US Air Pollution Analysis

DATA:

- Source of Data: The data is available on one of the open source platforms – Kaggle.com and downloaded from the following link https://www.kaggle.com/datasets/alpacanonymous/us-pollution-20002021

- Data Collection: This dataset deals with pollution in the U.S., that has been well documented by the U.S. EPA (Environmental Protection Agency). There are many more columns in the original dataset, however for the purpose of this project, only 4 major pollutants (Nitrogen Dioxide, Sulphur Dioxide, Carbon Monoxide and Ozone) are taken into account.

- Data Content: The data contains 48 states, 136 counties and 148 cities. The NO2 Units and SO2 Units are expressed in parts per billion and O3 Units and CO Units in parts per million.

- Data Profile:
    1. The data has 24 columns and 608699 rows.
    2. There are no missing values.
    3. Found 1479 rows of duplicate values. After removing them, the data has 607220 rows and 24 columns.

Index	**Columns**	**Description**	**Time Variant/Invariant**	**Data Type**
1	Date	Date of month data was collected	Variant	Qualitative
2	Year	Year data was collected	Variant	Qualitative
3	Month	Month the data was collected	Variant	Qualitative
4	Day	Day of month data was collected	Variant	Qualitative
5	Address	Location of the data point	Invariant	Qualitative
6	State	Area of the data point	Invariant	Qualitative
7	County	Area of the data point	Invariant	Qualitative
8	City	Area of the data point	Invariant	Qualitative
9	O3 Mean	Mean of Ozone gas readings	Variant	Quantitative
10	O3 1st Max Value	Maximum value of ozone recorded in a day	Variant	Quantitative
11	O3 1st Max Hour	Hour in which max value of ozone is recorded	Variant	Qualitative
12	O3 AQI	Ground level ozone concentration in atmosphere	Variant	Quantitative
13	CO Mean	Mean of Carbon Monoxide gas readings	Variant	Quantitative
14	CO 1st Max Value	Maximum value of CO recorded in a day	Variant	Quantitative

15	CO 1st Max Hour	Hour in which max value of CO is recorded	Variant	Quantitative
16	CO AQI	Ground level CO concentration in atmosphere	Variant	Quantitative
17	SO2 Mean	Mean of Sulphur Dioxide gas readings	Variant	Quantitative
18	SO2 1st Max Value	Maximum value of SO2 recorded in a day	Variant	Quantitative
19	SO2 1st Max Hour	Hour in which max value of SO2 is recorded	Variant	Quantitative
20	SO2 AQI	Ground level SO2 concentration in atmosphere	Variant	Quantitative
21	NO2 Mean	Mean of Nitrogen Dioxide gas readings	Variant	Quantitative
22	NO2 1st Max Value	Maximum value of NO2 recorded in a day	Variant	Quantitative
23	NO2 1st Max Hour	Hour in which max value of NO2 is recorded	Variant	Quantitative
24	NO2 AQI	Ground level NO2 concentration in atmosphere	Variant	Quantitative

- Limitations and ethics:
    1. Limitations: There are only 4 major gases in the data. Atmosphere has large number of gases which contribute to air pollution. Even though the 4 most contributors are there in the data, we would get more elaborated results with all gases data.
    2. Ethics: The data contains no personal information or indiscriminate variable, so no (Privacy Level Agreement) PLA Security is required.

- Questions to explore:
    1. Where - Which states have the highest pollutants?
    2. Why - What are the causes of certain pollutants in the atmosphere? Industrial? Vehicular? Household?
    3. When - What are the pollution trends across 2 decades? Monitor effects of COVID-19 lockdown, fuel prices, etc.
    4. How can these pollutants be reduced to improve air quality?
    5. What is the threshold of safe air quality?
    6. Recommendations for the general public based on AQI.