Naan Mudhalvan Project

Air Quality Analysis in Tamil Nadu

Phase 3

Team Members :

1. Hematharshini E – 2021115041
   Email - hemae2512@gmail.com
2. Sabitha S – 2021115087
   Email - sabitha.suresh15@gmail.com
3. Sandhya Shankar – 2021115090
   Email - sandhya.shankar.2002@gmail.com
4. Sanmitha V.S – 2021115092
   Email - sanmithasadhishkumar@gmail.com
5. Akash P – 2021115314
   Email - akashpanneer2004@gmail.com

Phase 3: Development Part I

Overview:

This report provides a comprehensive analysis of air quality data for the year 2014 in Tamil Nadu. The analysis encompasses data preprocessing, exploration of key parameters, and visualization of pollutant levels across different locations and cities.

Data Loading and Preprocessing:

The data was loaded from the CSV file 'cpcb_dly_aq_tamil_nadu-2014.csv'. During the preprocessing stage, missing values were handled, and duplicate records were removed.

- **Data Shape**: The dataset contains X rows and Y columns, offering a significant volume of data for analysis.
- **Missing Values**: Null values in the PM2.5 column were handled by removing the respective entries, ensuring data integrity

```
print("INFO:")
print(df.info())

print("\nDescribe:")
print(df.describe())

print("\nShape")
print(df.shape)
```

```
<bound method NDFrame.head of      Stn Code Sampling Date       State City/Town/Village/Area  ...   SO2   NO2 RSPM/PM10  PM 2.5
0          38      01-02-14  Tamil Nadu               Chennai  ...  11.0  17.0      55.0     NaN
1          38      01-07-14  Tamil Nadu               Chennai  ...  13.0  17.0      45.0     NaN
2          38      21-01-14  Tamil Nadu               Chennai  ...  12.0  18.0      50.0     NaN
3          38      23-01-14  Tamil Nadu               Chennai  ...  15.0  16.0      46.0     NaN
4          38      28-01-14  Tamil Nadu               Chennai  ...  13.0  14.0      42.0     NaN
...       ...           ...         ...                   ...  ...   ...   ...       ...     ...
2874      773      12-03-14  Tamil Nadu                Trichy  ...  15.0  18.0     102.0     NaN
2875      773      12-10-14  Tamil Nadu                Trichy  ...  12.0  14.0      91.0     NaN
2876      773      17-12-14  Tamil Nadu                Trichy  ...  19.0  22.0     100.0     NaN
2877      773      24-12-14  Tamil Nadu                Trichy  ...  15.0  17.0      95.0     NaN
2878      773      31-12-14  Tamil Nadu                Trichy  ...  14.0  16.0      94.0     NaN

[2879 rows x 11 columns]>
```

```
Describe:
          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("\nREMOVING COLUMNS WITH NULL VALUES\n ")
df = df.drop('PM 2.5', axis=1)
df.dropna(inplace=True)
```

```
REMOVING COLUMNS WITH NULL VALUES
```

```
print("\nDROPPING DUPLICATE ROWS:\n")
df.drop_duplicates(subset=None, inplace=True)
print(df.head)
```

```
DROPPING DUPLICATE ROWS:

<bound method NDFrame.head of       Stn Code Sampling Date       State City/Town/Village/Area  ...                    Type of Loc
ation   SO2   NO2  RSPM/PM10
0           38      01-02-14  Tamil Nadu               Chennai  ...                    Industrial Area  11.0  17.0       55.0
1           38      01-07-14  Tamil Nadu               Chennai  ...                    Industrial Area  13.0  17.0       45.0
2           38      21-01-14  Tamil Nadu               Chennai  ...                    Industrial Area  12.0  18.0       50.0
3           38      23-01-14  Tamil Nadu               Chennai  ...                    Industrial Area  15.0  16.0       46.0
4           38      28-01-14  Tamil Nadu               Chennai  ...                    Industrial Area  13.0  14.0       42.0
...        ...           ...         ...                   ...  ...                                ...   ...   ...        ...
2874       773      12-03-14  Tamil Nadu                Trichy  ...  Residential, Rural and other Areas  15.0  18.0      102.0
2875       773      12-10-14  Tamil Nadu                Trichy  ...  Residential, Rural and other Areas  12.0  14.0       91.0
2876       773      17-12-14  Tamil Nadu                Trichy  ...  Residential, Rural and other Areas  19.0  22.0      100.0
2877       773      24-12-14  Tamil Nadu                Trichy  ...  Residential, Rural and other Areas  15.0  17.0       95.0
2878       773      31-12-14  Tamil Nadu                Trichy  ...  Residential, Rural and other Areas  14.0  16.0       94.0

[2862 rows x 10 columns]>

CONVERTING TO DATE-TIME FORMAT

d:\nm_dsc\preair.py:21: UserWarning: Could not infer format, so each element will be parsed individually, falling back to `dateut
il`. To ensure parsing is consistent and as-expected, please specify a format.
  df['Sampling Date'] = pd.to_datetime(df['Sampling Date'])

Head after preprocessing:
<bound method NDFrame.head of       Stn Code Sampling Date       State City/Town/Village/Area  ...                    Type of Loc
ation   SO2   NO2  RSPM/PM10
0           38    2014-01-02  Tamil Nadu               Chennai  ...                    Industrial Area  11.0  17.0       55.0
1           38    2014-01-07  Tamil Nadu               Chennai  ...                    Industrial Area  13.0  17.0       45.0
2           38    2014-01-21  Tamil Nadu               Chennai  ...                    Industrial Area  12.0  18.0       50.0
3           38    2014-01-23  Tamil Nadu               Chennai  ...                    Industrial Area  15.0  16.0       46.0
```

Data Exploration:

Summary Statistics:

- **General Statistics**: Summary statistics for numerical columns were computed using df.describe(). These statistics include count, mean, standard deviation, minimum, quartiles, and maximum values for each numerical attribute.

Unique Locations and Cities:

- **Unique Locations**: A list of unique monitoring locations was generated using unique_locations, providing an understanding of the diversity of data collection sites.

- **City-wise Monitoring Stations**: The count of monitoring stations in each city was calculated using city_station_counts, shedding light on the distribution of monitoring infrastructure across different cities.

```
unique_locations = df['Location of Monitoring Station'].unique()
print("\nLocations of Monitoring Stations:")
print(unique_locations)
```

```
Locations of Monitoring Stations:
['Kathivakkam, Municipal Kalyana Mandapam, Chennai'
 'Govt. High School, Manali, Chennai.' 'Thiruvottiyur,  Chennai'
 'Thiyagaraya Nagar, Chennai' 'Anna Nagar, Chennai' 'Adyar, Chennai'
 'Kilpauk, Chennai' 'Madras Medical College, Chennai'
 'Thiruvottiyur Municipal Office, Chennai' 'NEERI, CSIR Campus Chennai'
 'Poniarajapuram, On the top of DEL, Coimbatore'
 'SIDCO Office, Coimbatore' "Distt. Collector's Office, Coimbatore"
 'Eachangadu Villagae'
 'District Environmental Engineer Office, Imperial Road, Cuddalore'
 'SIPCOT Industrial Complex, Cuddalore'
 'Highway (Project -I) Building, Madurai'
 'Fenner (I) Ltd. Employees Assiciation Building Kochadai, Madurai'
 'Kunnathur Chatram East Avani Mollai Street, Madurai'
 'Raman Nagar, Mettur' 'SIDCO Industrial Complex, Mettur'
 'Sowdeswari College Building, Salem' 'Fisheries College, Tuticorin'
 'AVM Jewellery Building, Tuticorin' 'Raja Agencies, Tuticorin'
 'Gandhi Market, Trichy' 'Main Guard Gate, Tirchy'
 'Bishop Heber College, Tirchy' 'Golden Rock, Trichy'
 'Central Bus Stand, Trichy']
```

```
city_station_counts = df.groupby('City/Town/Village/Area')['Location of Monitoring Station'].count().reset_index()

city_station_counts.columns = ['City', 'Number of Monitoring Stations']

print("\nCity-wise Number of Monitoring Stations:")
print(city_station_counts)
```

```
City-wise Number of Monitoring Stations:
          City  Number of Monitoring Stations
0      Chennai                            995
1   Coimbatore                            289
2    Cuddalore                            294
3      Madurai                            294
4       Mettur                            205
5        Salem                            131
6  Thoothukudi                            290
7       Trichy                            364
```

```
location_counts = df.groupby(['City/Town/Village/Area', 'Location of Monitoring Station']).size().reset_index()
location_counts.columns = ['City', 'Location', 'Number of Rows']

print("\nLocation-wise Number of Rows with City:")
print(location_counts)
```

```
Location-wise Number of Rows with City:
          City                                          Location  \
0      Chennai                                     Adyar, Chennai
1      Chennai                                 Anna Nagar, Chennai
2      Chennai              Govt. High School, Manali, Chennai.
3      Chennai   Kathivakkam, Municipal Kalyana Mandapam, Chennai
4      Chennai                                   Kilpauk, Chennai
5      Chennai                  Madras Medical College, Chennai
6      Chennai                     NEERI, CSIR Campus Chennai
7      Chennai          Thiruvottiyur Municipal Office, Chennai
8      Chennai                           Thiruvottiyur,  Chennai
9      Chennai                       Thiyagaraya Nagar, Chennai
10   Coimbatore            Distt. Collector's Office, Coimbatore
11   Coimbatore     Poniarajapuram, On the top of DEL, Coimbatore
12   Coimbatore                        SIDCO Office, Coimbatore
13    Cuddalore  District Environmental Engineer Office, Imperi...
14    Cuddalore                               Eachangadu Villagae
15    Cuddalore             SIPCOT Industrial Complex, Cuddalore
16      Madurai  Fenner (I) Ltd. Employees Assiciation Building...
17      Madurai           Highway (Project -I) Building, Madurai
18      Madurai  Kunnathur Chatram East Avani Mollai Street, Ma...
19       Mettur                               Raman Nagar, Mettur
20       Mettur                 SIDCO Industrial Complex, Mettur
21        Salem               Sowdeswari College Building, Salem
22  Thoothukudi                AVM Jewellery Building, Tuticorin
23  Thoothukudi                     Fisheries College, Tuticorin
24  Thoothukudi                         Raja Agencies, Tuticorin
25       Trichy                    Bishop Heber College, Tirchy
26       Trichy                       Central Bus Stand, Trichy
27       Trichy                           Gandhi Market, Trichy
28       Trichy                             Golden Rock, Trichy
29       Trichy                         Main Guard Gate, Tirchy
```

Pollution Levels:

- **Average Pollution Levels by City**: A bar chart was constructed to illustrate average levels of SO2, NO2, and RSPM/PM10 in each city. This offers a comparative view of pollution across various cities.

```
summary = df.groupby(['City/Town/Village/Area', 'Location of Monitoring Station'])[['SO2', 'NO2', 'RSPM/PM10']].agg(['sum', 'mean']).reset_index()

summary.columns = ['City', 'Location', 'SO2 Sum', 'SO2 Average', 'NO2 Sum', 'NO2 Average', 'RSPM/PM10 Sum', 'RSPM/PM10 Average']

print("\nSummary of SO2, NO2, and RSPM/PM10 Levels by Location:")
print(summary)
```

```
Summary of SO2, NO2, and RSPM/PM10 Levels by Location:
          City                                          Location  SO2 Sum  \
0      Chennai                                    Adyar, Chennai   1215.0
1      Chennai                                  Anna Nagar, Chennai   1527.0
2      Chennai                  Govt. High School, Manali, Chennai.   1213.0
3      Chennai   Kathivakkam, Municipal Kalyana Mandapam, Chennai   1215.0
4      Chennai                                   Kilpauk, Chennai   2231.0
5      Chennai                      Madras Medical College, Chennai    638.0
6      Chennai                           NEERI, CSIR Campus Chennai    516.0
7      Chennai              Thiruvottiyur Municipal Office, Chennai    719.0
8      Chennai                               Thiruvottiyur,  Chennai   1249.0
9      Chennai                             Thiyagaraya Nagar, Chennai   2114.0
10   Coimbatore                Distt. Collector's Office, Coimbatore    405.0
11   Coimbatore        Poniarajapuram, On the top of DEL, Coimbatore    425.0
12   Coimbatore                              SIDCO Office, Coimbatore    482.0
13    Cuddalore  District Environmental Engineer Office, Imperi...    802.0
14    Cuddalore                                  Eachangadu Villagae   1144.0
15    Cuddalore                SIPCOT Industrial Complex, Cuddalore    690.0
16      Madurai  Fenner (I) Ltd. Employees Assiciation Building...   1378.0
17      Madurai              Highway (Project -I) Building, Madurai   1147.0
18      Madurai  Kunnathur Chatram East Avani Mollai Street, Ma...   1391.0
19       Mettur                                   Raman Nagar, Mettur    780.0
20       Mettur                    SIDCO Industrial Complex, Mettur    948.0
21        Salem                  Sowdeswari College Building, Salem   1063.0
22  Thoothukudi                   AVM Jewellery Building, Tuticorin    893.0
23  Thoothukudi                        Fisheries College, Tuticorin   1351.0
24  Thoothukudi                            Raja Agencies, Tuticorin   1521.0
25       Trichy                        Bishop Heber College, Tirchy    826.0
26       Trichy                           Central Bus Stand, Trichy   1351.0
27       Trichy                               Gandhi Market, Trichy   1269.0
28       Trichy                                 Golden Rock, Trichy    853.0
29       Trichy                             Main Guard Gate, Tirchy   1268.0
```

Data Visualization

Pollutant Levels by City:

Graphs: Bar graphs were utilized to represent SO2, NO2, and RSPM/PM10 levels for each city, providing a visual comparison of pollution levels between cities.

> **Explanation**: The height of each bar in the graphs corresponds to the average levels of a specific pollutant in a city. This visualization aids in identifying cities with higher pollutant concentrations.

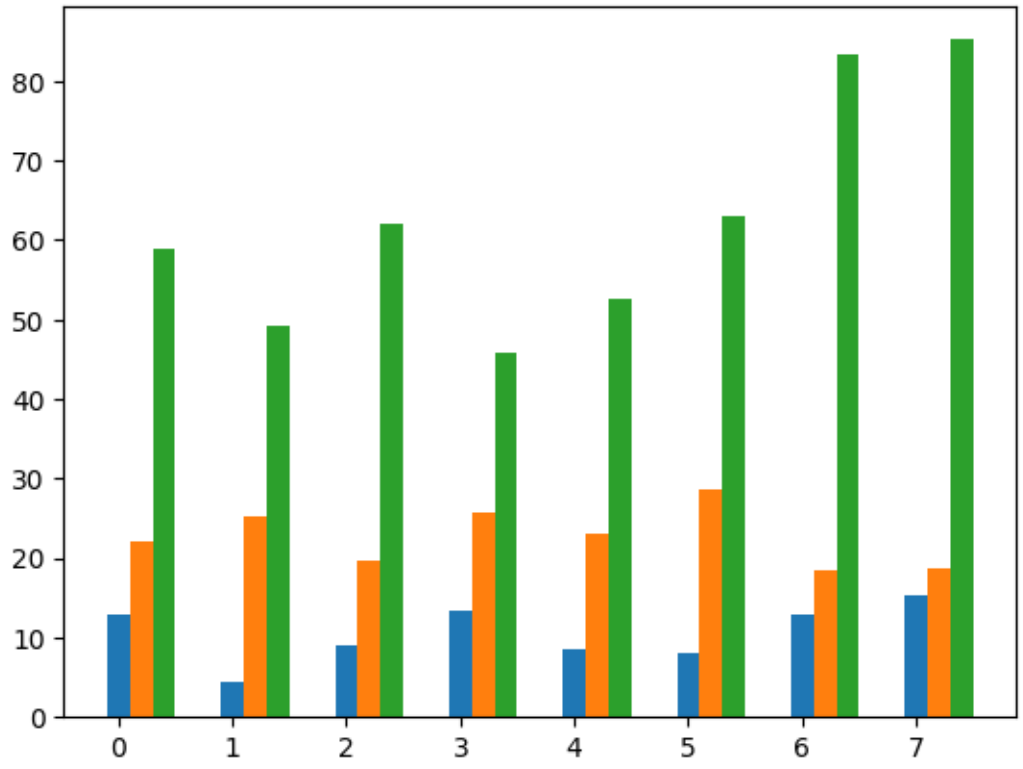
Pollutant Levels by Location:

- **Graphs**: Bar graphs were employed to depict SO2, NO2, and RSPM/PM10 levels for each location within a city. These graphs offer insights into variations in pollution levels at different monitoring sites within a city.

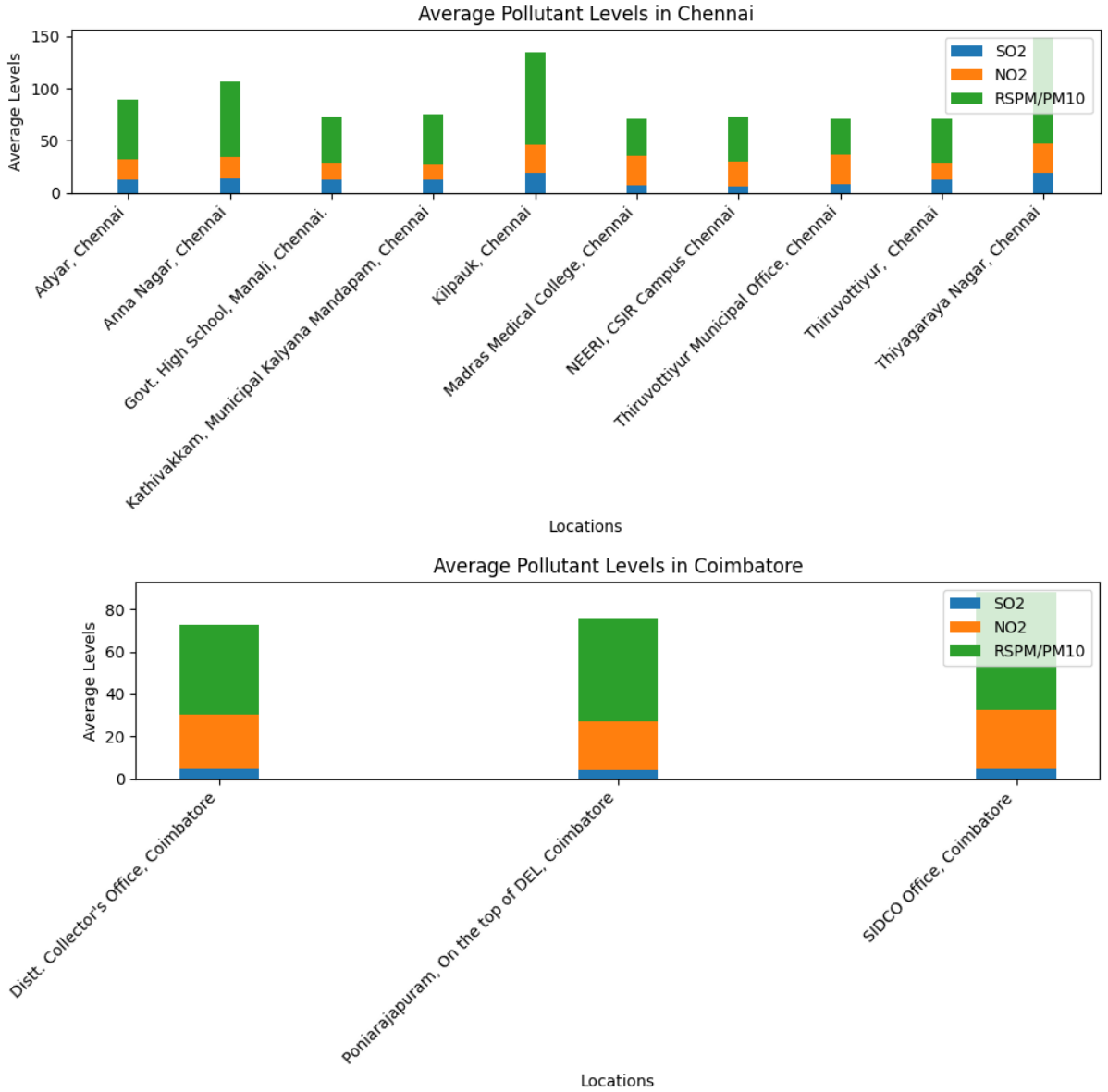
```
cities = city_avg['City']
so2_avg = city_avg['SO2 Average']
no2_avg = city_avg['NO2 Average']
rspm_avg = city_avg['RSPM/PM10 Average']
```

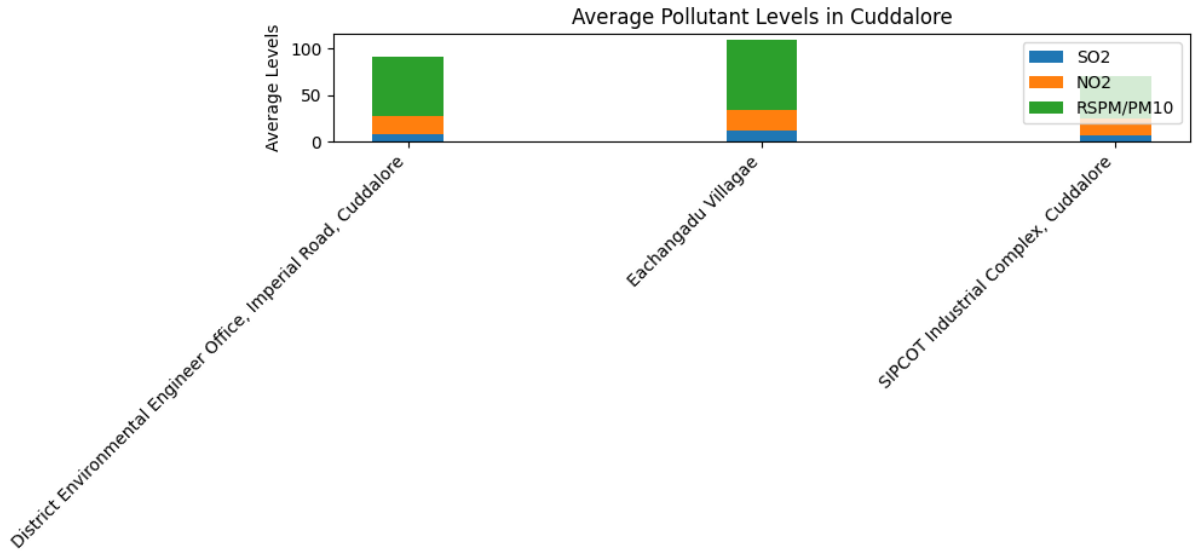
```
bar_width = 0.2

r1 = range(len(cities))
r2 = [x + bar_width for x in r1]
r3 = [x + bar_width for x in r2]
plt.bar(r1, so2_avg, width=bar_width, label='SO2')
plt.bar(r2, no2_avg, width=bar_width, label='NO2')
plt.bar(r3, rspm_avg, width=bar_width, label='RSPM/PM10')
```

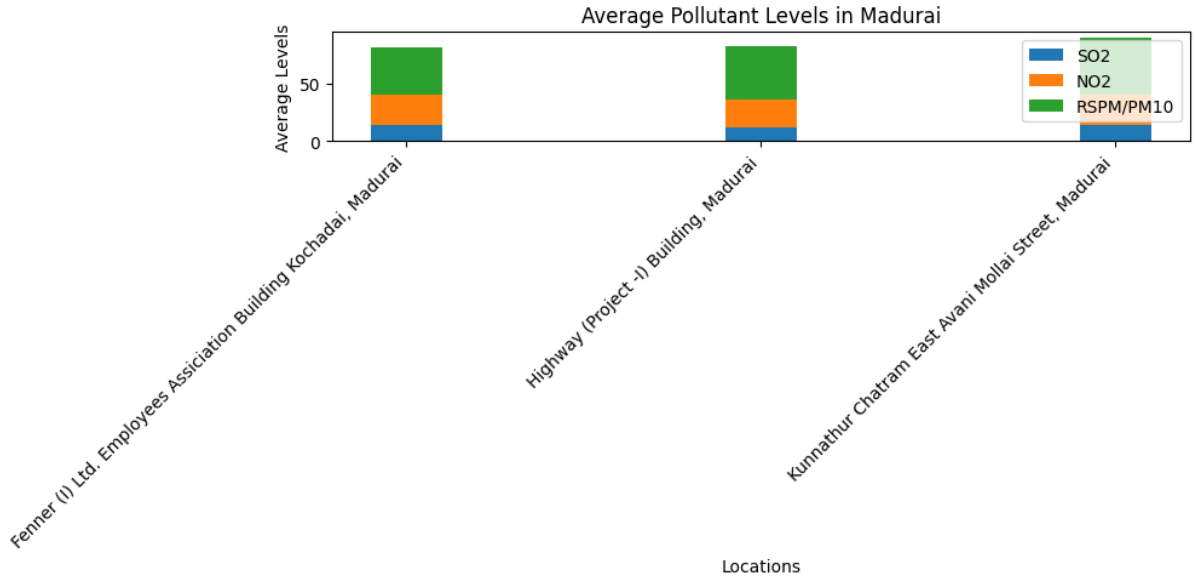
```
plt.xlabel('Cities')
plt.xticks([x + bar_width for x in r1], cities, rotation=90)

plt.ylabel('Average Levels')

plt.title('Average SO2, NO2, and RSPM/PM10 Levels by City')

plt.legend()

plt.tight_layout()
plt.show()
```

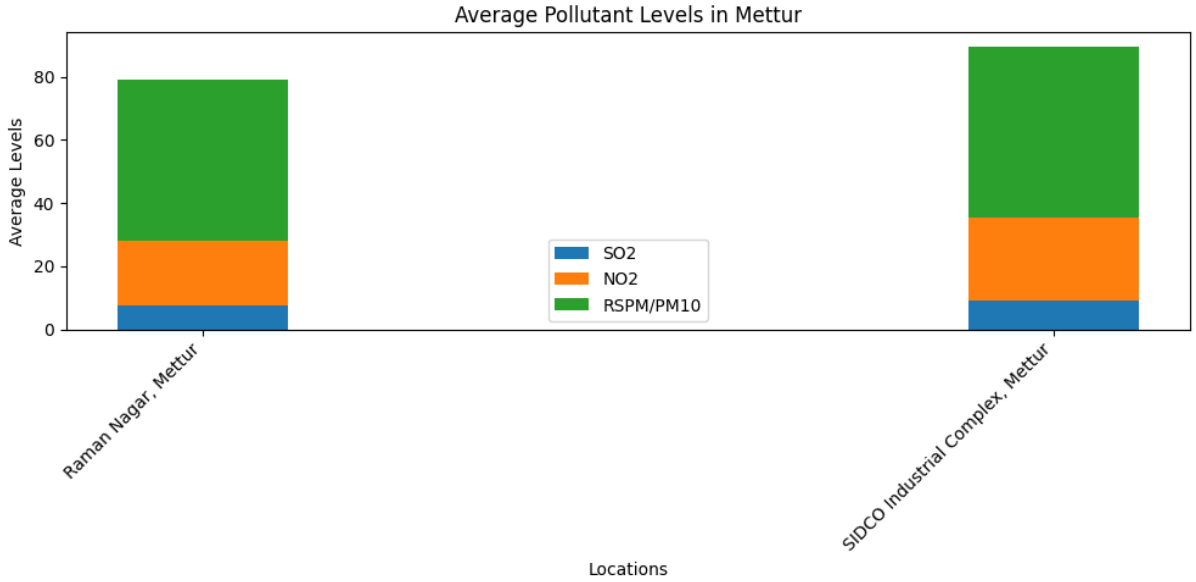


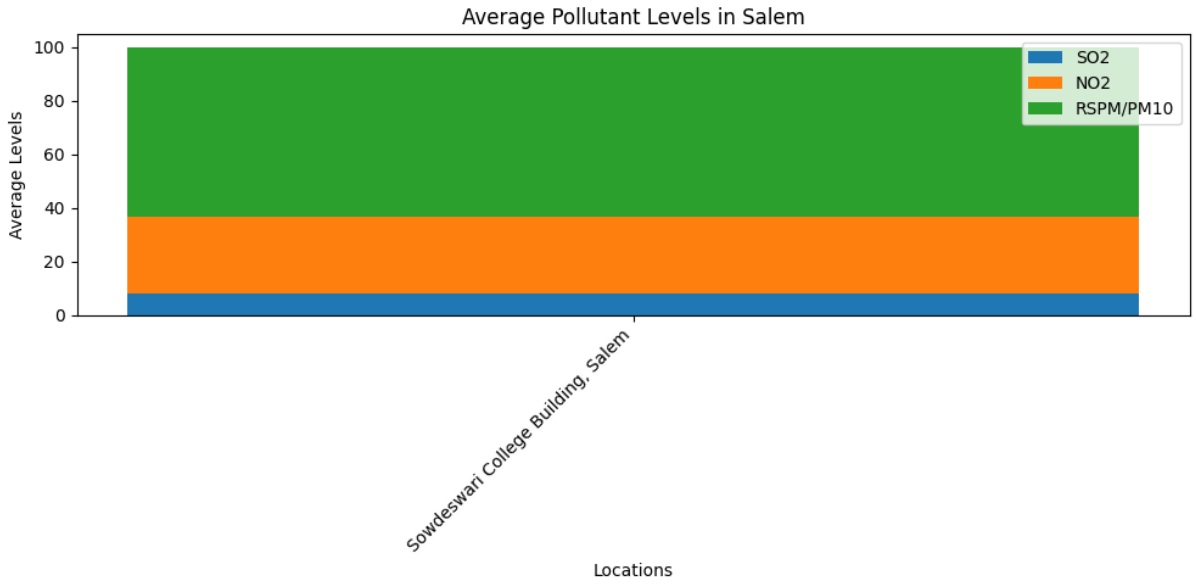
```python
import matplotlib.pyplot as plt

unique_cities = summary['City'].unique()

for city in unique_cities:
    city_data = summary[summary['City'] == city]

    locations = city_data['Location']
    so2_avg = city_data['SO2 Average']
    no2_avg = city_data['NO2 Average']
    rspm_avg = city_data['RSPM/PM10 Average']

    plt.figure(figsize=(10, 5))
    plt.bar(locations, so2_avg, width=0.2, label='SO2')
    plt.bar(locations, no2_avg, width=0.2, label='NO2', bottom=so2_avg)
    plt.bar(locations, rspm_avg, width=0.2, label='RSPM/PM10', bottom=so2_avg + no2_avg)

    plt.xlabel('Locations')
    plt.xticks(rotation=45, ha='right')

    plt.ylabel('Average Levels')

    plt.title(f'Average Pollutant Levels in {city}')

    plt.legend()

    plt.tight_layout()
    plt.show()
```

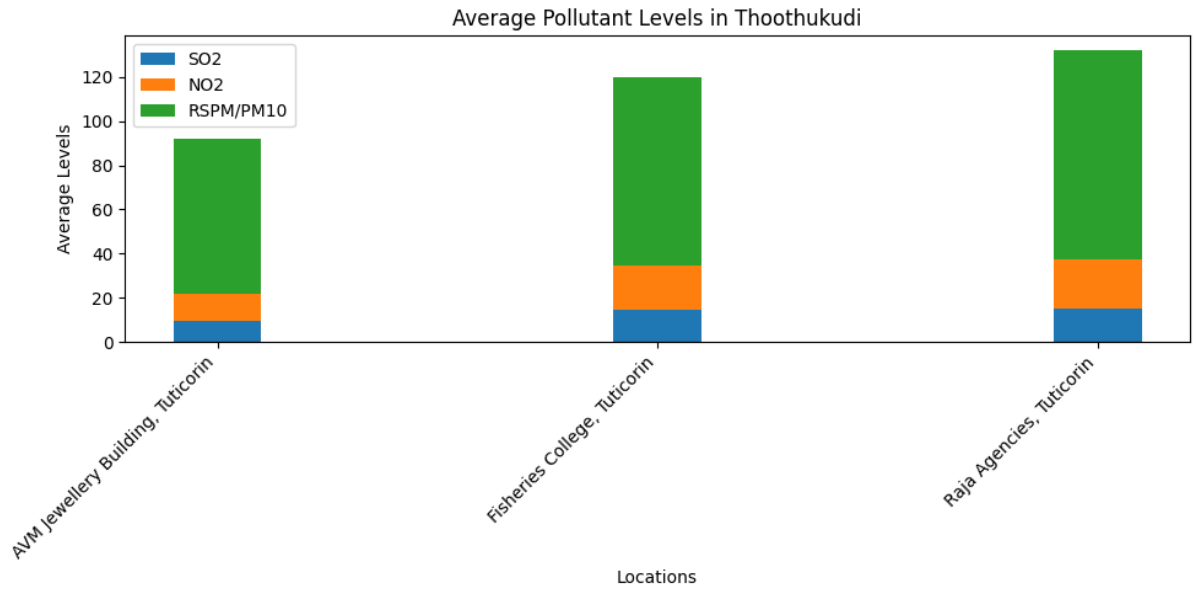


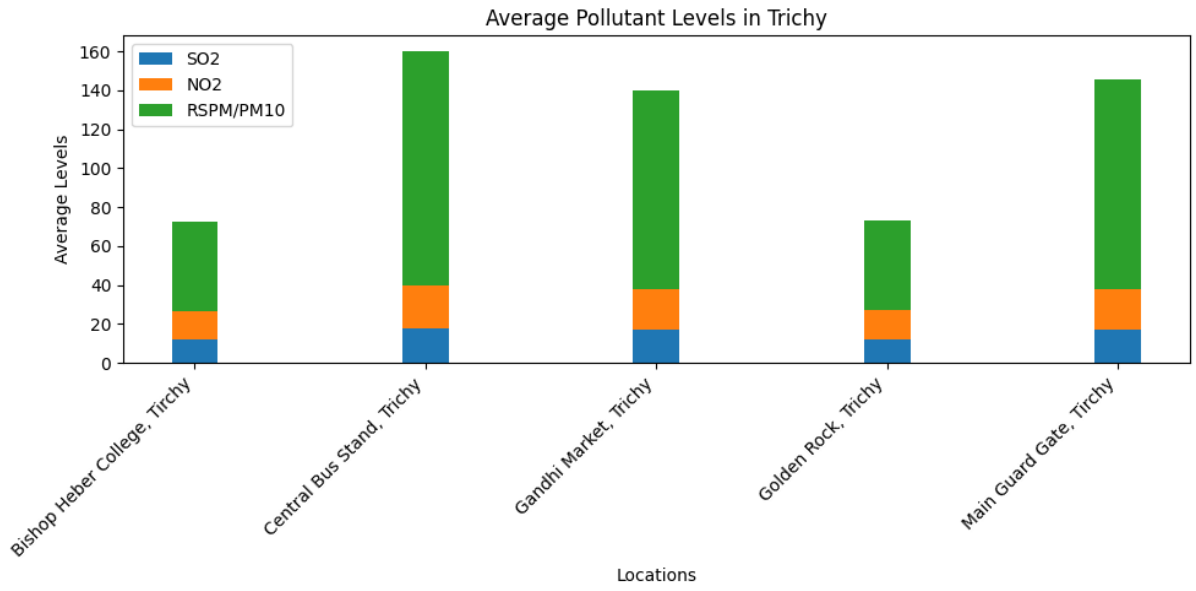

Average Pollutant Levels in Cuddalore
Average Levels
100
50
0
SO2
NO2
RSPM/PM10
District Environmental Engineer Office, Imperial Road, Cuddalore
Eachangadu Villagae
SIPCOT Industrial Complex, Cuddalore
Locations


Average Pollutant Levels in Madurai
Average Levels
50
0
SO2
NO2
RSPM/PM10
Fenner (I) Ltd. Employees Assiciation Building Kochadai, Madurai
Highway (Project -I) Building, Madurai
Kunnathur Chatram East Avani Mollai Street, Madurai
Locations


Average Pollutant Levels in Mettur
Average Levels
80
60
40
20
0
SO2
NO2
RSPM/PM10
Raman Nagar, Mettur
SIDCO Industrial Complex, Mettur
Locations


Average Pollutant Levels in Salem
SO2
NO2
RSPM/PM10
Average Levels
0
20
40
60
80
100
Sowdeswari College Building, Salem
Locations


Average Pollutant Levels in Thoothukudi
SO2
NO2
RSPM/PM10
Average Levels
0
20
40
60
80
100
120
AVM Jewellery Building, Tuticorin
Fisheries College, Tuticorin
Raja Agencies, Tuticorin
Locations


Average Pollutant Levels in Trichy
SO2
NO2
RSPM/PM10
Average Levels
0
20
40
60
80
100
120
140
160
Bishop Heber College, Tirchy
Central Bus Stand, Trichy
Gandhi Market, Trichy
Golden Rock, Trichy
Main Guard Gate, Tirchy
Locations

Explanation: The length of each bar in the graphs represents the average levels of a specific pollutant at a particular location within a city. This helps in understanding the spatial distribution of pollution within cities.

Conclusion:

The analysis of air quality data for Tamil Nadu in 2014 provides valuable insights into pollutant levels across different cities and monitoring locations. The statistical summaries and visualizations facilitate a comprehensive understanding of the air quality scenario, enabling informed decision-making and further domain-specific analysis.