University of São Paulo
Institute of Mathematics and Statistics
Bachelor of Computer Science

An evidence-based bicycle modal shift potential index

A case study of São Paulo, Brazil

Pedro Gigeck Freire

Final Essay

mac 499 — Capstone Project

Supervisor: Dr. Leticia Lemos
Co-supervisor: Prof. Dr. Higor Souza
Co-supervisor: Prof. Dr. Fabio Kon

São Paulo
December 24, 2021

All models are wrong, but some are useful.
— George E. P. Box

Acknolegments

First and foremost, I would like to express my gratitude and admiration to family for their unwavering support and affection at every moment, during and before the development of this work. Katia, Silvio, and Bianca, thank you for everything!

Also, I would like to thank my supervisors Dr. Letícia Lemos, and Professors Higor Amario and Fabio Kon. This work would not be possible without your valuable guidance. Thank you!

At last but not least, I would like to thank the professors and colleagues from the Institute of Mathematics and Statistics (IME-USP). The friendly environment you provided, even during the pandemics, was essential to the conclusion of this Capstone Project.

Resumo

Pedro Gigeck Freire. **Um índice baseado em evidências para migração de modal para bicicleta: *Um estudo de caso da cidade de São Paulo, Brasil***. Monografia (Bacharelado). Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2021.

Este trabalho visa criar um modelo baseado em evidências para indicar locais com maior potencial de uso de bicicletas. O índice *SP Cycling Potential* foi criado a partir de revisões bibliográficas e análises estatísticas da última pesquisa de mobilidade urbana de São Paulo para modelar o potencial de migração para bicileta das viagens. As variáveis consideradas são a distância e inclinação das rotas, assim como a idade e gênero dos indivíduos obtidas da pesquisa Origem Destino 2017. Os métodos consistem em aproximações contínuas das distribuições dos dados dos ciclistas. Assim, identificamos regiões da cidade de São Paulo com pouca infrastrutura cicloviária mas com alto potencial de adoção de bicicletas. Esperamos que o modelo e ferramentas analíticas descritos neste trabalho influenciem as políticas de mobilidade com bicicletas.

Palavras-chave: Bicicleta. Migração de Modal. Potencial Ciclável. Mobilidade Urbana. Ciência de Dados.

Abstract

Pedro Gigeck Freire. **An evidence-based bicycle modal shift potential index: *A case study of São Paulo, Brazil***. Capstone Project Report (Bachelor). Institute of Mathematics and Statistics, University of São Paulo, São Paulo, 2021.

This work aims to create an evidence-based model to indicate places with the most significant potential for bicycles' use. We developed the SP Cycling Potential Index using literature reviews and statistical analyzes from the latest São Paulo Urban Mobility Survey to model the potential for migration to cycling. The variables we considered are the distance and slope of the routes, as well as the age and gender of individuals. We modeled continuous approximations of the distributions of the cyclists' data to compose the model. Thus, we identified regions of the city with little cycling infrastructure but with a high potential for adopting bicycles. We hope that the model and analytical tools described in this work will influence bicycle mobility public policies.

Keywords: Bicycle. Modal Migration. Cycling Potential. Mobility. Data Science.

Abbreviation List

API	Application Programming Interface
ACP	Analysis of Cycling Potential
BH	Belo Horizonte
BSS	Bike Sharing System
CDF	Cumulative Density Function
CPI	Cycling Potential Index
CPT	Cycling Potential Tool
CSS	Cascade Style Sheets
CSV	Comma-separated Values
JS	JavaScript
OD	Origin-Destination
OLSR	Ordinary Least Squares Regression
PCT	Propensity to Cycle Tool
PDF	Probability Density Function
SP	São Paulo
SPMA	São Paulo Metropolitan Area
SRTM	Shuttle Radar Topography Mission

List of Figures

List of Tables

List of Programs

Contents

Chapter 1

Introduction

The increase in motorized transportation modes in large cities in the last decades has become a challenge as traffic congestion, noise and air pollution reach unsustainable levels. In this context, stimulating alternative commuting modals is a relevant topic faced by many cities worldwide. Recent international initiatives establish goals and develop strategies for the urgent necessity of reductions in carbon emissions (Bank and KEMA, 2014; Watts, 2018).

The bicycle modal represents a vital alternative for polluting vehicles due to its advantages for user health, low costs, air and noise pollution reduction, and others. Many studies analyzed a large amount of evidence and produced homogeneous conclusions regarding the benefits of cycling (for both individuals and cities) (Nazelle *et al.*, 2011). However, the use of bicycles is still quite limited in some regions. In the city of São Paulo, the bicycle modal corresponds to only 0.8% of the daily trips, even after the deployment of about 600 km of specialized lanes for bicycles in the past decade (METRÔ, 2021).

In cities where the mobility policies traditionally focused on motorized vehicles, such as in the São Paulo study case (Lemos, 2021), developing efficient strategies to promote cycling is an urgent and challenging topic. The objective of this work is developing an evidence-based model to identify locations where investments in cycling infrastructure could have a greater impact. With this model, the municipality can prioritize cycling infrastructure based on scientific evidence of cycling demand. We also intend to apply data science techniques to understand better the bicycle dynamics in the city of São Paulo and its metropolitan region.

This report continues in Chapter 2 describing the data sources and technologies we utilized during the model development. Chapter 3 details the literature research we conducted to understand the limitations for cycling and analyze related works. In Chapter 4, we analytically examine the data regarding the cycling environment in São Paulo. Chapter 5 specifies the concepts and implementation of the final model. Chapter 6 relates some validation techniques applied to measure the model's consistency, including the results of a user survey with 89 participants. Before concluding the report, Chapter 7 depicts the results and final analyses that the model propitiated regarding the cycling potential in São Paulo.

Chapter 2

Fundamental Concepts

This chapter describes how this work relates to the *BikeScience* project and details some technical aspects, e.g., the data source structures and the used programming languages.

2.1 BikeScience

The *BikeScience* is a data science tool that aims to build open source solutions to promote cycling in urban contexts. It is part of the *InterSCity* research project (interscity.org), which supports the "development of robust, integrated, sophisticated applications for the smart cities of the future".

The tool has several functionalities to analyze cycling data geographically and statistically. We highlight the analyses of *flows*, i.e., abstractions to report how many bike trips occur between two regions (administrative zones or grid cells). Kon *et al.* (2021) used the BikeScience to analyze bike-sharing systems in a Boston case study. The São Paulo Traffic Engineering Company (CET) also uses this tool for cycling mobility analyses (Oliveira Vianna *et al.*, 2021).

This way, we implemented the model and its software programs within the BikeScience tool. In the same way, the tool's software background (e.g., concerning map drawing and patterns for data treatment) supported the development of this model.

2.2 Technologies

Next, we list the technologies used for the implementations we will describe in the following chapters.

First, the basis technology for our implementation - as well as the entire BikeScience tool - is the *Python* programming language (www.python.org). *Python* is an open-source interpreted language broadly used for data science applications. We used this language in isolated modules for basic algorithms that many functionalities could reuse. Also, for the final applications with specific contexts, we utilized *Jupyter notebooks* to develop

interactive solutions. The *notebooks* are frameworks developed by the *Jupyter* project (jupyter.org) to support scientific computing.

In the *Python* context, we utilized some established libraries for data science, including *Pandas* (pandas.pydata.org) for data manipulation; *Numpy* (numpy.org) for mathematical calculus; and *Matplotlib* (matplotlib.org) for graph plotting. To support geoprocessing, we adopted the *Geopandas* (geopandas.org/) and *Shapely* (https://pypi.org/project/Shapely/ libraries, which implement geometrical structures (points, lines, polygons) and convenient functions (intersection, area, length, etc.). Also, the *Folium* tool (https://python-visualization.github.io/folium) provided the implementations for map drawings.

To develop of an online user survey (described in Chapter 6), we adapted the *Python* libraries for a web environment. The online form development counted with the traditional framework for front-end HyperText Markup Language (HTML), Cascading Style Sheets (CSS) and Javascript (JS), along with the Bootstrap toolkit (https://getbootstrap.com/) and the Leaflet JS library (https://leafletjs.com/) for the interactive maps. The back-end utilizes the HyperText Preprocessor language (PHP) to save and report the answers using the Javascript Object Notation (JSON) format.

During all development, we used the *git* (https://git-scm.com) and *GitLab*[1] (https://gitlab.com) systems for code versioning.

2.3 OD17

The data source for the analyses and modeling we propose in this work is the 2017 Origin-Destination survey from São Paulo (OD17). These Origin-Destination (OD) surveys are common in many cities to examine their transport dynamics.

In São Paulo, the subway company (Metrô) in partnership with other transport authorities have conducted an OD survey once every ten years since 1967 (METRÔ, 2021). This way, it collects data about the citizens' daily trips (e.g., from home to school, from work to restaurants). The trips' data include origin and destination points (latitude and longitude), vehicle, duration, and other socioeconomic data.

The methodology to obtain the data consist of interviews in domiciliary visits. For each, the survey estimates an *expansion factor* based on statistical concepts. This factor corresponds to the number of trips that the answer represents. For example, suppose that a trip has an expansion factor of 200, then we interpret that 200 trips connect that origin to that destination each day. The survey calculates this expansion factor relating the trips to data that indicates their representativeness, e.g., the number of people using the same bus line or the number of cars passing through one specific avenue. This way, it is essential to ensure that all algorithms and analyses consider this expansion factor in generating consistent results regarding the city's mobility.

[1] The BikeScience GitLab page can be accessed in https://gitlab.com/interscity/bike-science

Chapter 3

Literature Review

Research on modal shift to cycling has grown considerably in the past decade. Much of this work focuses on the effects of newly implemented bike-sharing systems (BSS) (Fishman *et al.*, 2014; Shaheen *et al.*, 2013; Ma *et al.*, 2020). After being implemented, these systems seem to significantly impact the modal share, increasing car-to-bike modal shift by 20% (Fishman *et al.*, 2014).

Along with the analysis of BSS impacts, other studies have investigated which socioeconomic and environmental factors impact bicycle usage. Based on the United Kingdom Census data, Parkin *et al.* (2007) indicate which conditions negatively impact cycling. The author suggests that car ownership, longer distances, steeper slopes, traffic volume, and road pavement quality are relevant for cycling. Tyndall (2020) argues that social and cultural aspects appear to be more relevant to cycling than environmental aspects such as weather or hilliness.

Larsen *et al.* (2013) debate recent studies that compare pro-cycling policies to determine which ones have the more significant potential to increase cycling. The authors concluded that segregated bicycle lanes positively affect cycling increase, and this effect is strengthened when combined with marketing and education campaigns. Cycling infrastructure has also been an object of the literature, focusing on technical aspects, e.g., the best material, signaling, lane width) (AASHTO, 2012).

While there is a considerable debate about the most effective policies to promote a modal shift to cycling and how to implement them, there is little research on prioritizing locations to implement them (Larsen *et al.*, 2013; Silva *et al.*, 2019). More recently, some authors proposed methods to indicate specific regions in a city (e.g., districts, neighborhoods) with better conditions for cycling. These tools intend to define attributes that may impact a modal shift to cycling by determining a *cycling potential* for each area, i.e., where cycling policies would be more effective in promoting this modal shift.

3.1 Related Work

Both academic researchers and governmental organizations developed methods to estimate cycling potential, and the studies vary according to the available data of the studied

region. Next, we describe different proposals to calculate cycling potential analyzed in our research.

3.1.1 Copenhagenize Index (Copenhagenize, 2011)

The Copenhagenize Index is one of the most established to assess the cycling environment of a city (Copenhagenize, 2011; Silva *et al.*, 2019). This Index rates about 600 cities globally, giving points for their efforts towards reestablishing the bicycle as a "feasible, accepted and practical form of transport" and determining the most bicycle-friendly cities. The index has been updated biannually since 2011, and in the last update in 2019, the better-ranked cities were Copenhagen (Denmark), Amsterdam, and Utrecht (Netherlands).

The methodology considers three groups of variables: streetscape, culture, and ambition. The first considers the existence and quality of cycling infrastructure, as well as measures for traffic calming. The second evaluates the gender balance of cyclists, recent modal share increase, safety indicators, and bicycle image (i.e., whether the bicycle is a "respected, accepted and normal" form of transport). The last one considers administrative actions, e.g., urban planning, politics, cycling organizations, and bike-sharing programs.

The Index is a valuable resource to understand and compare success strategies implemented by the top-ranked cities. Also, the multiple analyzed variables provide a broad panorama of the cycling situation. However, the methodology to calculate the points for each variable is not available. The research and assessment are done exclusively by the Copenhagenize Company team. For this reason, it is not possible to reproduce this Index.

3.1.2 Prioritization Index (Larsen *et al.*, 2013)

Larsen *et al.* (2013) proposed the Prioritization Index to estimate where to prioritize the investments in cycling infrastructure in a city. The authors applied this method in Montreal Island, Canada, as a case study. This Index uses five variables to define the priority for areas in the city. Two variables consider existing trips, retrieved from the Montreal OD survey: (i) trips already made with bicycles; and (ii) trips defined by the authors as feasible with the bicycle. The latter are trips made with cars shorter than 75% of the observed cycling trips, corresponding to less than two kilometers for the Montreal context. The authors indicate that future studies could improve this criterion for potential trips, considering other parameters alongside the distance, such as sociodemographic variables.

The third variable is the number of cycling collisions, retrieved from the public automobile insurance agency, and considers incidents of cyclists with all other modes. The fourth is the cyclists' opinion about which roads the municipality should prioritize in deploying cycling infrastructure, obtained from an online survey with 3000 cyclists. Finally, the index considers the connectivity of the cycling infrastructure, calculated by counting the "dangling nodes," i.e., points where bicycle lanes end abruptly.

The method divides the studied region into a 300-meter grid. It then calculates a

"Priority Index" (PI) for each cell, based on the arithmetic mean of percentages for each of the first four variables, i.e., the Priority Index of a cell i is

$$PI_i = \frac{o_i + p_i + col_i + pri_i}{4},$$

where o_i denotes the percentage of observed cycling trips passing through grid cell i, p_i represents the percentage of potential trips, col_i is the percentage of collisions in that cell, and pri_i is the percentage of times a cyclist cited grid cell i in the opinion survey.

For example, if a cell i contains 5% of the observed cycling trips, 8% of the potential trips, 20% of the collisions, and was present in 3% of the cyclists answers, then it receives a value of $x_i = (0.05 + 0.08 + 0.20 + 0.01)/4 = 0.09$. This value defines a "priority percentage" of this cell concerning the rest of the region. The grid cells with a higher value are those with priority to receive cycling infrastructure.

The Index does not aggregate the 'dangling nodes' variable by grid cell. The authors argued that the number of dangling nodes is not sufficient to indicate that a grid cell should be prioritized, since connecting cycling infrastructure makes more sense in some locations than in others, depending on their context. The Index only displays these nodes in the grid, helping transportation analysts visualize the bicycle lanes' connectivity.

3.1.3 Willingness Index (Zhang *et al.*, 2014)

Similar to Larsen *et al.* (2013), Zhang *et al.* (2014) proposed a method to prioritize regions to receive investments in cycling infrastructure. Their approach aims at identifying where, in the studied city, people would be most willing to use a bicycle. The study case was Belo Horizonte (BH), Brazil, where cycling infrastructure has expanded between 2010 and 2014.

They then applied an *ordered outcome model*[1] to the collected data. The model estimated the relationship between the influential variables and the respondents' willingness to commute by bicycle - this estimative generated coefficients for each variable. The higher the coefficient, the higher is the effect of the variable on the use of bicycles. Inversely, the smaller the coefficient, the stronger the variable influences people not to cycle. The model revealed the strongest correlations in three variables: (i) commuting time (higher commuting time had a negative influence); (ii) monthly income (the group with medium household income presented a greater willingness to cycle); and (iii) current mode (walking showed the most substantial positive influence).

The final Willingness Index applied the coefficients, modeled with the survey answers, on the BH districts data. The variables' values, obtained from the Census and OD survey, were multiplied by the coefficients and then summed for each district. Thus, the regions with better values on the variables with higher coefficients had a better Willingness Index (e.g., districts with plenty of trips with lower commuting time).

[1] An ordered outcome model is a linear regression adaptation generally used to process user answers with 7-point scale options, e.g., "strongly disagree" to "strongly agree", where the underlying metric is not necessarily the same as the linear numerical metric (Jackman, 2000).

It is noteworthy that the relations between bicycle use and sociodemographic factors are country-specific and unlikely causal (Parkin et al., 2007; Zhang et al., 2014). Therefore, to apply this method in other cities, conducting a similar survey to collect local data would be necessary.

3.1.4 Steer Cycling Potential Index (Steer, 2015)

Steer is a consultancy that develops tools for complex urban problems. In 2015, the company developed the Cycling Potential Index (CPI) (Steer, 2015) that ranks administrative regions in England and Wales regarding cycling investment effectiveness.

This Index considers three variables, defined by the study as "dimensions": hilliness, sociodemographics, and trip length. The hilliness is calculated as the standard deviation of the heights in the region, using a regular grid of 90 meters resolution. The second dimension – sociodemographic – identifies the predominant lifestyle in the region based on the *MOSAIC*[2] demographic classification. The Index compares the classes with travel survey data to calculate a cycling potential for each lifestyle class. Finally, the authors identified that 88% of work-related British bicycle trips on the local OD survey were shorter than eight kilometers. Based on this information, they defined this distance as a threshold for cycling trips.

The Index ranks the regions based on the weighted average of the three dimensions' ranks. The hilliness and sociodemographic dimensions received a weight of 1, while the trip-length dimension received 0.5. The authors defined this difference to consider that they based this variable on work-related trips, and the travel distance may be longer for non-work-related trips.

3.1.5 Analysis of Cycling Potential (TFL, 2017)

Analysis of Cycling Potential (ACP) is a tool developed by Transport for London – the local government body responsible for most of the transport network in London. It identifies cyclable trips in London based on a dataset from the London Travel Demand Survey from 2012 to 2020. According to the ACP tool, a trip is cyclable if it satisfies the following four criteria: (1) the person making the trip is carrying no encumbrance – for example, heavy tools or pushchairs; (2) the trip length is not longer than three kilometers for those older than 80 years old, five kilometers for those between 65 and 79 years old, and ten kilometers for people between 5 and 64 years old; (3) it is possible to substitute the current mode for the bicycle (is not boat or plane, for example); and (4) the trip is not part of a wider chain of trips that can not be done entirely by cycling.

Transport for London calculated the number of potentially-cyclable trips in each London sub-region using these variables and analyzed their main characteristics. The analysis demonstrated a concentration of cyclable trips in central London. Also, the age, gender, and income of potential cyclists contrasted with those of the current cyclists'

[2] MOSAIC is a customer classification model which segments the population into 66 types. It considers economic, social, cultural, demographic, and geographical aspects (The Audience Agency, 2020).

profiles. Based on this information, they then set key strategies to increase cycling in London.

3.1.6 Propensity to Cycle Tool (Lovelace *et al.*, 2017)

The Propensity to Cycle Tool (PCT) is an online (www.pct.bike) open-source planning support system developed by the UK's Department for Transport, a national department responsible for the English transport network, and a limited number of transport matters in Scotland, Wales, and Northern Ireland. The tool aimed to identify parts of England with a great propensity to cycle.

The PCT divided the studied area (England and Wales) into administrative regions and each administrative region into smaller zones, called MSOAs, with the same number of residents (around 7500 individuals each). It used the UK 2011 Census data to obtain the bicycle trips for commuting between MSOAs, and the online API *CycleStreets* (www.cyclestreets.net) to calculate routes between the zones. They used population-weighted centroids as the reference point to define the start and end of trips.

The tool considers only two variables: routes distance and slope. The model consists of a logistic regression to estimate the influence of the variables in the proportion of cycling. It considers the distance, distance squared, and the distance square root to capture nonlinear influences. Next, for each variable, the regression produces coefficients $p_i \in [0, 1]$ used as weights to estimate the proportion of bicycle use.

The formula $c_i + t_i * p_i$ defines the final "propensity to cycle" P_i of a region i, where c_i is the current number of cyclists, t_i is the total number of trips, and p_i is the cycling proportion calculated by the regression. P_i measures the number of cyclists that the region would have in the future based on the geographical characteristics of the trips.

Regarding the impact of the PCT, Weir *et al.* (2020) reported that 81 organizations had used the tool in some way by 2020, including transport authorities and district councils. In the Northwest and Southeast regions of England, more than 70% of the local authorities have used the PCT.

3.1.7 Cycling Potential Tool (Phillips and Range, 2017)

The governmental organization Cycling Scotland developed the Cycling Potential Tool (CPT) to identify areas in Scotland where cycling investments would bring higher benefits. The tool divides each studied city into a grid of one square meter cells, producing results on a local scale, down to a street by street level.

They used data provided by local authorities and geographic information systems (GIS) and considered eight variables: (1) the existence of physical barriers; (2) population density; (3) slope; (4) distance to cycling infrastructure; (5) average road speed; (6) average distance to work or school; (7) cyclists proportion; and (8) access to services (hospitals, schools, etc.).

Then, the range of values is reclassified to a standard scale from 1 to 10, where 1 represents the lower cycling potential and ten the highest. Finally, they defined the cycling

potential for each cell using the arithmetic mean of all variables. Five Scottish cities used the tool to examine the strengths and weaknesses of existing cycle networks and analyze the impacts of the proposed new infrastructure.

3.1.8 Potential for Cycling Assessment Method (Silva *et al.*, 2019)

Silva *et al.* (2019) developed the Potential for Cycling Assessment Method. The authors based their proposal on a review of the existing literature about the variables that affect cycling. They focused on low-cycling cities that lack cycling tradition and data about cyclists' profiles.

This method does not consider trip data, but it evaluates more than 20 variables for each point in the city. They grouped the variables into three dimensions: (1) target population, (2) target area, and (3) political commitment. The first dimension considers average age, population density, educational level, and car ownership. The target area dimension identifies geographic attributes that benefit cycling, including streets inclination, and distance to educational and commercial facilities. The last dimension evaluates the existence of local policies to stimulate cycling, such as bicycle parking, integration with public transport, and cycling infrastructure coverage.

For each region, the variables received values varying from 1 to 5 (from most to least cyclable), and the weighted average of all variables defines the final potential of the region. The classification and weights were defined based on an extensive literature review on relevant attributes for cycling. Public authorities in eight Portuguese cities applied the Assessment Method. They used the generated maps and cycling potential values to direct investments in cycling.

3.1.9 Increasing Cycling in Canada (Verlinden *et al.*, 2019)

Verlinden *et al.* (2019) developed a guide that reports the effectiveness of different actions to encourage cycling in large cities. Analyzing the cycling potential to identify regions with more significant bicycle demand is one of the suggested strategies. The guide does not provide a precise method to calculate a cycling potential. However, it proposes essential variables for this analysis.

The variables cited as relevant for the cycling potential include: (1) trips shorter than eight kilometers; (2) average number of trips per person; (3) average number of cars per person; (4) cycling infrastructure density; (5) current cycling mode share in adjacent neighborhoods; (6) weather (number of snow days); and hilliness (maximum change of elevation). The authors analyze these variables separately. For instance, in a Toronto case study, the specialists generated a map for the short trips and another for cycling mode share in the city, identifying higher potential regions for each of the two variables.

3.1.10 Data Science Framework (Olmos *et al.*, 2020)

Recently, Bogota received massive investments to support cycling (Olmos *et al.*, 2020), appearing for the first time in the Copenhagenize Index in 2019 in the 12th position

(Copenhagenize, 2011). In this context, Olmos *et al.* (2020) developed an open source framework to identify trips that could benefit from bicycle paths.

The framework considers two variables: trip distance and bicycle infrastructure connectivity. The authors obtained the trips data from popular bike apps. Using a shortest path algorithm and prioritizing bicycle lanes, they generated the routes with the Open Street Routing Machine. It did not consider the route inclination since Bogota is generally flat.

The methodology identifies the proportion of bike trips in each length range (e.g., 8% of trips have two to three km) and uses it as weights in a percolation model. The model considers bicycle infrastructure connectivity, predicting the behavior of a network when links are added. In other words, a trip has more significant cycling potential if it is short or if it connects isolated clusters of bike lanes.

3.1.11 Nodal Approach (Hitge and Joubert, 2021)

Hitge and Joubert (2021) investigate the Cape Town cycling potential by filtering the population that corresponds the most with cyclists' profiles worldwide. The authors retrieved the profiles from a literature review on the individual factors that influence choosing bicycles for transport. The existing profiles indicate that the potential cyclist (1) is studying, (2) belongs to low-middle income economic classes, (3) does not live in an informal settlement, and (4) has no access to cars. Nonetheless, the authors consider that it is conservative to consider only the most probable people that would migrate to cycle.

After, each potential cyclist receives a weight considering the gender, age, and the number of family members, that defines the probability of adopting bicycles. For example, suppose an individual receives a 0.7 weight. In that case, it means that he/she will be considered a potential cyclist with a 70% chance of choosing the bicycle. This individual approach shows where the likely cyclists are concentrated, grouping the regions (called nodes) by interest points (e.g., schools and hospitals). It identifies the nodes that would benefit from investments in cycling mobility.

3.2 Method Comparison

In this section, we summarize the main similarities and differences among the methods presented in Section 3.1. We analyze and compare three aspects of the studied works: the scale, the used variables, and the methodologies to aggregate the variables. We will also emphasize some ideas that influenced the index developed in this work. In the end of the Section, Table 3.1 sums up the characteristics of each Index we analyzed.

3.2.1 Scale

The first comparison we propose here is regarding the scale of the methods. We can group the studied works in three geographical levels: macro, mezzo, and micro. The macro-level refers to a more comprehensive analysis of the cycling conditions within a nation or

an entire city. On this level, there is no indication or prioritization of policies or places to implement them. The mezzo-level is more detailed, bringing the analysis to areas within the cities. This level may serve as general guidelines to understand cycling dynamics and demands within the municipalities.

The micro-level presents the highest level of detail. Looking at the street level, it provides a more nuanced perception of possible interventions. This scale allows transport authorities to identify streets inside zones with more significant cycling potential, which would benefit the most from policies. This information may inform local authorities of locations to implement bicycle lanes and parking, for example. In this sense, it may guide public authorities in their investments. The model described in Chapter 5 adopts this scale, dividing the studied city into administrative zones.

Examples of macro-scale are the Copenhagenize Index, the Steer CPI, and one of the PCT functionalities. Figure 3.1a exemplifies the latter. These methods do not aim at specifying cycling potential on streets or neighborhoods. Instead, they show macro-regions with generally good conditions for cycling, e.g., large flat areas, bike-friendly policies, and proper sociodemographics characteristics.

The Willingness Index, applied to the city of BH (Figure 3.1b), Verlinden *et al.* (2019), and the ACP tool are examples of methods that adopt the mezzo-scale, identifying cycling potential inside municipal limits. Some of these tools partition their studied cities into smaller areas or districts. The PCT, in turn, divides the zones into MSOAs (zones with a similar number of citizens), and Hitge and Joubert (2021) divide Cape Town by interest points. Both approaches work similarly to the district partition.

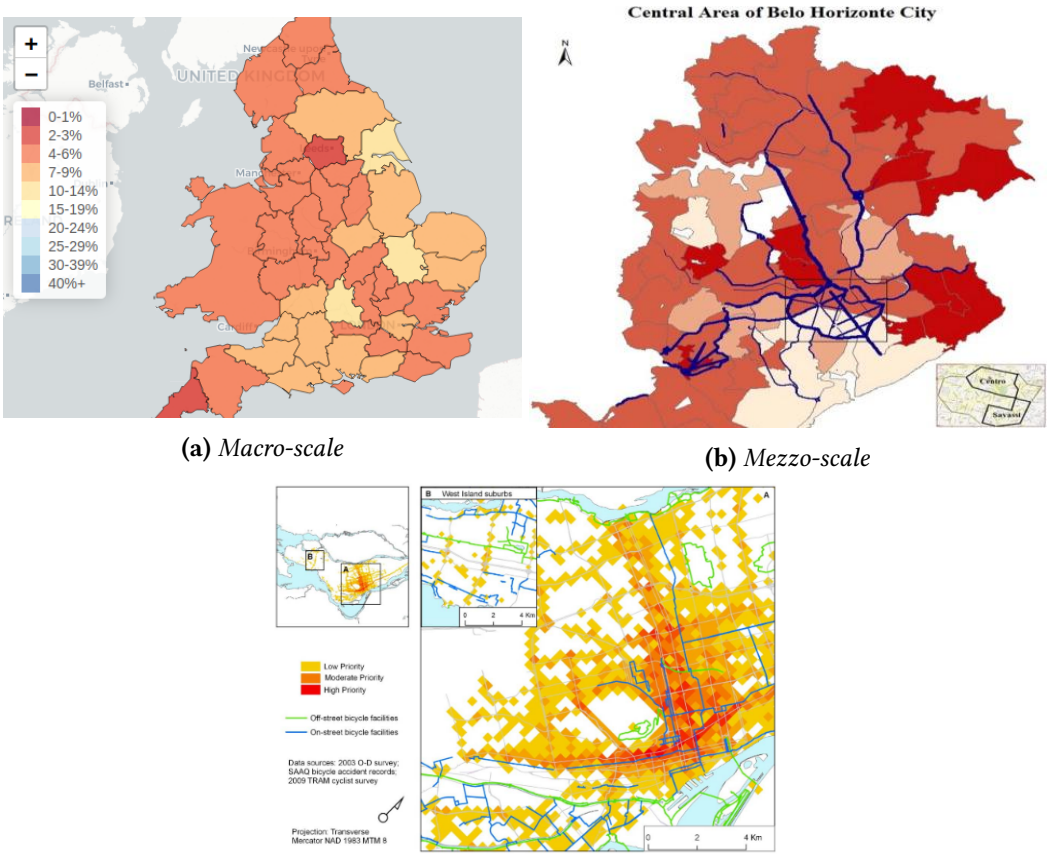
As for the micro-scale, the Prioritization Index, shown in Figure 3.1c, and the CPT divide the study area into grid cells. The Assessment Method considers a continuous model for the geography of the studied cities, making precise maps of cycling potential (Silva *et al.*, 2019). Although this approach requires very detailed data sets and more processing power, it can produce the most accurate models.

3.2.2 Variables

Lovelace *et al.* (2017) categorize the variables considered by the related work into three classes: individual-based measures (demographic data), route-based measures, and area-based measures (environmental data). Even if it is possible to use variables from all classes, as in the Assessment Method (Silva *et al.*, 2019), the methods generally focus only on one of the classes.

The individual-based variables include *income*, *car ownership*, *educational level*, *age*, and others. This type of personal characteristics seems to be relevant for the choice to cycle (Tyndall, 2020), so these variables may be helpful to locate people that would likely adopt the bicycle to commute. The proposals of Zhang *et al.* (2014) and Hitge and Joubert (2021) are examples of this approach.

However, these individual variables are rarer since transportation data sources usually do not aggregate demographic information. Besides, using this type of data to indicate cycling potential may favor social groups that already use bicycles rather than new cyclist



(a) *Macro-scale*

(b) *Mezzo-scale*

(c) *Micro-scale*

Figure 3.1: *Example of methods with different scales. (Source: Lovelace et al. (2017), Zhang et al. (2014), Larsen et al. (2013))*

profiles.

The most common route-based variables are the *distance* and the *slope*. The *distance* variable appears in 9 of the 11 studied methods, being the most frequent one. This is an expected fact, since there is plenty of evidence that people tend not to use the bicycle to travel longer distances (Parkin *et al.*, 2007; Broach *et al.*, 2012). Besides, it is relatively simple to obtain this data, using online services to trace routes between any two points, e.g., origin and destination of trips, two regions in the city, etc. An alternative to calculating routes is to consider the time spent on the trip. This information is easier to obtain while working with user surveys as in the Willingness Index (Zhang *et al.*, 2014). The best distances for cycling considered by the works are concentrated between 2 and 5 km, and TFL (2017) and Verlinden *et al.* (2019) consider that trips up to 8km also could migrate to bicycle.

The *slope* variable is also very relevant to cycling potential (Parkin *et al.*, 2007; Braun *et al.*, 2016). However, it is usually complex to process this data. Geoprocessing is necessary to localize hilly terrains, highlight steeper inclination areas, and determine the appropriate granularity. Steer (2015) and Phillips and Range (2017) use the standard deviation of altitudes as the slope measure, Lovelace *et al.* (2017) calculate the average slope of a trip, and Silva *et al.* (2019) uses the slope as a parameter to calculate the time necessary to

travel a route.

The last class of variables - area-based measures - consists of environmental information that cannot be grouped by people or trips, e.g., cycling infrastructure, demographic density, proximity to interest places. The Prioritization Index and the CPT use mainly these area-based variables.

Within this concept, some methods analyze the existing cycling conditions of the studied region. For example, it is known that better *cycling infrastructure* is a strong stimulus for trip migration (Parkin *et al.*, 2007; Broach *et al.*, 2012; Braun *et al.*, 2016), thus this variable is considered by seven methods. Similarly, regions that already have a higher *cyclist proportion* may indicate the existence of other implicit factors benefiting bike use.

Other variables are less commonly used, mainly because the available data vary among the studied cities. Additionally, using a large number of variables or very specific data may prejudice the reproduction of the model for other regions.

3.2.3 Methodology

Another comparison we can make is concerning the methodologies each study applies to assemble all the variables. We can arrange these methodologies in three classes: filters, weighted means, and regression models. The filters methodology is a binary approach. It defines a set of conditions that trips or people should satisfy to be considered cyclable. This methodology is convenient when the study sets specific targets, such as when studying cycling potential within a particular transport mode or age group.

The weighted mean and the regression models methodologies are similar. Both ponder the variables by weights. The weighted mean approach defines these ponderers based on technical decisions. This methodology aims to appraise the importance of each variable in the studied context or to compensate for problems with data reliability.

On the other hand, the regression models methodology defines the weights of the variables based exclusively on the available information. It designs models to quantify the relevance of each variable to the current cycling activity. For example, suppose the provided data reveals that most cyclists have a high educational level. The regression model will then find a strong correlation between the use of bicycles and the educational level variable. Thus this variable will automatically receive a higher weight.

Although these models provide solid statistical evidence for their analysis, they require substantial data so that the regression produces reliable and consistent results concerning the importance of the variables. Also, they may privilege regions that already have good cycling conditions since the basis data references the current cycling situation.

The ACP tool and the Nodal Approach described in 3.1.11 apply filters to aggregate their final cycling potential. The first filters trip attributes (e.g., distance and transport mode). The latter applies filters by individual characteristics (income, educational level, car ownership) and then combines with a second probabilistic filter using gender, age, and the number of family members as probabilities.

The Prioritization Index, Steer CPI, the CPT, and the Assessment Method by Silva *et al.* (2019) are examples of the weighted mean methodology. In the Assessment Method, each of the 22 variables receives a weight varying from 1 to 5. The variables with higher weights were those with more robust literature evidence of their stimulus to bicycle use. The Steer CPI considers lower importance for the distance variable due to the trips' biased source. The CPT tool uses a simple arithmetic mean of the ten variables, i.e., all variables have equal weight.

Lastly, the Willingness Index, the PCT, and the Data Science Framework by Olmos *et al.* (2020) apply regression models. The Willingness Index used this methodology to process the answers from a user survey. The PCT ran a regression model to calibrate the weights of the distance and slope variables based on OD data from the UK Census. Similarly, Olmos *et al.* (2020) used a percolation model (regression adaptation) to associate the trip length variable with the cycling infrastructure connectivity, combining the modeled weights for these variables.

		Methods										
		3.1.1	3.1.2	3.1.3	3.1.4	3.1.5	3.1.6	3.1.7	3.1.8	3.1.9	3.1.10	3.1.11
Scale	Macro	✓			✓	✓						
	Mezzo			✓		✓	✓			✓	✓	✓
	Micro		✓					✓	✓			✓
Method	Filters	NA				✓				NA		✓
	Means	NA	✓		✓			✓	✓	NA		
	Regression	NA		✓			✓			NA	✓	
Individual	Age					✓			✓			✓
	Income			✓								✓
	Gender	✓										✓
	Car ownership			✓					✓	✓		✓
	Education			✓					✓			✓
	House style			✓								✓
	Family Size			✓								✓
	Number of trips									✓		
	Life style				✓							
Route	Trip Distance		✓	✓	✓	✓	✓	✓	✓	✓	✓	
	Encumbrance					✓						
	Vehicle modal			✓		✓						
	Average speed	✓						✓	✓			
	Slope				✓		✓	✓	✓			
Environmental	Physical barriers							✓				
	Cycling network	✓	✓	✓				✓	✓	✓	✓	
	Cyclists share	✓	✓				✓	✓		✓		
	Cyclists opinion	✓										
	Safety	✓	✓						✓			
	Popup. density							✓	✓			
	Job diversity								✓			
	Public transport								✓			
	Interest places							✓	✓			✓

Table 3.1: *Comparison of the studied methods to calculate cycling potential. The variables grouping follows the classification proposed by* Lovelace *et al. (2017)*

Chapter 4

Statistical Analysis of the Bicycle Usage in São Paulo

The use of bicycles for commuting may present some limitations. In this sense, which trips are cyclable depends on certain conditions of the trip and the person traveling. How many kilometers do people travel by bike? What are the age groups most inclined to cycle? Are there socioeconomic barriers suppressing the use of bicycles? The answers to these and other similar questions vary with location (Parkin *et al.*, 2007; Zhang *et al.*, 2014). For example, people may choose to cycle longer distances to avoid slopes in a region with rough terrain. In contrast, bicycles may be less common in higher-income groups in urban areas, where cars are strongly associated with social status.

In this chapter, we analyze daily cycling the São Paulo using the data from the last mobility survey conducted in 2017 (OD17). We then discuss the prevalent patterns that support our model for cycling potential in the city.

The São Paulo metropolitan area (SPMA) is one of the largest urban conglomerates in the world, with 21.5 million inhabitants - 12 million living in the city alone (IBGE, 2021). The OD17 estimated that about 42 million trips occur in the SPMA daily. Like other large cities in emerging countries, the transportation network in São Paulo is saturated and congested. Motorized vehicles (e.g., car, motorcycle, bus, train) are responsible for 67.3% of this total, representing 28.3 million trips. Active modes (walking and cycling) correspond to 13.7 million journeys (32.7%). However, cycling alone corresponds to only 0.37 million journeys, i.e., 0.9% of daily trips (METRÔ, 2021).

This relatively small share of bicycle usage in the city has diverse explanations. São Paulo has often been characterized as a non-bicycle-friendly city due to its heavy traffic, hilly topography, and unstable weather condition (Benedini *et al.*, 2020; Malatesta, 2014). However, other essential factors help to explain this situation. Historically, the local policies have prioritized infrastructure for motorized vehicles (Malatesta, 2014; Lemos, 2021). The municipality started to implement cycle infrastructure in the past decade, which the 2013-2016 administration intensified. Until 2012, São Paulo had built up to 68 km of segregated bike lanes, while in 2021, there were 684 km (CET, 2021).

Besides, economic and governmental actors frequently block and delay the imple-

mentation of new bicycle lanes, being an object of a conflicting and polarized political context (Lemos, 2021). Other authors also discusses the bicycle's social role in São Paulo, arguing that private motorized vehicles are culturally associated with power and richness status (Sheller and Urry, 2000). At the same time, the bicycle is a cheap alternative for precarious public transport conditions.

The following sections use the OD17 data to discuss the general characteristics of the bicycle users and trips. We also report the cycling infrastructure distribution in São Paulo.

4.1 Cyclists Profile

According to OD17, 102 thousand people make 370 thousand trips by bicycle every day in the São Paulo Metropolitan Area. The survey indicates a reasonably homogeneous group of cyclists in the SPMA: 89.5% are men (Figure 4.1), 50% have a family income lower than one minimum wage (Figure 4.2a), and 57.2% have no cars in the household (figure 4.3a). Some possible explanations for the higher share of male cyclists are (1) the necessity of more intense physical effort due to topography, (2) unsafe and hazardous traffic conditions, and (3) the unequal gender division of the labor, since most of the SPMA bike trips are for work purposes and women are more overloaded with reproductive labor. (Malatesta, 2014; Harkot, n.d.).

According to a survey with 605 cyclists in São Paulo, the newly implemented cycling infrastructure helped increase cycling among women (Benedini *et al.*, 2020). The authors showed that the group that started cycling after the deployment of cycling infrastructure had 40% more women than the more experienced groups. This study indicates that building segregated bicycle lanes is a relevant resource for promoting gender equality among cyclists in a low-cycling environment.

In contrast to gender, figure 4.1 shows that the age distribution among SPMA bicyclists is less concentrated. Most cyclists are between 20 and 40 years old, while the most frequent age group is between 18 and 20, representing 10.48%. This group includes college students that are generally more willing to cycle (Braun *et al.*, 2016). Also, Benedini *et al.* (2020) show that the recently implemented bicycle lanes have stimulated younger people to adopt bicycles. In their research, the group that started cycling after the deployment of infrastructure had 59% more young cyclists (with age between 21 and 40 y/o) than the groups that cycle before this implementation. As for women, the most frequent age group is between 25 and 30 years old. The lack of safety may discourage less experienced women, while maternity demands might keep older women from cycling. The OD17 also indicates that cycling decreases with age: 38% of cyclists are 15 to 30 years old, 30% are 31 to 45 years old, 22% are 46 to 60 y/o, and only 2% are older than 61.

Concerning socioeconomic attributes, the bicycle is most frequently used by low-income cyclists. Malatesta (2014) indicates that these individuals tend to avoid the precarious conditions of the transit system and adopt the bicycle as a cheaper alternative to motorized vehicles. Figure 4.2a presents the distribution of cyclists according to family income. The OD survey indicates that 50% of bicyclists have a family income lower than R$ 1.000 per month (U$ 301 in 2017).

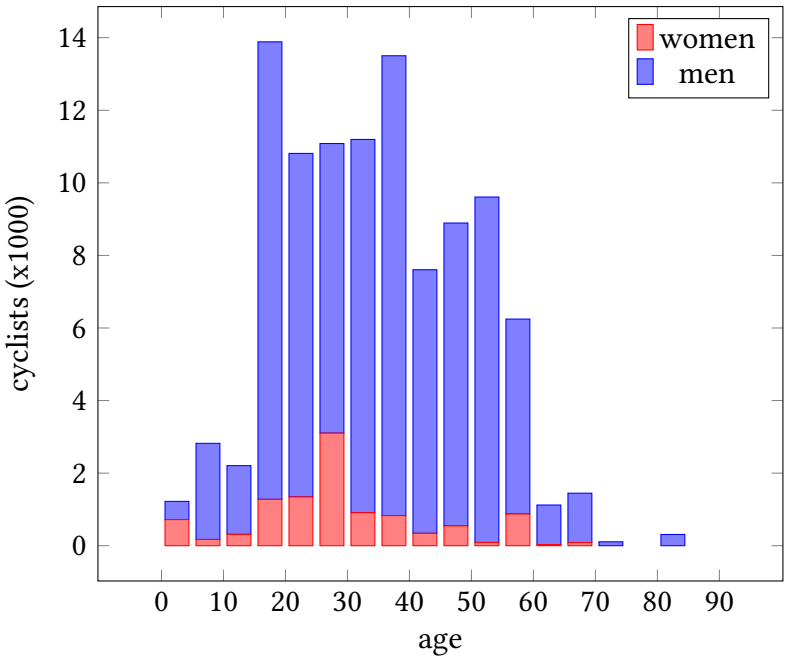
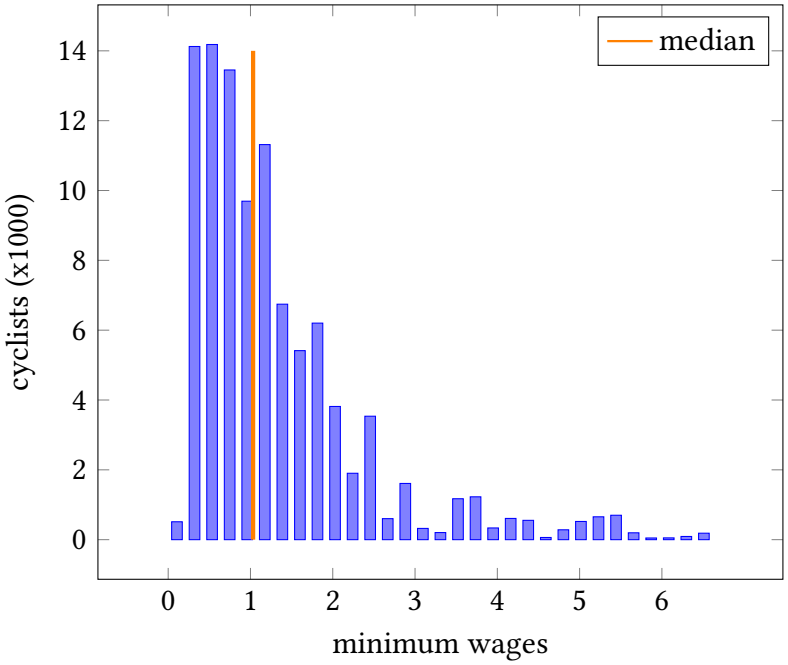
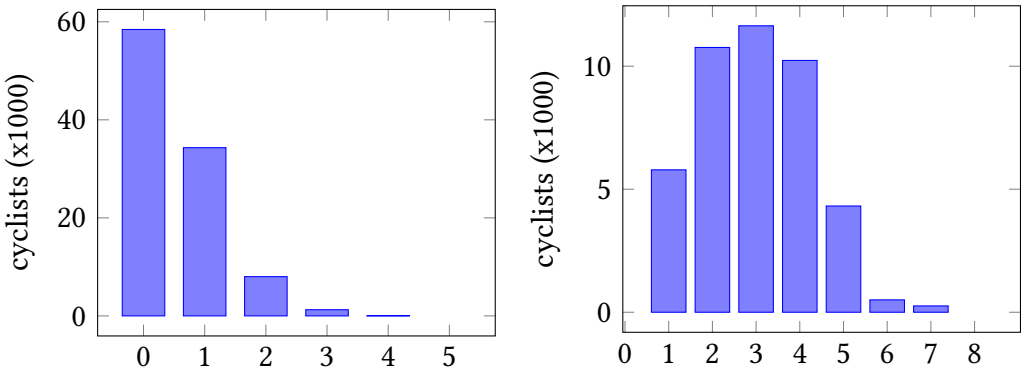


Figure 4.1: *Cyclists age by gender.*



(a) *Monthly family income divided by family members.*

Following the low-income predominance, the share of individuals without a car is 57.2%, as shown by Figure 3.1c. When we consider the number of cars per family member, access to these vehicles is even scarcer, as illustrated by figure 3.1d. Among the individuals possessing car(s), 61.8% share one car for three or more members, while only 13.6% have one or more cars per family member. Thus, even when the families have a car, some family members must use other modes, such as bicycles. Benedini *et al.* (2020) argue that it is not possible to determine if the bicycle is contributing to a decrease in vehicle ownership, or if this mode is just a lower-cost alternative.



(a) *Number of cars that cyclists possess*

(b) *Family members per car*

Figure 4.3: *SPMA cyclists car ownership*

Another issue that relates to income is the level of education. According to the OD17, the minority of cyclists hold college degrees (15%). In contrast, 44.5% have not finished high school, and 40.5% completed high school. In total, 21.2% of the cyclists are currently studying. The proportion reflects the general educational level of Brazilians and the shares observed for users of other transport modes. Therefore, the educational level of the individuals does not seem to be a proxy for using bicycles in the context of São Paulo.

4.2 Trips Profile

The OD17 provides crucial data to understand the patterns for cycling in the region. This section analyzes the trips' purposes, period of the day, distance, slope, and geographical distribution. Some sectors of society public authorities have historically perceived the bicycle as a vehicle for leisure activities only. Thus, people would only cycle in parks during weekends and not on daily commuting trips. This framework explains why the municipality only built bike paths and lanes inside parks until around 2008 (Lemos, 2021).

Nonetheless, the OD17 indicates that most daily trips (69.2%) are work-related, followed by education purposes (13.1%), as shown in Figure 4.4. Leisure is the third most frequent trip motive, corresponding to 5.1% of the trips. When considering the last trip of the day, 79.5% of cycling trips are returning home. The OD survey considers only working days; hence it does not provide data to compare the bicycle use on non-working days. However, it does indicate that the bicycle is indeed used for commuting.

The period of the day in which cyclists make their trips follows a distribution that is typical for all transport modes: peaks at the beginning and end of the day, shown in Figure

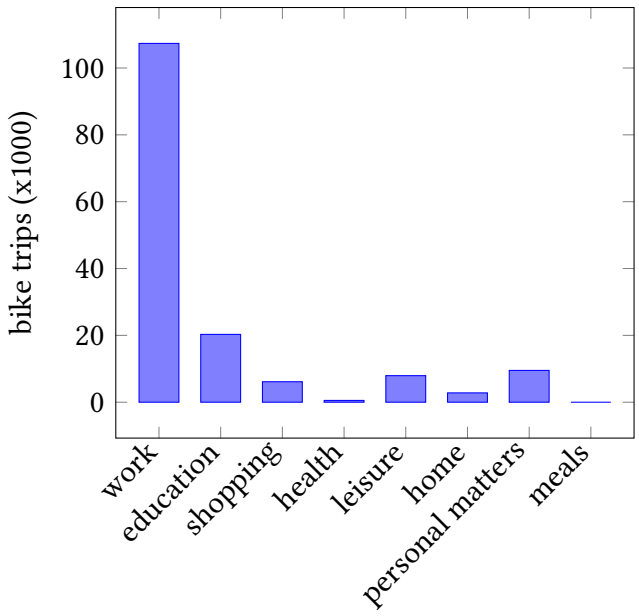


Figure 4.4: *SPMA bike trips reasons (first trip of the day)*

4.5. There are two peaks with most of the trips: (1) between 6 and 9 a.m., corresponding to people going to work or school; and (2) between 5 and 7 p.m., which is the flow of people returning home. We can observe a slight rise in the middle of the day, corresponding to lunchtime. These trips around noon are shorter in the distance, hence this peak is more accentuated when looking at walking trips (METRÔ, 2021).

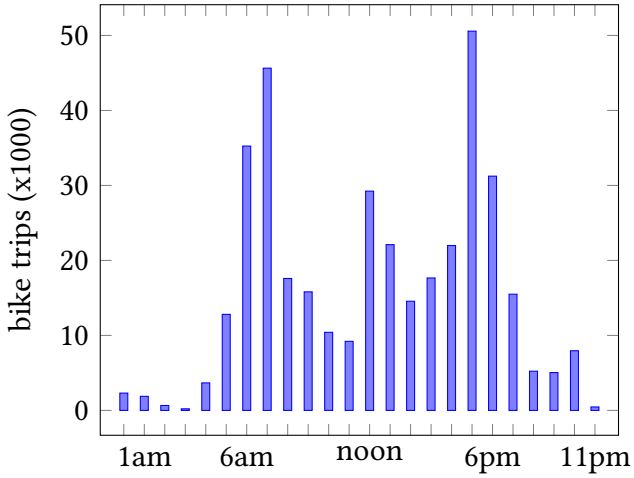


Figure 4.5: *SPMA bike trips period of the day*

4.2.1 Route Properties

We used GraphHopper Routing API (https://docs.graphhopper.com) to define the routes based on the origin and destination coordinates provided by OD17. We then used the itineraries indicated by the tool to analyze the distance and slope studies under the supposition that cyclist had traveled through these routes. As discussed in Chapter 1, these

routes may not represent the exact traveled path but are a better alternative to using a straight line or the shortest path algorithm, because they consider topography and traffic conditions.

Figure 4.6 shows the distribution of bicycle trips distance based on GraphHopper's routes and separated by gender. Most trips presented distances between 1 and 4 kilometers, while fewer trips were under 1 km (12%) and over 9 km (10%). This 1 to 4 km range befits what the studied methods pointed as most favorable to cycling (as discussed in 3.2.2). At the same time, walking is usually preferred for short distances, while motorized modes are used for longer distances.

This 1 to 4 km range reflects the male pattern, with trips more widely distributed. In contrast, women seem to travel shorter routes, 2 km or less. The gendered division of labor explains this difference: women tend to undertake more chained trips, which tend to be shorter, to deal with reproductive demands (McGuckin and Murakami, 1999).

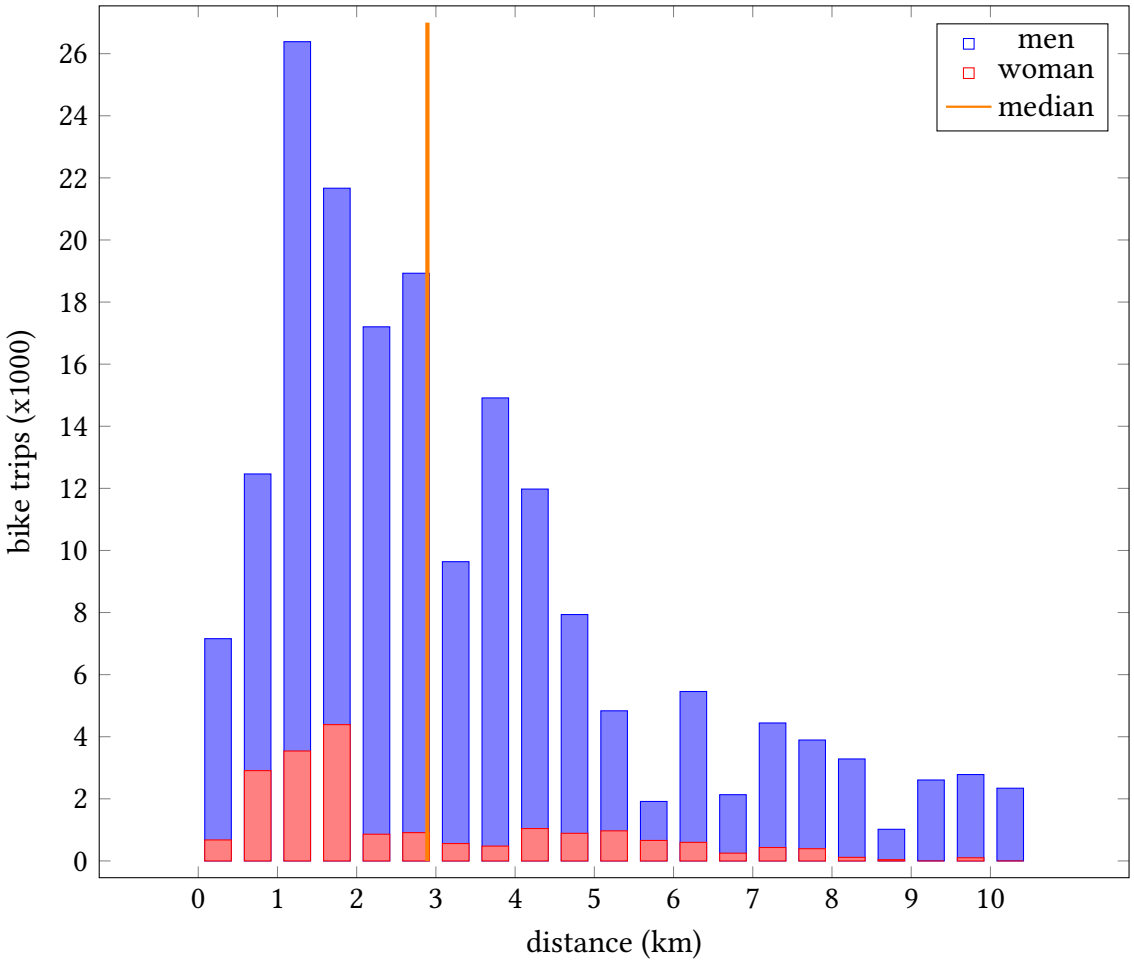


Figure 4.6: *SPMA bike trips distance*

As for the slope of a route, we considered methods applied in other studies. For example, Steer (2015) considers the standard deviation of altitudes: the smaller the deviation is, the closer the altitudes are to each other, thus indicating a flat terrain. Parkin *et al.* (2007) developed a complex matrix method to define slope by comparing the heights of each

cell with the adjacent ones. Tyndall (2020) considers the sum of altitude variation in 10 equally spaced points in a route. Although interesting, this method does not consider how long the slope occurs. For this reason, it may produce similar results for two routes with different topography, e.g., a path with one very steep hill and another composed of a smooth longer ramp. For the case of São Paulo, the method by Tyndall (2020) indicates that bicycle trips decrease with more significant altitude variations (Figure 4.7a).

Lovelace *et al.* (2017) use the mean percentage slope of a route to measure the inclination. use the mean percentage slope of a route to measure the inclination. The percentage slope refers to the relation between the increased altitude, and the distance traveled horizontally. For example, a slope of 5% indicates a 5m gain in altitude for every 100m traveled horizontally. Conversely, a -5% means descending 5m for every 100m.

We applied a similar method to the bicycle trips in São Paulo, considering only the uphill sections, i.e., only the positive slopes. Figure 4.7b presents this result and shows that most routes in São Paulo (70%) have mean gradients of around 1%, which is virtually flat. In comparison, 22.5% have a mean inclination steeper than 2%. The mean slope of a trip tends to evaluate to smaller values (closer to 0%). This concentration happens because uphill parts are generally just a minor part of the journey. In contrast, most of it is flat, which pushes the average to 0. If we consider negative slopes, the concentration next to 0% would be more intense, since downhill segments balance the positive slope values.

Another approach to analyzing the slope impact on cycling is the maximum slope (in percentage) of the bicyclists' routes. Figure 4.7c shows this data about the bike trips in São Paulo. The maximum slope of most trips is less than 10%, indicating that people would not use bicycles in routes with slopes steeper than that. However, the maximum inclination analysis does not provide details about the length of the slopes, i.e., it may correspond to a lengthy uphill street or just a short segment.

The methods we presented to analyze the trips' slope provide essential insights. When applied to São Paulo bike trips, they consistently indicate that cyclists more commonly use bicycles in flat areas. However, it is crucial to relate the distance factor to the hilliness, i.e., the inclination of one specific uphill may be cyclable or not, depending on its length.

To bring together the distance and the slope variables, we propose another approach to the slope analysis, adapted from AASHTO (2012). It considers how many meters the person traveled in each slope percentage along the route. With this approach, we can differentiate a path with a very steep hill from a not very inclined ramp and a long slope from a short one. AASHTO (2012) aims to orientate technical properties for bicycle lanes in the USA. It defines distance limits for each slope percentage, e.g., in a 6% slope, cyclists would travel up to 240m, otherwise they would use other vehicles. We collected the bike trips' slope data to better understand these limits in São Paulo.

Figure 4.8 shows the distribution of distances traveled on each slope percentage (in the form of boxplots). As expected, it indicates that people tend to travel shorter distances on steeper gradients. For example, 50% of trips present more than 200 meters on the 2% slope level. In contrast, for the 5% slope level, only 10% of the trips ride this distance.

Additionally, Figure 4.9 presents the mean distance traveled in each slope level by the bike trips in São Paulo. It reveals that the proportion of uphill and downhill stretches are

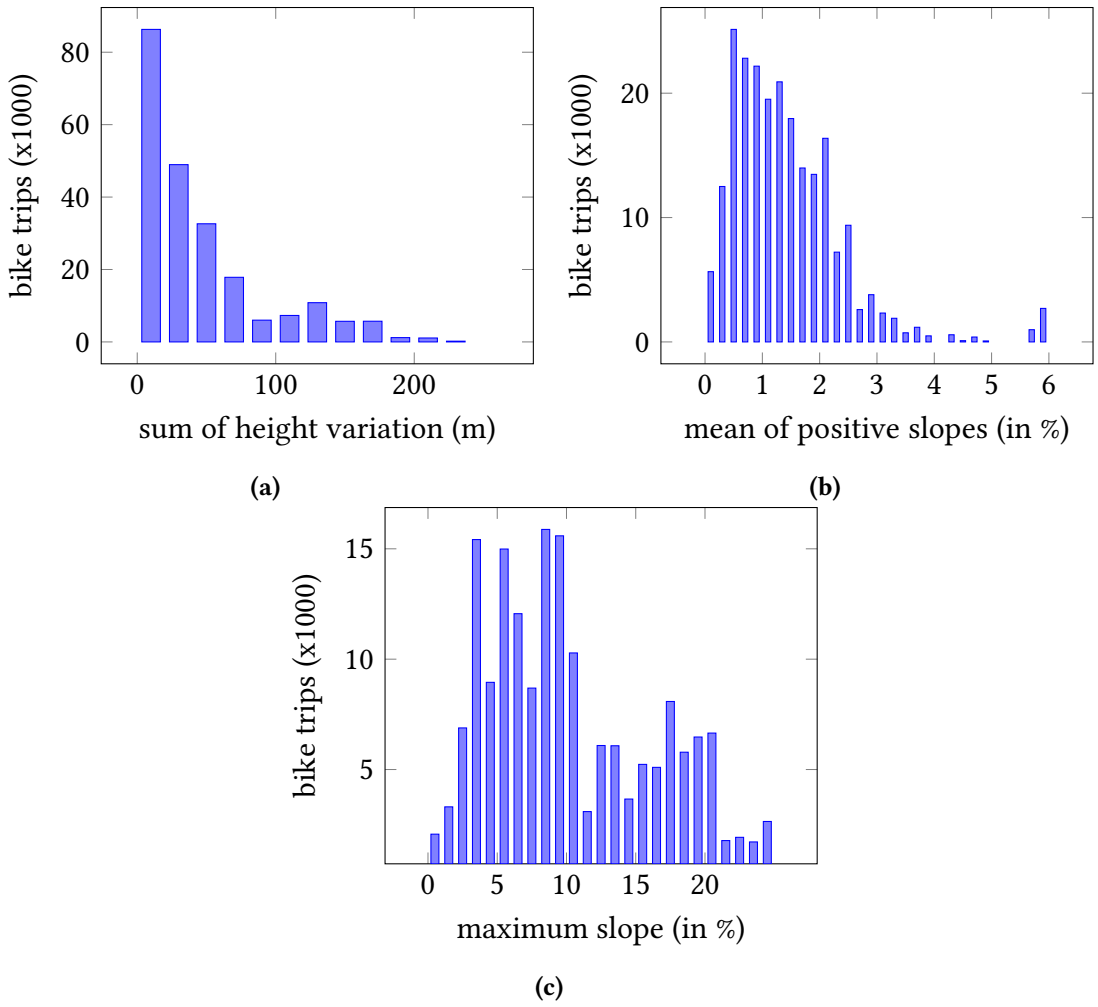


Figure 4.7: *Slope analysis proposals*

similar. This similarity may occur because most cyclists might duplicate the routes in opposite directions, e.g., using the same route to work in the morning and returning home later. We will use this analysis (meters traveled per slope percentage) to model the cycling potential based on a route topography.

4.2.2 Geographical Distribution

Some districts of São Paulo and cities in the metropolitan region present a more considerable use of bicycles. Figure 4.10 depicts a map with the number of bicycle trips in the SPMA municipalities. It shows that the cities east of the city of São Paulo present more bicycle trips. *Guarulhos*, a city in the northeast, concentrates the most number of trips. This city is the second most populated in the SPMA and is located on the floodplains of the Tietê River.

Figure 4.11 presents only the districts inside the city of São Paulo. We observe an intense concentration of bicycle trips in the central-west districts of *Itaim Bibi* and *Pinheiros*, with 25 and 21 thousand bike trips, respectively. These districts are economic poles of São Paulo, and they concentrate trips with all modes. Besides, as we will see in the following

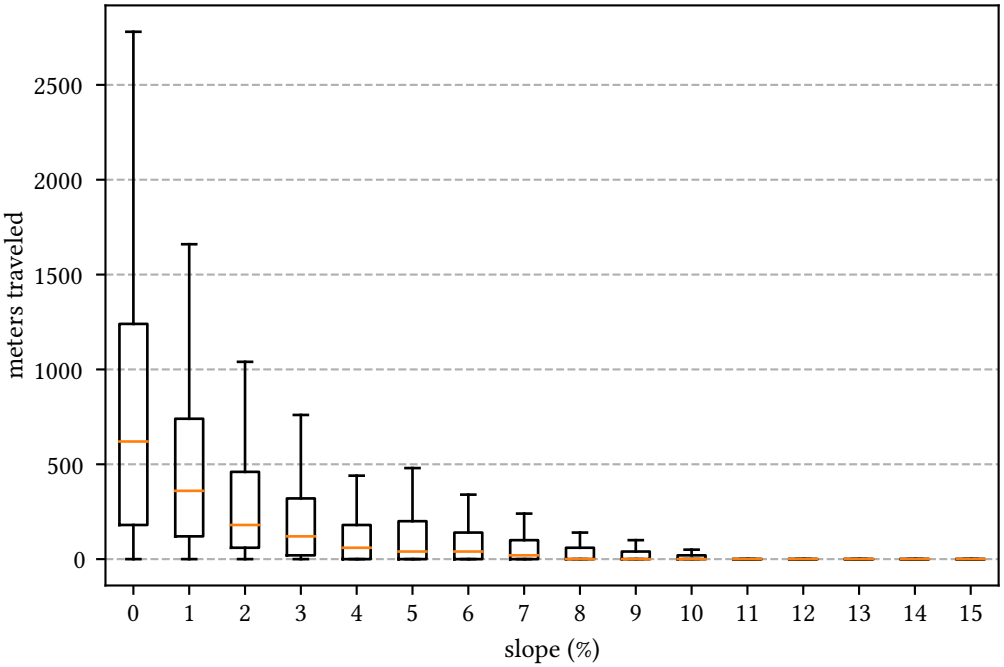


Figure 4.8: *Distance distribution per slope level*

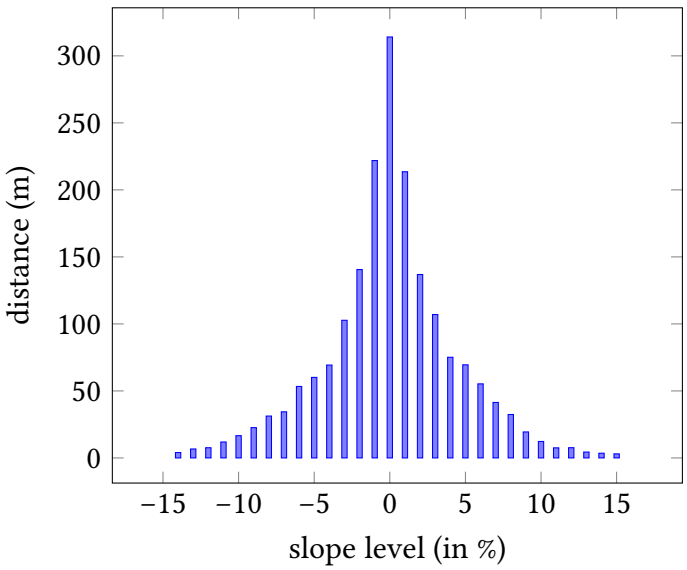


Figure 4.9: *Mean distance on each slope level*

subsection, this region is well served with cycling infrastructure. The deployment of bike paths and lanes seems to have impacted the distribution of trips in São Paulo. As shown by Lemos, Harkot, *et al.* (2017), the cyclyng trips in the Origin-Destination survey carried out in 2007 were primarily distributed in the peripheral areas of the SPMA.

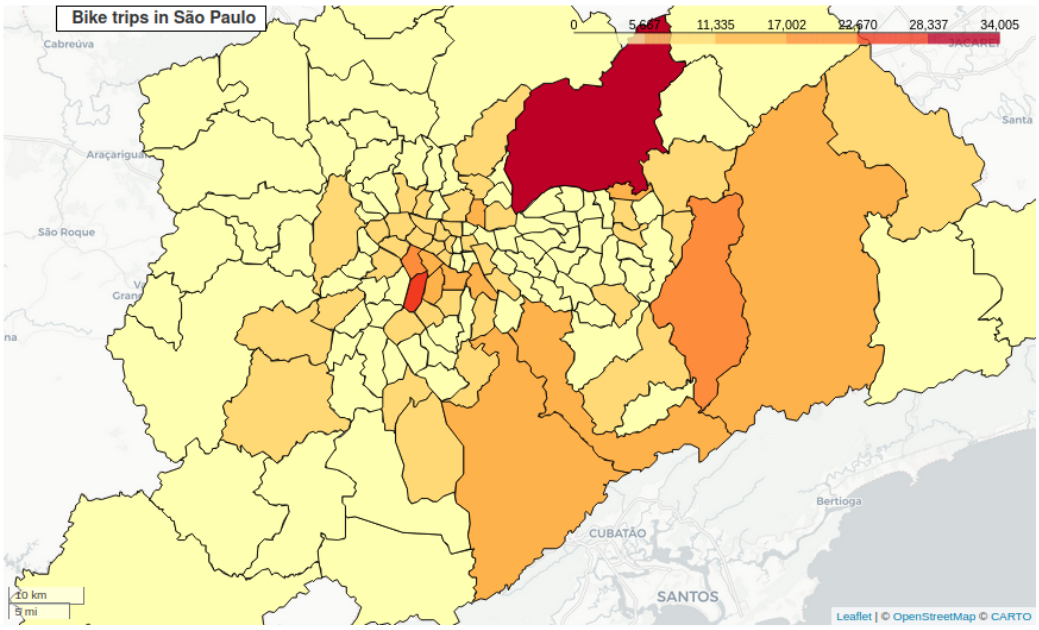


Figure 4.10: *Absolute number of bicycle trips in the SPMA*

Complementing the report about the number of trips per zone, we inspect the density of bicycle trips, i.e., the proportion of cycling to all trips in each region. This way, we avoid over-highlighting populated zones that usually have more trips in general. Figure 4.12 shows that the municipalities in the far-east of the SPMA present a more density of cycling trips. Notably, in *Guararema*, the municipality with the highest proportion of bicycle trips, people use bicycles in 12% of the trips. Inside the city of São Paulo, the districts that proportionally use more bicycles are *Jardim Helena* in the extreme east and *Vila Guilherme* to the north of the central region. They both present around 4% of trips made with bikes (Figure 4.13).

4.3 Cycling Infrastructure

In this section, we examine the situation of bicycle infrastructure in the city of São Paulo as of 2021. Because the only data available was from the municipal traffic engineering company (CET), the analyses presented here will be limited to the city of São Paulo and will not include the metropolitan region.

We will consider two types of cycle infrastructure for this analysis: protected and unprotected. The first type segregates cyclists from motorized traffic. It is either built on the sidewalks or the median strip of avenues. The second consists of a painted lane on the roadbed, thus sharing the same physical space used by cars. We did not consider shared lanes, also called “bike routes”. These interventions consist of specific road signs that should only be implemented on streets with low traffic volume and be complemented by

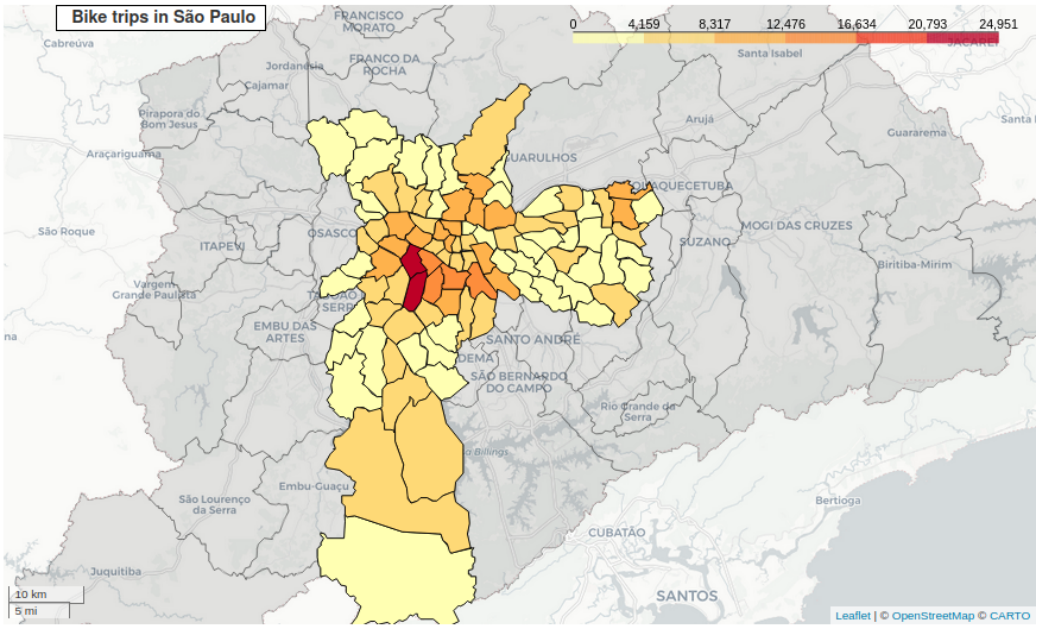


Figure 4.11: *Absolute number of bicycle trips in São Paulo*

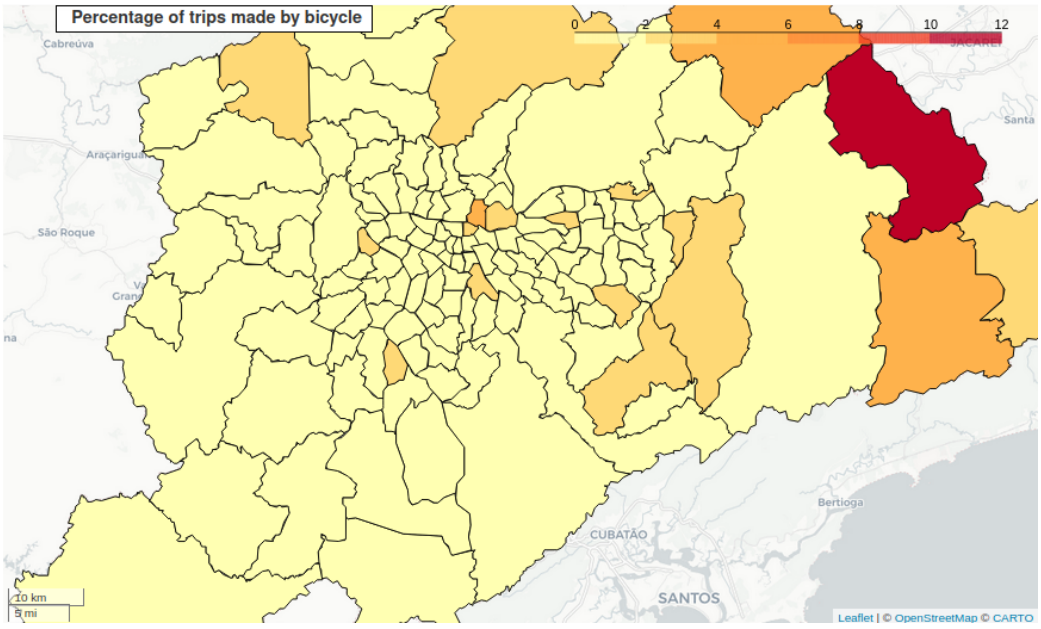


Figure 4.12: *Proportion (%) of bicycle trips regarding all trips in the SPMA*

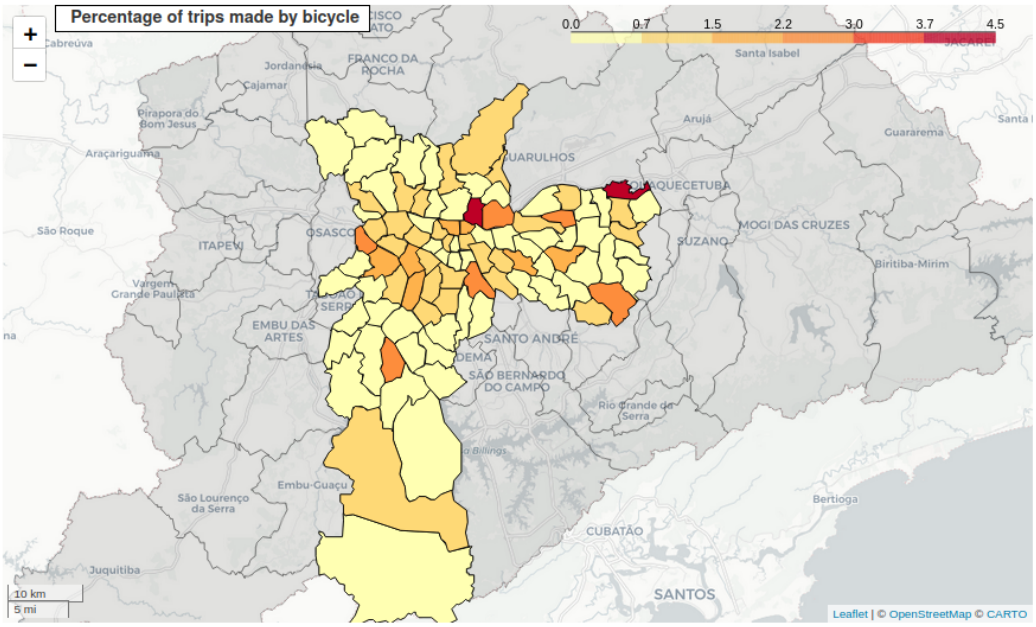


Figure 4.13: *Proportion (%) of bicycle trips regarding all trips in São Paulo*

lowering the speed limits. However, bicycles and other vehicles share the physical space without any segregation. Also, "bike routes" in São Paulo did not necessarily follow these conditions. They were frequently implemented without any intervention other than road signs (Lemos and Neto, 2014).

Since 2013, the municipality of São Paulo has deployed hundreds of kilometers of bike lanes and paths, summing up to 663.1 km (protected or not). Nevertheless, cycling infrastructure is limited in peripheral zones (Figure 4.14). In general, the cycling network lacks connectivity. However, the lack of bicycle connections to the central region from the far north, south, and east regions is remarkable. There are numerous lanes in local streets and avenues that do not connect with other bike lanes. In addition, protected bike paths (in red) are scarcer (9.5% of the total), implemented in a limited number of long avenues such as the ones next to the Pinheiros River in the west (*Marginal Pinheiros* and *Brigadeiro Faria Lima* avenues). Although these paths bring more safety to the cyclist, they are more expensive. Also, they are usually located on the median strip, thus dependent on avenues with such a condition.

Map 4.15 presents the bike lanes per district, counting the length (in km) of the cycling network in each district. This map highlights localities with bicycle lanes along extensive avenues crossing them, such as in the northeast (*Cangaíba* and *Emerlino Matarazzo* districts, contemplating the *Doutor Assis Ribeiro* street). Other regions in the south area (e.g., Santo Amaro, Ipiranga) also have a considerable extension of bike lanes, corresponding to local roads with a denser cycling network. One attention point with this analysis is that the CET geographical data distinguish the stretches of constructed lanes on both roads directions. This way, when we group by district, one street with 5 km of bike lanes on both sides would count with 10 km of cycling infrastructure. Thus the total kilometers of lanes shown in the map nearly doubles the actual extension.

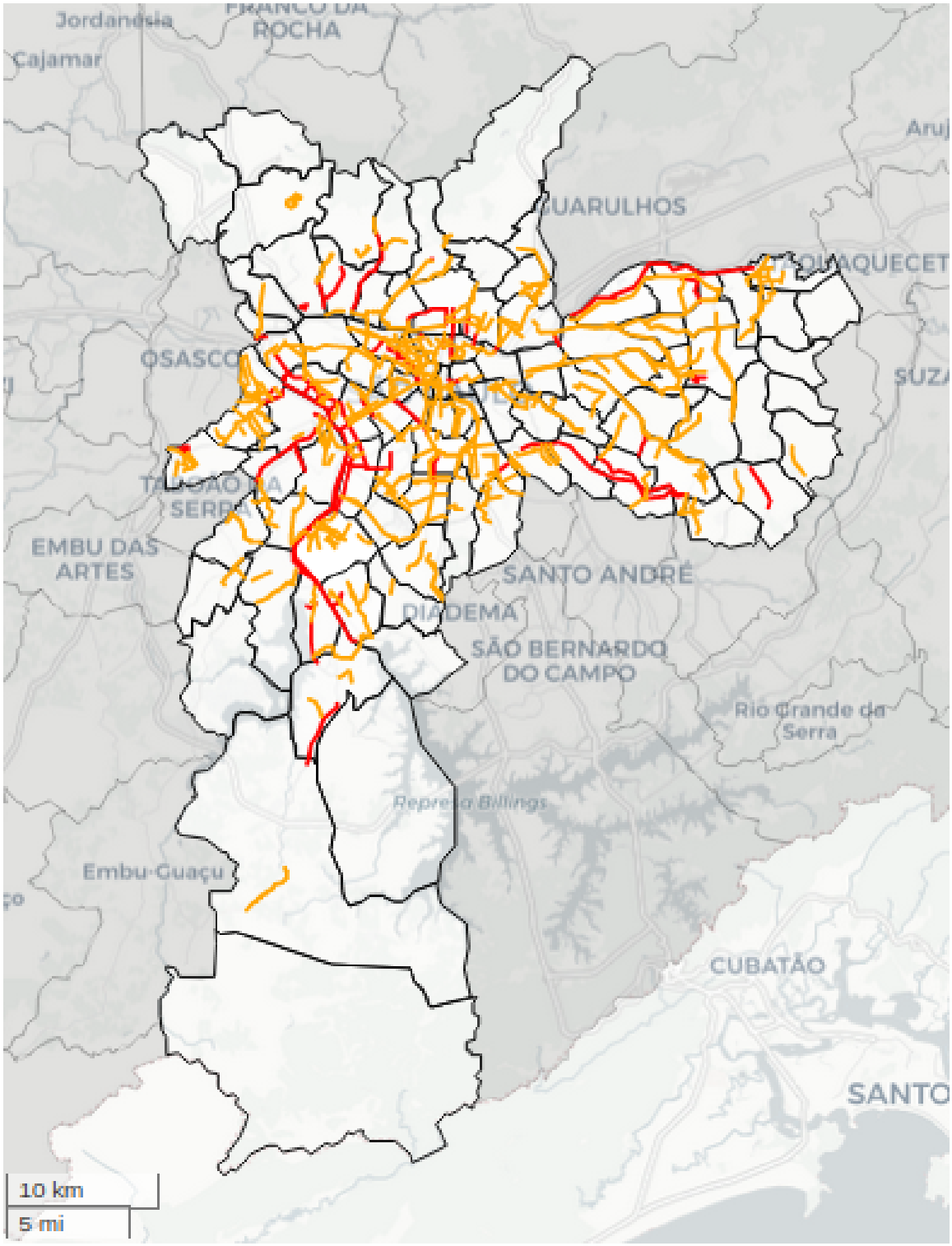


Figure 4.14: *São Paulo cycling network in 2021 (red represents the protected lanes and orange the unprotected bike lanes)*

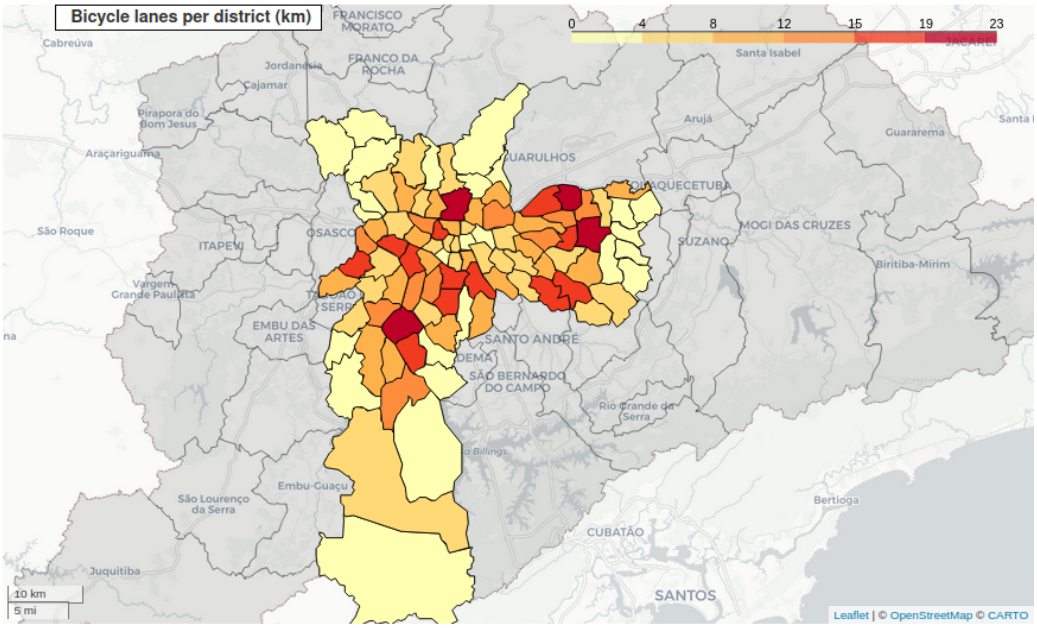
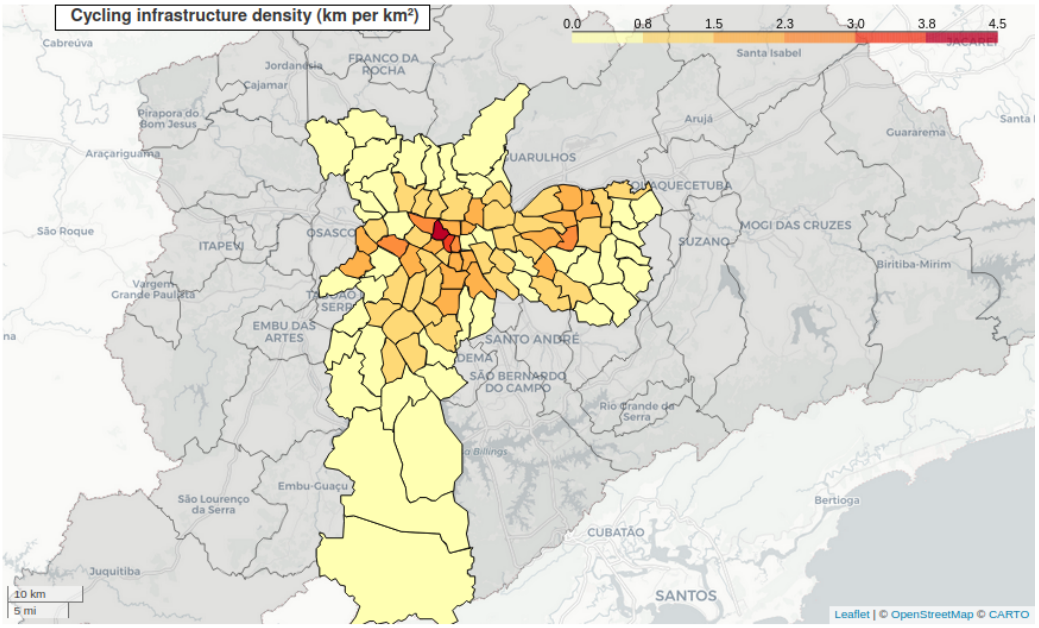
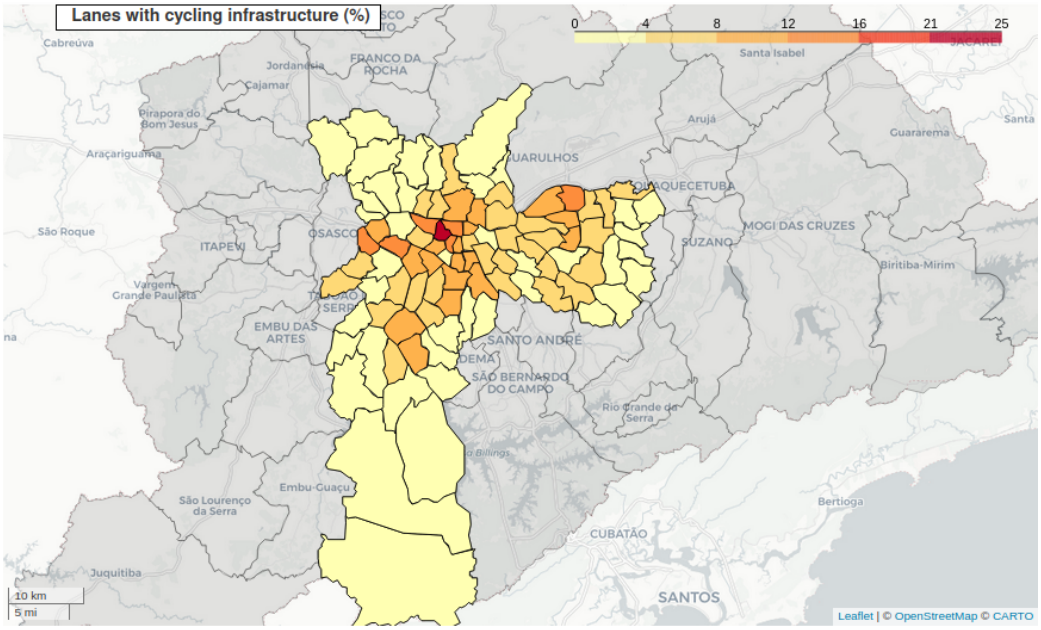


Figure 4.15: *Bike lanes length per district*

Beyond the extension of the cycling infrastructure, it is relevant to analyze the network density among the districts, avoiding the bias of larger districts aggregating more bicycle lanes. Map 4.16a shows the extension of bicycle lanes (in kilometers) divided by the district area (in km2). Another form to analyze the density is the percentage of roads served with bike infrastructure, as shown in Figure 4.16b. These maps show that the central regions aggregate more infrastructure since they have many bike paths and lanes in smaller areas. For example, the *Santa Cecília* region, downtown, is the densest district, with 4.54 kilometers of bicycle lanes per square kilometer, having proper space for bicycles in 25% of the streets and avenues. Appendix A details the cycling infrastructure extension in each district.



(a) *cycling network density by area*



(b) *percentage of roads with bike lines*

Figure 4.16: *São Paulo cycling network density*

Chapter 5

SP Cycling Potential

This chapter describes the variables and methodology we used to model the São Paulo cycling potential. It details the computing methods to process the data and examples of the resulting model. This proposal aims to estimate the *cyclability* of a trip by comparing its similarity with the current profile of bike trips in São Paulo reported by the OD17 survey. If a non-cycling journey is similar to the existing pattern of trips carried out with bicycles, it could potentially shift mode to cycling. In contrast, if its characteristics are too different, the shift would be less likely to occur.

5.1 Variables

One of the first steps to design a model is to define what data this model should consider. As discussed in Chapter 3, previous research examined several factors relevant to cycling potential, emphasizing the following: distance, slope, speed limits, the existence of bicycle lanes, individuals' age, educational level, and car ownership.

This study adopts a more restricted set of variables. We intend to create a replicable model that does not depend on specific complex data. The *SP Cycling Potential* relies upon four variables of a given trip: **distance, slope, age,** and **gender**.

The trip distance is one of the most relevant variables to determine the potential to shift to bicycles (Parkin *et al.*, 2007). Of the 11 methods discussed in Chapter 2, 9 consider this variable. With this variable, we can identify, for example, regions with short motorized trips or longer walking paths. A typical scenario of these cases corresponds to "first/last-mile integration", i.e., trips linking the mass transport networks (subways, trains) to the final destination. When there is proper space for bicycle parking or bike-sharing systems nearby, these integrations frequently represent success cases of modal shift to bike (Ma *et al.*, 2020).

In addition to distance, the slope variable seems to impose a more significant barrier to cycling. Broach *et al.* (2012) tracked the paths of 160 American cyclists during eight months and concluded that they would prefer to cycle 1.73 miles on flat terrain than only 1 mile on a route with a slope of 2% to 4%.

The age variable is the only individual characteristic we will utilize to define the cycling potential of trips. As debated in the previous Chapters, data shows that cycling is more widespread among younger people due to the physical limitations of advanced age. In addition, São Paulo has a very unbalanced age geographical distribution: the average age for districts in the center is around 40 y/o, while in peripheral regions, it is about 30 (SEADE, 2014). Hence, it might be necessary to consider that areas with older populations might have a lower propensity to cycle.

As for gender, we used this information only to categorize the other three variables. For example, because the OD17 shows that the cyclists' average age is different between men and women, our model will indicate values for age depending on gender. This way, we avoid modeling a scenario that privileges only the male patterns since there are much more data about the men's bicycle trips.

Other socioeconomic variables influence the choice to cycle, for example, income and number of cars in the household. Nevertheless, we decided not to use them to broaden the target group of the policies that the SP Cycling Potential will potentially inform. For example, although low-income people tend to cycle more, we aim to promote cycling among all economic classes. For this reason, the tool considers only the physical limitations to cycling, including distance, slope, and age, for each gender.

Similarly, we did not consider the existing cycling infrastructure to identify potential trips regardless of proximity to bike lanes or paths. The tool's objective is not to understand the impact of infrastructure on the propensity to use the bicycle but rather to indicate places where the municipality should build infrastructure to stimulate a modal shift. Nonetheless, we will compare the modeled cycling potential with the existence of cycling infrastructure in Chapter 7.

5.2 The Model

The methodology we designed to calculate the cycling potential of a trip consists of modeling continuous functions that approximate the OD17 data, separating by gender. The resulting method is more complex for the slope model since it is a two-dimensional variable (the slope level and the distance per slope level). The age and distance variables resulted in simpler real functions.

5.2.1 Continuous Distributions

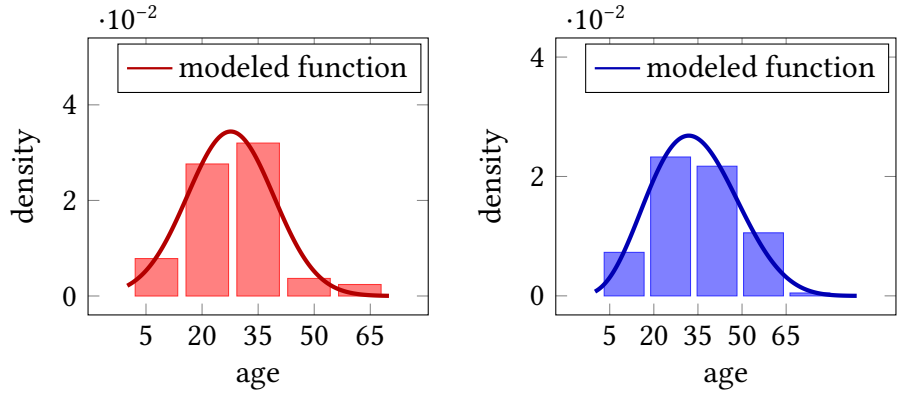
We calculated the probability density functions (p.d.f.) for the age and distance variables that best fit the data distribution. Olmos *et al.* (2020) utilized the same approach to analyze the distance of bike trips in Bogotá.

The BikeScience tool already contains an algorithm implementing this method, used in previous studies. The algorithm uses the Python library *SciPy*[1], which provides many statistical utilities, including 90 continuous probability distributions. This way, it iterates over all p.d.f., testing different parameters and finding the p.d.f. that produced the minor

[1] https://docs.scipy.org/doc/scipy/

difference from the data histogram, using the residual sum of squares technique. The algorithm receives the dataset and the number of groups to separate the data. These groups correspond to the histograms' vertical bars (also called bins) that will be approximated.

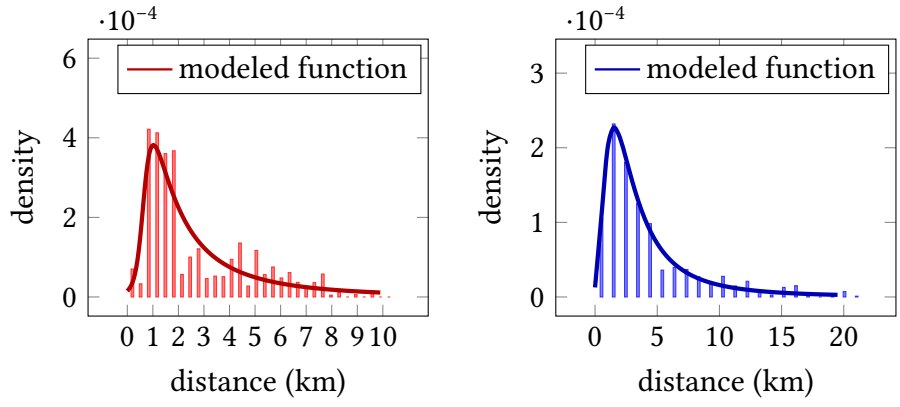
Figure 5.1 shows this continuous modeling for the age variable. The random variable best suiting men's age data follows a *Gauss Hypergeometric* distribution. For women, it was the *Log Gamma* density function. We used five bins for the age variable, producing more regular results. Since the age ranges are small (from 0 to about 80), the lack of female cyclists' data impedes using more bins.



(a) *age modeling for women* **(b)** *age modeling for men*

Figure 5.1: *Age modeling*

Regarding the length of the trips distances, the density functions follow an *Inverted Weibull* distribution for men and a *Johnson* S_U distribution for women. We show the suitability of the functions in Figure 5.2.



(a) *distance modeling for women* **(b)** *distance modeling for men*

Figure 5.2: *Distance modeling*

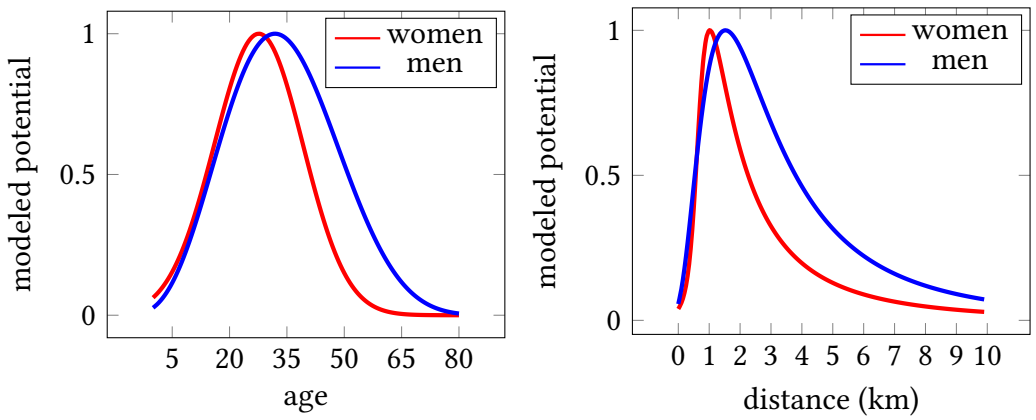
After calculating the continuous distributions, we scale the functions to the interval [0, 1], ensuring that all variables have the same range of values for both genders. Therefore, the smaller values (closer to 0) represent the least frequent data (e.g., older people or longer

distances). The most frequent data, in turn, receive values approaching 1, e.g., the distances up to 2 km.

The scaling process divides each density function f by its maximum value, i.e., we define a new function

$$g : \mathbb{R} \longrightarrow [0, 1]$$
$$x \longmapsto \frac{f(x)}{\max_{a \in \mathbb{R}} f(a)}.$$

For simplicity, we numerically approximated these maximum values by sampling points and shrinking the domain to reach them. Since the density functions are well-behaved functions, we can disregard the error of this numerical scaling process. Figure 5.3 draws the final function graphs for the age and distance variables, scaled to the [0, 1] range.



(a) *Age variable*

(b) *Distance variable*

Figure 5.3: *Final cycling potential model for the age and distance variables*

5.2.2 Slope Modeling

As discussed previously (in Subsections 3.2.2 and 4.2.1), processing the slope of a route or region is a complex task. In the *SP Cycling Potential*, we process each slope percentage separately and then aggregate them to obtain a final value. The idea is assigning lower cycling potential to routes that travel uphill segments, weighted by the distance and the slope gradient. For example, suppose that a route R_1 travels through a flat path, a route R_2 has 50m in an uphill street, and a route R_3 presents 200m in this same slope level. One may conclude that route R_1 requires less effort to cycle than R_2, and R_2 is, in turn, more bicycle-friendly than R_3. Now, suppose that the 50m uphill street of R_2 occurs in an 8% slope level, while the R_3's 200m ascending segment occurs in a slope of 5%. Which route is more likely to attract cycling? We now detail our algorithmic proposal to answer this question based on the São Paulo bike trips data.

We first calculated a continuous function to model the distribution of distance of bicycle trips in each slope percentage. This approach is the same we applied for the distance variable, but now grouping by six inclination intervals:]0%, 2%],]2%, 4%],]4%, 6%],]6%, 8%],]8%, 10%], and]10%, +∞). Figure 5.4 shows the resulting functions for these

slope degrees. The functions in this Figure relate to the boxplots from Figure 4.8. Both Figures show that people travel shorter distances in steeper slope levels.

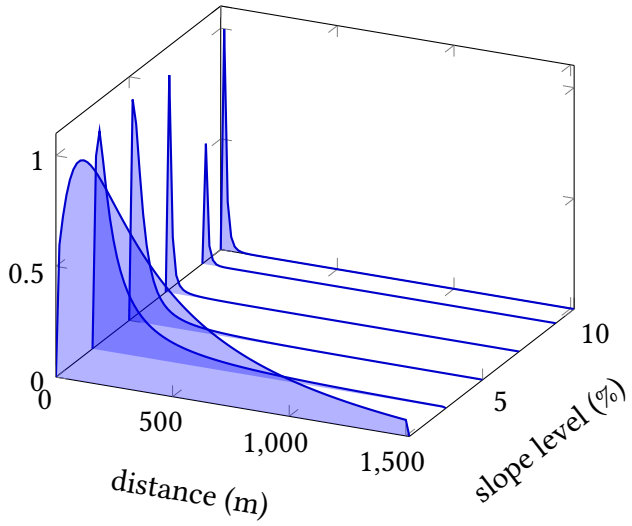


Figure 5.4: *Continuous density functions for the distance traveled in each slope percentage interval*

However, these continuous density functions presented an undesirable property for the slope model: they are not ascending. This way, a route that passes through a shorter uphill distance may receive a lower "cycling potential" than a route with long uphill segments. This way, for each variable X, instead of using the density function f_X, we adopted the cumulative density function (c.d.f.) F_X defined as

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt.$$

Each density function f_X is related to one unique cumulative function F_X that has convenient properties, e.g., ascending and assuming values in the interval $[0, 1]$. Figure 5.5 shows these cumulative functions for each slope interval.

With this approach, a route receiving a 0.5 value means that it presents a shorter distance in that slope level than 50% of the bike trips. Similarly, suppose that a path receives a 0.1 (10%) value, then it would have fewer uphill stretches than 90% of the trips. Hence, the smaller the value, the more bicycle-friendly the route is.

After estimating these continuous approximations, the next step was combining the data from all six slope levels to define a unique "slope potential". We adopt a weighted average for each slope level for this combination, with the slope value as the weights. This way, an uphill stretch on slope 10% would count five times more than if it was on a 2% slope. Figure 5.6 shows the final model values, grouped by gender.

We detail the computational method for this modeling in Program 5.1. The algorithm's input is a set of points $\mathcal{P}$ representing one route and the size of this set $n \in \mathbb{Z}$. Each pair of consecutive points determines a segment of the route, usually corresponding to a street. These segments' mean length is 62 meters, determining the level of detail that this slope processing reaches. The points have three dimensions: latitude (x), longitude

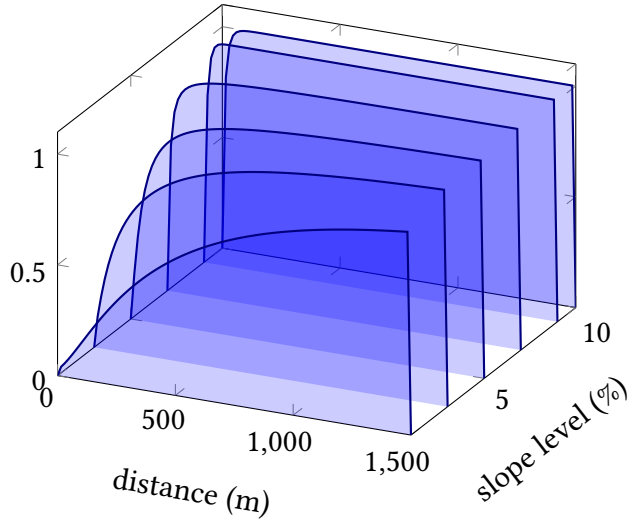


Figure 5.5: *Continuous cumulative distribution functions for the distance traveled in each slope level*

(y), and altitude (z). The algorithm's output is the trip's potential for the slope variable $s(t) \in [0, 1]$. For consistency, we implemented the algorithm so that the values closer to 0 represent worse cycling potential, similar to the age and distance models.

Program 5.1 The Slope Cycling Potential Processing

```
1   FUNCTION Slope_Potential(P, n)
2       slope_levels ← (0, 2, 4, 6, 8, 10)
3       distances ← (0, 0, 0, 0, 0, 0)   ▷ distances travelled in each slope
4       for i ← 2 to n do
5           p1 ← P[i − 1]
6           p2 ← P[i]
7           dist ← sqrt((pi.x − pj.x)^2 + (pi.y − pj.y)^2)   ▷ Euclidean distance
8           slope ← ⌈(pi.z − pj.z)/dist⌉
9
10          for s ← 6 descending to 1 do
11              if slope > slope_levels[s] then
12                  distances[s] ← distances[s] + dist
13      mean ← 0
14      for s ← 1 to 6 do   ▷ for each slope interval
15          weight ← slope_levels[s] + 2   ▷ slope between 0% and 2% has weight 2, and so on
16          potential ← Fs(distances[s])   ▷ Fs is the c.d.f. for the s^th slope interval
17          mean ← mean + potential * weight
18      mean ← mean / 42   ▷ 42 = sum of weights
19
20      return 1 − mean   ▷ invert final value so that 1 indicates good potential
```

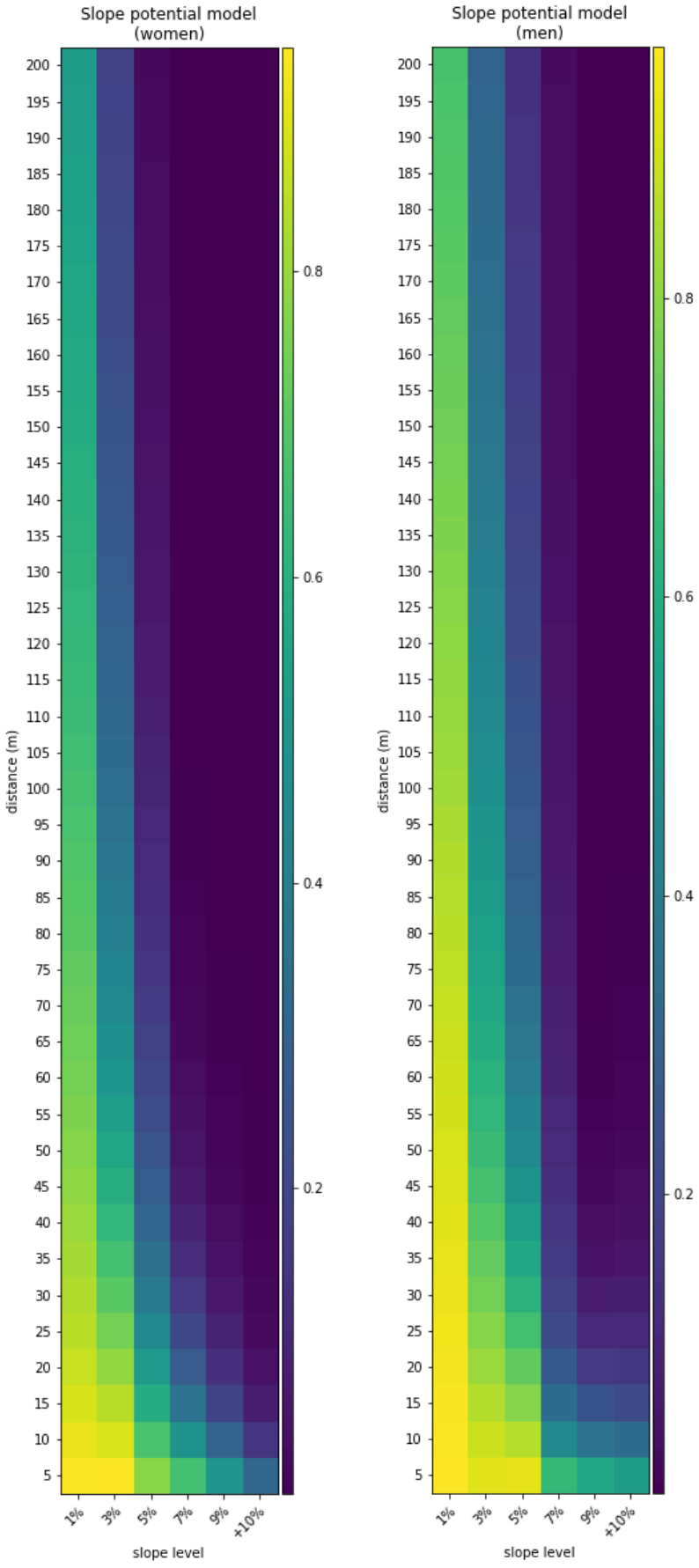


Figure 5.6: *Cycling potential model for the slope variable*

5.2.3 Variables Aggregation

Now, we discuss the methodology adopted to aggregate the three variables for the *SP Cycling Potential*: age, distance, and slope.

Among the three approaches discussed in Subsection 3.2.3 (filters, weighted average, and regression models), we first ran a regression model to understand the effect of each variable on the choice to cycle. A regression model aims at analyzing a set of *independent variables* to predict the value of a *dependent variable*.

The employed regression model was an Ordinary Least Squares Regression (OLSR), implemented by the *statsmodels* Python library. We used a data set of trips with four variables. The dependent variable "is_bike" assumes a binary value: 1 if the record corresponds to a bicycle trip and 0 for other vehicles. The three independent variables consisted of the age, distance, and slope potentials. The OLSR results are shown in Table 5.1.

	coef	std err	t	P-value	95% conf. interval
age potential	0.0162	0.001	21.753	$1.12 * 10^{-104}$	[0.015, 0.018]
distance potential	0.0094	0.001	10.484	$1.05 * 10^{-25}$	[0.008, 0.011]
slope potential	-0.0033	0.001	-3.911	$9.19 * 10^{-5}$	[-0.005, -0.002]

Table 5.1: *Regression results to identify each variable's effect on the choice to cycle*

We highlight two of the observed results. The first result regards the low P-value of all three variables. The P-value denotes the probability of the data behaving in the input distribution by chance, without relation among the variables. In other words, lower P-values represent that the independent variables are strongly related to the dependent variable. Thus, in this case, age, distance, and slope potentials have an intense correlation with the choice to adopt a bicycle instead of other vehicles. This strong correlation is natural because we modeled the potential using the distribution of the current bike trips.

The second result we emphasize is the scale of the coefficients ("coef" column in the table). None of the three coefficients evaluates to a significant value. The values lie around 0.01, while the variables range is the interval $[0, 1]$. This way, although the P-values indicate that all variables strongly influence the choice to cycle, their effect does not distinguish them from each other.

Based on the regression results, we adopted the simple **arithmetic mean** of the three variables as the final cycling potential of a trip, i.e., let t be a trip retrieved from OD17, $a(t)$ be the age potential of t, $d(t)$ be the distance potential, and $s(t)$ be the slope potential. Then, the final cycling potential $P(t)$ of trip t is defined by

$$P(t) := \frac{a(t) + d(t) + s(t)}{3}.$$

This way, we have that $P(t) \in [0, 1]$, where values closer to 1 indicate trips more likely to migrate to cycling.

5.3 Trip Examples

This section will present three OD17 trips to exemplify the *SP Cycling Potential* model. The first example in Figure 5.7 shows a trip with significantly high cycling potential. This example represents a typical trip at noon connecting close points in the central region of São Paulo. In addition to being short, the route follows a flat path. The uphill segments are not extensive and occur in less steep slope levels. The individual who made this trip is a 42-year-old woman, thus receiving a 0.46 potential since it is not the most common age group among cyclists. The distance variable has a potential of 0.94, and the slope variable value is 0.92, representing very favorable cycling physical conditions. The final cycling potential for this route is (0.46 + 0.94 + 0.92)/3 = 0.77.

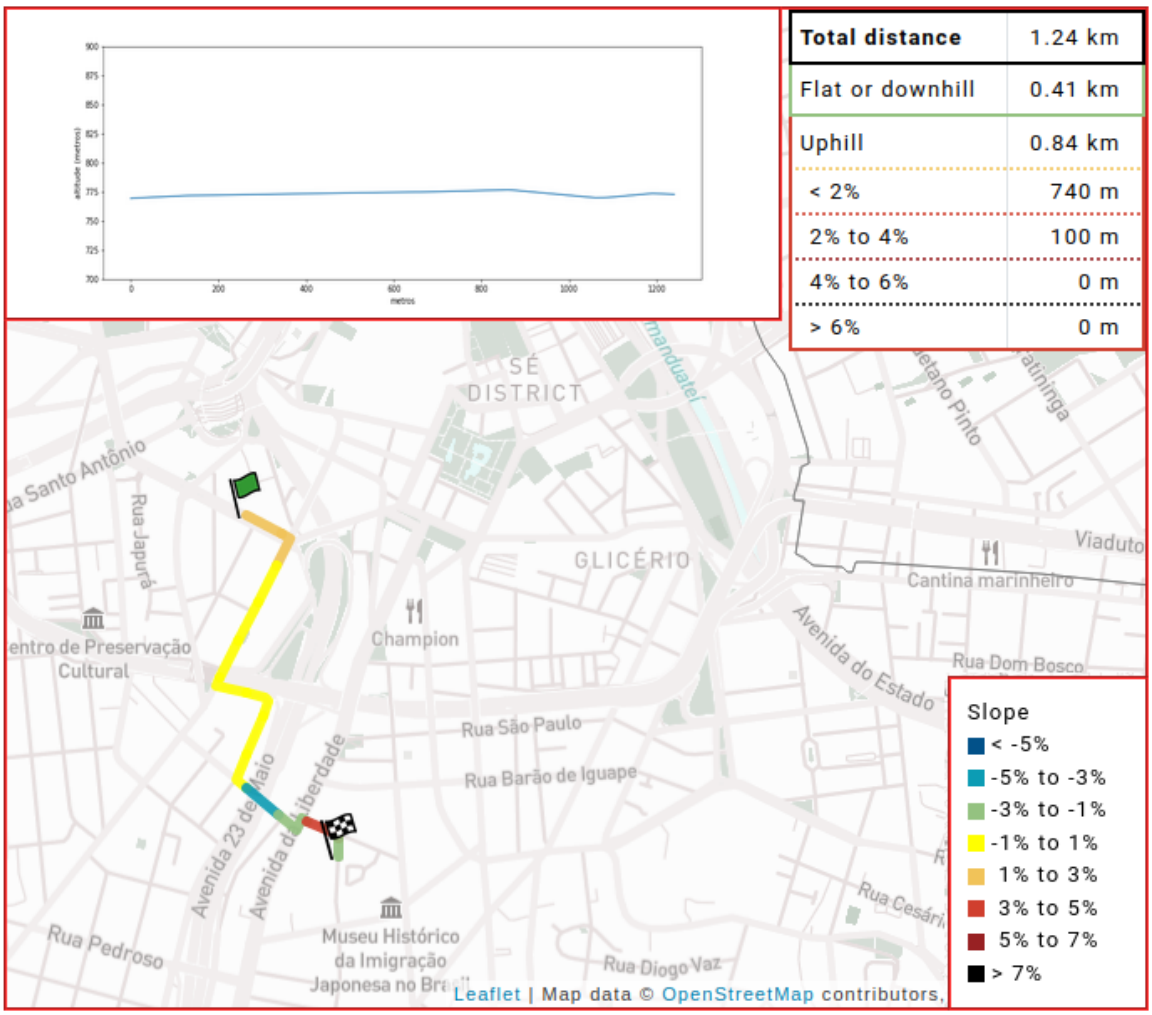


Figure 5.7: *High cycling potential example.*

Figure 5.8 illustrates a route having a moderate cycling potential. This trip connects a region on lower altitudes - next to the Tietê River - to a higher location - *Alto da Lapa*. The man responsible for this route is 54 years old, receiving 0.41 for the age variable. The value for the distance is 0.56, corresponding to 3.5 km, and the slope potential evaluates to 0.38 since it travels over some steep uphill streets. The final cycling potential for this trip is 0.45.

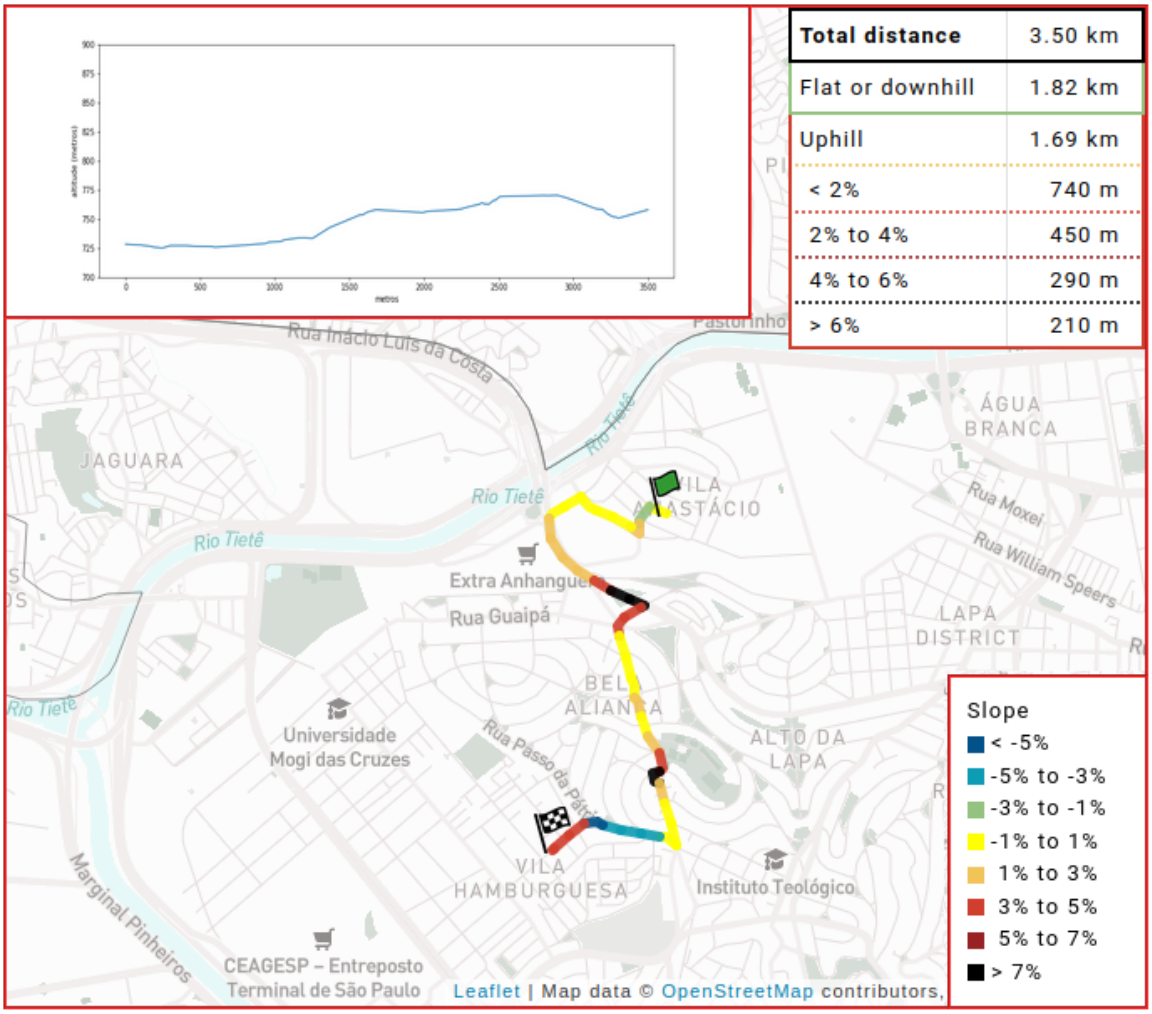


Figure 5.8: *Moderate cycling potential example.*

Figure 5.9 presents a trip not as bicycle-friendly as the previous ones as the last example, with a cycling potential of 0.35. This trip occurs in a hilly area in the north region of the city (*Serra da Cantareira*). As we can see in the topography profile shown in the Figure's upper right corner, there are many up and downhill segments, including slopes steeper than 10%. Although the age variable has a higher potential (0.80 corresponding to a 20 y/o woman), the assessment for the other two variables is lesser: 0.18 for the distance and 0.08 for the slope. Table 5.2 summarizes the three examples' information and potential.

Examining examples helps to understand the general ideas of the model. However, they provide particular perceptions that may not hold for all data. In the next chapter, we apply validation methods to estimate the model's consistency.

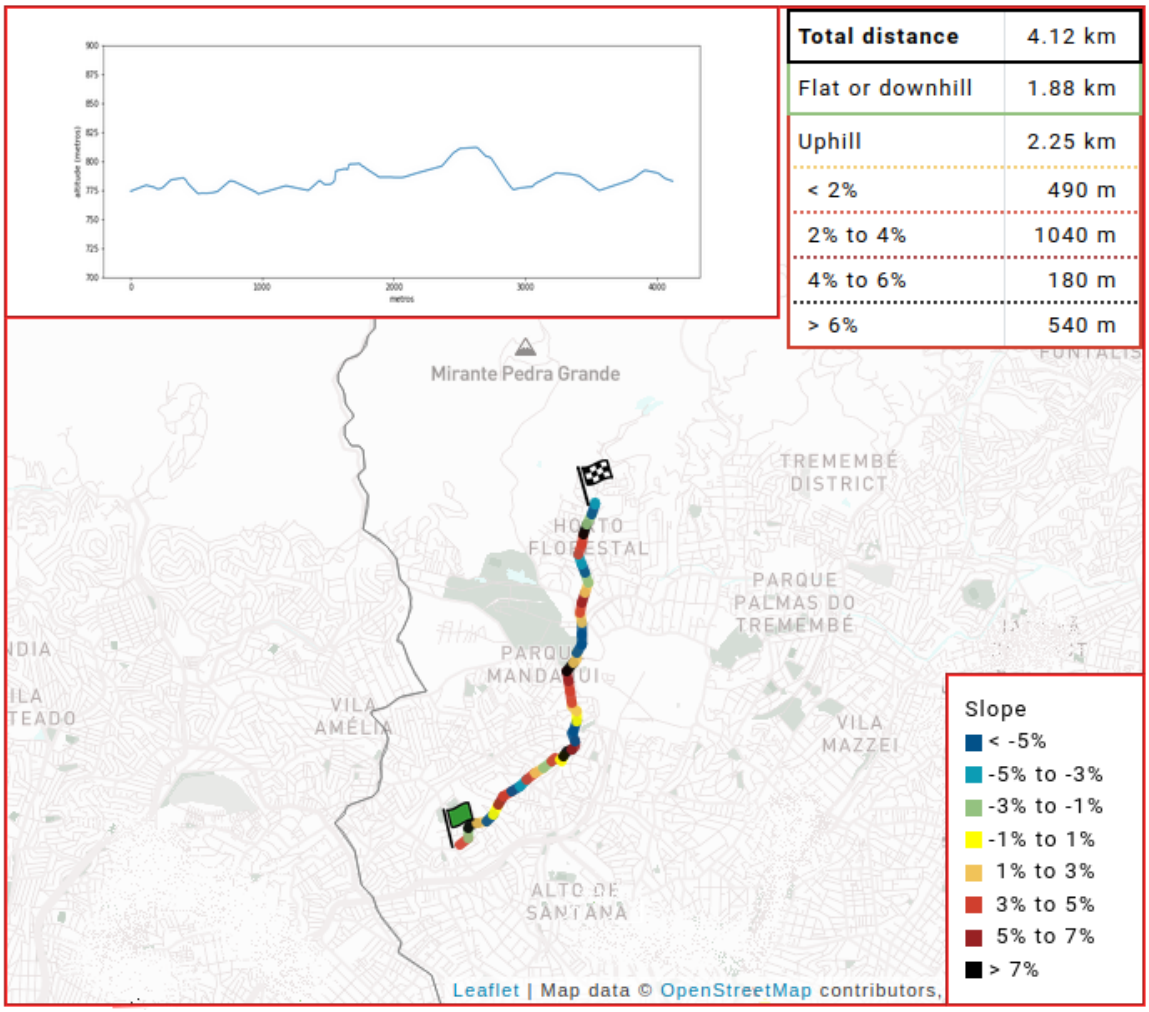


Figure 5.9: *Low cycling potential example.*

Example	Gender	Age	Distance (km)	Cycling Potential			
				age	distance	slope	final
5.7	F	42	1.24	0.46	0.94	0.92	0.77
5.8	M	54	3.50	0.41	0.56	0.38	0.45
5.9	F	20	4.12	0.80	0.18	0.08	0.35

Table 5.2: *Examples of the SP Cycling Potential*

Chapter 6

Model Validation

Another vital step in the conception of data science models is the validation process. One common technique to perceive consistency is looking at visual examples, as we did in 5.3. With the examples, one can easily verify the presence of sharp irregularities during the data cleaning and treatment operations. Likewise, a common approach to validate results is to detach a representative subset of the data universe and test the model to analyze this set's qualitative and quantitative measures.

In the absence of pre-established cycling potential measures for the trips of São Paulo, we conducted a survey targeting different profiles of cyclists from the city, asking their opinion about the *cyclability* trips instances. This way, we can compare the *SP Cycling Potential* values to the users' answers, measuring the model's adherence. In the following sections, we first perform general validations on intuitive properties that the model needs to satisfy, and then describe the implementation and results of the user survey.

6.1 Primary Verifications

One relevant but straightforward analysis consists of calculating the cycling potential of the bicycle trips. Naturally, bike trips tend to have the most favorable cycling conditions, so the model should assign them a high cycling potential. Figure 6.1 compares the cycling potential of the car, walking, and cycling trips. We observe that most bicycle trips (54%) evaluated to values higher than 0.65. Regarding the walking modal, the share of trips in this range of values was 35%, and only 15% for the car modal. Indeed, bicycle trips receive more significant cycling potential than other modals. However, because we used the set of OD17 bike trips to design the functions determining the *SP Cycling Potential*, these higher values were expected. Future research can improve this validation using bicycle trips from other sources, such as cycling apps and bike-sharing systems.

Another technical validation we conducted relates to slope processing. While analyzing route examples, we observed routes that did not correspond to their regions' actual hilliness i.e., flat locations where the model assigned low values for the slope variable. This way, we mapped the slope of route samples in different levels of detail (Figure 6.2a), and compared these slope profiles with the São Paulo topography data (Figure 6.2b) retrieved from the

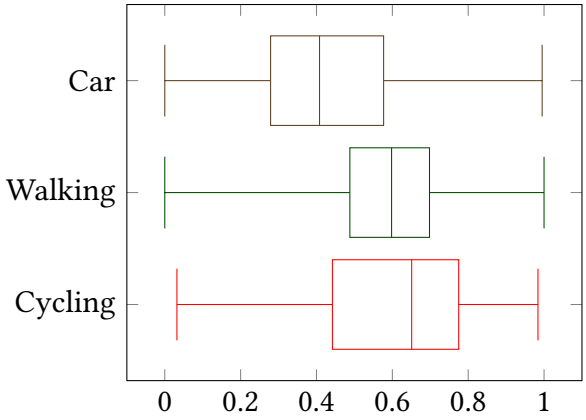
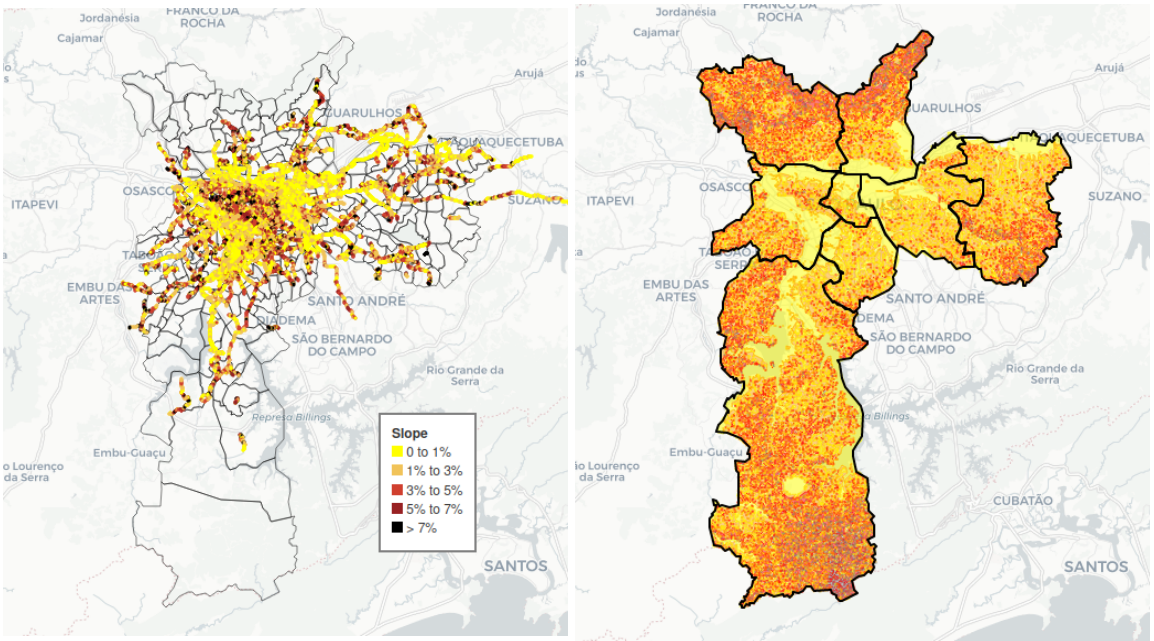


Figure 6.1: *Cycling potential comparison among car, walking, and cycling modals*

municipality geographical service *GeoSampa* (geosampa.prefeitura.sp.gov.br). Figure 6.3 shows one route with inconsistent slope (*São Gabriel* Avenue), i.e., the model indicated uphill segments (in black), while the route region does not present height variations.



(a) *routes slope sample*

(b) *GeoSampa topography data*

Figure 6.2: *Comparison between the modeled slope with the GeoSampa topography data.*

This issue's source is the altitude of the points the GraphHopper API returns. The API retrieves the elevation information from the SRTM Data (srtm.csi.cgiar.org/srtmdata/) (GraphHopper, 2015), a space shuttle mission to shape the Earth's surface using radars. Because the SRTM Data implements a discretized model, discontinuity points originate these inconsistencies. Fortunately, they are rare and the routes affected are exceptions since the SRTM resolution is about 30 meters, sufficient to identify topography hindrances for cycling. Figure 6.2 indicates that in general the city topography coincides with the

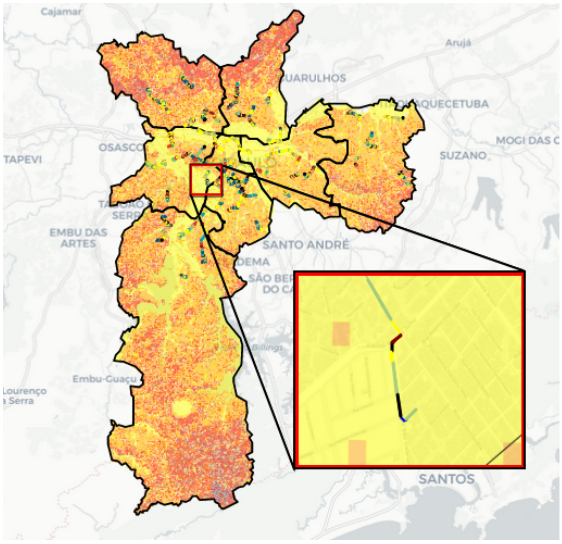


Figure 6.3: *Example of a route with incorrect slope data (indicating uphill segments in a flat area).*

slope of the routes, i.e., the flat segments occur in flat areas, and the up and downhill segments occur normally in hilly regions.

To help mitigate this problem for future works, we accessed the OpenStreetMap (OSM) content (www.openstreetmap.org), which is the tool that supplies data about roads and public transport for the GraphHopper API and many other geographic systems. In the OSM, we added tags detailing the correct altitude of some points in the inconsistent areas, based on topographical charts from GeoSampa. With these tags, other cartography services can produce more precise topography reports (OpenStreetMap, 2021).

6.2 The User Survey

The objective of the user survey was to collect data about people's notion of distance, slope, and general conditions of routes that would limit their cycling activity. This information helps to identify whether the model is too optimistic, pessimistic, or proposes realist indications of cycling opportunities in the city.

We divided the survey into two parts: personal data and routes evaluation. The respondents answered the first group of questions about their **age, gender, cycling frequency** (daily, monthly, etc.), **cycling experience time,** and **bicycle trip reason** (education, work, leisure, work tool, or others). Also, participants select one region of the city they want to evaluate the routes. The idea is that the individual will analyze familiar locations, helping the visualization, for example, of the slope degree in that path.

In the second part, users visualize a map with **five routes** from the selected region. After clicking over each trip course, the map focuses on that trip and shows two icons indicating the beginning (green flag) and the end (checkered flag) points of the trip. Also, details about distance and topography appear, as shown in Figure 6.4. It is also possible to change the map style to a satellite view. Appendix B shows all questions from the user survey (in Portuguese).

There are four questions for each of the five routes:

1. How do you assess this route distance?
2. How do you assess this route slope?
3. If you needed to travel through this route and it had a bike lane or path, what is the chance of making it with a bike?
4. Comments about the route (optional).

We used a five-point *Likert* scale to gather the answers to the first three questions, where the first item (1) denoted the less proper condition for cycling (bad distance, bad slope, or low chance to travel by bicycle), and 5 represents the most bike-friendly answer.

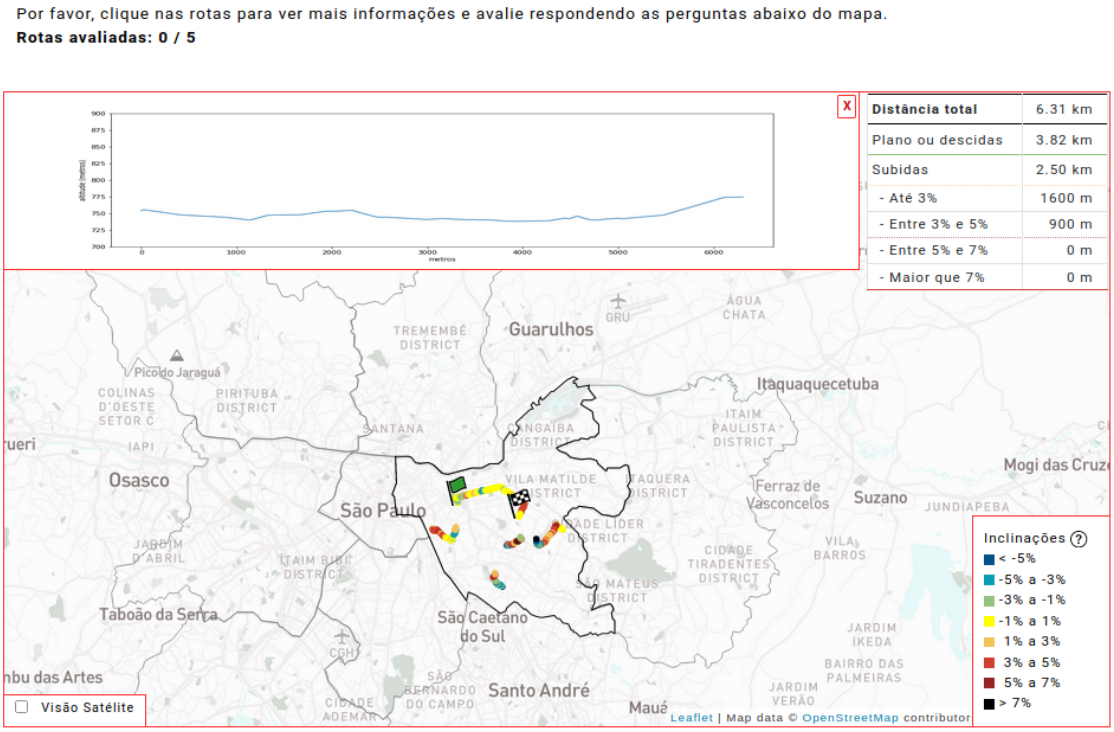


Figure 6.4: *Map from user survey (text in Portuguese)*

We applied a partially randomized process to select the routes for the user survey. We grouped the universe data set (170 thousand route records) into five classes, combining different characteristics from distance and slope: (1) good distance and slope; (2) good distance, bad slope; (3) moderate distance and slope; (4) bad distance, good slope; (5) bad distance and slope. We consider a variable's value *good* if the potential is in the interval [0.65, 1], *moderate* if in [0.35, 0.65], and *bad* for the values in [0, 0.35]. Then, we processed the districts' geographical boundaries to separate the trips among the eight city regions (North 1, North 2, West, Central, East 1, East 2, South 1, and South 2)[1].

Next, we randomly selected one route from each of the classes for each region. This way, people would analyze routes of all the different profiles independent of the region

[1] The municipality frequently uses this division for administrative councils (São Paulo, 2015).

they selected. To improve the visualization in the map, we manually excluded some trips from the sampling process: routes shorter than 300 m, longer than 10 km, routes that intersect others, and paths already fully covered by cycling lanes, preventing the existence of infrastructure from influencing the evaluation of the physical aspects.

6.3 Survey Results

The survey was available for one month, from November to December 2021. We disclosed the study among cycling non-governmental organization (NGO's) mailing lists, cycling groups and pages in social networks, and individuals related to urban mobility research. In total, 89 respondents evaluated five routes, resulting in 445 evaluations.

6.3.1 Respondents profile

We observed a male predominance of 78% regarding the respondents profile. Although undesirable, this predominance is less intense than the general gender balance among São Paulo cyclists - where only 10% are women. As for the age factor, Figure 6.5 shows the respondents' distribution, which is similar to the São Paulo cyclists age distribution with a higher dispersion for women.

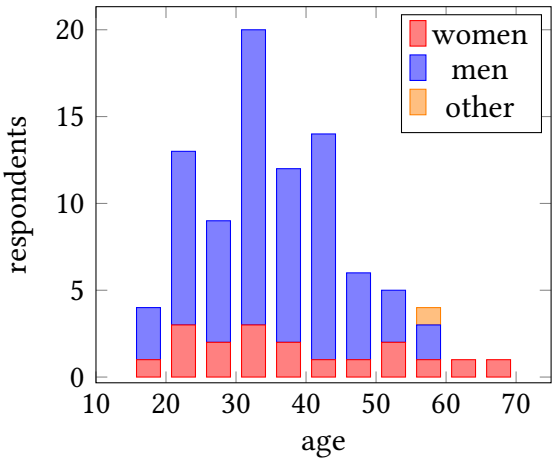


Figure 6.5: *User survey - respondents age*

The respondents' experience also presented uneven results, as depicted in Figures 6.6 and 6.7. Generally, our survey has reached cyclists who use bicycles daily or at least once a week (85%) and for a long time. Therefore, this audience presents a relevant bias for the evaluation of the routes.

The trip reason question accepted multiple answers. This way, 85% of the surveyed reported using the bicycle for leisure, 66% to go to work, 36% for educational purposes, and 7% use the bicycle as a work tool, such as for food delivery.

Lastly, the concentrated distribution of the users' regions may have produced less reliable results in locations with fewer evaluations. Figure 6.8 shows that the West and Central regions have more numerous responses (33 and 16 respectively), while the North 1, North 2, and East 2 achieved a more limited number of users.

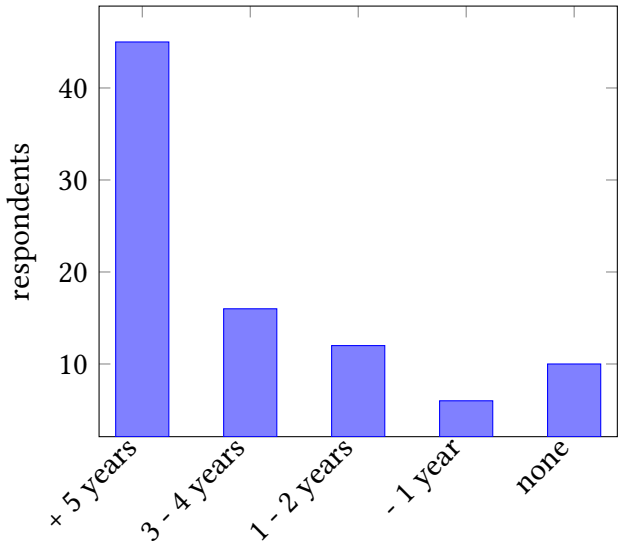


Figure 6.6: *User survey - time of experience as a cyclist*

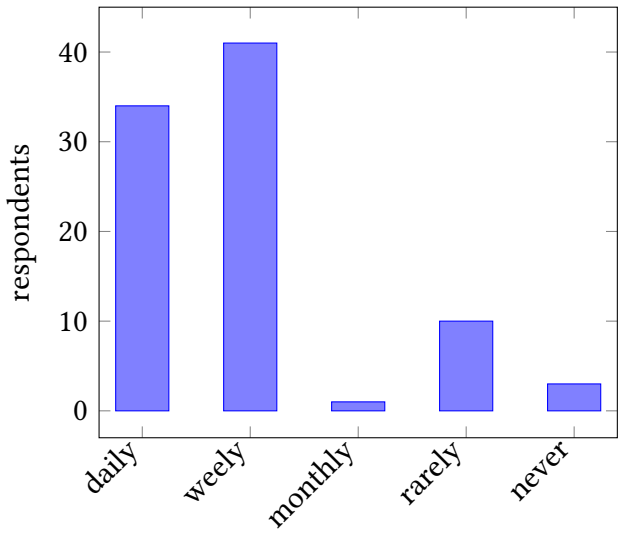


Figure 6.7: *User survey - cycling frequency*

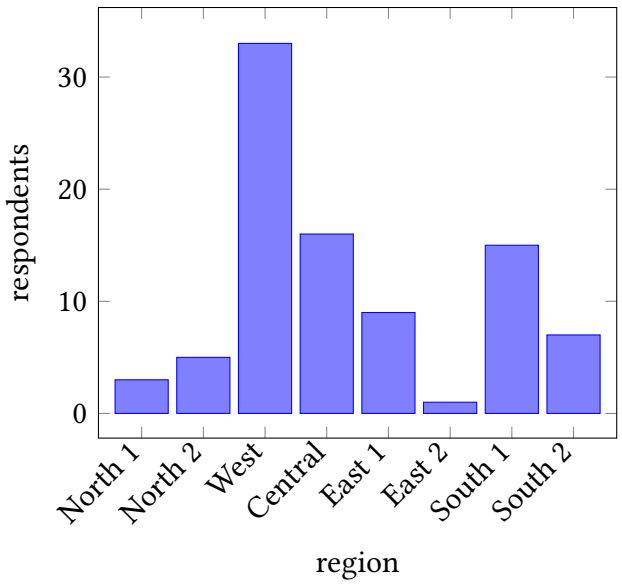


Figure 6.8: *Regions selected*

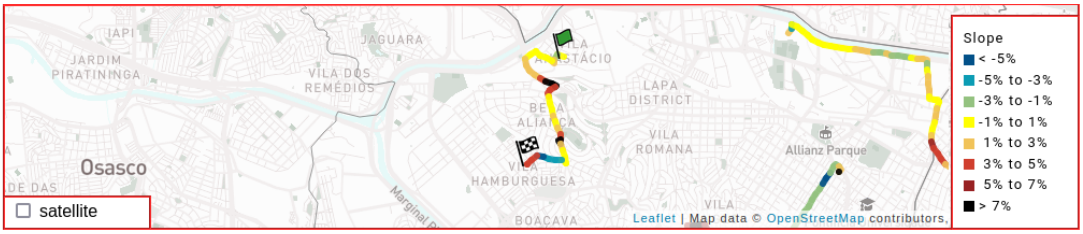
6.3.2 Routes evaluation

Now, we will apply a few techniques to compare the user interpretations of the routes to the *SP Cycling Potential* model, investigating the notable differences and similarities. We will conduct this analysis by presenting one example to illustrate the general users' answers. The reflections we highlight in the example will reflect more general scenarios. In this sense, Figure 6.9 reports the 33 evaluations that the *moderate* route from Figure 5.8 received.

The first point we emphasize in the example is the intense concentration of the answers on the higher values. 27 of the 33 respondents considered that the 3.5 km distance is proper for cycling, assessing with a grade of 4 or 5, while none responded with 1. This concentration also happens for the other two questions. Our model, in turn, calculated a 0.5 value for the distance variable.

This distribution of the distance answers is typical among all sampled trips. Figure 6.10 compares the mean of the answers about the distance with the modeled cycling potential value for each route - in the figure, each point represents a route. If the model were entirely adherent to the users' answers, we would observe the points forming an inclined line (positive correlation), implying that low cycling potential routes would correspond to low values of the user answers. However, we observe that the points approximate a horizontal line, indicating that this variable did not influence the user responses. Indeed, the users grade all routes with 3 or more, on average.

We discuss two possible explanations for this appearing inconsistency. The first, and likely the most relevant, refers to the users' profile bias. As Benedini *et al.* (2020) reported, experienced cyclists travel longer distances than those who recently started cycling. Similarly, people who cycle every day are used to traveling more than those migrating to the bicycle modal. The second cause of this difference may relate to a change in cyclists'



ID: 46177

	Distance	Slope
Potential	0.561	0.381

Route's evaluations

	1 (bad)	2	3	4	5 (good)
Distance	0	1	5	12	15
Slope	0	4	8	15	6
Would you cycle?	1	2	7	10	13

Figure 6.9: *Example of the user survey results.*

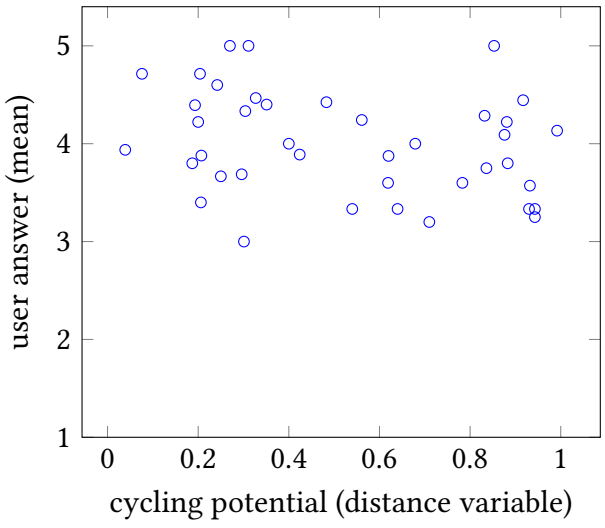


Figure 6.10: *Comparison among the user survey answers and the cycling potential for the distance variable*

profiles from 2017 - when the OD survey collected the data we used - to 2021. This way, the deployment of more cycling infrastructure and new dock and dockless bike-sharing systems implemented in 2018 may have stimulated cyclists to use the bicycle even on longer trips. Hence, the OD17 data seem to have conducted the *SP Cycling Potential* to have a conservative approach regarding the distance variable, i.e., it may indicate trips which some groups could accept to migrate to bicycle as *too long* .

Returning to the example, we observe another interesting fact concerning the better adherence of the slope variable. The respondent answers assigned lower scores for the route slopes compared to the distance. Only six out of 33 respondents indicated that that trip had an entirely suitable topography. Although there are no 1-valued answers, we note that the distribution for this question concentrates in less bike-friendly values. In general, the relation between the users' opinion and the developed model is more intense for the slope variable, indicated by Figure 6.11. In the figure, we can note that routes with lower slope potential also received less favorable answers from the respondents, with some exceptions. We emphasize the trips from the west region in red since more respondents evaluated these routes, producing more consistent indications of this slope relation.

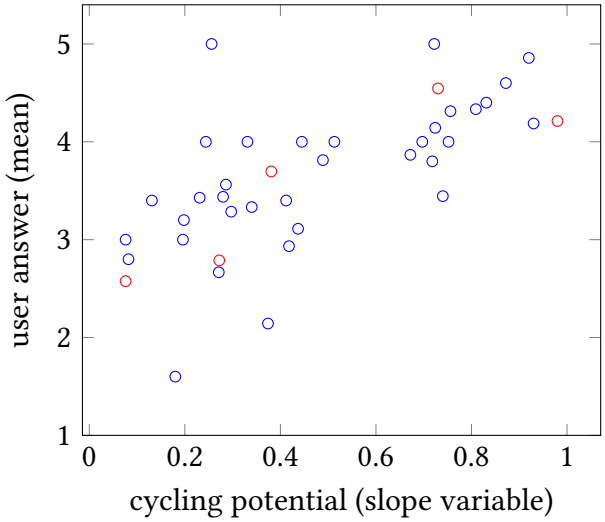


Figure 6.11: *Comparison among the user survey answers and the cycling potential for the slope variable - highlighting the trips from the west region in red*

Regarding the optional comments about the routes, there were 103 comments among all route evaluations. From these, 55 were explanations for the slope and distance answers, e.g., explaining that a route segment is too steep or that a course is too short. Other 15 comments justified the answers using other factors beyond the slope and distance, mainly the safety of the streets and traffic intensity. This type of answer may indicate other variables that the mobility policies should consider in addition to the physical characteristics. Other 20 comments mentioned alternatives for the routes, suggesting that they would deviate the path to avoid hilly areas or shorten the length. Lastly, 10 participants reported curiosities about the routes without influencing the answers.

Although the distribution of the respondents' was unequal among regions and cycling experience, the significant number of answers (445 routes evaluated) allowed valid and

relevant analyses of the model consistency. With these validations, we can suggest that the *SP Cycling Potential* applies technically consistent approaches that capture some limitations for the cycling activity. Like many data science models, it uses available data to propose and test different methods that can - and must - be improved and updated to generate even more precise models.

Chapter 7

Cycling Potential Analyses

This Chapter describes the software solution implemented within the *BikeScience* project to study the cycling potential of all non-cycling OD17 trips. We examine the geographical dispersion of the cycling potential on the SPMA municipalities and inside the city of São Paulo. Then, we compare the indicated regions with significant cycling potential to the existing cycling infrastructure, showing areas with great opportunities for cycling infrastructure investments where the modal shift to bicycle could be more effective.

The input dataset of this analytical tool is all 41.6 million non-cycling trips of OD17, concerning all SPMA. However, we divided each trip by the different vehicles composing them. For example, if a person travels by bus and then uses a subway line, there will be two records for this trip: one with the route for the bus part and the other for the subway part. We can precisely analyze each modal's cycling potential with this strategy, identifying the correct fragments of trips that could migrate to cycling. Hence, the final dataset contains 47.4 million trips.

The tool filters the routes data set and displays choropleth maps with the routes' geographical distributions. Figure 7.2 shows the interactive settings to generate the maps. It is possible the determine the range of values for the routes cycling potential. The users can also select each variable's lower and upper limits (for age, distance, and slope). The "Modal" option filters the trips according to their original vehicles - the alternatives are walking, car, motorcycle, subway, train, and "all", which considers the totality of trips.

The "Trip Reference Point" option characterizes the way we aggregate the routes among the districts - the choices are *origin, destination*, and *whole trip*. Suppose the analyst selects the origin option, then only the area where a route starts will count this trip (Figure 7.2a). Analogously, in the destination option, only the end district will consider this trip (Figure 7.2b). The whole trip alternative sets that all districts where this route passes through will account for that journey (Figure 7.2c).

The binary option "Show routes" defines whether the map will contain only the districts or also the routes paths. Similarly, if the "Plot SPMA" is checked, the map will calculate and present data about the entire the metropolitan area, otherwise, it will restrict to São Paulo districts. Lastly, the "Density" option indicates whether the map will show the

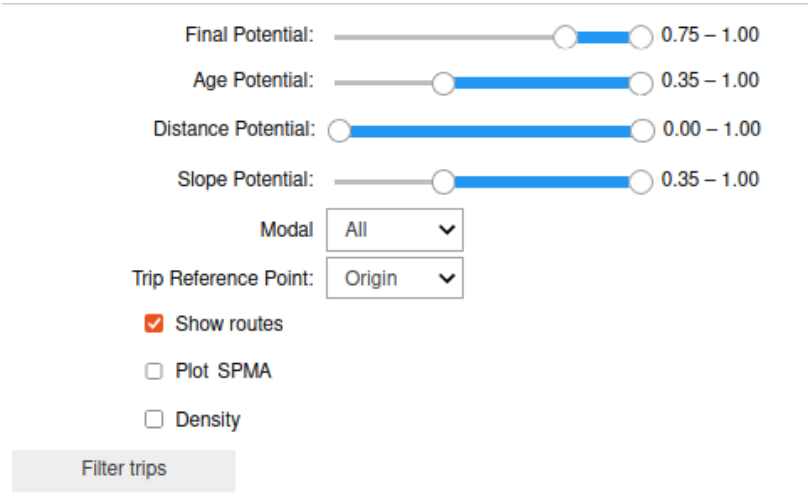
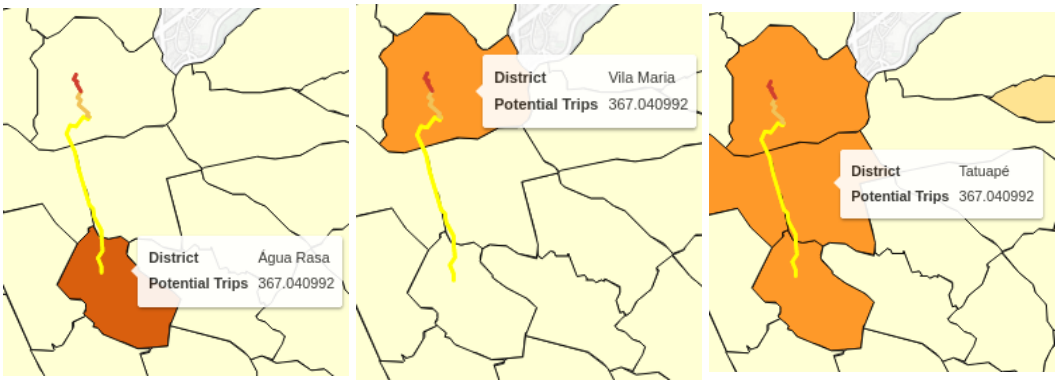


Figure 7.1: *SP Cycling Potential maps settings*



(a) *origin* **(b)** *destination* **(c)** *whole trip*

Figure 7.2: *Trip reference points*

absolute number of trips or the percentage of trips that satisfy the specified filters. For example, if a district has 100 thousand trips, including 10 thousand with a specific range of cycling potential, the default setting (not-density) would present 10 thousand potential trips, and the density option would show the district with a value of 1%, corresponding the proportion of trips in that region satisfy that filter.

7.1 Cycling Potential Distribution

Many cycling potential analyses are possible and relevant in several contexts. Now, we detach some maps with interesting results that may indicate directions for the cycling investments in São Paulo.

First, Figure 7.3 reports the distribution of cycling potential among all trips. Most trips (51%) received an intermediate cycling potential - between 0.3 and 0.6, generally representing trips with some obstacle for cycling in one or more variables. Nevertheless, we highlight a very significant number of trips with high cycling potential: 26% of the trips have a cycling potential greater than or equal to 0.60. It means that, in an optimistic scenario,

one quarter of the SPMA daily trips could migrate to cycling with the implementation of adequate policies.

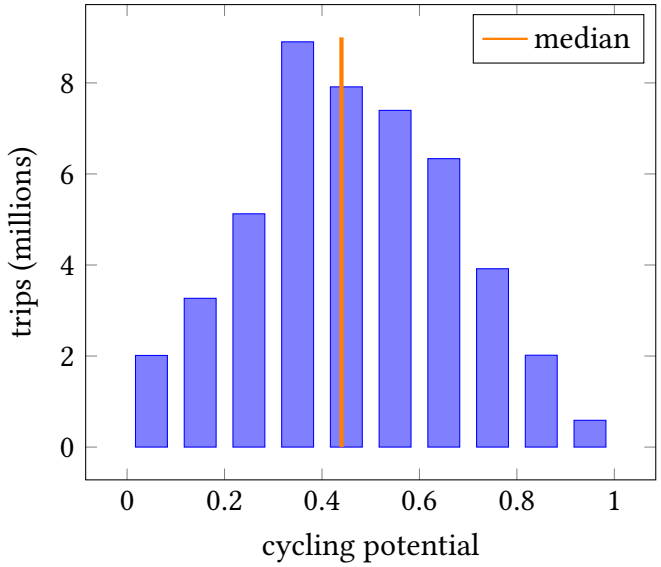


Figure 7.3: *Cycling potential distribution of all OD17 non-cycling trips*

Our analysis will consider a more restricted scenario - trips with cycling potential higher than 0.8. We will refer to the trips in this range of cycling potential as "potential trips" or "cyclable trips". This 0.8 limit implies that the trips received a high value for at least two of the three variables and that all variables evaluated to a reasonable value - discarding any low cycling potential situations. The number of trips within this range is 2.6 million, corresponding to 5.4% of all trips in the SPMA. If all these trips were migrated to cycling, the number of bicycle trips would be seven times greater than the current share.

For the geographical distribution, we used the "whole trip" option for the reference point, i.e., all districts where the trip passes consider it. However, the other options (origin, destination) produced similar results because these routes were short, thus passing through fewer districts.

Figure 7.4 shows the number of cyclable trips on the SPMA, highlighting *Guarulhos* - in the northeast - with a significant number of potential trips (150 thousand, corresponding to 5% of all trips in that city). Figure 7.5 indicates that the cities in the extreme east (*Guararema, Santa Isabel, Salesópolis*) have a greater density (around 10% of potential trips). This geographical panorama is curiously similar to the current bicycle usage in the São Paulo metropolitan area discussed in 4.2.2. These coincident patterns suggest that the current number of bicycle trips may be a cycling potential indicator, i.e., a region with greater bike adoption may naturally indicate proper cycling conditions.

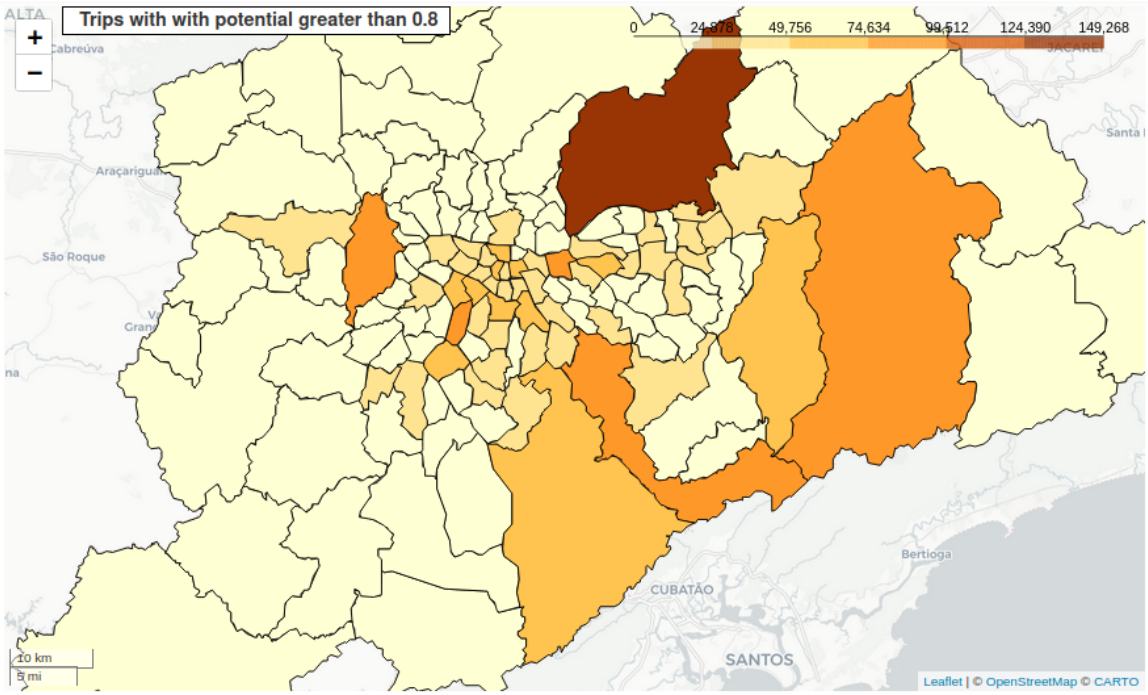


Figure 7.4: *São Paulo potential trips by district.*

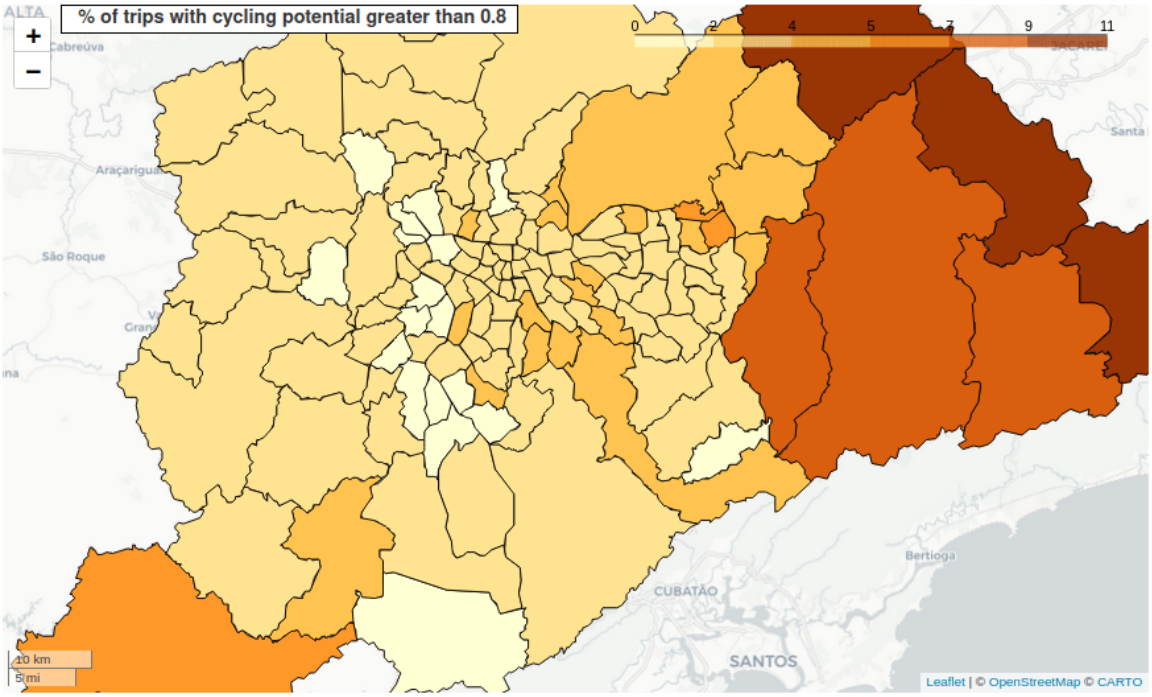


Figure 7.5: *São Paulo potential trips density by district.*

The similarities between the potential trips and the current cycling distribution also occur inside the city of São Paulo. As shown in Figure 7.6, the districts in the southwest region along the Pinheiros River (*Itaim Bibi, Santo Amaro, Pinheiros*) concentrate a substantial number of potential trips - comparable with the bicycle trips in Figure 4.11. An

exception appears in the *Tatuapé* district, in the central-east region. This region is an important economic pole with flat topography and have mass transport infrastructure (subway and train lines). However, it currently has a low share of cyclists and fewer bicycle lanes, despite presenting plenty of potential trips (88 thousand).

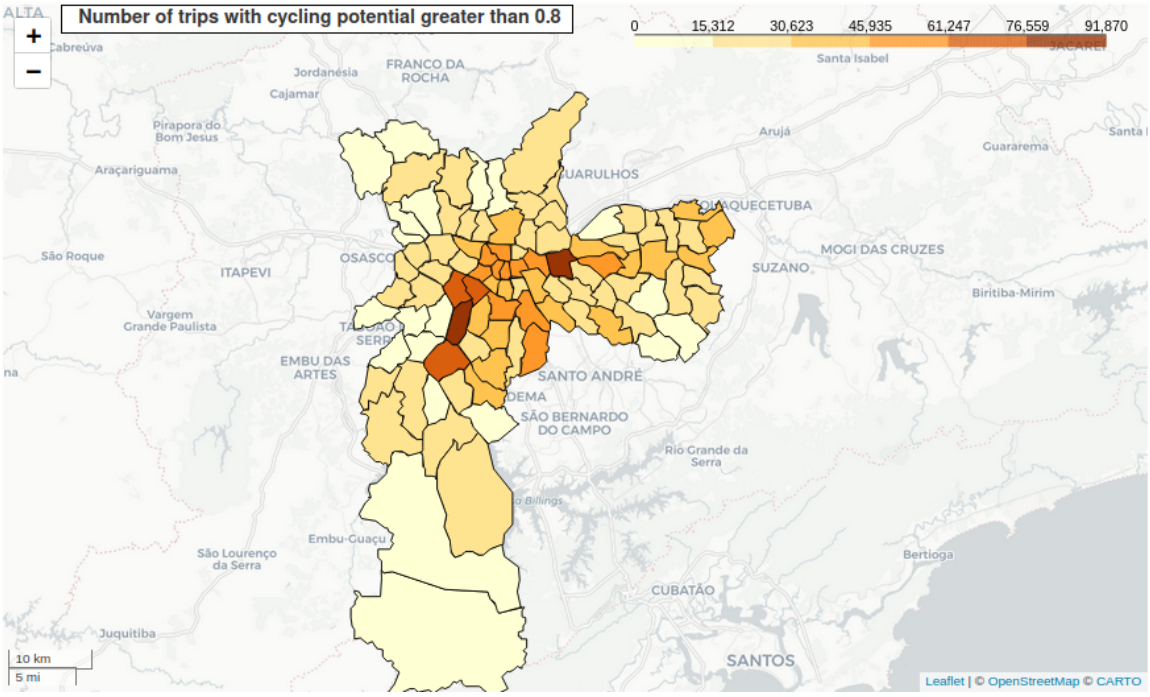


Figure 7.6: *São Paulo potential trips.*

Regarding the cycling potential density among São Paulo districts, Figure 7.7 reports that the *Itaim Paulista* and *Jardim Helena* districts, in the far east region, have the most significant proportion of cycling trips (5.9% and 5.3%, respectively). These peripheral zones have a younger population (SEADE, 2014), are located in the flat terrain next to Tietê River, and have short trips connecting train stations to residential areas. We also observe that the whole east region has a considerable cyclable trips proportion (more intense orange colors on the map).

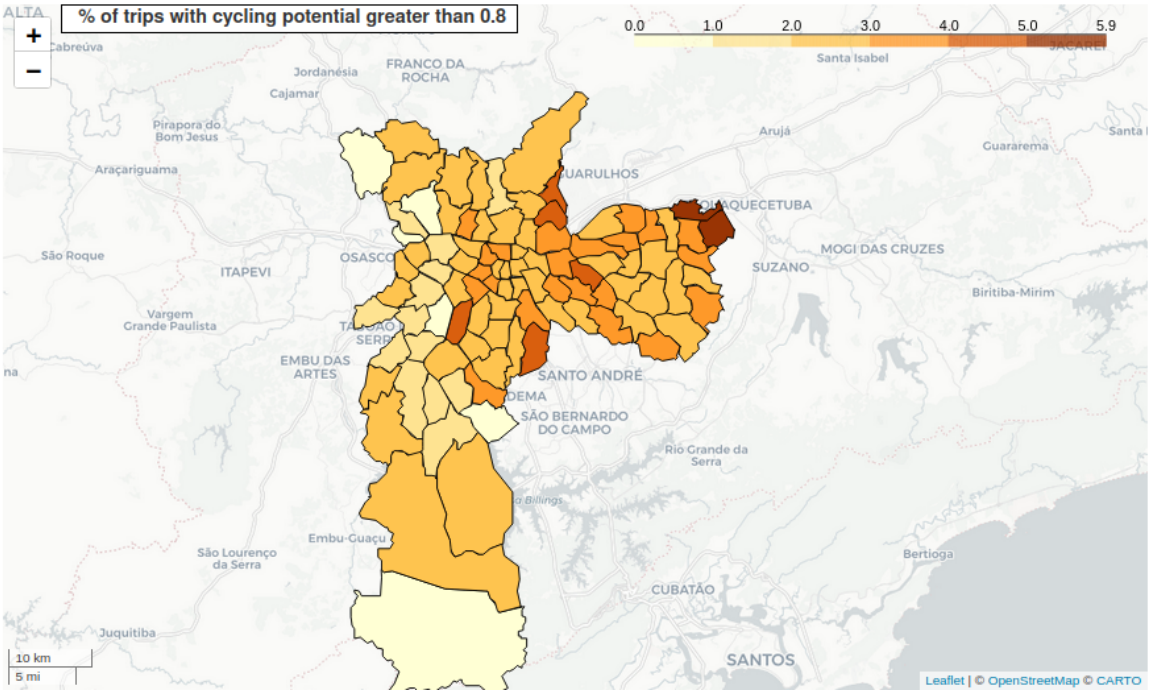


Figure 7.7: *São Paulo potential trips density.*

Next, we group the potential trips by their modal. Figure 7.8 shows the share of each vehicle in the 2.6 million cyclable trips. The walking modal is the most cyclable modal, with 1.38 million trips, followed by the car modal (0.63 million). The walking trips usually have shorter lengths and do not travel in longer uphill segments. On the other hand, public transportation (bus, subway, train) generally crosses longer paths, thus having fewer trips with high cycling potential.

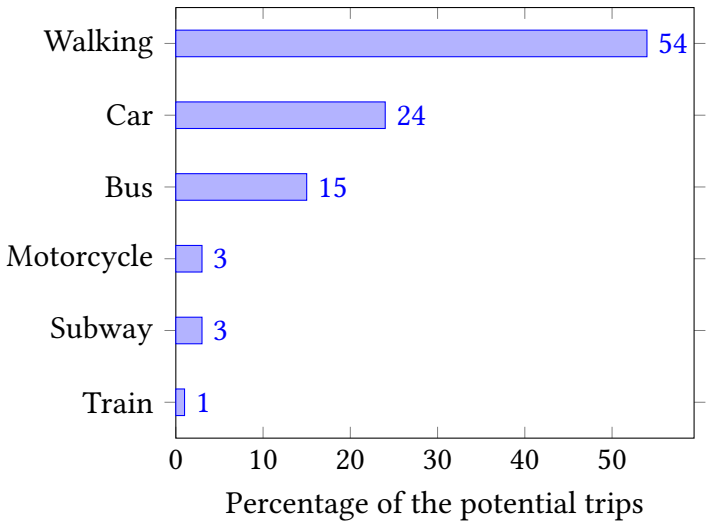


Figure 7.8: *Cycling potential modal share.*

The car-to-bike modal shift is the one that can provide noteworthy positive environmental and traffic impacts. Concerning the geographical distribution, car cyclable trips are

more homogeneous, i.e., located in different regions, more significantly in the expanded central area. The total number of car trips with cycling potential greater than 0.8 is 629 thousand trips, corresponding to 4.8% of all car trips and 900 thousand traveled kilometers per day. If all these trips migrated to bicycle, this would reduce carbon emissions by 108 tons per day (EEA, 2021).

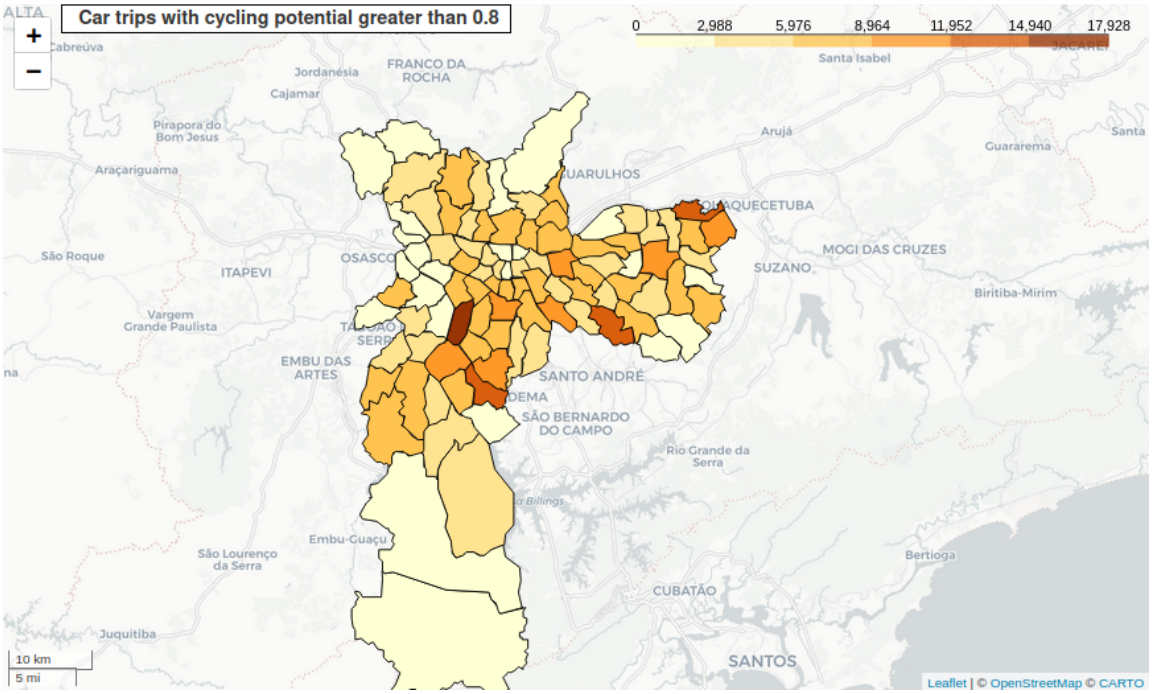


Figure 7.9: *São Paulo car trips with cycling potential greater than 0.8 - considering the "whole trip" as reference.*

Lastly, exemplifying studies with the bike routes, Figure 7.10 shows the number of cyclable trips made originally by train - along with the proposed paths. This analysis can indicate roads for bike lanes and stations that could receive proper parking for bicycles. Using our tool, analysts can also filter the routes by the slope variable, showing appropriate paths for bike lanes' deployments

7.2 High Potential, Low infrastructure

This section summarizes some of the analysis presented in the previous section, qualitatively comparing with the cycling infrastructure geographical data (shown in Section 4.3).

The first region we highlight is São Paulo extreme east - *Jardim Helena* and *Itaim Paulista* districts. This area presents both high cycling potential (5% of trips are cyclable) and cycling share (cyclists make 4% of the current trips). However, there is a notable deficiency of cycling infrastructure in those districts: in *Jardim Helena*, there are local bike lanes that do not connect to the major avenues leading downtown; in *Itaim Paulista*, in its turn, there is no cycling infrastructure.

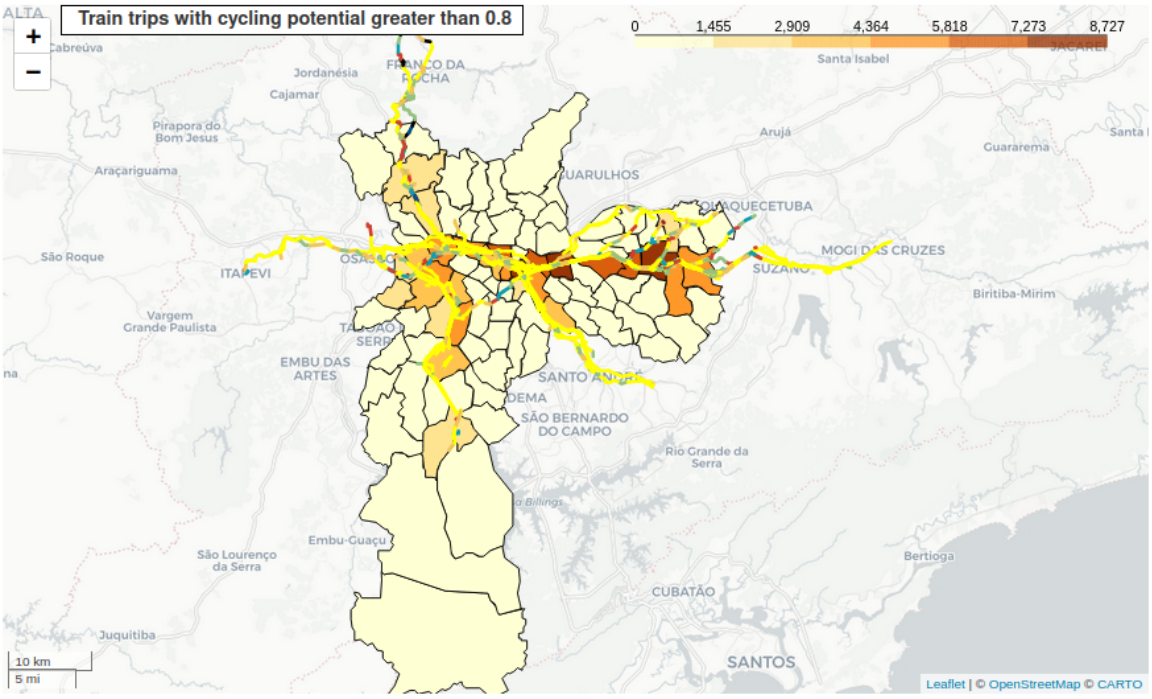


Figure 7.10: *São Paulo train trips with cycling potential greater than 0.8 - showing routes.*

Another relevant point is the *Tatuapé* region. As discussed previously, this region is an important regional pole. It is essential for urban mobility connecting the most populated zones (in the east) with the city's downtown. Nonetheless, this region contemplates only one bicycle lane in its major road, *Radial Leste*, lacking connections to other adjacent districts and to the dense cycling network in the central areas.

Lastly, we emphasize the Pinheiros River east margin districts (*Pinheiros, Itaim Bibi, Santo Amaro*). These areas are examples of prosperous cycling environments. They cover many bicycle trips and have solid cycling infrastructure, including the rare segregated bicycle paths. In addition, we observe that the cycling potential in these districts is also higher, indicating relevant opportunities for the expansion of the cycling share. The investments include not only bike lanes, but also other policies such as reduction of speed limits, marketing campaigns, bike-sharing systems, safety improvements, and others.

The discussion we provided in this chapter are a sample of the applications of the *SP Cycling Potential* tool. Local authorities, policymakers, and mobility researchers may conduct more complex studies by analyzing each variable separately, exporting the filtered data (in a CSV format), and aggregating them in other data engines. With these analyses, public policies on cycling investments may follow a data-driven approach in a worldwide smart cities context.

Chapter 8

Conclusion

This work developed a data science model to identify locations with proper conditions for cycling. The model used evidence from the most recent mobility survey to propose a data-driven approach for prioritizing cycling investments.

The *SP Cycling Potential* index considered a detailed literature review to identify convenient methodologies and relevant data that we should consider (see Chapter 3). Along with the literature concepts, we conducted statistical analysis to understand the current limitations for cycling in our case study. Thus, it was possible to propose evidence-based models to locate regions in the city that could have much more bicycle trips than they currently have. This model can support more effective public policies for urban mobility, indicating areas where investments in cycling infrastructure would promote a significant modal shift to bicycle.

Although the index presented satisfactory consistency, the user survey validation showed relevant directions for its improvement (see Chapter 6). The analyses and implementations developed in the previous chapters present relevant opportunities for additional studies. Next, we discuss topics for future work and the limitations faced during the development of this study.

8.1 Future work

A central topic we highlight regards the data source used in the modeling. There is space for valuable complementary data about urban mobility in São Paulo in a worldwide Big Data context. For example, it is possible to use GPS information from bike-sharing systems (BSS) to track the cycling routes instead of modeling likely paths. The *BikeScience* project conducted previous research in American cities using BSS data (Kon *et al.*, 2021), and the tool also has access to private BSS data from São Paulo. Thus, aggregating this detailed GPS information would be a natural next step. Other sources for modeling daily commutes may include trip counts (manual and automatic) by the municipal traffic engineering company, cycling apps, such as the voluminous dataset from *strava* (https://www.strava.com/), routing apps, and others.

Regarding the OD17 particularities, one relevant attention point is that it estimates one

weekday of the SPMA mobility system, i.e., we overlook data about weekend dynamics. Also, we are analyzing a 2017 scenario that may be significantly different from the current (2021) panorama. In the past four years, the cycling network grew about 50%, and the COVID-19 pandemic abruptly changed all traffic patterns in the city. The bicycle use may have increased as an alternative to agglomeration in mass transports. With updated sources, we could have observed less restrictive limitations for cycling.

As for the validation process, there are possibilities for many other analytical methods. One could relate the question regarding general conditions to cycle with the final cycling potential, using the age of the respondents as a parameter. It is also possible to detail the answers of slope and distance variables to identify the existing correlations. In addition, with more extensive data sets, i.e., more respondents, we could segregate this analysis by region or cycling experience, thus providing a better model fitness for each group.

Future work could also extend the results' analyses by applying other tools from the *BikeScience* project. For instance, it is possible to group the trips by origin and destination zones or grid-cells, abstracting them into *flows*. Also, we could implement functionalities of the *SP Cycling Potential* in the website *BikeScienceWeb* (http://bikescienceweb.interscity.org/), which promotes the tool for a broader public, that does not need to install any software or interact with programming languages. Lastly, researchers can replicate this open-source model in other cities with similar OD surveys - or use different sources to retrieve trips data.

With this work, in a collaboration context between the *BikeScience* project and the municipal traffic engineering company (CET), we hope to contribute to the city's future cycling policies. This way, public investments can better understand the particularities of the bicycle modal and promote a quality leap in São Paulo's cycling environment.

Appendix A

Cycling Infrastructure in São Paulo Districts

The following table complements the discussion of Subsection 4.3, presenting the extension of cycling infrastructure in all São Paulo districts.

District	Area (km^2)	Population	Roads (km)	Protected bike lanes (km)	Unprotected bike lanes (km)	Sharrow bike lanes (km)
Água Rasa	7.15	84963	151.47	0.0	7.57	0.47
Alto de Pinheiros	7.51	43117	143.55	9.22	8.68	0.0
Anhanguera	33.4	65859	195.12	0.0	0.0	0.0
Aricanduva	6.86	89622	147.71	1.48	9.81	0.0
Artur Alvim	6.53	105269	140.96	0.0	16.2	0.0
Barra Funda	5.9	14383	89.95	1.89	12.01	0.87
Bela Vista	2.77	69460	55.48	0.0	2.21	0.0
Belém	6.14	45057	92.82	0.25	1.58	1.65
Bom Retiro	4.27	33892	74.02	2.88	6.72	0.0
Brás	3.65	29265	60.62	0.57	1.99	0.0
Brasilândia	21.06	264918	279.51	2.22	4.29	0.0
Butantã	12.92	54196	188.67	4.11	3.55	0.0
Cachoeirinha	13.59	157408	163.73	0.0	3.7	0.0
Cambuci	3.94	36948	64.6	0.0	7.77	0.0
Campo Belo	8.86	62530	160.96	0.0	7.1	0.0
Campo Grande	13.02	100713	199.05	5.31	13.72	0.0
Campo Limpo	12.61	216098	257.8	0.0	7.3	0.0

Cangaíba	13.79	136623	217.2	6.26	12.71	0.0
Capão Redondo	13.82	275230	296.56	0.0	2.68	0.0
Carrão	7.8	83281	156.12	0.0	10.49	0.0
Casa Verde	7.17	85624	133.72	2.07	7.08	0.0
Cidade Ademar	12.26	266681	270.92	0.0	0.82	0.0
Cidade Dutra	27.79	203473	385.74	4.18	7.67	0.0
Cidade Líder	10.58	126597	192.99	0.0	3.97	0.0
Cidade Tiradentes	14.95	211501	207.2	0.0	1.87	0.0
Consolação	3.8	57365	68.14	0.82	6.64	0.0
Cursino	12.08	109088	172.25	0.0	1.75	0.0
Ermelino Matarazzo	9.43	113615	152.81	6.4	13.35	0.0
Freguesia do Ó	11.06	142327	226.76	6.19	2.88	0.0
Grajaú	92.79	500787	657.16	0.0	0.0	0.0
Guaianases	8.91	103996	141.43	0.0	0.0	0.0
Iguatemi	19.53	127662	238.66	2.66	1.91	0.0
Ipiranga	11.05	106865	177.09	0.85	18.2	0.0
Itaim Bibi	10.05	92570	212.37	11.27	3.67	2.51
Itaim Paulista	12.25	278026	247.58	0.0	0.0	0.0
Itaquera	14.73	204871	304.15	1.34	18.17	0.0
Jabaquara	14.08	223780	287.33	0.0	11.05	0.0
Jaçanã	7.51	94609	143.8	0.0	1.4	0.0
Jaguara	4.54	24895	79.21	0.07	1.89	0.0
Jaguaré	6.57	49863	98.08	1.95	11.82	0.0
Jaraguá	28.2	281824	287.84	0.0	3.44	0.0
Jardim Ângela	36.82	295434	376.37	0.0	0.0	0.0
Jardim Helena	9.15	135043	155.79	0.61	8.1	0.0
Jardim Paulista	6.27	88692	108.04	2.19	6.97	7.35
Jardim São Luís	25.97	267871	326.29	1.49	6.93	0.0
José Bonifácio	14.45	124122	150.51	0.0	0.02	0.0
Lajeado	8.92	185184	180.34	0.0	0.0	0.0
Lapa	10.3	65739	169.37	0.0	4.33	3.9

Liberdade	3.64	69092	63.77	0.0	6.96	0.92
Limão	6.42	80229	132.95	0.0	6.2	0.0
Mandaqui	13.23	107580	147.16	0.1	7.52	0.0
Marsilac	207.96	8258	278.77	0.0	0.0	0.0
Moema	9.13	83368	172.01	1.99	10.28	3.28
Mooca	7.95	63133	129.14	0.0	6.07	1.82
Morumbi	11.49	46957	185.22	3.37	7.88	0.0
Parelheiros	151.8	202321	622.98	0.31	3.89	0.0
Pari	2.73	17299	47.83	0.37	5.05	0.0
Parque do Carmo	15.61	68258	151.93	0.0	9.93	0.0
Pedreira	18.51	158656	210.26	0.0	1.03	0.0
Penha	11.46	127820	214.39	0.0	11.52	0.0
Perdizes	6.33	111161	122.21	0.08	7.91	2.48
Perus	23.69	80187	175.24	0.0	0.0	0.0
Pinheiros	8.29	65364	160.78	7.71	10.46	1.55
Pirituba	17.09	167931	283.53	2.51	2.28	0.0
Ponte Rasa	6.57	93894	146.82	0.0	14.32	0.0
Raposo Tavares	12.19	100164	157.14	1.48	6.38	0.0
República	2.33	56981	62.31	0.32	7.37	0.0
Rio Pequeno	9.78	118459	211.26	0.25	15.79	0.0
Sacomã	14.64	247851	276.9	0.0	7.85	0.0
Santa Cecília	3.71	83717	68.41	2.83	14.01	0.0
Santana	13.13	118797	209.52	5.24	14.08	0.0
Santo Amaro	15.97	71560	264.56	6.74	16.43	6.36
São Domingos	9.94	84843	163.95	0.0	1.93	0.0
São Lucas	9.69	142347	193.94	4.66	1.54	0.0
São Mateus	12.79	158533	254.85	2.5	14.67	0.0
São Miguel	8.68	92081	156.95	1.73	5.75	0.0
São Rafael	13.08	151017	225.81	0.0	3.98	0.0
Sapopemba	13.65	296042	316.42	10.47	5.9	0.0
Saúde	9.26	130780	183.06	1.1	14.35	0.0
Sé	2.18	23651	58.4	0.36	5.64	0.0
Socorro	12.29	37783	149.85	4.81	6.26	0.0

Tatuapé	8.51	91672	153.28	0.26	6.2	0.0
Tremembé	57.83	197258	382.88	0.0	0.37	0.0
Tucuruvi	9.47	98438	173.61	0.0	0.0	0.0
Vila Andrade	10.33	127015	165.68	3.66	9.46	0.0
Vila Curuçá	9.52	162486	200.8	0.0	2.26	0.0
Vila Formosa	7.52	94799	156.94	0.0	4.75	0.0
Vila Guilherme	7.25	54331	116.91	2.16	8.91	0.0
Vila Jacuí	8.23	167965	188.06	1.3	11.49	0.0
Vila Leopoldina	7.0	39485	102.16	2.52	6.68	0.1
Vila Maria	11.82	113463	230.12	0.47	12.13	0.0
Vila Mariana	8.54	130484	167.83	3.17	13.38	7.85
Vila Matilde	8.92	104947	193.26	0.0	14.54	0.0
Vila Medeiros	7.89	129919	176.5	0.0	8.8	0.0
Vila Prudente	9.6	104242	181.37	1.56	7.88	0.0
Vila Sônia	10.09	108441	184.23	4.49	2.01	0.0

Appendix B

User Survey Screens

This appendix shows the screens developed for the user survey. The text of the questions are in Portuguese. Figure B.1 shows the personal questions (age, gender, frequency, experience, trip reasons, and email). The questions regarding the routes' region is shown in Figure B.2.

Figure B.3 shows the questions that appear when the user selects each route. The questions are about the respondent perception for distance, slope, as well as general potential to cycle and optional comments of the route. Lastly, the satellite view available in the survey maps is show in Figure B.4.

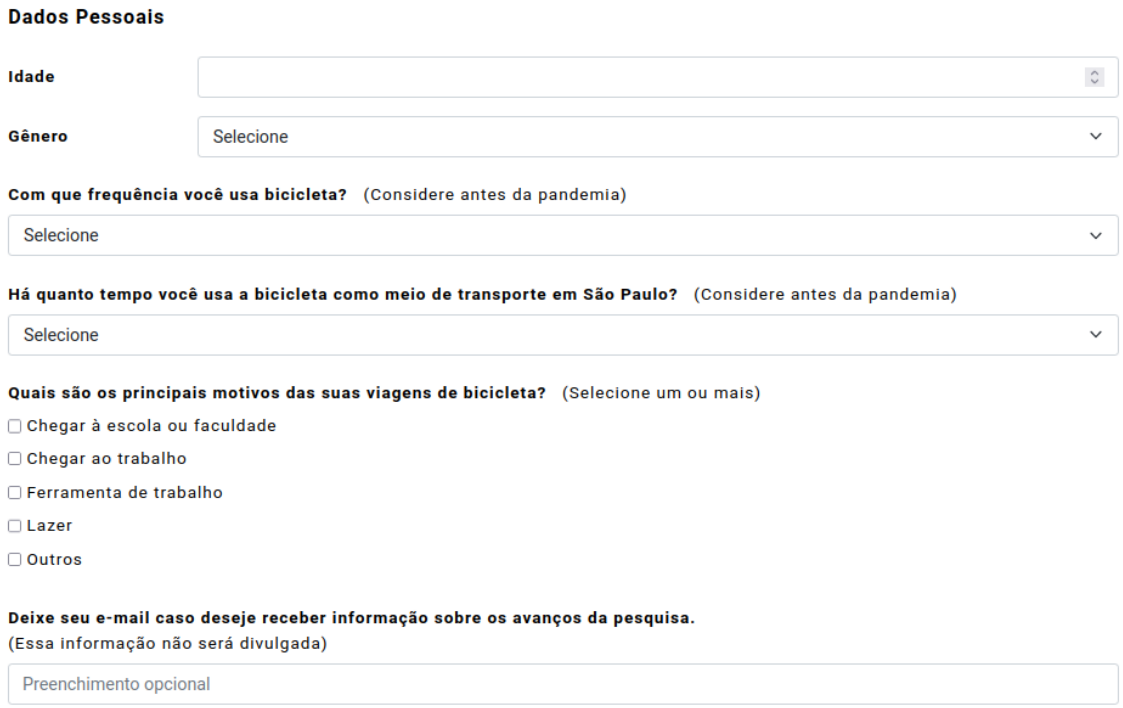


Figure B.1: *User survey personal questions (in Portuguese)*

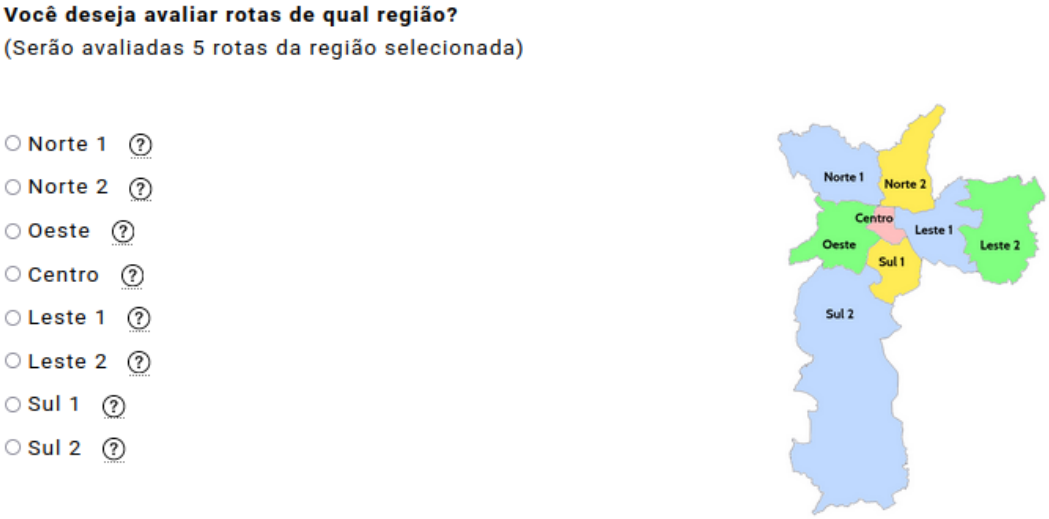


Figure B.2: *User survey region options*

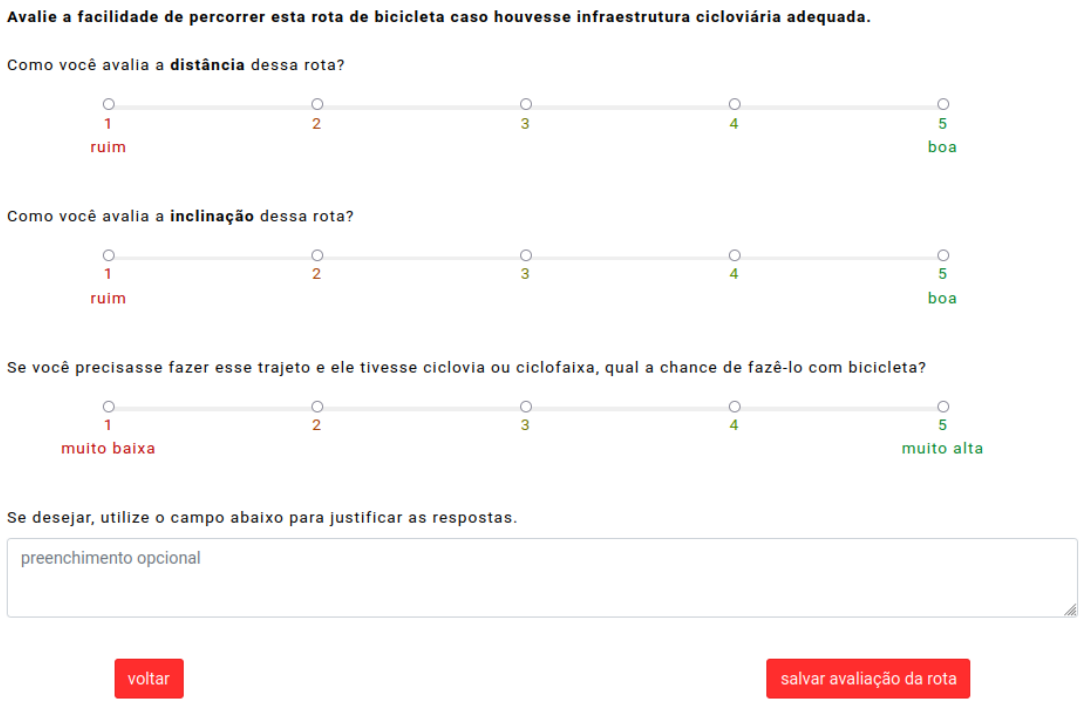


Figure B.3: *User survey questions for each route*

Por favor, clique nas rotas para ver mais informações e avalie respondendo as perguntas abaixo do mapa.
Rotas avaliadas: 0 / 5

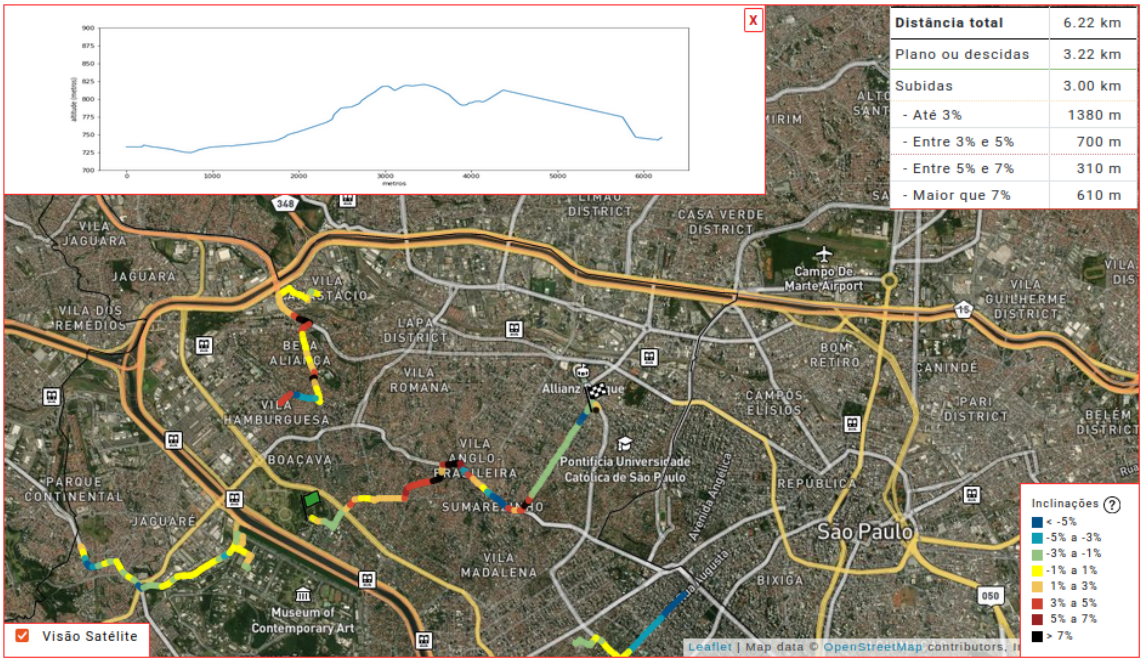


Figure B.4: *User survey map - satellite view*

References

[Benedini *et al.* 2020] Débora J. Benedini, Patrícia S. Lavieri, and Orlando Strambi. "Understanding the use of private and shared bicycles in large emerging cities: the case of sao paulo, brazil". In: *Case Studies on Transport Policy* 8 (2 June 2020). issn: 2213624X. doi: 10.1016/j.cstp.2019.11.009 (cit. on pp. 17, 18, 20, 51).

[Braun *et al.* 2016] Lindsay M. Braun *et al.* "Short-term planning and policy interventions to promote cycling in urban centers: findings from a commute mode choice analysis in barcelona, spain". In: *Transportation Research Part A: Policy and Practice* 89 (July 2016). issn: 09658564. doi: 10.1016/j.tra.2016.05.007 (cit. on pp. 13, 14, 18).

[Broach *et al.* 2012] Joseph Broach, Jennifer Dill, and John Gliebe. "Where do cyclists ride? a route choice model developed with revealed preference gps data". In: *Transportation Research Part A: Policy and Practice* 46 (10 Dec. 2012), pp. 1730–1740. issn: 09658564. doi: 10.1016/j.tra.2012.07.005 (cit. on pp. 13, 14, 33).

[Copenhagenize 2011] Copenhagenize Design Company. *Our Methodology*. 2011. url: https://copenhagenizeindex.eu/about/methodology (visited on 11/30/2021) (cit. on pp. 6, 11).

[EEA 2021] EEA - European Environment Agency. *CO2 performance of new passenger cars in Europe*. 2021. url: https://www.eea.europa.eu/ims/co2-performance-of-new-passenger (visited on 12/20/2021) (cit. on p. 61).

[CET 2021] CET - Companhia de Engenharia de Tráfego. *Mapa de Infraestrutura Cicloviária*. 2021. url: http://www.cetsp.com.br/consultas/bicicleta/mapa-de-infraestrutura-cicloviaria.aspx (visited on 12/02/2021) (cit. on p. 17).

[Fishman *et al.* 2014] Elliot Fishman, Simon Washington, and Narelle Haworth. "Bike share's impact on car use: evidence from the united states, great britain, and australia". In: *Transportation Research Part D: Transport and Environment* 31 (Aug. 2014). issn: 13619209. doi: 10.1016/j.trd.2014.05.013 (cit. on p. 5).

[IBGE 2021] IBGE - Instituto Brasileiro de Geografia e Estatística. *Cidades: São Paulo*. 2021. url: https://cidades.ibge.gov.br/brasil/sp/sao-paulo/panorama (visited on 12/02/2021) (cit. on p. 17).

[GraphHopper 2015] GraphHopper API. *Elevation data and OpenStreetMap*. 2015. URL: https://www.graphhopper.com/blog/2015/04/21/elevation-data-and-openstreetmap/ (visited on 12/15/2021) (cit. on p. 46).

[Harkot n.d.] Marina Kohler Harkot. "A bicicleta e as mulheres: mobilidade ativa, gênero e desigualdades socioterritoriais em São Paulo". DOI: 10.11606/d.16.2018.tde-17092018-153511. URL: https://doi.org/10.11606/d.16.2018.tde-17092018-153511 (cit. on p. 18).

[Hitge and Joubert 2021] Gerhard Hitge and Johan W. Joubert. "A nodal approach for estimating potential cycling demand". In: *Journal of Transport Geography* 90 (Jan. 2021). ISSN: 09666923. DOI: 10.1016/j.jtrangeo.2020.102943 (cit. on pp. 11, 12).

[Jackman 2000] Simon Jackman. "Models for ordered outcomes: standford political science lecture notes". In: (2000) (cit. on p. 7).

[Kon *et al.* 2021] Fabio Kon *et al.* "Abstracting mobility flows from bike-sharing systems". In: *Public Transport* (Mar. 2021). ISSN: 1866-749X. DOI: 10.1007/s12469-020-00259-5 (cit. on pp. 3, 63).

[Larsen *et al.* 2013] Jacob Larsen, Zachary Patterson, and Ahmed El-Geneidy. "Build it. but where? the use of geographic information systems in identifying locations for new cycling infrastructure". In: *International Journal of Sustainable Transportation* 7 (4 June 2013). ISSN: 1556-8318. DOI: 10.1080/15568318.2011.631098 (cit. on pp. 5–7, 13).

[Lemos 2021] Letícia Lindenberg Lemos. "Política, mobilidade e espaço: a bicicleta na cidade de São Paulo". Universidade de São Paulo, May 2021. DOI: 10.11606/T.16.2021.tde-16072021-212501 (cit. on pp. 1, 17, 18, 20).

[Lemos, Harkot, *et al.* 2017] Letícia Lindenberg Lemos, Marina Kohler Harkot, Paula Freire Santoro, and Isis Bernardo Ramos. "Mulheres, por que não pedalam? por que há menos mulheres do que homens usando bicicleta em são paulo, brasil?" In: *Revista Transporte Y Territorio* (2017), pp. 68–92 (cit. on p. 26).

[Lemos and Neto 2014] Letícia Lindenberg Lemos and Hélio Wicher Neto. "Cycling infrastruture in são paulo: impacts of a leisure-oriented model". In: *Spinoffs of Mobility: Technology, Risks & Innovation* (2014) (cit. on p. 28).

[Lovelace *et al.* 2017] Robin Lovelace *et al.* "The propensity to cycle tool: an open source online system for sustainable transport planning". In: *Journal of Transport and Land Use* 10 (1 Jan. 2017). ISSN: 1938-7849. DOI: 10.5198/jtlu.2016.862 (cit. on pp. 9, 12, 13, 16, 23).

[Ma *et al.* 2020] Xinwei Ma, Yufei Yuan, Niels Van Oort, and Serge Hoogendoorn. "Bike-sharing systems' impact on modal shift: a case study in delft, the netherlands". In: *Journal of Cleaner Production* 259 (June 2020). ISSN: 09596526. DOI: 10.1016/j.jclepro.2020.120846 (cit. on pp. 5, 33).

[Malatesta 2014] Maria Ermelina Brosch Malatesta. "A bicicleta nas viagens cotidianas do Município de São Paulo". Universidade de São Paulo, Mar. 2014. doi: 10.11606/T.16.2014.tde-04062014-102731 (cit. on pp. 17, 18).

[McGuckin and Murakami 1999] Nancy McGuckin and Elaine Murakami. "Examining trip-chaining behavior: comparison of travel by men and women". In: *Transportation Research Record: Journal of the Transportation Research Board* 1693 (1 Jan. 1999), pp. 79–85. issn: 0361-1981. doi: 10.3141/1693-12 (cit. on p. 22).

[METRÔ 2021] METRÔ – Companhia do Metropolitano de São Paulo. *Pesquisa de Mobilidade da Região Metropolitana de São Paulo: Síntese das Informações*. 2021. url: https://transparencia.metrosp.com.br/dataset/pesquisa-origem-e-destino/resource/b3d93105-f91e-43c6-b4c0-8d9c617a27fc (visited on 12/02/2021) (cit. on pp. 1, 4, 17, 21).

[Nazelle *et al.* 2011] Audrey de Nazelle *et al.* "Improving health through policies that promote active travel: a review of evidence to support integrated health impact assessment". In: *Environment International* 37.4 (May 2011), pp. 766–777. doi: 10.1016/j.envint.2011.02.003. url: https://doi.org/10.1016/j.envint.2011.02.003 (cit. on p. 1).

[Oliveira Vianna *et al.* 2021] Edison de Oliveira Vianna, Higor Amario de Souza, Edlene, Carneiro de Souza, and Fabio Kon. "Implantação e uso da ferramenta de análise de mobilidade de bicicletas bikescience na cet: identificando caminhos cicláveis em são paulo". In: *Revista UniCET* (Mar. 2021), pp. 21–43 (cit. on p. 3).

[Olmos *et al.* 2020] Luis E. Olmos *et al.* "A data science framework for planning the growth of bicycle infrastructures". In: *Transportation Research Part C: Emerging Technologies* 115 (June 2020). issn: 0968090X. doi: 10.1016/j.trc.2020.102640 (cit. on pp. 10, 11, 15, 34).

[OpenStreetMap 2021] OpenStreetMap. *Altitude*. 2021. url: https://wiki.openstreetmap.org/wiki/Altitude (visited on 12/15/2021) (cit. on p. 47).

[Parkin *et al.* 2007] John Parkin, Mark Wardman, and Matthew Page. "Estimation of the determinants of bicycle mode share for the journey to work using census data". In: *Transportation* 35 (1 Nov. 2007). issn: 0049-4488. doi: 10.1007/s11116-007-9137-5 (cit. on pp. 5, 13, 14, 17, 22, 33).

[Phillips and Range 2017] Luke Phillips and Andrew Range. "Development of a gis based toolbox for mapping cycling potential across scotland". In: 2017 (cit. on pp. 9, 13).

[São Paulo 2015] Dados Abertos - Prefeitura de São Paulo. *Região 8 - Divisão do Município de São Paulo em Oito Regiões*. 2015. url: http://dados.prefeitura.sp.gov.br/pt_PT/dataset/regiao-8-divisao-do-municipio-de-sao-paulo-em-oito-regioes (visited on 12/16/2021) (cit. on p. 48).

[SEADE 2014] SEADE – Fundação Sistema Estadual de Análise de Dados. "Perspectivas demográficas dos distritos do município de são paulo: o rápido e diferenciado processo de envelhecimento". In: *Resenha de Estatísticas Vitais do Estado de são Paulo* (Jan. 2014). URL: https://transparencia.metrosp.com.br/dataset/pesquisa-origem-e-destino/resource/b3d93105-f91e-43c6-b4c0-8d9c617a27fc (visited on 12/02/2021) (cit. on pp. 34, 59).

[Shaheen *et al.* 2013] Susan Shaheen, Elliot Martin, and Adam Cohen. "Public bikesharing and modal shift behavior: a comparative study of early bikesharing systems in north america". In: *International Journal of Transportation* 1 (1 Dec. 2013). ISSN: 22877940. DOI: 10.14257/ijt.2013.1.1.03 (cit. on p. 5).

[Sheller and Urry 2000] Mimi Sheller and John Urry. "The city and the car". In: *International Journal of Urban and Regional Research* 24 (4 Dec. 2000), pp. 737–757. ISSN: 0309-1317. DOI: 10.1111/1468-2427.00276 (cit. on p. 18).

[Silva *et al.* 2019] Cecília Silva, João Teixeira, and Ana Proença. "Revealing the cycling potential of starter cycling cities". In: *Transportation Research Procedia* 41 (2019). ISSN: 23521465. DOI: 10.1016/j.trpro.2019.09.113 (cit. on pp. 5, 6, 10, 12, 13, 15).

[AASHTO 2012] AASHTO - American Association of State Highway and Transportation Officials. "Guide for the development of bicycle facilities". In: (2012) (cit. on pp. 5, 23).

[Steer 2015] Steer. *Cycling Potential Index*. 2015. URL: https://www.yumpu.com/en/document/view/41572043/1-cycling-potential-index-steer-davies-gleave (visited on 11/15/2021) (cit. on pp. 8, 13, 22).

[The Audience Agency 2020] The Audience Agency. *Explanation: Mosaic*. 2020. URL: https://www.theaudienceagency.org/insight/mosaic (visited on 11/30/2021) (cit. on p. 8).

[Bank and KEMA 2014] The World Bank and DNV KEMA Energy and Sustainability. *THE LOW CARBON CITY DEVELOPMENT PROGRAM (LCCDP) GUIDEBOOK A systems approach to low carbon development in cities*. 2014. URL: https://www.eea.europa.eu/ims/co2-performance-of-new-passenger (visited on 12/20/2021) (cit. on p. 1).

[TFL 2017] Transport for London. *Analysis of Cycling Potential 2016*. Mayor of London, 2017 (cit. on pp. 8, 13).

[Tyndall 2020] Justin Tyndall. "Cycling mode choice amongst us commuters: the role of climate and topography". In: *Urban Studies* (Oct. 2020), p. 004209802095758. ISSN: 0042-0980. DOI: 10.1177/0042098020957583 (cit. on pp. 5, 12, 23).

[Verlinden *et al.* 2019] Yvonne Verlinden *et al. Increasing Cycling in Canada: A guide to what works*. 2019 (cit. on pp. 10, 12, 13).

[Watts 2018] Mark Watts. *How walking & cycling is transforming cities*. 2018. URL: https://www.c40.org/news/how-walking-cycling-is-transforming-cities (visited on 12/20/2021) (cit. on p. 1).

[Weir *et al.* 2020] Holly Weir, Asa Thomas, and Rachel Aldred. *The Propensity to Cycle Tool: Impact Report 2020*. 2020 (cit. on p. 9).

[Zhang *et al.* 2014] Dapeng Zhang, David José Ahouagi Vaz Magalhães, and Xiaokun (Cara) Wang. "Prioritizing bicycle paths in belo horizonte city, brazil: analysis based on user preferences and willingness considering individual heterogeneity". In: *Transportation Research Part A: Policy and Practice* 67 (Sept. 2014). ISSN: 09658564. DOI: 10.1016/j.tra.2014.07.010 (cit. on pp. 7, 12, 13, 17).