Air Quality Prediction And Analysis In Tamilnadu

Phase -4

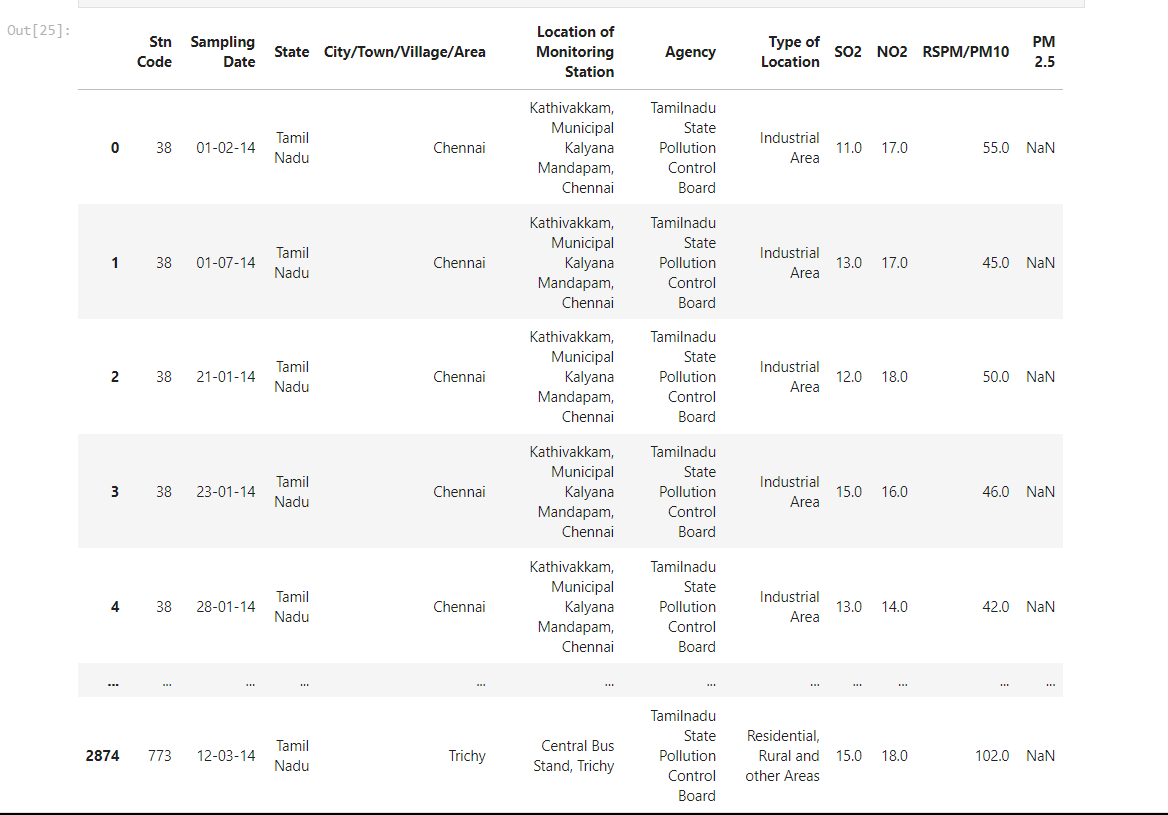
Development part-2

Shastivelan.G(team leader)

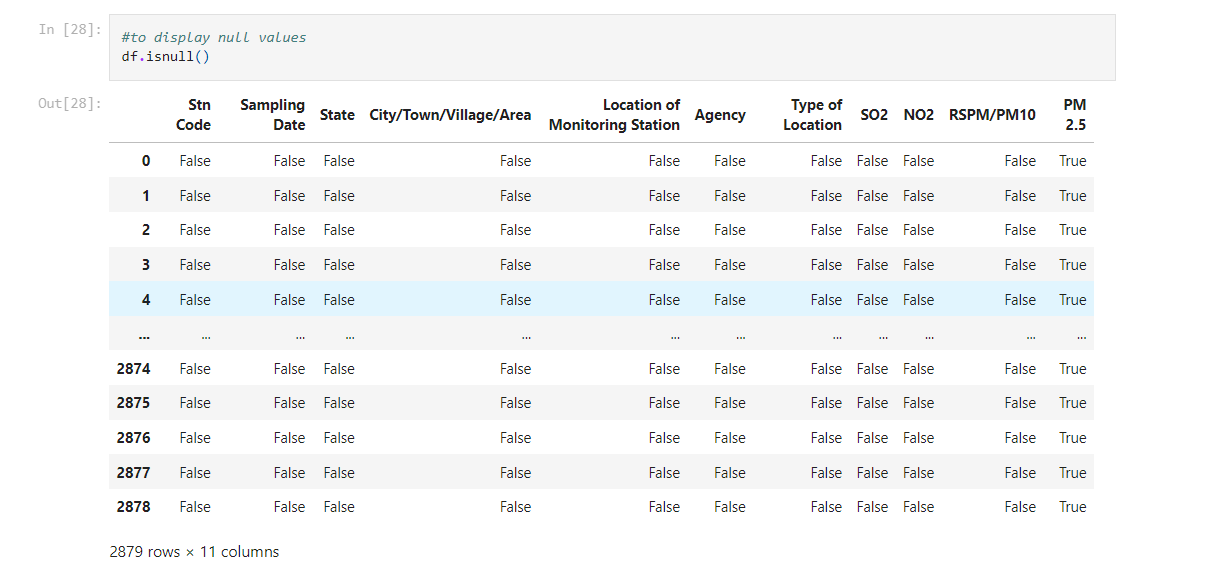
**Explanation for Development part-1:**

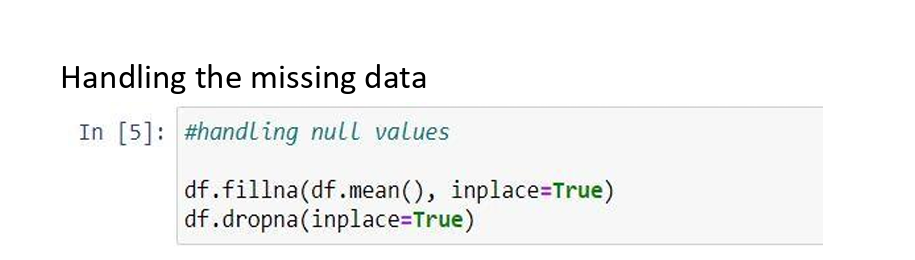
Step 1: Importing the required libraries and loading the dataset



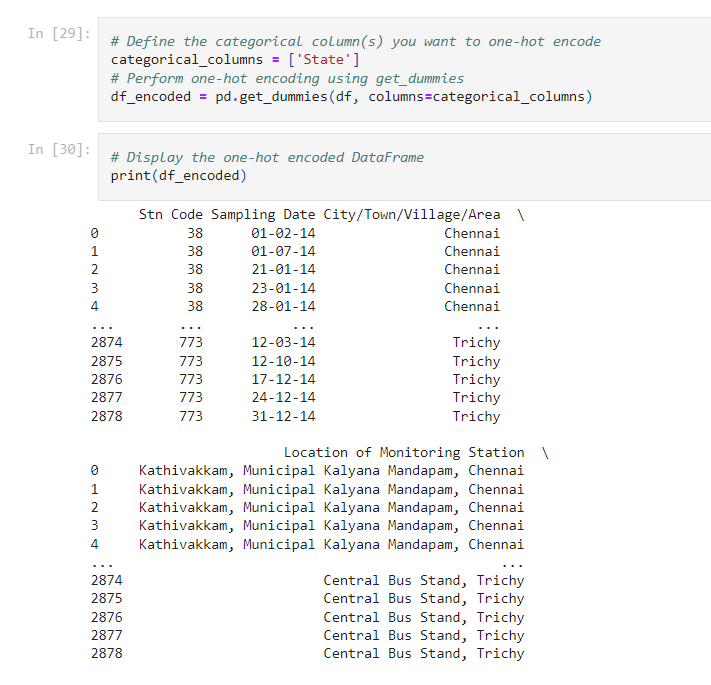


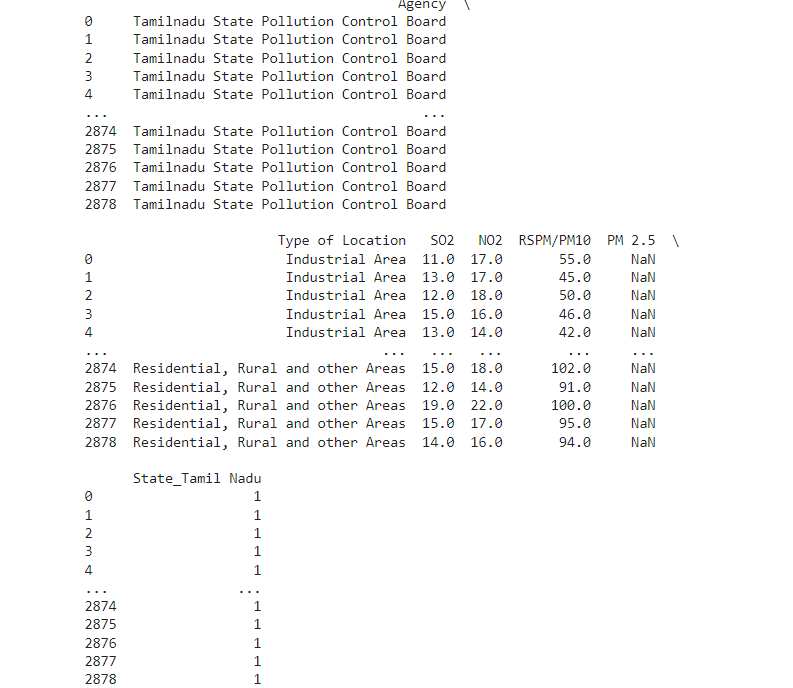
**Step 2: Handling Missing Data**



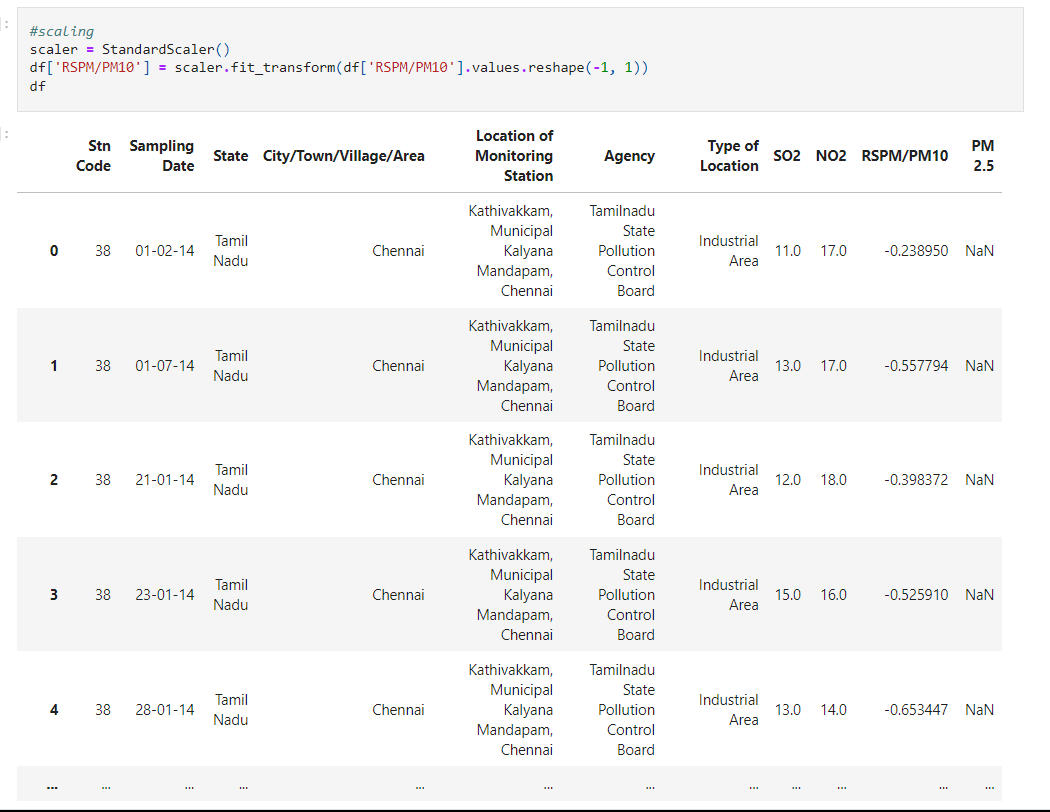


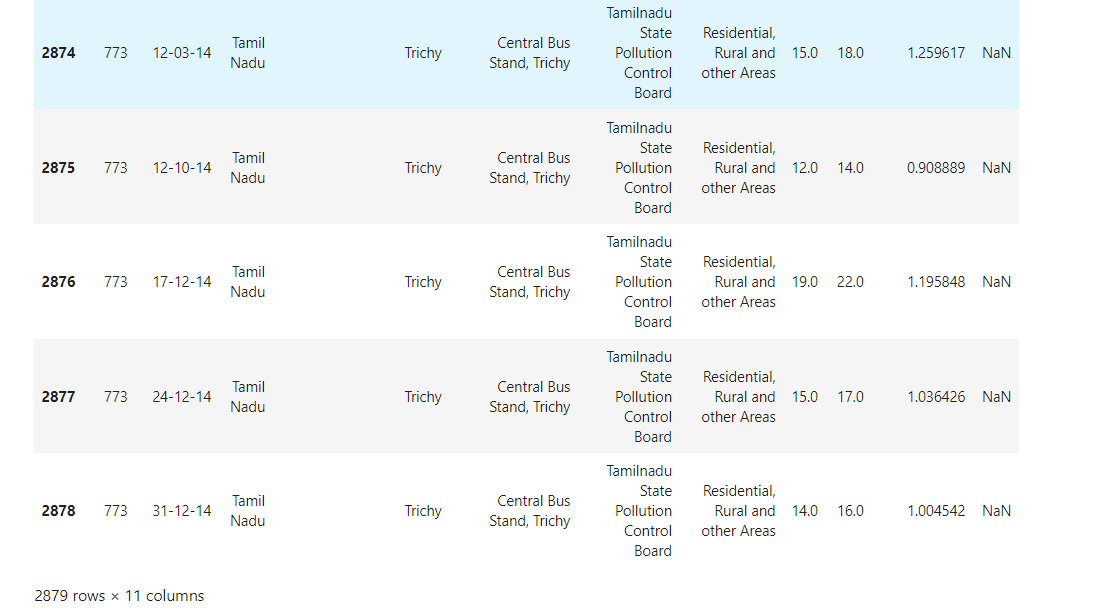
**Step:3 label Encoder**



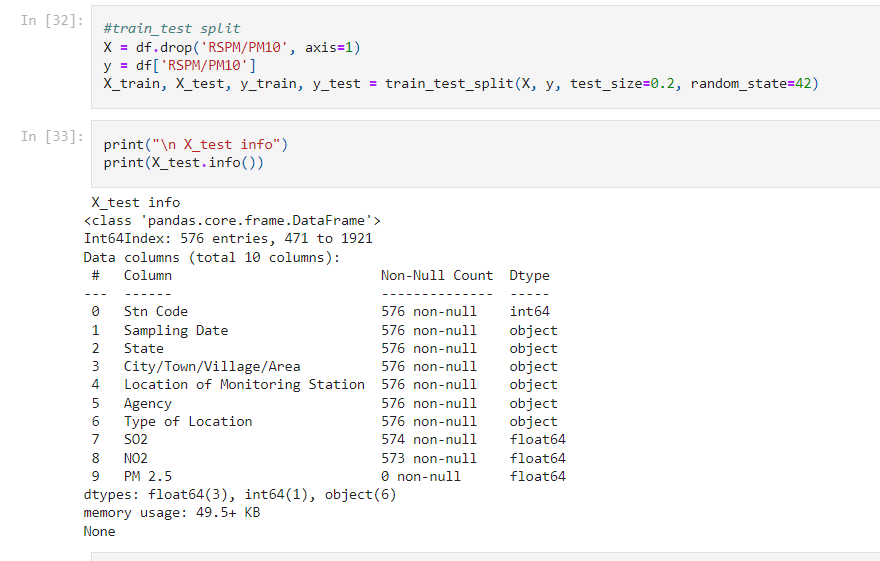


**Step 4: Feature Scaling using StandardScaler**





**Step 5: Splitting the data into a training set and a test**

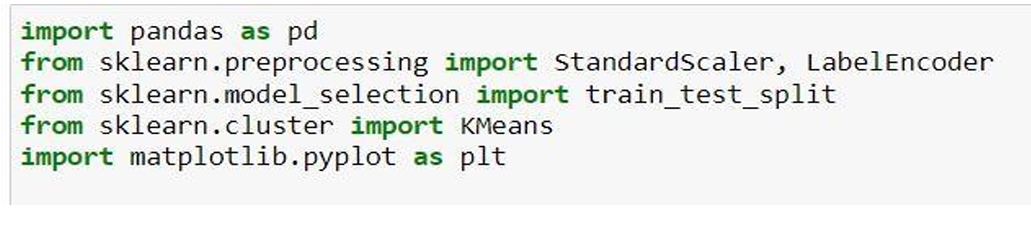
****

**Algorithm for Air Quality Prediction and analysis :**

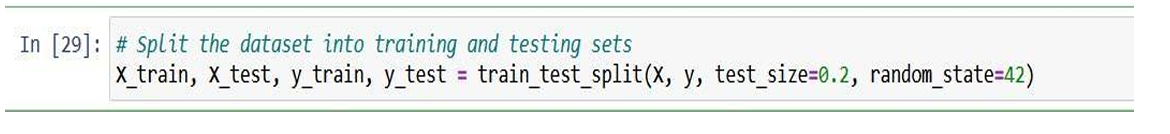
* **Objective:** This algorithm aims to guide the development of a customer segmentation using the provided dataset. It covers essential steps, including feature engineering, model training, and evaluation, to ensure accurate predictions
* **Import necessary libraries:-** Import essential Python libraries, including pandas, scikit-learn, and matplotlib, for data manipulation, clustering, and visualization
* **Suppress FutureWarnings**:- Configure the system to suppress FutureWarnings to prevent unnecessary warning messages
* **Read the dataset:-** Load the customer data from a CSV file located at a specified file path.- Use the specified encoding (ISO-8859-1) to read the data
* **Data Exploration:-** Display the DataFrame (`df`) to inspect the loaded data.- Check the data's information, including data types and missing values.- Display the first few rows of the dataset for a quick overview
* **Handling Missing Values:-** Check for missing values within the dataset.- Fill missing values with the mean of the respective columns.- Drop rows with any remaining missing values.
* **Label Encoding:-** Apply label encoding to the 'state' column to convert categorical values into numerical values (e.g., 0 and 1)
* **Feature Scaling:** Use StandardScaler to scale the 'RSPM/PM10' column to have a mean of 0 and a standard deviation of 1. Standardization helps ensure that features with different scales contribute equally to clustering.
* **Data Splitting:--** Split the dataset into features (X) and the target variable (y). Divide the data into training and testing sets using train\_test\_split, with a specified test size and randomseed
* **K-Means :-** Define the number of clusters (k) for K-Means clustering (in this case, k=5).- Apply K-Means clustering to the feature data (X) to segment customers into 'k' clusters.- Assign cluster labels to the data points
* **Visualization:-** Create ascatter plot to visualize the clusters.

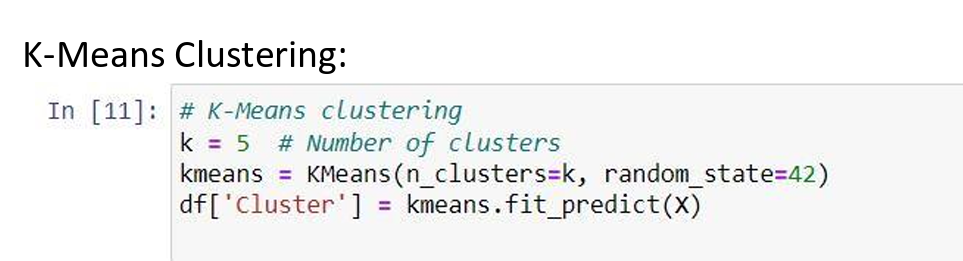
**Execution of the K-Means:**

Importing the necessary libraries:

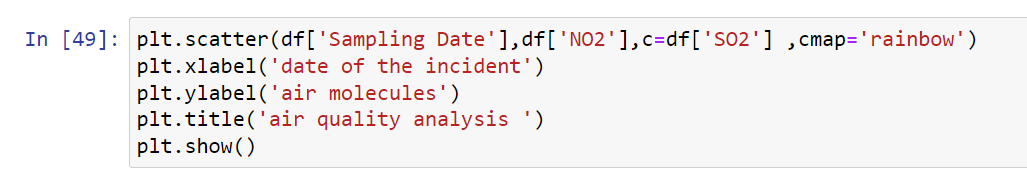
****

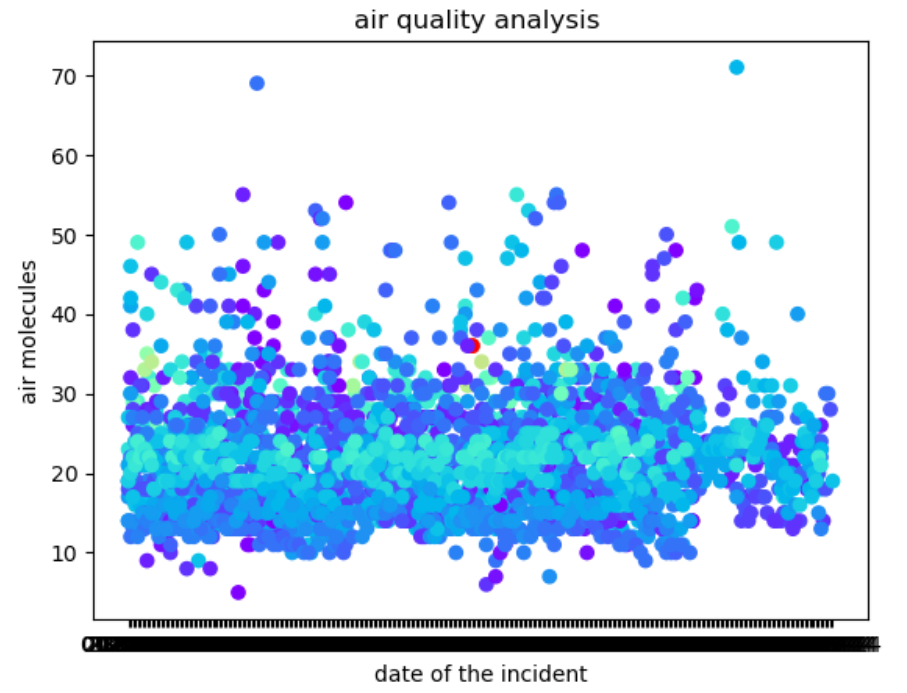
**Train test split:**

****

****

**Visualization of the result:**

****

****