INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Impact of COVID-19 Pandemic Restrictions in Urban Mobility and Air Pollution in Lisbon, Portugal

Alexandre Sadik Vieira Juma

Master Degree in Computer Engineering

Supervisor:
PhD Luís Miguel Martins Nunes, Associate Professor,
ISCTE-IUL

Co-Supervisor:
PhD João Carlos Amaro Ferreira, Assistant Professor with Habilitation,
ISCTE-IUL

October, 2021

TECNOLOGIAS
E ARQUITETURA

Department of Information Science and Technology

Impact of COVID-19 Pandemic Restrictions in Urban Mobility and Air Pollution in Lisbon, Portugal

Alexandre Sadik Vieira Juma

Master Degree in Computer Engineering

Supervisor:
PhD Luís Miguel Martins Nunes, Associate Professor,
ISCTE-IUL

Co-Supervisor:
PhD João Carlos Amaro Ferreira, Assistant Professor with Habilitation,
ISCTE-IUL

October, 2021

Acknowledgements

To my dissertation coordinators, Prof. Dr. Luís Nunes and Prof. Dr. João Ferreira, I would like to thank you for all your guidance and support throughout this journey.

I'd like to thank my family and friends who've supported me in many different ways during these two intense years I've taken to complete my MSc.

To the course colleagues with whom I've had the chance to work closely for two years during my MSc, I'd like to thank your availability, commitment, and excellence in achieving our common goals, as well as for the good times despite the long hours.

I would like to thank the entities that contributed with the required data and responded to both business and technical enquiries that helped me to complete this research work, namely CML/EMEL, IPMA, APA, CCDR-LVT, EEA, ESA, Metropolitano de Lisboa and ANA Aeroportos de Portugal.

I would also like to acknowledge my appreciation for the Lisbon urban mobility data provided by private enterprises, such as Google, Apple, Moovit and Waze, for COVID-19 pandemic related research.

To all that I enumerate my sincere "Thank you".

This work was partly funded through national funds by FCT - Fundação para a Ciência e Tecnologia, I.P. under project UIDB/04466/2020 (ISTAR).

Resumo

O presente trabalho relata os impactos na mobilidade urbana e qualidade do ar em Lisboa, Portugal, como consequência das restrições impostas para conter a transmissão do vírus SARS-CoV-2, causador da doença COVID-19, onde durante o primeiro período de emergência nacional (18-03-2020 a 03-05-2020) as reduções acentuadas nas atividades antropogénicas, nomeadamente o tráfego rodoviário, resultaram na redução generalizada das concentrações dos principais poluentes atmosféricos medidos nas seis estações de monitorização da qualidade do ar em Lisboa quando comparados ao período homólogo de 2013-2019, sendo o NO_2 o poluente atmosférico mais impactado com uma redução média de 54.35% nas estações de tráfego e 28.62% nas estações de fundo. Uma exceção a esta tendência foi o aumento observado na concentração de O_3 de 12.89% nas estações de tráfego potencialmente devido a mudanças na relação NO_x:COV e redução da ação de redução de O_3 por reação com NO como resultado da redução acentuada da concentração de NO_x nas zonas habitualmente mais poluídas da cidade. Este fenómeno reforça a necessidade de medidas que mitiguem o aumento da poluição de O_3 no âmbito do plano de melhoria da qualidade do ar de Lisboa e Vale do Tejo que visa a redução das concentrações de NO_2, nomeadamente medidas específicas de gestão de COV. O indicador de mobilidade da Google para o comércio local em Lisboa foi identificado como a atividade antropogénica mais relevante com uma correlação moderada e positiva com a concentração NO_2 (r=+0.54). A velocidade média do vento foi identificada como a atividade natural mais relevante com uma correlação moderada e negativa com a concentração NO_2 (r=-0.53). Foi treinada uma *ML pipeline* para prever a concentração NO_2 que teve como entradas os dados de atividade antropogénica, meteorológica e qualidade do ar desde Março/2020 a Março/2021, obtendo R^2=0.925 no conjunto de teste. A análise de importância dos atributos identificam as variáveis antropogénicas como responsáveis por 41.19% da concentração NO_2 enquanto que as variáveis naturais respondem por 58.81%.

Palavras-Chave: Mobilidade urbana; qualidade do ar; COVID-19; aprendizagem automática

Abstract

The present work reports the impacts on urban mobility and air quality in Lisbon, Portugal as a consequence of the imposed restrictions to curb the transmission of SARS-CoV-2 virus which causes COVID-19 disease. During the first national emergency period (18-03-2020 to 03-05-2020) the sharp reductions in anthropogenic activities, most importantly road traffic, resulted in generally reduced criteria air pollutant concentration when compared to a homologous baseline from 2013-2019 measured in the six air quality monitoring stations throughout the city. The most negatively impacted air pollutant was NO_2 with a reduction of 54.35% in traffic stations and 28.62% reduction in background stations. An exception to this trend was the observed O_3 concentration increase of 12.89% in traffic stations which is potentially due to changes in the NO_x:VOC ratio and reduced O_3 titration by NO as a result of sharp decrease of NO_x emissions in the usually most polluted city hotspots. This phenomenon raises the need of additional measures to mitigate O_3 pollution increases as part of the Lisbon and Tagus Valley air quality improvement plan which aims to reduce NO_2 concentrations, namely specific measures for VOC management. Google mobility indicator for local commerce was found to be the main anthropogenic activity indicator for Lisbon with a moderate and positive correlation with NO_2 concentration (r=+0.54), whereas the average wind speed was the most relevant natural phenomena contributing to NO_2 concentration with a moderate and negative correlation (r=-0.53). A regressor ML pipeline was trained to predict NO_2 concentration with the available anthropogenic activity, weather, and air pollutant inputs from March/2020 to March/2021, achieving R^2=0.925 on the test set and subsequent feature importance analysis uncovered that anthropogenic features contribute to 41.19% of NO_2 concentrations and natural phenomena features contribute to 58.81%.

Keywords: urban mobility; air quality; covid-19; machine-learning; automl

Index

Acknowledgements ..... **i**
Resumo ..... **ii**
Abstract ..... **iii**
Index ..... **iv**
Tables Index ..... **vi**
Figures Index ..... **viii**
List of abbreviations ..... **xi**
Chapter 1 – Introduction **13**
1.1. Topic Context 13
1.2. Motivation and topic relevance 14
1.3. Research Questions 16
1.4. Objectives 17
1.5. Methodologic approach 18
Chapter 2 – Literature review **22**
2.1. Literature review process 23
2.2. Atmospheric chemistry of Nitrogen Oxides and Ozone 24
2.3. Relevant variables for estimating urban air pollutants 27
2.4. Impact of COVID-19 restrictions on urban air quality 29
2.5. Machine Learning based Air Pollutant Concentration Prediction 30
2.6. Literature Review Summary 32
2.7. Differentiated approach 34
Chapter 3 – Analysis and results presentation **35**
3.1. Business Understanding 35
3.2. Available data sources 38
3.2.1. Lisbon Ground-based Meteorological Data Source 39
3.2.2. Lisbon Mobility Indexes Data Sources 40
3.2.3. Lisbon Subway Ridership Data Source 41
3.2.4. Lisbon International Airport Traffic Data Source 42
3.2.5. Lisbon Air Pollutant Concentration Data Source 42
3.2.6. Portugal COVID-19 Pandemic Report 44
3.2.7. Lisbon and Tagus Valley Mortality Report 45
3.2.8. Bike Lane Counter (Av. Duque de Ávila) 45
3.2.9. Shared Bike Trips (GIRA) 46
3.2.10. Data Source Summary 47

3.2.11. Consolidated Data Model 48
3.3. Temporal and Spatial Scope of the Analysis 48
3.4. Mobility Analysis during COVID-19 Pandemic 51
3.5. Nitrogen Dioxide (NO2) performance during COVID-19 Pandemic..... 61
3.6. Tropospheric Ozone (O3) performance during COVID-19 Pandemic..... 69
3.7. Additional criteria pollutants and overall impact to air pollution..... 77
3.8. Relevant variable bivariate analysis..... 82
3.9. Modeling NO2 Concentration with AutoML 91
3.10. Anthropogenic and Meteorological contributions to NO2 Concentration . 103
Chapter 4 – Conclusions and future work 111
4.1. Main conclusions 111
4.2. Research limitations..... 113
4.3. Future research proposals 113
References..... 116
Annexes and Appendices 122
Appendix A – Data tables related to figures..... 122
Appendix B – Data Source Metadata 130

Tables Index

Table 1 – Literature Review Summary of Main Articles 32
Table 2 – WHO Air Pollutant Safety Threshold Guidelines and EU regulated air quality thresholds for air pollutants in scope for this work 37
Table 3 –EU regulated air quality thresholds for public announcement 38
Table 4 – Approached entities, requested data and request success 38
Table 5 – Data Sources 47
Table 6 – List of meteorological and air quality monitoring station 49
Table 7 – Description of the main phases of the pandemic management in the Lisbon Metropolitan Area 50
Table 8 – List of data features 93
Table 9 – Performance metrics of the Initial exploratory NO2 prediction pipeline on the test set 95
Table 10 – NO2 concentration interpretable model XGBRegressor optimized hyperparameters 100
Table 11 – Performance metrics of the NO2 concentration interpretable model on the test set 100
Table 12 – NO2 causes model XGB Regressor Hyperparameters found by T-POT 103
Table 13 – Performance metrics of the NO2 cause model on the test set 104
Table 14 – All Lisbon stations monthly median NO2 concentration (µg/m3) for the years 2013-2019 and 2020. 122
Table 15 –Road traffic Lisbon monitoring stations monthly median NO2 concentration (µg/m3) for the years 2013-2019 and 2020. 122
Table 16 – Background Lisbon monitoring stations monthly median NO2 concentration (µg/m3) for the years 2013-2019 and 2020. 123
Table 17 – Hourly NO2 median concentration in Traffic Stations for the National Confinement period (14-03-2020 to 03-05-2020) and homologous 2013-2019 period. 123
Table 18 – Hourly NO2 median concentration in Background Stations for the National Confinement period (14-03-2020 to 03-05-2020) and homologous 2013-2019 period. 124
Table 19 – NO2 median concentration in Traffic Stations per day of the week during the National Emergency period and Calamity State (14-03-2020 to 01-07-2020) and homologous 2013-2019 period. 124
Table 20 – NO2 median concentration in Background Stations per day of the week during the National Emergency period and Calamity State (14-03-2020 to 01-07-2020) and homologous 2013-2019 period. 125
Table 21 – NO2 annual average NO2 concentration in all Lisbon stations from 2013 to 2020. 125
Table 22 –Road traffic Lisbon monitoring stations monthly median O3 concentration (µg/m3) for the years 2013-2019 and 2020. 125
Table 23 –Background Lisbon monitoring stations monthly median O3 concentration (µg/m3) for the years 2013-2019 and 2020. 126
Table 24 – Hourly O3 median concentration in Traffic Stations for the National Confinement period (14-03-2020 to 03-05-2020) versus 2013-2019 Interquartile range, for the same year period. 126
Table 25 – Hourly O3 median concentration in Background Stations for the National Confinement period (14-03-2020 to 03-05-2020) versus 2013-2019 Interquartile range, for the same year period. 127

Table 26 – O3 median concentration in Traffic Stations per day of the week during the National Emergency period and Calamity State (14-03-2020 to 01-07-2020) versus 2013-2019 Interquartile range, for the same year period. 127
Table 27 – O3 median concentration in Background Stations per day of the week during the National Emergency period and Calamity State (14-03-2020 to 01-07-2020) versus 2013-2019 Interquartile range, for the same year period. 128
Table 28 – Traffic Lisbon monitoring stations monthly median CO concentration (μg/m3) for the years 2013-2019 and 2020. 128
Table 29 – Background Lisbon monitoring stations monthly median CO concentration (μg/m3) for the years 2013-2019 and 2020. 128
Table 30 – Traffic Lisbon monitoring stations monthly median PM2.5 concentration (μg/m3) for the years 2013-2019 and 2020. 129
Table 31 – Background Lisbon monitoring stations monthly median PM2.5 concentration (μg/m3) for the years 2013-2019 and 2020. 129
Table 32 – Traffic Lisbon monitoring stations monthly median PM10 concentration (μg/m3) for the years 2013-2019 and 2020. 130
Table 33 – Background Lisbon monitoring stations monthly median PM10 concentration (μg/m3) for the years 2013-2019 and 2020. 130
Table 34 – IMPA Meteorology data source metadata 130
Table 35 – APA Air Pollution data source metadata 131
Table 36 – DGS COVID-19 dataset metadata 131
Table 37 – Apple Mobility Trend Report data source metadata 134
Table 38 –Moovit Insights Public Transit Index data source metadata 134
Table 39 – Google COVID-19 Community Mobility Report data source metadata 134
Table 40 – Waze COVID-19 Impact data source metadata 134
Table 41 – Lisbon international airport traffic data source metadata 135
Table 42 – Lisbon Subway ridership data source metadata 135
Table 43 – Death certificates data source metadata 135
Table 44 – Bike Lane counter (Av. Duque de Ávila) data source metadata 135
Table 45 – GIRA Trips data source metadata 135

Figures Index

Figure 1 – Pasteur Quadrant [29] 18
Figure 2 – CRISP-DM Methodology Phases and Flow [30] 19
Figure 3 – Air Pollutant lifecycle schematic [30] 22
Figure 4 – Peak ozone isopleths generated from initial mixtures of VOC and NOx [31] 25
Figure 5 – Generalized NOx – O3 lifecycle 26
Figure 6 – Schematic of chemical and transport processes related to atmospheric composition. These processes link the atmosphere with other components of the Earth system, including the oceans, land, and terrestrial and marine plants and animals [48] 36
Figure 7 – Lisbon worst station yearly average concentration per air pollutant 44
Figure 8 – Simplified analytical data model built in the present work 48
Figure 9 – Lisbon municipality geography boundaries, its reduced pollution emission zones and air pollution, weather monitoring stations. 49
Figure 10 – Main phases of the pandemic management in the Lisbon Metropolitan Area. 50
Figure 11 – Moovit Insights Public Transit Index for Lisbon, 2020 52
Figure 12 – Apple Mobility Trend Report for Lisbon, 2020 53
Figure 13 – Waze COVID-19 Impact Index for Lisbon, 2020 54
Figure 14 – Google Community Mobility Report for Lisbon, 2020 55
Figure 15 – Lisbon Subway Ridership for 2019 and 2020 56
Figure 16 – Lisbon Subway (a) YoY monthly percentual ridership change 2019 and 2020 and (b) monthly percentual ridership change against January 2020 57
Figure 17 – Lisbon GIRA shared-bike ridership for 2019 and 2020 58
Figure 18 – Lisbon international airport non-transferred passengers for 2019 and 2020 59
Figure 19 – Moovit and Google public transportation 2020 Mobility Indexes against an Independent Variable (Subway ridership) 59
Figure 20 – Google, Apple, Moovit and Waze mobility index change during the 1st national emergency state period from 18-03-2020 to 03-05-2020 (1) and first national calamity state period from 04-05-2020 to 01-07-2020 (2) against a pre-pandemic baseline 60
Figure 21 – Lisbon and Tagus Valley Monthly Average of Total NO2 vertical column (μmol/m2) measured by the TROPOMI instrument onboard Sentinel-P5 satellite. 61
Figure 22 – All Lisbon stations monthly median NO2 concentration (μg/m3) for the years 2013-2019 and 2020. 62
Figure 23 –Road traffic Lisbon monitoring stations monthly median NO2 concentration (μg/m3) for the years 2013-2019 and 2020. 63
Figure 24 – Background Lisbon monitoring stations monthly median NO2 concentration (μg/m3) for the years 2013-2019 and 2020. 63
Figure 25 – Road traffic Lisbon monitoring stations NO2 concentration (μg/m3) 14-day Moving Aver-age per day of year for all years from 2013 to 2020. 64
Figure 26 – Hourly NO2 median concentration in Traffic Stations for the National Confinement period (14-03-2020 to 03-05-2020) and homologous 2013-2019 period. 65
Figure 27 – Hourly NO2 median concentration in Background Stations for the National Confinement period (14-03-2020 to 03-05-2020) and homologous 2013-2019 period. 66
Figure 28 – NO2 median concentration in Traffic Stations per day of the week during the National Emergency period and Calamity State (14-03-2020 to 01-07-2020) and homologous 2013-2019 period. 66

Figure 29 – NO2 median concentration in Background Stations per day of the week during the National Emergency period and Calamity State (14-03-2020 to 01-07-2020) and homologous 2013-2019 period 67
Figure 30 – Annual average NO2 concentration in all Lisbon stations from 2013 to 2020 and the current violation threshold in place (40 µg/m3) 68
Figure 31 – Count of yearly 1-hour mean violations (200 µg/m3) per year in Lisbon stations and allowed exceedances per year (18) 68
Figure 32 –Road traffic Lisbon monitoring stations monthly median O3 concentration (µg/m3) for the years 2013-2019 and 2020. 70
Figure 33 –Background Lisbon monitoring stations monthly median O3 concentration (µg/m3) for the years 2013-2019 and 2020. 71
Figure 34 – Hourly O3 median concentration in Traffic Stations for the National Confinement period (14-03-2020 to 03-05-2020) versus 2019 and 2013-2019 Interquartile range, for the same year period. 72
Figure 35 – Hourly O3 median concentration in Background Stations for the National Confinement period (14-03-2020 to 03-05-2020) versus 2019 and 2013-2019 Interquartile range, for the same year period. 73
Figure 36 – O3 median concentration in Traffic Stations per day of the week during the National Emergency period and Calamity State (14-03-2020 to 01-07-2020) versus 2013-2019 Interquartile range, for the same year period. 74
Figure 37 – O3 median concentration in Background Stations per day of the week during the National Emergency period and Calamity State (14-03-2020 to 01-07-2020) versus 2013-2019 Interquartile range, for the same year period. 74
Figure 38 – Count of Long-term O3 concentration violation for human health per year and maximum O3 8hr-average concentration median concentration from 2013 to 2020 76
Figure 39 – Informational O3 threshold violations from 2013 to 2020 77
Figure 40 – Road traffic (1) and Background Lisbon (2) monitoring stations monthly median CO concentration (µg/m3) for the years 2013-2019 and 2020. 78
Figure 41 – Road traffic (1) and Background Lisbon (2) monitoring stations monthly median PM10 concentration (µg/m3) for the years 2013-2019 and 2020. 79
Figure 42 – Road traffic (1) and Background Lisbon (2) monitoring stations monthly median PM2.5 concentration (µg/m3) for the years 2013-2019 and 2020. 80
Figure 43 – Criteria air pollutant median concentrations homologous analysis of the 1st national emergency state period from 18-03-2020 to 03-05-2020 (1) and 1st national calamity situation period from 04-05-2020 to 01-07-2020 (2) against a same-period from 2013 to 2019 baseline for background and traffic stations in Lisbon, Portugal. ... 81
Figure 44 – NO2 and O3 Pearson Correlation Coefficients 82
Figure 45 – Relationship between Waze Mobility Index and NO2 Concentration (14d Moving Averages) 83
Figure 46 – Relationship between Daily %Change in driven KMs (Waze) and Daily Mean NO2 Concentration 84
Figure 47 – Relationship between Wind Intensity and NO2 Concentration (14d Moving Averages) 85
Figure 48 – Relationship between Daily Mean Wind Intensity and Daily Mean NO2 Concentration 86
Figure 49 – Wind Direction and Speed relationship with NO2 and O3 concentrations homologous analysis (2019 vs 2020) in traffic stations for the 1st national emergency state and 1st national calamity state periods (18-03-2020 to 01-07-2020) 87

Figure 50 – Road traffic Entrecampos monitoring station NO2 and O3 concentrations (µg/m3) 14-day Moving Average for 2020 and 2013-2019. Shaded with 2020 Daily Air Temperature Mean 14d Moving Average 88
Figure 51 – Background monitoring stations NO2 and O3 concentrations (µg/m3) 14-day Moving Average for 2020 and 2013-2019 89
Figure 52 – Hourly O3 and NO2 median concentration in Traffic Stations for the National Confinement period (14-03-2020 to 03-05-2020) versus 2013-2019..... 90
Figure 53 – Hourly O3 and NO2 median concentration in Background Stations for the National Confinement period (14-03-2020 to 03-05-2020) versus 2013-2019..... 91
Figure 54 – ML Pipeline steps automated by TPOT AutoML Framework 92
Figure 55 – Initial exploratory AutoML based NO2 prediction performance 96
Figure 56 – Initial exploratory AutoML based NO2 prediction pipeline stages and hyperparameters..... 97
Figure 57 –TPOT AutoML training efficiency 98
Figure 58 – NO2 concentration interpretable model performance on the test set..... 100
Figure 59 – NO2 concentration interpretable model feature importance analysis..... 102
Figure 60 – NO2 cause model performance on the test set..... 104
Figure 61 – Summary of feature importance analysis..... 105
Figure 62 – Summary of feature importance analysis (mean absolute SHAP)..... 107
Figure 63 – SHAP Values for the two most important independent variables (Daily Wind mean intensity and Daily Avg Google Local Commerce Activity Index) 108
Figure 64 – Top 10 contributing variables for the exceptional low NO2 concentration registered in 05-04-2020..... 108
Figure 65 – Top 10 contributing variables for the higher NO2 concentration registered in 12-09-2020 109
Figure 66 – All variables contribution to NO2 concentration registered in a regular day (11-11-2020)..... 110

List of abbreviations

AML – Lisbon Metropolitan Area

APA – Agência Portuguesa do Ambiente (Portuguese Environment Agency)

AVOC – Anthropogenic Volatile Organic Compound

BVOC – Biogenic Volatile Organic Compound

CAMX – Comprehensive Air Quality Model with Extensions

CMAQ – Community Multiscale Air Quality Modelling System

CML – Câmara Municipal de Lisboa (Lisbon City Council)

CO – Carbon Monoxide

CRISP-DM – Cross Industry Standard Process for Data Mining

EC – European Community

EEA – European Environment Agency

EEC – European Economic Community

ETL – Extract, Transform and Load

EU – European Union

IPMA – Instituto Português do Mar e Atmosfera (Portuguese Agency for the Sea and Atmosphere)

LPG - Liquefied petroleum gas

LTV – Lisbon and Tagus Valley

ML – Machine Learning

NO – Nitrogen Oxide

NO2 – Nitrogen Dioxide

NOx – Nitrogen Oxides (NO + NO2)

O3 – Ozone

PM – Particulate Matter

PM10 – Particulate Matter with diameter of less than 10 micrometers

PM2.5 – Particulate Matter with diameter of less than 2.5 micrometers

SARS – Severe Acute Respiratory Syndrome

VOC – Volatile Organic Compound

WHO – World Health Organization

ZER – Reduced Emissions Zone

Chapter 1 – Introduction

1.1. Topic Context

The COVID-19 disease, caused by the SARS-CoV-2 virus, initially detected in Wuhan, China in December 2019, made its way to Europe arriving in the end of January 2020 with the first reported cases in France on the 24th of January 2020. The first cases reported in Portugal on the 2nd of March 2020 happened while some other countries, such as Italy, were already in the first exponential COVID-19 case curve.

The World Health Organization (WHO) declared COVID-19 a pandemic on the 11th of March 2020 and the Portuguese government started imposing pre-emptive restrictive measures on Monday, 16th of March, which made Friday, 13th of March, the last day of in-person work and study for many Portuguese, followed by the formal declaration of the national state of emergency in Portugal on the 18th of March 2020 until the 2nd of May 2020 which introduced a general mandatory confinement to curb the spread of the disease. This resulted in an unprecedent reduction of anthropogenic activities responsible for the emission of several air pollutants.

Immediately after the end of the (first) national emergency state, the national calamity situation was declared where a gradual lockdown easing plan, re-evaluated every two weeks, was executed between the 2nd of May and 15th of June for Lisbon Metropolitan Area (AML), restarting many in-person public, economic, educative, social and cultural activities with increased sanitary rules while some riskier activities remained prohibited like mass public events, nightclubs and public in football stadiums, and telework was no longer mandatory.

The national calamity situation ended on the 1st of July for most of Lisbon Metropolitan Area (19 parishes of 5 municipalities remained in calamity situation due to the high incidence of COVID-19 cases) and gave way for the local contingency situation, while the rest of the country entered the alert situation. The local contingency situation was declared for the remaining Lisbon Metropolitan Area 19 parishes on the 30th of July.

In April 2020, the month with most severe levels of confinement, the consumption of liquid fuels suffered unprecedent homologous reductions [1], -58.6% for Gasoline and -47.0% for Diesel, and electrical energy consumption had an homologous drop of -13.8%, whereas 69% of the production was of renewable sources, 14% was imported and the remaining 17% of non-renewable sources was produced solely with Natural Gas with no

coal being burned since the two existing coal power plants in Portugal were commissioned in 1985 [2]. Changes in economic activity and telework promoted drastic changes in energy consumption [3], namely domestic energy consumption which in April 2020 had a homologous increase of 31%, while industry energy consumption dropped 17% and services dropped 43%. In April 2020 the maritime port of Lisbon registered a decrease of 47.70% in ships docked [4] and Lisbon International Airport registered a drop in 95.88% aircraft movements [5].

Throughout the most severe lockdown periods, government agencies, environmental associations and the press reported exceptional improvement of air quality in Lisbon and the cause was associated to the sharp reduction of anthropogenic pollutant activities. It is of the utmost importance to study the phenomena of air pollution response to variation of anthropogenic activities to better understand and quantify its causal relation and to support the decision process related to environmental policies.

1.2. Motivation and topic relevance

In a typical workday in 2019, Lisbon saw more than half a million road vehicles [6] transiting through the city. Around 370.000 vehicles came in from the periphery [7] adding to the roughly 213.000 vehicles from within the city. In 2019 Lisbon had a Tom-Tom congestion Index of 33% [8], which means that it takes in average 33% more time per trip than in normal traffic conditions, with rush-hour drivers spending an average of 43 extra minutes stuck in traffic per workday, thus increasing vehicle operation time, engine idle time, slower speeds, and lower gear operation, all contributing to increased fuel consumption and therefore pollutant emissions [9].

In Portugal in 2019, of all registered vehicles [10], diesel engines account for 65.47%, gasoline engines for 32.46%, LPG engines for 0.83% and 1,24% for other motorization types where all types of electric vehicles are included [11]. Moreover, in 2019, 62.02% of the Portuguese vehicle fleet is over 10 years [12], whose engines implement worse European emission standards than newer engines [13].

The 2017 Mobility Survey for the Lisbon Metropolitan Area [14] reports that 60.8% of all trips are done using private passenger vehicles (cars and motorbikes) and the occupation ratio of passenger vehicles is 1,60 persons. On the other hand, only 15.8% of

trips are done using public transportation and 23.5% of trips done with soft transportation means from which only 0.5% of trips are done using bicycles.

The low electrification of the Portuguese vehicle fleet, the greater share of diesel engines, the large share of older vehicles (pre-EURO 5/6), high urban road congestion levels and low adoption of public transport make urban air pollution concentrations, namely NO_x and PM pollutants, in an urban setting such as Lisbon a tough phenomenon to tackle [15].

Since 2011, the Lisbon municipality has been implementing a multi-stage roadmap coined Reduced Emission Zones (ZER) [7] to remove older vehicles, which implement worse emission standards, in the most affected downtown areas and main road arteries affected by increased air pollution. Currently in phase 3 since January 2015, all vehicles in ZER Zones 1 (downtown) must adhere to EURO 3 standard and EURO 2 in ZER Zones 2 (Urban ring consisting of 1/3 of the area of Lisbon municipality). As of June 2021, ZER Phase 4 has been postponed [16] due to the COVID-19 pandemic and, when implemented, will implement stricter EURO standards, reduce maximum speeds and implement additional human and electronic means of enforcing the new rules which are currently largely unenforced as regularly reported by the press and environmental associations [17] [18].

Additional efforts by the municipality and government in recent years to reduce urban pollution [19] include expanding bike lanes, investment in mass public transportation systems and soft transportation means, reducing the prices of monthly passes of public transportation, introducing additional BUS lanes and suppressing car lanes, creation of deterrent car parking space near inter-modal public transport stations, creating pedestrian areas, adjusting speed limits and introducing speed radars, introducing parking meters with differentiated prices per zone and resident parking badges, creating a public bike shared mobility program (GIRA), financing the purchase of bikes for personal use and installing charging spots for electric vehicles.

Simultaneously, an estimated increase in motorization rate of 14.89% in AML from 2015 to 2019 [20] for passenger vehicles, resident population increase of 1.27% in AML from 2011 to 2019 [21], with a +3.03% contribution (+69.147 hab.) from peripheric municipalities and -6.07% contribution (-32.925 hab.) from Lisbon municipality, historical increase of usage of private vehicle for pendular movements in AML from

12.7% in 1981 to 53.82% in 2011 [22], act as opposing forces to efforts of air pollutant reduction in the city of Lisbon.

In the Lisbon and Tagus Valley area the main contributing anthropogenic sources of primary air pollutants are predominantly road transport vehicles [23] with NO_x contribution of 63%, CO contribution of 78% and PM_{10} contribution of 62%, as well as minor contribution from other forms of transportation like air and sea transport, which were all gravely impacted by mobility restrictive measures to control the COVID-19 pandemic.

Air pollution in urban settings is a major source of concern due to adverse effects on human health leading to increased respiratory and cardiovascular disease development and premature death. Other impacts of air pollution include increases in sick leaves for workers and students, lower standards of living for vulnerable groups, such as asthmatic and elderly citizens, as well as added cost burden to public health systems. WHO estimates 4.2 million yearly deaths were related to air pollution in 2016 [24], EEA estimates that 6.690 premature deaths in Portugal (2018) are caused due to three main air pollutants (NO_2, $PM_{2.5}$ and O_3) [25] and an 858 city study in Europe for the year of 2015 estimates that 1.837 premature yearly deaths in Lisbon Metropolitan Area are related to two main air pollutants (NO_2, $PM_{2.5}$) [26], ranking 116th worst position for NO_2 related premature deaths and 514th worst position for $PM_{2.5}$ related premature deaths.

The COVID-19 pandemic management by the Portuguese government and local Lisbon Metropolitan Area authorities included highly restrictive measures that severely reduced urban mobility in Lisbon for all transportation forms and for a large period of time. It is a once in a lifetime opportunity to measure and study the effects of reduced pollutant anthropogenic activities in an urban setting, which were previously only possible to simulate using atmospheric chemistry simulation techniques, in order to aid the definition and prioritization of air pollution reduction policies.

1.3.Research Questions

The main questions that should be answered by this work are:

- How did primary and secondary air pollutant concentrations in Lisbon City change throughout the COVID-19 pandemic period?

- How did urban mobility patterns change in Lisbon City throughout the COVID-19 pandemic period in response to COVID-19 pandemic related restrictions?
- Is it possible to train a machine-learning model that can correctly predict concentration values for an air pollutant with the available data?
- Can this machine-learning model identify the variable(s) that contribute the most to air pollutant concentration values and extrapolate the anthropogenic contributions to the air pollutant concentration from this model?

1.4.Objectives

The performance of the most relevant criteria primary and secondary air pollutants (NO_2, CO, C_6H_6, O_3, PM_{10}, $PM_{2.5}$ and SO_2) will be analyzed in depth for the Lisbon municipality before, during and after the main COVID-19 pandemic management milestones, namely confinement periods where anthropogenic activity was severely affected, in order to quantify the impacts of the COVID-19 pandemic restrictions on urban air quality parameters.

It is also in scope to analyze multiple public transportation measures and mobility indexes in an attempt to explain and quantify routine changes in urban mobility in the city like vehicle driving, traffic jams, public transport rides or shared mobility modes, and quantify the contribution of these anthropogenic urban activities changes to the changes in urban air quality parameters.

Finally, it is also an objective to train an air pollutant concentration prediction model, using atmospheric monitoring measures, air quality monitoring measures and urban mobility indexes, which are proxies for mobile anthropogenic pollutant emitters, in order to understand their predictive power and utility in monitoring air quality in an urban setting.

Hopefully, the insights revealed by the present work on urban mobility dynamics and its impact on urban air quality will provide decision-makers with additional tools to aid future policy making regarding urban air quality management, as well as demonstrating that alternative data-driven approaches to modeling localized urban air pollutant concentrations are possible and provide accurate results.

1.5. Methodologic approach

This dissertation is a hybrid between a basic research work and an applied research project [27], since it is related to a data-driven challenge sponsored by Lisbon LX Datalab challenge #49/2020 "Determining the COVID-19 pandemic impact on urban mobility and environment" [28] which intends to acquire specific insights and solve a specific problem, and a basic research work that aims to expand knowledge on urban air pollution modeling techniques.

Therefore, the present work should be considered a use-inspired basic research work (Pasteur type research) where basic research, i.e: localized urban air pollution modeling, is motivated by ultimate application, i.e: implementing specific measures to solve air quality problems in the city, as shown in the Pasteur Quadrant depicted in Figure 1.

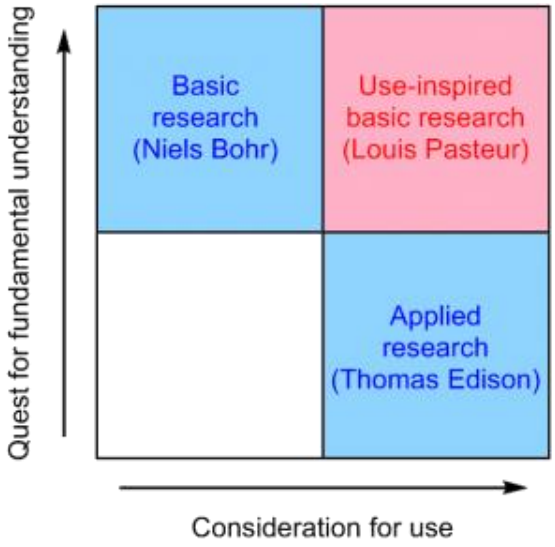


Figure 1 – Pasteur Quadrant [29][1]

Due to the data science nature of the work carried out in this dissertation, a suitable methodology was needed to help structure and execute the required work. The CRISP-DM process model was selected due to its maturity and because its iterative approach simplifies the bootstrap of the research work, when there are still a large number of unknowns, allowing for initial simpler and quicker business and data understanding, data wrangling and modeling, learning along the way and iterating several times gaining

[1] Image obtained via: http://blogs.nature.com/thescepticalchymist/2013/06/speaking-frankly-the-allure-of-pasteurs-quadrant.html

deeper knowledge of both the business and available data until the objectives have been met and the model performance is acceptable.

Additionally, taking into consideration that the present work relates to an ongoing pandemic, its management and how urban dynamics respond to change, a flexible iterative process model like CRISP-DM supports fail-fast iterations where a specific novel experiment is to be conducted midway along the dissertation work without any complexity or limitation introduced by the framework lifecycle. This iterative process also provides several milestones to iterate result evaluation with relevant stakeholders in order to gather feedback and be sure the work is going in the right direction.

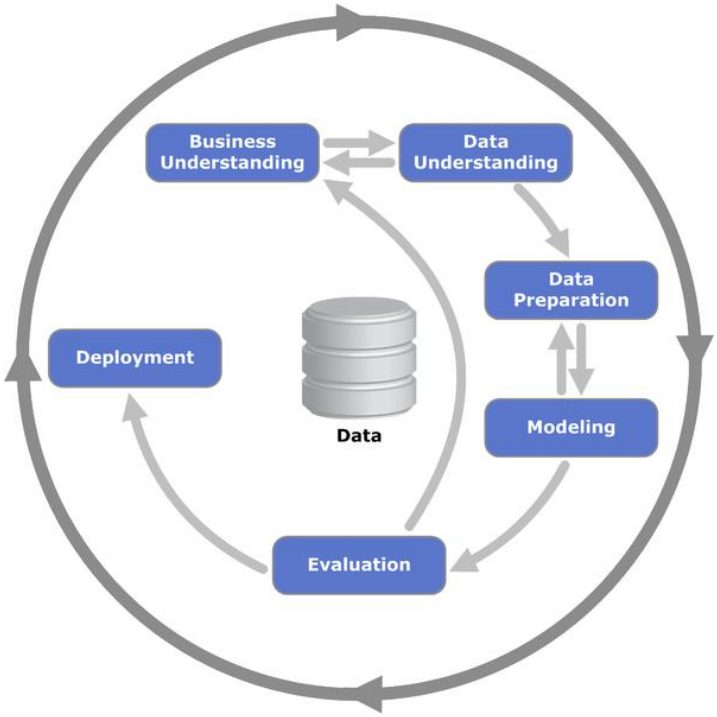


Figure 2 – CRISP-DM Methodology Phases and Flow [30][2]

The present work was conducted broadly following the CRISP-DM phases depicted in Figure 2:

1. Business Understanding: As part of the Lx Datalab challenge #49 (Impact of COVID-19 pandemic on mobility and the environment) [28], several introductory materials were provided and discussions with the challenge sponsor were carried

[2] Imagem retirada do site: https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining

out. With the business objectives understood and an initial understanding of the problem posed by the LX Datalab challenge, the initial iteration of literature review was concluded. In different phases of this work there was the need to interact with subject matter experts from several entities, such as CML, IPMA, APA and EEA, to further clarify business (or science) and data related questions as well as to perform literature review of additional subjects that were included in this dissertation scope. Most of the business understanding phase was conducted in Chapter 3.1 and from Chapter 2.2 to Chapter 2.6.

2. Data Understanding: As part of the Business Understanding phase the required data to carry out this work was assessed. In this phase data was procured and acquired from several public and private entities. Several interactions were needed to match the data needs with the available data, and to properly understand and document the data provided, including quality issues, and fit for purpose analysis. An exploratory data analysis was conducted to get acquainted with the data structure and main characteristics such as missing or corrupted data, basic statistical analysis of the data features and identifying how to relate all datasets in the following phase. The bulk of the data understanding work was done from Chapter 3.2.1 to Chapter 3.2.9.

3. Data Preparation: After the Data Understanding phase, an ETL process was created to integrate, transform, and load the available data into a multidimensional model in PowerBI. Several transformations to normalize dates, measurement units and data fill-in or correction were needed, multiple hourly fact data sources had to be aggregated to the model temporal grain (daily), dimensional tables for the monitoring stations were fitted into the model and a central date dimension was introduced to relate all the fact tables. Several logical calculated fields were added to several fact tables to simplify many of the analysis, such as moving averages, homologous analysis, or regulatory compliancy validations. While the ETL process is not extensively reported, the resulting data model is documented in Chapter 3.2.11 and aspects such as filtered data, data quality issues and aggregations performed were documented from Chapter to Chapter 3.2.1 to Chapter 3.2.9.

4. Modeling: With the multidimensional model built and available for querying with PowerBI, two very distinct modeling activities were carried out. First, a statistical

analysis on the air pollution and urban mobility impact before and during the COVID-19 pandemic, which is still ongoing as of the time of the present work, to quantify the impact of specific pandemic management phases and the implemented public policies. Secondly, a denormalized dataset was extracted from PowerBI with several relevant data features for the modeling of urban air pollution with ML techniques. Machine learning models were trained to predict air pollutant concentrations using a machine learning genetic optimization library so feature importance analysis could be conducted to perform air pollutant source appointment analysis in a data-driven fashion. The bulk of modeling phase work was carried out from Chapter 3.4 to Chapter 3.10.

5. Evaluation: For the statistical analysis phase, extensive graphical and statistical analysis was done to make sense of the impact of the different pandemic management phases on the main air pollutant and mobility indicators, thus allowing the correlation and discuss potential cause-effect of reduction of anthropogenic activities in air pollutant concentration. For the machine learning modeling phase, the genetic optimization library evaluates model performance using K-Fold cross-validation, which in the present work uses 5 folds. Since air pollutant concentration prediction is a regression problem, MSE (mean squared error) loss function was used as the model scoring function. The model performance score is then given by the average of each of the cross-validation 5 folds. Additionally, for the proposed model, the other traditional regression metrics, such as MAE (mean absolute error), RMSE (root mean squared error), MSLE (mean squared log error) and R^2 (coefficient of determination), were also computed to analyse the model performance. As with modeling phase work, the bulk of evaluation phase work was carried out from Chapter 3.4 to Chapter 3.10.

6. Deployment: Since the present work aims to provide insights into the impact of the reduction of anthropogenic activities associated with COVID-19 restrictions on urban mobility and the environment, no system or framework was produced that could be deployed. In this sense, this phase concerns the actual use of the critical knowledge attained by the present work by public authorities for decision making and by the scientific community. The operationalization of a fully functioning system based on the present work is proposed as future work in Chapter 4.3.

Chapter 2 – Literature review

Three different research topics are explored independently. The first is related to all relevant variables, direct and indirect, that can play a part in inhibiting, emitting, transforming, transporting, trapping, deposition and re-suspension of pollutant gases and aerosols. This air pollutant lifecycle, as depicted in Figure 3, includes complex non-linear relationships of different mechanisms related to nature of emissions, weather patterns, photochemical reactions in the atmosphere and the topography, where resulting pollution effects can be deferred in time and in places far away from where the main contributing anthropogenic or biogenic emissions occurred.

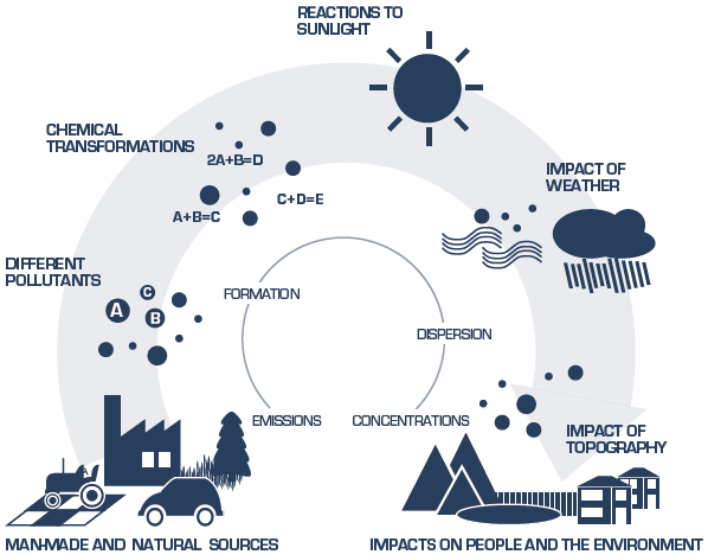


Figure 3 – Air Pollutant lifecycle schematic [30][3]

The second is related to machine learning techniques applied to model air pollution, namely in uncovering predictor power of variables, engineering data features and training models capable of capturing the air pollutant phenomena and comparing machine learning techniques in terms of weaknesses and strengths.

The last is related to analysis of air pollutant performance during COVID-19 restrictions in which anthropogenic activities were severely impacted.

[3] Image obtained from: https://www.eea.europa.eu/media/infographics/many-factors-contribute-to-air-1/view

Before these three topics, a brief description of the troposphere chemistry involving Nitrogen Oxides (NO_x) and Ozone (O_3) is described to further contextualize the phenomena being studied in the present work.

2.1. Literature review process

Following the CRISP-DM process model, the literature review process followed an iterative approach as well. Two initial research topics, "Machine Learning approaches to air pollutant concentration modeling" and "Impact of COVID-19 restrictions on urban air pollution", were explored up to the required depth that could unlock an iteration where analysis and experiments were carried out, leading to either further refinement of the literature review for the two existing research topics or opening up way to new research topics, namely "Relevant variables for estimating urban air pollutant concentration" which was included at a later phase of the present work.

In order to search existing relevant scholarly literature, Google Scholar was used. Taking into consideration that Google Scholar is not a scholarly literature database, but rather an academic search engine, it does not have an editorial review board and content quality is solely evaluated by means of specialized algorithms, so special care was taken to assess the source and quality of the papers, namely if they come from peer-reviewed journals as well as sorting out conference proceedings, reports, and other documents as grey literature.

Literature for three different relevant research topics were searched separately. The literature research topics are as follows:

- Relevant variables for estimating urban air pollutant concentration
    - Search query: (O3 OR NO2 OR SO2 OR CO OR PM2.5 OR PM10 OR C6H6) AND (Air Pollution) AND (variables OR factors OR predictors OR correlation)
    - Time filter: 1990-01-01 until now
    - Language: English, Portuguese
    - Full-text articles analyzed: 21
- Machine learning methods for air pollutant concentration prediction

    - Search query: (O3 OR NO2 OR SO2 OR CO OR PM2.5 OR PM10 OR C6H6) AND (Air Pollution) AND (Artificial Intelligence OR Machine Learning)
    - Time filter: From 2010-01-01 until now
    - Language: English, Portuguese
    - Full-text articles analyzed: 11
- COVID-19 pandemic restrictions impact on urban air pollutant concentration
    - Search query: (O3 OR NO2 OR SO2 OR CO OR PM2.5 OR PM10 OR C6H6) AND (Air Pollution) AND COVID-19
    - Time filter: From 2020-01-01 until now
    - Language: English, Portuguese
    - Full-text articles analyzed: 17

In addition to scholarly literature, several searches for grey literature, in the form of official technical and statistical reports, as well as other publications like fact sheets, were conducted in trusted and/or official Portuguese, European Union and International agencies related to environment, demography, mobility, etc. Finally, press articles related to pandemic management and environment in Lisbon were also searched through Google News.

2.2. Atmospheric chemistry of Nitrogen Oxides and Ozone

In urban centers, Nitrogen Oxides (NO_x) emissions are mostly associated to internal combustion powered vehicles. In Lisbon and Tagus Valley, internal combustion vehicles associated with road traffic are estimated to be responsible for 63% of all emissions, and when air and maritime transportation are included, the contribution increases to 71% [23].

Nitrogen dioxide (NO_2) is produced both directly and indirectly by the combustion of fuel at high temperatures in the engine which emit both Nitrogen dioxide (NO_2) and Nitric Oxide (NO) in different ratios depending on the engine size and fuel. Directly emitted Nitric Oxide (NO) will then react with atmospheric oxidizing agents to effectively produce Nitrogen dioxide (NO_2) indirectly.

Simplified set of chemical reactions describing daytime NO_x atmospheric chemistry [31]:

Anthropogenic sources of NO:

(1) $N_2 + O_2 \rightarrow 2NO$ [At high temperatures during combustion process]

Main NO_2 formation mechanisms:

(2) $NO + O_3 \rightarrow NO_2 + O_2$ [O_3 titration by NO into NO_2can be a net destroyer of O_3in highly NO polluted scenarios, even during daytime]

(3) $RO_2 + NO \rightarrow NO_2 + RO$ [NO oxidation by free radicals generated by daytime VOC photolysis]

Tropospheric O_3 formation:

(4) $NO_2 + hv(\lambda < 420nm) \rightarrow NO + O$ [Daytime photolysis of NO_2]

(5) After (4), $O + O_2 \rightarrow O_3$ [O_3 production]

(6) Downwards O_3 flux from the stratosphere

Tropospheric O_3 destruction besides (2):

(7) $O_3 + hv(\lambda < 300nm) \rightarrow O + O_2$ [Daytime photolysis of O_3]

This cycle is further replenished by the oxidation of the NO molecule formed in reaction (4) into NO_2 by free radicals produced VOC photochemical complex reactions which means that tropospheric Ozone (O_3) is a result of both NO_x and VOC pollution.

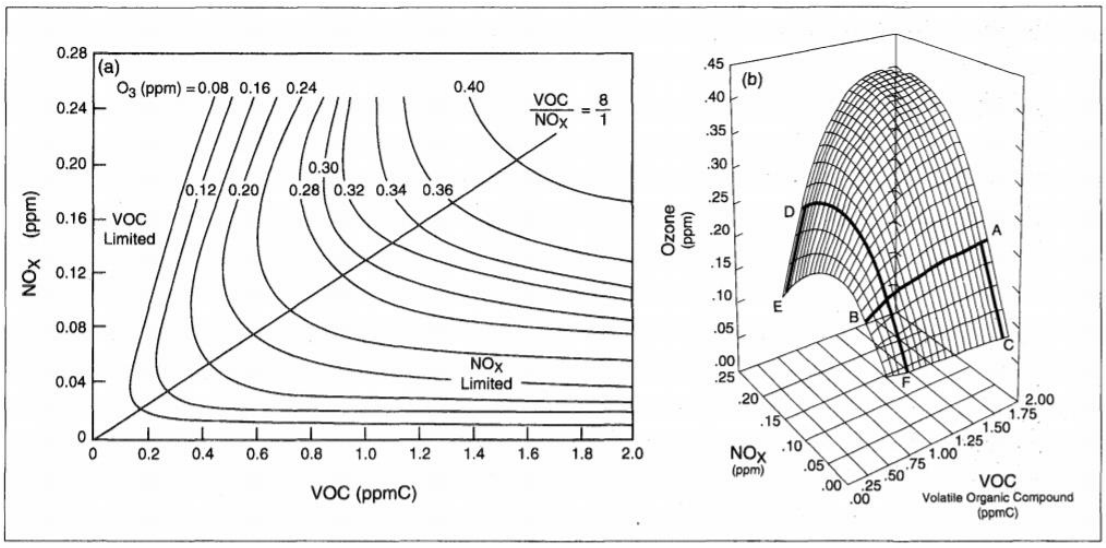


Figure 4 – Peak ozone isopleths generated from initial mixtures of VOC and NOx [31][4]

[4] Image obtained via: https://www.tandfonline.com/doi/pdf/10.1080/1073161X.1993.10467187

The Ozone formation efficiency is related to the ratio of NO_x to VOCs since the NO_2 recycling process is highly dependent on VOCs to oxidize the NO molecule resulting from the NO_2 photolysis process (4). This ratio, as depicted by the NO_x:VOC isopleths in Figure 4 , defines if the tropospheric ozone production is limited by VOC or NO_x concentrations. In VOC-limited scenarios such as urban centers, sharp drops in NO_x concentrations, such as the ones during COVID-19 lockdown restriction periods, might result in additional O_3 production while rebalancing towards a NO_x limited scenario (i.e: assuming a VOC-stable scenario, a NO_x decrease from point E to point D in Figure 4 will increase Ozone production efficiency) as well as reduced O_3 titration by NO.

The present work is focused on analyzing NO_2 (Nitrogen dioxide) as a primary pollutant and O_3 (Tropospheric Ozone) as secondary pollutant, omitting the role of NO_2 as a precursor to other important secondary pollutants such as Particulate Matter (PM), Nitric acid (HNO_3) or Peroxyacetyl nitrate (PAN).

Finally, as already stated before, both NO_2 (precursor) and O_3 (secondary) can be transported to places far away from its formation by air currents, which generates pollution events far from the places where the air pollutants were emitted or produced.

In this sense, a simplified generalization of the Nitrogen Oxides (NO_x) and tropospheric Ozone (O_3) urban pollution can be described as depicted in Figure 5:

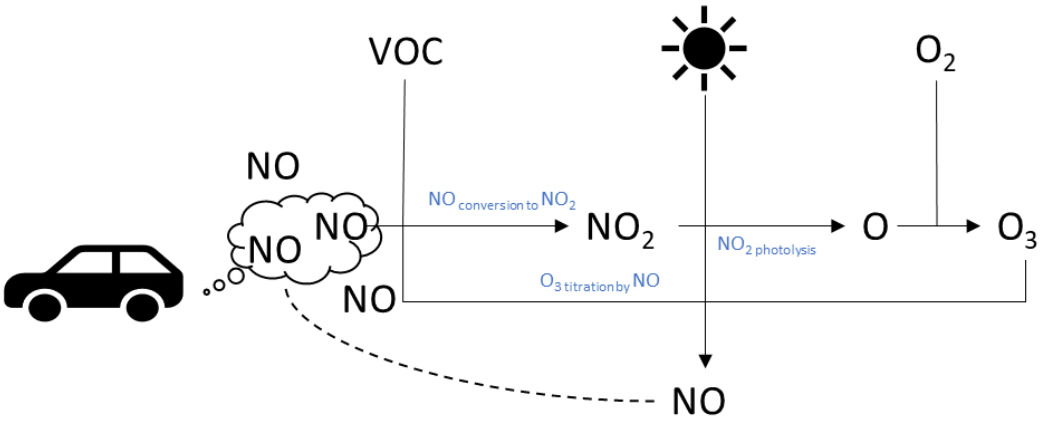


Figure 5 – Generalized NOx – O3 lifecycle

2.3. Relevant variables for estimating urban air pollutants

Atmospheric pollution depends on several anthropogenic and biogenic processes responsible for primary pollutant emission, chemical reactions in the atmosphere that lead to secondary pollutants, meteorological conditions which help potentiate such chemical reactions, the geography and topographic features, the long-range transportation of aerosols and gases as well as pollutant deposition and resuspension [32].

Zhang *et al.* [33] investigated the main meteorological factors affecting the concentrations of particulate matter in Wuhan, China in the 2013-2016 period using data sourced from nine air quality monitoring stations in the city. High wind intensity was found to be linearly correlated with $PM_{2.5}$ concentration reduction, measuring 60% reduction in concentrations when wind speed is up to 6m/s while PM_{10} was found to have a non-linear relationship with 15% concentration reduction when wind speed is up to 6m/s, potentially due to resuspension of coarse particulate matter during high-wind intensity episodes. Regarding precipitation, the particulate matter wet deposition (washout) and resuspension preventive action was also measured as being more relevant from PM_{10} than $PM_{2.5}$ with rainy days measuring 69% and 72% mean concentration respectively when compared to non-rainy days.

Regarding Ozone secondary pollutant, Zoran *et al.* [34] measures the relationship between meteorological factors with ground-level Ozone formation in Milan, Italy in the period of January to April 2020. Air temperature and ground level Ozone formation is related since solar radiation, which is key to drive air temperature, increases photochemical performance of Ozone formation, having resulted in a positive correlation of R^2= 0.84. It is also suggested that particulate matter with aerodynamic diameter of 2.5 microns ($PM_{2.5}$) and 10 microns (PM_{10}) negatively affects ground level Ozone (O_3) production due to increased sunlight reflection on particulate matter thus slowing down photochemical reactions that produce Ozone. The negative correlation of ground level Ozone and $PM_{2.5}$ measured R^2= −0.63 and PM_{10} measured R^2= −0.61. The relationship of ground level Ozone is also related to relative humidity since high relative humidity is associated with cloud formation, which also reduces the sunlight reaching the ground, washout effect of primary precursor pollutant NO_2 when precipitation occurs, and thus also slowing down photochemical reactions that produce Ozone having measured a negative correlation (R^2= −0.79). Finally, the dispersion of ground level Ozone is also affected by the Planetary Boundary Layer (PBL) height where higher PBL height, along

with stronger convective flows, is correlated with lower ground level Ozone concentrations, and conversely lower PBL height is correlated with higher ground level Ozone concentrations, effectively trapping pollutants near the surface, having measured a positive correlation (R^2= 0.74).

Lecocq *et al.* [35] used seismometer data from 268 monitoring stations to model a proxy of anthropogenic activities, such as road traffic, trains, aircraft movements or industries, where during the COVID-19 pandemic related lockdowns it was measured a global reduction of seismic noise of 50%. For the specific case of Brussels, Belgium this reduction was validated against other independent sources of anthropogenic activities, namely audible noise road traffic monitoring and Google mobility reports (Transportation) for the city of Brussels, having achieved strong positive correlations of 0.93 for and 0.86 respectively, which suggests seismic signal can be used as a proxy for anthropogenic activities.

Bonet-Solà *et al.* [36] used data from Barcelona acoustic monitoring network used to monitor urban noise pollution associated with anthropogenic activities such as airport traffic, road traffic, industry, shopping, and leisure throughout the COVID-19 pandemic related lockdown in the city comparing multiple lockdown stages and 2018 to 2019 homologous periods. A 49.56% reduction in noise levels was measured throughout the most severe lockdown period ranging from March to April 2020 where NO_2 concentrations are also reported to have decreased 46%. Fridays and Saturdays were found to be the weekdays with a wider gap to the baseline and daytime leisure, restaurants and nightlife being the areas with wider gaps to the baseline, high road traffic zone noise levels reduction were milder than moderate traffic zones which also experienced relevant noise reductions during the afternoon. Such findings shows that acoustic noise in urban settings is a potential proxy for anthropogenic activities.

Gualtieri *et al.* [37] used a municipality level mobility indicator from connected vehicles, maps and navigation system provided by EnelX & Here which was confirmed to be highly correlated with independent traditional traffic counters in Milan, Italy, achieving R^2=0.972, mean bias of 0.8% and RMSE of 5.7% for the analysis period of 7th of February to 30th of April 2020 and was therefore considered a good proxy for road traffic so its contribution to the air pollution reduction during the COVID-19 pandemic restrictions in Italy could be extrapolated.

2.4. Impact of COVID-19 restrictions on urban air quality

The impact of sharp reductions of anthropogenic activities throughout many countries during COVID-19 pandemic associated lockdowns were widely analyzed as to its impact and contributions to primary and secondary air pollutant concentrations. Rana *et al.* [38] carried out a systematic literature review of the impact of COVID-19 restrictions on air quality in China using PRISMA guidelines where 35 studies out of 396 met the eligibility criteria and were thoroughly analyzed. All articles were published in 2020 and included articles using both satellite (12) and ground monitoring (23) air pollutant concentration data, with temporal comparison to both pre and post lockdown periods in 2020 (8), comparison to the same period in 2019 (15) and the remainder compared using period ranges from 2015 to 2019 (12). The most studied air pollutant was NO_2 (80%) followed by $PM_{2.5}$ (75%) and the authors estimate that during COVID-19 lockdown periods the mean reduction of air pollutant concentrations reported was 45.1% for NO_2, 26.6% for $PM_{2.5}$, 31.4% for PM_{10}, 31.3% for SO_2, 20.7% for CO and 21.7% for AQI.

Highlighting some additional studies, Bauwens *et al.* [39] used satellite observations to acquire NO_2 vertical column measures from two satellite instruments, TROPOMI (TROPOspheric Monitoring Instrument) on board of Sentinel P-5 satellite and OMI (Ozone Monitoring Instrument) on board of Aura satellite, to reveal a sharp reductions of NO_2 concentrations throughout the COVID-19 related lockdown phases in multiple cities throughout the world when compared to pre-lockdown periods and homologous periods in 2019. Notable examples include -43% to -57% in Wuhan, China, and -31% to -32% in Barcelona, Spain.

Connerton *et al.* [40] used air quality data from ground monitoring stations in four megacities during the initial COVID-19 lockdown period in March 2020, and conducted a statistical analysis having measured pollutant concentration reductions, for instance in Paris, France, of 67% for CO, 39% for NO_2 and 29% for $PM_{2.5}$ when compared to a 2015-2019 air pollutant baseline. In order to extrapolate the contribution of the reduction of anthropogenic activities to the air pollutant measured concentration changes during the pandemic period, thus accounting for natural atmospheric phenomena that can disperse air pollutants or facilitate atmospheric chemical reactions that consumes said air pollutant, such as wind speed, air temperature and relative humidity, a General Linear Model was fitted with meteorology measures and a lockdown indicator variable as independent variables and air pollutant concentration as dependent variables, where it was found that

while both anthropogenic activities and meteorology significantly influence air pollutant concentrations, anthropogenic activities reduction contribution was heavier.

Sicard *et al.* [41] investigated the effect of sharp reductions of NO_x emissions on O_3 concentrations during the COVID-19 related lockdowns in European and Chinese cities where it was suggested that local urban O_3 concentrations greatly increased as NO_x emissions decreased due to reduced O_3 titration by NO thus raising the need of controlling VOC emissions to balance the NO_x:VOC ratio which is key in tropospheric O_3 formation. For instance, during the 2020 COVID-19 related lockdown period in Wuhan, China daily mean O_3 concentration increased by 36% while NO_2 concentrations decreased 57%, when compared to a computed 2017-2019 baseline for the same period.

2.5.Machine Learning based Air Pollutant Concentration Prediction

Advanced atmospheric chemical transport frameworks [42] to model air pollution, such as CAMX or CMAQ, use as inputs a weather forecast model, an emission inventory for anthropogenic and biogenic sources and using a chemical transport model are able to output an 3D air pollution concentration map and compute source appointment analysis. On the other hand, instead of a framework capable of simulating complex atmospheric chemical interactions, which can rely on outdated or incorrect estimated emission stocks, there are other data-driven techniques which use statistical and machine learning techniques to model air pollutant concentrations, capture the complex non-linear relationship of the several variables that determine the concentration of a given air pollutant at a specific place and time. Rybarczyk and Zalakeviciute [43] have performed a systematic review of machine learning approaches to outdoor air pollution modeling following the PRISMA guidelines where 46 out of 103 papers published from 2010 to 2018 met the eligibility criteria. It was found that the volume of publications of applied machine learning techniques related to air quality has ramped up from 2016 onwards, mostly published with regards to northern hemisphere geographies, and are mostly related to criteria pollutants (NO_2, SO_2, CO, PM_{10}, $PM_{2.5}$ and O_3) or the Air Quality Index (AQI), whereas the largest group of publications are related to identifying relevant predictor variables and modeling non-linear relationship of variables in air pollution. The main algorithm classes ordered by prevalence found in the publications were Ensembles (mainly tree-based predictors), ANNs, SVMs and LRs.

Further analyzing more recent publications, Vu *et al.* [44] trained a Random Forest ML algorithm and compared prediction results for 2017 $PM_{2.5}$ concentrations in Beijing with those outputted by a WRF-CMAQ model and was able to produce a slightly more accurate value for the yearly mean $PM_{2.5}$ concentration, 61.8 – 62.4 μg/m3 for the WRF-CMAQ model and 61.0 μg/m3 for the RF ML model, whereas the observed value was 58.0 μg/m3. At the month granularity for 2017, the WRF-CMAQ model concentration predictions ranged from 3% to 33.6% difference when compared to the observed values, a mean difference of 7.8%, whereas the RF ML model predictions ranged from 0.4% to 7.9%, a mean difference of 1.5%. Castelli *et al.* [45] employed SVR (Support Vector Regression) algorithm to predict air pollutant concentrations such as NO_2, CO, SO_2, O_3 and $PM_{2.5}$ in California for years 2016 to 2018, using meteorological measures, pollutant concentration rolling means and timeseries features as predictors, achieving R^2=0.937 on the validation set using the RBF (Radial Basis Function) kernel for NO_2 forecasting. Luna *et al.* [46] also employed an SVR (Support Vector Regression) algorithm to predict O_3 concentrations in Rio de Janeiro using ground monitoring stations data from 2011 and 2012, using as data features the chemical precursors, such as NO, NO_2, NO_x and CO, as well as meteorological factors, namely wind speed, solar radiation, air temperature and relative humidity. This model achieved R^2=0.912 on the validation set, having also trained an ANN (Artificial Neural Network) using the same data source to solve the same problem and achieving R^2=0.915 on the same validation set.

2.6. Literature Review Summary

A high-level summary of the most relevant papers on the three research topics analyzed in the literature review can be found in Table 1.

Table 1 – Literature Review Summary of Main Articles

Research topic	Author(s)	Title	Source	Main conclusions
Machine learning based air pollutant concentration prediction	Rybarczyk and Zalakeviciute (2018)	Machine Learning Approaches for Outdoor Air Quality Modelling: A Systematic Review	Applied Sciences (Q2 Journal)	Volume of publications of applied machine learning techniques related to air quality has ramped up from 2016 onwards and are mostly related to criteria pollutants (NO_2, SO_2, CO, PM_{10}, $PM_{2.5}$ and O_3) or the Air Quality Index (AQI). The main algorithm classes ordered by prevalence found in the publications were Ensembles (mainly tree-based predictors), ANNs, SVMs and LRs.
	Vu *et al.* (2019)	Assessing the impact of clean air action on air quality trends in Beijing using a machine learning technique	Atmospheric Chemistry and Physics (Q1 Journal)	Results from a Random Forest ML model trained with 2017 data to predict $PM_{2.5}$ concentrations in Beijing slightly more accurate than those outputted by a WRF-CMAQ model.
	Castelli *et al.* (2020)	A Machine Learning Approach to Predict Air Quality in California	Complexity (Q1 Journal)	A Support Vector Regression model trained to predict air pollutant concentrations in California for years 2016 to 2018 achieved R^2=0.937 on the validation set using the Radial Basis Function kernel for NO_2 forecasting.
	Luna *et al.* (2020)	Prediction of ozone concentration in tropospheric levels using artificial neural networks and support vector machine at Rio de Janeiro, Brazil	Atmospheric Environment (Q1 Journal)	Two models trained to predict ground-level Ozone concentrations in Rio de Janeiro using data from 2011 and 2012 achieved R^2=0.912 (Support Vector Regression) and R^2=0.915 (Artificial Neural Network) on the validation set.
Impact of COVID-19 restrictions on urban air quality	Rana *et al.* (2021)	A Systematic Literature Review of the Impact of COVID-19 Lockdowns on Air Quality in China	Aerosol and Air Quality Research (Q2 Journal)	The most studied air pollutant was NO_2 (80%) followed by $PM_{2.5}$ (75%) and the authors estimate that during the COVID-19 lockdown periods the mean reduction of air pollutant concentrations reported were 45.1% (NO_2), 26.6% ($PM_{2.5}$), 31.4% (PM_{10}), 31.3% (SO_2), 20.7% (CO) and 21.7% (AQI).
	Bauwens *et al.* (2020)	Impact of Coronavirus Outbreak on NO_2 Pollution Assessed Using TROPOMI and OMI Observations	Geophysical Research Letters (Q1 Journal)	NO_2 vertical column measures from the TROPOMI instrument aboard Sentinel P-5 satellite show a sharp reduction of NO_2 concentrations during COVID-19 restrictions in multiple cities around the world when compared to pre-lockdown periods and homologous periods in 2019. Notable examples include -43% to -57% in Wuhan, China, and -31% to -32% in Barcelona, Spain.
	Connerton *et al.* (2020)	Air Quality during COVID-19 in Four Megacities: Lessons and Challenges for Public Health	International Journal of Environmental Research and Public Health (Q2 Journal)	Analysis of four megacities during the initial COVID-19 lockdown period in March 2020 measured pollutant concentration reductions, for instance in Paris, France, of 67% for CO, 39% for NO_2 and 29% for $PM_{2.5}$ when compared to a 2015-2019 air pollutant baseline

	Sicard *et al.* (2020)	Amplified ozone pollution in cities during the COVID-19 lockdown	Science of The Total Environment (Q1 Journal)	Sharp reductions of NO_x emissions augmented O_3 concentrations during the COVID-19 related lockdowns in European and Chinese cities due to reduced O_3 titration by NO. For instance, during the 2020 COVID-19 related lockdown period in Wuhan, China, daily mean O_3 concentration increased by 36% while NO_2 concentration decreased 57% when compared to a 2017-2019 baseline for the same period.
Relevant variables for estimating urban air pollutants	Zhang *et al.* (2018)	Influences of wind and precipitation on different-sized particulate matter concentrations ($PM_{2.5}$, PM_{10})	Meteorology and Atmospheric Physics (Q3 Journal)	A study in Wuhan, China for the 2013-2016 period found that wind intensity linearly correlated with $PM_{2.5}$, measuring 60% reduction when wind speed is up to 6m/s while PM_{10} has a non-linear relationship with a 15% concentration reduction when wind speed is up to 6m/s, potentially due to resuspension of coarse particulate matter during high-wind intensity episodes. Particulate matter wet deposition (washout) and resuspension preventive action was also measured as being more relevant from PM_{10} than $PM_{2.5}$ with rainy days measuring 69% and 72% mean concentration respectively when compared to non-rainy days.
	Zoran *et al.* (2020)	Assessing the relationship between ground levels of ozone (O_3) and nitrogen dioxide (NO_2) with coronavirus (COVID-19) in Milan, Italy	Science of The Total Environment (Q1 Journal)	A study in Milan, Italy for the period of January to April 2020 found that sunlight reflection on particulate matter results in lowered tropospheric Ozone production having measured a negative correlation of ground level Ozone and $PM_{2.5}$ ($R^2 = -0.63$) and PM_{10} ($R^2 = -0.61$). High relative humidity increases cloud cover, which reflects sunlight, and Ozone precursor washout which results in a negative correlation ($R^2 = -0.79$). Trapping effect of lower PBL height also results in increased Ozone concentrations resulting in a positive correlation ($R^2 = 0.74$).
	Lecocq *et al.* (2020)	Global quieting of high-frequency seismic noise due to COVID-19 pandemic lockdown measures	Science (Q1 Journal)	Correlation analysis of urban seismic signal and audible noise road traffic monitoring, as well as Google mobility reports (Transportation), for the city of Brussels, Belgium during COVID-19 restrictions, achieved strong positive correlations of 0.93 for and 0.86 respectively, which suggests seismic signal can be used as a proxy for pollutant anthropogenic activities.
	Bonet-Solà *et al.* (2021)	The Soundscape of the COVID-19 Lockdown: Barcelona Noise Monitoring Network Case Study	International Journal of Environmental Research and Public Health (Q2 Journal)	Data from Barcelona, Spain acoustic monitoring network shows a 49.56% reduction in urban noise during COVID-19 restrictions in 2020 when compared to a homologous 2018-2019 baseline, which was found to be highly correlated to NO_2 reductions in the same period which was of 46%.
	Gualtieri *et al.* (2020)	Quantifying road traffic impact on air quality in urban areas: A Covid19-induced lockdown analysis in Italy	Environmental Pollution (Q1 Journal)	A municipality level mobility indicator from connected vehicles, maps and navigation system provided by EnelX & Here was confirmed to be highly correlated with independent traditional traffic counters in Milan, Italy, achieving R^2=0.972, mean bias of 0.8% and RMSE of 5.7% for the analysis period of 7th of February to 30th of April 2020 and was therefore considered a good proxy for road traffic.

2.7. Differentiated approach

Several differentiated approaches are to be taken in throughout this dissertation which can uncover academic and business implications.

1. There is no comprehensive study on the impact of COVID-19 restrictions in urban air quality and mobility for Lisbon, Portugal geography which has particular topographic and orographic characteristics, as well as different weather patterns.

2. Several articles refer to the usage of machine learning techniques to model air pollution but are usually fitted with only weather and air pollutant concentration data as independent variables. In an attempt to improve model performance and increase the interpretability of air pollution phenomena, the current work proposes the usage of multiple urban mobility indicators, covering different aspects of mobility modes in the city, as a group of independent variables that are proposed to be proxies for pollutant anthropogenic activities.

3. Several articles refer to the usage of AutoML frameworks to model air pollution, but the results are not well discussed and not compared to applied traditional ML techniques when it comes to training efficiency, model performance and interpretability.

Chapter 3 – Analysis and results presentation

3.1.Business Understanding

The primary objective of the current work is to study the impact of the COVID-19 pandemic restrictions imposed by the government and changes in anthropogenic activities in the urban environment and mobility in the city of Lisbon. Besides the immediate interest in learning the impacts of COVID-19 related restrictions on the urban mobility and the environment in Lisbon, the cause-effect relationship between the reduction of heavy emitter anthropogenic activities, such as road traffic, and the changes in NO_x and O_3 concentrations can be studied in a real-life laboratory with a months-long field experiment instead of simulated experiments or sub-optimal conditions such as the weekend effect or the holiday periods.

For this problem to be correctly modelled in a data-driven way, common air pollutants concentration resultant of anthropogenic urban activity, atmospheric conditions, and mobility indexes per type of urban transportation mode, namely private vehicle means, must be available at a proper time granularity for the period in analysis (before, during and after the pandemic related restriction periods) and, where needed, for baseline periods that are used for comparison.

Anthropogenic emissions related to urban mobility are not the solely responsible for the different types of air pollution experienced in the city. The long-range transport of aerosols (i.e: Sahara Desert dust, sea-salt spray), biogenic emissions (i.e: different type of plants), tropospheric ozone transported from other origins, sulphur dioxide being released naturally by a volcano or by being in the downwind path of an industrial complex can contribute to the air quality degradation in any specific geography as depicted in Figure 6. The convention on Long-Range Transboundary Air Pollution (LRTAP) and its protocols were put in place to investigate and propose solutions for these phenomena [47]. Nevertheless, as stated in the introduction, road transportation is generally considered the most relevant contributor to urban air pollution and the scope of the work is narrowed down to Lisbon, Portugal geographical context.

Specifically, for the scope of this work, the focus is placed on the NO_2 primary pollutant, which in an urban setting is mostly associated with road transport, and the O_3 secondary pollutant for which NO_2 is a precursor, along with VOCs, whereas the impact of COVID-19 restrictions on the remainder criteria pollutants (CO, SO_2, PM_{10}, $PM_{2.5}$ and C_6H_6) is

only analyzed in a high-level fashion. The atmospheric chemistry process associated with NO_2 and O_3 is described in section 2.2.

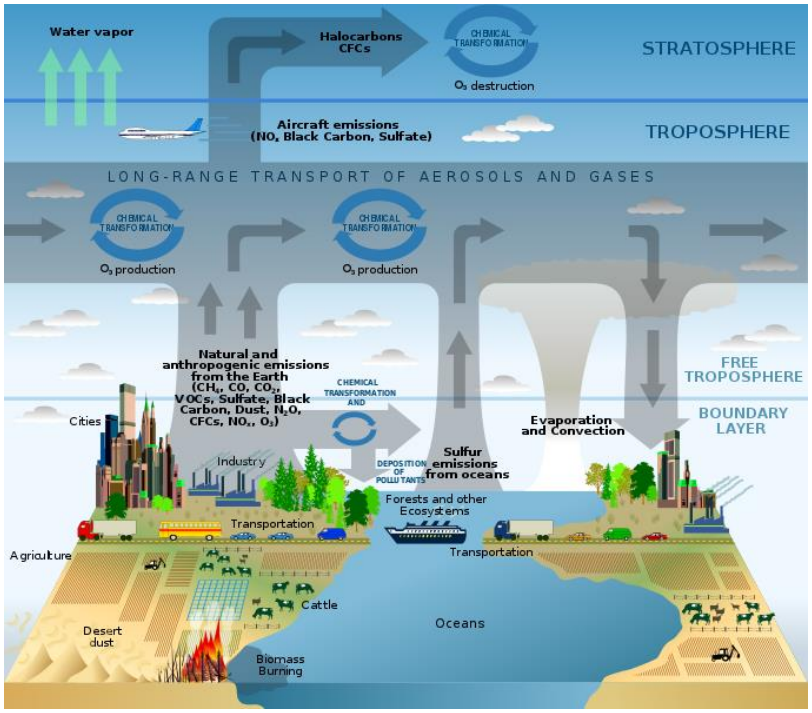


Figure 6 – Schematic of chemical and transport processes related to atmospheric composition. These processes link the atmosphere with other components of the Earth system, including the oceans, land, and terrestrial and marine plants and animals [48][5]

Regulation is in place in the European Union, namely Decree Law No.102/2010 of 2010-09-23 (transposition of EU Directives 2008/50/CE and 2004/107/CE), to control air pollutant thresholds [49] in order to safeguard human health, which in Portugal is monitored and enforced by APA and CCDR-LVT. Additionally, WHO also has guidelines with the recommended thresholds for outdoor pollutants [24] which are usually tighter since they are forward looking and serve as reference for future legislation. Both EU and WHO defined threshold can be found in Table 2.

[5] Image obtained from: https://www.globalchange.gov/browse/multimedia/chemical-and-transport-processes-related-atmospheric-composition

Table 2 – WHO Air Pollutant Safety Threshold Guidelines and EU regulated air quality thresholds for air pollutants in scope for this work

Air Pollutant	**Entity**	**Threshold**	**Notes**
NO_2	WHO	40 μg/m3 annual mean 200 μg/m3 1-hour mean	
CO	WHO	*20 mg/m3 1-hour mean 10 mg/m3 8-hour mean	Exposure at these concentrations should be no longer than the indicated times and should not be repeated within 8 hours.
$PM_{2.5}$	WHO	10 μg/m3 annual mean 25 μg/m3 24-hour mean	
PM_{10}	WHO	20 μg/m3 annual mean 50 μg/m3 24-hour mean	
O_3	WHO	100 μg/m3 8-hour mean	
SO_2	WHO	20 μg/m3 24-hour mean 500 μg/m3 10-min mean	
NO_2	EU	40 μg/m3 annual mean *200 μg/m3 1-hour mean	*Permitted yearly exceedances: 18
CO	EU	10 mg/m3 8-hour mean	
$PM_{2.5}$	EU	25 μg/m3 annual mean	
PM_{10}	EU	40 μg/m3 annual mean *50 μg/m3 24-hour mean	*Permitted yearly exceedances: 35
O_3	EU	*120 μg/m3 8-hour mean	*Permitted yearly exceedances: 25 days (averaged over 3 years)
SO_2	EU	*125 μg/m3 24-hour mean **350 μg/m3 1-hour mean	*Permitted yearly exceedances: 3 **Permitted yearly exceedances: 24
C_6H_6	EU	5 μg/m3 annual mean	

Additionally, the thresholds defined in Table 3 are used to issue public announcements which were introduced by Decree Law No.102/2010 of 2010-09-23 [30] to warn the population about severe pollution episodes so that specific mitigation measures can be implemented to avoid harm to public health, namely to the most fragile population groups affected by respiratory disease induced by poor air quality such as asthma.

These regulatory thresholds are used throughout the statistical analysis of air pollutant concentration performance throughout COVID-19 pandemic lockdowns to understand how the compliance has been affected for the Lisbon geography.

Table 3 –EU regulated air quality thresholds for public announcement

Air Pollutant	**Entity**	**Threshold**	**Notes**
NO_2	EU/PT	Alert: 400 µg/m3 1-hour mean	** For 3 consecutive hours either in an area of 100km2 or entirety of a populational agglomeration
O_3	EU/PT	Info: 180 µg/m3 1-hour mean Alert: 240 µg/m3 1-hour mean	
SO_2	EU/PT	Alert: 500 µg/m3 1-hour mean	For 3 consecutive hours either in an area of 100km2 or entirety of a populational agglomeration

3.2.Available data sources

In order to pursue this work objectives and according to the literature review process, several data sources are required to properly model the urban air pollution phenomena: (1) Air pollutant concentration, (2) meteorological parameters and (3) anthropogenic activity direct or indirect indicators.

Table 4 – Approached entities, requested data and request success

Entity	**Status**	**Dataset**
APA/EEA	Data available (Online)	Air pollutant concentration (ground)
ESA/Copernicus	Data available (Online)	Air pollutant concentration (satellite)
IPMA	Data available (On-demand)	Metereological parameters
Metropolitano de Lisboa	Data available (On-demand)	Subway ridership
Carris	Data unavailable	Urban bus ridership
Comboios de Portugal	Data unavailable	Suburban train ridership
ANA	Data available (On-demand)	Lisbon airport volume of aircraft movements and passengers transported
Google	Data available (Online)	Lisbon Mobility Report
Apple	Data available (Online)	Lisbon Mobility Report
Moovit	Data available (Online)	Lisbon Public Transport Mobility Index
Waze	Data available (Online)	Lisbon Waze App driven distances; Lisbon Waze Jams
SIBS	No response	Volume of Point-Of-Sales transactions.
Via Verde	No response	Volume of highway toll passages; Volume of parking operations
PSE	Data unavailable	Lisbon Mobility Report
CML/EMEL	Data available (On-demand)	GIRA docks status and ridership; car Parking statistics; bike lane traffic counters
DGS	Data available (Online)	COVID-19 pandemic daily reports
SICO/EVM	Data available (Online)	Lisbon and Tagus Valley mortality

In this sense, apart from data available online, several public entities and companies were directly contacted to acquire such data as described in Table 4.

As part of the data understanding phase of CRISP-DM, all the data and respective metadata were analyzed, exploratory data analysis was performed, quality problems identified and finally loaded to the data analysis tool (PowerBI).

The following sub-chapters describe the data sources used throughout this dissertation in a compact fashion while the full metadata for datasets with a large number of attributes is provided in the annexes of the present work. Additionally, many of the attributes are named in Portuguese and have not been translated due to source metadata consistency purposes. Nevertheless, the field description is properly detailed in English.

3.2.1. Lisbon Ground-based Meteorological Data Source

Data Source Description: Hourly weather measures were sourced from IPMA (Portuguese Institute for Sea and Atmosphere) for the Lisbon Gago Coutinho weather station (lat: 38,76620278, lon: -9,12749444) for the 2019 and 2020 years.

The averaged hourly measures include wind speed (m/s) and direction (º), relative humidity (%), air temperature (ºC) and sea level air pressure (hPa). Hourly cumulative measures include total global radiation (KJ/m2) and accumulated precipitation (mm)

Data Quality issues: Atmospheric condition measures for Lisboa Gago Coutinho station sourced from IPMA has missing data for the day 26/03/2020 which was filled-in with Lisboa Geofísico station data.

Aggregation performed: The data-source reports weather parameters at the hour granularity. To integrate in the consolidated data model, a temporal aggregation to the daily grain was conducted. The resulting metrics include the minimum, maximum, average, quartile 1, quartile 2 (median) and quartile 3 concentrations values at the daily level for all metrics except the additive metrics total global radiation and accumulated precipitation which were additionally summed.

Filtered Data: None

Data Source Metadata: See Appendix B – Data Source Metadata

3.2.2. Lisbon Mobility Indexes Data Sources

Four different data sources describing urban mobility patterns in Lisbon, Portugal were used as proxy to measure changes in anthropogenic activities.

Data Source Description: Google provides a community mobility report [50] which measures the per-weekday percentual change in user activity per location type (Retail & Recreation, Grocery & Pharmacy, Parks, Transit Stations, Workplace and Residential) of users with Google location services enabled in their mobile devices against a per-weekday baseline computed between 03-01-2020 and 06-02-2020 (pre-lockdown).

Apple provides a mobility trend report [51] which measures the daily percentual change of Apple maps application walking and driving direction requests in the city of Lisbon, Portugal against the baseline day of 13-01-2020 (a pre-lockdown Monday).

Moovit Insights Public Transit Index [52] provides a Lisbon 7-day public transportation ridership percentual change, measured with the Moovit App, from a baseline week computed for the week of 08-01-2020 to 15-01-2020.

Waze provides COVID-19 Impact dashboard [53] which measures the percentual change of kilometers driven per day measured with the Waze App within the city of Lisbon against a pre-computed baseline scoped to the weekday for the period between 11-02-2020 and 25-02-2020 (pre-lockdown). It is unknown if the Waze App user base in Lisbon is representative of all traffic types such as cargo heavy vehicles or light passenger vehicles, and in this sense, it might overestimate or underestimate the relationship of this indicator with expected road-based air pollutant contribution.

All four mobility indexes suffer from a number of problems that cause some bias in the reported urban mobility change.

Data Quality Issues All used mobility indexes suffer from a yearly seasonal bias because the baseline periods were computed for a period of time during winter weeks (January/February 2020) in which most of Lisbon population is usually working on a typical year. This baseline periods are then used to compare mobility pattern changes throughout the entire year of 2020, namely during usual vacation periods in summer months when urban mobility drops sharply, leading to distortions in reported mobility pattern changes.

Additionally, Apple mobility trend report suffers from an additional weekly seasonal bias because the baseline was computed from a single weekday (Monday).

Filtered Data: None

Aggregation performed: None. The original data-source is already at the daily temporal granularity.

Data Source Metadata: See Appendix B – Data Source Metadata

3.2.3. Lisbon Subway Ridership Data Source

Data Source Description: Lisbon daily subway ridership data per type of pass (occasional, regular, child, etc) was sourced from Metropolitano de Lisboa, the company who operates the Lisbon subway network, for the years of 2019 and 2020.

Data Quality Issues: Lisbon subway ridership data missed data between 17-03-2020 and 02-05-2020 because ticket validation machines were disabled to decrease queues, thus social contact, and interaction with ticketing machines and ticket validation machines. In this sense, those days were filled in by computing a per weekday baseline from the week 02-03-2020 to 08-03-2020, normalized against the Google Public Transit Lisbon Index, which was then used to infer the filled-in values.

Filtered Data: None

Aggregation performed: None. The original data-source is already at the daily temporal granularity.

Data Source Metadata: See Appendix B – Data Source Metadata

3.2.4. Lisbon International Airport Traffic Data Source

Data Source Description: Lisbon international airport (Humberto Delgado Airport) daily cumulative aircraft movements and non-transferred passengers were sourced from ANA (Aeroportos de Portugal) for the years 2019 and 2020.

Non-transferred passengers better describe passengers that either arrive in Lisbon as their last flight leg or departure from Lisbon as their initial flight leg and therefore discarding passengers that were in-transfer and therefore neutral for this analysis.

Data Quality Issues: None

Filtered Data: None

Aggregation performed: None. The original data-source is already at the daily temporal granularity.

Data Source Metadata: See Appendix B – Data Source Metadata

3.2.5. Lisbon Air Pollutant Concentration Data Source

Data Source Description: Hourly averaged air pollution measures were sourced from APA (Portuguese Agency for the Environment) via the EEA (European Environment Agency) for six urban air pollution stations, including three in high road traffic locations, from years 2013 to 2020.

The measures include concentration values (μg/m3) for primary pollutants NO_2 (Nitrogen dioxide), CO (Carbon monoxide), SO_2 (Sulphur dioxide) and secondary pollutants PM_{10}

Particulate Matter < 10μm of diameter, $PM_{2.5}$ (Particulate Matter < 2.5μm of diameter), O_3 (Ozone) and C_6H_6 (Benzene).

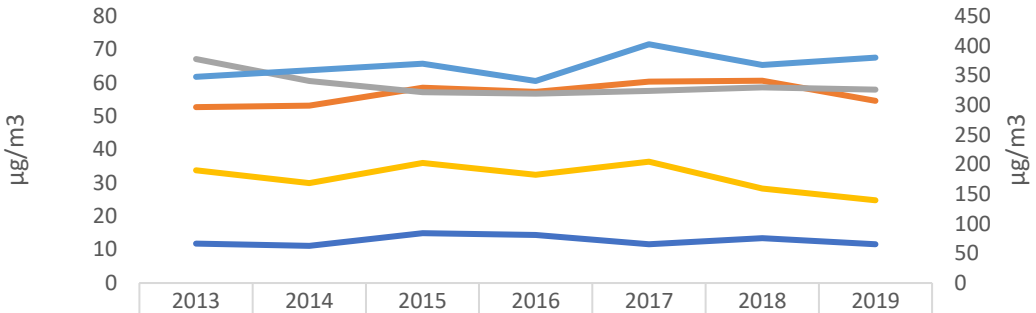
Data Quality issues: In the APA (Portuguese Environment Agency) air pollution dataset, for the $PM_{2.5}$ aerosol on background pollution station there's no data for January 2020, for the CO pollutant there's no data for 2014. Furthermore, Benzene measurements in Lisbon monitoring stations have multiple massive gaps throughout multiple years and months and therefore will not be reported in detail.

A number of comparisons of 2020 air pollution levels with previous years (2013-2019 baseline) are carried out in order to understand the magnitude of air pollutant concentration change during the several stages of the COVID-19 pandemic handling by Portuguese authorities.

To understand any potential bias of this type of comparative analysis, for instance if air pollution levels changed significantly throughout the 2013 to 2019 period due to environmental policies, an historical worst station yearly average concentration per air pollutant is provided (Figure 7).

Most air pollutants have had irregular trends in the 2013-2019 period with the exception of PM_{10} which appears to follow a decrease trend since years 2018 and 2019 and Ozone (O_3) which decreased in 2013 and 2014 and stabilized in the following years. It may affect comparative analysis since an organic downward trend was already in progress.

Moreover, a similar yearly worst station average per air pollutant analysis is provided by CCDR-LVT (Lisbon and Tagus Valley Regional Development Coordination Commission) with additional temporal span (2001-2019) which yields the same conclusion [10].



	2013	2014	2015	2016	2017	2018	2019
NO2	52.766651	53.237609	58.626446	57.344602	60.399414	60.687698	54.611514
O3	67.186	60.623399	57.254034	56.802762	57.663646	58.664607	58.03802
PM10	33.795172	29.925618	35.981193	32.425602	36.400589	28.376249	24.815447
PM25	11.825775	11.147168	14.948739	14.411583	11.599504	13.486851	11.654915
CO	348.15308		369.92151	340.90194	402.9269	367.71354	380.21559

Figure 7 – Lisbon worst station yearly average concentration per air pollutant

Filtered Data: For NO_2, SO_2, O_3, PM_{10} and $PM_{2.5}$, measures with concentration equal to 0 µg/m3 were filtered from the analysis. For all pollutants, only samples with validation status equal to 1 were kept.

Aggregation performed: The data-source reports air pollutant concentration sampling at the hour granularity. To integrate in the consolidated data model, a temporal aggregation to the daily grain was conducted. The resulting metrics include the minimum, maximum, average, quartile 1, quartile 2 (median) and quartile 3 concentrations values at the daily level.

Data Source Metadata: See Appendix B – Data Source Metadata

3.2.6. Portugal COVID-19 Pandemic Report

Data Source Description: COVID-19 pandemic related health indicators

The data is sourced from DGS (Direção Geral de Saúde) but it is currently gathered, integrated, and managed by a DSSG-PT (Data Science for Social Good Portugal) in their public repository [54]. It includes indicators such as the number of confirmed COVID-19 cases, COVID-19 patients admitted to hospital wards and ICUs, COVID-19 related deaths and many other indicators. The data has a daily temporal granularity and also has

many of the indicators broken down by geographical regions, spanning from the very beginning of the pandemic in Portugal until the present date of this work.

Data Quality Issues: None

Filtered Data: None

Aggregation performed: None. The original data-source is already at the daily temporal granularity.

Data Source Metadata: See Appendix B – Data Source Metadata

3.2.7. Lisbon and Tagus Valley Mortality Report

Data Source Description: Death certificate information system

The data is sourced from DGS (Direção Geral de Saúde) SICO/EVM system and provides consolidated data on all deaths registered in Portugal as result of the digitalization of death certificates spanning a period including the years from 2014 to 2020. This data-source allows the analysis of registered deaths by type and geography, which for the present work is the Lisbon and Tagus Valley health region, namely excess mortality, or effectiveness of restrictive measures in the context of the COVID-19 pandemic.

Data Quality Issues: None

Filtered Data: Only date for LVT health region was kept.

Aggregation performed: None. The original data-source is already at the daily temporal granularity.

Data Source Metadata: See Appendix B – Data Source Metadata

3.2.8. Bike Lane Counter (Av. Duque de Ávila)

Data Source Description: Bi-directional bike lane counter in Av. Duque de Ávila

The data source is the result of the count of bike passages through a sensor owned by the municipality in each of the directions (east/west) of a busy bike lane in the center of Lisbon, which aims to be a sample of the usage of bikes in the remainder of the city, in the period spanning January 2019 to October 2020.

Data Quality Issues: Data gap from 16-04-2019 to 04-07-2019 doesn't allow a complete homologous analysis. The expanding bike lane network, owned bicycles and shared bike programs might impact the homologous analysis since the baseline bike users are different.

Filtered Data: None.

Aggregation performed: None. The original data-source is already at the daily temporal granularity.

Data Source Metadata: See Appendix B – Data Source Metadata

3.2.9. Shared Bike Trips (GIRA)

Data Source Description: Total number of GIRA trips by day

The data source provides a daily trip count of GIRA rides spanning from January 2019 to December 2020. GIRA is a municipality electric and regular shared bike service with pick-up/drop-off stations spanning large swaths of Lisbon municipality for regular and occasional bikers.

Data Quality Issues: There are three days with data gaps (20/03/2020, 09/05/2020 and 10/05/2020) The expanding bike lane network, owned bicycles and shared bike programs might impact the homologous analysis since the baseline bike users are different.

Filtered Data: None.

Aggregation performed: None. The original data-source is already at the daily temporal granularity.

Data Source Metadata: See Appendix B – Data Source Metadata

3.2.10. Data Source Summary

A high-level summary of the most relevant data-sources used in the present work can be found in Table 5.

Table 5 – Data Sources

Entity	Dataset	Grain	Period	Short description
IPMA	Meteorological parameters	Hour	2019-2021*	The measures include wind speed (m/s) and direction (º), relative humidity (%), air temperature (ºC), sea level air pressure (hPa), total global radiation (KJ/m2) and accumulated precipitation (mm)
APA/EEA	Air Pollutants Concentration	Hour	2013-2021*	Concentration values (µg/m3) for primary pollutants NO_2 (Nitrogen Dioxide), CO (Carbon Monoxide), SO_2 (Sulphur Dioxide) and secondary pollutants PM_{10} Particulate Matter < 10µm of diameter, $PM_{2.5}$ (Particulate Matter < 2.5µm of diameter) and O_3 (Ozone).
Apple	Apple mobility trend report	Daily	2020-2021*	Daily percentual change of Apple Maps App walking and driving direction requests in the city of Lisbon, Portugal against the baseline day of 13-01-2020 (a pre-lockdown Monday).
Google	Community mobility report	Daily	2020-2021*	Per-weekday percentual change in user activity per location type (Retail & Recreation, Grocery & Pharmacy, Parks, Transit Stations, Workplace and Residential) of users with Google location services enabled in their mobile devices against a per-weekday baseline computed between 03-01-2020 and 06-02-2020 (pre-lockdown).
Moovit	Moovit Insights Public Transit Index	Daily	2020-2021*	Lisbon 7-day public transportation ridership percentual change, measured with the Moovit App, from a baseline week computed for the week of 08-01-2020 to 15-01-2020 (pre-lockdown).
Waze	COVID-19 Impact dashboard	Daily	2020-2021*	Percentual change of kilometers driven per day measured with the Waze App within the city of Lisbon against a pre-computed baseline scoped to the weekday for the period between 11-02-2020 and 25-02-2020 (pre-lockdown).
Metropolitano de Lisboa	Lisbon subway rides	Daily	2019-2020	Lisbon subway network daily cumulative rides (one-way trips) per type of pass (occasional, regular, child, etc)
ANA Portugal	Lisbon international airport traffic	Daily	2019-2020	Lisbon international airport (Humberto Delgado Airport) daily cumulative aircraft movements (inbound, outbound) and non-transferred passengers (passengers just transiting).
DGS	COVID-19 pandemic report	Daily	2020-2021*	COVID-19 Pandemic related health indicators communicated by public entities broken down by geographical context.
SICO/EVM	Lisbon and Tagus Valley Mortality	Daily	2020-2021*	Consolidated data on all deaths registered in Portugal broken down by death type, age group and by geography.
EMEL	GIRA bike ridership	Daily	2019-2020	Daily Lisbon GIRA shared bike program ridership.

* Incomplete year

3.2.11. Consolidated Data Model

As previously described in chapter 1.5, after the data extraction, transformation and loading process, the resulting analytical data model loaded into PowerBI was used to support all of the visualizations and consolidated data extractions for machine learning purposes. A simplified diagram of the data model can be found in Figure 8.

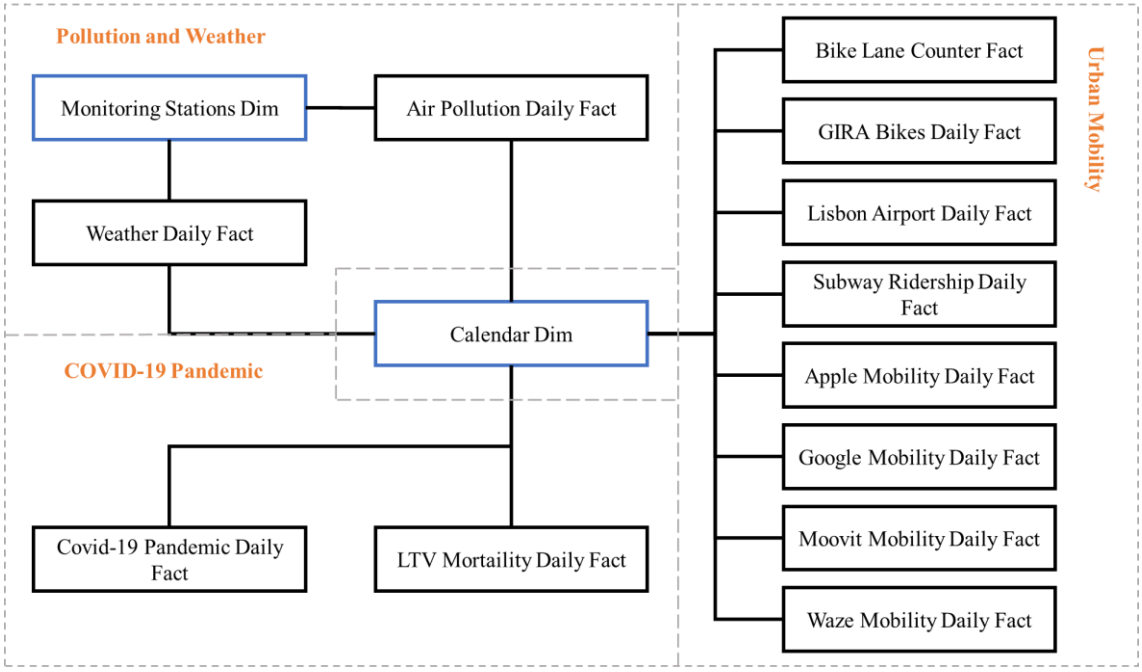


Figure 8 – Simplified analytical data model built in the present work

3.3. Temporal and Spatial Scope of the Analysis

The present work is based on several data sources describing weather, air pollution and urban mobility during the pre-pandemic and pandemic period in Lisbon Metropolitan Area, Portugal. The weather and air quality monitoring stations that were used in this work are depicted in Figure 9 and described in Table 6.

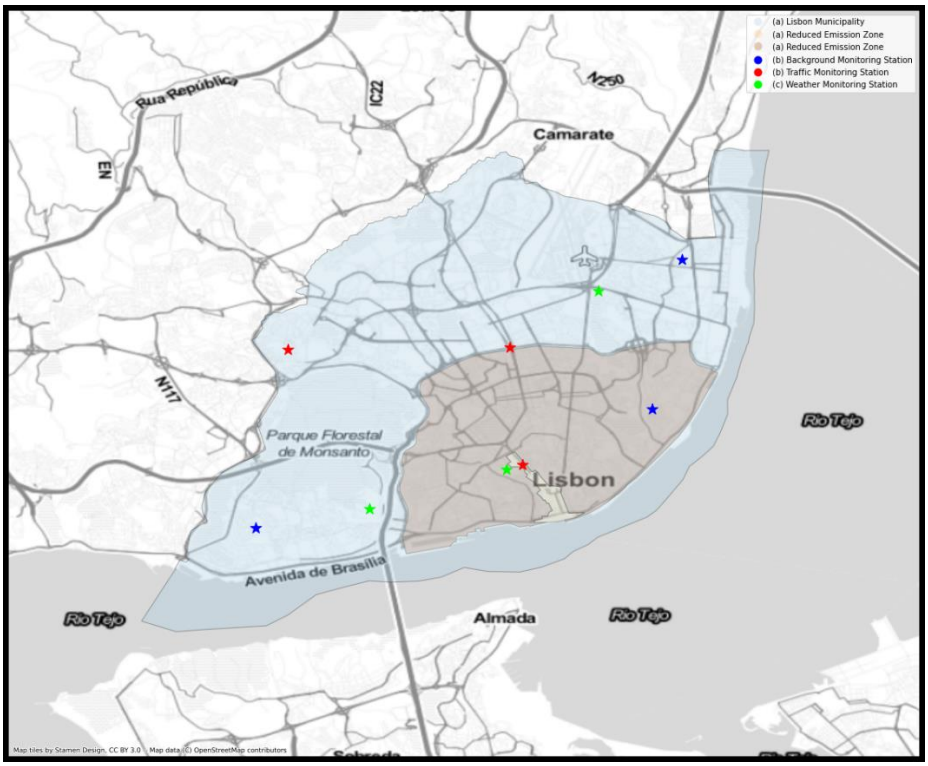


Figure 9 – Lisbon municipality geography boundaries, its reduced pollution emission zones and air pollution, weather monitoring stations.

Table 6 – List of meteorological and air quality monitoring station

Station	Entity	Altitude	Location (Lat, Lon)	Station Type
Benfica (T)	APA	76 m	38.748787, -9.201764	Traffic Air pollutant monitoring station
Av. Liberdade (T)	APA	44 m	38.721149, -9.146152	Traffic Air pollutant monitoring station
Entrecampos (T)	APA	86 m	38.748567, -9.149012	Traffic Air pollutant monitoring station
Beato (B)	APA	56 m	38.733686, -9.114497	Background Air pollutant monitoring station
Olivais (B)	APA	32 m	38.769783, -9.107292	Background Air pollutant monitoring station
Restelo (B)	APA	143 m	38.705738, -9.209461	Background Air pollutant monitoring station
Gago Coutinho	IPMA	140 m	38,766202,-9,127494	Metereological station
T. da Ajuda	IPMA	62 m	38,709561,-9,182825	Metereological station
I. Geofísico	IPMA	77 m	38,719077,-9,149722	Metereological station

As for the temporal scope, the Portuguese government enacted different lockdown restrictions in several periods and geographies throughout the year of 2020 and 2021 to contain the spread of COVID-19 pandemic. Each type of restriction implements different mechanisms that aim to reduce the spread of the disease. These types of restrictions have different effects on anthropogenic activities that produce air pollutants such as driving internal combustion vehicles, heating, industrial or energy production.

The present work mainly focuses on the months between March and July 2020, where the two main confinement periods occurred, to understand the impacts of restriction measures introduced by the authorities to manage the pandemic in the urban environment and mobility as summarized in Table 7. The preemptive approach to the management of the pandemic by the Portuguese government, which included strict lockdowns, resulted in limited spread of COVID-19 in the Lisbon and Tagus Valley health region as depicted in Figure 10.

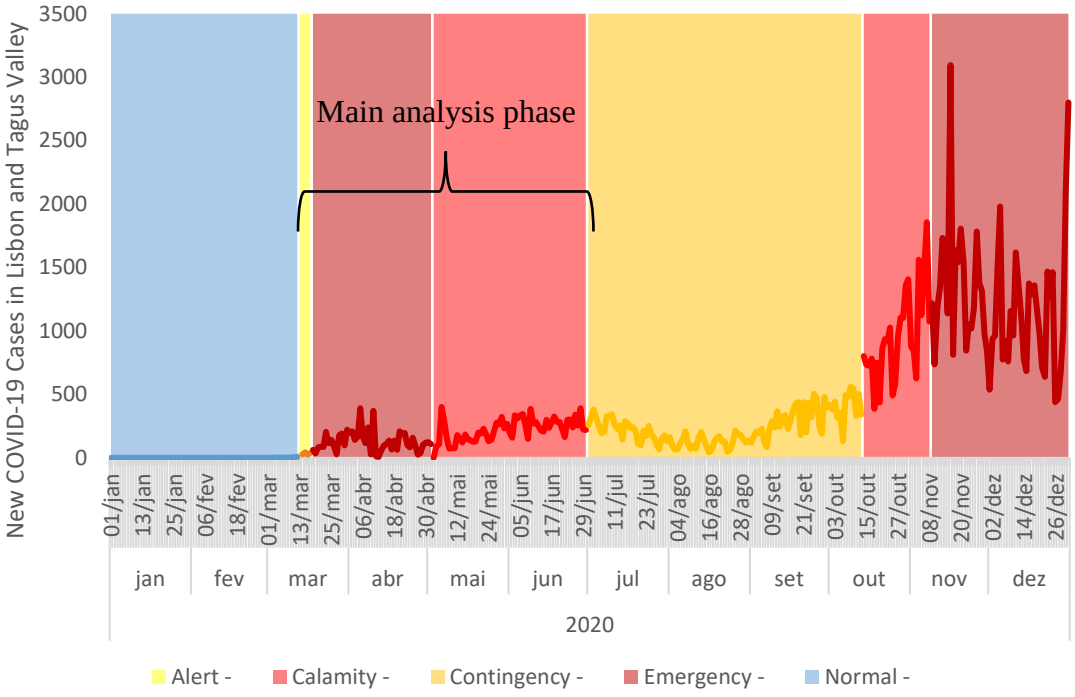


Figure 10 – Main phases of the pandemic management in the Lisbon Metropolitan Area.

Table 7 – Description of the main phases of the pandemic management in the Lisbon Metropolitan Area

Phase	**Start / End**	**Description**
Pre-confinement	04-02-2020 to 13-03-2020	Since the first suspected cases in early February 2020 until generalized lockdown in mid-March 2020, several governmental decisions restricting international travel, large scale public events, some public services and hospital visits. Several reported companies started moving to a telework even before it was mandated by the government.

1st National General Confinement	14-03-2020 to 03-05-2020	Schools were closed in 12-03-2020 and telework was generalized in both public and private sector where 13-03-2020 was the last office day for most workers. National Alert state decreed in 13-03-2020 and shortly after the Emergency State was declared in 18-03-2020 where the government enacted a number of restrictive measures to stop the spread of COVID-19. With most in-person public services, social, cultural, and economic activities related to tourism, restaurants, shopping, and general leisure closed, as well as limitation of mobility without proper authorization, urban mobility declined sharply.
1st National Calamity State	04-05-2020 to 01-07-2020	A two week 3-phase, gradual reopening plan was implemented where mask mandates were introduced in public transportation, public services, schools, commerce, restaurants, and cafes reopened with additional sanitary rules, occupancy limitations and limited schedules. Telework, where possible, was still mandatory. High risk activities such as football fans in the stadiums, concerts, and nightclubs, were not allowed. Large social agglomerations were still not allowed (and dispersed if needed) and civic duty of confinement was recommended to every citizen.

3.4.Mobility Analysis during COVID-19 Pandemic

In March 2020, the Portuguese government mandated lockdowns required most of the population to remain at home. In Lisbon, public transportation ridership and trips using private car and soft transport means, such as bicycle or walking, were severely impacted as a result.

The Moovit App, which is used to navigate metropolitan mass public transportation networks, provides a rolling 7-day ridership percentual change for Lisbon city measured with data from Moovit App usage, against a baseline week computed for the week of 08-01-2020 to 15-01-2020 (pre-lockdown).

According to this public transportation mobility index depicted in Figure 11, the most affected period was 09-04-2020 to 16-04-2020, where the 7-day rolling percentual change against the baseline estimated a 78.30% decrease in ridership. Even after the first state of

emergency ended on the 2nd of May 2020, demand for public transportation recovered only slightly to values far lower than from pre-pandemic periods, potentially due to general practice of telework, large scale lay-offs, reduced offer, and capacity of PT equipments, inexistent tourism as well as fear of contagium due to the enclosed nature of the public transportation equipments. During the first national emergency period (18-03-2020 to 03-05-2020) the mean reduction in public transportation demand is estimated at 75.63% against the index baseline. The index would only come to reach the pre-pandemic baseline levels on the 13th of July 2021.

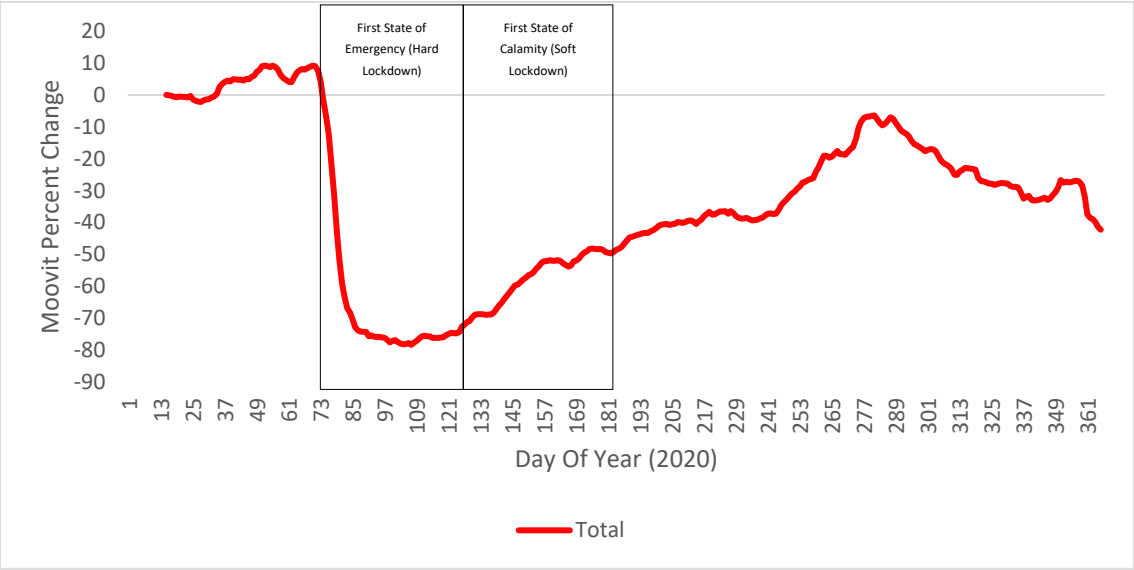


Figure 11 – Moovit Insights Public Transit Index for Lisbon, 2020

According to the Apple Mobility Trend Report, which uses Apple Maps data from the navigation feature, the daily percentual change of direction requests in the city of Lisbon, Portugal against the baseline day of 13-01-2020 (a pre-lockdown Monday) provides details on other urban mobility means, namely walking and driving, which are two of the Apple Maps app modes.

According to this index, the day with largest reduction of walking navigation requests was on the 29th of March 2020 with a reduction of 92.60% and on the 12th of April for driving navigation requests with a reduction of 86.02% as depicted in Figure 12. These results must be interpreted cautiously since the baseline period for comparison is a pre-lockdown Monday and both the dates with largest reductions during the lockdowns are

Sundays which are days where there's naturally reduced urban mobility associated to the weekends.

Nevertheless, the results are consistent with the high initial adherence to lockdowns by the population, where in the first national emergency state (18-03-2020 to 03-05-2020) walking trips reduced in 89% and driving trips reduced in 78.84% against baseline, while during the first national calamity state (04-05-2020 to 01-07-2020) walking trips reduced in 72% and driving trips reduced in 43.05% against baseline. From this moment onwards, lockdown erosion and continuous de-escalation of confinement contributed to a steady increase in both mobility patterns until the summer.

From the 1st of August 2020 onwards, the remaining 19 AML parishes still in calamity state, which still had special mobility restrictions due to the virus incidence, joined the rest of the AML to the lower state of contingency. This appears to be correlated with a sharp increase in walking trips starting in early August, with additional potential contributions from tourism, which peaked in August 2020, and the vacation periods that might have been spent closer to home.

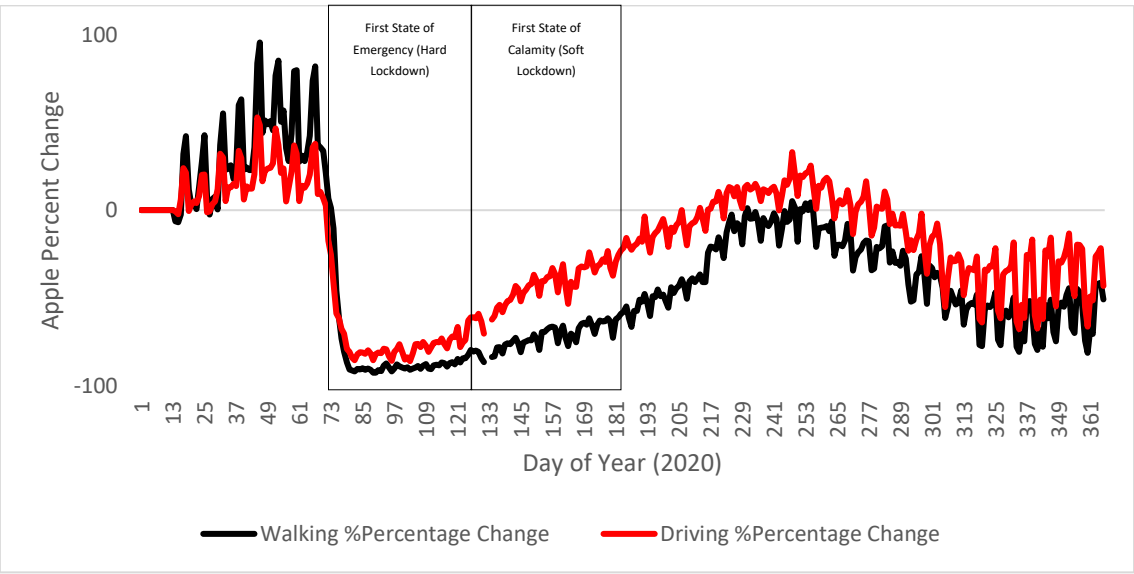


Figure 12 – Apple Mobility Trend Report for Lisbon, 2020

While Apple Mobility Trend focus on the number of direction requests by users, the Waze App COVID-19 Impact Index calculates the percentual change of kilometers driven per day measured with the Waze App within the city of Lisbon against a pre-computed

baseline scoped to the weekday for the period between 11-02-2020 and 25-02-2020 (pre-lockdown). An important aspect of a measure related to driven distance is that it is expected to correlate better to vehicle related pollution emissions than the number of trips.

According to this index, the day with the highest reduction of driven kilometers when compared to the respective baseline was the 10th of April 2020 with a reduction of 94% in driven kilometers as depicted in Figure 13. It then follows a similar distribution to the Apple Driving index but with a negative offset that widens from the beginning of the lockdown period until the summer period, potentially due to the fact that it measures driven kilometers and not number of trips, whereas it is possible that the average driven kilometers per trip during the lockdown periods were overall lower due to the more direct trips and non-leisure use of the private vehicle.

During the first national emergency period (18-03-2020 to 03-05-2020), it registered a 75.63% decrease in the daily driven kilometres registered with the Waze App while the first state of calamity (04-05-2020 to 01-07-2020), which was marked by lockdown erosion and continuous de-escalation of confinement, registered a 56.13% decrease against the baseline.

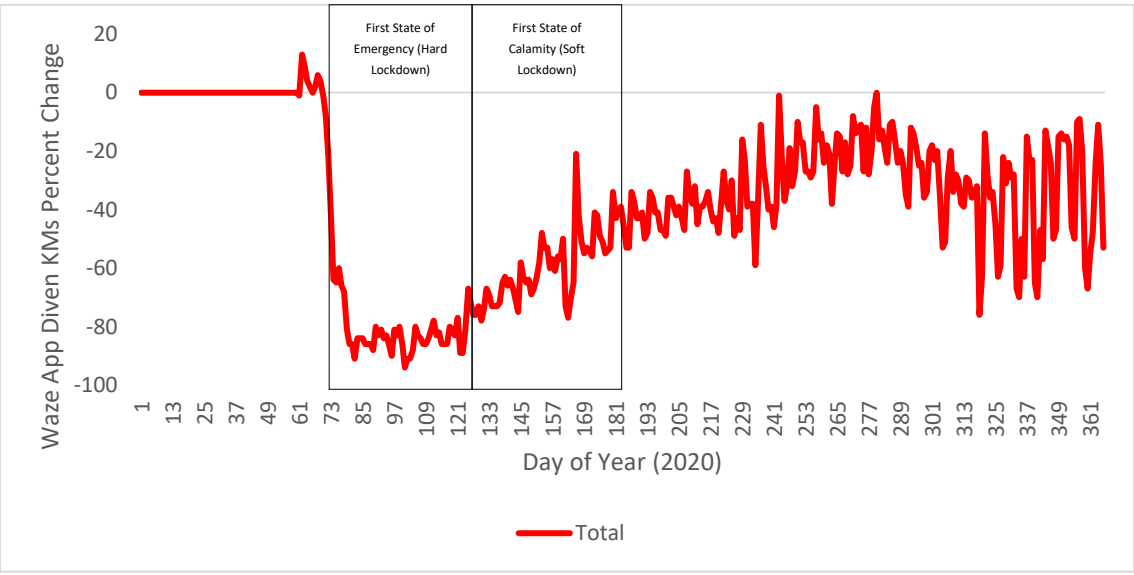


Figure 13 – Waze COVID-19 Impact Index for Lisbon, 2020

Google Community Mobility Index is a different mobility index type since it associates user activity to a location type instead of movement distances or the number of

movements. It derives a per-weekday percentual change in user activity per location type (Retail & Recreation, Grocery & Pharmacy, Parks, Transit Stations, Workplace and Residential) from data of users with Google location services enabled in their mobile devices against a per-weekday baseline computed between 03-01-2020 and 06-02-2020 (pre-lockdown).

According to this index depicted in Figure 14, the days with the sharpest reduction of activity in Grocery and Pharmacy (Commerce) was the 12th of April 2020 with a decrease in activity of 84%, for Workplace was the 10th of April 2020 with a decrease of 89%, for Parks was the 5th of April 2020 with a decrease of 91%, for Retail and Leisure was the 12th of April 2020 with a decrease of 91% and for Public Transportation was the 10th of April 2020 with a decrease of 90%. On the other hand, for Residential activity there was a general increase peaking on the 10th of April 2020 with a maximum of +47%.

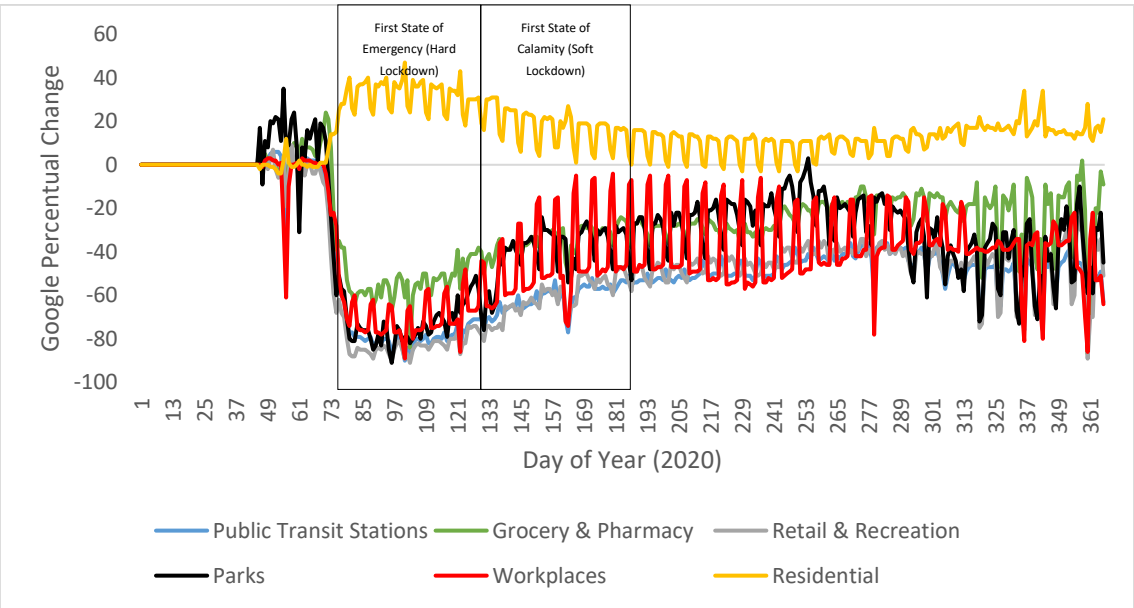


Figure 14 – Google Community Mobility Report for Lisbon, 2020

Lisbon subway network ridership was severely reduced as a consequence of COVID-19 restrictions beginning in March 2020 as depicted in Figure 15. As one of the three main public transportation modes used in the city, it showcases the dramatic effect of the lockdown in the city.

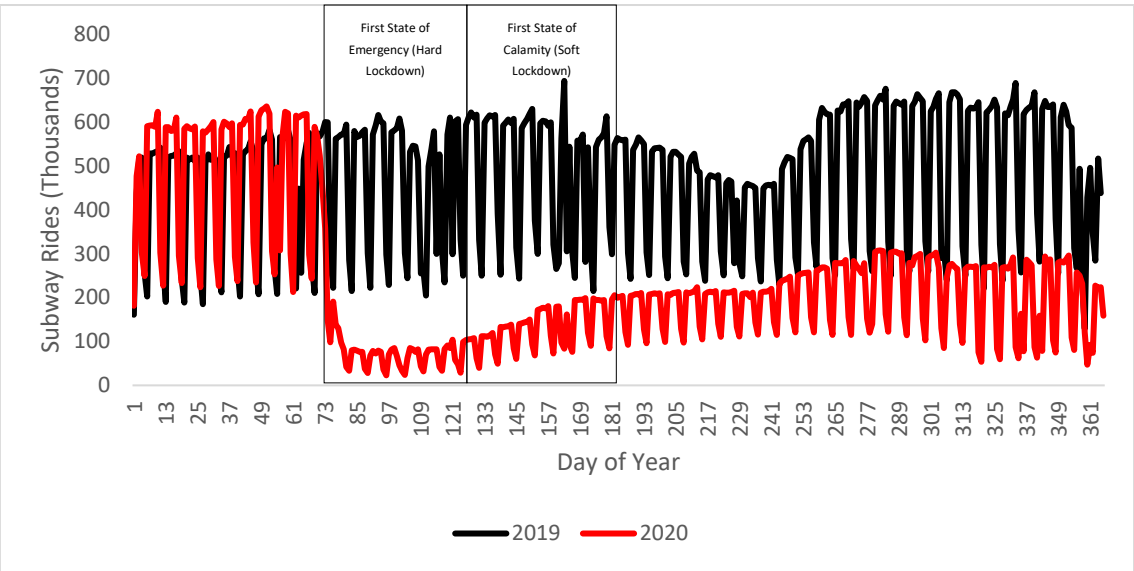


Figure 15 – Lisbon Subway Ridership for 2019 and 2020

The YoY analysis should be done with care since from September 2019 onwards a change in the prices of recurrent titles [55] increased the baseline usage of the subway network starting in April 2019. The first two months of 2020 where no restrictions were in-place, when compared to the homologous month in 2019 show that regular passes had an increase in ridership of 23%, free child passes 39% and overall ridership of 13%. In this sense two analysis are provided, an YoY monthly percentual change for 2019 and 2020, as well as a monthly percentual change against a static baseline (January/2020).

As depicted in Figure 16, the most impacted month was April/2020, with a homologous decrease in total ridership of 84% when compared to 2019. For the same month, when compared to January/2020 baseline, there's an 85% decrease in total ridership. The most impacted type of trip was the occasional title which measured a homologous ridership decrease of 92% and a decrease of 89% when compared to January/2020. This was expected since leisure and tourist ridership were virtually halted during the initial lockdowns. As detailed in section 3.2.3, the subway ridership data for April/2020 had to be filled-in via extrapolation since ticket validation was halted during the peak of the pandemic and thus should not be considered factual.

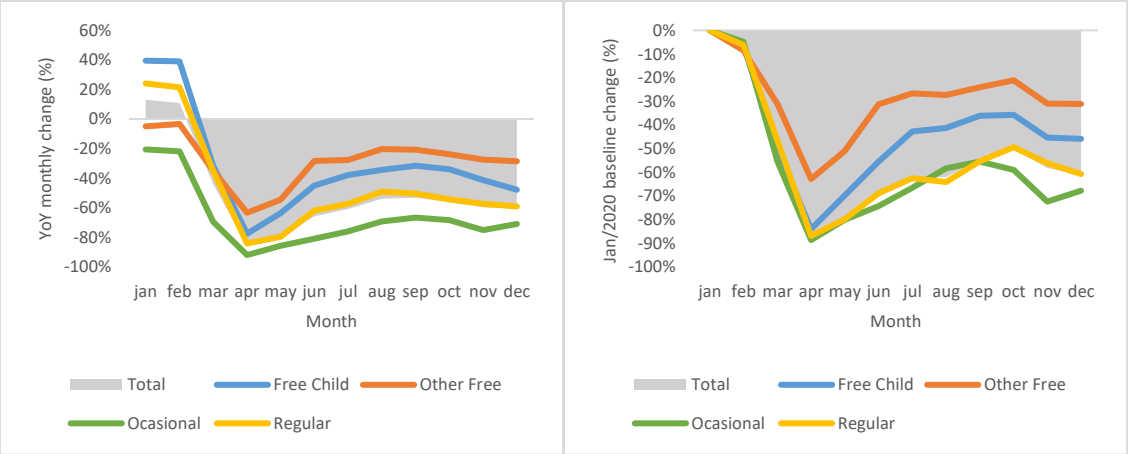


Figure 16 – Lisbon Subway (a) YoY monthly percentual ridership change 2019 and 2020 and (b) monthly percentual ridership change against January 2020

In the first state of emergency (18-03-2020 to 03-05-2020), GIRA bike trips fell by 68.47% when compared to the same period in 2019 but recovered during the first state of calamity (04-05-2020 to 01-07-2020) measuring an increase of 22.59% in ridership when compared to the same period in 2019 as depicted in Figure 17. After the first state of emergency, the usage of the shared GIRA bike during the first state of calamity increased in greater order of magnitudes than other types of urban transportation, namely the subway which still measured a 72.56% retraction against the same period in 2019. This could be attributed to increased demand for soft transportation means due to fear of contagium or because GIRA bike might also be used as an outdoor leisure activity and not only as a mean of transportation.

International non-resident business or leisure visitors make their way to Lisbon from their home countries primarily by means of air transportation [56]. Besides aircraft emitted air and noise pollution, visitors temporarily enlarge Lisbon population and therefore the human footprint in terms of anthropogenic pollutant activities and additional pressure on public transportation systems.

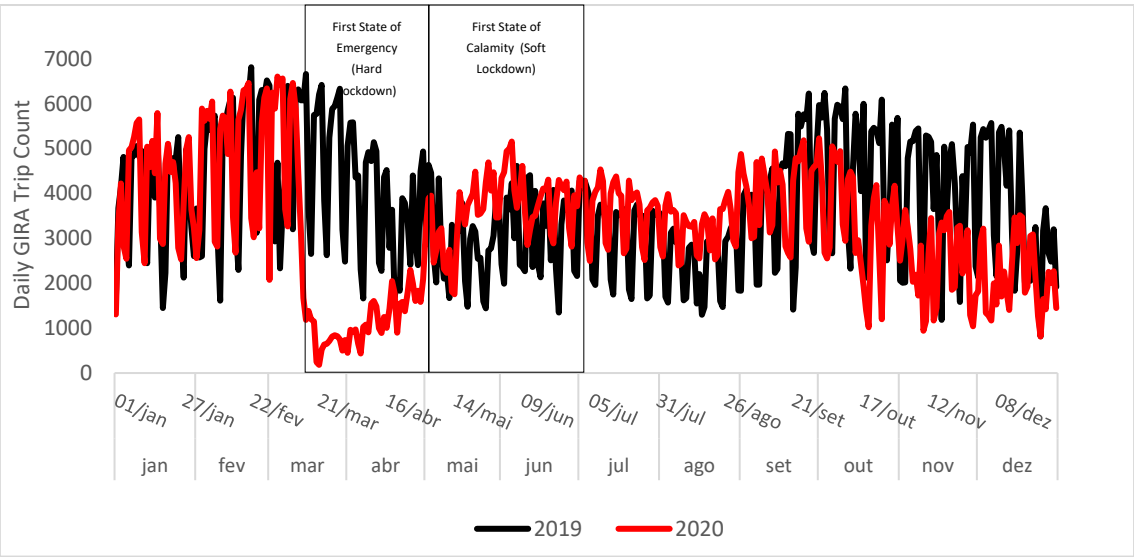


Figure 17 – Lisbon GIRA shared-bike ridership for 2019 and 2020

As depicted in Figure 18, the total non-transferred passengers in Lisbon airport during COVID-19 restrictions had a homologous reduction of 99.11% in the 15th week of the year and was measured to be only 4.423 passengers in 2020 (2020-04-05 to 2020-04-11) against 498.565 passengers in 2019 (2019-04-07 to 2019-04-13). Regarding the number of total aircraft movements, it had a homologous reduction of 96.44% in the same analysis period wherein 2020 there was only 152 movements against 4273 movements in 2019 and some of the few flights were related with citizen repatriation efforts.

The partial recovery from July 2020 onwards is related to the lifting of self-imposed travel restrictions by the Portuguese government [47] and additional relaxation of traveling rules from August 2020 onwards also had an effect [57]. The overall volume of passengers was also affected by travel restrictions imposed by the countries of origin of tourists or expatriates that visit Portugal (England, Germany, France, Italy, Spain, The Netherlands, etc.) and reciprocate restrictions of not allowing Portuguese citizens to travel to their countries.

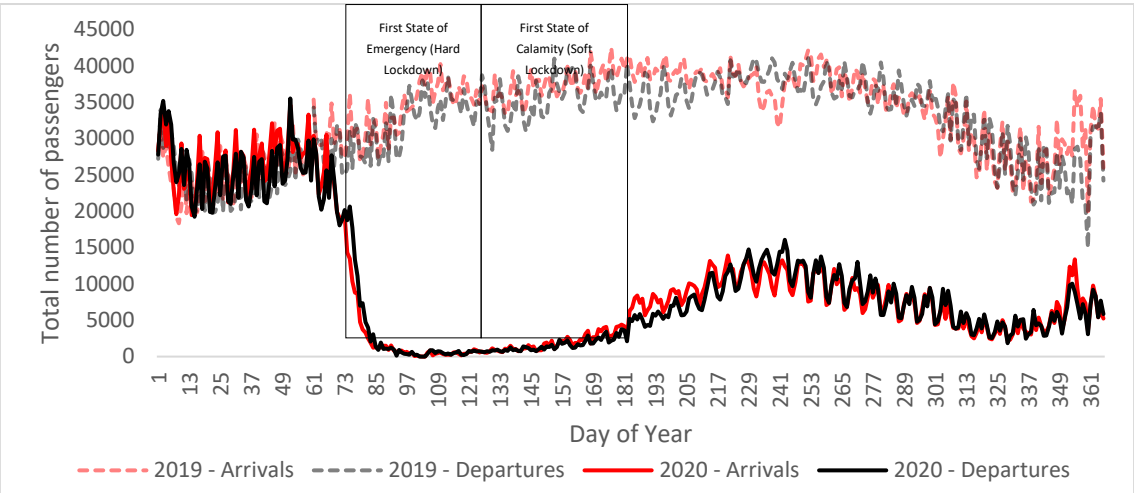


Figure 18 – Lisbon international airport non-transferred passengers for 2019 and 2020

By computing the Pearson Correlation of both public transportation mobility indexes for the year 2020 with an independent public transportation variable (subway ridership), it is found that Google Mobility Index for Public Transport correlates better than Moovit COVID-19 Public Transport Impact Index (r=0.98598 vs r=0.90812) and is thus a mobility index that correctly models public transport commuter routines in the city as shown in Figure 19. This could be useful not only to gauge changes in public transportation usage but to be used as a proxy indicator for public transportation usage in, for instance, the training of machine learning models.

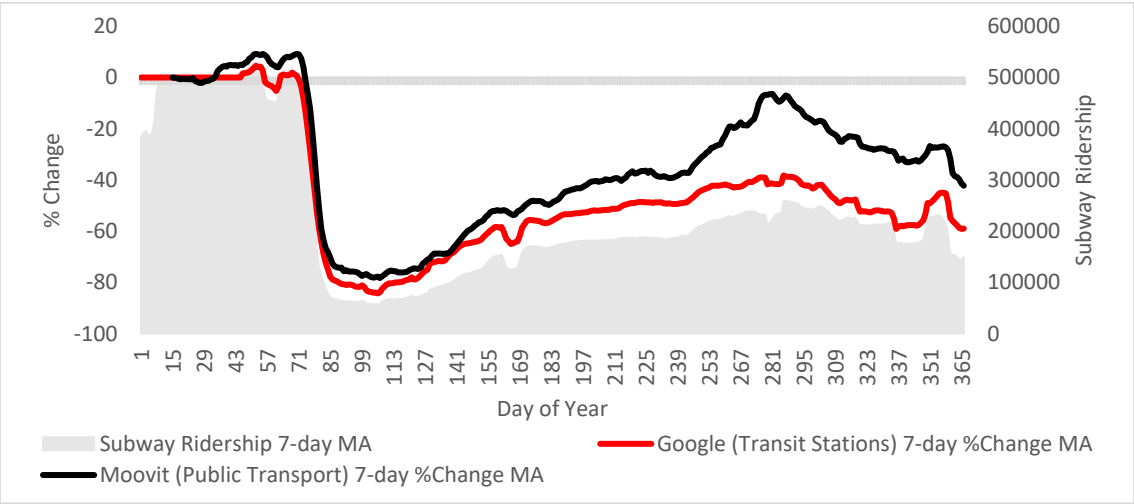


Figure 19 – Moovit and Google public transportation 2020 Mobility Indexes against an Independent Variable (Subway ridership)

A visual summary of the main impacts on urban mobility indicators in Lisbon during the first national emergency state period from 18-03-2020 to 03-05-2020 and during the first national calamity state period from 04-05-2020 to 01-07-2020 can be found in Figure 20.

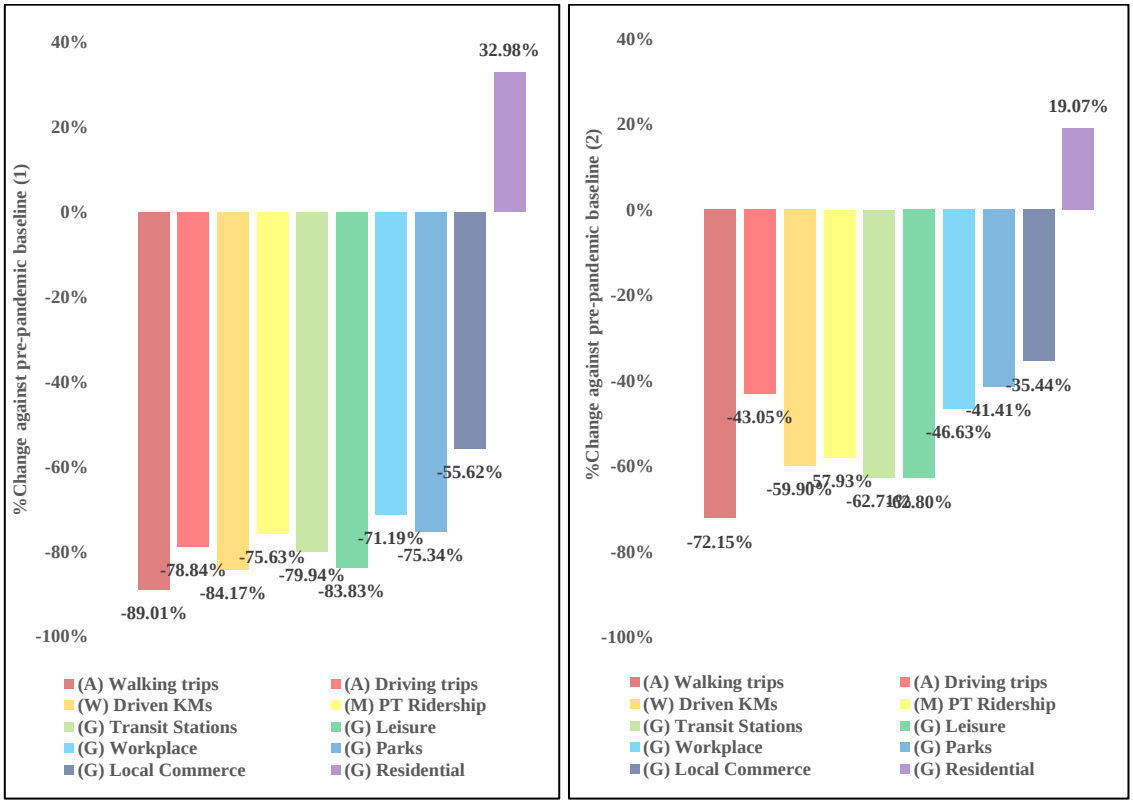


Figure 20 – Google, Apple, Moovit and Waze mobility index change during the 1st national emergency state period from 18-03-2020 to 03-05-2020 (1) and first national calamity state period from 04-05-2020 to 01-07-2020 (2) against a pre-pandemic baseline

3.5.Nitrogen Dioxide (NO2) performance during COVID-19 Pandemic

While the present work used in-situ measures acquired by ground air quality monitoring stations, one can use satellite observations to have a high-level view of the impact of the confinement measures on Nitrogen dioxide (NO_2), a pollutant emitted directly and indirectly by internal combustion vehicles. During the first national emergency state lockdown and the first half of the first calamity state (March-June), where the most restrictive measures affecting anthropogenic activities were in-place, a clear reduction on the density of the NO_2 vertical column measured by the TROPOMI instrument aboard Sentinel-P5 satellite when compared to the same period in 2019 as shown in Figure 21.

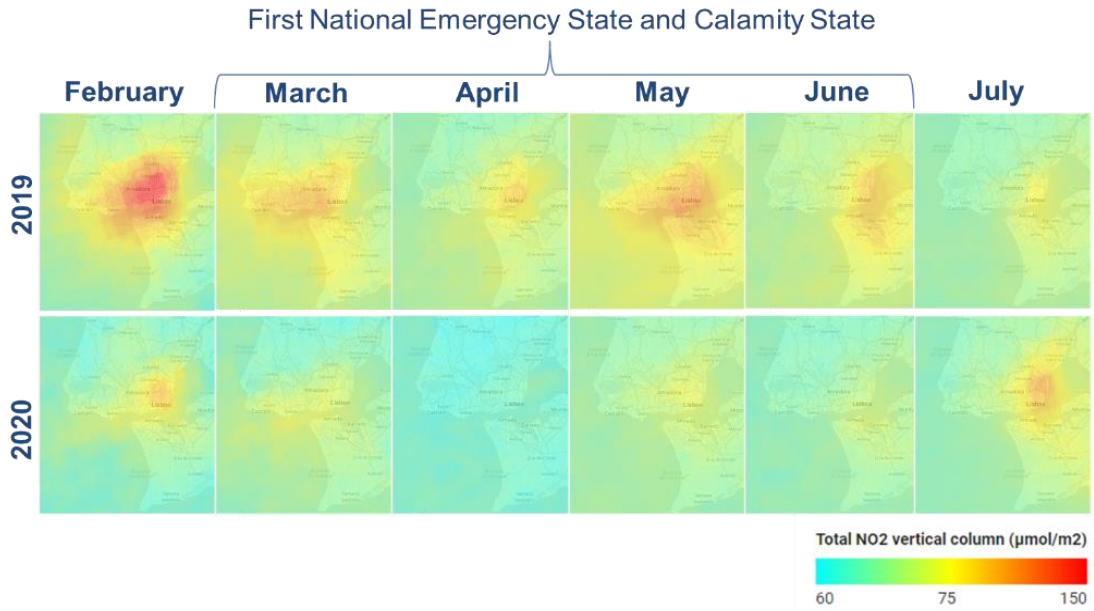


Figure 21 – Lisbon and Tagus Valley Monthly Average of Total NO2 vertical column (µmol/m2) measured by the TROPOMI instrument onboard Sentinel-P5 satellite.

For the same period, irrespective of the type of ground monitoring station (Background and Traffic), similar results can be observed. During the first national emergency lockdown and the first half of the first calamity state (March-June) NO_2 concentration levels measured at ground-level were severely reduced, having April/2020 reached a -46,96% difference when compared to the median NO_2 concentration for April 2013-2019 as depicted in Figure 22.

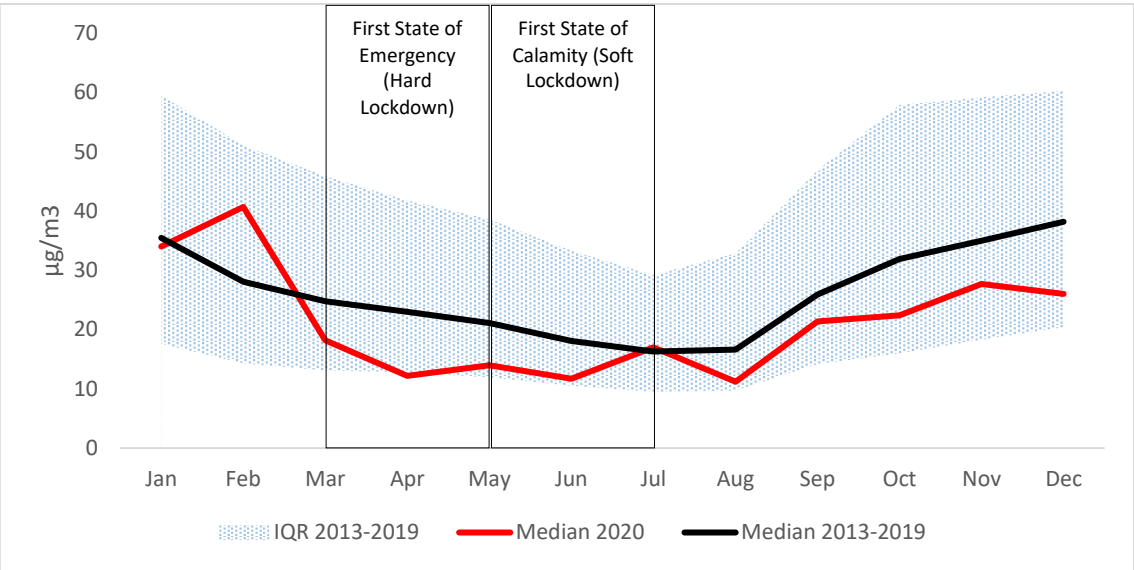


Figure 22 – All Lisbon stations monthly median NO2 concentration (µg/m3) for the years 2013-2019 and 2020.

When taking in consideration Lisbon air pollutant road traffic measurement stations it is also clear that during the first national emergency lockdown and the first half of the calamity state (March-June) NO_2 concentration levels measured at ground-level were severely reduced having April/2020 reached a -56,30% difference when compared to the median NO_2 concentration for April 2013-2019 (Figure 23). The drop is coherent with a sharp decrease in road transport contributions due to the imposed mobility restrictions.

Regarding Lisbon air pollutant background measurement stations during the first national emergency lockdown and the first half of the calamity state (March-June), NO_2 concentration levels measured at ground-level were moderately reduced, having April/2020 reached a -32,24% difference when compared to the median NO_2 concentration for April 2013-2019 (Figure 24).

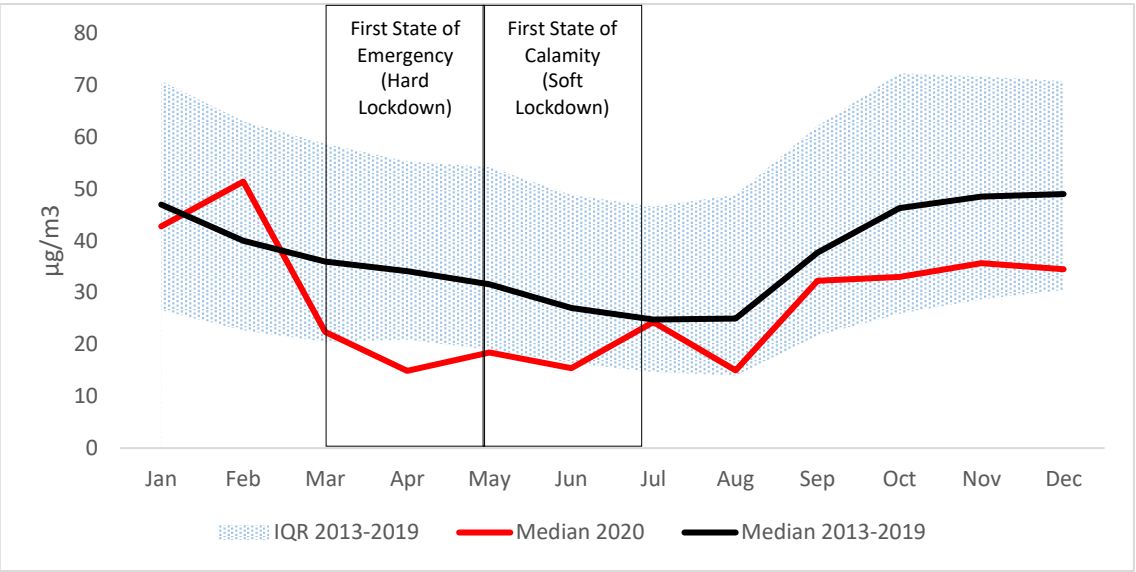


Figure 23 –Road traffic Lisbon monitoring stations monthly median NO2 concentration (µg/m3) for the years 2013-2019 and 2020.

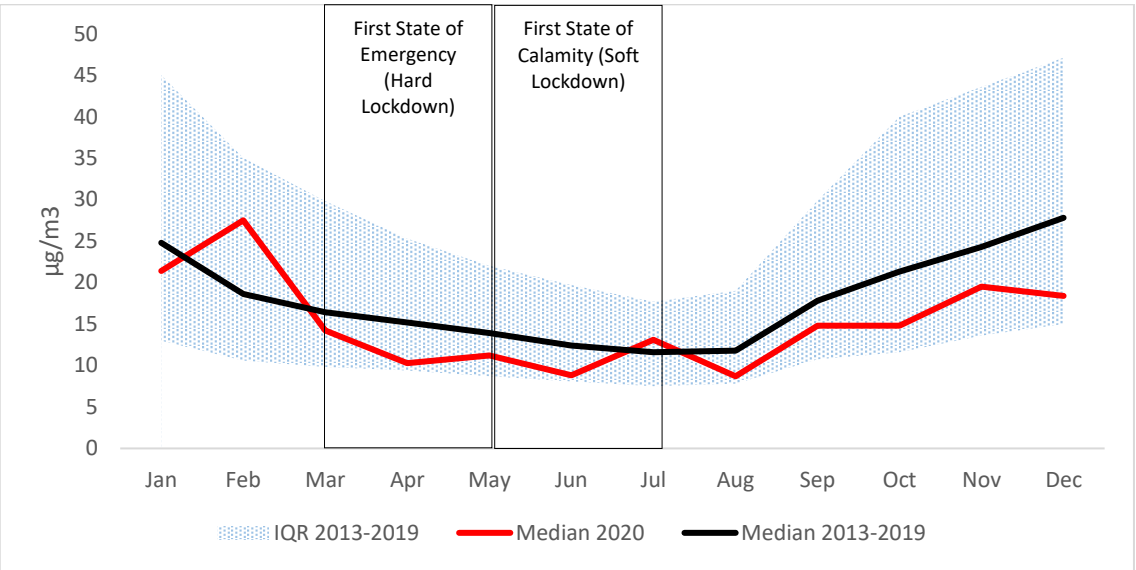


Figure 24 – Background Lisbon monitoring stations monthly median NO2 concentration (µg/m3) for the years 2013-2019 and 2020.

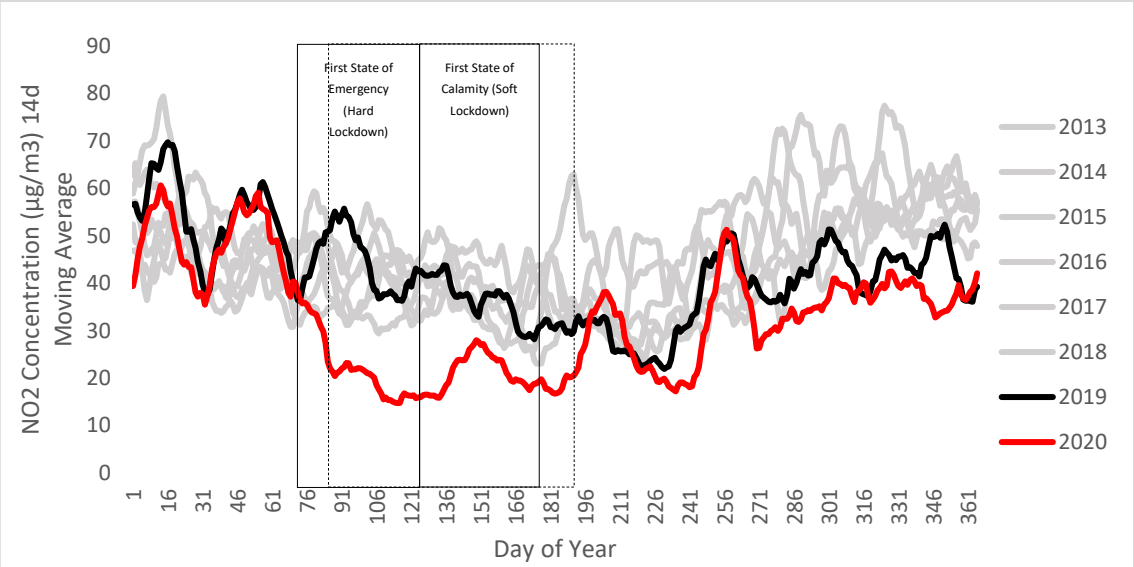


Figure 25 – Road traffic Lisbon monitoring stations NO2 concentration (µg/m3) 14-day Moving Aver-age per day of year for all years from 2013 to 2020.

The unlocking phase must be analyzed with care since the general end of lockdown entered the seasonal vacation phase of most workers and students (June-September) where historically NO_2 levels are lower. This means that two opposed forces are driving the NO_2 concentrations. A lower granularity multi-year comparative analysis with a daily 14-day moving average NO_2 concentration on Traffic monitoring stations allows a glance at the NO_2 performance during lockdown (Figure 25).

Anthropogenic activities are not the sole driver of NO_2 concentrations. Seasonal weather patterns, as well as anomalous weather-related episodes, contribute to higher or lower NO_2 concentrations and these phases are explained in Chapter 3.8 to better understand NO_2 concentrations throughout the pandemic phases under analysis.

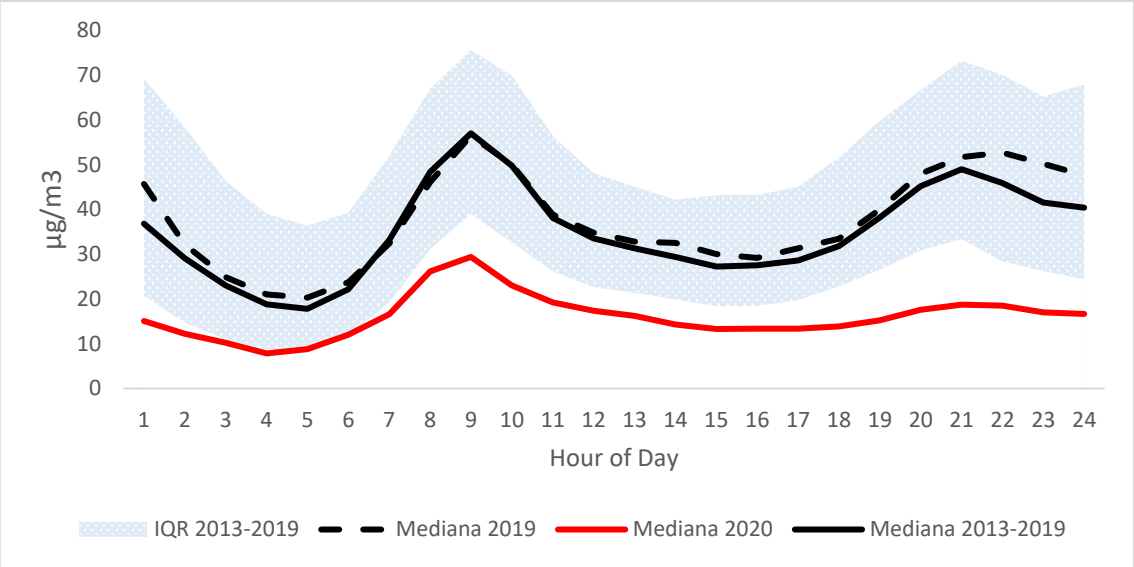


Figure 26 – Hourly NO2 median concentration in Traffic Stations for the National Confinement period (14-03-2020 to 03-05-2020) and homologous 2013-2019 period.

Because NO_2 concentrations in cities and towns are mostly related to internal combustion vehicles, the hourly NO_2 average concentration profile in Lisbon closely follows road traffic patterns where the most obvious feature is the home-work-home commuting dynamics, both intra-city and pendular, in the morning period when people go to work and in the end of the day when people go back home.

The hourly median NO_2 concentration profile during the 1st national lockdown period in traffic air quality monitoring stations is significantly lower than the same period in 2019 and 2013-2019 baseline period, peaking in excess of -60% between 18:00h and 20:00h. This happens because the curve itself changed to having a nearly flat post-morning concentration instead of a steep uptick at the end of the day when usually commuters are driving back home from work due to telework and layoff and general closure of in-person economic, social, cultural, and administrative activities. The morning commuting period is also significantly lower, peaking at -53,67% at 09:00h but still maintaining the uptick shape contributed by the regular non-road traffic sources and residual vehicle traffic pertaining to essential workers and services, goods deliveries, and general deliveries commerce provisioning (Figure 26).

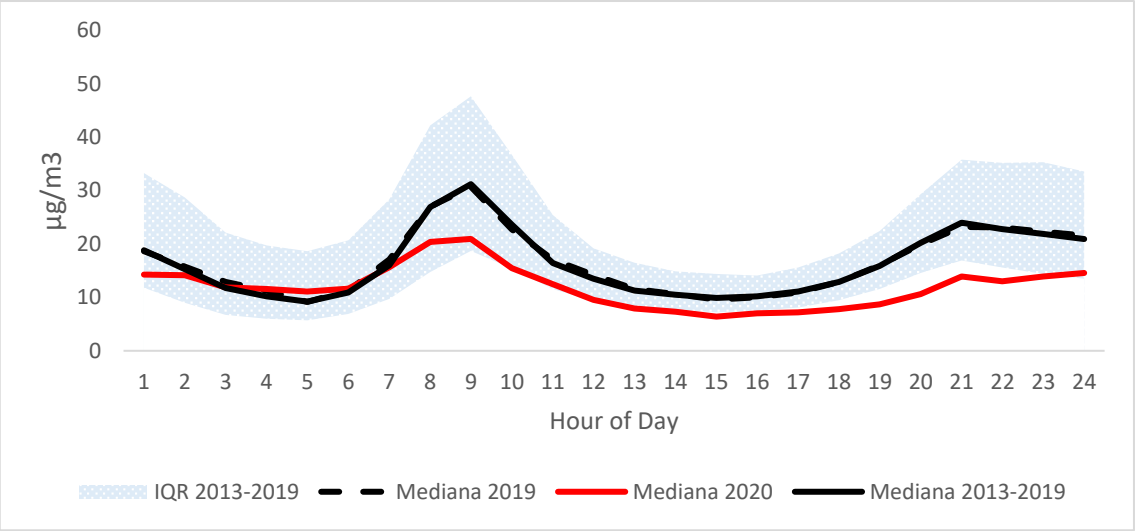


Figure 27 – Hourly NO2 median concentration in Background Stations for the National Confinement period (14-03-2020 to 03-05-2020) and homologous 2013-2019 period.

Regarding the background monitoring stations, the hourly median NO_2 concentration profile during the 1st national lockdown period is moderately lower when compared to the traffic stations as expected but still peaking below -40% between 18:00h and 21:00h when compared to the same period in 2019 and 2013-2019 baseline period (Figure 27).

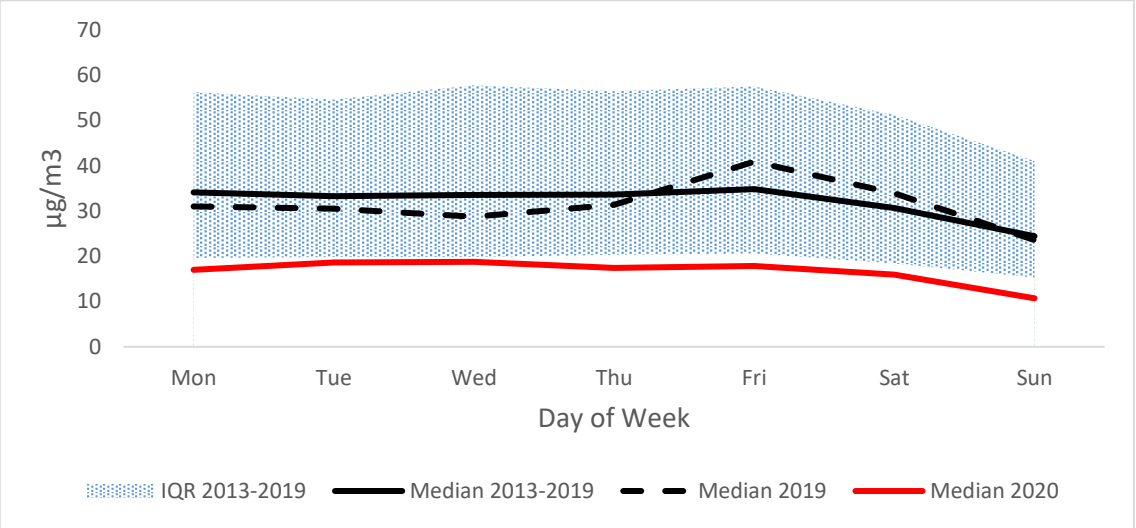


Figure 28 – NO2 median concentration in Traffic Stations per day of the week during the National Emergency period and Calamity State (14-03-2020 to 01-07-2020) and homologous 2013-2019 period.

Regarding the week-day profile of NO_2 concentrations, a larger period of time consisting of the first national emergency state and calamity situation, where both hard and soft lockdown measures were in place, was used to increase the analysis's statistical significance. As depicted in Figure 28, when comparing NO_2 median concentrations per weekday in traffic monitoring stations between the 2013-2019 baseline versus 2020 lockdown periods, it is possible to identify Sunday as the day with the largest difference peaking at -56,15% difference. This could be attributed to the general closure of supermarkets and other allowed commerce on Sundays, leaving little allowed reasons for driving in the city. Additional insight is related to the NO_2 concentration differences between weekdays and weekends, having Sundays in the lockdown period in analysis -40.26% than the average weekday of the same period, comparing to -27.90% during the same year period from the 2013-2019 baseline. By contrast, Saturdays in the lockdown periods in analysis were -11.22% of the average weekday of the same period, comparing to -9.57% during the same year period from the 2013-2019 baseline. This means that NO_2 concentrations during weekends were much lower compared to the baseline and the gap between weekend and working weekdays was widened. The same pattern can be generally verified for background stations although concentrations are much lower, as expected (Figure 29).

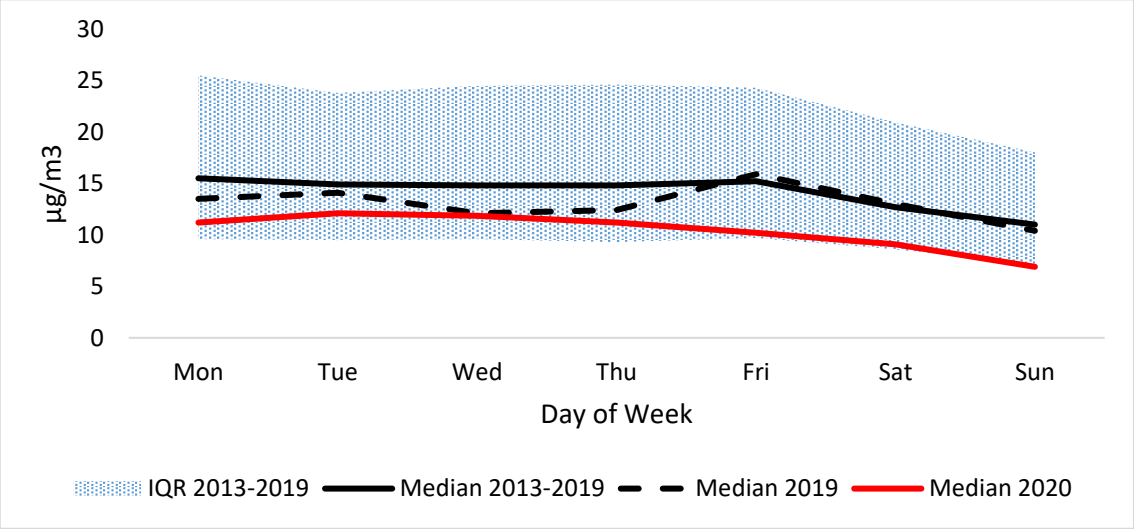


Figure 29 – NO2 median concentration in Background Stations per day of the week during the National Emergency period and Calamity State (14-03-2020 to 01-07-2020) and homologous 2013-2019 period

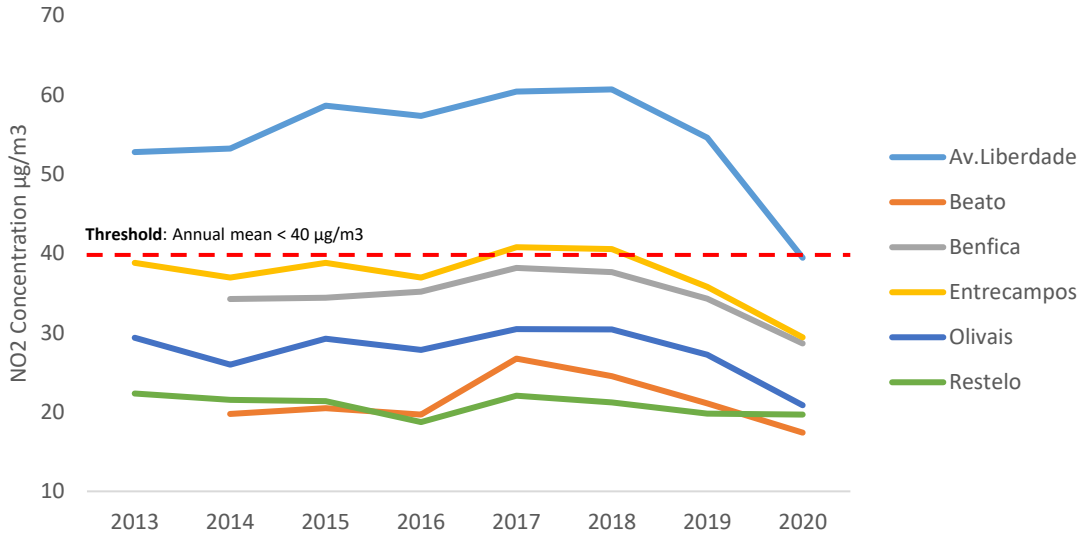
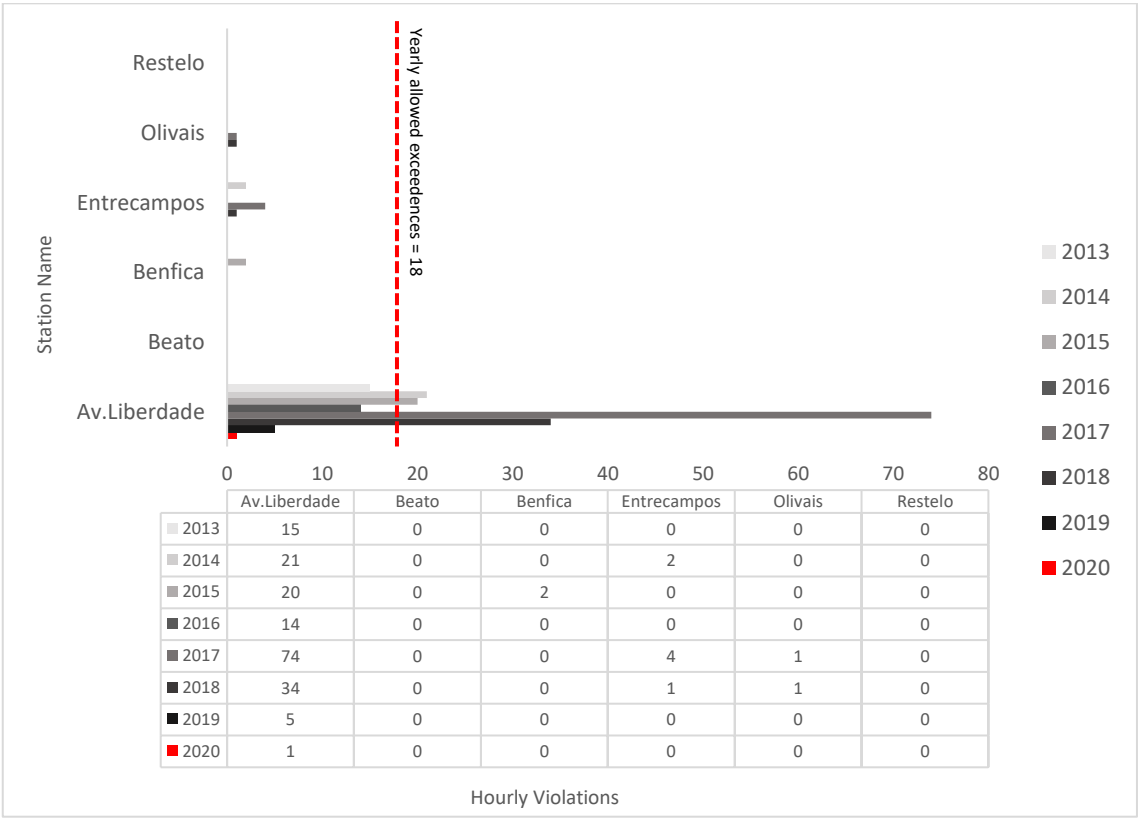


Figure 30 – Annual average NO2 concentration in all Lisbon stations from 2013 to 2020 and the current violation threshold in place (40 µg/m3)



	Av.Liberdade	Beato	Benfica	Entrecampos	Olivais	Restelo
2013	15	0	0	0	0	0
2014	21	0	0	2	0	0
2015	20	0	2	0	0	0
2016	14	0	0	0	0	0
2017	74	0	0	4	1	0
2018	34	0	0	1	1	0
2019	5	0	0	0	0	0
2020	1	0	0	0	0	0

Figure 31 – Count of yearly 1-hour mean violations (200 µg/m3) per year in Lisbon stations and allowed exceedances per year (18)

Regarding the Decree Law No.102/2010 of 2010-09-23 (transposition of EU Directives 2008/50/CE and 2004/107/CE), where air pollutant safety thresholds are regulated, in the pandemic year of 2020, the compliance of the annual mean NO_2 concentration safety threshold of 40 µg/m3 was met by all six monitoring stations in Lisbon including the Av. da Liberdade station which, as far as the available data concerns, had never been able to meet this regulatory threshold (Figure 30).

According to the same law, it is also defined a maximum number of eighteen yearly exceedances of NO_2 hourly-mean concentrations of 200 $\mu g/m^3$. As shown in Figure 31, during the year of 2020 there was a single occurrence in Av. da Liberdade where this threshold was exceeded, and it was well before the pandemic period (05-02-2020) when mobility restrictions associated to the COVID-19 pandemic were in place in Lisbon. As a measure of comparison, the maximum registered NO_2 concentration during COVID-19 lockdown periods in the Av. Liberdade monitoring station, namely during 1st emergency state and 1st calamity situation (14-03-2020 to 01-07-2020), was 99.7 µg/m3 on the 3rd of April 2020 compared to the same period in 2019 where it was measured 213.8 µg/m3 on the 1st of June 2019 or compared to the same period in 2018 where it was measured 237.7 µg/m3 on the 18th of June 2018.

3.6.Tropospheric Ozone (O3) performance during COVID-19 Pandemic

During the initial 2020 lockdown months, the median concentration of tropospheric Ozone measured in traffic stations, where NO_x concentrations are also usually higher due to vehicle emissions, has had a sequential increase of 204.55% from February 2020 to March 2020 whereas on the 2013-2019 baseline, from February to March, it should only increase 28.93% as per the usual yearly seasonality. As shown in Figure 32, from March 2020 onwards, Ozone concentrations remained at higher levels than the 2013-2019 baseline peaking with a 2013-2019 baseline homologous 18.37% increase in March 2020, following a similar but inverse trend of NO_x concentrations. This could indicate that there's change in NO_x:VOC ratio and reduced effect of O_3 titration by NO, resulting in generally higher concentrations of O_3 in the more polluted centre of the city which is usually VOC-limited. Other involved factors such as meteorology and other atmospheric parameters cannot explain this increase since a general increase in O_3 in the traffic stations during lockdown period appears to exist during different types of weather, both beneficial and prejudicial to O_3 formation, albeit in different magnitudes of increase.

Comparative analysis of tropospheric Ozone with the 2013-2019 baseline should be done with care since the 2013-2019 baseline includes two years (2013, 2014) where Ozone concentrations were generally higher with yearly averages of 67 µg/m3 and 60 µg/m3 respectively, against 56-58 µg/m3 yearly averages from 2015 onwards. This means that the actual homologous increase of Ozone concentration in traffic areas should be greater. To further aid this exercise, the 2019 median Ozone concentration has been included where the gap is significantly higher during the lockdown period.

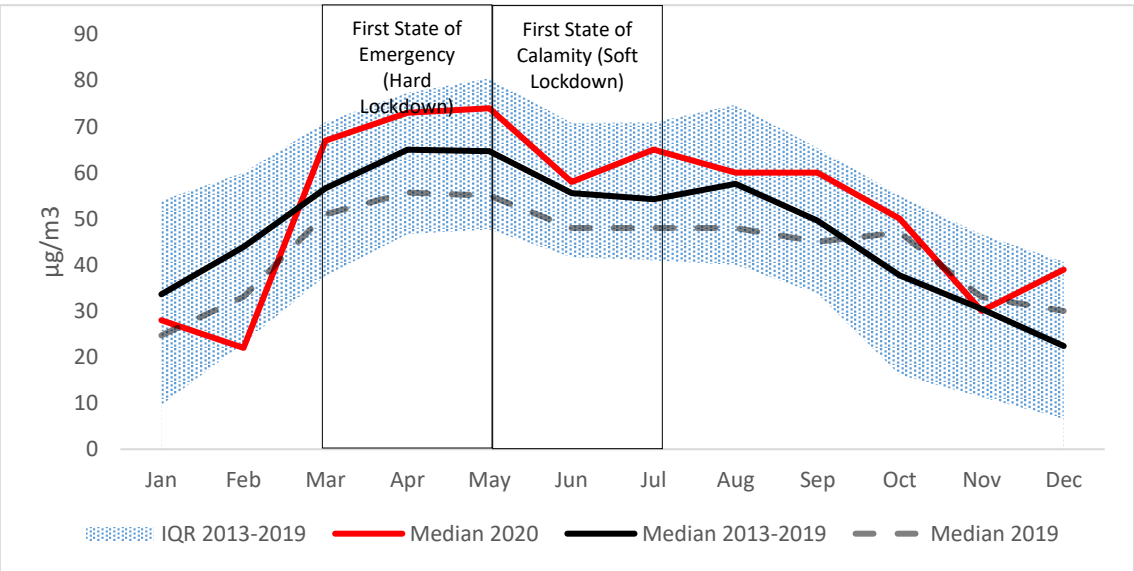


Figure 32 –Road traffic Lisbon monitoring stations monthly median O3 concentration (µg/m3) for the years 2013-2019 and 2020.

As for the background stations, which usually present higher O_3 concentrations due a different NO_x:VOC regime and with less O_3 titration by NO, as in the more polluted city centre, the comparative analysis of 2020 tropospheric Ozone concentration with the 2013-2019 baseline during COVID-19 lockdowns appears to show a slight negative change between -1% to -8% with some months within the statistical error margin (Figure 33). Since background stations are outside the high intensity traffic areas of the city and usually have much lower NO_x concentrations (NO_x limited), the additional reduction in NO_x concentration during lockdowns should justify this light O_3 concentration reduction in the background stations.

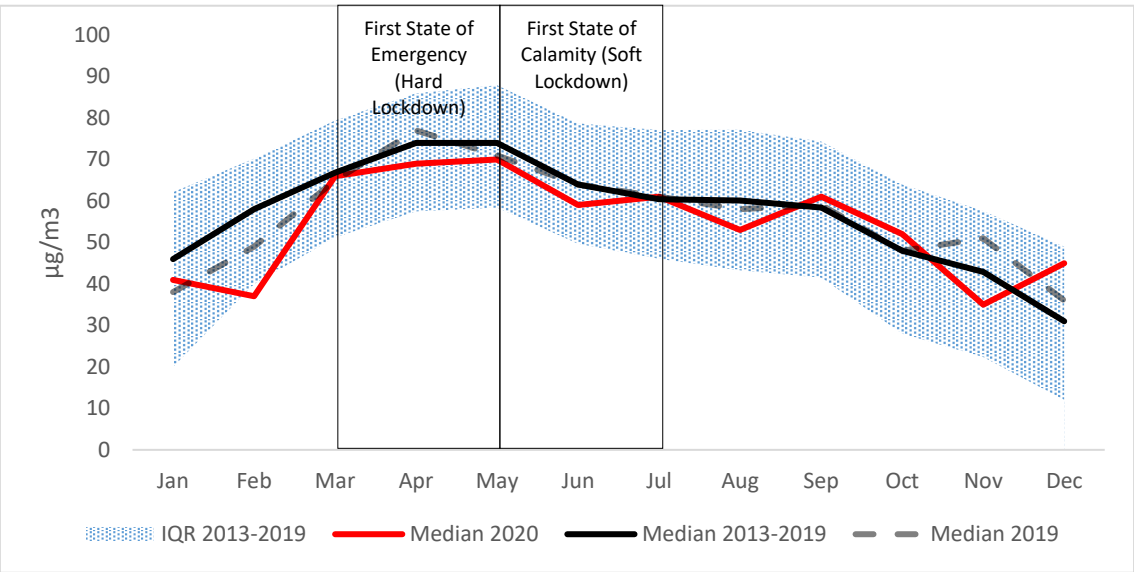


Figure 33 –Background Lisbon monitoring stations monthly median O3 concentration (µg/m3) for the years 2013-2019 and 2020.

The hourly profile of tropospheric Ozone (O_3) in traffic monitoring stations is highly influenced by high NO_x concentrations emitted by internal combustion engine vehicles. Since NO_2 photolysis is the main net contributor of Ozone in these conditions, it is usually expected to have a similar, but inverse, distribution of O_3 concentration when compared to NO_2 concentration during daytime. During night-time, without NO_2 photolysis to fuel the production of additional Ozone, it is slowly destroyed by titration of NO and might also be transported by air currents to other locations.

Specifically, during COVID-19 lockdown periods, thus a NO_x deprived period due to the reduction of emissions, in traffic monitoring stations, night-time periods clearly show slower destruction of Ozone concentrations potentially due to lower availability of NO for O_3 titration which results in less O_3 night scavenging resulting in overall increased O_3 concentrations. During the daylight period, the sharp reduction of NO_x concentrations appears to be correlated to the general increase of O_3 concentrations in urban traffic monitoring stations due to sharp NO_x reduction in a usually VOC-limited regime, shifting the regime to a more NO_x limited regime with higher O_3 production, and reduced effect of O_3 titration by NO associated with usually polluted locations in urban centres. The period with the largest change O_3 concentration change against the 2013-2019 baseline is the 7AM to 10AM period with a homologous 51.43% (07:00h), 46.45% (08:00h), 39.21%

(09:00h) and 20.07% (10:00h) increase of Ozone concentration, followed by the whole daylight period until 19:00h with O_3 concentration increases against the 2013-2019 baseline ranging from 6.78% to 17.55% and finally with the night time period between 20:00h and 01:00h where O_3 concentrations increase the gap against the baseline with positives changes between 16.67% and 24.79% (Figure 34).

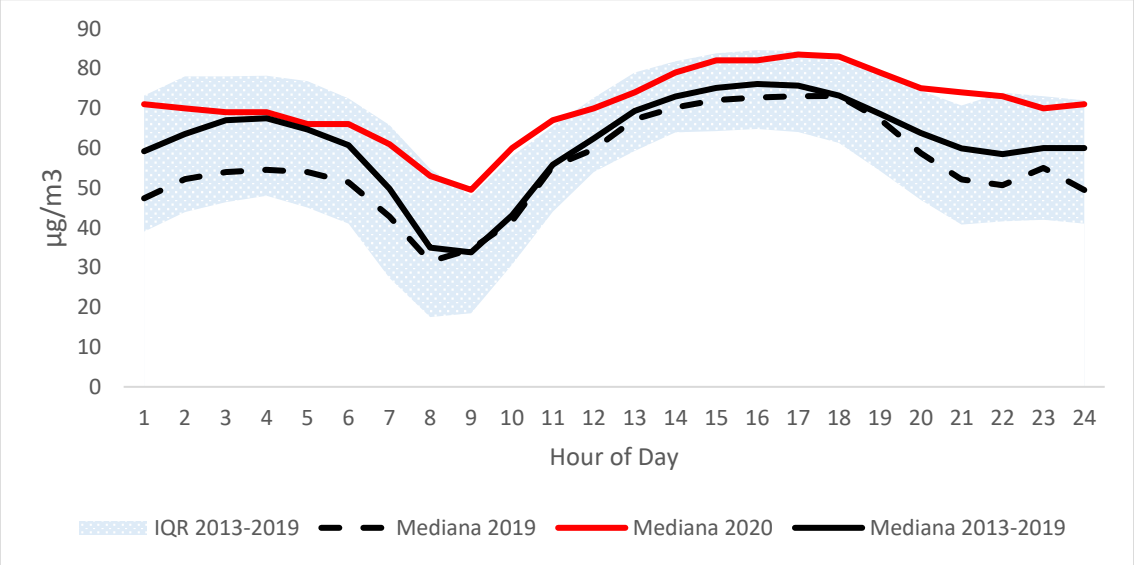


Figure 34 – Hourly O3 median concentration in Traffic Stations for the National Confinement period (14-03-2020 to 03-05-2020) versus 2019 and 2013-2019 Interquartile range, for the same year period.

As for the background stations, the hourly profile doesn't change much when compared against the 2013-2019 baseline, as already had been verified in the overall O_3 concentrations in background stations during the lockdown periods, but it shows a slight decrease. The largest decrease was 11.43% at 03:00h and 9.15% at 11:00h following a similar reduction on NO_x concentration registered in background stations whose already NO_x limited regime should justify the slight O_3 reduction (Figure 35).

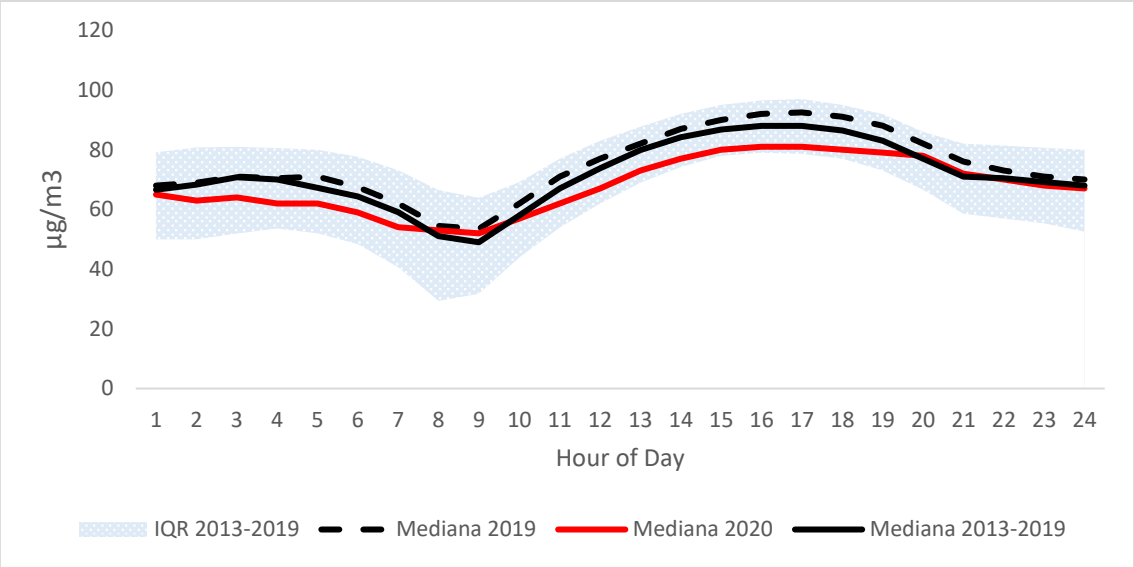


Figure 35 – Hourly O3 median concentration in Background Stations for the National Confinement period (14-03-2020 to 03-05-2020) versus 2019 and 2013-2019 Interquartile range, for the same year period.

Regarding the weekly profile of tropospheric Ozone (O_3) in traffic monitoring stations during lockdown periods, all days of the week show an increase on the O_3 concentration (Figure 36). The days with a higher gap to the baseline are Mondays (+20.40%), Thursdays (+19.83%) and Fridays (+18.15%) which are in line with the weekdays that had the greatest drops on NO_2 concentrations in the weekly profile analysis, further establishing the relationship between O_3 increase and NO_2 decrease.

Another interesting aspect is that due to the routine changes during lockdown periods the main precursor NO_2 has widened the weekend to weekday gap as discussed in Chapter 3.5, which should in theory further increase the O_3 weekend effect resulting in higher O_3 concentrations during the weekend days. This is not clearly noticeable, even with Sunday being the day of the week with the highest O_3 median concentration just slightly above other days of the week, which could be attributed by meteorological factors, the short analysis period resulting in less statistical significance or changes in the O_3 weekend effect during much lower NO_x concentrations.

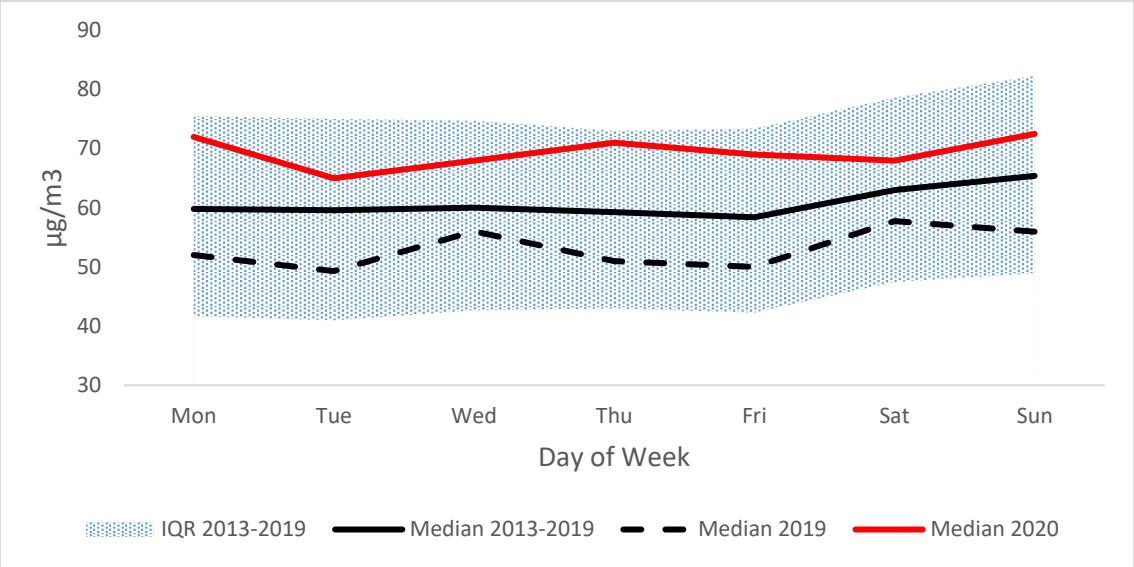


Figure 36 – O3 median concentration in Traffic Stations per day of the week during the National Emergency period and Calamity State (14-03-2020 to 01-07-2020) versus 2013-2019 Interquartile range, for the same year period.

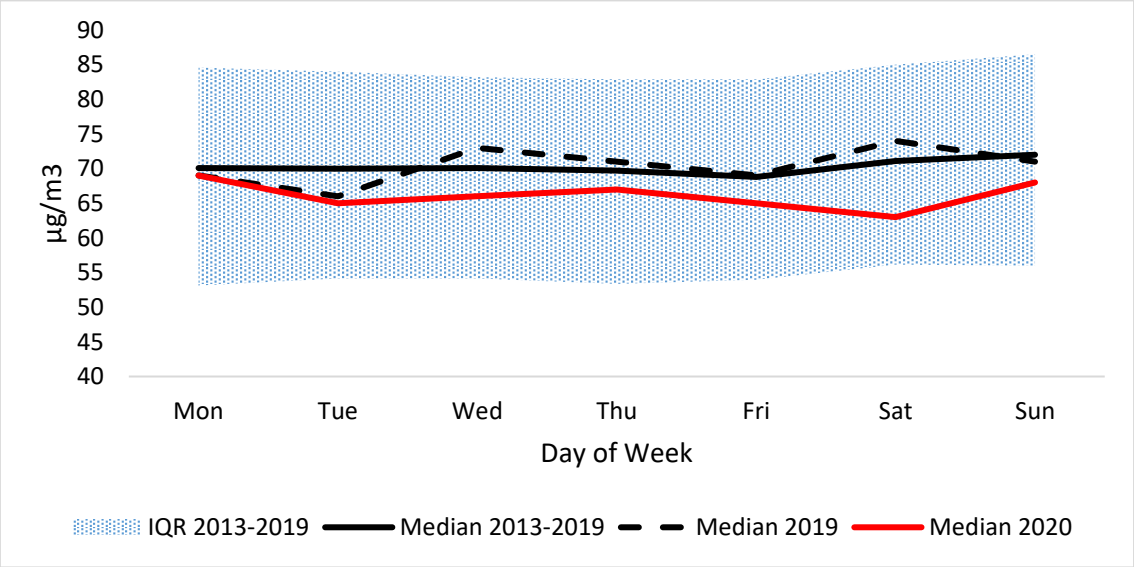


Figure 37 – O3 median concentration in Background Stations per day of the week during the National Emergency period and Calamity State (14-03-2020 to 01-07-2020) versus 2013-2019 Interquartile range, for the same year period.

Regarding the weekly profile of tropospheric Ozone in background stations during lockdown periods, all days of the week show slight decrease on the O_3 concentration (Figure 37) with a somewhat irregular pattern irrespective of the usual O_3 weekend effect. The day with the higher gap against the baseline was Saturday (+11.39%).

As for regulatory compliance of tropospheric Ozone (O_3) concentrations impacts on human health, Lisbon monitoring stations have been compliant to the mandatory regulatory indicator stipulated by the Decree Law No.102/2010 of 2010-09-23 (transposition of EU Directives 2008/50/CE and 2004/107/CE), where air pollutant safety thresholds are regulated, since the beginning of its measurement. The regulatory indicator defines that the daily maximum of 8hr-averages should not exceed 120 µg/m3 O_3 concentration more than 25 days per year, evaluated by a 3-year average. To assess changes introduced by COVID-19 lockdown periods in 2020, the simpler long-term O_3 concentration objective indicator has been used for the analysis. This indicator stipulates that no 8h-average of O_3 concentration shall exceed 120 µg/m3 in all days of the year.

The number of 120 µg/m3 exceedances measured in traffic stations, in this case the Entrecampos station, increased in 2020, following the already identified general O_3 concentration potentially changes in NO_x:VOC regime and to the reduced O_3 titration by NO generated by severed NO_x emissions from vehicle traffic and other anthropogenic activities impacted by the COVID-19 restrictions. The exceedances were registered during three heatwave periods where high air temperature and solar radiation, as well as low average wind speed, during the months of May, July and September 2020 helped sustain higher O_3 concentrations (Figure 38). These months registered percentual increases of the median O_3 concentration measured in the Entrecampos monitoring station against the 2013-2019 baselines of 18.37%, 19.82% and 20.97% respectively.

In May 2020, a heatwave from 26th to 29th registered an average of maximum daily air temperatures well over 30ºC and low average wind speeds measuring between 1.43 m/s and 2.89 m/s. It was the hotter May since 1931, with generally high air temperature averages and stable atmosphere influenced by the Azores anticyclone. There were O_3 exceedances of 120 µg/m3 on the 28th and 29th of May 2020 in both background and traffic monitoring stations.

In July 2020, a heatwave from the 9th to the 18th registered an average of maximum daily air temperatures near 35ºC, including two days (16th and 17th) where it was well over 35ºC, registering as well generally low to moderate average wind speeds. It was the hotter July since 1931, with generally high air temperature averages and stable atmosphere influenced by the Azores anticyclone. There were O_3 exceedances of 120 µg/m3 on the 11th, 13th to 17th and 22nd of July 2020 in both background and traffic monitoring stations.

In September 2020, a heatwave from the 3rd to the 13th registered an average of maximum daily air temperatures between 31ºC and 34ºC, registering as well generally low average wind speeds. It was the 11th hotter September since 1931, when taking in consideration the average maximum daily air temperature, with generally high air temperature averages and stable atmosphere in the first half of the month, influenced by the Azores anticyclone.

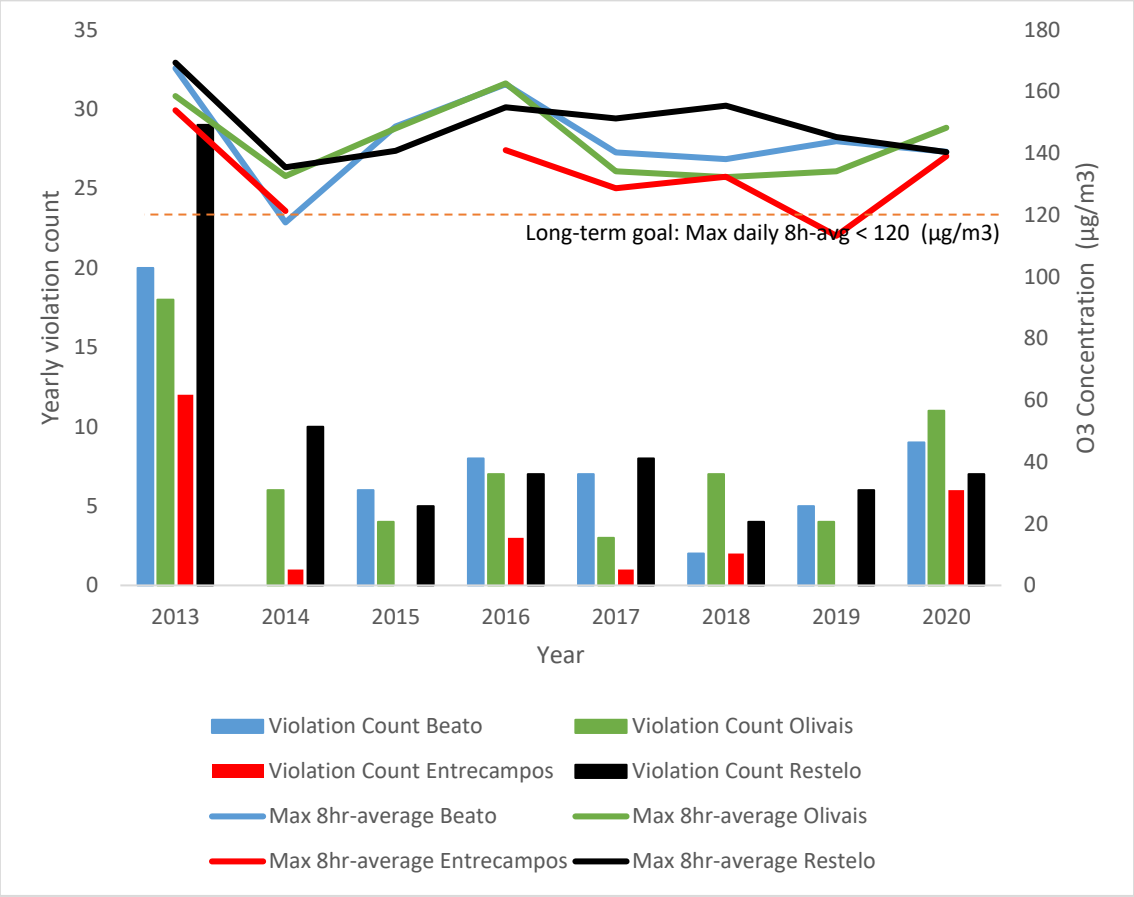


Figure 38 – Count of Long-term O3 concentration violation for human health per year and maximum O3 8hr-average concentration median concentration from 2013 to 2020

As for the public alert of high O_3 concentrations to allow the mitigation of health impacts on the populations, the two defined thresholds are one hour-mean of 180 µg/m3 (Informational) and one hour-mean 240 µg/m3 (Alert). During the year of 2020, there was one day (13/07/2020) in one specific hour (14:00h) when the Informational threshold was crossed in two background monitoring stations, Restelo and Olivais, when the O_3 concentrations reached 190 µg/m3 and 195 µg/m3 respectively (Figure 39). This specific

event occurred during the already described heatwave in July 2020 with optimal conditions for formation of high O_3 concentrations.

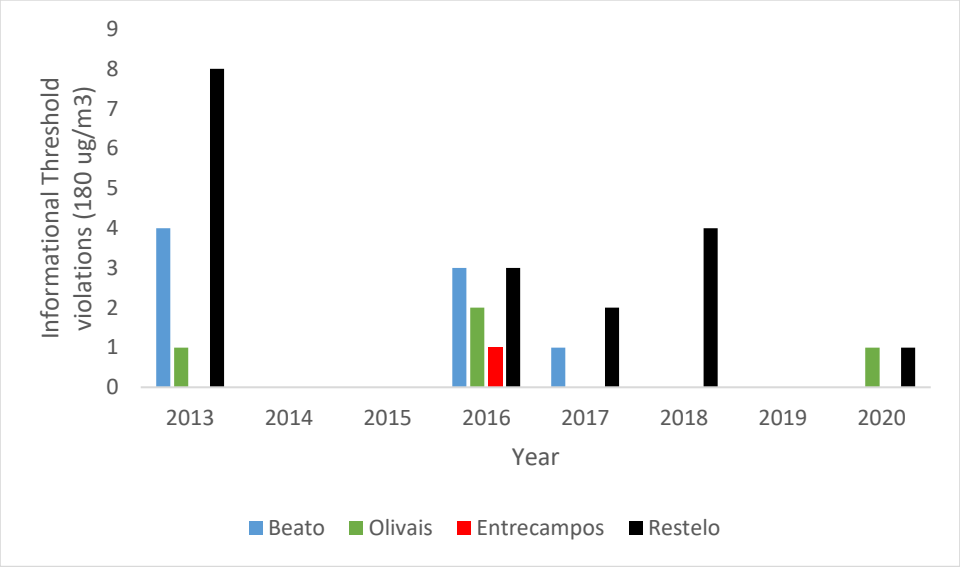


Figure 39 – Informational O3 threshold violations from 2013 to 2020

3.7.Additional criteria pollutants and overall impact to air pollution

While the present work is mainly focused on the response of the NO_2 and O_3 pollutants during the main confinement phases of the COVID-19 pandemic in Lisbon, Portugal, a short analysis on the remaining criteria pollutants, CO, $PM_{2.5}$ and PM_{10}, is provided to complete the high-level criteria pollutant homologous analysis. The SO_2 criteria pollutant is not reported in this work since, for the two monitoring stations in Lisbon reporting SO_2, this air pollutant concentration has been decreasing over the past few years to very low values due to the reduction of the usage of coal and fuel oil for power generation and general desulfurization of liquid fuels. This results in very low and potentially inaccurate readings due to low floating-point precision (1) in the measures recorded by monitoring stations. Over 25% of all 2019-2020 readings were below 0,1 µg/m3.

As a by-product of the incomplete combustion of fossil fuels, road transport is estimated to contribute 78% of CO emissions in Lisbon and Tagus Valley which is slightly above the contribution of road transport to NO_x concentrations (63%) [23]. But unlike NO_2, the median concentration of CO in traffic stations during the first national emergency state (18-03-2020 to 03-05-2020) only registered a decrease of 16.61% (Figure

40) against the 2013-2019 baseline opposed to the drop of 54.35% decrease for NO_2. It is not clear why CO concentrations have not decreased as much as NO_2 during lockdown period since road transport contributes heavily to both air pollutants. A possible cause could be that CO, being a primary pollutant, which is not regenerated by photochemistry processes such as NO_2, has had a lower drop than NO_2 when comparing to the baseline. Other possible factors could be related to the fact that CO is lighter than NO_2, making it easier to disperse in the atmosphere, or due to the fact the vehicle mix during the lockdown periods were different than the baseline (i.e: diesel vs petrol; light vs heavy; newer vs older). Nevertheless, this disproportional drop in NO_2 and CO concentrations should be further investigated.

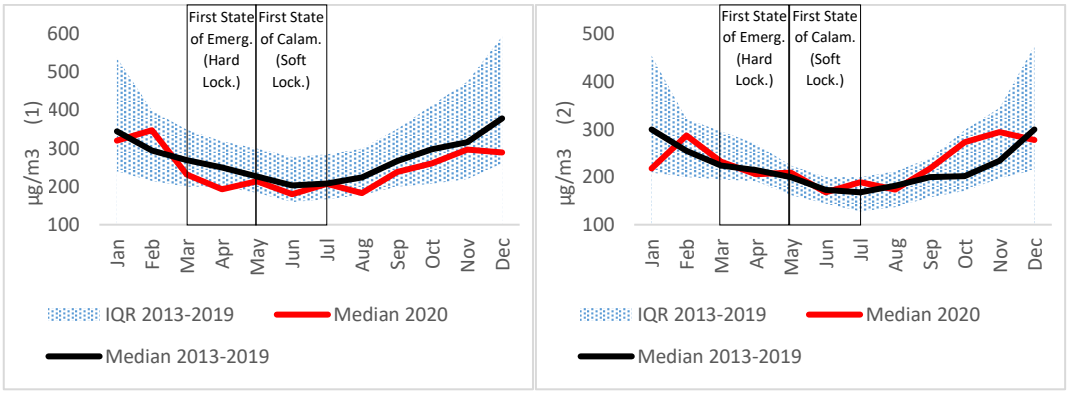


Figure 40 – Road traffic (1) and Background Lisbon (2) monitoring stations monthly median CO concentration (µg/m3) for the years 2013-2019 and 2020.

Particulate Matter with a diameter below 10 microns (PM_{10}), also known as coarse particles, is a major source of urban pollution that causes serious health issues to humans namely respiratory and cardiovascular morbidity, such as aggravation of asthma, respiratory symptoms, and an increase in hospital admissions [58]. In an urban environment the main primary anthropogenic contributions to coarse particles are related to the internal combustion engine in road transport, tyre, and brake wear, as well as abrasion of the asphalt, whereas secondary anthropogenic contributions are mostly related to resuspension of dust and particles by vehicles.

In the case of Lisbon and Tagus Valley region, road transport is estimated to account for 62% of anthropogenic PM_{10} concentrations, with additional relevant contributions by the industry (26%) and production of electricity (9%). Biogenic causes for PM_{10} are also

relevant and, for the Lisbon and Tagus Valley, are mostly related to episodes of dust transportation from North Africa deserts which introduces another variable in the analysis [23]. During the first national confinement period (18-03-2020 to 03-05-2020) the reduction of the median PM_{10} concentration against the 2013-2019 baseline for the same period in traffic stations suffered a retraction of 46.69% while in background monitoring stations the reduction was measured at 14.32% (Figure 41). The results are consistent with the reduction in vehicle traffic during the lockdown periods.

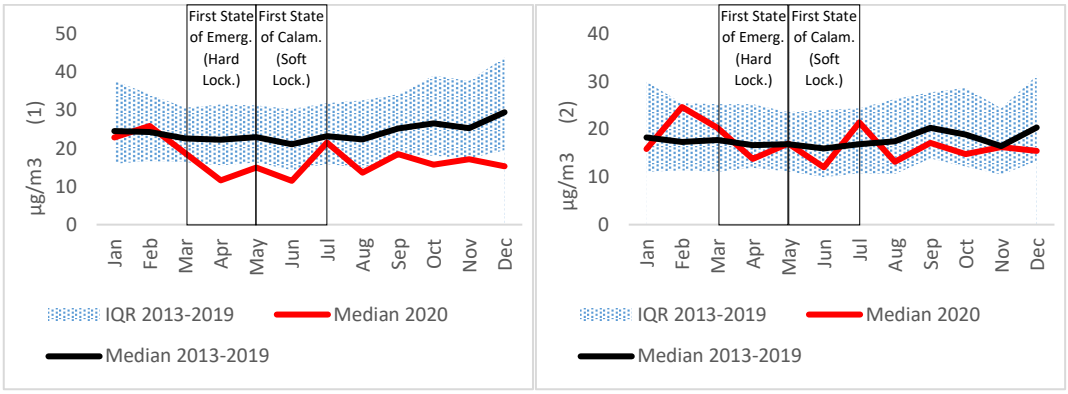


Figure 41 – Road traffic (1) and Background Lisbon (2) monitoring stations monthly median PM10 concentration (µg/m3) for the years 2013-2019 and 2020.

Particulate Matter with a diameter below 2.5 microns ($PM_{2.5}$) is especially dangerous to human health since they can penetrate deeper into the respiratory system and, when exposed to high concentrations for large periods of time, can cause serious respiratory and cardiovascular diseases and lung cancer [58]. The smaller particulate matter in urban environment is often related to secondary pollution formed in the atmosphere involving NO_x, SO_2 and COV, such as nitrates or sulphates, contrary to coarse particulate matter which is mostly related to primary pollution [23]. During the first national emergency (18-03-2020 to 03-05-2020) the reduction of the median $PM_{2.5}$ concentration against the 2013-2019 baseline for the same period in traffic stations dropped 30.28% while in background monitoring stations the reduction was measured at 8.20% (Figure 42).

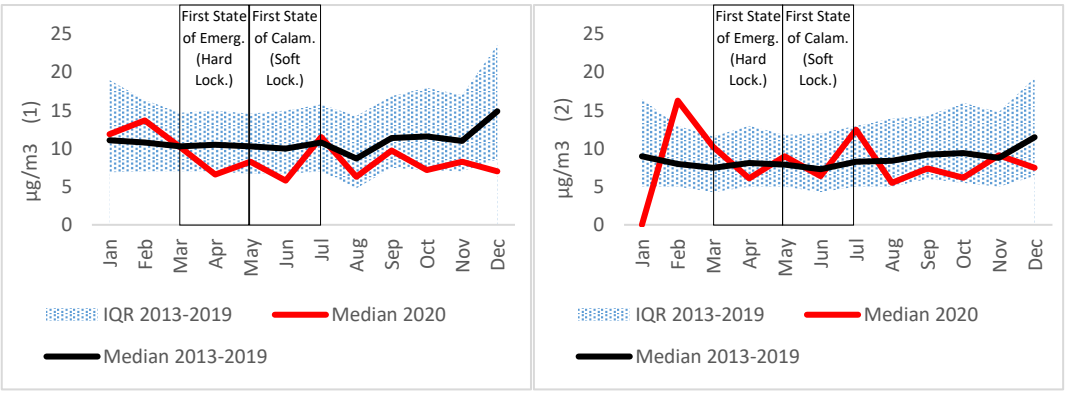


Figure 42 – Road traffic (1) and Background Lisbon (2) monitoring stations monthly median PM2.5 concentration (µg/m3) for the years 2013-2019 and 2020.

The last regulated criteria pollutant is Benzene, which along with toluene, ethylbenzene and xylene make up the VOC "BTEX" aromatic group. Benzene is released into the atmosphere by evaporation of fossil fuels, paints, solvents, and other petroleum products, namely during production, transport and usage, energy production and heating. Additionally, in an urban setting it is estimated that 80% of concentrations are the result of the incomplete combustion process in internal combustion engines of vehicles making it a good indicator for urban anthropogenic VOCs. Benzene is a carcinogenic compound and its concentration is subject to regulation but in recent years the concentrations in Lisbon have been far from the human health threshold (yearly average higher than 5 µg/m3) due to introduced regulatory limitations in Benzene concentrations in liquid fuels. The lifecycle of Benzene in the atmosphere can span up to several days and therefore can be transported by air currents and cause pollution episodes far from the emission location. The main Benzene removal mechanism from the atmosphere is through photochemical oxidation and, as with several other VOCs, a precursor of Ozone. [25]. In Lisbon, due to reduced data quality in Benzene (C_6H_6) concentrations measurements in monitoring stations, the sole monitored VOC could not be extensively reported as the other criteria pollutants.

Nevertheless, comparing the available data from the Entrecampos traffic station during first state of emergency period in 2020 (66% samples available) with homologous periods in 2013 (0% samples available), 2014 (0% samples available), 2015 (100% samples available), 2016 (100% samples available), 2017 (26% samples available), 2018 (0% samples available) and 2019 (73% samples available), a 64.47% reduction in Benzene

concentrations is measured. Regarding the Olivais background monitoring station, the available data for the same period in 2020 (88% samples available) when compared to the single available year of 2019 (100% samples available) shows a 16.75% Benzene concentration reduction.

As for the first calamity period in 2020 (100% samples available), comparing the available data from the Entrecampos traffic station to homologous periods in 2013 (0% samples available), 2014 (0% samples available), 2015 (100% samples available), 2016 (54% samples available), 2017 (0% samples available), 2018 (0% samples available) and 2019 (100% samples available), a 58.55% reduction in Benzene concentration is identified. Regarding the Olivais background monitoring station, the available data for the same period in 2020 (100% samples available), when compared to the single available year of 2019 (100% samples available), shows a 14.57% Benzene concentration reduction.

A summary of the homologous change of the criteria air pollutants concentrations for traffic and background stations in Lisbon during the first national emergency state period from 18-03-2020 to 03-05-2020 (1) and the first national calamity situation period from 04-05-2020 to 01-07-2020 (2) when compared to the 2013-2019 baseline (Figure 43).

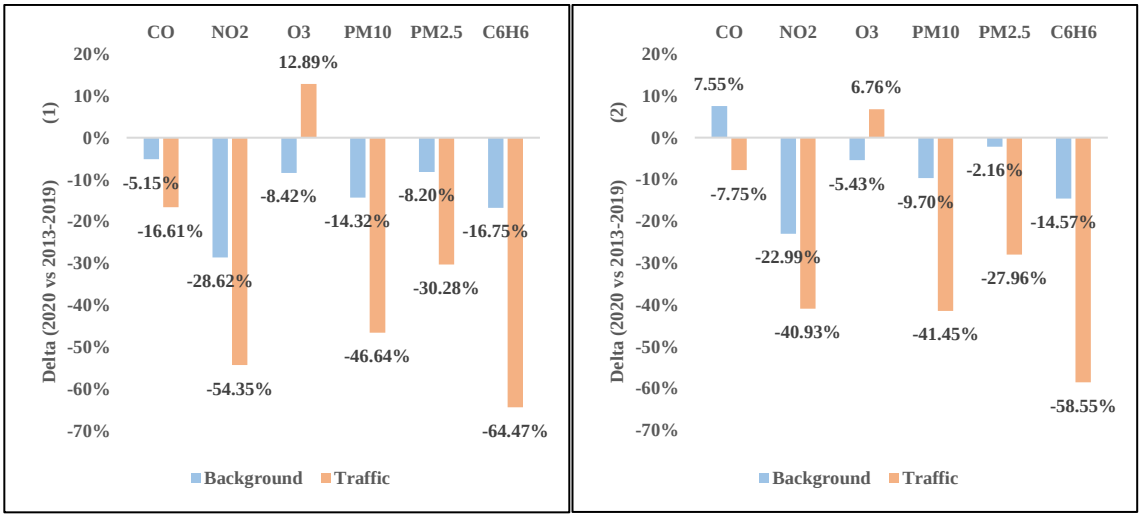


Figure 43 – Criteria air pollutant median concentrations homologous analysis of the 1st national emergency state period from 18-03-2020 to 03-05-2020 (1) and 1st national calamity situation period from 04-05-2020 to 01-07-2020 (2) against a same-period from 2013 to 2019 baseline for background and traffic stations in Lisbon, Portugal.

3.8. Relevant variable bivariate analysis

For the initial stage in the variable analysis a Pearson correlation between the main dependent variables in scope, NO_2 (Nitrogen dioxide) and O_3 (Ozone), and all other independent variables was computed taking into consideration the urban traffic monitoring station air pollutant measures for the period between March/2020 to March/2021. All variables related to air pollution, weather and mobility indexes were included in the correlation matrix.

Variable	NO2 Correlation	O3 Correlation
SUM_RADIAÇÃO_GLOBAL_TOTAL	-0.273188	0.454226
AVG_HUMIDADE_RELATIVA_MEDIA	-0.090687	-0.38972
AVG_TEMPERATURA_MEDIA	-0.248014	0.383737
MIN_TEMPERATURA	-0.35166	0.383405
MAX_TEMPERATURA	-0.160445	0.351638
AVG_PRESS_ATMOSFERICA_ESTACAO_MEDIA	0.3682	-0.410395
AVG_PRESS_ATMOSFERICA_MAR_MEDIA	0.374044	-0.419959
AVG_INTENSIDADE_VENTO_MEDIA	-0.527739	0.387259
MIN_INTENSIDADE_VENTO_MEDIA	-0.43581	0.267402
MAX_INTENSIDADE_VENTO_MEDIA	-0.513174	0.41352
SUM_PRECIPITACAO_ACUMULADA	-0.112309	0.05736
AVG_CHG_WAZE_KM	0.455335	-0.30106
MED_TR_NO2	1	-0.541007
MED_TR_O3	-0.541007	1
MED_TR_CO	0.771071	-0.641279
MED_TR_PM10	0.544638	-0.301011
MED_TR_PM25	0.564506	-0.357821
G_RETAIL_RECREATION	0.497434	-0.304965
G_GROCERY_PHARMACY	0.536579	-0.310879
G_PARKS	0.469234	-0.167697
G_TRANSIT_STATIONS	0.468611	-0.287488
G_WORKPLACE	0.206122	-0.205377
G_RESIDENTIAL	-0.234332	0.145957

Figure 44 – NO2 and O3 Pearson Correlation Coefficients

As a result, several of the relationships already identified during the literature review are now identified and quantified as to the strength of their correlation (Figure 44). Some of

the most obvious ones are the positive moderate correlations between NO_2 and several urban mobility indexes such as Waze driven kilometers (r=0.46) or Google Grocery and Pharmacy (r=0.54). As for weather related variables, NO_2 has a moderated negative correlation with the Average wind speed (r=-0.53) and as for the relationship with other air pollutants, NO_2 is moderately negatively correlated to O_3 which is a byproduct of NO_2 itself (r=-0.54) and is strongly correlated with CO (r=0.77) as they are usually co-linear in urban environments since they are both product of the combustion process in vehicle engines.

In urban centers, Nitrogen Oxides (NO_x) emissions are often associated to internal combustion powered vehicles and in Lisbon and Tagus Valley internal combustion vehicles associated with road traffic are estimated to be responsible for 63% of all emissions [23]. With the COVID-19 pandemic related lockdowns, road traffic is estimated to have been greatly reduced with the worst month being April 2020 measuring an average reduction of 84.13% of driven Kms registered by Waze App while the median NO_2 concentration in traffic monitoring stations registered a reduction of 56.30% for the same month when compared to the 2013-2019 baseline.

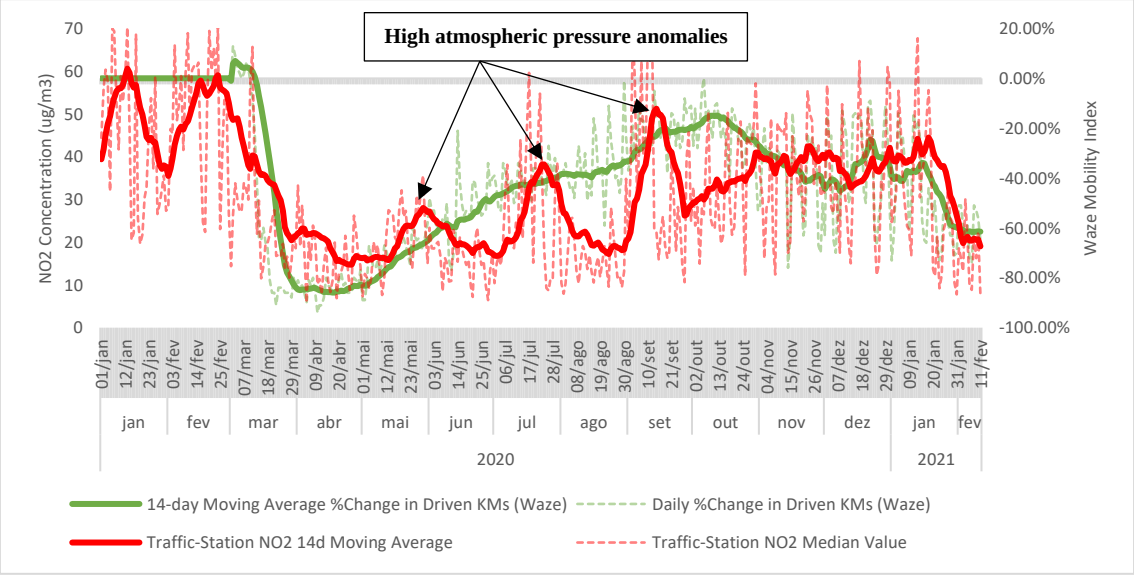


Figure 45 – Relationship between Waze Mobility Index and NO2 Concentration (14d Moving Averages)

A positive Pearson correlation of +0.46 is measured in the relation between daily NO_2 median concentration and the Waze daily mobility index (Figure 46), which suggests that

with higher levels of vehicle mobility there's also higher NO_2 concentration levels. During periods of optimal meteorological conditions that improve the atmospheric stability, general sharp NO_2 concentration upticks can be identified (Figure 45). These events occurred in 2020 at the end of May, mid-July, and early September, and have already been described in section 3.6.

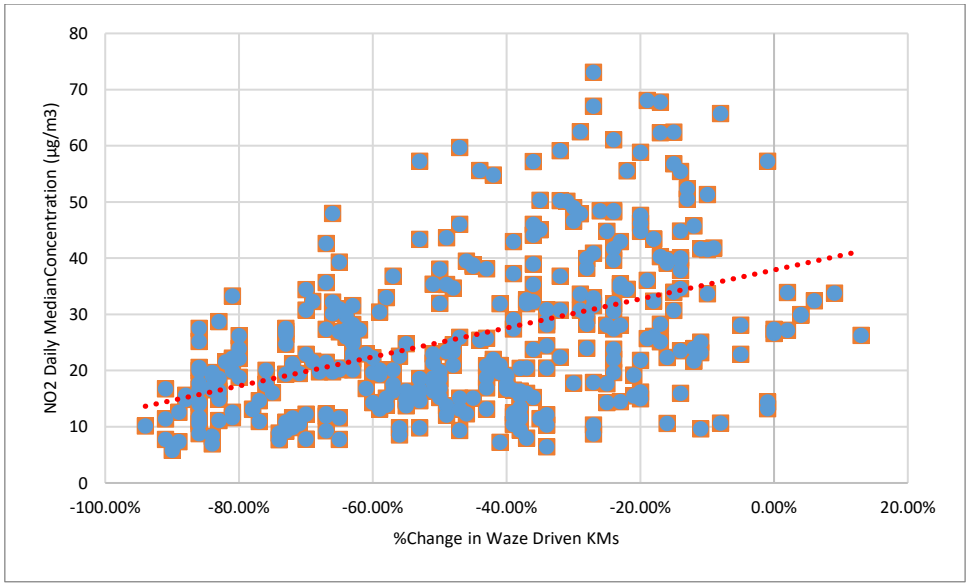


Figure 46 – Relationship between Daily %Change in driven KMs (Waze) and Daily Mean NO2 Concentration

It is well established that stable atmospheric conditions during high pressure events with low intensity wind are optimal conditions for primary air pollutants such as NO_x, CO, and VOC to locally linger near its emission sites and react to form secondary pollutants such as O_3, when in presence of increased solar radiation and temperature, or Particulate Matter. Additionally, during these atmospheric stable periods, because there's also less chance of rain there will be less air pollutant wet deposition and since there's also less chance for cloud formation more solar radiation will reach the low troposphere thus fueling photochemical reactions associated with some secondary air pollutants such as O_3. Additionally, high pressure events, can facilitate the occurrence of thermal inversions in which colder air is trapped beneath a warmer air mass thus reducing air convection and pollutant dispersion. On the other hand, turbulent atmospheric conditions and high intensity winds will disperse and transport air pollutants via air currents and vertically via convection. During low pressure events, rain is also much more probable which will result

in additional air pollutant wet deposition and since these events are also commonly associated with cloud formation, less solar radiation reaches the low troposphere thus slowing down photochemical reactions that lead to pollution.

In Lisbon, a moderate negative Pearson correlation of -0.53 is measured in the relationship between daily NO_2 median concentration and daily mean wind intensity for the year 2020 (Figure 48), which suggests that higher wind intensity tend to reduce NO_2 concentration levels by means of atmospheric turbulence, air current transport and vertical convection.

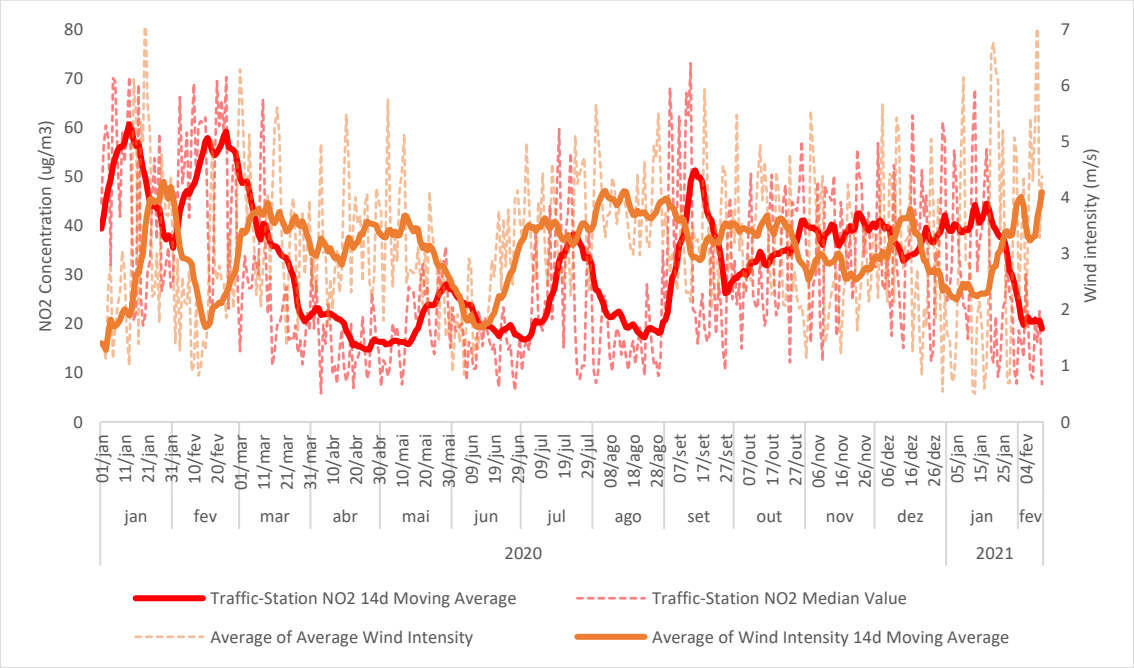


Figure 47 – Relationship between Wind Intensity and NO2 Concentration (14d Moving Averages)

During the first national emergency phase covering the entirety of April/2020, where most of the severe lockdowns happened, the main meteorological parameters for Lisbon were not beneficial to NO_2 build-ups since it was a rainy month with only a slight mean air temperature positive anomaly where the synoptic analysis show that most of the month was depressionary. These wet and windy conditions favour lower NO_2 concentrations which in conjunction with highly reduced anthropogenic activities might have contributed to unprecedent low NO_2 concentrations in Lisbon throughout this period (Figure 47).

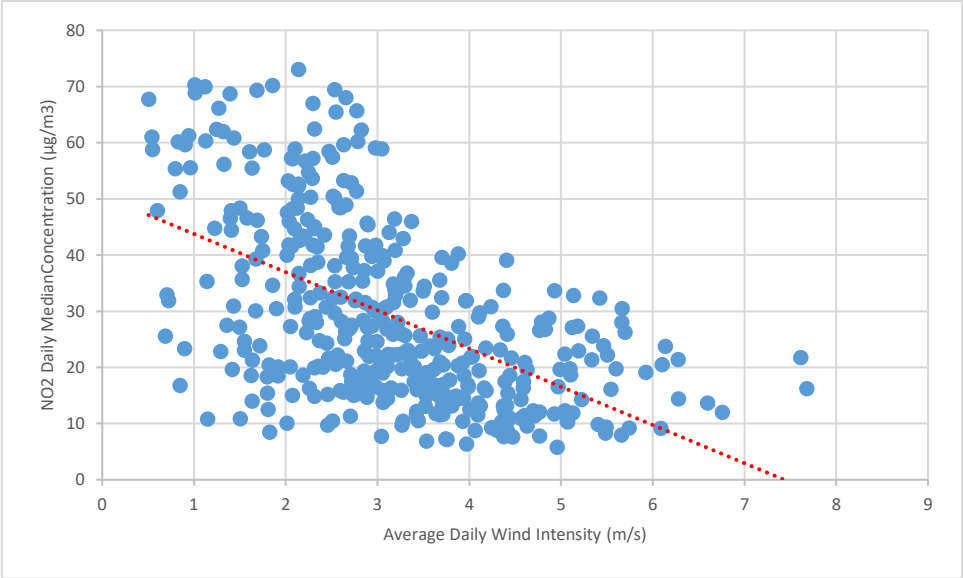


Figure 48 – Relationship between Daily Mean Wind Intensity and Daily Mean NO2 Concentration

During the main lockdown periods in analysis, from 18-03-2020 to 01-07-2020, the wind direction in Lisbon blew predominantly from NW to N direction with occasional strong winds (> 7 m/2) blowing from SW to W, as well as from NE (Figure 49). The relationship between wind speed and lower NO_2 concentrations can also be analyzed taking into consideration hourly wind direction patterns in Lisbon geographic region. For the period under analysis, winds in the morning rush hour, when usually higher NO_2 concentrations build up in the city center due increased vehicle traffic, were usually concentrated in quadrant 330º – 70º with predominantly low wind speeds with a mean interval of 2.1 m/s - 2.6m/s from 07:00h to 10:00h and a mean NO_2 concentration of 26.7 µg/m3 to 30.9 µg/m3. Towards the end of the day, even as the NO_2 buildup potential was lower due to the reduced pendular trips during the lockdown periods, the winds were predominantly concentrated in quadrant 220º - 340º with generally higher wind speeds with a mean interval of 4.1m/s – 4.5m/s from 17:00hM to 20:00h and a mean NO_2 concentration of 19.9 µg/m3 to 22.0 µg/m3. As for O_3, since it is mostly dependent on precursor availability, solar radiation, temperature, and stable enough atmosphere for the photochemical reactions that lead to O_3 production to occur throughout the day, there's a tendency for O_3 concentrations to peak in the afternoon which is also when, specifically in Lisbon, wind speeds are usually higher. This is a possible reason for the O_3 correlation coefficient to wind speed to be positive and not an indirect or direct causation effect.

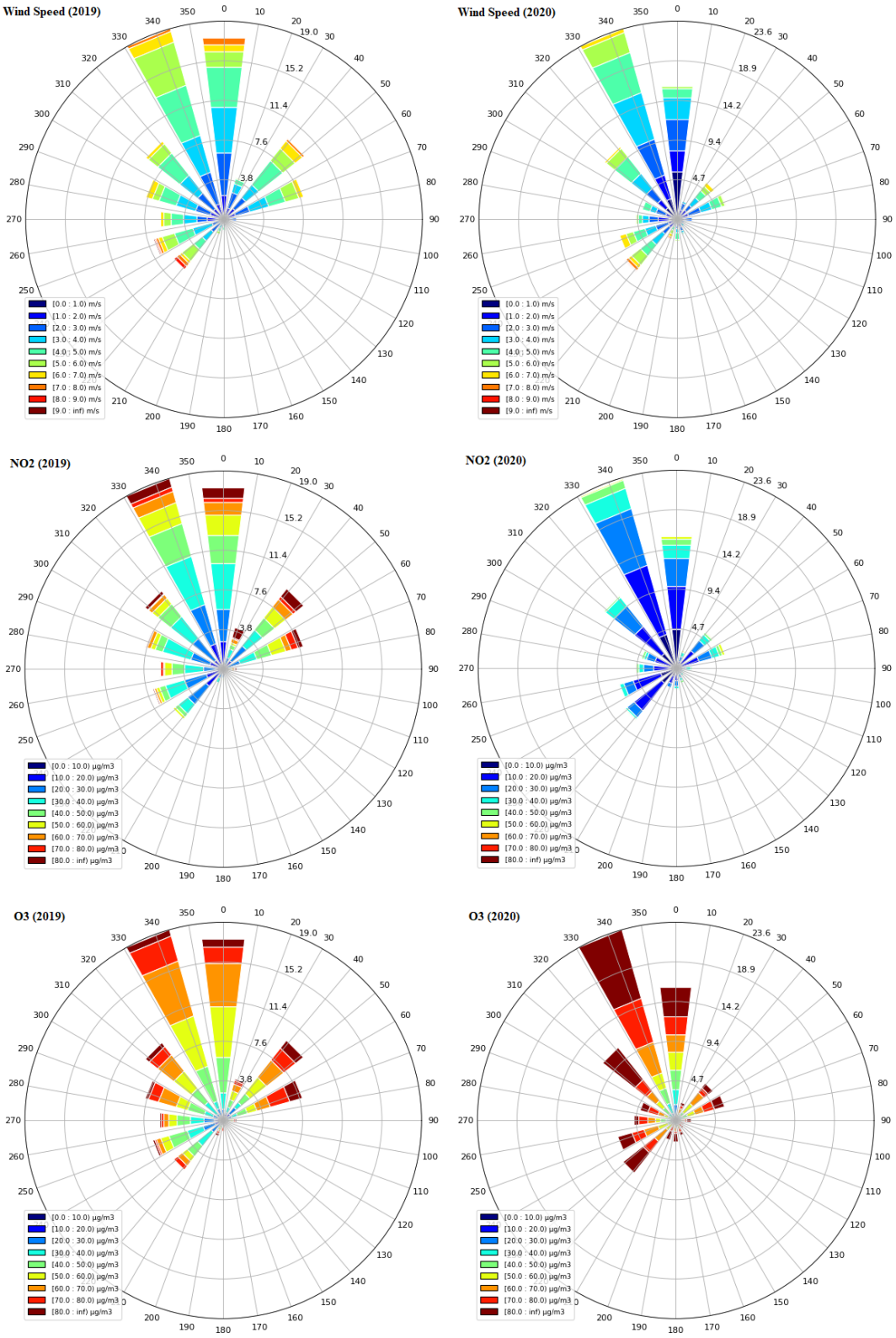


Figure 49 – Wind Direction and Speed relationship with NO2 and O3 concentrations homologous analysis (2019 vs 2020) in traffic stations for the 1st national emergency state and 1st national calamity state periods (18-03-2020 to 01-07-2020)

During the 2020 COVID-19 lockdown related periods, the sharp reduction of NO_x concentrations due to the reduction of emissions by road traffic appears to be correlated to the general increase of O_3 concentrations in urban traffic monitoring stations due to sharp NO_x reduction in a VOC-limited environment as discussed in section 3.6. During periods of optimal meteorological conditions that improve the photochemical reactions efficiency, general sharp upticks can be identified. These were described as well in section 3.6 and occurred in 2020 at the end of May, mid-July, and early September, and can be inspected on the graphic by the respective Average Air Temperature peaks (Figure 50).

From a source appointment point of view, the higher O_3 concentrations can be related to improved meteorological conditions, such as higher solar radiation, higher air temperatures and lower wind speed, and/or changes in precursors concentration and other compounds that are part of the complex photochemical reactions that create Ozone as described in 2.2.

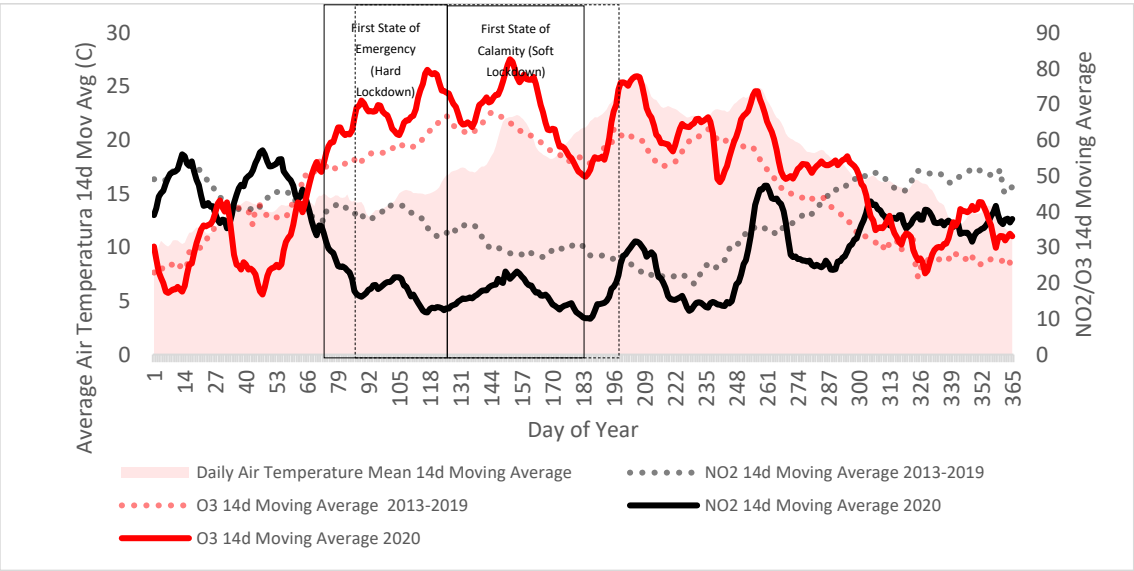


Figure 50 – Road traffic Entrecampos monitoring station NO2 and O3 concentrations (µg/m3) 14-day Moving Average for 2020 and 2013-2019. Shaded with 2020 Daily Air Temperature Mean 14d Moving Average

For instance, in April/2020 there doesn't seem to have occurred exceptional conditions that could improve Ozone formation. This specific month was rainy and only slightly

hotter than the historical baseline with a minimal anomaly on the mean air temperature (+0.76C) but the month was characterized by a 12.31% increase in the median O_3 concentration in the traffic monitoring station of Entrecampos. For the same period, background stations registered a reduction of -6.76% in the median O_3 concentration, the exact opposite trend. This behaviour appears to favour the hypotheses that the sharp NO_x reduction at the Entrecampos traffic monitoring station in 2020 (-56.30%) had an effect in the O_3 concentration increases measured at the same site, namely a change in the usual VOC-limited regime characteristic of traffic intensive spots in urban centres and the reduction of O_3 titration by NO (Figure 50).

On the other hand, that doesn't happen in background stations within the city that usually have a lower NO_x concentration profile and operates in a more NO_x limited regime. The smaller decrease in NO_x concentration during COVID-19 restrictions in background stations, when comparing April 2020 to the 2013-2019 baseline, was -32.24% which caused a slight decrease in O_3 concentrations as it deepened even further into the NO_x limited region (Figure 51).

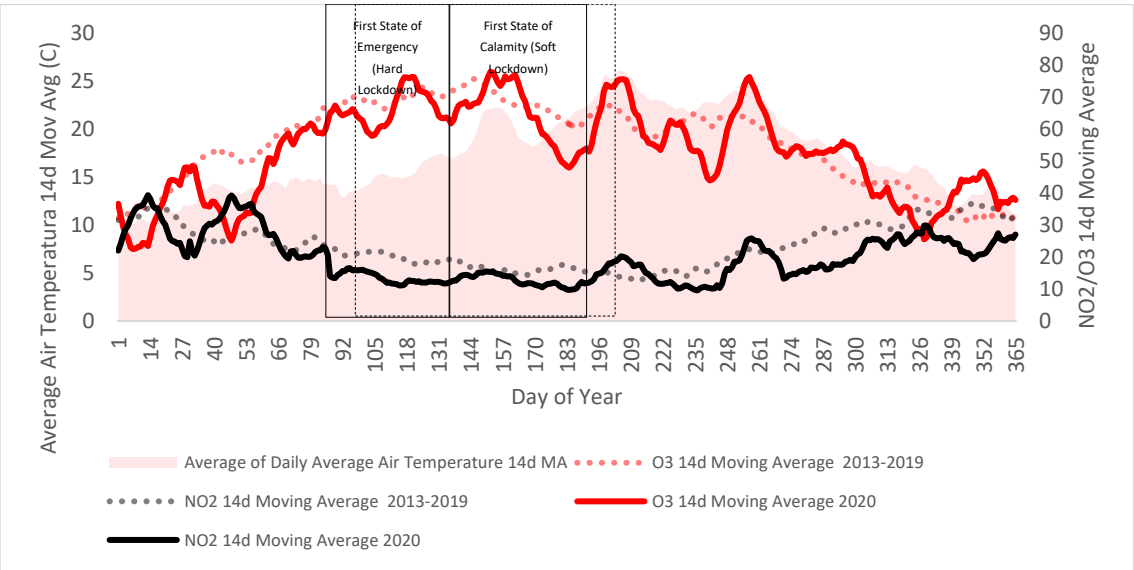


Figure 51 – Background monitoring stations NO2 and O3 concentrations (µg/m3) 14-day Moving Average for 2020 and 2013-2019

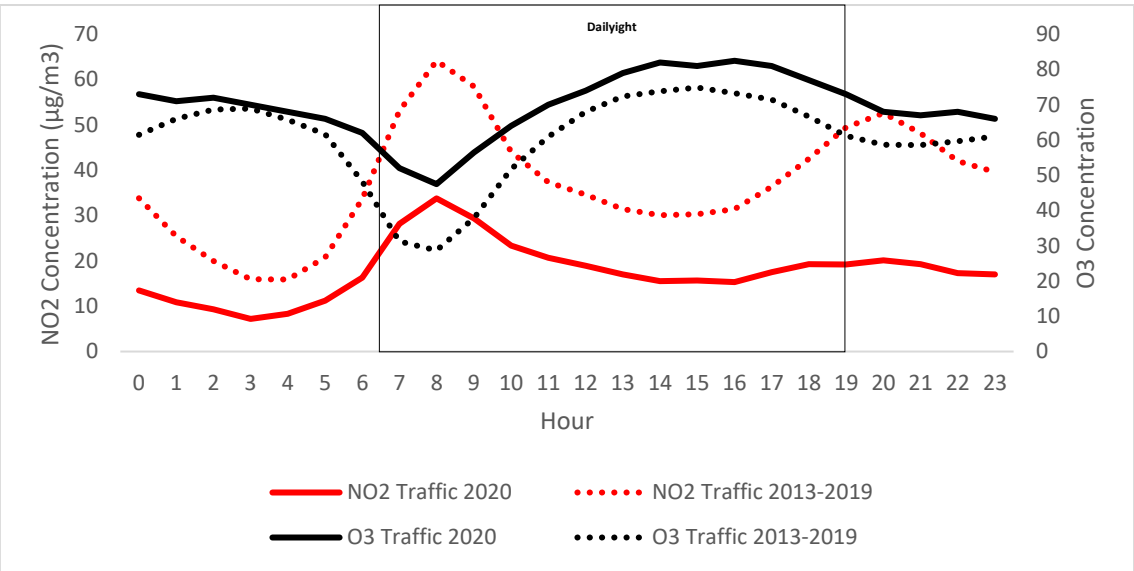


Figure 52 – Hourly O3 and NO2 median concentration in Traffic Stations for the National Confinement period (14-03-2020 to 03-05-2020) versus 2013-2019

The diurnal pattern of NO_2 and O_3 concentrations during the most severe lockdown period (1st emergency state) show clear differences to the 2013-2019 baseline. For the traffic monitoring stations, the morning rush NO_2 peak has a -53.67% drop at 09:00h against the 2013-2019 baseline and the end of day NO_2 peak there's a -61.80% drop at 20:00h. Conversely, O_3 concentrations increased by 66.40% at 07:00h against the 2013-2019 baseline and it remains 10% to 20% higher during the rest of the day (Figure 52). With such a drastic negative change in NO_x concentrations, the NO_x:VOC regime also changed resulting in generally higher O_3 concentrations even with reduced NO_x concentrations, as already discussed in section 3.6. O_3 concentrations are also further impacted by the reduced effect of O_3 titration by NO, which is characteristic of polluted environments such as the centre of large urban areas with large number of vehicles, and generally helps lowering O_3 concentrations. Regarding the background stations, which usually register lower NO_2 concentrations and thus operate in more NO_x limited regime, a moderate reduction of NO_x concentrations during the lockdown periods have caused little change to O_3 concentration. The usual NO_2 peak hour in background stations is at 08:00h and it suffered a -32.85% concentration reduction against the 2013-2019 baseline. Ozone concentrations registered a slight increase for the same hour of 6.12% while during the rest of the day O_3 concentrations remain slightly lower than the 2013-2019 baseline as NO_x concentrations remain at exceptional low concentrations levels.

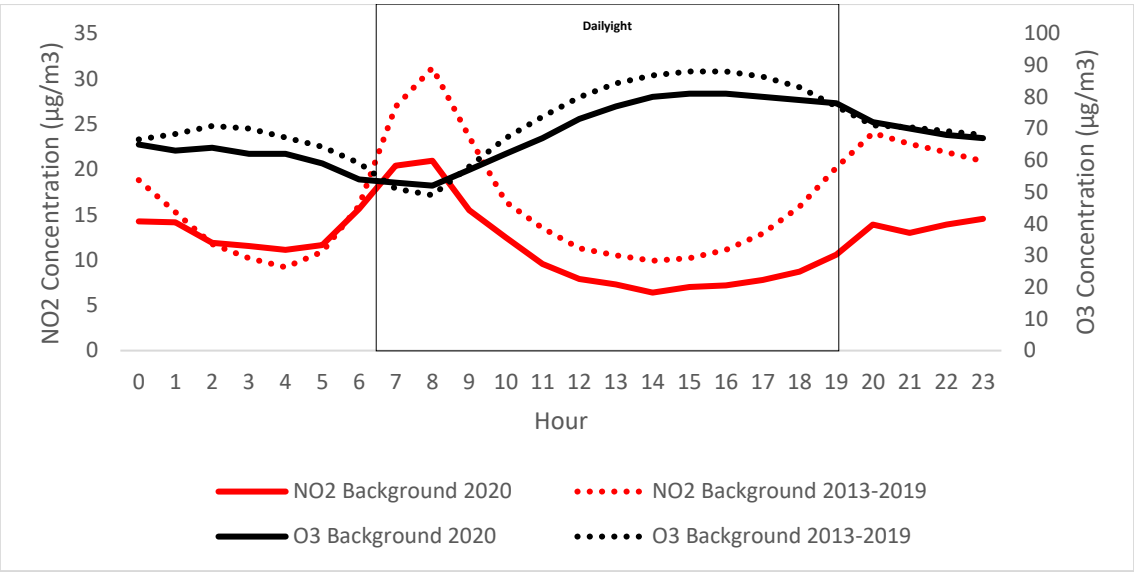


Figure 53 – Hourly O3 and NO2 median concentration in Background Stations for the National Confinement period (14-03-2020 to 03-05-2020) versus 2013-2019

3.9. Modeling NO2 Concentration with AutoML

Apart from an expert analysis of weather and pollutant emission data using traditional statistical analysis to assess the relationship between independent variables and the dependent variable, and, in the case of air pollution modeling, the usage of advanced and complex atmospheric chemical transport simulation models, it is also possible to use Machine Learning techniques to model physical and chemical phenomena involved in air pollution. In this work the AutoML framework TPOT [59] is used to discover a performant and interpretable machine learning model capable of explaining the relationship between independent and dependent variables as well as their strength.

The TPOT AutoML framework uses a genetic optimization algorithm to search for an optimal solution for the stacking of pipeline stages (i.e: add a feature normalization stage, feature encoding stage, dimensionality reduction stage, regressor stage, etc.) and for each stage the optimal parameters of the processor (i.e: selector or regressor hyperparameters) as depicted in Figure 54. This approach accelerates the first iterations of modeling and evaluation from the CRISP-DM framework, which is traditionally carried out by Data Scientists in a mostly manual exploratory fashion and allows the identification of what algorithms and approaches yield promising results before investing in more advanced feature engineering, and overall model building and tuning in subsequent iterations. The

present work intends to train a sufficiently accurate model capable of identifying the most relevant independent data features used to model the NO_2 urban concentration phenomena and to identify any relation between independent features that might be of interest.

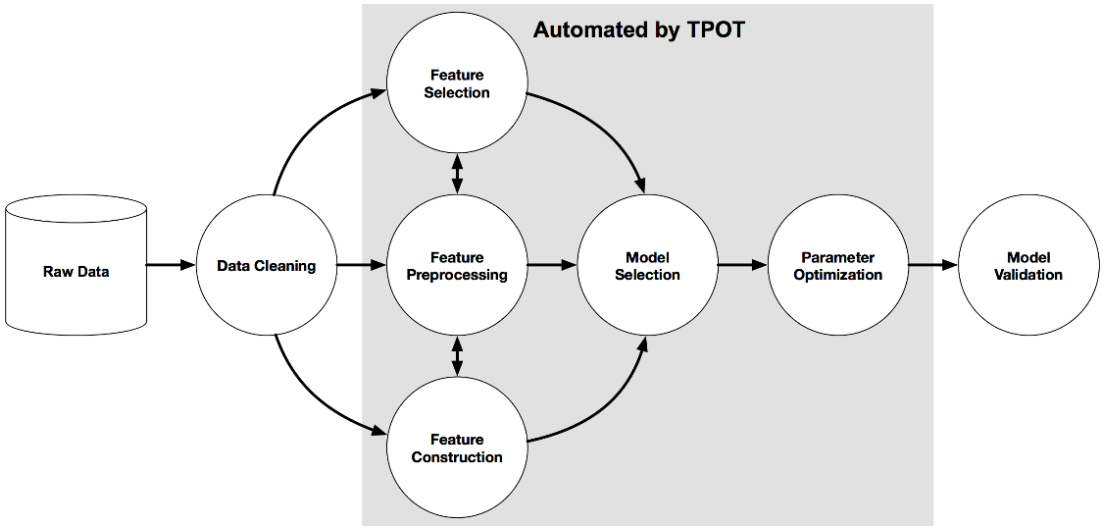


Figure 54 – ML Pipeline steps automated by TPOT AutoML Framework [6]

The list of initial raw data features used to model NO_2 concentration (Table 8) were selected from the datasets described in section 3.2 taking in consideration known variables that potentiate the formation and destruction of the NO_2 air pollutant also described in the literature, as well in the findings described in Chapter 3.8.

Since one of the main group of data features pertain to urban mobility indicators from Google, Waze, and Apple, which are proxies for anthropogenic activities, and these only started being collected in March 2020, the dataset used for modeling NO_2 daily median concentration contains 381 daily samples.

[6] Image obtained from http://epistasislab.github.io/tpot/

Table 8 – List of data features

Data Feature	Short Description
SUM_RADIAÇÃO_GLOBAL_TOTAL	Daily sum of total solar radiation (kj/m2)
SUM_RADIAÇÃO_GLOBAL_TOTAL_LAG_1D	The same variable with a 24h lag
AVG_DIRECAO_VENTO_MEDIA	Daily average wind direction (north degrees)
AVG_DIRECAO_VENTO_MEDIA_LAG_1D	The same variable with a 24h lag
AVG_HUMIDADE_RELATIVA_MEDIA	Daily average relative humidity (percentage)
AVG_HUMIDADE_RELATIVA_MEDIA_LAG_1D	The same variable with a 24h lag
AVG_TEMPERATURA_MEDIA	Daily average air temperature (ºC)
AVG_TEMPERATURA_MEDIA_LAG_1D	The same variable with a 24h lag
MIN_TEMPERATURA	Daily minimum air temperature hourly mean (ºC)
MIN_TEMPERATURA_LAG_1D	The same variable with a 24h lag
MAX_TEMPERATURA	Daily maximum air temperature hourly mean (ºC)
MAX_TEMPERATURA_LAG_1D	The same variable with a 24h lag
AVG_PRESS_ATMOSFERICA_MEDIA	Daily average atmospheric pressure at station level (hPa)
AVG_PRESS_ATMOSFERICA_MEDIA_LAG_1D	The same variable with a 24h lag
AVG_PRESS_ATMOSFERICA_MAR_MEDIA	Daily average atmospheric pressure at sea level (hPa)
AVG_PRESS_ATMOSFERICA_MAR_MEDIA_LAG_1D	The same variable with a 24h lag
AVG_INTENSIDADE_VENTO_MEDIA	Daily average wind intensity (m/s)
AVG_INTENSIDADE_VENTO_MEDIA_LAG_1D	The same variable with a 24h lag
MIN_INTENSIDADE_VENTO_MEDIA	Daily minimum wind intensity hourly mean (m/s)
MIN_INTENSIDADE_VENTO_MEDIA_LAG_1D	The same variable with a 24h lag
MAX_INTENSIDADE_VENTO_MEDIA	Daily maximum wind intensity hourly mean (m/s)
MAX_INTENSIDADE_VENTO_MEDIA_LAG_1D	The same variable with a 24h lag
SUM_PRECIPITACAO_ACUMULADA	Daily sum of accumulated precipitation (mm)
SUM_PRECIPITACAO_ACUMULADA_LAG_1D	The same variable with a 24h lag
AVG_CHG_WAZE_KM	Daily change of driven kilometers measured by Waze App against a pre-pandemic baseline (percentage)
AVG_CHG_WAZE_KM_LAG_1D	The same variable with a 24h lag
AVG_G_RETAIL_RECREATION	Daily change of user activity in retail and recreation measured by Google against a pre-pandemic baseline (percentage)
AVG_G_RETAIL_RECREATION_LAG_1D	The same variable with a 24h lag
AVG_G_GROCERY_PHARMACY	Daily change of user activity in local commerce measured by Google against a pre-pandemic baseline (percentage)
AVG_G_GROCERY_PHARMACY_LAG_1D	The same variable with a 24h lag
AVG_G_PARKS	Daily change of user activity in parks measured by Google against a pre-pandemic baseline (percentage)
AVG_G_PARKS_LAG_1D	The same variable with a 24h lag
AVG_G_TRANSIT_STATIONS	Daily change of user activity in transit stations measured by Google against a pre-pandemic baseline (percentage)
AVG_G_TRANSIT_STATIONS_LAG_1D	The same variable with a 24h lag
AVG_G_WORKPLACE	Daily change of user activity in workplace areas measured by Google against a pre-pandemic baseline (percentage)
AVG_G_WORKPLACE_LAG_1D	The same variable with a 24h lag
AVG_G_RESIDENTIAL	Daily change of user activity in residential areas measured by Google against a pre-pandemic baseline (percentage)
AVG_G_RESIDENTIAL_LAG_1D	The same variable with a 24h lag
AVG_A_DRIVING	Daily change of Apple maps driving requests measured by Google against a pre-pandemic baseline (percentage)
AVG_A_DRIVING_LAG_1D	The same variable with a 24h lag
AVG_A_WALKING	Daily change of Apple maps walking requests measured by Google against a pre-pandemic baseline (percentage)
AVG_A_WALKING_LAG_1D	The same variable with a 24h lag
MED_TR_NO2	Daily NO2 median concentration (µg/m3)
MED_TR_NO2_LAG_1D	24h lag variable of NO2 concentration (µg/m3)
MED_TR_O3	Daily O3 median concentration (µg/m3)
MED_TR_O3_LAG_1D	24h lag variable of O3 concentration (µg/m3)
MED_TR_CO	Daily CO median concentration (µg/m3)
MED_TR_CO_LAG_1D	24h lag variable of CO concentration (µg/m3)
MED_TR_PM25	Daily PM2.5 median concentration (µg/m3)
MED_TR_PM25_LAG_1D	24h lag variable of PM2.5 concentration (µg/m3)
MED_TR_PM10	Daily PM10 median concentration (µg/m3)
MED_TR_PM10_LAG_1D	24h lag variable of PM10 concentration (µg/m3)

To assess the relationship between relevant factors in the NO_x air pollutant lifecycle described in the literature, a model with target dependent variable NO_2 was fitted including both primary urban air pollutants that occur nearly co-linearly with NO_x, such as CO, and secondary pollutants for which NO_x are precursors, such as O_3 and PM_{10}. It was also fitted with additional data features such as meteorological parameters and mobility indexes as a proxy for anthropogenic activities.

To perform this initial analysis, TPOT genetic algorithm optimizer ran 100 generations and 1000 retained individuals in every generation with a total of 101.000 model fits evaluations using sklearn KFold cross-validation with 5 folds on top of 75% of the original data (N=381) whereas the remainder 25% were kept for model performance testing purposes. In this initial run no specific template was used and TPOT ran with full autonomy, using all the 30 processors in the default Regressor configuration [60], to build a pipeline with whatever stages and with whatever hyperparameters that maximizes the NO_2 model accuracy. One of the downsides of this approach is that there's a good chance that complex pipelines and ensembles comprised of multiple stages pre-processing, feature engineering and selection, and potentially the stacking of multiple regressors estimators, which will pass on their outputs as synthetic features to the next processor, will render the built model uninterpretable. Nevertheless, it is a very good way to initially assess the predictive capability of the available data and have not only a model performance baseline but also an idea of the types of processors that work best for the problem at hand.

One important aspect of AutoML is the relationship between training efficiency and achieved scores since the complexity to find a performant pipeline requires not only experimenting with different pipelines, which are composed of a number of sequential, stacked and unioned processors for the several modeling stages, but also the hyperparameter optimization of each of the processors. Several mechanisms in TPOT AutoML Framework allow the management of this trade-off namely the definition of the genetic optimization algorithm parameters such as the number of generations, which is the number of optimization iterations, population size, which is the number of individuals retained in every generation, and the other usual genetic algorithm parameters such as the offspring size, mutation rate and crossover rate. These parameters will ultimately define the number of model fits if the optimization process is not stopped midway in a manual or automatic fashion. A full optimization run is neither a guarantee that an optimal

solution will be found or that an optimal solution will not be found at an earlier phase of the optimization process. In order to optimize the process, one can use several mechanisms to limit how long the optimization process takes namely a global budget of time that overrides the number of generations, thus emitting the best pipeline in the Pareto front with the best score, or a maximum budget to optimize a single pipeline, thus avoiding that complex pipelines monopolize the optimization process time budget. Finally, in the case where an early near-optimal solution is found, the optimization process will naturally not be able to improve the performance of the pipelines for a number of generations which might be a condition for an automated early stop which is configured as the number of optimization generations where the Pareto front score has not improved.

From an infrastructure perspective, the parallelization of workloads is also extremely important to speed up the optimization process which can be done using multiple CPU cores in a single machine or in multiple machines using a distributed task scheduler. Since there's a high probability that the same pipeline will be scheduled for training by the genetic optimization algorithm over time in different generations, caching the pipeline training results will avoid recomputation. To avoid losing results in case of failures before the end of the optimization process it is also possible to checkpoint intermediate results of the optimization process to be able to resume it at a later time.

Table 9 – Performance metrics of the Initial exploratory NO2 prediction pipeline on the test set

Metric	**Result**
Mean Absolute Error (MAE):	2.768893312153063
Mean Squared Error (MSE):	11.927811359229048
Root Mean Squared Error (RMSE):	3.4536663647823667
Mean Squared Log Error (MSLE):	0.01841957017242524
R^2 Score (R2):	0.9246146971059643

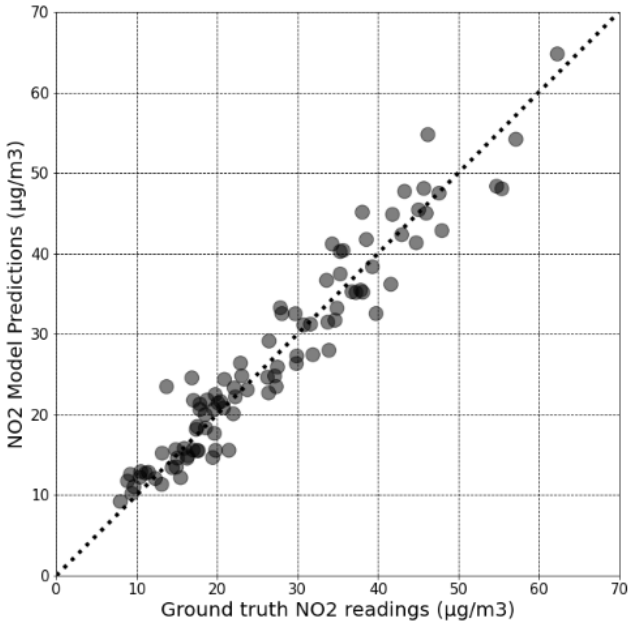


Figure 55 – Initial exploratory AutoML based NO2 prediction performance

This initial TPOT execution took roughly six days and two hours in a D4a_v4 Azure Virtual Machine (4 vCores, 16GB RAM and SSD Volume) using all the machine cores, keeping a cache of intermediate pipeline results to avoid unnecessary recomputations and periodically checkpointing the optimization process progress. This execution resulted in a complex pipeline (Figure 56) with nine stages, four pre-processors, one feature-selector and four stacked regressors, with a resulting average KFold Cross-Validation score on the training set of 12.87 (MSE) and 11.92 (MSE) on the 25% of records that was held for model testing. Having a similar MSE on both the CV (train + validate set) and test set is a good sign that the model is not overfitting or underfitting. While there was another pipeline that performed better on the training set with MSE=12.32, the reported pipeline performed better on the test set (Figure 55 and Table 9) and was therefore selected as the reference NO_2 prediction regressor ML pipeline.

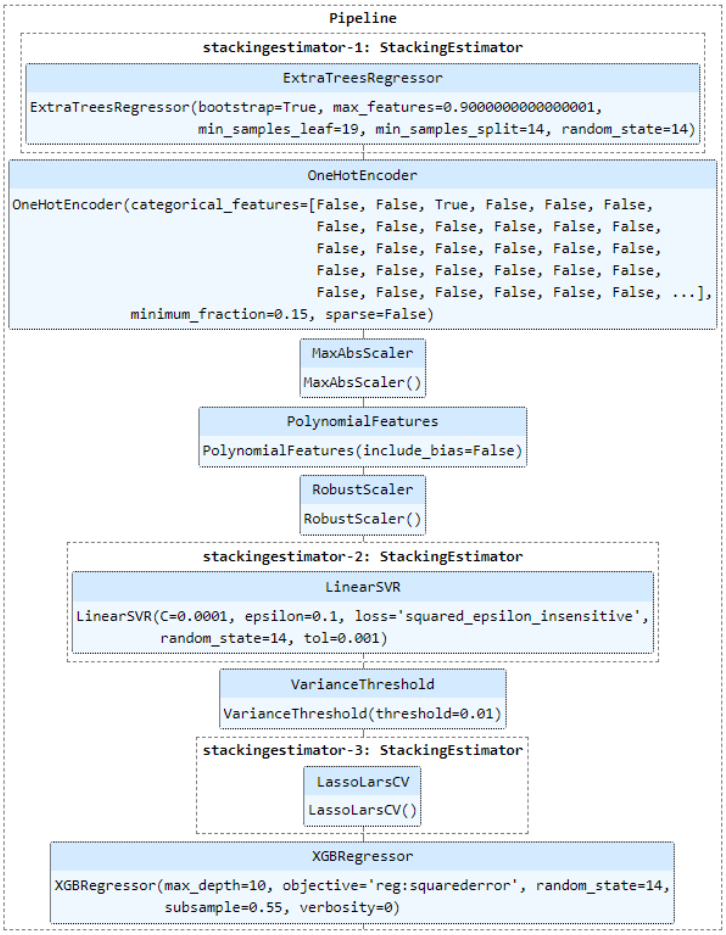


Figure 56 – Initial exploratory AutoML based NO2 prediction pipeline stages and hyperparameters

During this first run, the pipeline cross-validation score evolved rapidly during the first phases of the optimization process progress and slowed down on later phases with each generation taking longer to complete which is expectable since the average complexity of the pipelines increases over time with additional pipeline stages, unions, and stacks, that take longer to compute. The time required to complete a generation stabilized at around generation 40 since the optimizer focused on a similar cost and complexity search space for the remainder of the optimization process. It took 87 generations in 100 to achieve the best score, which took 122 hours in a total of 146 hours, and it took 36 generations in 100 to achieve 80% of the optimization from a min-max perspective, which took 32 hours in a total of 146 hours. A depiction of how this process unfolded throughout the optimization process can be found in Figure 57 with the highlighted area representing the fulfilment of 80% of the training CV score optimization.

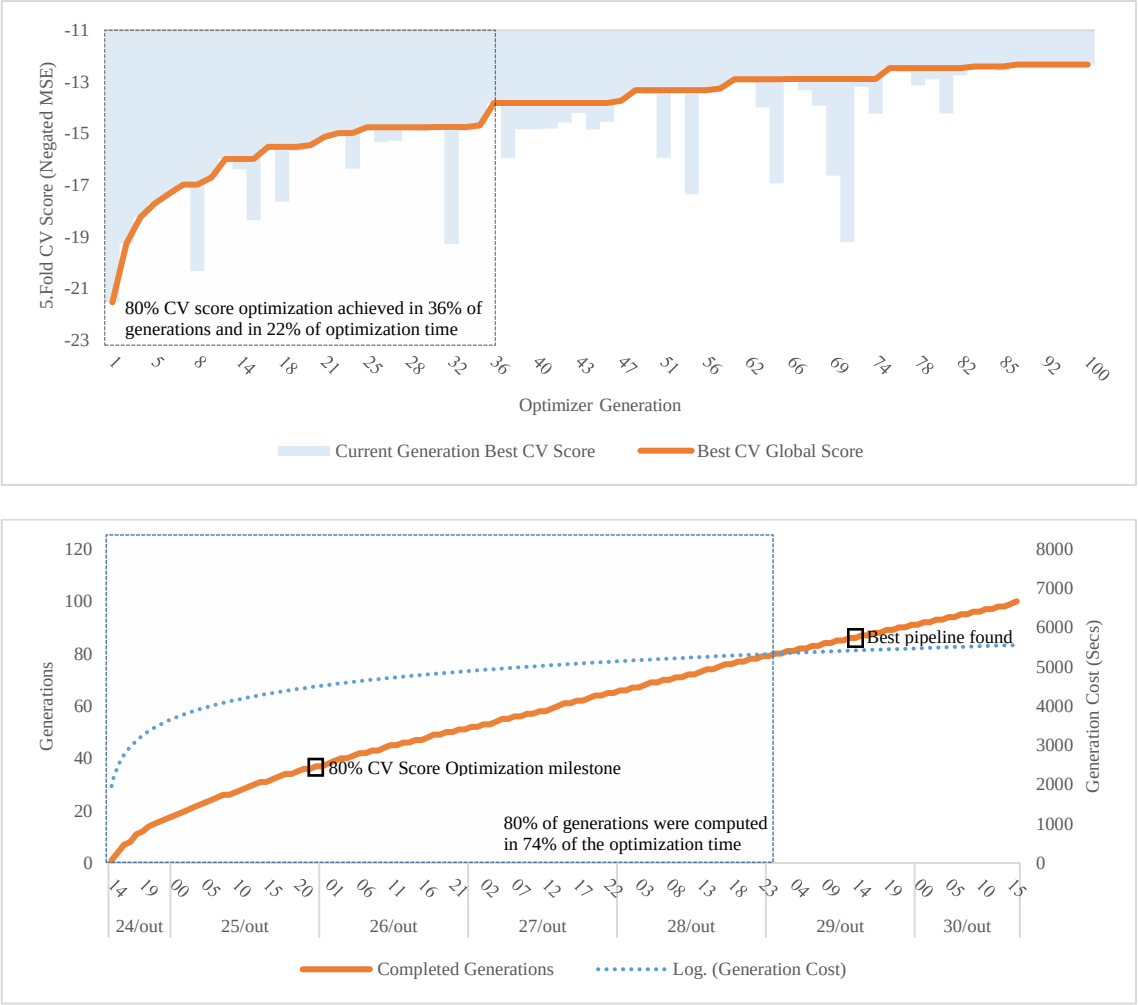


Figure 57 –TPOT AutoML training efficiency

As referred earlier, the best TPOT AutoML produced pipeline from this experiment, while a valid model capable of predicting NO_2 concentration with a reasonable performance, is quite hard to interpret since the initial data features have been transformed multiple times by several processors in different stages of the pipeline. This only leaves the option of trying to breakdown the stages of the pipeline and diving into a rabbit-hole-like analysis on the internals of the underlying machine learning library in an attempt to map the original data features to their intermediate forms and then assess regressor feature importance methods. To further explore feature importance analysis and their contribution to the NO_2 concentration, two other simpler approaches can be taken. Firstly, the TPOT AutoML framework allows the usage of templates that, when properly setup, will generate only interpretable pipelines. Secondly, there is also the possibility to analyze the top performing pipelines of the initial TPOT execution and figure out the main pre-

processing, selectors and regressors, along with their hyperparameters, that yield the best results and remove the stages that reduce interpretative capacity. Both these approaches are part of the interpretability versus performance trade-off and are expected to reduce the model accuracy but, since we have a base model to compare to, it is possible to measure the performance gap and assess its impact on the analysis.

Analyzing the top 20 pipelines generated by TPOT sorted by Cross-Validation score, thirteen had as main or final regressor the XGBRegressor (eXtreme Gradient Boosting Regressor) from the xgboost python library [61], including the topmost performant already reported above, and seven had as the main or final regressor the LassoLarsCV from the sklearn library [62]. Besides scoring slightly better, all resulting XGBRegressor pipelines used a tree-based booster, which is not sensitive to monotonic transformations, so no data feature standardization is required and since all raw variables are numerical and continuous, no special encoding is also needed. On the interpretability side, XGBRegressor is considered an interpretable algorithm out-of-the-box, with specific methods to detail feature importance with different measures which are common in tree-based algorithms namely the feature weight which is the number of times a feature is used to split the data across all trees, feature cover which is the average coverage across all splits the feature is used in and feature gain which is the average gain across all splits the feature is used in [63]. Additional advanced algorithms to assess feature importance for XGBRegressor are available and will be used in this work.

In order to train an interpretable and performant NO_2 prediction model, another TPOT execution was done but in this case a template was used to force the AutoML framework to build and optimize a simple pipeline that consists of a single XGBRegressor processor. Apart from configuring TPOT to optimize a pipeline consisting of a single XGBRegressor processor stage, the rest of the parameters of the optimizer were kept as in the first experiment, meaning that there will be a total of 101.000 model fits. No change has been done on the original data set and this pipeline was trained with the same 75% original observations, which is enforced by the usage of the same initialization seed number used in the train-test split method, whereas the remainder of the 25% of the original observations were retained to further assess the performance of the model in a train + validate (5-Fold CV) holding out an additional test set. The selected hyperparameters are described in Table 10.

Table 10 – NO2 concentration interpretable model XGBRegressor optimized hyperparameters

Hyperparameter	**Value**
objective	'reg:linear'
learning_rate	0.03
max_depth	4
min_child_weight	1
subsample	0.4
colsample_bytree	0.8
n_estimators	500

The average R^2 scores on the left-out validation folds of the K-Fold cross validation is 0.893 and the R^2 score on the remainder 25% test set was 0.889, which seems to have produced a valid model (Table 11 and Figure 58), only slightly less accurate than the original TPOT generated pipeline which had R^2=0.925 on the test set.

Table 11 – Performance metrics of the NO2 concentration interpretable model on the test set

Metric	**Result**
Mean Absolute Error (MAE):	3.2180649300416277
Mean Squared Error (MSE):	17.16391991334616
Root Mean Squared Error (RMSE):	4.142936146423954
Mean Squared Log Error (MSLE):	0.021681443615292815
R^2 Score (R2):	0.8890452484554431

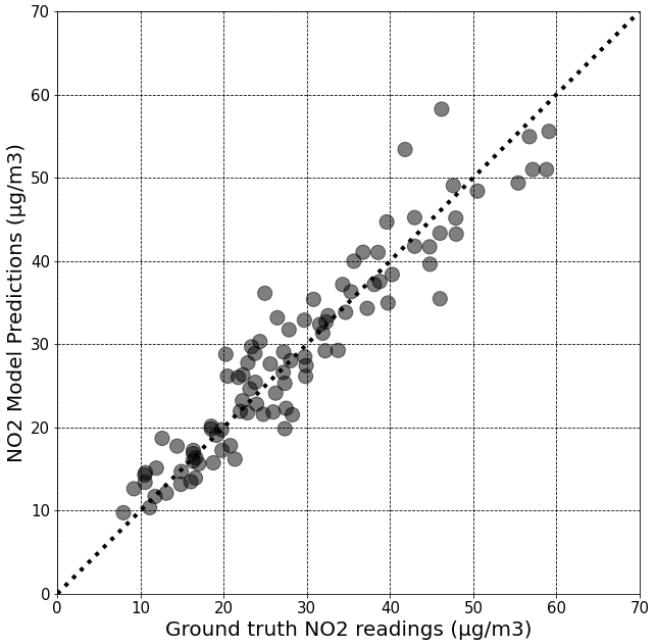


Figure 58 – NO2 concentration interpretable *model performance on the test set*

Feature importance analysis in the present work is done using the cross-algorithm game theory based approach SHAP (SHapely Additive exPlanations) [64], which connects optimal credit allocation with local explanations using the classic Shapley values from game theory and their related extensions, to assess the impact of one independent data feature on the regressor output, as well as allowing to study pairwise relationships interactions to uncover more complex and rich relationships between independent variables on the dependent variable. The SHAP implementation used is a polynomial time complexity implementation specific for tree-based algorithms or ensembles [65] such as the XGBRegressor being used in the present work. Some of the advantages of using SHAP is that we can have single prediction level feature importance analysis, as well as global importance feature analysis, instead of global-only feature importance analysis such as the weight, cover and gain metrics reported by the XGBRegressor. Additionally, the SHAP implementation removes the complexity of analyzing three different XGBRegressor feature importance metrics (gain, weight, and cover), which are sometimes inconsistent between themselves and are biased towards some particularity of the metric and data features, and instead allows the usage a unified approach to a more accurate and consistent feature importance in Machine Learning tree-based algorithms.

To interpret the SHAP values computed from the regressor predictions, assume that SHAP values computed for an independent variable (i.e: Air Temperature) from a list of predictions (each point) yield a negative value if it pushes down the predicted value for the dependent variable or up if the SHAP value is positive. In pairwise analysis, the red-blue color gradient is associated with the additional independent variable domain value range and not the main independent variable SHAP value.

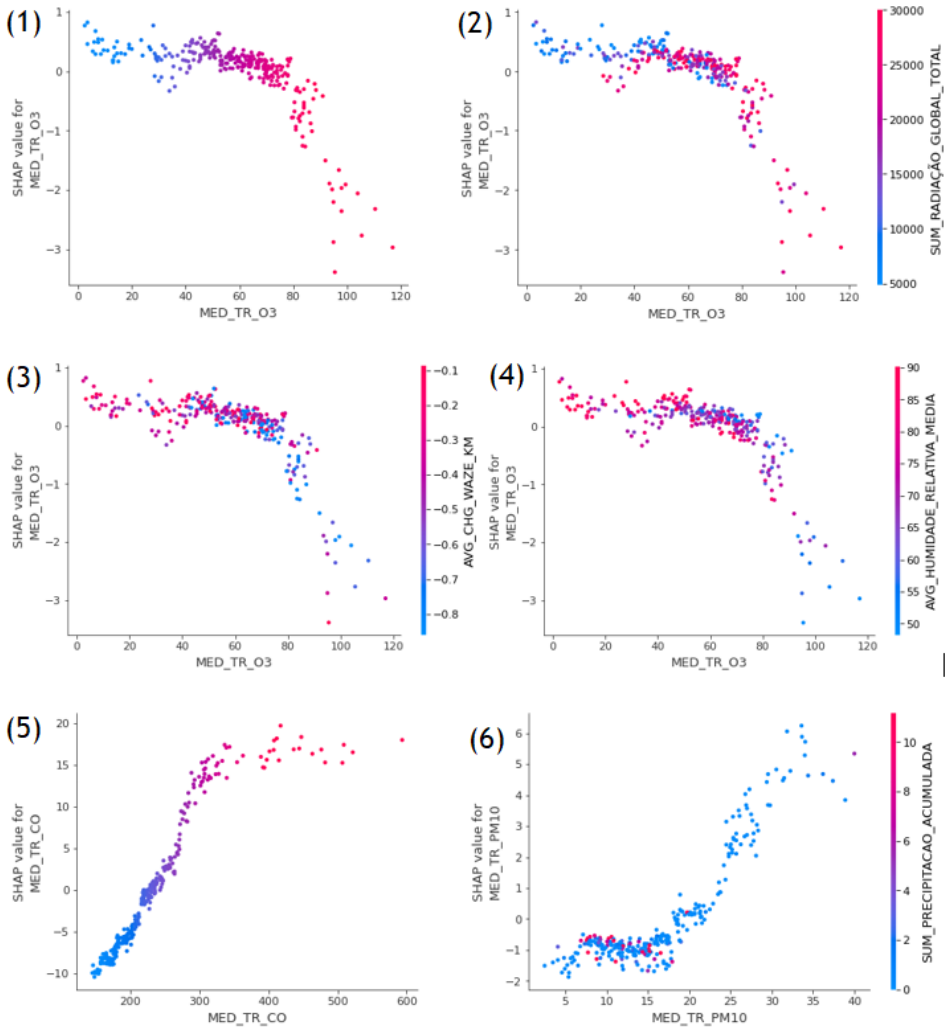


Figure 59 – NO2 concentration interpretable model *feature importance analysis*

Regarding the under the sunlight chemical reaction of NO_2 with O_3, in Figure 59 it is clear that the model favors positive NO_2 concentrations under lesser concentrations of O_3 (1) and that the interaction of the Global Radiation (kJ/m2) variable plays a role in the relationship between NO_2 and O_3 (2), where in the presence of high solar radiation it favors greater O_3 concentration and less NO_2 concentration as described in the literature. Additionally, the model favors higher O_3 concentrations under reduced traffic days (3) potentially due to the reduction of O_3 titration by NO, and it also favors higher O_3 concentration and lower NO_2 concentration with lower relative humidity (4), potentially due to the usual cloud overcast associated with higher humidity levels and the enhanced effect of O_3 dry deposition under higher humidity levels [66]. As for CO, a pollutant that as NO_x is also emitted as part of the combustion process, it expectedly increases nearly

linearly as NO_2 increases (5). Finally, a more complex relationship, where PM_{10} increases as NO_2 increases since NO_2 is a precursor to secondary PM_{10} and also because primary PM_{10} is also caused by vehicles due to dust resuspension, asphalt, tyre, and brake abrasion as well as during fuel combustion process [67], thus co-occurring with NO_x emissions, further more related to the particle wet deposition mechanism that comes along with increased precipitation where it is clear that rainy days yield less PM_{10} concentrations (6).

3.10. Anthropogenic and Meteorological contributions to NO2 Concentration

In this experiment, we no longer include air pollutants for which NO_2 is a precursor, such as O_3 and secondary PM, and nearly co-linear primary pollutants such as CO, thus focusing on the underlying meteorological and anthropogenic factors that contribute directly to urban NO_x concentrations in an attempt to understand the contribution of such factors and analyze any pairwise interaction between several independent variables.

The XGBRegressor hyperparameters in Table 12 were optimized with TPOT genetic algorithm optimizer using 100 generations and 1000 retained individuals in every generation with a total of 101.000 model fits evaluations using sklearn KFold cross-validation with 5 folds.

Table 12 – NO2 causes model XGB Regressor Hyperparameters found by T-POT

Hyperparameter	Value	Hyperparameter	Value	Hyperparameter	Value
base_score	0.5	learning_rate	0.08	objective	reg:linear
booster	gbtree	max_delta_step	0	reg_alpha	0
colsample_bylevel	1	max_depth	3	reg_lambda	1
colsample_bynode	1	min_child_weight	3	scale_pos_weight	1
colsample_bytree	0.7	missing	None	seed	0
gamma	0	n_estimators	1000	silent	1
importance_type	gain	nthread	1	subsample	0.4

This model is less accurate (R^2=0.889) on the test set than the model trained in the previous section (R^2=0.836), which in turn is also less accurate than the original TPOT NO_2 Concentration prediction model with R^2=0.925, due to the fact that the data features representing air pollutants emitted simultaneously with NO_2, such as CO, and those for which NO_2 is a precursor, such as O_3, $PM_{2.5}$ and PM_{10}, were removed so the anthropogenic contribution of the NO_2 lifecycle can be studied without being overshadowed by other factors which are also good NO_2 predictors but of no interest in this exercise. (Table 13 and Figure 60)

Table 13 – Performance metrics of the NO2 cause model on the test set

Metric	**Result**
Mean Absolute Error (MAE):	4.619360131633524
Mean Squared Error (MSE):	31.216223957493977
Root Mean Squared Error (RMSE):	5.587148105920764
Mean Squared Log Error (MSLE):	0.05503345698523074
R^2 Score (R2):	0.836004877566092

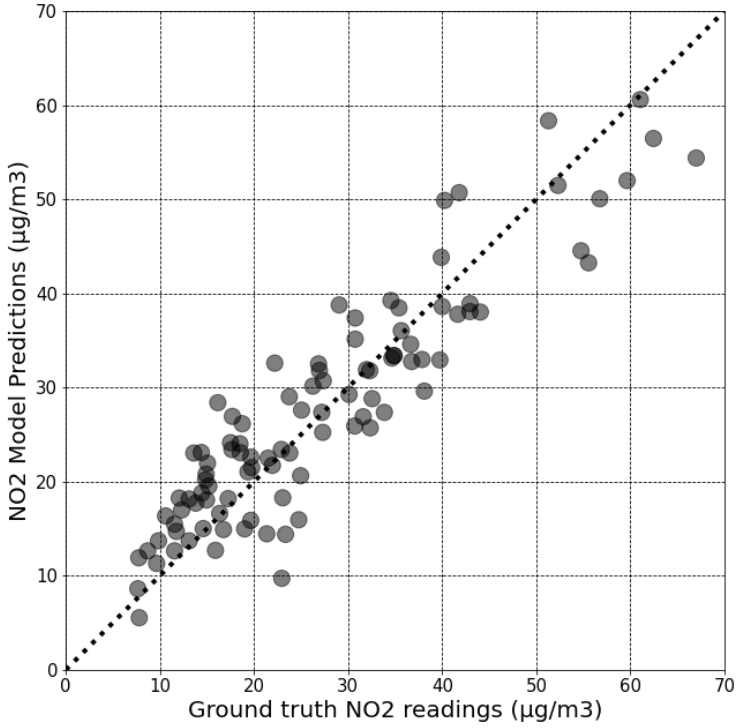


Figure 60 – NO2 cause model performance on the test set

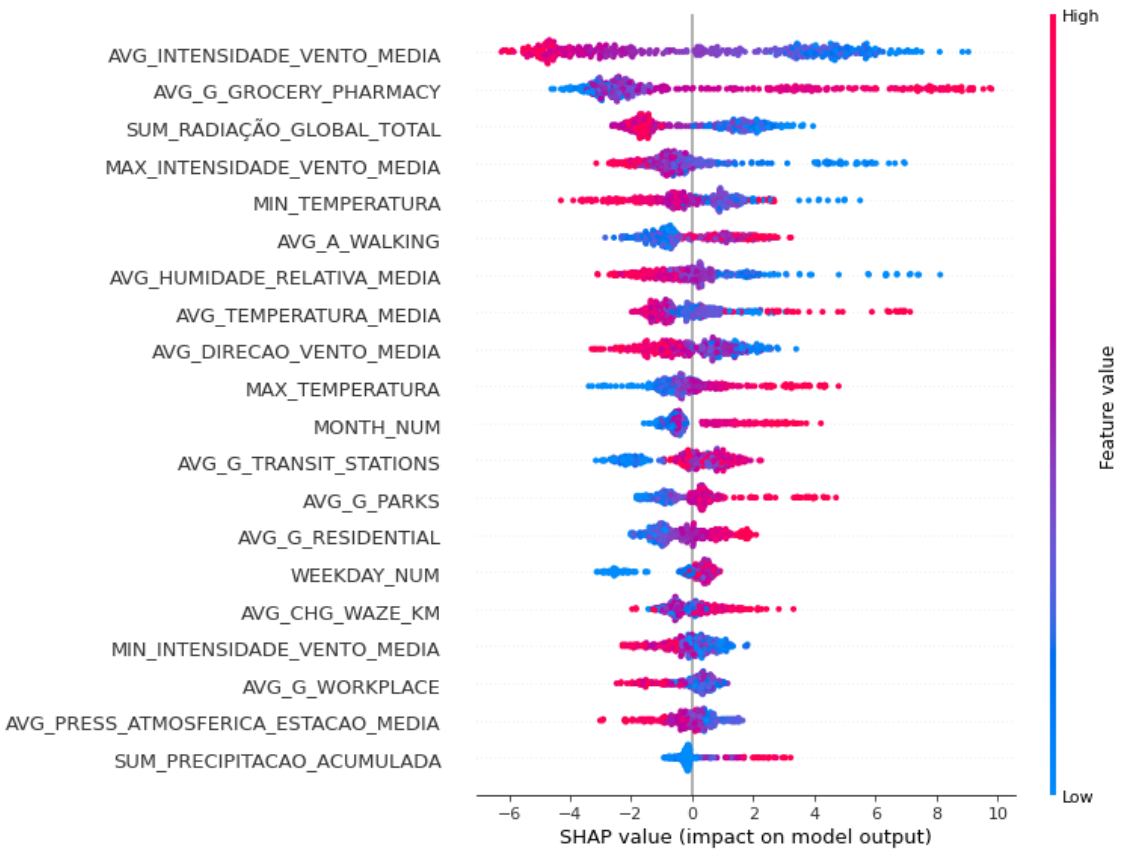


Figure 61 – Summary of feature importance analysis

The SHAP value per data feature for all model train observations can be observed in the summary plot depicted in Figure 61. The main associations found during the statistical analysis in section 3.8 can be mapped to the importance the model has associated to the anthropogenic and meteorological data features. For instance, the daily average driven kilometers measured with Waze App and Google Local commerce mobility index impacts NO_2 concentration predictions negatively (lower SHAP values) when the data feature value is lower (bluer predictions), indicating that less driven kilometers or lower mobility levels are associated to lower NO_2 air pollutant concentration, and vice-versa. Similarly, the daily average wind intensity negatively impacts NO_2 concentration when the wind intensity is higher, meaning that high winds are associated with lower concentrations and relatively stagnant air (0.0-2.5 m/s) is associated with higher NO_2 concentrations.

The mean absolute value of SHAP values, depicted in Figure 62, tries to improve the information provided by averaging all prediction SHAP values by its absolute value so that negative contributions are not cancelled out by positive contributions and vice-versa.

Take in consideration for example weekday number, there's a lot of predictions where this data feature caused an increase (positive SHAP values) of NO_2 whereas in other cases it caused a decrease (negative SHAP value), for instance weekdays versus weekends. Taking this into consideration, as shown in Figure 62,the model finds the anthropogenic activity proxy data feature, measured by Google as user activity levels on Local Commerce, as the topmost contributing independent variable from an anthropogenic source with a mean absolute SHAP value of +3.45, followed by the daily average wind intensity, which is a natural air pollutant dissipation mechanism, with a mean absolute SHAP value of +3.70, and considered the topmost contributing independent variable from a natural source, and then by the daily total solar radiation with +1.66 which is a very important potentiator of photochemical reactions in the atmosphere, namely the breakdown of NO_2 and O_3 production cycle.

By grouping all independent variables in either the anthropogenic or natural class, the mean absolute SHAP values can be summarized to assess which group contributes the most to NO_2 concentration. This model attributes a total of +13.98 summed mean absolute SHAP values to the natural class and +9.79 to the anthropogenic class, which suggests that natural phenomena are responsible for a greater share of NO_2 concentration formation. For this analysis the month number was not taken into consideration since it might relate to anthropogenic phenomena (holiday periods, school periods, etc) and natural phenomena (weather patterns change throughout the year).

Zooming into the two most important independent variables, a Google mobility activity indicator which acts as an anthropogenic pollutant activity proxy variable, and the daily average wind speed, which is a natural air pollution dissipation mechanism, it is clear the correlation between their values and the negative or positive contribution to the NO_2 concentration value for each of the predictions. For the Google mobility indicator, higher values contribute to higher NO_2 concentrations while the opposite is verifiable for the average wind speed where higher values contribute to lower NO_2 concentrations. As a complex and non-linear phenomenon, NO_2 concentration in urban environment depends on a number of variables that together can generate severely polluted or exceptionally clean air, and this appears to have happened in the most severe periods of the COVID-19 pandemic lockdowns, namely in several periods of April/2020, when pollutant anthropogenic activities dropped severely as a result of the lockdowns and the weather parameters were not particularly favorable for NO_2 buildups to occur, namely with

depressionary atmosphere with moderate winds and rain, leading to unprecedented low NO_2 concentrations.

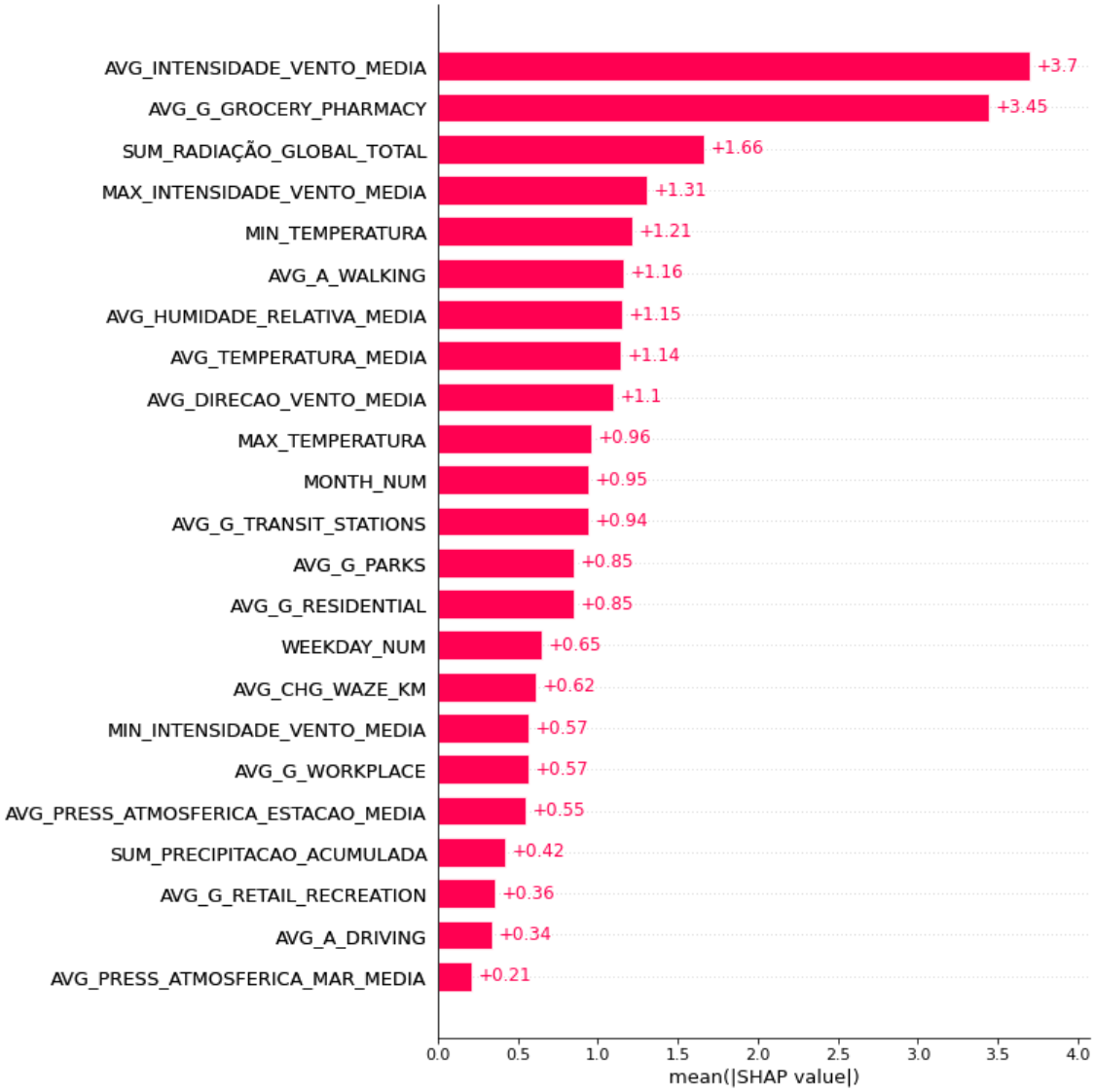


Figure 62 – Summary of feature importance analysis (mean absolute SHAP)

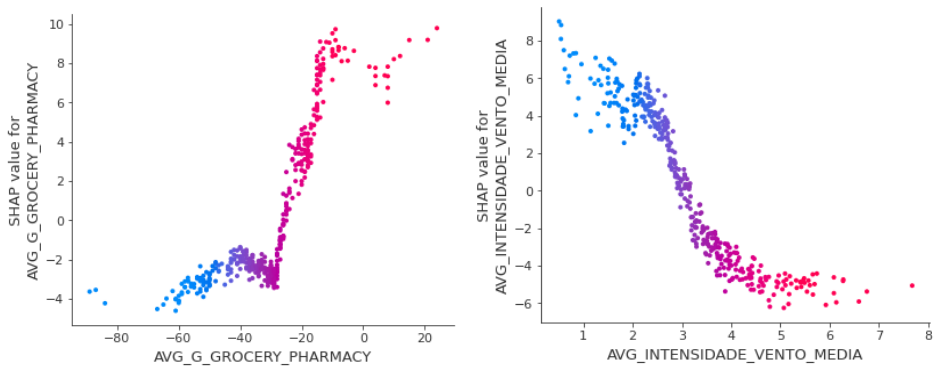


Figure 63 – SHAP Values for the two most important independent variables (Daily Wind mean intensity and Daily Avg Google Local Commerce Activity Index)

To further analyze particular predictions, let's take into consideration the days with the lowest and highest NO_2 concentrations from the March/2020 to March/2021 dataset, 5th of April 2020 with a median NO_2 concentration of 5.80 µg/m3 and 12th of September 2020 with a median NO_2 concentration of 73.05 µg/m3. In these two particular days, the NO_2 prediction model estimates 5.69 µg/m3 and 73.02 µg/m3 respectively with an error of -0.11 µg/m3 and -0.03 µg/m3. The detailed feature importance analysis at the prediction level allows a concrete explanation of the factors leading up to these values.

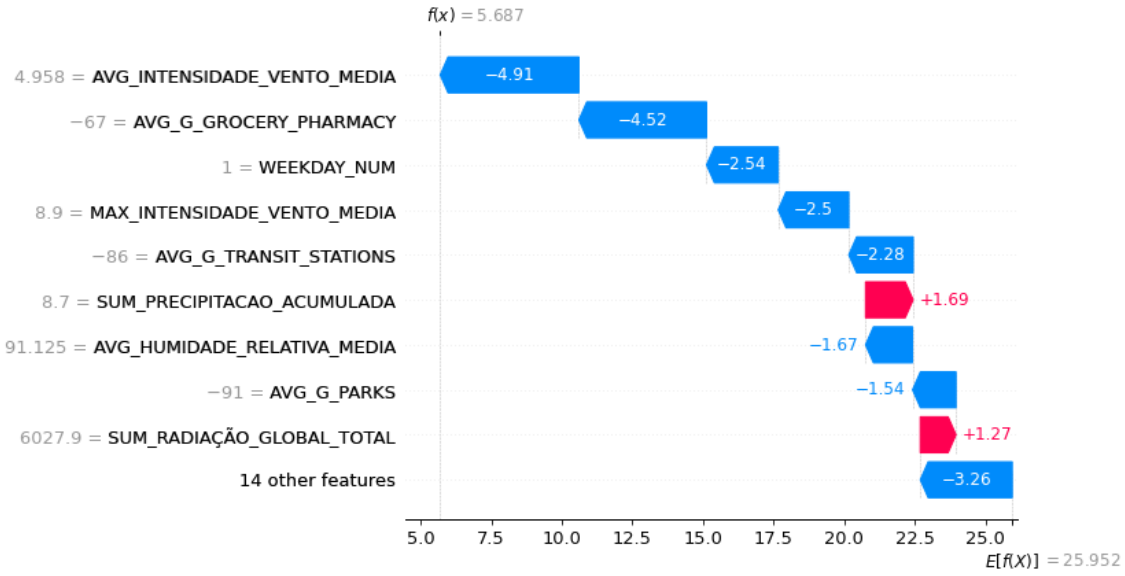


Figure 64 – Top 10 contributing variables for the exceptional low NO2 concentration registered in 05-04-2020

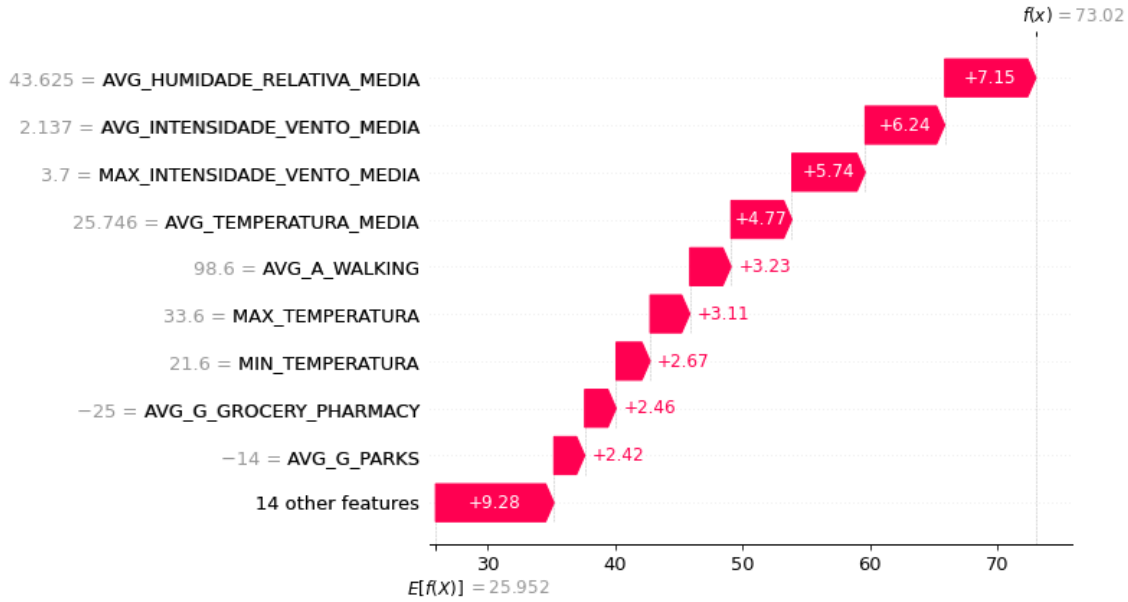


Figure 65 – Top 10 contributing variables for the higher NO2 concentration registered in 12-09-2020

As for the 5th of April 2020 day, which was deep into the first national emergency period where mandatory confinement was in place, it was a particular windy day as depicted in Figure 64 with an average daily wind speed of 5.0 m/s and maximum hourly wind speed of 8.9 m/s which acts as an effective dispersion mechanism of NO_2 air pollution. This already NO_2 sub-optimal condition was further aided by severely reduced anthropogenic activities measured by urban mobility indicators due to the hard confinement phase and because it was a Sunday. This combination of negative contributions to NO_2 concentrations made this particular day the least NO_2 polluted day in Lisbon from 2013 to early 2021.

A completely opposite example is the 12th of September 2020 which, while not even close to the most NO_2 polluted days in Lisbon on the records, is a good example of a moderately NO_2 polluted day. It was a hot Saturday with stagnated air and moderate cloud cover as depicted in Figure 65, measuring minimum hourly wind speed of 0 m/s, maximum hourly wind speed of 3.7 m/s and average daily wind speed of 2.1 m/s, which doesn't help disperse NO_2, higher urban mobility activity, meaning higher NO_x emissions, and sub-optimal total solar radiation that didn't allow NO_2 to O_3 conversion to happen at a greater rate, generating moderately high NO_2 concentrations.

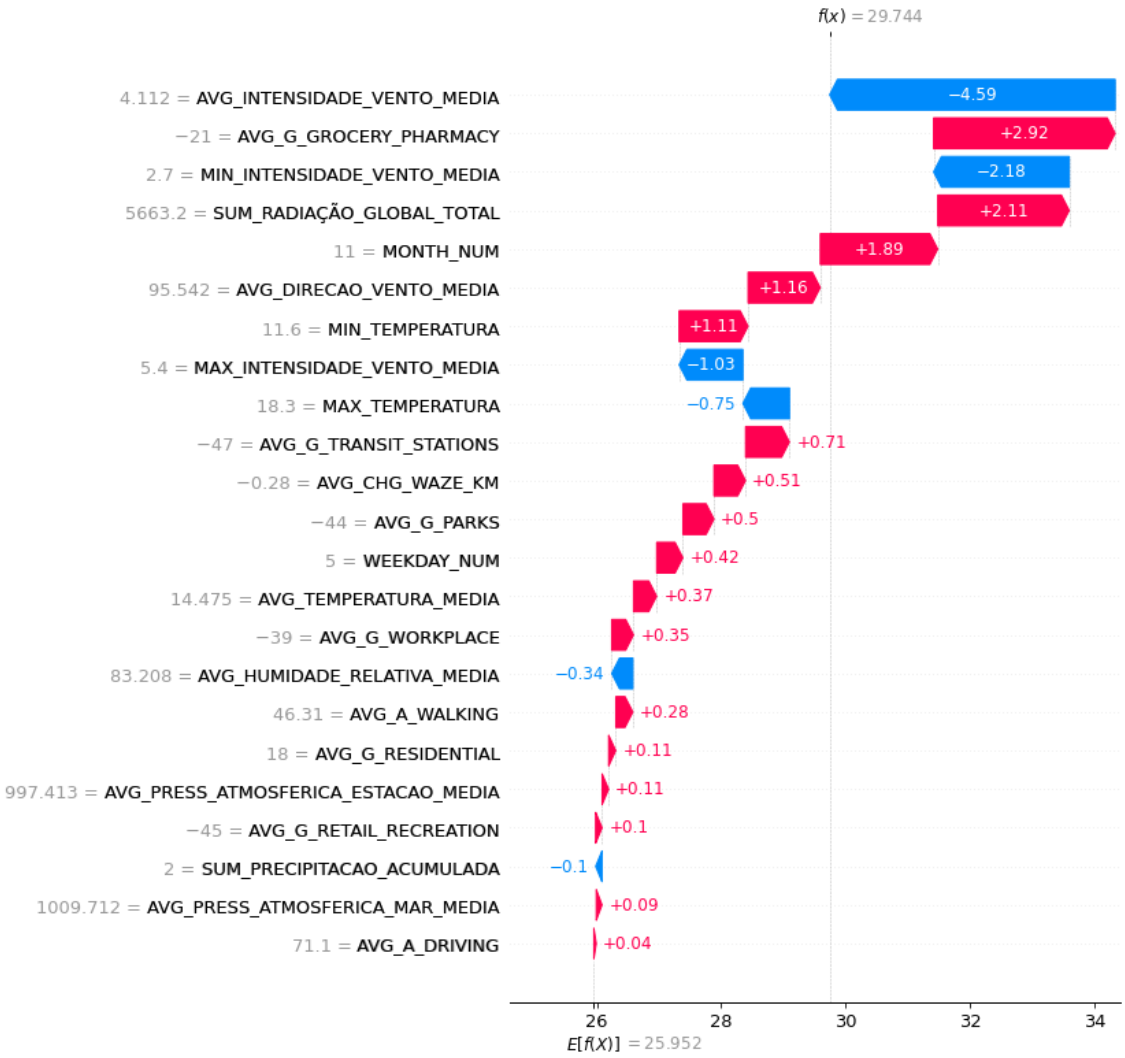


Figure 66 – All variables contribution to NO2 concentration registered in a regular day (11-11-2020)

While these two extreme examples seem rather intuitive and simple to analyze, by picking up a day with NO_2 concentration not to low and not to high, which means opposing forces drive the pollutant value to a more moderate value, it is clear how non-linear the NO_2 estimation is to the model when inspecting the feature contribution analysis. To inspect such a case, the 11th of November 2020 was chosen for the exercise having measured a NO_2 concentration of 29.75 µg/m3 with the NO_2 concentration prediction model estimating 29.74 µg/m3 with a measured error of -0.01 µg/m3. As depicted in Figure 66, this particular day NO_2 median concentration is affected by multiple natural and anthropogenic phenomena, each with its positive or negative contribution, with higher or lower strength, thus allowing for an in-depth analysis of, for instance, a

specific severe air pollution episode limited in time that require a detailed diagnosis. The main variables contributing to the NO_2 concentration in this particular day were that it was a moderately windy day, with an average daily wind speed of 4.1 m/s, maximum hourly wind speed of 5.4 m/s, and minimum hourly wind speed of 2.7 m/s, meaning that it is not particularly optimal for NO_2 buildups. On the other hand, urban mobility indicators show that it was a day with higher activity levels which translates to increased NO_2 pollutant anthropogenic activities such as road traffic, aided by the fact that atmospheric photochemistry performance was reduced due to lower total solar radiation meaning that there's reduced NO_2 photolysis which drives concentrations up.

Chapter 4 – Conclusions and future work

4.1. Main conclusions

As for the impact of COVID-19 pandemic related restrictions in Lisbon on the urban mobility, it is estimated that during the first national emergency period (18-03-2020 to 03-05-2020) public transportation ridership has suffered a homologous drop of 75% to 80% while during the first calamity state period (04-05-2020 to 01-07-2020) it is estimated to have dropped 57% to 62%. Regarding the usage of private vehicles, during the first national emergency period, it is estimated to have suffered a reduction of 78% to 84% while during the first calamity state period it is estimated to have dropped 43% to 59%. Residential area mobility activity has increased 33% during the first national emergency period, decreasing to 19% during the first calamity state period. The urban mobility indicators used in the present work correlate moderately with the NO_2 air pollutant which is usually associated to anthropogenic activity in the city, whereas the local commerce indicator (Google) has the strongest Pearson correlation (r=0.54) and has also been identified as the main anthropogenic data feature contributing to NO_2 concentration by the trained NO_2 concentration prediction pipeline.

Regarding the impact of COVID-19 pandemic related restrictions in Lisbon on the air quality, during the first national emergency period (18-03-2020 to 03-05-2020) and subsequent calamity state period (04-05-2020 to 01-07-2020) the main criteria air pollutants have generally decreased in both urban background and urban traffic stations when compared to a 2013-2019 baseline with the exception of Ozone (O_3) in urban traffic stations which have increased. With a sharp reduction in anthropogenic activities, most

importantly in road traffic, and with depressionary weather suboptimal for NO_2 build-ups in the first phase, NO_2 registered a 54.35% drop in the first phase and 40.39% in the second phase in urban traffic stations, while in background stations it dropped 28.62% and 22.99%, respectively. Such unprecedented drops in NO_2 concentrations caused the usually polluted Av. Liberdade station to comply with the regulatory yearly 40 µg/m3 mean threshold value in 2020 which, as far as the available data, this particular station had never been able to meet. Regarding O_3 secondary pollutant, during the first phase there was an increase of 12.89% in urban traffic stations and 6.76% in the second phase, while in background stations it decreased 8.42% in the first phase and 5.43% in the second phase. Tropospheric Ozone concentration increases in urban traffic monitoring stations suggests relevant changes in the NO_x:VOC ratio and reduced O_3 titration by NO, a result of sharp decrease of NO_x emissions in the usually most polluted and VOC-limited city hotspots. This finding raises the need of additional measures to mitigate O_3 pollution increases as part of the Lisbon and Tagus Valley air quality improvement plan which aims to reduce NO_2 concentrations, namely specific measures for VOC management.

An AutoML framework (TPOT) was used to build, train, and optimize a regressor ML pipeline to predict NO_2 concentration with the available anthropogenic activity, weather, and air pollutant inputs from March/2020 to March/2021, achieving R^2=0.925 out of the box on the test set. This is an acceptable result for an AutoML approach when compared to other recently purpose-built NO_2 prediction models such as R^2=0.920 [68], R^2=0.890 [69] and R^2=0.937 [45]. Further model performance improvements could be achieved by tackling some of the research limitations documented in Chapter 4.3. An interpretable NO_2 prediction model was trained to perform feature importance analysis which, from a global perspective, uncovered that anthropogenic features contribute to 41.19% of NO_2 concentrations and natural phenomena features contribute to 58.81%, with the average daily wind speed feature as the most important feature, followed by the Google daily local commerce activity indicator. At the individual prediction level, it was also provided a technique to intuitively inspect the contributing factors and their strengths leading to that particular estimation which is highly relevant to perform the diagnosis of a particular severe pollution episode.

4.2. Research limitations

The mobility impact assessment during COVID-19 pandemic restrictions is missing data for other main mass public transit operators in the city such as bus, train, and boat operators (Carris, CP Suburbano and Fertagus, Transtejo and Soflusa) for which the data could not be sourced and thus not reported.

All mobility indexes used in the present work were provided by entities that have access to mobility data, such as Google, Apple, Moovit and Waze, which in the wake of the COVID-19 pandemic, in an attempt to help public authorities make better decisions on pandemic management, started collecting, computing, and reporting these mobility indicators. For this reason, all these indicators, which are good proxies for anthropogenic pollutant activities, were unavailable before January 2020 and therefore limited the amount of data available to train urban air pollution concentrations ML models.

The present work did not include any VOC concentration data in most analysis due to data quality issues and unavailability of VOC monitoring data. VOCs are known to be an important factor in O_3 production and while the emissions of anthropogenic VOC in Lisbon, from both mobile and fixed sources, is assumed to have been reduced in wake of the impact of lockdown restrictions, it is assumed that it is not a significant driver in changes to the NOx/VOC mixing ratio since biogenic VOCs largely outweigh anthropogenic VOCs, 89% versus 9% of all non-methane VOC emissions respectively in LMA [68], and the analysis period is not large enough to consider significant changes in biogenic emissions.

Due to the unavailability of suitable hardware and time constraints, no Neural Network or GPU-assisted cuML/DMLC estimators were used and the hyperparameter search space was not as expansive as it could have been during the automatic machine-learning model genetic optimization process.

4.3. Future research proposals

As to proposed future work, the following items could be further explored to enhance the present or related research.

The findings of the present work, namely the contribution analysis of road traffic to the NO_2 concentration in Lisbon, could be compared with the proposed mitigative actions

to reduce NO_2 and PM_{10} concentrations in the LMA air quality improvement plan [47] that involve manipulating the Mean Daily Traffic (MDT) to achieve specific concentration objectives.

Acquire data and report the COVID-19 pandemic restrictions impact on the remaining Lisbon public transportation modes, such as train, bus, and boat, as well as other shared transportation modes such as taxi or ride-sharing operators and, finally, on shared soft transportation methods such as bikes, e-mopeds, and e-scooters, in order to have a complete picture of the urban mobility dynamics during the pandemic periods.

Update all descriptive statistical analysis and machine learning models for the remaining COVID-19 pandemic phases in Lisbon, more specifically the 2nd National Emergency phase that occurred from the fall of 2020 into spring of 2021, since seasonal effects on air pollution and meteorology between July and March were not analyzed in the present work, and in order to document urban mobility and air pollution impacts throughout the entirety of the COVID-19 pandemic.

It is not clear why CO concentrations have not decreased as much as NO_2 during lockdown period since road transport contributes heavily to both air pollutants. A possible cause could be that CO, being a primary pollutant, which is not regenerated by photochemistry processes such as NO_2, has had a lower drop than NO_2 when comparing to the baseline. Other possible factors could be related to the fact that CO is lighter than NO_2, making it easier to disperse in the atmosphere, or due to the fact the vehicle mix during the lockdown periods were different than the baseline (i.e: diesel vs petrol; light vs heavy; newer vs older). Nevertheless, this disproportional drop in NO_2 and CO concentrations should be further investigated to fully understand this phenomenon.

Investigate noise pollution changes during the most restrictive COVID-19 pandemic phases, namely near Lisbon Airport, main road traffic arteries and intersections, including its effect in citizens well-being.

Furthermore, it would also be interesting to study the effects of lockdown fatigue throughout COVID-19 lockdowns on the effectiveness of pandemic management policies on urban mobility and the spread of the disease in order to aid decision makers in the case of a future pandemic scenario.

Identify any changes on the urban mobility routines by its citizens, workers/students, and visitors' routines after the end of COVID-19 lockdowns in 2021 and its impact to the

commuting efficiency in the city. Propose improvements in the urban mobility offerings to improve efficiency and seize the opportunity to rebalance transportation means dominated by private vehicles during the pre-pandemic period.

Re-run the ML experiments with increased amount of collected data, with additional GPU-assisted computational power to use additional estimators, namely Neural Networks and cuML/DMLC, and to search over an expanded hyperparameter search space in order to increase model performance.

Additionally, as future work, it would be interesting to operationalize the present work into a fully functioning city-agnostic system. Define generic data interfaces for simplified data integration, uplift the ETL process to allow for delta loads, dimension change management and update of the multidimensional model, update visualizations in a public dashboard solution and continuously retrain machine learning models, enabling the interpretation of feature contribution for the past and present assisted by purpose-built visualizations, as well as allowing for short future horizon air pollutant concentration interpretable predictions via dashboard and an inference API.

References

[1] Observatório da Energia, “Consumo de produtos do petróleo: Abril 2020,” 2020. [Online]. Available: https://www.observatoriodaenergia.pt/pt/comunicar-energia/post/8256/consumo-de-produtos-do-petroleo-abril-2020/.

[2] REN, “Consumo de energia elétrica recua 12% em abril,” 2020. [Online]. Available: https://www.ren.pt/pt-PT/media/comunicados/detalhe/consumo_de_energia_eletrica_recua_12__em_abril_3.

[3] Observatório da Energia, “Consumo de Energia Elétrica: Abril 2020,” 2020. [Online]. Available: https://www.observatoriodaenergia.pt/pt/comunicar-energia/post/8254/consumo-de-energia-eletrica-abril-2020/.

[4] Porto de Lisboa, “Estatísticas Porto de Lisboa,” 2020. [Online]. Available: https://www.portodelisboa.pt/estatisticas.

[5] INE, “Movimento Aeronaves no Aeroporto de Lisboa,” 2020. [Online]. Available: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&contecto=pi&indOcorrCod=0000861&selTab=tab0.

[6] Jornal de Notícias, “Mais de meio milhão de carros entope Lisboa diariamente,” *Jornal de Notícias,* p. 1;22, 30 10 2019.

[7] CML, “CML - Relatório Zona Emissões Reduzidas,” 2020. [Online]. Available: https://zer.lisboa.pt/Relatorio_ZER.pdf.

[8] Tom Tom, “Lisbon traffic - 2019 historical traffic data,” 2020. [Online]. Available: https://www.tomtom.com/en_gb/traffic-index/lisbon-traffic/.

[9] M. Nasir, M. A. Kalam, B. Masum e R. Md. Noor, “Reduction of Fuel Consumption and Exhaust Pollutant Using Intelligent Transport System,” *The Scientific World Journal,* 2014.

[10] INE - Estatísticas do Parque de Veículos Rodoviários, “Veículos rodoviários motorizados em circulação: total e por tipo de combustível,” 13 05 2021. [Online]. Available: https://www.pordata.pt/DB/Portugal/Ambiente+de+Consulta/Tabela/5822450.

[11] IMT - Instituto da Mobilidade e Transportes, “Veículos elétricos matriculados,” IMT - Instituto da Mobilidade e Transportes, 2020. [Online]. Available: http://www.imt-ip.pt/sites/IMTT/Portugues/Observatorio/Estatisticas/OutrasInformacoes/Documents/2020/Ve%C3%ADculos%20eletricos%20matriculados_acumulado%20e%20por%20ano.pdf. [Acedido em 14 05 2021].

[12] INE - Estatísticas do Parque de Veículos Rodoviários, “Veículos rodoviários motorizados de passageiros em circulação: total e por idade do veículo,” PORDATA, 20 11 2020. [Online]. Available: https://www.pordata.pt/DB/Portugal/Ambiente+de+Consulta/Tabela/5822451. [Acedido em 13 05 2021].

[13] European Commission, “Commission Regulation (EU) No 459/2012 of 29 May 2012 amending Regulation (EC) No 715/2007 of the European Parliament and of the Council and Commission Regulation (EC) No 692/2008 as regards emissions from light passenger and commercial vehicles (Euro 6),” The European Commission, 01 06 2012. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2012/459/oj. [Acedido em 14 05 2021].

[14] INE, “Mobilidade e funcionalidade do território nas Áreas Metropolitanas do Porto e de Lisboa : 2017,” 2018. [Online]. Available: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_publicacoes&PUBLICACOESpub_boui=349495406&PUBLICACOESmodo=2&xlang=pt.

[15] N. Hooftman, L. Oliveira, M. Messagie, T. Coosemans e J. V. Mierlo, “Environmental Analysis of Petrol, Diesel and Electric Passenger Cars in a Belgian Urban Setting,” *Energies,* vol. 9, nº 2, p. 84, Jan 2016.

[16] Público, “Medina recua e remete restrições ao trânsito na Baixa de Lisboa para próximo mandato,” 2 10 2020. [Online]. Available: https://www.publico.pt/2020/10/02/local/noticia/medina-recua-remete-restricoes-transito-baixa-lisboa-proximo-mandato-1933749.

[17] A. Zero, “Zero apoia fortemente ideia da nova ZER em Lisboa mas quer maior coerência e mais ambição,” 2020. [Online]. Available: https://zero.ong/zero-apoia-fortemente-ideia-da-nova-zer-em-lisboa-mas-quer-maior-coerencia-e-mais-ambicao/. [Acedido em 18 05 2021].

[18] Jornal Público, “Restrições a carros antigos em Lisboa são “mais ficção que realidade”, diz pai da medida,” 2017. [Online]. Available: https://www.publico.pt/2017/01/14/local/noticia/restricoes-a-carros-antigos-em-lisboa-sao-mais-ficcao-que-realidade-diz-pai-da-medida-1758290. [Acedido em 05 18 2021].

[19] CML, “Lisbon Application Form for the European Green Capital Award 2020 (Chapter 3. Sustainable Urban Mobility),” 2018. [Online]. Available: https://ec.europa.eu/environment/europeangreencapital/wp-content/uploads/2018/07/Indicator_3_Lisbon_EN.pdf.

[20] ASF, “Parque automóvel seguro,” 2019. [Online]. Available: https://www.asf.com.pt/ISP/Estatisticas/seguros/estatisticas_anuais/historico/PA%202019.xlsx.

[21] INE, “População residente (N.º) por Local de residência (NUTS - 2013), Sexo e Grupo etário; Anual - INE, Estimativas anuais da população residente,” 2019. [Online]. Available: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0008273.

[22] N. M. d. Costa, “Acessibilidade e Transportes,” em *Atlas Digital da Área Metropolitana de Lisboa*, Lisboa, 2016.

[23] Comissão de Coordenação e Desenvolvimento Regional de Lisboa e Vale do Tejo, “Avaliação da qualidade do ar ambiente na região de Lisboa e Vale do Tejo,” 2019. [Online]. Available: http://www.ccdr-lvt.pt/files/acd442d5056c22381724d85aa9ffe3f179b49fe3.pdf. [Acedido em 18 05 2021].

[24] World Health Organization, “Ambient (outdoor) air pollution fact sheet,” 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health. [Acedido em 29 9 2021].

[25] EEA, “Air quality in Europe — 2018 report,” 2018. [Online]. Available: https://www.eea.europa.eu/publications/air-quality-in-europe-2018.

[26] S. Khomenko, M. Cirach, E. Pereira-Barboza, N. Mueller, J. Barrera-Gómez, D. Rojas-Rueda, K. d. Hoogh, G. Hoek e M. Nieuwenhuijsen, “Premature mortality due to air pollution in European cities: a health impact assessment,” *The Lancet Planetary Health*, vol. 5, nº 3, pp. e121 - e134, 2021.

[27] M. Saunders, P. Lewis e A. Thornhill, Research methods for business students, 6th ed., Upper Saddle River, NJ: Prentice Hall, 2012.

[28] CML Lisboa Lx Datalab, “Determinação do impacte da pandemia por COVID 19 na mobilidade e ambiente,” [Online]. Available: https://lisboainteligente.cm-lisboa.pt/lxdatalab/desafios/determinacao-do-impacte-da-pandemia-por-covid-19-na-mobilidade-e-ambiente/. [Acedido em 03 10 2021].

[29] D. E. Stokes, Pasteur's Quadrant: Basic Science and Technological Innovation, Washington, DC: Brookings Institution Press, 1997.

[30] K. Jensen, “Cross Industry Standard Process for Data Mining,” 2012.

[31] B. Finlayson-Pitts e J. P. Jr., “Atmospheric Chemistry of Tropospheric Ozone Formation: Scientific and Regulatory Implications,” *Air & Waste*, vol. 43, nº 8, pp. 1091-1100, 1993.

[32] C. Guerreiro, V. Foltescu e F. de Leeuw, Air Quality in Europe - 2013 Report, 2013.

[33] B. Zhang, L. Jiao, G. Xu, S. Zhao, X. Tang, Y. Zhou e C. Gong, “Influences of wind and precipitation on different-sized particulate matter concentrations (PM 2.5, PM 10, PM 2.5--10),” *Meteorology and Atmospheric Physics*, vol. 130, nº 3, pp. 383--392, 2018.

[34] M. A. Z. Tautan, R. S. Savastru, D. M. Savastru e M. N., “Assessing the relationship between ground levels of ozone (O3) and nitrogen dioxide (NO2) with coronavirus (COVID-19) in Milan, Italy,” *Science of The Total Environment*, vol. 740, p. 140005, 2020.

[35] T. Lecocq, S. P. Hicks, K. Van Noten, K. Van Wijk, P. Koelemeijer, R. S. De Plaen, F. Massin, G. Hillers, R. E. Anthony e M.-T. Apoloner, “Global quieting of high-frequency seismic noise due to COVID-19 pandemic lockdown measures,” *Science*, vol. 369, nº 6509, pp. 1338--1343, 2020.

[36] D. Bonet-Solà, C. Martínez-Suquía, R. Alsina-Pagès e P. Bergada, “The Soundscape of the COVID-19 Lockdown: Barcelona Noise Monitoring Network Case Study,” *International Journal of Environmental Research and Public Health*, vol. 18, 2021.

[37] G. G. Gioli, L. Brilli, F. Carotenuto, C. Vagnoli, A. Zaldei e B. Gioli, “Quantifying road traffic impact on air quality in urban areas: A Covid19-induced lockdown analysis in Italy,” *Environmental Pollution*, vol. 267, p. 115682, 2020.

[38] R. H. Rana, S. A. Keramat e J. Gow, “A Systematic Literature Review of the Impact of {COVID}-19 Lockdowns on Air Quality in China,” *Aerosol and Air Quality Research*, vol. 21, 2021.

[39] M. Bauwens, S. Compernolle, T. Stavrakou, J.-F. Müller, J. van Gent, H. Eskes, P. F. Levelt, R. van der A, J. P. Veefkind, J. Vlietinck, H. Yu e C. Zehner, “Impact of Coronavirus Outbreak on NO2 Pollution Assessed Using TROPOMI

and OMI Observations," *Geophysical Research Letters,* vol. 47, nº 11, p. e2020GL087978, 2020.

[40] P. Connerton, J. Vicente de Assunção, R. Maura de Miranda, A. Dorothée Slovic, P. José Pérez-Martínez e H. Ribeiro, "Air Quality during COVID-19 in Four Megacities: Lessons and Challenges for Public Health," *International Journal of Environmental Research and Public Health,* vol. 17, nº 14, 2020.

[41] P. S. Calatayud, A. {. Marco}, E. Agathokleous, Z. Feng, X. Xu, E. Paoletti, J. J. D. Rodriguez e V. Calatayudf, "Amplified ozone pollution in cities during the COVID-19 lockdown," *Science of The Total Environment,* vol. 735, p. 139542, 2020.

[42] EPA, "Photochemical Air Quality Modeling," EPA, [Online]. Available: https://www.epa.gov/scram/photochemical-air-quality-modeling. [Acedido em 09 07 2021].

[43] Y. Rybarczyk e R. Zalakeviciute, "Machine Learning Approaches for Outdoor Air Quality Modelling: A Systematic Review," *Applied Sciences,* vol. 8, p. 2570, 2018.

[44] T. V. Vu, Z. Shi, J. Cheng, Q. Zhang, K. He, S. Wang e R. M. Harrison, "Assessing the impact of clean air action on air quality trends in Beijing using a machine learning technique," *Atmospheric Chemistry and Physics,* vol. 19, nº 17, pp. 11303--11314, 2019.

[45] M. Castelli, F. Clemente, A. Popovič, S. Silva e L. Vanneschi, "A Machine Learning Approach to Predict Air Quality in California," *Complexity,* vol. 2020, pp. 1-23, 2020.

[46] A. Luna, M. Paredes, G. Oliveira e S. Correa, "Prediction of ozone concentration in tropospheric levels using artificial neural networks and support vector machine at Rio de Janeiro, Brazil," *Atmospheric Environment,* vol. 98, 2014.

[47] European Monitoring and Evaluation Programme, "Convention on Long-range Transboundary Air Pollution," [Online]. Available: https://www.emep.int/. [Acedido em 03 10 2021].

[48] U.S. Climate Change Science Program, "Strategic Plan for the U.S. Climate Change Science Program," U.S. Climate Change Science Program, Washington, D.C, 2003.

[49] CCDR LVT, "Regulamentação - Qualidade do Ar," 2016. [Online]. Available: http://www.ccdr-lvt.pt/pt/regulamentacao---qualidade-ar/8083.htm.

[50] Google, "Google Community Mobility Report for Lisbon," Google, [Online]. Available: https://www.google.com/covid19/mobility/. [Acedido em 30 09 2021].

[51] Apple, "Mobility Trend Report for Lisbon," Apple, [Online]. Available: https://covid19.apple.com/mobility. [Acedido em 30 09 2021].

[52] Moovit, "Moovit COVID-19 public transit data," Moovit , [Online]. Available: https://moovitapp.com/insights/en/Moovit_Insights_Public_Transit_Index-countries. [Acedido em 30 09 2021].

[53] Waze, "Waze COVID-19 Impact Dashboard," Waze, [Online]. Available: https://www.waze.com/en-GB/covid19. [Acedido em 30 09 2021].

[54] DSSG-PT, "Dados relativos à pandemia COVID-19 em Portugal," [Online]. Available: https://github.com/dssg-pt/covid19pt-data. [Acedido em 2021 10 07].

[55] M. d. Lisboa, “A partir de abril, o seu passe custa menos,” Metropolitano de Lisboa, [Online]. Available: https://www.metrolisboa.pt/2019/03/18/novos-passes-aml/. [Acedido em 30 09 2021].

[56] SEF/GEPF, “Relatório de Imigração, Fronteiras e Asilo 2019,” SEF, Oeiras, 2020.

[57] Governo de Portugal, “Regras de tráfego aéreo de e para Portugal a partir de 1 de agosto,” 2020. [Online]. Available: https://www.portugal.gov.pt/pt/gc22/comunicacao/comunicado?i=regras-de-trafego-aereo-de-e-para-portugal-a-partir-de-1-de-agosto.

[58] World Health Organization, “Health effects of particulate matter,” World Health Organization, 2013.

[59] TPOT, “TPOT Home Page - Data Science Assistant,” 23 11 2020. [Online]. Available: https://github.com/epistasislab/tpot/. [Acedido em 29 09 2021].

[60] University of Pennsylvania, “TPOT - Default Regression configuration,” [Online]. Available: https://github.com/EpistasisLab/tpot/blob/master/tpot/config/regressor.py.

[61] Distributed (Deep) Machine Learning Community, “XGBoost Regressor,” [Online]. Available: https://xgboost.readthedocs.io/en/latest/python/python_api.html#module-xgboost.sklearn.

[62] Scikit-learn Project, “Cross-validated Lasso, using the LARS algorithm.,” [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LassoLarsCV.html.

[63] Distributed (Deep) Machine Learning Community, “XGBoost - Tree model Importance types,” [Online]. Available: https://xgboost.readthedocs.io/en/latest/python/python_api.html#xgboost.Booster.get_score.

[64] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M, Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal e S.-I. Lee, “From local explanations to global understanding with explainable AI for trees,” *Nature Machine Intelligence,* vol. 2, nº 1, pp. 56-67, 2020.

[65] S. Lundberg, “SHapley Additive exPlanations - Github,” [Online]. Available: https://github.com/slundberg/shap. [Acedido em 29 09 2021].

[66] O. E. Clifton, A. M. Fiore, W. J. Massman, C. B. Baublitz, M. Coyle, L. Emberson, S. Fares, D. K. Farmer, P. Gentine, G. Gerosa, A. B. Guenther, D. Helmig, D. L. Lombardozzi, J. W. Munger e Edward, “Dry Deposition of Ozone over Land: Processes, Measurement, and Modeling,” *Rev Geophys,* vol. 58, nº 1, 2020.

[67] M. Penkała, P. Ogrodnik e W. Rogula-Kozłowska, “Particulate Matter from the Road Surface Abrasion as a Problem of Non-Exhaust Emission Control,” *Environments,* vol. 5, nº 1, p. 9, 2018.

[68] A. Rahimi, “Short-term prediction of NO2 and NO x concentrations using multilayer perceptron neural network: a case study of Tabriz, Iran,” *Ecological Processes,* vol. 6, nº 1, p. 4, 2017.

[69] S. R. Shams, S. K. Ali Jahani, M. Moeinaddini e N. Khorasani, “Artificial intelligence accuracy assessment in NO2 concentration forecasting of metropolises air,” *Scientific Reports,* vol. 11, nº 1, p. 1805, 2021.

[70] CCDR-LVT, “Inventário de emissões atmosféricas da região de Lisboa e Vale do Tejo (2011-2014),” 2017.

[71] L. A. Oliveira, Dissertação e Tese em Ciência e Tecnologia Segundo Bolonha, Lisboa: Lidel, 2011.

[72] IMT - Instituto da Mobilidade e dos Transportes, “Anuário Estatístico da Mobilidade e dos Transportes 2019,” Nov 2020. [Online]. Available: http://www.imt-ip.pt/sites/IMTT/Portugues/IMTT/relatoriosectoriais/Documents/Anu%C3%A1rioEstat%C3%ADsticoMobilidadeTransportes_2019.pdf. [Acedido em 2021 05 14].

[73] M. Bauwens, S. Compernolle, T. Stavrakou, J.-F. Müller, J. van Gent, H. Eskes, P. F. Levelt, R. van der A, J. P. Veefkind, J. Vlietinck, H. Yu e C. Zehner, “Impact of Coronavirus Outbreak on NO2 Pollution Assessed Using TROPOMI and OMI Observations,” *Geophysical Research Letters,* vol. 47, nº 11, p. e2020GL087978, 2020.

[74] Governo de Portugal, “Governo prorrogou várias medidas restritivas do tráfego aéreo com destino e a partir de Portugal,” 2020. [Online]. Available: https://www.portugal.gov.pt/pt/gc22/comunicacao/comunicado?i=governo-prorrogou-varias-medidas-restritivas-do-trafego-aereo-com-destino-e-a-partir-de-portugal.

[75] T. Brown, H. LeMay Jr. e B. Bursten, Chemistry: The central science (8th ed.), New Jersey: Prentice Hall, 2000.

[76] CCDR-LVT, “Plano de melhoria da qualidade do ar da região de Lisboa e Vale do Tejo para os poluentes partículas PM10 e NO2 nas aglomerações da área metropolitana de Lisboa Norte e área metropolitana de Lisboa Sul,” 2017.

Annexes and Appendices

Appendix A – Data tables related to figures

Table 14 – All Lisbon stations monthly median NO2 concentration (µg/m3) for the years 2013-2019 and 2020.

Month	Median 2013-2019 (CI95)	Median 2020 (CI95)	Difference	N (2013-2019)	N (2020)
Jan	35,5±0,35 (µg/m3)	34±0,85 (µg/m3)	-4,23%	28326	3706
Feb	28,1±0,34 (µg/m3)	40,7±0,87 (µg/m3)	44,84%	26153	3992
Mar	24,8±0,31 (µg/m3)	18,2±0,57 (µg/m3)	-26,61%	27810	4102
Apr	23±0,28 (µg/m3)	12,2±0,38 (µg/m3)	-46,96%	27673	4236
May	21,1±0,3 (µg/m3)	14±0,39 (µg/m3)	-33,65%	27048	4155
Jun	18,1±0,28 (µg/m3)	11,7±0,34 (µg/m3)	-35,36%	26467	4305
Jul	16,3±0,27 (µg/m3)	17±0,59 (µg/m3)	4,29%	26925	4440
Aug	16,6±0,31 (µg/m3)	11,2±0,39 (µg/m3)	-32,53%	26684	4097
Sep	25,9±0,34 (µg/m3)	21,4±0,76 (µg/m3)	-17,37%	26441	4278
Oct	31,9±0,4 (µg/m3)	22,4±0,64 (µg/m3)	-29,78%	27844	4417
Nov	35±0,36 (µg/m3)	27,7±0,65 (µg/m3)	-20,86%	26905	4316
Dec	38,2±0,35 (µg/m3)	26±0,65 (µg/m3)	-31,94%	28014	4323

Data table related to Figure 22

Table 15 –Road traffic Lisbon monitoring stations monthly median NO2 concentration (µg/m3) for the years 2013-2019 and 2020.

Month	Median 2013-2019 (CI95)	Median 2020 (CI95)	Difference	N (2013-2019)	N (2020)
Jan	47±0,52 (µg/m3)	42,8±1,13 (µg/m3)	-8,94%	14493	2191
Feb	40±0,5 (µg/m3)	51,4±1,23 (µg/m3)	28,50%	13202	2086
Mar	36±0,48 (µg/m3)	22,4±0,87 (µg/m3)	-37,78%	14189	2231
Apr	34,1±0,43 (µg/m3)	14,9±0,61 (µg/m3)	-56,30%	14117	2080
May	31,6±0,46 (µg/m3)	18,45±0,66 (µg/m3)	-41,61%	14447	1926
Jun	27±0,46 (µg/m3)	15,4±0,54 (µg/m3)	-42,96%	13514	2148
Jul	24,8±0,47 (µg/m3)	24,3±0,98 (µg/m3)	-2,02%	13313	2227
Aug	25±0,51 (µg/m3)	15±0,62 (µg/m3)	-40,00%	13949	2225
Sep	37,7±0,54 (µg/m3)	32,3±1,2 (µg/m3)	-14,32%	13486	2158
Oct	46,3±0,62 (µg/m3)	33±0,94 (µg/m3)	-28,73%	13600	2224
Nov	48,5±0,55 (µg/m3)	35,65±0,91 (µg/m3)	-26,49%	13010	2158
Dec	49±0,53 (µg/m3)	34,5±0,94 (µg/m3)	-29,59%	13932	2162

Data table related to Figure 23

Table 16 – Background Lisbon monitoring stations monthly median NO2 concentration (µg/m3) for the years 2013-2019 and 2020.

Month	Median 2013-2019 (CI95)	Median 2020 (CI95)	Difference	N (2013-2019)	N (2020)
Jan	24,8±0,41 (µg/m3)	21,4±1,06 (µg/m3)	-13,71%	13833	1515
Feb	18,6±0,39 (µg/m3)	27,5±1,03 (µg/m3)	47,85%	12951	1906
Mar	16,4±0,33 (µg/m3)	14,2±0,63 (µg/m3)	-13,41%	13621	1871
Apr	15,2±0,28 (µg/m3)	10,3±0,44 (µg/m3)	-32,24%	13556	2156
May	13,9±0,25 (µg/m3)	11,2±0,39 (µg/m3)	-19,42%	12601	2229
Jun	12,4±0,22 (µg/m3)	8,8±0,34 (µg/m3)	-29,03%	12953	2157
Jul	11,6±0,18 (µg/m3)	13,1±0,51 (µg/m3)	12,93%	13612	2213
Aug	11,8±0,25 (µg/m3)	8,7±0,31 (µg/m3)	-26,27%	12735	1872
Sep	17,8±0,32 (µg/m3)	14,8±0,72 (µg/m3)	-16,85%	12955	2120
Oct	21,3±0,41 (µg/m3)	14,8±0,72 (µg/m3)	-30,52%	14244	2193
Nov	24,3±0,4 (µg/m3)	19,5±0,84 (µg/m3)	-19,75%	13895	2158
Dec	27,8±0,41 (µg/m3)	18,4±0,8 (µg/m3)	-33,81%	14082	2161

Data table related to Figure 24

Table 17 – Hourly NO2 median concentration in Traffic Stations for the National Confinement period (14-03-2020 to 03-05-2020) and homologous 2013-2019 period.

Hour	Median 2013-2019 (CI95)	Median 2020 (CI95)	Difference	N (2013-2019)	N (2020)
0	36,8±2,02 (µg/m3)	15,1±2,84 (µg/m3)	-58,97%	959	143
1	29,1±1,96 (µg/m3)	12,25±2,47 (µg/m3)	-57,90%	957	142
2	23±1,77 (µg/m3)	10,2±1,94 (µg/m3)	-55,65%	957	140
3	18,8±1,57 (µg/m3)	7,85±1,56 (µg/m3)	-58,24%	952	144
4	17,8±1,42 (µg/m3)	8,8±1,19 (µg/m3)	-50,56%	953	142
5	22,2±1,31 (µg/m3)	12±1,54 (µg/m3)	-45,95%	958	142
6	33,2±1,41 (µg/m3)	16,6±2,11 (µg/m3)	-50,00%	959	145
7	48,4±1,56 (µg/m3)	26,2±2,67 (µg/m3)	-45,87%	960	145
8	57±1,66 (µg/m3)	29,4±2,83 (µg/m3)	-48,42%	962	145
9	49,75±1,54 (µg/m3)	23,05±2,39 (µg/m3)	-53,67%	960	144
10	38,1±1,33 (µg/m3)	19,2±2,05 (µg/m3)	-49,61%	953	142
11	33,5±1,2 (µg/m3)	17,35±1,9 (µg/m3)	-48,21%	954	142
12	31,35±1,14 (µg/m3)	16,2±1,42 (µg/m3)	-48,33%	956	143
13	29,4±1,05 (µg/m3)	14,3±1,32 (µg/m3)	-51,36%	957	143
14	27,25±1,05 (µg/m3)	13,3±1,18 (µg/m3)	-51,19%	956	144
15	27,55±1,2 (µg/m3)	13,35±1,24 (µg/m3)	-51,54%	952	144
16	28,6±1,34 (µg/m3)	13,35±1,14 (µg/m3)	-53,32%	950	144
17	31,8±1,61 (µg/m3)	13,85±1,39 (µg/m3)	-56,45%	959	144
18	38,2±1,68 (µg/m3)	15,2±1,52 (µg/m3)	-60,21%	961	145
19	45,2±1,73 (µg/m3)	17,6±1,71 (µg/m3)	-61,06%	961	143
20	48,95±1,8 (µg/m3)	18,7±2,14 (µg/m3)	-61,80%	960	143
21	45,9±1,9 (µg/m3)	18,5±2,81 (µg/m3)	-59,69%	959	143
22	41,55±1,85 (µg/m3)	17±2,79 (µg/m3)	-59,09%	960	143
23	40,4±1,94 (µg/m3)	16,65±2,8 (µg/m3)	-58,79%	959	144

Data related to Figure 26

Table 18 – Hourly NO2 median concentration in Background Stations for the National Confinement period (14-03-2020 to 03-05-2020) and homologous 2013-2019 period.

Hour	Median 2013-2019 (CI95)	Median 2020 (CI95)	Difference	N (2013-2019)	N (2020)
0	18,8±1,39 (µg/m3)	14,25±2,3 (µg/m3)	-24,20%	913	134
1	15,3±1,36 (µg/m3)	14,15±2,04 (µg/m3)	-7,52%	916	134
2	11,75±1,24 (µg/m3)	11,9±1,69 (µg/m3)	1,28%	914	133
3	10,2±1,11 (µg/m3)	11,55±1,52 (µg/m3)	13,24%	913	134
4	9,2±1,01 (µg/m3)	11,1±1,32 (µg/m3)	20,65%	914	134
5	10,9±1 (µg/m3)	11,65±1,45 (µg/m3)	6,88%	913	134
6	16±1,18 (µg/m3)	15,65±1,9 (µg/m3)	-2,19%	915	134
7	26,9±1,34 (µg/m3)	20,4±2,16 (µg/m3)	-24,16%	914	134
8	31,2±1,31 (µg/m3)	20,95±2,02 (µg/m3)	-32,85%	914	134
9	23,6±1,08 (µg/m3)	15,5±1,7 (µg/m3)	-34,32%	913	133
10	16,4±0,85 (µg/m3)	12,5±1,52 (µg/m3)	-23,78%	906	133
11	13,5±0,67 (µg/m3)	9,55±1,39 (µg/m3)	-29,26%	906	134
12	11,3±0,55 (µg/m3)	7,9±1,2 (µg/m3)	-30,09%	908	135
13	10,5±0,49 (µg/m3)	7,3±1,05 (µg/m3)	-30,48%	910	135
14	9,9±0,48 (µg/m3)	6,4±0,87 (µg/m3)	-35,35%	909	135
15	10,2±0,46 (µg/m3)	7±0,62 (µg/m3)	-31,37%	907	135
16	11,1±0,46 (µg/m3)	7,2±0,78 (µg/m3)	-35,14%	909	135
17	12,9±0,56 (µg/m3)	7,8±0,84 (µg/m3)	-39,53%	914	135
18	15,9±0,66 (µg/m3)	8,7±0,86 (µg/m3)	-45,28%	912	135
19	20,2±0,83 (µg/m3)	10,6±1,07 (µg/m3)	-47,52%	913	135
20	24±1,14 (µg/m3)	13,9±1,61 (µg/m3)	-42,08%	913	135
21	22,8±1,27 (µg/m3)	13±1,87 (µg/m3)	-42,98%	913	135
22	21,9±1,26 (µg/m3)	13,9±2,21 (µg/m3)	-36,53%	913	135
23	20,9±1,34 (µg/m3)	14,55±2,36 (µg/m3)	-30,38%	913	134

Data related to Figure 27

Table 19 – NO2 median concentration in Traffic Stations per day of the week during the National Emergency period and Calamity State (14-03-2020 to 01-07-2020) and homologous 2013-2019 period.

Week Day	Median 2013-2019 (CI95)	Median 2020 (CI95)	Difference	N (2013-2019)	N (2020)
Mon	34,1±0,63 (µg/m3)	17±0,83 (µg/m3)	-50,15%	7308	1097
Tue	33,2±0,62 (µg/m3)	18,6±0,84 (µg/m3)	-43,98%	7271	1110
Wed	33,5±0,66 (µg/m3)	18,75±0,79 (µg/m3)	-44,03%	7257	1124
Thu	33,6±0,64 (µg/m3)	17,4±0,85 (µg/m3)	-48,21%	7247	1036
Fri	34,8±0,66 (µg/m3)	17,8±0,97 (µg/m3)	-48,85%	7237	1002
Sat	30,6±0,61 (µg/m3)	15,9±0,8 (µg/m3)	-48,04%	7294	1078
Sun	24,4±0,55 (µg/m3)	10,7±0,57 (µg/m3)	-56,15%	7329	1074

Data related to Figure 28

Table 20 – NO2 median concentration in Background Stations per day of the week during the National Emergency period and Calamity State (14-03-2020 to 01-07-2020) and homologous 2013-2019 period.

Week Day	Median 2013-2019 (CI95)	Median 2020 (CI95)	Difference	N (2013-2019)	N (2020)
Mon	15,5±0,37 (µg/m3)	11,2±0,55 (µg/m3)	-27,74%	6863	1098
Tue	14,9±0,35 (µg/m3)	12,1±0,58 (µg/m3)	-18,79%	6763	1103
Wed	14,8±0,39 (µg/m3)	11,85±0,55 (µg/m3)	-19,93%	6678	1104
Thu	14,8±0,37 (µg/m3)	11,2±0,54 (µg/m3)	-24,32%	6694	1029
Fri	15,2±0,38 (µg/m3)	10,2±0,61 (µg/m3)	-32,89%	6772	1045
Sat	12,7±0,38 (µg/m3)	9,1±0,63 (µg/m3)	-28,35%	6832	1103
Sun	11±0,33 (µg/m3)	6,9±0,46 (µg/m3)	-37,27%	6895	1104

Data related to Figure 29

Table 21 – NO2 annual average NO2 concentration in all Lisbon stations from 2013 to 2020.

Year	Av. Liberdade	Beato	Benfica	Entrecampos	Olivais	Restelo
2013	52.77	No Data	No Data	38.80	29.39	22.34
2014	53.24	19.75	34.27	36.95	25.98	21.53
2015	58.63	20.47	34.42	38.81	29.23	21.39
2016	57.34	19.70	35.20	36.95	27.82	18.74
2017	60.40	26.74	38.18	40.79	30.46	22.06
2018	60.69	24.54	37.66	40.54	30.43	21.22
2019	54.61	21.08	34.29	35.79	27.24	19.82
2020	39.46	17.41	28.65	29.42	20.86	19.68

Data related to Figure 30

Table 22 –Road traffic Lisbon monitoring stations monthly median O3 concentration (µg/m3) for the years 2013-2019 and 2020.

Month	Median 2013-2019 (CI95)	Median 2020 (CI95)	Difference	N (2013-2019)	N (2020)
Jan	33,6±0,72 (µg/m3)	28±1,87 (µg/m3)	-16,67%	4171	560
Feb	43,9±0,8 (µg/m3)	22±1,76 (µg/m3)	-49,89%	3232	684
Mar	56,6±0,76 (µg/m3)	67±1,47 (µg/m3)	18,37%	3558	744
Apr	65±0,76 (µg/m3)	73±1,52 (µg/m3)	12,31%	3452	717
May	64,7±0,77 (µg/m3)	74±1,54 (µg/m3)	14,37%	3597	744
Jun	55,6±0,81 (µg/m3)	58±1,39 (µg/m3)	4,32%	3129	719
Jul	54,25±0,79 (µg/m3)	65±1,98 (µg/m3)	19,82%	3722	720
Aug	57,6±0,79 (µg/m3)	60±1,61 (µg/m3)	4,17%	4308	495
Sep	49,6±0,75 (µg/m3)	60±2,02 (µg/m3)	20,97%	4015	719
Oct	37,7±0,72 (µg/m3)	50±1,76 (µg/m3)	32,63%	4119	744

Nov	30,4±0,72 (µg/m3)	30±1,65 (µg/m3)	-1,32%	3080	720
Dec	22,4±0,69 (µg/m3)	39±1,81 (µg/m3)	74,11%	3466	703

Data related to Figure 32

Table 23 –Background Lisbon monitoring stations monthly median O3 concentration (µg/m3) for the years 2013-2019 and 2020.

Month	Median 2013-2019 (CI95)	Median 2020 (CI95)	Difference	N (2013-2019)	N (2020)
Jan	46±0,41 (µg/m3)	41±1,17 (µg/m3)	-10,87%	13477	1395
Feb	58±0,38 (µg/m3)	37±1,07 (µg/m3)	-36,21%	13367	2037
Mar	66,8±0,37 (µg/m3)	66±0,79 (µg/m3)	-1,20%	13650	2229
Apr	74±0,38 (µg/m3)	69±0,84 (µg/m3)	-6,76%	14080	2158
May	74±0,36 (µg/m3)	70±0,92 (µg/m3)	-5,41%	14633	2230
Jun	64±0,38 (µg/m3)	59±0,83 (µg/m3)	-7,81%	14882	2086
Jul	60,4±0,4 (µg/m3)	61±1,26 (µg/m3)	0,99%	15056	2088
Aug	60,1±0,44 (µg/m3)	53±0,88 (µg/m3)	-11,81%	15076	2232
Sep	58,4±0,42 (µg/m3)	61±1,18 (µg/m3)	4,45%	14823	2154
Oct	48±0,38 (µg/m3)	52±0,93 (µg/m3)	8,33%	14789	2231
Nov	42,9±0,35 (µg/m3)	35±0,9 (µg/m3)	-18,41%	14489	2147
Dec	31±0,37 (µg/m3)	45±0,91 (µg/m3)	45,16%	13618	2096

Data related to Figure 33

Table 24 – Hourly O3 median concentration in Traffic Stations for the National Confinement period (14-03-2020 to 03-05-2020) versus 2013-2019 Interquartile range, for the same year period.

Hour	Median 2013-2019 (CI95)	Median 2020 (CI95)	Difference	N (2013-2019)	N (2020)
0	59,2±2,99 (µg/m3)	71±6,23 (µg/m3)	19,93%	237	49
1	63,5±3,1 (µg/m3)	70±6,01 (µg/m3)	10,24%	237	49
2	67±3,08 (µg/m3)	69±5,81 (µg/m3)	2,99%	237	49
3	67,5±2,96 (µg/m3)	69±5,9 (µg/m3)	2,22%	237	49
4	64,7±2,89 (µg/m3)	66±5,31 (µg/m3)	2,01%	237	49
5	60,7±2,98 (µg/m3)	66±5,65 (µg/m3)	8,73%	237	49
6	49,8±3,16 (µg/m3)	61±6,1 (µg/m3)	22,49%	238	49
7	35±2,96 (µg/m3)	53±6,01 (µg/m3)	51,43%	237	49
8	33,8±2,63 (µg/m3)	49,5±5,75 (µg/m3)	46,45%	238	48
9	43,1±2,38 (µg/m3)	60±4,8 (µg/m3)	39,21%	238	48
10	55,8±2,15 (µg/m3)	67±4,81 (µg/m3)	20,07%	236	49
11	62,4±1,94 (µg/m3)	70±4,56 (µg/m3)	12,18%	237	49
12	69,3±1,94 (µg/m3)	74±4,3 (µg/m3)	6,78%	237	49
13	72,95±1,95 (µg/m3)	79±4,13 (µg/m3)	8,29%	238	49
14	75,1±1,95 (µg/m3)	82±3,81 (µg/m3)	9,19%	238	49
15	76,1±1,98 (µg/m3)	82±3,8 (µg/m3)	7,75%	237	49
16	75,7±2,08 (µg/m3)	83,5±3,35 (µg/m3)	10,30%	235	48
17	73,2±2,25 (µg/m3)	83±3,67 (µg/m3)	13,39%	236	49
18	68,6±2,49 (µg/m3)	79±3,89 (µg/m3)	15,16%	237	49
19	63,8±2,57 (µg/m3)	75±4,26 (µg/m3)	17,55%	238	49

20	59,9±2,81 (µg/m3)	74±4,93 (µg/m3)	23,54%	239	49
21	58,5±2,95 (µg/m3)	73±5,71 (µg/m3)	24,79%	239	49
22	60±2,91 (µg/m3)	70±5,84 (µg/m3)	16,67%	239	49
23	60±2,91 (µg/m3)	71±6,03 (µg/m3)	18,33%	237	49

Data related to Figure 34

Table 25 – Hourly O3 median concentration in Background Stations for the National Confinement period (14-03-2020 to 03-05-2020) versus 2013-2019 Interquartile range, for the same year period.

Hour	Median 2013-2019 (CI95)	Median 2020 (CI95)	Difference	N (2013-2019)	N (2020)
0	66,55±1,45 (µg/m3)	65±3,24 (µg/m3)	-2,33%	946	147
1	68,3±1,45 (µg/m3)	63±3,14 (µg/m3)	-7,76%	953	147
2	70,8±1,44 (µg/m3)	64±3,07 (µg/m3)	-9,60%	953	147
3	70±1,4 (µg/m3)	62±2,95 (µg/m3)	-11,43%	951	147
4	67,2±1,37 (µg/m3)	62±2,73 (µg/m3)	-7,74%	950	147
5	64,3±1,36 (µg/m3)	59±2,86 (µg/m3)	-8,24%	950	147
6	59,1±1,46 (µg/m3)	54±3,22 (µg/m3)	-8,63%	949	147
7	51±1,5 (µg/m3)	53±3,21 (µg/m3)	3,92%	950	147
8	49±1,35 (µg/m3)	52±2,73 (µg/m3)	6,12%	950	147
9	58±1,21 (µg/m3)	57±2,75 (µg/m3)	-1,72%	947	146
10	67±1,13 (µg/m3)	62±2,67 (µg/m3)	-7,46%	941	145
11	73,75±1,04 (µg/m3)	67±2,57 (µg/m3)	-9,15%	946	146
12	79,9±0,96 (µg/m3)	73±2,53 (µg/m3)	-8,64%	951	147
13	84,2±0,95 (µg/m3)	77±2,35 (µg/m3)	-8,55%	951	147
14	86,8±0,94 (µg/m3)	80±2,32 (µg/m3)	-7,83%	946	147
15	88±0,95 (µg/m3)	81±2,18 (µg/m3)	-7,95%	946	147
16	88±0,97 (µg/m3)	81±2,13 (µg/m3)	-7,95%	953	147
17	86,4±0,98 (µg/m3)	80±2,22 (µg/m3)	-7,41%	953	147
18	83±0,99 (µg/m3)	79±2,38 (µg/m3)	-4,82%	954	147
19	77±1,04 (µg/m3)	78±2,59 (µg/m3)	1,30%	953	147
20	71±1,21 (µg/m3)	72±2,74 (µg/m3)	1,41%	953	147
21	70,45±1,33 (µg/m3)	70±2,99 (µg/m3)	-0,64%	952	147
22	69,2±1,36 (µg/m3)	68±3,24 (µg/m3)	-1,73%	953	147
23	68±1,42 (µg/m3)	67±3,38 (µg/m3)	-1,47%	952	147

Data related to Figure 35

Table 26 – O3 median concentration in Traffic Stations per day of the week during the National Emergency period and Calamity State (14-03-2020 to 01-07-2020) versus 2013-2019 Interquartile range, for the same year period.

Week Day	Median 2013-2019 (CI95)	Median 2020 (CI95)	Difference	N (2013-2019)	N (2020)
Mon	59,8±1,11 (µg/m3)	72±1,9 (µg/m3)	20,40%	1766	384
Tue	59,6±1,11 (µg/m3)	65±1,97 (µg/m3)	9,06%	1766	383
Wed	60±1,11 (µg/m3)	68±2,08 (µg/m3)	13,33%	1728	384
Thu	59,25±1,02 (µg/m3)	71±2 (µg/m3)	19,83%	1756	358
Fri	58,4±1,04 (µg/m3)	69±2,41 (µg/m3)	18,15%	1785	359
Sat	63±1,11 (µg/m3)	68±2,41 (µg/m3)	7,94%	1788	384
Sun	65,4±1,12 (µg/m3)	72,5±2,17 (µg/m3)	10,86%	1782	384

Data related to Figure 36

Table 27 – O3 median concentration in Background Stations per day of the week during the National Emergency period and Calamity State (14-03-2020 to 01-07-2020) versus 2013-2019 Interquartile range, for the same year period.

Week Day	Median 2013-2019 (CI95)	Median 2020 (CI95)	Difference	N (2013-2019)	N (2020)
Mon	70,1±0,54 (µg/m3)	69±1,09 (µg/m3)	-1,57%	7475	1123
Tue	70±0,52 (µg/m3)	65±1,16 (µg/m3)	-7,14%	7379	1141
Wed	70,1±0,52 (µg/m3)	66±1,17 (µg/m3)	-5,85%	7338	1152
Thu	69,7±0,51 (µg/m3)	67±1,09 (µg/m3)	-3,87%	7387	1079
Fri	68,8±0,5 (µg/m3)	65±1,38 (µg/m3)	-5,52%	7453	1078
Sat	71,1±0,54 (µg/m3)	63±1,37 (µg/m3)	-11,39%	7511	1139
Sun	72±0,52 (µg/m3)	68±1,22 (µg/m3)	-5,56%	7528	1128

Data related to Figure 37

Table 28 – Traffic Lisbon monitoring stations monthly median CO concentration (µg/m3) for the years 2013-2019 and 2020.

Month	Median 2013-2019 (CI95)	Median 2020 (CI95)	Difference	N (2013-2019)	N (2020)
Jan	345,1±6,12 (µg/m3)	321±13,03 (µg/m3)	-6,98%	11530	2041
Feb	295±4,39 (µg/m3)	348±10,04 (µg/m3)	17,97%	10326	2032
Mar	269,7±2,81 (µg/m3)	232±3,64 (µg/m3)	-13,98%	12260	2085
Apr	250±1,98 (µg/m3)	193±2,74 (µg/m3)	-22,80%	12559	2157
May	226,4±1,71 (µg/m3)	214±2,35 (µg/m3)	-5,48%	13188	2120
Jun	203±1,83 (µg/m3)	180±2,25 (µg/m3)	-11,33%	11680	2117
Jul	208±1,58 (µg/m3)	207±3,42 (µg/m3)	-0,48%	13105	2212
Aug	224,15±2,02 (µg/m3)	183±2,34 (µg/m3)	-18,36%	12766	2227
Sep	267,3±2,31 (µg/m3)	238±4,75 (µg/m3)	-10,96%	12231	2159
Oct	298±3,76 (µg/m3)	261±6,06 (µg/m3)	-12,42%	12941	2228
Nov	316,1±4,96 (µg/m3)	297±10,11 (µg/m3)	-6,04%	12408	2143
Dec	379±6,84 (µg/m3)	290±10,29 (µg/m3)	-23,48%	12803	2160

Table 29 – Background Lisbon monitoring stations monthly median CO concentration (µg/m3) for the years 2013-2019 and 2020.

Month	Median 2013-2019 (CI95)	Median 2020 (CI95)	Difference	N (2013-2019)	N (2020)
Jan	300±9,34 (µg/m3)	218±12,27 (µg/m3)	-27,33%	4402	31
Feb	255±6,5 (µg/m3)	287±14,79 (µg/m3)	12,55%	3978	695
Mar	224±3,38 (µg/m3)	232±4,55 (µg/m3)	3,57%	4128	739
Apr	215±2,82 (µg/m3)	206±4 (µg/m3)	-4,19%	4246	720
May	200±2,05 (µg/m3)	209±2,27 (µg/m3)	4,50%	3899	741

Jun	173±1,68 (µg/m3)	168±2,16 (µg/m3)	-2,89%	3985	720
Jul	168±1,93 (µg/m3)	189±3,27 (µg/m3)	12,50%	3714	743
Aug	182±2,85 (µg/m3)	174±2,4 (µg/m3)	-4,40%	3977	742
Sep	200±2,58 (µg/m3)	218±6,03 (µg/m3)	9,00%	4282	720
Oct	202±4,57 (µg/m3)	273±9,36 (µg/m3)	35,15%	4434	628
Nov	234±7,4 (µg/m3)	294,5±16,82 (µg/m3)	25,85%	4299	720
Dec	300±11,66 (µg/m3)	278±16,83 (µg/m3)	-7,33%	3983	712

Table 30 – Traffic Lisbon monitoring stations monthly median PM2.5 concentration (µg/m3) for the years 2013-2019 and 2020.

Month	**Median 2013-2019 (CI95)**	**Median 2020 (CI95)**	**Difference**	**N (2013-2019)**	**N (2020)**
Jan	11,1±0,35 (µg/m3)	11,9±0,82 (µg/m3)	7,21%	4193	735
Feb	10,8±0,31 (µg/m3)	13,7±0,54 (µg/m3)	26,85%	3487	689
Mar	10,3±0,21 (µg/m3)	10,2±0,36 (µg/m3)	-0,97%	4240	740
Apr	10,5±0,19 (µg/m3)	6,6±0,28 (µg/m3)	-37,14%	4232	714
May	10,3±0,19 (µg/m3)	8,3±0,35 (µg/m3)	-19,42%	4275	735
Jun	10±0,21 (µg/m3)	5,8±0,32 (µg/m3)	-42,00%	3974	708
Jul	10,8±0,21 (µg/m3)	11,5±0,42 (µg/m3)	6,48%	4038	736
Aug	8,75±0,26 (µg/m3)	6,3±0,26 (µg/m3)	-28,00%	3880	734
Sep	11,4±0,23 (µg/m3)	9,75±0,42 (µg/m3)	-14,47%	3665	718
Oct	11,6±0,27 (µg/m3)	7,2±0,31 (µg/m3)	-37,93%	4067	722
Nov	11±0,29 (µg/m3)	8,3±0,54 (µg/m3)	-24,55%	3642	634
Dec	14,9±0,36 (µg/m3)	7±0,41 (µg/m3)	-53,02%	4256	688

Table 31 – Background Lisbon monitoring stations monthly median PM2.5 concentration (µg/m3) for the years 2013-2019 and 2020.

Month	**Median 2013-2019 (CI95)**	**Median 2020 (CI95)**	**Difference**	**N (2013-2019)**	**N (2020)**
Jan	9±0,32 (µg/m3)	NA	NA	4345	NA
Feb	8±0,26 (µg/m3)	16,3±0,55 (µg/m3)	103,75%	3959	619
Mar	7,5±0,19 (µg/m3)	10,2±0,36 (µg/m3)	36,00%	4875	726
Apr	8,1±0,18 (µg/m3)	6,1±0,27 (µg/m3)	-24,69%	4714	670
May	7,9±0,16 (µg/m3)	9±0,44 (µg/m3)	13,92%	4678	740
Jun	7,3±0,18 (µg/m3)	6,4±0,26 (µg/m3)	-12,33%	4679	680
Jul	8,3±0,17 (µg/m3)	12,5±0,42 (µg/m3)	50,60%	4892	702
Aug	8,4±0,25 (µg/m3)	5,5±0,18 (µg/m3)	-34,52%	4771	740
Sep	9,2±0,2 (µg/m3)	7,4±0,35 (µg/m3)	-19,57%	4664	712
Oct	9,4±0,25 (µg/m3)	6,2±0,21 (µg/m3)	-34,04%	5063	726
Nov	8,8±0,24 (µg/m3)	9,1±0,47 (µg/m3)	3,41%	4695	713
Dec	11,5±0,32 (µg/m3)	7,5±0,43 (µg/m3)	-34,78%	4728	681

Table 32 – Traffic Lisbon monitoring stations monthly median PM10 concentration (µg/m3) for the years 2013-2019 and 2020.

Month	Median 2013-2019 (CI95)	Median 2020 (CI95)	Difference	N (2013-2019)	N (2020)
Jan	24,55±0,34 (µg/m3)	22,9±0,58 (µg/m3)	-6,72%	12084	2108
Feb	24,3±0,31 (µg/m3)	25,9±0,59 (µg/m3)	6,58%	10762	2083
Mar	22,6±0,26 (µg/m3)	18,8±0,4 (µg/m3)	-16,81%	12156	2227
Apr	22,3±0,24 (µg/m3)	11,6±0,25 (µg/m3)	-47,98%	11573	2152
May	23±0,22 (µg/m3)	15±0,31 (µg/m3)	-34,78%	11575	2163
Jun	21,1±0,24 (µg/m3)	11,5±0,27 (µg/m3)	-45,50%	11271	2133
Jul	23,2±0,22 (µg/m3)	21,5±0,39 (µg/m3)	-7,33%	11071	2221
Aug	22,4±0,29 (µg/m3)	13,6±0,29 (µg/m3)	-39,29%	11557	1886
Sep	25,2±0,24 (µg/m3)	18,5±0,4 (µg/m3)	-26,59%	11574	2138
Oct	26,6±0,34 (µg/m3)	15,7±0,31 (µg/m3)	-40,98%	12227	2191
Nov	25,3±0,32 (µg/m3)	17,2±0,5 (µg/m3)	-32,02%	12182	2144
Dec	29,5±0,37 (µg/m3)	15,3±0,47 (µg/m3)	-48,14%	12835	2054

Table 33 – Background Lisbon monitoring stations monthly median PM10 concentration (µg/m3) for the years 2013-2019 and 2020.

Month	Median 2013-2019 (CI95)	Median 2020 (CI95)	Difference	N (2013-2019)	N (2020)
Jan	18,3±0,42 (µg/m3)	15,9±1,8 (µg/m3)	-13,11%	5179	21
Feb	17,4±0,37 (µg/m3)	24,7±0,77 (µg/m3)	41,95%	5710	685
Mar	17,8±0,3 (µg/m3)	20,3±0,6 (µg/m3)	14,04%	6169	738
Apr	16,7±0,28 (µg/m3)	13,8±0,42 (µg/m3)	-17,37%	5457	682
May	16,9±0,26 (µg/m3)	17±0,47 (µg/m3)	0,59%	5603	744
Jun	16±0,28 (µg/m3)	12,05±0,39 (µg/m3)	-24,69%	5670	718
Jul	16,9±0,26 (µg/m3)	21,5±0,59 (µg/m3)	27,22%	6292	744
Aug	17,5±0,36 (µg/m3)	13,2±0,36 (µg/m3)	-24,57%	5855	744
Sep	20,3±0,29 (µg/m3)	17,15±0,66 (µg/m3)	-15,52%	5744	714
Oct	18,9±0,35 (µg/m3)	14,8±0,45 (µg/m3)	-21,69%	6374	724
Nov	16,5±0,31 (µg/m3)	16,3±0,79 (µg/m3)	-1,21%	6259	713
Dec	20,4±0,41 (µg/m3)	15,5±0,7 (µg/m3)	-24,02%	6205	681

Appendix B – Data Source Metadata

Table 34 – IMPA Meteorology data source metadata

Field	Data Type	Description
ESTACAO	INT	Station Code
ANO	INT	Sample year
MS	INT	Sample month of year
DI	INT	Sample day of month
HR	INT	Sample hour of day

P_E_MD	FLOAT	Hourly average station level air pressure (hPa)
P_M_MD	FLOAT	Hourly average sea level air pressure (hPa)
T_MED	FLOAT	Hourly average of air temperature (Celsius)
HR_MED	FLOAT	Hourly average of air relative humidity (percentage)
DD_MED	FLOAT	Hourly average wind direction (north degree
FF_MED	FLOAT	Hourly average wind speed (m/s)
RG_TOT	FLOAT	Hourly total global radiation (KJ/m2)
PR_QTD	FLOAT	Hourly total precipitation (mm)

Table 35 – APA Air Pollution data source metadata

Field	**Data Type**	**Description**
Countrycode	STRING	Country or territory ISO2 code.
Namespace	STRING	Inspire identifier/namespace of reporting entity, given by data provider.
AirQualityNetwork	STRING	Inspire identifier (Local Id) of air quality measurement network, given by data provider.
AirQualityStation	STRING	Inspire identifier (Local Id) of air quality measurement station, given by data provider.
AirQualityStationEoICode	STRING	EoI code of air quality measurement station (as in AirBase).
SamplingPoint	STRING	Inspire identifier (Local Id) of sampling point, given by data provider.
SamplingProcess	STRING	Inspire identifier (LocalId) of sampling process (procedure), given by data provider.
Sample	STRING	Inspire identifier (LocalId) of sample (featureofinterest), given by data provider.
AirPollutant	STRING	Air polluting substance, level of which is measured and reported to the EEA (see notation in Data Dictionary: http://dd.eionet.europa.eu/vocabulary/aq/pollutant).
AirPollutantCode	STRING	Air polluting substance, level of which is measured and reported to the EEA (see id in Data Dictionary: http://dd.eionet.europa.eu/vocabulary/aq/pollutant).
AveragingTime	STRING	Averaging time/frequency of reported air quality values (see in Data Dictionary: http://dd.eionet.europa.eu/vocabulary/aq/primaryObservation).
Concentration	FLOAT	Measured concentration of air polluting substance.
UnitOfMeasurement	STRING	Unit of concentration of air polluting substance (see in Data Dictionary: http://dd.eionet.europa.eu/vocabulary/uom/concentration).
DatetimeBegin	DATETME	Date-time begin of measurement (UTC+1 for hourly data and original time zone for other averaging times).
DatetimeEnd	DATETIME	Date-time end of measurement (UTC+1 for hourly data and original time zone for other averaging times).
Validity	INT	Information about data validity, given by data provider (see in Data Dictionary: http://dd.eionet.europa.eu/vocabulary/aq/observationvalidity).
Verification	INT	Information whether data have been verified by data provider (see in Data Dictionary: http://dd.eionet.europa.eu/vocabulary/aq/observationverification).

Table 36 – DGS COVID-19 dataset metadata

Field	**Data Type**	**Description**

data	DATE	Date when this data was added to the dataset
data_dados	DATETIME	Date to which the measures refer to
confirmados	INT	Confirmed cases
confirmados_arsnorte	INT	Confirmed cases in ARS Norte
confirmados_arscentro	INT	Confirmed cases in ARS Centro
confirmados_arslvt	INT	Confirmed cases in ARS Lisboa e Vale do Tejo
confirmados_arsalentejo	INT	Confirmed cases in ARS Alentejo
confirmados_arsalgarve	INT	Confirmed cases in ARS Algarve
confirmados_acores	INT	Confirmed cases in Região Autónoma dos Açores
confirmados_madeira	INT	Confirmed cases in Região Autónoma da Madeira
confirmados_estrangeiro	INT	Confirmed cases abroad
confirmados_novos	INT	Difference between the confirmed cases in the day before and the actual day.
recuperados	INT	Total recoveries
obitos	INT	Total deaths
internados	INT	Number of patients admitted to hospital wards
internados_uci	INT	Number of patients admitted to the ICU
lab	INT	Number of suspected cases awaiting laboratory confirmation
suspeitos	INT	Number of suspected cases (definition changed in 29/02/2020) since 01/01/2020
vigilancia	INT	Number of cases under surveillance by health authorities
n_confirmados	INT	Number of suspected cases which did not result in infection
cadeias_transmissao	INT	Number of active contagium chains
transmissao_importada	INT	Number of confirmed cases via transmission of infected foreign nationals
confirmados_0_9_f	INT	Number of confirmed cases for the female gender from 0 to 9 years
confirmados_0_9_m	INT	Number of confirmed cases for the male gender from 0 to 9 years
confirmados_10_19_f	INT	Number of confirmed cases for the female gender from 10 to 19 years
confirmados_10_19_m	INT	Number of confirmed cases for the male gender from 10 to 19 years
confirmados_20_29_f	INT	Number of confirmed cases for the female gender from 20 to 29 years
confirmados_20_29_m	INT	Number of confirmed cases for the male gender from 20 to 29 years
confirmados_30_39_f	INT	Number of confirmed cases for the female gender from 30 to 39 years
confirmados_30_39_m	INT	Number of confirmed cases for the male gender from 30 to 39 years
confirmados_40_49_f	INT	Number of confirmed cases for the female gender from 40 to 49 years
confirmados_40_49_m	INT	Number of confirmed cases for the male gender from 40 to 49 years
confirmados_50_59_f	INT	Number of confirmed cases for the female gender from 50 to 59 years
confirmados_50_59_m	INT	Number of confirmed cases for the male gender from 50 to 59 years
confirmados_60_69_f	INT	Number of confirmed cases for the female gender from 60 to 69 years
confirmados_60_69_m	INT	Number of confirmed cases for the male gender from 60 to 69 years
confirmados_70_79_f	INT	Number of confirmed cases for the female gender from 70 to 79 years
confirmados_70_79_m	INT	Number of confirmed cases for the male gender from 70 to 79 years

confirmados_80_plus_f	INT	Number of confirmed cases for the female gender from 80+ years
confirmados_80_plus_m	INT	Number of confirmed cases for the male gender from 80+ years
sintomas_tosse	FLOAT	Percentage of cases that report symptom: cough. (Sampled)
sintomas_febre	FLOAT	Percentage of cases that report symptom: fever. (Sampled)
sintomas_dificuldade_respiratoria	FLOAT	Percentage of cases that report symptom: difficulty breathing. (Sampled)
sintomas_cefaleia	FLOAT	Percentage of cases that report symptom: headache. (Sampled)
sintomas_dores_musculares	FLOAT	Percentage of cases that report symptom: muscular pain. (Sampled)
sintomas_fraqueza_generalizada	FLOAT	Percentage of cases that report symptom: weakness. (Sampled)
confirmados_f	INT	Total number of females confirmed cases
confirmados_m	INT	Total number of male confirmed cases
obitos_arsnorte	INT	Total deaths in ARS Norte
obitos_arscentro	INT	Total deaths in ARS Centro
obitos_arslvt	INT	Total deaths in ARS Lisboa e Vale do Tejo
obitos_arsalentejo	INT	Total deaths in ARS Alentejo
obitos_arsalgarve	INT	Total deaths in ARS Algarve
obitos_acores	INT	Total deaths in Região Autónoma dos Açores
obitos_madeira	INT	Total deaths in Região Autónoma da Madeira
obitos_estrangeiro	INT	Total deaths abroad
recuperados_arsnorte	INT	Total recoveries in ARS Norte
recuperados_arscentro	INT	Total recoveries in ARS Centro
recuperados_arslvt	INT	Total recovieres in ARS Lisboa e Vale do Tejo
recuperados_arsalentejo	INT	Total recoveries in ARS Alentejo
recuperados_arsalgarve	INT	Total recoveries in ARS Algarve
recuperados_acores	INT	Total recovieres in Região Autónoma dos Açores
recuperados_madeira	INT	Total recovieres in Região Autónoma da Madeira
recuperados_estrangeiro	INT	Total recoveries abroad
obitos_0_9_f	INT	Total deaths female from 0 to 9 years
obitos_0_9_m	INT	Total deaths male from 0 to 9 years
obitos_10_19_f	INT	Total deaths female from 10 to 19 years
obitos_10_19_m	INT	Total deaths male from 10 to 19 years
obitos_20_29_f	INT	Total deaths female from 20 to 29 years
obitos_20_29_m	INT	Total deaths male from 20 to 29 years
obitos_30_39_f	INT	Total deaths female from 30 to 39 years
obitos_30_39_m	INT	Total deaths male from 30 to 39 years
obitos_40_49_f	INT	Total deaths female from 40 to 49 years
obitos_40_49_m	INT	Total deaths male from 40 to 49 years
obitos_50_59_f	INT	Total deaths female from 50 to 59 years
obitos_50_59_m	INT	Total deaths male from 50 to 59 years
obitos_60_69_f	INT	Total deaths female from 60 to 69 years
obitos_60_69_m	INT	Total deaths male from 60 to 69 years
obitos_70_79_f	INT	Total deaths female from 70 to 79 years
obitos_70_79_m	INT	Total deaths male from 70 to 79 years
obitos_80_plus_f	INT	Total deaths female from 80+ years
obitos_80_plus_m	INT	Total deaths male from 80+ years

obitos_f	INT	Total deaths female
obitos_m	INT	Total deaths male
confirmados_desconhecidos_m	INT	Number of confirmed cases for the male gender with unknown age
confirmados_desconhecidos_f	INT	Number of confirmed cases for the female gender with unknown age
ativos	INT	Number of active cases
internados_enfermaria	INT	Number of patients admitted in wards but not in ICU
confirmados_desconhecidos	INT	Number of confirmed cases with unknown gender

Table 37 – Apple Mobility Trend Report data source metadata

Field	Data Type	Description
Date	DATE	Date
Driving Trips Change	FLOAT	Daily percentual change of Apple maps application driving direction requests in the city of Lisbon compared against a pre-pandemic baseline
Walking Trips Change	FLOAT	Daily percentual change of Apple maps application walking direction requests in the city of Lisbon compared against a pre-pandemic baseline

Table 38 –Moovit Insights Public Transit Index data source metadata

Field	Data Type	Description
Date	DATE	Date
Mobility Index Change	FLOAT	7-day public transportation ridership percentual change, measured with the Moovit App, against a pre-pandemic baseline.

Table 39 – Google COVID-19 Community Mobility Report data source metadata

Field	Data Type	Description
Data	DATE	Date
Retalho e lazer	FLOAT	Mobility trends for places like restaurants, cafes, shopping centers, theme parks, museums, libraries, and movie theaters, compared to a pre-pandemic baseline.
Mercearias e farmácias	FLOAT	Mobility trends for places like grocery markets, food warehouses, farmers markets, specialty food shops, drug stores, and pharmacies, compared to a pre-pandemic baseline.
Parques	FLOAT	Mobility trends for places like national parks, public beaches, marinas, dog parks, plazas, and public gardens, compared to a pre-pandemic baseline.
Estações transporte públicos	FLOAT	Mobility trends for places like public transport hubs such as subway, bus, and train stations, compared to a pre-pandemic baseline.
Locais de trabalho	FLOAT	Mobility trends for places of work, compared to a pre-pandemic baseline.
Residencial	FLOAT	Mobility trends for places of residence, compared to a pre-pandemic baseline.

Table 40 – Waze COVID-19 Impact data source metadata

Field	Data Type	Description
Date	DATE	Date

Mobility Index Change	FLOAT	Percentual change of kilometers driven per day measured with the Waze App within the city of Lisbon against a pre-pandemic baseline

Table 41 – Lisbon international airport traffic data source metadata

Field	Data Type	Description
Centro Negócios	STRING	Airport name
Date	DATE	Date
Movement type	INT	Arrival / Departure indicator
No-transfer passengers	INT	Number of passengers either boarding in Lisbon or deboarding in Lisbon for the given date. Transferred passengers not included.
Movement Count	INT	Count of aircraft movements for the given date

Table 42 – Lisbon Subway ridership data source metadata

Field	Data Type	Description
Dia de Exploração	DATE	Date
Passe	INT	Regular title validation (i.e: monthly pass)
Ocasional	INT	Occasional ticket validation
Gratuito Criança	INT	Child free validations (up to 12yrs)
Outros Gratuitos	INT	All other free validations
Soma	INT	Sum of all validations in the given day

Table 43 – Death certificates data source metadata

Field	Data Type	Description
Data	DATE	Date
Área de saúde	STRING	Health region (i.e: Lisbon and Tagus Valley)
Mortes	INT	Total number of deaths

Table 44 – Bike Lane counter (Av. Duque de Ávila) data source metadata

Field	Data Type	Description
Time	DATETIME	Date
Piloto Lx	INT	Total number of bikes passing through the sensor
Piloto Lx Ciclistas Entradas	INT	Total number of bikes passing east bound
Piloto Lx Ciclistas Saídas	INT	Total number of bikes passing west bound

Table 45 – GIRA Trips data source metadata

Field	Data Type	Description
Data	DATE	Date
Viagens	INT	Number of GIRA trips in the given day