Project Title: Air Quality Assessment TN

Problem Definition:

The project aims to analyze and visualize air quality data from monitoring stations in Tamil Nadu. The objective is to gain insights into air pollution trends, identify areas with high pollution levels, and develop a predictive model to estimate RSPM/PM10 levels based on SO2 and NO2 levels. This project involves defining objectives, designing the analysis approach, selecting visualization techniques, and creating a predictive model using Python and relevant libraries.

Design Thinking:

1. Project Objectives:

Objective 1: Analyze Air Quality Trends

- **Approach:** We will analyze the historical air quality data to identify trends, seasonal variations, and any significant changes in pollution levels. This will involve statistical analysis and time-series visualization.

Objective 2: Identify Pollution Hotspots

- **Approach:** We will use geospatial analysis techniques to pinpoint regions with consistently high pollution levels. Heatmaps and spatial visualizations will help in identifying pollution hotspots.

Objective 3: Build a Predictive Model

- **Approach:** To create a predictive model for RSPM and PM10 levels, we will perform regression analysis. The model will be trained using historical data, and features such as SO2 and NO2 levels will be used as predictors.

Analysis Approach:

1. **Data Collection:**
    - We will gather the air quality dataset from the provided link: Dataset Link.
    - Data will be downloaded in a suitable format (e.g., CSV) for further analysis.
2. **Data Preprocessing:**
    - Clean and preprocess the dataset to handle missing values, outliers, and inconsistencies.
    - Convert relevant columns to appropriate data types.
    - Create derived features, such as pollutant concentration indices.
3. **Exploratory Data Analysis (EDA):**
    - Perform EDA to understand the basic statistics and distributions of air quality parameters.
    - Identify trends, seasonality, and correlations between variables.

4. **Data Visualization:**

   - Utilize various visualization techniques, including line charts, bar plots, heatmaps, and geospatial plots, to effectively communicate findings.

5. **Pollution Hotspot Analysis:**

   - Employ geospatial analysis tools (e.g., GIS) to create heatmaps and identify areas with high pollution levels.

6. **Predictive Modeling:**

   - Split the dataset into training and testing sets.
   - Build regression models (e.g., linear regression, random forest) to predict RSPM and PM10 levels based on SO2 and NO2 levels.
   - Evaluate model performance using appropriate metrics (e.g., RMSE, R-squared).

7. **Documentation and Reporting:**

   - Document all steps, code, and findings in a clear and organized manner.
   - Create a comprehensive report summarizing the analysis, visualizations, and model results.

Visualization Selection:

1. **Air Quality Trends:**

   - Line charts to show the variation in air quality parameters (RSPM, PM10, SO2, NO2) over time (monthly or seasonal trends).
   - Time-series decomposition plots to visualize seasonality and trend components.

2. **Pollution Hotspot Identification:**

   - Heatmaps to represent spatial variations in pollution levels.
   - Geospatial plots with color-coded markers to pinpoint pollution hotspots on a map of Tamil Nadu.

3. **Predictive Model Evaluation:**

   - Scatter plots to visualize the predicted vs. actual values of RSPM and PM10.
   - Residual plots to assess model performance and identify any patterns in prediction errors.

By following this structured approach and design thinking principles, we aim to gain valuable insights into air quality in Tamil Nadu, identify areas in need of environmental intervention, and provide a predictive model for estimating pollutant levels based on key parameters.