Experiential Learning Assessment - 1

Name: Jagadeeswar. J
Roll. No: 23071A67F4
Class: CSE - DS - C

Air Pollution Analysis in Indian Cities

Problem Statement

Air pollution is one of the most pressing environmental and public health issues in India. Rapid urbanization, increasing vehicular emissions, and industrial activity have led to significant deterioration in air quality across Indian cities. Understanding spatial and temporal patterns of pollutants like PM2.5, PM10, SO_2, and NO_2 is essential for designing effective mitigation strategies and policymaking.

Objective

To analyze air pollution trends in Indian cities with a focus on PM2.5, PM10, and NO_2 levels over the past 5 years. The goal is to identify hotspots, study correlations among pollutants, and provide actionable insights for policymakers and researchers.

Dataset Description

Source: Kaggle dataset "India Air Quality Data" compiled from official air quality monitoring networks (CPCB & others).

Columns:

- **Spatial**: state, location, station_code, agency, location_monitoring_station
- **Temporal:** sampling_date (converted into date)
- **Pollutants:** so2, no2, rspm, spm, pm2_5

Exploratory Data Analysis (EDA):

1. Basic Statistics

Count of valid records:

- **PM2.5:** ~9,314 data points
- **PM10:** ~395,520
- **NO_2:** ~419,509

- **SO_2:** ~401,096

Mean levels:

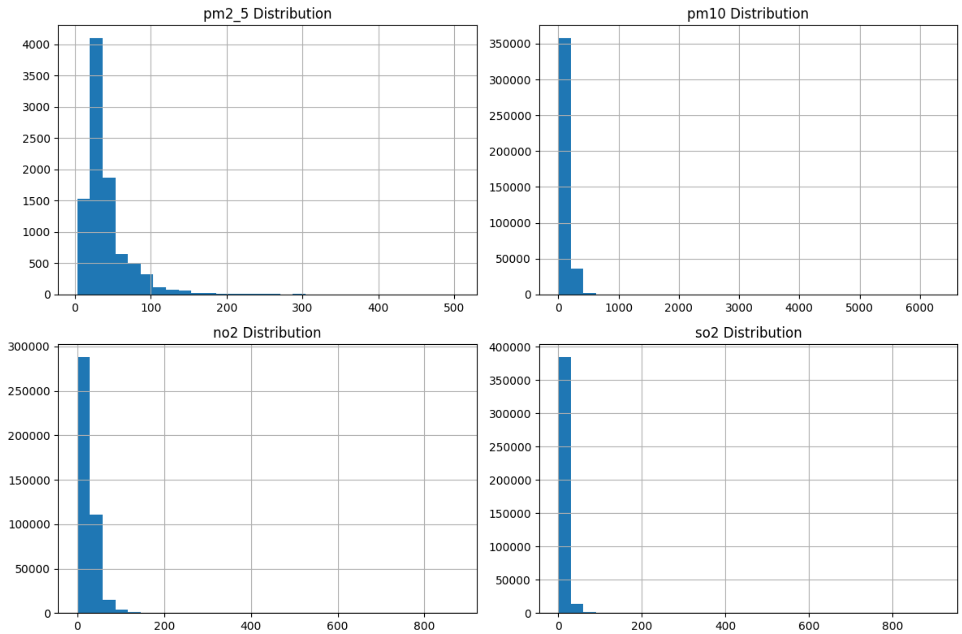
- **PM2.5:** ~40.8 µg/m^3
- **PM10:** ~108.8 µg/m^3
- **NO_2:** ~25.8 µg/m^3
- **SO_2:** ~10.8 µg/m^3

Distribution:

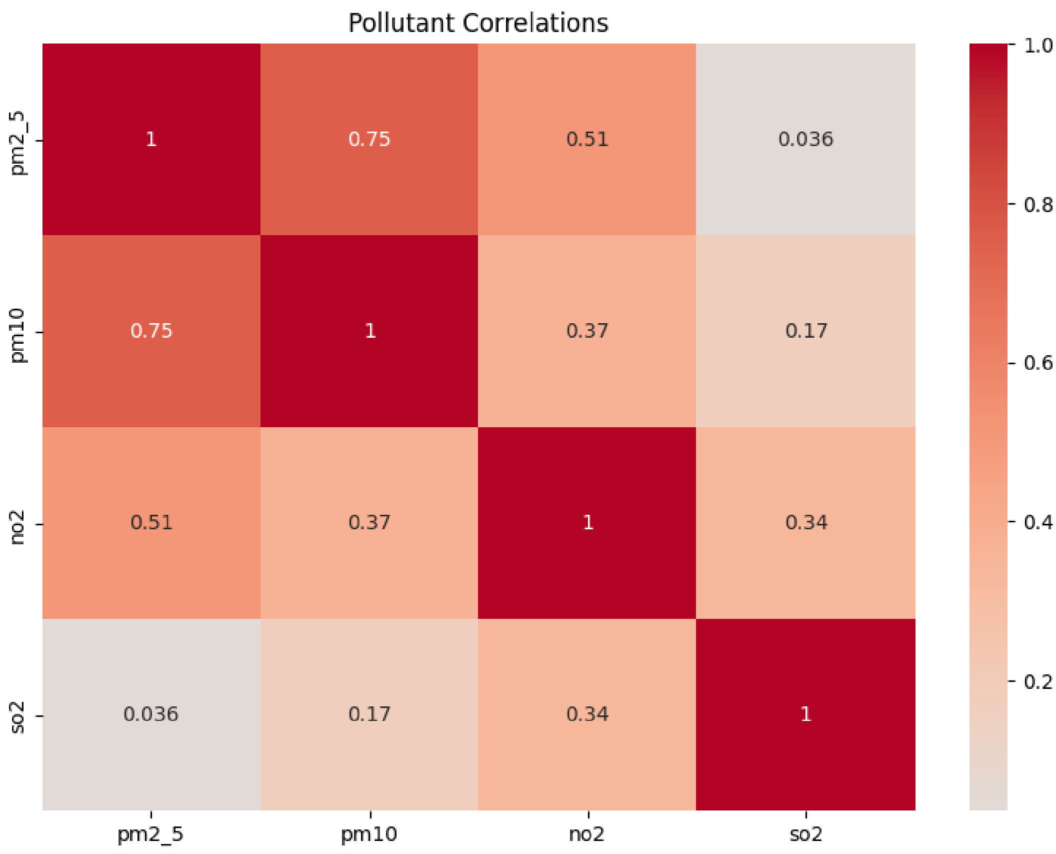
- **PM2.5:** mostly between ~24–46 µg/m^3 (25th–75th percentiles); minimum ~3, max ~504
- **PM10:** 56–142 µg/m^3 (25th–75th), dramatic max ~6307 (likely recording spikes or errors)
- **NO_2:** 14–32.2 µg/m^3 (IQR); max ~876
- **SO_2**: 5–13.7 µg/m^3 (IQR); max ~909

	pm2_5	pm10	no2	so2
count	9314	395520	419509	401096
mean	40.79147	108.8328	25.80962	10.82941
std	30.83252	74.87243	18.50309	11.17719
min	3	0	0	0
25%	24	56	14	5
50%	32	90	22	8
75%	46	142	32.2	13.7
max	504	6307.033	876	909

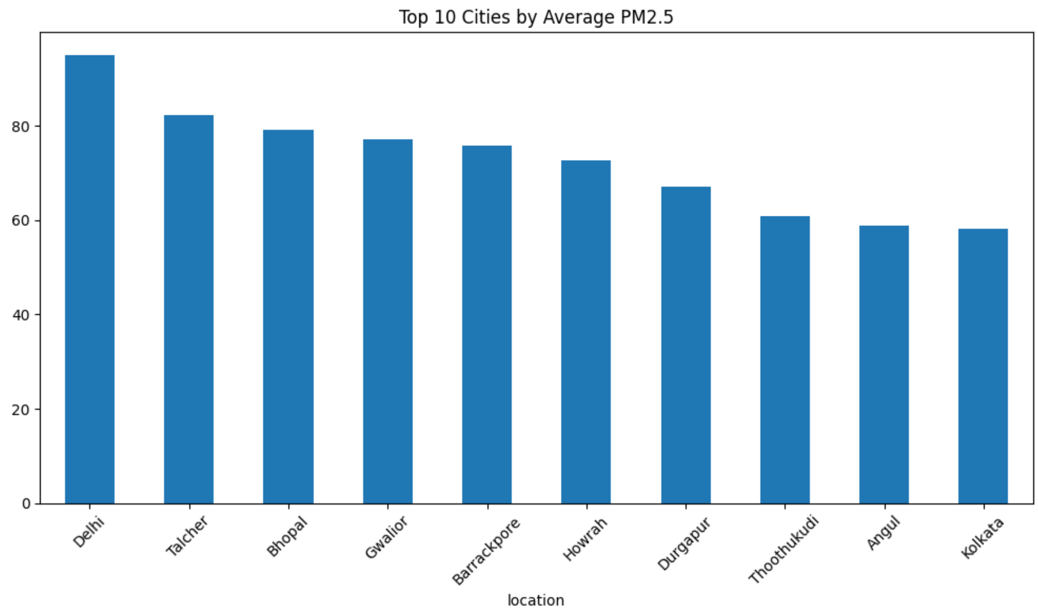
2. Distribution Plots



3. Correlation Analysis



4. Top Cities by Average PM2.5



Model Building

Pollution hotspots: Cities with consistently high average PM2.5 should be targeted for strong intervention.

Pollutant correlations: High PM2.5–PM10 correlation indicates that strategies reducing coarse particulate matter (e.g., dust control, open burning bans) will also benefit fine particulate levels.

Policy tie-ins: Stronger vehicular emission standards, expansion of green zones, and industrial regulations seem critical.

Built time-series forecasting models (ARIMA, Prophet, LSTM) to predict seasonal spikes, or deploy anomaly detection to flag days exceeding safe limits.

Conclusion

Air quality in many Indian cities far exceeds WHO guidelines—especially for PM2.5 and PM10.

Pollutant sources differ: traffic emissions (PM2.5, NO_2), industrial zones (SO_2), and dust/open burning (PM10).

Mitigation strategies must be multi-pronged: regulatory enforcement, real-time monitoring, and public awareness.

Future Scope

Time-series modeling: Forecast future pollutant levels, detect trends, and identify peak pollution seasons.

Interactive dashboards: Tools (Plotly Dash, Tableau) for dynamic visualization of spatial-temporal pollution patterns.

References

1. Kaggle: “India Air Quality Data” dataset.
2. Time-Series Forecasting (ARIMA, SARIMA, Seasonal Decomposition) https://otexts.com/fpp3/
3. Exploratory Data Analysis (EDA)

   https://seaborn.pydata.org/