AIR Q ASSESMENT TN

Phase 5: Project Documentation& Submission:

Project objectives:

• The project involves analyzing air quality data to assess the suitability of air for specific purposes, such as breathing.

• The objective is to identify potential issues or deviations from regulatory standards and determine air probability based on various parameters.

• This project involves analysis objectives , collecting air quality data, designing relevant visualizations, building a predictive model.

- Define objectives such as analyzing air quality trends, identifying pollution hotspots and building a predictive model for RSPM/PM10 levels using air quality dataset .

Dataset Link: https://tn.data.gov.in/resource/location-wise-daily-ambient-air-quality-tamil-nadu-year-2014

Analyzing approach:

- Plan the steps to load the dataset and preprocess , visualizing the dataset and visualizing air quality data using different visualizing techniques.

• Using air quality dataset. This dataset may include information about air pollutants like PM2.5, PM10, CO, SO2, NO2, temperature, humidity, etc.

• Preprocessing the data: Handling missing values, outliers, and ensure it's in a suitable format machine learning.

Example: Checking for missing values using isnull() and notnull()

Filling for missing values using fillnull() and replace()

Loading the dataset:

Load your dataset into a pandas DataFrame, as we discussed in the previous response.

```
import pandas as pd
```

```
Import matplotlib.pyplot as plt

Import seaborn as sns

df=pd.read_csv("C:\\Air Q TN.csv")

print(df.head())
```

This will display the first few rows of the dataset, including column names and some sample data.

Output:

```
        Stn Code Sampling Date       State City/Town/Village/Area  \
0         38      01-02-14  Tamil Nadu                Chennai
1         38      01-07-14  Tamil Nadu                Chennai
2         38      21-01-14  Tamil Nadu                Chennai
3         38      23-01-14  Tamil Nadu                Chennai
4         38      28-01-14  Tamil Nadu                Chennai

                    Location of Monitoring Station  \
0  Kathivakkam, Municipal Kalyana Mandapam, Chennai
1  Kathivakkam, Municipal Kalyana Mandapam, Chennai
2  Kathivakkam, Municipal Kalyana Mandapam, Chennai
3  Kathivakkam, Municipal Kalyana Mandapam, Chennai
4  Kathivakkam, Municipal Kalyana Mandapam, Chennai

                                  Agency Type of Location   SO2   NO2  \
0  Tamilnadu State Pollution Control Board  Industrial Area  11.0  17.0
1  Tamilnadu State Pollution Control Board  Industrial Area  13.0  17.0
2  Tamilnadu State Pollution Control Board  Industrial Area  12.0  18.0
3  Tamilnadu State Pollution Control Board  Industrial Area  15.0  16.0
4  Tamilnadu State Pollution Control Board  Industrial Area  13.0  14.0

   RSPM/PM10  PM 2.5
0       55.0     NaN
1       45.0     NaN
2       50.0     NaN
3       46.0     NaN
4       42.0     NaN
```

```
print(df.info())
```

Output:

```
      <class 'pandas.core.frame.DataFrame'>
RangeIndex: 2879 entries, 0 to 2878
Data columns (total 11 columns):
```

```
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   Stn Code                        2879 non-null   int64
 1   Sampling Date                   2879 non-null   object
 2   State                           2879 non-null   object
 3   City/Town/Village/Area          2879 non-null   object
 4   Location of Monitoring Station  2879 non-null   object
 5   Agency                          2879 non-null   object
 6   Type of Location                2879 non-null   object
 7   SO2                             2868 non-null   float64
 8   NO2                             2866 non-null   float64
 9   RSPM/PM10                       2875 non-null   float64
 10  PM 2.5                          0 non-null      float64
dtypes: float64(4), int64(1), object(6)
memory usage: 247.5+ KB
None
```

print(df.describe())

Output:

```
       Stn Code          SO2          NO2    RSPM/PM10  PM 2.5
count  2879.000000  2868.000000  2866.000000  2875.000000     0.0
mean    475.750261    11.503138    22.136776    62.494261     NaN
std     277.675577     5.051702     7.128694    31.368745     NaN
min      38.000000     2.000000     5.000000    12.000000     NaN
25%     238.000000     8.000000    17.000000    41.000000     NaN
50%     366.000000    12.000000    22.000000    55.000000     NaN
75%     764.000000    15.000000    25.000000    78.000000     NaN
max     773.000000    49.000000    71.000000   269.000000     NaN
```

print(df.isnull().sum())

Output:

```
Stn Code                          0
Sampling Date                     0
State                             0
City/Town/Village/Area            0
Location of Monitoring Station    0
Agency                            0
Type of Location                  0
SO2                              11
NO2                              13
RSPM/PM10                         4
PM 2.5                         2879
dtype: int64
```

print(df['SO2'].mean())

Output:

```
11.503138075313808
```

```
print(df['SO2'].median())
```

Output:

```
12.0
```

Visualization:

The graphic depiction of data and information is known as data visualization. Converting unstructured data into visual representations like maps, charts, graphs, and infographics helps users comprehend the data's patterns, trends, and relationships.

Histogram

Barchart

Heatmap

Linechart

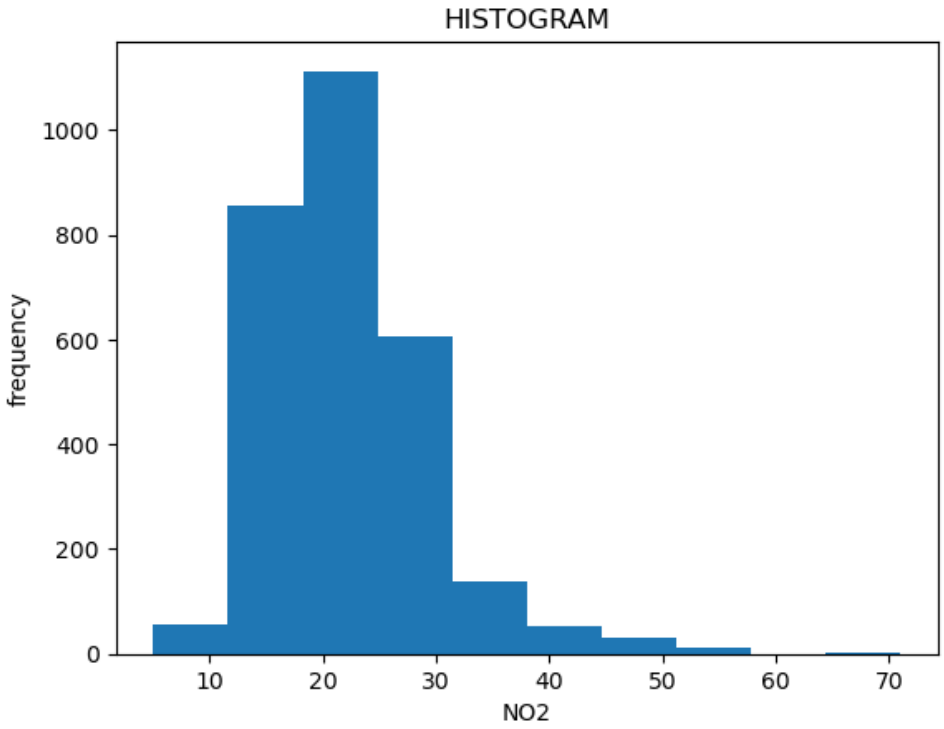
Scatter plot

Histogram:

A histogram is a graph that shows the frequency of numerical data using rectangles. The height of a rectangle (the vertical axis) represents the distribution frequency of a variable (the amount, or how often that variable appears).

```
plt.title("HISTOGRAM")
plt.hist(df["NO2"])
plt.xlabel("NO2")
plt.ylabel("frequency")
plt.show()
```

Output:

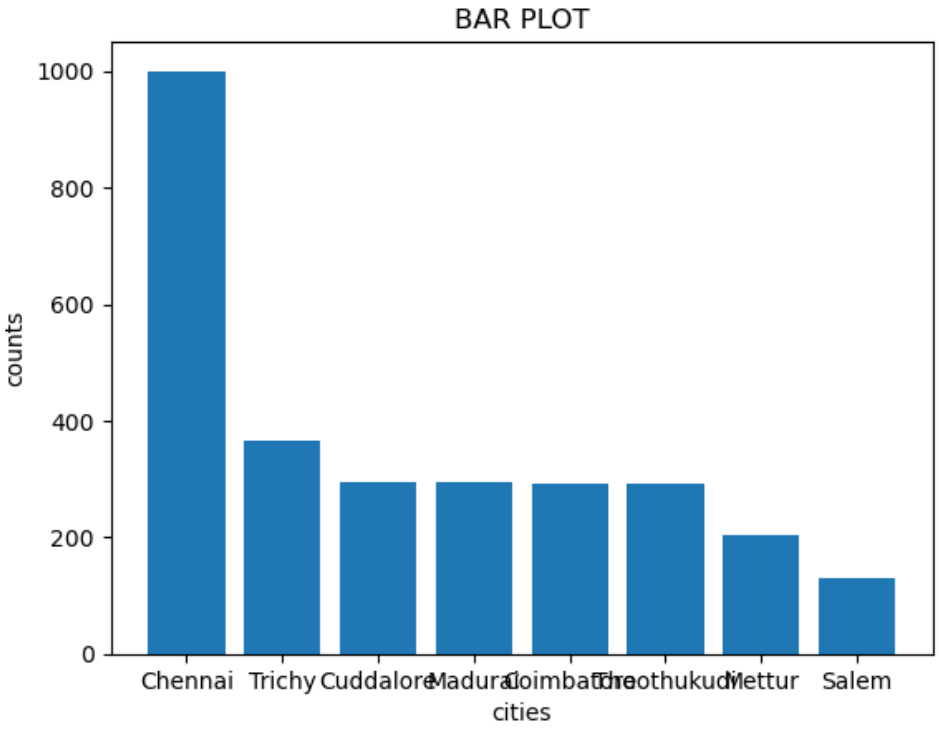


Barchart:

A bar chart or bar graph is a chart or graph that presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally. A vertical bar chart is sometimes called a column chart.

```
plt.title("BAR PLOT")
 x=df["City/Town/Village/Area"].value_counts().nlargest(10)
  plt.bar(x.keys(),x.values)
   plt.xlabel("cities")
   plt.ylabel("counts")
```

Output:

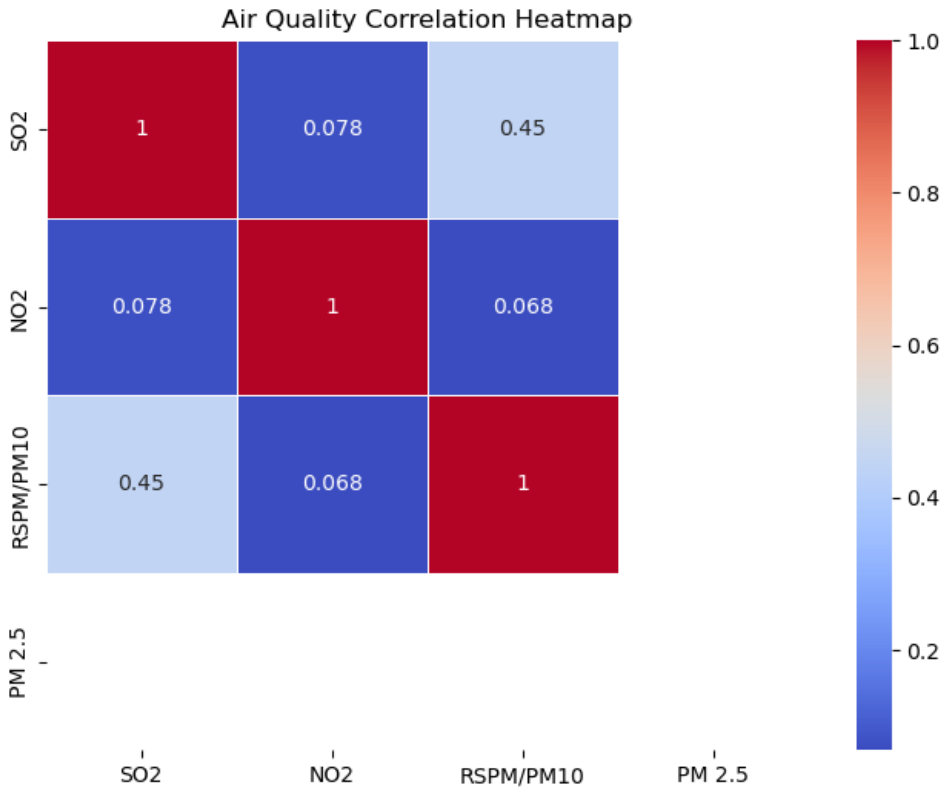


Heatmap:

A heatmap is a graphical representation of data that uses a system of color coding to represent different values. Heatmaps are used in various forms of analytics but are most commonly used to show user behavior on specific web pages or webpage templates

```
numeric_columns = ['SO2', 'NO2', 'RSPM/PM10', 'PM 2.5']

correlation_matrix = df[numeric_columns].corr()

plt.figure(figsize=(8, 6))

sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', linewidths=0.5)

plt.title('Air Quality Correlation Heatmap')

plt.show()
```

Output:

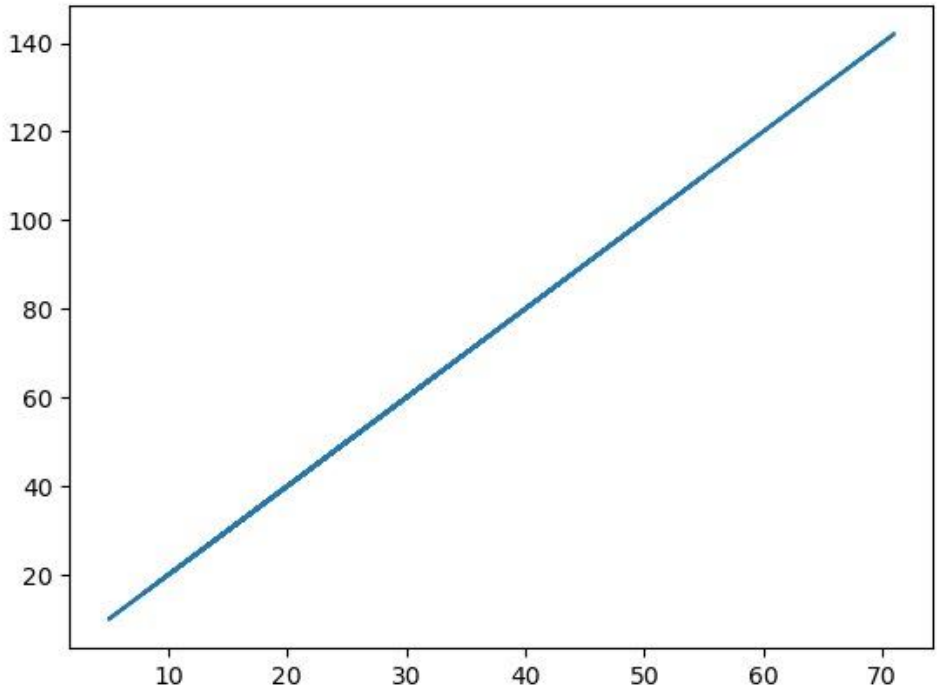


Line plot:

Line plots are used to display numerical, discrete data only, not the continuous data. Line plots organize the data by indicating the occurrences of each value on a number line. These graphs are easily constructed with small data sets, and allow for interpretation based on the frequency patterns that are revealed .

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
df=pd.read_csv("C:\\Air Q TN.csv")
x=df["NO2"]
y=x*2
plt.plot(x,y)
plt.show()
```

Output:

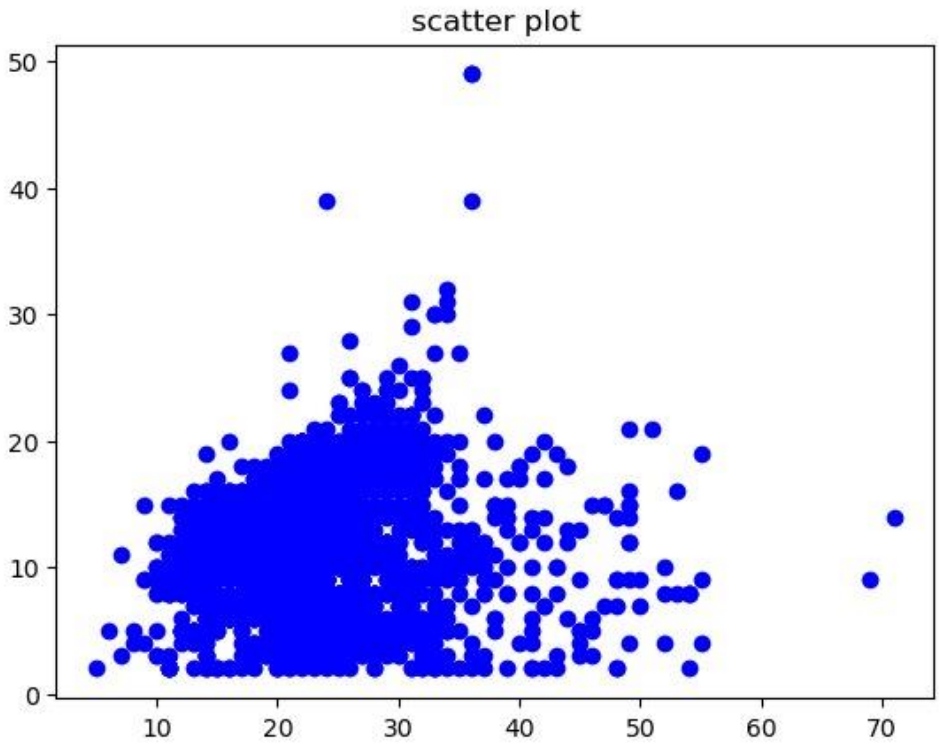


Scatter plot:

The collected data of the temperature and humidity can be presented in the form of a scatter plot. Temperature is marked on the x-axis and humidity is on the y-axis. To calculate the humidity at a temperature of 60 degrees Fahrenheit, we need to first draw a line of best fit.

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
df=pd.read_csv("C:\\Air Q TN.csv")
plt.title("scatter plot")
plt.scatter(df["NO2"],df["SO2"],color='blue')
plt.show()
```

Output:



Conclusion:

Concluding a project of air quality analysis involves the project definition and design thinking with various methods and techniques along the subject of the given project.

Thus we have outlined our project objectives with project definition,outlining the analysis approach by selecting the appropriate visualization techniques.