

Report On Air Quality

S-18-824

January 30, 2024

Contents

1	Introduction	3
2	Literature Review	4
3	Data set	5
3.1	Data Description	5
3.2	Data Exploration	5
4	Results and Discussion	6
4.1	Exploratory Data Analysis	6
5	Conclusion	11

1 Introduction

Air quality has an impact on public health and well-being. Air pollution is a major concern in many cities across the world because it is responsible for million deaths yearly and it can cause adverse health effects, such as asthma, allergies, cancer and infections due to rapid growth of population, increase in number of vehicles and industrialization.

The health - related issues are correlated with air pollution, particularly in traffic. Air pollutants such as nitrite oxides (NO_2 , NO and NO_x), sulphur dioxide (SO_2), ozone (O_3) and particulate matter ($PM_{2.5}$, PM_{10}) are very harmful.

Environmental Protection Agency and World Health Organization have acknowledged the material impact of air quality on public health and policies to regulate and improve air quality.

Various studies conducted in London at various sites suggests that pollution levels varies significantly in different areas with respect to its location, time, period of sampling and climatic conditions.

In this project, we aim to analyze and discover the trends or underlying principles in air quality across London city based on data collected at 36 air monitoring sites located in London from 01/01/2022 to 31/12/2023.

I am writing this report to share my approach and what I analyzed from the data and also make people aware of the enormous problem of the country is facing.

2 Literature Review

Many searches were conducted to identify studies that had measured or modeled air pollution concentrations in other countries. The following evidence is one of that,

A past research regarding an air pollution have analyzed the proportion of various air pollutants (NO , NO_2 , CO , PM_{10} and SO_2) with respect to the time of the day and the day of the week and estimated the effect of environmental parameters as temperature, wind speed and humidity on the air pollutants mentioned above with the help of data obtained from India.

According to the past studies related to air quality they consider over all harmful air pollutants and they predict future values based regression analysis and time series analysis.

Here, only focusing on some exploratory data analysis and insights we can gain from that in the harmful air pollutants.

3 Data set

3.1 Data Description

Hourly measurement of NO_2 , NO_x , NO , O_3 , SO_2 , PM_{10} , $PM_{2.5}$ are taken from the sites and saved as "London local data 2022" data set. The details of each monitoring sites are given in another data set including the following variables.

1. code - identification code of the site.
2. site - name of the site.
3. latitude - latitude of the site.
4. longitude - longitude of the site.
5. parameter name - name of the substance measured.

The data obtained were transformed into a form suitable for processing by the R program and used to derive predictions based on that air pollutants.

3.2 Data Exploration

The data set contains **289,069** rows and **10** columns with variables such as site, code, date, NO_2 , NO_x , NO , O_3 , SO_2 , PM_{10} , $PM_{2.5}$ including some missing values also. We can see only few non - null values in the column containing $PM_{2.5}$, ozone and SO_2 .

Among the 36 monitoring sites in London some are having nitrite oxides and particulate matter 2.5 in high amount and some have ozone and sulfur dioxide in high concentration. So, it is difficult to analyze the whole data set to find the air pollutants. Use some data wrangling techniques and extract useful data.

From the summary statistics we can find mean concentration, median, max, min, quartile and missing values(NA). Data is summarize in the following table.

Pollutant's name	Mean	No.of NA's
NOx	62.81	46861
NO2	33.15	46861
NO	19.34	46860
PM10	19.47	121636
O3	49.61	269738
PM2.5	10.6	272308
SO2	3.08	280607

Table 1: Summary of the data set

This shows sulfur dioxide, ozone and particulate matter 2.5 are found in few sites. Among the pollutants nitrite oxide (NO_x) is found in high concentration in air.

4 Results and Discussion

4.1 Exploratory Data Analysis

The exploratory data analysis showed that there were missing values in the data set. By using the R codes we can remove those missing observations and discuss the effects of the several harmful gases such as sulfur dioxide, nitrite oxides and other particulate matters.

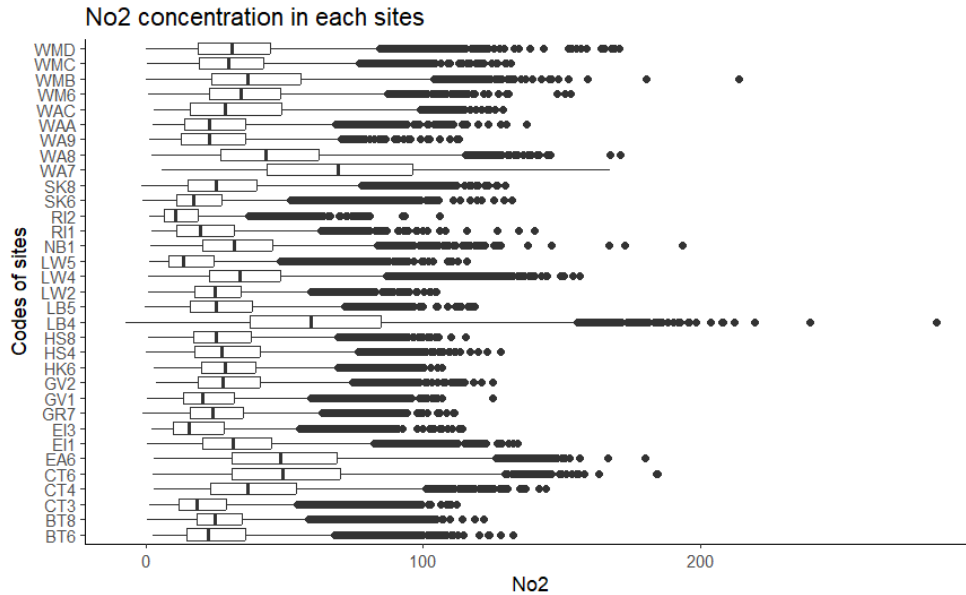


Figure 1: Box plot of NO₂ with each sites

From the above figure, we can see that how concentration of NO_2 varies with monitoring sites in London. This graph shows so many outliers in that data set. We have to handle this in proper way to analyze the whole data set.

Let's have a look on how NO_2 concentration varies using a bar plot.

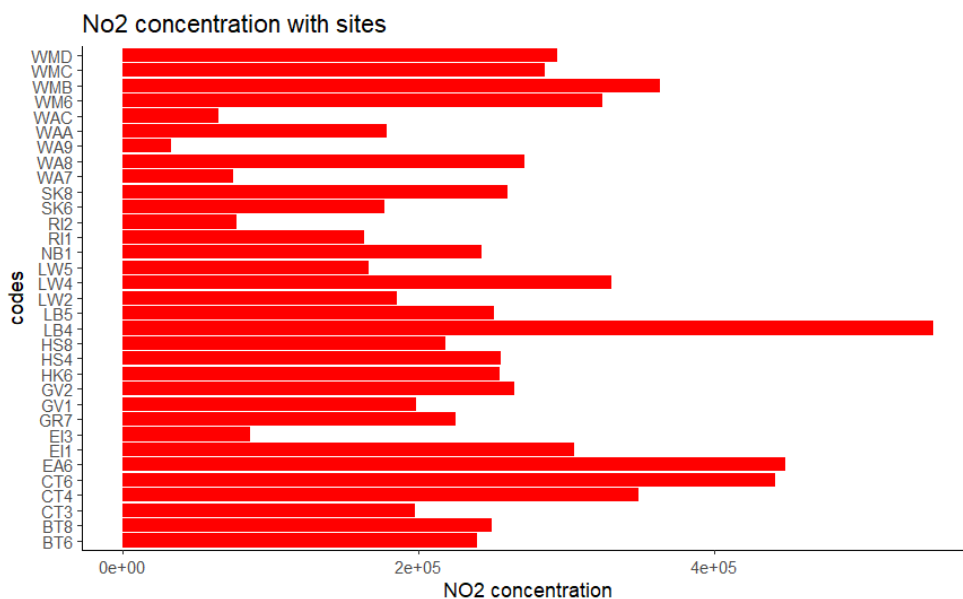


Figure 2: Distribution of NO2 with each sites

According to this figure 2 Lambeth - Brixton Road has high concentration of nitrogen dioxide among 33 monitoring sites whereas Wandsworth - Putney High Street has low concentration. Nitrogen dioxide is harmful gas so we have to take necessary actions to reduce that amount in Lambeth in order to get rid of health issues.

Variation of particulate matter with each month in an year

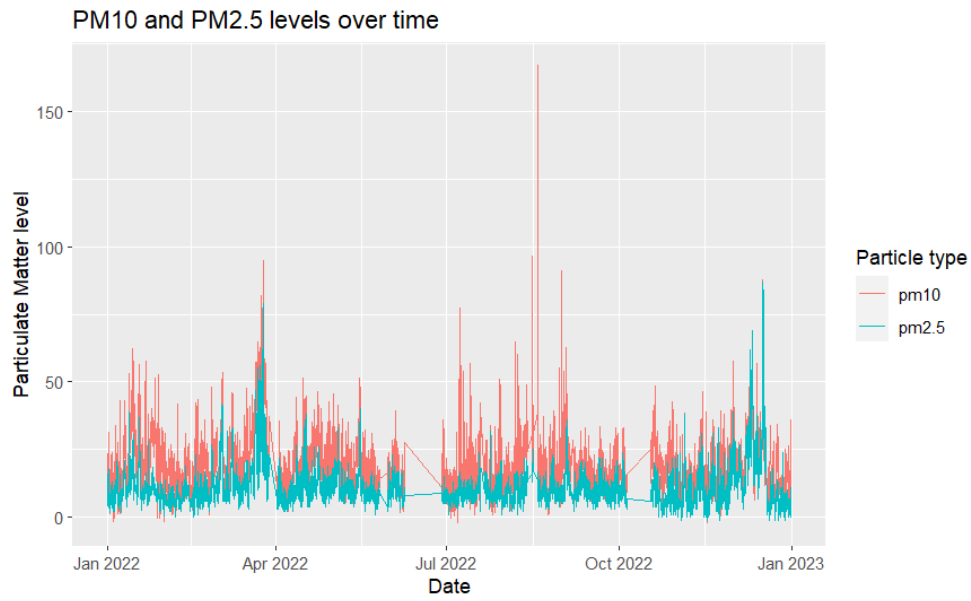


Figure 3: Distribution of particulate matter with year

This graph shows the variation of particulate matter 10 and particulate matter 2.5 in each sites in the year 2022. Particulate matter 10 is in high amount than that of 2.5 in whole year. In the end of November particulate matter 2.5 is in high concentration than particulate matter 2.5

Particulates are the deadliest form of air pollution due to their ability it penetrate deep into lungs and causing respiratory disease and heart attacks. Now we can separately analyze both particulate matters and see whether how their effects in sites located in London.

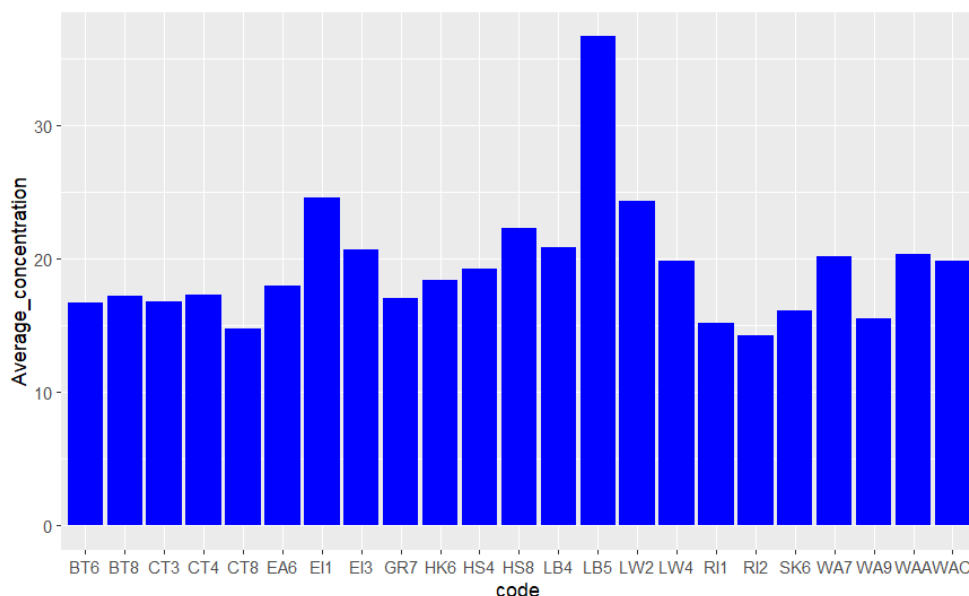


Figure 4: Distribution of particulate matter 10 with monitoring sites

From the above figure 4 we can see that particulate matter 10 is present in high amount in Lambeth city while in other sites its concentration is nearly same.

Let's see the variation of particulate matter 2.5 by month wise in all monitoring sites in London.

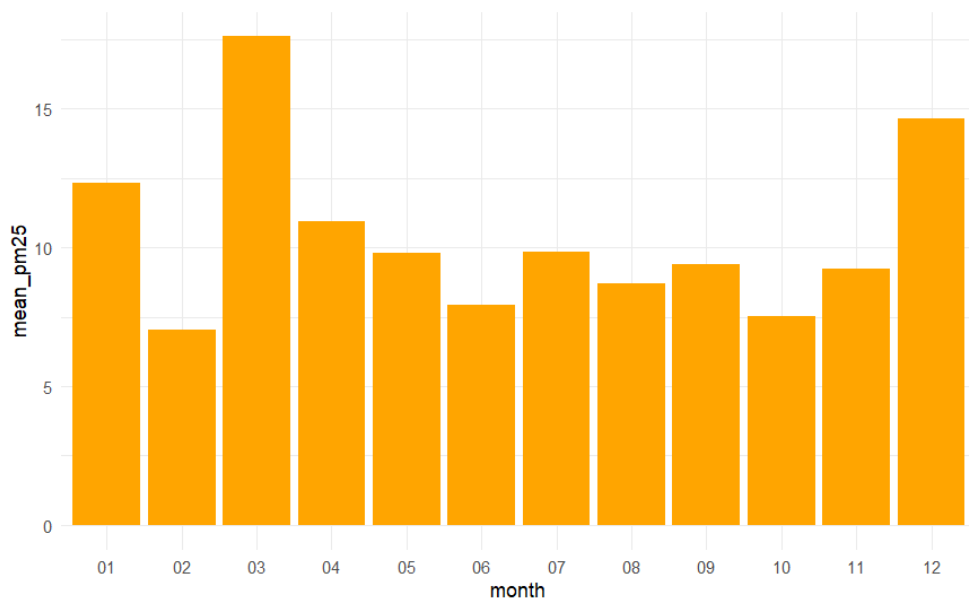


Figure 5: Distribution of particulate matter 2.5

This plot clearly illustrates the variation of particulate matter 2.5 in each month. In March $PM_{2.5}$ concentration is nearly high.

We can check whether there is any relationship between two particulates using scatter plot given below.

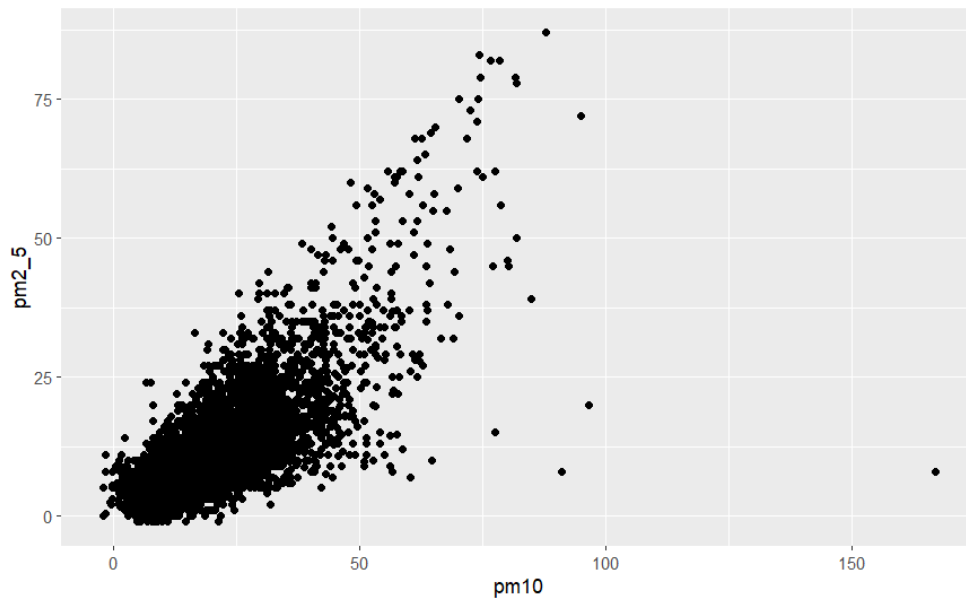


Figure 6: Scatter plot between particulates

We can assume there is a strong positive relationship between the particulate matter 10 and 2.5

When we analyze ozone concentration, it is only available in 3 monitoring sites such as Hackney - Old Street, Richmond Upon Thames - Barnes Wetlands and Southwark - Elephant and Castle. The distribution is given in the following figure.

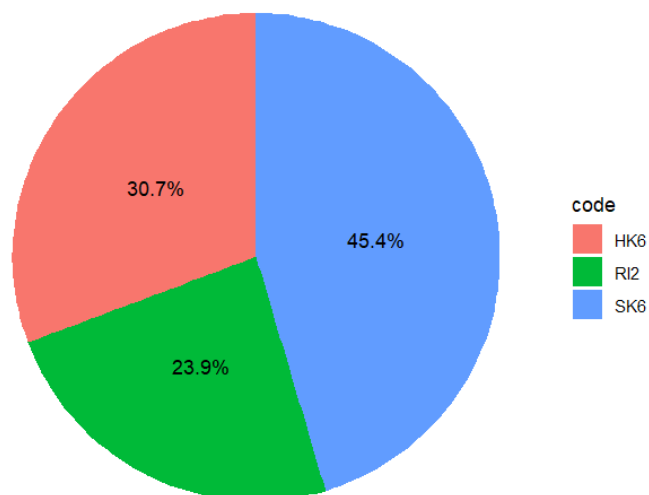


Figure 7: Ozone concentration with sites

In Southwark site, percentage of ozone concentration is nearly high and Richmond has low concentration .

5 Conclusion

We can analyze the effect of all the pollutants which are very dangerous using the above plots which were generated using r codes. You can get some insights from those plots and you can see how they effect .

From the overall analysis, in the Lambeth city concentration of nitrite oxide and particulate matters is very high when compared to other cities in London. Both are very dangerous to human heath. So, it is necessary to take severe actions in those sites to prevent from dangerous situations in near future.

In conclusion, the analysis showed that there were variation in air quality across different areas of London, some sites have higher risk of harmful gases while some sites have low effect and also seasonal variations in air quality. The findings of this study is useful to reduce the air pollution levels.