**Programming For Data Science**

**Subject Code: CSL 225**

**Project Report**



**Faculty name:**

**Ms. Akanksha Kaushik**

**Team:**

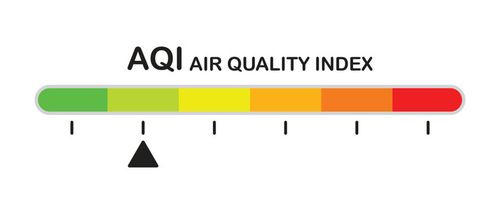
* **Rujul Bakshi(23csu269)**
* **Saanvi Vashist (23csu270)**
* **Rahul Thakan (23csu256)**

**Department of Computer Science and Engineering The NorthCap University, Gurugram- 122001, India**

Project Title:

Air Quality Index Exploration





**INDEX**

**1. Abstract**

**2. Objectives**

**3. Step-wise tasks for completing this project**

**4. Summary**

**5. Tools used**

**ABSTRACT**

This project provides an in-depth analysis of air quality data across Indian cities, focusing on key air pollutants such as PM2.5, PM10, NO2, SO2, and O3, along with the Air Quality Index (AQI). The objective of the project is to explore the distribution of pollutants, identify pollution hotspots, and understand regional variations in air quality. Using two datasets, one containing city-level air quality measurements and the other providing geographic and demographic details of Indian cities, we perform various analyses to reveal trends and insights related to air pollution.

The findings from this project highlight the cities and states with the worst air quality, identifying the dominant pollutants responsible for poor air quality. The analysis helps in understanding the underlying patterns and trends, offering valuable insights for policymakers and environmental agencies to take targeted action towards improving air quality.

**OBJECTIVES**

1. Data Collection and Preprocessing
2. Exploration of Pollutant Distribution
3. Dominant Pollutant Identification
4. Geographic Analysis of Pollution
5. State-wise Pollutant and AQI Comparison

**TASKS**

**Task 1: Data Import and Exploration**

1. **Imported and examined** the structure of the city-level air quality dataset (city\_day.csv) and the Indian cities geographic dataset (indian\_cities.csv).
2. **Handled missing values** by filling numerical columns with mean values and assigning placeholders to categorical columns

**Task 2: AQI and Pollutant Distribution Analysis**

1. **Analysed the distribution of AQI and pollutants** (PM2.5, PM10, NO2, SO2, O3) to explore their spread across cities.
2. **Examined pollutant distribution by AQI buckets** to understand how different levels of pollutants correspond to AQI categories

**Task 3: Geographic Analysis**

1. **Mapped the geographic distribution of AQI** by visualizing longitude and latitude data to identify pollution hotspots across cities.
2. Adjusted data indices and ensured consistency.

**State-wise Pollution Analysis**

1. **Calculated state-wise average pollutant levels** (PM2.5, PM10, NO2, SO2, O3) to compare air quality across states.

**Task 5: Pollutant Correlation and Insight Extraction**

1. **Analysed the correlation** between different pollutants and AQI to identify the most impactful pollutants on air quality.

**SUMMARY**

This project analyses air quality data for Indian cities, focusing on the relationship between key pollutants (PM2.5, PM10, NO2, SO2, O3) and the Air Quality Index (AQI). Using datasets on pollutant levels and city geography, the project preprocesses the data, addresses missing values, and aggregates pollutants by city.

It explores AQI and pollutant distributions, identifies pollution hotspots through geographic visualizations, and compares state-wise pollutant levels. The project also calculates the correlation between different pollutants and AQI, providing insights into regional air quality variations.

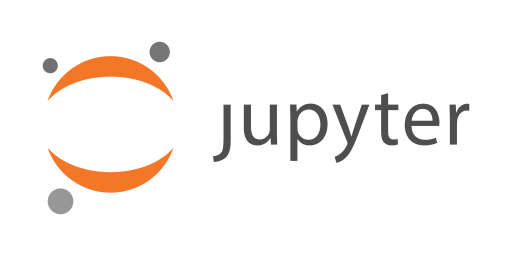
The findings aim to offer a clearer understanding of air quality trends across Indian cities, supporting efforts to improve pollution control and public health.

**TOOLS USED**

THE TOOLS THAT WE HAVE USED IN THIS PROJECT ARE AS FOLLOWS:

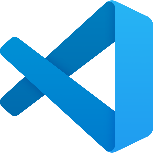
1. JUPYTER NOTEBOOK

FOR CODE COMPILATION



2. VSCODE

ADDITIONAL CODING TASKS AND DEBUGGING



3. MICROSOFT EXCEL

PREPROCESSING DATASETS



4. GITHUB

CREATING REPOSITORY AND DOCUMENTING THE PROJECT

