College code: 4212

Register numb: 421221243008

# AIR QUALITY ANALYSIS IN TAMIL NADU

**PHASE3-DATA ANALYTICS WITH COGNOS: GROUP2**

This phase involves in designing of the steps that defining in each phase of the previous documentation this involves importing necessary functions, data processing and so on in this phase we have to begin our project by loading and pre-processing the dataset.

The IBM suggests using the Jupiter notebook for loading and pre-process the dataset: Here for this project title we need to define the loading the libraries, and Perform the air quality analysis and create visualizations.

PROBLEM: Calculate average SO2,NO2, and RSPM/PM10 levels across different monitoring stations, cities or area . Identify pollution trends and areas with high pollution levels.

2. Create Visualization using data Visualization Libraries (eg; Matplotlib seaborn).

For this certain inputs are defined for this project.in this phase each of the input lines of the project is given as follows:

Phase4

October 31, 2023

[86]:

**import**

**numpy**

**as**

**np**

**import**

**pandas**

**as**

**pd**

**import**

**seaborn**

**as**

**sns**

**import**

**matplotlib**

**.**

**pyplot**

**as**

**plt**

[87]:

df

=

pd

.

read\_csv(

'

cpcb\_dly\_aq\_tamil\_nadu-2014.csv

'

)

[88]:

df

.

head()

[88]:

Stn Code Sampling Date

State City/Town/Village/Area \

1. 38 01-02-14 Tamil Nadu Chennai
2. 38 01-07-14 Tamil Nadu Chennai
3. 38 21-01-14 Tamil Nadu Chennai
4. 38 23-01-14 Tamil Nadu Chennai
5. 38 28-01-14 Tamil Nadu Chennai

Location of Monitoring Station \

1. Kathivakkam, Municipal Kalyana Mandapam, Chennai
2. Kathivakkam, Municipal Kalyana Mandapam, Chennai
3. Kathivakkam, Municipal Kalyana Mandapam, Chennai
4. Kathivakkam, Municipal Kalyana Mandapam, Chennai
5. Kathivakkam, Municipal Kalyana Mandapam, Chennai

Agency Type of Location SO2 NO2 \

1. Tamilnadu State Pollution Control Board Industrial Area 11.0 17.0
2. Tamilnadu State Pollution Control Board Industrial Area 13.0 17.0
3. Tamilnadu State Pollution Control Board Industrial Area 12.0 18.0
4. Tamilnadu State Pollution Control Board Industrial Area 15.0 16.0
5. Tamilnadu State Pollution Control Board Industrial Area 13.0 14.0

RSPM/PM10 PM 2.5

1. 55.0 NaN
2. 45.0 NaN
3. 50.0 NaN
4. 46.0 NaN
5. 42.0 NaN

[89]:

df

.

shape

[89]: (2879, 11)

[90]:

df

.

info

[90]: <bound method DataFrame.info of 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 |
| … | … | … … | … |  |
| 2874 | 773 | 12-03-14 Tamil Nadu |  | Trichy |
| 2875 | 773 | 12-10-14 Tamil Nadu |  | Trichy |
| 2876 | 773 | 17-12-14 Tamil Nadu |  | Trichy |
| 2877 | 773 | 24-12-14 Tamil Nadu |  | Trichy |
| 2878 | 773 | 31-12-14 Tamil Nadu |  | Trichy |

Location of Monitoring Station \

* 1. Kathivakkam, Municipal Kalyana Mandapam, Chennai
  2. Kathivakkam, Municipal Kalyana Mandapam, Chennai
  3. Kathivakkam, Municipal Kalyana Mandapam, Chennai
  4. Kathivakkam, Municipal Kalyana Mandapam, Chennai
  5. Kathivakkam, Municipal Kalyana Mandapam, Chennai

… …

* 1. Central Bus Stand, Trichy
  2. Central Bus Stand, Trichy
  3. Central Bus Stand, Trichy
  4. Central Bus Stand, Trichy
  5. Central Bus Stand, Trichy

Agency \

* 1. Tamilnadu State Pollution Control Board
  2. Tamilnadu State Pollution Control Board
  3. Tamilnadu State Pollution Control Board
  4. Tamilnadu State Pollution Control Board
  5. Tamilnadu State Pollution Control Board

… …

* 1. Tamilnadu State Pollution Control Board
  2. Tamilnadu State Pollution Control Board
  3. Tamilnadu State Pollution Control Board
  4. Tamilnadu State Pollution Control Board
  5. Tamilnadu State Pollution Control Board

Type of Location SO2 NO2 RSPM/PM10 PM 2.5

|  |  |  |
| --- | --- | --- |
| 0 Industrial Area 11.0 17.0 | 55.0 | NaN |
| 1 Industrial Area 13.0 17.0 | 45.0 | NaN |
| 2 Industrial Area 12.0 18.0 | 50.0 | NaN |
| 3 Industrial Area 15.0 16.0 | 46.0 | NaN |
| 4 Industrial Area 13.0 14.0 | 42.0 | NaN |
| … … … … … | … |  |
| 2874 Residential, Rural and other Areas 15.0 18.0 | 102.0 | NaN |
| 2875 Residential, Rural and other Areas 12.0 14.0 | 91.0 | NaN |
| 2876 Residential, Rural and other Areas 19.0 22.0 | 100.0 | NaN |
| 2877 Residential, Rural and other Areas 15.0 17.0 | 95.0 | NaN |
| 2878 Residential, Rural and other Areas 14.0 16.0 | 94.0 | NaN |

* 1. rows x 11 columns]>

[91]:

df

.

columns

[91]: Index(['Stn Code', 'Sampling Date', 'State', 'City/Town/Village/Area',

'Location of Monitoring Station', 'Agency', 'Type of Location', 'SO2',

'NO2', 'RSPM/PM10', 'PM 2.5'], dtype='object')

[92]:

df

.

describe()

[92]: 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 |

[93]:

column\_name

=

'

NO2

'

[94]:

print

(

df[column\_name

]

.

mean())

22.136775994417306

[95]:

column\_name

=

'

SO2

'

[96]:

print

(

df[column\_name

]

.

mean())

11.503138075313808

[97]:

x

=

df[

'

City/Town/Village/Area

'

]

y

=

df[

'

SO2

'

]

plt

.

figure(figsize

=

(

10

,

6

))

plt

.

plot(x, y, marker

=

'

o

'

, linestyle

=

'

-

'

, color

=

'

b

'

, label

=

'

Pollution Trends

'

)

plt

.

title(

'

Pollution Trends

'

)

plt

.

xlabel(

'

City/Town/Village/Area

'

)

plt

.

ylabel(

'

SO2

'

)

plt

.

legend()

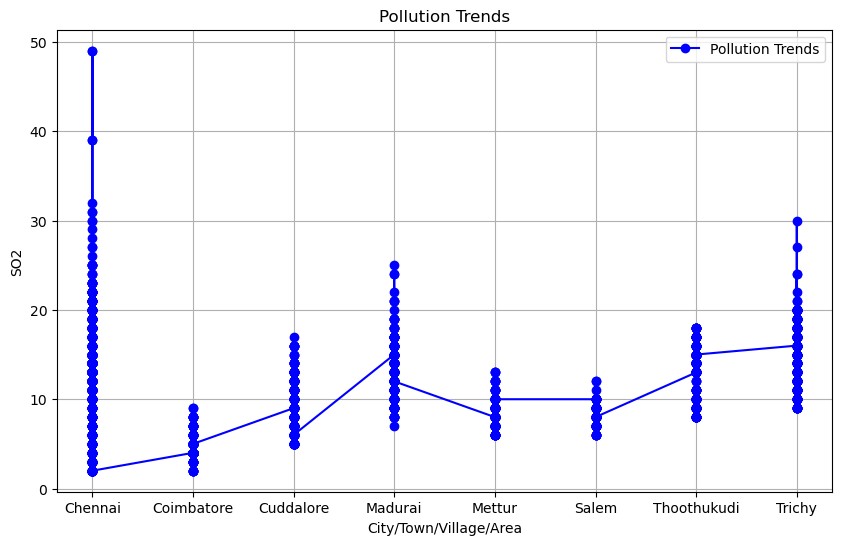
plt

.

grid(

**True**

)



[98]:

plt

.

figure(figsize

=

(

12

,

6

))

plt

.

xticks(rotation

=

75

)

plt

.

title(

'

High pollution levels

'

)

plt

.

xlabel(

'

City/Town/Village/Area

'

)

plt

.

ylabel(

'

Location of Monitoring Station

'

)

sns

.

barplot(df)

C:\ProgramData\anaconda3\lib\site-packages\seaborn\algorithms.py:98:

RuntimeWarning: Mean of empty slice boot\_dist.append(f(\*sample, \*\*func\_kwargs))

C:\ProgramData\anaconda3\lib\site-packages\numpy\lib\nanfunctions.py:1559:

RuntimeWarning: All-NaN slice encountered r, k = function\_base.\_ureduce(a,

[98]: <Axes: title={'center': 'High pollution levels'}, xlabel='City/Town/Village/Area\t', ylabel='Location of Monitoring Station\t'>

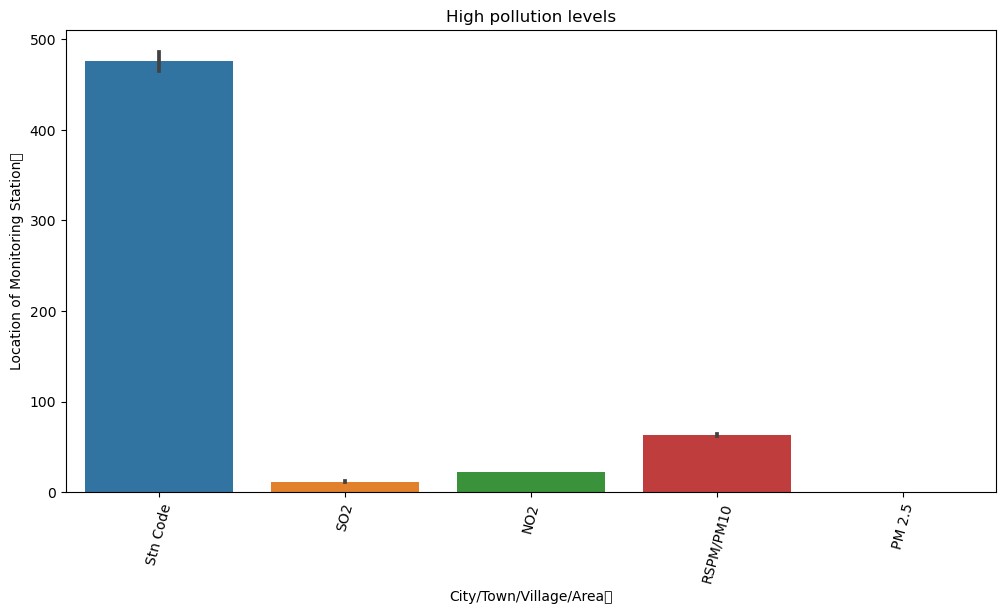
C:\ProgramData\anaconda3\lib\site-packages\IPython\core\events.py:89:

UserWarning: Glyph 9 ( ) missing from current font.

func(\*args, \*\*kwargs)

C:\ProgramData\anaconda3\lib\site-packages\IPython\core\pylabtools.py:152:

UserWarning: Glyph 9 ( ) missing from current font. fig.canvas.print\_figure(bytes\_io, \*\*kw)



[120]:

df

=

pd

.

DataFrame(df)

plt

.

figure(figsize

=

(

8

,

6

))

plt

.

bar(df[

'

Sampling Date

'

]

, df

[

'

SO2

'

]

, color

=

'

blue

'

)

plt

.

title(

'

SO2 Levels Over Time

'

)

plt

.

xlabel(

'

Sampling Date

'

)

plt

.

ylabel(

'

SO2 Level

'

)

plt

.

xticks(rotation

=

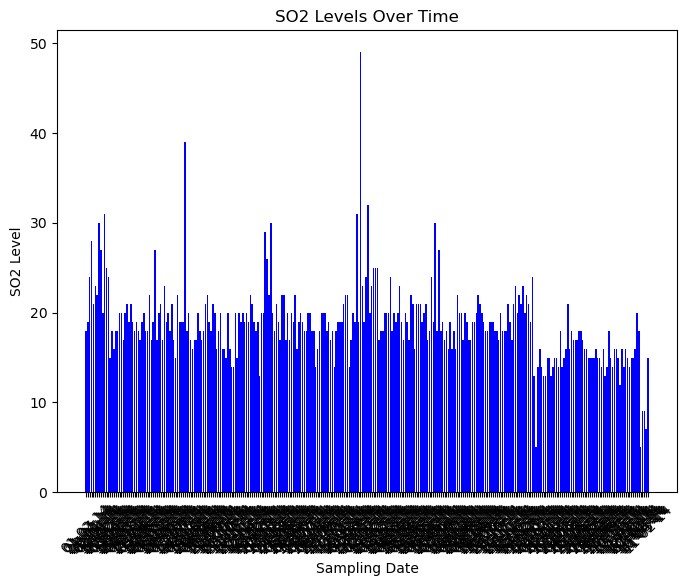
45

)

plt

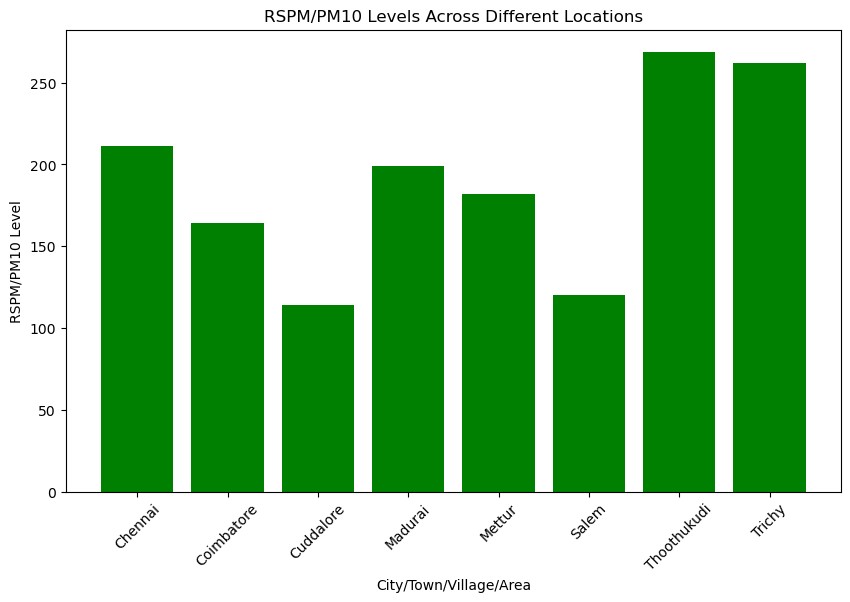
.

show()



[121]: df = pd.DataFrame(df) plt.figure(figsize=(10, 6)) plt.bar(df['City/Town/Village/Area'], df['RSPM/PM10'], color='green') plt.title('RSPM/PM10 Levels Across Different Locations')

plt.xlabel('City/Town/Village/Area') plt.ylabel('RSPM/PM10 Level') plt.xticks(rotation=45) plt.show()



[ ]: