# Analysis of Pollution
## In
## USA
### Reports within the Last
## 50 years of data



# By - Lalith Varma

**Problem Statement :**

Pollution is an alarming and critical environmental issue that the United States is currently facing. It is caused by numerous industries that emit harmful pollutants which negatively affect the quality of air, water, and land resources. Although the country has made significant strides in regulating pollution through the implementation of laws such as the Clean Air Act and the Clean Water Act, there is a growing concern regarding climate change. The challenge of addressing pollution necessitates a multifaceted approach that involves not only regulatory measures but also technological innovation and public education and engagement. Collaborative efforts must be undertaken to mitigate the harmful impacts of pollution and ensure the well-being of the planet and its inhabitants.

**Project Goals and Conditions :**

The goal of this project is to provide the government with evidence-based recommendations on how to tackle the pressing issue of pollution in the United States. With the help of AQI data spanning over four decades, we intend to identify the most significant contributors to pollution and assess their impact on air, water, and land resources. Our analysis will involve evaluating the effectiveness of existing regulations, such as the Clean Air Act and the Clean Water Act, and identifying areas where regulatory measures can be strengthened or implemented. We will also explore emerging technologies and their potential to reduce pollution and enhance environmental sustainability. Additionally, we will consider the role of public education and engagement in promoting sustainable practices and reducing pollution.

Through this project, we aim to provide a comprehensive set of recommendations that are feasible, effective, and aligned with the goals of environmental sustainability and public health. Our ultimate objective is to help the government safeguard the health of its citizens, mitigate the negative impacts of pollution on the environment, and contribute to global efforts to combat climate change.

**Business Understanding :**

The Air Quality Index (AQI) is a scientific tool used to assess the level of air pollution in a given area. It takes into consideration various pollutants such as ozone, particulate matter, and carbon monoxide, to provide a comprehensive evaluation of air quality. The AQI levels range from 0 to 500, with higher values indicating more polluted air. The AQI is determined by calculating the pollutant concentrations in the air, which are compared to the National Ambient Air Quality Standards (NAAQS) established by the Environmental Protection Agency (EPA). The AQI value is based on the highest pollutant level measured, with each pollutant having its own standard.

The AQI values are divided into six categories ranging from "good" to "hazardous," with corresponding health messages provided to inform people of potential health risks. It is essential for individuals, businesses, and government agencies to monitor the AQI to make informed decisions about activities that can affect air quality. For instance, an AQI value in the "unhealthy" range might prompt a school to cancel outdoor activities or a construction company to reschedule work that generates dust or emissions. By using the AQI to monitor air quality, individuals and organizations can take measures to protect their health and minimize their environmental impact.

**Data Understanding :**

The data for the pollution analysis report includes four datasets. The first dataset, "aqi_daily_1980_to_2021.csv," contains the daily Air Quality Index (AQI) measurements for the United States from 1980 to 2021. The second dataset, "US_AQI.csv," contains individual AQI measurements for each state in the United States. The third and fourth datasets are "WDICountry.csv" and "WDIDATA_T.csv," which contain global data from the World Bank on various factors, such as population, pollution, poverty, hunger, and more. The AQI datasets will be used to analyze the air quality trends in the United States, both on a national and state level. The global datasets will provide additional context and allow for comparisons between the United States and other countries. This information can be used to identify areas where pollution is particularly severe and to explore potential causes and solutions. By combining these datasets and analyzing the trends and patterns, we can gain insights into the impact of pollution on the environment and human health. The report will help policymakers and organizations make informed decisions to mitigate the negative effects of pollution and protect the well-being of the planet and its inhabitants.

**Data Preparation :**

In the Data Preparation step, the US_AQI.csv file was loaded using the pandas library, and the required information was extracted from it. Then, a rectangle plot was created to display the AQI values in a visually simple and clean manner. This was done by defining six rectangles with specific dimensions and colors, representing different levels of air quality. These levels are labeled as "GOOD," "MODERATE," "UNHEALTHY FOR SENSITIVE PEOPLE," "UNHEALTHY," "VERY UNHEALTHY," and "HAZARDOUS." The rectangles were then added to the plot in the appropriate segments and colored accordingly. Finally, the average AQI values for the United States over time were plotted on the x-axis, with the AQI values on the y-axis. This visualization allows for easy identification of trends in air quality over time and can aid in making informed decisions regarding activities that may affect air quality.

Lets see the output we get through this for the average AQI of the USA from the year 1980 to the 2021.
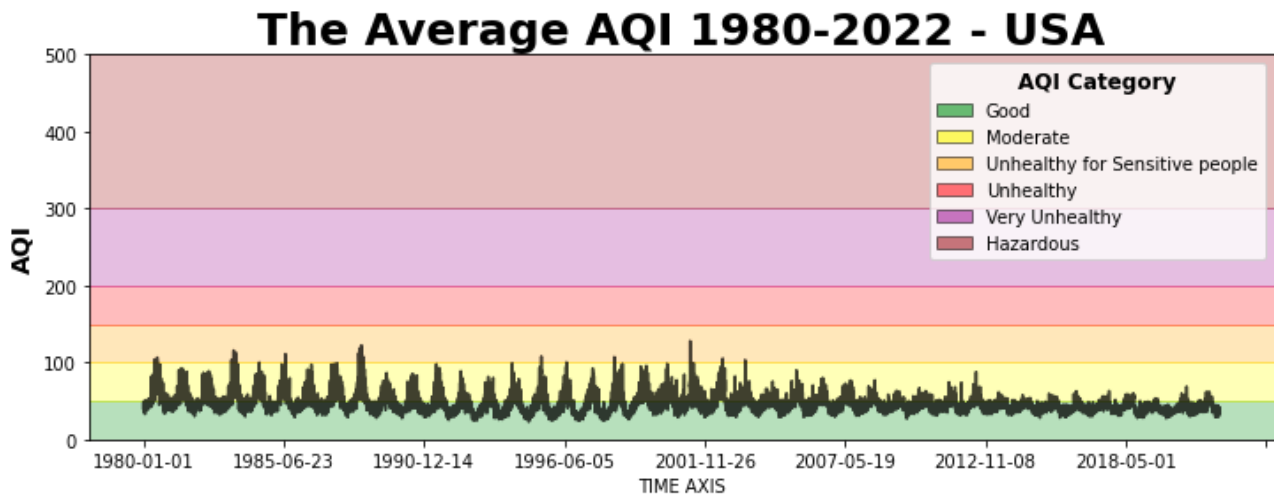


From the above information is a bit clear that the concentration of the pollution is reduced to a good extent over the time and the moves which are made are showing some results but how ever the results might be diverse and cannot be considered as a cumulative result as the contribution of the individual state might be different so lets just have a look at the data of the AQI for the individual states in the USA.

Now let's start with the next data set we have about the AQI of the different states in the United States. The dataset " aqi_daily_1980_to_2021.csv" is provided with the information about the states impact on the ozone and the AQI value on the day to day basis for all the states. so, we will load that particular dataset using pandas and will use the information which we needed from the dataset to have our view on the AQI Over the time. The process will take some time to generate all the information's visual view but how ever we can also pass the list of some states alone we want to check and limit the time if wanted. Now after the response from the plotting done over the data we can see that the majority of the states are having a good decline in the pollution over the time comparative to the 1980's to 2020's. But there are some states in the list which are not meeting the required numbers and are in a constant Pace with the Air Quality. For now lets have a look at the difference between the states AQI graph for the best and worst performers.

Maryland performing as one of the top :



AQI for Maryland
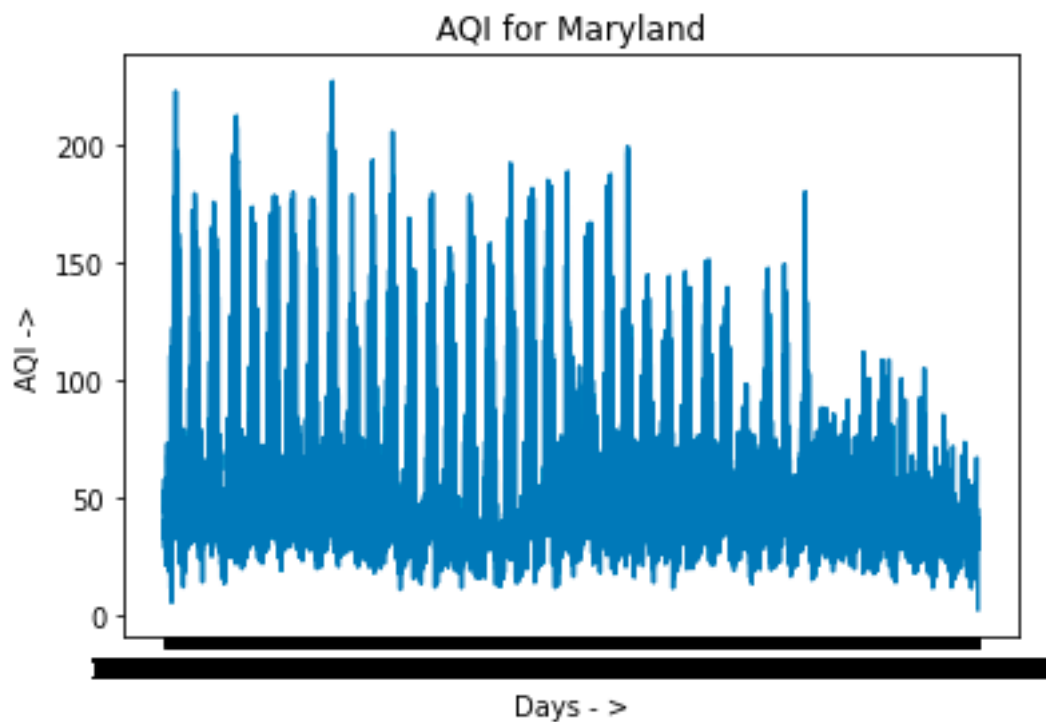
California being in the same state and causing more contribution :



AQI for California
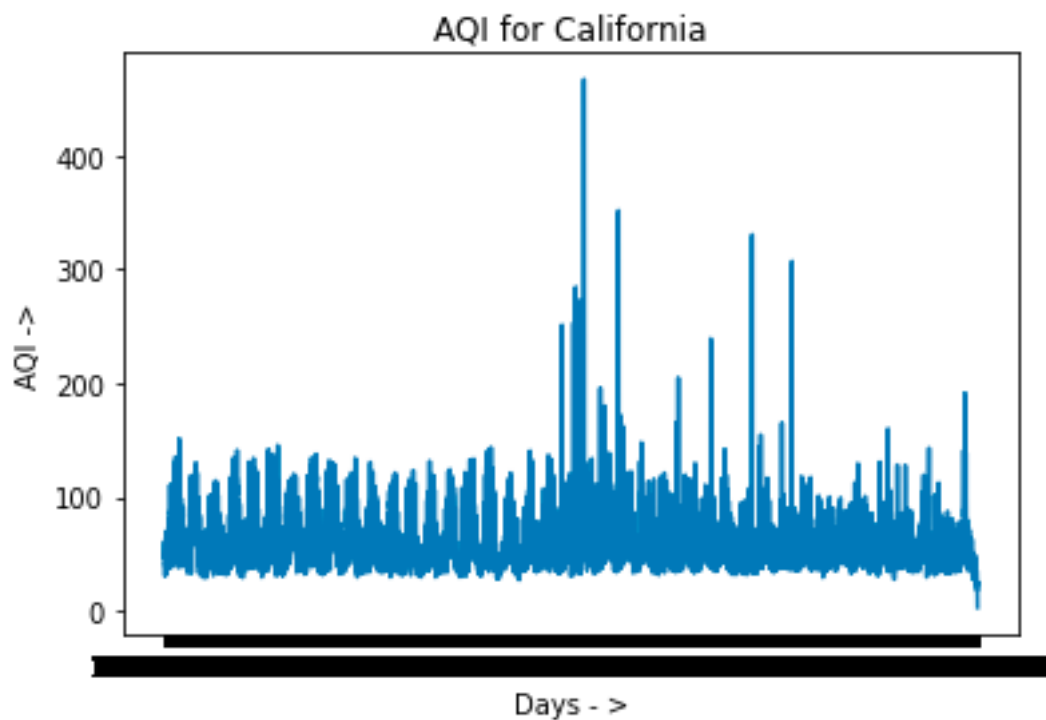
So leaving these independent datasets now lets check the over all status again in more particular way, where we check with the data from the most renounced survey which is done from the world bank to measure the different metrics from the entire dataset.

By load the other datasets from the list mentioned above which are: "WDIData_T.csv" and "WDICountry.csv" where we can cumulatively filter the data accordingly to the required instinct which is the data related to the USA . Just make it into a different variable as we will perform the slicing of the data by taking the following metrics into the consideration.
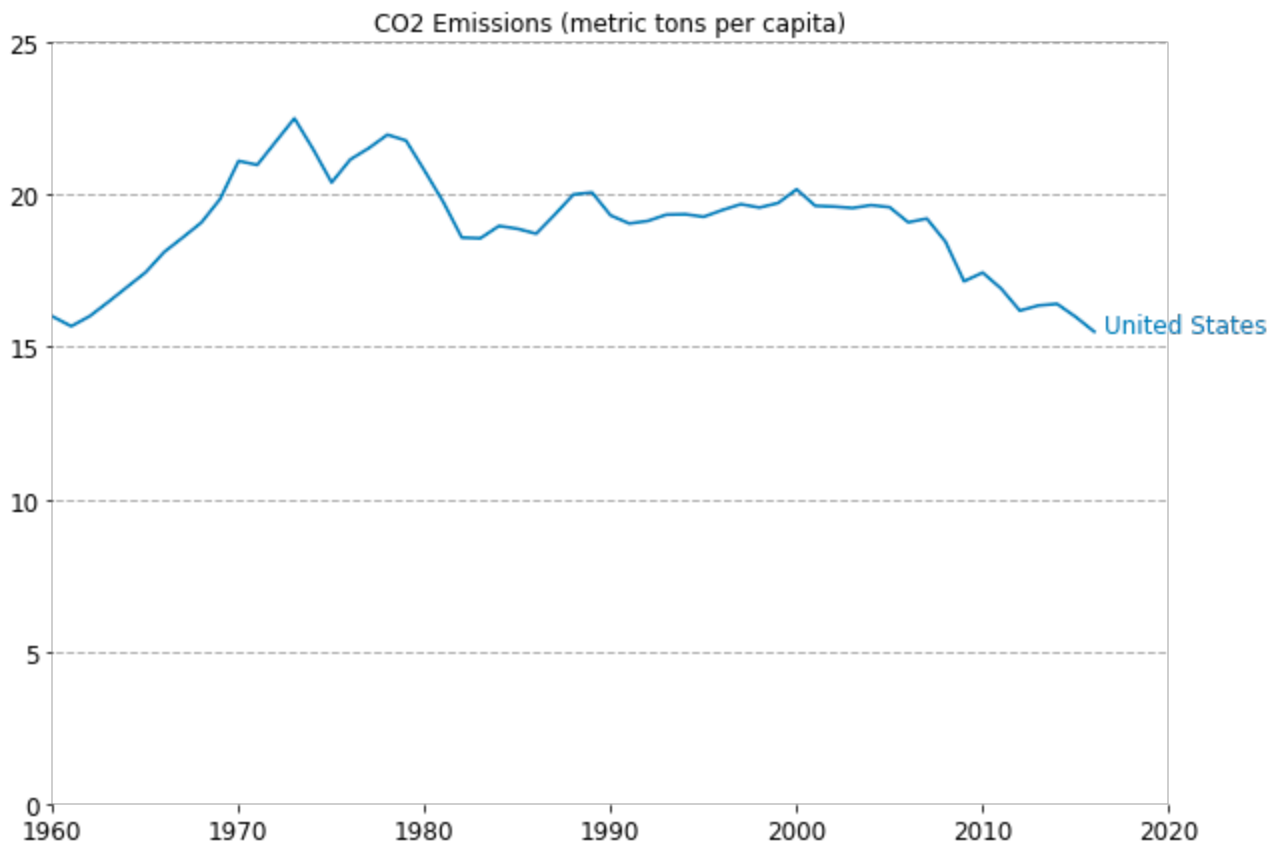
The parameter which we can consider as the cause of pollution are :
• CO2 emissions\metric, CO2 emissions\kt
• Total greenhouse gas emissions\kt of CO2
• CO2 emissions from gaseous fuel consumption
• CO2 emissions from liquid fuel consumption
• CO2 emissions from solid fuel consumption

These are the parameters we choose to consider as these affect the AQI directly and we have filtered the information from the dataset which ends with the following columns:
• 'CountryName'
• 'CountryCode'
• 'IndicatorName'
• 'IndicatorCode'
• 'Year'
• 'Value'

In the column 'IndicatorName' we have specified the the above list of parameters as a filter to get the specific data and see weather the emission from these root causes are tackled or not . The carbon emission per capita is measure as a whole now to generalise the pollution caused due to the carbon emission, as a result we can plot the graph for the carbon emission per capita over the time axis which will tell us weather The AQI is in good state or not in an indirect manner. The result of the line plot for the above information is shown below, which shows the very good decrease in the percapita of the carbon emission.

CO2 Emissions (metric tons per capita)

Therefore there is a good decline in the carbon emission in the USA comparatively and the air quality is directly affected by this improvement in the emission rates.

**Conclusion :**

Based on the graph analysis, we can infer that the CO2 emissions from various metrics such as CO2 emissions\metric, CO2 emissions\kt, Total greenhouse gas emissions\kt of CO2, CO2 emissions from gaseous fuel consumption, CO2 emissions from liquid fuel consumption, and CO2 emissions from solid fuel consumption are well-adjusted and under control. These metrics show a declining trend over the time axis. Therefore, the government's recommended action would be to continue regulating the consumption of CO2 emissions from gaseous fuel, liquid fuel, and solid fuel, as they are major contributors to pollution, according to World Bank statistics. By doing so, we can strive towards achieving a better environment than the peak pollution levels of the 1970s.

Note: Please go through the python notebook for more information and clarity towards the data and the visualisation part as the decision made here in report are very much clear from the analysis done there.