination and there is a shortage of docking stations in that station, the user needs to deliver the bike to another station, but the IT system will not record the failed delivery operation in the first station chosen.

1.2.5 Station state

Let $L_{s,t,d}$ be the number of bikes in station s, at time t of day d, and let B^N be the total number of bikes in the BSS.

The *station state*, or station level, is the current number of bikes at the station. For the purpose of this thesis, three different cases can occur in a station:

- state lower than 0: the station state can be negative when the station is empty and, virtually, one or more users pick-up a bike;
- state between 0 and the station capacity: the station state between these boundaries does not report any problem;
- state greater than the station capacity: the station state can be greater than the station capacity when the station is full and, virtually, one or more users delivery a bike.

The *initial station state* $L_{s,1,d}$ is the number of bikes at station s at the beginning of the day d, i.e. when the time interval t is equal to 1.

The *final station state* $L_{s,T^N,d}$ is the number of bikes at station s at the end of the day d, i.e. when the time interval t is equal to T^N.

The *current station state* $L_{s,t,d}$ is the number of bikes at station s in a determined time interval t. The current station state can be calculated as the initial station state $L_{s,1,d}$ plus the cumulative users demand at the time interval t:

$$L_{s,t,d} = L_{s,1,d} + D^{cum}_{s,t,d}.$$

The *docks per bike ratio* R^{DpB} is the ratio between the total number of docking stations and the total number of bikes. In a BSS, the number of docking stations should be larger than the number of bikes in the system, to ensure enough parking spaces available to deliver bikes at the end of the trip. According to the *Bike Sharing Planning Guide* (*Gauthier 2013*), the docks per bike ratio should be a number between 2 and 2.5. We can calculate the ratio between the total number of docking stations and the total number of bikes available in the BSS, R^{DpB}, as:

$$R^{\mathrm{DpB}} = \frac{C^N}{B^N}$$

where C^N is the total number of docking stations (see Section 1.2) and B^N is the total number of bikes in the BSS.

The *standard initial station state* $L^S_{s,1,1}$ is defined as the minimum value between

$$\left\lfloor \frac{C^N_s}{R^{\mathrm{DpB}}} \right\rfloor$$

and

$$\left\lfloor \frac{C^N_s}{2.25} - 0.5 \right\rfloor + 1.$$

Hence, the *standard initial station state* $L^S_{s,1,1}$ is calculated as:

$$L^S_{s,1,1} = \min\left(\left\lfloor \frac{C^N_s}{R^{\mathrm{DpB}}} \right\rfloor, \left\lfloor \frac{C^N_s}{2.25} - 0.5 \right\rfloor + 1 \right) \tag{1.1}$$

This method can be seen as a standard criteria used by the service provider to relocate the bikes between the stations, according to *Bike Sharing Planning Guide* (*Gauthier 2013*), ensuring that the proportion of bikes respect to the docking stations is a number close to 2.25, according to *Bike Sharing Planning Guide* (*Gauthier 2013*).

The *optimal initial station state* $L^O_{s,1,d}$ can be seen as an upgrade of $L^S_{s,1,d}$, when the expected users demand is taken into consideration. $L^O_{s,1,d}$ is the optimal number of bikes at station s at the beginning of the day d, to better meet the users demand of bikes and parking spaces, taking system resources into account.

The *minimum station state* $L^{\min}_{s,d}$ is the minimum number of bikes in station s of day d:

$$L^{\min}_{s,d} = \min_t L_{s,t,d} = L_{s,1,d} + D^{cum;\min}_{s,d}.$$

The minimum daily station state can be a useful measure to understand if that stations needs to be enlarged in size or if it needs less or more bikes in the station at the beginning of the day.

The *maximum station state* $L^{\max}_{s,d}$ is the maximum number of bikes present in station s of day d. In this case, we assume an unbounded station size:

$$L^{\max}_{s,d} = \max_t L_{s,t,d} = L_{s,1,d} + D^{cum;\max}_{s,d}.$$

The maximum daily station state can be a useful measure to understand if that stations needs to be enlarged in size or if it needs less or more bikes in the station.

1.2.6 Station size

The *station size* C^N_s is the capacity of station s, i.e. the number of docking stations installed in that station. The station size in the docking stations BSS is fixed, although, in some BSS of new generation, it can be flexible.

The *standard station size* C^{NS}_s is the capacity actually installed in station s of the BSS.

The *optimal station size* C_s^{NO} is the optimal capacity of station s, given a certain number of docking stations available in the BSS.

The *required station size* $C_{s,d}^{NR}$ is the size needed to meet all the users demand, for station s in day d. This quantity is obtained as the difference between the maximum station state $L_{s,d}^{\max}$ and the minimum station state $L_{s,d}^{\min}$:

$$C_{s,d}^{NR} = L_{s,d}^{\max} - L_{s,d}^{\min}.$$

The *extra required docking stations* $C_{s,d}^{NE}$ are the extra docking stations that should be added to station s in day d to reach the required station size:

$$C_{s,d}^{NE} = \max(C_{s,d}^{NR} - C_s^N, 0).$$

The *utopic station size* C_s^{NU} is the capacity of station s such that no extra docking stations are required. It can be calculated as the standard station size plus the maximum extra required docking station in a certain time period of days D.

$$C_s^{NU} = C_s^N + \max_d C_{s,d}^{NE}.$$

Figure 1.3 shows an example of the the size of a specific station s in the course of one day, with the related required and extra required docking stations.

1.2.7 Service vehicles

We recall that R is the set of vehicles (see Section 1.4). Let R^N be the total number of vehicles and let a re-balancing operation be the addition or removal of bikes in a station by a service vehicle, in case of shortage of docking stations or bikes, respectively. Re-balancing operations are needed when the system is not able to

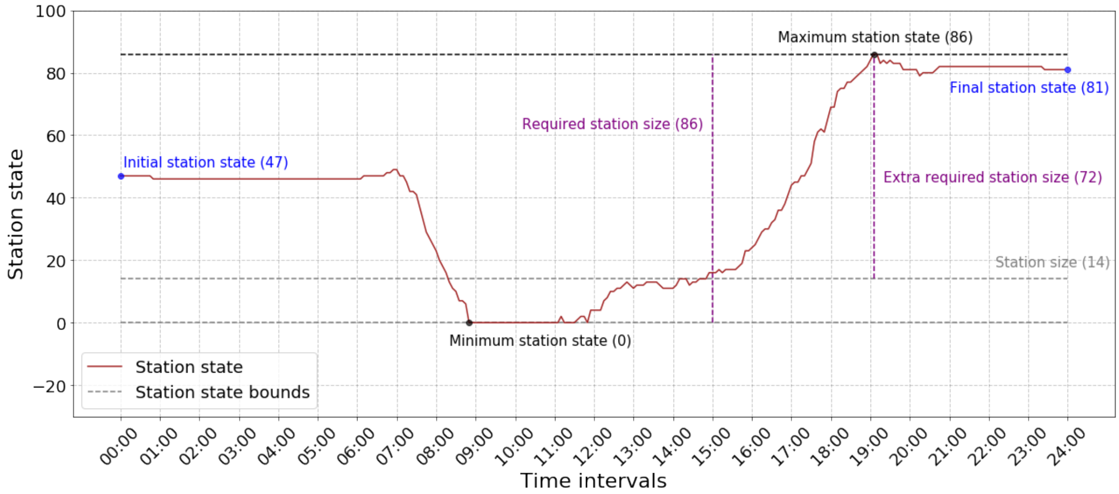

Figure 1.3: An example of the size of a specific station s in the course of one day, with the related required and extra required docking stations.

meet a high percentage of the users demand. In this situation, the system is called unbalanced.

A *service vehicle* $r \in R$, also called re-balancing vehicle, is a mode of transport, usually a small truck, used to re-balance the BSS. A service provider can have a fleet of one or more service vehicles.

The *vehicle capacity* C^r, is the maximum number of bikes that vehicle r can load. Usually, the service vehicle leaves the depot to start the re-balancing route with some bikes already carried on it.

The *vehicle initial load* L_0^r is the number of bikes carried by vehicle $r \in R$ when it starts the route.

The *vehicle load* L_s^r is the number of bikes carried by the vehicle after visiting station s.

1.3 Measures and Performance indicators

The deviation of the stations state

For a station s and a day d, a deviation $q_{s,T^N,d}$ is the difference between the final station state $L_{s,T^N,d}$ and the optimal or standard initial station state for the beginning of the next day $L_{s,1,d+1}$:

$$q_{s,T,d} = L_{s,T^N,d} - L_{s,1,d+1}.$$

It is useful to know the deviation of a station to better understand if there is a necessity of an addition or removal of bikes.

The violations in a station

The number of *empty violations* $V^E_{s,t,d}$ is a measure that evaluates the number of potential users that did not benefit from the BSS service, as a shortage of bikes occurred in station s at time t of day d. This measure is used to understand if there is a necessity of more bikes or docking stations at the stations. Empty violations $V^E_{s,t,d}$ occur when the station is empty and one or more users would like to pick-up a bike. $V^E_{s,t,d}$ is equal to the absolute value of the minimum between the station state $L_{s,t,d}$ and 0:

$$V^E_{s,t,d} = |\min(L_{s,t,d}, 0)|$$

where:
$s \in S$ refers to the station;
$t \in \{1, ..., T^N\}$ refers to the time interval;
$d \in D$ refers to the day.

The *maximum empty violation* $V^{E;\max}_{s,d}$ is the maximum empty violation that occurs during a day d in station s, given the minimum station state $L^{\min}_{s,d}$ (as defined in Section 1.6):

$$V^{E;\max}_{s,d} = |\min(L^{\min}_{s,d}, 0)|.$$

A full violation occurs at station s, and at time t of day d, when the station is full and one or more users try to deliver a bike. The number of *full violations* $V^F_{s,t,d}$ is equal to the maximum value of the difference between the station state $L_{s,t,d}$ and the station size C^N_s, and 0:

$$V^F_{s,t,d} = \max(L_{s,t,d} - C^N_s, 0).$$

The *maximum full violation* $V^{F;\max}_{s,d}$ is the maximum full violation that occurs during a day d in station s, given the maximum station state $L^{\max}_{s,d}$ and the station size C^N_s:

$$V^{F;\max}_{s,d} = \max(L^{\max}_{s,d} - C^N_s, 0).$$

The *total number of violations* $V^T_{s,t,d}$ that occur in station s, at time t of day d, is equal to the sum of the empty violations $V^E_{s,t,d}$ and the full violations $V^F_{s,t,d}$:

$$V^T_{s,t,d} = V^F_{s,t,d} + V^E_{s,t,d}.$$

The *maximum total violation* $V^{T;\max}_{s,d}$ is the sum of the maximum empty violation and the maximum full violation that occurs during a day d in station s:

$$V^{T;\max}_{s,d} = V^{E;\max}_{s,d} + V^{F;\max}_{s,d}.$$

The *final number of violations* V are the violations that occur in a defined time period of several days. The final number of violation is one of the key performance indicators that is taken into account to compare different scenarios that will be introduced in Chapter 5:

$$V = \sum_{d \in D} \sum_{s \in S} \sum_{t \in T} V_{s,t,d}.$$

Figure 1.4 shows an example of the state of a specific station s in the course of one day, with the related maximum full and empty violations.

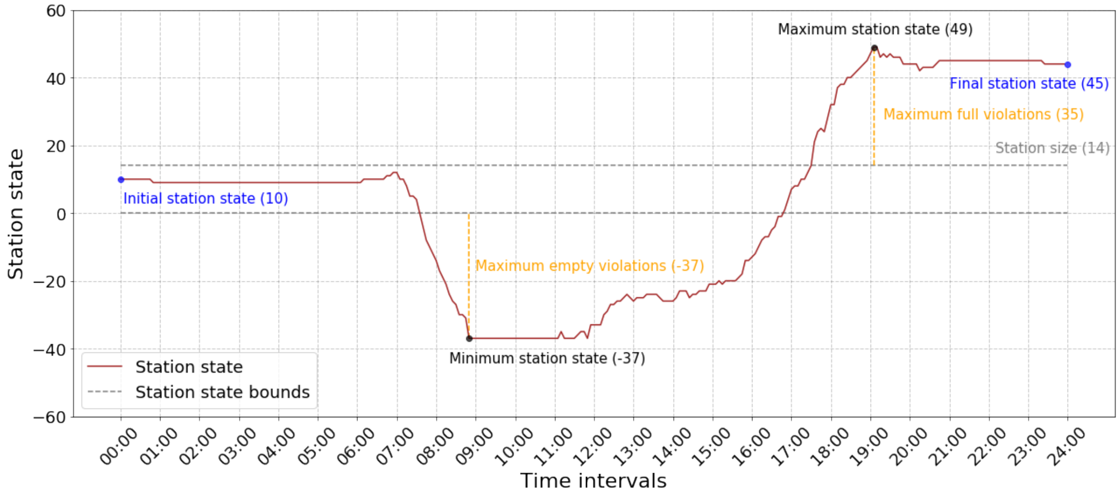

Figure 1.4: An example of the state of a specific station s in the course of one day, with the related maximum full and empty violations.

Chapter 2

Optimization problems in Bike Sharing Systems and state of the art

We dedicate this chapter to the exploration of various problems faced in a BSS, introducing the literature that contributed to solve these problems.

2.1 Optimization problems in Bike Sharing Systems

Several challenges are faced in a docking stations BSS and can be formulated as optimization problems. We can group the problems faced in a BSS in three hierarchical planning levels: strategic, tactical and operational (*Neumann-Saavedra et al. 2015*). We will use the label P_n^l to distinguish the problems, where $l \in \{S, T, O\}$ identify the planning level (Strategic, Tactical or Operational) and n enumerates problems on the same planning level. The classification of the planning levels is illustrated in Figure 2.1: the first planning level is the strategic level, when the BSS is not yet active, then we have the tactical level to find the optimal allocation of docking stations and bikes, and we finish the planning with the operational level, where we re-balance the system using service vehicles.

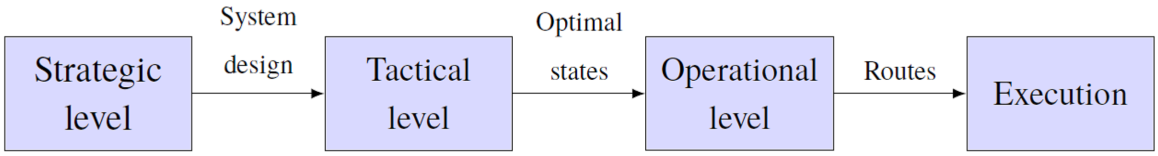

Figure 2.1: The planning levels of a docking stations BSS (*Kristianslund et al. 2016*).

2.2 Strategic level

The strategic level contains the challenges to face when constructing the BSS and a big role is played by the political parties. A successful BSS needs a strong political support and coordination between the city agencies (*Gauthier 2013*). Below, we introduce the main challenges to face at the strategic level.

2.2.1 P_1^S: determining the number of stations

Let S^N be the number of stations. The stations of the BSS should cover the more populated areas of the city but at the same time they should be close between each other to ensure convenience and reliability for the users. Generally, small pilots of BSS are not successful because there is a limit of usability due to the poor coverage area or number of bikes. According to the *Bike Sharing Planning Guide* (*Gauthier 2013*), a BSS is recommended to locate between 10 and 16 stations per km^2. It is important to create a dense network to offer a complete service to the potential users. *Lin et al. 2011* create a model that attempts to determine the number and location of docking stations, using a network structure of bike paths connected between the stations and travel paths for users between each pair of origins and destinations.

2.2.2 P_2^S: determining the location of the stations

The second challenge to face in the planning of a BSS is to define the coverage area where the BSS will be operative. Usually, the choice of the location of stations for the BSS area goes simultaneously with the choice of the number of stations S^N and the choice in this first phase is to select the areas with the more residential population, where there is already a strong cycle infrastructure. Also, the stations should be near to other transportation hubs. According to the *Bike Sharing Planning Guide* (*Gauthier 2013*), a BSS is recommended to cover at least 10 km^2. In this phase it is crucial to locate the stations in the key areas of the city and to guarantee the connection between them, so that the potential users are more encouraged to use the service. *Romero et al. 2012* present a method that considers the interaction between private cars and public bicycle transport modes to optimize the location of docking stations in a BSS. *Martinez et al. 2012* use an optimization algorithm to establish the location of the docking stations in the city to maximise the net revenue in a case study for the BSS of Lisbon. According the *Bike Share Station Siting Guide* (*City Transportation Officials 2016*), ensuring that bicycle stations are placed within 3-5 minute walking distance of one another, throughout a dense network, is the guarantee for a successful and sustainable Bike Sharing program. *Garcıéa-Palomares et al. 2012* use a GIS-based method to calculate the spatial distribution of the potential demand for trips to locate stations using location - allocation models.

2.2.3 P_3^S: determining the number of bikes

Let B^{NS} be the number of bikes in the BSS. Before deciding how many bikes to add into the system, an accurate analysis of the potential demand needs to be done. According to the *Bike Sharing Planning Guide* (*Gauthier 2013*), a BSS is recommended to have between 10 and 30 bikes for every 1,000 residents within the coverage area, but also the high number of commuters and/or tourists needs to be taken into account. Once understood the potential market, a metric to monitor is the daily number of trips per bike. If the number of trips per bike is too high, the system has too few bikes to meet demand. This problem leads to disservices and a smaller impact on the city's objectives. On the other side, if the number of trips per bike is too low, the service is oversized and the BSS could be seen as a wrong investment by the citizens. *Sayarshad et al. 2012* propose a mathematical model to find the minimum required bikes fleet size that simultaneously minimizes unmet demand, unused bikes, and the need to transport empty bikes between rental stations to meet demand.

2.2.4 P_4^S: determining the number of docking stations

Let C^{NS} be the number of docking stations in the BSS. Usually, the service provider would like to ensure that there will be enough bikes and enough docking stations in every station. According to the *Bike Sharing Planning Guide* (*Gauthier 2013*), a BSS is recommended to have a docks per bike ratio R^{DpB} between 2 to 2.5 (see Section 1.6). Hence, the choice of the number of docking stations C^{NS} in P_4^S is a direct consequence of the choice of the number of bikes B^{NS} in P_3^S. It is important to ensure enough docking stations for each station to minimize the violations (see Section 1.9) that may occur, but the cost-benefit of adding more docking stations in the BSS should be also considered.

2.2.5 P_5^S: determining the starting stations size

Let C_s^{NS} be the size of station s, i.e. the number of docking stations in station s (see Section 1.2.6). The direct action that follows the choice of C^{NS} is the distribution of the docking stations C_s^{NS} for each station s. The station size should be proportional to the potential demand of trips for that station. The factors to consider are, for example, the density of population in the area or the presence of facilities such as universities, offices and restaurants. Another important factor to consider is the presence of train or bus stations close to the bicycle stations, that may significantly increase the potential users demand. A problem that can occur in a BSS is to undersize the stations in the lower density residential area, while many commuters would like to end their trips there to reach facilities. According to the *Bike Sharing Planning Guide* (*Gauthier 2013*), a consistent uniform station size or at least a minimum station size for all the stations is crucial to create a system in which the users can rely on. The stations size should vary from having ten docking stations in the lower density area to hundreds of parking spaces per station in the very high density area. In a docking stations BSS in which the stations are not modular, a wrong decision of the stations size may have a large impact on the lost demand.

Instead, the advantage of a modular station is that, once a station is built, the station can easily be relocated to a place with higher demand (*Gauthier 2013*), or, if the station is too small in size, it is possible to scale it up.

2.2.6 P_6^S: determining the starting station state

Let $L_{s,1,1}$ be the starting state of station s (see Section 1.2.5). Once defined the total number of bikes B^N and the stations size C_s^N, the problem to face is to find the correct number of bikes for every station that minimizes the potential disservices for the users, i.e. the times in which there is a shortage of bikes or docking stations. At the strategic planning, when the system is not yet active, it is not possible to calculate the expected users demand for every time interval and no studies have been found for the choice of the starting stations state. Hence, we can impose as starting station state the standard initial station state (see Section 1.6), and this choice can be updated when historical data become available.

2.2.7 P_7^S: determining the number of service vehicles

Let R^N be the number of service vehicles (see Section 1.2.7). In a BSS, one or more service vehicles are critical to ensure the redistribution of the bikes between the stations. This operation is called re-balancing operation and it is one of the greatest challenges of a BSS to ensure the minimization of the lost users demand. Surprisingly, no studies have been found on the ideal size of the fleet of service vehicles, which should be, at a first analysis, roughly proportional to the number of stations to serve.

2.2.8 P_8^S: the maximum number of users

Usually, the users of a BSS are split in two categories: the occasional users and the long-term users, based on their subscription is shorter or longer than seven days. According to *Shaheen et al. 2013*, it is recommended to have 10 long-term users for every bike in the system to create a well-functioning system and, according to the *Bike Sharing Planning Guide* (*Gauthier 2013*), it is recommended to have about

4 to 8 daily users for every bike in the system. The maximum number of users is strongly related with the number of bikes in the system. If the number of daily users per bike is larger than 8, it means that there are few bikes to meet the users demand, so it is recommended to invest in the infrastructure, adding more stations and bikes in the system. If the number of daily users per bike is lower than 4, it means that the investment in the BSS may give a low economic return. A solution may be the limitation of the number of subscriptions or the increase of the number of bikes in the BSS.

2.3 Tactical level

At the tactical level, the parameters from the strategic level are already set and the BSS is considered active. The objective, at this stage, is to find the desired number of docking stations and bikes among the stations, taking the historical users demand into account.

2.3.1 P_1^T: user demand forecast

The user demand forecast is the prediction of the number of users that pick-up or delivery a bike from a station s in each time interval t of the day d. External factors affect the bikes usage as the lack of bikeways, weather conditions or season factors. Furthermore, the demand varies depending on the day of the week, special events, national public holidays, steep slopes or availability of other inter-modal services, as train and bus stations. An accurate forecast of the users demand is crucial in the daily re-balancing operations to anticipate the problems that may occur in some stations. For example, if a station is almost full and the forecast suggests that other users are likely to deliver bikes in that station in the next time intervals, that station needs to be emptied to avoid full violations. A real-time-monitoring based method cannot tackle the problem because it may be too late to re-balance the stations after observing the full or empty violations. According to *Li et al. 2019*, an accurate prediction of the bike usage in the city is very challenging for three reasons:

- large fluctuations of the demand: the human behaviours can largely depends on the time, the meteorology or special events, as traffic accidents or music festivals, for example;

- randomness: a large part of the historical trips is random because to reach a certain area of the city there are different combinations of starting and ending stations that are possible to ride;

- unbalanced external factors: rare events, as the snow hours, a festival, or a wind speed concentrated in a small range of time, are events that are difficult to predict and this problem can lower the accuracy of the predictions.

According to *Goh et al. 2019* other two problems to take into consideration are:

- demand censoring: it occurs when, in some time intervals, stations are out of bikes or docking stations available, so the demand is apparently null. In the course of the thesis we refer to this problem as an empty violation;

- choice substitution: it occurs when a user may decide to pick-up or delivery a bike to a station next to the one that had problems or when a user decides for another mode of transportation. In the course of the thesis we refer to this problem as a full violation.

Hulot et al. 2018 reduces the problem using a singular value decomposition to predict behaviors of the aggregate stations instead of the specific stations, and they apply different machine learning models to predict the users demand for each hour of the day. *Lin et al. 2018* build a convolutional neural network model that can learn hidden heterogeneous pairwise correlation between stations. *Goh et al. 2019* estimate the demand taking into account the choice substitution effects. *Li et al. 2019* propose a hierarchical approach to forecast the entire city demand, cluster of stations and finally, each specific station.

2.3.2 P_2^T: determining the optimal stations size

Let C_s^{NO} be the optimal size of station s. Given a set of stations and a limited number of docking stations $C^{NO} = C^{NS}$, the challenge of P_2^T is to find the optimal size C_s^{NO} for each station s to minimize the re-balancing actions that are needed when there is a shortage of bikes or docking stations. According to *Büttner et al. 2011*, the re-balancing costs account for as much as 30 percent of operating costs in European BSSs (between 1500 and 2500 € per bike each year). For the purpose of this thesis, the stations are assumed to be modular, i.e. equipped with the possibility to be resized, in the medium term, to adapt to possible new trends of the users demand. Also, the flexibility of the station size can be crucial to promptly fix possible mistakes of the first configuration of the station size (C_s^{NS}), if there is a mismatch between the potential and the real users demand. When the analysis of the users demand shows that there are systematic problems of shortage of bikes or docking stations in the course of the day, the service provider could consider to resize the stations. For example, if the size of a station is small and the users demand is high, the station will be quickly emptied or filled of bikes, causing a disservice for the users. Utopically, we would like to have stations with an infinite size to support every possible user delivery of bikes in the most frequented points of the city, but in a world with limited resources we have to adapt to the urban and economic constraints of the city and to optimize the distribution of the docking stations with reasonable criteria. We can consider the challenge to update the stations size as a problem that should be solved for example, once a year, or when it is crucial for the service provider. An utopic stations size C^{NU} may be calculated in absence of constraints related to the total number of docking stations available.

Fricker et al. 2016 study the stations where most problems arise in the BSS of Paris, quantifying the influence of changing the station capacities, and computing the optimal number of bikes in the system that minimizes the proportion of problematic stations, i.e. the stations that are often empty or completely full. According to *Fricker et al. 2016*, the docks per bike ratio R^{DpB} should be around 2 plus a few more docking stations in each station.

2.3.3 P_3^T: determining the optimal initial stations state

Let $L_{s,1,d}^{O}$ be the optimal initial state of station s. Then, let B^{NS}, B^{NO} and B^{NU} be the total number of bikes in the system with the standard, optimal, or utopic station size, respectively. Given the stations size C_s^{NS} (or C_s^{NO}, if the stations size have been updated), and the total number of bikes B^{NS} (or B^{NO}), the challenge to face in this section is to find the optimal allocation of the bikes among the stations at the beginning of the day, to reduce the unbalance of the BSS. Hence, the objective is to have the stations ready in the early morning to meet the users demand during the course of the day. In particular, the objective is to meet the pick-up users demand with enough bikes and the delivery users demand with enough docking stations available. Utopically, we would like to have infinite bikes to meet all the pick-up users demand, but in a world with limited resources, we have to obtain the more efficient allocation of the bikes among the stations. Hence, we can consider finding $L^O_{s,1,d}$ as a problem that should be considered every day, adding or removing bikes during the night, when the users activity is low. The optimal initial stations state could be also calculated in absence of resources constraints, i.e. given the utopic stations size C^{NU} and the utopic number of bikes B^{NU}, to have an utopic performance bound.

Vogel et al. 2011 use data mining and knowledge discovery in databases to gain insight into the bike activity patterns and use this information to find the optimal station state. *Raviv et al. 2013a* introduce a user dissatisfaction function (UDF) based on the occurrence of violations. Using discrete time periods and a Poisson process gives an approximation of the UDF that can be used to find optimal station state in the case study of the BSS of Tel Aviv. According to *Raviv et al. 2013a*, in the long term the number of bicycles that are available should be increased by ordering more bicycles, since the cost of the bicycles in a BSS is small compared to the infrastructure cost. *Schuijbroek et al. 2017* model stochastic demand using a queuing system to obtain the optimal state that minimizes the re-balancing costs. *Parikh et al. 2015* develop Mixed Integer Program model to predict the optimal stations state using the historical trips of the BSS of Antwerp. The users demand at the stations is modelled as a Markov process and the expected penalty values are calculated for the different initial stations state. The objective is to minimize the expected violations for all stations over the planning horizon. *O'Mahony et al. 2015* analyse the users demand in the BSS of New York and group the stations into clusters by similar distribution of the users demand. Then, they label each cluster with a desired level of bikes.

2.4 Operational level

Finally, when the strategical and tactical configuration of the system are given, in the operational level the re-balancing of the system needs to be planned. Hence, the operational level includes the challenges to face in the course of the day to ensure the minimization of the disservices for the users and, in the night, to bring back the stations state to a desired level. In many cases, to ensure the service, a redistribution of the bikes has to be done by the service provider, with one or more vehicles that load and unload bicycles from the full stations to the empty ones. Situations where the stations are empty or full have to be avoided because a user wants to rely on a system in which he/she finds a bicycle to start the trip and available docking stations to finish it (*Alvarez-Valdes et al. 2016*). Therefore, operators may also measure the fraction of time that the stations are full or empty (*Schuijbroek et al. 2017*). In most BSSs, the re-balancing happens both during the day, when the bikes are used, and/or during the night when the system is closed, or scarsely used.

The problem of defining the trips of a service vehicle to re-balance a BSS can be seen as a Vehicle Routing Problem (VRP). In accordance with *Toth et al. 2014* and *Psaraftis et al. 2016*, we classify the VRPs as either static or dynamic, and stochastic or deterministic, as we can see in Figure 2.2. A VRP is dynamic if the input of the problem is received and updated concurrently with the determination of the route. If all problem inputs are received before route determination, the VRP is static. A VRP is deterministic if all inputs are known with certainty and there are no stochastic inputs. The stochastic VRP on the other hand, is characterized as a VRP where one or more parameters are stochastic, i.e. some future events are represented by random variables with given probability distributions. A solution contains optimized routes for all service vehicles, including the number of bikes that should be picked up and delivered at each station. Using graph theory, each station could be considered as a node, and the trips between each station pair as an arc.

	Deterministic	Stochastic
Static	*Static and Deterministic.* All input parameters are known with certainty and in advance, and are assumed not to change during operation. The problem may be solved once and before the beginning of the planning period	*Static and Stochastic.* Some input parameters are random or stochastic, and the actual values are revealed during the execution of the routing process.
Dynamic	*Dynamic and Deterministic.* Some or all input data are not known beforehand, but become available over time. Only probabilistic information is available for future events. Optimization can only be performed as new information arrives.	*Dynamic and Stochastic.* Dynamic problems where part of the unknown input data is in the form of stochastic information (e.g., forecasts, range values, and prescribed distributions).

Figure 2.2: The VRP classification (*Toth et al. 2014*).

2.4.1 P_1^O: the optimal night route

Given a set of stations, a fleet of re-balancing vehicles and the final stations state, the main objective of the service provider, during the night, is to refill the stations, riding the shortest route to satisfy the stations request. In this case, the station request is the gap, or deviation (see Section 1.9), between the final station state and the initial station state at the beginning of the next day (defined by $L^S_{s,1,d}$ or $L^O_{s,1,d}$). The re-balancing problem during the night, in most of the cases, is considered as static and deterministic, since the deviation is known and the route does not change because of new available information. Hence, the problem to solve is the Static Bike-sharing Rebalancing Problem (SBRP), static and deterministic, where an optimal route for the entire planning horizon can be determined at the beginning of the period. The SBRP can be classified as a variant of the "static many-to-many one-commodity pickup and delivery problem with selective pickup and selective delivery" (1-PDTSP) introduced by *Hernández-Pérez et al. 2003*. The main differences are that the SBRP allows the use of more than one vehicle and that it not always minimizes the length of the route, but it can also minimize the deviation (see Section 1.9). The flexibility for the deviation to be different from zero allows to have a trade-off between the satisfaction of the station requests and the choice of a shorter route. For example, if the delivery of 1 bike to one station adds 10 kilometers to the vehicle route, we can ignore the delivery for a better trade-off between costs and quality of the re-balancing service. The SBRP allows only one visit in each station and the route that the service vehicles rides, is measured by the driving distance in kilometres between each couple of stations. We recall that the operations that the service vehicle can do to satisfy the station request, i.e. to fill the deviation, are limited by the service vehicle capacity.

Benchimol et al. 2011 study the case in which there is only one vehicle with limited capacity and it is allowed to split the deliveries, i.e. it is allowed to visit the stations more than once, under the assumption that the sum of all requests is equal to zero. *Chemla et al. 2013* study the case in which there is only one vehicle with limited capacity, each station can be visited several times and also it can be used as a buffer in which bikes can be stored for a later visit. *Raviv et al. 2013b* define the SBRP in which they minimize the length of the re-balancing route with two vehicles, but they also minimize the users dissatisfaction through the expected number of shortage of bikes or docking stations available. *Dell'Amico et al. 2014* allow only one vehicle that start from a Depot, it can visit only once each station and it satisfies all the stations request in one route. *Ho et al. 2014* allow only one vehicle that can start from one of the stations in the system and it can visit only once each of the other stations to minimize a penalty function that depends on the stations state. *Schuijbroek et al. 2017* group the BSS stations in different areas, called clusters. The authors allow only one vehicle that can start from one of the stations in the system and can visit only one time each of the other stations to satisfy a certain service level requirement minimizing the maximum length of the route, in a certain rolling horizon.

2.4.2 P_2^O: the optimal daily route

In the operational planning, the operator can choose to re-balance the system during the day, when the users demand can influence the routing plan of the operator. In the daily re-balancing operations, one or more vehicles travel throughout the stations prioritizing some stations to minimize the violations. During the day, the predictions of the users demand are taken into account and the operator needs to prioritize the re-balancing of the stations that may quickly become full or empty. During the day, the problem is thus dynamic, as new information of the stations state is received during the planning route. If, while operating the chosen route, some other stations become empty or full, the route can be updated. The vehicles' routes can start and finish anywhere in the system, with any feasible number of bikes on board. Hence, the problem to face during the day is called the Dynamic Bike-sharing Rebalancing Problem (DBRP), and it has been introduced into scientific literature by *Contardo et al. 2012*. In DBRP, the demand throughout the planning period and the driving time between stations are taken into account, returning, as output solution, a complete route with specific loading instructions for the service vehicle operator. The model does not however consider the stochasticity in demand and it does not take into account the lost demand when the stations are completely full or empty. For the DBRP, there is an estimated demand for bikes at all stations, that can be positive or negative. If, in the next time periods, the prediction of the pick-up demand $D^P_{s,t,d}$ is high, and the station is empty, there is a concrete possibility that an empty violation $V^E_{s,t,d}$ (see Section 1.9) may happen in that station. If the prediction of the delivery demand $D^D_{s,t,d}$ is high and the station is full, there may be a full violation $V^F_{s,t,d}$. The user that experiences a full violation delay the trip to find another station where to deliver the bike and after this experience would be probably feel discouraged to use the service again in the future.

Caggiani et al. 2012 propose a decision support system based on neural networks to forecast the users demand in each time interval (10 minutes) and to find a good route, with the assumption that the vehicle makes only one movement within one interval. Also, *Caggiani et al. 2013* assume that the users satisfaction is proportional to the probability to find a bike or a free docking station. The objective of the model is to minimize the re-balancing costs for the bike-sharing operator aiming at an high level of users satisfaction. *Vogel et al. 2014* model the DBRP by discretizing the day in long time intervals (one hour) where the bikes to move are determined and there is no connection between the routes of two different time intervals. *Kloimüllner* *et al. 2014* model the DBRP using continuous time where one vehicle starts and ends at the depot. The user demands over time is assumed to be the expected cumulated demand from the beginning of the re-balancing to the end of the planning time horizon. The objective is to minimize the lost demands as well as to minimize the deviation from the target stations state. *Regue et al. 2014* split the DBRP into tactical problems, to re-balance once a week the BSS. The tactical problems include the users demand forecast and an optimal inventory interval for each station in every time interval (20 minutes). The operational problems are solved by minimizing the expected violations that may occur in the stations and, in particular, the problem is resolved every time new data of the current stations state become available. *Brinkmann et al. 2015* propose a short term relocation (STR) strategy looking only one move ahead visiting the nearest, feasible, unbalanced station. *Brinkmann et al. 2016* present a DBRP model with discrete time-intervals where the objective is to minimize the squared gap between the final state and target interval at each station for each time interval.

2.5 The case study of Padua: relevant problems

In this thesis, we assume as given all the settings of the strategic level, i.e. the current configuration of an active BSS. The BSS taken as example for the application of our study is the Padua BSS, a small service active since 2013. Hence, we will focus on the tactical and operational level problems, to question if the current BSS needs more support in terms of bikes, docking stations and service vehicles. In particular, we see the applications of four optimization models to:

- the optimal station size (problem P_2^T);
- the optimal initial station state (problem P_3^T);
- the optimal night route (problem P_1^O);
- the optimal daily route (problem P_2^O).

Citing other case studies in literature, among others, *Saltzman et al. 2016* emulate the San Francisco BSS, generating the users demand according to a Poisson distribution. The authors focuses on the five stations where most of the shortage of bikes or docking stations occurs, simulating a BSS in which bikes and docks are added to the system, underlying that a short addition of resources (+3-4 % of bikes and docks) can lead a great decrease of the number of violations (-30 %). In this case, no studies about the impact of the re-balancing vehicles are reported.

Soriguera et al. 2018 emulate the Barcelona BSS, generating the users demand according to a Poisson distribution. The authors assume that a user starts from an origin location and, if it is closer than 750 meters to one BSS station, he/she will take a bike to go to the destination location. In this case, they use 13 vehicles to re-balance the BSS but they did not provide the model used for the choice of the vehicle route.

In Chapter 3, we see more in details the optimization models to solve the four problems above. In Chapter 4, we introduce a new and sistematic methodology based on historical data our applying the optimization models and simulation based evaluation to improve the performance of a BSS, with application to the case study of Padua. The believe that moves us is that an efficient use of the limited resources, as bikes, docking stations and service vehicles, can contribute to increase the usage of bike sharing, leading to a greener city and healthier population.

Chapter 3

Mathematical models for Bike Sharing Systems

This chapter is devoted to possible mathematical formulations of the optimization problems related to solve the two tactical and the two operational problems introduced in Sections 2.3.2, 2.3.3, 2.3.4 and 2.3.5.

3.1 Mathematical programming

The optimization models that we will describe in the following sections belong to the class of linear programming models. Linear Programming (LP) is a branch of mathematical optimization and it belongs to the applied mathematics field of Operation Research (OR). The objective of LP is to achieve the best outcome (such as maximum profit or lowest cost) in a mathematical model subject to requirements which are represented by linear relationships (inequalities). We introduce the Mixed Integer Linear Programming (MILP) problem where the problem solution is bound to be integer (or just a subset of its components is bound to be integer).

A simple exampl of a LP is a problem of the following form:

$$\min x - 3y$$

subject to

$x \geq 2$

$x \leq 7$

$y \geq 3$

$y \leq 5$

where the problem is to minimize the linear function $x - 3y$, called *objective function* (OF), changing the values of the variables x and y, called decision variables, subject to four inequalities, called constraints. Clearly, in this case, the best solution is $x - 3y = -13$, given by decision variables $x = 2$ and $y = 5$ that respect all the constraints and minimize the objective function. In general, solving an LP problem is not obvious, and we can use the simplex method, a famous algorithm for LP problems (see Bartels et al. 1969, among others).

Sometimes, variables represent indivisible goods, or decision variables are binary, so, in this case, the problem is called Integer Linear Program (LP). Problems in which only a subset of variables is bound to be integral are called Mixed Integer Linear Programs (MILP). At first, IP might be solved simply enumerating all the feasible integer points and then pick a point that results in the lowest value of the OF. This is of course possible in theory, but is not practical when the size of the problem are large. Another idea is to to solve the problem with the simplex method without setting the integrability constraints, i.e. a problem relaxation. The main algorithms used by MILP solver that based themselves on this relaxation idea are the Cutting Planes Method, the Branch & Bound Method and the Branch & Cut Method (see *Dantzig et al. 1997*, *Schrijver 1998*, *Conforti et al. 2014*, among others).

3.2 The optimal station size

In this section, we will see how to calculate, for each station s, the optimal size C_s^{NO}, to see which station needs more docking stations with respect to the standard configuration of the station size, denoted by C_s^{NS}. We recall that the required station size $C_{s,d}^{NR}$ for station s in day d (see Section 1.7), is calculated as:

$$C_{s,d}^{NR} = L_{s,d}^{\max} - L_{s,d}^{\min} \qquad \forall s \in S, d \in D$$

and the extra required docking stations $C_{s,d}^{NE}$ as:

$$C_{s,d}^{NE} = \max\{C_{s,d}^{NR} - C_s^N, 0\} \qquad \forall s \in S, d \in D$$

where:

$L_{s,d}^{\max}$ is the maximum station state;

$L_{s,d}^{\min}$ is the minimum station state;

$C_{s,d}^{NR}$ is the required station size;

$C_{s,d}^{NE}$ are the extra required docking stations;

C_s^N is the station size.

3.2.1 An optimization model

We propose an LP model where the decision variable are the optimal stations size C_s^{NO}.

We can formulate the optimization model as:

$$\min \sum_{s\in S} \sum_{d\in D} C_{s,d}^{NE} \qquad s \in S,\ d \in D \quad (3.1)$$

subject to

$$C_{s,d}^{NE} = \max\{C_{s,d}^{NR} - C_s^{NO}, 0\} \qquad s \in S,\ d \in D \quad (3.2)$$

$$\sum_{s\in S} C_s^{NO} \leq C^{NO} \quad (3.3)$$

$$C_s^{NO} \in \mathbb{Z}_+ \qquad s \in S \quad (3.4)$$

where the following data are given constants: $C_{s,d}^{NE}$ are the extra required docking stations for station s in day d;

$C_{s,d}^{NR}$ is the required station size;

C_s^{NO} is the optimal size of station s;

C^{NO} is the number of docking stations available in the BSS in the optimal stations size configuration.

Objective function (3.1) minimizes the extra required docking stations. Constraints (3.2) define the calculation of the extra required docking stations. Notice that these constraints are not linear, but they can be easily linearized. Constraint (3.3) imposes that the total number of docking stations can be at most the number of docking stations available in the BSS. Constraints (3.4) impose that the optimal number of docking stations is a non negative integer.

3.2.2 An example

Given two stations, the stations size, the number of docking stations, the maximum and the minimum stations size for a period of four days, the objective is to find the optimal stations size in a scenario with limited resources and the utopic stations size (see Section 1.2.6) in a scenario with unlimited resources.

Problem data

$S = \{1, 2\}$

$C_1^{NS} = 15$

$C_2^{NS} = 10$

$C^{NS} = C^{NO} = 25$

$D = \{1, 2, 3, 4\}$

$L_{1,d}^{\max} = [24, 18, 20, 18]$

$L_{1,d}^{\min} = [-2, -4, 0, 2]$

$L_{2,d}^{\max} = [10, 7, 8, 8]$

$L_{2,d}^{\min} = [-2, 1, 2, 0]$

Application

Starting from the data available, it is possible to calculate the required station size:

$C_{1,d}^{NR} = [26, 22, 20, 16]$ for Station 1;

$C_{2,d}^{NR} = [12, 6, 6, 8]$ for Station 2.

From these results we can deduce that:

- Station 1 needs a minimum of 16 docking stations every day and at least 26 to satisfy all the user demand;

- Station 2 needs a minimum of 6 docking stations every day and at least 12 to satisfy all the user demand.

If we do not optimize the stations size, given 25 docking stations, the size is $C_1^{NS} = 15$ for Station 1 and $C_2^{NS} = 10$ for Station 2. Given the current station size, it is possible to calculate the extra required docking stations:

$C_{1,d}^{NE} = [11, 7, 5, 1]$ for Station 1;

$C_{2,d}^{NE} = [2, 0, 0, 0]$ for Station 2.

We can calculate the total number of extra required docking stations as:

$$\sum_{s=1}^{2}\sum_{d=1}^{4} C_{s,d}^{NE} = 11 + 7 + 5 + 1 + 2 + 0 + 0 + 0 = 26$$

If we apply the optimization model, we obtain the optimal stations size, given 25 docking stations, the solution is $C_1^{NO} = 19$ for Station 1 and $C_2^{NO} = 6$ for Station 2. Given the optimal station size, it is possible to calculate the extra required docking stations:

$C_{1,d}^{NE} = [7, 3, 1, 0]$ for Station 1;

$C_{2,d}^{NE} = [6, 0, 0, 2]$ for Station 2.

We can calculate the total number of extra required docking stations as:

$$\sum_{s=1}^{2}\sum_{d=1}^{4} C_{s,d}^{NE} = 7 + 3 + 1 + 0 + 6 + 0 + 0 + 2 = 19$$

If we want to compare the previous solutions with an utopic situation, we can determine the utopic stations size, given unlimited docking stations, the solution is $C_1^{NU} = 26$ for Station 1 and $C_2^{NU} = 12$ for Station 2. The total number of extra required docking stations in the utopic stations size configuration is, of course, by definition (see Section 1.2.6), equal to:

$$\sum_{s=1}^{2}\sum_{d=1}^{4} C_{s,d}^{NE} = 0$$

Results

In conclusion, we can summarize that:

- with the standard stations size and $C^{NS} = 25$, 26 extra docking stations are required;
- with the optimal stations size and $C^{NO} = 25$, 19 extra docking stations are required;
- with the utopic stations size and $C^{NU} = 38$, no extra docking stations are required.

In Chapter 7, we will apply the optimization model to the case study of Padova and we will compare the overall performance of different solutions among them.

3.3 The optimal initial station state

Given the stations size, the other decision that the operator needs to make, at the tactical level, is the choice of the optimal initial station state. In this case, to calculate the optimal initial station state $L_{s,1,d}^{O}$, we measure the maximum empty violation $V_{s,d}^{E;\max}$ calculated as:

$$V_{s,d}^{E;\max} = |\min(L_{s,d}^{\min}, 0)|$$

and the maximum full violation $V_{s,d}^{F;\max}$ calculated as:

$$V_{s,d}^{F;\max} = \max(L_{s,d}^{\max} - C_{s}^{N}, 0)$$

that occur at station s in day d, as explained in Section 1.9. Given these quantities, we can calculate the maximum total violation $V_{s,d}^{T;\max}$ as the sum of the maximum empty violation and the maximum full violation:

$$V_{s,d}^{T;\max} = V_{s,d}^{E;\max} + V_{s,d}^{F;\max}$$

The optimal initial station state can assume two different values for each station s:

- $L_{s,1,d}^{O,W}$ if $d \in D_W$, i.e. if d is a working day;

- $L_{s,1,d}^{O,H}$ if $d \in D_H$, i.e. if d is a national holiday or a day of the weekend.

3.3.1 An optimization model

We propose an ILP where the decision variables are the optimal initial stations state $L_{s,1,d}^{O}$.

We can formulate the optimization model as:

$$\min \sum_{s \in S} \sum_{d \in D} V_{s,d}^{T;\max} \qquad s \in S,\ d \in D \quad (3.5)$$

subject to

$$V_{s,d}^{T;\max} = V_{s,d}^{E;\max} + V_{s,d}^{F;\max} \qquad s \in S,\ d \in D \quad (3.6)$$

$$V_{s,d}^{E;\max} = |\min(L_{s,d}^{\min}, 0)| \qquad s \in S,\ d \in D \quad (3.7)$$

$$V_{s,d}^{F;\max} = \max(L_{s,d}^{\max} - C_s^{NO}, 0) \qquad s \in S,\ d \in D \quad (3.8)$$

$$L_{s,d}^{\min} = L_{s,1,d}^{O} + D_{s,d}^{cum;\min} \qquad s \in S,\ d \in D \quad (3.9)$$

$$L_{s,d}^{\max} = L_{s,1,d}^{O} + D_{s,d}^{cum;\max} \qquad s \in S,\ d \in D \quad (3.10)$$

$$\sum_{s \in S} L_{s,1,d}^{O} \leq B^{NO} \qquad s \in S,\ d \in D \quad (3.11)$$

$$L_{s,1,d}^{O} \leq C_s^{NO} \qquad s \in S,\ d \in D \quad (3.12)$$

$$L_{s,1,d}^{O} \in \mathbb{N} \qquad s \in S,\ d \in D \quad (3.13)$$

where (see Chapter 1):

$V_{s,d}^{T;\max}$ are the maximum total violations for station s in day d;

$V_{s,d}^{E;\max}$ are the maximum empty violations;

$V_{s,d}^{F;\max}$ are the maximum full violations;
$L_{s,1,d}^{O}$ is the optimal state of station s;
$L_{s,d}^{\min}$ is the minimum state of station s;
$L_{s,d}^{\max}$ is the maximum state of station s;
C_{s}^{NO} is the size of station s;
$D_{s,t,d}^{cum}$ is the cumulative users demand for station s at time t in day d;
B^{NO} is the number of bikes available in the BSS in the optimal stations size configuration.

Objective function (3.5) minimizes the maximum total violations. Constraints (3.6), (3.7), (3.8), (3.9) and (3.10) are quantities already introduced in Chapter 1 and they have been formally added in the model. Notice that constraints (3.7) and (3.8) are not linear, but they can be easily linearized. Constraint (3.11) imposes that the number of bikes can be lower or equal to the number of bikes available in the BSS B^{N}. Constraints (3.12) impose that the station state can be lower than or equal to the station size C_{s}^{N}. Constraints (3.13) impose that all the stations states are non negative integers.

3.3.2 An example

Given two stations, the stations size, the total number of bikes, minimum and maximum cumulative users demand for a period of four days, the objective is to find the standard and the optimal stations state.

Problem data

$S = \{1, 2\}$
$C_{1}^{NS} = 20$
$C_{2}^{NS} = 10$
$C^{NS} = 30$
$B^{NS} = 12$
$D = \{1, 2, 3, 4\}$
$D_{1,d}^{cum;\max} = [18, 12, 17, 18]$
$D_{1,d}^{cum;\min} = [-8, -5, -5, -6]$

$D^{cum;\max}_{2,d} = [14, 7, 8, 10]$

$D^{cum;\min}_{2,d} = [-6, -8, -3, -4]$

Application

Starting from the available data , it is possible to calculate the docks per bike ratio R^{DpB} (see Section 1.6) as:

$$R^{\mathrm{DpB}} = \frac{C^{NS}}{B^{NS}} = \frac{30}{12} = 2.5$$

and according to formula (1.1) it, it is possible to calculate the standard initial stations state as:

$$L^S_{1,1,d} = \min\left(\left\lfloor \frac{C^{NS}_1}{R^{\mathrm{DpB}}} \right\rfloor, \left\lfloor \frac{C^{NS}_1}{2.25} - 0.5 \right\rfloor + 1\right) = \min\left(\left\lfloor 8 \right\rfloor, \left\lfloor 8.39 \right\rfloor + 1\right) = 8$$

$$L^S_{2,1,d} = \min\left(\left\lfloor \frac{C^{NS}_2}{R^{\mathrm{DpB}}} \right\rfloor, \left\lfloor \frac{C^{NS}_2}{2.25} - 0.5 \right\rfloor + 1\right) = \min\left(\left\lfloor 4 \right\rfloor, \left\lfloor 3.94 \right\rfloor + 1\right) = 4$$

If we do not optimize the stations state, but simply keep the suggest size without optimization, we have the following: given 12 bikes, the solution is $L^S_{1,1,d} = 8$ and $L^S_{2,1,d} = 4$. Given the standard initial stations state and the stations size it is possible to calculate:

- the minimum and maximum stations state

$L^{\min}_{1,d} = [0, 3, 3, 2]$

$L^{\max}_{1,d} = [26, 20, 25, 26]$

$L^{\min}_{2,d} = [-2, -4, 1, 0]$

$L^{\max}_{2,d} = [18, 11, 12, 14]$

- the maximum empty and full violations

$V^{E;\max}_{1,d} = [0, 0, 0, 0]$

$V^{F;\max}_{1,d} = [6, 0, 5, 6]$

$V^{E;\max}_{2,d} = [2, 4, 0, 0]$

$V^{F;\max}_{2,d} = [8, 1, 2, 4]$

- the maximum total violation

$\sum_s \sum_d V_{s,d}^{T;\max} = 38$

If we use the optimization model to calculate the optimal stations state, we have: given 12 bikes, the solution is $L_{1,1,d}^O = 5$ and $L_{2,1,d}^O = 3$. Given the optimal initial stations state and the stations size it is possible to calculate:

- the minimum and maximum stations state

$L_{1,d}^{\min} = [-3, 0, 0, -1]$

$L_{1,d}^{\max} = [23, 17, 22, 23]$

$L_{2,d}^{\min} = [-3, -5, 0, -1]$

$L_{2,d}^{\max} = [17, 10, 11, 13]$

- the maximum empty and full violations

$V_{1,d}^{E;\max} = [3, 0, 0, 1]$

$V_{1,d}^{F;\max} = [3, 0, 2, 3]$

$V_{2,d}^{E;\max} = [3, 5, 0, 1]$

$V_{2,d}^{F;\max} = [7, 0, 1, 3]$

- the maximum total violation

$\sum_s \sum_d V_{s,d}^{T;\max} = 32$

Results

In conclusion, we can summarize that:

- with the standard station size, the standard initial stations state, $B^{NS} = 12$ and $C^{NS} = 30$, there are in total 38 maximum total violations;
- with the standard station size, the optimal initial stations state, $B^{NS} = 12$ and $C^{NS} = 30$, there are in total 32 maximum total violations.

3.4 The optimal night route

In this section, we see problem P_1^O introduced in Section 2.4.1 and we apply an optimization model developed by *Dell'Amico et al. 2014* to determine the shortest route to rebalance the BSS during the night, i.e. to bring back the stations state from the final station state to the initial station state for the next day. Consider a directed graph $G = (V, A)$, where the vertices

$$V = \{0\} \cup S = \{0\} \cup \{1, ..., P_1^S\} = \{0, ..., P_1^S\}$$

are the set of stations S, including the depot (vertex 0) and an arc $(s_1, s_2) \in A$ represents the trip between station s_1 and station s_2, to which it is associated a travelling cost $d(s_1, s_2) \in A$, e.g. the driving distance in kilometres between each pair of stations. Each arc is associated with a binary variable $w(s_1, s_2) \in A$ that takes value 1 if that arc is used in the optimal route, 0 otherwise. Each station s has a deviation $q_{s,T^N,d}$ that is the difference between the final station state $L_{s,T^N,d}$ and the given initial station state for the beginning of the next day $L_{s,1,d+1}$. We recall that the deviation (see Section 1.9) is determined as:

$$q_{s,T^N,d} = L_{s,T^N,d} - L_{s,1,d+1}$$

If $q_{s,T^N,d} \geq 0$, then s is a pickup station, where $q_{s,T^N,d}$ bikes should be removed; if $q_{s,T^N,d} \leq 0$ then s is a delivery station, where $q_{s,T^N,d}$ bikes should be supplied, for $s \in S$. The bikes removed from pickup stations can either go to a delivery station or back to the depot. Bikes supplied to delivery stations can either come from the depot or from pickup stations. A fleet of R^N vehicles of capacity C^r, $r \in \{1, ..., R^N\}$, is available at the depot. A decision variable L_s^r is added to the model and depicts the load of a vehicle r after it visited the station s, for $s \in V$. The load L_s^r needs to be updated along the route taking into consideration that, if a vehicle r travels along arc (s_1, s_2), then $L_{s_2}^r$ should be equal to $L_{s_1}^r + q_{s_2,T^N,d}$. These "conditional" constraints can be put in a linear form with the "big M" method, as:

$$L_{s_2}^r \geq L_{s_1}^r + q_{s_2,T^N,d} - M(1 - w(s_1, s_2)), \qquad s_1 \in V, s_2 \in S$$

$$L_{s_1}^r \geq L_{s_2}^r - q_{s_2,T^N,d} - M(1 - w(s_1, s_2)), \qquad s_1 \in S, s_2 \in V$$

where M is a large integer constant used to linearise the constraints.

We want to minimize the total cost, while ensuring that the following constraints are not violated:

- each vehicle r performs a route that starts and ends at the depot;

- each vehicle r starts from the depot with a number of bikes that vary from 0 to C^r, the maximum capacity of the vehicle;

- each station s is visited exactly once and its deviation $q_{s,T,d}$ is completely fulfilled by the vehicle visiting it;

- the sum of requests of the visited stations plus the initial load is never negative or greater than C^r in the route performed by a vehicle;

- a station s with deviation $q_{s,T,d} = 0$ must be visited, even if it implies that no bike has to be dropped off or picked up there. This case arises when the driver of the vehicle is supposed to check that the station is correctly working. The case in which the stations with null requests are skipped can be obtained by simply removing those stations from the set of vertices.

Note that we do not impose the sum of redistributed bikes to be null, and hence there can be a positive or a negative flow of bikes involving the depot. This consideration is useful to model cases in which some bikes enter or leave the depot at the end or beginning of the route.

3.4.1 An optimization model

We report the optimization model developed by *Dell'Amico et al. 2014*, formulated as:

$$\min \sum_{s_1} \sum_{s_2} d(s_1, s_2) w(s_1, s_2) \tag{3.14}$$

subject to

$$\sum_{s_1} w(s_1, s_2) = 1 \qquad s_1 \in V, s_2 \in S \tag{3.15}$$

$$\sum_{s_1} w(s_2, s_1) = 1 \qquad s_1 \in V, s_2 \in S \tag{3.16}$$

$$\sum_{s} w(0,s) \leq R^N \qquad s \in S \quad (3.17)$$

$$\sum_{s} w(0,s) = \sum_{s} w(s,0) \qquad s \in S \quad (3.18)$$

$$\sum_{s_1}\sum_{s_2} w(s_1,s_2) \leq |S| - 1 \qquad s_1 \in S, s_2 \in S \quad (3.19)$$

$$L^r_{s_2} \geq L^r_{s_1} + q_{s_2,T^N,d} - M(1 - w(s_1,s_2)) \qquad s_1 \in V, s_2 \in S, r \in R \quad (3.20)$$

$$L^r_{s_1} \geq L^r_{s_1} - q_{s_2,T^N,d} - M(1 - w(s_1,s_2)) \qquad s_1 \in S, s_2 \in V, r \in R \quad (3.21)$$

$$\max(0, q_{s,T^N,d}) \leq L^r_s \leq \min(C^r, C^r + q_{s,T^N,d}) \qquad s \in V \quad (3.22)$$

$$w(s_1,s_2) \in \{0,1\} \qquad s_1, s_2 \in V \quad (3.23)$$

where:

- $d(s_1, s_2)$ is the distance between each pair of stations;
- $w(s_1, s_2)$ is the binary variable that value 1 if the optimal route solution includes arc (s_1, s_2);
- R^N is the number of service vehicles;
- L^r_s is the load of the vehicle after visiting station s;
- $q_{s,T^N,d}$ is the deviation in station s;
- M is a large integer constant;
- C^r is the capacity of vehicle r.

Objective function (3.14) minimizes the travelling cost. Constraints (3.15) and (3.16) impose that every station but the depot is visited exactly once. Constraints (3.17) and (3.18) ensure, respectively, that at most R^N vehicles leave the depot, and that all vehicles that are used return to the depot at the end of their route. Constraints (3.19) are the sub tour elimination constraints that impose the connectivity of the solution. Constraints (3.20) e (3.21) impose the load of the vehicles to be updated after visiting each station. Constraints (3.22) give the lower and upper bound of the loads. Constraints (3.23) impose the variables $w(s_1, s_2)$ to be binary.

3.4.2 An example

Given three stations, the distances between each pair of stations, the deviations and one service vehicle, the objective is to find the shortest feasible route to fulfil the stations request. For the sake of simplicity, in this example the distance between two stations will be the same in both directions: $d(s_1, s_2) = d(s_2, s_1)$. Notice that, in general, due e.g. to road constraints, the distance between two stations may be different depending on the direction.

Problem data

$S = \{1, 2, 3\}$

$V = \{0, 1, 2, 3\}$

$R^N = 1$

$d(0, 1) = d(1, 0) = 2$ km

$d(0, 2) = d(2, 0) = 4.5$ km

$d(0, 3) = d(3, 0) = 3$ km

$d(1, 2) = d(2, 1) = 3.5$ km

$d(1, 3) = d(3, 1) = 2.5$ km

$d(2, 3) = d(3, 2) = 1$ km

$q_{1,T^N,1} = -4$

$q_{2,T^N,1} = 6$

$q_{3,T^N,1} = 8$

$C^1 = 20$

$L_0^1 = 10$

Application

By considering all the permutation of the stations, it is possible to calculate six possible routes to visit all the stations:

- $route_1 = \{0, 1, 2, 3, 0\}$
  $L_s^1 = [10, 6, 12, 20, 20]$
  $cost_1 = d(0, 1) + d(1, 2) + d(2, 3) + d(3, 0) = 9.5\text{km}$

- $route_2 = \{0, 1, 3, 2, 0\}$
  $L_s^1 = [10, 6, 14, 20, 20]$
  $cost_2 = d(0, 1) + d(1, 3) + d(3, 2) + d(2, 0) = 10\text{km}$

- $route_3 = \{0, 2, 3, 1, 0\}$
  $L_s^1 = [10, 16, 24, 20, 20]$
  $route_3$ is not feasible because $L_3^1 > C^R$

- $route_4 = \{0, 2, 1, 3, 0\}$
  $L_s^1 = [10, 16, 12, 20, 20]$
  $cost_4 = d(0, 2) + d(2, 1) + d(1, 3) + d(3, 0) = 13.5\text{km}$

- $route_5 = \{0, 3, 2, 1, 0\}$
  $L_s^1 = [10, 18, 24, 20, 20]$
  $route_5$ is not feasible because $L_2^1 > C^R$

- $route_6 = \{0, 3, 1, 2, 0\}$
  $L_s^1 = [10, 18, 14, 20, 20]$
  $cost_6 = d(0, 3) + d(3, 1) + d(1, 2) + d(2, 0) = 13.5\text{km}$

Results

From the calculations above, it is possible to observe that $route_1$, $route_2$, $route_4$ and $route_6$ are feasible routes, whereas $route_3$ and $route_5$ are not feasible, because during the route the vehicle load becomes greater than C^1. Hence, the optimal night route is $route_1$ because it has the minimum cost between all the feasible routes: $cost_1 = 9.5$. Notice that the same optimal route can be obtained by solving the proposed ILP model (see Section 3.1) with more effective algorithms than enumerating all the proposed solutions, which become impractical for large instances.

3.5 The optimal daily route

The last step for the service provider is to rebalance the BSS during the day, introduced in Section 2.4.1 as problem P_3^O. The optimization model that we will use to find the optimal daily route is developed by *Contardo et al. 2012*. The objective of the model is to minimize the sum of violations at the end of the planning period, given the current stations state and the users demand forecast for all the stations. Full violations are often considered worse than empty violations, and could therefore be given more weight. Additionally, there could be given different weights to violations at different stations based on their importance.

Let R be the set of vehicles, decided by solving the strategic problem P_7^S and let C^r be the vehicles capacity. Let L_0^r be the initial load of a vehicle $r \in R$ at the beginning of the time horizon. Let S be the set of stations in the network. The time horizon is discretized into a set of time intervals $t \in T_H^N$. This is done to explicitly take into account the possibility of visiting the same station at different times. We assume that the time intervals are indexed from 1 to T_H^N. We consider a set of states P, each state is a pair (a, t), where a is a vehicle or a station and $t \in T_H^N \cup \{0\}$. In particular, P is composed of:

- the states representing the initial positions of the vehicles at time 0, $\{(u^r, 0) : r \in R\}$;
- the states representing the stations at the different time periods, $\{(s, t) : s \in S, t \in T_H^N\}$;
- a dummy state denoted ϕ to represent the end of a route in the planned schedule.

We denote by $s(p)$ and $t(p)$ the station and time interval corresponding to state $p \in P \backslash \{\phi\}$. We denote by P_S the subset of states composed of the pairs (s, t), $s \in S$, $t \in T^H$. For the given state $p \in P_S$ we define $pred(p) = (s(p), t(p) - 1)$ if $t(p) \geq 2$ and $succ(p) = (s(p), t(p) + 1)$ if $t(p) \leq T_H^N - 1$. With each station $s \in S$ we associate a capacity C_s^N, and, for each state $p \in P_S$, a number of bikes $L_{s(p),t(p),d}$ in station $s(p)$ at time $t(p)$. Also, with each state $p \in P_S$ we associate a users demand $D_{s(p),t(p),d}$ for bikes ($D_{s(p),t(p),d} \geq 0$ if the pick-up demand is higher than the delivery demand, and $D_{s(p),t(p),d} \leq 0$ if the delivery demand is higher than the pic-up demand). Let us consider a graph $G = (P, A)$, where the nodes correspond to states and the arc set A is composed of three types of arcs:

- all feasible direct trips between a pair of states, i.e., all arcs $(p, p') \in P \times P$ such that $t(p') - t(p) \geq d(s(p), s(p')) > t(p') - t(p) - 1$, where $d(s(p), s(p'))$ is the distance between two nodes;

- all arcs $(p, succ(p))$ for $p \in P_S$ such that $t(p) \leq T_H^N - 1$;

- all arcs (p, ϕ) representing the end of the vehicle routes.

Figure 3.1 illustrates a network with two vehicles and three stations, and the dynamic network resulting after a time discretization into five periods of one unit each. The number of edges in the dynamic network is much larger than in the original graph, i.e. a graph in which we consider only the states, not the different time periods. Also, note that the dynamic network allows only arcs that go forward in time. For

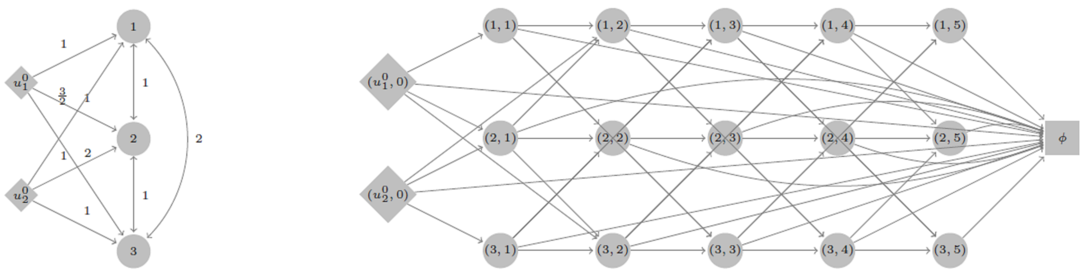

Figure 3.1: Original network versus dynamic network (*Contardo et al. 2012*)

each state $p \in P$, let $y_p^+, y_p^- \geq 0$ be two variables representing a shortage and excess of bikes at state p. Let $z_p \geq 0$ be a variable representing the number of bikes left at state p. For each arc $a \in A$ and vehicle $r \in R$, let w_a^r be a binary variable equal to 1 if vehicle r traverses arc a in its route, and let $x_a^r \geq 0$ be a continuous variable equal to the load of vehicle r along arc a. For state $p \in P$ we denote by $\delta^+(p)$ the set of arcs ending at p and by $\delta^-(p)$ the set of arcs starting at p.

3.5.1 An optimization model

Below, we introduce the mathematical programming formulation of the problem, proposed by *Contardo et al. 2012*:

$$\min \sum_{p\in P_S} (y_p^+ + y_p^-) \tag{3.24}$$

subject to

$$\sum_{r\in R}\sum_{a\in\delta^+(p)} x_a^r - \sum_{r\in R}\sum_{a\in\delta^-(p)} x_a^r - z_p + y_p^+ - y_p^- = D_{s(p),t(p),d} - L_{s(p),t(p),d} \quad p\in P_S, t(p)=1 \tag{3.25}$$

$$\sum_{r\in R}\sum_{a\in\delta^+(p)} x_a^r - \sum_{r\in R}\sum_{a\in\delta^-(p)} x_a^r + z_{pred(p)} - z_p + y_p^+ - y_p^- = D_{s(p),t(p),d} \quad p\in P_S, t(p)\geq 2 \tag{3.26}$$

$$\sum_{r\in R}\sum_{a\in\delta^+(p)} w_a^r \leq 1 \quad p\in P_S \tag{3.27}$$

$$0 \leq z_p \leq C^N_{s(p)} \quad s\in P_S \tag{3.28}$$

$$x_a^r \leq C^r w_a^r \quad r\in R, a\in A \tag{3.29}$$

$$\sum_{a\in\delta^-(p)} w_a^r - \sum_{a\in\delta^+(p)} w_a^r = 0 \quad r\in R, s\in P_S \tag{3.30}$$

$$\sum_{a\in\delta^-(u^r,0)} w_a^r = 1 \quad r\in R \tag{3.31}$$

$$\sum_{a\in\delta^-(u^r,0)} x_a^r = L_0^r \quad r\in R \tag{3.32}$$

$$y_p^+, y_p^- \geq 0 \quad s\in P_S \tag{3.33}$$

$$x_a^r \geq 0 \quad r\in R, a\in A \tag{3.34}$$

$$w_a^r \in \{0,1\} \quad r\in R, a\in A \tag{3.35}$$

where:

$D_{s(p),t(p),d}$ is the users demand in station $s(p)$ at time $t(p)$ in day d;

$L_{s(p),t(p),d}$ is the station state in station $s(p)$ at time $t(p)$ in day d;

$C^N_{s(p)}$ is the station size of station $s(p)$;

$\delta^+(p)$ is the set of arcs ending at p;

$\delta^-(p)$ is the set of arcs starting at p;

y_p^+ is the shortage of bikes in state p;

y_p^- is the excess of bikes in state p;

x_a^r is the load of the vehicle r along arc a;

z_p is the number of bikes left at state p;

w_a^r is binary variable taking value 1 if vehicle r use arc a, 0 otherwise;

L_0^r is initial load of vehicle r;

C^r is the service vehicle capacity.

The objective function (3.24) represents the total unmet demand, i.e., the number of users who try to collect bikes from empty stations or to deliver bikes to full stations. Constraints (3.25) and (3.26) are the flow conservation constraints at each station for every time period. For each $p \in P_S$, the first sum indicates the number of bikes entering the station at time $t(p)$, while the second sum indicates the number of bikes that come out from the station at time $t(p)$. The role of the variables y_p^+, y_p^- is to compensate for the imbalance of the network. In a balanced network, these quantities will always be zero. Constraints (3.27) ensure that each node is visited at most once in a time period. Constraints (3.28) are the non-negativity constraints for the variables z_p and the capacity constraints of the stations in every time period. Constraints (3.29) link the use of each arc to the maximum allowable load on the vehicle traversing that arc. Constraints (3.30) are the vehicle-flow conservation constraints: they force vehicles to leave the stations previously visited. Constraints (3.31) ensure that every vehicle is used exactly once. Constraints (3.32) ensure that vehicles leave their starting positions with their current loads. Constraints (3.33) are the non-negativity constraints for the variables y_p^+, y_p^-. Constraints (3.34) are the non-negativity constraints for the vehicle loads on arcs. Finally, constraints (3.35) ensure that the vehicle-flow variables w_a^r is binary. Notice that integrability of variable x and z is implied by the model.

3.5.2 An example

Given two stations and a depot, one vehicle with a limited capacity and an initial load, a time horizon of three time intervals of 5 minutes each, the time distance in minutes between the two stations and between each station and the depot, the users demand in the next three time intervals, the stations state and the stations size, the objective is to find the optimal daily route that minimize the expected number of violations.

Problem Data

$S = \{1, 2\}$

$V = \{0, 1, 2\}$

$R = \{1\}$

$D = \{1\}$

$C^1 = 20$

$L_0^1 = 10$

$T_H^N = \{1, 2, 3\}$

$d(0, 1) = d(1, 0) = 7\,\mathrm{min}$

$d(0, 2) = d(2, 0) = 4\,\mathrm{min}$

$d(1, 2) = d(2, 1) = 3\,\mathrm{min}$

$D_{1,1,1} = 2$

$D_{1,2,1} = 3$

$D_{1,3,1} = 2$

$D_{2,1,1} = -3$

$D_{2,2,1} = -4$

$D_{2,3,1} = -2$

$L_{1,0,1} = 7$

$L_{2,0,1} = 6$

$C_1^N = 10$

$C_2^N = 10$

Application

First of all, we need to define the set of states and arcs.

The states P are:

$P = \{(u^1, 0), (1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3), \phi\}$

The arcs A are:

$A = \{\{(u^1, 0), (2, 1)\}, \{(2, 1), (2, 2)\}, \{(2, 1), (1, 2)\}\{(2, 2), (2, 3)\}, \{(2, 2), (1, 3)\},$
$\{(u^1, 0), (1, 2)\}, \{(1, 2), (1, 3)\}, \{(1, 2), (2, 3)\}, \{(u^1, 0), \phi\}, \{(1, 2), \phi\}, \{(1, 3), \phi\},$
$\{(2, 1), \phi\}, \{(2, 2), \phi\}, \{(2, 3), \phi\}\}$

Notice that arc $\{(u^1, 0), (1, 1)\}$ does not exist because as there is a time distance of 7 minutes between the depot and Station 1, we need 2 time intervals to reach Station 1. Also, we can see that the expected users demand in the next three time periods is positive for Station 1 and negative for Station 2.

In this case, by the application of the optimization model, the optimal route is $route = \{(u^1, 0), (2, 1), (1, 2), \phi\}$, where:

- $w\{(u^1, 0), (2, 1)\} = 1$, $w\{(2, 1), (1, 2)\} = 1$, $w\{(1, 2), \phi\} = 1$ and any other $w_a = 0$;
- $x\{(u^1, 0), (2, 1)\} = 10$, $x\{(2, 1), (1, 2)\} = 7$, $x\{(1, 2), \phi\} = 11$ and any other $x_a = 0$.

Results

We compare the optimal solution with the base case where the service operator does not operate during the day, so it is possible to calculate the stations state as

$L_{1,1,1} = \max(\min(L_{1,0,1}, C_1^N), 0) + D_{1,1,1} = 7 + 2 = 9$

$L_{1,2,1} = \max(\min(L_{1,1,1}, C_1^N), 0) + D_{1,2,1} = 9 + 3 = 12$

$L_{1,3,1} = \max(\min(L_{1,2,1}, C_1^N), 0) + D_{1,3,1} = 10 + 2 = 12$

$L_{2,1,1} = \max(\min(L_{2,0,1}, C_2^N), 0) + D_{2,1,1} = 6 - 3 = 3$

$L_{2,2,1} = \max(\min(L_{2,1,1}, C_2^N), 0) + D_{2,2,1} = 3 - 4 = -1$

$L_{2,3,1} = \max(\min(L_{2,2,1}, C_2^N), 0) + D_{2,3,1} = 0 - 2 = -2$

where the number of violations is

- $L_{1,1,1} = 0$, $L_{1,2,1} = 2$, $L_{1,3,1} = 2$

- $L_{2,1,1} = 0,\ L_{2,2,1} = 1,\ L_{2,3,1} = 2$

and the total number of violations is

$$\sum_{s=1}^{S} \sum_{t=1}^{T_H^N} = 0 + 2 + 2 + 0 + 1 + 2 = 7$$

For the case of the optimal daily route, the service operator operates during the day to minimize the number of violations, hence, according to the results of the model, it is possible to calculate the stations state as

$L_{1,1,1} = \max(\min(L_{1,0,1}, C_1^N), 0) + D_{1,1,1} = 7 + 2 = 9$

$L_{1,2,1} = \max(\min(L_{1,1,1}, C_1^N), 0) + D_{1,2,1} + x\{(2,1),(1,2)\} - x\{(1,2),\phi\} = 9 + 3 + 7 - 11 = 8$

$L_{1,3,1} = \max(\min(L_{1,2,1}, C_1^N), 0) + D_{1,3,1} = 8 + 2 = 10$

$L_{2,1,1} = \max(\min(L_{2,0,1}, C_2^N), 0) + D_{2,1,1} + x\{(u^1,0),(2,1)\} - x\{(2,1),(1,2)\} = 6 - 3 + 10 - 7 = 6$

$L_{2,2,1} = \max(\min(L_{2,1,1}, C_2^N), 0) + D_{2,2,1} = 6 - 4 = 2$

$L_{2,3,1} = \max(\min(L_{2,2,1}, C_2^N), 0) + D_{2,3,1} = 2 - 2 = 0$

, where the number of violations as

- $L_{1,1,1} = 0,\ L_{1,2,1} = 0,\ L_{1,3,1} = 0$
- $L_{2,1,1} = 0,\ L_{2,2,1} = 0,\ L_{2,3,1} = 0$

and the total number of violations as

$$\sum_{s=1}^{S} \sum_{t=1}^{T_H^N} = 0 + 0 + 0 + 0 + 0 + 0 = 0$$

In summary, the number of violations without any rebalancing of the system is 7, while with the usage of a service vehicle and the above introduced optimization models, the number of violations is 0.

Chapter 4

A scenario based process for the configuration of Bike Sharing Systems

In this chapter, we elaborate a decision-making process to configure and evaluate a BSS. As discussed in Chapter 2, there are various decisions to make in the configuration of a BSS and we will focus on sizing the resources available, such as bikes, docking stations and service vehicles. Let us call *configuration scenario* a case defined by the use of specific stations size and state, among the alternatives introduced in Section 1.2.5 (standard and optimal initial stations state) and 1.2.6 (standard, optimal and utopic stations size), and by the use, or not, the re-balancing routes introduced in Section 2.4.1 and 2.4.2.

In the next sections we better define the configuration scenario, we describe the steps of a methodology that may be followed by the service provider to evaluate the BSS under given configuration scenarios, in order to support her/his decision-making process. Also, we classify and select our subset of different relevant configuration scenarios that can be applied to evaluate and compare the performance of the BSS service.

4.1 Configuration scenarios

A configuration scenario can be used to assist the service provider in a decision-making process, simulating the impact of different choices on the functionality of a BSS. There is a large number of choices that a service provider can make to improve the BSS, but not all of them are covered in the thesis. Hence, some parameters are given as common to all the scenarios selected for the simulations, as the number and location of the stations and the number of service vehicles.

Once the common parameters fixed, the service provider can choose to use the resources currently available in the system, or to increase them. This choice impacts the number of bikes and docking stations available in each station. To this end, different choices for the configuration of stations size and state can be made. Finally, the service provider should decide if, during the night or the day, there is a necessity to re-balance the stations. In Section 4.3, we describe more in details the criterion followed for the creation of each configuration scenario.

4.2 Decision-making process

In this section, we propose a methodology to evaluate the performance of a configuration scenario, with the aim of supporting the decision-making process of a BSS operator. The evaluation is obtained by the steps described below.

4.2.1 Step 1: common parameters

As first step, the service operator sets all the strategic level parameters that will be used for all the configuration scenarios: the number and location of the stations and the number of service vehicles.

4.2.2 Step 2: setting the time period of analysis

The period of time chosen for the evaluation of the models needs to be decided. Each day in this period is split in discrete time intervals of few minutes. Consequently, the users demand from historical data is discretized. For the case study of Padua, we have chosen to split each day in time intervals of 5 minutes, with a total of 288 time intervals per day, and we used a time period D of 365 days, from the 1st January 2018 to the 31st December 2018.

4.2.3 Step 3: users demand forecast

For example, for the chosen time period, a forecast of the users demand needs to be computed. We will see this step more in detail in Section 5.7.

4.2.4 Step 4: setting the stations size

There are three alternatives, for each station s, for the choice of the stations size:

- standard: the standard station size C_s^{NS}, e.g. the current setting;
- optimal: the optimal station size C_s^{NO}. To this end, we consider the solution of the model of Section 3.2;
- utopic: the utopic station size C_s^{NU}. This is equal to the maximum required station size in the time period D, able to avoid any extra required docking station, under the hypothesis of unlimited number of docking stations (see Section 1.7).

For each scenario, the choice of the stations size will be the same during all the time period.

4.2.5 Step 5: setting the stations state

Given the stations size chosen in Step 4, there are two alternatives for the choice of the stations state:

- standard: the initial station state is the standard initial station state $L_{s,1,d}^{S}$;
- optimal: the initial station state is the optimal initial station state $L_{s,1,d}^{O}$ given by the model of Section 3.3. For the optimal stations state solution, we recall that we have to make distinction between the configurations of working days and holidays.

For each scenario, the choice of the stations state will be the same during all the time period.

4.2.6 Step 6: night route

There are two alternatives for the night route:

- no night route: the service operator does not operate during the night, hence, the station state at the end of the day is not adapted for the next day;
- optimal night route: the service operator brings back the stations state to desirable stations state chosen at Step 5.

4.2.7 Step 7: daily route

There are two alternatives for the daily route:

- no daily route: the service operator does not operate during the day, hence, the station state changes only if users pick-up or delivery bikes;
- optimal daily route: the service operator operates during the day with the fleet of vehicles given by Step 1, according to the operations proposed by the solution of the dynamic model (see Section 3.5).

4.2.8 Step 8: evaluation by simulation

The evaluation of each configuration scenario is done by a *simulation* as we will detail in Section 5.1. The simulation consists on the emulation of the BSS operation with the settings selected above, and based on the real data of the trips occurred in 2018. Hence, the simulation starts from the first day of the time period defined in Step 2. For each time interval, the stations state are updated depending on the pick-up or delivery operations done by the users from one station to another. Each time interval a station state is negative, one or more empty violations occurred, they are counted for the final violation performance indicator and the station state is reset to 0. Each time interval a station state is larger than the station size, one or more full violations occurred, they are counted for the final violation index and the station state is reset to the value of the station size.

4.2.9 Step 9: integrating the simulation of daily operations

Step 9 enters in the process only if at Step 7 the daily re-balancing has been chosen. For this step, we have to introduce two important concepts: the rolling horizon and the route operations penalization.

Rolling horizon

A rolling horizon is a methodology to split a large time window into smaller time windows, called time horizons, to decrease the computational time. On top of that, a roll period, equal or usually shorter than the time horizon, is used to update the model as new information becomes available. In our case, the large time window consists on the time in which the re-balancing vehicle is active during the day. We introduce the following definitions:

- the starting time interval t_{SDR} is the starting time for the daily re-balancing operations, i.e. the interval in which the re-balancing vehicle begins its service around the city;
- the ending time interval t_{EDR} is the ending time for the daily re-balancing operations, i.e. the interval in which the re-balancing vehicle ends its service to go back to the Depot;
- the time horizon T_H^N ($\leq t_{EDR} - t_{SDR}$) is the temporal length of the current re-balancing route plan. The time horizon should not be too short, as there will be few possibilities to include the more distant stations in the solution route, but also it should not be too long, as the complexity of the model to calculate the route sensibly increases. Once defined the time horizon, the route of the re-balancing vehicle is planned for the time intervals between t and $t+T_H^N$. We would like to dynamically update the re-balancing route, as new information are available from the IT system. For example, a station that 10 minutes

ago did not need to be refilled, may suddenly become empty as several users pick-up bikes. To this end, we define a roll period to update the route;

- a roll period T_R^N is the time interval between two consecutive update of the re-balancing route. The length of the roll period T_R^N is equal or shorter than the length of the time horizon T_H^N. Hence, the route is planned for the time intervals between a time t (starting from t_{SDR}) and the time $t + T_H^N$, but the vehicle operates only until time $t + T_R^N$, then new information is given by the IT system, and the route is updated.

This approach allows to plan the re-balancing route on a long period, by considering a sequence of problems on smaller time periods, and it also adds the possibility to update the route if new stations need to be refilled.

We assume that, once the vehicle starts moving toward a station, it cannot be diverted. For this reason, if at time $t + T_R^N$ the vehicle is not visiting any station, the route is updated in the first time interval, after time $t+T_R^N$, in which the vehicle visit a station.

For example:

- let be $t_{SDR} = 10$, $t_{EDR} = 30$, $T_R^N = 3$ and $T_H^N = 6$;
- the re-balancing service starts its route at time $t = t_{SDR} = 10$;
- the first route is planned for the time intervals between time 10 (t) and time 16 ($t + T_H^N$). The vehicle operates until time 13 ($t + T_R^N$), when new information is given by the IT system;
- then t is updated to the current time interval 13;
- the second route is planned for the time intervals between time 13 (t) and time 19 ($t+T_H^N$). The vehicle operates until time 16 ($t+T_R^N$), then new information is given by the IT system;
- then t is updated to the current time interval 16, and so on, until the ending time interval t_{EDR}.

Route operations penalization

Another aspect to take into consideration is the objective function of the optimization model of Section 3.5 (Formula 3.24), that we recall here:

$$\min \sum_{p \in P_S} (y_p^+ + y_p^-)$$

where:

y_p^+ is the shortage of bikes in state p;

y_p^- is the excess of bikes in state p.

Hence, the objective of this function is to minimize the number of expected violations that may occur in the next time horizon, which may be obtained by different re-balancing operations. For example, the optimization model may give as solution a route in which 10 bikes are moved between two stations without increasing or decreasing the number of violations. Thus, an equivalent solution exists with less re-balancing operations, which may be preferred. To this end, we add a penalization in the objective function to avoid extra operations. A way to do it is to penalize the binary variable w_a^r, that takes value 1 if vehicle r use arc a, 0 otherwise, i.e. to penalize the number of stations visited by the service vehicles.

Hence, the objective function of the optimization model becomes:

$$\min \sum_{p \in P_S} (y_p^+ + y_p^-) + \gamma \sum_r \sum_a w_a^r \tag{4.1}$$

where the value of the parameter γ can be seen as a threshold: $w_a^r = 1$ only if the marginal decrease of the number of violations after travelling arc a with vehicle r is equal to or greater than γ.

For example, if $\gamma = 2$, the vehicle r will include arc a only if this action will benefit at least in a marginal decreasing of 2 units in the number of violations.

In summary, in order to apply Step 9, we have to:

- define a starting time intervals t_{SDR} for the service vehicle beginning of the re-balancing around the city;

- define a ending time intervals t_{EDR} for the service vehicle to end the re-balancing actions;
- define a time horizon t_H^N for the route plan;
- define a roll period t_R^N to update the route;
- define a penalization γ to avoid solutions with unnecessary vehicle moves.

4.2.10 Step 10: preparing the simulation of a new day

Step 10 concludes the day and introduces the next day. At midnight, if in Step 6 the night route has been chosen, the stations state is set to the desirable level of Step 5 and all the following steps are followed for each day until the end of the time period.

4.2.11 Step 11: collecting results at the end of the period

At the end of the time period, the scenario simulation is concluded and it is ready to be evaluated. The key performance indicator to compare different scenarios is the total number of violations (see Section 1.9) occurred in the analysed time period. We recall that we historical data of users demand are used to calculate the total number of violations, for each scenario.

4.3 A selected set of configuration scenarios

In this section, we go through the selection of 15 different configuration scenarios that are relevant for the operation of a docking stations BSS. All the scenarios considered are summarized in Figure 4.1.

4.3.1 Assumptions

Common parameters

There are some common parameters that are assumed as given, for all the scenarios, at the strategic level decision, in particular:

- the number of stations;
- the location of the stations;
- the number of service vehicles;
- the maximum number of users. In this case there are no limitation of the number of users that can be registered to the BSS.

Notice that the number of bikes and docking stations, and their allocation, given by the strategic level, are not common to all the scenarios, as in the scenarios with unlimited resources the number of bikes and docking stations in the system may increase.

4.3.2 Scenarios classification

A hierarchical method has been used to classify the scenarios into groups with a similar BSS structure, following the order of which decision needs to be made by the service provider (see Figure 4.1).

RESOURCES	LIMITED												UNLIMITED		
STATION SIZE	STANDARD						OPTIMAL						UTOPIC		
NIGHT ROUTE	NO		YES				NO		YES				NO	YES	
INITIAL STATION STATE	STD		STD		OPT		STD		STD		OPT		STD	STD	OPT
DAILY ROUTE	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO	NO	NO
SCENARIO	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Figure 4.1: The structure of the scenarios classification.

First criterion: resources

The first criterion is between cases with:

- limited resources: the total number of bikes and docking stations are given by the strategic level (see Section 2.2.3 and 2.2.4);
- unlimited resources: the number of docking stations depends on the utopic stations size C_s^{NU}, and the number of bikes B^{NU} comes from the docks per bike ratio that we have in the scenarios with limited resources. The docks per bike ratio is equal to:

  $$R^{DpB} = \frac{C^{NS}}{B^{NS}}$$

  Hence, the number of bikes B^{NU} is equal to:

  $$B^{NU} = \left\lfloor \frac{C^{NU}}{R^{DpB}} \right\rfloor$$

Second criterion: stations size

The second criterion distinguish between cases with:

- standard stations size C_s^{NS};
- optimal stations size C_s^{NO};
- utopic stations size C_s^{NU}.

Third criterion: night route

The third criterion distinguish between cases with:

- no night route (the state of a station does not change during the night);
- optimal night route.

Fourth criterion: stations state

In case the night route is included in the scenario, the fourth criterion distinguish between cases with:

- standard initial stations state $L_{s,d}^{S}$;
- optimal initial stations state $L_{s,d}^{O}$.

In case that there is no night route, the standard stations state is chosen as initial stations state for the first day of the time period of simulation.

Fifth criterion: daily route

The fifth criterion distinguish between cases with:

- no daily route;
- optimal daily route.

4.3.3 Excluded scenarios

From Figure 4.1, we notice that some scenarios are not taken into consideration because they are assumed:

- redundant: in the scenarios in which there is not a night route, the stations state needs to be set only for the first day of the time period of simulation. Hence, the cases in which we use the optimal stations state are considered similar to the cases in which we use the standard stations state, because the only difference is in the first day of simulation. For this reason, when the night route is not applied, only the standard stations state is used;
- too unrealistic: in the scenarios in which there are unlimited resources, the utopic stations size configuration is a sensible investment for the service provider. Hence, we decided to not consider scenarios in which there is also a daily rebalancing service.

Chapter 5

Case study: data analysis

In this chapter, we discuss the method and the software and hardware resources used to perform all the analysis included in this thesis and we apply them to the historical data of the Bike Sharing System of Padua. The data have been kindly made available by the service provider Bicincittà S.R.L., with permission from the Municipality of Padua. We report some descriptive statistics of the current Padua BSS for the problems at the strategic level (see Section 2.2) and we analyse the user demand and its relation to time intervals, weekdays, holidays, months, seasons or weather conditions. Finally, we focus on the prediction of the pick-up and delivery demand for each station of the BSS in each time interval of the day, which is useful for the evaluation process described in Chapter 4.

5.1 Implementation

5.1.1 Technological tools

The software, the work environments and the hardware used for the analysis and to solve the optimization models are listed in Table 5.1.

Excel (*Microsoft 2020*) has been used to calculate the solution for the optimal stations size and optimal stations state models, using the *Solver* tool.

R (*R 2020*) has been used to calculate the driving distances in kilometres and the driving distance in seconds between all pairs of stations, using an API that extracted

Software (Enviroment)	Version	Tasks
Microsoft Excel	Office16	Optimal stations size and optimal stations state
R (Rstudio)	3.5.1	Driving distance and driving time between stations
Google Maps API	-	Driving distance and driving time between stations
CPLEX Studio IDE	12.9.0	Optimal daily route
Python (Jupyter Notebook)	3.7.6	Data analysis, forecast and simulations
Hardware	RAM	Processor
Window 7 64 bit	8 GB	Intel Core i7 2.30 GHz

Table 5.1: The software and hardware used to perform the data analysis and to solve the optimization models.

this information from Google Maps (*Google 2020*). In particular, we used the library `gmapsdistance` (*Melo et al. 2018*) in R. We gave as input the coordinates of each pair of stations and we received back the driving distance between them.

CPLEX Studio IDE (*IBM 2020*) is a MILP solver used to solve optimization models. In this thesis, CPLEX has been used to solve the dynamic model introduced by *Contardo et al. 2012* for the optimal daily route. We also implemented the static model for the night route introduced by *Dell'Amico et al. 2014*, but we decided not to include it in the simulation process, since as using only one service vehicle for the case study of Padua is not always feasible.

Python (*Python 2020*) has been used to analyse the historical data of the case study of Padua and to run the simulations for the different scenarios introduced in Section 4.2. The choice of Python comes from its simplicity of use and the possibility to use an API toward CPLEX. The connection between Python and CPLEX is one of the crucial aspects of the thesis, because it allowed to calculate the optimal daily route in CPLEX and to save the solution routes in Python, for all the times that was needed during the simulation. The connection between the two software allowed the automation of the scenarios evaluation process. Furthermore, we want to underline that Microsoft Excel and CPLEX Studio IDE are available in their free version thanks to a student license. On the other side Google Maps API, R, Python, Rstudio and Jupyter Notebook, are available as open source software.

Scenario	Daily Route	Scenarios 5 min	1 hour	1 day	1 year
1, 3, 5, 7, 9, 11, 13, 14, 15	No	0.00s	0.00s	0.01s	2.94s
2, 4, 6, 8, 10, 12	Yes	0.22s	2.62s	1m 3s	6h 23m

Table 5.2: The computational time for the simulation of one scenario for different units of time.

5.1.2 Computational time

The simulation of the Padua BSS in different scenarios has been implemented by Python and calling a CPLEX API when the optimal daily route needed to be calculated, following the evaluation process of Chapter 4. The computational time of the scenarios simulation is a crucial aspect of the analysis and it strongly depends on the optimal daily route. In Table 5.2 the scenarios are split between the scenarios in which a daily rebalancing is done and the ones in which it is not done. For each group of scenarios, it is reported the average computational time to simulate 5 minutes, one hour, one day and one year, of BSS activity. The computational time to simulate all the 15 scenarios is of about 38 hours.

5.2 Data gathering

5.2.1 Time Period

The historical data available for the analysis contain all the information from the start of the service in Padua, July 2013, up to May 2019. In this thesis, we refer our analysis to data from 1st January 2014 to 31st December 2018 to maintain only full year data.

5.2.2 Datasets

Three datasets are available for the analysis:

- trips dataset: the dataset has made available by the service provider and contains:

  - the bikes ID for all the trips;
  - anonymous ID and subscription period for all the users;
  - starting and ending station for all the trips;
  - starting and ending time (date, hour, minute) for all the trips.

- stations dataset: the dataset has been made available by the service provider and contains, for each station:
  - station number;
  - station name;
  - station location (latitude, longitude);
  - station size.

- weather dataset: historical weather information has been gathered from the website `https://www.3bmeteo.com` (*Meteosolutions 2020*) for each day of the time period between 2014 and 2018. In particular the information available are:
  - weather time (date);
  - minimum temperature (°C);
  - mean temperature (°C);
  - maximum temperature (°C);
  - rainfall (mm);
  - weather conditions (clear, cloud, snow, rain, thunderstorm).

Driving distance between stations

As during the night we minimize the length of the route to visit all the stations that needs to be rebalanced (see Section 2.3), we considered the driving distance among stations. In Table 5.3, we can see some of the distances and we can notice that, for example, from Station 0 to Station 28 the road distance is 4.204 km, while the opposite route is of 3.346 km, due to asymmetric road constraints, e.g. one-way streets.

Station	0	1	2	$\cdots$	28
0	0	4.685	5.032	$\cdots$	4.204
1	4.434	0	0.347	$\cdots$	2.518
2	4.368	0.375	0	$\cdots$	2.452
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
28	3.346	2.533	2.880	$\cdots$	0

Table 5.3: The stations driving distances matrix: each cell represents the distance in kilometers from the station in the row to the station in the column.

Station	0	1	2	$\cdots$	28
0	0	490	547	$\cdots$	445
1	514	0	56	$\cdots$	417
2	511	67	0	$\cdots$	415
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
28	414	482	538	$\cdots$	0

Table 5.4: The stations driving time matrix: each cell represents the time in seconds from the station in the row to the station in the column.

Driving time between stations

Regarding the daily routing of the vehicle (see Section 2.4), it is crucial to know how much time is needed to go from a station to another (see Table 5.4), since we want to anticipate the user demand of parking spaces or bikes in each station.

5.2.3 Training, validation and test set

In Section 5.7, the forecast of the users demand will be done using statistical models that needs to be evaluated in a dataset that is not used for the training of the model. Hence, the dataset available have been split in three sets:

- training set: the data from 1st January 2014 to 31st December 2015 are used as training set;
- validation set: the data from 1st January 2016 to 31st December 2016 are used as validation set to choose the best hyper-parameters for some of the models;
- test set: the data from 1st January 2017 to 31st December 2017 are used as testing set to evaluate the accuracy of different models.

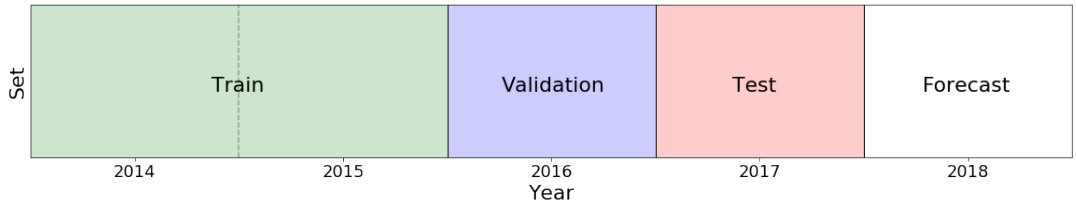

Figure 5.1: The split in train, validation, test and forecast set with the related time period.

Notice that the data from 1st January 2018 to 31st December 2018 have not been included in the evaluation of the models because this time period will be used to simulate different scenarios and we do not want to bias the choice of the model that will forecast the data for this time period. Figure 5.1 shows the split of the dataset with the related time period.

5.2.4 Explanatory variables and response variables set

Another subdivision of the datasets is between the explanatory variables and the response variables:

- explanatory variables set: it contains all the features of time and weather;
- response variables set: it contains the pick-up and delivery operations for each station.

5.3 Data pre-processing

Time intervals

The trips dataset contains all the users trips in chronological order from the first trip occurred in 2014 but we need to aggregate the users demand by time intervals, as the optimization model of Section 3.5, for the daily re-balancing operations, needs the forecast of the users demand in time intervals. The demand for bikes and docking stations changes in the course of the day, depending on the routine of the users. The time intervals of these predictions need to be of a few minutes, because if the BSS is active during the day, the timing is fundamental for the service vehicle to anticipate possible disservices at the stations. Previous studies in the literature suggest a one hour time step aggregating all the pick-up or delivery operations occurred in the same one hour interval of the day (*Rudloff et al. 2014*). The author reports that the one hour interval is a good trade-off between overfitting and precision of the user demand predictions. For example, if use a 5 minutes time interval, the predictive models does not follow only the pattern of the historical data, but also the noise due to some peaks intervals in the past (training set), that could not represent a peak in the future (test set), overfitting the data. On the other side, if we use a 4 hours time interval, the predictive models can do a good estimation of the users demand in these large time intervals, but they are not accurate to represent the peak hours, underfitting the data. For this reason, it is suggested to use a time interval between 10 minutes and 1 hour. So, for example, if we need a short term prediction in 5 minutes intervals but the model just provides a long term prediction in one hour intervals, we can divide the long term prediction in 12 equal short term 5 minutes intervals.

We decided to aggregate the users demand in time interval of 5 minutes, which should be both long enough to give time to the re-balancing vehicle to reach a station, and short enough to reach more than one station in each time horizon. For the prediction models, this short time interval can lead to a problem of complexity, as it means having a categorical variable with 288 labels (12 times 5 minutes for each of the 24 hours of a day). For this reason, if necessary, the users demand can be aggregated in time intervals of 10, 30 or 60 minutes, thus decreasing the complexity of the prediction models. We will choose the 10 minutes time interval, as we will see in Section 5.7.

5.4 Feature selection

As first step to prepare the dataset for the predictive models, we merge the feature matrices from the weather and time data. The weather data gathered (See Section 5.1) are the observed weather data of the same day. However, it is not realistic to use, for example, the millimetres of rain of the day, to predict for example the users demand in the morning, because we use “future”data to predict the past. The quantitative weather data in this case are called leaker variables and cannot be used for the predictions. Only the feature that contains a general textual information about the weather of the day (cloud, clear, snow, rain, thunderstorm), has been assumed as obtained from the weather predictions of the previous day and it was kept for the models. Then, from the date, some time features have been extracted as the day of the week, the month, the season, the year and the national holidays. A categorical variable to distinguish weekends and national holidays from working days has been created. Finally, the time of the day has been considered as categorical, with time intervals of 5 minutes.

With respect to the response variables, the pick-up and delivery operations have been analysed separately, so that, for each station, two columns have been added to the response variables dataset, each aggregating all the related operations occurred in the same one hour interval. In Table 5.5 and Table 5.6, we can see the explanatory variables set and the response variables set, respectively. Table 5.5 is a matrix with 525888 rows (1826 days per 24 hours per 12 time intervals of 5 minutes each) and 7 columns (the index is not counted). If we transform the categorical variables in binary variables, the matrix has 525888 rows and 77 columns. Table 5.6 is a matrix with 525888 rows (1826 days per 24 hours per 12 time intervals of 5 minutes each) and 56 columns (two for each station).

Index	**Time**	**Holiday**	**Weekday**	**Year**	**Month**	**Season**	**Weather**
2014-01-01 00:00	00:00	True	Wednesday	2014	January	Winter	clear
2014-01-01 00:05	00:05	True	Wednesday	2014	January	Winter	clear
2014-01-01 00:10	00:10	True	Wednesday	2014	January	Winter	clear
...	...	...	...	...	...	...	...
2018-12-31 23:55	23:55	False	Monday	2018	December	Winter	cloudy

Table 5.5: The explanatory variables dataset.

Index	Pick-up 1	Delivery 1	···	Pick-up 28	Delivery 28
2014-01-01 00:00	0	0	···	0	1
2014-01-01 00:05	1	0	···	0	1
2014-01-01 00:10	0	2	···	1	0
···	···	···	···	···	···
2018-12-31 23:55	0	1	···	0	0

Table 5.6: The response variables dataset.

5.5 Descriptive analysis

5.5.1 The number of stations and their location

The dataset is provided with the latitude and longitude for all the stations and the depot. The stations coordinates are interesting to understand the geographical distribution of the stations inside the city and they are fundamental at the operational level to plan the vehicle route during the re-balance of the system. The Padua BSS is composed of 28 stations located around the city center, except for four stations in the suburban area, plus one depot, or Station 0, as shown in Figure 5.2.

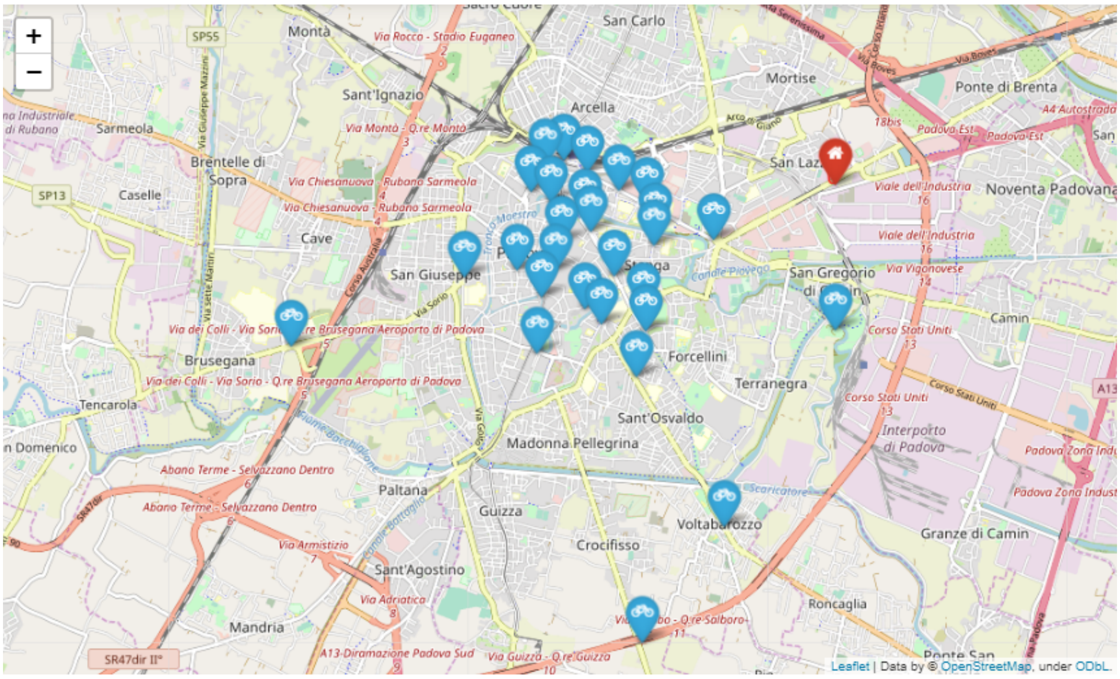

Figure 5.2: The stations location of the Padua BSS: the blue markers depict the stations, the red marker depicts the depot.

5.5.2 The number of bikes

The service provider did not provide the number of bikes available in the BSS but it is possible to have an estimation of this parameter, by counting the bikes IDs. In Figure 5.3, it is possible to see the average number of bikes used in the BSS each month, for the time period considered in the analysis. The number of used bikes varies during the time period, probably for maintenance, vandalism or renewal of the bike fleet. From the time series it is possible to extract the average number of bikes in the BSS, that, in this case, is 168 bikes, used as fixed parameter for the optimal stations state model of problem P_3^T ($B^N = 168$).

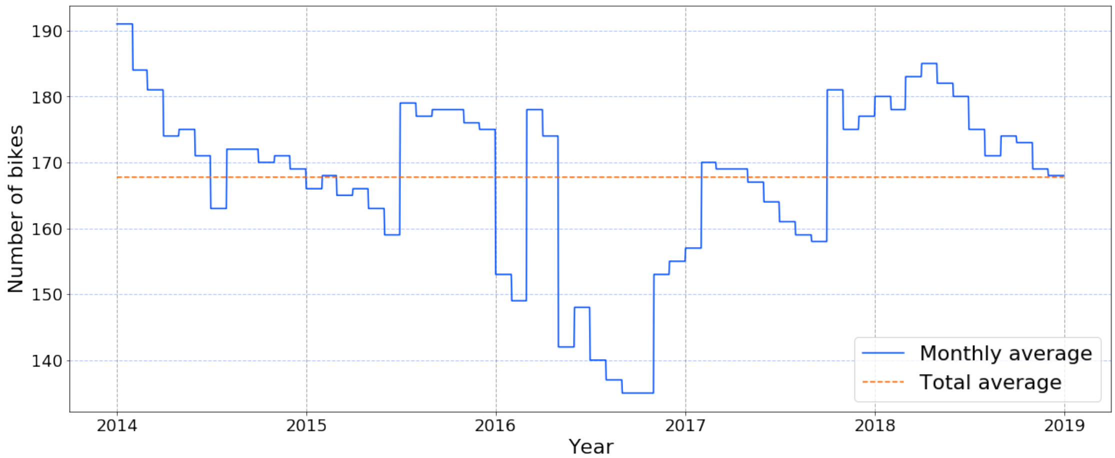

Figure 5.3: The average number of bikes per month.

5.5.3 The number of docking stations, the starting stations size and the starting stations state

Data about the stations size, and consequently the number of docking stations, are available from the web page of the BSS of the city of Padua, `http://www.goodbikepadova.it` (*Bicincittà 2020*). The docking stations in the current BSS of Padua are 350 distributed among 28 stations. The starting stations state have been calculated using the docks per bike ratio:

$$R^{\text{DpB}} = \frac{C^{NS}}{B^{NS}} = \frac{350}{168} = 2.08$$

where B^{NS} is the total number of bikes, estimated as 168 from the available datasets and C^{NS} is the total number of docking stations. In this case, following the formula introduced in Section 1.2.5, it is possible to calculate the standard initial station state $L^S_{s,1,1}$ as:

$$L^S_{s,1,1} = \min\left(\left\lfloor \frac{C^{NS}_s}{R^{\text{DpB}}} \right\rfloor, \left\lfloor \frac{C^{NS}_s}{2.25} - 0.5 \right\rfloor + 1\right).$$

Table 5.7 shows the stations size, the standard initial stations state and the standard stations size. The stations size varies from 5 to 20, except for Station 6, no longer operational since January 2018. The depot is assumed to have an infinite size and a number of bikes $L^S_{0,1}$ equal to the difference between the total number of bikes available B^{NS} and the ones in the stations:

$$L^S_{0,1,1} = B^{NS} - \sum_s L^S_{s,1,1}, \qquad \forall s \in S$$

5.5.4 The number of service vehicles

In the case of Padua, there is only one vehicle available for the rebalancing operations. The starting position of the vehicle, both for the night and the daily route, is the depot. The capacity C^r of the service vehicle r is assumed to be of 20 bikes.

5.5.5 The number of users

Trips per member

To understand the growth and the evolution of the Padua BSS, we calculate the number of users with an active subscription in the BSS in each day of the time period under analysis. Figure 5.4 shows the time series of the number of members registered in the BSS compared with the moving average of the number of trips per day. The moving average is calculated as the average of the number of daily trips in the previous 365 days, to remove the seasonality from the time series. In Figure 3.1, we can see how, after an initial enthusiasm for the service that got a peak of members subscribed in June 2014, the number of members and the number of trips have been in a decreasing trend for the following years. Notice that the number of

Station	**Name**	**Std Size**	**Std State**
1	Sarpi	14	6
2	Mazzini	13	6
3	Giotto	14	6
4	Gasometro	13	6
5	Venezia Colombo	14	6
6	Marzolo	0	0
7	Morgagni	14	6
8	Altinate	14	6
9	Duomo	11	5
10	Antenore	14	6
11	Cesarotti	18	8
12	Orsini	13	6
13	Gattamelata	12	5
14	Pontecorvo	14	6
15	Stazione	10	4
16	Venezia 2	15	7
17	I Colli	15	7
18	Della Valle	20	9
19	Park Bembo Est	15	7
20	Piovese	13	6
21	Falloppio	12	5
22	Fiere	15	7
23	Riviera Tito Livio	14	6
24	Tribunale	14	6
25	Stazione 2	14	6
26	La Fenice	5	2
27	Facciolati	5	2
28	Nazareth	5	2
	Total	**350**	**154**

Table 5.7: The standard stations size and the standard initial stations state of the BSS of Padua.

trips per day decreases more slowly than the number of members per day, which could mean that the members that are still registered in the BSS are the ones that take more advantage of the service. Indeed, Figure 5.5 shows the time series of the daily number of trips per member, which has a growing trend over time. Although the trend conveys hope for the creation of a more loyal group of members, Figure 5.5 shows that the number of trips per member is still lower than the ideal range, according to the *Bike Sharing Planning Guide* (*Gauthier 2013*).

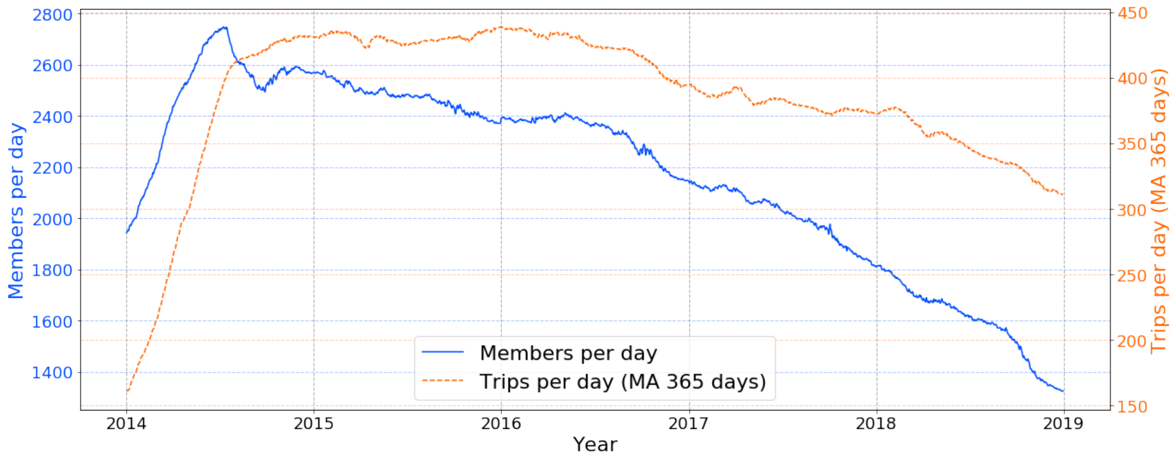

Figure 5.4: The number of members registered per day compared with the number of trips per day.

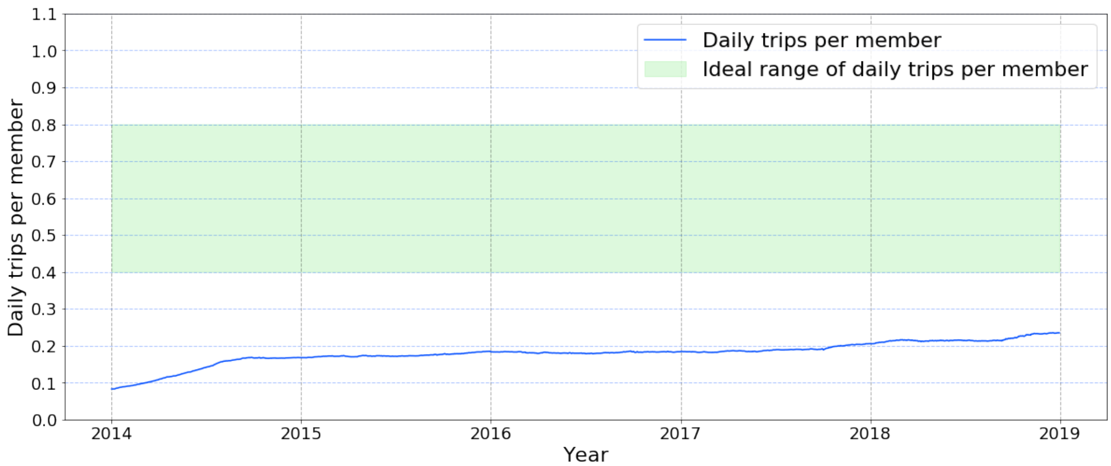

Figure 5.5: The time series of the number of trips per member.

Members per bike

The ratio between the number of members per bike, according to the *Bike Sharing Planning Guide* (*Gauthier 2013*), should be around 10, for an efficient use of the BSS. In Figure 5.6, we can see the number of members per bike in the analysed period. In this case, we can see that between February 2014 and September 2017 the number of members per bike is above the ideal ratio, constantly decreasing until December 2018.

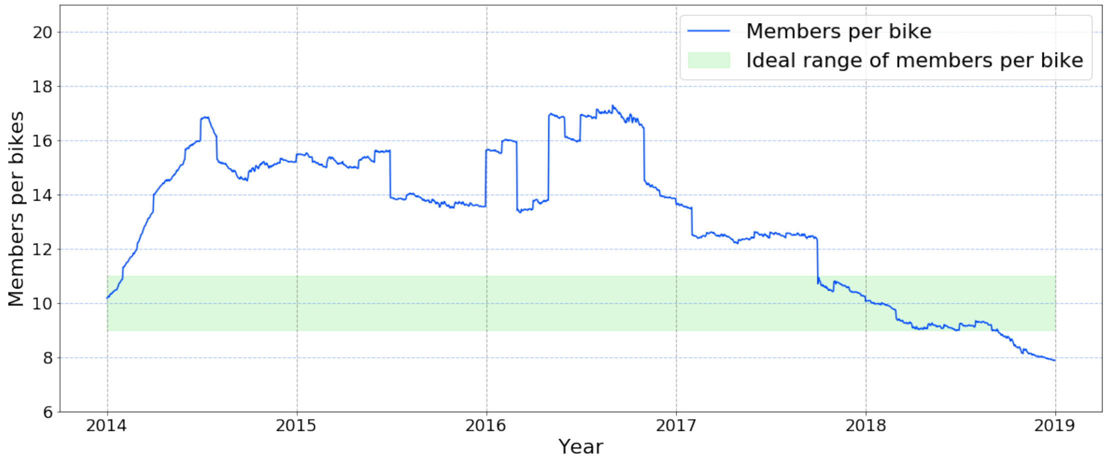

Figure 5.6: The time series of the number of members per bike.

Trips per bike

The ratio between the number of daily trips per bike, according to the *Bike Sharing Planning Guide* (*Gauthier 2013*), should be between 4 and 8, for an efficient use of the BSS by the users and a return of investment for the service provider. For the Padua BSS, it seems that the service is not used in its complete potential, maybe for a lack of cycling infrastructure or an inefficient distribution of the docking stations between the stations. In Figure 5.7, we can see the number of daily trips per bike in the analysed period.

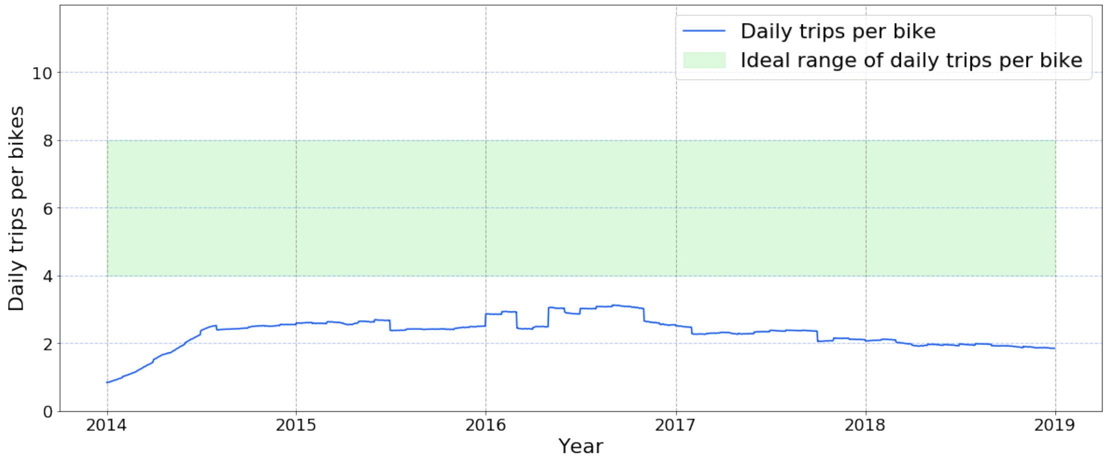

Figure 5.7: The time series of the number of trips per bike.

5.6 Users demand analysis

The users demand analysis is important to see the relation between the users demand and the other explanatory variables. The analysis is done in a subset of the original dataset, to remove the observations of the year 2018, in which we evaluate the different scenarios.

The users activity

Figure 5.8 shows the users activity (as defined in Section 1.2.4) in each time interval of the day. We can notice that the majority of the users activity is between 7:00 and 21:00, with a peak of 9.52 operations per day in the time interval 08:15-08:20. We can also notice that there are three time windows of high activity:

- in the morning between 7:30 and 9:30;
- in the afternoon between 12:00 and 14:30;
- in the evening between 16:30 and 19:00.

Figure 5.9 shows the cumulative users activity in the course of the day (see Section 1.2.4). In the period 2014-2017 the users activity has been of 817.6 operations per day, on average.

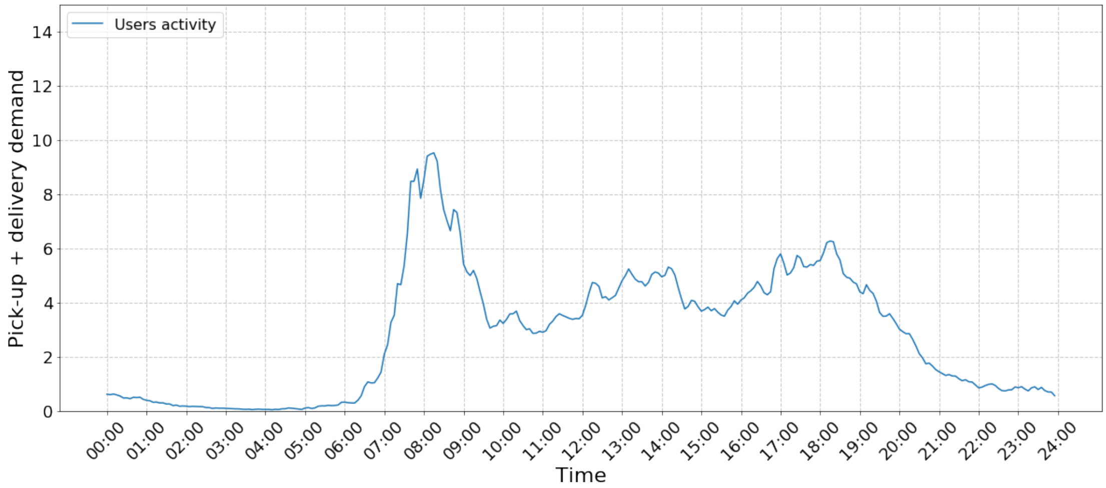

Figure 5.8: The users activity.

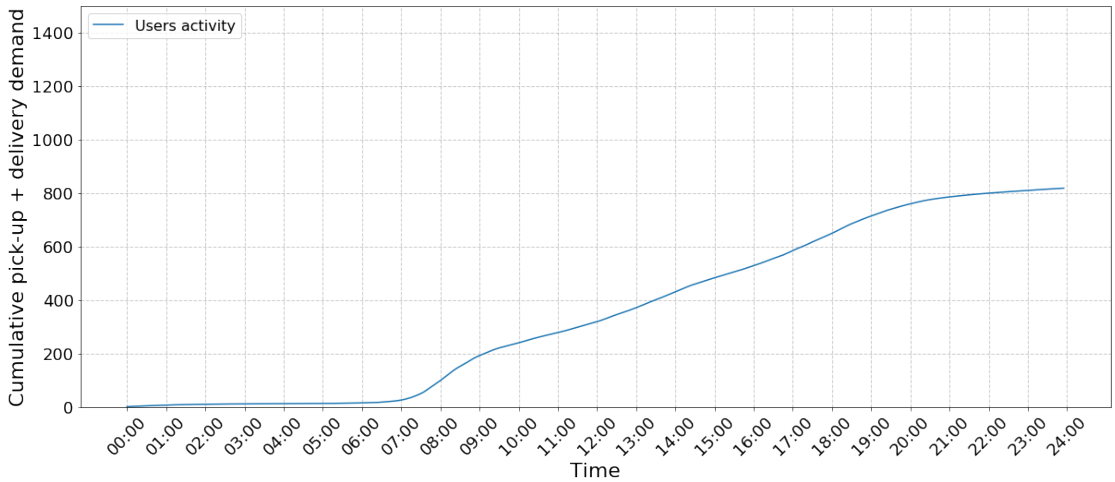

Figure 5.9: The cumulative users activity.

The users activity grouped by month

Figure 5.10 shows the average users activity in each time interval of the day, grouped by month of the year. We can notice that in August there is the lowest activity, which is likely caused by the shutdown of some companies and the university. Then, we notice in October the higher activity , which is likely due to the good weather and the beginning of a new university year. In Figure 5.11, we can see the cumulative users activity. In the figure it is clear that August (-40% than the average daily activity) has the lowest cumulative users activity, followed by January (-23%), while the months with the highest activity are May (+22%) and October (+26%). The number of operations per day in October (1031.8) are 2.1 times greater than in August (493.9).

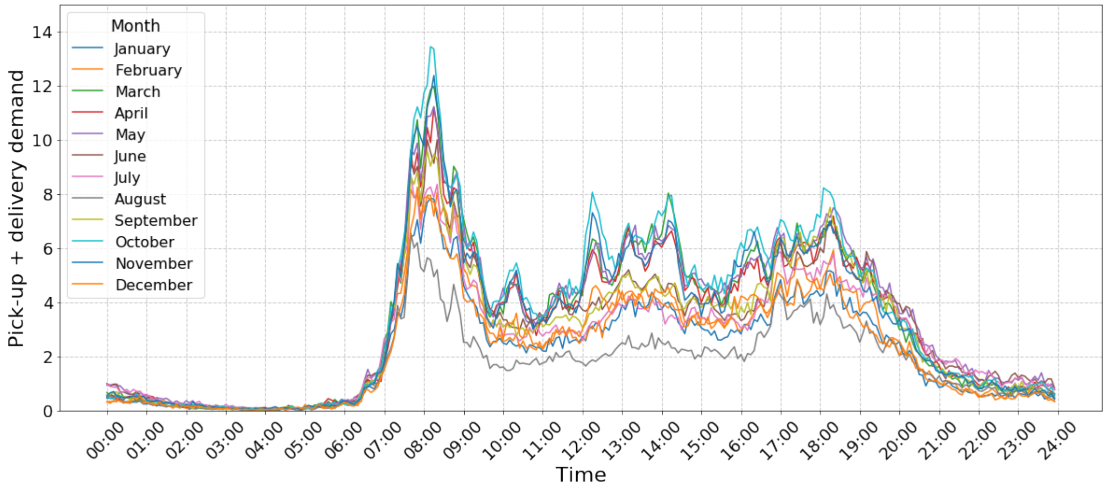

Figure 5.10: The users activity grouped by month.

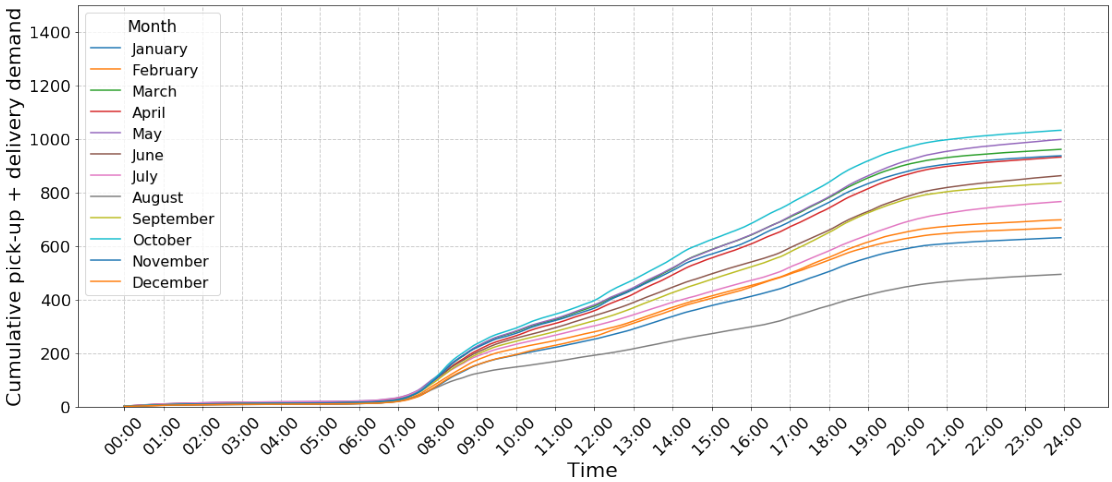

Figure 5.11: The cumulative users activity grouped by month.

The users activity demand grouped by weekday

Figure 5.12 shows the average users activity demand in each time interval of the day, grouped by weekday. In this case we can notice similar users activity for the working days from Monday to Friday, a significantly lower activity on Saturdays and the lowest activity on Sundays. In Figure 5.13, we depict the cumulative users activity by weekday. In the figure it is more clear as the highest activity is on Tuesdays

(+29.8% than the average daily activity), Wednesdays (+31.4%) and Thursdays (+30.3%), then the users activity is slightly lower on Mondays (+16.2%) and Fridays (+15.5%), while the weekdays with the lowest activity are Saturdays (-54.1%) and Sundays (-68.8%). The number of operations per day on Wednesdays (1074.4) are 4.2 times greater than on Sundays (254.8).

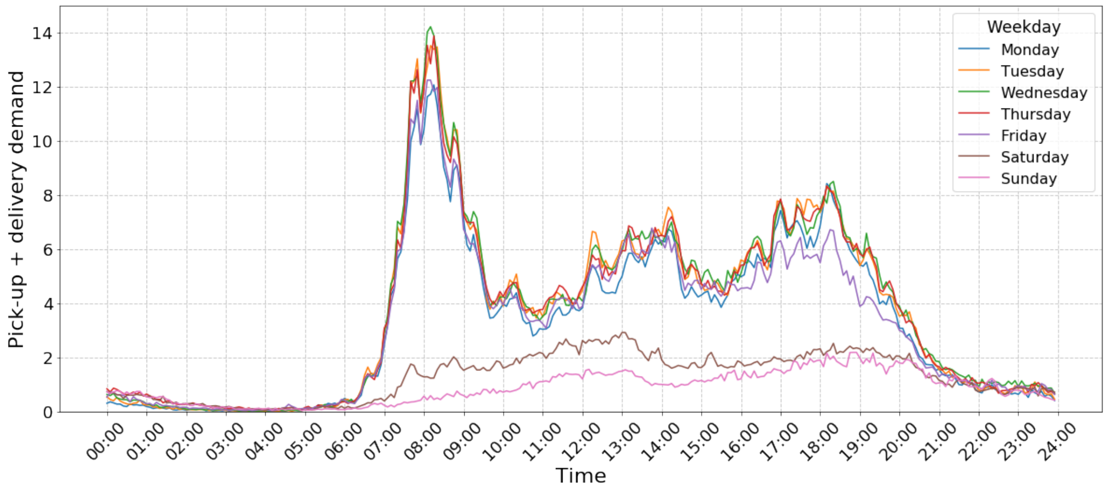

Figure 5.12: The users activity grouped by weekday.

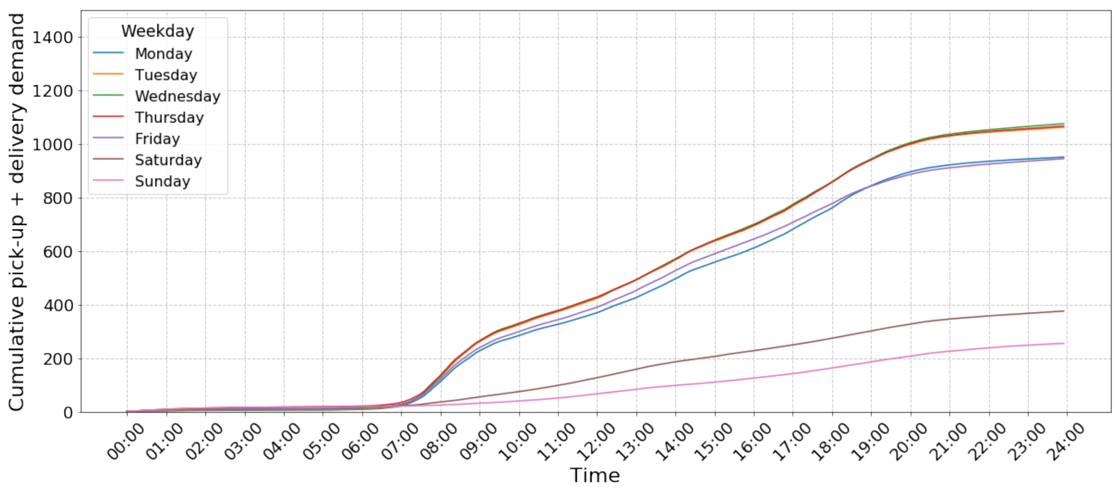

Figure 5.13: The cumulative users activity grouped by weekday.

The users activity demand grouped by season

Figure 5.14 shows the average users activity in each time interval of the day, grouped by season. In this case, we notice that the highest users activity is in Spring (+14.6%

than the average daily activity) and in Autumn (+15.5%), with very close distribution between them during all the time intervals of the day. On the other side, in Summer (-14.4%) and in Winter (-15.1%), due to extreme weather conditions or holidays, have the lowest activity. The two seasons groups are also confirmed by the cumulative plot of Figure 5.15. The number of operations per day in Spring (936.6) and in Autumn (944.1) are 1.3 times greater than in Winter (694.4) and Summer (699.7).

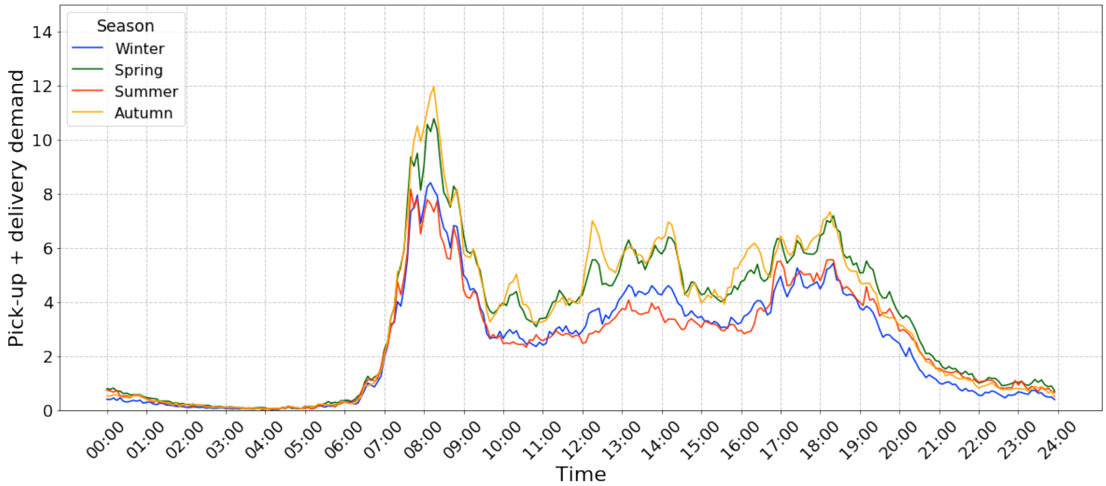

Figure 5.14: The users activity grouped by season.

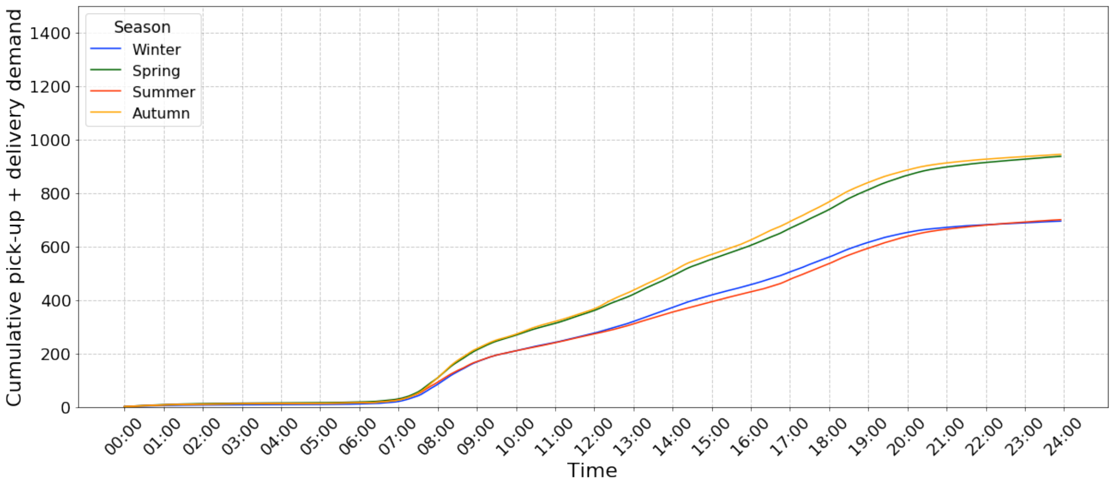

Figure 5.15: The cumulative users activity grouped by season.

The users activity demand grouped by workdays and holidays

Figure 5.16 shows the average users activity in each time interval of the day, grouped by workdays and holidays. In this case we see an high users activity during the workdays (+28.3% than the average daily activity), from Monday to Friday, while during weekends and Italian festivities the users activity is sensibly lower (-62.7%). The differences are also confirmed by the cumulative plot of Figure 5.17. The number of operations per day on workdays (1048.6) is 3.4 times greater than during the holidays (305.2).

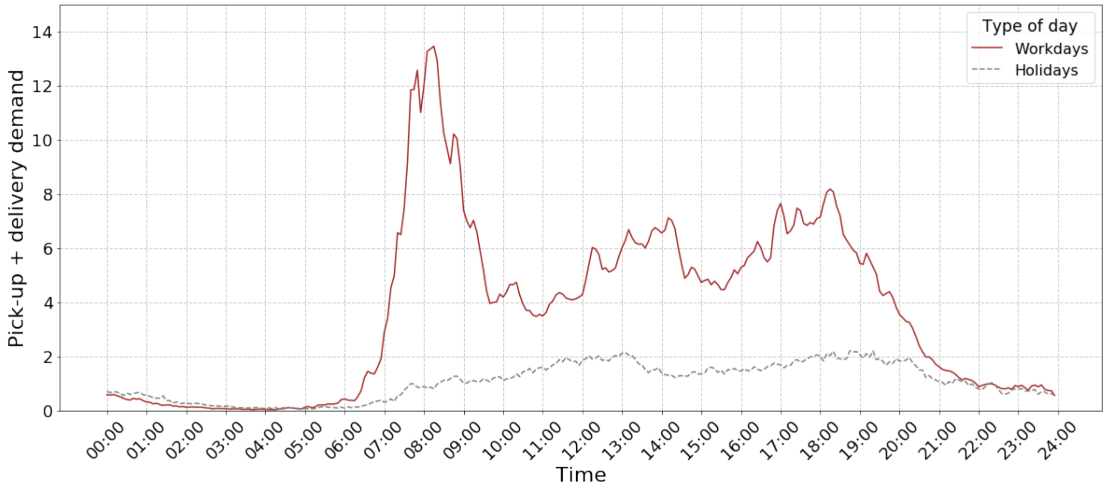

Figure 5.16: The users activity grouped by workdays and holidays.

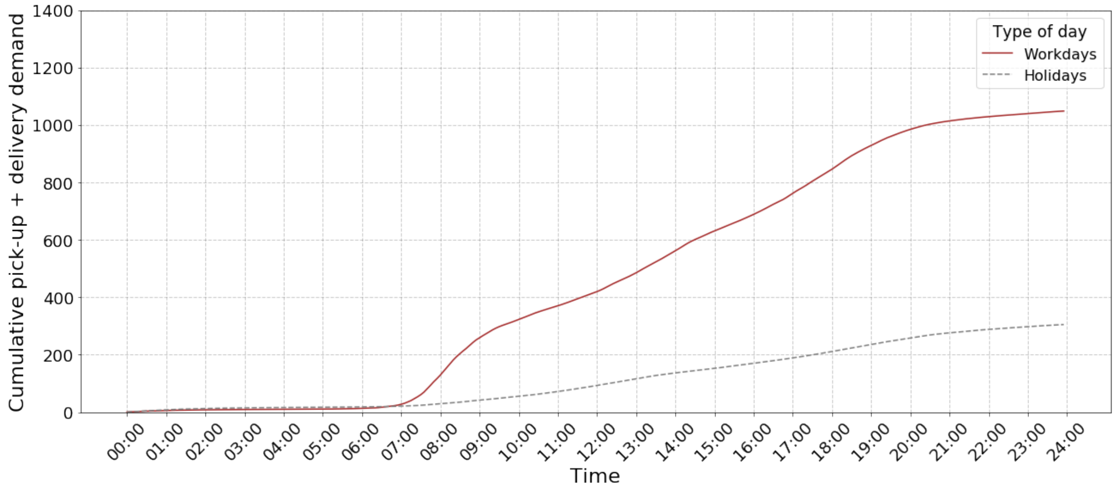

Figure 5.17: The cumulative users activity grouped by workdays and holidays.

The users activity demand grouped by weather conditions

Figure 5.18 shows the average users activity in each time interval of the day, grouped by weather conditions. In this case we see higher users activity on clear (+5.4% than the average daily activity) and cloudy (+6.1%) day, while when there is rain (-9.5%) or thunderstorm (-16.2%) the users activity is lower. Then, as expected, the lowest users activity occurs during the days of snow (-58.3%). The differences are also confirmed by the cumulative plot of Figure 5.19. Notice that the number of operations per day on a clear day (862.1) is only 1.2 times greater than on a rainy day (739.6).

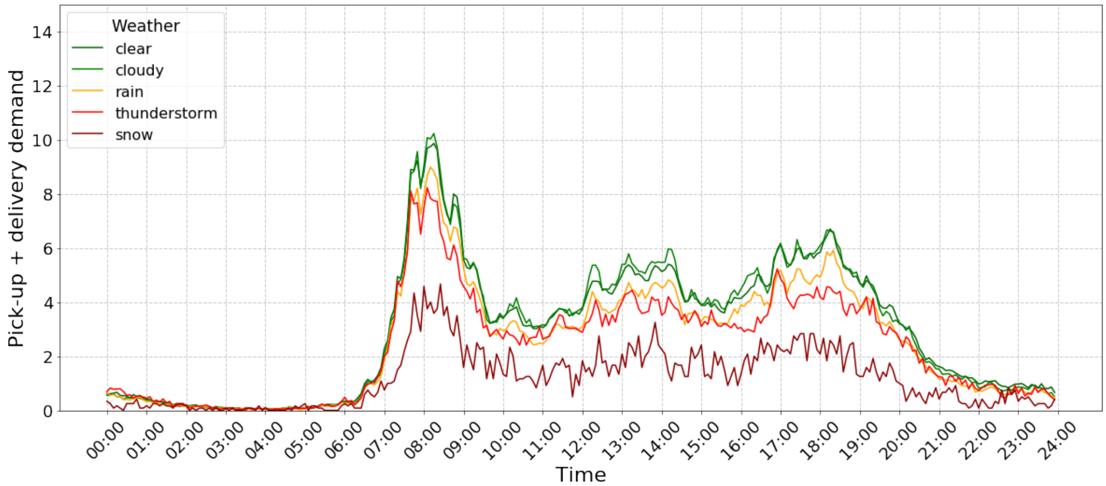

Figure 5.18: The users activity grouped by weather conditions.

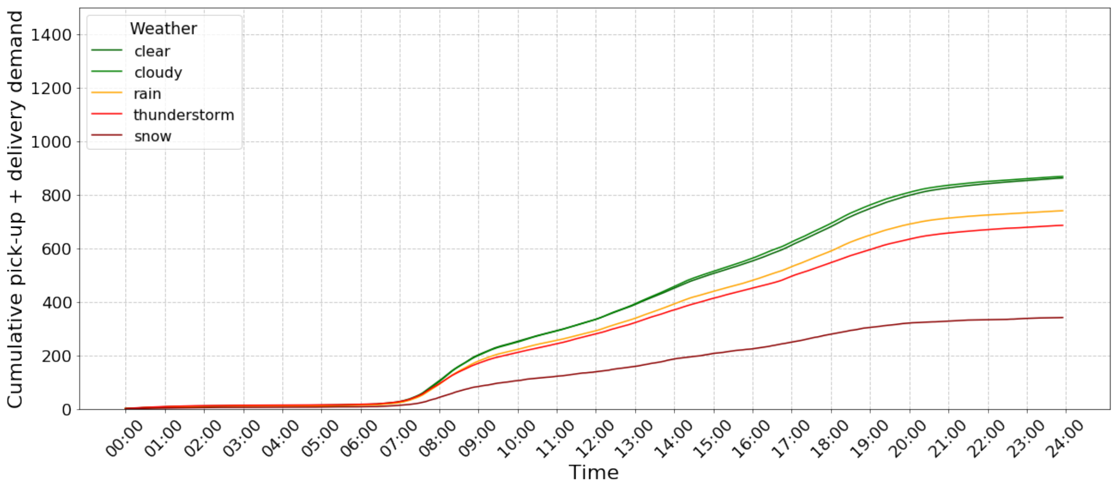

Figure 5.19: The cumulative users activity grouped by weather conditions.

5.7 Users demand forecast

The users demand forecast

The users demand forecast helps the service provider, during the rebalancing operations, to anticipate and prevent possible violations in the BSS stations. In this section, we focus on the prediction of the pick-up and delivery demand for each station of the BSS in each time interval of the day. The time interval of the day is considered as a categorical variable, so, the more the time interval is short, the more the number of categories increases. The great number of categories increases the computational time and the complexity of the models. For this reason, we decided to try three different length for the time intervals: 60 minutes, 30 minutes or 10 minutes (see Section 5.3). As the thesis is more focused on the optimization models and on the decision-making process, the models to forecast the users demand have been chosen for their simplicity and interpretability. The models that will be evaluated are the Linear Model (with two variables or complete), the Poisson Model, the Decision Tree and the Random Forest. We do not doubt that future works may give more attention to further models, like time series models or neural networks. Each model has been trained on the training set, regularized on the validation set and evaluated on the test set (see Section 5.1.3). In particular, we have decided to add the interaction variable between the time of the day and the holidays variable, since we decided to treat differently working days and holidays (Figure 5.16 supports this decision).

Evaluation metric

To evaluate the accuracy of the prediction models, the following metric has been used:

- Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}$$

  where y_i are the real values, $\hat{y}_i$ the predicted values and n the total number of observations.

Prediction models

We decided to evaluate a simple model to have a base:

1 - Benchmark model: a linear model using, as explicative variables, the categorical variables of the time of the day and the day of the week (workdays or holidays). This model is considered as base for its simplicity and interpretability.

Hence, we can compare the predictions accuracy (see Section 5.7.2) of the benchmark model with other more complex models. Notice that the hyper-parameters for the Decision Tree and the Random Forest are the ones that minimized the Root Mean Squared Error (RMSE) on the validation set. Then, all the models have been trained with the train and validation set together and the evaluation is obtained on the test set. Hence, we estimated the following models:

2 - Linear model: a linear model for each station using all the explanatory variables plus the interaction between the time and the holiday variable;

3 - Poisson model: a Poisson model for each station using all the explanatory variables plus the interaction between the time and the holiday variable;

4 - Decision Tree model: the decision tree for each station has been trained with a max depth of 16 and a minimum of 15 samples per leaf, using all the variables plus the interaction between the time and the holiday variable;

5 - Random Forest model: the random forest for each station has been trained with 16 trees, a max depth of 30, a minimum of 15 samples per leaf and all the features plus the interaction between the time and the holiday variable for each tree;

6 - Ensemble model: an ensemble of model 2, 3, 4 and 5, using the average of the predictions of these models to avoid the overfitting and to increase the accuracy on the test set.

Model	RMSE 60 minutes	RMSE 30 minutes	RMSE 10 minutes
1 - Benchmark	0.10457	0.10716	0.09986
2 - Linear model	0.10460	0.10478	0.09988
3 - Poisson model	0.10475	-	-
4 - Decision tree	0.10531	0.10751	0.10574
5 - Random forest	0.10476	0.10524	0.10304
6 - Ensemble model	0.10465	-	-

Table 5.8: The accuracy of the predictive models in the test set for the users demand for different time intervals of the variable time of the day. Note: the symbol $-$ means that it was not possible to estimate that model, due to its complexity.

For each model, three different cases have been tested with respect to the time intervals (60 minutes, 30 minutes or 10 minutes).

Models evaluation

The models have been evaluated using the RMSE metric, previously introduced. In Table 5.8, we can see the RMSE of the different models in test set for the users demand. In this case, we can see that the Benchmark model obtain the lowest RMSE, using the time as a categorical variable with 144 different intervals of 10 minutes, better than all the other models. Finally, we can notice that, for some models, due to the complexity of the model when the variables increased, it was not possible to calculate the accuracy. Hence, the Linear model with time intervals of 10 minutes and the holiday variable is chosen to forecast the users demand in the time period of year 2018. These predictions will be later used during the simulation of the different scenarios with the daily rebalancing route (see Section 4.2).

Chapter 6

Case study: initial system settings

6.1 Stations size

In this section, we set the stations size according to the three different definitions: standard, optimal or utopic stations size (see Section 1.7 and Section 4.1). For the BSS of Padova, the standard stations size is given by the current configuration of the BSS.

Regarding the optimal and utopic stations size, we should use the predictions of the users demand for the year 2018. In this case, the predictions of the Linear model selected in Section 5.7, are more suitable to predict the time intervals during the day, but not to predict the maximum or minimum stations state, as it is needed in phase of configuration of the optimal and utopic stations size. As the peaks of the users demand are difficult to predict, the linear model tends to underestimate them, suggesting that only one stations needs to be oversized (Station 25). Furthermore, if we use the predictions to calculate the utopic stations size, we obtain a solution with only 158 docking stations, 55% lower than the standard stations size, that has 350 docking stations. For this reason, we decided to analyse the historical trips data of the Padua BSS for the period between 2014 and 2017, assuming that they are a suitable representation of the peaks that may occur in the stations in the following years. Hence, the data are gathered from the response variables dataset for the time period from 1st January 2014 to 31st December 2017, in time intervals of 5 minutes. The optimal stations size is determined by the optimization model of Section 3.2, while the utopic stations size is determined as indicated in Section 1.7.

Table 6.1 shows the results for the standard, optimal and utopic stations size, with the related total number of docking station. Also, Table 6.1 shows the percentage of docking stations added or removed in each station in the optimal and utopic solutions with respect to the standard station size.

The table shows that the utopic stations size configuration has a total of 543 docking stations in the system, 193 more than the other two configurations (+55% than the standard and optimal stations size solutions). If we consider more in detail the specific stations, we can see that the optimal stations size solution suggests to undersize some stations in favour of others that have more necessity of permanent extra docking stations. For example, we can see that Station 25 should be enlarged (from 14 docking stations in the standard configuration to 41, +193%) and Station 11 could be lessened (from 18 to 11, -39%). Regarding the utopic configuration, we can notice that the stations that needs to be enlarged the most are Station 5 (from 14 to 40, +186%), Station 15 (from 10 to 26, +160%) and Station 25 (from 14 to 56, +400%). On the other side, some stations could be lessened, as Station 17 (from 15 to 10, -33%) and Station 19 (from 15 to 8, -47%). In Chapter 7, we will see if a change in the configuration of the stations size lead to a decrease of the number of violations in the BSS. Figure 6.1 shows the comparison between the three different choices of station size for each station of the BSS. The boxplots indicate the distribution of the stations shortage in the time period between 2014 and 2017, the pink background depicts the standard station state and the markers represents the suggested number of docking stations per station for the optimal and utopic stations size solution. Figure 6.1 shows that the standard size for Station 25 (14 docking stations) is lower than the median required docking stations per day (20 docking stations). Other critical problems can be identified for Station 5, 15, 27 and 28, in which the current station size is lower than the third quartile of the required docking stations distribution per day. Furthermore, we can notice that, for the optimal solution, the ideal number of docking stations per station is between the third quartile and the maximum value of the required size per day, for all the stations.

Station	Station Size			Station Size Changes (%)		
Number	**Standard**	**Optimal**	**Utopic**	**Standard**	**Optimal**	**Utopic**
1	14	9	19	0 %	-36 %	+36 %
2	13	12	18	0 %	-8 %	+38 %
3	14	9	13	0 %	-36 %	-7 %
4	13	11	19	0 %	-15 %	+46 %
5	14	26	40	0 %	+86 %	+186 %
6	0	0	0	0 %	0 %	0 %
7	14	13	21	0 %	-7 %	+50 %
8	14	19	25	0 %	+36 %	+79 %
9	11	13	19	0 %	+18 %	+73 %
10	14	17	26	0 %	+21 %	+86 %
11	18	11	18	0 %	-39 %	0 %
12	13	9	16	0 %	-31 %	+23 %
13	12	11	19	0 %	-8 %	+58 %
14	14	12	18	0 %	-14 %	+29 %
15	10	19	26	0 %	+90 %	+160 %
16	15	16	22	0 %	+7 %	+47 %
17	15	5	10	0 %	-67 %	-33 %
18	20	13	18	0 %	-35 %	-10 %
19	15	5	8	0 %	-67 %	-47 %
20	13	7	12	0 %	-46 %	-8 %
21	12	17	26	0 %	+42 %	+117 %
22	15	17	27	0 %	+13 %	+80 %
23	14	10	17	0 %	-29 %	+21 %
24	14	11	19	0 %	-21 %	+36 %
25	14	41	56	0 %	+193 %	+300 %
26	5	1	4	0 %	-80 %	-20 %
27	5	8	15	0 %	+60 %	+200 %
28	5	8	12	0 %	+60 %	+140 %
Total	350	350	543	0 %	0 %	+55 %

Table 6.1: The stations size (standard, optimal and utopic), with the related total number of docking stations and the percentage difference with respect to the standard stations size.

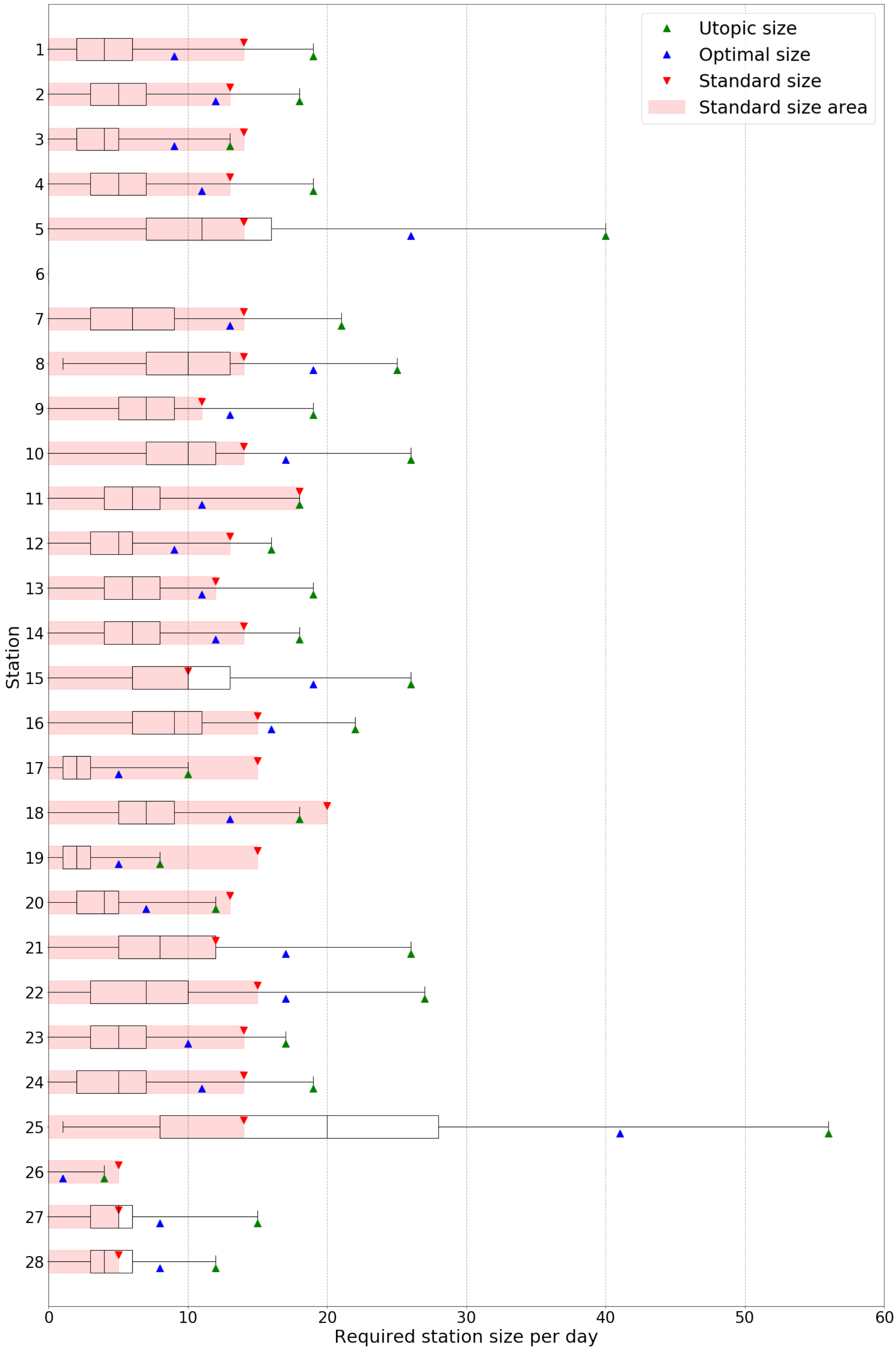

Figure 6.1: The distribution of the required stations size in the time period between 2014 and 2017. The markers indicated the standard, optimal and utopic station size.

6.2 Stations state

In this section, we examine the different configurations for the initial stations state, given the stations size values obtained in Section 6.2. For each choice of the stations size, there are two possibilities for the service provider: the standard initial stations state (see Section 1.2.5), or calculate the optimal stations state based on the historical demand of bikes and docking stations in each BSS station (see Section 4.1.5).

We recall that the desired stations state are effective only if the night rebalancing route is done (see Section 4.1.6). The standard initial station state is determined by the docks per bike ratio R^{DpB} (see Section 1.2.5), while the optimal station state is determined by the optimization model of Section 3.3, applied to the historical trips data of the Padua BSS of the period 2014-2017 instead of the predictions of 2018, for the same reasons discussed in Section 6.1.

Table 6.2 shows the configuration of the standard, optimal and utopic stations size, with the related standard and optimal stations state and the total number of bikes and docking stations. The last row of Table 6.2 shows that in the standard stations size configuration, the standard stations state has only 154 bikes, out of 168 available (see Section 5.2.2). The use of less bikes in the standard stations state solution is due to the docks per bike ratio R^{DpB}, that tends to suggest more than 2 docks per bike (2.25). This conservative criteria is not applied for the optimal stations state solution, that in fact uses all the 168 bikes, in the working days, and 167 bikes, in the holidays. More in detail, the R^{DpB} is equal to 2.27 in the standard solution, while it is only 2.09 in the optimal solution. The decrease of the number of bike impacts on the number of full violations, because with more bikes in the network, the probability that a station is full increases. For this reason, in future work, a weighted sum could be applied in the objective function of the optimization model of Section 3.3. This weighted sum may give more importance to some types of violations with respect to others.

In the optimal stations size configuration, the standard and optimal stations state are very similar, with the optimal solution that tends to add more bikes per

	Standard Size			**Optimal Size**			**Utopic Size**		
Station Num.	**Size**	**State** Std.	Opt.	**Size**	**State** Std.	Opt.	**Size**	**State** Std.	Opt.
1	14	6	6 : 7	9	4	4 : 4	19	8	8 : 7
2	13	6	7 : 6	12	5	6 : 6	18	8	9 : 8
3	14	6	8 : 6	9	4	5 : 5	13	6	7 : 6
4	13	6	7 : 5	11	5	5 : 5	19	8	9 : 9
5	14	6	8 : 7	26	12	14 : 12	40	18	20 : 21
6	0	0	0 : 0	0	0	0 : 0	0	0	0 : 0
7	14	6	8 : 6	13	6	7 : 6	21	9	12 : 9
8	14	6	6 : 7	19	8	9 : 9	25	11	11 : 12
9	11	5	6 : 6	13	6	7 : 6	19	8	10 : 10
10	14	6	6 : 7	17	8	8 : 9	26	12	12 : 13
11	18	8	9 : 9	11	5	6 : 6	18	8	9 : 9
12	13	6	6 : 7	9	4	4 : 4	16	7	8 : 9
13	12	5	6 : 6	11	5	6 : 5	19	8	10 : 9
14	14	6	8 : 7	12	5	7 : 6	18	8	10 : 9
15	10	4	5 : 6	19	8	9 : 10	26	12	13 : 13
16	15	7	8 : 7	16	7	8 : 8	22	10	11 : 10
17	15	7	6 : 7	5	2	2 : 3	10	4	4 : 5
18	20	9	9 : 10	13	6	6 : 6	18	8	8 : 9
19	15	7	6 : 3	5	2	2 : 3	8	4	4 : 4
20	13	6	7 : 6	7	3	4 : 4	12	5	6 : 6
21	12	5	6 : 6	17	8	9 : 8	26	12	13 : 14
22	15	7	5 : 7	17	8	6 : 7	27	12	9 : 13
23	14	6	8 : 7	10	4	6 : 6	17	8	10 : 9
24	14	6	7 : 7	11	5	5 : 6	19	8	9 : 11
25	14	6	2 : 6	41	18	15 : 16	56	25	24 : 21
26	5	2	2 : 3	1	0	0 : 0	4	2	1 : 2
27	5	2	3 : 3	8	4	4 : 4	15	7	8 : 7
28	5	2	3 : 3	8	4	4 : 4	12	5	6 : 6
Total	350	154	168 : 167	350	156	168 : 168	543	241	261 : 261

Table 6.2: The configuration of the standard, optimal and utopic stations size, with the relative standard and optimal stations state. For the optimal stations state, the symbol : distinguish between working days and holiday station state.

station. Indeed, the total number of bikes is 156 in the standard solution and 168 for the optimal solution. About the specific stations, we can see a small difference of stations state for Station 5 that starts with 12 bikes in standard solution, while in the optimal solution it starts with 14 bikes in the working days and with 12 bikes in the holidays. Station 25 starts with 18 bikes in standard solution, while in the optimal solution it starts with 15 bikes in the working days and with 16 bikes in the holidays. Finally, we can observe the different solutions in the utopic stations size configuration. In this case, we assume to have unlimited resources, so that given the total number of docking stations $C^{NU} = 543$, we estimated the number of bikes available as the docks per bike ratio that we have in the scenarios with limited resources. We can calculate the docks per bike ratio as:

$$R^{DpB} = \frac{C^{NS}}{B^{NS}} = \frac{350}{168} = 2.08$$

where C^{NS} and B^{NS} are the total number of docking stations and the total number of bikes available in the standard stations size configuration. The number of bikes available B^{NU} in the utopic stations size configuration are calculated as:

$$B^{NU} = \left\lfloor \frac{C^{NU}}{R^{DpB}} \right\rfloor = \left\lfloor \frac{543}{2.08} \right\rfloor = 261$$

In summary, the utopic stations size configuration we have 543 docking stations and 261 bikes available and in this case, the standard stations state solution uses only 241 bikes, while the optimal stations state solution uses all the 261 bikes.

6.3 Night route: initial load

The night route is run by one re-balancing vehicle that starts its route from the Depot. This vehicle has a specific capacity and the initial load has to be assigned. Although the vehicle capacity is fixed, and assumed to be of 20 bikes, the vehicle initial load depends on the total number of bikes at the Depot. The number of bikes available impacts the night route of the vehicle. For example, if no bikes are left in the Depot, the route always starts from the visit of a station in which a pick-up

Stations					Depot	Vehicle
Type	**Size**	**Bikes**	**State**	**State**	**State**	**Load (L_0^r)**
Standard	350	168	Standard	154	14	10
			Optimal	168 : 167	0 : 1	0 : 1
Optimal	350	168	Standard	156	12	10
			Optimal	168 : 168	0 : 0	0 : 0
Utopic	543	261	Standard	241	20	10
			Optimal	261 : 261	0 : 0	0 : 0

Table 6.3: The vehicle load at the beginning of the rebalancing operations, for different configurations of stations size and state. For the optimal stations state, the symbol : distinguish between working days and holiday.

operation is needed, though there could have been another shorter route that starts from a station in which a delivery operation needs to be done. We decided to set a maximum initial load of 10 bikes, that is the half of the vehicle capacity. Hence, the initial load is the minimum between 10 and the number of bikes available in the Depot at day d:

$$L_0^r = \min(L_{0,1,d}, 10)$$

Table 6.3 shows the value of the initial load L_0^r for the different configuration of the stations size and states, as from Section 6.3.

6.4 Daily route: initial settings

In the case study of the city of Padua, the re-balancing operations during the day are managed by one rebalancing vehicle, that travels around the city, trying to avoid the disservices for the users. The service provider did not give details about the actual methodology to refill the stations, so it is not possible to make a comparison between the application of the optimization model of Section 3.5 and the current criterion. One goal of this thesis is to evaluate the daily operations of a service vehicle, that follows the instruction given by the optimal daily route model, under certain conditions of stations state, capacity and initial load of the rebalancing vehicle, as in the night route (see Section 6.3). Before applying the model to a real dataset, some parameters needs to be set (see Section 3.5). For the case study of the city of Padua, the parameters are the following:

- $r = 1$ is the number of vehicles;
- $C^1 = 20$ is the capacity of vehicle 1;
- L_0^1 is the initial load of the vehicle. See Table 6.3 for its value in the different configurations;
- $T^N = 288$ is the total number of time intervals per day, as t has a length of 5 minutes;
- $T_H^N = 6$ is the length of the time horizon included in the model, i.e. the next 30 minutes (6 time intervals of 5 minutes, see Section 4.1.9);
- $T_R^N = 3$ is the length of roll period to update the rebalancing route, i.e. after 15 minutes (3 time intervals of 5 minutes, see Section 4.1.9);
- u^1 is the station in which the vehicle is located. In the first time horizon, u^1 is the Depot (or Station 0), but the next u^1 will be updated depending on the current station of the vehicle (the last visited station);
- $D_{s,t,d}$ is the forecast of the users demand. We use the prediction obtained by the Benchmark model in Section 5.7;
- t_{SDR} and t_{SDR} are the starting and ending time intervals in which the vehicle operates during the day (see Section 4.1.9). It has been chosen a set of time intervals between 7:00 and 21:00, based on the historical users trips activity during the day. Hence, $t_{SDR} = 84$ (the time interval that starts at 07:00 in the morning) and $t_{EDR} = 252$ (the time interval that finish at 21:00 in the evening).

Chapter 7

Case study: results and discussion

In this chapter, we see the information that is possible to extract in the simulations, we analyse the performances of the scenarios and we analyse the tour of the service vehicle during the night and daily re-balancing.

7.1 Output data from simulation

The scenarios have been simulated using the resources introduced in Section 5.1 and the historical data on operations available from the analysed time period (year 2018). The output data are recorded for the analysis and comparison of the scenarios.

For each scenario, these output data are gathered for each time interval of 5 minutes, for all the time period of simulation (year 2018). Hence, the output data, for each time interval and station, are:

- the station state;
- the number of empty, full and total violations;
- the vehicle operations.

Then, from the output data above, it is possible to extract other useful information, such as:

Violations

- the average number of empty, full and total violations per day;

- the average number of empty, full and total violations per day grouped by station, by month, by weekday or by time interval;
- the percent decrease of the number of violations with respect to Scenario 1. We recall that this configuration scenario is characterized by the current setting active in the Padua BSS, hence we have the standard stations size, the standard initial stations state and no rebalancing operations are done in the course of the time period of simulation. To this end, Scenario 1 is taken as base scenario.
- the percentage of violations on the total number of operations;
- the percentage of time in which a station had a shortage of bikes or docking stations;

Vehicle

- the average number of visits per day done by the service vehicle in the night and in the daily tour;
- the average number of operations per day done by the service vehicle in the night and in the daily tour;
- the average travel time spent per day by the service vehicle in the night and in the daily tour;
- the average distance travelled per day by the service vehicle in the night and in the daily tour.

In the next section, we will discuss the metrics listed above, for the case of Padua, and it may be possible to find some suggestions applicable to other BSSs, depending on the resources available.

7.2 Violations

In the year 2018, 113525 trips occurred in the Padua BSS, i.e. 311.03 trips a day, on average. As each trip is composed by one pick-up and one delivery operation, the

	Station		Route		Violations			
Scenario	Size	State	Night	Day	Empty	Full	Total	Total %
1	Standard	Standard	-	-	26.28	26.33	52.61	0.00 %
2		Standard	-	Yes	9.81	9.88	19.70	-62.55 %
3		Standard	Yes	-	22.21	23.11	45.32	-13.86 %
4		Standard	Yes	Yes	7.48	8.52	16.00	-69.59 %
5		Optimal	Yes	-	22.29	25.41	47.70	-9.33 %
6		Optimal	Yes	Yes	8.15	8.29	16.44	-68.75 %
7	Optimal	Standard	-	-	14.50	14.49	28.99	-44.90 %
8		Standard	-	Yes	5.64	5.78	11.42	-78.29 %
9		Standard	Yes	-	13.14	7.27	20.41	-61.21 %
10		Standard	Yes	Yes	5.87	4.20	10.07	-80.86 %
11		Optimal	Yes	-	13.42	8.69	22.11	-57.97 %
12		Optimal	Yes	Yes	6.92	4.37	11.29	-78.54 %
13	Utopic	Standard	-	-	10.65	10.61	21.26	-59.59 %
14		Standard	Yes	-	3.93	0.82	4.75	-90.97 %
15		Optimal	Yes	-	3.78	1.10	4.88	-90.72 %

Table 7.1: The average number of empty, full, total violations per day and the percent decrease of the number of total violations with respect to Scenario 1. The symbol - means that no re-balancing operations have been done.

number of total operations occurred is 227050, i.e. 622.06 operations per day, on average. We can take this metric as a base for the following analysis. In this section, we analyse the average number of violations per day (see Section 1.9), a first key performance indicator of the quality of a BSS: the lower the number of violations, the higher the probability for a user to find bikes and docking stations.

7.2.1 Violations per day

Table 7.1 shows the average number of empty, full and total violations per day in the 15 scenarios, and the percent decrease of the number of violations with respect to Scenario 1, taken as benchmark. In Figure 7.1, we can see the boxplots of the number of violations per day in 2018.

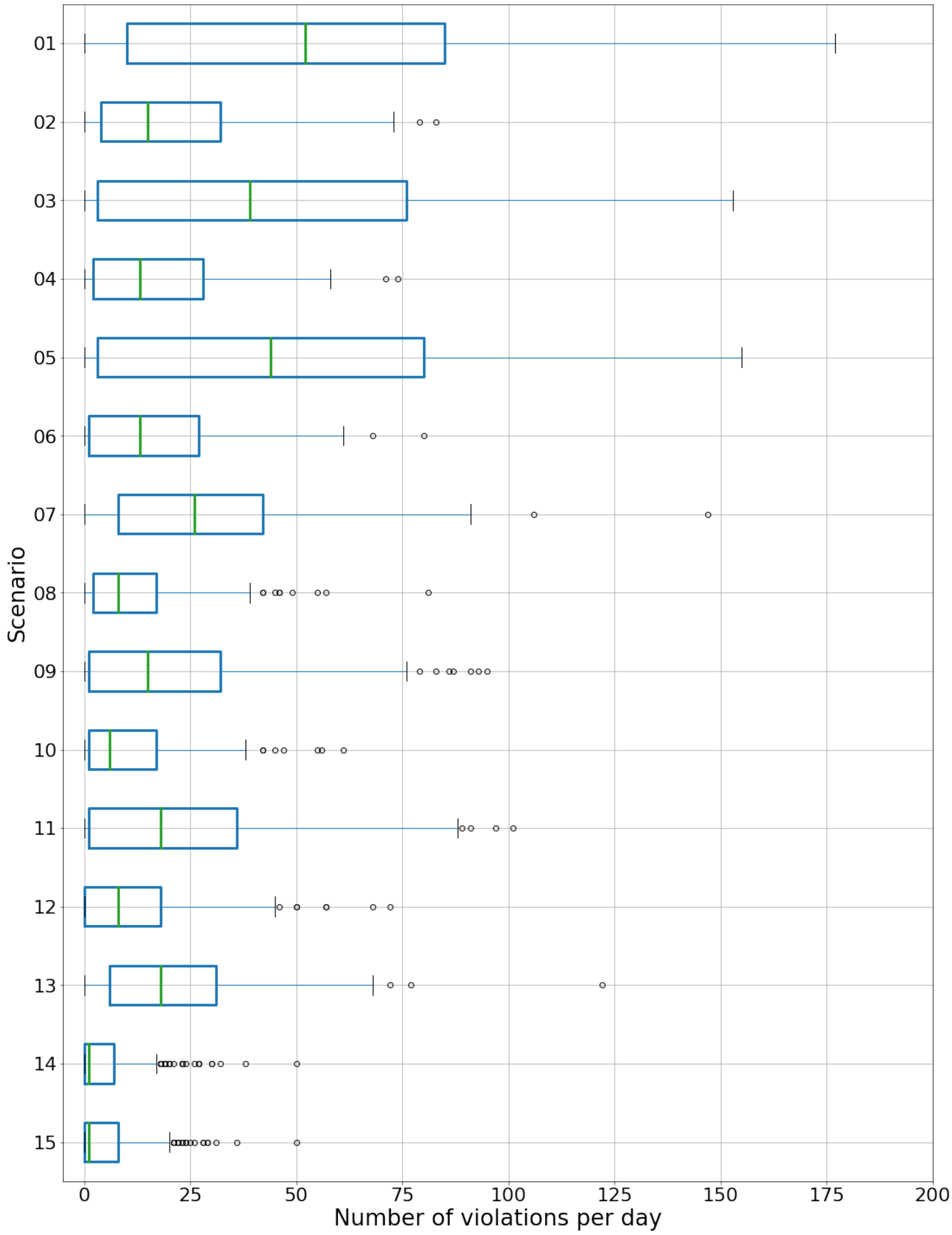

Figure 7.1: The distribution of the number of violations per day in the different scenarios.

Standard stations size

In Scenario 1, the BSS has the standard stations size configuration, the standard stations state and no re-balancing operations are done. In this scenario, we have an average of 52.61 violations per day, the worst performance among all the scenarios, that can be broken down into 26.28 empty violations and 26.33 full violations. If we compare the number of violations per day with the number of operations per day calculated above, we obtain the percentage of violations per operation, that is 8.46% for Scenario 1 (52.61/622.06).

In Scenario 2, we add the daily re-balancing with respect to Scenario 1, and the average number of violations per day decreases to 19.70 (-62.55% of violations with respect to Scenario 1). This result is crucial, as it suggests that for the current configuration of the Padua BSS, a service vehicle active during the day is fundamental to decrease the disservices. Figure 7.2 shows the average number of violations in Scenario 1 and Scenario 2 for each time interval of the day.

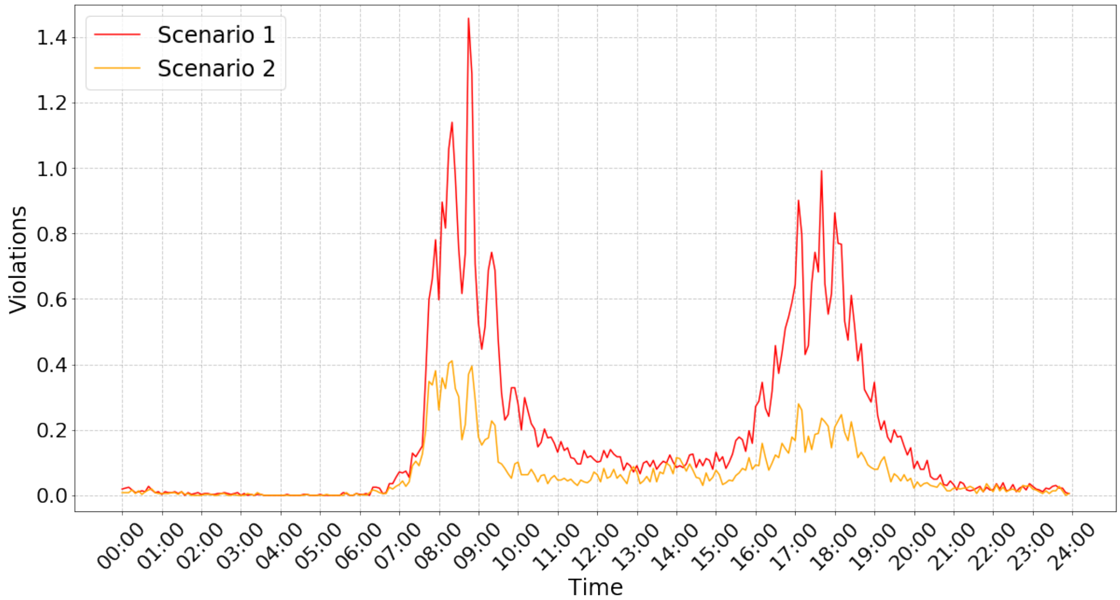

Figure 7.2: The average number of violations per day in Scenario 1 and Scenario 2 for each time interval of the day.

In Scenario 3 and Scenario 5, we add the night re-balancing with respect to Scenario 1, and we re-balance the system in the night using the standard and the optimal stations state, respectively. In these scenarios the number of violations per day decreases to 45.32 (-13.86%) and 47.70 (-9.33%), on average. This result suggest that, with the standard stations size configuration, the night re-balancing can help to re-balance the system but it has a lower impact than the daily re-balancing (-56.53% and -58.70% of violations in Scenario 2 with respect to Scenario 3 and Scenario 5, respectively). We recall that the night re-balancing consists of one vehicle that does only one visit per station, while the daily re-balancing consists of one vehicle active during 14 hours to refill the stations, so a comparison between these two services is not completely fair. Furthermore, in Scenario 5, that has an optimal stations state, we have 5.25% more violations per day than Scenario 3 that has a standard stations state. In particular, the number of full violations in Scenario 5 is 9.95%, hence greater than in Scenario 3, while the number of empty violations in Scenario 5 is only 0.36% greater than in Scenario 3. In fact, in Scenario 5 there are 14 bikes more than in Scenario 3, that lead to an increase of the shortage of docking stations. Another reason of the better performances of Scenario 3 can be related to the fact that the optimal stations state, calculated using the historical data for the period 2014-2017, may not correctly represent the behaviour of the users demand for 2018. In future works, we will better focus on the forecast of the users demand. Figure 7.3 shows the average number of violations in Scenario 3 and Scenario 5 for each time interval of the day.

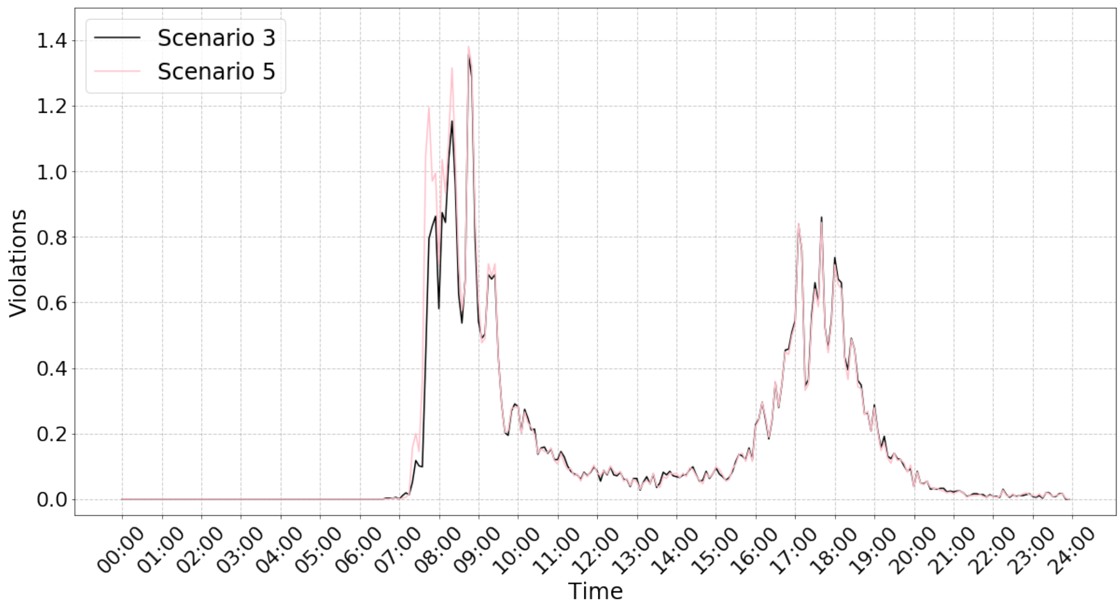

Figure 7.3: The average number of violations per day in Scenario 3 and Scenario 5 for each time interval of the day.

In Scenario 4 and Scenario 6, we add the night and the daily re-balancing with respect to Scenario 1 and the performances are very satisfying, as expected. In particular, for Scenario 4 the number of violations decreases to 16.00 (-69.59%), while for Scenario 6 it decreases to 16.44 (-68.75%). As in the previous comparison, the standard stations state performs slightly better than the optimal stations state.

In fact, in Scenario 6 we have 2.75% more violations per day with respect to Scenario 4. Figure 7.4 shows the average number of violations in Scenario 4 and Scenario 6 for each time interval of the day.

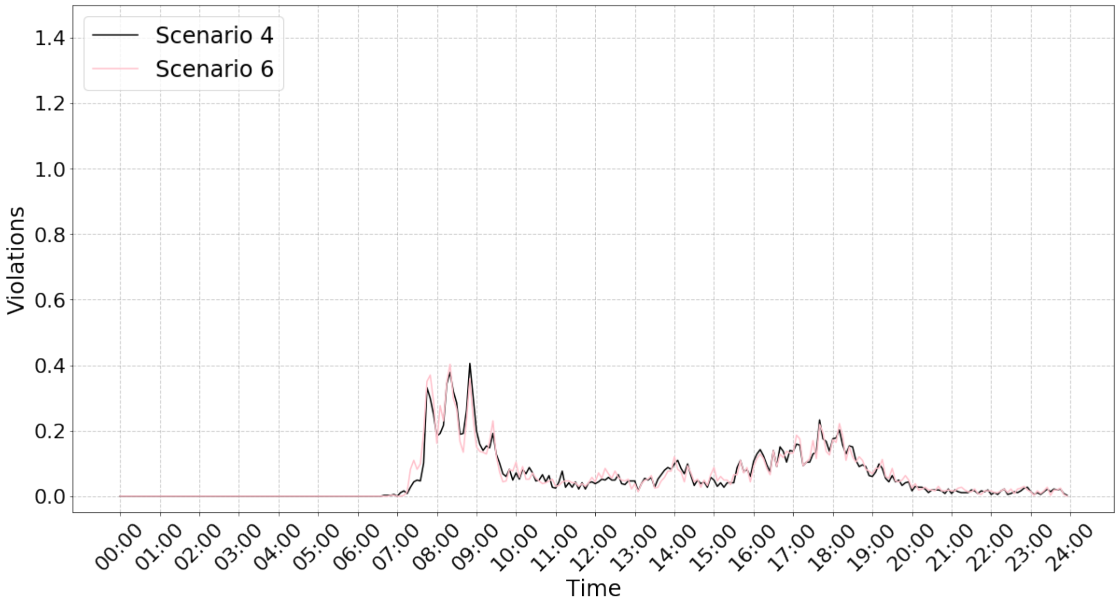

Figure 7.4: The average number of violations per day in Scenario 4 and Scenario 6 for each time interval of the day.

Optimal stations size

In Scenario 7, we change the BSS structure using the optimal stations size. We recall that Scenario 7 has the same number of docking stations and bikes available with respect to Scenario 1, but in this case the average number of violations decreases to 28.99 (-44.90%). The difference of performance between these two scenarios is large, so the application of the optimization model introduced in Section 3.2 to resize the BSS, original contribution of this thesis, may have a sensible impact on the decreasing of the violations.

In Scenario 8, we add the daily re-balancing with respect to Scenario 7 and the number of violations decreases to 11.42 (-78.29% with respect to Scenario 1, -60.61% with respect to Scenario 7). From this performance indicator, we can notice that the daily re-balancing has a similar impact both in the configuration with standard stations size, where in Scenario 2 we decrease the violations by 62.55% with respect to Scenario 1, and in the configuration with optimal stations size, where in Scenario 8 we decrease the violations by 60.61% with respect to Scenario 7. Figure 7.5 shows the average number of violations in Scenario 7 and Scenario 8 for each time interval of the day.

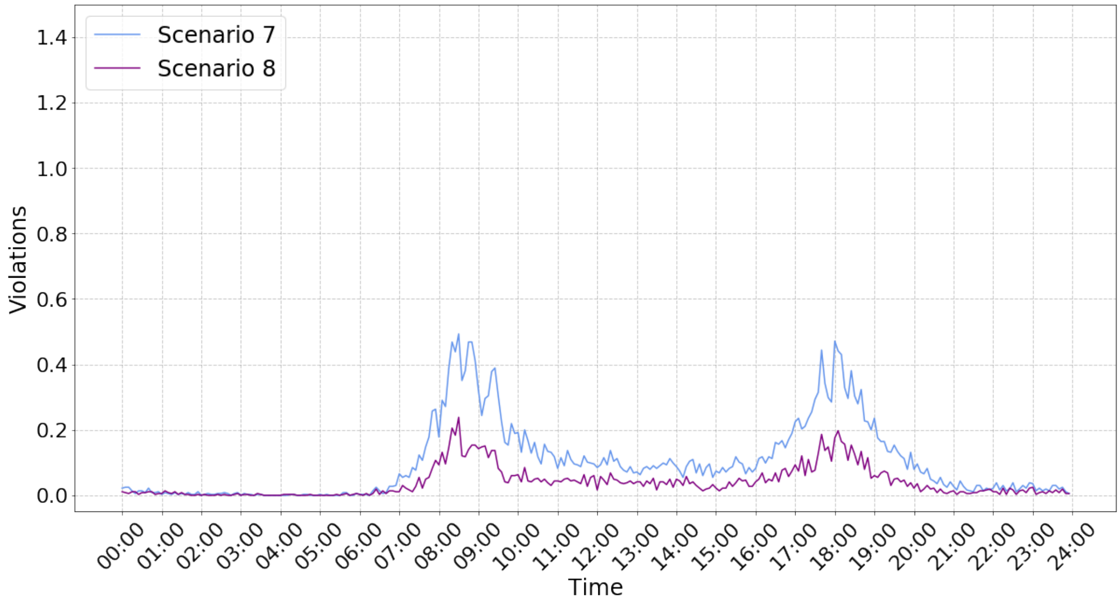

Figure 7.5: The average number of violations per day in Scenario 7 and Scenario 8 for each time interval of the day.

In Scenario 9, we add the night re-balancing with respect to Scenario 7 and the number of violations decreases to 20.41 (-61.21% with respect to Scenario 1). The night re-balancing, in the optimal stations size configuration, has a greater impact on the number of violations of Scenario 7 (-29.60%) with respect to the impact of Scenario 3 on Scenario 1 (-13.86%). This result suggests that, if the stations size is more adapt to match the users demand, the night re-balancing can play a greater role of the BSS. Figure 7.6 shows the average number of violations in Scenario 7 and Scenario 9 for each time interval of the day.

Finally, the comparison between Scenario 9 and Scenario 11, and between Scenario 10 and 12, confirms the better performance of the standard stations state with respect to the optimal stations state.

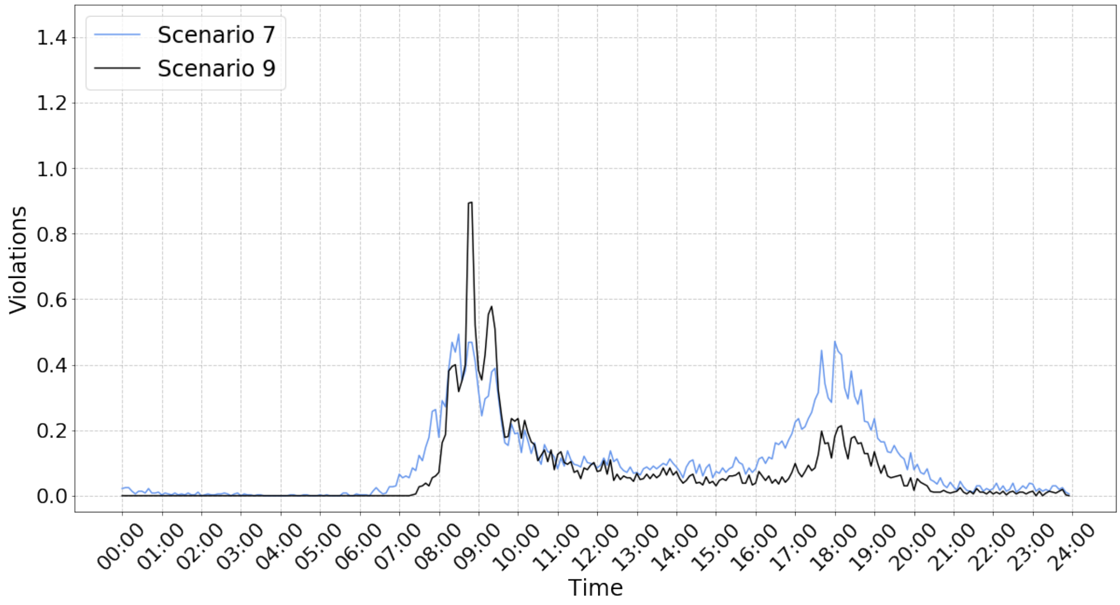

Figure 7.6: The average number of violations per day in Scenario 7 and Scenario 9 for each time interval of the day.

Utopic stations size

In Scenario 13, we change the BSS structure and we use the utopic stations size. We recall that Scenario 13 has 193 more docking stations (+55.14%) and 93 more bikes (+55.36%) than Scenario 1 and Scenario 7. The great investment to enlarge the stations shows its benefits, in fact, the average number of violations per day in Scenario 13 is 21.26 (-59.59% with respect to Scenario 1, -26.66% with respect to Scenario 7). We can notice that the performances of Scenario 13 are only slightly worst than the ones in Scenario 2, so the service provider could consider to oversize the stations, if, in the long term, the cost of new bikes and docking stations is lower than the cost of the daily re-balancing. Figure 7.7 shows the average number of violations in Scenario 1, Scenario 7 and Scenario 13 for each time interval of the day.

Finally, in Scenario 14 and 15, we add the night re-balancing with respect to Scenario 13 and we obtain the best performance over all the scenarios. In Scenario 14, the average number of violations per day is 4.75 (-90.97%) and in Scenario 15 they are 4.88 (-90.72%). In this case, the night re-balancing crucially impacts the BSS, as the number of violations in Scenario 14 decreases by -77.66% with respect

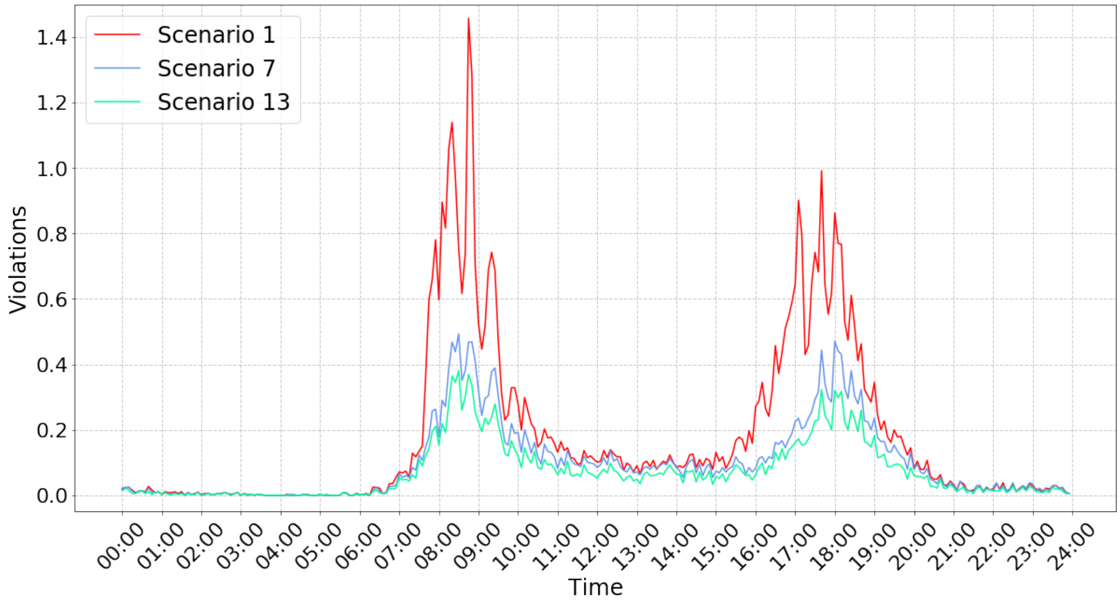

Figure 7.7: The average number of violations per day in Scenario 1, Scenario 7 and Scenario 13 for each time interval of the day.

to Scenario 13, and the number of violations in Scenario 15 decreases by -77.05% with respect to Scenario 13. Figure 7.8 shows the average number of violations in Scenario 13 and Scenario 14 for each time interval of the day.

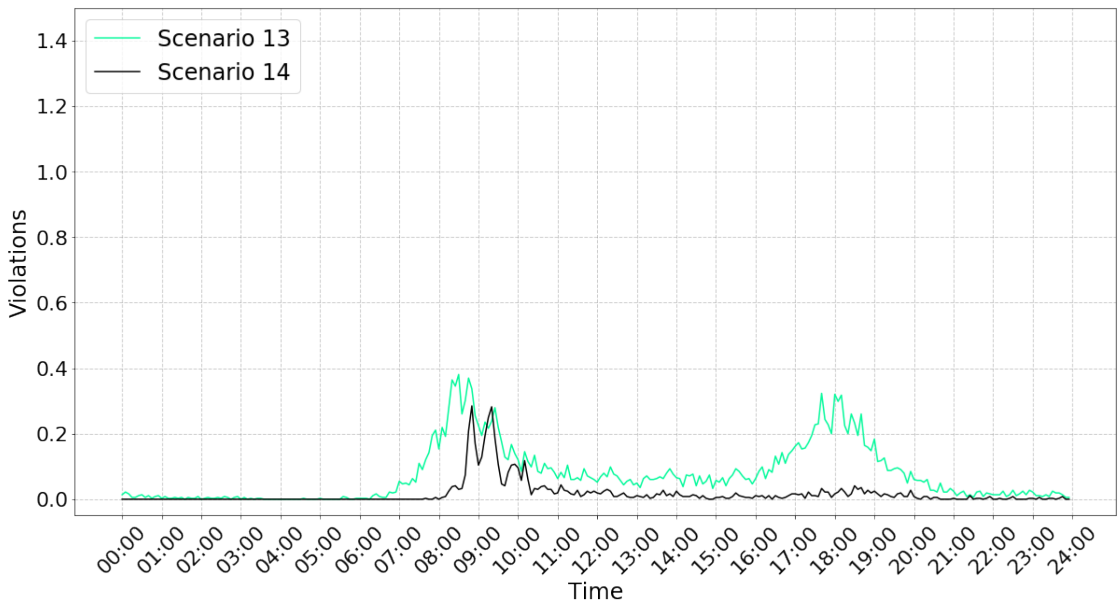

Figure 7.8: The average number of violations per day in Scenario 13 and Scenario 14 for each time interval of the day.

7.2.2 Violations per day grouped by working days and holidays

In the course of the thesis, we underlined the differences between working days and holidays. Indeed, in the case study of Padua, the users demand of 2018 drops dramatically from 802.46 operations per day in the working days, to 214.52 operations per day in the holidays. So, in the working days there are 274% more trips than in the holidays. Table 7.2 and Table 7.3 show the number of violations per day in the working days and holidays, respectively.

In Scenario 1, we can notice that the average number of violations per day is 72.83 in the working days and 6.94 in the holidays. As just mentioned above, the users demand is greater in the working days (274% more trips), but the number of violations in the working days is much larger (949% more violations). This result suggests that, as the users demand increases, the unbalance of the BSS increases more than proportionally.

Working days

If we focus on the working days, from Monday to Friday, we can notice that, the average number of violations is slightly greater than the aggregate results of Table 7.1, but the impact of the specific models remains similar to the previous results.

Also, Table 7.2 shows clearly the better performance of the standard stations state with respect to the optimal stations state, as for example in Scenario 3 (64.54 violations, standard stations state) and Scenario 5 (68.06 violations, optimal stations state). In Figure 7.9 we can see the boxplots of the number of violations per day in the working days of 2018.

Holidays

During the holidays (weekends and national holidays) the number of violations is sensibly lower than during the working days. Unexpectedly, in Table 7.3 we can notice that the night re-balancing has a greater impact on the number of violations with respect to the daily re-balancing, and in Figure 7.10 we can see the boxplots

	Station		**Route**		**Violations**			
Scenario	Size	State	Night	Day	Empty	Full	Total	Total %
1	Standard	Standard	-	-	36.23	36.6	72.83	0.00 %
2		Standard	-	Yes	13.60	13.62	27.23	-62.61 %
3		Standard	Yes	-	31.41	33.13	64.54	-11.38 %
4		Standard	Yes	Yes	10.58	11.97	22.55	-69.04 %
5		Optimal	Yes	-	31.89	36.18	68.06	-6.55 %
6		Optimal	Yes	Yes	11.58	11.69	23.28	-68.04 %
7	Optimal	Standard	-	-	19.46	19.70	39.15	-46.24 %
8		Standard	-	Yes	7.62	7.86	15.48	-78.75 %
9		Standard	Yes	-	18.68	10.43	29.11	-60.03 %
10		Standard	Yes	Yes	8.34	5.90	14.23	-80.46 %
11		Optimal	Yes	-	19.21	12.39	31.60	-56.61 %
12		Optimal	Yes	Yes	9.87	6.22	16.09	-77.91 %
13	Utopic	Standard	-	-	14.24	14.43	28.67	-60.63 %
14		Standard	Yes	-	5.66	1.18	6.83	-90.62 %
15		Optimal	Yes	-	5.45	1.59	7.04	-90.33 %

Table 7.2: The average number of empty, full, total violations per working day and the percent decrease of the number of total violations with respect to Scenario 1. The symbol - means that no re-balancing operations have been done.

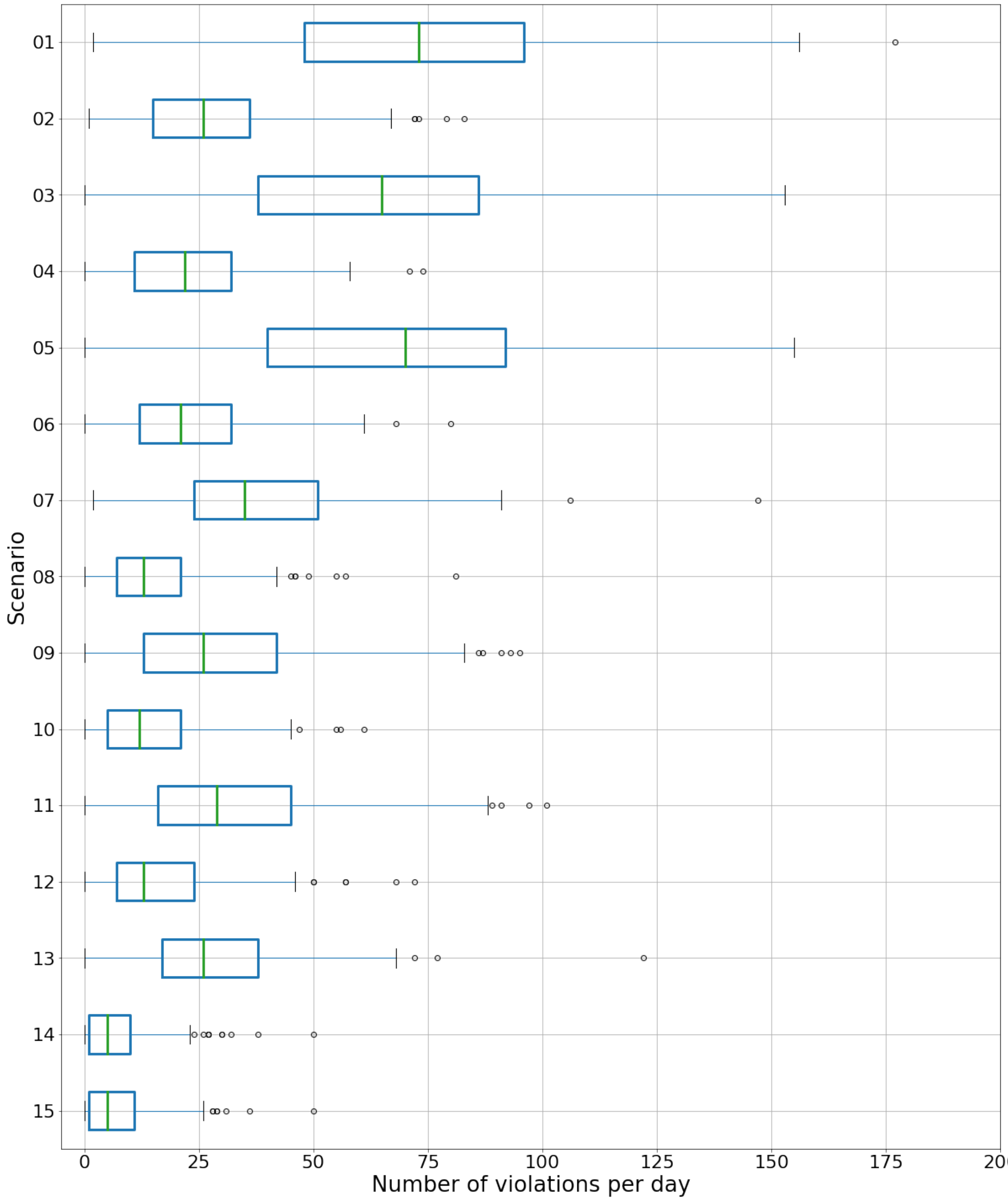

Figure 7.9: The distribution of the number of violations per workday in the different scenarios.

of the number of violations per day in the holidays of 2018.

In particular, in Scenario 2, we add the daily re-balancing to Scenario 1 decreasing the violations to 2.69 (-61.24% with respect to Scenario 1), while, in Scenario 3, we add the night re-balancing to Scenario 1 decreasing the violations to 1.92 (-72.33%). The worst performance of the daily rebalancing during the holidays may be caused by the unpredictability of users demand in this days, as the trips that are not related to the working or university schedule time. However, if we analyse more in details the results of Table 7.2 and Table 7.3, we can notice that the impact of the daily rebalancing, for example in Scenario 2, does not change between working days and holidays (-62.61% and -61.24%, respectively). The difference is on the night re-balancing that has a low impact on the working days but a huge impact on the holidays, as we can see in Scenario 3 (-11.38% and -72.33%, respectively). Hence, the better performance of Scenario 3 during the holidays, means that, once the stations are balanced during the night, it is difficult that they can have a shortage of bikes or docking stations during the weekends, as the users activity is sensibly lower than the working days.

Finally, utopic Scenario 14 and Scenario 15 confirm the best performances among all the scenarios with a low number of violations, 0.04 (-99.42%) and 0.01 (-99.86%), respectively. In particular, the addition of the night re-balancing impacts Scenario 13 by a decrease of 99.11% of the number of violations, for Scenario 14, and by a decrease of 99.78%, for Scenario 15.

In conclusion, Figure 7.11 shows the average number of violations per day in the working days and in the holidays, for each scenario, in addition to the total number of violations per day.

	Station		Route		Violations			
Scenario	Size	State	Night	Day	Empty	Full	Total	Total %
1	Standard	Standard	-	-	3.81	3.12	6.94	0.00 %
2		Standard	-	Yes	1.26	1.43	2.69	-61.24 %
3		Standard	Yes	-	1.45	0.47	1.92	-72.33 %
4		Standard	Yes	Yes	0.47	0.74	1.21	-82.56 %
5		Optimal	Yes	-	0.62	1.09	1.71	-75.36 %
6		Optimal	Yes	Yes	0.38	0.61	0.98	-85.88 %
7	Optimal	Standard	-	-	3.30	2.73	6.04	-12.97 %
8		Standard	-	Yes	1.17	1.08	2.25	-67.58 %
9		Standard	Yes	-	0.62	0.12	0.74	-89.34 %
10		Standard	Yes	Yes	0.30	0.36	0.66	-90.49 %
11		Optimal	Yes	-	0.32	0.33	0.65	-90.63 %
12		Optimal	Yes	Yes	0.27	0.19	0.46	-93.37 %
13	Utopic	Standard	-	-	2.54	1.97	4.51	-35.01 %
14		Standard	Yes	-	0.04	0.00	0.04	-99.42 %
15		Optimal	Yes	-	0.01	0.00	0.01	-99.86 %

Table 7.3: The average number of empty, full, total violations per holiday and the percent decrease of the number of total violations with respect to Scenario 1. The symbol - means that no re-balancing operations have been done.

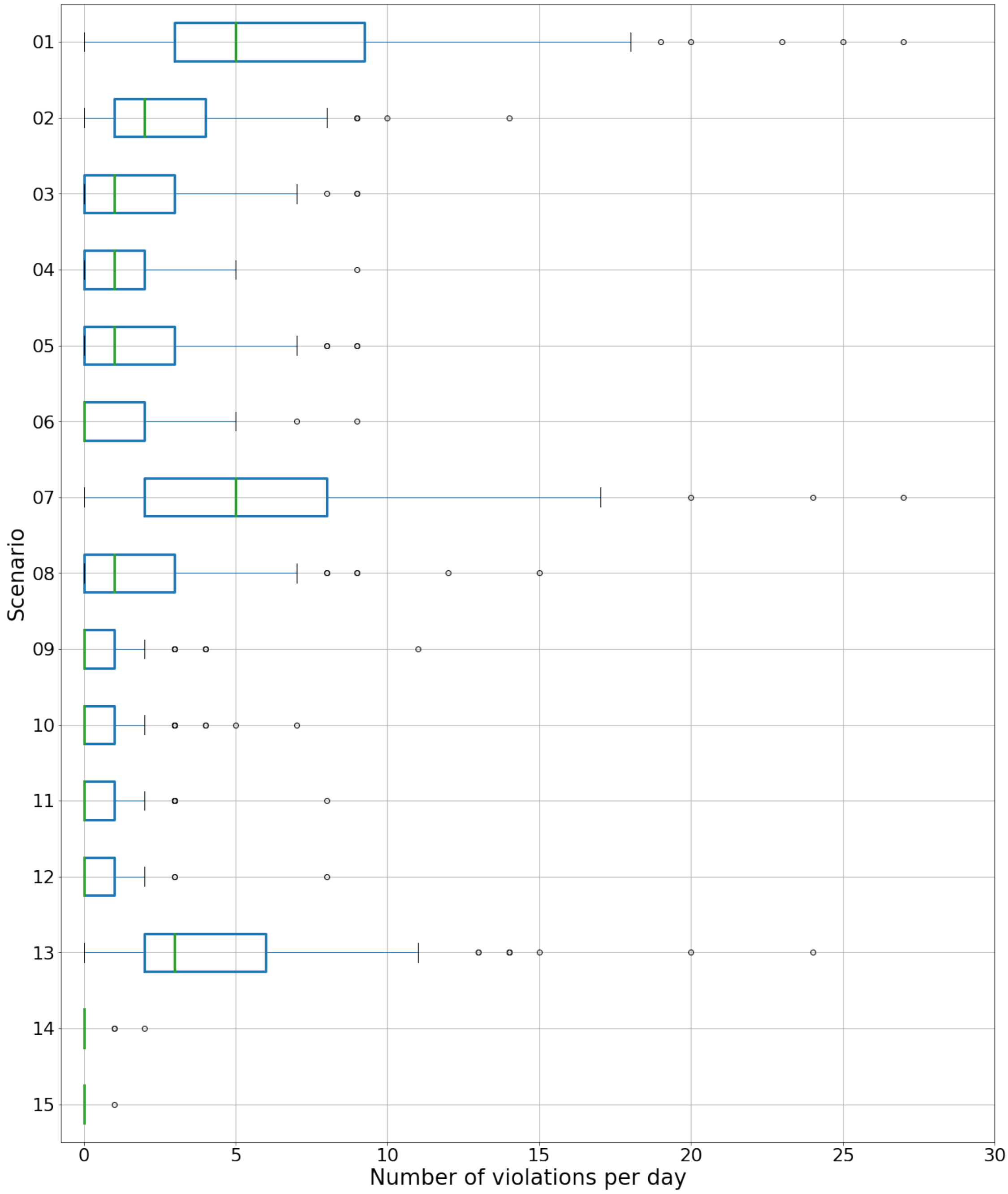

Figure 7.10: The distribution of the number of violations per holiday in the different scenarios.

7.2.3 Violations per day grouped by month

Another criterion that could be taken into consideration is the performance in the different months, to see if there are months in which the rebalancing operations may be more crucial. Table 7.4 shows the average number of violations per day grouped by month in the various scenario. In Scenario 1, we can notice that the average

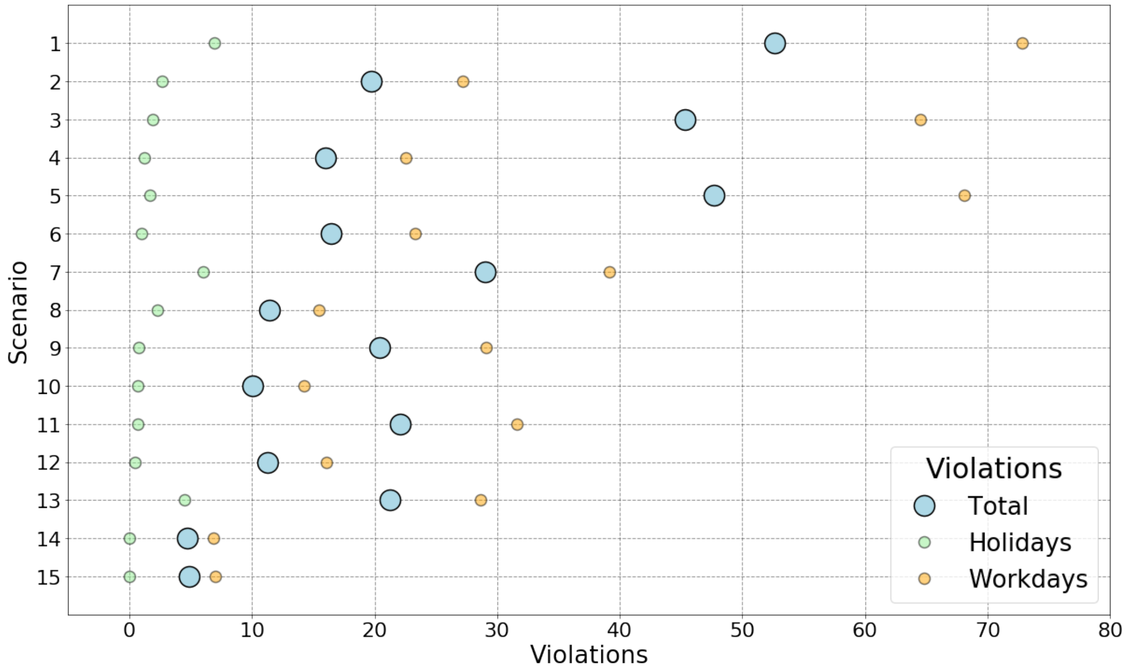

Figure 7.11: The average number of total violations per day. In addition, for each scenario, the number of violations in the working days and in the holidays is represented.

number of violations per day is high in March (73.26, +39% than the average), April (65.87, +25%) and May (67.52, +28%), while the number of violations is sensibly lower in August (27.97, -47%), as the users demand is also low. Figure 7.12 shows the average number of violations in Scenario 1, Scenario 2 and Scenario 3 for each month of the year. In this figure, we clearly see the larger benefits of the daily re-balancing with respect to the night re-balancing. Also, Figure 7.13 shows the average number of violations in Scenario 1, Scenario 7 and Scenario 13 for each month of the year. This figure shows the benefits of the stations resizing in the three different configurations: standard, optimal and utopic stations size.

From this analysis, the service provider may organize the resources to efficiently manage the BSS problems. For example, it is possible to choose the more suitable months to reduce or to increase the fleet of rebalancing vehicles. Table 7.5 shows the percent decrease of the number of total violations per day with respect to Scenario 1.

	Station		Route		Month											
Scenario	Size	State	Night	Day	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	Standard	Standard	-	-	56.00	49.18	73.26	65.87	67.52	41.87	43.55	27.97	44.50	56.61	57.77	46.87
2		Standard	-	Yes	19.71	15.86	26.52	25.90	29.48	18.33	14.16	7.48	17.10	26.03	20.80	14.71
3		Standard	Yes	-	50.16	39.75	68.71	61.80	60.23	36.20	33.48	18.65	36.17	51.58	48.87	37.81
4		Standard	Yes	Yes	17.81	12.54	24.10	21.27	23.23	12.83	10.71	5.16	12.53	21.19	17.00	13.29
5		Optimal	Yes	-	52.84	42.82	71.32	63.50	62.52	38.37	35.84	21.26	38.60	53.65	51.07	40.23
6		Optimal	Yes	Yes	18.23	13.14	24.13	22.97	24.03	14.10	10.16	5.55	12.63	22.74	16.70	12.55
7	Optimal	Standard	-	-	28.45	24.43	40.42	34.63	41.19	19.97	26.97	18.13	23.80	33.35	30.77	25.13
8		Standard	-	Yes	9.90	7.04	16.52	17.23	20.77	9.13	7.10	3.13	10.03	14.94	13.20	7.77
9		Standard	Yes	-	22.06	13.75	35.68	31.27	32.84	13.33	12.65	4.71	13.57	27.35	20.50	16.42
10		Standard	Yes	Yes	10.26	5.64	16.48	16.60	17.26	7.03	4.84	2.06	7.30	14.74	10.63	7.58
11		Optimal	Yes	-	23.81	15.43	37.65	33.53	33.52	15.6	14.90	6.13	14.77	28.45	22.77	18.03
12		Optimal	Yes	Yes	11.23	7.50	20.48	17.77	18.74	8.70	5.48	2.77	7.80	15.58	10.53	8.52
13	Utopic	Standard	-	-	20.29	17.50	29.81	23.77	27.48	12.07	22.45	15.39	16.70	25.26	24.30	19.45
14		Standard	Yes	-	5.74	3.18	10.39	6.67	7.06	2.23	2.29	0.26	2.67	7.65	4.43	4.16
15		Optimal	Yes	-	5.90	3.39	10.29	7.10	7.10	2.43	2.61	0.29	3.03	7.42	4.33	4.48

Table 7.4: The average number of total violations per day grouped by month for the different scenarios.

	Station		**Route**		**Month**											
Scenario	Size	State	Night	Day	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	Standard	Standard	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2		Standard	-	Yes	-64.80	-67.75	-63.80	-60.68	-56.34	-56.22	-67.49	-73.26	-61.57	-54.02	-64.00	-68.62
3		Standard	Yes	-	-10.43	-19.17	-6.21	-6.18	-10.80	-13.54	-23.12	-33.32	-18.72	-8.89	-15.41	-19.33
4		Standard	Yes	Yes	-68.20	-74.50	-67.10	-67.71	-65.60	-69.36	-75.41	-81.55	-71.84	-62.57	-70.57	-71.64
5		Optimal	Yes	-	-5.64	-12.93	-2.65	-3.60	-7.41	-8.36	-17.70	-23.99	-13.26	-5.23	-11.60	-14.17
6		Optimal	Yes	Yes	-67.45	-73.28	-67.06	-65.13	-64.41	-66.32	-76.67	-80.16	-71.62	-59.83	-71.09	-73.22
7	Optimal	Standard	-	-	-49.20	-50.33	-44.83	-47.43	-39.00	-52.30	-38.07	-35.18	-46.52	-41.09	-46.74	-46.38
8		Standard	-	Yes	-82.32	-85.69	-77.45	-73.84	-69.24	-78.19	-83.70	-88.81	-77.46	-73.61	-77.15	-83.42
9		Standard	Yes	-	-60.61	-72.04	-51.30	-52.53	-51.36	-68.16	-70.95	-83.16	-69.51	-51.69	-64.51	-64.97
10		Standard	Yes	Yes	-81.68	-88.53	-77.50	-74.80	-74.44	-83.21	-88.89	-92.63	-83.60	-73.96	-81.60	-83.83
11		Optimal	Yes	-	-57.48	-68.63	-48.61	-49.10	-50.36	-62.74	-65.79	-78.08	-66.81	-49.74	-60.59	-61.53
12		Optimal	Yes	Yes	-79.95	-84.75	-72.04	-73.02	-72.25	-79.22	-87.42	-90.10	-82.47	-72.48	-81.77	-81.82
13	Utopic	Standard	-	-	-63.77	-64.42	-59.31	-63.91	-59.30	-71.17	-48.45	-44.98	-62.47	-55.38	-57.94	-58.5
14		Standard	Yes	-	-89.75	-93.53	-85.82	-89.87	-89.54	-94.67	-94.74	-99.07	-94.00	-86.49	-92.33	-91.12
15		Optimal	Yes	-	-89.46	-93.11	-85.95	-89.22	-89.48	-94.20	-94.01	-98.96	-93.19	-86.89	-92.50	-90.44

Table 7.5: The percent decrease of the number of total violations per day grouped by month with respect to Scenario 1.

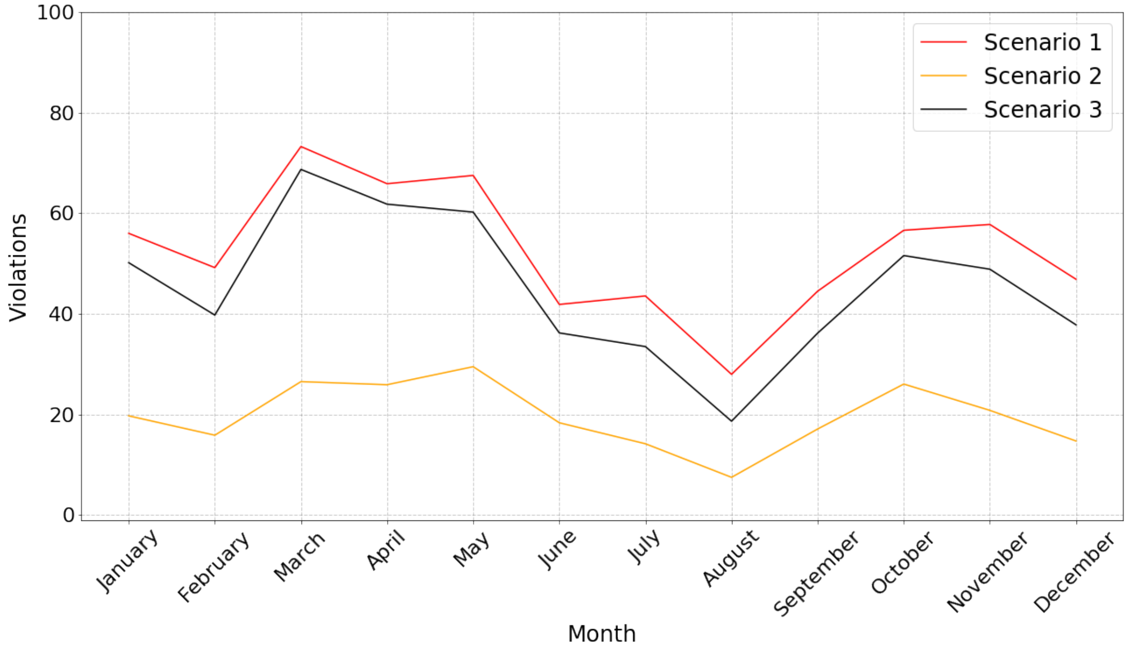

Figure 7.12: The average number of violations in Scenario 1, Scenario 2 and Scenario 3 for each month of the year.

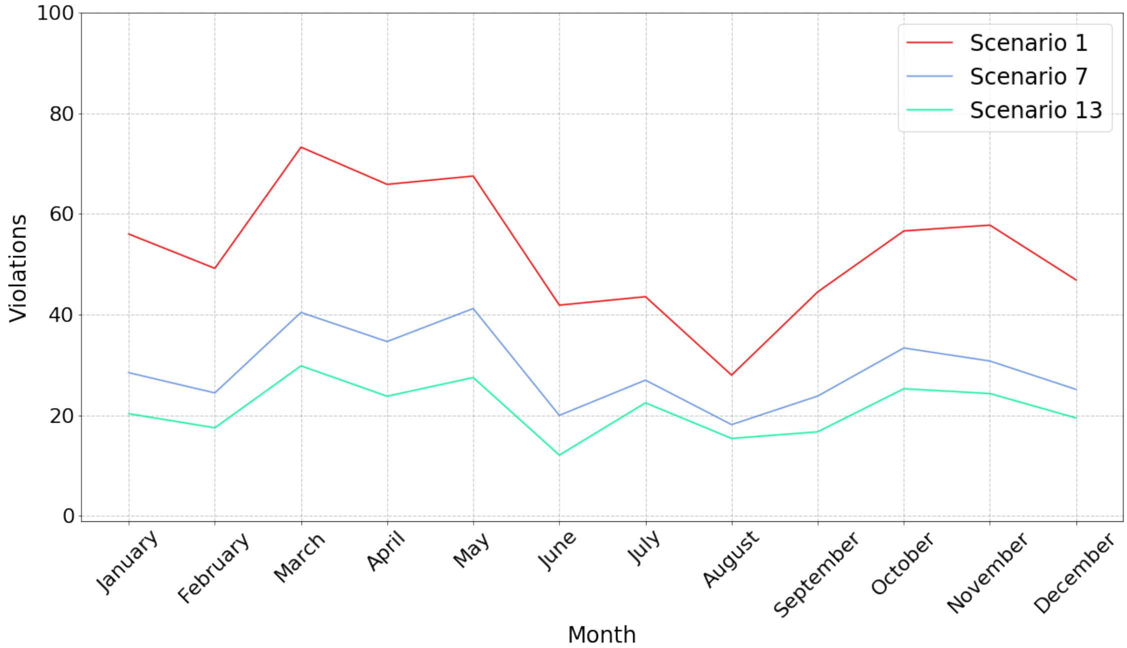

Figure 7.13: The average number of violations in Scenario 1, Scenario 7 and Scenario 13 for each month of the year.

7.2.4 Violations per day grouped by time interval

Another analysis that can give some insights to the service provider, is the number of violations per day in the different time intervals of the day, to see if there are hours in which it is necessary to give more attention to the re-balancing operations. Table 7.6 shows the average number of violations per day grouped by time intervals of two hours in the different scenarios. Table 7.7 shows the percent decrease of the number of total violations per day grouped by time intervals of two hours with respect to Scenario 1. Also, we recall Figure 7.2, that shows the number of violations in Scenario 1 and Scenario 2 for each time interval of 5 minutes.

We recall, from the users demand analysis of Section 5.6, that the greater users activity is:

- in the morning between 7:30 and 9:30;
- in the afternoon between 12:00 and 14:30;
- in the evening between 16:30 and 19:00.

In Scenario 1, no action is taken to improve the service and we can notice that the intervals in which there are more problems are between 08:00 and 10:00, with an average of 16.57 violations per day (31.5% of the total number of violations per day), between 16:00 and 18:00 with an average of 12.75 violations per day (24.2%) and between 18:00 and 20:00 with an average of 8.65 (16.4%). In these three time intervals, that correspond to just 6 hours of the day, there are 72.2% of the total violations.

	Station		Route		Time											
Scenario	Size	State	Night	Day	00-02	02-04	04-06	06-08	08-10	10-12	12-14	14-16	16-18	18-20	20-22	22-24
1	Standard	Standard	-	-	0.27	0.07	0.03	3.42	16.57	3.96	2.45	2.86	12.75	8.65	1.05	0.52
2		Standard	-	Yes	0.18	0.04	0.02	1.95	5.44	1.30	1.50	1.65	3.83	2.81	0.59	0.39
3		Standard	Yes	-	0.00	0.00	0.00	3.35	15.97	3.42	1.58	2.11	10.91	7.05	0.67	0.26
4		Standard	Yes	Yes	0.00	0.00	0.00	1.20	4.75	1.18	1.23	1.46	3.16	2.28	0.41	0.33
5		Optimal	Yes	-	0.00	0.00	0.00	5.10	17.05	3.33	1.56	2.11	10.77	6.87	0.62	0.28
6		Optimal	Yes	Yes	0.00	0.00	0.00	1.59	4.51	1.19	1.25	1.53	3.11	2.42	0.47	0.37
7	Optimal	Standard	-	-	0.24	0.07	0.04	1.63	7.79	2.96	2.18	1.98	4.97	5.64	0.95	0.54
8		Standard	-	Yes	0.16	0.03	0.03	0.58	2.96	1.14	0.96	0.78	2.00	2.16	0.34	0.28
9		Standard	Yes	-	0.00	0.00	0.00	0.25	9.17	2.97	1.58	1.19	1.94	2.70	0.41	0.21
10		Standard	Yes	Yes	0.00	0.00	0.00	0.14	3.60	1.13	0.98	0.72	1.35	1.70	0.26	0.18
11		Optimal	Yes	-	0.00	0.00	0.00	0.38	11.42	3.12	1.58	1.08	1.51	2.44	0.35	0.23
12		Optimal	Yes	Yes	0.00	0.00	0.00	0.25	4.54	1.16	0.94	0.73	1.49	1.71	0.29	0.18
13	Utopic	Standard	-	-	0.18	0.06	0.02	1.30	5.80	2.09	1.56	1.48	3.72	3.99	0.69	0.37
14		Standard	Yes	-	0.00	0.00	0.00	0.01	2.55	0.83	0.32	0.19	0.28	0.46	0.07	0.04
15		Optimal	Yes	-	0.00	0.00	0.00	0.01	2.87	0.89	0.33	0.17	0.20	0.35	0.05	0.03

Table 7.6: The average number of total violations per day grouped by time intervals of two hours for the different scenarios.

Scenario	**Station**		**Route**		**Time**											
	Size	State	Night	Day	00-02	02-04	04-06	06-08	08-10	10-12	12-14	14-16	16-18	18-20	20-22	22-24
1	Standard	Standard	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2		Standard	-	Yes	-33.33	-42.86	-33.33	-42.98	-67.17	-67.17	-38.78	-42.31	-69.96	-67.51	-43.81	-25.00
3		Standard	Yes	-	-100.00	-100.00	-100.00	-2.05	-3.62	-13.64	-35.51	-26.22	-14.43	-18.50	-36.19	-50.00
4		Standard	Yes	Yes	-100.00	-100.00	-100.00	-64.91	-71.33	-70.20	-49.8	-48.95	-75.22	-73.64	-60.95	-36.54
5		Optimal	Yes	-	-100.00	-100.00	-100.00	+49.12	+2.90	-15.91	-36.33	-26.22	-15.53	-20.58	-40.95	-46.15
6		Optimal	Yes	Yes	-100.00	-100.00	-100.00	-53.51	-72.78	-69.95	-48.98	-46.50	-75.61	-72.02	-55.24	-28.85
7	Optimal	Standard	-	-	-11.11	0.00	+33.33	-52.34	-52.99	-25.25	-11.02	-30.77	-61.02	-34.80	-9.52	+3.85
8		Standard	-	Yes	-40.74	-57.14	0.00	-83.04	-82.14	-71.21	-60.82	-72.73	-84.31	-75.03	-67.62	-46.15
9		Standard	Yes	-	-100.00	-100.00	-100.00	-92.69	-44.66	-25.00	-35.51	-58.39	-84.78	-68.79	-60.95	-59.62
10		Standard	Yes	Yes	-100.00	-100.00	-100.00	-95.91	-78.27	-71.46	-60.00	-74.83	-89.41	-80.35	-75.24	-65.38
11		Optimal	Yes	-	-100.00	-100.00	-100.00	-88.89	-31.08	-21.21	-35.51	-62.24	-88.16	-71.79	-66.67	-55.77
12		Optimal	Yes	Yes	-100.00	-100.00	-100.00	-92.69	-72.60	-70.71	-61.63	-74.48	-88.31	-80.23	-72.38	-65.38
13	Utopic	Standard	-	-	-33.33	-14.29	-33.33	-61.99	-65.00	-47.22	-36.33	-48.25	-70.82	-53.87	-34.29	-28.85
14		Standard	Yes	-	-100.00	-100.00	-100.00	-99.71	-84.61	-79.04	-86.94	-93.36	-97.80	-94.68	-93.33	-92.31
15		Optimal	Yes	-	-100.00	-100.00	-100.00	-99.71	-82.68	-77.53	-86.53	-94.06	-98.43	-95.95	-95.24	-94.23

Table 7.7: The percent decrease of the number of total violations per day grouped by time intervals of two hours with respect to Scenario 1.

7.2.5 Violations per day grouped by station

A detailed analysis of the performance of the different scenarios can be done on a station level. Table 7.8 shows the average number of violations per day grouped by station in the different scenarios.

In Scenario 1, we can notice that the the station with more violations per day is Station 25, the train station of the city, in which the average number of violations is 24.66 per day, i.e. 46.87% of the violations. Also, we can notice a great number of violations in Station 5 (4.68 violations per day), Station 15 (2.42), Station 21 (2.26) and Station 22 (2.45). In Scenario 1, these five stations count for 69.32% of the total violations. In Scenario 2, we can notice that the daily re-balancing is a good service to reduce the violations of these stations, but there are still 6.30 violations per day in Station 25. On the other side, in Scenario 7, we can see that the oversize of Station 25 shows its results, decreasing the violations to 5.59. We recall that the docking stations added to Station 25 are removed from other stations, in fact, in most of the other stations the number of violations slightly increase with respect to Scenario 1. Anyway, in Scenario 7, the choice of the optimal stations size shows an important decrease of the total number of violations (29.04, -44.80% with respect to Scenario 1). Finally, in Scenario 14, we have the utopic stations size, the night rebalancing and standard stations state, that lead to the best performance overall the scenarios. It is noteworthy to see that, despite the oversize of Station 25, from 14 to 56 docking stations, this station has most of the problems of this scenario, counting for 58.23% of the total violations. Figure 7.14 shows the average number of violations per day for each station in the different scenarios.

In conclusion, the results of Table 7.8 show the crucial role of the station size and suggest to focus more on the stations in which the users activity is particularly high, as these stations causes most of the problems in the BSS.

	Scenario														
Size	Standard						Optimal						Utopic		
State	Std				Opt		Std				Opt		Std		Opt
Night	-	-	Yes	Yes	Yes	Yes	-	-	Yes	Yes	Yes	Yes	-	Yes	Yes
Daily	-	Yes	-	Yes	-	Yes	-	Yes	-	Yes	-	Yes	-	-	-
Station	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.90	0.59	0.57	0.44	0.57	0.49	1.28	0.59	1.34	0.69	1.34	0.77	0.82	0.23	0.23
2	0.59	0.28	0.25	0.20	0.16	0.19	0.64	0.30	0.38	0.19	0.26	0.18	0.47	0.07	0.04
3	1.20	0.21	0.11	0.10	0.04	0.08	1.25	0.36	0.43	0.24	0.26	0.22	1.21	0.11	0.06
4	0.65	0.27	0.12	0.21	0.17	0.25	0.75	0.34	0.26	0.22	0.26	0.23	0.46	0.01	0.01
5	4.68	1.98	4.78	1.63	5.65	1.74	1.60	0.75	1.01	0.53	1.49	0.74	1.05	0.08	0.14
6	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.11	0.12	0.13	0.14
7	0.83	0.37	0.19	0.14	0.42	0.30	0.87	0.33	0.30	0.25	0.43	0.24	0.67	0.01	0.06
8	1.29	0.88	0.63	0.53	0.65	0.53	0.88	0.42	0.13	0.13	0.19	0.16	0.70	0.02	0.02
9	0.56	0.32	0.27	0.20	0.27	0.18	0.42	0.18	0.12	0.07	0.14	0.11	0.23	0.02	0.02
10	1.80	1.10	1.32	0.85	1.30	0.85	1.39	0.64	0.70	0.57	0.70	0.55	1.10	0.08	0.08
11	0.51	0.27	0.06	0.03	0.08	0.07	0.99	0.51	0.59	0.33	0.70	0.51	0.51	0.06	0.08
12	0.38	0.15	0.09	0.08	0.09	0.12	0.57	0.23	0.45	0.24	0.45	0.23	0.31	0.03	0.01
13	0.84	0.46	0.44	0.23	0.37	0.30	0.95	0.47	0.54	0.26	0.53	0.36	0.58	0.07	0.02
14	0.91	0.28	0.26	0.11	0.12	0.13	1.00	0.33	0.43	0.17	0.22	0.13	0.81	0.07	0.04
15	2.42	1.76	2.32	1.35	2.21	1.35	1.02	0.51	0.34	0.23	0.32	0.19	0.73	0.06	0.07
16	1.96	1.05	1.35	0.76	1.55	0.89	1.78	0.97	1.04	0.77	1.18	0.82	1.25	0.25	0.27
17	0.14	0.02	0.00	0.00	0.00	0.00	0.34	0.12	0.19	0.12	0.19	0.12	0.21	0.01	0.01
18	0.50	0.12	0.03	0.02	0.03	0.02	0.68	0.24	0.23	0.17	0.23	0.19	0.52	0.05	0.05
19	0.22	0.04	0.00	0.00	0.00	0.00	0.31	0.07	0.30	0.12	0.29	0.15	0.27	0.02	0.02
20	0.16	0.07	0.12	0.08	0.05	0.05	0.43	0.21	0.99	0.47	0.58	0.38	0.18	0.26	0.12
21	2.26	1.19	2.81	1.09	3.33	1.24	1.30	0.51	1.28	0.72	1.66	0.88	0.88	0.12	0.20
22	2.45	0.61	1.74	0.67	1.00	0.47	2.35	0.61	1.30	0.59	0.72	0.40	2.21	0.17	0.08
23	0.76	0.22	0.06	0.04	0.04	0.04	0.88	0.32	0.31	0.18	0.26	0.15	0.74	0.01	0.01
24	0.67	0.40	0.19	0.26	0.15	0.26	0.81	0.44	0.37	0.29	0.37	0.31	0.52	0.04	0.03
25	24.66	6.30	25.46	5.61	28.13	6.15	5.59	1.67	7.00	2.13	8.97	2.89	4.07	2.83	3.17
26	0.02	0.00	0.00	0.00	0.00	0.00	0.06	0.11	0.04	0.10	0.04	0.10	0.02	0.00	0.00
27	0.70	0.46	0.90	0.79	0.64	0.41	0.50	0.15	0.09	0.10	0.09	0.13	0.48	0.00	0.00
28	0.55	0.31	1.23	0.58	0.68	0.32	0.34	0.07	0.22	0.21	0.22	0.14	0.26	0.05	0.01
Total	52.61	19.72	45.32	16.03	47.74	16.48	29.04	11.52	20.46	10.18	22.19	11.39	21.38	4.86	4.99
Mean	1.88	0.70	1.62	0.57	1.71	0.59	1.04	0.41	0.73	0.36	0.79	0.41	0.76	0.17	0.18

Table 7.8: The average number of total violations per day grouped by station for the different scenarios.

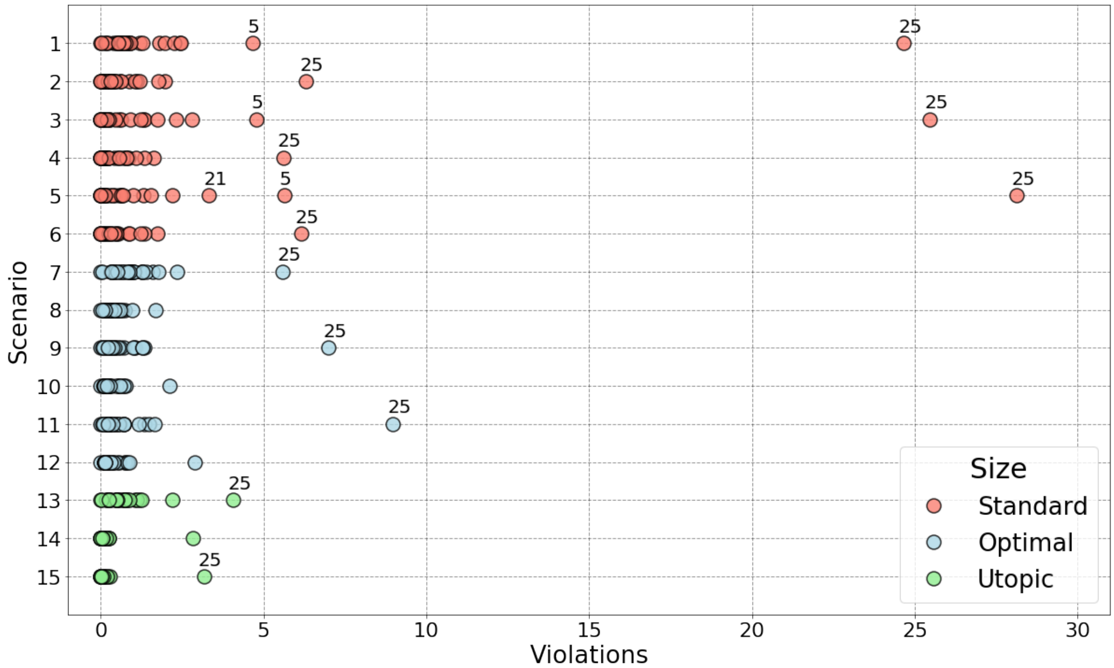

Figure 7.14: The average number of violations per day for each station in the different scenarios.

7.3 Vehicle

The last aspect that we take into account is the rebalancing cost, in terms of amount of time and kilometres spent by the service vehicle to re-balance the BSS. Also, we take into account the number of visits and operations done by the service vehicle in the stations. For this analysis, only the scenarios in which night or daily rebalancing is applied are taken into consideration. Hence, Scenario 1, 7 and 13 are excluded from the analysis of this section.

7.3.1 Night rebalancing costs

The night operations grouped by station

Table 7.9 shows the average number of operations per day grouped by station, for the night re-balancing operations done by the service vehicle in 10 scenarios. The table shows the average number of operations obtained as sum between pickup and delivery operations. We can notice that the more the size of the stations increases, the more the number of night operations increases. Indeed, the smallest number of operations occurs in Scenario 3, that has the standard stations size configuration. In this scenario the average number of operations is 51.48, that is 1.84 operations per station per day, on average. On the other side, in Scenario 14 and Scenario 15, that have the utopic stations size configuration, the average number of operations per day is 62.25 and 65.54, respectively, that are 2.22 and 2.34 operations per station per day, on average. The increase of the number of operations as we enlarged the problematic stations, is particularly relevant in Station 15 and Station 25, in which the number of docking stations increased by 160% and 300%, respectively (see Section 6.1). This result can be a contradiction, as an increase in the number of docking stations leads to an increase of the re-balancing costs, however the larger number of available docking stations allows the possibility that the station state, during the evening, is sensibly different from the station state required for the morning. Hence, the deviation between final station state and initial station state of the next day may be greater with the utopic configuration, leading to a greater number of night re-balancing operations.

The night route

As mentioned in Section 5.1, the night rebalancing model has not been implemented in the simulations, due to problems related to the feasibility of the solution route when the stations size increases. Nevertheless, we calculated the shortest route to visit all the stations by solving the well known Travelling Salesman Problem (TSP), and we obtained the results of Table 7.10, where we can see the driving distance and driving time to complete the night route. Hence, for all the scenarios with the night rebalancing, we assume that the service vehicle travels 41.15 km and it spend one hour and 39 minutes to complete the route. Also, for a costs analysis, the time spent to rebalance the stations should be assumed to be proportional to the number of bikes added to, or removed, from the station.

	Scenario									
Size	Standard				Optimal				Utopic	
State	Std		Opt		Std		Opt		Std	Opt
Night	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Daily	-	Yes	-	Yes	-	Yes	-	Yes	-	-
Station	3	4	5	6	9	10	11	12	14	15
1	1.85	2.38	1.85	2.45	1.76	1.68	1.76	1.76	1.90	1.97
2	1.62	1.90	2.02	2.30	1.53	1.55	1.60	1.60	1.78	1.93
3	1.99	2.06	2.51	2.77	1.72	1.59	1.87	1.89	1.99	2.07
4	2.10	2.37	2.60	2.79	1.98	1.85	1.98	2.06	2.22	2.22
5	2.64	2.99	3.22	2.79	3.23	3.53	3.48	3.77	3.55	3.62
6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7	1.94	2.16	2.12	2.42	1.86	1.95	1.90	2.21	2.08	2.62
8	2.72	2.55	2.72	2.63	3.06	3.01	3.05	3.06	3.16	3.16
9	1.96	1.89	1.98	1.90	2.09	2.02	2.11	2.05	2.19	2.20
10	2.92	2.80	3.04	2.93	3.27	3.34	3.37	3.38	3.58	3.67
11	2.15	2.17	2.15	2.22	1.91	1.79	2.00	1.99	2.15	2.15
12	1.54	1.66	1.66	1.81	1.48	1.55	1.48	1.56	1.58	1.74
13	1.69	2.28	1.75	1.99	1.64	1.83	1.76	1.90	1.96	2.02
14	2.08	2.05	2.14	2.14	1.94	1.93	2.04	2.04	2.25	2.21
15	2.07	2.10	2.11	2.20	3.19	3.10	3.13	3.00	3.45	3.45
16	2.65	2.80	2.71	2.84	2.67	2.67	2.82	2.91	3.14	3.05
17	0.82	0.83	1.01	1.01	0.80	0.78	0.97	0.91	0.82	1.00
18	2.10	2.12	2.09	2.09	2.04	2.12	2.04	2.11	2.09	2.09
19	0.50	0.52	1.31	1.31	0.49	0.80	0.74	0.84	0.48	0.48
20	1.10	1.23	1.16	1.21	1.27	1.75	1.06	1.20	1.10	1.09
21	2.51	2.31	2.93	2.47	2.60	2.73	2.84	3.07	2.54	2.61
22	2.28	2.50	2.66	2.92	2.46	2.55	2.74	2.81	3.13	3.59
23	2.16	2.22	2.22	2.28	1.93	1.90	2.03	2.04	2.21	2.23
24	1.89	2.37	1.91	2.21	1.76	1.86	1.76	1.81	2.00	2.12
25	3.38	3.18	5.10	3.91	8.72	8.69	10.03	9.55	8.19	9.12
26	0.06	0.06	0.35	0.35	0.05	0.04	0.05	0.03	0.06	0.35
27	1.47	1.43	1.11	0.98	1.36	1.42	1.36	1.34	1.40	1.49
28	1.29	1.35	1.03	0.84	1.11	1.36	1.11	1.15	1.25	1.29
Total	51.48	54.28	57.46	57.76	57.92	59.39	61.08	62.04	62.25	65.54
Mean	1.84	1.94	2.05	2.06	2.07	2.12	2.18	2.22	2.22	2.34

Table 7.9: The number of operations per day in each station in different scenarios during the night re-balancing.

From station		To station		Resources	
Number	Name	Number	Name	Distance	Time
0	Depot	16	Venezia 2	2.31 km	3m 29s
16	Venezia 2	22	Fiere	0.88 km	1m 37s
22	Fiere	5	Venezia Colombo	0.42 km	1m 39s
5	Venezia Colombo	24	Tribunale	1.12 km	3m 02s
24	Tribunale	25	Stazione 2	0.97 km	1m 48s
25	Stazione 2	1	Sarpi	0.54 km	2m 08s
1	Sarpi	2	Mazzini	0.35 km	0m 56s
2	Mazzini	3	Giotto	0.32 km	0m 48s
3	Giotto	15	Stazione	0.82 km	3m 21s
15	Stazione	12	Orsini	3.20 km	6m 47s
12	Orsini	17	I Colli	2.43 km	4m 33s
17	I Colli	9	Duomo	3.29 km	7m 03s
9	Duomo	18	Della Valle	1.86 km	7m 41s
18	Della Valle	19	Park Bembo Est	4.14 km	8m 56s
19	Park Bembo Est	20	Piovese	2.65 km	3m 44s
20	Piovese	27	Facciolati	1.98 km	4m 40s
27	Facciolati	14	Pontecorvo	0.74 km	2m 27s
14	Pontecorvo	11	Cesarotti	0.33 km	1m 06s
11	Cesarotti	23	Riviera Tito Livio	0.66 km	2m 56s
23	Riviera Tito Livio	10	Antenore	0.31 km	1m 26s
10	Antenore	8	Altinate	0.46 km	2m 14s
8	Altinate	21	Falloppio	0.75 km	3m 35s
21	Falloppio	7	Morgagni	0.58 km	1m 34s
7	Morgagni	4	Gasometro	0.17 km	0m 31s
4	Gasometro	6	Marzolo	1.26 km	3m 41s
6	Marzolo	13	Gattamelata	1.43 km	3m 27s
13	Gattamelata	28	Nazareth	0.41 km	1m 05s
28	Nazareth	26	La Fenice	3.39 km	6m 31s
26	La Fenice	0	Depot	3.39 km	6m 14s
Total				41.15 km	1h 39m

Table 7.10: The shortest route to visit all the stations in the BSS of Padova.

7.3.2 Daily rebalancing costs

The daily operations grouped by station

Table 7.11 shows the average number of operations per day grouped by station. In particular, in Table 7.11 we can see the average number of bikes added or removed in each station. For example, for Scenario 2 we can notice that the service vehicle added or removed in average 11.52 bikes in Station 5 and 30.93 bikes in Station 25. If we compare these results with Scenario 8, where the only difference is in the stations size, set to the optimal configuration, we can notice that the number of bikes moved strongly decrease to 3.57 for Station 5 (-69%) and 8.32 for Station 25 (-73.1%). The difference between these two scenarios is very important because, with the optimal stations size, we can decrease the daily rebalancing operations by 45.4% (from 111.98 to 61.14 operations per day). Furthermore, we can notice that Scenario 10 decreases the daily rebalancing operations by 45.1% with respect to Scenario 4. (from 96.81 to 53.11 operations per day) and Scenario 12 decreases the daily rebalancing operations by 43.7% with respect to Scenario 6 (from 99.04 to 55.75 operations per day). These results demonstrate the impact of the optimal stations size configuration, that not only sensibly decreases the number of violations per day, but also decreases the re-balancing costs.

The daily visits grouped by station

Another interesting metric is the average number of visits per station done by the service vehicle during the day. Table 7.12 shows the average number of visits per day grouped by station for the various scenarios. This metric is directly comparable with the night rebalancing visits, that we assumed to be equal to the number of stations, that is 28, or 1 per station. In this case, we can see that the average number of visits per day of the different scenario is close to 28 (between 21.5 and 36.5), i.e. close to one visit per day for each station, as in the night rebalancing visits. In particular we can see that for Scenario 2, 4 and 6, with the standard stations size, the average number of visits per station is 1.30, 1.11 and 1.14, respectively, so some stations are visited more than one time. On the other side, in Station 8, 10 and 12, the number

	Scenario					
Size	Standard			Optimal		
State	Std		Opt	Std		Opt
Night	-	Yes	Yes	-	Yes	Yes
Day	Yes	Yes	Yes	Yes	Yes	Yes
Station	2	4	6	8	10	12
1	5.60	3.76	3.75	3.15	2.96	2.79
2	2.79	1.87	2.41	1.85	1.06	1.25
3	2.06	1.48	1.80	1.84	1.47	1.53
4	3.08	2.96	3.70	1.93	1.36	1.88
5	11.52	9.65	11.46	3.57	2.96	3.79
6	0.00	0.00	0.00	0.00	0.00	0.00
7	3.16	1.46	2.90	2.65	2.05	2.52
8	1.75	0.85	1.01	1.44	0.51	0.56
9	1.22	0.72	0.72	0.80	0.45	0.46
10	3.09	2.44	2.19	1.98	1.28	1.36
11	2.00	0.42	0.35	2.18	1.40	1.48
12	1.19	0.67	0.88	1.32	1.26	1.04
13	4.18	3.10	3.08	3.35	2.32	2.69
14	1.51	0.65	0.37	1.77	0.93	0.56
15	6.77	5.99	5.60	2.31	1.39	1.15
16	5.79	4.99	5.55	3.64	2.35	4.35
17	0.49	0.06	0.02	0.81	0.78	0.79
18	0.70	0.07	0.10	2.01	0.79	1.12
19	0.59	0.05	0.03	0.79	1.75	1.02
20	0.45	0.47	0.42	1.17	3.26	1.59
21	6.35	6.47	7.25	3.33	3.26	4.26
22	6.14	5.39	4.41	3.52	2.96	2.75
23	1.44	0.63	0.30	1.98	1.04	0.58
24	4.62	2.95	3.38	2.21	1.38	1.65
25	30.93	30.92	32.74	8.32	9.56	11.36
26	0.07	0.01	0.01	1.06	0.99	0.85
27	2.17	4.21	2.18	1.13	1.50	1.16
28	2.33	4.57	2.45	1.04	2.07	1.22
Total	111.98	96.81	99.04	61.14	53.11	55.75
Mean	4.00	3.46	3.54	2.18	1.90	1.99

Table 7.11: The number of operations per day in each station in different scenarios during the daily re-balancing.

of visits decreases to 0.92 (-29.2% with respect to Scenario 2), 0.77 (-30.6% with respect to Scenario 4) and 0.80 (-29.8% with respect to Scenario 6), respectively. More in detail, we can notice that for the first three scenarios the number of visits in Station 25 is more than 8 per day, due to the small station size of 14 docking stations. In the last three scenarios this metric decreases to 2-3 visits per day, as the station size increases to 41 docking stations, leading to considerable potential savings in operational costs.

The daily operations grouped by time interval

During the day, from 07:00 to 21:00, the re-balancing vehicle is active to refill the stations that have a shortage of bikes or docking stations. Table 7.13 shows the average number of operations per day done by the service vehicle. These operations are grouped by time intervals of two hours. In Scenario 2, 4 and 6, we can notice an high number of operations in the morning in the interval between 7:00 and 9:00, that is on average 35.05, 27.19 and 32.92 operations, respectively. On the other side, in Scenario 8, 10 and 12, in the same time interval, the number of operations decreases to 14.08 (-59.8% with respect to Scenario 2), 13.54 (-50.2% with respect to Scenario 4) and 15.88 (-51.8% with respect to Scenario 6) operations, respectively. Other time intervals in which the service vehicle is particularly active are between 15:00 and 17:00, and between 17:00 and 19:00, while it seems that in the other time intervals the stations have a good self rebalance. In conclusion, the results of Table 7.13 confirms that the optimal stations size configuration may sensibly decrease the costs of the daily re-balancing, and service provider may consider limiting the working hours of the vehicle to the peak periods. However, a further analysis should be conducted in this topic, to determine the impact of shorter rebalancing time intervals on the total number of violations.

The daily route

Table 7.14 shows the average number of time spent by the service vehicle to rebalance the BSS, and the related number of kilometres travelled. Also, we have a summary of the number of operations and the number of visits, as previously determined.

	Scenario					
Size	Standard			Optimal		
State	Std		Opt	Std		Opt
Night	-	Yes	Yes	-	Yes	Yes
Day	Yes	Yes	Yes	Yes	Yes	Yes
Station	2	4	6	8	10	12
1	1.70	1.45	1.39	1.52	1.43	1.46
2	1.03	0.73	0.96	0.79	0.56	0.57
3	0.71	0.51	0.56	0.88	0.71	0.78
4	1.19	1.02	1.39	1.04	0.69	0.89
5	3.04	2.56	2.78	1.11	0.79	1.11
6	0.00	0.00	0.00	0.00	0.00	0.00
7	0.96	0.52	0.84	1.16	0.77	1.00
8	0.76	0.39	0.42	0.55	0.20	0.18
9	0.55	0.30	0.31	0.40	0.24	0.21
10	1.06	0.75	0.74	0.90	0.56	0.56
11	0.60	0.14	0.13	1.12	0.67	0.73
12	0.48	0.26	0.35	0.66	0.60	0.58
13	1.41	0.92	1.06	1.58	1.07	1.32
14	0.57	0.23	0.18	0.75	0.47	0.30
15	2.74	2.42	2.30	0.78	0.45	0.43
16	1.53	1.10	1.34	1.31	0.84	1.33
17	0.16	0.02	0.01	0.55	0.50	0.52
18	0.31	0.03	0.04	0.76	0.38	0.45
19	0.20	0.02	0.01	0.51	0.92	0.75
20	0.19	0.20	0.12	0.71	1.42	0.89
21	2.06	2.05	2.24	1.15	1.19	1.41
22	1.75	1.61	1.36	1.29	1.08	0.89
23	0.71	0.22	0.18	1.08	0.55	0.40
24	1.51	1.12	1.22	0.98	0.61	0.72
25	8.38	8.30	8.95	1.98	2.27	2.81
26	0.04	0.01	0.01	1.06	0.99	0.85
27	1.35	2.05	1.40	0.62	0.68	0.65
28	1.53	2.22	1.62	0.64	0.84	0.68
Total	36.50	31.14	31.91	25.88	21.46	22.45
Mean	1.30	1.11	1.14	0.92	0.77	0.80

Table 7.12: The average number of visits per day in different stations during the daily rebalancing.

	Scenario					
Size	Standard			Optimal		
State	Std		Opt	Std		Opt
Night	-	Yes	Yes	-	Yes	Yes
Day	Yes	Yes	Yes	Yes	Yes	Yes
Time	2	4	6	8	10	12
07-09	35.05	27.19	32.92	14.08	13.54	15.88
09-11	14.52	13.17	11.48	10.05	12.29	12.68
11-13	6.00	4.35	4.30	5.24	4.13	3.59
13-15	8.69	6.50	5.60	4.99	4.81	4.09
15-17	17.21	16.93	16.77	7.15	5.10	4.86
17-19	23.96	23.22	22.58	14.64	9.75	10.75
19-21	6.56	5.45	5.39	4.99	3.49	3.89
Total	111.98	96.81	99.04	61.14	53.11	55.75

Table 7.13: The average number of bikes operations per day in different time intervals of the day during the daily rebalancing.

The last row of the table shows the percentage of time spent in the re-balancing operations on the total of 14 hours available. Notice that the amount of resources is underestimated, since it does not take into account the time spent in the stations to add and remove bikes, and the time in which the vehicle is available but there is no need of it. Indeed, the service vehicle is active 14 hours, even if in Table 7.14, we see that the active working time is between 2 hours and 26 minutes in Scenario 12 (25.4% of the available time) to 3 hours and 31 minutes in Scenario 2 (25.1%).

7.4 Summary results

Table 7.15 shows a summary of all the performance indicators introduced in the previous sections and it reports, starting from the first left column:

- the scenario number;
- the total number of docking stations available;
- the total number of bikes available;
- the number of docking stations added in the BSS;

	Scenario					
Size	Standard			Optimal		
State	Std		Opt	Std		Opt
Night	-	Yes	Yes	-	Yes	Yes
Day	Yes	Yes	Yes	Yes	Yes	Yes
Indicator	2	4	6	8	10	12
Operations	111.98	96.81	99.04	61.14	53.11	55.75
Visits	36.50	31.14	31.91	25.88	21.46	22.45
Distance (km)	85.44	68.38	68.84	76.72	68.12	65.93
Time	3h 31m	2h 49m	2h 50m	2h 58m	2h 28m	2h 26m
Time Active (%)	25.1 %	20.1 %	20.2 %	21.2 %	17.6 %	17.4 %

Table 7.14: The key metrics of the daily re-balancing in the different scenarios.

- the number of docking stations removed from the BSS;
- the average number of operations during the night;
- the average number of operations during the day;
- the average number of operations aggregating night and day;
- the estimated number of kilometres travelled by the vehicle in the night re-balancing route;
- the average number of kilometres travelled by the vehicle in the daily re-balancing route;
- the average number of kilometres travelled by the vehicle, aggregating night and daily routes;
- the estimated hours spent by the vehicle in the night re-balancing route;
- the average hours spent by the vehicle in the daily re-balancing route;
- the average hours spent by the vehicle, aggregating night and daily routes;
- the average number of violations per day.

Clearly, every community desires a service with a high number of bikes, stations and re-balancing vehicles that assists during all the days, but this is not feasible in all the BSSs. Each choice of the service provider affects the quality of the service and the investments should ensure a benefit in the system. A trade-off between costs and disservices should be found, taking into consideration all the possible costs that are included in the different scenarios. The proposed analysis and a comparison of the selected 15 configuration scenarios is a step forward in this direction, and it should be completed by taking economical aspect into consideration in future works.

Summary	**BSS**				**Vehicle**									**Metrics**
	Structure		**Docks**		**Bikes moved per day**			**Distance per day**			**Time per day**			Final
Scenario	Docks	Bikes	Added	Removed	Night	Day	Total	Night	Day	Total	Night	Day	Total	Violations
1	350	168	-	-	-	-	-	-	-	-	-	-	-	52.61
2	350	168	-	-	-	111.98	111.98	-	85.44	85.44	-	3 h 31 m	3 h 31 m	19.70
3	350	168	-	-	51.48	-	51.48	41.15	-	41.15	1 h 39 m	-	1 h 39 m	45.32
4	350	168	-	-	54.28	96.81	151.09	41.15	68.38	109.53	1 h 39 m	2 h 49 m	4 h 28 m	16.00
5	350	168	-	-	57.46	-	57.46	41.15	-	41.15	1 h 39 m	-	1 h 39 m	47.70
6	350	168	-	-	57.76	99.04	156.80	41.15	68.84	109.99	1 h 39 m	2 h 50 m	4 h 29 m	16.44
7	350	168	72	72	-	-	-	-	-	-	-	-	-	28.99
8	350	168	72	72	-	61.14	61.14	-	76.72	76.72	-	2 h 58 m	2 h 58 m	11.42
9	350	168	72	72	57.92	-	57.92	41.15	-	41.15	1 h 39 m	-	1 h 39 m	20.41
10	350	168	72	72	59.39	53.11	112.50	41.15	68.12	109.27	1 h 39 m	2 h 28 m	4 h 07 m	10.07
11	350	168	72	72	61.08	-	61.08	41.15	-	41.15	1 h 39 m	-	1 h 39 m	22.11
12	350	168	72	72	62.04	55.75	117.79	41.15	65.93	107.08	1 h 39 m	2 h 26 m	4 h 05 m	11.29
13	543	261	210	17	-	-	-	-	-	-	-	-	-	21.26
14	543	261	210	17	62.25	-	62.25	41.15	-	41.15	1 h 39 m	-	1 h 39 m	4.75
15	543	261	210	17	65.54	-	65.54	41.15	-	41.15	1 h 39 m	-	1 h 39 m	4.88

Table 7.15: The summary of the key metrics for the scenarios analysed in the theisis.

Chapter 8

Conclusions

The analysis conducted in this thesis have the purpose to support the configuration of a Bike Sharing System (BSS). Each of the methodologies introduced in the thesis shows both costs and benefits, which should be carefully evaluated by the service provider, toward proper choices that avoid extra costs or unpleasant disservices for the users of the BSS.

There are several problems in configuring and managing a BSS and, in this thesis, we have seen how optimization models can be applied to obtain a more efficient use of the available resources, such as bikes, docking stations and service vehicles. In particular, we proposed two optimization models to find a more efficient allocation of docking stations and bikes among the stations. We also introduced the static optimization model developed by *Dell'Amico et al. 2014* for the night re-balancing route, and the dynamic optimization model developed by *Contardo et al. 2012* for the daily re-balancing route.

We elaborated a process to support the configuration of a docking station BSS based on the evaluation of different scenarios by simulation. The decisions to make include the allocations of docking stations and bikes among the stations, and the possible use of a fleet of service vehicles to re-balance the stations, during the night or the day.

We selected 15 different scenarios and we implemented a simulation process to evaluate the impact of the optimization models and different configuration choices on the total number of violations. Finally, we evaluated the performances of the different scenarios on a real case study, namely the historical trips of the Padua BSS.

The conclusions that can be drawn from this work are related to different aspects of a BSS, we have tried to summarize the main ones:

- the BSS is more unbalanced during the working days, between Monday and Friday, as the users activity is sensibly greater with respect to the weekends (and national holidays). In particular, the most active days are Tuesday, Wednesday and Thursday and the time intervals in which is more crucial to balance the stations are 08:00-10:00, 16:00-18:00 and 18:00-20:00. Since an unbalanced systems leads to more violations, the service provider should take into consideration the use of more resources to manage the BSS in these days and time intervals;

- the initial configuration of a BSS with fixed docking stations is crucial to determine its success in the city. We have noticed that few stations causes most of the problems, as the trips, in particular small BSSs as the Padua's one, are mostly connected to these stations. Hence, an accurate study of the potential users demand should be conducted in phase of configuration of the stations size to avoid unpleasant disservices for the users. In case it is possible to resize the stations later on, a study of the historical users trips should be conducted and the stations should be resized as necessary. We have seen that a reallocation of the available docking stations can lead to a decrease by 45% of the number of violations. Furthermore, an increase by 55% of the number of docking stations in the system can lead to a decrease by 60% of the number of violations;

- the night re-balancing can be useful to refill the stations, but it has an important impact if the configuration of the stations size fits the users demand. In fact, in the current Padua BSS configuration, the night re-balancing decreases the number of violations by 14%, while in the optimal stations size configuration it impacts by 30%, and in the utopic stations size configuration it impacts by 78%. Also, a further investigation should involve the optimization model, as in the application of the static model developed by *Dell'Amico et al. 2014*, the night route is not always feasible when only one vehicle is available;

- the optimal initial stations state configuration, obtained by the application of an ILP optimization model (see Section 3.3) to the historical data, shows a slightly worst performance of the standard initial stations state, in terms of number of violations. We noticed that the ILP solution suggests a larger number of bikes in the system, namely 168, with respect to the standard initial stations state configuration, that suggests 154 bikes. The increase of the number of bikes leads to an increase of the probability that there is a shortage of docking stations. For this reason, other criterion should be investigated to improve the optimization model;

- the daily re-balancing plays an important role on the BSS operation, as a vehicle that refills the stations during the day can anticipate possible shortage of bikes and docking stations. The dynamic model developed by *Contardo et al. 2012* shows great benefit, if it is used together with a rolling horizon algorithm, to update the route as soon as new information becomes available. Moreover, the introduction of a route penalization to avoid any unnecessary visits to the stations has shown significant potential cost savings. The daily re-balancing in the case study of Padua shows a decrease of the number of violations by at least -60%. Also, the use of the daily re-balancing in addition to the night re-balancing decreases the number of violations by at least -50%, with respect to the usage of the night re-balancing alone;

- the re-balancing costs accounts for the 30% of all the costs on European BSSs, according to the *Bike Sharing Planning Guide* (*Büttner et al. 2011*). The results of Chapter 7 shows a sensible decrease of the daily re-balancing costs in the configuration with optimal stations size with respect to the configuration with standard stations size. In particular, the time spent by the vehicle in the re-balancing route decreases by 14%, the number of visits to the stations decreases by 30% and the number of operations decreases by 45%, in the Padua BSS.

In conclusion, optimization models and an accurate users demand analysis as well as a methodology to take advantage of them, are fundamental to make efficient decisions in a BSS with limited resources. An efficient allocation of the resources decreases the disservices for the users, without wasting more resources than necessary. The assistance of a service vehicle is often needed, through a night or daily re-balancing, to refill the stations. But service vehicles are not always the solution, as there are problems that persist in a long period of time. Hence, it is necessary to further analyse the historical demand, as to reallocate the docking stations among the stations. Responsible decisions of the service provider can lead to great benefits for the users and, in the long term, these decisions make the difference for success or decline of a Bike Sharing System.

References

Alvarez-Valdes, Ramon et al. (2016). Optimizing the level of service quality of a bike-sharing system. *Omega* 62, pp. 163–175.

Bartels, Richard et al. (1969). The simplex method of linear programming using LU decomposition. *Communications of the ACM* 12.5, pp. 266–268.

Bellon, Antonio (2018). A mathematical programming model for air traffic flow management with horizontal and vertical separation constraints. MA thesis. Università degli Studi di Padova.

Benchimol, Mike et al. (2011). Balancing the stations of a self service "bike hire" system. *RAIRO-Operations Research* 45.1, pp. 37–61.

Bicincittà (2020). *GoodBike Padova*. URL: `http://www.goodbikepadova.it`.

Brinkmann, Jan et al. (2015). Short-term strategies for stochastic inventory routing in bike sharing systems. *Transportation Research Procedia* 10, pp. 364–373.

— (2016). Inventory routing for bike sharing systems. *Transportation research procedia* 19, pp. 316–327.

Büttner, Janett et al. (2011). *Optimising bike sharing in European cities-a handbook.*

Caggiani, Leonardo et al. (2012). A modular soft computing based method for vehicles repositioning in bike-sharing systems. *Procedia-Social and Behavioral Sciences* 54, pp. 675–684.

— (2013). A dynamic simulation based model for optimal fleet repositioning in bike-sharing systems. *Procedia-Social and Behavioral Sciences* 87, pp. 203–210.

Chemla, Daniel et al. (2013). Bike sharing systems: Solving the static rebalancing problem. *Discrete Optimization* 10.2, pp. 120–146.

City Transportation Officials, National Association of (2016). *Bike Share Station Siting Guide.*

Conforti, Michele et al. (2014). *Integer programming.* Vol. 271. Springer.

Contardo, Claudio et al. (2012). *Balancing a dynamic public bike-sharing system.* Vol. 4. Cirrelt Montreal, Canada.

Dantzig, George et al. (1997). *Linear Programming. 1, Introduction {Springer Series in Operations Research}*. Springer-Verlag New York Incorporated.

Dell'Amico, Mauro et al. (2014). The bike sharing rebalancing problem: Mathematical formulations and benchmark instances. *Omega* 45, pp. 7–19.

DeMaio, Paul (2009). Bike-sharing: History, impacts, models of provision, and future. *Journal of public transportation* 12.4, p. 3.

Fishman, Elliot et al. (2014). Bike share's impact on car use: Evidence from the United States, Great Britain, and Australia. *Transportation Research Part D: Transport and Environment* 31, pp. 13–20.

Frade, Inês et al. (2014). Bicycle sharing systems demand. *Procedia-Social and Behavioral Sciences* 111.1, pp. 518–527.

Fricker, Christine et al. (2016). Incentives and redistribution in homogeneous bike-sharing systems with stations of finite capacity. *Euro journal on transportation and logistics* 5.3, pp. 261–291.

Garcıéa-Palomares, Juan Carlos et al. (2012). Optimizing the location of stations in bike-sharing programs: A GIS approach. *Applied Geography* 35.1-2, pp. 235–246.

Gauthier, Aimee (2013). *The bike-share planning guide.* ITDP Institute for Planning & Development Policy.

Goh, Chong Yang et al. (2019). Estimating Primary Demand in Bike-sharing Systems.

Google (2020). *Google Maps.* URL: `https://www.google.com/maps`.

Hernández-Pérez, Hipólito et al. (2003). The one commodity pickup-and-delivery travelling salesman problem. *Combinatorial Optimization—Eureka, You Shrink!* Springer, pp. 89–104.

Ho, Sin C et al. (2014). Solving a static repositioning problem in bike-sharing systems using iterated tabu search. *Transportation Research Part E: Logistics and Transportation Review* 69, pp. 180–198.

Hulot, Pierre et al. (2018). Towards station-level demand prediction for effective rebalancing in bike-sharing systems. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 378–386.

IBM (2020). *IBM ILOG CPLEX Optimization Studio*. URL: `https://www.ibm.com/products/ilog-cplex-optimization-studio`.

Kloimüllner, Christian et al. (2014). Balancing bicycle sharing systems: an approach for the dynamic case. *European Conference on Evolutionary Computation in Combinatorial Optimization*. Springer, pp. 73–84.

Kristianslund, Johannes et al. (2016). Optimal Repositioning in Bike Sharing Systems. MA thesis. NTNU.

Li, Yexin et al. (2019). Citywide bike usage prediction in a bike-sharing system. *IEEE Transactions on Knowledge and Data Engineering*.

Lin, Jenn et al. (2011). Strategic design of public bicycle sharing systems with service level constraints. *Transportation research part E: logistics and transportation review* 47.2, pp. 284–294.

Lin, Lei et al. (2018). Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach. *Transportation Research Part C: Emerging Technologies* 97, pp. 258–276.

Lozano, Álvaro et al. (2018). Multi-agent system for demand prediction and trip visualization in bike sharing systems. *Applied Sciences* 8.1, p. 67.

Martinez, Luis M et al. (2012). An optimisation algorithm to establish the location of stations of a mixed fleet biking system: an application to the city of Lisbon. *Procedia-Social and Behavioral Sciences* 54, pp. 513–524.

Melo, Rodrigo et al. (2018). *Distance and Travel Time Between Two Points from Google Maps*. URL: `https://github.com/rodazuero/gmapsdistance`.

Meteosolutions (2020). *3BMeteo*. URL: `https://www.3bmeteo.com`.

MetroBike, LLC (2011). *The bike sharing blog*. URL: `http://bike-sharing.blogspot.com/`.

Microsoft (2020). *Microsoft Excel*. URL: `https://www.microsoft.com/it-it/microsoft-365/excel`.

Neumann-Saavedra, Bruno et al. (2015). Anticipatory service network design of bike sharing systems. *Transportation Research Procedia* 10, pp. 355–363.

Nielsen, Brigitte Høj et al. (1993). *The bicycle in Denmark: present use and future potential.* Ministry of Transport.

Notebook, Jupyter (2020). *Jupyter Notebook.* URL: https://jupyter.org/.

O'Mahony, Eoin et al. (2015). Data analysis and optimization for (citi) bike sharing. *Twenty-ninth AAAI conference on artificial intelligence.*

Parikh, Pulkit et al. (2015). Estimation of optimal inventory levels at stations of a bicycle sharing system. *Transportation Research Board Annual Meeting.* Vol. 15.

Psaraftis, Harilaos et al. (2016). Dynamic vehicle routing problems: Three decades and counting. *Networks* 67.1, pp. 3–31.

Python (2020). *Python.* URL: https://www.python.org/.

R (2020). *The R Project for Statistical Computing.* URL: https://www.r-project.org/.

Raviv, Tal et al. (2013a). Optimal inventory management of a bike-sharing station. *IEEE Transactions* 45.10, pp. 1077–1093.

Raviv, Tal et al. (2013b). Static repositioning in a bike-sharing system: models and solution approaches. *EURO Journal on Transportation and Logistics* 2.3, pp. 187–229.

Regue, Robert et al. (2014). Proactive vehicle routing with inferred demand to solve the bikesharing rebalancing problem. *Transportation Research Part E: Logistics and Transportation Review* 72, pp. 192–209.

Romero, Juan et al. (2012). A simulation-optimization approach to design efficient systems of bike-sharing. *Procedia-Social and Behavioral Sciences* 54, pp. 646–655.

RStudio (2020). *RStudio.* URL: https://rstudio.com/.

Rudloff, Christian et al. (2014). Modeling demand for bikesharing systems: neighboring stations as source for demand and reason for structural breaks. *Transportation Research Record* 2430.1, pp. 1–11.

Saltzman, Robert et al. (2016). Simulating a more efficient bike sharing system. *Journal of Supply Chain and Operations Management* 14.2, p. 36.

Sayarshad, Hamidreza et al. (2012). A multi-periodic optimization formulation for bike planning and bike utilization. *Applied Mathematical Modelling* 36.10, pp. 4944–4951.

Schrijver, Alexander (1998). *Theory of linear and integer programming*. John Wiley & Sons.

Schuijbroek, Jasper et al. (2017). Inventory rebalancing and vehicle routing in bike sharing systems. *European Journal of Operational Research* 257.3, pp. 992–1004.

Shaheen, Susan et al. (2010). Bikesharing in Europe, the Americas, and Asia: past, present, and future. *Transportation Research Record* 2143.1, pp. 159–167.

Shaheen, Susan et al. (2013). Public bikesharing in North America: early operator understanding and emerging trends. *Transportation research record* 2387.1, pp. 83–92.

Soriguera, Francesc et al. (2018). A simulation model for public bike-sharing systems. *Transportation Research Procedia* 33, pp. 139–146.

Toth, Paolo et al. (2014). *Vehicle routing: problems, methods, and applications*. SIAM.

Van Heijningen, Hélène Margarethe Corine (2016). Exploring the design of urban bike sharing systems intended for commuters in the Netherlands. MA thesis. Technische Universiteit Delft.

Vogel, Patrick et al. (2011). Understanding bike-sharing systems using data mining: Exploring activity patterns. *Procedia-Social and Behavioral Sciences* 20, pp. 514–523.

Vogel, Patrick et al. (2014). A hybrid metaheuristic to solve the resource allocation problem in bike sharing systems. *International Workshop on Hybrid Metaheuristics*. Springer, pp. 16–29.