# **EE798Q**- **Foundations of inferential** [****](https://github.com/jetharam171/EE798Q)**[**[**Code link**](https://github.com/jetharam171/EE798Q)**] statistics and automation** [FISA]

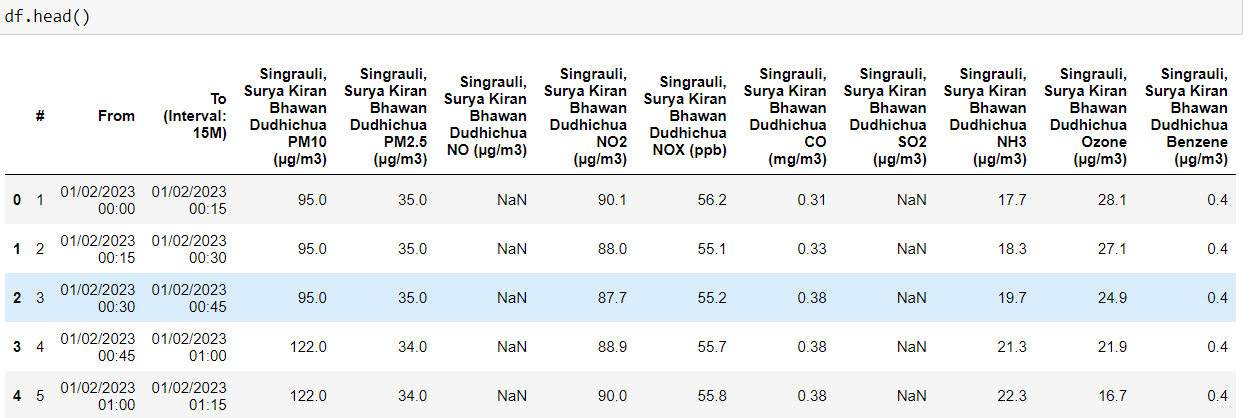
# **Assignment-1**

Name – **Jetha Ram**

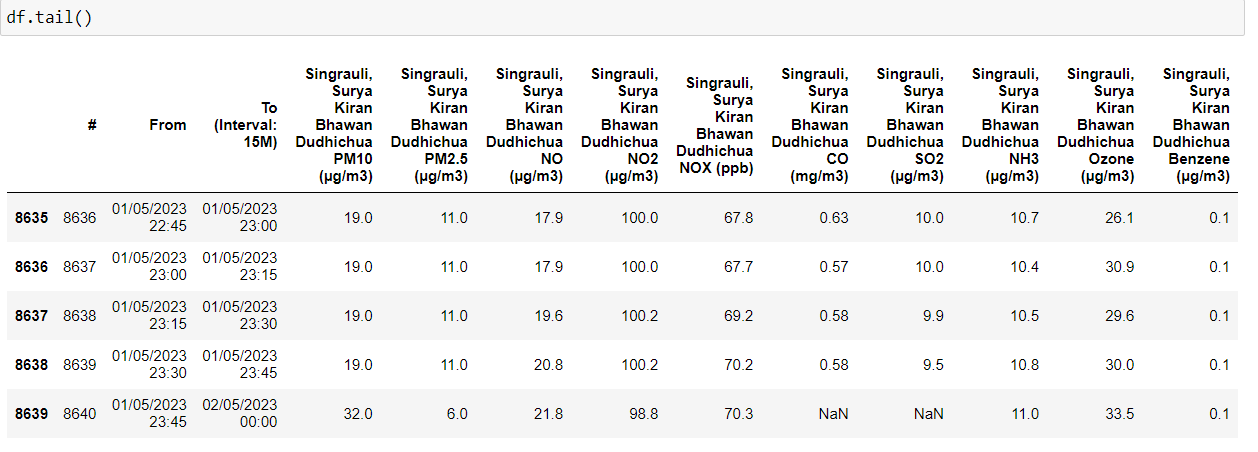
Roll No.- **210473**

Analysis of given data-

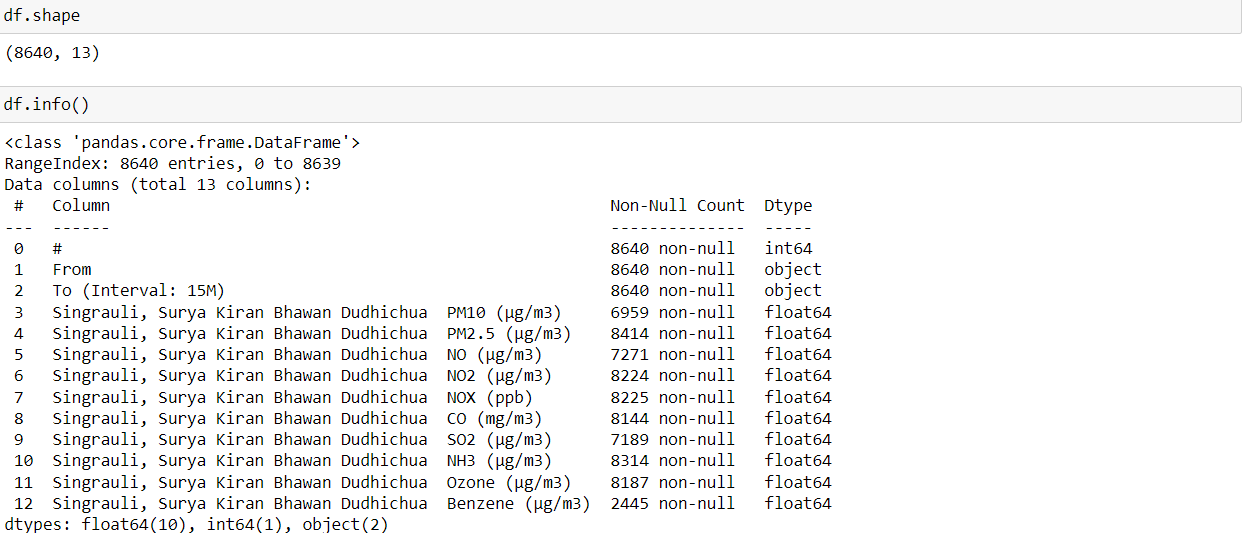
Head 5 element-df=Given data



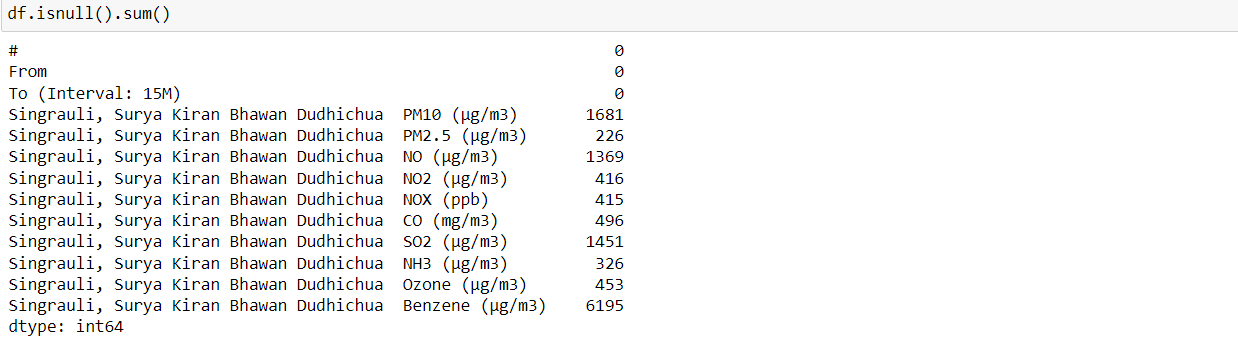
Tail 5 element-



Columns name and type

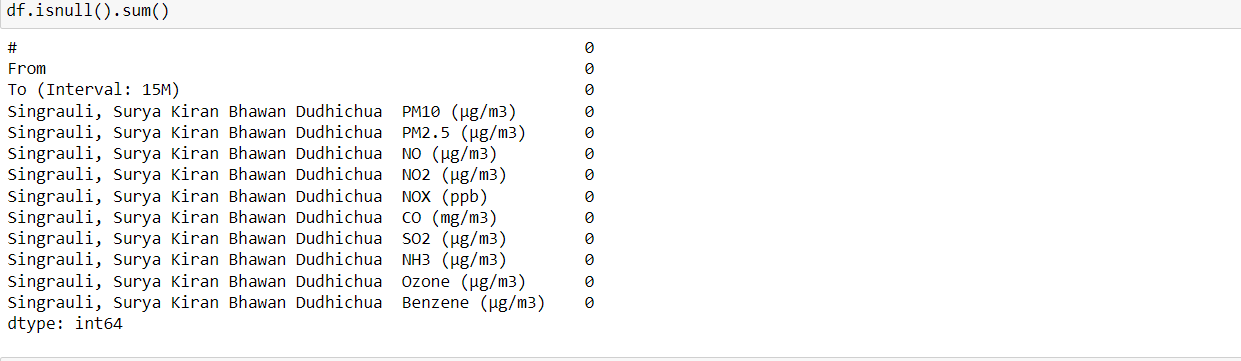


Now no. of empty columns here are



Here may be because of technical error in data collecting device there some data is missing in given file so for reduce our error we will fill that corresponding pollutant by Linear interpolation.

By fill up all space with Linear interpolation of all columns-

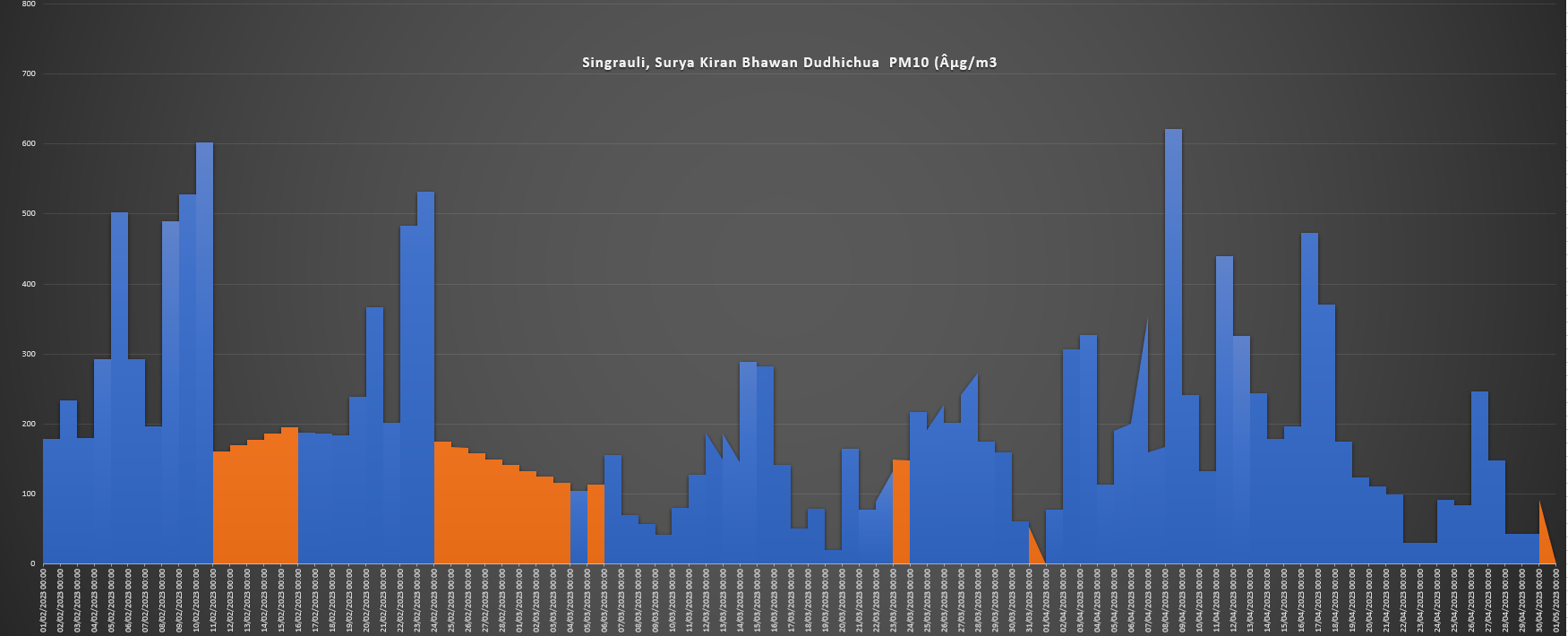


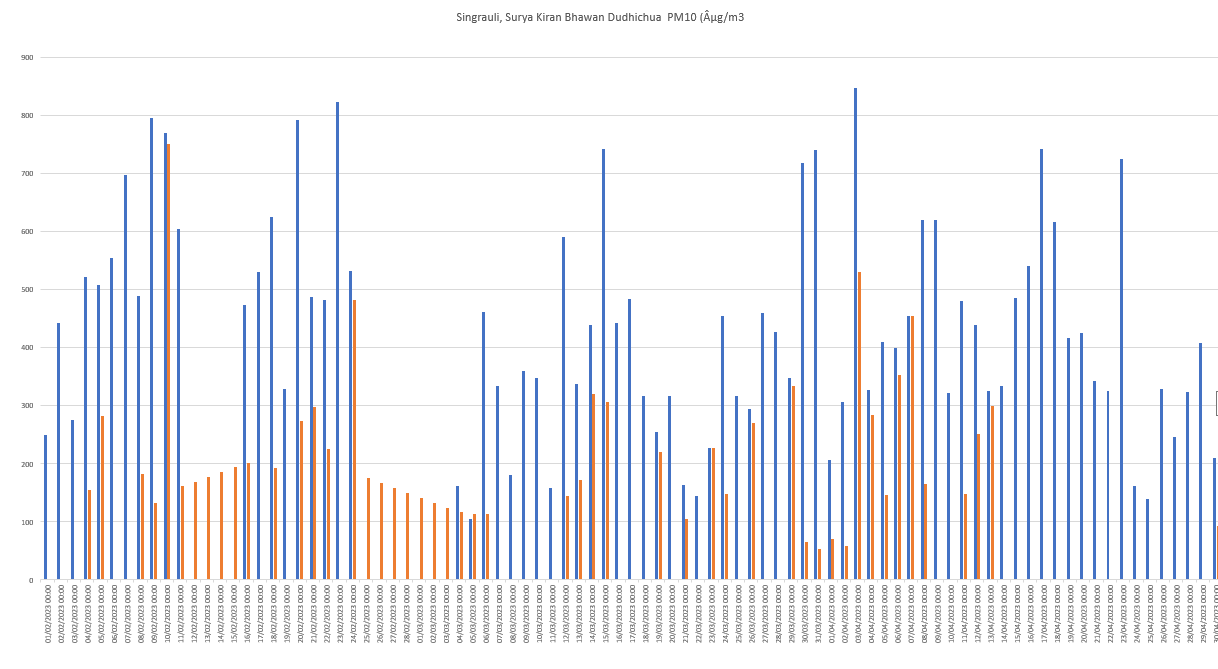
Now, Time series analysis of all pollutants:-

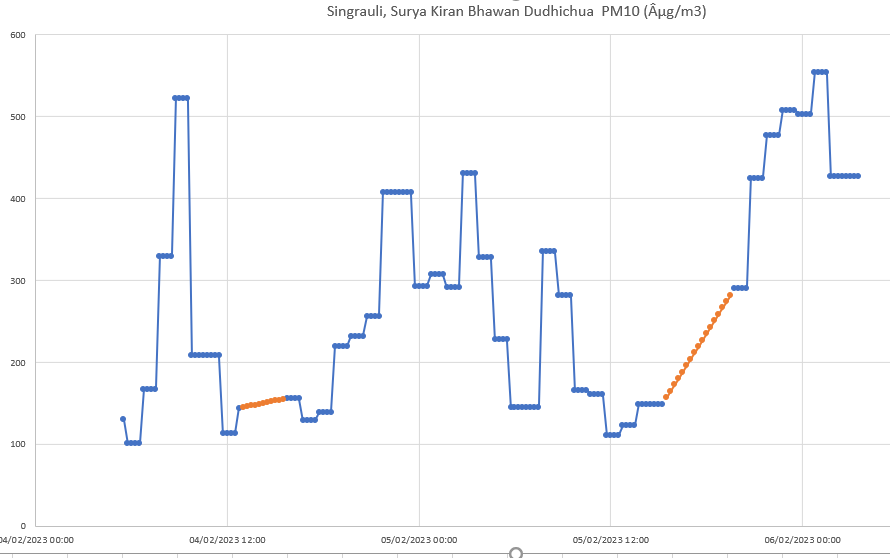
Legend-Blue- Given Value

Saffron-For Forcasted value by interpolation

1. PM10

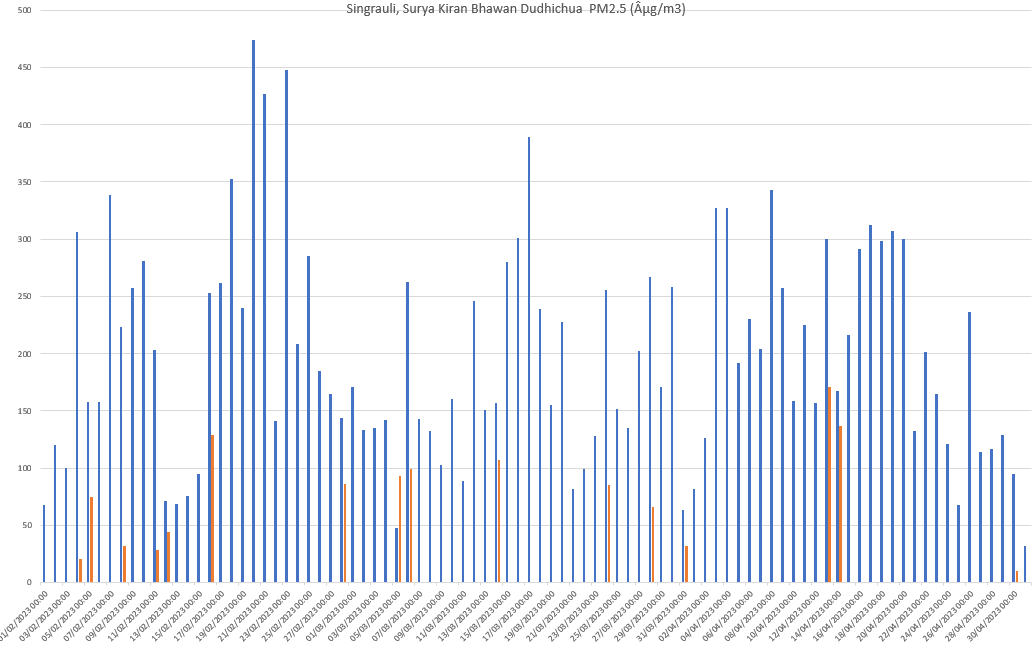


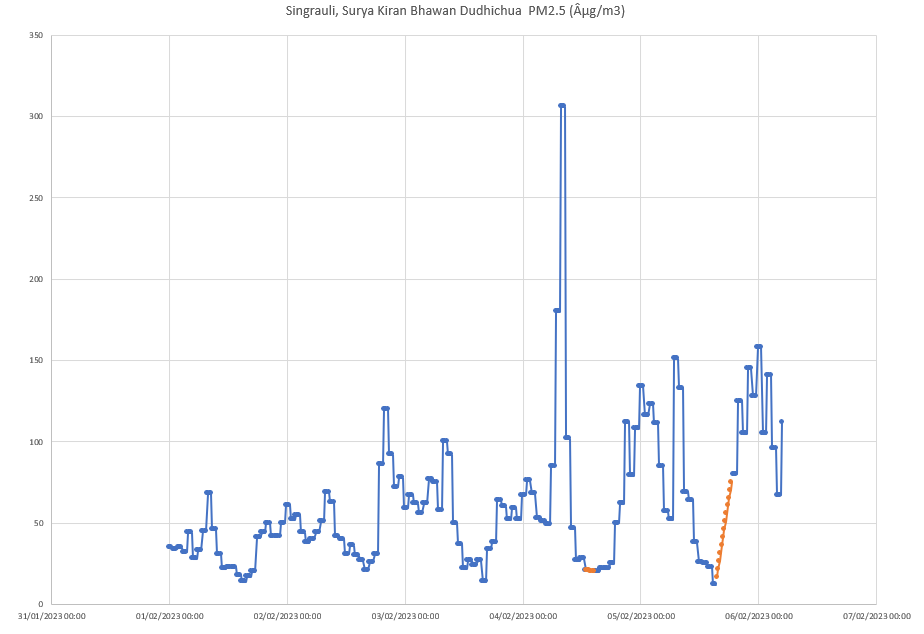




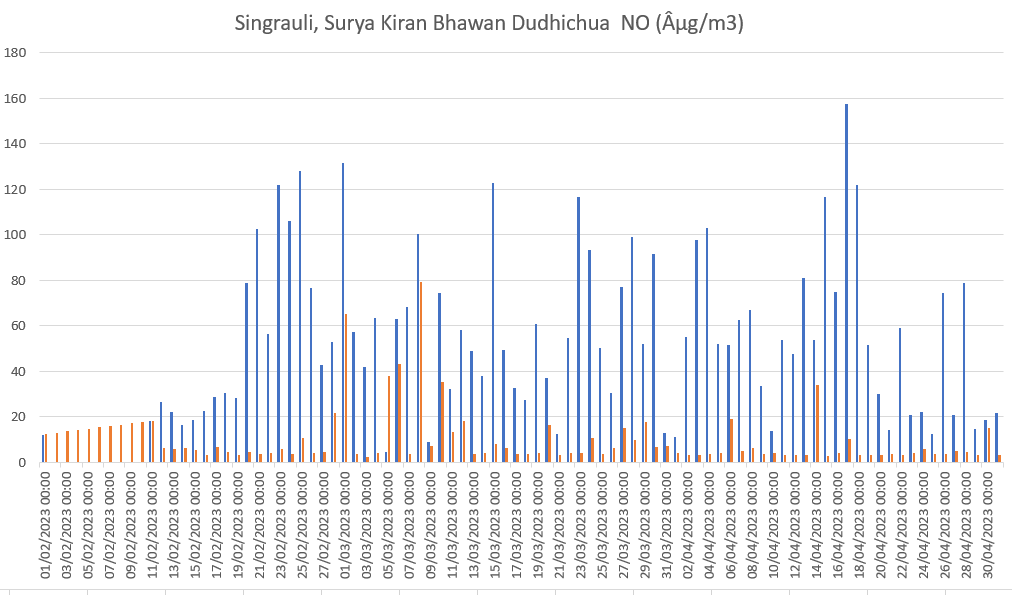
Time series analysis for 2 days with forecasted value for better view

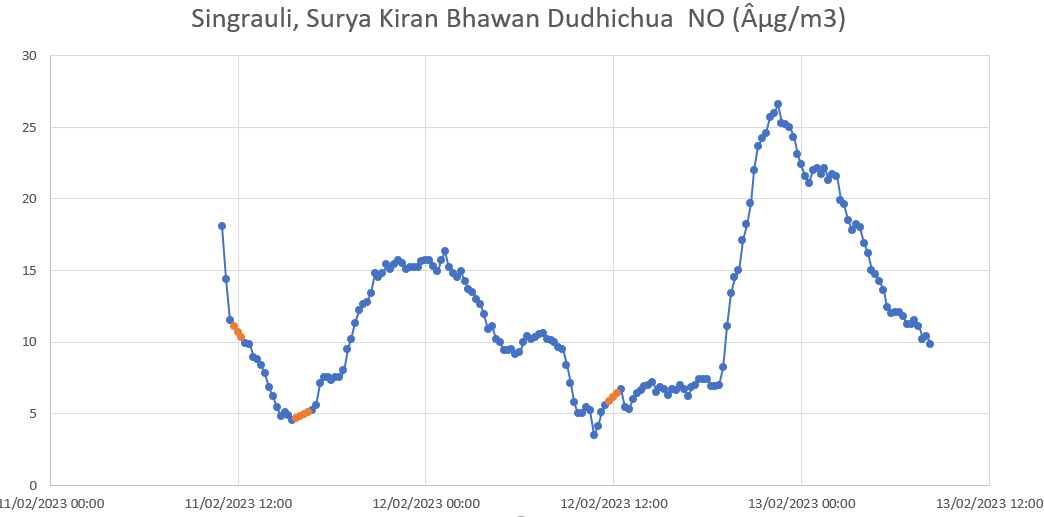
2. PM2.5



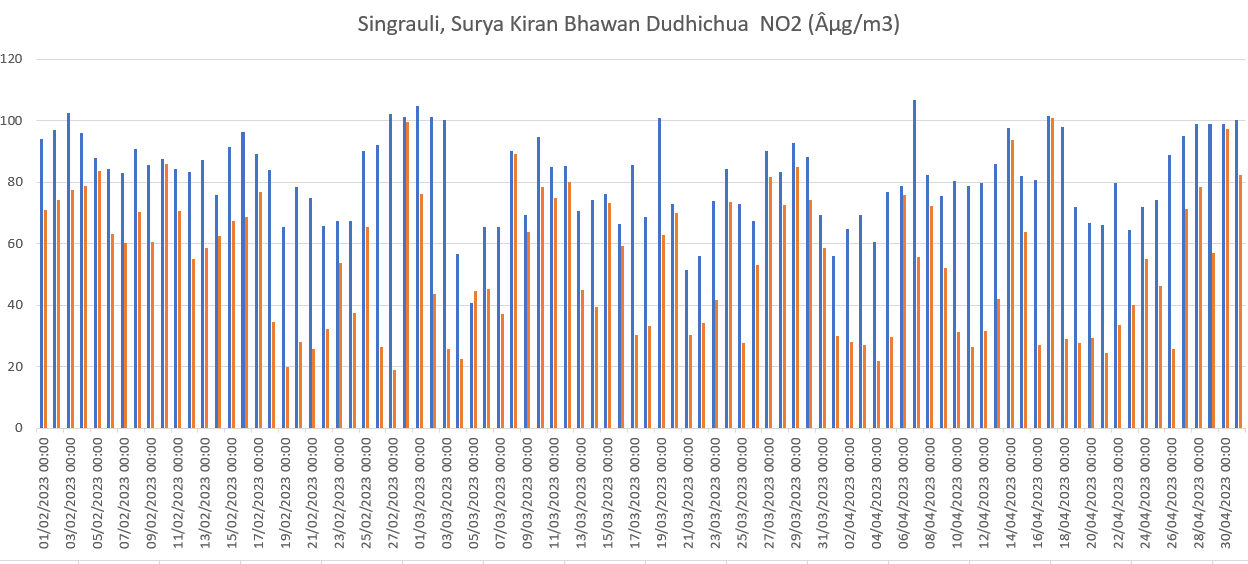


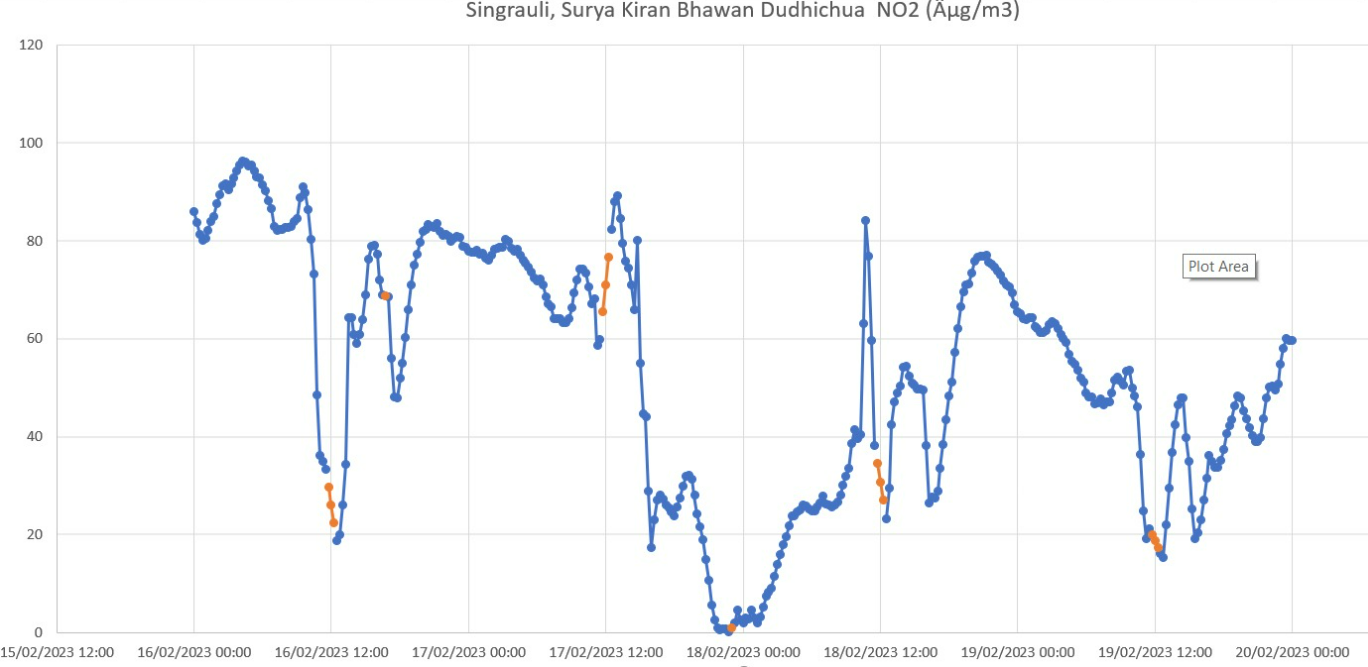
3. NO



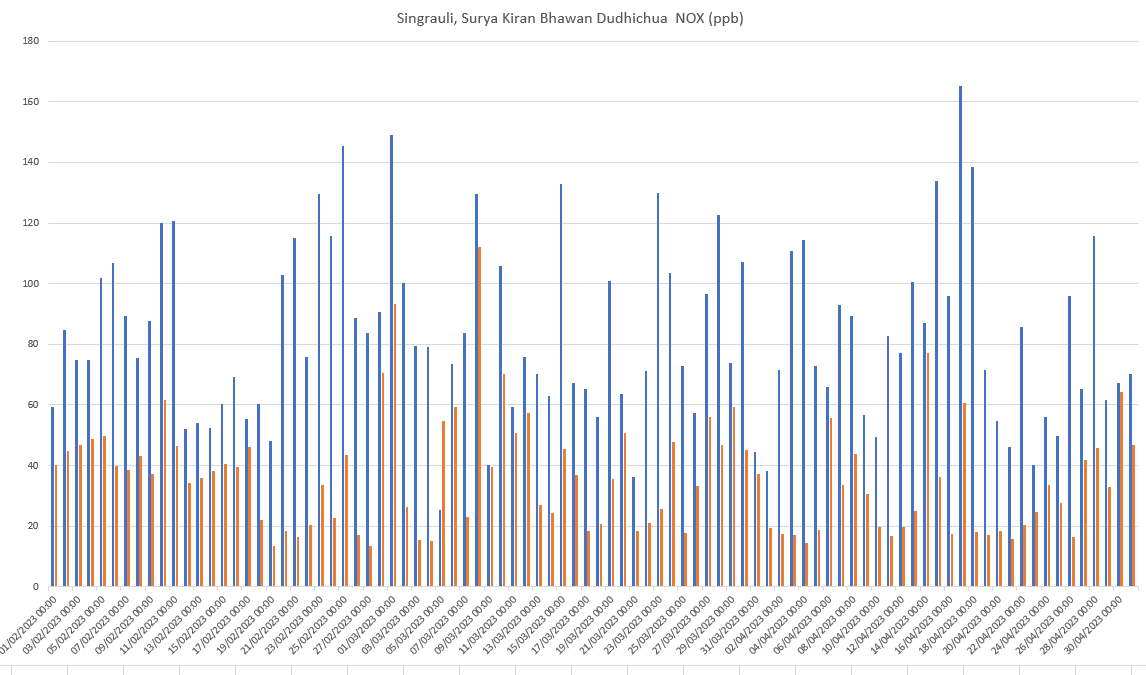


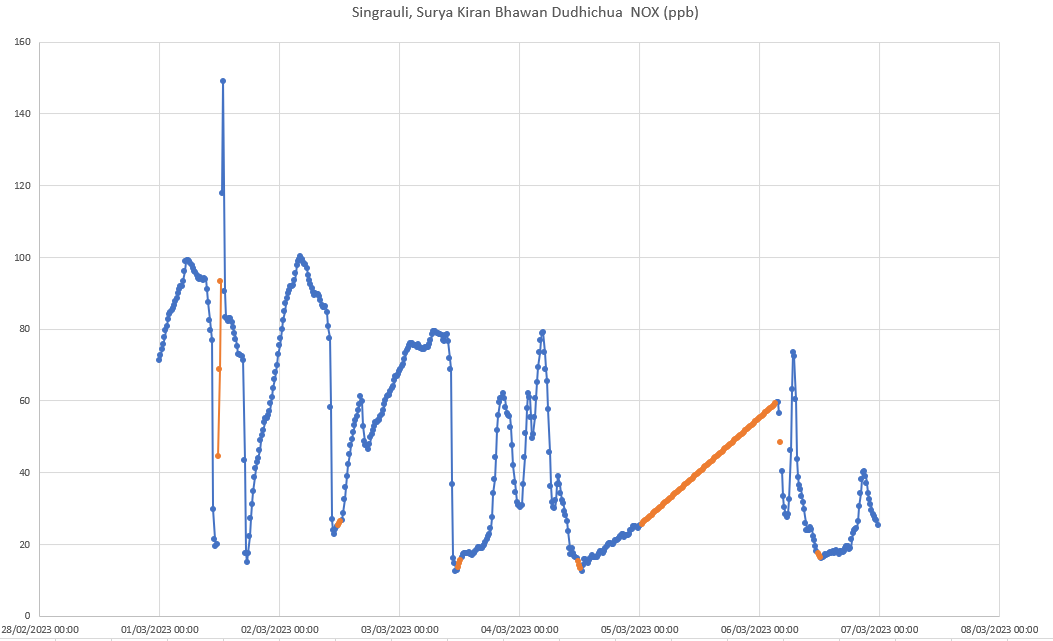
4. NO2



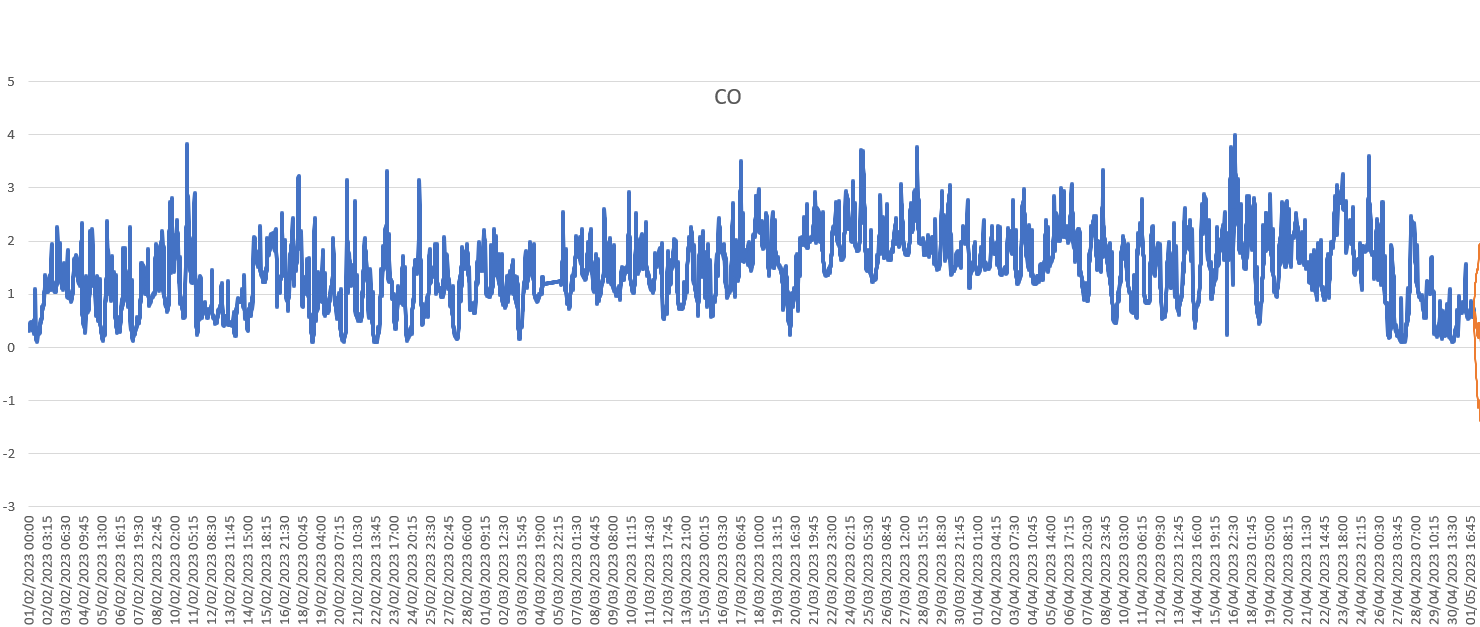


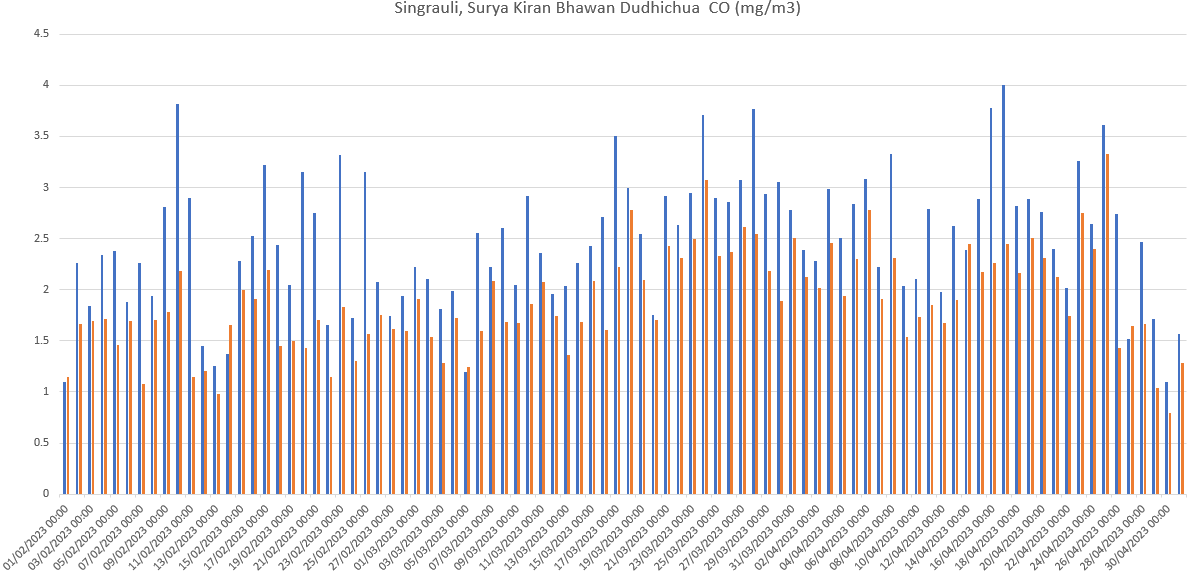
5. NOX

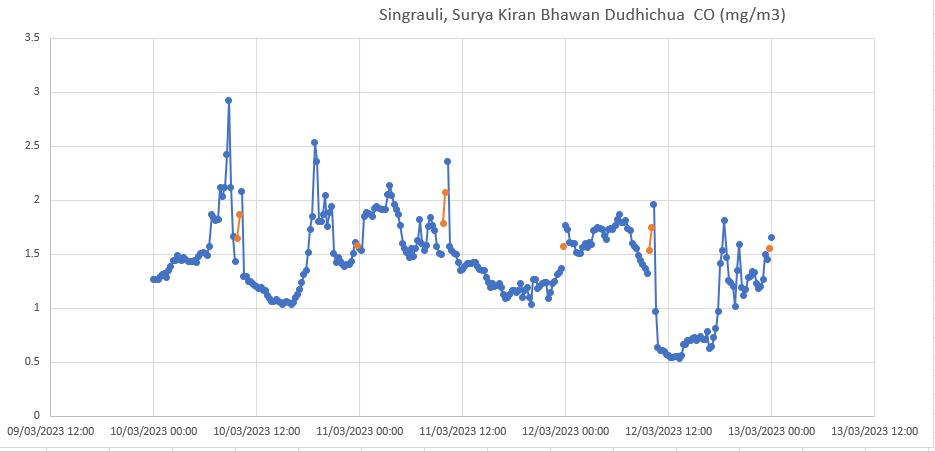




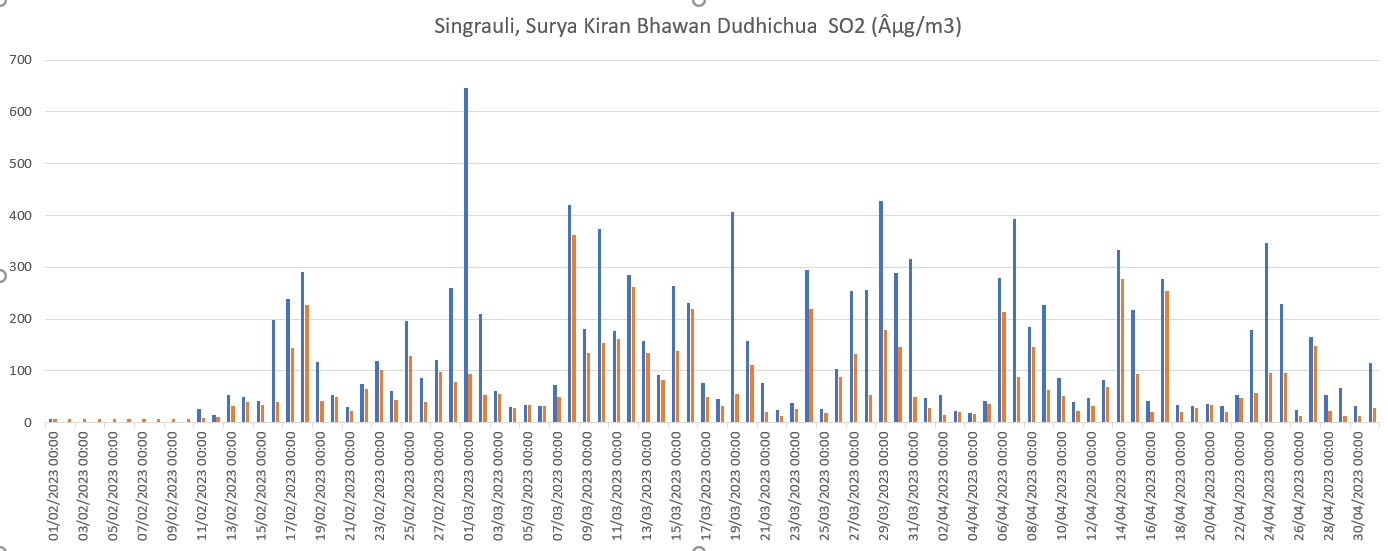
6. CO

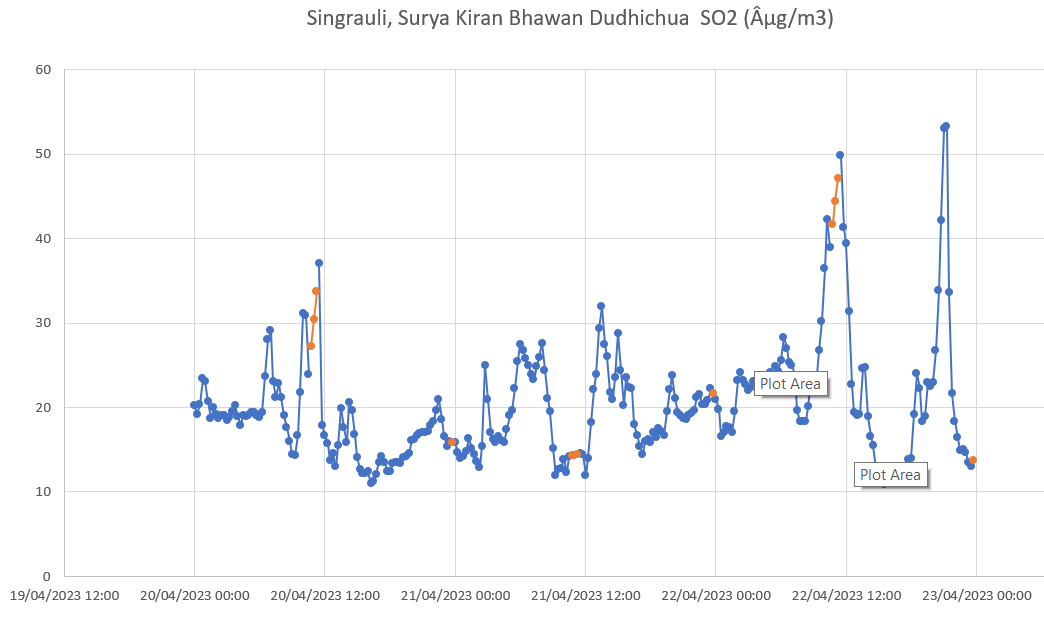




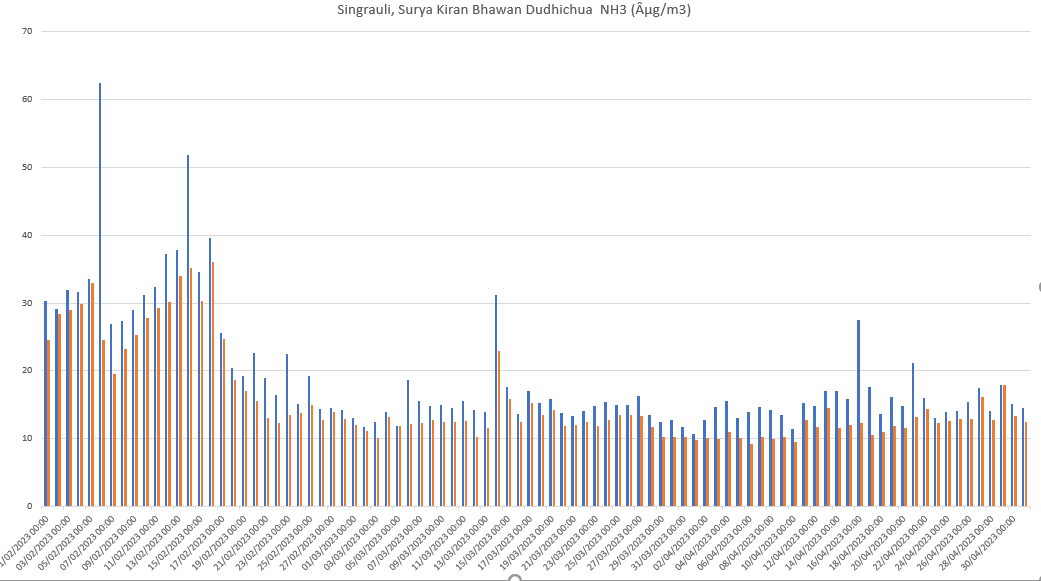


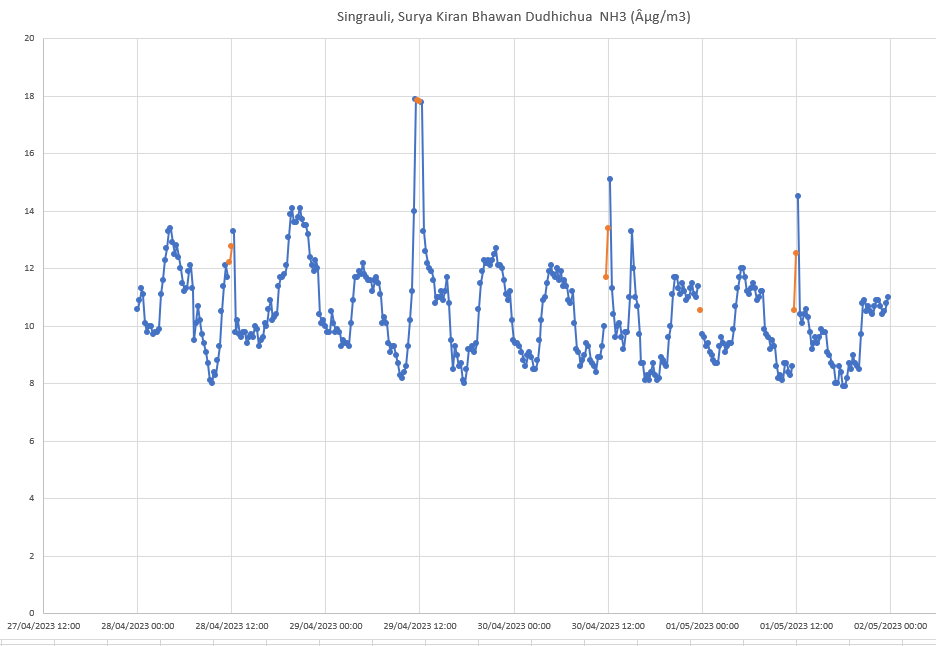
7. SO2



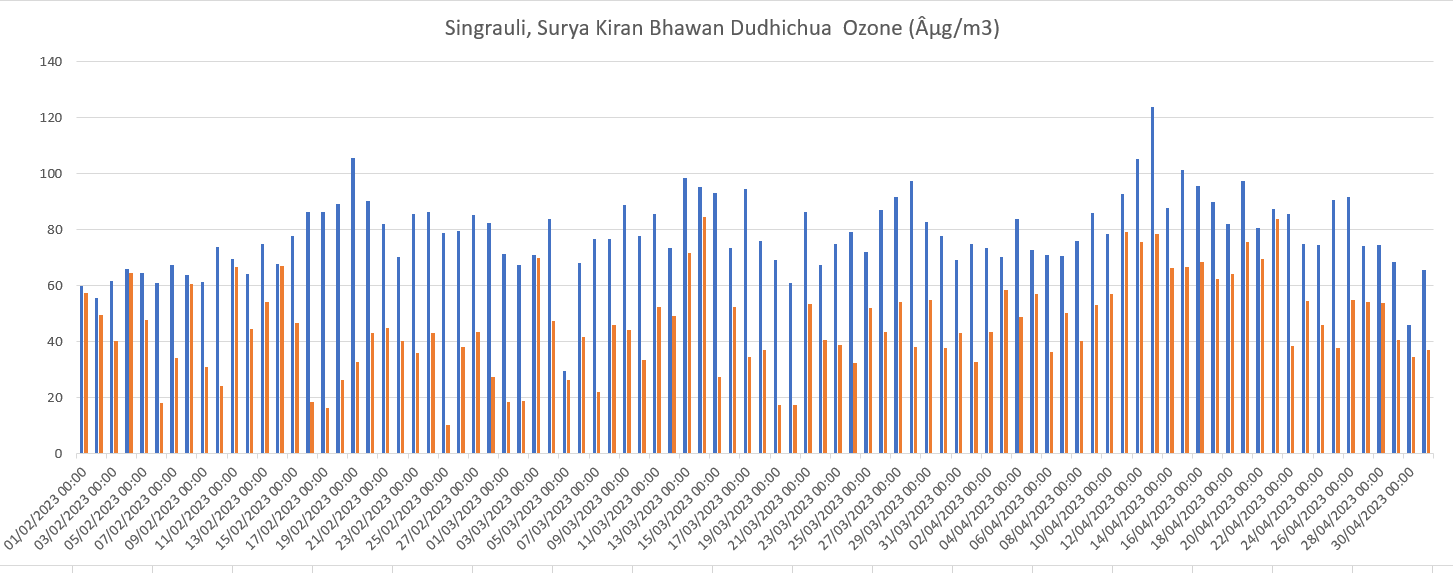


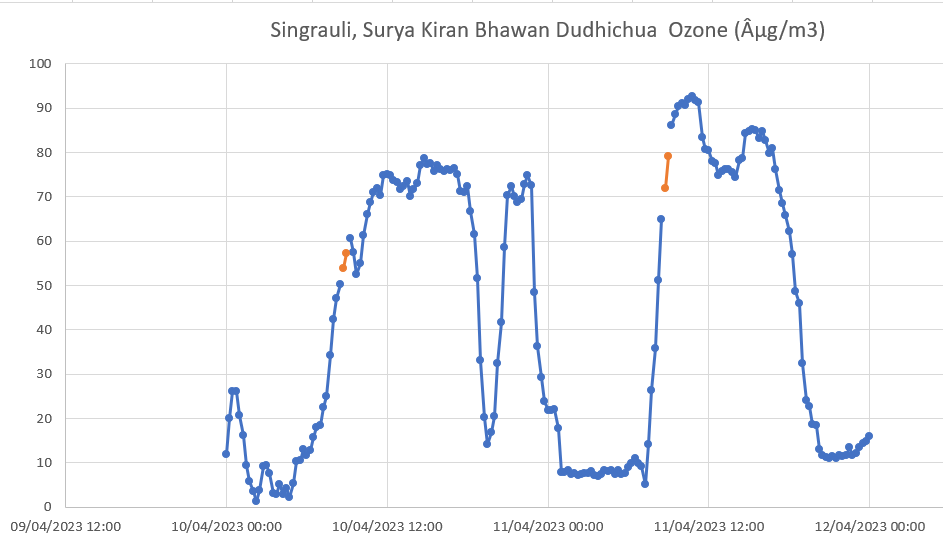
8. NH3



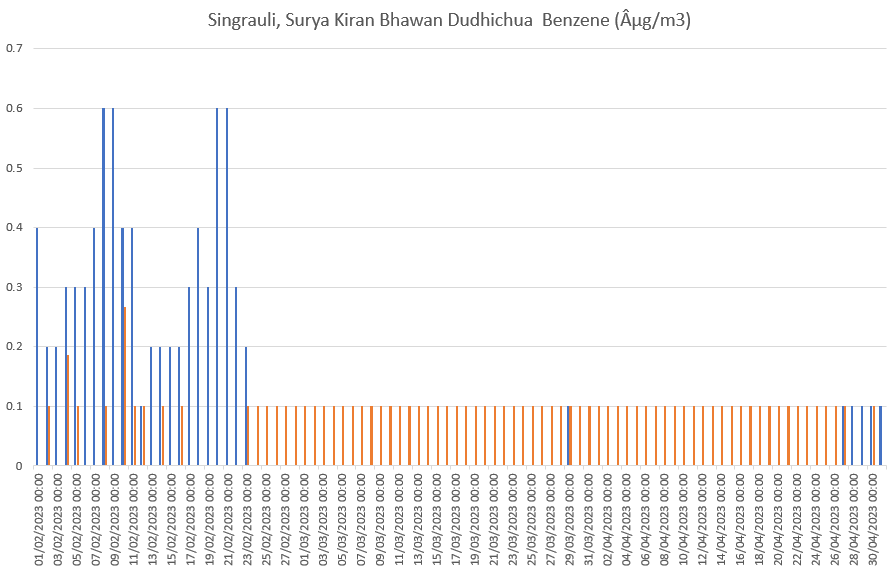


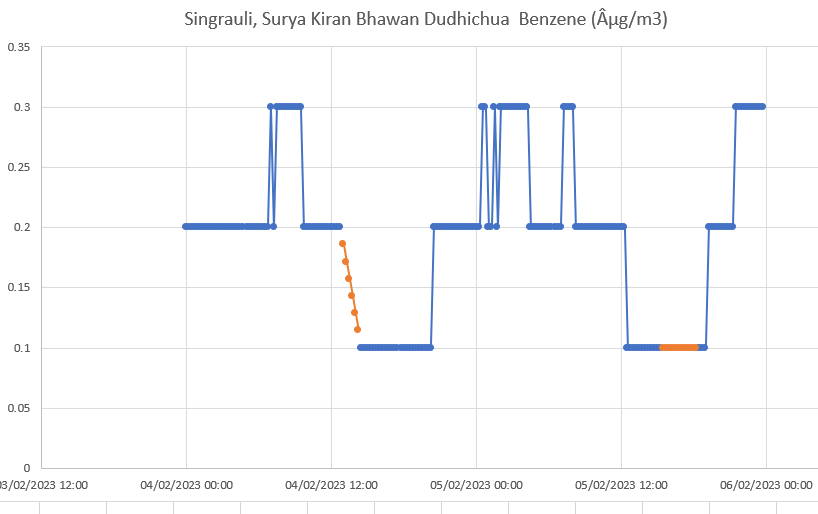
9. OZONE





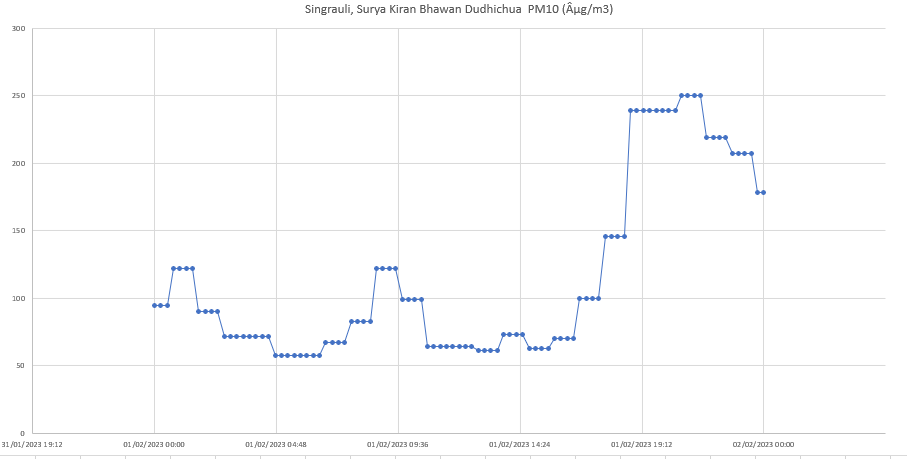
10. BENZENE

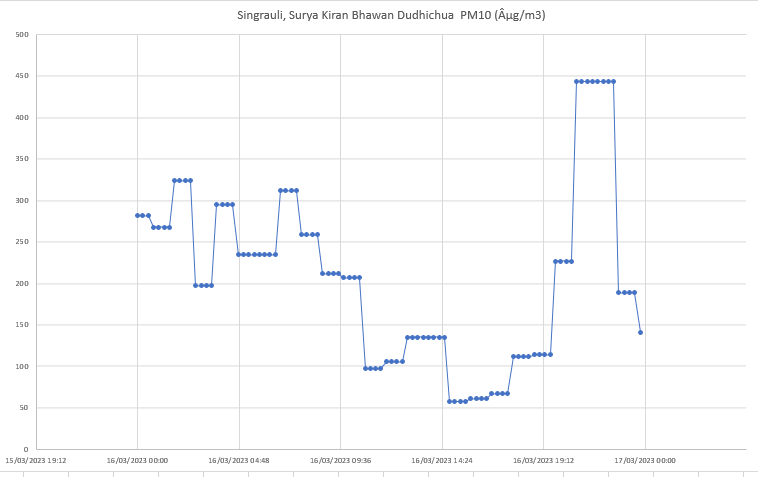


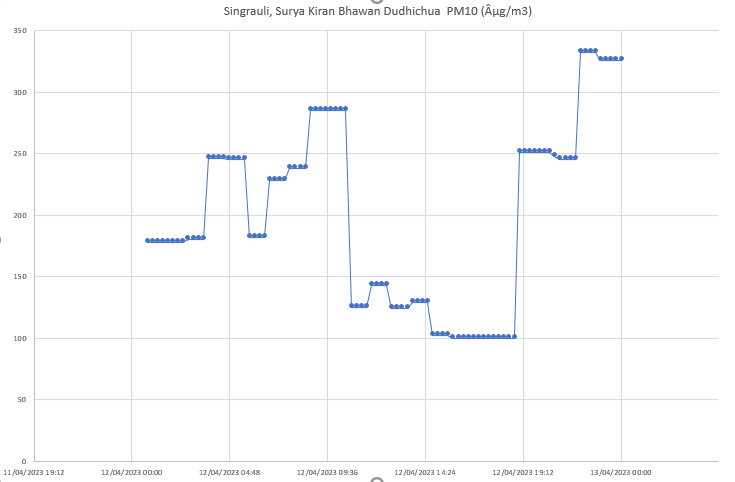


As we see here, we replace blank space replace by interploting value so continuity so graph remain, so this is better than 0.

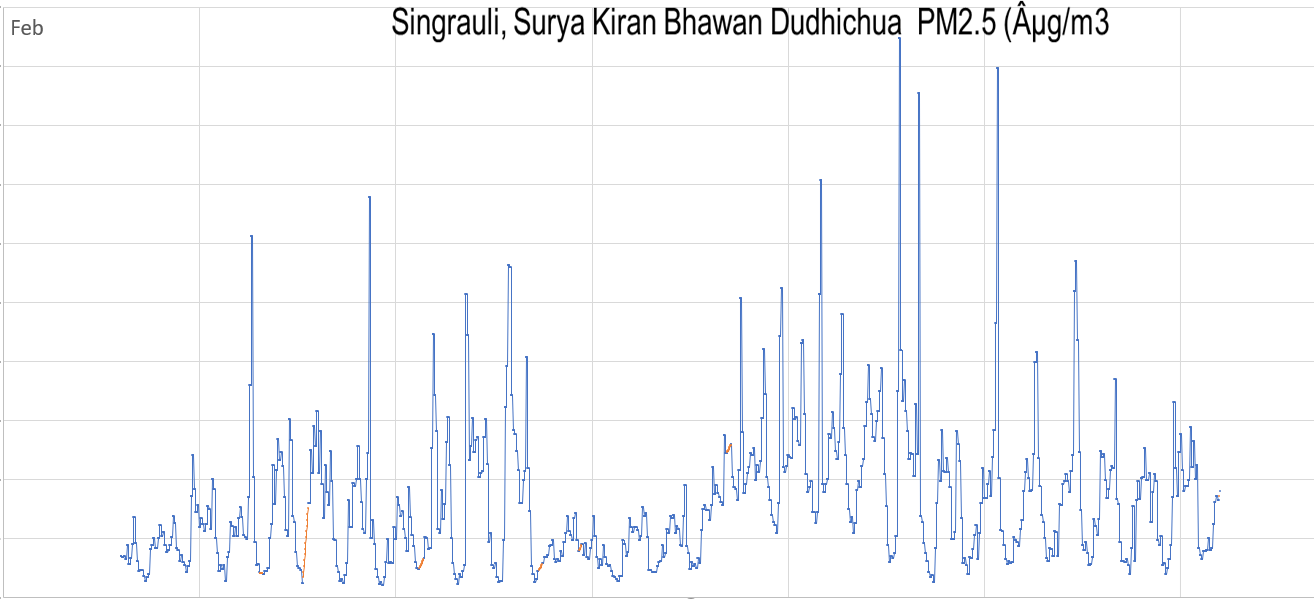
Now for claim blasting time in coal India is 13:45 to 14:45 we picked 3 random day and observed their air pollutant component for PM10

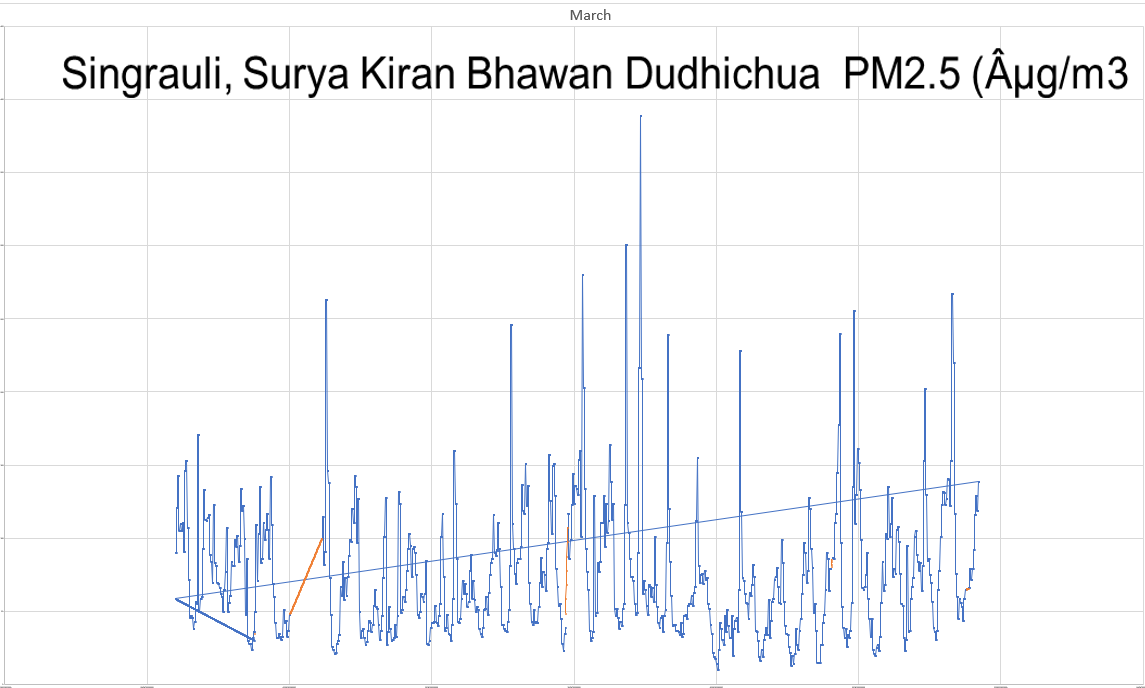


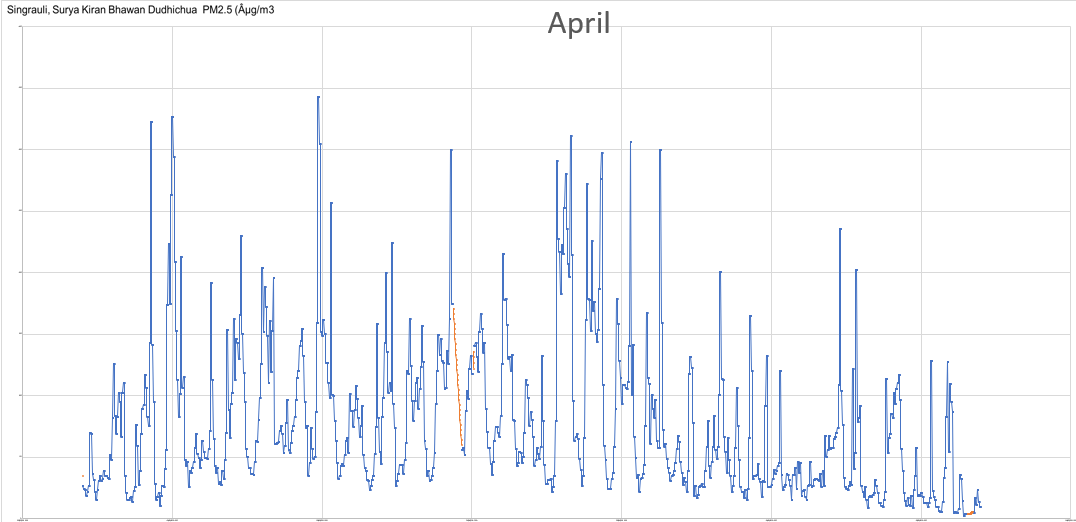




Here we can see after 14:30 there components of pollutant suddenly increase.







So, now we came to the Auto regressive model for forecasting

So here I initially plot data for initially 20 days for PM 2.5 pollutant as per given data

So now I am plotting data for 15 feb to 19 feb by using forecasting by observing all trend between 1 feb to 15 Feb.

Here we can clearly see from plot that by using Auto regressive model over forecasting is similar as per given data and all trend are also follow up in forecasted data

Similarly, we are observing last 15 days data of NH3.

Now as per given data following is plot

Now by using forecasting for last 5 days we observed following graph

Here also our forecasted data using past data trend/pattern is also same.

Now we came to point when our major part of data missing, and we want fill it up with interpolation and Auto Regressive moving average

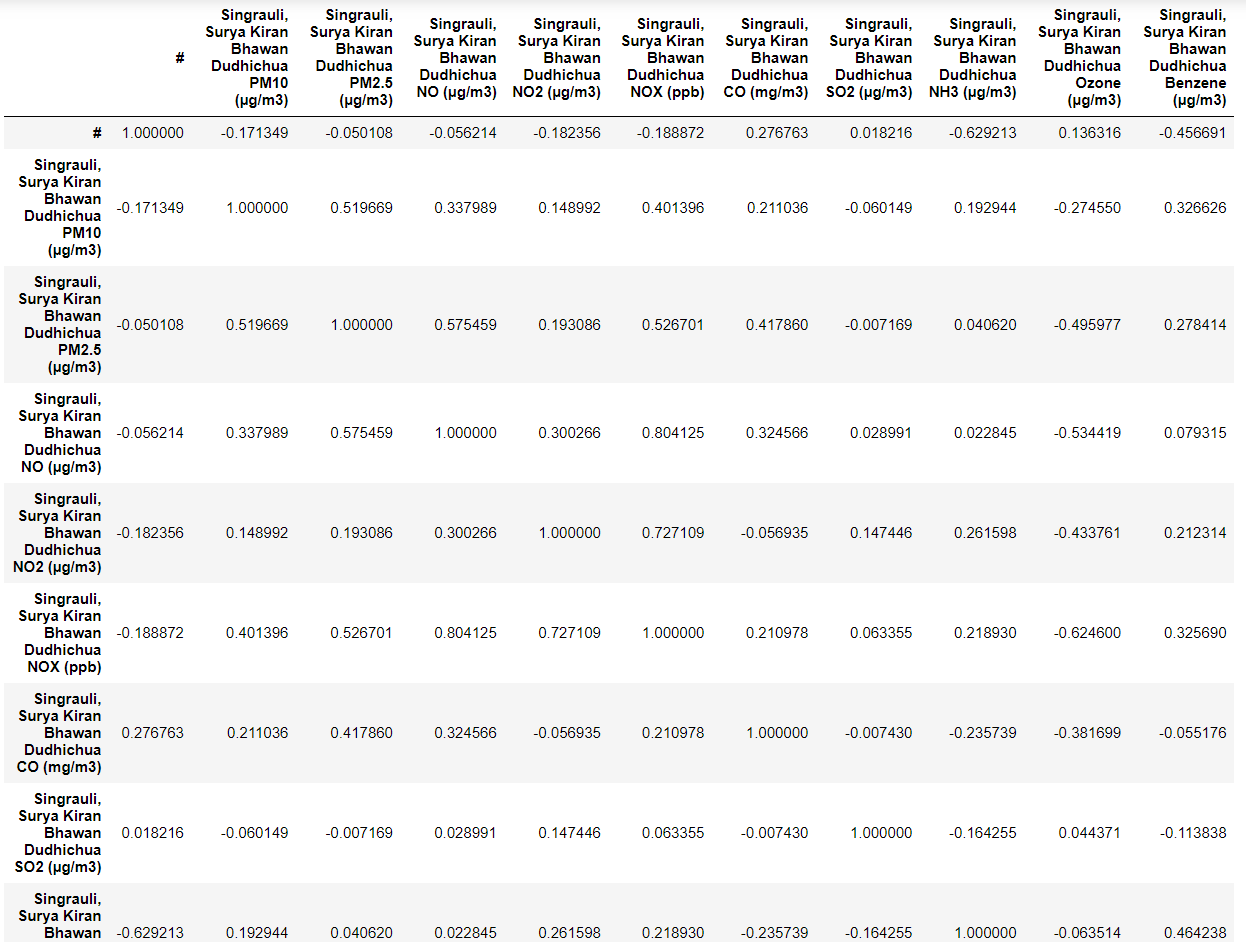
So, for example for pollutant PM 10 our data for 10 Feb to 15 Feb is missing so as per interpolation we got this plot.

and

This plot observed for part 10 Feb to 15 Feb by forecasted data using ARMA so we can clearly see that our trend of peaks and down followed very carefully.

So, we can say that ARMA is better approach for interpolation if there large part of data is missing otherwise for some value interpolation also a good approach.

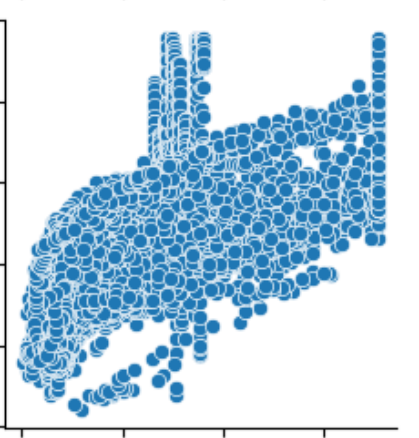
So, Now came to the point that is there any correlation between different pollutant components?



A picture containing text, screenshot, font, line

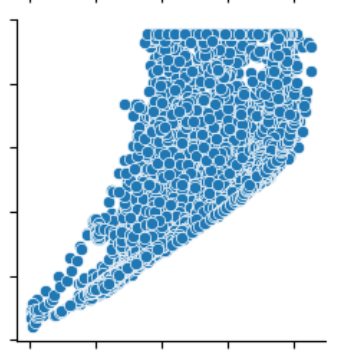
Description automatically generated

Here we can see NOX and NO correlate by 0.80 factor which is much significant and NOX and NO2 also correlate by 0.72 factor which is also easily justified as NOX and NO and NO2 have same component Nitrogen so they must be correlate to each other.



This correlated graph is for NOX on y-axis and NO on x-axis.

And

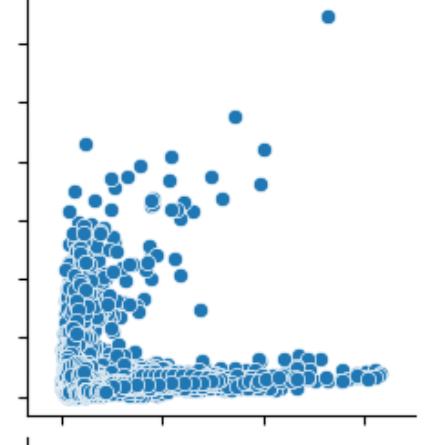


This correlated graph is for NOX on y-axis and NO2 on x-axis.

These graphs showing how increase in value also increase in other components value.

And we can see NO and SO2 correlate by 0.02 factor which is very low.

And here we can conclude that if correlation factor high that mean they are somehow dependent on each other otherwise they are independent

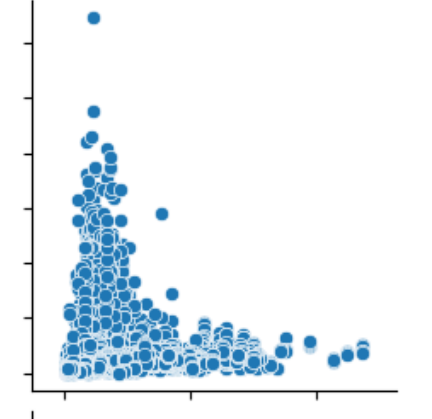


This correlated graph is for SO2(y-axis) and NO(x-axis) .

And we can clearly see for large value of SO2 there is little NO.

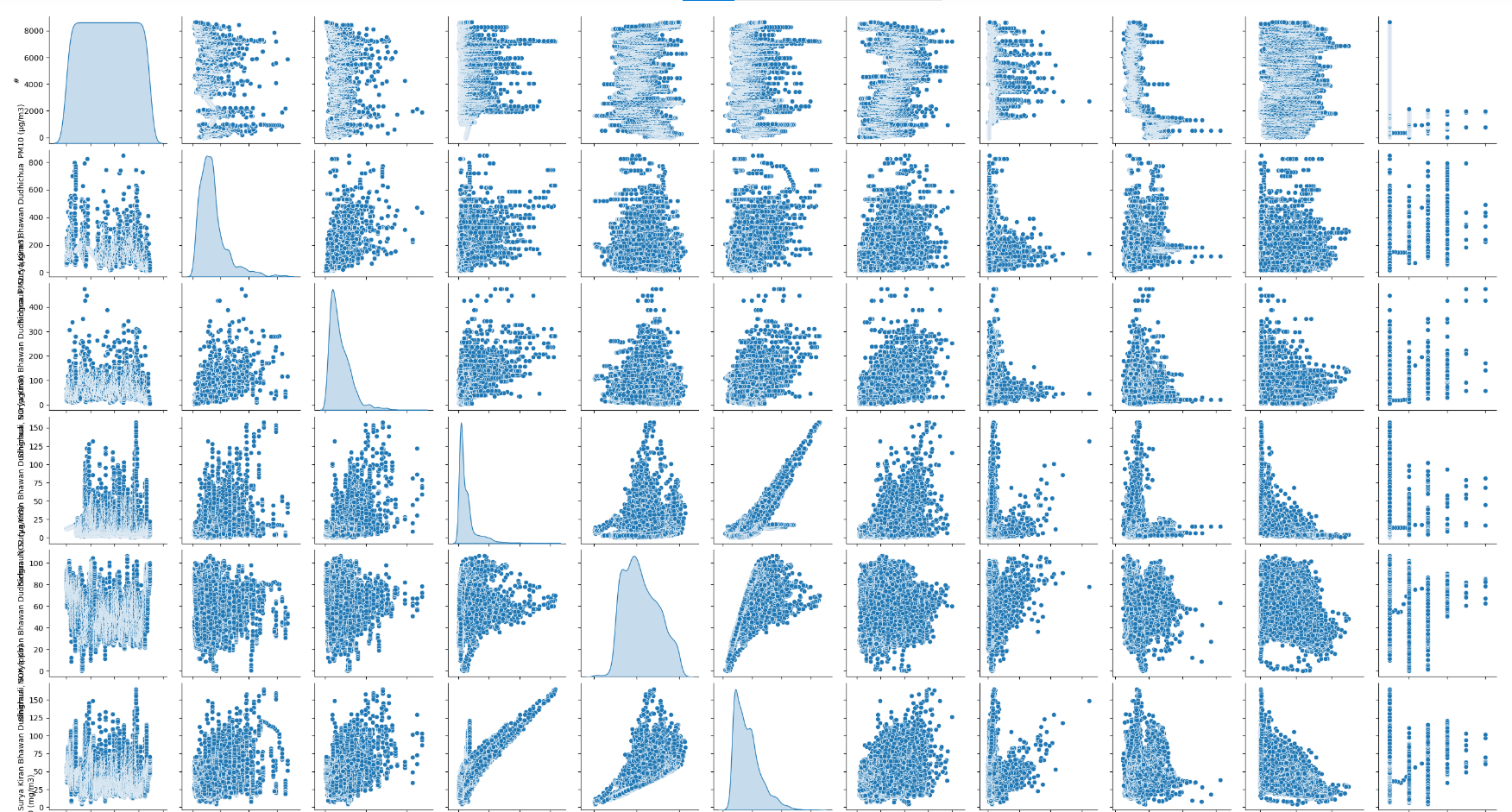
and vice-versa for large value of NO these is no SO2.

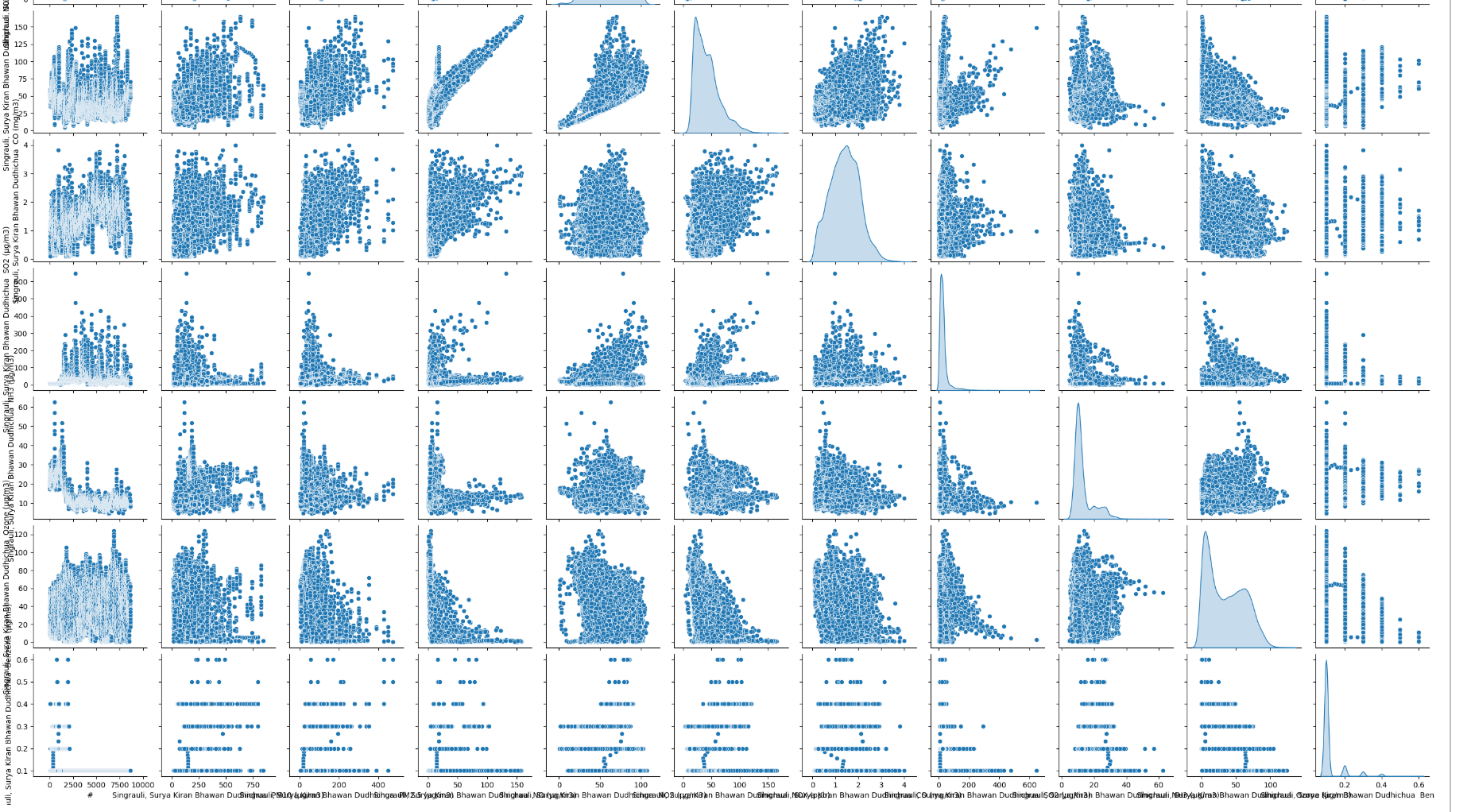
And Correlation factor for SO2 and PM2.5 is 0.007 which is also very low.



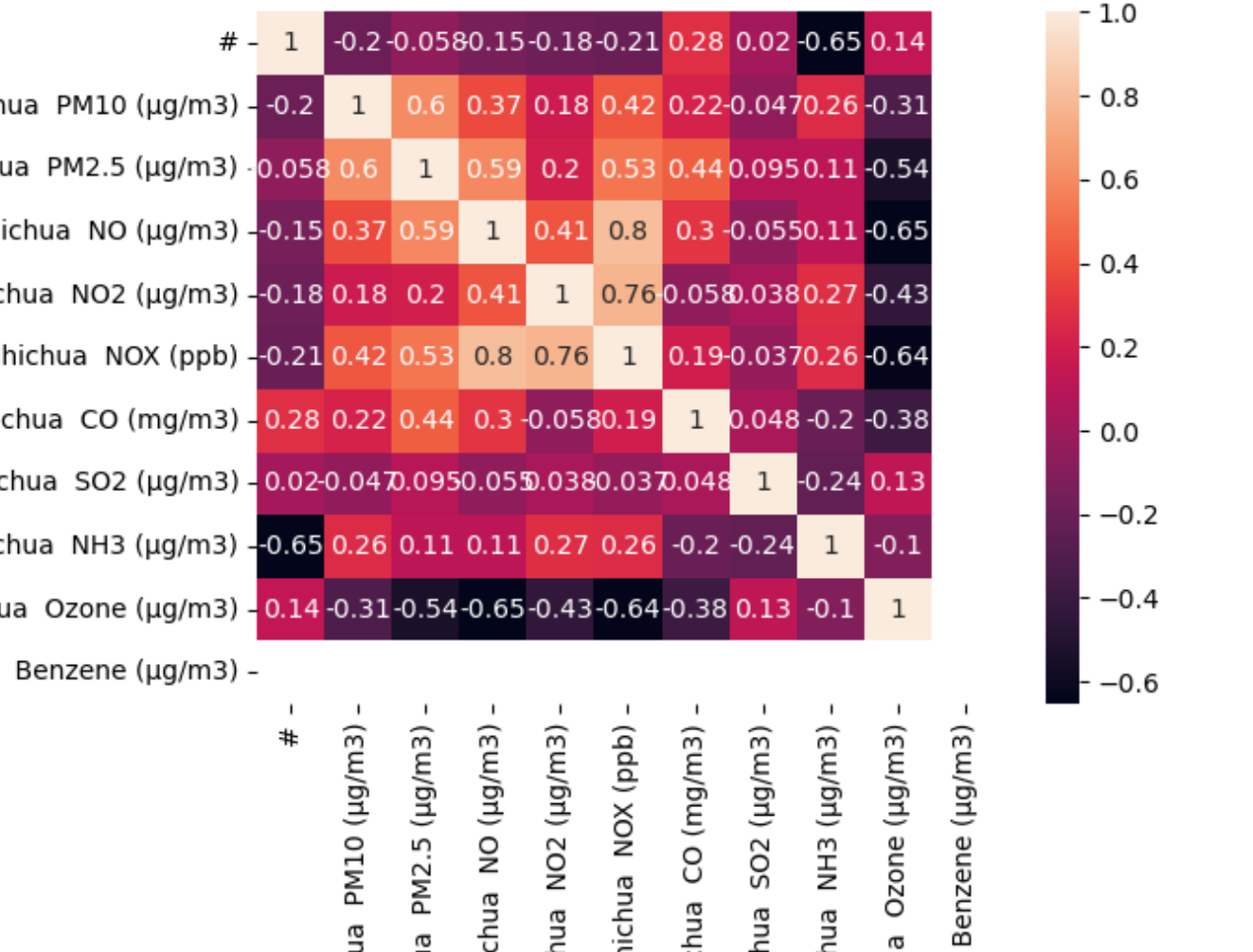
This graph is for SO2 and PM2.5 correlation here also we can see for large value of SO2 no PM2.5 present and vice versa for large value of PM2.5 no SO2 present.

Complete correlation graph is as below

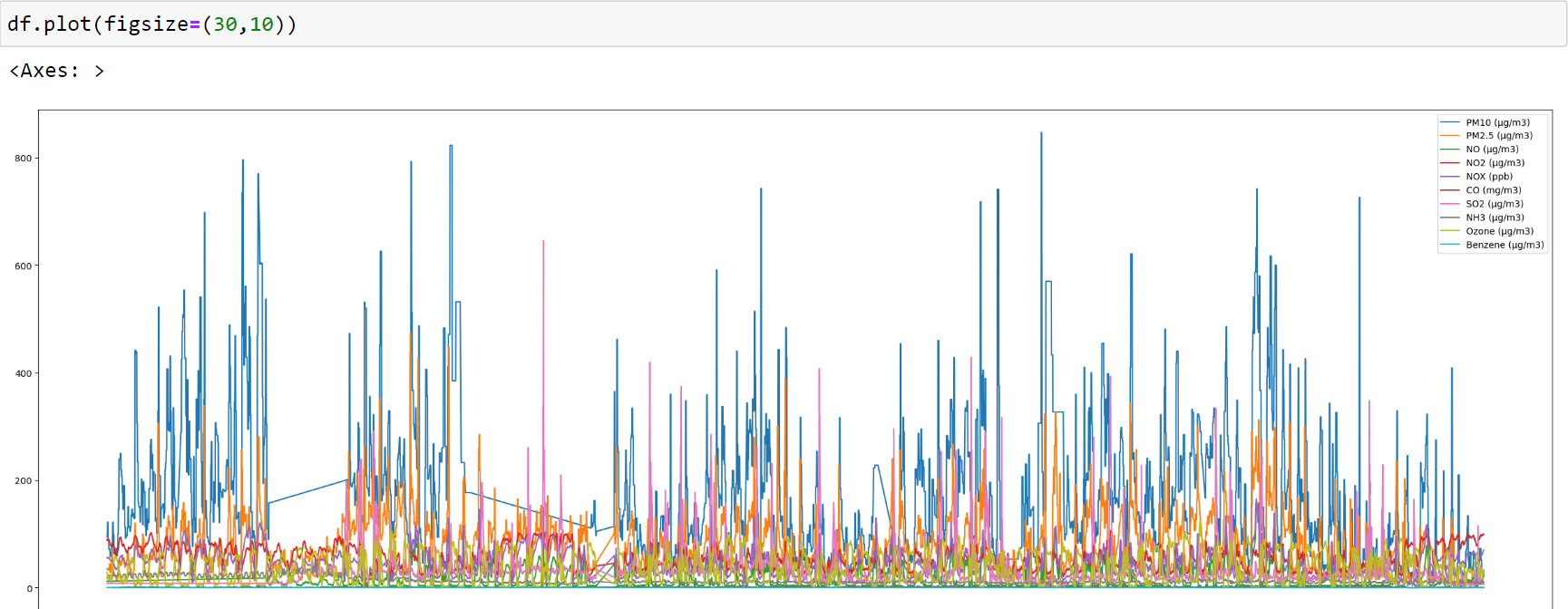




And for corresponding correlations graph heat amp as per given below



All Pollution data on same curve



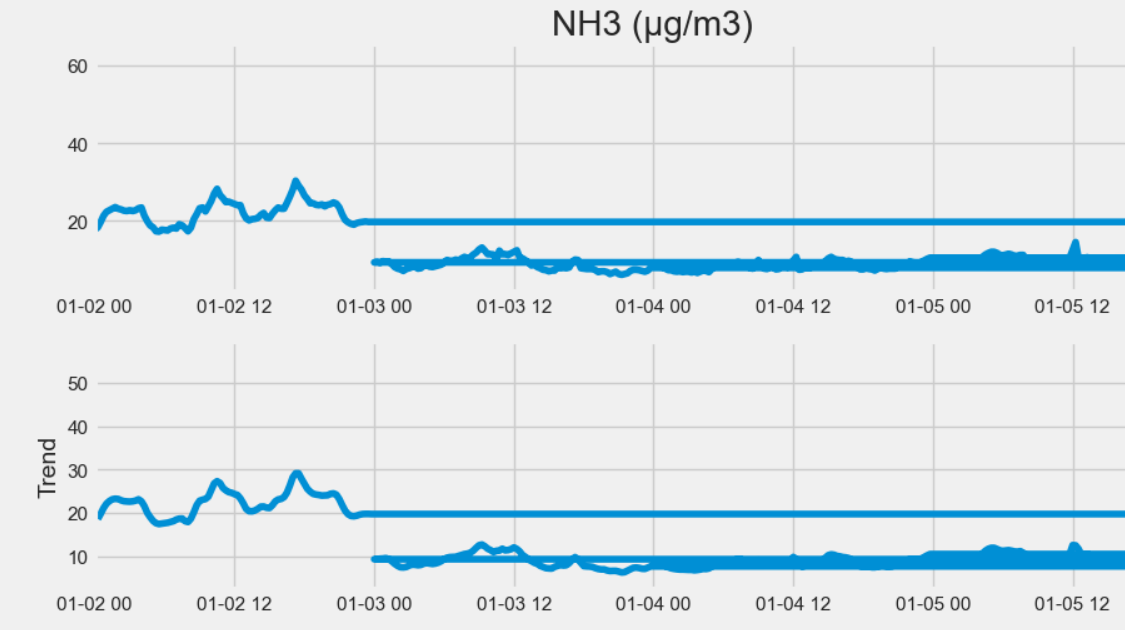
Now we came to cyclic and trend analysis

So let’s I take NH3 pollutants

A picture containing text, screenshot, line, plot

Description automatically generated

This curve is for NH3 pollutant.

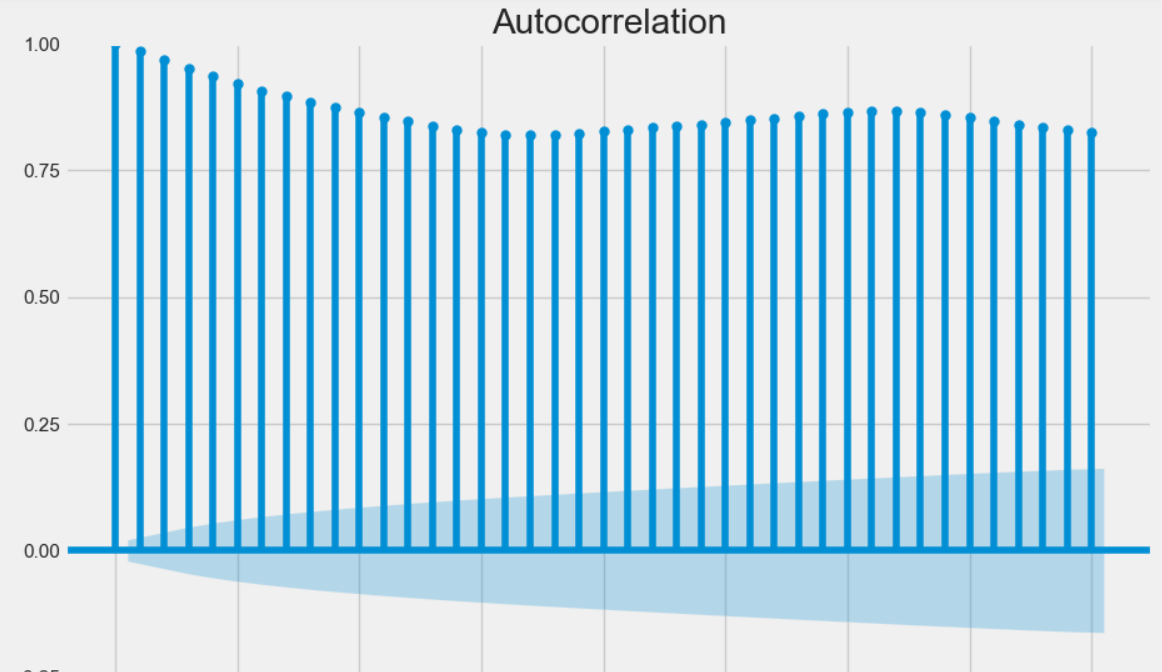


And this curve is for NH3 Trend in 3 months

A picture containing text, line, font, plot

Description automatically generated

These two curves for after observe in seasonal and trend

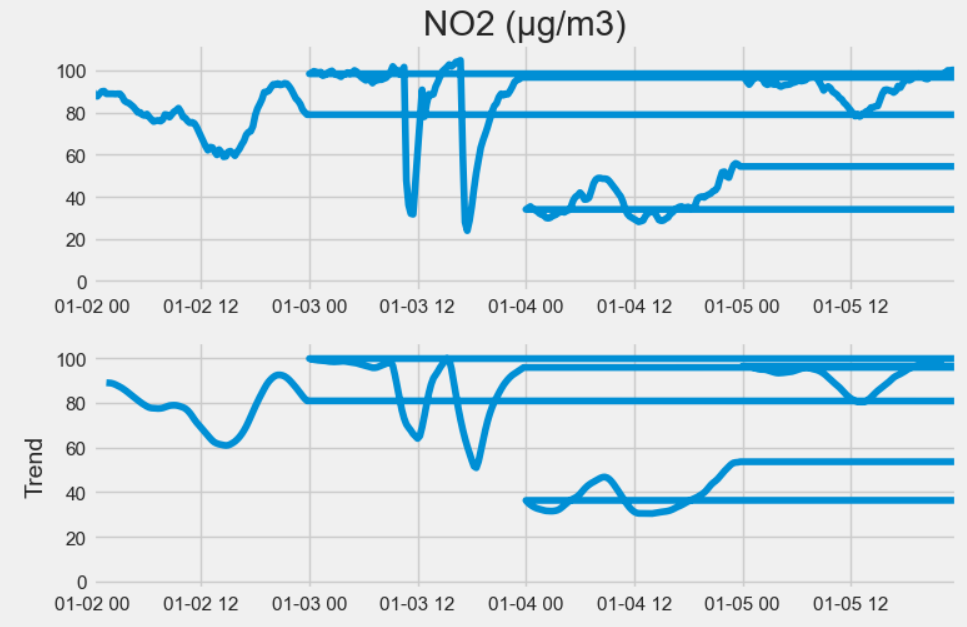


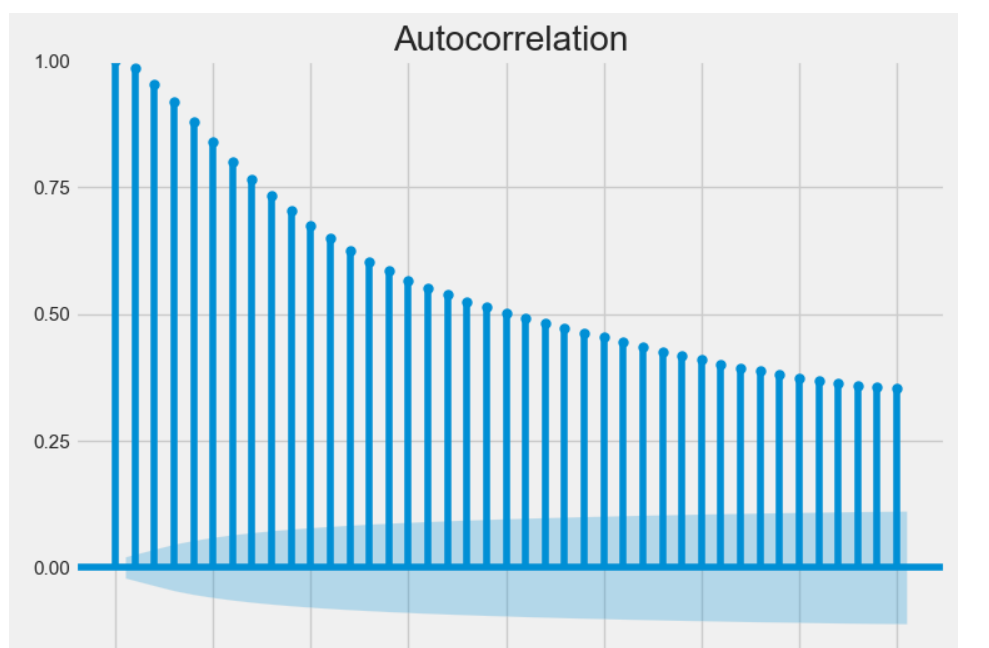
A picture containing line, plot, diagram

Description automatically generated

These two for auto correlation and partial auto correlation for NH3

Now, let’s same observe for NO2



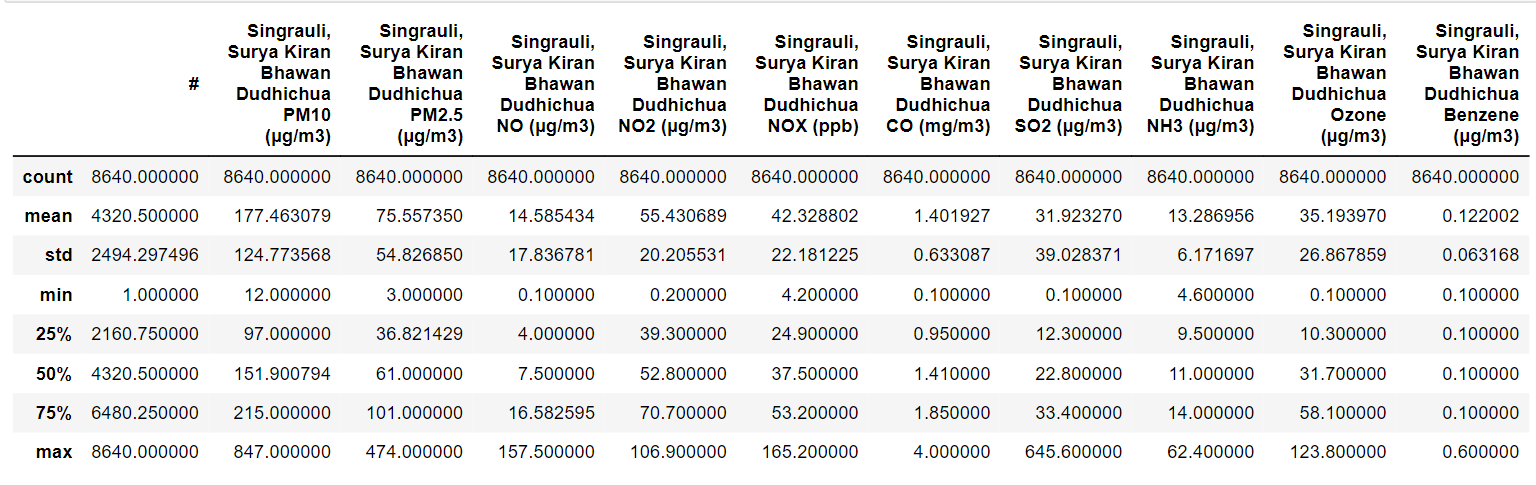


Here we can see for NO2 for month feb and march and April different trend following.

In the feb month there was high proportion of NO2 but in month march and April its decrease.

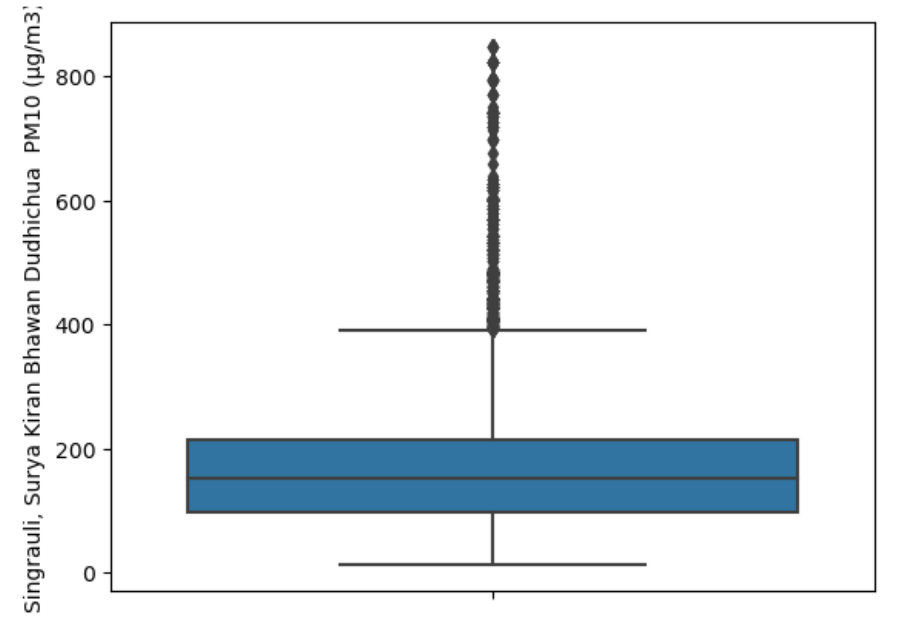
And, can observe in correlation graph of NO2.

Descriptive analysis as follow



Here for all 3 months data, we can see mean, standard deviation, min-max value observed during that between 3-month period of time and also we can see QQ value for 25%,50%,75%.

Lets plot QQ plot for PM2.5 and which is look as follows



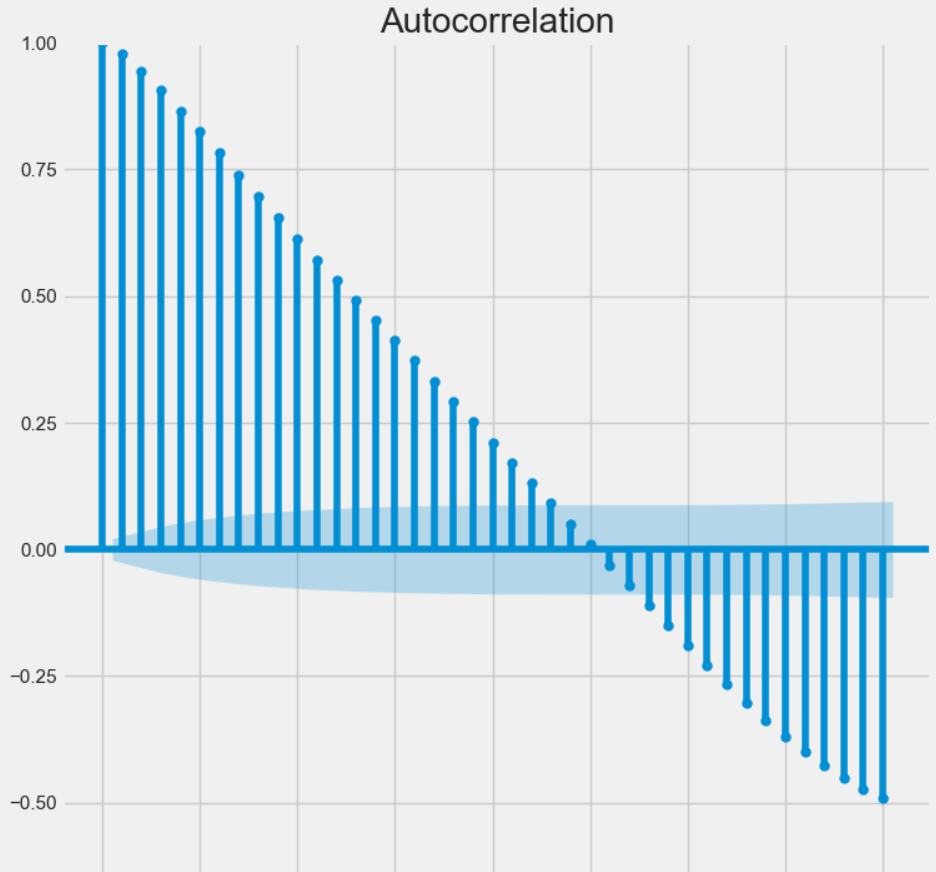
In this graph we can say that point above 85-90% are outliers for our and we can remove them (as they are less in number)

And if plot QQ for Ozone which is look like as follows

A picture containing rectangle, screenshot, line, text

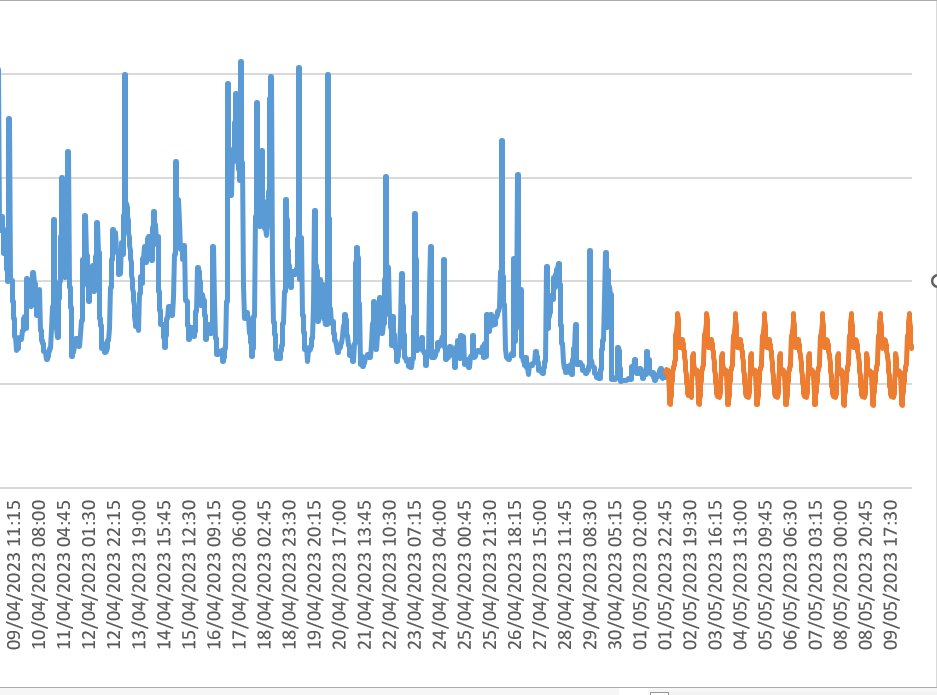
Description automatically generated

Here we see no significant outliers are and we also see correlation graph for Ozone as per follows



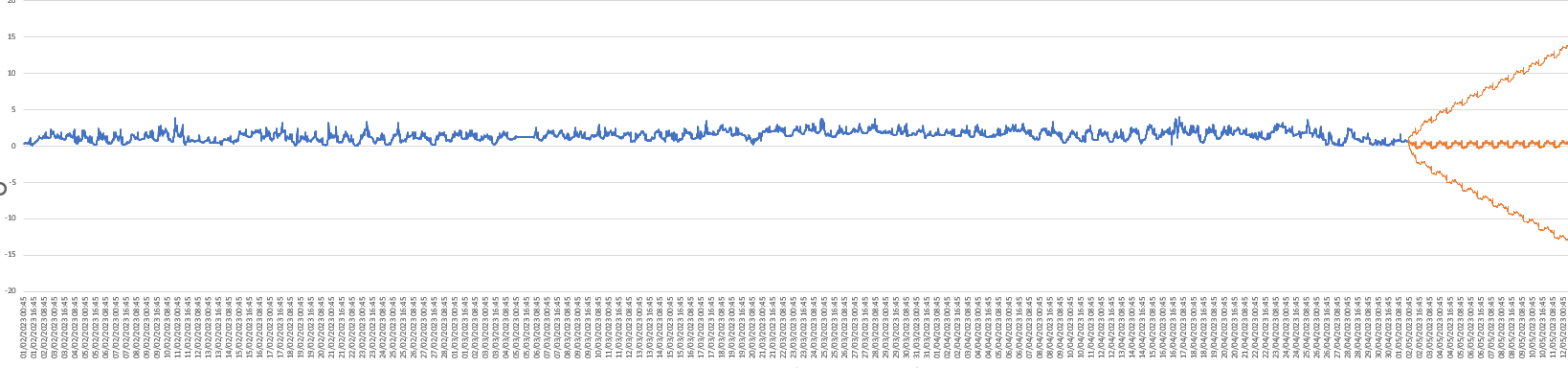
For future pollution prediction we can use ARMA/ARMIA model to forecast future pollution

For PM2.5 pollutants forecast value for 2 may to 10 may as per below



It’s following the trend that each day from 13:00 to 15:00 there is a peak in pollutant.

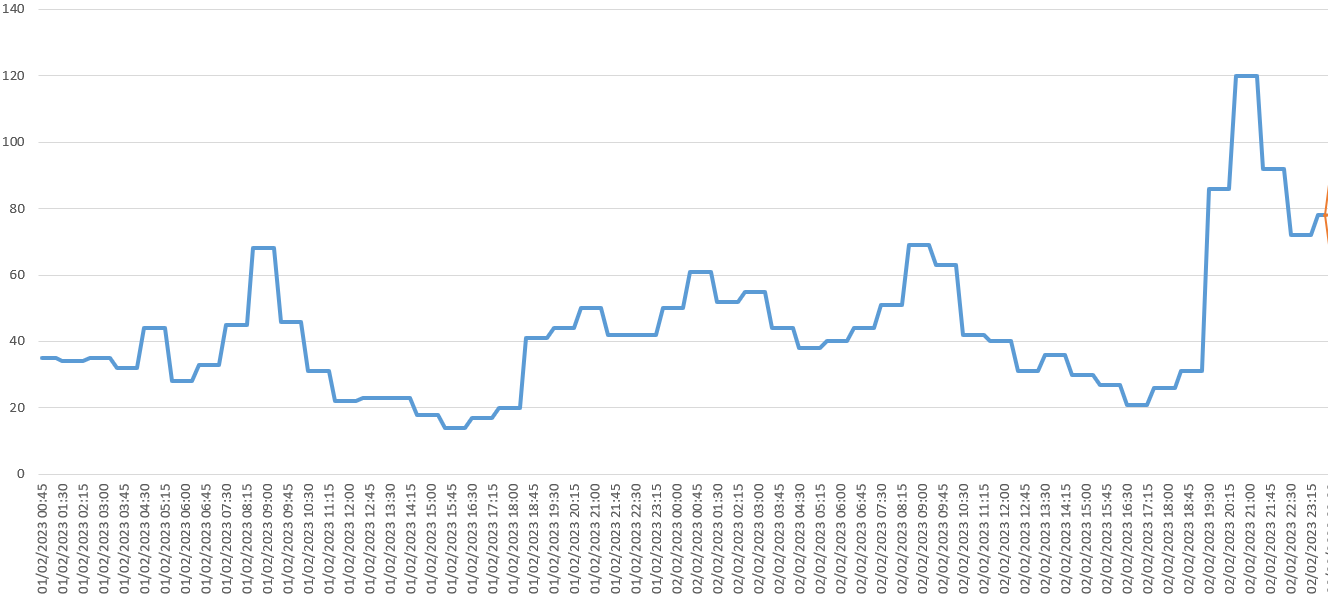
And same for let’s CO forecast for 2 may to 12 may

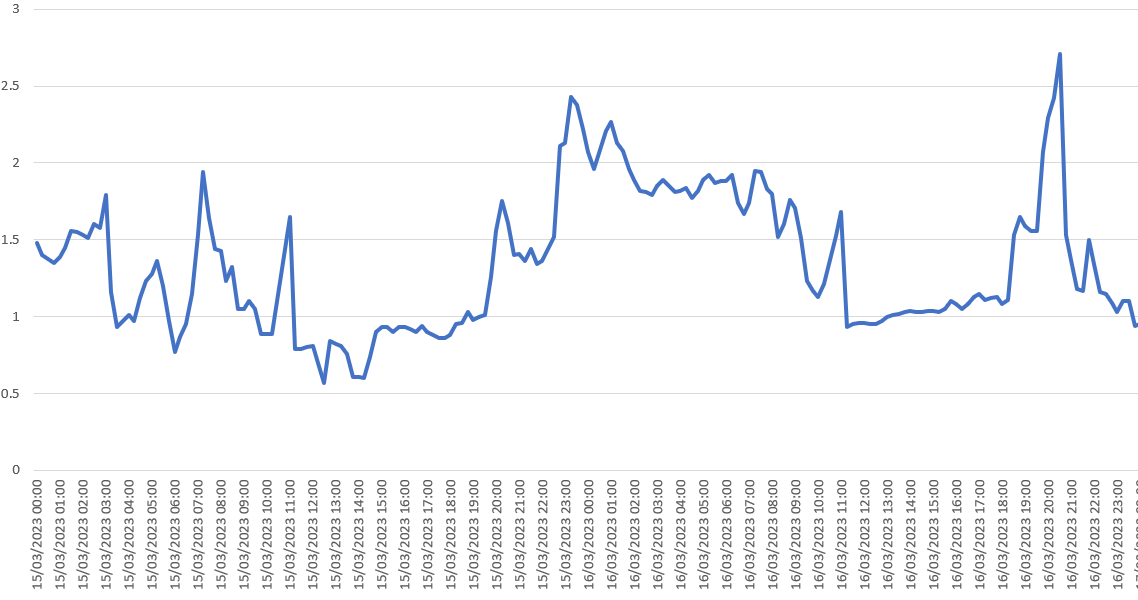


It can also predict with some lower confidence and high confidence value.

And one point we should keep in mind that when we forecasting we should forecast for short period of time.

Segmentation- Now let’s observe two days for PM2.5 and we can after 2-3 hours of blasting at around 14:00 we can there is rise in pollutants.



And For supports our point we randomly pick two days and pollutant let CO pollutants and two days as 15 and 16 of march

In this also we can observe same pattern that after 2-3 hours of blasting there is rise in pollutant in air.

Thanks-

Name- Jetha Ram

Roll no. -210473