SKILL UP

IBM

AIR QUALITY ANALYSIS IN TAMIL NADU

DAC_Phase1

Team Members:

1. **Nithyasri V G**
2. **Sakshi Rajesh Bhavsar**
3. **Mathew P**
4. **Sarfraj Ansari**

AIR QUALITY ANALYSIS IN TAMIL NADU

Project Definition:

The primary objective of this project is to conduct a comprehensive assessment and visualization of air quality data collected from monitoring stations situated across Tamil Nadu. Through this project, we aim to extract valuable insights into the prevailing patterns of air pollution, identify regions exhibiting elevated levels of pollution, and establish a predictive model capable of estimating RSPM/PM10 levels. This analysis will be based on the concentrations of SO2 and NO2.

To achieve these objectives, the project encompasses several key phases. Firstly, we will precisely define the specific goals and targets we intend to accomplish through this study. Next, we will carefully devise the approach to be used in our analytical process. This will involve selecting appropriate statistical techniques and data processing methods. Additionally, we will thoughtfully choose visualization methods that effectively convey the patterns and trends identified in the air quality data. Finally, a predictive model will be developed using the Python programming language along with relevant libraries, allowing for accurate assessments of RSPM/PM10 levels based on the concentrations of SO2 and NO2. This comprehensive approach will enable us not only to ascertain the current state of air quality in Tamil Nadu but also to provide valuable insights for potential interventions and policy decisions aimed at mitigating air pollution.

Design Thinking:

Project Objectives:

Our objectives include analysing air quality trends, pinpointing pollution hotspots, and constructing a predictive model for RSPM/PM10 levels. We aspire to obtain comprehensive insights into the state of air pollution, thereby strengthening our capacity to make well-informed decisions concerning air quality management and the formulation of effective mitigation strategies.

Analysis Approach:

To plan the steps for loading, preprocessing, analyzing, and visualizing air quality data in Python, we follow this approach:

- **Import Relevant Libraries**:
  Begin by importing necessary libraries such as Pandas for data handling and Matplotlib/Seaborn for visualization.
- **Load Data:**
  Retrieve the air quality data from its source, which could be a CSV file or a database, using Pandas' data import functions.
- **Data Preprocessing:**
  In this stage, perform tasks like handling missing values, converting data types, and potentially scaling or normalizing numerical features.
- **Exploratory Data Analysis (EDA):**
  - ✓ Conducting exploratory data analysis to gain insights from the data.
  - ✓ Calculating descriptive statistics using functions like describe() in Pandas.
  - ✓ Creating visualizations with libraries like Matplotlib and Seaborn, including line graphs for trends, histograms for distributions, and scatter plots for relationships.
  - ✓ If necessary, performing statistical tests using modules like Scipy to uncover patterns and relationships.

- **Feature Selection:** Begin by identifying relevant features, such as SO2 and NO2 levels, that will be used in the modeling process.
- **Model Selection:** Choose appropriate regression or machine learning models for predicting RSPM/PM10 levels based on the selected features.
- **Model Training and Validation:** Split the data into training and testing sets, train the selected model, and validate its performance on the testing data
- **Final Reporting:** Ultimately, document and present the results of the analysis. This can be done using tools like Jupyter Notebook or by generating reports in formats such as PDF or HTML.

Visualization Selection:

Determining appropriate visualization techniques, such as line charts, heatmaps, scatter plots, and geographic maps, to effectively represent air quality trends and pollution levels. Additionally, considering radar charts, box plots, and bubble charts in specific contexts for a comprehensive understanding of air quality dynamics.