FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Understanding the impact of COVID-19 on mobility behavior of public transport passengers through data extraction

Hugo Daniel Gonçalves Fernandes

Mestrado em Engenharia Informática e Computação

Supervisor: Marta Campos Ferreira

Second Supervisor: Thiago Sobral

July 5, 2022

Understanding the impact of COVID-19 on mobility behavior of public transport passengers through data extraction

Hugo Daniel Gonçalves Fernandes

Mestrado em Engenharia Informática e Computação

Approved by . . . :

President: Rosaldo Rossetti
Referee: Marco Veloso
Referee: Marta Campos Ferreira

July 5, 2022

Resumo

A pandemia de COVID-19 impactou severamente a mobilidade das pessoas em todo o mundo, em particular o setor de transporte público. Por exemplo, na capital chilena de Santiago, durante a segunda semana de março de 2020, houve uma diminuição no número de viagens feitas em meios de transporte público de 55% e 45% para metro e autocarro, respetivamente [5]. Os mandatos de isolamento e trabalho em casa que causaram essas diminuições eram comuns em todo o mundo e ainda estão presentes em 2022 devido às ondas contínuas de infeções por COVID-19.

A pandemia provocou uma mudança nos hábitos de deslocação das pessoas, seja por necessidade ou por sensação de segurança. Se ou quando a pandemia desaparecer, não se sabe quantos desses cidadãos retornarão aos seus hábitos anteriores de deslocamento. O planeamento das redes de transporte público depende da procura estimada. As mudanças no comportamento dos cidadãos trazem novas incertezas a essas estimativas e tornam o processo de planeamento muito menos preciso. Verificar se essa tendência será difícil de reverter é vital para todos, porque os meios de transporte coletivo nos centros urbanos são uma das formas mais eficazes de evitar o excesso de emissões de carbono [34].

Esta dissertação visa proporcionar uma melhor compreensão dos comportamentos de mobilidade dos passageiros durante as diferentes fases da pandemia, utilizando o caso de estudo da Área Metropolitana do Porto (AMP), Portugal. Para tal, os dados de Automated Fare Collection (AFC) do primeiro ano de pandemia, 2020, foram analisados e comparados com o período homólogo do ano pré-pandémico de 2019. Os dados foram analisados de diversas formas. Foi realizada uma análise temporal a um nível mensal, diário e horário para compreender melhor a evolução da recuperação, destacando-se as diminuições imediatas do congestionamento dos transportes públicos nas horas de ponta após o confinamento e o gradual abrandamento da recuperação até à segunda vaga, que inverteu progressos de recuperação. Análise geográfica foi utilizada para determinar quais pontos da AMP foram mais afetados, mostrando um impacto mais percetível na zona mais utilizada, PRT1, e o crescimento de zonas mais pequenas. A informação das assinaturas e descontos foi também utilizada para analisar impactos em diferentes demografias, tais como idade e rendimento, o que revelou que os jovens com menos de 25 anos reduziram significativamente as suas viagens. Foram criadas representações visuais e utilizadas para analisar os conjuntos de dados para extrapolar estas e outras conclusões. Os resultados desta análise podem ser idealmente utilizados para o planeamento e alocação de recursos dos transportes públicos na área, quer em preparação no caso de outra pandemia ou crise de saúde pública, quer para informar as decisões de operações quotidianas ordinárias.

Abstract

The COVID-19 pandemic has severely impacted the mobility of people worldwide, in particular the public transport sector. For example, in Chile's capital city of Santiago, during the second week of March 2020, there was a decrease in the number of trips taken using public transport modes of 55% and 45% for metro and bus, respectively [5]. The isolation and work-from-home mandates that caused these decreases were commonplace worldwide and are still present in 2022 due to the continued waves of COVID-19 infections.

The pandemic caused a change in people's commuting habits, either due to necessity or a feeling of safety. If or when the pandemic subsides, how many of these citizens will return to their previous commuting habits is unknown. The planning of public transport networks depends on estimated demand. The changes in the behavior of the citizens bring new uncertainty to these estimations and make the planning process much less accurate. Verifying if this trend will be hard to reverse is vital for everyone, as collective transport modes in urban centers are one of the most effective ways to prevent excess carbon emissions [34].

This dissertation aims to provide a better understanding of the mobility behaviors of passengers during different stages of the pandemic, using the case study of the Area Metropolitana do Porto (AMP), Portugal. As such Automated Fare Collection (AFC) data from the first year of the pandemic, 2020, were analyzed and compared with the homologous period for the pre-pandemic year of 2019. The data was analyzed in diverse ways. Temporal analysis on a monthly, daily, and hourly level was performed to understand better how recovery evolved, highlighting the immediate decreases in rush-hour congestion of public transport after the lockdown and the slowing of recovery until the second wave, which reversed recovery progress. Geographical analysis to determine what different points of the AMP were affected, showing a more noticeable impact felt in the most significant zone, PRT1, and growth from smaller regions. The discount information was also used to analyze impacts on different demographics such as age and income, which revealed that young people under 25 significantly reduced their travel. Visual representations were created and used to analyze the data sets to extrapolate these conclusions and others. The results of this analysis can ideally be used for the planning and resource allocation of public transport in the area both in preparation in case of another pandemic or public health crisis or to inform ordinary everyday operation decisions.

Acknowledgements

I want to thank my supervisors, Prof. Marta Campos Ferreira and Prof. Thiago Sobral, as well as Prof. Teresa Galvão, for all their help and support. Completing this project on time would not have been possible without them. All the professors ever showed me was kindness and willingness to help whenever needed and motivated me to work on the days where motivation was lacking. I could not have asked for better supervision; thank you.

I also want to thank my parents for all the support they have given me during this project and throughout my academic life. Thank you for all the sacrifices made and for being there all the times I needed something urgent from you, from car trips to print material. And last but not least, thank you for financially supporting my studies all the way to the end of my Master's Degree; I would not have accomplished the things I have if it weren't for you.

Finally, I want to thank my friends, from those I made at FEUP to those from before I enrolled here, for all the good moments, the laughing and the crying, the good and the bad. You are the family I chose, so thank you for putting up with me because I know it's not always easy. Among them, I would like to particularly thank Sofia Oliveira, João Malheiro, and João Magalhães for all your support this semester.

Thank you for the all-nighters writing the thesis Sofia and for being one of the most amazing people I've had the pleasure of knowing. Thank you, J. Magalhães, not only for being the fantastic friend you have always been but specifically for the company during all the afternoons spent in different study halls, reading articles or implementing code with Henrique Santos and Francisco Cardoso, none of you are in the field of I.T. but having you there helped a lot. And thank you, J. Malheiro for being a fantastic friend and an excellent example of footsteps to follow, an engineer with a heart of gold.

All of you were tremendously important to me during this project, and I would not have finished it were it not for you. I can't thank you enough.

Hugo Fernandes

*"Musashi... **You** are the man who drives himself to all manner of extremes in the pursuit of the way of the sword.*
*But **you** are also the man who dreams of Otsū and agonizes himself over it. And the man who pulls pranks on me, that is also Musashi.*
*All those men are aspects of **yourself**. They are all part of **you**.*
*Be aware of **yourself**. And accept yourself as **you** are. That's were your training should begin."*

Takehiko Inoue

Contents

1 Introduction **1**
- 1.1 Context . . . 1
- 1.2 Portuguese Context . . . 3
  - 1.2.1 Restrictive States . . . 6
  - 1.2.2 Porto public transport . . . 8
- 1.3 Motivation and Objectives . . . 12
- 1.4 Outline . . . 13

2 State of the Art **14**
- 2.1 Literature Review . . . 14
  - 2.1.1 Methodology . . . 15
  - 2.1.2 Lockdown Effects . . . 16
  - 2.1.3 Relevant factors . . . 17
  - 2.1.4 Modal Shift . . . 18
  - 2.1.5 Trip Chaining . . . 19
  - 2.1.6 Perspective for the future . . . 20
  - 2.1.7 Work in Portugal . . . 20
- 2.2 Automated Fare Collection Systems . . . 21
- 2.3 Chapter Summary . . . 21

3 Data and Materials **22**
- 3.1 Data Sources . . . 22
- 3.2 RR & ROD data . . . 23
  - 3.2.1 Andante 2019 Expansion . . . 23
  - 3.2.2 Structure of RR data . . . 25
  - 3.2.3 Structure of ROD data . . . 27
  - 3.2.4 Significant differences and their provided value . . . 27
  - 3.2.5 October and November 2020 . . . 28
- 3.3 Auxiliary datatables . . . 28
- 3.4 Google Places data . . . 29
- 3.5 Chapter Summary . . . 29

4 Approach and Methods **31**
- 4.1 Problem . . . 32
- 4.2 Approach . . . 32
- 4.3 Data storage . . . 34
  - 4.3.1 Setting up the database environment . . . 34
  - 4.3.2 Problems . . . 34

4.3.3 Solutions . . . 35
4.4 Data filters . . . 36
4.4.1 Zone converter filter . . . 36
4.4.2 RR data filter . . . 38
4.4.3 ROD data filter . . . 40
4.4.4 Execution of the filtering process . . . 40
4.5 Main data queries . . . 41
4.5.1 Trips by zone per month . . . 41
4.5.2 Trips by day of the month . . . 41
4.5.3 Trips by weekday . . . 42
4.5.4 Trips by hour . . . 43
4.5.5 Trips by operator . . . 43
4.5.6 Trips by discount . . . 44
4.5.7 Most common destination zone . . . 44
4.6 Kepler.gl . . . 45
4.7 Identifying Stops in the datasets . . . 47
4.7.1 Filtering the auxiliary table . . . 48
4.7.2 Cross referencing of auxiliary data with AFC entries . . . 49
4.8 Google Places . . . 51
4.9 Chapter Summary . . . 52

5 Analysis **53**
5.1 Overall . . . 54
5.2 Zones . . . 57
5.2.1 The importance of PRT1 . . . 57
5.2.2 Behaviour observed . . . 58
5.3 Days of the Month . . . 60
5.3.1 March 2020 . . . 61
5.3.2 May 2020 . . . 61
5.3.3 July & August . . . 62
5.3.4 September 2020 . . . 65
5.3.5 December 2020 . . . 66
5.4 Days of the week . . . 67
5.4.1 Total Entries per weekday . . . 71
5.5 Hours of the day . . . 71
5.6 Operators . . . 73
5.7 Discounts . . . 77
5.7.1 Grouping similar discounts for redundancy and overall analysis of alternatives . . . 77
5.7.2 Discount Percentage . . . 80
5.7.3 Ages of users . . . 83
5.7.4 Estimated income . . . 85
5.7.5 Ages and income estimation . . . 86
5.8 Most common destination zone . . . 89
5.9 Stops & Google Places Data . . . 92
5.9.1 Metro stops . . . 92
5.9.2 CP stops . . . 93
5.9.3 STCP stops . . . 94
5.9.4 Google Places hot spots analysis . . . 95

5.10 Chapter Summary 105

6 Results and Discussion **106**
6.1 Effects of the public safety measures on total ridership 106
6.2 Recovery 108
6.3 Impact on transport use throughout the day 109
6.4 Impact on different demographics 110
6.5 Impact on different modes of public transport 111
6.6 August Decreases 112
6.7 Chapter Summary 113

7 Conclusions **114**
7.1 Conclusions and expected behaviours 114
7.2 Limitations 115
7.3 Future work 117

References **120**

A Code Snipets **123**

B Kepler Maps **126**

List of Figures

1.1 Hierarchy of States with indication of the issuing authority 6
1.2 Porto Metro Grid with color coded indications for each of its 6 service lines[13] . 10
1.3 Comboios de Portugal's AMP train grid stretching from the district of Porto to the districts of Aveiro and Braga[27] 11
1.4 STCP Daytime Bus Grid representing hundreds of bus stops present in the city[32] 12

2.1 PRISMA diagram 15

3.1 Andante Region before April 2019 with CP service line highlighted 24
3.2 New Andante Region limits and zones in operation since April 2019 25

4.1 Entity-relationship diagram for PostgreSQL database 34
4.2 Map generated from the results of the code present in listing 4.15 47

5.1 Plot of the number of trips each month 2019-20 55
5.2 Diagram showing the percentage shifts in total trips per month for the first half of the year 2020 56
5.3 Diagram showing the percentage shifts in total trips per month for the second half of the year 2020 56
5.4 Comparative plot of the number of trips for each month for 2019 and 2020 57
5.5 3D Map Representing the total number of trips per zone for January 2020 58
5.6 3D Map Representing the percentage of variations in total trips between 2019 and 2020 for each zone 59
5.7 Plot showing the total number of trips for the PRT1 zone per month for 2019 and 2020 60
5.8 Plot showing the total number of trips for the PV_VC zone per month for 2019 and 2020 60
5.9 Bar chart of trips by day of the month for March 2020 61
5.10 Bar chart of trips by day of the month for May 2020 62
5.11 Bar chart of trips by day of the month for July 2020 63
5.12 Bar chart of trips by day of the month for August 2020 63
5.13 Bar chart of trips by day of the month for July 2019 64
5.14 Bar chart of trips by day of the month for August 2019 65
5.15 Bar chart of trips by day of the month for September 2020 66
5.16 Bar chart of trips by day of the month for December 2020 66
5.17 Pie-chart of trips by weekday for 2019 67
5.18 Pie-chart of trips by weekday for 2020 68
5.19 Pie-chart of trips by weekday for the 3 month period of Jan-Mar 2019 68
5.20 Pie-chart of trips by weekday for the 8 month period of May-Dec 2019 69

5.21 Pie-chart of trips by weekday for the 3 month period of Jan-Mar 2020 70
5.22 Pie-chart of trips by weekday for the 8 month period of May-Dec 2020 70
5.23 Progression of entries per month for each day of the week over the two-year period 71
5.24 Trips per hour for each month of 2019 72
5.25 Trips per hour for each month of 2020 72
5.26 Total trips per hour for each of the years 73
5.27 Pie-chart of trips by operator for the 3 month period of Jan-Mar 2019 75
5.28 Pie-chart of trips by operator for the 8 month period of May-Dec 2019 76
5.29 Pie-chart of trips by operator for the 3 month period of Jan-Mar 2020 76
5.30 Pie-chart of trips by operator for the 8 month period of May-Dec 2020 77
5.31 Pie chart of Similar Discount demographics May-Dec 2019 79
5.32 Pie chart of Similar Discount demographics Jan-Mar 2020 80
5.33 Pie chart of Similar Discount demographics May-Dec 2020 80
5.34 Pie chart of Discounts grouped by Percentages May-Dec 2019 81
5.35 Pie chart of Discounts grouped by Percentages Jan-Mar 2020 82
5.36 Pie chart of Discounts grouped by Percentages May-Dec 2020 82
5.37 Pie chart of Discounts grouped by age of users May-Dec 2019 83
5.38 Pie chart of Discounts grouped by age of users Jan-Mar 2020 84
5.39 Pie chart of Discounts grouped by age of users May-Dec 2020 84
5.40 Pie chart of Discounts grouped by estimation of user income May-Dec 2019 . . . 85
5.41 Pie chart of Discounts grouped by estimation of user income Jan-Mar 2020 . . . 85
5.42 Pie chart of Discounts grouped by estimation of user income May-Dec 2020 . . . 86
5.43 Pie chart of Discounts grouped trough a mixture of age and income brackets May-Dec 2019 87
5.44 Pie chart of Discounts grouped trough a mixture of age and income brackets Jan-Mar 2020 88
5.45 Pie chart of Discounts grouped trough a mixture of age and income brackets May-Dec 2020 89
5.46 Metro stops heat-map for July 2020 93
5.47 CP stations heat-map for July 2020 94
5.48 STCP stops heat-map for July 2020 95
5.49 Google Places Data 96
5.50 Google places and Stop data: Hospital S. João May 2020 97
5.51 Google places and Stop data: Hospital S. João July 2020 97
5.52 Google places and Stop data: Hospital S. João September 2020 98
5.53 Google places and Stop data: Hospital S. João September 2019 98
5.54 Google places and Stop data: College campus May 2020 99
5.55 Google places and Stop data: College campus July 2020 99
5.56 Google places and Stop data: College campus September 2020 100
5.57 Google places and Stop data: College campus September 2019 100
5.58 Google places and Stop data: Downtown Porto June 2020 101
5.59 Google places and Stop data: Downtown Porto August 2020 101
5.60 Google places and Stop data: Downtown Porto August 2019 102
5.61 Google places and Stop data: Campo Alegre campus September 2020 102
5.62 Google places and Stop data: Campo Alegre campus July 2020 103
5.63 Google places and Stop data: Campo Alegre campus September 2019 103
5.64 Google places and Stop data: Matosinhos Coast August 2020 104
5.65 Google places and Stop data: Matosinhos Coast August 2019 104

B.1 2D Color Coded Trip per Zone January 2020 Map 127
B.2 2D Color Coded Trip per Zone February 2020 Map 128
B.3 2D Color Coded Trip per Zone March 2020 Map 129
B.4 2D Color Coded Trip per Zone May 2020 Map 130
B.5 2D Color Coded Trip per Zone June 2020 Map 131
B.6 2D Color Coded Trip per Zone July 2020 Map 132
B.7 2D Color Coded Trip per Zone August 2020 Map 133
B.8 2D Color Coded Trip per Zone September 2020 Map 134
B.9 2D Color Coded Trip per Zone October 2020 Map 135
B.10 2D Color Coded Trip per Zone November 2020 Map 136
B.11 2D Color Coded Trip per Zone December 2020 Map 137

List of Tables

1.1 Timeline of the main government policies implemented in 2020 [11, 19, 33, 26, 31] 3
1.2 Timeline of the main government policies implemented in 2021-2022 [11, 19, 26] 5

3.1 RR Dataset Structure 26
3.2 ROD Dataset Structure 27

5.1 Times each zone was the most common destination 2019-20 89
5.2 Times each zone was the most common destination in 2020 90
5.3 Times each zone was the most common destination in 2019 91

6.1 Comparison between July-September 2019 and 2020 112

Abbreviations

AFC	Automated Fare Collection
AMP	Área Metropolitana do Porto
CP	Comboios de Portugal
CSI	Complemento Solidário de Idosos
FEUP	Faculdade de Engenharia da Universidade do Porto
MP	Metro do Porto
STCP	Sociedade de Transportes Colectivos do Porto
TIP	Transportes Intermodais do Porto
WHO	World Health Organization
WoS	Web of Science

Chapter 1

Introduction

This chapter explains the context that led to this dissertation, both globally and locally, its motivation, its importance, the proposed approach, and the expected results. Section 1.1 introduces the global context of the COVID-19 pandemic concerning the public transport sector. Section 1.2 introduces the local pandemic context of Portugal, the measures implemented, and timelines corresponding to such measures. Section 1.3 explains the motivation behind the project and states some general objectives. Section 1.4 presents the structure of this document.

1.1 Context

In the last days of December 2019, the World Health Organization (WHO) saw the first reports of an outbreak of 'viral pneumonia' through a media statement by the Wuhan Municipal Health Commission. On the 1st of January 2020 the WHO activated the emergency response framework, over the following weeks the confirmed cases increased, the virus was identified as Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-Cov-2). By the end of January, the situation escalated. The WHO director-general declared the novel coronavirus a public health emergency of international concern (PHEIC), the highest level of alarm. The virus had spread to 18 countries, and human-to-human transmission was confirmed. Two weeks into February, the WHO named the new disease COVID-19, after which, it kept encouraging countries to stop the spread of the virus. Despite the efforts of the affected countries, the virus continued to increase its reach so much so that on the 11th of March of 2020, COVID-19 was officially declared a pandemic [22].

In the months that followed all across the world, infections were registered, some countries were severely affected, such as Spain and Italy [3]. Due to their preemptive measures, others managed to minimize the impacts of the first wave of infections. This wave was met with various governmental measures, from lockdowns and work-from-home mandates to cancellation of public events, travel bans mandatory mask-wearing, among others. Some countries started to see the end of the first wave, and the easing of restrictions began. However, the first wave of the virus was

not the only wave the world would have to deal with the appearance of other variants such as the delta variant, which appeared in late 2020, and the omicron variant that appeared in late 2021. Many countries were forced to re-impose public health measures after removing them whenever a significant wave of new cases presented itself.

During this two-year time frame from January 1st 2020 to January 1st 2022, the virus caused over 280 million infection cases and close to 5.5 million deaths all over the world[1]. It is safe to say that all sectors of the economy were affected by this pandemic.

This dissertation will focus in particular on the sector of public transport, as highlighted by Christidis [8] the transport sector is connected to the overall economy. As the world recovers from the virus, how it recovers will impact the sector, specifically mentioning that a "V-shaped" fast recovery would be ideal for minimizing effects and that an "L-shaped" slow recovery could cause changes to the supply and demand structure. Prior to the COVID-19 pandemic, public transport focused on efficiency and sustainability; however, this is now difficult. The focus shifted from transporting the maximum number of passengers per vehicle to ensuring that the vehicles were not full due to contagion fears. This necessary shift made transport providers have to re-adapt their planning [30].

Some countries even restricted or outright banned the use of public transportation during lockdown periods. In extreme cases, for example, India and Egypt, all transportation was banned, private, public, and mobility services as well [29, 24]

Contagion fears and governmental measures are not the only problems created by the pandemic for the public transport sector.

> ”The pandemic crisis is already considered as one of the greatest shocks in the last 60 years, strong enough to modify future needs and social values. Several of the changes in personal priorities may persist in time, even after the eventual recovery.” [8]

The mainstream adoption of remote working and video-conferencing, both for education and work, can result in a larger percentage of the working population and companies embracing those work methodologies, which could result in a reduction of overall mobility affecting the sector. Mentality shifts coupled with an increase in the prominence of e-commerce are expected to have a large impact and shift demand to privately owned vehicles; this undoubtedly would harm new mobility as a Service(MaaS) developments and businesses[8].

Considering these factors, eventual long-term impacts are highly likely. However, since currently, we cannot guess what they will be like, what can be done is to track how the effects, such as reduction of passengers or the disappearance of some types of trip-chains, change with decreases of cases and ’downtimes’ in between the waves and try to ascertain how the eventual post-COVID19 recuperation will be. During April 2020 in Australia, a high point for COVID-19 cases in the country, an 80% reduction in public transport when compared to pre-COVID19 periods, similar reductions are expected to have happened in most countries worldwide[20].

[1] WHO Corona Dashboard - https://covid19.who.int/ Accessed on February 2022

1.2 Portuguese Context

As of February 2022, Portugal has had over 3.2 million confirmed cases of COVID-19 and over 21 thousand deaths caused by the virus[2] a large percentage of the population has been affected by the pandemic directly, and all of the population's movements have been affected indirectly.

Portugal remained free of COVID-19 longer than most other countries in Europe. The first two cases were detected in Porto on the 2nd of March 2020; the first being a doctor returning from a vacation in Italy, and the second being a worker from Spain[3] [4]. After the first cases presented themselves, the country quickly implemented measures to flatten the infection curve. The main policies put in place are present in Tables 1.1 and 1.2. As we can see from the Tables, Portugal has been through multiple waves of infections and had restrictions imposed, taken away, and imposed multiple times again.

Table 1.1: Timeline of the main government policies implemented in 2020 [11, 19, 33, 26, 31]

Measure	Description	Date
Alert Level Declaration	On this day the government declared the highest level of alert possible for the COVID-19 pandemic	13th March 2020
Social Restriction	Restrictions started to be imposed on bars and restaurants limiting closing times and level of occupancy, nightclubs were closed	14th-16th March 2020
Travel Restrictions	Border with Spain is closed only remaining open for the transportation of goods, flights to and from Italy are cancelled.	13th-16th March 2020
Closing of Schools	Primary schools, middle schools, secondary school and universities are closed.	15th March 2020
State of Emergency	The country declares its first state of emergency, starting its first lockdown.	19th March 2020
Public Gatherings	Gatherings of more than five people are prohibited	2nd April 2020

Continued on next page

[2] Direção Geral de Saude https://covid19.min-saude.pt/wp-content/uploads/2022/02/723_DGS_boletim_20220223.pdf - Accessed on February 2022

[3] Diário de Notícias article - https://www.dn.pt/pais/confirmados-dois-primeiros-casos-de-coronavirus-em-portugal-11805920.html Accessed on February 2022

[4] Portuguese American Journal article https://portuguese-american-journal.com/porto-first-two-confirmed-cases-of-coronavirus-portugal/ Accessed on February 2022

Table 1.1 – continued from previous page

Measure	Description	Date
State of Calamity	The country begins to transition to a state of calamity, easing measures slowly	3rd May 2020
Easing of Restrictions	Small stores started to re-open, public transport is back with 2/3 capacity, work from home when possible, prohibition of gatherings over 10 people, libraries and public archives re-open, individual outdoor sports activities allowed	4th May 2020
Easing of Restrictions	The last 2 years of secondary school reopened along with restaurants, cafes, medium-sized stores and nurseries, all these with mandatory mask usage, museums monuments, art galleries allowed to open	18th May 2020
Beach Limitations	A law went into effect to combat overcrowding in beaches during the summer of 2020	25th May 2020
Easing of restrictions and public events	Community Religious ceremonies, work-from-home becomes partial, large scale stores open, kindergartens and day-cares re-open, cinemas, theaters, public showrooms re-open, official national football activities re-instated	30th May - 1st June 2020
Easing of restrictions	The country begins to transition to a state of contingency and alert at more local levels continuing to ease more measures.	1st August 2020

Continued on next page

Table 1.1 – continued from previous page

Measure	Description	Date
State of Contingency	With an rise in cases the country returns to a country-wide state of contingency, multiple measures are re-instated such as a reduction in maximum number of people allowed in indoor gatherings.	15th September 2020
In-person schooling	Even with an increase in cases the government deemed it necessary that schools returned to in-person classes, while universities mainly resorted to mixed teaching.	17th September 2020
State of Calamity	Reinstating of the state of calamity	15th October 2020
State of Emergency	Reinstating of the state of emergency	24th November 2020

Table 1.2: Timeline of the main government policies implemented in 2021-2022 [11, 19, 26]

Measure	Description	Date
Lockdown	Start of the second nationwide lockdown	15th January 2021
Closure of Schools	Government closes all school for at least 15 days. Classes resume on the 8th of February remotely.	21st January 2021
End of Lockdown	The country starts to lift lockdown measures slowly.	13th March 2021
Easing of Restrictions	Middle school resumes classes in person, restaurants re-open as well as cafes but with group restrictions	1st April 2021
End of State of Emergency	The state of emergency is ended.	30th April 2021
Time Restrictions	In some cities across the country restrictions are placed on being outside after 23:00 due to an increase in the number of cases	1st-30th July 2021
Removal of Restrictions	Bars nightclubs commerce and other public gathering events can now function as normal.	1st August 2021

Continued on next page

Table 1.2 – continued from previous page

Measure	Description	Date
Removal of Restrictions	Removal of the mandatory mask usage when walking the street	12th September 2021
Special Restriction period	Taking into account the festivities of the end of the year some restrictions were put in place which included mandatory work from home, closing of nurseries , bars and nightclubs, reduced capacity for stores and the prohibition of gatherings of 10 or more people in public for new year.	25th-10th December - January 2022
Removal of Restrictions	Removal of most restrictions still in place	17th February 2022

1.2.1 Restrictive States

As seen in Tables 1.1 and 1.2 the country transitioned multiple times through states of emergency, calamity, contingency and alert.

Portuguese law predicts three possible states to address catastrophes or severe accidents in the law of civil protection bases(law nº 27/2006) in conjunction with the state of emergency proclaimed in the Portuguese constitution article nº138. They are ranked in order of magnitude in Figure 1.1 their hierarchical structure is represented, as well as the entity responsible for issuing each state [7][23].

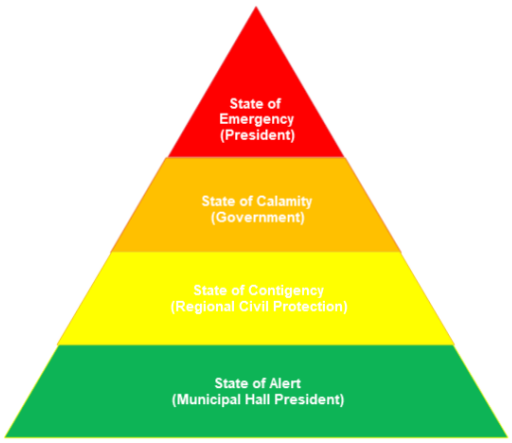

Figure 1.1: Hierarchy of States with indication of the issuing authority

During periods where one of these states has been issued, citizens that disobey orders from the authorities can be charged with a crime of disobedience which incurs a maximum sentence of one

year and four months of jail time; beyond this, each of the four states has particular characteristics and limits of power[7][23]:

- **State of Alert** - As the lowest level of severity, this state indicates the recognition of the necessity to adopt preemptive measures at the municipal level to deal with the distress situation. The issuing of this state falls to the municipal hall president or civil protection entity responsible for the area. The proclamation initiates the mobilization of institutional and political coordinating structures of civil protection at a municipal level, complemented by establishing the outstanding obligation for public communication outlets, television and radio, in particular, to divulge the relevant information relative to the situation[23].

- **State of Contingency** - As the next step in the hierarchy, this state can only be issued when it is determined that measures beyond the municipal level need to be taken to address the situation. The issuing of this state is the responsibility of the entity responsible for civil protection in the encompassing area. The declaration of a state of contingency immediately initiates the mobilization of institutional and political coordinating structures of civil protection at the respective territorial level and sees to the automatic activation of the civil protection emergency plans for the specific region[23].

- **State of Calamity** - As the highest of the three states present in the law of civil protection bases, the state of calamity allows for harsher public orders, such as restricting the movement of citizens or vehicles in certain zones, quarantining areas, reallocation of civilians for determined periods of time, rationing of essential goods such as water, electricity, food, etc. It can be declared on a national or a regional level; however, due to its more significant impact, only the government can declare a state of calamity. During a state of calamity, multiple extraordinary consequences exist: civil protection officers have the right to access private property and utilize private natural or energy resources strictly for use in the fulfilling of their actions; allowing for the temporary requisition of goods and services when urgency is verified for the public or national interest; allows for the dismissal of workers, officers or other public administration workers that also are civil protection or rescue officers; awards the government member in charge of health to take immediate emergency measures including the suspension of activities, separation of healthy and sick people, merchandise or means of transport that have been contaminated, these measures must respect the fundamental rights liberties present in the Portuguese constitution and law[23].

- **State of Emergency** - As the pinnacle of administrative measures that can be taken, declaring a state of emergency is reserved for only the direst of public calamities, having been issued for the first time in Portuguese democracy in March 2020 because of the pandemic. Proclamation of the state is done by the President after a hearing with the government and approval from the Portuguese Parliament and allows for a partial suspension of rights and liberties of citizens. If necessary, the armed forces can be used to reinforce civil authorities and provide support. The declaration also facilitates public authorities to take all necessary

provisions with highly reduced to no level of bureaucratic formalities. Public entities should, however, exercise restraint only, issuing necessary measures. During the state of emergency, some rights cannot be overridden by the declaration; these are the rights of superior constitutional dignity, rights such as the right to equality and non-discrimination, the right to life, the right to free speech, the right to personal integrity, the right to personal identity, among others. The state is valid for up to 15 days, after which it must be renewed[23].

1.2.2 Porto public transport

A previous study of the impacts of COVID-19 on the public transport usage patterns in Lisbon noted that a general decrease in demand is observed in both metro and bus/train networks for the year 2020. It also found a correlation between the decreases and the geographical positioning of different stations. Stations present in lower-income areas or periphery areas were less affected by the pandemic[4].

To the best of my knowledge, no study of this kind currently exists for the city of Porto. The city of Porto is the second-largest city in the country with its metropolitan area of over 2000 square kilometers [5], home to over 1.7 million people[6] in its district.

Porto's public transport sector comprises an extensive network of metro, bus, and train services. Analyzing how the pandemic impacted the behaviors of the city is essential to regulate urban planning for the city. The metro service alone provides trips for tens of millions of people every year, having reached a record 71 million trips in the year 2019; this number is followed closely by the number of trips registered for the leading bus provider of the region Sociedade de Transportes Coletivos do Porto[STCP] with over 69 million trips recorded, private transport operators that also use the Andante system had over 26 million trips recorded, and lastly, Comboios de Portugal(CP) registered over 8 million train trip validations [7]. Besides the three leading operators, others exist; notably, there are multiple smaller but prominent bus providers in the region, to name a few; Viabus, Transcovizela, Transdev, Mondinense, ETG, etc. Back in the end of 2002 Metro do Porto, STCP and CP joined together to create a business group called Transportes Intermodais do Porto(TIP), which is responsible for the management of the Andante ticketing system.

> " This grouping aims to guarantee the implementation of a ticketing and billing system common and exclusive to the members of the group and other adhering operators. Being also responsible for the revenue sharing model and the implementation and management of a communication and information system for the infrastructures and transport methods used " - Transportes Intermodais do Porto[8]

[5] Vila do Conde City Hall website regarding the Porto metropolitan area - https://www.cm-viladoconde.pt/pages/1230 Accessed on March 2022

[6] Jornal de Notícias https://www.jn.pt/nacional/distrito-do-porto-perde-mais-de-30-mil-residentes-em-13-dos-18-municipios-13983767.html Accessed on February 2022

[7] Management Report of Transportes Intermodais do Porto(TIP) - https://www.linhandante.com/uploadFiles/RC_TIP_2019-20200713.pdf Accessed on February 2022

[8] TIP About us page - https://www.linhandante.com/web/tip/tip-transportes-intermodais-do-porto Accessed on June 2022

On Figures 1.2, 1.3 and 1.4 are shown the grids for the city's metro service, main bus operator service and train service, respectively. All the aforementioned transport services can be used through the Andante card, an Automated Fare Collection System, the area that is covered by these networks, is subdivided into zones, when users charge the Andante trip-cards they choose what type of ticket they want based on how many zones they will travel through (Z2 for two, Z3 for three, etc.), for example, to complete a trip that crosses four zones a passenger would need to buy a Z4 ticket. There are multiple different types of ticket types and cards which in the data are referred to as tariffs[9][10] :

- Ocassional Ticket - individual paper ticket purchased directly from the bus driver that is only valid for that particular bus and does not allow for changing vehicle.
- Blue Andante Card and Título de Viagem - a rechargeable non-personalized paper card that allows for the user to purchase and charge multiple individual trip tickets for a type of zone, for example, a user could have 10 Z4 tickets in a card they could not. However, have 5 Z4 and 5 Z3 tickets, these cards are for sale in a large number of metro stops or ticket lines, each ticket as a lifetime of 1 hour minimum (varies with the number of the Z) during this time the passenger can use as many services as he desires without crossing the limit of zones from the one where the first validation happened.
- Andante 24 - a 24-hour ticket that allows the user to validate it as many times as he wants for 24 hours, akin to a daily ticket.
- Andante Pass- a personalized subscription-based card, paid monthly and non-transmissible. It can include all zones in the metropolitan area or just a combination of 3 zones for a smaller price. One must go to either Andante Stores or CP ticket lines to obtain this card.

There are records for all trips done using these cards; this is the data that will be analyzed and worked on within this project. One major limitation of the AFC system is its entry-only validation. This identifies when users entered and even change vehicles, but it does not tell us what the exit stop was.

Each trip also contains information about the type of discount each card is connected to. Discounts are applied to monthly subscriptions based on different criteria. Values range from no discount for normal monthly signatures to completely free for children under 13 years old. In between those extremes are three other levels of discounts. The widest-reaching discount bracket available is 25%, for children, students, unemployed adults, the elderly, and people that receive the Social+ B level of public support; The 50% discount is for adults who are recipients of Social+ A, the highest level of public support. And lastly, 60% this discount is for children and students under 23 years old whose household is eligible for Social+ A support.

[9] Andante ocasional tickets - https://www.linhandante.com/ocasionais.asp Accessed on March 2022
[10] Andante Cards - https://www.linhandante.com/cartoes.asp Accessed March 2022

There is nuance to these discounts, and when they can be applied, they have also changed during the period of 2019 and 2020 on two separate occasions making the analysis more complicated. Further explanation is given in Section 5.7.

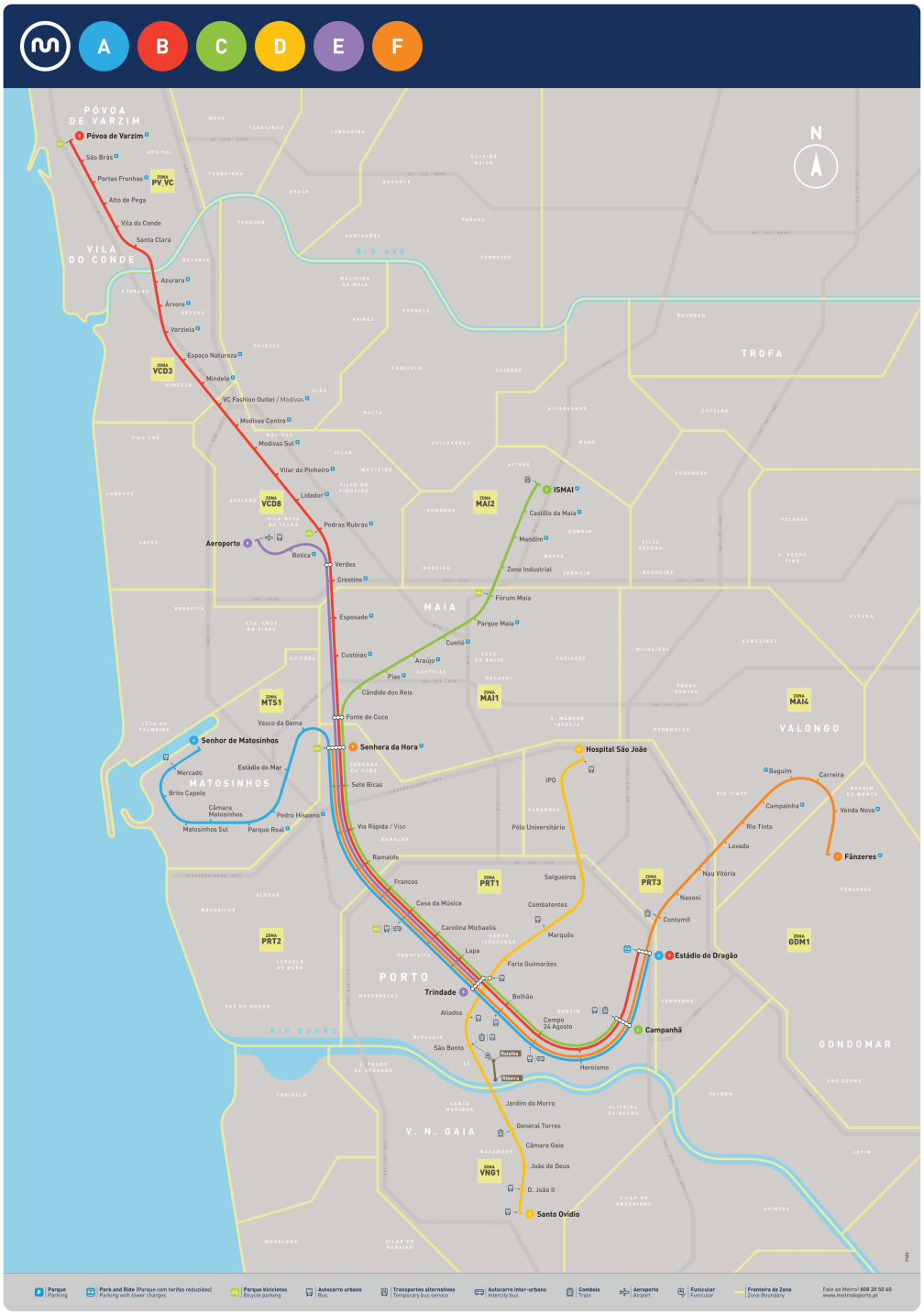

Figure 1.2: Porto Metro Grid with color coded indications for each of its 6 service lines[13]

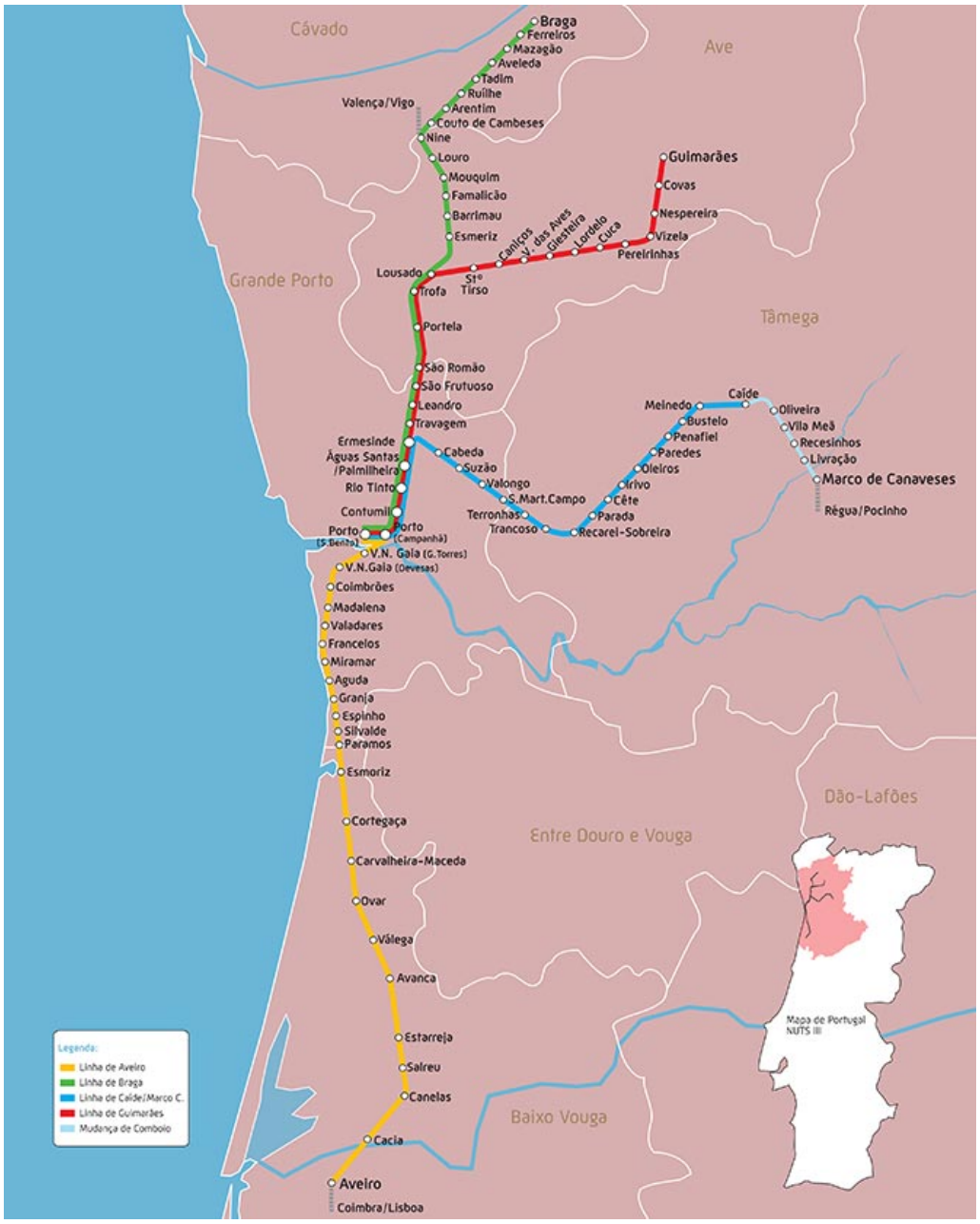

Figure 1.3: Comboios de Portugal's AMP train grid stretching from the district of Porto to the districts of Aveiro and Braga[27]

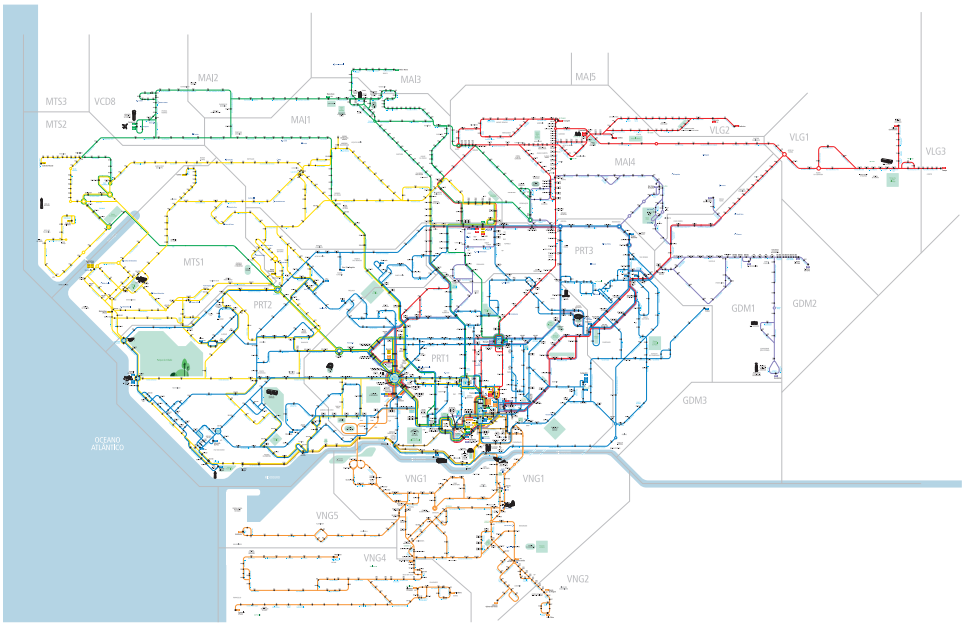

Figure 1.4: STCP Daytime Bus Grid representing hundreds of bus stops present in the city[32]

1.3 Motivation and Objectives

The COVID-19 pandemic has impacted the mobility of people who depend on public transport. Major cities like Santiago, Melbourne, and Taipei experienced decreased public transport usage [5, 10, 21]. For example, in Chile's capital city of Santiago, during the second week of March 2020, there was a decrease in the number of trips taken using public transport modes of 55% and 45% for metro and bus, respectively [5].

Lockdowns and work-from-home mandates that caused such decreases were commonplace worldwide and are still present in 2022 due to the continued waves of infections. If or when the pandemic subsides, how many of these citizens will return to their previous commuting habits is unknown. The planning of public transport depends on estimated demand. The changes in the behavior of citizens bring new uncertainty to the planning process; as such, verifying how difficult these decreases will be to revert is vital for the sector[34].

In order to understand the impact of COVID-19 on the mobility behavior of public transport passengers, a systematic review of existing literature was conducted following the PRISMA statement guidelines and checklist. Several focus points were present in the examined literature. Some studies focused on analyzing data and displaying emerging patterns, sometimes highlighting concepts such as trip chaining or missing-link trips and how these behaviors changed during the pandemic, other studies focused on the changes observed in the travelers' personal habits. Multiple different justifications based on different fields of study are conceivable and discussed in these papers, ranging from psychological and sociological reasons to geographical and economic ones.

Data used in the studies varies from case to case. Some data comes from surveys of the public transport users and some from smart-card data or route planner apps.

The literature analysis facilitates the identification of changes in the mobility patterns of public transport passengers. By identifying these changes and their effects on the sector, important insights can be gained to help improve public transport planning, prevent the worst effects in similar future situations, and deepen the understanding of what drives citizens to use public transport.

This dissertation aims to provide a better understanding of the mobility behaviors of passengers through analysis of data from the Automated Fare Collection (AFC) system of the Porto Metropolitan Area during different stages of the pandemic and pre-pandemic operations.

Data from 2019 and 2020 from the AFC system will be cleaned and filtered to ensure the best possible version of the data to prevent errors in the outcomes and to prepare the data for the exploratory analysis. In conjunction with the AFC data Point of Interest(POI) data will also be filtered and checked before initiation of the exploratory analysis. When both datasets are properly prepared exploratory analysis is performed on the data. The results of this analysis can ideally be used for the planning and resource allocation of public transport in the area.

By assessing the impact of COVID-19 in public transport use during the different stages of pandemic operations from the early days before lockdown measures to during lockdowns and to periods between waves of infections, we expect to find geographical heterogeneity in the effects of the pandemic, with urban centers feeling harsher effects than suburban and rural areas in the vicinities of cities due to the heightened use of public transport to access urban centers being highly affected by the mandates. Different types of trip-chains are expected to have been impacted differently as well. Presenting long-term results is still not possible due to the ongoing pandemic state; the pandemic will have to be over for a substantial time period before the long-term effects are accurately documented.

1.4 Outline

Beyond this introduction, this document has six chapters. Chapter 2 includes a literature review on the effects of COVID-19 on public transport and an introduction to such concepts as AFC systems. Chapter 3 presents the data used in detail. Chapter 4 explains the approach taken to implement all the relevant data and analysis environments and code. Chapter 5 presents the multiple different types of analysis performed, the most important images were selected and are described briefly. Chapter 6 discusses the results from the analysis and concludes the different scopes of each analysis type. Finally, in Chapter 7 to conclude this document, some of the main conclusions are summarized, the limitation the work faced are discussed, and possible opportunities for future work are presented.

Chapter 2

State of the Art

2.1 Literature Review 14
2.1.1 Methodology 15
2.1.2 Lockdown Effects 16
2.1.3 Relevant factors 17
2.1.4 Modal Shift 18
2.1.5 Trip Chaining 19
2.1.6 Perspective for the future 20
2.1.7 Work in Portugal 20
2.2 Automated Fare Collection Systems 21
2.3 Chapter Summary 21

This chapter presents the state of the art for this project and is comprised of a systematic review in Section 2.1 followed by an introduction and explanation of AFC systems in Section 2.2.

2.1 Literature Review

In order to understand the impact of COVID-19 on the mobility behavior of public transport passengers, a systematic review of existing literature was conducted. Several focus points were present in the examined literature. Some studies' focus was similar to this project's analyzing AFC data sets and seeing how the COVID-19 pandemic impacted a city. Other studies focused on the changes observed in the travelers' personal habits overall, giving different forms of reasoning to explain their changes in behavior ranging from psychological and sociological reasons to geographical and economic ones. Additionally, at times studies highlight concepts such as trip-chaining or missing-link trips and seek to evaluate how the behaviors surrounding them changed

during the pandemic. Data used in the studies varies from case to case. Some data comes from surveys of the public transport users and some from smart-card data or route planner apps.

2.1.1 Methodology

This review follows the PRISMA statement guidelines and checklist[25]. The databases used to conduct the search for articles were Web of Science (WoS) and Scopus. The search was performed in all fields of the papers indexed to the databases, using the following combination of terms: ("Public Transport*" OR Ridership OR "Travel Behavior*" OR "Urban Mobility" OR "Travel Pattern?") AND ("Covid-19" OR Sars-cov-2 OR "Covid" OR Coronavirus OR Pandemic OR "cov-19" OR "2019-ncov").

Two filters were applied to limit the search documents written in English and published between January 2020 and December 2021. There were 423 papers obtained from WoS and 451 from Scopus through this selection, resulting in 874 documents to scan. After removing duplicates and screening, 69 papers were selected and then read and examined in depth; besides these 69, there were three other studies from other sources. In Figure 2.1 this selection process is summarized in a PRISMA diagram.

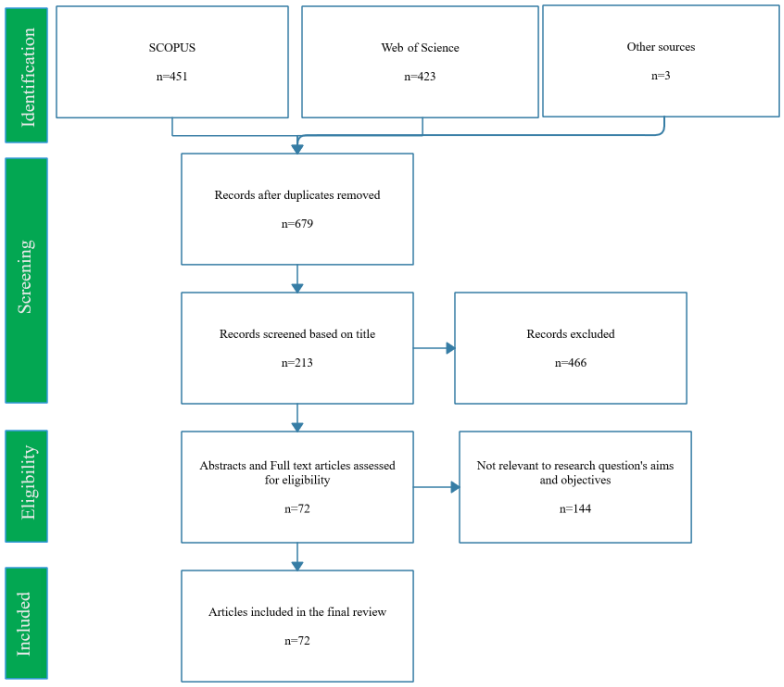

Figure 2.1: PRISMA diagram

At the time, not all of the 72 documents have been analyzed, only 23, in each of the following subsections, the more relevant points of said 23 studies will be approached. The full list of articles were not analysed in depth due to a lack of time. However they analysis is going to be done as

future work and presented at the 2022 edition of the European Transport Conference in Milan, Italy at the start of September.

Multiple diverse focuses were identified in the reviewed studies, from studies that focused on the effects of lockdowns, both during and after their conclusion, to studies that focused on trip chaining changes and modal shifts across transportation as a whole and studies that focused on future predictions.

2.1.2 Lockdown Effects

Lockdowns, naturally, severely diminish public transport use, when a city or nation is in a lockdown, Aloi et al. mention in their study on Santander Spain that a decrease of 80 to 90 % public transport use was present in multiple European cities from the 15th of March 2020 onwards, roughly the time at which lockdowns were widely imposed in Europe[3]. As for the city of Santander itself, an overall decline of 76% in mobility was identified. As for public transport, this decline was up to 93% [3]. The values of overall decreases can vary depending on the country due to the culture of the region in question or even the trust placed on the government orders. Such is the case identified in the study of Haddawy et al. as it was found that in a rural region of northern Thailand when a nationwide lockdown was issued, there was a 90% reduction in overall mobility [17]. Reductions of such magnitude are not limited only to rural areas. A study by Nailly et al. that analyzed nine years of ticketing data at a major transport hub in Paris highlighted the fact that during the period between 17th of March 2020 and the 11th of May 2020, there was an almost total loss of ridership[12].

The reductions to each mode of travel are also different in every situation, cities with a much broader and well-developed metro service than bus service might see similar effects in regards to percentages, but the impact on the metro would be much more expansive. A study by Astroza et al. on the impacts of the policies implemented by the Chilean government at the start of the pandemic noted an overall decrease of 44% in mobility, with the most significant decreases by mode of travel coming from metro services(55%), ride-hailing services(51%) and bus services(45%)[5]. As expected, reductions were also registered on automobile and motorcycle usage among other modes; however, these were the smallest reductions with 34 and 28 % respectively[5].

Different types of policies will also impact mobility differently. A study of over 41 cities in 22 countries by Vannoni et al.[36] found different percentages of declines correlated to what types of measures were in place. The results showed a 10% decline in mobility when school closures were enforced, 15% for workplace closures, 7% for the cancellation of public events, 18% for the closing of public transport services, 13% for the restriction of internal movements, and 5% for international travel restrictions. However, the authors found no relation found between monetary measures and mobility changes [36].

2.1.3 Relevant factors

Multiple studies have analyzed how the characteristics of a population influenced its response to COVID-19's effects on public transport. A Stockholm-based study by Almlöf et al. found that gender impacts the probability of stopping public transport usage and that areas with more significant percentages of males were more likely to stop using public transport. This study also noted that it found connections between age and a reduction in public transport use; however, as the study states, there are conflicting reports from other studies on the relevance of age as a factor, which leads to the belief that it might be different in different locations due to mentality and cultural shifts[2].

Evidence to support the claim that culture might have an impact on how these factors change across the world is present in a study by Politis et al. of the behavior changes caused by the COVID-19 lockdowns in Greece[24]. The study identifies evidence that before COVID, there was no noticeable difference in trip frequency between men and women, but within the context of the pandemic, men are more likely to perform trips, 30.3% more likely in cities with over 100 thousand citizens and 37.5% in cities with less than 100 thousand citizens[24]. This noticeable behavior change is attributed to some factors by the authors, such as women taking on roles of caregivers for sick family members or children who could not attend school and industries that are "women-dominated" being severely impacted by COVID, as is the case with air travel, tourism, retail, accommodation services, etc[24]. While the study by Almlöf et al. found that older people reduced their mobility more than younger people, Politis et al. find the opposite in their analysis of Greece, younger(18-25) people's mobility decreased more than people between the ages of 26 and 64. Of note is that people over 65 seemed not to decrease their travel even though they are more at risk of suffering severe consequences from the virus; this could be due to a low level of information or due to a low sample size of older people in the study data [24]. Finally, a connection was found between income levels and reductions in mobility; people with higher income were more likely to reduce their travels during the study period. However, there was no connection found between education level and reduction in travel, which is at odds with multiple other studies findings, the authors conclude that due to the impact of these factors, populations should not be considered homogeneous for future studies and research [24].

In contrast with Politis et al., Astroza et al. found that higher educated individuals were more likely to be able to work from home and, as such, reduced their travels more [5]. Beyond those findings, the authors also ascertain that larger households are more likely to have people working away from home and, to do so, engage in more travel. Their findings support the claims of Politis et al. with regards to women being more prone to working from home and correlate those findings to not only the impact COVID had on women-dominated industries but also the fact that men tend to engage in more health-related risks than women both during and prior to the pandemic[5]. This study also found that people over 60 reduced their trips much less than younger people, once again highlighting that even though they are an at-risk population, many older people still do not restrict their travels[5].

A study in Pakistan by Abdullah et al. found that in their model, gender, income, education, trip frequency, profession, ownership of a car or motorbike, and a safety-precautions underlying factor were all critical determinants of choice between public transport or solo modes of travel, and also found that both age and marital status were not significant[1].

It is apparent that some factors impact changes from location to location, as such, analyzing the study by Aparício et al.[4] done in the city of Lisbon is very important as with Lisbon being also a large city in the same country as our target city Porto it shares many cultural factors with it as well. Although this study is, much like ours, is more focused on the analysis of trip density rather than the characteristics of the population, it does conclude that higher-income areas of the city saw a much more accentuated decrease in usage when compared to areas outside the city center and as such public transport highlighted many socioeconomic disparities[4].

In a 2020 study, Tirachini and Cats[34] highlighted the impacts the pandemic can have on social equity, saying that it is expected to exacerbate disparities at a national but also international level. The emphasis was also placed on the income disparities in the use of public transport as wealthier individuals abandoned it, the authors said

> "(...) the vision of public transportation as a motor of social integration rather than of social segregation seems more distant today than ever." [34]

finally, the authors mention the risk that if the public transport sector is seen as an unhealthy mode of travel, this belief could linger even after the COVID-19 pandemic and hurt the sector's bottom line.

In an analysis of the smart card data for metro usage in Taipei, Mützel and Scheiner[21] found that the demand for metro services decreased strongly and seemed to remain lower after COVID waves had passed. The decreases were of particular note when comparing weekends to workdays (Monday through Friday). They noted that the decline of usage was considerably more significant for weekends and also noted that for workdays, the declines were smaller during rush hours and more prominent during the periods of lower traffic (10-16 and 20-01)[21]. Leading to the idea that non-essential trips were heavily reduced while work commutes remained imperative; this also is detrimental to overcrowding during rush hours.

2.1.4 Modal Shift

Reductions in public transport usage that are not proportional to the overall reductions in mobility mean that people swapped from using public transport to other modes of travel. The immediate thought that the swap to individual travel modes such as cars or motorbikes would be a prudent move is logical after all; it minimizes contact with other people, but those are not the only modes of transportation to benefit from a modal shift away from public transport, walking cycling and ride-sharing or ride-hailing are all modes that can see growth due to the pandemic.

In a study based on self-reported expectations of commuting in the city of Melbourne, Australia, Currie et al.[10] finds evidence to suggest that although the levels of use will recover, it will

not be a full recovery. An overall decrease of 6% in commuting overall and 20% for public transport(with bus notably only declining 4%) is what the authors expect. For the commercial business district of Melbourne specifically, a 20% decrease is expected to occur, but this is accompanied by a modal shift that sees the population move towards car driving, increasing the number of cars on the streets and worsening pollution fears, and creating increased traffic congestion [10]. The two leading causes identified for this reduction are the increase in work-from-home jobs and the fear of infection [10]. Accompanying the main changes, the study also finds that post-pandemic, there is an increase in the usage of bikes of over 40% and ride-hailing by over 10%; however, it is expected that walking, which saw an increase during the pandemic, decreases slightly[10]. A 2021 study by Munawar et al.[20] backs this claim, finding that a 46% decline in public transport was the catalyst not only for a shift towards cars but also walking.

An online survey was conducted from March 2020 to May 2020 in the city of Sicily, Italy, for a study by Campisi et al. [6] in this study; the authors found that women were 1.5 times more likely to reduce their walking frequency after the pandemic and that respondents that believed in the increase of micromobility options post lockdown were 2.8 times more likely to increase their cycling frequency respondents who supported remote work were even more likely to increase their bike-riding (4.2 times more likely than respondents that were neutral on remote work) [6]. The study finds that women also showed a smaller inclination to leave the house after the pandemic due to fear and to the state of the economy [6]. Finally, this study points to the need to create policies that push people away from car usage overall, but especially in city centers, [6].

The argument could be made that perhaps the decline in public transport usage is due to not only a decline in demand but also a considerable drop in supply; however, a study was conducted by Konečný and Brídziková[18] to analyze the changes in supply and demand due to the state of emergency. The main conclusion reached in this paper is that a reduction of 40 to 50% on demand was apparent, but in supply, the highest value registered for a region was 5% [18]. These findings may not be directly transversal to all areas of the globe; however, it brings some added context that allows us to conclude that the problem most likely is lack of demand and not supply.

To complement the findings of the previously mentioned study, the work of Šinko et al.[30] regarding the impacts of COVID-19 on modal shift can be used. In the paper, the results indicate that 40% of the participants recognized that the pandemic made them change their habits; also, 40% of participants said that they avoided public transport [30]. Finally, the study concludes that the overall modal shift is away from sustainable solutions and towards individual vehicles.

2.1.5 Trip Chaining

A trip chain is a travel sequence comprised of multiple destinations which begins and ends at one point of origin. In a study from 2011, Currie and Delbosc [9] traffic congestion and car parking were found to be determining factors for trip chaining to city centers. The study concludes that the car has many advantages the more complex a trip becomes; however, public transport would be a sensible option for city centers on some occasions.

To highlight the impact that increasing a trip chain's complexity can have on the feasibility of completing said chain by public transport, the paper by Walle and Steenberghen[37] will be referenced. In this paper, the authors conclude that due to the high number of "missing link" trips, which are trips where one of the destinations is not reachable by public transport, the car has significant advantages in fulfilling trip chains [37]. When initiating a trip chain by public transport, one has, according to Walle and Steenberghen, a roughly 50% chance to be able to complete the trip using solely public transport, this number drops to 40% for up to 5 trips and drops to 25% for trip chains larger than five stops long[37].

2.1.6 Perspective for the future

As pointed out by van Wee and Witlox [35] changes can be made to try and make public transport more appealing for commuters, from investing in extending existing links to increase the services reach to reducing crowding by having higher trip frequency [35].

A literature review conducted in 2020 Gkiotsalitis and Cats[15] concludes that strategical planning changes are likely to be few and that the focus should lay in operational and tactical changes from service frequencies to station limits being imposed. The authors highlight the paradigm shift that COVID-19 caused, changing the industry's focus from maximum efficiency by placing as many passengers into a vehicle as possible to spreading demand over time and, as such, diminishing the number of passengers present in the vehicle at a time.

A European Union comprehensive survey-based study of all 27 member states by Christidis et al. [8] showed positive results that point towards the eventual rebound of public transport usage; the study does mention that not all 27 member states will recover at the same rate, but it does present evidence as to why all likely will return to their pre-pandemic levels also stating that this rebound might take a while suggesting around three years after the pandemic is over [8]. Major threats to this recovery were identified, four in total they are: Mobility restrictions; Economic activity; Underlying trends in modal preferences, and Post-Pandemic behavioral change. If these factors are kept in check, the level of public transport ridership should return to normal.

2.1.7 Work in Portugal

The two most similar projects to this one are the analysis of AFC card data for the city of Lisbon to identify trends by Aparício et al.[4] and the analysis of AFC data to identify trends in Taipei by Mützel et al.[21]. Both projects utilize data sets comprised of AFC data; although they approach the data differently, both projects extensively provide visualizations of the trends identified and similar analysis of the results as to the expected results for this project.

Notably, the paper focused on the city of Lisbon is particularly important because the target demographic is similar to the target demographic of this dissertation, citizens of Lisbon, while different culturally from citizens of Porto, the overall Portuguese culture is a stable base across both groups. One notable finding by Aparício et al. that may be verified in the region of Porto is the noticeably more significant reductions in public transport use for higher-income zones [4].

Of note in Portugal, there have been two studies, the aforementioned Lisbon AFC card data study by Aparício et al.[4] and a study that uses a data-driven approach using the results of their mathematical analysis to attempt to understand the mobility patterns of the population by Tamagusko and Ferreira[33], in the latter study the change points were calculated for the mobility patterns. It was observed that even before the actual lockdown started, the Portuguese people were already in the process of changing their mobility patterns and changed rapidly again once the lockdown was lifted, which not only supports the claim by Christidis et al.[8] that the rebound from COVID-19 will happen but also shows that Portugal may have a fast to adapt population which could speed up the rebound process[33].

2.2 Automated Fare Collection Systems

Fare collection has evolved over the years from requiring people to buy tickets at a counter to magnetic stripe cards or even contact-less solutions such as allowing direct payment through bank cards or phone apps. These cards then are used in conjunction with read/write media, which can be in the form of ticket office terminals, ticket vending machines, fare gates, stand-alone validators, on-vehicle validators, or inspector validators. The information is transferred to back-office servers where the fare and media ownership management are handled[28] [1].

The existence of these systems is making life easier for consumers, reducing the time it takes for the process of purchasing trip tickets and everyday use of public transport, among other benefits for consumers as stated by Rubiano et al.,

> "AFC technology is making it much easier for commuters to access targeted benefit schemes, including subsidies toward specific population groups (lower-income users, the elderly, students), setting differentiated fares by time of day, introducing weekly transit passes, unlocking fare gates for disabled users, restoring the unused balance on a lost transport card, and even providing a credit trip when card balance falls to zero." [28]

2.3 Chapter Summary

In this chapter, we began by presenting the methodology behind the literature review. Withing the review, we presented the relevant information found in the articles for multiple topics, from the direct or indirect effects of lockdowns, relevant contextual factors, overall shifts in modal behavior, effects on trip chains, and outlooks for the future. We highlighted similar work previously performed in Portugal. Finally, we briefly explained Automated Fare Collection systems.

[1] Grand View Research - https://www.grandviewresearch.com/blog/automated-fare-collection-systems-ultramodern-to-the-core Accessed on February 2022

Chapter 3

Data and Materials

3.1 Data Sources 22
3.2 RR & ROD data 23
3.2.1 Andante 2019 Expansion 23
3.2.2 Structure of RR data 25
3.2.3 Structure of ROD data 27
3.2.4 Significant differences and their provided value 27
3.2.5 October and November 2020 28
3.3 Auxiliary datatables 28
3.4 Google Places data 29
3.5 Chapter Summary 29

This chapter presents the datasets and supplementary tables provided and used in this project. In Section 3.1 the origins of the data are explained in further detail, followed by an in-depth explanation of the two datasets that contain the AFC entries in Section 3.2 in which their differences are highlighted, and the relevance of each of them is explained. Followed by a look at the supplementary data Tables in Section 3.3 and how they provided valuable information to reach conclusions. Finally, the data from google places is addressed, and its use is explained in Section 3.4

3.1 Data Sources

The majority of the data utilized in this project was supplied by TIP. Being responsible for managing the Andante ticketing and billing system, TIP has the data of every entry created by an AFC card. The project analyzed AFC data from 2019 and 2020 to identify any type of valuable information related to COVID-19-caused behaviors for the recovery period after the first lockdown.

The data and materials used can be separated into these 3 groups:

- AFC Entries - These were the primary datasets used on the project. Each entry represents a travel validation using Andante cards(see Section 1.2) including multiple types of relevant information such as the entry zone, what kind of subscription the card responsible for the validation had, among others. There are two datasets with these entries ROD and RR.
- Auxiliary Materials - Group of materials encompassing all the extra data tables provided for better understanding of different parts of the AFC entries datasets, such as tables that allowed, through cross-referencing, for the identification of the stop where the validation occurred for STCP entries.
- Google Places data - Due to Google removing the option to use the Places API for academic purposes freely, the data used was from the previous work of João Gonçalves[16]. The data provides a way to attempt to find connections between the types of services present in an area and the changes in the use of public transport in the area before, during, and after high points of the COVID-19 pandemic.

3.2 RR & ROD data

The RR and ROD datasets are not two distinct parts of the total data; each has the totality of the data for most of the months in the two-year period. The exception is the start of 2019 for the ROD dataset; in 2019, the Andante network was expanded and re-structured, and the new fare zones and nomenclature came into operation in April 2019; until this month, the ROD datasets do not possess the complete information.

3.2.1 Andante 2019 Expansion

In April 2019, the Andante system was expanded to include all of the Área Metropolitana do Porto(AMP); this included the addition of multiple municipalities, redrawing zone limits for existing zones, and a change in the naming scheme of the zones for the whole region. The previous naming scheme was changed from one letter followed by a number shown in Figure 3.1.

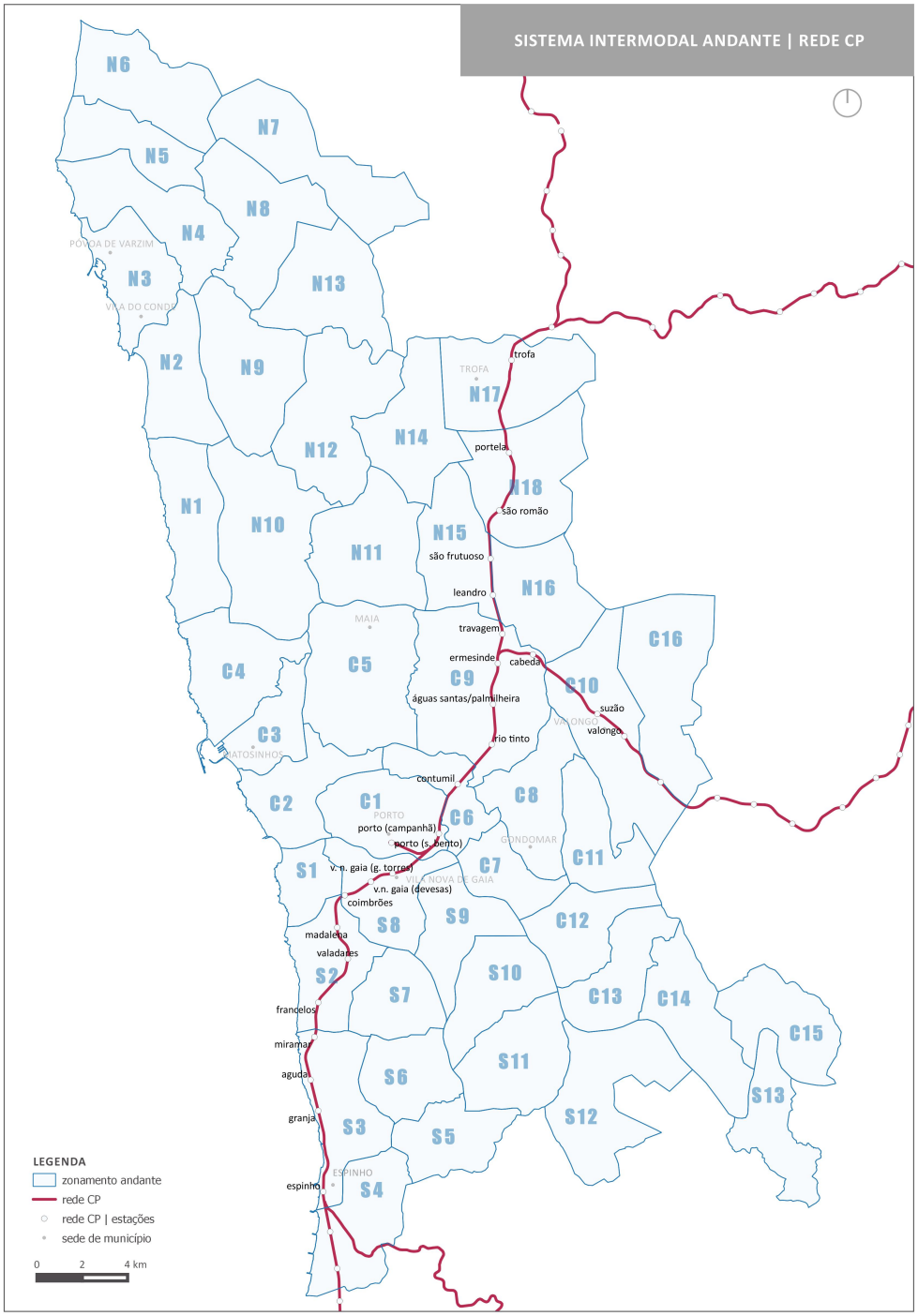

Figure 3.1: Andante Region before April 2019 with CP service line highlighted

The new more intuitive nomenclature is comprised of three initials for the area and a number as shown in Figure 3.2[1]. The number of zones also considerably increased due to the geographical expansion from 47 to 124.

[1]Portal AMP Expansão Andante - http://portal.amp.pt/pt/noticia/481 Access on June 2022

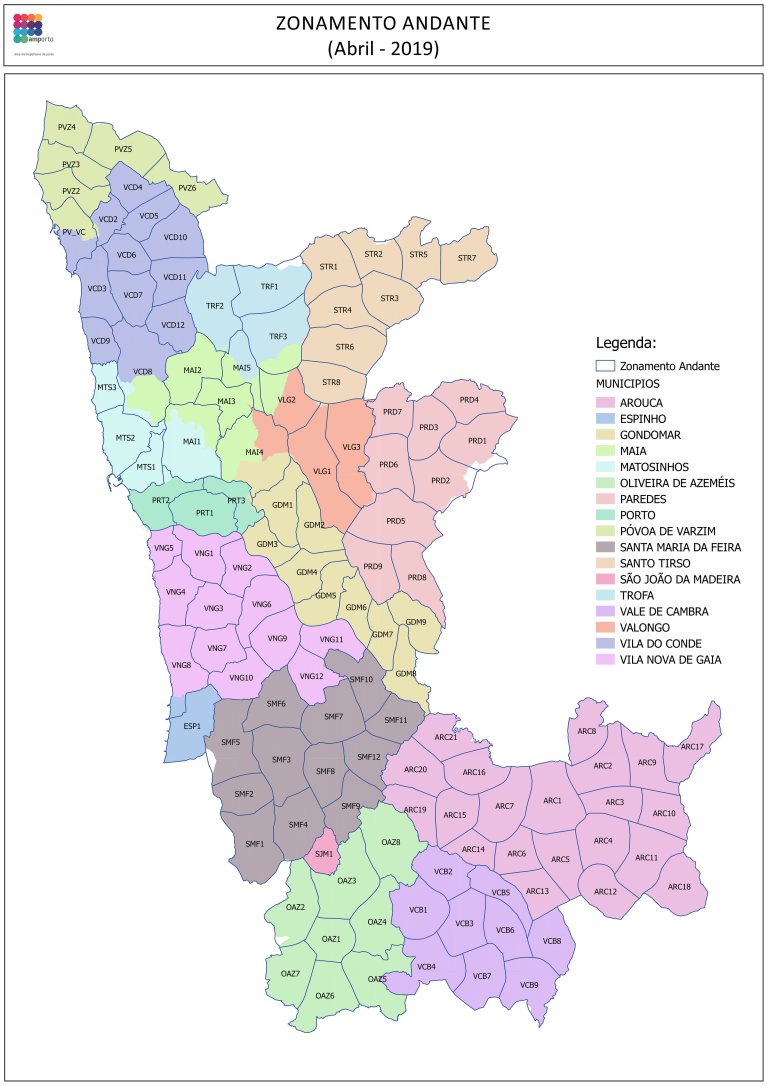

Figure 3.2: New Andante Region limits and zones in operation since April 2019

3.2.2 Structure of RR data

The RR dataset is the largest of the two AFC datasets, with 28 data columns. However, of the 28 columns, not all are significant. A large number of columns always present NULL values; these are signaled with an asterisk and are columns specifically directed at data from entries not from AFC cards but from the ANDA android application [14] [2]. The application data is not within the scope of this project and created data privacy problems; as such, it was not provided, which led to multiple columns that are of no value and would be filtered out eventually. The columns of the data set are presented and described in Table 3.1 in order[3].

[2] ANDA application - https://www.linhandante.com/web/tip/anda Accessed on June 2022

[3] The column names signaled with an asterisk(*) are the aforementioned application-specific columns

Table 3.1: RR Dataset Structure

Column	Description
Code	Ten digit numerical code that identifies the specific validation
Card ID	Hexadecimal code that identifies the specific card used
Validador	Ten digit hexadecimal code
Provider code	Numerical code that identifies the service provider for the transport method used in this entry
Discount code	Numerical code that identifies the discount or lack thereof the user possesses
Tariff code	Numerical code that identifies the kind of tariff paid for the trip(i.e. single trip ticket, 24 hour ticket)
Zoning	Number(2 or 5) that identifies if the entry was from before or after the Andante expansion
Zonemask	Numerical code to identify the combination of zones the users plan covers. This is of low relevance and as such is mainly ignored and even filtered out
DateTime of validation	Timestamp of the validation from the year down to the second
Entry zone	Zone where the validation occurred always present in the entries
Entry station	Station where the trip started, this value is NULL for all entries that are not for train or metro services
Exit zone	Zone of exit from the trip
Customer name*	String indicative of the customers name
Email*	email of the user
Phone*	phone number of the app user
Model*	model of the phone using the app
OS version*	version of android the phone has
Manufacturer*	phone manufacturer
Brand*	brand of the phone
Commercial model*	commercial model name of the phone
Operator*	service operator for the sim card present on the phone
Line*	name and number that indicates what line the trip belongs too
Direction*	string indicating which way the vehicle was going for the trip
Variant*	identifier of a trip variant if there is one
Vehicle*	vehicle identifier
Exit stop A*	id of the stop the user exits at
Exit zone A*	zone of the exit
End time A*	time at which the trip ended

3.2.3 Structure of ROD data

The ROD dataset is smaller than the previous one and does not have columns specific to the ANDA application, meaning fewer columns with no value. Table 3.2 shows the columns of the dataset and their respective small descriptions.

Table 3.2: ROD Dataset Structure

Column	Description
Code	Ten digit numerical code that identifies the specific validation
Card ID	Hexadecimal code that identifies the specific card used
Validador	Ten digit hexadecimal code
Vehicle	Identifier of the vehicle
Provider code	Numerical code that identifies the service provider for the transport method used in this entry
Discount code	Numerical code that identifies the discount or lack thereof the user possesses
Tariff code	Numerical code that identifies the kind of tariff paid for the trip(i.e. single trip ticket, 24 hour ticket)
Zoning version	Number(2 or 5) that identifies if the entry was from before or after the Andante expansion
Zone mask	Numerical code to identify the combination of zones the users plan covers. This is of low relevance and as such is mainly ignored and even filtered out
Internal line	Identifier of the internal line
Variant	identifier of a trip variant if there is one
Direction	code that indicates which direction is being traveled inbound or outbound
Validation stop	code indicative of stop of validation(STCP only)
DateTime of trip start	Timestamp of trip start mostly unusable
DateTime of validation	Timestamp of the validation from the year down to the second
Zone	Zone where the validation occurred always present in the entries

3.2.4 Significant differences and their provided value

Looking at Tables 3.1 and 3.2 clear resemblances between the datasets are seen. The columns: Code; Card ID; Validador; Provider code; Discount code; Tariff code; Zoning; Zonemask; DateTime of validation; Entry zone/Zone. They are the same across both datasets; however, enough differences are still present that make both datasets valuable and worth analyzing.

While both datasets have a column to indicate the zone where the passenger first validated the card in the form of Entry zone(RR) and Zone(ROD), only the RR dataset has a column that expresses an estimated exit zone. This extra information can be used in conjunction with the entry zone to analyze travel patterns and provides a unique value to the RR dataset, which was later made use of through a query to analyze the most common destination zone for each starting zone, see Section 4.5.7.

The second significant difference between the datasets is regarding stops. While RR has data that shows the stops for metro and train, the stops for bus services are not present and appear as NULL, whereas for the ROD dataset, the stops for bus services provided by STCP are indicated through a number that can be correlated to a stop code, but the stops for metro, train, and other bus services are NULL. With this in mind, to obtain data from both bus and metro services, both datasets are required, even if for bus services, in particular, only the STCP entries will provide stop information.

3.2.5 October and November 2020

As mentioned previously, the two datasets are not two parts of one whole but two different representations of the same entries. As such, in the case of a gap in the data provided, utilizing the other dataset as a replacement in queries that involve information shared across both datasets is a viable solution. Such is the case for October and November 2020.

The ROD dataset provided is missing entries for November 2020, and the RR dataset is missing the entries for October 2020. In some visual representations and data analyses, the ROD entries from October 2020 will be used in conjunction with the remainder of the RR dataset to provide the best representation of the complete entries for the two years.

3.3 Auxiliary datatables

To help interpret the AFC entries, supplementary data tables that can be cross-referenced with the codes from different columns were provided.

The columns relating to zonemasks and zoning had supplementary tables provided; however, since these columns were deemed of low significance and later filtered out, their auxiliary tables are also non-significant. Similarly, a supplementary table was meant to identify individual lines, which was deemed non-significant.

Several columns equal to both datasets utilize numerical codes to identify relevant information. The types of discount, tariff paid, and service provider are identified in this way.

Each entry's discount is one of 23 possible alternatives; these are identified with non-sequential numbers ranging from 0 to 35. Among the alternatives, some subgroups later can be joined together for analysis due to relevance.

Tariffs have 11 possible alternatives. While also being described in numbers, the codes fluctuate from 1 to 152. Tariffs are significantly less relevant than discounts as they only show the user's type of service plan.

For the AMP area, there are 30 service providers. In particular, bus service providers differ substantially in terms of operations size and the area they operate in. Analyzing this data can provide helpful information for revenue-sharing management alternatives showing which providers have higher usage.

Due to the 2019 Andante expansion(see Section 3.2.1) there were several changes to the zones system that existed before, such as names changing and borders being adjusted as such a table exists that allows for the most accurate conversion of the old zones to their new names or current equivalent, some cases are not precisely the same due to the border changes.

Identifying the STCP stops that each code refers to was an essential task in analyzing the ROD data; towards such a goal, there exists a table with all STCP stops and their numeric codes. However, this table's data has not been correctly treated for this purpose and does not contain geographical coordinates for the position of each stop which would be a necessity for visual representation.

To attempt to solve the problem with the previous table, an extra data table was acquired that related all stops in the AMP with their geographical coordinates. However, this table did not contain the STCP-specific identification codes.

3.4 Google Places data

The google places data comes in the form of .csv files that contain four columns:

- name column stating the name of each place
- type column classifying the place in one of several categories
- latitude column
- longitude column

The data files from the previous work of João Gonçalves encompassed around 75 thousand places for a rectangular area that covered most of the AMP[16].

With the help of code from the previous project, the existing file with all the data could be turned into several smaller files with data separated by the type assigned to each place. These could then be directly imported into a Kepler.gl map and represented graphically.

Although this data is not from the same period as the AFC data, instead of from a year later, the most relevant places that will be examined are long-standing institutions that would not be affected by such a gap in time as not to exist as they still exist today(i.e., S. João Hospital, Piolho coffee shop, etc.)

3.5 Chapter Summary

In this chapter, we addressed the data and materials used in the project, beginning with explaining the origins of the different data groups. After, we described in detail the composition of the AFC

entry datasets. A clarification of the auxiliary data tables is presented, and what they were utilized to achieve is discussed. Finally, the Google Places data is explained.

Chapter 4

Approach and Methods

4.1 Problem **32**
4.2 Approach **32**
4.3 Data storage **34**
4.3.1 Setting up the database environment 34
4.3.2 Problems 34
4.3.3 Solutions 35
4.4 Data filters **36**
4.4.1 Zone converter filter 36
4.4.2 RR data filter 38
4.4.3 ROD data filter 40
4.4.4 Execution of the filtering process 40
4.5 Main data queries **41**
4.5.1 Trips by zone per month 41
4.5.2 Trips by day of the month 41
4.5.3 Trips by weekday 42
4.5.4 Trips by hour 43
4.5.5 Trips by operator 43
4.5.6 Trips by discount 44
4.5.7 Most common destination zone 44
4.6 Kepler.gl **45**
4.7 Identifying Stops in the datasets **47**
4.7.1 Filtering the auxiliary table 48
4.7.2 Cross referencing of auxiliary data with AFC entries 49
4.8 Google Places **51**

4.9 Chapter Summary 52

This chapter addresses the approach used in this project to tackle the problem. The first Section 4.1 explains the problem and the need for this project. Following, Section 4.2 presents the devised plan of action and tasks. Section 4.3 is a detailed look into the project's data storage, the chosen approach's problems, and how they were solved. Section 4.4 explains the data filtering process performed. Section 4.5 explains the main database queries utilized to extract information from the data, their importance, and implementation. In Section 4.6 the visualization tool Kepler.gl is demonstrated. Section 4.7 explains the process behind extracting the information about Metro and STCP stops and train stations. Finally, Section 4.8 addresses the approach to using Google Places data from the previous work of João Gonçalves[16].

4.1 Problem

As mentioned previously in Sections 1.2 and 1.3, identifying mobility trends is vital for appropriately planning and structuring any public transport network. COVID-19 has impacted the sector harshly and continues to do so in some regions of the globe. Cities like Porto have to deal with the pandemic's effects on their service networks in the best way possible, and public transport providers need as much information as is available to distribute their routes, stops, and vehicles better.

With a population of around 1.7 million people, the public transport network of the greater district of Porto serves many people. Inefficient service or lousy route planning would affect thousands of lives. With the data available from the Andante system, mapping the routes and stops more critical to the population should be possible, as well as determining which routes are more or less critical in a pandemic situation and can be down-scaled or up-scaled in a lockdown situation.

When used in conjunction with the POI data identification of what effects recovery had on places in the city, for example, besides the massive drop in general use, the lines to commercial centers of the city and university campus would likely see an even more significant decrease in use due to closure of stores and video-classes during some periods as seen on Table 1.1.

Both datasets to be used in this project are extensive collections of data; this by itself introduces the need for special care when filtering and analyzing the data; as a complete human analysis of each entry is entirely impossible, big data approaches will have to be present in both parts of the process.

4.2 Approach

The primary data used in this project are two years' worth of AFC entries for the Andante system of the AMP; as such, the first step in our approach to this project is creating a support structure for the data in the form of a PostgreSQL database.

The data can have some inconsistent or irrelevant entries as the raw data supplied was not extensively filtered. Once the database was completed, filtering and cleaning the AFC data was the next step.

Once the data is appropriately checked and ready for use, work can commence on the exploratory analysis. Through specific queries to the database, results were obtained that could then be used for an exploratory analysis, which allows for the identification of patterns and reaching of conclusions; this will be the most significant part of this project. The outline of the approach is then as follows:

- Creation of the Database environment - Through analysis of the datasets provided, create an entity diagram and set up a database accordingly
- Filtering of the datasets - To handle the errors and internal test entries from the Andante System, among other problems, the data needs to be filtered
- Creation of the main queries - Creating, implementing, and running the main queries is the most significant step in the whole development process, as, without them, none of the analysis can be performed
- AFC data temporal analysis - Analysing how the AFC entries changed over the two years and correlating this with implemented policies
- AFC data spatial analysis - Analysis of how different regions of the public transport grid were affected by the pandemic using the Kepler.gl tool and Google Places data to look at specific points of interest.
- AFC data Spatio-temporal analysis - Crossing of the geographical analysis with the analysis of overtime effects and drawing conclusions about overall regions affected for each period by using the spatial representations created with Kepler.gl for different months and comparing them to each other
- Identifying the use of stops - Through the use of auxiliary data, the identification of entry and sometimes exit stops is possible for some AFC entries, although only possible for STCP, Metro, and CP services, these insights can provide valuable information
- Visual Representation of Google Places and Stops Data - Using the Kepler.gl tool to represent the stops data obtained as well as the data from Google Places provided
- Incorporate the visual analysis of AFC and the Places and Stops data - In this task, the analysis of AFC data and the stops and Places data will be related to determine possible impacts of the effects of COVID-19 over time for specific points of interest

Finally, with statistical and visual representations created, conclusions were drawn on what is believed to be some of the impacts of the COVID-19 pandemic, particularly the impact of the lockdowns and public safety measures.

4.3 Data storage

Proper data management is dependent on the working context; with this in mind and considering the type of data and physical hardware available, the conclusion was reached that a PostgreSQL database was the best way to perform operations on the data effectively.

4.3.1 Setting up the database environment

In order to set up the database environment, both primary datasets and most auxiliary materials were considered, and an entity-relationship diagram was created; this was then edited, considering which columns were relevant to the analysis. The final product can be seen in Figure 4.1, which was used as a base for the database environment.

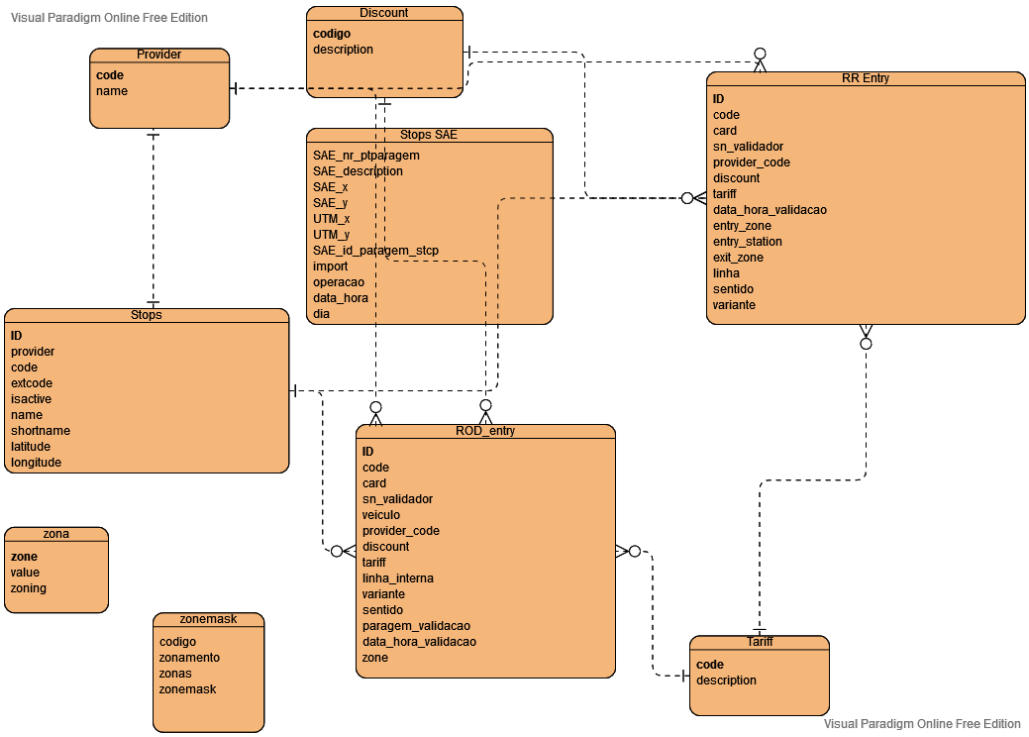

Figure 4.1: Entity-relationship diagram for PostgreSQL database

4.3.2 Problems

When accounting for both AFC datasets and all months, the data analyzed in this project is over 120GB in size. When working with datasets of this magnitude, multiple factors can introduce points of failure. Trying to access too much data at once can cause the computer running the scripts to crash or become extremely slow to the point of being unusable while running the program and increasing the total run-time exponentially.

During the implementation and initial testing of the database, it was clear that entirely uploading all the AFC entry data at once to the database would not be a good approach. Although the

PostgreSQL technology is and was valuable, the fact it stores the database on a local hard drive was a bottleneck due to hardware limitations. The only available disk with enough space to store the data was a Hard Disk Drive(HDD) with average read and write speeds; this means that running a query on a database that had too many entries could easily use up all available resources and effectively making the device unusable for the duration.

Although the increases in execution time were more pronounced on queries, mainly queries that selected data from all months of one of the data types, the effects were felt in every operation with the database, even during the upload of AFC entries. Uploading an average month of AFC entries would increase its run-time exponentially when more data was present, sometimes going from times of under ten minutes to over two hours.

These effects were confirmed using the *time* and *datetime* packages by adding snippet A.1 into all scripts.

As mentioned in Section 3.2, the two datasets contain mostly duplicate entries as of May 2019; however, this was only identified at the very end of the development and writing of the dissertation. Had this been clear from the start, it is possible the work would have been more streamlined, and the problem created by the size of the totality of the data would have had less impact on how the solutions were implemented.

4.3.3 Solutions

By the time the similarities between the two datasets were discovered, it was too late to adapt the initial solution. In this subsection are the steps taken to minimize the effects of the extensive database.

Since most of the main queries that were planned to be run could be done on a month-by-month basis instead of using all the data at once and then, if needed, using the results of said queries demonstrations for the overall progression over the two years could be created, the focus was shifted from creating a database and uploading all the entries into it to creating fast and easy ways to set up the database, upload the desired AFC entries for a specific month and year and then remove the data stored.

The first step was to create the PostgreSQL database using a python script, and this is a relatively simple and quick process, requiring some familiarizing with the use of the syntax used to perform actions[1]. In the A.2 code snippet, we see the import of the required package, the creation of the connection and cursor used to execute actions, the creation of the RR entries table, and the final closure of the cursor and commit of the changes.

In the same way, as the RR entries table was created in the snippet, the rest of the structure presented in 4.1 is created in the same python script; this allows for the quick setup of the entire database structure. Although the RR and ROD AFC entry tables were large, the same was not the case for the supplementary tables; a python script named bootstrapper was created in which the data for these tables was uploaded from their original files into the database. In the A.3 code

[1] Python Connect To PostgreSQL Server - https://www.postgresqltutorial.com/postgresql-python/connect/ Accessed in March 2022

snippet, as an example, the discount table that was previously created is populated with the values present in the original .txt file. Of note is that the connection to the database and cursor must be reinitialized at the start of each script since the connection should also be closed at the end of script execution; however, for all other scripts, that will be omitted.

After running both scripts, the database is ready to receive AFC entries from both data types. To upload the entries, another script was created entitled AFC entries uploader, which asks the user from which year he wishes to upload data 2019, 2020, or both, followed by asking the month of data to upload and then finally the data type. This script must consider the previously mentioned differences between the two data types and the fact that for the RR data entries, May and October of 2019 are special cases since the entries for said months are split into two files. In the A.4 below are some parts of the code from the uploader script, specifically both of the SQL queries and an example of the formatting of the strings for the cursor execution call for both years(ROD case) and for the May special case for only 2019(RR case). Of note is the need to use a LATIN9 encoding for the ROD data because some entries have special characters.

Finally, the delete script is the last step in creating the easy to set up, populate and delete database. Two options were created, a script to remove AFC entries without deleting the database, with options to delete one month of entries from either type of data, all entries from one of the types, or simply all entries, and a script to entirely delete the database. The A.5 code snippet is an example of the "Delete from" command being used in the first script. In the A.6 snippet are the instructions to delete the entire database.

The main use for the entries remover became occasions in which only part of the uploaded entries needed to be removed due to the entire database deleting and re-creating process was faster than deleting all entries with the first script in most situations.

Utilizing these scripts and creating the queries accordingly, we managed to circumvent the problems presented in the previous subsection to the best possible degree with the hardware in hand and our understanding of the two datasets at the time.

4.4 Data filters

As mentioned in chapter 3, both AFC entry datasets have columns that provide little to no value, coupled with entries that had errors or were test values remaining from internal system tests the datasets originate from. Due to the combination of those factors, filters were created for each dataset to remove all entries with incorrect values or test information.

4.4.1 Zone converter filter

Beyond each dataset's problems, there was the common necessity to convert the zones from the first four months of data from the old nomenclature to the new one to properly perform the analysis further in the project.

Initially, this conversion was idealized as a query that ran on the data on upload to the database; however, this proved extremely slow and was then replaced with a script that directly read from the files and converted the entries into a new file.

To achieve this, two lists of strings were created using the auxiliary conversion table, the lists of old and new zones. The new zones are ordered in the list to match the old versions at the respective index.

```
old_zones = ["C1","C2","C3","C4","C5","C6","C7","C8",...]
new_zones = ["PRT1","PRT2","MTS1","MTS2","MAI1","PRT3","GDM3","GDM1",...]
```

Listing 4.1: Lists of old and new zones ordered for conversion

After creating the lists, a small function seen in A.7 was created to compare the string passed as an argument with the old list one by one. When the same one is found, use the index and replace it with the element in the same position in the new zones list.

For the RR files, while there are still lines in the file, per line, the script splits the line by tab and gets the position of the entry zone and exit zone, removing the white spaces to get the zone codes seen in A.8.

Since RR files have two zones to convert, the associated logic is more complicated than for ROD files seen in. Verifications still have to be conducted: entry zone is a new zone or an old zone; entry zone is a code that is not a valid zone; exit zone being NULL a new zone or an old zone; exit zone is a non-valid zone code. Multiple different combinations of these verifications have to be accounted for, resulting in an extensive code listing:

```
if zone_2 == 'NULL': #if there is no exit zone
    if zone_1 in new_zones: #if entry zone is already correct
        file2.write(line)
    elif zone_1 in old_zones: #if entry zone is an old zone
        new_zone_1 = get_new_zone(zone_1)
        ...
    else: #if zone has an error does not exist
        print("zone doesn't exist")
else:
    if zone_1 in new_zones: #if entry zone is already correct
        if zone_2 in new_zones: #if exit zone is already correct
            file2.write(line)
        elif zone_2 in old_zones: #if only exit zone is an old zone
            new_zone_2 = get_new_zone(zone_2)
            ...
        else: #if exit zone is not a zone
            new_zone_2 = 'NULL'
            ...
    elif zone_1 in old_zones: #if entry zone is an old zone
        if zone_2 in new_zones: #if exit zone is already correct
```

```
            new_zone_1 = get_new_zone(zone_1)
            ...
        elif zone_2 in old_zones: #if exit zone is an old zone
            new_zone_1 = get_new_zone(zone_1)
            new_zone_2 = get_new_zone(zone_2)
            ...
        else: #if exit zone is not a zone
            new_zone_1 = get_new_zone(zone_1)
            ...
    else: #if entry zone isn't a zone
        if zone_2 in new_zones: #if the exit zone is already correct
            new_zone_1 = 'NULL'
            ...
        elif zone_2 in old_zones: #if the exit zone is an old zone
            new_zone_1 = 'NULL'
            new_zone_2 = get_new_zone(zone_2)
            ...
        else: #if the old zone doesn't exist
            new_zone_1 = 'NULL'
            new_zone_2 = 'NULL'
            ...
```

Listing 4.2: The if/else logic tree for RR conversion of entry and exit zones

Since there is only one zone in need of conversion on the ROD dataset, the logic behind obtaining the new line is much simpler:

```
    # ROD version
    zone_field = line_split[15]
    zone = zone_field.split(' ')[0]

    if zone in new_zones: #if zone is already correct
        file2.write(line)
    elif zone in old_zones: #if entry zone is an old zone
        new_zone = get_new_zone(zone)
        ...
    else: #if entry zone isn't a zone
        print("stop doesnt exist")
```

Listing 4.3: Extracting entry zone from ROD file line and if/else logic for entry zone conversion

This script was run before the individual data type filters.

4.4.2 RR data filter

As the RR dataset contains multiple columns related to ANDA data for each of these columns mentioned in Section 3.2.2 a check was performed on each entry; if any of the columns had a non-NULL value, the line was filtered out. The column would then be removed after this verification,

regardless of the result.

```
write = True
customer_name = line_split[12]
if customer_name != "NULL":
    write = False
...
line_split.pop(12)
```

Listing 4.4: Filtering out lines with information in the application columns

Besides the application columns, the RR filter removes the columns for "zonemask" and "zoning" due to lack of relevance for future analysis.

System test entries had a particular value attributed to SN_Validador to check if an entry was originated by a test the SN_Validador column was checked for the string "FFFFFFFFFF". Beyond tests, the provider, discount, and tariff values were checked with lists containing all possible valid values of each.

```
sn_validador = line_split[2]
if sn_validador == "FFFFFFFFFF":
    write = False
provider = int(line_split[3])
if provider not in providers:
    write = False
discount = int(line_split[4])
if discount not in discounts:
    write = False
tariff = int(line_split[5])
if tariff not in tariffs:
    write = False
```

Listing 4.5: Verification of valid SN_Validador; provider; discount; tariff

Afterward date of validation is verified, the year and the month are compared to the expected year and month of the file being filtered if they do not match, the flag is raised.

```
date = line_split[6].split(' ')[0]
if date.split('-')[0] != year or date.split('-')[1] != month:
    write = False
```

Listing 4.6: Verification of accurate date

Finally, the entry station is verified. If this entry is NULL, nothing happens besides the increment of a counter; if there is an entry with the "/" character, this is changed for "-CP" to indicate this stop is a train stop for future data analysis.

If, after all verifications, the write flag is still TRUE, the line is written into the new file.

4.4.3 ROD data filter

The ROD data filter performs some of the same operations as the RR filter, specifically, the verifications for: SN_Validador; provider; discount; tariff; date. Moreover, it does not need to perform the operations for the ANDA application-specific columns.

The ROD filtering removes the columns for "zonemask", "zoning" and "DateTime of trip start" due to lack of relevance for the analysis since the DateTime variable only indicates when the bus began its current trip, and these entries are often erroneous values.

Finally, the unique validation needed in this filter is to validate the stop codes present in some entries against the database of STCP stops. To validate this, first, a list of all STCP stop codes is imported, against which each line's stop code is compared; if the value is not viable, it is replaced with NULL.

```
1  while True:
2      line = codes_file.readline()
3      if not line:
4          print("End of codes parsing")
5          break
6      code = int(line)
7      codes.append(code)
8  ...
9  stop = line_split[10]
10 if stop == 'NULL':
11     count_nulls += 1
12 elif int(stop) not in codes:
13     line_split[10] = 'NULL'
```

Listing 4.7: Importing STCP stop codes from auxiliary materials and checking existence of the stop code

4.4.4 Execution of the filtering process

Having developed the zone converting filter and both data filters, the process of executing the filtering was set underway. With millions of entries per file for a total of 44 files(excluding both versions of April 2020), the process was long, many times causing hardware problems as it was attempted to be conciliated with other operations.

As mentioned in Section 4.4.1, the possibility of performing the filtering on the data in the database environment was discarded due to not only the increase in execution time created by a storage drive bottleneck but also the need to perform this after each upload if the original files uploaded were not changed. New versions of the files were created with the epithet "_filtered"

added to the end, and these became the new files that were uploaded by the scripts in Section 4.3.3.

4.5 Main data queries

Having resorted to a database environment to hold the data, the extraction of information was performed mainly through queries to the database. The queries used were considered likely to produce statistically relevant data for analyzing behavioral changes over the two years.

4.5.1 Trips by zone per month

Obtaining data on how many trips there are per month in each zone allows for the subsequent mapping of variations along the two-year period, which is instrumental in assessing the impacts of the virus and the public safety measures in each zone; as well as showing which areas of the AMP have a higher volume of trips.

Having been the first main query to be implemented, it works slightly different from other queries, as it does not inquire which type of dataset the user wishes to query and instead queries the database for both types one at a time for a specific month for both years. Other main scripts allow the query to be directed at a single year or dataset. This difference is one of the reasons why knowing that ROD and RR are duplicate entries represented differently would have extremely reduced execution times as this query was the highest in terms of execution time.

```
1 sql_rr = '''SELECT COUNT(*) FROM rr_entries
2 WHERE EXTRACT(YEAR FROM data_hora_validacao) = {}
3 AND EXTRACT(MONTH FROM data_hora_validacao) = {}
4 AND entry_zone ~ '{}';'''
5 ...
6 for zone in zones:
7     cursor.execute(sql_rr.format(year, input_month, zone))
8     result = cursor.fetchone()[0]
```

Listing 4.8: Code Snippet from trips by zone per month query

4.5.2 Trips by day of the month

Beyond the primary information on each zone, the previous query offered data in the scope of a whole month. For some aspects of analysis, a higher level of granularity is needed.

When multiple impactful public safety measures are lifted in the same month, such as in May 2020, as shown in Table 1.1, simply with the information from the whole month, not much information can be attained regarding the individual measures. With this in mind, a query was created showing the number of trips registered daily.

```
sql_rr = '''SELECT COUNT(*)
FROM  rr_entries
WHERE EXTRACT(YEAR FROM data_hora_validacao) = {}
AND EXTRACT(MONTH FROM data_hora_validacao) = {}
AND EXTRACT(DAY FROM data_hora_validacao) = {};'''
...
def trip_day(data, y, m):
    if y == 2020 & m == 2:
        days = 29
    else:
        days = number_days_per_month[m - 1]
    ...
    while i < days:
        i += 1
        cursor.execute(sql_rod.format(y, m, str(day)))
        result = cursor.fetchone()[0]
    ...
```

Listing 4.9: SQL Query and use implemented for the trips per day script.

4.5.3 Trips by weekday

As mentioned in chapter 2 differences were found when comparing workday usage to weekends [21]. To assess if the recovery in the case of the AMP happened uniformly for different days of the week, a python script was created that returned the number of trips for each day of the week.

```
def trip_weekday(data, y, m):
    record = [0, 0, 0, 0, 0, 0, 0]
    if y == 2020 & m == 2:
        days = 29
    else:
        days = number_days_per_month[m - 1]
    i = 0
    while i < days:
        i += 1
        if data == 1:
            cursor.execute(sql_rod.format(y, m, str(day)))
        elif data == 2:
            cursor.execute(sql_rr.format(y, m, str(day)))
        result = cursor.fetchone()[0]
        index = int(datetime.datetime(y, m, day).strftime("%w"))
        record[index] += result
        ...
```

Listing 4.10: Obtaining the number of trips per weekday through the use of datetime

4.5.4 Trips by hour

After presenting some preliminary conclusions of this project in the 14th edition of the annual CITTA conference on planning research[2], receiving some feedback from researchers in the area and attending presentations of other projects. The analysis of how the behaviors of passengers changed in terms of travel density during the day, and if some particular times of day were impacted more harshly by the lockdown and public safety measures was deemed valuable to analyse.

As such, we utilize the DateTime column with the validation timestamp and compare its year, month, and hour of recording and count the number of trips that match each hour for all the days in the month.

```
sql_rr = '''SELECT COUNT(*)
FROM  rr_entries
WHERE EXTRACT(YEAR FROM data_hora_validacao) = {}
AND EXTRACT(MONTH FROM data_hora_validacao) = {}
AND EXTRACT(HOUR FROM data_hora_validacao) = {};'''
...
def trip_hour(data, y, m):
    record = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
    i = 0
    while i <= 23:
        if data == 1:
            cursor.execute(sql_rod.format(y, m, str(i)))
        elif data == 2:
            cursor.execute(sql_rr.format(y, m, str(i)))
        result = cursor.fetchone()[0]
        record[i] += result
        ...
```

Listing 4.11: Obtaining and storing the number of trips for each hour of a day for a month

4.5.5 Trips by operator

The Andante network being comprised of multiple service providers means that understanding if any of them were disproportionally affected by the pandemic recovery is essential. The total number of trips per month for each operator was collected for each month to assess the impacts.

Most operators only provide services in subsections of the AMP, which entails that there might be some correlation between the impact felt on a particular zone and the impact felt by a particular operator.

```
sql_rr = '''SELECT p.name as provider_name, r.provider_code as code,
  COUNT(r.provider_code) AS trips
```

[2]CITTA Conference on Planning Research - https://paginas.fe.up.pt/ citta2022/ Accessed June 2022

```
FROM  rr_entries r, providers p
WHERE
  r.provider_code = p.id
  AND EXTRACT(YEAR FROM data_hora_validacao) = {}
  AND EXTRACT(MONTH FROM data_hora_validacao) = {}
GROUP BY
  provider_name, r.provider_code
ORDER BY
  trips DESC;'''
...
```

Listing 4.12: SQL query to obtain the number of trips for each provider code in a specific month

4.5.6 Trips by discount

The different discount alternatives are provided to people in need of social aid or from specific age groups; the discounts provide an opportunity to analyze some social impacts.

Through the information extracted from the total number of trips performed by people in each discount category, we can obtain valuable insights into the impact of COVID-19 and the lockdown recovery on different social groups. By analyzing if there were unexpected increases or decreases and investigating the social context in that particular time window to identify possible relevant factors.

```
# SELECT RR_ENTRIES by month and zone
sql_rr = '''SELECT d.description as discount_name, r.discount as code,
  COUNT(r.discount) AS trips
FROM
  rr_entries r, discounts d
WHERE
  r.discount = d.code
  AND EXTRACT(YEAR FROM data_hora_validacao) = {}
  AND EXTRACT(MONTH FROM data_hora_validacao) = {}
GROUP BY
  discount_name, r.discount
ORDER BY
  trips DESC;'''
```

Listing 4.13: SQL query to obtain the total number of trips for each discount alternative in a specific month

4.5.7 Most common destination zone

As mentioned in Sections 3.2.2 and 4.4.2, the RR dataset has a unique property in some entries, including the start and end zone of the trip. Using these columns, we can produce a query to

analyze which zone is the most common destination for trips from all other zones. These, however not present for all trip entries, and the reasons why some trips have this information and some do not is not fully understood; we believe it to be representative of the overall trip tendencies still and, as such, developed this query to evaluate points of centralization of traffic.

```
sql = '''SELECT exit_zone,
  COUNT(exit_zone) AS value_occurrence
FROM
  rr_entries
WHERE entry_zone ~ '{}'
AND EXTRACT(YEAR FROM data_hora_validacao) = {}
AND EXTRACT(MONTH FROM data_hora_validacao) = {}
GROUP BY
  exit_zone
ORDER BY
  value_occurrence DESC
LIMIT 1;'''
...
for zone in zones:
    cursor.execute(sql.format(zone, year, input_month))
    if cursor.rowcount == 0:
        print('The zone ' + zone + ' has no uses')
    else:
        destination = cursor.fetchone()
        if destination[0] is None:
            print('The zone ' + zone + ' has no uses')
        else:
            result_zone = destination[0].split(' ')[0]
            number_trips = destination[1]
    ...
```

Listing 4.14: Most common destination query (RR dataset only)

4.6 Kepler.gl

The Kepler.gl tool has many uses and applications; for this thesis, its primary use will be to serve as the visualization platform for geographical data, such as the number of trips in a zone or locations of stops. Through the use of 2D and 3D maps, the tool allows for a much more straightforward interpretation of the data during the project analysis and in reading the report.

When utilizing a tool after importing the necessary packages, a .geojson file was used to outline the AMP with each zone border indicated; this could then be used in conjunction with data from the queries performed to create Kepler maps that provided a valuable visual representation, as seen in 4.2

```
#importing and treating a set of data
df_1 = pd.read_table("../Data/Results/trips_by_zone_per_month/rod_2019_May.txt",
    delimiter="\t")
df_1['Zone'] = df_1['Zone'].str.lower()
df_1.rename(columns = {'Zone': 'code'}, inplace= True)

porto = geopandas.read_file('../Data/andante_geo_2019.geojson')
porto = porto.merge(df_1, on='code')

map_1 = KeplerGl(height = 700, config = porto_config_1)
map_1.add_data(data=porto, name='porto') #merging the dataset with the geographical
     polygons

map_1.save_to_html(file_name='maps/rod/05_2019_2D.html') #export the map
```

Listing 4.15: Utilizing Kepler.gl and data from May 2019 to create a map

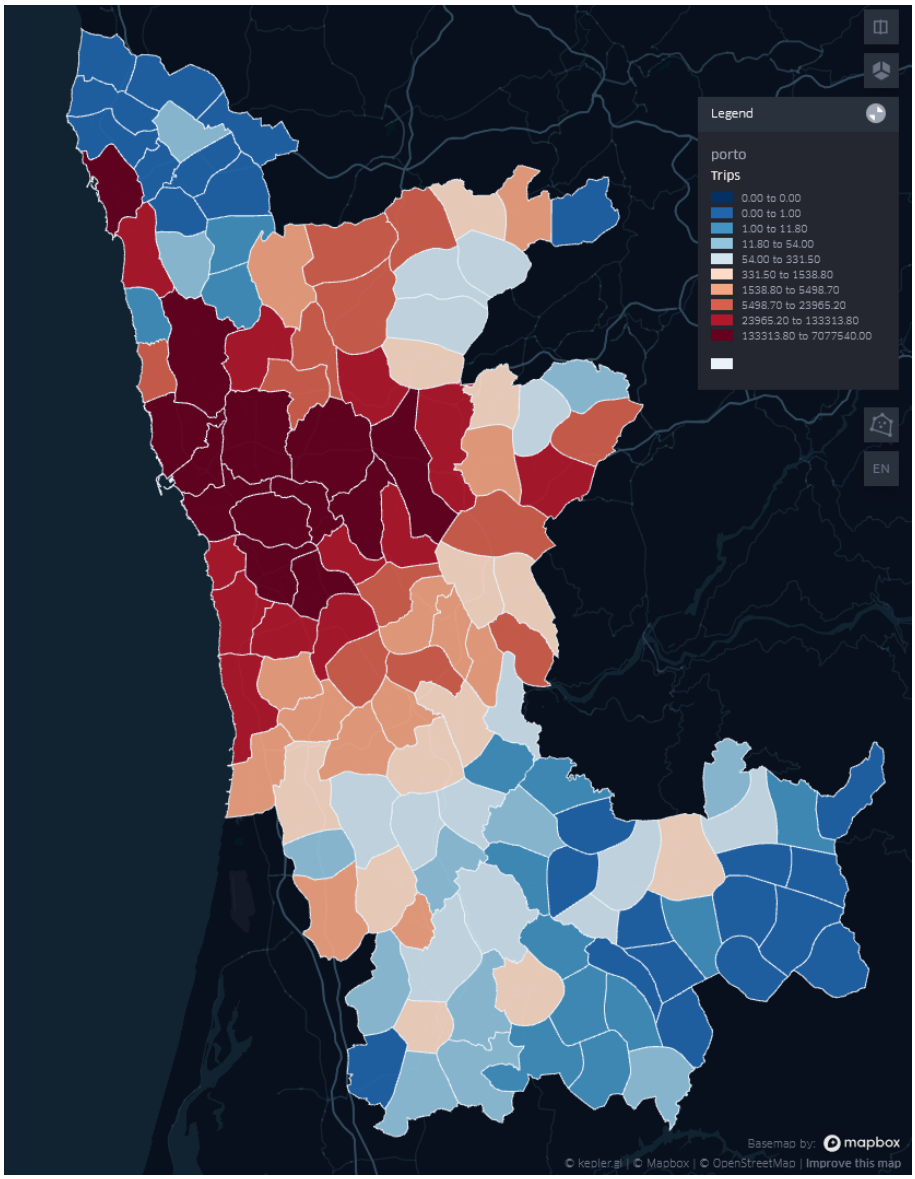

Figure 4.2: Map generated from the results of the code present in listing 4.15

4.7 Identifying Stops in the datasets

Creating a way to identify the stops of entry from the data tables we possess, through filtering repeated entries in an auxiliary material table and cross-matching with other tables provided, lists of all the stops for STCP busses, Metro, and CP trains were created. These tables also included thanks to cross-matching with other tables, the geographic coordinates of the stops; these were then used to obtain the number of trips started at each one of these points and placed them in Kepler maps to be analyzed

4.7.1 Filtering the auxiliary table

Originally the auxiliary materials provided had several problems that stopped us from achieving individual and geographical recognition of stops from the ROD entries. The table of STCP stops contained multiple lines for each stop that depended on when changes were done to the stop previously; for example, the S. Roque stop had had entries from 2009 and multiple other years, and small pieces of information changed in each of these entries, so we needed to guarantee that we used the most recent one. Another problem was that not all entries for the same stop were not together in the original file.

To tackle both problems, the following script was used; seeing as the highest individually numbered entry was around 6200, a list was used with 7000 open spots to record the most chronologically recent entry for each number, then the entire file was filtered at each line placing the entry into the list on the index of the id number within the entry. This resulted in a finished file with only the most recent entry for each STCP stop and their assigned code that could be matched to the data in the AFC entries.

```
numbers = ['None'] * 7000
count = 0
while True:
    if count == 0:
        line = file1.readline()
        file2.write(line)
        count += 1
        line = file1.readline()
    else:
        line_1 = line
        id_1 = line_1.split('\t')[0]
        numbers[int(id_1)] = line
        line = file1.readline()
        if not line:
            for i in numbers:
                if i != 'None':
                    file2.write(i)
                    count += 1
            break
```

Listing 4.16: Filtering the original auxiliary file to obtain the most recent entry for each stop

After obtaining a file with only one entry for each stop, we could address the second problem mentioned in Section 3.3. The original file did not have latitude and longitude values attributed to each line. To solve this, the new file created was merged with the file that does not have the identification codes to create a new file that possesses both identification codes and geographical indicators. Some columns that provided non-valuable information were also removed.

```
df_paragens_sae_stcp = pd.read_table("../Data/Tabelas de apoio/Paragens_SAE_stcp -
    redux.txt", delimiter="\t", encoding="LATIN9")
df_stops_amp = pd.read_csv("../Data/Tabelas de apoio/stopsAMP.csv")
del df_paragens_sae_stcp["dia"]
...
df_paragens_sae_stcp.rename(columns = {'SAE_id_paragem_stcp': 'Code'}, inplace=
    True)
df_stops = df_stops_amp.merge(df_paragens_sae_stcp, on='Code')
df_stops = df_stops[df_stops["Name"].str.contains("TESTE")==False]
del df_stops['IsActive']
...
df_stops.to_csv('../Data/Tabelas de apoio/stops_final.csv',index=False)
```

Listing 4.17: Creating the final STCP stops file through editing merging the two incomplete ones

4.7.2 Cross referencing of auxiliary data with AFC entries

Having obtained the proper final file, the next step was using the file to identify and count the number of entries for each stop per month. This operation was done in memory in a Jupyter notebook environment instead of in the database as the number of queries would be so large that it would create a massive bottleneck of time, and as this was one of the last practical tasks done in the thesis time was crucial.

To avoid a RAM overload, the data files were loaded into memory only one at a time, taking each other's place when loaded, through the use of a list of paths and importing the file at each path in the list inside a for loop, and then performing the operations, finishing by exporting the results to a file for later use.

```
for p in paths:
    df_1 = pd.read_table(path.format(p), delimiter="\t", header=None)
    df_1.columns = ['Code', 'Card', 'SN_Validador','Veiculo','Provider_Code','
        Discount','Tariff','Linha_Interna','Variante','Sentido','Stop','
        Data_Hora_Validacao','Zone']
    del df_1['SN_Validador']
    ...
    df_1 = df_1[df_1['Stop'].notna()]

    df_counts = pd.DataFrame(df_1['Stop'].value_counts())
    df_counts['Stop_nr'] = df_counts.index
    df_counts.rename(columns = {'Stop': 'Trips'}, inplace= True)
    df_stop_counts = df_stops.merge(df_counts, on='Stop_nr')
    df_stop_counts.sort_values(['Trips'], inplace=True, ascending = False)
    df_stop_counts.reset_index(drop=True,inplace = True)
    ...
    df_stop_counts.to_csv(export_path.format(export_paths[i]),index=False)
```

Listing 4.18: Cross-referencing the ROD AFC entry files with the final STCP stops file to obtain total trips per month for each stop

For the RR datasets, the process was more straightforward than for ROD. The counting of trips per stop was directly done off of the dataset since previously, when filtering the data, an identifier had been placed in all CP stops in the form of "-CP" added to their name at the end. By filtering the entries that had this in their station columns, the entries in CP stops for the month were obtained; then, by filtering out the NULL values from the remaining entries, all that would be left were the Metro station entries. On those two subsets of entries, the number of occurrences of each name was counted and stored in a file.

The RR dataset did, however, present a problem ROD did not, since there were two special cases of May and November 2019 in which the data for the whole month was in two separate files and, as such, could not be handled in the for loop a garbage collector had to be used to clear the data from memory before handling each special case to avoid errors due to lack of volatile memory to store the datasets.

```
for p in paths:
    df_1 = pd.read_table(path.format(p), delimiter="\t",header=None)
    del df_1['Code']
    ...
    df_1 = df_1[df_1['Entry_Station'].notna()]
    df_cp_trips = df_1[df_1['Entry_Station'].str.contains('-CP')]
    df_metro_trips = df_1[df_1["Entry_Station"].str.contains("-CP")==False]

    df_cp_counts = pd.DataFrame(df_cp_trips['Entry_Station'].value_counts())
    df_cp_counts['Stop'] = df_cp_counts.index
    df_cp_counts.reset_index(drop=True,inplace = True)
    df_cp_counts.rename(columns = {'Entry_Station': 'Trips'}, inplace= True)
    df_cp_counts.to_csv(export_path.format(export_paths[i] + "_cp"),index=False)

    df_metro_counts = pd.DataFrame(df_metro_trips['Entry_Station'].value_counts())
    df_metro_counts['Stop'] = df_metro_counts.index
    df_metro_counts.reset_index(drop=True,inplace = True)
    df_metro_counts.rename(columns = {'Entry_Station': 'Trips'}, inplace= True)
    df_metro_counts.to_csv(export_path.format(export_paths[i] + "_metro"),index=
        False)
    ...
del [[df_1,df_cp_trips,df_metro_trips,df_cp_counts,df_metro_counts]]
gc.collect()
```

Listing 4.19: Separating the AFC entries for CP and Metro stops (RR dataset)

With the number of entries per stop and their names already being provided, since the RR stops had no identification codes, only names, we could use the file previously used to provide the

geographical coordinates to the STCP file to obtain the complete geographical information for all Metro stops and CP train stations with no need for extra cross-referencing.; to do this, the two data files were merged on the station names, and columns of no value were removed, originating two final files for each month, one for the entries of CP stops and one for Metro stops.

```
df_stops__cp_local = pd.read_csv("../Data/Tabelas de apoio/stops_cp.csv",header=
    None)
df_stops__metro_local = pd.read_csv("../Data/Tabelas de apoio/stops_metro.csv",
    header=None)
del df_stops__cp_local['Provider']
...
del df_stops__metro_local['Provider']
...
for p in paths_2:
    df_cp = pd.read_csv(import_cp.format(p))
    df_metro = pd.read_csv(import_metro.format(p))
    ...
    list_metro = df_metro['Stop'].values.tolist()
    list_cp = df_cp['Stop'].values.tolist()
    for met in list_metro:
        if len(df_stops__metro_local[df_stops__metro_local['Stop'].str.contains(met
            .split('-')[0].strip())]) == 1:
            stop_1 = df_stops__metro_local[df_stops__metro_local['Stop'].str.
                contains(met.split('-')[0].strip())]
            df_metro.loc[i,"Latitude"]= stop_1["Latitude"].iloc[0]
            df_metro.loc[i,"Longitude"]= stop_1["Longitude"].iloc[0]
        ...
    df_metro.to_csv(import_metro.format(p),index=False)
    for cb in list_cp:
        if len(df_stops__cp_local[df_stops__cp_local['Stop'].str.contains(cb.split(
            '-')[0].strip())]) == 1:
            stop_2 = df_stops__cp_local[df_stops__cp_local['Stop'].str.contains(cb.
                split('-')[0].strip())]
            df_cp.loc[i,"Latitude"]= stop_2["Latitude"].iloc[0]
            df_cp.loc[i,"Longitude"]= stop_2["Longitude"].iloc[0]
        ...
    df_cp.to_csv(import_cp.format(p),index=False)
```

Listing 4.20: Adding geographical coordinates to the files counting the number of entries for Metro and CP

4.8 Google Places

In this thesis, the Google Places data we had was from a previous dissertation project, and this was due to Google places changing its policy regarding free access to its API for students, so we

utilized data previously extracted by a FEUP student for a dissertation. There is a slight risk of some entries not being accurate in the year 2022; however, they would have been more accurate for the year 2020 since the data is from the 2020-2021 year.

In the previous work of João Gonçalves, sub-sets were created from the primary dataset by filtering the types[16]. To work with the visual representations of the data, we utilize Kepler maps and introduce the different sub-sets into the AMP outline.

```
df_education = pd.read_csv(path.format("educationlist"))
...
porto = geopandas.read_file('../Data/andante_geo_2019.geojson')
map_1 = KeplerGl(height = 700)
map_1.add_data(data=df_education, name='education')
...
map_1.add_data(data=porto, name='porto')
map_1.save_to_html(file_name=export_path.format(p_rod))
```

Listing 4.21: Importing the Google Places sub-sets created by the previous work

4.9 Chapter Summary

In this chapter, we begin by introducing the problem the thesis aims to address. After, we present the approach and its structure before going into more detail on each part. We then explain in detail the approach taken for data storage, problems encountered, and solutions implemented. Afterward, we discuss the data filters created and the specific details of each filter depending on its purpose and the filtering process. Following, we introduce each of the main queries and dive into the implementation details of each of them, from the logic behind them to their importance and implementation code. After, we introduce the Kepler.gl tool and its capabilities. Then, we explain the process needed to identify stops and stations due to the data provided. Finally, we discuss the Google Places data uploads to Kepler.gl maps.

Chapter 5

Analysis

5.1 Overall **54**

5.2 Zones **57**

5.2.1 The importance of PRT1 57

5.2.2 Behaviour observed 58

5.3 Days of the Month **60**

5.3.1 March 2020 61

5.3.2 May 2020 61

5.3.3 July & August 62

5.3.4 September 2020 65

5.3.5 December 2020 66

5.4 Days of the week **67**

5.4.1 Total Entries per weekday 71

5.5 Hours of the day **71**

5.6 Operators **73**

5.7 Discounts **77**

5.7.1 Grouping similar discounts for redundancy and overall analysis of alternatives 77

5.7.2 Discount Percentage 80

5.7.3 Ages of users 83

5.7.4 Estimated income 85

5.7.5 Ages and income estimation 86

5.8 Most common destination zone **89**

5.9 Stops & Google Places Data **92**

5.9.1 Metro stops 92

5.9.2 CP stops 93
5.9.3 STCP stops 94
5.9.4 Google Places hot spots analysis 95
5.10 Chapter Summary 105

This chapter presents the analysis performed on the data, relevant tables, graphs, and maps created and explains some parts of the data interpretation. The first Section 5.1 provides an overall analysis of the data for the two years, looking at total trip numbers. Following, Section 5.2 presents the data analysis for the query "trips by zone per month." Section 5.3 explains the analysis of the results produced by the query "trips by day of the month." Section 5.4 presents the analysis of the different behaviors observed in the data for each day of the week. Section 5.5 introduces the analysis of the results from the "trips by hour" query. In Section 5.6 the analysis of the results related to the operators is presented. Section 5.7 explores the analysis of the results from the "trips by discount" query. Following, Section 5.8 presents the cumulative results of the "most common destination" query. Finally, Section 5.9 presents the analysis of the bus, metro, and train grids and then analyzes some relevant geographical locations using the combination of the grids and Google Places data.

5.1 Overall

To perform an overall analysis of the progression of the recovery during the two-year time period, a graph plotting a line of the total trips for each month was created. Looking at Figure 5.1 it is apparent that even in standard years such as 2019, there are high and low points for the use of public transport by the population. In 2019 after a seemingly promising start to the year numbers-wise, with the advent of the pandemic, usage plummeted to 0 during April and started to recuperate slowly, never reaching the numbers of 2019 for the rest of 2020.

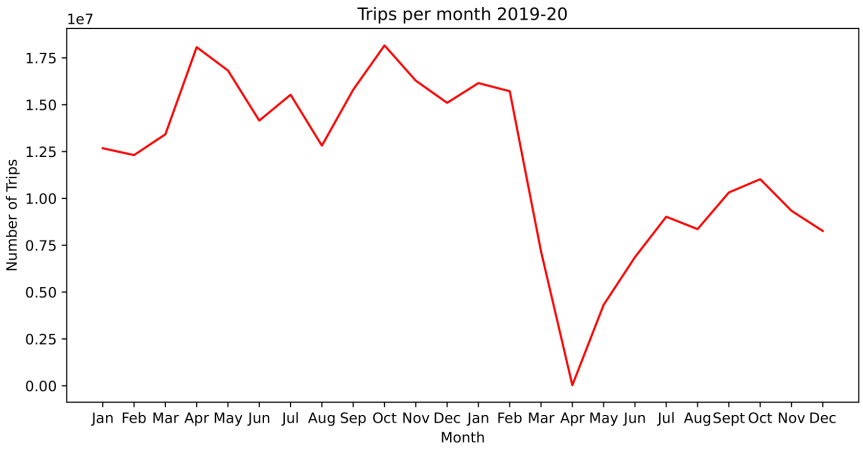

Figure 5.1: Plot of the number of trips each month 2019-20

The fluctuation of trip totals during 2020 is represented in more detail in Figure 5.2 and Figure 5.3, each containing half of the year.

Observing the totals shown for the first months of the year shows substantial growth compared with the previous year. An initial dip in total trips is felt in February; however, this is consistent with the previous year's behavior. However, from February to March, a substantial decrease is registered; this was due to the start of lockdown in the middle of March. Initially, March was on track to have an increase from February's 15.7 million trip validations.

The month of April is an outlier whose data is mainly ignored for all operations of the project; its contextual significance is relevant but not its entries. The 99.6% decrease from March and the 14279% increase to May are not relevant hence the displaying of the value of 39.7% on the image, the reduction in total trip entries from March to May.

Starting with May, the recovery process is initiated, with May registering under a third of the number of trips from February, the last whole month of trip validations. However, from May to June, a significant increase of 58.8% was registered; this would, however, be the most significant single increase in trip numbers.

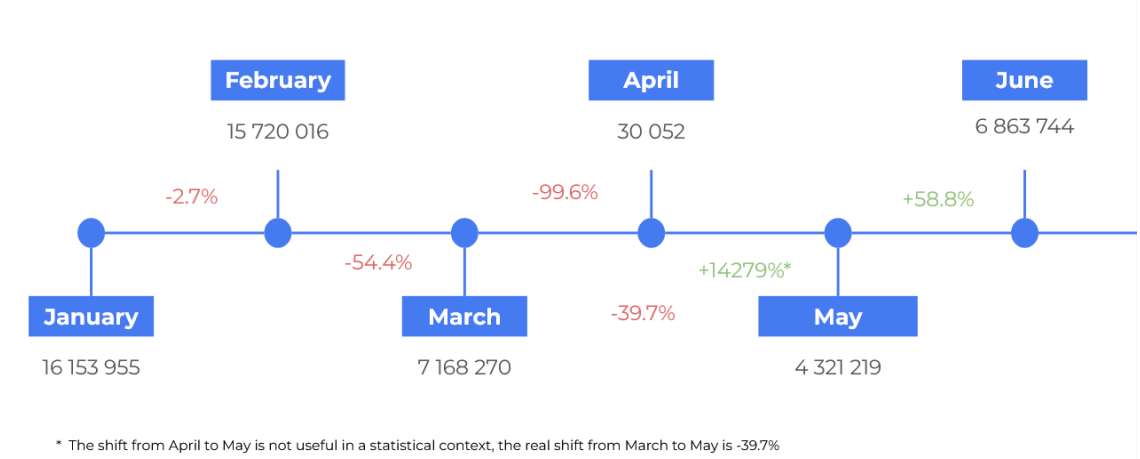

Figure 5.2: Diagram showing the percentage shifts in total trips per month for the first half of the year 2020

After the massive increase to June, proceeding to July, the recovery slowed down to a total increase of 31.4% total trips, still a considerable increase but significantly lower than the previous.

From July to August, the first decrease in trip numbers since the end of lockdown was registered, a small decrease of only 7.3% but a decrease nonetheless. After the small setback of August, the number of trips increased again by 23.4%, which would indicate that the August decrease wasn't caused by the slowing of the recovery but by extenuating circumstances.

September to October presents the last increase in trip numbers for 2020, the smallest one so far of only 6.4%. After October, the number of cases in the country started to rise, and during November, the public safety measures started to return and led to a 15.3% decrease in trip totals, followed by an 11.5% decrease in December due to the changes in states in the country at the end of November in combination with a large number of public holidays being present in the month of December.

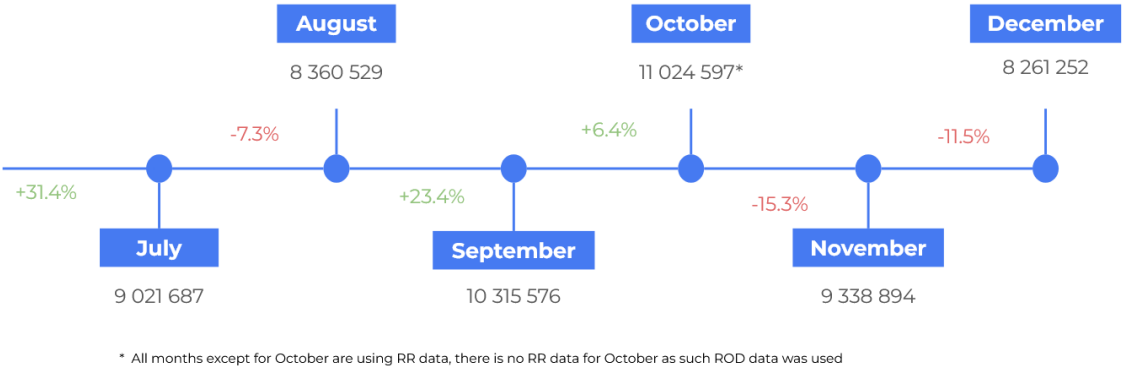

Figure 5.3: Diagram showing the percentage shifts in total trips per month for the second half of the year 2020

By examining Figure 5.4 some things of note can be detected. Firstly as previously mentioned, assuming the pattern present in the year 2019 is a normal trip fluctuation pattern for a regular year, the year 2020 was on track to reach higher totals than 2019. Secondly, the pattern of fluctuation present in 2019 is also reflected in 2020 after the month of May. In particular, by looking at the months from the second half of the year, a more homogenized version of the same pattern seems to manifest, which also supports the statement that the August decrease was not related to the COVID-19 pandemic.

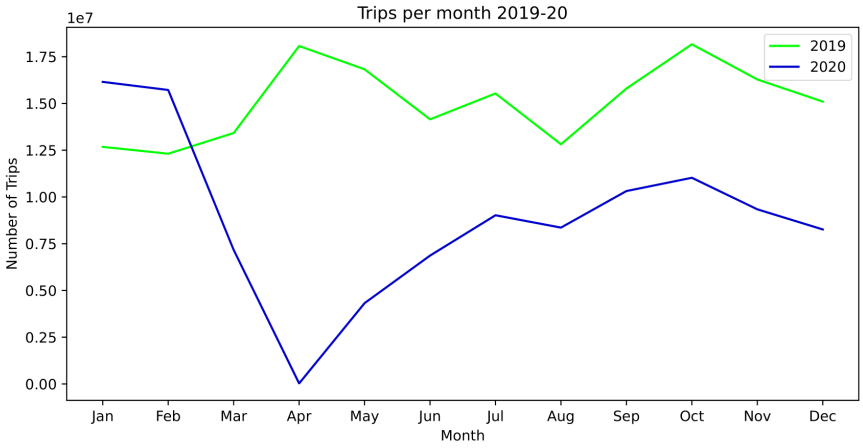

Figure 5.4: Comparative plot of the number of trips for each month for 2019 and 2020

5.2 Zones

5.2.1 The importance of PRT1

PRT1 is the zone that encapsulates central Porto and by far the most statistically significant zone in the whole AMP; in each month, roughly around 40% of all trips are started in the PRT1 zone.

Looking at Figure B.1, which contains data from January 2020, we can see a centralized area around downtown Porto with a high density of trips; however, in Figure 5.5, a 3D map representation, we can see how much PRT1 separates itself from the rest to the point it somewhat skews the intervals for the coloring on the 2D maps.

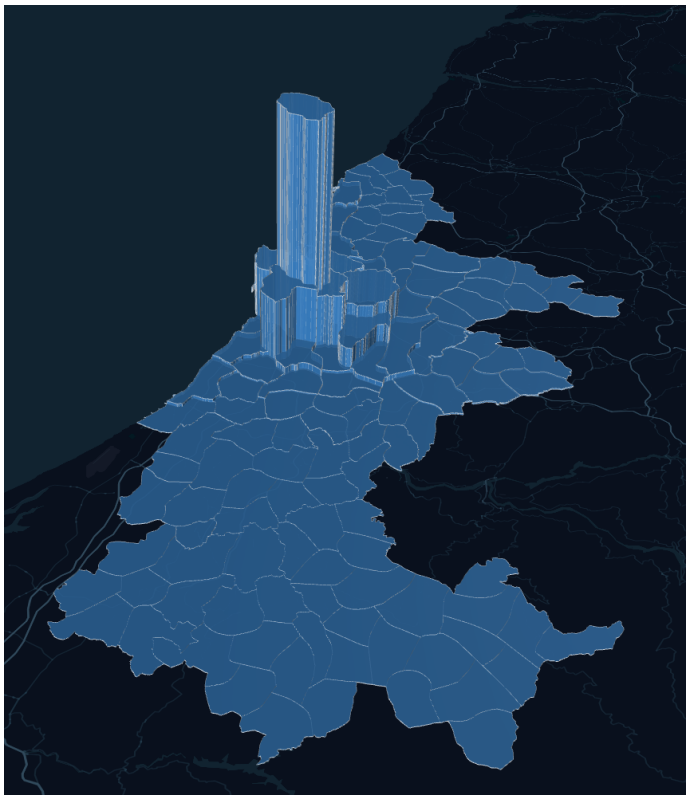

Figure 5.5: 3D Map Representing the total number of trips per zone for January 2020

This is something that was expected, and it was known that downtown Porto attracts a large amount of traffic in public transport, much more than in any other zone.

Part of it is the fact it includes the largest outlier of all stops, the Trindade metro station. Alone this station has over 1 million trip validations per month before COVID, making it the biggest stop of any kind, and the geographical positioning of Trindade, as well as being the only stop in which all lines of the Metro meet, creates this perfect storm.

Besides Trindade, PRT1 also has the S. Bento and Campanhã train stations and bus stops that attract a lot of traffic, such as the ones at Campo 24 de Agosto or Bolhão.

When discussing the zone data on multiple occasions, PRT1 will be used as a measuring reference because generally, a decrease in PRT1's percentage of total trips means the majority of other smaller zones saw increases.

5.2.2 Behaviour observed

Due to the expansion of the Andante system, all the major zones from before lost some percentage points. Most of the percentages for the biggest zones are down significantly at the start of 2020, PRT1 the biggest zone is down three whole percentage points; in January 2019, it was the origin of 44.2% of all trips, but in January 2020, it had decreased to 41.1%. Similar decreases happened to PRT2 from 9.5% to 8,3%, PRT3 from 9.0% to 8.2%. The exception to the rule was VNG1,

which increased from 10.8% to 11.6%, all the while most smaller and new zones have increased in percentage, even if, in the grand scheme of things, the values are very small. Together they removed equity from the larger zones.

The first two months of 2020, similarly to the first two months of 2019, saw small changes in percentages for each zone, but these were of not much relevance.

The month of March was the month that saw the start of the lockdown, so for half the month, there were no entries. The month was on track to result in PRT1 having a percentage of the total trips below 40%, and smaller zones growing.

When looking at the period of May 2020, a decrease in percentage for PRT1 can be seen when compared to February and January. This could be expected as the population is expected to still be adverse to travel for most recreational reasons, and PRT1 containing the downtown region of the city of Porto would be considerably affected by this

From the months ranging from May to September, slowly, the peripheral zones of the AMP that in multiple months registered no trips began to recover, and their entry numbers began to grow. Looking at the map represented in Figure 5.6 we can see the changes in percentages between the two years for each zone.

Figure 5.6: 3D Map Representing the percentage of variations in total trips between 2019 and 2020 for each zone

By the end of the year, public safety measures started to be implemented again, and in the last months of the year, however, the behavior did not change much regarding zone use.

Looking at Figures 5.7 and 5.8, we can see that PRT1 is much more closely related to the final graphical representation of the fluctuation of total trip number during the years in Figure 5.1, "PV_VC" was chosen as an example of a medium sized zone.

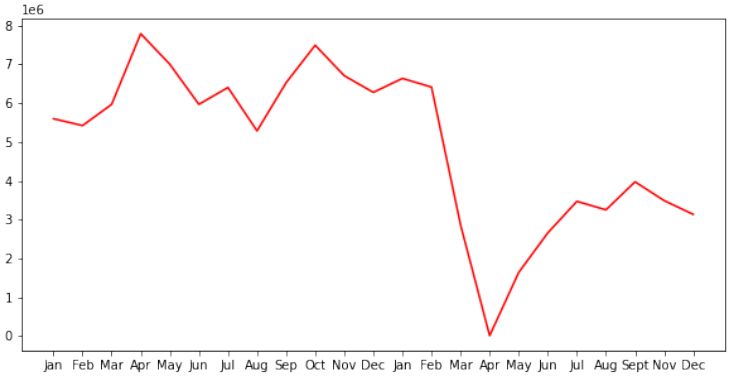

Figure 5.7: Plot showing the total number of trips for the PRT1 zone per month for 2019 and 2020

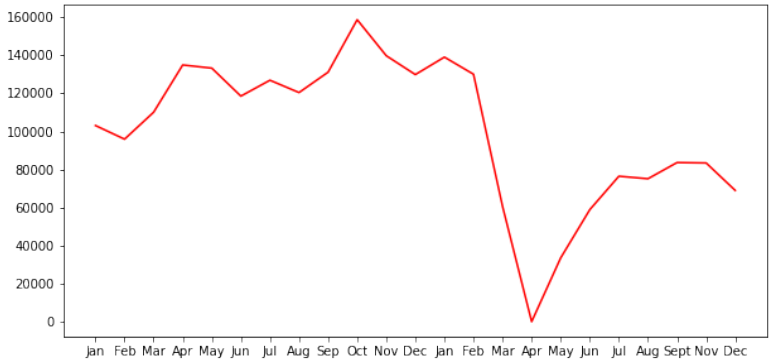

Figure 5.8: Plot showing the total number of trips for the PV_VC zone per month for 2019 and 2020

In the Annex B of this document, there is an image of a map for each of the months of 2020 except April, color-coded to indicate trip density.

5.3 Days of the Month

The results produced by the query are extensive, having at least two graphical representations for each month for both datasets. A select few found to be of more relevance will be presented in this section.

5.3.1 March 2020

Although being one of the last European nations to start having COVID-19 cases, Portugal quickly transitioned from detecting the first cases inside the borders to a nationwide lockdown, and all this was in the month of March 2020.

In Figure 5.9 we can see that the detection of the first cases did not have an impact on usage for the days immediately after the news broke. This is followed by a drastic decrease in usage starting on Tuesday, the 10th of March 2020; this was when the lockdown was announced for the next week.

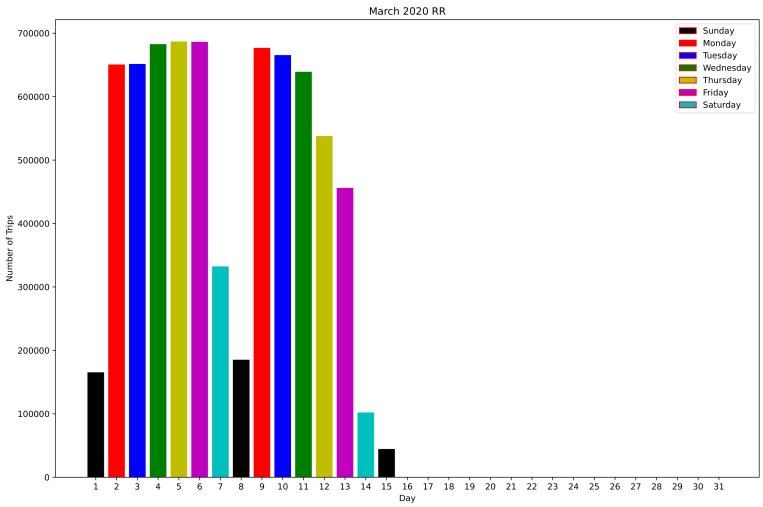

Figure 5.9: Bar chart of trips by day of the month for March 2020

5.3.2 May 2020

In May 2020, the country-wide lockdown was lifted, and the recuperation of the public transport service started. Observing Figure 5.10 the pattern of growth is clear throughout the month, seemingly indicating a steady overall recovery.

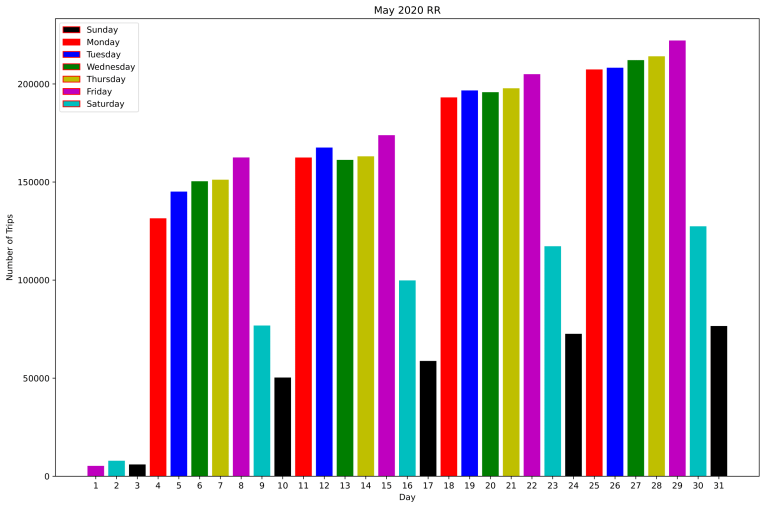

Figure 5.10: Bar chart of trips by day of the month for May 2020

Similar results were seen for the month of June to the point where presenting the graph would be redundant. The growth remained steady, excluding public holidays throughout the two months.

5.3.3 July & August

By the month of July, the recovery continued; nonetheless, it was at a much slower rate than in May and June. The increase is still seen throughout the days of the month in Figure 5.11, the number of trips per day was clearly increasing.

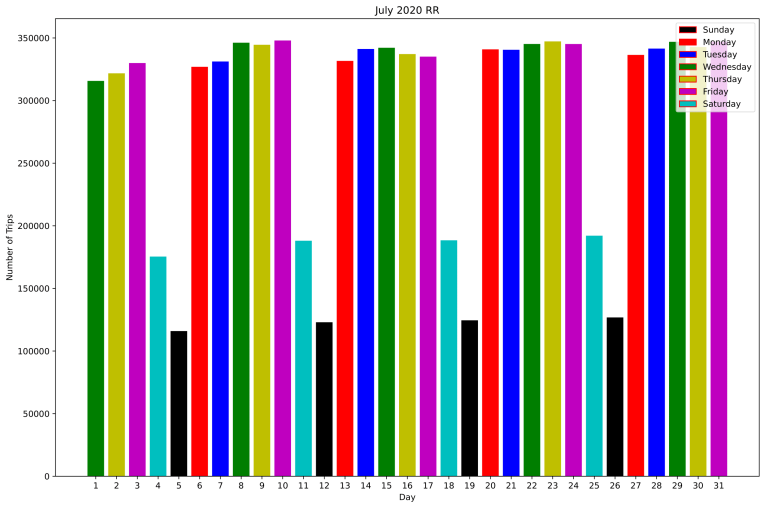

Figure 5.11: Bar chart of trips by day of the month for July 2020

When the month of August started, it continued to show growth in line with the growth from July, as seen in Figure 5.12. However, this growth seemed to halt during the second week of the month, and during the third and fourth weeks, a decrease was registered.

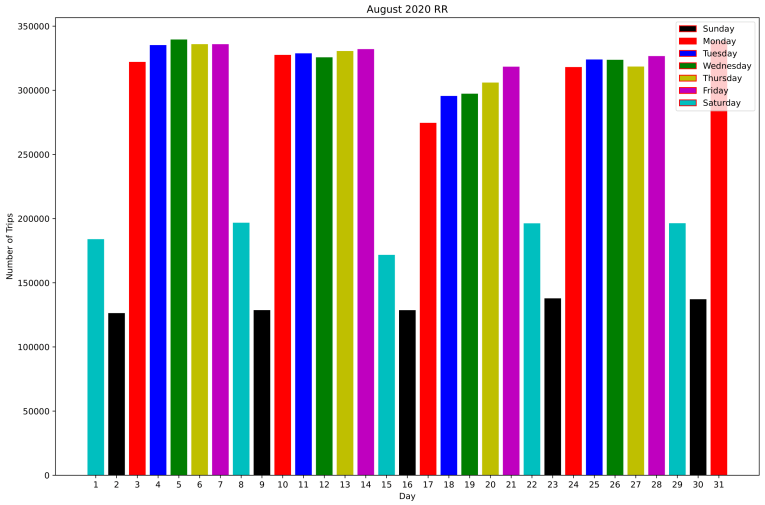

Figure 5.12: Bar chart of trips by day of the month for August 2020

These decreases seemed strange as the number of COVID-19 cases had remained similar to the previous month, and the public safety measures were not increased. For more context, a look was

taken into the months of July and August 2019 to see if a decrease in this time period is present in a year outside of a pandemic context.

Looking at Figure 5.13 the start of a decrease can be seen in after the second week of July; as expected, the total numbers of trips are much larger, but relative to the context, a decrease presents itself sooner than in 2020.

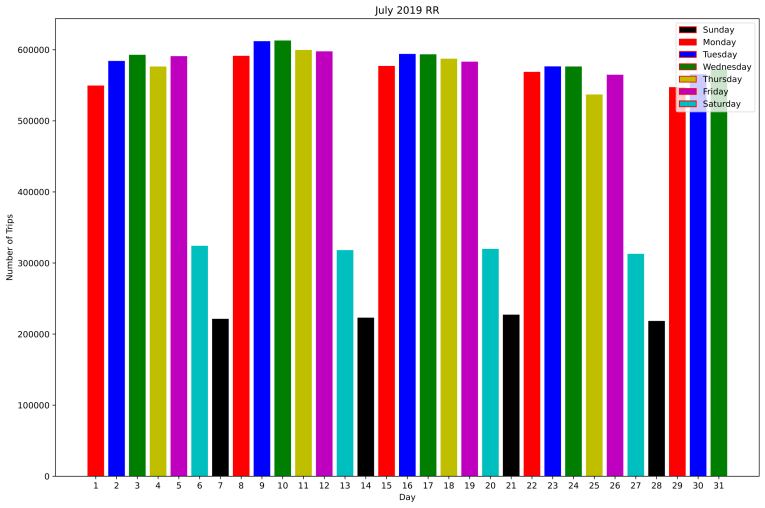

Figure 5.13: Bar chart of trips by day of the month for July 2019

Analyzing the data for August 2019 in Figure 5.14 the continuation of the decrease from July 2019 is clear, reaching the lowest point in trips per day on the third and fourth weeks of August. Although the decrease started sooner in July, the lowest point seems to be consistent across the years.

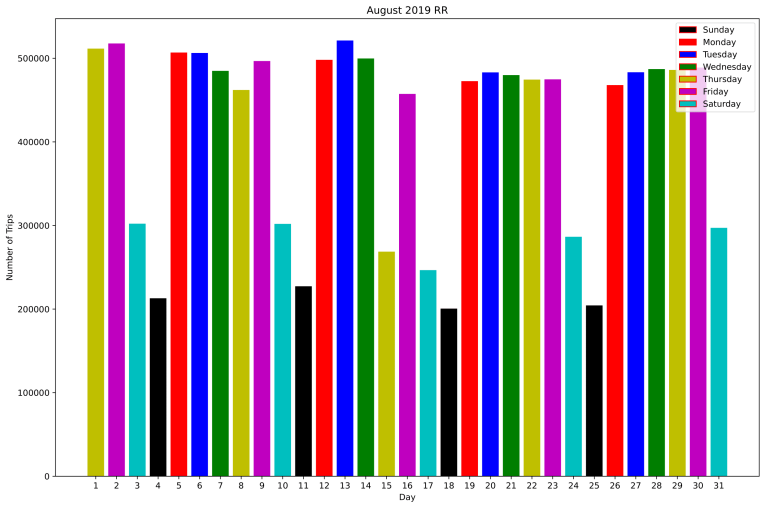

Figure 5.14: Bar chart of trips by day of the month for August 2019

5.3.4 September 2020

In September, in-person schooling started partially for universities and fully for public schools, as mentioned in Table 1.1; this created the expectation of an increase in total trips, as students present a considerable percentage of public transport users.

Looking at the data presented in Figure 5.15, the expected result can be seen, the week after the restarting of in-person classes, the number of total trips per day saw a considerable jump, and the last days of the month indicated a continued increase to come.

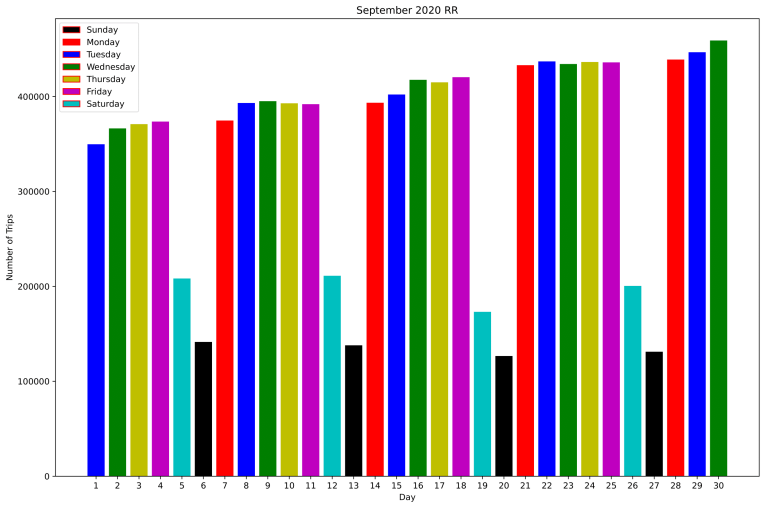

Figure 5.15: Bar chart of trips by day of the month for September 2020

5.3.5 December 2020

By the end of the year, the number of cases had increased, a second outbreak was starting, and the government was slowly increasing public safety measures again starting in November.

By looking at Figure 5.16 a considerable decrease in trips per day can be identified on the weekdays at the start and at the end of the month.

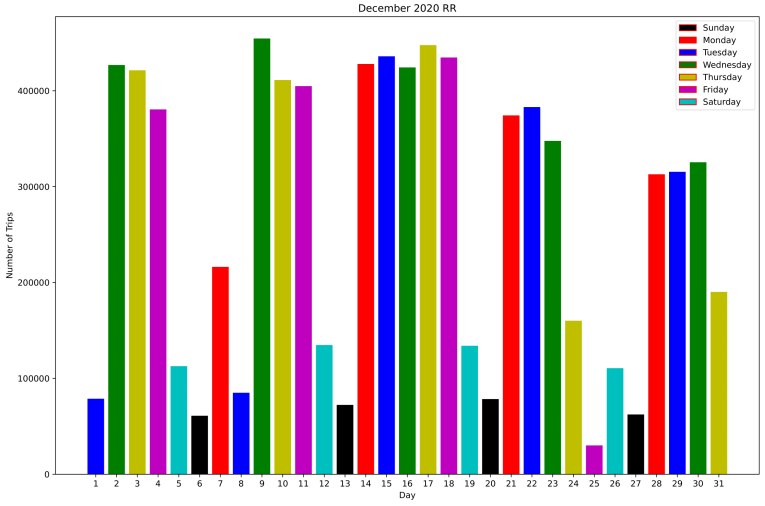

Figure 5.16: Bar chart of trips by day of the month for December 2020

5.4 Days of the week

As seen in the previous section, the different days of the week have different levels of trips associated with them; in particular, weekends generally present much lower trip numbers when compared to weekdays.

To assess how the distribution of trips per weekdays progressed over the course of the two-year period, data was split on a per year basis and also in the form of the first three months of each year as well as the last eight.

In Figure 5.17 are the percentages for the distribution of trips per weekday for the complete year of 2019. Immediately what was previously observed is corroborated with Saturdays and Sundays having much lower percentages than all weekdays, and these having small variations amongst themselves.

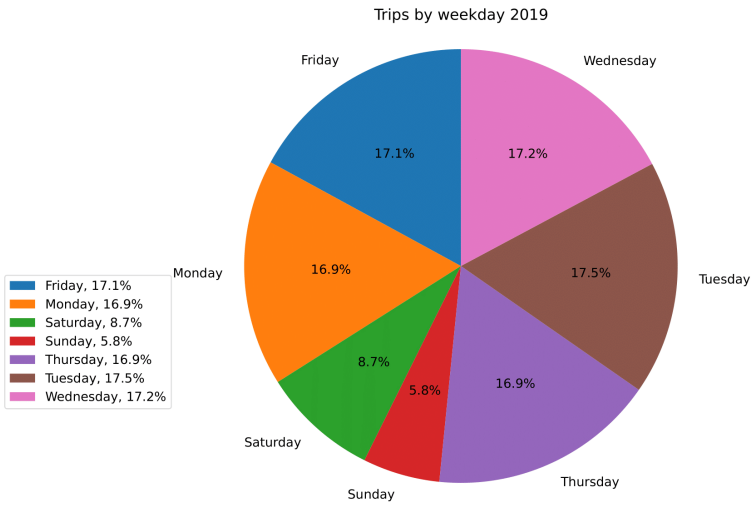

Figure 5.17: Pie-chart of trips by weekday for 2019

Observing Figure 5.18 with the data for the year 2020, and the even larger disparity between weekdays and weekends is observed, a considerably decrease, particularly on Saturdays, is of note, as well as the decrease in values for Tuesdays, these were accompanied by large increases for Thursdays, Wednesdays, and Fridays.

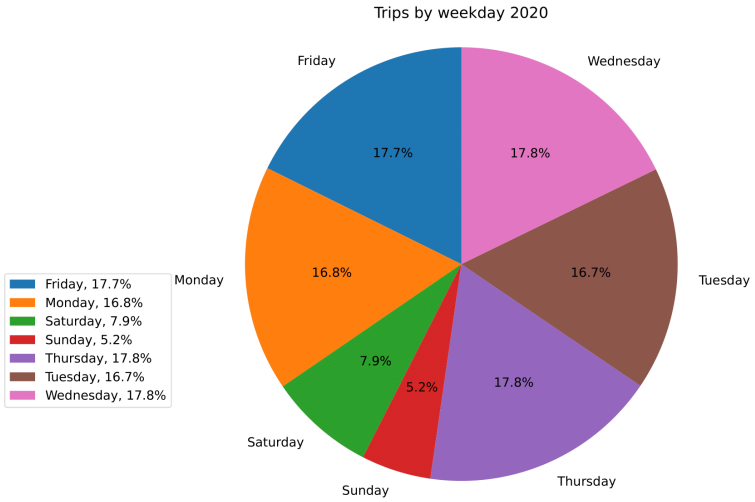

Figure 5.18: Pie-chart of trips by weekday for 2020

Taking a more focused look, by looking at smaller intervals of the two years, in Figure 5.19 the data for the first three months of 2019 is presented.

At the start of 2019, there was a disparity among weekdays, with Thursdays, Wednesdays, and Fridays having a more significant gap between them and Mondays and Tuesdays than in the previously examined graphs.

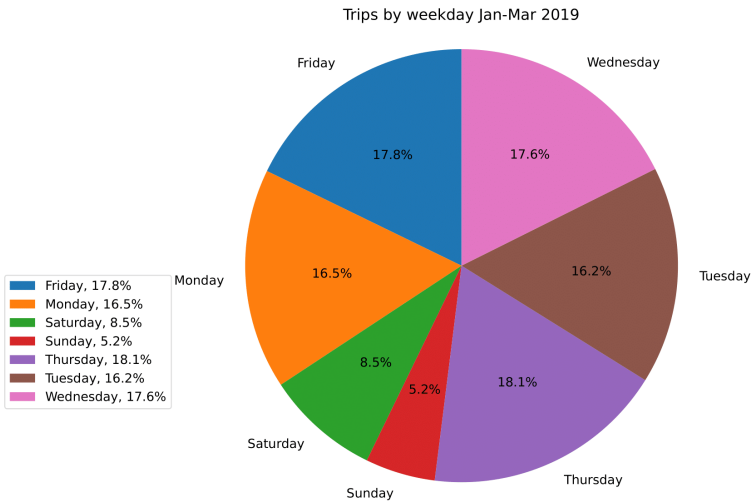

Figure 5.19: Pie-chart of trips by weekday for the 3 month period of Jan-Mar 2019

When examining the data from the latter half of 2019 in Figure 5.20 the gap previously identified in the first three months is seen to have narrowed considerably; this was expected, as the conjunction of these two datasets, plus the month of April, generated Figure 5.17.

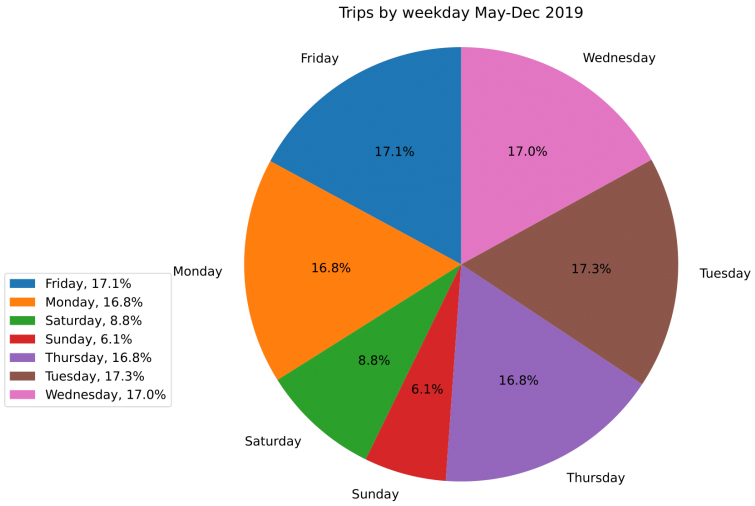

Figure 5.20: Pie-chart of trips by weekday for the 8 month period of May-Dec 2019

Moving to the start of the year 2020, in Figure 5.21, notable shifts are identified. Firstly Friday emerges above Thursday as the highest percentage for the first time with an increase to 18.3%, the previously mentioned gap between the three last weekdays and the first two grew once more. Saturday decreased in percentage considerably, and Sunday returned to the numbers of the start of 2019.

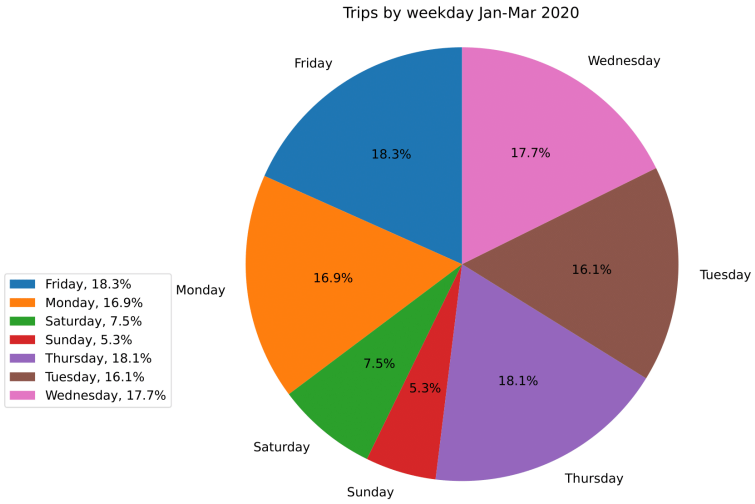

Figure 5.21: Pie-chart of trips by weekday for the 3 month period of Jan-Mar 2020

In Figure 5.22 the data for the months of 2020 after the lockdown period can be seen. The gap that grew at the start of the year was reduced once more with decreases in the percentages of the last three days of the week and increases in the first two. Saturday also saw an increase in the numbers of late 2019, and Sunday saw minor decreases.

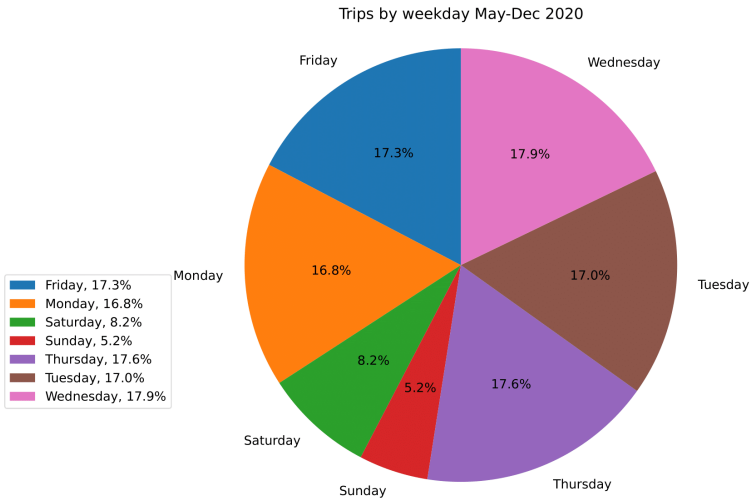

Figure 5.22: Pie-chart of trips by weekday for the 8 month period of May-Dec 2020

5.4.1 Total Entries per weekday

After looking at the percentages values in detail, we looked at the recovery rate for each weekday concerning total trip numbers.

Focusing on Figure 5.23, it is clear that at no point did any day of the week have a recovery to the average level before the pandemic; however, some points at which each day reached pre-pandemic numbers can be seen, although these mainly coincide with the lowest pre-pandemic results.

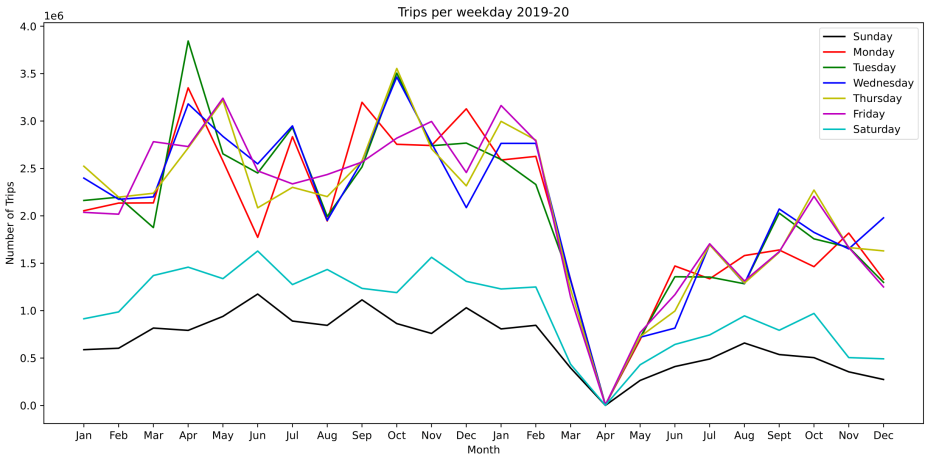

Figure 5.23: Progression of entries per month for each day of the week over the two-year period

The graph shows that the values for Sundays reach their post-lockdown peak in August 2020, and at this time, they coincide with the numbers from February 2019. For Monday, the post lockdown peak was in November and coincided with numbers from June 2019. Tuesday and Wednesday saw their peaks in September 2020, and these numbers were on the same level as August and March 2019 for Tuesday, and December and August 2019, for Wednesday. Thursday and Friday peaked in October 2020, and their peaks coincided with the numbers from June and August 2019 for Thursday and February 2019 for Friday. Finally, Saturday peaked in August and October 2020, reaching the values of February 2019.

5.5 Hours of the day

The distribution of trips during the day is vital for service management and load distribution. Using the data obtained, visual representations of these distributions were created. Analyzing these graphs can allow us to understand the impacts COVID-19 recovery had.

Observing Figure 5.24, in which the total number of trips for each hour per month of 2019 is represented, a clear trend can be identified. The data shows a minimal number of trips in the early hours; as the morning starts, the number of trips starts to grow exponentially, peaking at

around 8:00, then progressively decreasing. The second increase starts to manifest close to 12:00 and 13:00, followed by another decrease. Depending on the month, this decrease changes in magnitude more than in the morning. At the end of the afternoon, the number of trips rises once more, around 17:00 to 18:00, the peak is reached, followed by a decline into the night hours.

The two more prominent peaks of the day can be observed in the morning around 8:00 and the afternoon around 17:00, with a smaller one occurring at 12:00.

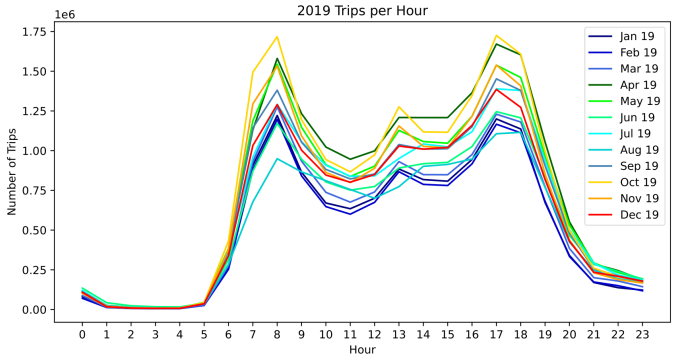

Figure 5.24: Trips per hour for each month of 2019

Having seen from the previous figure what the distribution looks like in a typical year such as 2019, we can look to Figure 5.25 to compare the data from 2020 to the expected behaviors.

Immediately when looking at Figure 5.25 the first two months of the year stand out, and their presence on the graph shows the lockdown's impact. January and February show very similar lines to the months from Figure 5.24, and to a lesser extent, so does March, although with a reduced number of total trips. The graph clearly shows that the distribution of trips across the day has been highly homogenized. The two prominent peaks from 2019 are not as far removed from the numbers during other times of day, although still clearly present.

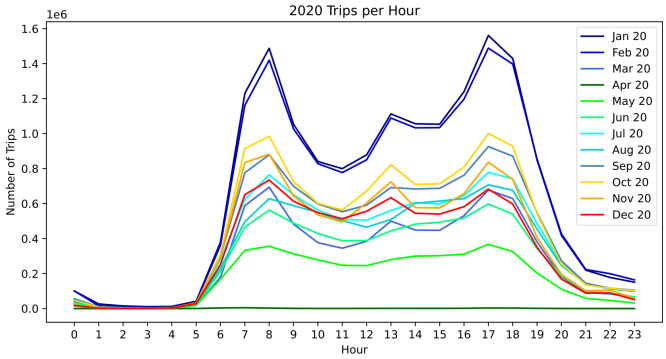

Figure 5.25: Trips per hour for each month of 2020

Finally, the impacts can be better understood by examining Figure 5.26 which shows total trips per hour for each of the years.

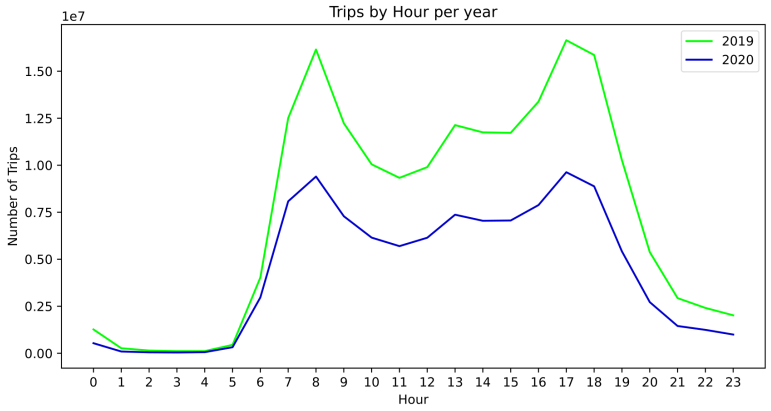

Figure 5.26: Total trips per hour for each of the years

Even with January and February being included in the 2020 line numbers, it is clear that the distribution is much less uneven and skewed towards the rush hours of 8:00 and 17:00.

5.6 Operators

By identifying the operators in the Andante system and analyzing the impacts the pandemic had on their total number of trips, areas of higher impact can be identified, and performance can be considered.

The complete list of operators present in the data is:

- **TIP** - As previously mentioned, TIP is a business group that involves Metro do Porto, CP and STCP. As all three of those operators are also present on the list, it is not clear which trips were classified with TIP's code or why
- **Metro do Porto** - Metro provider for the region of Porto the map of the full grid can be seen in Figure 1.2
- **STCP** - Largest bus operator in the AMP, STCP services the center of the AMP around the city of Porto. A diagram of the grid can be seen in Figure 1.4
- **Resende** - bus provider ensures urban transportation throughout the municipality of Matosinhos and passing through other municipalities[1]
- **CP-Porto** - The train service provider for Porto, the full grid can be seen in Figure 1.3

[1] Resende webpage - http://resendeturismo.com/en/history/ Accessed on July 2022

- **Funicular dos Guindais** - The funicular is an on-rail service that runs from Ribeira to Batalha along a hill in downtown Porto[2]
- **VALPI** - bus provider that works in the municipalities of Porto, Penafiel, Valongo, Paredes, Lousada, Amarante, and Matosinhos[3]
- **Espírito Santo** - a bus service provider that operates in Gaia[4]
- **Gondomarense** - bus provider that operates in Gondomar and Porto[5]
- **Maia Transportes** - bus provider that services multiple lines originating in the Maia municipality and provides transport to and from Porto, Matosinhos, Trofa e Valongo[6]
- **OFR** - bus provider for whom not much information was found
- **ARRIVA** - bus provider in the AMP area, it stopped operation at the end of 2021[7]
- **Pacense** - bus provider that operates in mutiple municipalities of the AMP specifically Lousada, Paredes, Penafiel, Paços de Ferreira, Maia, Valongo, Porto, Santo Tirso, Famalicão, Trofa, Vila do Conde, Povoa de Varzim[8]
- **Nogueira da Costa** - small-scale bus provider, provides service to some lines in Porto and Maia[9]
- **Landim** - bus provider in the northern region of the country[10]
- **Rodonorte** - bus provider in the north of the country, specifically the districts of Vila Real, Bragança, Viseu, Porto e Braga[11]
- **Carvalhos** - bus provider in multiple regions of the country[12]
- **Caima** - bus provider subsidiary of Transdev operating in São João da Madeira[13]
- **Transdev Norte** - Transdev sub-group that operates in the north of the country[14]

[2]Funicular dos Guindais on a travel website - https://www.introducingporto.com/funicular-dos-guindais Accessed on July 2022
[3]Valpi website about us - https://www.valpibus.pt/ Accessed on July 2022
[4]Espírito santo transportes lines map - https://www.carreiras.espiritosanto.com.pt/carreiras/mapa/ Accessed on July 2022
[5]Gondomarense(ETG) website - https://www.gondomarense.com/ Accessed 2022
[6]Maia Transportes website - https://maiatransportes.pt/linhas/ Access July 2022
[7]ARRIVA website - https://www.arriva.pt/ Accessed July 2022
[8]Pacense about us page - http://www.avpacense.pt/empre.html Accessed on July 2022
[9]Nogueira da Costa website - https://www.nogueiradacosta.pt/transporte-regular Accessed on July 2022
[10]Landim about us - http://www.avlandim.pt/empre.html Accessed on July 2022
[11]Rodonorte about us - https://www.rodonorte.pt/pt/empresa/sobre-nos/ Accessed July 2022
[12]Carvalhos group about us - https://utc.pt/about-us-2/ Accessed July 2022
[13]Transportes de Portugal page for Caima transports - https://transporteportugal.com/empresa-caima-transportes-em-sao-joao-da-madeira-2259 Accessed July 2022
[14]Transdev about us page - https://www.transdev.pt/mobility-company/ha-mais-20-anos-investir-portugal Accessed July 2022

- **Transdev Douro** - Transdev sub-group providing travel along the Douro river

As we could see in the previous list, some redundancy exists with regards to operators; some operators are sub-divisions of others, others are not clear which role they perform, and for some, the areas of operation could not be ascertained. Multiple operators were also present in the list of codes provided as auxiliary data with no actual entries and, as such, are not present in the list.

Due to April 2020 having no data of value for this analysis, the data was split into four groups, which were the data from the first three months of each year, 2019 and 2020, and then the data corresponding to the data from May to December for each of the years.

In Figure 5.27 the number of trips for each operator during the first three months of 2019 is represented in a pie chart. The graph clearly shows the two largest operators are Metro do Porto and STCP. During this time, the expansion of the Andante system had still not happened, as these values are expected to diminish after May 2019.

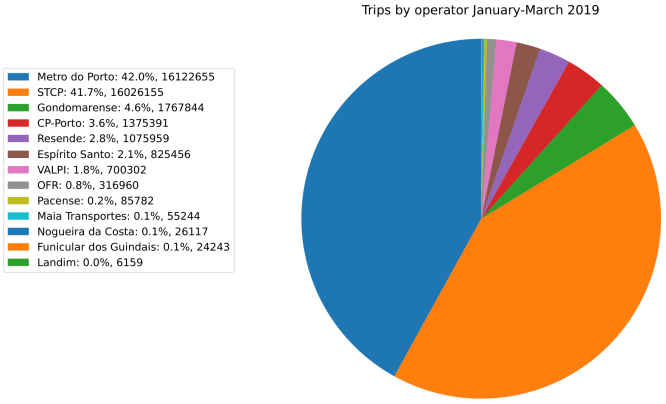

Figure 5.27: Pie-chart of trips by operator for the 3 month period of Jan-Mar 2019

Looking at Figure 5.28 as expected after the expansion of the Andante system, the total percentage of the largest operators was reduced. A considerable increase for CP-Porto was also noted, with multiple new providers being represented in the graph, most of which only represent a small number of total trips.

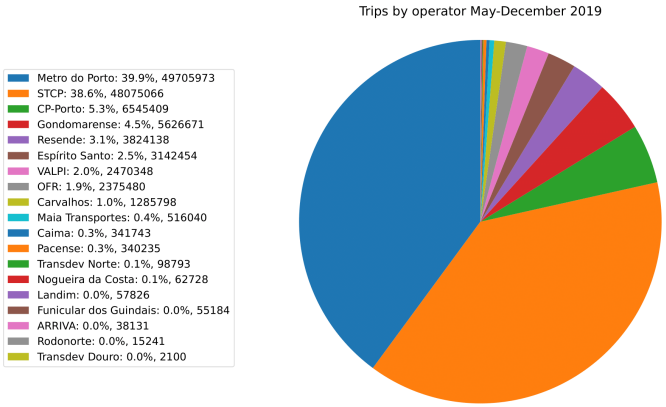

Figure 5.28: Pie-chart of trips by operator for the 8 month period of May-Dec 2019

When analyzing Figure 5.29, it is clear that between the later months of 2019 and the starting months of 2020, only minor changes were identified, the most notable being a decrease of almost 1 percentage point in the percentages of total trips for STCP.

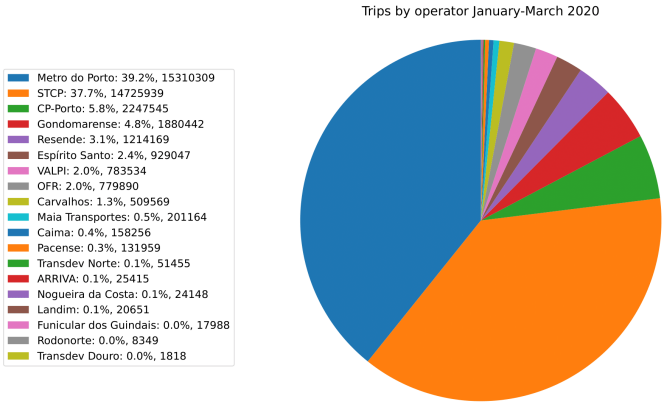

Figure 5.29: Pie-chart of trips by operator for the 3 month period of Jan-Mar 2020

Comparing the previous figures with the results present in Figure 5.30 the impacts of COVID-19 recovery on the operators can be visualized.

While most of the small bus providers were not heavily affected, Metro do Porto saw a massive impact, decreasing from 39.2% in the first three months of 2020 and 39.9% in the later eight months of 2019 to 35.8%. The primary benefactor from the post lockdown environment was STCP, increasing to close to 43%, the highest percentage in the four-time periods analyzed. The train

service provider saw a slight reduction of 0.3 percentage points, similarly to the most prominent bus providers besides STCP, which are Gondomarense, Resende, Espírito Santo, VALPI, and OFR; all of these saw slight decreases.

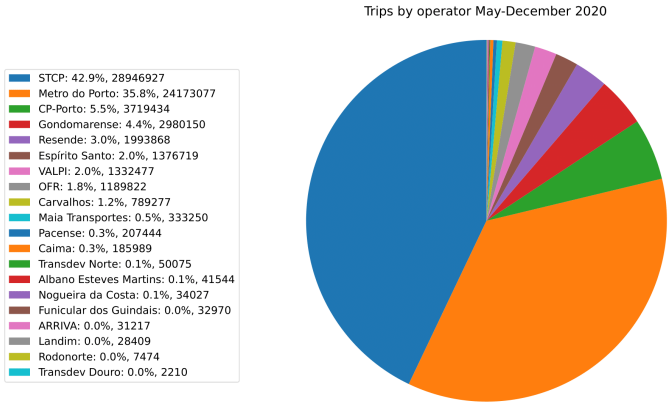

Figure 5.30: Pie-chart of trips by operator for the 8 month period of May-Dec 2020

It would appear that bus and train use were not too affected by the COVID recovery, but Metro was greatly affected and presented considerable decreases.

5.7 Discounts

Over the course of the two years, many changes happened to the Andante system, some of which were multiple pricing changes. The pricing details for the period before May 2019 were unavailable, as such data from the first four months of 2019 was ignored for this discount-based section. The changes happened in May 2019 and then in January 2020, with the first of each month being the starting point for implementing the new structures.

There are multiple relevant ways to group the discount data. There were five identified ways to group the data that we believed to be important: similar data brackets to avoid redundancy; by discount percentages; by ages of the users; by incomes estimated, and by ages and estimated incomes in conjunction.

5.7.1 Grouping similar discounts for redundancy and overall analysis of alternatives

The entries had multiple redundancies in discount alternatives, discounts that had the same meaning and should be together but initially were not. These were joined and then represented graphically; the relevant alternatives are:

- **Normal** - This type of classification is present in entries that have no discount whatsoever and indicates someone that paid full price for the subscription
- **Elderly** - This discount is attributed to elderly people through a monthly subscription and provides a 25% reduction in the price of the subscription
- **Social+ B** - Destined for people with reduced income within a predetermined bracket, this is the lowest of two possible discounts of its type and provides a 25% reduction in the subscription price
- **Social+ A** - This discount is the highest of the two Social+ discounts destined for people with an even worse financial situation than Social+ B and provides a 50% subscription price discount
- **Sub 23** - A discount destined for young adults that are pursuing higher education, this discount is only available to people between the ages of 18 and 23 that are enrolled in an institution that the ministry of education has recognized specifically to access this program, this provides a 25% discount
- **Sub 23 A** - When a student that fulfills the criteria for Sub 23 is also a benefactor from a Social+ A level of social support, this discount is available. This is the highest possible level of total subscription discount, providing a 60% total price reduction
- **Sub 13** - A child under 13 years old has, since the start of 2020, being able to get a subscription for free for the whole AMP; before 2020, the child was entitled to a discount of 25%
- **13_18** - This discount falls under the umbrella of the 4_18 discounts, destined for children between the ages of 13 to 18 who have school-provided trip assistance
- **13_18 A** - This discount falls under the 4_18 A discount. If a child falls within the criteria of the previous discount and is also a recipient of Social+ A support, then they will receive a discount of 60%, the highest possible and the same as Sub 23 A
- **4_12** - Destined for children that are given trip support from their schools and as such are not eligible for the Sub 13 free travel subscriptions, this discount belongs to the 4_18 bracket and has a monetary value of 25% total subscription reduction
- **4_12 A** - This discount falls under the 4_18 A discount and is destined for children that fit the criteria for 4_12 but are also recipients of Social+ A and, as such, are given a 60% total subscription discount
- **Student** - This discount is destined for students under the age of 25 that either no longer belong in the sub-23 or 4_18 categories due to age or due to the institution they are a part of not being recognized by the ministry of education

- **Family** - This category started appearing in the data for the later eight months of 2020 and is part of a family package promotion in which by having three or more people signed into the subscription, a reduced total price for the subscriptions is provided

Among these, 4_12 and 13_18 were sometimes joined into 4_18 when relevant.

In Figure 5.31 we can see the representation of each of the alternatives with the total number of entries each possessed for the eight months between May and December 2019 and the relative percentages of each.

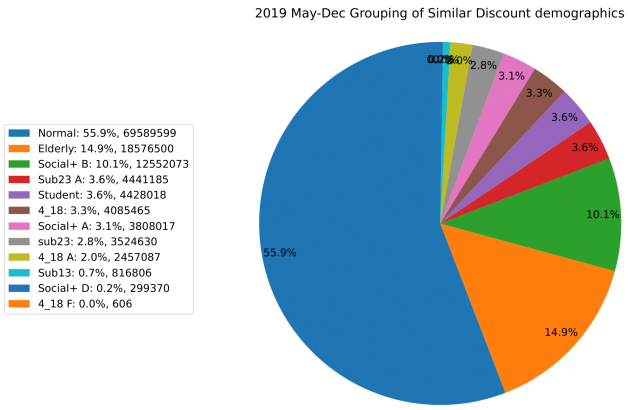

Figure 5.31: Pie chart of Similar Discount demographics May-Dec 2019

Figure 5.32 then shows the totals for the first three months of 2020 before including March, in which the lockdown started. Some differences in the pricing scheme for 2020 can already be seen here. With Sub 13 being made accessible for the whole AMP instead of only on a municipal level, it has increased from a total of just over 800 thousand entries in the eight months of 2019, corresponding to 0.7% of the total trips, to a value of close to 700 thousand in just three months of 2020, corresponding to a percentage of 1.7%.

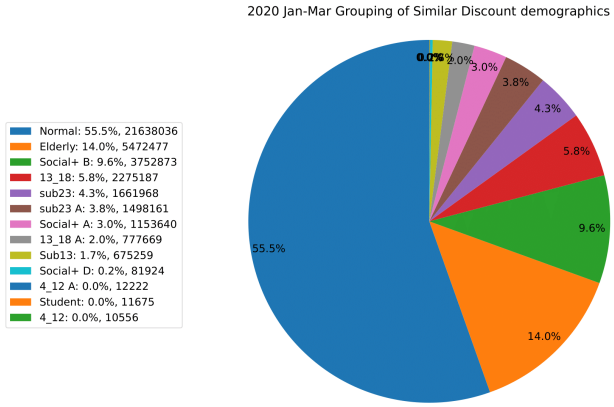

Figure 5.32: Pie chart of Similar Discount demographics Jan-Mar 2020

Since the country was on total lockdown for the whole month of April, it has been excluded from our analysis. Figure 5.33 represents the discount groups for the remaining eight months of 2020. In this immediately, multiple changes can be seen in different groups; notably, all groups including students, decrease in percentage.

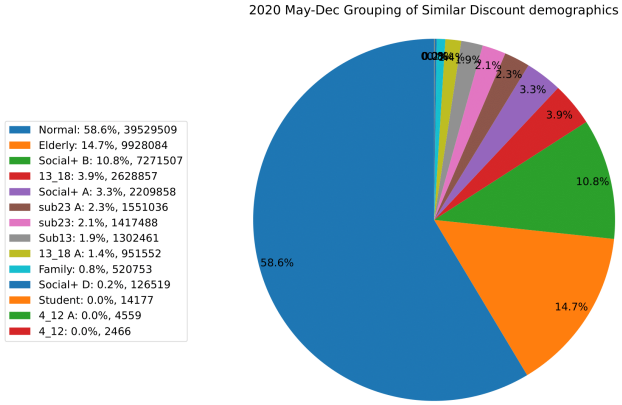

Figure 5.33: Pie chart of Similar Discount demographics May-Dec 2020

5.7.2 Discount Percentage

Following the analysis of the entries by the discount group, we thought it relevant to analyze the impacts felt by each discount bracket. Among all discount categories, there are five levels of discount:

- 0% Discount - the normal price of a monthly subscription, the entries corresponding to this bracket are all entries in the Normal category
- 25% Discount - this discount bracket covers most of the different categories presented above specifically, 4_18, Elderly, Sub 23, Social+ B, Student
- 50% Discount - this discount is for people receiving Social+ A level support
- 60% Discount - this discount is applied specifically in cases where students from the Sub 23 and 4_18 groups are also recipients of Social+ A support
- 100% Discount - Only applied to children under 13 this discount, 2020 this was only applied at a municipal level, since then it has been applied for all of the AMP

Analyzing which brackets saw the most changes could help understand the impacts of COVID-19 on the people covered by each percentage. Figure 5.34 presents the distribution of the percentages for the later eight months of 2019.

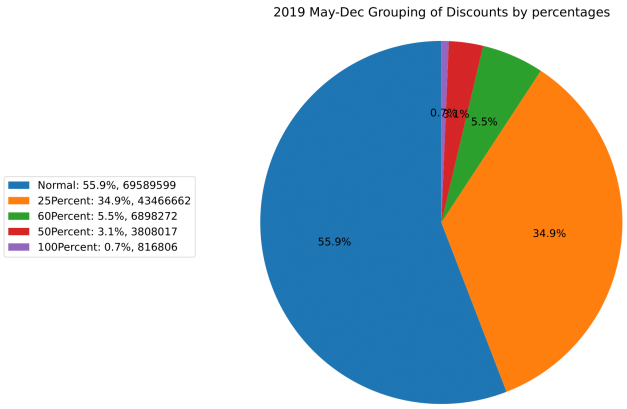

Figure 5.34: Pie chart of Discounts grouped by Percentages May-Dec 2019

When comparing Figure 5.34 to the following Figure 5.35, we do not notice much shift, besides the increase in the 100% bracket, that coincides with the increase in Sub 13 users previously detected. This was expected as no significant change happened between these two periods, only minor changes in the pricing system.

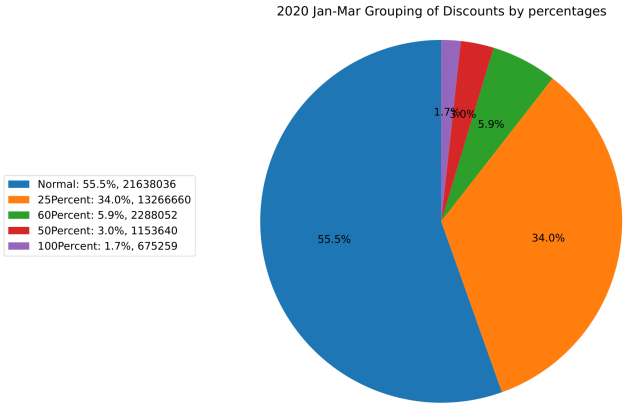

Figure 5.35: Pie chart of Discounts grouped by Percentages Jan-Mar 2020

In contrast, when analyzing Figure 5.36, apparent shifts are visible in the three most prominent brackets of ordinary, 25% and 60%, as well as the addition of the Family bracket that is a different kind of discount not easily quantifiable. These shifts are related to the impact COVID-19 had and show that the brackets of 25 and 60% were severely impacted; these ranges contain multiple alternatives that have students and elderly people of risk, which may explain the decrease due to online classes and health concerns.

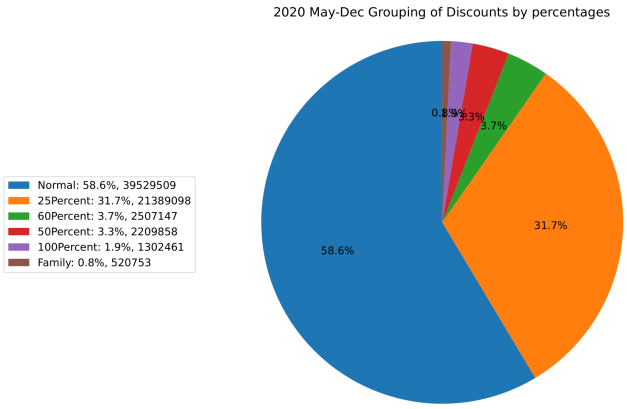

Figure 5.36: Pie chart of Discounts grouped by Percentages May-Dec 2020

5.7.3 Ages of users

With the available categories, it is possible to create groupings of entries based on the users' ages; using these groups, the impacts of the lockdown on different demographics can be analyzed.

There were three main groups identified:

- Under 25 - grouping comprised of all categories that included children or students under the age of 25
- Adults - grouping comprised of the normal category, the unemployed category, and other social categories that were not destined specifically for children or students, such as Social+ B and A
- Elderly - grouping comprised of the elderly category.

This division is not perfect. For example, a person above 65 or below 25 can still have a Normal monthly subscription or support from Social+ categories depending on their circumstances. It was, however, the best possible grouping available to analyze the data according to age.

In Figure 5.37 the overwhelming percentage of adult use of the Andante system is clear. This value fluctuates somewhat throughout the year but consistently crosses the 70% threshold in 2019.

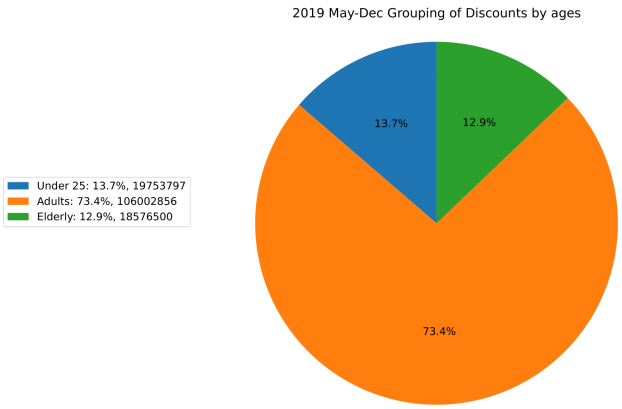

Figure 5.37: Pie chart of Discounts grouped by age of users May-Dec 2019

Looking to the first three months of 2020 in Figure 5.38 a shift can immediately be noticed in the percentages of use obtained by the Adult grouping, as well as a significant increase in the Under 25 grouping as well as a slight increase in the Elderly. The Sub 13 changes seen in previous Figures 5.31,5.32 and 5.33 were undoubtedly connected to this and the trend of decreases in the values of categories present inside the Adult grouping.

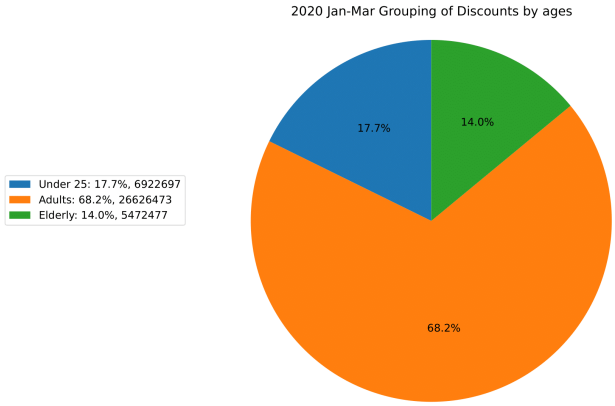

Figure 5.38: Pie chart of Discounts grouped by age of users Jan-Mar 2020

In Figure 5.39 the data indicates a massive decrease in the percentages of the Under 25 group. As expected, the lockdown and online classes significantly impacted this group, mainly using public transport for education.

What was not expected, however, was for the group containing the Elderly population not only not to decrease significantly but to increase. It would seem the pandemic had no impact on the growth of this group that had been seen in the first months of the year.

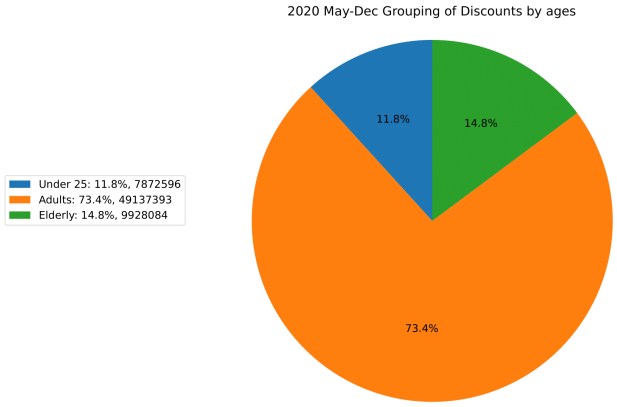

Figure 5.39: Pie chart of Discounts grouped by age of users May-Dec 2020

5.7.4 Estimated income

In order to estimate the impact COVID-19 had along income lines, we divided the data into two groups, entries from people with no social backing were classified as normal income, and entries from individuals that had received support from Social+ governmental programs were classified as low income.

Analyzing Figure 5.40 an overwhelming majority of 81.1% to 18.9% can be observed between the Normal and Low-income groups.

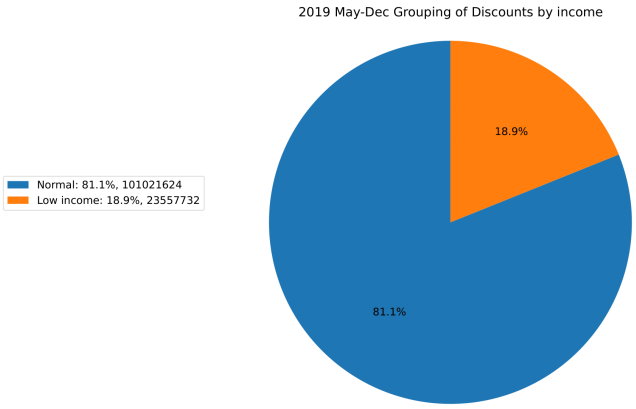

Figure 5.40: Pie chart of Discounts grouped by estimation of user income May-Dec 2019

With 2019's numbers as a base, looking at Figure 5.41, it is possible to identify a slight growth in the Normal group percentage in 2020 before the impact of COVID-19.

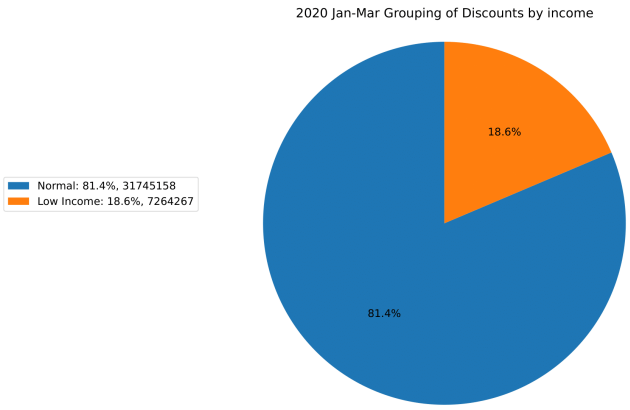

Figure 5.41: Pie chart of Discounts grouped by estimation of user income Jan-Mar 2020

When looking at Figure 5.42, an unexpected decrease in the Low-income group can be seen. Seemingly the upward trend identified between the first two periods remained constant despite the impacts of the pandemic, which, as mentioned in chapter 2, typically impacts Low-income families harder.

Factors in this decrease could be the previously mentioned decrease in student ridership, as the Social+ programs support many students.

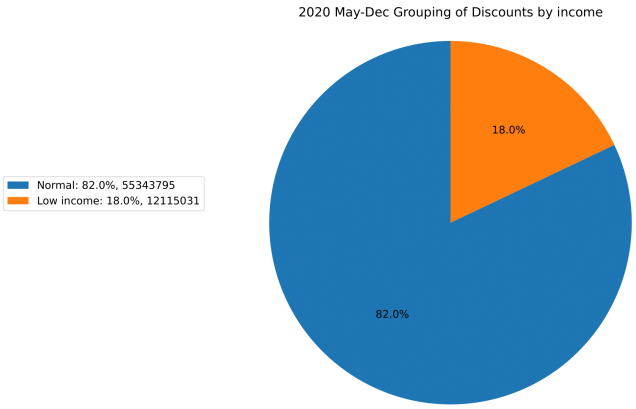

Figure 5.42: Pie chart of Discounts grouped by estimation of user income May-Dec 2020

5.7.5 Ages and income estimation

In order to analyze the impact of COVID on users of different financial brackets, we divided the discount entries into six sub-groups, with which analysis can be performed to ascertain different impacts.

The groups created were:

- Under 25 Low Income - the entries from children and students that received social support were grouped in this sub-group
- Under 25 Normal Income - this sub-group contains the remaining entries from children and students
- Adults Low Income - this sub-group was comprised of the Social+ B and A categories
- Adults Normal Income - the sub-group was comprised of only the Normal category
- Elderly Low Income - this sub-group was comprised of the entries for retirees that received social support
- Elderly Normal Income - this group contains the Elderly category of entries except for the retirees that had previously been added to it

The family data category present in the data for the latter half of 2020 was not included in any of the groups due to encompassing adults and children on many occasions.

Due to the data, it is challenging to perform these groupings accurately; as an example, Social+ A can be attributed to Elderly people through the Complemento Solidário de Idosos(CSI) governmental fund, but these are not identified in any way as such, the Elderly Low-Income group is severely under-represented, in Figure 5.43 all this is identifiable, especially the severely low number of entries for the Elderly Low-income group. The dominance of the Adults with normal income group is likely to be over-represented as well; however, due to a lack of a better alternative for creating the sub-groups, this was the closest to a real case scenario possible.

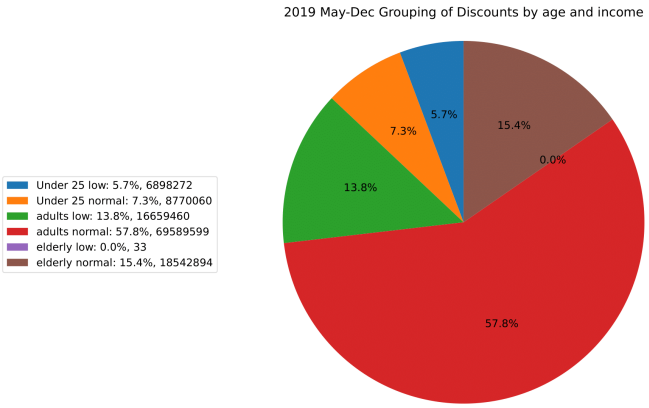

Figure 5.43: Pie chart of Discounts grouped trough a mixture of age and income brackets May-Dec 2019

In Figure 5.44 multiple differences to the 2019 data can be seen. A considerable increase in the number of Under 25 Normal income trips of over 4 percentage points, as well as a decrease in the Adult Normal income group of 2.3 percentage points and decreases of 1.4 and 1 percentage points in the Elderly Normal and Adult Low-income groups.

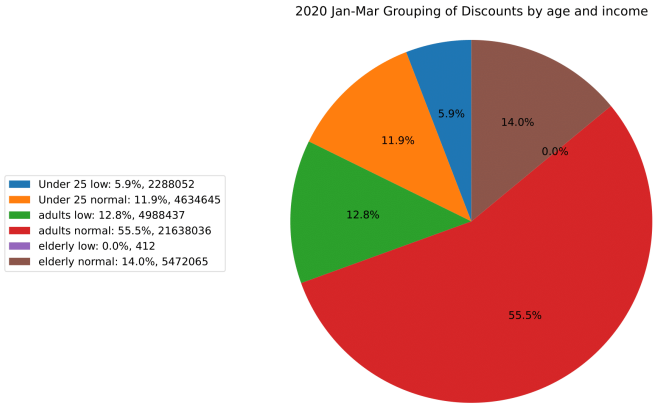

Figure 5.44: Pie chart of Discounts grouped trough a mixture of age and income brackets Jan-Mar 2020

For the months following the lockdown, significant impacts can be seen in Figure 5.45. Considerable decreases are seen in both the Low and Normal income Under 25 groups, of 2.2 and 3.9 percentage points respectively, and these are coupled with increases in both Adult income groups, the Low group saw an increase of 1.6 percentage points and the Normal group saw the most significant increase of all increasing by 3.6 percentage points lastly an increase was also seen in the Elderly Normal income group of 0.8 percentage points.

The number of Elderly Low-income entries increased through all the periods, although their relevance is diminished, as previously stated.

The evidence is consistent with the decrease in use by young people, also showing that this was common for both low and normal income groups with a more considerable impact on the normal income group.

The COVID-19 lockdown and public safety measures appear to have reversed the trends of slowly decreasing the number of adult ridership for low-income and normal-income individuals.

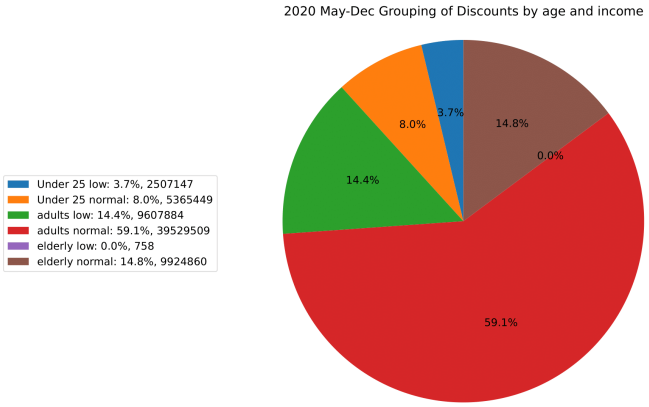

Figure 5.45: Pie chart of Discounts grouped trough a mixture of age and income brackets May-Dec 2020

5.8 Most common destination zone

By utilizing the columns in the RR entries that refer to the entry and exit zones, the most common destinations for trips originating from each zone were acquired for each month and then compiled into tables that indicated how many times each zone was the most common destination for other zones.

Table 5.1 are the numbers indicating how often each zone was the most common for the entire two-year period. As expected, the dominance of PRT1 in trip numbers is reflected in this table, with almost one hundred more times as the most common zone. The second zone, VNG1, is also within expected results as it is a zone with a high ridership level, as seen in Section 5.2. The zone with the third-highest number was somewhat surprising as SJM1 was not expected to reach these results. The fourth place, however, was more expected as most of the trips from the zones spanning the Póvoa de Varzim line of the Metro either go to Póvoa de Varzim or Porto.

Table 5.1: Times each zone was the most common destination 2019-20

Zone	Times most common
PRT1	563
VNG1	365
SJM1	266
PV_VC	143
VCB1	75
STR1	72

Continued on next page

Table 5.1 – continued from previous page

Zone	Times most common
OAZ1	71
VLG1	46
MTS1	44
MAI4	43
ARC1,MAI1	42
MAI2	33
PRD1	25
VNG8,GDM2	22
GDM1,MTS2	21
PRD5	17
ARC14,PRD2	7
VCD3,SMF1,GDM6	4
PRD3,OAZ6,VCB6, OAZ2	3
SMF6,VNG6,OAZ4	2
PRT3,STR5,VLG3, MAI5,OAZ3,OAZ8, SMF2,SMF4,STR8	1

Looking at the results from only the year of 2020 in Table 5.2 we see a strong resemblance with the total results presented previously.

Table 5.2: Times each zone was the most common destination in 2020

Zone	Times most common
PRT1	285
VNG1	193
SJM1	146
PV_VC	82
STR1	43
OAZ1	37
VCB1	35
VLG1	23
MAI4	21
MAI1,MTS1	20
ARC1	19

Continued on next page

Table 5.2 – continued from previous page

Zone	Times most common
PRD1	15
MAI2	14
VNG8	11
GDM1,GDM2,MTS2	10
PRD5	9
GDM6,ARC14,SMF1	4
PRD3,OAZ6,VCB6	3
PRD2,VLG3,OAZ3, STR5,OAZ8,VCD3, SMF2,OAZ4,VNG6, OAZ2,MAI5	1

In Table 5.3 with the results from only 2019 a larger difference from the total results in Table 5.1 is clear. We can see that after the fifth position on the table, the number of times as most common is higher compared to 2020, indicating that in 2019 there was a higher concentration of trips towards the central zones. There were also some shifts. From 2019 to 2020, many of the zones after "PV_VC" swapped somewhat, while the top four remained the same across both years.

Table 5.3: Times each zone was the most common destination in 2019

Zone	Times most common
PRT1	278
VNG1	172
SJM1	120
PV_VC	61
VCB1	40
OAZ1	34
STR1	29
MTS1	24
ARC1,VLG1	23
MAI1,MAI4	22
MAI2	19
GDM2	12
MTS2,GDM1,VNG8	11
PRD1	10
PRD5	8
PRD2	6

Continued on next page

Table 5.3 – continued from previous page

Zone	Times most common
VCD3,ARC14	3
OAZ2,SMF6	2
STR8,SMF4,OAZ4, VNG6,PRT3	1

5.9 Stops & Google Places Data

Analyzing the patterns that are apparent on the transportation grids of STCP, CP, and Metro, as well as showing the highest demand zones, is essential. Using the grids for each provider and the Google Places data, hubs for specific types of services can be identified and analyzed to see if there is a possible connection between the places and the ridership levels.

5.9.1 Metro stops

The importance of the Trindade and other downtown Porto stations mentioned before can be seen in Figure 5.46 with the heat-map data for July 2020. Besides previously determining factors, we can see Povoa de Varzim and Vila do Conde on the top left of the map have a significantly larger number of trips than the stations that follow them, which leads to the thought that there is a large number of people from those regions that use the metro and train services to go into teh center of the AMP and back while rarely stopping at stations in between. Specifically, a higher level of congestion is also seen in the Matosinhos line in the summer months due to the recreational beach-going.

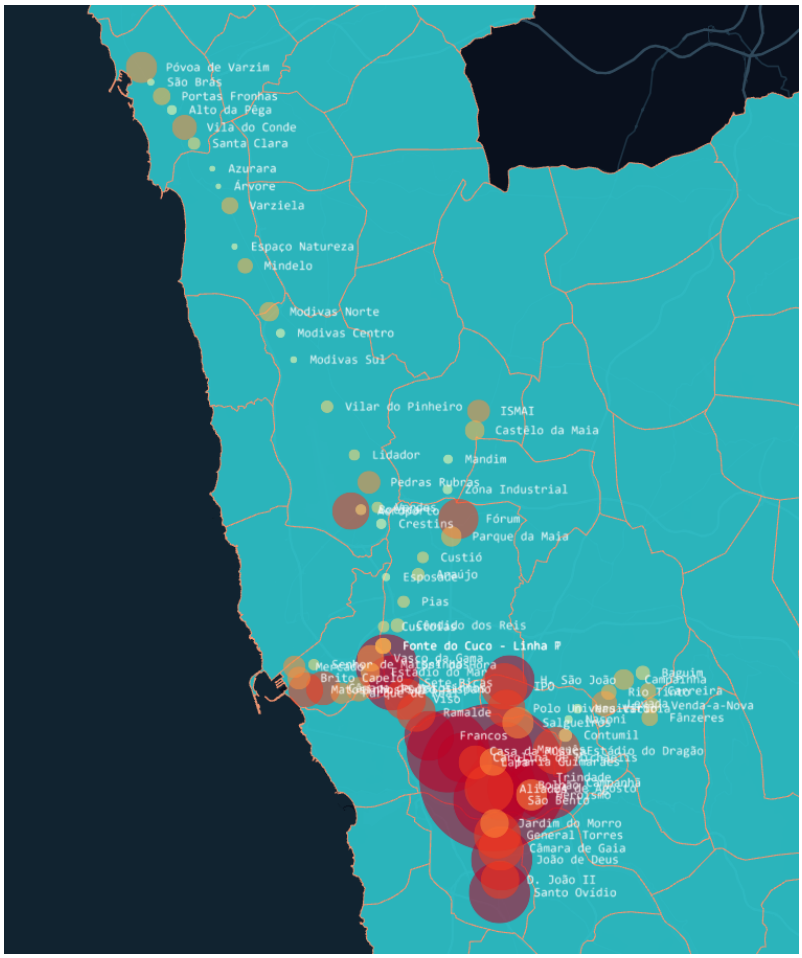

Figure 5.46: Metro stops heat-map for July 2020

5.9.2 CP stops

When looking at a train stop in Figure 5.48, it is crucial to remember that these graphs are scaled individually across transport types. The number of trips for the metro services far outweighs the number of trips by train, even tho the circles for S. Bento and Campanha seem larger than Trindade in Figure 5.46.

Espinho presents a similar situation to what Póvoa de Varzim and Vila do Conde did previously. While Metro has the Trindade station outlier, the train service appears to have three stations with considerably more traffic than all others, including S. Bento Campanhã and Ermesinde.

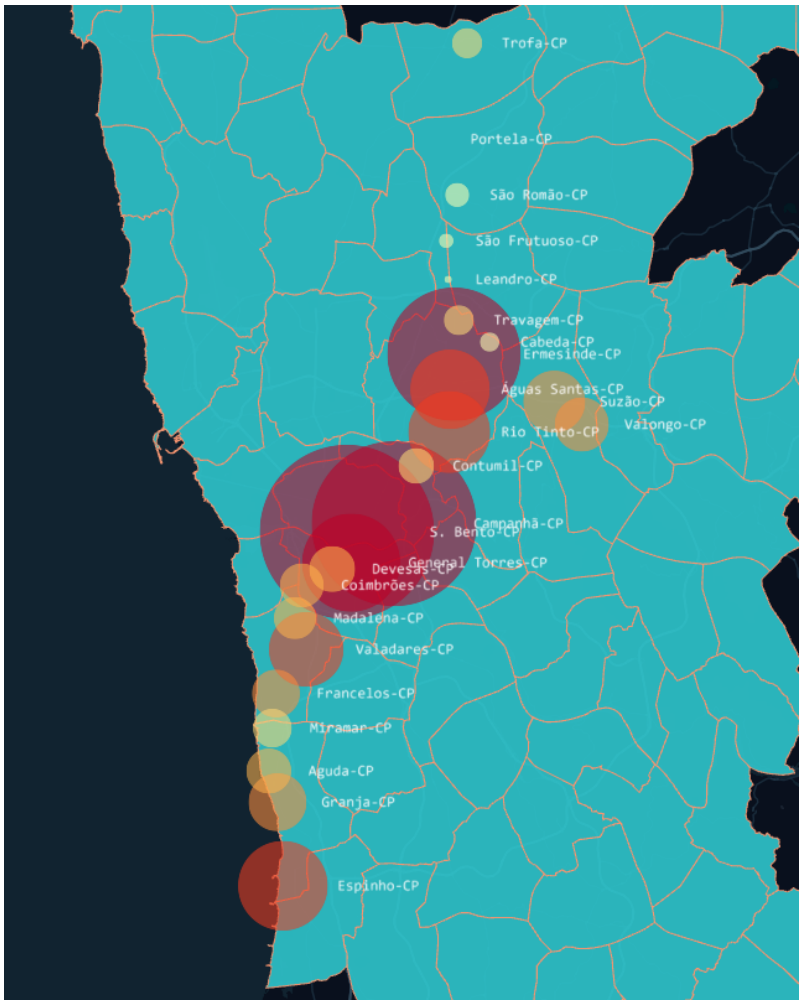

Figure 5.47: CP stations heat-map for July 2020

Ridership is highly concentrated on the Aveiro line that leads to Espinho, and inside the upwards suburbs of Porto around Ermesinde, the rest of the stations further up north have little to no travel validations.

5.9.3 STCP stops

As previously stated, unfortunately, we can only calculate the numbers for the stops from STCP services; this covers a relatively small area of the AMP but the area with the most consistent traffic.

In contrast with metro and train, we can observe, in Figure 5.48, a pattern of routes to downtown Porto with high concentrations, but where most stops have similar numbers of trips, there is not a substantially larger stop at the end or the start of a route.

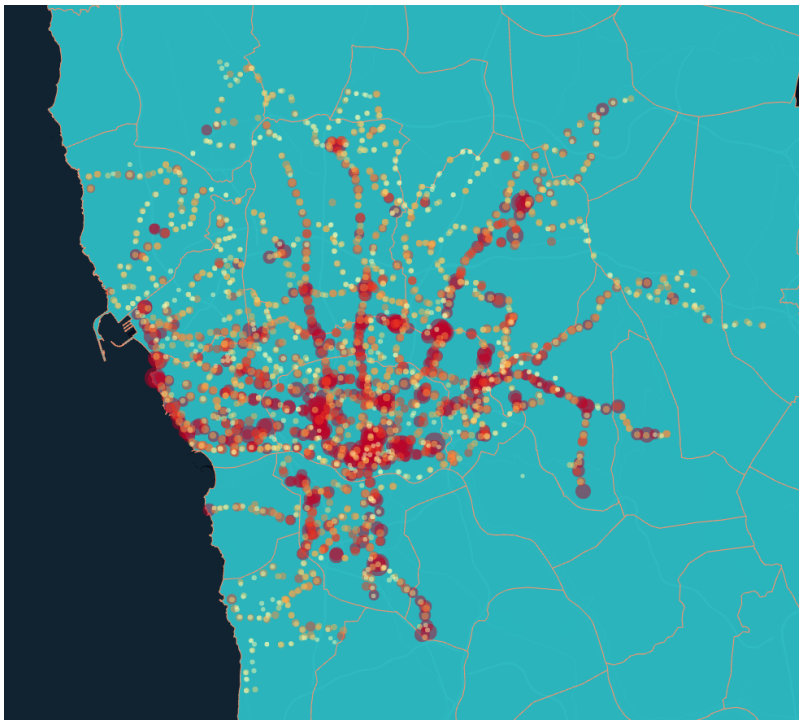

Figure 5.48: STCP stops heat-map for July 2020

We also see that the routes leading to downtown Porto have a considerably larger number of users than those connecting to other locations. Forming what appears to be an outwards spreading web from PRT1 through the main lines to the suburbs.

5.9.4 Google Places hot spots analysis

In this subsection, a look at some regions of particular relevance in the Porto area; due to time constraints, an in-depth analysis of all the regions that warranted it was not possible, so the focus was directed at the ones deemed more exciting and relevant. In Figure 5.49 we can see all the Places in the data we had, color-coded by typing: Education in purple; Food in blue; Government in red; Health in dark green; Leisure in yellow; Lifestyle in orange; Lodging in light green; and services in pink.

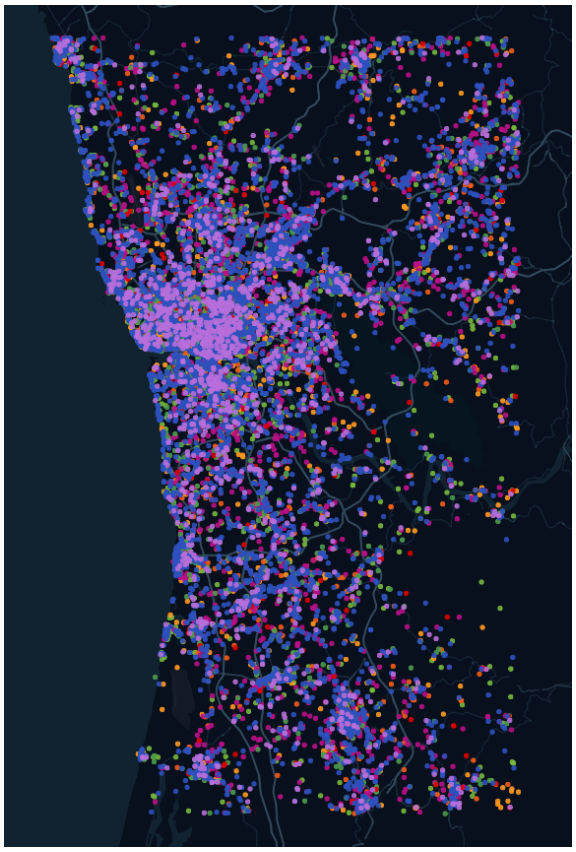

Figure 5.49: Google Places Data

Health services were affected by the pandemic in a particular way, and many people had trouble reaching public health services due to the number of people sick with COVID-19. To visualize some of the impacts of COVID on Hospital S. João, we utilized the data from Metro and STCP as well as places data. This representation needs to be analyzed with an understanding of the caveat that the area around the Hospital contains multiple universities, gyms, and restaurants; as such, the impacts are not only relevant to the use of health services, but an isolated example is non-existent.

Periods found to be important for the behaviors around the Hospital were: the month of May 2020, immediately after the lifting of the lockdown, to analyze what the base number of trips were; July 2020, in this period, some restrictions had already been lifted, and the initial spikes in recovery had happened, it was deemed relevant to analyze the ridership levels before the start of the school year in September but also avoid the dip in total ridership of August; September 2020 was the start of some in-person activities in universities, and was used to compare the numbers of 2020 to those of 2019 for the start of the school year.

In conjunction with the area around the Hospital, the area around the southern part of the University campus was also analyzed for the same periods.

In the period immediately after the lockdown was lifted, May 2020, we can see in Figure 5.50, that trips to the main bus(yellow) and metro(red) stops for the Hospital Area were significant.

Also, we can see some stops typically associated with other colleges in the area presenting minimal numbers in the area of a couple hundred.

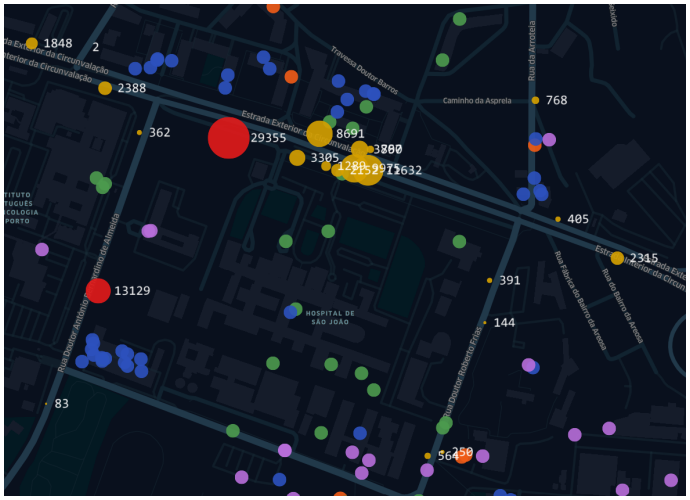

Figure 5.50: Google places and Stop data: Hospital S. João May 2020

Examining Figure 5.51 we find that after two months of the recovery process, the number of trips has gone up to more than double for some stops, showing just how much the population that usually would use these services to reach the hospital area had avoided doing so in the first months out of the lockdown, these increases in usage are not guaranteed to be related to the use of health services.

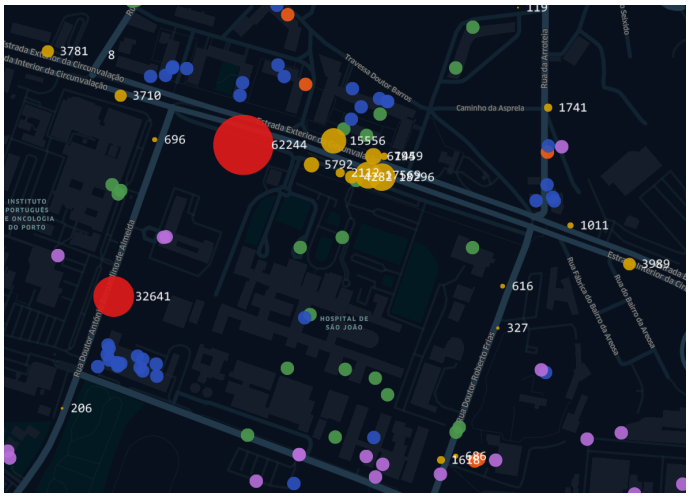

Figure 5.51: Google places and Stop data: Hospital S. João July 2020

Analyzing Figure 5.52 presenting the data for September 2020, an increase in the number of trips for both metro and bus is apparent, believed to be related to the return of some university in-person activities. These numbers are still inferior to the ones from 2019, as seen in Figure 5.53,

These impacts are not distributed similarly over all the bus and metro stops, and some suffer more than others.

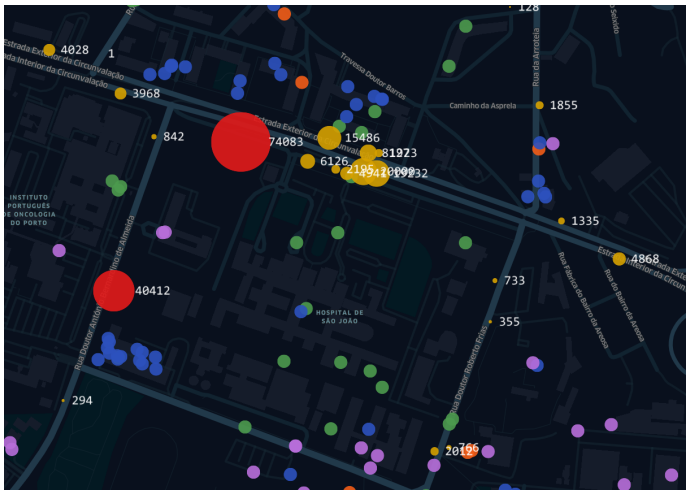

Figure 5.52: Google places and Stop data: Hospital S. João September 2020

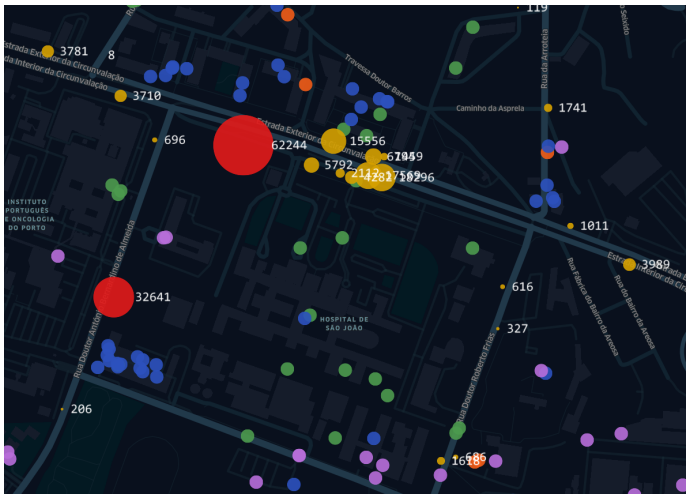

Figure 5.53: Google places and Stop data: Hospital S. João September 2019

Regarding the southern part of the university campus, as expected for May 2020, the number of trips to the bus stops directly in front of Universities is significantly reduced; however, the number of trips in the Polo Universitario metro station has not suffered from this, as can be seen in Figure 5.54, probably due to the stop being very close to the UPTEC start-up hub and other places of work, these would already possibly have reinstated in-person work.

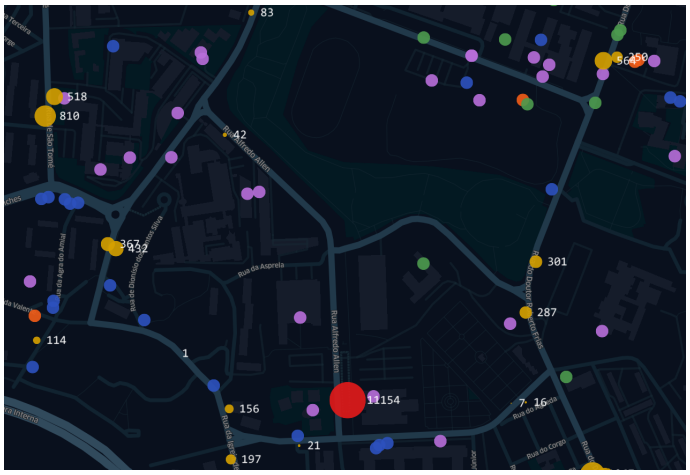

Figure 5.54: Google places and Stop data: College campus May 2020

From May to July, an increase similar to the one identified in the hospital area was seen in Figure 5.55 for the stops surrounding the Polo area, most zones increasing to close to double the number of trips they had in May.

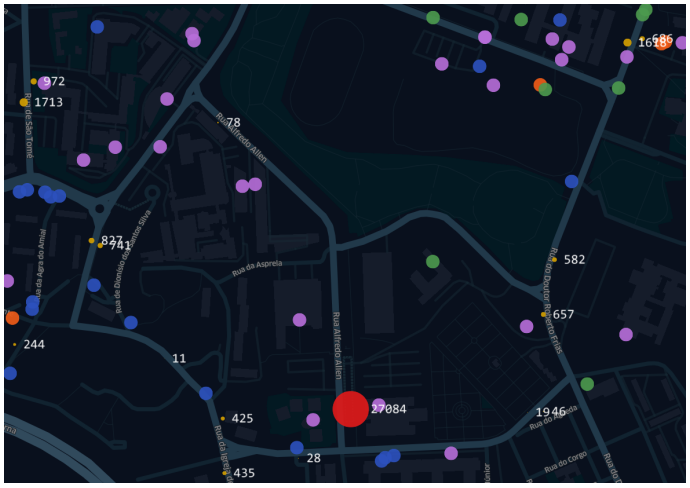

Figure 5.55: Google places and Stop data: College campus July 2020

In September, the increase predicted in trips for both metro and bus was verified. In Figure 5.56 we can see the larger values across the board compared to July; as with the hospital area, the return of in-person activities to some of the universities is believed to have played a significant role in these increases.

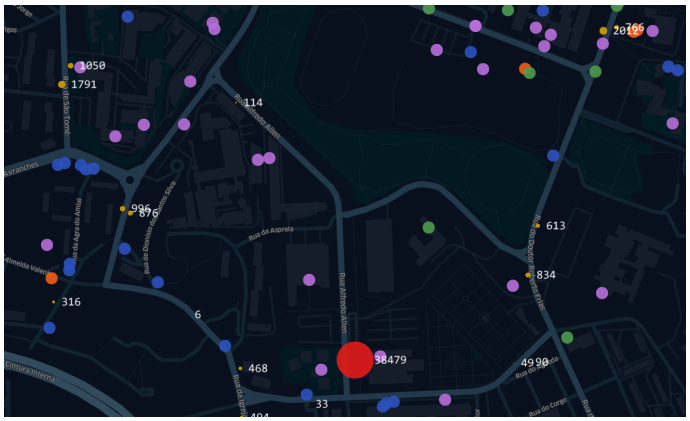

Figure 5.56: Google places and Stop data: College campus September 2020

Polo sees similar changes in the Hospital area when comparing 2019 to 2020 but on a lower intensity, as can be observed when comparing the previous figure to Figure 5.57.

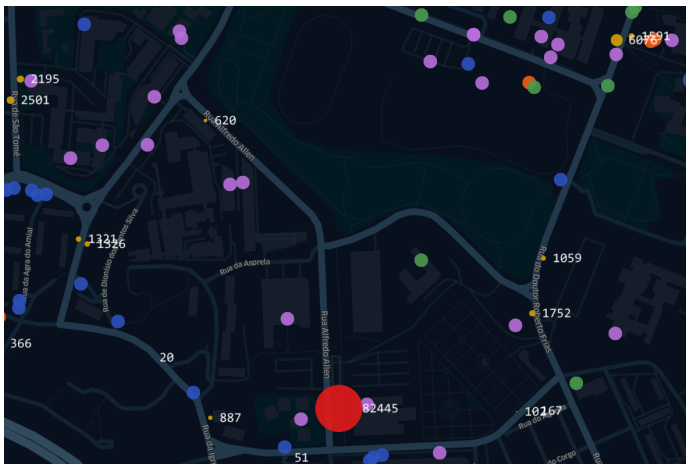

Figure 5.57: Google places and Stop data: College campus September 2019

Downtown Porto was the most significant cluster of restaurants and services places available in the usable zone; as such, it was selected to analyze the recovery patterns of travel to areas of more superfluous needs; however, no perfect use case exists for this since Porto is a very condensed city, all the downtown region has Many places of work as well as other types of points of interest.

We compare the month of June in Figure 5.58 during which restaurants were open but with a maximum number of people limitations to August in Figure 5.59 when most of these limitations had been lifted, then compare the distribution of trips in August 2020 with August 2019 in Figure 5.60.

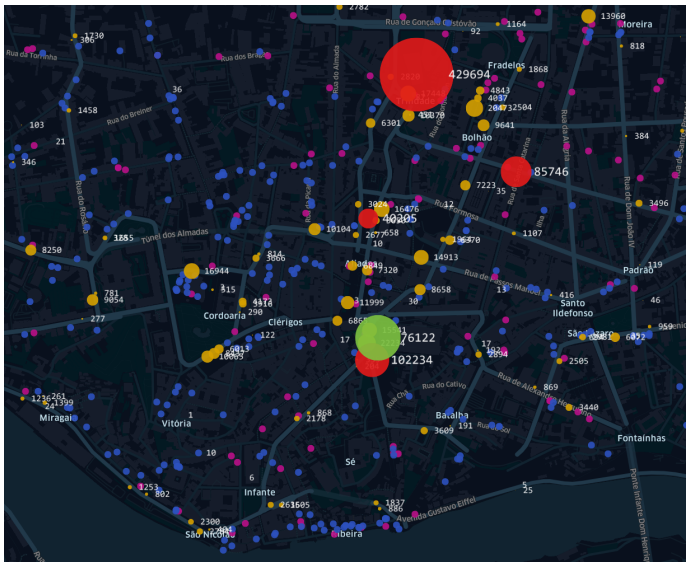

Figure 5.58: Google places and Stop data: Downtown Porto June 2020

As was expected, the numbers for August present significant increases from the June numbers; however, the increases are more significant in the large metro and train stations than on the bus stops closest to the areas with a high density of restaurants.

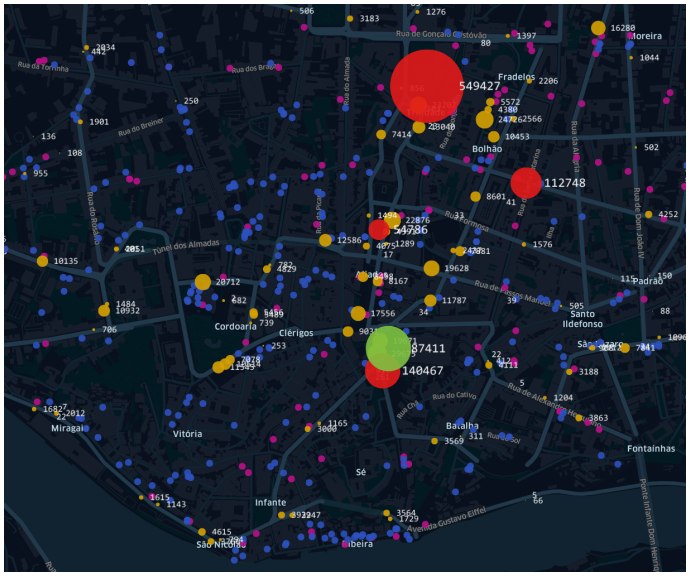

Figure 5.59: Google places and Stop data: Downtown Porto August 2020

When comparing the numbers for trips each year, the Metro services are shown to be impacted more than buses or train services, and some bus stations even managed near-total recovery in August 2020 to the numbers present in the previous year.

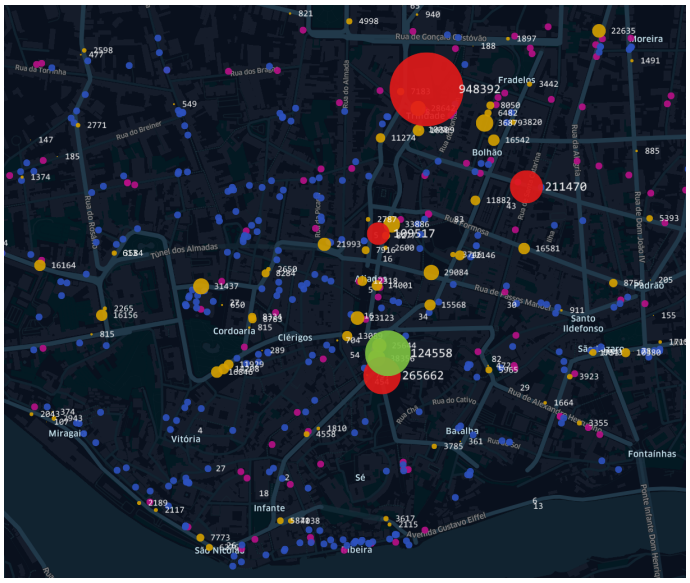

Figure 5.60: Google places and Stop data: Downtown Porto August 2019

Besides the university campus near hospital S. Jõao, another campus is closer to the center of Porto, near Campo Alegre, home to the Faculty of Sciences, and Architecture, among others. This campus is not serviced by Metro or Train lines, only by some bus lines; it is also more isolated, which means it is a relevant area to analyze to ascertain the impact that school starting had on public transport.

We looked at September 2020 and compared it to the numbers from July 2020 and September 2019. When considering the data from Figures 5.61 and 5.62, slight growth is identified; however, the magnitude of this growth is within expected values due to the natural recovery of the sector; it does not appear to be thanks to influence from school restarting.

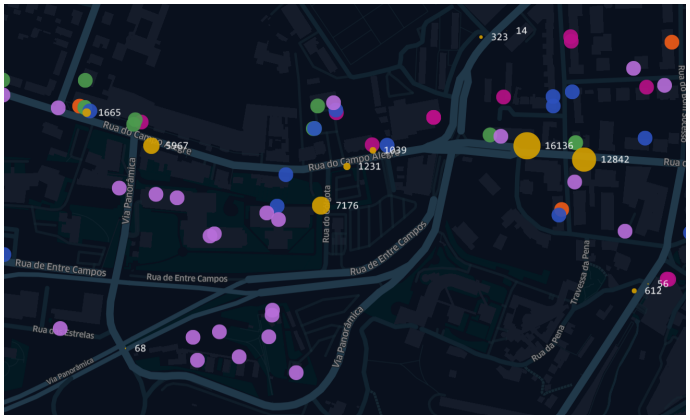

Figure 5.61: Google places and Stop data: Campo Alegre campus September 2020

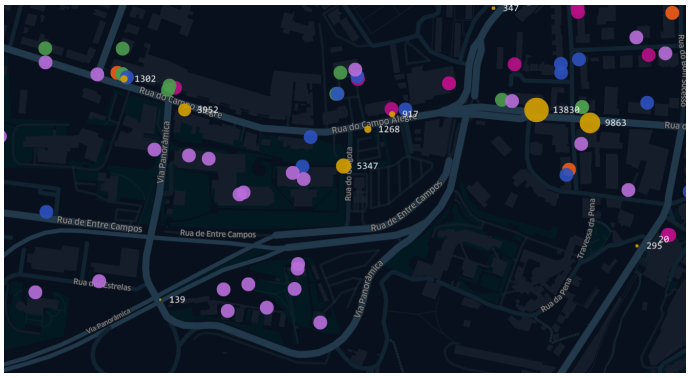

Figure 5.62: Google places and Stop data: Campo Alegre campus July 2020

When comparing the data from both years, a disparity in decreases can be found. Analyzing Figure 5.63 it is clear that multiple bus stops were impacted in different magnitudes, some now becoming less utilized than others that were not used before, and some even have managed to recover their trip numbers to a level nearing that of 2019.

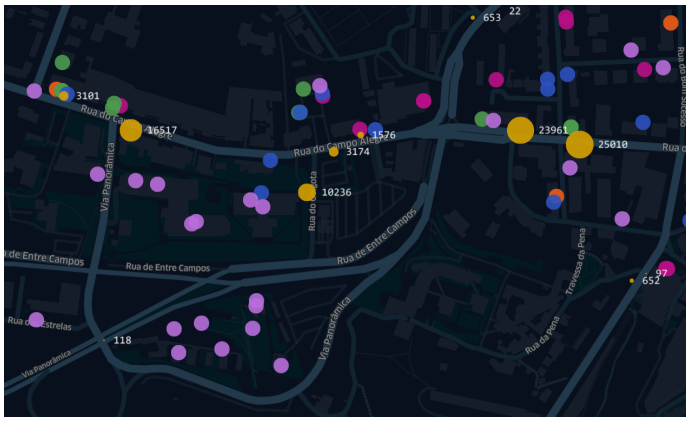

Figure 5.63: Google places and Stop data: Campo Alegre campus September 2019

As referenced before, during the summer period the are positive shifts in congestion for the region of Matosinhos due to it being a popular destination for young people to frequent the beach; as such, it was deemed a relevant point of interest for analysis specifically comparing both years and seeing if the distribution of trips across them was similar, although severely reduced for 2020.

Looking at Figures 5.64 and 5.65, we can see the expected reduction in trips for 2020. However, the metro stops seem to have been affected more harshly. Additionally, some bus stops seem to have recuperated almost completely, specifically the stops on the lower left of the map on the roundabout.

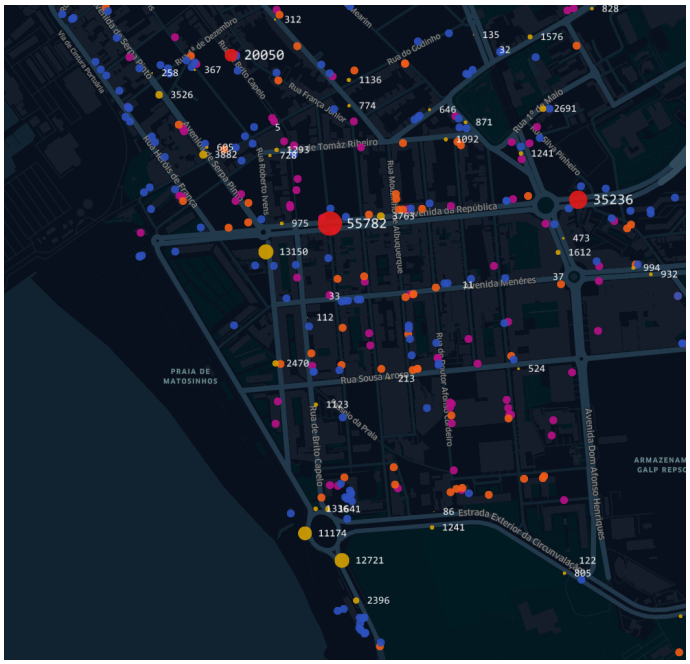

Figure 5.64: Google places and Stop data: Matosinhos Coast August 2020

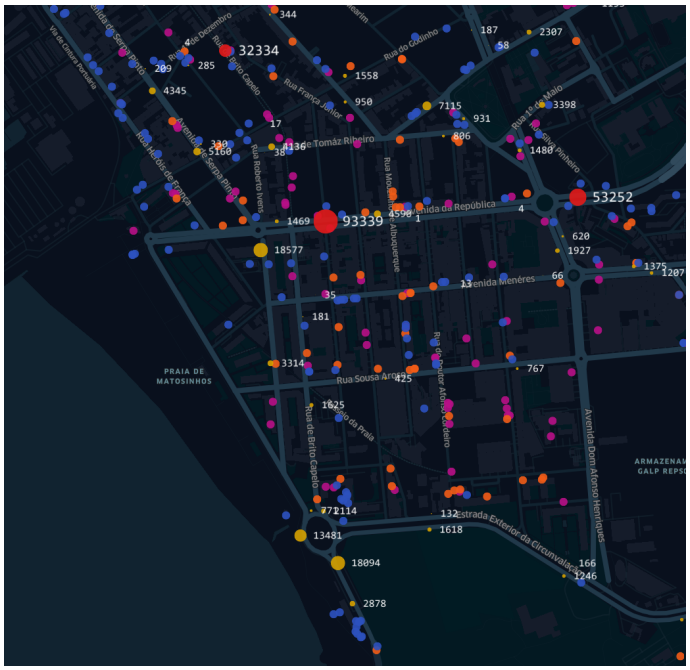

Figure 5.65: Google places and Stop data: Matosinhos Coast August 2019

5.10 Chapter Summary

In this chapter, we provide the principal analysis of the document. We are starting with a look at the overall trip values for both years and the trends. After, we look into the impacts on a fare zone level. Then, we look at the most valuable information extracted when analyzing trips by day of the month. Following, we look at the impacts on trip numbers by weekday. Afterward, we analyze the trip numbers for each time of day to showcase the changes in congestion throughout the day. Then, we look at the effects of the pandemic on the AMP service operators. Following, we utilize the discount data to create sub-groups based on different demographics and analyze the impacts across multiple criteria such as age, discounts, and income. After, we look at the most common estimated destination zones. Finally, we present the data for stops and stations and then cross-reference it and the Google Places data for specific points of relevance in the AMP.

Chapter 6

Results and Discussion

6.1 Effects of the public safety measures on total ridership 106
6.2 Recovery 108
6.3 Impact on transport use throughout the day 109
6.4 Impact on different demographics 110
6.5 Impact on different modes of public transport 111
6.6 August Decreases 112
6.7 Chapter Summary 113

This chapter draws from the analysis of the previous chapter to discuss some interpreted results. Firstly, in Section 6.1 the observed effects of public safety measures are discussed. Section 6.2 discusses the trajectory and overall recovery in the post-lockdown period. This will be followed by assessing the impacts on transport use throughout the day in Section 6.3. Section 6.4 discusses the effects on different demographics identified. In Section 6.5 the impacts on different modes of public transport are discussed. In Section 6.6, using the results previously presented are used to attempt to find an explanation for the decrease in trip numbers for August 2020.

6.1 Effects of the public safety measures on total ridership

From the second week of March 2020 until the time of writing, July 2022, there have been restrictions due to public health and safety concerns caused by the pandemic. The level of strictness of the restrictions wildly fluctuated during this period. This project specifically focused on the year of 2020 and the restrictions implemented and their impacts.

Looking at the year of 2020 chronologically, the first restrictions were social gathering and travel restrictions implemented in the second week of March. However, these would soon be overshadowed by the implementation of the first lockdown. The lockdown lasted until the 4th of

May when restrictions started to be reduced, and the everyday use and regular record taking for public transport restarted, but the limits of passengers per bus was limited to two-thirds of the total standard capacity.

The first and second large-scale easing of restrictions were only separated by two weeks, as seen in Table 1.1. Restrictions lifted on the 4th of May allowed for the return of public transport, although with reduced maximum capacity, the re-opening of small stores and libraries as well as the increasing the limit for public gatherings to a maximum of 10 people. Restrictions lifted on the 18th of May built on the previous ones, allowing for re-opening medium-sized stores, restaurants, cafes, museums, and others, and resuming in-person classes for the last two years of secondary school while requiring mandatory mask usage for all of these. The time between the restrictions being lifted was so reduced that accurately assessing the impact of each is speculative at best, especially considering the lockdown was just lifted and there was no expected behavior data to compare the numbers to.

On the 25th of May, a new restriction was put into law to combat overcrowding at beaches during the summer, and this law may have impacted public transport use to typical beach destinations such as the Matosinhos coast.

The third lifting of restrictions occurred by the end of May and the start of June. This time allows large-scale stores to open, no longer requiring that work-from-home be used whenever possible but only advisable, kindergartens, day-cares, theaters, and cinemas are also allowed to resume operations, with mask restrictions, and national football activities were resumed. These restrictions were further from the end of lockdown, and by comparing the numbers of May to those of June, the results seem to indicate a significant impact. June 2020 saw an increase of over 58% in public transport usage, and it is crucial to take into consideration that the growth tendencies shown in May throughout the month are steady increases and that within May itself, the number of trips per day increased from under 150 thousand to close to 250 thousand. In contrast, there is no doubt that the constant removal of restrictions contributed to the recovery happening linearly for May and early June, as mentioned in 5.3.2.

After the third easing of restrictions, none more would be lifted during June and July. From the third lifting onward until August, steady growth in trips per day was present. The recovery details during this period are looked at in more detail in Section 6.2.

At the start of August, a final easing of restrictions occurs, transitioning the country to states of contingency and alert depending on the region and lifting low-level restrictions. Despite this lifting of restrictions, for the first time since the lockdown, the number of total trips saw a slight decrease in August. However, this decrease is not believed to be related to any particular public health and safety measure or pandemic. This is further explained in Section 6.6.

With a slight rise in cases with the Autumn season approaching and the start of in-person schooling for all schools, with universities being on a case-by-case basis, the government decided to place the country in a state of contingency on the 15th of September, re-instating multiple restrictions, for example, public gatherings being once more limited to only six people.

On the 17th of September, in-person schooling was immediately reflected in the data; increases were registered in total numbers, and the discount groups were populated by students. The restart of schooling possibly diminished the impact of the increase in restrictions two days prior, but this is a double-edged sword. The restrictions also diminished the increase the restart of classes could have produced.

The government's fears of a second wave striking during the Autumn and Winter started to be validated in the middle of October when the number of cases started to rise more rapidly. The country's state was increased from contingency put into place in September to calamity. However, some restrictions were not reflected in the October data, as it still showed increases in total numbers.

By the end of November, the second wave was undeniable, so the state of emergency was re-instated, and restrictions increased once more as case numbers reached record numbers at the time. This would be the last increase in restrictions for the year of 2020. The month of November had already seen a decrease in total trips due to what is believed to be a mixture of the increase of restrictions in October and the worsening case numbers.

By looking at the data from December 2020 present in Figure 5.16, the impact of the measures implemented in November should be noticed. This was the only relevant period of time to analyze the impact, as the increases in measures were done in the latter half of November.

However, context is relevant in this particular case. December contains multiple public holidays during the month, as well as Christmas and the New year period, decreases in public transport travel are expected in these portions of the year as Portugal is a mainly Christian country, and the large majority of the population celebrates Christmas, and schools and universities are on break. As such, there is no accurate way to evaluate these measures' impact on the ridership levels.

6.2 Recovery

Immediately after the end of the lockdown, the trajectory identified in the data was one of steady recovery. However, the initial numbers of trips for the month of May 2020 were considerably low at 4.3 million trips; it was less than a third of the last whole month of data not affected by COVID-19, February 2020. There are important details to consider in these numbers. As mentioned above, the maximum capacity of buses was limited to two-thirds of the regular maximum. Furthermore, only two years of school had returned to in-person school to prepare for the national exams.

The recovery rate was at its highest for the period of June, benefiting from the lifting of measures in May continued the steady recovery process into July.

During these two months, the recovery trajectory slowed down, with July presenting an increase of just over 31%, twenty-seven percentage points lower than the previous month. However, the speed of recovery was expected to slow down as more and more users returned. Additionally, the total number of trips increased in the two months was very close to each other. With June showing a recovery rounding 2.5 million trips and July 2.1 Million, the percentage increase was much smaller, but the same was not true for the total number of trips.

July, although lowering the number of trips recovered slightly still ensured recovery remained at a somewhat steady rate. From July to August, the first decrease in total trip numbers was observed; however, this was quickly counteracted by an increase of 23% in total trip numbers from august to September, which raised the total trip number 1 million above what it was in July.

Due to a worsening context for October, the trip increase was only close to 700 thousand or 6%. At this point, the recovery of the public transport sector from the first lockdown and first wave stops, with the imminent start of the second wave trip numbers, started to go down and would follow this trend until the end of the year.

Recovery was highly accentuated in the first few months, slowly tapering off into lower percentages. From what is observed in the data, it is believed the recovery would have lasted longer and maintained higher numbers of trip recovery were it not for the worsening of the pandemic in the last trimester of the year. However as shown in Figure 5.4 the pattern of fluctuation is similar for the months of 2020 after May and the corresponding months of 2019, but with lower numbers of total trips and less accentuated rises and falls which indicates a lower turnover rate of customers, people that use public transport are being more consistent in their use.

6.3 Impact on transport use throughout the day

In Section 5.5 the graphs and results are presented to outline the drastic effects the pandemic had on the trip distribution by the time of day.

The observed results clearly show that in a typical year such as 2019, an average day would have two significant times of congestion.

In the morning, from 6:00 to 8:00, an accentuated rise in travel numbers is seen peaking at 8:00 and steadily diminishing in the following hours. The time when the number of trips starts to increase dramatically coincides with the start of regular public transport operations at 6:00. The first peak coincides with the morning rush hour, typically caused by most jobs starting at 8:00 or 9:00.

After the first significant peak in congestion, a smaller peak is seen around 13:00, and this would coincide with lunchtime for many jobs but more relevantly with the end of morning classes in many schools and universities.

The second major time of congestion is seen between 17:00 and 18:00. Typically these are the hours most people leave their 8-hour jobs after starting at 8:00 or 9:00. As such, this peak very closely mirrors the morning one as most trips are return trips in the opposite direction.

When looking at 2020, the same three spikes in congestion are seen. However, the magnitude of these is substantially reduced. Overall the representation of the trips during the day is much more homogeneous.

Multiple factors are believed to be significant influences in lowering these peaks. The widespread adoption of work-from-home and virtual classes effectively removed the most prominent reasons why the two most significant congestion points existed. This can be supported by Figure 5.25. In this graphical representation, the month of May shows the most homogeneous line. During the

first half of this month, work-from-home was still considered a must for any non-essential case, and only some secondary school students attended school in person. During the following summer months, when schools and universities were on break and work-from-home was now only recommended, the lines became more heterogeneous. However, the biggest argument towards these being the causes are the lines for the months of September, October, and November, and once in-person classes resumed, an evident change is present in the lines that accentuated the peaks that had already been growing in the previous months. December serves to confirm this point, as during the month, most schools and universities are on vacation for one or two weeks and this month once more reverts to a much more homogeneous state. It is important to note that in December, the country was back in a state of emergency, and work-from-home returned to being a recommendation in all non-essential cases. Also, the line for March should be ignored in the graph, and it ends up together with the relevant months due to the volume of trips in this month.

6.4 Impact on different demographics

A pandemic's impacts on different demographics are expected to differ considerably while maintaining the overall negative general impact. No demographic saw growth in the total number of trips, the analysis did, however, show changes in the percentages distribution for different subgroups.

Following the previously identified impact that school returning had, the data for the eight months after lockdown showed that the number of children and young adults utilizing public transport decreased considerably compared to the first three months of the year. The decrease was nearly 6 percentage points for the percentage of trips performed by people under 25 years old. This reversed the previous trends. From the latter eight months of 2019 to the starting three of 2020, the percentage of trips done by under 25-year-olds had increased by precisely 4 percentage points.

The percentage then missing from under 25-year-olds was mainly gained by adults with an increase of over 5 percentage points and the elderly age group with a slight increase of 0.8 percentage points. The pandemic was expected to impact adults' percentages negatively due to the implementation of widespread work-from-home, and the elderly, considered people at risk, were expected to decrease their public transport usage. However, these factors proved less relevant than the lack of in-person school activities.

Besides, age analysis was performed to see how the income of users influenced the impacts experienced. It was expected that the lower-income people would see a rise in percentages due to having no other alternative to public transport, as people with average or high incomes were expected to abandon public transport at least temporarily. However, the results show that low-income people were the ones that saw their percentages decrease, but only slightly, around 0.6 percentage points.

Finally, when crossing the financial and age analyses, the data revealed the effects were particularly devastating for the under 25 normal income sub-group with a decrease of nearly 3 percentage points. However, low-income under 25-year-olds were also negatively affected with decreases of over 2 percentage points.

As a result, the adult percentage increased, specifically the normal income adults sub-group increased by 3.6 percentage points, followed by the low-income adult group with a 1.6 percentage points increase in trip percentages.

This was reflected in the analysis of trip percentages by discount percentage. The percentages that saw significant decreases were those typically associated with young riders and young riders with low-income support, the 25% and 60% brackets, the second of which is only available to students who receive financial aid.

6.5 Impact on different modes of public transport

Due to the specific characteristics of each mode of travel, different impacts were expected to affect each of the three, bus, metro, and train.

The train services provided by CP in the AMP were the least affected, percentage-wise of the three, in the eight months following the lockdown, only showing a minimal decrease of 0.3 percentage points from the first three months of 2020 and bolstering higher numbers than in the same eight-month period of 2019 where the percentage was 0.2 percentage points lower. The low level of fluctuation in these percentages is believed to be caused by a significant percentage of the trips performed by train being trips in which are destined for out of the AMP or to and from regions for which the only other alternative is traveling by car, in sum the trips done by train are less likely to be replaced or outright removed from due to being superfluous.

Metro saw substantial travel percentage reductions in the latter eight months of 2020, with a reduction of over 4 percentage points total share. With work-from-home becoming widely used during the post-lockdown period, the metro service, which is heavily centralized around downtown Porto and focused on providing connections from peripheral locations to the center of the city, saw a loss of relevance because these office spaces did not need employees to travel to work in-person.

Bus services were the biggest winners in market share. As expected, STCP stands out from the rest of the providers, increasing its total share by over 5 percentage points. Other bus service providers had mixed results, Gondomarense saw a slight decrease, but Espírito Santo saw slight increases, among others. Bus service providers were expected to benefit from metro and train losing percentage points. Due to the nature of the bus as a mode of transport, its level of adaptation is far superior to metro and train, not requiring so much infrastructure and having the ability to change routes to adapt to circumstances, and this is one reason believed to have helped this mode increase its percentages. However, the biggest reason is believed to be the negative impact suffered by the metro. With downtown Porto not being as relevant of a destination, the increased speed of travel of the metro starts to be less valuable than the higher adaptability of using bus lines.

This general analysis extracted from the trips per operator data was corroborated by the data of Section 5.9, the images for downtown Porto and Matosinhos clearly show a substantially more significant impact on the metro, showing even multiple bus stops which managed to recover close to the 2019 total number of trips for the months in question.

6.6 August Decreases

In August, we see the first decrease in total trips since the end of the lockdown. However, as the number of cases was similar during this month and the previous month and the restrictions in place were reduced on the 1st of August, there appears to be no logical reason why COVID-19 would cause this decrease.

The hypothesis found to be most plausible is that this decrease was not related to the pandemic. Typically August is the month most Portuguese workers opt to take their vacation time. Even with overseas travel restrictions, travel inside the country was not prohibited at the time. This would cause reductions in the number of public transport users present in the region during some part of the month and lead to diminished numbers of public transport use. The most significant decreases seen within the month are in the third and fourth weeks, with the fourth week showing the start of a recovery in numbers as seen in Figure 5.12.

The idea that 2020's August decrease was not COVID-related is supported by the data from 2019. If a comparison is made between August 2019 and July 2019, a decrease in total trips can also be seen. This decrease is even more significant than in 2020. In 2019, the numbers decreased by 17%, and in 2020, they decreased by 7%.

So the results point to the conclusion that not only was this decrease not related to COVID, but it was smaller than what it usually would be for the time of the year due to the ongoing recovery of the public transport system.

In Table 6.1, it is clear that when comparing the intervals from July to September in both years, 2020 presented a much lower decrease in total trips during august and a substantial increase in trips after the presumed vacation period. While, 2019 showed a significantly meager increase in trips of around 1%. If the typical decrease for a standard year can be considered to be around 17% then the post lockdown recovery context roughly reduced these adverse effects by 10% and managed to produce a 14% increase by September,

Table 6.1: Comparison between July-September 2019 and 2020

	July	**August**	**September**	**Change Jul-Sep**
2019	15 533 451	12 815 512	15 798 424	+1.7%
2020	9 021 687	8 360 529	10 315 576	+14.3%

6.7 Chapter Summary

In this chapter, we discuss the analysis results and reach our main conclusions. It starts with a look at the effects of the public safety measures on total ridership. Followed by a look at the recovery trends during the eight months of data from the time after the lockdown. Afterward, we draw conclusions from the data obtained for trips in each hour of the day and relate them to the times of year and restrictions in place. Then, we look at the impacts felt by each demographic group previously established and compare it to what was initially expected through the popular public understanding. After, we analyze the impact felt by each of the different modes of public transport(bus, metro, train). Finally, we utilize the data recovered from multiple queries to explain why the decrease in trips in August 2020 was not related to COVID-19 but to the underlying context of the time of year and habitual behavior of the Portuguese workforce.

Chapter 7

Conclusions

The impact of COVID-19 reverberates through our society daily, and the public transport sector is affected by it. This work aims to identify some of the effects caused by the pandemic. This chapter concludes this document, reflecting on the totality of the project and the conclusions drawn from the work done in Section 7.1. Discussing the limitations of this work in Section 7.2. And finally outlining some possibilities for future work in Section 7.3

7.1 Conclusions and expected behaviours

This project's primary goal was to identify behaviors induced by the pandemic context, lockdowns, and restrictions through data analysis for the years 2019 and 2020. Using the provided data, analysis was performed on different geographical, demographic, temporal, economic, and contextual aspects of the post-lockdown recovery period.

Geographical data analysis showed that the aftermath of the lockdown caused the largest zone in the AMP, PRT1, to lose some of its total trip shares. Utilizing the data from Google Places and analyzing stops and stations data allowed it to be determined that different regions of the AMP recovered slightly differently at different points of the year.

The same data in conjunction with the operator-oriented analysis showed that metro services had been more significantly affected than bus services in the AMP, recovering slower in particular for the downtown of Porto, where even in the single largest metro station, Trindade, this slower recovery was apparent. Bus services recovered faster, and depending on the region and the stops themselves, near full recovery was possible. Train services saw proportional decreases in usage, not having suffered particularly significantly.

By looking at the temporal information in the data, we could ascertain that in the recovery period, times of congestion present in 2019 on regular days were drastically affected. The congestion peaks throughout the day were flattened, and the distribution of trips per hour was homogenized.

However, as the months passed, the behavior slowly reverted to the distribution seen in 2019. It was particularly notable after in-person classes started again in September.

In September, with the return of in-person classes, a clear impact was observed in the data, increasing the growth factor temporarily. This occurrence, in conjunction with the discount data analysis, made it clear that children and young adults under 25 were the most impacted group. The group saw a significant drop in trip shares compared to adults and the elderly, which was unexpected. Furthermore, it was observed that the impact was more significant for under 25-year-olds who were not recipients of social support, but those who were recipients were also impacted. Adults saw a rise in ride share percentage, and the elderly barely saw any impacts.

When analyzing the data, it became clear that the public transport sector did not have enough time to recover properly between the first lockdown and the second wave, which eventually led to the second lockdown in early 2021. In October, the number of cases started to increase slowly. For the rest of the year, the impacts on the sector would be nothing but negative. By the end of November, the number was rising rapidly and setting new records every day. December saw the lowest point in trips since July.

Although we do not have the data for 2021 from what we saw in 2020, the assumption can be made that the recovery from January's lockdown would be similar. Starting with a significant base number of trips somewhere from a third to half of the early 2020 numbers and then moving into an accentuated recovery for the first months that would eventually slow down and reach an upper ceiling probably still below the numbers from before COVID but may be higher than the numbers for the first recovery period of around 11 million, a slower easing of restrictions in 2021 could lead to the recovery starting at a lower month by month increase percentage but remain at a more steady rate.

This project achieved its goal of identifying multiple relevant behaviours and impacts of the COVID-19 pandemic and the 2020 lockdown specifically. Some preliminary results were presented in the 14th annual edition of the CITTA conference on Planning Research. Furthermore, the literature review as been accepted into the Association of European Transport's (AET) European Transport Conference(ETC) and will be presented in Milan Italy, in a completed state on September 2022,

7.2 Limitations

Throughout its planning, implementation, and writing, this project faced problems and impediments of different forms that hindered the smooth progression of work and limited what could be achieved.

As alluded to in chapter 3 the sheer amount of data provided was tremendously large; when handling such quantities of data, multiple difficulties can arise. For the entirety of this project, the hardware used to store the data, implement the data structure, run operations and generate all the results was a personal computer. As such, the size of the data caused severe problems, the impossibility to place the totality of the data in the PostgreSQL database environment, and multiple

reoccurring crashes of the system due to query operations that were too demanding. When the demand was not to the point of causing a complete system crash, it caused the computer to be highly bottle-necked in terms of storage, hard drive access, and processing ability which meant the system was functionally unusable for the duration of the operation, some times these were operations that lasted over two hours. Eventually, workarounds were implemented that streamlined the process substantially. However, this was at the cost of doing the operation in a non-ideal way. In sum, the combination of the large amounts of data and hardware limitations slowed the implementation process tremendously, especially in the project's first tasks before forcing the solution to adapt in whatever way possible.

Besides the amount of data being troublesome, the data itself had many different problems that created a need for specific care or a change of strategy from what would be ideal. Across the two data datasets, there were thousands of error entries and system test entries. The test entries were known and were filtered out as planned. The unexpected entries that had errors required some specific analysis to identify common points to create the possibility of filtering them out. Another problem with the data was the inconsistency in the information available. Even though the andante system has a centralized collection of entries from all providers, they do not all supply the same information necessarily, which means some rather detrimental data was missing, for example, the only bus provider that had information about the stop in which a trip was validated was STCP, this meant it was impossible to analyze the use levels of stops from all other providers. In the same vein is the problem of the information related to metro and train stops only being present in the RR dataset and the STCP stop information only being available in the ROD dataset. There were also missing months in the data. The ROD dataset was missing the month of November 2020 and the RR dataset the month of October 2020.

Nevertheless, the biggest problem with the dataset was a lack of clarity. It was only within the last month of the project that it was discovered that the entries in the ROD and RR datasets are not complementary to each other but instead two different forms of representing the same entries. This was after months of implementation and analysis of the data done with the idea that analyzing both datasets was needed to get the total number of trips in the AMP. It cannot be stressed enough how much this problem impacted the work; tens of hours were spent running redundant scripts and incorrectly performing analysis until this was found out. Although some operations would still have been necessary for each dataset despite redundancy to solve the problem previously mentioned of only ROD containing the stop information for STCP buses and RR the train and metro station information, not nearly as much time would have been expended to perform the needed operations. After this realization, another look at the analysis performed in multiple queries had to be taken, and many planned points of the final document structure had to be adapted. However, this allowed for solving the problem involving missing months in each dataset. To a certain extent, we could use the datasets to complement each other except for the stop and station information.

Although not directly a problem for the project completion, the fact that the data provided only covered the years 2019 and 2020 caused multiple limitations on what could be achieved. As

seen in chapter 2 Portugal's most prominent peak in COVID-19 was at the beginning of 2021, the country had one of the highest cases by day per ten thousand people numbers, in the world, and the recovery from the second lockdown was much more gradual. Having dealt with a lockdown situation in 2020, the government was much better prepared to gradually implement measures and remove them after they were no longer needed. The more gradual implementation and retraction of restrictions coupled with the gravity of the situation being significantly higher would have made the 2021 data extremely valuable to analyze. When many public safety measures are implemented in a very short interval of time, it is very difficult or even impossible to determine what the impact of each measure was, and this is what happened in 2020, immediate implementation of a lockdown and a fast lifting of restrictions which made it difficult to analyze that aspect of the impacts. Also, only having access to eight months of data after the first lockdown limits the long-term conclusions that can be drawn.

The problem of the sheer amount of graphical data and representations produced in this project is directly related to the writing of this document. In writing chapter 5 the most essential and condensed versions of the graphs were utilized, and restraint had to be exercised; otherwise, there would be over one hundred relevant figures in the document. To avoid this, a more condensed version of the analysis focusing on larger groupings of months was written instead of presenting a month-by-month analysis with all relevant graphs for each month. Due to the nature of the project, this would be expected. The possibilities for data representations were highly diverse and could allow for more in-depth representations in future work.

Finally, a problem that plagued the project was lousy timing estimation and management. Due to inexperience working with some tools and performing some of the tasks needed in the course of this project, some parts were expected to last roughly one to two weeks but, in turn, lasted a month, and some tasks were initially thought to be exceptionally long turned out to be quicker than expected when done in different manners. Those time wrong time estimations, coupled with the revelation about the datasets, made the deadlines tighter than they should have been and limited what could be achieved in the time frame of the semester.

7.3 Future work

Analyzing transportation patterns in response to crisis and planning accordingly is highly nuanced and complex. When some results of this project were presented in the 14th edition of the CITTA conference, there were opportunities to analyze all sorts of projects around the subject, allowing for a better understanding of the field.

In direct continuation of this project, some clear avenues exist for further work. One could perform a more in-depth analysis of one or multiple aspects analyzed here, for example, the correlation of the recovery progression in an area with the characteristics of the services in the area. This project looked superficially at some areas of interest only due to time restrictions. However, a more widespread and in-depth analysis could be performed in future work, focusing more on correlation with Google Places information.

Another example of performing a more in-depth analysis of a specific point of this project could be a deeper dive into the impacts of COVID-19 on the travel behaviors of different demographics of people, possibly even coupled with large-scale questionnaires to public transport users.

Analyzing how the new zones added in 2019 handled the integration into the system while dealing with a pandemic context could also be researched. The impact the recovery had on a subset of zones specifically, for example, the smallest zones or all the zones in a specific municipality, could reveal interesting information.

Another avenue would be continuing the work performed here with new or extra data resources. As mentioned in Section 7.2, the addition of the datasets of 2021 could introduce a whole new level of analysis and relevant information, for example.

Unfortunately, due to the expansion in 2019, using previous years' data as a baseline for overall behaviors would not be correct. However, they could be used as a baseline for the behavior of the zones that were part of the old Andante system grid.

However, the addition of new years ideally would be coupled with solving some of the problems present in the datasets, the most prescient being the lack of entry stop information for all bus service providers except STCP, in order to analyze the entire AMP Andante grid correctly, this is necessary, as metro and train are much more limited in their geographical reach than bus services.

With the advent of extra data for the region being incorporated, perhaps an attempt could be made to produce prediction algorithms for recovery. Utilizing more in-depth analysis of the effects each type of restriction has on ridership, more trip data possibly from 2021 and further years, and maybe even data from other similar urban environments in Europe, a machine learning and artificial intelligence approach could be used to try and create algorithms that could predict how the public sector of a city would recover from an outbreak of a disease. This could be relevant in the future if a health crisis of an epidemic or pandemic level happens and would reduce the unpredictability of the situation, allowing for better resource allocation during the crisis.

Using data from subsequent years coupled with the 2020 data from after the first lockdown, it becomes possible to research the effects of the population living within the pandemic context for an extended amount of time on post-lockdown recovery. With the pandemic dragging on and extending its impact on society for more than just a couple of months, it is reasonable to consider the possibility that the population would start to get accustomed to living with the inherent restrictions and dangers. This could lead to later instances of lockdowns seeing less drastic reductions in public transport use after being lifted or perhaps seeing more steady recovery curves. With the pandemic lasting two years and multiple lockdowns being put in place in several regions all over the globe, this research could be feasible. It would fall between data analysis and psychology to understand and explain why this is or is not the case.

The planning of routes is exceptionally complex, and to achieve the closest to an ideal result, a lot of expertise and analysis must be present and performed. Studies in the same vein of this project are incredibly relevant to understanding the tendencies and behaviors of the public transport using members of a population; as seen in the results of this project, some zones of congestion exist in

the AMP, and during a typical year such as 2019 severe points of congestion existed throughout the day at peak hours, this was not seen as much for the year of 2020. However, nothing can guarantee that as the public returns to in-person work, these will not manifest again, certainly not to the same extent due to the popularization of remote work, but an increase from the 2020 results should already be present in the 2022 data. Dedicated research should be done on how best to handle these congestions with human and mechanical resource allocation.

Research into possible new routes connecting unserviced parts of the AMP to areas of high relevance, such as downtown Porto or Vila Nova de Gaia, should be a focus. The expansion of the Andante system removed some entry barriers for users from regions that previously only had their local services providers. With these barriers removed, new routes could be a viable investment. However, due to the level of investment required to expand the grids, this possibility must be researched thoroughly.

Finally, as future work already initiated the literature review present in Chapter 2 will be completed and written into an article to be presented in September in Italy and published afterwards.

References

[1] Muhammad Abdullah, Nazam Ali, Muhammad Ashraf Javid, Charitha Dias, and Tiziana Campisi. Public transport versus solo travel mode choices during the covid-19 pandemic: Self-reported evidence from a developing country. *Transportation Engineering*, 5:100078, 9 2021.

[2] E Almlöf, I Rubensson, M Cebecauer, and E Jenelius. Who continued travelling by public transport during covid-19? socioeconomic factors explaining travel behaviour in stockholm 2020 based on smart card data. *European Transport Research Review*, 13, 2021.

[3] A Aloi, B Alonso, J Benavente, R Cordera, E Echaniz, F Gonzalez, C Ladisa, R Lezama-Romanelli, A Lopez-Parra, V Mazzei, L Perrucci, D Prieto-Quintana, A Rodriguez, and R Sanudo. Effects of the covid-19 lockdown on urban mobility: Empirical evidence from the city of santander (spain). *SUSTAINABILITY*, 12, 2020.

[4] Joao T Aparicio, Elisabete Arsenio, Rui Henriques, and Juan Gómez. Understanding the impacts of the covid-19 pandemic on public transportation travel patterns in the city of lisbon. *SUSTAINABILITY*, 2021.

[5] Sebastian Astroza, Alejandro Tirachini, Ricardo Hurtubia, Juan Antonio Carrasco, Angelo Guevara, Marcela Munizaga, Macarena Figueroa, and Valentina Torres. Mobility changes, teleworking, and remote communication during the covid-19 pandemic in chile. *Findings*, 7 2020.

[6] T Campisi, S Basbas, A Skoufas, N Akgun, D Ticali, and G Tesoriere. The impact of covid-19 pandemic on the resilience of sustainable mobility in sicily. *SUSTAINABILITY*, 12, 2020.

[7] Sílvia Caneco. Qual, afinal, a diferença entre estado de alerta, de contingência e de calamidade? Available at https://visao.sapo.pt/atualidade/politica/2020-10-14-qual-afinal-a-diferenca-entre-estado-de-alerta-de-contingencia-e-de-calamidade/, 10 2020.

[8] P Christidis, A Christodoulou, E Navajas-Cawood, and B Ciuffo. The post-pandemic recovery of transport activity: Emerging mobility patterns and repercussions on future evolution. *SUSTAINABILITY*, 13, 2021.

[9] Graham Currie and Alexa Delbosc. Exploring the trip chaining behaviour of public transport users in melbourne. *Transport Policy*, 18:204–210, 1 2011.

[10] Graham Currie, Taru Jain, and Laura Aston. Evidence of a post-covid change in travel behaviour – self-reported expectations of commuting in melbourne. *Transportation Research Part A: Policy and Practice*, 153:218–234, 11 2021.

[11] Diário da Républica. Legislação covid-19. Available at https://dre.pt/dre/geral/legislacao-covid-19. Accessed: 2022-02-23.

[12] Paul de Nailly, Etienne Côme, Allou Samé, Latifa Oukhellou, Jacques Ferriere, and Yasmine Merad-Boudia. What can we learn from 9 years of ticketing data at a major transport hub? a structural time series decomposition. *https://doi.org/10.1080/23249935.2021.1948626*, 2021.

[13] Metro do Porto. Metro grid. Available at https://www.metrodoporto.pt/uploads/document/file/471/EUT_____PMV_CMYK.pdf. Accessed: 2022-06-05.

[14] Marta Campos Ferreira, Teresa Galvão Dias, and João Falcão e Cunha. Anda: An innovative micro-location mobile ticketing solution based on nfc and ble technologies. *IEEE Transactions on Intelligent Transportation Systems*, pages 1–10, 2021.

[15] Konstantinos Gkiotsalitis and Oded Cats. Public transport planning adaption under the covid-19 pandemic crisis: literature review of research needs and directions. *https://doi.org/10.1080/01441647.2020.1857886*, 41:374–392, 2020.

[16] João Angélico Marta Campos Gonçalves and Teresa Galvão Dias. Understanding urban mobility patterns of tourists through data extraction. 2021.

[17] P Haddawy, S Lawpoolsri, C Sa-Ngamuang, M S Yin, T Barkowsky, A Wiratsudakul, J Kaewkungwal, A Khamsiriwatchara, P Sa-Angchai, J Sattabongkot, and L Cui. Effects of covid-19 government travel restrictions on mobility in a rural border area of northern thailand: A mobile phone tracking study. *PLoS ONE*, 16, 2021.

[18] V Konečný and M Brídziková. The impact of the state of emergency on the supply of services and passenger demand for public transport. *LOGI - Scientific Journal on Transport and Logistics*, 11:56–67, 2020.

[19] Instituto Montaigne. Europe versus coronavirus - portugal: Health crisis in check. next up, political crisis? https://www.institutmontaigne.org/en/blog/europe-versus-coronavirus-portugal-health-crisis-check-next-political-crisis. Accessed: 2022-02-23.

[20] H S Munawar, S I Khan, Z Qadir, A Z Kouzani, and M A P Mahmud. Insight into the impact of covid-19 on australian transportation sector: An economic and community-based perspective. *SUSTAINABILITY*, 13, 2021.

[21] Christian Martin Mutzel and Joachim Scheiner. Investigating spatio-temporal mobility patterns and changes in metro usage under the impact of covid-19 using taipei metro smart card data. *Public Transport*, 8 2021.

[22] World Health Organization. Who director-general's opening remarks at the media briefing on covid-19 - 11 march 2020. Available at https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020, 3 2020.

[23] VdA Legal Partners. Estados de calamidade e de emergência. Available at https://www.vda.pt/pt/publicacoes/insights/estados-de-calamidade-e-de-emergencia/21881/, 5 2022.

[24] I Politis, G Georgiadis, A Nikolaidou, A Kopsacheilis, I Fyrogenis, A Sdoukopoulos, E Verani, and E Papadopoulos. Mapping travel behavior changes during the covid-19 lock-down: a socioeconomic analysis in greece. *European Transport Research Review*, 13, 2021.

[25] PRISMA. Prisma checklist. Available at http://prisma-statement.org/documents/PRISMA_2020_checklist.pdf. Accessed: 2022-03-01.

[26] Público. Esperança e sobressalto. https://www.publico.pt/2021-ano-da-mascara/cronologia. Accessed: 2022-02-23.

[27] Sara Riobom. Porto train grid. Available at https://portoalities.com/pt/estacoes-comboio-no-porto-guia/. Accessed: 2022-06-05.

[28] Leonardo Rubiano and Georges Darido. The ticket to a better ride: How can automated fare collection improve urban transport? *Transport for Development*, 2 2019.

[29] Sam Schwartz. Global mobility response to covid-19 how cities are responding, recovering, and reopening transportation systems around the world, 2020.

[30] S Sinko, K Prah, and T Kramberger. Spatial modelling of modal shift due to covid-19. *SUSTAINABILITY*, 13, 2021.

[31] S.N.S. Plano de desconfinamento. Available at hhttp://arquivo2020.cm-montemornovo.pt/pt/site-municipio/amunicipal/COVID-19/Medidas%20de%20Preven%C3%A7%C3%A3o%20e%20Conten%C3%A7%C3%A3o/Paginas/Plano-de-Desconfinamento.aspx, 5 2020.

[32] STCP. Stcp general day grid. Available at https://www.stcp.pt/pt/viajar/mapas/procurar-mapas/?a=redes-download&rede=4. Accessed: 2022-06-05.

[33] Tiago Tamagusko and Adelino Ferreira. Data-driven approach to understand the mobility patterns of the portuguese population during the covid-19 pandemic. *SUSTAINABILITY*, 12, 2020.

[34] Alejandro Tirachini and Oded Cats. Covid-19 and public transportation: Current assessment, prospects, and research needs. *Journal of Public Transportation*, 6 2020.

[35] Bert van Wee and Frank Witlox. Covid-19 and its long-term effects on activity participation and travel behaviour: A multiperspective view. *Journal of Transport Geography*, 95:103144, 7 2021.

[36] Matia Vannoni, Martin Mckee, Jan C Semenza, Chris Bonnel, and David Stuckler. Using volunteered geographic information to assess mobility in the early phases of the covid-19 pandemic: a cross-city time series analysis of 41 cities in 22 countries from march 2nd to 26th 2020. *GLOBALIZATION AND HEALTH*, 16, 2020.

[37] Stefaan Vande Walle and Therese Steenberghen. Space and time related determinants of public transport use in trip chains. *Transportation Research Part A: Policy and Practice*, 40:151–162, 2 2006. Data source: Belgica 3 regioes com diferentes caracteristicas (1998.1999)
.

Appendix A

Code Snipets

```
import time
import datetime
start_time = time.time()
#### SCRIPT CODE ####
t = time.time() - start_time
print("--- TOTAL TIME: %s ---" % str(datetime.timedelta(seconds=t)))
```

Listing A.1: Timing the execution of a script

```
import psycopg2
# Database connection
conn = psycopg2.connect(database="postgres", user="********", password="*******",
    host="127.0.0.1", port="5432")
cursor = conn.cursor()
cursor.execute('''CREATE TABLE RR_ENTRIES
      (ID SERIAL,
      CODE BIGINT NOT NULL,
      CARD TEXT NOT NULL,
      ...
      PRIMARY KEY(ID,CODE,CARD));''')
conn.commit()
conn.close()
```

Listing A.2: Creation of the database connection and tables

```
cursor.execute('''COPY discounts(code,description)
FROM '************/Data/TABELA~1/Desconto.txt'
WITH DELIMITER '\t'
CSV HEADER
NULL 'NULL'
```

```
ENCODING 'LATIN9';''')
```

Listing A.3: Populating of the discount table

```
###inside of an if else that checks for data type input ###
    # ROD Entries SQL input query
    sql_rod_entries = '''COPY rod_entries(code,CARD,SN_VALIDADOR,veiculo,
        provider_code,discount,tariff,
    linha_interna,variante,sentido,paragem_validacao,data_hora_validacao,zone)
    FROM '***********/Data/ROD/ROD_20{}_Filtered.txt'
    WITH DELIMITER '\t'
    CSV HEADER
    NULL 'NULL'
    ENCODING 'LATIN9';'''
    ### If else to check if the month number is < or > then 10 ###
        ### if to check desired input year ###
            cursor.execute(sql_rod_entries.format('190' + input_a))
            print('ROD ' + month + ' 2019 added successfully')
            cursor.execute(sql_rod_entries.format('200' + input_a))
            print('ROD ' + month + ' 2020 added successfully')
    # RR Entries SQL input query
    sql_rr_entries = '''COPY rr_entries(code,CARD,SN_VALIDADOR,PROVIDER_CODE,
        DISCOUNT,tariff,
    DATA_HORA_VALIDACAO,ENTRY_ZONE,ENTRY_STATION,EXIT_ZONE,LINE,SENTIDO,VARIANTE)
    FROM '***********/Data/RR/RR_20{}_filtered.txt'
    WITH DELIMITER '\t'
    CSV HEADER
    NULL 'NULL'
    ENCODING 'UTF8';'''
    ### If else to check the number of the month and handle special cases ###
        ### if to check the desired input year for May special case ###
            cursor.execute(sql_rr_entries.format('1905_1'))
            print('RR ' + month + ' 2019 added successfully')
            cursor.execute(sql_rr_entries.format('1905_2'))
            print('RR ' + month + ' 2019 added successfully')
```

Listing A.4: AFC entries uploading

```
    sql_rod = '''DELETE FROM rod_entries
    WHERE EXTRACT(YEAR FROM data_hora_validacao) = {}
    AND EXTRACT(MONTH FROM data_hora_validacao) = {};'''

    year = 2019
    cursor.execute(sql_rod.format(year, input_month))
    print('ROD ' + month + ' 2019 removed successfully')
```

Listing A.5: Use of 'DELETE FROM' in AFC entries remover

```
cursor.execute('''DO $$ DECLARE
  r RECORD;
BEGIN
  FOR r IN (SELECT tablename FROM pg_tables WHERE schemaname = current_schema())
      LOOP
    EXECUTE 'DROP TABLE ' || quote_ident(r.tablename) || ' CASCADE';
  END LOOP;
END $$;''')
```

Listing A.6: SQL command to fully delete the database

```
def get_new_zone(zone):
    count = 0
    for zones in old_zones:
        if zone == zones:
            zone = new_zones[count]
            return zone
        else:
            count += 1
```

Listing A.7: Method for obtaining the new zone from the list

```
while True:
    line = file1.readline()
    if not line:
        print("End of file parsing")
        break
    line_split = line.split('\t')

    # RR version
    zone_field_1 = line_split[9]
    zone_field_2 = line_split[11]
    zone_1 = zone_field_1.split(' ')[0]
    zone_2 = zone_field_2.split(' ')[0]
```

Listing A.8: Reading lines and zones from RR files

Appendix B

Kepler Maps

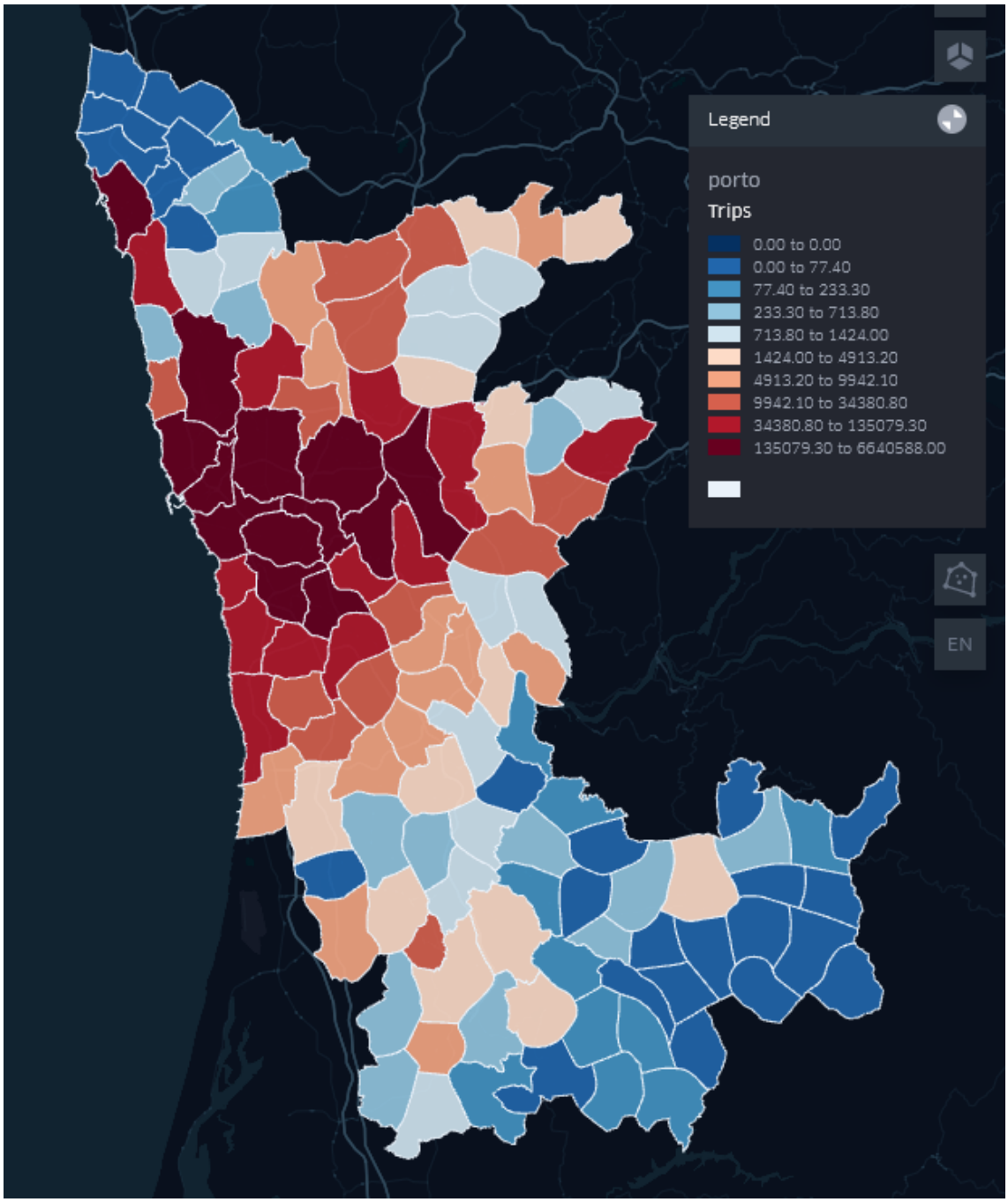

Figure B.1: 2D Color Coded Trip per Zone January 2020 Map

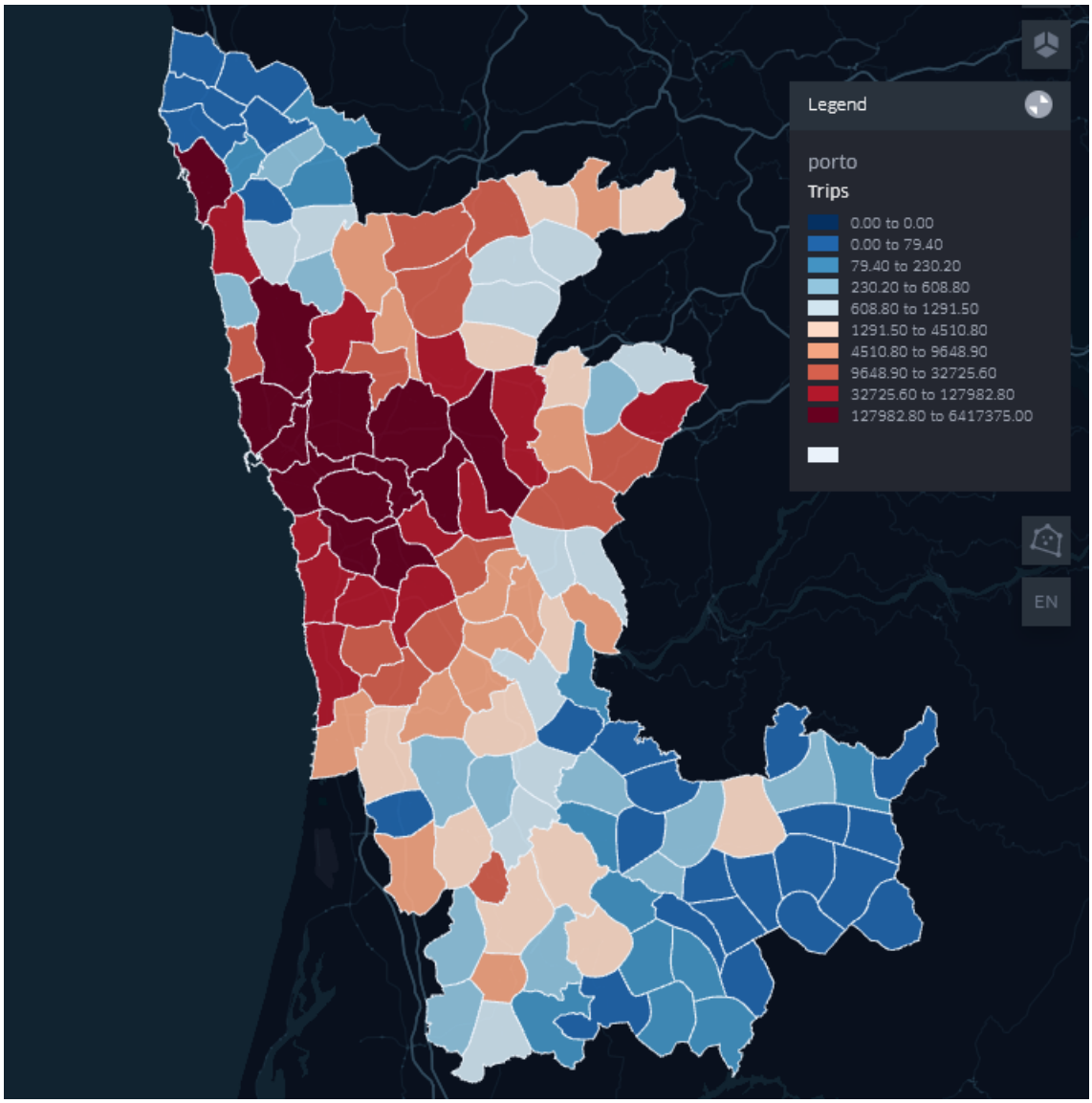

Figure B.2: 2D Color Coded Trip per Zone February 2020 Map

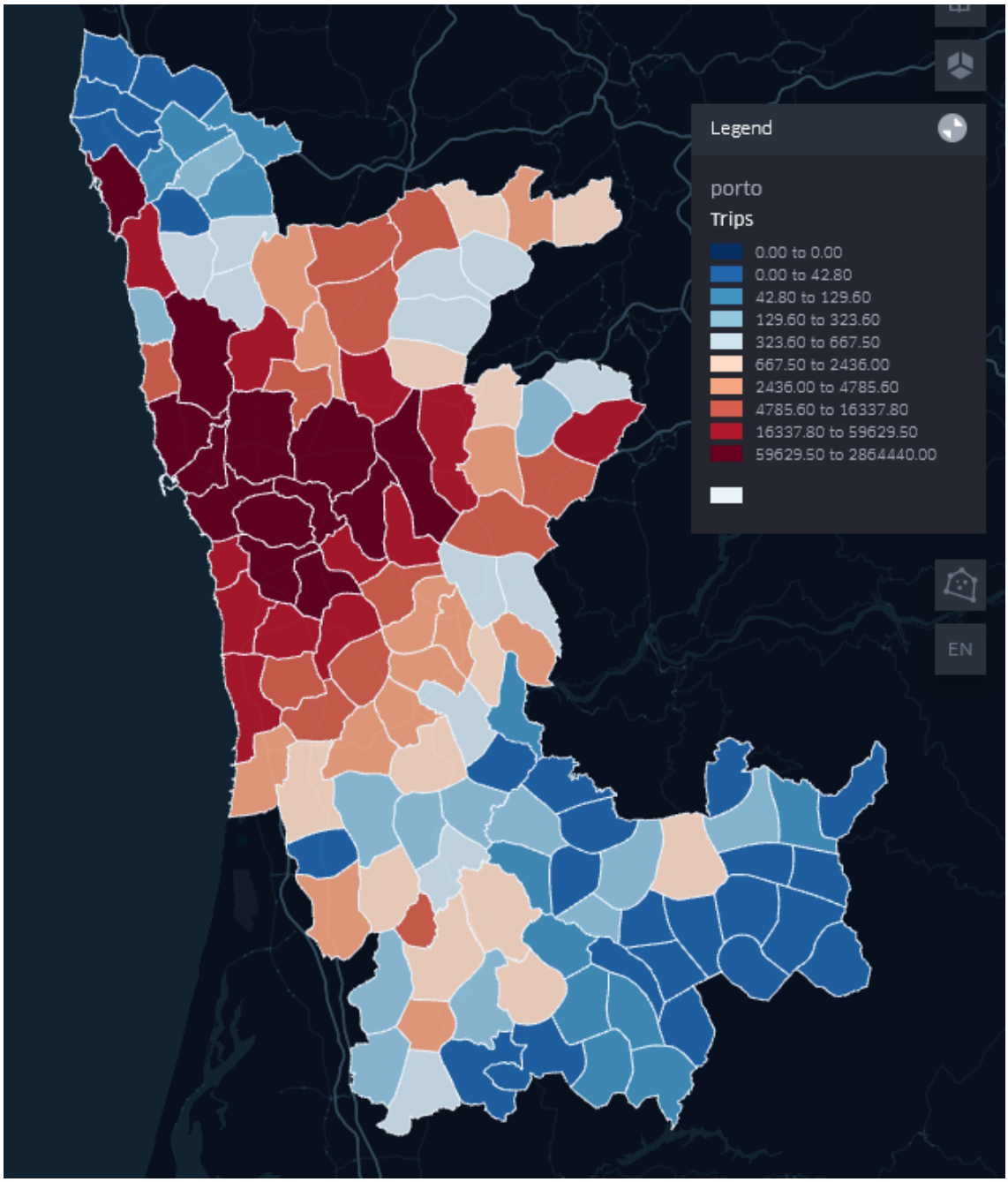

Figure B.3: 2D Color Coded Trip per Zone March 2020 Map

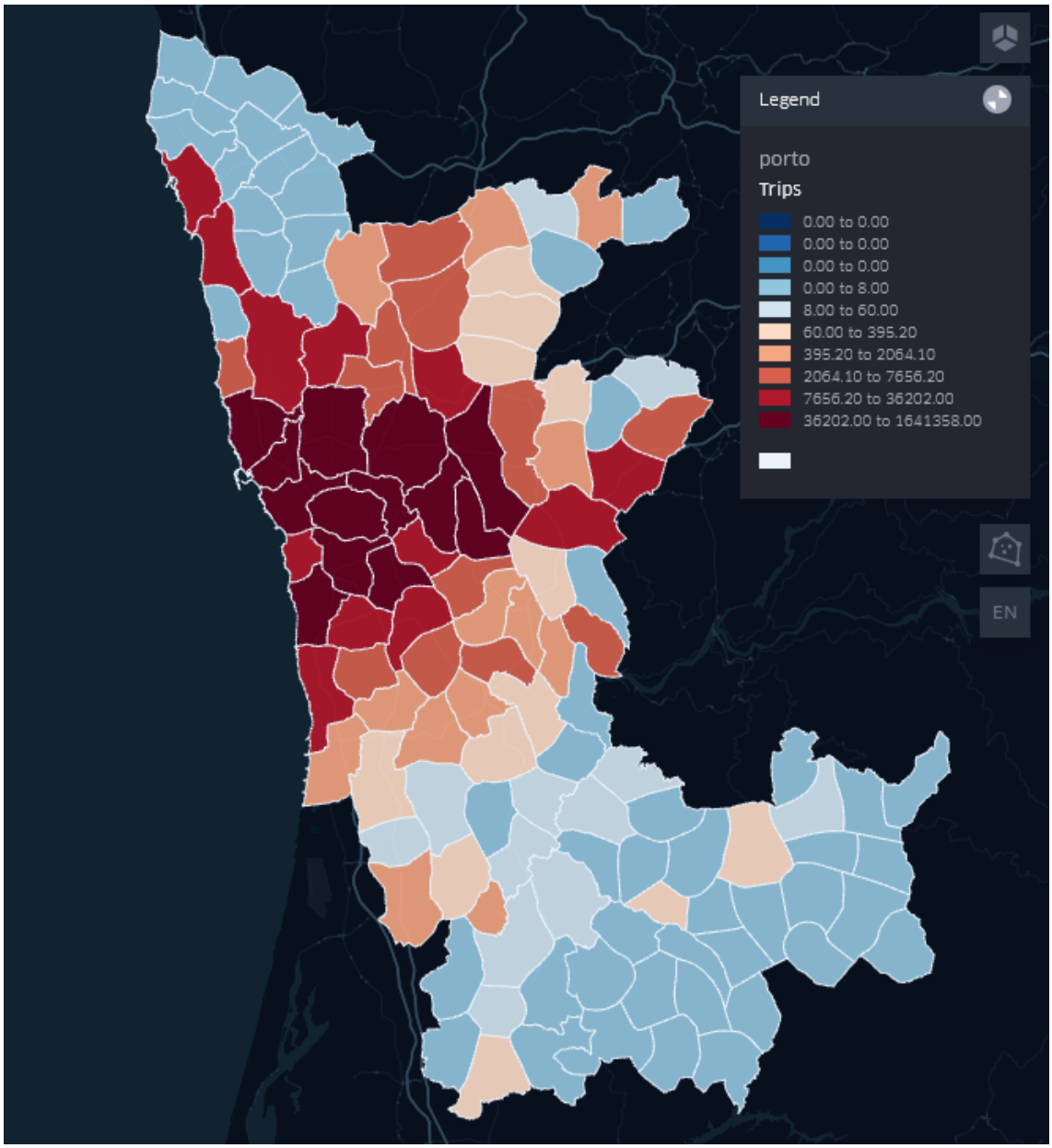

Figure B.4: 2D Color Coded Trip per Zone May 2020 Map

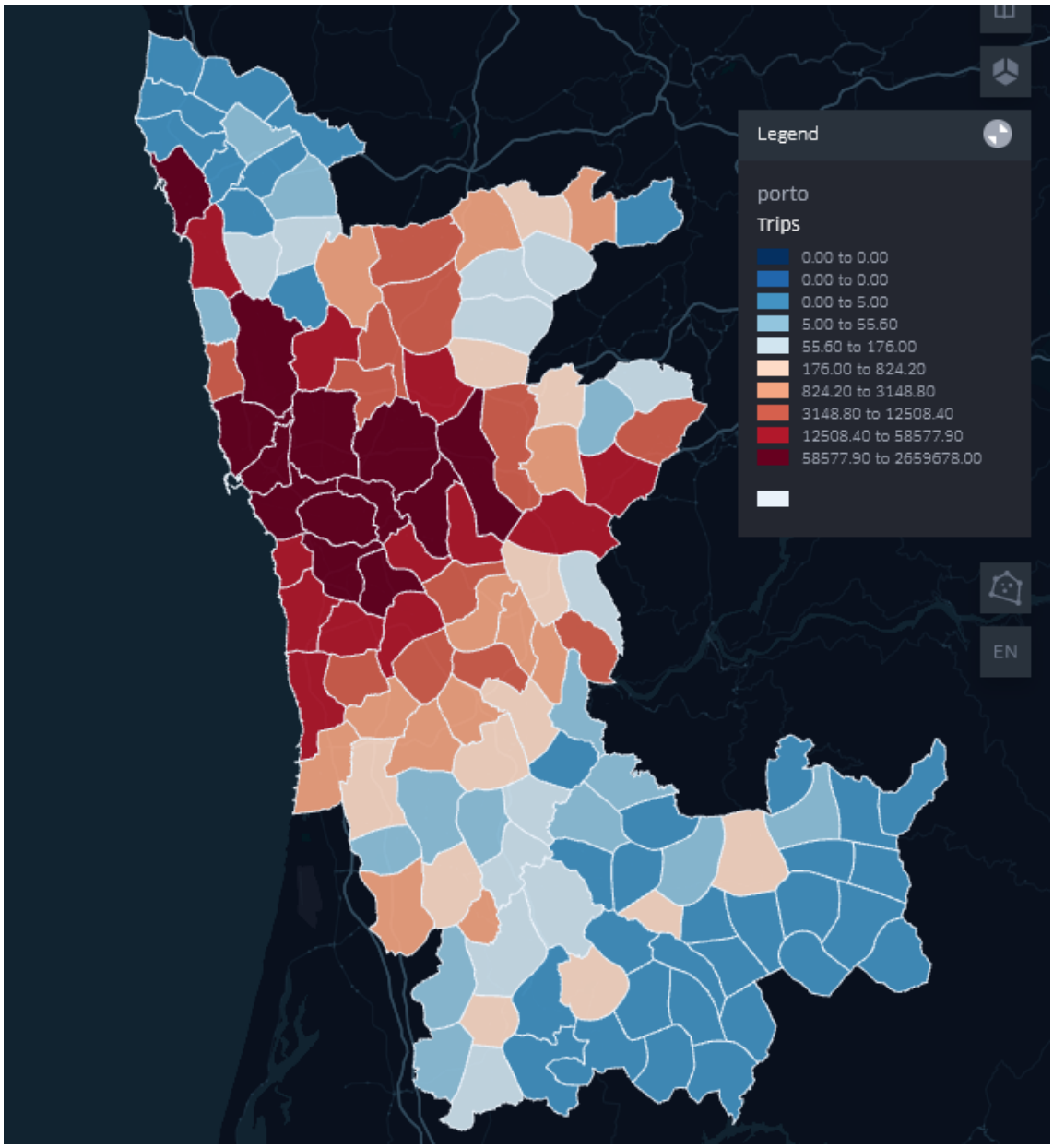

Figure B.5: 2D Color Coded Trip per Zone June 2020 Map

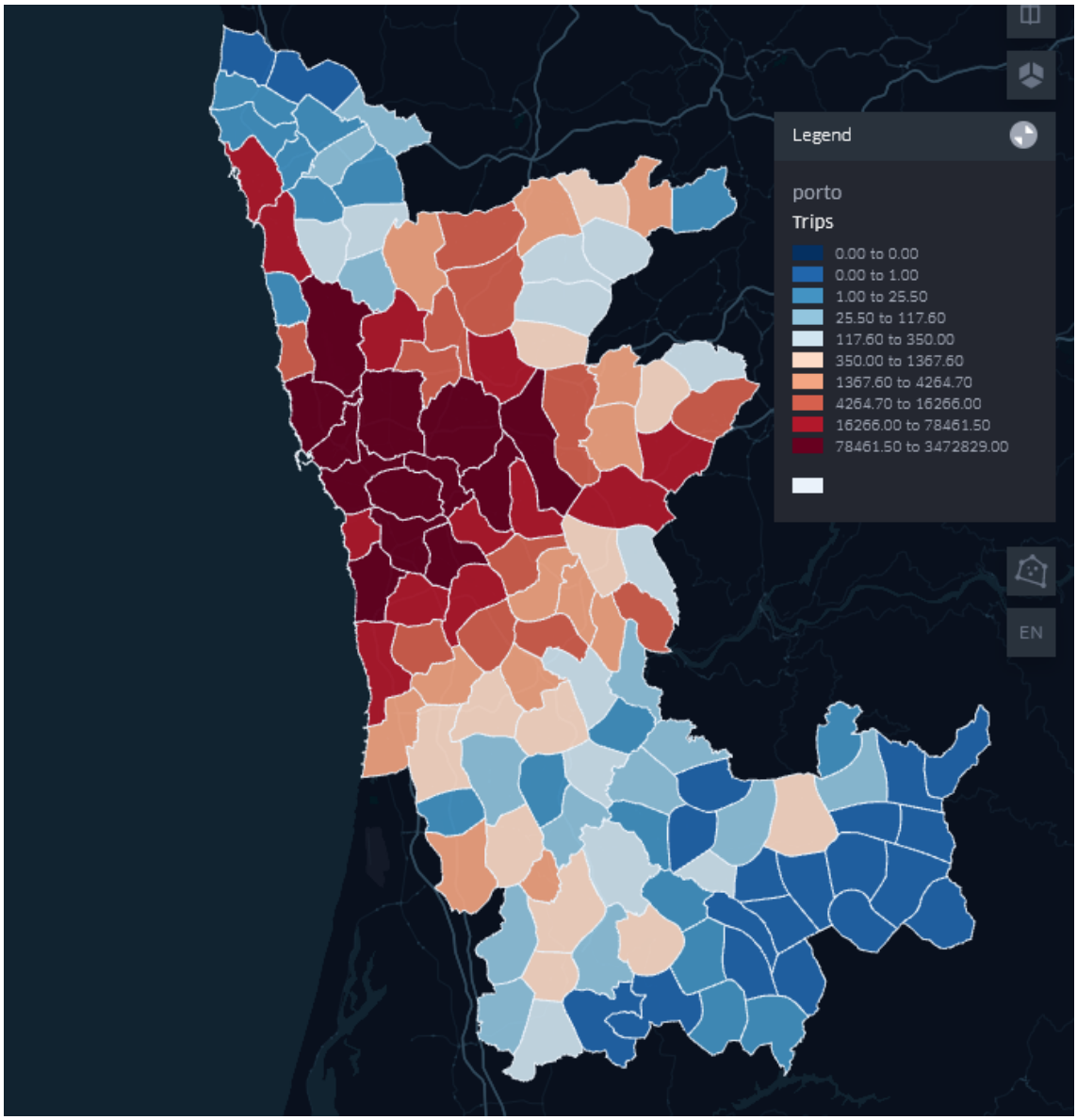

Figure B.6: 2D Color Coded Trip per Zone July 2020 Map

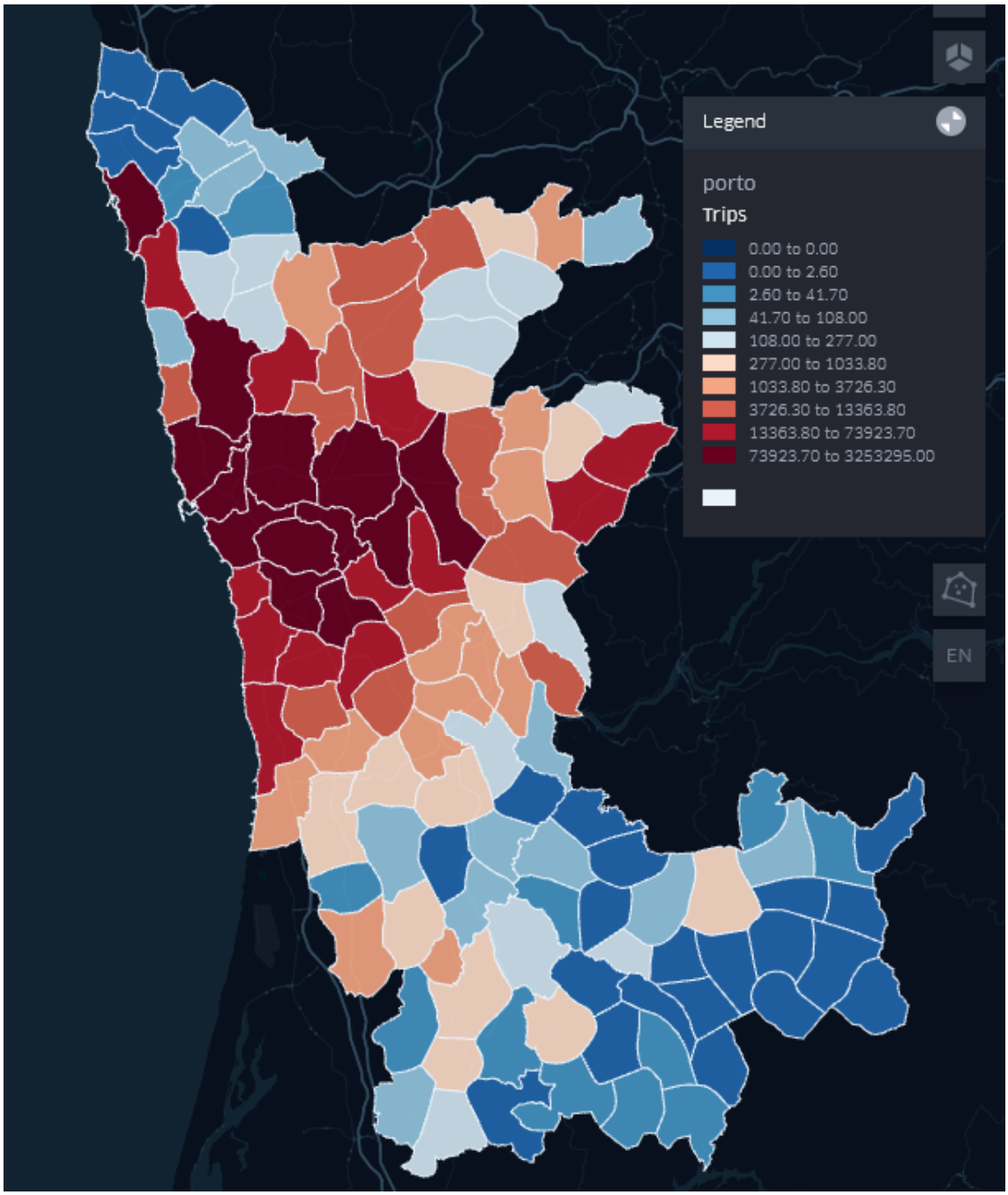

Figure B.7: 2D Color Coded Trip per Zone August 2020 Map

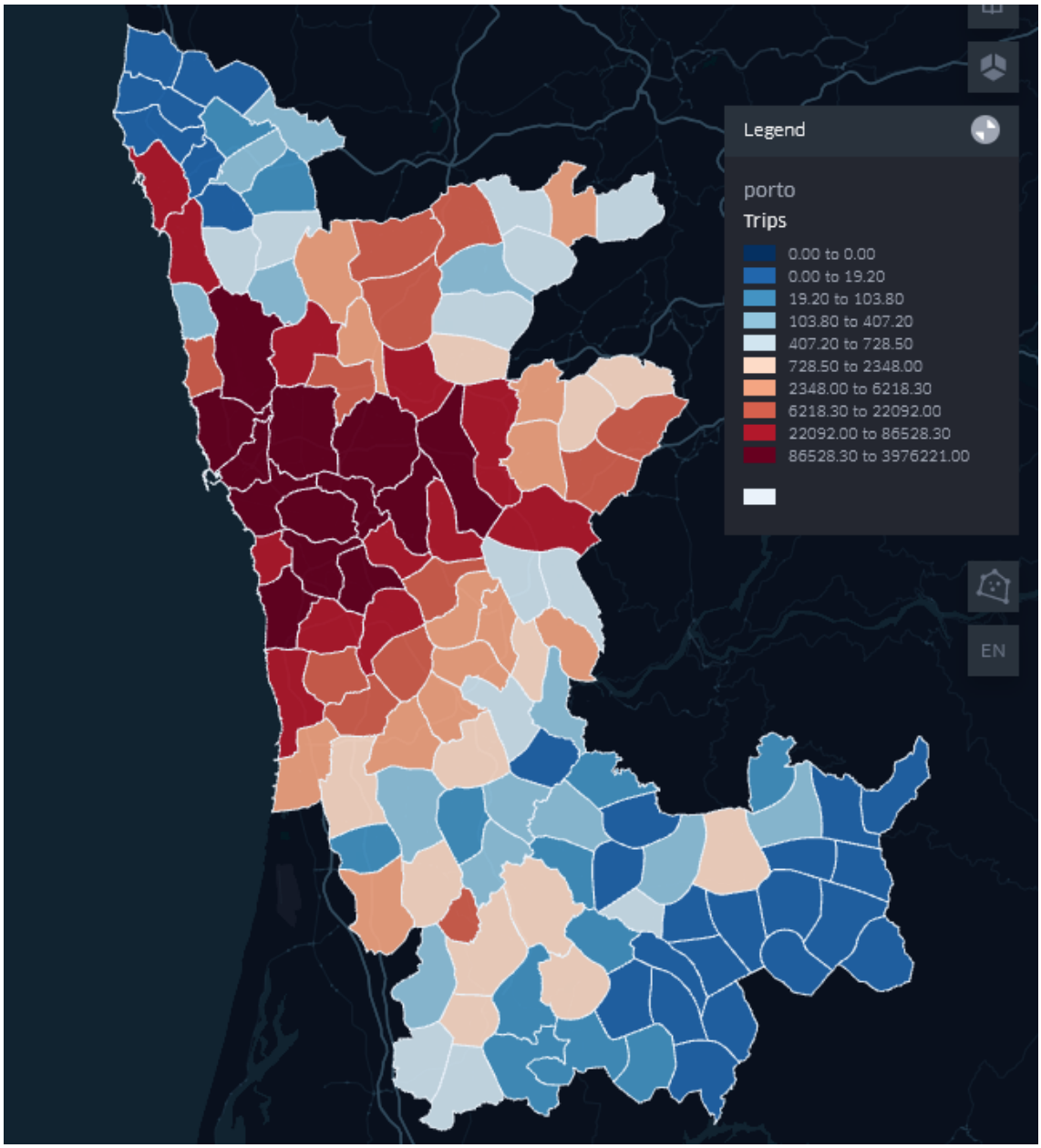

Figure B.8: 2D Color Coded Trip per Zone September 2020 Map

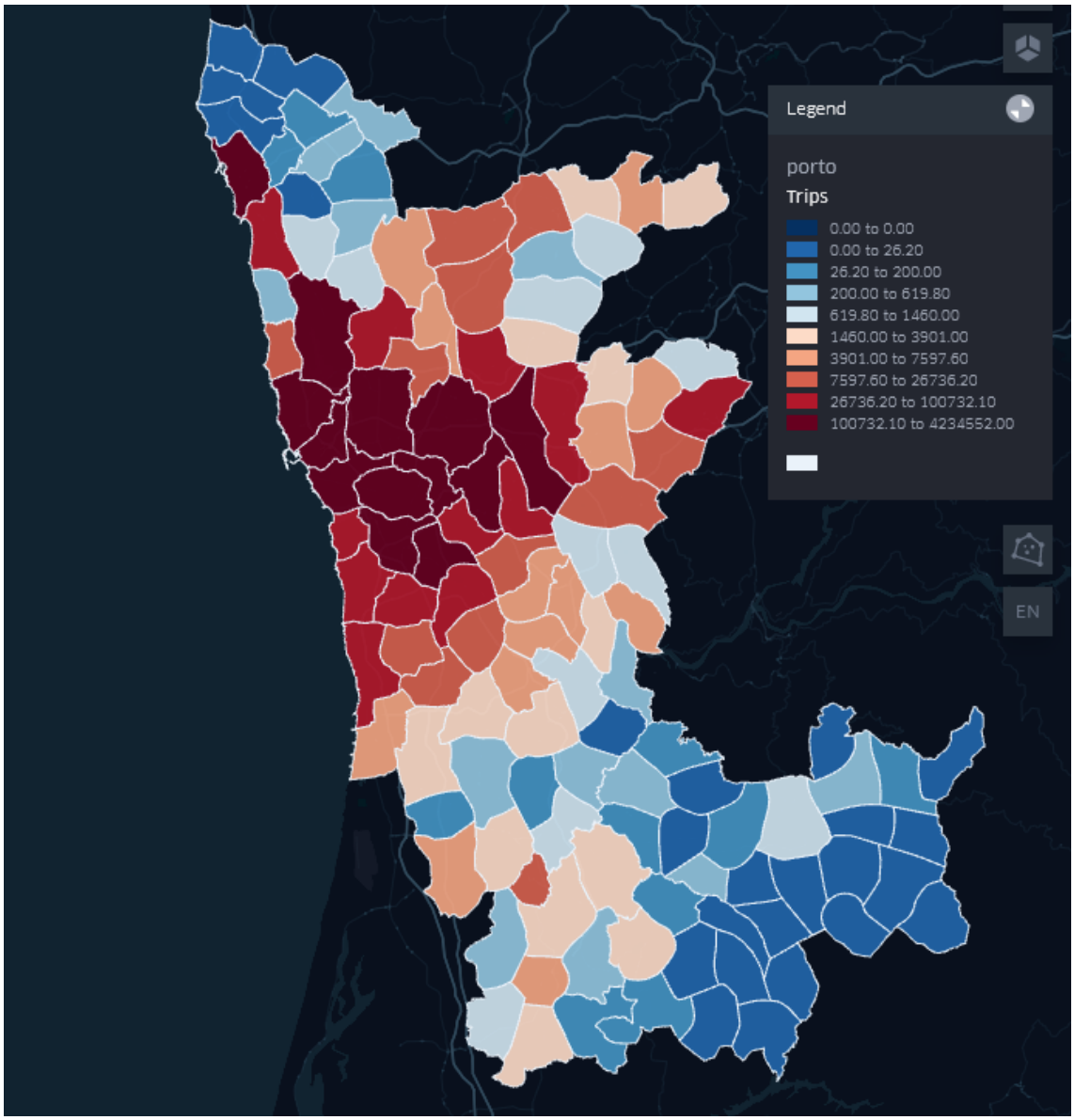

Figure B.9: 2D Color Coded Trip per Zone October 2020 Map

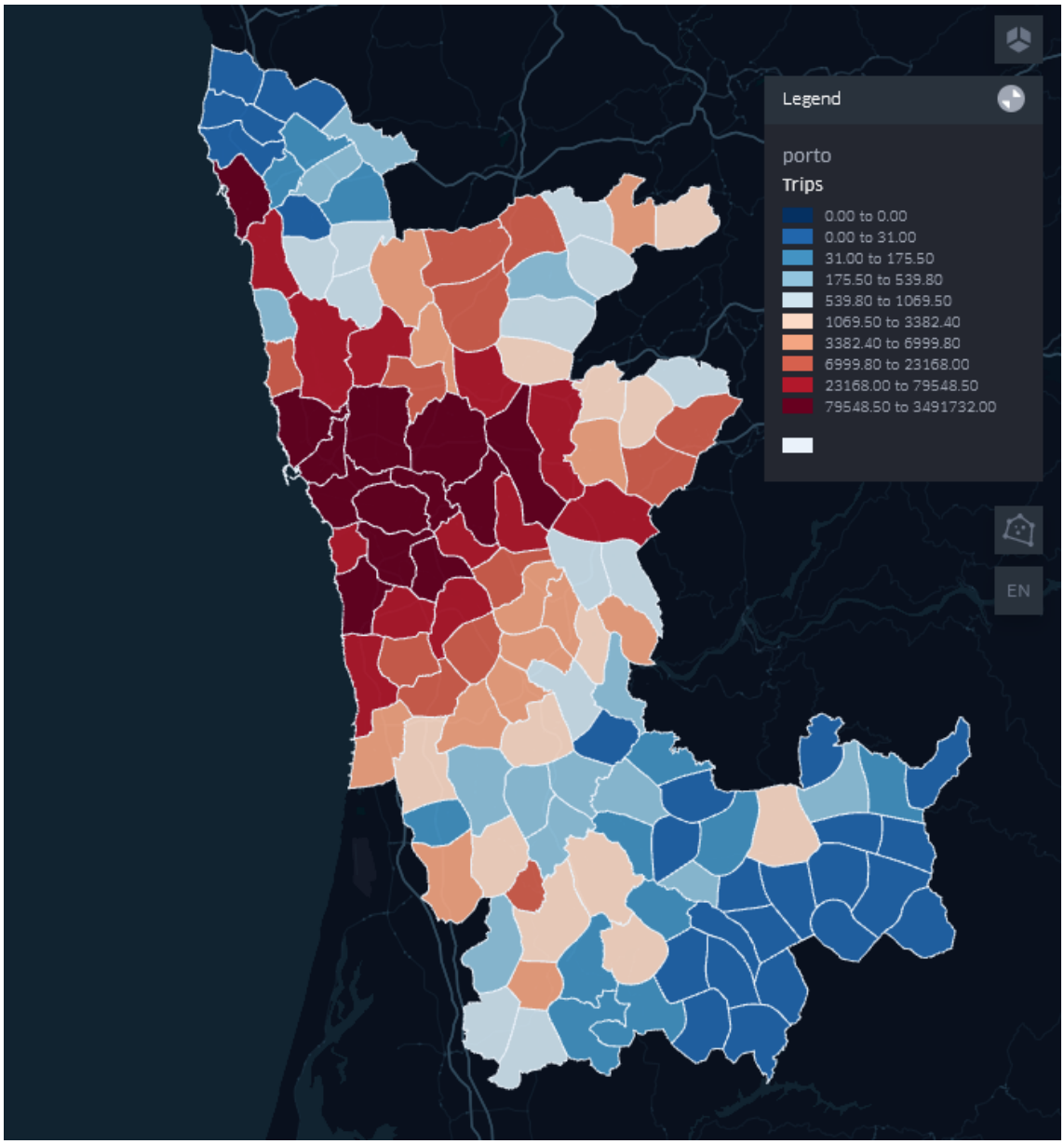

Figure B.10: 2D Color Coded Trip per Zone November 2020 Map

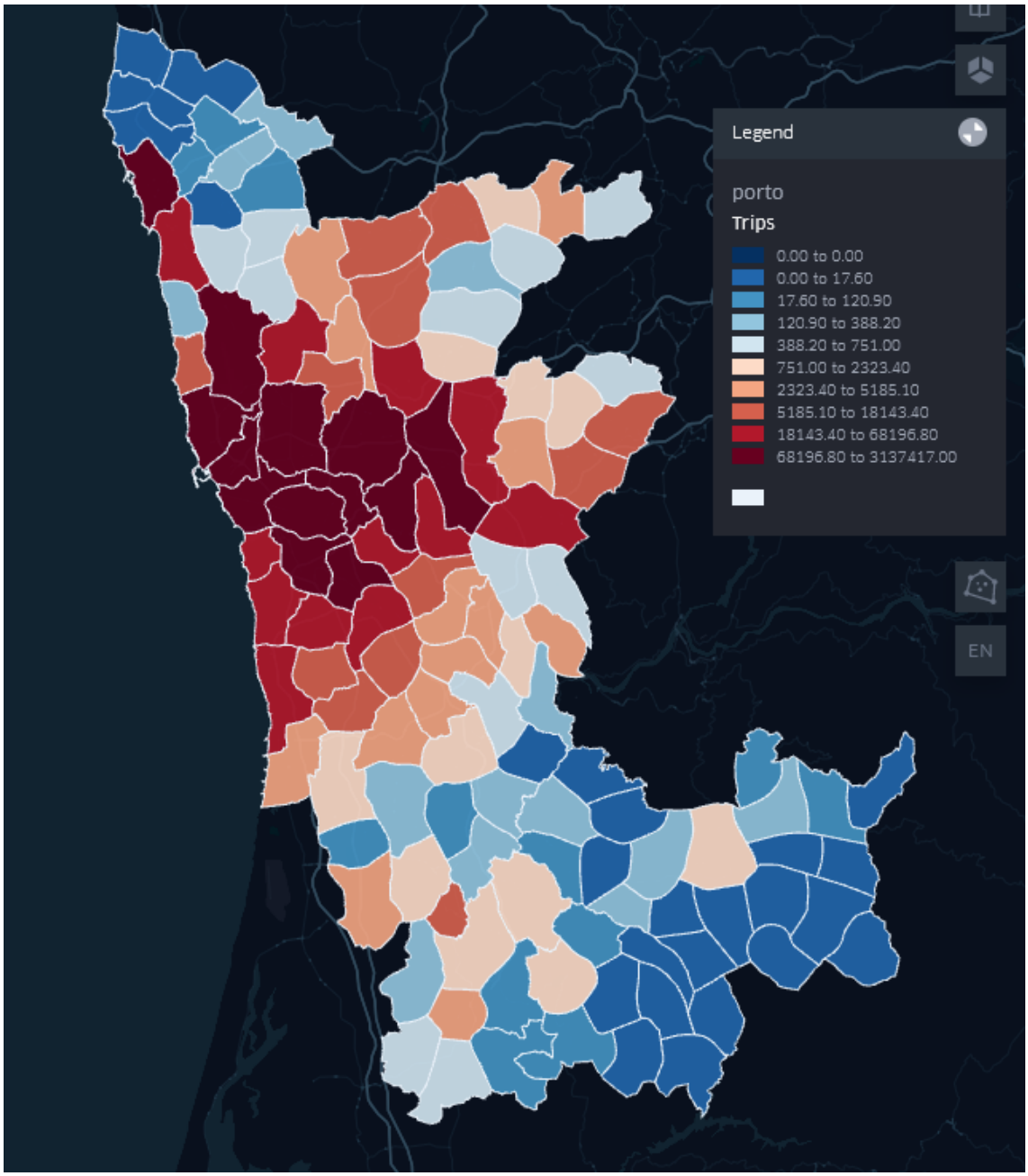

Figure B.11: 2D Color Coded Trip per Zone December 2020 Map