DATA ANALYTICS ON CORONAVIRUS

IMPACT OF AIR POLLUTION ON OUR LIVES

Introduction

In these recent times, humanity is threatened by a pandemic caused by a new coronavirus. Early this year, after an outbreak in China, WHO identified this new type of coronavirus as SARS-Cov-2 belonging to one of the seven coronavirus. Covid-19 is the disease caused by SARS-Cov-2, which is a respiratory disease and can affect both the upper and lower respiratory tracts. To fight the spread of this virus, almost every other major city in the world has been locked down and people are requested to maintain social distancing. As expected, the economy has crashed and people belonging to the lower income band are affected the most. Government and other officials are trying hard to revive the situation but experts predict it's quite a time before we get back to normal. Countless people have lost their lives and millions are infected. Every day health officials are working extremely hard to save the ones infected. But among all these bad news, Covid-19 has helped to revive our mother nature a bit. Though the lockdown has cost many people their jobs, the lockdown has exponentially reduced the pollution levels. Be it air, water or soil, the cities under lockdown have seen significant reduction in their pollution level. Though the damage caused by Covid-19 is pretty bad, on the other side it tried to heal our environment.

Objective

The primary objective of this report is to analyse air quality index trends of various air pollutants both pre and during lockdown period. The city that I chose for the analysis is London, UK. The main focus in this report is on the pollutants emitted by vehicles and industries.

Data

Dataset Description

The data that I used for the analysis was open source data available on the aqicn.org. The data can be found here. The website is maintained by the World Air Quality Index project. The data set contains 77 months of daily historical air quality index of air pollutants such as particulate matters (pm 2.5, pm 10), Oxides of Carbon, Nitrogen and Sulphur (CO, NO, SO_2) and Ozone (O_3). The dataset has 2326 observations and all are in accordance with the United States Environmental Protection Agency (EPA). Taking the maximum value of the AQI, we get an average AQI for the past 77 months to be “64.42”. According to EPA standards, this value represents London air to be of ‘Moderate Level of Health Concern’. The EPA standards chart can be found here.

Data Preprocessing

The first discrepancy that I found in the data was NaN. Days when a particular pollutant wasn’t recorded. On diving deeper, the number of missing observations were as follows.

```
so2     69
co      54
o3      43
no2     42
pm10     9
pm25     9
date     0
dtype: int64
```

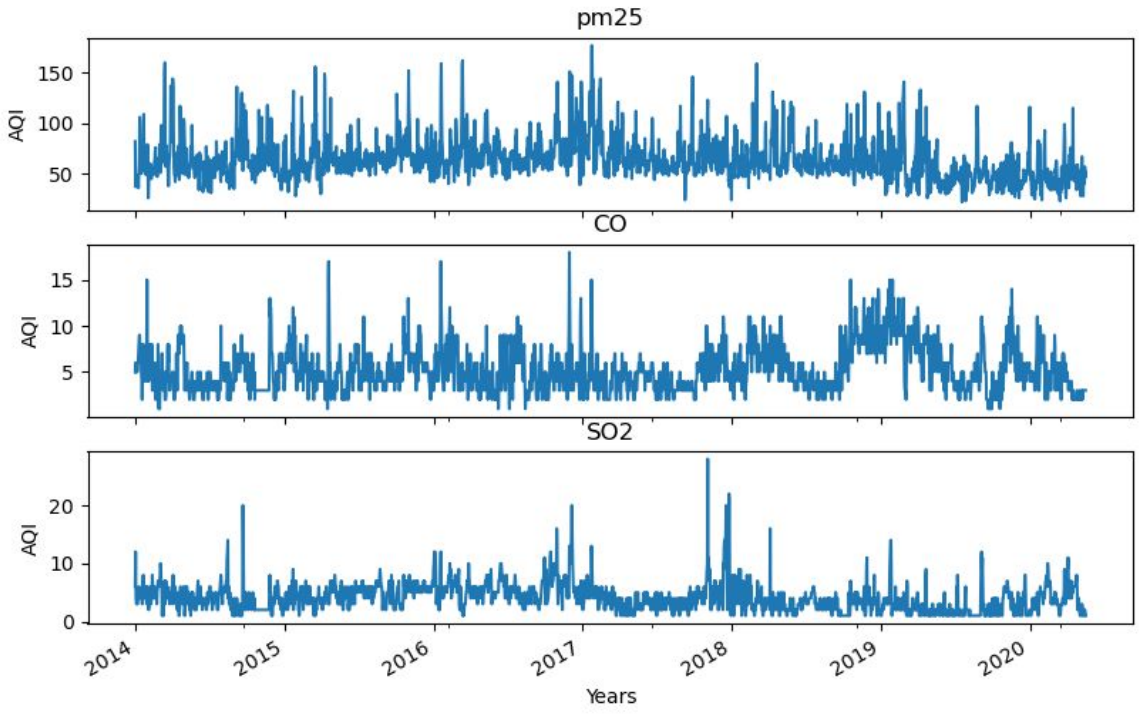
I filled the missing values using the ‘ffill’ method of pandas library. It fills the cell with the previous observed value. The next thing I did was to make columns for days, months and years to study the trend of the air quality index and infer useful insights.

Analysis

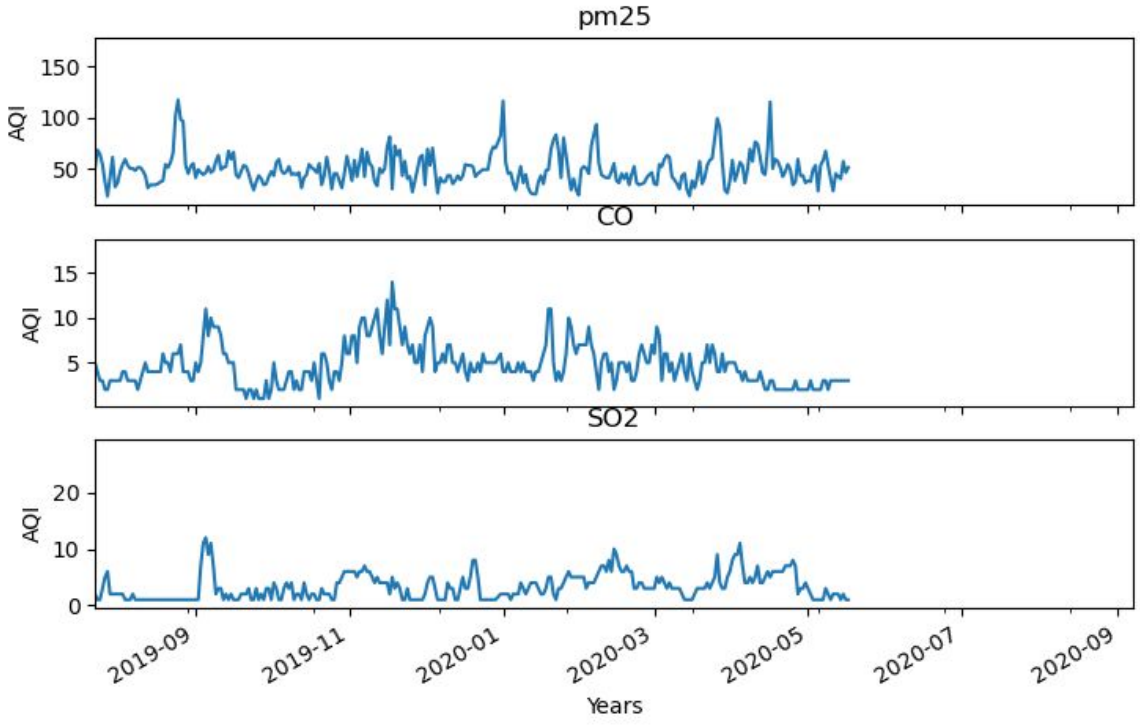
The data for analysis comprises the analysis of the major vehicular emissions viz. CO and pm2.5 (vehicles also emit other pollutants, but they are in abundance and more common) and the major industrial air pollutant, SO_2. Most of the following analysis will be graphical because it is easier to read a graph than a continuous set of data values.

Vehicular and Industrial Emissions Analysis (pm2.5, CO and SO_2)

Let's first plot the variations of these 3 pollutants over time.

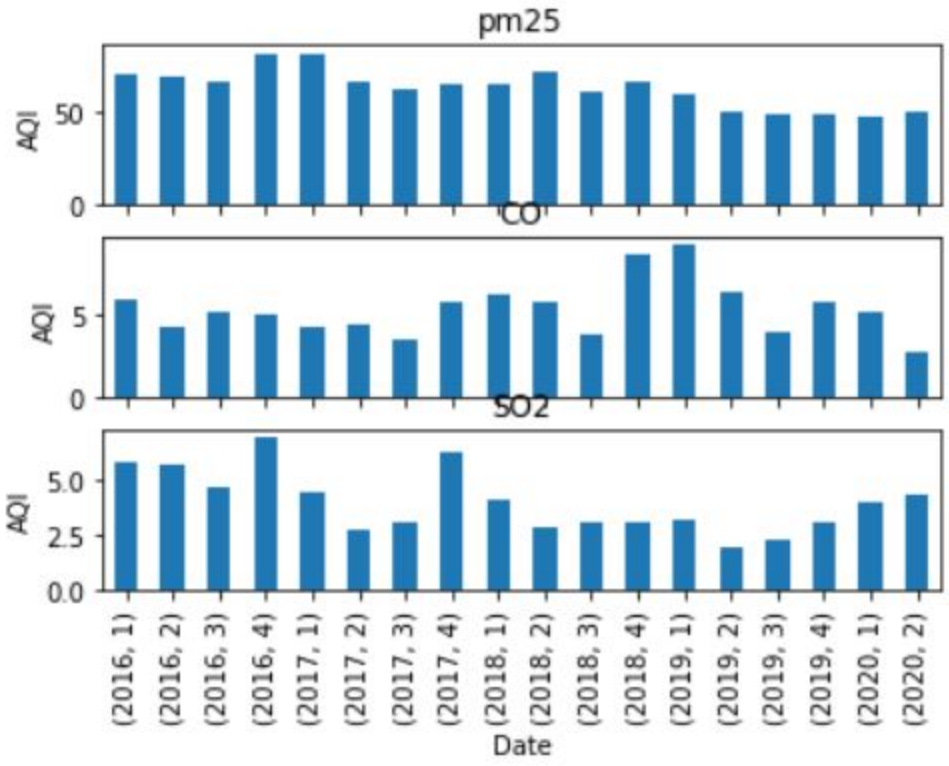


No definite inference could be drawn from this data at first look. But let's zoom in the 2020 period to see if we can find any trend.



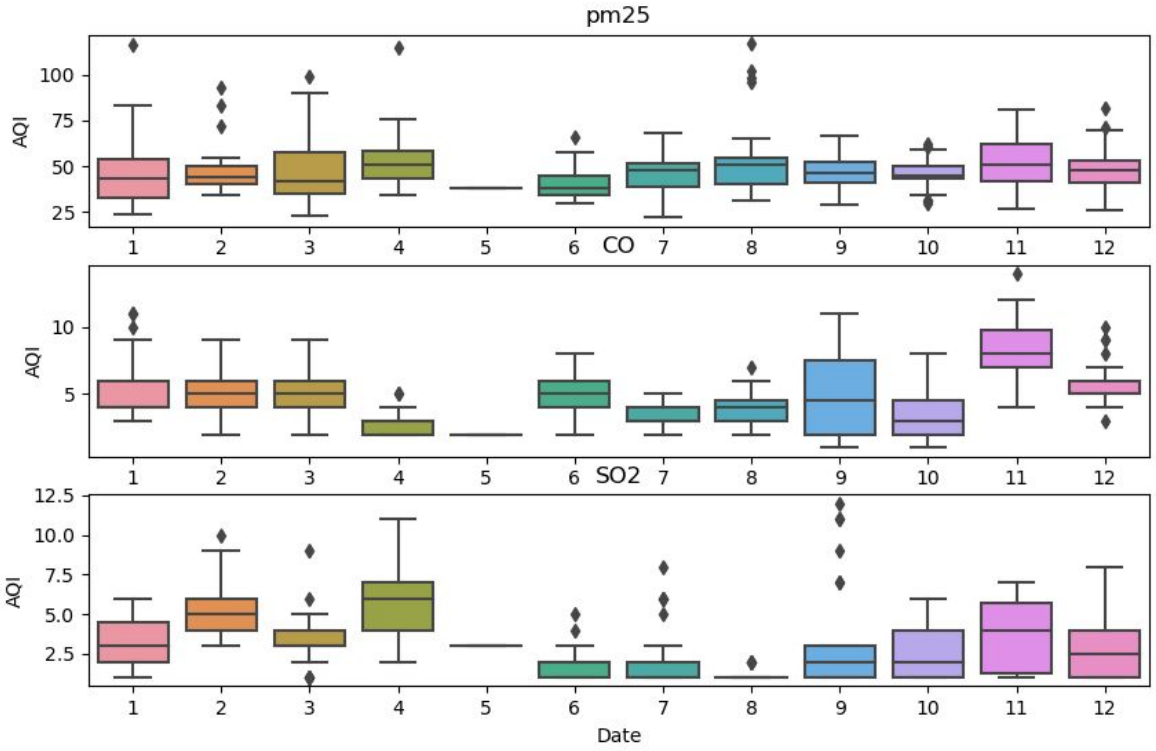
And yes. As I anticipated, there is a downward trend in the first quarter of the year 2020.

Now let's dive deep into further analysis.



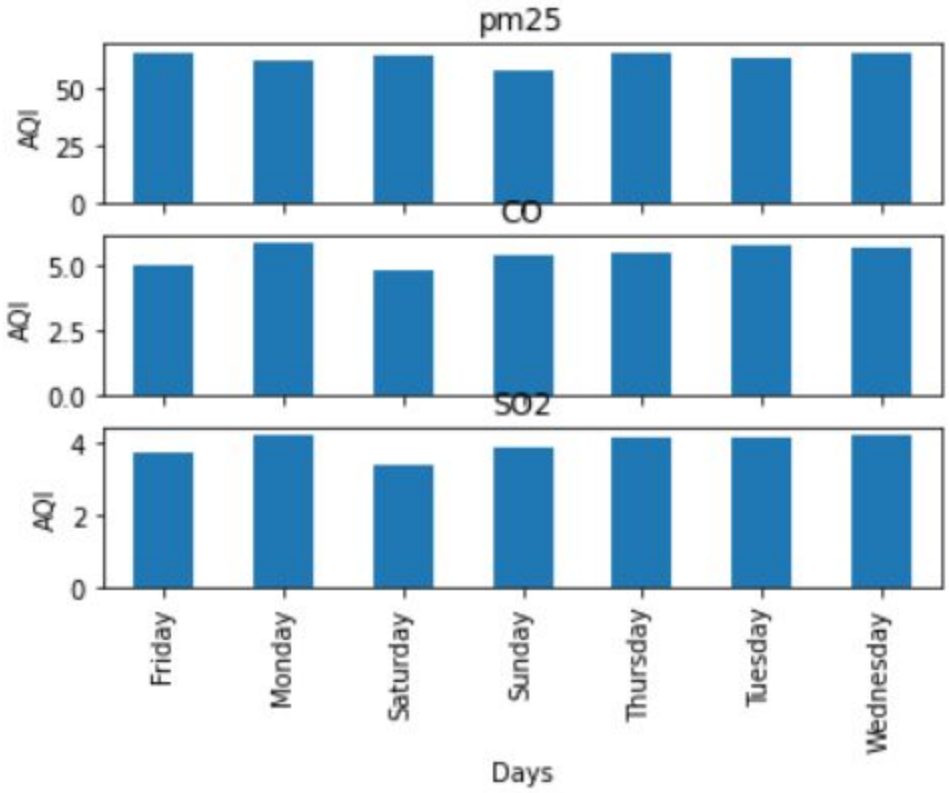
Here, I've grouped both pm2.5 and co mean AQIs with respect to the quarters of the year, since 2016. If we check the 2nd quarter of each year, since it's that time when lockdown was declared. But one thing that's to be noted here is that the data for 2nd quarter of 2020 is not complete since it only has records till 1st of May, 2020.

Till now, nothing concrete was inferred. Let's further analyse our data. For better understanding the effect of lockdown, let's analyse the AQIs for the past 1 year starting from "01-06-2019" to "01-05-2020".



The above figure is a box plot of the AQIs. As we can see, the month of April (denoted by 4 on the x-axis) has the lowest CO emission because CO is mainly emitted from vehicles and in April 2020 the city saw very few to no vehicles on the street. As of pm2.5, we can't account it's measure solely to cars therefore there is this discrepancy. While SO_2 has also shown discrepancy, I don't know the exact reason but can be ignored for the time being.

Another interesting trend I tend to observe from the data is how the value changes on weekends and weekdays.



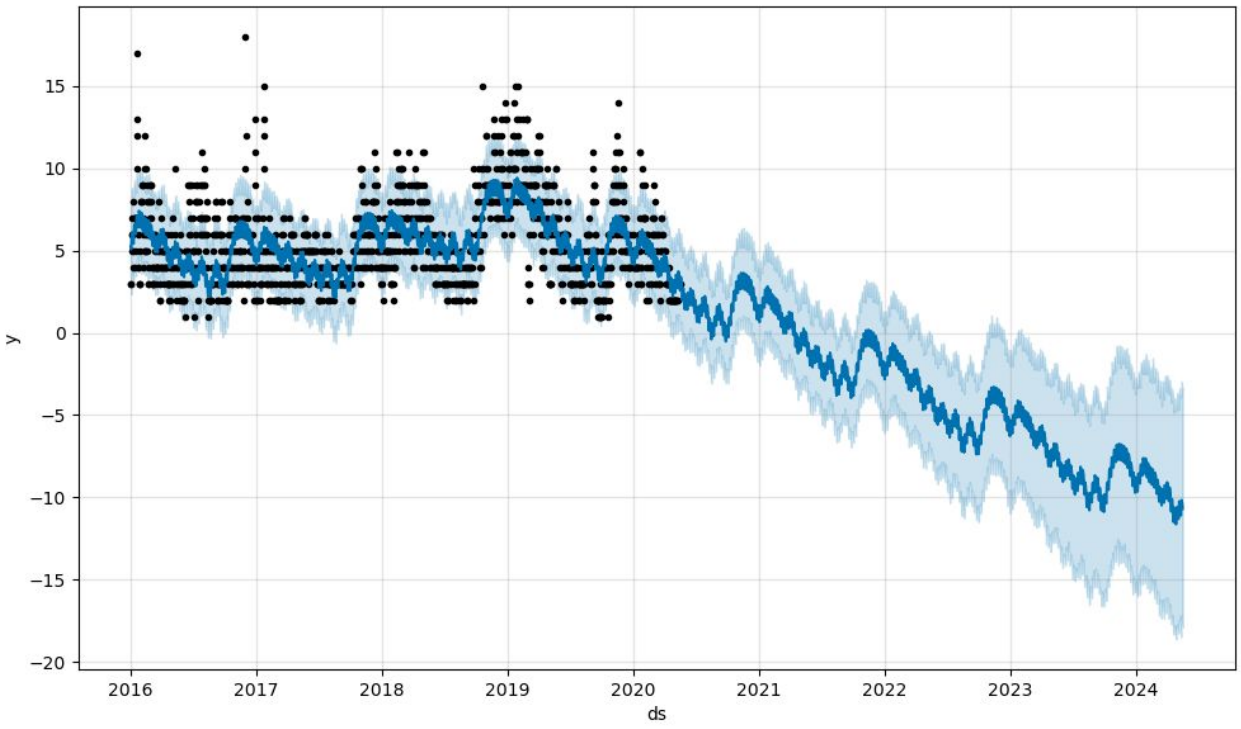
Surprisingly, Sundays have more CO emission than Fridays. Maybe Londoners love to travel on Sunday! Anyway, the Saturday figures justify my previous assumption of lower emission on weekends.

Finally, I'd like to conclude that the AQI values have significantly reduced during the lockdown. In the next section I'd like to forecast this trend and predict the values of pm2.5, CO and SO_2.

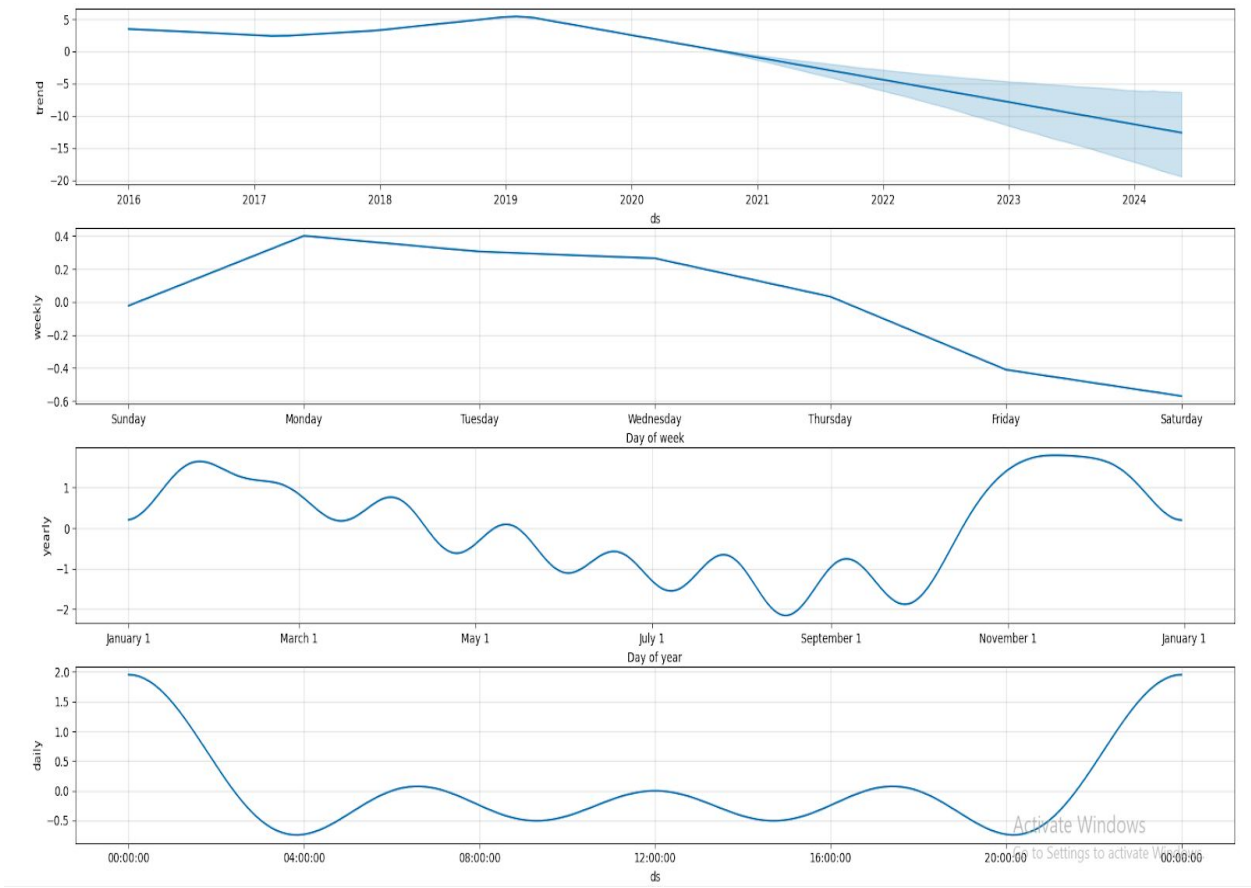
Forecasting

For the forecasting of this data I've used the fbprophet tool developed by Facebook core data science team, to forecast time series data. The official documentation of the Prophet can be found here. I chose fbprophet over other statistical models such as ARIMA and MA because it's easier to implement and can be tweaked easily to get better results.

Here follows the forecast of CO AQIs for the next 4 years. I chose only CO for forecasting because as evident from the previous data analysis, CO levels are most affected due to Covid-19 lockdown.



The black dot denotes the actual values for the past 4 years. As we can see, there's a decreasing trend of "CO" AQI values for the next 4 years, if we follow the current situation. This can be further clarified by looking at the plot of this forecast features.



As we can see from the plots, the trend seems to fall from 2020 onwards and a possible explanation for this is the lockdown, but I'm not fully sure because of minimum available data and

Although the model might have neglected the lockdown period discrepancy values as an outlier, if we continue the lockdown restrictions to check our vehicular and industrial use, we can achieve even lesser CO AQIs in future.

Conclusion

In conclusion, I'd like to point out the fact that if we check our air pollutant emissions, be it vehicular, industrial, etc. we can build a better environment to breathe for humans and animals of this and upcoming generations. Lastly, I am confident that we all will fight together against this virus and get back to our normal lives.

References

1. World Air Quality Index Project

   Link - https://aqicn.org/

2. Air Quality Index

   Link - https://en.wikipedia.org/wiki/Air_quality_index

3. Fbprophet official documentation

   Link - https://facebook.github.io/prophet/docs/quick_start.html

Code Repository

The code for this analysis can be found in my GitHub account. The link to this repository is as follows :

Link - https://github.com/IP1102/Covid_AirPollution_Analysis