Air Quality Analysis In TamilNadu

In this part to begin building our project by loading and preprocessing the dataset. Begin the analysis by loading and preprocessing the air quality dataset.

To begin our analysis and development project by loading and preprocessing an air quality dataset, I can follow these steps using Python and the Pandas library. In this example, I'll assume the given CSV file containing the air quality data. I can adapt the code to your specific dataset format.

Loading and Preprocessing Methods:

* Import Libraries
* Load the Dataset
* Explore the Dataset
* Handling Missing Data
* Data Cleaning
* Data Transformation
* Feature Engineering
* Exploratory Data Analysis (EDA)
* Save Preprocessed Dataset

### From The Given Dataset:

In this section I Use The Csv File Air Quality Analysis In Tamil Nadu Dataset From IBM Naan Mudhalvan

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

# Import Libraries:

Import the necessary libraries, primarily Pandas for data manipulation and NumPy for numerical operations. You may also need other libraries based on the specifics of your dataset.

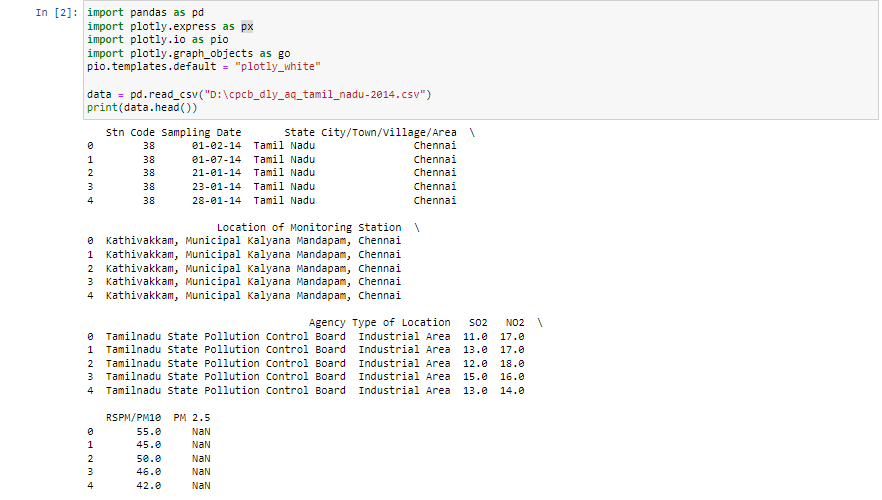
* *import pandas as pd*
* *import numpy as np*

## Load the Dataset:

Load your dataset into a Pandas DataFrame. Replace 'your\_dataset.csv' with the actual file path or URL of your dataset.

data = pd.read\_csv('Example.csv')

If you have a different format (e.g., Excel, JSON), you can use appropriate Pandas functions like pd.read\_excel() or pd.read\_json().

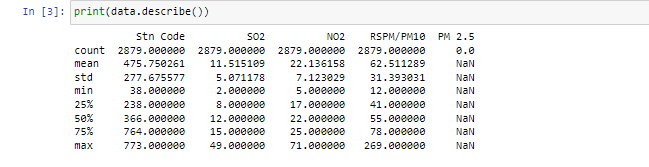


### Explore the Dataset:

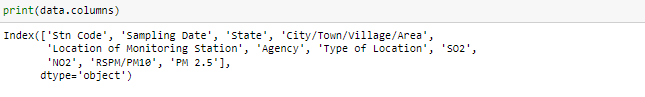
Begin by getting an overview of your dataset. Check the first few rows, column names, and data types.

Ex:

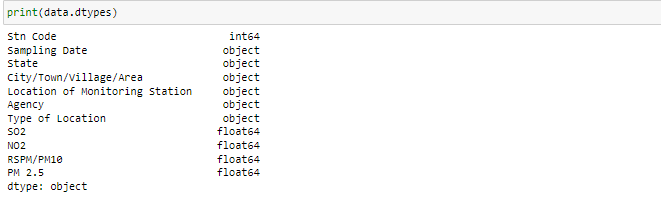
* print(df.head()) # Display the first few rows



* print(df.columns) # List of column names



* print(df.dtypes) # Data types of each column



### Handling Missing Data:

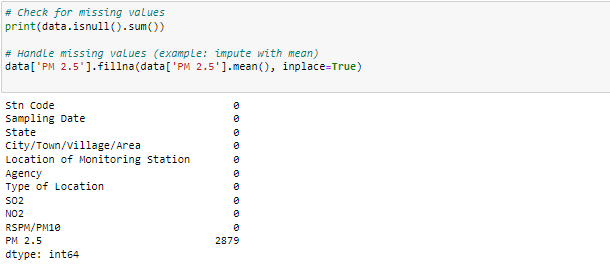
Identify and handle missing data, which could involve removing rows with missing values or imputing missing values.

# Check for missing values

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

# Handle missing values (example: impute with mean)

df['column\_name'].fillna(df['column\_name'].mean(), inplace=True)



### Data Cleaning:

Clean the data by addressing any data anomalies, inconsistencies, or outliers.

### Data Transformation:

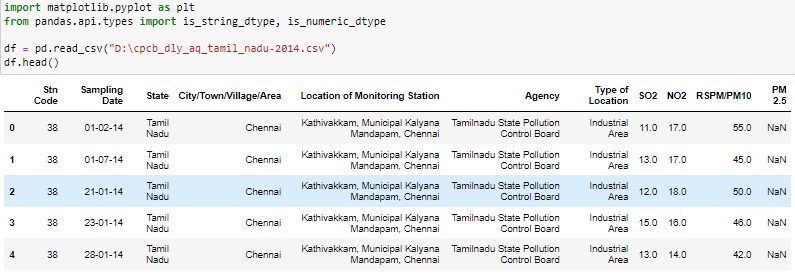
Depending on your project's requirements, you may need to transform the data. This could include converting date columns to datetime objects, encoding categorical variables, or scaling numerical features.

import matplotlib.pyplot as plt

from pandas.api.types import is\_string\_dtype, is\_numeric\_dtype

df = pd.read\_csv("../input/marketing-data/marketing\_data.csv")

df.head()



### Feature Engineering:

Create new features or modify existing ones to improve your dataset's quality.

### Exploratory Data Analysis (EDA):

Perform exploratory data analysis using visualizations (e.g., Matplotlib or Seaborn) to gain insights into your data.

### Save Preprocessed Dataset:

Once you've completed preprocessing, save the cleaned and transformed dataset to a new file for future use.

df.to\_csv('preprocessed\_dataset.csv', index=False)

6.PNG

These steps provide a general guideline for loading and preprocessing a dataset. The specifics may vary depending on your dataset, project goals, and data quality

------------------------.-------.--------------.------------------------