Project Report: Air Quality Analysis in India

1. Introduction

Air pollution is a major environmental and public health concern in India. High levels of pollutants such as PM2.5, NO2, and SO2 contribute to respiratory diseases, climate change, and reduced quality of life. This project aims to analyze real-time air quality data using OpenWeatherMap's Air Pollution API to identify trends, correlations, and anomalies in air pollution levels over time.

2. Problem Definition

Objective

The goal of this project is to:

- Fetch real-time air quality data for India.
- Analyze the levels of pollutants over time.
- Identify correlations between different pollutants.
- Detect anomalies in air quality levels.
- Provide insights that can be useful for policymakers, researchers, and environmentalists.

Research Questions

- How does the Air Quality Index (AQI) fluctuate over time?
- What are the trends in major pollutants (PM2.5, NO2, etc.)?
- Are certain pollutants strongly correlated?
- Can we detect unusual pollution spikes?

3. Methodology

3.1 Data Collection

- **Source:** OpenWeatherMap Air Pollution API.
- **Parameters:** AQI and concentrations of CO, NO, NO2, O3, SO2, PM2.5, PM10, and NH3.
- **Storage:** Data is stored in a CSV file (india_air_quality.csv).

3.2 Data Preprocessing

- Convert timestamps to human-readable datetime format.
- Handle missing values.
- Convert numerical data to appropriate formats.

3.3 Exploratory Data Analysis (EDA)

- **Statistical Summary**: Compute mean, median, min, max, and standard deviation.

- **Visualizations:**
    - Line plots to analyze trends over time.
    - Box plots to detect outliers.
    - Correlation heatmaps.

3.4 Anomaly Detection

- Compute **Z-scores** for AQI values.
- Identify anomalies where |Z-score| > 3.

3.5 Trend Analysis

- Compute **7-day rolling averages** for pollutants.

4. Analysis & Results

4.1 AQI Trends

- AQI fluctuates significantly, with periodic peaks.
- Certain pollutants (PM2.5, NO2) show consistently high concentrations.

4.2 Correlation Between Pollutants

- Strong correlation observed between **PM2.5 and NO2**.
- CO and NH3 have weaker correlations with AQI.

4.3 Anomalies in Air Quality

- Detected extreme pollution spikes using Z-score analysis.
- Potential causes include industrial emissions and seasonal weather changes.

4.4 Visualizations

- Line plots show temporal changes in pollutants.
- Correlation heatmap highlights relationships between pollutants.
- Rolling averages smooth out short-term fluctuations.

5. Conclusion

This project successfully analyzed real-time air quality data and identified key trends and anomalies.

Key Takeaways:

- **High pollution levels require continuous monitoring**.
- **Certain pollutants are strongly correlated, indicating common sources**.
- **Anomaly detection can help identify critical pollution events**.

Future Work

- Expanding the dataset over multiple months for seasonal analysis.
- Predicting future air quality using machine learning.
- Developing an interactive web-based dashboard for public awareness.

Appendix

Technologies Used

- Python (Pandas, Matplotlib, Seaborn, Requests)
- OpenWeatherMap API
- Data Processing & Visualization Tools

References

- OpenWeatherMap API Documentation
- Indian Environmental Reports on Air Pollution

Author: Ritalee Monde