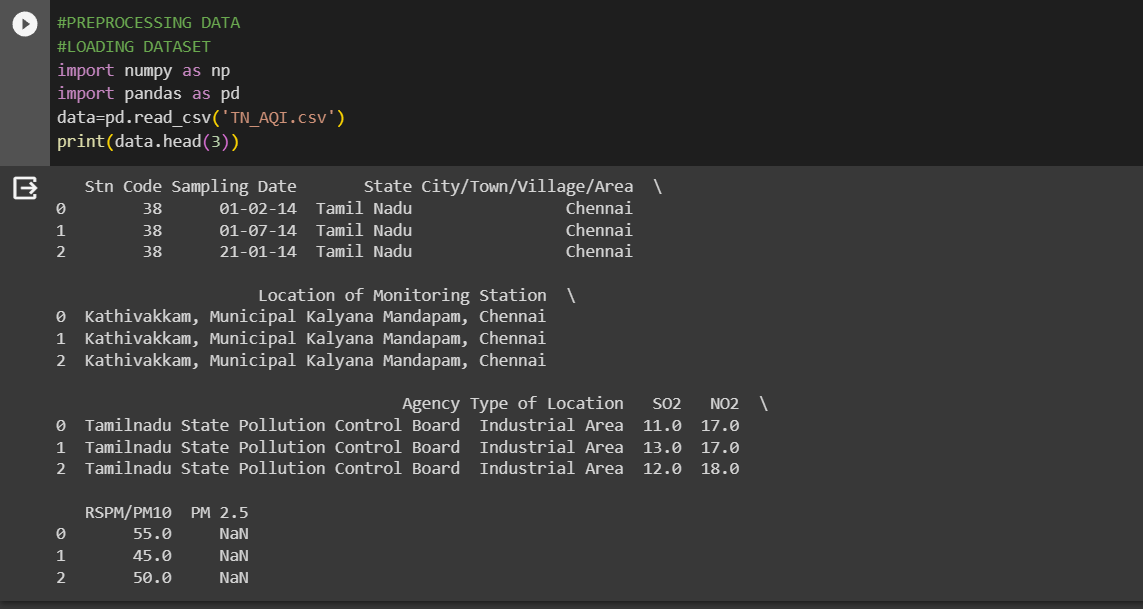
**AIR QUALITY ANALYSIS IN TAMILNADU**

**PHASE 3: DEVELOPMENT PART 1**

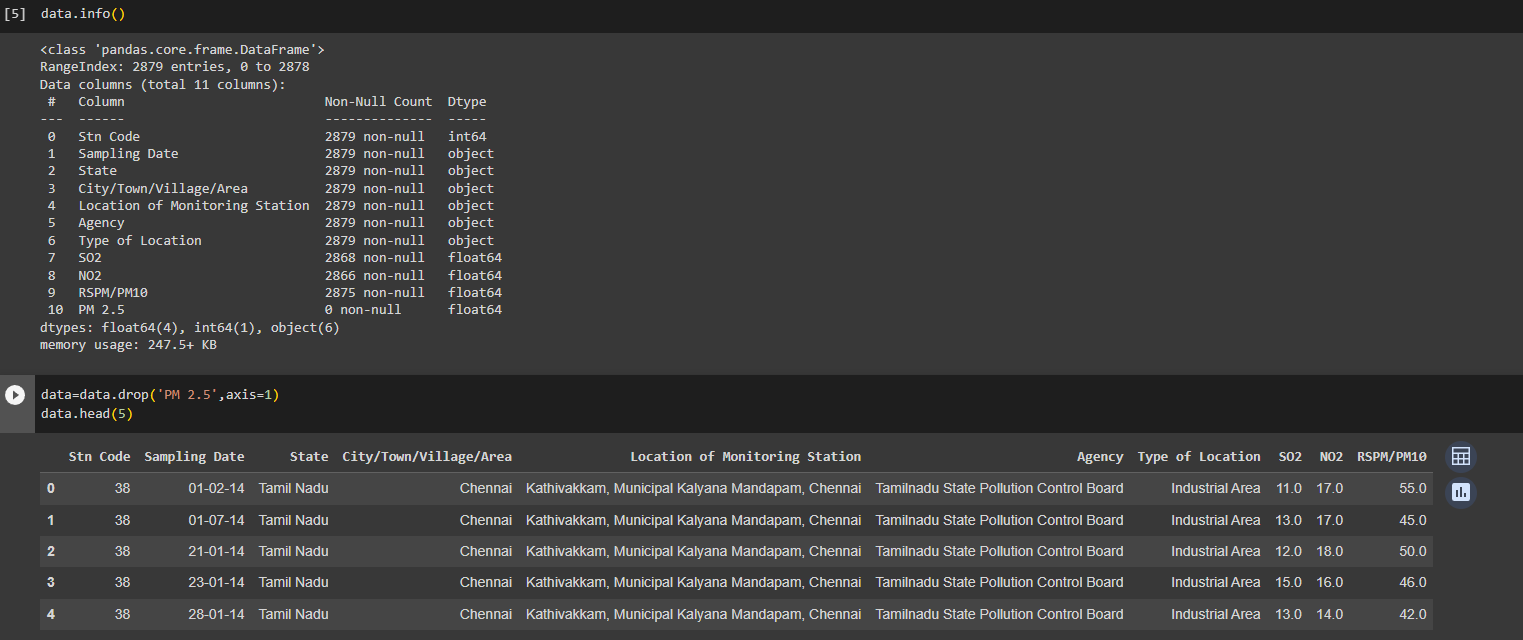
***DATA PREPROCESSING:***

1.Loading The Dataset:

We are using google colab python for data preprocessing. First we need to import necessary libraries like pandas for loading the dataset. Checking few rows of data using head() function.



## **2. Drop Columns That Aren’t Useful: We shall drop the PM 2.5 column since it contains zero non-null values and we require only three pollutant’s sub-index for finding AQI(Air Quality Index).**





## **3. Drop Rows With Missing Values: We can drop rows that have Null values. When rows are dropped in large amount,data is being wasted.To avoid this we can fill them with better values like median,average or interpolate methods.**

## **Anyhow we have dropped only 0.6% of the data and have 2862 rows of Non-Null values. So dropping the missing values is considered best in this scenario to avoid miscalculations.**

## 

## *VISUALIZATION WITH IBM COGNOS:*

## Let us use some visualization techniques to overview or understand the data easily.

## Here we have showed the values of the pollutants like NO2, SO2, PM 10 and where they range. Also a comparison between those pollutants is given. The values of pollutants over a period of time is visually represented using Line and Barplot as below.

## *CONCLUSION:*

## Thus ,data pre-processing and visualization using IBM Cognos has been executed and documented successfully.