United States International University-Africa

Education to take you places

ANALYZING GLOBAL AIR QUALITY INDEX (AQI) DATA

Group 5

Tanveer Omar 670752

- Ambachow Kahsyay 670550
- Peter Kidiga 671341
- Mitchel Makena 667413

Instructor Name: Dr. Alu

Due Date: [illegible]

01-Introduction and Data Collection

02 - Data importing and cleaning

03 - Data Exploration and Analysis

04 - Visualization and Interpretation

Introduction

Understanding the structure and contents of the dataset is key to resolving issues and proceeding with analysis and visualization.

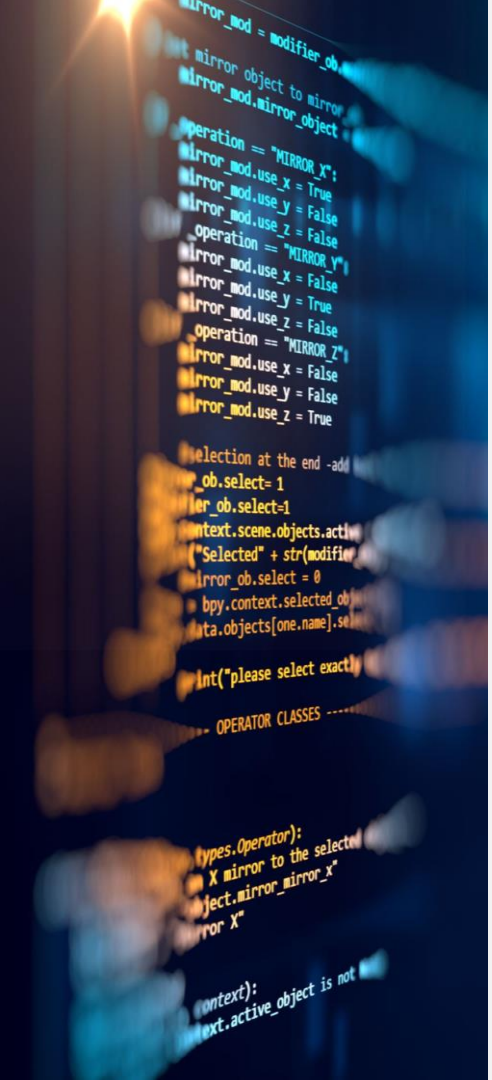
What is Data

The term "data" refers to the raw information you have in your dataset.

- Data: Raw information collected for analysis. This can include various types of data such as numerical values, text, dates, etc.

- DataFrame: A two-dimensional, size-mutable, and potentially heterogeneous tabular data structure with labeled axes (rows and columns) used in libraries like pandas.

So, the name of the **columns are variables** of the dataset, and the **rows are values of a variable**. The value can be numeric or non-numeric. If value is numeric, it is quantitative and if not, it is qualitative type of the data. So Qualitative and quantitative are the two broad classification of data.



Important libraries and functions

Pandas

Used for data manipulation and analysis. Key functions include reading and writing data, data cleaning, aggregation, and transformation

- **pd.read_csv(), pd.read_excel():** Reading data from files
- **pd.to_datetime():** Converting columns to datetime format
- **groupby(), pivot_table():** Aggregation and reshaping data

Matplotlib

A plotting library for creating static, animated, and interactive visualizations in Python.

- plt.plot(), plt.bar(), plt.pie(), plt.hist(): Different types of plots
- plt.savefig(), plt.show(): Saving and displaying plots
- plt.figure(), plt.subplot(): Creating figure and subplots

seaborn

A statistical data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics

- sns.heatmap(), sns.histplot(), sns.boxplot(), sns.scatterplot(): Creating various types of plots with enhanced aesthetics

DATA CLEANING AND PREPARATION

- **Data Cleaning Steps:**
  - **Handling Missing Values:** No Recorded Missing Values
  - **Removing Duplicates:** Mostly unique Data
  - **Data Transformation:** Converted date and time columns for consistency.

Importing Data and Cleaning

Data Structure: Before

Name of the data frame: data_combined

consists *of 64019 Rows, 15 columns*

The names of the column are:

'location_id', 'location_name', 'parameter', 'value', 'unit', 'datetimeUtc', 'datetimeLocal', 'timezone', 'latitude', 'longitude', 'country_iso', 'isMobile', 'isMonitor', 'owner_name', 'provider']

Focus: location_name', 'parameter', 'value', 'unit', 'datetimeUtc', 'datetimeLocal',

Location_name: is a categorical type of data that consists of cities; ***Bucharest', 'Los Angeles', 'Madrid', 'Nairobi', 'Rome***

parameter: It is a categorical data consists of the name of the pollutants namely; ***'pm10', 'so2', 'co', 'no2', 'no', 'o3', 'nox', 'pm25']***

Value: value is numeric type of data that consists of each records of the measures of pollutants

Unit: it is categorical type of data that consists of unit measurement of each pollutants [ppm and μg/m^3]

datetimeUtc/datetimeLocal: It is object type of data and then it will be converted time format that means every value is collocted at per/1hr.

Importing Data and Cleaning

Data Structure: plan (for better analysis we will add the following new columns

DateUtc/local: a date format [2024-02-06} data that consists of each dates of the data recorded

DayOfWeek: The date data that consists of days of the week

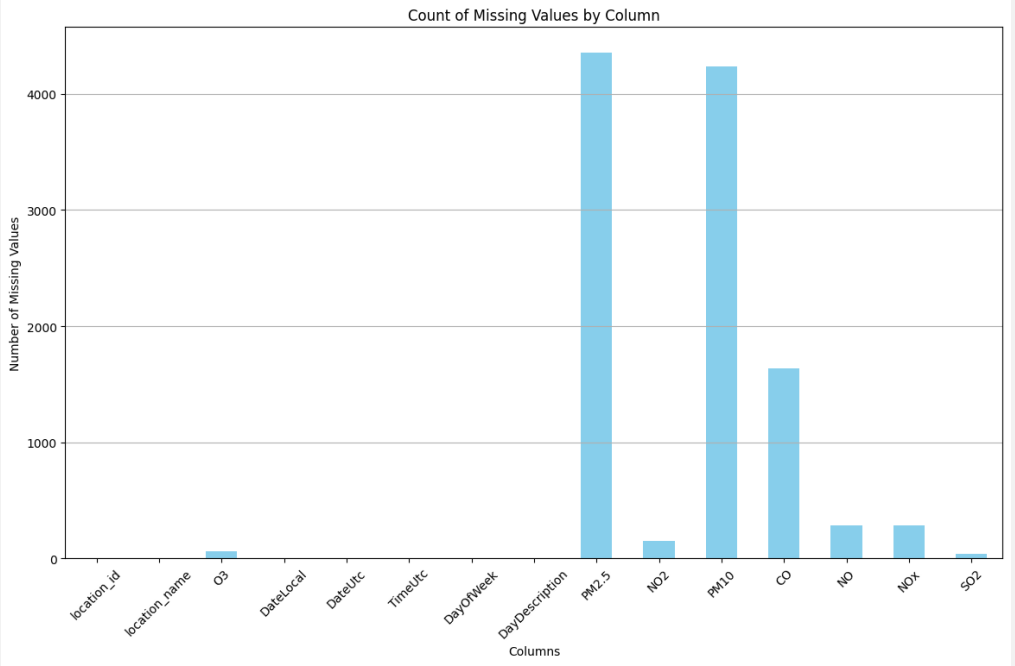
Day Description: a categorical data that consists of workday, Holiday name or weekend.

O3 /PM10/ PM2.5/ NO2/CO/NO/Nox/O2 (for raw and AQI value)

Importing Data and Cleaning

Handling missing Values: Using Pandas

- Checking for missing values
- But be care full the data that seems missing value is not missed pivoting by itself can create NaN values
- So first we checked the type of pollutants for each city
- Nairobi has only PM2.5 reading so we filled the other 0.To the same goes to others
- The rest using linear interpolation method
- For time data and a value of parameter recorded per hour interpolation is very efficient.
- And in third stage backfill method.



DATASET AT A GLANCE

Locations:

- Bucharest(Romania)
- Los Angeles(North Hollywood)
- Molina De Segura(Spain
- Madrid
- Nairobi
- Rome

Time Frame:

- January 2024 to June 2024

Pollutants:

- PM2.5 (Particulate Matter less than 2.5 micrometers)
- PM10 (Particulate Matter less than 10 micrometers)
- CO (Carbon Monoxide)
- NO2 (Nitrogen Dioxide)
- SO2 (Sulfur Dioxide)
- O3 (Ozone)

PM2.5 (PARTICULATE MATTER LESS THAN 2.5 MICROMETERS)

PM2.5 refers to particulate matter with a diameter of 2.5 micrometers or smaller. These tiny particles are capable of penetrating deep into the respiratory system and can even enter the bloodstream.

Implications:

- **Health Effects:** Exposure to PM2.5 can lead to serious health problems, including respiratory and cardiovascular diseases, reduced lung function, and aggravation of pre-existing conditions like asthma and heart disease.
- **Environmental Impact:** PM2.5 can affect visibility, contribute to haze, and impact ecosystems by settling on soil and water bodies

PM10 (PARTICULATE MATTER LESS THAN 10 MICROMETERS)

- PM10 refers to particulate matter with a diameter of 10 micrometers or smaller. These particles are larger than PM2.5 but can still be inhaled and cause health problems.
- **Implications:**
- **Health Effects:** PM10 can cause respiratory issues, exacerbate asthma, and increase the risk of cardiovascular diseases. It may also contribute to lung infections and other chronic respiratory conditions.
- **Environmental Impact:** Like PM2.5, PM10 can reduce visibility and affect air quality. It can also have detrimental effects on vegetation and water bodies.

CO (CARBON MONOXIDE)

Carbon monoxide is a colorless, odorless gas produced by incomplete combustion of carbon-containing fuels. It is commonly emitted from vehicle exhausts and industrial processes.

Implications:

- **Health Effects:** CO interferes with the blood's ability to carry oxygen, leading to symptoms like headaches, dizziness, and in severe cases, can cause unconsciousness or death. Chronic exposure can lead to long-term cardiovascular and neurological problems.
- **Environmental Impact:** CO does not have significant direct environmental impacts compared to other pollutants, but it can contribute to the formation of ground-level ozone and smog.

NO2 (NITROGEN DIOXIDE)

- Nitrogen dioxide is a reddish-brown gas with a characteristic sharp odor, produced from combustion processes, particularly from vehicles and power plants.
- **Implications:**
- **Health Effects:** NO2 can irritate the respiratory system, reduce lung function, and increase susceptibility to respiratory infections. It also contributes to the development of asthma and chronic bronchitis.
- **Environmental Impact:** NO2 contributes to the formation of ground-level ozone and particulate matter. It can also lead to acid rain, which harms ecosystems and water bodies.

SO2 (SULFUR DIOXIDE)

- Sulfur dioxide is a colorless gas with a pungent smell, produced primarily from burning fossil fuels containing sulfur, like coal and oil.
- **Implications:**
- **Health Effects:** SO2 can cause respiratory problems, aggravate asthma, and reduce lung function. Short-term exposure can lead to coughing and throat irritation.
- **Environmental Impact:** SO2 contributes to the formation of acid rain, which can damage buildings, soil, and water bodies. It can also affect vegetation and contribute to the formation of fine particulate matter.

O3 (OZONE)

- Ozone is a gas composed of three oxygen atoms. It occurs naturally in the upper atmosphere (stratosphere) where it protects life from harmful ultraviolet (UV) radiation. However, at ground level (troposphere), it is a major component of smog and a harmful air pollutant.
- **Implications:**
- **Health Effects:** Ground-level ozone can cause respiratory issues, aggravate asthma, and reduce lung function. Long-term exposure can lead to chronic respiratory diseases and reduced lung development in children.
- **Environmental Impact:** Ozone can damage crops, other vegetation, and materials like rubber and plastics. It also contributes to the formation of smog, which affects overall air quality and visibility.

DATA COLLECTION AND SOURCES

EXPLORATORY DATA ANALYSIS

- Distribution of Pollutants
- Average pollutant Values

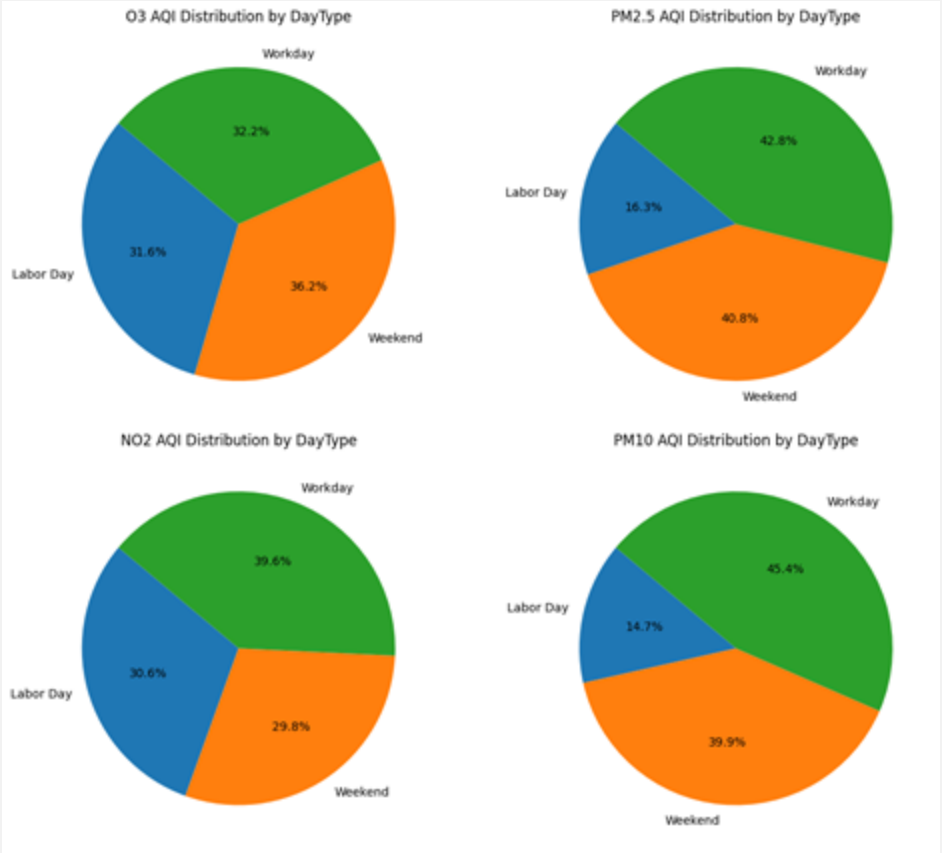
SUMMARY STATISTICS OF NAIROBI CITY

- **The table above visualizes the summary statistics for PM2.5 levels in Nairobi**. The statistics include:
- **Count**: The number of data points (3319).
- **Mean**: The average PM2.5 value (49.36).
- **Std Dev**: The standard deviation, indicating variability (65.62).
- **Min**: The minimum PM2.5 value recorded (0.0).
- **25th Percentile**: The value below which 25% of the data falls (7.7).
- **Median (50th Percentile)**: The middle value (12.0).
- **75th Percentile**: The value below which 75% of the data falls (75.5).
- **Max**: The maximum PM2.5 value recorded (210.0)

Statistic	Value
Count	3319.0
Mean	49.36
Std Dev	65.62
Min	0.0
25th Percentile	7.7
Median	12.0
75th Percentile	75.5
Max	210.0

MAJOR POLLUTANTS BY CITY

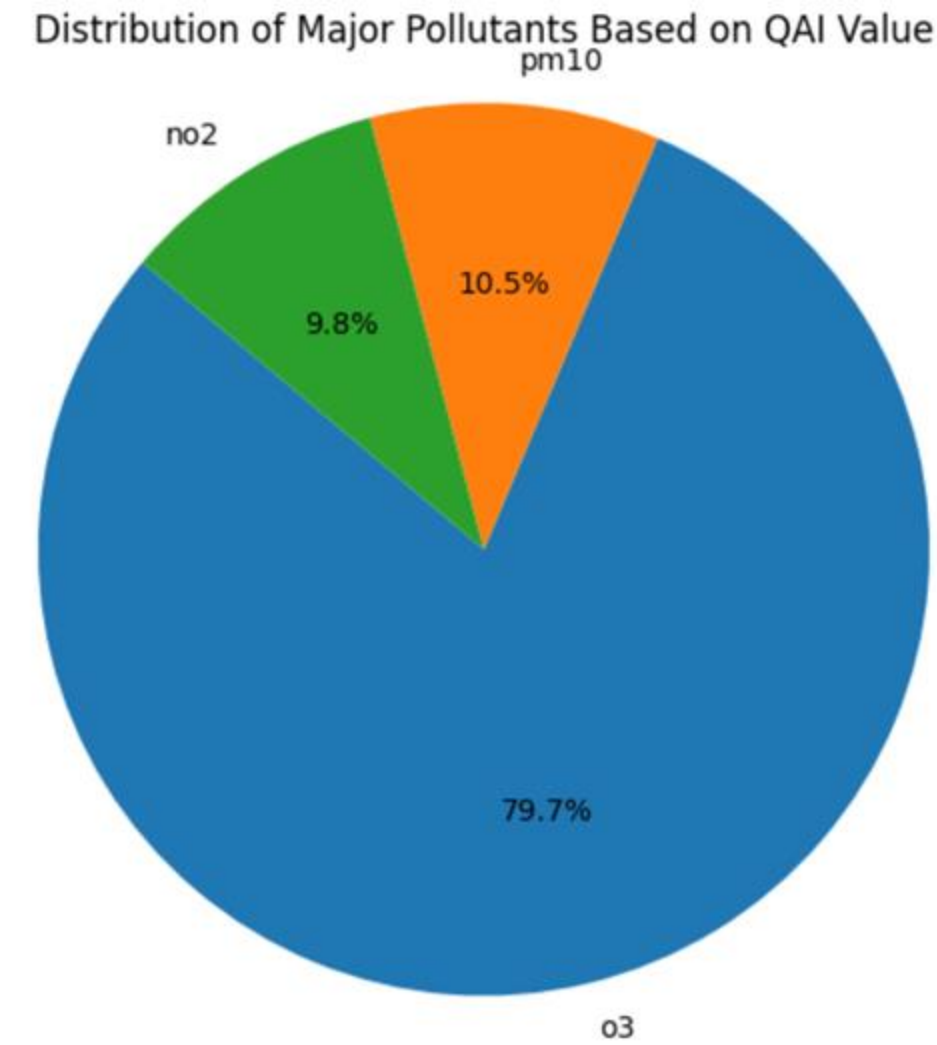
- **Bucharest**: CO (Carbon Monoxide)
- **Los Angeles:** PM2.5
- **Madrid:** O3 (Ozone) & PM10
- **Nairobi:** PM2.5
- **Rome:** O3 (Ozone, PM10 and PM2.5:



MOLINA AQI DISTRIBUTION BY DAY TYPE

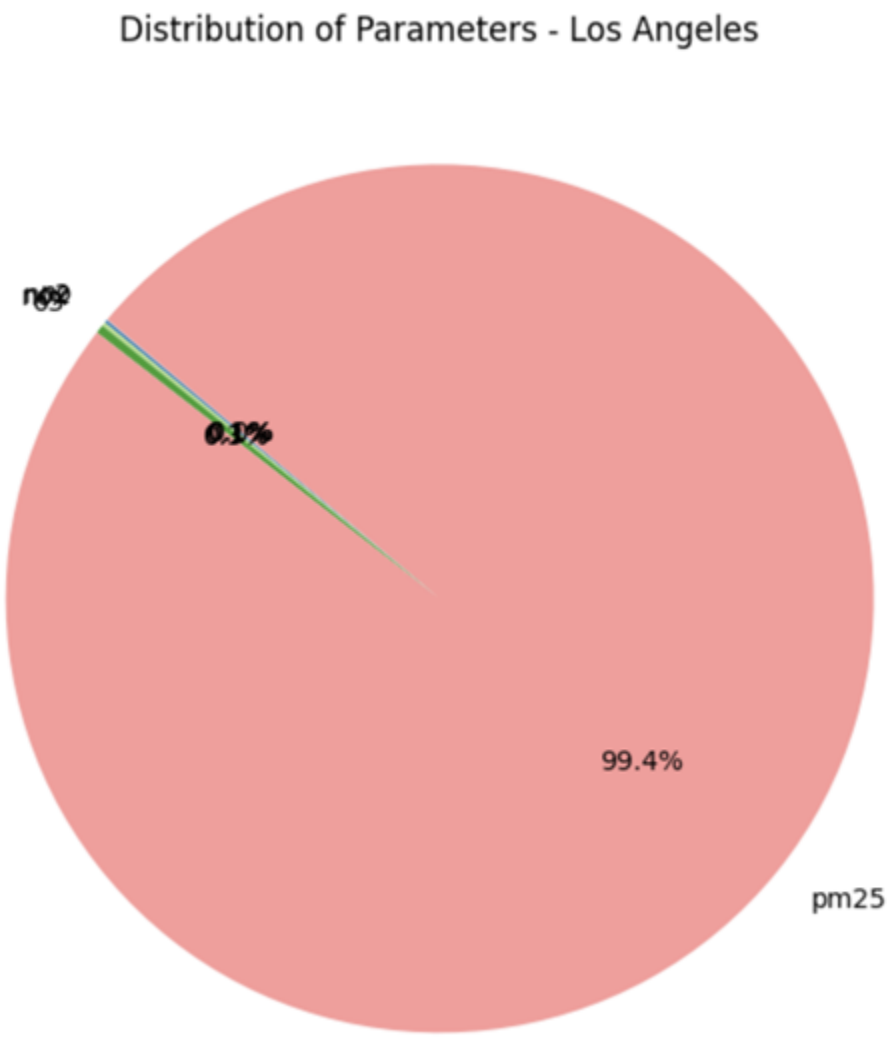
The O3 AQI distribution was highest during the weekend(36.2%) but not significantly from the workday and labor day.

The AQI distribution of PM2.5,PM10 and NO2 show that the highest percentage is during workdays.



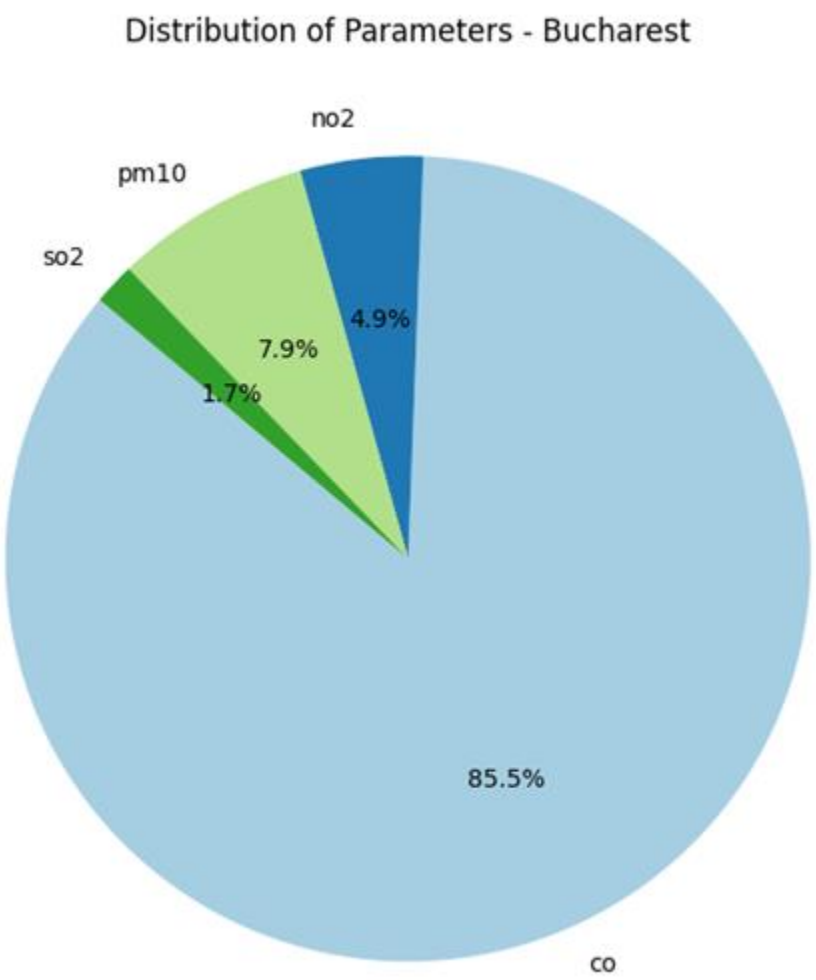
DISTRIBUTION OF MAJOR POLLUTANTS
MOLINA DE SEGURA(SPAIN)

- Analysis:
  - **O3 (Ozone)**: 79.7%
  - **PM10 (Particulate Matter 10 micrometers or less)**: 10.5%
  - **NO2 (Nitrogen Dioxide)**: 9.8%



DISTRIBUTION OF MAJOR POLLUTANTS
B) LOS ANGELES

- **Analysis**
- Dominance of PM25: there is an overwhelming amount of pm25 compared to the rest of the pollutants which may mean they are insignificant to the analysis



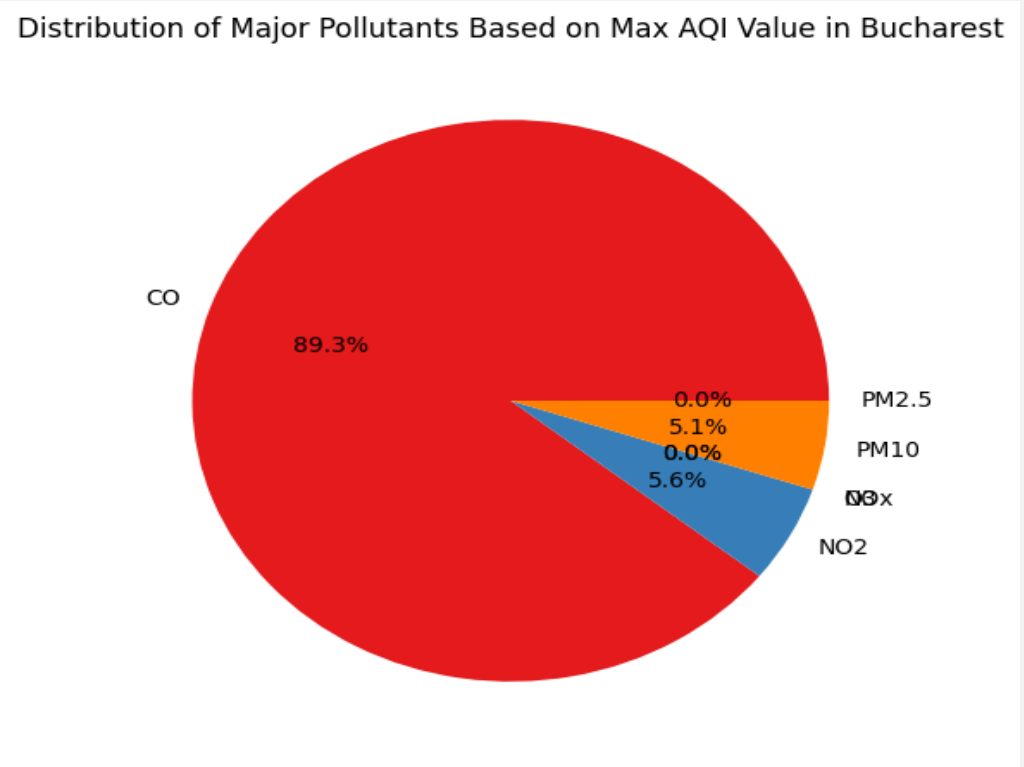
DISTRIBUTION OF MAJOR POLLUTANTS
C) BUCHAREST

- **Analysis**
  - Dominance of CO: The chart indicates that carbon monoxide is the most prevalent pollutant, making up 85.5% of the total. This could be due to high traffic emissions, industrial activities, or other combustion processes in Bucharest.
  - NO2 Levels: Nitrogen dioxide, which constitutes 4.9%, is typically associated with vehicle emissions and industrial processes. Elevated NO2 levels can contribute to respiratory problems and other health issues.
  - SO2 and PM10: Both sulfur dioxide and particulate matter (PM10) are present at 1.7%. SO2 is often produced by burning fossil fuels, while PM10 can come from various sources, including construction sites, unpaved roads, and industrial emissions.

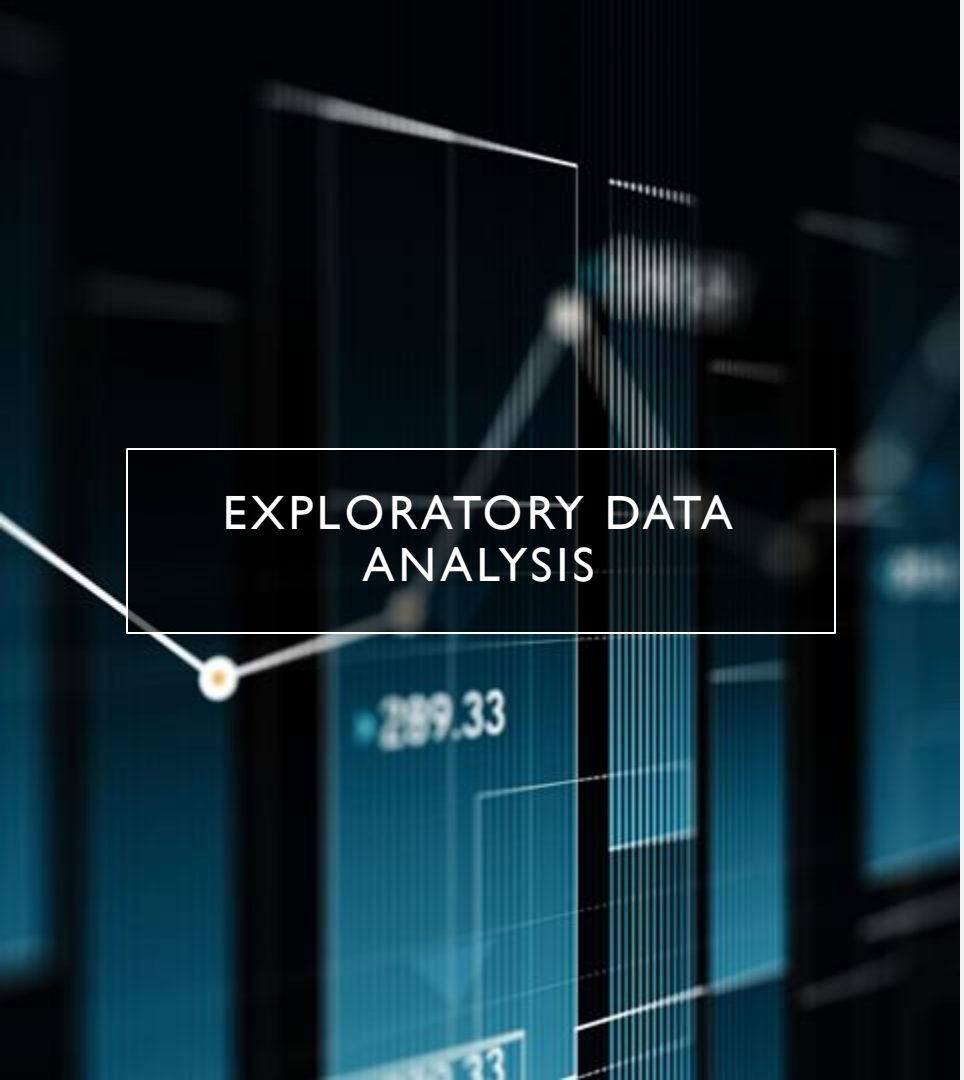
Distribution of Major pollutants by Max Value

The pie chart highlights that Carbon Monoxide (CO) is the primary pollutant influencing the AQI in Bucharest, followed by Nitrogen Dioxide (NO2) and Nitrogen Oxides (NOx).

What about in Comparison to all cities




DAILY AIR
QUALITY
TRENDS



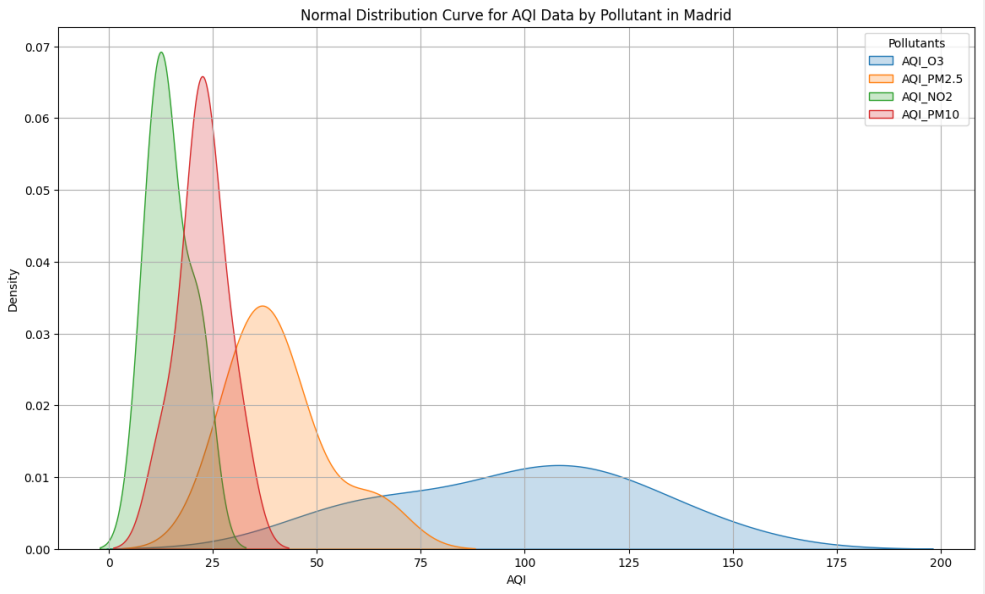
EXPLORATORY DATA ANALYSIS

- **Summary Statistics:**
  - **Central Tendency:** Mean AQI in Los Angeles was 85 (moderate), while in Bucharest it was 78 (moderate).
  - **Variability:** Los Angeles had a wider range of AQI values compared to Bucharest.
- **Initial Exploration:**
  - **Distribution of AQI Values:** AQI values ranged widely, with noticeable peaks.
  - **Categorical Data:** Dates were treated as categorical for month-wise analysis.

Observations from the graph

•**Ozone (AQI_O3)**:
•The distribution is more spread out, with a peak around 75 AQI and values extending up to 200 AQI. This indicates that ozone levels vary significantly, with occasional high pollution events.
•**PM2.5 (AQI_PM2.5)**:
•The distribution is relatively narrow, with a peak around 30 AQI. This suggests that PM2.5 levels are generally consistent but can reach higher levels occasionally.
•**Nitrogen Dioxide (AQI_NO2)**:
•This distribution is very narrow with a sharp peak around 25 AQI, indicating that nitrogen dioxide levels are fairly consistent with fewer fluctuations.
•**PM10 (AQI_PM10)**:
•The distribution shows a peak around 25 AQI and extends up to 75 AQI. Similar to PM2.5, PM10 levels are generally consistent but with some higher values observed.

NORMAL DISTRIBUTION CURVE FOR AQI DATA BY POLLUTANT IN MADRID



Normal Distribution Curve for AQI Data by Pollutant in Madrid

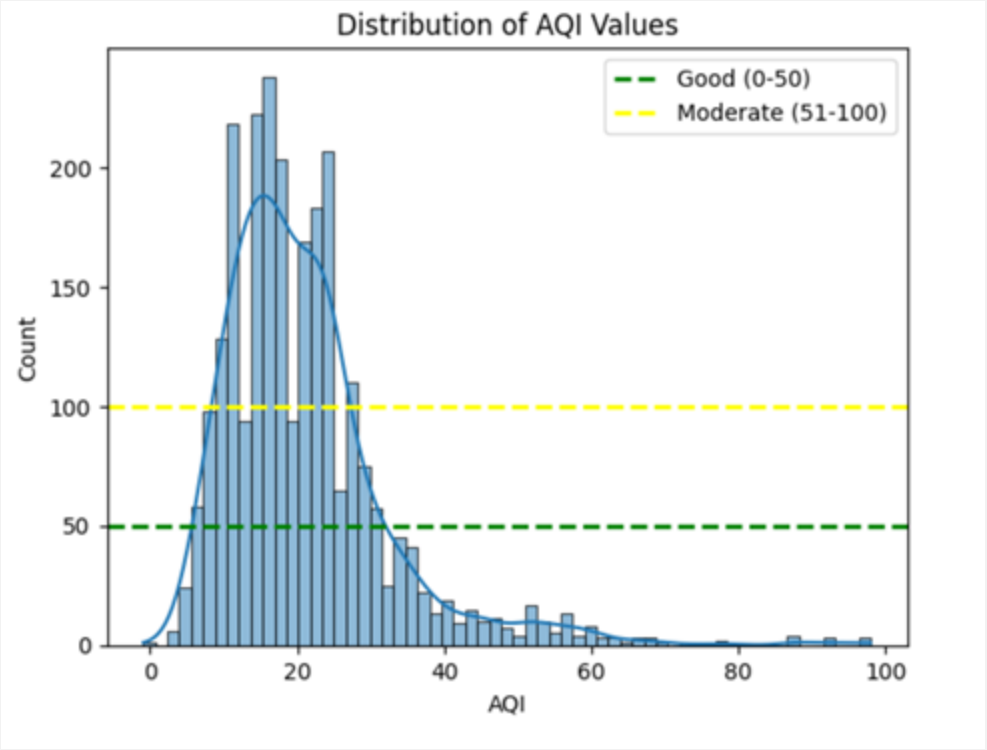
Observations from the graph

•**Ozone (AQI_O3)**:
•The distribution is more spread out, with a peak around 75 AQI and values extending up to 200 AQI. This indicates that ozone levels vary significantly, with occasional high pollution events.
•**PM2.5 (AQI_PM2.5)**:
•The distribution is relatively narrow, with a peak around 30 AQI. This suggests that PM2.5 levels are generally consistent but can reach higher levels occasionally.
•**Nitrogen Dioxide (AQI_NO2)**:
•This distribution is very narrow with a sharp peak around 25 AQI, indicating that nitrogen dioxide levels are fairly consistent with fewer fluctuations.
•**PM10 (AQI_PM10)**:
•The distribution shows a peak around 25 AQI and extends up to 75 AQI. Similar to PM2.5, PM10 levels are generally consistent but with some higher values observed.

DISTRIBUTION OF AQI BUCHAREST

The distribution of AQI is skewed to the right.

Most of the AQI values are below 50. This means the air quality is good.

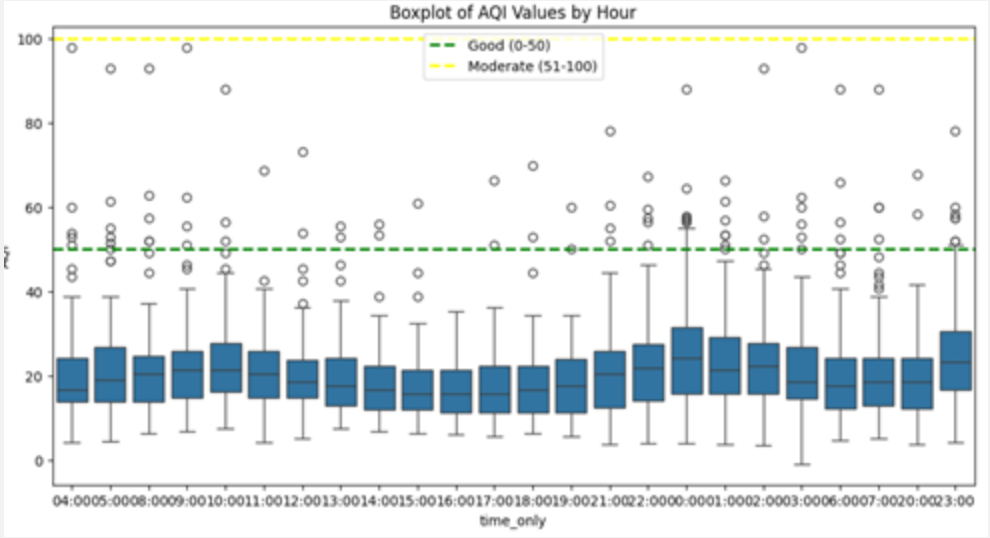


DISTRIBUTION OF AQI BUCHAREST BY HOUR

There are notable outliers over time.

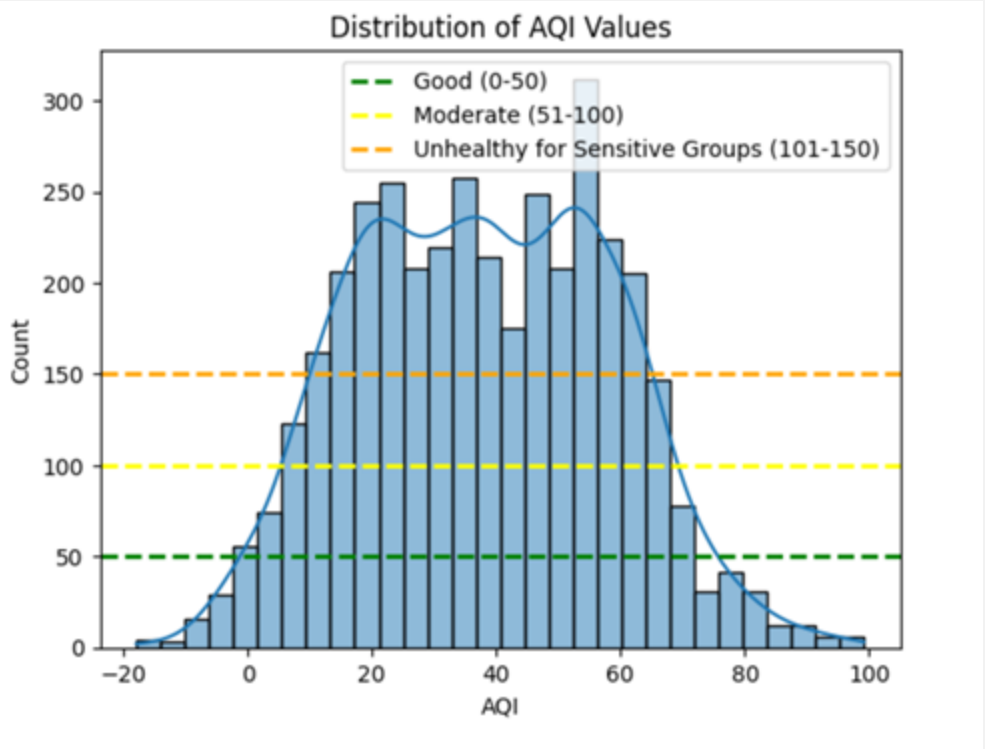
The AQI values are mainly below 50. This indicates good air quality.

The AQI was slightly higher between the 2000 hours and 2200 hours.



DISTRIBUTION OF AQI LOS ANGELES

The normal distribution of AQI values in Los Angeles suggests that pollution levels were significant throughout the time period.

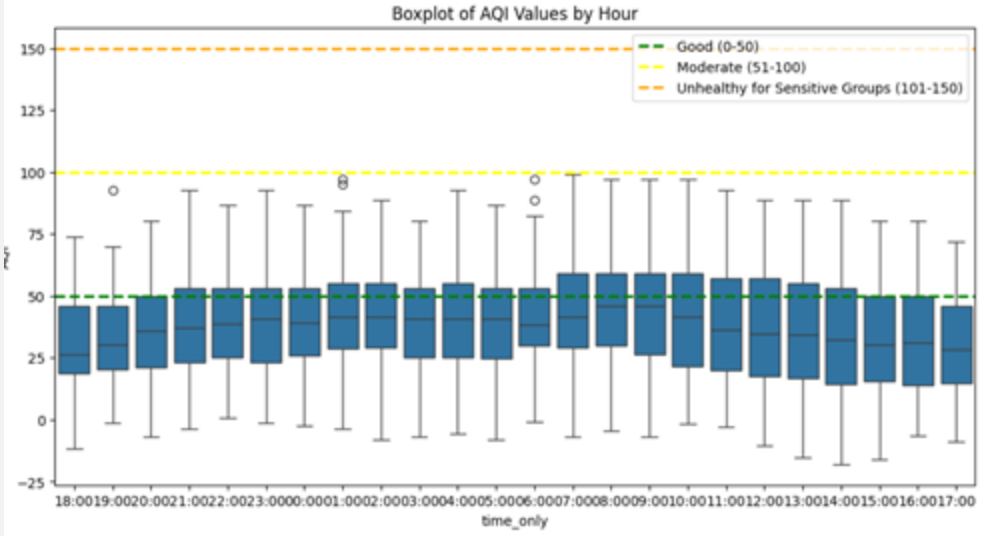


DISTRIBUTION OF AQI LOS ANGELES BY HOUR

There is little notable outliers.

Most AQI values are below or slightly above the 50 mark.

The AQI values are slightly above on the 0900 hour.

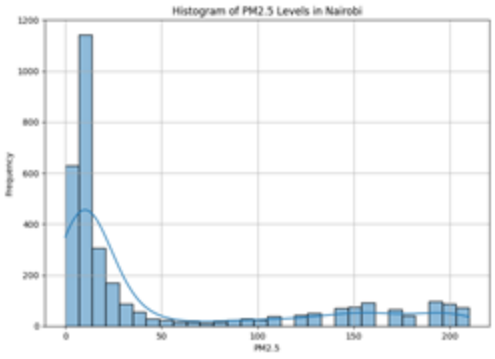


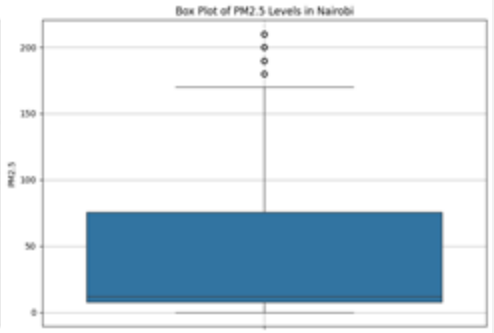
DISTRIBUTION OF AQI NAIROBI

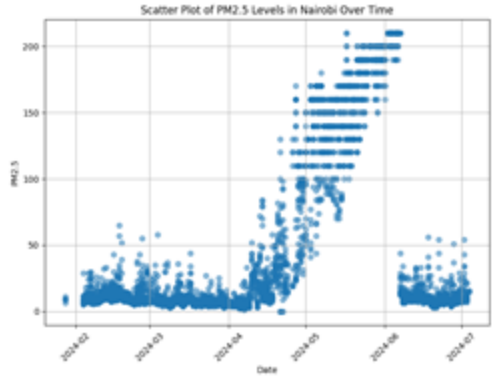
Histogram: The AQI values are skewed to the right. Most values are below 50 meaning the air quality is good.

Boxplot: The AQI values are positively skewed due to the large distance between the median and the maximum point. This also indicates that the air quality is good.

Scatter plot: The graph shows that between 2024-05 and 2024-06, there was an increase to around 200μg/m³. This is unhealthy air quality because it's above 150 but below 300.







Some major event occurred in April that led to sharp increase!

DISTRIBUTION OF AQI NAIROBI, Continued.

- **Histogram of PM2.5 Levels in Nairobi**: This chart shows the distribution of PM2.5 levels. The majority of the data points are clustered at lower levels, with a long tail extending towards higher values, indicating occasional spikes in PM2.5 levels.

- **Box Plot of PM2.5 Levels in Nairobi**: This chart displays the spread and central tendency of PM2.5 levels. The box represents the interquartile range, while the whiskers show the range of the data excluding outliers, which are marked as individual points above the whiskers.

- **Scatter Plot of PM2.5 Levels in Nairobi Over Time**: Ta chart tracks the PM2.5 levels over a specific period, showing how these levels fluctuate over time. There seems to be a noticeable increase in PM2.5 levels around May 2024, followed by a decrease.

TIME-SERIES ANALYSIS

- **Objective:**
Analyze AQI trends over time for Los Angeles and Bucharest.

- **Visualizations:**
  - **Time Series Plots:** Los Angeles showed seasonal peaks in AQI during summer months, while Bucharest had more stable levels with occasional spikes.
  - **Findings:** Seasonal variations in Los Angeles were linked to higher traffic and industrial activities during summer. Bucharest's AQI stability suggests less fluctuation in pollution sources.

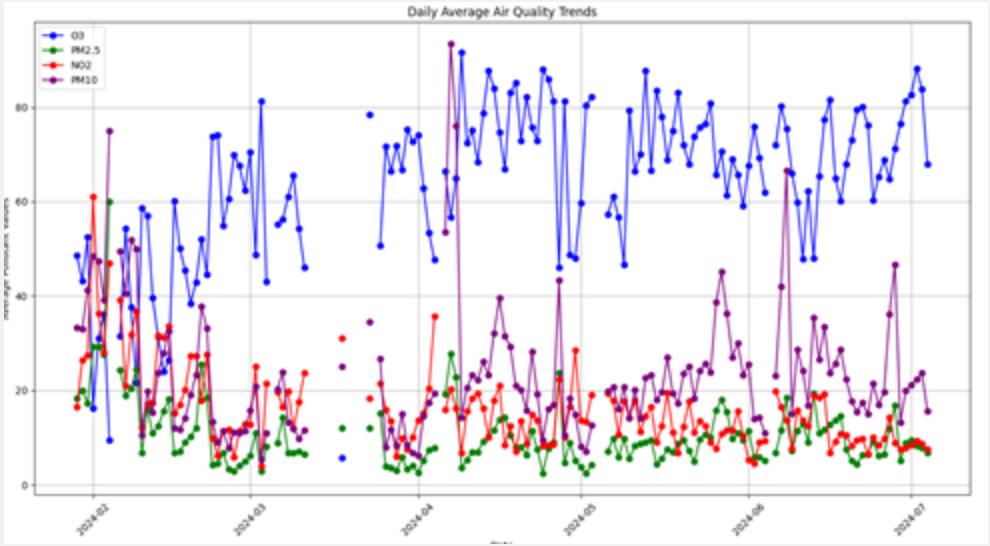
- **Analysis:**
  - **Los Angeles:** Significant AQI increases during summer and specific pollution events.
  - **Bucharest:** Stable AQI with occasional spikes linked to local events.

MOLINA DE SEGURA: AVERAGE AIR QUALITY TREND

This graph indicates that the air pollutant O3 is the most produced from the date

2024-03 to 2024-07.

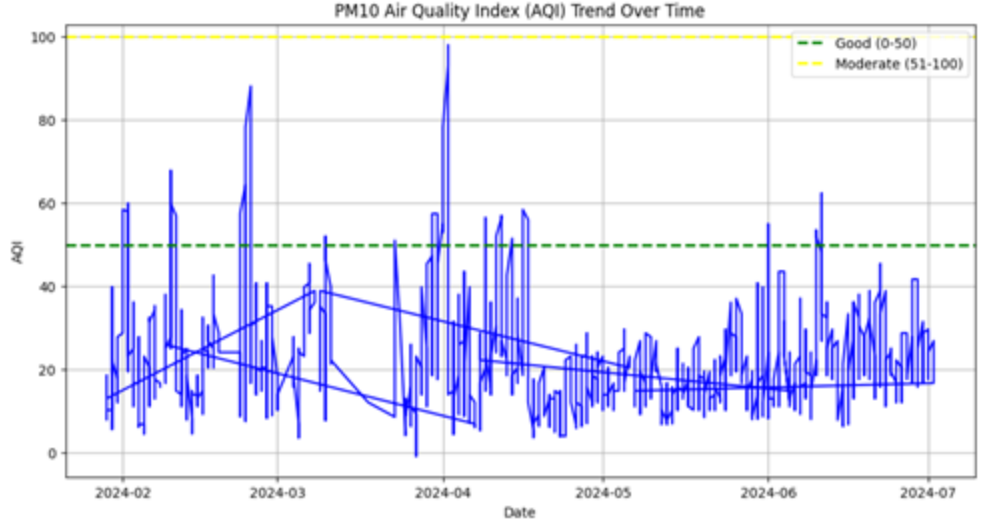
Also, between 2024-04 and 2024-05, the air pollutant PM10 was highly produced.



BUCHAREST: AIR QUALITY INDEX PER PM10

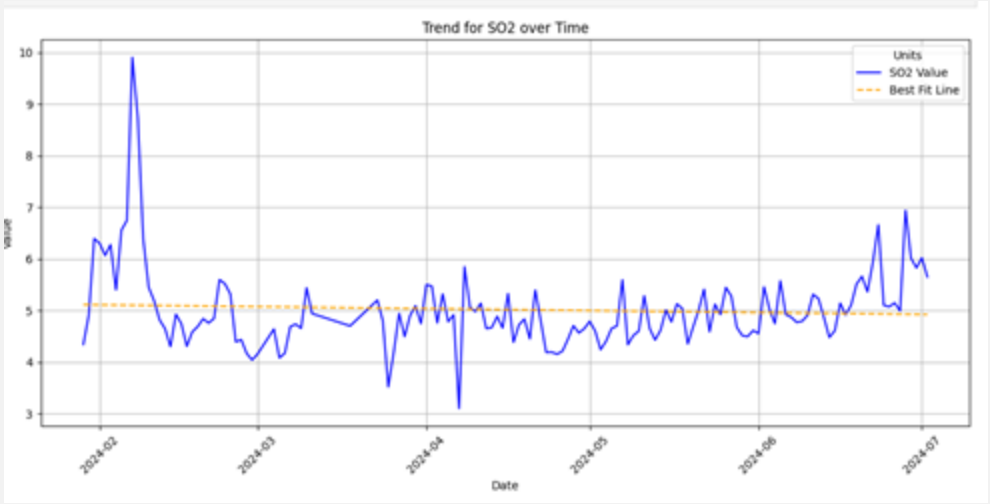
This graph shows that most PM10 AQI is below 50. This indicates good air quality.

At 2024-04, the AQI shoot up close to 100. This shows moderate air quality because it's above 50 but below 100.



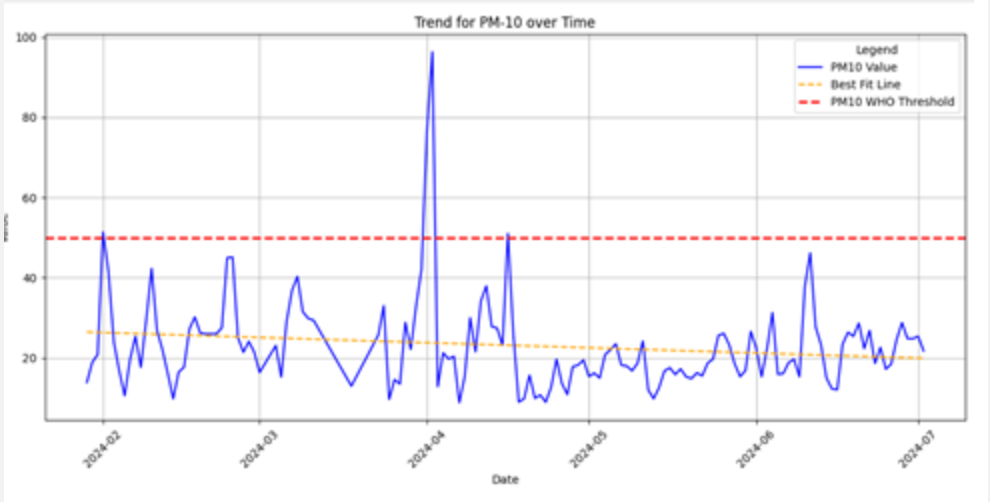
BUCHAREST: AVERAGE AIR QUALITY TREND 6-MONTH SULPHUR DIOXIDE

The SO_2 trend graph shows fluctuating levels in $\mu g/m^3$, with significant peaks on January 29th, February 18th, March 11th, April 10th, April 30th, May 20th, June 9th, and June 29th. These peaks may be due to increased energy consumption from burning fossil fuels, industrial emissions, volcanic eruptions, or weather conditions like wind patterns and temperature inversions affecting pollutant dispersion.



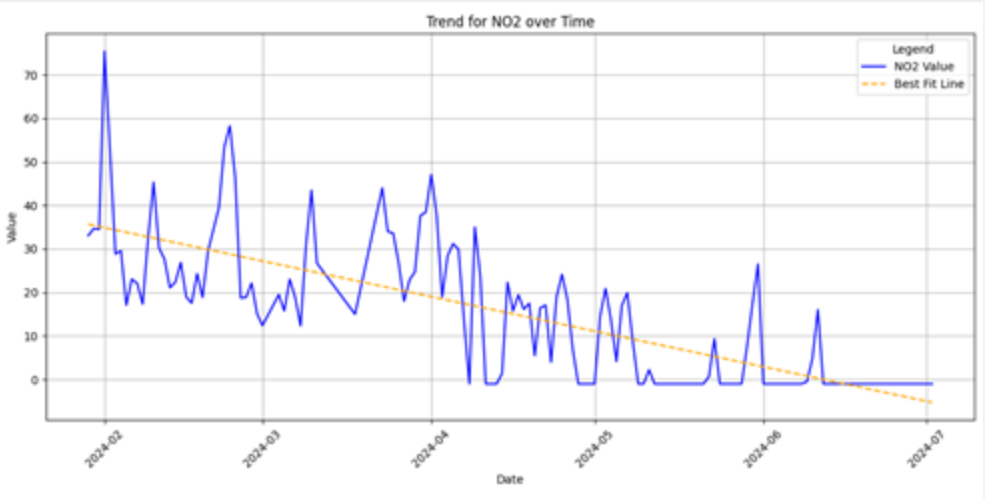
BUCHAREST: AVERAGE AIR QUALITY TREND 6-MONTH PM-10

The PM10 trend graph shows fluctuations in µg/m³, with peaks on January 29th, February 18th, March 11th, April 10th, April 30th, May 20th, June 9th, and June 29th, surpassing the WHO threshold of 50 µg/m³. These peaks may result from heavy traffic, industrial activities, weather conditions such as temperature inversions and wind patterns, or seasonal variations. Chronic exposure to PM10 can negatively impact respiratory health, making monitoring and mitigation efforts essential for air quality.



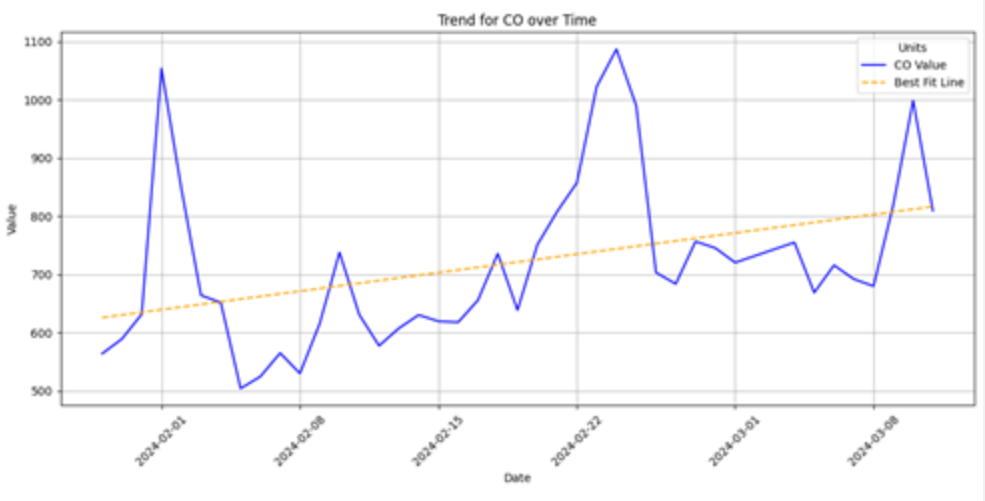
BUCHAREST: AVERAGE AIR QUALITY TREND 6-MONTH NITROGEN DIOXIDE

Data shows that NO2 levels, measured in µg/m³, decrease overall from late January to June, starting around 70 µg/m³ with significant peaks on January 29th, February 2nd, and March 6th. Levels drop to near 0 µg/m³ by June. Fluctuations are likely due to traffic patterns, industrial activities, weather conditions like temperature inversions, and seasonal variations, which influence NO2 levels throughout the year.



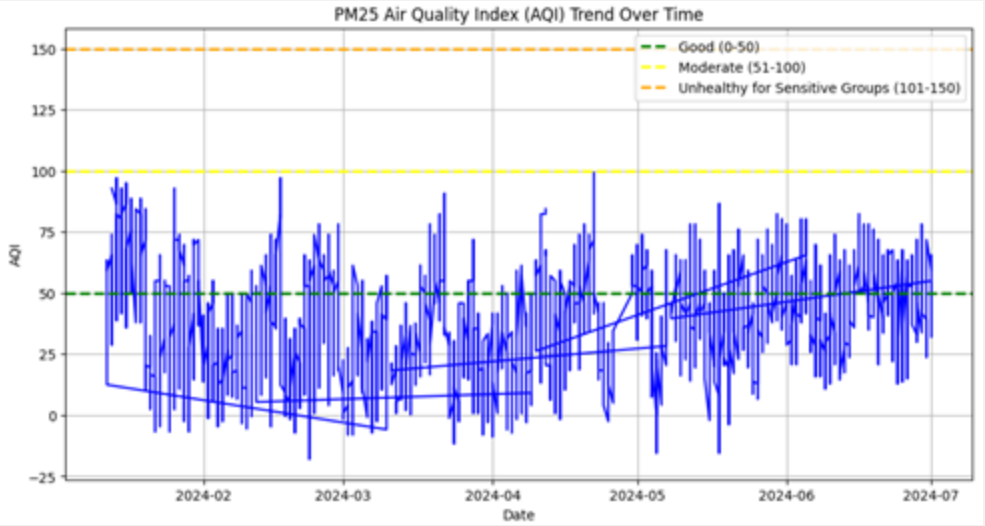
BUCHAREST: AVERAGE AIR QUALITY TREND 6-MONTH CARBON MONOXIDE

Data shows CO levels, in µg/m³, vary significantly. Levels start below 700 µg/m³ in late January, with notable spikes to over 1000 µg/m³ on February 2nd and March 6th. Fluctuations are influenced by traffic, industrial activity, weather conditions, and residential heating.



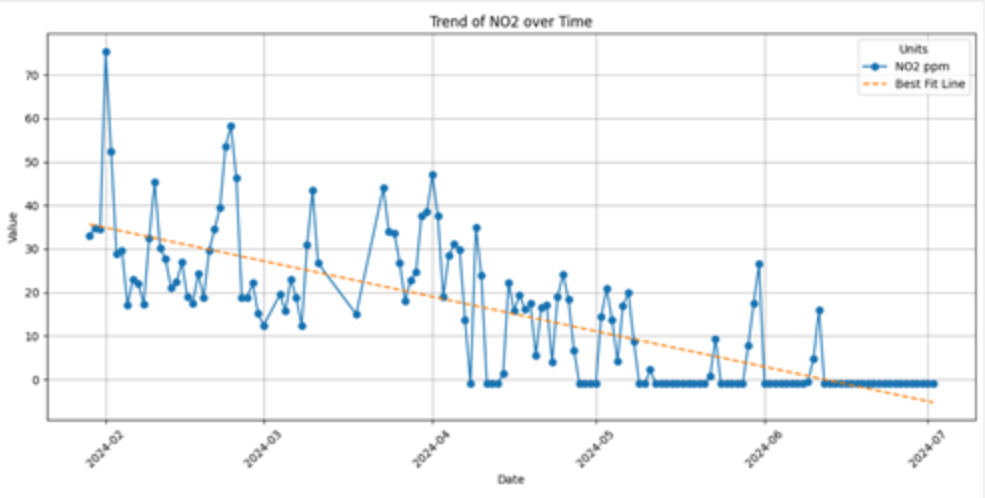
LOS ANGELES: AIR QUALITY INDEX PER PM25

Some of the AQI values of PM25 are below or above 50 but not passed 100.This shows that the air quality is acceptable.



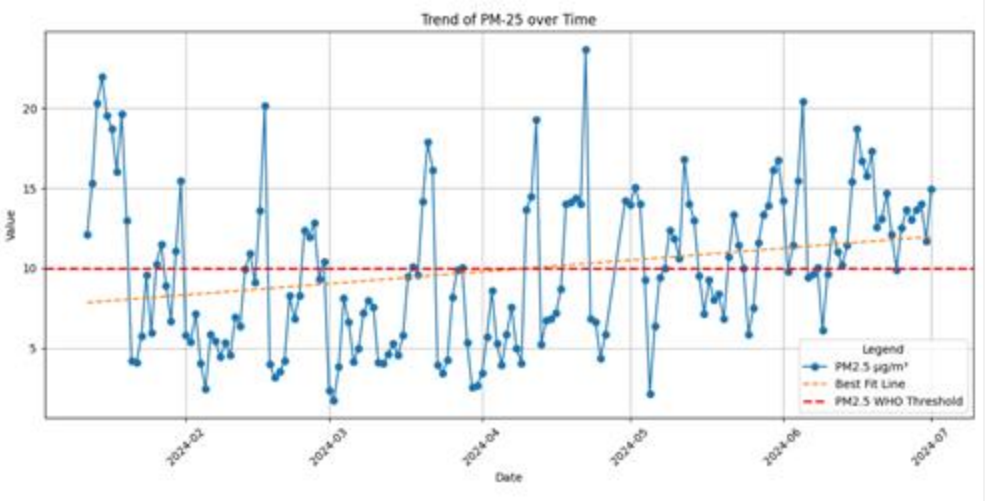
LOS ANGELES: AVERAGE AIR QUALITY TREND 6-MONTH NITROGEN DIOXIDE

- The NO2 concentration trends show fluctuations over time. The blue line represents NO2 in ppm, ranging from 0.000 to 0.030. The steady decline is most likely due to it moving from winter to summer and a reduction in heating equipment usage, which is a major source of air contamination.



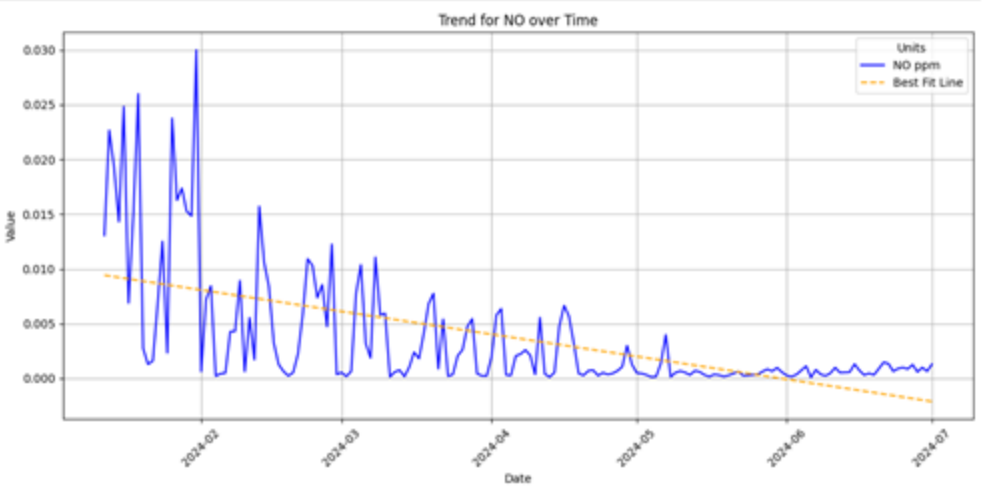
LOS ANGELES: AVERAGE AIR QUALITY TREND 6-MONTH PM-25

The plot displays PM2.5 levels (µg/m³) over time, with the WHO threshold of 10 µg/m³ marked by a red dashed line. PM2.5 values fluctuate significantly, with several peaks surpassing the threshold, indicating potentially hazardous air quality. Peaks in PM2.5 levels are likely linked to increased vehicle and industrial emissions, seasonal changes, local events such as wildfires, and weather patterns like temperature inversions. High activity periods, such as commuting and industrial operations, may also contribute to elevated PM2.5 levels.



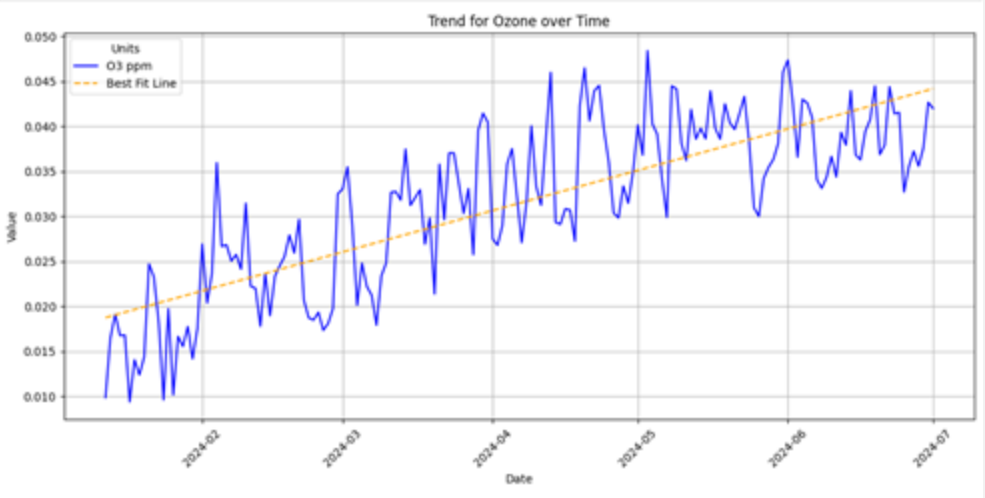
LOS ANGELES: AVERAGE AIR QUALITY TREND 6-MONTH NITROGEN OXIDE

The NO concentration trends show fluctuations over time. The blue line represents NO in ppm, ranging from 0.000 to 0.030. The steady decline is most likely due to it moving from winter to summer and a reduction in heating equipment usage, which is a major source of air contamination.



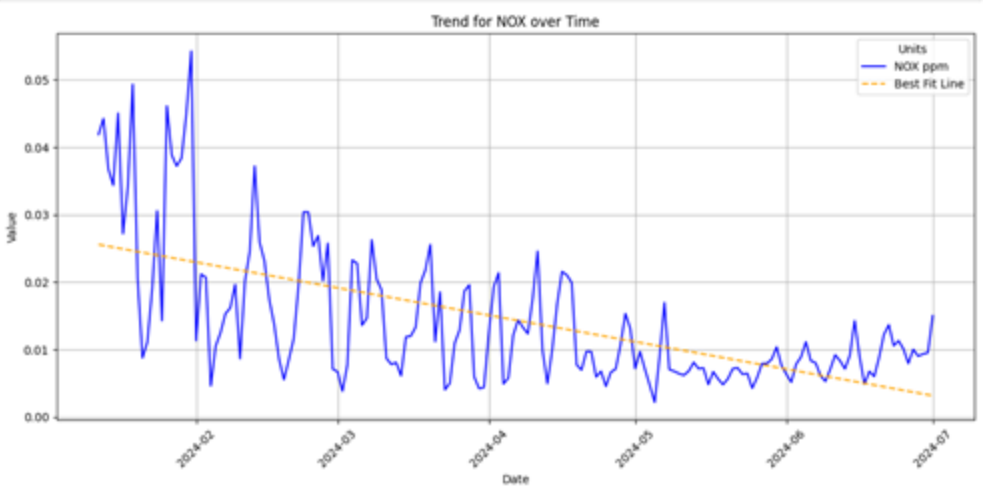
LOS ANGELES: AVERAGE AIR QUALITY TREND 6-MONTH OZONE

- Tropospheric ozone levels have increased over the observed period, with significant rises in the lower troposphere (950–800 hPa) in regions such as East Asia, the Persian Gulf, India, northern South America, the Gulf of Guinea, and Malaysia/Indonesia. In contrast, weaker trends are noted above North America, Europe, and high latitudes in both hemispheres.



LOS ANGELES: AVERAGE AIR QUALITY TREND 6-MONTH NITRIC OXIDE

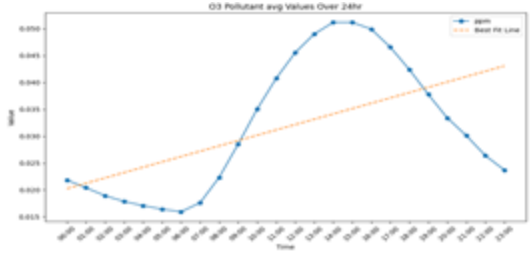
The NOX concentration trends show fluctuations over time. The blue line represents NOX in ppm, ranging from 0.000 to 0.050. The steady decline is most likely due to it moving from winter to summer and a reduction in heating equipment usage, which is a major source of air contamination.

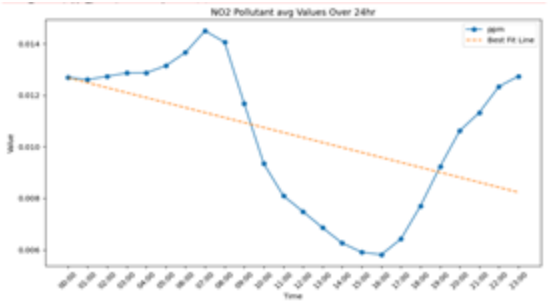


HOURLY PATTERNS AND EVENTS ANALYSIS

- **Findings:**
  - Higher AQI levels observed during peak hours (morning and evening) in Los Angeles.
  - Increased AQI levels during major events like New Year's Eve and local festivals.

- **Hourly Variability:**
  - AQI peaks during rush hours.
  - **Event Impact:** Significant AQI increases during holidays and special events due to higher activity levels





HOURLY ANALYSIS LA POLLUTANTS

-NO2 (µg/m³):
- Peaks slightly above 11 µg/m³ early in the day.
- Decreases to around 8 µg/m³ in the late afternoon.
- Rises again towards the evening.

PM2.5 (µg/m³):
- Peaks slightly above 11 µg/m³ early in the day.
- Drops to about 8 µg/m³ in the afternoon.
- Increases again towards the evening.

Ozone (O3) (ppm)
- Starts around 0.02 ppm.
- Peaks at about 0.05 ppm from 06:00 to 12:00.
- Decreases back to 0.02 ppm by evening.

NOx (ppm):
- Starts at approximately 0.015 ppm.
- Peaks around 0.03 ppm early in the morning.
- Decreases to a low of 0.01 ppm by 13:00.
- Rises again towards the evening.

Nitric Oxide (NO) (ppm):
- Starts around 0.015 ppm.
- Peaks at about 0.03 ppm from 05:00 to 07:00.
- Drops to around 0.01 ppm by 13:00.
- Increases towards the evening.

POSSIBLE REASONS

NO2:	- Higher levels due to increased traffic and industrial activities during the day.	- Influenced by temperature inversions and wind patterns.	PM2.5:	- Fluctuates with temperature inversions trapping pollutants and wind speeds dispersing them.
Ozone (O3):	- Rises due to sunlight-driven photochemical reactions.	- Decreases as sunlight fades, with precursor pollutants from traffic and industry contributing.	NOx:	- Peaks in the morning due to high traffic and industrial emissions.
- Rises again in the evening with increased traffic and heating activities.	Nitric Oxide (NO):	- Increases in the morning with high traffic and industrial activities.	- Decreases through the day, then rises again in the evening due to traffic and heating.	

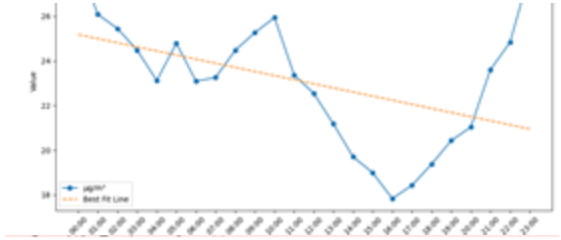
HOURLY ANALYSIS OF BUCHAREST POLLUTANTS

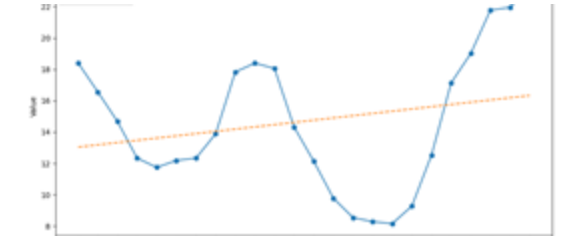
Trends:

- CO: Peaks around 8:00 AM and midnight; low in early morning and late evening.
- NO2: High at midnight, drops in early morning, peaks around 10:00 AM, rises by midnight.
- PM10: Peaks around 10:00 AM and midnight; lowest in early afternoon.
- SO2: Starts high at midnight, dips in early morning, peaks in afternoon, and stabilizes with minor fluctuations.

Likely Causes:

- CO: Morning rush hour, temperature inversions, and wind speed impact dispersion.
- NO2: Morning rush hour, temperature inversions, and wind variations.
- PM10: Increased vehicle and industrial emissions, temperature inversions, and wind speed changes.
- SO2: Afternoon emissions, morning temperature inversions, and wind speed effects.





HOURLY ANALYSIS OF ALL POLLUTANTS: MOLINA DE SEGURA

Trends:

Ozone (O3): Peaks sharply around 14:00, with AQI above 80.

PM2.5: Shows a smaller peak around midday, not exceeding an AQI of 20.

NO2: Remains stable and low throughout the day.

PM10: Also remains stable and low throughout the day.

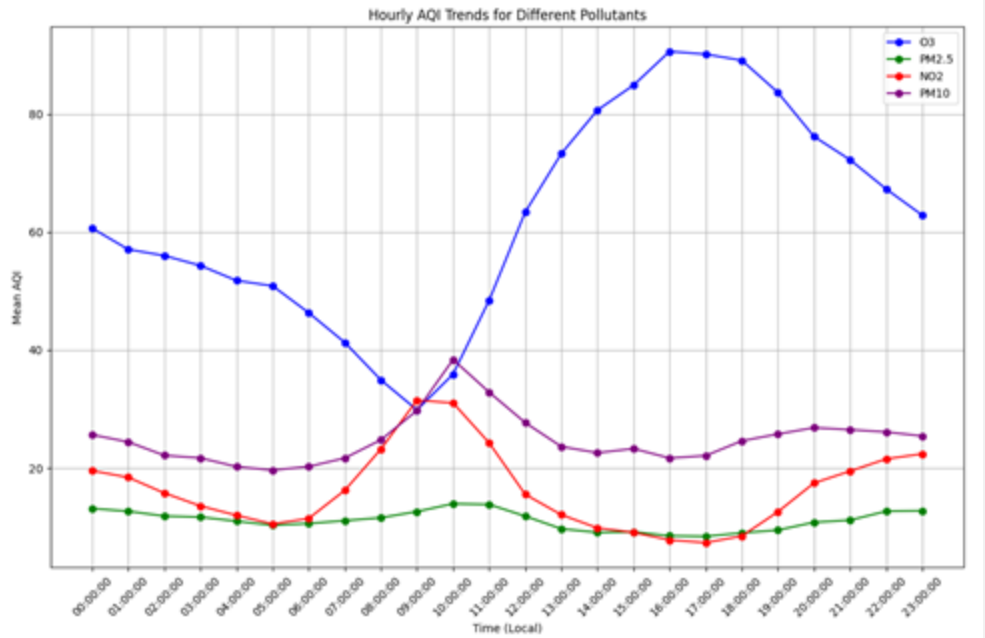
Likely Causes:

Ozone (O3): Midday peak due to sunlight-driven photochemical reactions with NOx and VOCs.

PM2.5: Slight midday peak from increased vehicular and industrial emissions.

NO2: Consistent levels due to steady sources like traffic and industrial operations.

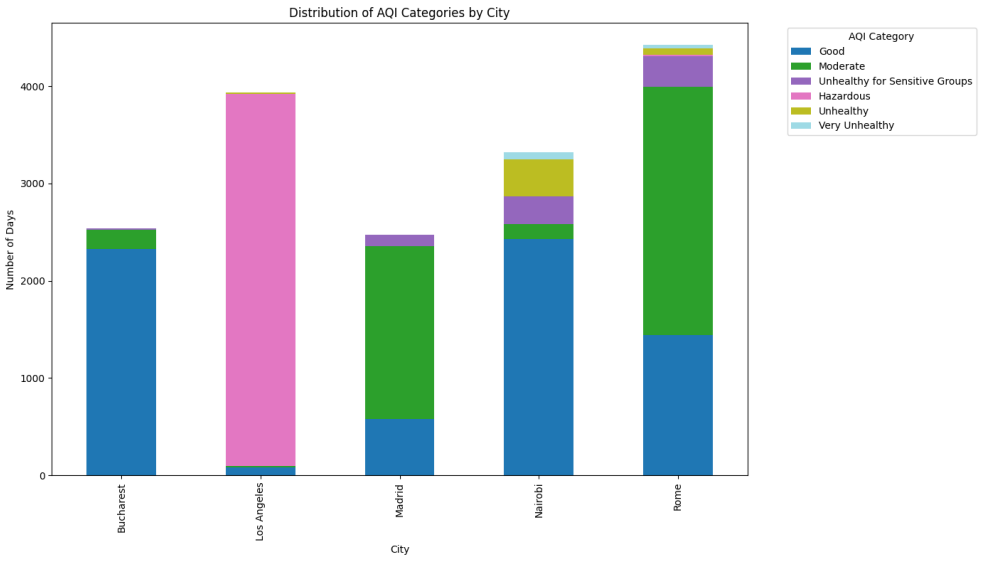
PM10: Stable levels from ongoing sources such as construction and road dust.



COMPARATIVE ANALYSIS ACROSS CITIES

DISTRIBUTION OF AQI BY CITY

- **Bucharest**: Predominantly in the 'Good' category with a small proportion of 'Moderate' days.
- **Los Angeles**: Mostly 'Hazardous' days, with minor contributions from other categories
- **Madrid**: Largely 'Moderate', followed by 'Good' and a smaller number of 'Unhealthy for Sensitive Groups' days.
- **Nairobi**: Primarily 'Good', with significant 'Moderate' days and a few days in each of the higher risk categories
- **Rome**: Mixed distribution with a large proportion of 'Moderate' days, followed by 'Good', and some days in the higher risk categories.

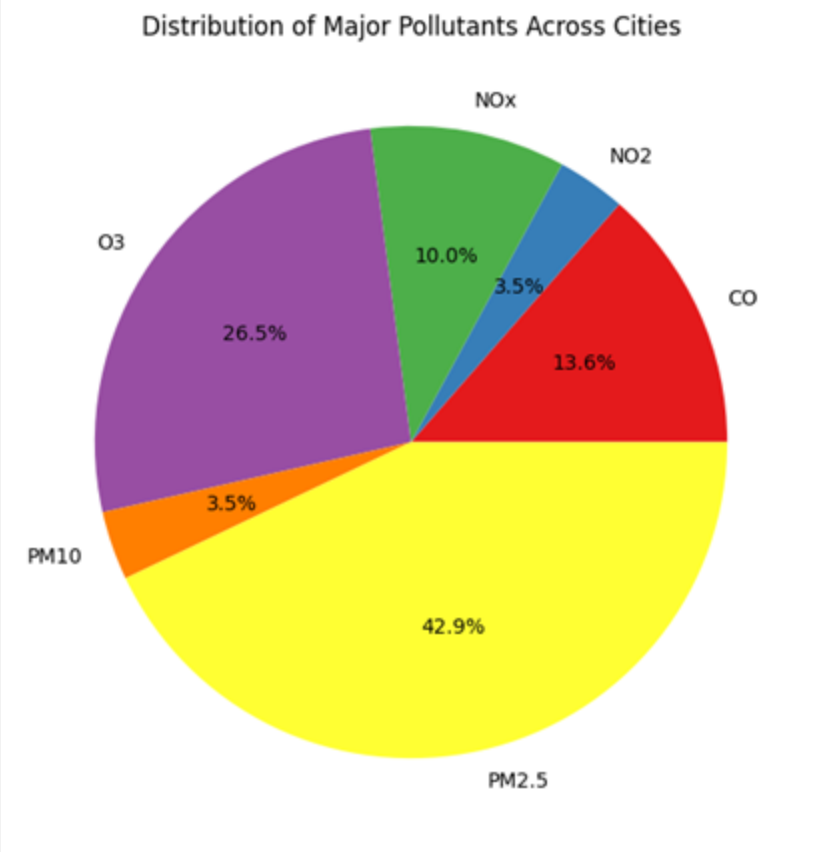


DISTRIBUTION OF MAJOR POLLUTANTS ACROSS CITIES

The pie chart shows the percentage of distribution.

The distribution shows that PM2.5 is the most produced pollutant across cities with a 42.9%.

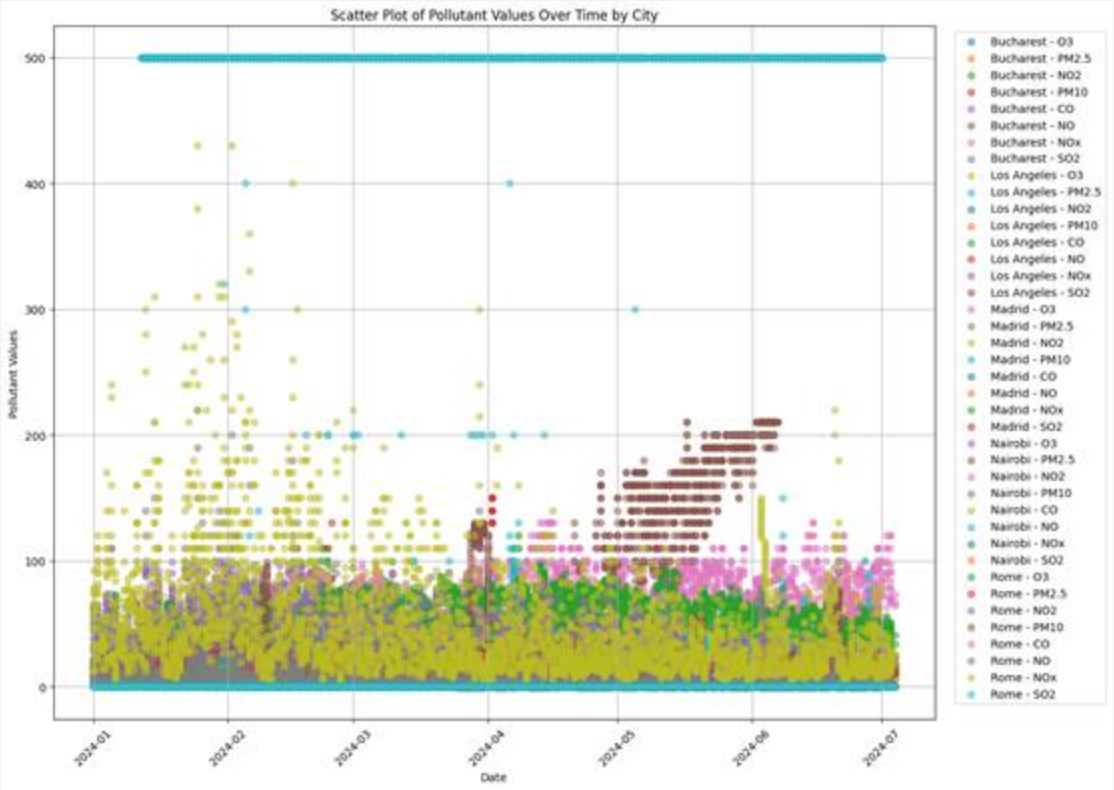
Whereas, the least produced pollutant are both PM10 and NO2 with 3.5%

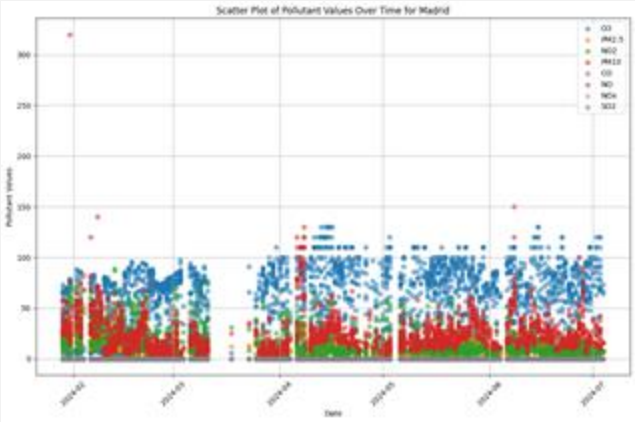


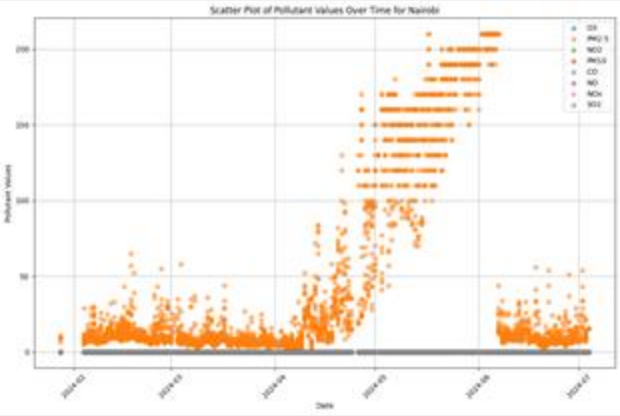
POLLUTANT VALUES ACROSS CITIES OVER TIME

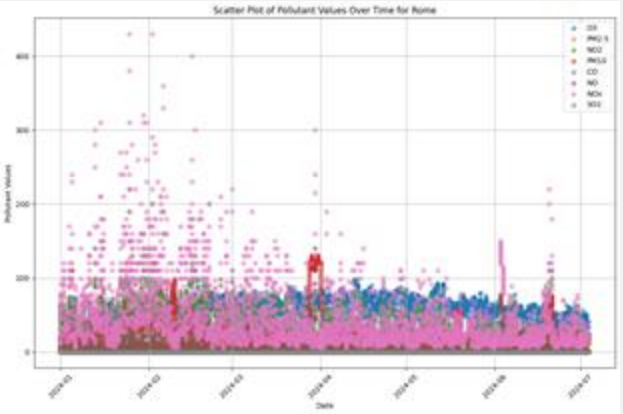
The distribution shows most cities have acceptable air quality as their most of their air quality values are between 0 to 100.

Just a few are above 100. This indicates unhealthy air quality that may cause health problems.









POLLUTANT VALUES ACROSS ROME, MADRID AND NAIROBI OVER 6 MONTHS

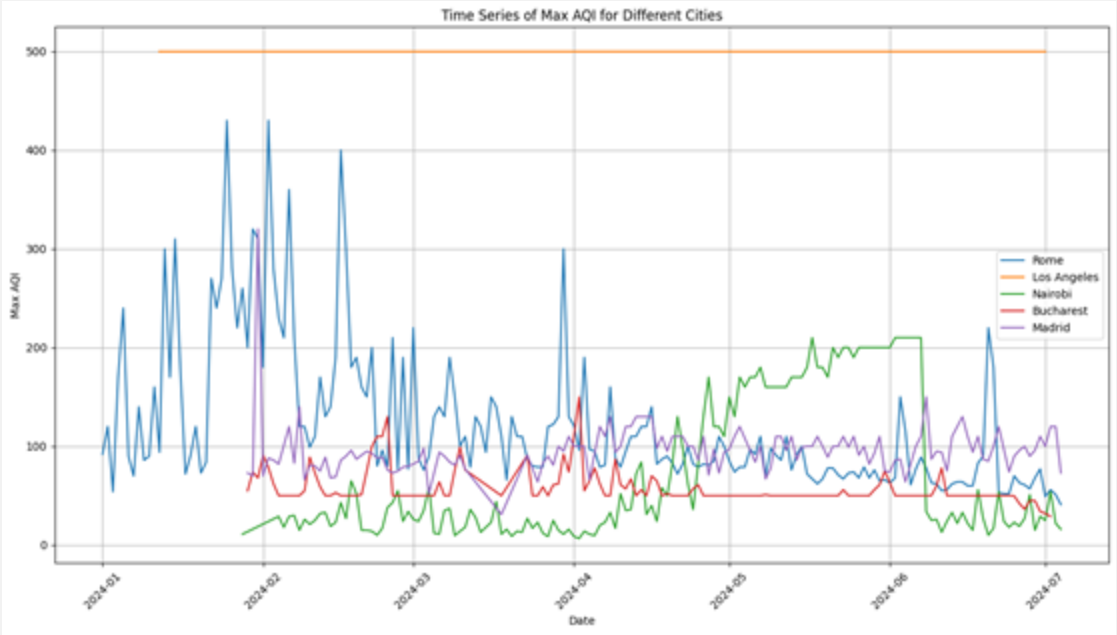
Madrid and Rome have most for its pollutant values below 150 and 200 respectively. This indicates that the cities air quality shows some members of the public might experience health effects.

For Nairobi, from 2024-02 to 2024–05, have good air quality because most AQI values are below 50. However, between 2024-05 and 2024-06, there was an increase of the AQI values to 200, this could cause the public to experience health effects as the air quality is unhealthy. In addition,after 2024-06, it drops back to below 50.

06- TIME SERIES ANALYSIS

TIME SERIES MAX AQI FOR COMBINED CITIES

The time series graph show that between 2024-01 and 2024-03, the AQI for Rome was above 300(Mostly on 2024-02). This is hazardous air quality and everyone in the city is more likely to affected.



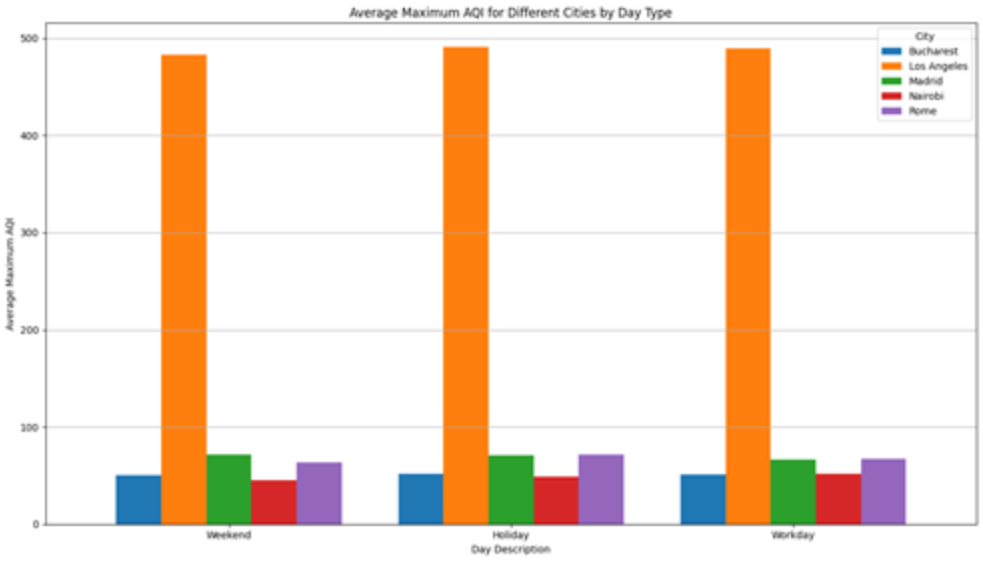
RELATIVE BETTER AIR QUALITY BASED ON THE TSA

- **Rome and Los Angeles** show the most concerning air quality trends, with frequent high AQI values that indicate significant health risks.
- **Nairobi** experiences variability in air quality but generally maintains lower AQI values than Rome and Los Angeles.
- **Bucharest and Madrid** maintain relatively better air quality but still experience occasional periods of elevated AQI values
- ***Recommendations***
- **Rome & Los Angeles**: Implement stricter emission controls, enhance air quality monitoring, issue public health advisories.
- **Nairobi**: Address industrial emissions and traffic congestion.
- **Bucharest & Madrid**: Maintain monitoring, target interventions during pollution spikes.

AVERAGE MAXIMUM AQI FOR DIFFERENT CITIES BY DAY TYPE

The bar chart shows that the city LA, has high average maximum AQI in all day descriptions(Weekend,Holiday and Weekday).

Whereas, Nairobi has the lowest.



HEATMAP OF ALL AREAS

- **Heatmap of Average Monthly AQI_PM2.5 by City**:
  1. **Bucharest**: Consistently shows zero AQI_PM2.5 values, which might indicate a lack of PM2.5 data collection or negligible PM2.5 levels.
  2. **Los Angeles**: Exhibits very high AQI_PM2.5 levels consistently, with values ranging from approximately 473.5 to 499.7, indicating significant PM2.5 pollution.
  3. **Madrid**: Shows moderate AQI_PM2.5 levels, with the highest value in April 2024 (75.4) and a notable peak in May 2024 (207.3).
  4. **Nairobi**: Displays varying AQI_PM2.5 levels, with a peak in May 2024 (207.3), suggesting an instance of high pollution during that month.
  5. **Rome**: Exhibits relatively lower AQI_PM2.5 levels, with the highest value in January 2024 (60.5).

HEATMAP OF AVERAGE MONTHLY AQI_PM10 BY CITY

- **Implications**:
  - **Air Quality Concerns**: Cities like Los Angeles and Nairobi show periods with very high PM2.5 pollution levels, which could have serious health implications.
  - **Data Completeness**: The zero values for Bucharest suggest either exceptionally low PM2.5 pollution or potential gaps in data reporting.
  - **Seasonal Variations**: Some cities exhibit seasonal patterns, such as higher PM2.5 levels in certain months (e.g., Nairobi in May and Madrid in May), which could be due to specific events or conditions affecting air quality.

INTERPRETATION AND DISCUSSION

For sure if we give a chance for you to ask the first question will be, which city is better to live?

Bucharest: Predominantly 'Good' days.

Los Angeles: Mostly 'Hazardous' days.

Madrid: Largely 'Moderate', followed by 'Good' and a smaller number of 'Unhealthy for Sensitive Groups' days.

Nairobi: Primarily 'Good', with significant 'Moderate' days and a few days in each of the higher risk categories.

Rome: Mixed distribution with a large proportion of 'Moderate' days, followed by 'Good', and some days in the higher risk categories.

So where??? As data analyst our task is to give you a data driven knowledge; we don't force anyone.

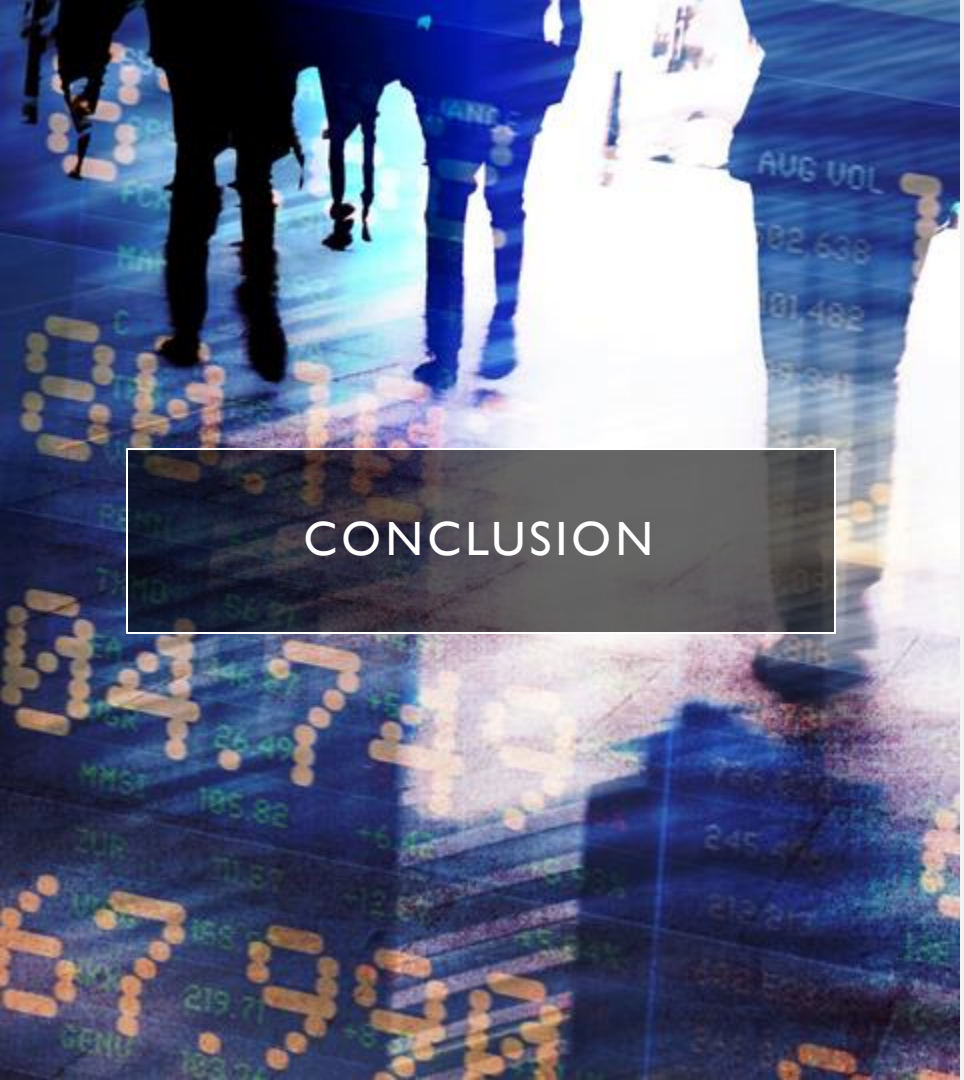
We can suggest recommendations.

RECOMMENDATIONS:

- **Los Angeles**:
- **Reduce PM2.5 Levels**: Introduce stricter emission controls for industries and vehicles, promote the use of cleaner fuels and technologies, and encourage the use of air purifiers indoors.
- **Bucharest**:
- **Reduce CO Emissions**: Implement measures to reduce carbon monoxide levels, such as improving vehicle emission standards, promoting the use of electric vehicles, and regulating industrial emissions.
- **Madrid**:
- **Reduce Ozone and PM10 Levels**: Implement measures to reduce precursor emissions (NOx and VOCs), promote public transportation and non-motorized transport, and enhance green spaces to mitigate ozone formation.
- **Nairobi**:
- **Reduce PM2.5 Levels**: Regulate industrial emissions, improve waste management practices to reduce burning, and promote cleaner cooking methods to reduce indoor and outdoor PM2.5 levels.
- **Rome**:
- **Reduce Ozone and NOx Levels**: Implement low-emission zones, encourage the use of public transportation, and promote electric and hybrid vehicles to reduce nitrogen oxides and ozone levels.

RECOMMENDATIONS

- **Air Quality Management:**
  - **Suggested Actions:** Implement stricter traffic regulations and industrial controls in Los Angeles. Promote cleaner technologies and better traffic management.
  - **Policy Implications:** Recommendations for improved air quality management based on findings.
- **Recommendations for Improving Air Quality:**
  - **Public Health Measures:** Enhance public awareness and implement measures to reduce pollution sources.
  - **Regulatory Changes:** Advocate for policies aimed at reducing industrial emissions and traffic congestion.



CONCLUSION

- **Summary of Key Findings:**
  - **Recap:** Los Angeles shows higher and more variable AQI levels compared to Bucharest, influenced by seasonal and event-based factors.
- **Implications:**
  - **Public Health:** Higher AQI levels in Los Angeles pose greater health risks.
  - **Policy:** Findings highlight the need for targeted air quality management strategies.
- **Final Thoughts:**
  - **Significance:** The study underscores the importance of addressing air quality issues through informed policies and public awareness.

WHAT WE LEARN FROM THE PROJECT

- Every group member knows at least
  - How to download data from open sources like OpenAQ
  - How to import pandas
  - How to install important packages and then how to import important libraries to work with Data Frame (Pandas, matplotlib and seaborn
  - Data Exploration and Analysis [Calculating summary statistics and understanding the distribution of data]
  - Visualization and Interpretation of findings

- Working in collaboration in remote GitHub
- A good Clue about remote work!!!!