Applied DataScience :

Project Name: Air Quality Analysis and Prediction in Tamilnadu

Project Description: To Develop a Machine learning algorithms like random forest and Using Pandas and Numpy Libraries to predict ,calculate and using matplotlib ,seaborn to visualize the air quality in Tamilnadu.

Phase 4 : Development Part 2

Description :

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

* Create visualizations

Create visualizations using data visualization libraries (e.g., Matplotlib, Seaborn).

Load the Air Quality dataset and preprocess the data for analysis.

Dataset Link: <https://tn.data.gov.in/resource/location-wise-daily-ambient-air-quality-tamil-nadu-year-2014>

Working Procedure :

To load and preprocess the Air Quality dataset in 2014 from Kaggle, we can use the following steps:

Step 1:

Install the necessary Python libraries

Step 2 :

Load the dataset

# Load the dataset from the Device

Step 3 :

Explore the dataset

# Print the first 5 rows of the dataset

# Print the basic information about the dataset

Step 4 :

Preprocess the data

Handle missing values: There are no missing values in the dataset.

Convert categorical features to numerical features:

# Define a function to convert categorical features to numerical features

# Encode the Genre feature

# Encode the Language feature

Step 5 :

Scale the numerical features

# Define a function to scale numerical features

Step 6 :

Split the dataset into training and test sets.

Step 7 :

Using matplotlib to visualize the air quality with Histogram,lineplot,scatterplot and more than many plots.

Conclusion:

We have now loaded and preprocessed the Air Quality Analysis dataset for analysis. Data Visualisation libraries to visualize the SO2,NO2 and RSPM/PM Levels.

Program for an above steps :

In[1] : import pandas as pd

# Load the dataset from the Kaggle website

In [2] : air\_quality= pd.read\_csv(<https://tn.data.gov.in/resource/location-wise-daily-ambient-air-quality-tamil-nadu-year-2014>.csv)

Out[2] :

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Stn Code | Sampling Date | State | City/Town/Village/Area | Location of Monitoring Station | Agency | Type of Location | SO2 | NO2 | RSPM/PM10 | PM 2.5 |
| 38 | 1/2/2014 | Tamil Nadu | Chennai | Kathivakkam, Municipal Kalyana Mandapam, Chennai | Tamilnadu State Pollution Control Board | Industrial Area | 11 | 17 | 55 | NA |
| 38 | 1/7/2014 | Tamil Nadu | Chennai | Kathivakkam, Municipal Kalyana Mandapam, Chennai | Tamilnadu State Pollution Control Board | Industrial Area | 13 | 17 | 45 | NA |
| 38 | 21-01-14 | Tamil Nadu | Chennai | Kathivakkam, Municipal Kalyana Mandapam, Chennai | Tamilnadu State Pollution Control Board | Industrial Area | 12 | 18 | 50 | NA |
| 38 | 23-01-14 | Tamil Nadu | Chennai | Kathivakkam, Municipal Kalyana Mandapam, Chennai | Tamilnadu State Pollution Control Board | Industrial Area | 15 | 16 | 46 | NA |
| 38 | 28-01-14 | Tamil Nadu | Chennai | Kathivakkam, Municipal Kalyana Mandapam, Chennai | Tamilnadu State Pollution Control Board | Industrial Area | 13 | 14 | 42 | NA |
| 38 | 30-01-14 | Tamil Nadu | Chennai | Kathivakkam, Municipal Kalyana Mandapam, Chennai | Tamilnadu State Pollution Control Board | Industrial Area | 14 | 18 | 43 | NA |

... ... ... ... ... ... ...

In [3] : air\_qualitu.head()

Out [3] :

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Stn Code | Sampling Date | State | City/Town/Village/Area | Location of Monitoring Station | Agency | Type of Location | SO2 | NO2 | RSPM/PM10 | PM 2.5 |
| 38 | 1/2/2014 | Tamil Nadu | Chennai | Kathivakkam, Municipal Kalyana Mandapam, Chennai | Tamilnadu State Pollution Control Board | Industrial Area | 11 | 17 | 55 | NA |
| 38 | 1/7/2014 | Tamil Nadu | Chennai | Kathivakkam, Municipal Kalyana Mandapam, Chennai | Tamilnadu State Pollution Control Board | Industrial Area | 13 | 17 | 45 | NA |
| 38 | 21-01-14 | Tamil Nadu | Chennai | Kathivakkam, Municipal Kalyana Mandapam, Chennai | Tamilnadu State Pollution Control Board | Industrial Area | 12 | 18 | 50 | NA |
| 38 | 23-01-14 | Tamil Nadu | Chennai | Kathivakkam, Municipal Kalyana Mandapam, Chennai | Tamilnadu State Pollution Control Board | Industrial Area | 15 | 16 | 46 | NA |
| 38 | 28-01-14 | Tamil Nadu | Chennai | Kathivakkam, Municipal Kalyana Mandapam, Chennai | Tamilnadu State Pollution Control Board | Industrial Area | 13 | 14 | 42 | NA |
| 38 | 30-01-14 | Tamil Nadu | Chennai | Kathivakkam, Municipal Kalyana Mandapam, Chennai | Tamilnadu State Pollution Control Board | Industrial Area | 14 | 18 | 43 | NA |

<class 'pandas.core.frame.DataFrame'

# Print the basic information about the dataset

In [4] : air\_quality.info()

Out [4] :

Data columns (total 6 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 stncode 105 non-null object

1 date 105 non-null int64

2 State 105 non-null object

3 V

illage 105 non-null object

4 SO2 105 non-null int64

5 NO2e 105 non-null float64

dtypes: float64(1), int64(2), object(3)

memory usage: 5.0+ KB

# Check for missing values

In [5] : netflix\_originals.isnull().sum()

Out [5] :

--- ------ -------------- -----

Import matplotlib.pyplot as plt

# Create a figure and axis

ax = plt.subplots()

# Plot PM2.5 data

Ax.plot(dates, pm25\_values, label=’PM2.5’, marker=’o’)

# Plot PM10 data

Ax.plot(dates, pm10\_values, label=’PM10’, marker=’x’)

# Set labels and title

Ax.set\_xlabel(‘Date’)

Ax.set\_ylabel(‘Air Quality Index’)

Ax.set\_title(‘Air Quality Over Time’)

Ax.legend()

# Rotate x-axis labels for better readability

Plt.xticks(rotation=45)

# Show the plot

Plt.tight\_layout()

Plt.show()

Histogram :

Plt.hist(air\_quality\_data, bins=10, edgecolor=’k’, alpha=0.7)

# Set labels and title

Plt.xlabel(‘PM2.5 Levels’)

Plt.ylabel(‘Frequency’)

Plt.title(‘Air Quality Histogram’)

# Show the plot

Plt.show()

Scatterplot :

Plt.scatter(pm25\_values, pm10\_values, c=’b’, marker=’o’, label=’Air Quality Data’)

# Set labels and title

Plt.xlabel(‘PM2.5 Levels’)

Plt.ylabel(‘PM10 Levels’)

Plt.title(‘Air Quality Scatterplot’)

# Show the plot

Plt.legend()

Plt.grid(True)

Plt.show()