Report

Air Quality Assessment of TamilNadu

Project Definition:

The project aims to analyze and visualize air quality data from monitoring stations in Tamil Nadu. The objective is to gain insights of air pollution trends, identify areas with high pollution levels and develop a predictive model to estimate RSPM/PM10 levels based on SO2 and NO2 levels. This project involves defining objectives, designing the analysis approach , selecting visualization techniques and creating a predictive model using python and relevant libraries. Collecting a air quality dataset, designing relevant visualizations in IBM Cognas and deriving insights from the data.

Project Objective:

The specific objectives of analyzing and estimating RSPM/PM10 levels based on SO2 and NO2 levels using the air quality datasets of Tamil Nadu.

Public Health Protection: One of the primary objectives of air quality assessment is to safeguard public health by identifying and monitoring harmful pollutants in the air. This is crucial for preventing respiratory diseases, cardiovascular issues, and other health problems related to poor air quality.

Environmental Protection: Assessing air quality helps in understanding the impact of pollutants on the environment. This knowledge can inform policies and initiatives aimed at preserving the local ecosystem, including plant and animal life, as well as natural resources.

Compliance with Regulatory Standards: Monitoring air quality allows authorities to ensure compliance with local, national, and international regulatory standards. These standards often include limits on pollutant concentrations that are deemed safe for human health and the environment.

Policy Formulation: Data from air quality assessments can guide the formulation of policies aimed at reducing air pollution. These policies might include measures to control industrial emissions, promote the use of cleaner energy sources, and encourage the adoption of sustainable transportation systems.

Awareness and Education: Assessments of air quality can help raise public awareness about the impacts of air pollution on health and the environment. This can lead to educational campaigns and community initiatives that

promote sustainable practices and encourage behavioral changes to reduce pollution.

Long-Term Planning: By continuously monitoring air quality, policymakers can develop long-term strategies to improve air quality and mitigate the impacts of pollution. Such planning might involve the implementation of sustainable urban development, green infrastructure, and the promotion of renewable energy sources. **Risk Assessment and Management**: Air quality assessments aid in identifying high-risk areas and populations vulnerable to the adverse effects of air pollution. This information enables authorities to implement targeted measures to reduce exposure and manage risks effectively.

International Commitments: Many countries, including India, are signatories to international agreements and protocols that require regular monitoring and reporting of air quality data. Assessments are crucial for fulfilling these commitments and participating in global efforts to combat air pollution and climate change.

Project Summary:

The air quality assessment of tamilnadu began with loading the datset and preprocessing it . The dataset used for air quality analysis project is a csv file with 11 columns represents station code, sampling date, state, city , location of agency,SO2 and NO2 content and RSPM/PM rates. Clean and preprocess the collected data. This includes handling missing values, removing duplicates, and converting data into a suitable format for analysis. Consider normalizing or scaling the data if the features have different scales.

Then several visualization techniques have been on the dataset.

It includes usage of python manuplation libraries :

Matplotlib: Matplotlib is a widely-used Python library for creating static, animated, and interactive visualizations in a wide range of formats. It is highly versatile and can be used in various domains, including data science, scientific research, engineering, and more.

Pandas: Pandas is a popular Python library for data manipulation and analysis. It provides data structures and functions for working with structured data, such as tables, spreadsheets, and time series data.

Seaborn: Seaborn is a python data visualization library built on top of Matplotlib.It provides a high-level interface

for creating atrractive and informative statistical graphics Seaborn is particularlyuseful for visualizing complex relationships in your data.

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

How to clean a dataset with coding:

Loading and preprocessing the dataset

Stn Code	Sampling	State	City/Towr	Location c	Agency	Type of Lc	SO2	NO2	RSPM/PM	SPM
38	5/1/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	9.6	17.16667	73.33333	149.6667
38	7/1/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.15	20.28333	61.33333	150.3333
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	12.45	20.51667	75	114.6667
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	10.75	18.18333	120	197.6667
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	9.78	17.32	96.5	216
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	17.11667	20.86667	81.33333	159.3333
38	2/2/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.48333	20.66667	82.33333	185.6667
38	4/2/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	10.93333	18.65	79	179.3333
38	9/2/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.36667	19.5	96.66667	181.3333
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.16667	18.38333	52.66667	125.6667
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	10.18333	18.63333	50.66667	128.3333
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.56667	21.48333	85	157
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.55	19.78333	88.66667	272.3333
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	14.71667	22.33333	83	197.3333
38	2/3/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	9.633333	16.61667	94.66667	193.3333
38	4/3/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.38333	16.58333	118.3333	335.3333
38	9/3/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	12.7	22.13333	56	173.6667
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	8.65	14.36667	53.33333	107.3333
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.71667	19.36667	69.66667	167.3333
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	10.18333	16.6	111	298.6667
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.41667	20.15	58	222.3333
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	15.35	20.5	67.66667	279.3333

air quality

Printing the Intial structure of the Dataset

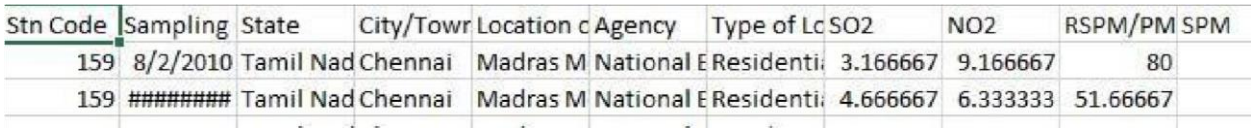
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	16	19	117	NA	
773	5/6/2015	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	20	23	120	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	10	14	106	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	17	20	120	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	22	24	123	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	19	22	118	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	13	20	124	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	15	22	141	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	13	20	137	NA	
773	8/5/2015	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	15	23	141	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	14	23	144	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	13	20	134	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	14	20	136	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	NA	NA	102	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	13	22	142	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	11	18	80	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	8	17	116	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	13	24	121	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	13	22	118	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	NA	NA	104	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	14	24	132	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	15	25	147	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	15	24	138	NA	

air quality

8332	237	1/8/2014	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	35	35	NA
8333	237	5/8/2014	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	30	48	NA
8334	237	8/8/2014	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	32	41	NA
8335	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	26	41	NA
8336	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	27	47	NA
8337	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	25	45	NA
8338	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	7	23	59	NA
8339	237	2/9/2014	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	29	55	NA
8340	237	5/9/2014	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	30	51	NA
8341	237	9/9/2014	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	34	35	NA
8342	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	28	40	NA
8343	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	25	40	NA
8344	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	25	43	NA
8345	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	27	78	NA
8346	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	25	47	NA
8347	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	33	37	NA
8348	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	23	40	NA
8349	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	26	47	NA
8350	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	22	50	NA
8351	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	24	43	NA
8352	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	29	59	NA
8353	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	25	46	NA
8354	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	6	28	46	NA

air quality

Introduction to the attribute given in the datasets TYPE OF PARAMETER GIVEN IN AIR QUALITY DATASETS OF TAMIL NADU:



Stn Code	Sampling	State	City/Towr	Location c	Agency	Type of Lc	SO2	NO2	RSPM/PM	SPM
159	8/2/2010	Tamil Nad	Chennai	Madras M	National E	Residenti	3.166667	9.166667	80	
159	########	Tamil Nad	Chennai	Madras M	National E	Residenti	4.666667	6.333333	51.66667	

1 .STN CODE: It refers to station code of the Tamil Nadu

District.

1. 0-150 : Unhealthy and senstive

2. 150-200:unhealthy

3. 201-300: very unhealthy

1. 2. SAMPLING Date: It refers to the sampling date at which No2 and so2 measured.

3. STATE: It refers to the Tamil Nadu.

4. City/Town/village: List of village givens in the datasets.

1. Chennai
2. Coimbatore
3. Cuddalore

4. Madurai
5. Thoothukudi
6. Salem
7. Trichy
8. Mettur

LOCATION: It refers to the Location the so2 and no2 measured.

Agency: Refers to the Industry emitting No2 and so2

Type of Location: It refers to the type of location such resident and street .

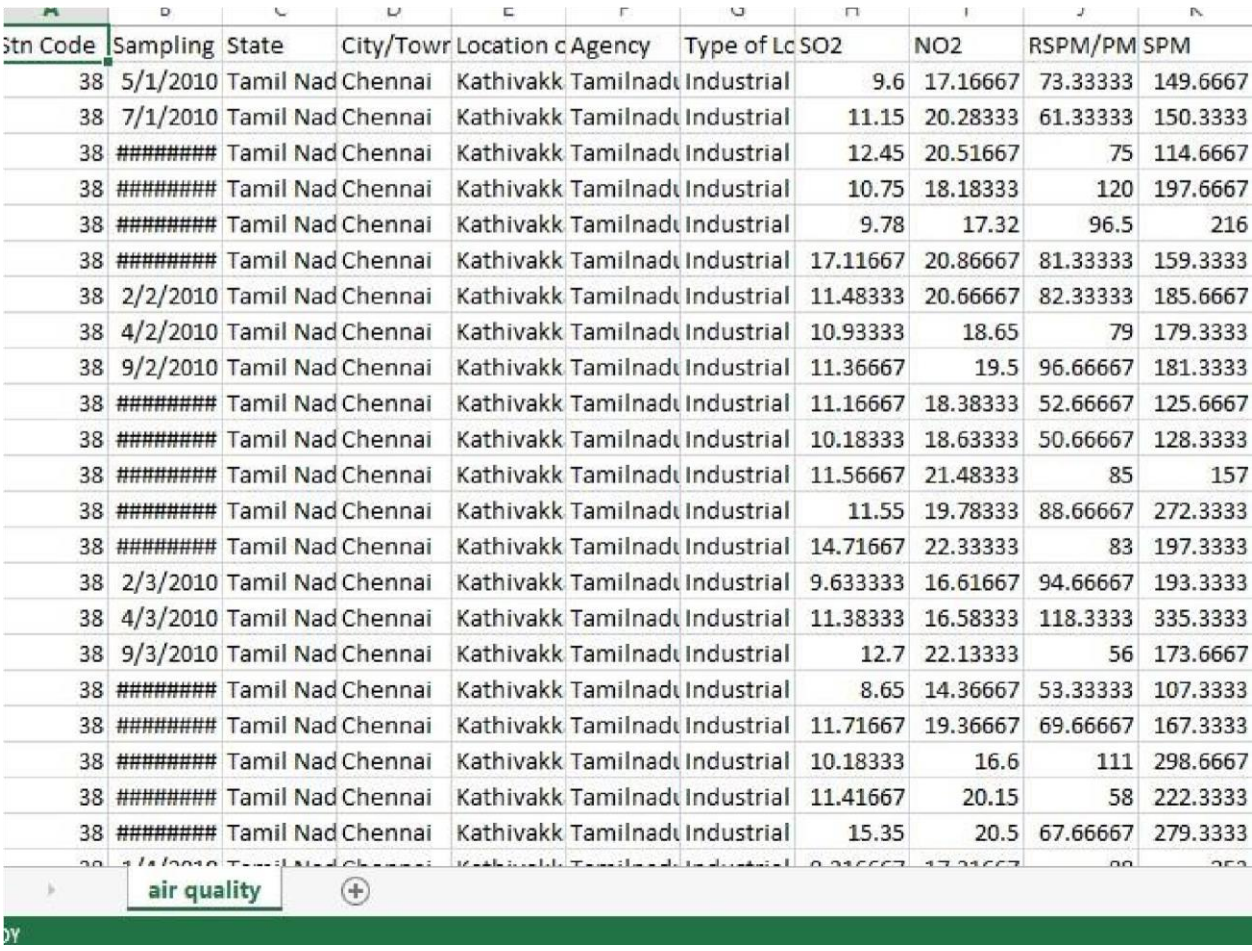
SO2: It refers to the sulphur dioxide.

NO2: It refers to the Nitrogen dioxide.

RSPM: Respirable suspended Particulate Matter.

SPM: Suspended Particulate Matter.

Loading the Intial structure of the Datasets without preprocessing



Stn Code	Sampling	State	City/Towr	Location c	Agency	Type of Lc	SO2	NO2	RSPM/PM	SPM
38	5/1/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	9.6	17.16667	73.33333	149.6667
38	7/1/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.15	20.28333	61.33333	150.3333
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	12.45	20.51667	75	114.6667
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	10.75	18.18333	120	197.6667
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	9.78	17.32	96.5	216
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	17.11667	20.86667	81.33333	159.3333
38	2/2/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.48333	20.66667	82.33333	185.6667
38	4/2/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	10.93333	18.65	79	179.3333
38	9/2/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.36667	19.5	96.66667	181.3333
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.16667	18.38333	52.66667	125.6667
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	10.18333	18.63333	50.66667	128.3333
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.56667	21.48333	85	157
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.55	19.78333	88.66667	272.3333
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	14.71667	22.33333	83	197.3333
38	2/3/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	9.633333	16.61667	94.66667	193.3333
38	4/3/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.38333	16.58333	118.3333	335.3333
38	9/3/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	12.7	22.13333	56	173.6667
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	8.65	14.36667	53.33333	107.3333
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.71667	19.36667	69.66667	167.3333
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	10.18333	16.6	111	298.6667
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.41667	20.15	58	222.3333
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	15.35	20.5	67.66667	279.3333

air quality

773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	16	19	117	NA	
773	5/6/2015	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	20	23	120	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	10	14	106	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	17	20	120	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	22	24	123	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	19	22	118	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	13	20	124	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	15	22	141	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	13	20	137	NA	
773	8/5/2015	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	15	23	141	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	14	23	144	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	13	20	134	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	14	20	136	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	NA	NA	102	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	13	22	142	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	11	18	80	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	8	17	116	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	13	24	121	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	13	22	118	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	NA	NA	104	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	14	24	132	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	15	25	147	NA	
773	########	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	15	24	138	NA	

air quality

8332	237	1/8/2014	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	35	35	NA
8333	237	5/8/2014	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	30	48	NA
8334	237	8/8/2014	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	32	41	NA
8335	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	26	41	NA
8336	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	27	47	NA
8337	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	25	45	NA
8338	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	7	23	59	NA
8339	237	2/9/2014	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	29	55	NA
8340	237	5/9/2014	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	30	51	NA
8341	237	9/9/2014	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	34	35	NA
8342	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	28	40	NA
8343	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	25	40	NA
8344	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	25	43	NA
8345	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	27	78	NA
8346	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	25	47	NA
8347	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	33	37	NA
8348	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	23	40	NA
8349	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	26	47	NA
8350	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	22	50	NA
8351	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	24	43	NA
8352	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	29	59	NA
8353	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	25	46	NA
8354	237	########	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	6	28	46	NA

air quality

Preprocessing of the datsets using python Manipulation Libraries Using pandas and Other libraries:

```
mport pandas as pd


# Load the dataset df =
pd.read_csv('air quality.csv')


# Display the initial structure of
the dataset print("Initial structure
of the dataset:") print(df.head())


# Handling missing values
# Drop rows with missing
values
df.dropna(inplace=True)


# Removing duplicates
df.drop_duplicates(inplace=True)


# Dealing with outliers (you can use
statistical methods like Z-score or IQR)

# Example: Removing outliers in a specific
column
```

```
# Standardizing data formats (converting to appropriate data types if necessary)
# Example: Converting a column to a datetime format df['Sampling Date']
= pd.to_datetime(df['Sampling Date'])
# After cleaning, you might want to save the cleaned dataset to a new file.
df.to_csv('cleaned_dataset.csv', index=False)
```

```
Initial structure of the dataset:
  Stn Code Sampling Date       State City/Town/Village/Area  \
0     38      5/1/2010  Tamil Nadu                 Chennai
1     38      7/1/2010  Tamil Nadu                 Chennai
2     38     12/1/2010  Tamil Nadu                 Chennai
```

3 38 1/19/2010 Tamil Nadu

Chennai 4 38 1/21/2010 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 9.60 17.166667

1 Tamilnadu State Pollution Control Board Industrial Area 11.15

```
  20.283333
2 Tamilnadu State Pollution Control
  Board  Industrial Area  12.45
  20.516667
3 Tamilnadu State Pollution Control
  Board  Industrial Area  10.75
  18.183333
4 Tamilnadu State Pollution Control
  Board  Industrial Area   9.78
  17.320000


    RSPM/PM10         SPM
0      73.333333  149.666667
1      61.333333  150.333333
2      75.000000  114.666667
3      120.000000  197.666667   4   96.500000
216.000000   Cleaned dataset:
   Stn Code Sampling Date       State
City/Town/Village/Area  \
0      38    2010-05-01  Tamil Nadu
Chennai
1      38    2010-07-01  Tamil Nadu
Chennai
2      38    2010-12-01  Tamil Nadu
```

Chennai
3 38 2010-01-19 Tamil Nadu
Chennai 4 38 2010-01-21
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 9.60
17.166667
1 Tamilnadu State Pollution Control

```
  Board  Industrial Area  11.15
  20.283333
2 Tamilnadu State Pollution Control
  Board  Industrial Area  12.45
  20.516667
3 Tamilnadu State Pollution Control
  Board  Industrial Area  10.75
  18.183333
4 Tamilnadu State Pollution Control
  Board  Industrial Area   9.78
  17.320000

    RSPM/PM10        SPM
0  73.333333  149.666667
1  61.333333  150.333333
2  75.000000  114.666667
3  120.000000  197.666667
4  96.500000  216.000000
```

Preprocess Information after removing Null Values and NAN Values:

```
<class
'pandas.core.frame.DataFrame'>
```

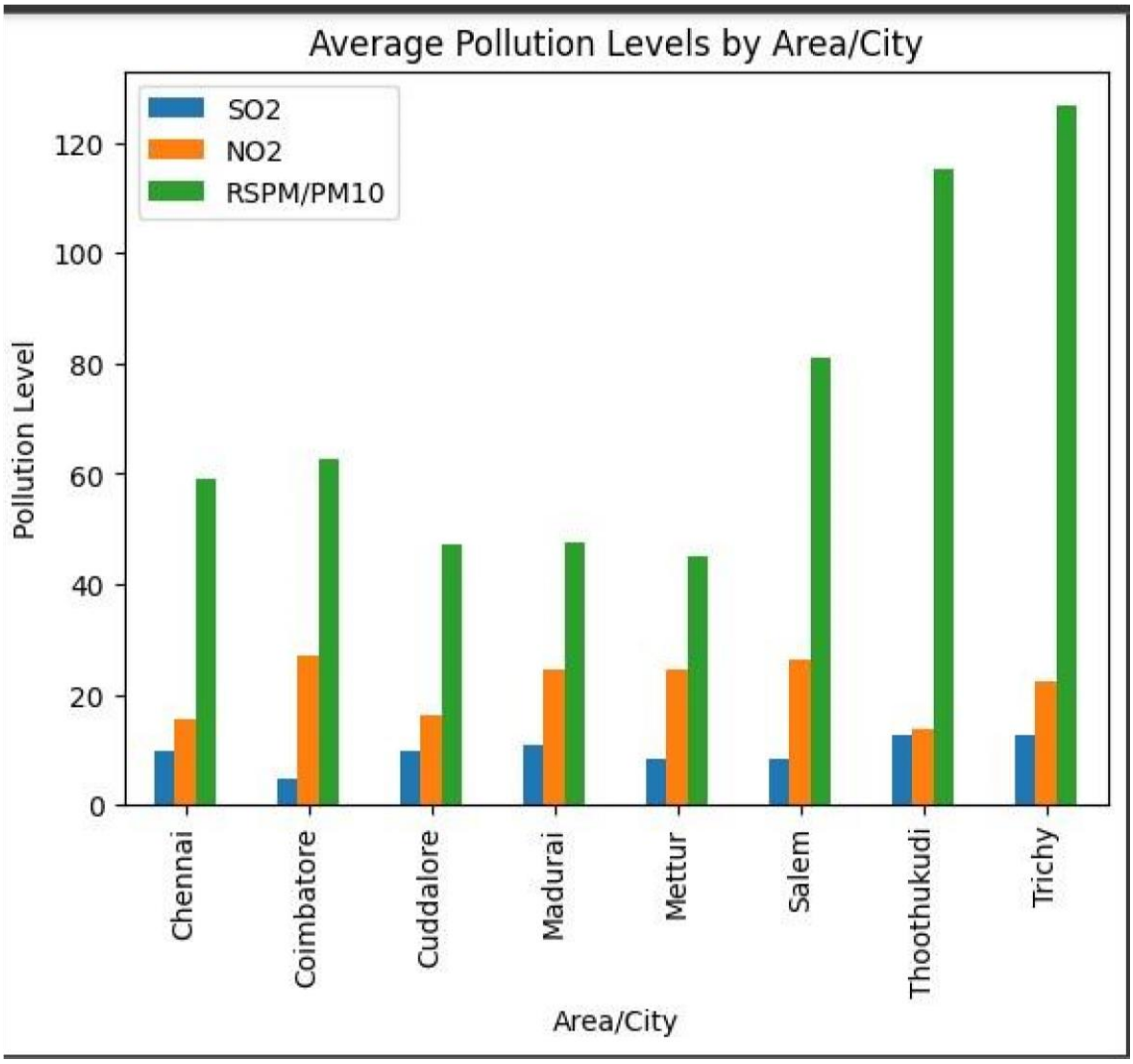
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1811 entries, 0 to 12230
Data columns (total 11 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   Stn Code                        1811 non-null   int64
 1   Sampling Date                   1811 non-null   datetime64[ns]
 2   State                           1811 non-null   object
 3   City/Town/Village/Area          1811 non-null   object
 4   Location of Monitoring Station  1811 non-null   object
 5   Agency                          1811 non-null   object
 6   Type of Location                1811 non-null   object
```

```
7  SO2                             1811
   nonnull   float64
8  NO2                             1811
   nonnull   float64
9  RSPM/PM10                       1811
   nonnull   float64          10  SPM
   1811 non-null   float64         dtypes:
   datetime64[ns](1), float64(4), int64(1),
   object(5)
```

Type of visualization used in project and the usage of the visualization Bar Chart:

- Bar charts are used to compare categories of data. They can be vertical or horizontal and are often used for showing frequency, distribution, and comparisons between different categories.

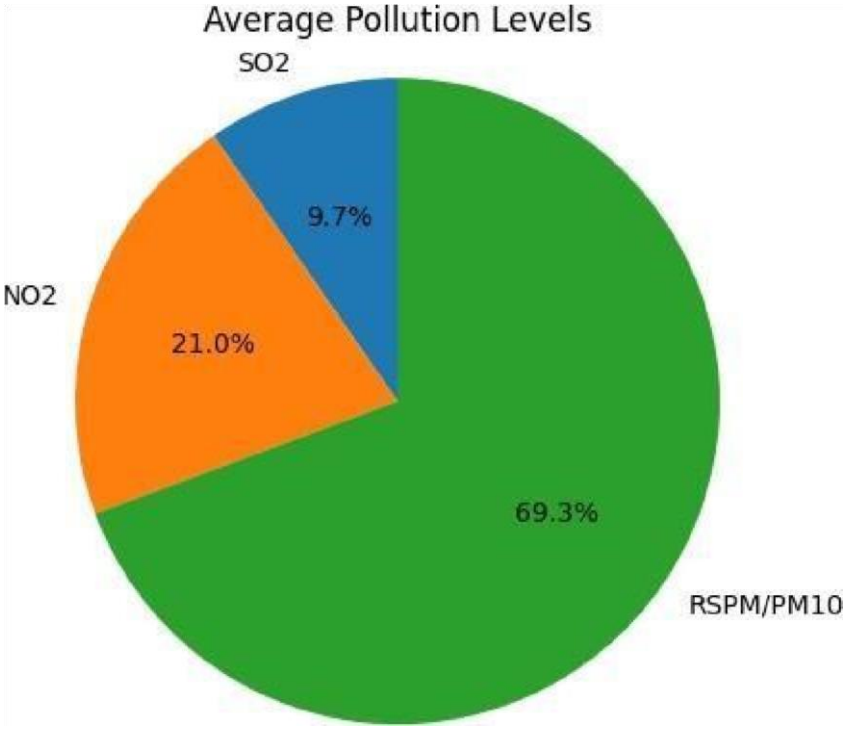
- Here in our visualization we use the bar chart to describe the **Highest rate of SO2 ,NO2, plotted against the Type of Monitoring Station,city/area/village.**



.

Pie Chart:

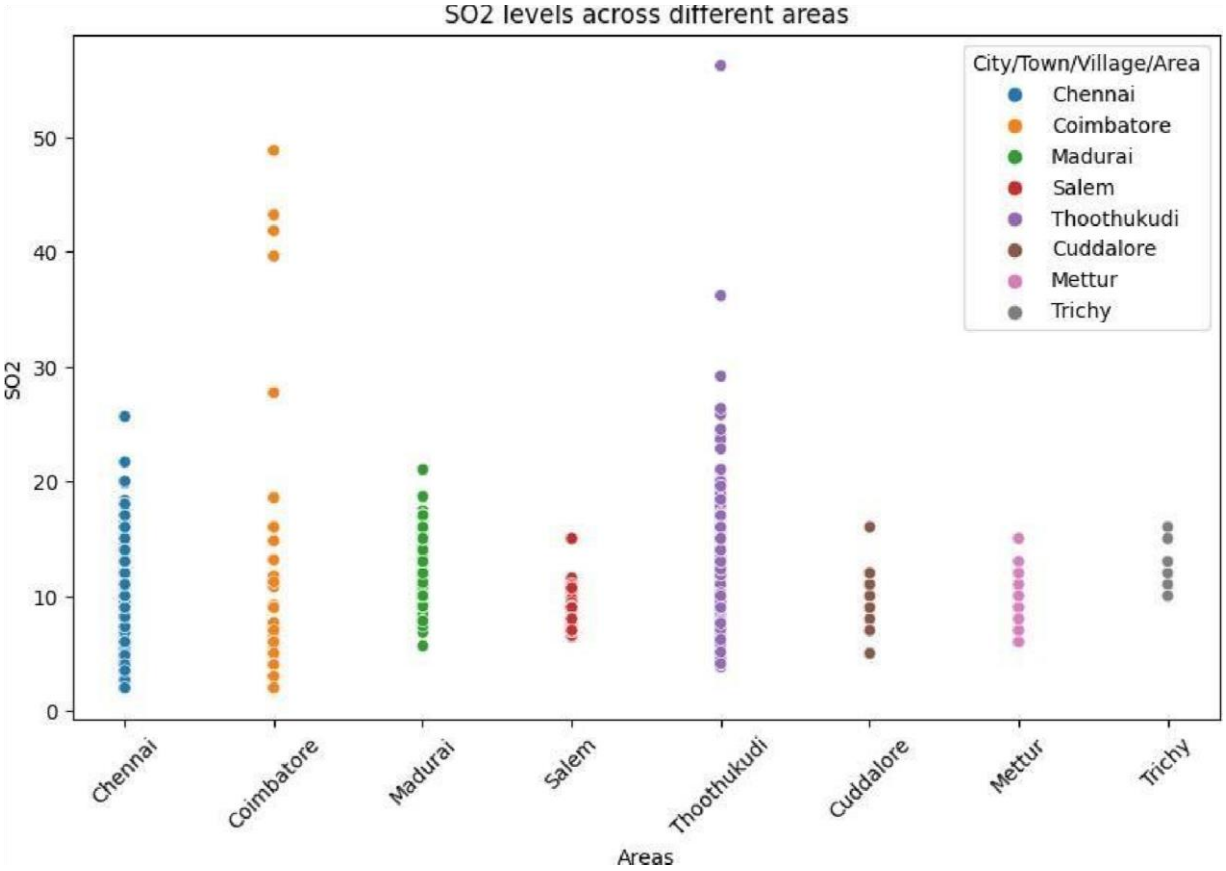
- Pie charts represent parts of a whole. They are useful for showing the distribution of categories as a percentage of the total.
- In our visualization we use the **pie chart to describe the overall percentage of the NO2,SO2,RSPM range across different area and monitoring station.**

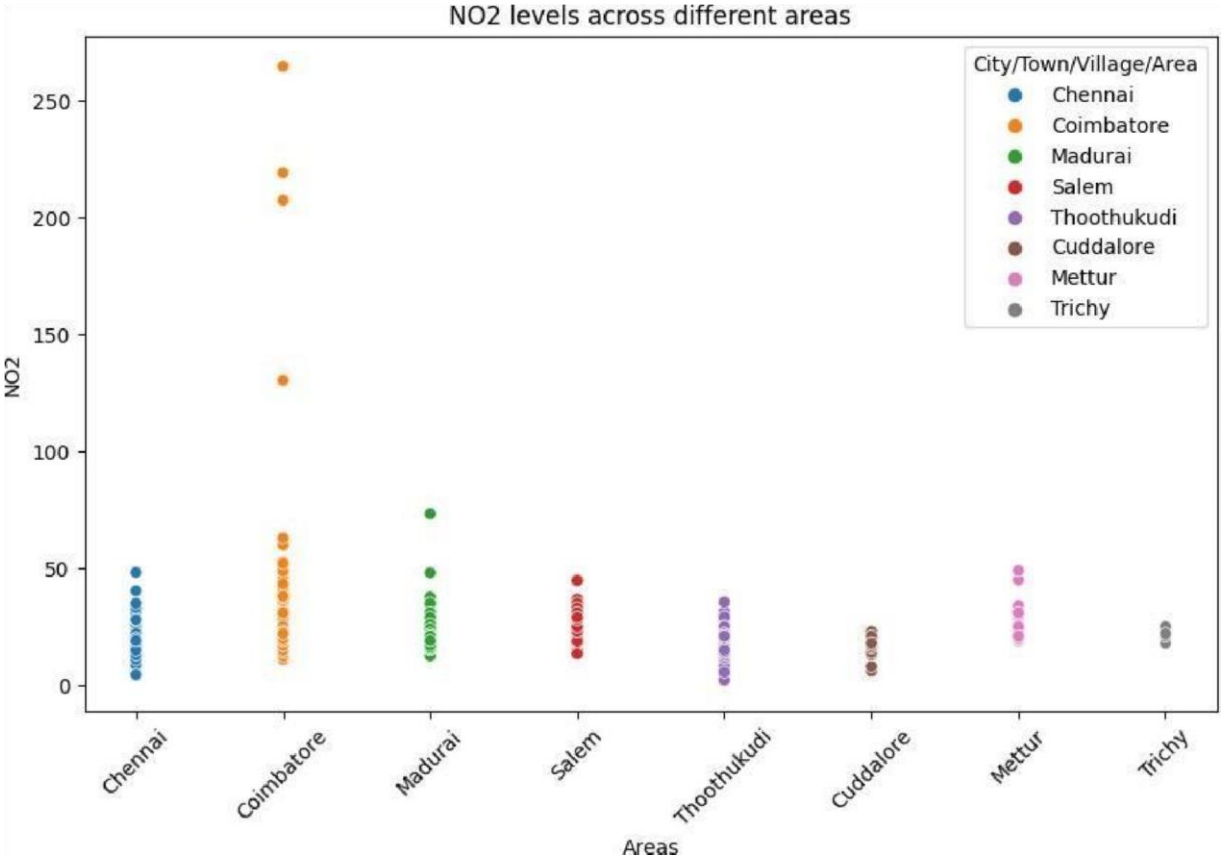


Scatter Plot:

- Scatter plots are used to visualize the relationship between two variables. Each data point is represented as a dot on the chart, making it easy to identify patterns or correlations.

- Scatter plots used to describe the highest rate of No2,SO2,RSPM.
- This would depict the range of NO2,SO2,RSPM/PM10 spread across the different region.

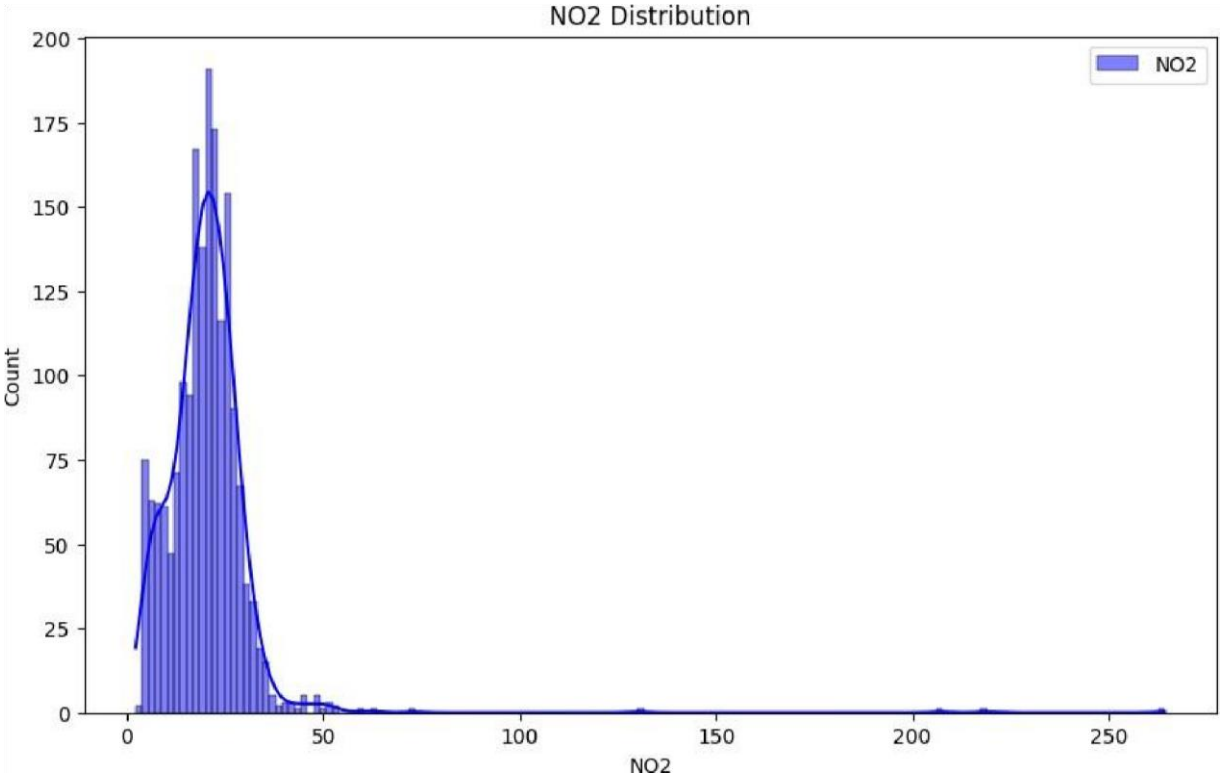


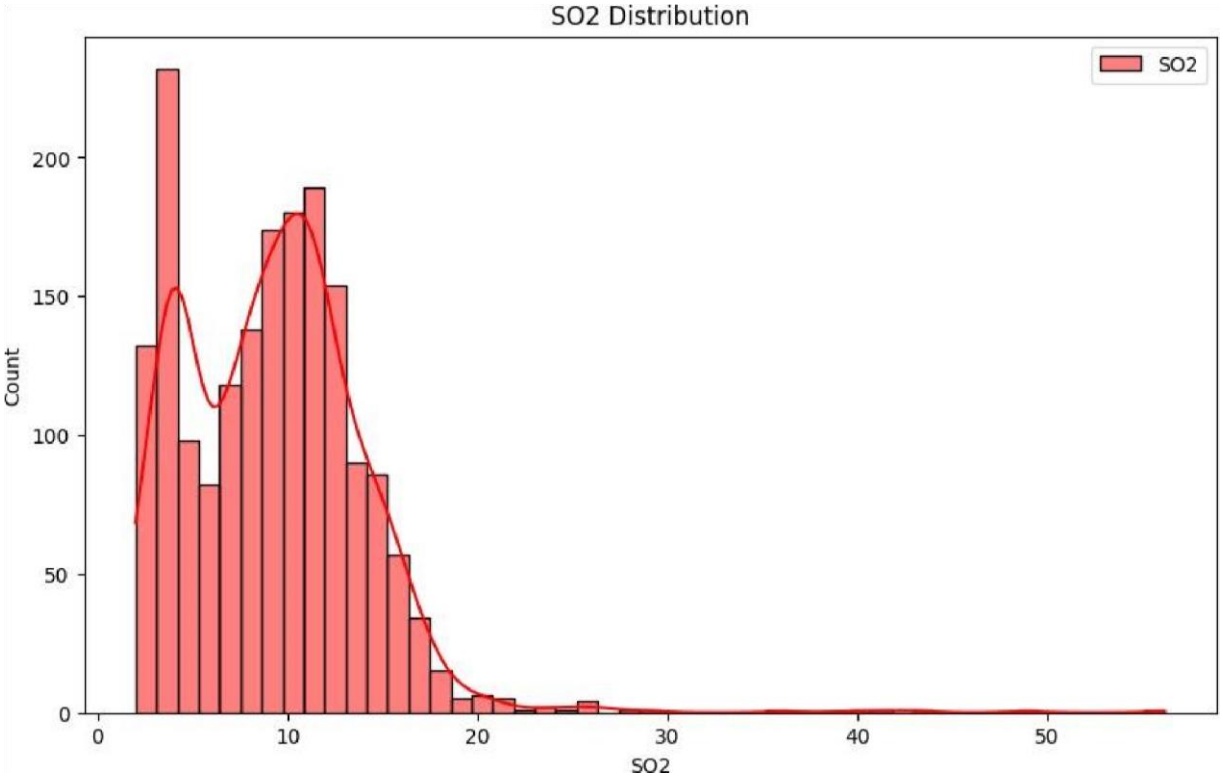


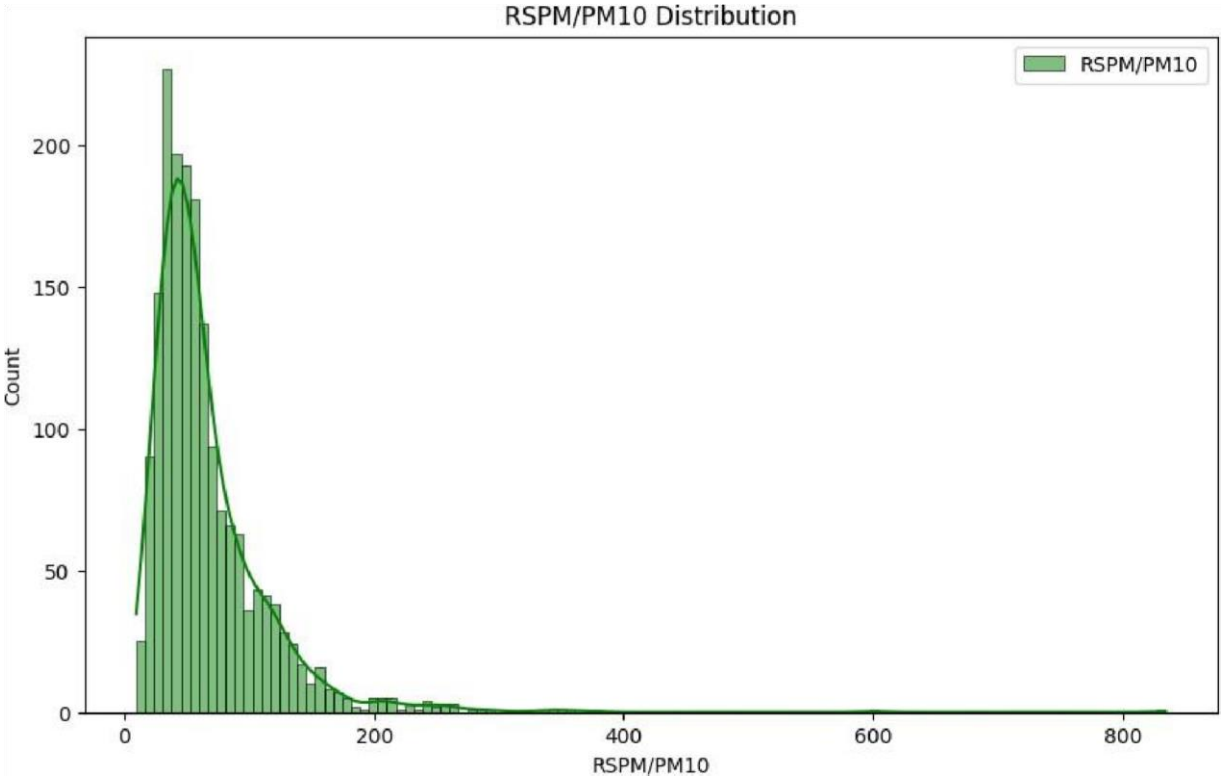
Histogram:

- Histograms display the distribution of continuous data by grouping it into bins or intervals. They are useful for understanding data distribution and identifying outliers.

- The overall percentage of NO2,SO2,RSPM range of the cities, type of Monitoring Station

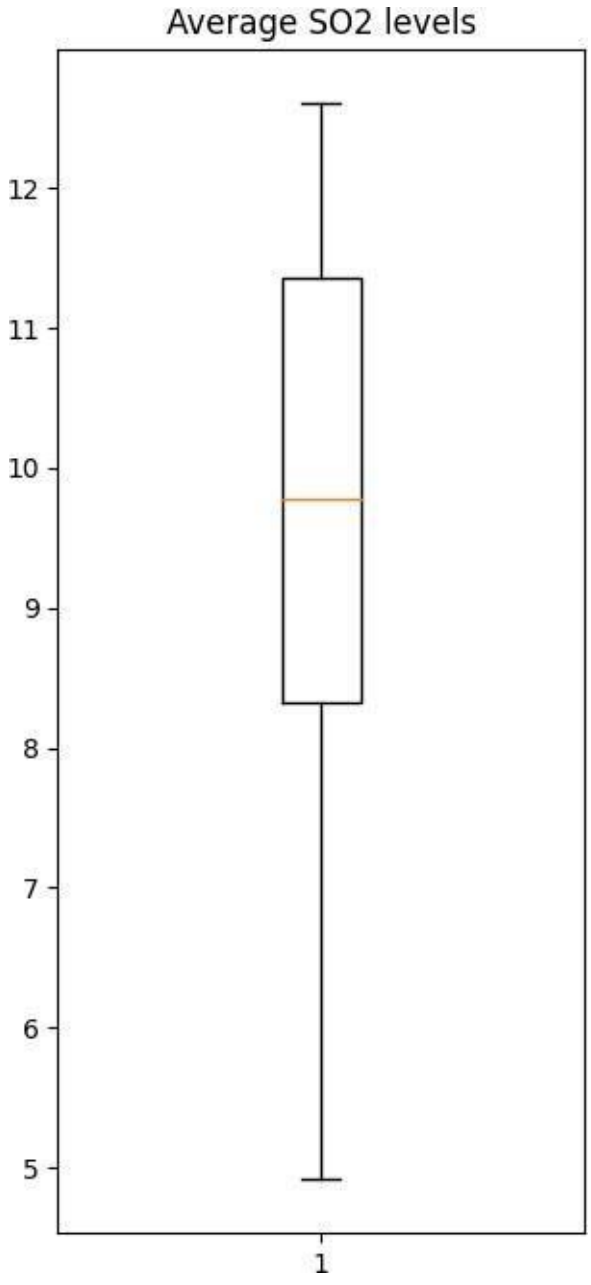


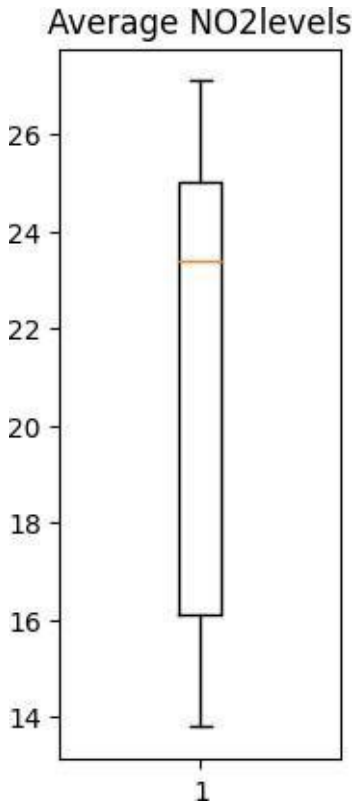




Box Plot (Box-and-Whisker Plot):

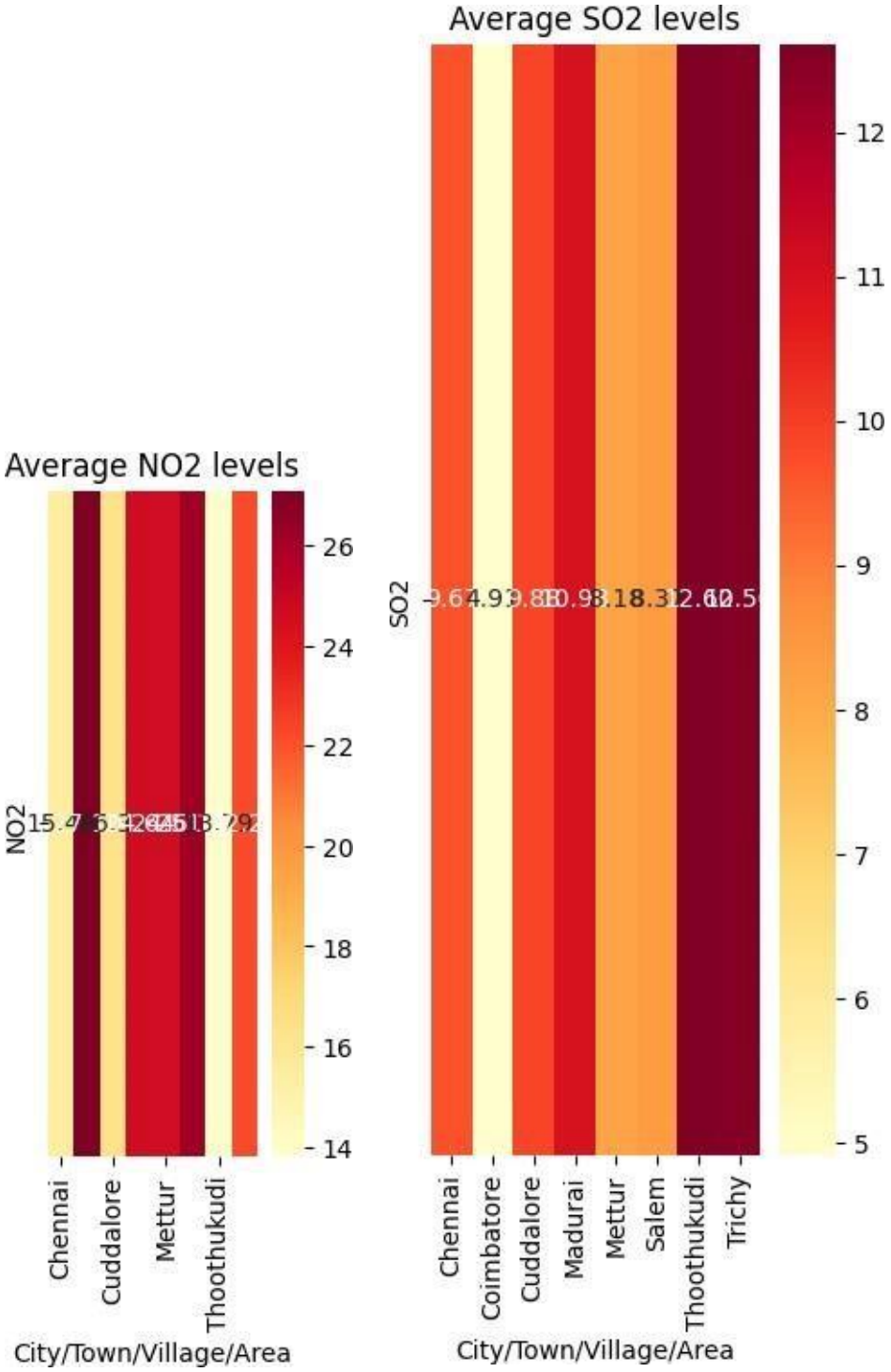
- Box plots are great for visualizing the distribution, spread, and skewness of data. They provide information about the median, quartiles, and potential outliers in the data.
- The overall concentration and average value of the NO2,SO2 ,Rspm value of the Spread across the area.





Heat Map:

- Heat maps use color to represent the magnitude of values in a matrix or table. They are often used to show correlations or patterns in large datasets.
- The overall structure of the Heat map the level of concentration of NO2,SO2,RSPM/PM10



CREATING A VISUALIZATION USING

COGNOS

TYPE OF PARAMETER GIVEN IN AIR QUALITY DATASETS OF TAMIL NADU:

Stn Code	Sampling	State	City/Towr	Location c	Agency	Type of Lc	SO2	NO2	RSPM/PM	SPM
159	8/2/2010	Tamil Nad	Chennai	Madras M	National E	Residenti	3.166667	9.166667	80	
159	########	Tamil Nad	Chennai	Madras M	National E	Residenti	4.666667	6.333333	51.66667	

1 .**STN CODE:** It refers to **station code of the Tamil Nadu District.**

1. **0-150 : Unhealthy and senstive**
2. **150-200:unhealthy**
3. **201-300: very unhealthy**

1. 2. **SAMPLING Date:** It refers to the **sampling date at which No2 and so2 measured.**

3. **STATE**: It refers to the **Tamil Nadu.**

4. **City/Town/village: List of village givens in the datasets.**

1. **Chennai**
2. **Coimbatore**

3. **Cuddalore**
4. **Madurai**
5. **Thoothukudi**
6. **Salem**
7. **Trichy**
8. **Mettur**

LOCATION: It refers to the Location the so2 and no2 measured.

Agency: Refers to the Industry emitting No2 and so2

Type of Location: It refers to the type of location such resident and street .

SO2: It refers to the **sulphur dioxide**.

NO2: It refers to the **Nitrogen dioxide.**

RSPM: Respirable suspended Particulate Matter.

SPM: Suspended Particulate Matter.

AIM: To create a **BAR GRAPH VISUALIZATION i**dentify

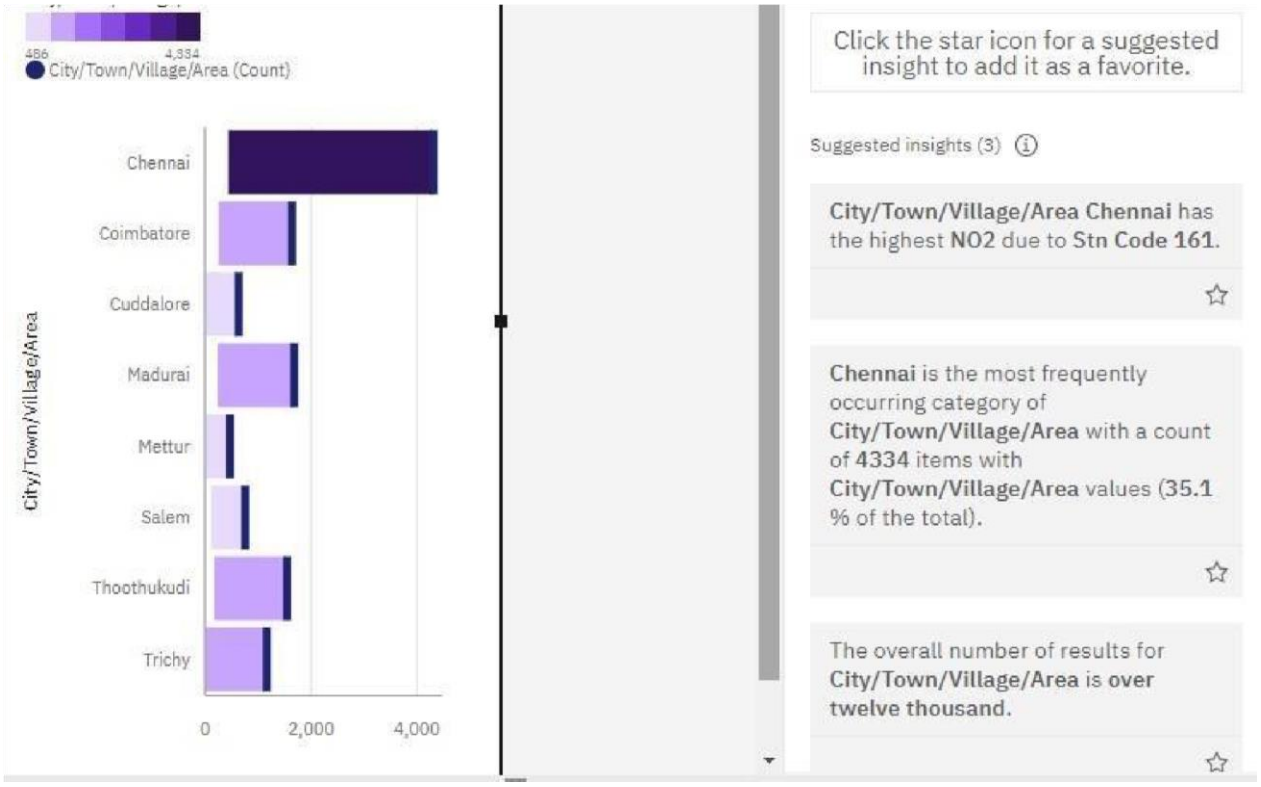
SO2 , NO2, RSPM and SPM

containing the **highest**

Parameter Needed : x-axis: city-village

Y-axis: SO2 , NO2, RSPM and SPM

NO2:



Insights from the data:

1. **Chennai has the highest**

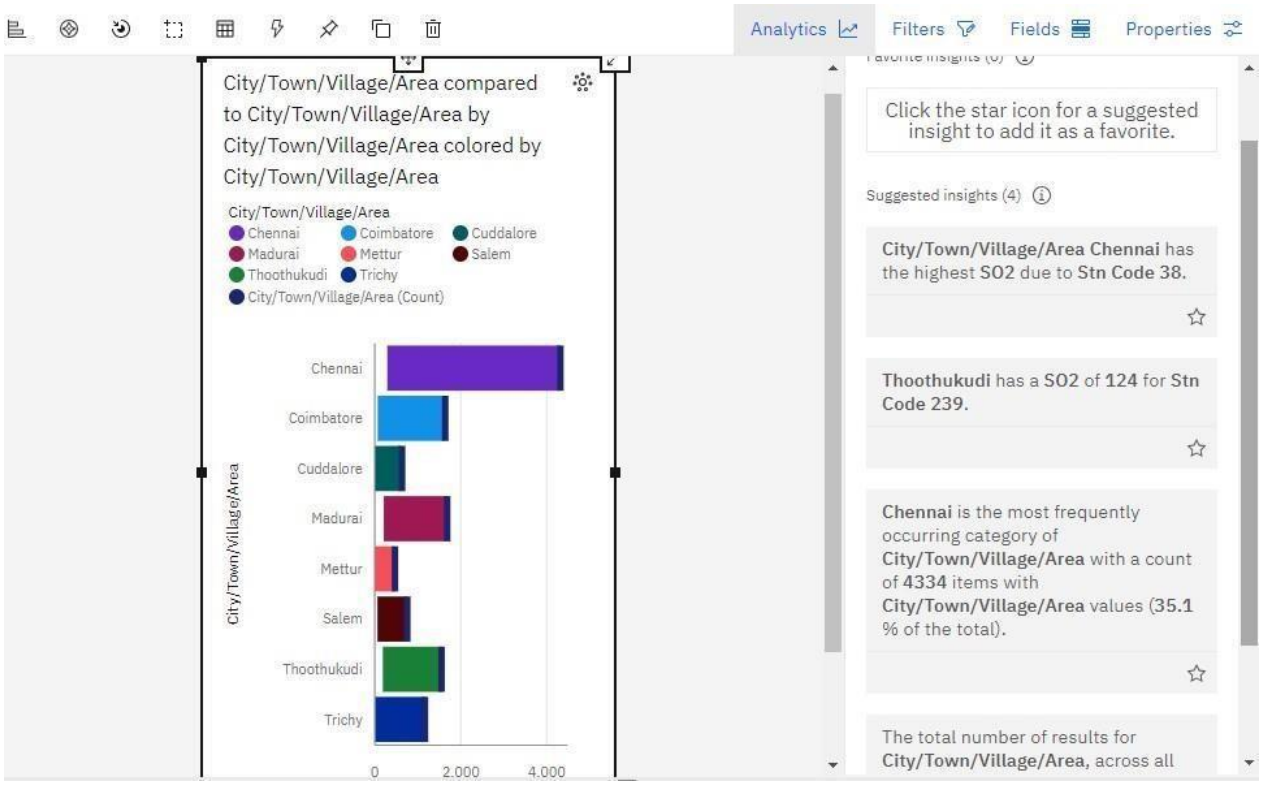
NO2 level-58%. Due to station code

159 High pollution due to No2

2. **Mettur has the lowest NO2 level-**

10%. It has normal air quality Index

SO2:



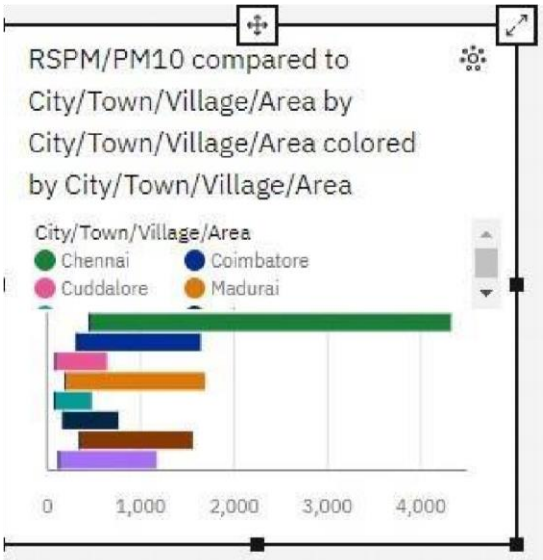
INSIGHT

S: **1. Chennai has the highest SO2 due to highest air quality index 159 and Thoothukudi has the second highest air quality Index 140.**

2. Mettur and cuddalore has the lowest so2 content.

RSPM:

1. **Chennai and Thoothukudi has the highest RSPM/PM10 because high industry and vehicle Range that causes by high concentration of <u>NO2 and SO2</u> std code 138 and 240 respectively.**



City/Town/Village/Area Chennai has the highest **RSPM/PM10** due to **Stn Code 38.**

Thoothukudi has a **RSPM/PM10** of **256** for **Stn Code 240.**

Chennai is the most frequently occurring category of **City/Town/Village/Area** with a count of **4334** items with **City/Town/Village/Area** values (**35.1** % of the total).

The total number of results for **City/Town/Village/Area**, across all **City/Town/Village/Area**, is **over twelve thousand.**

SPM:

1. **Chennai and salem has the content of SPM compared to other district**

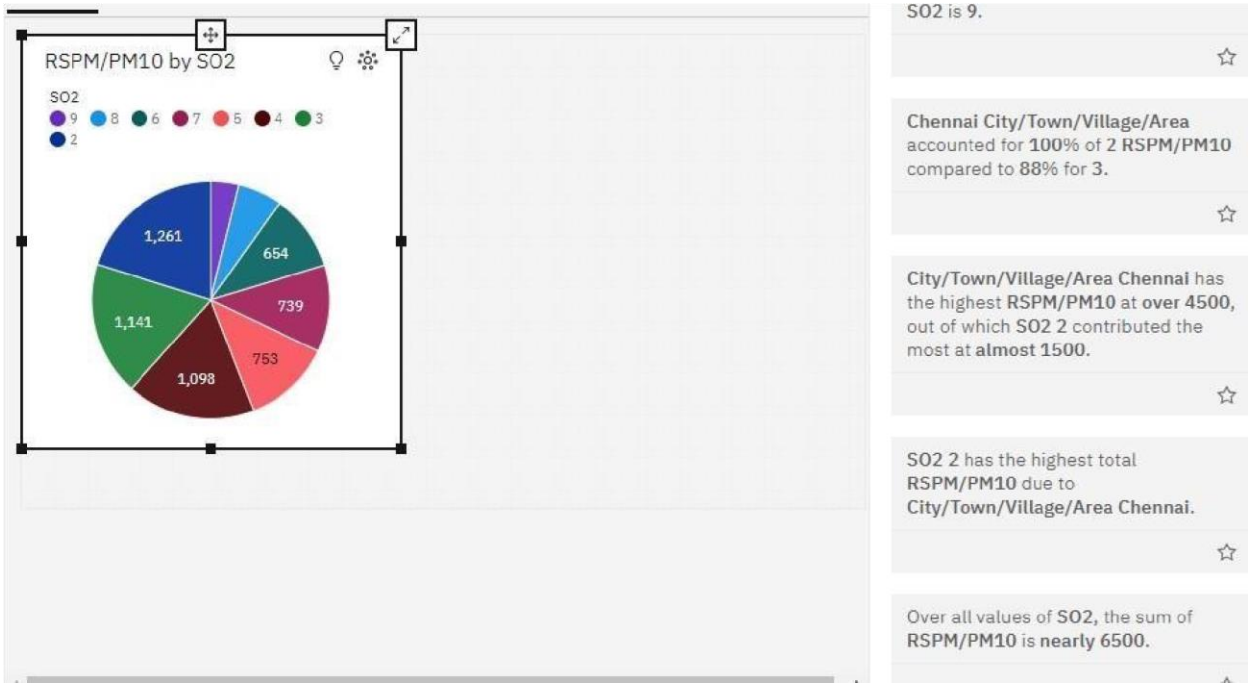
AIM: To create pie chart to visualize the so2 and no2 based on RSPM and SPM

Segments: SO2,

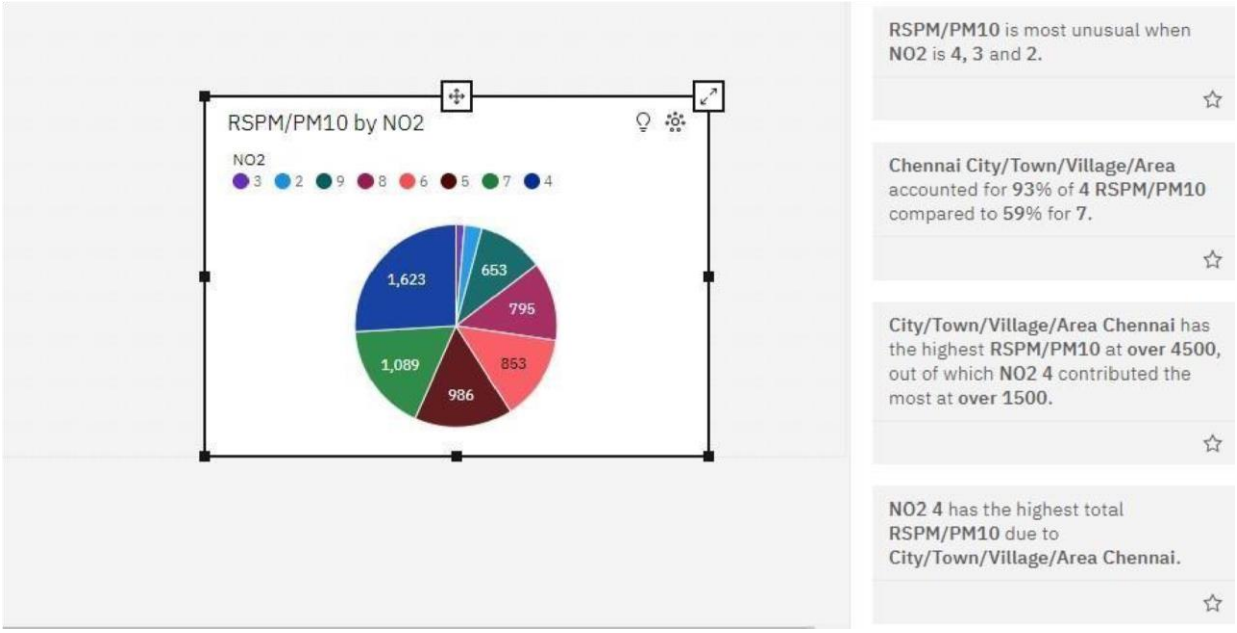
NO2 SIZE: RSPM

Case1: RSPM LEVEL Based on So2:

1. **Chennai has the RSPM Level.**
2. **Chennai contributed 4500 RSPM level over 1500 of SO2.**



Case2: RSPM LEVEL Based on No2



Insights:

1. **Chennai has the highest RSPM caused by 59%of NO2.**
2. **It is due to high population in Chennai area.**

AIM: Creating a box plot to identify the cause of SPM due to

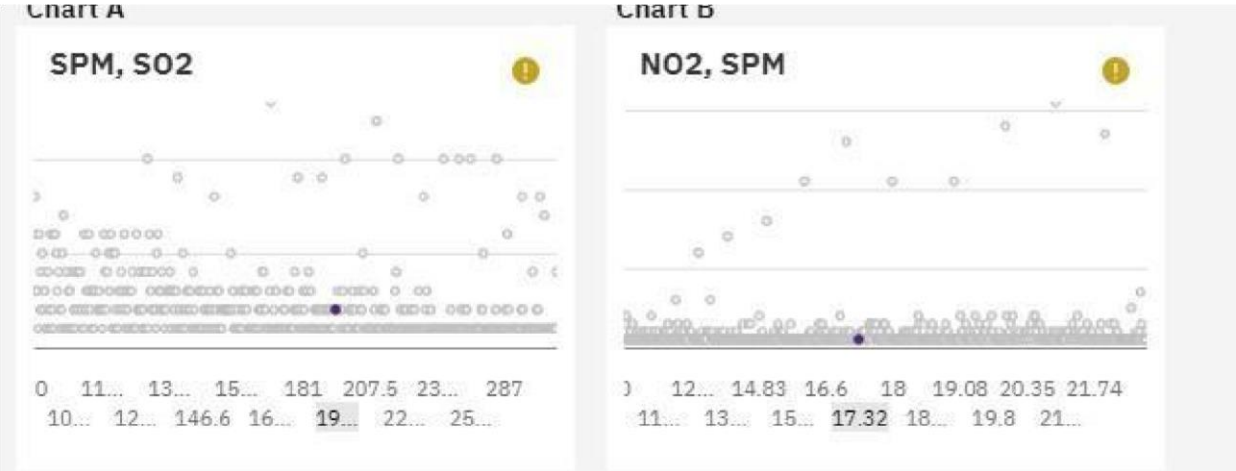
NO2 and

SO2

Requirements:

X - axis: SPM

Y - axis: So2,NO2

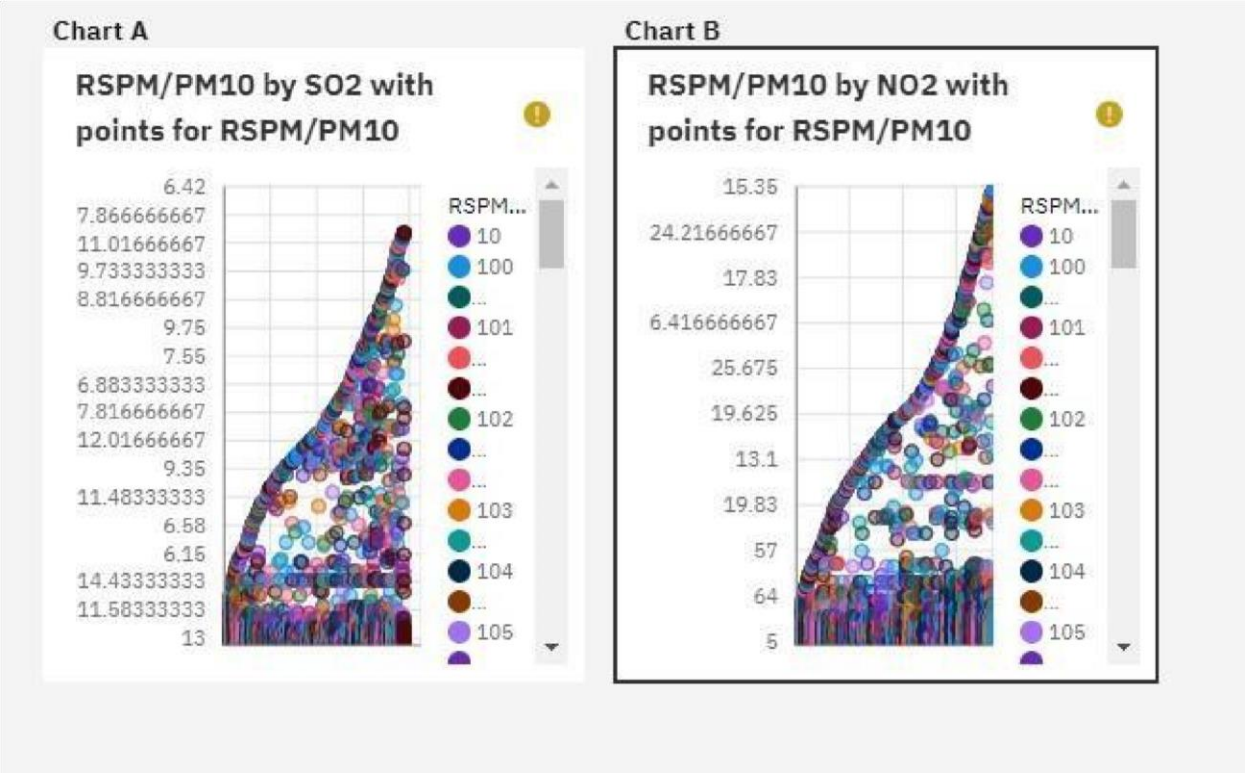


Summary	Chart A : SO2	Chart B : SPM	Combined
Minimum	1	1	-
Maximum	57	82	-
Chart perc... data set	100%	100%	-
Chart total	639	761	-
Difference...rt totals			-

INSIGHTS:

1. **SPM is caused by both so2 and No2.**
2. **So2 is contributed most of the SPM**

Aim: To create a scatter plot to identify the cause RSPM range caused by NO2 and So2

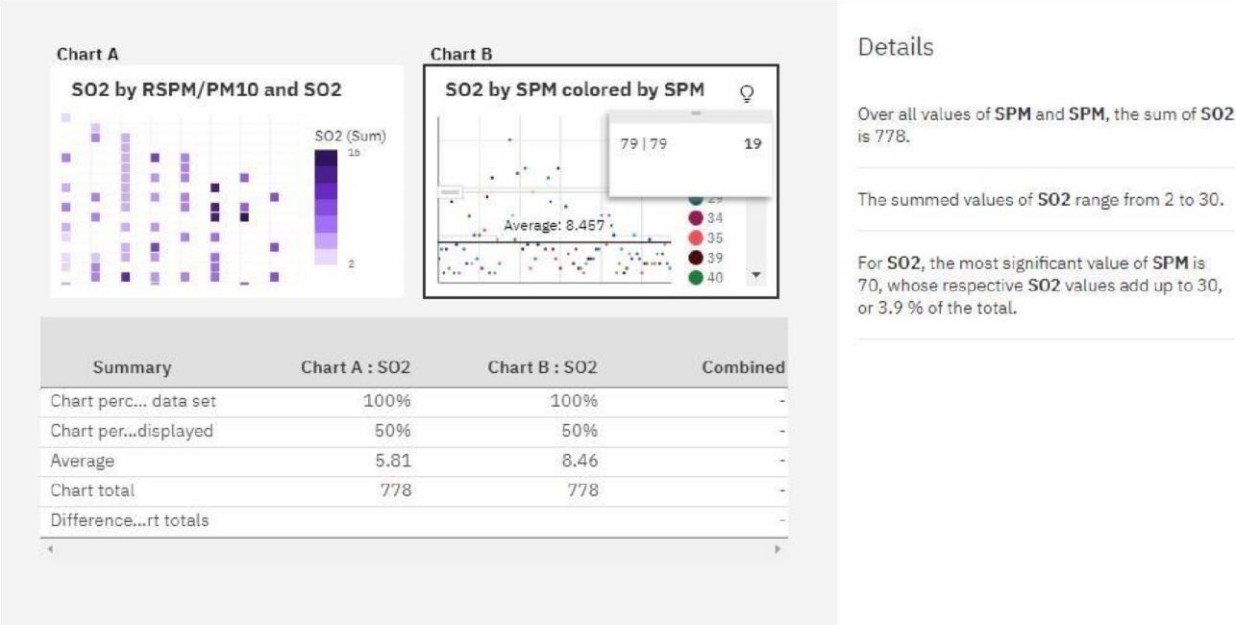


1 .Chennai contributed most so2 and No2 that causes RSPM

LEVEL.

Aim: To create different visualization to obtain predictive analysis of RSPM by SO2

Same predicting value.



Summary	Chart A : SO2	Chart B : SO2	Combined
Chart perc... data set	100%	100%	-
Chart per...displayed	50%	50%	-
Average	5.81	8.46	-
Chart total	778	778	-
Difference...rt totals			-

Insight:

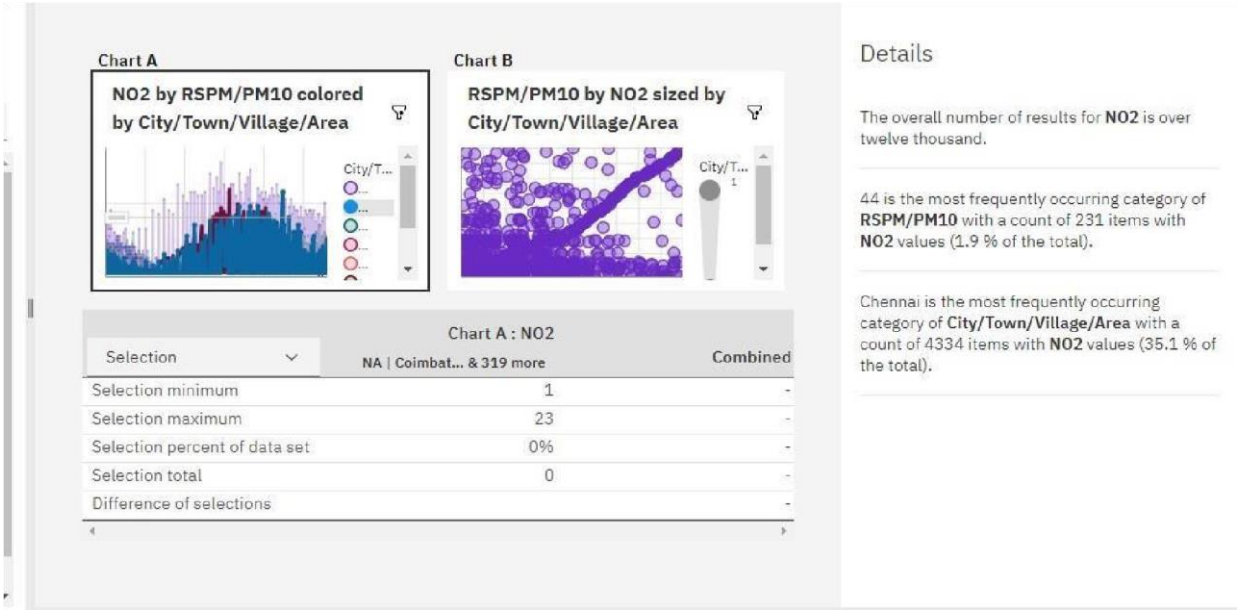
- **Comparing the value of so2 using heat map scatter plot.**
- **The value of SO2 has the range and chennai the highest percentange of so2 content of 8**

AIM: To create a different visualization comparing the

value of NO2 of different location.

Insights:

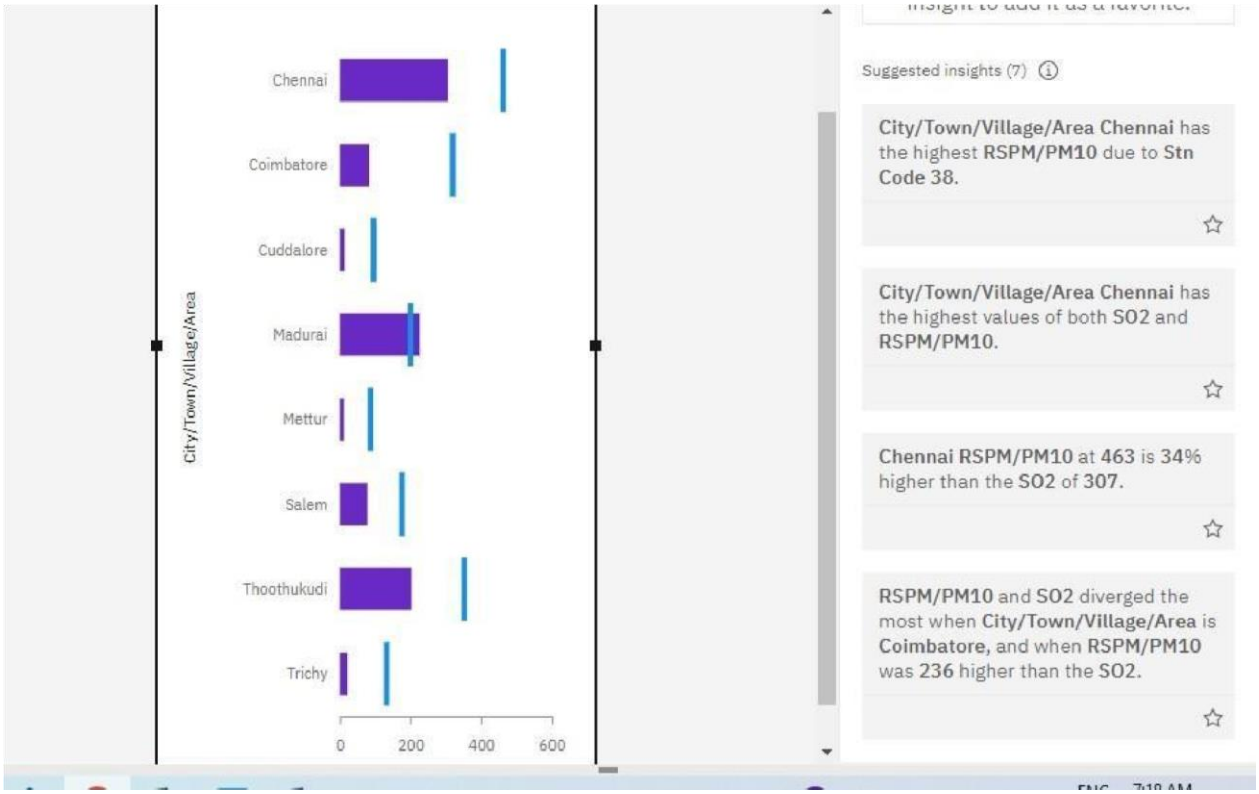
- **Comparing the value of So2 and No2 and RSPM Range using Line plot and column plot. The value of No2 has highest range of RSPM in Chennai.**



Aim: To create a column plot to identify the highest range of so2 plotted over a RSPM

X- Axis: RSPM Range

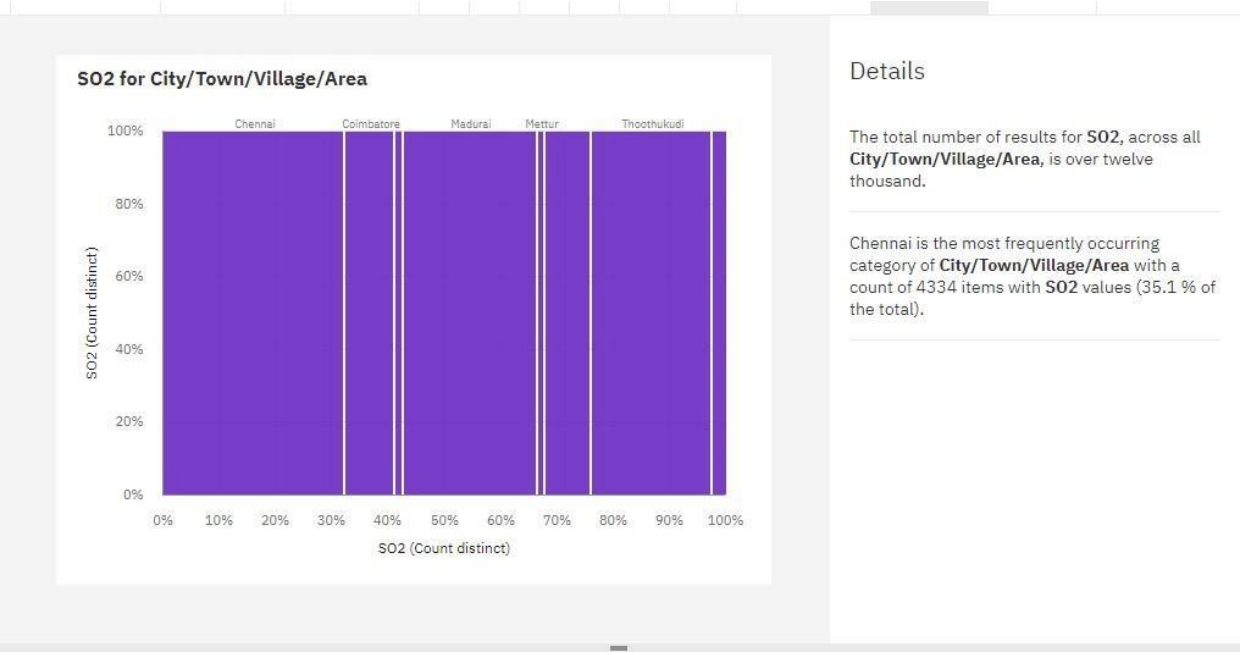
Y-Axis: So2 Range



Insights:

1. **Chennai has the highest range of the RSPM/PM is 34% and contributed the so2 range of 307.**
2. **Chennai is considered to be a highest pollutated area among the given plot.**

AIM: To create a Marimeeko Plot to identify the NO2 and so2 range plotting against City and Town

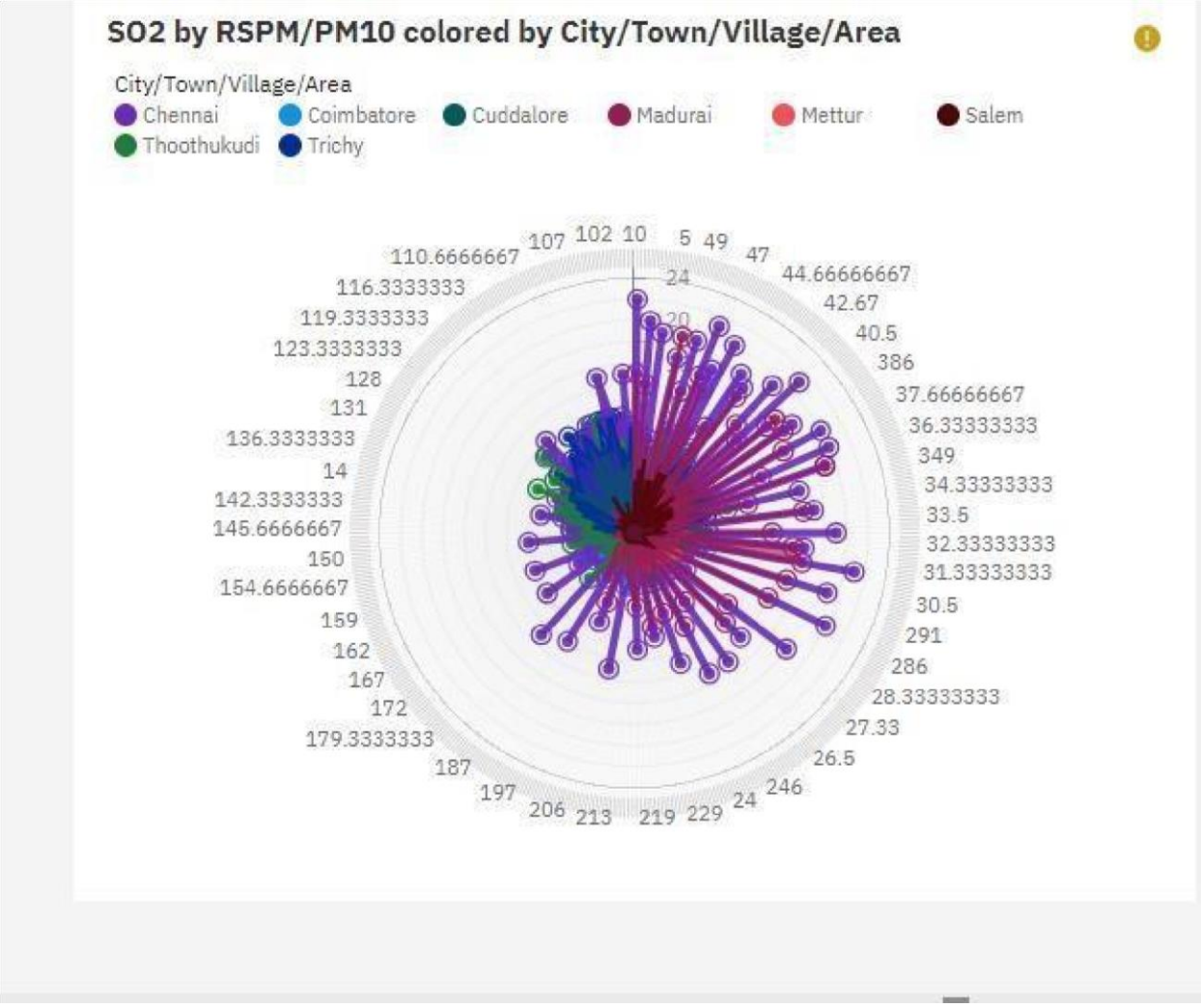


H

Insight:

1. **Chennai and thoothukudi contributed RSPM range across all the district.**
2. **They have largest city and town and s0 of 35.1%.**

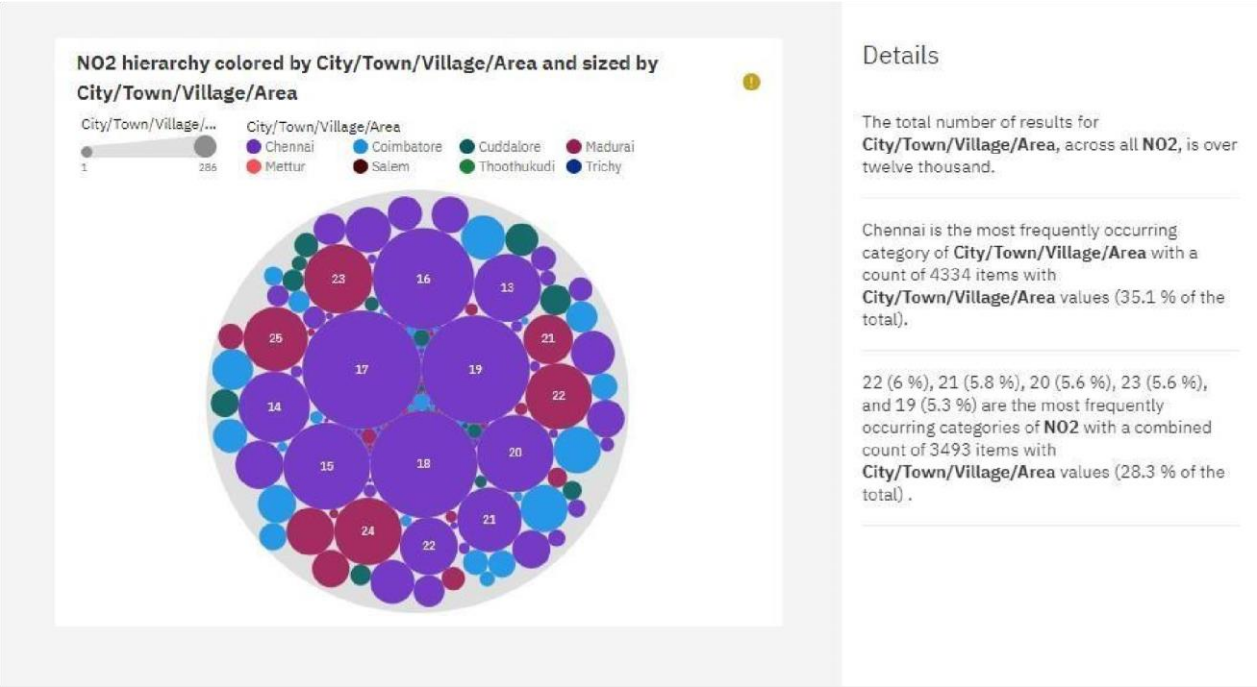
AIM: To create a radar method to find out the insights of so2 to RSPM used to Highest range of polluted area



INSIGHTS:

1. **It shows a highly polluted region among all region.**
2. **It shows SO2 increases the range of RSPM also increases.**

AIM: To find the concentration of NO2 using a hierarchy bubble to find out the insights of city that is affected by NO2

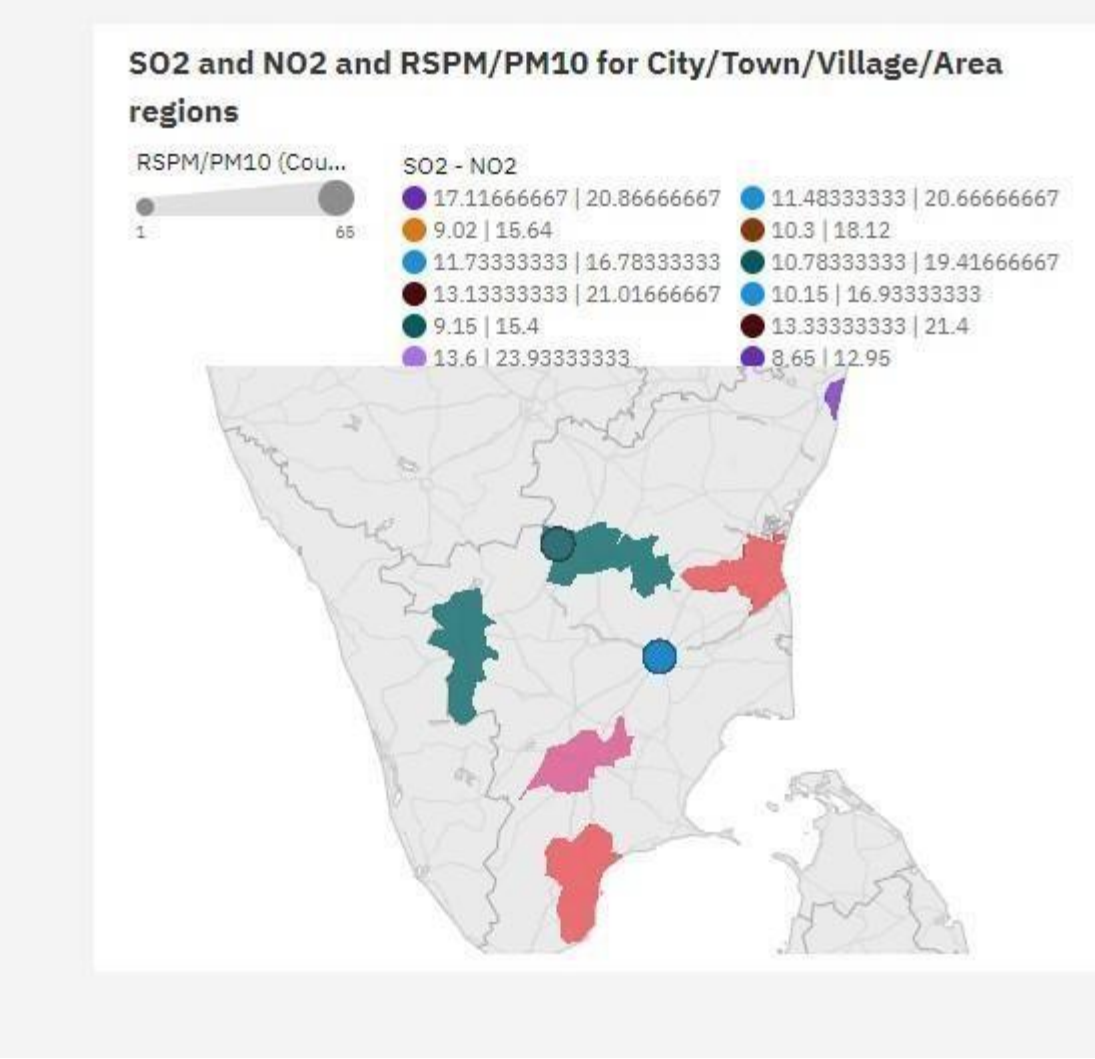


Insights:

1. **The total amount NO2 ocurred in the Chennai due to high traffic area.**

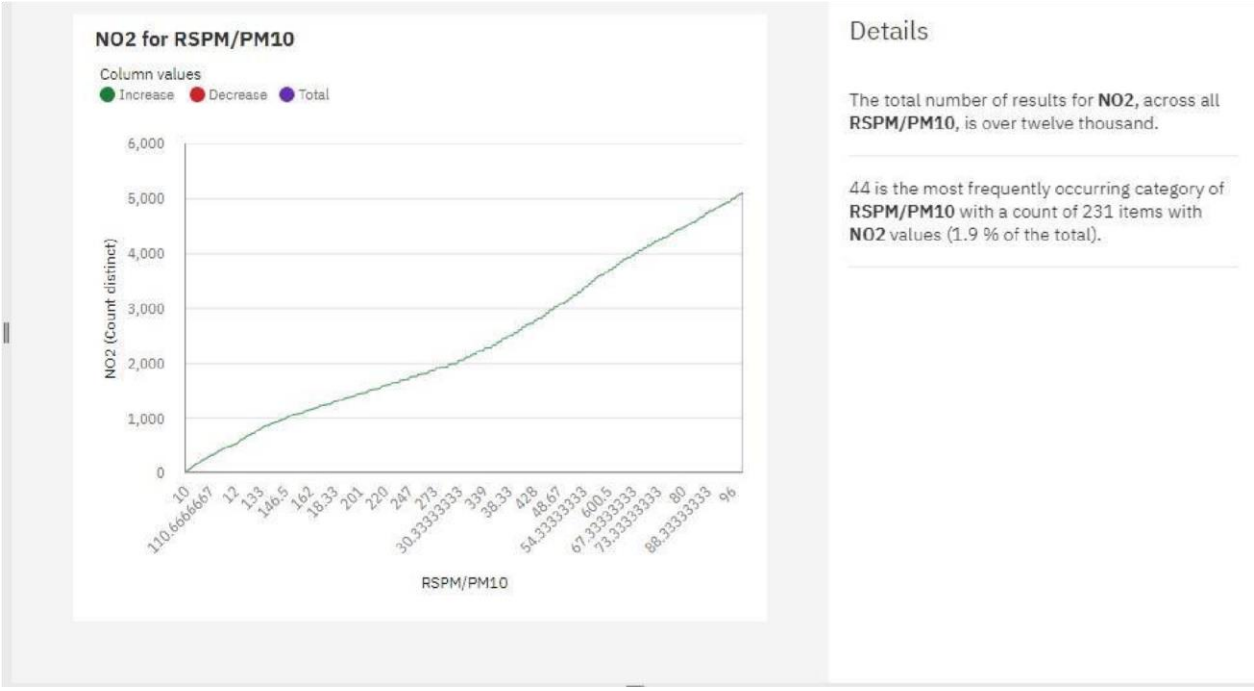
2. **The total value of NO2 has the value concentration has 12000.**

Aim: To represent high area pollution area using map projections



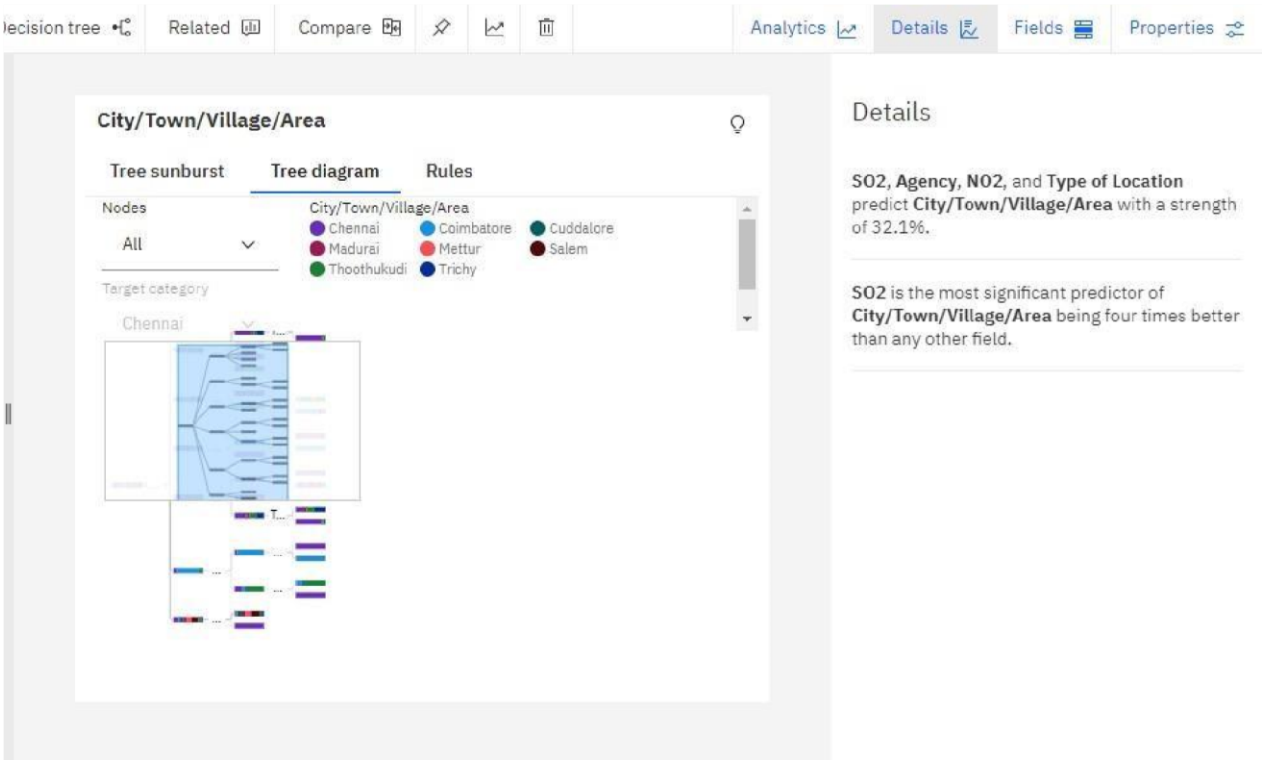
AIM: To create a waterfall visualization to identify the

NO2



1. This shows the NO2 increase Rspm also Increases.

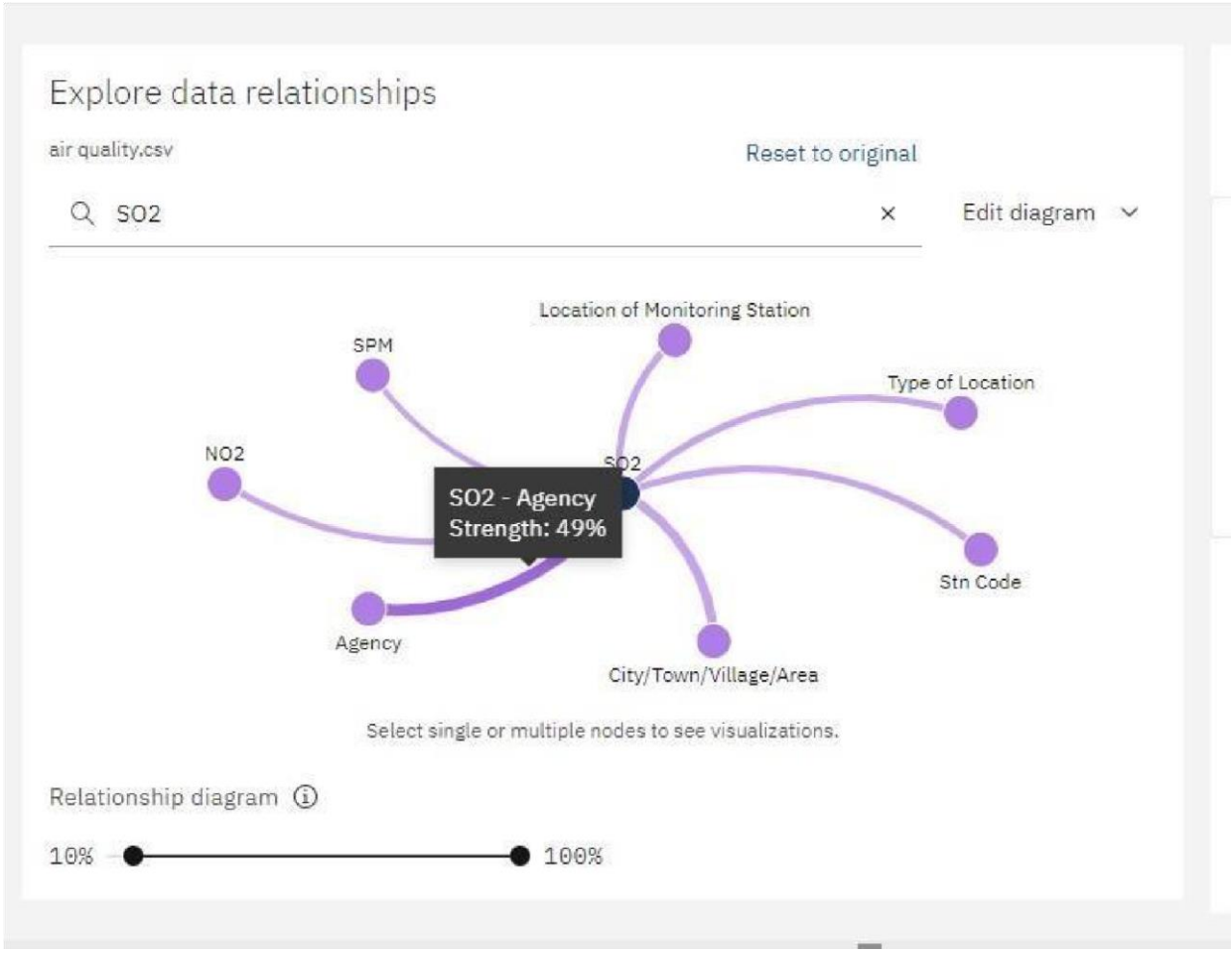
Aim: To create a decision tree to find out the insights of SO2 over RSPM .



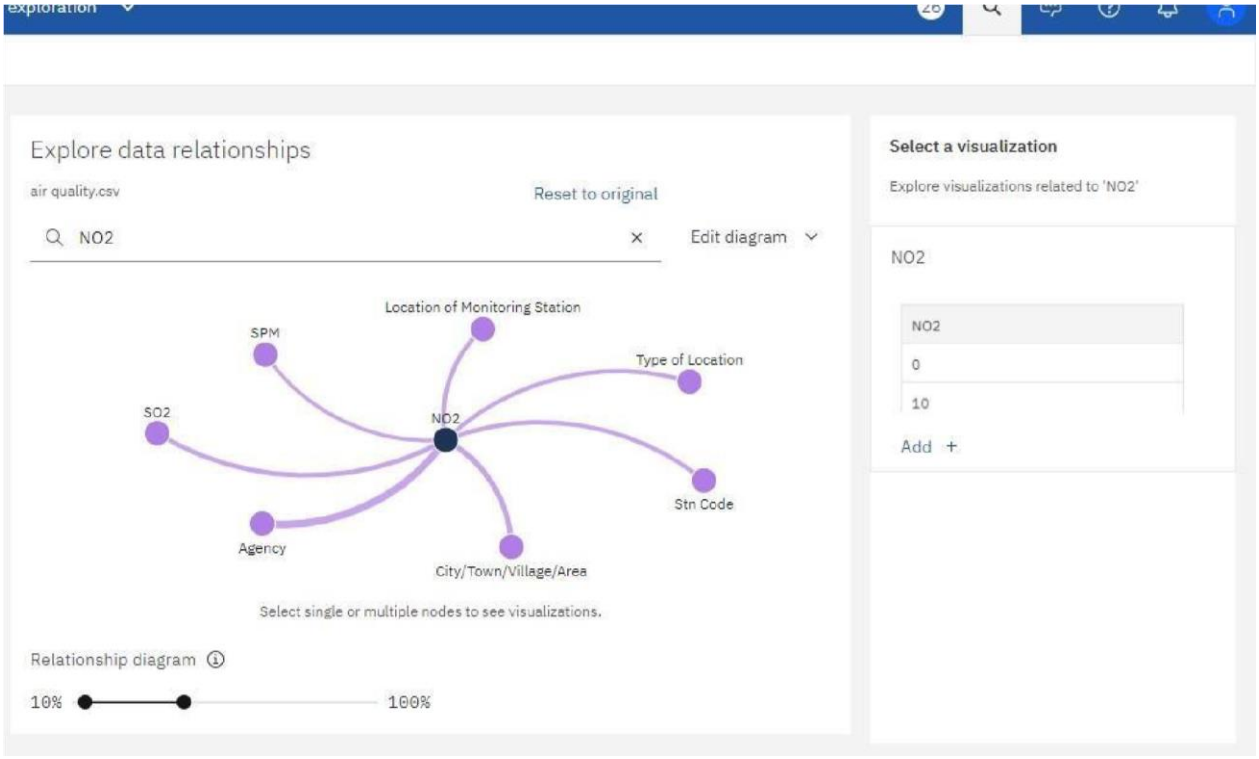
1. So2 has the highest percentage of 50%

AIM:

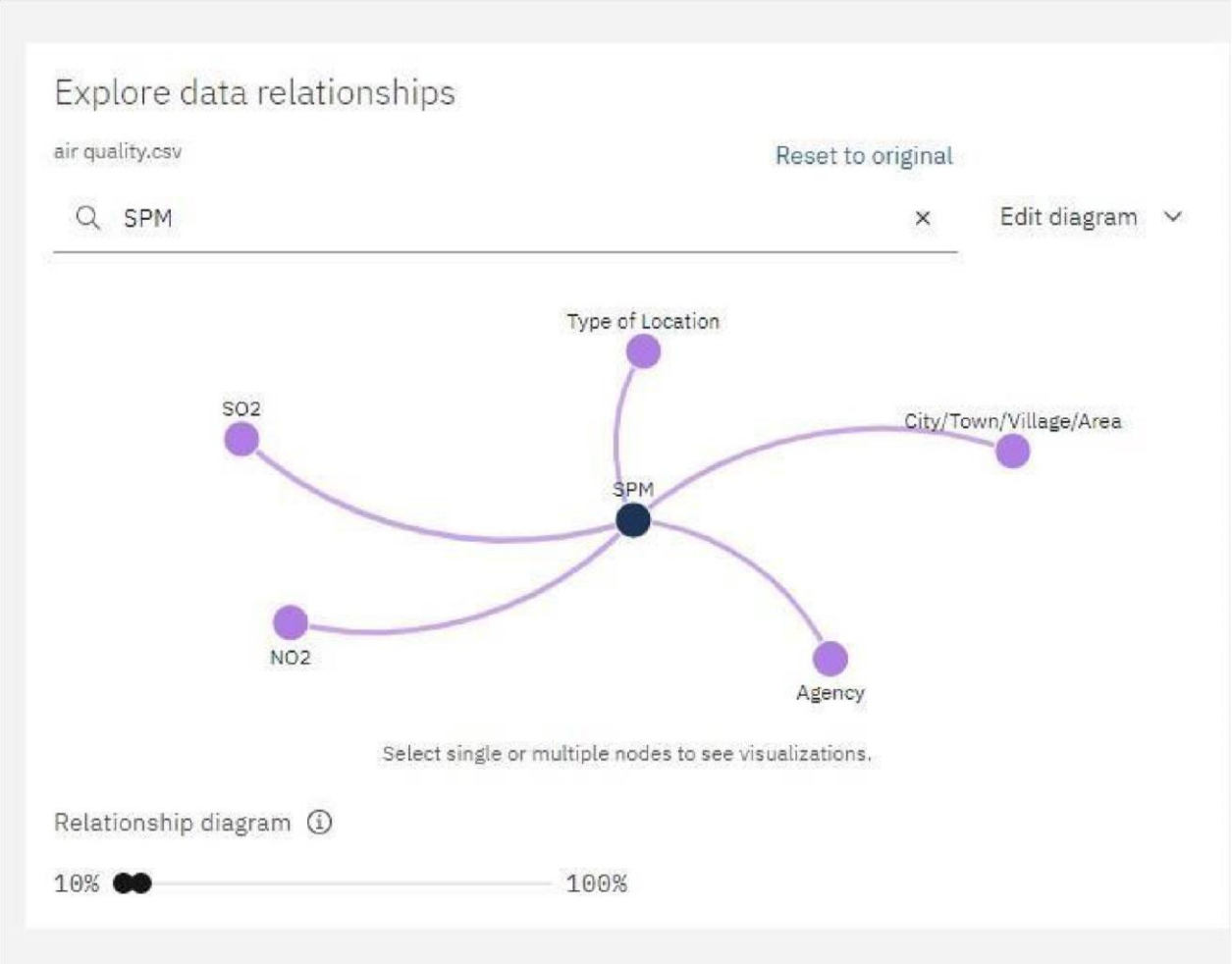
To find out the relationship of NO2,So2, RSPM and SPM Insights.


Explore data relationships
air quality.csv
Reset to original
SO2
Edit diagram
Location of Monitoring Station
SPM
Type of Location
NO2
SO2
SO2 - Agency
Strength: 49%
Stn Code
Agency
City/Town/Village/Area
Select single or multiple nodes to see visualizations.
Relationship diagram
10%
100%

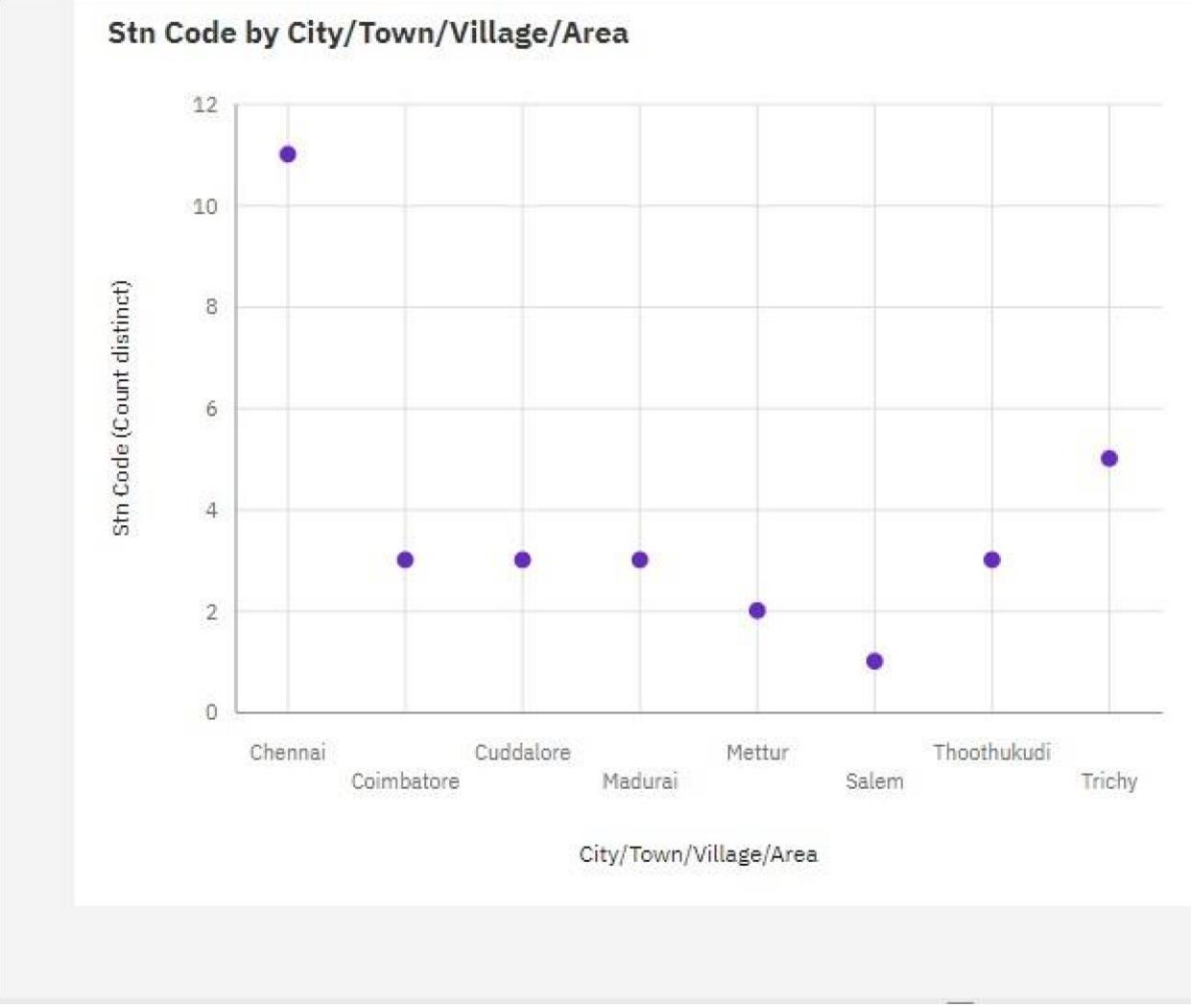
NO2


exploration
Explore data relationships
air quality.csv
Reset to original
NO2
Edit diagram
Location of Monitoring Station
SPM
Type of Location
SO2
NO2
Stn Code
Agency
City/Town/Village/Area
Select single or multiple nodes to see visualizations.
Relationship diagram
10%
100%
Select a visualization
Explore visualizations related to 'NO2'
NO2
NO2
0
10
Add +

SPM:



Aim: To identify the highly polluted area using visualization using line plot .



With these libraries visualizations such as histogram,barchart,heatmap,bloxplot,atc have been created. This is to Visualize the air quality predictions and make them easily accessible to the public and decisionmakers. Effective communication of the data and its implications is crucial for public awareness and policy decisions.

A predictive model is use for calculating Air Quality Index(AQI).The Air Quality Index (AQI) is a standardized system used to communicate and assess air quality based on the concentration of specific air pollutants. The AQI is typically calculated for pollutants such as particulate matter (PM2.5 and PM10), ground-level ozone (O3), sulfur dioxide (SO2), nitrogen dioxide (NO2), carbon monoxide (CO), and others. The AQI provides a simple way to inform the public about the current air quality conditions and potential health risks associated with air pollution.

Calculating the air quality Index of the SO2,NO2, RSPM Level

Python coding for calculating SO2 AQI level

```
import pandas as pd   import
matplotlib.pyplot       as    plt   import
seaborn as sns


# Read the data into a DataFrame df =
pd.read_csv("updated_project.csv")

# Check the information about the
DataFrame print(df.info())

# Display the first few rows of the
DataFrame def calculate_si(SO2):
    si=0
    if(SO2<= 40):
        si=SO2*(50/40)

    if(SO2>40 and SO2<=80):
```

calculating NO2 AQI Level

```
def
```

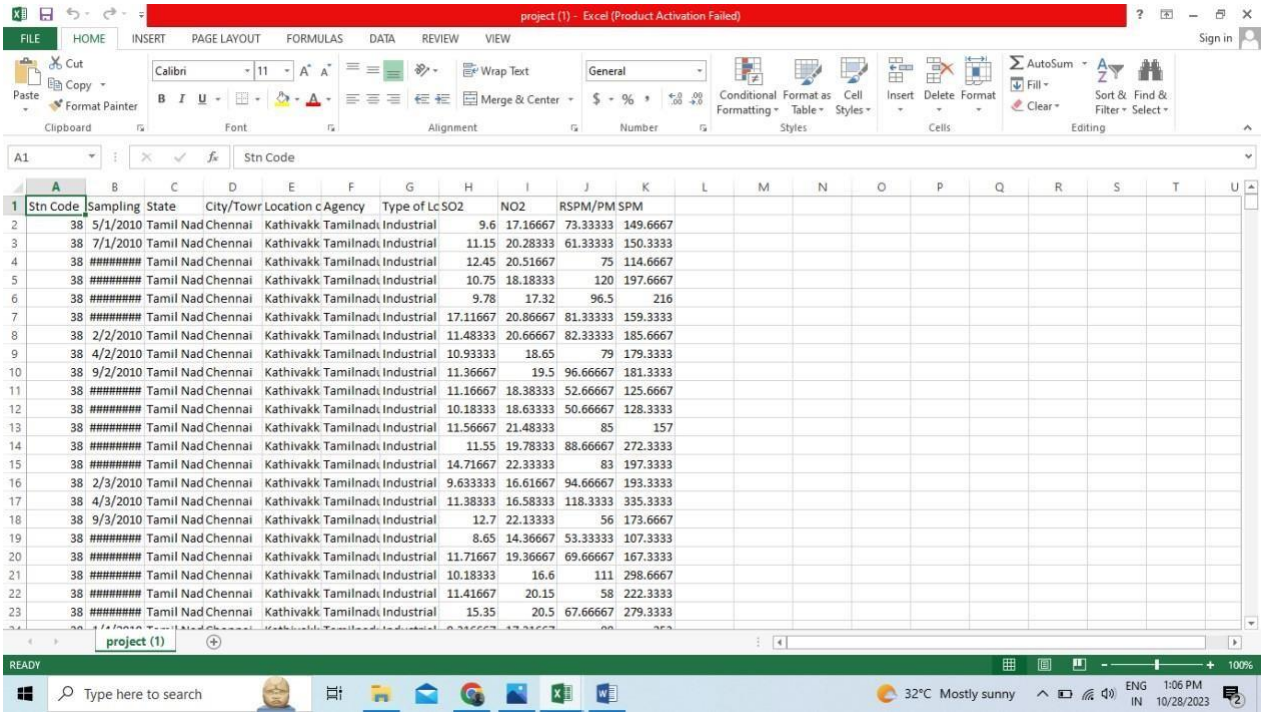
```
calculate_ni(N
ni=NO2*(50/40)

if(NO2>40 and NO2<=80): ni=50+(NO2-
40)*(50/40)

if(NO2>80 and NO2<=380): ni=100+(NO2-
80)*(100/300)
```

calculating RSPM AQI level

```
ef
  calculate_ri(RSPM_
  PM10): ri=0
  if(RSPM/PM10<=
    40):
    ri=RSPM/PM10
    *(50/40)

  if(RSPM/PM10>40 and
    RSPM/PM10<=80): ri=50+(RSPM/PM10-
    40)*(50/40)
data=df[['RSPM_PM10',
'ri']]print (data)
```

output: It is calculated and store in csv file

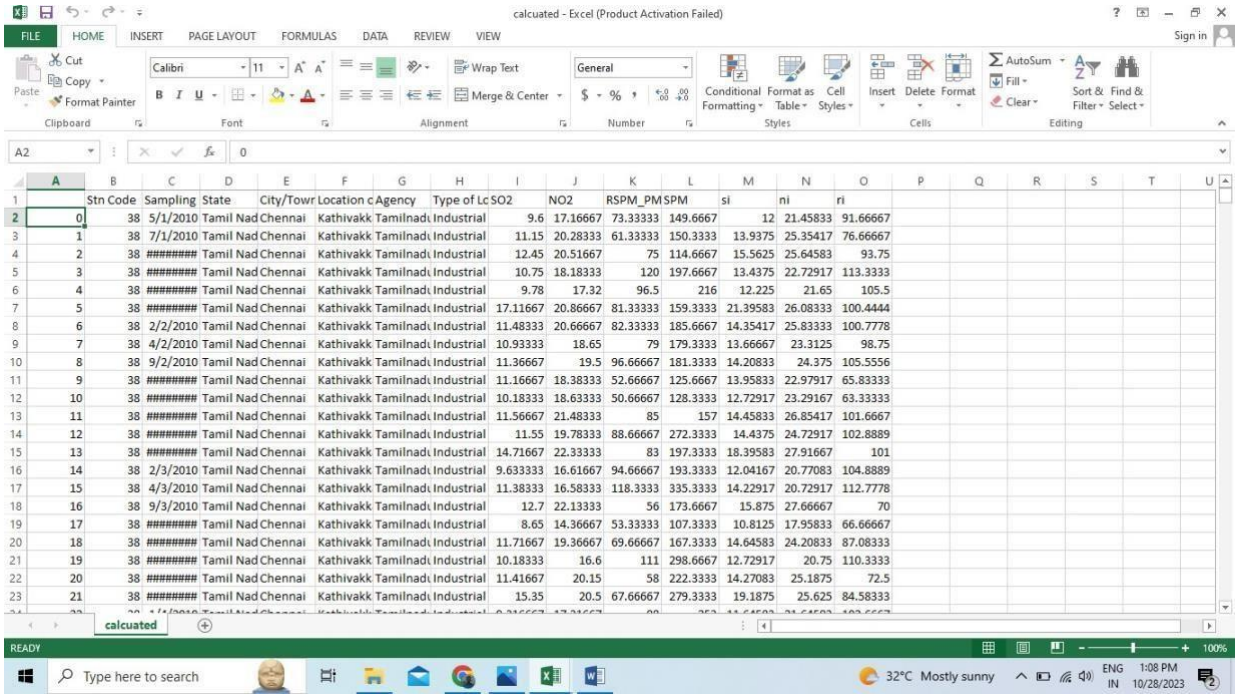
Before calcualting



Stn Code	Sampling	State	City/Tow	Location c	Agency	Type of Lc	SO2	NO2	RSPM/PM	SPM
38	5/1/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	9.6	17.16667	73.33333	149.6667
38	7/1/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.15	20.28333	61.33333	150.3333
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	12.45	20.51667	75	114.6667
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	10.75	18.18333	120	197.6667
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	9.78	17.32	96.5	216
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	17.11667	20.86667	81.33333	159.3333
38	2/2/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.48333	20.66667	82.33333	185.6667
38	4/2/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	10.93333	18.65	79	179.3333
38	9/2/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.36667	19.5	96.66667	181.3333
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.16667	18.38333	52.66667	125.6667
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	10.18333	18.63333	50.66667	128.3333
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.56667	21.48333	85	157
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.55	19.78333	88.66667	272.3333
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	14.71667	22.33333	83	197.3333
38	2/3/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	9.633333	16.61667	94.66667	193.3333
38	4/3/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.38333	16.58333	118.3333	335.3333
38	9/3/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	12.7	22.13333	56	173.6667
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	8.65	14.36667	53.33333	107.3333
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.71667	19.36667	69.66667	167.3333
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	10.18333	16.6	111	298.6667
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.41667	20.15	58	222.3333
38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	15.35	20.5	67.66667	279.3333

Calculate value of ni,si,ri respective value of AQI level of

SO2,NO2,RSPM/PM10



calcuated - Excel (Product Activation Failed)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		Stn Code	Sampling	State	City/Towr	Location c	Agency	Type of Lc	SO2	NO2	RSPM_PM	SPM	si	ni	ri
2	0	38	5/1/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	9.6	17.16667	73.33333	149.6667	12	21.45833	91.66667
3	1	38	7/1/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.15	20.28333	61.33333	150.3333	13.9375	25.35417	76.66667
4	2	38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	12.45	20.51667	75	114.6667	15.5625	25.64583	93.75
5	3	38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	10.75	18.18333	120	197.6667	13.4375	22.72917	113.3333
6	4	38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	9.78	17.32	96.5	216	12.225	21.65	105.5
7	5	38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	17.11667	20.86667	81.33333	159.3333	21.39583	26.08333	100.4444
8	6	38	2/2/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.48333	20.66667	82.33333	185.6667	14.35417	25.83333	100.7778
9	7	38	4/2/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	10.93333	18.65	79	179.3333	13.66667	23.3125	98.75
10	8	38	9/2/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.36667	19.5	96.66667	181.3333	14.20833	24.375	105.5556
11	9	38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.16667	18.38333	52.66667	125.6667	13.95833	22.97917	65.83333
12	10	38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	10.18333	18.63333	50.66667	128.3333	12.72917	23.29167	63.33333
13	11	38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.56667	21.48333	85	157	14.45833	26.85417	101.6667
14	12	38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.55	19.78333	88.66667	272.3333	14.4375	24.72917	102.8889
15	13	38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	14.71667	22.33333	83	197.3333	18.39583	27.91667	101
16	14	38	2/3/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	9.633333	16.61667	94.66667	193.3333	12.04167	20.77083	104.8889
17	15	38	4/3/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.38333	16.58333	118.3333	335.3333	14.22917	20.72917	112.7778
18	16	38	9/3/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	12.7	22.13333	56	173.6667	15.875	27.66667	70
19	17	38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	8.65	14.36667	53.33333	107.3333	10.8125	17.95833	66.66667
20	18	38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.71667	19.36667	69.66667	167.3333	14.64583	24.20833	87.08333
21	19	38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	10.18333	16.6	111	298.6667	12.72917	20.75	110.3333
22	20	38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.41667	20.15	58	222.3333	14.27083	25.1875	72.5
23	21	38	########	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	15.35	20.5	67.66667	279.3333	19.1875	25.625	84.58333

Predictive analysis and find out the Accuracy of the Model using svm

Python code for to check the accuracy of the model

```
from sklearn import datasets

from sklearn.model_selection import train_test_split from sklearn.preprocessing import StandardScaler from sklearn.svm import
```

```
SVC from sklearn.metrics import
accuracy_score



# Load the dataset (you will need to replace this with your own
data)
iris = datasets.load_iris() X
= iris.data y =
iris.target

# Split the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.3, random_state=42)



# Feature scaling for normalization sc
=
StandardScaler()
```

```
X_train = sc.fit_transform(X_train)

X_test = sc.transform(X_test)


# Train the SVM model svm_model =

SVC(kernel='linear', random_state=42)

svm_model.fit(X_train, y_train)


# Make predictions y_pred = svm_model.predict(X_test)



# Calculate accuracy accuracy =

accuracy_score(y_test, y_pred)

print(f'Accuracy: {accuracy}')
```

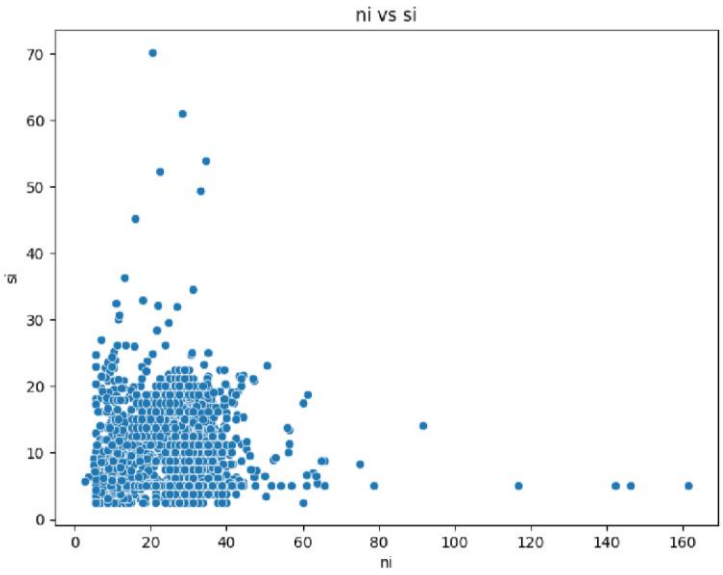
Accuracy of the Model**:0.70**

Creating a web application that would generate the visualization automatically for the given input data

```
from flask import Flask, render_template, request
import pandas as pd import matplotlib.pyplot as
plt
import seaborn as sns import
os


app = Flask(__name__)


UPLOAD_FOLDER = 'uploads'
app.config['UPLOAD_FOLDER'] = UPLOAD_FOLDER


def create_visualization(data, x, y):    if x in
data.columns and y in data.columns:
    plt.figure(figsize=(8, 6))
sns.scatterplot(x=x, y=y, data=data)
plt.title(f'{x} vs {y}')       plt.xlabel(x)
plt.ylabel(y)
    plt.savefig('static/visualization.png')
else:
    return "Required columns not found in the dataset."
```

```
@app.route('/')
def index():
   return render_template('index.html') @app.route('/upload',
methods=['POST']) def upload_file():     if request.method ==
'POST':        file = request.files['file']        x = request.form['x']
y = request.form['y']        if file:
          df = pd.read_csv(file)
          result = create_visualization(df, x, y)
          if result == "Required columns not found in the
dataset.":
               return result
else:
               return render_template('visualization.html')
return "Upload failed"


if __name__ == '__main__':     if not
os.path.exists(UPLOAD_FOLDER):
       os.makedirs(UPLOAD_FOLDER)
app.run(debug=True) output:
```

AQI Visualization



Outcome of the project:

1.The assessment provides the AQI values for various regions in Tamil Nadu, indicating the overall air quality and pollutant concentrations.

2. Information on the levels of specific pollutants such as PM2.5, PM10, NO2, SO2, CO, and O3, which are commonly measured to assess air quality.

3.Assessments often include data on the potential health effects associated with exposure to air pollution, including respiratory and cardiovascular issues.This helps to reduce the pollution in the respected areas.

4. This assessment offer recommendations for improving air quality, such as policy changes, emission controls, and public awareness campaigns.