DATA ANALTYICS WITH COGNOS

PROJECT: AIR QUALITY ANALYSIS TAMILNADU

PHASE 4 PROJECT

TEAM MEMBERS:

1. S. BALAKANDAN – 820421205012 – balakandan50@gmail.com
2. R. BHARATH RAJ – 820421205015 – bharathrajravi2004@gmail.com
3. V. ARUN KUMAR – 820421205008 – arurarun0088@gmail.com
4. M. MUKILAN – 820421205042 – mukilmuruga07@gmail.com
5. T. LOGESH – 820421205039 – logeshselvan6@gmail.com

Table of Contents:

1. Introduction:

Recently, much has been discussed about air pollution and its consequences on the environment. These discussion always gain prominence when some of their consequences haunt the world and leave us wondering what will be of future generation. Air quality is a critical concern for public health and environmental well-being. Tamil Nadu have PM 2.5 of 28.2 microgram/m3 levels. This Project emphasis by analysing and pre-processing the air quality dataset which is essential for informed decision-making and effective pollution control measures.

2. Problem Statement:

In this part we will build our project to:

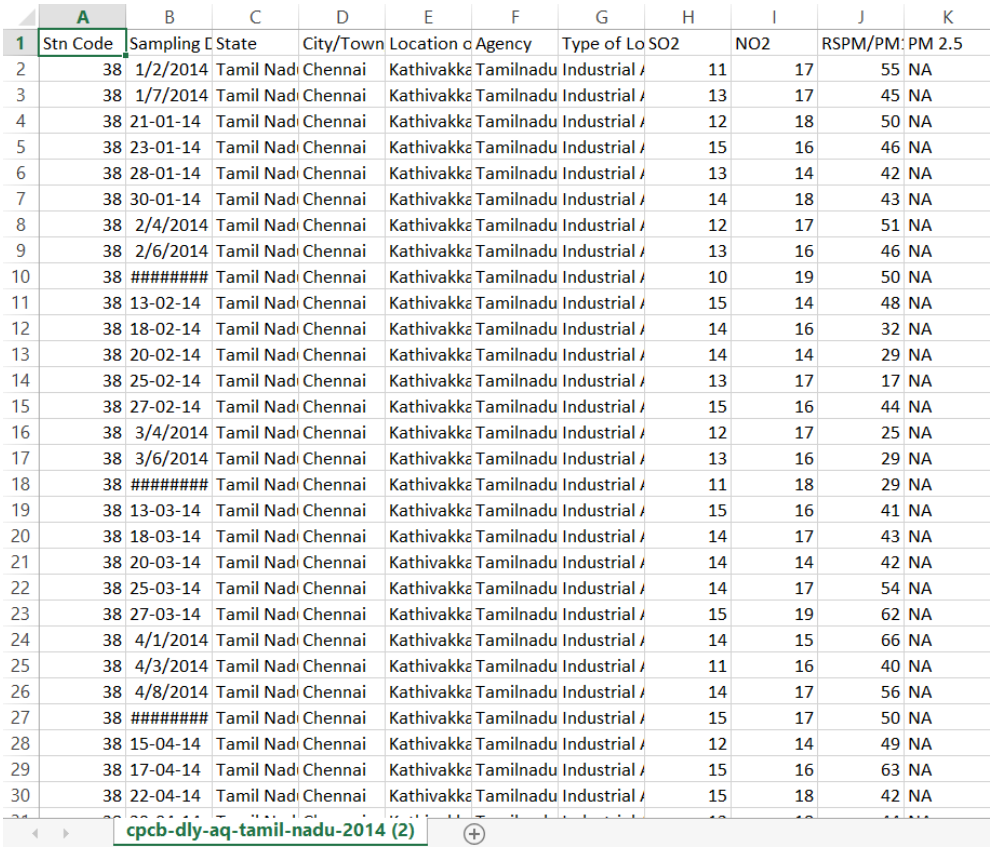
- Perform the air quality analysis and create visualizations.
- Calculate average SO2, NO2, and RSPM/PM10 levels across different monitoring stations, cities, or areas. Identify pollution trends and areas with high pollution levels.
- Create visualizations using data visualization libraries (e.g., Matplotlib, Seaborn)

3. Data Collection:

The dataset containing location-wise daily ambient air quality records for Tamil Nadu in the year 2014 has been obtained from the below datalink.

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

Dataset:



	A	B	C	D	E	F	G	H	I	J	K
1	Stn Code	Sampling D	State	City/Town	Location o	Agency	Type of Lo	SO2	NO2	RSPM/PM	PM 2.5
2	38	1/2/2014	Tamil Nad	Chennai	Kathivakka	Tamilnadu	Industrial A	11	17	55	NA
3	38	1/7/2014	Tamil Nad	Chennai	Kathivakka	Tamilnadu	Industrial A	13	17	45	NA
4	38	21-01-14	Tamil Nad	Chennai	Kathivakka	Tamilnadu	Industrial A	12	18	50	NA
5	38	23-01-14	Tamil Nad	Chennai	Kathivakka	Tamilnadu	Industrial A	15	16	46	NA
6	38	28-01-14	Tamil Nad	Chennai	Kathivakka	Tamilnadu	Industrial A	13	14	42	NA
7	38	30-01-14	Tamil Nad	Chennai	Kathivakka	Tamilnadu	Industrial A	14	18	43	NA
8	38	2/4/2014	Tamil Nad	Chennai	Kathivakka	Tamilnadu	Industrial A	12	17	51	NA
9	38	2/6/2014	Tamil Nad	Chennai	Kathivakka	Tamilnadu	Industrial A	13	16	46	NA
10	38	########	Tamil Nad	Chennai	Kathivakka	Tamilnadu	Industrial A	10	19	50	NA
11	38	13-02-14	Tamil Nad	Chennai	Kathivakka	Tamilnadu	Industrial A	15	14	48	NA
12	38	18-02-14	Tamil Nad	Chennai	Kathivakka	Tamilnadu	Industrial A	14	16	32	NA
13	38	20-02-14	Tamil Nad	Chennai	Kathivakka	Tamilnadu	Industrial A	14	14	29	NA
14	38	25-02-14	Tamil Nad	Chennai	Kathivakka	Tamilnadu	Industrial A	13	17	17	NA
15	38	27-02-14	Tamil Nad	Chennai	Kathivakka	Tamilnadu	Industrial A	15	16	44	NA
16	38	3/4/2014	Tamil Nad	Chennai	Kathivakka	Tamilnadu	Industrial A	12	17	25	NA
17	38	3/6/2014	Tamil Nad	Chennai	Kathivakka	Tamilnadu	Industrial A	13	16	29	NA
18	38	########	Tamil Nad	Chennai	Kathivakka	Tamilnadu	Industrial A	11	18	29	NA
19	38	13-03-14	Tamil Nad	Chennai	Kathivakka	Tamilnadu	Industrial A	15	16	41	NA
20	38	18-03-14	Tamil Nad	Chennai	Kathivakka	Tamilnadu	Industrial A	14	17	43	NA
21	38	20-03-14	Tamil Nad	Chennai	Kathivakka	Tamilnadu	Industrial A	14	14	42	NA
22	38	25-03-14	Tamil Nad	Chennai	Kathivakka	Tamilnadu	Industrial A	14	17	54	NA
23	38	27-03-14	Tamil Nad	Chennai	Kathivakka	Tamilnadu	Industrial A	15	19	62	NA
24	38	4/1/2014	Tamil Nad	Chennai	Kathivakka	Tamilnadu	Industrial A	14	15	66	NA
25	38	4/3/2014	Tamil Nad	Chennai	Kathivakka	Tamilnadu	Industrial A	11	16	40	NA
26	38	4/8/2014	Tamil Nad	Chennai	Kathivakka	Tamilnadu	Industrial A	14	17	56	NA
27	38	########	Tamil Nad	Chennai	Kathivakka	Tamilnadu	Industrial A	15	17	50	NA
28	38	15-04-14	Tamil Nad	Chennai	Kathivakka	Tamilnadu	Industrial A	12	14	49	NA
29	38	17-04-14	Tamil Nad	Chennai	Kathivakka	Tamilnadu	Industrial A	15	16	63	NA
30	38	22-04-14	Tamil Nad	Chennai	Kathivakka	Tamilnadu	Industrial A	15	18	42	NA

cpcb-dly-aq-tamil-nadu-2014 (2)

4. Data Cleansing:

Data cleaning is one of the important parts of machine learning. It plays a significant part in building a model.

Data pre-processing is a crucial step in preparing the dataset for machine learning.

Data Cleansing involves:

- Identifying and removing any missing, duplicate or irrelevant data.
- Handling missing data.
- Removing outliers.
- Scaling the data.

Using Methods Such as:

- dropna()
- drop()
- drop_duplicates()
- scale()
- get_dummies()

5. Air Quality Index:

Air quality is a measure of how clean or polluted the air is. Monitoring air quality is important because polluted air can be bad for our health—and the health of the environment. Air quality is measured with the Air Quality Index, or AQI. The AQI works like a thermometer that runs from 0 to 500 degrees. However, instead of showing changes in the temperature, the AQI is a way of showing changes in the amount of pollution in the air.

Air Quality Index Levels of Health Concern	Numerical Value	Meaning
Good	0 to 50	Air quality is considered satisfactory, and air pollution poses little or no risk.
Moderate	51 to 100	Air quality is acceptable; however, for some pollutants there may be a moderate health concern for a very small number of people who are unusually sensitive to air pollution.
Unhealthy for Sensitive Groups	101 to 150	Members of sensitive groups may experience health effects. The general public is not likely to be affected.
Unhealthy	151 to 200	Everyone may begin to experience health effects; members of sensitive groups may experience more serious health effects.
Very Unhealthy	201 to 300	Health alert: everyone may experience more serious health effects.
Hazardous	301 to 500	Health warnings of emergency conditions. The entire population is more likely to be affected.

6. Program:

- First, import the required Python libraries. As given below.

```
import pandas as pd
import plotly.express as px
import plotly.io as pio
import plotly.graph_objects as go
```

✓ 1.2s

- To import the database into the jupyter notebook, you can use the following Python code given below.

```
# In[2]:data =pd.read_csv(r"C:\Users\Administrator\Downloads\airquality.csv")


# In[3]:print(data.head())
```

 # OP [3]:

```
   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
```

- Data Cleansing and Transformation are done by following Python Code.
  - drop() - To drop an entire Column.
  - dropna() - To drop a the NaN Values.
  - drop_duplicates() - To drop the Duplicates in the Dataset.
  - scale() - In cases where all the columns have a significant difference in their scales, are needed to be modified in such a way that all those values fall into the same scale.
  - get_dummies() - indicate whether each row in the original dataset belongs to a particular category or not.

```
# In[3]: df=data.drop(['PM 2.5'],axis=1)
# In[4]: df.head()

# OP[4]:
```

	Stn Code	Sampling Date	State	City/Town/Village/Area	Location of Monitoring Station	Agency	Type of Location	SO2	NO2	RSPM/PM10
0	38	01-02-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	11.0	17.0	55.0
1	38	01-07-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	13.0	17.0	45.0
2	38	21-01-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	12.0	18.0	50.0
3	38	23-01-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	15.0	16.0	46.0
4	38	28-01-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	13.0	14.0	42.0

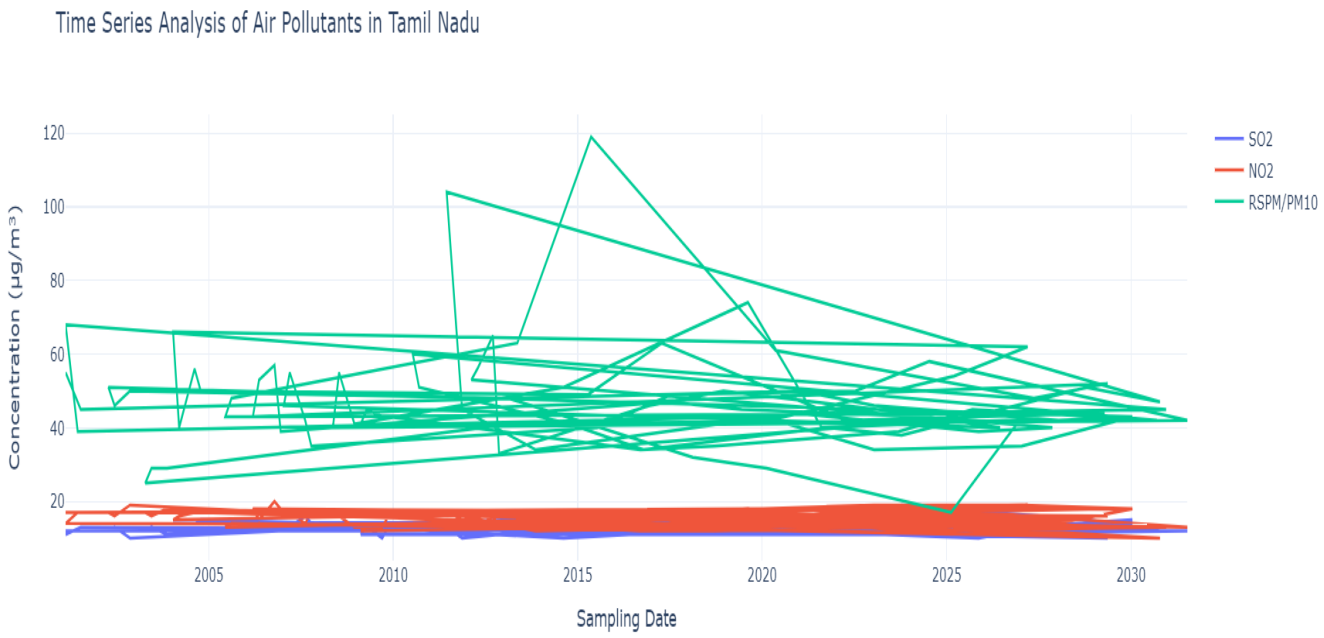
```
# In[5]: newf=df.dropna()
         newd=newf.drop_duplicates()

# In[6]: newf.head()
         newf.scale()
         newf.get_dummies()

# OP[6]:
```

Stn Code	Sampling Date	State	City/Town/Village/Area	Location of Monitoring Station	Agency	Type of Location	SO2	NO2	RSPM/PM10	
0	38	01-02-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	11.0	17.0	55.0
1	38	01-07-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	13.0	17.0	45.0
2	38	21-01-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	12.0	18.0	50.0
3	38	23-01-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	15.0	16.0	46.0
4	38	28-01-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	13.0	14.0	42.0

- Constructing a Time series plot for each Air Pollution using the data visualization library-pyplot:

```
# time series plot for each air pollutant
fig = go.Figure()

for pollutant in ['NO2', 'SO2', 'RSPM/PM10']:
    fig.add_trace(go.Scatter(x=data['date'], y=data[pollutant], mode='lines',
                             name=pollutant))

fig.update_layout(title='Time Series Analysis of Air Pollutants in Delhi',
                  xaxis_title='Date', yaxis_title='Concentration (µg/m³)')
fig.show()
```

Time Series Analysis of Air Pollutants in Tamil Nadu:



- Calculating Air Quality Index using RSPM/PM10 and Categorizing Air Quality Index :

```
# Define AQI breakpoints and corresponding AQI values
aqi_breakpoints = [
    (0, 12.0, 50), (12.1, 35.4, 100), (35.5, 55.4, 150),
    (55.5, 150.4, 200), (150.5, 250.4, 300), (250.5, 350.4, 400),
    (350.5, 500.4, 500)
]

def calculate_aqi(pollutant_name, concentration):
    for low, high, aqi in aqi_breakpoints:
        if low <= concentration <= high:
            return aqi
    return None

def calculate_overall_aqi(row):
    aqi_values = []
    pollutants = ['NO2', 'SO2', 'RSPM/PM10']
    for pollutant in pollutants:
        aqi = calculate_aqi(pollutant, row[pollutant])
        if aqi is not None:
            aqi_values.append(aqi)
    return max(aqi_values)

# Calculate AQI for each row
data['AQI'] = data.apply(calculate_overall_aqi, axis=1)

    # Define AQI categories
    aqi_categories = [(0, 50, 'Good'), (51, 100, 'Moderate'), (101,
150, 'Unhealthy forSensitive Groups'),
    (151, 200, 'Unhealthy'), (201, 300, 'Very Unhealthy'), (301, 500,
'Hazardous')
]

def categorize_aqi(aqi_value):
    for low, high, category in aqi_categories:
        if low <= aqi_value <= high:
            return category
    return None

# Categorize AQI
data['AQI Category'] = data['AQI'].apply(categorize_aqi)
print(data.head())
```

AQI Categories:

```
   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  AQI                    AQI_Category
0       55.0  150  Unhealthy for Sensitive Groups
1       45.0  150  Unhealthy for Sensitive Groups
2       50.0  150  Unhealthy for Sensitive Groups
3       46.0  150  Unhealthy for Sensitive Groups
4       42.0  150  Unhealthy for Sensitive Groups
```

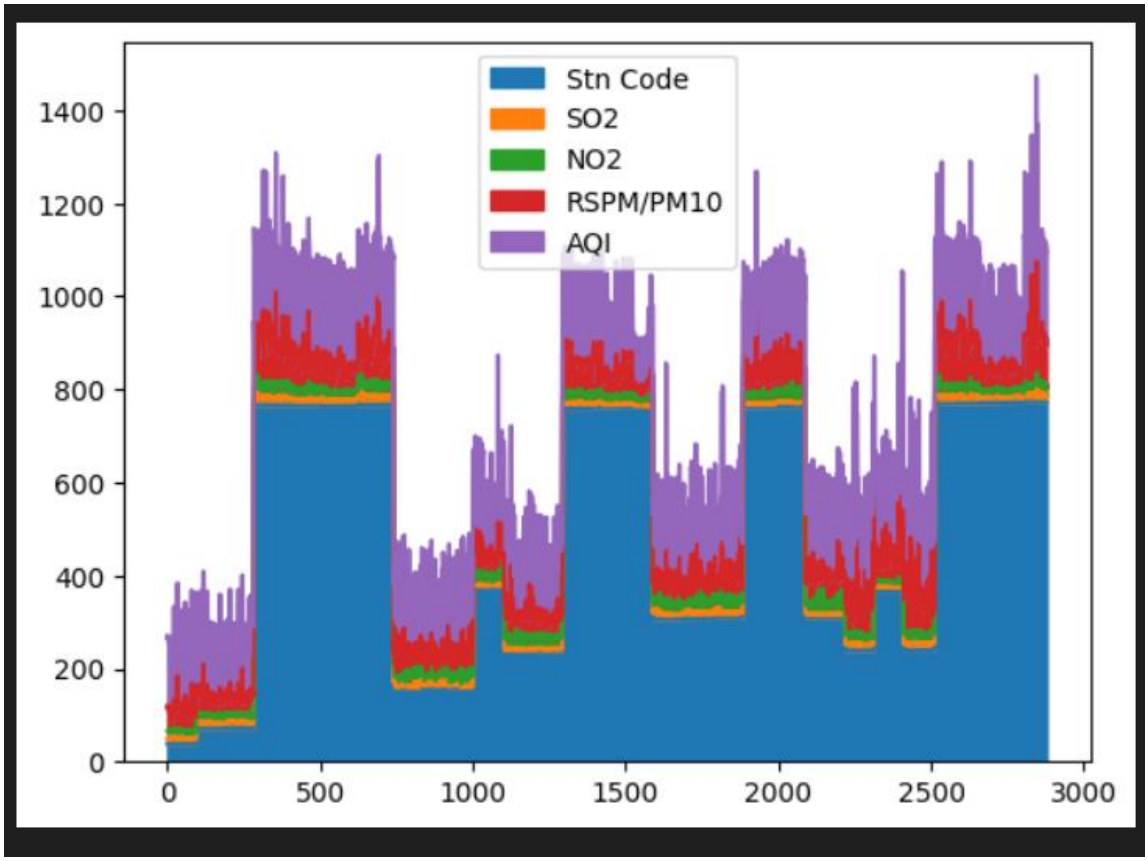
- Identify pollution trends and areas with high pollution levels by highlight function:

```
def highlight_unhealthy(val):
    color = 'red' if val == 'Unhealthy' else ''
    return f'background-color: {color}'

# Apply the highlight function to the 'AQI Category' column
styled_df1= newd.style.apply({'AQI Category': highlight_unhealthy}, subset=pd.IndexSlice[newd[newd['AQI Category'] == 'Unhealthy'].index, 'AQI Category'])

# Display the styled DataFrame
styled_df1
```

Areas with High Pollution Levels:

20	38	25-03-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	14.000000	17.000000	54.000000	150	Unhealthy for Sensitive Groups
21	38	27-03-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	15.000000	19.000000	62.000000	200	Unhealthy
22	38	04-01-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	14.000000	15.000000	66.000000	200	Unhealthy
23	38	04-03-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	11.000000	16.000000	40.000000	150	Unhealthy for Sensitive Groups
24	38	04-08-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	14.000000	17.000000	56.000000	200	Unhealthy
25	38	04-10-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	15.000000	17.000000	50.000000	150	Unhealthy for Sensitive Groups
26	38	15-04-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	12.000000	14.000000	49.000000	150	Unhealthy for Sensitive Groups
27	38	17-04-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	15.000000	16.000000	63.000000	200	Unhealthy
28	38	22-04-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	15.000000	18.000000	42.000000	150	Unhealthy for Sensitive Groups
29	38	29-04-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	12.000000	18.000000	44.000000	150	Unhealthy for Sensitive Groups
30	38	05-06-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	13.000000	13.000000	43.000000	150	Unhealthy for Sensitive Groups
31	38	05-08-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	14.000000	14.000000	48.000000	150	Unhealthy for Sensitive Groups
32	38	13-05-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	14.000000	13.000000	63.000000	200	Unhealthy
33	38	15-05-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	12.000000	15.000000	119.000000	200	Unhealthy
34	38	20-05-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	13.000000	18.000000	61.000000	200	Unhealthy

- Area plot for the Dataset



9.Conclusion:

The proposed approach aims to enhance the accuracy of predictive models for ambient air quality in Tamil Nadu through the incorporation of machine learning algorithms. The success of this project will lead to better air quality predictions, enabling more effective pollution control measures and safeguarding public health and the environment.