Assignment-1

What are open-pit coal mines?

- Open-pit coal mines, also known as open-cast or open-cut mines, are surface mines where coal is extracted from the earth's surface by removing the overlying layers of soil, rock, and vegetation. These mines are characterized by their large open pits or excavations, as opposed to underground mines where coal is extracted from below the surface.

What is Blasting?

- Blasting is a process used in mining, construction, and demolition activities that involves the use of explosives to break or fragment rock, concrete, or other materials.

Blasting from open-pit coal mines causing massive air pollution

What are those pollutants?

- Particulate Matter (PM)
- Nitrogen Oxides (NOx)
- Sulphur Dioxide (SO2)
- Carbon Monoxide (CO)

Singrauli is a region in the state of Madhya Pradesh, India, known for its rich coal reserves and extensive open-pit coal mining operations. It is one of the largest coalfields in India and is home to several major coal mines.

The Singrauli coalfield is part of the larger Son-Mahanadi coalfield, which is spread across the states of Madhya Pradesh and Uttar Pradesh. The coalfield covers an area of approximately 2,200 square kilometres and contains significant reserves of coal.

Open-pit mining, also known as **open-cast mining or surface mining**, is the predominant method used for coal extraction in Singrauli.

What Open-pit mining?

- This means that pollutants are released for sure.

The two main air pollutants in NCL coal fields are suspended particulate matter (SPM) and reparable particulate matter (RPM). Air quality monitoring is regularly carried out at both dust generating and non-generating locations in the vicinity in order to evaluate the particulate pollution in and around the opencast mining projects of the Singrauli coalfield. SPM and RPM concentrations are

predominated at coal working surfaces, coal yards, coal handling facilities, and haul roads used to transport coal, as well as close to drilling sites, in overburden, and on such haul roads.

Data was recorded for the following pollutants:

PM10, PM2.5, SO2, NO2, NOx, CO, NH3, O3 and Benzene.

And the data for above pollutants is stored under following columns

Singrauli, Surya Kiran Bhawan Dudhichua PM10 (µg/m3)
Singrauli, Surya Kiran Bhawan Dudhichua PM2.5 (µg/m3)
Singrauli, Surya Kiran Bhawan Dudhichua NO (µg/m3)
Singrauli, Surya Kiran Bhawan Dudhichua NO2 (µg/m3)
Singrauli, Surya Kiran Bhawan Dudhichua NOX (ppb)
Singrauli, Surya Kiran Bhawan Dudhichua CO (mg/m3)
Singrauli, Surya Kiran Bhawan Dudhichua SO2 (µg/m3)
Singrauli, Surya Kiran Bhawan Dudhichua NH3 (µg/m3)
Singrauli, Surya Kiran Bhawan Dudhichua Ozone (µg/m3)
Singrauli, Surya Kiran Bhawan Dudhichua Benzene (µg/m3)

How can you plot the time-series?

- We could plot time-series with the help of following python libraries:
  - Matplotlib
  - Pandas
  - Seaborn

  There are more libraries other than the above ones to plot time-series data.

How NA values are interfering with plotting?

- If there are NA values in your time-series data, most plotting libraries will create a gap in the line plot at the positions where the NA values occur. This can result in discontinuities or brea ks in the line connecting the data points.

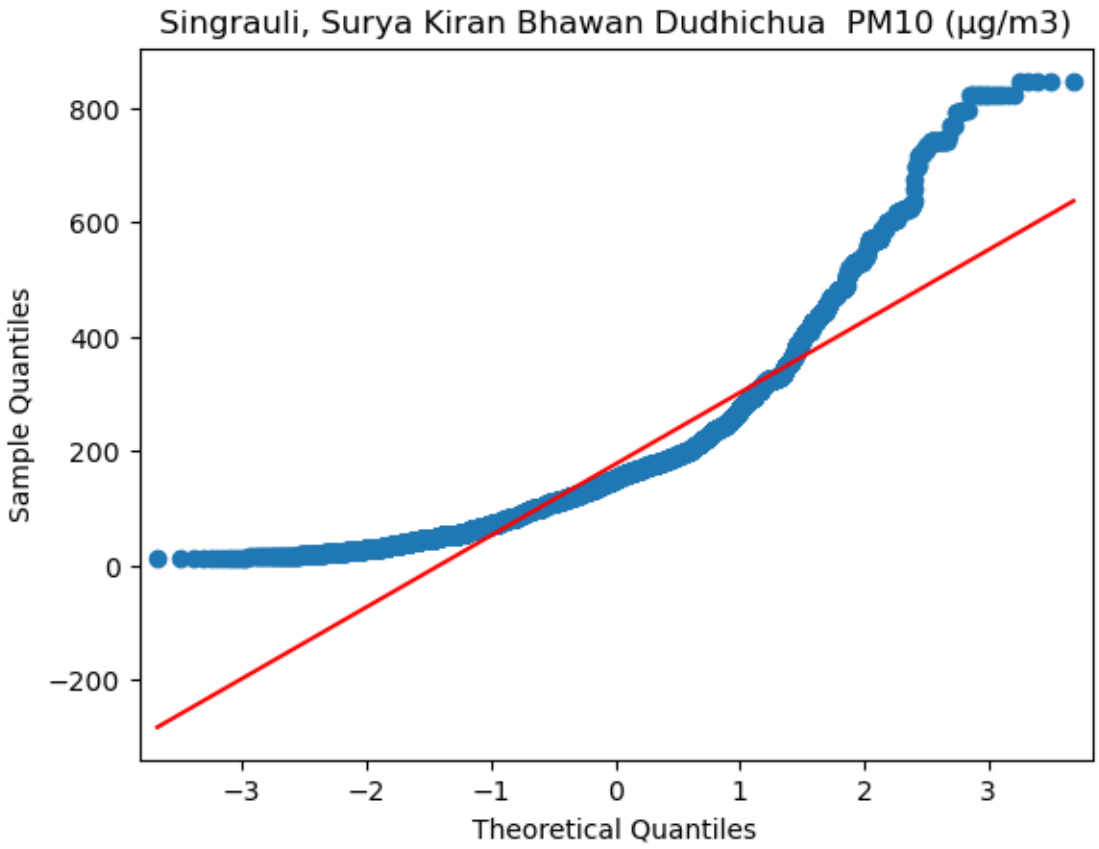
Can we just replace NA values with 0 values?

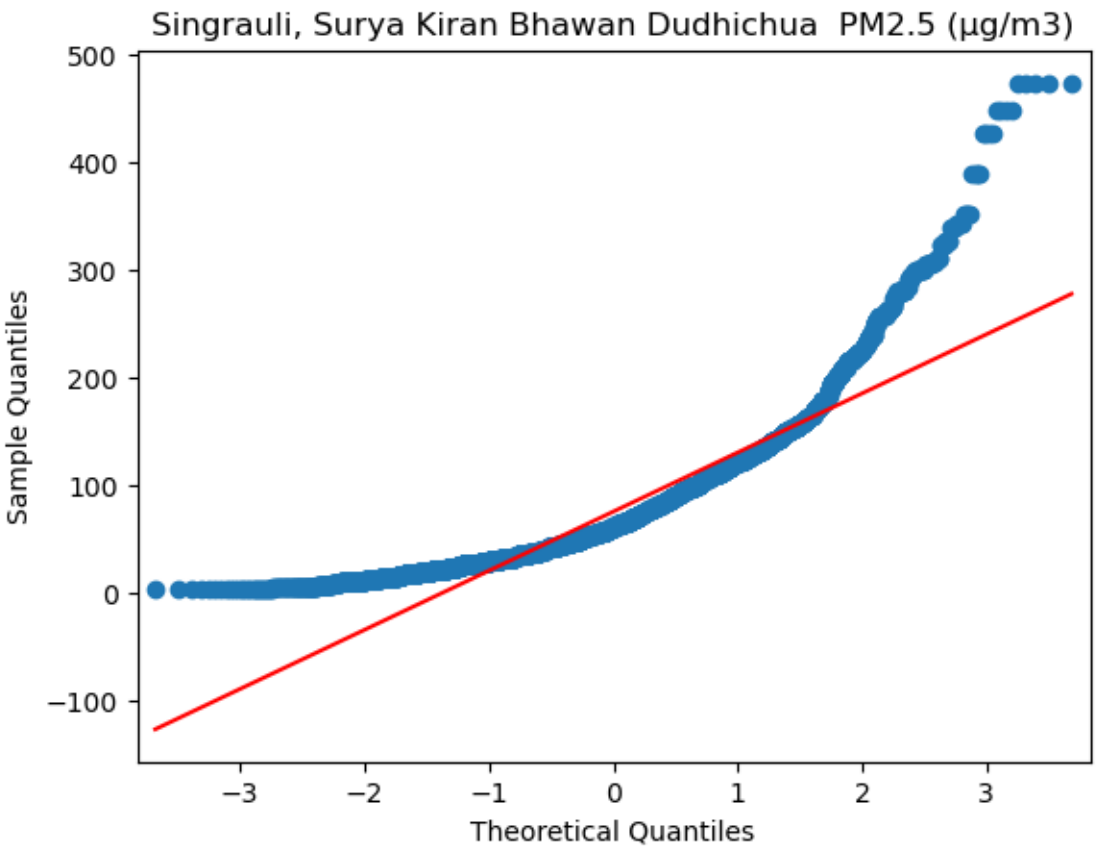
It is a simplistic method that does not consider the underlying patterns or relationships in the data an d can lead to the following:

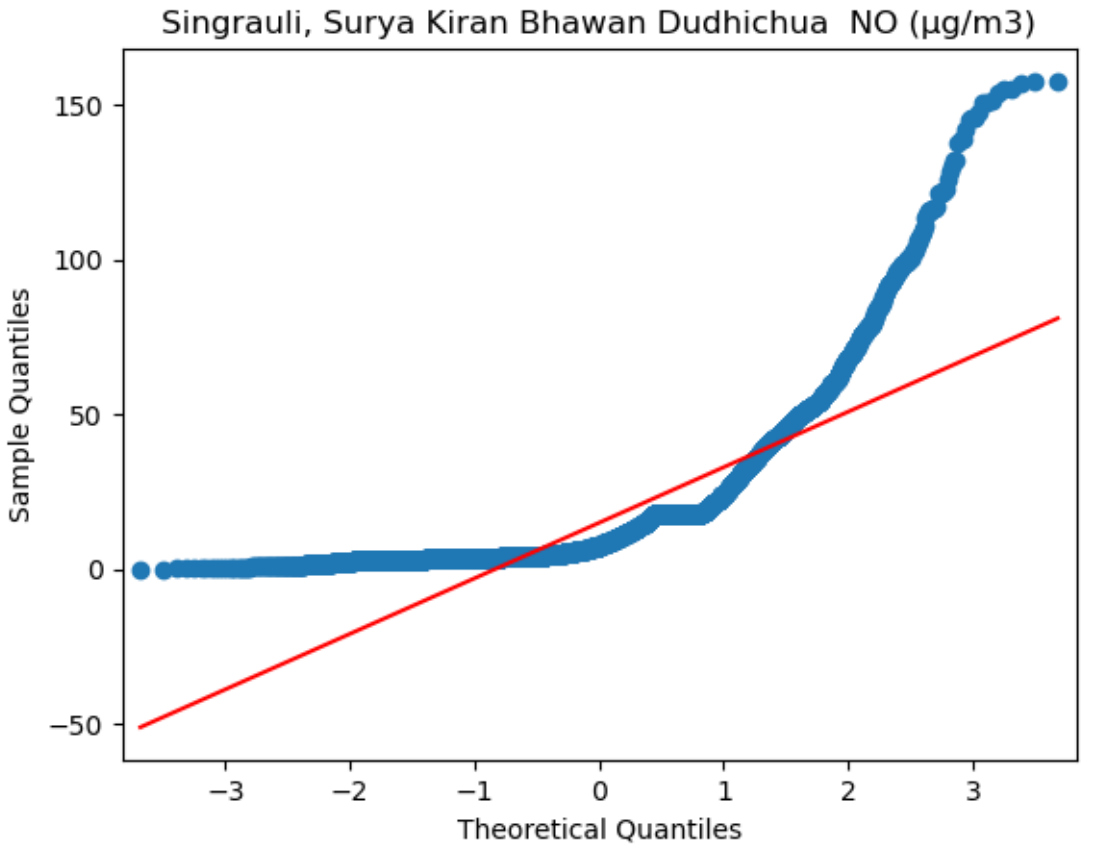
- Replacing NA values with 0 can distort the interpretation of the data, especially if zero is not a meaningful or accurate representation of the missing values. It is crucial to understand the context and nature of the missing data before replacing it with a specific value.
- Replacing NA values with 0 can impact the statistical analysis and computations. It can influe nce measures such as mean, variance, and correlation. If the missing values are not truly zer o, this can lead to biased estimates and incorrect inferences.
- When plotting data with replaced NA values, it can create misleading visualizations. A line or bar plot, for example, may show a continuous line or bar connecting or representing the repl aced values, which may not reflect the actual data.

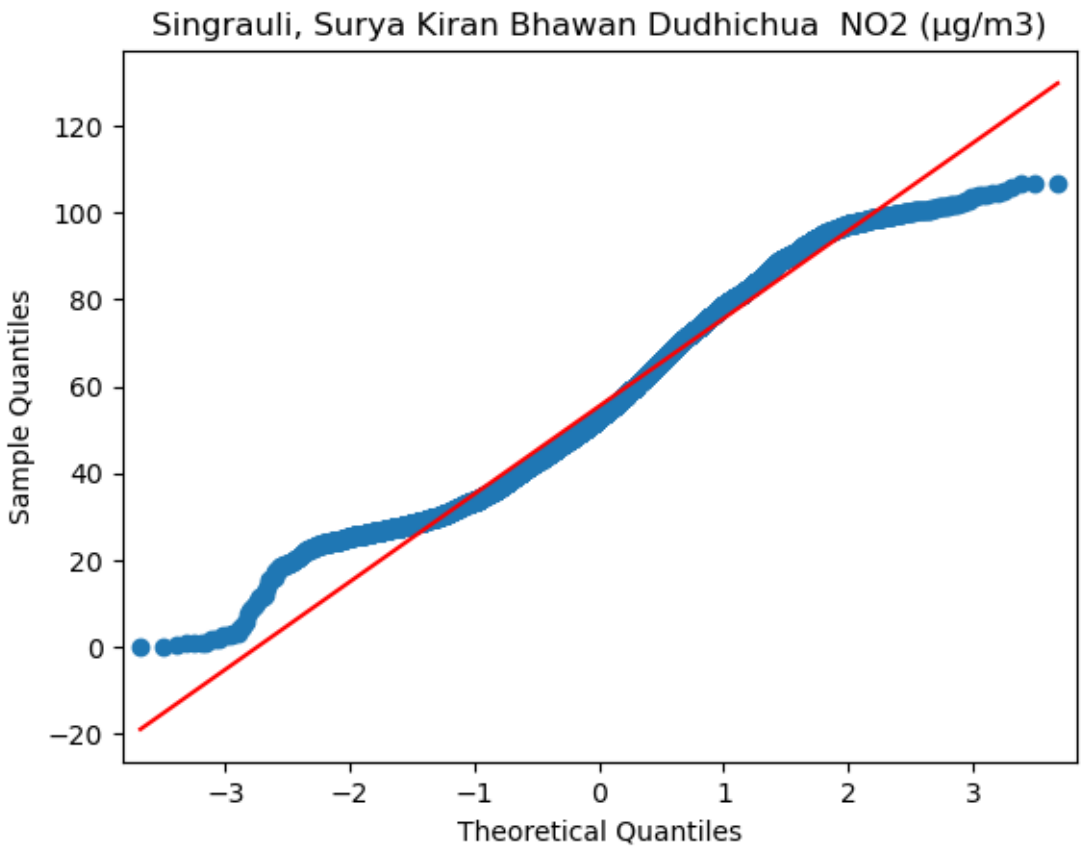
STATISTICAL INFERENCE

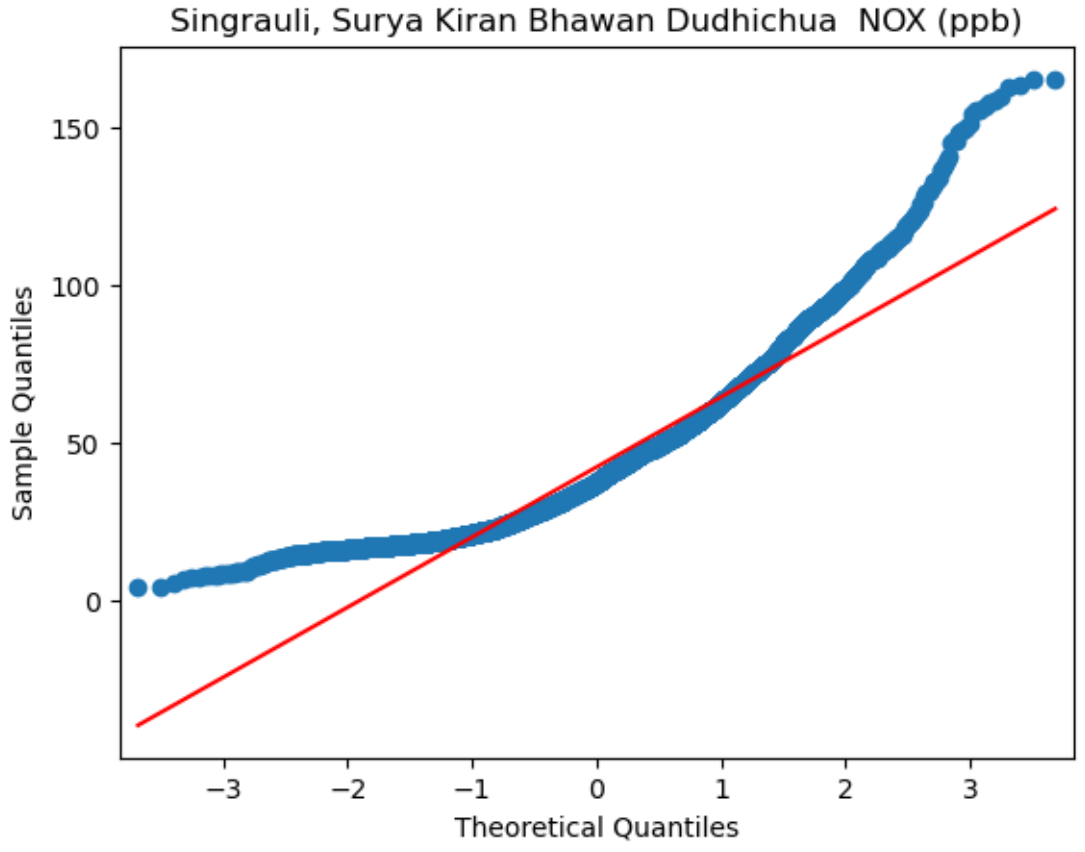
The majority of the graphs demonstrate positive skewness, which is also evident fro m the Q-Q plots. Certain graphs, such as the ones depicting SO2 and CO, do exhibit a normal distribution pattern

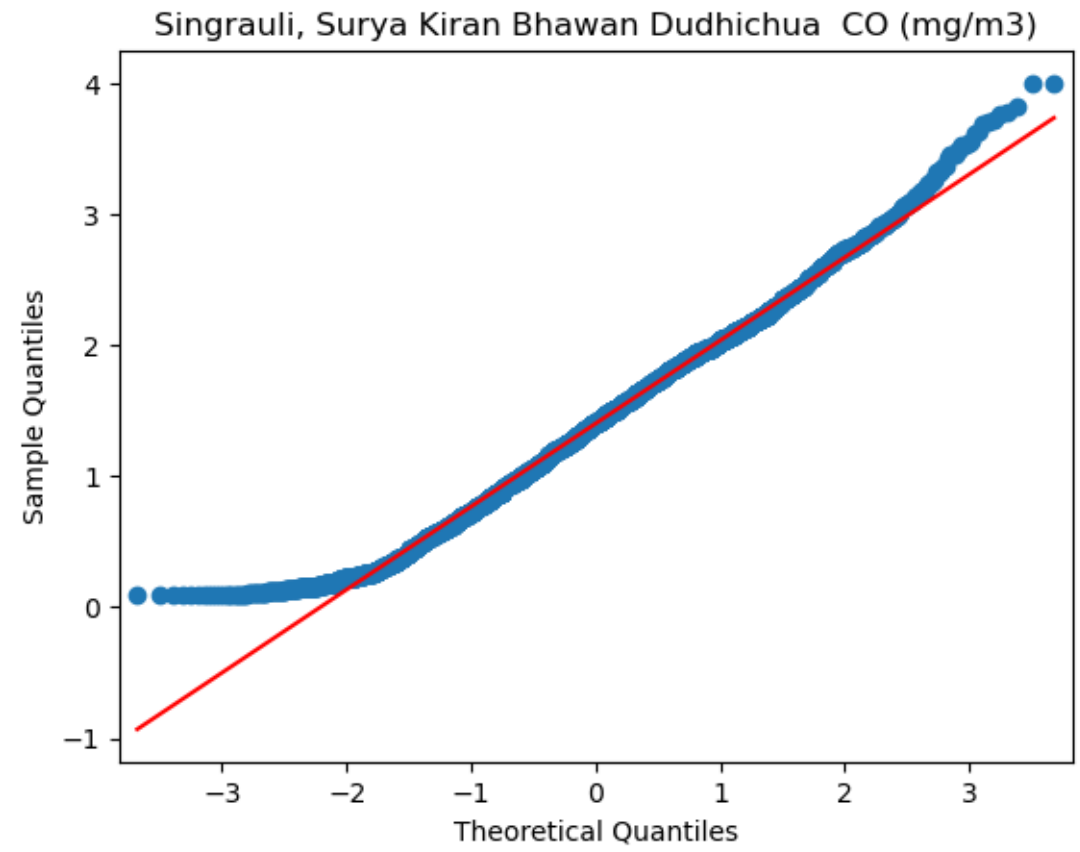


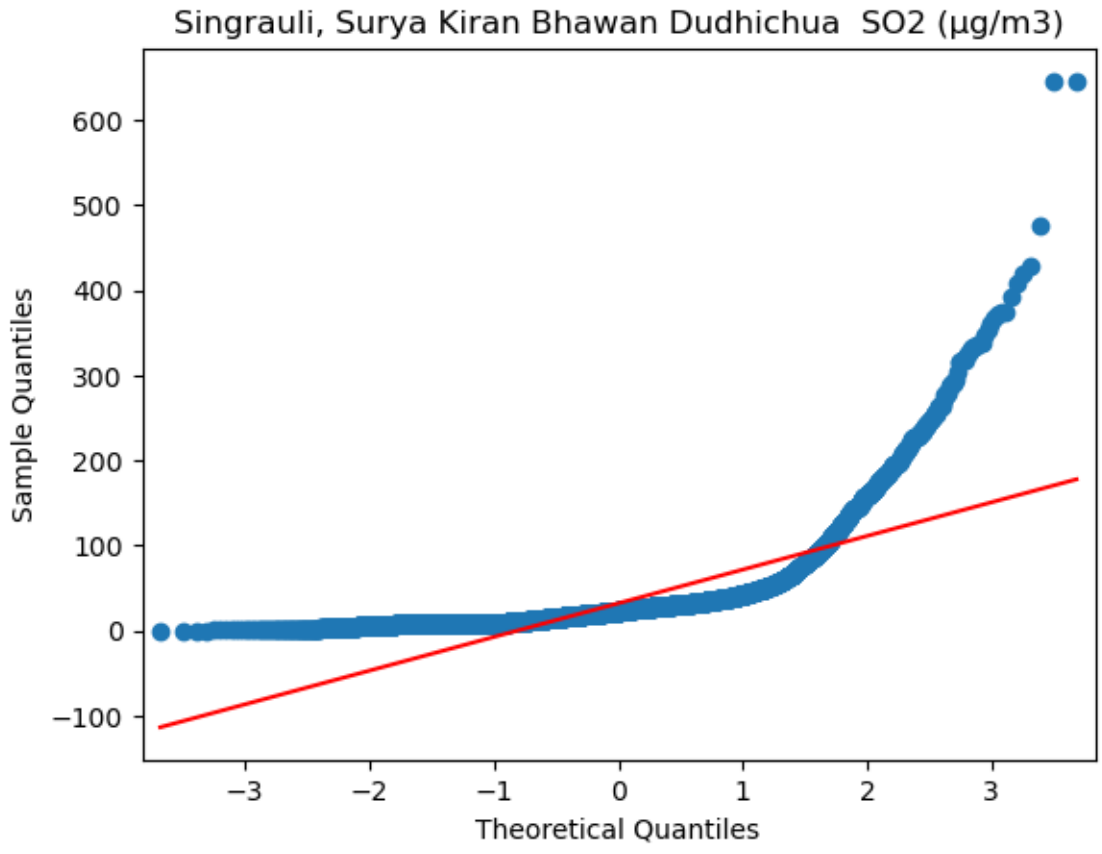

Singrauli, Surya Kiran Bhawan Dudhichua PM2.5 (µg/m3)
Sample Quantiles
Theoretical Quantiles
500
400
300
200
100
0
−100
−3
−2
−1
0
1
2
3

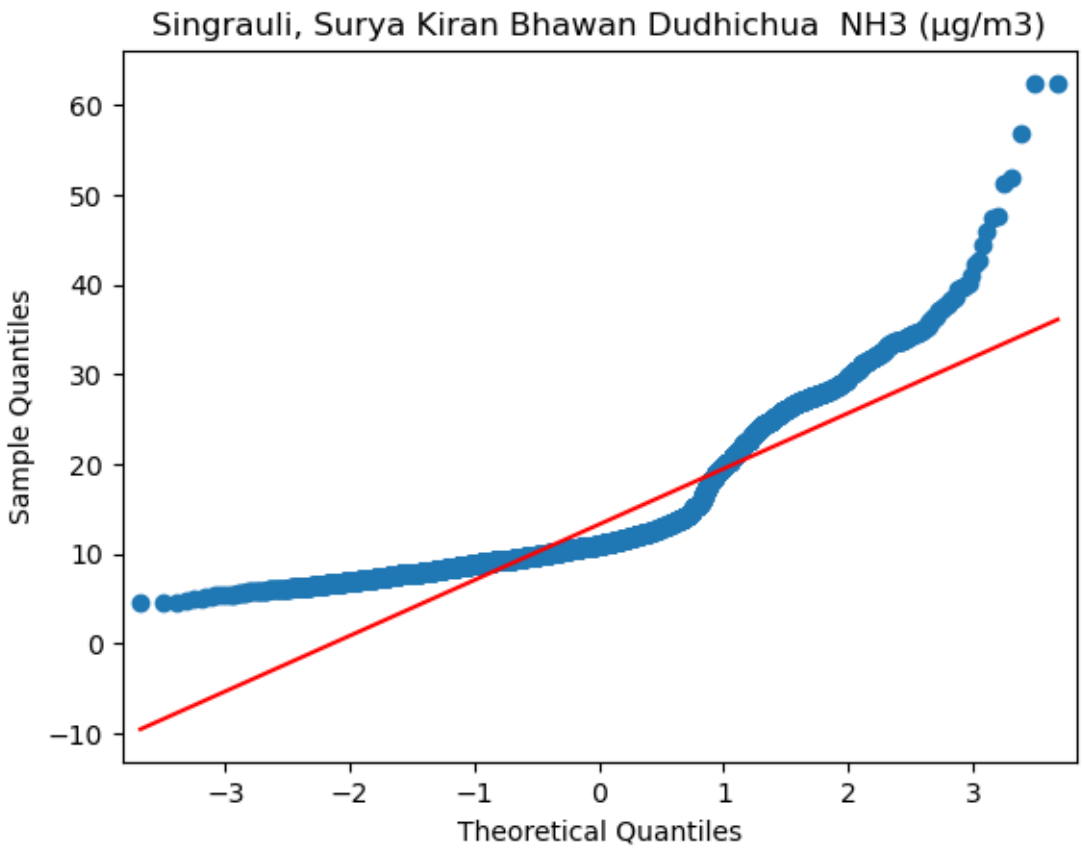

Singrauli, Surya Kiran Bhawan Dudhichua NO (µg/m3)
Sample Quantiles
Theoretical Quantiles
150
100
50
0
−50
−3
−2
−1
0
1
2
3

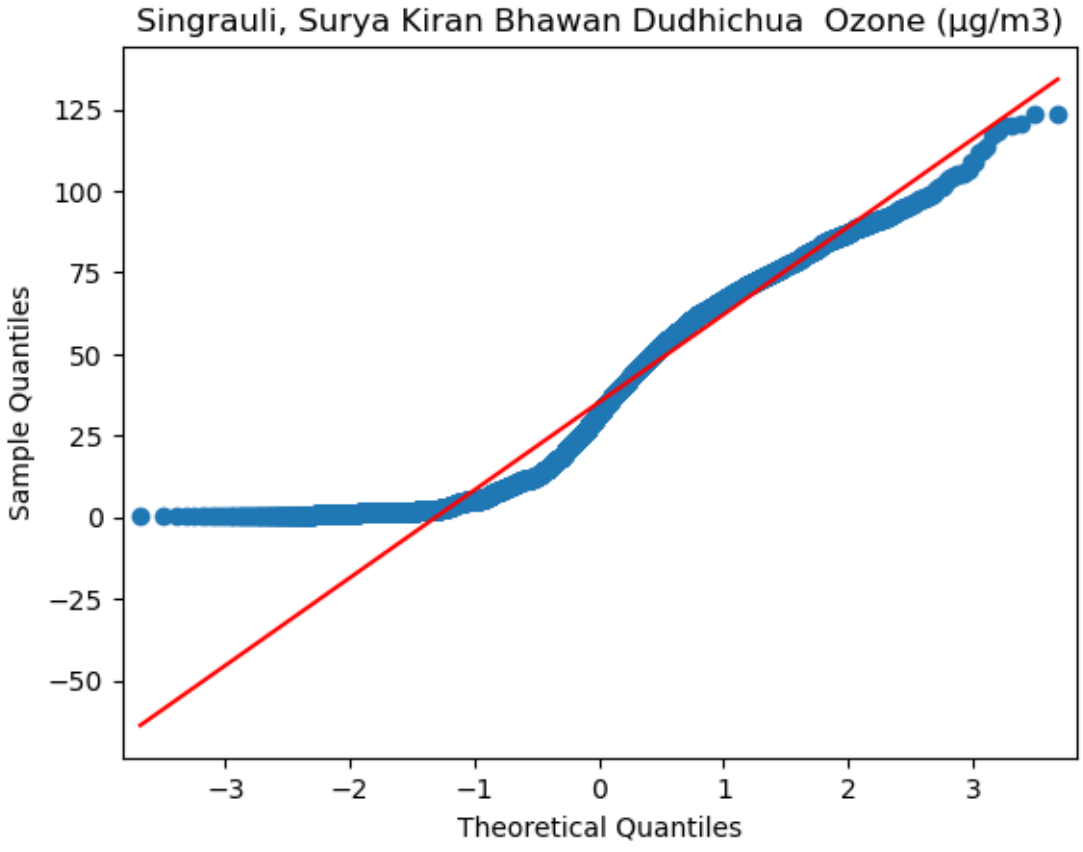

Singrauli, Surya Kiran Bhawan Dudhichua NO2 (µg/m3)
Sample Quantiles
Theoretical Quantiles
120
100
80
60
40
20
0
−20
−3
−2
−1
0
1
2
3

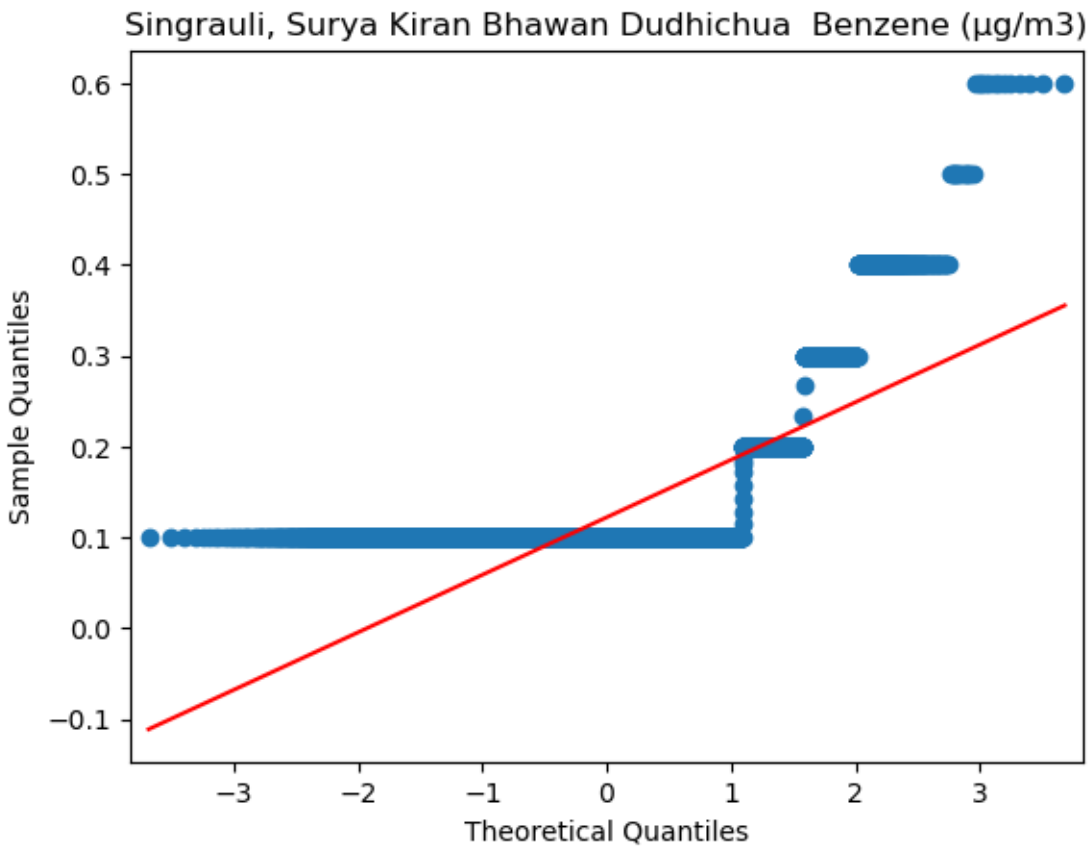

Singrauli, Surya Kiran Bhawan Dudhichua NOX (ppb)
Sample Quantiles
Theoretical Quantiles
150
100
50
0
−3
−2
−1
0
1
2
3


Singrauli, Surya Kiran Bhawan Dudhichua CO (mg/m3)
Sample Quantiles
Theoretical Quantiles
4
3
2
1
0
−1
−3
−2
−1
0
1
2
3


Singrauli, Surya Kiran Bhawan Dudhichua SO2 (µg/m3)
Sample Quantiles
Theoretical Quantiles
600
500
400
300
200
100
0
−100
−3
−2
−1
0
1
2
3


Singrauli, Surya Kiran Bhawan Dudhichua NH3 (µg/m3)
Sample Quantiles
Theoretical Quantiles
60
50
40
30
20
10
0
−10
−3
−2
−1
0
1
2
3


Singrauli, Surya Kiran Bhawan Dudhichua Ozone (µg/m3)
Sample Quantiles
Theoretical Quantiles
125
100
75
50
25
0
−25
−50
−3
−2
−1
0
1
2
3



Problem Setting and Prediction

Classification of time-series data:

Time-series data can be classified as the following:

- Stock time series data refers to measuring characteristics at a specific moment, much like a st atic image of the data as it was. Another definition of stock time series data, it refers to a coll ection of historical stock market data that records the prices and other relevant information of a particular stock over a specific period of time. This data is typically presented in a seque ntial order, with each data point representing a specific time interval.

- Flow time series data refers to a collection of sequential observations or measurements of a variable that represents the flow of something over time. It could be the flow of water in a ri ver, the flow of traffic on a road, the flow of customers in a store, or any other measurable fl ow.

What is the data type of Pollution data?

- Air pollution data is typically considered a type of flow time series data. It represents the acti vity or variation of pollutant concentrations over a specific period, such as hourly, daily, mon thly, or yearly measurements. The data reflects the changes in pollutant levels and provides i nformation about the temporal patterns, trends, and fluctuations in air quality. To classify ai r pollution data, we can consider different attributes, such as pollutant concentrations (e.g.,

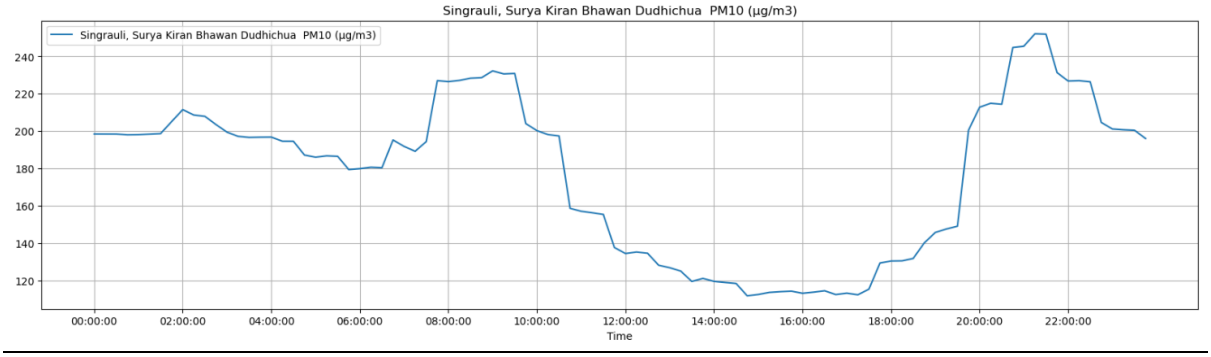
PM10, PM2.5, SO2, NO2), meteorological factors (e.g., temperature, humidity, wind speed), and temporal features (e.g., time of day, day of the week, season).
It's worth noting that air pollution data classification can serve different purposes, such as id entifying pollution events, assessing the severity of air quality, identifying pollution sources, or predicting future pollution levels. The specific classification approach would depend on th e objectives and the available data.

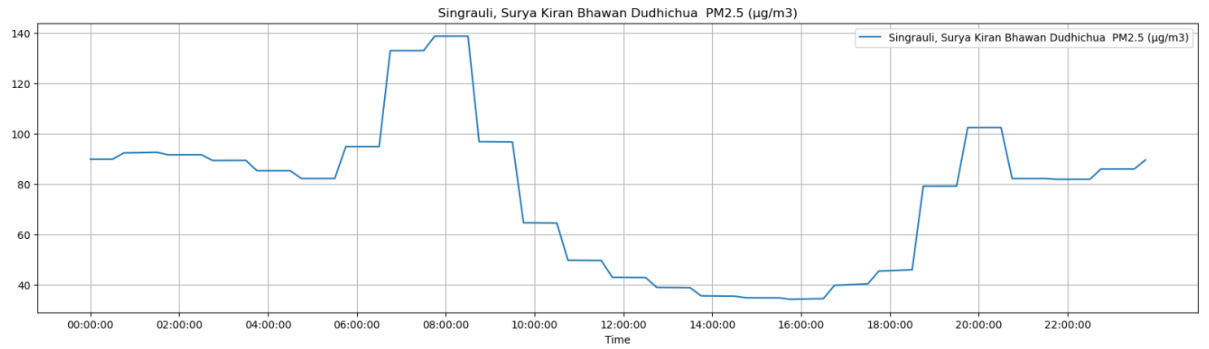
Identify patterns in time series data at the time of coal India open-pit blasting effect, in co al India blasting effect time is 13:45 pm to 14:45 pm

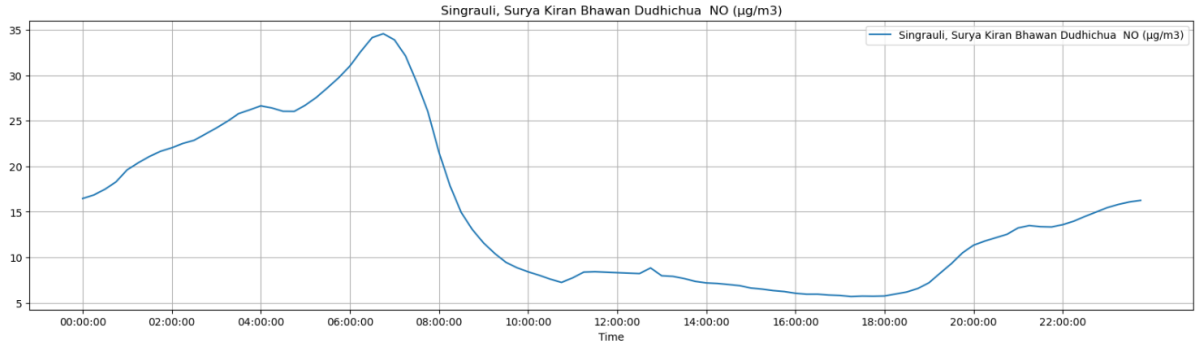
Code: -

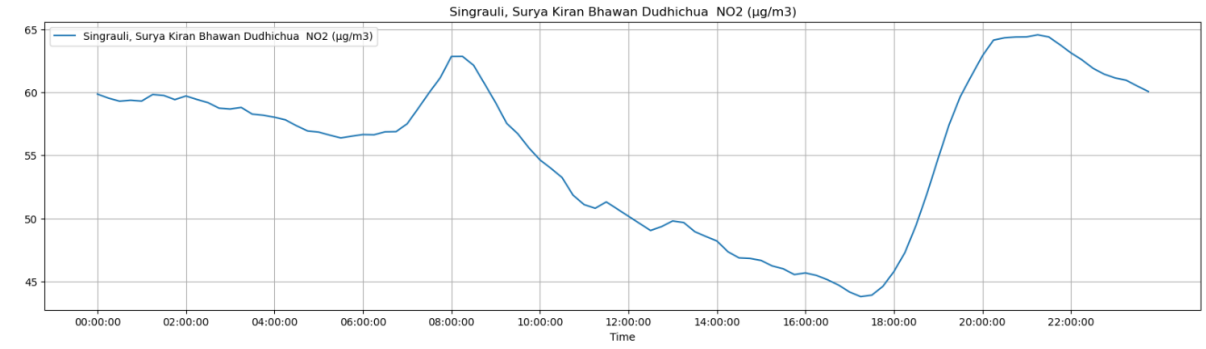
```
df1 = dataSet.copy()
df1['From'] = pd.to_datetime(df1['From'][:8640])
df1['Time'] = df1['From'].dt.strftime('%H:%M:%S')

df1_means = df1.groupby('Time').mean().reset_index()

plt.figure(figsize= (20, 15))
for column in columns:
    plt.plot(df1_means['Time'] ,df1_means[column] , label = column)
tick_frequency = 8

x_ticks = range(0, len(df1_means['Time']), tick_frequency)

x_labels = df1_means['Time'].iloc[x_ticks]
plt.xticks(x_ticks, x_labels)

plt.title("Combined Graph")
plt.xlabel('Time')
plt.grid(True)
plt.legend()
plt.show()
```

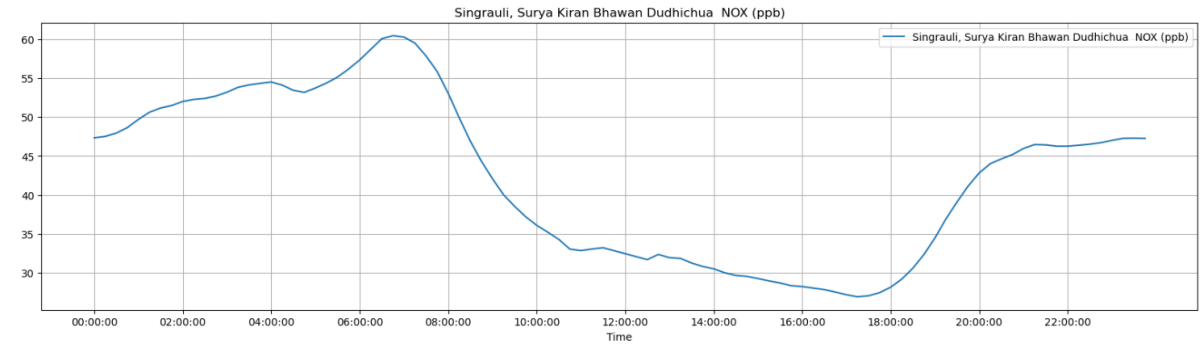
Graphs: -

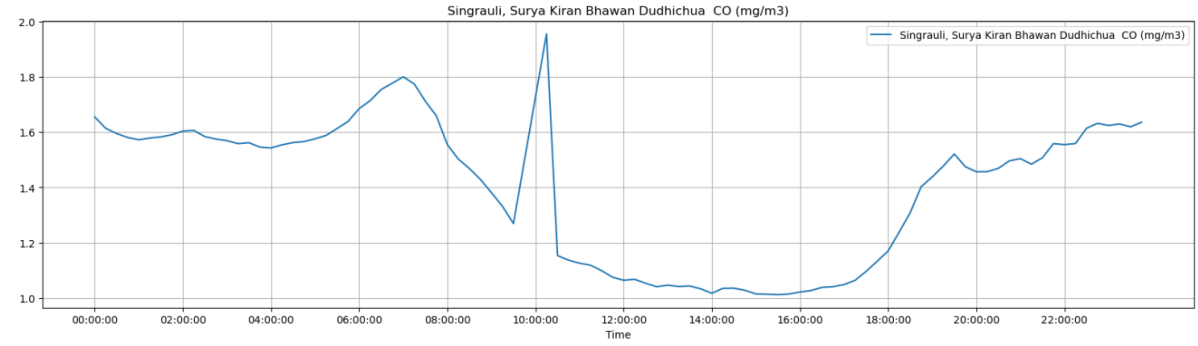


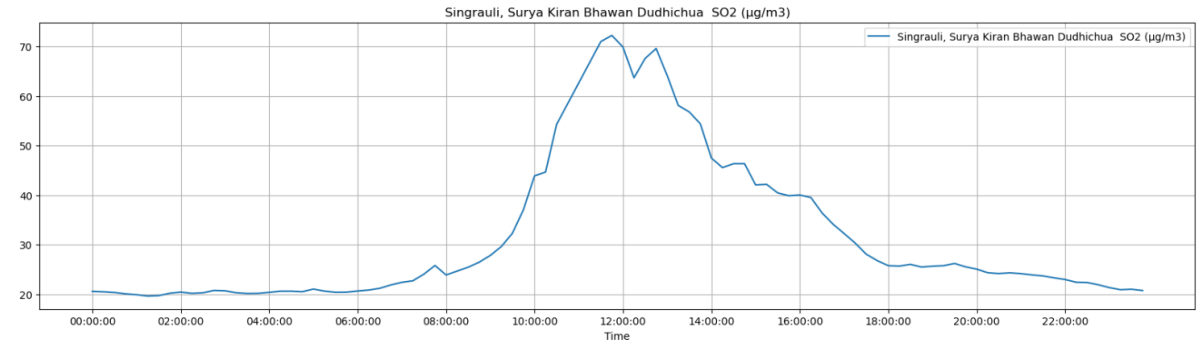


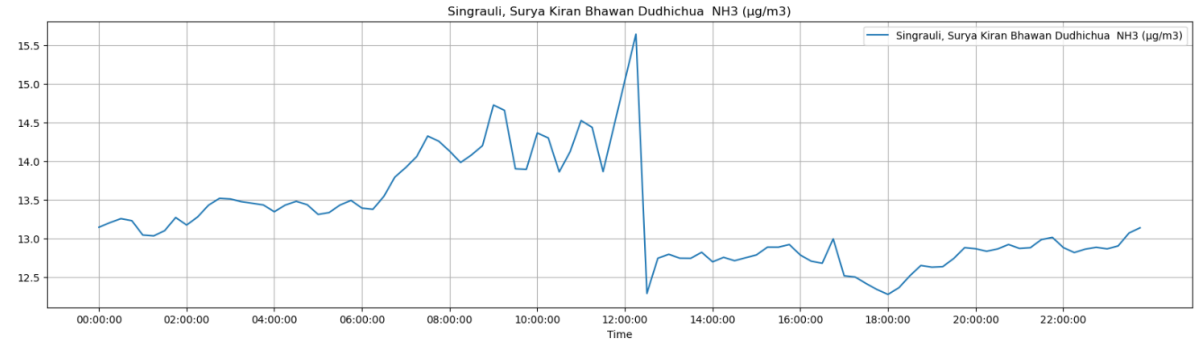

Singrauli, Surya Kiran Bhawan Dudhichua NO (µg/m3)
Singrauli, Surya Kiran Bhawan Dudhichua NO (µg/m3)
35
30
25
20
15
10
5
00:00:00
02:00:00
04:00:00
06:00:00
08:00:00
10:00:00
12:00:00
14:00:00
16:00:00
18:00:00
20:00:00
22:00:00
Time

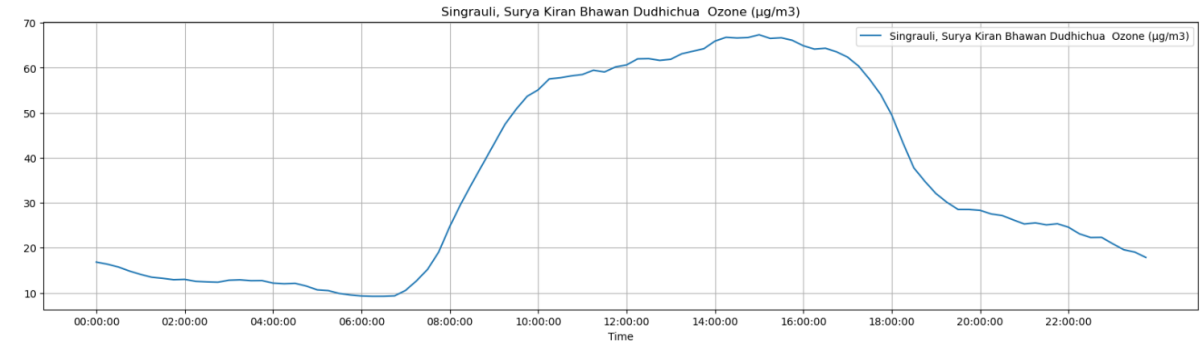

Singrauli, Surya Kiran Bhawan Dudhichua NO2 (µg/m3)
Singrauli, Surya Kiran Bhawan Dudhichua NO2 (µg/m3)
65
60
55
50
45
00:00:00
02:00:00
04:00:00
06:00:00
08:00:00
10:00:00
12:00:00
14:00:00
16:00:00
18:00:00
20:00:00
22:00:00
Time

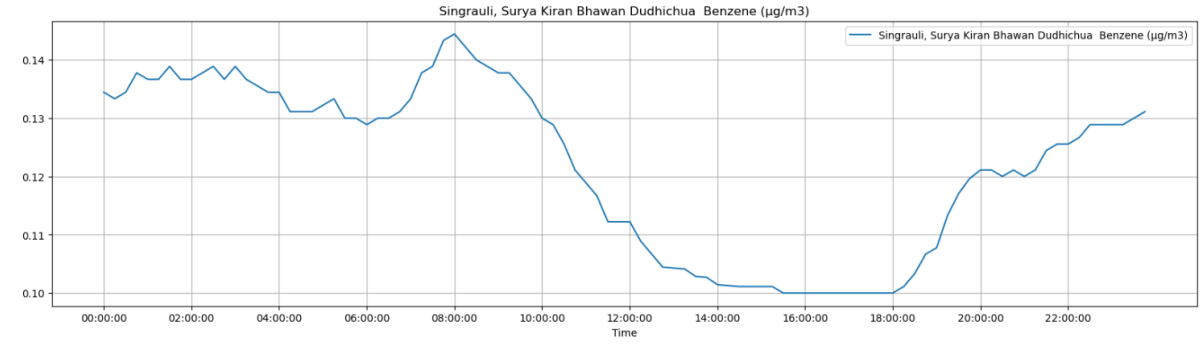

Singrauli, Surya Kiran Bhawan Dudhichua NOX (ppb)
Singrauli, Surya Kiran Bhawan Dudhichua NOX (ppb)
60
55
50
45
40
35
30
00:00:00
02:00:00
04:00:00
06:00:00
08:00:00
10:00:00
12:00:00
14:00:00
16:00:00
18:00:00
20:00:00
22:00:00
Time

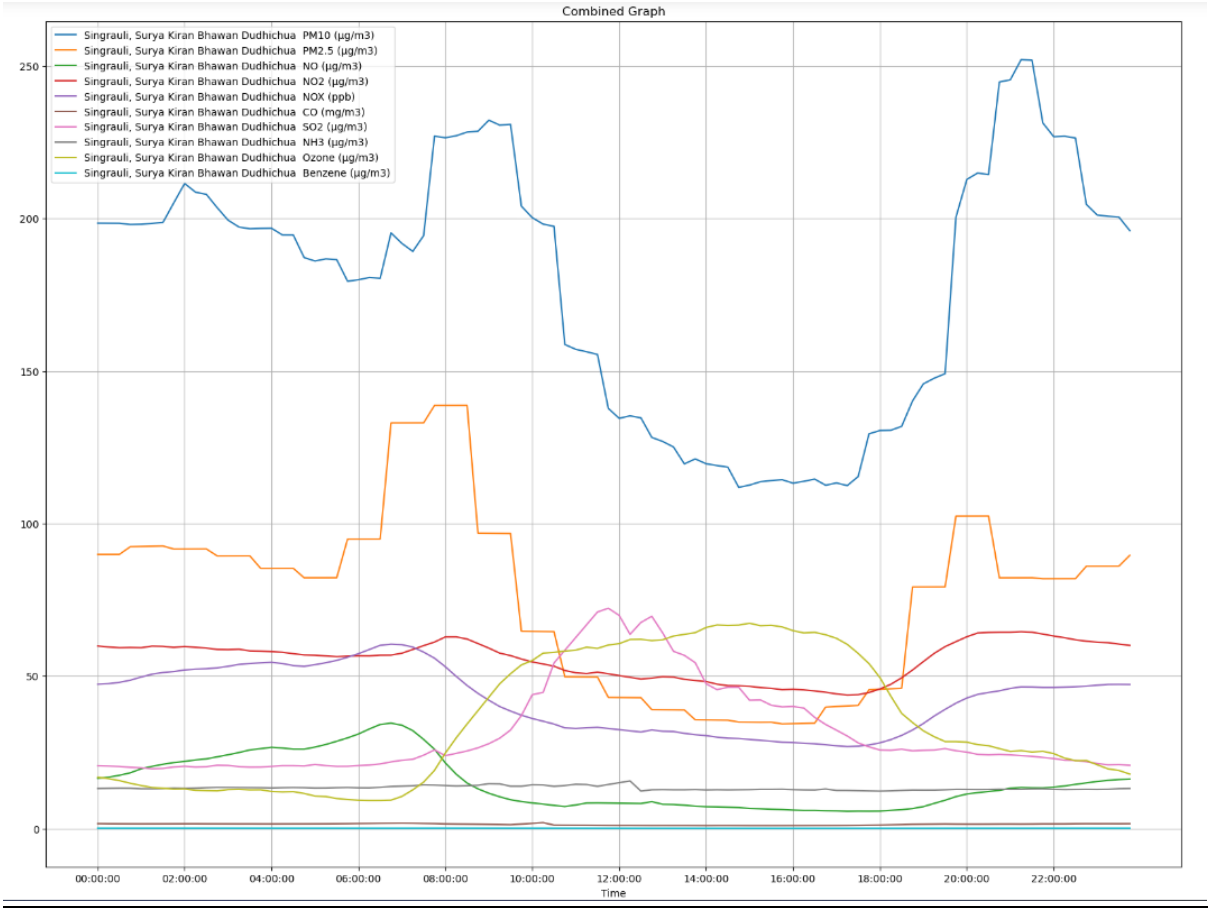

Singrauli, Surya Kiran Bhawan Dudhichua CO (mg/m3)
Singrauli, Surya Kiran Bhawan Dudhichua CO (mg/m3)
2.0
1.8
1.6
1.4
1.2
1.0
00:00:00
02:00:00
04:00:00
06:00:00
08:00:00
10:00:00
12:00:00
14:00:00
16:00:00
18:00:00
20:00:00
22:00:00
Time


Singrauli, Surya Kiran Bhawan Dudhichua SO2 (µg/m3)
Singrauli, Surya Kiran Bhawan Dudhichua SO2 (µg/m3)
70
60
50
40
30
20
00:00:00
02:00:00
04:00:00
06:00:00
08:00:00
10:00:00
12:00:00
14:00:00
16:00:00
18:00:00
20:00:00
22:00:00
Time


Singrauli, Surya Kiran Bhawan Dudhichua NH3 (µg/m3)
Singrauli, Surya Kiran Bhawan Dudhichua NH3 (µg/m3)
15.5
15.0
14.5
14.0
13.5
13.0
12.5
00:00:00
02:00:00
04:00:00
06:00:00
08:00:00
10:00:00
12:00:00
14:00:00
16:00:00
18:00:00
20:00:00
22:00:00
Time


Singrauli, Surya Kiran Bhawan Dudhichua Ozone (µg/m3)
Singrauli, Surya Kiran Bhawan Dudhichua Ozone (µg/m3)
70
60
50
40
30
20
10
00:00:00
02:00:00
04:00:00
06:00:00
08:00:00
10:00:00
12:00:00
14:00:00
16:00:00
18:00:00
20:00:00
22:00:00
Time


Singrauli, Surya Kiran Bhawan Dudhichua Benzene (µg/m3)
Singrauli, Surya Kiran Bhawan Dudhichua Benzene (µg/m3)
0.14
0.13
0.12
0.11
0.10
00:00:00
02:00:00
04:00:00
06:00:00
08:00:00
10:00:00
12:00:00
14:00:00
16:00:00
18:00:00
20:00:00
22:00:00
Time



WHAT DO WE INFER FROM THE ABOVE GRAPHS?

By observing the above graphs, we can find patterns and trends in them. At around 08:00:00 , the majority of the pollutants exhibit a downward slope, indicating a decrease in their conc entrations. This decline could be due to the processes done before the blasting process in w hich there is spraying of water jets and dust suppressants to reduce the concentration of PM and other pollutants. Dust suppressants are substances applied to mitigate the dispersion of pollutants, resulting in a reduction in their quantities in the air.
Between 14:00:00 and 17:00:00, majority of graphs display a relatively stable trend. This tim e period coincides with the time period of open-pit blasting. During blasting, the release of p ollutants from the process overrides the effects of the dust suppressants and water sprayed, rendering their impact less significant. As a result, the pollutant concentrations remain relati vely constant during this time period.
Around 18:00:00, the graphs exhibit an increasing slope, indicating a rise in pollutant levels. This can be attributed to the settled dust particles and chemicals becoming unsettled and m obile once again due to factors such as wind and demographic changes. The re-suspended d ust, combined with other pollutants, leads to an increase in their concentrations.

In summary, the observations from the above graphs illustrate the temporal variations in pol lutant levels resulting from the combined influences of water, dust suppressants, blasting ac tivities, and other environmental factors.

IS THERE ANY TREND OR SEASONALITY IN THE ABOVE DATA?

To find the answer of the above problem I ran the below code snippet to get the statistical v alues to find whether the data given have seasonality or trend.

Code: -

```
from statsmodels.tsa.stattools import adfuller
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
for column in columns:
    print("\n----",column, "----\n")
    result_trend = adfuller(dataSet[column], autolag = 'AIC')
    print("ADF Statistic (Trend):", result_trend[0])
    print("p-value (Trend):", result_trend[1])
    print("Critical Values (Trend):\n", result_trend[4])

    result_seasonality = adfuller(dataSet[column].diff().dropna(), autolag='AIC')
    print("\nADF Statistic (Seasonality):", result_seasonality[0])
    print("p-value (Seasonality):", result_seasonality[1])
    print("Critical Values (Seasonality):\n", result_seasonality[4])
```

Result: -

---- Singrauli, Surya Kiran Bhawan Dudhichua PM10 (µg/m3) ----

```
ADF Statistic (Trend): -9.121219449062005
p-value (Trend): 3.2013749669733968e-15
Critical Values (Trend):
{'1%': -3.4311100803593018, '5%': -2.8618759043136137, '10%': -2.5669487969323943}

ADF Statistic (Seasonality): -22.81205689558395
p-value (Seasonality): 0.0
Critical Values (Seasonality):
{'1%': -3.4311100803593018, '5%': -2.8618759043136137, '10%': -2.5669487969323943}
```

---- Singrauli, Surya Kiran Bhawan Dudhichua PM2.5 (µg/m3) ----

```
ADF Statistic (Trend): -11.308820139207697
p-value (Trend): 1.2537252172855157e-20
Critical Values (Trend):
{'1%': -3.4311100803593018, '5%': -2.8618759043136137, '10%': -2.5669487969323943}

ADF Statistic (Seasonality): -22.492383939614022
p-value (Seasonality): 0.0
Critical Values (Seasonality):
{'1%': -3.4311100803593018, '5%': -2.8618759043136137, '10%': -2.5669487969323943}
```

---- Singrauli, Surya Kiran Bhawan Dudhichua NO (µg/m3) ----

```
ADF Statistic (Trend): -15.277730599551038
p-value (Trend): 4.62304322438328e-28
Critical Values (Trend):
{'1%': -3.4311077024748293, '5%': -2.861874853582867, '10%': -2.566948237620683}

ADF Statistic (Seasonality): -25.305816804767986
p-value (Seasonality): 0.0
Critical Values (Seasonality):
{'1%': -3.4311084934481215, '5%': -2.861875203095297, '10%': -2.5669484236687063}
```

---- Singrauli, Surya Kiran Bhawan Dudhichua NO2 (µg/m3) ----

```
ADF Statistic (Trend): -10.816645346489949
p-value (Trend): 1.8499525756607362e-19
Critical Values (Trend):
```

{'1%': -3.431108053814755, '5%': -2.8618750088316998, '10%': -2.5669483202607704}

ADF Statistic (Seasonality): -20.782590433098875
p-value (Seasonality): 0.0
Critical Values (Seasonality):
{'1%': -3.4311102570927527, '5%': -2.8618759824079, '10%': -2.566948838502564}

---- Singrauli, Surya Kiran Bhawan Dudhichua NOX (ppb) ----

ADF Statistic (Trend): -12.837143779865558
p-value (Trend): 5.7086693795782825e-24
Critical Values (Trend):
{'1%': -3.4311091097922146, '5%': -2.8618754754431404, '10%': -2.566948568641467}

ADF Statistic (Seasonality): -25.183320772483192
p-value (Seasonality): 0.0
Critical Values (Seasonality):
{'1%': -3.4311084934481215, '5%': -2.861875203095297, '10%': -2.5669484236687063}

---- Singrauli, Surya Kiran Bhawan Dudhichua CO (mg/m3) ----

ADF Statistic (Trend): -10.027782087337366
p-value (Trend): 1.6148117194937135e-17
Critical Values (Trend):
{'1%': -3.4311076146907284, '5%': -2.861874814793139, '10%': -2.5669482169726283}

ADF Statistic (Seasonality): -18.248402989134863
p-value (Seasonality): 2.3460195741433872e-30
Critical Values (Seasonality):
{'1%': -3.4311102570927527, '5%': -2.8618759824079, '10%': -2.566948838502564}

---- Singrauli, Surya Kiran Bhawan Dudhichua SO2 (µg/m3) ----

ADF Statistic (Trend): -17.429106443278652
p-value (Trend): 4.778186924950188e-30
Critical Values (Trend):
{'1%': -3.4311074391835352, '5%': -2.8618747372406377, '10%': -2.5669481756908668}

ADF Statistic (Seasonality): -21.236824805111628
p-value (Seasonality): 0.0
Critical Values (Seasonality):
{'1%': -3.4311102570927527, '5%': -2.8618759824079, '10%': -2.566948838502564}

---- Singrauli, Surya Kiran Bhawan Dudhichua NH3 (µg/m3) ----

ADF Statistic (Trend): -3.1619872160575313
p-value (Trend): 0.02229316792019156
Critical Values (Trend):
{'1%': -3.4311099037080006, '5%': -2.8618758262556225, '10%': -2.566948755381546}

ADF Statistic (Seasonality): -22.124780252647
p-value (Seasonality): 0.0
Critical Values (Seasonality):
{'1%': -3.4311099037080006, '5%': -2.8618758262556225, '10%': -2.566948755381546}

---- Singrauli, Surya Kiran Bhawan Dudhichua Ozone (µg/m3) ----

ADF Statistic (Trend): -20.908288586331228
p-value (Trend): 0.0
Critical Values (Trend):
{'1%': -3.4311099920233863, '5%': -2.8618758652800826, '10%': -2.566948776154556}

ADF Statistic (Seasonality): -19.068012935111845
p-value (Seasonality): 0.0
Critical Values (Seasonality):
{'1%': -3.4311102570927527, '5%': -2.8618759824079, '10%': -2.566948838502564}

---- Singrauli, Surya Kiran Bhawan Dudhichua Benzene (µg/m3) ----

```
ADF Statistic (Trend): -9.064518331267331
p-value (Trend): 4.469950691074848e-15
Critical Values (Trend):
{'1%': -3.431110168715755, '5%': -2.8618759433562184, '10%': -2.5669488177150632}

ADF Statistic (Seasonality): -19.504808351025634
p-value (Seasonality): 0.0
Critical Values (Seasonality):
{'1%': -3.431110168715755, '5%': -2.8618759433562184, '10%': -2.5669488177150632}
```

A brief about ADF Test:

Augmented Dickey-Fuller (ADF) test, which is a statistical test used in econometrics to deter mine whether a time series has a unit root. The ADF test is commonly employed to test for t he presence of a trend in a time series data and assess its stationarity.

The results of an Augmented Dickey-Fuller (ADF) test provide insights into the stationarity of a time series and the presence of a unit root. The test generates a test statistic and critical va lues, which are used to determine whether the time series is stationary or exhibits a unit roo t.

- Rejecting the null hypothesis: If the test statistic is less than the critical value, the null hypot hesis is rejected. This suggests that the time series is stationary, indicating the absence of a u nit root. In this case, it implies that the time series does not have a long-term trend.

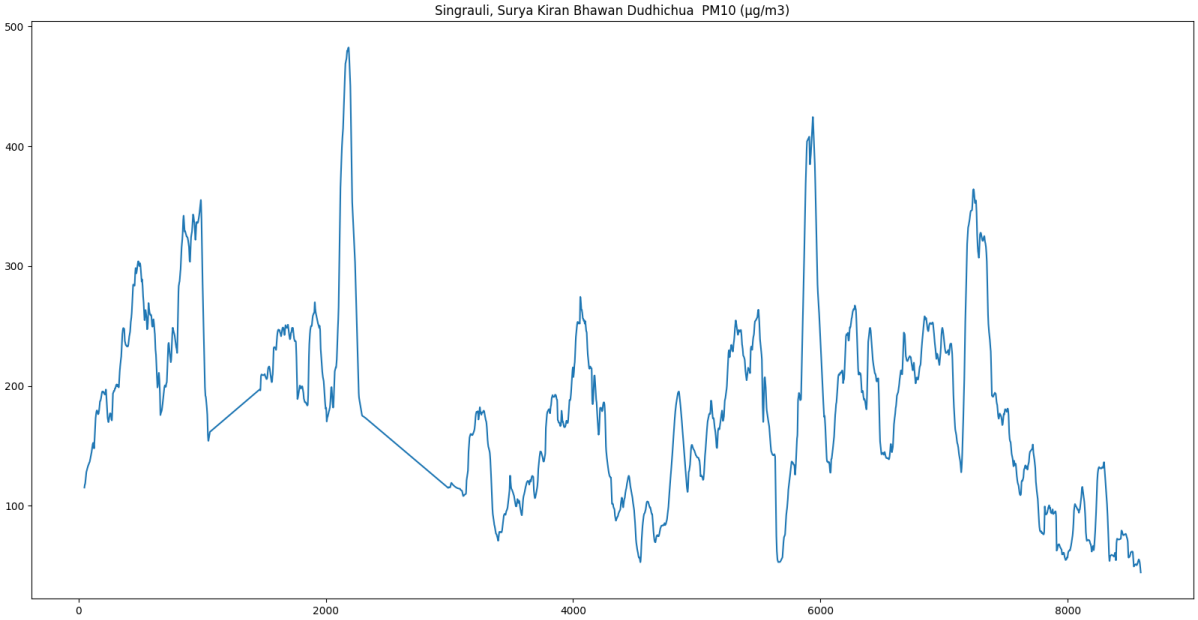
- Failing to reject the null hypothesis: If the test statistic is greater than the critical value, the n ull hypothesis is not rejected. This indicates that the time series is non-stationary, implying t he presence of a unit root. In other words, the time series likely exhibits a long-term trend.

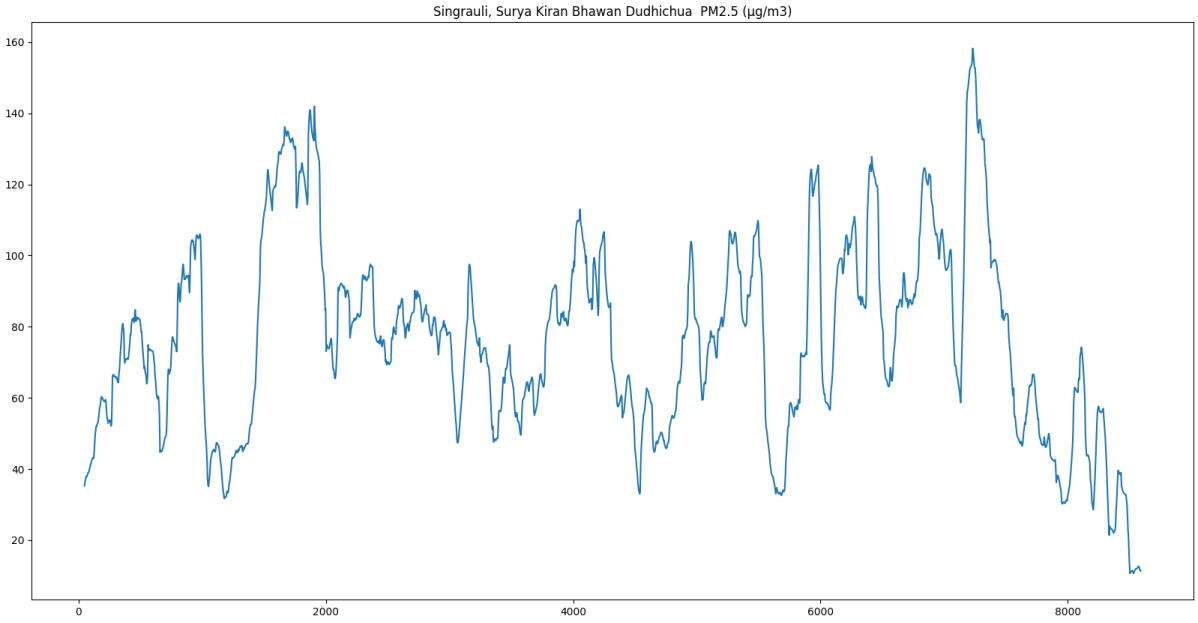
- Inconclusive results: If the test statistic is close to the critical value, the results may be inconc lusive, and it becomes difficult to determine whether the time series is stationary or non-sta tionary. In such cases, further analysis or alternative tests may be required to obtain conclusi ve results.

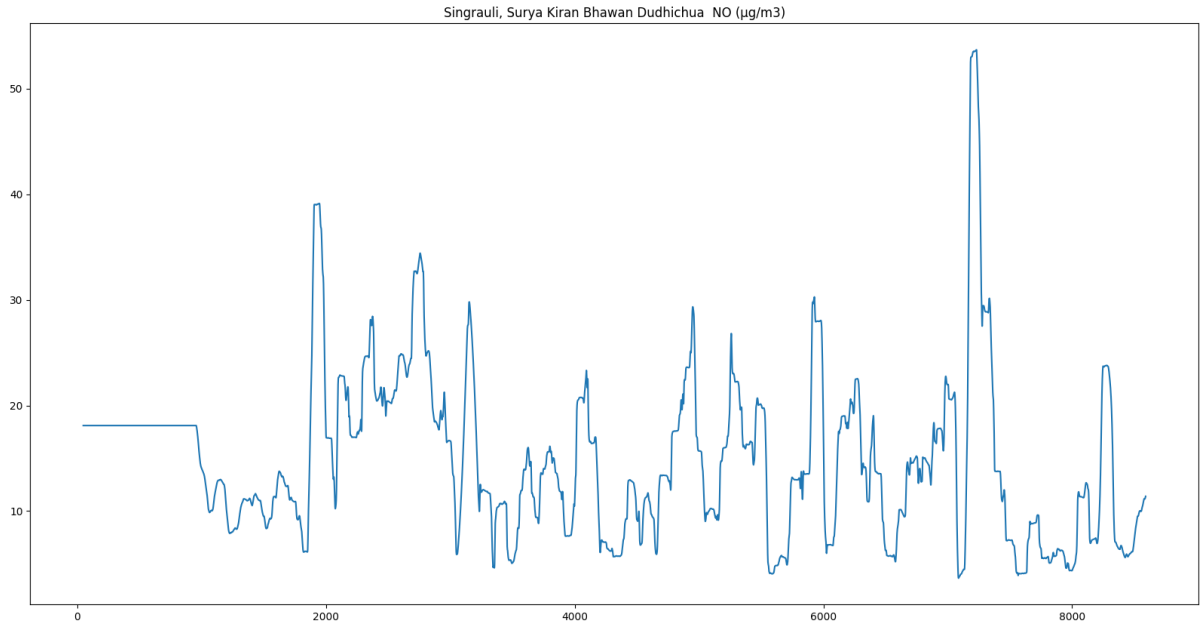
From the above **result** and after knowing how the test works, I come to the conclusion that g iven data doesn't have trend and seasonality because the ADF statistics of all pollutants is m uch below than the critical values of 1%, 5% and 10% significance levels. Also, the p-value of all pollutants is much below 0.05 meaning that data doesn't possesses trend or seasonality.

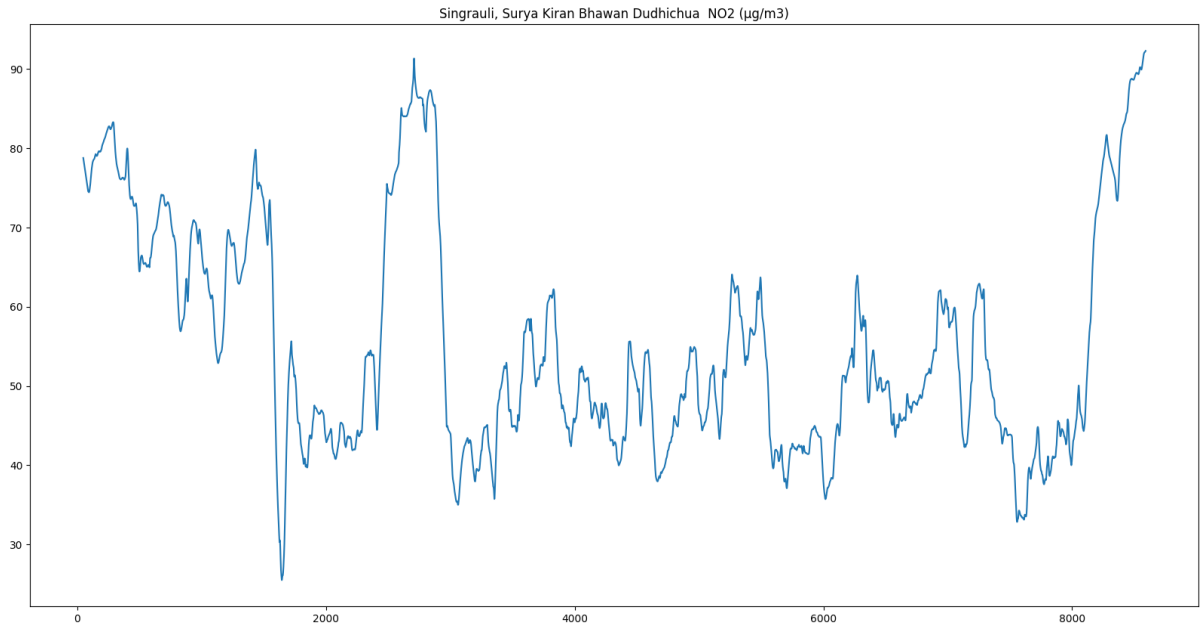
We can also observe the trend plot of pollutants.

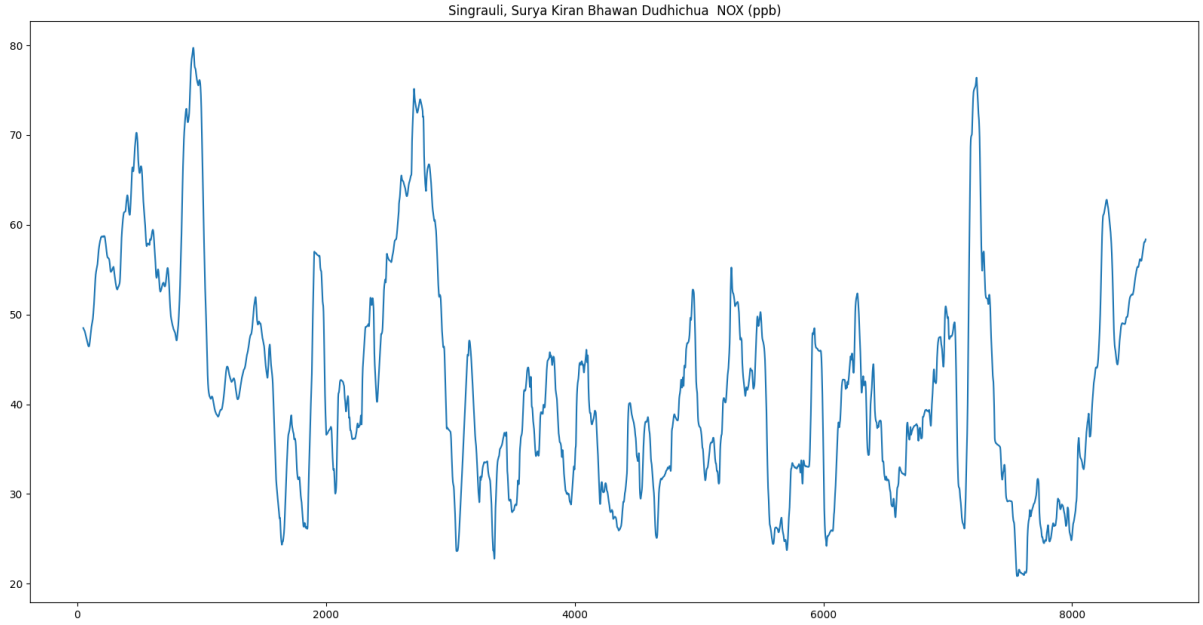
Trend plot of pollutants: -

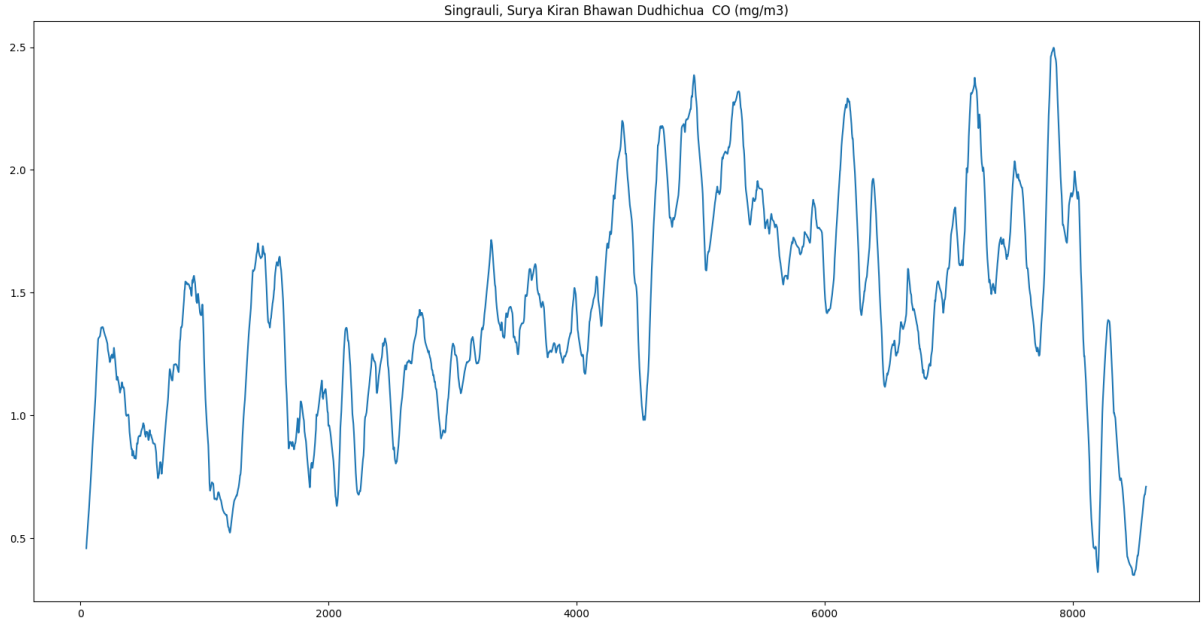

Singrauli, Surya Kiran Bhawan Dudhichua PM10 (µg/m3)
500
400
300
200
100
0
2000
4000
6000
8000

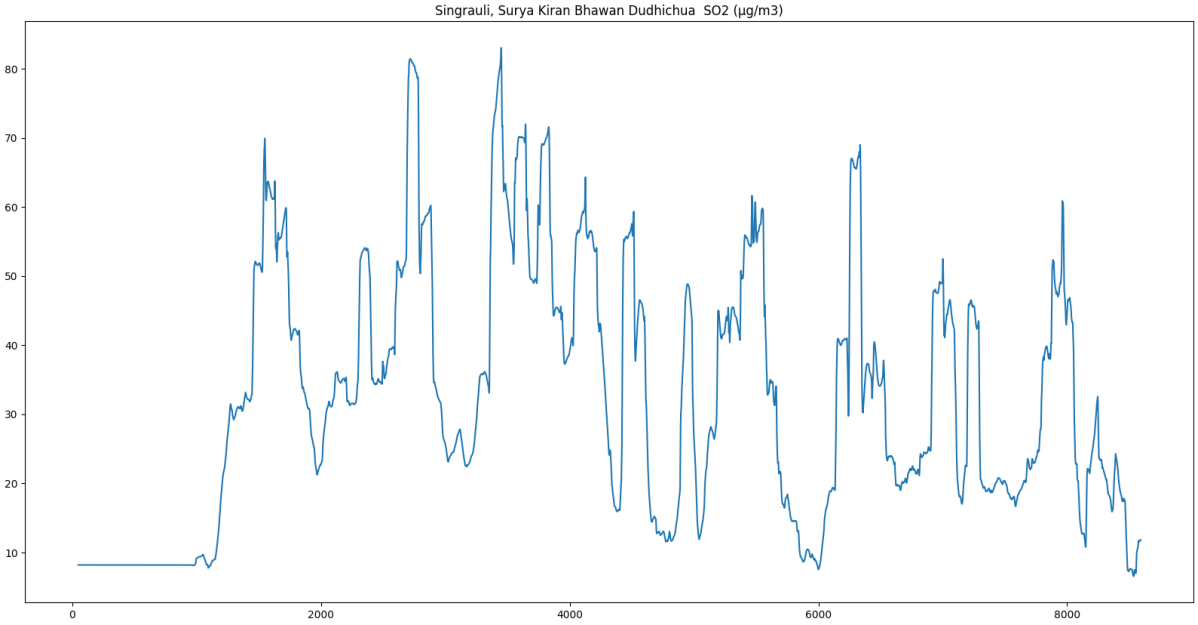

Singrauli, Surya Kiran Bhawan Dudhichua PM2.5 (µg/m3)
160
140
120
100
80
60
40
20
0
2000
4000
6000
8000

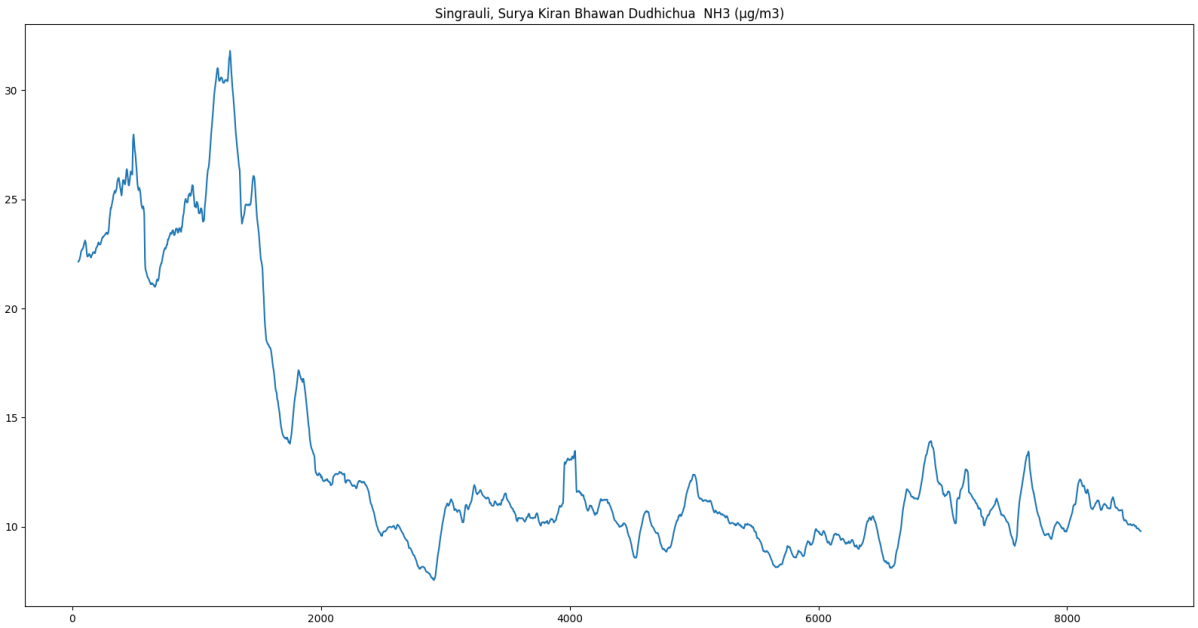

Singrauli, Surya Kiran Bhawan Dudhichua NO (µg/m3)
50
40
30
20
10
0
2000
4000
6000
8000

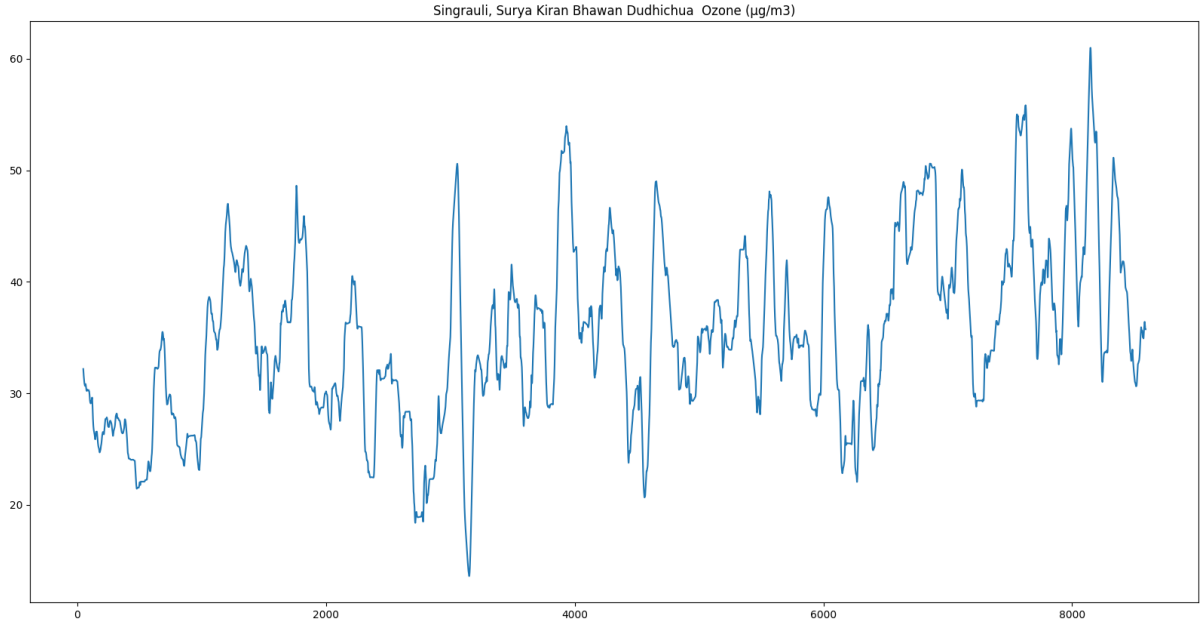

Singrauli, Surya Kiran Bhawan Dudhichua NO2 (µg/m3)
90
80
70
60
50
40
30
0
2000
4000
6000
8000

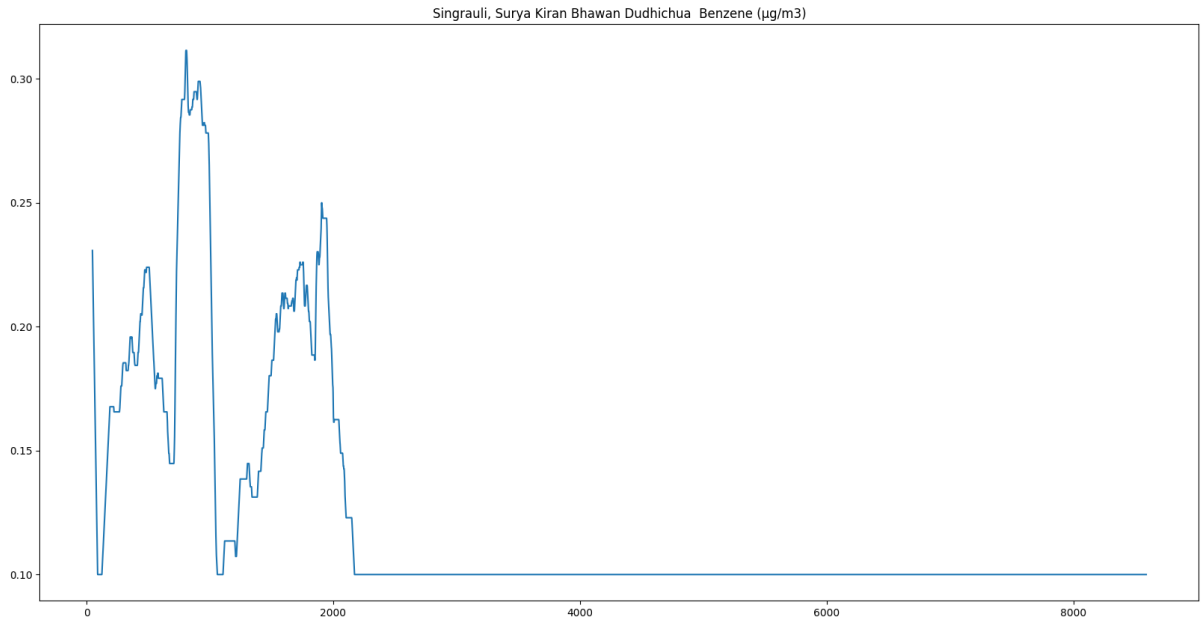

Singrauli, Surya Kiran Bhawan Dudhichua NOX (ppb)
80
70
60
50
40
30
20
0
2000
4000
6000
8000


Singrauli, Surya Kiran Bhawan Dudhichua CO (mg/m3)
2.5
2.0
1.5
1.0
0.5
0
2000
4000
6000
8000


Singrauli, Surya Kiran Bhawan Dudhichua SO2 (µg/m3)
80
70
60
50
40
30
20
10
0
2000
4000
6000
8000


Singrauli, Surya Kiran Bhawan Dudhichua NH3 (µg/m3)
30
25
20
15
10
0
2000
4000
6000
8000





After observing the trend plots, my conclusion is that there is no trend in the data of polluta nts.

<u>Do Descriptive analysis can be categorized into four types which are measures of frequency, centr al tendency, dispersion or variation, and position of air pollution data?</u>

Yes, descriptive analysis can be categorized into four types when analysing air pollution data. These categories are measures of frequency, measures of central tendency, measures of dis persion or variation, and measures of position.

- **Measures of Frequency:** These measures provide information about the occurrence or frequ ency of different values or categories within the air pollution data. Common measures of fre quency include counts, frequencies, percentages, and proportions. They help to understand the distribution of different pollution levels or categories.

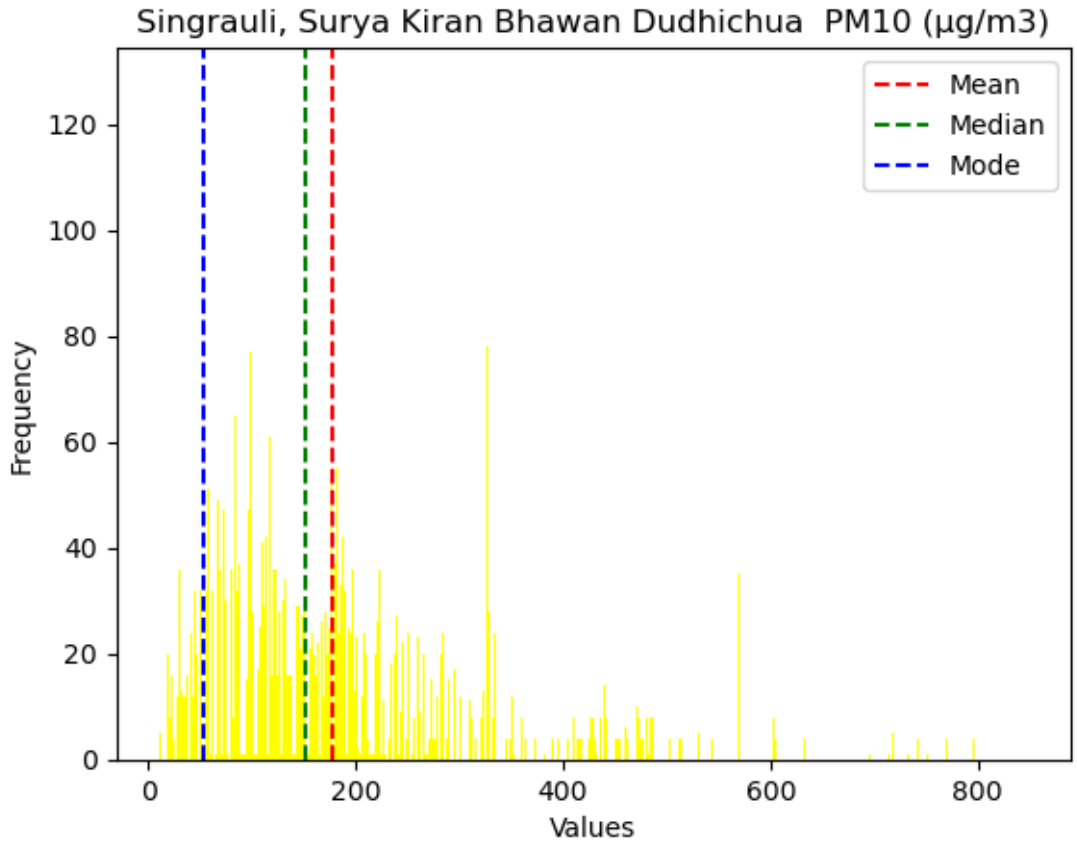
- **Measures of Central Tendency:** These measures provide information about the central or ty pical value of the air pollution data. The three main measures of central tendency are:

  Mean: It is the arithmetic average and is calculated by summing all the values and dividing b y the total number of observations. The mean gives an idea of the overall average pollution l evel.

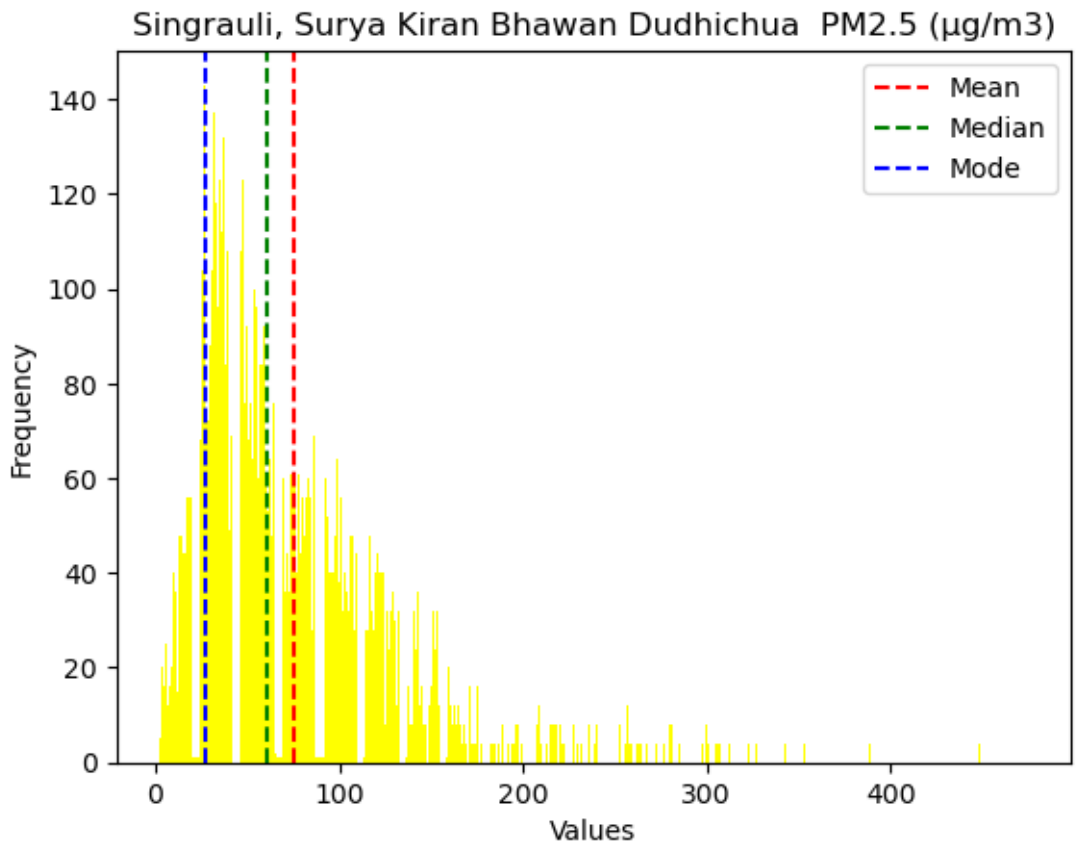
  Median: It is the middle value in a sorted list of observations. It represents the value below which 50% of the data falls. The median is less influenced by extreme values and provides a measure of the central tendency.

  Mode: It represents the most frequently occurring value in the data. It is useful for identifyin g the most common pollution level or category.

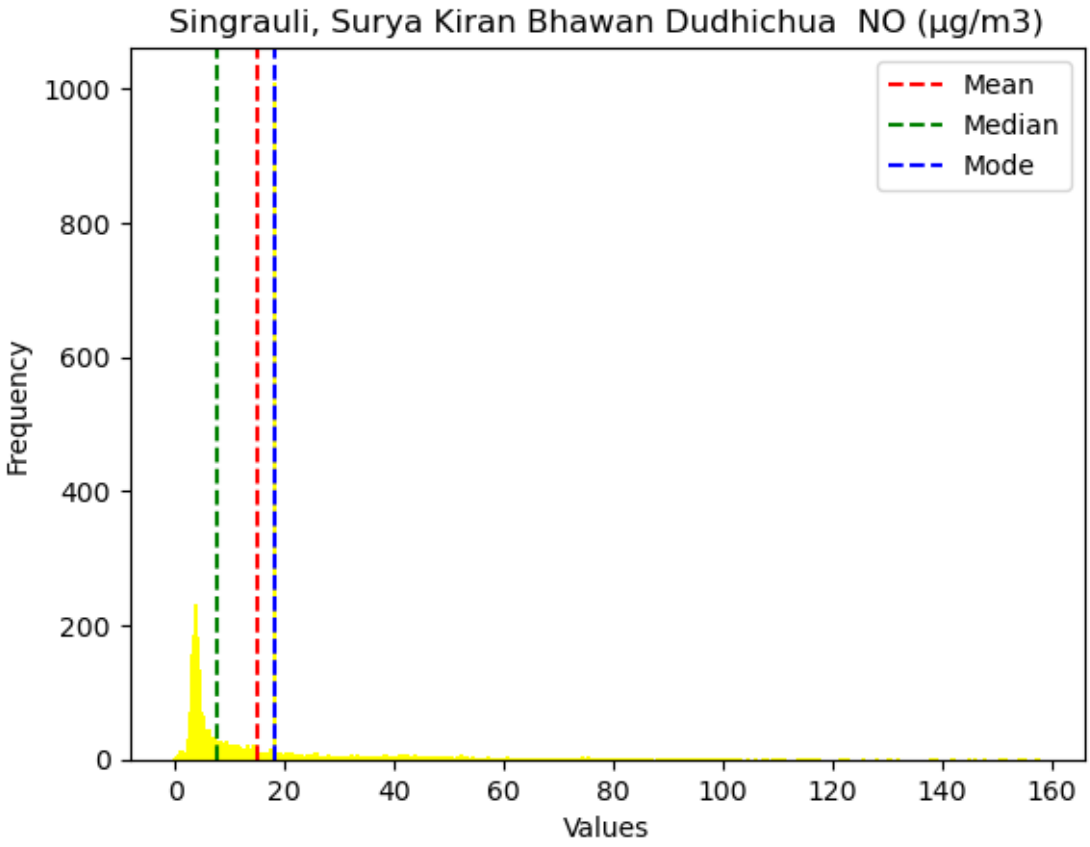
  Below are the frequency distribution graphs of pollutants and the mean, mode and median o f their data is marked on the graph.



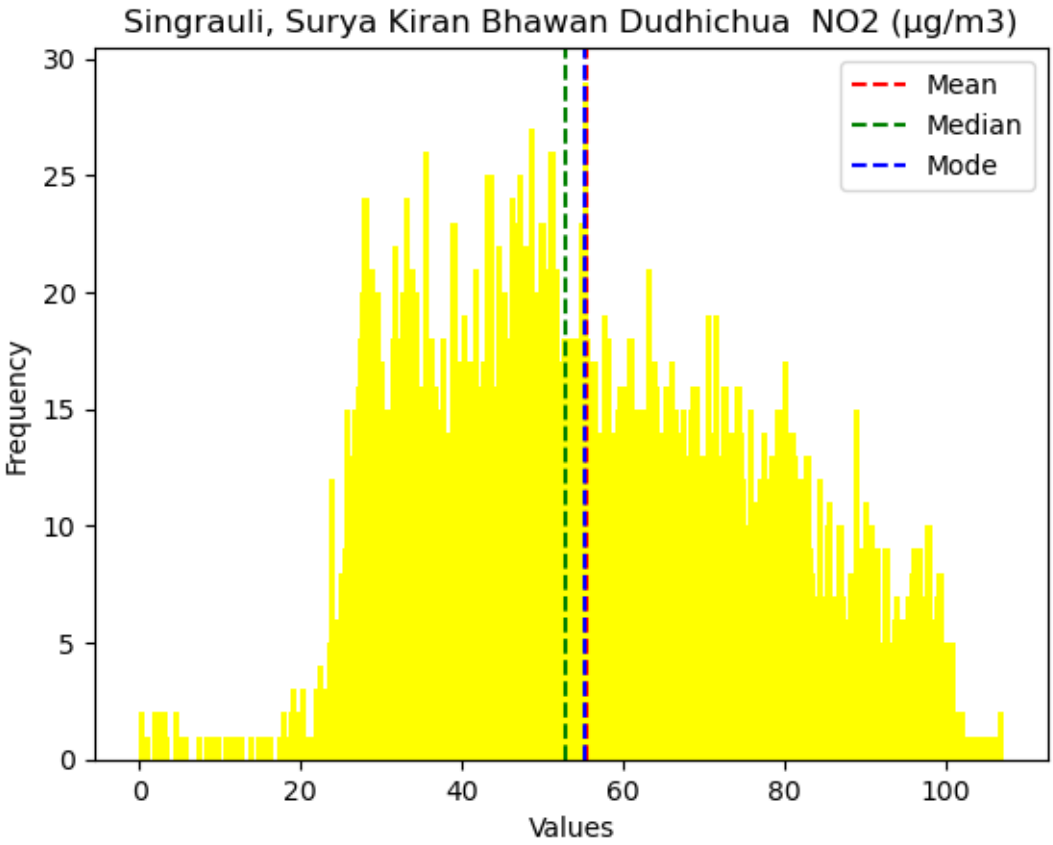

```
Mean = 177.52185699409927 , Median = 151.94444444444446 , Mode = 53.0
```



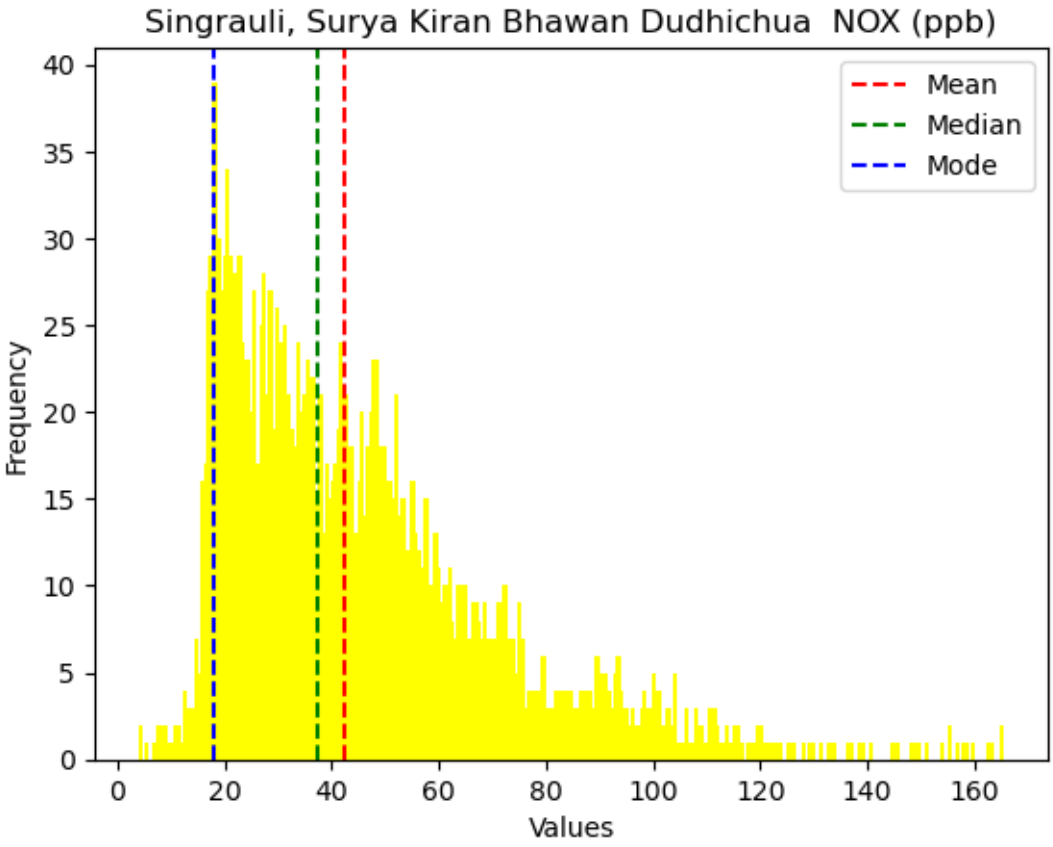
Mean = 75.59506999884299 , Median = 61.0 , Mode = 27.0



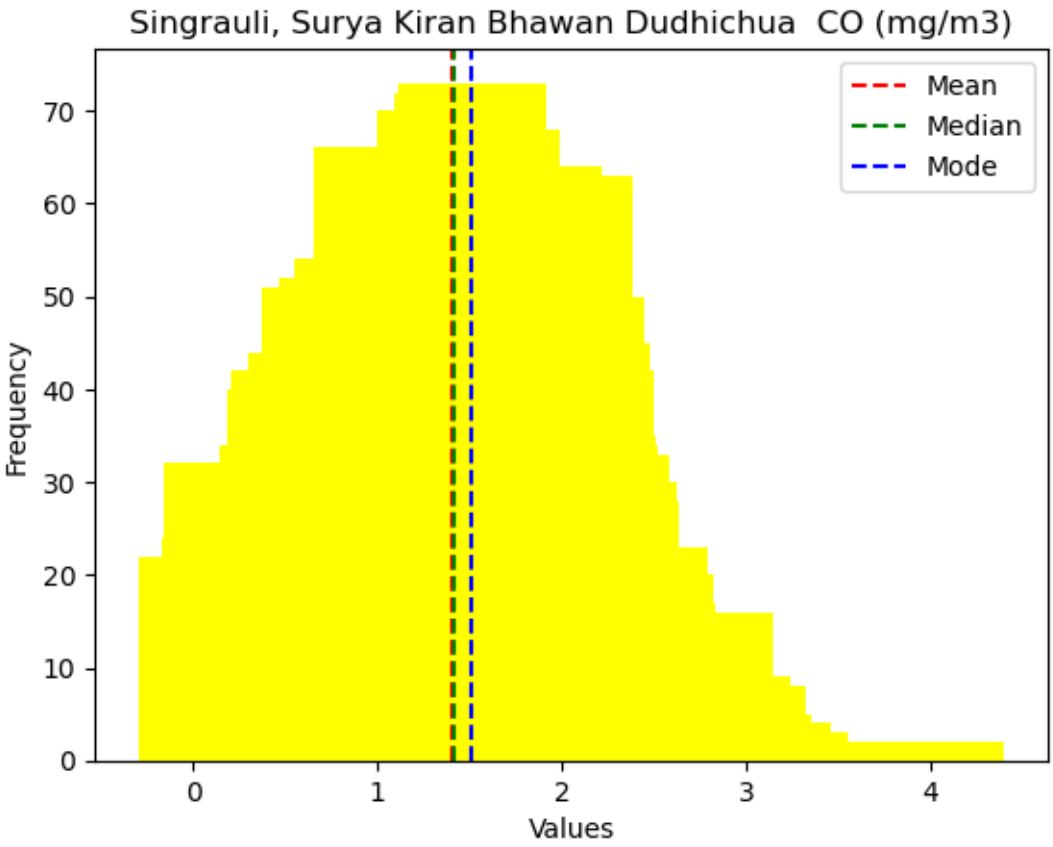
Mean = 14.95495198426471 , Median = 7.5 , Mode = 18.1



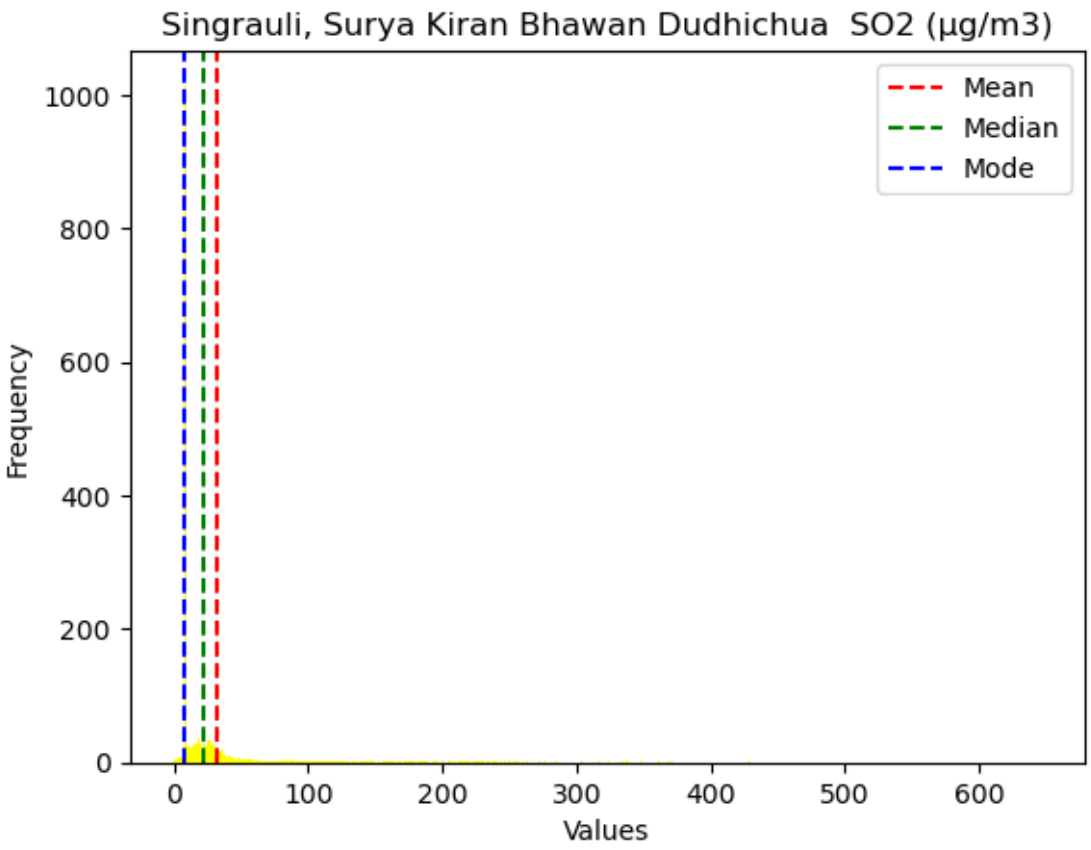
Mean = 55.43029156542875 , Median = 52.8 , Mode = 55.3



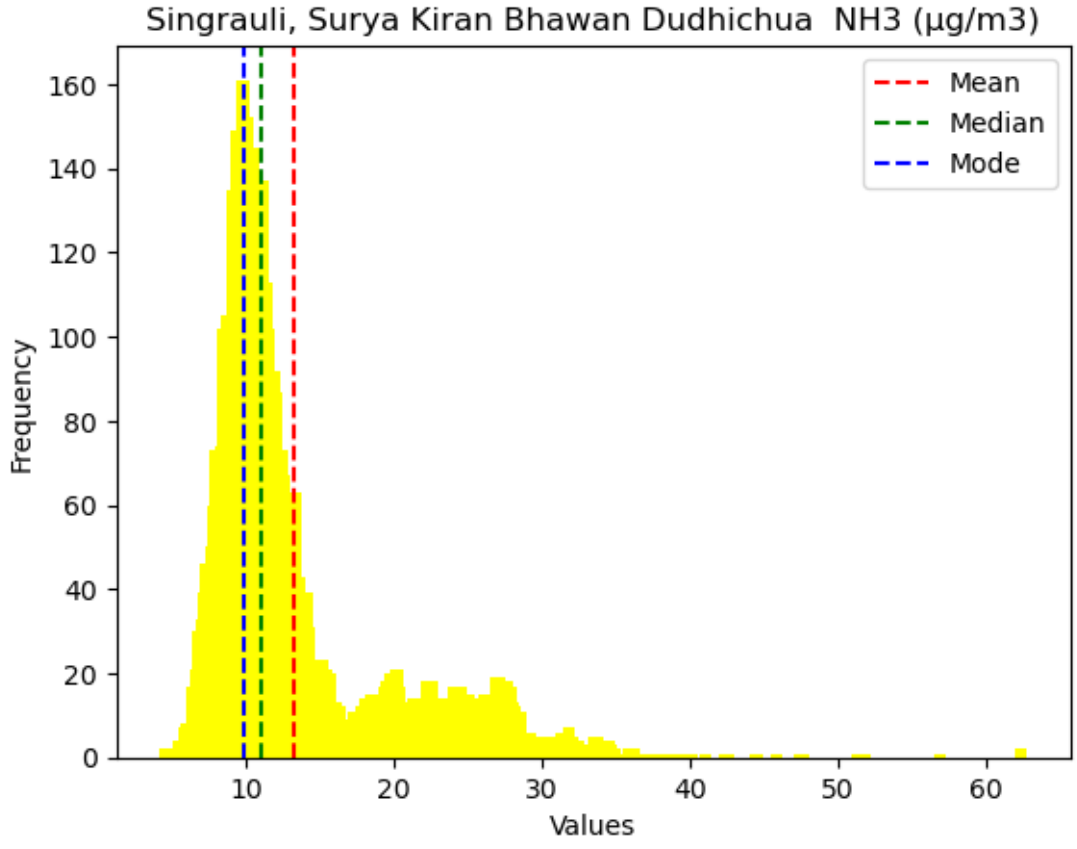
Mean = 42.33864630336682 , Median = 37.5 , Mode = 18.1



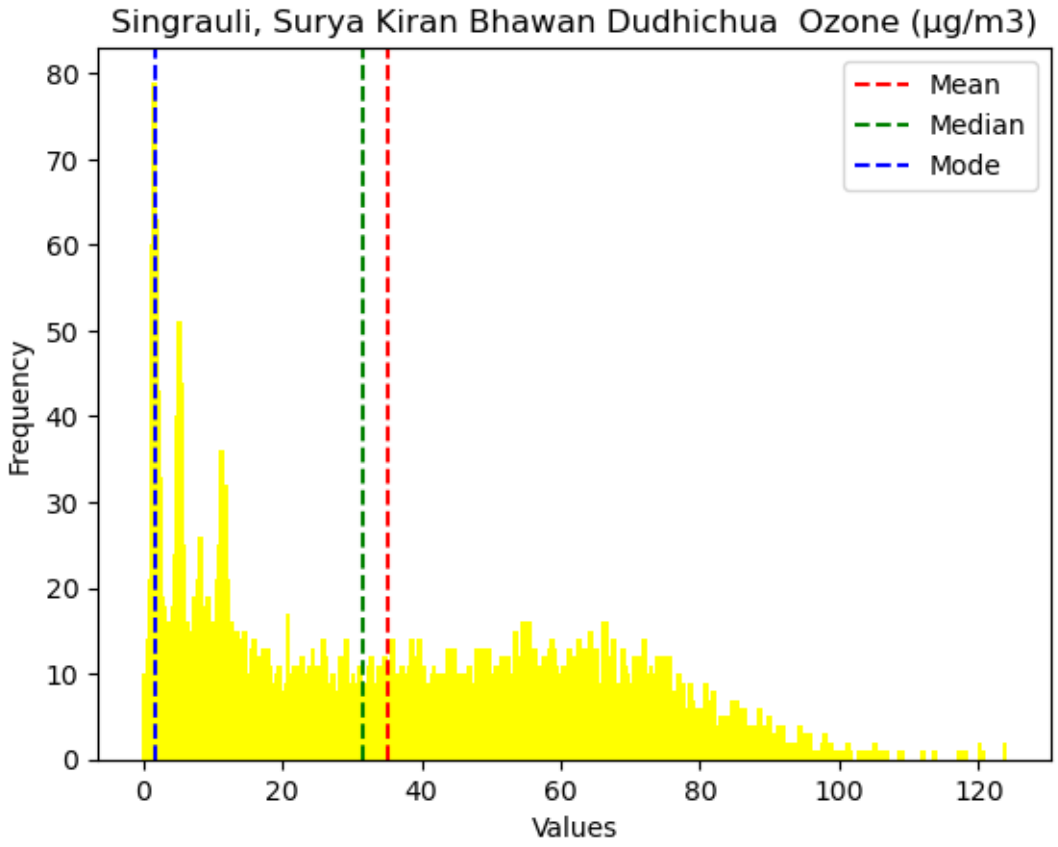
Mean = 1.402049635543208 , Median = 1.41 , Mode = 1.51



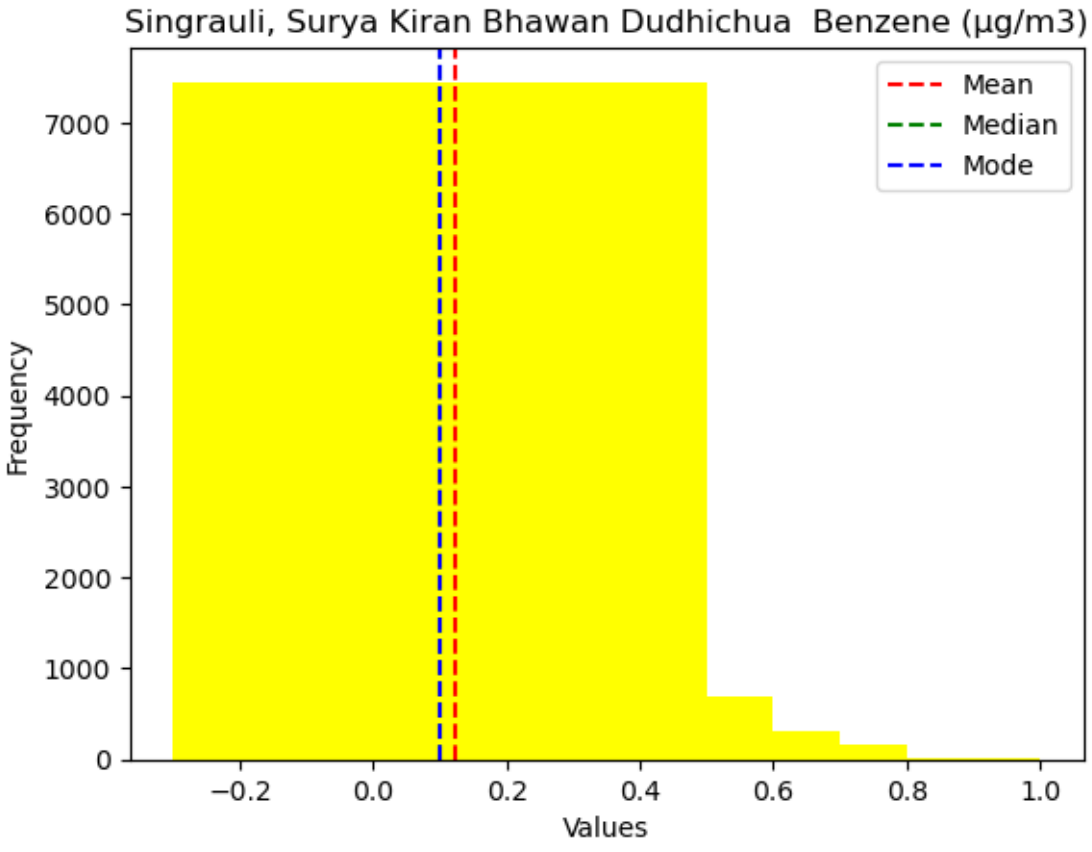
Mean = 31.9903135485365 , Median = 22.8 , Mode = 8.2



Mean = 13.29162790697677 , Median = 11.0 , Mode = 9.8



Mean = 35.200211732037424 , Median = 31.7 , Mode = 1.7



```
Mean = 0.12206178410275549 , Median = 0.1 , Mode = 0.1
```

- **Measures of Dispersion or Variation:** These measures provide information about the spread or variability of the air pollution data. They help assess how much the data values deviate fro m the central tendency. Common measures of dispersion include:

  Range: It is the difference between the maximum and minimum values in the dataset. It pro vides an idea of the overall spread of pollution levels.

  Variance: It measures the average squared deviation of each data point from the mean. It q uantifies the overall variability of the data.

  Standard Deviation: It is the square root of the variance and represents the average amount by which data points deviate from the mean. It provides a measure of dispersion that is in th e same units as the original data.

Singrauli, Surya Kiran Bhawan Dudhichua PM10 (µg/m3) -

```
Range = 835.0
Variance = 15616.270302349527
Standard Deviation = 124.96507633074741
```

Singrauli, Surya Kiran Bhawan Dudhichua PM2.5 (µg/m3) -

```
Range = 471.0
Variance = 3023.5682353370844
Standard Deviation = 54.98698241708745
```

Singrauli, Surya Kiran Bhawan Dudhichua NO (µg/m3) -

```
Range = 157.4
Variance = 321.3043064520576
Standard Deviation = 17.92496322038228
```

Singrauli, Surya Kiran Bhawan Dudhichua NO2 (µg/m3) -

```
Range = 106.7
Variance = 408.73398088196217
Standard Deviation = 20.217170446973093
```

Singrauli, Surya Kiran Bhawan Dudhichua NOX (ppb) -

```
Range = 161.0
Variance = 493.69391594600773
Standard Deviation = 22.219224017638595
```

Singrauli, Surya Kiran Bhawan Dudhichua CO (mg/m3) -

```
Range = 3.9
Variance = 0.4016429823988621
Standard Deviation = 0.6337530926148306
```

Singrauli, Surya Kiran Bhawan Dudhichua SO2 (µg/m3) -

```
Range = 645.5
Variance = 1566.2218029429519
Standard Deviation = 39.57552024854445
```

Singrauli, Surya Kiran Bhawan Dudhichua NH3 (µg/m3) -

```
Range = 57.8
Variance = 38.36001130834236
Standard Deviation = 6.193545939794292
```

Singrauli, Surya Kiran Bhawan Dudhichua Ozone (µg/m3) -

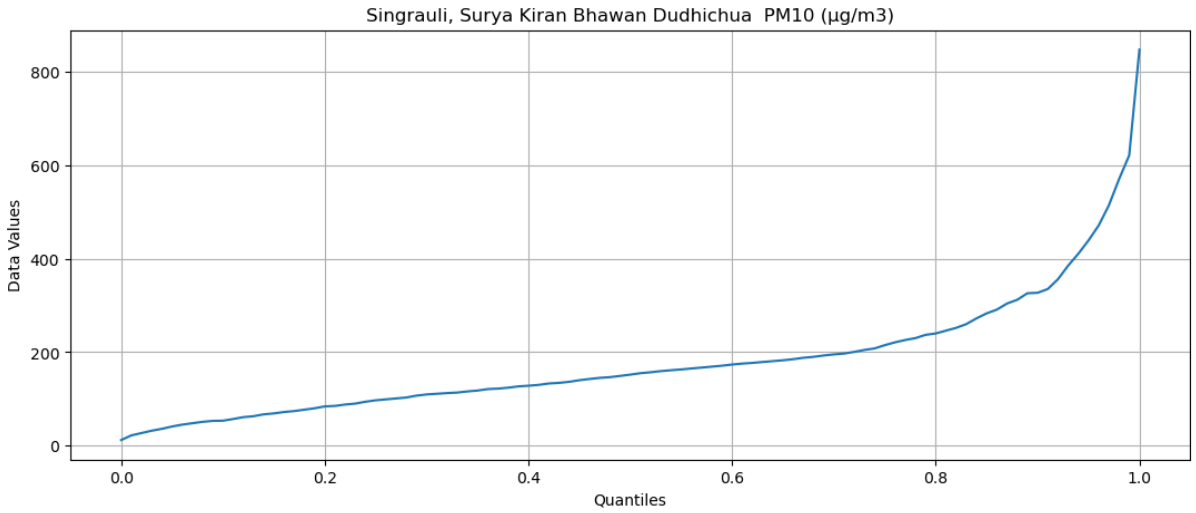
```
Range = 123.7
Variance = 722.598617600535
Standard Deviation = 26.881194497278855
```

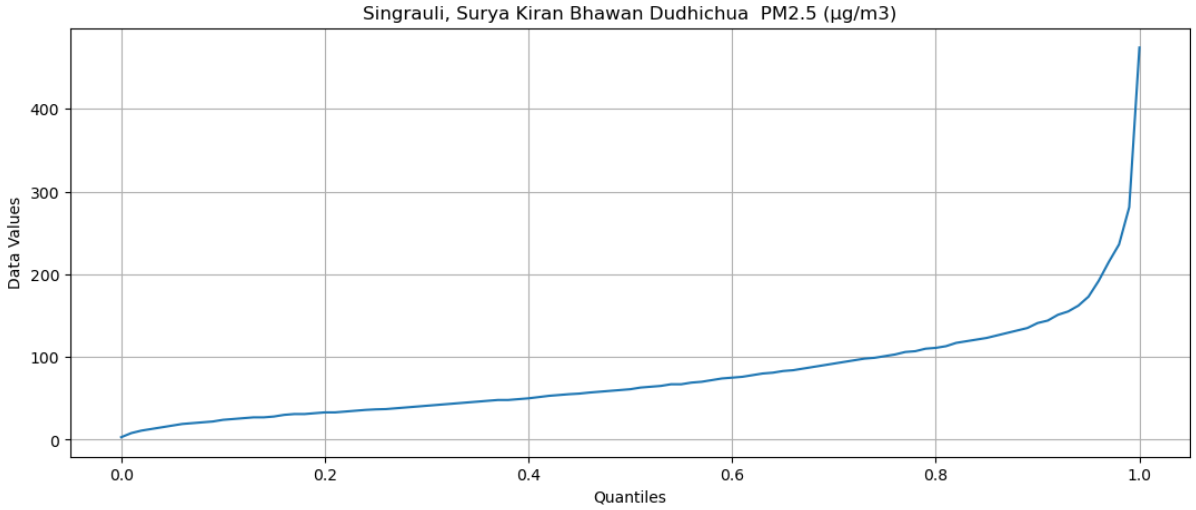
Singrauli, Surya Kiran Bhawan Dudhichua Benzene (µg/m3) -

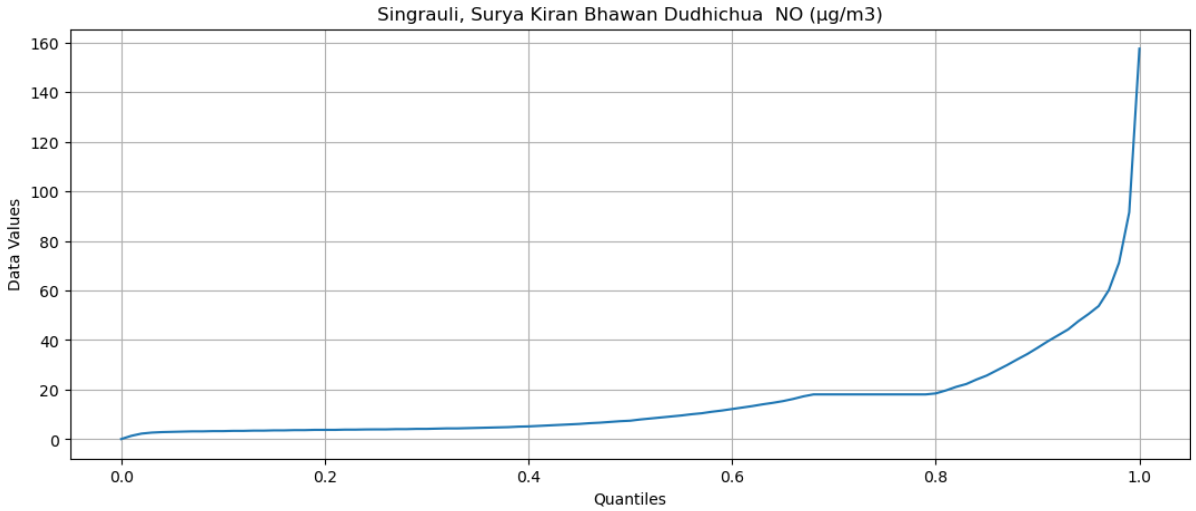
```
Range = 0.5
Variance = 0.004015247713919568
Standard Deviation = 0.06336598230848764
```

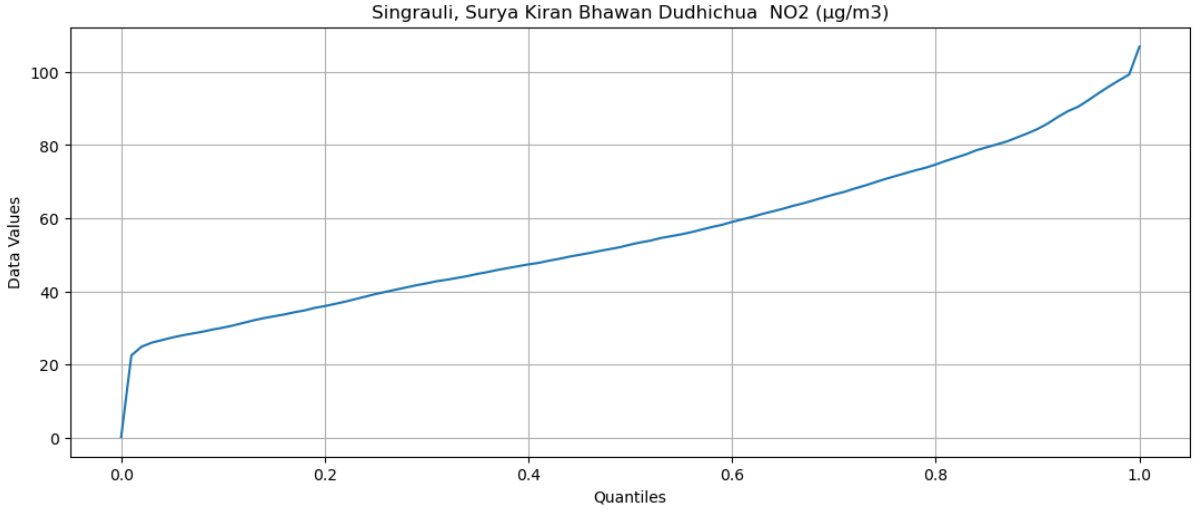
- **Measures of Position:** These measures provide information about the relative position of sp ecific values within the air pollution data. They help identify the percentile or rank of a partic ular observation.

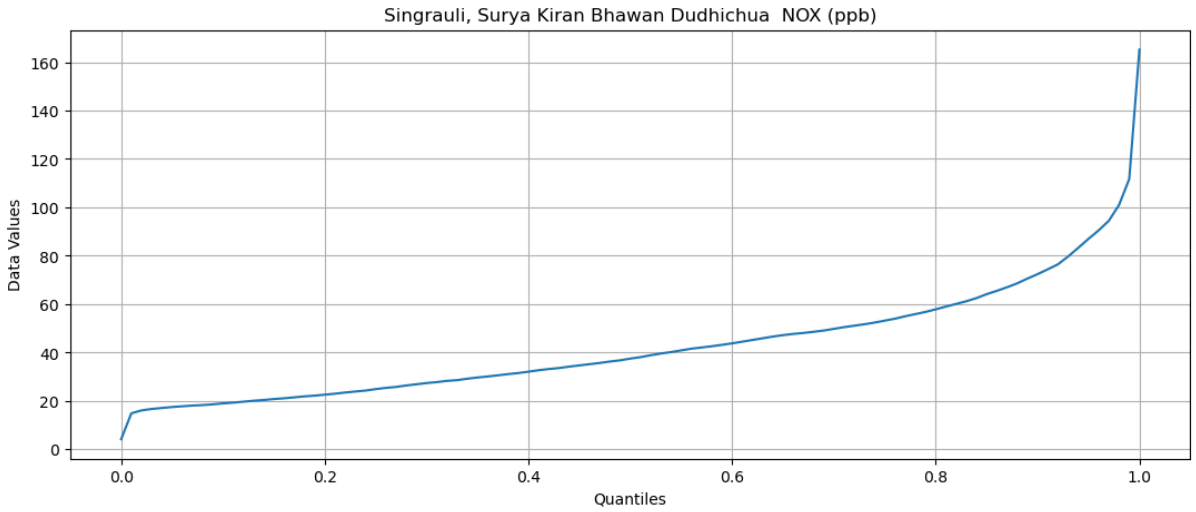
Quantile plot of pollutants: -

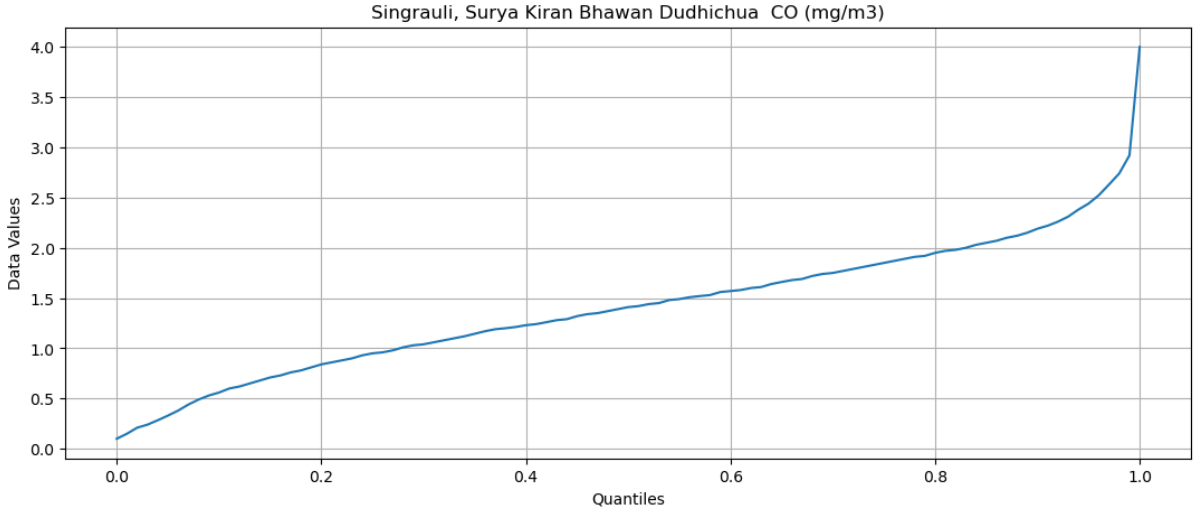


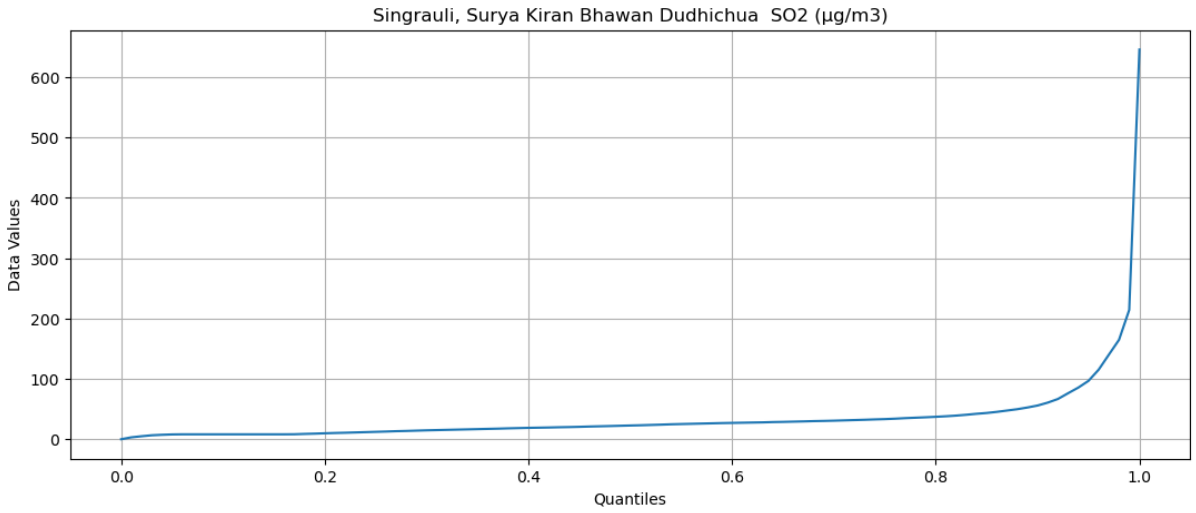

Singrauli, Surya Kiran Bhawan Dudhichua PM2.5 (µg/m3)
Data Values
Quantiles
0
100
200
300
400
0.0
0.2
0.4
0.6
0.8
1.0

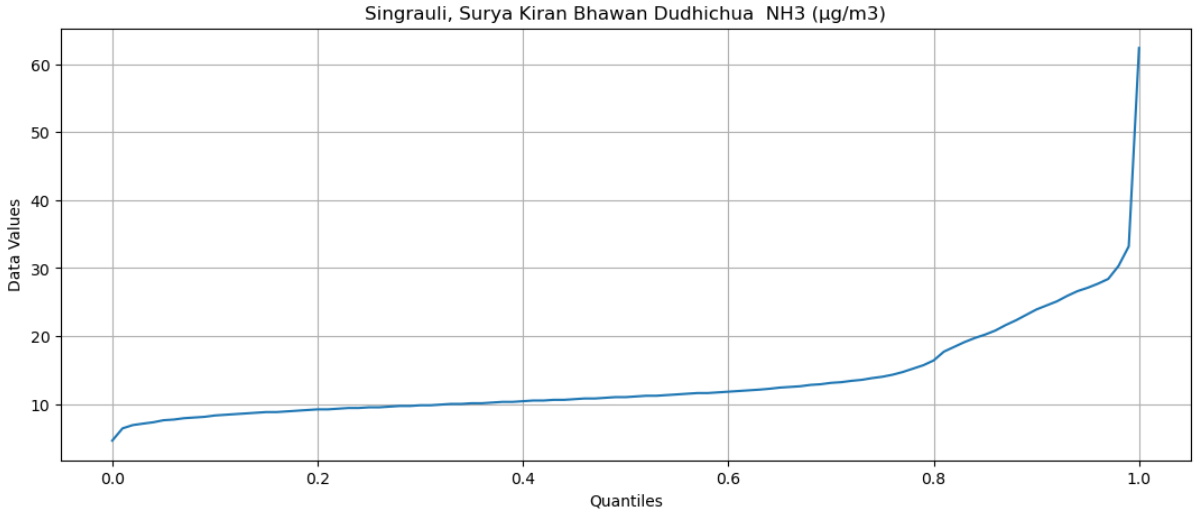

Singrauli, Surya Kiran Bhawan Dudhichua NO (µg/m3)
Data Values
Quantiles
0
20
40
60
80
100
120
140
160
0.0
0.2
0.4
0.6
0.8
1.0

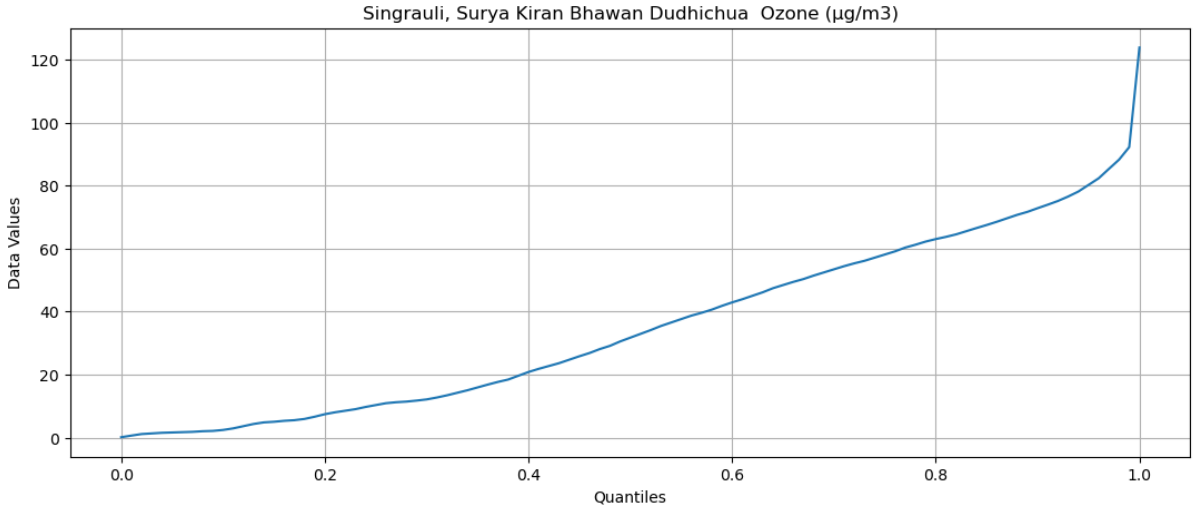

Singrauli, Surya Kiran Bhawan Dudhichua NO2 (µg/m3)
Data Values
Quantiles
0
20
40
60
80
100
0.0
0.2
0.4
0.6
0.8
1.0

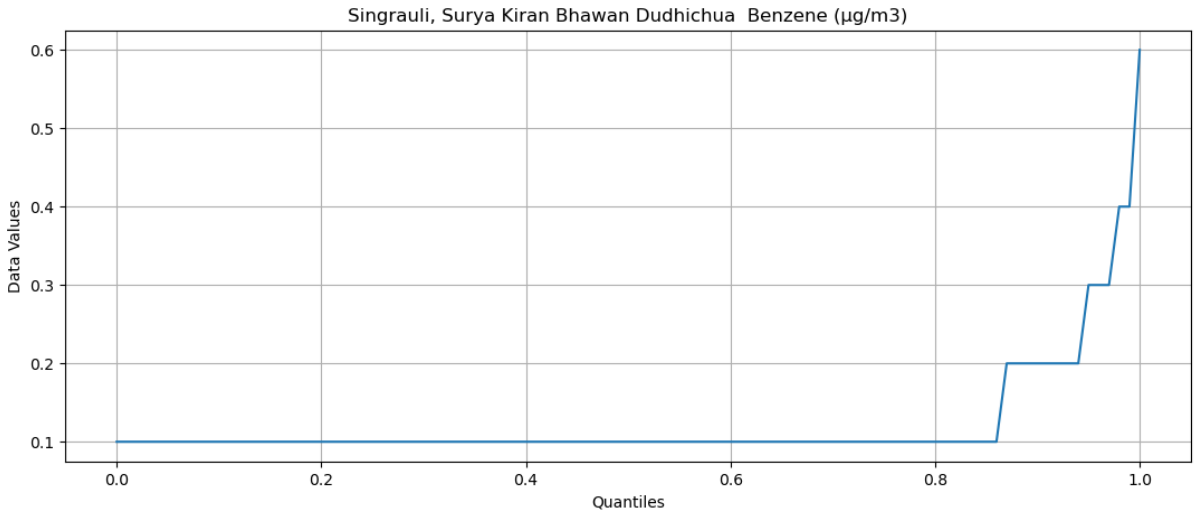

Singrauli, Surya Kiran Bhawan Dudhichua NOX (ppb)
Data Values
Quantiles
0
20
40
60
80
100
120
140
160
0.0
0.2
0.4
0.6
0.8
1.0


Singrauli, Surya Kiran Bhawan Dudhichua CO (mg/m3)
Data Values
Quantiles
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
0.0
0.2
0.4
0.6
0.8
1.0


Singrauli, Surya Kiran Bhawan Dudhichua SO2 (µg/m3)
Data Values
Quantiles
0
100
200
300
400
500
600
0.0
0.2
0.4
0.6
0.8
1.0



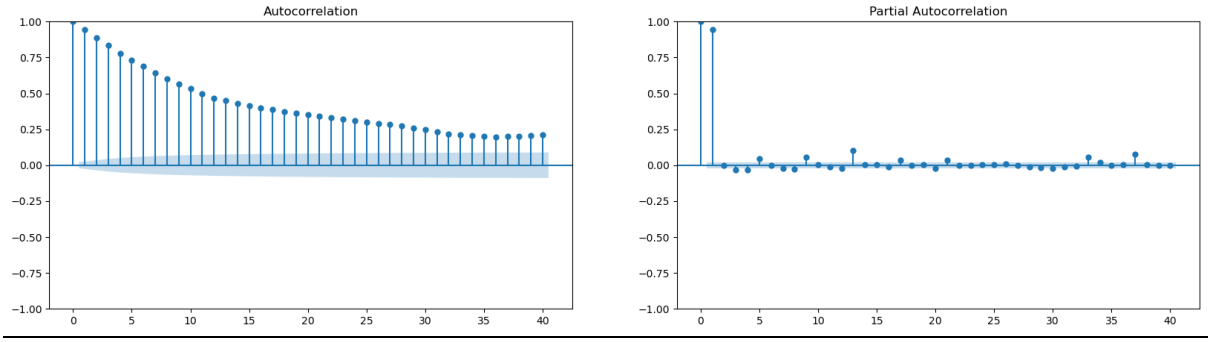




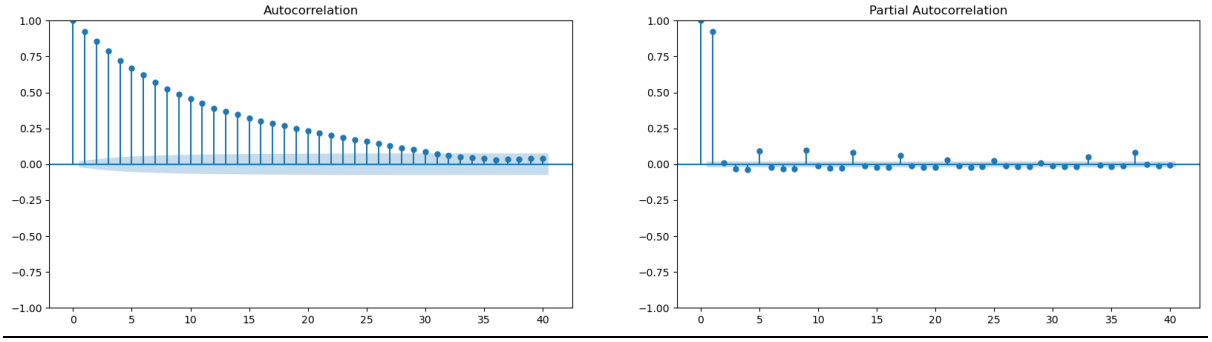
AUTOCORRELATION AND PARTIAL AUTO CORRELATION OF DATA

- **Autocorrelation Function (ACF):** ACF measures the correlation between the current observa tion and past observations at various lags. Look for significant spikes or decaying patterns in the ACF plot.
- **Partial Autocorrelation Function (PACF):** PACF measures the correlation between the curren t observation and past observations, removing the effects explained by intervening lags. Loo k for significant spikes or decaying patterns in the PACF plot.

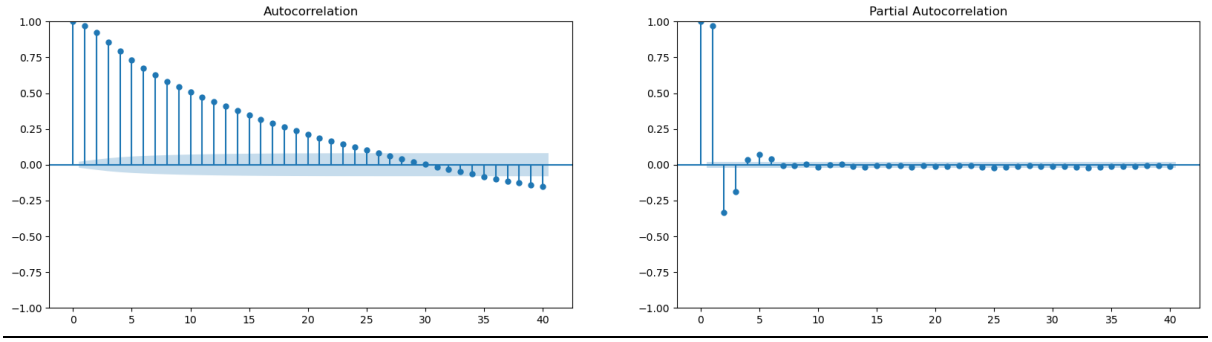
Singrauli, Surya Kiran Bhawan Dudhichua PM10 (µg/m3)



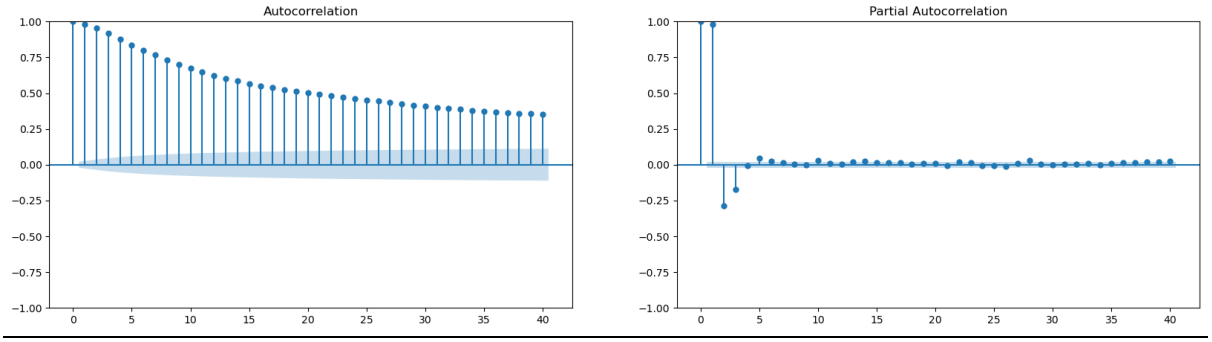
Singrauli, Surya Kiran Bhawan Dudhichua PM2.5 (µg/m3)



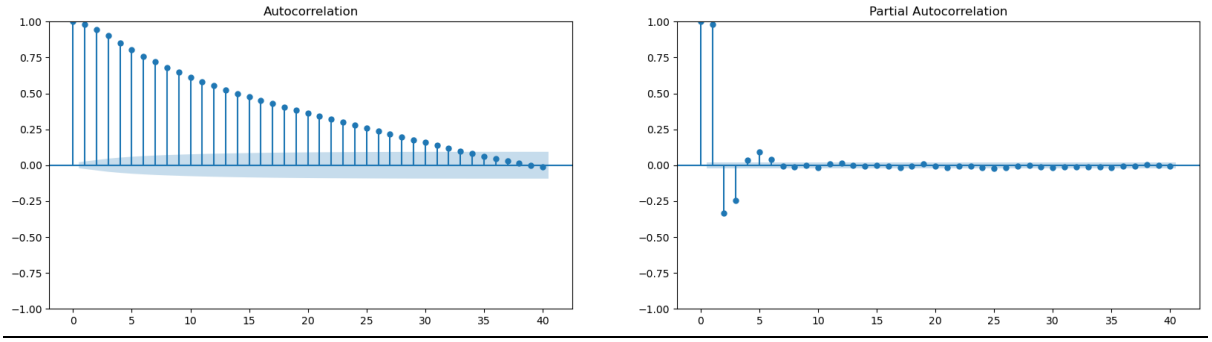
Singrauli, Surya Kiran Bhawan Dudhichua NO (µg/m3)



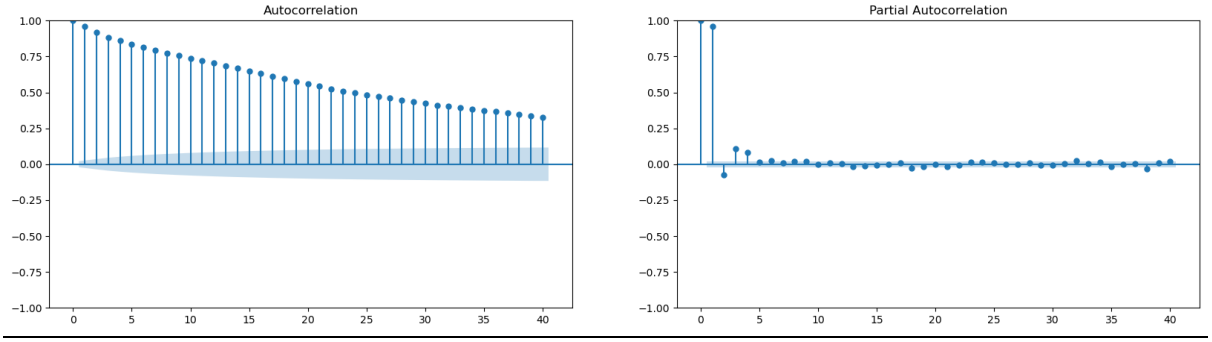
Singrauli, Surya Kiran Bhawan Dudhichua NO2 (µg/m3)



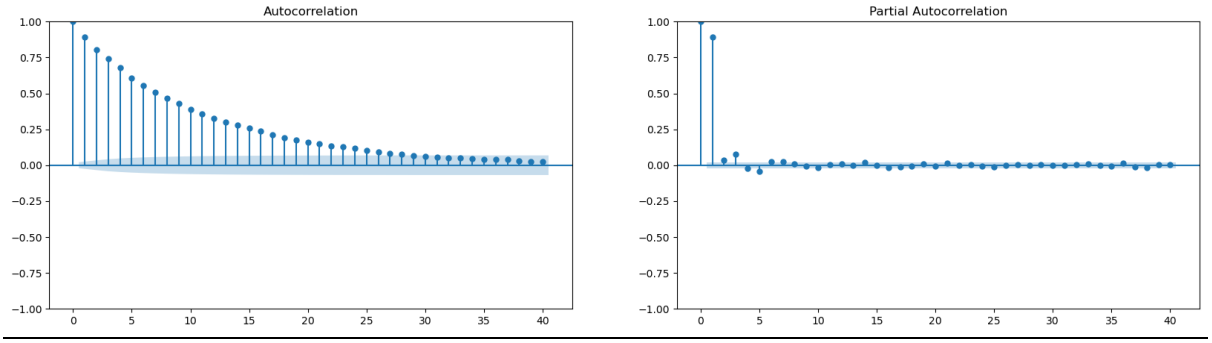
Singrauli, Surya Kiran Bhawan Dudhichua NOX (ppb)



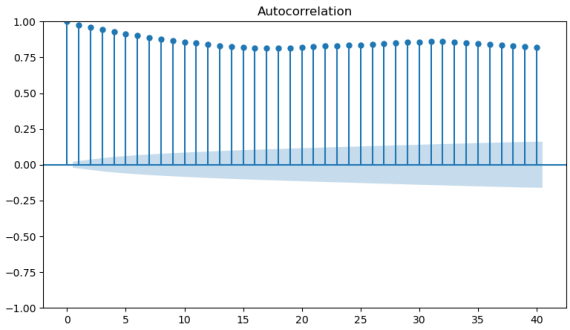
Singrauli, Surya Kiran Bhawan Dudhichua CO (mg/m3)

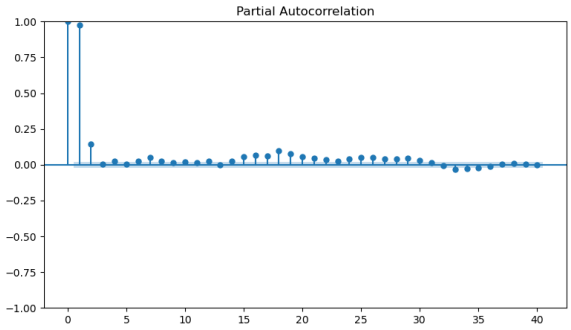


Singrauli, Surya Kiran Bhawan Dudhichua SO2 (µg/m3)

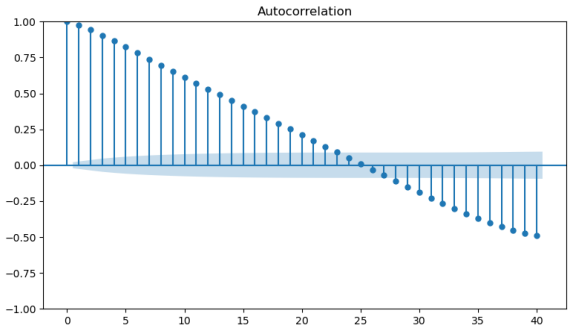


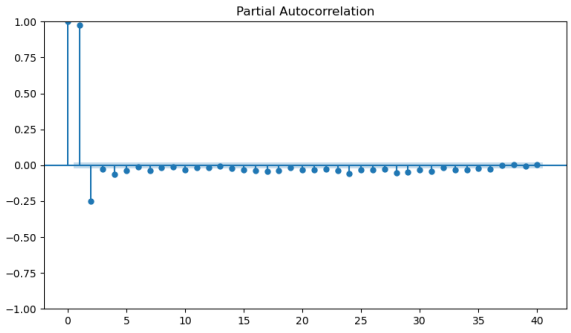
Singrauli, Surya Kiran Bhawan Dudhichua NH3 (µg/m3)



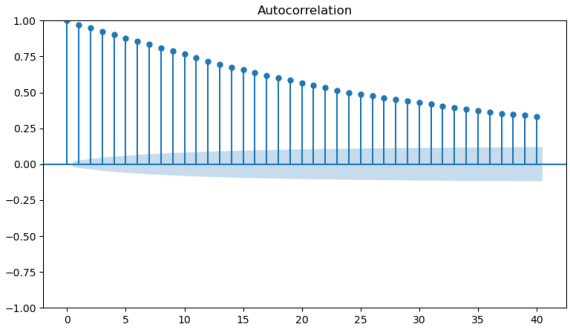


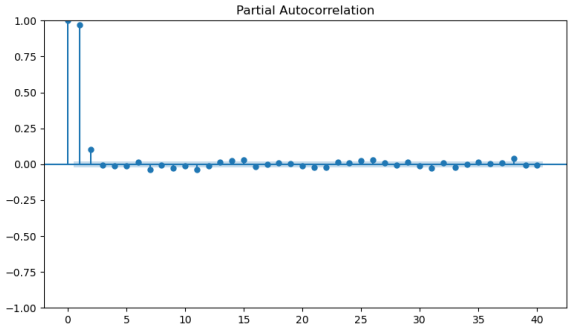
Singrauli, Surya Kiran Bhawan Dudhichua Ozone (µg/m3)





Singrauli, Surya Kiran Bhawan Dudhichua Benzene (µg/m3)





How do we select model on the basis of ACF and PACF plots of data

- **AR Model (AR):** If the ACF plot shows a gradual decay and cuts off after a certain lag, while t he PACF plot decays more quickly or cuts off after a certain lag, it suggests an Autoregressive (AR) model. The lag at which the PACF cuts off represents the order of the AR model.

- **MA Model (MA):** If the PACF plot shows a gradual decay and cuts off after a certain lag, whil e the ACF plot decays more quickly or cuts off after a certain lag, it indicates a Moving Avera ge (MA) model. The lag at which the ACF cuts off represents the order of the MA model.

- **ARMA Model (ARMA):** If both the ACF and PACF plots show gradual decay and cut off after c ertain lags, it suggests the need for both autoregressive and moving average components. In this case, an Autoregressive Moving Average (ARMA) model may be appropriate. The orders of the AR and MA components can be determined from the lags at which the ACF and PACF cut off, respectively.

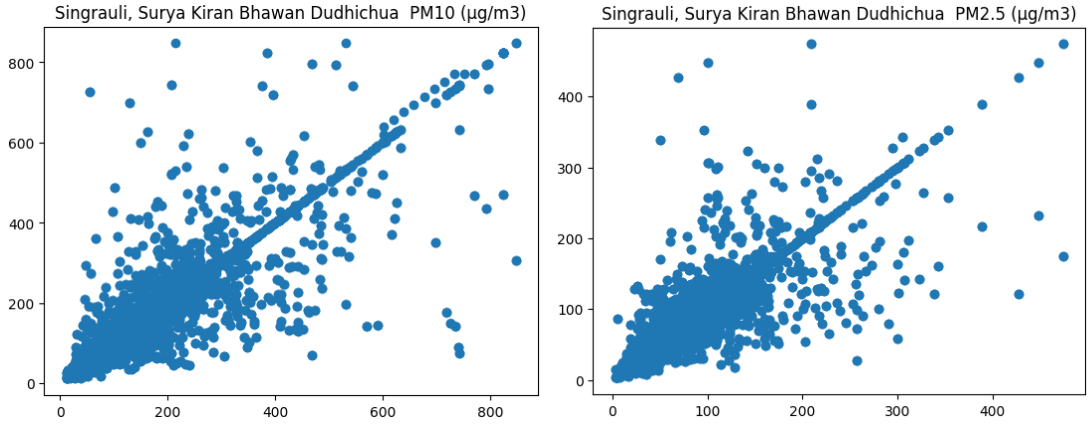
- **ARIMA Model (ARIMA):** If the ACF plot shows a gradual decay, while the PACF plot has a sig nificant spike at lag 1 followed by a gradual decay, it suggests the presence of an underlying trend in the data. In such cases, differencing the data to remove the trend and applying an A RMA model to the differenced data may be suitable. This is known as the Autoregressive Int egrated Moving Average (ARIMA) model. The order of differencing (d), the order of the AR c omponent (p), and the order of the MA component (q) can be determined based on the num ber of differencing operations needed and the lags at which the ACF and PACF cut off.

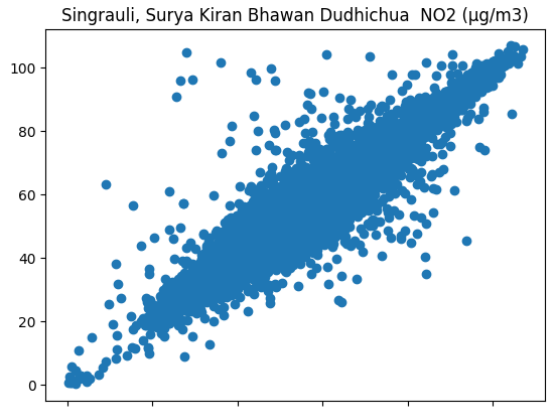
  Here, after observing the ACF and PACF plots carefully, I have arrived at a conclusion that in the plots of ACF and PACF, there is a gradual decay in autocorrelation while partial autocorre lation decays much faster than autocorrelation.
  So, we will use AR model.

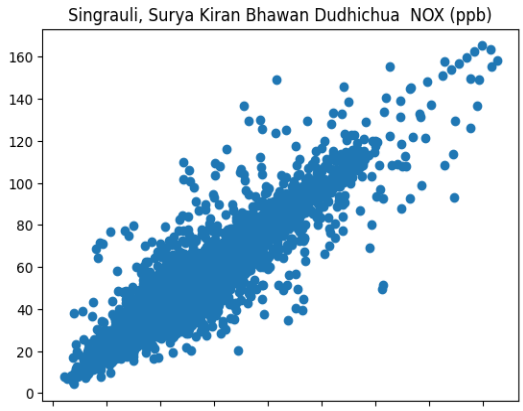
  Based on these findings, we can conclude that the most suitable model for the given data is an Auto Regressive (AR) model with a lag value, denoted as "p," equal to 2. This means that the current observation is dependent on the two previous observations. Therefore, we select the **AR (2)** model as the most appropriate choice to capture the underlying dynamics of the data.

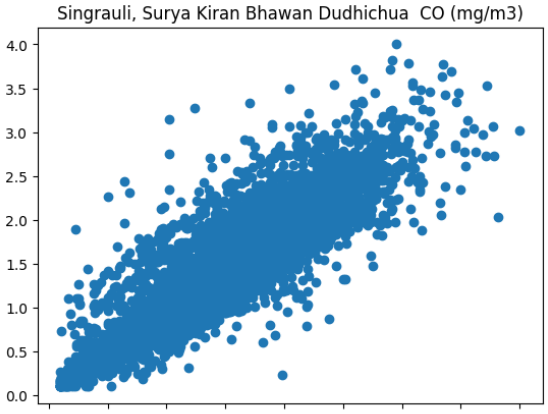
  The lag graphs provide evidence that the data exhibits a high level of correlation for a value of p equal to 2. This observation strongly supports the reasonableness of selecting p = 2 as an appropriate choice.

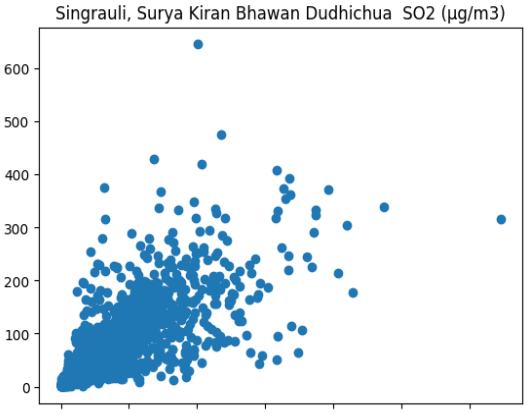
  Lag Graph: -

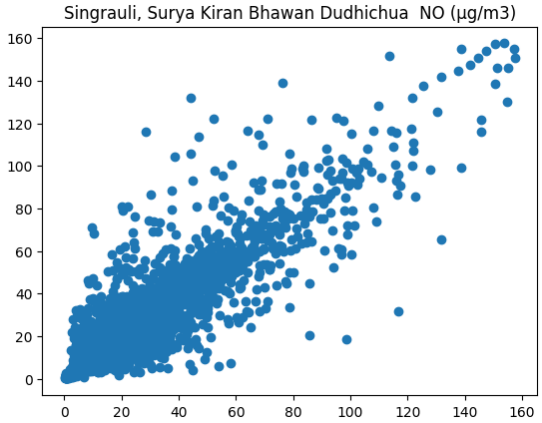


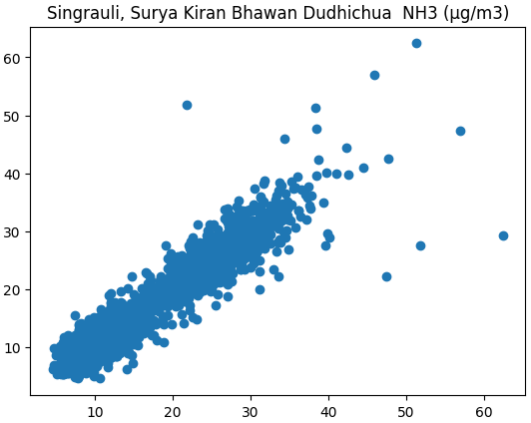

Singrauli, Surya Kiran Bhawan Dudhichua NO2 (µg/m3)

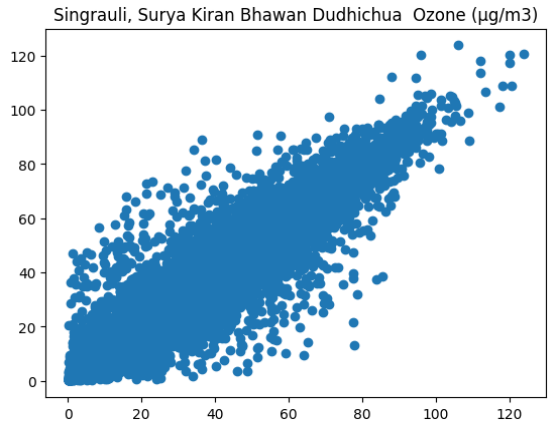

Singrauli, Surya Kiran Bhawan Dudhichua NOX (ppb)

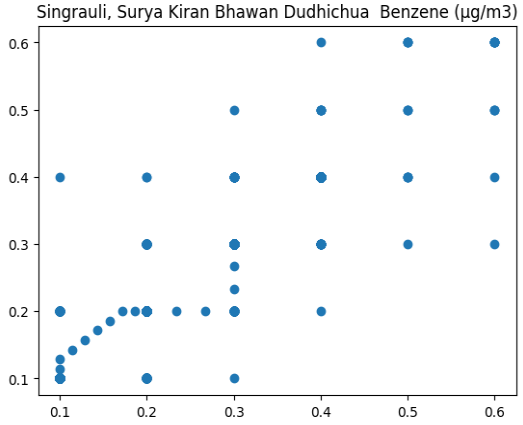

Singrauli, Surya Kiran Bhawan Dudhichua CO (mg/m3)


Singrauli, Surya Kiran Bhawan Dudhichua SO2 (µg/m3)


Singrauli, Surya Kiran Bhawan Dudhichua NO (µg/m3)


Singrauli, Surya Kiran Bhawan Dudhichua NH3 (µg/m3)


Singrauli, Surya Kiran Bhawan Dudhichua Ozone (µg/m3)


Singrauli, Surya Kiran Bhawan Dudhichua Benzene (µg/m3)

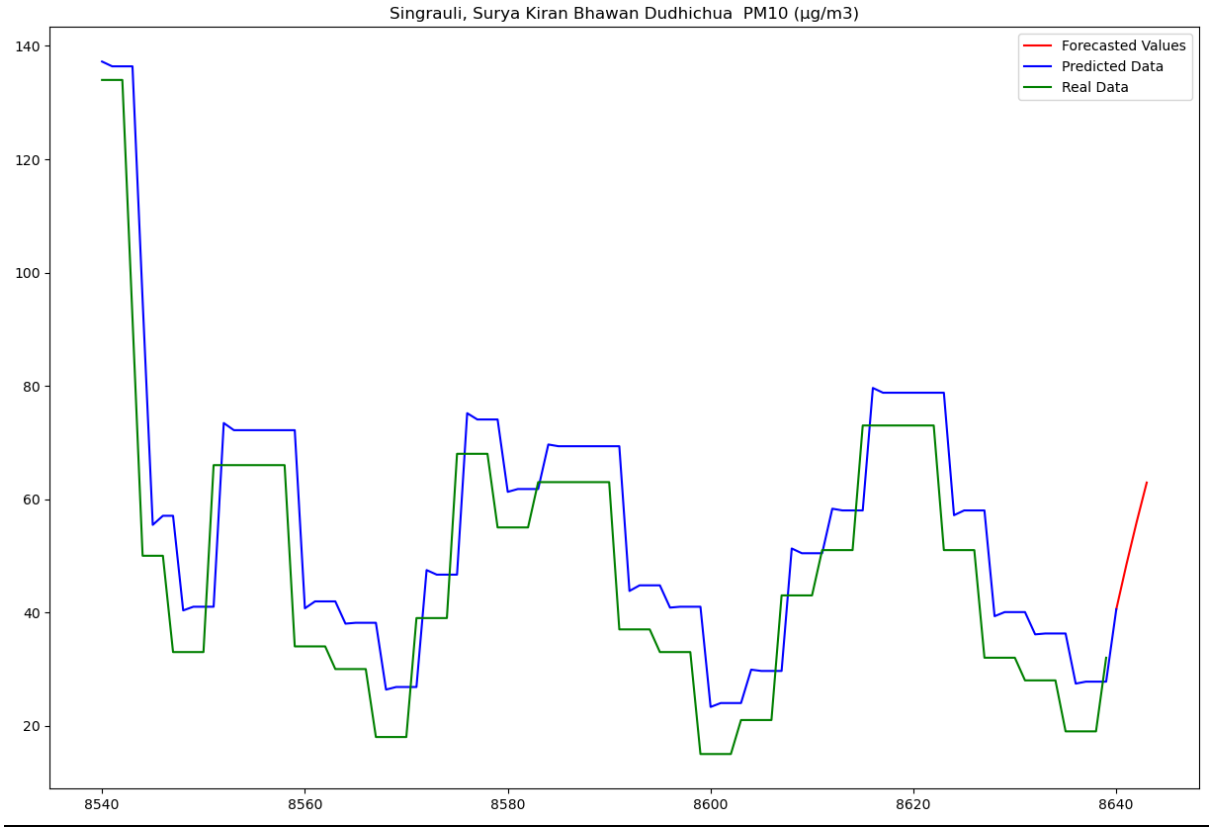
IMPLEMENTING AR MODEL ON DATA AND PREDICTING FUTURE VALUES USING THAT MODEL

Code: -

```
from sklearn.metrics import mean_absolute_error
for column in columns:
    data = dataSet[column][:8640]
    train_data = data[:-100]
    test_data = data[-100:]
    ar_model = AutoReg(data, lags = 2).fit()
    print(ar_model.summary())
    #print(len(train_data) , len(data))
    pred = ar_model.predict(start = len(train_data), end = len(data), dynamic=False)
    forecast = ar_model.predict(start = len(data), end = len(data)+3, dynamic=False)
    plt.figure(figsize = (15,10))
    plt.title(column)
    plt.plot(forecast , color ="red", label="Forecasted Values")
    plt.plot(pred, color = "blue", label="Predicted Data")
    plt.plot(test_data, color = "green", label="Real Data")
    plt.legend()
    plt.show()
    pred = ar_model.predict(start = len(train_data), end = len(data)-1, dynamic=False)
    rmse = sqrt(mean_squared_error(pred, test_data))
    mean = data.mean()
#     print("Mean :",mean)
    print("Mean Absolute Error:", mean_absolute_error(pred,test_data))
    print("Root Mean Squared Error:",rmse)
```

Results: -

```
                                AutoReg Model Results
==============================================================================
Dep. Variable:     Singrauli, Surya Kiran Bhawan Dudhichua  PM10 (µg/m3)   No. Observations:                 8640
Model:                                                        AutoReg(2)   Log Likelihood              -44166.077
Method:                                                   Conditional MLE   S.D. of innovations             40.208
Date:                                                    Tue, 27 Jun 2023   AIC                         88340.154
Time:                                                            21:47:21   BIC                         88368.410
Sample:                                                                 2   HQIC                        88349.788
                                                                     8640
==============================================================================
                                                                coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const                                                          9.8291      0.759     12.947      0.000       8.341      11.317
Singrauli, Surya Kiran Bhawan Dudhichua  PM10 (µg/m3).L1     0.9828      0.011     91.410      0.000       0.962       1.004
Singrauli, Surya Kiran Bhawan Dudhichua  PM10 (µg/m3).L2    -0.0382      0.011     -3.556      0.000      -0.059      -0.017
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1            1.0613           +0.0000j            1.0613            0.0000
AR.2           24.6460           +0.0000j           24.6460            0.0000
-----------------------------------------------------------------------------
```



Mean Absolute Error: 9.997905641686028
Root Mean Squared Error: 12.807041015837923

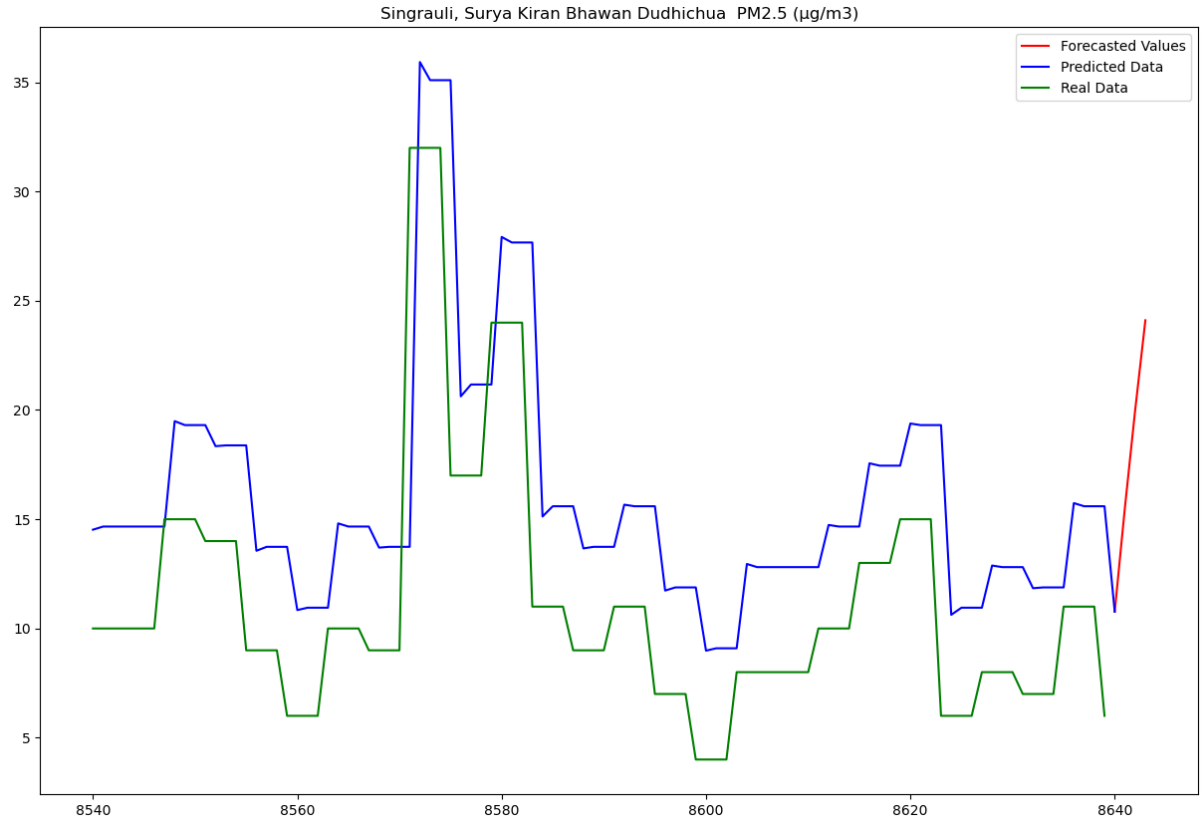
AutoReg Model Results

Dep. Variable:	Singrauli, Surya Kiran Bhawan Dudhichua PM2.5 (µg/m3)	No. Observations:	8640
Model:	AutoReg(2)	Log Likelihood	-38120.032
Method:	Conditional MLE	S.D. of innovations	19.968
Date:	Tue, 27 Jun 2023	AIC	76248.064
Time:	22:00:26	BIC	76276.320
Sample:	2	HQIC	76257.698
	8640		

	coef	std err	z	P>|z|	[0.025	0.975]
const	5.3773	0.370	14.528	0.000	4.652	6.103
Singrauli, Surya Kiran Bhawan Dudhichua PM2.5 (µg/m3).L1	0.9649	0.011	89.735	0.000	0.944	0.986
Singrauli, Surya Kiran Bhawan Dudhichua PM2.5 (µg/m3).L2	-0.0361	0.011	-3.356	0.001	-0.057	-0.015

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	1.0800	+0.0000j	1.0800	0.0000
AR.2	25.6547	+0.0000j	25.6547	0.0000



Mean Absolute Error: 5.0359820715358525
Root Mean Squared Error: 5.739367454946405

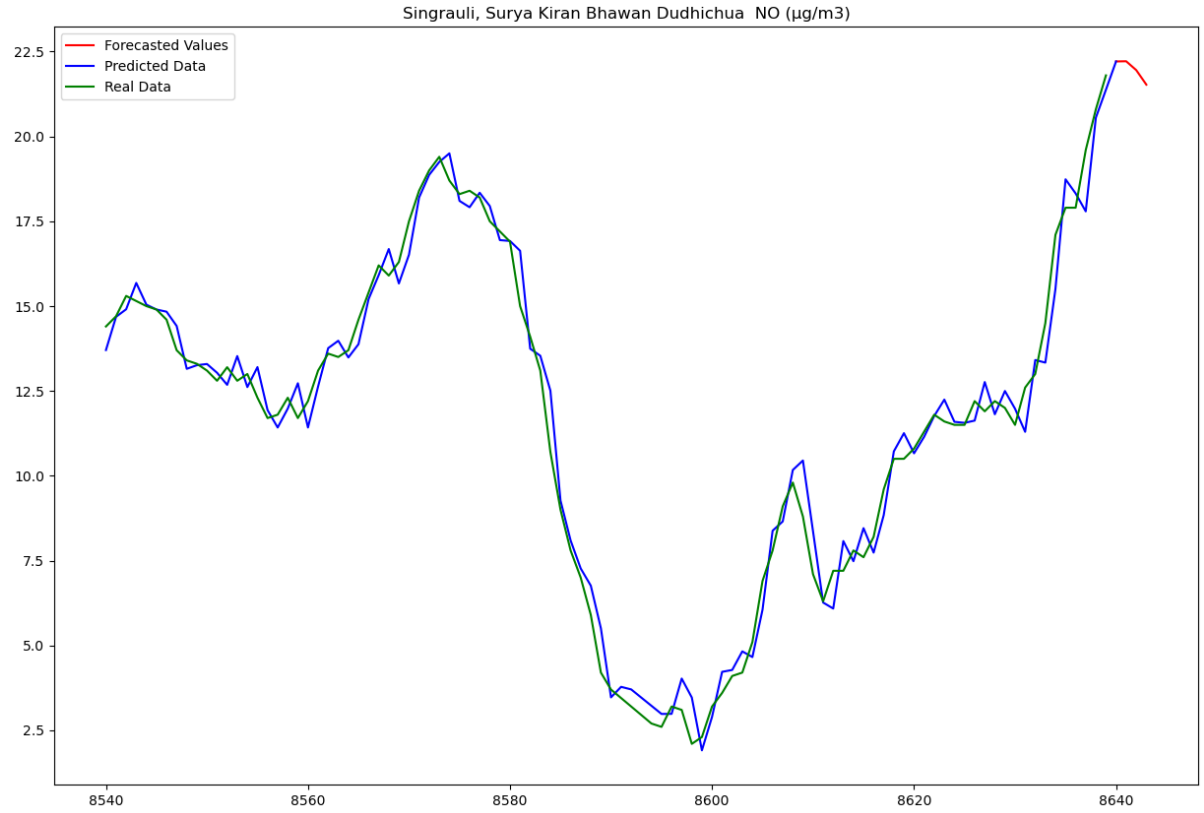
AutoReg Model Results

Dep. Variable:	Singrauli, Surya Kiran Bhawan Dudhichua NO (µg/m3)	No. Observations:	8640
Model:	AutoReg(2)	Log Likelihood	-21144.423
Method:	Conditional MLE	S.D. of innovations	2.798
Date:	Tue, 27 Jun 2023	AIC	42296.847
Time:	22:00:26	BIC	42325.103
Sample:	2	HQIC	42306.481
	8640		

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.5448	0.039	13.850	0.000	0.468	0.622
Singrauli, Surya Kiran Bhawan Dudhichua NO (µg/m3).L1	1.6233	0.008	200.773	0.000	1.607	1.639
Singrauli, Surya Kiran Bhawan Dudhichua NO (µg/m3).L2	-0.6598	0.008	-81.601	0.000	-0.676	-0.644

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	1.2302	-0.0475j	1.2311	-0.0061
AR.2	1.2302	+0.0475j	1.2311	0.0061



Mean Absolute Error: 0.5385183487941294
Root Mean Squared Error: 0.6791086790313348

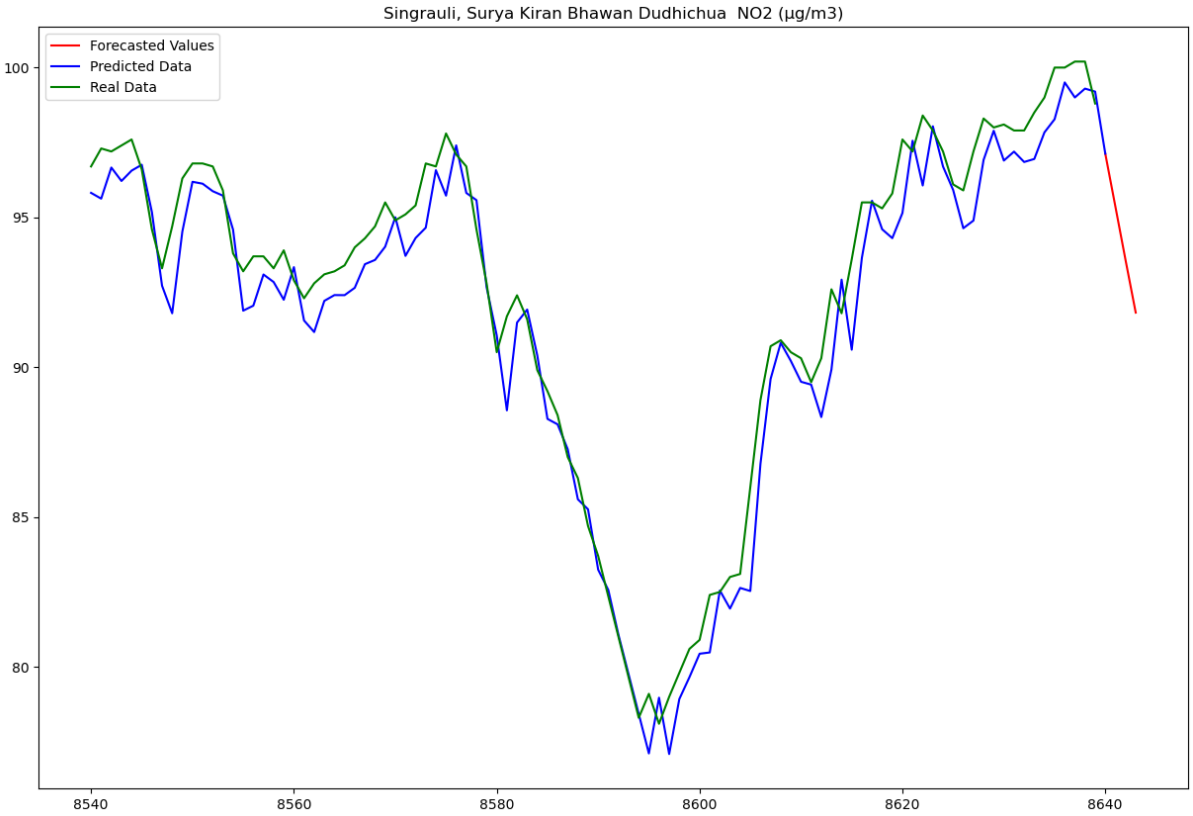
AutoReg Model Results

Dep. Variable:	Singrauli, Surya Kiran Bhawan Dudhichua NO2 (µg/m3)	No. Observations:	8640
Model:	AutoReg(2)	Log Likelihood	-21776.975
Method:	Conditional MLE	S.D. of innovations	3.011
Date:	Tue, 27 Jun 2023	AIC	43561.949
Time:	22:00:27	BIC	43590.205
Sample:	2	HQIC	43571.584
	8640		

	coef	std err	z	P>|z|	[0.025	0.975]
const	1.2420	0.095	13.084	0.000	1.056	1.428
Singrauli, Surya Kiran Bhawan Dudhichua NO2 (µg/m3).L1	1.4793	0.009	158.933	0.000	1.461	1.498
Singrauli, Surya Kiran Bhawan Dudhichua NO2 (µg/m3).L2	-0.5017	0.009	-53.895	0.000	-0.520	-0.483

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	1.0497	+0.0000j	1.0497	0.0000
AR.2	1.8989	+0.0000j	1.8989	0.0000



Mean Absolute Error: 1.0149966459763704
Root Mean Squared Error: 1.2730811601111378

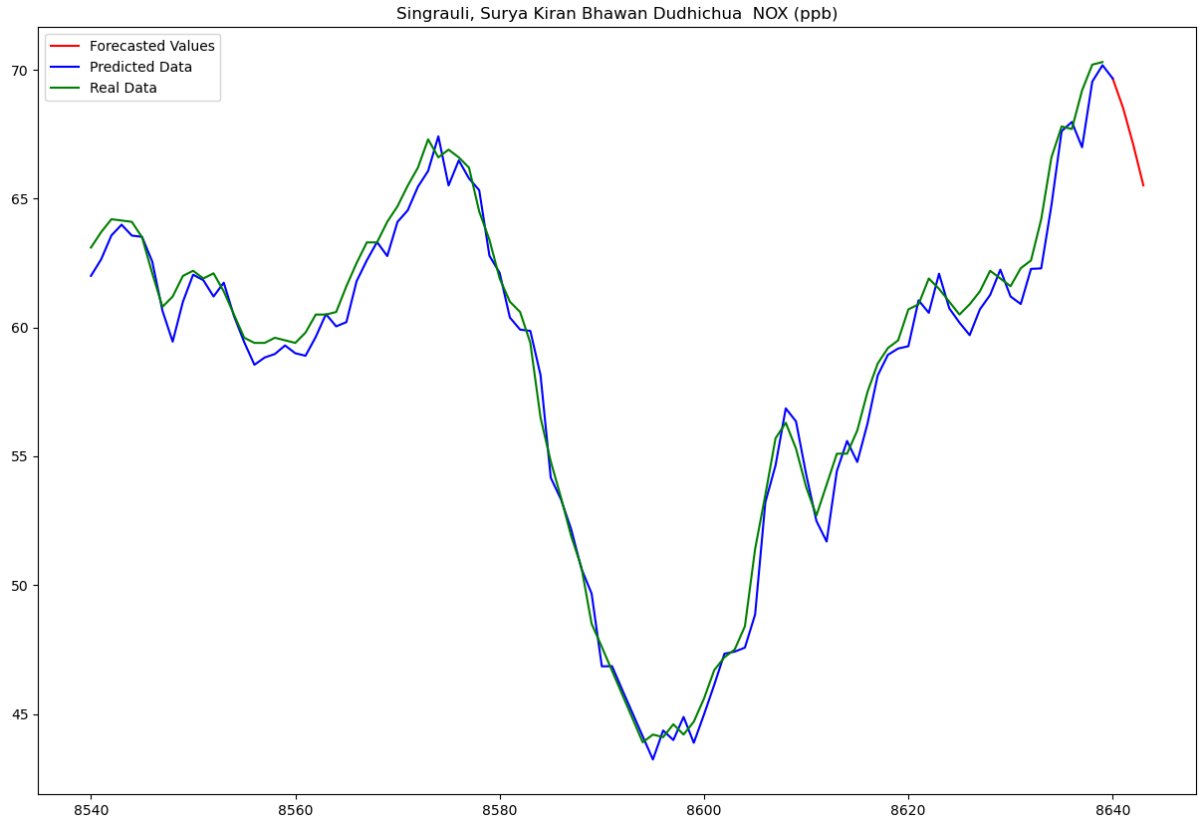
AutoReg Model Results

Dep. Variable:	Singrauli, Surya Kiran Bhawan Dudhichua NOX (ppb)	No. Observations:	8640
Model:	AutoReg(2)	Log Likelihood	-21178.954
Method:	Conditional MLE	S.D. of innovations	2.809
Date:	Tue, 27 Jun 2023	AIC	42365.907
Time:	22:00:27	BIC	42394.163
Sample:	2	HQIC	42375.542
	8640		

	coef	std err	z	P>|z|	[0.025	0.975]
const	1.0703	0.065	16.387	0.000	0.942	1.198
Singrauli, Surya Kiran Bhawan Dudhichua NOX (ppb).L1	1.6553	0.008	209.979	0.000	1.640	1.671
Singrauli, Surya Kiran Bhawan Dudhichua NOX (ppb).L2	-0.6806	0.008	-86.329	0.000	-0.696	-0.665

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	1.1183	+0.0000j	1.1183	0.0000
AR.2	1.3138	+0.0000j	1.3138	0.0000



Mean Absolute Error: 0.680213076132455
Root Mean Squared Error: 0.8627181893328912

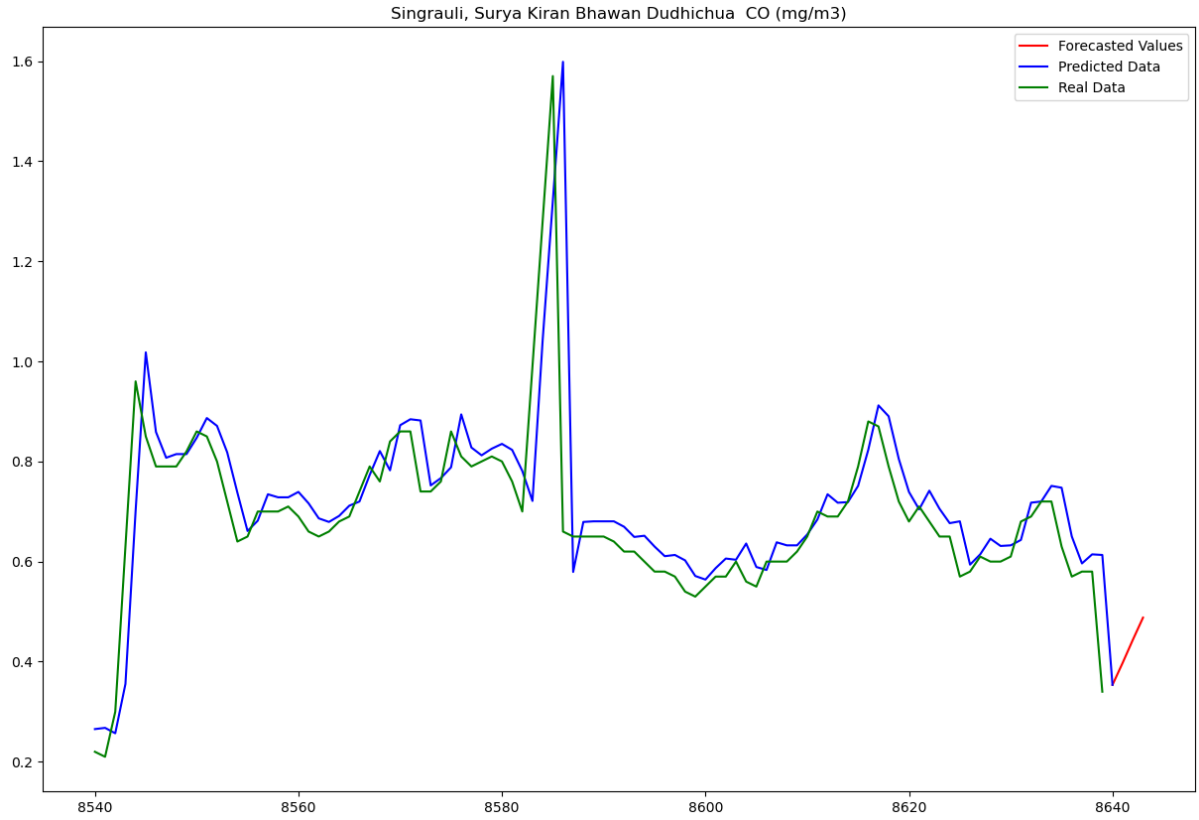
AutoReg Model Results

Dep. Variable:	Singrauli, Surya Kiran Bhawan Dudhichua CO (mg/m3)	No. Observations:	8640
Model:	AutoReg(2)	Log Likelihood	3196.420
Method:	Conditional MLE	S.D. of innovations	0.167
Date:	Tue, 27 Jun 2023	AIC	-6384.841
Time:	22:00:28	BIC	-6356.585
Sample:	2	HQIC	-6375.206
	8640		

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.0567	0.004	12.866	0.000	0.048	0.065
Singrauli, Surya Kiran Bhawan Dudhichua CO (mg/m3).L1	1.0812	0.011	101.226	0.000	1.060	1.102
Singrauli, Surya Kiran Bhawan Dudhichua CO (mg/m3).L2	-0.1216	0.011	-11.388	0.000	-0.143	-0.101

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	1.0486	+0.0000j	1.0486	0.0000
AR.2	7.8409	+0.0000j	7.8409	0.0000



Mean Absolute Error: 0.06349337473985234
Root Mean Squared Error: 0.12402669237122571

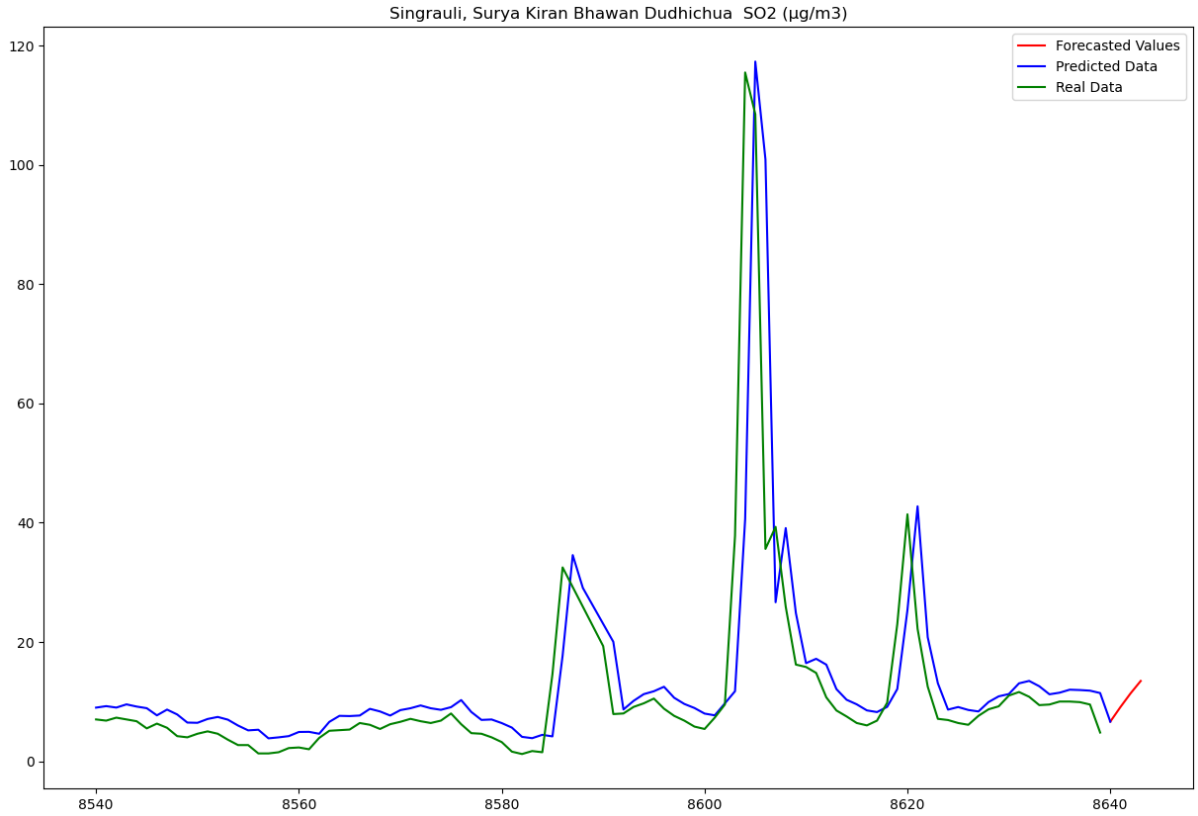
AutoReg Model Results

Dep. Variable:	Singrauli, Surya Kiran Bhawan Dudhichua SO2 (µg/m3)	No. Observations:	8640
Model:	AutoReg(2)	Log Likelihood	-35690.845
Method:	Conditional MLE	S.D. of innovations	15.073
Date:	Tue, 27 Jun 2023	AIC	71389.691
Time:	22:00:28	BIC	71417.947
Sample:	2	HQIC	71399.325
	8640		

	coef	std err	z	P>|z|	[0.025	0.975]
const	2.8104	0.211	13.304	0.000	2.396	3.224
Singrauli, Surya Kiran Bhawan Dudhichua SO2 (µg/m3).L1	1.0302	0.011	96.426	0.000	1.009	1.051
Singrauli, Surya Kiran Bhawan Dudhichua SO2 (µg/m3).L2	-0.1183	0.011	-11.069	0.000	-0.139	-0.097

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	1.1128	+0.0000j	1.1128	0.0000
AR.2	7.5986	+0.0000j	7.5986	0.0000



Mean Absolute Error: 5.139238616358898
Root Mean Squared Error: 11.41102988858203

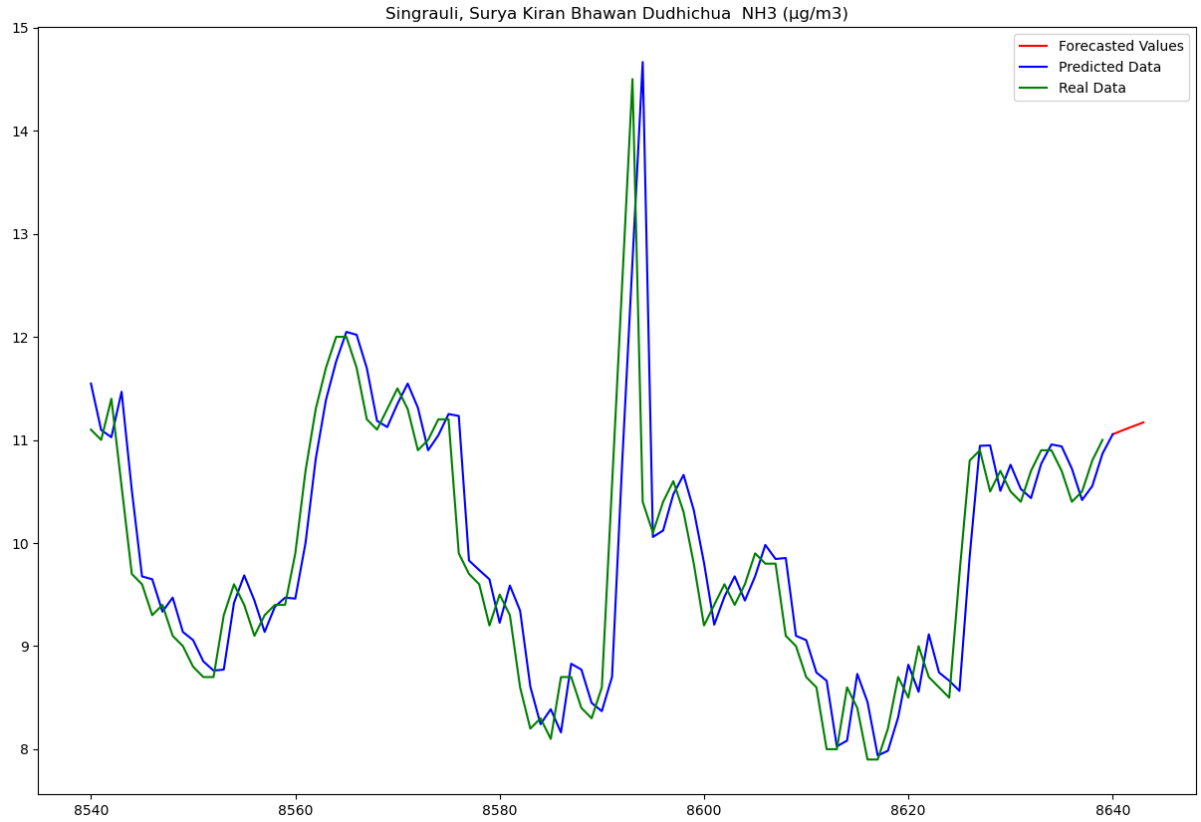
AutoReg Model Results

Dep. Variable:	Singrauli, Surya Kiran Bhawan Dudhichua NH3 (µg/m3)	No. Observations:	8640
Model:	AutoReg(2)	Log Likelihood	-12626.145
Method:	Conditional MLE	S.D. of innovations	1.044
Date:	Tue, 27 Jun 2023	AIC	25260.290
Time:	22:00:29	BIC	25288.546
Sample:	2	HQIC	25269.925
	8640		

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.2109	0.027	7.888	0.000	0.159	0.263
Singrauli, Surya Kiran Bhawan Dudhichua NH3 (µg/m3).L1	1.0784	0.011	100.676	0.000	1.057	1.099
Singrauli, Surya Kiran Bhawan Dudhichua NH3 (µg/m3).L2	-0.0943	0.011	-8.805	0.000	-0.115	-0.073

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	1.0179	+0.0000j	1.0179	0.0000
AR.2	10.4164	+0.0000j	10.4164	0.0000



Mean Absolute Error: 0.38601713221981465
Root Mean Squared Error: 0.6506204060723245

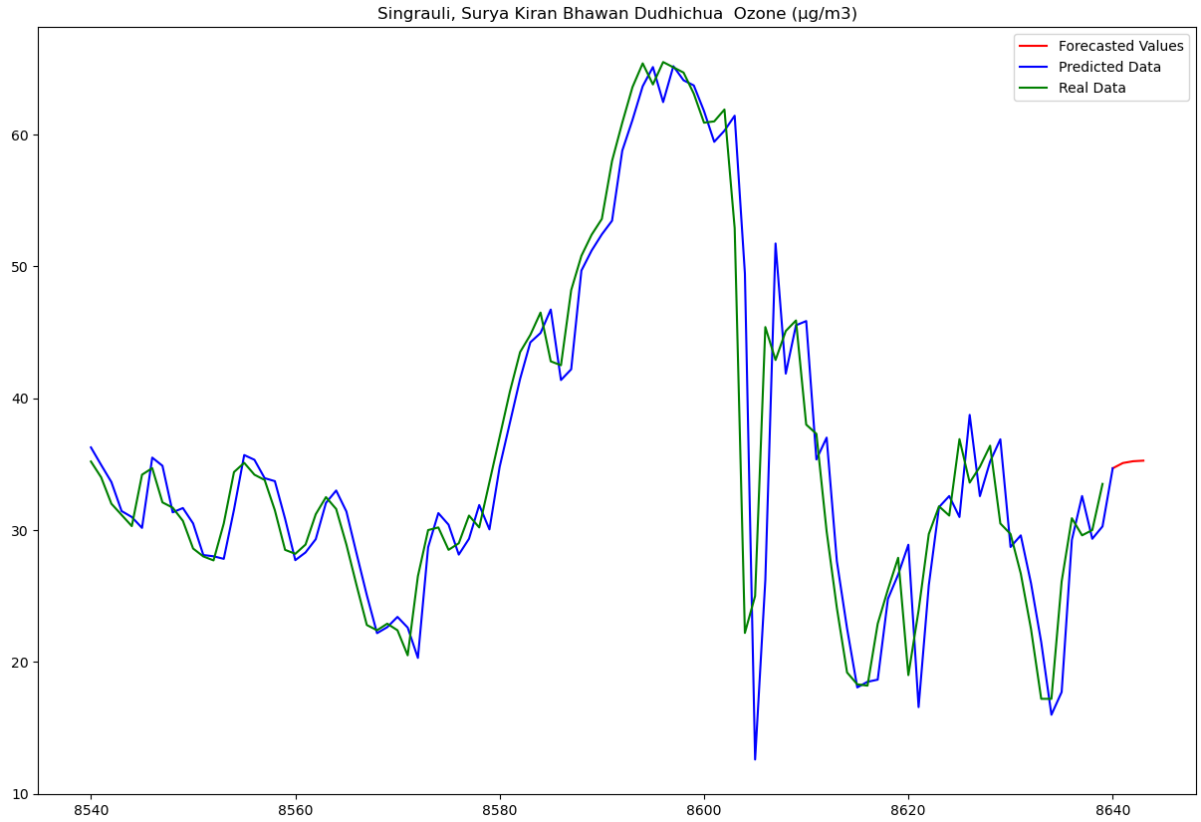
AutoReg Model Results

Dep. Variable:	Singrauli, Surya Kiran Bhawan Dudhichua Ozone (µg/m3)	No. Observations:	8640
Model:	AutoReg(2)	Log Likelihood	-26584.645
Method:	Conditional MLE	S.D. of innovations	5.252
Date:	Tue, 27 Jun 2023	AIC	53177.289
Time:	22:00:29	BIC	53205.545
Sample:	2	HQIC	53186.923
	8640		

	coef	std err	z	P>|z|	[0.025	0.975]
const	1.0079	0.093	10.785	0.000	0.825	1.191
Singrauli, Surya Kiran Bhawan Dudhichua Ozone (µg/m3).L1	1.2963	0.010	127.393	0.000	1.276	1.316
Singrauli, Surya Kiran Bhawan Dudhichua Ozone (µg/m3).L2	-0.3250	0.010	-31.934	0.000	-0.345	-0.305

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	1.0453	+0.0000j	1.0453	0.0000
AR.2	2.9440	+0.0000j	2.9440	0.0000



Mean Absolute Error: 2.8940956929890205
Root Mean Squared Error: 4.775819407486162

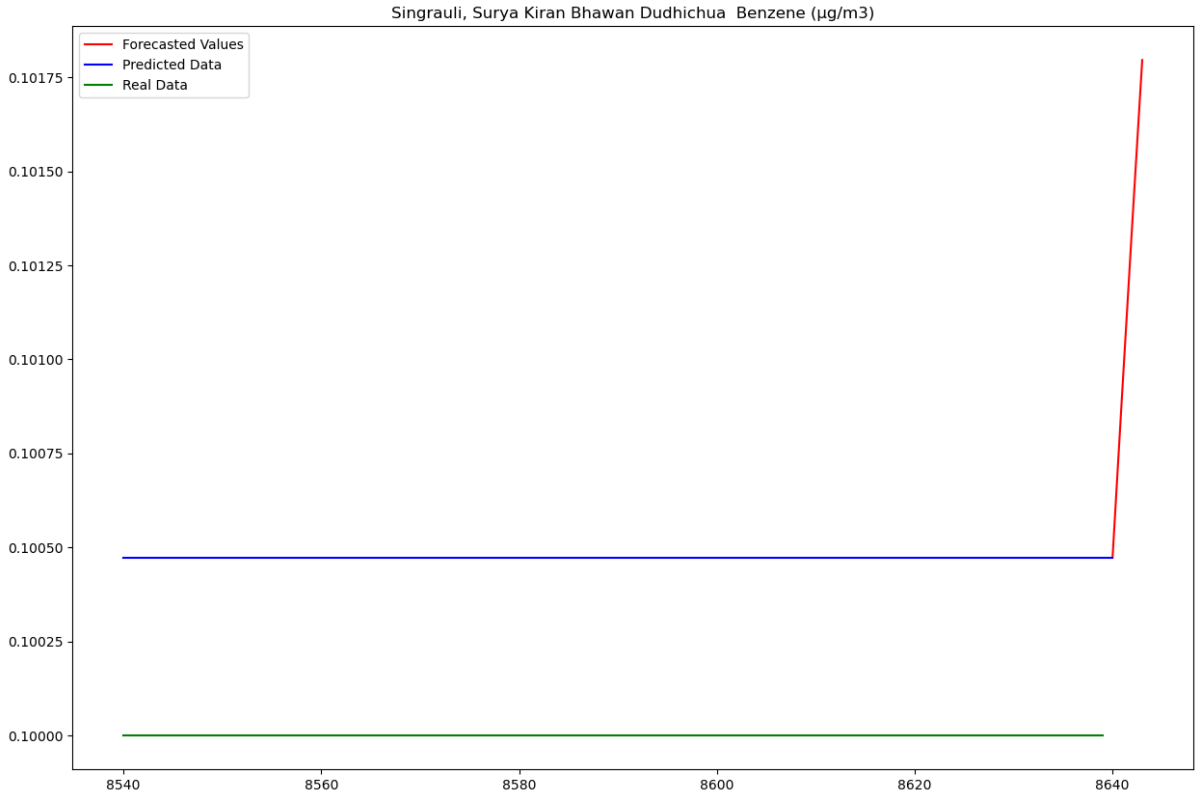
AutoReg Model Results

Dep. Variable:	Singrauli, Surya Kiran Bhawan Dudhichua Benzene (µg/m3)	No. Observations:	8640
Model:	AutoReg(2)	Log Likelihood	25041.140
Method:	Conditional MLE	S.D. of innovations	0.013
Date:	Tue, 27 Jun 2023	AIC	-50074.281
Time:	22:00:29	BIC	-50046.025
Sample:	2	HQIC	-50064.646
	8640		

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.0028	0.000	8.906	0.000	0.002	0.003
Singrauli, Surya Kiran Bhawan Dudhichua Benzene (µg/m3).L1	0.9515	0.011	88.459	0.000	0.930	0.973
Singrauli, Surya Kiran Bhawan Dudhichua Benzene (µg/m3).L2	0.0254	0.011	2.361	0.018	0.004	0.046

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	1.0231	+0.0000j	1.0231	0.0000
AR.2	-38.5316	+0.0000j	38.5316	0.5000




```
Mean Absolute Error: 0.00047358621369454623
Root Mean Squared Error: 0.0004735862136945462
```

HOW TO CHECK WHETHER OUR MODEL IS GOOD OR NOT?

- **Residual Analysis:**
  One way to assess the goodness of an AR model is by analysing the residuals, which are the differences between the predicted values and the actual values of the time series. Plotting t he residuals over time can help identify any patterns or trends that may suggest inadequacie s in the model. Ideally, the residuals should be random with zero mean and constant varianc e.
  You can also examine the autocorrelation of the residuals using the autocorrelation function (ACF) plot. In a well-fitted AR model, the autocorrelation of the residuals should be close to z ero for all lags.

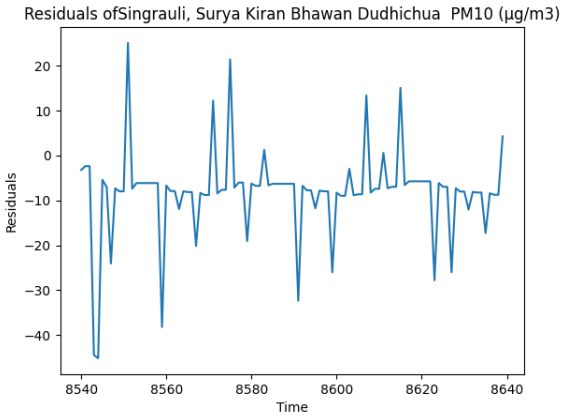
- **Model Fit and Prediction Accuracy:**
  Another way to evaluate the goodness of an AR model is by examining how well it fits the tra ining data and its predictive accuracy on unseen data.
  You can compute various metrics such as the mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), or mean absolute percentage error (MAPE) to qua ntify the model's prediction performance.
  Splitting your data into a training set and a test set can help assess the model's ability to gen eralize and make accurate predictions on new data.

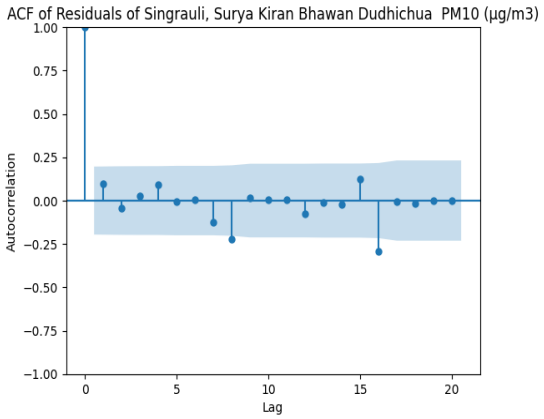
- **Information Criteria:**
  Information criteria such as Akaike Information Criterion (AIC) or Bayesian Information Criter ion (BIC) can be used to compare the goodness of fit between different AR models.

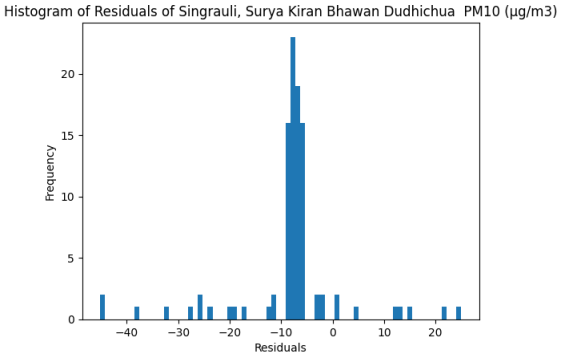
Lower values of AIC or BIC indicate a better fit, considering the trade-off between model co mplexity and goodness of fit.

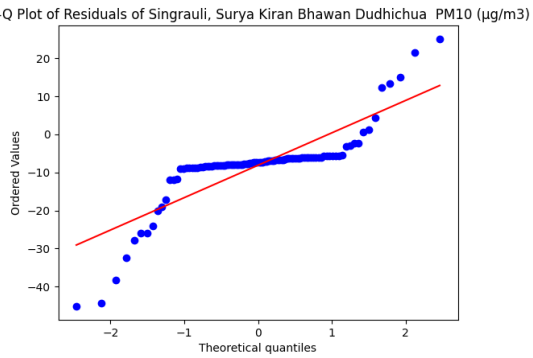
ANALYSING RESIDUALS

- **PM 10**

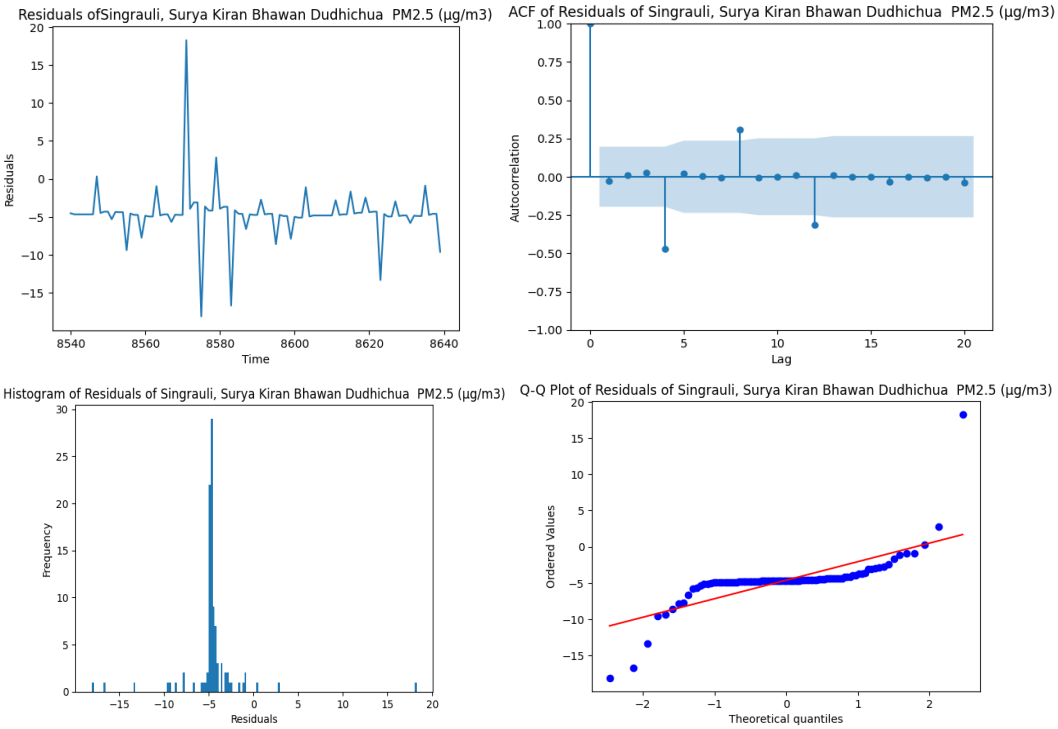




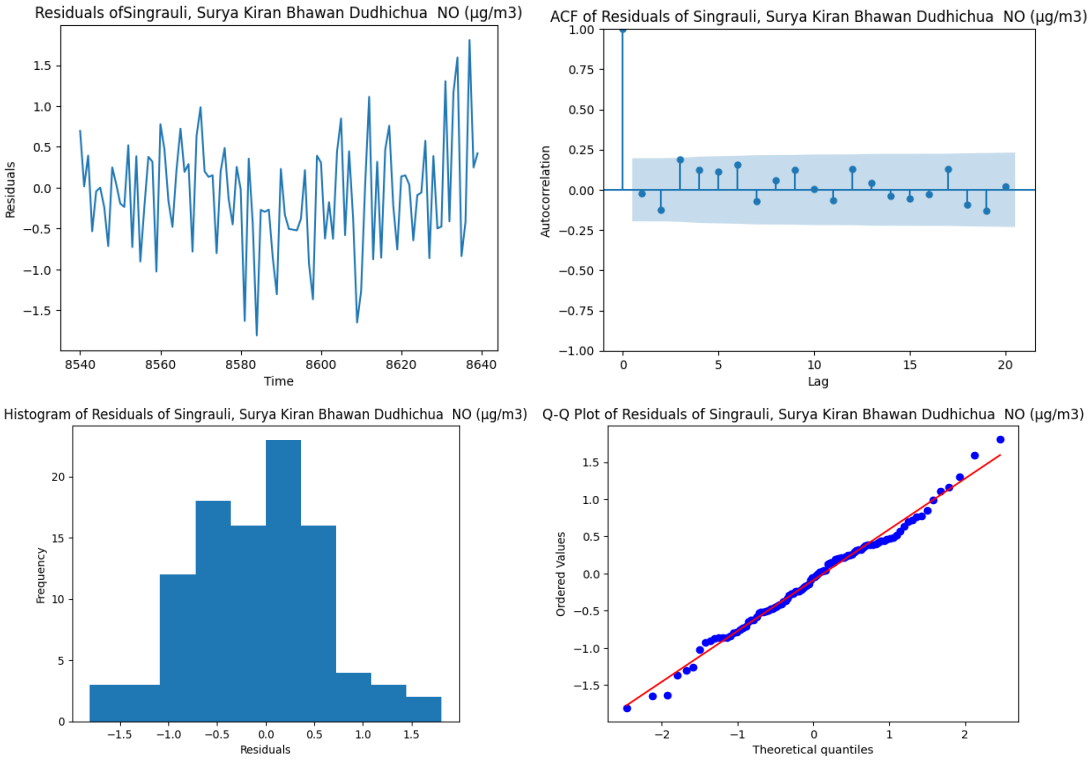




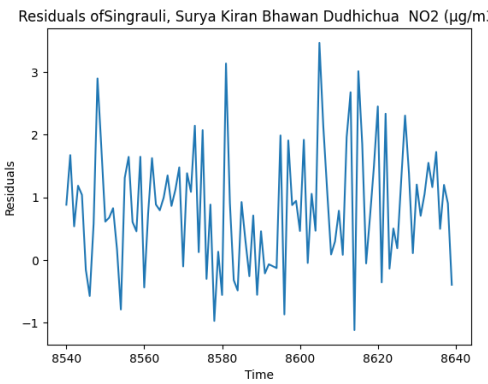
- **PM 2.5**

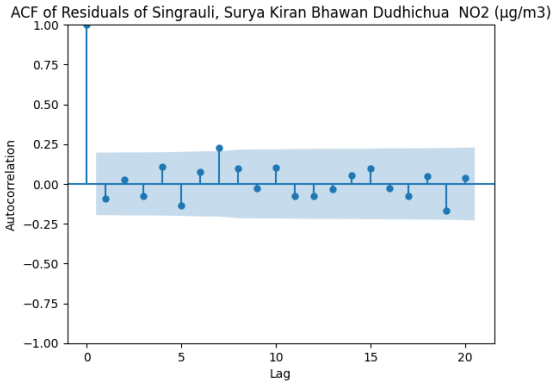

Residuals ofSingrauli, Surya Kiran Bhawan Dudhichua PM2.5 (µg/m3)
Residuals
Time
ACF of Residuals of Singrauli, Surya Kiran Bhawan Dudhichua PM2.5 (µg/m3)
Autocorrelation
Lag
Histogram of Residuals of Singrauli, Surya Kiran Bhawan Dudhichua PM2.5 (µg/m3)
Frequency
Residuals
Q-Q Plot of Residuals of Singrauli, Surya Kiran Bhawan Dudhichua PM2.5 (µg/m3)
Ordered Values
Theoretical quantiles

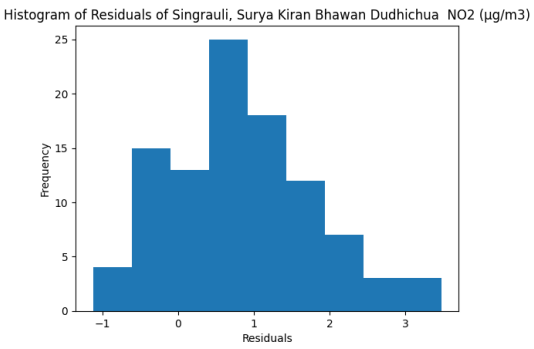
- **NO**

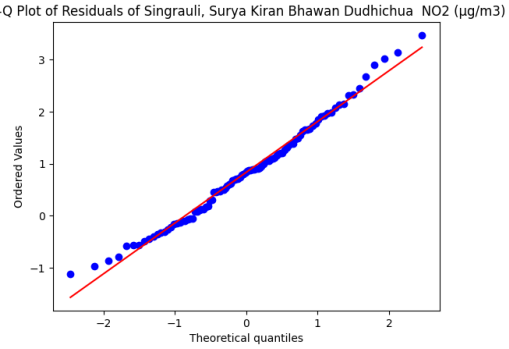

Residuals ofSingrauli, Surya Kiran Bhawan Dudhichua NO (µg/m3)
Residuals
Time
ACF of Residuals of Singrauli, Surya Kiran Bhawan Dudhichua NO (µg/m3)
Autocorrelation
Lag
Histogram of Residuals of Singrauli, Surya Kiran Bhawan Dudhichua NO (µg/m3)
Frequency
Residuals
Q-Q Plot of Residuals of Singrauli, Surya Kiran Bhawan Dudhichua NO (µg/m3)
Ordered Values
Theoretical quantiles

- **NO2**

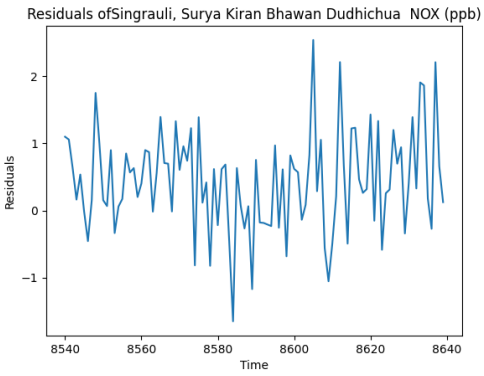

Residuals ofSingrauli, Surya Kiran Bhawan Dudhichua NO2 (µg/m3)
Residuals
Time

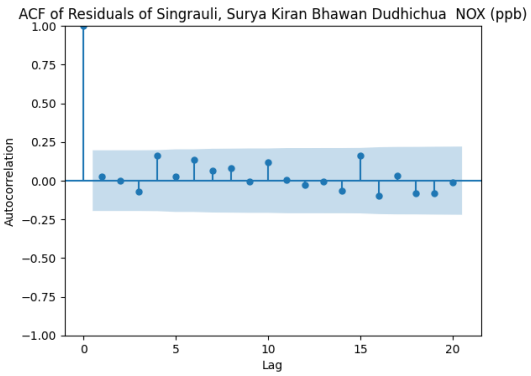

ACF of Residuals of Singrauli, Surya Kiran Bhawan Dudhichua NO2 (µg/m3)
Autocorrelation
Lag

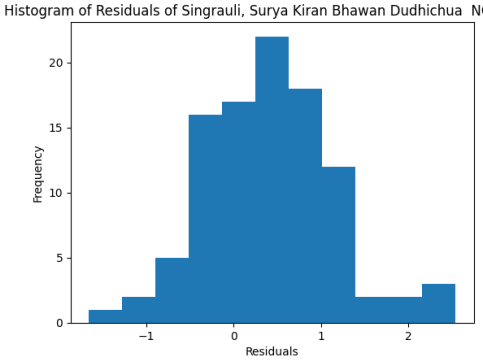

Histogram of Residuals of Singrauli, Surya Kiran Bhawan Dudhichua NO2 (µg/m3)
Frequency
Residuals

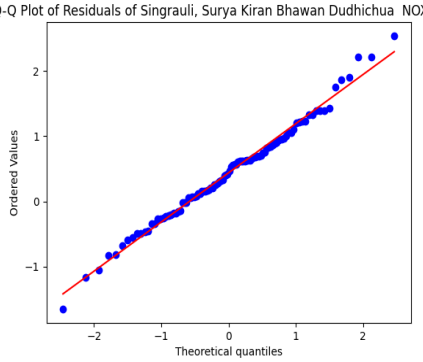

Q-Q Plot of Residuals of Singrauli, Surya Kiran Bhawan Dudhichua NO2 (µg/m3)
Ordered Values
Theoretical quantiles

- **NOX**

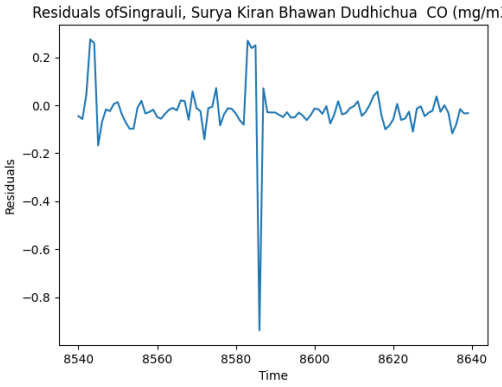

Residuals ofSingrauli, Surya Kiran Bhawan Dudhichua NOX (ppb)
Residuals
Time

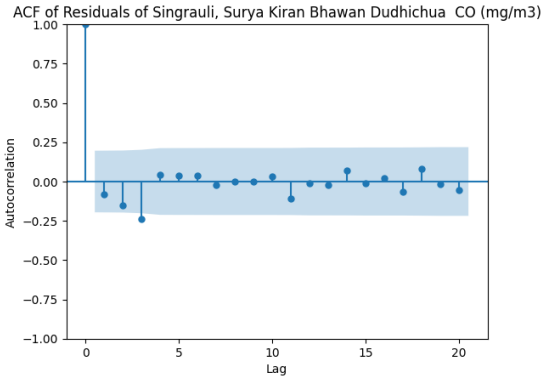

ACF of Residuals of Singrauli, Surya Kiran Bhawan Dudhichua NOX (ppb)
Autocorrelation
Lag

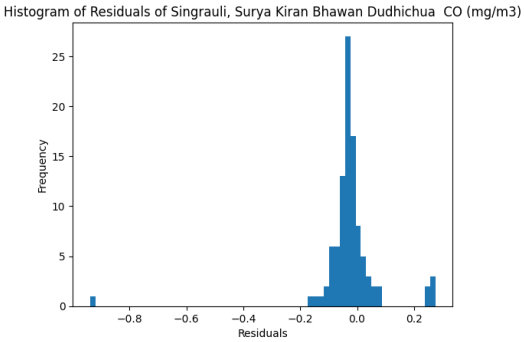

Histogram of Residuals of Singrauli, Surya Kiran Bhawan Dudhichua NOX (ppb)
Frequency
Residuals

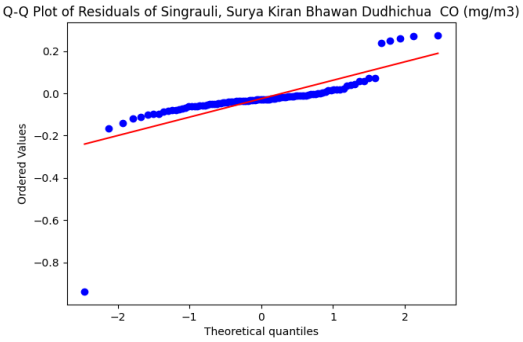

Q-Q Plot of Residuals of Singrauli, Surya Kiran Bhawan Dudhichua NOX (ppb)
Ordered Values
Theoretical quantiles

- **CO**

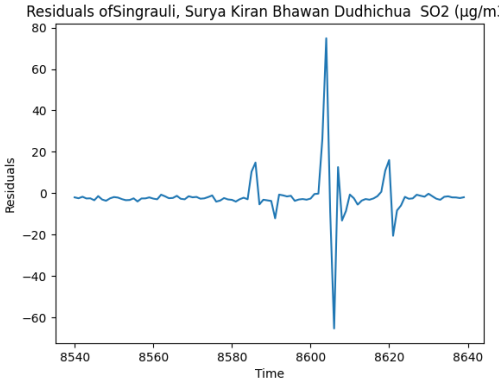

Residuals ofSingrauli, Surya Kiran Bhawan Dudhichua CO (mg/m3)
Residuals
Time

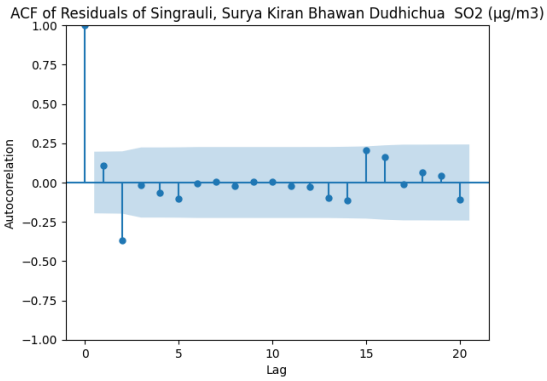

ACF of Residuals of Singrauli, Surya Kiran Bhawan Dudhichua CO (mg/m3)
Autocorrelation
Lag

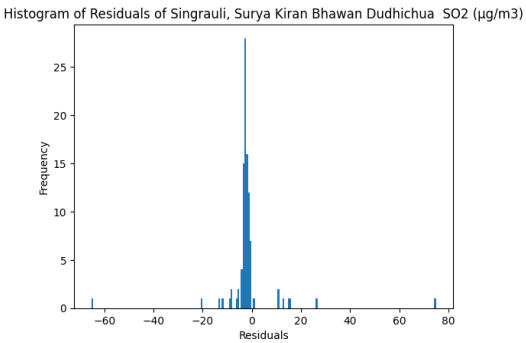

Histogram of Residuals of Singrauli, Surya Kiran Bhawan Dudhichua CO (mg/m3)
Frequency
Residuals

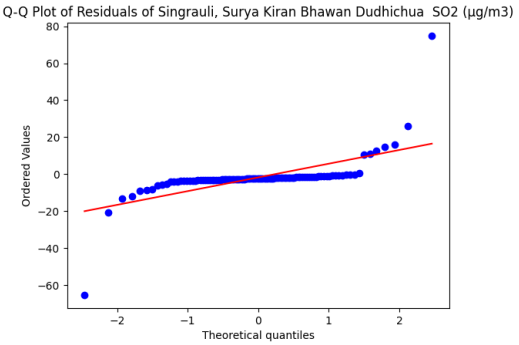

Q-Q Plot of Residuals of Singrauli, Surya Kiran Bhawan Dudhichua CO (mg/m3)
Ordered Values
Theoretical quantiles

- **SO2**

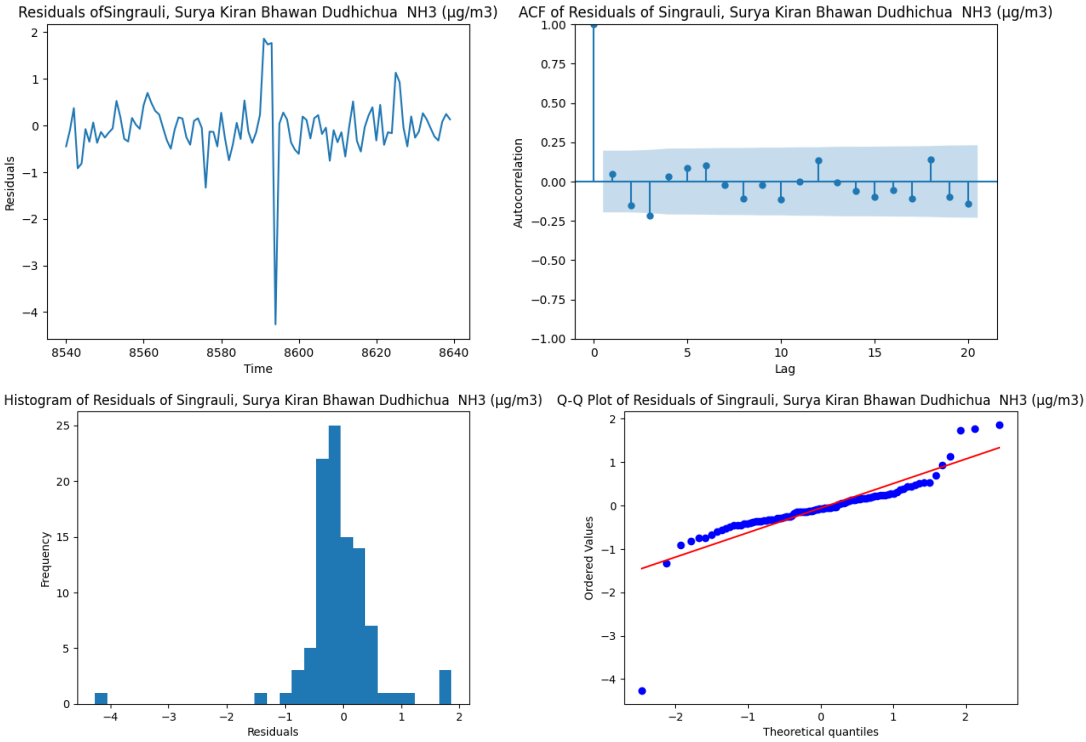

Residuals ofSingrauli, Surya Kiran Bhawan Dudhichua SO2 (µg/m3)
Residuals
Time


ACF of Residuals of Singrauli, Surya Kiran Bhawan Dudhichua SO2 (µg/m3)
Autocorrelation
Lag

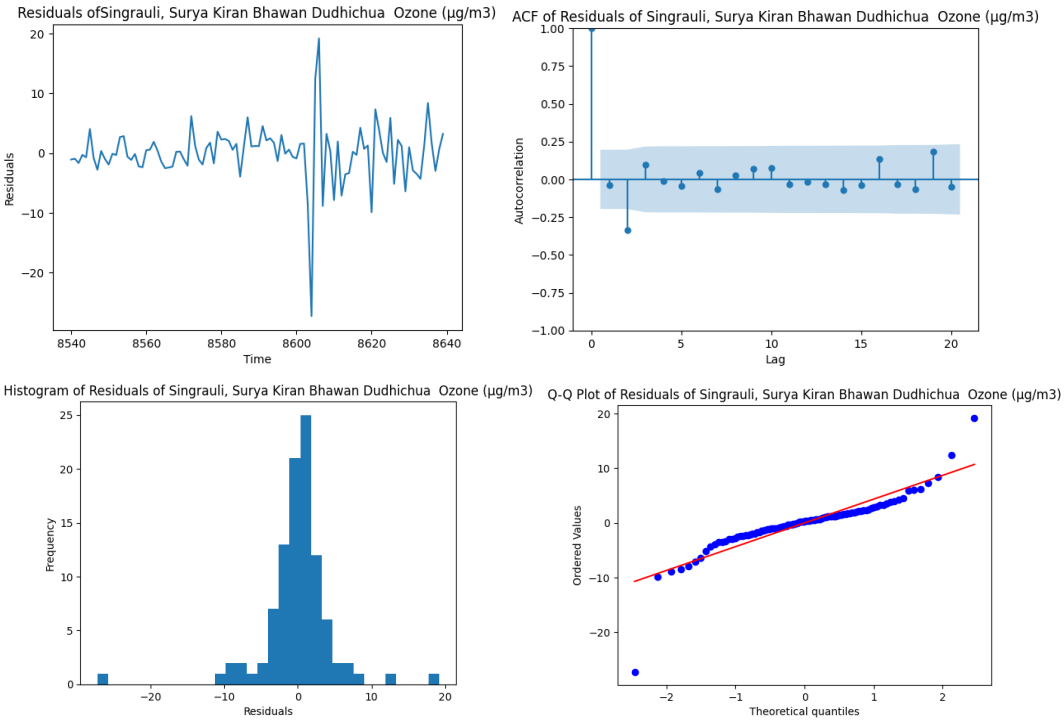

Histogram of Residuals of Singrauli, Surya Kiran Bhawan Dudhichua SO2 (µg/m3)
Frequency
Residuals


Q-Q Plot of Residuals of Singrauli, Surya Kiran Bhawan Dudhichua SO2 (µg/m3)
Ordered Values
Theoretical quantiles

- **NH3**



- **Ozone**



Based on the analysis of the above graphs:

- **Residual Graph:** The graph indicates that the mean of the residuals is in close proximity to zero, suggesting that the model captures the data's central tendency effectively.

- **Autocorrelation Graph:** The graph reveals that the residuals exhibit no significant correlation with each other, implying that the model adequately captures the temporal dependencies in the data.
- **Histogram and Q-Q Plot:** These graphs illustrate that the residuals closely align with the curve of a normal distribution, indicating that the model's residuals follow the expected distribution.

Considering these graphical representations and observations, it can be confidently concluded that the model performs well and provides a good fit for the data.

Information Criteria: Based on the AutoRegressive (AR) model report, compelling evidence emerges to support the effectiveness of the model:

- **AIC and BIC**: Both AIC and BIC exhibit exceptionally low values, indicating a superior model performance. This suggests that the AR model outperforms alternative models when considering the trade-off between goodness-of-fit and complexity.

- **Log likelihood value:** The log likelihood value is notably high, indicating a strong fit between the AR model and the observed data. This supports the notion that the model accurately captures the underlying patterns and dynamics present in the dataset.

Considering these inferences, it is evident that the AR model stands out as a superior choice among other models. The combination of low AIC and BIC values, along with a high log likelihood value, underscores the model's favorable performance.