By-ANUPAM SRIVASTAVA

INTERNSHIP REPORT

Data Analytics on Air Pollution

Topic:-IMPACT OF AIR POLLUTION ON OUR LIVES

Introduction

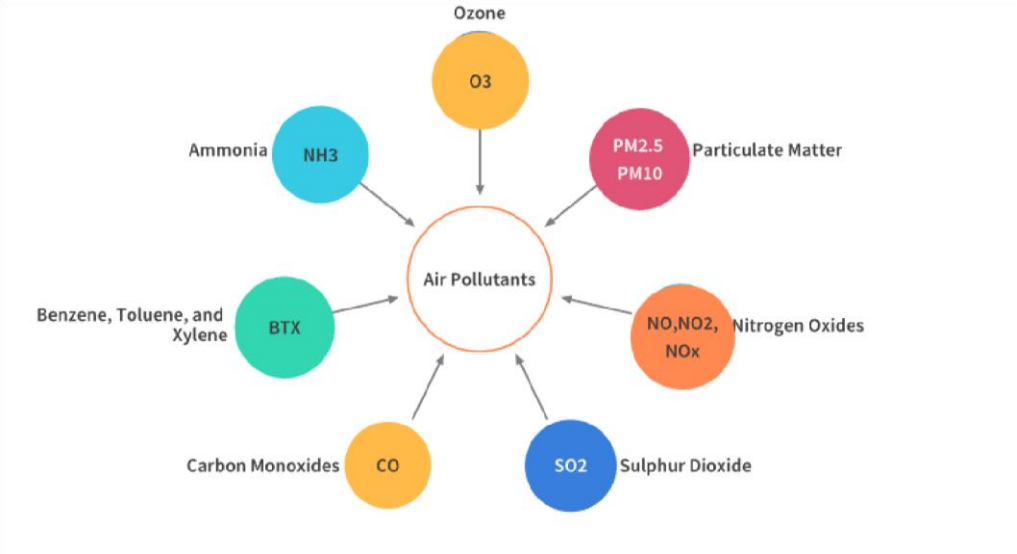
In these ongoing occasions, humankind is compromised by a pandemic brought about by another coronavirus. Early this year, after an episode in China, WHO recognized this new kind of coronavirus as SARS-Cov-2 having a place with one of the seven coronavirus. Covid-19 is the malady brought about by SARS-Cov-2, which is a respiratory sickness and can influence both the upper and lower respiratory tracts. To battle the spread of this infection, pretty much every other significant city on the planet has been secured and individuals are mentioned to keep up social separating. True to form, the economy has smashed and individuals having a place with the lower pay band are influenced the most. Government and different authorities are making a decent attempt to resuscitate the circumstance yet specialists foresee it's a significant time before we return to typical. Innumerable individuals have lost their lives and millions are contaminated. Consistently wellbeing authorities are endeavoring to spare the ones contaminated. Be that as it may, among all these terrible news, Covid-19 has assisted with restoring our the unstoppable force of life a piece. In spite of the fact that the lockdown has cost numerous individuals their employments, the lockdown has exponentially decreased the contamination levels. Be it air, water or soil, the urban areas under lockdown have seen critical decrease in their contamination level. Despite the fact that the harm brought about by Covid-19 is entirely terrible, on the opposite side it attempted to recuperate our condition.

Objective

The primary objective of this report is to analyse air quality of various air pollutants both previously and after lockdown period. The city that I choose for the analysis is India –all cities include(2015-2020). The main focus in this report is on the pollutants emitted by vehicles and industries taken as key points.

Types of Air Pollutants

Let's first try and understand the various types of air pollutants in the datasets. On a broader level, these pollutants can be classified as:



- **Particulate matter (PM2.5 and PM10)** > Particulate matter is a mix of solids and liquids, including carbon, complex organic chemicals, sulphates, nitrates, mineral dust, and water suspended in the air. PM varies in size. Some particles, such as dust, soot, dirt or smoke are large or dark enough to be seen with the naked eye. But the most damaging particles are the smaller particles, known as PM10 and PM2.5. Source. The following diagram will help to understand the concept more concretely.

- **Nitrogen Oxides (NO, NO2, NOx)** > Nitrogen oxides are a group of seven gases and compounds composed of nitrogen and oxygen, sometimes collectively known as NOx gases. The two most common and hazardous oxides of nitrogen are nitric oxide (NO) and nitrogen dioxide (NO2)

- **Sulphur Dioxide (SO2)** > Sulphur dioxide, or SO2 is a colourless gas with a strong odour, similar to a just-struck match. It is formed when fuel containing sulphur, such as coal and oil, is burned, creating air pollution.

- **Carbon Monoxide (CO)** > Carbon monoxide is a colourless, highly poisonous gas. Under pressure, it becomes a liquid. It is produced by burning gasoline, natural gas, charcoal, wood, and other fuels.

- **Benzene, Toluene and Xylene (BTX)** > Benzene, toluene, xylene, and formaldehyde are well-known indoor air pollutants, especially after house

decoration. They are also common pollutants in the working places of the plastic industry, chemical industry, and leather industry

- Ammonia(NH3) >Ammonia pollution is pollution by the chemical ammonia (NH3) – a compound of nitrogen and hydrogen which is a by-product of agriculture and industry.

- Ozone(O3) >Ground-level ozone is a colourless and highly irritating gas that forms just above the earth's surface. It is called a "secondary" pollutant because it is produced when two primary pollutants react in sunlight and stagnant air. These two primary pollutants are nitrogen oxides (NOx) and volatile organic compounds (VOCs)

Causes Of these pollutants are:-

- Vehicles: Transportation majorly emits Carbon Monoxide (CO) and Nitrogen Oxide (NO). It minorly emits Ozone(O3) and Particulate Matter (PM2.5 and PM10).

- Industries: Industries majorly emit Sulphur Dioxide (SO2), Carbon Monoxide (CO), Ammonia (NH3) and Particulate Matter (PM2.5 and PM10), BTX (Benzene, toluene, xylene)

Data

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.

Xylene 16807
PM10 10766
NH3 9847
Toluene 7555
Benzene 5287
PM2.5 4289
AQI_Bucket 4282
AQI 4282
NOx 4043

```
O3          3660
SO2         3544
NO          3233
NO2         3217
CO          1961
Date           0
City           0
dtype: int64
```

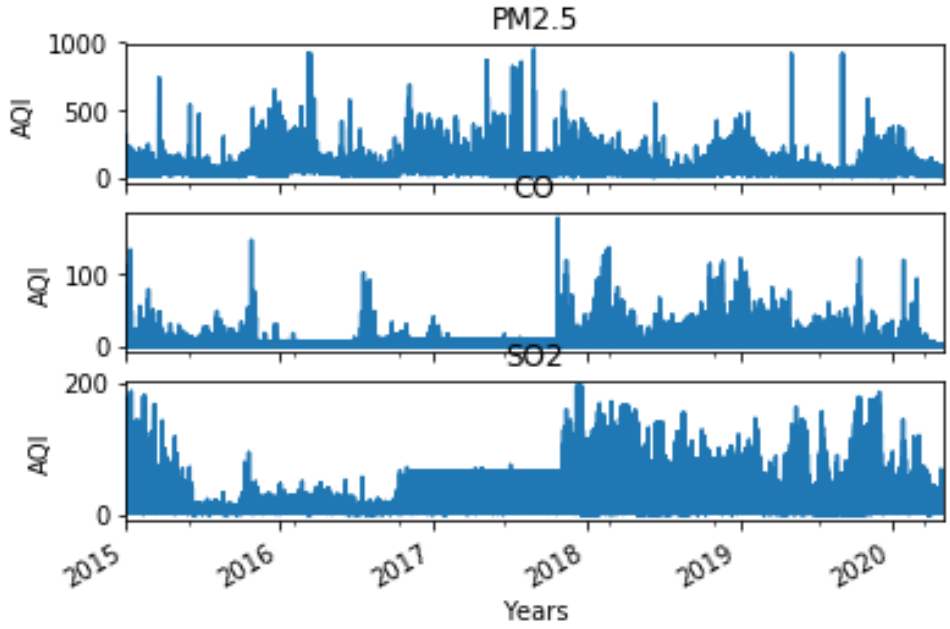
I filled the missing values using the 'ffill' method of pandas library. It fills the cell with the previous observed value. The next thing I did was to make columns for days, months and years to study the trend of the air quality index and infer useful insights.

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 tim

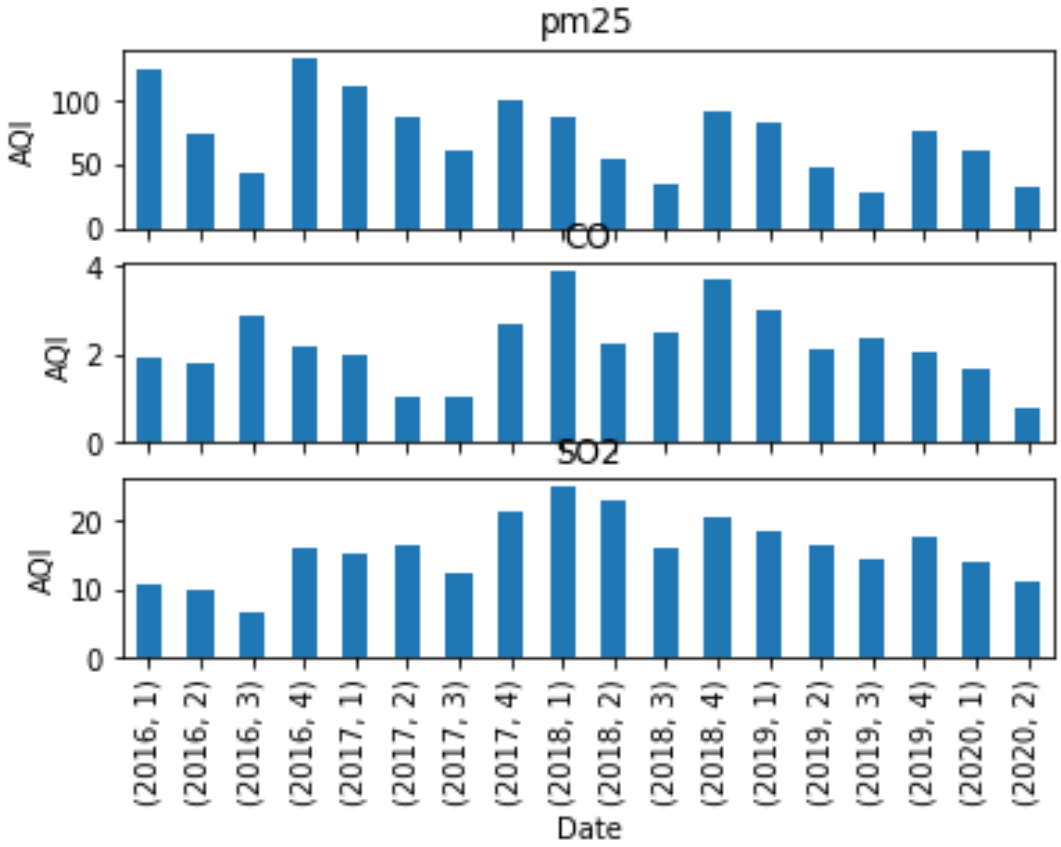


e.

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.

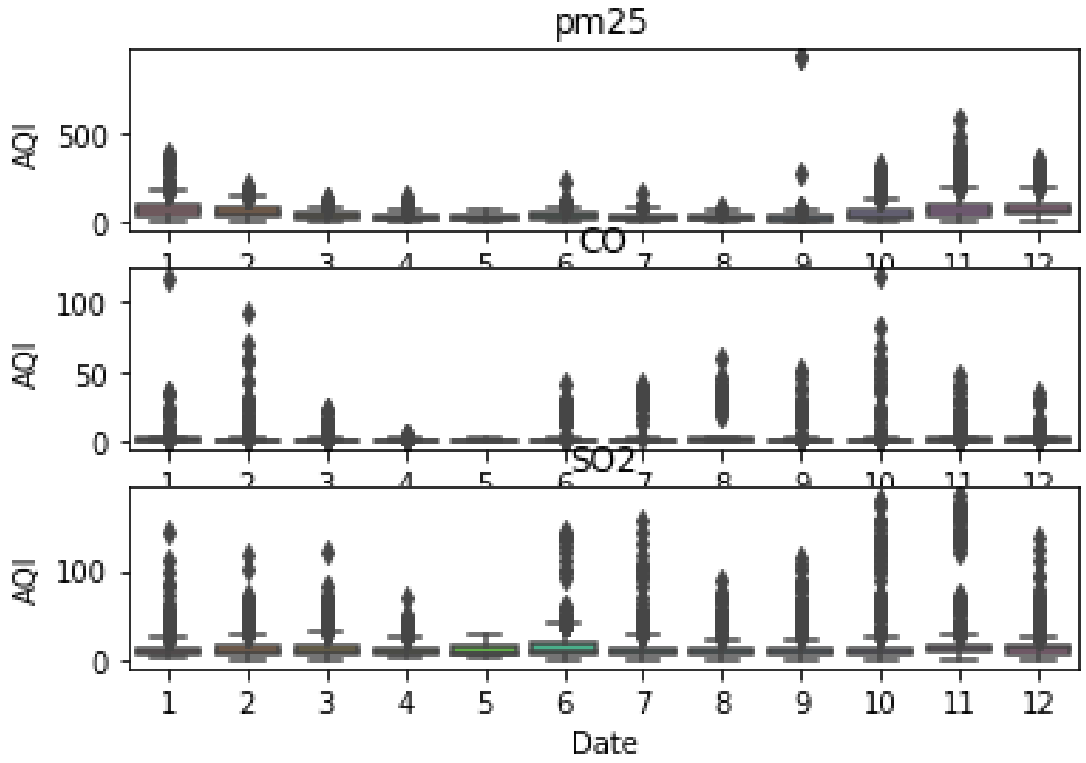
And yes. As I anticipated, there is a downward trend in the first quarter of the year 2020.

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 2015. 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.



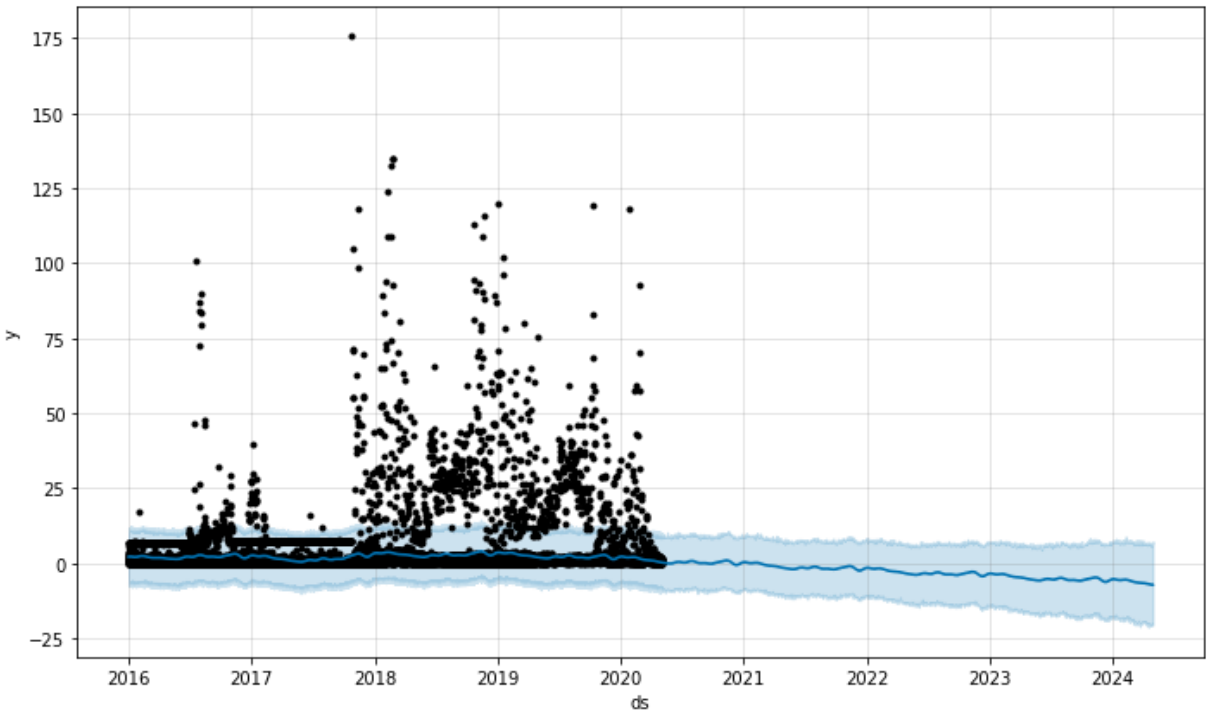
The above figure is a box plot of the AQIs. As we can see, the month of April (denoted by 4 on the x-axis) has the lowest CO emission because CO is mainly emitted from vehicles and in April 2020 the city saw very few to no vehicles on the street. As of pm2.5, we can't account it's measure solely to cars 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.

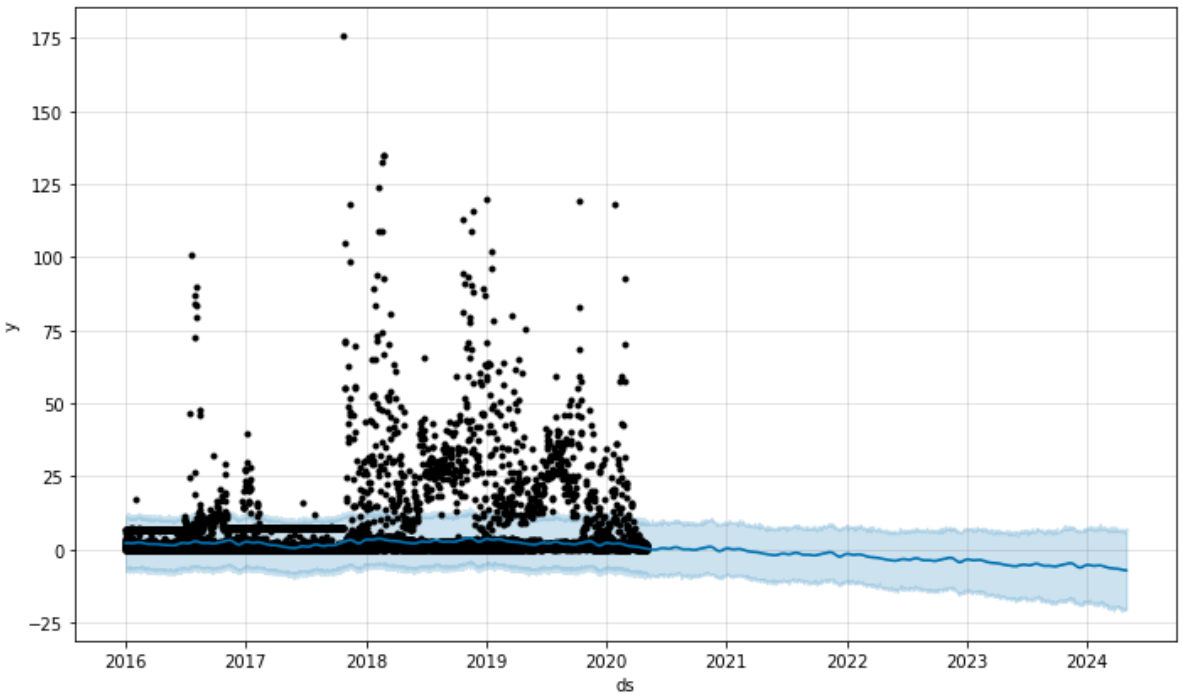
Another interesting trend I tend to observe from the data is how the value changes on weekends and weekdays.

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 .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.

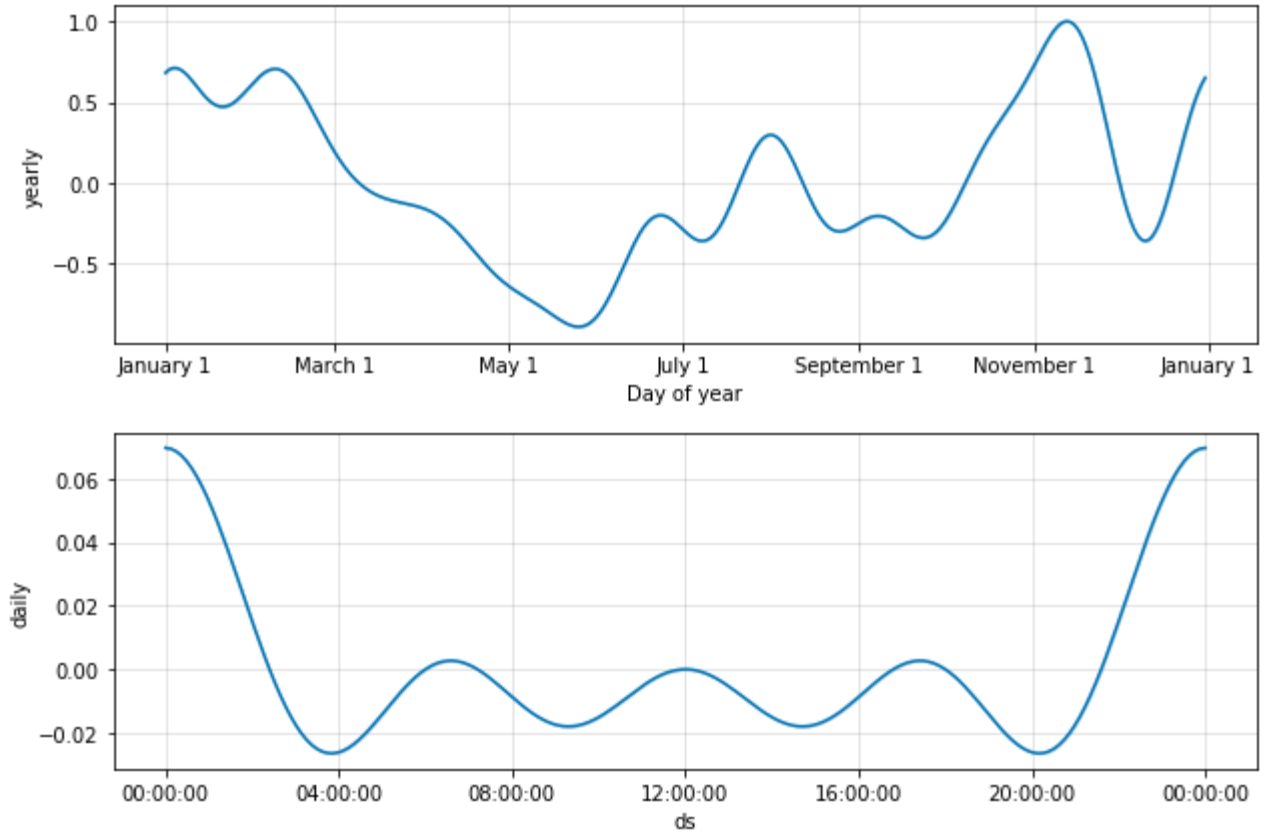
Here follows the forecast of CO AQIs for the next 4 years. I chose only CO for forecasting because as evident from the previous data analysis, CO levels are most affected due to Covid-19 lockdown.

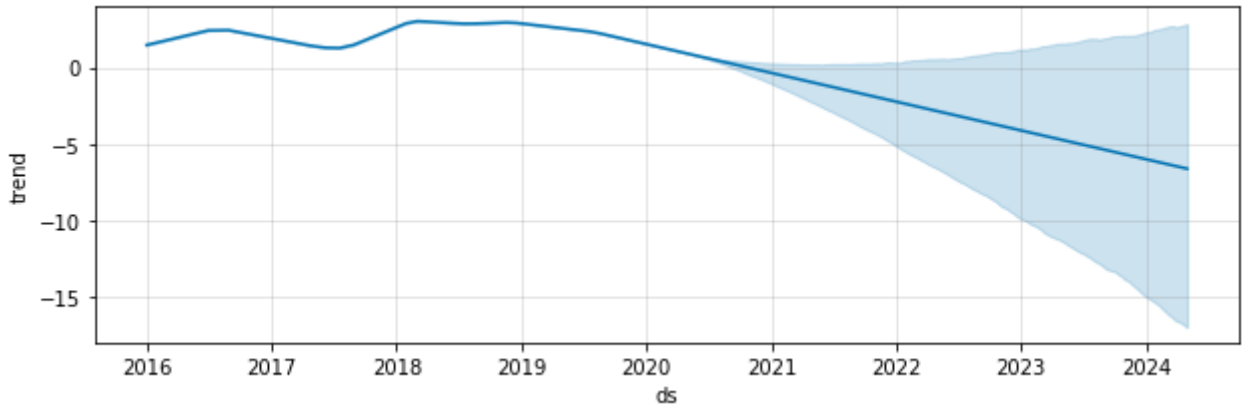


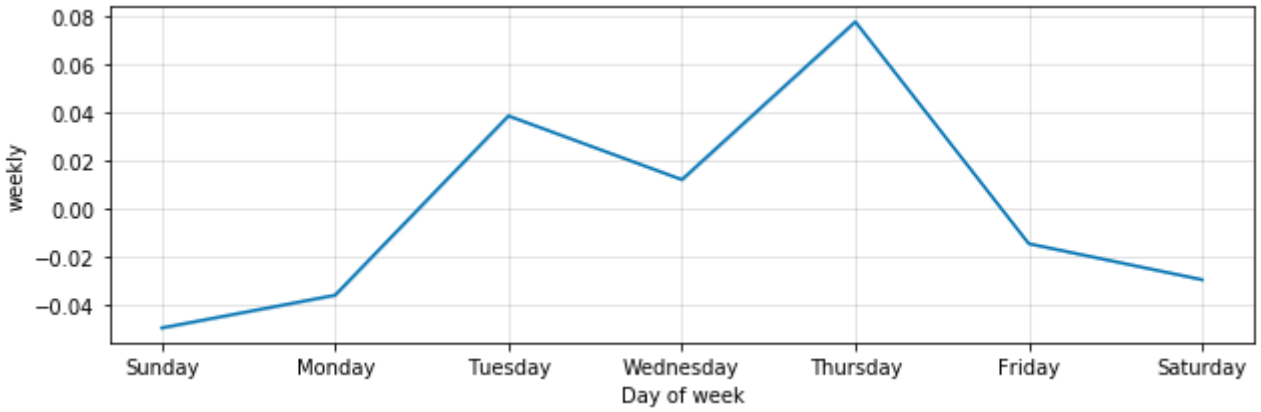


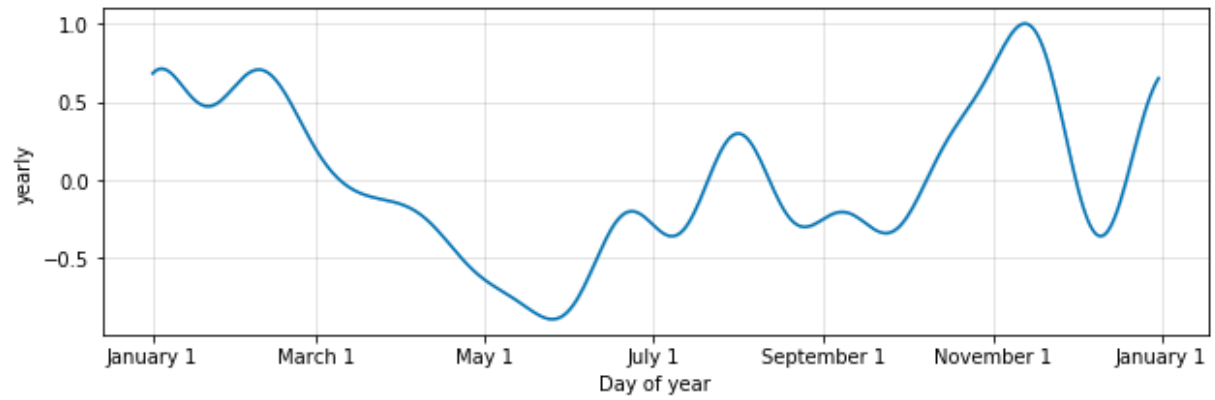
The black dot denotes the actual values for the past 5 years. As we can see, there's a decreasing trend of "CO" AQI values for the next 5 years, if we follow the current situation.

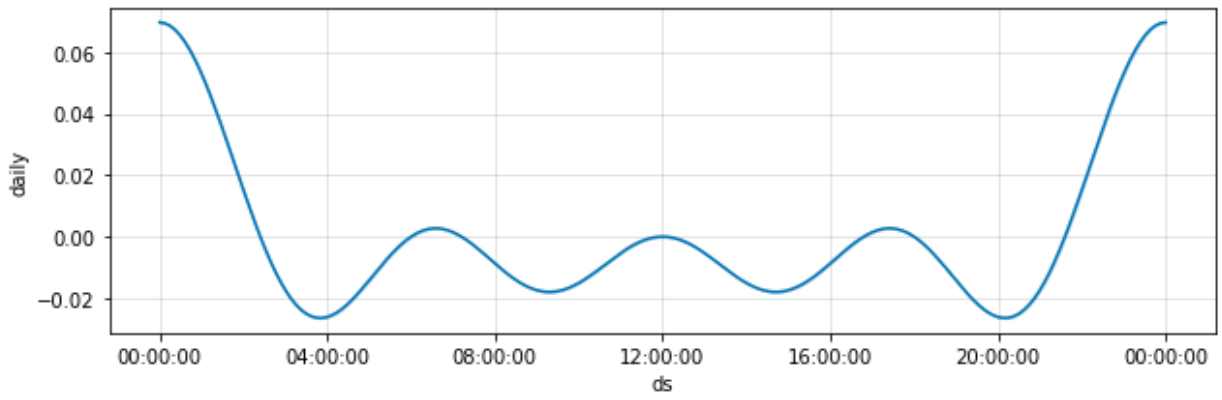
This can be further clarified by looking at the plot of this forecast features.

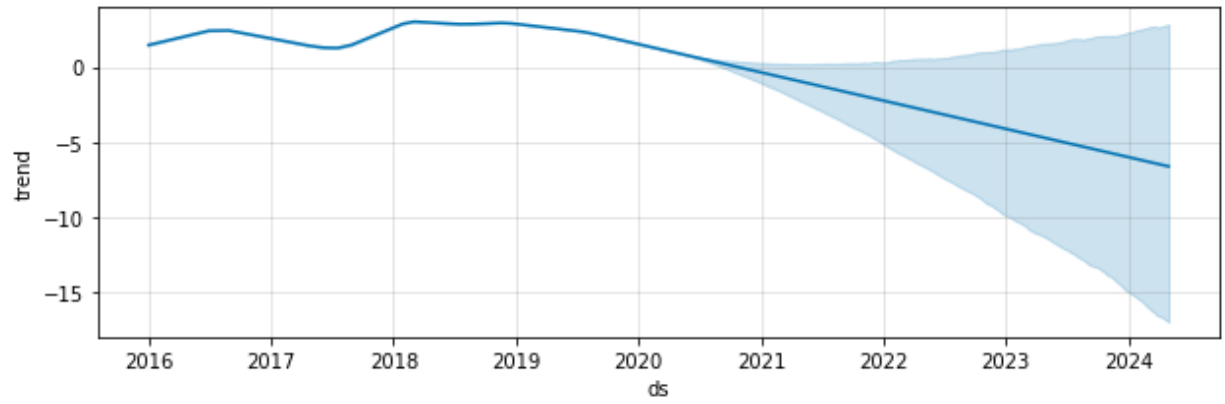


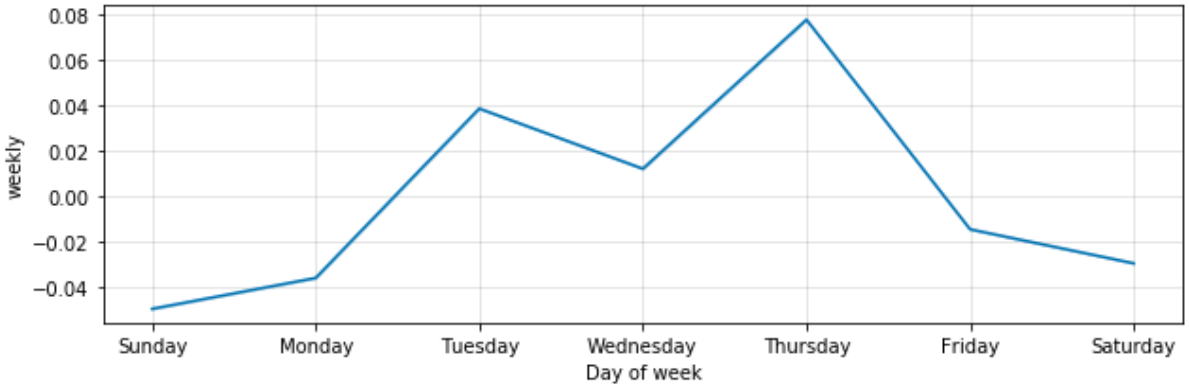

trend
0
−5
−10
−15
2016
2017
2018
2019
2020
2021
2022
2023
2024
ds


weekly
0.08
0.06
0.04
0.02
0.00
−0.02
−0.04
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Day of week


yearly
1.0
0.5
0.0
−0.5
January 1
March 1
May 1
July 1
September 1
November 1
January 1
Day of year


daily
0.06
0.04
0.02
0.00
−0.02
00:00:00
04:00:00
08:00:00
12:00:00
16:00:00
20:00:00
00:00:00
ds





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. Data source: https://www.kaggle.com/rohanrao/air-quality-data-in-india
2. Information source: Wikipedia and Google

For more data visualization and data analysis code visit my github repo.: https://github.com/anupamsri/-Data-Analytics-on-Impact-of-Air-Pollution-on-our-Lives