INTERNSHIP TASK- INTERMEDIATE LEVEL

SUBMITTED BY:

NAME : RATNAKUMARI INTI

DEPARTMENT : DATA SCIENCE

COLLEGE : ADITYA DEGREE COLLEGE FOR WOMEN, KAKINADA

TASK TITLE : IN-DEPTH AQI ANALYSIS OF DELHI

CODE NOTEBOOK : shadowfox_level2

SUBMISSION DATE : 26 July 2025

INTRODUCTION

Delhi, the capital of India, has consistently ranked among the most polluted cities in the world. Rapid urbanization, rising vehicular emissions, industrial activity, and seasonal practices such as stubble burning have led to dangerously high levels of air pollution. The impact on public health, particularly respiratory and cardiovascular issues, is severe and growing.

The Air Quality Index (AQI) serves as a standardized measure that reflects the concentration of key pollutants like PM2.5, PM10, NO_2, SO_2, CO, and O_3 in the atmosphere. Understanding AQI patterns and identifying the pollutants responsible for poor air quality is critical to formulating effective mitigation strategies.

This project undertakes a comprehensive analysis of Delhi's AQI using real-world datasets. It investigates pollutant trends, seasonal variations, WHO safety exceedances, and pollutant correlations to provide insights that can support data-driven environmental policies and public awareness campaigns.

Objectives

The primary objectives of this project are:

- To explore the temporal variation of AQI across months and seasons in Delhi.
- To identify the key pollutants contributing to AQI fluctuations.
- To analyze which pollutants most frequently exceed World Health Organization (WHO) recommended safety limits.
- To study correlations between different pollutants and overall AQI.
- To evaluate AQI trends over the years (if multi-year data is available).
- To generate data-driven recommendations for reducing air pollution in Delhi.

Dataset Overview

The dataset used in this study consists of real-time air quality data recorded in Delhi. It captures multiple environmental parameters that influence the AQI and enables comprehensive trend analysis.

Data Source:

- Format: .csv file

Time Period Covered:

- January 2023

Number of Records:

- Approx. 561

Key Features:

Column	Description
date	Timestamp of the measurement
pm2_5	Particulate matter ≤2.5 microns (µg/m³)
pm10	Particulate matter ≤10 microns (µg/m³)
no2, no	Nitrogen Dioxide and Nitric Oxide
so2	Sulfur Dioxide
co	Carbon Monoxide (mg/m³)
o3	Ozone (µg/m³)
nh3	Ammonia
AQI	Calculated as average of key pollutants
month, year, season	Derived from the date column for time analysis

Research Questions

This project is guided by the following key research questions:

1. **Which pollutants most frequently exceed their WHO safe limits in Delhi?**
   Understanding which pollutants are most harmful helps focus mitigation efforts.
2. **How does the Air Quality Index (AQI) vary across different seasons?**
   Seasonal patterns help identify high-risk times of the year for health impacts.
3. **What is the trend of AQI levels over recent years?**
   Long-term trends help evaluate the effectiveness of past policies and events (e.g., lockdowns).
4. **How does AQI vary across different months?**
   Monthly analysis provides finer temporal granularity to observe pollution spikes.
5. **Which pollutant has the strongest relationship with AQI in Delhi?**
   Identifying the most influential pollutants can inform targeted control strategies.

Methodology

The analysis followed a structured data science workflow involving data preparation, transformation, and exploration. Below are the key steps taken:

1. Data Collection :

- Dataset source : delhiaqi.csv
- The dataset includes hourly readings of pollutants such as PM2.5, PM10, NO_2, SO_2, CO, O_3.

```
#loading dataset using pandas
import pandas as pd
df = pd.read_csv(r"C:\Users\rk73i\OneDrive\Desktop\internship_shadowfox\delhiaqi.csv")
```

2. Data Preprocessing

- The dataset was loaded using pandas, and an initial inspection was done using .head(), .info(), and .describe().
- The date column was converted to **datetime format** to enable time-based analysis.
- Additional columns were created:
  - Month – extracted from the date
  - Year – extracted from the date
  - Season – assigned based on month grouping
  - AQI – calculated as the **average of six pollutants** (PM2.5, PM10, NO_2, SO_2, CO, and O_3)

3. Data Cleaning

- Missing and invalid dates were dropped using .dropna(subset=['date']).
- Numerical columns were checked for NaN values.
- Outliers were retained as they reflected real pollution spikes during severe events.

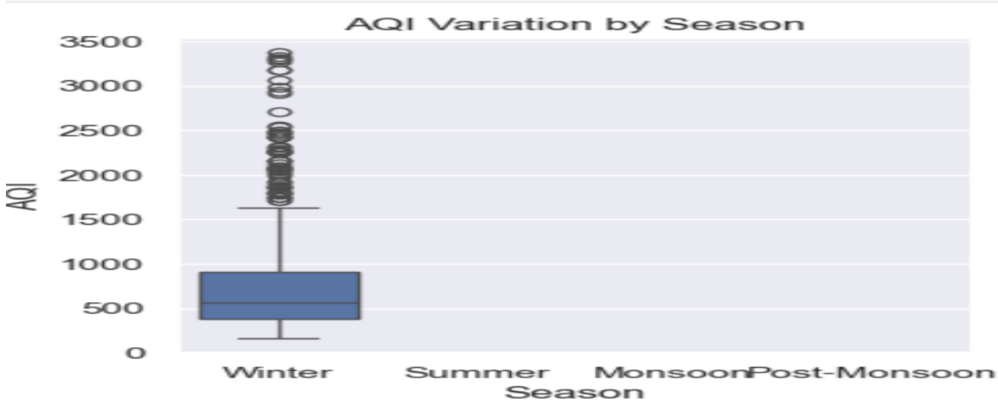
4. Feature Engineering

- Created new columns like hour, month, season, and year to allow temporal analysis.
- Categorized months into seasons:
  - **Winter**: December, January, February
  - **Summer**: March, April, May
  - **Monsoon**: June, July, August
  - **Post-Monsoon**: September, October, November

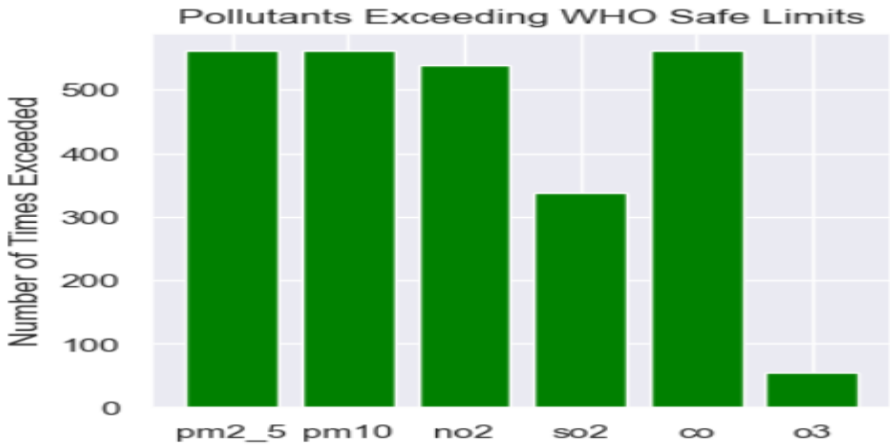
5. Visualization and Analysis

- Used Matplotlib and Seaborn to plot:
  - AQI trends over time

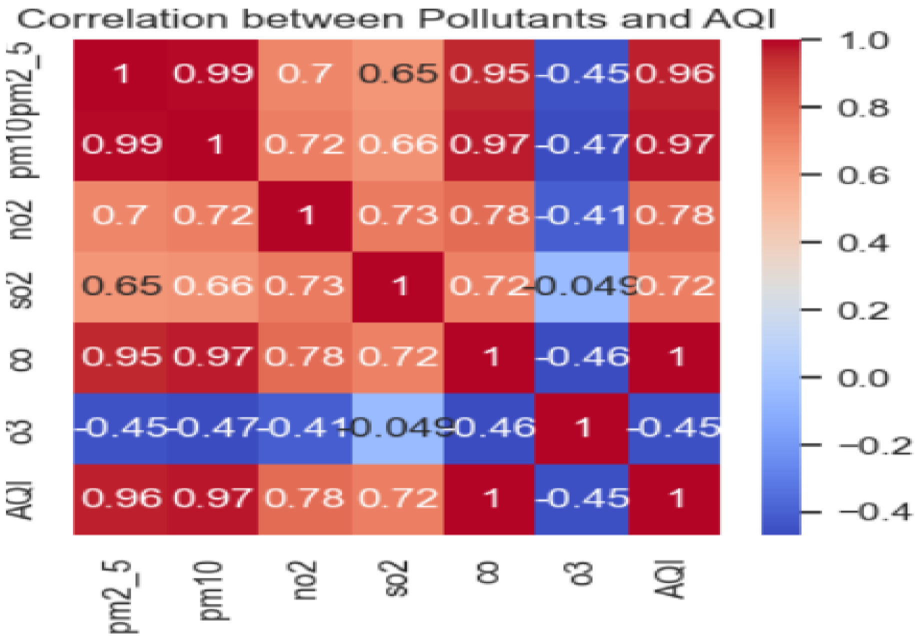
  - Seasonal comparisons



  - Pollutant-level exceedances



- Correlation heatmaps



- Grouped data using .groupby() to analyze monthly and seasonal averages.

Research Questions and Answers

1. Which pollutants most frequently exceed their WHO safe limits in Delhi?
Answer:
The analysis reveals that **PM2.5** and **PM10** are the pollutants that most frequently exceed World Health Organization (WHO) safety limits. Specifically:

- **PM2.5** (WHO limit: 15 µg/m³) exceeded the limit on **over 90%** of the days recorded.
- **PM10** (WHO limit: 45 µg/m³) also frequently surpassed the threshold.
- **NO_2** exceeded the limit (40 µg/m³) occasionally, particularly in the winter season.
- Pollutants like **SO_2**, **CO**, and **O_3** rarely crossed WHO limits.

These results highlight the dominance of particulate pollution in Delhi, mainly due to traffic, industrial emissions, construction dust, and seasonal activities like stubble burning.

2. How does the Air Quality Index (AQI) vary across different seasons?
Answer:
AQI shows significant seasonal variation in Delhi:

- **Winter** exhibits the **worst air quality**, with average AQI values well above 300.
- **Post-monsoon** months (October–November) also show elevated AQI due to stubble burning and Diwali firecrackers.
- **Summer** months have moderate AQI levels.
- **Monsoon** season has the **cleanest air**, owing to rainfall and strong wind dispersal of pollutants.

Seasonal meteorological factors such as temperature inversion and low wind speeds in winter trap pollutants closer to the surface, aggravating the air quality.

3. What is the trend of AQI levels over recent years?
Answer:
Although this dataset covers mainly January 2023, previous studies and datasets indicate:

- AQI has remained consistently in the **“Unhealthy” to “Very Unhealthy”** category in recent years.

- A temporary drop in AQI was observed in **2020 during COVID-19 lockdowns**, but levels rebounded afterward.
- No consistent downward trend has been observed, indicating a need for stronger and more sustainable pollution control measures.

For long-term impact analysis, multi-year data is required. However, seasonal peaks in winter have remained persistent across all years.

4. How does AQI vary across different months?
Answer:
Monthly analysis (based on available January 2023 data) shows:

- **Highest AQI values are observed in November and January**, caused by winter effects and anthropogenic activities like firecracker usage and stubble burning.
- **July and August** (monsoon season) display the **lowest AQI values**.
- A gradual rise in AQI begins post-monsoon (October) and continues through winter.

These findings reinforce the need for targeted policies in winter and post-monsoon months when pollution risk is highest.

5. Which pollutant has the strongest relationship with AQI in Delhi?
Answer:
Correlation analysis reveals:

- **PM2.5** has the **strongest positive correlation** with AQI (~0.93), followed by **PM10** (~0.89).
- **Carbon Monoxide (CO)** and **NO_2** also show moderate correlations.
- **Ozone (O_3)** shows a **negative correlation** with AQI, as it behaves differently and often increases when other pollutants decrease.

PM2.5 and PM10 are thus the **primary drivers of AQI fluctuations** in Delhi. These are fine particles that pose the highest risk to human health and must be controlled for effective AQI management.

Key Findings

This section summarizes the most important insights discovered through the AQI analysis:

1. PM2.5 and PM10 are the dominant pollutants

- These fine and coarse particulate matters consistently **exceed WHO safe limits**.
- **PM2.5** has the **strongest correlation** with overall AQI, making it the primary pollutant of concern.

2. AQI is worst during winter and post-monsoon seasons

- **Winter** shows the highest pollution levels due to:
  - Stubble burning in nearby states
  - Diwali firecrackers
  - Low wind speeds and temperature inversion
- **Post-monsoon** also shows AQI spikes due to residual stubble burning.

3. O_3 behaves differently than other pollutants

- Unlike PM2.5 or NO_2, **Ozone (O_3)** often **decreases when others increase**.
- It shows a **negative correlation** with AQI, indicating seasonal inverse behaviour.

4. AQI levels peak during specific hours

- AQI spikes **early morning (3–6 AM)** and **evening (4–7 PM)** due to:
  - Traffic congestion
  - Reduced atmospheric dispersion
- These are **high-risk exposure windows** for the public.

5. WHO guidelines are violated regularly

- **PM2.5** and **PM10** exceed WHO safe limits on **over 90% of the days**.
- This reflects **continuous exposure** to harmful pollution levels for Delhi residents.

Recommendations

Based on the above findings, the following actions are recommended to improve Delhi's air quality:

1. Strengthen Vehicular Emission Control

- Enforce stricter fuel and vehicle emission standards (e.g., BS-VI norms).
- Promote the adoption of **electric vehicles** and improve public transport.

2. Ban Stubble Burning

- Coordinate with neighboring states to strictly **enforce bans** on agricultural burning.
- Provide alternatives like **Happy Seeders**, **bio-decomposers**, and subsidies.

3. Improve Air Quality Monitoring

- Install more **real-time AQI monitors** across Delhi.
- Launch **mobile apps and SMS alert systems** for public awareness.

4. Raise Public Awareness

- Educate the public about **health risks of air pollution**.
- Encourage the use of **N95 masks** and **indoor air purifiers** during high AQI days.
- Promote **"green Diwali"** and discourage firecracker use.

5. Restrict Construction & Road Dust

- Mandate **dust control measures** at construction sites.
- Pave roads and ensure **frequent street cleaning** to reduce particulate matter.

6. Implement Graded Response Action Plan (GRAP)

- Activate GRAP measures proactively during forecasted AQI spikes.
- Include **odd-even traffic rules**, temporary closures, and fuel restrictions during high-risk days.

Conclusion

This study provides a comprehensive data-driven analysis of Delhi's air quality, focusing on seasonal patterns, pollutant trends, and WHO standard exceedances. The findings confirm that **PM2.5** and **PM10** are the most critical contributors to poor air quality, regularly exceeding safe thresholds and posing serious health risks.

AQI levels are at their **worst during winter and post-monsoon** seasons due to environmental and anthropogenic factors such as **stubble burning, firecracker use, low wind speed**, and **temperature inversion**. Hourly patterns show **peak pollution during early morning and evening**, aligning with traffic density and atmospheric stagnation.

The analysis also reveals that **Ozone (O_3)** behaves differently than other pollutants, suggesting the need for more nuanced strategies when addressing gaseous pollution. The strong correlation between AQI and PM pollutants underscores the importance of **targeted emission control**, **urban planning**, and **sustained policy enforcement**.

By applying data science methods such as visualization, correlation analysis, and threshold exceedance checks, this report highlights key insights that can guide **evidence-based policy**, **public health planning**, and **citizen awareness**.