

AIR QUALITY ASSESSMENT-TAMILNADU

TEAM MEMBER

962221106075 : Mohammad Roshan

Phase-3 (DEVELOPMENT)

INTRODUCTION:

In this phase we will begin our project by loading and preprocessing the air quality dataset. Dataset will be loaded in python and the preprocessing will be done as follows (data manipulation libraries will be installed).

LIBRARIES USED FOR PREPROCESSING:

```
In [1]: #import required libraries  
  
In [2]: import numpy as np  
  
In [3]: import seaborn as sns  
  
In [4]: import pandas as pd
```

NumPy: NumPy stands for Numerical Python. It is a fundamental package for numerical computations in Python. It provides support for arrays and matrices, as well as a large collection of high-level mathematical functions to operate on these data structures. NumPy is widely used in scientific and engineering applications for tasks involving numerical operations.

Pandas: Pandas is a data manipulation and analysis library. It provides data structures like Series (1-dimensional) and DataFrame (2-dimensional), which are highly efficient and designed for working with structured data. Pandas allows for easy data ingestion, cleaning, transformation, and analysis.

Seaborn: Seaborn is a statistical data visualization library based on Matplotlib. It provides a high-level interface for creating informative and attractive statistical graphics. Seaborn simplifies the process of creating complex visualizations and is especially useful for exploring relationships between variables in datasets.

DATASET COLLECTED: Recent data(5th august 2023)

FILE

HOME

INSERT

PAGE LAYOUT

FORMULAS

DATA

REVIEW

VIEW

PROTECTED VIEW

Be careful—files from the Internet can contain viruses. Unless you need to edit, it's safer to stay in Protected View.

Enable Editing

I13

X

✓

Satisfactory

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
|----|------|---------------------|----------|-----|-----|-------|------|------------|------------|-----------|------|---|---|---|---|
| 1 | S.No | District (Location) | S02 | N02 | co | PM2.5 | PMIO | AQI Index | AQI | Prominent | | | | | |
| 2 | | | | | | | | | Value | Pollutant | | | | | |
| 3 | 1 | Ariyalur | 11 | 14 | 0.4 | 16 | 37 | Good | 37 | PMIO | | | | | |
| 4 | 2 | Chengalpattu (Vand | 13 | 18 | 0.8 | 20 | 96 | Satisfacto | 96 | PMIO | | | | | |
| 5 | 3 | Kodungaiy | 3 | 16 | 0.7 | 17 | 68 | Satisfacto | 76 | pM10 | | | | | |
| 6 | 4 | Koyambee | 4 | 13 | 0.4 | 28 | 67 | Satisfacto | 62 | pMIO | | | | | |
| 7 | 5 | Chennai | perungud | 3 | 23 | 0.5 | 17 | 89 | Satisfacto | 89 | PMIO | | | | |
| 8 | 6 | Royapuram | 3 | 24 | 0.6 | 19 | 72 | Satisfacto | 72 | PMIO | | | | | |
| 9 | 7 | Kuruchi-S | 6 | 12 | 0.3 | 21 | 38 | Good | 38 | PMIO | | | | | |
| 10 | | Coimbatore | | | | | | | | | | | | | |
| 11 | 8 | PSG Collag | 4 | 9 | 0.2 | 10 | 33 | Good | 33 | PMI O | | | | | |
| 12 | 9 | Semmend | 6 | 12 | 0.3 | 20 | 29 | Good | 29 | PMIO | | | | | |
| 13 | 10 | Cuddalore SIPCOT | 17 | 13 | 0.5 | 34 | 43 | Satisfacto | 43 | PMIO | | | | | |
| 14 | 11 | Dindigul | | | | | ND | | | | | | | | |
| 15 | 12 | Hosur | 6 | 3 | o.i | 24 | 45 | | 45 | PMIO | | | | | |
| 16 | 13 | Kanchipuram | 1 | 2 | o.i | 24 | 49 | Good | 49 | pMIO | | | | | |
| 17 | 14 | Karur | 16 | 19 | 0.6 | 29 | 43 | Good | 43 | ptv410 | | | | | |
| 18 | 15 | Madurai | 2 | 4 | 0.5 | 20 | 41 | Good | 41 | PMIO | | | | | |
| 19 | 16 | Nagapattinam | 17 | 19 | 0.5 | 24 | 15 | | 24 | PM2.5 | | | | | |
| 20 | 17 | Namakkal | | | | | ND | | | | | | | | |
| 21 | 18 | Ooty | 13 | 16 | 0.3 | 12 | 30 | Good | 30 | PMIO | | | | | |
| 22 | 19 | Perundurai | 9 | 14 | 0.5 | 23 | 38 | Good | 38 | PMIO | | | | | |
| 23 | 20 | Pudukkottai | 24 | 26 | 0.9 | 21 | 49 | Good | 49 | pMIO | | | | | |
| 24 | 21 | Ramanathapuram | 7 | 3 | 0.4 | 11 | 51 | Satisfacto | 51 | PMIO | | | | | |
| 25 | 22 | Ranipet, SIPCOT | 20 | 21 | 0.3 | 11 | 35 | Good | 35 | PMIO | | | | | |
| 26 | 23 | Salem | 12 | 16 | 0.8 | 22 | 39 | Good | 39 | PMIO | | | | | |

Loading of data in python and preprocessing:

```
In [1]: #import required libraries
```

```
In [2]: import pandas as pd
```

```
In [3]: import numpy as np
```

```
In [4]: import seaborn as sns
```

```
In [11]: data = pd.read_excel("C:\\Users\\pc\\Pictures\\New folder\\airquality.xlsx")
```

```
In [12]: data.head()
```

```
Out[12]:
```

| | S.No | District (Location) | Unnamed: 2 | S02 | N02 | co | PM2.5 | PMIO | AQI Index | AQI | Prominent |
|---|------|-------------------------|-------------|------|-----|-----|-------|------|--------------|-------|-----------|
| 0 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | Value | Pollutant |
| 1 | 1.0 | Ariyalur | NaN | 11.0 | 14 | 0.4 | 16 | 37 | Good | 37 | PMIO |
| 2 | 2.0 | Chengalpattu (Vandalur) | NaN | 13.0 | 18 | 0.8 | 20 | 96 | Satisfactory | 96 | PMIO |
| 3 | 3.0 | NaN | Kodungaiyur | 3.0 | 16 | 0.7 | 17 | 68 | Satisfactory | 76 | pM10 |

```
In [13]: pd.read_excel("C:\\Users\\pc\\Pictures\\New folder\\airquality.xlsx")
```

Out[13]:

| | S.No | District (Location) | Unnamed: 2 | S02 | N02 | co | PM2.5 | PMIO | AQI Index | AQI | Prominent |
|----|------|-------------------------|---------------|------|-----|-----|-------|------|--------------|-------|-----------|
| 0 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | Value | Pollutant |
| 1 | 1.0 | Ariyalur | NaN | 11.0 | 14 | 0.4 | 16 | 37 | Good | 37 | PMIO |
| 2 | 2.0 | Chengalpattu (Vandalur) | NaN | 13.0 | 18 | 0.8 | 20 | 96 | Satisfactory | 96 | PMIO |
| 3 | 3.0 | NaN | Kodungaiyur | 3.0 | 16 | 0.7 | 17 | 68 | Satisfactory | 76 | pM10 |
| 4 | 4.0 | NaN | Koyambedu | 4.0 | 13 | 0.4 | 28 | 67 | Satisfactory | 62 | pMIO |
| 5 | 5.0 | Chennai | perungudi | 3.0 | 23 | 0.5 | 17 | 89 | Satisfactory | 89 | PMIO |
| 6 | 6.0 | NaN | Royapuram | 3.0 | 24 | 0.6 | 19 | 72 | Satisfactory | 72 | PMIO |
| 7 | 7.0 | NaN | Kuruchi-SIDCO | 6.0 | 12 | 0.3 | 21 | 38 | Good | 38 | PMIO |
| 8 | NaN | Coimbatore | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 9 | 8.0 | NaN | PSG Collage | 4.0 | 9 | 0.2 | 10 | 33 | Good | 33 | PMI O |
| 10 | 9.0 | NaN | Semmendalam | 6.0 | 12 | 0.3 | 20 | 29 | Good | 29 | PMIO |
| 11 | 10.0 | Cuddalore | SIPCOT | 17.0 | 13 | 0.5 | 34 | 43 | Satisfactory | 43 | PMIO |
| 12 | 11.0 | Dindigul | NaN | NaN | NaN | NaN | NaN | ND | NaN | NaN | NaN |

Basic info from preprocessing our data:

```
In [14]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 36 entries, 0 to 35
Data columns (total 11 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   S.No                  34 non-null    float64
1   District (Location)  26 non-null    object
2   Unnamed: 2           13 non-null    object
3   S02                   31 non-null    float64
4   N02                   30 non-null    object
5   co                    31 non-null    object
6   PM2.5                31 non-null    object
7   PMIO                 34 non-null    object
8   AQI Index            28 non-null    object
9   AQI                  32 non-null    object
10  Prominent             32 non-null    object
dtypes: float64(2), object(9)
memory usage: 3.2+ KB
```

Conclusion:

In this phase we have loaded our collected data in the python and the preprocessing importations such as pandas,numpy and seaborn are done.let us make our analysis in the upcoming phases.