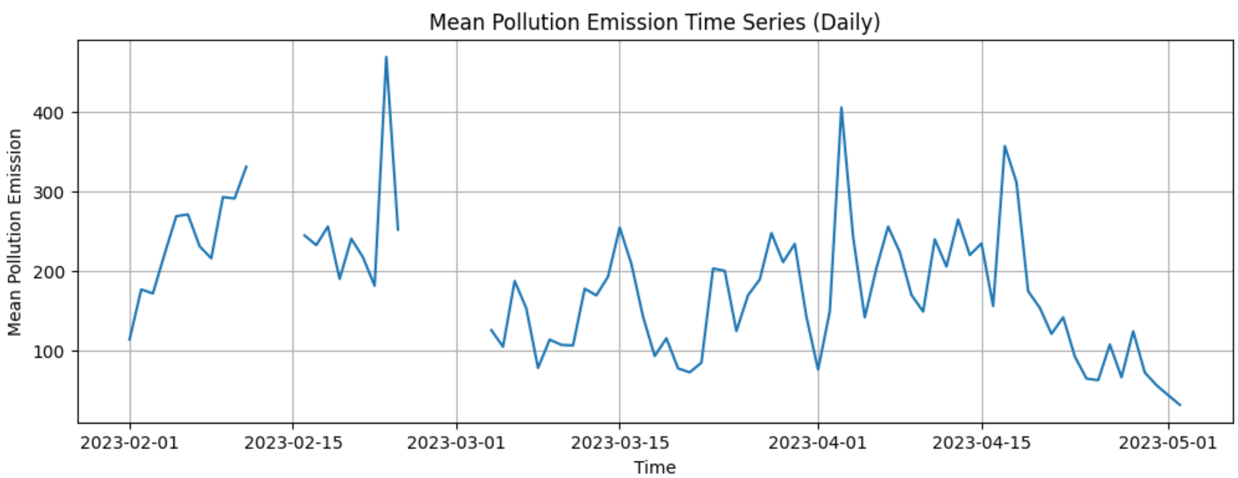
EE798: *FISA* Assignment

India's coal production industry is predominantly dominated by open-pit mining, a method known for its high productivity and cost-effectiveness. However, the future of coal demand is expected to face significant challenges due to mounting environmental concerns. The emission of particulate matter and other gaseous pollutants from diverse mining operations has contributed to worsening air quality and raised concerns about public health and climate change.

This analysis aims to provide a comprehensive understanding of India's coal production landscape, and the pressing environmental concerns associated with it. By leveraging statistical techniques, we can derive meaningful insights from the data and contribute to informed decision-making in energy and environmental policy domains.

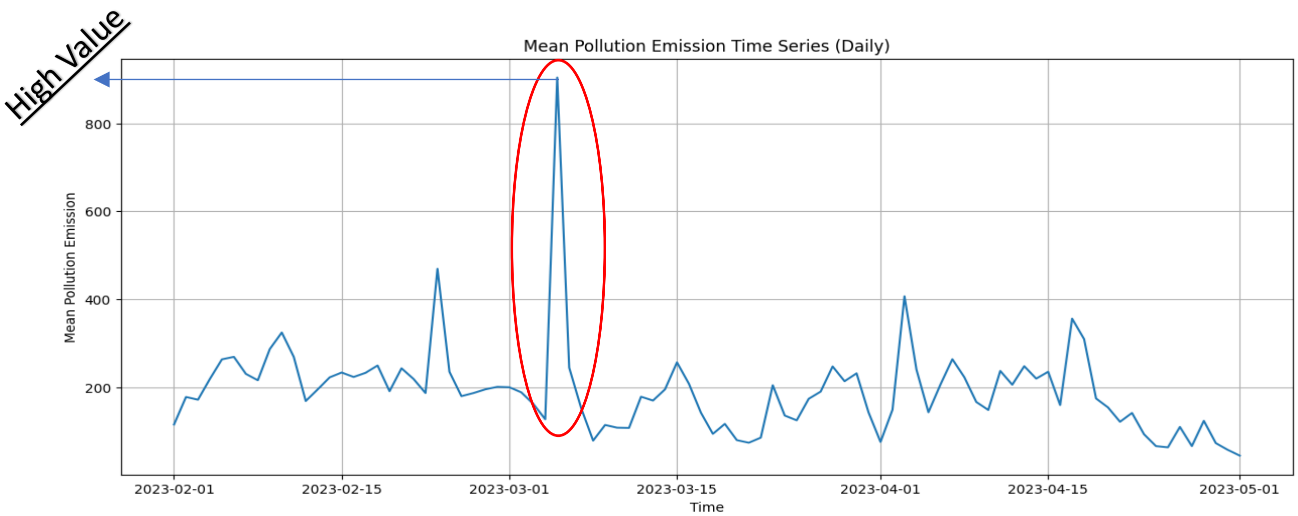
To begin with, there is a lot of null data due to some technical glitches. I have taken the daily average to plot the graph.



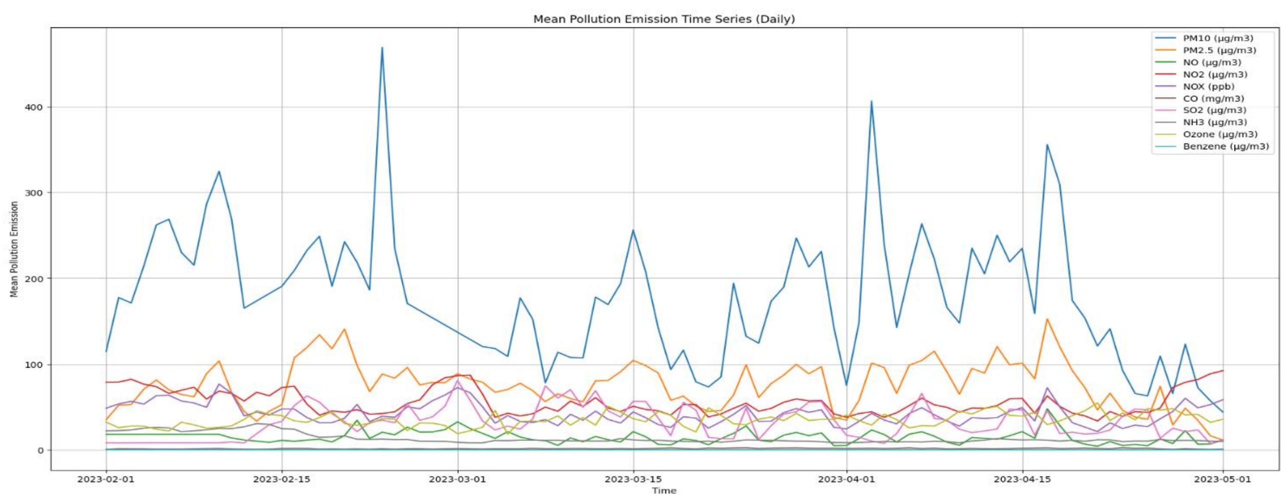
we set the null values to zero may not be the best strategy. It will change the meaning of entire data. Replacing NA values with 0 could introduce bias and affect the interpretation of the data.

Interpolation

First, tried cubic interpolation to fill the null values. It goes negative or very high values.

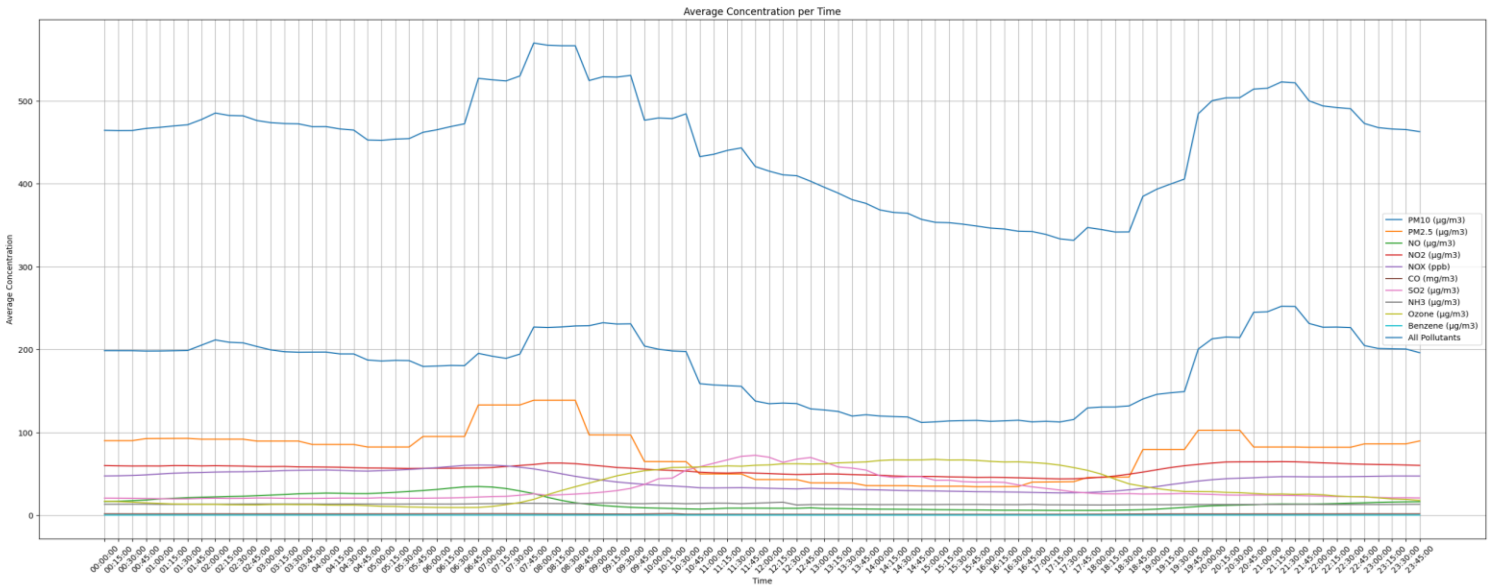


Here is the linear interpolation for filling the null values.



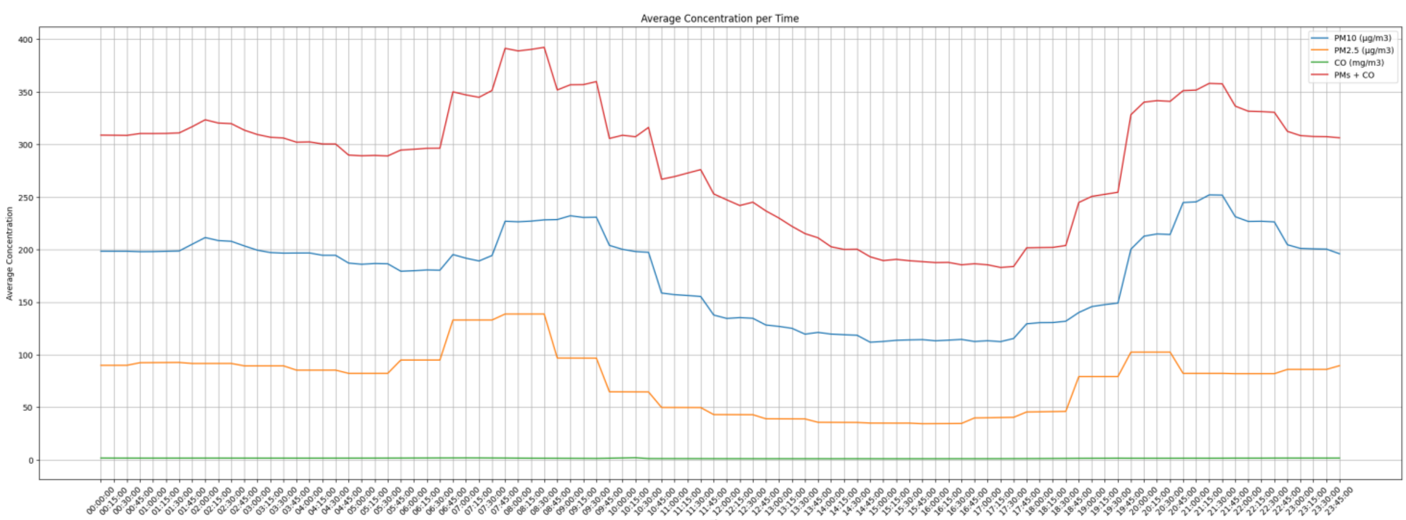
Statistical Inference

I have taken an average of time interval for the prediction of blasting time to validate the given data.

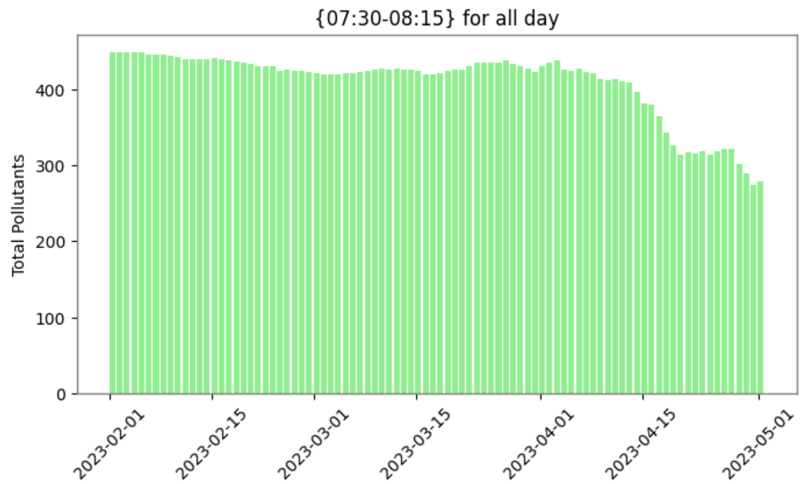


Here All Pollutants are the plot of summation of all pollutants.

When the pit blasting is done there is emission of Particulate matters and CO in large amounts. So further, I plot the graph to see when they are at peak to determine the time of pit blasting.



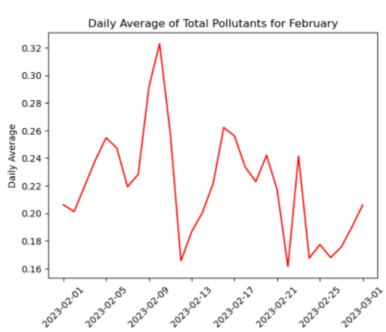
by above those graphs, there is possibility that Blasting time is in the interval of 07:15 – 08:15

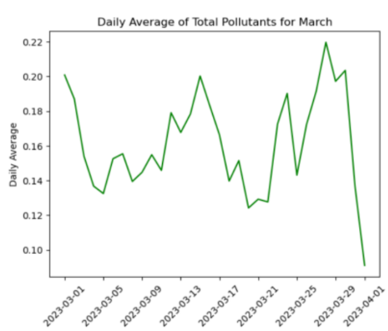


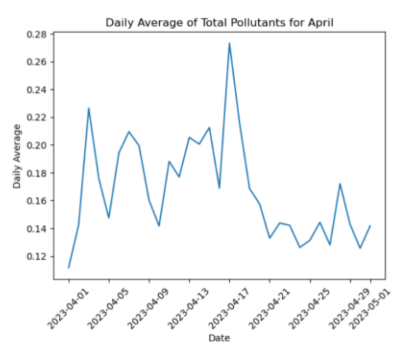
Here, by plotting the histogram for from inferred time of blasting, I can see that the everyday in this interval of time pollution rises to a constant level.
But, also there is decline in this pollution peak about in first week of April.

It is possibly due to changes in weather patterns such as increased wind speed or rainfall, which can help disperse pollutants and reduce their concentration in the air.
It is also possible that coal mine in Singrauli implement some emission control measures, or there might be some maintenance work going on in some part of ***Coal-mines****.*

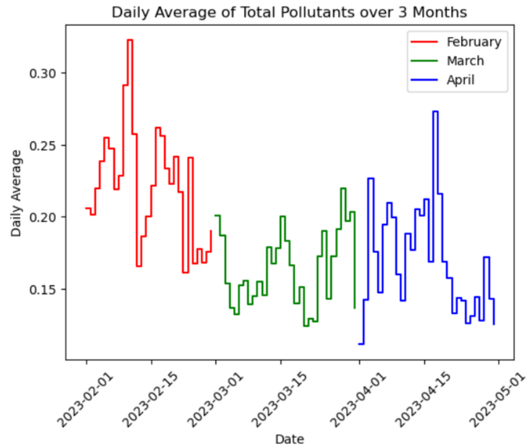
By assigning weights to pollutants and calculating their weighted averages after normalizing the data, daily averages were obtained for a period of three months. Graphs were plotted to depict the trends observed during each month, providing insights into the variations in pollutant levels. This approach offers a comprehensive understanding of the overall pollutant levels and their temporal pattern within the studied area. We notice that there are no obvious patterns in the pollutant levels. The graphs are irregularly shaped.







By combining the individual graphs into a single visualization, it becomes evident that the month of March exhibits the lowest average concentration of total pollutants, characterized by relatively lower peak values. The month of April closely follows March in terms of lower pollutant levels, while February has the highest levels. This consolidated view provides a clear comparison of the average pollutant levels across the three months, highlighting the contrasting pollution patterns.



Weighted average of all pollutants during February = 0.22061

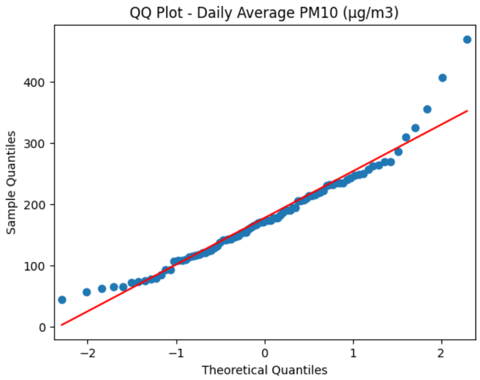
Weighted average of all pollutants during March = 0.1638

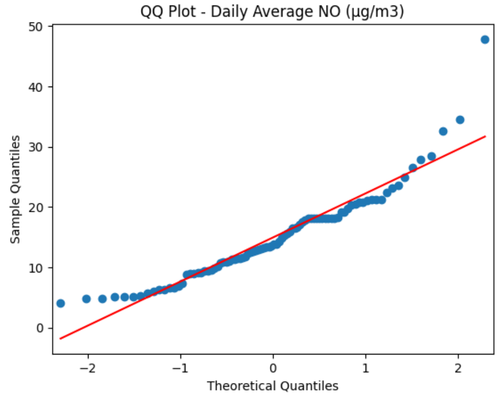
Weighted average of all pollutants during April = 0.1689

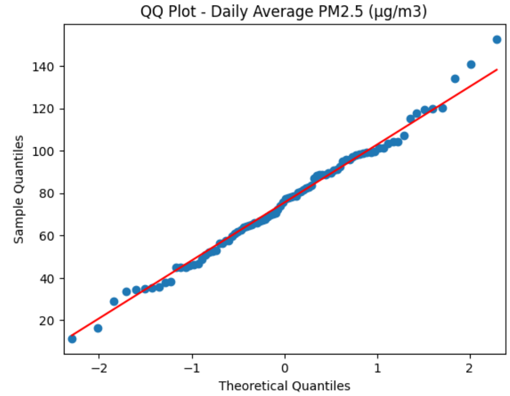
The last graph reveals significant variance in pollutant levels during the months of February and April, while March exhibits lower variance. Factors such as weather events, energy demand, emissions, and specific pollution sources can contribute to these variations. Weather conditions, including temperature inversions and stagnant air masses, may have influenced higher variance in February and April. The higher pollutant concentrations in February can be attributed to increased energy demand for heating purposes, while the transition to spring in March may have facilitated better pollutant dispersion and lower emissions. Other factors like industrial activities, transportation, nearby construction projects, agricultural practices, and localized emissions could also contribute to the observed differences.

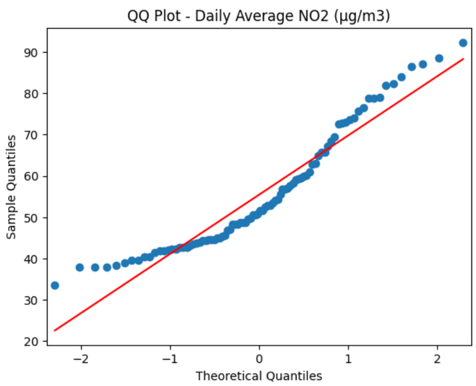
It's important to note that these factors are general and can vary depending on the location and specific pollutants of interest. Conducting a more comprehensive analysis, considering additional data sources, and consulting domain experts can provide a deeper understanding of the underlying reasons for the variance in pollutant levels.

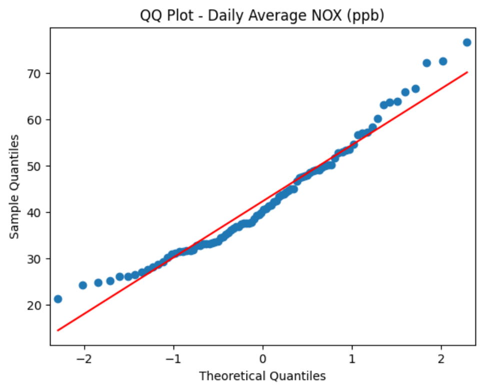
QQ-Plot

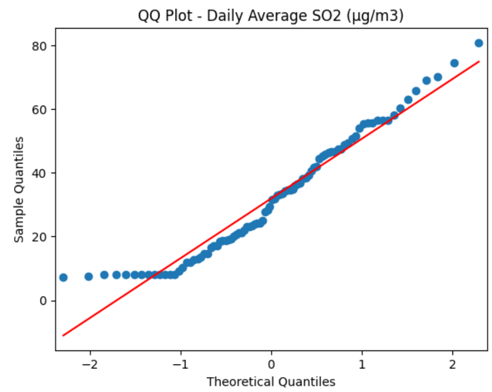


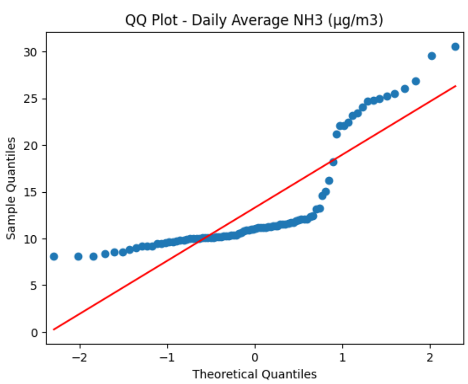


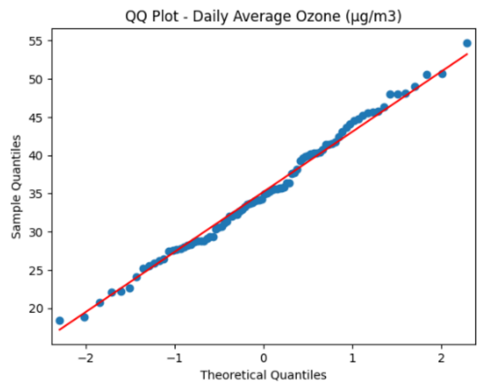


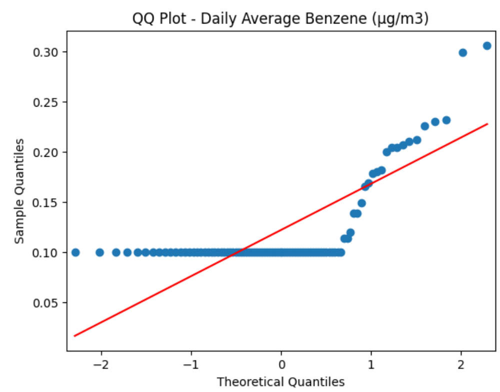


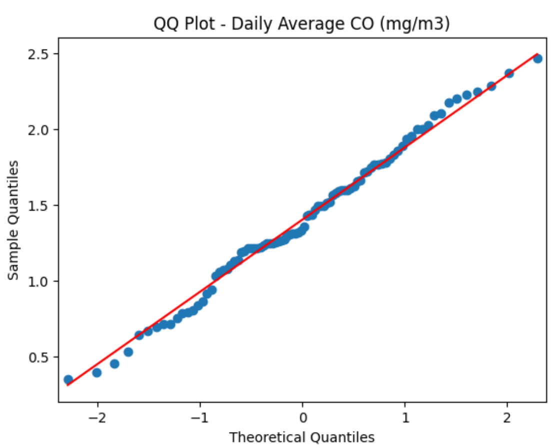












*By these QQ-Plots of all pollutant **Carbon Monoxide**, **Ozone** and **PM2.5** follows Normal distribution .*

PM2.5 is the only pollutant which has higher emission level due to pit blasting and shows normal distribution.

It is perfect for determining the timing of pit blasting by ***hypothesis test*** ***and*** ***p-value.***

Considering time interval (7:15-8:15)

$\bar{X} = 75.55$ $\sigma = 27.50$ $n = 90$

Avg conc. of PM2.5 (7:15-8:15) $= 137.29$

Null Hypothesis : $H_0 : \mu = 137.29$

Alternate Hypothesis : $H_A : \mu \neq 137.29$

$$t\text{-score} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \cancel{75.55 - 137.29}$$

$$= \frac{75.55 - 137.29}{27.5 / \sqrt{90}} = -21.29$$

P value : 0.0000

which is << 0.05

Degree of freedom = 89

one tail

Significance level : 0.05

$\therefore$ we can reject H_0

$\Rightarrow$ there is large ~~in~~ increase in the concentration during pit blasting.

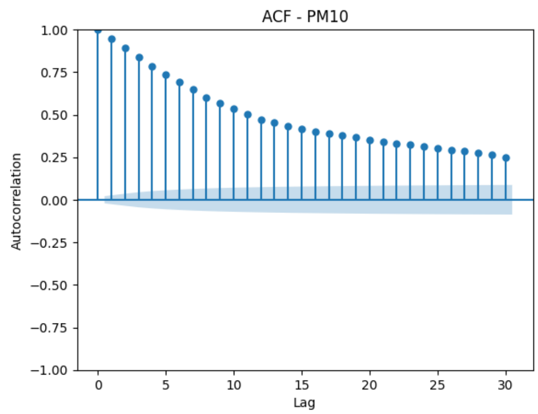
I took ***one-tail*** *analysis because I am just considering it to be increasing or remain-same not decreasing.*

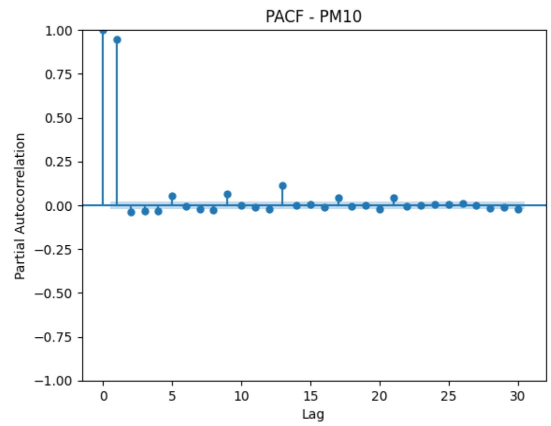
FORCASTING: ***I have done forecasting only for the next 2 days.***

*I have only used **ARMA, AR, MA** as there are no trends in the given data. Time series data which provided is already in stationary (mean, variance is unchanged).*

PM10

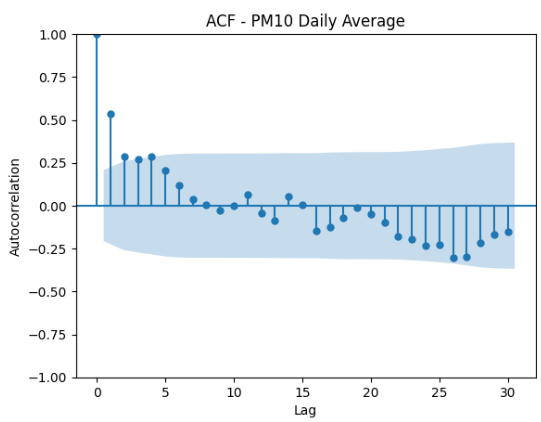
Initially I tried to make PM10 ACF and PACF for every 15 min interval.

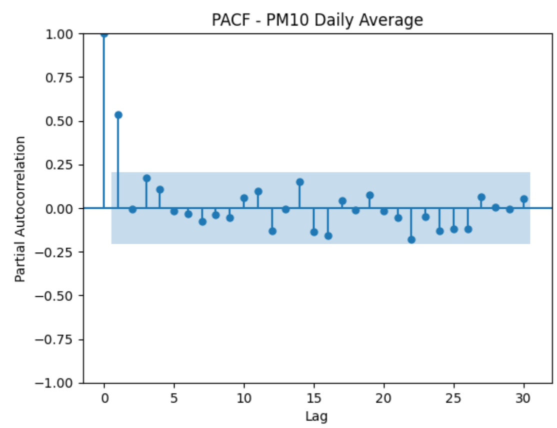




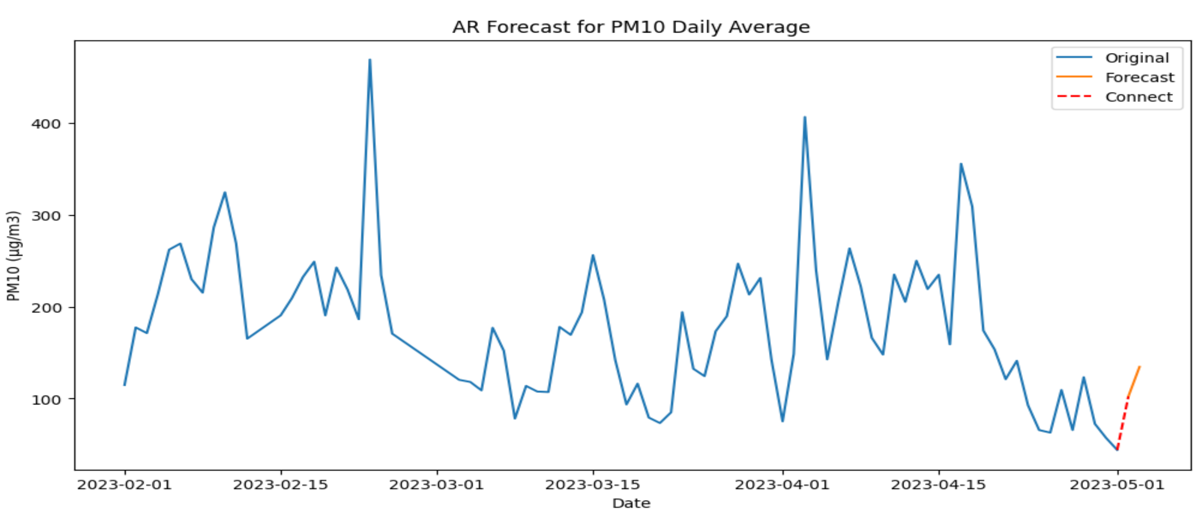
It is hard to analyze due to low visibility of significant spikes(lollipop). Which is necessary to determine which Time Series Model will apply for FORECAST. So, here I processed the data to daily average to increase the accuracy in applying the time series model.

So, further I have made ACF and PACF for Daily Average.





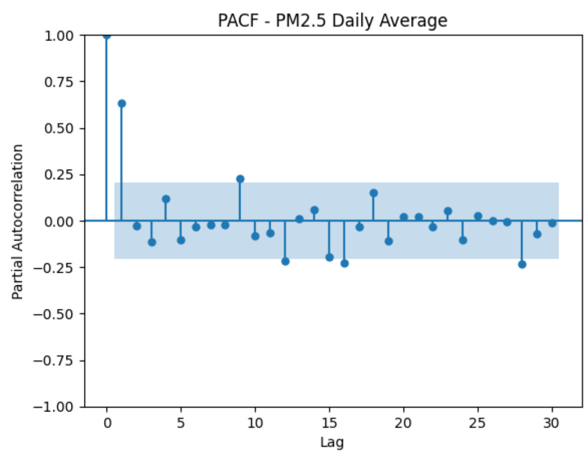
*By, this ACF and PACF of PM10 I can infer the forecast would be of **AR order 1 model**.*

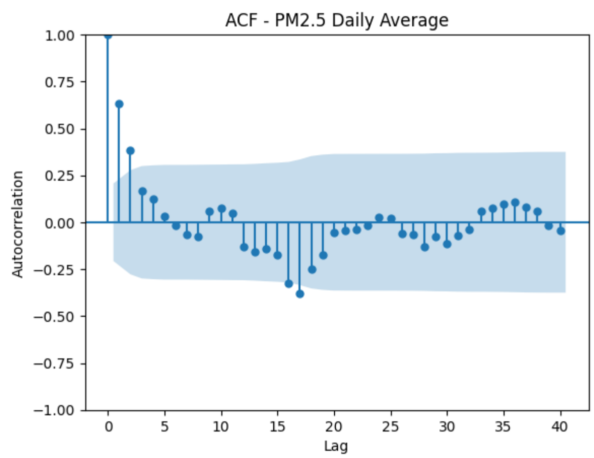


Forecasted data *2023-05-02* *<u>102.239089</u>*

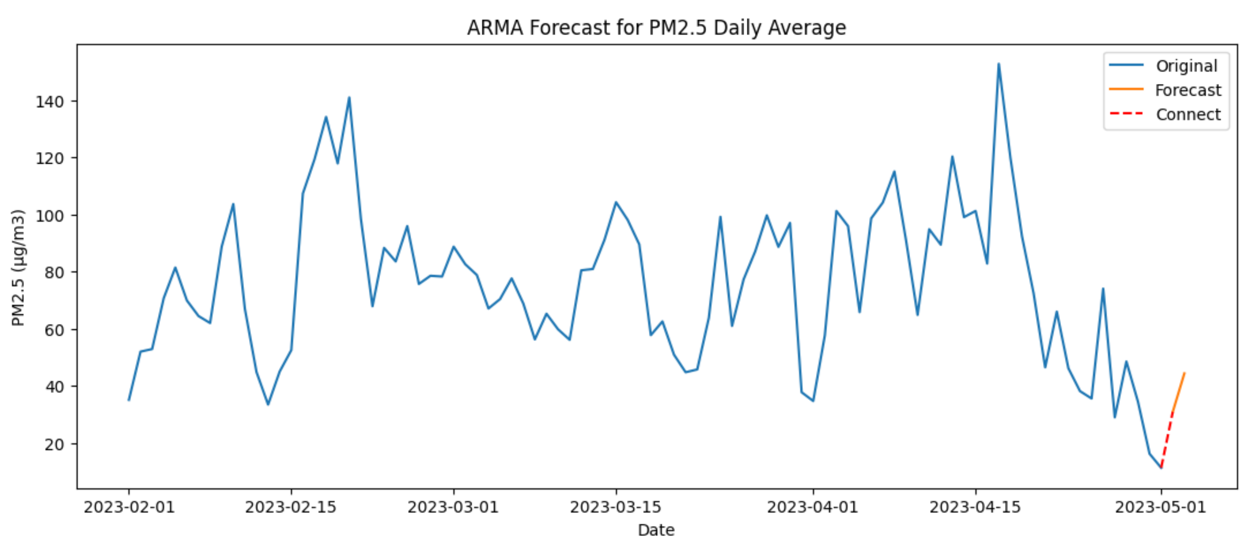
2023-05-03 *<u>134.542063</u>*

PM2.5





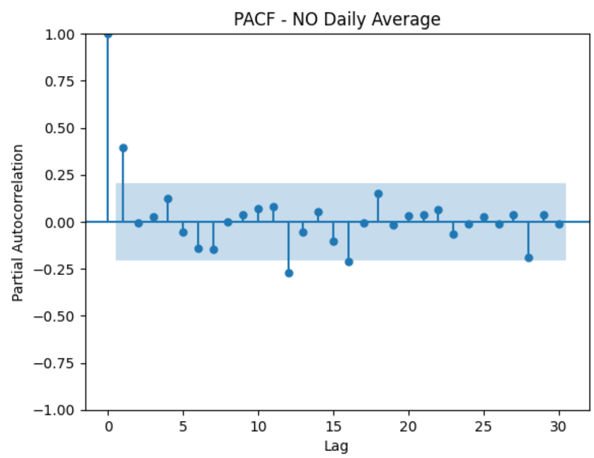
Here, in both ACF and PACF there are many significant spikes, so we use ***<u>ARMA (1,1).</u>***

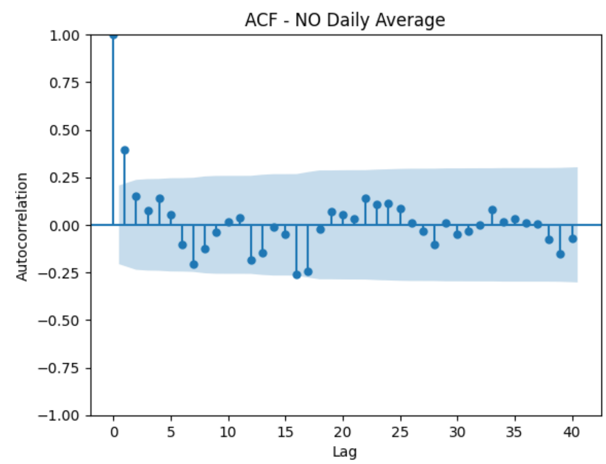


Forecasted data *2023-05-02* *<u>30.939603</u>*

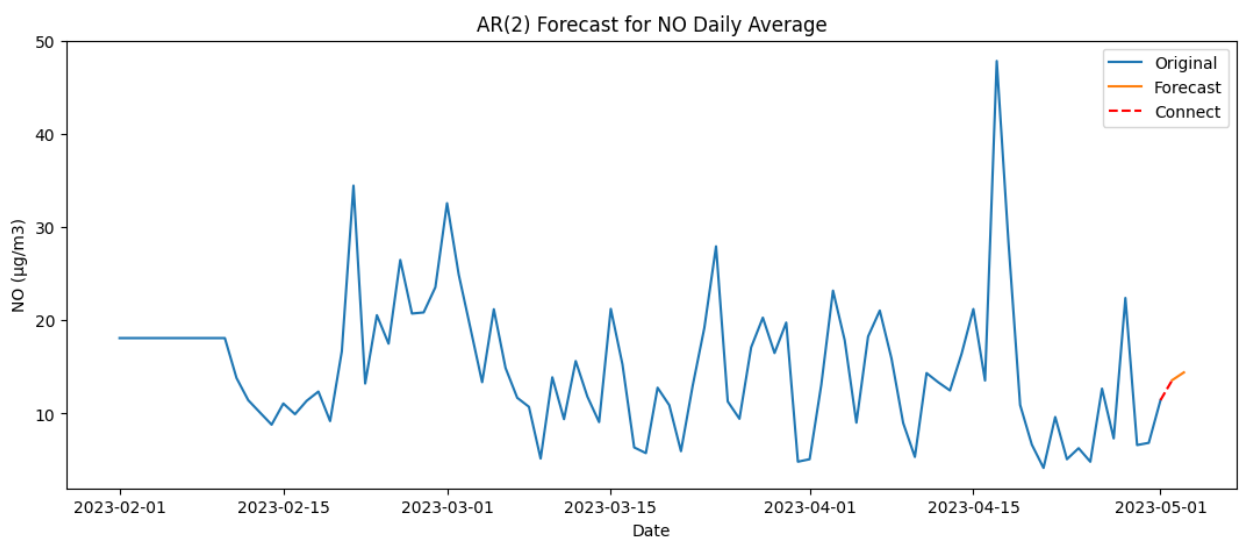
2023-05-03 *<u>44.407397</u>*

NO



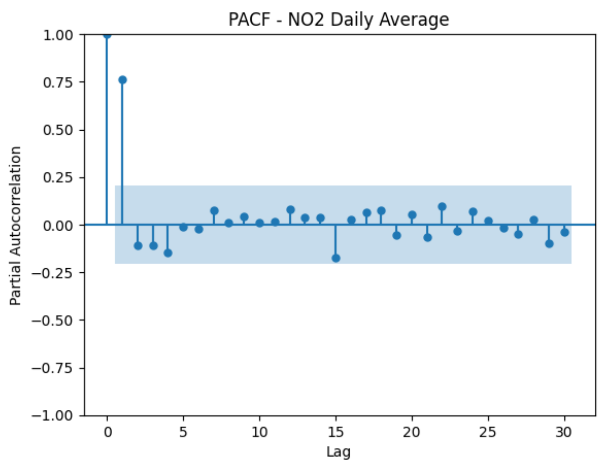


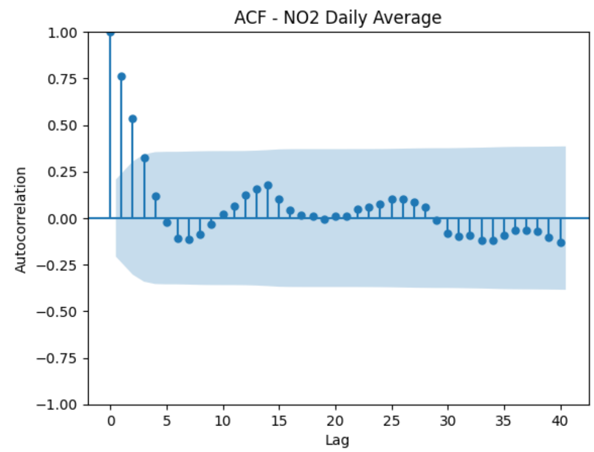
*Here, there are two significant spikes in PACF, therefore **AR (2)**.*



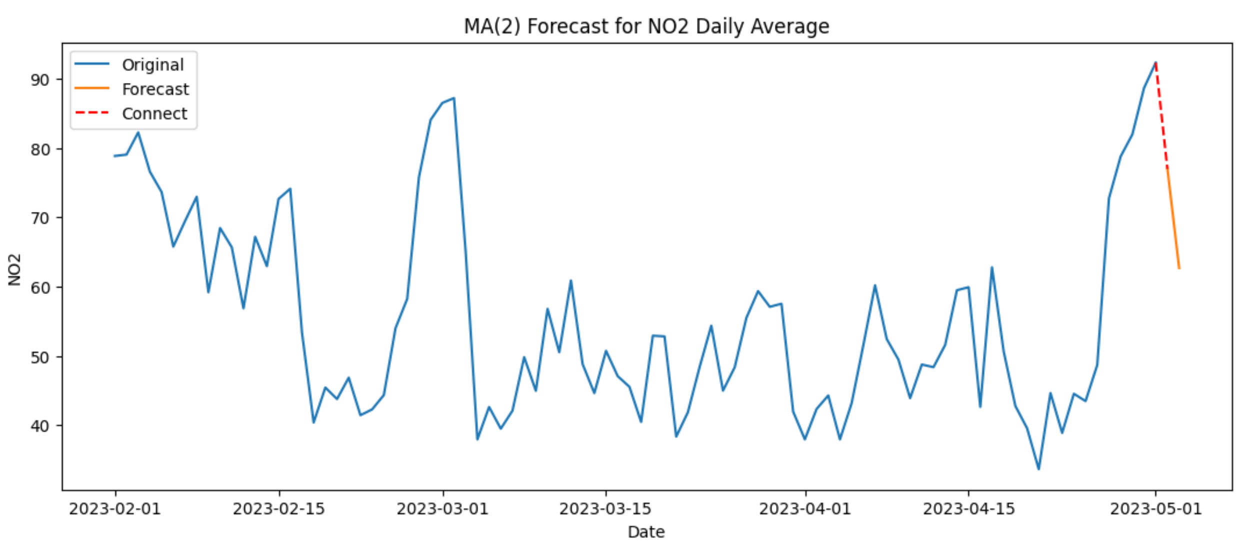
Forecasted data	*2023-05-02*	*13.583929*
	2023-05-03	*14.415644*

NO_2





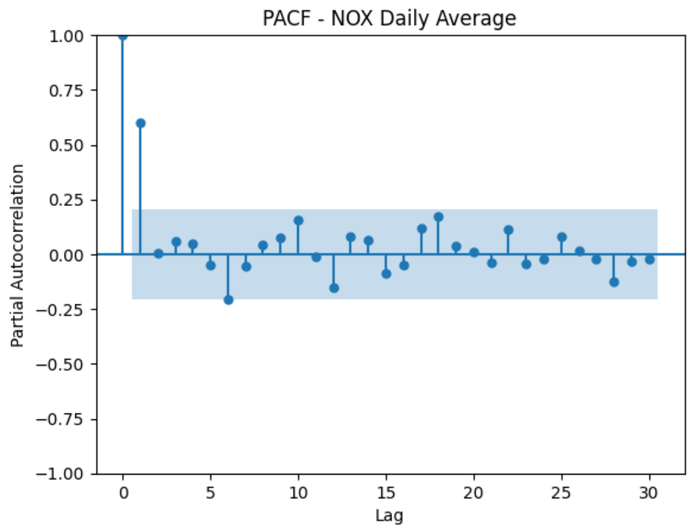
Here, instant decay in PACF and 2 significant spikes in ACF. So, ***MA (2).***

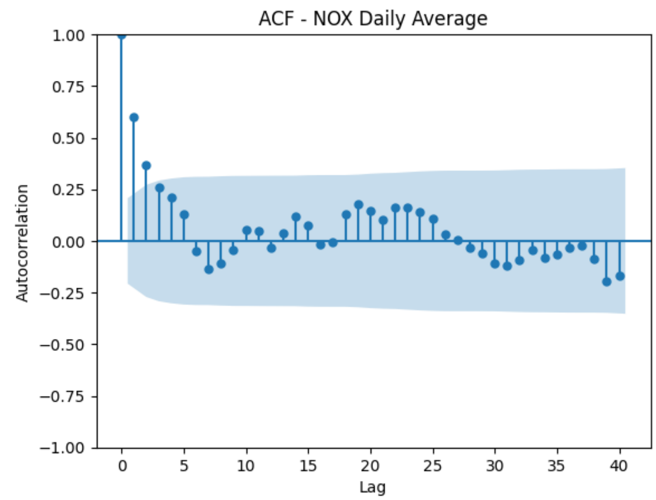


Forecasted data *2023-05-02* *76.478983*

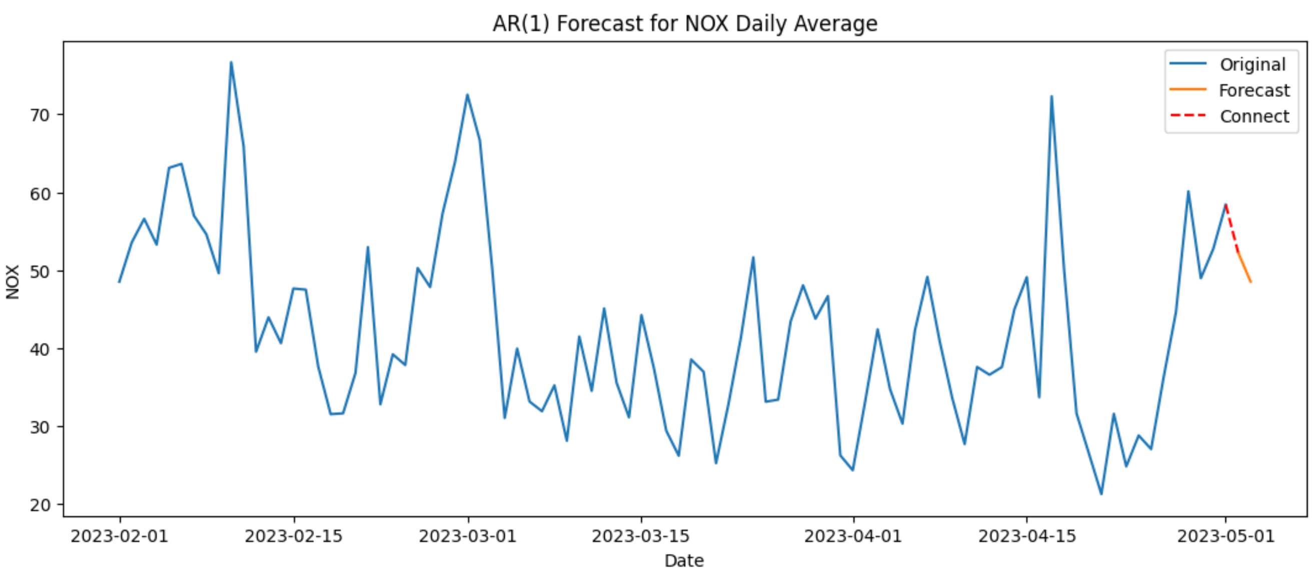
2023-05-03 *62.727091*

NOX



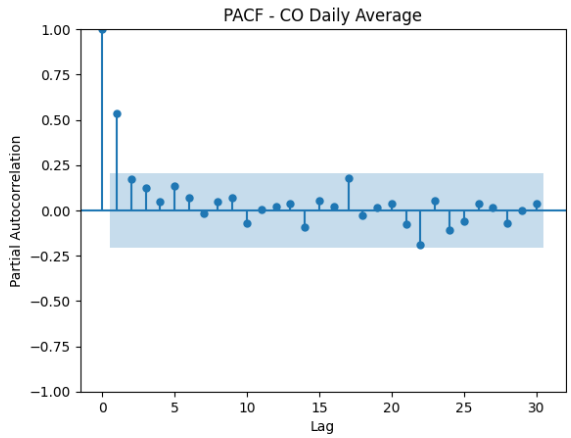


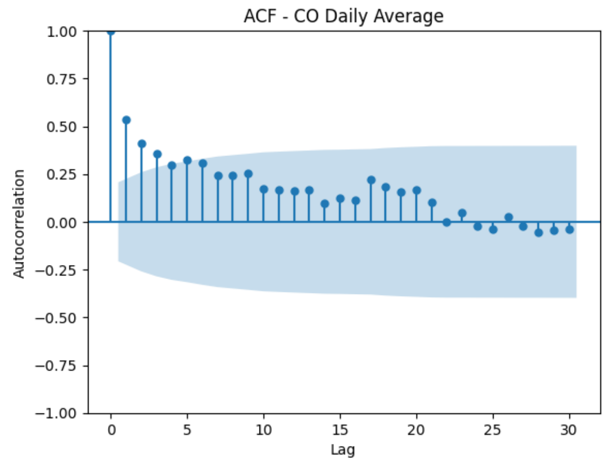
*As ACF suddenly goes to zero and significant spikes in PACF is one **AR (1)***



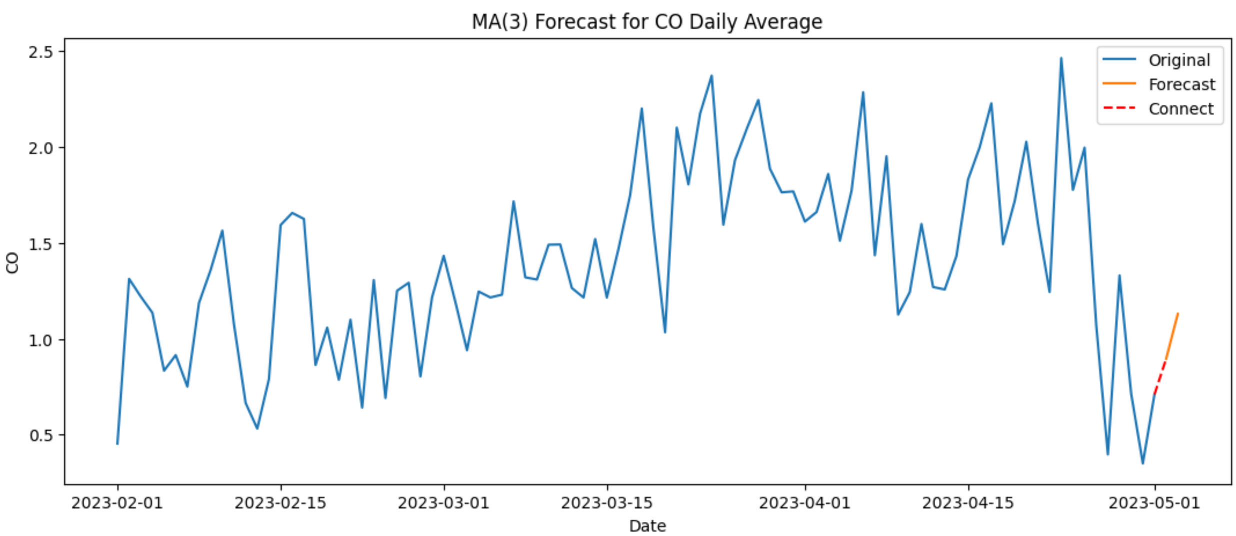
Forecasted data	*2023-05-02*	*52.272248*
	2023-05-03	*48.436304*

CO





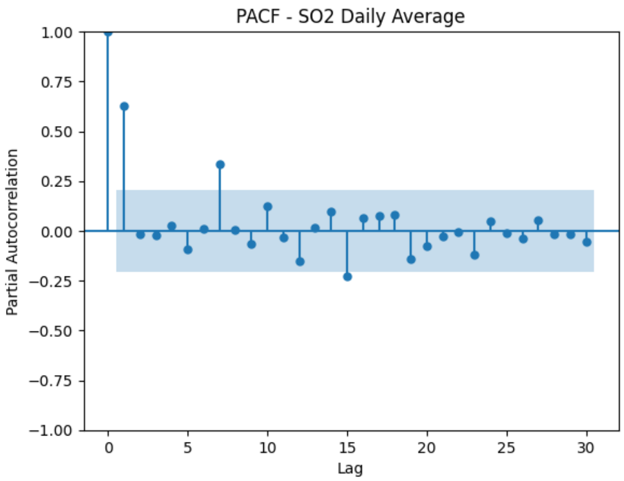
As PACF suddenly goes to zero and significant spikes in ACF are 3, ***MA (3)***

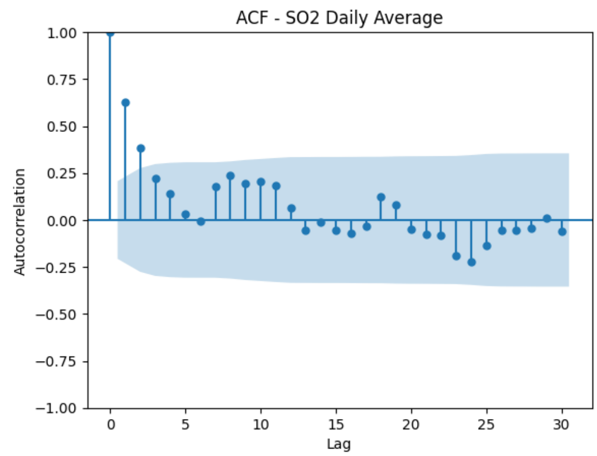


Forecasted data *2023-05-02* *0.895462*

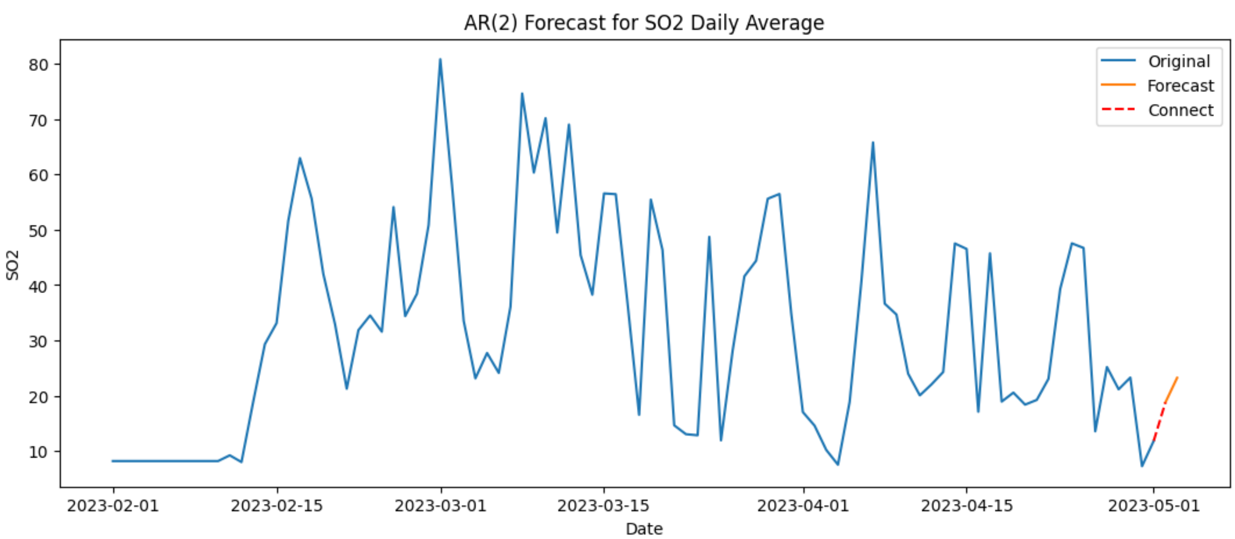
2023-05-03 *1.129399*

SO_2





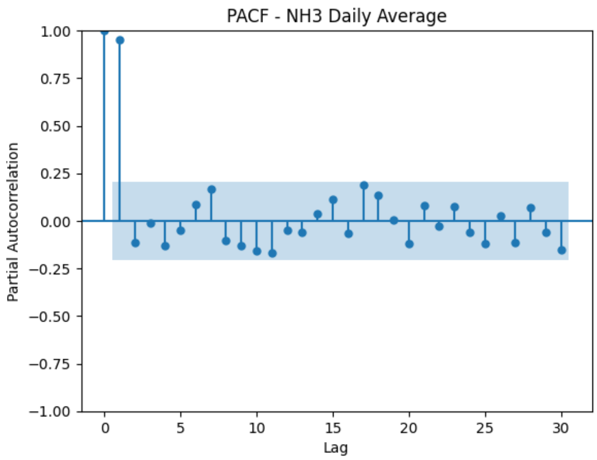
AR (2)

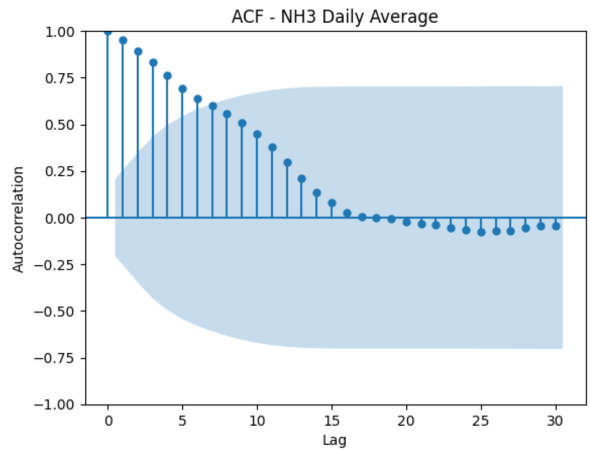


Forecasted data *2023-05-02* *18.802814*

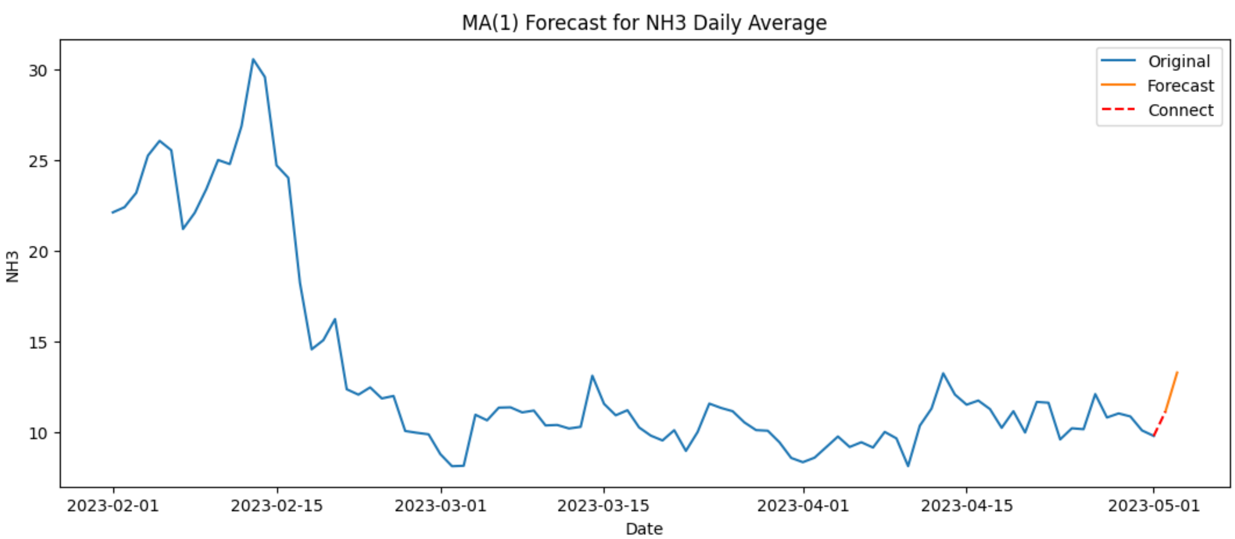
2023-05-03 *23.239386*

NH$_3$





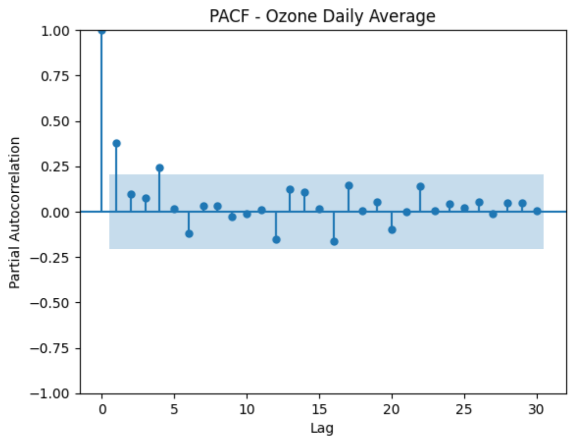
MA (1)

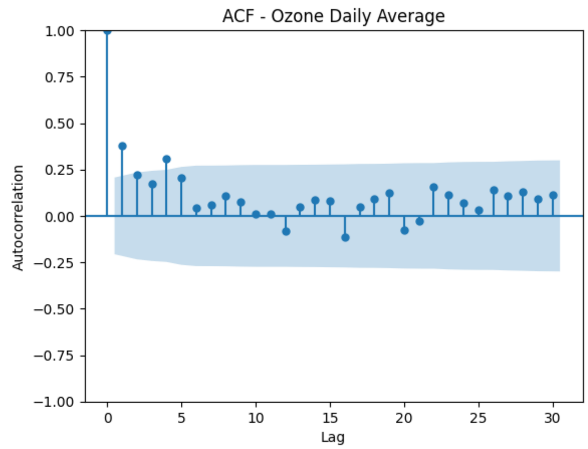


Forecasted data *2023-05-02* *11.142814*

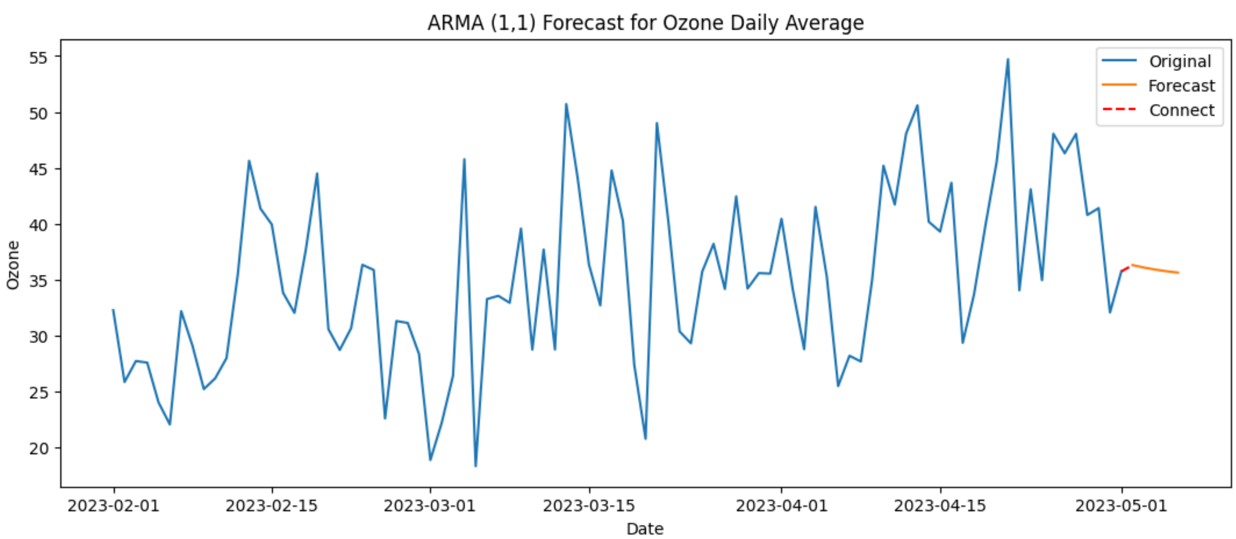
2023-05-03 *13.299386*

Ozone





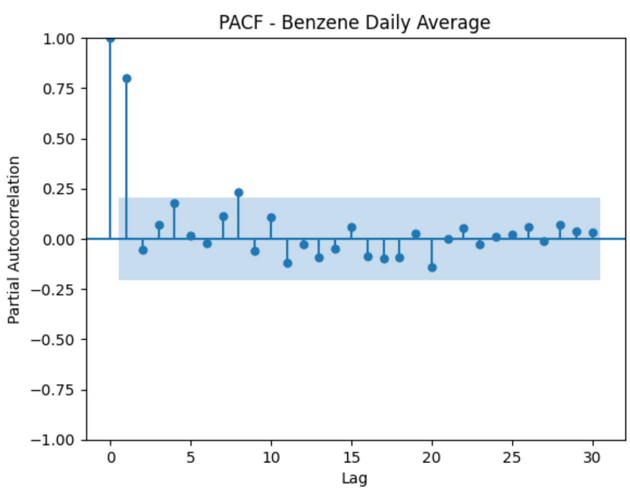
ARMA (1,1)

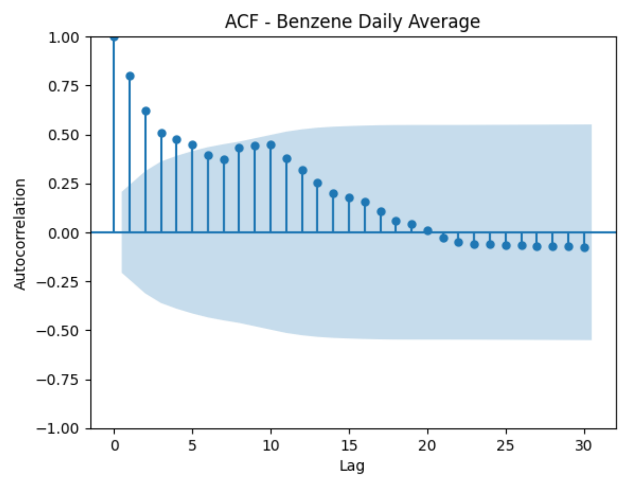


Forecasted data *2023-05-02* *36.309333*

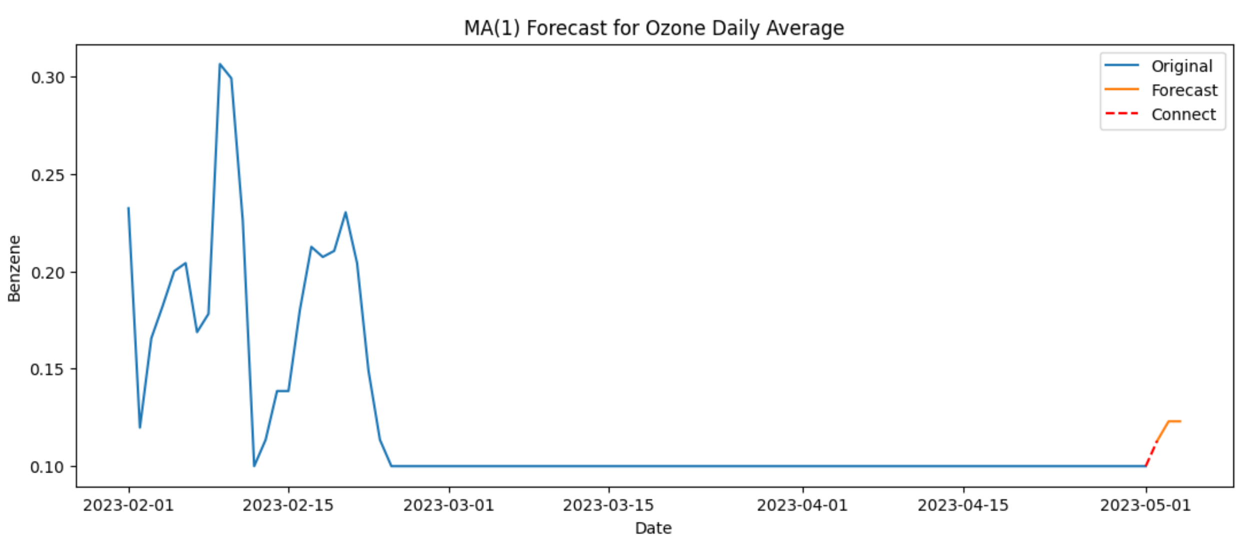
2023-05-03 *36.090294*

Benzene





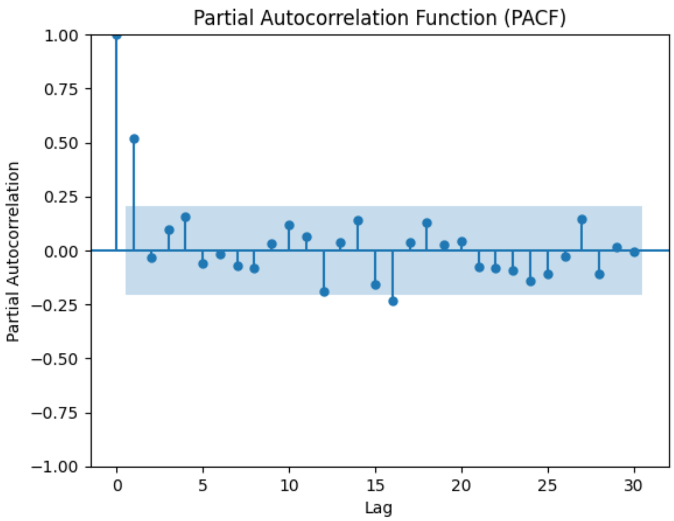
MA (1)

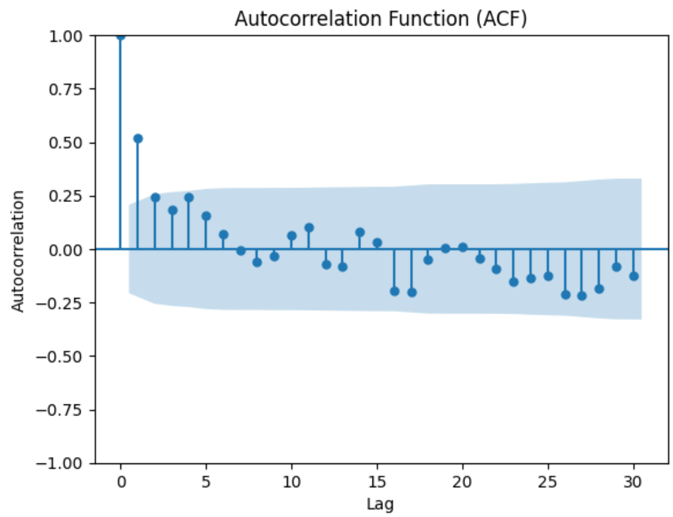


Forecasted data *2023-05-02* *0.11299*

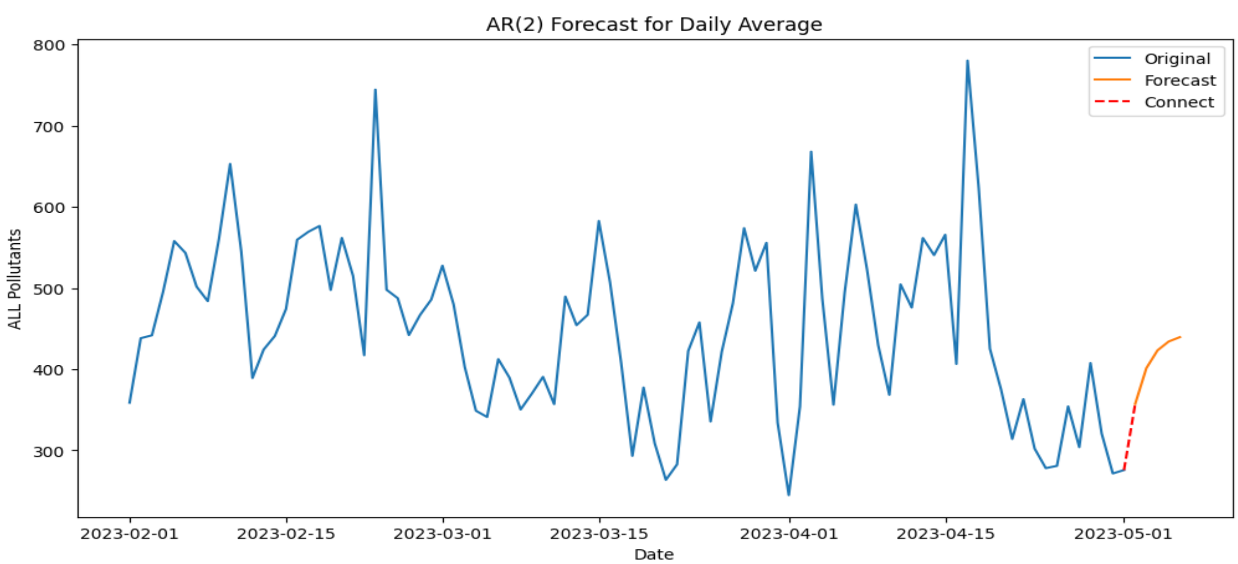
2023-05-03 *0.12299*

All Pollutants





AR (2)



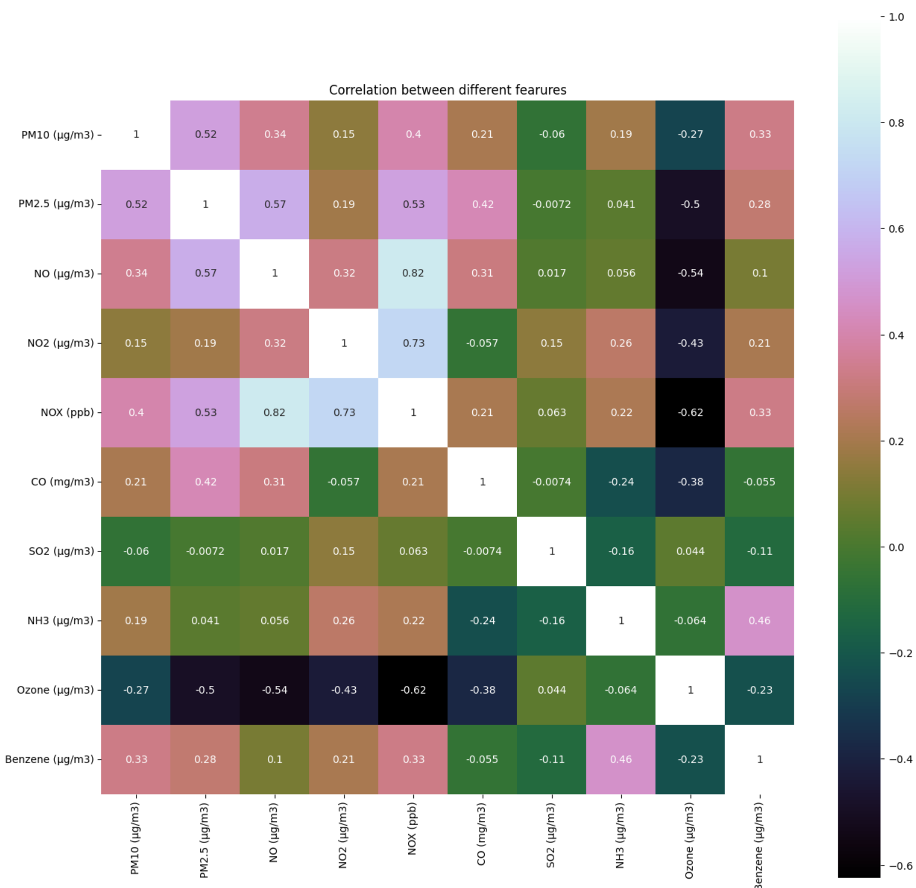
```
2023-05-02 00:00:00     39.772099
2023-05-02 00:15:00     47.128934
2023-05-02 00:30:00     54.092693
2023-05-02 00:45:00     60.684378
2023-05-02 01:00:00     66.923869
2023-05-02 01:15:00     72.829983
2023-05-02 01:30:00     78.420533
2023-05-02 01:45:00     83.712380
2023-05-02 02:00:00     88.721482
2023-05-02 02:15:00     93.462949
2023-05-02 02:30:00     97.951078
2023-05-02 02:45:00    102.199406
```

Forecasted data

Seasonal Variations: *Pollutant concentrations may exhibit seasonal patterns due to factors such as weather conditions, temperature changes, or human activities. The AR model may capture and reflect these patterns, resulting in an increase in pollutant concentrations during certain seasons.*

Lagged Effects*: The AR model considers the lagged values of the pollutant concentrations as predictors. If there are lagged effects or dependencies among the pollutant concentrations, an increase in the current concentration may be influenced by past values. The AR model can capture these dependencies and result in an increasing trend in the forecasted concentrations.*

Heatmap



A correlation coefficient of 0.82 between NO (nitric oxide) and NOX (nitrogen oxides) in a heatmap suggests a strong positive correlation between these two variables. A correlation coefficient of 0.82 indicates a high degree of linear association between the NO and NOX measurements. It means their concentration varies with a similar pattern. These gases are generated during the explosive reactions and combustion processes occurring during blasting. Most importantly both gases emitted together as the byproduct of explosive reactions. Hence there is so much correlation between NO and NOX.