College code: 4212

Register numb: 421221243034

# 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

da

taset: Here for this project title we need to define the loading the libraries,

understand the data and visualize the missing values.

For this certain inputs are defined for this project.in this phase each of the

input lines of the project is given as foll

ows:

## p3

October 18, 2023

[1]:

**import**

**pandas**

**as**

**pd**

**import**

**numpy**

**as**

**np**

[2]:

df

=

pd

.

read\_csv(

'

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

'

)

[3]:

df

.

head()

[3]:

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

[4]:

df

.

shape

[4]: (2879, 11)

[5]:

df

.

info

[5]: <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

* 1. Industrial Area 11.0 17.0 55.0 NaN
  2. 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]>

[6]:

df

.

columns

.

values

[6]: array(['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)

[7]:

df

.

dtypes

[7]: Stn Code int64

Sampling Date object

State object

City/Town/Village/Area object

Location of Monitoring Station object

Agency object

Type of Location object

SO2 float64

NO2 float64

RSPM/PM10 float64

PM 2.5 float64

dtype: object

[8]:

df

=

df

.

drop([

'

Sampling Date

'

]

, axis

=

1

)

df

.

head()

[8]: Stn Code State City/Town/Village/Area \

1. 38 Tamil Nadu Chennai
2. 38 Tamil Nadu Chennai
3. 38 Tamil Nadu Chennai
4. 38 Tamil Nadu Chennai
5. 38 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

[9]:

df[df[

'

SO2

'

]

==

0

]

.

index

[9]: Int64Index([], dtype='int64')

[10]: df.drop(labels=df[df['SO2'] == 0].index, axis=0, inplace=**True**) df[df['SO2'] == 0].index

[10]: Int64Index([], dtype='int64')

[11]:

df

.

fillna(df[

"

NO2

"

]

.

mean())

|  |  |  |
| --- | --- | --- |
| [11]: | Stn Code State City/Town/Village/Area \ | |
| 0 | 38 Tamil Nadu | Chennai |
| 1 | 38 Tamil Nadu | Chennai |
| 2 | 38 Tamil Nadu | Chennai |
| 3 | 38 Tamil Nadu | Chennai |
| 4 | 38 Tamil Nadu | Chennai |
| … | … … | … |
| 2874 | 773 Tamil Nadu | Trichy |
| 2875 | 773 Tamil Nadu | Trichy |
| 2876 | 773 Tamil Nadu | Trichy |
| 2877 | 773 Tamil Nadu | Trichy |
| 2878 | 773 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 22.136776 |
| 1 Industrial Area 13.0 17.0 | 45.0 22.136776 |
| 2 Industrial Area 12.0 18.0 | 50.0 22.136776 |
| 3 Industrial Area 15.0 16.0 | 46.0 22.136776 |
| 4 Industrial Area 13.0 14.0 | 42.0 22.136776 |
| … … … … … | … |
| 2874 Residential, Rural and other Areas 15.0 18.0 | 102.0 22.136776 |
| 2875 Residential, Rural and other Areas 12.0 14.0 | 91.0 22.136776 |
| 2876 Residential, Rural and other Areas 19.0 22.0 | 100.0 22.136776 |
| 2877 Residential, Rural and other Areas 15.0 17.0 | 95.0 22.136776 |
| 2878 Residential, Rural and other Areas 14.0 16.0 | 94.0 22.136776 |

[2879 rows x 10 columns]

[12]:

df

.

isnull()

.

sum()

[12]: Stn Code 0

State 0

City/Town/Village/Area 0

Location of Monitoring Station 0

Agency 0

Type of Location 0

SO2 11

NO2 13

RSPM/PM10 4

PM 2.5 2879

dtype: int64

[13]:

numerical\_cols

=

[

'

SO2

'

,

'

RSPM/PM10

'

,

'

NO2

'

]

df[numerical\_cols]

.

describe()

[13]: SO2 RSPM/PM10 NO2 count 2868.000000 2875.000000 2866.000000

|  |  |  |  |
| --- | --- | --- | --- |
| mean | 11.503138 | 62.494261 | 22.136776 |
| std | 5.051702 | 31.368745 | 7.128694 |
| min | 2.000000 | 12.000000 | 5.000000 |
| 25% | 8.000000 | 41.000000 | 17.000000 |
| 50% | 12.000000 | 55.000000 | 22.000000 |
| 75% | 15.000000 | 78.000000 | 25.000000 |
| max | 49.000000 | 269.000000 | 71.000000 |

[14]:

df

.

dtypes

[14]: Stn Code int64

State object

City/Town/Village/Area object

Location of Monitoring Station object

Agency object

Type of Location object

SO2 float64

NO2 float64

RSPM/PM10 float64

PM 2.5 float64

dtype: object