GOVERNMENT COLLEGE OF TECHNOLOGY, COIMBATORE-13

PROJECT NAME: Air Quality Assessment TN

Team Members:

1. Kokulavenkat K
2. Keerthana R
3. Kayalvizhi T
4. Karunasri A

Phase 4: PROJECT DEVELOPMENT LEVEL 2

PROJECT DESCRIPTION:

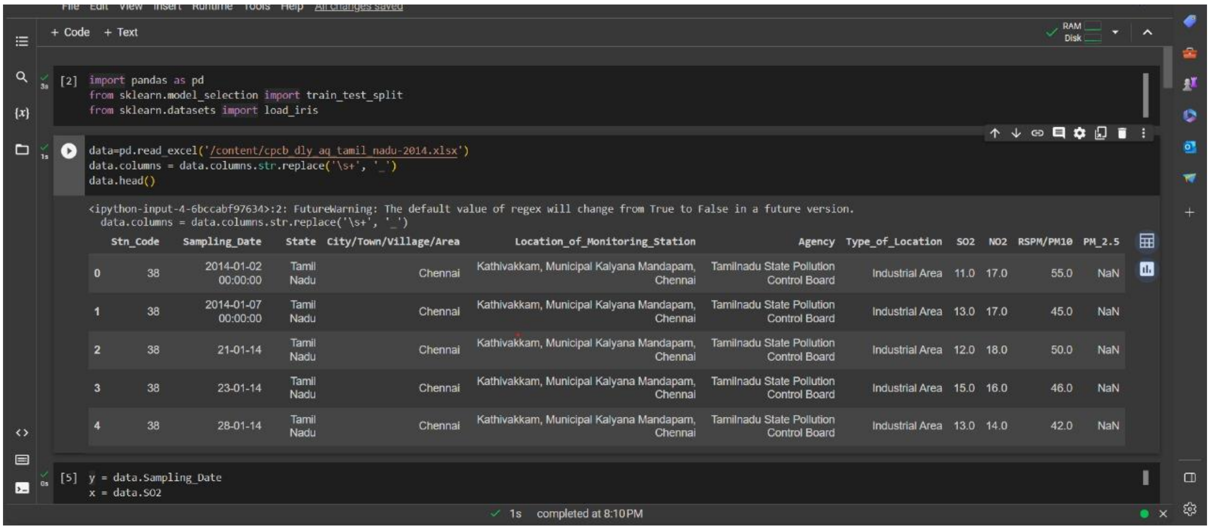
Gather air quality data from monitoring stations across Tamil Nadu. This data may include measurements of pollutants like PM2.5, PM10, nitrogen dioxide (NO2), sulphur dioxide (SO2), ozone (O3), carbon monoxide (CO), and other relevant parameters.

Ensure the data is structured, accurate, and includes timestamps for each measurement. Aggregate data if multiple stations provide data for the same location. To visualise the data including line charts, bar graphs, heatmaps, and scatter plots and compare different pollutants. Geographic visualizations using tools like GIS software or Python libraries like Folium to map air quality across different regions in Tamil Nadu.

Consider using color-coding to represent air quality levels and station locations on the map. Calculate AQI based on the collected data and the relevant formula for Tamil Nadu. To visualise the air quality find how different pollutants relate to each other and apply time series analysis techniques, autoregressive integrated moving average (ARIMA) or prophet for forecasting air quality.

LOADING AN DATASET:

- Load a dataset into a pandas DataFrame using `pd.read_csv()` for CSV files or `pd.read_excel()` for Excel files.
- Ensure that 'data' is correctly loaded before applying the provided code for preprocessing.



```
[2] import pandas as pd
    from sklearn.model_selection import train_test_split
    from sklearn.datasets import load_iris
```

```
data=pd.read_excel('/content/cpcb_dly_aq_tamil_nadu-2014.xlsx')
data.columns = data.columns.str.replace('\s+', '_')
data.head()
```

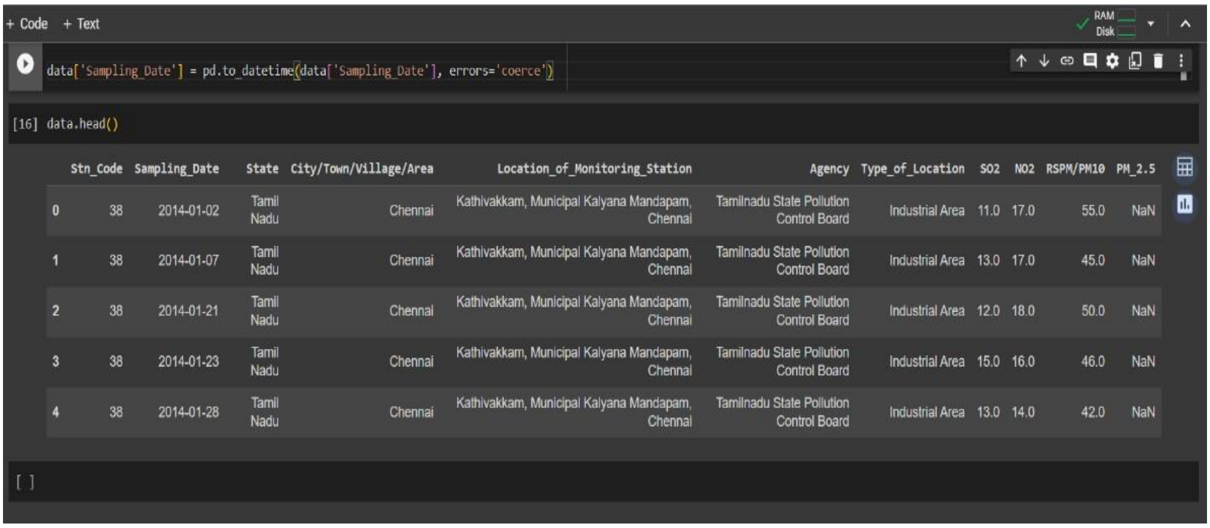
```
<ipython-input-4-6bccabf97634>:2: FutureWarning: The default value of regex will change from True to False in a future version.
  data.columns = data.columns.str.replace('\s+', '_')
```

	Stn_Code	Sampling_Date	State	City/Town/Village/Area	Location_of_Monitoring_Station	Agency	Type_of_Location	SO2	NO2	RSPM/PM10	PM_2.5
0	38	2014-01-02 00:00:00	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	11.0	17.0	55.0	NaN
1	38	2014-01-07 00:00:00	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	13.0	17.0	45.0	NaN
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	NaN
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	NaN
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	NaN

```
[5] y = data.Sampling_Date
    x = data.SO2
```

DATA CONVERSION:

Convert the 'Sampling_Date' column to a datetime format, handling errors by using NaT.



```
data['Sampling_Date'] = pd.to_datetime(data['Sampling_Date'], errors='coerce')
```

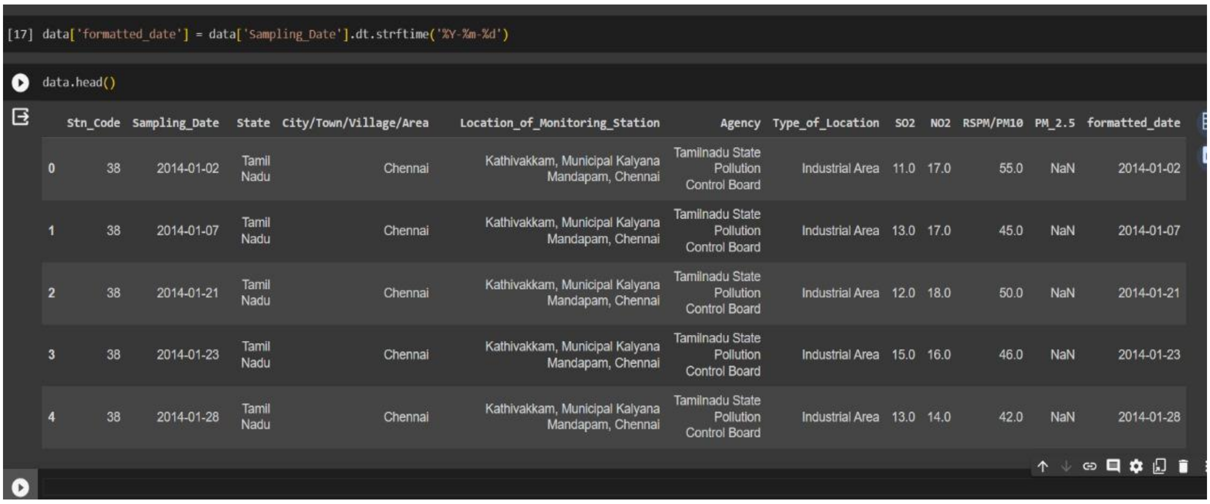
```
[16] data.head()
```

	Stn_Code	Sampling_Date	State	City/Town/Village/Area	Location_of_Monitoring_Station	Agency	Type_of_Location	SO2	NO2	RSPM/PM10	PM_2.5
0	38	2014-01-02	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	11.0	17.0	55.0	NaN
1	38	2014-01-07	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	13.0	17.0	45.0	NaN
2	38	2014-01-21	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	12.0	18.0	50.0	NaN
3	38	2014-01-23	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	15.0	16.0	46.0	NaN
4	38	2014-01-28	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	13.0	14.0	42.0	NaN

```
[ ]
```

DATA FORMATTING:

Create a new 'formatted_date' column with a specific date format (YYYY-MM-DD).



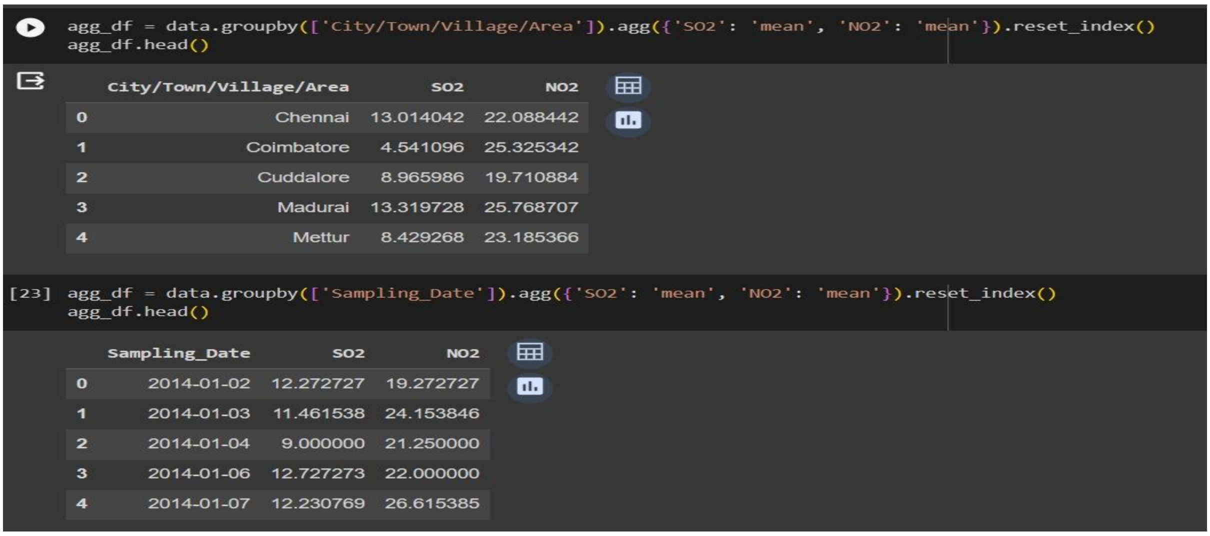
```
[17] data['formatted_date'] = data['Sampling_Date'].dt.strftime('%Y-%m-%d')
```

```
data.head()
```

	Stn_Code	Sampling_Date	State	City/Town/Village/Area	Location_of_Monitoring_Station	Agency	Type_of_Location	SO2	NO2	RSPM/PM10	PM_2.5	formatted_date
0	38	2014-01-02	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	11.0	17.0	55.0	NaN	2014-01-02
1	38	2014-01-07	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	13.0	17.0	45.0	NaN	2014-01-07
2	38	2014-01-21	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	12.0	18.0	50.0	NaN	2014-01-21
3	38	2014-01-23	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	15.0	16.0	46.0	NaN	2014-01-23
4	38	2014-01-28	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	13.0	14.0	42.0	NaN	2014-01-28

DATA AGGREGATION:

Aggregate data by 'City/Town/Village/Area' and 'Sampling_Date', calculating mean values for 'SO2' and 'NO2' for each group.



```
agg_df = data.groupby(['City/Town/Village/Area']).agg({'SO2': 'mean', 'NO2': 'mean'}).reset_index()
agg_df.head()
```

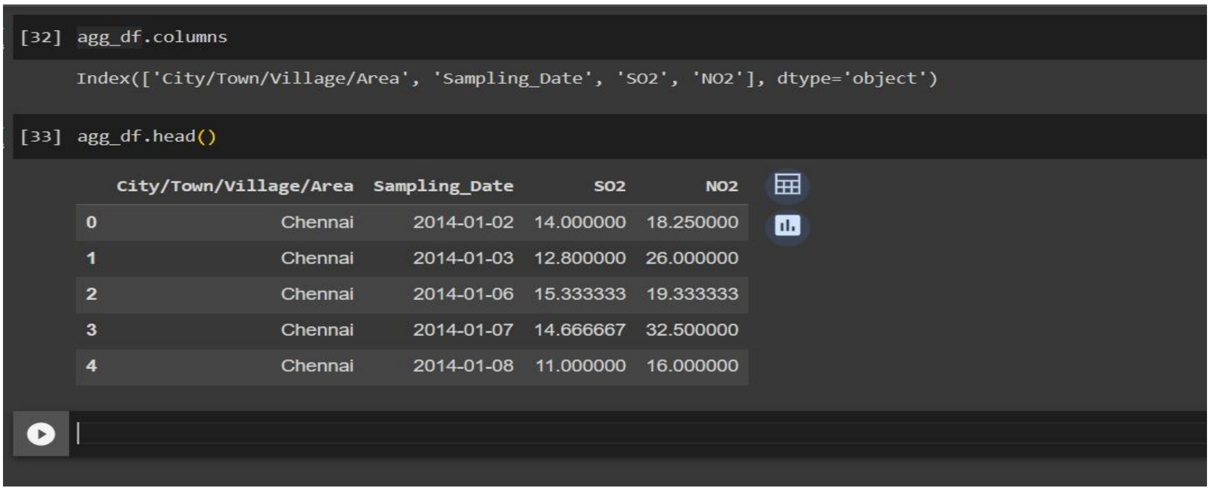
	City/Town/Village/Area	SO2	NO2
0	Chennai	13.014042	22.088442
1	Coimbatore	4.541096	25.325342
2	Cuddalore	8.965986	19.710884
3	Madurai	13.319728	25.768707
4	Mettur	8.429268	23.185366

```
[23] agg_df = data.groupby(['Sampling_Date']).agg({'SO2': 'mean', 'NO2': 'mean'}).reset_index()
     agg_df.head()
```

	Sampling_Date	SO2	NO2
0	2014-01-02	12.272727	19.272727
1	2014-01-03	11.461538	24.153846
2	2014-01-04	9.000000	21.250000
3	2014-01-06	12.727273	22.000000
4	2014-01-07	12.230769	26.615385

FILTERING UNWANTED DATA:

- Use double square brackets, like `[['NO2', 'SO2', 'City', 'Sampling_Date']]`.
- This creates a new DataFrame with only the selected columns.
- Useful for focusing on specific columns in analysis or data processing tasks.



```
[32] agg_df.columns

     Index(['City/Town/Village/Area', 'Sampling_Date', 'SO2', 'NO2'], dtype='object')

[33] agg_df.head()

        City/Town/Village/Area  Sampling_Date        SO2        NO2
     0                 Chennai     2014-01-02  14.000000  18.250000
     1                 Chennai     2014-01-03  12.800000  26.000000
     2                 Chennai     2014-01-06  15.333333  19.333333
     3                 Chennai     2014-01-07  14.666667  32.500000
     4                 Chennai     2014-01-08  11.000000  16.000000
```

ANALYSING AND VISUALISATION:

Creating a heatmap for air quality data can provide a clear visualization of the variations in SO2 and NO2 content across different cities and sampling dates. Since as mentioned that the city column has repeated city names, we may want to aggregate the data before creating the heatmap. Here's a step-by-step guide:

Aggregate Data:

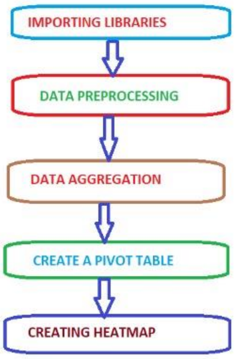
Calculate the average SO2 and NO2 content for each city on each sampling date. This will help in reducing repeated city names and provide a representative value for each city-date combination.

Pivot Table:

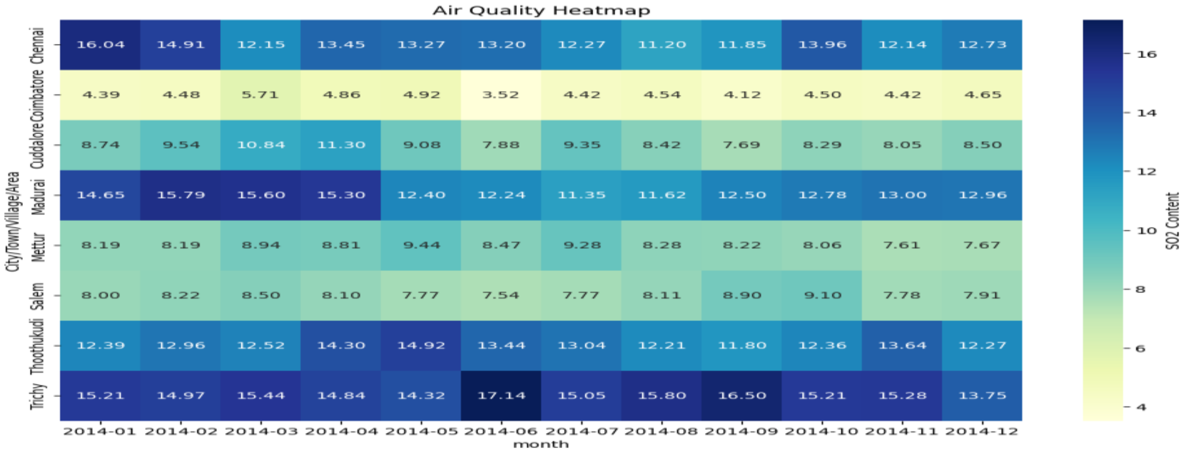
Create a pivot table where cities are along the rows, sampling dates along the columns, and the values are either average SO2 or NO2 content.

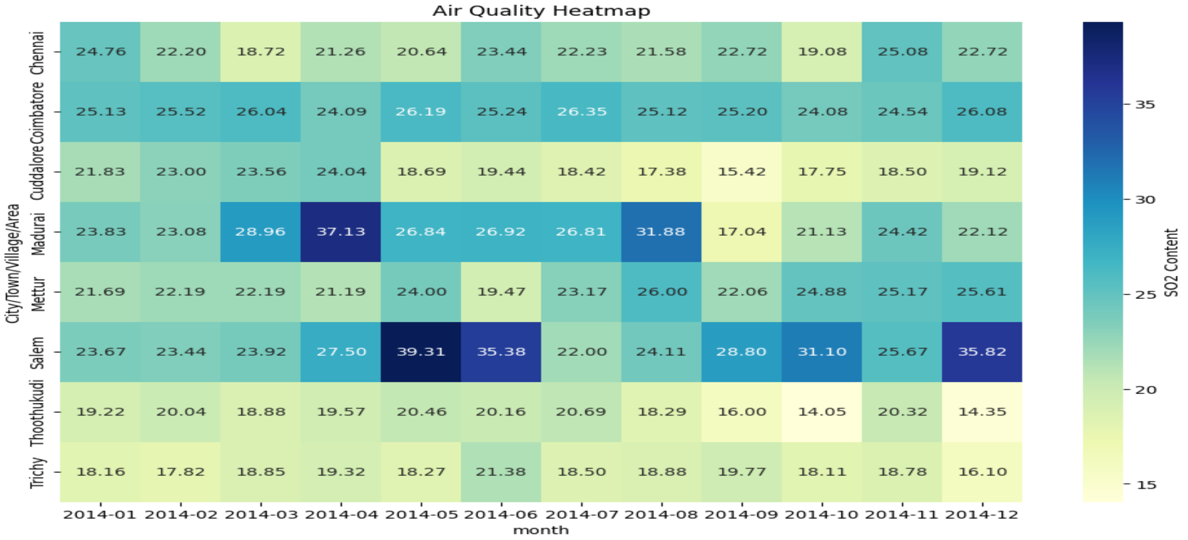
Heatmap:

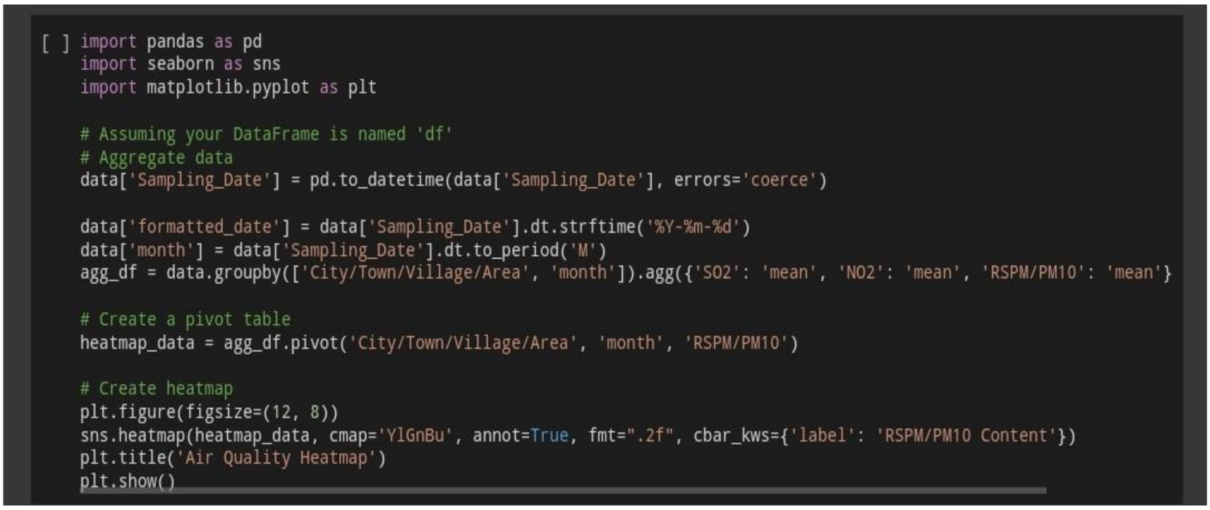
Use a heatmap to visualize the data. You can use tools like Python with libraries such as Matplotlib and Seaborn or other data visualization tools like Tableau or Excel.



The code preprocesses air quality data by converting dates and creating a formatted date column. It then aggregates the data by calculating mean 'SO2' and 'NO2' values for different locations and dates. Finally, it visualizes the 'SO2' and 'NO2' levels across locations and dates using a heatmap with color-coded values and a color bar, making air quality trends easily interpretable.



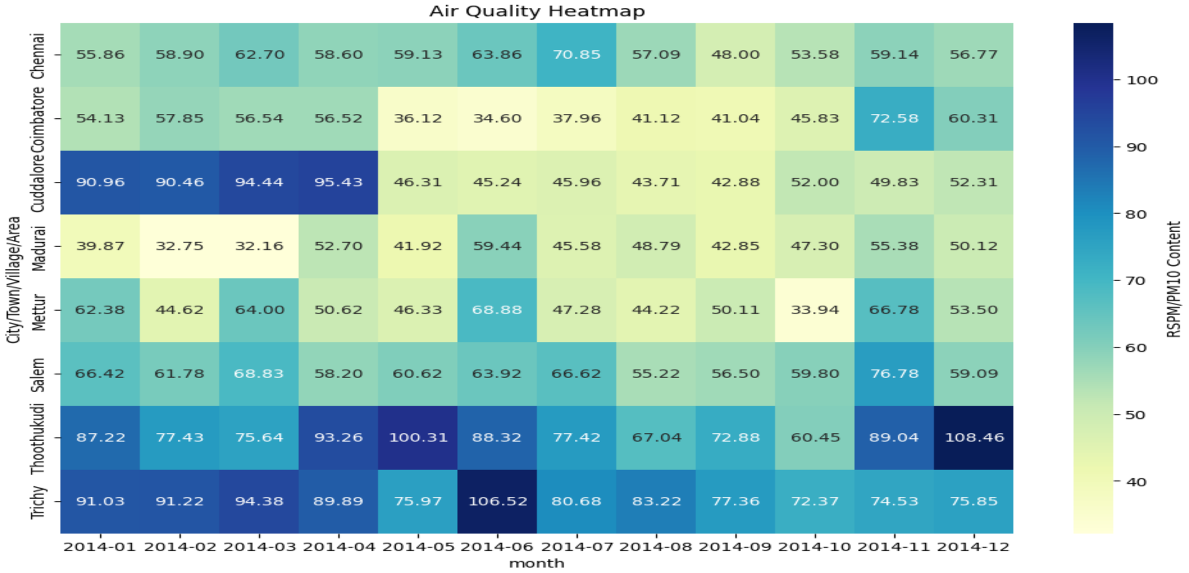




```
[ ] import pandas as pd
    import seaborn as sns
    import matplotlib.pyplot as plt

    # Assuming your DataFrame is named 'df'
    # Aggregate data
    data['Sampling_Date'] = pd.to_datetime(data['Sampling_Date'], errors='coerce')

    data['formatted_date'] = data['Sampling_Date'].dt.strftime('%Y-%m-%d')
    data['month'] = data['Sampling_Date'].dt.to_period('M')
    agg_df = data.groupby(['City/Town/Village/Area', 'month']).agg({'SO2': 'mean', 'NO2': 'mean', 'RSPM/PM10': 'mean'}

    # Create a pivot table
    heatmap_data = agg_df.pivot('City/Town/Village/Area', 'month', 'RSPM/PM10')

    # Create heatmap
    plt.figure(figsize=(12, 8))
    sns.heatmap(heatmap_data, cmap='YlGnBu', annot=True, fmt=".2f", cbar_kws={'label': 'RSPM/PM10 Content'})
    plt.title('Air Quality Heatmap')
    plt.show()
```

This code snippet performs the following actions:

- Imports necessary libraries: `pandas`, `seaborn`, and `matplotlib.pyplot`.
- Converts a date column in a DataFrame ('data') to a datetime format.
- Extracts the month and creates a formatted date column.
- Aggregates data by grouping it based on 'City/Town/Village/Area' and 'month', calculating the mean values of 'SO2', 'NO2', and 'RSPM/PM10'.
- Creates a pivot table from the aggregated data with 'City/Town/Village/Area' as rows and 'month' as columns, with 'RSPM/PM10' values in the cells.
- Plots a heatmap using Seaborn to visualize the 'RSPM/PM10' content in the DataFrame. It includes annotations with two decimal places and a colorbar with a label


Air Quality Heatmap
City/Town/Village/Area
Chennai 55.86 58.90 62.70 58.60 59.13 63.86 70.85 57.09 48.00 53.58 59.14 56.77
Coimbatore 54.13 57.85 56.54 56.52 36.12 34.60 37.96 41.12 41.04 45.83 72.58 60.31
Cuddalore 90.96 90.46 94.44 95.43 46.31 45.24 45.96 43.71 42.88 52.00 49.83 52.31
Madurai 39.87 32.75 32.16 52.70 41.92 59.44 45.58 48.79 42.85 47.30 55.38 50.12
Mettur 62.38 44.62 64.00 50.62 46.33 68.88 47.28 44.22 50.11 33.94 66.78 53.50
Salem 66.42 61.78 68.83 58.20 60.62 63.92 66.62 55.22 56.50 59.80 76.78 59.09
Thoothukudi 87.22 77.43 75.64 93.26 100.31 88.32 77.42 67.04 72.88 60.45 89.04 108.46
Trichy 91.03 91.22 94.38 89.89 75.97 106.52 80.68 83.22 77.36 72.37 74.53 75.85
month
RSPM/PM10 Content