MSC THESIS – Final version
23/06/2020

Comparing the performance characteristics of the Public Bike-Sharing Systems of São Paulo and Rio de Janeiro

Paul Schilte
University of Twente
The Netherlands

Comparing the performance characteristics of the Public Bike-Sharing Systems of São Paulo and Rio de Janeiro

Department of Civil Engineering and Management (CEM)

Faculty of Engineering Technology (ET)

University of Twente

Author:
P.G. (Paul) Schilte

p.g.schilte@student.utwente.nl

S1470485

Supervisors and examination committee
Prof. Dr. Ing. K.T. Geurs (Karst)
Prof. Dr. M.A. Giannotti (Mariana)
Dr. K. Gkiotsalitis (Konstantinos)

UNIVERSITY OF TWENTE.

Escola Politécnica
Universidade de São Paulo

Preface

In the concluding period as a student, I have had the privilege to work and live in São Paulo for six months. During this time, I have been investigating the bicycle-sharing systems of São Paulo and Rio de Janeiro. Before presenting this final work as a student, I would like to take the opportunity to thank the people that made this thesis possible.

I want to thank my supervisors Prof. Dr. Ing. Karst Geurs, Prof Dr. Mariana Giannotti and Dr. Konstantinos Gkiotsalitis for taking away doubts and giving me the confidence to eventually deliver this final product. The meetings we had were always pleasant and the feedback was always helpful. During my stay in São Paulo I worked in the LabGEO, in an open atmosphere and with colleagues that were willing to help me. This made it a very amiable place to work, thank you for everything.

I would also like to thank the 'Van Eesteren-Fluck & Van Lohuizen Stichting', a foundation that was established by the founders of modern urban design. The foundation allocates subsidies to researches in the field of landscape architecture and urban planning. Their scholarship for my work allowed me to go to Brazil.

Additionally, I want to thank TemBici and especially Renata for always being responsive to my emails and helping me out with the questionnaire. Finally, I want to thank my parents, sisters and friends for their emotional and scientific support and interest.

Paul Schilte

Assen, June 2020

Research summary

The shift from individual motorized transport to more sustainable transport has been one of the main topics for many transport planners in the past decades. One of the emerging developments that are partly contributing to moving towards low carbon mobility are Public Bike-Sharing Systems (PBSS). One of the advantages of these systems is that they reduce congestion, which also implies less greenhouse gas emissions. Besides, choosing the bicycle over the car improves the physical health of the user. A PBSS is recognized by the solid stations from which the user can unlock a bicycle, cycle to the desired destination, and lock the bicycle to another dock close by. In recent years, the available systems have been exponentially growing all over the world and so has the academic interest to research these systems. However, the majority of the PBSS are found in Europe, North-America and Asia. Correspondingly, most of the academic research about PBSS comes from these areas. One of the countries where PBSS are relatively new in Brazil. Therefore, little research about the use of such systems in Brazil has been performed. Currently, there are seven bike-sharing systems operated by TemBici. The PBSS of São Paulo and Rio de Janeiro are the main subjects in this research. The research objective in this thesis is: To examine the spatial inequality in user access to the Public Bike-Sharing Systems of São Paulo and Rio de Janeiro, investigate the possible factors that influence the average station departures in these systems and explain the differences between the two systems.

The research is divided into two parts to achieve this objective. At first, the inequality in user access to the systems is analysed. The goal was to investigate if there exists a divergence between the population that lives within the so-called service area of the PBSS and those who live outside this designated area. The spatial inequality in user access is evaluated by comparing the average income, human development index (HDI) and the education level of the population living within the station's catchment area with the municipal averages. The results show a contrast between the areas; residents of the service area appeared to be wealthier, more developed and higher educated than the average inhabitant in both cities. This difference is particularly strong in São Paulo, where all the compared statistics of the service areas are superior to the surrounding areas, with the exception of a few high income neighbourhoods. In Rio de Janeiro, the service areas are also predominantly located in the wealthy and developed parts of the city. However, a significant number of stations, mainly located in and around the historic city centre, are located in 'middle-class' neighbourhoods with comparable income levels to the municipality. This raises the question if there are significant differences in bicycle use between the relatively wealthy and deprived neighbourhoods. The second research question further explored this issue.

The second part examines the factors which are explaining the station departures of both PBSS and what differences exist between the two systems. This is done by developing a prediction model for the number of departures per station. The objective is to include and test independent variables and examine which are significant contributors to the prediction model. In total, twenty variables were found in the appropriate spatial disaggregation and could be included in both models. This resulted in the involvement of many varieties of variables which can be divided into three categories. The first group of variables embodies the population characteristics such as the population density, ethnicity and the income per capita. The second group relates to the station, for instance, the capacity of the station and the station density. The last set of parameters relates to the presence of bicycle infrastructure close to the station. The available trip data from TemBici between April 2018 and September 2019 was used as the dependent variable for the model. In total, 2,9 million trips of BikeSampa and 10 million trips for BikeRio were utilized to calculate the average daily departures per station. Interestingly, the system in Rio de Janeiro generates nearly 3,5 times more trips on average in the same period. The fact that more residents are living in the service area and that the system of Rio de Janeiro was completed earlier, partially explain the higher values. The results of the prediction models for both PBSS helped to clarify other possible reasons. The final models reached similar values for the determination coefficient. Nevertheless, both of the models for BikeSampa ($R^2 = 0{,}42$) and BikeRio ($R^2 = 0{,}45$) have different significant independent variables. One interesting finding from the prediction model for BikeRio is that when average income increases in the service area, the use of the PBSS decreases. Considering that the majority of the stations is located in wealthy neighbourhoods, the choice of the locations of the stations might not have been appropriate for generating the optimum number of trips. On the basis of both

prediction models is also discovered that a higher percentage of black inhabitants in a service area has a significant positive relationship to the number of station departures. The service area of BikeSampa is inhabited for 9% belonging to this group, while the municipal average is 36%. The reason for the higher trip generation of the stations of BikeRio might be partly due to the higher number of black inhabitants in the service areas, which is set at 20%. Yet, this is less than half of the municipal average (46%). One significant variable for both cities was the proximity to public transport. When a bicycle station is located within 150 meters of a metro or train station, the number of departures is more than twice the average, which suggests that many users utilize the PBSS to cover the last part of their trip. The data was scrutinized by clustering some significant independent variables or data attributes. Ergo, the data was distinguished by day of the week and the service areas were grouped by their primary land use. Furthermore, the stations were clustered by capacity, station density and the average departures to improve the prediction. Nevertheless, the data clustering did not result in improved prediction models.

After answering the two research question is became apparent why particular stations perform better than other stations and which of the twenty tested variables are related to that difference. The recommendations followed from these findings can, when applied, have a positive impact on the average number of generated trips. Some limitation were that not all the desired variables were available in the appropriate resolution and many of the tested variables were found to be insignificant predictors. Besides, the service areas, especially in São Paulo, have homogeneous characteristics with low numbers of departures. This complicated the building of reliable prediction models. Furthermore, both of the PBSS were still in the developing phase when the research was conducted such that the station data was not always continuous and commensurate. The author recommends using the presented findings (considering the limitations) as a foundation to investigate other PBSS in Brazil, which are located in cities that are characterized by lower average income and a higher relative number of black residents, two characteristics of stations that generated more departures in São Paulo and Rio de Janeiro.

Table of contents

List of figures

List of tables

Research jargon

Banco Itaú = The financier of TemBici and the largest bank of Brazil

BikeRio = Public Bike-Sharing System of Rio de Janeiro, operated by TemBici

BikeSampa = Public Bike-Sharing System of São Paulo, operated by TemBici

Ciclofaixa = A lane for cyclists exclusively, but located next to the road. Often separated by lines or little cubes

Ciclorrota/ Via compartilhada = A road where bicycles and cars share the road. The maximum allowed speed for all vehicles is set at 30 km/h

Ciclovia = A separate lane, which is only allowed to use by cyclists

HDI = Human Development Index is a statistical composite index that measures life expectancy, education and income for a country or area.

Pardo = Brazilian of mixed ethnic ancestries. Usually a mixture of Europeans, Africans and/or Native Brazilians

PBSS = Public Bike-Sharing System

Real/Reais = Is the currency of Brazil. As of June 2020; €1 = R$ 5,95 / $1 = R$ 5,32

Service area (SA) = The area that is being served by a station. It is assumed that the inhabitants living in this service area will use the station, which is most likely located in the centre of the service area. The catchment area of one station is set to be a maximum of 10-minute walking to each station.

TBD = Number of Trips per Bike per day

TemBici = The operators of the PBSS in São Paulo, Rio de Janeiro and four other cities.

TPR = Number of Trips per resident

1. Introduction

Public Bike-Sharing Systems (PBSS) have been widely introduced in many countries during the last years. PBSS help with the reduction of congestion, since it can be seen as an alternative transport mode. Besides, choosing the bicycle over the car reduces the emission of greenhouse gases and improves the physical health of the user. These advantages make the (development of) PBSS essential for public transport systems in cities. The majority of these systems are found in developed regions such as Europe, North America and also China, but the PBSS have been on the rise for several years in Latin-American countries including Brazil. Few kinds of research have been performed about the PBSS in this part of the world and this thesis will aim attention at one Latin-American country, Brazil and two cities in specific: Rio de Janeiro and São Paulo. Both cities currently have an operating PBSS. The assumption that a PBSS in Brazil yields the same results as a comparable system in Europe or the US does not hold. Brazil has more problems relating to poverty, violence and income inequality, and therefore, the country may need a different approach when implementing a PBSS. Thereby, little quantitative research about PBSS in Brazil has been done so far. The research gap that this study aims to fill is to expand the knowledge about PBSS in Brazil and analyse which variables contribute to a well-functioning PBSS in São Paulo and Rio de Janeiro, the two biggest cities of the country.

Historically, there have been three generations of PBSS (DeMaio, 2009). The first generation was developed in Amsterdam in 1965. These so-called 'Witte Fietsen' (white bikes) were introduced to reduce car emissions and consumerism (Médard de Chardon, 2016). The municipality of Amsterdam painted ordinary bicycles white and provided them for public use. The idea was that the user takes the bicycle and ride to his or her destination and leave it for the next user. This initiative failed within days because the bicycles were utilised for private use or thrown in the canal (DeMaio, 2009). Despite the failure of the PBSS in Amsterdam, some other cities in Europe introduced the same concept. The results were predominantly the same; program failure. Introduced in 1995 in Copenhagen, the Bycyklen, or City Bikes, was the first large-scale urban bike-sharing program in the second generation (DeMaio, 2009). The second-generation distinguished itself with 'Coin-Deposit Systems'. The user had to pay a small amount to unlock the bicycle. The deposit was retrieved when the user returned the bicycle to one of the docking stations, but since the user remained anonymous, the bicycle theft continued to be a problem for the second generation (Médard de Chardon, 2016). Technological improvements led to the introduction of the third-generation PBSS. Users had to use mobile phones, smartcards and credit cards to rent a bicycle which also meant they no longer remained anonymous. The first widely recognized PBSS of the third generation was developed in Lyon in 2005 and led to global implementation of bike-sharing systems (DeMaio, 2009). The last decennium, the number PBSS has been grown steadily. In 2014, PBSS were operating in 50 countries spread over five continents and 712 cities. A total of approximately 806.000 bicycles are operating between 37.500 stations (Marchuk et al., 2016). As of December 2016, there were around 1000 cities with a PBSS (Gutman, 2016).

1.1 The Public Bike-Sharing System of TemBici

The research will focus on the Public Bike-Sharing Systems (PBSS) operated by TemBici. In total, Tembici operates PBSS in seven cities, of which six are located in Brazil and one in Chile (Bikeitaú, 2019). Two of those systems, located in São Paulo (BikeSampa) and Rio de Janeiro (BikeRio) will be addressed in this study. BikeSampa and BikeRio are operated by TemBici since 2018. However, both cities have seen similar bicycle sharing systems before. In the case of São Paulo, a previous system, with a different operator, was providing shared bicycles from 2012 to 2017. The system eventually failed because it was not robust enough to sustain. TemBici has changed the type of bicycles and chose different locations for the stations to assure a better performing system and built a new 'network' of stations from scratch. TemBici works in cooperation with Banco Itaú (Itaú Bank), which is the largest Brazilian bank. One can easily recognize the bicycles by the orange colour and the logo of Itaú flaunts on the fender of each bicycle (Figure 1). With their robust construction, the bicycles are designed to ensure a long lifespan. Furthermore, the bicycles are equipped with GPS to be able to track them down when they are lost or stolen. The other characteristic of the PBSS are the stations, where the bicycles can be picked up and returned to. The system was built using the PBSC technology that has three main components; the solar panel (1) ensures that the station is self-sufficient and doesn't need an external power source. (2) is a payment terminal where the user can process payments wireless. The actual dock (3) is where the bicycle can be picked up and returned to. This can be done using the user card, station code and mobile phone application (Rabello, 2019).

Figure 1: The bicycle and the components of the bicycle stations (Bikeitaú, 2019)

To be able to utilise the bicycles, one has to download the application for the mobile phone and connect his/her credit card. The application shows the location of the station and how many bicycles and free docks are available at that moment. Furthermore, the application is used to unlock the bicycle from the station. The price of using the service depends on the subscription the customers chose. All the systems have the same tariffs with the cheapest being a daily subscription for R$ 8[1], followed by a three-day plan for R$ 15. To use the system for one month will cost the customer R$ 20, three months for R$ 50 and the price tag for a yearly subscription is R$ 160. The programmes are not tied to a city, which means that one can use the system in all the seven cities.

The network of the PBSS is formed by designated stations that are placed across a service area, which is usually a city. Ideally, the stations are easily accessible and connected by proper bicycle infrastructure, which allows the user to travel from A to B as convenient and fast as possible. However, in reality, the situation is often different, as many factors can negatively or positively influence the performance of PBSS. These factors can affect the performance of a whole PBSS or just one station or area. For instance, the climate of a city influences the performance of the system as a whole. On a rainy day, people are less likely to cycle compared to a sunny day. Examples of factors that affect the local performance are the availability of bicycles at each station and to which services the station provides access to. Moreover, having a bicycle station close does not necessarily mean that potential users have uncomplicated access to the network. For instance, poorly maintained infrastructure or too few available bicycles can exclude stations and people from a network. These examples of factors can vary significantly within and between cities and partly indicate how many trips a PBSS generates. Besides, an existing network is often not comprehensive enough to serve a whole city. Therefore, some neighbourhoods or city-districts are not connected to the network, which can result in differences in access to the bicycle network between residents or population groups. Notably, areas with limited transportation opportunities could benefit more from access to a PBSS. For this matter, the data of the PBSS in São Paulo and Rio de Janeiro will be analysed and compared.

[1] 1 R$ = €0,17 / 1 R$ = $0,19 (June 2020)

1.2 Research objectives and questions

The research will analyse the PBSS of São Paulo and Rio de Janeiro by comparing which population groups have access to the system and which characteristics and variables are important in reaching a PBSS that generates more trips. Therefore, the main research objective is:

To examine the spatial inequality in user access to the Public Bike-Sharing Systems of São Paulo and Rio de Janeiro, investigate the possible factors that influence the average station departures in these systems and explain the differences between the two systems.

The report is divided into two parts to answer the objective. The first part explores the possible inequality in user access to the system and the second part investigates the available data of the stations' surrounding and seeks to find relationships between the data and the number of trips a station generates and compare the two systems. Even though both municipalities are located in Brazil, the history of both cities is fairly different. São Paulo and Rio de Janeiro have been developing in their own way over the past centuries, which has resulted in different demographics. Ergo, variables such as income, quality of life and the culture of the inhabitants are different. The factors that influence the use of bicycles, in combination with the external factors, are included to compare the number of departing trips between the stations and cities.

One of the advantages of a station-based type PBSS is that the collection of data about the origin, destination and the time that was needed to succeed the trip is relatively easy to collect and analyse. Data about the stations' demographics and other attributes still had to be collected. The goal is to include as many variables as possible to ensure a robust and reliable model. Another focus point in this thesis is to examine equality in user access to each system. To meet this objective, the relative use of the PBSS for a particular population group will be analysed. The results will gain insight into possible excluded groups, that may have a high potential to use the system when having access.

Part of the research also examines the strategies and operations of the operator, TemBici. The approach of the operator for each city, and decisions relating to the location of the stations will also be addressed. The information regarding these decisions was collected through an interview. The findings from the interview and the other research goals were combined to be able to contribute to the understanding in the PBSS operate. It is expected that revealing the strengths and weaknesses of each city could contribute to recommendations for future PBSS or possible guidelines for the expansion of the existing systems. This has resulted in the following two research questions:

1. *How is the spatial inequality in user access to the PBSS inside and between the systems of São Paulo and Rio de Janeiro?*

2. *What are the factors which are explaining the station departures of the PBSS in São Paulo and Rio de Janeiro and what are the differences between these two cities?*

The research questions are answered by starting with literature research (chapter 2) that discusses the possible influential factors on bicycle use. The literature study also explores the relevance of this topic for Brazil and the possible methods to predict the number of departures using other studies. Chapter 3 elaborates on the methods that were applied during this thesis. Followed by chapter 4, which describes the case study. In chapter 5, the results of the analysis are given and the research questions are answered. The last chapters contain the discussion of the results, recommendations for further research and the final conclusions.

2. Literature study

This chapter elaborates on a broader assessment of the available literature regarding the core aspects of this research. The first research question examines and compares the spatial inequality in user access to the system. The factors that describe specifically the spatial inequality in user access are discussed in subsections 2.1.1 and 2.1.5. The rest of this chapter and section 2.1 are outlining some relevant literature regarding the answering of the second research question. The possible factors that influence the use of (public) bicycles that will later be tested in the prediction models are described in section 2.1. This will guide the process of determining which data is useful to collect and, therefore, could contribute to a more advanced and precise prediction model. 2.2 explores the different trip purposes and how this can affect the use of the system. The methods to measure and model the performance of PBSS are illustrated in sections 2.3 and 2.4 and the last section, 2.5, gives a summary of the chapter

2.1 Factors that influence bicycle use

This section will resume on the previous research relating the factors that influence bicycle use. The paragraphs are divided into different topics and use various studies to examine the impact of the variable on the use of bicycles in general and more specifically on the trip generation for PBSS. The variables that turn out to be important influencers will be used in the later developed prediction model. The discussed variables in this section are depicted in Figure 2 and the following subsections will chronologically illustrate the influence of the schematized factors. Furthermore, the researched papers often presented whether the expected influence of the variable on bicycle use is positive or negative. The framework of the factor is coloured green if the expected effect on bicycle use is positive, e.g. if the population density increases so do the expected number of trips. The orange colour is used for the variables where the literature could not rule out the existence of both a positive and a negative relationship. A clear example of such a variable is the average income. Finally, a few variables are framed in red, they are expected to be negatively correlated with the trip generation, e.g. more slopes will result in a lower average number of trips.

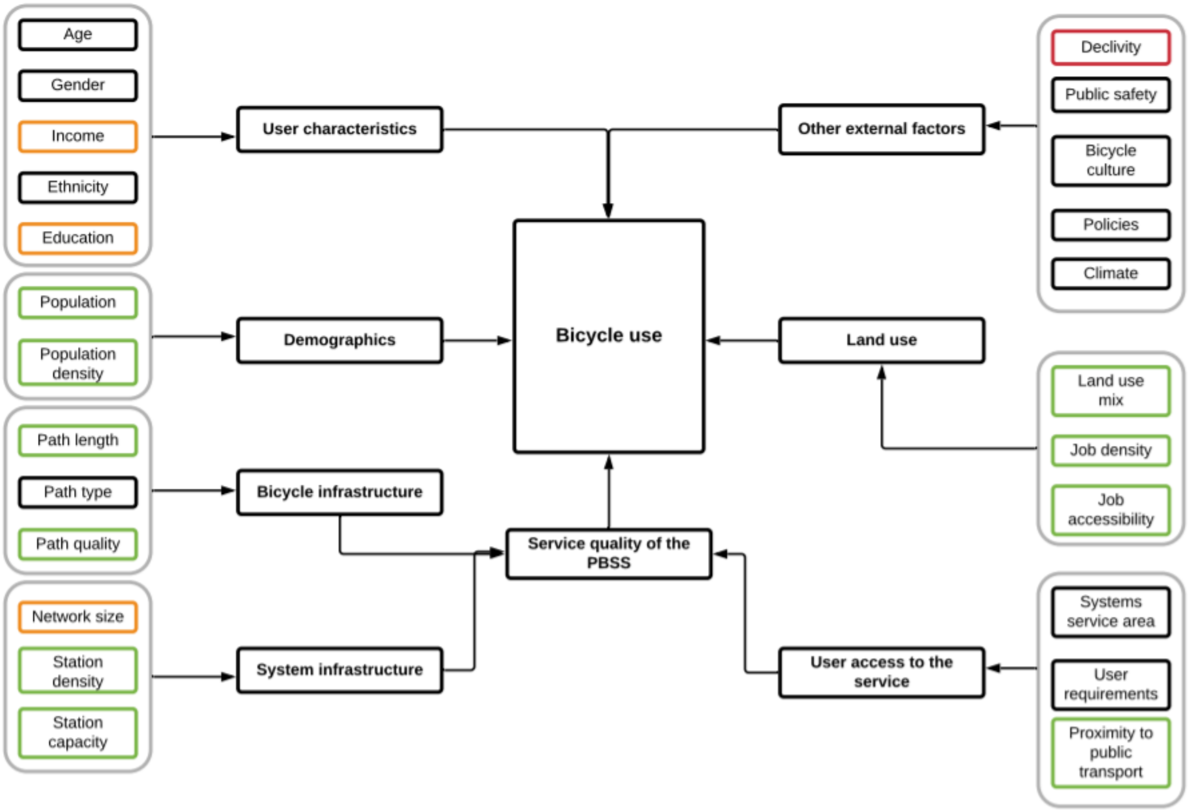


Figure 2: Factors that influence bicycle use

2.1.1 User characteristics

Socio-demographic variables such as age, gender, income and ethnicity are important travel behaviour determinants. In terms of gender, in countries with low overall cycling levels, the majority of the total bicycle trips, around 80%, is made by males (Harms et al., 2014). In countries with relatively high cycling levels, the cycling rate between males and females seems to be more equally distributed. The differences in cycling share per age group also depend on the total cycling share. In countries with low overall cycling levels, mostly young adult males are using the bicycle, wherein countries with relative high cycling levels children and elderly are also cycling (Garrard et al., 2012). The relationship between average income and cycling level is rather contradictory. Firstly, a higher income means people can spend more money on a bicycle, resulting in higher cycling rates. On the other hand, people with higher income have higher rates of car ownership, which has a negative effect on cycling rates (Pucher & Buehler, 2008). The ethnic background is a socio-cultural aspect that also influences cycling levels. In the United States, which is a typical country with low overall cycling levels, the use of bicycles differs per ethnic group. White Americans are less likely to cycle African- or Latin-Americans. The main explanation is the lower average income and therefore, lower rates of car ownership. Another possible reason the spatial clustering of migrants groups in urban areas such that the average distance shortened, which encourages bicycle use (Smart, 2010). Research conducted in the Netherlands, a country with high cycling levels, found opposing results. Non-western immigrants cycle less and shorter distances than the native Dutch. They also have lower levels of car use than their native-born counterparts. Instead, they travel more by public transport (Harms, 2007). Various researches about user characteristics of PBSS show similar results. Even though the main function of the PBSS is expected to encourage social equity because of the low user costs, the actual users are observed to be wealthier, white, younger and male (Fuller et al., 2011; Marmot et al., 2010; Parkin et al., 2008; Steinbach et al., 2011) and even more likely to possess a car (Fishman, 2016).

2.1.2 Demographics

The size of the city in terms of the number of inhabitants does affect the share of the bicycle trips. In Brazil, the National Association of Public Transport (Antp) developed a report about this matter. If the number of inhabitants in an urban area increases, the relative bicycle use decreases. Cities with between 60 and 100 thousand inhabitants have a share in bicycle trips of 13% of the total trips. For larger cities (more than 1 million inhabitants) the share of total trips drops to 1% (Antp, 2012). However, at the time that this research was conducted, PBSS were yet to be implemented in Brazil. In the present day, PBSS are found in the larger cities of the country and might have led to an increased share of total trips. The included cities for this research are the two largest cities of the country with the largest being São Paulo with little over 12 million inhabitants and followed by Rio de Janeiro with 6.7 million residents. In addition to population size, population density also influences bicycle use in the city. A higher population density is linked to higher system performance (J. Zhao et al., 2014). However, when the population density reaches a certain level, pedestrians and cyclists have to cope with congestion as well (Krygsman et al., 2004). Furthermore, the variable population density does not consider tourists and commuters, who can represent an important proportion of users (Gauthier, 2013).

2.1.3 Bicycle infrastructure

The presence of cycling infrastructure promotes cycling. In Seville, a fully segregated bicycle network was developed between 2006 and 2011. The goal was to encourage bicycle mobility in a city without a cycling culture. The new infrastructure made cycling not just safe, but also easy and comfortable. The results were predominantly positive, the number of bicycle trips per day increased from 13 thousand in 2006 to almost 73 thousand in 2011(Marqués et al., 2015). Despite, the absolute number of bicycle crashes per year increased, the relative amount of bicycle traffic injuries decreased, making it safer to cycle in the city. Controversially, in developing countries, like Brazil, the lack of bicycle infrastructure prevents potential users from choosing the bicycle as a mode of transport. If present, the infrastructure is often a shared lane, which can be considered too dangerous by potential users (de Souza et al., 2017). A safe bicycle network is especially essential for women (Daley et al., 2007). Females are less likely to use the bicycle as a mode of transport in countries where cycling has a low modal share of transport trips (Garrard et al., 2012). Unlike males, females prefer off-road (segregated) bicycle infrastructure over bicycle paths located alongside the road (Garrard et al., 2008). A survey from Brazil found that 95.4% of the respondents consider that it is important to have dedicated cycle lanes (Freitas & Maciel, 2017a). There are three main types of infrastructure found in both São Paulo and Rio de Janeiro, shown in Figure 3. The ‘ciclovia’ is a separate lane for cyclists exclusively. The ‘ciclofaixa’ is a path located next to the road, often separated by lines of little cubs. The ‘ciclorrota’ is a road where bicycles and cars share the road and the maximum allowed speed for all vehicles is set a 30 km/h.

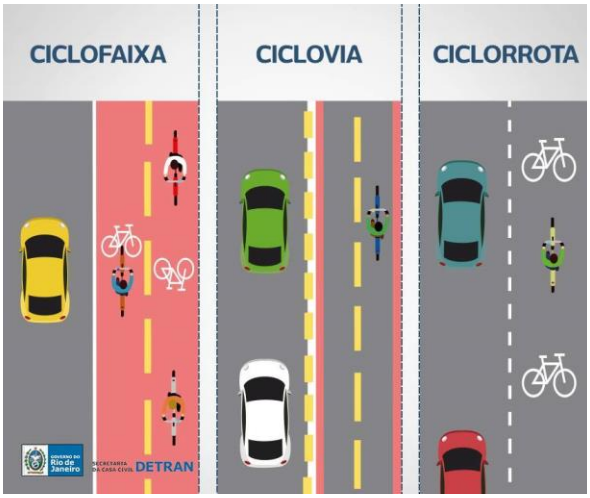


Figure 3: Types of bicycle paths in Brazil (DETRAN, 2016)

2.1.4 System infrastructure

Médard de Chardon et al. (2017) researched the determinants that define a well-performing bicycle sharing system for 75 PBSS across five continents. One argument that is often unjustly used by policymakers is that solely large systems with many stations can generate large numbers of trips. The author assessed the PBSS by calculating the number of Trips per Bike per day. From the 75 BSS that were analysed, some of the systems with the highest TDB were found in ‘small’ systems, such as Ljubljana, Dublin and Vilnius, with 33, 49 and 33 stations. It was found that when certain characteristics, such as a high variety of land-use and population density are present, the bicycles of the BSS are used more often. In contrast, some of the big BSS in Brussels, Minneapolis and Brisbane, with 323, 169 and 151 stations respectively reach 8 to 15 times less TBD than the small systems (Médard de Chardon et al., 2017). Furthermore, the station density increased the performance per station by 4 – 32% per square kilometre. Higher station density also decreases the distance between the closest bicycle station and the true origin or destination of the user. Ultimately, the distance that users have to span if their desired station is full or empty is lessened when the station density is higher. So station density is also a measure of resiliency and reliability of the system (Médard de Chardon, 2016).

2.1.5 User access to the service

The coverage area of a PBSS includes a 500-meter radius around each station (Gauthier, 2013). Ideally, the reflection of the socio-demographic groups is equitable within the coverage area. This subject was researched for London (Ogilvie & Goodman, 2012). The results proved the opposite; the PBSS were more often located in wealthier neighbourhoods. Therefore, fewer people from deprived areas lived close to a station and accordingly, fewer are registered to the PBSS. Among those who did register, the usage of the bicycles was higher, which indicates that there is an unmet need for cycling in the deprived areas of London. Stewart et al. (2013) pointed out that credit card requirements exclude people from participating. Especially people who live in low-income areas are less likely to own a credit card. In the case of São Paulo and Rio de Janeiro, economic goals rule the implementation of the system too. The bicycles are sponsored by a private entity, meaning the bicycles share scheme could be used for brand promotion. As a result, improving public transport and accessibility on city-level is often not the primary goal of the system (Duarte, 2016).

The access and egress to and from public transportation stations can be done with different modes, i.e. by bicycle. A PBSS offers an environmentally friendly alternative solution for the 'last mile' problem, which describes the short distance between home/work and the public transport system that have to be bridged. This distance may be too far to walk, and bike-sharing could play an important role to cover this gap (Shaheen et al., 2010). When properly integrated with public transport systems, the bicycle is an efficient way to increase the catchment area of a public transport service (de Souza et al., 2017). People are willing to cycle between 2 or 3 km to reach the bus or metro (Martens, 2004). The geographical location of the station does affect the temporal pattern of the trip too (McBain & Caulfield, 2018). Bicycle stations located adjacent to transit hubs show higher peaks in morning and afternoon demand, as these stations are used as feeder stations for the public transportation network. On the other hand, stations located near a park show higher usage of bicycles on the weekends (Médard de Chardon, 2016). Principally, the number of trips from and to a zone or specific station are dependent on the demographic, socio-economic and land use characteristics of each zone or station (Tsekeris & Tsekeris, 2011).

2.1.6 Land use

Neighbourhood environment characteristics can encourage or discourage cycling levels. Several studies found that areas with higher population densities, mixed land-use and high connectivity to public transport also see a higher share of non-motorized travel (Muhs & Clifton, 2016; Nielsen & Skov-Petersen, 2018; Saelens et al., 2003). Contrary, low density and single land use neighbourhoods, in which a large share of the United States' population lives, are associated with low levels of walking and cycling (Saelens et al., 2003). Since a large share of bicycle trips are for commuting purposes, the density of jobs, and more specifically, the jobs-housing balance are important factors to encourage commuting by bicycle (P. Zhao, 2014). Moreover, closer proximity or accessibility to services and jobs increase the levels of walking and cycling (Kockelman, 1997; Schneider, 2011). Accordingly, the effect of land-use on the use of the PBSS can be measured with the land-use mix index, average job density and the job accessibility by bicycle or public transport for each service area.

2.1.7 External factors

Not all the influencing factors that were found in the literature can be measured and therefore, included in the model. The author chose to put a concise description of these factors below.

Policies: Certain policies implemented by a governing body can influence the use of the PBSS. For instance, the obligation to wear a helmet while utilizing the PBSS results in fewer trips, mainly because people have to bring their own helmet (Basch et al., 2014; Fishman et al., 2014). However, in Brazil, there is no such thing as helmet obligation for using a bicycle.

Public safety: Public safety is has been a continuous problem in Brazil and the rest of Latin America. Especially woman are less likely to cycle in unsafe environments (Emond et al., 2009). If one is cycling, he or she is more exposed to threats and this could lead to an unsafe feeling. Hence this person is less likely to use the bicycle. Thereby, the annual bicycle theft rates in Brazil are significantly higher than in Europe and North America (8.1% versus 3.2%) (Kahn et al., 2002). These facts make implementing a lucrative PBSS in Brazil more complicated.

Climate: High levels of humidity and high temperatures decrease the likelihood of choosing the bicycle as a mode of transport. Having a comfortable climate helps to develop a bicycle culture and eases infrastructure maintenance (Médard de Chardon et al., 2017). The impact of weather and climate on the use of PBSS have been researched various times (Corcoran et al., 2014; Gebhart & Noland, 2014). The defined indicators are relative humidity, precipitation, wind and temperature. The effect of weather and climate on bicycle-commuting is influenced by both (short-term) weather conditions and (long-term) seasonal variations.

2.2 Trip purpose

In general, there are two distinguished trip purposes for PBSS; commuting and recreational. During the weekdays, relatively more commuting trips are made and during the weekend, recreational trips will have a higher share. It is important to understand the patterns in the trip purpose of the user when designing or expanding new bike programs (Fishman et al., 2013). Murphy (2010) points out that 70% of the trips made with the Dublin bike share program were work or education-related. A survey executed in four North-American cities also found that commuting was the most common trip purpose (Shaheen et al., 2012). These results are in line with the survey results of a medium-sized Brazilian city, however, a significant percentage of respondents used the bicycle for recreational purposes (Freitas & Maciel, 2017b). In general, the daily patterns can be divided into two groups, the weekdays and the weekends. This subject has been researched for the PBSS in São Paulo by Engels (2019). For weekdays, two peak moments in bicycle usage appeared; the morning rush hour (7:00 – 9:30) and the evening rush hour (17:00 – 20:00). Tough, the evening peak was notably higher than the morning peak. During the weekends, the usage of the system was relatively equally distributed, with Sundays being reasonably busier than Saturdays. The difference in the number of bicycle trips between the weekend days is partly due to an initiative from the government to make some of the principal roads of the city car-free on Sundays. The average renting period in the weekends (97 min.) is somewhat higher compared with weekdays (82 min.) (Engels, 2019).

2.3 Measuring the performance of PBSS

The most obvious way of measuring the performance of a PBSS is analysing the average number of trips in a system or per station. Various researches including (Buck & Buehler, 2012; Daddio, 2012; Maurer, 2012; Rixey, 2013) evaluated an existing PBSS or performed a feasibility study using the average, often departing, trips per station or area as a measure. Another more specific performance measure was introduced by IDTP (2013) and stated that the efficiency of the system could be measured with two critical performance metrics. The first one being trips per bike per day (TBD), a lucrative PBSS needs four to eight TBD. Fewer than four can result in a low cost-benefit ratio. More than eight daily uses will limit bicycle availability, especially during peak hours. The second metric describes the market penetration and is measured by the average daily trips per resident. Ideally, one daily trip per twenty to forty residents is needed to achieve this. High quantity of uses among the population within the coverage area (the area within 500 metres of the station) supports to reach the primary goals of PBSS. These two performance metrics are inversely related. The reason why systems have a high average daily use per bicycle could come from the fact that there are too little bicycles in circulation. On the other side, there are systems with high market penetration, but very few uses per bike, which could result in a low cost-benefit ratio. The planning of a PBSS must be carefully computed to assure that the performance is within the optimum range for both metrics (Gauthier, 2013).

Médard de Chardon et al. (2017) estimated the overall performance of PBSS for 75 stations. With this data, prediction models for the TBD for each station was built. Many of the independent variables that were used to create this model are described in section 2.1. However, no attention is paid to the type of cycling infrastructure. Yet, the use of PBSS is dependent on the available infrastructure. Type and quality of infrastructure significantly increased or decreased the performance of a station (Garrard et al., 2008; Marqués et al., 2015; Mateo-Babiano et al., 2016).

The profitability of bike-sharing systems for various cities in China was researched by J. Zhao et al. (2014). He concluded that the turnover rate is found when the bike-user ratio is approximately 0.2. I.e. each public bike should have at least 5 potential users. Nevertheless, the relationship between more public bicycles and higher bike-sharing ridership level is present, but too many public bikes can significantly decrease the system's effectiveness, which corresponds to the conclusions drawn in the ITDP planning guide for PBSS (Gauthier, 2013).

2.4 Forecasting the number of trips

Previous studies on station-level forecasting have been carried out with different approaches. Rixey (2013) investigated the effects on bike-sharing ridership near bike-sharing stations for three operational PBSS in the USA, using multivariate linear regression. The three PBSS that were included in the research already had been analysed individually and Rixey paid particular attention to the data quality and consistency issues that he considered more relevant in this multicity analysis. Maurer (2012) studied the feasibility of a bike-sharing program in Sacramento, California and emphasised his prediction model on reaching the highest values of R^2. Therefore, the linear regression model included 16 variables and no attention was paid to the significance of the individual variables. As a result, some independent variables had counterintuitive coefficients. Similar research conducted by Daddio (2012) included 14 variables to analyse and predict the departing number of trips for Capital Bikeshare (CaBi) in Washington D.C. Another prediction model for the daily number of departing trips for CaBi, emphasising on the influence of bicycle paths, was developed by Buck and Buehler (2012). The final model had a lower Adjusted R^2 value then Daddio's model, 0,66 compared with 0,78, respectively, but only included four significant (90% level) variables. The methods used in the studies of Buck and Buehler (2012), Maurer (2012) and Daddio (2012) were used by Rixey as input for a new model with improved applicability for other U.S. communities. Resulting into three prediction models that were developed using multivariate linear regression with the main focus lying on ensuring the statistical significance of the individual independent variables instead of maximizing the value of the (adjusted) R^2. The author considered a 10% level ($p < 0.1$) as significant, but most included independent variables were significant at 1% level ($p<0.01$). Despite not being the main purpose, the prediction models acquired a strong value of R^2 with the explained variation between 0,75, 0,80 and 0,80 respectively. Bivariate correlations between each independent variable and the dependent variable where calculated to establish which parameters should be included in the regression analysis.

Médard de Chardon et al. (2017) predicted the TBD using OLS and robust regression. Ordinary least squares (OLS) is a method for finding population parameters in a linear regression model. This method minimizes the sum of the vertical distances between the observed responses in the sample and the responses in the model. The resulting parameter can be expressed through a simple formula, especially in the case of a single regressor. Robust regression is a statistical procedure that aims to perform a regression analysis if the data set is contaminated with some points that do not belong to a (multivariate) normal distribution. This method differentiates from normal linear regression analysis, which usually performed using the least-squares method. A problem here is that the solution is sensitive to errors and deviations in the data. In a regression analysis in multiple dimensions, an outlier will sometimes look very harmless due to the projection used in graphic inspection. Therefore, there is a need for a method that identifies and neutralizes the outliers. The study included 75 cities spread over the world and the prediction models using OLS and robust regression reached R^2 values between 0,42 and 0,49. The number of bicycles and the number of stations were left out as predictors due to the high correlation with the dependent variable. The table below provides a summary of the analysed literature regarding station-based forecasting.

Author & year	Case	Type of models	Objective	R^2 value	Limitations
Maurer, 2012	Sacramento, CA	Multiple linear regression	maximize R^2	/	Counterintuitive coefficients because the significance of the individual variables was not emphasised
Daddio, 2012	Washington, D.C.	Multiple linear regression	maximize R^2	0,80 - 0,82	The author didn't find accurate data on job density
Buck & Buehler, 2012	Washington, D.C.	Multiple linear regression	statistical significance	0,66	One cannot determine causality from this analysis (a general problem in this field)
Rixey, 2013	Washington, Minneapolis/St. Paul, Denver	Multiple linear regression	statistical significance	0,75 - 0,80	Gathering comparable variables across the three systems
Médard de Chardon, 2017	75 cities over 5 continents	OLS and robust regression	Compare influential variables for BSS in various systems	0,42 - 0,49	Pays no attention to bicycle infrastructure

Table 1: An overview of the literature found on trip forecasting

2.5 Summary of the literature research

The first part of the literature study provided an overview of the factors that influence the use of bicycles in general and the use of PBSS in particular. Previous literature pointed out that users of a PBSS tend to be younger, wealthier and male. A possible reason is that PBSS are mainly found in richer neighbourhoods. The use in deprived areas is lower because the PBSS does not serve these neighbourhoods. The first research question will elaborate on this subject in the case of São Paulo and Rio de Janeiro. The objective is to include the utmost number of discussed variables in the models. This will depend on the availability of the data.

In most of the related literature, the performance of a PBSS is measured as the average number of departing trips per station per time unit. This method can be used when one has the possession of the trip data that includes the departing station of the trips. A second, more specific method has two performance metrics, the Trips per Bike per Day and the Trips per Resident per day. This method measures both the system's efficiency and market penetration. To succeed in calculating the TBD, one needs data on the available number of bicycles at a station per time slot. This specific data was not available for BikeSampa and BikeRio, so the author chose to use the average daily departing trips per station as a measure to analyse the influential variables on the use of the PBSS of São Paulo and Rio de Janeiro. The final part of the literature study explored the methods that were applied to model the performance of the PBSS. All the case studies that were evaluated used a form of linear regression. Since the objective is to include and test the utmost available independent variables, and multivariate linear regression seems to be the best regression method to succeed in this goal. Furthermore, the literature also pointed out that acceptable values of R-squared can be reached with this method.

3. Methodology

This chapter describes which methods were applied to meet the research objective. Firstly, the procedures to calculate the stations' service area are explained. Both research questions used the same service areas and the same conversion method to calculate the census averages per area. Section 3.2 outlines the specific applied methods for answering the first research question and section 3.3 describes the practised processes to succeed in answering the second research question. The final part of this chapter describes the interview with TemBici, provides the research framework and summarized the assumptions which were made during this research.

The figure below is a revised version of Figure 2 and depicts the variables that were available in the appropriate spatial disaggregation. The figure reveals that not all the illustrated variables from the literature study could be included in this study. Notwithstanding, it was possible to collect at least one variable for each overarching category. The variables related to the station, such as station density, station capacity and proximity to public transport are already applicable for prediction models. The variables that describe the average over a certain area have to be shaped to service area averages, which is explained in 3.1.1

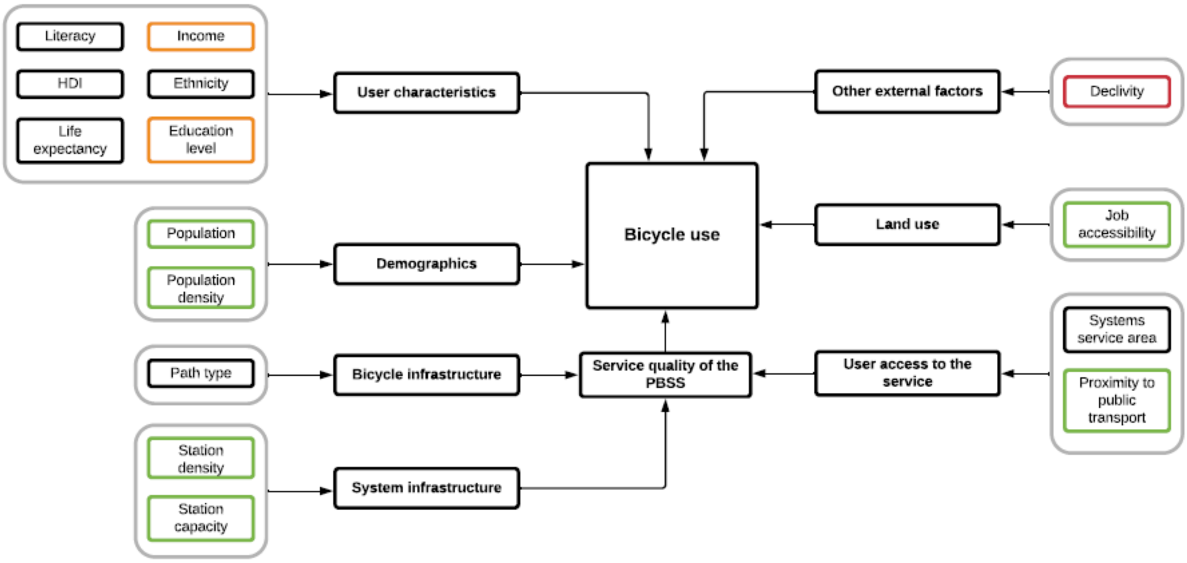


Figure 4: Available and included variables for the prediction models

3.1 Estimating the service areas of BikeSampa and BikeRio

Before it is possible to answer any of the research questions, a service area has to be calculated. The available data on the variables which influence the number of trips has to be aggregated to an average value per station. Therefore, the catchment area around each of the stations has to be determined. In the case of a bicycle station, this is generally a small area because people are more likely to travel to and from the station on foot. Literature research suggests a service area with a perimeter of 500 meters around each station (Médard de Chardon, 2016). The availability of the road network with intersections made it possible to calculate the service area with higher accuracy by approaching the maximum walking distance to and from a station using the Network Analyst in ArcMap. It was chosen to take a maximum walking time of 10 minutes as the boundary of the service area. When taking the intersections into account, a 10 minute would be equal to an average walking distance of 500 meters. If an area has more than one bicycle station within the acceptable walking time, the closest station is considered.

3.1.1 Calculating census averages of the service areas

With the service areas being determined, the comparison with the rest of the municipality can be made. The objective is to analyse the (in)equality in user access within the network, but most of all with the rest of the city. Therefore, the census data have to be aggregated from the city blocks to averages per service area, which was done using the formula below.

$$Census\ data\ (service\ area) = \sum Census\ data\ (census\ area) * (\frac{service\ area \cap census\ area}{census\ area})$$

The formula sums up the percentual parts of the available data that are located in one service area to calculate an average value for this service area. A simplified example for calculating the average income of a service area of 4km^2; half of the service area is located in a neighbourhood with an average income of R$ 6000 and the other half has an average income of R$ 4000. Meaning that the average calculated income inside this area is R$ 5000.

3.2 Research question 1

The first research question: *'How is the spatial inequality in user access to the PBSS inside and between the systems of São Paulo and Rio de Janeiro?'.* The objective is to compare the calculated averages of the service areas with the municipal averages and find possible inequalities is user access to the systems. To analyse the spatial inequalities, four social characteristics of the inhabitants will be compared; the ethnic background of the population, the average monthly income per capita, HDI, and the percentages of inhabitants that graduated for medium education and superior education.

ArcGIS was used to compare and visualise the characteristics, but the level of detail and differences in sizes of the polygons, make the maps rather tricky to interpret accurately. Therefore, boxplots and tables were added to simplify the interpretation of the results. The significance of the statistical differences between the service areas and the municipality averages is tested using the t-test. This parametrical statistical test is used to determine whether the averages of the service area significantly deviates from the municipal average. The commonly used border values for t is set at 0.05 (5%) and will also be applied in this thesis. Section 5.2 describes and discusses the results of the analysis. The ethnicity of the population, income, HDI and education level will be compared. The comparison of education level is solely done for São Paulo due to the unavailability of such data for Rio de Janeiro. The author chose to include both the boxplot and summarizing table to provide a clear view of how the tested characteristics are distributed. Additionally, figures regarding the ethnicity and the income per service area are included. Presenting the data as such, makes it accessible to compare the results among and between the cities.

3.3 Research question 2

The second research question: *'What are the factors which are explaining the station departures of the PBSS in São Paulo and Rio de Janeiro and what are the differences between these two cities?'*. To answer this question, two main steps were taken. First, the TemBici trip data has to be aggregated to the number of departures per station per day, which is the dependent variable in the prediction models. Next, a regression method to predict the average departures has to be chosen and the final paragraphs elaborate on the clustering techniques that were applied to group the data based on specific attributes.

3.3.1 Preparing the trip data

The objective of TemBici to build 260 stations for both BikeRio and BikeSampa was achieved in February 2019 in Rio de Janeiro and in September 2019 in São Paulo. The number of stations that is available for analysis is increasing per month. BikeSampa started running at the end of January 2018 with 43 bicycle stations. The last months of included data are from September 2019 with 260 operating stations in both systems. BikeRio started to operate a few weeks later in February but had already 71 active bicycle stations in this month. The lack of operating stations and trips for BikeSampa in the beginning months is the main argument not to include the first three months in the analysis. TemBici delivered the trip data per month in CSV-format. Each row described a single trip, and the columns provided information (trip duration, origin, start date and time, destination, end date and time, user type, the birth year of the user and gender) about each trip. The complicated names of a number of the stations caused some problems in the encoding. To solve this, some manual changes in the names had to be made to assure that the trips were counted for the specific and unique bicycle stations. The next step is to manipulate the individual trips to daily overviews. To do so, the trip data was aggregated to the average number of daily departing trips per origin station per day, see Table 2. The output gave a matrix where the stations filled the rows and the days of the months the columns. To determine the TBD the number of available bicycles at the station for a given time moment is needed. Unfortunately, this information was not available, so an estimation had to be made.

Station ID	Station name	1-8-2019	2-8-2019	……..	29-8-2019	30-8-2019
1	Largo da Batata	1165	1089	…	1269	1244
2	CPTM Pinheiros	38	32	…	44	42
…	………………	…	…	…	…	…
234	R. Tripoli	7	8	…	4	11
235	Edward Weston	13	13	…	10	7

Table 2: The output matrix for the month August 2019 of BikeSampa

3.3.2 Building the regression models

A mathematical model has to be built to study the collected data to a further extent. The objective is to build a model that epitomises a series of statistical assumptions. One way of accomplishing this is through regression analysis. In short, this method seeks to estimate the trend in the data by approximating a 'line of best fit'. A regression analysis scrutinises the relationships between dependent variables and independent variables using a set of statistical processes. A dependent variable is a parameter for which statistics and scientific research make a prediction to test hypotheses, while an independent variable or predictor is a parameter which is used to base these predictions on. Since this research seeks to analyse the PBSS of São Paulo and Rio de Janeiro, the dependent variable is the number of trips generated per station. To predict the number of departing trips per station, the characteristics of the stations and the surroundings will be used as the independent variables. The relation between the dependent and independent variables was tested in both the linear and logistic approach. The goal was to examine whether the logistic approach generated better results than the commonly used linear approach. Table 17 in Appendix C gives the results and comparison between the two tested methods for all the included independent variables. The logistic regression results did not show improvement over the linear approach. Therefore, the author chose to continue utilizing linear regression, because this method is applied in many papers and is easier to facilitate when including many independent variables. The following paragraphs are explaining this method in detail.

3.3.3 Multivariate linear regression

In linear regression analysis, the approximated 'line of best fit' is a straight (linear) line. This method is the most straightforward and most commonly used to predict a certain variable. This regression method was applied to predict trip generation in various studies, including Engels (2019), Médard de Chardon (2016), Daddio (2012), Maurer (2012), Buck and Buehler (2012) and Rixey (2013). The objective is to predict one dependent variable, the average daily departing trips per station using various independent variables (Table 9). Multiple Linear Regression is used to meet this objective and has the following equation:

$$y_i = b_0 + \sum_{j=1}^{p} b_j x_{ij} + \varepsilon_i$$

y_i is the predicted value for the i^{th} observation, b_0 the regression intercept, b_j is the j^{th} predictor's regression slope, x_{ij} is the j^{th} predictor for the i^{th} observation and ε_i is a Gaussian error term.

The prediction models generate variables which could be used to validate and analyse the model. For each developed model, two tables are shown. The first table is a summary of the model, the determinative variable is the R^2, which is the measure for 'goodness-of-fit,' i.e. the explained variation of the complete model. The R^2 value for the prediction model of the average trips for BikeSampa is 0.419, which means that the variables explain 41.9 % of the total variation. This is expressed by the following equations:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

R^2 (or determination coefficient) is the measure for goodness-of-fit and the coefficient of multiple determination, with $0 \leq R^2 \leq 1$. The amount of variation in y_i that is explained by the linear relationships $x_{i1}, \dots, x_{ip}$.

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

The Sum-of-Squares Total (SST), where y_i is the value for the i^{th} observation and $\bar{y}$ is the overall mean of the observed data.

$$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

The Sum-of-Squares Regression (SSR), where $\hat{y}_i$ is the estimated average for the i^{th} observation and $\bar{y}$ is the overall mean of the observed data.

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

The Sum-of-Squares Error (SSE), where y_i is the value for the i^{th} observation and $\hat{y}_i$ represents the estimated average for the i^{th} observation.

Besides the R^2, two other values are depicted in the upcoming prediction models. The Beta coefficient and the level of significance. The Beta shows the linear trend between the dependent and the individual independent variable. A positive value Beta signifies a positive impact of the independent variable to the expected value. The significance, shown in the column next to Beta value, is the plausibility that a correlation is statistically significant and not based on coincidence. When the significance level is below a certain value, the null-hypothesis is rejected, meaning the variable is significantly contributing to a better prediction. The intended goal to solely accept models and variables with a significance level lower than 0.05 (5%) and still obtain significant prediction models, could not be accomplished. Therefore, the significance level of lower than 0.10 (10%) is also accepted in the models.

3.3.4 Data clustering

The data is further explored by clustering certain attributes. The main goal of the clustering was to improve the value of R^2. The table below gives a summary of the clustered attributes and the corresponding cluster method.

Clustered attribute	**Cluster method**
Days of the week	Distinguished between weekdays and weekend days
Land-use around the station	Distinguished between residential and commercial areas
The capacity of the station	K-means clustering (two clusters)
The station density	K-means clustering (two clusters)
Average departures	K-means clustering (two clusters for BikeSampa and three clusters for BikeRio

Table 3: Clustered attributed and the clustering methods

The first cluster distinguishes the week and weekend days, because the number of departures is expected to vary between the week and weekend. The data was aggregated to departures per station per day.

Accordingly, the week and weekend days were filtered out and put in the right cluster. The land-use was clustered by comparing the land-use per service area. A 'commercial' service area has most of its area assigned to offices and other commerce-related buildings and a 'residential' service area is predominantly located in residential areas. Initially, a third cluster with leisure areas was developed, but the number of stations included in this cluster was insufficient for testing in the prediction models. Therefore, these 'leisure' stations are added to the residential clusters of São Paulo and Rio de Janeiro. The three independent variables were clustered to further explore the influence of these variable to the number of departures. They have all been clustered using k-means clustering. In short, this method randomly selects, in this case, two data points (1) and assigns the other data points to the closest cluster (2). Then, it calculates the mean value of each cluster, these are the new cluster centres (3). Steps (2) and (3) are repeated iteratively such that the within-cluster sum of squares is minimized. Convergence is achieved when the maximum absolute coordinate change for any centre between the current and previous iteration is 0.000, i.e. the cluster means do not change anymore.

3.4 Interview with TemBici

So far, primarily quantitative data has been used to approach and answer the research questions. However, the operators' preferences and choices can also be decisive in the way the PBSS are performing. Therefore, a small qualitative part in the form of a questionnaire with one of the employees from TemBici is added. The goal of this questionnaire is to present and discuss some the results from the spatial inequality analysis and the prediction models. Furthermore, to gain more insight in the history of PBSS in both cities and the choices and decisions that are made for the current systems. Since there will only one respondent, the results are not included in the quantitative part of the research nor will be leading in the final conclusions. Nonetheless, the answers could still be a valuable contribution to the overall idea of how the PBSS in São Paulo and Rio de Janeiro operate and serve as a guidance the explain the model results.

3.5 Summary methodology

This chapter described the methods and strategies that were applied to meet the objective by answering the two research questions. The figure below gives a summary of what has been discussed in this chapter and how the different tasks are linked. The boxes surrounded by the blue dashed line outline research question one and the boxes inside the green dashed line outline research question two.

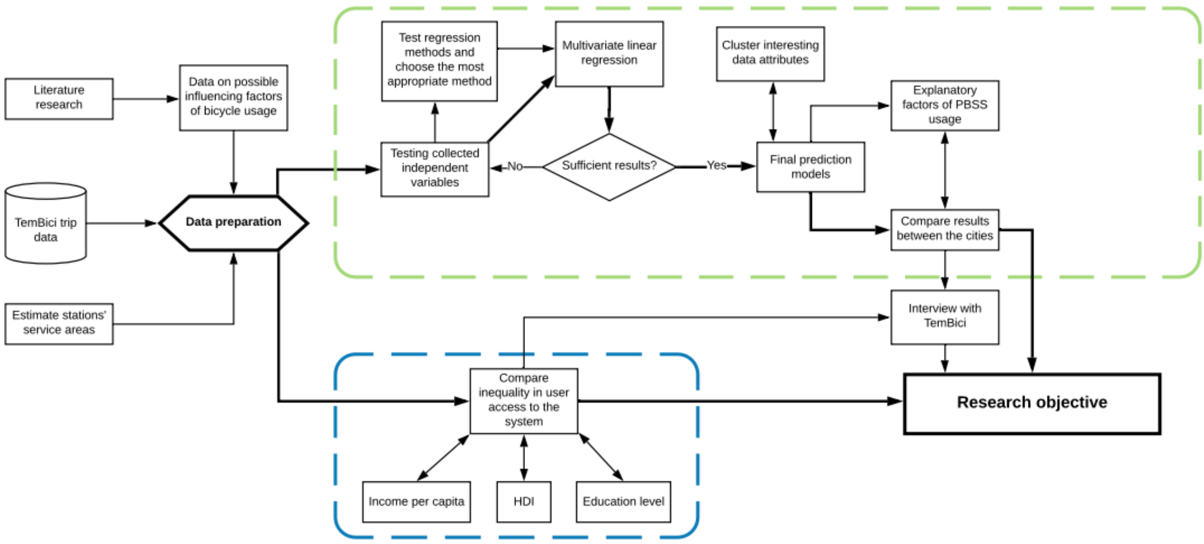


Figure 5: Conceptual research framework

3.5.1 Assumptions

The assumptions that were made during this research are summarized below

Assumptions regarding the service areas

- The boundary is set at a maximum walking distance of 10 minutes
- The potential user walks to the closest station when there happen to be more bicycle stations within 10 minutes walking
- Walking speed is set a 5 km/h
- The included variables are equally distributed over the available resolution, such that the averages per service area can be calculated

Assumptions regarding the trip data

- The first month with trip data is removed for each station

Assumptions regarding the prediction model

- All public transport stations are considered equal, meaning there is no distinction made between the number of passengers or connectivity with other stations.

4. Introduction to Brazil and the two researched cities

This chapter will provide background information about the two researched cities and the country Brazil in general. Since this thesis emphasises on social issues, the current challenges relating to social inequality and poverty that Brazil is facing are also introduced and discussed in this chapter. Furthermore, the upcoming paragraphs also argue how the potential utility of PBSS could help to reduce these problems. The final part introduces the two researched cities, São Paulo and Rio de Janeiro.

4.1 Social inequality in Brazil

Social inequality has been a continuous problem everywhere in the world. Brazil and many other Latin-American countries are known to have significant inequality related problems such as a high GINI-index, which describes income inequality (Gini, 1936). As a consequence, Brazil also experiences a relatively high number of homicides. In the case of Brazil, the GINI has consistently been among the highest in the world over the past decades. In 2017, Brazil had an index value of 0.47. Within Brazil the differences are notable, the GINI-index for the state of São Paulo is 0.41, which is notably lower than the countries' average. Nevertheless, the city sees large differences in income per neighbourhood or area of the city. Therefore, the municipality of São Paulo has an index value of 0.63. In the state of Rio de Janeiro, the GINI is slightly above the national average, reaching 0.49. The municipality copes with a similar index as São Paulo, 0.62 (Brasil, 2013). Like in Brazil and many other countries, areas located in the centre of the city have higher land-value than areas situated in the periphery. In the case of São Paulo, the stations of the bicycle share system from TemBici are located in neighbourhoods known for their high standards of living. Thus, the values of the socio-economic differences between the service area of the system and the rest of the city are expected to be different. The first research question will further explore the inequalities inside and between the cities by comparing the ethnic background of the inhabitants, the income per capita, human development and education level.

One advantage of PBSS is the improvement of the accessibility for people who do not have the financial resources to buy a car and live far away from other public transportation opportunities. In Brazil, many people live under these conditions. In 2017, 21% of the population of Brazil lived below the 'upper-middle-income poverty line' which is less than $5.50 per day (WorldBank, 2018). This problem is also the reality in São Paulo and Rio de Janeiro. Many people continue to live in favela's, which are often located on the edges of cities. The accessibility between the periphery to the city centre, where most of the jobs are found, is not always sufficient, especially when one does not possess a car. A well organised and connected PBSS could be a help for this population group.

4.2 A brief introduction of the two researched cities

BikeSampa is located in São Paulo, a city located in the eponymous state in the south of Brazil, pictured blue in Figure 6. As of 2018, the municipality has a population of 12.2 million. The metropolitan region of São Paulo has a population of 21.5 million, making it the largest urban agglomeration of South America and the 12th largest in the world. Despite not being the capital of Brazil, São Paulo is considered the 'financial capital of Brazil'. The city is located on a plateau and has an average elevation of 799 meters above sea level. The lowest points of the city are found around the Tietê river and its tributary, the Pinheiros river. The central part of the city is relatively flat, while the periphery has a hilly landscape. The average temperature is 19.2 ºC, and the annual precipitation is close to 1500 mm, making the climate humid subtropical (IBGE, 2018; INMET, 2019).

BikeRio is located in Rio de Janeiro, a city in the southeast of Brazil and pictured red in Figure 6. The city is the capital of the state of Rio de Janeiro and was the capital of Brazil until 1960. As of 2019, the municipality has 6.7 million inhabitants, the metropolitan area is home to 13 million people, making it both the second biggest municipality and metropolitan area of the country, after São Paulo. The city also has the second-biggest economy of Brazil with headquarters of various (state-owned) companies. However, the city is most famous for its beautiful landscapes and beaches, making it the most visited city in Brazil and one of the most visited cities in the southern hemisphere. The city is located at the Atlantic coast between the offshoots of the Serra do Mar mountain range. The city experiences a tropical savanna climate with hot and wet summers and warm, sunny winters. The average annual temperature is 23.8 ºC, and the average annual precipitation is 1070 mm (IBGE, 2018; INMET, 2019).

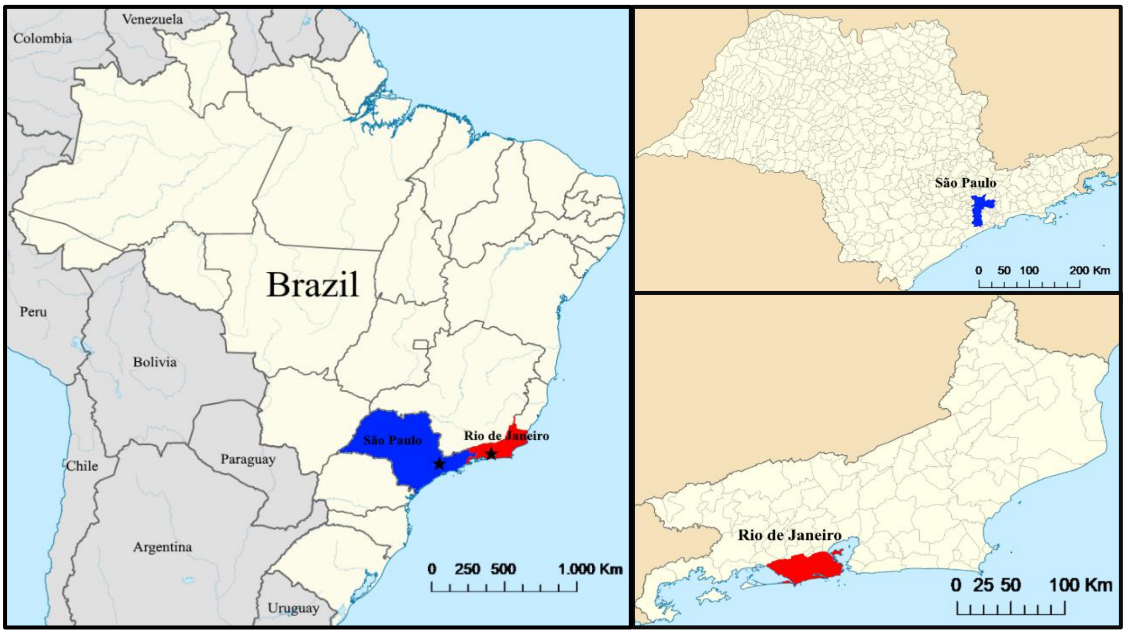


Figure 6: Location of the researched municipalities

5. Results

This chapter contains the results of the two research questions. At first, section 5.1 depicts the stations' service areas, which is an essential step for answering both research questions. Next, in section 5.2, research question one is answered by comparing population characteristics between the services areas and with the municipal averages. Concerning research question two, the results of the trip aggregation, final prediction models, the clustering analysis and a summary of the questionnaire are presented in section 5.3.

5.1 The refined service areas of the bicycle stations

5.1.1 The service areas of BikeSampa

The location of the stations has changed over the years. At the moment, all the stations are located east of the Pinheiros river and west of the historic city centre, stretching from Vila Leopoldina in the north to Brooklin in the south. The station data was updated in May 2019 and consisted of 235 bicycle stations, which means the system was unfinished at that time. The catchment area of BikeSampa comprises 4% (59 km^2) of the total area of the city. In general, the stations located at the edge of the network have a larger service area, especially on the east side of the network.

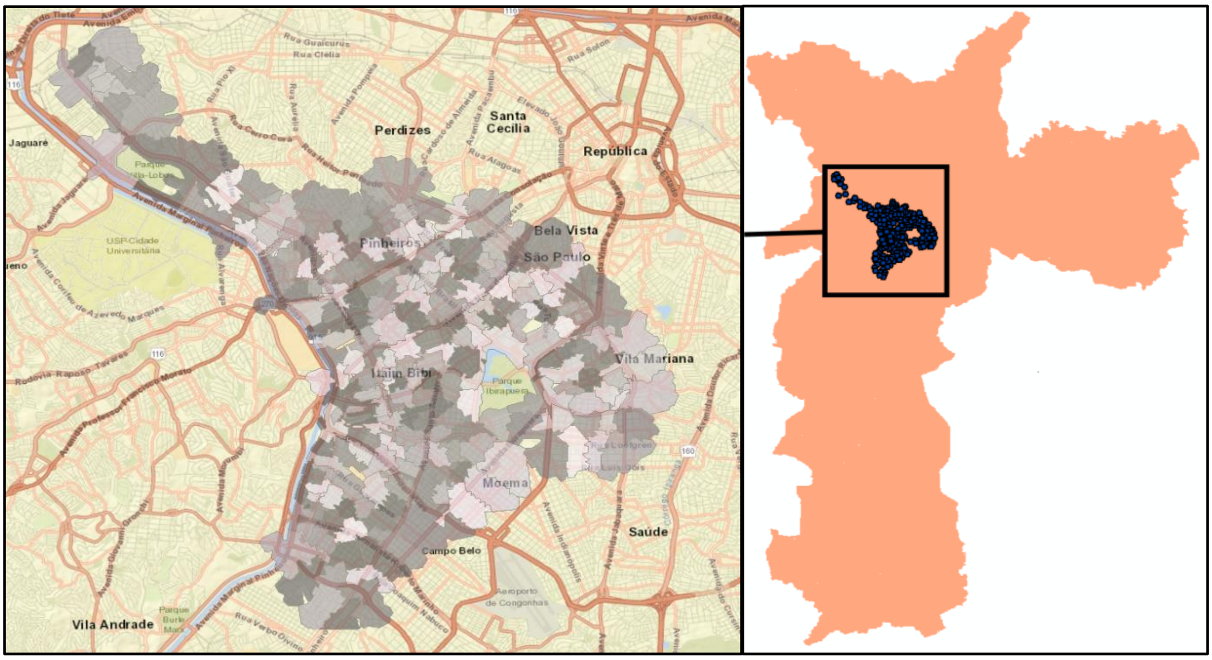


Figure 7: Location of the stations and the calculated service areas of BikeSampa

5.1.2 The service areas of BikeRio

BikeRio has currently 260 working stations used by 2600 bicycles. Unlike BikeSampa, the stations of BikeRio are spread over the city such that various groups of stations are found throughout the city, primarily along the coast as depicted in Figure 8. The catchment area of BikeRio was determined in the same way as BikeSampa, as described in section 3.1. The station data originates from May 2019, at that time, all the 260 stations were built. The first group of seven stations is located in the western edge of the municipality in a neighbourhood called Receiro dos Bandeirantes. This residential suburb is located approximately 40 kilometres from the historic centre of Rio. The second group, with around 30 bicycle stations in Barra da Tijuca, is located 30 kilometres from Rio's historic centre. The final group of stations can be considered the core BikeRio with more than 200 bicycle stations. It is crescent-shaped with bicycle stations from Tijuca/Grajaú through the historical and financial centre and moves around the hills to the south zone of the city, where the famous beaches of Flamengo, Copacabana and Ipanema are located.

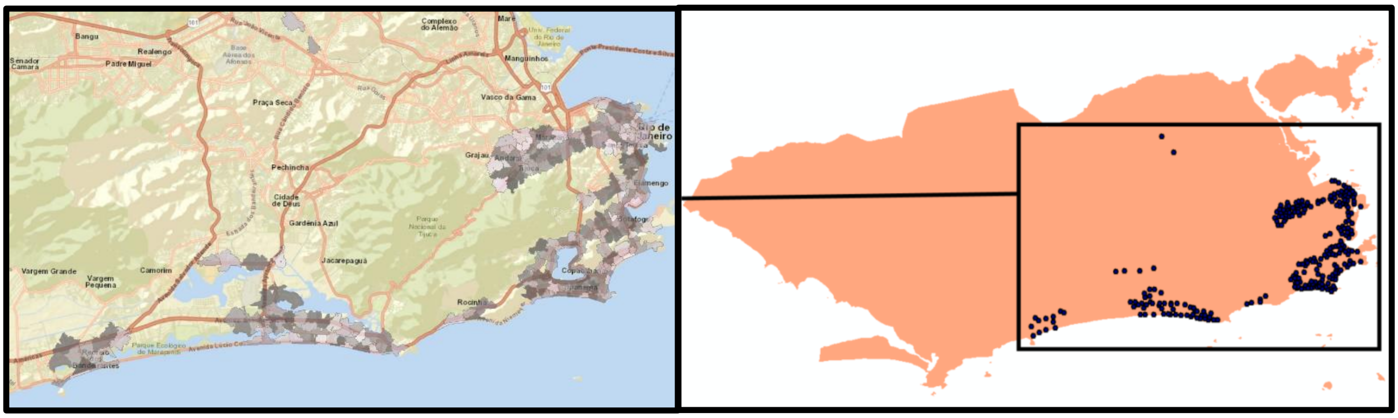


Figure 8: Location of the stations and the calculated service areas of BikeRio

5.2 Research question 1

The results of research question one: *'How is the spatial inequality in user access to the PBSS inside and between the systems of São Paulo and Rio de Janeiro?'* are presented in the upcoming paragraphs. First, the results of the comparison between the service area and the municipal averages are shown per analysed variable. The final two sections combine the results and answer the first research question.

5.2.1 Population characteristics of both researched cities

The population characteristics were calculated based on research performed by Pereira et al. (2020) that adopts hexagons containing information about the number of inhabitants and ethnic background of the population. Each hexagon has the same size, which is roughly 0.1 km^2. The table below is a summary of the ethnological data which was available per hexagon. This spatial resolution also allows one to calculate the distribution of the ethnic groups on a service area level. This data will also be used and implemented in the prediction models in the upcoming chapters. This section will solely elaborate on the distribution of ethnic groups on the municipal level and how it is related to the demographics in the service areas.

São Paulo	**Total population**	**White/Pardo**	**%**	**Black**	**%**	**Asian**	**%**	**Indigenous**	**%**
Municipality	11.044.837	6.697.926	60,64	3.978.327	36,02	243.939	2,21	11.554	0,10
Service area BikeSampa	601.181	515.904	85,82	53.032	8,82	30.730	5,11	396	0,07
Rio de Janeiro									
Municipality	6.119.399	3.134.615	51,22	2.796.055	45,69	43.343	0,71	5.647	0,09
Service area BikeRio	975.242	751.597	77,07	190.992	19,58	6.367	0,65	1.348	0,14

Table 4: Population and distribution of ethnic groups in São Paulo and Rio de Janeiro

São Paulo, with little over 11 million people living within the city limits, has a majority of white or pardo residents (61%), followed by black residents (36%) and two minorities which are Asians (2.2%) and a tiny number of Indigenous residents (0.1%). The service area of BikeSampa has a different population distribution. The vast majority is white or pardo (86%). With almost 9%, the proportion of black residents is four times smaller in the service area. The main reason for this change is that the majority of black people live in peripheral areas around the city. On the other hand, the proportion of Asians descendants in the service area is larger, mainly because the population of Asians is largely concentrated south of the city centre and BikeSampa covers part of this area. The table also points out that the majority of inhabitants does not live within a 10-minute walking distance from one of the stations. 5% of the population of São Paulo lives inside the service area of BikeSampa.

The demographic built-up of Rio de Janeiro is slightly different. Little over half of the population is white or pardo (51%) and close to half of the population is black (46%). Other minorities, such as Asian descendants and Indigenous people are less than 1% of the total population. Inside the service area of BikeRio, the majority of the population is white or pardo (77%) and the proportion of black residents is significantly smaller (20%). In terms of access to the bicycle share system, 16% of the total population lives within a 10-minute walking distance from a station.

The two main ethnic groups among the residents in both cities are black and white. Figure 9 reveals how these two ethnicities are distributed over the service areas. The figure is an addition to Table 4 and shows that almost all of BikeSampa's service area is inhabited by more than 80% of white or pardo and less than 20% of black residents. BikeRio's service areas have a less homogeneous distribution of the population. However, the majority of residents is also white or pardo and the highest concentrations can be found at the south coast and in the two suburbs 'Barra da Tijuca' and 'Recreio dos Bandeirantes'. The small number of service areas with predominantly black residents are located in the historic city centre, in the north-east part of the map. Later in this research, the percentages are included as independent variables for the prediction models, which might reveal more information if there exist differences in the use of the system between the population groups. Since the service areas have different population totals, the relative share of these groups inside the service area i.e. the percentage of black and white or pardo people living in the stations' service areas will be explored and is depicted as such in the figure below.

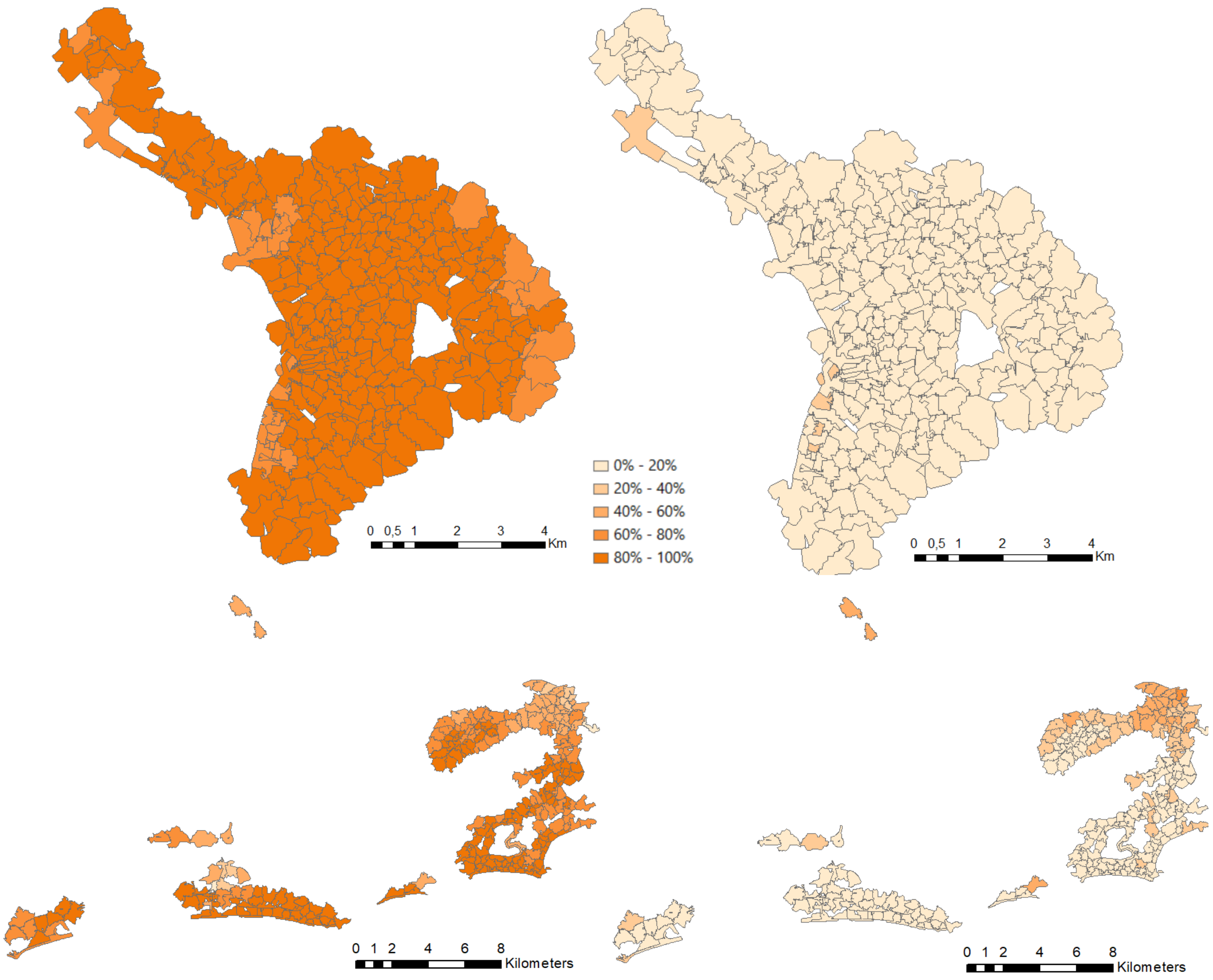


Figure 9: Black and white ethnicities in the service areas of BikeSampa and BikeRio. Upper left: White or Pardo population(%) in SP. Upper right: Black population (%) in SP. Lower left: White or Pardo population (%) in Rio. Lower right: Black population (%) in Rio.

5.2.2 Distribution of monthly income per capita

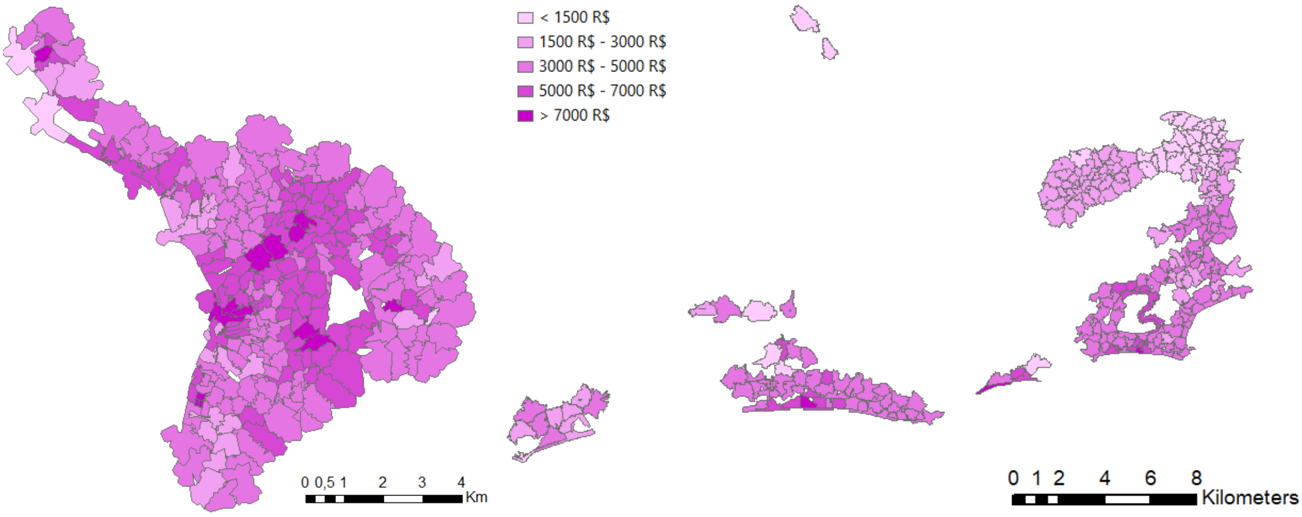


Figure 10: Distribution of monthly income over the service areas. Left: BikeSampa. Right: BikeRio

São Paulo: The PBSS of São Paulo is located in some of the wealthiest areas of the city. Figure 10 shows the distribution of income within the system and in Figure 11, the distribution of the average monthly income is visualised. The dark blue boxplot represents the service area and the light blue boxplot depicts the distribution of income for the municipality. The striking differences in income between the central part of the city, where the system is located, with the rest show how the high and low-income neighbourhoods are distributed within the city. Most of the inhabitants in the north, east and south part of the city have a personal income below 1500 R$ per month while in the central part of the city the income is regularly more than R$ 5000 per person. The boxplots for both areas display how diversified the two areas are. Table 5 reveals that the average income inside the service area lies around 4700 R$ whilst the average income for the whole municipality is 1500 R$.

Rio de Janeiro: The average monthly income in Rio de Janeiro vastly differs inside the service area (see: Figure 10), but also with the averages of the municipality, displayed in Figure 10. In brief, the parts located on the southern coast of the city have the highest average income and the city districts located inland receive the lowest average salary. A large share of the BikeRio stations are located along or close to the coast and in prosperous neighbourhoods. The northern located bicycle stations are serving the less wealthy areas. The large variation of income in the service area is also the result of the many favela's which are located on the hillslopes. The mountainous landscape of Rio allows favela development throughout the city. Therefore, favela's can still be found next to highly developed neighbourhoods. Notwithstanding, the difference in income between the service area and the rest of the city is significant since the income of the service area (R$ 3024) is almost three times higher than the municipal average (R$1189).

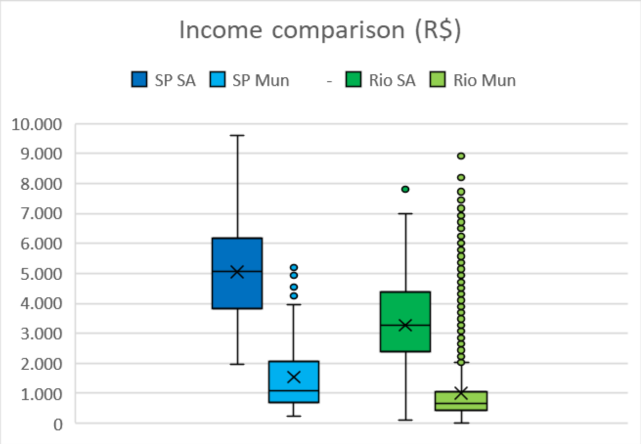


Figure 11: The distribution of income in Reais for the service areas and the municipal averages

Income (R$)	São Paulo		Rio de Janeiro	
	Service area	Municipality	Service area	Municipality
minimum	1046	245	99	1
maximum	11767	5310	7869	20358
mean	4706	1262	3024	1189
st dev	1526	1195	1510	1154
t-test	0.000		0,000	

Table 5: Descriptive statistics of the income

5.2.3 Distribution of human development

São Paulo: The Human Development Index (HDI) measures the quality of life by combining data on life expectancy, education and income. As of 2017, the HDI for the state of São Paulo is 0.826 and the municipality has an index of 0.843. The most recent data on HDI per neighbourhood originates from 2010 and therefore, shows lower index values. The data also explains that Brazil is still a developing country and that the quality of life is improving over the years. The differences in HDI maybe even more outstanding than the income differences (Figure 12). The index values for the service area are all above the 0.900. Such values are exclusively found in highly developed countries in the western world. The range of the boxplot for the municipality describes the significant variation of the human development inside the municipality. The small proportion of the inhabitants that enjoy the life standard of western European countries are outweighed by a larger group that stay around or below the Brazilian average. The service area boxplot shows noticeably little variation in the results meaning that the inequality in HDI is small, which is in contrast unequal division of HDI in the municipal area.

Rio de Janeiro: The large income differences between the service area and the rest of the city do not result in equally large differences in the HDI, which were found in São Paulo. The average values for HDI are 0.935 for the service area and 0.842 for the municipality. The most developed districts are located in the south zone around Lagoa Rodrigo de Freitas and have an HDI of at least 0.950. The south coast, in general, is highly developed with values of at least 0.900. The lowest HDI's in the service area are found close to the historic city centre in the hills of Santo Cristo, where some favelas are located.

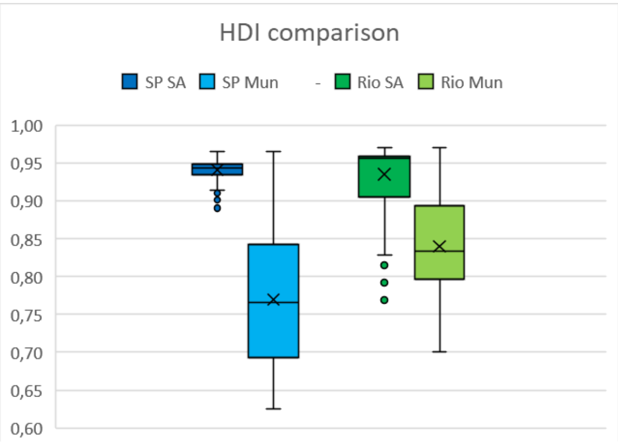


Figure 12: The distribution of HDI for the service areas and the municipal averages

HDI	São Paulo		Rio de Janeiro	
	Service area	Municipality	Service area	Municipality
minimum	0,890	0,625	0,769	0,700
maximum	0,970	0,965	0,970	0,970
mean	0,940	0,843	0,935	0,842
st dev	0,010	0,090	0,03	0,067
t-test	0,000		0,000	

Table 6: Descriptive statistics of the HDI

5.2.4 Education level São Paulo

The percentages of people with medium and superior education level are compared. In Brazil, a medium education is equal to a degree in high school and a superior education level means one has a university degree. The results are in the line of expectation; people from the service area have a higher education than the cities' average. On average, 62% of the population from the service area has a superior degree, compared with a city average 15%. The range for the medium educations is smaller, yet significant. 80% of the service area has a high school degree, and the cities' average is around 40%. One can say that the service area is also one of the highest educated areas in São Paulo. From the ten areas with the highest percentages of medium educated inhabitants are nine located inside the service area and one outside. The complete top ten areas with the highest percentage of a degree in superior education are located inside the service area.

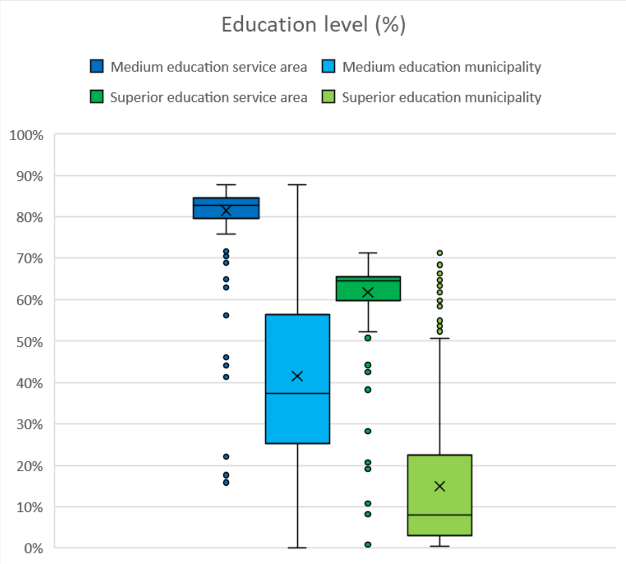


Figure 13: A comparison of the percentage of inhabitants with a degree in medium or superior education in São Paulo

	Medium education (%)		Superior education (%)	
	service area	municipality	service area	municipality
minimum	15,82	11,66	0,88	0,39
maximum	87,69	87,69	71,32	71,32
mean	81,50	41,50	61,73	14,98
st dev	7,10	18,70	8,83	15,83
t-test	0.000		0.000	

Table 7: Descriptive statistics on the level of education

5.2.5 Summary of the comparisons

São Paulo: Summarising, the results of the comparison between the municipality and the service area confirm the immense inequalities in social-economic characteristics. Inhabitants from the service area are the most developed, educated and have the highest income of any citizen in the municipality of São Paulo. Consequently, the t-tests show statistical differences between the service areas and the municipality for all the researched variables. It can be concluded that the operator has chosen to offer their service to inhabitants of the higher social classes. Unfortunately, there was no detailed user info available which could further analyse the social characteristics of the user of the system instead of the averages per service area. Various literature studies have pointed out that PBSS are more often located in wealthier neighbourhoods and that actual users are observed to be wealthier, male and white (Fishman, 2016; Ogilvie & Goodman, 2012). Although the last conclusion about the users cannot be validated due to a lack of data, BikeSampa users likely have similar characteristics. The inequality in user access could be one of the consequences of having a private system operator for a service that is expected to encourage social equity. The boxplots point out that, in contrast to the large variations in socio-economic characteristics in the municipality, the service areas show similar values for the tested variables.

Rio de Janeiro: The bicycle stations of BikeRio are mainly located along the coast. The city has a highly urbanised centre located in the east, but the mountains and lakes, which are mainly found in the central part of the municipality, do not allow urban development at every place. The result is a spread-out city with urban expansions in and around the mountains. Turning now to the comparison of the socio-economic characteristics, as explained in the previous paragraphs, it is clear that people that live within the boundaries of the service area are more privileged than residents outside the service area. This finding is strengthened by the t-tests, which show a strong statistical difference between all the tested variables. The differences in average income are particularly distressing. Although some bicycle stations are located in impoverished neighbourhoods, the average salary of a person living close to a bicycle station is nonetheless nearly three times the salary of a person that does not have the benefit of the bicycle share system close by. The city districts with the highest HDI have access to BikeRio and the least developed part of Rio, the western side of the city, is not covered by the system at all. This results in an uneven distribution of the human development in the city and thus between the service area and the rest of the municipality.

5.2.6 Concluding remarks on the spatial inequality in user access

The first part of this research sought to examine the selected locations of the stations of both BikeSampa and BikeRio and if there exist spatial inequality in user access. In São Paulo, 5.5% of the total municipal population lives inside the service area. BikeRio scores better, with almost 16% of the municipal population living inside a service area. Thus, the spatial inequality in user access is present since the majority of the inhabitants live outside the service area. The main objective was to evaluate the user characteristics of the service area and compare this with the rest of the city. Previous literature research reported that users of a PBSS are most likely to be wealthier, younger, white and male because the bike-sharing systems are principally located in wealthy neighbourhoods (Fishman, 2016). It became clear that stations are located in areas with predominantly white or pardo residents, while close to half of the average population in both municipalities is black.

The results of this comparison indicate a similar trend as found in the literature. The findings support the idea that certain population groups have easier access to the bike-sharing system than other groups. The unequal division is the most distinct in São Paulo since the stations of BikeSampa are solely located in wealthy neighbourhoods. BikeRio shows many resemblances with BikeSampa when comparing the service area averages and the municipal averages. However, the contrast is less extreme than in São Paulo. A significant number of stations are located in and around the historic centre of the city. This area is considerably less developed than the south zone, where the rest of the system is located. This raises the question if there are significant differences in bicycle use between the relatively wealthy and deprived neighbourhoods. The upcoming chapters will elaborate further on this matter and seek to answer this. Nevertheless, the majority of the station are located along the wealthy south-coast of the city. The size of the area that is served by both systems is different. The stations of BikeSampa form one cluster and the station are located close together, resulting in a relatively small average service area per station. The stations of BikeRio, on the other hand, form multiple clusters which are spread over a larger area of the city. As a consequence, the station's density is smaller, the service areas are larger and more inhabitants are living in one of the service areas.

Concluding, in both cities, user access to a TemBici station is far from equally distributed. People living in wealthy neighbourhoods have a much higher chance of living close to a TemBici station. The average inhabitant in one of the systems' service area earns a multiple of someone outside the service area. The placement of the stations has almost exclusively taken place in areas with a very high HDI (> 0.9). Though in Rio de Janeiro, the situation is slightly different, since the system does not exclusively operate in wealthy, developed neighbourhoods. The differences in socio-economic variables between the service areas give incentives to further study the trip generation of the stations located in the service areas and seek to explain why certain stations generate more trips than others. These results raise the question of what the influence of these tested variables is on the actual use of the system. For instance, if there exists a relationship between income or distribution of ethnicity in a service area and the number of departures that is generated by the station in this particular service area. The upcoming paragraphs will answer the second research question, where this question will also be addressed.

5.3 Research question 2

The results of the second research question: *'What are the factors which are explaining the station departures of the PBSS in São Paulo and Rio de Janeiro and what are the differences between these two cities?'* are presented in the following paragraphs. First, the aggregated trip data and some explanatory figures about the performance of BikeSampa and BikeRio are revealed. Afterwards, the definite included independent variables and the final models are exhibited. The results of the cluster analysis are presented next and the final part of this section provides a summary of the main findings from the questionnaire.

5.3.1 TemBici trip data

The second part of this research will mainly focus on building the prediction models for the average trip departures per station. The trip data is essential to achieve this. TemBici has kept track of individual trips since the beginning of 2018. Both the PBSS in Rio de Janeiro and São Paulo that are active today are relatively new, originating from the 20th of February and the 26th of January of 2018 respectively. This section will subsequently describe the data preparation, followed by some descriptive about how both systems have performed over the analysed period.

The number of stations of both systems has increased between April 2018 and September 2019. The graphs below show two curves per city, one is the total number of stations/trips that were recorded. The other curves show the number of stations/trips that were included in the model. Ideally, all the trips and station are included in the model, but because of two reasons, not all the stations and trips could be involved in the models. The first reason is that TemBici not only added stations but also removed stations that did not perform adequately and moved them to another location and gave them a new name. The station data, on which also the service areas are based, originates from April 2019. Therefore, any changes in the location of the station before and after this period influence the number of included station and trips. An example is found in the last months, which clearly shows a decrease in the number of included stations, even though the total number of stations remained stable. The second reason being that the first month of data for each station has been omitted because in many cases, the stations started operating in the middle or end of the months. Both systems show a similar pattern in the number of operating stations. Nevertheless, BikeRio has more operating stations for all the analysed months. Figure 14 also substantiates the reason why the first included months is April 2018. In the first three months of 2018, the number of operating stations increased gently to 43 in São Paulo and 71 in Rio de Janeiro. Upward from April, the number of operating stations grew faster and at the beginning of 2019, BikeRio reached the target value of 260 stations. In the meantime, BikeSampa reached the 260 stations in September of 2019.

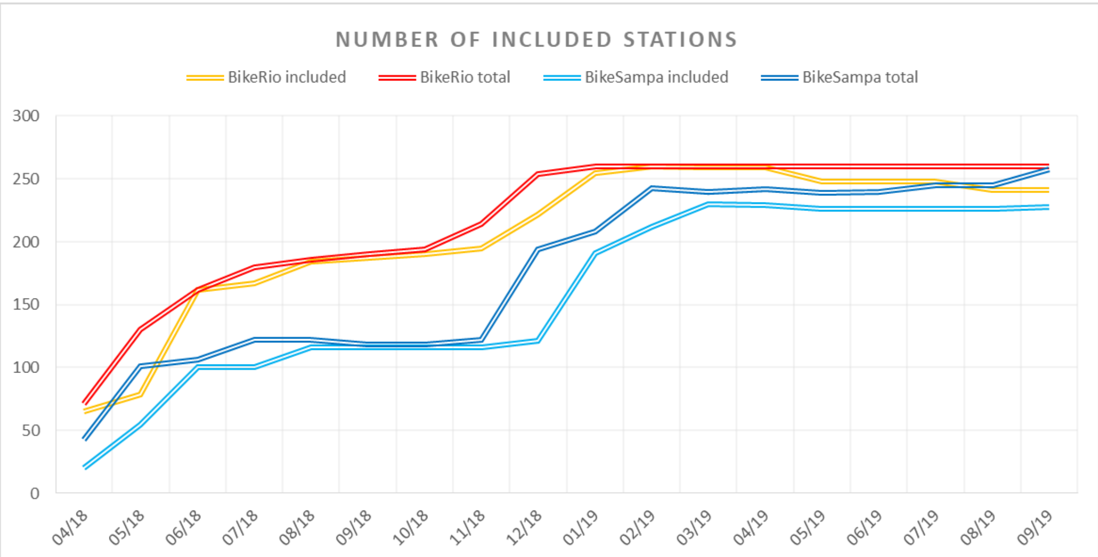


Figure 14: The number of included stations and the total number of stations for both cities over the months

The total number of trips per month (Figure 15) increases along with the rising amount of operating bicycle stations. The difference between both systems is visible as BikeRio, while having a comparable number of stations and bicycles, has nearly four times more monthly trips than BikeSampa. A reason for this large difference in the total number of trips can be partly addressed to the difference in the total population of the service area, which is 975241 inhabitants for Rio de Janeiro and 601181 for São Paulo. The prediction models, which will be described later in the thesis, are expected to provide a more thorough explanation for this considerable variation in the number of trips.

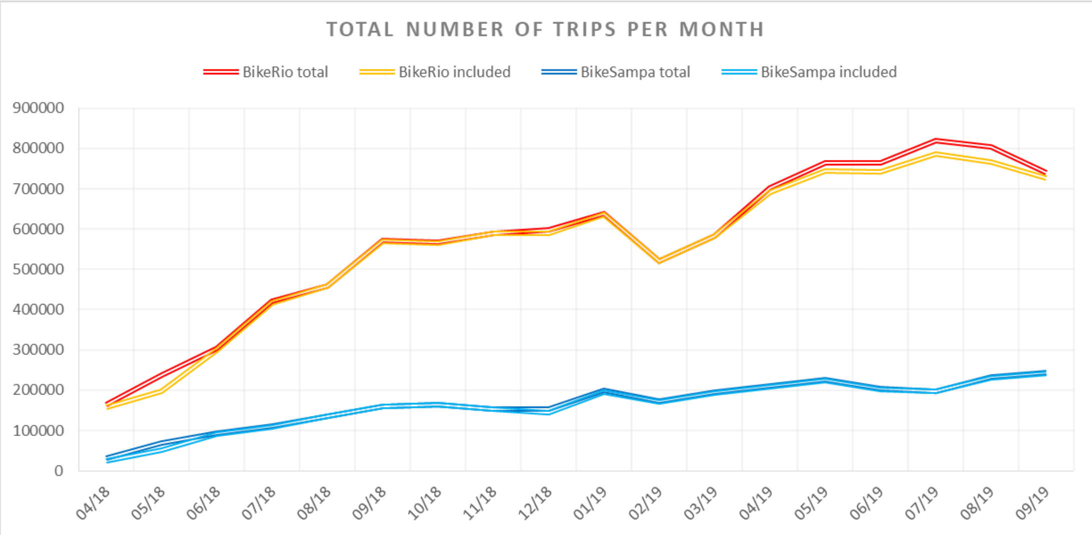


Figure 15: The number of included bicycle trips and the total number of trips for both cities over the analysed months

The total span of the analysed data is 18 months, from April 2018 until September 2019. However, many stations still had to be developed during this period and are not included for the full period. This is one of the limitations in the data analysis. The figure below is showing the number of months included in the models, cumulatively. The months with the sharpest inclination in the curves are most represented. For instance, the blue curve, defining São Paulo, increases from zero to little over a hundred stations between the months seven and nine. This indicates that half of the analysed stations for BikeSampa have only seven to nine months of data for the model. On average, BikeSampa has twelve months of data and a BikeRio station has almost fourteen months of data to work with.

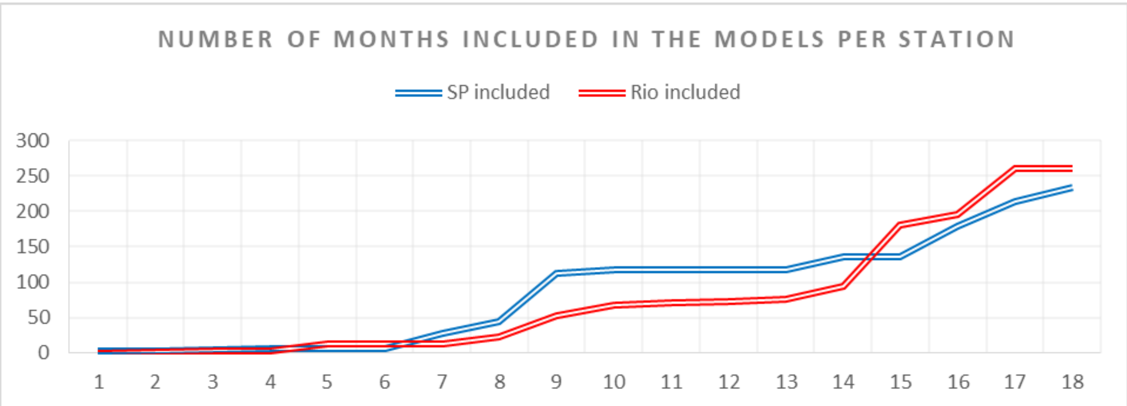


Figure 16: An overview of the number of months that the stations have been analysed

5.3.2 Daily trips per station

Figure 17 shows the average number of daily departing trips per station for both cities. A detailed version of the average daily departures per station can be found in Appendix B, as Figure 23 and Figure 24. The difference in total trips between the cites was already pointed out and this map shows the spatial distribution in the cities. Remarkable is that more than 200 BikeSampa station do not surpass 50 daily trips and large areas of the total service area show low numbers of daily trips. City districts Pinheiros, Itaim Bibi and Vila Olímpia are an exception and have some stations with more than 100 daily trips. These stations are predominantly located in business districts and next to the ciclovia of the Avenida Brigadeiro Faria Lima. The only station with more than 300 daily departures is located next to Faria Lima metro station, an important transit hub between the metro and train networks. BikeRio performs better and the average daily trips per station are more equally distributed over the stations. However, there exist spatial differences between the city's districts. The worst performing stations are principally located in Receiro dos Bandeirantes and Barra da Tijuca, both eastern suburban areas. The stations located in the central and south zone of the city generate the highest number of trips, especially the stations situated along the coast. A summary of the classification per city is given in the table.

Average daily trips per station	BikeRio	BikeSampa
less than 50	93	203
50 – 100	99	22
100 - 200	47	6
200 - 300	16	1
more than 300	5	1

Table 8: Average daily departures categorized

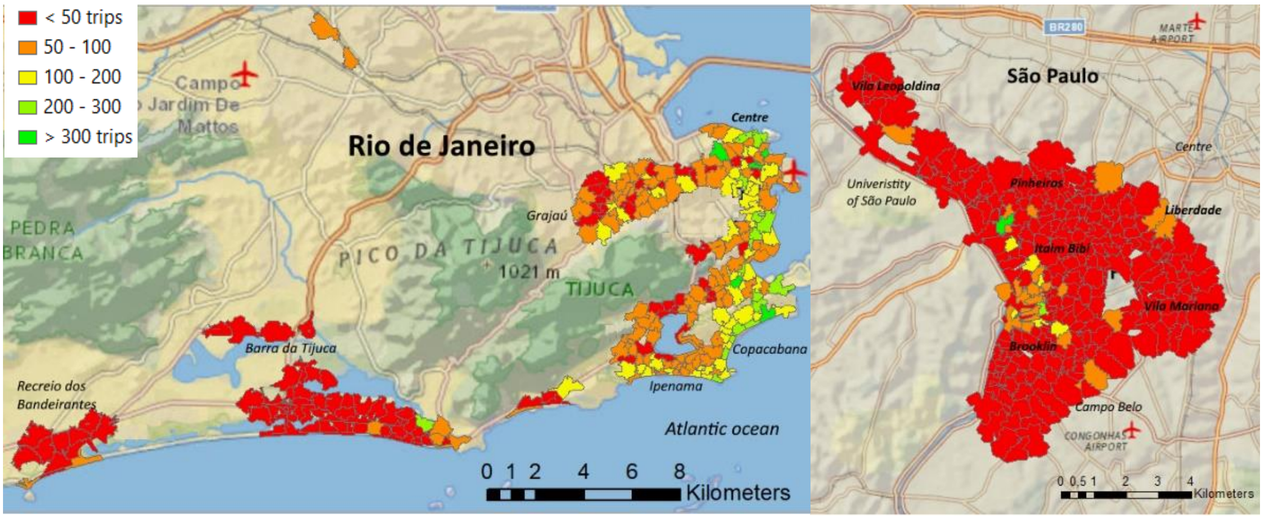


Figure 17: The average number of daily departing trips per service area

5.3.3 Collected and tested independent variables

Table 9 provides an overview of the collected variables (also schematically presented in Figure 4) which will be tested as predictors for the average number of departing trips. The initial intention was to include only variables that were available for both cities, but this was dropped because some variables were only accessible for one city could have a significant positive contribution to the model. The life expectancy and literacy rate are examples of variables that were exclusively available for Rio de Janeiro and the declivity and education level were only available for São Paulo. The rest of the included parameters are tested for both PBSS. In general, they can be divided into three groups. The first group stores variables related to socio-economic characteristics of the people living in the service area. The second set of variables is related to the bicycle station and provides information such as capacity, proximity to other stations and public transport facilities, but also the job accessibility by bicycle from a station. The last group of variables that will be tested are infrastructure-related. The table describing the variables consists of three columns; first, the variable name followed by a short description. The last columns give the expected relationship with the average number of departing trips. For instance, the positive relation of the variable 'Population' signifies that when the population of the service area increases, so will the expected number of departing trips.

Variable name	**Variable description**	**Exp**
Population	Number of inhabitants living inside the service area	+
Population density	Average population density in each service area (ppl/km^2)	+
HDI	Average HDI of the service area	±
White residents (%)	Percentage of white or pardo inhabitants living in each service area	+
Black residents (%)	Percentage of black inhabitants living in each service area	±
Income	Average monthly income per capita in Reais	±
Life expectancy (only Rio)	The average life expectancy of a person living in the service area	+
Literacy rate (only Rio)	The average percentage of inhabitants that can read and write (over 15 years of age)	+
Med_edu (only SP)	Percentage of the population of the service area with a degree in high school	±
Sup_edu (only SP)	Percentage of the population of the service area with a university degree	±
Declivity (only SP)	Scaled average declivity in each service area	-
Capacity	Total number of docks per station	+
Station density	Number of other stations in a radius of 1km^2	+
Infrastructure	Presence of any bicycle path within 50 meters of the station	+
Ciclovia	Presence of a ciclovia within 50 meters of the station	+
Ciclofaixa	Presence of a ciclofaixa within 50 meters of the station	+
Ciclorrota	Presence of a ciclorrota within 50 meters of the station	+
Faixa compartilhada (only Rio)	Presence of a faixa compartilhada within 50 meters of the station (São Paulo does not this type of bicycle infrastructure)	+
Metro 300	Presence of a metro/train station within 300 meters of the station	+
Metro 150	Presence of a metro/train station within 150 meters of the station	+
Cycling 15 min.	Accessible jobs within 15 minutes cycling (%)	+
Cycling 30 min.	Accessible jobs within 30 minutes cycling (%)	+
Cycling 60 min.	Accessible jobs within 60 minutes cycling (%)	+

Table 9: The included variables and the expected impact on the average departures

5.3.4 Final models

In this section, the final prediction models are developed and elaborated on. This chapter is divided into three parts; the first part discussed the results of the final prediction model. Next, the variables that were found to be divergent are investigated and the last part reveals the results of the cluster analysis. The final prediction models for both cities are depicted in Table 10. The simple linear regression indicates the predicting power of the individual variables for the dependent variable. This first assessment is a linear curve estimation with the independent variable on the x-axis and the dependent variable on the y-axis. In total, 8 of the 20 tested independent variables for BikeSampa are statistically significant without controlling for other variables. BikeRio has a better score with 13 of the 20 tested variables being individually significant for the predictor. The next column represents the multivariate linear regression, where all the independent variables, significant or not, are forced in the same prediction model. The final column for each city represents the final models, where the insignificant variables are removed and the prediction model is solely run with the significant variables. Ergo, the values of beta and the level of significance are slightly different for the reduced models. Eliminating the insignificant variables had a minimum negative influence on the determination coefficient. The statistical significance of a model or variable is demonstrated by means of an asterisk next to the significance. Three levels of statistical significance are distinguished, 10% (*), 5% (**) and 1% (***).

	BikeSampa (N=233)						**BikeRio (N=260)**					
	Simple linear regression		**Multivariate linear regression**		**Final model**		**Simple linear regression**		**Multivariate linear regression**		**Final model**	
Total model			R^2	sig.	R^2	sig.			R^2	sig.	R^2	sig.
			0,433	0,459	0,419	0,003***			0,467	0,322	0,445	0,000***
Independent variables	R^2	sig.	beta	sig.	beta	sig.	R^2	sig.	beta	sig.	beta	sig.
Population	0,009	0,152	0,069	0,429			0,040	0,001***	0,027	0,739		
Population density	0,007	0,195	0,026	0,746			0,092	0,000***	0,179	0,056*	0,187	0,001***
White/Pardo residents (%)	0,011	0,103	0,198	0,070*	0,214	0,018**	0,018	0,033**	-0,001	0,988		
Black residents (%)	0,033	0,005***	0,199	0,070*	0,188	0,035**	0,091	0,000***	0,356	0,002***	0,298	0,000***
HDI	0,000	0,852	-0,012	0,911			0,001	0,623	0,211	0,363		
Income	0,010	0,121	-0,029	0,658			0,071	0,000***	-0,445	0,017**	-0,341	0,001***
Literacy	-	-	-	-			0,000	0,936	-0,199	0,236		
Life expectancy	-	-	-	-			0,017	0,041**	-0,502	0,003***	0,411	0,000***
Med edu	0,000	0,776	0,034	0,802			-	-	-	-		
Sup edu	0,002	0,472	0,050	0,742			-	-	-	-		
Declivity	0,033	0,005***	-0,080	0,258			-	-	-	-		
Capacity	0,302	0,000***	0,425	0,000***	0,424	0,000***	0,136	0,010***	0,162	0,004***	0,171	0,001***
Stationdensity	0,035	0,004***	0,109	0,159	0,140	0,009***	0,025	0,010***	-0,035	0,638		
Any type of infrastructure	0,054	0,000***	0,328	0,000***			0,000	0,724	0,272	0,220	0,254	0,000***
Ciclovia	0,153	0,000***	0,254	0,000***	0,249	0,000***	0,002	0,508	-0,062	0,740		
Ciclofaixa	0,003	0,417	-0,026	0,644			0,000	0,966	-0,147	0,201	-0,160	0,004***
Faixa comparthilhada	-	-	-	-			0,002	0,523	0,048	0,667		
Ciclorrota	0,004	0,311	-0,005	0,928			0,009	0,122	-0,02	0,815		
Metro 300m	0,057	0,000***	0,016	0,812			0,086	0,000***	-0,030	0,667		
Metro 150m	0,049	0,001***	0,213	0,001***	0,249	0,000***	0,132	0,000***	0,303	0,000***	0,274	0,000***
Cycling 15 min.	0,000	0,734	0,086	0,385			0,112	0,000***	0,103	0,437		
Cycling 30 min.	0,009	0,160	-0,145	0,359			0,100	0,000***	-0,155	0,304		
Cycling 60 min.	0,007	0,192	0,087	0,505			0,096	0,000***	-0,028	0,859		

Table 10: The simple, multivariate and final linear regression models for BikeSampa and BikeRio

5.3.4.1 Prediction models BikeSampa

The first column shows the results of the simple linear regression. The variables are either insignificant or significant at 1% level. Six of the eight significant independent variables explain between 3% and 6% of the variation of the average departing trips. Aside from the percentage of black residents in the service area, no independent variables about socio-economic characteristics are proven to be significant. The two variables with the highest R-squared values are station capacity ($R^2 \approx 0{,}30$) and presence of a ciclovia ($R^2 \approx 0{,}15$). Consequently, the two variables are also significant in the full and final prediction models. The full prediction model, presented in the next column, does not consider the variables individually, and this gives some changes in the levels of significance, mainly because of correlations between the independent variables. For instance, the station density is not significant anymore. The full prediction model has an R-squared of 0,433, which is slightly higher than the final model that reached a determination coefficient of 0,419. This indicates that the majority of the variation in the number of departures is not explained by the model. The socio-economic variables such as income, HDI and education level continued to be insignificant predictors. A possible explanation is the lack of variations for these variables between the service areas in São Paulo. As mentioned in sections 5.2.2, 5.2.3 and 5.2.4, the socio-economic variables for the different catchment areas of the stations are compatible. However, the ethnicity of the inhabitants is significant both for the percentage of white and black inhabitants. Surprisingly both ethnic groups have positive B-values while they show a strong negative correlation of -0,813 (for all correlations between the variables, see Appendix A) which is rather counter-intuitive. A possible explanation is, again, the small variation for both variables when compared between the service areas. The maximum percentual range for the distinguished ethnicities lies around 25%. The evidently strongest predictor is the station capacity. The stations will be clustered by their capacity later this paper. Other important predictors include the presence of ciclovia's and a metro station close to the docking station. Surprisingly, the job accessibility index for the bicycle has been tested for fifteen, thirty and sixty minutes, but did not improve the prediction model significantly. The simple regression results for the job accessibility reveals that the individual influence on the predictor stays under 1% for all three indices. Aside from the job accessibility index, the average income and population density also show little to no relationship with the dependent variable. Section 5.3.5 elaborates further on these unexpected results.

5.3.4.2 Prediction models BikeRio

The results of the linear curve estimation for the data of Rio de Janeiro gave 13 significant individual predictions with values of R-squared ranging from 2% to 14%. In contrast to the results for São Paulo, the differences between the significant variables are smaller. The station capacity is the most powerful predictor, shortly followed by the presence of a metro station within 150 meters. The statistics relating to the characteristics of the population, such as income, population density and ethnicity per service area, explain the variation in the average trips significantly. Combining the variables in the full model gave an R-squared value of 46% and the reduced model has an R-squared value of 45%. Ergo the predicting power is only slightly higher than the model for BikeSampa. The higher diversity of socio-economic variables in the service areas in Rio de Janeiro fed the suggestion that the model for BikeRio would be more reliable. One of the strongest and most noteworthy predictors is the average monthly income. Service areas with a relatively high income generate a lower number of trips compared with their relatively low-income counterparts. Furthermore, service areas with higher percentages of black residents also generated more departing trips. The presence of a metro within 150 meters of a station does also increase the number of departing trips. Noteworthy is that for both Rio de Janeiro and São Paulo the presence of a metro station within 300 meters does not significantly improve the prediction, indicating that users prefer to walk only short distances between the metro/train and the bicycle stations. The variables relating to job accessibility by bicycle were not significant predictors in the full and final models, despite all of them being significant predictors in the simple regression. The influence of bicycle infrastructure on the number of departures little compared to BikeSampa, section 5.3.5 will discuss this relationship more in detail.

5.3.4.3 Summary and the main findings regarding the final model results

The two final prediction models show that both PBSS operate differently. In total, three independent variables are significant for both models, namely the percentage of black residents, the capacity of the station and the presence of a metro/train station within 150 meters of the station. The rest of the discussed values are solely significant for one of the two final models. The results of the simple regression reveal that the majority of the socio-economic related variables are correlated with the departures of BikeRio; this is not the case for the results of BikeSampa. The possible reason for this, the homogeneous character of the service areas of BikeSampa, is discussed in 5.2.6.

Some of the regression results were counter-intuitive and differed with the expectations of the literature study. Section 5.3.5 will further explore why these variables did not behave as expected. Those are the variables related to population density, income and cycling accessibility for the model of BikeSampa and the variables related to the bicycle infrastructure for BikeRio.

The models also provided valuable information about the possible reasons why the average use of BikeRio is more than three times higher than that of BikeSampa. It seems that income and ethnicity are important predictors for the average departures. The differences between the average income and ethnicity (especially the percentage of black inhabitants) between the two cities' service areas is significant (Figure 9). The wealthy and predominantly white service areas of São Paulo and Rio de Janeiro are also home to the stations that generate low numbers of trips. In contrast, the lower-income and coloured service areas found in Rio de Janeiro have higher use of the system.

5.3.5 Examination on opposed regression results

This section of the report elaborates on some specific independent variables from which the linear regression results were opposed to the findings in the literature. A more in-depth examination of these variables will provide more information and possible reasons why these specific variables do not behave as one would expect. The simple regression results are plotted to investigate possible correlations with the average daily departures. The logistic curve estimation is also added to compare with the linear relationship. Especially socio-economic variables tend to have a logistic rather than a linear relationship (Field, 2009). An example is the average income per capita; the bicycle use might be quite different between a person from the middle and lower income-class, yet between a millionaire and billionaire, the expected difference is smaller, despite the enormous variation in income. A second illustration where the logistic estimation might be a better fit than the linear estimation is the population density. A higher population density around a bicycle station is linked to more station departures, but to a certain extend. At some point, the population density will reach a level where the departures curve will flatten, because the number of docks stays equal and the system reaches its capacity. The upcoming paragraphs will determine whether these expectations also apply to Rio de Janeiro and São Paulo. The plots with the linear and logistic regression results for all the included variables are presented in Appendix C. The tables with the curve estimations for both the linear and logistics relationship do not indicate one superior method, because the results are similar. Aside from the plots, some relevant descriptive statistics regarding the discussed variables are visualized in Table 11 and Table 12.

5.3.5.1 São Paulo

Most of the socio-economic related variables were showing contradictory results in the form of no linear relationship. The variables related to the demographics and job accessibility were not confirmed as significant linear predictors. The author has chosen to further examine four of these unexpected results. A concise overview with some descriptive results and the linear and logistic curve estimations can be found in the table below.

					Linear regression		Logistic regression	
Independent variable	**Average**	**St. dev**	**minimum**	**maximum**	**R-squared**	**sig.**	**R-squared**	**sig.**
Income (R$)	4706	1533	1046	11767	0,010	0,121	0,003	0,414
Population density (ppl/km2)	10961	8991	344	46624	0,009	0,152	0,001	0,568
Job accessibility 15 min cycling	0,0472	0,0185	0,0022	0,0977	0,000	0,734	0,002	0,458
Job accessibility 60 min cycling	0,5552	0,0793	0,2498	0,6287	0,007	0,192	0,006	0,258

Table 11: Descriptive statistics of some insignificant predictors

The figures below plot the average daily departure (y-axis) against the average income and population density (x-axis) for BikeSampa. The literature on the relationship between income and trips generation did not find consensus. However, the simple regression results for BikeRio (Appendix C) show a clear negative impact from income on the daily departures. In São Paulo, there exists a fragile relationship, even considering that eight of the ten best stations with the most departures have a lower than average income, the determination coefficient for both curves lies below 1%, which is far from significant. Most of the service areas have an average income per capita between the R$3000 and R$6000 and the average is R$4700. The linear and logistic relationship between the population density and the average departure is, surprisingly non-existent. Stations with a low number of departing trips are also found in residential neighbourhoods, where the population density is higher. The plots show that the stations with a higher number of departures are located in service areas with a below-average ($\bar{x} = 10.961$) population density because they are located in commercial areas where relatively few people live. In general, the socio-economic related independent variables are not proven to be reliable predictors for the departures of BikeSampa. The apparent haphazardness of the relationship between the variables and the departures result in R-squared values below 1% for all the tested parameters, except the black population (R^2 = 3,3%) and white or pardo population (R^2 = 1,1%). Fitting a logistic curve for these variables did not result in improved values for R-squared.

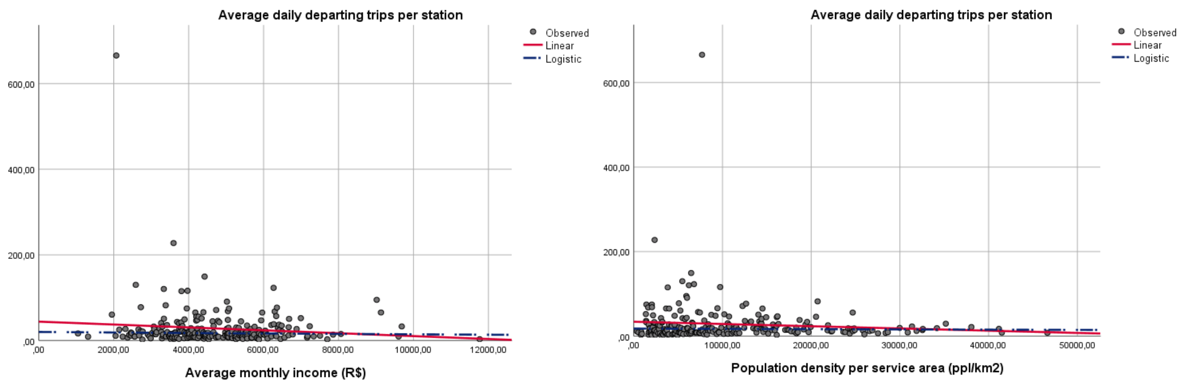


Figure 18 - Curve estimations for average departures of BikeSampa. Left: Average income (R$). Right: Population density

Another missing relationship that was rather surprising is that the job accessibility index by Pereira et al. (2020) does not seem to explain anything about the variation in the average departing trips. The R-squared for all three tested cycling times (15, 30 and 60 minutes) was not significant for both the linear and logistics curve estimations. This is in contrast with BikeRio, where the job accessibility index reached significant R-squared values for both linear (11%, 10%, 9,6%) and the logistic (12%, 16%, 20%). The stations in BikeSampa do not seem sensible for this index even while the index results per service area are certainly not unanimous (Figure 19)

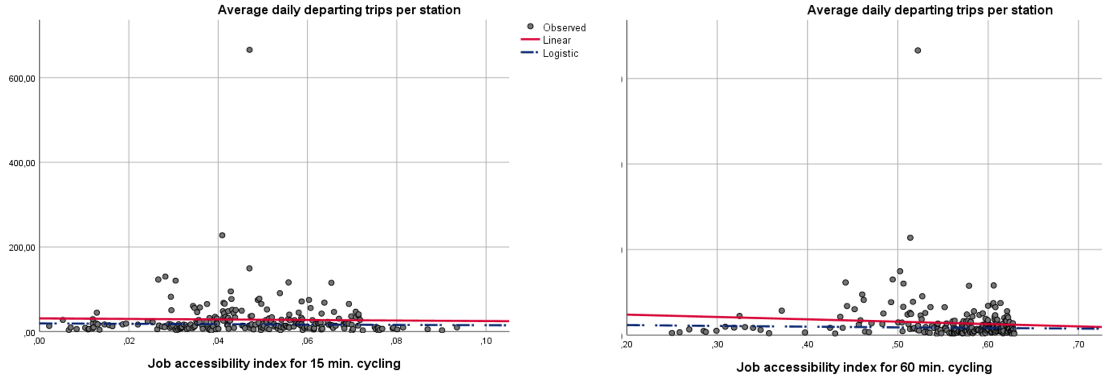


Figure 19 - Curve estimations for average departures of BikeSampa. Left: 15 min. cycling. Right: 60 min. cycling

5.3.5.2 Rio de Janeiro

In general, the simple regression results for BikeRio are in line with the findings from the literature study. The majority of the tested socio-economic variables are significant contributors for the departures for both the linear and logistic regression. The same can be concluded about the independent variables related to the station (e.g. station density, station capacity), the proximity to public transport stations and the job accessibility by bicycle. However, the variables related to the bicycle paths are not showing any linear neither logistic link with the departures. Table 12 provides a concise summary of the independent variable that are further examined in this section.

Independent variable	**Yes (%)**	**No (%)**	**Linear regression**		**Logistic regression**	
Ciclovia	21,5	78,5	0,002	0,508	0,001	0,638

Table 12: Descriptive statistics of the ciclovia

Interestingly, proximity to a ciclovia is a powerful predictor in the models for BikeSampa. A possible explanation for these differences can be argued when comparing the locations of the bicycle paths. The types of bicycle paths for both cities are mapped in Appendix B. Figure 21 and Figure 22 of the annex demonstrate that the infrastructure in Rio de Janeiro is not inferior to the infrastructure in São Paulo. However, a large share of the paths, especially the ciclovias, are located in the wealthy western suburbs, where the cycling rate is low. On the other hand, areas with a higher cycling rate, e.g. the historic centre of Rio de Janeiro, barely have bicycle paths. This means that the bicycle infrastructure is present, and a relatively large proportion of stations is connected to a path, but the location of the paths does not seem to match the location with higher cycling rates. It also indicates that, even with infrastructure, the cycling rates in the wealthy parts of the city are not increasing. The results of the curve estimation are depicted in the figure below and do not show any linear of logistic relationship between the presence of infrastructure and the average departures.

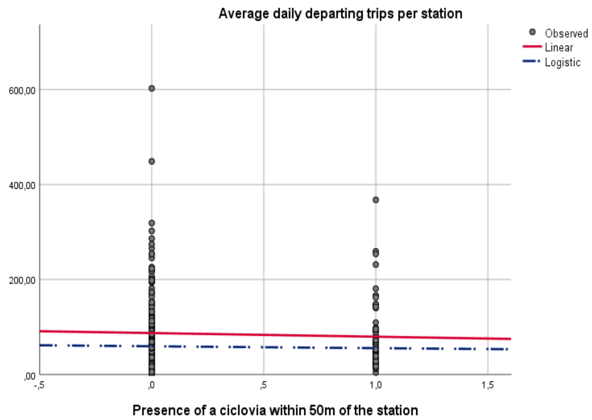


Figure 20: Curve estimations for average departures of BikeRio against presence of a ciclovia

5.3.5.3 Main findings from the validation of the opposed results

- The majority of the tested variables showed similar results for the linear and logistic curve estimations. Nonetheless, the variables related to the population characteristics tend more towards a logistic curve, and variables related to the proximity of public transport have a stronger linear relationship.
- The independent variables show either the expected relation or no relation with the dependent variable. There are no cases where the independent variables show significant relationships that are opposite to the expectations derived from the literature study.
- The research will continue to use multivariate linear regression to model the average departures because the results are similar and the multivariate linear regression is widely used in the analysed literature

5.3.6 Data clustering

The final reduced models, as presented in the previous paragraphs, include all the service areas and stations. The R-squared values of the two models did not surpass the value of 0.5, meaning that the majority of the variation remains unexplained. This section seeks to improve the values of R-squared by aggregating the data into clusters based on a certain independent variable or attribute. In total, five attributes have been clustered into smaller groups. The data is clustered based on the day of the week, land-use characteristics of the service area, station capacity, station density and average departing trips per station. The model results for the division between weekday and weekend are presented and discussed in the next paragraph. The summarized results of the other cluster analysis are shown in Table 14, where each clustered variable has a distinguished colour. The clustering did not have the desired results, and in many cases, the values for R-squared even deteriorated. Ergo, a more detailed analysis of the results has been relocated the Appendix D**.**

5.3.6.1 Prediction models for weekday and weekend

The trip data is separated by the day of the week. The departures from Monday to Friday are assigned to the weekdays' cluster and the departures on Saturday and Sunday are modelled in the weekend cluster. During the weekdays, one would expect a higher number of trips in stations located in commercial areas with a morning and evening peak. During the weekends, stations located close to parks are expected to have their peak in departures. The descriptive statistics of Table 13 show that the average number of departures is significantly higher during the weekdays, especially in São Paulo. The statistics also show the large differences in departures between the two cities. Even during the calmer weekend, the stations of BikeRio generate on average twice the number of trips than during a weekday in São Paulo. The next section discusses the results of the prediction models.

	BikeSampa				BikeRio			
	Weekdays		Weekend		Weekdays		Weekend	
Descriptives	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ
Average trips	31,98	67,17	16,03	21,00	83,21	81,69	66,62	63,84
Total model	R^2	Sig.	R^2	Sig.	R^2	Sig.	R^2	Sig.
	0,414	0,002***	0,180	0,082*	0,443	0,000***	0,362	0,000***
Independent variables	beta	sig.	beta	sig.	beta	sig.	beta	sig.
Population			0,112	0,073**				
Population density					0,113	0,040**	0,306	0,000***
White/Pardo residents (%)	0,212	0,019**						
Black residents (%)	0,199	0,026**			0,327	0,000***	0,266	0,008***
Income					-0,315	0,001***	-0,637	0,000***
Life expectancy					0,400	0,000***	0,570	0,000***
Med edu			0,117	0,060*				
Capacity	0,445	0,000***	0,143	0,030**	0,145	0,007***	0,274	0,000***
Station density	0,148	0,006***						
Any type of infrastructure					0,153	0,005***	0,170	0,004***
Ciclovia	0,205	0,000***	0,328	0,000***				
Faixa comparthilhada							0,191	0,004***
Metro 150m	0,219	0,000***	0,131	0,036**	0,309	0,000***		
Cycling 15 min.							-0,286	0,005***
Cycling 60 min.							-0,315	0,001***

Table 13: Prediction models for weekdays and weekend

Separating the weekdays and weekend days departures of BikeSampa demonstrated that the system is primarily used during the week, likely to commute. Consequently, the prediction model for the weekdays has the same independent variables as the final model of the city (Table 10). The determination coefficient is with 0,414 also nearly the same as the 0,419 of the final model. In contrast, the model for the weekend has a significantly lower value for R-squared of 0,180. This makes it precarious to draw conclusions about trip generation during the weekends because the vast majority of the variation in the data remains unexplained.

Both the prediction models for the weekdays and weekend for BikeRio have similar significant independent variables. The value for R-squared for the weekdays model is with 0,443 higher than the 0,362 that was achieved for the weekend days. However, still lower than the final prediction model for BikeRio, which reached and R-squared of 0.445 respectively. One interesting result is that the job accessibility index by bicycle negatively influences the weekend departures, which indicates that commercial and job-rich areas seem to be avoided by cyclists.

Summarizing, clustering the trip data to weekdays and weekend did not improve the prediction models, since all four new models have a lower value for R-squared. In both cases, the weekdays' models have a higher R-squared value than the weekend models. This is likely because more trips during the week are work-related and during the weekend, people use the system for recreational purposes. The tested independent variables are more descriptive for commuting trips rather than leisure-related trips. For instance, the proximity to public transport, bicycle infrastructure and job accessibility is included. On the other hand, the proximity to parks, the beach or landmarks, which could be explanatory indicators for weekend trips, are not included as independent variables in the analysis.

5.3.6.2 Summarized results of the other clusters

The summarized results of the other cluster analysis are shown in Table 14, where each clustered variable has a distinguished colour. The clustering did not have the desired results, and in many cases, the values for R-squared even deteriorated. Ergo, a more detailed analysis of the results have been relocated the Appendix D**.**

	BikeSampa					**BikeRio**					
	Land-use										
	Residential (N=160)		**Work (N=71)**			**Residential (N=219)**		**Work (N=28)**			
Final model	R^2	Sig.	R^2	Sig.		R^2	Sig.	R^2	Sig.		
	0,264	0,013**	0,558	0,040**		0,456	0,000***	0,565	0,057*		
	Capacity										
	Capacity < 20 (N=181)		**Capacity ≥ 20 (N=51)**			**Capacity < 20 (N=210)**		**Capacity ≥ 20 (N=50)**			
Final model	R^2	Sig.	R^2	Sig.		R^2	Sig.	R^2	Sig.		
	0,260	0,000***	0,596	0,096*		0,467	0,000***	0,293	0,107		
	Station density										
	Density < 7 (N=135)		**Density ≥ 7 (N=99)**			**Density < 7 (N=210)**		**Density ≥ 7 (N=50)**			
Final model	R^2	Sig.	R^2	Sig.		R^2	Sig.	R^2	Sig.		
	0,309	0,003***	0,684	0,038**		0,479	0,000***	0,327	0,096*		
	Average trips										
	Trips < 50 (N=203)		**50 < Trips < 150 (N=28)**			**Trips < 50 (N=93)**		**50 < Trips < 150 (N=129)**		**Trips > 150 (N=38)**	
Final model	R^2	Sig.	R^2	Sig.		R^2	Sig.	R^2	Sig.	R^2	Sig.
	0,293	0,001***	0,160	0,000***		0,449	0,031*	0,154	0,012**	0,413	0,073*

Table 14: Results of the cluster analysis

5.3.6.3 Main findings from the cluster analysis

- Against the expectations, most clusters had unimproved values for R-squared. Clustering by a significant predictor results, in most cases, in an inferior prediction model
- The homogeneity of the stations in São Paulo (population predominantly wealthy, white) could not be filtered out through the clustering. The vast majority of these stations were also residential, low capacity, had a low number of departure and had a low station density. Consequently, the R-squared values of these clusters remained low.
- The literature pointed out that higher density and capacity should result in more trips, this is line with the descriptive statistics of the model. Furthermore, it seems that these clusters of stations are 'easier' to predict since the determination coefficient is higher.

5.3.7 Summary of the interview

The questionnaire provided information about the operators' view of both systems. The questions are designed to obtain some qualitative data as an addition to the quantitative data that was collected for the models. The combination will help to achieve a broader perspective on how the PBSS of São Paulo and Rio de Janeiro are working in practice. The findings from the questionnaire might also be a valuable help to declare and justify the prediction models results and vice versa. The complete questionnaire with the given answer can be found in Appendix E. This section gives a summary of the main findings and interesting discoveries.

The decision process for the location of the stations goes in cooperation with the municipality, Banco Itaú and cycle activists of each city. The role of TemBici is to carry out the study and planning of the system. Banco Itaú was responsible for connecting TemBici with the cycle activists. Their opinion is highly valued during the planning process since they have the most experience and knowledge of cycling in the city. This also proves that the principal purpose is indeed to improve the overall mobility rather than being used as a means of advertisement for the bank. The role of the municipality is to evaluate the feasibility of implementing the proposed bicycle station, considering the current traffic and legislations.

The core values of TemBici in deciding the locations of the stations are discussed in question four (Appendix E). Evidently, the potential number of users for the system are important. The proximity to infrastructure is also decisive. The operator seeks to implement stations close to metro and train stations to integrate with the city's public transport system. Approximately 70% of the users do not live inside the coverage area of the system. Proximity to bicycle infrastructure is also important. TemBici pursues to build the stations around existing infrastructure. Thereby, the company seeks to find information about new infrastructure plans to be able to adapt to these future developments. Essential in the decision process is the concentration of activities around the station, such as the density of jobs, as they seek to build a dense network of stations in the core areas of the cities. Ultimately, the visibility of the station is considered very valuable since it might attract potential users. Therefore, the stations are often located next to the entrances of public transport and are visible form bicycle paths.

The main motives of TemBici are in line with the findings from the literature study. One difference is that TemBici does emphasise on the concentration of activities around a station and not so much on the population characteristics of the service areas. This explains why the stations are primarily located in neighbourhoods with a high job density, close to metro and train stations. Consequently, the land value is higher, and the residents of these areas are wealthier than average. Important included values for the operator that were measured in this study, such as proximity to public transport and the presence of bicycle infrastructure, are also significant predictors for both final models. Station density was only found a significant predictor for BikeSampa. Job accessibility per bicycle, which gives information about the job density, was not significant in any of the final models.

6. Discussion

The results of this study leave room for interpretation and therefore, have to be carefully evaluated. This section of the report consists of four parts. The first section gives an in-depth analysis of the results of the prediction models and how to interpret them. Next to this, the applicability of the results for other cities and areas is discussed. The last part of this chapter sheds light on the limitations and uses these to recommend changes and possibilities for further research on this subject.

6.1 Interpretation of the results

The expectation that the PBSS of São Paulo and Rio de Janeiro do not have similar characteristics turned out to be correct. Part of the explanation for the different model results is explained in the first research question where equality in user access is evaluated. It became evident that BikeSampa offers its service primarily to wealthy and white residents, and BikeRio serves a more divergent population. As a result, unlike the final model for BikeSampa, the prediction models for BikeRio had a stronger emphasis on the socio-economic variables, which were significant predictors. The hypothesis that the prediction models for Rio de Janeiro would reach higher values of R^2, because of the larger differences within the researched areas, was inconclusive. In fact, the values of the determination coefficient for most of the developed models, including the two final models, were below the value of 0.5. Part of the reason is that many of the described influential factors from the literature study could not be included, because they were not accessible, available in the right spatial disaggregation or not available at all. Furthermore, some available variables turned out to be insignificant contributors for the prediction. For instance, four out of the fourteen tested independent variables for both systems were not significant in the two final prediction models. The population of the service area, the average HDI, the proximity of a ciclorrota and having a metro station within 300 metres were not significant in both final models. The latter two variables are binary, and the majority of the stations did not meet the criteria, meaning the data was a bit fickle. A possible explanation for the HDI not being significant might be the small variation of this parameter between the service areas; this is especially the case in São Paulo. What is genuinely surprising is that the population of the service area does not improve the prediction. A higher population, and population density, was expected to generate more trips (J. Zhao et al., 2014). The first possible explanation for this counterintuitive result is that a substantial portion of the stations is located in commercial areas, with a high number of departing trips, yet low populations. However, this argument is invalidated since the population is still not a significant variable if the stations are divided by their land-use. The second more plausible explanation is that the share of the departing trips is minimum compared with the populations of these areas.

The majority of the literature about PBSS originates from countries in Asia, Europe and North America. The papers did not mention problems relating to societal inequalities such as criminality rates, violence against woman and other problems relating safety, at least not in the magnitude that these issues are present in Brazil and other Latin-American countries. Literature about countries that are more similar to Brazil (e.g. other Latin-American countries) was scarce, due to the relative novelty of such systems in that part of the world. Safety issues do negatively influence the use of bicycles in general and also create a climate which complicates the implementation of PBSS (Emond et al., 2009; Kahn et al., 2002). Moreover, finding quantitative data on crime-related actions per neighbourhood was unsuccessful, while these issues might play a role in the explanation of the number of station departures.

6.2 Evaluation of the used regression methods

The final prediction models were developed using multivariate linear regression. This method was also applied in the literature studies which developed similar models. However, some of the tested independent variables did not have the expected linear relationship, nor this relation was convincing. Therefore, the logistic curve estimation was also tested and expected to be a better predictor for primarily the socio-economic related independent variables, which are often non-linear. Appendix C provides an overview of the linear and logistic relationship for all the tested independent variables. A number of variables with counterintuitive results are discussed in more detail in section 6.1. It became clear that the linear and logistic line approximations were comparable. In general, the socio-economic variables showed slightly better R-squared values for the logistic approach, and the variables related to the (system) infrastructure had higher R-squared values for the linear curve estimation. Ergo, a change in approach was not considered necessary such that the final models remained linear.

The developed final models reached R-squared values between 0,40 and 0,45, respectively. The summary of comparable studies in Table 1 shows that some studies from the USA had better regression results. One of the reasons might be the difference in data quality, which was considered crucial in the analysed works. The data quality and resolution were not optimal in this research due to limitations in time and available data. Furthermore, the fact that the systems of BikeSampa and BikeRio were still being built and developed during this research did also not positively contribute to the data quality. The second reason for the lower determination coefficient in the final models might be the relatively high amount of included socio-economic related independent variables. In general, it is more complicated to find (linear)relationships using socio-economic data as independent variables, especially when considering averages of a service area, which assigns a large group of people with a single station. At last, the analysed cities in these studies are all located in the USA, which makes it precarious to compare the models one-on-one. On the contrary, the most extensive study from Médard de Chardon (2016) including 75 PBSS, did have similar values for R-squared. This study developed regression models with the TBD as the dependent variable. The study also included BikeSampa and BikeRio, but the individual models of the two cities were not presented in the report. However, the paper revealed that the R-squared of the final models lies between 0,42 and 0,49, which is comparable to the results in this thesis.

6.3 Applicability of the results

The R-squared values of the final models disclosed that the majority of the variation remained unexplained. Nevertheless, the results, i.e. the significant independent variables, still illustrate what is important for a PBSS in Brazil to generate departing trips from a station. In Brazil, the development of station-based PBSS is still in its infancy, since six cities are using such a service. The fact that Brazil has over 200 million inhabitants, of which 87% lives in urban areas gives a high potential number of users of a PBSS (CIA, 2020). A logically emerging question is how similar systems would perform in other Brazilian cities, which makes it interesting to look at the applicability of the developed prediction model. Since both cities have specific characteristics and model results, the suitability will depend on the urban shape of the city and the socio-economic characteristics of the inhabitants. Most of the urban development in Brazil is along the Atlantic coast, where most large- and medium-sized cities are located. In general, the coast-line is flat, which makes the development of bicycle infrastructure along the coast attractive. However, it is rather difficult to compare both Rio de Janeiro and São Paulo with other urban areas in Brazil. São Paulo, for instance, is the largest, wealthiest and most developed city of the country. The types of neighbourhoods that BikeSampa serves are not found in other cities. Areas with similar HDI, average income and division of ethnicity are uncommon in Brazil. Some cities such as Curitiba, Belo Horizonte and Brasilia have neighbourhoods with similar statistics, but the size of these neighbourhoods is most likely not big enough to sustain a whole PBSS. Moreover, it appeared that considering the average departures for theses specific stations of BikeSampa, the inhabitants of such neighbourhoods are not likely to use the system at all. Aside from these exceptions, the socio-economic characteristics of the service areas of BikeSampa are superlative other Brazilian cities. The average income of an inhabitant of São Paulo is three the average of an inhabitant of Salvador. In popular culture, São Paulo is called the 'New York City' of Brazil, because of the cosmopolitan character of the city and the large number of banks and other financial institutions in both cities. Rio de Janeiro shows more resemblances in terms of ethnological distribution with other Brazilian cities. However, the beautiful sceneries and famous landmarks has made Rio a popular destination for (international) tourists. Furthermore, as the old capital of the country, many important companies are still located in the city. These two reasons also make Rio de Janeiro a significantly wealthier Brazilian city, the average income per capita is two times higher than cities such as Salvador, Recife and Fortaleza (IBGE, 2016). The primary reason why the models for Rio de Janeiro are more applicable than the models for São Paulo is also shown in research question one. The characteristics or the service areas for BikeSampa are quite homogeneous and BikeRio's service areas have more variety in the analysed parameters. For instance, the average variation between the service area for the two primary ethnic groups is around 25 percentage points in São Paulo and 65 percentage points in Rio de Janeiro.

It is rather difficult to base conclusions regarding the applicability for other cities solely based on the results of the models for Rio and São Paulo. TemBici also operates in three smaller cities; Porto Alegre, Salvador and Recife with the number of inhabitants ranging from 1.5 million in Porto Alegre till 2.8 million in Salvador. The TemBici systems in these cities are smaller with the number of stations ranging from 40 till 80 and the number of operating bicycles from 400 till 800. Especially Recife and Salvador, which are both located at the coast in the north-east of Brazil, can be good examples for implementing a similar PBSS. The cities' size, climate and the socio-economic characteristics, i.e. the distribution of ethnic groups and a lower average income per capita, make Salvador and Recife interesting cases of how a PBSS operates in cities with these attributes. Accordingly, it would be interesting to further explore the trip data of these cities.

6.4 Limitations in the research

Some limitations and complications were faced during the course of this research. The main limitations are listed below and clarified with a brief elaboration.

- **The number of included stations:** Both PBSS were still being developed/built while performing this research. TemBici started building both systems in January of 2018 and finished building the desired 260 stations in February 2019 for BikeRio and September 2019 for BikeSampa. The research analysed a total of eighteen months, starting in April 2018 until September 2019 to develop the prediction model. As a result, the number of included stations/service areas remained low in the first months.

- **Developed stations were not always definite:** The operator also removed/replaced existing stations, which resulted in continuous station data. The latest available sheet with the stations and corresponding geographical coordinates was published in April 2019. Thus, any changes in the location of the stations before and after these months are not covered in the analysis. In some cases, stations appeared on the sheet and could disappear a few months later, which causes a distorted picture for the station, because some data for a certain period was missing.

- **Clusters with a low number of average departing trips.** A significant amount of stations, mostly located in São Paulo, had a small average number of departing trips. 203 of the 234 stations did not produce more than fifty daily departures, which means the stations were barely used. This made it complicated to develop reliable and significant prediction models because a small variation in the dependent variable impedes the chances of creating a distinctive model. Consequently, in the results, the clusters that embody the characteristics of these 'low-performing' stations (low capacity and located in a residential area) also have low values for R-squared and a few decisive independent variables. On the other hand, the described stations also have rather homogeneous independent variables. Remarkably, according to the literature, most users of a PBSS are wealthier, white and well-educated. The results for São Paulo are in sheer contrast with the findings from the literature since the underperforming neighbourhoods and service areas are primarily inhabited by white, wealthy and well-educated residents.

- **Comparing the models.** The initial intention was to build a model for São Paulo and implement the significant independent variables in the models for Rio de Janeiro and compare the results of the two cities. However, the cities' characteristics turned out to be quite different in most aspects, which made this objective unrealistic. Thereby assembling the exact same list of variables for both cities would mean that possible significant variables had to be left out. The emphasis changed from building an overarching model to developing separate models for both cities. Nonetheless, comparing the results and the significant variables of the separate models remained a part of the research. Accordingly, the process of developing the clusters was done joined, which resulted in equal clusters that could be compared.

- **Alteration of the dependent variable.** Initially, the idea was to compute a prediction model for the two performance metrics of a PBSS as described in the ITDP planning guide (2017). The Trips per Bike per Day (TBD) to estimate the cost-benefit ratio of the stations and the Trips per Resident to evaluate the market penetration of the stations. However, there was no available data on the number of available bicycles per station per timestamp. Therefore, a reliable prediction model for TBD could not be realized. The author chose to change the course of the thesis and model and decided to predict the average number of departing trips per station instead.

6.5 Recommendations for further research

- A number of limitations of this research were because the systems were still in the developing phase. The full potential of 260 operating stations was reached in April 2019 for BikeRio and September 2019 for BikeSampa, the last months included in the prediction models. The prediction models are likely to be more accurate when the network of stations is fixed. Therefore, performing similar research when the stations are definite will most certainly produce better prediction models.

- As mentioned in the discussion, the initial objective was to build prediction models of the Trips per Bike per Day (TBD) and the Trips per Resident (TPR). Unfortunately, there was no historical data available on the available number of bicycles at the station at a given time, and therefore it was impossible to calculate the TBD. The live station data is accessible and simply needs to be stored somewhere. Consequently, the prediction models for the TBD can be built, which can provide more specific information about the cost-benefit ratio of the systems.

- TemBici has the same type of PBSS in Salvador, Recife and Porto Alegre. The initial intention to include these cities did not work out due to time limitations. Future research that also includes these cities will give a broader view of how PBSS operate in Brazil. The author suggests using the models for BikeRio as leading because the cities show more affinity (in term of socio-economic characteristics) with Rio de Janeiro than with São Paulo. The results of the prediction models suggest that area with a lower income and a higher proportion of black residents generate more departing trips. The other cities, especially Salvador and Recife, have a notable lower average income and also have larger proportions of black residents. Ergo, it would be interesting to perform a similar analysis on these cities to evaluate if the findings from São Paulo and Rio de Janeiro also apply in these cases.

- The clustering of the independent variables did not improve the determination coefficients. However, there might be a spatial component that influences the variation, which has not been addressed in this study. The two western clusters of stations in BikeRio (see Figure 17) are generating few numbers of departures while clusters of stations in the city centre are achieving much higher cycling rates, this could be further explored by performing a spatial clustering on the data of BikeRio

7. Conclusions

This chapter presents the most important conclusions that could be derived from the results of this research. The structure of this chapter is as follows; first, the main conclusion from the two research questions are presented. The final paragraph of this chapter will recall the main research objective and draw the general conclusions of this thesis.

7.1 First research question

How is the spatial inequality in user access to the PBSS inside and between the systems of São Paulo and Rio de Janeiro?

- There are significant differences between the service areas and the municipal averages concerning the income per capita, HDI and education level. The inhabitants of the service areas are, wealthier, more developed and higher educated than the average resident of the city. These differences are specifically high in São Paulo. In Rio de Janeiro, several stations are located in 'middle-class' neighbourhoods, where the analysed variables are equivalent to the municipal averages.

- Both systems are primarily located in neighbourhoods with a majority of white or pardo residents, while the models show a positive relationship between the percentage of black residents with the number of departures per station. On average, the stations of BikeRio generate more departures than the stations of BikeSampa, and this could be partly explained of the higher number of black residents in the service area of the city. Section 5.2.1 shows that the relative proportion of residents from black ethnicities is significantly higher in the municipalities. The proportional differences in population ethnicities between the service areas and the municipal averages are significant. Therefore, there might be a mismatch between the areas in which the PBSS is located and the areas that can really benefit from the service.

7.2 Second research question

What are the factors which are explaining the station departures of the PBSS in São Paulo and Rio de Janeiro and what are the differences between these two cities?

- The main difference between the two systems is that the system of Rio de Janeiro produces more than three times the number of departures of São Paulo. Therefore, the prediction models of BikeRio explain more about which independent variables are important to bicycle stations that generate more trip and perform better.

- The service areas of BikeSampa had many similar characteristics with limited variation between the researched catchment areas. Particularly, the socio-economic parameters indicated that the service areas were almost entirely located in the wealthy neighbourhoods of the city. The homogeneity of the data made it difficult to develop reliable prediction models. As a result, many variables and especially the socio-economic relating parameters were not found significant predictors for the models. This also implies that the models for BikeSampa are not very applicable to other cities because the service areas of BikeSampa, and most of the service areas of BikeRio, are among the wealthiest and most developed places in Brazil.

- The determination coefficient of the final models for both cities was similar and remained below the 0,5, which means that most of the variation in the number of departures per station remains unexplained. Accordingly, not all the factors which are explaining the use of PBSS are included in the prediction models. Nevertheless, the included and tested independent variables provided explanations and interesting differences between the researched cities. In the case of

Rio de Janeiro, the presence of bicycle infrastructure did not significantly influence the number of departures, while in São Paulo, the presence of infrastructure improves the number of station departures. On the contrary, the tested socio-economic variables are significant predictors for BikeRio, whereas they are insignificant for BikeSampa.

- The models for BikeRio indicate that service areas with a lower average income generate more trips; the stations with the lowest number of departures are found in the wealthy parts of the city. Income is not a significant variable in the prediction models for BikeSampa, likely because even the service areas with the lowest average income still earn twice the municipal average. Therefore it cannot be excluded that a similar trend exists within the borders of São Paulo. (2)

- Proximity to metro and train is considered very important by the operator and it does indeed generate more trips. The distance between the bicycle station and the metro or train station is important in this matter. The final prediction models for both cities have the presence of a metro or train station within 150 meters as a significant predictor that positively influences the number of departures. However, having a metro or train station within 300 meters is not a significant predictor, indicating that only bicycle stations in the near perimeter are significantly benefitting from access and egress to public transport.

7.3 General conclusion

To examine the spatial inequality in user access to the Public Bike-Sharing Systems of São Paulo and Rio de Janeiro, investigate the possible factors that influence the average station departures in these systems and explain the differences between the two systems.

Summarizing, the user access to the stations is not equally distributed since relatively more people with white or pardo ethnicity live close to a bicycle station. Adding to that, the average service area is located in more developed areas, where wealthier and higher educated people live. In fact, the average income per capita in the service areas of BikeSampa and BikeRio is three times higher than the surrounding municipality. Interestingly, the developed prediction models found that percentage of black residents is a significant positive predictor for the number of departures. Furthermore, the final prediction model for BikeRio shows a strong negative relationship between average income in a service area and the number of station departures, which suggest that there might be a mismatch between the areas in which the PBSS is located and the areas that can really benefit from the service. An average BikeRio station generates more than three times the number of departures than a BikeSampa station, where many stations are producing less than fifty daily departures. Most of BikeRio's stations with a high number of departures are located in service areas with a relatively low average income and a relatively high percentage of black residents. The stations of BikeSampa that produced a relatively high number of departures were often located in commercial areas next to a ciclovia, which is, therefore, an important predictor in the models. In the case of BikeRio, also stations in residential neighbourhoods, often located along the coast, generated a high number of departures. Therefore, on the contrary with the models from BikeSampa, the socio-economic neighbourhood characteristics, such as income, population density and life expectancy are significant predictors for the models of BikeRio.

References

Antp. (2012). *Sistema de Informações da Mobilidade Urbana*.

Basch, C. H., Ethan, D., Rajan, S., Samayoa-Kozlowsky, S., & Basch, C. E. (2014). Helmet use among users of the Citi Bike bicycle-sharing program: a pilot study in New York City. *Journal of community health, 39*(3), 503-507.

Bikeitaú. (2019). Retrieved from https://bikeitau.com.br/

Brasil, A. (2013). Atlas do desenvolvimento humano no Brasil. *www. atlasbrasil. org. br/2013, consultado em, 2*(3), 2015.

Buck, D., & Buehler, R. (2012). *Bike lanes and other determinants of capital bikeshare trips.* Paper presented at the 91st Transportation research board annual meeting.

CIA. (2020). Urbanization rate per country. *The World Factbook*.

Corcoran, J., Li, T., Rohde, D., Charles-Edwards, E., & Mateo-Babiano, D. (2014). Spatio-temporal patterns of a Public Bicycle Sharing Program: the effect of weather and calendar events. *Journal of Transport Geography, 41*, 292-305.

Daddio, D. W. (2012). Maximizing bicycle sharing: an empirical analysis of capital bikeshare usage.

Daley, M., Rissel, C., & Lloyd, B. (2007). All dressed up and nowhere to go?: a qualitative research study of the barriers and enablers to cycling in Inner Sydney. *Road & Transport Research: A Journal of Australian and New Zealand Research and Practice, 16*(4), 42.

de Souza, F., Puello, L. L. P., Brussel, M., Orrico, R., & Van Maarseveen, M. (2017). Modelling the potential for cycling in access trips to bus, train and metro in Rio de Janeiro. *Transportation Research Part D: Transport and Environment, 56*, 55-67.

DeMaio, P. (2009). Bike-sharing: History, impacts, models of provision, and future. *Journal of Public Transportation, 12*(4), 3.

DETRAN. (2016). Retrieved from http://www.detran.rj.gov.br/index.asp

Duarte, F. (2016). Disassembling bike-sharing systems: Surveillance, advertising, and the social inequalities of a global technological assemblage. *Journal of Urban Technology, 23*(2), 103-115.

Emond, C. R., Tang, W., & Handy, S. L. (2009). Explaining gender difference in bicycling behavior. *Transportation Research Record, 2125*(1), 16-25.

Engels, J. (2019). *Forecasting spatial and temporal variations in OD-pairs: a case study from Sao Paulo.* University of Twente.

Field, A. (2009). Understanding statistics using SPSS: Sage Publications Inc.: Thousand Oaks, CA.

Fishman, E. (2016). Bikeshare: A review of recent literature. *Transport Reviews, 36*(1), 92-113.

Fishman, E., Washington, S., & Haworth, N. (2013). Bike share: a synthesis of the literature. *Transport Reviews, 33*(2), 148-165.

Fishman, E., Washington, S., Haworth, N., & Mazzei, A. (2014). Barriers to bikesharing: an analysis from Melbourne and Brisbane. *Journal of Transport Geography, 41*, 325-337.

Freitas, A. L. P., & Maciel, A. B. L. (2017a). Assessing cyclists' perceptions, motivations and behaviors: an exploratory study in Brazil. *Procedia engineering, 198*, 26-33.

Freitas, A. L. P., & Maciel, A. B. L. (2017b). Cycling in a Brazilian city. *Procedia engineering, 198*, 411-418.

Fuller, D., Gauvin, L., Kestens, Y., Daniel, M., Fournier, M., Morency, P., & Drouin, L. (2011). Use of a new public bicycle share program in Montreal, Canada. *American journal of preventive medicine, 41*(1), 80-83.

Garrard, J., Handy, S., & Dill, J. (2012). *Women and cycling* (Vol. 2012): MIT Press Cambridge, MA.

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

Gauthier, H., Christopher Kost, Shanshan Li, Clarisse Linke, Stephanie Lotshaw, Jacob Mason, Carlosfelipe Pardo, Clara Rasore, Bradley Schroeder, and Xavier Treviño. (2013). *The bike-share planning guide*. Retrieved from New York: https://itdpdotorg.wpengine.com/wp-content/uploads/2014/07/ITDP_Bike_Share_Planning_Guide.pdf

Gebhart, K., & Noland, R. B. (2014). The impact of weather conditions on bikeshare trips in Washington, DC. *Transportation, 41*(6), 1205-1225.

Gini, C. (1936). On the measure of concentration with special reference to income and statistics. *Colorado College Publication, General Series, 208*, 73-79.

Gutman, D. (2016). Will helmet law kill Seattle's new bike-share program? *The Seattle Times*.

Harms, L. (2007). Mobilität ethnischer Minderheiten in den Stadtgebieten der Niederlande.

Harms, L., Bertolini, L., & Te Brömmelstroet, M. (2014). Spatial and social variations in cycling patterns in a mature cycling country exploring differences and trends. *Journal of Transport & Health, 1*(4), 232-242.

IBGE. (2016). PIB por Município. Retrieved from https://www.ibge.gov.br/estatisticas/economicas/contas-nacionais/9088-produto-interno-bruto-dos-municipios.html?t=pib-por-municipio&c=4205407

IBGE. (2018). *Estimativas de População*. Retrieved from Rio de Janeiro: ftp://ftp.ibge.gov.br/Estimativas_de_Populacao/Estimativas_2018/

INMET. (2019). Gráficos Climatológicos Retrieved from http://www.inmet.gov.br/portal/index.php?r=clima/graficosClimaticos

Kahn, T., Besen, J., & Custódio, R. (2002). Pesquisa de Vitimização 2002 e Avaliação do Plano de Prevenção da Violência Urbana-PIAPS. *ILANUD, FIA-USP, Gabinete de Segurança Institucional*.

Kockelman, K. (1997). Travel behavior as function of accessibility, land use mixing, and land use balance: evidence from San Francisco Bay Area. *Transportation Research Record, 1607*(1), 116-125.

Krygsman, S., Dijst, M., & Arentze, T. (2004). Multimodal public transport: an analysis of travel time elements and the interconnectivity ratio. *Transport Policy, 11*(3), 265-275.

Marchuk, M., Shkompletova, A., & Boyarskaya, A. (2016). Bicycle sharing system.

Marmot, M., Allen, J., Goldblatt, P., Boyce, T., McNeish, D., Grady, M., & Geddes, I. (2010). The Marmot review: Fair society, healthy lives. *London: UCL*.

Marqués, R., Hernández-Herrador, V., Calvo-Salazar, M., & García-Cebrián, J. (2015). How infrastructure can promote cycling in cities: Lessons from Seville. *Research in Transportation Economics, 53,* 31-44.

Martens, K. (2004). The bicycle as a feedering mode: experiences from three European countries. *Transportation Research Part D: Transport and Environment, 9*(4), 281-294.

Mateo-Babiano, I., Bean, R., Corcoran, J., & Pojani, D. (2016). How does our natural and built environment affect the use of bicycle sharing? *Transportation Research Part A: Policy and Practice, 94*, 295-307.

Maurer, L. K. (2012). *Feasibility study for a bicycle sharing program in Sacramento, California*.

McBain, C., & Caulfield, B. (2018). An analysis of the factors influencing journey time variation in the Cork Public Bike System. *Sustainable Cities and Society, 42*, 641-649.

Médard de Chardon. (2016). *A geographical analysis of bicycle sharing systems.* (PhD), Université du Luxembourg, Luxembourg.

Médard de Chardon, Caruso, G., & Thomas, I. (2017). Bicycle sharing system 'success' determinants. *Transportation Research Part A: Policy and Practice, 100,* 202-214.

Muhs, C. D., & Clifton, K. J. (2016). Do characteristics of walkable environments support bicycling? Toward a definition of bicycle-supported development. *Journal of Transport and Land Use, 9*(2), 147-188.

Murphy, H. (2010). Dublin bikes: An investigation in the context of multimodal transport. *Dublin: MSc Sustainable Development, Dublin Institute of Technology*.

Nielsen, T. A. S., & Skov-Petersen, H. (2018). Bikeability–Urban structures supporting cycling. Effects of local, urban and regional scale urban form factors on cycling from home and workplace locations in Denmark. *Journal of Transport Geography, 69,* 36-44.

Ogilvie, F., & Goodman, A. (2012). Inequalities in usage of a public bicycle sharing scheme: socio-demographic predictors of uptake and usage of the London (UK) cycle hire scheme. *Preventive medicine, 55*(1), 40-45.

Parkin, J., Wardman, M., & Page, M. (2008). Estimation of the determinants of bicycle mode share for the journey to work using census data. *Transportation, 35*(1), 93-109.

Pereira, R. H., Braga, C. K. V., Serra, B., & Nadalin, V. G. (2020). Desigualdades socioespaciais de acesso a oportunidades nas cidades brasileiras–2019.

Pucher, J., & Buehler, R. (2008). Making cycling irresistible: lessons from the Netherlands, Denmark and Germany. *Transport Reviews, 28*(4), 495-528.

Rabello, R. C. (2019). *Sistema público de bicicletas compartilhadas: a disputa do espaço urbano.* Universidade de São Paulo.

Rixey, R. A. (2013). Station-level forecasting of bikesharing ridership: station network effects in three US systems. *Transportation Research Record, 2387*(1), 46-55.

Saelens, B. E., Sallis, J. F., & Frank, L. D. (2003). Environmental correlates of walking and cycling: findings from the transportation, urban design, and planning literatures. *Annals of behavioral medicine, 25*(2), 80-91.

Schneider, R. J. (2011). *Understanding sustainable transportation choices: Shifting routine automobile travel to walking and bicycling.* UC Berkeley.

Shaheen, S. A., Guzman, S., & Zhang, H. (2010). Bikesharing in Europe, the Americas, and Asia: past, present, and future. *Transportation Research Record, 2143*(1), 159-167.

Shaheen, S. A., Martin, E. W., Cohen, A. P., & Finson, R. S. (2012). *Public bikesharing In North America: early operator and user understanding*. Retrieved from

Smart, M. (2010). US immigrants and bicycling: Two-wheeled in Autopia. *Transport Policy, 17*(3), 153-159.

Steinbach, R., Green, J., Datta, J., & Edwards, P. (2011). Cycling and the city: a case study of how gendered, ethnic and class identities can shape healthy transport choices. *Social science & medicine, 72*(7), 1123-1130.

Stewart, S. K., Johnson, D. C., & Smith, W. P. (2013). Peer Reviewed: bringing bike share to a low-income community: lessons learned through community engagement, Minneapolis, Minnesota, 2011. *Preventing chronic disease, 10*.

Tsekeris, T., & Tsekeris, C. (2011). Demand forecasting in transport: Overview and modeling advances. *Economic research-Ekonomska istraživanja, 24*(1), 82-94.

WorldBank. (2018). POVERTY TREND (BY INTERNATIONAL STANDARDS). *Poverty & Equity Data Portal.* Retrieved from http://povertydata.worldbank.org/poverty/country/BRA

Zhao, J., Deng, W., & Song, Y. (2014). Ridership and effectiveness of bikesharing: The effects of urban features and system characteristics on daily use and turnover rate of public bikes in China. *Transport Policy, 35*, 253-264.

Zhao, P. (2014). The impact of the built environment on bicycle commuting: Evidence from Beijing. *Urban Studies, 51*(5), 1019-1037.

Appendixes

Appendix A – Pearson's correlations between the included dependent and independent variables

Pearson's correlations BikeSampa

	Average trips	Population	Popdensity	HDI	White (%)	Black (%)	Income	Med edu	Sup edu	Declivity	Capacity	Station density	Metro 300m	Metro 150m	Infrastructure	Ciclovia	Ciclofaixa	Ciclorota	Cycling 15 min.	Cycling 30 min.	Cycling 60 min.
Population	-0,085	-	0,596	-0,072	-0,079	-0,174	-0,120	0,015	-0,020	0,563	-0,068	-0,393	0,161	0,195	-0,072	-0,161	0,035	0,057	-0,239	-0,225	-0,204
Popdensity	-0,094	0,596	-	0,067	0,074	-0,304	0,020	0,187	0,153	0,401	-0,105	0,108	0,001	0,041	-0,119	-0,277	0,051	0,115	-0,126	-0,118	-0,092
HDI	-0,012	-0,072	0,067	-	0,328	-0,266	0,393	0,788	0,815	-0,014	-0,160	0,304	-0,143	-0,043	-0,081	-0,111	-0,042	0,059	-0,044	-0,129	-0,089
White (%)	-0,107	-0,079	0,074	0,328	-	-0,807	0,435	0,441	0,504	-0,095	-0,215	0,092	-0,214	-0,199	-0,164	-0,210	-0,056	0,064	-0,057	-0,061	-0,034
Black (%)	0,182	-0,174	-0,304	-0,266	-0,807	-	-0,321	-0,409	-0,481	-0,144	0,204	-0,001	0,171	0,108	0,118	0,239	-0,047	-0,068	0,088	0,065	0,088
Income	-0,102	-0,120	0,020	0,393	0,435	-0,321	-	0,290	0,444	-0,092	-0,219	0,143	-0,194	-0,143	0,024	-0,062	0,057	0,065	0,064	-0,030	-0,031
Med edu	-0,019	0,015	0,187	0,788	0,441	-0,409	0,290	-	0,892	0,061	-0,172	0,291	-0,057	-0,002	-0,109	-0,185	-0,027	0,095	-0,251	-0,257	-0,155
Sup edu	-0,047	-0,020	0,153	0,815	0,504	-0,481	0,444	0,892	-	0,024	-0,205	0,225	-0,074	-0,034	-0,113	-0,176	-0,004	0,049	-0,228	-0,289	-0,207
Declivity	-0,183	0,563	0,401	-0,014	-0,095	-0,144	-0,092	0,061	0,024	-	-0,145	-0,412	0,085	0,185	-0,024	-0,138	0,052	0,091	-0,082	-0,044	-0,028
Capacity	0,550	-0,068	-0,105	-0,160	-0,215	0,204	-0,219	-0,172	-0,205	-0,145	-	0,149	0,274	0,119	0,226	0,342	-0,021	-0,048	-0,059	-0,105	-0,157
Station density	0,188	-0,393	0,108	0,304	0,092	-0,001	0,143	0,291	0,225	-0,412	0,149	-	-0,128	-0,150	-0,031	-0,005	-0,063	0,025	0,056	-0,019	0,019
Metro 300m	0,232	0,161	0,001	-0,143	-0,214	0,171	-0,194	-0,057	-0,074	0,085	0,274	-0,128	-	0,482	0,085	0,081	0,053	-0,022	-0,218	-0,288	-0,292
Metro 150m	0,391	0,195	0,041	-0,043	-0,199	0,108	-0,143	-0,002	-0,034	0,185	0,119	-0,150	0,482	-	0,082	-0,015	0,207	-0,071	-0,095	-0,093	-0,087
Infrastructure	-0,053	-0,072	-0,119	-0,081	-0,164	0,118	0,024	-0,109	-0,113	-0,024	0,226	-0,031	0,085	0,082	-	0,621	0,481	0,394	-0,031	-0,021	-0,012
Ciclovia	-0,067	-0,161	-0,277	-0,111	-0,210	0,239	-0,062	-0,185	-0,176	-0,138	0,342	-0,005	0,081	-0,015	0,621	-	-0,141	-0,116	0,092	0,049	0,010
Ciclofaixa	0,239	0,035	0,051	-0,042	-0,056	-0,047	0,057	-0,027	-0,004	0,052	-0,021	-0,063	0,053	0,207	0,481	-0,141	-	-0,089	-0,067	-0,047	-0,011
Ciclorrota	0,221	0,057	0,115	0,059	0,064	-0,068	0,065	0,095	0,049	0,091	-0,048	0,025	-0,022	-0,071	0,394	-0,116	-0,089	-	-0,108	-0,053	-0,024
Cycling 15 min.	-0,022	-0,239	-0,126	-0,044	-0,057	0,088	0,064	-0,251	-0,228	-0,082	-0,059	0,056	-0,218	-0,095	-0,031	0,092	-0,067	-0,108	-	0,827	0,733
Cycling 30 min.	-0,092	-0,225	-0,118	-0,129	-0,061	0,065	-0,030	-0,257	-0,289	-0,044	-0,105	-0,019	-0,288	-0,093	-0,021	0,049	-0,047	-0,053	0,827	-	0,904
Cycling 60 min.	-0,086	-0,204	-0,092	-0,089	-0,034	0,088	-0,031	-0,155	-0,207	-0,028	-0,157	0,019	-0,292	-0,087	-0,012	0,010	-0,011	-0,024	0,733	0,904	-

Table 15: Pearson's correlations between the variables of BikeSampa

Pearson's correlations BikeRio

	Average trips		Population	Popdensity	HDI	White (%)	Black (%)	Income	Life expectancy	Literacy	Capacity	Station density	Metro 300m	Metro 150m	Infrastructure	Ciclovia	Ciclofaixa	Faixa compartilhada	Via compartilhada	Cycling 15 min.	Cycling 30 min.	Cycling 60 min.
Population	0,200		-	0,693	0,001	0,118	0,051	-0,160	-0,100	-0,061	0,015	-0,085	0,147	0,144	-0,091	-0,176	0,074	-0,121	0,176	-0,072	0,122	0,268
Popdensity	0,303		0,693	-	0,201	0,307	-0,114	-0,071	0,132	0,107	0,155	0,305	0,195	0,217	-0,066	-0,176	0,142	-0,132	0,194	-0,033	0,188	0,368
HDI	-0,031		0,001	0,201	-	0,571	-0,713	0,788	0,893	0,868	-0,005	0,123	-0,064	-0,051	0,158	0,140	-0,018	0,062	0,049	-0,437	-0,355	-0,266
White (%)	-0,133		0,118	0,307	0,571	-	-0,685	0,501	0,364	0,340	-0,020	0,073	-0,077	-0,032	0,265	0,283	0,048	-0,049	0,069	-0,565	-0,433	-0,280
Black (%)	0,301		0,051	-0,114	-0,713	-0,685	-	-0,761	-0,455	-0,453	0,134	0,048	0,161	0,078	-0,347	-0,280	-0,008	-0,185	-0,049	0,719	0,623	0,467
Income	-0,267		-0,160	-0,071	0,788	0,501	-0,761	-	0,638	0,686	-0,041	-0,130	-0,150	-0,100	0,236	0,210	-0,152	0,294	-0,026	-0,661	-0,635	-0,642
Life expectancy	0,118		-0,100	0,132	0,893	0,364	-0,455	0,638	-	0,880	0,090	0,249	-0,020	-0,028	0,044	0,057	-0,058	0,050	0,001	-0,134	-0,125	-0,120
Literacy	-0,005		-0,061	0,107	0,868	0,340	-0,453	0,686	0,880	-	0,043	0,202	-0,033	-0,041	0,013	-0,008	-0,001	0,060	-0,046	-0,182	-0,099	-0,066
Capacity	0,369		0,015	0,155	-0,005	-0,020	0,134	-0,041	0,090	0,043	-	0,071	0,271	0,269	0,067	0,005	0,056	0,098	0,028	0,104	0,116	0,151
Station density	0,159		-0,085	0,305	0,123	0,073	0,048	-0,130	0,249	0,202	0,071	-	0,214	0,112	-0,195	-0,270	0,170	-0,121	-0,003	0,416	0,382	0,435
Metro 300m	-0,022		0,147	0,195	-0,064	-0,077	0,161	-0,150	-0,020	-0,033	0,271	0,214	-	0,640	-0,135	-0,179	0,127	-0,101	-0,040	0,203	0,254	0,203
Metro 150m	-0,041		0,144	0,217	-0,051	-0,032	0,078	-0,100	-0,028	-0,041	0,269	0,112	0,640	-	-0,150	-0,147	0,040	-0,064	-0,059	0,139	0,180	0,153
Infrastructure	0,000		-0,091	-0,066	0,158	0,265	-0,347	0,236	0,044	0,013	0,067	-0,195	-0,135	-0,150	-	0,696	0,362	0,305	0,279	-0,313	-0,275	-0,233
Ciclovia	0,096		-0,176	-0,176	0,140	0,283	-0,280	0,210	0,057	-0,008	0,005	-0,270	-0,179	-0,147	0,696	-	-0,106	-0,120	-0,017	-0,296	-0,385	-0,409
Ciclofaixa	-0,040		0,074	0,142	-0,018	0,048	-0,008	-0,152	-0,058	-0,001	0,056	0,170	0,127	0,040	0,362	-0,106	-	-0,063	0,018	0,070	0,203	0,312
Faixa compartilhada	0,293		-0,121	-0,132	0,062	-0,049	-0,185	0,294	0,050	0,060	0,098	-0,121	-0,101	-0,064	0,305	-0,120	-0,063	-	-0,048	-0,158	-0,069	-0,063
Via compartilhada	0,364		0,176	0,194	0,049	0,069	-0,049	-0,026	0,001	-0,046	0,028	-0,003	-0,040	-0,059	0,279	-0,017	0,018	-0,048	-	-0,068	-0,075	-0,034
Cycling 15 min.	0,335		-0,072	-0,033	-0,437	-0,565	0,719	-0,661	-0,134	-0,182	0,104	0,416	0,203	0,139	-0,313	-0,296	0,070	-0,158	-0,068	-	0,822	0,641
Cycling 30 min.	0,317		0,122	0,188	-0,355	-0,433	0,623	-0,635	-0,125	-0,099	0,116	0,382	0,254	0,180	-0,275	-0,385	0,203	-0,069	-0,075	0,822	-	0,855
Cycling 60 min.	0,309		0,268	0,368	-0,266	-0,280	0,467	-0,642	-0,120	-0,066	0,151	0,435	0,203	0,153	-0,233	-0,409	0,312	-0,063	-0,034	0,641	0,855	-

Table 16: Pearson's correlations between the variables of BikeRio

Appendix B – Additional maps of the service areas

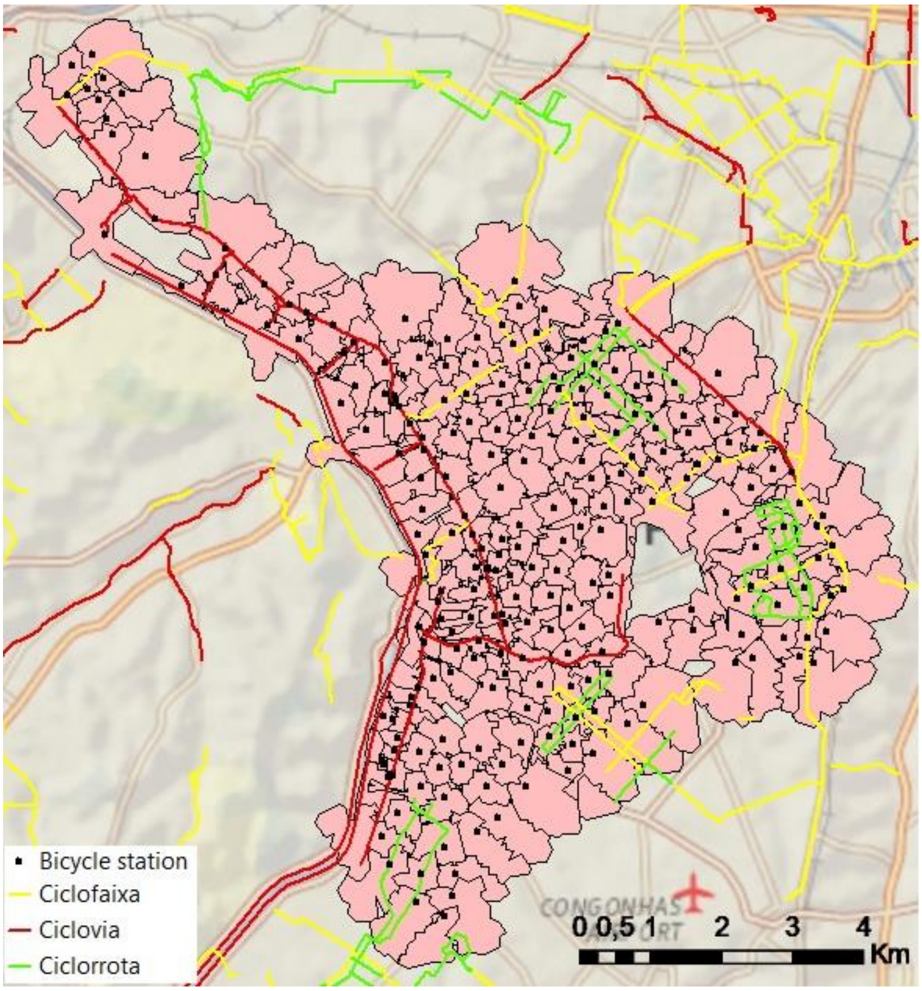


Figure 21: Location and types of bicycles path in São Paulo

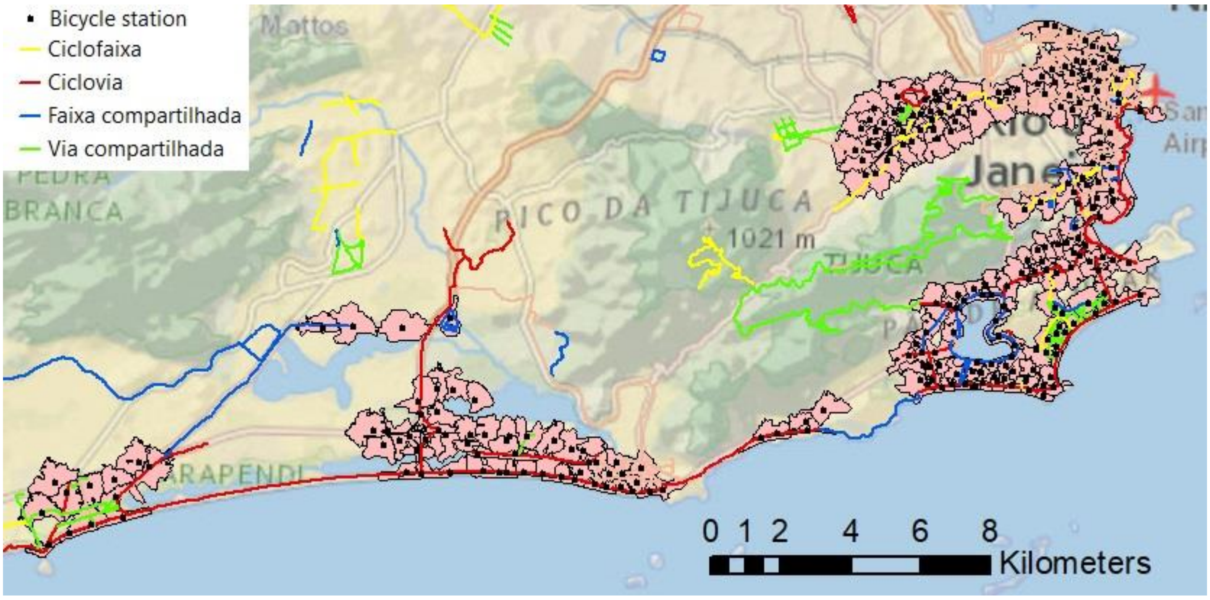


Figure 22: Location and types of bicycle paths in Rio de Janeiro

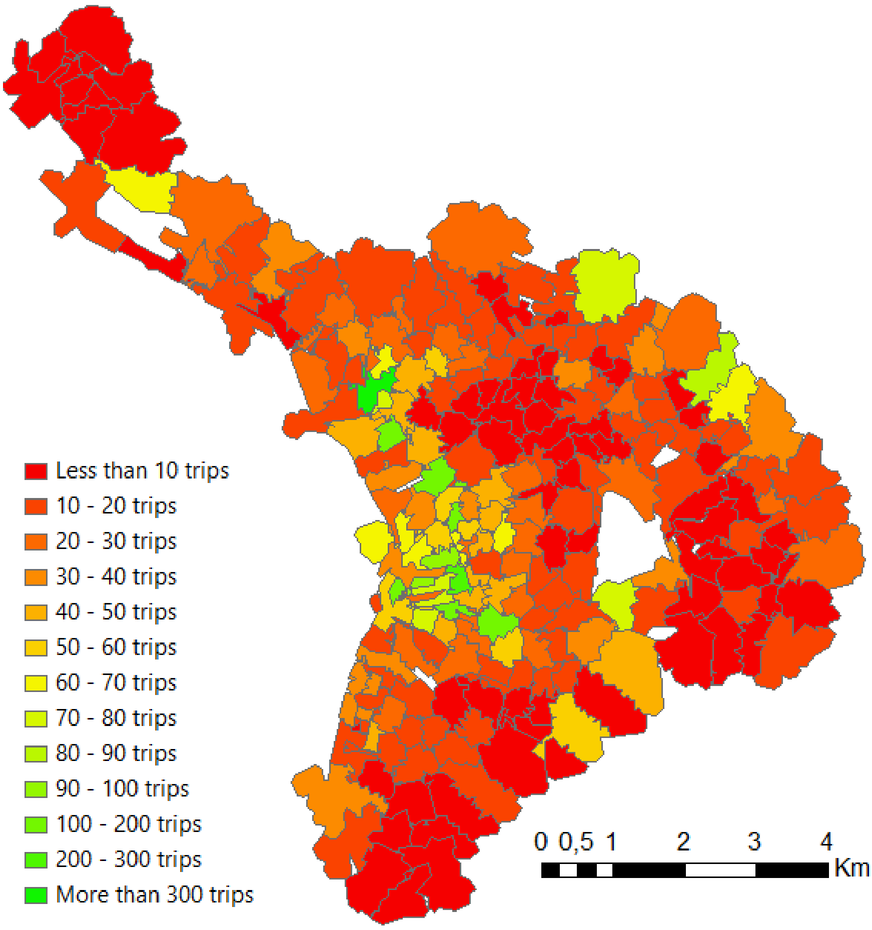


Figure 23: Detailed version of average daily departures of BikeSampa

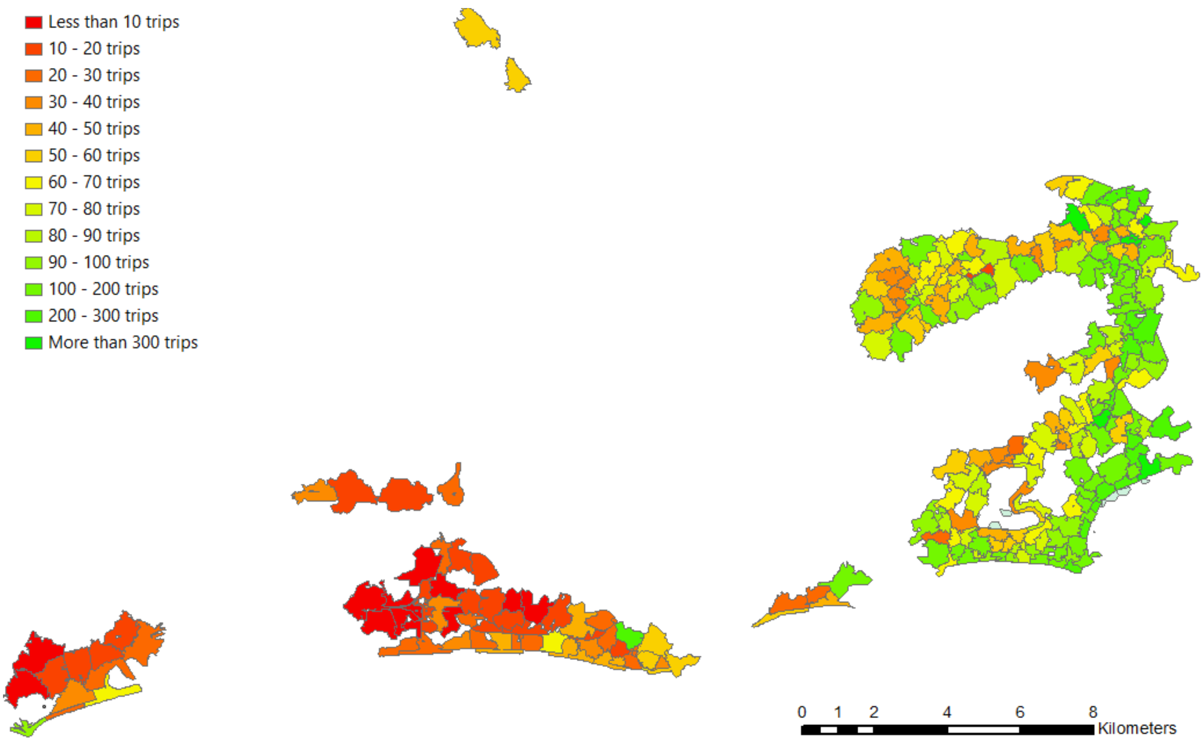


Figure 24: Detailed version of average daily departures of BikeRio

Appendix C – Linear and Logistics curve estimations

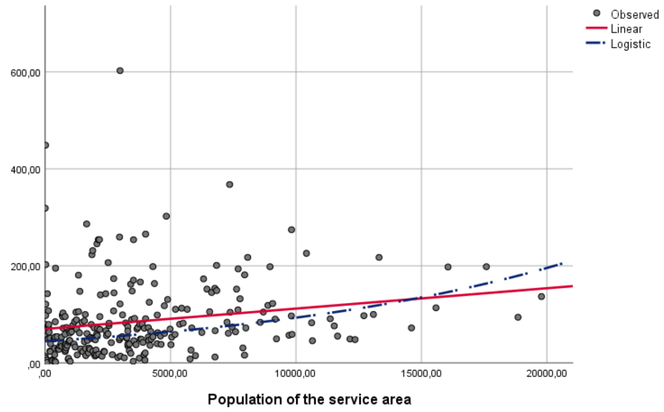
This appendix is a follow up on chapter 5.3.4 and provides a more elaborate view on the individual relationship between the independent variables and dependent variable. Furthermore, the comparison between the linear and logistic regression curve is made. The table below presents a summary of the results for those two methods. Consequently, the remaining independent variables which were not evaluated in section 5.3.5 are plotted to further clarify on their relationship with the dependent variable

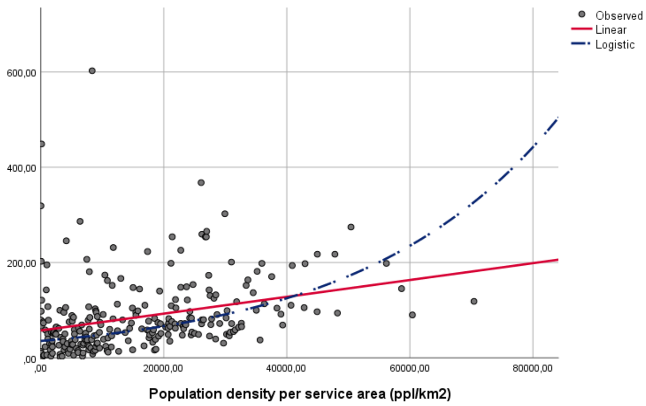
Summary table

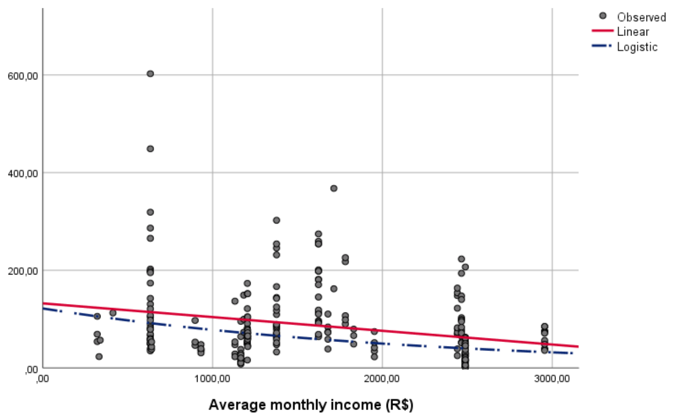
	BikeSampa				**BikeRio**			
	Linear		Logistic		Linear		Logistic	
	R^2	sig.	R^2	sig.	R^2	sig.	R^2	sig.
Population of the service area	0,007	0,195	0,001	0,562	0,040	0,001***	0,078	0,000***
Population density of the service area (ppl/km^2)	0,009	0,152	0,001	0,568	0,092	0,000***	0,183	0,000***
Inhabitants of white/pardo ethnicity (%)	0,011	0,103	0,008	0,187	0,018	0,033**	0,001	0,664
Inhabitants of black ethnicity (%)	0,033	0,005***	0,031	0,007**	0,091	0,000***	0,101	0,000***
The average Human Development Index per service area	0,000	0,852	0,020	0,029**	0,001	0,623	0,003	0,357
The average monthly income (R$)	0,010	0,121	0,003	0,414	0,071	0,000***	0,113	0,000***
Inhabitants over 15 years that can read and write	-	-	-	-	0,000	0,936	0,001	0,591
The average life expectancy in years per service area	-	-	-	-	0,017	0,041**	0,012	0,082*
Inhabitants with a degree in medium education (%)	0,000	0,776	0,009	0,159	-	-	-	-
Inhabitants with a degree in superior education (%)	0,002	0,472	0,002	0,495	-	-	-	-
Scaled average declivity in each service area	0,033	0,005***	0,026	0,014**	-	-	-	-
Station capacity	0,302	0,000***	0,219	0,000***	0,136	0,010***	0,195	0,000***
Station density (station/km^2)	0,035	0,004***	0,081	0,000***	0,025	0,010***	0,077	0,000***
Any type of bicycle infrastructure within 50m of a station	0,054	0,000***	0,076	0,000***	0,000	0,724	0,001	0,611
Presence of a ciclovia within 50m of a station	0,153	0,000***	0,199	0,000***	0,002	0,508	0,001	0,638
Presence of a ciclofaixa within 50m of a station	0,003	0,417	0,002	0,520	0,000	0,966	0,006	0,197
Presence of a faixa comparthilhada within 50m of station	-	-	-	-	0,002	0,523	0,000	0,842
Presence of a ciclorrota within 50m of a station	0,004	0,311	0,006	0,235	0,009	0,122	0,002	0,528
Presence of a metro/train station within 300m of the station	0,057	0,000***	0,018	0,041**	0,086	0,000***	0,079	0,000***
Presence of a metro/train station within 150m of the station	0,049	0,001***	0,010	0,120	0,132	0,000***	0,076	0,000***
Job accessibility index for 15 min. cycling	0,000	0,734	0,002	0,458	0,112	0,000***	0,123	0,000***
Job accessibility index for 30 min. cycling	0,009	0,160	0,004	0,324	0,100	0,000***	0,163	0,000***
Job accessibility index for 60 min. cycling	0,007	0,192	0,006	0,258	0,096	0,000***	0,203	0,000***

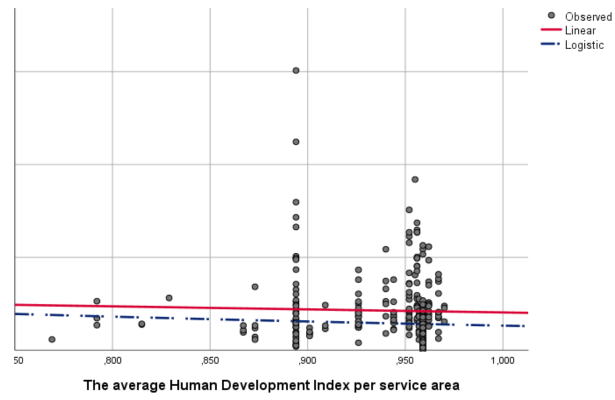
Table 17: Linear and logistic curve estimation of all included variables

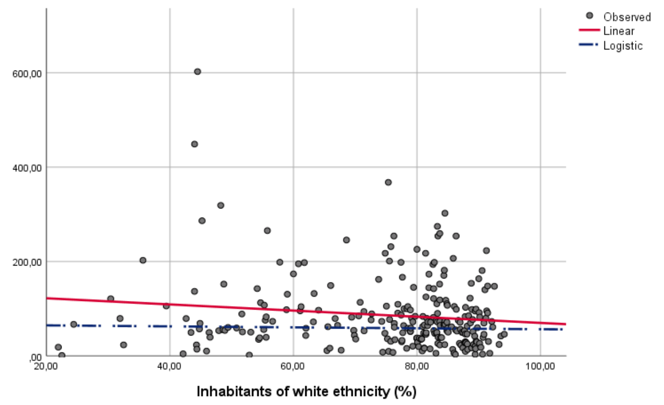
Curve estimations for BikeRio

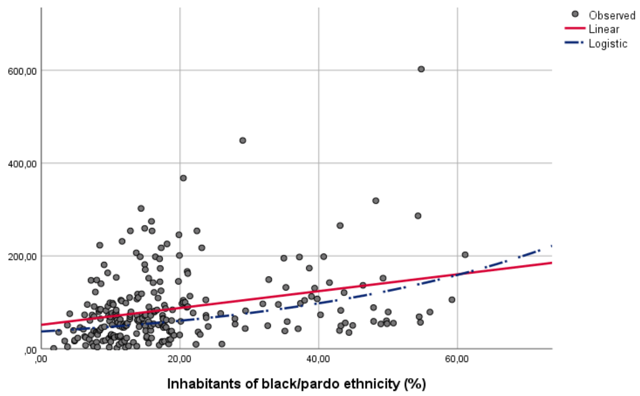


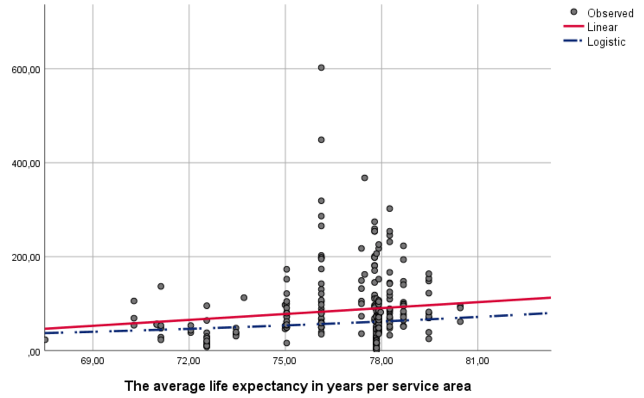


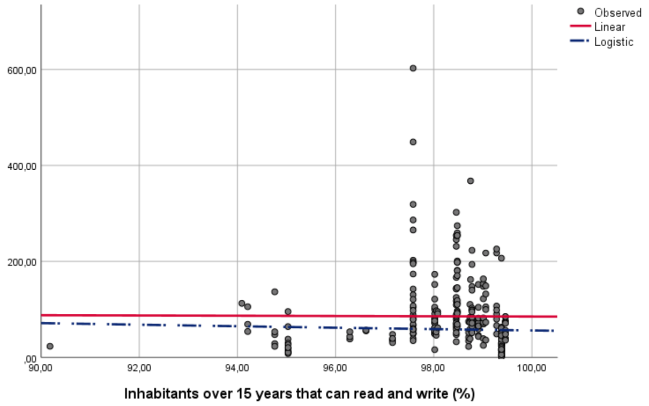

Observed
Linear
Logistic
Average monthly income (R$)

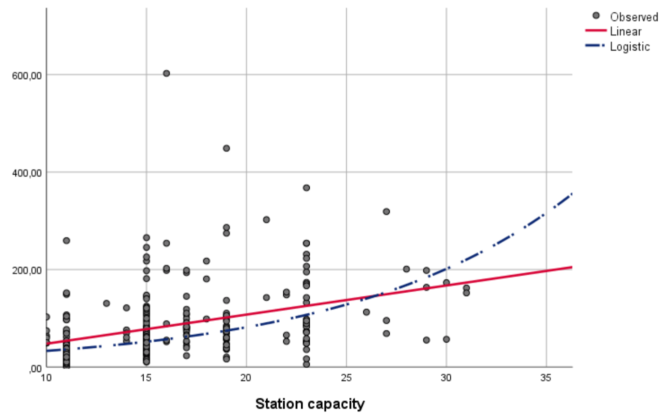

Observed
Linear
Logistic
The average Human Development Index per service area

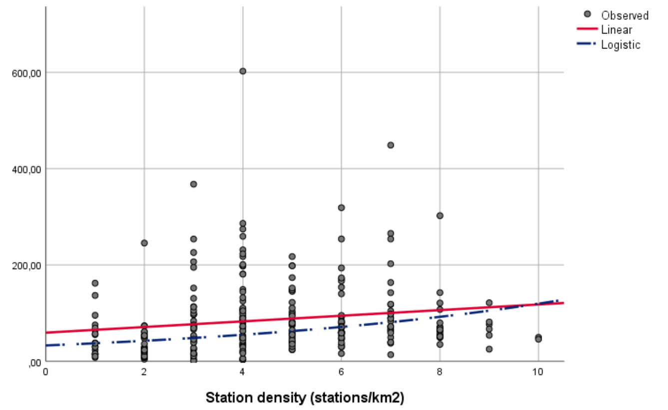

Observed
Linear
Logistic
Inhabitants of white ethnicity (%)

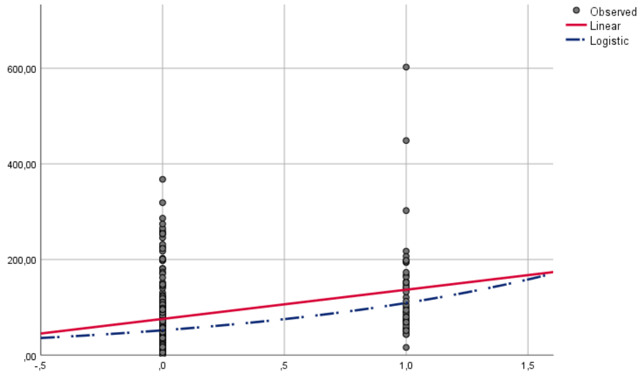

Observed
Linear
Logistic
Inhabitants of black/pardo ethnicity (%)


Observed
Linear
Logistic
The average life expectancy in years per service area

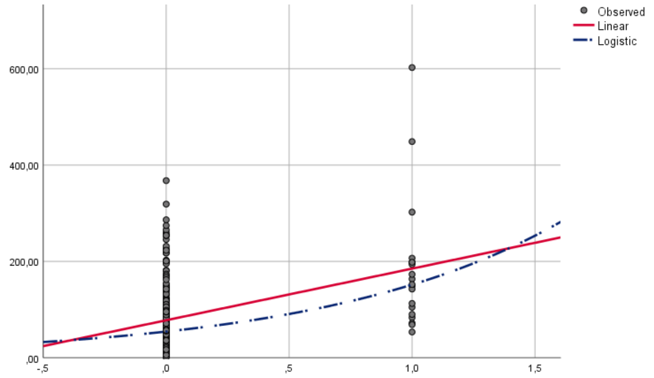

Observed
Linear
Logistic
Inhabitants over 15 years that can read and write (%)


Observed
Linear
Logistic
Station capacity

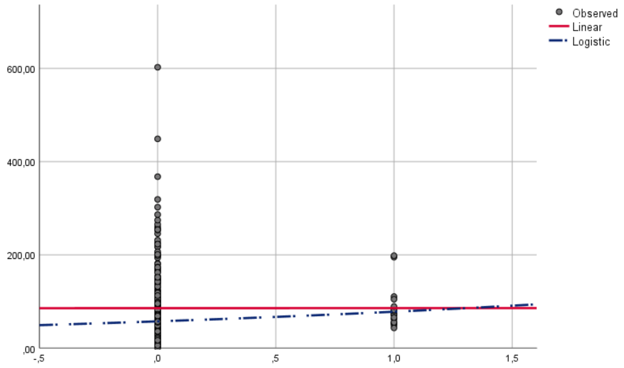

Observed
Linear
Logistic
Station density (stations/km2)



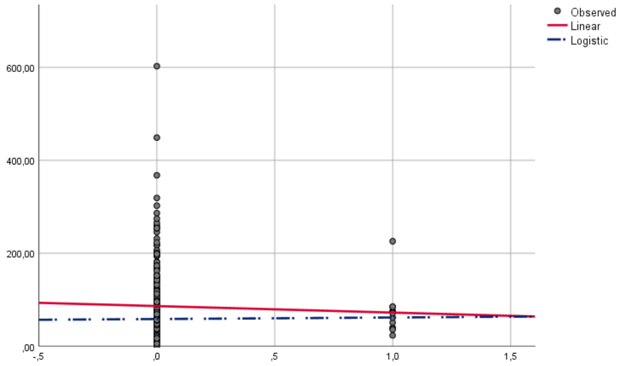
Presence of a metro/train station within 300m of the station



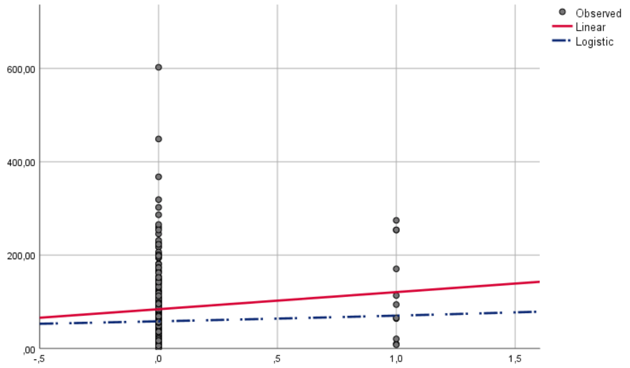
Presence of a metro/train station within 150m of the station



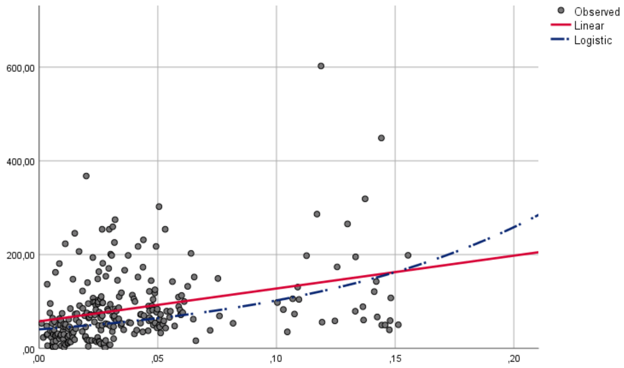
Presence of a ciclofaixa within 50m of the station



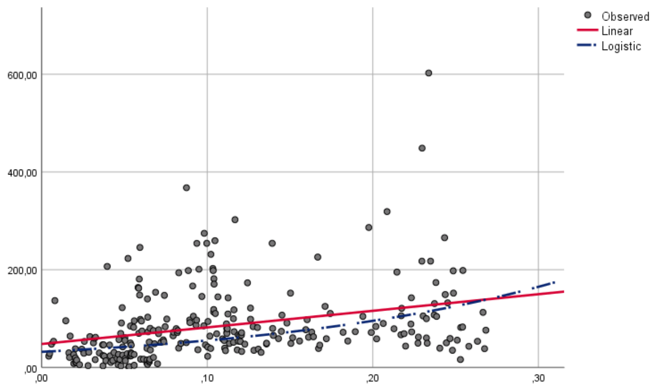
Presence of a faixa compartilhada within 50m of the station



Presence of a via compartilhada within 50m of the station

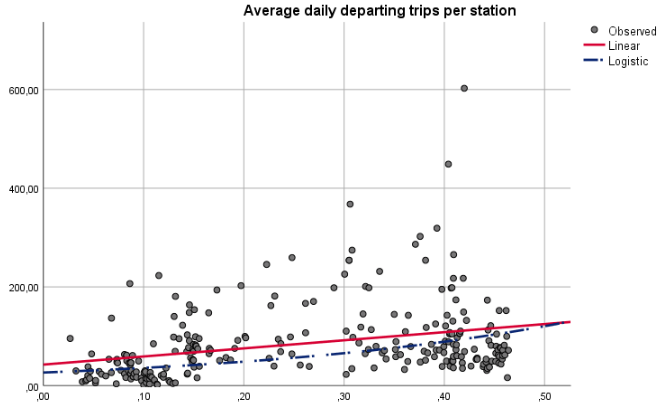


Job accessibility index for 15 min. cycling



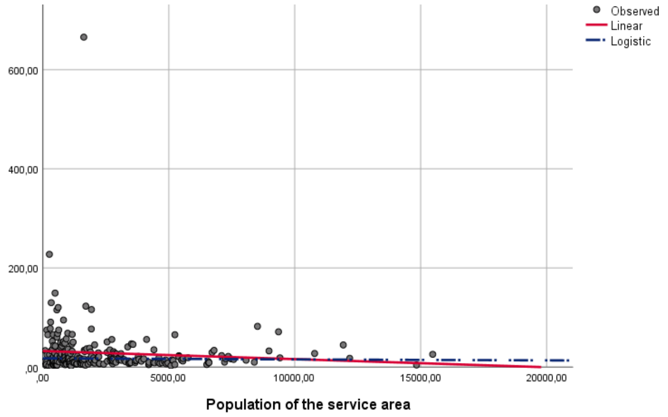
Job accessibility index for 30 min. cycling

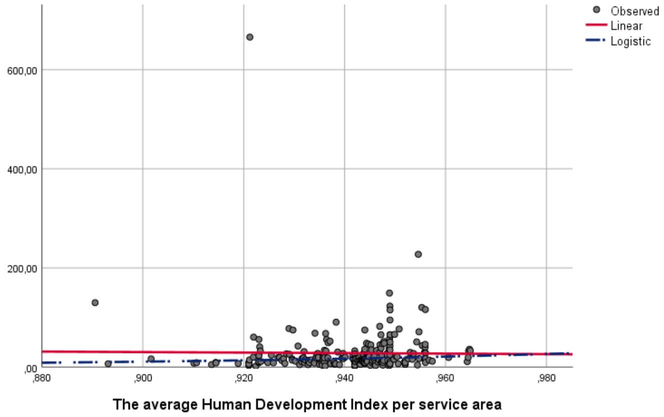
Average daily departing trips per station

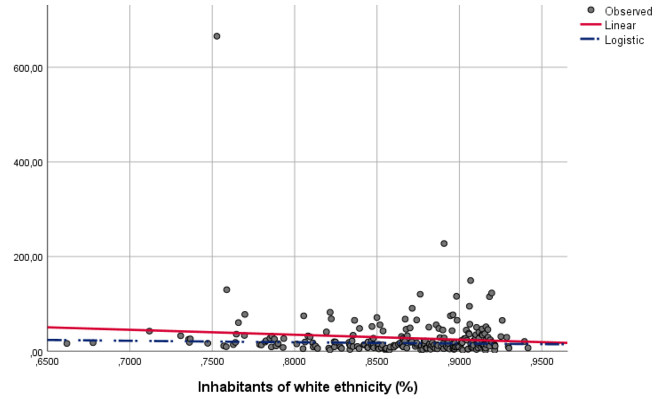


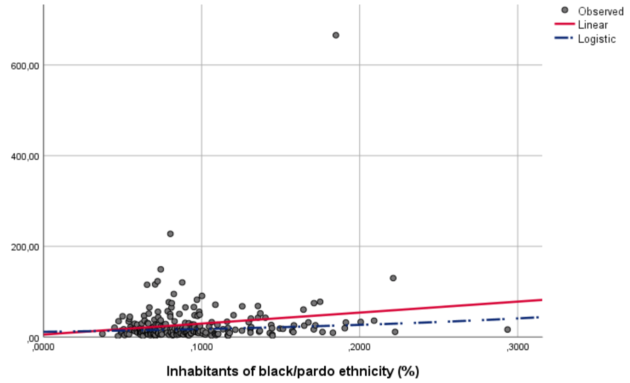
Job accessibility index for 60 min. cycling

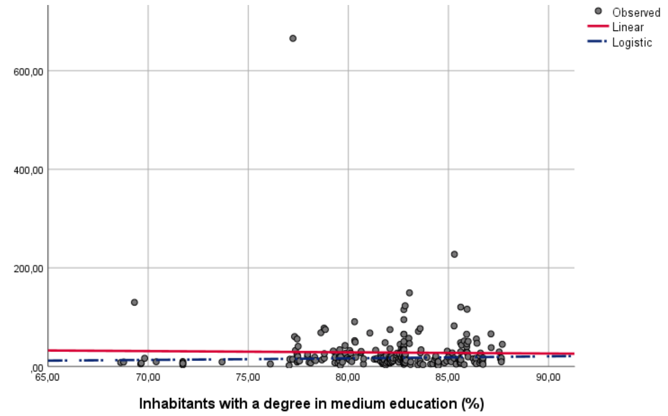
Curve estimations for BikeSampa

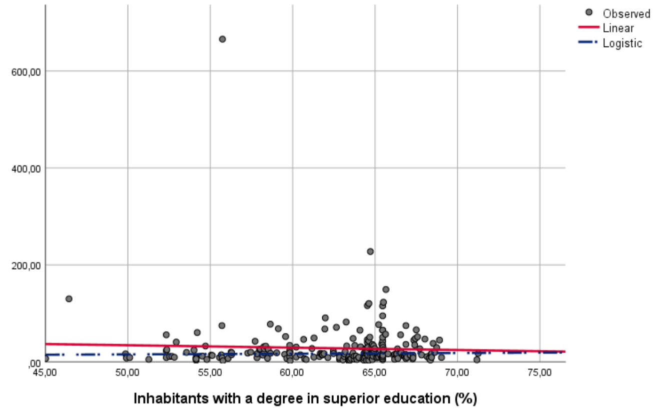


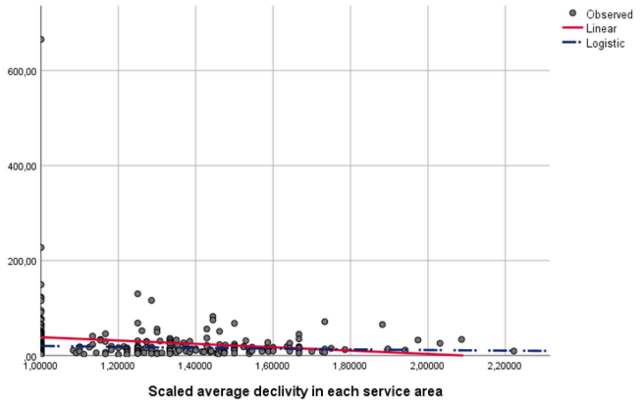


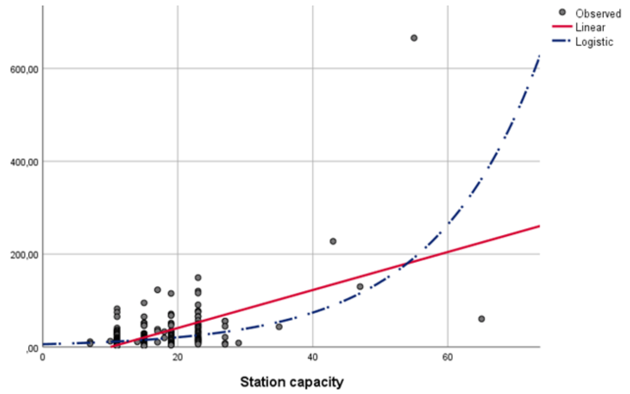


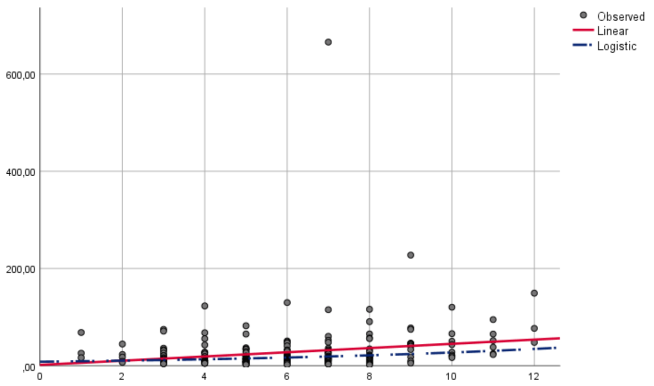




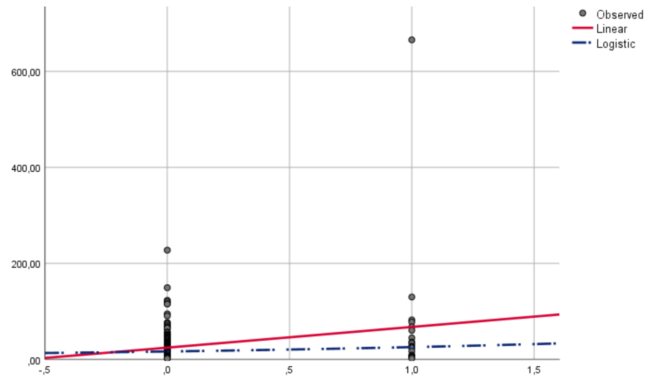




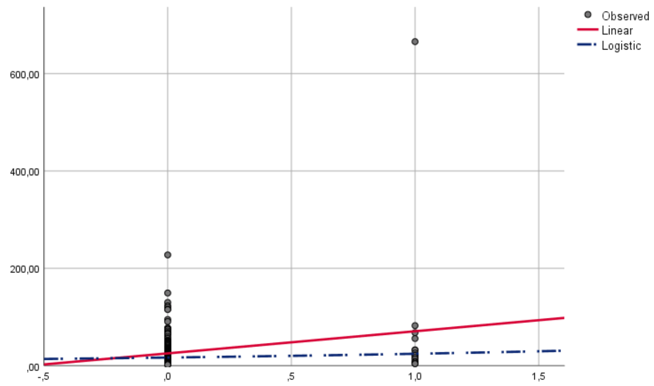




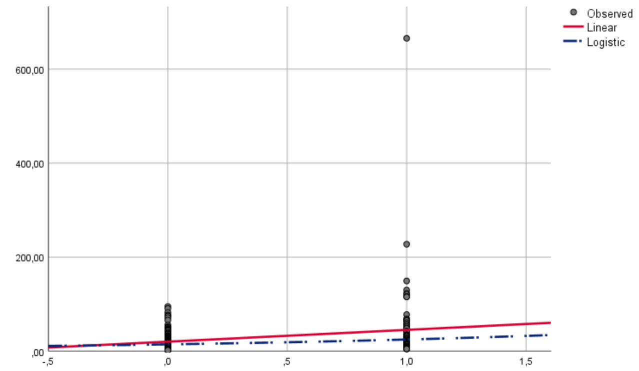
Station density (stations/km2)



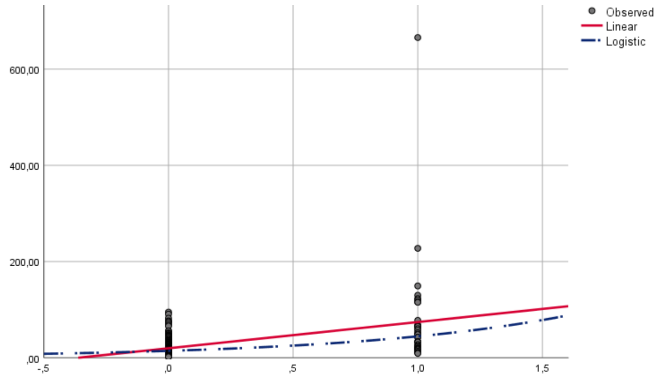
Presence of a metro/train station within 300m of the station



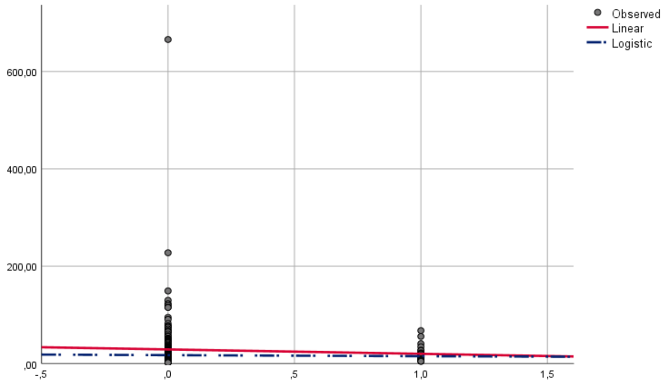
Presence of a metro/train station within 150m of the station



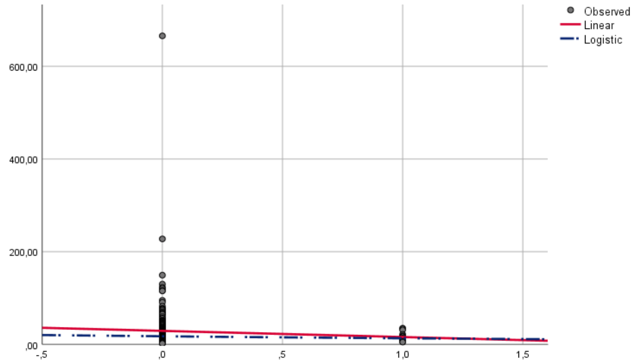
Presence of any bicycle infrastructure 50m of the station



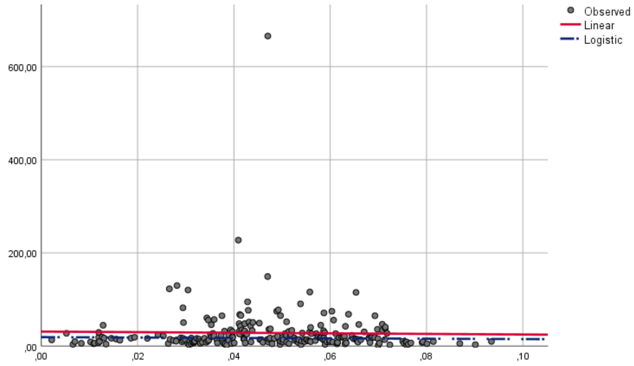
Presence of a ciclovia within 50m of the station



Presence of a ciclofaixa 50m of the station



Presence of a ciclorota 50m of the station



Job accessibility index for 15 min. cycling

Appendix D– Clustering

Land-use characteristics

The first clustered variable is land-use. The author has chosen to assign a service area as residential or commercial, which are the two primary land-use attributes in both of the catchment areas. Thereby, the literature research suggested two distinguished trip purposes for PBSS; commuting and recreational. In general, commuting related trips occur in areas with many offices and commercial buildings. Recreational trips occur in parks and residential neighbourhoods. The following prediction model distinguished primarily work-related service areas with stations located in residential neighbourhoods and parks. In the case of São Paulo; 160 service areas are positioned in residential areas and 71 service areas are situated in work areas. In Rio de Janeiro, the vast majority of the stations, 218, are located in residential neighbourhoods while 28 stations are located in commercial areas. The results are depicted in the table below, which is divided into two parts. The top part provides a summary of three important variables; the average number of departing trips, the stations capacity and the station density. The objective to examine how these variables differ among the clusters. The bottom part of the table portrays the final prediction models for the four clusters. First, a summary of the final prediction model is presented, followed by the significant independent variables. In contrast to the results displayed in Table 10, solely the independent variables that are significant for one of the clusters are shown, which implies that six of the seventeen parameters were not significant for one of the four clusters.

	BikeSampa				**BikeRio**			
	Residential (N=160)		**Work (N=71)**		**Residential (N=219)**		**Work (N=28)**	
Descriptives	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ
Average trips	18,56	18,83	50,00	83,31	83,39	72,94	103,33	103,19
Capacity	15,77	4,77	19,57	9,44	16,39	4,86	16,17	3,83
Station density	5,73	1,996	6,96	2,35	4,37	2,01	5,38	2,43
Total model	R^2	sig.	R^2	sig.	R^2	sig.	R^2	sig.
	0,264	0,013**	0,558	0,040**	0,456	0,000***	0,565	0,057*
Independent variables	beta	sig.	beta	sig.	beta	sig.	beta	sig.
Population density					0,223	0,000***	0,323	0,023**
HDI	0,176	0,016**						
Black residents (%)					0,363	0,000***		
Income					-0,370	0,001***		
Life expectancy (only Rio)					0,436	0,000***		
Capacity	0,190	0,008***	0,563	0,000***	0,143	0,011**	0,448	0,003***
Station density	0,182	0,015**						
Infrastructure					0,276	0,000***		
Ciclovia	0,395	0,000***	0,158	0,086*				
Ciclofaixa					-0,155	0,002***		
Metro 150			0,372	0,000***	0,268	0,000***	0,431	0,005***
Cycling 15			0,286	0,032**				
Cycling 30			-0,217	0,099*	-0,161	0,039**		

Table 18: Prediction models after clustering the land-use characteristics

The descriptive statistics show that especially stations located in residential areas in São Paulo generate a low average number of departing trips. The argument that average station capacity and average station density are responsible for these low values does not hold since these averages are comparable with the other clusters. In general, stations located in commercial areas generate more trips. Finally, the large values for the standard deviation for both clusters in Rio de Janeiro and the ‘work’ cluster in São Paulo indicate significant differences in the number of departing trips between stations within the same clusters.

Separating the models by land-use allowed a more specific insight into how residential and work areas differs in terms of generated trips. Apart from the residential stations of BikeSampa, the determination coefficients have improved. The low R-squared value (0.264) for the 'residential' stations in São Paulo is a result of the small average number of departing trips coming from these stations. Thereby, the low standard variation in departing trips in this cluster makes it challenging to find clear relationships between the independent variables and the dependent variable. The models also point out that the 'work' stations, with R-squared values over the 0.50 in both cities, seem to have higher predictability than 'residential' stations. Furthermore, the work and residential stations have different significant predictors; work stations are primarily predicted by variables related to infrastructure and facilities in and around the stations. The residential stations are also dependent on the composition of the population in the service area and their attributes. For instance, life expectancy and the percentage of black people are significant predictors for the residential stations in Rio de Janeiro. A remarkable result in both the 'work' clusters is that the job accessibility index for up to thirty minutes cycling has a significant negative impact on the dependent variable.

Station capacity

The decision to cluster the stations based on station capacity has two reasons. At first, all the six developed prediction models have the station capacity as a significant predicting variable. A higher capacity means more possible locations to get and return the bicycle and therefore, are likely to generate more trips. Secondly, the stations capacity variable has the highest Pearson's correlation coefficient with the average number of departing trips, 0.55 for BikeSampa and 0.369 for BikeRio, respectively (see Table 9). Two clusters were developed for each system using K-means clustering. Accordingly, the stations are divided in two groups. Stations with a low capacity have up till 20 docks, and high capacity stations have up to 65 docks (BikeSampa) and up to 31 station docks (BikeRio). The table below shows the descriptives for the average number of departing trips and the average station density for each cluster. The division of the number of stations in each cluster is similar for BikeSampa and BikeRio, yet the model results are not alike. The clustering also did not result in improved values for R-squared. Hence, the results are discussed briefly.

The descriptive statistics are in line with the expectations and the previous models, a higher capacity leads to more average departing trips. The low capacity cluster for BikeSampa has a remarkably low determination coefficient, similar to the cluster of the residential stations presented in the previous paragraph. It appears that the clusters are corresponding since 80% of the stations that fall under the low capacity cluster are located in a residential neighbourhood.

Station capacity	**BikeSampa**				**BikeRio**			
	Low (N=181)		**High (N=51)**		**Low (N=210)**		**High (N=50)**	
Descriptives	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ
Average trips	19,77	19,31	58,24	96,51	75,67	72,13	129,68	82,11
Station density	5,91	1,960	6,88	2,747	4,40	2,103	4,80	1,98
Total model	R^2	sig.	R^2	sig.	R^2	sig.	R^2	sig.
	0,260	0,000***	0,596	0,096*	0,467	0,000***	0,293	0,107

Table 19: Prediction models after clustering the station capacity

Station density

The station density is divided into two categories using K-means clustering. The low-density clusters have a station density up to six stations per km^2 and the high-density clusters have up to twelve stations in the near perimeter. According to the literature, higher station density is linked to more trips. The descriptive statistics show similar results, the average number of departing trips is higher when the station density increases, while the capacity of the stations remains similar. The corresponding prediction models of the clusters do not show satisfactory R-squared values aside from one positive outlier. Interestingly, the high-density cluster from BikeSampa has a prediction model that explains almost 70% of the total variation. It seems possible that these results are due to the high standard deviation, which is set at 73 is nearly twice the value for the average departing trips. The significant predictors for this cluster are sup-edu, declivity, capacity, metro 150, ciclovia and ciclofaixa

Station density	**BikeSampa**				**BikeRio**			
	Low (N=135)		**High (N=99)**		**Low (N=210)**		**High (N=50)**	
Descriptives	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ
Average trips	19,94	20,67	39,44	72,71	81,11	77,87	95,39	74,31
Capacity	16,07	5,33	18,13	8,246	16,06	4,79	17,05	4,63
Total model	R^2	sig.	R^2	sig.	R^2	sig.	R^2	sig.
	0,309	0,003***	0,684	0,038**	0,479	0,000***	0,327	0,096*

Table 20: Prediction models after clustering the station density

Average departing trips

The fourth and final cluster divided the dependent variable and sought to demonstrate the possible differences in modelling stations with a lower and higher number of departing trips. In this case, stations that do not surpass the fifty daily departing trips are combined in one cluster. The station between fifty and hundred and fifty daily trips are put together. The last cluster has stations with over a hundred and fifty daily departing trips and since BikeSampa has two stations that fit this condition, this cluster is exclusively made for BikeRio. The prediction models of all five clusters did not have improved values for R-squared. A thought-provoking result is that the second cluster of both cities have a particularly low value for R-squared

Average Trips	**BikeSampa**				**BikeRio**					
	Trips < 50 (N=203)		**50 < Trips < 150 (N=28)**		**Trips < 50 (N=93)**		**50 < Trips < 150 (N=129)**		**Trips > 150 (N=38)**	
Descriptives	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ
Station density	5,91	2,00	7,39	2,97	3,77	2,06	4,93	2,10	4,71	1,59
Capacity	15,88	4,66	22,29	10,82	13,78	3,52	16,98	4,21	20,55	5,47
Total model	R^2	sig	R^2	sig	R^2	sig	R^2	sig	R^2	sig
	0,293	0,001***	0,160	0,000***	0,449	0,031*	0,154	0,012**	0,413	0,073*

Table 21: Prediction models after clustering the average departures

Conclusions clusters

- Against the expectations, most clusters had unimproved values for R-squared. Clustering by a significant predictor results, in most cases, in an inferior prediction model
- The homogeneity of the stations in São Paulo (population predominantly wealthy, white) could not be filtered out through the clustering. The vast majority of these were residential, low capacity, low number average number of departing trips and had a low station density. Consequently, the R-squared values of these clusters were also low.
- The literature pointed out that higher density and capacity should result in more trips, this is line with the descriptive statistics of the model. Furthermore, it seems that these clusters of stations are ‘easier’ to predict, the determination coefficient is higher

Appendix E– Questionnaire

The detailed version of the questionnaire is presented below. The questions are answered by Renata Rabello, an employee of TemBici at the 22nd of May 2020. The questions are displayed in blue and the answers as given by Ms Rabello in red and cursive letters. The objective of this questionnaire is to gain more insight into the operations and considerations made from the perspective of Tembici. The questionnaire has a total of nine questions with different subjects:

1. History of the systems in São Paulo and Rio de Janeiro
2. Theft and vandalisation of the bicycles
3. Rebalancing of the bicycles
4. The core values in deciding where to place a station
5. The approach per city
6. The decision-making power of the involved actors
7. The interaction with other modes of transport
8. Questions regarding the results and findings of the thesis
9. The future of TemBici

Questions

1. History of the systems in São Paulo and Rio de Janeiro

The first part of the questionnaire will help me understand the bicycle system and the history better. I came to notice that I still have many trivial questions relating both public bike-sharing systems.

A. In São Paulo, a previous system failed because of robbery and vandalisation of the public bicycles. What is the history of TemBici in Rio de Janeiro?
    - Was there any system active before the implementation of TemBici as it is today?

    ***1a**: The story is very similar. TemBici bought the company Samba that operated the systems in São Paulo, Rio de Janeiro, Porto Alegre, Salvador and the State of Pernambuco in 2017. And started operating these systems firstly with the old technology (from the company Samba) and then, we changed every system's technology for PBSC. I explain this process in more detail in my master's dissertation. https://teses.usp.br/teses/disponiveis/16/16135/tde-05112019-164700/pt-br.php*

B. What are the primary changes that were made to prevent this from happening again?

    ***1b**: Yes, we used, as an example, the systems that already operated PBSC technology, such as London, Chicago, Guadalajara, and now Barcelona. The stations and bicycles are more robust, and developed specifically for bicycle sharing. They are not ordinary bikes adapted for shared use. Thus, the parts are unique and unconventional to prevent theft, and the locking system is reinforced. And bikes have GPS to help us recover, in case of robbery.*

C. The last week I was at USP, in the beginning of March, I saw new TemBici stations being placed at the campus. I heard that this is not the first time TemBici has been operating at USP, but the previous stations were removed about 1 year ago I believe. What was the reason to place stations at the campus again? Which differences are there with the previous time?

    ***1c**:Actually, there was an event in 2019, when we installed some stations for 1 week, just for the duration of the event. It has always been our interest to have stations at USP in Cidade Universitária from the beginning, due to the potential of trips, between the internal stations of the Campus, and as a connection with the Butantã Metro and the Cidade Universitária train station. But it is interesting to tell the story of Mauricio Villar, one of the directors of Tembici (our COO). He was a student at Poli-USP and his final project was the Pedalusp System, implemented in Cidade Universitária 10 years ago. Therefore, there was already a system with a few stations within the Campus and in the Butantã metro, but it was closed due to lack of financing.*

2. Theft and vandalisation of the bicycles

As far as I've heard, theft and vandalisation is a significant problem for operators. I would like to know a bit more about the numbers and how is dealt with this problem

A. How many bicycles are there on average stolen, vandalised or declared unusable per week or month?
   - Does this fall within the expectation, or does this exceed this?

   2a: *The financial and operational model forecasts 1% per month for vandalism and on average we are below that. But there was an exception to overcoming this expectation in the case of the Buenos Aires System (Ecobici). The vandalism rate was higher than expected in the first year. The system is subsidised by the government, in addition to having a sponsor, and is free for users. Because of the gratuity, many users did not return the bike because there was no penalty for charging excessive use.*

B. Are the numbers of theft and vandalisation for the current bicycle system less than the previous system?

   2b: *The current vandalism rates are certainly lower than that of the old technology, which presented rates of approximately 10% a month*

C. What is the percentage of costs incurred by theft and vandalisation?

   2c: *Unfortunately, we consider costs as sensitive data*

3. The rebalancing of the bicycles

A. When is rebalancing happening in general? Day/night?

   3a: *Rebalancing happens all day, even at dawn, to organise the system to start the day balanced. There is also a great effort to remove the bicycles marked as inoperative by the users, so that they can be taken to the warehouse, repaired and then returned back to the system.*

B. Does the process follow some sort of schedule or is it determined per day how things are going?

   3b: *An artificial intelligence system is used that calculates the demand for each station at each hour of the day and directs the logistical operation to supply each point with the number of bicycles needed.*

C. Are there many lost sales due to inadequate rebalancing? (meaning that people cannot use the system because there are no available bicycles)

   3c: *Especially during peak travel times in cities, the demand may be greater than the supply of bicycles. For this reason, we use different strategies in order to increase the offer in the stations at specific times. We have valet stations in specific points of the city, such as in metro stations, with the offer of more bikes than the number of available docks. In addition, the system is planned with a dense network of stations approximately 400 meters apart between each one, to ensure a good user experience, so that if he needs to return or collect a Bicycle and does not have spaces or bicycles available at the station he's at, there are locations nearby as a second option for his need.*

4. The core values in deciding where to place a station (idea: scaling of what is more and less important (1 = unimportant / 5 = important)

During the development of the system, various considerations have been made.

Consideration for placing a bicycle station	Importance (1 to 5)
Number of potential users of the system	*5*
Potential customers of Banco Itaú	*1*
Areas with low criminality rates/safe to cycle for the customers	*3*
Proximity to bus/train/metro station	*5*
Proximity to bicycle infrastructure	*5*
The proximity between other stations/ station density	*5*
Docking stations are easy to access (by car) for maintenance or rebalancing	*3*

Table 22: Core values of TemBici

Any additional comment relating to the core values

- *The visibility is also really important, so we always try to locate stations at intersections, next to the entrances of public transportation, visible from the cycleways. People can be 'invited' to use a bike if they remember that it exists and that it is in their path. If they are not visible, they might not be useful for the user.*

- *About bicycle paths, we always seek information about existing bicycle paths and about plans for new proposals to adapt our system to this infrastructure.*

- *Density is essential for the system, and it always starts where there is a greater concentration of jobs, bicycle infrastructure and integration with the city's public transport system. Therefore, the coverage area of the stations is defined, in brief, considering these three factors and is limited to the number of stations available for each city. In cases of expansion, the expansion of the coverage area adjacent to the existing locations is always considered. Through modal integration, it is possible to offer shared bicycles to the population of the entire city, to those who live in peripheral regions but who work in central regions. In Tembici systems, approximately 70% of users live outside the coverage area. This demonstrates that it is not only the people who live in the central regions that use the system.*

5. The approach per city

Both cities are quite different in build-up and the way they operate. For instance, São Paulo has more commercial areas, and Rio de Janeiro has more tourists.

A. Do both systems have different/specific target users?

- How can this difference be seen in the development of both systems?

5a: *The two cities have their differences, however, in both, the subscribers, who travel for both work and leisure, are the vast majority, and the company's main focus.*

B. BikeRio also covers the historic city centre, while BikeSampa removed stations from their historic city centre. Both areas are not considered safe. What is the main reason why for this difference in approach?

5b: *BikeSampa did not remove the stations from the historic centre. Obtaining authorisation from the municipality took longer, but as soon as we had the license, the stations were installed.*

6. The decision-making power of the involved actors

A. Which actors are involved in the process of deciding where the station will be placed and the number of docs per stations and how are they involved? *Please give a short statement of the involvement of actors below and how they are involved.*
   - Municipality of São Paulo:
   - Banco Itaú:
   - TemBici:
   - Others??

   6a: *In each city there is a different involvement and process for the licensing of stations. Tembici carries out the study and planning of the system. This plan is presented both to Itaú (for science) and to municipality in the cities.*

B. Is there communication/collaboration between the parties in deciding where to build bicycle paths?

   6b: *At the beginning (when we reviewed all the systems for the exchange of technology) Itaú was responsible for connecting us with cycle activists of every city, because their opinion (of itau) about our proposal was going to be based on the opinion of the cycle activists (due to their knowledge and their experience of cycling in each city). This was really important because it meant that the principal goal of the bike sharing system for Itaú was to be useful as a means of transportation, and it wasn't related to where they thought the publicity would be better. So, they gave our urban planner's team all their confidence of our technical opinion in developing and planning each station location.*

C. Is there a superior party in the decision-making process and who is this?

   6c: *The municipality analyses each location to identify the feasibility of implementing the bike station, considering the current traffic and urban planning rules.*

7. Interaction with other modes of transport

A. Is there any communication and/or collaboration with other modes of transport in São Paulo?
   - Train & Metro?:
     My research showed that bicycle stations within 150 meters of a metro or train station generate significantly more trips, for stations within 300 meters, this was not the case

   7a: *We always seek to set up stations close to the city's metro and train stations, as in fact there is a lot of demand for modal integration.*

B. Other PBSS, such as Yellow bike & CicloSampa did you make any agreements with those parties?

   7b: *There was a dialogue, with Yellow, when we verified the existence of station projects for the same spaces.*

C. **For the future:** A smart card that could be used for all modes of transport?

   I remember to hear there was already a project to integrate the SMART card
   - *How far is it?*
   - *Do you think it will be possible to use the SMART card somewhere in the next five years?*

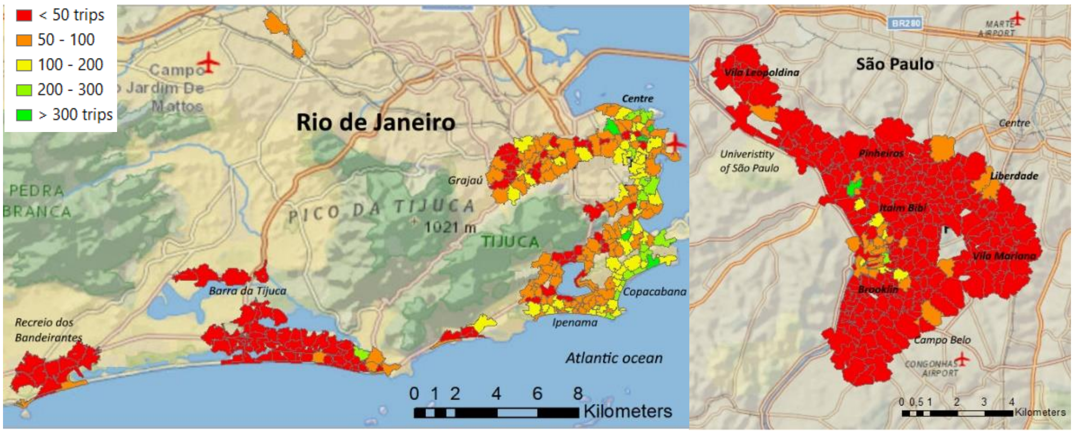
   7c: *Currently, it is possible to register the Bilhete Único, in São Paulo, to use the system, but it is not yet possible to use it as a means of payment. I have no information on when this will be possible, as it does not depend on Tembici.*

8. Questions regarding the results and findings of the thesis

Some findings:
I analysed the data since April 2018 and it struck me that the locations of (some of the) stations are constantly changing, especially in São Paulo

I explained about the original project for Bike Sampa, and about the difficulty to have the authorisations in my thesis. Therefore, firstly, we installed where the licenses came first, and then we started relocating to the locations of the original project (USP and Subprefeitura da Sé). We are also always analysing the performance of the systems, and replanning things if necessary. It happens with all the systems (not only Rio and SP).

- *I compared the service area with the municipal averages:*
    - *Income and HDI of the service area are much higher than the municipal averages*
    - *Stations are located in 'White', wealthy neighbourhoods, especially in São Paulo, but most of these stations generate very few trips (see figure and table below)*



Average daily trips per station	BikeRio	BikeSampa
less than 50	93	203
50 – 100	99	22
100 - 200	47	6
200 - 300	16	1
more than 300	5	1

A. The prediction model (for Rio) shows that neighbourhood with a higher percentage of black population and a lower average monthly income generates more trips.
   - Were you aware of this trend?
   - Do these results make sense to you? Is it line with your own experience?

   ***8a:** What was considered for your prediction model? Could you show me the map with the result, in order for me to see if the results make sense? We are also working in a prediction model, and we have an understanding of where would generate more trips (in case of an expansion), but we analyse many urban aspects, not only the higher percentage of black population and lower average monthly income.*

B. Stations located in neighbourhoods with lower income per capita generate more trips on average. This might also be a reason why the system in SP is performing inadequate, since *"all the stations are located in wealthy neighbourhoods and serve people that do not need this service".*
   - How do you look at this statement?

   ***8b:** Completing what has already been explained in the answer to question number 6, our cities have the common characteristic of having a large concentration of jobs and infrastructure in the central region and a large dispersion of housing in peripheral regions. The floating population in the central region is very high, which is why the system started to be implemented in this region, with the aim of reaching a larger number of people. In order to maintain density, we continue to implement the network of stations respecting the System's starting point. São Paulo and Rio de Janeiro have many urban differences, so it is difficult to compare the use of both. One of the biggest differences, which in my opinion is one of the reasons for the greater number of Bicycle trips in Rio de Janeiro than in São Paulo, is the hourly travel graph. In São Paulo, there are two major peaks on the way to and from work with a valley at lunchtime and in Rio de Janeiro, the use remains intense throughout the day. Some hypotheses for this* *behaviour may be the greater offer of leisure spaces in Rio de Janeiro, such as the bike lanes on the seashore, which function as an extensive linear park, while in São Paulo there are only a few punctual parks distributed in the coverage area of the System. Or it can also be a behavioural issue in Rio de Janeiro, of using the bicycle for various activities throughout the day, not as concentrated in the use for work, as it seems to be in São Paulo.*

C. Do you think that the choice of the locations of the stations is rather to advertise (by Banco Itaú) than to improve the overall mobility and accessibility of the cities?
8c: *No, I don't think the choice of the locations of the stations is rather to advertise than to improve mobility and accessibility of the cities. Find below the answer for question 6:*
At the beginning (when we reviewed all the systems for the exchange of technology) Itaú was responsible for connecting us with cycle activists of every city, because their opinion (of itau) about our proposal was going to be based on the opinion of the cycle activists (due to their knowledge and their experience of cycling in each city). This was really important because it meant that the principal goal of the bike sharing system for Itaú was to be useful as a means of transportation, and it wasn't related to where they thought the publicity would be better. So, they gave our urban planner's team all their confidence of our technical opinion in developing and planning each station location

9. The future of TemBici
The last set of questions are focusing of the future and possible new developments.

A. Are there any plans to extend BikeRio and BikeSampa inside both cities?
- *Both systems cover only a small part of both cities. As a consequence, the majority of inhabitants doesn't have access. 5.5% of Sampa's population and 16% of Rio's population live within a 10-minute walking distance of a station (these are the areas that are shown in the figure, the so-called stations' service areas of each station). Meaning inside the cities are many potential locations to implement stations*
- *Include **less developed regions**, which probably benefit more*

9a: *Yes, there are plans to extend both cities. It is really important to remember that modal integration is one of the key elements for the bike sharing system, and that more than the population that live near the stations use the system. In my thesis, you can find a map of the resident location of the users of BikeSampa. Do you know about the 'Estação Bike of Cidade Tiradentes'? – you can find information in my thesis too. Is a bike parking with BikeSampa bikes in a peripheral neighbourhood of the city.*

B. Do you know about any ideas of expanding to more Brazilian cities next to the five that have your service now?
Many cities have a (much) lower average income than SP and Rio and are located on the coast, which means they are mostly flat (especially close to the coast)
9b: *We also have a system in Vila Velha, called Bike VV. We are always studying municipal biddings, looking for more opportunities in other Brazilian cities.*