Project -> Comparing Air Quality of USA and Delhi[INDIA Capital] to Study Climate change from Combustion of Fossil Fuels [Petroleum and Coal]

Data Collection 🡪 the datasets used were found from the **official gov. sites for env. Protection** of USA and INDIA.

Data extracted from Following Sites:

1. <https://www.epa.gov/transportation-air-pollution-and-climate-change/accomplishments-and-success-air-pollution-transportation>
2. [https://www.eia.gov/totalenergy/data/browser/?tbl=T06.02 - /?f=A](https://www.eia.gov/totalenergy/data/browser/?tbl=T06.02#/?f=A)
3. <https://www.epa.gov/transportation-air-pollution-and-climate-change/carbon-pollution-transportation>
4. [https://www.epa.gov/automotive-trends/explore-automotive-trends-data - SummaryData](https://www.epa.gov/automotive-trends/explore-automotive-trends-data#SummaryData)
5. <https://www.epa.gov/air-emissions-inventories/2017-national-emissions-inventory-nei-data>
6. [https://cfpub.epa.gov/ghgdata/inventoryexplorer/ - transportation/entiresector/allgas/category/all](https://cfpub.epa.gov/ghgdata/inventoryexplorer/#transportation/entiresector/allgas/category/all)
7. Delhi air pollution from 🡪 <https://data.gov.in/>
8. Other Datasets Used Were Found from the same sources

Data Cleaning 🡪

The data Extracted from above had almost zero or few missing values , but the format of these datasets was very inappropriate and had ENCODING different from standard UTF-8 ,

We Managed to read the datasets after cleaning and now

The **Cleaned Data Is stored in .csv** formats and now can be used in different data softwares [which earlier was not possible].

Data VIZ🡪

After cleaning the mini datasets, we have made small graphs showing relationships between pollution and other various factors which affected it over the years …..

Different GRAPHS CAN BE **FOUND ON TABLEAU AND IPYNB** file ATTACHED

MAIN DATASET—

The data set showing usa air quality over the years showed us the quality of air improved over the years because of Measures from EPA and improvement in better transportation technology over the years , this dataset had features which were not good for DATA MODELLING process , therefore I converted all of them from **object 🡪 Int32 or float64** datatypes

DATA MODELLING 🡪 now since the data is cleaned we will be making splits in the datasets for training and testing sets ….

[this will be done today…]

Thinking of a very simple regression model , if this gives a good acc. We will stay on this , else we can sort help of other advanced regression techniques , OK ……