Time Series Project Proposal

Student Name:	Sujil Kumar K.M	**Student ID:**	D00242726

Project Title:

Air Pollution in US

1. **Background Research / Problem Description**

   Do you know, as per WHO every year seven million people die from breathing polluted air and also, during world environmental day the WHO has warned that nine out of ten people on earth breathe polluted air(10 facts about air pollution on World Environment Day | World Economic Forum n.d.)? There is also a huge problem as we are all aware of happening around the world in terms of global warming. Gases like carbon potentially damage the earth's ozone layer. This is something that can destroy humankind by the hit of even a small asteroid. As responsible data engineers, what we can contribute to society in this regard? There could be so many topics I could talk about, but I strongly believe it's time for us, data engineers to make use of technology to deep dive into the data to explore it further and make predictions from it. We can make use of these predictions to focus on those counties or towns so that govt. can put more regulations in place to protest against the unrestrained use of plastic or fossil fuel vehicles (Hertwich et al. 2001). Such measures are inevitable in this era to mitigate the air pollution in the country and around the world. Even though several studies like (Roberts 2004) have been done in the past, In my view, this analysis is expected to make accurate and more efficient models which can impact the society to act differently by using more eco-friendly vehicles and materials in day to day life leveraging the air pollution in all the cities of US.

   In addition, time-series analysis is incorporated with resampling, window/smoothing/transform methods, and interpolation techniques to explore the data and make precise forecasts.

2. **List the research questions related to the problem statement. Subsequently list out your project goals.**

   a. Predict the change in atmospheric gases over the coming years.

   b. Which counties are going to generate more pollution coming years?

   c. Which gases would be most generated which could potentially cause global warming in coming years?

   d. Check and find out if there is any seasonality in air pollution changes?

   Project Goals

   a. Learn more about Air pollution happening in the US and check its impacts over the years

   b. Understand different models and check for accuracies to find out the optimum model?

   c. Gain Hands-on experience with resampling, window/smoothing/transform methods, and interpolation techniques.

Data Source : click here to check the source of data.

Air pollution dataset contains total 29 Variables and 129810 rows are there.

Data Description

Nominal Data: - Address, state, county, city,

Continuous Data :- 'NO2 Units', 'NO2 Mean', 'NO2 1st Max Value', 'NO2 1st Max Hour', 'NO2 AQI', 'O3 Units', 'O3 Mean', 'O3 1st Max Value', 'O3 AQI', 'SO2 Units', 'SO2 Mean', 'SO2 Mean', 'SO2 1st Max Value', 'SO2 1st Max Hour', 'SO2 AQI', 'CO Units', 'CO Mean', 'CO 1st Max Value', 'CO 1st Max Hour', 'CO AQI'

Discrete Data: - State Code, 'County Code', 'Site Num', Date Local

Tools Used

Visual Studio Code, Jupyter Notebook, Git, Google Colab, Anaconda Navigator, MS Office.

3.   **Gantt Chart**
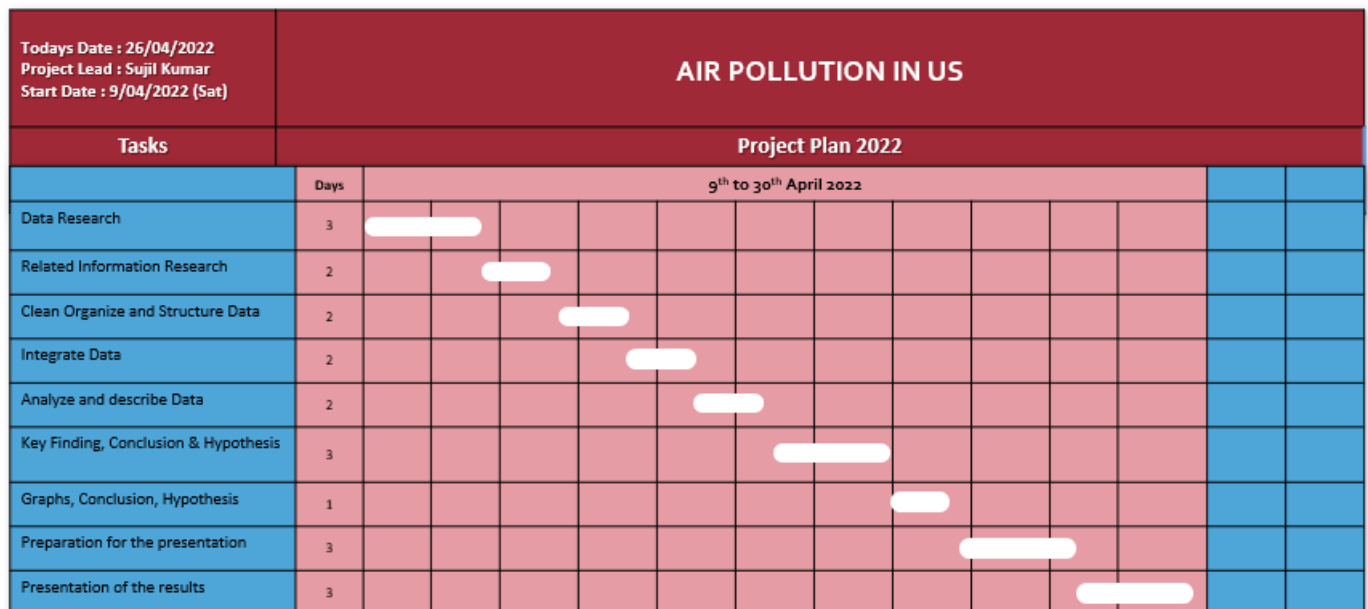
4.   **Special requirements and deliverables of the project**

   No special requirements for the project.

5.   **Risks – What risks can you identify? What will be the impact if the risk becomes a reality? What can you do to minimize the impact?**

   - Contingency plan in case you do not meet your project goals – prioritize the goals.
   - Unexpected output – any output – you should report it.
   - Loss of data/coding due to system failure – work on cloud/back up your work.
   - System compatibility – look for virtual machine options.
   - Ethics risks – work under the ethical guidelines detailed on DKIT/ Data protection website.

6. References

10 facts about air pollution on World Environment Day | World Economic Forum. Available from: https://www.weforum.org/agenda/2019/06/10-facts-about-air-pollution-on-world-environment-day [accessed 27 April 2022].

Hertwich, E.G., Mateles, S.F., Pease, W.S. and McKone, T.E. (2001). Human toxicity potentials for life-cycle assessment and toxics release inventory risk screening. *Environmental Toxicology and Chemistry* [online], 20(4), pp.928–939.

Roberts, S. (2004). Interactions between particulate air pollution and temperature in air pollution mortality time series studies. *Environmental Research* [online], 96(3), pp.328–337.