Air Quality Analysis and Prediction in Tamil Nadu

Phase 1

4-10-2023

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.

**Design Thinking:**

The project's primary goal is to examine the air quality data collected from monitoring stations located in Tamil Nadu.

The main objectives include uncovering patterns and trends in air pollution, pinpointing regions with elevated pollution levels, and constructing a predictive model for estimating RSPM/PM10 levels based on the concentrations of sulfur dioxide (SO2) and nitrogen dioxide (NO2).

Project Objectives:

* Analyze the air quality in Tamil Nadu.
* Examine the changes in air quality over time.
* Extract insights from the data.
* Predict future trends in air quality and RSPM/PM10 levels based on SO2 & NO2 Levels.

Analysis Approach:

1. Clean and preprocess the data.
2. Drop unwanted columns such as stn Code, State, etc...
3. Change the Sampling Date data type to datetime object.
4. Rename the Column names (Removing spaces & capital letters)
5. Group the data by Area.
6. Conduct EDA on the data.
   1. Perform Various Visualizations of the Data
7. Extract insights from the data.
8. Build a predictive model.
   1. Pick Machine Learning Models based on the extracted insights and train and test them with the data.
9. Present the results, insights, and findings.

Visualization Selection:

* Overall Data Distributions of SO2, NO2 & RSPM/PM10 (Using Violin Chart
* Histogram to Visualize the frequency distribution of pollutant levels
* Data Distribution of Every Area and location type by Month (Using Violin Chart)
* Data Distribution of Every Area and location type by Weekdays (Using Violin Chart)
* Line chart of SO2, NO2, & PM10 in Every Area and Location Type, With Important Holidays Marked
* Heat Map of Correlations Matrix (Correlation between SO2, NO2 & PM10).
* Regression Plots of SO2, NO2 & PM10 for every Area and Location Type.
* Visualizing Spatial Air Quality Distribution in Tamil Nadu Using Heatmaps.