DATA ANALYTICS

Phase 1: Problem Definition and Design Thinking

Team:

BENI HARRISON V - (211521106021)
ANBUCHEZHIYAN D - (211521106009)
KISHORE D - (211521106084)
GOKULAVASAN S - (211521106046)
BHAVAN KARTHIK S - (211521106022)

PROJECT TITLE:

The Air Quality Assessment in Tamil Nadu (TN)

PROJECT 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 pollutions 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.

Main reasons for air quality changes:

In recent times, Tamil Nadu has seen a significant boom in industries. However, this growth has resulted in the mixing of chemical effluents and toxic gases with the air, water, soil, and more. The presence of toxic gases like SO2, NO2, and fly ash in the air can lead to air pollution, which has a direct impact on the health and well-being of humans. This is a cause for concern that needs to be addressed urgently.

It's important to be aware that vehicles can have a significant impact on air quality. This is because they release harmful gases such as carbon dioxide, carbon monoxide, sulphur oxide, nitrogen dioxide, and more. It's worth noting that carbon monoxide is especially dangerous, as it is a poisonous gas. We should all do our part to reduce our vehicle emissions and help keep the air we breathe clean and safe.

Scope of our project:

From our project, we can easily find out the current air quality index of our state and anticipate the future air quality index of Tamil Nadu. This will assist us in pinpointing the pollution hotspots and taking the necessary precautions to regulate harmful substances in the air at specific locations with great care.

DESIGN THINKING:

Project Objectives:

Analyze Air Quality Trends: **This objective will involve examining historical air quality data to identify patterns and trends over time. You can use statistical methods and time series analysis to achieve this.**

Identify Pollution Hotspots: **To accomplish this, we can create visualizations that display pollution levels across different monitoring stations in Tamil Nadu. Heatmaps or geographical maps with color-coding can be effective for this purpose.**

Build a Predictive Model: **Developing a predictive model for RSPM/PM10 levels based on SO2 and NO2 levels is a complex but valuable goal. This will require data preprocessing, feature engineering, and machine learning techniques. Regression algorithms like linear regression or more advanced methods like random forests or neural networks could be considered.**

Analysis Approach:

Data Loading: Begin by gathering the air quality data from monitoring stations in Tamil Nadu. Ensure that the data is structured and clean for analysis.

Data Preprocessing: Prepare the data by handling missing values, outliers, and any data quality issues. Normalize or scale the data as needed for modeling.

Data Analysis: Utilize statistical analysis techniques to gain insights into air quality trends. You can calculate summary statistics, create distribution plots, and perform hypothesis tests if relevant.

Data Visualization: Visualize the data to make it more understandable. Use line charts to show time-series trends, heatmaps to identify pollution hotspots geographically, and scatter plots to explore relationships between variables.

Visualization Selection:

Line Charts: Line charts are suitable for displaying time-series data, making them a good choice for showing air quality trends over time. We can have separate charts for different pollutants or monitoring stations.

Heatmaps: Heatmaps can effectively represent geographical variations in air quality. we can use color gradients to show pollution levels across different regions of Tamil Nadu.

Scatter Plots: Scatter plots can help visualize relationships between variables. For instance, we can create scatter plots to explore the correlations between SO2 and NO2 levels and RSPM/PM10 levels.

In addition to these design elements, consider the following:

Data Sources: Ensure we have access to reliable and up-to-date air quality data from monitoring stations in Tamil Nadu.

We can able to access the previous data of air quality index in this government website: https://tn.data.gov.in/resource/location-wise-daily-ambient-air-quality-tamil-nadu-year2014

Python Libraries: Identify and list the specific Python libraries we plan to use for data analysis, visualization, and machine learning. Common libraries include pandas, matplotlib, seaborn, scikit-learn, and TensorFlow or PyTorch for machine learning.

Evaluation Metrics: Define how we will evaluate the performance of our predictive model. Common metrics for regression tasks include Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R2).

Ethical Considerations: Consider the ethical implications of our project, especially if it involves sensitive data or has potential environmental and public health implications.

By following this well-structured approach, We should be able to effectively analyze and visualize air quality data in Tamil Nadu and build a predictive model to estimate RSPM/PM10 levels based on SO2 and NO2 levels.

AIR QUALITY ASSESSMENT IN TAMIL NADU: