Pollution Data Analysis Report

# 1. INTRODUCTION

This project focuses on analyzing air pollution data from various Indian cities and states. Using Python for data cleaning, transformation, and visualization, the study aims to uncover trends in pollutant levels, identify the most polluted regions, and understand pollutant distribution across locations.

# 2. SOURCE OF DATASET

The dataset, provided as a CSV file, includes the following fields:

- City, State, and Station name

- Pollutant ID, along with minimum, maximum, and average pollutant readings

-The data reflects pollution monitoring metrics across various urban and rural locations.

# 3. EXPLORATORY DATA ANALYSIS (EDA)

The data was first loaded and inspected for missing values. Entries with missing pollutant readings (pollutant\_min, pollutant\_max, pollutant\_avg) were removed to ensure data integrity.

Python libraries used:

- pandas for data handling

- matplotlib and seaborn for data visualization

# 4. ANALYSIS ON DATASET

General Description:

The dataset was analyzed to highlight pollution levels by city, state, and station, as well as to examine correlations and pollutant distribution.

Steps and Observations:

1. Top 10 Most Polluted Cities

2. Pollutant Type Frequency

3. Correlation between Pollutant Metrics

4. Top 10 States by Average Pollution

# 5. VISUALIZATIONS

- Bar Charts: Top cities, states, stations by pollution levels

- Heatmap: Correlation between min, max, and average pollutant values

# 6. CONCLUSION

This pollution data analysis revealed critical information on which cities and states face higher pollution risks. The correlation analysis confirmed consistency in the pollutant readings, and the frequency plots helped identify the most tracked pollutants. Such insights are vital for environmental monitoring and policy-making.

# 7. FUTURE SCOPE

- Include time-based trend analysis to observe pollution changes over months or seasons

- Use machine learning to forecast pollution spikes or drops

- Build interactive dashboards for real-time pollution tracking across locations

# 8. REFERENCES

[1] Dataset Source: https://www.data.gov.in/catalog/real-time-air-quality-index

[2] Matplotlib & Seaborn Documentation