**Air Pollution Data Tracker**

**Project Overview**

**Dataset Description**

**Business Problem Statement:** Comprehensive Analysis and Mitigation Strategy Development for Air Pollution in the United States

**Background:** Amidst growing concerns about environmental health and air quality, there is a need for an in-depth analysis of pollution trends and their impacts across different regions of the United States. Understanding the variations in key pollutants like Nitrogen Dioxide (NO2), Sulphur Dioxide (SO2), Carbon Monoxide (CO), and Ozone (O3) is crucial for developing targeted environmental policies and health advisories.

**Title:** Air Quality Analysis in the United States (2000-2022)

**Scope:** This dataset provides an extensive analysis of air quality in the United States, focusing on key pollutants such as Nitrogen Dioxide (NO2), Sulphur Dioxide (SO2), Carbon Monoxide (CO), and Ozone (O3). It covers a period from the year 2000 to 2022.

**Size:** Approximately 647,251 observations across 21 columns.

**Data Sources:**

* **Primary Source:** U.S. Environmental Protection Agency (EPA).
* **Additional Contributions:** Data enhancements by Kagglers BrendaSo and ANGELA KIM.

**Purpose:** The dataset is designed for comprehensive research and analysis of air quality trends, pollution levels, and environmental health studies in the United States.

**Data Dictionary**

* **Date:** Date of data collection.
* **Address:** Specific location of data collection.
* **State:** U.S. state where data was collected.
* **County:** County within the state of data collection.
* **City:** City where data was collected.
* **O3 Mean:** Average Ozone level for the day.
* **O3 1st Max Value:** Highest Ozone level for the day.
* **O3 1st Max Hour:** Hour of highest Ozone level.
* **O3 AQI:** Air Quality Index for Ozone.
* **CO Mean:** Average Carbon Monoxide level for the day.
* **CO 1st Max Value:** Highest Carbon Monoxide level for the day.
* **CO 1st Max Hour:** Hour of highest Carbon Monoxide level.
* **CO AQI:** Air Quality Index for Carbon Monoxide.
* **SO2 Mean:** Average Sulphur Dioxide level for the day.
* **SO2 1st Max Value:** Highest Sulphur Dioxide level for the day.
* **SO2 1st Max Hour:** Hour of highest Sulphur Dioxide level.
* **SO2 AQI:** Air Quality Index for Sulphur Dioxide.
* **NO2 Mean:** Average Nitrogen Dioxide level for the day.
* **NO2 1st Max Value:** Highest Nitrogen Dioxide level for the day.
* **NO2 1st Max Hour:** Hour of highest Nitrogen Dioxide level.

**Resources link**

**Resources Link: Air Quality Dataset**

* **Dataset Link:** [Air Quality Dataset](https://www.kaggle.com/datasets/sirajahmad/acciojob-ml-capstone-air-pollution)

**Introduction to the Dataset**

The provided dataset on Kaggle offers comprehensive insights into air quality analysis in the United States from 2000 to 2022. It contains detailed information on key pollutants such as Nitrogen Dioxide (NO2), Sulphur Dioxide (SO2), Carbon Monoxide (CO), and Ozone (O3). This dataset is sourced from the U.S. Environmental Protection Agency (EPA) and includes additional enhancements by Kagglers BrendaSo and ANGELA KIM. Students can leverage this dataset for in-depth research, analysis, and exploration of air quality trends and environmental health studies.

**5. Self-Analysis and Problem Identification**

**Task Overview**

For this step, delve into the dataset provided in the resources section. Your objective is to perform a comprehensive self-analysis of the dataset and articulate your findings. This analysis will lay the groundwork for your project direction by identifying specific problem statements you aim to address.

**Instructions**

1. **Dataset Review**: Thoroughly examine the dataset shared in the resources section. Familiarize yourself with its contents, structure, and any relevant metadata provided.
2. **MECE Breakdown**:
   * Employ the MECE (Mutually Exclusive, Collectively Exhaustive) framework to break down the dataset into distinct, non-overlapping categories. These categories should cover different aspects or segments of the data.
   * Create a structured breakdown using bullet points, tables, or diagrams to illustrate your MECE analysis clearly.
3. **Problem Statement Formulation**:
   * Based on your MECE breakdown, craft **15 problem statements** that encapsulate the key issues, challenges, or areas of improvement within the dataset.
   * Each problem statement should be concise, specific, and address a distinct aspect of the data. Aim for diversity in the problem statements, covering various facets of the dataset.