PHASE No. : 3

### Project Title : Air Quality Analysis

PROJECT BUILDING

Introduction:-

In this part you will begin building your project by loading and preprocessing the dataset.

Begin the analysis by loading and preprocessing the air quality dataset.

Load the dataset using Python and data manipulation libraries

Collect Data:

First, you need an air quality dataset. You can find such datasets on websites like Kaggle, government environmental agencies, or through research institutions.

Import Libraries:

Import the necessary Python libraries, such as Pandas for data manipulation and Matplotlib or Seaborn for data visualization. You may also need NumPy for numerical operations.

Load the Dataset:

Load your air quality dataset into a Pandas DataFrame. You can use the pd.read\_csv() function if your data is in a CSV file.

CODE:

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

# URL of the dataset

url = 'https://tn.data.gov.in/resource/location-wise-daily-ambient-air-quality-tamil-nadu-year-2014'

# Load the dataset

df = pd.read\_csv("N:\cpcb\_dly\_aq\_meghalaya-2014.csv")

# Display the first few rows of the dataframe

print(df.head())

# Basic info about the dataset

print(df.info())

# Summary statistics of the dataset

print(df.describe())

# Check for missing values

print(df.isnull().sum())

# Plotting a histogram for a column (replace 'column\_name' with your column of interest)

for column in df.columns:

plt.figure(figsize=(10,5))

sns.histplot(data=df, x=column, kde=True)

plt.title(f'Distribution of {column}')

plt.show()

OUTPUT

Stn Code Sampling Date State City/Town/Village/Area \

0 588 01-07-14 Meghalaya Dawki

1 588 01-08-14 Meghalaya Dawki

2 588 14-01-14 Meghalaya Dawki

3 588 15-01-14 Meghalaya Dawki

4 588 21-01-14 Meghalaya Dawki

Location of Monitoring Station \

0 Terrace building, Dawki, Jaintia Hills Distric...

1 Terrace building, Dawki, Jaintia Hills Distric...

2 Terrace building, Dawki, Jaintia Hills Distric...

3 Terrace building, Dawki, Jaintia Hills Distric...

4 Terrace building, Dawki, Jaintia Hills Distric...

Agency \

0 Meghalaya State Pollution Control Board

1 Meghalaya State Pollution Control Board

2 Meghalaya State Pollution Control Board

3 Meghalaya State Pollution Control Board

4 Meghalaya State Pollution Control Board

Type of Location SO2 NO2 RSPM/PM10 PM 2.5

0 Residential, Rural and other Areas 2 11 52.0 NaN

1 Residential, Rural and other Areas 2 12 48.0 NaN

2 Residential, Rural and other Areas 2 12 52.0 NaN

3 Residential, Rural and other Areas 2 11 50.0 NaN

4 Residential, Rural and other Areas 2 8 47.0 NaN

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 609 entries, 0 to 608

Data columns (total 11 columns):

# Column Non-Null Count Dtype

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

0 Stn Code 609 non-null int64

1 Sampling Date 609 non-null object

2 State 609 non-null object

3 City/Town/Village/Area 609 non-null object

4 Location of Monitoring Station 609 non-null object

5 Agency 609 non-null object

6 Type of Location 609 non-null object

7 SO2 609 non-null int64

8 NO2 609 non-null int64

9 RSPM/PM10 608 non-null float64

10 PM 2.5 0 non-null float64

dtypes: float64(2), int64(3), object(6)

memory usage: 52.5+ KB

None

Stn Code SO2 NO2 RSPM/PM10 PM 2.5

count 609.000000 609.000000 609.000000 608.000000 0.0

mean 505.371100 4.875205 10.261084 56.856908 NaN

std 206.545538 7.843925 5.849186 35.511526 NaN

min 120.000000 2.000000 5.000000 12.000000 NaN

25% 340.000000 2.000000 5.000000 35.000000 NaN

50% 588.000000 2.000000 10.000000 42.000000 NaN

75% 698.000000 2.000000 14.000000 83.000000 NaN

max 699.000000 42.000000 30.000000 181.000000 NaN

Stn Code 0

Sampling Date 0

State 0

City/Town/Village/Area 0

Location of Monitoring Station 0

Agency 0

Type of Location 0

SO2 0

NO2 0

RSPM/PM10 1

PM 2.5 609

dtype: int64







 



    