Air Quality Index

In the past few years, the problem of air pollution in the United States has been increasing for several factors such as the damage caused by the ozone hole and the increase in the density of harmful particles from the air such as carbon monoxide, nitrogen dioxide, carbon dioxide and fine particles, so the United States is keen to study the level of air quality Throughout the day by air quality index, AQI is the primary way to measure the current quality of the air. AQI values range from 0-500 with 0 being perfectly healthy and 500 being extremely hazardous. This issue take a lot of time during all the day to measure, through this project will study the probability of prediction of air quality index depend on old measurement during last forty years to save time and cost.

- **Description**:

Link: [Kaggle](#)

AQI or Air Quality Index is the primary way to measure the current quality of the air. AQI values range from 0-500 with 0 being perfectly healthy and 500 being extremely hazardous. AQI values are derived from moving averages/current values of PM2.5 (particulate matter), PM10, Ozone, Carbon Monoxide, Sulfur Dioxide, and Nitrogen Dioxide levels. Info:

Data contain 14 feature

Data Shape (5617325, 15)

Column description:

CBSA Code => The core-based statistical area (CBSA) code. CBSA = a U.S. geographic area defined by the Office of Management and Budget (OMB) that consists of one or more counties.

Date => The day of measurement.

AQI => The average air quality index (AQI) value for the day.

Category => The category of air quality ranging from "Good" to "Hazardous".

Defining Parameter => One of PM2.5 (particulate matter), PM10, Ozone, Carbon Monoxide, Sulfur Dioxide, or Nitrogen Dioxide which has the highest concentration.

Number of Sites Reporting => The number of stations used to make the data aggregation.

city_ascii => Name of the city where the measurement was taken.

state_id => Abbreviation of the state where the measurement was taken.

state_name => The state where the measurement was taken.

lat => The latitude where the measurement was taken.

lng => The longitude where the measurement was taken.

population => The population of the region where the measurement was taken.

density => The population per square kilometer where the measurement was taken.

timezone => The time zone of the region where the measurement was taken.

- **Analysis questions**
  - What's the average of AQI for last 10 year in California for each City?
  - What's the highest parameter concentration along months in California?
  - Compare between Riverside city and Los Angeles city which one have better AQI?
  - Does the area of the city in square kilometers effect on the Number of Sites Reporting needed to measure AQI (in California)?
  - What's the reason behind the increasing in number of testing in last decade?