📘 ASSIGNMENT – 4

AIR QUALITY DATA VISUALIZER PROJECT

Submitted by: Arpit Patni

Roll No: 2501730111

Class: B.Tech CSE (AIML)

Section: A

Submitted to: Dr. Sameer Farooq

1. Introduction

This project is developed as part of Assignment-4, where the objective was to build a Weather Data Visualizer.
However, the dataset downloaded from Kaggle contained air quality pollutant measurements (PM2.5, PM10, NO2, AQI, etc.) instead of temperature and rainfall.

Therefore, the assignment was adapted into an Air Quality Data Visualizer, keeping all steps and learning outcomes identical:

Data loading

Data cleaning

Monthly / seasonal analysis

Visualizations

Statistical summaries

The project uses Python, Pandas, Matplotlib, and Jupyter Notebook for analysis and visualization.

2. Dataset Description

The dataset used is the Kaggle Air Quality Data (Indian Cities) which contains:

City

Date

Pollutants: PM2.5, PM10, NO, NO2, NOx, NH3, CO, SO2, O3

Volatile compounds: Benzene, Toluene, Xylene

AQI & AQI_Bucket

The analysis focuses on AQI, PM2.5, PM10, and NO2, as they are key measures of air pollution levels.

3. Data Cleaning & Preprocessing

Steps performed:

Loaded dataset from /data/raw_weather.csv

Converted Date column to datetime

Set Date as index

Sorted values by date

Handled missing values by replacing with median

Selected important pollutants for analysis

Cleaned columns used:

PM2.5

PM10

NO2

AQI

4. Statistical Analysis

✓ Daily AQI Statistics:

Mean AQI: 158.78

Standard Deviation: 130.27

Minimum AQI: 13

Maximum AQI: 2049

✓ Monthly Aggregation (PM2.5, PM10, AQI):

Data was grouped month-wise using:

df.resample('ME').mean()

Revealing seasonal pollution patterns such as:

Higher PM2.5 in winter months

Lower pollution in monsoon

✓ Seasonal AQI Classification Used:

Months, Season

Dec–Feb, Winter

Mar–May, Summer

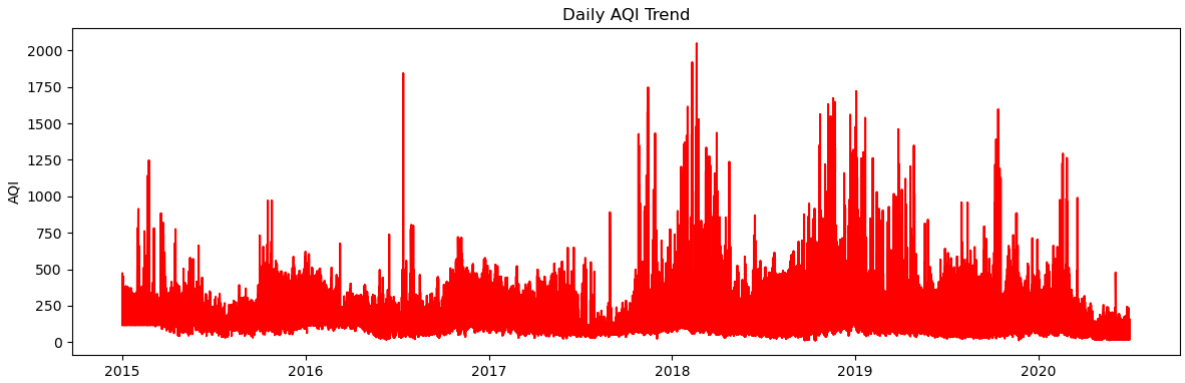
Jun–Aug, Monsoon

Sep–Nov, Post-Monsoon

5. Visualizations

Below are the graphs generated using Matplotlib.

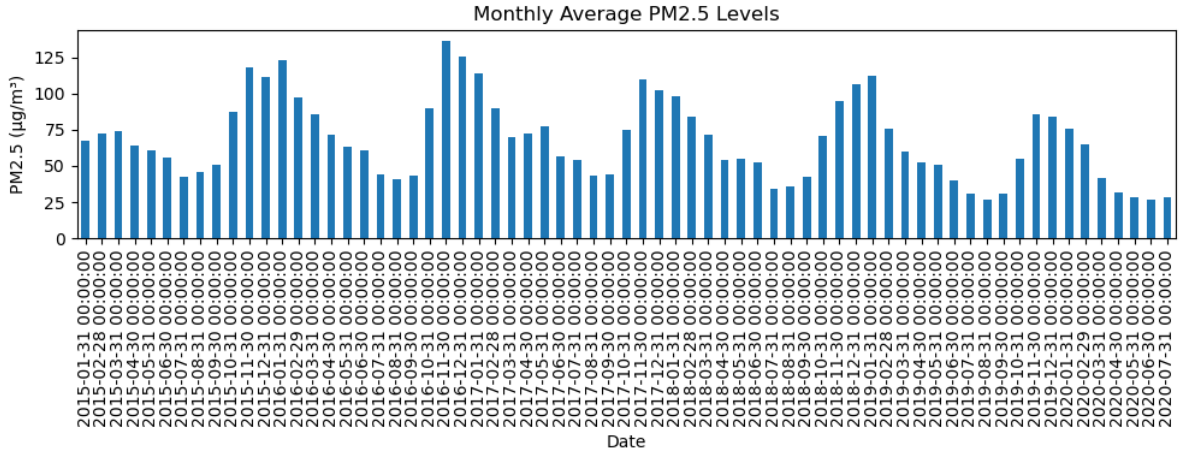
5.1 Daily AQI Trend



This plot shows the day-to-day variation of AQI from 2015 to 2020.

Major spikes observed around 2017–2019 indicate pollution peaks, especially winter months

5.2 Monthly Average PM2.5 Levels



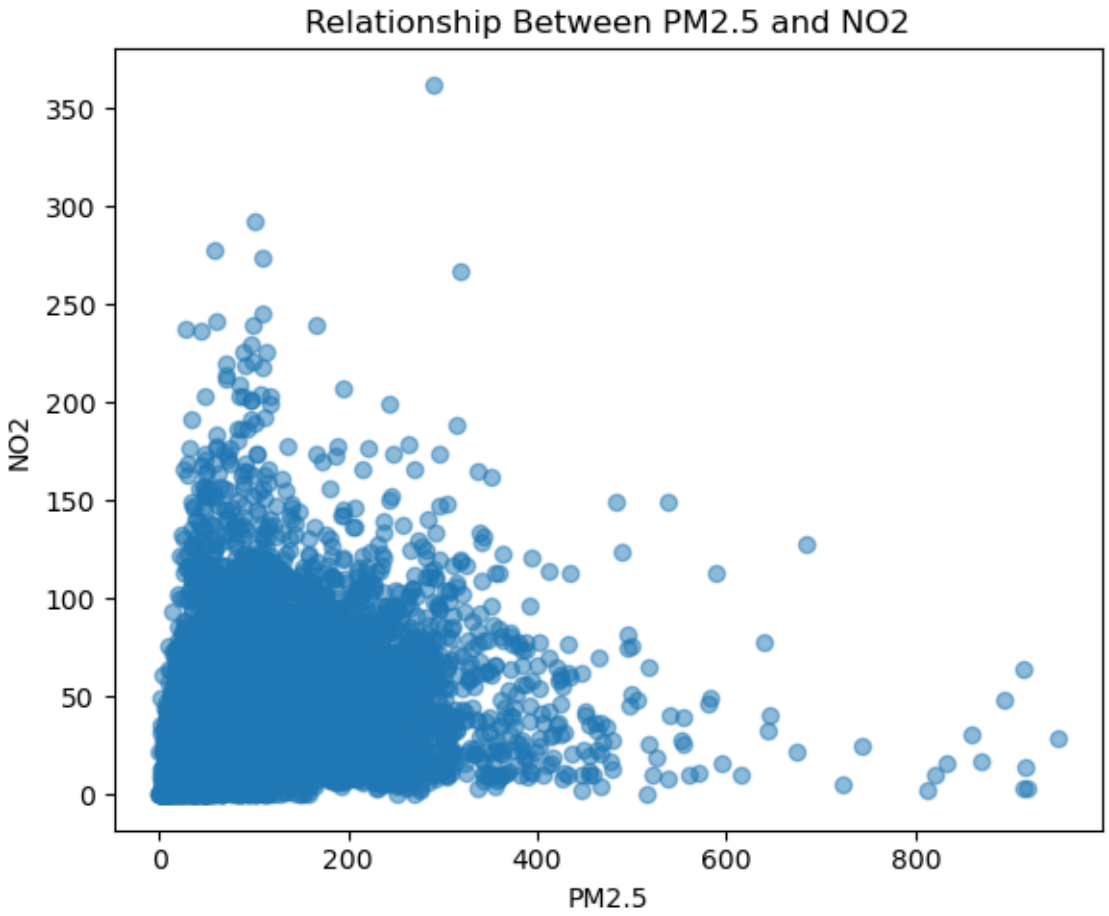
This chart shows clear seasonal patterns:

Higher PM2.5 levels during winter

Lower values during monsoon months

Repeating yearly cycle

5.3Scatter Plot: PM2.5 vs NO2



This shows a positive correlation:
Higher PM2.5 often corresponds to higher NO2, indicating vehicular/industrial emissions.

6. Observations & Insights

✓ AQI shows large fluctuations but clear seasonal cycles
✓ Winter months show the highest pollution
✓ Monsoon reduces particulate levels due to rainfall
✓ PM2.5 and NO2 are moderately correlated
✓ Air quality declined significantly between 2017–2019

7. Folder Structure

Assignment_4_Weather_Visualizer/ │

├── data/ → raw dataset

├── src/ → analysis, cleaning, visualization modules
├── notebooks/ → main Jupyter Notebook

├── outputs/ → plots and cleaned CSV

└── README.md → documentation

8. Conclusion

This project successfully demonstrates:

Real-world dataset processing

Air quality trend analysis

Pollution visualization

Seasonal evaluation

Correlation study between pollutants

Even though the assignment originally required weather data, the adaptation to air quality data maintains all objectives and enhances environmental insight.

9. GitHub Links

Repository LINK:

https://github.com/arpitpaatni2007-jpg/Python-assissgnments/tree/main/Assisgnment-4

Profile LINK:

https://github.com/arpitpaatni2007-jpg