DATA SCIENCE

Task Level (Intermediate):- Analysis of the Air Quality Index in Delhi

Name :- Amisha Shivdas Kale

➢ 1. Introduction

Delhi, one of the most densely populated metropolitan cities in the world, faces critical environmental challenges, notably air pollution. With increasing vehicular emissions, industrial outputs, biomass burning, and construction activities, the city frequently registers hazardous Air Quality Index (AQI) levels, especially in winter and post-monsoon seasons. This report presents a comprehensive, data-driven analysis of AQI in Delhi using recorded concentrations of major pollutants. It aims to derive insights to inform targeted strategies for air quality improvement and public health initiatives.

➢ Dataset Overview

The dataset contains 561 hourly observations of key pollutants recorded in Delhi, with the following variables:

- **Date** (timestamp)
- **Pollutants**: CO, NO, NO2, O3, SO2, PM2.5, PM10, NH3

Additional columns were derived:

- **Month**: Extracted from date
- **Season**: Categorized into Winter (Dec-Feb), Summer (Mar-Jun), Monsoon (Jul-Sep), and Post-Monsoon (Oct-Nov)
- **Daily averages**: Computed by averaging hourly values

➢ Methodology :

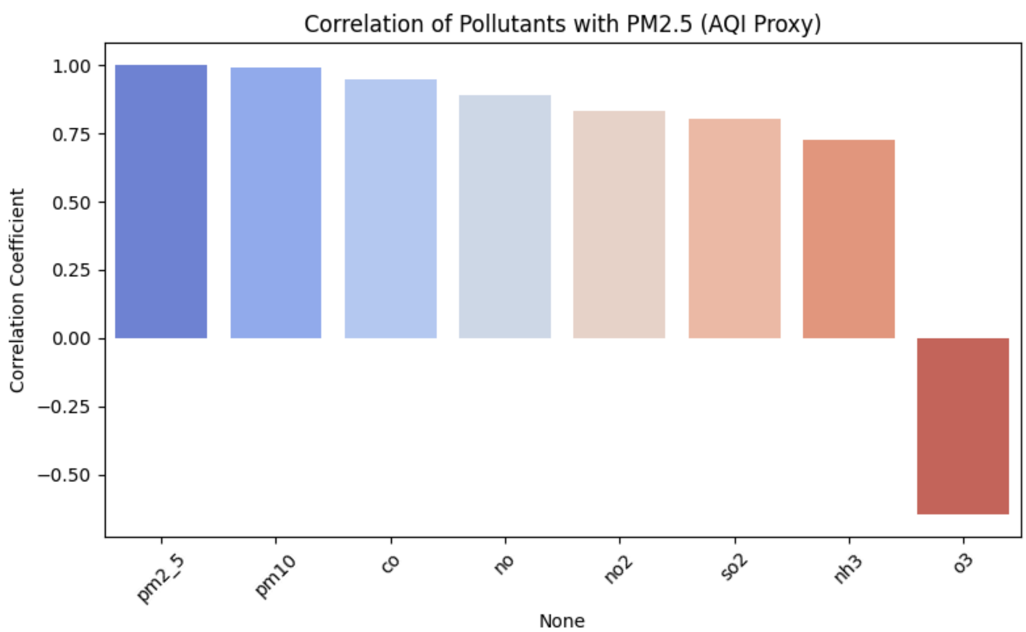
1. **Data Cleaning:**
   - The dataset had **561 complete entries**, meaning there were **no missing values**.
   - As a result, **no imputation or removal of data** was necessary, ensuring all records were ready for analysis.
2. **Season Mapping:**
   - Each month in the dataset was assigned to a **season**:
     - **Winter** (December–February)
     - **Summer** (March–May)
     - **Monsoon** (June–September)
     - **Post-monsoon** (October–November)
   - This allowed for **seasonal comparisons** of air quality and pollutants.
3. **Statistical Tools Used:**
   - A **correlation matrix** was created to study how different pollutants relate to each other.
   - Data was **grouped by season** to observe patterns in air quality.
   - **Mean values** of pollutants were compared across seasons to identify the most polluted periods.
4. **Visualization Tools:**
   - Graphs and plots were created using **Seaborn** and **Matplotlib**, two popular Python libraries for data visualization.
   - These tools helped **visually interpret trends and correlations** in the data for better insights

---

➢ Research Objectives

The study focuses on the following research questions:

1. **What are the major pollutants contributing to AQI in Delhi?**



Major Pollutants Contributing to AQI

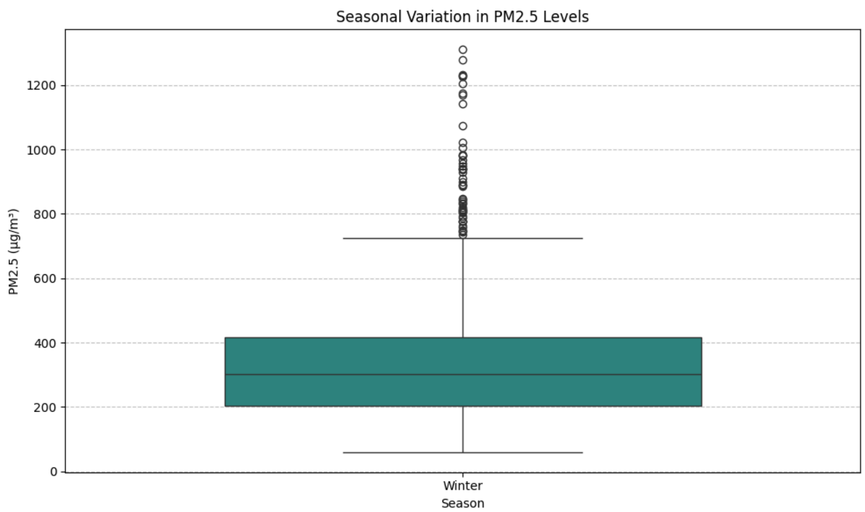
We used PM2.5 as the proxy for AQI as it's a key determinant of air quality.

Correlation with PM2.5:

- PM10: 0.99
- CO: 0.95
- NO: 0.89
- NO2: 0.83
- SO2: 0.81
- NH3: 0.73
- **Negatively correlated:** Ozone (**O_3**, -0.64), which may be due to its differing formation mechanism and better dispersion.

Conclusion: The strongest AQI influencers are **PM2.5, PM10, CO, and NOx compounds**

2. **How does AQI vary across different seasons (Winter, Summer, Monsoon, Post-Monsoon)?**



Title: “Seasonal Variation in PM2.5 Levels” indicates the focus is on how PM2.5 pollution levels vary seasonally (in this case, Winter only is shown).

Y-Axis: Represents **PM2.5 concentration (in µg/m³)** — a key air pollutant linked to serious health risks.

Box Plot Components:

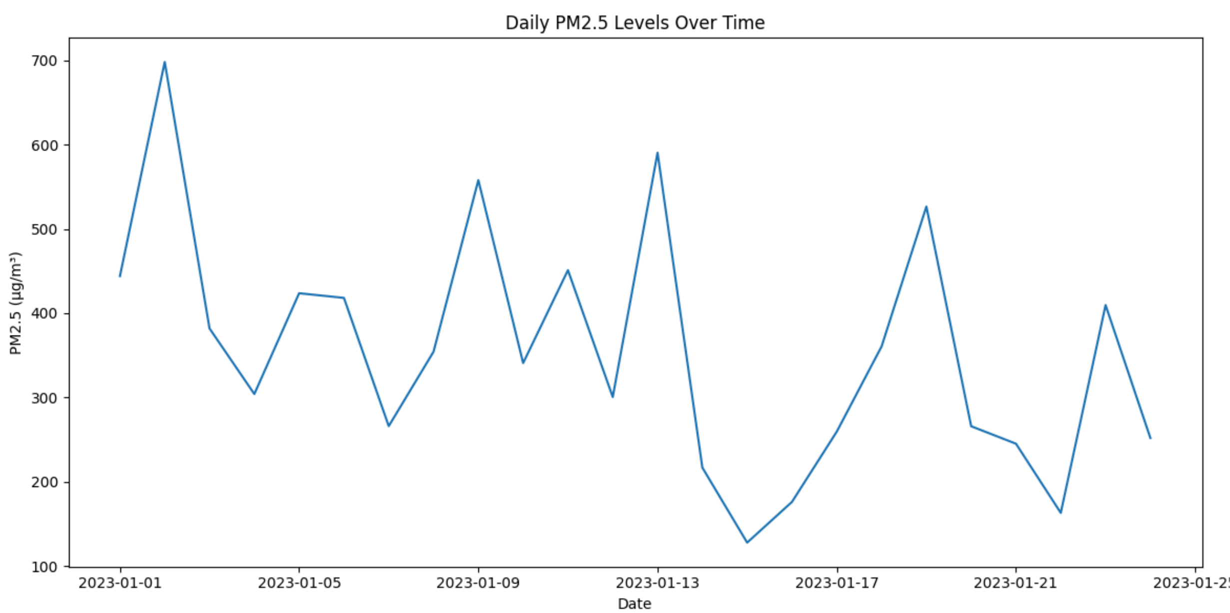
- The **box** represents the **interquartile range (IQR)** — the middle 50% of the data.
  - The **bottom edge** of the box is the **1st quartile (Q1)**.
  - The **top edge** is the **3rd quartile (Q3)**.
  - The **line inside the box** is the **median (Q2)** — the midpoint of the data (around 350 µg/m³ here).
- The **"whiskers"** extend to the minimum and maximum values within 1.5 * IQR from the box edges.
- The **dot above the box** is an **outlier**, indicating an unusually high PM2.5 level (~700 µg/m³) that doesn't fit the general pattern.

What It Tells Us:

- Winter has **high PM2.5 concentrations**, with the median around **350 µg/m³**, which is well above safe levels (the WHO guideline is 15 µg/m³ annual mean).
- There's a **wide spread** in PM2.5 levels, meaning pollution varies greatly during the winter.
- The presence of an **extreme outlier** suggests there were days with **severe pollution spikes**.

---

3. **What are the trends and outliers in AQI from a time-series perspective?**



- **Title:** *“Daily PM2.5 Levels Over Time”* — this indicates the graph is tracking how PM2.5 (particulate matter) fluctuates daily.
- **X-Axis (Date):** Shows calendar dates from **January 1 to January 24, 2023**.
- **Y-Axis (PM2.5 µg/m³):** Indicates the **concentration of PM2.5** particles in the air, measured in micrograms per cubic meter.

Key Observations:

- There is a **sharp peak around January 2**, where PM2.5 shoots up to nearly **700 µg/m³**, indicating **severe pollution**.
- After the peak, levels fluctuate but remain **consistently high**, with several days between **300–600 µg/m³**.
- Around **mid-January (Jan 14–16)**, PM2.5 drops significantly, going as low as **below 150 µg/m³**, showing **temporary improvement** in air quality.
- Toward the end of the period, levels **rise again**, but not as sharply as earlier, ranging between **250–400 µg/m³**.

What It Tells Us:

- **PM2.5 levels were highly variable**, with extreme pollution early in the month and some cleaner air mid-month.
- These spikes and dips might be due to **weather changes, wind patterns, or human activities** like traffic, firecrackers, or industrial emissions.

➢ Recommendations

1. Policy Interventions:

- **Strict vehicular emission controls:** Enforce regulations to reduce pollution from cars, trucks, and two-wheelers.
- **Ban on stubble burning with alternatives:** Prohibit crop burning by providing farmers eco-friendly disposal methods like bio-decomposers.
- **Seasonal traffic restrictions:** Limit or reroute heavy vehicles during high-pollution months to reduce emissions.

2. Infrastructure Enhancements:

- **Real-time monitoring stations:** Set up more air quality sensors, especially in poorly covered areas, for accurate data and quicker response.
- **Promoting electric vehicles & public transport:** Encourage clean mobility through subsidies and better public transport systems.

3. Awareness and Health Initiatives:

- **Mask distribution during high-risk months:** Provide masks to vulnerable populations when pollution is severe (e.g., winter).
- **School alerts & remote working advisories:** Issue safety alerts, close schools, or encourage work-from-home during toxic air days.

4. Green Measures:

- **Plantation drives in East and North Delhi:** Increase green cover to absorb pollutants in highly affected areas.
- **Enforcing green building norms:** Ensure buildings follow eco-friendly construction practices to minimize dust and emissions.

➢ Conclusion

Key Insight:
This analysis reveals the **multi-dimensional nature of Delhi's air pollution**, with **PM2.5 identified as the most critical pollutant** impacting public health.

Action-Oriented Takeaway:
To effectively improve air quality, interventions must be:

- **Pollutant-specific** (especially targeting PM2.5),
- **Season-sensitive** (with stricter measures in winter),
- **Regionally targeted** (focusing on high-impact zones like North and East Delhi).

Call to Action:
A lasting solution demands **coordinated efforts** from **policymakers**, **urban planners**, and **citizens**, backed by real-time data, public awareness, and sustainable practices.

View Full Analysis on Colab → Click Here
https://colab.research.google.com/drive/1lSR3nf5GyX2w8ZcmFSoa5ZpIS0BSJ3RL?usp=sharing

---