**DAC Phase 3:**

**Problem Statement: Air Quality Assessment of TamilNadu**

**Loading and Pre-processing of data:**

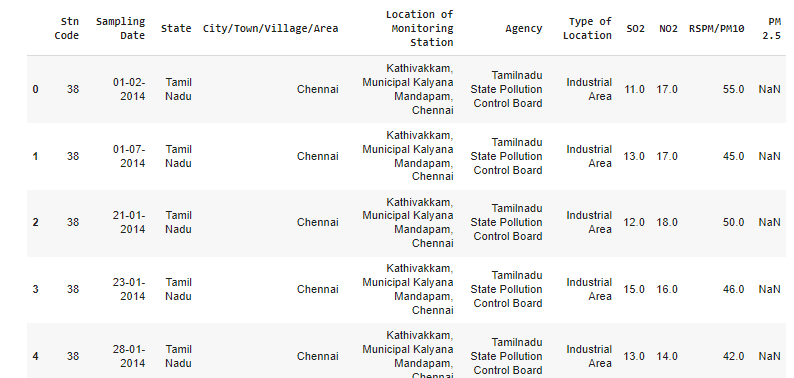
**Loading data import pandas as pd**

import pandas as pd

import numpy as np

data = pd.read\_csv('/content/Air\_quality.csv')

data.head(5)



data.describe()

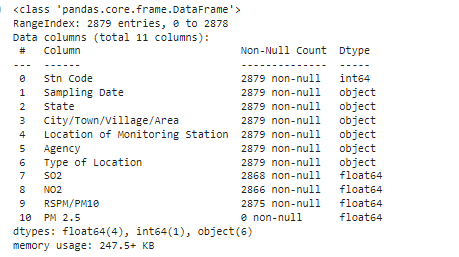


This command is used to view a concise summary of the dataset, including important statistical parameters such as percentiles, standard deviation, mean, minimum, and maximum values for each column, along with a count of data points in each column

data.shape

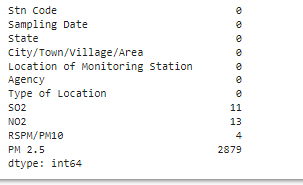


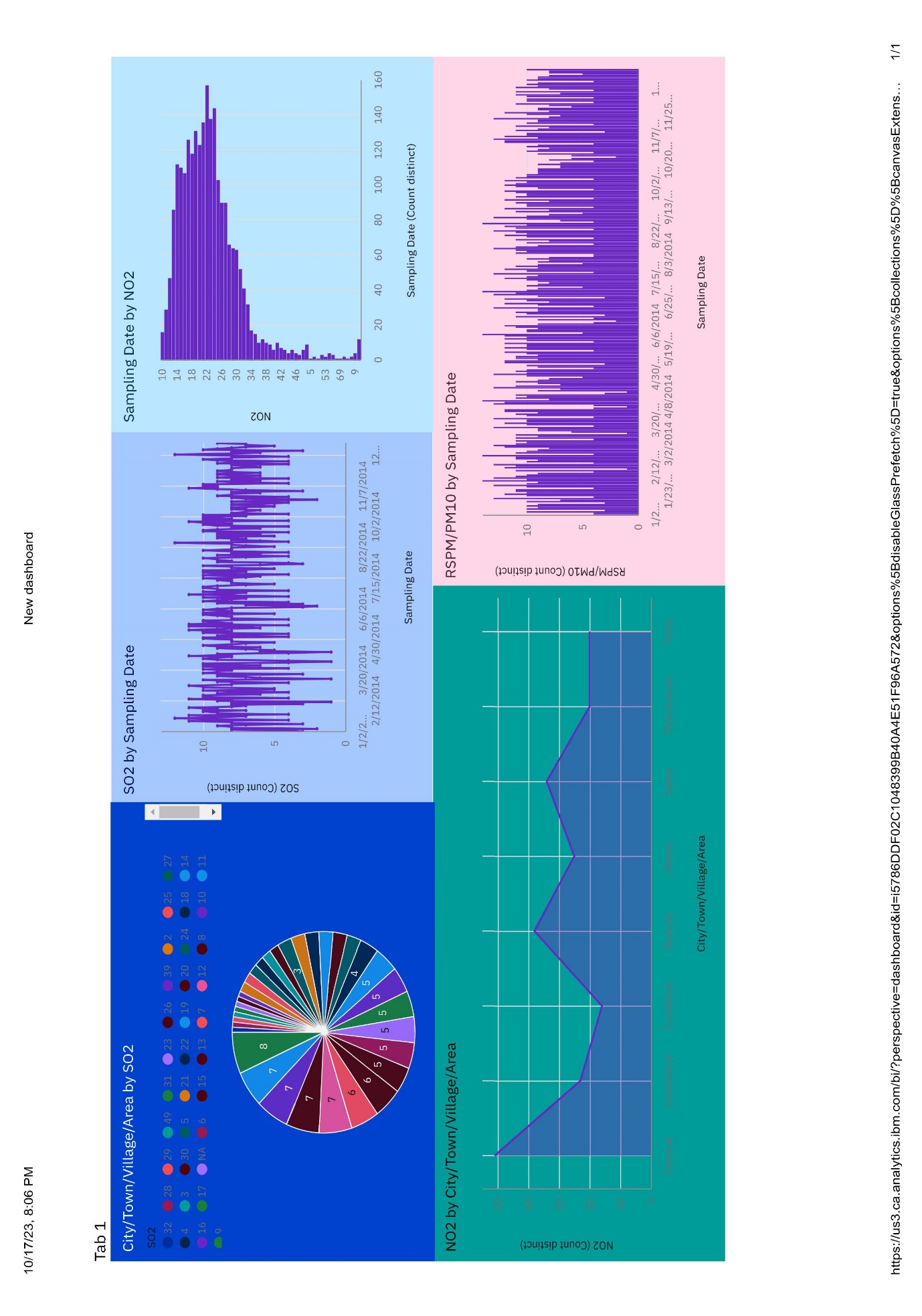
data.info()



The 'info' command is used to check the data type of each column and the count of non-null values in each column. The main difference between 'describe()' and 'info()' is that 'describe()' provides statistical parameters such as mean and standard deviation, while 'info()' does not include these mathematical statistics

data.isna().sum()



The command above is utilized to detect null values in each column. It is evident that there are null values present in columns like SO2, NO2, and RSPM. It is imperative to address and rectify these null values in the dataset.

mean\_no2 = data['NO2'].mean()

data['NO2'] = data['NO2'].fillna(mean\_no2)

mean\_so2 = data['SO2'].mean()

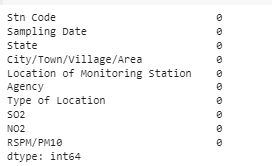
data['SO2'] = data['SO2'].fillna(mean\_so2)

mean\_rspm = data['RSPM/PM10'].mean()

data['RSPM/PM10'] = data['RSPM/PM10'].fillna(mean\_rspm)

data.drop('PM 2.5',axis=1,inplace=True)

data.isna().sum()

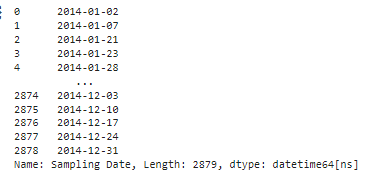


The fillna() method is utilized to replace missing or null values with the mean of the corresponding column.

**Converting the date column to date format from object**

data['Sampling Date'] = pd.to\_datetime(data['Sampling Date'])

data['Sampling Date'].dtype



The data type of the 'Sampling Date' column was initially 'object,' which is not suitable for training a model or analyzing the dataset. Therefore, the data type of the column was converted to pandas date and time using the pandas.to\_datetime() function.

**Insights**

Chennai has the highest RSPM/PM10 level at 654, with SO2 contributing the most at 59.

Coimbatore has an RSPM/PM10 level of 61.

From January 30, 2014, to January 31, 2014, the RSPM/PM10 level in location 10 increased by 300%.

Chennai has the highest SO2 levels, with Station Code 161 being the primary contributor.

Chennai is the most frequently occurring City/Town/Village/Area category, accounting for 1000 items with RSPM/PM10 values (34.7% of the total).

The total number of results for RSPM/PM10 across all City/Town/Village/Area categories is nearly three thousand.