Phase 1:

Problem Definition:

The project aims to analyze and visualize air quality data from monitoring stations in Tamil Nadu. The objective is to gain insights into air pollution trends, identify areas with high pollution levels, and develop a predictive model to estimate RSPM/PM10 levels based on SO2 and NO2 levels. This project involves defining objectives, designing the analysis approach, selecting visualization techniques, and creating a predictive model using Python and relevant libraries.

Design Thinking:

Project Objectives: Define objectives such as analyzing air quality trends, identifying pollution hotspots, and building a predictive model for RSPM/PM10 levels.

Analysis Approach: Plan the steps to load, preprocess, analyze, and visualize the air quality data.

Visualization Selection: Determine visualization techniques (e.g., line charts, heatmaps) to effectively represent air quality trends and pollution levels.

.....

Phase 2:

Project Name: Air Quality Analysis and prediction in Tamil Nadu

Project Description: Develop a machine learning model predict a Air quality such as So2 ,No2 and RSPM .This prediction make a Air quality in Normal Flow .

Phase 2: Innovation

Description: Using a Random Forest algorithm to predict the air Quality and Find a accuracy for a prediction.

Air Quality Analysis and Prediction in Tamilnadu

1.Collect and prepare data:

Gather a dataset of Air quality with their So2,NO2 And RSPM rates. Preprocess the data by cleaning the text, removing stop words, and stemming or lemmatizing the words.

2. Choose a Random Forest Technique:

There are many random forest techniques used to predict the value of Air quality.

3. Train Random Forest model.

Feed the prepared data to the model and allow it to learn the relationships between the features and the target variable (Accuracy of SO2 and NO2).

4. Evaluate the model.

Evaluate the performance of the model on a held-out test set. This will give an idea of how well about the model will generalize to new data.

5. Deploy the model.

Once you are satisfied with the performance of the model, it can deploy it to production. This may involve saving the model to a file or deploying it to a cloud-based platform.

Here is an example of how to train a simple feedforward Machine Learning Algorithms such as Random Forest Technique

Random forest in Python:

```
Import pandas as pd
```

From sklearn.model_selection import train_test_split

From sklearn.ensemble import RandomForestRegressor

From sklearn.metrics import mean_squared_error

Import matplotlib.pyplot as plt

Load your dataset

Data = pd.read_csv('cpcb_dly_aq_tamil_nadu-2014.csv')

Split the dataset into features (X) and target (y)

X = data.drop('AQI', axis=1)

Y = data['AQI']

Split the data into training and testing sets

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)s

Create and train the Random Forest model

Rf_model = RandomForestRegressor(n_estimators=100, random_state=42)

Rf_model.fit(X_train, y_train)

Make predictions on the test set

Y_pred = rf_model.predict(X_test)
Calculate the Mean Squared Error (MSE) to evaluate the model
Mse = mean_squared_error(y_test, y_pred)
Print(f'Mean Squared Error: {mse}')
Visualize the feature importances
Feature_importances = rf_model.feature_importances_
Feature_names = X.columns
Plt.barh(feature_names, feature_importances)
Plt.xlabel('Feature Importance')
Plt.ylabel('Feature Name')
Plt.show()
`
Phase 3:
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 and calculate the air quality in Tamilnadu.
Phase 3: Development Part 1
Description :
Begin building the Air quality prediction model by loading and preprocessing the dataset.

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 Kaggle website
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.
Conclusion:
We have now loaded and preprocessed the Air Quality Analysis dataset for analysis. The next step is to build a machine learning model to predict Air quality
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-year2014)

Out[2]:

Stn Code Sampling Date State City/Town/Village/AreaLocation 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/AreaLocation 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'<="" td=""></class>
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 Village 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] :

0 stncode 0

1 date 0
2 State 0
3 Village 0
4 SO2 0
5 NO2. 0
This means that the training set contains SO2samples and the test set contains NO2 samples.
Phase 4:
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/AreaLocation 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

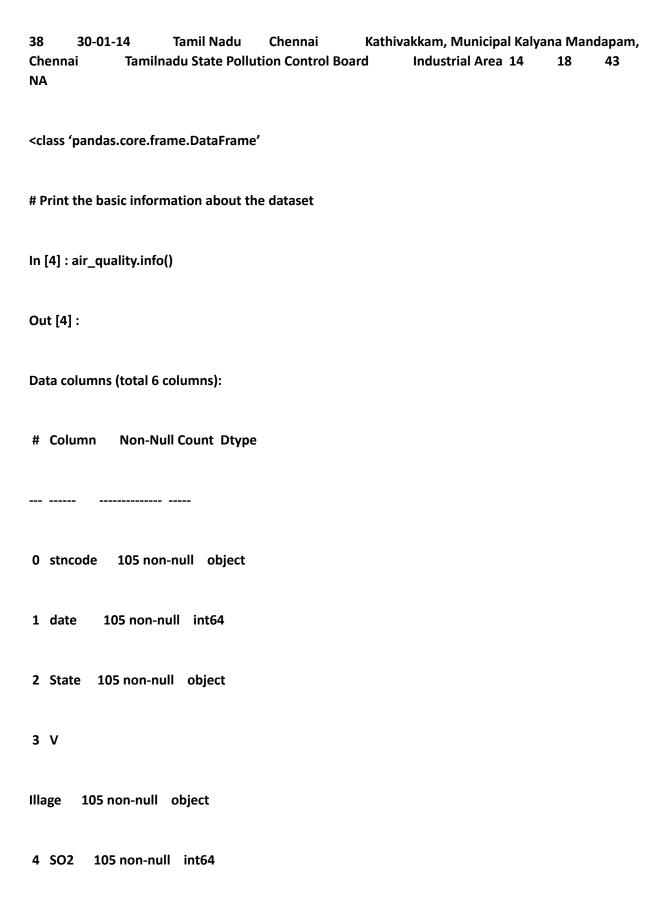
...

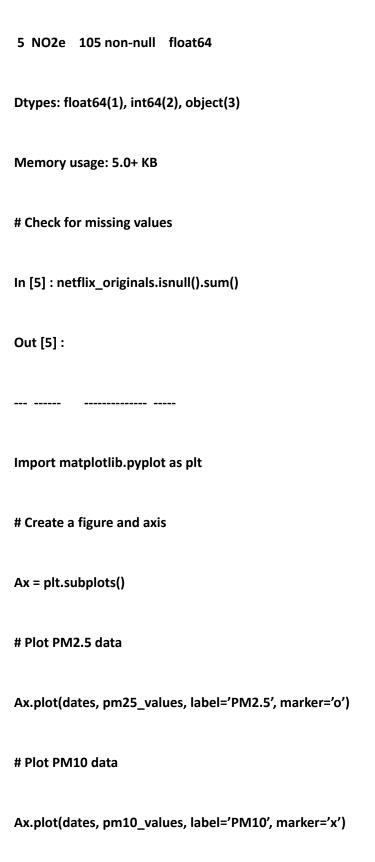
In [3]: air_qualitu.head()

Out [3]:

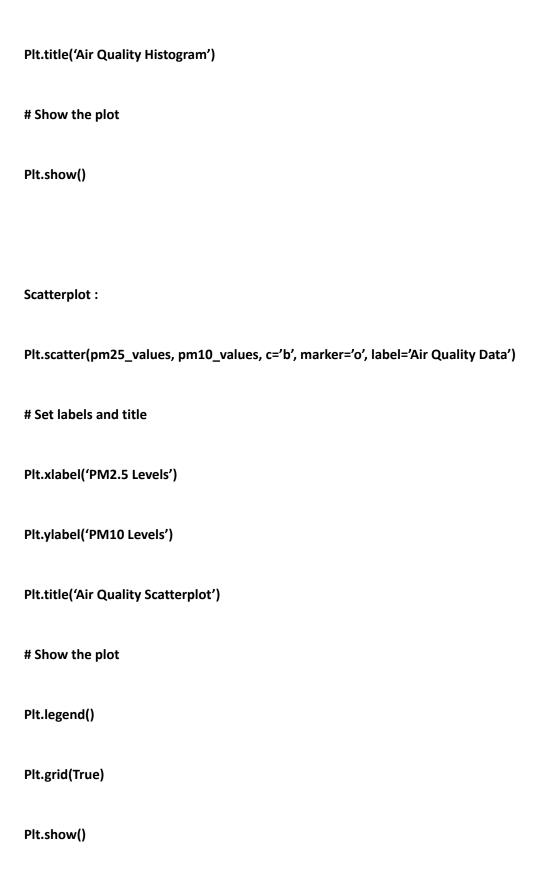
Stn Code Sampling Date State City/Town/Village/AreaLocation of Monitoring Station Agency Type of Location SO2 NO2 RSPM/PM10 PM 2.5

- 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





```
# 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')
```



```
Coding:
Import pandas as pd
From sklearn.model_selection import train_test_split
From sklearn.ensemble import RandomForestRegressor
From sklearn.metrics import mean_squared_error
Import matplotlib.pyplot as plt
# Load your dataset
Data = pd.read_csv('cpcb_dly_aq_tamil_nadu-2014.csv')
# Split the dataset into features (X) and target (y)
X = data.drop('AQI', axis=1)
Y = data['AQI']
# Split the data into training and testing sets
```

X = data.drop('AQI', axis=1)
Y = data['AQI']

Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)s

Create and train the Random Forest model
Rf_model = RandomForestRegressor(n_estimators=100, random_state=42)
Rf_model.fit(X_train, y_train)

Make predictions on the test set
Y_pred = rf_model.predict(X_test)

Calculate the Mean Squared Error (MSE) to evaluate the model
Mse = mean_squared_error(y_test, y_pred)
Print(f'Mean Squared Error: {mse}')

Visualize the feature importances

Feature_importances = rf_model.feature_importances_
Feature_names = X.columns
Plt.barh(feature_names, feature_importances)
Plt.xlabel('Feature Importance')
Plt.ylabel('Feature Name')
Plt.show()