MGI

Master Degree Program in
Information Management

Air Quality Monitoring in Lisbon

Business Intelligence approach to Smart Sustainable Cities

Ana Catarina Maia Ribeiro

Project Work

presented as partial requirement for obtaining the Master Degree Program in Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

AIR QUALITY MONITORING IN LISBON

By

Ana Catarina Maia Ribeiro
R20181158

Project Work presented as partial requirement for obtaining the Master's degree in Information Management, with a specialization in Knowledge Management and Business Intelligence

Supervisor: Prof. Dr. Miguel de Castro Neto

Co-Supervisor*:* Prof. Bruno Jardim

July 2023

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledge the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Corroios, July 13th, 2023

DEDICATION

Aos meus pais, que sempre me encorajaram a atingir os meus objetivos e sempre acreditaram em mim. Sem vocês, nada disto seria possível.

Ao meu irmão, obrigada pela persistência e paciência e por me fazeres querer ser sempre melhor.

Ao Diogo, obrigada pela paciência e carinho, por me fazeres acreditar em mim e por me teres dado o equilíbrio que precisei durante esta jornada.

Aos meus amigos, obrigado pelo vosso apoio e amizade, estejam perto ou longe.

ABSTRACT

Rapid population growth has had numerous consequences for our environment, namely in the air quality of urban areas. In order to achieve smarter and more sustainable cities, we need to act quickly to reduce the emission of pollutants into the atmosphere, through daily and real-time control. With the rise of innovative technologies and new ways of generating real-time data through sensors and cameras, smart cities began to take shape, using new ways of managing their resources, assets, and services more efficiently. New studies began to emerge in the Business Intelligence field in which the behavior of citizens was analyzed, helping cities in their short-term decision-making. However, these studies sometimes do not include important variables that could influence air quality, such as traffic, industrial production, and meteorological data, and fail to tackle the challenges related to big data generated by sensors. This study seeks to implement a Business Intelligence solution that allows the measurement of air quality of Lisbon in real-time, through data generated by sensors, relating it to the behavior of communities and other environmental metrics. This solution aims to rank the areas of Lisbon according to the air quality, understand if air pollution is influenced by traffic and weather, and create an alert system that informs the relevant authorities whenever a given environmental metric exceeds healthy levels. The development of this solution was achieved in Microsoft Power BI, where several dashboards were constructed to analyze real-time data, uncover hidden patterns in historical data, and visualize the correlation between air quality, traffic, and meteorological variables. This study demonstrated the importance of using Business Intelligence in the integration, processing, and analysis of large amounts of sensor data, i.e. big data. Thanks to the Power BI dashboards created, it was possible to determine which areas of Lisbon displayed the highest level of pollutants, an information which is useful to authorities, who can use it to better prevent and tackle unhealthy levels of air pollution. This solution can be extended beyond Lisbon and adapted for application in other cities within Portugal or even globally.

KEYWORDS

Business Intelligence; Air Quality; Smart city; Sustainable city; Sensor Data; Air Pollution

Sustainable Development Goals (SGD):

INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

AQG	Air Quality Guidelines
AQI	Air Quality Index
BI	Business Intelligence
BDA	Big Data Analytics
C_6H_6	Benzene
CO	Carbon Monoxide
DW	Data Warehouse
ETL	Extract, Transform and Load
ICT	Information and Communication Technology
IoT	Internet of Things
KPI	Key Performance Indicator
NO	Nitrogen Monoxide
NO_2	Nitrogen Dioxide
O_3	Ozone
PM	Particulate Matter
PM_{10}	Particulate matter with a diameter smaller than 10 micrometers
$PM_{2.5}$	Particulate matter with a diameter smaller than 2.5 micrometers
SO_2	Sulfur Dioxide
WHO	World Health Organization

1. INTRODUCTION

Rapid population growth has had numerous consequences for our environment, namely in the air quality of urban areas. With a higher population density, there are more vehicles around the city and, therefore, more greenhouse gases being emitted into the atmosphere (Shahid et al., 2021). In order to achieve smarter and more sustainable cities, we need to act quickly to reduce these emissions through daily and real-time control (Shahid et al., 2021; Xiao & Ji, 2009).

With the rise of modern technologies and new ways of generating real-time data through sensors and cameras, smart cities began to take shape, using new ways of managing their resources, assets, and services more efficiently. New studies began to emerge in the Business Intelligence (BI) field in which the behavior of citizens was analyzed, helping cities in their short-term decision-making (Jardim et al., 2022) to become smarter and more sustainable. More specifically, the number of studies on atmospheric gases in cities has been increasing not only because of the rising popularity of environmental issues, such as global warming, but also because of the emergence of diseases associated with this phenomenon, such as asthma and cardiovascular disorders (WHO, 2021). Notwithstanding the relevance of these studies, they sometimes do not include important variables that could influence air quality, such as traffic and industrial production (Taborda et al., 2020) or meteorological data (Xiao & Ji, 2009), and fail to tackle the challenges related with big data generated by sensors (Rios & Diguez, 2014).

This study seeks to implement a BI solution that allows the measurement of air quality of Lisbon in real-time, through data generated by sensors, relating it to the behavior of communities and other environmental metrics. This solution aims to rank the areas of Lisbon according to the air quality, understand if air pollution is influenced by traffic and weather, and create an alert system that informs the relevant authorities whenever a given environmental metric exceeds healthy levels.

To achieve these objectives, a BI solution was developed in which environmental parameters, provided by the Lisbon City Council, and traffic data, provided by Waze, were integrated, and processed using Microsoft Power BI. Following the data transformation, several dashboards were constructed using this tool to analyze real-time data, uncover hidden patterns in historical data, and visualize the correlation between air quality, traffic, and meteorological variables. Moreover, this BI tool, has the potential of provide up-to-date access to Lisbon's air quality measurements, and support real-time decision-making. The analysis provided by the dashboards enables the access to relevant information for the authorities of the Lisbon City Council, e.g., the awareness of the areas with the most deteriorated air quality, and which are the main causes of this phenomenon. With this information, it will be possible to implement policies to control those causes, therefore contributing to a smarter and more sustainable city.

This study demonstrated the importance of using BI in the integration, processing, and analysis of large amounts of sensor data, i.e. big data. Thanks to the Power BI dashboards created, it was possible to:

- Determine which areas of Lisbon displayed the highest level of pollutants.
- Identify which pollutants do not meet the healthy concentration limits and how severely.
- Identify air quality trends over time.

- Identify periods during day, week, month, quarter, and year when air quality is more deteriorated.
- Identify how traffic and weather data impact air pollution levels.

This solution can be extended beyond Lisbon and adapted for application in other cities within Portugal or even globally.

The next chapter of this master thesis comprises a literature review exploring Smart and Sustainable Cities, as well as air quality monitoring studies worldwide, with a specific focus on Lisbon, Portugal. The "Data" chapter presents an overview of the data sources used for the BI solution. The "Methodology" chapter describes the framework's construction process, while the "Results and Discussion" chapter analyzes the created dashboards and highlights key findings regarding Lisbon's atmospheric quality. The "Conclusions and Future Works" chapters cover project implications, limitations, main conclusions, and future steps.

2. LITERATURE REVIEW

2.1. SMART AND SUSTAINABLE CITIES

Bibri & Krogstie (2017) defined Smart and Sustainable cities as "a new techno-urban phenomenon" and as the convergence of sustainability awareness, urban growth, and technological development to provide society with a better quality of life. These cities are extremely supported by the presence and use of Information and Communication Technologies (ICT).

2.1.1. Smart cities

When defining a Smart City, Bibri & Krogstie (2017), Chourabi et al. (2012), and Caragliu et al. (2009) agreed in the fact that although the use of the term "smart city" has been increasing throughout the years, there is still no consensus about its meaning. It is a vague notion. Nonetheless, each author tried to find the best description for the concept of smart cities.

Caragliu et al. (2009) based the definition of a smart city on 6 dimensions: smart economy; smart mobility; smart environment; smart people; smart living; and, smart governance. He described it as "*…when investments in human and social capital and traditional (transport) and modern (ICT) communication infrastructure fuel sustainable economic growth and a high quality of life, with a wise management of natural resources, through participatory governance.*".

Bibri & Krogstie (2017) described a smart city similarly but divided it into 2 approaches: 1) the technology and ICT-oriented approach and 2) the people-oriented approach. The former focuses on the advancement of technologies and hard infrastructure (transport, energy, communication, waste, water, etc.) using ICTs, and the latter focuses on enhancing citizens' quality of life.

Chourabi et al. (2012), on the other hand, depreciates the second approach in his definition of smart cities, describing them as the combination of 4 cores: digital telecommunication networks, ubiquitously embedded intelligence, sensors and tags, and, lastly, software.

2.1.1.1. Information and Communication Technology (ICT)

ICTs are the group of infrastructures, applications, data analytic capabilities, and services that enable smart governance of cities (Chourabi et al., 2012), based on smart computing technologies. They are considered crucial for a city to be smart (Bibri & Krogstie, 2017).

In many smart cities, the main focus seems to be the role of ICTs, which are the key drivers and the core of smart city initiatives (Caragliu et al., 2009; Chourabi et al., 2012). These technologies provide data that helps the development of urban intelligence functions and the formulation of powerful and predictive insights, crucial to strategic decision-making (Bibri & Krogstie, 2017). Therefore, cities rely on ICTs to remain sustainable and address the challenges related to urban growth (Bibri & Krogstie, 2017; Chourabi et al., 2012).

Urban ICT encompasses hardware and software components, including:

- Sensors
- Computers and terminals

- Smartphones
- Internet Infrastructure
- Wireless communication networks
- Telecommunication systems
- Database systems
- Cloud computing infrastructure
- Middleware architecture

Middleware Architecture includes all software that runs on top of the hardware systems, with big data analytics capabilities, database integration, visualization methods, real-time operation methods, and decision support systems, among others (Bibri & Krogstie, 2017).

2.1.1.2. Internet of Things (IoT) for Smarter Cities

IoT is the main component of ICT, empowering its capacity to improve environmental sustainability. This concept is defined by the connectivity that exists between the physical world and the informational world, integrated by the Internet infrastructure, and supported by a wide range of smart data-sensing devices. These networks between physical objects (people, railways, machines, animals, air, etc.) and smart data-sensing devices (accelerometers, laser scanners, infrared sensors, etc.) have been rapidly evolving throughout time and becoming increasingly sophisticated. Its main goal is to extract information and control, manage, and plan things around cities (Bibri, 2018).

2.1.2. Sustainable cities

In parallel to the definition of "Smart Cities", the same dilemma appears when it comes to finding a common definition of what a "Sustainable City" is. Bibri & Krogstie (2017) characterized it as the adoption of sustainable development strategies that contribute, on the one hand, to environmental quality and protection (diminish the environmental impacts caused by the intensive consumption of energy), and on the other hand, to social equity and quality of life (social justice, safety, and stability). Moreover, concerning environmental impacts, sustainable cities focus on the maximization of the efficiency of energy and resources, promote zero-waste systems and use of renewable energy production, minimize pollution, and transportation needs, and preserve ecosystems.

2.1.3. Air Quality

Nowadays, with the constant urban growth, cities have to deal with tremendous difficulties regarding waste management, resource scarcity, air pollution, human health, traffic congestion, among others, triggering cities around the world to become smarter and more sustainable (Chourabi et al., 2012). Additionally, as discussed in section 2.1.2, for a city to be sustainable, it must make a constant effort in minimizing the environmental impacts caused by its activities, which includes the management of the environmental issues listed above.

More specifically, one of the main concerns of smart sustainable cities is the monitoring of air quality. Air pollution is a global problem, particularly in recent years, as it is one of the leading causes of mortality worldwide, causing millions of deaths every year (WHO, 2021) and mainly impacting low-to-middle-income countries (Ritchie & Roser, 2017).

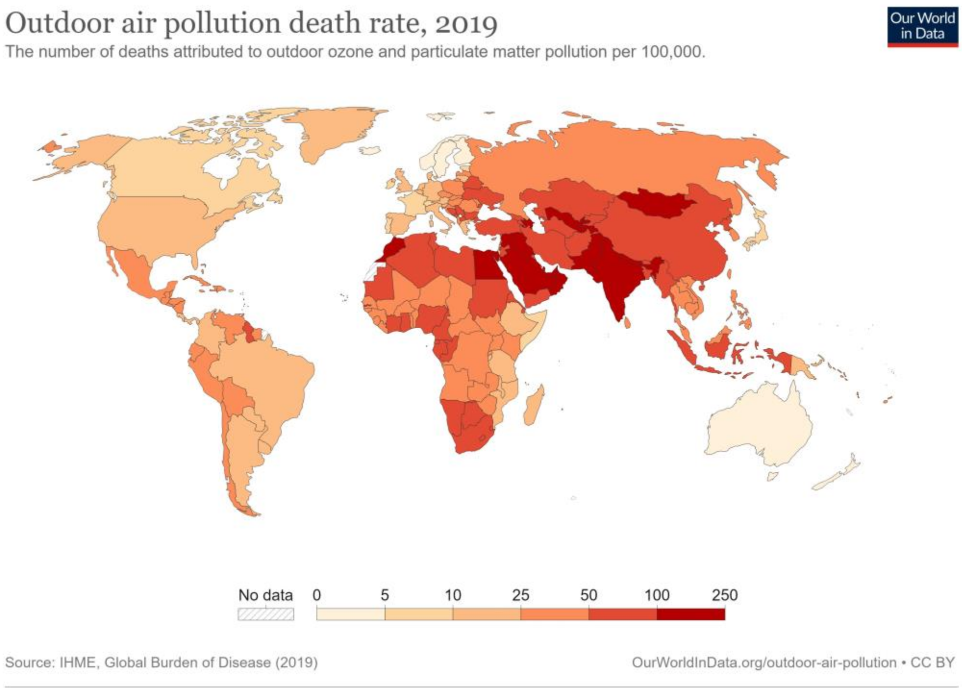


Figure 2.1 – Outdoor air pollution death rate in 2019 (adapted from Ritchie & Roser, 2017)

Recent research on the consequences of air pollution has revealed that exposure to pollutants such as Particulate Matter (PM) with a diameter of less than 10 µm is associated with life expectancy shortening, lung cancer, and non-malignant cardiopulmonary mortality. Other studies have found a connection between air pollution, especially gases emitted by traffic, and lung cancer (Brunekreef & Holgate, 2002). Furthermore, according to the WHO[1], the combined impacts of outdoor and indoor pollution *"cause about 7 million premature deaths every year from increased mortality from stroke, heart disease, chronic obstructive pulmonary disease, lung cancer, and acute respiratory infections."*

In addition to all the public health consequences, there are serious environmental effects on the climate and ecosystems globally (Manisalidis et al., 2020; WHO, 2021). Since greenhouse gases emitted by the extraction and combustion of fossil fuels are the primary cause of atmospheric pollution and climate change, all policies that reduce the high concentrations of these gases in the atmosphere (e.g. public transportation, renewable energies) benefit not only air quality but also public health (Ramanathan & Feng, 2009). Thus, it is of the utmost importance to monitor and manage air pollution levels to develop and implement new policies to address this issue.

The WHO provides worldwide evidence-based guidelines for the maximum concentrations of air pollutants that all countries should follow to achieve good air quality. These standards have been modified several times throughout the years, with the most recent amendment being made in 2021 in response to the real and continuous threat of air pollution to public health.

[1] https://www.who.int/teams/environment-climate-change-and-health/air-quality-and-health/health-impacts/exposure-air-pollution

Table 2.1 – WHO recommended Air Quality Guidelines (AQG) levels in µg/m^3 (adapted from WHO, 2021)

Measurement indicator	Acronym	Period	AQG Level
Carbon Monoxide	CO	daily average	4000
Ozone	O_3	8-hour average	100
Particulate Matter	PM_{10}	daily average	45
Sulfur Dioxide	SO_2	daily average	40
Nitrogen Dioxide	NO_2	daily average	25
Particulate Matter	$PM_{2.5}$	daily average	15

In Portugal, the Lisbon City Council also established Environmental Emission Index Criteria that informs if the values of certain environmental factors exceed the normal level, following the *Agência Portuguesa do Ambiente* index (Câmara Municipal de Lisboa, 2021).

Table 2.2 – Lisbon's Environmental Emission Index Criteria in µg/m^3 (adapted from Câmara Municipal de Lisboa, 2021)

Measurement indicator	Acronym	Period	Normality	Lisbon environmental index		
			Green	Yellow	Orange	Red
Carbon Monoxide	CO	8-hour average	10000	-	-	-
Sulfur Dioxide	SO_2	hourly average	200	201-180	351-500	>501
Nitrogen Dioxide	NO_2	daily average	100	101-200	201-400	>401
Ozone	O_3	hourly average	100	101-200	181-240	>241
Particulate Matter	PM_{10}	hourly average	35	36-50	51-100	>101
Particulate Matter	$PM_{2.5}$	hourly average	20	21-25	26-50	>51

The main pollutants contained in the air we breathe, as shown in both tables above, are particulate matter, ozone, nitrogen dioxide, and sulfur dioxide, which are originated from a variety of sources such as industrial chimneys, traffic, power generation, agriculture, desert dust, among others (Lelieveld et al., 2015). The selection of these pollutants was based on their global significance, considering factors such as the strength of evidence, severity of health outcomes, burden of disease, and other relevant considerations (WHO, 2021).

Nitrogen dioxide is a highly reactive gas that contributes to acid rain formation and is primarily caused by road traffic through the combustion of fossil fuels (Brunekreef & Holgate, 2002), with serious consequences for the health of the most vulnerable population, especially respiratory disorders (WHO, 2021). It also contributes to the development of ozone, which influences both air quality and climate (Santos et al., 2019).

Ozone is classified as a secondary pollutant since it originates from a chemical interaction between other pollutants, such as nitrogen oxide and organic compounds. When present in the "ozone layer," ozone is good for our planet because it protects us from dangerous ultraviolet radiation; but, when detected in the lower atmosphere, in the air we breathe, it is harmful to health, causing lung inflammation (Kampa & Castanas, 2008), among other problems. Additionally, this gas is found in higher quantities on days with stronger solar radiation, such as during summer, and low wind intensity (Câmara Municipal de Lisboa, 2020).

Sulfur oxides are formed during energy production and industrial processes, such as the burning of sulfur-containing materials like coal and oil (Guarnieri & Balmes, 2014). This is another pollutant that, in addition to having severe consequences for the cardiovascular system, also causes respiratory disorders, since it is soluble in water, present in the air we breathe, and can produce irritation and mucus hypersecretion in the lungs.

Particulate matter is a combination of solid and liquid particles. Large particles originate from the friction of larger particles, e.g. from the erosion of road surface, whereas small ones are mostly formed from gases (Brunekreef & Holgate, 2002). Particulate Matter might come from solid-fuel combustion, industrial and agricultural operations, or even from traffic (Santos et al., 2019). Exposure to particulate matter has been linked to cardiovascular morbidity and death in the elderly (Gordon & Reibman, 2000; Samet et al., 2000). Furthermore, one of the natural events with the largest impact on Particulate Matter levels is dust storms, driving these levels far beyond their healthy ranges. These phenomena have an impact on numerous urban areas across the world, in particular Southern Europe, Eastern Asia, Australia, West Africa, North America, and the Middle East, and worsen various health problems such as respiratory and cardiovascular disorders, skin irritations, among others (Goudie, 2014).

Carbon monoxide is produced by the incomplete combustion of fossil fuels, with traffic being the primary source of this pollutant. Regarding its consequences to health, this chemical can progressively replace the oxygen in the blood's hemoglobin, lowering its capacity to transport oxygen and causing breathing difficulties, impaired concentration, slow reflexes, and, in extreme situations, asphyxia (Kampa & Castanas, 2008).

2.1.3.1. Air quality monitoring around the world

One way of monitoring air pollution is achieved by IoT technology, e.g. through sensors installed around the city that collect different pollutant concentrations and store them for further analysis (Kingsy et al., 2016; Suakanto et al., 2013; Xiaojun et al., 2015). These sensors generate large volumes of data, at high speed, which needs to be managed and integrated in real-time so that cities can extract useful insights and act rapidly. Nevertheless, once big data becomes too complex and difficult to handle by traditional data processing tools, a new challenge arises: understanding how to collect, store, and process big data (Cai et al., 2017).

This challenge can be approached by using BI, which, based on data management and warehousing, can transform big data into valuable insights through its analytical and “anywhere, anytime” support. According to Chen et al. (2012) it is composed of a set of business reporting functions, including:

1. The design of data marts and tools to integrate data through Extract, Transform and Load (ETL) processes.
2. The database query, OLAP, and reporting tools to explore and analyze data.
3. Business Performance Management (BPM) that resorts to dashboards and scorecards to enhance the understanding of the data.
4. Data Mining techniques to extract meaningful insights.

Given the importance of studying air quality in urban areas and the necessity of analyzing it in real-time to make cities smarter and more sustainable, many authors decided to approach this subject, using many different techniques. A brief summary of the main studies published on this topic is displayed in Table 2.3, and a more detailed description is presented afterwards.

Table 2.3 – Summary of previous studies on air pollution worldwide

Research objective	Approach	Reference
Studied air pollution by evaluating air quality status, analyzing its spatio-temporal characteristics, forecasting, and providing decision-making support	Implementation of a prototype system of air quality Data Warehouse (DW) and a web spatial OLAP system to enhance the ability of spatial analysis and visualization	(Xiao & Ji, 2009)
Model the spatial correlation between the air quality of various locations and model the temporal dependency of air quality in a location	Semi-supervised learning based on a co-training framework	(Zheng et al., 2013)
Integrate Big Data tools for the gathering, storage, and analysis of data generated by sensors that monitor air pollution levels in a city	Use of big data tools such as Hadoop and Storm for data processing, storage, and analysis.	(Rios & Diguez, 2014)
Process and analyze the high-volume data generated by gas sensors in an air pollution monitoring system	Enhanced K-Means clustering algorithm to analyze the air pollution data	(Kingsy et al., 2016)
Enhance the well-being of citizens in the urban environment by monitoring air pollution through the construction of dashboards and extracting useful insights about pollutant concentration at different time granularity levels	Data analysis engine, based on BI methodologies to support air pollution data analysis	(Baralis et al., 2016)

Xiao & Ji (2009), used Oracle 10g and ArcGIS 9.0 to create a prototype of an air quality DW to examine the air quality of 86 cities in China between 2000 and 2007, using a star schema model to show a dynamic technological association between data and spatial geometry. This framework gathered and stored real-time air quality data in a data mart associated with each city, where ETL techniques were later applied. Next, under a star schema approach, all data marts were combined into a single DW. Moreover, an OLAP system was created to enhance the spatial visualization of the data and analyze it at various degrees of granularity.

Zheng et al. (2013), used real-time (hourly) and historical data on the concentration of several pollutants (SO_2, NO_2, $PM_{2.5}$, and PM_{10}) from a small number of sensors, as well as meteorological data (temperature, humidity, barometric pressure, wind speed, and weather), traffic flow, human mobility, the road network structure, and Points of Interest (POIs) in order to understand their correlations. It was also found that atmospheric pressure enhances air quality while PM_{10} concentration increases with humidity and decreases with wind speed. Additionally, when the pressure is low and the temperature is high and vice versa, air quality also improves. Traffic speeds was shown to be directly correlated with nitrogen dioxide concentrations, more specifically, lower car speeds (<20 km/h) were associated with better air quality, and higher speeds (>40 km/h) displayed the contrary. Moreover, it was shown that levels of PM_{10} and NO_2 pollution worsen when human mobility increases. Finally, different POIs display different values for air quality, according to their nature; for example, the air quality is better around parks, but worse in areas around factories, due to the industrial production linked to air pollution. Following the identification of the different features, a co-training-based semi-supervised learning approach was used. Two classifiers—spatial and temporal data—were employed, and the outcomes of this approach outperformed other well-known supervised learning models also used in this study in terms of accuracy.

Rios & Diguez (2014), employed a wireless system network with sensors to gather and store air quality data in a repository and used a distributed system to handle big data created by the sensors. This system was split into two parts: data acquisition and data processing. During data acquisition, a prototype node recorded variables such as temperature, humidity, CO and O_3 and performed some basic pre-processing techniques. Following that, the data was transferred to a gateway node, which connected to the rest of the system before reaching the data processing module. In this second module, data was handled in two steps: one based on Storm, which performed stream processing and allowed for real-time alerts and warnings, and another in Hadoop with MapReduce algorithms, which performed more sophisticated analyses on historical data.

Kingsy et al. (2016), compared the accuracy of an unsupervised approach, the improved K-means clustering, with another algorithm, the Possibilistic Fuzzy C-Means (PFCM), to describe real-time air quality data collected by sensors (Particular Matter, SO_2, O_3, CO, NO). Using the same algorithms, the correlation of atmospheric gas concentration variables with environmental variables such as wind speed and wind direction was also computed. The primary goal was to determine which of the algorithms was more efficient and resulted in higher accuracy when computing the air quality index (AQI), and it was concluded that K-means performed better.

Baralis et al. (2016), performed an air pollution analysis (APA) in Milan using BI techniques on real-time data. The air concentration of several pollutants (PM_{10}, $PM_{2.5}$, CO, O_3, NO_2, C_6H_6, SO_2) was collected, along with weather data (temperature, relative humidity, precipitation level, wind speed, and atmospheric pressure) and traffic data (number of vehicles entering a given urban area, hourly) to determine the impact of these variables on air pollution. After collecting the data, it was pre-processed and stored in a DW, and KPIs were set up to access the concentration of a given atmospheric pollutant, which would later be visually represented in the dashboards built in the APA's final phase. These dashboards supplied in-depth and valuable analyses of air quality and how human activity might influence atmospheric pollution.

More specifically, in Lisbon, some studies have been made, regarding air quality. Ferreira et al. (2000) presented a comprehensive methodology in his study, which integrates an air quality monitoring framework specific to Lisbon, using sensor-collected data, and establishes correlations with traffic data. This investigation provided valuable insights into the predominant vehicle types contributing to specific pollutant levels. To ease the analysis of air quality within the expansive urban area of Lisbon, the study employed spatio-temporal kriging as a powerful tool. The findings conclusively demonstrated that traffic represents a primary factor influencing air quality conditions. Moreover, the study revealed that passenger vehicles exhibit the highest pollutant emissions outside of peak hours, whereas buses contribute the most during these periods. In light of these observations, several potential mitigation measures were proposed to address atmospheric pollution in Lisbon, encompassing enhancements in traffic management, the establishment of speed limits, the promotion of cycling as an alternative mode of transportation, and other relevant interventions.

Jorge (2009) devised a robust methodology in his research aimed at forecasting the maximum hourly concentration of O_3 and the daily average concentration of Particulate Matter for the following day at air quality monitoring stations situated within the Lisbon metropolitan area. To achieve this, he employed statistical models grounded in multiple regression analysis, and classification and regression tree analysis. The application of these models yielded remarkable success in accurately predicting the daily average levels of the targeted pollutants.

Taborda et al. (2020) developed a platform to visualize the air quality in Lisbon, employing interactive dashboards. These visualizations were constructed using real-time sensor data, specifically targeting $PM_{2.5}$, along with environmental contextual data, including spatio-temporal mobility data. The resulting dashboard facilitated the visualization of particulate matter concentration severities by employing color-coded areas on the map. The European Union Air Quality Index thresholds were used for this purpose, categorizing the severities as follows: 1) Good (Green), 2) Fair, 3) Moderate, 4) Poor (Yellow), 5) Very poor, and 6) Extremely poor (Red). The outcome of this study yielded a user-friendly dashboard that enables convenient access to a comprehensive overview of air quality in Lisbon.

In conclusion, the use of BI techniques and big data tools is not only important to enhance the integration, analysis, and reporting of air quality data (Chen et al., 2012), but also to deal with the volume and velocity of sensor data. However, many studies on the air quality around the world, while dealing with sensor data, ignore the limitations of traditional data processing tools or do not use context variables important to understand the behavior and causes of air pollution.

3. DATA

The data used to generate the BI solution was obtained from "Lisboa Aberta," an open data strategy developed by the Lisbon City Council to promote a transparent policy for the development of inhabitants and the provision of improved services. This project began in 2017 and will allow for an environmental diagnostic of the city, mapping out priority areas for intervention, investing in public information, and implementing measures to combat air pollution (Câmara Municipal de Lisboa, 2021). More information on Lisbon's City Council website[2].

The data is collected by eighty distinct sensor stations that are evenly distributed around the city and mounted on light poles, as can be seen in Figure 3.1. These sensors are designed to detect environmental quality parameters such as air quality, noise, weather, and traffic in real-time. They monitor multiple parameters including pollutant concentrations ($PM_{2.5}$, PM_{10}, CO, SO_2, NO_2, NO, O_3, C_6H_6), sound level, atmospheric pressure, relative humidity, strong wind, precipitation, temperature, ultraviolet radiation, and hourly traffic volume per direction. However, not all the eighty sensor locations allow measurement of all parameters listed above. The variables from this source used in this study are the ones related to air quality and weather, meaning that sound level and hourly traffic volume variables were disregarded.

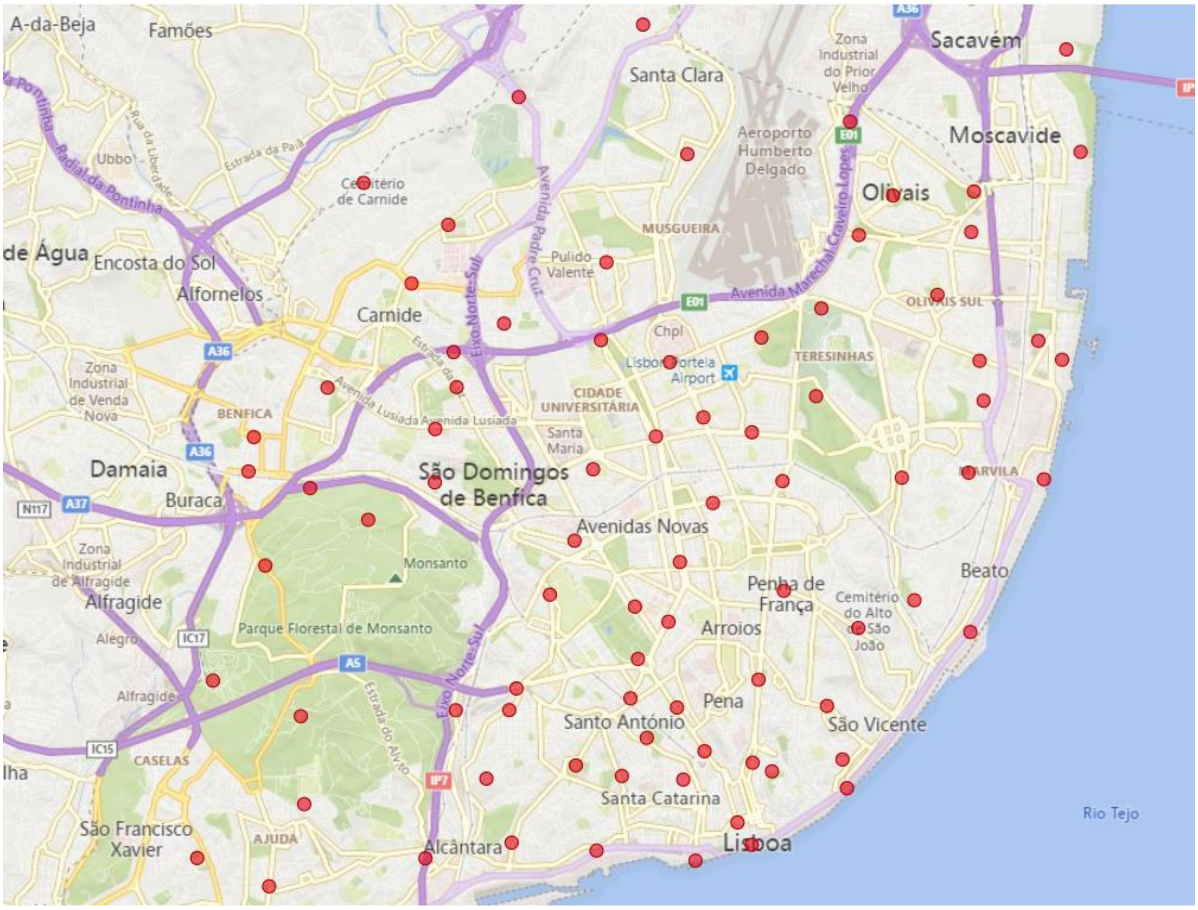


Figure 3.1 – Sensor's Locations in Lisbon

[2] https://www.lisboa.pt/cidade/ambiente/qualidade-ambiental/ar

The data for analysis is generated hourly by aggregating (averaging) measurements obtained every 15 minutes. It can be accessed via the "Lisboa Aberta" website[3], through an API link, which allows users to access two types of data: real-time data, corresponding to a dataset with the most recent measurements of the parameters in all stations, and historical data, which must be filtered by location, parameter, and period. As can be seen in Table 3.1, this source allowed the collection of important variables for the study of air pollution in Lisbon.

Table 3.1 – Description of the most relevant Environmental Parameters in the source data

Variable	Description
ID	The 10-digit measurement ID is composed by the indicator (2 digits), parameter (4 digits), and location (4 digits) of that measurement.
Value	Sensor's measured value of a specific parameter, for a specific hour of the day, in each geographic location
Unit	Measurement unit of the measured parameter (e.g., $\mu g/m^3$, mbar, %)
Address	The name of the street on which the sensor is located
Coordinates (lat, lng)	Coordinates of the location of the sensor that took the measurement

In Figure 3.2 a snippet of the data can be seen.

id	avg	date	value	unit
QAPM250001	1h	202302030000	13	µg/m3
QAPM250001	1h	202302022300	14	µg/m3
QAPM250001	1h	202302022200	17	µg/m3
QAPM250001	1h	202302022100	20	µg/m3
QAPM250001	1h	202302021900	18	µg/m3
QAPM250001	1h	202302021800	18	µg/m3
QAPM250001	1h	202302021700	13	µg/m3
QAPM250001	1h	202302021600	11	µg/m3
QAPM250001	1h	202302021500	9	µg/m3
QAPM250001	1h	202302021400	9	µg/m3
QAPM250001	1h	202302021300	10	µg/m3
QAPM250001	1h	202302021200	9	µg/m3

Figure 3.2 – Environmental parameters' data source snippet

Since the traffic data present in the source of Environmental Parameters was very incomplete, an additional source was collected to better understand the influence of traffic on air quality in Lisbon, provided by *NOVA Cidade*, the Urban Analytics Lab of NOVA IMS, that has a protocol with Waze. Like the environmental parameters source data, the traffic data was also generated hourly, along with other important variables, as can be seen in Table 3.2. However, this source only included data from June 22nd, 2019 to September 22nd, 2022.

[3] https://lisboaaberta.cm-lisboa.pt/index.php/pt/dados/conjuntos-de-dados

Table 3.2 – Description of the most relevant parameters in the source traffic data

Variable	Description
Parish	Parish where the measurement was taken
Geometry	Group of coordinates that formed a polygon cell, representing a street in Lisbon, where the measurement was taken
Average Level	Average level measured
Average Speed	Average speed measured
Total Delay	Total delay (in minutes) measured
Number of Jams	Number of traffic jams (scale from 1 to 5) measured

In Figure 3.3 a snippet of the data can be seen.

date	hour	parish	geometry	avg_level	avg_speed	total_delay	n_jams
16/04/2019	12	Arroios	POLYGON ((-1016536.6788137308 4655382.80	5,00	0,00	-1	1
16/04/2019	12	Avenidas Novas	POLYGON ((-1018073.8344124897 4658465.82	5,00	0,00	-1	1
16/04/2019	12	Lumiar	POLYGON ((-1017661.2594884973 4663118.36	5,00	0,00	-3	3
17/04/2019	3	Avenidas Novas	POLYGON ((-1018073.8344124897 4658465.82	5,00	0,00	-1	1
19/04/2019	3	Campolide	POLYGON ((-1019433.2535644913 4657591.47	5,00	0,00	-1	1
20/04/2019	3	Ajuda	POLYGON ((-1023402.6963003477 4652152.83	5,00	0,00	-4	4
20/04/2019	3	Santa Clara	POLYGON ((-1017434.4207863944 4664408.00	5,00	0,00	-2	2
22/04/2019	3	Carnide	POLYGON ((-1021380.4499381933 4663478.03	5,00	0,00	-1	1
22/04/2019	4	SÃ£o Domingos de Benfica	POLYGON ((-1020766.0695576698 4660215.67	5,00	0,00	-1	1
22/04/2019	6	Alvalade	POLYGON ((-1016188.2184109117 4661025.46	4,00	4,89	120	1
22/04/2019	6	Avenidas Novas	POLYGON ((-1018073.8344124897 4658465.82	3,00	8,61	253	2
22/04/2019	6	Campo de Ourique	POLYGON ((-1019069.3316519783 4654272.69	3,00	5,77	278	2
22/04/2019	6	Campolide	POLYGON ((-1019433.2535644913 4657591.47	3,00	8,61	253	2
22/04/2019	6	Olivais	POLYGON ((-1016501.5591892828 4664059.33	3,33	6,83	405	3
22/04/2019	6	Santo AntÃ³nio	POLYGON ((-1017588.2130630513 4654470.91	3,00	4,99	93	1
22/04/2019	7	Ajuda	POLYGON ((-1023402.6963003477 4652152.83	3,00	20,56	118	1

Figure 3.3 – Traffic Data source snippet

Many Power BI Dashboards were created to assess the air quality in Lisbon. As a result, it was possible to retrieve the status of the air pollutants at the current time while also analyzing their history and correlating it with weather and traffic data.

The Lisbon City Council has established an Environmental Emission Index Criteria that informs if the values of certain environmental factors exceed the normal level, which are included in our source's metadata (see Table 2.2). This data will be represented in our BI solution through KPIs that will graphically display an alert anytime an air pollutant concentration exceeds healthy limits.

The dashboards created will allow for interactive data analysis, making them an important monitoring tool for the control of air quality in Lisbon. Hence, it will be possible not only to study possible correlations between parameters (e.g. lower pollutant concentration during peak hours), but also determine when the parameters' levels are critical. As a result, it will be possible to identify the main contributors to air pollution and decide on the best strategy to tackle this issue.

4. METHODOLOGY

Nowadays, many businesses are taking advantage of BI solutions, where the usage of Decision Support Tools is more essential than ever for improving their competitiveness (Petrini & Pozzebon, 2009). The benefits of adopting these technologies, however, are not limited to businesses; they are also tremendously useful to the government. The adoption of Big Data Analytics (BDA) by the government will significantly reduce bureaucracy, enable the identification of areas of underperformance, and demonstrate where resources should be focused, thus enhancing overall performance (Höchtl et al., 2016).

The application of BI methodologies is critical in the construction of a tool that enables strategic operational decision-making based on data, particularly when dealing with big data, as in the case in this work. Only then will it will be possible to use visualizations to integrate, analyze, and display huge amounts of data (Tavera Romero et al., 2021).

A BI architecture is generally made up of a back-end and front-end component. The former handles the integration, cleansing, and transformation of data from various data sources to feed a DW. This is known as "ETL" (Extract, Transform, and Load). The latter is where a variety of querying, reporting, and analysis tools are used (Dayal et al., 2009).

Even though the traditional structure of a BI solution consists of a set of well-defined steps and processes (Tavera Romero et al., 2021), the BI solution developed in this work does not involve the creation of a DW. As a result, the data was imported using Power BI's tools and applications, and the relational model was fully generated in the same context.

Regarding the back-end portion, the environmental parameters source data was obtained via connection to the "Lisboa Aberta" API, which is available in JSON format. In the case of historical data, the API connection had to be managed by parametrizing the web link with the parameters and period desired for analysis:

http://opendata-cml.qart.pt/measurements/PARAMETERID?startDate= AAAAMMDDHHMM&endDate= AAAAMMDDHHMM

The "PARAMETERID" is a 10-digit word, as mentioned in Table 3.1, and the "AAAAMMDDHHMM" corresponds to the year, month, day, hour, and minute desired for analysis.

To ensure the comprehensive import of all historical data measurements into Power BI, the start date was defined as June 1st, 2021—the date of the earliest recorded measurement. Furthermore, the end date was parametrized and set as February 3rd, 2023. However, the user can edit this date and filter for the date range needed. Regarding the *parameterID*, a parameter was also created, with the list of all *parameterID* measured historically, and used in the Power Query's "invoke custom function" tool, which imported all tables, in which the API link contained the specified parameterID enlisted in the parameter list. This enabled the creation of two fact tables, namely *FACT_HistoricalData*, for air pollution parameters, and *FACT_Weather*, for meteorological parameters.

The traffic data was obtained through 42 distinct CSV files, with each file representing a different month of measurement. The multiple files were subsequently merged into a single file and underwent several modifications applied to the data, using Python. More specifically, given that the location of the Waze traffic sensors was in "polygon" format (set of coordinates that form a road, in this case), as opposed to individual coordinate pairs, an approach was devised to link the traffic data to the nearest environmental sensor. Consequently, by aligning both tables to a consistent geographic format, it was possible to analyze the correlation between traffic data patterns and environmental pollutant levels.

Then, all the necessary transformations for these three sources (real-time and historical environmental data, and traffic data) were performed in Power Query. These transformations encompassed:

- Modification of variables' data types.
- Split of "ID" column (see Table 3.1), to get information about the Parameter and sensor's location.
- Elimination of nonessential columns deemed irrelevant for the analysis.
- Replacement of values "-99" with null in environmental parameters' data, in line with the guidelines stipulated in the metadata provided by Lisbon City Council (Câmara Municipal de Lisboa, 2021).

Following the inclusion of the aforementioned data sources, contextual information encompassing location, date, time, and environmental parameters were included. This contextual data was imported with Excel files, leveraging the provided metadata. Notably, the Date table itself was generated entirely through Power Query features. After this data transformation step, the data model was created.

4.1. MULTIDIMENSIONAL MODEL

According to Kimball & Ross (2013), the primary goals of data warehousing and BI are as follows:

1. Providing easier access to data
2. Resenting data consistently
3. The system being prepared for changes
4. The system presenting information in a timely manner
5. The system serving as a reliable source for decision making

Hereafter, the most important aspect is that the data is moved from an operational environment to a data warehousing environment via an integration process (Inmon, 2005). One of the most important steps in this process is the database design. There are two basic database design models in the literature: the *relational model* introduced by Inmon, and the *multidimensional model* developed by Kimball. Each strategy has its advantages and disadvantages, but in some cases, one approach is better suited than the other. In the context of the current solution, the multidimensional model is the most appropriate, since it is used for short-term DWs and where there is limited scope for the DW, i.e., there are few processing needs. The relational architecture, on the other hand, is more suited to long-term data warehousing and where a true enterprise approach is required, which is not the case in this solution (Inmon, 2005).

In the *star join model*, a fact table is at the center of the multidimensional model, surrounded by dimension tables that define features of the fact table (Inmon, 2005). The fact table holds data about an organization's business process events; that is, each line of the fact table represents a measured event with a certain degree of detail, called grain. Dimension tables represent the context of the business process event measured by the fact table, defining the "who”, “what”, “where”, “when”, “how”, and “why" of that event (Kimball & Ross, 2013).

Kimball & Ross (2013) developed an effective and well-known four-step technique for the dimensional design process:

1. Select the business process
2. Declare the grain
3. Identify the dimensions
4. Identify the facts

Starting with the first step of the dimensional design process, namely the selection of the business process, the focus of this project was directed toward the measurement of each environmental parameter. Consequently, the identification of the business process in this context reflects into the systematic monitoring and recording of various environmental parameters. This first step is essential because it allows the creation of a specific target design and determines the granularity, fact tables, and dimension tables. Besides the aforementioned business process, two additional processes can be identified, specifically pertaining to the measurement of traffic data and meteorological data. However, given that these processes are not the primary focus of this project, they have not been highlighted.

The next step of the dimensional design process encompasses the grain declaration, i.e. the level of detail at which the measurements are recorded in the dimensional model. It was determined that each individual row contained within the fact table corresponds to a measurement denoting a specific environmental parameter recorded at hourly intervals and associated with a specific point defined by geographic coordinates.

Regarding the identification of dimension tables, in the third step, which provides contextual information for our business process event, the present solution encompasses the inclusion of the dimensions described in Table 4.1.

Table 4.1 – Description of the dimension tables

Dimension Table	**Primary Key**	**Description**
DIM Parameters	*ParameterID*	Information about the parameter measured, including more detail about its designation and measurement unit.
DIM Date	*Date*	Date information enabling the analysis of data at various temporal resolutions, such as year, quarter, month, week, and day. It also allows the identification of weekends and holidays for deeper analysis.
DIM Time	*Time*	Time information that enables data analysis at different periods of the day
DIM Geography	*LocationID*	Geographic information about the sensor's location with descriptive attributes such as address, parish, and coordinates which enables the visualization of location-based data

Within the *DIM Date* and *DIM Geography* tables, it was possible to create hierarchical relationships between variables. This implementation has significantly enhanced the depth and breadth of data analysis, enabling roll-up and drill-down operations. By incorporating these hierarchical structures, the dimensional model enabled the navigation and aggregation of data across various levels of granularity, empowering users to gain comprehensive insights into the underlying data.

Following Kimball's four-step methodology, after determining the different dimensions that form the multidimensional model, the fact tables must be identified. In a fact table, each row consists of a pointer, i.e. a foreign key, that points to each dimension table, which in turn provides multidimensional coordinates (Chaudhuri et al., 2011). In this BI solution, four fact tables were created, as can be seen in Table 4.2.

Table 4.2 – Description of the fact tables

Fact Table	**Description**
FACT LastMeasurements	Air pollution data from the last hour measured by the sensor
FACT HistoricalData	Air pollution historical data, since June 1st, 2021, resulting from the aggregation of all the "last measurements" tables that existed in the past.
FACT Weather	Weather historical data, since June 1st, 2021, resulting from the aggregation of all the "last measurements" tables that existed in the past.
FACT Traffic	Lisbon traffic data generated hourly, measuring variables such as number of jams, average level, average speed, and total delay

With the presence of four fact tables, the model produced a snowflake structure (Inmon, 2005), that is, a structure with multiple fact tables that share dimensions between them.

In Figure 4.1, the outcome of the dimensional model regarding the *FACT HistoricalData* table can be seen. However, it is important to note, that to maintain simplicity, the diagram displayed is not exhaustive. The comprehensive diagrams of the remaining fact tables are presented in Appendix A.

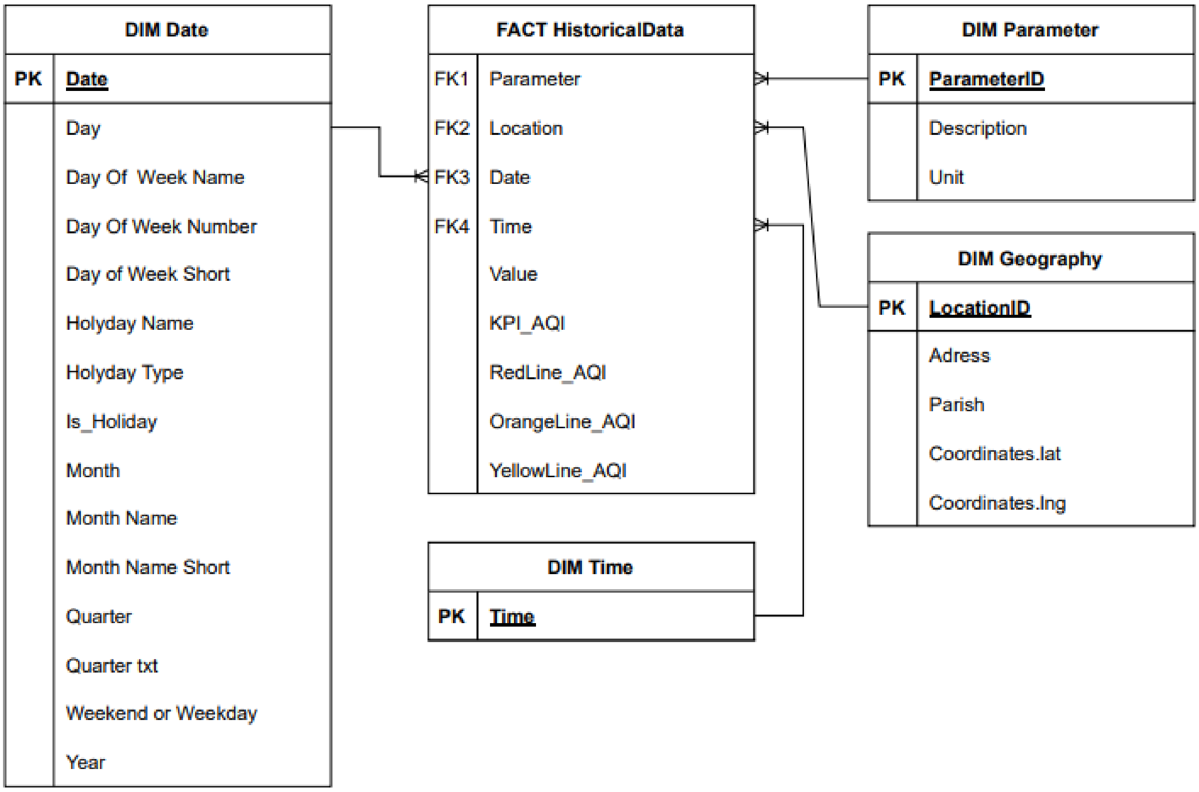


Figure 4.1 – Relational Model for *FACT HistoricalData*

Subsequently, Key Performance Indicators (KPIs) were developed to determine whether the measurements were compliant with the predefined environmental quality thresholds established by the Lisbon City Council (see Table 2.2). The purpose of constructing these KPIs was to gain a comprehensive understanding of the extent to which the specified limits for environmental quality were being respected. For this purpose, the KPI_AQI column was added to both the *FACT HistoricalData* and *FACT LastMeasurements* tables. This column determined the threshold limits categorized as green, yellow, orange, and red for each specific pollutant, as described in detail in Table 2.2. In addition to this KPI, three additional columns (YellowLine_AQI, OrangeLine_AQI, RedLine_AQI) were added to the *FACT HistoricalData* table for visual representation purposes within the dashboards. Each of these supplementary columns corresponded to the yellow, orange, and red limits established for individual pollutants. It should be noted that all these columns were created using the DAX programming language, a feature incorporated into the Power BI environment.

For Carbon Monoxide, the Table 2.2 does not provide specific color-coded limits. Instead, it only specifies an 8-hour average limit of 10 mg/m^3. To establish additional limits, a common-sense approach was adopted. The yellow limit was associated with the provided 8-hour average limit of 10 mg/m^3. The orange and red limits were set at 13 mg/m^3 and 16 mg/m^3, respectively, to provide a gradual progression of lower concentration thresholds. Regarding Nitrogen Dioxide, the dashboard presents limits based on daily averages rather than hourly averages. However, for simplicity and consistency, it was decided to treat these limits as hourly limits.

Lastly, the DAX language was also employed to create metrics that summarize and support the generated visualizations, enhancing the interpretive aspects of the data, can be seen in Table 4.3.

Table 4.3 – Description of Metrics

Metrics	**Description**
SelectedMetric	Air quality metric selected in a given filter, to support the calculation of KPIs according to the chosen selection
NumberOfSensors	Number of different sensors, showing, for a given filtered pollutant, how many sensors participated in its measurement.
YellowCount_LM	Number of times that a specific pollutant, or set of pollutants, exceeded the yellow environmental limit.
OrangeCount_LM	Number of times that a specific pollutant, or set of pollutants, exceeded the orange environmental limit.
RedCount_LM	Number of times that a specific pollutant, or set of pollutants, exceeded the red environmental limit.

Regarding de front-end part of the BI solution, when dealing with big data, visualizations are highly important, as they allow the communication of findings in a short amount of time (Africk & Levy, 2021). They can be used to discover intricate relationships that would otherwise be difficult to find with algorithms. Consequently, visualizations grant meaning to data, making it more intelligible to the human eye (Aragona & Rosa, 2018), thus supporting data-driven decision-making and avoiding the development of insights based on gut intuition (Lavalle et al., 2019). Nevertheless, obtaining the right visualization, according to Lavalle et al. (2019), is extremely challenging because, on the one hand, business users are not skilled in data visualization and do not know exactly what information they want to extract, and on the other hand, the goals of the dashboard designs are frequently not clearly defined. The frequent incidence of these problems may result in misinterpreted graphs and incorrect findings, jeopardizing a company's success.

The primary goal of this project is to depict the current and historical situation of air quality in Lisbon through real-time visualizations. To achieve this, the project will leverage the tools and applications available within the same environment used for data extraction and preprocessing, namely Power BI.

5. RESULTS AND DISCUSSION

As previously mentioned, the primary goal of this BI project is to represent the air quality in Lisbon through the creation of Power BI dashboards. Thus, after using various data transformation techniques, it was possible to create visualizations that allowed the analysis of environmental indicators gathered by sensors spread across seventy-two geographic areas of Lisbon. This analysis highlighted an overview analysis of real-time and historical data and detailed analyses of air pollution contributors. The dashboards were published and made publicly available[4].

5.1. AIR QUALITY REAL-TIME DATA

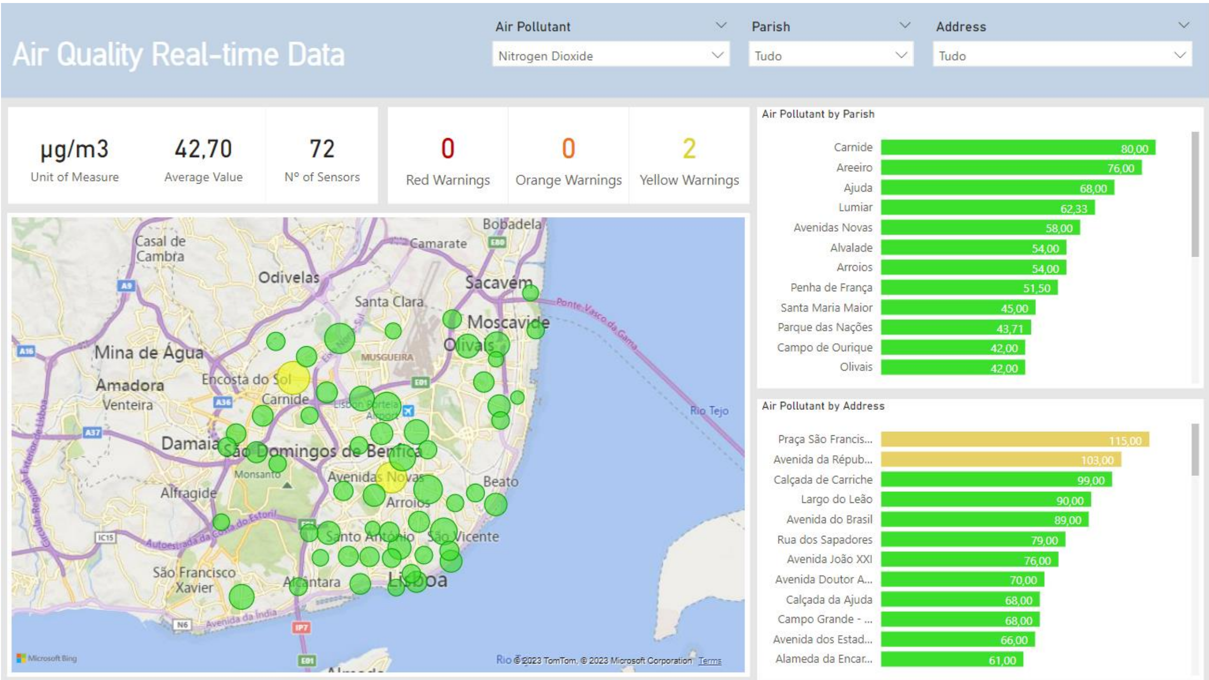


Figure 5.1 – Air Quality Real-time Data Dashboard

The first tab, titled "Air Quality Real-time Data," illustrated in Figure 5.1, provides users with access to the most recent hourly measurements captured by the sensors. The main objective of this tab is to present a quick and concise overview of the present air quality conditions in Lisbon.

Thus, to analyze a specific air pollutant, users can use the slicer to select the desired metric for analysis, dynamically filtering the graphs to display the latest data associated with the chosen metric. In addition to this filter, the user can select which parish or address to explore in further detail or interact directly with the dashboard visuals, for example, by clicking on an area of the map and evaluating it in more detail.

This dashboard helps the comprehension of geographically prominent areas with high pollution levels through a map visualization that provides an overview, and a clustered bar chart that offers more detailed insights into streets and parishes. Additionally, Key Performance Indicators (KPIs) are presented in graphs (through green, yellow, orange, and red colors) and cards, indicating the locations

[4] https://tinyurl.com/32um8745

and total number of sensors that detected excessive concentrations of pollutants in the air, respectively, in accordance with the limits established in Table 2.2. Furthermore, at the top of the page, supplementary details about the selected air pollutant are available, including its unit of measurement, average value, and the number of sensors involved in its measurement.

This form of analysis serves as a valuable tool for the city of Lisbon to accurately assess the present air quality status. By doing so, the city can rapidly implement real-time intervention measures based on facts, to minimize the adverse effects of atmospheric pollution on the health of its residents. Examples of possible measures include improve public transportation, promote cycling and walking, implement car-free zones, among others. Given the severe repercussions of continued exposure to air pollutants on public health, swift action and prompt communication of the current atmospheric conditions to relevant authorities are of the utmost importance. Through the utilization of yellow, orange, and red environmental alerts generated by the dashboard, real-time notifications can be accessed, enabling the identification of parishes and, more specifically, streets that need intervention by the relevant authorities. This mechanism eases prompt and targeted actions to address environmental concerns at specific locations, aiding in the efficient allocation of resources and prompt response to mitigate potential issues.

5.2. AIR QUALITY HISTORICAL DATA

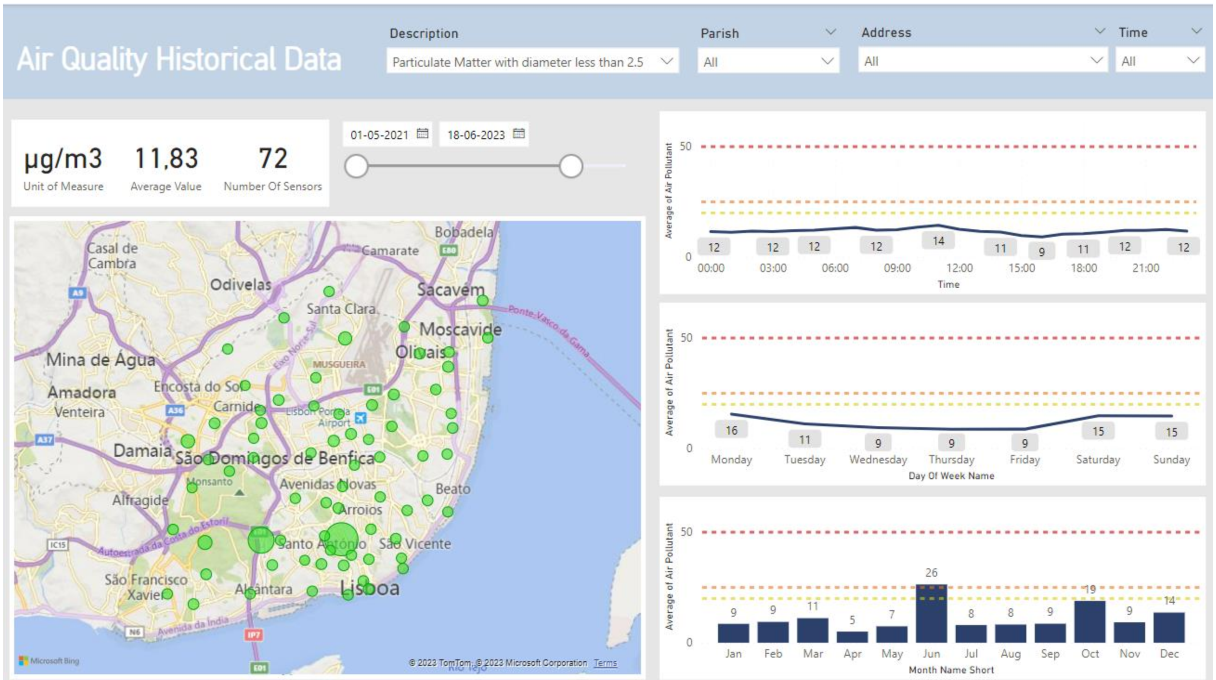
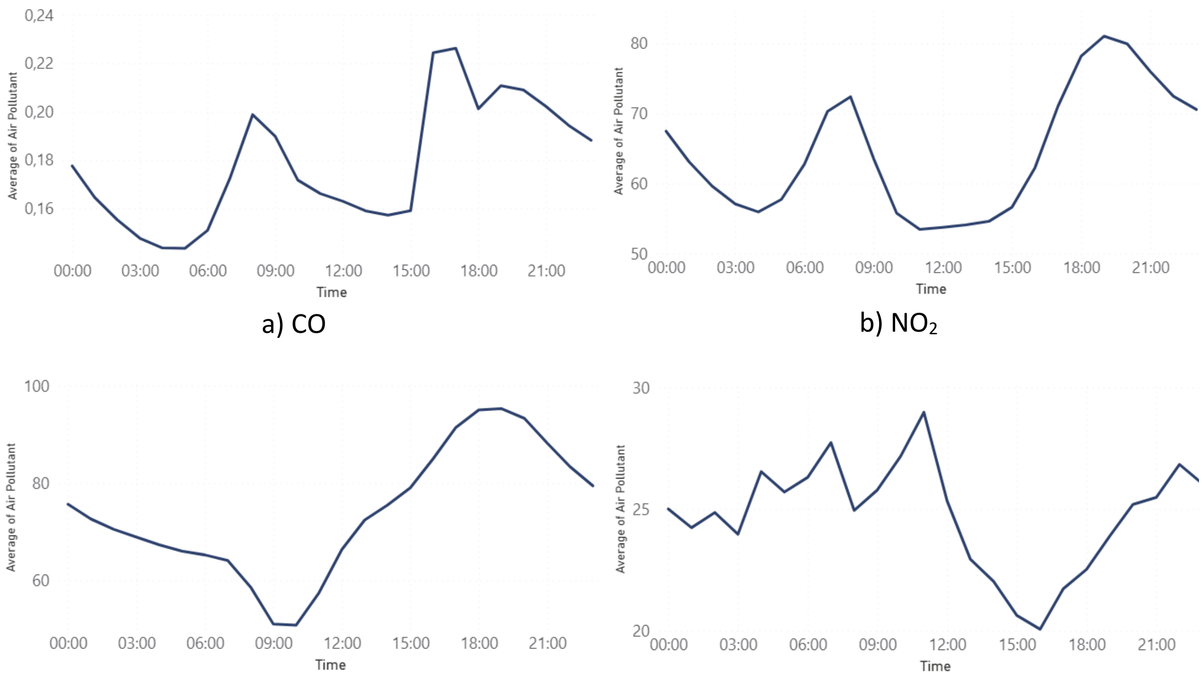


Figure 5.2 – Air Quality Historical Data Dashboard

The second tab, titled "Air Quality Historical Data", illustrated in Figure 5.2, serves a similar purpose to the first tab. However, it aims to provide a comprehensive summary of historical air quality data in Lisbon. Like the first tab, users have the flexibility to select a specific air pollutant, which dynamically filters the remaining graphs for a more focused analysis. In addition to this slicer, users can further refine their analysis by filtering data based on parish, address, hour, and date range. This dashboard effectively displays the spatial distribution of air pollutant values, allowing users to identify geographic locations with the highest concentrations. It also facilitates an understanding of how these values fluctuate across different time granularities, providing insights into temporal patterns and trends. These graphs also have constant lines on the air pollutant concentrations axis, which reflect the KPI regarding the environmental index limits, with the colors yellow, orange, and red, in order to enrich the analysis and show in which periods the pollutants exceeded these limits. To provide additional details about the selected air pollutant, three informative cards are displayed in the top right corner, including its unit of measure, average value, and the number of sensors that measured that particular pollutant.



a) CO

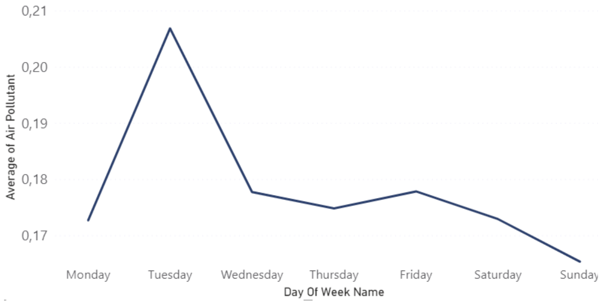
b) NO_2

c) O_3

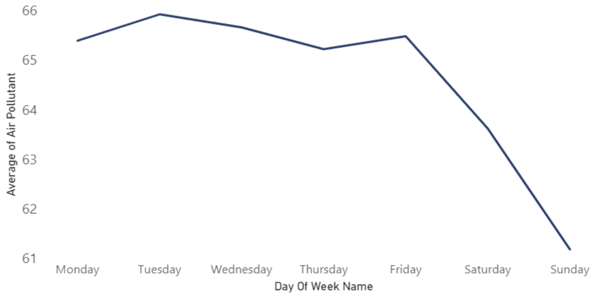
d) PM_{10}

Figure 5.3 – Pollutants' hourly concentrations a) CO, b) NO_2, c) O_3, d) PM_{10}

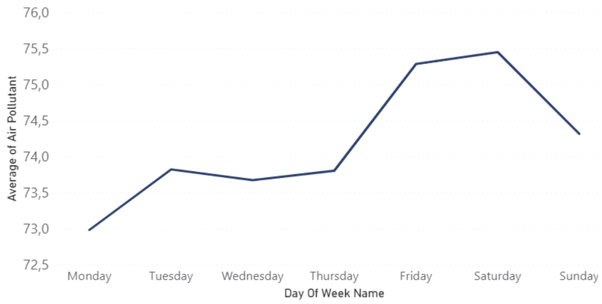
Through a comprehensive analysis of the pollutants depicted on the dashboard, significant conclusions have been drawn. The examination of different time periods throughout the day, as can be seen in Figure 5.3, has revealed noteworthy patterns. Pollutants such as CO, NO_2, PM_{10} exhibit higher concentrations during peak hours, coinciding with the times when Lisbon's residents commute between their homes and workplaces. This observation can potentially be attributed to the heightened usage of combustion engine vehicles, which contributes to the emission of these pollutants. Furthermore, the deterioration of road surfaces caused by these vehicles also plays a role in the elevated levels of particulate matter. In contrast, O_3 concentrations display an upward trend between 9 am and 6 pm. This phenomenon can be attributed to the presence of solar radiation, as O_3 formation is facilitated by sunlight, so it is expected that O_3 concentrations would rise during daylight hours.



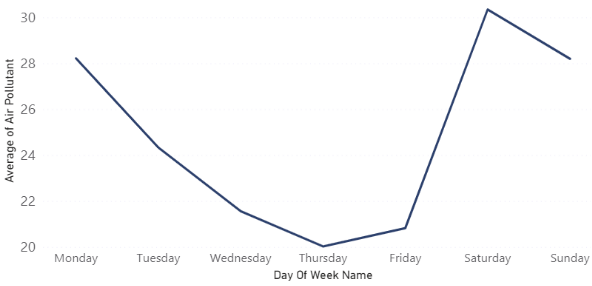
a) CO



b) NO_2



c) O_3



d) PM_{10}

Figure 5.4 – Pollutants' weekly concentrations a) CO, b) NO_2, c) O_3, d) PM_{10}

Regarding the different days of the week, in Figure 5.4, it has been observed that the concentrations of CO and NO_2 pollutants tend to be higher during weekdays, specifically during periods when residents are working and use motor vehicles for commuting between their residences and workplaces. This finding suggests that the elevated levels of these pollutants during weekdays are likely attributable to the combustion of fossil fuels, as concluded in the previous analysis. In contrast, the concentrations of O_3 and PM pollutants exhibit higher levels during weekends. Regarding O_3, one possible explanation for this phenomenon is the reduced industrial activity and traffic typically observed during weekends. Industrial activities and vehicles often emit pollutants that contribute to the depletion of O_3 in the atmosphere, thus reducing its concentrations. With PM, higher concentrations during weekends may be influenced by outliers present in the variable measurements, which will be further analyzed.

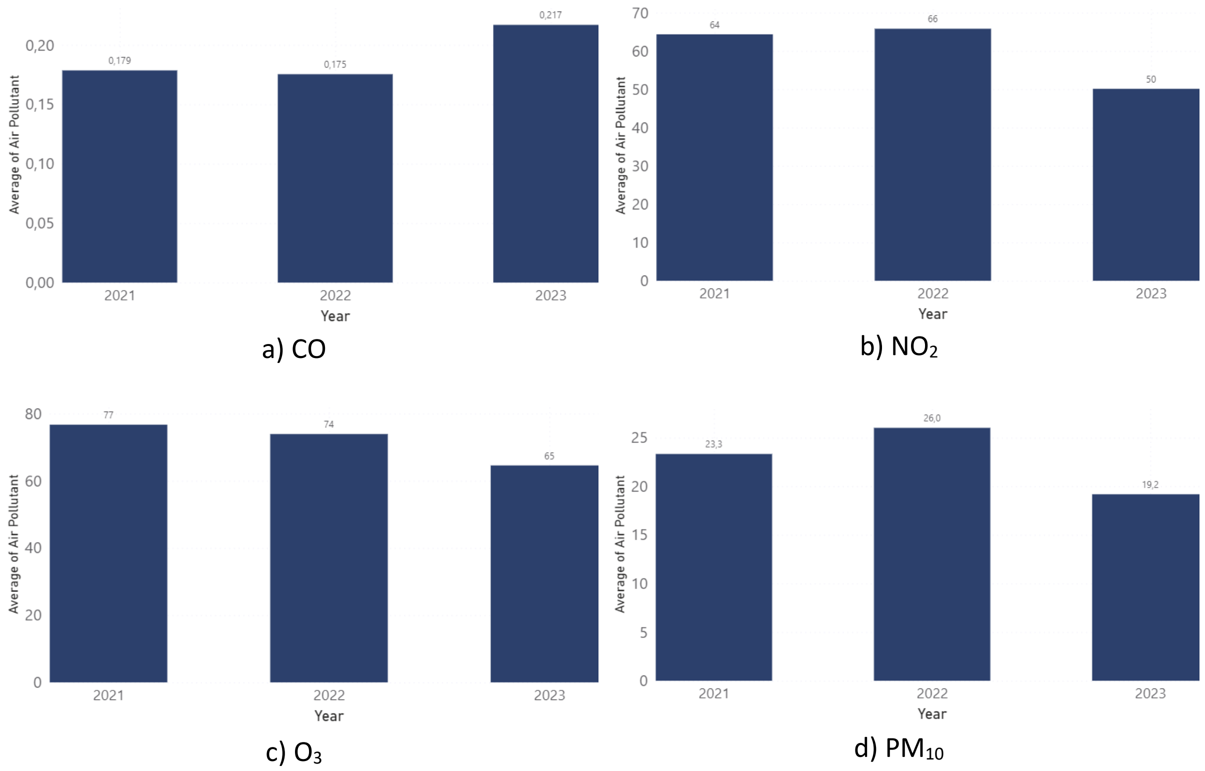
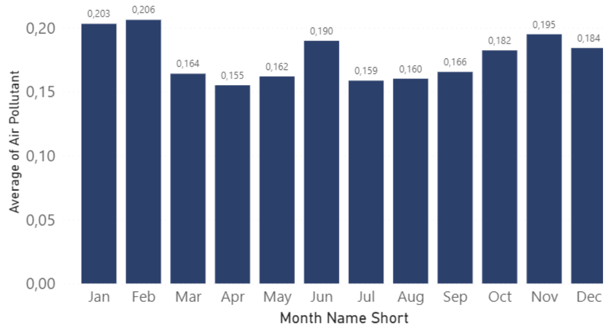
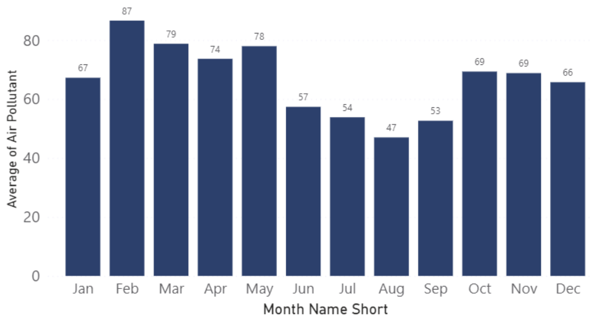


Figure 5.5 – Pollutants' yearly concentrations a) CO, b) NO_2, c) O_3, d) PM_{10}

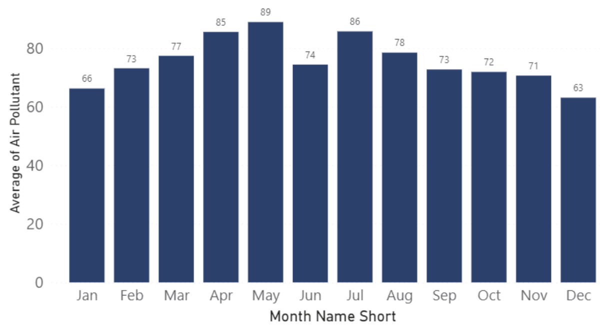
In the annual analysis, in Figure 5.5, only CO consistently showed high concentrations since 2021. The other variables, had similar values between 2021 and 2022, with a decrease in 2023. However, the analysis of 2023 is based on limited data. Concentrations of PM, stood out with significantly higher values in 2022, likely due to the occurrence of specific events or incidents that led to exceptionally high values on specific dates during that year. As stated before, further analysis is needed to understand these events and their implications.



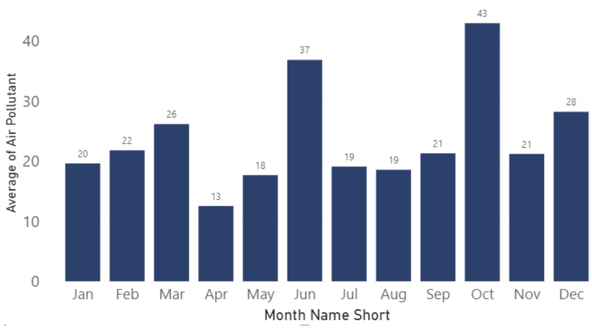
a) CO



b) NO_2



c) O_3



d) PM_{10}

Figure 5.6 – Pollutants' monthly concentrations a) CO, b) NO_2, c) O_3, d) PM_{10}

Upon analyzing each quarter and month in greater detail, in Figure 5.6, it was observed that concentrations of CO and NO_2 decreased between May and October. This coincides with the holiday period, when many residents travel outside of Lisbon, resulting in a reduced number of vehicles on the roads and thus lower combustion of fossil fuels that contribute to the production of these pollutants. In contrast, O_3 concentrations are higher during the same period. This may be because from May to October, days have longer periods of sunlight, which aids in the production of O_3. As for PM, as previously mentioned, no justifiable pattern can be observed due to the presence of outliers that require further analysis.

It is important to note that, the exclusion of colored thresholds in the figures above is intentional to enhance the interpretability of the different patterns. Additionally, the graphs for $PM_{2.5}$ were omitted since they exhibited an almost identical pattern to PM_{10}.

Understanding these temporal variations in pollutant levels provides valuable insights into the sources and drivers of air pollution in Lisbon, enabling the formulation of targeted interventions and policies to mitigate its adverse effects. For example, higher concentrations of pollutants such as CO, NO_2, PM_{10}, and $PM_{2.5}$ during peak hours of commuting, weekdays, and non-vacation months, suggest the need for the implementation of measures to reduce traffic congestion, promote public transportation, and encourage alternative modes of transport to improve air quality during these periods.

5.3. Air Quality Overview

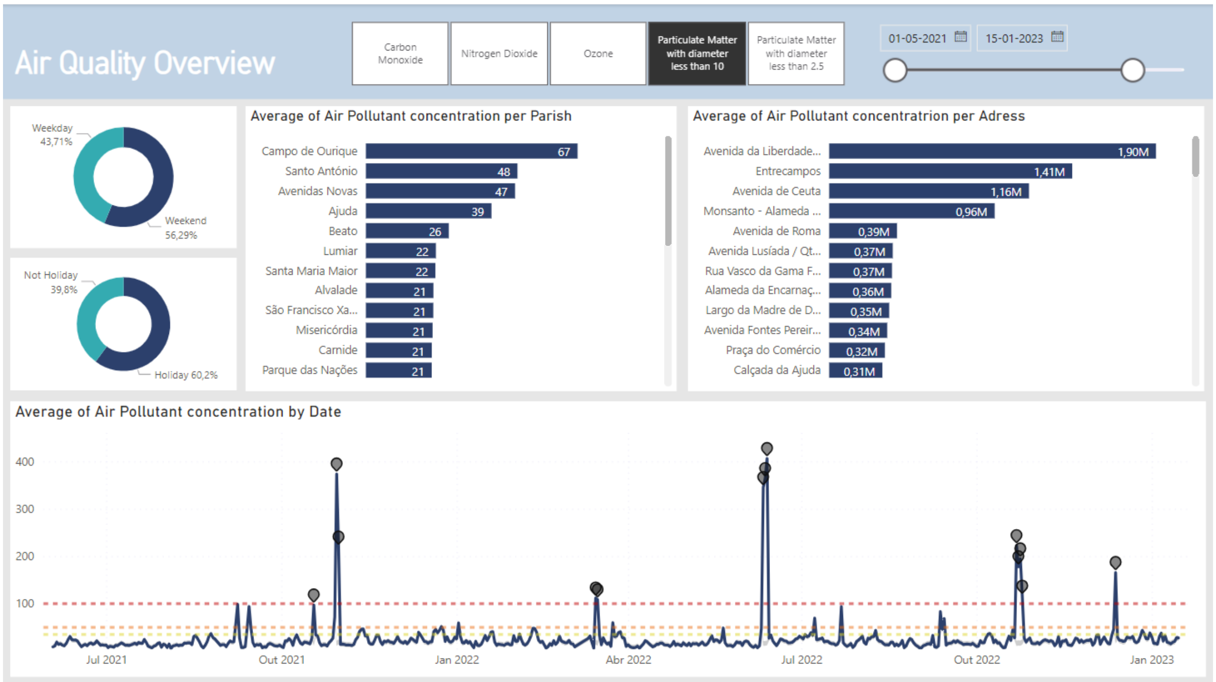


Figure 5.7 – Air Quality Overview Dashboard

The third tab, titled "Air Quality Overview", illustrated in Figure 5.7, provides a comprehensive analysis of the air quality measurements for the entire timeseries, aiming to uncover underlying trends and patterns. This dashboard offers users the ability to select and analyze specific air pollutants and time periods in two different slicers. Additionally, filtering can be performed by interacting with individual graphs, allowing for the selection of specific days, such as holidays or weekdays, to further refine the analysis. Within this dashboard, detailed information is available on the air quality status for each pollutant, at both the parish and street levels. This enables a focused examination of the geographic areas with the most concerning air quality conditions. The two pie charts on the left side depict the concentration of pollutants on different days, highlighting the variations between holidays and normal days, as well as weekends and working days. At the bottom of the dashboard, the selected air pollutant's behavior over time is displayed, allowing users to observe the overall trend in air quality. Anomaly points are highlighted to indicate days when the average value deviates from the expected range. To facilitate a quick and thorough analysis, colored lines representing the various environmental index limits are also incorporated, aiding in the identification of air quality conditions.

Upon closer examination of the graphs that show the patterns of the air quality metrics in more detail, it is possible to see that only particulate matter is influenced by weekends and holidays, with higher values on these days. However, these values may be influenced by the outliers present in these variables, as will be shown below in Figure 5.11.

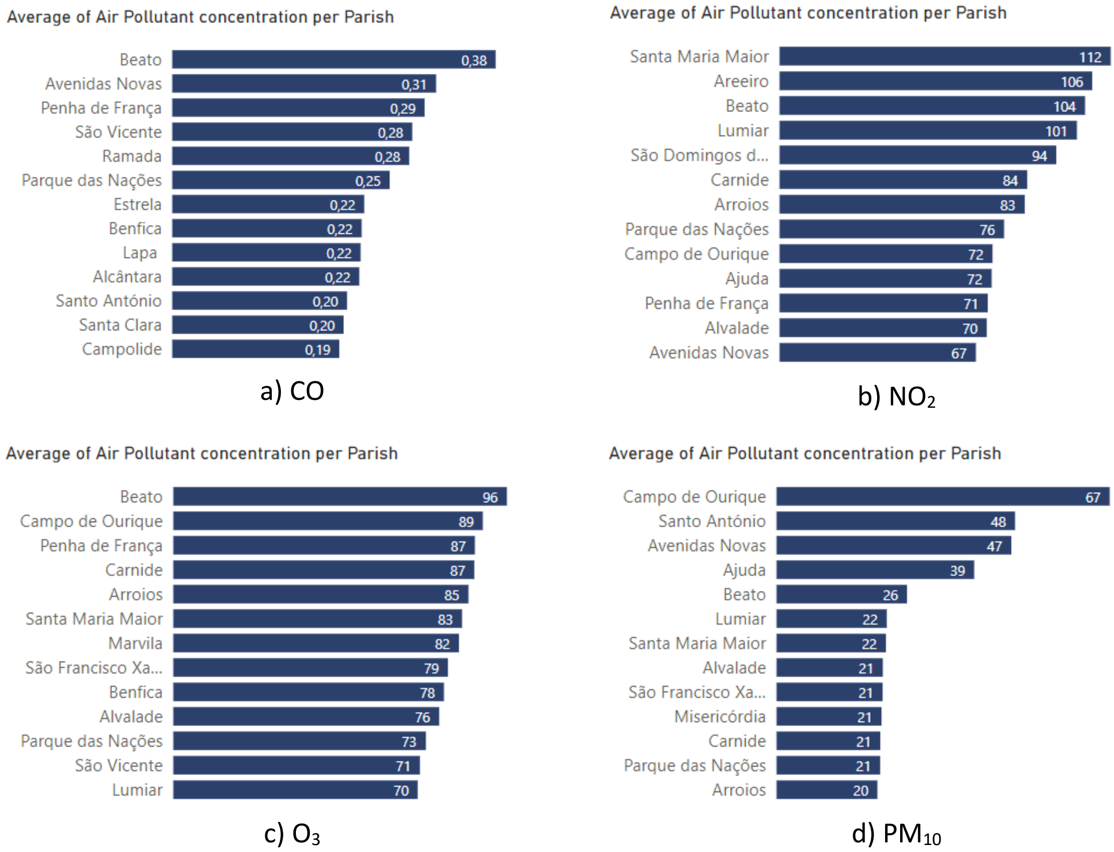


Figure 5.8 – Pollutants' concentrations per Parish a) CO, b) NO_2, c) O_3, d) PM_{10}

By analyzing the concentrations of various pollutants across different parishes in Figure 5.8, it becomes evident that "Beato" generally exhibits the highest concentration of air pollutants, except for PM, where "Campo de Ourique" and "Santo António" emerge as the locations with the highest concentrations for this pollutant.

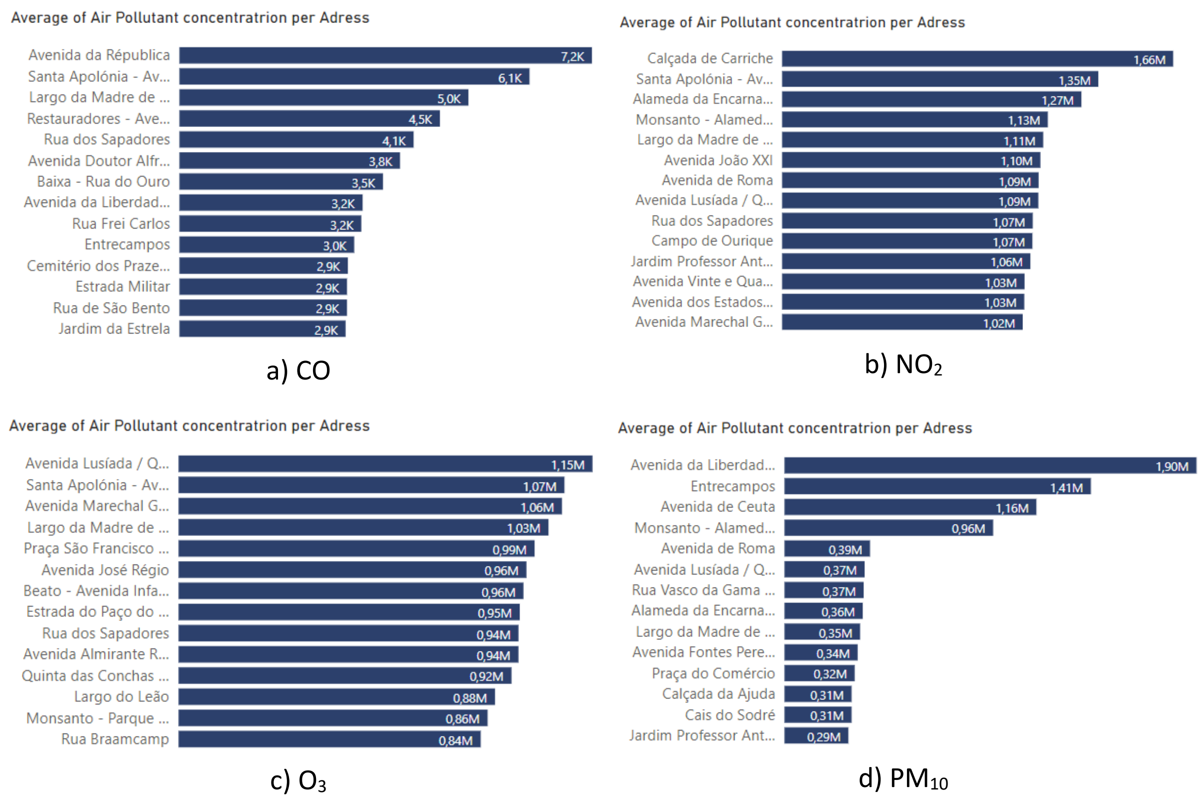
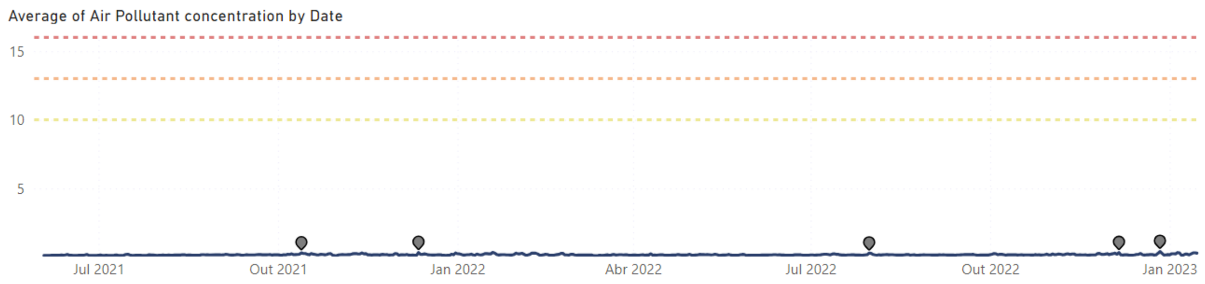


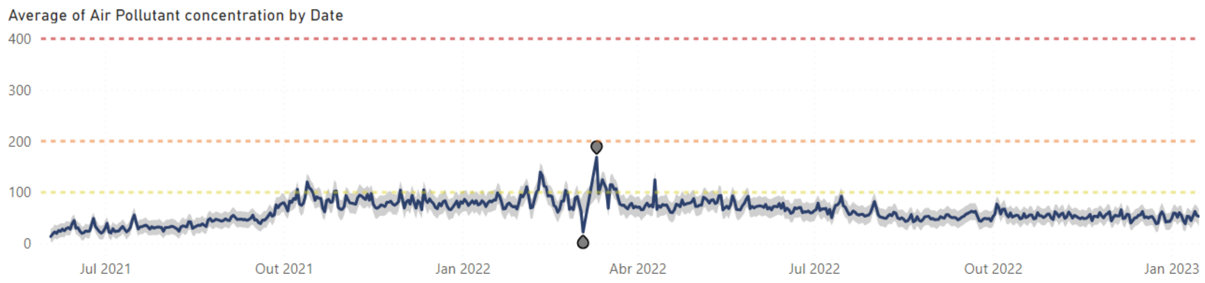
Figure 5.9 – Pollutants' concentrations per Address a) CO, b) NO_2, c) O_3, d) PM_{10}

Specifically, the addresses that notably stand out in Figure 5.9 are the main avenues of Lisbon, including "Avenida da República", "Calçada de Carriche", "Avenida Lusíada", and "Avenida da Liberdade", as each one demonstrate the highest concentrations for each respective pollutant. This observation can be attributed to the significant vehicular traffic in these areas, as these streets and parishes serve as major transportation routes in Lisbon, thereby contributing to the emission of these pollutants.

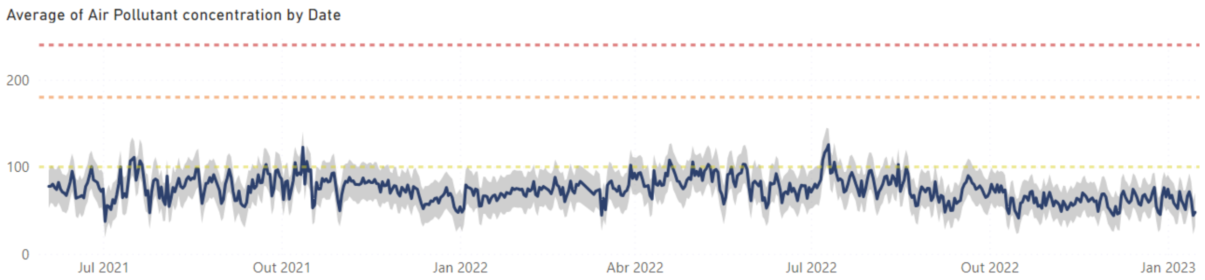
It is important to note that, like in the previous section, the graphs illustrated above for $PM_{2.5}$ were omitted since they exhibited an almost identical pattern to PM_{10}.



a) CO



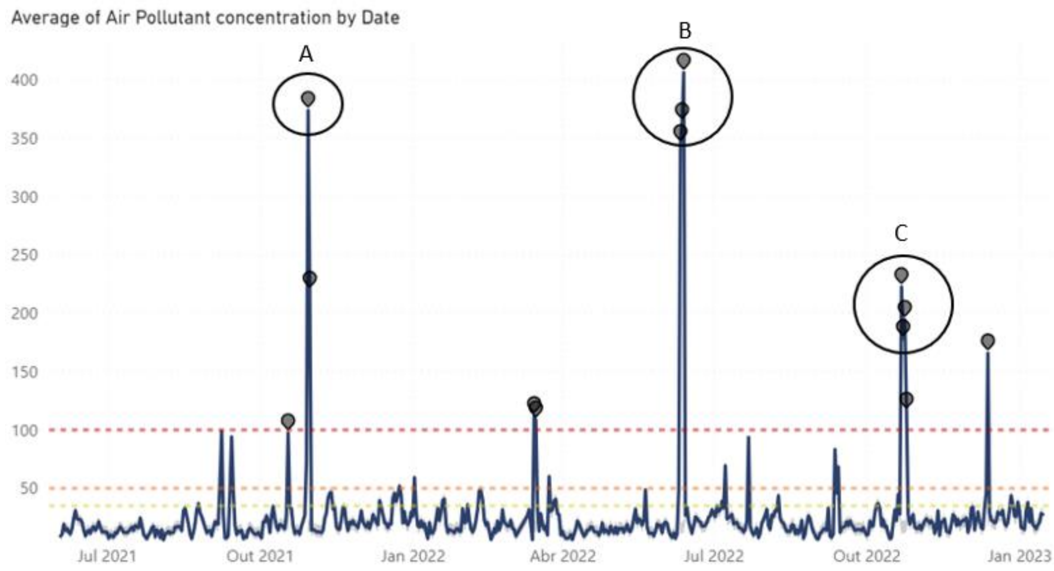
b) NO_2



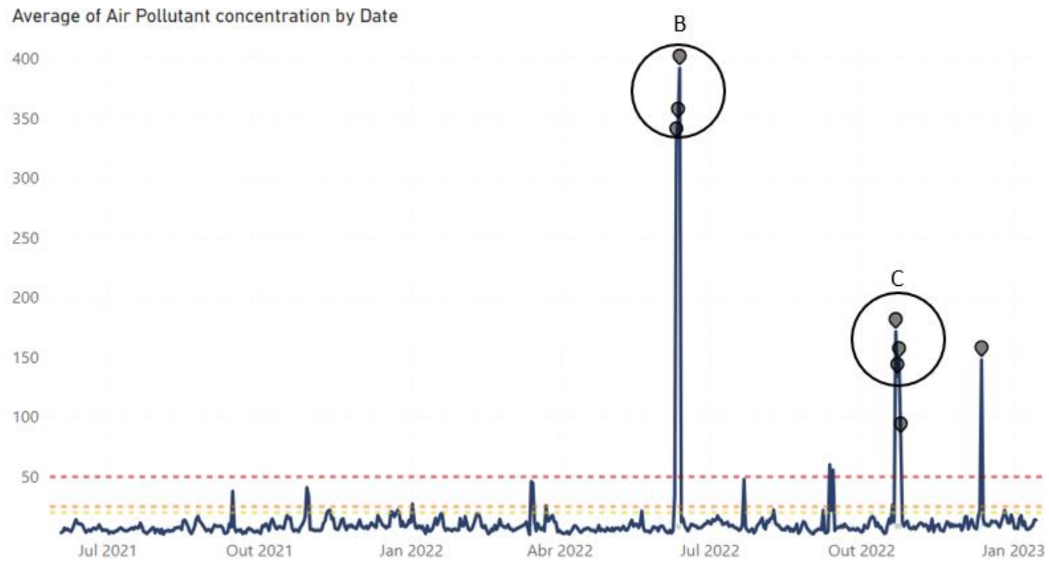
c) O_3

Figure 5.10 – Pollutants' trends throughout time a) CO, b) NO_2, c) O_3, d) PM_{10}

When conducting a detailed analysis of each pollutant and their daily average values over 2 and a half years, in Figure 5.10, several noteworthy observations were made. In the case of CO, the recorded values do not deviate significantly from the defined environmental index limits, indicating no immediate environmental concerns for the city of Lisbon. However, it is worth noting that carbon monoxide concentrations have shown a slight upward trend since concentrations were first recorded for this pollutant, which might be explained by the end of COVID-19 pandemic. Concentrations of NO_2, on the other hand, frequently surpasses the yellow environmental index limit, particularly in the early months of 2022. However, these values have stabilized and remained within the normal environmental limit in subsequent periods. Similarly, O_3, repeatedly shows values exceeding the yellow environmental index limit, although no specific annual pattern is observed. Notably, O_3 concentrations have experienced a slight decline since 2021. In the case of Particulate Matter, both PM_{10} and $PM_{2.5}$, there are instances of exceptionally high peaks occurring for three to four consecutive days at various times over the years. Unfortunately, these concentration values significantly surpass the red environmental index limit, indicating a pressing need for special attention.



d) PM_{10}



d) $PM_{2.5}$

Figure 5.11 – Pollutants' concerning peaks since 2021 a) PM_{10}, b) $PM_{2.5}$

In the case of PM_{10}, in Figure 5.11, three specific dates stand out: October 30th, 2021 (point A), June 13th, 2022 (point B) and October 22nd, 2022 (point C), represented by the three circles. Similar patterns are observed for the $PM_{2.5}$ variable, except for the first date, marked as point A. The first date, in point A, coincides with the eruption of the *Cumbre Vieja* volcano in the Canary Islands[5] that started few days before. The resulting smoke plumes from the volcanic eruption could have potentially affected the atmospheric conditions in Portugal. In June (APA, 2022a) and October (APA, 2022b) 2022, the elevated concentrations of PM_{10} and $PM_{2.5}$ could be attributed to phenomena related to the transport of air masses originating from the deserts of North Africa, carrying dust. These events, influenced by atmospheric patterns, may have contributed to the observed increase in pollutant levels. It is

[5] https://www.nationalgeographic.pt/meio-ambiente/o-pulso-do-vulcao-palma_2929

important to note that the remaining peaks, which were not specifically analyzed, are likely associated with similar events and atmospheric phenomena.

These findings highlight the impact of external factors, such as volcanic eruptions and the transport of air masses, on the concentrations of PM_{10} and $PM_{2.5}$ in Lisbon. Understanding these events and their implications is crucial for developing effective strategies to mitigate the adverse effects of air pollution in the city.

5.4. AIR QUALITY VS TRAFFIC DASHBOARD

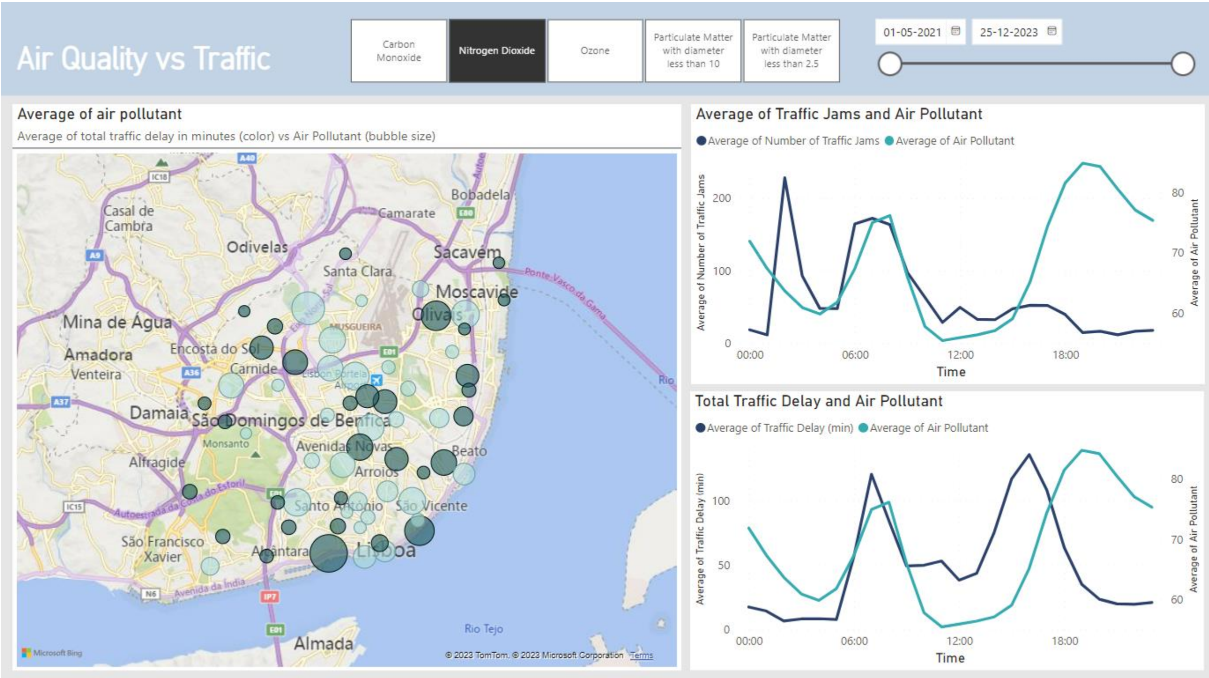
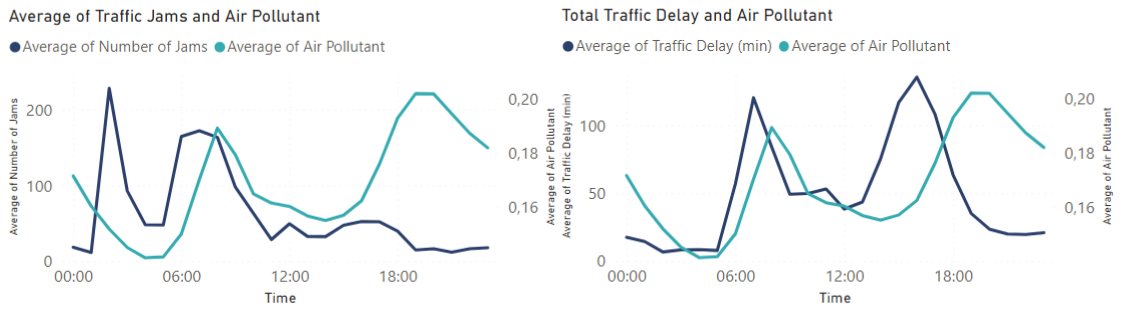


Figure 5.12 – Air Quality vs Traffic Dashboard

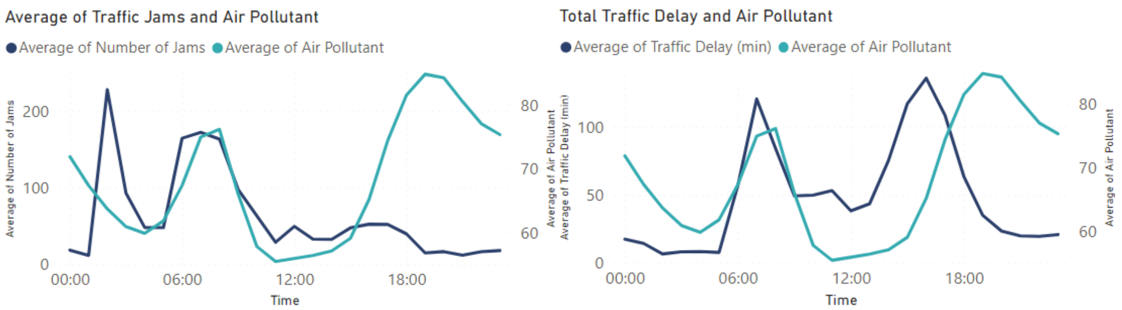
The fourth dashboard, titled " Air Quality vs Traffic Dashboard", illustrated in Figure 5.12, serves as a comprehensive tool to examine the influence of traffic on each pollutant. Similar to the previous dashboard, users can filter the graphs based on the desired pollutant and time period for analysis. Additionally, direct interaction with the charts is also possible, enhancing the flexibility of data exploration. On the left side of the dashboard, the concentration of pollutants, represented by the bubble size, is compared with the total delay, represented by color gradient, where light blue and dark blue represent lower and higher values of traffic delay, respectively. This provides insights into the relationship between pollutant levels and traffic congestion across different locations. On the right side, two line graphs depict the variations in pollutant concentrations in contrast to traffic jams and traffic delays at various times of the day. This facilitates a deeper understanding of how traffic conditions impact the levels of pollutants in the air.

It is important to note that this dashboard exclusively analyzed data from June 1st, 2021, to September 22nd, 2022. This time frame was chosen because it represents the period during which both data sources coexisted, allowing for a comprehensive analysis and comparison of the available information.

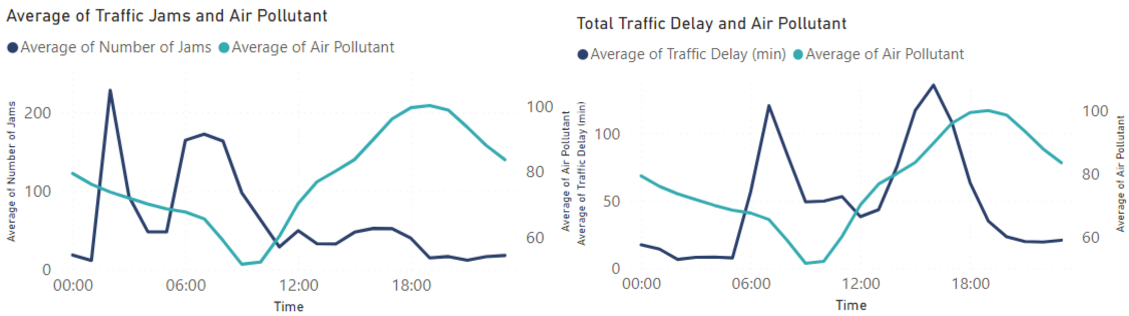
When analyzing the map, the correlation between the size and the color of the bubbles seems evident, however, it has to be evaluated in more detail. Still, this implies that areas with higher concentrations of air pollutants also exhibit longer traffic delays. This correlation can be attributed to the fact that increased traffic delays are indicative of higher traffic volumes, resulting in a greater number of vehicles on the road that contribute to pollutant emissions through the combustion of fossil fuels.



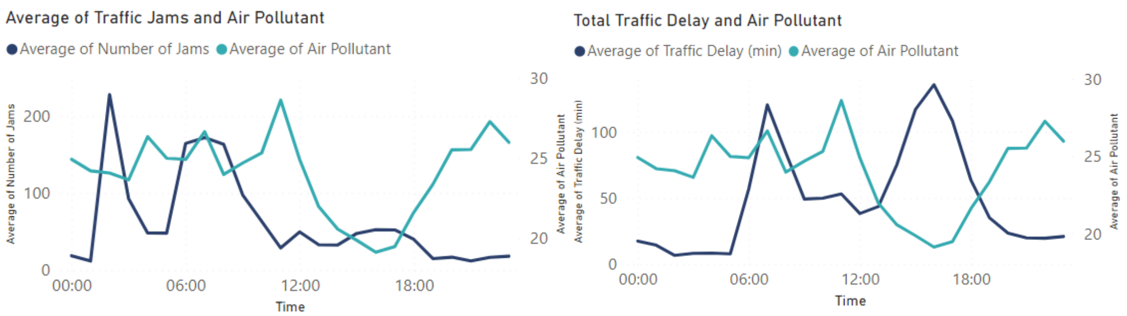
a) CO



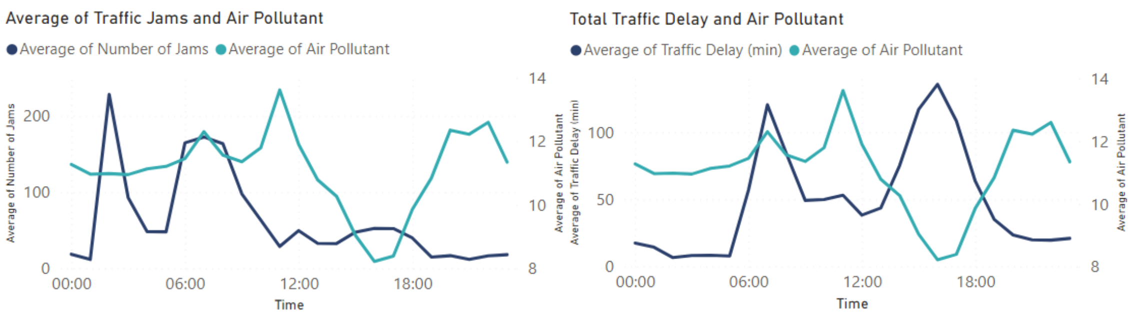
b) NO_2



c) O_3



d) PM_{10}



d) $PM_{2.5}$

Figure 5.13 – Pollutants' concentrations vs Traffic a) CO, b) NO_2, c) O_3, d) PM_{10}, d) $PM_{2.5}$

Considering the line graphs, in Figure 5.13, CO and NO_2 exhibit a positive correlation with the number of traffic jams and traffic delays. The observed correlation can be attributed to the increased traffic volume during peak hours, when there is a frequent stop-and-go movement of vehicles, leading to an extended duration of pollutant emissions. Regarding O_3, a negative correlation is observed during certain periods of the day, particularly between 7 am and 11 am. This decline in O_3 concentrations during these periods may be attributed to the increased production of pollutants resulting from the combustion of fossil fuels, which can react with O_3, leading to its breakdown and subsequent reduction in O_3 concentrations, as mentioned previously. Similarly, both variables of particulate matter display a negative correlation with traffic jams and traffic delays. This can be attributed to the stop-and-go driving patterns associated with traffic, which may facilitate the dispersion of particulate matter due to increased mixing with the surrounding air. Furthermore, the erosion of road surfaces, which contributes to the production of particulate matter, is reduced when vehicles operate at lower speeds during traffic congestion.

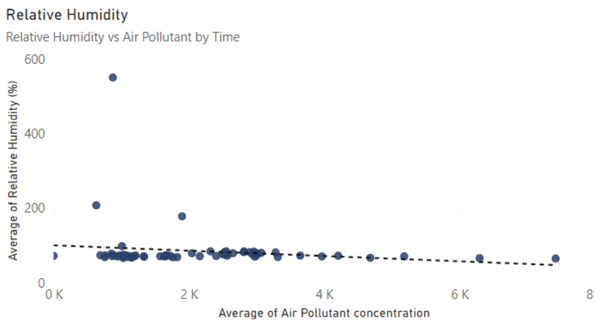
It is imperative to highlight that there is a noteworthy occurrence of traffic congestion reaching its highest levels at 2 a.m., that is consistent throughout the measurement period. It is plausible that these values might be attributed to measurement inaccuracies, as it is not conventionally anticipated for a substantial volume of vehicles to be in circulation during the early hours of dawn. Conversely, one would expect a surge in the number of traffic jams between 5 and 7 p.m., which corresponds to the rush hour in Lisbon when residents typically commute home from work. Surprisingly, this anticipated increase in traffic congestion during the aforementioned period does not occur. Therefore, both scenarios significantly undermine the validity of the parameter measurements and warrant a thorough and comprehensive analysis to ascertain the underlying factors contributing to these observations.

By analyzing the relationship between traffic patterns and pollutant concentrations, policymakers and stakeholders in Lisbon can gain valuable insights to guide the development of targeted interventions and strategies for mitigating air pollution. Understanding how traffic dynamics influence pollutant levels can help identify high-pollution areas and prioritize measures such as traffic management, promoting alternative transportation modes, and implementing emission reduction initiatives. This comprehensive approach acknowledges the significant impact of traffic on air quality and empowers decision-makers to take effective action in addressing air pollution in Lisbon.

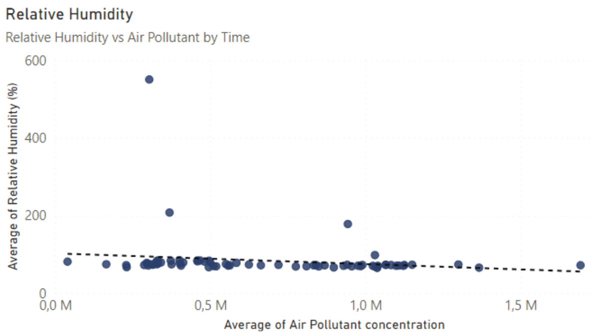
5.5. AIR QUALITY VS WEATHER

Figure 5.14 – Air Quality vs Weather Dashboard

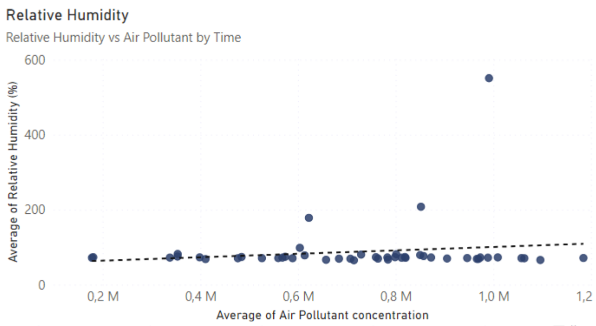
The fifth and final dashboard, named "Air Quality vs Weather," illustrated in Figure 5.14, aims to examine the influence of meteorological conditions on the concentrations of the various pollutants. Similar to the previous dashboard, users can filter the graphs based on the specific pollutant and time period to analyze. To depict the correlation between several weather factors and pollutant concentration, four scatter charts have been created, containing trend lines to facilitate the interpretation. However, it is important to note that the variable "wind intensity" had to be excluded from this analysis due to limited data availability, as it was measured by only two sensors.



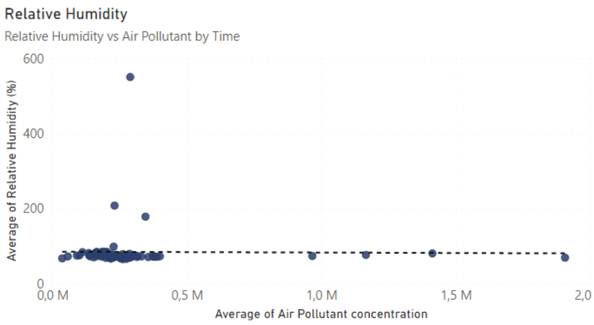
a) CO



b) NO_2



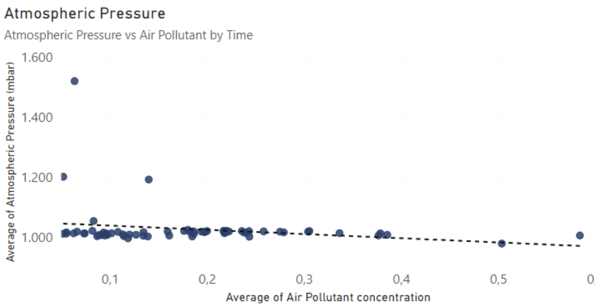
c) O_3



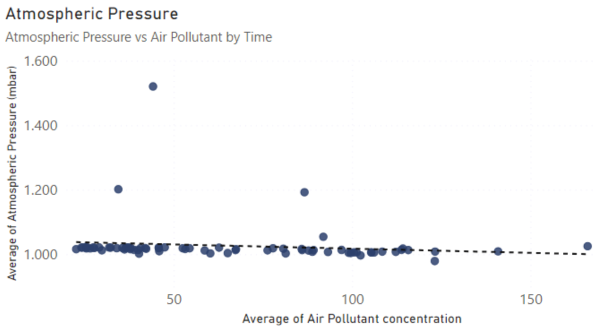
d) PM_{10}

Figure 5.15 – Pollutants' concentrations vs Relative Humidity a) CO, b) NO_2, c) O_3, d) PM_{10}

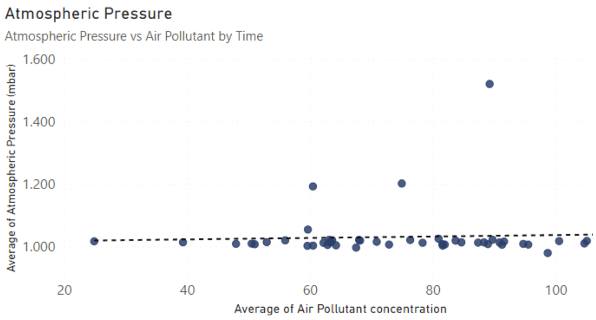
When analyzing the various pollutants in relation to different meteorological variables, several noteworthy conclusions were drawn. In terms of relative humidity, in Figure 5.15, a slight positive correlation can be observed with CO and NO_2, while a slight negative correlation is observed with O_3. Regarding NO_2, higher levels of moisture in the atmosphere help the formation of other nitrogen-containing compounds, leading to a reduction in NO_2 concentrations. Similarly, for CO, increased humidity promotes the formation of hydroxyl radicals that contribute to break down this chemical, resulting in lower concentrations of this pollutant in the atmosphere. In contrast, humidity has a positive influence on ozone concentrations, due to the fact that moist environments facilitate the reaction between water vapor and nitrogen oxides, ultimately promoting the formation of ozone. Particulate matter did not have significant correlations with relative humidity.



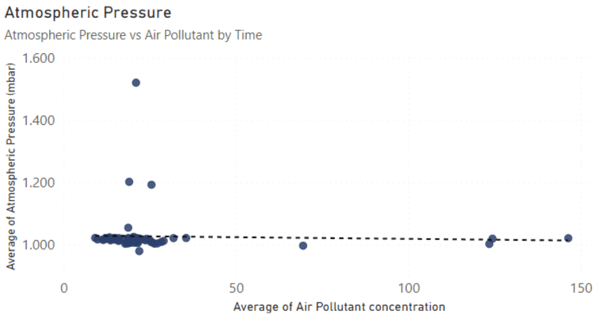
a) CO



b) NO_2



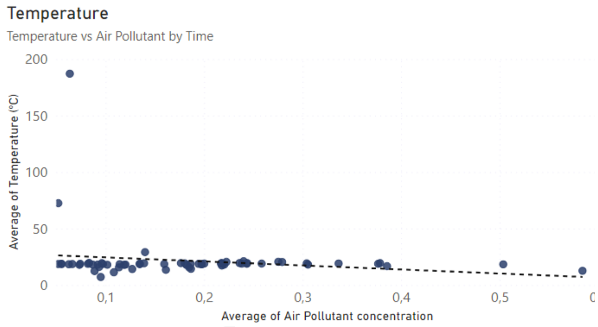
c) O_3



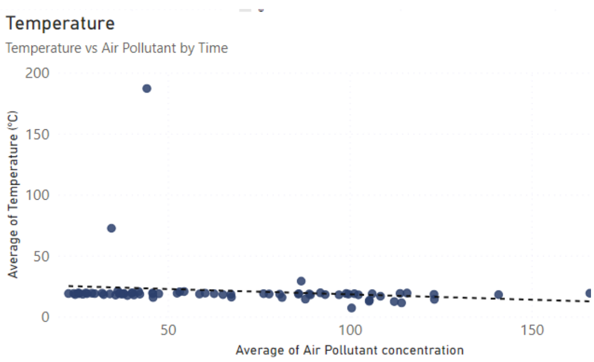
d) PM_{10}

Figure 5.16 – Pollutants' concentrations vs Atmospheric Pressure a) CO, b) NO_2, c) O_3, d) PM_{10}

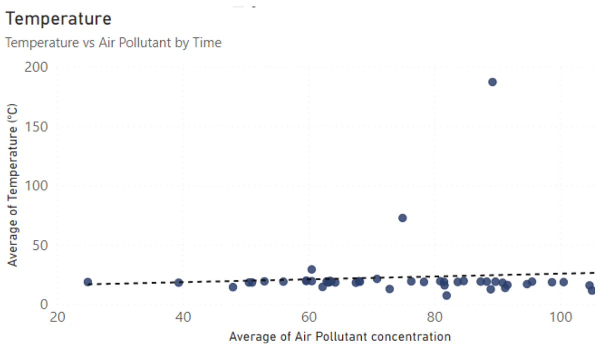
When considering atmospheric pressure, in Figure 5.16, a slight positive correlation was observed between O_3 and this meteorological variable. It should be noted that atmospheric pressure does not have a direct impact on atmospheric gases as it is not directly involved in chemical reactions or physical processes that affect their formation. However, higher atmospheric pressure is often associated with stable weather conditions and reduced mixing of air masses. This limited mixing leads to less dispersion of gases in the atmosphere and, consequently, higher concentrations. Therefore, the positive correlation between O_3 and atmospheric pressure suggests that calmer weather conditions associated with higher pressure contribute to the accumulation of ozone in the atmosphere. The remaining pollutants did not have significant correlations with atmospheric pressure.



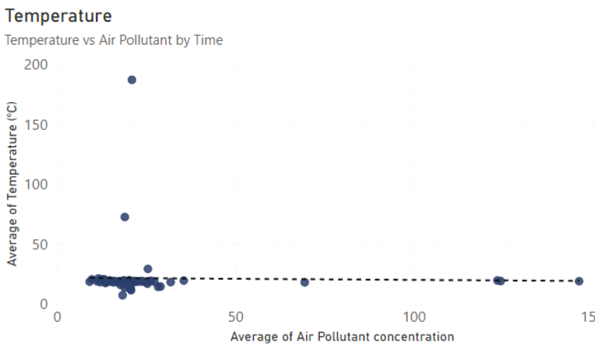
a) CO



b) NO_2



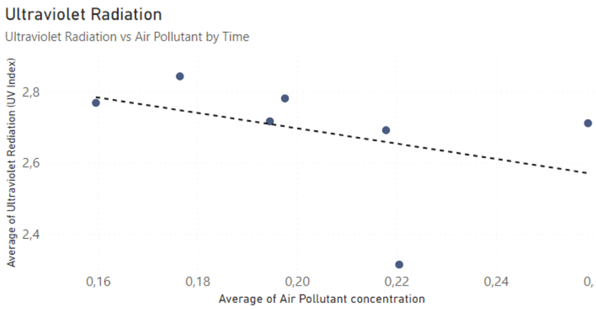
c) O_3



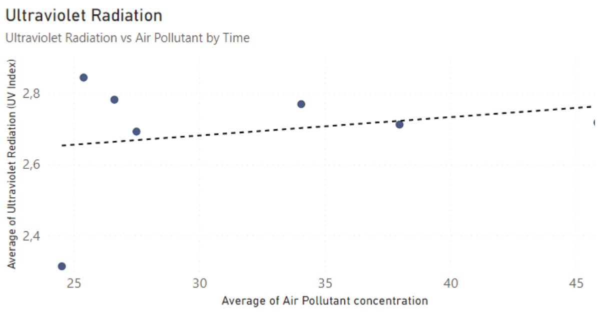
d) PM_{10}

Figure 5.17 – Pollutants' concentrations vs Temperature a) CO, b) NO_2, c) O_3, d) PM_{10}

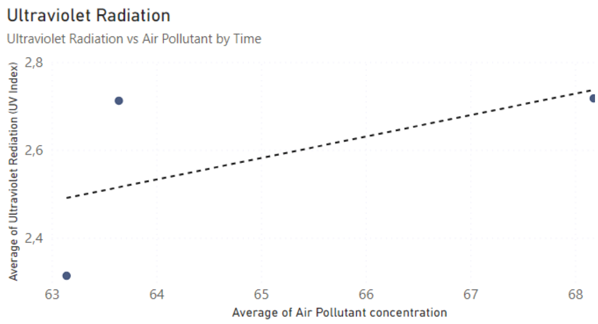
Regarding temperature, in Figure 5.17, it is expected that the production of air pollutants increases, since temperature influences the rate of chemical reactions, influencing the formation of secondary pollutants such as O_3. This can explain the slight positive correlation between this pollutant and temperature. However, temperature can also enhance atmospheric mixing and dilution, resulting in lower concentrations of atmospheric gases. Hence, the slight negative correlation between temperature and CO and NO_2 can be explained by this phenomenon. Concentrations of PM_{10} did not exhibit significant correlations with temperature.



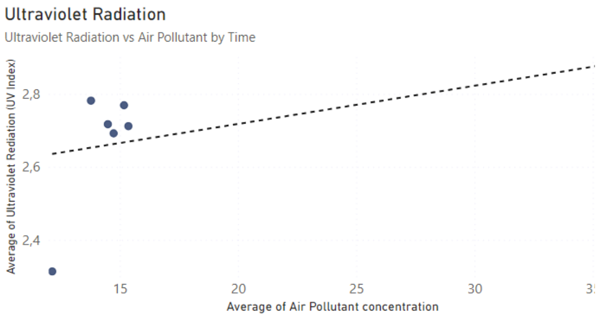
a) CO



b) NO_2



c) O_3



d) PM_{10}

Figure 5.18 – Pollutants' concentrations vs Ultraviolet Radiation a) CO, b) NO_2, c) O_3, d) PM_{10}

Regarding ultraviolet radiation, in Figure 5.18, it does not exert a significant influence on the concentrations of pollutants in the atmosphere, except for O_3. Ultraviolet radiation actively participates in the chemical reactions that facilitate the formation of O_3. Therefore, the positive correlation between ozone and ultraviolet radiation can be caused by this phenomenon. Given that it is not anticipated for Ultraviolet Radiation to have no impact on the formation of the remaining pollutants, the effect of this environmental parameter on their concentration has not been highlighted.

Nevertheless, it is crucial to emphasize that the measurements pertaining to Ultraviolet Radiation were significantly restricted, having few sensors measuring this environmental parameter. Therefore, it becomes imperative to do more measurements encompassing a greater number of locations. This is essential to ensure that the conclusions drawn from the data are precise, and capable of providing an accurate representation of the prevailing Ultraviolet radiation levels.

Overall, the plots revealed the presence of certain outliers that have the potential to introduce bias to the results and subsequent conclusions drawn from this analysis. Consequently, it is important for future investigations to carefully scrutinize the data and, if necessary, exclude these outliers to determine the robustness and validity of the conclusions made.

Furthermore, the measurements pertaining to $PM_{2.5}$ were omitted from the plots as they exhibited patterns similar to those observed for PM_{10}.

The city of Lisbon can use the findings from this dashboard to develop strategies and interventions to combat air pollution based on weather conditions. By monitoring meteorological conditions alongside air pollutant concentrations, the city can effectively manage air quality and implement targeted measures. This comprehensive approach allows for a better understanding of the factors influencing pollutant levels and enables initiative-taking interventions to reduce pollution.

6. CONCLUSIONS AND FUTURE WORK

The implemented BI solution has provided an extensive overview of air quality in Lisbon, encompassing a detailed analysis of its trends and patterns over the past 2 and a half years. It is important to note that this solution can be extended beyond Lisbon and adapted for application in other cities within Portugal or even globally, certainly with adjustments in the data sources and relational models. By employing this BI solution, the city of Lisbon will have a valuable and user-friendly tool for analyzing air quality conditions. This tool enables the identification of pollution hotspots that exceed acceptable limits for public health, as well as the comprehension of air quality trends and its correlation with traffic and weather, supporting data-driven decision-making. Local authorities can access the most up-to-date information regarding current air quality conditions, facilitating the development of proactive measures aimed at preventing prolonged exposure to harmful pollutant concentrations. This may involve the creation of contingency plans or the implementation of more efficient traffic management strategies. Furthermore, the historical data analysis and trend evaluation provided by this solution empower authorities to assess the impact and effectiveness of implemented measures in the past and make data-informed adjustments, optimizing the allocation of strategies and resources in areas that require immediate attention. Additionally, transparent, and accessible communication of the current air quality status can raise awareness among local authorities and promote a collaborative approach with the residents of Lisbon, working towards improving air quality in the city.

During the development of this solution, several limitations were encountered. Firstly, due to missing data, it was not possible to analyze all the collected environmental parameters, including air pollutants such as Benzene and Sulfur Dioxide. Additionally, the analysis of certain parameters, like wind intensity, was constrained by the limited number of sensors available to measure this parameter. Another limitation refers to the timeframe of the analyzed data, which only covers the year 2021 and onwards, which coincides with the period after the outbreak of the COVID-19 pandemic. As the lockdown measures implemented during this period had a significant impact on air quality, including reduced transportation, industrial activities, agricultural production, among others, it would be valuable to compare this data with pre-pandemic data. Such a comparison would provide insights into the long-term trends and determine whether air quality in the city has been improving or deteriorating. Furthermore, the correlation analysis between air pollution and traffic data was limited to a period when both datasets had concurrent measurements. This constraint resulted in a restricted time window for analysis and limited the ability to explore the long-term relationship between air pollution and traffic dynamics. With respect to the measurements of variables, both traffic and weather data exhibited discrepancies. The number of jams variable in the traffic data displayed an abnormal pattern deviating from the expected hourly traffic behavior. Furthermore, the scatter plot between weather data and air pollutants showed some outliers. Both of these events significantly undermine the validity of the measurements taken and consequently call into question the reliability of the conclusions drawn from the analyses. Moreover, the method of data importation via a JSON file prevented the implementation of a scheduled automatic update for real-time data on the dashboard, after publishing the dashboards in Power BI online. This prevented the display of the latest average hourly measurements on the real-time data dashboard and the addition of new measurements to the historical data dashboards.

In terms of future developments, to enhance the utility of this solution, several improvements can be made. Firstly, the collection of data could be enriched with the installation of additional sensors in more geographical points of the city. Furthermore, improvements to existing sensors should be pursued to enable the measurement of a wider range of environmental parameters, thus providing a more comprehensive understanding of air quality in all locations. Furthermore, conducting a more comprehensive analysis of the traffic data is necessary to provide further support and understanding for the unexpected events observed in the analysis. Also, it would be of interest to investigate the influence of weather conditions on air pollutants by conducting an analysis that excludes the observed outliers in the scatter plots. Moreover, incorporating predictive algorithms would be beneficial to enable the prediction of air quality in Lisbon. This predictive capability would assist authorities in constructing preventive measures to address potential air quality issues proactively. Additionally, expanding the scope of data analysis to include variables that impact air quality, such as industrial production, agricultural activities, electricity generation, city topography, would be valuable. This broader analysis would facilitate a more comprehensive understanding of the external factors influencing air quality, thereby enabling authorities to take appropriate actions based on these insights. Finally, the implementation of automatic updates for the dashboards, ensuring that new data measured by the sensors is integrated, would ensure that the dashboards always display the most up-to-date information available. This timely and accurate data presentation is crucial for the fulfillment the intended purpose of the dashboards.

BIBLIOGRAPHICAL REFERENCES

Africk, E., & Levy, Y. (2021). An examination of historic data breach incidents: What cybersecurity big data visualization and analytics can tell us? *Online Journal of Applied Knowledge Management (OJAKM)*, *9*(1), 31–45. https://doi.org/10.36965/OJAKM.2021.9(1)31-45

APA. (2022a). *Previsão de transporte de partículas naturais com origem em regiões áridas*. https://qualar.apambiente.pt/download/documentos.ficheiro.aa0cff9f5415ee63.UFJFVklTQU9fRU5fMjAyMl8wNl8xMy5wZGY.pdf

APA. (2022b). *Previsão de transporte de partículas naturais com origem em regiões áridas*. https://qualar.apambiente.pt/download/documentos.ficheiro.8b1827ce41d218fe.UFJFVklTQU9fRU5fMjAyMl8xMF8yNi5wZGY.pdf

Aragona, B., & Rosa, R. D. (2018). Big data in policy making. *Mathematical Population Studies*. https://www.tandfonline.com/doi/full/10.1080/08898480.2017.1418113

Baralis, E., Cerquitelli, T., Chiusano, S., Garza, P., & Kavoosifar, M. R. (2016). Analyzing air pollution on the urban environment. *2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 1464–1469. https://doi.org/10.1109/MIPRO.2016.7522370

Bibri, S. E. (2018). The IoT for smart sustainable cities of the future: An analytical framework for sensor-based big data applications for environmental sustainability. *Sustainable Cities and Society*, *38*, 230–253. https://doi.org/10.1016/j.scs.2017.12.034

Bibri, S. E., & Krogstie, J. (2017). Smart sustainable cities of the future: An extensive interdisciplinary literature review. *Sustainable Cities and Society*, *31*, 183–212. https://doi.org/10.1016/j.scs.2017.02.016

Brunekreef, B., & Holgate, S. T. (2002). Air pollution and health. *The Lancet*, *360*(9341), 1233–1242. https://doi.org/10.1016/S0140-6736(02)11274-8

Cai, H., Xu, B., Jiang, L., & Vasilakos, A. V. (2017). IoT-Based Big Data Storage Systems in Cloud Computing: Perspectives and Challenges. *IEEE Internet of Things Journal*, *4*(1), 75–87. https://doi.org/10.1109/JIOT.2016.2619369

Câmara Municipal de Lisboa. (2020). *25 dicas sobre a qualidade do ar*. https://cidadania.lisboa.pt/fileadmin/atualidade/publicacoes_periodicas/ambiente/25_dicas_qualidade_do_ar.pdf

Câmara Municipal de Lisboa. (2021). *Metainformação relativa à Monitorização de Parâmetros Ambientais da Cidade de Lisboa*. https://www.lisboa.pt/fileadmin/informacoes_servicos/user_upload/Metainformacao_relativa_a_Monitorizacao_de_Parametros.pdf

Caragliu, A., Del Bo, C., & Nijkamp, P. (2009). Smart Cities in Europe. *VU University Amsterdam, Faculty of Economics, Business Administration and Econometrics, Serie Research Memoranda*, *18*. https://doi.org/10.1080/10630732.2011.601117

Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An overview of business intelligence technology. *Communications of the ACM*, *54*(8), 88–98. https://doi.org/10.1145/1978542.1978562

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly: Management Information Systems*, *36*(4), 1165–1188. Scopus. https://doi.org/10.2307/41703503

Chourabi, H., Nam, T., Walker, S., Gil-Garcia, J. R., Mellouli, S., Nahon, K., Pardo, T. A., & Scholl, H. J. (2012). Understanding Smart Cities: An Integrative Framework. *2012 45th Hawaii International Conference on System Sciences*, 2289–2297. https://doi.org/10.1109/HICSS.2012.615

Dayal, U., Castellanos, M., Simitsis, A., & Wilkinson, K. (2009). *Data integration flows for Business Intelligence*. 1–11. Scopus. https://doi.org/10.1145/1516360.1516362

Ferreira, F., Tente, H., Torres, P., Cardoso, S., & Palma-Oliveira, J. M. (2000). Air Quality Monitoring and Management in Lisbon. *Environmental Monitoring and Assessment*, *65*(1), 443–450. https://doi.org/10.1023/A:1006433313316

Gordon, T., & Reibman, J. (2000). Cardiovascular toxicity of inhaled ambient particulate matter. *Toxicological Sciences*, *56*(1), 2–4. Scopus. https://doi.org/10.1093/toxsci/56.1.2

Goudie, A. S. (2014). Desert dust and human health disorders. *Environment International*, *63*, 101–113. Scopus. https://doi.org/10.1016/j.envint.2013.10.011

Guarnieri, M., & Balmes, J. R. (2014). Outdoor air pollution and asthma. *The Lancet*, *383*(9928), 1581–1592. https://doi.org/10.1016/S0140-6736(14)60617-6

Höchtl, J., Parycek, P., & Schöllhammer, R. (2016). Big data in the policy cycle: Policy decision making in the digital era. *Journal of Organizational Computing and Electronic Commerce*, *26*(1–2), 147–169. Scopus. https://doi.org/10.1080/10919392.2015.1125187

Inmon, W. H. (2005). *Building the Data Warehouse* (Fourth Edition). Wiley Publishing, Inc.

Jardim, B., Castro Neto, M. de, Alpalhão, N., & Calçada, P. (2022). The daily urban dynamic indicator: Gauging the urban dynamic in Porto during the COVID-19 pandemic. *Sustainable Cities and Society*, *79*, 103714. https://doi.org/10.1016/j.scs.2022.103714

Kampa, M., & Castanas, E. (2008). Human health effects of air pollution. *Environmental Pollution*, *151*(2), 362–367. https://doi.org/10.1016/j.envpol.2007.06.012

Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit* (Third Edition). John Wiley & Sons, Inc.

Kingsy, G. R., Manimegalai, R., Geetha, D. M. S., Rajathi, S., Usha, K., & Raabiathul, B. N. (2016). Air pollution analysis using enhanced K-Means clustering algorithm for real time sensor data. *2016 IEEE Region 10 Conference (TENCON)*, 1945–1949. https://doi.org/10.1109/TENCON.2016.7848362

Lavalle, A., Maté, A., Trujillo, J., & Rizzi, S. (2019). Visualization Requirements for Business Intelligence Analytics: A Goal-Based, Iterative Framework. *2019 IEEE 27th International*

Requirements Engineering Conference (RE), 109–119. https://doi.org/10.1109/RE.2019.00022

Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., & Pozzer, A. (2015). The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature*, *525*(7569), Article 7569. https://doi.org/10.1038/nature15371

Manisalidis, I., Stavropoulou, E., Stavropoulos, A., & Bezirtzoglou, E. (2020). Environmental and Health Impacts of Air Pollution: A Review. *Frontiers in Public Health*, *8*. https://www.frontiersin.org/articles/10.3389/fpubh.2020.00014

Petrini, M., & Pozzebon, M. (2009). Managing sustainability with the support of business intelligence: Integrating socio-environmental indicators and organisational context. *The Journal of Strategic Information Systems*, *18*(4), 178–191. https://doi.org/10.1016/j.jsis.2009.06.001

Ramanathan, V., & Feng, Y. (2009). Air pollution, greenhouse gases and climate change: Global and regional perspectives. *Atmospheric Environment*, *43*(1), 37–50. https://doi.org/10.1016/j.atmosenv.2008.09.063

Rios, L. G., & Diguez, J. A. I. (2014). Big Data Infrastructure for analyzing data generated by Wireless Sensor Networks. *2014 IEEE International Congress on Big Data*, 816–823. https://doi.org/10.1109/BigData.Congress.2014.142

Ritchie, H., & Roser, M. (2017). Air Pollution. *Our World in Data*. https://ourworldindata.org/air-pollution

Samet, J. M., Dominici, F., Curriero, F. C., Coursac, I., & Zeger, S. L. (2000). Fine particulate air pollution and mortality in 20 U.S. cities, 1987-1994. *New England Journal of Medicine*, *343*(24), 1742–1749. Scopus. https://doi.org/10.1056/NEJM200012143432401

Santos, F. M., Gómez-Losada, Á., & Pires, J. C. M. (2019). Impact of the implementation of Lisbon low emission zone on air quality. *Journal of Hazardous Materials*, *365*, 632–641. https://doi.org/10.1016/j.jhazmat.2018.11.061

Shahid, N., Shah, M. A., Khan, A., Maple, C., & Jeon, G. (2021). Towards greener smart cities and road traffic forecasting using air pollution data. *Sustainable Cities and Society*, *72*, 103062. https://doi.org/10.1016/j.scs.2021.103062

Suakanto, S., Supangkat, S. H., Suhardi, & Saragih, R. (2013). Smart city dashboard for integrating various data of sensor networks. *International Conference on ICT for Smart Society*, 1–5. https://doi.org/10.1109/ICTSS.2013.6588063

Taborda, R., Datia, N., Pato, M. P. M., & Pires, J. M. (2020). Exploring air quality using a multiple spatial resolution dashboard—A case study in Lisbon. *2020 24th International Conference Information Visualisation (IV)*, 140–145. https://doi.org/10.1109/IV51561.2020.00032

Tavera Romero, C. A., Ortiz, J. H., Khalaf, O. I., & Ríos Prado, A. (2021). Business Intelligence: Business Evolution after Industry 4.0. *Sustainability*, *13*(18), Article 18. https://doi.org/10.3390/su131810026

WHO. (2021). *WHO global air quality guidelines: Particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*. https://apps.who.int/iris/bitstream/handle/10665/345329/9789240034228-eng.pdf

Xiao, Y., & Ji, C. (2009). Management of Air Quality Monitor Data with Data Warehouse and GIS. *2009 WRI World Congress on Computer Science and Information Engineering*, *4*, 148–152. https://doi.org/10.1109/CSIE.2009.280

Xiaojun, C., Xianpeng, L., & Peng, X. (2015). IOT-based air pollution monitoring and forecasting system. *2015 International Conference on Computer and Computational Sciences (ICCCS)*, 257–260. https://doi.org/10.1109/ICCACS.2015.7361361

Zheng, Y., Liu, F., & Hsieh, H.-P. (2013). *U-Air: When urban air quality inference meets big data*. 1436–1444. https://doi.org/10.1145/2487575.2488188

APPENDIX A

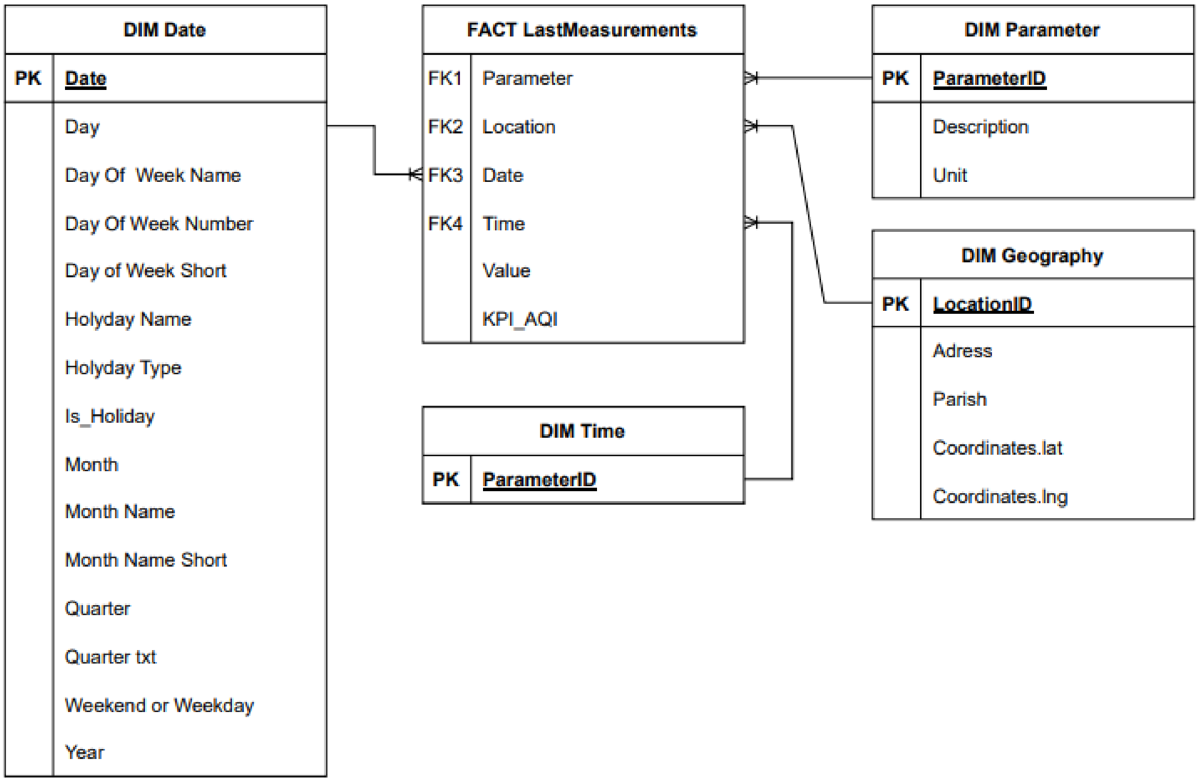


Figure A.1 – Relational Model for *FACT LastMeasurements*

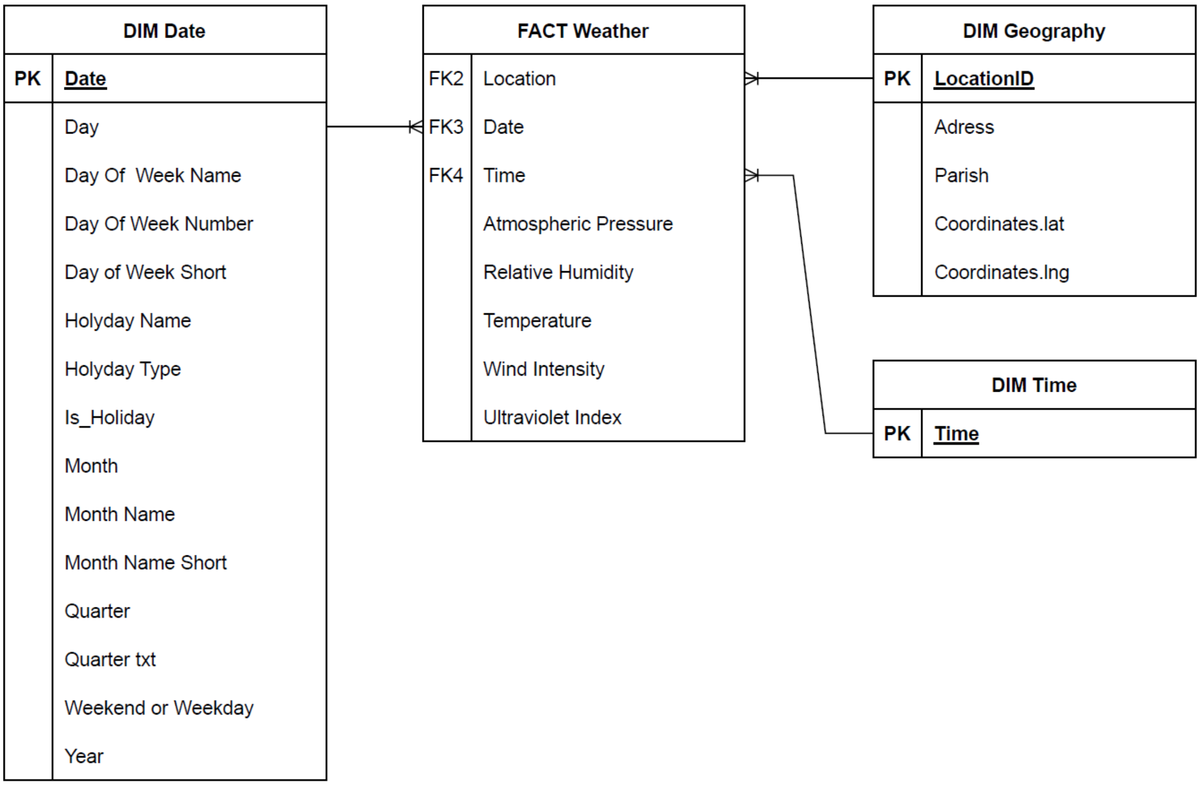


Figure A.2 – Relational Model for *FACT Weather*

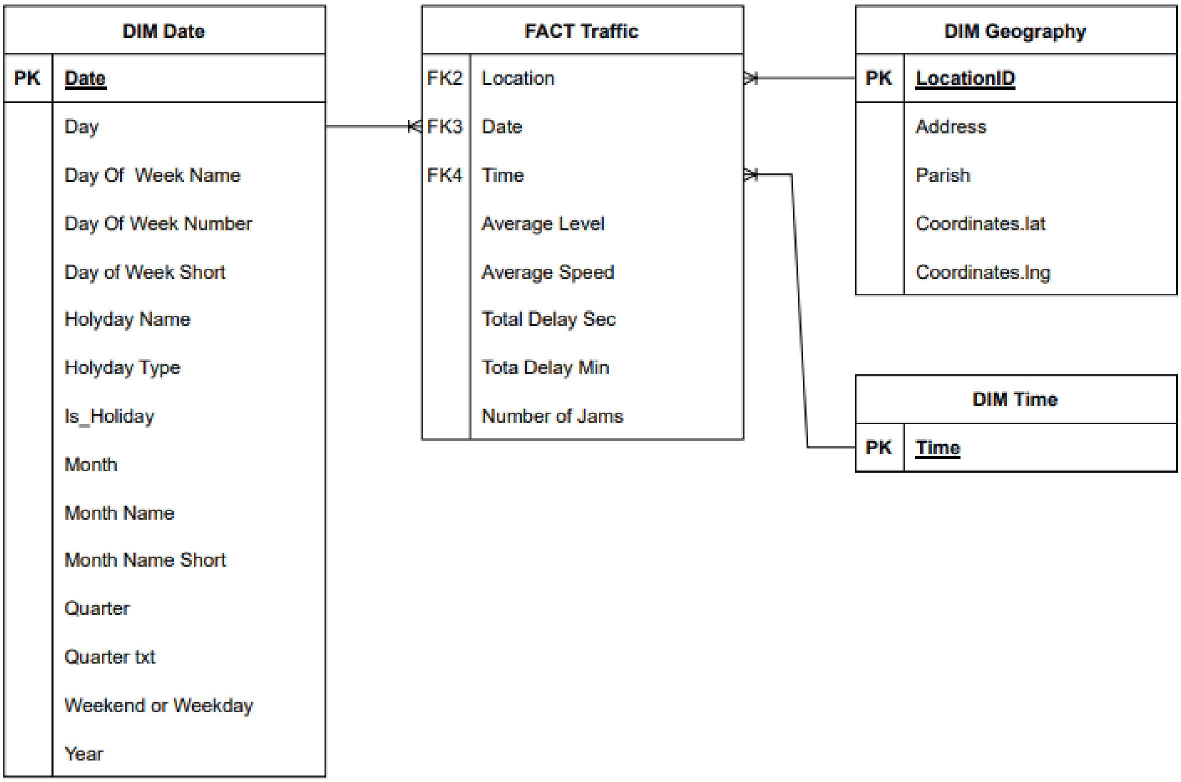


Figure A.3 – Relational Model for *FACT Traffic*