



**JOANA SOFIA RODRIGUES BARRADAS**  
Licenciada em Engenharia Informática

**DETERMINAÇÃO DO IMPACTO DA  
COVID-19 NA MOBILIDADE E QUALIDADE  
DO AR DE LISBOA**

MESTRADO EM ENGENHARIA INFORMÁTICA

Universidade NOVA de Lisboa  
Fevereiro, 2022



# DETERMINAÇÃO DO IMPACTO DA COVID-19 NA MOBILIDADE E QUALIDADE DO AR DE LISBOA

**JOANA SOFIA RODRIGUES BARRADAS**

Licenciada em Engenharia Informática

**Orientador:** João Carlos Gomes Moura Pires  
*Professor Associado, Universidade NOVA de Lisboa*

**Coorientadores:** Carlos Augusto Isaac Piló Viegas Damásio  
*Professor Associado, Universidade NOVA de Lisboa*  
Fernando Pedro Reino da Silva Birra  
*Professor Auxiliar, Universidade NOVA de Lisboa*

## Júri:

**Presidente:** João Baptista da Silva Araújo Júnior  
*Professor Associado, Universidade NOVA de Lisboa*

**Arguente:** Ana Raquel de Ponte Figueiras  
*Professora Auxiliar, Universidade NOVA de Lisboa*

**Orientador:** João Carlos Gomes Moura Pires  
*Professor Associado, Universidade NOVA de Lisboa*

## **Determinação do impacto da COVID-19 na mobilidade e qualidade do ar de Lisboa**

Copyright © Joana Sofia Rodrigues Barradas, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa.

A Faculdade de Ciências e Tecnologia e a Universidade NOVA de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

*Aos meus pais e namorado por todo o apoio incondicional.*

## AGRADECIMENTOS

Em primeiro lugar gostaria de agradecer ao meu orientador, Prof. João Moura Pires, por todo o apoio e disponibilidade ao longo desta dissertação. Agradeço por tudo o que me ensinou e pelas preciosas contribuições transmitidas no decorrer do trabalho desenvolvido.

Aos meus co-orientadores, Prof. Carlos Damásio e Prof. Fernando Birra, gostaria de agradecer por todas as sugestões e ideias transmitidas ao longo da realização do trabalho e, também, por todo o tempo e apoio disponibilizado.

Agradeço também à Câmara Municipal de Lisboa por toda a disponibilidade e colaboração durante a dissertação. Em especial, à Sra. Ana Vasconcelos e ao Sr. Pedro Oliveira, por todo o *feedback* prestado ao longo das reuniões realizadas.

Um grande obrigado à minha faculdade, NOVA School of Science and Technology, por todas as oportunidades que possibilitou ao longo do meu percurso académico. Sendo que este trabalho é apoiado pela NOVA LINCS (UIDB/04516/2020) com o apoio financeiro da FCT.IP, importa ainda agradecer a ambas estas organizações.

Aos meus pais agradeço por todos os esforços feitos que me possibilitaram chegar até aqui. Principalmente, por me terem sempre apoiado e incentivado a nunca desistir de lutar pelo meu futuro. Sem eles nunca teria chegado onde estou hoje, pois foram os valores que me transmitiram que me fizeram ser a pessoa que sou.

Um profundo agradecimento ao meu namorado, Diogo Silva, por ter estado sempre presente nos momentos mais difíceis e por me fazer acreditar constantemente nas minhas capacidades. Estou grata por me ter incentivado a dar sempre o melhor de mim e a nunca me ter deixado desistir. Sem a sua presença e apoio incondicional durante esta fase da minha vida nada teria sido igual.

Por último, mas não menos importante, queria agradecer a todos os meus amigos e colegas que acompanharam o meu percurso académico e que tornaram esta experiência tão gratificante.

## RESUMO

Em 2019 foi identificada, pela primeira vez, a doença infecciosa COVID-19 causada pelo Coronavírus da Síndrome Respiratória Aguda Grave 2 (SARS-CoV-2) em Wuhan, China. O surto desta doença rapidamente deu origem a uma pandemia a nível global, sendo que, no início de março apareceu o primeiro caso em Portugal. O aparecimento e o rápido alastramento desta doença fizeram com que houvesse a necessidade de ordens de confinamento, que tinham como objetivo reduzir a propagação do vírus entre as pessoas. Estas ordens levaram a que se sentisse um impacto na utilização de transportes e na poluição observada nas cidades.

Em Lisboa, após o aparecimento do surto desta doença, sentiu-se também um impacto no que concerne à mobilidade e à qualidade do ar. O principal objetivo desta dissertação é a criação de ferramentas interativas que permitam responder às questões levantadas em relação a esta temática e, ainda, que possibilitem a descoberta de correlações entre os dados. A maioria dos dados utilizados na criação destas ferramentas foram fornecidos pela Câmara Municipal de Lisboa e são relativos, principalmente, a medições de poluentes atmosféricos, ao estado da mobilidade por parte de automóveis e bicicletas e ao estado da meteorologia.

Para efetuar a medição e a caracterização do impacto deste surto em Lisboa foi necessário analisar os dados e verificar a existência de padrões e mudanças de comportamento nos mesmos. Sendo que as análises realizadas necessitaram ter em consideração componentes espaciais, temporais e espaço-temporais. Para além disso, e visto que a pandemia não afetou de igual forma todas as localizações, foi importante analisar não só a cidade como um todo, mas também certas zonas e pontos de interesse de forma mais localizada. Outro fator essencial consistiu na procura de correlação entre as mudanças ocorridas na mobilidade e na qualidade do ar, de modo a ser possível verificar se o tráfego automóvel afetou, de certo modo, a qualidade do ar da cidade de Lisboa.

**Palavras-chave:** COVID-19, Mobilidade, Qualidade do Ar, Visualização Interativa de Dados, Dados Espaciais, Dados Temporais, Análise de Dados

## ABSTRACT

In 2019, the infectious disease COVID-19, caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), was identified for the first time in Wuhan, China. The outbreak of this disease quickly turned into a global pandemic, and in early March the first case appeared in Portugal. The emergence and rapid spread of this disease led to the need for lockdown orders, which aimed to reduce the spread of the virus between people. These orders had an impact on the use of transportation and on the pollution observed in cities.

In Lisbon, after the outbreak of this disease, an impact was also felt in what concerns mobility and air quality. The main goal of this dissertation is the creation of interactive tools that allow answering the questions raised about this topic and also allow the discovery of correlations between data. Most of the data used in the creation of these tools were provided by the Lisbon City Hall and it is mainly related to measurements of air pollutants, the state of mobility by cars and bicycles, and the state of the weather.

In order to measure and characterize the impact of this outbreak in Lisbon it was necessary to analyze the data and verify the existence of patterns and behavioral changes in them. The analysis needed to take into account spatial, temporal and spatio-temporal components. Furthermore, since the pandemic did not affect all locations equally, it was important to analyze not only the city as a whole, but also certain zones and points of interest in a more localized manner. Another essential factor was the search for correlation between changes in mobility and air quality, in order to be able to verify if car traffic affected, in some way, the air quality of the city of Lisbon.

**Keywords:** COVID-19, Mobility, Air Quality, Interactive Data Visualization, Spatial Data, Temporal Data, Data Analysis

# ÍNDICE

<b>Índice de Figuras</b>	<b>xi</b>
<b>Índice de Tabelas</b>	<b>xv</b>
<b>Glossário</b>	<b>xvi</b>
<b>Siglas</b>	<b>xvii</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Contexto e Motivação . . . . .	1
1.2 Descrição do Problema . . . . .	2
1.3 Contributos . . . . .	3
1.4 Organização do Documento . . . . .	4
<b>2 Dados Disponibilizados pela CML e Adquiridos de Outras Fontes</b>	<b>6</b>
2.1 Introdução . . . . .	6
2.2 Conjuntos de Dados . . . . .	8
2.2.1 Estações Gira . . . . .	8
2.2.2 Número de viagens totais das Giras . . . . .	8
2.2.3 <i>Eco-Counter</i> . . . . .	8
2.2.4 Parques de Estacionamento . . . . .	8
2.2.5 Congestionamentos <i>Waze</i> . . . . .	9
2.2.6 NO <sub>2</sub> . . . . .	9
2.2.7 Ciclovias . . . . .	10
2.2.8 Meteorologia . . . . .	10
2.2.9 Poluentes Atmosféricos . . . . .	10
2.2.10 Medidas Governamentais . . . . .	11
2.2.11 Pontos de Interesse . . . . .	11
<b>3 Estado da Arte e Trabalhos Relacionados</b>	<b>12</b>

---

3.1	Visualização Analítica de Dados . . . . .	12
3.1.1	Técnicas de Visualização Temporais . . . . .	14
3.1.2	Técnicas de Visualização Espaciais . . . . .	17
3.1.3	Técnicas de Visualização Espaço-Temporais . . . . .	20
3.1.4	Interações com Visualizações de Dados . . . . .	21
3.2	Trabalhos Relacionados . . . . .	25
3.2.1	Impacto da COVID-19 na mobilidade . . . . .	25
3.2.2	Impacto da COVID-19 na qualidade do ar . . . . .	31
3.2.3	Conclusões . . . . .	36
<b>4</b>	<b>Trabalho Realizado</b> . . . . .	<b>39</b>
4.1	Metodologia Utilizada . . . . .	39
4.2	Ferramentas Utilizadas . . . . .	47
4.2.1	Tableau Desktop . . . . .	47
4.2.2	Tableau Prep Builder . . . . .	49
4.2.3	Tableau Online . . . . .	49
4.2.4	QGIS . . . . .	50
4.2.5	Spyder . . . . .	50
4.2.6	MySQL Workbench . . . . .	51
4.2.7	Geojson.io . . . . .	51
4.3	Processamento dos Dados . . . . .	52
4.3.1	Dados das Estações Gira . . . . .	52
4.3.2	Dados dos Parques de Estacionamento . . . . .	58
4.3.3	Dados dos Congestionamentos <i>Waze</i> . . . . .	61
4.3.4	Dados dos Poluentes Atmosféricos . . . . .	65
4.3.5	Dados da Meteorologia . . . . .	66
4.3.6	Dados das Medidas Governamentais . . . . .	68
4.3.7	Dados dos Pontos de Interesse . . . . .	68
4.3.8	Criação de Novos Conjuntos de Dados . . . . .	69
4.4	<i>Dashboards</i> Desenvolvidos . . . . .	70
4.4.1	Análise às bicicletas Gira da cidade de Lisboa . . . . .	73
4.4.2	Análise aos parques de estacionamento da cidade de Lisboa . . . . .	77
4.4.3	Análise geral dos congestionamentos detetados pelo <i>Waze</i> na cidade de Lisboa . . . . .	80
4.4.4	Análise dos congestionamentos detetados pelo <i>Waze</i> em algumas zonas de Lisboa . . . . .	83
4.4.5	Análise dos congestionamentos detetados pelo <i>Waze</i> em redor de pontos de interesse de Lisboa . . . . .	87
4.4.6	Análise da qualidade do ar da cidade de Lisboa . . . . .	89
4.4.7	Análise da qualidade do ar nas seis estações da cidade de Lisboa . . . . .	93

4.4.8	Comparação entre o comportamento dos poluentes atmosféricos medidos nas estações de qualidade do ar da cidade de Lisboa . . .	95
4.4.9	Procura de relação entre os congestionamentos detetados pelo Waze e a qualidade do ar de Lisboa . . . . .	98
4.5	Conclusões e Análises Feitas nas Ferramentas Desenvolvidas . . . . .	101
4.5.1	Bicicletas Gira . . . . .	103
4.5.2	Parques de estacionamento . . . . .	104
4.5.3	Congestionamentos . . . . .	105
4.5.4	Qualidade do ar . . . . .	108
4.5.5	Congestionamentos vs qualidade do ar . . . . .	111
<b>5</b>	<b>Avaliação do Trabalho Realizado</b>	<b>113</b>
5.1	Metodologia de Avaliação . . . . .	113
5.2	Participantes . . . . .	118
5.3	Resultados . . . . .	119
5.4	Conclusões . . . . .	129
<b>6</b>	<b>Conclusões e Trabalho Futuro</b>	<b>131</b>
	<b>Bibliografia</b>	<b>133</b>
	<b>Apêndices</b>	
<b>A</b>	<b>Lista de <i>Dashboards</i> Desenvolvidos</b>	<b>138</b>
<b>B</b>	<b>Questionário dos testes experimentais</b>	<b>150</b>

## ÍNDICE DE FIGURAS

2.1	Localização das estações Gira, estações meteorológicas, parques de estacionamento, estações de qualidade do ar e sensor <i>eco-counter</i> . . . . .	7
2.2	Ciclovias da cidade de Lisboa. . . . .	10
3.1	Tipos de atributos e direção dos atributos. Extraído de [31]. . . . .	13
3.2	Exemplo de Gráfico de Linhas. Extraído de [45]. . . . .	14
3.3	Exemplo de Gráfico de Barras. Extraído de [10]. . . . .	14
3.4	Exemplo de Mapa de Calor. Extraído de [13]. . . . .	15
3.5	Exemplo de <i>Horizon Graphs</i> . Extraído de [35]. . . . .	16
3.6	Exemplo de Gráfico Espiral. Extraído de [47]. . . . .	16
3.7	Exemplo de Mapa Coroplético. Extraído de [36]. . . . .	18
3.8	Exemplo de Mapa de Pontos. Extraído de [14]. . . . .	18
3.9	Exemplo de Mapa de Densidade de Pontos. Extraído de [1]. . . . .	18
3.10	Exemplo de Mapa de Símbolos Proporcionais. Extraído de [11]. . . . .	19
3.11	Exemplo de Mapa de Densidade. Extraído de [10]. . . . .	19
3.12	Exemplo de Mapa de Fluxo. Extraído de [49]. . . . .	20
3.13	Exemplo de Mapas de Diferença. Extraído de [21]. . . . .	21
3.14	Exemplo de Animação. Extraído de [43]. . . . .	21
3.15	Gráfico de barras que mede a frequência relativa de diferentes tipos transportes utilizados. Extraído de [24]. . . . .	26
3.16	Variação do número de viagens na <i>Citi Bike</i> 3.16(a) e no metro 3.16(b) vs o aumento no número de casos COVID-19, em março. Extraído de [45]. . . . .	27
3.17	Rácio de mobilidade por cada município dos US, em 24 de janeiro de 2020 e 17 de abril de 2020. Extraído de [8]. . . . .	28
3.18	Uso de bicicletas partilhadas e empresas tecnológicas. Extraído de [11]. . . . .	29
3.19	Tendências de mobilidade comparativamente a 2019 na cidade de Nova Iorque. Extraído de [10]. . . . .	30
3.20	Mapa de densidade do número total de viagens para as estações de destino <i>City Bike</i> . Extraído de [10]. . . . .	30

3.21	Interface DatAC. . . . .	31
3.22	<i>Bivariate polar plots</i> que ilustram o efeito conjunto da velocidade e direção do vento sobre o poluente NO <sub>2</sub> observado na estação de Rakah. Extraído de [6].	33
3.23	Diagramas de caixas de bigodes das concentrações médias diárias de NO <sub>2</sub> dentro dos períodos de pré-confinamento, confinamento e pós-confinamento. Extraído de [6]. . . . .	34
3.24	Índice de cores padrão suportado pela Organização Mundial de Saúde (OMS) para os intervalos dos valores do poluente PM2.5. Extraído de [13]. . . . .	35
4.1	Fluxo do modelo de processo <i>Cross-Industry Standard Process for Data Mining</i> (CRISP-DM). . . . .	40
4.2	Principais etapas realizadas em cada fase da metodologia CRISP-DM. . . . .	44
4.3	Avenidas de Lisboa. . . . .	46
4.4	Entradas/Saídas de Lisboa. . . . .	46
4.5	Exemplo de como é calculado os valores de ocupação das estações Gira. . . . .	55
4.6	Exemplo de como é calculado o atributo que quantifica a mudança ocorrida nos estados da pandemia das estações Gira. . . . .	57
4.7	Congestionamentos detetados pelo Waze na cidade de Lisboa. . . . .	62
4.8	<i>Dashboard</i> para a análise às bicicletas Gira da cidade de Lisboa. . . . .	73
4.9	Alteração na visualização V1 após clicar no “Estado de Calamidade” em F1. . . . .	75
4.10	Alteração na visualização V4 após alterar o parâmetro P1 para os dados do <i>eco-counter</i> . . . . .	75
4.11	<i>Dashboard</i> para a análise às bicicletas Gira da cidade de Lisboa - Após clicar no botão para ver análise por períodos temporais específicos. . . . .	76
4.12	<i>Dashboard</i> para a análise aos parques de estacionamento da cidade de Lisboa. . . . .	78
4.13	Exemplo de <i>tooltip</i> do <i>dashboard</i> dos parques de estacionamento. Referente à visualização V2, quando é feito <i>hover</i> no parque do Campo Grande. . . . .	79
4.14	Exemplo de <i>tooltip</i> do <i>dashboard</i> dos parques. Referente à visualização V4, quando é feito <i>hover</i> num dia. . . . .	79
4.15	<i>Dashboard</i> para a análise geral aos congestionamentos detetados pelo Waze na cidade de Lisboa. . . . .	81
4.16	<i>Dashboard</i> para a análise geral aos congestionamentos detetados pelo Waze em algumas zonas da cidade de Lisboa. . . . .	83
4.17	<i>Dashboard</i> para a análise dos congestionamentos detetados pelo Waze numa zona específica da cidade de Lisboa. . . . .	85
4.18	<i>Dashboard</i> para a análise dos congestionamentos detetados pelo Waze ao redor de pontos de interesse da cidade de Lisboa. . . . .	87
4.19	Visualização exibida após a mudança do parâmetro P3 para “Sem nível” no <i>dashboard</i> que analisa os congestionamentos ao redor dos pontos de interesse. . . . .	89
4.20	<i>Dashboard</i> para a análise da qualidade do ar da cidade de Lisboa. . . . .	90

4.21 Exemplo de <i>tooltip</i> do <i>dashboard</i> da qualidade do ar. Referente à visualização V1, quando é feito <i>hover</i> no mês de abril de 2020. . . . .	91
4.22 Exemplo de <i>tooltip</i> do <i>dashboard</i> da qualidade do ar. Apresentado quando é feito <i>hover</i> em V6. . . . .	91
4.23 <i>Dashboard</i> para a análise da qualidade do ar nas seis estações da cidade de Lisboa. . . . .	93
4.24 <i>Dashboard</i> para a comparação entre o comportamento dos poluentes atmosféricos medidos nas estações de qualidade do ar da cidade de Lisboa. . . . .	96
4.25 <i>Dashboard</i> para a procura de relação entre os congestionamentos detetados pelo Waze e a qualidade do ar de Lisboa. . . . .	99
4.26 Exemplo de análise feita no <i>dashboard</i> para a análise às bicicletas Gira da cidade de Lisboa. . . . .	103
4.27 Exemplo de análise feita no <i>dashboard</i> para a análise aos parques de estacionamento da cidade de Lisboa. . . . .	104
4.28 Exemplo de análise feita no <i>dashboard</i> para a análise geral aos congestionamentos detetados pelo Waze na cidade de Lisboa. . . . .	105
4.29 Exemplo de análise feita no <i>dashboard</i> para a análise geral aos congestionamentos detetados pelo Waze em algumas zonas da cidade de Lisboa. . . . .	106
4.30 Exemplo de análise feita no <i>dashboard</i> para a análise dos congestionamentos detetados pelo Waze numa zona específica da cidade de Lisboa. . . . .	107
4.31 Exemplo de análise feita no <i>dashboard</i> para a análise dos congestionamentos detetados pelo Waze ao redor de pontos de interesse da cidade de Lisboa. . . . .	107
4.32 Exemplo de análise feita no <i>dashboard</i> para a análise da qualidade do ar da cidade de Lisboa. . . . .	108
4.33 Exemplo de análise feita no <i>dashboard</i> para a análise da qualidade do ar nas seis estações da cidade de Lisboa. . . . .	109
4.34 Exemplo de análise feita no <i>dashboard</i> para a comparação entre o comportamento dos poluentes atmosféricos medidos nas estações de qualidade do ar da cidade de Lisboa. . . . .	110
4.35 Exemplo de análise feita no <i>dashboard</i> para a procura de relação entre os congestionamentos detetados pelo Waze e a qualidade do ar de Lisboa. . . . .	111
5.1 Resumo das questões utilizadas durante os testes experimentais. . . . .	117
5.2 Distribuição dos participantes por dados demográficos (a) e por experiência com visualização de dados (b). . . . .	119
5.3 Taxa de sucesso de resposta correta por cada questão. . . . .	120
5.4 Tempo médio de resposta a cada pergunta - Ordenado do maior tempo para o menor tempo. . . . .	121
5.5 Distribuição do tempo de resposta dos participantes por cada pergunta de cada <i>dashboard</i> . . . . .	121

5.6	Número de participantes que realizaram o número mínimo de ações de interação, mais ações de interação que o necessário e menos ações de interação que o necessário por questão. . . . .	123
5.7	<i>Scatter Plot</i> com o tempo médio de resposta, média de interações extra por questão e taxa de sucesso por utilizador. . . . .	127
5.8	<i>Dashboard</i> para a análise às bicicletas Gira da cidade de Lisboa - Após melhorias. . . . .	130
5.9	Gráfico com a média móvel diária (V3) do <i>dashboard</i> da qualidade do ar - Após melhorias. . . . .	130
5.10	<i>Heat table</i> (V4) do <i>dashboard</i> dos parques de estacionamento - Após melhorias. . . . .	130
A.1	<i>Dashboard</i> para a análise às bicicletas Gira da cidade de Lisboa - Análise por estados declarados durante a pandemia. . . . .	139
A.2	<i>Dashboard</i> para a análise às bicicletas Gira da cidade de Lisboa - Análise por períodos temporais específicos. . . . .	140
A.3	<i>Dashboard</i> para a análise aos parques de estacionamento da cidade de Lisboa. . . . .	141
A.4	<i>Dashboard</i> para a análise geral aos congestionamentos detetados pelo <i>Waze</i> na cidade de Lisboa. . . . .	142
A.5	<i>Dashboard</i> para a análise geral aos congestionamentos detetados pelo <i>Waze</i> em algumas zonas da cidade de Lisboa. . . . .	143
A.6	<i>Dashboard</i> para a análise dos congestionamentos detetados pelo <i>Waze</i> numa zona específica da cidade de Lisboa. . . . .	144
A.7	<i>Dashboard</i> para a análise dos congestionamentos detetados pelo <i>Waze</i> ao redor de pontos de interesse da cidade de Lisboa. . . . .	145
A.8	<i>Dashboard</i> para a análise da qualidade do ar da cidade de Lisboa. . . . .	146
A.9	<i>Dashboard</i> para a análise da qualidade do ar nas seis estações da cidade de Lisboa. . . . .	147
A.10	<i>Dashboard</i> para a comparação entre o comportamento dos poluentes atmosféricos medidos nas estações de qualidade do ar da cidade de Lisboa. . . . .	148
A.11	<i>Dashboard</i> para a procura de relação entre os congestionamentos detetados pelo <i>Waze</i> e a qualidade do ar de Lisboa. . . . .	149

## ÍNDICE DE TABELAS

2.1	Conjuntos de dados e onde foram adquiridos. . . . .	7
3.1	Oito variáveis visuais. . . . .	14
3.2	Taxonomia de técnicas de interação. Extraído de [20]. . . . .	22
3.3	Taxonomia de interação de visualizações espaço-temporais. Extraído de [39]. . . . .	24
4.1	Colunas iniciais dos dados das Estações Gira. . . . .	52
4.2	Principais colunas dos dados das Estações Gira após processamento. . . . .	58
4.3	Colunas iniciais dos dados dos Parques de Estacionamento. . . . .	58
4.4	Principais colunas dos dados dos Parques de Estacionamento após processamento. . . . .	60
4.5	Colunas iniciais dos dados dos congestionamentos <i>Waze</i> . . . . .	61
4.6	Principais colunas dos dados dos congestionamentos <i>Waze</i> após processamento. . . . .	64
4.7	Principais colunas dos dados dos poluentes atmosféricos após processamento. . . . .	65
4.8	Principais colunas dos dados dos poluentes atmosféricos, após processamento. Conjunto de dados utilizado para junção com os dados do <i>Waze</i> . . . . .	66
4.9	Colunas iniciais dos dados da Meteorologia. . . . .	66
4.10	Colunas dos dados da Meteorologia Global após processamento. . . . .	67
4.11	Colunas dos dados das Medidas Governamentais após processamento. . . . .	68
4.12	Colunas dos dados dos Pontos de Interesse. . . . .	69
4.13	Colunas dos dados do Sinal do Tráfego ao Redor dos POI. . . . .	70
4.14	<i>Dashboards</i> agrupados segundo a sua temática principal. . . . .	71
4.15	Conjuntos de dados utilizados em cada <i>dashboard</i> . . . . .	72
4.16	Classes usadas para cada parâmetro meteorológico. . . . .	98
4.17	Algumas das interações mais gerais presentes em cada <i>dashboard</i> . . . . .	102

## GLOSSÁRIO

- estações de tráfego** As estações de tráfego situam-se próximas de vias de tráfego intenso e permitem avaliar o risco máximo de exposição da população às emissões do tráfego automóvel. Esta exposição é, por norma, de curta duração, mas os níveis de poluição observados são normalmente elevados. As estações de tráfego presentes em Lisboa são as estações da Avenida da Liberdade, de Santa Cruz de Benfica e de Entrecampos [7].
- estações de fundo** As estações de fundo não se encontram sob a influência direta de vias de tráfego nem de qualquer fonte próxima de poluição. Estas estações permitem conhecer a exposição média da população aos fenómenos de poluição de fundo. As estações de fundo presentes em Lisboa são as estações do Beato, do Restelo e de Olivais [7].

## SIGLAS

<b>BSS</b>	<i>Bike Sharing System</i>
<b>CML</b>	Câmara Municipal de Lisboa
<b>CO</b>	Monóxido de Carbono
<b>CRISP-DM</b>	<i>Cross-Industry Standard Process for Data Mining</i>
<b>FCT NOVA</b>	Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa
<b>IPMA</b>	Instituto Português do Mar e da Atmosfera
<b>LxDataLab</b>	Laboratório de Dados Urbanos de Lisboa
<b>MTA</b>	<i>Metropolitan Transportation Authority</i>
<b>NO<sub>2</sub></b>	Dióxido de Azoto
<b>O<sub>3</sub></b>	Ozono
<b>OMS</b>	Organização Mundial de Saúde
<b>PGIL</b>	Plataforma de Gestão Inteligente de Lisboa
<b>PM<sub>10</sub></b>	Partículas com diâmetro inferior a 10 $\mu\text{m}$
<b>PM<sub>2.5</sub></b>	Partículas com diâmetro inferior a 2.5 $\mu\text{m}$
<b>POI</b>	Ponto de Interesse
<b>SARS-CoV-2</b>	Coronavírus da Síndrome Respiratória Aguda Grave 2
<b>SO<sub>2</sub></b>	Dióxido de Enxofre



# INTRODUÇÃO

COVID-19 é uma doença infecciosa causada pelo [Coronavírus da Síndrome Respiratória Aguda Grave 2 \(SARS-CoV-2\)](#). Esta doença foi identificada pela primeira vez em dezembro de 2019, num grupo de pacientes admitidos num hospital na cidade de Wuhan, China [40]. A propagação mundial desta doença fez-se a uma velocidade alarmante que, no dia 11 de março de 2020, a [OMS](#) declarou o surto de COVID-19 uma pandemia [34]. No dia 2 de março de 2020 foram confirmados em Portugal os dois primeiros casos desta doença [29] e, devido ao avanço significativo do número de casos, no dia 18 de março de 2020, foi declarado estado de emergência em Portugal [37].

O surto desta doença não afetou apenas a saúde e a economia, mas também expandiu os seus efeitos noutros aspetos, tais como o ambiente e a mobilidade urbana. As formas de deslocação por parte dos cidadãos mudaram, e isto deveu-se, principalmente, à alteração significativa do estilo de vida e do dia a dia de cada pessoa. Todas as mudanças ocorridas na mobilidade urbana acarretaram grandes benefícios no ambiente, essencialmente, na qualidade do ar.

## 1.1 Contexto e Motivação

Esta dissertação provém de uma colaboração entre a [Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa \(FCT NOVA\)](#) e a [Câmara Municipal de Lisboa \(CML\)](#), no âmbito do [Laboratório de Dados Urbanos de Lisboa \(LxDataLab\)](#). O LxDataLab foi criado com o objetivo de extrair valor da informação disponível da cidade de Lisboa. Este foca-se, em grande parte, na criação de soluções analíticas e de visualizações de dados que permitam melhorar o planeamento e a gestão operacional da cidade [26]. De forma a alcançar estes objetivos, foram lançados vários desafios sobre diferentes assuntos, sendo o tema desta dissertação um dos desafios propostos.

A principal motivação desta dissertação centra-se na avaliação do impacto que a COVID-19 teve na mobilidade urbana e na qualidade do ar da cidade de Lisboa. Isto ajuda a compreender quais foram os efeitos da pandemia nos meios de transporte considerados modos suaves, isto é, nas deslocações por meio de bicicletas. É ainda importante

perceber de que forma o surto desta doença afetou o tráfego geral e localizado da cidade de Lisboa e, também, a utilização dos parques de estacionamento desta cidade. Para além disso, é essencial entender quais foram os efeitos no que diz respeito à qualidade do ar, tentando constatar se existem alterações significativas em comparação com bases temporais anteriores a esta pandemia. Deste modo, o objetivo principal desta dissertação foca-se no desenvolvimento de ferramentas visuais interativas que permitam realizar este tipo de análises. No entanto, houve a preocupação que as ferramentas pudessem não só responder a questões relativas à pandemia, mas também mostrar um panorama geral da cidade de Lisboa.

### 1.2 Descrição do Problema

Como referido na secção 1.1, a motivação desta dissertação é, a partir dos dados fornecidos, criar ferramentas interativas que permitam analisar qual foi o impacto da pandemia no que diz respeito à mobilidade urbana e à poluição atmosférica da cidade de Lisboa.

Segundo os dados de 2019 do Parlamento Europeu, os transportes são responsáveis por quase 30% das emissões de dióxido de carbono na União Europeia, dos quais 72% provêm de transportes rodoviários [15]. De acordo com o *Emissions Gap Report 2019*, para limitar o aumento da temperatura global a 1.5 graus celsius é necessário reduzir 7.6% das emissões de gases com efeito de estufa por ano [16] entre 2020 e 2030. Devido à COVID-19, este valor pode, muito possivelmente, ter sido alcançado [33].

Com o aparecimento desta pandemia houve ainda um acréscimo na preocupação de estar longe de grandes aglomerados de pessoas. Para além disso, e devido às restrições que foram implementadas, os cidadãos sofreram uma grande alteração do seu dia a dia, havendo menos necessidade de saírem de casa todos os dias, o que, conseqüentemente, reduziu o uso dos automóveis.

Os dados que foram utilizados na realização desta dissertação foram cedidos pela [CML](#) e foram também obtidos, através de outras fontes, mais dados que são igualmente relevantes e que contribuem para uma melhor caracterização do impacto desta pandemia. Os dados fornecidos pela [CML](#) são referentes ao estado da ocupação das estações Gira, às viagens efetuadas diariamente na rede GIRA, às contagens diárias de entradas de bicicletas na Avenida Duque de Ávila (Saldanha), a dados geoespaciais das ciclovias existentes na cidade de Lisboa, ao estado da ocupação de um conjunto de parques de estacionamento da [CML](#), ao estado dos congestionamentos consoante o *Waze*, a informação meteorológica e aos valores medidos do dióxido de nitrogénio em Lisboa. Os dados obtidos posteriormente dizem respeito a medições das concentrações de outros poluentes atmosféricos, a dados relativos às medidas governamentais impostas durante a pandemia e a informação relativa a alguns pontos de interesse da cidade de Lisboa.

Através destes dados foram criadas ferramentas interativas que permitem um melhor entendimento e comunicação das informações retiradas dos dados. Estas possibilitam,

essencialmente, perceber de que forma a pandemia afetou a mobilidade urbana e a poluição atmosférica, sendo isto realizado através da caracterização e medição do impacto do confinamento e das restantes restrições. No fim, foram ainda criadas duas ferramentas que permitem elaborar algumas correlações entre os dados, nomeadamente, entre os diversos poluentes atmosféricos e entre os dados da qualidade do ar com os dos congestionamentos.

Nesta dissertação o interesse está em analisar as diferenças ambientais e de mobilidade que há na época pré-pandemia comparativamente à época durante a pandemia. E, através da análise destas diferenças, conseguir extrair conclusões importantes para o melhoramento da cidade de Lisboa. É importante salientar que nesta dissertação não existe um panorama completo dos modos suaves, pois podem existir pessoas que utilizam bicicletas individuais e não partilhadas, sendo que neste momento não existe maneira de contabilizar isso. E, para além disso, existem também outros operadores de bicicletas partilhadas e de trotinetes. Outro aspeto a salientar é que o foco desta dissertação passou pela criação de visualizações que permitam observar padrões e relações entre os dados, e não pela análise explicativa dos fenómenos.

### 1.3 Contributos

Tal como referido anteriormente, o objetivo desta dissertação é o desenvolvimento de ferramentas visuais e interativas que permitam determinar o impacto que a COVID-19 teve na mobilidade e na qualidade do ar da cidade de Lisboa. Assim, os contributos desta dissertação são os seguintes:

- **Relatório de qualidade dos dados:** O uso de dados tem um forte peso nesta dissertação então é necessário usar abordagens estruturadas para processos de *Data Mining* de forma a adquirir um maior conhecimento do domínio. Assim, com base na fase de *Data understanding* do [CRISP-DM](#), elaborou-se um relatório de qualidade dos dados que abrange cada conjunto de dados disponíveis. Este relatório beneficiará a tomada de decisões em análises futuras e pode ser consultado neste [documento](#).
- **Dados com melhor qualidade:** Após a análise da qualidade inerente a cada conjunto de dados, seguiu-se para o processamento de cada um deles de modo a melhorar a sua qualidade e a corrigir alguns problemas que tenham sido detetados na fase anterior. Este processamento fez com que os dados utilizados no desenvolvimento das ferramentas apresentassem uma melhor qualidade. Visto que os conjuntos de dados não são de domínio público, foi partilhado apenas com a [CML](#) uma pasta que contém todos os conjuntos de dados após terem sido processados.
- **Conjunto de questões analíticas:** Um dos resultados deste trabalho foi a licitação de um conjunto de questões analíticas. Estes grupos de questões foram estabelecidos mediante um processo iterativo sobre quais são as verdadeiras preocupações e

requisitos por parte da [CML](#).

- **Ferramentas para a análise e determinação do impacto da COVID-19 na mobilidade urbana:** Foi desenvolvido um conjunto de ferramentas visuais, sob o formato de *dashboards*, que permitem a análise e a determinação do impacto que a pandemia teve na mobilidade da cidade de Lisboa. A mobilidade urbana é avaliada tendo em consideração a utilização das bicicletas Gira, a utilização dos parques de estacionamento e os congestionamentos ocorridos em Lisboa. Estas ferramentas encontram-se disponíveis no Tableau Online para o uso por parte da [CML](#).
- **Ferramentas para a análise e determinação do impacto da COVID-19 na qualidade do ar:** Foi desenvolvido um conjunto de *dashboards* que permitem avaliar e determinar qual o impacto que a pandemia teve na qualidade do ar da cidade de Lisboa. Para os efeitos de análise foram considerados os poluentes NO<sub>2</sub>, O<sub>3</sub>, PM<sub>10</sub>, PM<sub>2.5</sub> e SO<sub>2</sub> medidos durante os anos 2017 até 2020. Estas ferramentas também se encontram disponíveis no Tableau Online para o uso por parte da [CML](#).
- **Ferramentas para a procura de relações e comparações entre os dados:** Após as ferramentas principais de análise estarem desenvolvidas, foi essencial a criação de ferramentas que permitissem a procura de relações entre os dados. Deste modo, foi desenvolvido um *dashboard* que possibilita a comparação entre o comportamento dos diversos poluentes atmosféricos e um *dashboard* que permite a procura de relação entre os congestionamentos detetados pelo *Waze* e a qualidade do ar da cidade de Lisboa. Estas ferramentas também se encontram disponíveis no Tableau Online para o uso por parte da [CML](#).
- **Disponibilização das ferramentas criadas para o *workshop* da Rede C40:** A partir de 2019 a cidade de Lisboa passou a integrar a rede C40, ou seja, o Grupo de Liderança Climática das Grandes Cidades, que é a mais importante rede de combate às alterações climáticas [27]. No dia 30 de novembro de 2021 foi realizado um *workshop* da rede C40 no qual os representantes da Câmara Municipal de Lisboa fizeram uso de algumas das ferramentas desenvolvidas durante a sua apresentação.

### 1.4 Organização do Documento

Este documento está dividido em seis capítulos que contêm os seguintes conteúdos:

- **Introdução** - No capítulo 1 são introduzidos o contexto e motivação do trabalho e também uma descrição geral do problema. Este capítulo apresenta ainda uma secção com as principais contribuições desta dissertação.
- **Dados Disponibilizados pela CML e Adquiridos de Outras Fontes** - Os dados disponibilizados pela Câmara Municipal de Lisboa e os dados adquiridos de outras

fontes que são utilizados no contexto desta dissertação são apresentados no capítulo 2. Cada conjunto de dados é apresentado e descrito brevemente, tendo em conta a sua descrição mais detalhada presente relatório de qualidade de dados. Este relatório encontra-se neste [documento](#), servindo de apêndice a esta dissertação.

- **Estado da Arte e Trabalhos Relacionados** - O capítulo 3 apresenta um resumo da pesquisa efetuada durante a dissertação através de um levantamento do estado da arte. É feita uma breve introdução à visualização analítica de dados e são descritas sucintamente algumas das principais técnicas existentes para a visualização de dados e algumas das principais interações utilizadas em visualizações. Neste capítulo são também apresentados trabalhos que analisam o impacto que a pandemia por COVID-19 teve na mobilidade e na qualidade do ar.
- **Trabalho Realizado** - O capítulo 4 apresenta todo o trabalho realizado no decorrer da dissertação. Começa inicialmente por apresentar a metodologia que foi seguida na realização do trabalho. Seguidamente, são descritas todas as ferramentas utilizadas, todo o processamento feito aos dados e todos os *dashboards* criados. Por fim, são mostradas algumas análises e conclusões retiradas dos *dashboards* desenvolvidos.
- **Avaliação do Trabalho Realizado** - O capítulo 5 apresenta a avaliação feita às ferramentas desenvolvidas. Neste capítulo é descrita qual foi a metodologia seguida no processo de avaliação e quais foram os resultados e as conclusões obtidas dos testes experimentais.
- **Conclusões e Trabalho Futuro** - No capítulo 6 são apresentadas as conclusões do trabalho desenvolvido, bem como qual poderá ser o trabalho futuro tendo em conta o contexto desta dissertação.

## DADOS DISPONIBILIZADOS PELA CML E ADQUIRIDOS DE OUTRAS FONTES

Este capítulo apresenta uma pequena síntese e aspetos importantes sobre os conjuntos de dados fornecidos pela Câmara Municipal de Lisboa e sobre os que foram adquiridos de outras fontes. Primeiramente, na secção 2.1, é feita uma breve descrição de quais são os dados disponíveis. De seguida, na secção 2.2, são apresentados sucintamente os diferentes conjuntos de dados.

De notar que, neste capítulo, são descritos os dados na sua forma inicial, ou seja, sem qualquer tipo de processamento ou manipulação. No capítulo 4 será detalhado todo o processamento feito nestes conjuntos de dados.

### 2.1 Introdução

Os conjuntos de dados fundamentais para a realização desta dissertação foram disponibilizados pela Câmara Municipal de Lisboa. Estes dados são relativos ao estado de cada estação Gira, ao número de viagens totais feitas nas bicicletas Gira, ao número de bicicletas que circularam na ciclovia da Avenida Duque de Ávila, a dados geoespaciais das ciclovias existentes em Lisboa, ao estado dos parques de estacionamento de Lisboa, ao estado dos congestionamentos segundo o *Waze*, aos valores medidos de **Dióxido de Azoto (NO<sub>2</sub>)** e ao estado de parâmetros meteorológicos segundo o **Instituto Português do Mar e da Atmosfera (IPMA)**. De referir ainda que, todos eles, são alusivos aos anos de 2019 e 2020.

Ao longo da realização da dissertação foram ainda incluídos mais conjuntos de dados inerentes à cidade de Lisboa. Nomeadamente, dados relativos a outros poluentes atmosféricos, que englobam os anos desde 2017 até 2020, dados relativos às medidas governamentais impostas durante a pandemia e informação acerca dos pontos de interesse da cidade de Lisboa.

A tabela 2.1 sintetiza quais os conjuntos de dados fornecidos pela **CML** e quais foram obtidos, posteriormente, por outras fontes. Sendo que, todos serão brevemente detalhados na secção 2.2. A figura 2.1 mostra os dados espaciais disponíveis que englobam os

conjuntos de dados das estações Gira, dos parques de estacionamento, da meteorologia, dos poluentes atmosféricos e do sensor *eco-counter*.

Para a realização desta dissertação é essencial conhecer e perceber os dados que irão ser usados para responder ao problema existente. Com este propósito, foi elaborado um relatório de qualidade dos dados, presente neste [documento](#), no qual é analisada a qualidade de cada conjunto de dados.

Tabela 2.1: Conjuntos de dados e onde foram adquiridos.

Conjuntos de Dados	Adquirido de
Estações Gira	CML
Número de viagens totais das Giras	CML
<i>Eco-Counter</i>	CML
Parques de Estacionamento	CML
Congestionamentos <i>Waze</i>	CML
NO <sub>2</sub>	CML
Meteorologia	CML
Ciclovias	CML
Poluentes Atmosféricos	Base de dados da European Environment Agency <sup>1</sup>
Medidas Governamentais	Repositório COVID-19 da DSSG-PT <sup>2</sup>
Pontos de Interesse	Colegas de Mestrado

<sup>1</sup> <https://www.eea.europa.eu/data-and-maps/data/aqereporting-9>

<sup>2</sup> <https://github.com/dssg-pt/covid19pt-data>

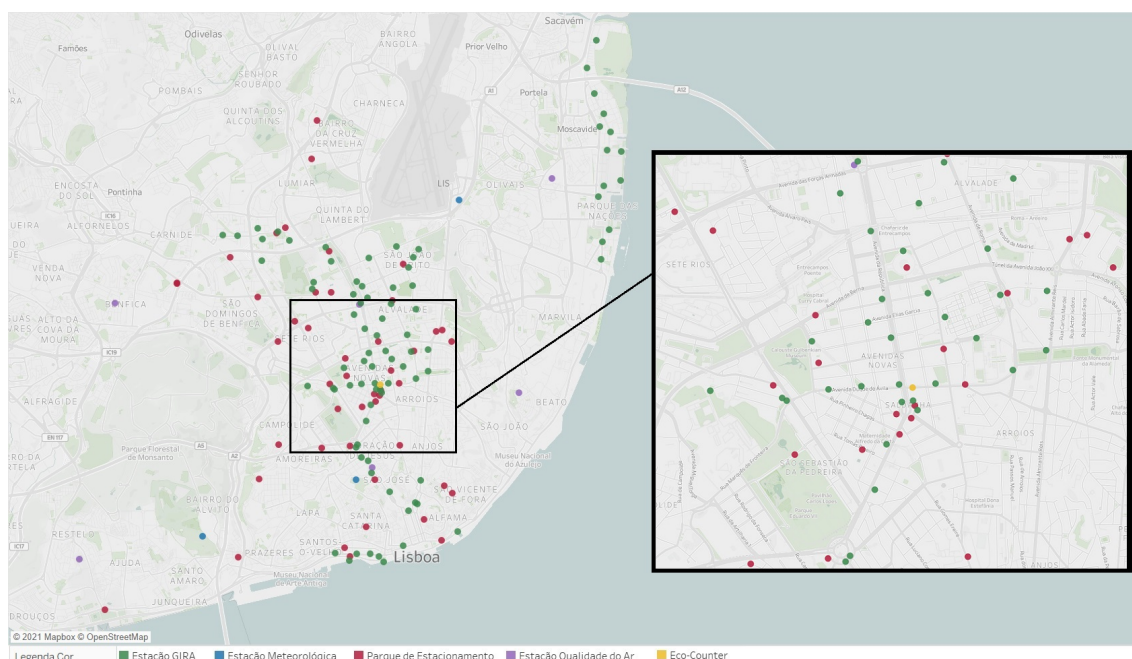


Figura 2.1: Localização das estações Gira, estações meteorológicas, parques de estacionamento, estações de qualidade do ar e sensor *eco-counter*.

## 2.2 Conjuntos de Dados

Nas subsecções que se seguem será feita uma breve descrição de cada conjunto de dados, tendo em consideração uma caracterização mais detalhada existente no relatório de qualidade dos dados.

### 2.2.1 Estações Gira

Este conjunto de dados diz respeito ao estado de 85 estações Gira entre os anos de 2019 e 2020, sendo que o ano de 2020 só apresenta dados até ao mês de setembro. Cada registo contém informação temporal e espacial, sendo esta a data e hora e um GeoJSON da localização de cada estação. Apresenta atributos que se referem ao nome da estação, à identificação única da mesma e ao seu estado, isto é, se a estação se encontra ativa ou em reparo. Os restantes atributos dizem respeito ao número de docas existentes em cada estação, ao número de docas que se encontram vazias, ao número de bicicletas na estação e à taxa de ocupação da estação a uma determinada hora específica.

Através do relatório de qualidade dos dados é ainda possível verificar que existem alguns dados que necessitaram ser melhorados. Para tal, estes dados foram processados e todo o procedimento encontra-se detalhado no capítulo 4.

### 2.2.2 Número de viagens totais das Giras

Este conjunto de dados fornece o número total de viagens diárias feitas nas bicicletas da rede GIRA entre os anos de 2019 e 2020. Cada registo contém informação acerca do número de viagens e informação temporal, mais precisamente, a data da contagem. Sendo que o primeiro registo deste conjunto de dados é no dia 01 de janeiro de 2019 e o último dia 31 de dezembro de 2020.

### 2.2.3 *Eco-Counter*

O conjunto de dados *Eco-counter* diz respeito à contagem diária de bicicletas que circularam na ciclovia da Avenida Duque de Ávila entre os anos de 2019 e 2020. O *Eco-counter* conta todas as bicicletas que passam nesta ciclovia e não apenas as bicicletas Gira. Cada registo contém o número de passagens de bicicletas na ciclovia na direção Este (em direção ao Instituto Superior Técnico), na direção Oeste (em direção ao El Corte Inglés) e o total de ambas as direções. Para além destes três valores existe ainda informação temporal, mais precisamente a data da contagem, sendo que o primeiro registo é no dia 01 de janeiro de 2019 e o último é no dia 01 de outubro de 2020.

### 2.2.4 Parques de Estacionamento

Este conjunto de dados é referente ao estado de 47 parques de estacionamento existentes na cidade de Lisboa entre os anos de 2019 e 2020. Não estando incluído neste conjunto de

dados os estacionamento de rua da EMEL. Cada registo apresenta informação temporal, mais concretamente a data e a hora, sendo que o primeiro registo é no dia 08 de Agosto de 2019 às 12:45 horas e o último é no dia 30 de Junho de 2020 às 15:10 horas. Contém também informação espacial alusiva à localização de cada parque de estacionamento. Cada registo apresenta um valor para a sua ocupação numa hora específica, e conta com informação relativa ao nome do parque, à sua capacidade máxima e ao seu tipo, isto é, se o mesmo é subterrâneo ou de superfície.

Este conjunto de dados possuía alguns dados que necessitavam ser corrigidos de forma a melhorar a qualidade dos mesmos, tal como é indicado no relatório de qualidade dos dados. Todo o processamento feito com base nesse melhoramento encontra-se detalhado no capítulo 4.

### 2.2.5 Congestionamentos *Waze*

O conjunto de dados dos congestionamentos segundo o *Waze* fornece informação acerca do estado do tráfego da cidade de Lisboa. Cada entrada representa um registo de congestionamento ocorrido em Lisboa entre os anos de 2019 e 2020. Contém informação temporal, mais concretamente, um *timestamp* em milissegundos. Existe também informação espacial para cada registo, nomeadamente informação sobre a rua onde ocorreu o congestionamento e informação sobre o cruzamento mais próximo do final do mesmo. Para além destes atributos espaciais, existe ainda um atributo no formato GeoJSON que contém as coordenadas alusivas a cada registo, mostrando desta forma qual foi local exato da via onde ocorreu o congestionamento.

Este conjunto de dados apresenta também atributos referentes ao atraso, em segundos, provocado no tráfego e ao tamanho, em metros, do congestionamento. Contém informação relativa ao nível do congestionamento, isto é, um indicativo entre 1 e 5 sobre a intensidade do mesmo, sendo que 5 significa que o trânsito ficou bloqueado. Inclui informação sobre o tipo de estrada onde ocorreu o congestionamento, por exemplo, se foi uma autoestrada, estrada primária, estrada secundária, entre outras. Os dados obtidos pelo *Waze* são analisados tendo por base dados históricos de velocidade para o mesmo dia da semana, hora e tipo de estrada, deste modo, é apresentado um atributo que refere qual o tipo de irregularidade detetada através desta análise. A irregularidade detetada pode ser uma irregularidade pequena, média, grande, enorme ou nenhuma.

### 2.2.6 NO<sub>2</sub>

Este conjunto de dados apresenta os valores de dióxido de nitrogénio medidos, de hora a hora, em seis estações diferentes de Lisboa entre os anos de 2019 e 2020. Um dos seus atributos refere-se à informação temporal, contendo a data e a hora em que foi feita a medição do valor do poluente atmosférico, sendo que o último registo é no dia 30 de Setembro de 2020 às 23:00 horas. Os restantes atributos referem-se ao valor medido em  $\mu\text{g}/\text{m}^3$  e ao nome da estação.

### 2.2.7 Ciclovias

Foi fornecido um *shapefile* que contém todas as ciclovias locais, complementares e principais da cidade de Lisboa. Para além da informação geoespacial, representada na figura 2.2, é possível ainda obter o nome da ciclovias, o eixo onde a mesma está inserida (por exemplo: eixo marginal, eixo circular exterior, etc) e o comprimento da mesma.

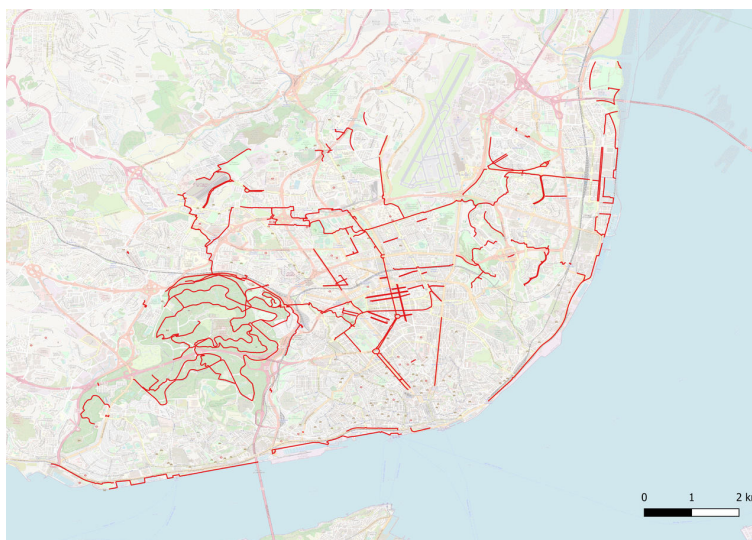


Figura 2.2: Ciclovias da cidade de Lisboa.

### 2.2.8 Meteorologia

Este conjunto de dados apresenta informação sobre os parâmetros meteorológicos medidos em três estações da cidade de Lisboa, para os anos de 2019 e 2020. Cada registo contém informação temporal, mais concretamente a data e a hora da medição, e informação espacial com a localização de cada estação. Nas três estações da cidade de Lisboa são feitas medições, de hora a hora, a seis parâmetros meteorológicos, mais concretamente à humidade, intensidade do vento, precipitação, pressão, radiação e temperatura.

Será necessário, tal como em alguns dos conjuntos de dados vistos anteriormente, realizar algum processamento nos dados para melhorar a qualidade dos mesmos.

### 2.2.9 Poluentes Atmosféricos

Através da base de dados da European Environment Agency foram adquiridos conjuntos de dados alusivos aos poluentes atmosféricos. Estes são idênticos ao conjunto de dados do  $\text{NO}_2$ , visto em 2.2.6, porém apresentam medições desde o início de 2017 até ao final de 2020. Os poluentes abrangidos são o Ozono ( $\text{O}_3$ ), Partículas com diâmetro inferior a  $2.5 \mu\text{m}$  ( $\text{PM}_{2.5}$ ), Partículas com diâmetro inferior a  $10 \mu\text{m}$  ( $\text{PM}_{10}$ ), Dióxido de Enxofre ( $\text{SO}_2$ ) e mais dados do poluente  $\text{NO}_2$ .

### 2.2.10 Medidas Governamentais

Este conjunto de dados, adquirido através do repositório COVID-19 da DSSG-PT, apresenta informação acerca das medidas governamentais impostas durante a pandemia por COVID-19 em Portugal. Cada registo contém informação temporal, mais concretamente a data em que a medida foi imposta, e ainda uma descrição da mesma.

Tal como visto no relatório de qualidade do dados, este *dataset* cobre um período temporal muito reduzido, então será necessário realizar o levantamento de mais medidas governamentais.

### 2.2.11 Pontos de Interesse

Este conjunto de dados, criado por colegas de mestrado, contém informação espacial, mais concretamente a latitude e longitude, de alguns dos pontos de interesse da cidade de Lisboa. Os pontos de interesse encontram-se divididos em categorias: transporte, ensino, saúde e lazer. E dentro dessas categorias existem ainda subcategorias.

Estes dados serão estendidos e atualizados, de forma a incluírem os pontos de interesse mais relevantes para o tema desta dissertação, como por exemplo farmácias.

## ESTADO DA ARTE E TRABALHOS RELACIONADOS

O presente capítulo faz o levantamento do estado da arte e apresenta trabalhos já realizados que se relacionam com o tema desta dissertação. Primeiramente, na secção 3.1, é feita uma introdução aos conceitos base da visualização analítica de dados. Nessa secção são ainda descritas algumas técnicas de visualização analítica de dados e algumas interações com visualizações que são consideradas pertinentes para este trabalho. De seguida, na secção 3.2, são apresentados os trabalhos relacionados no âmbito do impacto da COVID-19 na mobilidade, seguido pelo impacto da COVID-19 na qualidade do ar.

### 3.1 Visualização Analítica de Dados

Para Tamara Munzner a visualização de dados serve, principalmente, para ajudar as pessoas a efetuar tarefas de forma mais eficaz [31]. Ou seja, visualização não se trata somente de ver os dados, mas sim entendê-los e conseguir tomar decisões com base neles. Através da deteção de padrões nos dados e da transmissão de informação de forma mais direta é então possível aumentar a percepção e cognição retirada dos dados. A ideia central da visualização de dados baseia-se no mapeamento dos dados para o ecrã através de variáveis visuais. Deste modo, o utilizador pode interagir com as visualizações, o que lhe permite entender os dados e responder às questões formuladas acerca dos mesmos. A visualização analítica de dados baseia-se então no uso de *data mining* integrado com a visualização dos dados.

Um dos grandes desafios da visualização analítica de dados é a representação de grandes quantidades de dados que auxiliem as pessoas a interpretar e a compreender a informação apresentada. Então, advém a necessidade de empregar métodos de, por exemplo, agregação, sumarização e filtragem nos dados, de forma a ser possível descobrir padrões e tendências. As interações são então cruciais quando se está a construir ferramentas que lidam com uma grande complexidade. As capacidades das ferramentas que contêm visualizações interativas aumentam vastamente quando comparadas a ferramentas complexas que não apresentam interatividade [31]. *Dashboards* são ferramentas

que consistem numa coleção de múltiplas visualizações interligadas e interativas. Estes permitem assim analisar e comparar vários dados em simultâneo.

Tal como é possível observar através da figura 3.1, Tamara Munzner considera que existem dois tipos de atributos: categóricos e ordenados. Os dados ordenados podem ser do tipo ordinal (não suportam comparação aritmética completa) e quantitativo (suportam comparação aritmética). Para além disso, os dados ordenados podem variar sequencialmente entre um valor mínimo e um valor máximo, podem divergir em ambos os sentidos a partir de um ponto zero no meio de um intervalo, ou os seus valores podem ser cíclicos. Os dados categóricos não têm uma ordem implícita, mas têm frequentemente uma estrutura hierárquica [31].

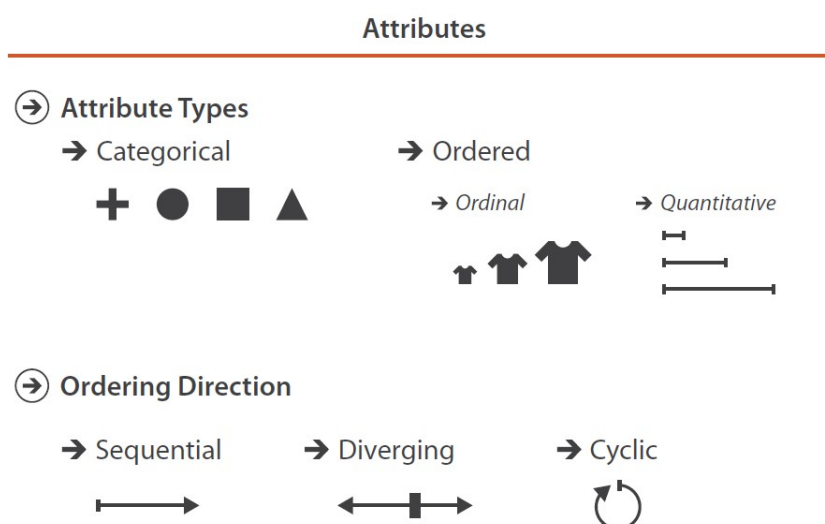


Figura 3.1: Tipos de atributos e direção dos atributos. Extraído de [31].

O mapeamento dos dados para o ecrã é feito, tal como já foi referido, através de variáveis visuais. Existem oito variáveis visuais diferentes [48], contudo as mais importantes são a posição e a forma, sendo que as restantes influenciam a maneira como as representações individuais destas duas são exibidas. Na tabela 3.1 são apresentadas e descritas brevemente as oito variáveis visuais.

Nesta secção são apresentadas técnicas de visualização de dados existentes na literatura consideradas relevantes para o problema em causa. Os principais dados utilizados na realização deste trabalho, descritos brevemente no capítulo 2, são principalmente de natureza temporal e espacial. Então, foi feito um levantamento de técnicas temporais (subsecção 3.1.1), espaciais (subsecção 3.1.2) e espaço-temporais (subsecção 3.1.3). Para além disso, e visto que o principal objetivo deste trabalho é a criação de ferramentas interativas, são apresentadas também algumas técnicas de interação com visualizações de dados (subsecção 3.1.4).

Tabela 3.1: Oito variáveis visuais.

Variável visual	Descrição
Posição	Posicionamento de visualizações dentro do ecrã, podendo ser em uma, duas ou três dimensões.
Forma	Uso de formas para representar graficamente os dados. As formas podem variar desde pontos, linhas, áreas, entre outros.
Tamanho	O tamanho é facilmente mapeado para variáveis com intervalos ou contínuas. Todavia, esta variável apenas pode ser usada para categorias de pequena cardinalidade.
Brilho	O brilho é apropriado para grandes intervalos ou para dados contínuos. Contudo, existe uma capacidade limitada das pessoas distinguirem entre muitos níveis de brilho.
Cor	A cor é apropriada para dados divergentes, sequenciais ou categóricos. Deve-se optar sempre por escolher cores intuitivas.
Orientação	A forma deve ter uma orientação clara ao longo da rotação de 360 graus. Orientação é adequada principalmente para formas que possuem apenas um eixo.
Textura	A textura é considerada a combinação de muitas das outras variáveis visuais, como por exemplo, cor, orientação e formas.
Movimento	O movimento pode ser associado a cada uma das outras sete variáveis visuais. É frequentemente usado através da variação da velocidade, de modo a ser possível observar a variação dos valores de uma variável ao longo do tempo.

### 3.1.1 Técnicas de Visualização Temporais

Uma das técnicas de visualização temporal mais utilizada consiste em **gráficos de linhas**, um exemplo disso está presente na figura 3.2. Estes são uma extensão aos gráficos de pontos onde os pontos nos dados são ligados com linhas que realçam a relação temporal dos dados.

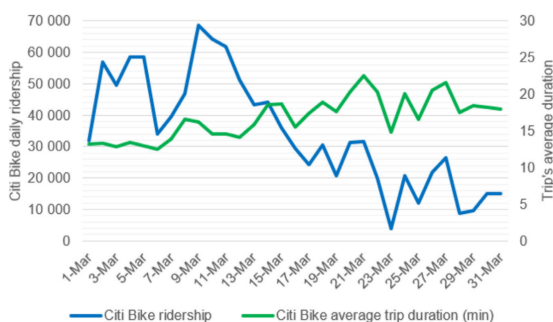


Figura 3.2: Exemplo de Gráfico de Linhas. Extraído de [45].

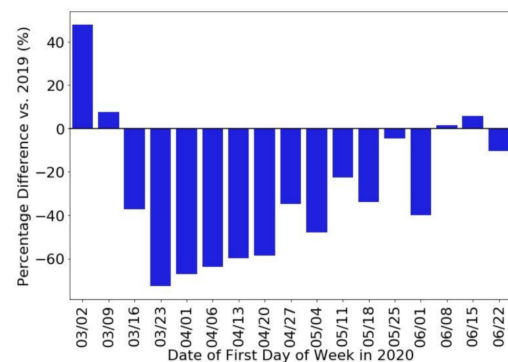


Figura 3.3: Exemplo de Gráfico de Barras. Extraído de [10].

Os **gráficos de pontos** são também uma forma muito utilizada de representar dados temporais. Nestes gráficos, à semelhança dos gráficos de linhas, o tempo é mapeado para

o eixo das variáveis independentes e o valor correspondente para o eixo das variáveis dependentes.

**Gráficos de barras** são técnicas temporais onde são destacados valores individuais, tal como acontece também nos gráficos de pontos. As barras são utilizadas para representar o valor associado à variável temporal, tornando as comparações entre variáveis bastante intuitivas e facilitando a observação das mudanças nas variáveis ao longo do tempo [41]. A figura 3.3 mostra um exemplo de gráfico de barras que analisa a diferença percentual semanal entre o número total de viagens na *Citi Bike* em 2019 e 2020.

Quando se está perante dados temporais é necessário ter em consideração qual a sua granularidade, de forma a escolher qual a visualização mais adequada. De acordo com o TimeViz Browser<sup>1</sup>, que é uma plataforma de pesquisa visual de técnicas de visualização de dados temporais, as visualizações são caracterizadas segundo os dados, o tempo e o tipo de visualização. Nesta plataforma é possível observar que os gráficos de linhas, gráficos de pontos e gráficos de barras são definidos com um arranjo temporal linear onde o tempo representa um instante, isto é, o tempo é representado como um processo linear entre o passado e o futuro e a variável temporal representa apenas um ponto único no tempo.



Figura 3.4: Exemplo de Mapa de Calor. Extraído de [13].

Numa das análises feitas pela empresa IQAir [13] foi utilizada uma técnica de visualização temporal que consiste num **mapa de calor**, representada na figura 3.4. Um mapa de calor é uma representação gráfica onde o valor dos dados é representado sob a forma de cor numa matriz. Através deste é possível identificar, muito rapidamente, zonas de altas concentrações de uma determinada variável. Na figura 3.4 é apresentado um mapa de calor seccionado por anos e meses, onde cada matriz está seccionada pelos dias do mês correspondente e a cor representa o valor medido de  $PM_{2.5}$ .

<sup>1</sup><http://www.timeviz.net>

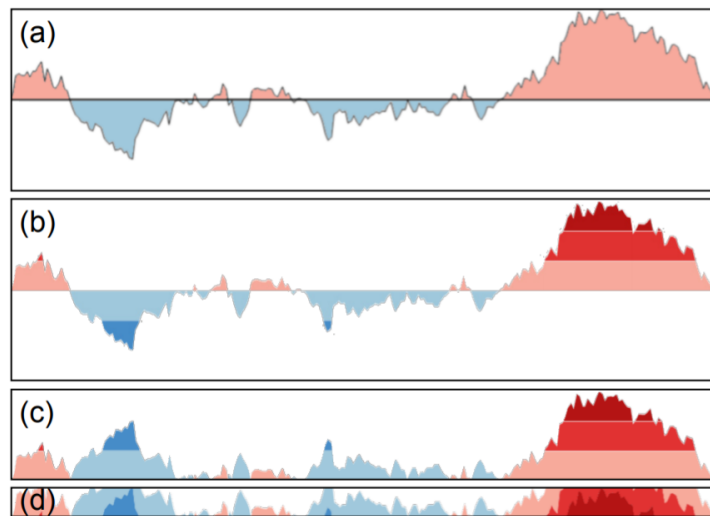


Figura 3.5: Exemplo de *Horizon Graphs*. Extraído de [35].

Com os *horizon graphs* é possível detetar padrões e comparar valores de um grande número de variáveis dependentes do tempo. São facilmente identificadas as tendências através da saturação da cor e, desde que se use a mesma escala horizontal, é possível comparar diversos gráficos de diferentes séries temporais [18]. Os *horizon graphs* utilizam dois parâmetros: o número de bandas e o valor da linha de referência horizontal, que separa horizontalmente o gráfico em valores positivos e negativos. Para construir este gráfico, começa-se inicialmente por traçar uma linha horizontal, e a parte superior desta linha é colorida de uma cor e a parte inferior doutra cor (figura 3.5 (a)). Após a coloração, são criadas bandas horizontais com diferentes saturações de cor, tendo por base a proximidade de cada banda à linha horizontal, sendo que as bandas mais distantes têm uma saturação maior (figura 3.5 (b)). De seguida, as bandas abaixo da linha de referência sofrem um efeito de espelho, sendo projetadas para a parte superior (figura 3.5 (c)). Por fim, as várias bandas são sobrepostas umas às outras, ficando unificadas num único estrato que ocupa o mesmo espaço (figura 3.5 (d)) [35].

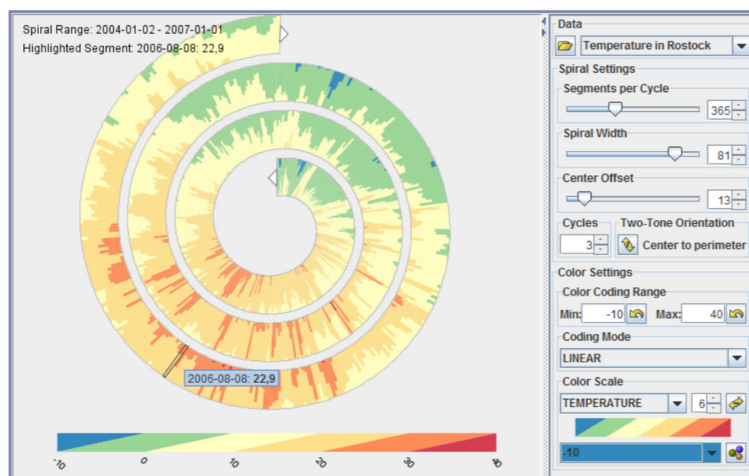


Figura 3.6: Exemplo de Gráfico Espiral. Extraído de [47].

Os **gráficos espirais**, que mostram o tempo de forma cíclica, são pertinentes quando é necessário descobrir padrões sazonais nos dados de séries temporais. Os dados temporais são mapeados ao longo da espiral, enquanto que os dados quantitativos são representados pela cor. Estes gráficos começam a partir do centro da espiral e avançam para fora dela. Sempre que se trabalha com o tempo de forma cíclica, é muito importante saber qual o período do ciclo representado, por exemplo, se é em meses, dias ou até mesmo horas [47]. Um exemplo deste tipo de gráficos é visto na figura 3.6, onde é mostrada uma *user interface* onde todos os parâmetros podem ser ajustados no painel do lado direito.

É importante ainda referir que em gráficos que usem duas métricas ao longo do tempo e onde cada uma delas representa apenas um sinal, existe o problema de ambas poderem possuir escalas completamente distintas. Como forma de solucionar este problema podem ser usados **gráficos com eixos duplos**, nos quais é adicionado um segundo eixo Y. Um exemplo disto está presente na figura 3.2, que mostra a variação da duração média diária, em minutos, de viagens em *Citi Bike* em relação ao ciclismo diário em *Citi Bike* ao longo de março de 2020. Este problema já não se verifica quando se está a usar duas métricas no mesmo gráfico em que ambas as métricas representam taxas de variação.

### 3.1.2 Técnicas de Visualização Espaciais

A análise espacial é, no geral, feita através do mapeamento de coordenadas geográficas para um mapa. Uma forma de fazer isso é através de mapas temáticos que se focam nos principais aspetos dos dados, ilustrando a distribuição geográfica de um tema particular [46]. Existem vários tipos de mapas temáticos e estes podem ser classificados consoante o número de variáveis visuais que utilizam para representar os dados: *univariate* (uma variável visual), *bivariate* (duas variáveis visuais) e *multivariate* (mais do que duas variáveis visuais).

**Mapas coropléticos** são um tipo de mapas temáticos para visualização de dados de área. Os valores de um atributo numérico são codificados no mapa como regiões coloridas ou sombreadas, permitindo estudar a forma como esse atributo evolui ao longo do território. Neste tipo de mapas, normalmente, utiliza-se a saturação da cor para representar as diferenças de intensidade do atributo, usando cores mais saturadas para valores mais elevados. É necessário ter em consideração que diferentes tipos de separação das classes de cores têm impacto no mapa gerado. Num mapa coroplético é assumido que o atributo mapeado está distribuído de forma uniforme em cada uma das regiões, sendo então impossível inferir, dentro de uma área geográfica, qual a contribuição de cada uma das suas zonas para o valor apresentado. Outro problema que este tipo de mapas levanta é que, frequentemente, os valores mais interessantes estão concentrados em áreas densamente povoadas mas com áreas pequenas e pouco visíveis, enquanto que os valores menos interessantes estão dispersos por áreas pouco povoadas mas com áreas grandes e visualmente dominantes. Na figura 3.7 encontra-se um exemplo deste tipo de mapas.

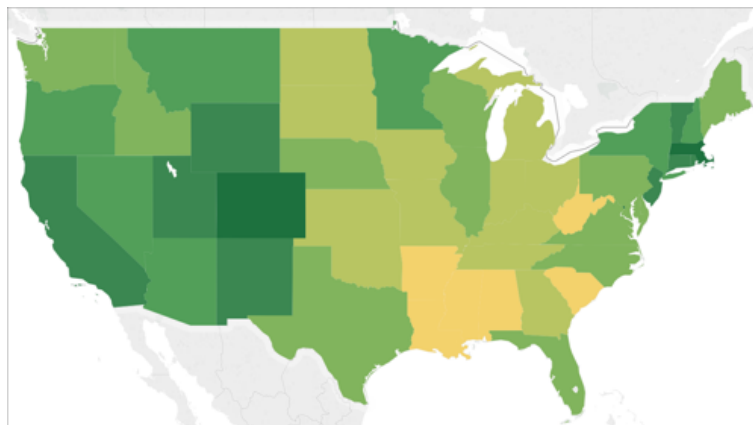


Figura 3.7: Exemplo de Mapa Coroplético. Extraído de [36].

Outro tipo de mapas temáticos são os **mapas de pontos**, que utilizam pontos para mostrar a ocorrência de um determinado fenômeno. Por norma cada ponto corresponde a uma ocorrência e, quando são representados grandes conjuntos de dados num mapa, ocorre um problema de sobreposição de pontos em certas áreas, enquanto que outras áreas ficam praticamente vazias. Existe também a possibilidade de um ponto representar um conjunto de ocorrências, neste caso o tipo de mapa é **mapa de densidade de pontos**. Neste mapa os pontos são posicionados de forma arbitrária dentro do conjunto de localizações das ocorrências de que esse ponto é constituído [2]. A figura 3.8 mostra um exemplo de mapa de pontos, onde cada ponto representa uma tempestade de granizo nos US entre 1955 e 2013. E a figura 3.9 mostra um exemplo de mapa de densidade de pontos para a população dos US em 2010, onde cada ponto representa 500 000 pessoas.

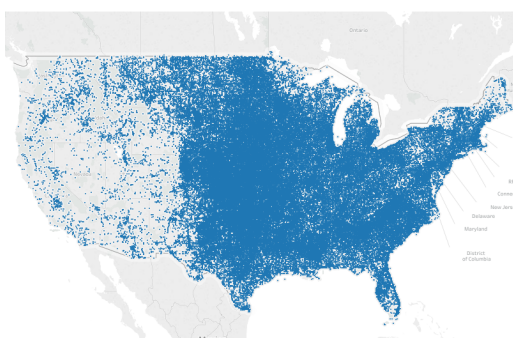


Figura 3.8: Exemplo de Mapa de Pontos. Extraído de [14].

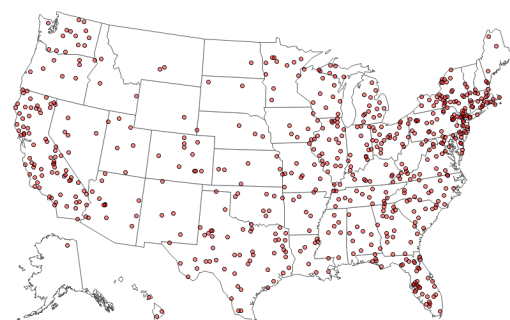


Figura 3.9: Exemplo de Mapa de Densidade de Pontos. Extraído de [1].

Os **mapas de símbolos proporcionais** são outro tipo de mapas temáticos. Nestes, o parâmetro quantitativo é mapeado para o tamanho do símbolo, sendo que as diferenças entre os tamanhos devem ser proporcionais ao atributo a mapear. Este tipo de mapas são bons para transmitir a localização exata e a quantidade associada. Porém, uma das

desvantagens que apresenta é que quando há sobreposição de símbolos e as zonas geográficas têm um tamanho pequeno em comparação com os símbolos, torna-se complicado distinguir a zona associada a cada símbolo [2]. No mapa de símbolos proporcionais da figura 3.10 está representada a relação entre os registos agregados de bicicleta e os casos confirmados de COVID-19 no dia 8 de fevereiro numa das subdivisões administrativas de Pequim.

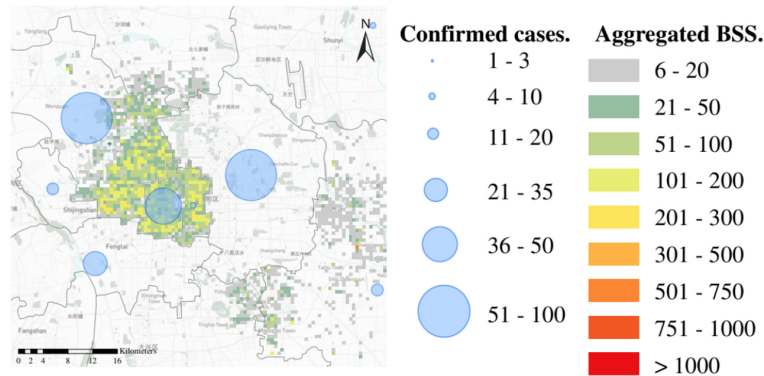


Figura 3.10: Exemplo de Mapa de Símbolos Proporcionais. Extraído de [11].

Uma forma de representar concentrações numa área é através de **mapas de densidade**, que podem ser gerados recorrendo a técnicas de agrupamento, como por exemplo, algoritmos de *kernel*. Estes mapas vêm a solucionar o problema, colocado pelos mapas de pontos, de haver sobreposição de pontos quando se está a trabalhar com conjuntos de dados de grande volume. A figura 3.11 mostra um mapa de densidade do número total de viagens de *Cite Bike*, na cidade de Nova Iorque, para cada destino em duas alturas temporais diferentes.

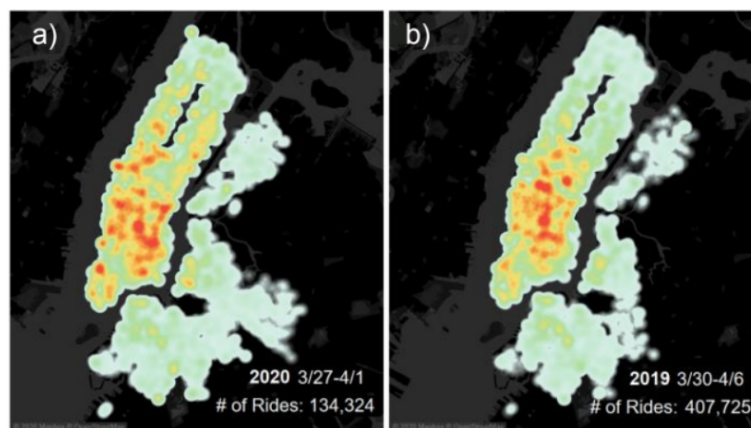


Figura 3.11: Exemplo de Mapa de Densidade. Extraído de [10].

Num **mapa de hexabins** o espaço geográfico é dividido através de *bins*, neste caso os *bins* são hexágonos. Estes hexágonos permitem que não existam espaços em branco entre cada um deles. Este tipo de mapa tem vantagens sobre o uso de mapas coropléticos pois, com este, é possível contornar o problema do uso de limites administrativos [23]. Nestes

é utilizada a cor para mostrar o número de registros agregados em cada hexágono.

**Mapas de linhas** são usados para representar dados espaciais que descrevam fenômenos lineares na superfície terrestre, fazendo uso de segmentos de linha entre pares de pontos especificados por latitude e longitude. Permitem ainda que os parâmetros dos dados sejam mapeados para a largura, cor ou rótulo da linha. **Mapas de Fluxo** são um tipo de mapas de linhas e são usados para dados que estejam relacionados com movimento. Na figura 3.12 é possível ver um exemplo de um mapa de fluxo onde foram mapeadas as rotas de corrida dos cidadãos de Paris.



Figura 3.12: Exemplo de Mapa de Fluxo. Extraído de [49].

### 3.1.3 Técnicas de Visualização Espaço-Temporais

Técnicas de visualização espaço-temporais têm como intuito analisar, simultaneamente, componentes espaciais e temporais. A **sequência de mapas temáticos** é uma técnica muito utilizada para representação de componentes espaço-temporais. Esta técnica consiste na ordenação de vários mapas em sequência, um por cada momento em análise, para transmitir a noção de evolução espaço-temporal dos fenômenos. Contudo, esta técnica é limitada, pois o número de períodos temporais a apresentar é restrito e por vezes torna-se complicado analisar a evolução gradual entre os mapas. Um exemplo de sequência de mapas temáticos é apresentado na figura 3.11 que mostra a densidade do número total de viagens de Cite Bike em dois períodos temporais distintos.

**Mapas de diferença** são outro tipo de técnica espaço-temporal. Nestes mapas é possível observar a diferença no valor de uma variável espacial ao longo do tempo, sendo que, por norma, a diferença é mapeada para a cor. Estes mapas permitem averiguar, com relativa facilidade, se os valores de determinada variável subiram ou desceram nesse intervalo de tempo. Na figura 3.13 é possível ver a diferença, em percentagem, nos valores medidos de O<sub>3</sub>, entre a primeira semana de quarentena com uma semana normal, entre a segunda semana de quarentena com a primeira semana, e entre a terceira semana de quarentena com a primeira semana em São Paulo.

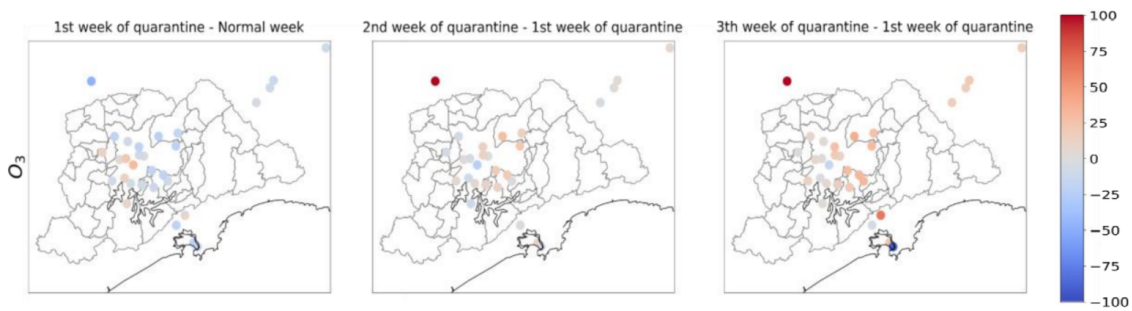


Figura 3.13: Exemplo de Mapas de Diferença. Extraído de [21].

Técnicas de **animação** são também usadas para transmitir a evolução dos fenómenos espaciais ao longo do tempo. Esta técnica é bastante usada, pois mapear as mudanças espaço-temporais através de mudanças reais na visualização é muito apelativo para os utilizadores, permitindo obter uma sensação de continuidade evolutiva. A figura 3.14 apresenta um fragmento de um *dashboard* feito pela equipa C2SMART que permite avaliar a velocidade dos carros na cidade. Esta tem a possibilidade de controlar os dados temporais através de uma sequência de momentos discretos no tempo regulados pelo *slider* existente, por filtragem de valores ou de forma automática.

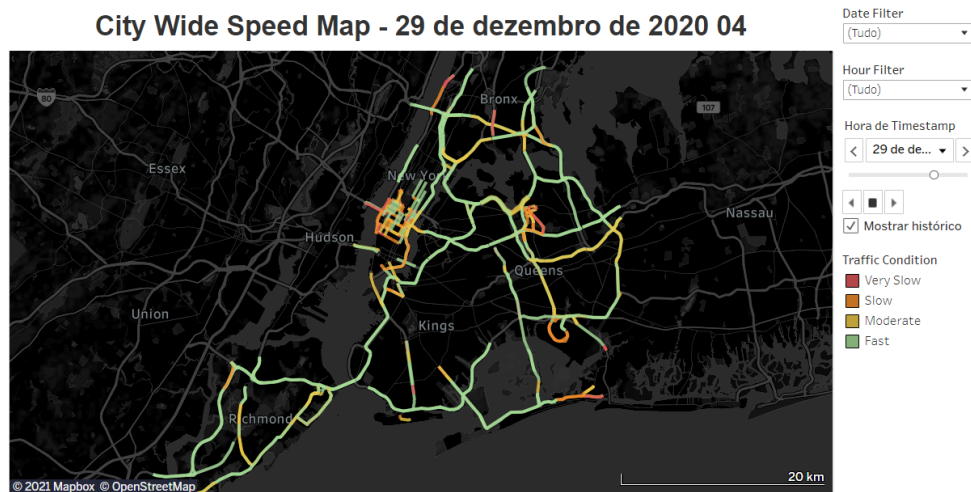


Figura 3.14: Exemplo de Animação. Extraído de [43].

### 3.1.4 Interações com Visualizações de Dados

Visualização de dados é considerado, principalmente, uma ferramenta de comunicação e apoio à tomada de decisões. Nestas ferramentas é necessário proporcionar meios para que os utilizadores obtenham respostas às suas questões sobre os dados envolventes. É frequente, quando os conjuntos de dados são complexos, a utilização de interatividade nas visualizações. Isto, pois com visualizações interativas é possível mostrar os dados em

quantidades flexíveis, melhorando a compreensão e análise dos mesmos devido à diminuição do seu grau de complexidade. Para além disso, visualizações interativas são mais envolventes e permitem ao utilizador encontrar respostas às suas questões de forma independente, aumentando assim a variedade de respostas analíticas a diferentes necessidades [20].

Tabela 3.2: Taxonomia de técnicas de interação. Extraído de [20].

<b>Interação</b>	<b>Descrição</b>
Filtrar	Mostrar apenas os dados nos quais estou interessado.
Selecionar	Marcar ou localizar dados nos quais estou interessado.
Abstrair/Elaborar	Ajustar o nível de abstração dos dados.
Visão geral e Explorar	Visão geral primeiro, zoom e filtro, depois detalhes a pedido.
Conectar/Relacionar	Mostrar como estes dados estão relacionados.
Reconfigurar	Dar uma disposição diferente dos dados.
Codificar	Dar uma representação diferente dos dados.
História	Permitir retraçar os passos que dei na exploração dos dados.
Extração de características	Permitir extrair os dados nos quais estou interessado.
Participação/ Colaboração	Permitir contribuir para os dados.
Gamificação	Mostrar os dados de uma forma mais lúdica.

Com o intuito de explorar os benefícios e objetivos de visualizações interativas, Ana Figueiras realizou um estudo onde fez o levantamento das técnicas de interatividade utilizadas atualmente. Desse estudo surgiram onze categorias: filtrar, selecionar, abstrair/elaborar, visão geral e explorar, conectar/relacionar, história, extração de características, reconfigurar, codificar, participação/colaboração e gamificação. A taxonomia de interação nas visualizações, proposta por Ana Figueiras em [20], encontra-se presente na tabela 3.2. Esta taxonomia foi escolhida devido à sua abrangência e ao facto de incluir técnicas de interação mais modernas comparativamente a outras taxonomias já existentes, como por exemplo a gamificação.

As técnicas de interação de **filtrar** e **selecionar** servem essencialmente para reduzir a complexidade dos dados, mostrando apenas os dados nos quais se está interessado. As formas de filtragem mais eficazes são as que permitem a atualização imediata da visualização, ou seja, fazendo uso de filtros dinâmicos que permitem ao utilizador uma maior concentração naquilo que realmente lhe interessa. Todavia, se existirem longos atrasos entre o *input* do utilizador e a resposta do sistema poderá ocorrer um impacto negativo na interpretação da visualização. As técnicas de seleção também funcionam como um filtro, porém, em vez de ocultar os restantes dados, colocam em evidência os dados de interesse, contrastando-os dos restantes [20].

Técnicas de **abstrair/elaborar** permitem que o utilizador ajuste, facilmente, o nível de abstração da visualização de dados às suas necessidades de interpretação, regulando a

quantidade de informação que é exibida ou enfatizada. Em [20] são vistos três exemplos muito comuns desta técnica de interação: o zoom, os detalhes sob demanda e a ligação. O **zoom** permite que o utilizador veja uma visão geral da visualização ou uma visão mais detalhada da mesma. Realizar *zooming-in* faz com que elementos de dados menores sejam ampliados, removendo ou reduzindo o tamanho de grandes dados nos quais o utilizador não está interessado. Contrariamente, *zooming-out* reduz os elementos de dados maiores. Esta técnica permite fornecer ao utilizador uma representação altamente comprimida dos dados, onde é possível observar a posição dos dados que se pretende analisar dentro do panorama geral, facilitando assim a descoberta de possíveis padrões. Por sua vez, a visão mais detalhada fornecerá uma visualização sem os dados que não são do interesse do utilizador. A técnica de **detalhes sob demanda** consiste em obter detalhes adicionais após a seleção de um item ou grupo de itens, utilizando, por exemplo, técnicas de *drill-down*, *tool-tips* ou *pop-ups*. Por fim, a técnica de **ligação** é utilizada para aceder a informação externa, por exemplo, aceder a uma visualização diferente ou a uma hiperligação [20].

A técnica de interação de **visão geral e explorar** consiste em fornecer ao utilizador uma visão geral de toda a coleção, de forma a dar contexto sobre o conjunto de dados. Porém, devido ao tamanho e complexidade de muitos conjuntos de dados, muitas das vezes é optado por mostrar apenas um número limitado de itens de cada vez. No entanto, toda a informação estará disponível para exploração que, ao ser realizada, levará à exibição de novos dados, fazendo-os entrar na visualização removendo os outros. *Panning* é o tipo mais comum de exploração, e consiste no movimento que o utilizador realiza, frequentemente através de *scrollbars*, para mudar de cena [20].

**Conectar/Relacionar** é uma técnica de interação que possibilita a visualização das relações entre os dados, sendo que as relações podem ser mostradas destacando as ligações entre os itens da visualização ou mostrando itens ocultos que são relevantes para outro item. Esta técnica ajuda na realização de comparações entre características de diferentes itens de dados na visualização [20].

A técnica de interação **história** consiste na disponibilização de maneiras para o utilizador repetir ou desfazer os passos que realizou na exploração dos dados. Desta forma, é possível aprender com os erros cometidos e refinar a exploração dos dados [20].

**Extração de características** consiste na capacidade de extrair os dados para que possam ser utilizados externamente, desta forma o utilizador pode realizar outras visualizações que melhorem a perceção [20].

A técnica interativa de **reconfigurar** traduz-se no fornecimento de disposições diferentes sobre o conjunto de dados, permitindo que o utilizador tenha perspetivas diferentes sobre os dados. Um exemplo disto é a possibilidade do utilizador escolher os atributos apresentados nos eixos dos gráficos e reorganizar a ordem das colunas [20]. **Codificar** é uma técnica que, tal como a técnica de reconfigurar, fornece uma perspetiva diferente sobre os dados. Todavia, esta técnica consiste no uso de representações completamente novas, sendo que as alterações na codificação podem ser em termos de tamanho, cor, forma, entre outros [20].

Tabela 3.3: Taxonomia de interação de visualizações espaço-temporais. Extraído de [39].

Dados	Interação	Dimensão	Descrição
ST	Filtrar	Tema	Mostrar apenas os itens com as propriedades em que estou interessado.
		Tempo	Mostrar apenas os pontos ou intervalos de tempo nos quais estou interessado.
		Espaço	Mostrar apenas os locais nos quais estou interessado.
	Selecionar	Tema	Marcar ou localizar os itens com as propriedades nas quais estou interessado.
		Tempo	Marcar ou localizar os pontos ou intervalos de tempo nos quais estou interessado.
		Espaço	Marcar ou localizar os locais nos quais estou interessado.
	Ampliar	Tempo	Variar o nível de abstração num intervalo de tempo.
		Espaço	Ampliar ou reduzir uma área.
	Conectar/ Relacionar	Tema	Relacionar dois ou mais itens de temas diferentes.
		Tempo	Relacionar dois ou mais pontos ou intervalos de tempo.
		Espaço	Relacionar dois ou mais locais.
	Reconfigurar	Tema	Dar uma disposição diferente dos dados.
		Tempo	Dar um arranjo diferente da estrutura temporal.
		Espaço	Dar uma disposição diferente dos locais.
	Codificar	Tema	Dar um código visual diferente para os componentes temáticos.
		Tempo	Dar um código visual diferente para a estrutura temporal.
		Espaço	Dar um código visual diferente para os componentes geográficos.
	Visão Geral	Primeiro	Dar primeiro uma visão geral de todas as ocorrências.
Último		Dar uma visão geral de todas as ocorrências por último.	
A pedido		Dar uma visão geral de todas as ocorrências em qualquer altura.	
Genéricos	Detalhes a pedido	Tema	Mostrar os detalhes de um item específico a pedido.
	Ligar	Tema	Levar a informações complementares.
	História	-	Permitir retrazar os passos que dei na exploração de dados.
	Extração de características	Tema	Permitir extrair os dados nos quais estou interessado.
	Participação/ Colaboração	Tema	Permitir contribuir para os dados.
	Gamificação	Tema	Mostrar os dados de uma forma mais lúdica.

A técnica interativa de **participação/colaboração** baseia-se na ideia de qualificar os utilizadores, isto é, os utilizadores têm o papel de consumidor de conteúdos e contribuem no conteúdo que consomem, criando desta forma um ambiente colaborativo. Isto pode ser feito através de etiquetas, *links*, marcadores de página, discussões duplamente ligadas, anotações gráficas, comentários tradicionais, entre outros [20].

Por fim, a técnica interativa de **gamificação** é considerada uma das mais complexas que podem ser adicionadas às visualizações de dados. As visualizações que possuem gamificação podem incluir todas as técnicas de interação vistas anteriormente, porém incluem também padrões de *design* e mecanismos de jogos [20].

Ana Figueiras e Sara Rodrigues exploraram em [39] quais as especificidades do uso de interação nas visualizações espaço-temporais, aplicando a taxonomia genérica vista na tabela 3.2. Esta análise e exploração, feita em 25 exemplos de visualizações espaço-temporais, resultou na conversão da taxonomia genérica para uma nova taxonomia especializada que considera a decomposição da maioria das técnicas de interação em: Tema, Tempo e Espaço. De notar que nem todas as 11 técnicas de interação podem ser decompostas nestas três dimensões espaço-temporais. Na tabela 3.3 encontra-se presente e descrita esta nova taxonomia.

## 3.2 Trabalhos Relacionados

Nesta secção são abordados os trabalhos relacionados com o tema desta dissertação. A subsecção 3.2.1 é referente a trabalhos que abordam o impacto que a pandemia teve na mobilidade e a subsecção 3.2.2 diz respeito ao impacto na qualidade do ar. Na subsecção 3.2.3 encontram-se resumidas as conclusões obtidas através do levantamento dos trabalhos relacionados.

### 3.2.1 Impacto da COVID-19 na mobilidade

Com o intuito de controlar a propagação da **SARS-CoV-2**, vários países decretaram diferentes regras de confinamento. Este confinamento aliado a vários outros fatores, nomeadamente, o sentimento de insegurança por parte das pessoas de estar em locais ou transportes públicos, veio a afetar significativamente a mobilidade humana.

Segundo o estudo feito por Jizhou Huang *et al.*, a China Continental teve, durante o período de pandemia, uma mudança significativa nos comportamentos relacionados com a mobilidade [24]. A análise realizada para compreender estes comportamentos usa o ano 2018 e 2019 como termos de comparação ao ano 2020, e baseia-se em cinco perspetivas: meio de transporte utilizado, tipos de locais visitados, horário de chegada ao local, distância entre a origem e o destino e padrões origem-transporte-destino.

Através de um gráfico de barras, presente na figura 3.15, Jizhou Huang *et al.* concluíram que o uso de transportes públicos diminuiu significativamente durante a pandemia. Contrariamente, o uso de bicicletas e veículos pessoais teve um aumento considerável

nos meses de pandemia. Fazendo uso da mesma técnica de visualização, constataram que houve um aumento significativo de pessoas em áreas residenciais, enquanto que houve um decréscimo acentuado de visitantes em locais de meios de transporte. Hospitais e farmácias têm um ligeiro aumento de visitas na altura da pandemia e os locais culturais e de lazer têm uma pequena diminuição de visitas [24].

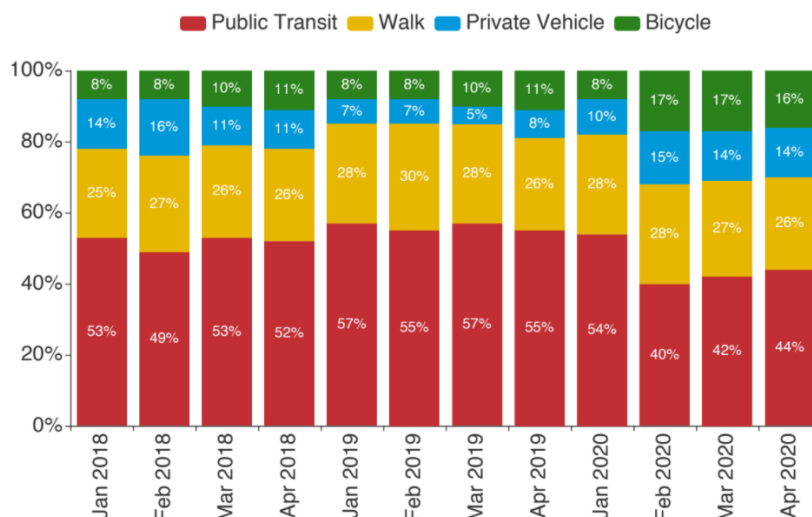


Figura 3.15: Gráfico de barras que mede a frequência relativa de diferentes tipos transportes utilizados. Extraído de [24].

A equipa de pesquisa C2SMART da Universidade de Nova Iorque tem vindo a investigar as mudanças, ao longo do tempo, nas tendências de mobilidade antes e depois das ordens de confinamento. O primeiro estudo conduzido por esta equipa centraliza a sua análise no impacto inicial da COVID-19 nos transportes [22]. À semelhança das conclusões extraídas do trabalho [24] de Jizhou Huang *et al.*, também esta equipa constatou que houve um decréscimo imediato no uso de transportes públicos, mais concretamente no uso dos serviços da *Metropolitan Transportation Authority (MTA)*. Houve, na área metropolitana de Nova Iorque, uma diminuição significativa do tráfego nas pontes e túneis e uma redução dos tempos médios de viagem. Isto teve como consequência uma redução do congestionamento e de acidentes de trânsito nesta cidade. Este primeiro relatório da equipa C2SMART evidencia ainda que, nos primeiros dez dias de março, houve um aumento de ciclistas, porém após a emissão da ordem *stay-at-home* este número caiu. No entanto, a *Citi Bike* reportou que as estações adjacentes aos hospitais tiveram um aumento no número de passageiros.

Como resultado do aumento do número de passageiros em estações adjacentes a hospitais, a *Citi Bike* lançou o programa *Critical Workforce Membership*. Os efeitos deste programa são analisados num estudo elaborado por João Teixeira e Miguel Lopes [45]. Através da análise de padrões geográficos constataram que, embora algumas estações de bicicletas tenham aumentado o número de passageiros, este aumento parece ocorrer apenas em eventos isolados. Isto é confirmado posteriormente com análises estatísticas,

que mostram que a distância das estações de bicicletas ao hospital mais próximo não tem efeitos nítidos no número médio diário de viagens. Através de um gráfico de linhas, presente na figura 3.16, que apresenta a variação no número de viagens diárias no metro e na *Citi Bike* vs o aumento no número de casos de COVID-19 no mês de março, chegaram a conclusões interessantes. Uma dessas conclusões é que o número de passageiros de *Citi Bike* aumentou 33% no dia que o número de passageiros do metro começou a baixar. Todavia, após a emissão da ordem *stay-at-home*, o número de passageiros começou a diminuir significativamente, como já tinha sido visto no estudo da equipa C2SMART.

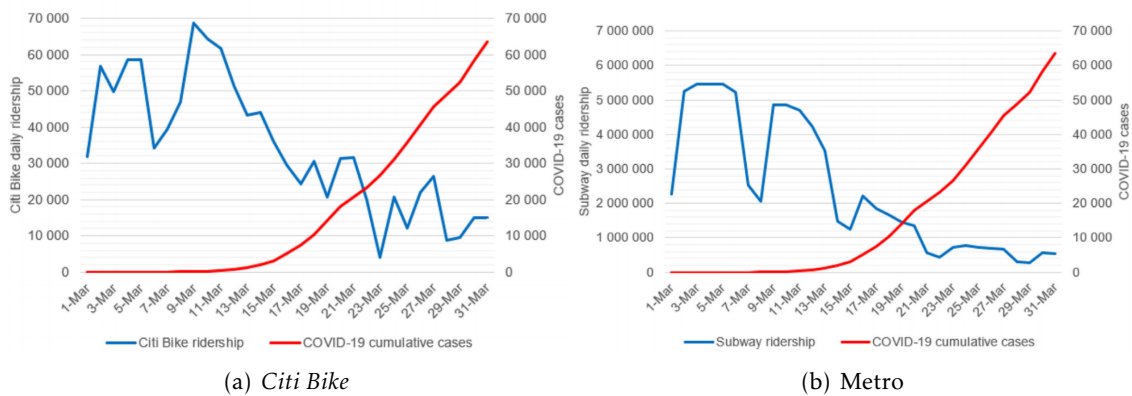


Figura 3.16: Variação do número de viagens na *Citi Bike* 3.16(a) e no metro 3.16(b) vs o aumento no número de casos COVID-19, em março. Extraído de [45].

O estudo feito por Hamada S Badr *et al.* [8], apesar de não ter como finalidade a determinação do impacto da COVID-19 na mobilidade, é de grande interesse ser considerado para efeitos de trabalhos relacionados. Neste trabalho é avaliado o impacto que a mobilidade teve no aumento de transmissão desta doença. Para tal, de forma a ser possível quantificar o total de distanciamento social em cada município dos Estados Unidos da América, definiram um rácio de mobilidade (MR) para cada dia e município. Este rácio quantifica a mudança nos padrões de mobilidade dos cidadãos, isto é, retrata a alteração no número de viagens individuais feitas em cada município por dia, comparativamente a uma base temporal anterior à pandemia. Esta medida é calculada pela soma do total de viagens num determinado dia, dividindo pela mesma medida num dia de referência anterior à pandemia.

Auxiliando-se de um mapa coroplético, e mapeando o rácio de mobilidade para a variável visual de cor (figura 3.17), Hamada S Badr *et al.* constataram que nos USA, entre 24 de janeiro de 2020 e 17 de abril de 2020 houve um decréscimo significativo na mobilidade humana. Semelhante às conclusões tiradas no estudo da equipa C2SMART [22], é também possível neste estudo constatar que Nova Iorque sofreu um grande impacto na mobilidade após decretado o confinamento. Sendo, juntamente com Nova Jersey e Massachusetts, os locais que tiveram um impacto mais acentuado [8].

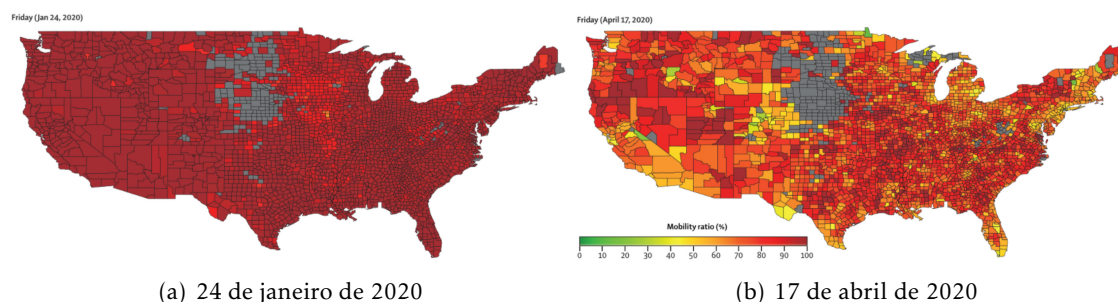


Figura 3.17: Rácio de mobilidade por cada município dos US, em 24 de janeiro de 2020 e 17 de abril de 2020. Extraído de [8].

Através do estudo feito por Edmilson D. Freitas *et al.*, é possível observar uma redução nos padrões de mobilidade dos cidadãos do Brasil e de São Paulo até o dia 23 de março de 2020 [21]. Este estudo, tendo por base os relatórios de mobilidade fornecidos pela Apple, dispõe de um gráfico de linhas para três modos diferentes de mobilidade, e tem como referência a base temporal de 13 de janeiro de 2020. Após o dia 23 de março de 2020 é identificada uma pequena subida nas tendências de mobilidade a nível nacional, todavia este aumento foi inferior à curva nacional em São Paulo. Esta avaliação feita à mobilidade ajudou, numa análise posterior, a constatar que a redução da mobilidade durante o confinamento ajudou na diminuição da concentração dos poluentes atmosféricos. O estudo feito a esta correlação será analisado em 3.2.2.

Xinwei Chai *et al.*, motivados em descobrir o padrão dos comportamentos espaço-temporais do uso de bicicletas partilhadas durante a pandemia em Pequim, realizaram uma análise da qual extraíram conclusões importantes [11]. O seu objetivo foca-se, principalmente, em medir as mudanças do comportamento dos cidadãos de forma a melhorar a gestão de emergência e a recuperação pós-pandemia da cidade. Para isso, e com base nos dados do *Bike Sharing System (BSS)*, realizaram uma primeira análise que permitiu ver quais os padrões de mobilidade entre o mesmo período de tempo de 2019 e 2020. Como o surto coincidiu com o feriado do Ano Novo Chinês de 2020, foi necessário remover o impacto do mesmo, de forma a ser possível avaliar com maior precisão os impactos da COVID-19. Assim, chegaram à conclusão que houve uma redução do uso do BSS por volta de 40%, comparativamente a 2019.

Neste trabalho [11] foram ainda mais ao detalhe, realizando uma identificação dos períodos chave da pandemia e classificando-os em três diferentes fases: antes da pandemia, durante a pandemia e após a mitigação da pandemia. Havendo esta clara separação entre as fases, e auxiliando-se de sequências de *heat maps*, realizaram uma análise espacial dos padrões de ciclismo e os diferentes pontos de interesse da cidade. Chegaram à conclusão que o uso de bicicletas diminuiu em todos os pontos de interesse escolhidos, sendo mais sentido perto das empresas tecnológicas, como pode ser observado na figura 3.18. Locais

como estações do metro e supermercados não sofreram uma diminuição tão acentuada, e a explicação dada para este fenómeno baseia-se no facto dos residentes terem de adquirir bens essenciais. Por fim, analisaram ainda a relação entre o uso das bicicletas partilhadas e os casos confirmados de COVID-19 durante a fase da pandemia. Foi notado que as áreas residenciais onde há casos de COVID-19 situam-se no centro da cidade, e são os sítios onde se sente mais a diminuição do uso de bicicletas partilhadas, inclusive depois da mitigação da pandemia.

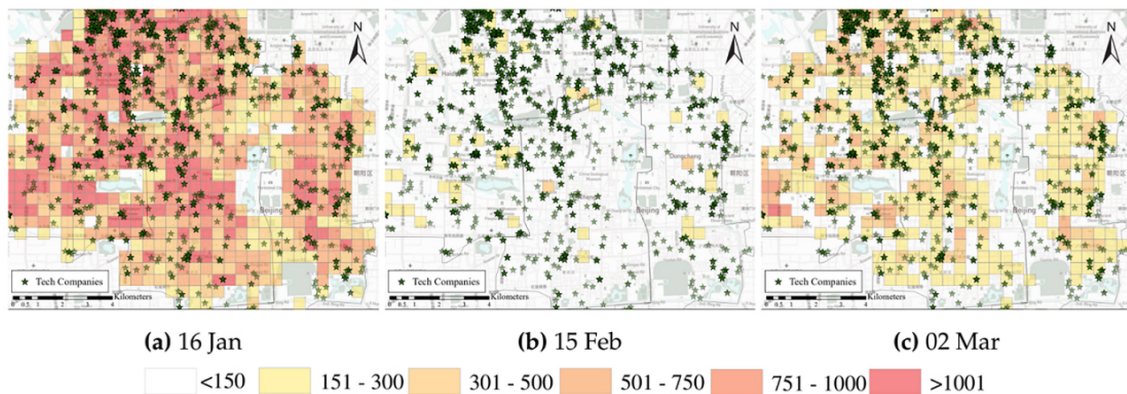


Figura 3.18: Uso de bicicletas partilhadas e empresas tecnológicas. Extraído de [11].

Noutro estudo, conduzido pela equipa de pesquisa C2SMART, é revelado quais foram as tendências de mobilidade na cidade de Nova Iorque até finais de junho de 2020 [10]. Apesar de haver uma pequena subida comparativamente aos números observados na altura da emissão da ordem *stay-at-home*, os números relativos às viagens de metro continuam baixos no mês de junho, cerca de 80% menos viagens em relação a 2019. No entanto, através do gráfico onde foram traçadas as tendências de mobilidade visto na figura 3.19, é possível ainda observar que o tráfego médio diário de veículos em pontes e túneis do MTA é restaurado para valores semelhantes aos pré-pandémicos. Para além disso, conclui-se que o número de viagens de autocarro teve uma boa recuperação, havendo somente uma queda de menos de 50% no número de viagens no final de junho, comparativamente a 2019. Concluíram ainda que em junho o número de viagens de autocarro atingiu cerca de 9% mais viagens que o metro.

Através da análise dos dados recolhidos por câmaras de tráfego, processados através de aprendizagem automática, são analisados por esta equipa os padrões de viagens em locais importantes de Nova Iorque. Neste relatório, [10], a equipa C2SMART explorou os dados referentes às caminhadas feitas pelos peões, às viagens de carro e às viagens de bicicleta feitas na *Fifth Avenue/42 Street*, Manhattan. Os gráficos elaborados revelam que houve um aumento gradual na densidade de carros e peões até junho, todavia estes valores ainda se apresentam abaixo dos medidos antes da pandemia. Contrariamente, a densidade de ciclistas nesta localização teve um grande aumento, aproximando-se dos valores pré-pandémicos.

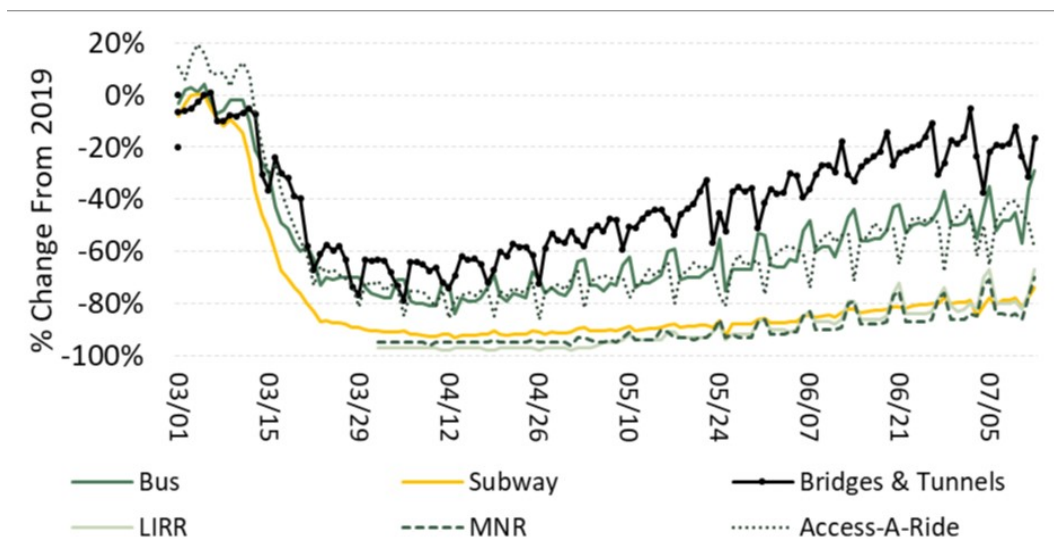


Figura 3.19: Tendências de mobilidade comparativamente a 2019 na cidade de Nova Iorque. Extraído de [10].

Tendo em conta os modos de transporte suaves, mais concretamente as bicicletas do *City Bike*, a equipa C2SMART tirou algumas conclusões importantes neste estudo [10]. Observou-se que na primeira semana de março houve um aumento de 40% das viagens feitas na *City Bike* em comparação com o mesmo período do ano 2019, tal como já tinha sido constatado no seu primeiro relatório [22]. Após a descida abrupta no número de viagens feitas no *City Bike*, aquando da emissão da ordem *stay-at-home*, começou no mês de abril o aumento no número de viagens. Existindo picos nos valores deste gráfico no mês de maio, a equipa decidiu explorá-los, usando para este efeito um gráfico de linhas que permite a correlação da temperatura com o número de viagens feitas. Foi verificado que estes picos no aumento de viagens coincidem com picos de temperaturas altas, o que pode indicar que o aumento das viagens poderá estar relacionado também com o clima.

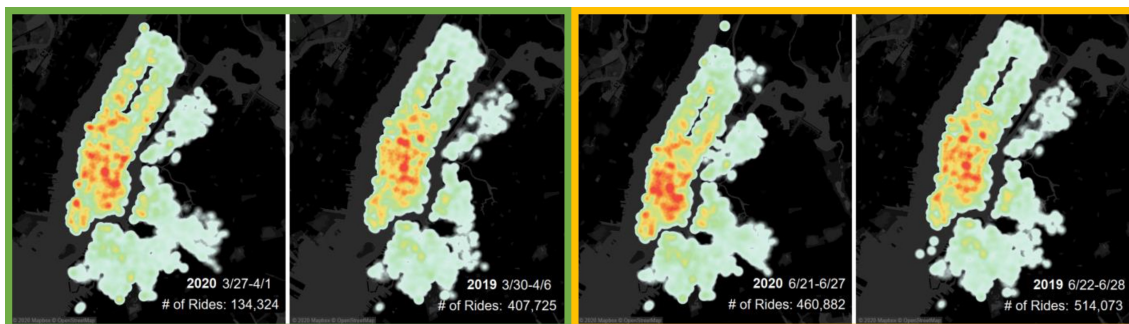


Figura 3.20: Mapa de densidade do número total de viagens para as estações de destino *City Bike*. Extraído de [10].

Esta equipa decidiu ainda explorar a concentração de viagens que terminam em cada estação *City Bike*, usando o ano 2019 como comparação. Para isto, foram elaborados mapas

de densidade, presentes na figura 3.20, que permitiram concluir que o número total de viagens para destinos diferentes é menor em 2020, todavia é mais espalhado pela cidade, tendo alguns *hotspots* perto de hospitais e parques.

### 3.2.2 Impacto da COVID-19 na qualidade do ar

Tal como foi visto em 3.2.1, vários países decretaram ações de contenção para minimizar os impactos da COVID-19. As medidas implementadas levaram a um abrandamento significativo do uso de veículos, transportes públicos e processos industriais. Isto levou a que se sentisse, por todo o mundo, um melhoramento na qualidade do ar. Dado a ter havido uma redução do tráfego rodoviário na maioria das cidades do mundo, é expectável que a concentração de  $\text{NO}_2$  no ar tenha diminuído também. Isto, pois o  $\text{NO}_2$  está, principalmente, relacionado com a queima de combustíveis fósseis e com o uso de veículos motorizados [9]. Este tem também um papel importante na formação de ozono troposférico [12] e têm influências negativas na saúde respiratória das pessoas [9].

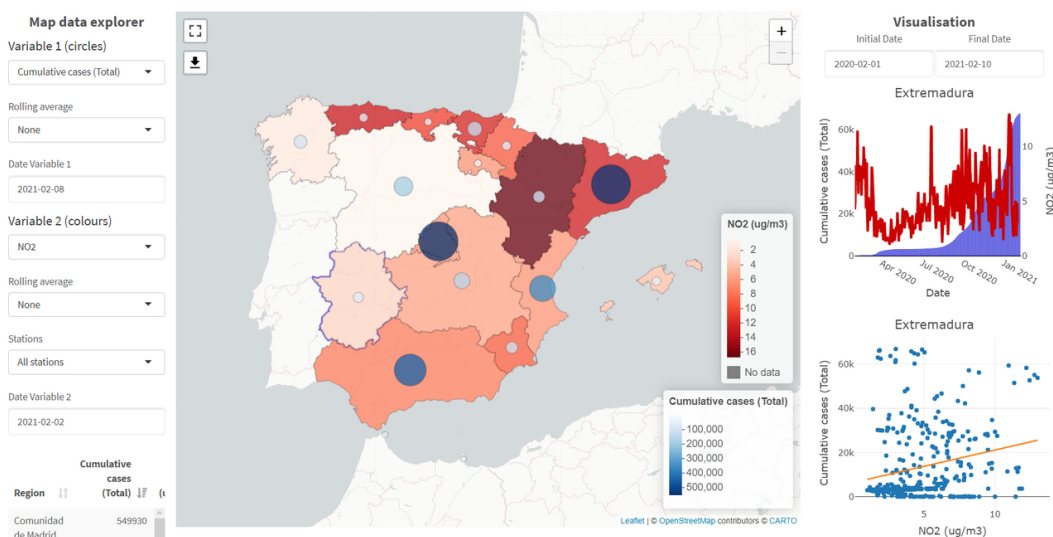


Figura 3.21: Interface DatAC<sup>2</sup>.

Jordi Martorell-Marugán *et al.* implementaram um site *web* intitulado DatAC<sup>2</sup>, que integra dados dos casos de COVID-19, dados meteorológicos, dados demográficos e dados da qualidade do ar das várias regiões espanholas. Estes dados são apresentados sob a forma de uma interface interativa, permitindo vários tipos de análises. Desta forma, proporciona ao utilizador a capacidade de efetuar correlações entre os dados disponíveis mencionados anteriormente. Existe ainda a possibilidade de dividir a informação entre áreas urbanas, suburbanas e rurais, e ainda escolher qual a variável que é mapeada para os círculos representados no mapa e para a cor das regiões do mapa. Esta plataforma incorpora visualização de dados com informações espaciais e temporais, tal como é possível observar na figura 3.21.

<sup>2</sup><https://covid19.genyo.es>

Avaliando as tendências e associações entre variáveis de qualidade do ar, do clima, demográficas e indicadores COVID-19, Jordi Martorell-Marugán *et al.* publicaram um estudo onde transmitem as conclusões extraídas da plataforma criada [30]. A sua análise baseou-se em três fases diferentes da pandemia em Espanha: antes do confinamento (23 de janeiro 2020 até 13 de março 2020), durante o confinamento estrito (14 de março 2020 até 3 de maio 2020) e meses após o relaxamento das medidas do confinamento (4 de maio 2020 até 20 de junho 2020). Constataram que todos os poluentes analisados ( $\text{NO}_2$ , **Monóxido de Carbono (CO)**,  $\text{PM}_{2.5}$ ,  $\text{PM}_{10}$ ,  $\text{SO}_2$ ), à exceção do  $\text{O}_3$ , caíram significativamente durante o confinamento estrito em todas as regiões urbanas de Espanha, especialmente o  $\text{NO}_2$  e  $\text{PM}_{10}$ . Por outro lado, os níveis de  $\text{O}_3$  subiram 50% neste período, o que pode ser explicado, principalmente, pela redução das emissões de óxidos de nitrogénio, levando a uma menor titulação de  $\text{O}_3$  pelo NO [30, 42]. Após o relaxamento das medidas de confinamento, os níveis das emissões dos poluentes CO,  $\text{PM}_{2.5}$ ,  $\text{SO}_2$  continuaram a descer, enquanto que os níveis de  $\text{NO}_2$ ,  $\text{PM}_{10}$  e  $\text{O}_3$  sofreram uma subida nos seus valores. Concluíram ainda que estas observações são semelhantes para zonas suburbanas. Nas zonas rurais, apesar dos valores serem semelhantes, as variações dos valores entre os três períodos é menor [30].

A empresa Suíça de tecnologia de qualidade do ar, IQAir, desenvolveu um relatório onde analisa o impacto da COVID-19 na qualidade do ar em dez grandes cidades a nível mundial [13]. Este relatório compara as medições nas concentrações do poluente  $\text{PM}_{2.5}$  antes e durante a pandemia. As análises são baseadas nas três semanas em que as medidas de confinamento foram mais rigorosas em cada cidade. Estas três semanas são analisadas para os anos 2016, 2017, 2018, 2019 e 2020. Decidiram analisar o mesmo período de tempo em todos os anos, pois os níveis de poluição do ar costumam variar muito com as mudanças sazonais no clima.

Uma das cidades analisadas no relatório da empresa IQAir é a cidade espanhola de Madrid [13]. Através de um gráfico que mostra os níveis anuais de  $\text{PM}_{2.5}$  ao longo das três semanas de confinamento constataram que, comparativamente ao ano de 2019, a cidade de Madrid teve um decréscimo na concentração deste poluente, mais precisamente uma redução de 11%. Todavia, este gráfico mostra que, em relação aos 4 anos anteriores, não existe uma tendência de queda tão clara como em comparação apenas com o ano 2019. Havendo até uma subida de 2% na concentração deste poluente em comparação com a média dos 4 anos anteriores. Verificou-se que Madrid tem, praticamente todos os anos, mais de 80% das horas com um valor de  $\text{PM}_{2.5}$  considerado dentro dos parâmetros de exposição médio anual ( $10 \mu\text{g}/\text{m}^3$ ) anunciado pela OMS [3]. Isto indica que, apesar de ter havido uma pequena melhoria na qualidade do ar sentida em Madrid na altura do confinamento, esta melhoria não foi significativa quando comparada com tendências a longo prazo. Contrariamente a este relatório, a análise feita no estudo de Jordi Martorell-Marugán *et al.* [30] só tem em conta uns meses antes do confinamento, apresentando assim resultados que consideram tendências a curto prazo.

Tal como visto em 3.2.1, Edmilson D. Freitas *et al.*, publicaram um estudo onde é

possível observar os padrões de mobilidade dos cidadãos de São Paulo e relacionar estes padrões com os valores registados para os parâmetros de qualidade do ar [21]. São elaboradas análises para os anos de 2019 e 2020, representadas em gráficos de linhas que medem a concentração de cada poluente ao longo dos dias. Estas visualizações mostram que, no geral, a concentração de todos os poluentes diminuiu. Todavia, existem três períodos de tempo onde as concentrações de todos os poluentes medidas em 2020 são superiores ao ano 2019, sendo esses períodos entre 18-21 de março, 26-27 de março e entre 5-8 de abril. Este facto parece estar relacionado com o aumento sentido na mobilidade nessas épocas.

São Paulo é uma das cidades analisadas no relatório da empresa IQAir, onde é possível observar que esta cidade tem sofrido, de ano para ano, uma subida nos valores de  $PM_{2.5}$ . Porém, entre 2019 e 2020 houve uma descida acentuada nas concentrações deste poluente, mais precisamente uma redução de 32%. Comparando com valores medidos desde 2017, observa-se uma redução de 26% nos valores de  $PM_{2.5}$ . É possível ainda averiguar que a média da concentração deste poluente, durante o confinamento de 2020, foi  $10.1 \mu\text{g}/\text{m}^3$  [13]. O que indica que o confinamento teve um impacto significativamente positivo na qualidade do ar da cidade de São Paulo.

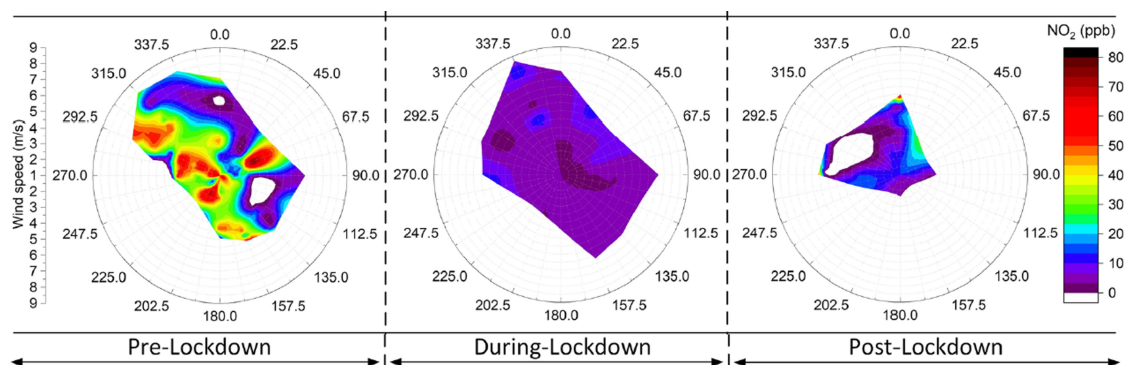


Figura 3.22: *Bivariate polar plots* que ilustram o efeito conjunto da velocidade e direção do vento sobre o poluente  $\text{NO}_2$  observado na estação de Rakah. Extraído de [6].

Ismail Anil e Omar Alagha publicaram um estudo no qual analisam qual foi o impacto da pandemia na qualidade do ar da Província Oriental na Arábia Saudita [6]. À semelhança de Jordi Martorell-Marugán *et al.* [30], também dividiram a sua análise em três épocas diferentes da pandemia: pré-confinamento (15 de setembro 2019 até 22 de março 2020), durante o confinamento (23 de março 2020 até 20 de junho 2020) e após o confinamento (21 de junho 2020 até 18 de julho 2020). Avaliando, portanto, os resultados tendo em conta as tendências sentidas a curto prazo. Visto que a Província Oriental na Arábia Saudita enfrenta severas tempestades de areia, decidiram criar gráficos de barras circulares, mais precisamente *rose plots*, para os valores da velocidade do vento. Desta forma puderam, posteriormente, em conjunto com *bivariate polar plots*, analisar os valores medidos dos parâmetros de qualidade do ar comparativamente aos valores da velocidade do vento. Na figura 3.22 são apresentados os *bivariate polar plots* referentes ao poluente

NO<sub>2</sub>. Recorreram a diagramas de caixa de bigodes para representar as concentrações médias dos parâmetros da qualidade do ar, divididos pelas três épocas diferentes e por estações. Os diagramas feitos para o poluente NO<sub>2</sub> estão representados na figura 3.23. Com estes diagramas foi possível observar a média, mediana, quartil inferior, quartil superior, intervalo interquartil e *outliers* de cada poluente, observado em cada estação para os três diferentes períodos.

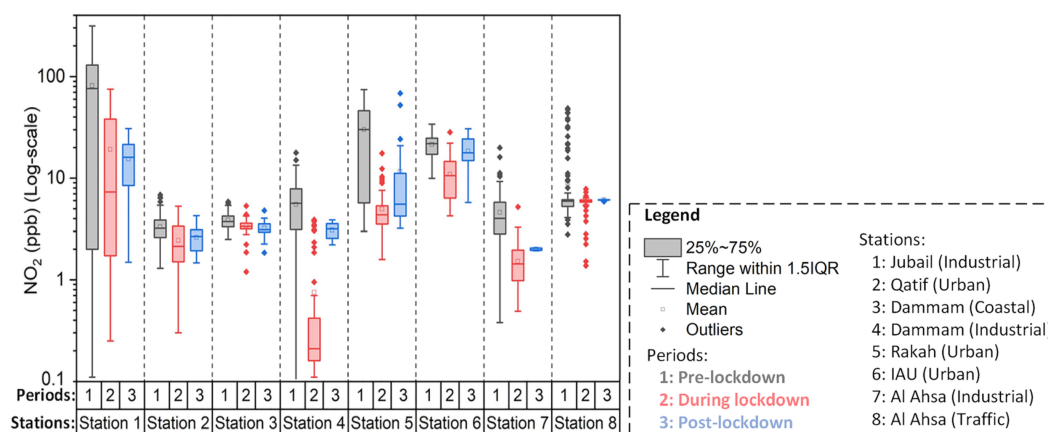


Figura 3.23: Diagramas de caixas de bigodes das concentrações médias diárias de NO<sub>2</sub> dentro dos períodos de pré-confinamento, confinamento e pós-confinamento. Extraído de [6].

Através da análise dos diagramas de caixa de bigodes, Ismail Anil e Omar Alagha, tiraram conclusões importantes em relação aos valores medidos por cada estação nas diferentes épocas da pandemia. Todavia, as conclusões convergem para o facto do NO<sub>2</sub> ter sido o poluente que respondeu melhor às medidas do confinamento, tendo diminuído as suas concentrações em todas as estações. Os restantes parâmetros, excluindo o O<sub>3</sub>, tiveram reduções em várias estações. Semelhante às conclusões tiradas por Jordi Martorell-Marugán *et al.* em [30], as concentrações de O<sub>3</sub> aumentaram em quase todas as estações de monitorização durante o período de confinamento. Foram ainda realizados gráficos de barras com o objetivo de ter uma melhor perceção da diferença percentual de cada poluente atmosférico, sentida em cada estação durante o período de confinamento e pós confinamento. Estes gráficos confirmaram as conclusões anteriores de uma forma mais simples e imediata [6].

Com os *bivariate polar plots*, Ismail Anil e Omar Alagha mediram o efeito da velocidade do vento nas concentrações dos poluentes atmosféricos, medidos na estação urbana de Rakah nas três diferentes épocas da pandemia. Mapearam a velocidade do vento para a variável visual raio, a direção do vento para o ângulo e a concentração do poluente para a cor. É possível observar que os ventos de norte com velocidade moderada e os ventos do sul com velocidade baixa, resultam numa maior concentração de CO e SO<sub>2</sub> antes e durante o confinamento. Após o confinamento, as concentrações destes poluentes atmosféricos aumentaram notavelmente, e têm associadas velocidades baixas de ventos soprados em

todas as direções. Os valores mais altos de  $\text{NO}_2$  parecem estar associados a ventos de sudoeste e nordeste com velocidades variáveis. As maiores concentrações de  $\text{PM}_{10}$  foram observadas quando o vento tinha velocidades superiores a 3 m/s [6]. Isto mostra que, apesar da COVID-19 ter tido um impacto na redução dos poluentes atmosféricos a curto prazo, as condições climáticas também desempenham um papel importante nos valores das concentrações destes poluentes [17].

Embora haja estudos por todo o mundo que mostram uma redução acentuada dos poluentes atmosféricos, é preciso ter em consideração tanto as mudanças a curto prazo na qualidade do ar, como as mudanças a longo prazo. As mudanças a curto prazo devem-se, essencialmente, a mudanças de estações do ano e condições meteorológicas. Já as mudanças a longo prazo podem estar relacionadas com regulamentos ambientais das regiões. Foi segundo esta doutrina que se basearam Shelby Zangari *et al.* na sua avaliação de como a COVID-19 afetou a qualidade do ar de Nova Iorque.

Shelby Zangari *et al.* analisaram as concentrações de  $\text{NO}_2$  e  $\text{PM}_{2.5}$  nas primeiras dezassete semanas de 2015-2020 (janeiro a maio) na cidade de Nova Iorque [50]. Foram elaborados gráficos de dispersão lineares para a média da concentração de cada poluente, pelos dias de cada ano. Analisando estas visualizações chegaram à conclusão que, a longo prazo, não houve um grande impacto na concentração de  $\text{PM}_{2.5}$  na região metropolitana de Nova Iorque. Isto, pois o ano 2020 teve uma diminuição semelhante à que aconteceu nos anos anteriores. Mesmo havendo uma diminuição na concentração média diária de  $\text{PM}_{2.5}$  de, aproximadamente, 36% nas primeiras dezassete semanas de 2020, é possível constatar que ocorreram reduções semelhantes no mesmo período de tempo para os anos anteriores. Conclusões semelhantes são tiradas para o poluente  $\text{NO}_2$  que, embora tenha tido uma redução de 51%, esta é semelhante à observada nos anos anteriores.



Figura 3.24: Índice de cores padrão suportado pela OMS para os intervalos dos valores do poluente  $\text{PM}_{2.5}$ . Extraído de [13].

Através do relatório realizado pela empresa IQAir [13] é possível constatar que, tal como dito por Shelby Zangari *et al.* em [50], as concentrações de  $\text{PM}_{2.5}$  têm vindo a diminuir ao longo dos anos em Nova Iorque. Havendo uma redução na concentração deste poluente de 25% comparativamente a 2019. Todavia, contrariamente à conclusão chegada por Shelby Zangari *et al.*, neste relatório, considerando apenas três semanas de cada ano, há uma redução de 29% no valor deste poluente comparativamente aos quatro anos anteriores. Indicando desta forma que, para este período de tempo em estudo, há uma melhoria significativa a longo prazo na qualidade do ar. É visto ainda que, durante a altura do confinamento, o valor médio deste poluente é  $4.4 \mu\text{g}/\text{m}^3$ , sendo considerado

um valor dentro dos *guidelines* da OMS. Sendo que as *guidelines* da OMS se encontram presentes na figura 3.24. Observa-se que 98% das horas em 2020 cumpriram as metas da OMS e os restantes 2%, embora não cumpram as metas da OMS, encontram-se no intervalo de valores considerados bons para este poluente. Os restantes anos tiveram sempre a cima de 87% das horas com valores considerados bons.

No relatório da empresa IQAir [13] é constatado que cidades com níveis historicamente mais altos de  $PM_{2.5}$  testemunharam quedas mais substanciais nos seus valores na altura do confinamento. Uma dessas cidades é Delhi, na Índia, onde é vista uma redução de 60% nos níveis deste poluente comparativamente a 2019 e uma redução de 55% comparativamente à média dos quatro anos anteriores. Nesta cidade, ano após ano, os valores deste poluente têm tido tendência a subir, porém de 2019 para 2020 os valores diminuíram drasticamente. Dias considerados “não saudáveis” (valores superiores a  $55.4 \mu\text{g}/\text{m}^3$ ) tiveram uma mudança de 68% em 2019 para 17% durante o confinamento. É observado ainda que 3% das horas tiveram qualidade do ar considerada boa em 2020, enquanto que em 2019 a qualidade do ar nunca foi considerada boa.

A cidade de Roma em Itália foi a única que, segundo o relatório da empresa IQAir [13], observou um aumento de 30% na poluição por  $PM_{2.5}$  em comparação aos níveis medidos em 2019. Contudo, houve um aumento das horas que os valores medidos nesta cidade estiveram dentro dos padrões considerados aceitáveis pela OMS.

### 3.2.3 Conclusões

As conclusões retiradas dos trabalhos analisados na subsecção 3.2.1 convergem todas para o facto dos padrões de mobilidade terem sofrido uma enorme redução após o surgimento desta nova pandemia. Contudo, nas cidades que possuem um sistema de bicicletas partilhadas, foi visto um aumento do uso das mesmas aquando da redução do uso de transportes públicos, o que pode, possivelmente, indicar uma migração no tipo de transporte usado pelos cidadãos. Este aumento na utilização de bicicletas partilhadas durou apenas até à emissão da ordem de *stay-at-home*, tendo a partir desse momento começado a diminuir o número de viagens.

Foi visto ainda a importância de definir pontos de interesse nas cidades, pois os padrões de mobilidade são, no geral, diferentes consoante a localização. Em todos os trabalhos analisados houve uma redução no número de viagens para todos os pontos de interesse, exceto para zonas de hospitais e farmácias. Tendo sido encontrados, em alguns trabalhos, *hotspots* em estações de bicicletas perto de zonas com atividades alusivas à saúde.

Nos trabalhos analisados na subsecção 3.2.1 foram utilizados diversos indicadores referentes à mobilidade, sendo que um dos mais utilizados foi relativo à contagem de viagens feitas nos diversos tipos de transporte analisados, tanto em períodos temporais durante a pandemia como em períodos temporais anteriores à pandemia. Tendo estas

contagens, é efetuado em alguns trabalhos a diferença entre os valores nos mesmos períodos temporais de anos diferentes, sendo que este valor por vezes é apresentado em percentagem. A percentagem de uso dos diferentes meios de transporte e a percentagem do número total de visitas aos diferentes pontos de interesse também são indicadores interessantes que foram utilizados em alguns trabalhos. Num dos trabalhos analisados em 3.2.1 é apresentado ainda um indicador chamado rácio de mobilidade. Este mostra a alteração no número de viagens individuais feitas em cada município dos Estados Unidos por dia, comparativamente a uma *baseline* anterior à pandemia. O seu cálculo é feito através da soma do total de viagens num determinado dia, dividido pela mesma medida do dia da *baseline*.

Na maioria dos lugares, a parcial paralisação da mobilidade e da indústria revelou um vislumbre de um mundo mais limpo e com menos poluição atmosférica. No entanto, a percepção visual, por si só, pode ser enganadora ao observar a poluição no ar. Para avaliar se a pandemia teve um impacto significativo nos valores dos poluentes atmosféricos não basta fazer apenas uma análise a curto prazo, é preferível considerar, sempre que possível, padrões a longo prazo.

É possível ainda constatar, através dos trabalhos analisados na subsecção 3.2.2, que as maiores melhorias na qualidade do ar foram encontradas em locais que tinham níveis mais elevados de poluentes atmosféricos antes da chegada da COVID-19, comparativamente com locais que apresentavam níveis de poluição atmosférica razoável. Todavia, no geral, todos os locais sofreram melhorias na qualidade do ar, sendo que o parâmetro que teve um decréscimo mais notório foi o  $\text{NO}_2$ . Em contrapartida, houve um poluente que, na maioria das cidades, sofreu um aumento do seu valor, sendo este o  $\text{O}_3$ . Os autores dos trabalhos analisados explicam que este fenómeno se deve à menor titulação entre o  $\text{O}_3$  pelo  $\text{NO}$ , que ocorre pela diminuição das emissões de óxidos de nitrogénio.

Em alguns dos trabalhos é visto ainda que as condições climáticas têm influência na qualidade do ar. Então é recomendado fazer comparações entre os mesmos meses de anos diferentes, ao invés de comparar apenas com meses anteriores ou posteriores.

Os indicadores utilizados nos trabalhos analisados na subsecção 3.2.2 são maioritariamente referentes à média das concentrações dos poluentes atmosféricos em diferentes períodos da pandemia e períodos anteriores à pandemia. É muito utilizado a percentagem da diferença entre a média da concentração do poluente de um ano com o ano anterior ou com a média de uma determinada *baseline* temporal. A percentagem do total de horas em que a concentração de um poluente esteve em cada classe do índice de qualidade do ar é uma métrica muito útil e utilizada em alguns destes trabalhos.

As técnicas de visualização utilizadas nos trabalhos relacionados são muito variadas e adequadas para os problemas em questão. A grande maioria utiliza gráficos de barras e gráficos de linhas, e estes são utilizados tanto em trabalhos relacionados com a mobilidade como em trabalhos relacionados com a qualidade do ar. Os mapas coropléticos e os mapas de densidade foram utilizados maioritariamente em trabalhos no contexto da mobilidade. Já os diagramas de caixa de bigodes foram utilizados, na sua maioria, em

trabalhos referentes ao impacto da pandemia na qualidade do ar. De modo a serem realizadas correlações, os autores dos trabalhos relacionados utilizaram *scatter plots*. Em trabalhos relacionados com a qualidade do ar foram utilizadas técnicas de visualizações mais complexas, nomeadamente *rose plots* e *bivariate polar plots*.

## TRABALHO REALIZADO

Este capítulo apresenta todo o trabalho realizado durante o desenvolvimento desta dissertação. Primeiramente, na secção 4.1, encontra-se detalhada a metodologia utilizada para alcançar os objetivos finais da dissertação. Seguidamente, na secção 4.2, serão apresentadas as tecnologias utilizadas para a criação de visualizações de dados, para a criação de *dashboards* e para o processamento dos respetivos dados. No relatório de qualidade dos dados, presente neste **documento**, foi visto que existem conjuntos de dados que requeriam de um processamento para melhorar a qualidade dos mesmos, então, na secção 4.3, será explicado todo o processamento feito em cada um deles. Na secção 4.3 é ainda detalhado o processo de criação de novos conjuntos de dados que foram essenciais no desenrolar do trabalho. De seguida, na secção 4.4, serão apresentados todos os *dashboards* desenvolvidos, bem como uma explicação detalhada das decisões de implementação em cada um deles. Por último, na secção 4.5, serão apresentadas algumas análises feitas, e conclusões tiradas, com os *dashboards* desenvolvidos.

### 4.1 Metodologia Utilizada

Para a realização desta dissertação começou-se, inicialmente, por definir uma metodologia que permitisse uma boa organização do trabalho. A metodologia utilizada consiste no *Cross-Industry Standard Process for Data Mining* (CRISP-DM) que, sendo um modelo de processo bastante flexível, permite a adaptação às necessidades desta dissertação. CRISP-DM é uma metodologia e um modelo de processo, utilizado em *data mining* onde são descritas as abordagens mais comuns utilizadas nesta área [25].

Esta metodologia está dividida em seis fases interligadas entre si, sendo que, devido à sua flexibilidade, é possível o salto entre etapas e a criação de um modelo que se adequa às necessidades de cada um [25]. Na figura 4.1 está presente o fluxo deste modelo de processo e as suas principais relações entre fases, e é possível constatar, através do ciclo exterior, que o processo é contínuo e não termina no final de uma iteração.

A fase de **entendimento do negócio** (*Business Understanding*) tem como principal objetivo identificar o problema a ser resolvido e entender a finalidade da análise. É, portanto,

nesta etapa que se define o contexto do problema e de que forma é que o projeto pode resolvê-lo, estabelecendo critérios de aceitação para definir o projeto como finalizado. É aqui também onde se conhecem os requisitos, as restrições e as questões que os parceiros de negócio pretendem ver respondidas no final do projeto. Por norma, termina-se esta fase com a elaboração de um plano de trabalho e com um conjunto de questões gerais que serão respondidas futuramente [25].

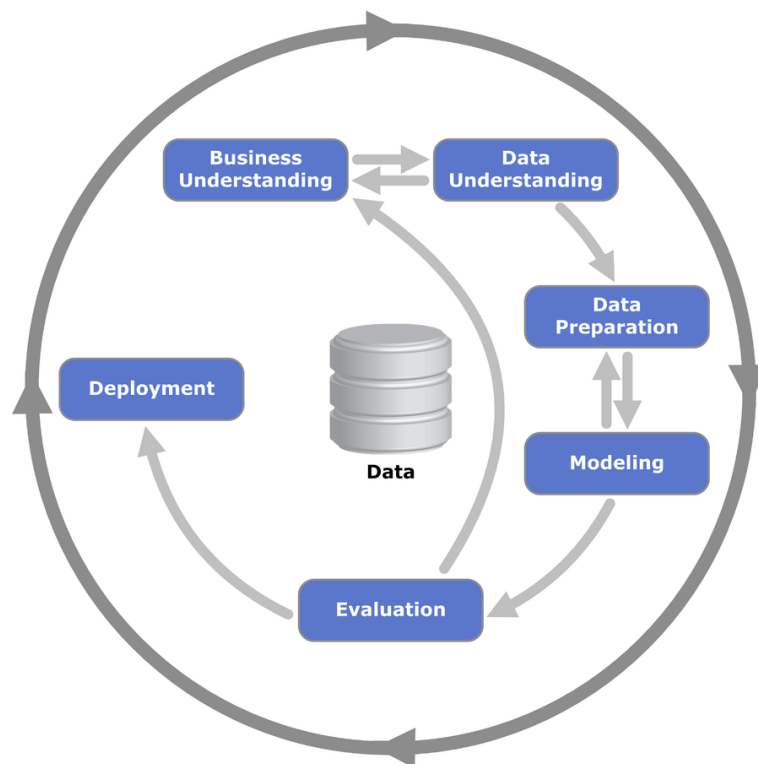


Figura 4.1: Fluxo do modelo de processo CRISP-DM.

O objetivo da fase de **entendimento dos dados** (*Data Understanding*) consiste na identificação e exploração dos dados disponíveis, bem como a verificação da qualidade dos mesmos, analisando a sua integridade e validade. São elaboradas análises preliminares onde são identificados os problemas nos dados e as suas possíveis soluções. Esta fase é fundamental, e deve ser conduzida com bastante critério, pois com ela é possível evitar possíveis problemas que possam surgir na fase seguinte - preparação dos dados - que é, tipicamente, a fase mais longa do projecto [25]. No final desta etapa, é elaborado um relatório de qualidade dos dados, organizado pelas etapas da fase de *Data Understanding*:

- Reunir os dados iniciais
- Descrição dos dados
- Exploração dos dados
- Verificar a qualidade dos dados

Quando se está na fase de entendimento dos dados é frequente surgirem dúvidas que apenas os parceiros de negócio, ou seja, os peritos na área saibam responder. Então é essencial que haja comunicação entre estas duas fases iniciais. Esta relação pode ser vista no diagrama de ciclo de vida do **CRISP-DM** presente na figura 4.1.

A fase de **preparação dos dados** (*Data Preparation*) é uma das fases mais importantes de *data mining* e também uma das que consomem mais tempo durante o projeto. É nela onde ocorre o processamento dos dados que serão utilizados nas fases seguintes, envolvendo por norma tarefas de seleção de dados, seleção de atributos, limpeza dos dados, integração de dados, criação de novos atributos, entre outras [25].

Na fase de **modelação** (*Modeling*) é onde se define qual o modelo a ser utilizado. É nesta fase que os resultados começam a dar respostas às questões gerais, colocadas durante a etapa de entendimento do negócio, através de modelação e análise estatística. Portanto, nesta fase é inicialmente escolhida a técnica de modelação e de seguida é gerado um desenho de teste onde são descritos os passos que serão tomados para testar os modelos produzidos. Seguidamente, é construído o modelo, testando sempre várias parametrizações, sendo que no final é então feita a classificação dos modelos [25].

É fundamental haver bastante interação entre as fases de modelação e de preparação dos dados, pois todo o processo é iterativo e, por vezes, é necessário regressar à fase de preparação dos dados para novas manipulações exigidas pelo modelo escolhido.

A próxima fase do **CRISP-DM** corresponde à **avaliação** (*Evaluation*) e é nesta fase onde se verifica se os critérios de sucesso do projeto, definidos anteriormente, foram atingidos com o modelo construído na fase de modelação. Ou seja, nesta etapa o modelo criado é avaliado, todas as fases efetuadas são revistas e é analisado se as questões gerais são respondidas. Por vezes, quando os critérios de sucesso do projeto não são cumpridos, é necessário voltar à etapa de entendimento do negócio, de forma a rever os conceitos e entender o que precisa ser ajustado [25].

Por fim, na última fase do processo, a fase de **desenvolvimento** (*Deployment*), é disponibilizado o modelo ao cliente para que o mesmo o possa utilizar [25].

A figura 4.2 mostra como foram divididas as principais etapas desta dissertação pelas fases da metodologia do **CRISP-DM**. Começou-se inicialmente pela fase do entendimento do negócio, que consistiu na realização de um momento de reunião com a **CML** e os *stakeholders* envolvidos, nomeadamente pessoal do departamento de mobilidade. Nesta reunião foi discutido o que se pretendia com o desenvolvimento deste trabalho e, a partir disso, foi possível fazer um levantamento de quais as questões analíticas mais gerais que serão respondidas no final do trabalho. Foram definidas 9 questões analíticas gerais, todas referentes à cidade de Lisboa. Cada uma das 9 questões analíticas mais gerais tem associado um conjunto de exemplos de questões mais específicas que lhes estão associadas. Todas as questões, gerais e mais específicas, têm um identificador, em que **QG01** significa **Questão Geral 01** e **Q01.02** significa **Questão 02** pertencente à questão geral **01**. A listagem das questões é a seguinte:

- **QG01.** Houve alterações no padrão de uso das bicicletas Gira durante a pandemia COVID-19?
  - **Q01.01.** “A taxa de ocupação ao longo das horas, na estação 103, é semelhante em 2019 e 2020 durante a altura do estado de emergência?”
  - **Q01.02.** “Relativamente ao estado de contingência, considera-se que, durante os dias de trabalho, houve alteração da hora com maior taxa de ocupação nas estações pertencentes à zona do eixo central?”
  - **Q01.03.** “Durante os estados da pandemia houve alguma alteração significativa no número diário de viagens feitas nas bicicletas da rede Gira?”
  
- **QG02.** Houve alterações no padrão de uso dos parques de estacionamento durante a pandemia COVID-19?
  - **Q02.01.** “O parque de estacionamento da Avenida Lusíada ficou, na sua maioria, mais cheio ou mais vazio após o aparecimento da COVID-19?”
  - **Q02.02.** “Relativamente à alteração média da taxa de ocupação por hora comparativamente com 14 dias atrás, qual é o comportamento observado no parque Avenida Lusíada durante o mês de junho?”
  - **Q02.03.** “Comparando a altura do estado de calamidade com o restante período temporal, considera-se que existiu alguma alteração no padrão da taxa de ocupação diária dos parques de estacionamento presentes no top 10?”
  
- **QG03.** O aparecimento da COVID-19 teve impacto nos congestionamentos?
  - **Q03.01.** “No geral, houve mais ou menos congestionamentos em 2020 na altura da COVID-19?”
  - **Q03.02.** “Comparativamente a 2019, considera-se que a densidade espacial dos congestionamentos na cidade de Lisboa se alterou durante as duas primeiras semanas do estado de calamidade?”
  - **Q03.03.** “Em 2020, durante as duas primeiras semanas do estado de emergência, existiram picos horários da média de quilómetros congestionados?”
  
- **QG04.** Quais foram as zonas da cidade de Lisboa que sofreram um maior impacto no número de congestionamentos após o aparecimento da COVID-19?
  - **Q04.01.** “Quais foram as três avenidas de Lisboa onde se sentiu um maior impacto da COVID-19 em relação ao número de congestionamentos?”
  - **Q04.02.** “Existiram mudanças significativas no número de congestionamentos diários ocorridos na Avenida da Liberdade após a declaração do estado de alerta?”
  - **Q04.03.** “Existiu alguma diferença na localização dos congestionamentos da Autoestrada do Norte (A1) durante as duas primeiras semanas do estado de emergência?”

- **QG05.** Como se comportaram os congestionamentos ao redor dos pontos de interesse após o aparecimento da COVID-19?
  - **Q05.01.** “Como se comportaram os congestionamentos em redor dos pontos de interesse da categoria ‘Saúde’ na cidade de Lisboa após o aparecimento da COVID-19?”
  - **Q05.02.** “Durante a primeira semana do estado de calamidade houve mais congestionamentos ao redor de estabelecimentos do ensino superior ou do ensino secundário?”
  - **Q05.03.** “Durante a primeira semana do estado de emergência, qual foi o dia que apresentou um maior número de congestionamentos ao redor do Hospital Santa Maria?”
  
- **QG06.** Após o aparecimento da COVID-19, existiram alterações nas concentrações dos poluentes atmosféricos medidos nas estações da cidade de Lisboa?
  - **Q06.01.** “Quais foram os três meses que, na estação da Avenida da Liberdade, apresentaram uma maior mudança no valor médio do poluente  $\text{NO}_2$  durante o ano 2020, comparativamente aos anos anteriores?”
  - **Q06.02.** “Para o conjunto de estações da Avenida da Liberdade e Entrecampos, qual foi o estado da pandemia que parece ter tido um maior impacto nos valores médios diários de  $\text{NO}_2$  em 2020?”
  - **Q06.03.** “Relativamente à estação de Entrecampos e durante o 1º estado de emergência, considera-se que a distribuição das concentrações de  $\text{NO}_2$  pelas classes do índice de qualidade do ar foram idênticas nos quatro anos?”
  
- **QG07.** O comportamento de cada poluente atmosférico foi semelhante entre todas as estações em períodos temporais anteriores e durante a pandemia?
  - **Q07.01.** “No geral, considera que o padrão da concentração do poluente  $\text{PM}_{2.5}$  é semelhante entre estações? Se sim, esse padrão manteve-se após o aparecimento da COVID-19?”
  - **Q07.02.** “Relativamente a quatro semanas a partir do dia 4 de abril de 2020, qual é a estação que mais diverge da sua média de concentração global de  $\text{NO}_2$ ?”
  - **Q07.03.** “No dia 31 de março de 2020 considera que o pico horário com maior valor da concentração do poluente  $\text{PM}_{2.5}$  é idêntico entre estações?”
  
- **QG08.** Existe alguma relação entre o comportamento dos diferentes poluentes atmosféricos?
  - **Q08.01.** “Após o aparecimento da COVID-19 e num panorama geral, qual foi a classe predominante do índice de qualidade do ar em todos os poluentes medidos nas diversas estações?”

- Q08.02. “Analisando as quatro primeiras semanas após o aparecimento da COVID-19, quais foram os poluentes que, em média, apresentaram um índice de qualidade do ar considerado apenas bom? E em que estações é que isso se verificou?”
- Q08.03. “Analisando as quatro primeiras semanas após o aparecimento da COVID-19, considera que existe alguma correlação entre os poluentes  $\text{NO}_2$  e  $\text{O}_3$  medidos na estação de Entrecampos?”
- QG09. Os congestionamentos têm impacto na concentração dos poluentes atmosféricos?
  - Q09.01. “Após a declaração do estado de emergência, qual foi o comportamento geral observado no número diário de congestionamentos e na média móvel da concentração do  $\text{NO}_2$ , comparativamente a períodos temporais anteriores?”
  - Q09.02. “Analisando três semanas após a declaração do estado de emergência, qual é a estação de qualidade do ar que apresenta uma maior média diária da concentração do poluente  $\text{NO}_2$ ?”
  - Q09.03. “Analisando oito semanas a partir do dia 20/02/2020, parece haver alguma correlação entre a concentração de  $\text{NO}_2$  e o número de congestionamentos num raio de 500 metros à volta da estação da Avenida da Liberdade?”

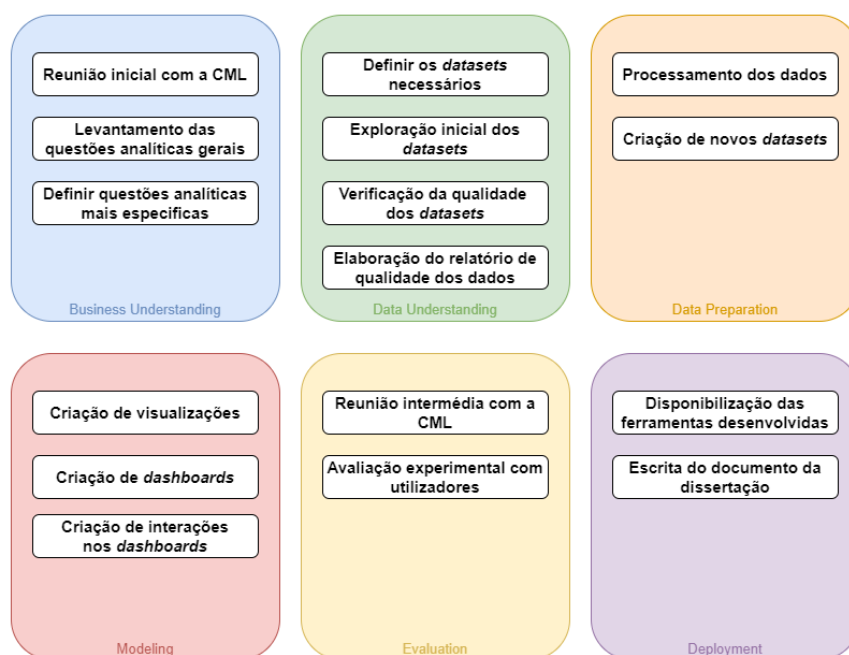


Figura 4.2: Principais etapas realizadas em cada fase da metodologia CRISP-DM.

Neste momento de reunião foram ainda dadas pistas sobre as análises que iriam ser feitas, como por exemplo, foi aconselhado definir uma área de influência (*buffer*) para cada estação de qualidade do ar, de forma a poder analisar o tráfego localmente e poder relacionar com os parâmetros medidos dos poluentes atmosféricos. Como consequência desta sugestão foi então pedido a opinião ao Professor Francisco Ferreira do Departamento de Ciências e Engenharia do Ambiente da FCT NOVA sobre qual seria o tamanho mais adequado desses *buffers*. Foi recomendado a utilização de diferentes áreas de influência para **estações de fundo** e **estações de tráfego**. Isto, pois as **estações de tráfego** localizam-se em áreas com gradientes de concentrações elevados, o que leva a serem do tipo *hotspot* pois têm áreas de representatividade mais reduzidas. As **estações de fundo**, por sua vez, têm uma maior área de representatividade, uma vez que são utilizadas para monitorizar os níveis de poluição médios, resultantes do transporte de poluentes atmosféricos de fora da zona urbana e de emissões com origem na própria cidade, não sendo diretamente influenciadas pelas fontes de emissão dominantes como o tráfego rodoviário. Assim sendo foi sugerido:

- **Buffers para estações de tráfego:** 50, 100, 250 e 500 metros;
- **Buffers para estações de fundo:** 500, 1 000, 1 500 e 2 000 metros.

Durante este primeiro momento de reunião com a CML, foi ainda sugerido trabalhar com mais detalhe em algumas zonas da cidade de Lisboa, nomeadamente algumas avenidas/eixos e entradas/saídas da cidade. Desta sugestão surgiu uma lista de oito avenidas e onze entradas e saídas da cidade de Lisboa que foram alvo de análise durante o decorso da dissertação. Nas figuras 4.3 e 4.4 é possível ver as avenidas e as entradas/saídas consideradas, sendo que a lista de todas as zonas é a seguinte:

**Avenidas/Eixos da cidade de Lisboa:**

- Alameda dos Oceanos
- Avenida 24 de Julho
- Avenida Almirante Gago Coutinho
- Avenida Almirante Reis
- Avenida da Liberdade
- Avenida das Forças Armadas
- Avenida dos Combatentes
- Eixo Central

**Entradas e Saídas da cidade de Lisboa:**

- A30/IC2
- A37/IC19
- Autoestrada Costa do Estoril (A5/IC15)
- Autoestrada do Norte (A1)
- Avenida Marginal (N6)

- Calçada de Carriche
- Eixo Norte-Sul (IP7)
- Estrada dos Cabos de Ávila (N117)
- IC16
- Ponte 25 de Abril (A2)
- Ponte Vasco da Gama (A12)

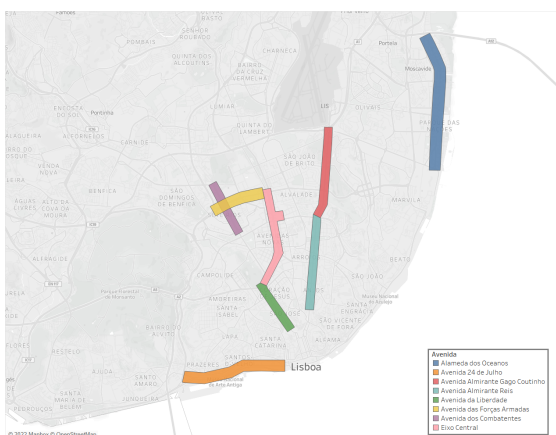


Figura 4.3: Avenidas de Lisboa.



Figura 4.4: Entradas/Saídas de Lisboa.

Após concluída a fase de identificação do problema a ser resolvido seguiu-se para a fase de entendimento dos dados. Durante esta fase foi então definido quais os conjuntos de dados necessários para a resolução do problema. E, para além dos conjuntos de dados fornecidos pela CML, foi necessário o acréscimo de dados sobre novos poluentes atmosféricos, que continham medições desde 2017 até 2020, dados das medidas governamentais implementadas durante a pandemia e dados relativos aos pontos de interesse da cidade de Lisboa. Tendo os conjuntos de dados definidos foi então conduzida uma exploração inicial de todos eles, de forma a verificar a sua qualidade para facilitar as análises futuras. Como resultado desta exploração foi elaborado um relatório de qualidade dos dados, presente neste [documento](#). Concluiu-se ainda, durante esta etapa, que era necessário dispor de dados espaciais referentes a algumas zonas da cidade de Lisboa e dados espaciais das entradas e saídas de cidade de Lisboa. Estes dados foram criados, posteriormente, durante a etapa de preparação dos dados.

Findada a fase de entendimento dos dados seguiu-se para uma das etapas mais cruciais e importantes do trabalho: a fase de preparação dos dados. Foi durante esta fase que se procedeu à preparação dos dados, isto é, à sua limpeza e processamento. Este processamento teve por base os dados vistos no capítulo 2 e os seus problemas encontrados durante a fase anterior. Foi também durante esta fase que se criaram novos conjuntos de dados essenciais para o sucesso do trabalho. Todo este processamento encontra-se detalhado na secção 4.3.

Sendo uma das vantagens do **CRISP-DM** a sua flexibilidade e adaptação às necessidades de cada trabalho, houve a necessidade de adaptar a fase de modelação. Isto deve-se ao facto do foco desta dissertação não ser o uso de técnicas de modelação, mas sim a criação de ferramentas de visualização e análise interativa de dados. Deste modo, durante esta etapa foram criadas visualizações e *dashboards* interativos.

Quando a maioria das ferramentas já se encontravam desenvolvidas surgiu um primeiro momento da fase avaliação. Este baseou-se na realização de uma reunião intermédia com a **CML** e os representantes do departamento de mobilidade, tendo como propósito apresentar os *dashboards* desenvolvidos até à data e receber *feedback* sobre os mesmos. Após o término do desenvolvimento de todas as ferramentas foi então realizada uma avaliação experimental com utilizadores. Os participantes desta avaliação tiveram que responder a um conjunto de questões. De modo a obterem as respostas corretas às questões, foi necessário a utilização dos *dashboards* desenvolvidos e das suas interações.

Por último, na fase de desenvolvimento, foi organizado todo o trabalho desenvolvido e iniciado o processo de escrita do documento da dissertação. Para além disso, as ferramentas desenvolvidas foram disponibilizadas, através do *Tableau Online*, aos *stakeholders* interessados.

## 4.2 Ferramentas Utilizadas

No decorrer desta dissertação utilizaram-se várias tecnologias que foram fundamentais para atingir os objetivos finais do trabalho. Nesta secção será explicado o porquê da escolha das ferramentas utilizadas, detalhando cada uma delas e enumerando quais as suas vantagens, o seu modo de utilização, as suas funcionalidades, entre outras.

Uma das principais finalidades destas ferramentas é o desenvolvimento de visualizações interativas de dados de forma rápida, intuitiva e fácil, onde é possível a utilização de grandes quantidades de dados sem que o desempenho das visualizações seja afetado. Para além disso, também foram utilizadas tecnologias que permitissem o processamento de dados, com o intuito de preparar os dados para as análises futuras.

### 4.2.1 Tableau Desktop

O *Tableau Desktop*, sendo uma ferramenta concebida para a criação de visualizações de dados e *dashboards*, permite a descoberta e exploração de dados de forma interativa. Esta foi a tecnologia mais essencial para o desenvolvimento desta dissertação. A sua escolha incide sobre diversas razões, mas principalmente sobre o facto deste *software* ser de uso bastante intuitivo, utilizando maioritariamente *drag and drop* e não ser necessário a escrita de muito código, o que torna a implementação de visualizações bastante fácil. Outro fator que contribuiu para a sua escolha foi o facto de ser uma ferramenta completa de análise e visualização de dados que tem uma forte componente interativa, sendo, deste modo, uma mais valia na criação de *dashboards*. A análise realizada a diversas ferramentas de

visualização de dados, feita pela colega de mestrado Maria Ferreira [19], foi outro aspeto que contribuiu para a escolha do *Tableau Desktop*, visto esta ferramenta ser a que apresenta mais vantagens neste tipo de trabalho.

Nesta ferramenta é possível fazer a conexão a dados que se localizem na *cloud* ou que estejam armazenados localmente, sendo que pode haver várias conexões a diversos conjuntos de dados. Ao estabelecer uma ligação a um conjunto de dados é necessário averiguar se os tipos dos dados estão corretos. Caso não estejam, basta identificar qual o significado de cada campo, isto é, se é uma *string*, número (decimal ou inteiro), booleano, data, data e hora ou campo geográfico. No caso deste último, é preciso ainda indicar qual o seu papel geográfico, por exemplo, se se refere a uma cidade, latitude, longitude, etc. O *Tableau Desktop* consegue-se também conectar a ficheiros geográficos, como é o caso de ficheiros *GeoJSON* e *Shapefile* [5].

Nesta ferramenta existem dois tipos diferentes de campos: as dimensões e as medidas, sendo que cada campo recebe automaticamente um tipo de dados, podendo ser modificado posteriormente. Uma dimensão corresponde a um campo que contém valores qualitativos, como por exemplo, nomes, datas ou dados geográficos. Este tipo de campo é utilizado, por norma, para categorizar, segmentar e revelar os detalhes dos dados, afetando o nível de detalhe da visualização. Por outro lado, as medidas dizem respeito a valores numéricos e quantitativos que podem ser medidos ou agregados. Existe também a noção de campos discretos e contínuos. Campos discretos são representados pelo *software* com uma cor azul, sendo que este tipo de campos são tratados como finitos e, geralmente, adicionam um cabeçalho discreto à visualização. Os campos contínuos, representados pela cor verde, por norma adicionam eixos à visualização [5].

Caso os dados a que se está conectado não incluam todos os campos necessários para a criação de uma visualização, é sempre possível criar novos campos no *Tableau*. Estes são chamados de campos calculados, e são criados a partir de fórmulas que usam os campos já existentes na fonte de dados [5].

A criação de visualizações no *Tableau Desktop* é muito fácil e intuitiva. De modo que, este *software* apresenta sugestões de quais os tipos de gráficos que melhor se ajustam aos campos selecionados. Sendo assim apenas necessário selecionar os campos que se pretende adicionar à visualização e escolher um gráfico dos sugeridos. A outra forma de criar visualizações é através de *drag and drop*, arrastando os campos do painel dos dados até à divisória pretendida, ou seja, a divisória das linhas ou das colunas. Com o auxílio dos atributos do conjunto de dados é ainda possível mapear variáveis visuais, codificando os dados com cor, tamanho, forma, texto e outras propriedades que adicionam contexto, detalhe e significado às marcas na visualização. Pode-se ainda adicionar filtros, incorporar parâmetros à visualização e usar o controlo de páginas, sendo que esta última tem como propósito mostrar mudanças ao longo do tempo de forma interativa. Sempre que for necessário mostrar mais detalhes sobre os dados pode-se adicionar *tooltips* aos gráficos, desta forma o utilizador coloca o ponteiro do rato sobre a marca e a *tooltip* exhibe os detalhes relevantes sobre a mesma [5].

Para criar *dashboards* basta colocar as visualizações no sítio desejado, e para isso é necessário ter criado primeiramente as visualizações que irão fazer parte da mesma. Para aprimorar os *dashboards* e adicionar interatividade nos mesmos pode-se adicionar ações, escolhendo as visualizações de origem e as de destino. Estas ações consistem em filtro, destaque, mudança de parâmetro, mudança de visualização, ir para um URL e mudança de um conjunto de valores [5].

É importante salientar que este *software* não é gratuito, tendo um custo de 70\$ mensais por utilizador. Todavia, são disponibilizadas licenças gratuitas renováveis, de um ano, para estudantes e para investigadores que dão acesso a todas as ferramentas do *Tableau*.

#### 4.2.2 Tableau Prep Builder

O *Tableau Prep Builder* pertence à mesma empresa que criou o *Tableau Desktop*, sendo que foi projetado para facilitar e tornar intuitiva a preparação dos dados. Este *software* foi escolhido pois oferece uma forma direta e visual de juntar, limpar e estruturar os dados, fazendo com que o início das análises comece com maior rapidez. O facto de não ser necessária a escrita de muito código para efetuar alterações nos dados também contribuiu positivamente para a escolha desta ferramenta.

Tal como no *Tableau Desktop*, no *Tableau Prep Builder* é também possível conectar a dados que estejam localizados na *cloud* ou armazenados localmente. Feita a conexão pode-se começar a criar uma cadeia de etapas, mais conhecidas como fluxos de trabalho. Estas etapas representam alterações nos dados, sendo que existem etapas de limpeza, agregação, trocar colunas para linhas ou vice-versa, união de colunas, união de linhas, *script* e saída. Esta última etapa, a saída, proporciona a exportação do resultado do fluxo num novo conjunto de dados processado e limpo [4].

Esta ferramenta permite ter uma imagem integral dos dados, isto é, é possível consultar e alterar os tipos de dados de cada campo, ver detalhes do tamanho dos dados, ver a distribuição de valores e valores exclusivos de cada campo, pesquisar campos e valores, classificar valores e campos, realçar campos e valores num fluxo, entre outros [4].

#### 4.2.3 Tableau Online

O *Tableau Online* pertence à mesma empresa que criou o *Tableau Desktop* e o *Tableau Prep Builder*. Esta tecnologia é uma plataforma de análise, totalmente hospedada na *cloud*, que permite partilhar o conteúdo criado no *Tableau Desktop*. A sua escolha incide na necessidade de um meio de partilha dos *dashboards* desenvolvidos para receber *feedback* sobre os mesmos. Através da publicação *online* dos *dashboards*, os interessados podem ver todo o seu conteúdo e interagir facilmente com o mesmo. O conteúdo é acedido via *browser*, não sendo necessário a instalação de novas ferramentas no computador. Porém, é necessário a criação prévia de contas, para que as entidades que irão explorar os *dashboards* tenham permissão para tal.

Nesta ferramenta também é possível criar e editar visualizações sendo, por isso, muito semelhante ao *Tableau Desktop*, porém o *Tableau Online* é completamente hospedado em *cloud*, podendo correr em qualquer *browser*. É importante salientar que o *Tableau Online* apresenta algumas limitações quando comparado à versão *Desktop*. Na versão *online* existem algumas funcionalidades que não se encontram tão bem desenvolvidas como na versão *desktop*, quando os dados são muito grandes a exportação dos *dashboards* para a versão *online* não é tão linear e a qualidade dos *dashboards*, por vezes, é menor na versão *online*.

### 4.2.4 QGIS

O *software* QGIS é um Sistema de Informação Geográfica (SIG) de código aberto que funciona em múltiplos sistemas operativos. Este é um sistema gratuito de visualização, edição, manipulação e análise de dados georreferenciados [38].

A conexão aos dados pode ser feita com dados armazenados localmente ou com dados presentes em bases de dados ou servidores externos, facilitando desta forma a elaboração de mapas. São suportados dados em formatos vetoriais, *raster*, *csv*, *shapefile*, *GeoJSON*, entre outros. E, a partir destes dados, é possível gerar mapas com diversas camadas, sendo que cada camada pode, ou não, ter uma diferente projeção. Estas camadas são geradas a partir dos dados e podem ser dimensionadas, medidas, unidas e remodeladas. O QGIS disponibiliza ainda *plugins* que acrescentam funcionalidades extra ao *software* [38].

Uma das funcionalidades que levou à escolha desta ferramenta foi a possibilidade de inserir dados georreferenciados no mapa e, através de algoritmos de interseção, identificar quais os pontos que se encontram dentro do polígono, ou quais as linhas que interseccionam o polígono. Quando o algoritmo acaba de correr pode-se ver quais os pontos resultantes da interseção no mapa, sendo que é criada uma nova camada com estes pontos e a mesma pode ser exportada para um ficheiro espacial ou *csv*. Desta forma a interseção espacial torna-se mais simples e leva a que não seja necessário escrever código. Ao longo da dissertação esta foi a funcionalidade mais utilizada do QGIS, sendo que facilitou imenso o processamento de dados e a criação de novos conjuntos de dados.

### 4.2.5 Spyder

*Spyder* é um ambiente de desenvolvimento integrado (IDE) gratuito do *software* Anaconda. Este IDE é totalmente escrito em *python* e foi concebido para o desenvolvimento de código *python*. Este facilita a criação de *scripts*, *debugging* e visualização de gráficos. Nesta dissertação foi utilizado durante o processamento de dados e a criação de novos conjuntos de dados. A sua utilização baseia-se em conhecimentos prévios deste IDE e, também, no facto da linguagem *python* ser bastante simples e flexível, tendo inúmeras bibliotecas que podem ser usadas no decorrer da escrita do código, o que foi bastante útil para o trabalho realizado.

O IDE *Spyder* apresenta várias componentes, nomeadamente, um editor que é o sítio onde o código é escrito, uma consola *IPython* que mostra a execução do código ou o resultado final após a execução do código, um explorador de variáveis que permite gerir interativamente os objetos criados após a execução do código, um *debugger* que permite localizar cada passo da execução do código de forma interativa e ainda um menu de ajuda. Conta ainda com a integração de várias bibliotecas *python* populares, como é o caso da biblioteca *pandas*, *NumPy* e *SciPy*, sendo ainda possível estender todas as funcionalidades através de *plugins* [44].

#### 4.2.6 MySQL Workbench

O *MySQL Workbench* é uma ferramenta gratuita que permite o desenho e a gestão de bases de dados. Da sua lista de funcionalidades destacam-se a criação de diagramas *enhanced entity-relationship* (EER), a criação de *scripts* SQL, o facto de ser possível ver informações sobre o objeto seleccionado, a gestão de ligações a servidores MySQL, a administração de utilizadores, entre outras [32].

O *MySQL Workbench* foi utilizado para realizar ordenações nos dados. Isto, pois durante a realização de algum do processamento nos dados em código *python* era favorável ter os conjuntos de dados ordenados. E quando estamos perante conjuntos de dados muito grandes torna-se muito mais rápido executar uma operação de ordenação através de uma *query* SQL do que num *script python*. Isto acontece pois as bases de dados estão otimizadas para trabalhar com dados e, dependendo da operação, irão obter, na maioria das vezes, um melhor desempenho.

A escolha desta ferramenta deveu-se ao facto de ser gratuita e ser de fácil utilização, contendo todas as funcionalidades necessárias para a realização deste trabalho.

#### 4.2.7 Geojson.io

O *geojson.io* é uma ferramenta *online* utilizada para a edição e criação de dados espaciais. Durante esta dissertação foi utilizada, maioritariamente, para a criação de ficheiros espaciais, como é o caso do ficheiro das zonas da cidade de Lisboa.

Esta ferramenta, de uso simples, rápido e intuitivo, permite criar e editar o GeoJSON através de uma interface gráfica de um mapa e através do código GeoJSON. É possível importar dados localizados localmente ou localizados num repositório *GitHub*, sendo que os dados podem estar no formato GeoJSON, CSV, KML, GPX, entre outros e, após editados podem ser exportados. Estas características, aliadas ao facto de ser uma ferramenta gratuita, levaram à escolha desta ferramenta durante a realização do trabalho.

Quando se pretende criar um ficheiro espacial basta clicar num dos utensílios de desenho posicionados sobre mapa, podendo-se desenhar pontos, polígonos, linhas e rectângulos. Após as formas estarem desenhadas existe a possibilidade de adicionar informação

a cada característica. Para isso basta clicar na forma e editar as características, ou simplesmente alterar a estrutura JSON gerada. Quando o processo de criação estiver terminado basta exportar os dados para um dos diversos formatos permitidos.

### 4.3 Processamento dos Dados

Os dados disponíveis para a realização desta dissertação encontram-se sucintamente sintetizados no capítulo 2. Sendo que, durante o processo de exploração dos dados e verificação da sua qualidade, foram levantados problemas em alguns dos conjuntos de dados. Toda a exploração feita e os problemas levantados podem ser consultados no relatório de qualidade dos dados presente neste [documento](#).

Após esta exploração, que está enquadrada na fase de entendimento dos dados do CRISP-DM, passou-se para a fase de preparação dos dados. Esta fase teve como intuito melhorar a qualidade dos conjuntos de dados, para que as análises feitas sejam mais precisas. Nas subsecções seguintes será detalhado o processamento feito nos conjuntos de dados que necessitavam. Será ainda detalhado o processo de criação de novos conjuntos de dados essenciais para o desenvolvimento do trabalho.

#### 4.3.1 Dados das Estações Gira

Este conjunto de dados, disponibilizado pela Câmara Municipal de Lisboa, diz respeito ao estado de cada estação Gira, mais concretamente, apresenta medições da ocupação de cada estação num determinado momento. Existem dados entre os anos 2019 e 2020, sendo que o último registo é no dia 23 de setembro de 2020.

Tabela 4.1: Colunas iniciais dos dados das Estações Gira.

Atributo	Descrição
ID	Identificador único da estação Gira
Designação Comercial	Designação da estação Gira
Localização	GeoJSON com a localização da Estação Gira
Posição	GeoJSON com a localização da estação Gira
Estado	Estado atual da estação Gira
Número de Bicicletas	Número de bicicletas existentes na estação Gira
Número de Docas	Número de docas existentes na estação Gira
Número de Docas Vazias	Número de docas vazias existentes na estação Gira
Rácio	Taxa de ocupação da estação Gira
Data	Data e hora da atualização da estação Gira
Data Update	Data e hora da próxima atualização da estação Gira
Serviços Disponíveis	Tipo de serviços disponíveis na estação Gira

Na tabela 4.1 estão representados todos os atributos deste conjunto de dados antes de lhes ter sido aplicado qualquer processamento. E, tal como é possível constatar, cada

registo apresenta informação espacial, temporal, informação do seu estado e das suas características.

Após a exploração inicial, detalhada no relatório de qualidade dos dados, foi possível obter um conhecimento mais profundo sobre os dados e verificar a sua qualidade. Durante o processamento dos dados foram solucionados os problemas encontrados e também criados novos atributos que serão interessantes utilizar para o desenvolvimento das ferramentas. Para os problemas mais fáceis de resolver foi utilizado o *Tableau Prep Builder* e para os problemas que requeriam um processamento mais exigente foi utilizado o *Spyder IDE*, o *MySQL Workbench* e o QGIS, sendo que estes últimos auxiliaram no desenvolvimento de código *python*, na ordenação do conjunto de dados e na interseção de ficheiros espaciais, respetivamente.

Foi decidido remover colunas que não mostravam ser interessantes para propósitos de análise futura. Uma dessas colunas foi a dos serviços disponíveis, pois esta apresentava todos os seus valores a nulo. A coluna da data e hora do próximo *update* também foi removida porque não acrescentava valor às análises que viriam a ser feitas. Visto existirem duas colunas com a mesma informação espacial relativa às estações Gira, foi também decidido remover uma delas.

Um dos problemas encontrados durante a fase de entendimento dos dados foi o facto de alguns registos apresentarem dois identificadores diferentes, um no campo "ID" e outro no campo "Designação Comercial". Então foi necessário processar esses registos e, visto que o identificador correto estava presente no campo "Designação Comercial", bastou substituir o valor no campo "ID" para o valor correto. Todavia, os registos associados ao identificador "1234" foram descartados pois aparentavam ser erros de inserção. Outro problema encontrado foi o facto de existirem estações Gira que continham números bastante reduzidos de registos e, dado o objetivo desta dissertação, não serviriam para propósitos de análise futura. Deste modo, foram então removidas as entradas referentes à estação 113 e a estação 207, ficando-se com um total de 83 estações Gira neste conjunto de dados. Constatou-se também que os valores do atributo "Designação Comercial" apresentavam problemas de escrita, estando um mesmo valor redigido de várias formas diferentes. Isto levava a que houvesse uma duplicação dos valores para cada estação, então o problema foi corrigido processando cada valor e escrevendo-o da forma correta.

Foram encontrados ainda problemas de completude no atributo "Rácio", o que significa que este continha entradas com o seu valor a nulo. Contudo, visto que este atributo é calculado através do número de bicicletas e de docas, este problema foi facilmente corrigido.

De forma a simplificar as análises, e visto que o *Tableau Desktop* mapeia coordenadas geográficas no mapa, foi extraído do campo da informação espacial as coordenadas da estação Gira. Este campo inicialmente continha uma *string* com uma estrutura GeoJSON da localização da estação e após a divisão das coordenadas geográficas ficou-se com dois campos distintos: o campo da latitude e o campo da longitude da estação Gira. Após este passo ter sido executado, foi descoberto outro problema nos dados. Este problema

consistia que, para uma estação, existiam diferentes valores nestes campos de informação espacial. Contudo, estes valores apenas diferiam, na sua maioria, no número de casas decimais apresentadas. Então foi necessário processar estes dados para corrigir este problema e cada estação Gira ficar apenas com um valor único de localização.

No que concerne aos dados temporais, foram também encontrados alguns problemas na qualidade dos mesmos, sendo que estes problemas exigiram um nível de processamento mais profundo e demorado. Começou-se, primeiramente, por fazer uma ordenação ascendente de todo o conjunto de dados no *software MySQL Workbench*. Sendo que esta ordenação foi realizada por identificador único da estação Gira e por data e hora. O objetivo desta ordenação recai no facto dos *scripts python*, feitos para o processamento dos problemas nos dados temporais, terem um aumento da sua eficiência e rapidez quando os dados se encontram ordenados. Visto que o *MySQL Workbench* está otimizado para trabalhar com dados, a operação de ordenação mostrou-se mais rápida neste *software* do que realizada em código *python*.

Tendo os dados ordenados, seguiu-se para a resolução do problema encontrado de haver vários registos repetidos, isto é, para uma mesma estação existiam, por vezes, várias medições de ocupação na mesma hora exata. O processamento deste problema teve duas resoluções distintas: uma para quando as medições era exatamente iguais, tratando-se, deste modo, de uma duplicação de linhas; e outra quando as medições eram diferentes. Quando as medições eram exatamente iguais, existindo apenas uma duplicação de linhas, foram apagadas as linhas a mais e deixado apenas um registo para cada estação a uma data e hora específica. Quando existiam várias medições de uma estação à mesma hora, porém com valores desiguais de ocupação, foi decidido que o valor da ocupação seria a moda dos valores. Foi escolhido a moda pois, por norma, a grande maioria dos valores de ocupação eram os mesmos, havendo apenas alguns que eram bastante diferentes dos restantes, aparentando serem erros ocorridos no sensor ou durante a inserção dos dados na plataforma.

Outro problema encontrado na qualidade destes dados foi o facto da diferença temporal entre registos consecutivos não ser sempre a mesma, existindo apenas 60% dos registos com diferença de 20 minutos entre si. E, visto estes dados serem referentes a medições feitas em cada estação Gira, foi decidido fazer um processamento para tentar atenuar este problema. Deste modo, a primeira coisa a ser feita foi a criação de um novo atributo que consistia numa data e hora, sendo que as horas se encontravam em intervalos de 20 em 20 minutos. Um exemplo de valores presentes neste campo são: “01/01/2019 11:00:00”, “01/01/2019 11:20:00”, “01/01/2019 11:40:00”, “01/01/2019 12:00:00”, e por aí adiante. Para a criação deste atributo foram percorridas todas as entradas do conjunto de dados e foi-lhes associado o intervalo de 20 minutos mais próximo. Ou seja, uma medição que tivesse sido feita no dia 01/01/2019 às 11:08:00 horas teria um valor de “01/01/2019 11:00:00” neste novo atributo. A decisão da criação de intervalos de 20 minutos consistiu no facto que, aproximadamente, 60% das medições feitas distam 20 minutos da medição seguinte.

Após ter sido criado este atributo, faltava ainda resolver o problema dos seus valores puderem não ter apenas um valor exclusivo de ocupação. Então, prosseguiu-se para a realização de um código *python* que teve como objetivo agrupar os valores de medições de uma estação Gira dentro de um intervalo de 20 minutos, de forma a que cada intervalo tivesse apenas um valor exclusivo de ocupação para uma estação. Para calcular qual seria o valor de ocupação em cada intervalo utilizou-se a média ponderada pelo tempo entre observações, desta forma é obtido um valor de ocupação mais preciso. Assim, considerando  $i$  como sendo um registo,  $n$  como sendo o número total de registos dentro de um intervalo de tempo,  $dataHora$  como sendo a data e a hora a que foi feita a medição e  $ocupacao$  o valor medido de ocupação, temos as seguintes fórmulas:

$$peso(i) = |dataHora_i - dataHora_{i+1}| \quad (4.1)$$

$$media\ com\ peso(i) = \frac{ocupacao_i + ocupacao_{i+1}}{2} * peso \quad (4.2)$$

$$ocupação\_final = \frac{\sum_{i=1}^{n-1} media\_com\_peso(i)}{\sum_{i=1}^{n-1} peso(i)} \quad (4.3)$$

A fórmula 4.2, que representa o passo intermédio neste cálculo, determina o valor médio de ocupação entre duas medições e, visto ter-se optado por utilizar uma média ponderada, é feita uma multiplicação por um fator *peso*. Este peso é o espaço temporal entre dois registos consecutivos, e a sua fórmula encontra-se representada em 4.1. Após ter sido calculado o passo intermédio entre todos os registos dentro de um intervalo temporal, pode-se, finalmente, calcular a ocupação final que ficará nesse intervalo. Isto é feito através da média ponderada, representada na fórmula 4.3.

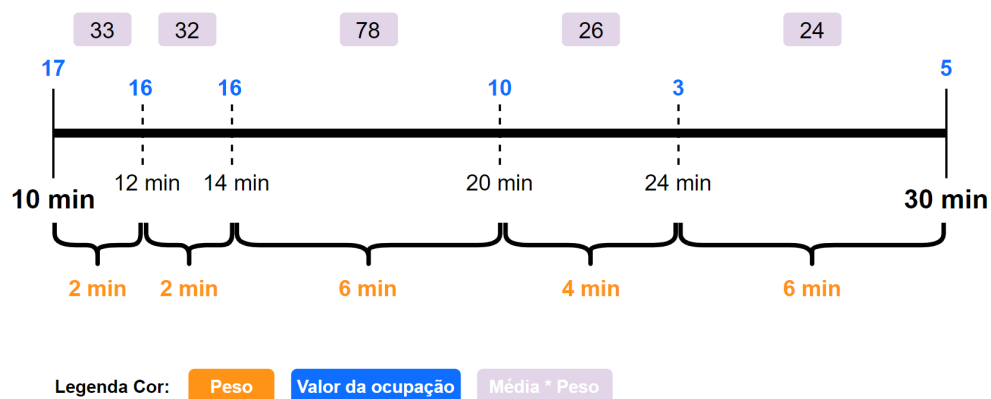


Figura 4.5: Exemplo de como é calculado os valores de ocupação das estações Gira.

Exemplificando, supondo que se quer saber qual a ocupação única representativa do intervalo temporal apresentado na figura 4.5, no qual cada traço representa um registo de

ocupação. Utilizando a fórmula 4.1 são calculados os pesos entre os registos consecutivos deste intervalo, por exemplo,  $|14 - 20| = 6$  min. Sendo que todos os valores dos pesos se encontram mapeados com a cor laranja na figura 4.5. De seguida é então calculado a média da ocupação entre registos multiplicada pelos pesos atribuídos, por exemplo,  $\frac{16+10}{2} * 6 = 78$ . Sendo que os valores deste passo intermédio estão mapeados na figura 4.5 através da cor rosa. Por fim, é então calculada a média ponderada da ocupação, que é neste caso:  $\frac{33+32+78+26+24}{2+2+6+4+6} = 9.65$ .

Findado este processamento, a maioria dos registos ficaram com um espaçamento entre medições de 20 minutos e com apenas um valor exclusivo de ocupação num intervalo temporal. Contudo, foi ainda necessário resolver o problema de haver falhas nos dados. Para isso, adicionou-se novas linhas sempre que após 20 minutos uma estação não tivesse registo de ocupação. Todavia, se a distância entre registos fosse superior a 24 horas não foram adicionadas linhas e foi considerado falha nos dados. O valor da ocupação destas novas linhas foi calculado fazendo interpolação de valores. Os restantes atributos, visto que se tratam de características da estação Gira, foram adicionados tendo em conta os restantes registos. Assim, este conjunto de dados apresenta agora apenas uma medição num espaçamento de 20 minutos em cada estação Gira. Esta característica apenas não se verifica quando existem falhas nos dados, ou seja, ausência de registos de medições há mais de 24 horas.

Para o desenvolvimento de alguns *dashboards* foi fundamental a criação de novos atributos nos conjuntos de dados. Nos dados das Gira houve a necessidade de ter uma coluna que indicasse qual a zona, neste caso avenida/eixo, a que pertencia a estação, sendo que a lista de zonas se encontra na secção 4.1. Para isto foi necessário, primeiramente, criar um ficheiro GeoJSON, através da ferramenta *geojson.io*, que contém polígonos para cada uma das zonas. O processo de identificar qual a zona correspondente a cada uma das 83 estações foi realizado no QGIS. Neste *software* começou-se por conectar o ficheiro espacial das zonas e o conjunto de dados das Gira e criaram-se duas camadas, uma com polígonos, ou seja as zonas de Lisboa, e outra com pontos, ou seja a localização das estações. De seguida, foi calculada a intersecção das duas camadas e, desta forma, cada estação ficou etiquetada com a zona correspondente. As estações que não pertenciam a nenhuma avenida/eixo foram etiquetadas com a zona “Outras”.

Foi criado ainda um novo campo, através de código *python*, que contém a diferença entre a taxa de ocupação do próprio registo com a taxa de ocupação de um registo 28 dias (4 semanas) anterior a esse. Isto foi útil pois nas ferramentas desenvolvidas conseguiu-se representar as mudanças de comportamento das estações, isto é, se existiu uma tendência de subida e/ou descida da sua taxa de ocupação ao longo das horas.

Por fim, sentiu-se a necessidade de criar um novo atributo que permitisse quantificar a diferença que houve no comportamento horário de uma estação Gira durante os estados da pandemia em 2020, comparativamente à mesma altura do respetivo estado em 2019. Isto permite mostrar nos *dashboards* criados, rapidamente, quais são as estações Gira que tiveram maiores alterações de comportamento, ao longo dos estados da pandemia, sem

ter que escolher estação a estação para efetuar essa verificação. Para isto, foi calculado em cada estação, em ambos os anos, uma média da sua taxa de ocupação por horas em todos os períodos temporais correspondentes aos diferentes estados da pandemia, tal como é possível ver no primeiro passo da figura 4.6. Para cada estação, em cada estado e em cada ano, calculou-se a variação da taxa de ocupação com a taxa de ocupação da hora anterior, exemplificado através das setas do primeiro passo da figura 4.6, sendo que o resultado se encontra no segundo passo da figura. Após isso, comparou-se então estados iguais de anos diferentes. Para tal, a cada valor de 2020 subtraiu-se o valor do mesmo estado de 2019, sendo que o resultado se encontra representado no terceiro passo da figura 4.6. No final, somou-se, em absoluto, todos estes valores horários de 2020 e obteve-se assim um atributo representativo da mudança ocorrida comparativamente a 2019, visível no último passo da figura 4.6. Um número mais alto significa que existiu uma grande diferença, e um número mais baixo significa que a diferença não foi tão sentida.

1º - Taxa de Ocupação	<b>Ano</b>	<b>Estado</b>	<b>ID Gira</b>	<b>0h</b>	<b>1h</b>	...	<b>22h</b>	<b>23h</b>
	2019	Emergência	101	22%	25%	...	80%	72%
	2020	Emergência	101	57%	52%	...	40%	43%
2º - Variação da taxa de ocupação com a hora anterior	<b>Ano</b>	<b>Estado</b>	<b>ID Gira</b>	<b>0h</b>	<b>1h</b>	...	<b>22h</b>	<b>23h</b>
	2019	Emergência	101	0%	3%	...	1%	-8%
	2020	Emergência	101	0%	-5%	...	4%	3%
3º - Diferença entre a variação da taxa com o ano anterior	<b>Ano</b>	<b>Estado</b>	<b>ID Gira</b>	<b>0h</b>	<b>1h</b>	...	<b>22h</b>	<b>23h</b>
	2020	Emergência	101	0%	-8%	...	3%	11%
								(em absoluto)
4º - Quantificação da mudança ocorrida	<b>Ano</b>	<b>Estado</b>	<b>ID Gira</b>	<b>Mudança</b>				
	2020	Emergência	101	67				

Figura 4.6: Exemplo de como é calculado o atributo que quantifica a mudança ocorrida nos estados da pandemia das estações Gira.

A decisão de comparar a variação na taxa de ocupação com a hora anterior ao invés de comparar o valor exato da taxa de ocupação consistiu no facto que podem existir vários fatores externos que explicassem a subida ou descida da taxa de ocupação, como por exemplo, mudança do rebalanceamento das estações. Ao comparar a variação da taxa de ocupação tem-se um indicativo mais preciso do comportamento da estação, verificando as alturas de chegadas e partidas de bicicletas e não os seus valores exatos de ocupação.

A tabela 4.2 apresenta os atributos presentes no *data source* do *Tableau Desktop* após o término do processamento feito a este conjunto de dados. Nesta tabela as linhas coloridas a cinzento representam os novos atributos criados durante o processamento. Este conjunto de dados ficou com um total de 3 244 007 linhas.

Tabela 4.2: Principais colunas dos dados das Estações Gira após processamento.

Atributo	Descrição
ID	Identificador único da estação Gira
Designação Comercial	Designação da estação Gira
Longitude	Longitude da estação Gira
Latitude	Latitude da estação Gira
Estado	Estado atual da estação Gira
Número de Bicicletas	Número de bicicletas existentes na estação Gira
Número de Docas	Número de docas existentes na estação Gira
Número de Docas Vazias	Número de docas vazias existentes na estação Gira
Rácio	Taxa de ocupação da estação Gira
Data-Hora	Data e hora (20 em 20 minutos) do registo de ocupação da estação Gira
Zona	Nome da zona a que pertence a estação Gira
Geometria Zona	Geometria da zona (em formato GeoJSON)
Diferença na taxa	Diferença com a taxa de ocupação de 28 dias atrás
Mudança de comportamento	Métrica que quantifica a diferença no comportamento da estação Gira num estado da pandemia

\* Novos atributos coloridos a cinzento.

### 4.3.2 Dados dos Parques de Estacionamento

Este conjunto de dados, disponibilizado pela Câmara Municipal de Lisboa, é referente ao estado da ocupação de alguns parques de estacionamento da cidade de Lisboa. Sendo que o primeiro registo é no dia 8 de agosto de 2019 e o último registo é dia 30 de junho de 2020.

Tabela 4.3: Colunas iniciais dos dados dos Parques de Estacionamento.

Atributo	Descrição
ID Parque	Identificador único do parque de estacionamento
ID Entidade	Identificador único da entidade do parque de estacionamento
Nome	Nome do parque de estacionamento
Tipo	Tipo do parque de estacionamento
Ocupação	Número de carros presentes no parque de estacionamento
Capacidade Máxima	Capacidade máxima do parque de estacionamento
Posição	GeoJSON com a localização do parque de estacionamento
Data de Ocupação	<i>Timestamp</i> da medição da ocupação do parque de estacionamento
Data Hora	Data e Hora da medição da ocupação do parque de estacionamento

Num geral, a premissa deste conjunto de dados é muito semelhante à do conjunto de dados das estações Gira, pois ambos dizem respeito ao estado da ocupação de algo. Desta forma, este *dataset* apresenta muitos dos problemas encontrados no *dataset* das estações Gira, como por exemplo, o facto dos valores dos atributos temporais não se encontrarem sempre à mesma distância temporal do registo consecutivo

Este conjunto de dados contém colunas com informação temporal, espacial e informação sobre o seu estado e as suas características, sendo que todas se encontram listadas na tabela 4.3. Estes atributos foram alvos de uma exploração inicial, detalhada no relatório de qualidade dos dados, e como resultado dessa exploração foram encontrados alguns problemas nos dados que foram resolvidos durante a etapa de processamento dos mesmos.

Durante o processamento dos dados foram removidas linhas que apresentavam todos os seus valores a nulo, linhas duplicadas e linhas que tinham o valor “Piscinas Campo de Ourique” no nome do parque, pois, em Lisboa, não existe nenhum parque de estacionamento com esse nome. Decidiu-se também apagar o atributo que contém o *timestamp* a que foi feita a medição da ocupação do parque, pois o mesmo continha imprecisão nos seus valores e, visto existir um atributo com a data e hora da medição, não existiam problemas em apagar esta coluna.

Existiam parques de estacionamento que apresentavam dois valores diferentes para o atributo “Capacidade Máxima” e, visto que cada parque tem apenas uma capacidade máxima, foram substituídos os valores que se encontravam errados pelos valores corretos. O atributo espacial apresentava problemas semelhantes, ou seja, existiam parques com duas localizações. Sendo que este problema foi resolvido de forma idêntica ao problema da capacidade máxima.

Outro problema encontrado nos dados foi o facto do atributo “Ocupação” apresentar valores negativos, ou seja, valores inválidos. Dado que estes valores apareciam pontualmente nos dados e eram uma percentagem não significativa, decidiu-se remover essas entradas. Observou-se ainda que alguns parques de estacionamento apresentavam quase sempre o mesmo valor de ocupação em todo período temporal abrangido por este *dataset*, o que se traduz em valores irrealistas neste contexto. Os parques da entidade 2 e 5, ou seja, da Empark e Dolce Vita Monumental, apresentam todos este comportamento fora do normal. Este problema aconteceu igualmente noutros parques de estacionamentos, contudo cobria um período temporal mais reduzido. Verificou-se que a maioria dos parques apresentava este comportamento nas últimas semanas do mês de março e iniciais de abril de 2020, que correspondem ao início da pandemia COVID-19 em Portugal e onde foi suspenso o pagamento do estacionamento na via pública de Lisboa e os residentes também puderam estacionar gratuitamente em alguns parques [28]. Foi então decidido apagar todos os parques de estacionamento da entidade 2 e 5 e também as entradas que tinham comportamentos fora do comum nos restantes parques, ficando, desta forma, a ser consideradas falhas nos dados. Para além disso, apagou-se ainda três parques que não apresentavam dados na altura do COVID-19, pois estes não seriam de interesse para as

análises desenvolvidas. No final deste processamento ficaram 28 parques de estacionamento neste conjunto de dados.

Os problemas encontrados no atributo temporal são idênticos aos encontrados para os dados das estações Gira, ou seja, a distância temporal entre registos consecutivos não é sempre a mesma e existem medições à mesma hora com valores de ocupação diferentes. Assim sendo, o processamento efetuado foi idêntico ao descrito na subsecção 4.3.1, começando pela ordenação dos dados no *software MySQL Workbench*, seguido pela limpeza de registos efetuados ao mesmo tempo mas que têm valores de ocupação distintos, pela criação de intervalos e agrupamento dos registos dentro de cada intervalo e, por fim, pela adição de linhas para preencher faltas de dados inferiores a um dia, interpolando os registos para obter a respetiva ocupação. Os intervalos utilizados neste conjunto de dados foram de 5 minutos, pois 43% dos registos distavam 5 minutos do registo seguinte.

O atributo espacial era representado por uma *string* de uma estrutura GeoJSON com a localização do parque de estacionamento, então foi decidido criar duas colunas com as coordenadas geográficas do parque, uma dessas colunas contém a latitude e a outra a longitude. Criou-se ainda um atributo para a taxa de ocupação do parque de estacionamento através da ocupação e da capacidade máxima do parque.

À semelhança do que foi feito nos dados das estações Gira, também nos parques de estacionamento se sentiu a necessidade da criação de um novo campo que mostrasse a diferença entre a taxa de ocupação do próprio registo com a taxa de ocupação de um registo 14 dias (2 semanas) anterior a esse. Para estes dados foi utilizado um valor de 14 dias atrás devido ao facto de existirem muitas falhas nos dados e dos dados abrangerem um período temporal mais reduzido.

Tabela 4.4: Principais colunas dos dados dos Parques de Estacionamento após processamento.

Atributo	Descrição
ID Parque	Identificador único do parque de estacionamento
ID Entidade	Identificador único da entidade do parque de estacionamento
Nome	Nome do parque de estacionamento
Tipo	Tipo do parque de estacionamento
Ocupação	Número de carros presentes no parque de estacionamento
Capacidade Máxima	Capacidade máxima do parque de estacionamento
Data Hora	Data e hora (5 em 5 minutos) da medição da ocupação do parque de estacionamento
Latitude	Latitude do parque de estacionamento
Longitude	Longitude do parque de estacionamento
Taxa de Ocupação	Taxa de ocupação do parque de estacionamento
Diferença na taxa	Diferença com a taxa de ocupação de 14 dias atrás

\* Novos atributos coloridos a cinzento.

Após terem sido resolvidos todos os problemas encontrados durante a fase de entendimento dos dados, relatados no relatório de qualidade dos dados, e criadas novas colunas, ficou-se com um total de 1 694 244 linhas e com os atributos principais descritos na tabela 4.4. Sendo que nesta tabela as linhas coloridas a cinzento representam os novos atributos criados durante todo o processamento.

### 4.3.3 Dados dos Congestionamentos Waze

O conjunto de dados dos congestionamentos segundo o Waze é o maior *dataset* fornecido pela Câmara Municipal de Lisboa, contando com mais de 12 milhões de linhas, o que equivale a mais de 5GB no tamanho do ficheiro. Este conjunto de dados fornece informação relativa ao estado do tráfego da cidade de Lisboa, apresentando dados para 2019 e 2020, sendo o seu último registo no dia 1 de julho de 2020.

Tabela 4.5: Colunas iniciais dos dados dos congestionamentos Waze.

Atributo	Descrição
Entity ID	Identificação única do congestionamento
Cidade	Cidade, freguesia ou zona
País	Código do país
Atraso	Atraso do congestionamento em comparação com a velocidade de fluxo livre (em segundos)
Nó Final	Cruzamento/Rua/Cidade mais próxima do final do congestionamento
Comprimento	Comprimento do congestionamento (em metros)
Nível	Nível do congestionamento
Tipo de estrada	Tipo de estrada onde ocorreu o congestionamento
Rua	Nome da rua
Tipo de curva	Tipo de curva onde ocorreu o congestionamento
Irregularidade	Tipo de irregularidade
Posição	Georreferenciação
<i>Timestamp</i> da entidade	Data de ocorrência (em <i>timestamp</i> )
<i>Timestamp</i> de validade	Data de arquivo na <a href="#">PGIL</a>

Cada entrada deste conjunto de dados representa um registo de congestionamento, sendo que um congestionamento único pode conter vários registos sobre o mesmo. Na tabela 4.5 encontram-se presentes os atributos iniciais deste conjunto de dados, antes de lhe ter sido aplicado qualquer tipo de processamento. Sendo possível constatar que existem atributos com informação espacial, temporal, informação do estado do tráfego e das características do congestionamento.

Após a exploração feita durante a etapa de entendimento dos dados surgiram alguns problemas na qualidade dos dados que careciam de processamento. Observou-se também, durante o decorrer do trabalho, a necessidade da criação de novos atributos que foram essenciais no desenvolvimento dos *dashboards*.



Outro processamento feito foi relativo ao atributo temporal que apresentava uma granularidade muito fina, isto é, era detalhado até aos segundos. E, visto que o objetivo desta dissertação é determinar qual o impacto que a COVID-19 teve na mobilidade, neste caso nos congestionamentos, não é necessário ter os dados com este nível de granularidade. Deste modo, e de forma semelhante ao que foi feito nos conjuntos de dados das estações Gira e parques de estacionamento, foram criados intervalos temporais e agrupados os congestionamentos únicos dentro desses intervalos. Analisou-se que, por norma, um registo de congestionamento era feito de 5 em 5 minutos, então a criação dos intervalos teve por base este valor. No final deste processamento cada congestionamento único apresenta no máximo um registo num intervalo de 5 minutos.

No decorrer do processamento dos dados foram ainda removidas algumas colunas que não mostravam ser interessantes para as análises futuras, nomeadamente as colunas “País”, “Tipo de curva”, “Cidade”, “Posição”, “*Timestamp* da entidade”, “Nó Final” e “*Timestamp* de validade”. Os dois primeiros atributos, “País” e “Tipo de curva”, apresentavam apenas um único valor. Já relativamente ao atributo “Cidade”, visto que o mesmo era representativo de vários valores, isto é, poderia conter o nome da cidade, da freguesia, ou até mesmo da zona do congestionamento, decidiu-se removê-lo e criar um novo atributo que contivesse apenas a freguesia do congestionamento. Sendo que a freguesia onde ocorreu o congestionamento foi descoberta através da interseção, no QGIS, de um polígono com todas as freguesias de Lisboa e as geometrias dos congestionamentos. Os atributos “Posição” e “*Timestamp* da entidade” deixaram de ser necessários visto que foi criado um novo atributo com a data e horas do congestionamento e outro que contém as geometrias dos mesmos. Por último, os atributos “Nó Final” e “*Timestamp* de validade” não se mostravam relevantes para propósitos de análises futuras.

Neste conjunto de dados houve também a necessidade da criação de um novo atributo que identificasse qual a avenida/eixo ou entrada/saída de Lisboa onde ocorreu o congestionamento. E, de forma semelhante ao que foi feito no conjunto de dados das estações Gira, foi realizado a interseção entre os polígonos das zonas de Lisboa com as geometrias dos congestionamentos. De modo que, para ser etiquetado com o valor da zona, o congestionamento não necessitava ter ocorrido em toda a sua extensão dentro da zona, bastava, portanto, interseção-la. Esta interseção foi realizada através do *software* QGIS e a lista de zonas encontra-se na secção 4.1.

Para ser possível analisar o tráfego localmente e correlacionar com a poluição atmosférica, foi essencial a criação de um novo atributo que permitiu, posteriormente, a união deste conjunto de dados com o conjunto de dados da qualidade do ar. A criação deste novo atributo teve por base a existência de zonas de influência (*buffers*) de cada estação de qualidade do ar, explicadas na secção 4.1, sendo que os valores do mesmo são os nomes das estações. Através do QGIS foram criados os *buffers* de tamanho máximo de cada estação, ou seja, de 500 metros para as [estações de tráfego](#) (Estação de Santa Cruz de Benfica, Estação de Entrecampos e Estação da Avenida da Liberdade) e de 2 000 metros para as [estações de fundo](#) (Estação do Restelo, Estação do Beato e Estação de Olivais). De seguida,

fez-se a interseção entre estes *buffers* e as geometrias dos congestionamentos, atribuindo-se uma etiqueta com o nome da estação correspondente a cada congestionamento que intersete a zona de influência.

Por último, foi criado um atributo que indica se o congestionamento intersesta o *buffer*, tendo em conta o raio do mesmo. Isto foi necessário para que nas ferramentas desenvolvidas se pudesse escolher, de forma interativa, o tamanho dos *buffers* que se pretende para cada tipo de estação. Ou seja, nesta coluna um valor de “Olivais/Olivais//” significa que esse registo de congestionamento intersestou o *buffer* da estação de Olivais a uma distância de 2 000 metros e 1 500 metros, e que a uma distância de 1 000 metros e 500 metros não intersestou a zona de influência da estação. Este processamento foi feito, novamente, com o auxílio da ferramenta QGIS na interseção das diferentes camadas espaciais.

No final do processamento do conjunto de dados dos congestionamentos *Waze* ficou-se com um total de 11 725 612 linhas, sendo que, no *data source* do *Tableau Desktop*, ficou-se com os principais atributos descritos na tabela 4.6. Nesta tabela as linhas coloridas a cinzento representam os novos atributos criados durante o processamento.

Tabela 4.6: Principais colunas dos dados dos congestionamentos *Waze* após processamento.

Atributo	Descrição
Entity ID	Identificação única do congestionamento
Atraso	Atraso do congestionamento em comparação com a velocidade de fluxo livre (em segundos)
Comprimento	Comprimento do congestionamento (em metros)
Nível	Nível do congestionamento
Tipo de estrada	Tipo de estrada onde ocorreu o congestionamento
Irregularidade	Tipo de irregularidade
Rua	Nome da rua onde ocorreu o congestionamento
Latitude Inicial	Latitude inicial do congestionamento
Longitude Inicial	Longitude inicial do congestionamento
Latitude Final	Latitude final do congestionamento
Longitude Final	Longitude final do congestionamento
Freguesia	Freguesia do congestionamento
Zona	Nome da zona onde ocorreu o congestionamento
Geometria Zona	Geometria da zona (em formato GeoJSON)
ID GeoJSON	Identificador único da geometria do congestionamento
Geometria Jam	Geometria do congestionamento (em formato GeoJSON)
Estação ( <i>buffer</i> )	Estação de qualidade do ar a que pertence o congestionamento
Estação Diferentes <i>buffers</i>	Atributo que identifica em que metros o congestionamento intersesta o <i>buffer</i>
Data Hora	Data e hora (5 em 5 minutos) em que ocorreu o registo de congestionamento

\* Novos atributos coloridos a cinzento.

#### 4.3.4 Dados dos Poluentes Atmosféricos

Para a realização desta dissertação a Câmara Municipal de Lisboa forneceu um conjunto de dados relativo ao poluente  $\text{NO}_2$ , para os anos 2019 e 2020. Posteriormente, foram adquiridos, através do *website* da European Environment Agency, novos conjuntos de dados referentes aos poluentes atmosféricos  $\text{O}_3$ ,  $\text{PM}_{2.5}$ ,  $\text{PM}_{10}$  e  $\text{SO}_2$ , para os anos 2017, 2018, 2019 e 2020. Sendo que, foram também adquiridos mais dados, para os anos 2017, 2018 e final de 2020, do poluente  $\text{NO}_2$ . Após o levantamento destes conjuntos de dados, foi feita a união de todos eles para um só ficheiro através do *software Tableau Prep Builder*.

As medições feitas aos parâmetros de qualidade do ar ocorrem, de hora a hora, em seis estações de qualidade do ar da cidade de Lisboa. E, é ainda importante salientar que nem todas as estações medem os mesmos poluentes.

Um problema que se detetou neste conjunto de dados foi a inexistência de um atributo espacial que permitisse mapear a estação no mapa. Deste modo, foi feito o levantamento das coordenadas geográficas de cada estação e foram criadas duas novas colunas: latitude e longitude. Criou-se também um novo atributo que indicava qual o tipo de estação, ou seja, se era estação de tráfego ou estação de fundo.

Tabela 4.7: Principais colunas dos dados dos poluentes atmosféricos após processamento.

Atributo	Descrição
Poluente	Nome do poluente atmosférico
Concentração	Concentração medida do poluente atmosférico (em $\mu\text{g}/\text{m}^3$ )
Estação	Nome da estação que fez a medição
Data Hora	Data e hora da medição do poluente
Tipo de Estação	Tipo da estação que fez a medição
Latitude	Latitude da estação
Longitude	Longitude da estação

\* Novos atributos coloridos a cinzento.

Foi ainda necessário ter dois conjuntos de dados com diferentes formatos, onde um deles facilitava a criação de *dashboards* exclusivamente sobre a qualidade do ar da cidade de Lisboa e o outro favorecia a junção do mesmo com o conjunto de dados do *Waze*. Deste modo, na tabela 4.7 encontram-se descritas as colunas do conjunto de dados para as ferramentas dos poluentes atmosféricos, onde cada entrada representa uma medição de um poluente atmosférico, a uma determinada hora, numa estação diferente. Sendo que, este conjunto de dados apresenta, no total, 1 051 090 linhas. Na tabela 4.8 estão detalhados os atributos do conjunto de dados que foi, posteriormente, unido no *Tableau Desktop* com os dados do *Waze*, e que contém, no total, 210 230 linhas. Neste conjunto de dados cada linha apresenta as medições feitas numa estação, a uma determinada hora, sendo que os valores de cada poluente se encontram na coluna respetiva.

Tabela 4.8: Principais colunas dos dados dos poluentes atmosféricos, após processamento. Conjunto de dados utilizado para junção com os dados do *Waze*.

Atributo	Descrição
NO2	Concentração medida do poluente NO2 (em $\mu\text{g}/\text{m}^3$ )
O3	Concentração medida do poluente O3 (em $\mu\text{g}/\text{m}^3$ )
PM10	Concentração medida do poluente PM10 (em $\mu\text{g}/\text{m}^3$ )
PM2.5	Concentração medida do poluente PM2.5 (em $\mu\text{g}/\text{m}^3$ )
SO2	Concentração medida do poluente SO2 (em $\mu\text{g}/\text{m}^3$ )
Estação	Nome da estação que fez a medição
Data Hora	Data e hora da medição
Tipo de Estação	Tipo da estação que fez a medição
Latitude	Latitude da estação
Longitude	Longitude da estação

\* Novos atributos coloridos a cinzento.

#### 4.3.5 Dados da Meteorologia

Quando se está a correlacionar a poluição atmosférica com os parâmetros observados no tráfego é fundamental ter também em consideração as alterações climáticas. Isto, pois existem poluentes em que a sua mudança de comportamento pode não estar necessariamente relacionada com o tráfego, mas sim com fatores meteorológicos. Posto isto, a Câmara Municipal de Lisboa disponibilizou dados meteorológicos do *IPMA*, para os anos 2019 e 2020.

Cada entrada neste conjunto de dados apresenta as medições feitas, a uma determinada hora, numa estação meteorológica. Na tabela 4.9 encontram-se descritos os atributos deste *dataset* antes de lhes ter sido aplicado qualquer processamento. Sendo possível constatar que existe informação temporal, informação espacial e informação do estado dos parâmetros climáticos.

Tabela 4.9: Colunas iniciais dos dados da Meteorologia.

Atributo	Descrição
Data	Data e Hora a que foi feita a medição
Estação	Identificador único da estação meteorológica
Posição	GeoJSON com a localização da estação meteorológica
Humidade	Humidade relativa média (em %)
Intensidade do Vento	Intensidade média do vento (em m/s)
Intensidade do Vento KM	Intensidade média do vento (em km/h)
Precipitação	Quantidade total de precipitação (em mm)
Pressão	Pressão média ao nível médio do mar (em hPa)
Radiação	Radiação Solar (em $\text{W}/\text{m}^2$ )
Temperatura	Temperatura média (em $^{\circ}\text{C}$ )

Após ter sido concluída a fase de entendimento dos dados, surgiram alguns problemas que necessitavam ser processados antes de serem realizadas as seguintes análises. Através da ferramenta *Tableau Prep Builder*, foram removidas entradas que não apresentavam quaisquer valores associados e, deste modo, alguns problemas de completude foram resolvidos. Foram ainda removidas entradas que pertenciam a uma estação fora da cidade de Lisboa, não se enquadrando, deste modo, dentro do contexto desta dissertação.

Foram encontrados também alguns problemas relacionados com o atributo temporal que precisaram de um processamento mais complexo, feito através de código *python*. O primeiro problema, e o mais fácil a ser resolvido, foi o facto de existirem várias linhas repetidas que necessitaram ser removidas. Seguidamente, constatou-se que a diferença temporal entre medições consecutivas tinha falhas, não apresentando sempre medições de hora a hora. Então foi necessário resolver este problema através da adição de linhas quando não existia medição na hora seguinte, sendo que os valores dos parâmetros meteorológicos foram calculados através de interpolação.

Contudo, este conjunto de dados apresentava ainda bastantes problemas de falta de dados e muitos valores -99. Sendo que este valor aparece quando uma estação não consegue recolher informação acerca dum parâmetro meteorológico, então é apresentado o valor -99 como valor *default* de que algo não correu bem. Visto existirem apenas três estações meteorológicas na área de Lisboa, e que os valores dos parâmetros climáticos medidos em cada uma delas, numa hora específica, não diferem muito entre si, decidiu-se criar um novo conjunto de dados que apresenta uma visão global da meteorologia na cidade de Lisboa. Os valores dos atributos deste novo conjunto de dados foram calculados através da média dos valores existentes nas três estações meteorológicas dos dados iniciais. Sendo que, no final, cada entrada contém um valor representativo do estado dos parâmetros meteorológicos, a uma hora específica, na cidade de Lisboa. Após este processamento conseguiu-se reduzir o número de valores *default* e melhorar o problema da falta de dados de algumas estações, ficando-se apenas com falta de dados entre 30 de outubro de 2019 e 15 de janeiro de 2020. Na tabela 4.10 encontram-se detalhados os atributos deste novo conjunto de dados, sendo que este conjunto de dados contém 17 562 linhas.

Tabela 4.10: Colunas dos dados da Meteorologia Global após processamento.

Atributo	Descrição
Data Hora	Data e Hora a que foi feita a medição
Humidade	Humidade relativa média (em %)
Intensidade do Vento	Intensidade média do vento (em km/h)
Precipitação	Quantidade total de precipitação (em mm)
Pressão	Pressão média ao nível médio do mar (em hPa)
Radiação	Radiação Solar (em W/m <sup>2</sup> )
Temperatura	Temperatura média (em °C)

### 4.3.6 Dados das Medidas Governamentais

De forma a melhorar as ferramentas desenvolvidas e, visto que o objetivo final desta dissertação é compreender o impacto que a COVID-19, e consequentes restrições, tiveram na mobilidade e na qualidade do ar de Lisboa, foi essencial ter um conjunto de dados que indicasse as fases da pandemia em Portugal.

O conjunto de dados encontrado era bastante limitado em termos temporais, então houve a necessidade de fazer o levantamento de mais medidas governamentais que fossem ao encontro do período temporal dos restantes conjuntos de dados, ou seja, terminando no dia 31 de dezembro de 2020. Deste modo, cada entrada do conjunto de dados é relativa a um dia durante a pandemia, apresentando, no total, 305 linhas.

Decidiu-se ainda separar o atributo referente à medida governamental em duas colunas, sendo uma referente ao estado da pandemia e outra relativa a medidas mais específicas implementadas nesse dia. Na tabela 4.11 encontram-se descritos os atributos que fazem parte deste conjunto de dados.

Tabela 4.11: Colunas dos dados das Medidas Governamentais após processamento.

Atributo	Descrição
Data	Data
Medida Governamental	Descrição da medida governamental implementada nesse dia
Estado	Estado da pandemia (Alerta, Emergência, Calamidade e Contingência)

\* Novos atributos coloridos a cinzento.

### 4.3.7 Dados dos Pontos de Interesse

À semelhança do que foi feito nos trabalhos [11, 24, 45], foram também utilizados neste trabalho pontos de interesse existentes na cidade de Lisboa. A utilização destes foi útil para distinguir os padrões dos comportamentos relativos à mobilidade nas redondezas dos mesmos.

Foi utilizado um conjunto de dados, criado por colegas de mestrado, que contém informação espacial de alguns pontos de interesse da cidade de Lisboa. Este conjunto de dados foi atualizado para o contexto desta dissertação, sendo que no final se ficou com 627 pontos de interesse divididos por categorias e subcategorias, sendo as mesmas:

- **Transporte**
  - Estações Ferroviárias
  - Estações Fluviais
  - Metro Lisboa

- **Ensino**
  - Secundário
  - Superior
- **Saúde**
  - Hospitais
  - Farmácias
- **Lazer**
  - Centros Comerciais
  - Marinas e Docas
  - Parques e Jardins
  - Terminal de Cruzeiros

A tabela 4.12 apresenta uma descrição das colunas presentes neste *dataset*.

Tabela 4.12: Colunas dos dados dos Pontos de Interesse.

Atributo	Descrição
Categoria	Categoria do ponto de interesse
Subcategoria	Subcategoria do ponto de interesse
Nome	Nome do ponto de interesse
Latitude	Latitude do ponto de interesse
Longitude	Longitude do ponto de interesse

#### 4.3.8 Criação de Novos Conjuntos de Dados

Para o desenvolvimento de alguns *dashboards* houve a necessidade da criação de um novo conjunto de dados. Este contém informação, de hora a hora, relativa ao estado do tráfego ao redor de 627 pontos de interesse da cidade de Lisboa.

A criação deste novo conjunto de dados foi realizada em código *python* e envolveu os dados do *Waze* e dos Pontos de Interesse da cidade de Lisboa. Cada registo deste novo conjunto de dados contém informação, numa hora específica, do estado do tráfego nas redondezas do [Ponto de Interesse \(POI\)](#).

Para a criação deste conjunto de dados foi necessário, primeiramente, definir um *buffer* relativo a cada subcategoria de ponto de interesse. Sendo que foram considerados três valores possíveis: 100 metros, 250 metros e 400 metros. Estes valores têm em consideração as dimensões do local, isto é, as farmácias como são edifícios, por norma, de pequenas dimensões têm associado um *buffer* de 100 metros. Por outro lado, os hospitais têm grandes dimensões, então têm associado um *buffer* de 400 metros. Em todos os restantes pontos de interesse foi atribuído um raio de 250 metros. Os atributos criados relativos ao estado do tráfego encontram-se representados na tabela 4.13.

Tabela 4.13: Colunas dos dados do Sinal do Tráfego ao Redor dos POI.

Atributo	Descrição
Nome	Nome do ponto de interesse
Latitude	Latitude do ponto de interesse
Longitude	Longitude do ponto de interesse
Data Hora	Data e Hora do registo do sinal do tráfego
Jams Diferentes	Número total de congestionamentos únicos naquela hora
Jams Diferentes (Nível 1)	Número total de congestionamentos únicos do nível 1 naquela hora
Jams Diferentes (Nível 2)	Número total de congestionamentos únicos do nível 2 naquela hora
Jams Diferentes (Nível 3)	Número total de congestionamentos únicos do nível 3 naquela hora
Jams Diferentes (Nível 4)	Número total de congestionamentos únicos do nível 4 naquela hora
Jams Diferentes (Nível 5)	Número total de congestionamentos únicos do nível 5 naquela hora

Contudo, foi ainda fundamental criar um novo conjunto de dados que apresentasse registos diários do estado do tráfego, isto é, o número total de congestionamentos diferentes num dia. Sendo que este foi, posteriormente, unido com o *dataset* que apresenta a informação horária do estado dos congestionamentos.

#### 4.4 Dashboards Desenvolvidos

Nesta secção serão apresentadas e detalhadas as *dashboards* desenvolvidas que permitem responder às questões gerais levantadas no início desta dissertação. Em cada subsecção será apresentado um *dashboard*, ou um conjunto de *dashboards*, mostrando quais as visualizações, *tooltips* e interações que o compõem. A exibição de cada *dashboard* será ainda acompanhada por uma listagem e breve explicação de questões analíticas mais específicas que podem ser respondidas através do mesmo.

Na tabela 4.14 é possível identificar quais os *dashboards* que serão apresentados nas seguintes subsecções, sendo que estes se encontram agrupados na tabela segundo a sua temática principal. Na primeira subsecção são apresentados dois *dashboards* referentes às bicicletas Gira e na segunda subsecção é apresentado um *dashboard* que tem como finalidade analisar os parques de estacionamento. Nas subsecções seguintes, ou seja desde a subsecção 4.4.3 até à subsecção 4.4.5 inclusive, são apresentados *dashboards* que têm como principal temática os congestionamentos detetados pelo Waze na cidade de Lisboa. Seguidamente, desde a subsecção 4.4.6 até à subsecção 4.4.8 inclusive, são apresentados *dashboards* relacionados com a qualidade do ar. Por fim, na subsecção 4.4.9 é apresentado um *dashboard* que inclui a integração de duas temáticas principais: os congestionamentos

detetados pelo *Waze* e a qualidade do ar da cidade de Lisboa.

Tabela 4.14: *Dashboards* agrupados segundo a sua temática principal.

<b>Grupo</b>	<b><i>Dashboards</i></b>
Bicicletas Gira	Giras - Estados da pandemia (secção 4.4.1)
	Giras - Datas específicas (secção 4.4.1)
Parques de Estacionamento	Parques de estacionamento (secção 4.4.2)
Congestionamentos Waze	Congestionamentos - Geral (secção 4.4.3)
	Congestionamentos - Todas as zonas (secção 4.4.4)
	Congestionamentos - Zona específica (secção 4.4.4)
	Congestionamentos - POI (secção 4.4.5)
Qualidade do Ar	Qualidade do ar (secção 4.4.6)
	Qualidade do ar - Por estações (secção 4.4.7)
	Comparação entre poluentes (secção 4.4.8)
Integração	Congestionamentos vs qualidade do ar (secção 4.4.9)

Sendo o principal objetivo desta dissertação o desenvolvimento de ferramentas que permitam analisar o impacto que a COVID-19 teve na mobilidade e na qualidade do ar de Lisboa, procurou-se sempre oferecer uma forte componente interativa nos *dashboards*. Tentando sempre que os mesmos fossem intuitivos e de uso fácil. Deste modo, pretendeu-se que as visualizações utilizadas nos *dashboards* não fossem muito complexas, optando, na maioria das vezes, por visualizações que fossem utilizadas no dia a dia.

O desenvolvimento de cada *dashboard* baseou-se num processo iterativo, tendo, portanto, várias iterações até atingirem a sua versão final. As ferramentas eram mostradas aos orientadores, colegas de mestrado e, em alguns momentos, aos *stakeholders* da Câmara Municipal de Lisboa, de forma a receber *feedback*. Deste modo, houve a necessidade de colocar o tamanho de cada *dashboard* como automático, para que estes se enquadrassem em qualquer ecrã onde fossem exibidos.

É de realçar ainda que, apesar das ferramentas terem sido desenvolvidas no contexto da determinação do impacto da COVID-19 na mobilidade e qualidade do ar de Lisboa, estas não estão limitadas a ver informação somente referentes a períodos temporais durante a pandemia. Sendo que, a grande maioria das ferramentas proporciona a possibilidade de responder a um vasto conjunto de questões analíticas que não impliquem a COVID-19.

No apêndice A estão presentes imagens de todos os *dashboards* desenvolvidos, num formato horizontal e de página inteira.

Em suma, na tabela 4.15 é possível observar quais os conjuntos de dados que foram utilizados em cada um dos *dashboards* apresentados nas subsecções seguintes. As denominações dadas aos conjuntos de dados presentes nas colunas da tabela são as seguintes:

- **CD1:** Estações Gira
- **CD2:** Número de viagens totais das Giras
- **CD3:** *Eco-Counter*
- **CD4:** Parques de Estacionamento
- **CD5:** Congestionamentos *Waze*
- **CD6:** Poluentes Atmosféricos
- **CD7:** Meteorologia
- **CD8:** Ciclovias
- **CD9:** Medidas Governamentais
- **CD10:** Pontos de Interesse
- **CD11:** Sinal de tráfego ao redor dos POIs

Tabela 4.15: Conjuntos de dados utilizados em cada *dashboard*.

<i>Dashboards</i>	Conjuntos de Dados										
	CD1	CD2	CD3	CD4	CD5	CD6	CD7	CD8	CD9	CD10	CD11
Giras - Estados da pandemia (4.4.1)	✓	✓	✓					✓	✓		
Giras - Datas específicas (4.4.1)	✓	✓	✓					✓			
Parques de estacionamento (4.4.2)				✓					✓		
Congestionamentos - Geral (4.4.3)					✓				✓		
Congestionamentos - Todas as zonas (4.4.4)					✓				✓		
Congestionamentos - Zona específica (4.4.4)					✓				✓		
Congestionamentos - POI (4.4.5)									✓	✓	✓
Qualidade do ar (4.4.6)						✓			✓		
Qualidade do ar - Por estações (4.4.7)						✓					
Comparação entre poluentes (4.4.8)						✓					
Congestionamentos vs qualidade do ar (4.4.9)					✓	✓	✓		✓		

#### 4.4.1 Análise às bicicletas Gira da cidade de Lisboa

Para responder à questão QG01, apresentada na secção 4.1, de “Houve alterações no padrão de uso das bicicletas Gira durante a pandemia COVID-19?” foi desenvolvida a ferramenta visual presente na figura 4.8. Sendo que, de forma a facilitar a descrição deste *dashboard*, serão utilizadas as denominações aos seus elementos presentes nessa figura. Para além disso, através deste vídeo é possível ver uma explicação das funcionalidades deste *dashboard*.

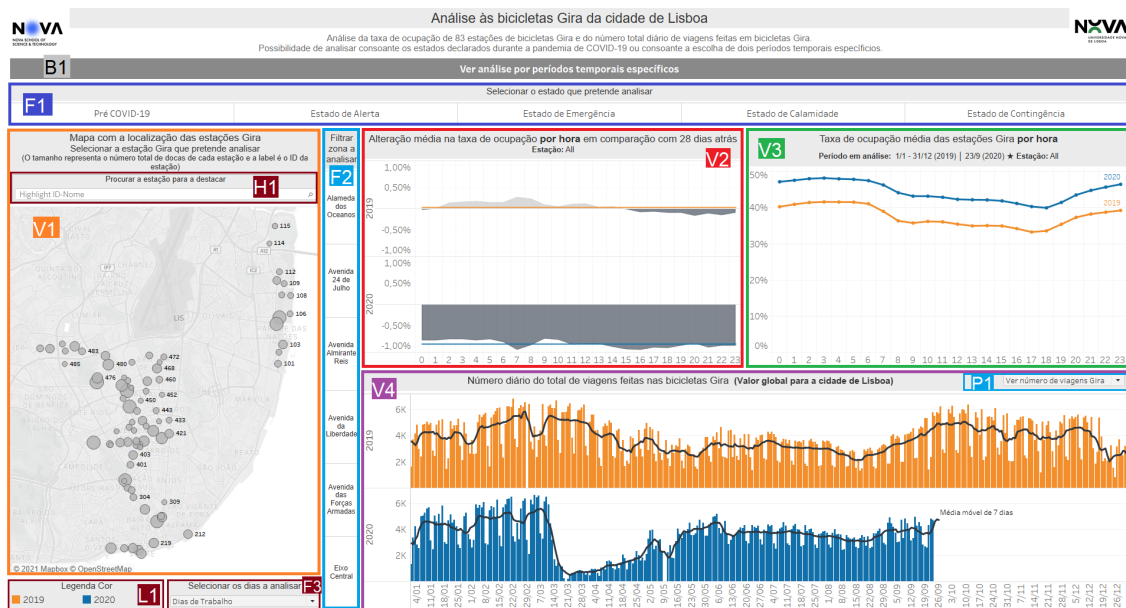


Figura 4.8: *Dashboard* para a análise às bicicletas Gira da cidade de Lisboa.

Com este *dashboard* é possível ter uma visão de como se comportou cada estação Gira durante os estados da pandemia. A mesma disponibiliza métodos de comparação do comportamento de cada estação relativamente ao mesmo período temporal do ano anterior. É ainda possível analisar o número total diário de viagens feitas na rede Gira da cidade de Lisboa e de entradas de bicicletas na ciclovía da Avenida Duque de Ávila. Para além disso, sentiu-se a necessidade de poder analisar segundo períodos temporais específicos, escolhendo um período para analisar e outro período como comparação. Esta funcionalidade é acedida através dum botão que redireciona o utilizador para outro *dashboard* idêntico ao da figura 4.8.

Através da figura 4.8 é possível observar que este *dashboard* exibe quatro visualizações, uma legenda interativa, três filtros, um parâmetro, um *highlighter* e um botão.

A visualização V1 é um mapa de símbolos proporcionais relativo à localização de cada estação Gira. Neste, o tamanho dos círculos representa o tamanho de cada estação, isto é, o número total de docs existentes. A visualização V2 é composta por um gráfico de área que apresenta, para o ano 2019 e 2020, a alteração média na taxa de ocupação por hora em comparação com 28 dias atrás. Isto significa que esta visualização permite avaliar

as mudanças de comportamento das estações, verificando se existiu uma tendência de subida e/ou descida da taxa de ocupação ao longo das horas. Para além disso, é ainda representado, para cada ano, uma linha com a média dos valores obtidos. Neste gráfico uma área colorida a cinzento escuro representa uma descida na taxa de ocupação e uma área colorida a cinzento claro representa uma subida. A visualização V3 é um gráfico de linhas que representa a evolução da taxa de ocupação, ao longo das horas, nos anos 2019 e 2020. Visto que este *dashboard* faz a comparação entre dois anos diferentes, houve a necessidade de associar cores distintas a cada um deles, estando a legenda destas cores presente em L1. Por fim, a visualização V4 é uma visualização de eixo Y duplo. Nas barras é possível observar o número total diário de viagens feitas na rede de bicicletas Gira e a linha representa a média móvel de 7 dias desse mesmo atributo.

Os filtros F1 e F2 têm como propósito filtrar as restantes visualizações, sendo que o filtro F1 apresenta, sob a forma textual, os estados da pandemia ocorridos durante o período temporal do conjunto de dados e o filtro F2 apresenta as zonas da cidade de Lisboa que foram consideradas relevantes para analisar. Neste *dashboard* é ainda possível filtrar os dias em análise, ou seja, ver apenas dias de trabalho, fins de semana ou ambos, sendo que isto é feito através do filtro F3. O parâmetro P1 permite escolher o que visualizar em V4 e V1, sendo que as escolhas possíveis são: “Ver número de viagens Gira” e “Ver sensor Eco-Counter”. A figura 4.10 mostra a visualização V4 quando se escolhe “Ver sensor Eco-Counter”.

Este *dashboard* apresenta ainda algumas *tooltips*, sendo que todas elas são exibidas ao fazer *hover* sobre um elemento da visualização e todas se encontram na forma textual. No filtro F1 é apresentado em *tooltip* a data de início e final do estado, bem como o número de dias entre essas datas. Na visualização V1 cada estação apresenta em *tooltip* informações sobre a mesma, ou seja, o seu identificador, o seu nome, o número de docas que apresenta e a zona a que pertence. No filtro F2 é mostrado, para cada zona, o número de estações que se encontram na mesma. Nas *tooltips* das visualizações V2, V3 e V4 são apresentados valores que já são mostrados nas visualizações, nomeadamente, a alteração média da taxa de ocupação, a média da taxa de ocupação, o número total de viagens nas bicicletas Gira, o número total de entrada de bicicletas contadas pelo *Eco-Counter* e ainda informação temporal para cada um deles.

A grande maioria das ferramentas desenvolvidas no âmbito desta dissertação apresentam uma forte componente interativa. Neste *dashboard* uma das interações mais importantes é a seleção do estado da pandemia que se pretende analisar. Esta seleção é feita clicando num dos estados presentes no filtro F1, sendo que todas as visualizações serão afetadas. Isto fará ainda com que a visualização V1 passe a apresentar, através da intensidade da cor, a mudança de comportamento de cada estação em 2020 comparativamente a 2019 (figura 4.9). Quando isto acontece, surge ainda uma pequena visualização flutuante que tem como propósito mostrar, em *tooltip*, informação acerca das cores mostradas no mapa, tal como é possível observar na figura 4.9. Ao fazer *hover* com o rato sobre o estado é ainda feito um destaque, na visualização V4, de qual o período temporal desse estado.

Através do filtro F2 é possível filtrar a zona que se pretende analisar, sendo que este filtro é aplicado às visualizações V1, V2 e V3. Clicando numa estação, presente na visualização V1, é feita uma ação de filtragem na visualização V2 e V3 para os valores referentes a essa estação. Caso se pretenda procurar por uma estação, de modo a poder fazer um destaque da mesma no mapa, basta escrever o ID da mesma em H1. Existe ainda o filtro F3, onde é possível escolher se se pretende analisar dias de trabalho ou fins de semana. Este filtro afeta as visualizações V2 e V3, e afeta também a visualização V1 quando esta apresenta as estações coloridas consoante a sua mudança de comportamento. Neste *dashboard* está ainda presente uma interação de destaque que é acionada clicando num ano, presente na legenda de cores L1, afetando, deste modo, as visualizações V2, V3 e V4.

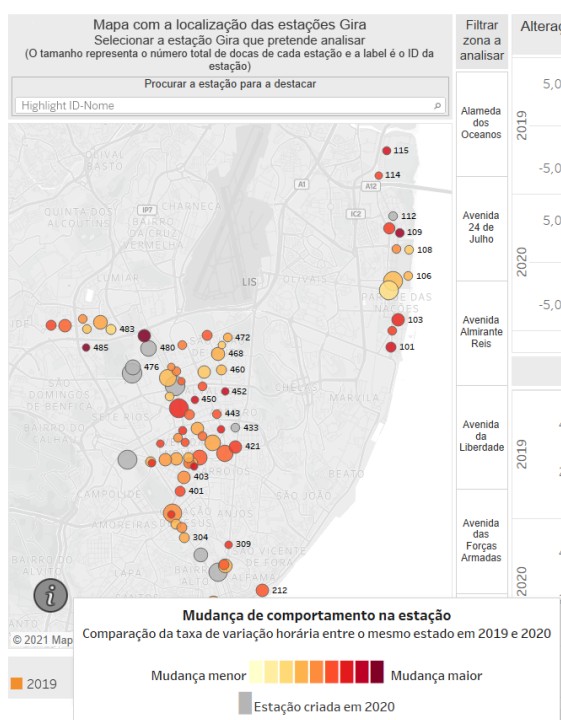


Figura 4.9: Alteração na visualização V1 após clicar no “Estado de Calamidade” em F1.

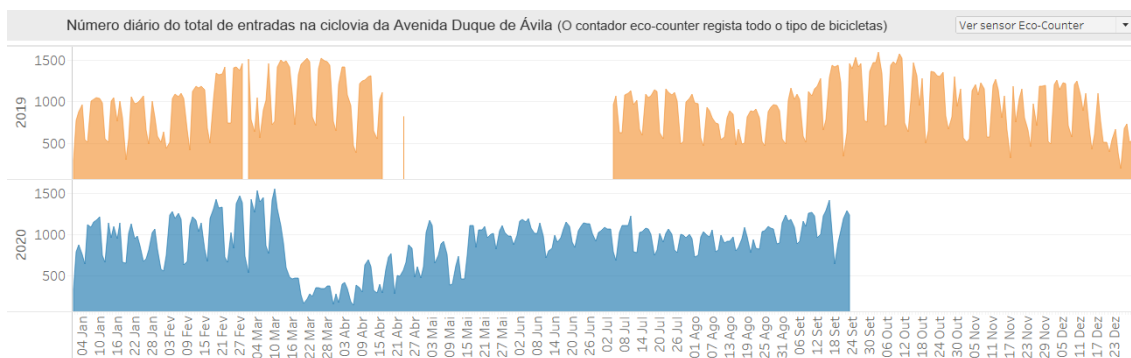


Figura 4.10: Alteração na visualização V4 após alterar o parâmetro P1 para os dados do *eco-counter*.

Este *dashboard* apresenta ainda a possibilidade de ver os dados referentes ao sensor *Eco-Counter*, instalado na ciclovia da Avenida Duque de Ávila. Para tal, é necessário selecionar, através do parâmetro P1, que se pretende ver estes dados. Esta seleção irá fazer com que as visualizações V1 e V4 sejam substituídas. A visualização V1 é substituída por um mapa que apresenta a localização da ciclovia onde está instalado o sensor e a visualização V4 é substituída por um gráfico de área que mostra o número total diário de entradas de bicicletas nessa ciclovia. Na figura 4.10 é possível observar a alteração ocorrida na visualização V4. Para além disso, todas as interações que afetam a visualização V4, afetam também esta nova visualização apresentada na mesma posição.

Existe ainda a possibilidade de dar uma maior liberdade de escolha ao utilizador dos períodos temporais que pretende analisar. Para isto, o mesmo pode clicar no botão B1 e será redirecionado para um *dashboard* semelhante a este. Neste *dashboard*, presente na figura 4.11, é possível escolher a data a partir da qual quer analisar e a data que quer que sirva de comparação, através dos parâmetros P1 e P2, respetivamente. Após a escolha das datas é possível escolher o intervalo de dias que será aplicado a ambas as datas escolhidas, sendo que isso é feito através do filtro F1. Estes dois períodos temporais escolhidos são representados pela cor vermelha e cinzenta em todo o *dashboard*, sendo possível ver uma legenda da cor em L1. Neste *dashboard* a visualização V4 difere da que foi apresentada anteriormente quando se analisa por estados declarados na pandemia. Aqui é exibido um gráfico de linhas em que o eixo X indica a ordem dos dias em análise, isto é, se é o primeiro dia, segundo dia, etc. O resto das visualizações, *tooltips* e interações presentes neste *dashboard* são idênticas às do *dashboard* anterior, com a exceção que neste não são apresentadas cores nas estações Gira que indiquem a mudança de comportamento das mesmas.

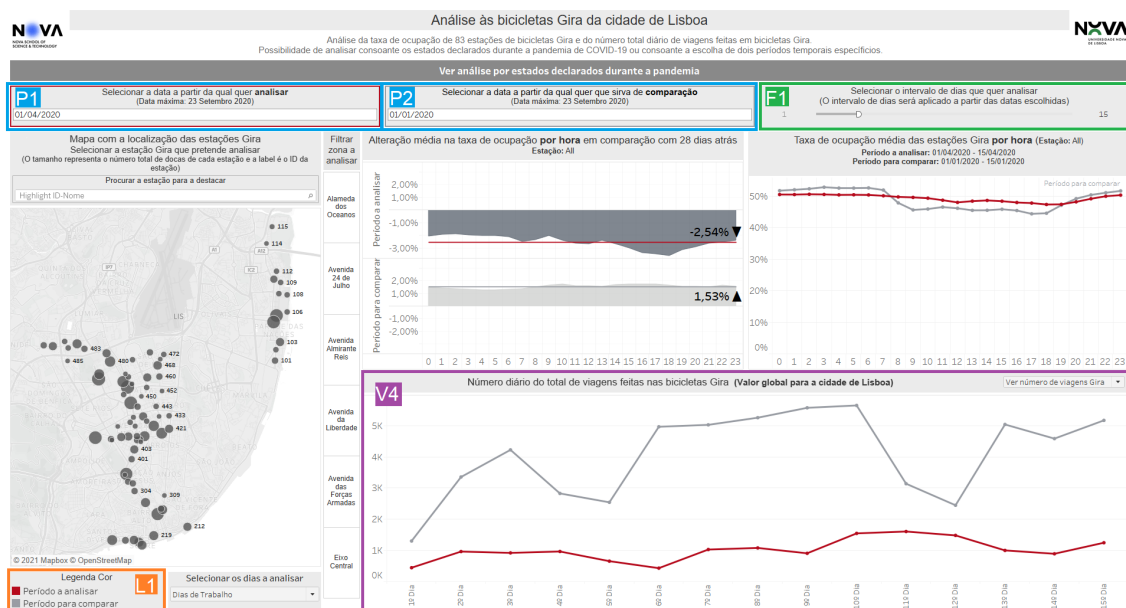


Figura 4.11: *Dashboard* para a análise às bicicletas Gira da cidade de Lisboa - Após clicar no botão para ver análise por períodos temporais específicos.

Concluindo, com estas ferramentas é possível responder a um vasto conjunto de questões analíticas mais específicas, como por exemplo as questões apresentadas na secção 4.1:

- **“A taxa de ocupação ao longo das horas, na estação 103, é semelhante em 2019 e 2020 durante a altura do estado de emergência?” (Q01.01).** É possível responder a esta questão através da seleção da estação Gira 103 no mapa e através da seleção do estado de emergência no filtro F1.
- **“Relativamente ao estado de contingência, considera-se que, durante os dias de trabalho, houve alteração da hora com maior taxa de ocupação nas estações pertencentes à zona do eixo central?” (Q01.02).** Para responder a esta questão é necessário selecionar o estado de contingência no filtro F1, alterar o filtro F3 de modo a apresentar os valores dos dias de trabalho e selecionar a zona do eixo central no filtro F2.
- **“Durante os estados da pandemia houve alguma alteração significativa no número diário de viagens feitas nas bicicletas da rede Gira?” (Q01.03).** Para obter a resposta a esta questão não é necessário realizar ações de filtragem, todavia ao fazer *hover* sobre os estados da pandemia no filtro F1 é possível ter-se uma maior perceção de quais foram os estados da pandemia onde ocorreu uma maior alteração.

#### 4.4.2 Análise aos parques de estacionamento da cidade de Lisboa

As mudanças na mobilidade ocorridas durante a pandemia COVID-19 afetaram também os parques de estacionamento da cidade de Lisboa. Deste modo, na secção 4.1, foi levantada a seguinte questão QG02: “Houve alterações no padrão de uso dos parques de estacionamento durante a pandemia COVID-19?”. Na figura 4.12 está uma imagem do *dashboard* que permite responder a esta questão, sendo que esta imagem se encontra acompanhada pelas denominações dadas a cada elemento do mesmo. Através deste [vídeo](#) é ainda possível ver uma explicação das funcionalidades do *dashboard*.

Através deste *dashboard* é possível observar o comportamento de cada um dos vinte e oito parques de estacionamento durante os estados da pandemia e, também, analisar o comportamento diário do top 10 parques de estacionamento com maior taxa de ocupação. Apesar deste *dashboard* apresentar algumas parecenças com a ferramenta desenvolvida para analisar as estações da rede Gira, foram tomadas diferentes decisões durante a sua implementação. Isto deveu-se ao facto do conjunto de dados dos parques de estacionamento abranger um período temporal reduzido e conter algumas alturas com falta de dados. Devido a este problema não foi possível conceber visualizações para comparar o comportamento entre a mesma altura do estado da pandemia no ano 2019 e 2020.

Este *dashboard* apresenta quatro visualizações, dois filtros, um *highlighter* e um parâmetro, tal como é possível observar na figura 4.12.

Na visualização V1 é possível observar a localização de cada parque de estacionamento através de um mapa de símbolos proporcionais, no qual o tamanho do símbolo representa

o tamanho do parque. A visualização V2 apresenta um gráfico de área que mostra a alteração média na taxa de ocupação horária em comparação com 14 dias atrás. Quando a variação é negativa a área do gráfico fica com uma coloração a cinzento escuro e quando é positiva fica a cinzento claro. Este gráfico apresenta ainda uma linha com o valor médio. A visualização V3 é constituída por um gráfico de barras, em que cada barra representa a taxa de ocupação média. Para esta visualização é possível mudar a granularidade temporal apresentada. Sendo que isto é feito através do parâmetro P1, e as opções disponíveis são: “Dias”, “Semanas” e “Meses”. Esta visualização faz ainda uso da cor para mapear os períodos temporais pré COVID-19 (a laranja) e durante a COVID-19 (a azul). Por último, na visualização V4 está um *heat table*, no qual é mostrado um top dos dez parques de estacionamento com maior taxa de ocupação por dia. Nesta, a intensidade da cor está mapeada para o valor da taxa de ocupação, de modo que os maiores valores apresentam uma coloração mais intensa. De notar que os valores a cinzento no *heat table* representam faltas de dados existentes no conjunto de dados dos parques de estacionamento.

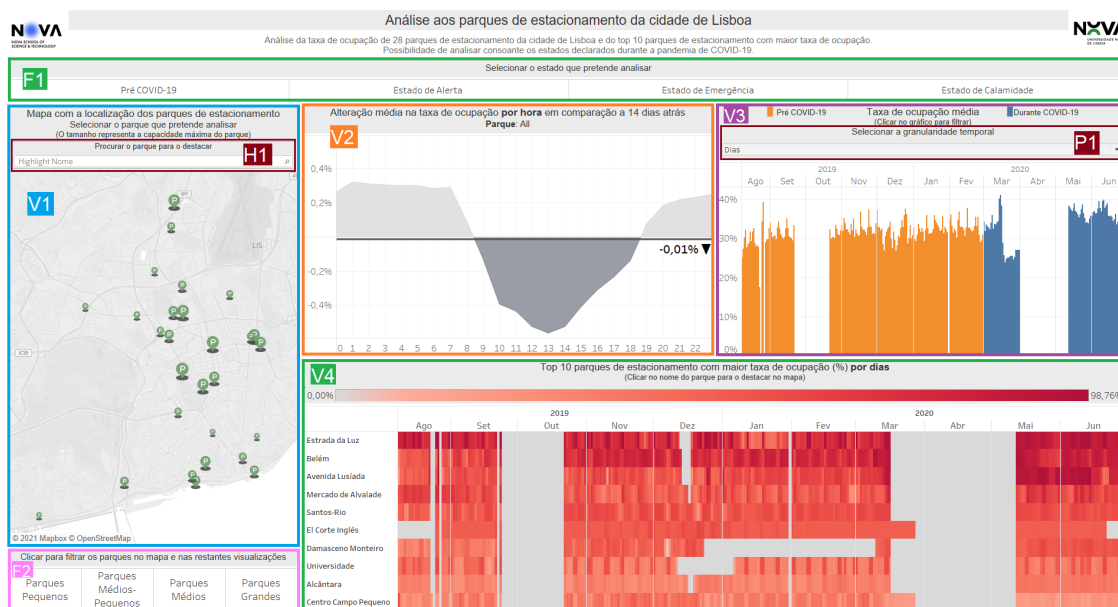


Figura 4.12: *Dashboard* para a análise aos parques de estacionamento da cidade de Lisboa.

Os filtro F1 e F2 têm como finalidade filtrar os dados das restantes visualizações. No filtro F1 é apresentado, sob a forma textual, os estados da pandemia que fazem parte deste conjunto de dados. No filtro F2 é apresentado, também sob a forma textual, quatro classes de tamanhos dos parques de estacionamento. Os limites de capacidade máxima das classes são as seguintes:

- **Parques Pequenos:** capacidade máxima  $\leq 100$  carros;
- **Parques Médios-Pequenos:**  $100 < \text{capacidade máxima} \leq 300$  carros;
- **Parques Médios:**  $300 < \text{capacidade máxima} \leq 1000$  carros;
- **Parques Grandes:** capacidade máxima  $> 1000$  carros;

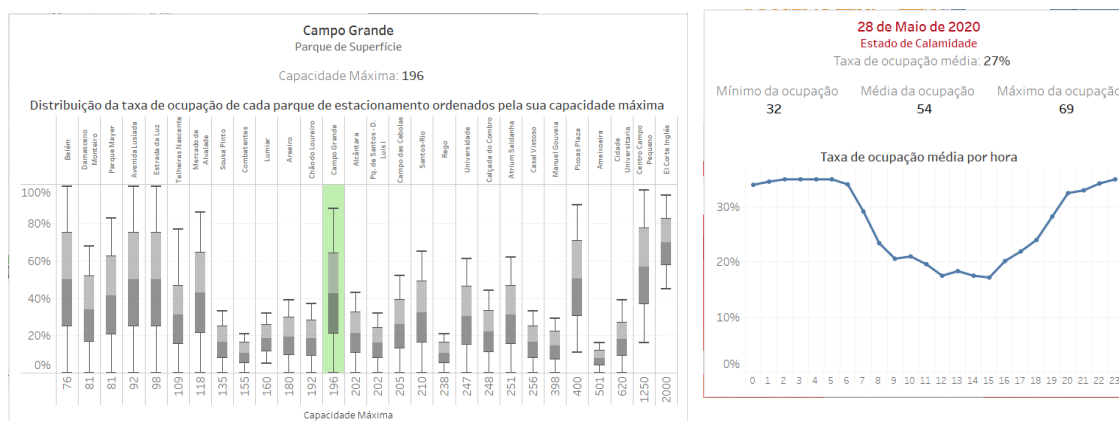


Figura 4.13: Exemplo de *tooltip* do *dashboard* dos parques de estacionamento. Referente à visualização V2, quando é feito *hover* no parque do Campo Grande.

Figura 4.14: Exemplo de *tooltip* do *dashboard* dos parques. Referente à visualização V4, quando é feito *hover* num dia.

Este *dashboard* apresenta dois tipos de *tooltips*: as que são apresentadas somente sob a forma textual e as que contém gráficos, sendo que todas elas são mostradas ao fazer *hover* num elemento das visualizações. Para a visualização V1 é mostrado, no formato textual, informação relativa ao parque, isto é, o seu nome, o seu tipo (superfície ou estrutura) e a sua capacidade máxima. Para além desta informação, é mostrado ainda um diagrama de caixa de bigodes no qual é possível observar a distribuição da taxa de ocupação de cada parque, sendo que estes se encontram ordenados no diagrama pela sua capacidade máxima. Deste modo, é possível compreender como se posiciona o parque que se pretende analisar relativamente aos restantes parques, pois o mesmo encontra-se destacado através da cor verde, tal como é possível ver na figura 4.13. Na visualização V3, quando a granularidade temporal escolhida é por dias, consegue-se observar em *tooltip* informações gerais sobre esse dia, nomeadamente, o valor mínimo, máximo e médio de ocupação e, ainda, o valor da taxa de ocupação média. Em adição a esta informação, é ainda apresentado um gráfico de linhas que contém a média da taxa de ocupação horária para aquele dia. Um exemplo desta *tooltip* está presente na figura 4.14. As *tooltips* dos restantes gráficos apresentam informação que já era possível observar através da visualização.

Muitas das interações presentes neste *dashboard* são equivalentes às interações do *dashboard* das bicicletas Gira. Após o utilizador selecionar um estado da pandemia no filtro F1, é realizada uma ação de filtragem na visualização V2 e, deste modo, a visualização V2 passa a mostrar informação relativa somente ao estado selecionado. Para além desta ação de filtro, é também realizado um destaque do período temporal relativo a esse estado da pandemia nas visualizações V3 e V4. Para analisar os dados de um parque específico, o utilizador pode clicar sobre ele em V1 e, após isso, as visualizações V2 e V3 são filtradas. Para além disso, é ainda feito um destaque do parque selecionado na visualização V4, isto caso o parque selecionado pertença ao top 10 parques de estacionamento com maior taxa de ocupação. Se for necessário procurar por um parque, basta escrever o nome do mesmo em H1 e este é destacado no mapa. Caso se pretenda analisar um conjunto de parques

consoante o seu tamanho, basta clicar sobre um elemento do filtro F2 e as visualizações V2 e V3 são filtradas. Para além destas, é também filtrado o mapa em V1, de modo a que o mesmo apresente somente os parques que pertencem ao grupo selecionado. Caso se pretenda analisar, para um período temporal específico, a alteração média da taxa de ocupação por hora comparativamente a 14 dias atrás, é possível clicar sobre o período pretendido na visualização V3 e a visualização V2 é filtrada para os valores correspondentes. Por fim, quando se clica sobre um parque presente no top 10 (em V4) é feito um destaque da sua localização no mapa em V1 e os gráficos V2 e V3 são filtrados.

Concluindo, com esta ferramenta é possível responder a várias questões analíticas mais específicas, como por exemplo as questões listadas na secção 4.1 :

- **“O parque de estacionamento da Avenida Lusíada ficou, na sua maioria, mais cheio ou mais vazio após o aparecimento da COVID-19?” (Q02.01).** Para responder a esta questão é necessário selecionar o parque da Avenida Lusíada no mapa, ou então ver a resposta diretamente através do *heat table* na visualização V4.
- **“Relativamente à alteração média da taxa de ocupação por hora comparativamente com 14 dias atrás, qual é o comportamento observado no parque Avenida Lusíada durante o mês de junho?” (Q02.02).** Para obter a resposta a esta questão é necessário selecionar o parque da Avenida Lusíada no mapa e o mês de junho na visualização V3.
- **“Comparando a altura do estado de calamidade com o restante período temporal, considera-se que existiu alguma alteração no padrão da taxa de ocupação diária dos parques de estacionamento presentes no top 10?” (Q02.03).** É possível responder a esta questão após a seleção do estado de calamidade no filtro F1.

#### 4.4.3 Análise geral dos congestionamentos detetados pelo Waze na cidade de Lisboa

Para além da análise feita às bicicletas Gira e aos parques de estacionamento, foram também efetuadas análises relativas aos congestionamentos na cidade de Lisboa. Todas estas análises permitem compreender como foi o panorama no que toca à mobilidade em alturas antes e durante a pandemia, de modo a averiguar se a COVID-19 teve, efetivamente, impacto na mobilidade da cidade de Lisboa.

Foi desenvolvido um *dashboard* que tem como intuito responder à seguinte questão QG03, levantada na secção 4.1: “O aparecimento da COVID-19 teve impacto nos congestionamentos?”. A figura 4.15 mostra o *dashboard* desenvolvido para responder a essa questão, sendo que na imagem estão também presentes denominações que servem de auxílio para a explicação do mesmo. Para além disso, este [vídeo](#) mostra uma explicação das funcionalidades presentes neste *dashboard*.

Para o desenvolvimento deste *dashboard* houve a necessidade de analisar quais seriam os melhores indicadores de tráfego para mostrar o comportamento dos congestionamentos na cidade de Lisboa. Visto que congestionamentos são eventos discretos, os mesmos

podem ser contabilizados de forma a mostrar as diferenças ocorridas. Para além disso, num dos momentos de reunião com a CML, foi mencionado que os indicadores que têm sido, atualmente, mais úteis na medição do tráfego na cidade são o número de congestionamentos únicos do nível 2 e 3 e do nível 4 e 5. O comprimento dos congestionamentos também se mostrou um bom indicador, principalmente quando utilizado conjuntamente com o número de congestionamentos únicos. Por outro lado, decidiu-se não utilizar o atraso provocado pelo congestionamento. Isto, pois este valor depende de vários fatores externos, nomeadamente, do tamanho da via e do congestionamento, de modo que um atraso de trinta minutos num congestionamento de vários quilómetros é diferente de um atraso de trinta minutos num congestionamento de poucos metros. Para além disso, os níveis de intensidade já têm em consideração o atraso provocado no tráfego.

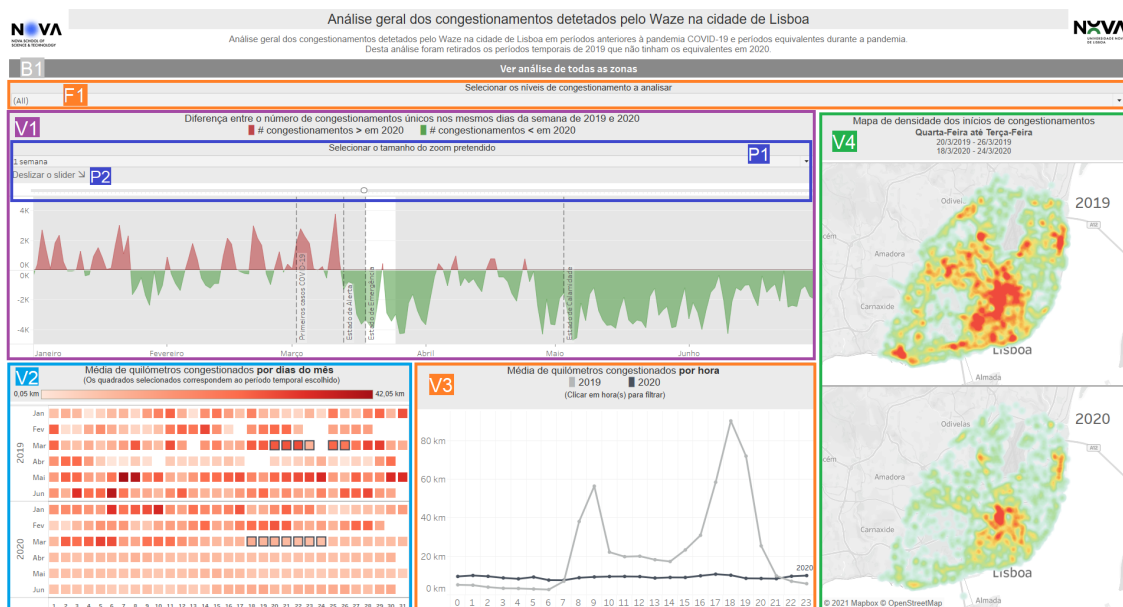


Figura 4.15: *Dashboard* para a análise geral aos congestionamentos detetados pelo Waze na cidade de Lisboa.

Como é possível observar através da figura 4.15, este *dashboard* apresenta quatro visualizações, dois parâmetros, um filtro e um botão.

A visualização V1 consiste num gráfico de área no qual é mapeado a diferença entre o número de congestionamentos únicos, nos mesmos dias da semana, entre 2019 e 2020. Deste modo, um valor positivo significa que em 2020 existiu um maior número de congestionamentos nesse dia da semana, sendo que é atribuído a essa área uma cor vermelha. Por outro lado, um valor negativo significa que em 2020 o número de congestionamentos foi menor, então é-lhe mapeada uma cor verde. Este gráfico, para além de mostrar um panorama geral de como se comportaram os congestionamentos, serve ainda como filtro das restantes visualizações. Isto, pois tem associado a ele dois parâmetros que permitem a escolha do período temporal a analisar com maior detalhe nas restantes visualizações. No parâmetro P1 é possível selecionar o tamanho do zoom pretendido e, com o parâmetro

P2, é possível escolher a data a partir da qual se pretende analisar, sendo que isto é feito deslizando o *slider* até à data escolhida. A visualização V2 consiste num *heat table* no qual é mostrado a média de quilómetros congestionados por dia do mês, sendo que uma cor mais intensa representa uma média maior. Nesta visualização é ainda possível observar que os quadrados realçados são referentes ao período temporal escolhido através dos parâmetros P1 e P2. Em V3 está representado um gráfico de linhas onde é mostrada a média de quilómetros congestionados por hora. Neste gráfico a linha cinzenta escura mapeia os valores do período temporal de 2020, selecionado através dos parâmetros P1 e P2, e a cinzenta clara mapeia os valores de 2019. Na visualização V4 encontram-se presentes dois mapas de densidade dos inícios de congestionamentos, ou seja onde começaram os congestionamentos. Sendo que o primeiro mapa é relativo ao período selecionado para 2019 e o segundo a 2020.

Todas as visualizações deste *dashboard* apresentam *tooltips* no formato textual. Em todas elas é apresentada informação temporal relativa a cada elemento e ainda informação que já era possível observar nas visualizações, como por exemplo, a diferença de congestionamentos e a média de quilómetros congestionados.

A maioria das interações presentes neste *dashboard* são feitas através de ações de filtragem. Com o filtro F1 é possível filtrar todas as restantes visualizações de modo a apresentarem dados de todos os congestionamentos, dados dos congestionamentos de nível 2 e 3 ou dados dos congestionamentos de nível 4 e 5. Tal como explicado anteriormente, a visualização V1, em conjunto com os parâmetros P1 e P2, permite escolher o período temporal apresentado nas restantes visualizações através de uma ação de filtragem. Sendo que na própria visualização é feito um destaque, através de bandas verticais de referência, do período temporal que está a ser filtrado nos restantes gráficos. Para além disso, o período temporal selecionado é também destacado em V2. Por último, existe ainda a possibilidade de filtrar a visualização V2 e V4 de modo a mostrarem apenas os dados relativos a uma hora específica. Para tal, é necessário clicar sobre uma hora, ou um conjunto de horas, na visualização V3. Após isto, é ainda mostrado, através de duas pequenas visualizações flutuantes, quais as horas dos dados que estão a ser exibidos em V4.

Este *dashboard* apresenta ainda um botão (B1) que, após ser clicado, faz com que o utilizador seja redirecionado para um novo *dashboard*. O *dashboard* para onde o utilizador é redirecionado permite ver uma análise relativa a todas as zonas consideradas importantes da cidade de Lisboa, sendo que será feita uma explicação mais detalhada ao mesmo na subsecção 4.4.4.

Concluindo, com esta ferramenta é possível responder a várias questões analíticas mais específicas, como por exemplo as questões analíticas formuladas na secção 4.1:

- **“No geral, houve mais ou menos congestionamentos em 2020 na altura da COVID-19?” (Q03.01).** Para responder a esta questão não é necessário aplicar ações de filtragem, basta simplesmente analisar para a visualização V1.

- “Comparativamente a 2019, considera-se que a densidade espacial dos congestionamentos na cidade de Lisboa se alterou durante as duas primeiras semanas do estado de calamidade?” (Q03.02). É possível responder a esta questão através da seleção de duas semanas no parâmetro P1 e através da seleção do primeiro dia do estado de calamidade no *slider* do parâmetro P2.
- “Em 2020, durante as duas primeiras semanas do estado de emergência, existiram picos horários da média de quilómetros congestionados?” (Q03.03). A resposta a esta questão obtém-se após a seleção de duas semanas no parâmetro P1 e após a seleção do primeiro dia do estado de emergência no *slider* do parâmetro P2.

#### 4.4.4 Análise dos congestionamentos detetados pelo Waze em algumas zonas de Lisboa

Após ter sido criada uma ferramenta que permite a análise geral dos congestionamentos na cidade de Lisboa, sentiu-se a necessidade de criar outra ferramenta que analise somente as zonas da cidade de Lisboa listadas em 4.1. Este *dashboard* permite responder à questão analítica QG04 da listagem feita em 4.1: “Quais foram as zonas da cidade de Lisboa que sofreram um maior impacto no número de congestionamentos após o aparecimento da COVID-19?”. Na figura 4.16 encontra-se presente uma imagem do *dashboard* criado, bem como as denominações aos seus elementos que servem de auxílio à explicação do mesmo. Através deste [vídeo](#) é possível ver uma explicação das funcionalidades presentes neste *dashboard* e das funcionalidades presentes no *dashboard* que será apresentado posteriormente nesta subsecção.

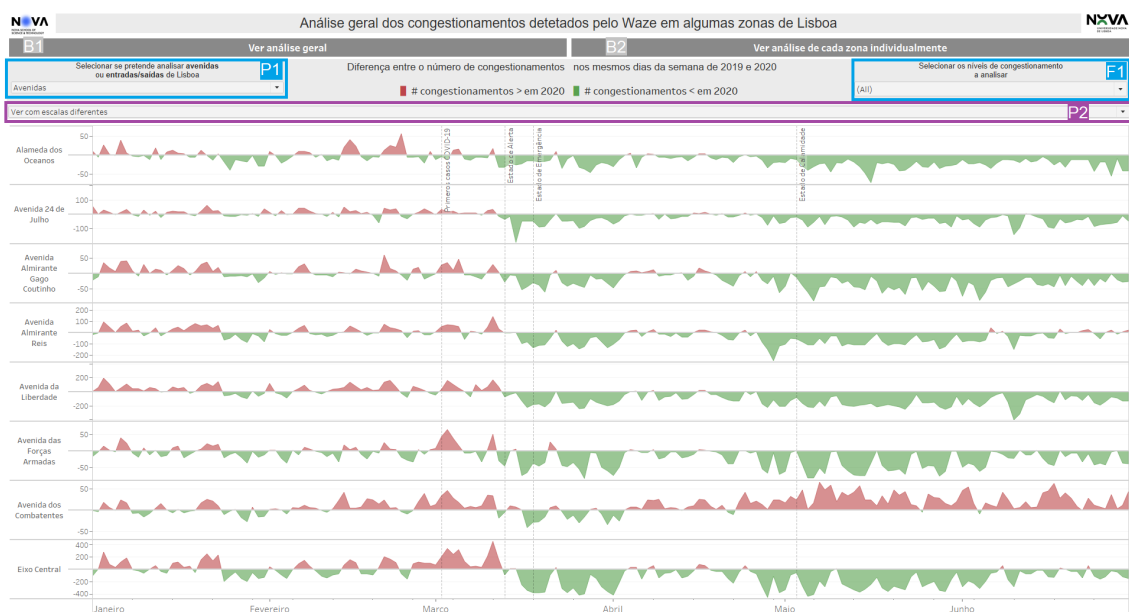


Figura 4.16: *Dashboard* para a análise geral aos congestionamentos detetados pelo Waze em algumas zonas da cidade de Lisboa.

Através da figura 4.16 constata-se que este *dashboard* exibe uma visualização, dois parâmetros, um filtro e dois botões. É importante referir que este *dashboard* é acessível através da ferramenta analisada na subsecção 4.4.3.

A visualização presente neste *dashboard* mostra, para todas as zonas, qual a diferença entre o número de congestionamentos nos mesmos dias da semana de 2019 e 2020. Isto é representado através de um gráfico de área, no qual valores negativos indicam que em 2020 houve menos congestionamentos e valores positivos indicam que em 2020 houve mais congestionamentos comparativamente ao mesmo dia da semana de 2019. As áreas referentes a alturas com mais congestionamentos em 2020 são mapeadas através da cor vermelha e as áreas com menos congestionamentos em 2020 são mapeadas através da cor verde. Existe mais uma visualização neste *dashboard* que, após a mudança do parâmetro P2, substitui a única visualização visível no *dashboard*. Com o parâmetro P2 é possível escolher o que se pretende observar, sendo que é possível escolher entre uma visualização que apresenta a mesma escala, no eixo Y, para todas as zonas e uma visualização que apresenta escalas diferentes para cada zona. Ambas as visualizações permitem observar fenómenos diferentes, por exemplo, quando as zonas apresentam escalas iguais é possível constatar qual delas é que apresentou uma maior mudança de comportamento, comparativamente às restantes. Por outro lado, quando as zonas apresentam escalas diferentes é possível observar com maior facilidade o comportamento isolado de cada zona.

As interações presentes neste *dashboard* dizem respeito apenas a ações de filtragem e mudança de parâmetros. Com o parâmetro P1 é possível escolher quais as zonas a ser mapeadas na visualização, sendo que as opções disponíveis são “Avenidas” ou “Entradas/Saídas”. O filtro F1 faz com que os valores mostrados na visualização sejam filtrados consoante os níveis que se pretende analisar, sendo possível escolher entre todos os níveis, níveis 2 e 3 e níveis 4 e 5.

A *tooltip* que é apresentada neste gráfico consiste em informação, na forma textual, sobre a diferença no número de congestionamentos entre 2020 e 2019 e informação temporal de quais os dias referentes a esse valor.

Por fim, existem ainda dois botões no topo deste *dashboard*. Ao clicar no botão B1 é feito um redirecionamento para o *dashboard* que proporciona uma análise geral dos congestionamentos ocorridos na cidade de Lisboa (descrito em 4.4.3). O botão B2 redireciona para um *dashboard* que permite a análise dos congestionamentos ocorridos em cada zona individual. Este *dashboard* encontra-se ilustrado na figura 4.17, bem como as denominações aos seus elementos que servem como auxílio na descrição dos mesmos.

O *dashboard* da figura 4.17 contém seis visualizações, três parâmetros, um filtro, uma legenda interativa e um botão.

Na visualização V1 consta um gráfico de áreas empilhadas, no qual é possível observar o número diário de congestionamentos únicos por níveis para os anos 2019 e 2020. Tendo em conta que o tráfego difere entre dias de trabalho e fins de semanas, sentiu-se a necessidade de fazer o alinhamento por dias da semana. Ou seja, no gráfico, a primeira segunda-feira de 2019 encontra-se alinhada com a primeira segunda-feira de 2020. As

áreas deste gráfico apresentam as cores vermelho e amarelo, sendo que a cor vermelho é referente a congestionamentos do nível 4 e 5 e amarelo a congestionamentos do nível 2 e 3. Na legenda L1 é possível ver o significado destas cores. Para além disso, a visualização V1, em conjunto com os parâmetros P2 e P3, serve ainda como filtro do período temporal exibido nas restantes visualizações. No parâmetro P2 é possível selecionar o tamanho do zoom pretendido e, deslizando o *slider* do parâmetro P3, é possível escolher a data a partir da qual se pretende analisar nas restantes visualizações. A visualização V2 consiste num gráfico de barras no qual é mostrado, através do tamanho e do rótulo da barra, a média diária do comprimento dos congestionamentos no período temporal selecionado, sendo que cada barra é relativa a um ano diferente. Na visualização V3 existe, sob o formato textual, informação acerca da média diária de congestionamentos no período temporal escolhido em 2019. Na visualização V4 encontra-se um mapa, no qual está mapeado a zona escolhida e também a localização dos congestionamentos no período temporal selecionado para 2019, sendo que a cor da linha indica o nível do congestionamento. No lado direito das visualizações V3 e V4 estão presentes duas visualizações idênticas, porém apresentam os dados do período selecionado para 2020. Existe ainda, no topo do *dashboard*, um parâmetro P1 e um filtro F1 que permitem escolher qual a zona em análise.

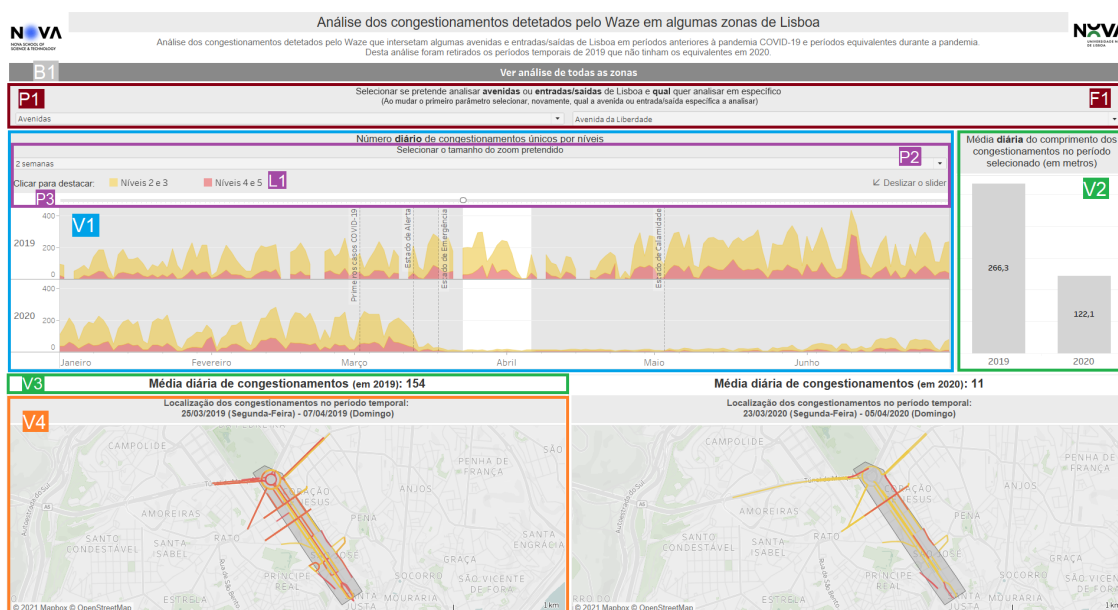


Figura 4.17: *Dashboard* para a análise dos congestionamentos detetados pelo Waze numa zona específica da cidade de Lisboa.

As *tooltips* existentes neste *dashboard* estão todas no formato textual e apresentam, na sua maioria, informação já visível nos gráficos e informação temporal. Na visualização V1 é possível ver em *tooltip* o número de congestionamentos diários existentes em cada conjunto de níveis. Para além disso, é também possível ver quantos desses congestionamentos são referentes a trânsito completamente parado. No mapa V4, e no mapa à sua direita, é possível observar em *tooltip* o nome da rua onde se localizou o congestionamento.

As interações presentes neste *dashboard* baseiam-se em ações de filtragem e de destaque. Através do filtro F1 é possível filtrar todas as visualizações de modo a mostrarem os dados relativos à zona escolhida. Os valores apresentados neste filtro diferem consoante a escolha do parâmetro P1. Ou seja, caso seja escolhido em P1 o valor “Avenidas”, o filtro F1 apresentará os nomes das avenidas disponíveis para análise. Caso o valor escolhido em P1 seja “Entradas/Saídas”, os valores apresentados em F1 serão os nomes das entradas/saídas disponíveis. Como visto anteriormente, através da visualização V1 e dos parâmetros P2 e P3 é possível filtrar todas as restantes visualizações de modo a mostrarem os valores do período temporal escolhido. Para além disso, é ainda feito um destaque em V1, através de bandas verticais de referência, do período temporal escolhido. Existe ainda uma interação de destaque que pode ser realizada clicando num elemento da legenda L1. Isto fará com que os congestionamentos do nível selecionado sejam destacados na visualização V1 e nas visualizações que apresentam mapas.

Por último, existe ainda um botão no topo do *dashboard* que redireciona para o *dashboard* da figura 4.16.

Concluindo, com estas duas ferramentas que possibilitam uma análise a algumas zonas da cidade de Lisboa, é possível responder a várias questões analíticas mais específicas, como por exemplo as questões listadas na secção 4.1:

- **“Quais foram as avenidas de Lisboa onde se sentiu um maior impacto da COVID-19 em relação ao número de congestionamentos?” (Q04.01).** Esta questão é respondida através do *dashboard* que permite a análise geral dos congestionamentos. Para responder a esta questão basta selecionar avenidas no parâmetro P1 e selecionar, no parâmetro P2, que se pretende ver o eixo do Y com escalas iguais.
- **“Existiram mudanças significativas no número de congestionamentos diários ocorridos na Avenida da Liberdade após a declaração do estado de alerta?” (Q04.02).** Esta questão é respondida através do *dashboard* que permite a análise dos congestionamentos numa zona específica. É possível responder a esta questão após selecionar, no parâmetro P1 e no filtro F1, que se pretende analisar avenidas, mais propriamente a Avenida da Liberdade.
- **“Existiu alguma diferença na localização dos congestionamentos da Autoestrada do Norte (A1) durante as duas primeiras semanas do estado de emergência?” (Q04.03).** Esta questão é respondida através do *dashboard* que permite a análise dos congestionamentos numa zona específica. Para obter a resposta a esta questão basta selecionar, no parâmetro P1 e no filtro F1, que se pretende analisar a entrada/saída da Autoestrada do Norte (A1). Após isso, é necessário selecionar no parâmetro P2 o valor de duas semanas e selecionar no *slider* do parâmetro P3 o primeiro dia do estado de emergência.

#### 4.4.5 Análise dos congestionamentos detetados pelo Waze em redor de pontos de interesse de Lisboa

A análise dos congestionamentos ao redor de alguns pontos de interesse da cidade de Lisboa também se mostrou uma mais valia para a determinação do impacto que a COVID-19 teve na mobilidade. As declarações dos estados de alerta, emergência, calamidade e contingência implicaram mudanças significativas no padrão de vida dos cidadãos, sendo que essas mudanças afetaram, de certa forma, o tráfego junto aos POI's.

Deste modo, foi desenvolvida uma ferramenta que permite responder à questão geral QG05, levantada na secção 4.1, de “Como se comportaram os congestionamentos ao redor dos pontos de interesse após o aparecimento da COVID-19?”. O *dashboard* desenvolvido encontra-se presente na figura 4.18, bem como as denominações aos seus elementos que servirão como auxílio na explicação dos mesmos. Neste vídeo é ainda possível ver uma explicação das funcionalidades presentes neste *dashboard*. Na criação desta ferramenta foram utilizados os conjuntos de dados descritos na subsecção 4.3.8 e também o conjunto de dados dos POI's, descrito em 4.3.7.

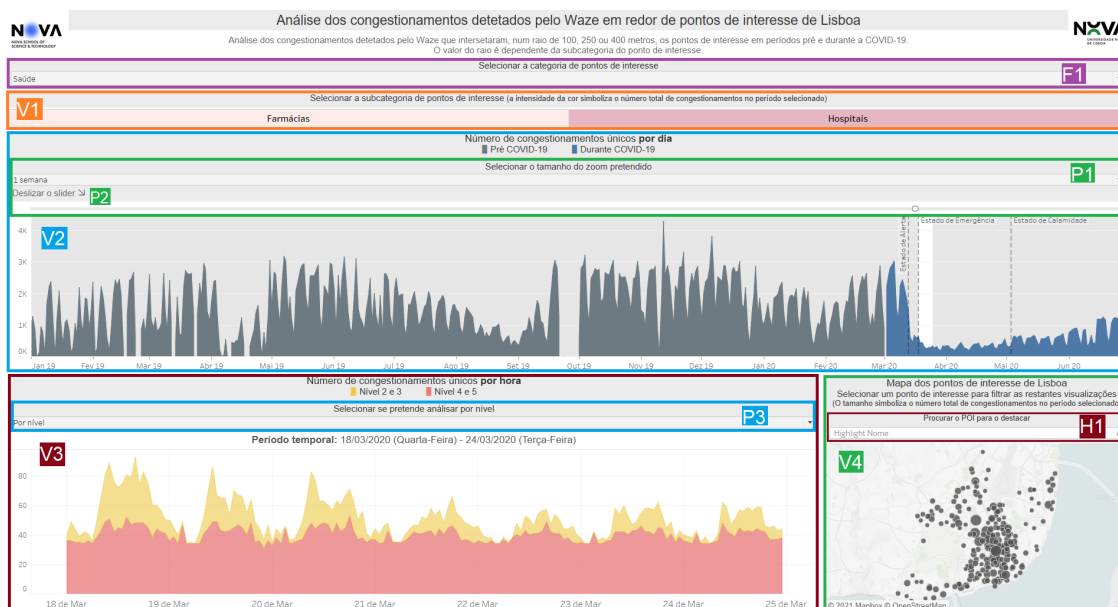


Figura 4.18: *Dashboard* para a análise dos congestionamentos detetados pelo Waze ao redor de pontos de interesse da cidade de Lisboa.

Através da figura 4.18 é possível constatar que este *dashboard* exhibe quatro visualizações, um filtro, três parâmetros e um *highlighter*.

A visualização V1 foi concebida com o propósito de servir, maioritariamente, de filtro às restantes visualizações. Sendo que nela são apresentadas, sob o formato textual, as subcategorias existentes dentro da categoria selecionada no filtro F1. Para cada elemento desta visualização é ainda mapeado uma cor, sendo que a intensidade da cor denota o número de congestionamentos ocorridos, em cada subcategoria, no período temporal

que será escolhido através dos parâmetros P1 e P2. Isto permite que o utilizador tenha, rapidamente, uma noção de qual foi a subcategoria que apresentou maior tráfego nas suas redondezas. A visualização V2 é constituída por um gráfico de área no qual é representado o número de congestionamentos únicos por dia. Neste gráfico a área a cinzento representa os períodos temporais anteriores à pandemia e a área a azul representa os períodos temporais durante a pandemia. Para além disso, e de modo semelhante às ferramentas vistas anteriormente, nesta visualização é ainda possível escolher, através dos parâmetros P1 e P2, o período temporal que é aplicado às restantes visualizações. Na visualização V3 é exibido um gráfico de áreas empilhadas que apresenta, para o período temporal selecionado em V2, o número de congestionamentos únicos por horas, ou seja, mostra em maior detalhe o que é visto em V2. Neste gráfico a área a vermelho representa o número de congestionamentos do nível 4 e 5 e a área a amarelo os congestionamentos do nível 2 e 3. A visualização V4 é constituída por um mapa que contém a localização dos pontos de interesse da categoria/subcategoria que se está a analisar. O tamanho dos círculos, que mapeiam a localização do POI, difere consoante o número de congestionamentos ocorridos nas suas redondezas durante o período temporal escolhido. Isto faz com que seja rapidamente perceptível quais foram os pontos de interesse que tiveram uma maior intensidade de tráfego ao seu redor.

As *tooltips* presentes nas visualizações deste *dashboard* são apresentadas ao fazer *hover* nos elementos das mesmas e todas são exibidas no formato textual. As *tooltips* das visualizações V2 e V3 apresentam informação temporal e sobre o número de congestionamentos. Na visualização V4 é mostrado o nome do ponto de interesse, a sua categoria e subcategoria, o número de congestionamentos ocorridos ao seu redor e o tamanho do *buffer* que foi utilizado durante o processo de cálculo dos congestionamentos que o interestavam.

De forma semelhante aos *dashboards* vistos anteriormente, este também apresenta uma forte componente interativa. Através do filtro F1 é possível selecionar qual a categoria de pontos de interesse que se pretende analisar, sendo que é possível escolher entre: “Ensino”, “Lazer”, “Saúde” e “Transporte”. A seleção deste filtro irá afetar todas as visualizações presentes neste *dashboard*. De seguida, é possível escolher, em V1, a subcategoria, ou conjunto de subcategorias, que se tenciona analisar com maior detalhe. Ao selecionar um elemento em V1 é feita uma ação de filtragem em todas as outras visualizações. Através dos parâmetros P1 e P2, associados à visualização V2, é possível filtrar o período temporal das visualizações V1, V3 e V4. Tal como já foi visto anteriormente, isto é feito através da escolha do tamanho do zoom em P1 e da escolha da data de início através do *slider* em P2. Esta ação fará com que seja realizado um destaque em V1, através de bandas verticais de referência, do período escolhido. Para além de ser possível analisar por categoria e subcategoria, é também possível analisar um ponto de interesse em específico. Para tal, basta clicar, em V4, sobre a localização do POI pretendido. Esta ação faz com que as visualizações V2 e V3 sejam filtradas de modo a mostrarem apenas os dados referentes ao POI escolhido. Caso se pretenda procurar por um ponto de interesse, basta escrever o nome do mesmo em H1 e é feito um destaque da sua localização no mapa.

Por fim, existe ainda a possibilidade de escolher a informação que é apresentada em V3. Isto é feito através da mudança do parâmetro P3, sendo que este permite escolher entre ver o número de congestionamentos separados por níveis ou sem estarem separados por níveis. A escolha do valor “Sem nível” no parâmetro P3 faz com que a visualização V3 seja substituída pelo gráfico da figura 4.19. Este é um gráfico de linhas que mostra, para o período temporal selecionado, o número total de congestionamentos por hora, sendo que o mapeamento da cor da linha é idêntico ao mapeamento feito em V2.



Figura 4.19: Visualização exibida após a mudança do parâmetro P3 para “Sem nível” no *dashboard* que analisa os congestionamentos ao redor dos pontos de interesse.

Concluindo, através desta ferramenta é possível responder a várias questões analíticas mais específicas, como por exemplo as questões apresentadas na secção 4.1:

- **“Como se comportaram os congestionamentos em redor dos pontos de interesse da categoria ‘Saúde’ na cidade de Lisboa após o aparecimento da COVID-19?” (Q05.01).** É possível responder a esta questão através da seleção da categoria saúde no filtro F1.
- **“Durante a primeira semana do estado de calamidade houve mais congestionamentos ao redor de estabelecimentos do ensino superior ou do ensino secundário?” (Q05.02).** A resposta a esta questão é encontrada após a seleção da categoria ensino no filtro F1, após a seleção no parâmetro P1 do valor de uma semana e após a seleção no *slider* do parâmetro P2 do primeiro dia do estado de calamidade.
- **“Durante a primeira semana do estado de emergência, qual foi o dia que apresentou um maior número de congestionamentos ao redor do Hospital Santa Maria?” (Q05.03).** É possível responder a esta questão selecionando a categoria saúde no filtro F1 e o Hospital Santa Maria no mapa.

#### 4.4.6 Análise da qualidade do ar da cidade de Lisboa

Para além da determinação do impacto da COVID-19 na mobilidade, é ainda crucial analisar o impacto que a COVID-19 teve na qualidade do ar da cidade de Lisboa. Para tal, foi desenvolvida uma ferramenta que permite responder à questão QG06: “Após o aparecimento da COVID-19, existiram alterações nas concentrações dos poluentes atmosféricos medidos nas estações da cidade de Lisboa?”, levantada na secção 4.1. Na figura

4.20 encontra-se uma imagem do *dashboard* desenvolvido e as denominações aos seus elementos, utilizadas durante o processo de explicação dos mesmos. Para além disso, neste *vídeo* é mostrado uma explicação das funcionalidades que este *dashboard* apresenta.

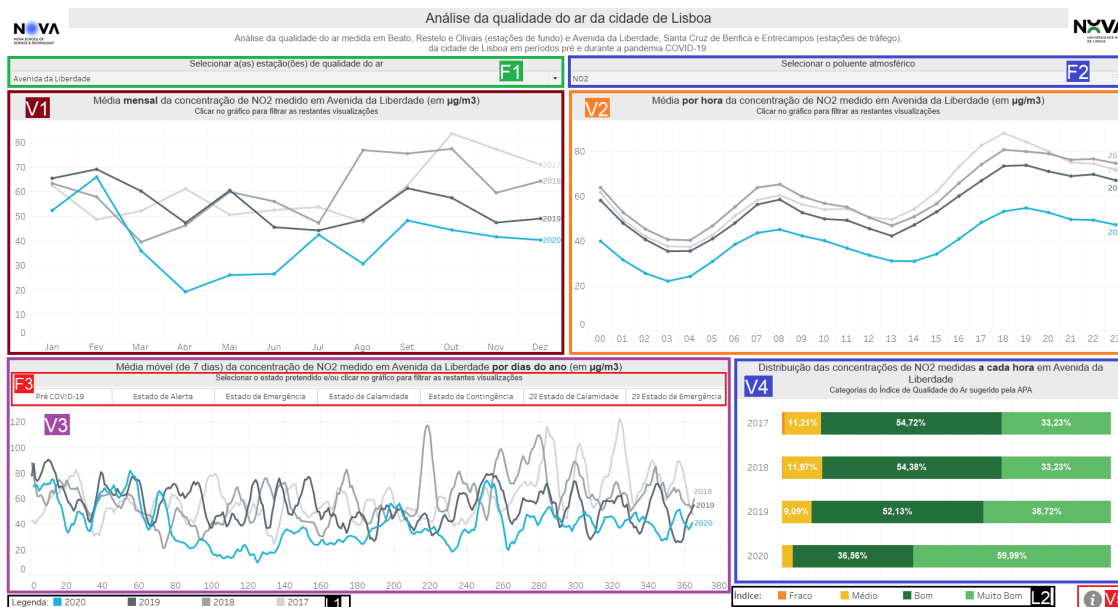


Figura 4.20: *Dashboard* para a análise da qualidade do ar da cidade de Lisboa.

Este *dashboard* apresenta cinco visualizações, três filtros e duas legendas interativas, tal como é possível observar a partir da figura 4.20.

A visualização V1 mostra a média mensal da concentração de um poluente medido numa estação, ou num conjunto de estações. O poluente em análise é selecionado através do filtro F2 e a estação em que o mesmo foi medido é escolhida através do filtro F1. Esta informação é mostrada através de um gráfico de linhas, sendo que cada linha representa um ano diferente, nomeadamente, 2017, 2018, 2019 e 2020. A visualização V2 é semelhante a esta, porém mapeia a média por hora da concentração do poluente medido na estação escolhida. A visualização V3 é, novamente, semelhante às anteriores, contudo nesta é apresentada a média móvel, de 7 dias, da concentração diária do poluente medido na estação escolhida. Optou-se por apresentar a média móvel pois desta forma é reduzido o ruído e as variações nas concentrações são acentuadas. Na legenda L1 é apresentado o mapeamento das cores feito nestas três visualizações. Associada à visualização V3 encontra-se ainda o filtro F3. Este filtro apresenta, no formato textual, os estados declarados durante a pandemia COVID-19. Sendo que o seu intuito é proporcionar ações de filtro e destaque nas restantes visualizações. A visualização V4 mostra a distribuição das concentrações horárias do poluente pelas categorias do índice de qualidade do ar. Isto é, um valor de 30% na categoria “Muito Bom” indica que 30% das horas desse ano estiveram com valores de concentração considerados muito bons. Esta visualização é composta por um gráfico de barras empilhadas, sendo que cada barra é referente a um ano diferente. Os valores representados nas barras correspondem à percentagem do total de cada célula,

ou seja, a conta é feita para cada ano individualmente, considerando, deste modo, um ano como sendo 100%. Cada classe do índice de qualidade do ar tem uma cor associada, sendo que é possível ver esse mapeamento de cores através da legenda L2. Para além disso, através da visualização V5 é possível ver informação acerca dos limites das concentrações de cada classe do índice de qualidade do ar.

Este *dashboard* contém dois tipos de *tooltips*: *tooltips* no formato textual e *tooltips* no formato gráfico. Na visualização V1, ao fazer *hover*, é mostrado em *tooltip* informação de cariz temporal e informação sobre a média da concentração do poluente nesse mês. Para além disso, é mostrado um gráfico de barras que apresenta a média, por dias da semana, do poluente medido na estação selecionada. Um exemplo desta *tooltip* está presente na figura 4.21. A visualização V5 apresenta, igualmente, uma *tooltip* no formato gráfico. E, tal como é possível observar através da figura 4.22, é exibido uma imagem que mostra, para cada poluente, os limites dos valores das suas concentrações em cada classe do índice de qualidade do ar. As *tooltips* das restantes visualizações têm somente informação no formato textual, nomeadamente, informação temporal e informação sobre a média do poluente na estação escolhida.

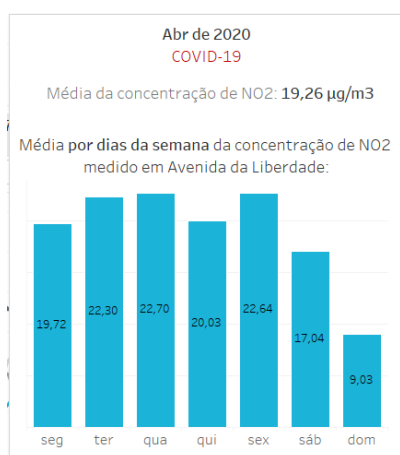


Figura 4.21: Exemplo de *tooltip* do *dashboard* da qualidade do ar. Referente à visualização V1, quando é feito *hover* no mês de abril de 2020.

Classes do Índice de Qualidade do Ar  
(valores em µg/m3)

Classificação	PM10	PM2.5	NO2	O3	SO2
Muito Bom	0-20	0-10	0-40	0-80	0-100
Bom	21-35	11-20	41-100	81-100	101-200
Médio	36-50	21-25	101-200	101-180	201-350
Fraco	51-100	26-50	201-400	181-240	351-500
Mau	101-1200	51-800	401-1000	241-600	501-1250

Índice: Fraco Médio Bom Muito Bom

Figura 4.22: Exemplo de *tooltip* do *dashboard* da qualidade do ar. Apresentado quando é feito *hover* em V6.

No que toca à interatividade, neste *dashboard* existem ações de filtragem e de destaque. Os filtros F1 e F2, presentes no topo do *dashboard*, fazem com que os dados das visualizações V1, V2, V3 e V4 sejam filtrados. Tal como já foi referido, no filtro F1 é possível escolher a estação, ou conjunto de estações, que se pretende analisar, sendo que as escolhas possíveis são: “Avenida da Liberdade”, “Beato”, “Entrecampos”, “Olivais”, “Restelo” e “Santa Cruz de Benfica”. No filtro F2 é possível escolher o poluente atmosférico a analisar, sendo que apenas se pode escolher entre os poluentes que são medidos na estação escolhida. Na visualização V1, ao clicar sobre um mês ou um conjunto de meses, é feito

uma ação de filtro nas visualizações V2, V3 e V4, sendo que estas passam a apresentar apenas os dados do mês escolhido. De forma semelhante, na visualização V2, existe a possibilidade de clicar sobre uma hora, ou um conjunto de horas, de modo a filtrar as visualizações V1, V3 e V4. Identicamente a estas, na visualização V3 também é possível clicar sobre um dia, ou um conjunto de dias, com o intuito de realizar uma ação de filtragem nas visualizações V2 e V4. Na visualização V3, ao fazer *hover* sobre um dia, é ainda feito um destaque desse mesmo dia nos restantes anos, de modo a ser possível comparar rapidamente com o valor dos outros anos. O filtro F3 possibilita, de uma forma rápida, a filtragem e o destaque dos dados das restantes visualizações para o período temporal referente aos estados declarados durante a pandemia. Assim sendo, ao fazer *hover* sobre um estado declarado durante a pandemia, é feito, na visualização V3, um destaque dos dias que fazem parte do mesmo. Para além disso, ao clicar sobre um estado, é feita uma filtragem aos dados presentes em V2, V3 e V4, de modo a apresentarem dados relativos ao período temporal do estado escolhido. Por fim, ao clicar sobre um elemento de cada legenda é feito um destaque nas respetivas visualizações. Sendo que, ao clicar sobre um ano da legenda L1, é feito um destaque na visualização V1, V2, V3 e V4. Ao clicar sobre uma classe de qualidade do ar na legenda L2, é feito, em V4, um destaque da mesma.

Concluído, com esta ferramenta é possível responder a um longo conjunto de questões analíticas mais específicas, como por exemplo as questões colocadas na secção 4.1:

- **“Quais foram os três meses que, na estação da Avenida da Liberdade, apresentaram uma maior mudança no valor médio do poluente NO<sub>2</sub> durante o ano 2020, comparativamente aos anos anteriores?” (Q06.01).** É possível responder a esta questão através da seleção da estação da Avenida da Liberdade no filtro F1 e do poluente NO<sub>2</sub> no filtro F2.
- **“Para o conjunto de estações da Avenida da Liberdade e Entrecampos, qual foi o estado da pandemia que parece ter tido um maior impacto nos valores médios diários de NO<sub>2</sub> em 2020?” (Q06.02).** Para responder a esta questão basta selecionar a estação da Avenida de Liberdade e de Entrecampos no filtro F1 e selecionar o poluente NO<sub>2</sub> no filtro F2. Ao fazer *hover* sobre os estados da pandemia no filtro F3 é possível ter-se uma maior perceção de qual foi o estado da pandemia onde ocorreu um maior impacto.
- **“Relativamente à estação de Entrecampos e durante o 1º estado de emergência, considera-se que a distribuição das concentrações de NO<sub>2</sub> pelas classes do índice de qualidade do ar foram idênticas nos quatro anos?” (Q06.03).** A resposta a esta questão é obtida através da seleção da estação de Entrecampos no filtro F1, do poluente NO<sub>2</sub> no filtro F2 e do estado de emergência no filtro F3.

#### 4.4.7 Análise da qualidade do ar nas seis estações da cidade de Lisboa

Para além de analisar a qualidade do ar de forma agrupada, sentiu-se a necessidade de desenvolver uma ferramenta que possibilitasse ver os valores da concentração de um poluente atmosférico em todas as estações ao mesmo tempo. Isto permite, não só ter uma visão global do comportamento do poluente em cada estação, mas também comparar se esse comportamento é semelhante entre estações. Deste modo, a ferramenta criada permite responder à questão geral **QG07**, levantada na secção 4.1: “O comportamento de cada poluente atmosférico foi semelhante entre todas as estações em períodos temporais anteriores e durante a pandemia?”. Na figura 4.23 é possível ver uma imagem deste *dashboard*, acompanhado pelas denominações dadas aos seus elementos, de forma a auxiliar na explicação dos mesmos. Neste [vídeo](#) é possível ver uma explicação das funcionalidades presentes neste *dashboard*.

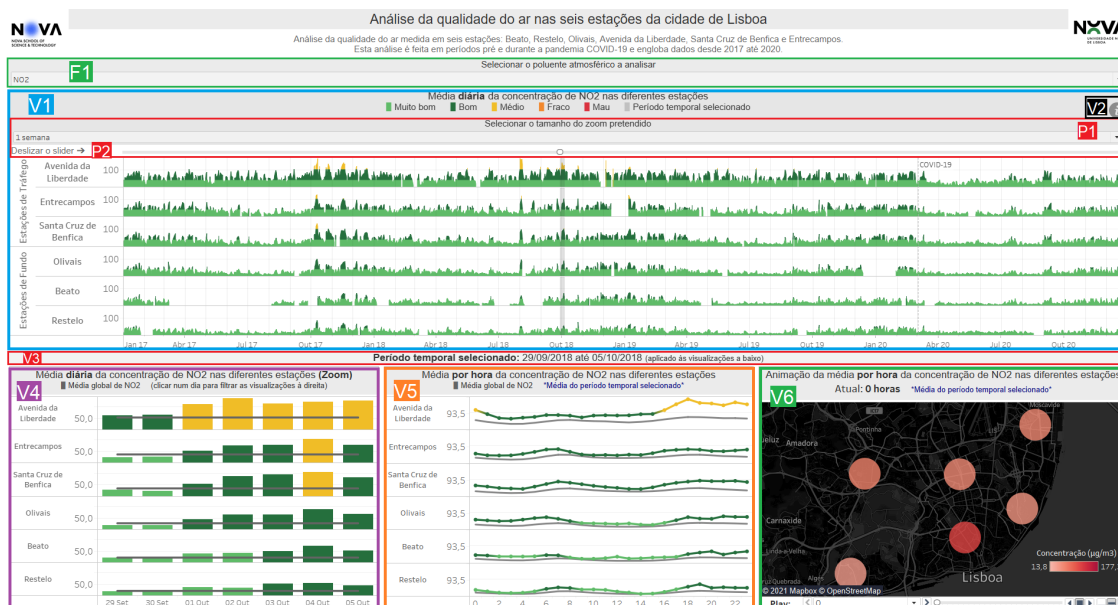


Figura 4.23: *Dashboard* para a análise da qualidade do ar nas seis estações da cidade de Lisboa.

Através da figura 4.23 é possível constatar que este *dashboard* apresenta seis visualizações, dois parâmetros e um filtro.

A visualização V1 mostra, para todas as estações medidas, a média diária da concentração do poluente selecionado no filtro F1. Esta informação é mapeada através de um gráfico de área, sendo que o mesmo se encontra dividido e colorido horizontalmente consoante o limite do poluente em cada classe do índice de qualidade do ar. Decidiu-se utilizar esta abordagem pois com o uso da cor é mais rapidamente perceptível quais foram os dias que estiveram acima de certos valores. Inicialmente pretendia-se que esta visualização fosse um gráfico de barras, contudo, devido ao comprimento do período temporal abrangido, as barras ficavam muito finas e tornavam-se todas cinzentas. Então,

a solução arranjada para conseguir ter um indicativo da cor foi através do gráfico de área dividido e colorido horizontalmente. Esta é a visualização mais geral do *dashboard*, de modo que permite, tal como em ferramentas vistas anteriormente, escolher o período temporal apresentado nas restantes visualizações. Esta escolha é feita com o auxílio dos parâmetros P1 e P2. À semelhança do que foi visto na ferramenta anterior, a visualização V2 apresenta, em *tooltip*, uma imagem informativa das classes do índice de qualidade do ar (figura 4.22). A visualização V3 apresenta, no formato textual, informação acerca de qual é o período temporal aplicado às visualizações V4, V5 e V6, sendo que este período temporal foi escolhido através dos parâmetros P1 e P2. A visualização V4 é um zoom da visualização V1, ou seja, ambas apresentam a mesma informação. Contudo, os dados em V4 apenas abrangem o período temporal escolhido e são apresentados através de um gráfico de barras, em que a cor de cada barra representa o índice de qualidade do ar correspondente. Na visualização V5 é apresentada a média por hora da concentração do poluente em todas as estações que o medem. À semelhança das restantes visualizações, é também mapeado, em cada hora, uma cor referente à classe do índice de qualidade do ar na qual esta se enquadra. Ambas as visualizações V4 e V5 apresentam, para cada estação, uma linha a cinzento que mostra a média global (diária ou por hora) do poluente, isto é, a média em todo o período temporal do *dataset*. Isto permite ver, rapidamente, quais foram os dias, ou horas, que apresentaram valores de concentrações acima, ou abaixo, da média global. Por fim, na visualização V6 é mostrado, para todas as estações, uma animação da média por hora da concentração do poluente selecionado. Esta tem por base um mapa que mostra, através de círculos, a localização de cada estação. Em cada círculo é mapeado, através da intensidade da cor, o valor da concentração média do poluente na hora corrente da animação. No fundo da visualização encontra-se o controlador da animação, sendo que este permite escolher uma hora específica a analisar, andar no tempo através de um *slider*, clicar no *play*, clicar na pausa e escolher a velocidade a que é apresentada a animação.

À exceção da *tooltip* da visualização V2, todas as restantes *tooltips* encontram-se no formato textual e mostram, maioritariamente, informação que já era perceptível através dos seus gráficos. Deste modo, a maioria das visualizações apresenta em *tooltip* informação de cariz temporal, informação sobre o nome da estação e informação sobre a média da concentração do poluente.

As interações existentes neste *dashboard* são feitas através de ações de filtro e de destaque. Através do filtro F1 é possível selecionar qual o poluente que se presente analisar nas visualizações V1, V4, V5 e V6. De forma idêntica ao que foi visto nas ferramentas anteriores, através do parâmetro P1 e P2 é possível filtrar o período temporal apresentado nas visualizações V4, V5 e V6. Sendo que, o período temporal escolhido é destacado em V1. Ao clicar sobre um dia na visualização V4, é realizada uma ação de filtragem que afeta as visualizações V5 e V6. Através desta ação é possível ver, em maior detalhe, o comportamento horário de um poluente num dia específico. Por fim, quando se faz *hover* sobre os dados de uma estação nas visualizações V1, V4 e V5, é feito um destaque em V6 da localização dessa estação.

Concluindo, através desta ferramenta é possível responder a um vasto conjunto de questões analíticas mais específicas, como por exemplo as questões levantadas na secção 4.1:

- “No geral, considera que o padrão da concentração do poluente  $PM_{2.5}$  é semelhante entre estações? Se sim, esse padrão manteve-se após o aparecimento da COVID-19?” (Q07.01). É possível responder a esta questão através da seleção do poluente  $PM_{2.5}$  no filtro F1.
- “Relativamente a quatro semanas a partir do dia 4 de abril de 2020, qual é a estação que mais diverge da sua média de concentração global de  $NO_2$ ?” (Q07.02). Para responder a esta questão basta selecionar o poluente  $NO_2$  no filtro F1, selecionar quatro semanas no parâmetro P1 e selecionar o dia 4 de abril de 2020 através do *slider* do parâmetro P2.
- “No dia 31 de março de 2020 considera que o pico horário com maior valor da concentração do poluente  $PM_{2.5}$  é idêntico entre estações?” (Q07.03). É possível responder a esta questão através da seleção do poluente  $PM_{2.5}$  no filtro F1 e da seleção de um período temporal que englobe o dia 31 de março de 2020. Este último é feito através dos parâmetros P1 e P2. Após isso, basta selecionar o dia 31 de março de 2020 na visualização V4.

#### 4.4.8 Comparação entre o comportamento dos poluentes atmosféricos medidos nas estações de qualidade do ar da cidade de Lisboa

Após terem sido criadas as duas ferramentas anteriores que analisam a qualidade do ar da cidade de Lisboa, passou-se para a criação de uma ferramenta que permite analisar, de forma simultânea, o comportamento de todos os poluentes nas diversas estações. Esta ferramenta permite então responder à questão QG08: “Existe alguma relação entre o comportamento dos diferentes poluentes atmosféricos?”, levantada na secção 4.1.

Para o desenvolvimento deste *dashboard* fez-se bastante uso da cor, nomeadamente, da cor correspondente às classes de qualidade do ar. Deste modo, o poluente  $SO_2$  não foi considerado alvo de análise pois o mesmo encontrava-se, em todo o período temporal, com valores considerados muito bons. Na figura 4.24 encontra-se uma imagem do *dashboard* desenvolvido, bem como as denominações que serão utilizadas no processo de explicação do mesmo. Para além disso, neste vídeo é possível ver uma explicação das funcionalidades presentes neste *dashboard*.

Este *dashboard* apresenta cinco visualizações, duas legendas interativas e quatro parâmetros, tal como é possível observar através da figura 4.24.

A visualização V1 mostra a média diária do índice de qualidade do ar de cada poluente nas respetivas estações. Esta informação é apresentada através de um *heat table* no qual é mapeado, para cada poluente de cada estação, a cor da classe do índice de qualidade do ar atribuída à média diária da sua concentração. Esta visualização permite ter, rapidamente,

uma visão do panorama geral e, de forma semelhante às ferramentas vistas anteriormente, é também possível escolher aqui, com auxílio dos parâmetros P1 e P2, o período temporal que se pretende analisar nas restantes visualizações. A legenda L1 apresenta o significado das cores que são mapeadas nesta visualização e também na visualização V4. De modo semelhante às duas ferramentas vistas anteriormente, a visualização V2 apresenta, em *tooltip*, uma imagem informativa sobre as classes do índice de qualidade do ar (figura 4.22). Em V3 é mostrado, no formato textual, qual é o período temporal que foi selecionado através dos parâmetros P1 e P2, sendo este o período que afeta as visualizações V4 e V5. A visualização V4 é um mapa onde é possível ver a localização de cada estação e informação acerca da média do índice de qualidade do ar do período temporal escolhido. A localização de cada estação é mapeada através de um gráfico de pizza, sendo que cada fatia é referente a um poluente. Deste modo, existem quatro fatias e a cor de cada uma é relativa à sua média do índice de qualidade do ar no período escolhido. A visualização V5 é um *scatter plot* no qual é mapeado a concentração de dois poluentes por dias e por horas, sendo que os poluentes mapeados são escolhidos através dos parâmetros P3 e P4. Cada círculo em V5 apresenta uma cor representativa da mistura dos índices de qualidade do ar de cada poluente. Nesta visualização é ainda possível ver uma linha de tendência linear que permite inferir a relação entre as duas variáveis escolhidas. Na legenda L2 é apresentado o significado das cores mapeadas em V5.

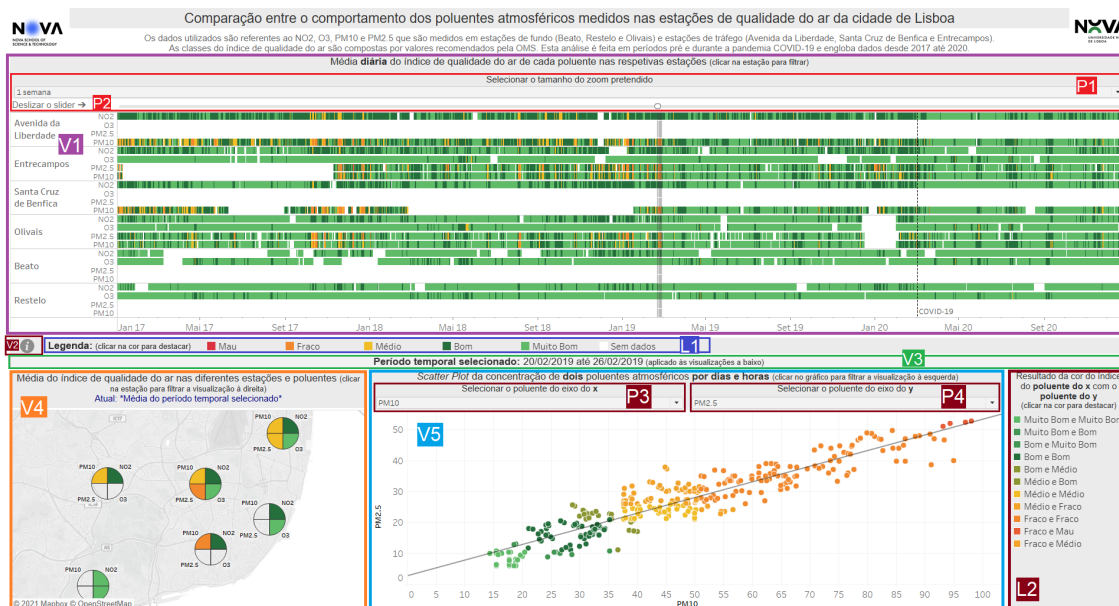


Figura 4.24: *Dashboard* para a comparação entre o comportamento dos poluentes atmosféricos medidos nas estações de qualidade do ar da cidade de Lisboa.

Com exceção da *tooltip* exibida em V2 (figura 4.22), todas as restantes *tooltips* presentes nas visualizações deste *dashboard* estão no formato textual. Estas apresentam informação que já era possível ver nas visualizações, nomeadamente, informação temporal, informação sobre o nome da estação correspondente, informação sobre qual o valor da

concentração do poluente e sobre qual o índice de qualidade do ar atribuído.

No que diz respeito às interações, este *dashboard* faz uso de ações de filtragem e ações de destaque. De forma idêntica a muitas das ferramentas vistas anteriormente, é possível filtrar os dados das visualizações V3, V4 e V5, de modo a conterem apenas o período temporal escolhido em P1 e P2. Este período temporal é destacado na visualização V1 através de bandas verticais. As visualizações V1 e V4 apresentam ações de filtragem e de destaque semelhantes, sendo que estas afetam a própria visualização V4 e a visualização V5. Ou seja, quando se faz *hover*, em V1 e V4, sobre os dados de uma estação é, de imediato, feito um destaque nos dados dessa estação presentes em V4 e V5. Porém, se se pretender que a visualização V5 apenas apresente dados referentes a uma estação específica, basta clicar, em V1 ou V4, sobre a estação escolhida. Com os parâmetros P3 e P4 é possível escolher quais os poluentes que estão a ser mapeados em V5, sendo possível escolher entre os valores “NO<sub>2</sub>”, “O<sub>3</sub>”, “PM<sub>10</sub>” e “PM<sub>2.5</sub>”. Adicionou-se ainda a possibilidade de, ao clicar sobre um elemento da visualização V5, o mapa em V4 ser filtrado para os dados referentes a esse dia e hora específica. Com isto é possível ver o comportamento de todos os poluentes, medidos nas diferentes estações, num período temporal específico. Por fim, ao clicar sobre um elemento da legenda L1, é efetuada uma ação de destaque nas visualizações V1 e V4. E, ao clicar numa cor da legenda L2, é realizada uma ação de destaque em V5.

Concluindo, com o auxílio deste *dashboard* é possível responder a um vasto conjunto de questões analíticas mais específicas, como por exemplo as questões levantadas na secção 4.1:

- **“Após o aparecimento da COVID-19 e num panorama geral, qual foi a classe predominante do índice de qualidade do ar em todos os poluentes medidos nas diversas estações?” (Q08.01).** Para responder a esta questão não é necessário realizar ações de filtragem, porém o destaque das cores presentes na legenda de cores L1 pode ajudar na resposta.
- **“Analisando as quatro primeiras semanas após o aparecimento da COVID-19, quais foram os poluentes que, em média, apresentaram um índice de qualidade do ar considerado apenas bom? E em que estações é que isso se verificou?” (Q08.02).** É possível responder a esta questão através da seleção de quatro semanas no parâmetro P1 e da seleção do dia em que apareceu a pandemia no *slider* do parâmetro P2.
- **“Analisando as quatro primeiras semanas após o aparecimento da COVID-19, considera que existe alguma correlação entre os poluentes NO<sub>2</sub> e O<sub>3</sub> medidos na estação de Entrecampos?” (Q08.03).** Para responder a esta questão basta selecionar quatro semanas no parâmetro P1 e selecionar o dia em que apareceu a pandemia COVID-19 no *slider* do parâmetro P2. Para além disso, é necessário selecionar a estação Entrecampos no mapa ou na visualização V1 e selecionar os poluentes NO<sub>2</sub> e O<sub>3</sub> nos parâmetros P3 e P4.

#### 4.4.9 Procura de relação entre os congestionamentos detetados pelo Waze e a qualidade do ar de Lisboa

Tal como foi visto através do levantamento do estado da arte, presente no capítulo 3, a melhoria dos valores observados nos parâmetros de qualidade do ar pode dever-se ao facto de ter havido um abrandamento significativo do uso dos veículos e, por consequente, um abrandamento dos congestionamentos. Para além disso, foi visto ainda que os parâmetros meteorológicos são essenciais, pois são eles que promovem a dispersão dos poluentes atmosféricos.

Deste modo, foi desenvolvida uma ferramenta que permite a procura de relação entre os congestionamentos ocorridos e a qualidade do ar da cidade de Lisboa, apresentando ainda informação acerca dos fatores climáticos. O *dashboard* criado permite responder à questão geral **QG09**: “Os congestionamentos têm impacto na concentração dos poluentes atmosféricos?”, levantada na secção 4.1. Na figura 4.25 encontra-se uma imagem do *dashboard* desenvolvido, acompanhado pelas denominações dadas aos seus elementos. Através deste [vídeo](#) é possível ver uma explicação das funcionalidades presentes neste *dashboard*.

Durante a criação deste *dashboard* sentiu-se a necessidade de dividir os valores de cada parâmetro meteorológico em classes com um significado único. Na tabela 4.16 encontram-se as classes que foram usadas para cada parâmetro meteorológico e os seus intervalos de valores, sendo que estes se basearam nas classificações feitas pelo [IPMA](#).

Tabela 4.16: Classes usadas para cada parâmetro meteorológico.

Parâmetro Meteorológico	Classes	Intervalos de Valores
Intensidade Média do Vento	Sem vento	0 km/h
	Vento fraco	1 km/h até 14 km/h
	Vento moderado	15 km/h até 35 km/h
	Vento forte	36 km/h até 55 km/h
	Vento muito forte	56 km/h até 75 km/h
	Vento excep. forte	A partir de 76 km/h
Temperatura média	Muito frio	Abaixo de 7°C
	Frio	8°C até 13°C
	Ameno	14°C até 18°C
	Agradável	19°C até 24°C
	Morno	25°C até 29°C
	Quente	30°C até 36°C
	Muito quente	A partir de 37°C
Precipitação	Sem chuva	0 mm
	Chuva fraca	0.01 mm até 0.49 mm
	Chuva moderada	0.5 mm até 4 mm
	Chuva forte	A partir de 4 mm
Humidade	Humidade baixa	Abaixo de 39%
	Humidade normal	40% até 70%
	Humidade alta	A partir de 71%

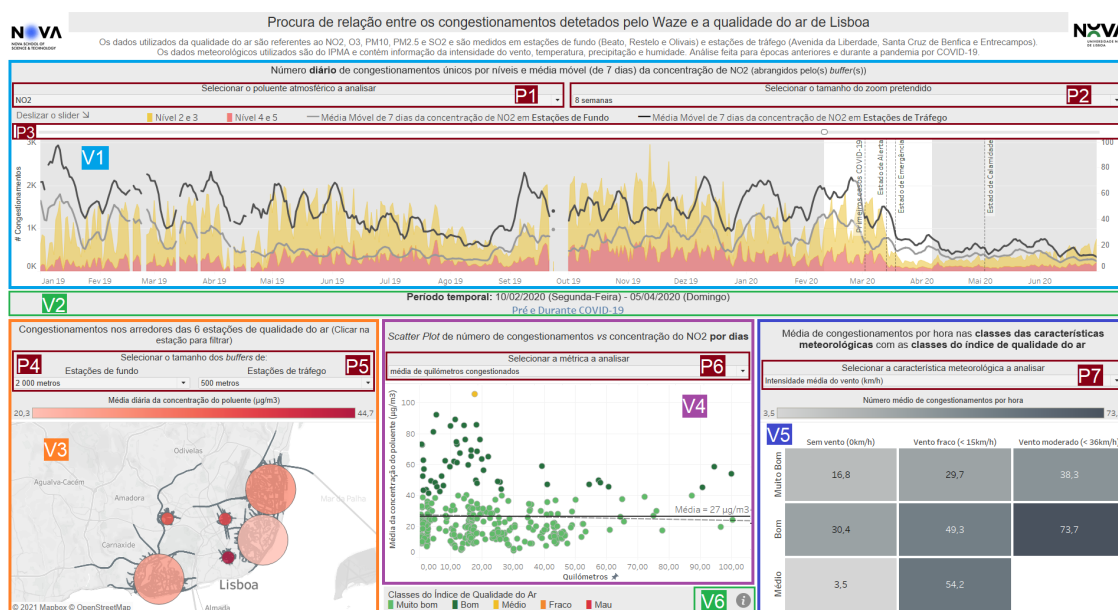


Figura 4.25: *Dashboard* para a procura de relação entre os congestionamentos detetados pelo *Waze* e a qualidade do ar de Lisboa.

Através da figura 4.25 é possível identificar que existem seis visualizações e sete parâmetros neste *dashboard*.

A visualização V1 mostra o número diário de congestionamentos únicos por níveis de intensidade e a média móvel, de 7 dias, da concentração do poluente selecionado através do parâmetro P1. Esta informação é apresentada através de um gráfico de eixo Y duplo. O eixo Y da esquerda é referente ao número diário de congestionamentos por níveis, sendo que esta informação é mapeada no gráfico através da área. Nesta área uma cor amarela denota congestionamentos dos níveis 2 e 3 e uma cor vermelha representa congestionamentos dos níveis 4 e 5. O eixo Y da direita é relativo ao valor da média móvel diária da concentração do poluente, sendo que estes dados são mapeados no gráfico através de duas linhas. A linha a cinzento claro representa os valores medidos em *estações de fundo*, enquanto que a linha a cinzento escuro mostra os valores medidos nas *estações de tráfego*. De forma semelhante às ferramentas anteriores, é possível escolher qual o período temporal a analisar em maior detalhe através dos parâmetros P2 e P3. Na visualização V2 é apresentado, no formato textual, qual o período temporal escolhido. Na visualização V3 é apresentado um mapa com duas camadas geográficas. Numa das camadas são mapeados os *buffers* das estações de qualidade do ar e na outra as geometrias dos congestionamentos que os interseccionam, sendo que o raio dos *buffers* é escolhido através dos parâmetros P4 e P5. Para além disso, é mapeado ainda em cada *buffer*, através da intensidade da cor, a média diária da concentração do poluente. Isto permite ver, rapidamente, qual foi a estação que apresentou concentrações mais elevadas no período temporal selecionado. A visualização V4 apresenta um *scatter plot* onde se encontra mapeado, para o período temporal escolhido, a concentração do poluente selecionado em P1 e uma das métricas escolhidas em

P6. No parâmetro P6 é possível escolher entre “média de quilômetros congestionados” e “número de congestionamentos”. Cada círculo deste gráfico representa um dia específico e a sua cor é relativa à classe do índice de qualidade do ar a que pertence a média da concentração do poluente. Nesta visualização é ainda possível observar uma linha de tendência linear a tracejado e uma linha com o valor médio das concentrações do poluente. A visualização V5 é um *heat table* onde é mostrado a média de congestionamentos por hora, referentes ao período temporal escolhido. Sendo que esta média é calculada tendo em conta a combinação das classes do parâmetro meteorológico (escolhido em P7) com as classes do índice de qualidade do ar. Esta visualização faz ainda uso da intensidade da cor de modo a mapear o número médio de congestionamentos por hora. Por fim, a visualização V6 mostra, em *tooltip*, uma imagem informativa dos intervalos de valores de cada classe do índice de qualidade do ar, tal como visto em ferramentas anteriores.

À exceção da *tooltip* existente em V6, todas as restantes *tooltips* presentes nas visualizações deste *dashboard* são apresentadas no formato textual. Na maioria destas é apresentada informação que já era parcialmente visível nos gráficos. Ao fazer *hover* sobre um elemento de V1 é possível observar informação temporal, informação sobre o número de congestionamentos em cada nível e informação sobre a média móvel da concentração do poluente. Cada *buffer* da visualização V3 apresenta em *tooltip* o nome e o tipo da estação a que se refere, a média diária do poluente selecionado e o número total de congestionamentos no período temporal escolhido. Em V4 é mostrado, essencialmente, informação temporal e informação sobre as métricas que estão a ser mapeadas no *scatter plot*. Em V5, para além de ser mostrada informação já visível na visualização, é exibido ainda quantas horas e congestionamentos houve, no período temporal escolhido, nas condições de meteorologia e qualidade do ar sobre as quais é feito *hover*.

Este *dashboard* apresenta, tal como todas as ferramentas vistas anteriormente, uma forte componente interativa. Através do parâmetro P1 é possível escolher qual o poluente que está a ser analisado nas visualizações V1, V3, V4 e V5, sendo que as opções de escolha são “NO<sub>2</sub>”, “O<sub>3</sub>”, “PM<sub>10</sub>”, “PM<sub>2.5</sub>” e “SO<sub>2</sub>”. À semelhança da maioria das ferramentas analisadas anteriormente, existe a possibilidade de filtrar, através dos parâmetros P2 e P3, o período temporal que está a ser analisado com maior detalhe. A alteração destes parâmetros filtra os dados das visualizações V2, V3, V4 e V5 e realiza uma ação de destaque na visualização V1. A alteração do raio dos *buffers*, nos parâmetros P4 e P5, faz com que uma ação de filtragem seja efetuada nas visualizações V1, V3, V4 e V5, de modo a contabilizar apenas os congestionamentos que estão a ser intersetados pelos *buffers*. Ao clicar, em V3, sobre o *buffer* de uma estação é também realizada uma ação de filtragem nas visualizações V1, V4 e V5, de forma a mostrar apenas os dados referentes a essa estação. A mudança do parâmetro P6 faz com que a métrica a ser analisada em V4 mude e seja, deste modo, possível explorar o comportamento da concentração do poluente com duas métricas diferentes relativas aos congestionamentos.

Concluindo, com esta ferramenta é possível responder a um vasto conjunto de questões analíticas mais específicas relacionadas, principalmente, com a correlação entre os

indicadores de tráfego e a qualidade do ar da cidade de Lisboa. De seguida encontram-se brevemente explicados alguns exemplos de questões analíticas mais específicas, nomeadamente as questões levantadas na secção 4.1:

- **“Após a declaração do estado de emergência, qual foi o comportamento geral observado no número diário de congestionamentos e na média móvel da concentração do NO<sub>2</sub>, comparativamente a períodos temporais anteriores?” (Q09.01).** Para responder a esta questão basta selecionar o poluente NO<sub>2</sub> no parâmetro P1.
- **“Analisando três semanas após a declaração do estado de emergência, qual é a estação de qualidade do ar que apresenta uma maior média diária da concentração do poluente NO<sub>2</sub>?” (Q09.02).** É possível responder a esta questão através da seleção do poluente NO<sub>2</sub> no parâmetro P1, da seleção de três semanas no parâmetro P2 e da seleção da data de início do estado de emergência no *slider* do parâmetro P3.
- **“Analisando oito semanas a partir do dia 20/02/2020, parece haver alguma correlação entre a concentração de NO<sub>2</sub> e o número de congestionamentos num raio de 500 metros à volta da estação da Avenida da Liberdade?” (Q09.03).** A resposta a esta questão é obtida através da seleção do poluente NO<sub>2</sub> no parâmetro P1, da seleção de oito semanas no parâmetro P2 e da seleção da data 20/02/2020 no *slider* do parâmetro P3. Para além disso, é necessário selecionar um raio de 500m no parâmetro P5, selecionar a estação da Avenida da Liberdade no mapa e selecionar a métrica do número de congestionamentos no parâmetro P6.

## 4.5 Conclusões e Análises Feitas nas Ferramentas Desenvolvidas

O objetivo fundamental desta dissertação passa pela criação de mecanismos que permitam, através dos dados fornecidos, analisar o impacto que a COVID-19 teve na mobilidade e na qualidade do ar da cidade de Lisboa. De modo a definir o projeto como finalizado foi necessário fazer o levantamento de questões gerais que se pretendiam ver respondidas, ou seja, foi estabelecido um critério de aceitação.

Através das ferramentas criadas foi possível responder a todas as questões gerais levantadas na secção 4.1. Para além disso, é ainda possível responder a um vasto conjunto de questões analíticas mais específicas sobre cada um dos *dashboards*. Deste modo, é então possível afirmar que o critério de aceitação foi atingido.

As ferramentas criadas oferecem todas uma forte componente interativa e, ao longo do desenvolvimento dos *dashboards*, tentou-se sempre seguir a abordagem de ter no topo das ferramentas uma visualização com uma visão geral dos dados e, por baixo dessa, várias outras visualizações que permitissem explorar o período temporal escolhido. Desta forma, a maioria dos *dashboards* tem associado a essa visualização mais geral um *slider* que permite escolher o período temporal que se pretende analisar. Através desta funcionalidade

é possível não só observar o comportamento e os padrões ocorridos nos dados antes e durante a pandemia, mas também analisar períodos que não estejam interligados com a COVID-19. Isto acaba por dar uma maior versatilidade às ferramentas criadas, sendo possível, com a maioria delas, responder a diversas questões fora do tema da pandemia.

Na tabela 4.17 é possível observar um resumo de algumas das interações gerais que estão presentes em cada um dos *dashboards* explicados na secção 4.4. Neste levantamento foram identificadas as interações que são consideradas mais importantes no âmbito deste trabalho. A última linha da tabela nomeada “Filtragem/Destaque por outros parâmetros” é relativa a cenários de, por exemplo, filtragem por níveis de congestionamento, filtragem por tamanho do parque de estacionamento, destaque de um POI no mapa, etc. As denominações dadas aos *dashboards* presentes nas colunas da tabela são as seguintes:

- **D01:** Giras - Estados da pandemia (subsecção 4.4.1)
- **D02:** Giras - Datas específicas (subsecção 4.4.1)
- **D03:** Parques de estacionamento (subsecção 4.4.2)
- **D04:** Congestionamentos - Geral (subsecção 4.4.3)
- **D05:** Congestionamentos - Todas as zonas (subsecção 4.4.4)
- **D06:** Congestionamentos - Zona específica (subsecção 4.4.4)
- **D07:** Congestionamentos - POI (subsecção 4.4.5)
- **D08:** Qualidade do ar (subsecção 4.4.6)
- **D09:** Qualidade do ar - Por estações (subsecção 4.4.7)
- **D10:** Comparação entre poluentes (subsecção 4.4.8)
- **D11:** Congestionamentos vs qualidade do ar (subsecção 4.4.9)

Tabela 4.17: Algumas das interações mais gerais presentes em cada *dashboard*.

Interações	Dashboards										
	D01	D02	D03	D04	D05	D06	D07	D08	D09	D10	D11
Filtragem temporal com o slider				✓		✓	✓		✓	✓	✓
Filtragem temporal por datas/horas específicas	✓	✓	✓	✓				✓	✓	✓	
Filtragem temporal por estados da pandemia	✓		✓					✓			
Filtragem espacial no mapa	✓	✓	✓				✓			✓	✓
Filtragem por zonas	✓	✓			✓	✓					
Filtragem por poluentes atmosféricos								✓	✓	✓	✓
Filtragem/Destaque por outros parâmetros	✓	✓	✓	✓	✓		✓	✓		✓	✓

## 4.5. CONCLUSÕES E ANÁLISES FEITAS NAS FERRAMENTAS DESENVOLVIDAS

No decorrer da dissertação e, muitas das vezes, durante os momentos de reunião com os *stakeholders* da CML, foram realizadas análises através dos *dashboards* detalhados em 4.4. Nas seguintes subsecções, que estão divididas por temáticas, serão apresentadas algumas dessas análises feitas em algumas ferramentas, de modo a ser possível, não só ilustrar o funcionamento das ferramentas criadas, mas também evidenciar alguns exemplos de descobertas feitas com as mesmas. Sendo que existe um vasto conjunto de análises que podem ser feitas nestes *dashboards*, dependendo sempre das questões analíticas que sejam formuladas.

### 4.5.1 Bicicletas Gira

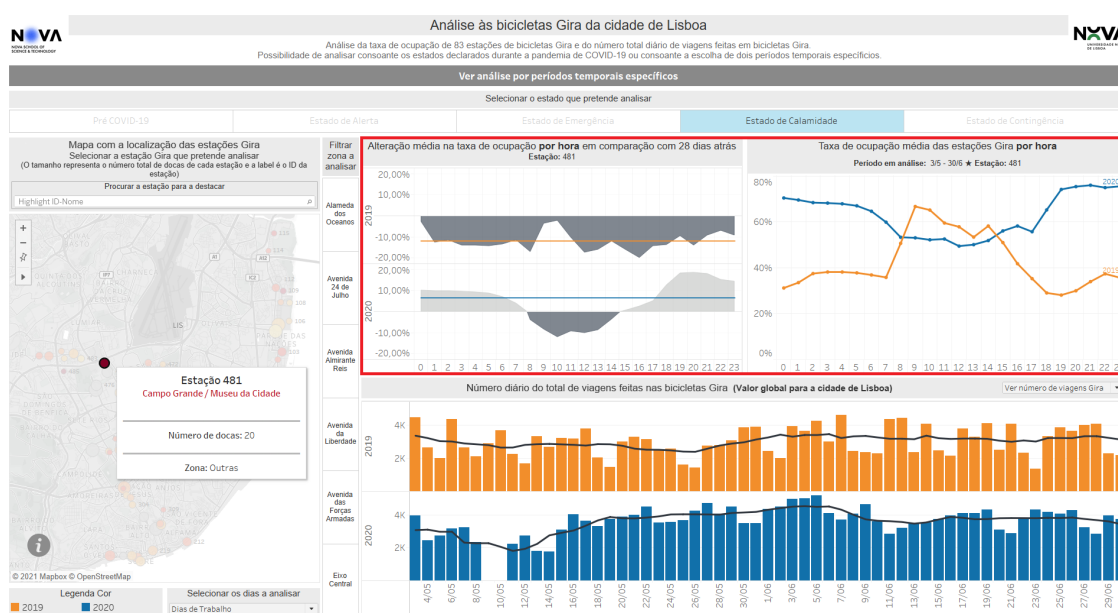


Figura 4.26: Exemplo de análise feita no *dashboard* para a análise às bicicletas Gira da cidade de Lisboa.

Na figura 4.26 é possível observar uma análise feita através do *dashboard* para a análise às bicicletas Gira da cidade de Lisboa. Através desta é possível constatar que existiram estações Gira que durante o estado de calamidade sofreram uma mudança no seu comportamento. A estação 481 é um exemplo onde ocorreu uma grande mudança. Em 2019, durante a altura do estado de calamidade, esta estação comportava-se, maioritariamente, como estação de chegada de bicicletas durante as horas de trabalho, ou seja, os valores de taxa de ocupação subiam durante estas horas. Em 2020, durante a mesma altura, esta estação apresentou um comportamento inverso, isto é, comportou-se maioritariamente como estação de partida de bicicletas. Também é possível constatar que, nesta estação, em 2019 existia uma tendência a haver menos bicicletas ao longo das horas do dia comparativamente aos 28 dias atrás. E que, em 2020, apenas no período temporal entre as 8h e as 14h é que houve uma redução da taxa de ocupação média comparativamente aos 28 dias atrás. Ou seja, um dos exemplos de análises que é possível fazer com esta ferramenta é,

não só analisar o comportamento que uma estação apresentava nos 28 dias atrás do estado escolhido, mas também verificar quais as estações que apresentaram maiores mudanças no seu comportamento comparativamente à mesma altura de 2019.

#### 4.5.2 Parques de estacionamento

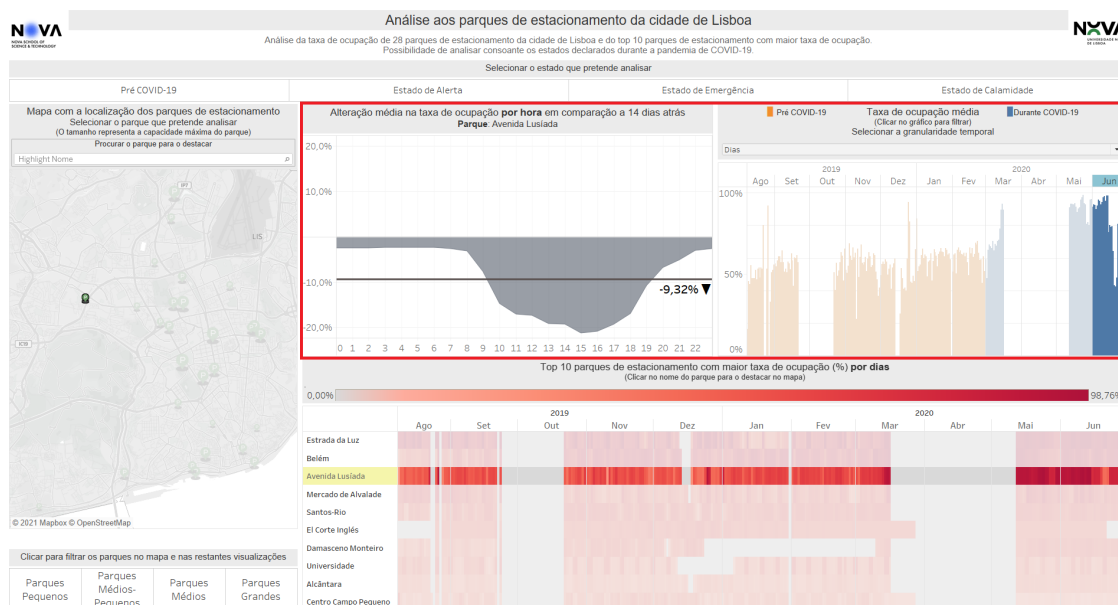


Figura 4.27: Exemplo de análise feita no *dashboard* para a análise aos parques de estacionamento da cidade de Lisboa.

Relativamente aos parques de estacionamento foi feita também uma análise interessante, visível na figura 4.27. Nesta é possível constatar que, à semelhança das estações Gira, houve também parques de estacionamento com alterações no seu comportamento durante a pandemia. Um destes parques foi o da Avenida Lusitana que, durante a altura da pandemia, ficou com valores diários de taxa de ocupação mais altos, ou seja, estavam estacionados lá mais carros. Isto pode ser um indicio de que este parque de estacionamento seja, maioritariamente, um parque residencial, pois durante a pandemia foram implementadas medidas de confinamento, levando as pessoas a sair menos da sua área residencial. Todavia, no final de junho começa-se a ver uma diminuição da taxa de ocupação diária e, analisando esse mês, é visto que a tendência dessa descida se deve, na sua maioria, às horas de trabalho. Isto, pois quando comparamos cada dia do mês de junho com os 14 dias atrás, vê-se que, em média, há uma descida dos valores de taxa de ocupação média. Esta descoberta vem a reforçar, ainda mais, a afirmação anterior deste parque ser residencial. Este exemplo de análise demonstra que é possível verificar o comportamento que cada parque de estacionamento apresentou em alturas anteriores e durante a pandemia, sendo possível verificar que os parques que são maioritariamente residenciais ficaram mais cheiros e os parques que não são residenciais ficaram mais vazios.

## 4.5. CONCLUSÕES E ANÁLISES FEITAS NAS FERRAMENTAS DESENVOLVIDAS

### 4.5.3 Congestionamentos

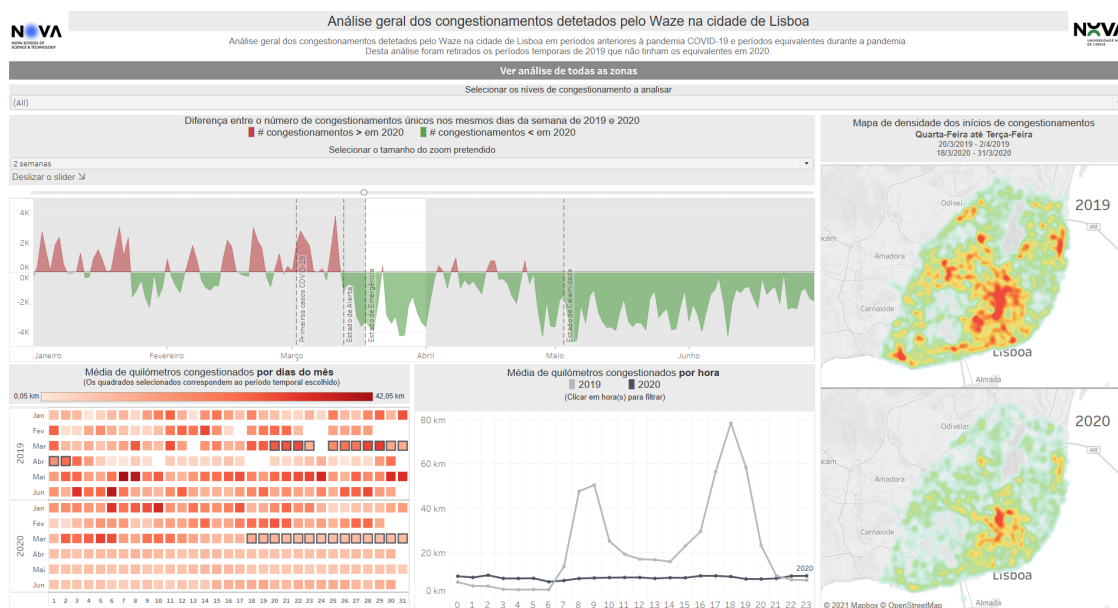


Figura 4.28: Exemplo de análise feita no *dashboard* para a análise geral aos congestionamentos detetados pelo *Waze* na cidade de Lisboa.

Na figura 4.28 é possível observar uma análise feita com o auxílio do *dashboard* para a análise geral aos congestionamentos detetados pelo *Waze* na cidade de Lisboa. Através desta é possível ver rapidamente que, após a declaração do estado de alerta, o número de congestionamentos em 2020 teve, na sua maioria, sempre mais baixo do que em 2019. Para além disso, analisando mais ao detalhe as duas primeiras semanas após a declaração do estado de emergência, é notório que a densidade espacial dos congestionamentos mudou comparativamente a 2019, havendo em 2020 menos concentrações de congestionamentos no período temporal escolhido. Nestas duas semanas observa-se ainda que a média de quilómetros congestionados por hora em 2020 é basicamente sempre constante, enquanto que em 2019 existiam picos nas horas de maior volume de trânsito. Por último, é ainda possível ter uma visão geral de como se comportou a média de quilómetros congestionados por dia do mês. Com esta é possível aferir que, a partir de meados de março de 2020, a média de quilómetros congestionados esteve, maioritariamente, sempre mais baixa do que nos dias anteriores. Esta ferramenta permite efetuar diversas análises de grande relevância que mostrem como esteve o estado do trânsito geral na cidade de Lisboa. E, visto que é possível escolher um período temporal para analisar em maior detalhe, podem-se estabelecer diferentes cenários onde se consegue observar e comparar o comportamento dos dois anos em questão.

Foram desenvolvidas ferramentas que permitissem realizar análises mais detalhadas em algumas zonas da cidade de Lisboa. Na figura 4.29 encontra-se um exemplo de uma análise realizada num dos *dashboards* desenvolvidos para essa finalidade. Nesta é possível

verificar que as duas avenidas onde parece ter ocorrido uma maior mudança no número de congestionamentos ao seu redor são o Eixo Central, que se localiza desde a Praça do Marquês de Pombal até ao Campo Grande, e na Avenida da Liberdade. Em ambas, o ano de 2020 apresenta, na grande maioria das vezes, menos congestionamentos do que no ano 2019. Sendo também possível observar que, comparativamente às restantes zonas, o Eixo Central é o que se destaca mais na sua mudança de comportamento.

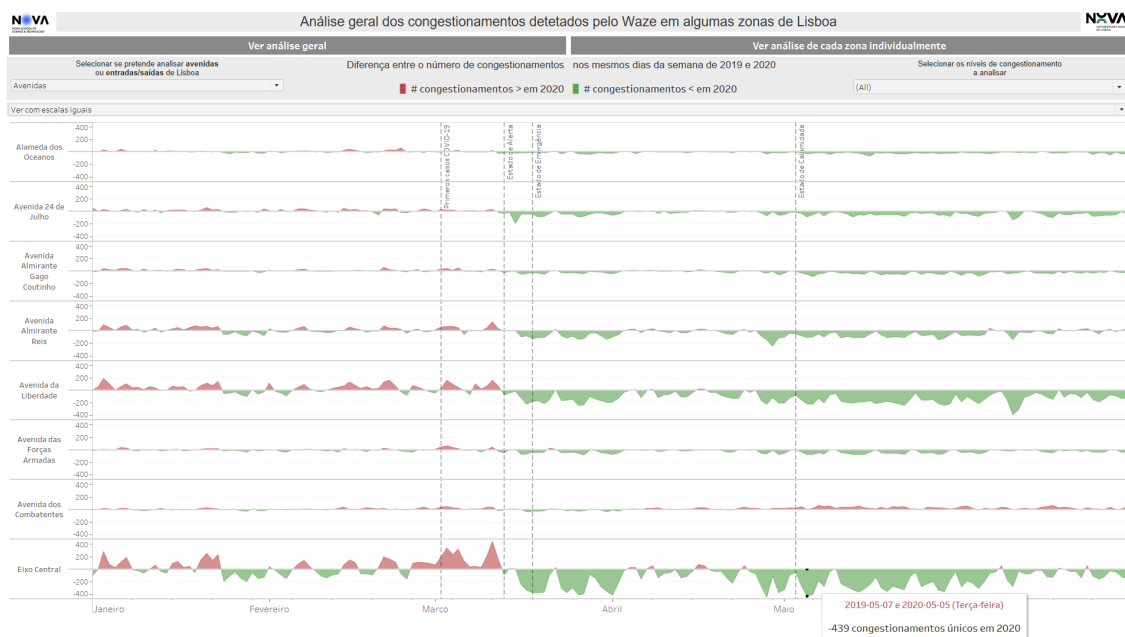


Figura 4.29: Exemplo de análise feita no *dashboard* para a análise geral aos congestionamentos detetados pelo *Waze* em algumas zonas da cidade de Lisboa.

Na figura 4.30 encontra-se uma análise realizada no *dashboard* feito para a análise dos congestionamentos numa zona específica de Lisboa. Neste é possível explorar cada zona isoladamente, sendo que neste caso a zona selecionada é o Eixo Central. É possível verificar que, a partir da declaração do estado de alerta, houve um grande decréscimo no número de congestionamentos diários de todos os níveis de intensidade ocorridos no Eixo Central e nas suas proximidades. Para além disso, e analisando mais detalhadamente as duas primeiras semanas após a declaração do estado de emergência, é possível constatar que em 2019 cada congestionamento tinha, em média, 275.5 metros, enquanto que em 2020 este valor reduziu para 98.2 metros. Em 2019, durante estas duas semanas houve, em média, 331 congestionamentos diários e em 2020 este valor desceu para uma média de 47 congestionamentos diários. Ou seja, um dos exemplos de análises que é possível realizar com esta ferramenta é a exploração mais detalhada do comportamento que cada zona de Lisboa apresentou, podendo também analisar mais detalhadamente um período temporal à escolha.

## 4.5. CONCLUSÕES E ANÁLISES FEITAS NAS FERRAMENTAS DESENVOLVIDAS

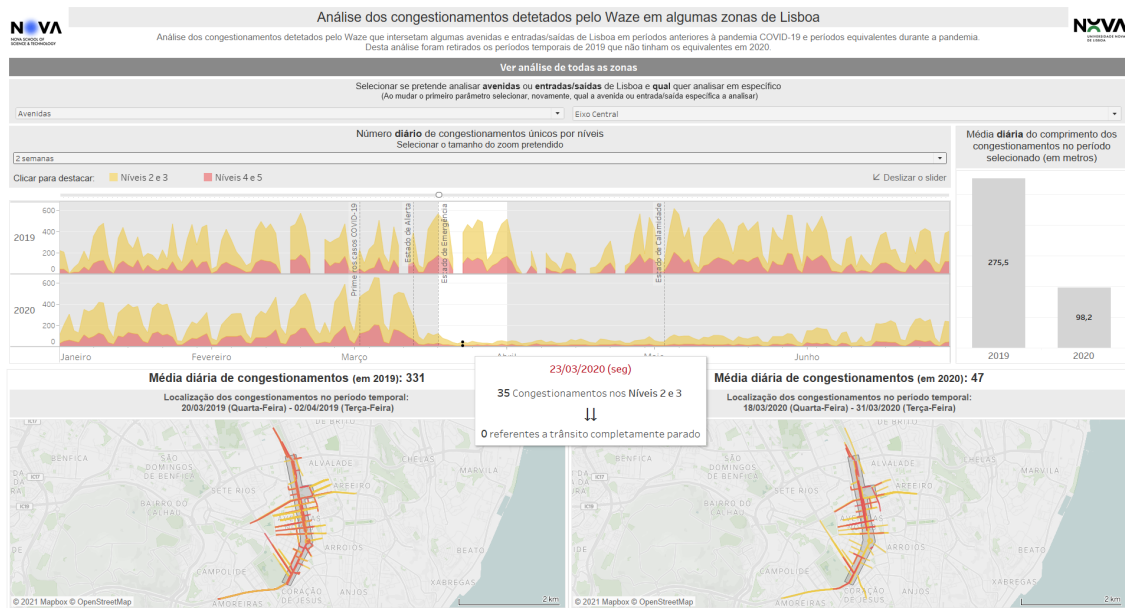


Figura 4.30: Exemplo de análise feita no *dashboard* para a análise dos congestionamentos detetados pelo Waze numa zona específica da cidade de Lisboa.

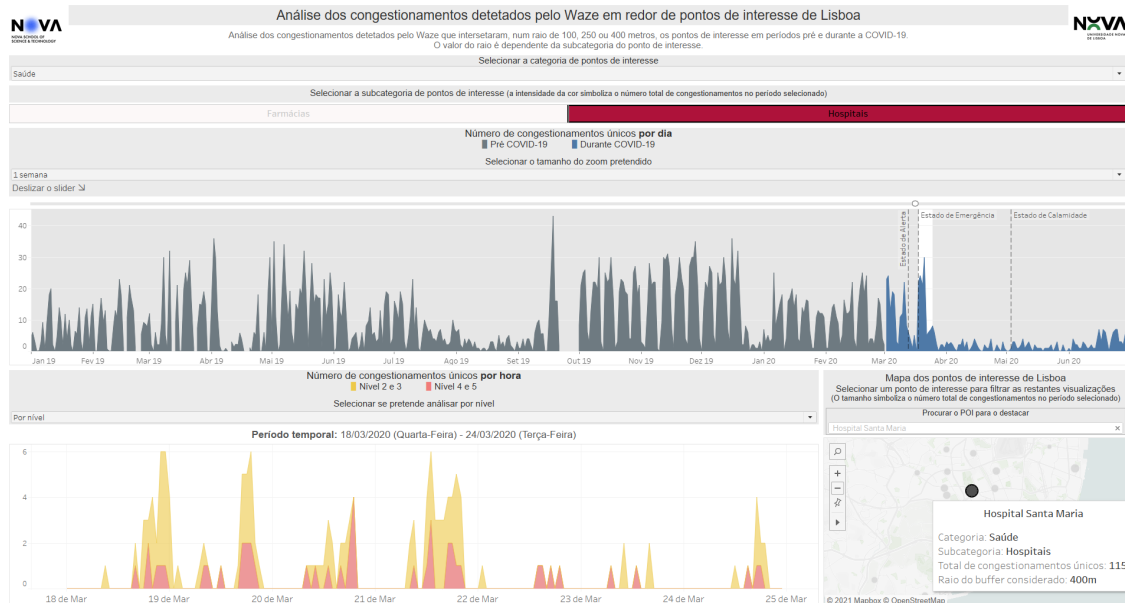


Figura 4.31: Exemplo de análise feita no *dashboard* para a análise dos congestionamentos detetados pelo Waze ao redor de pontos de interesse da cidade de Lisboa.

Ainda relativamente aos congestionamentos, foi desenvolvida uma ferramenta que permite analisar o tráfego ao redor de pontos de interesse de Lisboa. Sendo que os hospitais tiveram um papel muito importante durante a pandemia, e sendo o Hospital Santa Maria um dos maiores hospitais de Lisboa, decidiu-se analisar o comportamento dos congestionamentos ao seu redor. Através da figura 4.31 é possível ver a análise realizada

nesse *dashboard*. É observado que, nos primeiros quatro dias após a declaração do estado de emergência, houve um aumento no número de congestionamentos ao seu redor. Este revelou-se um comportamento inverso ao que se observa em muitos dos restantes POI pois, por norma, os congestionamentos ao redor dos mesmos reduzem após a declaração do estado de alerta. Analisando a primeira semana do estado de emergência mais ao detalhe, é possível constatar que existiram, nessa semana, 115 congestionamentos ao redor do Hospital Santa Maria. A maioria desses congestionamentos eram do nível 2 e 3, porém também existiram alguns congestionamentos do nível 4 e 5. Esta análise evidência o género de descobertas que podem ser feitas recorrendo a esta ferramenta, sendo que é possível analisar cada ponto de interesse isoladamente ou analisar os conjuntos de pontos de interesses pertencentes a cada categoria e/ou subcategoria.

#### 4.5.4 Qualidade do ar

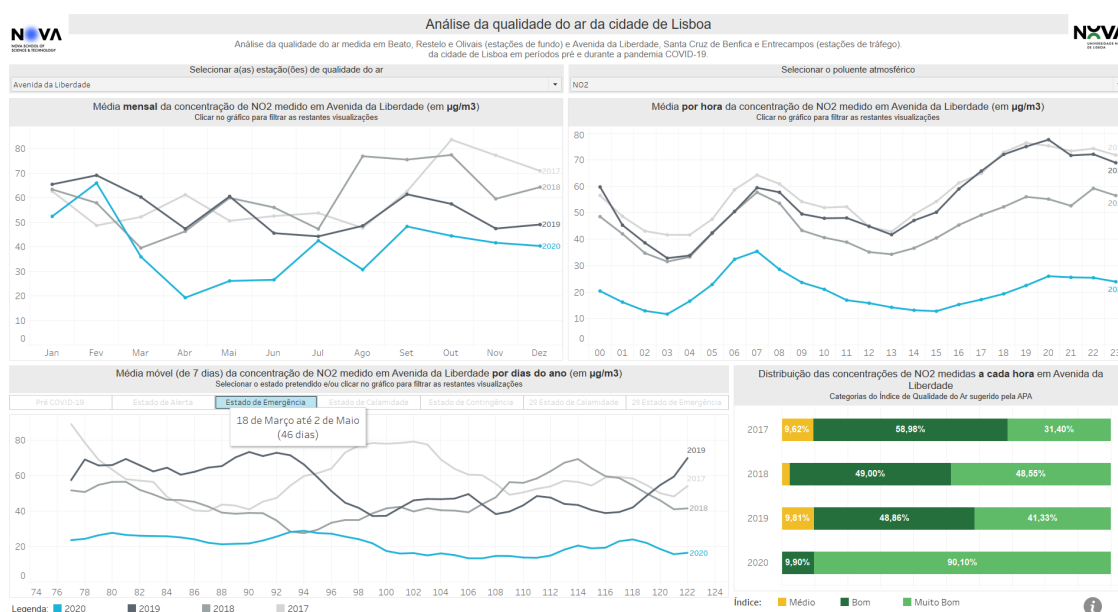


Figura 4.32: Exemplo de análise feita no *dashboard* para a análise da qualidade do ar da cidade de Lisboa.

Na figura 4.32 é possível observar uma análise feita no *dashboard* da qualidade do ar da cidade de Lisboa. Este exemplo de análise foi feito para a estação da Avenida da Liberdade pois esta é a estação que apresenta, por norma, medições de poluição mais elevadas. O poluente atmosférico escolhido para esta análise é o NO<sub>2</sub> devido ao tráfego rodoviário ser a principal fonte de emissão deste. É rapidamente perceptível que, no ano de 2020, houve uma descida significativa dos valores deste poluente a partir da altura que surgiu a COVID-19 em Portugal, sendo que esta descida é principalmente notória nos meses de abril, maio e junho. Dado que durante o estado de emergência foram implementadas medidas de confinamento, foi decidido analisar esse mês mais ao detalhe. Tal como é

## 4.5. CONCLUSÕES E ANÁLISES FEITAS NAS FERRAMENTAS DESENVOLVIDAS

possível constatar, durante o estado de emergência as concentrações deste poluente a nível horário também reduziram significativamente quando comparadas com os três anos anteriores. É visto ainda que a média móvel da concentração diária deste poluente também se encontra, a maioria das vezes, abaixo dos valores dos três anos anteriores. Por último, é possível ainda verificar que em 2020, durante a altura do estado de emergência, 90.10% das horas estiveram com concentrações de  $\text{NO}_2$  consideradas muito boas. Este é um valor surpreendente pois nos três anos anteriores nunca houve mais de 50% das horas com concentrações consideradas muito boas. Para além disso, em 2020 não foram registadas horas com concentrações em que o sua classe do índice de qualidade do ar fosse “Médio”, mas nos três anos anteriores foram. Este exemplo de análise exemplifica algumas das descobertas que podem ser feitas recorrendo a esta ferramenta.

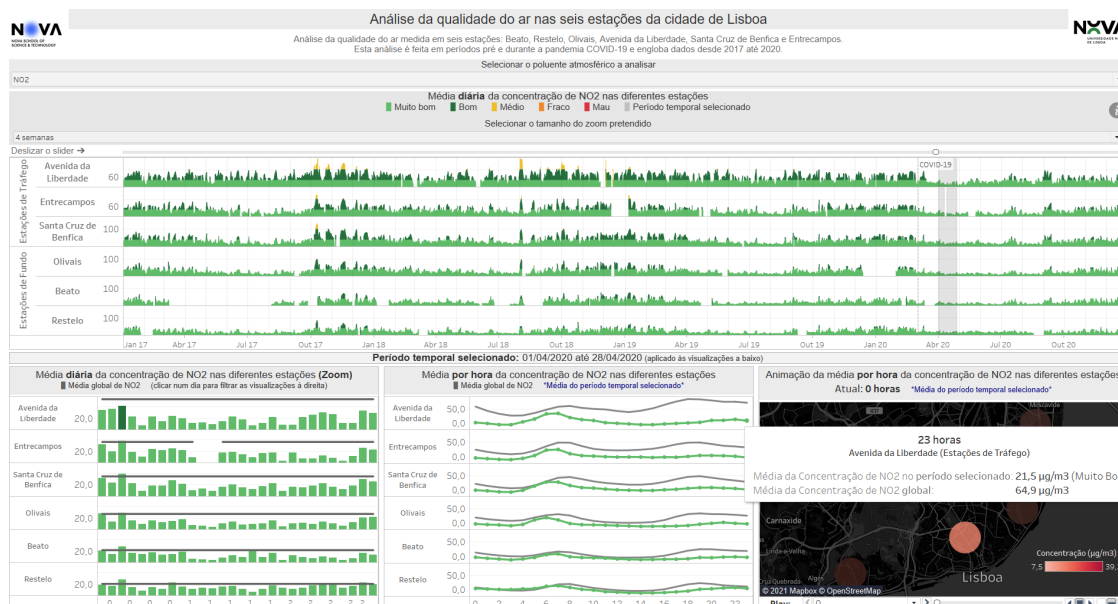


Figura 4.33: Exemplo de análise feita no *dashboard* para a análise da qualidade do ar nas seis estações da cidade de Lisboa.

A figura 4.33 mostra uma análise na qual é possível observar o comportamento nos valores medidos do poluente  $\text{NO}_2$  em todas as estações ao mesmo tempo. Com esta é possível detetar que parece haver um padrão na concentração deste poluente ao longo das estações, isto é, quando existe um pico do seu valor numa estação, por norma, as restantes estações também registam um pico, principalmente as estações da mesma categoria (i.e: tráfego ou fundo). Analisando mais ao detalhe as primeiras quatro semanas do mês de abril de 2020, é visto que basicamente todos os dias apresentam valores de concentração deste poluente considerados muito bons. Os valores médios por hora, durante estas quatro semanas, encontram-se também todos na categoria muito boa do índice de qualidade do ar. Para além disso, é ainda possível constatar que a estação que aparenta divergir mais da sua média global de concentração de  $\text{NO}_2$  é a estação da Avenida da Liberdade, tanto a nível diário, como a nível horário. Isto comprova que, tal como foi visto no capítulo 3,

as melhores melhorias na qualidade do ar acontecem em zonas que tinham níveis mais altos de concentrações dos poluentes antes da chegada da pandemia. Esta análise é um exemplo do que pode ser descoberto com esta ferramenta, sendo que é possível escolher entre os vários poluentes atmosféricos medidos nas estações de Lisboa.

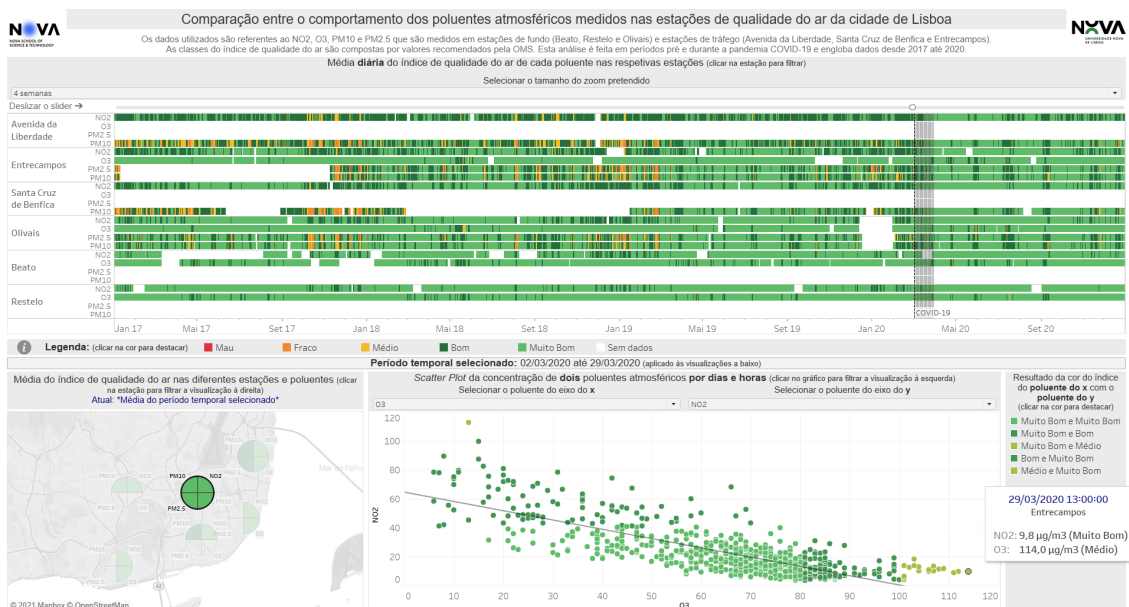


Figura 4.34: Exemplo de análise feita no *dashboard* para a comparação entre o comportamento dos poluentes atmosféricos medidos nas estações de qualidade do ar da cidade de Lisboa.

Através da figura 4.34 é possível verificar que, após a chegada da pandemia a Portugal, a média diária da concentração de todos os poluentes medidos em cada estação apresenta-se, maioritariamente, na classe “Muito Bom” do índice de qualidade do ar. Para além disso, após o aparecimento da COVID-19 não existe nenhum dia em que a média diária de um poluente seja considerada fraca. Analisando as quatro primeiras semanas após a chegada da pandemia, é possível constatar que a média da concentração de todos os poluentes em cada estação se encontra maioritariamente na classe “Muito bom”. Durante o levantamento do estado da arte, no capítulo 3, foi visto que vários autores relataram um aumento nas concentrações de O<sub>3</sub> quando as concentrações de NO<sub>2</sub> baixavam, indicando então um fenómeno de correlação negativa entre estes poluentes. Deste modo, através do *scatter plot*, podemos examinar que, na estação de Entrecampos, existe alguma correlação negativa entre os poluentes O<sub>3</sub> e NO<sub>2</sub> durante as quatro primeiras semanas após o aparecimento da pandemia. Ou seja, por norma quando os valores de NO<sub>2</sub> baixam é notado uma subida na concentração do poluente O<sub>3</sub>. Este é um exemplo de análise que é possível efetuar nesta ferramenta, sendo que a mesma apresenta a versatilidade para se poder escolher qualquer período temporal e, também, escolher qual o par de poluentes que se pretende analisar o seu comportamento em conjunto.

## 4.5. CONCLUSÕES E ANÁLISES FEITAS NAS FERRAMENTAS DESENVOLVIDAS

### 4.5.5 Congestionamentos vs qualidade do ar

Através da figura 4.35 é possível observar uma análise feita aos congestionamentos e à qualidade do ar da cidade de Lisboa. Esta análise focou-se na estação da Avenida da Liberdade, mais concretamente nos 500 metros que circundam esta estação e, também, nas medições feitas ao poluente  $\text{NO}_2$ . Após a declaração do estado de alerta é visto uma grande redução dos congestionamentos que intersejam o *buffer* desta estação, sendo que a concentração do poluente  $\text{NO}_2$  também sofreu uma descida. É visto que, nas dez semanas após o dia 10 de fevereiro de 2020, houve um total de 6 194 congestionamentos nas redondezas desta estação e que a média diária do poluente  $\text{NO}_2$  se encontra com o valor  $38.8 \mu\text{g}/\text{m}^3$ . Correlacionando, através do *scatter plot*, o número de congestionamentos à volta da estação da Avenida da Liberdade com os valores médios diários do poluente  $\text{NO}_2$ , é observado que existe alguma correlação positiva entre estas duas métricas, sendo que o valor do R-quadrado é 0.584.

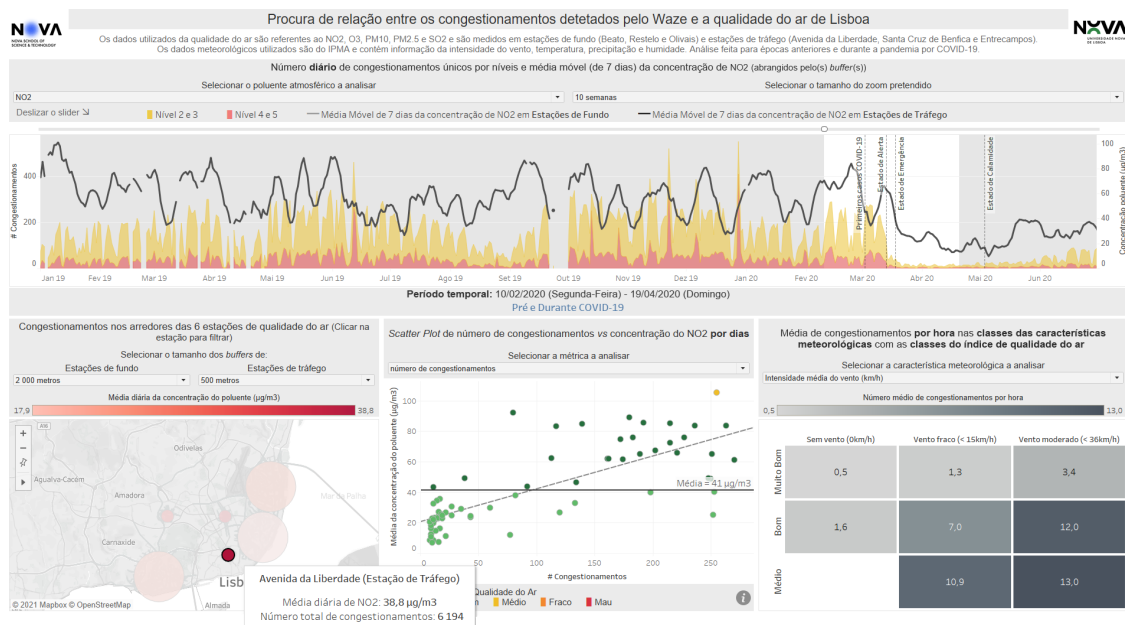


Figura 4.35: Exemplo de análise feita no *dashboard* para a procura de relação entre os congestionamentos detetados pelo *Waze* e a qualidade do ar de Lisboa.

Considerando as condições meteorológicas nestas dez semanas, mais propriamente a intensidade média do vento, é feita uma descoberta interessante. Sendo o tráfego rodoviário um dos fatores que contribuem para os valores elevados de  $\text{NO}_2$ , é espetável que quanto maior for a média de congestionamentos por hora pior fique a qualidade do ar. Contudo, em períodos de maiores velocidades do vento, consegue-se ver que é possível existir uma maior média de congestionamentos por hora em classes melhores do índice de qualidade do ar, comparativamente a dias sem vento que têm uma menor média de congestionamentos por hora. Neste período temporal, quando há vento moderado e existe uma média de 3.4 congestionamentos por hora é observado que a qualidade do ar ainda

era considerada muito boa. Contrariamente, quando não existiu vento bastou haver uma média de 1.6 congestionamentos por hora para que a qualidade do ar fosse classificada como boa. Esta descoberta pode revelar um indicio de que a velocidade do vento tem um impacto positivo na qualidade do ar.

## AVALIAÇÃO DO TRABALHO REALIZADO

Neste capítulo é apresentado o processo de avaliação experimental das ferramentas desenvolvidas. Na secção 5.1 é descrita a metodologia de avaliação utilizada durante todo o processo e na secção 5.2 são descritas algumas das características dos participantes. Na secção 5.3 são mostrados os resultados obtidos durante os testes experimentais e na secção 5.4 são discutidas algumas das conclusões retiradas e mostrados alguns aspetos melhorados nos *dashboards*.

### 5.1 Metodologia de Avaliação

A avaliação experimental das ferramentas criadas foi um processo iterativo, pois durante todo o tempo de desenvolvimento da dissertação foram realizadas várias reuniões com a CML e com os *stakeholders* interessados. Estas reuniões serviram, maioritariamente, para mostrar o trabalho desenvolvido até à data e receber sugestões por parte dos interessados. Através destas foi também possível ter uma ideia de quais seriam as questões mais relevantes e importantes a serem respondidas com a finalização do trabalho.

Quando todas as ferramentas já se encontravam desenvolvidas foi necessário realizar testes experimentais com utilizadores, de modo a avaliar se os *dashboards* estavam bem concebidos e se eram eficazes. A realização destes testes foi dividida em três fases, nomeadamente:

- Testes preliminares com dois utilizadores não entendidos na temática;
- Testes experimentais com utilizadores não entendidos na temática;
- Testes experimentais com os *stakeholders* entendidos na temática.

Na primeira fase dos testes, ou seja nos testes preliminares, foram apresentadas as ferramentas desenvolvidas a dois utilizadores, sendo que nenhum deles era entendido na temática do contexto desta dissertação, e apenas um tinha experiência com visualização de dados. Esta etapa preliminar foi crucial para receber *feedback* antes de se começar realmente a realizar os testes, sendo que dela surgiram sugestões e pequenas melhorias que foram feitas de modo a melhorar o sucesso da avaliação experimental.

As duas últimas fases da avaliação, ou seja, os testes experimentais com utilizadores entendidos e não entendidos na temática, basearam-se na resposta a um questionário para o qual era necessário consultar os *dashboards* de modo a encontrar as respostas corretas às questões. Para estes foi necessário definir, à priori, quais seriam as métricas alvo de análise durante o processo de avaliação. Deste modo, a principal métrica em análise foi a taxa de sucesso de resposta correta, por parte do utilizador, à questão colocada. Isto permite, não só avaliar se as visualizações presentes nas ferramentas são adequadas, mas também avaliar a taxa de sucesso das visualizações para responder às questões levantadas. O tempo que o utilizador demorou a responder à questão foi também uma métrica alvo de análise. Sendo que a recolha deste tempo destina-se, maioritariamente, para verificar se a ferramenta é explícita o suficiente de modo a que o utilizador consiga realizar rapidamente todos os passos necessários para responder à pergunta. Para além destas duas métricas, recolheu-se ainda o número total de ações de interação que os participantes efetuaram nos *dashboards* para responder a cada questão. Esta última métrica permite averiguar o comportamento que cada participante teve na interação com os *dashboards*.

Após estarem definidas as métricas que serão alvo de análise, foi necessário estabelecer quais as ferramentas submetidas para avaliação e quais as questões analíticas utilizadas em cada uma. Visto existirem, no total, onze *dashboards* foi fundamental realizar um processo de escolha destes. Foi decidido que seriam submetidos para análise dez *dashboards*, sendo que o único que não foi submetido foi o *dashboard* que permite analisar as bicicletas Gira da cidade de Lisboa por períodos temporais específicos. Isto, pois este é basicamente idêntico ao *dashboard* que analisa as bicicletas Gira por estados declarados na pandemia.

Relativamente às questões estabelecidas, foi decidido utilizar as questões analíticas mais específicas levantadas na secção 4.1. Contudo, visto que se pretendia que os testes experimentais não durassem muito mais do que uma hora a ser realizados, houve a necessidade de fazer uma seleção das questões que seriam mais interessantes considerar. Estimando uma média de dois minutos por questão e um total de vinte minutos para fazer breves apresentações e contextualizações iniciais a cada *dashboard*, estabeleceu-se que seriam feitas um total de vinte questões analíticas.

As vinte questões escolhidas encontram-se divididas em dez grupos no questionário, sendo que a maioria das questões é de natureza temporal pois, devido ao contexto desta dissertação, a grande parte das ferramentas desenvolvidas apresentam funcionalidades para escolher o período temporal em análise. Para além disso, existem também perguntas de natureza espacial, isto é, que incluem a seleção de algo no mapa ou então que podem ser respondidas através da observação de um mapa. De seguida é apresentada a lista de questões escolhidas por cada grupo, sendo que no apêndice B está presente uma listagem mais complexa onde são apresentadas as questões, as respostas possíveis e a resposta correta:

- **Bicicletas Gira (estados da pandemia):**
  - Q01.01. A taxa de ocupação ao longo das horas, na estação 103, é semelhante em 2019 e 2020 durante a altura do estado de emergência?
  - Q01.02. Relativamente ao estado de contingência, considera que, durante os dias de trabalho, houve alteração da hora com maior taxa de ocupação nas estações pertencentes à zona do eixo central?
  - Q01.03. Durante os estados da pandemia houve alguma alteração significativa no número diário de viagens feitas nas bicicletas da rede Gira?
- **Parques de estacionamento:**
  - Q02.01. O parque de estacionamento da Avenida Lusíada ficou, na sua maioria, mais cheio ou mais vazio após o aparecimento da COVID-19?
  - Q02.02. Relativamente à alteração média da taxa de ocupação por hora comparativamente com 14 dias atrás, qual é o comportamento observado no parque Avenida Lusíada durante o mês de junho?
- **Congestionamentos (geral)**
  - Q03.01. No geral, houve mais ou menos congestionamentos em 2020 na altura da COVID-19?
  - Q03.02. Comparativamente a 2019, considera que a densidade espacial dos congestionamentos na cidade de Lisboa se alterou durante as duas primeiras semanas do estado de calamidade?
  - Q03.03. Em 2020, durante as duas primeiras semanas do estado de emergência existiram picos horários da média de quilómetros congestionados?
- **Congestionamentos (todas as zonas):**
  - Q04.01. Quais foram as três avenidas de Lisboa onde se sentiu um maior impacto da COVID-19 em relação ao número de congestionamentos?
- **Congestionamentos (zona específica):**
  - Q04.02. Existiram mudanças significativas no número de congestionamentos diários ocorridos na Avenida da Liberdade após a declaração do estado de alerta?
- **Congestionamentos (POI):**
  - Q05.01. Como se comportaram os congestionamentos em redor dos pontos de interesse da categoria “Saúde” na cidade de Lisboa após o aparecimento da COVID-19?
  - Q05.03. Durante a primeira semana do estado de emergência, qual foi o dia que apresentou um maior número de congestionamentos ao redor do Hospital Santa Maria?

- **Qualidade do ar:**
  - **Q06.01.** Quais foram os três meses que, na estação da Avenida da Liberdade, apresentaram uma maior mudança no valor médio do poluente  $\text{NO}_2$  durante o ano de 2020, comparativamente aos anos anteriores?
  - **Q06.02.** Para o conjunto de estações da Avenida da Liberdade e Entrecampos, qual foi o estado da pandemia que parece ter tido um maior impacto nos valores médios diários de  $\text{NO}_2$  em 2020?
  - **Q06.03.** Relativamente à estação de Entrecampos e durante o 1º estado de emergência, considera que a distribuição das concentrações de  $\text{NO}_2$  pelas classes do índice de qualidade do ar foram idênticas nos quatro anos?
- **Qualidade do ar (por estações):**
  - **Q07.02.** Relativamente a quatro semanas a partir do dia 4 de abril de 2020, qual é a estação que mais diverge da sua média de concentração global de  $\text{NO}_2$ ?
- **Comparação entre poluentes:**
  - **Q08.01.** Após o aparecimento da COVID-19 e numa perspetiva diária, qual foi a classe predominante do índice de qualidade do ar em todos os poluentes medidos nas diversas estações?
  - **Q08.02.** Analisando as quatro primeiras semanas após o aparecimento da COVID-19, quais foram os poluentes que, em média, apresentaram um índice de qualidade do ar menos bom? E em que estações é que isso se verificou?
  - **Q08.03.** Analisando as quatro primeiras semanas após o aparecimento da COVID-19, considera que existe alguma correlação entre os poluentes  $\text{NO}_2$  e  $\text{O}_3$  medidos na estação de Entrecampos?
- **Congestionamentos vs qualidade do ar:**
  - **Q09.01.** Após a declaração do estado de emergência, qual foi o comportamento geral observado no número diário de congestionamentos e na média móvel da concentração do  $\text{NO}_2$ , comparativamente a períodos temporais anteriores?

Na figura 5.1 é possível observar um resumo de quais as questões específicas utilizadas durante os testes experimentais. É ainda possível observar quais os *dashboards* utilizados e qual o número mínimo de interações necessárias para encontrar a resposta correta a cada questão.

De forma a agilizar o processo de avaliação experimental, foi decidido que o questionário apresentado aos utilizadores seria realizado através de um *google forms* e seria no formato de escolha múltipla. Isto, pois se as perguntas fossem de resposta aberta os utilizadores poderiam responder de diversas formas e tornaria o processo mais difícil de avaliar. Para além disso, a ordem das questões foi igual entre todos os participantes pois não foi viável estar a alterar as condições em que todos os testes seriam realizados,

optando-se assim por fazer somente testes que tivessem sempre as mesmas condições entre si. Decidiu-se ainda gravar o ecrã do computador de modo a poder, posteriormente, recolher mais informações relativas à experiência. Durante a realização dos testes foi-se apontando o tempo que o utilizador demorava a responder a cada pergunta, o número total de interações realizadas no *dashboard* e também as dificuldades apresentadas durante a interação com as ferramentas. Para além disso, no final de cada teste experimental foi dada a possibilidade aos utilizadores de fornecerem *feedback* relativo às ferramentas desenvolvidas.

ID Dashboard	Dashboard	Questão Específica	Interações Mínimas
D01	Giras - Estados da pandemia	Q01.01	2
		Q01.02	3
		Q01.03	1
D02	Giras - Datas específicas	-	-
D03	Parques de estacionamento	Q02.01	0
		Q02.02	2
D04	Congestionamentos - Geral	Q03.01	0
		Q03.02	2
		Q03.03	2
D05	Congestionamentos - Todas as zonas	Q04.01	1
D06	Congestionamentos - Zona específica	Q04.02	2
D07	Congestionamentos - POI	Q05.01	1
		Q05.03	3
D08	Qualidade do ar	Q06.01	2
		Q06.02	3
		Q06.03	3
D09	Qualidade do ar - Por estações	Q07.02	3
D10	Comparação entre poluentes	Q08.01	0
		Q08.02	2
		Q08.03	5
D11	Congestionamentos vs qualidade do ar	Q09.01	1

Figura 5.1: Resumo das questões utilizadas durante os testes experimentais.

Os primeiros testes experimentais foram feitos com utilizadores não entendidos na temática, ou seja, foram realizados com alguns colegas universitários. Deste modo, houve a possibilidade de realizar a maioria dos testes em moldes presenciais. A abordagem seguida para a realização destes testes consistiu em começar inicialmente por dar um contexto geral da dissertação a todos os participantes. De seguida, foi-lhes fornecido um formulário que continha todas as vinte perguntas e foram-lhes disponibilizados os *dashboards* que iriam utilizar, de forma interativa, durante a realização da avaliação experimental. Visto que a cada grupo de perguntas corresponde um *dashboard* diferente, foi feito, sempre que o utilizador iniciava as questões de um novo grupo, uma breve descrição e contextualização do *dashboard* que este iria utilizar. Quando a breve descrição acabava,

foi pedido ao utilizador para analisar o *dashboard* e quando este achasse que estava pronto poderia então começar o questionário. Deste modo, durante os testes experimentais, os utilizadores leram e interpretaram a pergunta, analisaram e interagiram com o *dashboard* e procuraram a resposta à questão nas visualizações.

Os testes experimentais com *stakeholders* entendidos na temática foram realizados remotamente, sendo que a abordagem seguida durante os testes foi semelhante à abordagem seguida durante os testes presenciais. Desde modo, o processo de avaliação foi todo seguido remotamente através de uma reunião *Zoom*<sup>1</sup> e as ferramentas desenvolvidas foram consultadas pelos mesmos através do Tableau Online. Foi pedido aos participantes que partilhassem o seu ecrã durante a reunião, de forma a ser possível ver as interações que os mesmos tinham com os *dashboards*. Para além disso, a reunião foi gravada para eventual consulta futura.

Dado que a avaliação de forma remota é feita através de uma plataforma diferente, ou seja, é utilizado o Tableau Online ao invés do Tableau Desktop, foi necessário averiguar, à priori, se as condições seriam semelhantes entre as duas plataformas. Concluiu-se que, para a realização dos testes, não existem diferenças significativas no tempo de resposta nem na usabilidade entre as duas plataformas.

## 5.2 Participantes

Foram realizadas no total 16 sessões de testes experimentais, sendo que 13 dessas sessões foram com utilizadores não entendidos na temática (81%) e 3 dessas sessões foram com *stakeholders* da CML entendidos na temática (19%). Embora tenham sido realizadas diversas reuniões intermédias ao longo da dissertação, nenhum dos *stakeholders* tinha visto todos os *dashboards* antes da realização da avaliação experimental, nem tinham visto a versão final dos *dashboards*. Os restantes participantes, com exceção de uma colega e dos *stakeholders*, não tinham conhecimento prévio deste trabalho. Nenhum dos participantes teve nenhuma explicação adicional sobre o funcionamento dos *dashboards*, sendo que a breve contextualização feita no decorrer das sessões a cada *dashboard* foi idêntica para todos os participantes.

Durante a realização dos testes experimentais foram recolhidos os dados demográficos dos participantes. Com o auxílio da figura 5.2(a), é possível concluir que a maioria dos participantes, cerca de 63%, têm idades compreendidas entre 22-26 anos, sendo que neste intervalo de idades existiram participantes do género feminino e masculino. Os restantes participantes tinham idades compreendidas entre 20-22 anos, 26-28 anos, 42-44 anos e 48-50 anos. Pela figura 5.2(a) é ainda possível observar que 50% dos participantes são do sexo masculino e 50% do sexo feminino. É ainda possível constatar, através da figura 5.2(b), que cerca de 63% dos participantes afirmaram ter experiência com visualização de dados, enquanto 38% dos participantes afirmaram que não tinham experiência com

---

<sup>1</sup><https://zoom.us/>

visualização de dados. A recolha deste tipo de informação permite aferir se a experiência de cada participante tem influência no seu desempenho durante a realização da avaliação experimental.

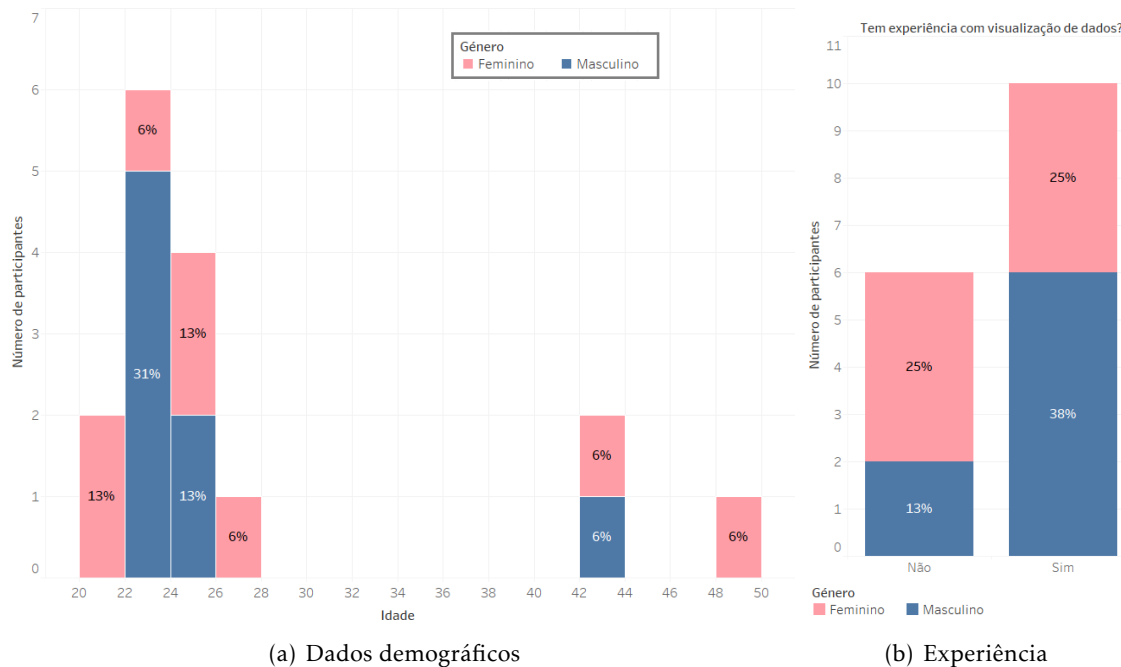


Figura 5.2: Distribuição dos participantes por dados demográficos (a) e por experiência com visualização de dados (b).

### 5.3 Resultados

Tal como referido na secção 5.1, as três principais métricas recolhidas para a análise dos resultados dos testes experimentais são:

- **Taxa de sucesso de resposta correta.**
- **Tempo médio de resposta a cada questão.**
- **Número total de ações de interação para responder a cada questão.** Sendo que ações de interação correspondem a filtragens, ações de destaque e mudanças de parâmetros. Clicar sobre um parque de estacionamento no mapa, fazer *hover* sobre um estado da pandemia, arrastar o *slider* e mudar o número de semanas associadas ao *slider* são exemplos de possíveis ações de interação nos *dashboards* desenvolvidos.

A primeira análise feita aos resultados de todos os testes experimentais foi relativamente à taxa de sucesso de resposta correta em cada questão. Esta primeira análise é visível na figura 5.3, na qual as questões se encontram ordenadas pela sua ordem de execução durante os testes. E, tal como é possível observar nesta figura, existiram 12

questões que tiveram sempre uma resposta correta por parte dos participantes, o que equivale a 60% das perguntas. Por outro lado, existiram 8 questões que não tiveram taxa de sucesso de 100%, o que equivale a 40% das perguntas. Sendo importante salientar que cada questão apresentou no máximo duas respostas erradas (cerca de 13%). As questões Q02.02, Q02.03 e Q09.01 foram erradas somente uma vez (cerca de 6%) e as questões Q02.01, Q03.03, Q04.01, Q06.03 e Q08.03 foram erradas duas vezes (cerca de 13%). Estas questões serão analisadas mais ao detalhe no decorrer desta subsecção.



Figura 5.3: Taxa de sucesso de resposta correta por cada questão.

Através da figura 5.3 é ainda possível observar que os *dashboards* que obtiveram 100% de taxa de sucesso durante todos os testes experimentais foram os *dashboards* das bicicletas Gira, dos congestionamentos (zona específica), dos congestionamentos (POI) e da qualidade do ar (por estações).

Visto que foram realizados testes com participantes que afirmaram ter experiência com visualização de dados, decidiu-se analisar se essa experiência influenciou a taxa de sucesso de resposta às questões. Foi observado que tanto os participantes com experiência em visualização de dados como os participantes sem experiência em visualização de dados apresentaram cerca de 4% de respostas erradas. Isto revela que o fator experiência não foi crucial para a taxa de sucesso de resposta correta e que qualquer utilizador consegue perceber os *dashboards* desenvolvidos com relativa facilidade.

Analisando os tempos médios de resposta a cada pergunta, visível na figura 5.4, é possível constatar que a questão que demorou, em média, mais tempo a ser respondida foi a questão Q01.02 (cerca de 99 segundos), enquanto que a questão Q03.02 foi a que demorou, em média, menos tempo a ser respondida (cerca de 39 segundos). Com o auxílio da figura 5.3 podemos verificar que nenhuma destas duas questões apresentou respostas erradas. Para além disso, é possível averiguar que as questões que apresentam respostas

erradas, ou seja, as questões Q02.01, Q02.02, Q03.01, Q03.03, Q04.01, Q06.02, Q08.03 e Q09.01 apresentaram, respetivamente, os tempos médios de 77 segundos, 54 segundos, 50 segundos, 42 segundos, 63 segundos, 80 segundos, 69 segundos e 74 segundos.

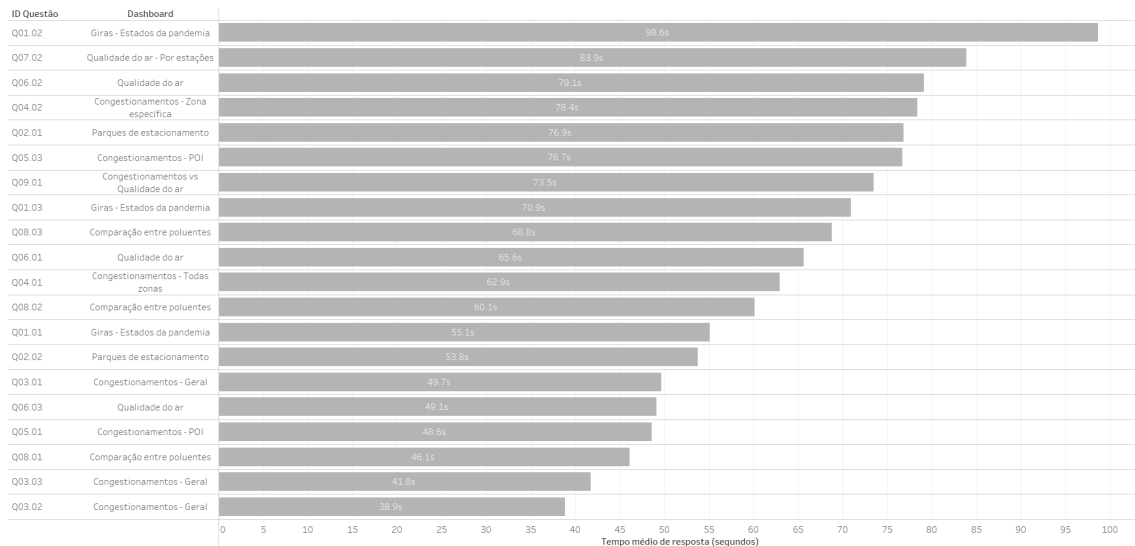


Figura 5.4: Tempo médio de resposta a cada pergunta - Ordenado do maior tempo para o menor tempo.

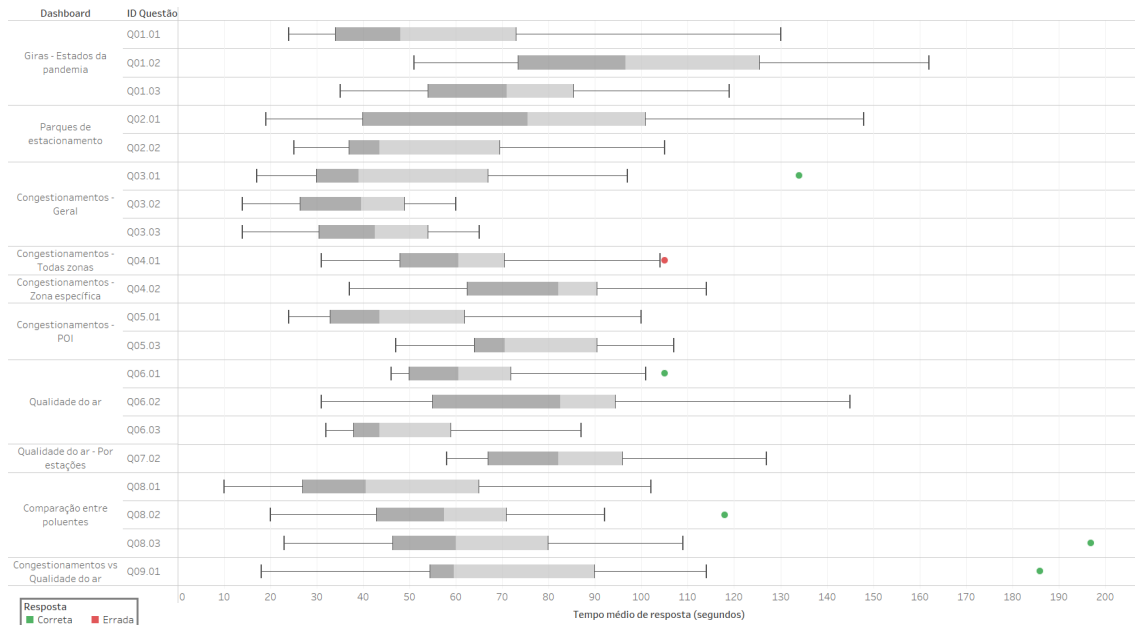


Figura 5.5: Distribuição do tempo de resposta dos participantes por cada pergunta de cada *dashboard*.

Com o intuito de analisar mais detalhadamente a distribuição dos tempos de resposta a cada pergunta por parte dos participantes, foi elaborado o diagrama de caixa de bigodes da figura 5.5. Este diagrama encontra-se ordenado segundo a ordem de execução de cada

pergunta durante os testes experimentais. O mesmo mostra que a distribuição dos tempos das respostas dos participantes variou consoante a pergunta. A questão que apresentou uma maior dispersão de valores entre o máximo e o mínimo foi a questão Q02.01 que, tal como é possível ver na figura 5.3, foi respondida de forma errada por dois participantes. Por outro lado, os *dashboards* que apresentaram uma maior dispersão de valores entre o máximo e o mínimo são os *dashboards* das bicicletas Gira e dos parques de estacionamento.

Uma curiosidade observada relativamente aos tempos das respostas erradas dos participantes é o facto de 75% destes valores se encontrarem acima do valor mediano de cada pergunta. Isto pode ser um indicativo de que esses participantes sentiram dificuldades durante a interação com o *dashboard* e essa dificuldade resultou na resposta incorreta a essas questões. Sendo que as questões que apresentam esses valores são a Q02.01, Q02.02, Q03.03, Q04.01, Q06.02 e Q08.03.

Tal como é possível observar na figura 5.5 existem seis questões que contém valores dos tempos de resposta considerados *outliers*. Desses seis valores existem dois mais evidentes, pertencentes às questões Q08.03 e Q09.01. O *outlier* pertencente à questão Q08.03 apresenta o valor de 197 segundos de tempo de resposta por parte do participante. Este valor mais elevado deveu-se ao facto do participante ter demonstrado alguma dificuldade na leitura da correlação entre os dois poluentes no *scatter plot*. Este mesmo participante sentiu a necessidade de trocar a ordem que escolheu inicialmente para os poluentes que são apresentados nos eixos X e Y. Isto permitiu-lhe reconfigurar os dados e obter uma perspetiva diferente sobre os mesmos de modo a responder corretamente à questão, todavia o seu tempo de resposta aumentou significativamente. A questão Q09.01 apresentou também um valor *outlier* referente a um participante que demorou 186 segundos a responder a essa mesma questão. Este participante demorou mais tempo a assimilar a questão e a compreender em que visualização do *dashboard* se encontrava a resposta à mesma. O participante justificou, no final da experiência, que demorou mais tempo pois o *dashboard* dos congestionamentos vs qualidade do ar é um pouco mais complexo do que os restantes e que, deste modo, precisou analisar mais detalhadamente cada visualização.

Por fim, é importante também analisar o único *outlier* que está associado a uma resposta errada na questão Q04.01. Este participante demorou 105 segundos a responder a esta questão pois, visto ter interpretado mal a questão, realizou uma ação de filtragem que não era suposto. Esta ação levou a uma resposta incorreta e também a um maior tempo de leitura do gráfico apresentado.

Para além da taxa de sucesso de resposta correta a cada questão e do tempo médio de resposta dos participantes, é também importante analisar o número total de ações de interação que os participantes realizaram nos *dashboards* para obter uma resposta a cada questão. Perante esta métrica surgiu a necessidade de definir, para cada questão, um número mínimo de ações de interação que são necessárias realizar de modo a ser possível obter uma resposta correta. Após este valor estar definido para cada questão, calculou-se qual foi o excedente/défice entre o mesmo e o número de ações de interação que cada utilizador efetuou em cada questão (fórmula 5.1). Isto permite, por exemplo, analisar

qual foi o comportamento que cada participante teve na interação com os *dashboards* e se as filtragens/destaques são perceptíveis para os utilizadores.

$$\text{excedente/déficit de interações} = \text{total\_interacoes} - \text{minimo\_interacoes} \quad (5.1)$$

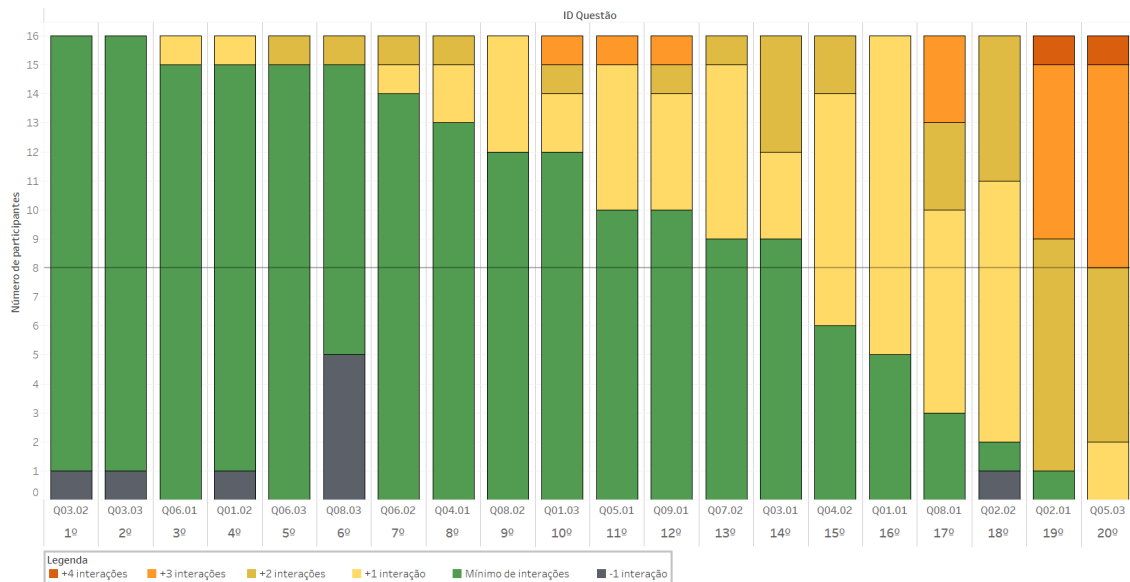


Figura 5.6: Número de participantes que realizaram o número mínimo de ações de interação, mais ações de interação que o necessário e menos ações de interação que o necessário por questão.

Na figura 5.6 é possível observar para cada questão quantos participantes é que realizaram o número mínimo necessário de ações de interação (mapeado a verde), quantos realizaram menos ações de interação que o mínimo necessário (mapeado a cinzento) e quantos realizaram mais ações de interação que o mínimo necessário (mapeado com tonalidades de amarelo e laranja). Esta figura encontra-se ordenada ascendentemente segundo o número de participantes que fizeram mais interações que as mínimas necessárias, ou seja, segundo o tamanho das barras mapeadas com tonalidades de amarelo e laranja. É possível observar que existiram 14 questões que tiveram mais de 50% dos utilizadores a realizarem o número mínimo de ações necessárias.

Outro aspeto importante que é visível na figura 5.6 foi o facto de existirem 5 questões onde os participantes não efetuaram uma ação de interação necessária para obter a resposta correta à respetiva questão. A questão Q03.02 (na 1ª posição) e a questão Q03.03 (na 2ª posição) são referentes ao *dashboard* geral dos congestionamentos e, para obter uma resposta correta, era necessário fazer uma **filtragem temporal com o slider** (tendo por base a tabela 4.17). O participante que em ambas as questões realizou menos uma ação de interação do que era suposto esqueceu-se de alterar o número de semanas associadas ao *slider*. Contudo, o número de semanas que estavam associadas ao *slider* não alteraram o resultado final e embora não tenham sido feitas todas as ações necessárias, a resposta foi

correta em ambas as questões. A questão Q01.02 (na 4ª posição), referente ao *dashboard* das bicicletas Gira, foi novamente um caso isolado de um utilizador que se esqueceu de realizar uma **filtragem temporal por datas/horas específicas**, mais concretamente filtrar por dias de trabalho. Porém, a falta desta ação de filtragem não mudou significativamente o resultado final da visualização, então o utilizador acertou a resposta à questão. Por outro lado, na questão Q02.02 (na 18ª posição), alusiva ao *dashboard* dos parques de estacionamento, a falta da ação de **filtragem temporal por datas/horas específicas** levou a que o utilizador respondesse erradamente à questão. Para esta questão era necessário que o utilizador escolhesse o parque da Avenida Lusíada e o mês de junho para filtrar as visualizações, todavia este utilizador esqueceu-se de selecionar o mês de junho e isto levou a uma alteração significativa da visualização onde era possível encontrar a resposta à questão. A questão Q08.03 (na 6ª posição), relativa ao *dashboard* da comparação entre poluentes, foi onde houve mais participantes a realizar menos uma interação que o mínimo necessário. A estes cinco participantes faltou a realização da ação de **filtragem/destaque por outros parâmetros**, mais concretamente a escolha da estação de qualidade do ar em análise. Contudo, a falta desta interação não causou uma mudança significativa na visualização, levando a que estes cinco participantes acertassem todos a questão.

É de salientar que houve seis questões em que a maioria dos participantes realizaram mais ações de interação que o necessário. Então houve a necessidade de analisar mais ao detalhes essas questões, tendo por base a tabela 4.17 das interações gerais (apresentada na secção 4.5):

- **Q05.03** (“Durante a primeira semana do estado de emergência, qual foi o dia que apresentou um maior número de congestionamentos ao redor do Hospital Santa Maria?”): Esta questão encontra-se na 20ª posição do gráfico da figura 5.6. Para obter uma resposta correta a esta pergunta bastavam apenas ser realizadas três interações. As interações necessárias seriam: **filtragem/destaque por outros parâmetros**, mais concretamente a seleção da categoria do **POI**, **filtragem/destaque por outros parâmetros**, mais precisamente o destaque do Hospital de Santa Maria no mapa e **filtragem espacial no mapa** do hospital. Nenhum dos participantes realizou apenas três interações, sendo que realizaram entre quatro a sete interações para responder a esta questão. A maioria sentiu a necessidade de efetuar uma **filtragem temporal com slider** o que causou um aumento de duas interações, isto pois alteraram o número de semanas associado ao *slider* e arrastaram o *slider* para o início do estado de emergência. Alguns participantes sentiram ainda a necessidade de efetuar outra **filtragem/destaque por outros parâmetros** que consistiu em filtrar a subcategoria dos pontos de interesse para hospitais. Isto mostra que os utilizadores, mesmo podendo fazer menos interações, preferem fazer mais do que o necessário de modo a terem a certeza que acertam a questão e de forma a filtrar mais os dados para reduzir a quantidade de informação apresentada.

- **Q02.01** (“O parque de estacionamento da Avenida Lusíada ficou, na sua maioria, mais cheio ou mais vazio após o aparecimento da COVID-19?”): Esta questão encontra-se na 19ª posição do gráfico da figura 5.6. Para responder a esta questão bastava apenas olhar para o *heat table* que continha o top 10 parques de estacionamento com maior taxa de ocupação por dias, ou seja, não eram necessárias interações. Contudo, os participantes realizaram entre zero a quatro interações para chegarem à resposta correta. A maioria dos participantes fez uma **filtragem/destaque por outros parâmetros**, mais precisamente o destaque do parque Avenida Lusíada no mapa, seguido de uma **filtragem espacial no mapa** desse parque. Isto é um indicativo de que os utilizadores tendem a olhar primeiramente para as visualizações que se encontram no topo do *dashboard*. Estes, ao verem o mapa, optam por filtrar o parque e obter a resposta através do gráfico de barras que apresenta a taxa de ocupação média do parque escolhido.
- **Q02.02** (“Relativamente à alteração média da taxa de ocupação por hora comparativamente com 14 dias atrás, qual é o comportamento observado no parque Avenida Lusíada durante o mês de junho?”): Esta questão encontra-se na 18ª posição do gráfico da figura 5.6. A resposta a esta questão era obtida após a realização de, pelo menos, duas interações. As interações mínimas necessárias eram: **filtragem/destaque por outros parâmetros**, mais concretamente a seleção do parque Avenida Lusíada no *heat table* e **filtragem temporal por datas/horas específicas** do mês de junho. Porém, alguns participantes realizaram mais uma ou duas interações que o mínimo necessário. Tal como foi referido anteriormente para a questão Q02.01, alguns participantes realizaram **filtragem/destaque por outros parâmetros**, mais precisamente o destaque do parque no mapa e seguidamente a **filtragem espacial no mapa**. Também alguns participantes mudaram a granularidade temporal associada ao gráfico de barras para meses, ou seja, realizaram mais uma **filtragem/destaque por outros parâmetros**.
- **Q08.01** (“Após o aparecimento da COVID-19 e numa perspetiva diária, qual foi a classe predominante do índice de qualidade do ar em todos os poluentes medidos nas diversas estações?”): Esta questão encontra-se na 17ª posição do gráfico da figura 5.6. De modo a obter uma resposta correta a esta questão bastava observar o *heat table* que mostra a média diária do índice de qualidade do ar de cada poluente nas respetivas questões, ou seja, não eram necessárias interações. Todavia, existiram participantes a realizar entre zero a três ações de interação. A maioria dos participantes sentiu a necessidade de realizar uma **filtragem/destaque por outros parâmetros**, mais precisamente o destaque da cor da classe do índice de qualidade do ar (clicando na legenda). Esta interação dava-lhes mais uma confirmação que respondiam corretamente à questão. Alguns participantes também realizaram uma **filtragem temporal com slider**, sendo que alguns arrastavam o *slider* para o início do aparecimento da COVID-19 e por vezes também alteravam o número de semanas

associado ao *slider*. Esta interação adverte, muito possivelmente, de um processo de aprendizagem e habituação associado aos *sliders* dos *dashboards* anteriores. Isto, pois como foi necessário realizarem algumas vezes esta interação nos *dashboards* anteriores, os participantes tiveram a tendência para a realizar novamente, contudo no final todos perceberam que não era necessário.

- **Q01.01** (“A taxa de ocupação ao longo das horas, na estação 103, é semelhante em 2019 e 2020 durante a altura do estado de emergência?”): Esta questão encontra-se na 16ª posição do gráfico da figura 5.6. Para responder corretamente a esta questão era necessário efetuar no mínimo duas ações de interação. As interações necessárias eram: **filtragem temporal por estados da pandemia** do estado de emergência e **filtragem espacial no mapa** da estação 103. Porém, houve 11 participantes que, para além destas duas interações, realizaram também uma **filtragem/destaque por outros parâmetros**, mais precisamente o destaque da estação no mapa. Isto é um indicativo de que os utilizadores consideraram mais prático a pesquisa da estação no parâmetro que faz o destaque no mapa, do que a sua procura através das *labels* presentes no mapa.
- **Q04.02** (“Existiram mudanças significativas no número de congestionamentos diários ocorridos na Avenida da Liberdade após a declaração do estado de alerta?”): Esta questão encontra-se na 15ª posição do gráfico da figura 5.6. Para responder corretamente a esta questão era preciso realizar no mínimo duas ações de interação. Era necessário realizar **filtragem por zonas** duas vezes, sendo que primeiramente era necessário selecionar que se pretende analisar avenidas e de seguida selecionar a Avenida da Liberdade. Houve contudo alguns participantes que, para além destas duas interações, realizaram mais uma ou duas interações. Estas interações excedentes foram a **filtragem temporal com slider**, sendo que alguns participantes arrastavam o *slider* para o início do estado de alerta e dois deles também alteraram o número de semanas associado ao *slider*. Isto mostra, novamente, que pode ter havido um processo de aprendizagem e habituação associado aos *sliders* dos *dashboards* anteriores.

É importante ainda realçar que, destas seis questões em que os participantes realizaram mais interações que o necessário, apenas as questões Q02.01 e Q02.02 tiveram respostas erradas. Sendo que, tal como é possível ver na figura 5.3, em dezasseis respostas, a questão Q02.01 apenas teve duas respostas erradas e a questão Q02.02 uma resposta errada.

Com os resultados vistos anteriormente, surgiu o interesse de analisar conjuntamente as três métricas usadas nas análises anteriores. Este interesse provém da curiosidade de procurar se existe algum padrão interessante entre estas métricas, de modo a ser possível tirar mais conclusões. Na figura 5.7 está ilustrado um *scatter plot* que apresenta no eixo do X o tempo médio de resposta por questão e no eixo do Y a média de interações extra

por questão (média do excedente/défice de interações apresentado na fórmula 5.1). Para além disso, este *scatter plot* contém ainda mapeado para a variável visual cor a taxa de sucesso do utilizador, de modo que cada círculo é referente a um utilizador diferente.

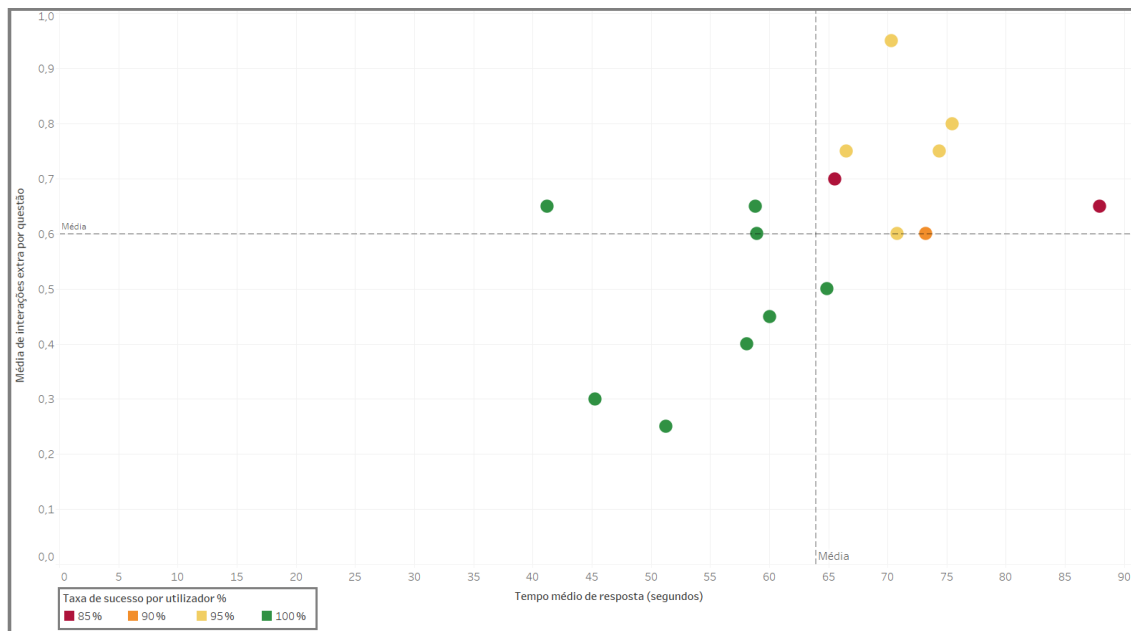


Figura 5.7: *Scatter Plot* com o tempo médio de resposta, média de interações extra por questão e taxa de sucesso por utilizador.

Através da figura 5.7 é possível detetar alguns fenómenos interessantes. O fenómeno que salta logo à vista é o facto de existirem semelhanças entre os comportamentos dos participantes que apresentam a mesma taxa de sucesso, havendo nitidamente dois grupos visíveis no gráfico. O primeiro desses grupos é referente aos participantes que tiveram 100% de taxa de sucesso, sendo que se conclui através do gráfico que os participantes que acertaram todas as questões tiveram todos um tempo médio de resposta inferior a 65 segundos. O outro grupo perceptível através deste gráfico é composto pela maioria dos participantes que obtiveram 95% de taxa de sucesso, tendo todos em média mais 0.75 interações extra por questão. Para além disso, todos os participantes que apresentaram uma taxa de sucesso inferior a 100% tiveram um tempo médio de resposta superior a 65 segundos. O gráfico da figura 5.7 mostra ainda que há alguma tendência para serem realizadas mais ações extra à medida que o tempo médio de resposta aumenta.

Em suma, os participantes que acertaram todas as questões foram mais rápidos e tiveram uma tendência para fazerem menos interações extra por questão. Os participantes que erraram pelo menos uma questão não são tão rápidos a responder e apresentam, por norma, mais interações extra por questão. Isto leva a crer que as respostas erradas às questões não provêm do facto dos participantes tentarem responder apressadamente, mas sim desses participantes apresentarem maiores dificuldades em interpretar a questão e/ou o *dashboard*, sendo que em alguns casos não chegaram à resposta correta.

Por último, falta analisar mais detalhadamente as questões que obtiveram uma maior percentagem de respostas incorretas. De seguida, é feito um levantamento das razões que levaram os participantes a responder erradamente, tendo em conta a análise feita às sessões experimentais realizadas com os utilizadores:

- **Q02.01** (“O parque de estacionamento da Avenida Lusíada ficou, na sua maioria, mais cheio ou mais vazio após o aparecimento da COVID-19?”): Os dois participantes que erraram esta questão responderam que o parque ficou mais vazio após o aparecimento da COVID-19. Isto deveu-se ao facto de ambos terem reparado que o *heat table* apresentava umas zonas a cinzento após o aparecimento da pandemia. A cor cinzento significa que existe falta de dados no *dataset* nessa altura, contudo os participantes associaram esta cor a uma taxa de ocupação mais baixa.
- **Q03.03** (“Em 2020, durante as duas primeiras semanas do estado de emergência existiram picos horários da média de quilómetros congestionados?”): Os dois participantes que erraram esta questão responderam que existiu, em 2020, um pico horário às 18h de 78km. Ambos os participantes erraram a questão pois não leram a legenda de cores do gráfico de linhas e, deste modo, olharam para a linha de 2019 ao invés da linha de 2020.
- **Q04.01** (“Quais foram as três avenidas de Lisboa onde se sentiu um maior impacto da COVID-19 em relação ao número de congestionamentos?”): Esta questão foi errada por dois participantes, sendo que ambos responderam: Eixo Central, Avenida da Liberdade e Avenida das Forças Armadas. A resposta correta a esta questão era: Eixo Central, Avenida da Liberdade e Avenida Almirante Reis. Ambos os participantes responderam errado pois tiveram uma interpretação diferente à pergunta. Isto levou-os a realizar uma ação de interação de mudança dos níveis de congestionamentos mostrados no gráfico, o que por sua vez os conduziu à resposta errada.
- **Q06.02** (“Para o conjunto de estações da Avenida da Liberdade e Entrecampos, qual foi o estado da pandemia que parece ter tido um maior impacto nos valores médios diários de  $\text{NO}_2$  em 2020?”): A resposta incorreta a esta questão deveu-se a uma interpretação diferente da mesma por parte dos dois participantes. Ambos os participantes responderam estados da pandemia que apresentavam picos de grandes concentrações de  $\text{NO}_2$  nos anos anteriores a 2020 ao invés de escolherem o estado que apresentava maior distância geral entre as concentrações diárias de 2020 com os restantes anos. Desta forma, a resposta correta à questão seria o estado de emergência.
- **Q08.03** (“Analisando as quatro primeiras semanas após o aparecimento da COVID-19, considera que existe alguma correlação entre os poluentes  $\text{NO}_2$  e  $\text{O}_3$  medidos na estação de Entrecampos?”): Os dois participantes que erraram esta questão efetuaram todas as interações necessárias para encontrar a resposta correta, porém

sentiram dificuldades na leitura da correlação através do *scatter plot*. Deste modo, ambos responderam que existe correlação positiva e a resposta correta era que existe correlação negativa entre as concentrações dos dois poluentes.

## 5.4 Conclusões

Os testes experimentais realizados com utilizadores foram um sucesso e através destes foi possível retirar algumas conclusões relativas aos *dashboards* desenvolvidos.

Primeiramente foi inferido que não existiu grandes diferenças relativamente à taxa de sucesso, tempo médio de resposta e número de interações entre os utilizadores não entendidos na temática e os *stakeholders* da CML entendidos na temática. Para além disso, o fator de experiência com visualização de dados e o género dos participantes também não influenciaram estas três métricas de avaliação medidas durante os testes experimentais. Estes resultados mostram que os *dashboards* e as suas visualizações são compreensíveis tanto para utilizadores com experiência e entendidos na temática como para utilizadores comuns.

Foi visto que muitas das questões erradas por parte dos utilizadores foram referentes a uma diferente interpretação das mesmas e também a alguma distração ou esquecimento da realização de interações. Reparou-se também que os participantes, por vezes, preferiram realizar mais interações que o necessário para responder à questão, de modo a conseguirem ter uma maior certeza que estão a responder corretamente. Foi gratificante perceber que os participantes conseguiram de facto utilizar com grande eficácia os *dashboards* e conseguiram compreender o papel das interações de forma a conseguirem responder às questões colocadas. Verificou-se que os participantes conseguiram perceber o objetivo das interações como um sendo mecanismo para os ajudar a extrair mais informações através das visualizações. Outra conclusão interessante foi o facto de existirem comportamentos semelhantes entre os participantes que apresentam a mesma taxa de sucesso.

Durante o decorrer dos testes experimentais, os participantes foram dando *feedback* e sugestões de pequenas melhorias dos *dashboards* desenvolvidos. A primeira sugestão foi referente ao *dashboard* das bicicletas Gira. Muitos dos participantes sugeriram que o filtro dos dias a analisar (dias de trabalho ou fins de semana) estivesse no topo das visualizações que são afetadas por este, sendo que esta mudança é vista na figura 5.8. Alguns participantes também sugeriram que a visualização da média móvel presente no *dashboard* da qualidade do ar apresentasse no eixo do X os dias dos meses ao invés de ser de 1 a 365 dias do ano (figura 5.9). Para além disso, houve ainda participantes que associaram a cor cinzenta, presente no *heat table* do *dashboard* dos parques de estacionamento, a uma menor taxa de ocupação ao invés de a falta de dados. Então foi decidido trocar a cor cinzenta pela cor branca, de modo a que a representação da falta de dados ficasse semelhante aos restantes gráficos do *dashboard* (figura 5.10).

## CAPÍTULO 5. AVALIAÇÃO DO TRABALHO REALIZADO

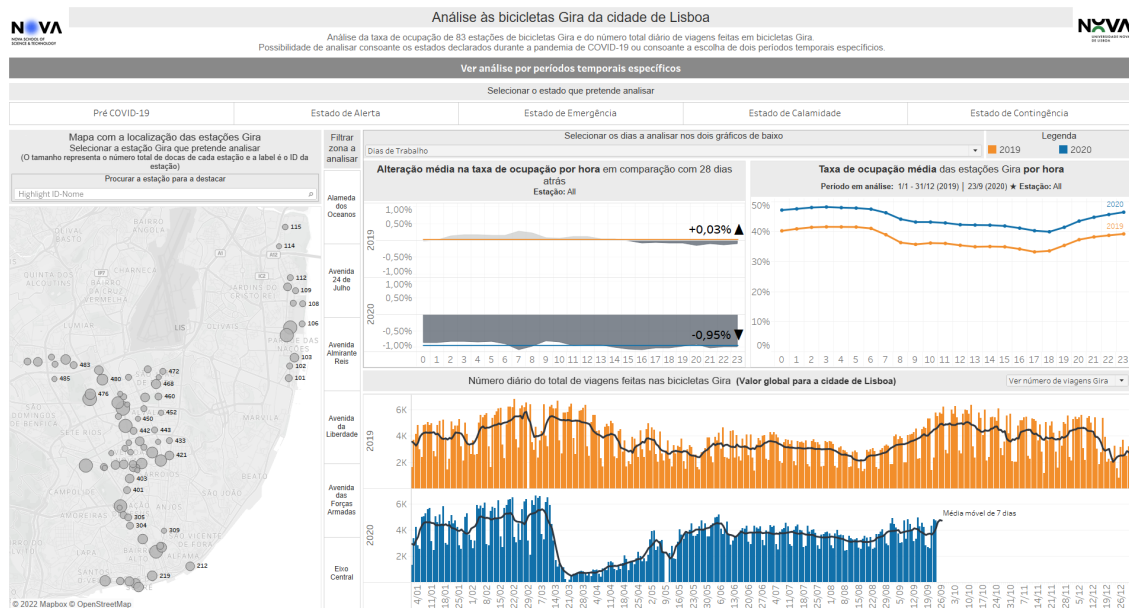


Figura 5.8: *Dashboard* para a análise às bicicletas Gira da cidade de Lisboa - Após melhorias.

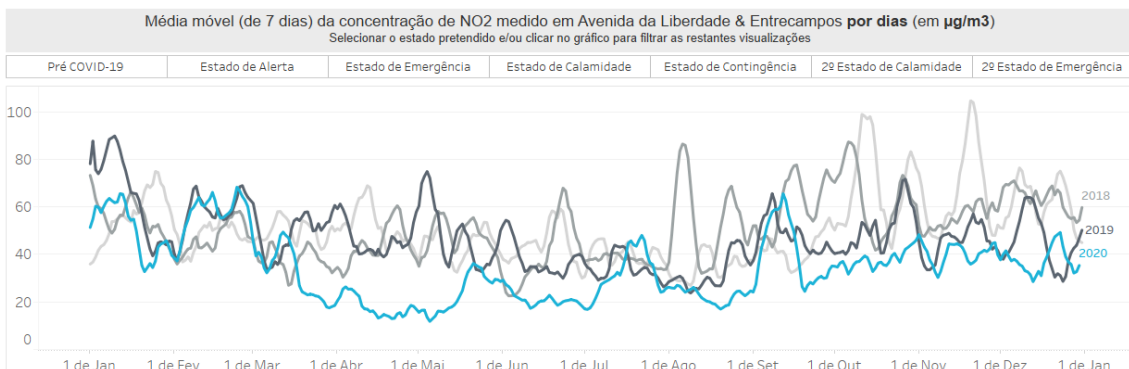


Figura 5.9: Gráfico com a média móvel diária (V3) do *dashboard* da qualidade do ar - Após melhorias.

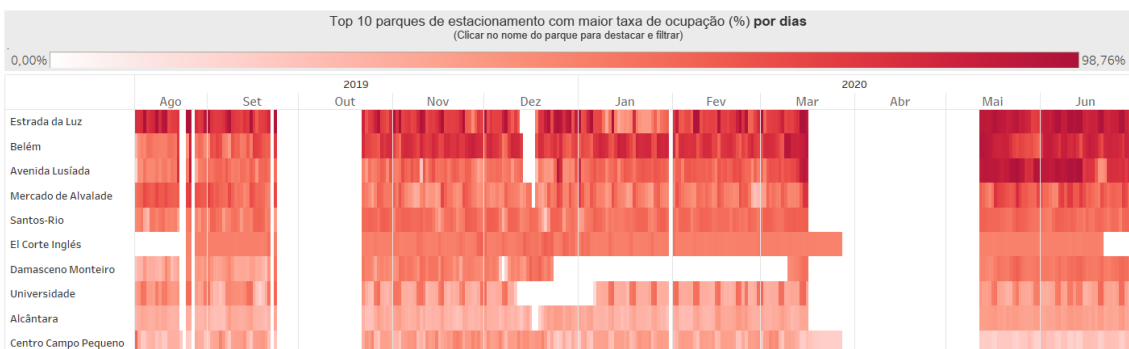


Figura 5.10: *Heat table* (V4) do *dashboard* dos parques de estacionamento - Após melhorias.

## CONCLUSÕES E TRABALHO FUTURO

Os principais objetivos traçados para esta dissertação foram cumpridos com sucesso. O primeiro grande objetivo centrou-se na criação de *dashboards* que permitissem analisar e determinar o impacto que a COVID-19 teve na mobilidade e na qualidade do ar da cidade de Lisboa. Cada uma das ferramentas desenvolvidas possibilita a resposta a uma questão analítica geral que foi levantada tendo em conta o *feedback* obtido por parte dos *stakeholders* da Câmara Municipal de Lisboa. Para além disso, com cada *dashboard* é ainda possível responder a um vasto conjunto de questões mais específicas. Outro objetivo delineado para esta dissertação foi a criação de ferramentas que possibilitassem a correlação entre métricas da qualidade do ar e métricas dos congestionamentos detetados pelo *Waze*, de forma a verificar se existe alguma relação entre elas. Em adição, foi ainda criada uma ferramenta que permitisse comparar o comportamento dos diferentes poluentes medidos em cada estação de qualidade do ar. Com todas as ferramentas desenvolvidas, não só é possível averiguar de forma interativa qual foi o impacto da pandemia na mobilidade urbana e na qualidade do ar, mas também avaliar ambas as temáticas num panorama geral, não considerando a pandemia. Todo o trabalho teve como base a metodologia **CRISP-DM** e durante o processo de criação de cada *dashboard* foram sempre seguidas boas práticas no que toca ao desenho dos *dashboards*.

No seguimento da dissertação surgiram também objetivos secundários, tais como a realização de um relatório que avaliasse a qualidade inerente a cada conjunto de dados e, conseqüentemente, o processamento dos dados de forma a ficarem com uma melhor qualidade. Houve ainda a oportunidade de contribuir com alguns dos *dashboards* desenvolvidos para um *workshop* a nível internacional da Rede C40. A Rede C40 é um grupo que reúne as grandes cidades do mundo empenhadas a debater contra as mudanças climáticas. O seu principal objetivo é criar soluções que melhorem a qualidade do ar e que permitam a redução dos gases de efeito de estufa. A contribuição com as ferramentas desenvolvidas para a apresentação da **CML** neste *workshop* foi uma mais valia para a divulgação do trabalho realizado nesta dissertação.

Apesar de terem sido cumpridos todos os objetivos delineados para esta dissertação, existem sempre aspetos a serem considerados para um trabalho futuro. Um dos aspetos

que poderá ser considerado futuramente é a atualização dos conjuntos de dados de modo a abrangerem um período temporal maior. Com dados do ano 2021 é possível, não só analisar o impacto que o confinamento e as suas restrições tiveram na mobilidade e qualidade do ar, mas também o impacto que as medidas de desconfinamento tiveram. Tendo os dados de 2021 integrados, iriam existir *dashboards* em que não era necessário fazer nenhum tipo de alterações e iriam existir *dashboards* onde era necessário fazer mudanças. Os *dashboards* que apresentam filtragem temporal por estados da pandemia precisariam de sofrer alterações, visto que atualmente a filtragem de estados é referente aos estados de 2020 e afeta o mesmo período temporal de 2019. Uma possível solução, tendo em conta que o ano 2021 apresenta outros estados da pandemia em datas diferentes, seria adicionar um parâmetro que permitia escolher se eram apresentados os estados de 2020 ou de 2021. Outros *dashboards* que precisariam de alteração são os que apresentam gráficos com a diferença entre o número de congestionamentos nos mesmos dias da semana de 2019 e 2020. Uma possível alteração seria, por exemplo, adicionar dois parâmetros nos quais se poderia escolher os dois anos que se pretendia comparar.

Para além disso, algo que poderia ser considerado no futuro seria estender as ferramentas criadas de modo a ser possível acomodar os dados da nova rede de monitorização em tempo real da qualidade do ar, níveis de ruído, meteorologia e de trânsito. Esta rede conta com oitenta estações e mais de seiscentos sensores que fazem a cobertura espacial de toda a cidade de Lisboa. Visto que a mesma foi inaugurada no mês de abril de 2021 não existem dados históricos anteriores à pandemia para servirem de termo de comparação, contudo seria interessante analisar o impacto que a pandemia teve a partir dos meses que existem dados.

Outro aspeto que poderia ser considerado num trabalho futuro seria a realização de uma nova avaliação experimental que permitisse analisar se existe algum processo de aprendizagem entre *dashboards* e também algum processo de aprendizagem relativo a cada *dashboard* específico. Para testar este aspeto seria necessário alterar as condições em que os testes são realizados, como por exemplo, fazer com que cada pergunta apareça numa ordem aleatória no questionário. Seria ainda necessário ter mais do que 16 participantes. Contudo, é de notar que não existem, ao longo das questões, evidências concretas de ter existido um processo de aprendizagem. Caso tenha havido foi apenas nas primeiras questões pois, tal como é possível verificar através da figura 5.5, o tempo médio de resposta e a sua dispersão dependeu da execução de cada pergunta e não da posição em que a mesma foi realizada. O único aspeto onde porventura existiu um processo de aprendizagem foi na utilização do *slider* para a realização da filtragem temporal.



- [10] S. D. Bernardes et al. “NYC Recovery at a Glance – The Rise of Buses and Micromobility”. en. Em: (2020), p. 7 (ver pp. 14, 19, 29, 30).
- [11] X. Chai et al. “Analysis of Spatial-temporal Behavior Pattern of the Share Bike Usage during COVID-19 Pandemic in Beijing”. Em: *arXiv e-prints*, arXiv:2004.12340 (abr. de 2020), arXiv:2004.12340. arXiv: 2004. 12340 [physics.soc-ph] (ver pp. 19, 28, 29, 68).
- [12] *Chemistry in the Sunlight*. URL: [https://earthobservatory.nasa.gov/features/ChemistrySunlight/chemistry\\_sunlight3.php](https://earthobservatory.nasa.gov/features/ChemistrySunlight/chemistry_sunlight3.php) (acedido em 30/01/2021) (ver p. 31).
- [13] *COVID-19 Air Quality Impact In 10 Major Cities*. URL: <https://www.iqair.com/blog/air-quality/report-impact-of-covid-19-on-global-air-quality-earth-day> (acedido em 01/02/2021) (ver pp. 15, 32, 33, 35, 36).
- [14] *Create Maps that Highlight Visual Clusters of Data in Tableau*. URL: [https://help.tableau.com/current/pro/desktop/en-us/maps\\_howto\\_pointdistribution.htm](https://help.tableau.com/current/pro/desktop/en-us/maps_howto_pointdistribution.htm) (acedido em 12/02/2021) (ver p. 18).
- [15] *Emissões de dióxido de carbono nos carros: factos e números (infografia): Atualidade: Parlamento Europeu*. Abr. de 2019. URL: <https://www.europarl.europa.eu/news/pt/headlines/society/20190313ST031218/emissoes-de-dioxido-de-carbono-nos-carros-factos-e-numeros-infografia> (acedido em 19/01/2021) (ver p. 2).
- [16] U. Environment. *Visual feature: The Emissions Gap Report 2019*. URL: <https://www.unenvironment.org/interactive/emissions-gap-report/2019/> (acedido em 19/01/2021) (ver p. 2).
- [17] N. Fann et al. *Ch. 3: Air Quality Impacts*. Abr. de 2016. URL: <https://health2016.globalchange.gov/air-quality-impacts> (acedido em 01/02/2021) (ver p. 35).
- [18] P. Federico et al. “Qualizon Graphs: Space-Efficient Time-Series Visualization with Qualitative Abstractions”. Em: *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*. AVI '14. Como, Italy: Association for Computing Machinery, 2014, pp. 273–280. ISBN: 9781450327756. DOI: 10.1145/2598153.2598172. URL: <https://doi.org/10.1145/2598153.2598172> (ver p. 16).
- [19] M. d’Amaral Ferreira. “Visualização Interativa de Dados para Projetos de Monitorização Fitossanitária”. Tese de mestrado. Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa, 2021 (ver p. 48).
- [20] A. Figueiras. “Towards the understanding of interaction in information visualization”. Em: *2015 19th International conference on information visualisation*. IEEE. 2015, pp. 140–147 (ver pp. 22, 23, 25).

- [21] E. D. Freitas et al. *Mobility Restrictions and Air Quality under COVID-19 Pandemic in São Paulo, Brazil*. en. preprint. EARTH SCIENCES, abr. de 2020. DOI: [10.20944/preprints202004.0515.v1](https://doi.org/10.20944/preprints202004.0515.v1). URL: <https://www.preprints.org/manuscript/202004.0515/v1> (ver pp. 21, 28, 33).
- [22] J. Gao et al. “Initial Impacts of COVID-19 on Transportation Systems: A Case Study of the U.S. Epicenter, the New York Metropolitan Area”. en. Em: (2020), p. 4 (ver pp. 26, 27, 30).
- [23] Y. H. Healy e Conor. *Hexbin map*. URL: <https://www.data-to-viz.com/graph/hexbinmap.html> (acedido em 12/02/2021) (ver p. 19).
- [24] J. Huang et al. “Understanding the Impact of the COVID-19 Pandemic on Transportation Related Behaviors with Human Mobility Data”. Em: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery amp; Data Mining. KDD '20*. Virtual Event, CA, USA: Association for Computing Machinery, 2020, pp. 3443–3450. ISBN: 9781450379984. DOI: [10.1145/3394486.3412856](https://doi.org/10.1145/3394486.3412856). URL: <https://doi.org/10.1145/3394486.3412856> (ver pp. 25, 26, 68).
- [25] IBM Corporation. *IBM SPSS Modeler CRISP-DM Guide*. (pp. 1-50) (ver pp. 39–41).
- [26] *Laboratório de Dados Urbanos de Lisboa*. URL: <https://lisboainteligente.cm-lisboa.pt/lxi-iniciativas/lx-data-lab/> (acedido em 18/01/2021) (ver p. 1).
- [27] *Lisboa Integra Grupo da Liderança Climática das Grandes Cidades - C40*. Jan. de 2019. URL: <https://www.lisboa.pt/atualidade/noticias/detalhe/lisboa-integra-grupo-da-lideranca-climatica-das-grandes-cidades-c40> (acedido em 28/12/2021) (ver p. 4).
- [28] Lusa. *Suspenso pagamento de estacionamento na via pública em Lisboa*. Mar. de 2020. URL: [https://sicnoticias.pt/especiais/coronavirus/2020-03-16-Suspenso-pagamento-de-estacionamento-na-via-publica-em-Lisboa?fbclid=IwAR3jrrydTkcjyuDY7QdeVG5MmauslickuF\\_HfnqDywa3idKn49b0JQZtntA](https://sicnoticias.pt/especiais/coronavirus/2020-03-16-Suspenso-pagamento-de-estacionamento-na-via-publica-em-Lisboa?fbclid=IwAR3jrrydTkcjyuDY7QdeVG5MmauslickuF_HfnqDywa3idKn49b0JQZtntA) (ver p. 59).
- [29] A. Maia. *Coronavírus: há dois casos confirmados em Portugal*. Mar. de 2020. URL: <https://www.publico.pt/2020/03/02/sociedade/noticia/coronavirus-ha-dois-infectados-portugal-1905823> (acedido em 17/01/2021) (ver p. 1).
- [30] J. Martorell-Marugán et al. “DatAC: A visual analytics platform to explore climate and air quality indicators associated with the COVID-19 pandemic in Spain”. en. Em: *Science of The Total Environment* 750 (jan. de 2021), p. 141424. ISSN: 00489697. DOI: [10.1016/j.scitotenv.2020.141424](https://doi.org/10.1016/j.scitotenv.2020.141424). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0048969720349536> (ver pp. 32–34).
- [31] T. Munzner. *Visualization analysis and design*. CRC press, 2014 (ver pp. 12, 13).
- [32] *MySQL Workbench*. URL: <https://www.mysql.com/products/workbench/> (acedido em 20/10/2021) (ver p. 51).

- [33] E. Newburger. *Covid pandemic drove a record drop in global carbon emissions in 2020*. Dez. de 2020. URL: <https://www.cnn.com/2020/12/11/covid-record-drop-global-carbon-emissions-2020.html> (acedido em 23/02/2021) (ver p. 2).
- [34] W. H. Organization et al. "Coronavirus disease 2019 (COVID-19): Situation Report - 51". Em: (2020). URL: [https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200311-sitrep-51-covid-19.pdf?sfvrsn=1ba62e57\\_10](https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200311-sitrep-51-covid-19.pdf?sfvrsn=1ba62e57_10) (ver p. 1).
- [35] C. Perin, F. Vernier e J.-D. Fekete. "Interactive Horizon Graphs: Improving the Compact Visualization of Multiple Time Series". Em: New York, NY, USA: Association for Computing Machinery, 2013, pp. 3217–3226. ISBN: 9781450318990. URL: <https://doi.org/10.1145/2470654.2466441> (ver p. 16).
- [36] *Personalizar a aparência do seu mapa*. URL: [https://help.tableau.com/current/pro/desktop/pt-br/maps\\_options.htm](https://help.tableau.com/current/pro/desktop/pt-br/maps_options.htm) (acedido em 28/12/2021) (ver p. 18).
- [37] Portugal.gov.pt. *Decreto do Governo que regulamenta o estado de emergência*. URL: <https://www.portugal.gov.pt/pt/gc22/comunicacao/documento?i=decreto-do-governo-que-regulamenta-o-estado-de-emergencia-> (acedido em 17/01/2021) (ver p. 1).
- [38] Qgis. URL: [https://qgis.org/pt\\_PT/site/index.html](https://qgis.org/pt_PT/site/index.html) (acedido em 20/10/2021) (ver p. 50).
- [39] S. Rodrigues e A. Figueiras. "There and then: interacting with spatio-temporal visualization". Em: *2020 24th International Conference Information Visualisation (IV)*. IEEE. 2020, pp. 146–152 (ver pp. 24, 25).
- [40] H. A. Rothan e S. N. Byrareddy. "The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak". Em: *Journal of Autoimmunity* 109 (2020), p. 102433. ISSN: 0896-8411. DOI: <https://doi.org/10.1016/j.jaut.2020.102433>. URL: <http://www.sciencedirect.com/science/article/pii/S0896841120300469> (ver p. 1).
- [41] Shyam, Lenin e Bethany. *Visualizing Time Series Data: 7 Types of Temporal Visualizations - Atlan: Humans of Data*. Ago. de 2019. URL: <https://humansofdata.atlan.com/2016/11/visualizing-time-series-data/> (acedido em 02/02/2021) (ver p. 15).
- [42] P. Sicard et al. "Amplified ozone pollution in cities during the COVID-19 lockdown". Em: *Science of The Total Environment* 735 (2020), p. 139542. ISSN: 0048-9697. DOI: <https://doi.org/10.1016/j.scitotenv.2020.139542>. URL: <http://www.sciencedirect.com/science/article/pii/S004896972033059X> (ver p. 32).

- 
- [43] *Tableau Dashboard - C2SMART*. URL: [https://public.tableau.com/profile/c2smart.center#!/vizhome/Home\\_15949538300480/Dashboard1](https://public.tableau.com/profile/c2smart.center#!/vizhome/Home_15949538300480/Dashboard1) (acedido em 12/02/2021) (ver p. 21).
- [44] S. Team. URL: <https://www.spyder-ide.org/> (acedido em 20/10/2021) (ver p. 51).
- [45] J. F. Teixeira e M. Lopes. "The link between bike sharing and subway use during the COVID-19 pandemic: The case-study of New York's Citi Bike". en. Em: *Transportation Research Interdisciplinary Perspectives* 6 (jul. de 2020), p. 100166. ISSN: 25901982. DOI: 10.1016/j.trip.2020.100166. URL: <https://linkinghub.elsevier.com/retrieve/pii/S2590198220300774> (ver pp. 14, 26, 27, 68).
- [46] *Thematic map*. Jan. de 2021. URL: [https://en.wikipedia.org/wiki/Thematic\\_map](https://en.wikipedia.org/wiki/Thematic_map) (acedido em 11/02/2021) (ver p. 17).
- [47] C. Tominski e H. Schumann. "Enhanced interactive spiral display". Em: *SIGRAD 2008. The Annual SIGRAD Conference Special Theme: Interaction; November 27-28; 2008 Stockholm; Sweden*. 034. Linköping University Electronic Press. 2008, pp. 53–56 (ver pp. 16, 17).
- [48] *Visual Variables - University of Calgary in Alberta*. URL: [https://pages.cpsc.ucalgary.ca/~saul/hci\\_topics/pdf\\_files/visual-variables.pdf](https://pages.cpsc.ucalgary.ca/~saul/hci_topics/pdf_files/visual-variables.pdf) (acedido em 07/01/2022) (ver p. 13).
- [49] N. Yau. *Where People Run in Major Cities*. Jul. de 2020. URL: <https://flowingdata.com/2014/02/05/where-people-run/> (acedido em 12/02/2021) (ver p. 20).
- [50] S. Zangari et al. "Air quality changes in New York City during the COVID-19 pandemic". en. Em: *Science of The Total Environment* 742 (nov. de 2020), p. 140496. ISSN: 00489697. DOI: 10.1016/j.scitotenv.2020.140496. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0048969720340183> (ver p. 35).

## LISTA DE *DASHBOARDS* DESENVOLVIDOS

Neste apêndice é possível observar uma listagem de todos os *dashboards* desenvolvidos no âmbito desta dissertação.

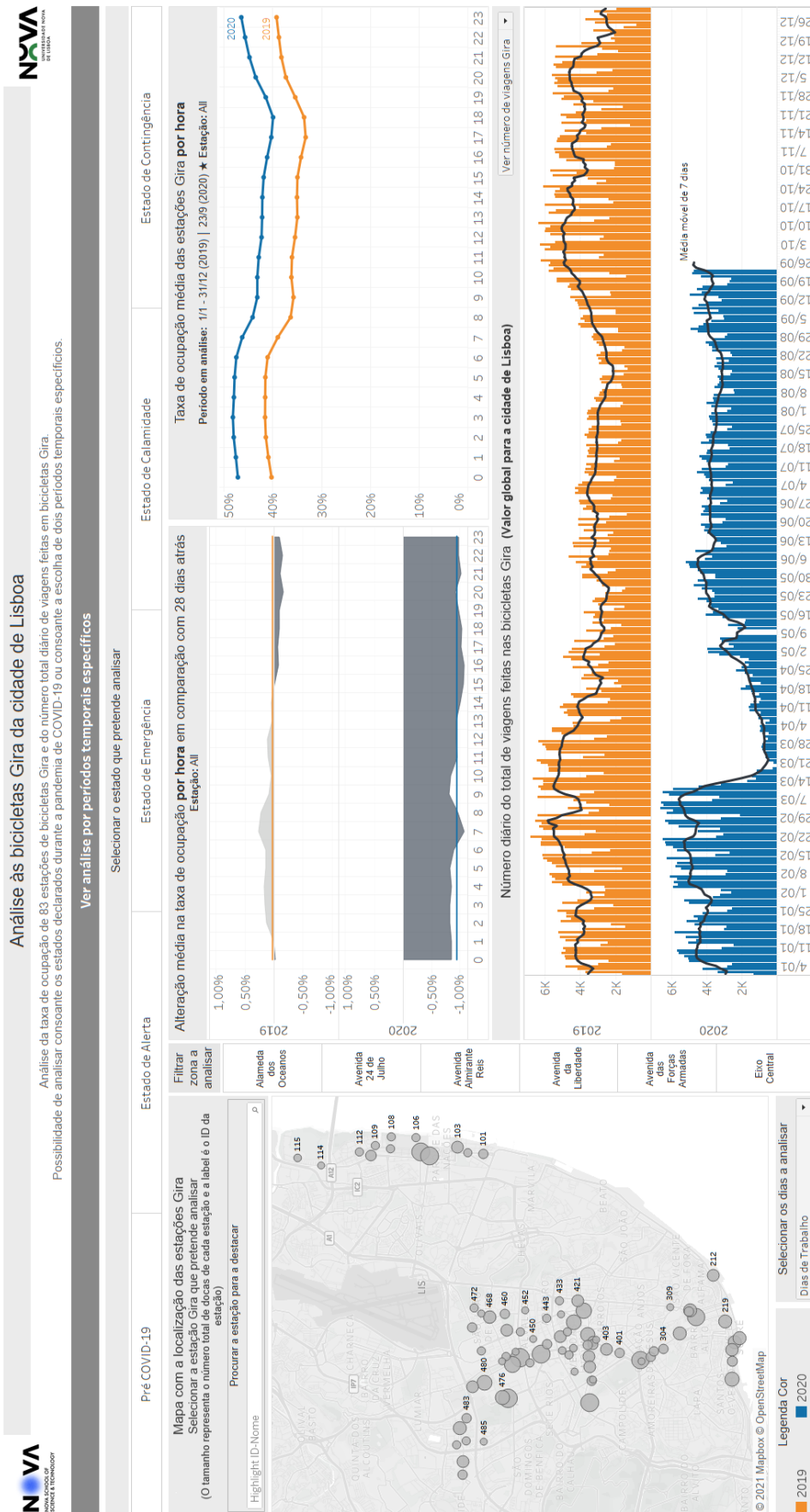


Figura A.1: *Dashboard* para a análise às bicicletas Gira da cidade de Lisboa - Análise por estados declarados durante a pandemia.



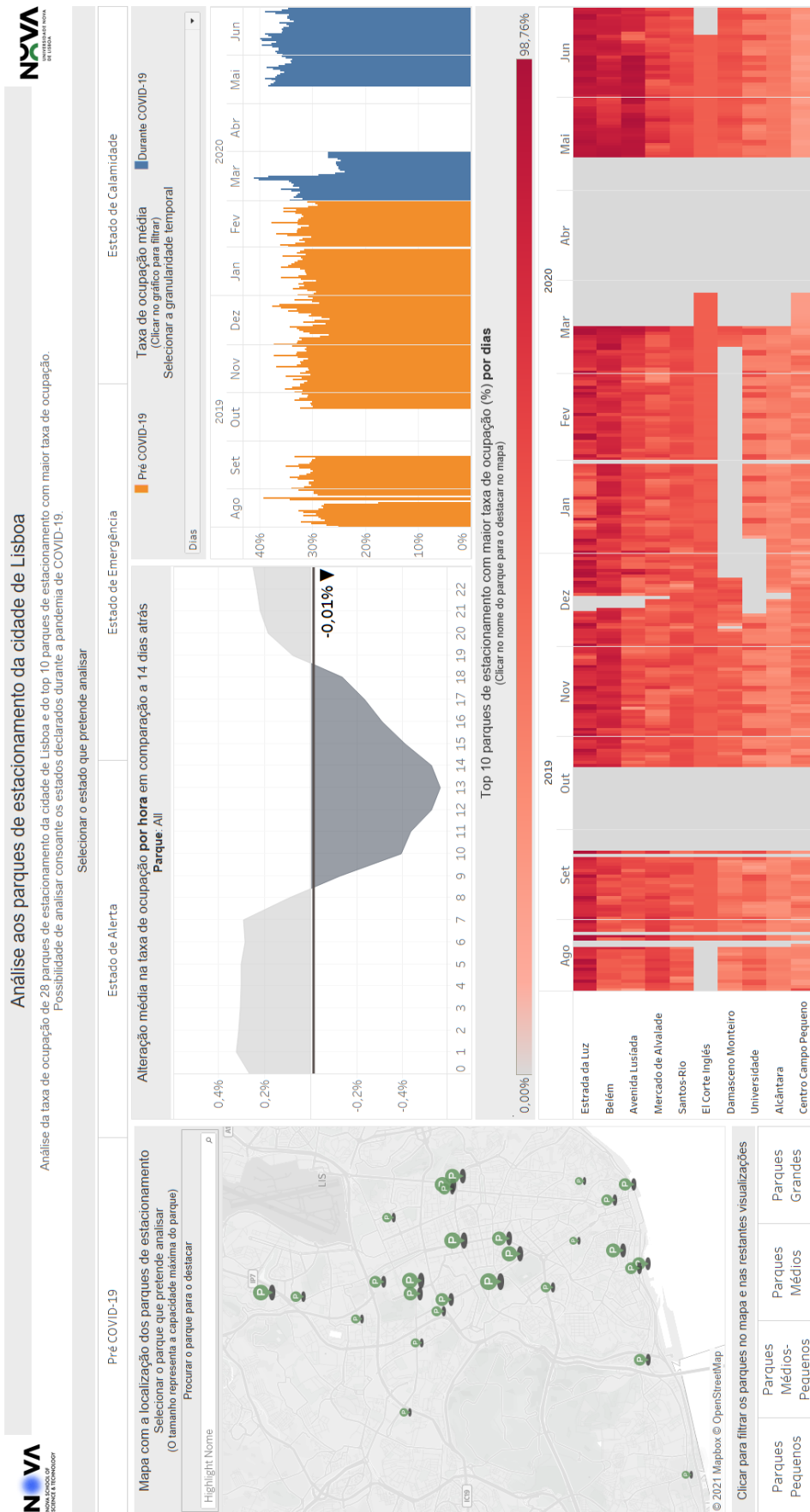


Figura A.3: *Dashboard* para a análise aos parques de estacionamento da cidade de Lisboa.

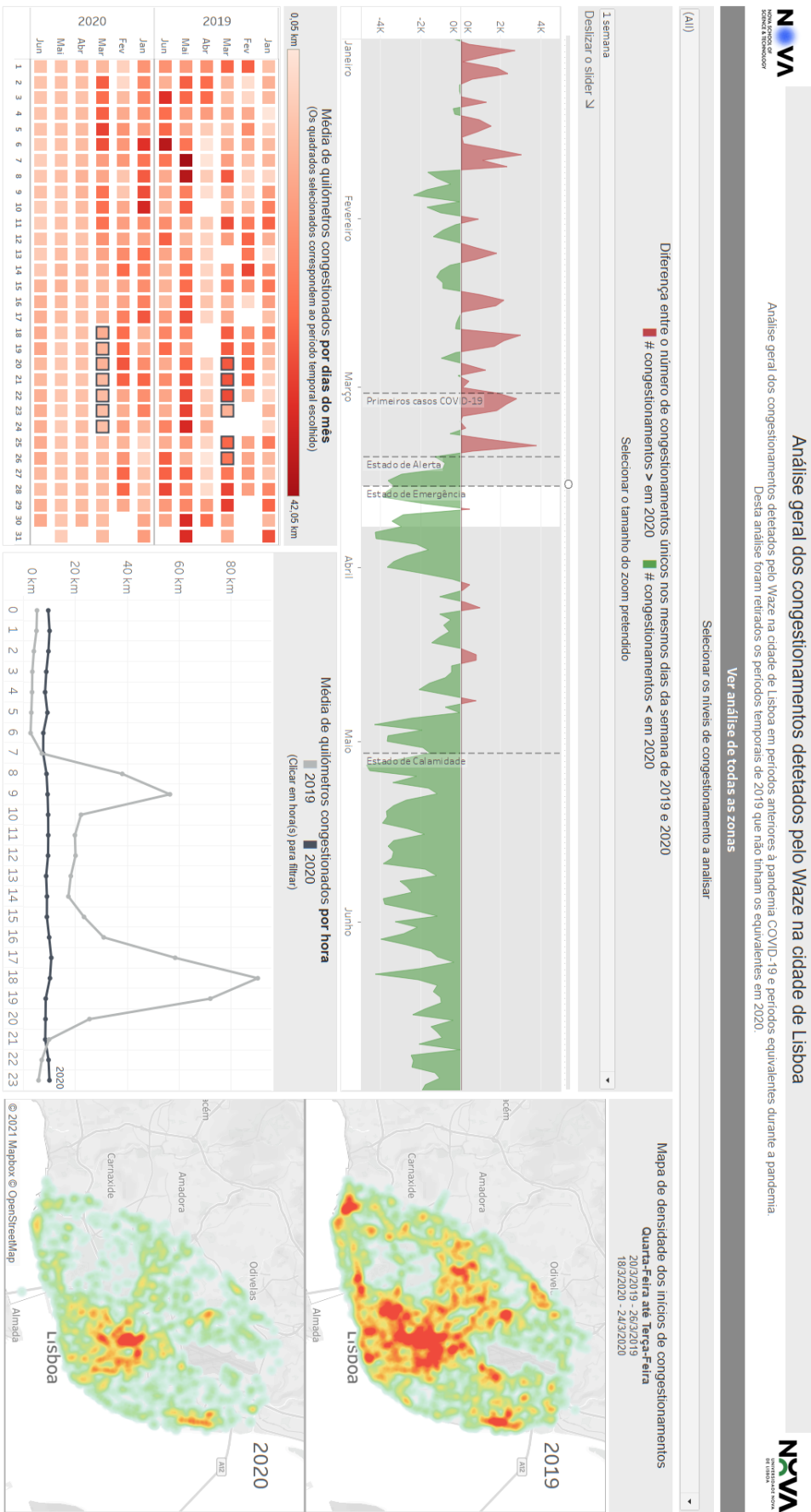


Figura A.4: Dashboard para a análise geral aos congestionamentos detetados pelo Waze na cidade de Lisboa.

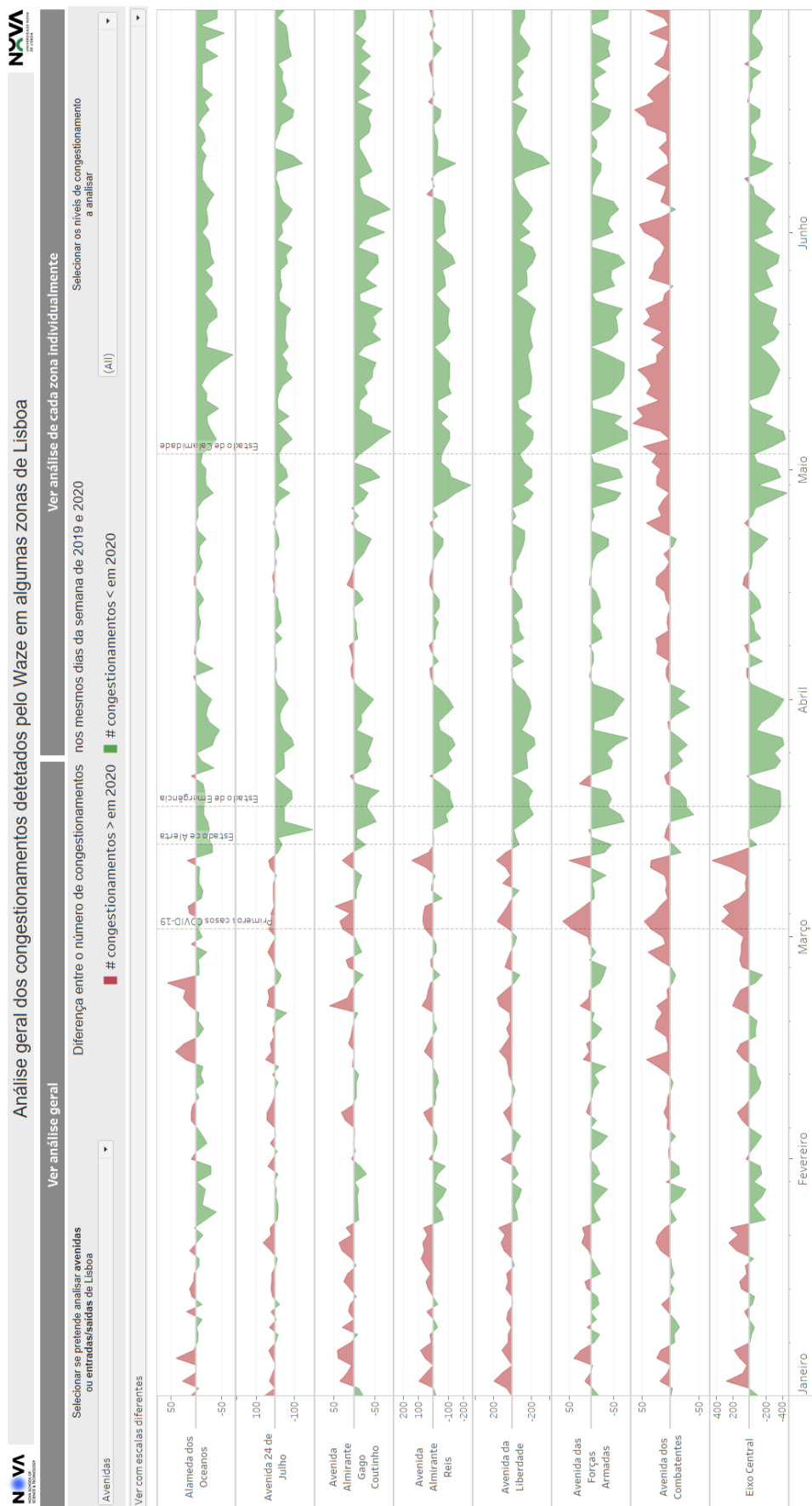


Figura A.5: Dashboard para a análise geral aos congestionamentos detetados pelo Waze em algumas zonas da cidade de Lisboa.



Figura A.6: Dashboard para a análise dos congestionamentos detetados pelo Waze numa zona específica da cidade de Lisboa.

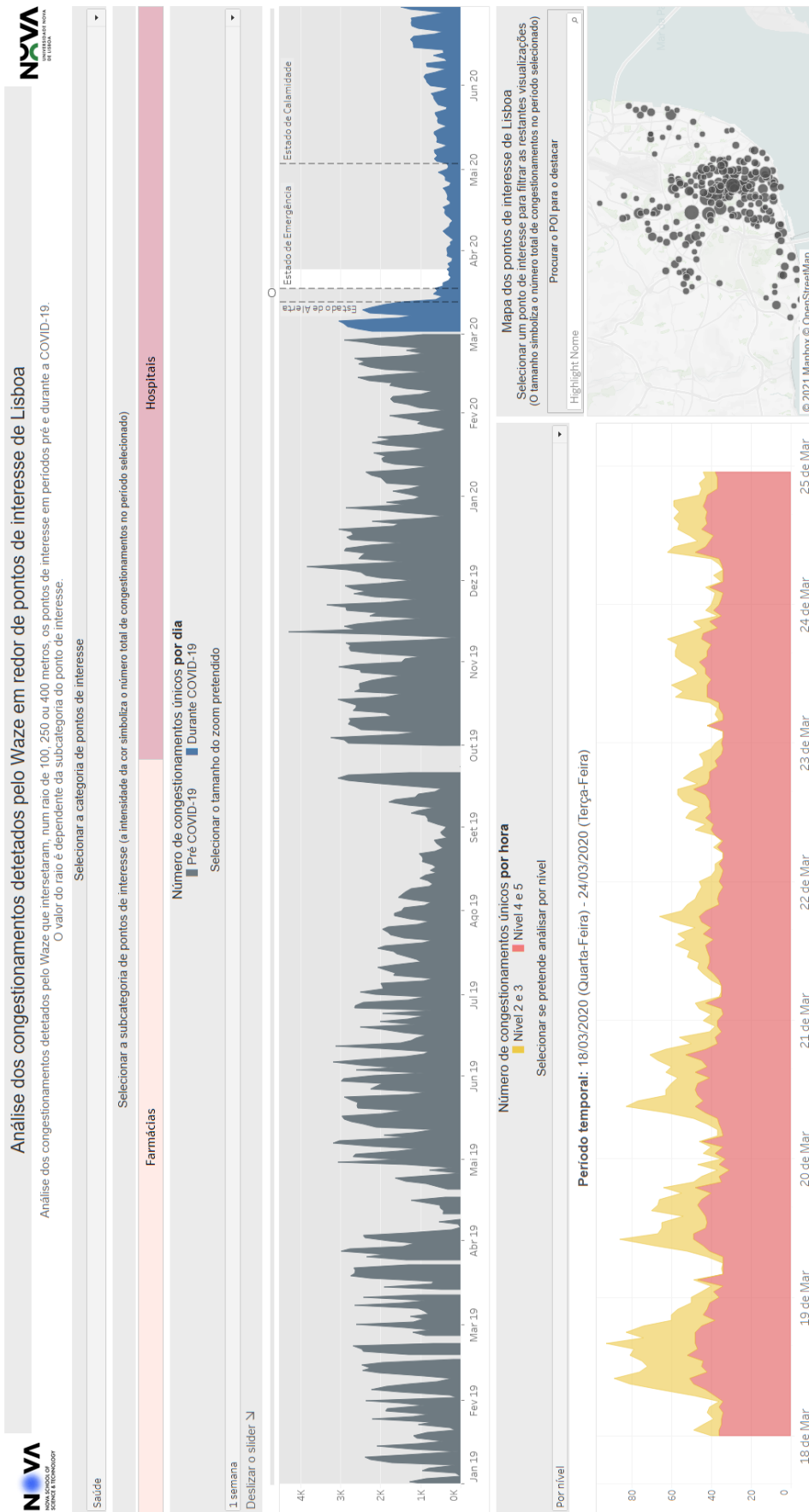


Figura A.7: *Dashboard* para a análise dos congestionamentos detetados pelo *Waze* ao redor de pontos de interesse da cidade de Lisboa.



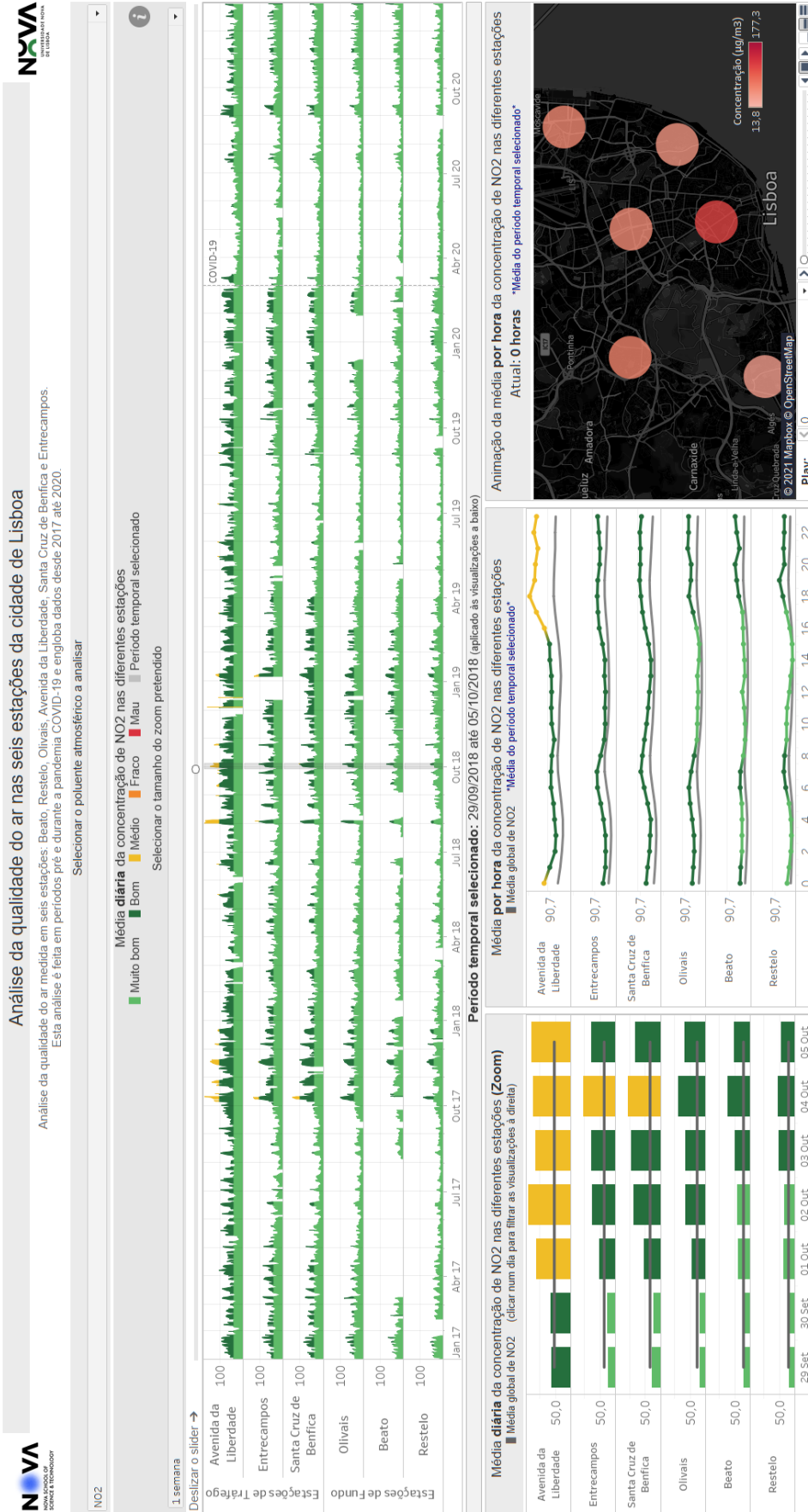


Figura A.9: Dashboard para a análise da qualidade do ar nas seis estações da cidade de Lisboa.



Figura A.10: *Dashboard* para a comparação entre o comportamento dos poluentes atmosféricos medidos nas estações de qualidade do ar da cidade de Lisboa.

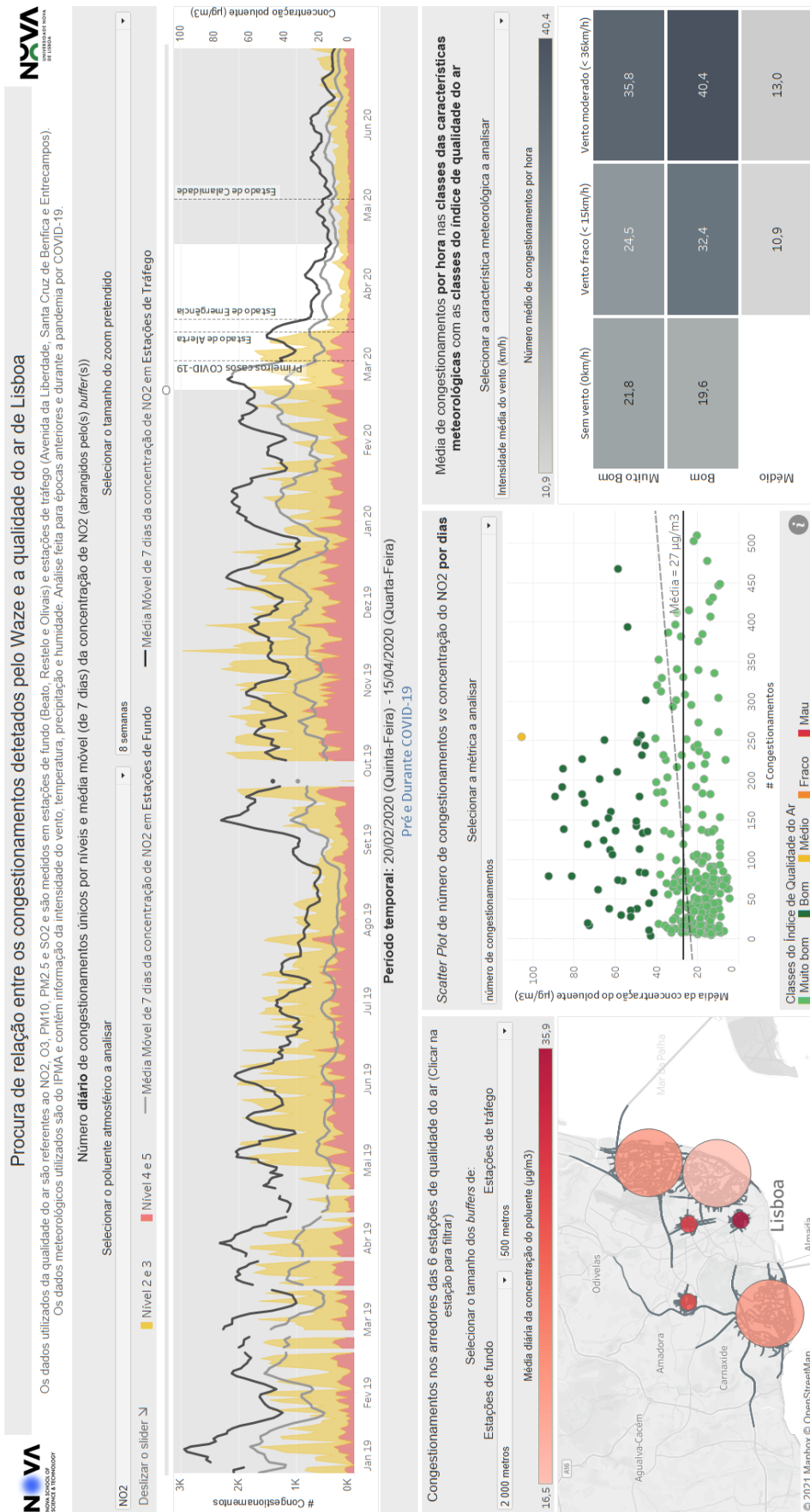


Figura A.11: Dashboard para a procura de relação entre os congestionamentos detetados pelo Waze e a qualidade do ar de Lisboa.

## QUESTIONÁRIO DOS TESTES EXPERIMENTAIS

- **Bicicletas Gira (estados da pandemia):**
  - **SQ0101. A taxa de ocupação ao longo das horas, na estação 103, é semelhante em 2019 e 2020 durante a altura do estado de emergência?**
    - A) Sim
    - B) Não ✓
  
  - **SQ0102. Relativamente ao estado de contingência, considera que, durante os dias de trabalho, houve alteração da hora com maior taxa de ocupação nas estações pertencentes à zona do eixo central?**
    - A) Sim, em 2019 a hora de maior taxa de ocupação era às 9h e em 2020 às 19h
    - B) Não, em ambos os anos a hora de maior taxa de ocupação é às 10h
    - C) Sim, em 2019 a hora de maior taxa de ocupação era às 17h e em 2020 às 9h
    - D) Não, em ambos os anos a hora de maior taxa de ocupação é às 9h ✓
  
  - **SQ0103. Durante os estados da pandemia houve alguma alteração significativa no número diário de viagens feitas nas bicicletas da rede Gira?**
    - A) Sim, houve um decréscimo significativo em 2020 durante o estado de alerta e de emergência ✓
    - B) Sim, houve um decréscimo em 2020 durante o estado de calamidade
    - C) Sim, houve um decréscimo em 2020 apenas durante o estado de emergência
    - D) Não houve alterações

- 
- **Parques de estacionamento:**
    - **SQ0201. O parque de estacionamento da Avenida Lusíada ficou, na sua maioria, mais cheio ou mais vazio após o aparecimento da COVID-19?**
      - A) Ficou, maioritariamente, mais vazio
      - B) Ficou, maioritariamente, mais cheio ✓
      - C) Não apresentou mudanças de comportamento
  
    - **SQ0202. Relativamente à alteração média da taxa de ocupação por hora comparativamente com 14 dias atrás, qual é o comportamento observado no parque Avenida Lusíada durante o mês de junho?**
      - A) Houve um aumento geral, mas principalmente desde as 8h até às 13h
      - B) Houve um aumento geral, mas principalmente desde as 20h até às 22h
      - C) Houve uma diminuição geral, mas principalmente desde as 8h até às 22h ✓
      - D) Não apresentou mudanças de comportamento
  
  - **Congestionamentos (geral)**
    - **SQ0301. No geral, houve mais ou menos congestionamentos em 2020 na altura da COVID-19?**
      - A) Menos, começou a reduzir logo após o aparecimento da COVID-19
      - B) Mais, começou a aumentar após a declaração do estado de alerta
      - C) Mais, começou a aumentar após a declaração do estado de emergência
      - D) Menos, começou a reduzir após a declaração do estado de alerta ✓
  
    - **SQ0302. Comparativamente a 2019, considera que a densidade espacial dos congestionamentos na cidade de Lisboa se alterou durante as duas primeiras semanas do estado de calamidade?**
      - A) Sim ✓
      - B) Não
  
    - **SQ0303. Em 2020, durante as duas primeiras semanas do estado de emergência existiram picos horários da média de quilómetros congestionados?**
      - A) Sim, houve um pico às 18h de 78km
      - B) Sim, houve um pico às 14h de 10km
      - C) Não, manteve-se constante ao longo das horas, com valores inferiores a 20km ✓
      - D) Não, manteve-se constante ao longo das horas, com valores superiores a 40km

- **Congestionamentos (todas as zonas):**
  - **SQ0401. Quais foram as três avenidas de Lisboa onde se sentiu um maior impacto da COVID-19 em relação ao número de congestionamentos?**
    - A) Alameda dos Oceanos
    - B) Avenida 24 de Julho
    - C) Avenida Almirante Gago Coutinho
    - D) Avenida Almirante Reis ✓
    - E) Avenida da Liberdade ✓
    - F) Avenida das Forças Armadas
    - G) Avenida dos Combatentes
    - H) Eixo Central ✓
  
- **Congestionamentos (zona específica):**
  - **SQ0402. Existiram mudanças significativas no número de congestionamentos diários ocorridos na Avenida da Liberdade após a declaração do estado de alerta?**
    - A) Sim, houve uma diminuição comparativamente aos dias anteriores e um aumento comparativamente a 2019
    - B) Sim, houve uma diminuição comparativamente aos dias anteriores e uma diminuição comparativamente a 2019 ✓
    - C) Sim, houve um aumento comparativamente aos dias anteriores e um aumento comparativamente a 2019
    - D) Não, o comportamento manteve-se sempre igual
  
- **Congestionamentos (POI):**
  - **SQ0501. Como se comportaram os congestionamentos em redor dos pontos de interesse da categoria “Saúde” na cidade de Lisboa após o aparecimento da COVID-19?**
    - A) Houve uma redução dos congestionamentos apenas a partir do estado de calamidade
    - B) Houve um aumento dos congestionamentos a partir do estado de emergência
    - C) Houve uma redução dos congestionamentos a partir do estado de alerta ✓
    - D) Não houve mudança no comportamento

---

– SQ0503. Durante a primeira semana do estado de emergência, qual foi o dia que apresentou um maior número de congestionamentos ao redor do Hospital Santa Maria?

- A) Dia 21/03/2020 com 30 congestionamentos ✓
- B) Dia 24/03/2020 com 50 congestionamentos
- C) Dia 18/03/2020 com 13 congestionamentos
- D) Dia 19/03/2020 com 27 congestionamentos

• Qualidade do ar:

– SQ0601. Quais foram os três meses que, na estação da Avenida da Liberdade, apresentaram uma maior mudança no valor médio do poluente  $\text{NO}_2$  durante o ano de 2020, comparativamente aos anos anteriores?

- A) Janeiro, fevereiro e março
- B) Abril, maio e junho ✓
- C) Setembro, outubro e novembro
- D) Não existiram mudanças de comportamento entre os meses

– SQ0602. Para o conjunto de estações da Avenida da Liberdade e Entrecampos, qual foi o estado da pandemia que parece ter tido um maior impacto nos valores médios diários de  $\text{NO}_2$  em 2020?

- A) Principalmente o estado de emergência ✓
- B) Principalmente o estado de alerta
- C) Principalmente o 2º estado de emergência
- D) Principalmente o estado de contingência

– SQ0603. Relativamente à estação de Entrecampos e durante o 1º estado de emergência, considera que a distribuição das concentrações de  $\text{NO}_2$  pelas classes do índice de qualidade do ar foram idênticas nos quatro anos?

- A) Sim, as distribuições são idênticas nos 4 anos
- B) Não, em 2020 houve uma diminuição das horas com concentrações muito boas
- C) Não, em 2020 houve um aumento das horas com concentrações fracas
- D) Não, em 2020 houve um aumento das horas com concentrações muito boas ✓

- **Qualidade do ar (por estações):**
  - **SQ0702. Relativamente a quatro semanas a partir do dia 4 de abril de 2020, qual é a estação que mais diverge da sua média de concentração global de NO<sub>2</sub>?**
    - A) Avenida da Liberdade ✓
    - B) Entrecampos
    - C) Santa Cruz de Benfica
    - D) Olivais
    - E) Beato
    - F) Restelo
  
- **Comparação entre poluentes:**
  - **SQ0801. Após o aparecimento da COVID-19 e numa perspetiva diária, qual foi a classe predominante do índice de qualidade do ar em todos os poluentes medidos nas diversas estações?**
    - A) Classe: Mau
    - B) Classe: Fraco
    - C) Classe: Médio
    - D) Classe: Bom
    - E) Classe: Muito Bom ✓
  
  - **SQ0802. Analisando as quatro primeiras semanas após o aparecimento da COVID-19, quais foram os poluentes que, em média, apresentaram um índice de qualidade do ar menos bom? E em que estações é que isso se verificou?**
    - A) NO<sub>2</sub> e O<sub>3</sub> na estação de Olivais
    - B) PM<sub>10</sub> na estação da Avenida da Liberdade e PM<sub>10</sub> e PM<sub>2.5</sub> na estação de Olivais ✓
    - C) NO<sub>2</sub> na estação da Avenida da Liberdade e PM<sub>10</sub> e PM<sub>2.5</sub> na estação do Beato
    - D) O<sub>3</sub> e PM<sub>10</sub> na estação de Entrecampos

---

– **SQ0803.** Analisando as quatro primeiras semanas após o aparecimento da COVID-19, considera que existe alguma correlação entre os poluentes **NO<sub>2</sub>** e **O<sub>3</sub>** medidos na estação de Entrecampos?

- A) Sim, quando os valores de NO<sub>2</sub> baixam os valores de O<sub>3</sub> sobem ✓
- B) Sim, quando os valores de NO<sub>2</sub> sobem os valores de O<sub>3</sub> sobem
- C) Sim, quando os valores de NO<sub>2</sub> baixam os valores de O<sub>3</sub> baixam
- D) Não existe correlação

• **Congestionamentos vs qualidade do ar:**

– **SQ0901.** Após a declaração do estado de emergência, qual foi o comportamento geral observado no número diário de congestionamentos e na média móvel da concentração do **NO<sub>2</sub>**, comparativamente a períodos temporais anteriores?

- A) O número de congestionamentos subiu e a concentração de NO<sub>2</sub> desceu
- B) O número de congestionamentos desceu e a concentração de NO<sub>2</sub> subiu
- C) Ambos subiram
- D) Ambos desceram ✓

