**Air Quality Analysis and Prediction in Tamil Nadu**

**Phase 3- Loading Pre-Processing Of Dataset**

P. Kishore (411621104025)

Harish Prabhavathy (411621104014)

Kakimetla Prashant (411621104018)

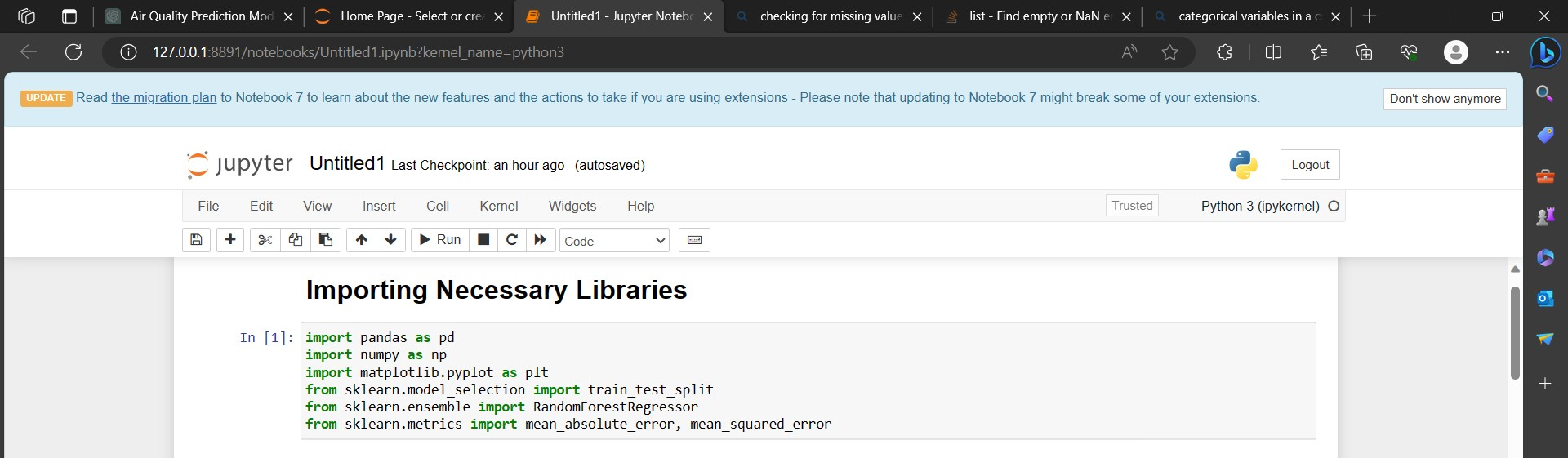
A. Kaaleshwaran (411621104019)

G. Gokula Prasadth (411621104012)

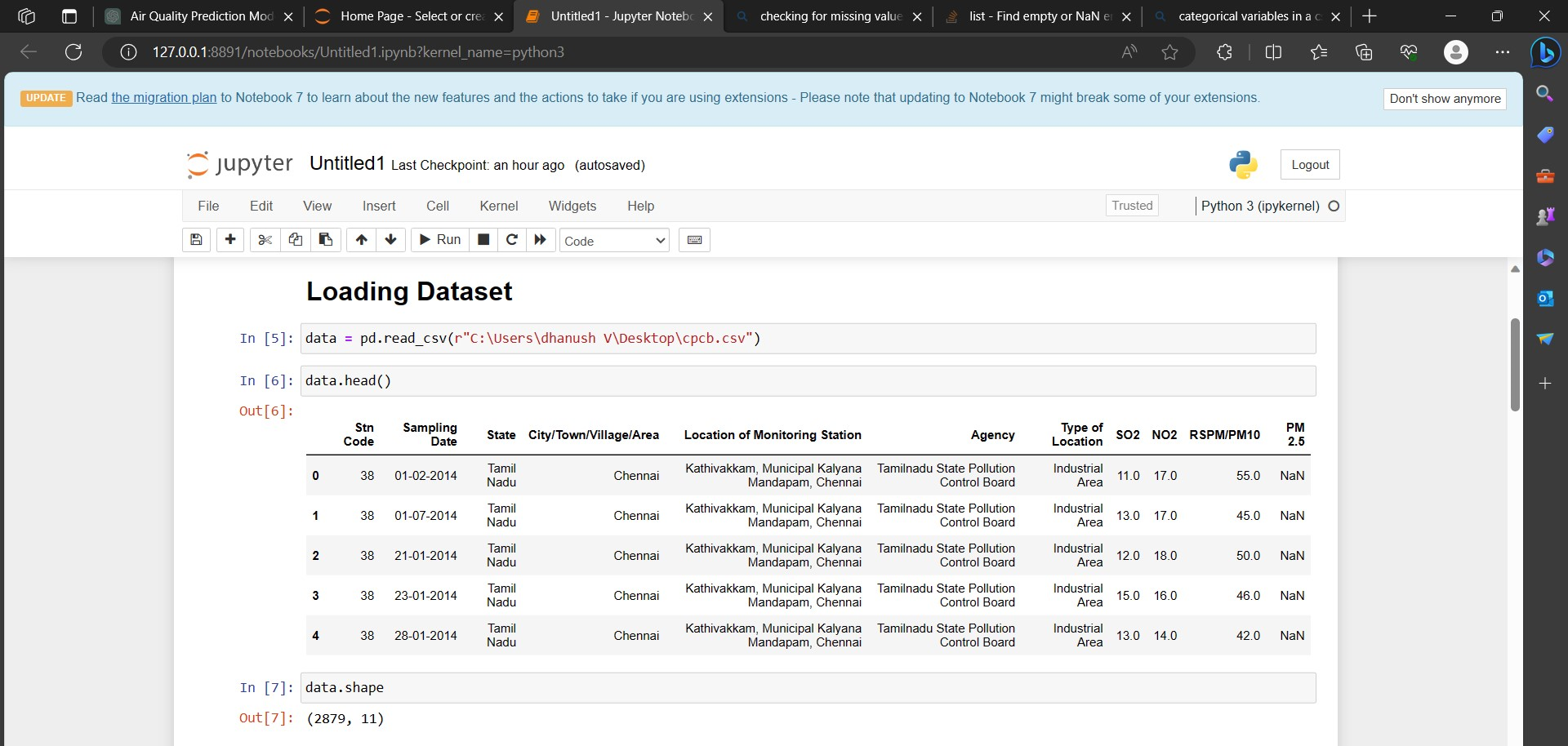
**Introduction:**

Air quality is a crucial aspect of environmental health, impacting the well-being of individuals and communities. This project focuses on the development of an Air Quality Analysis and Prediction model using data science techniques in Python. As a student passionate about environmental sciences and data analytics, I embarked on this project to contribute to a cleaner and healthier environment.

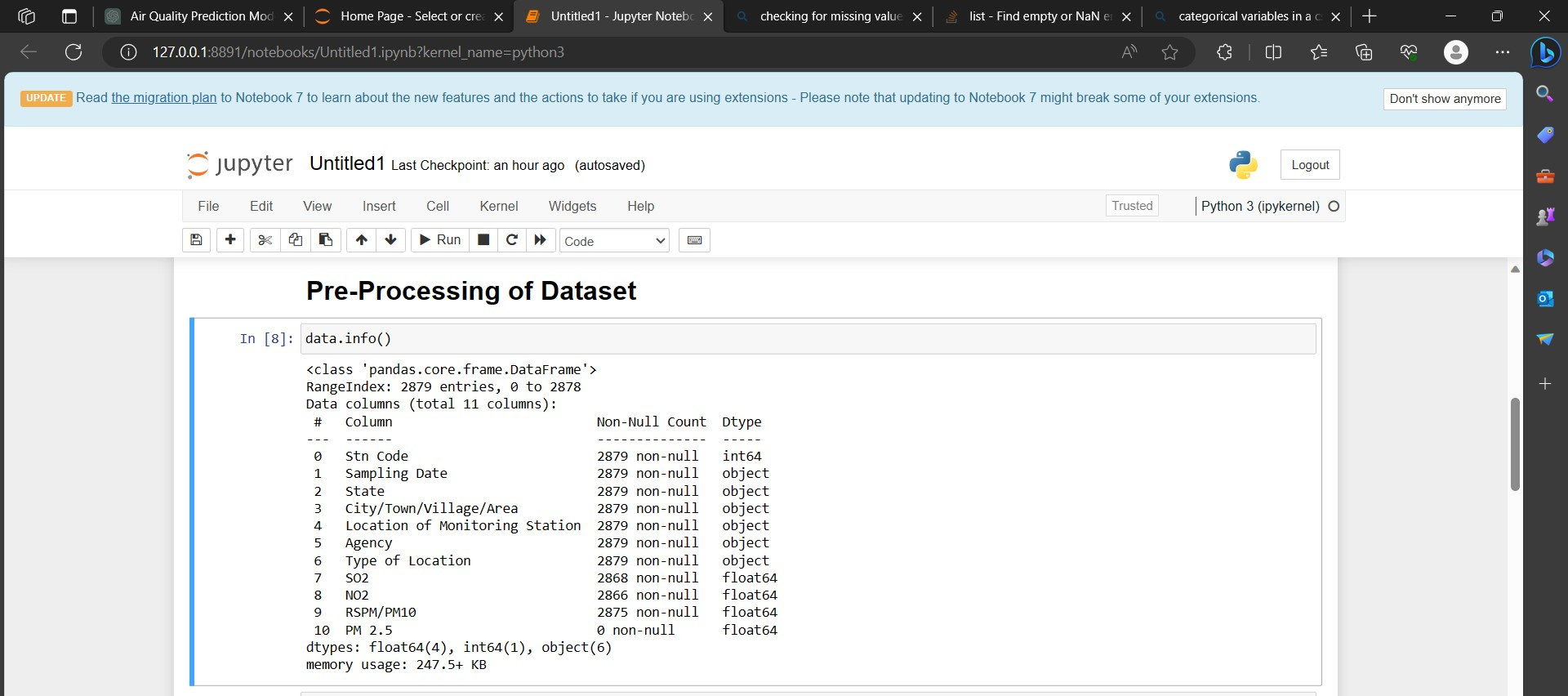
**Importing Necessary Libraries:**

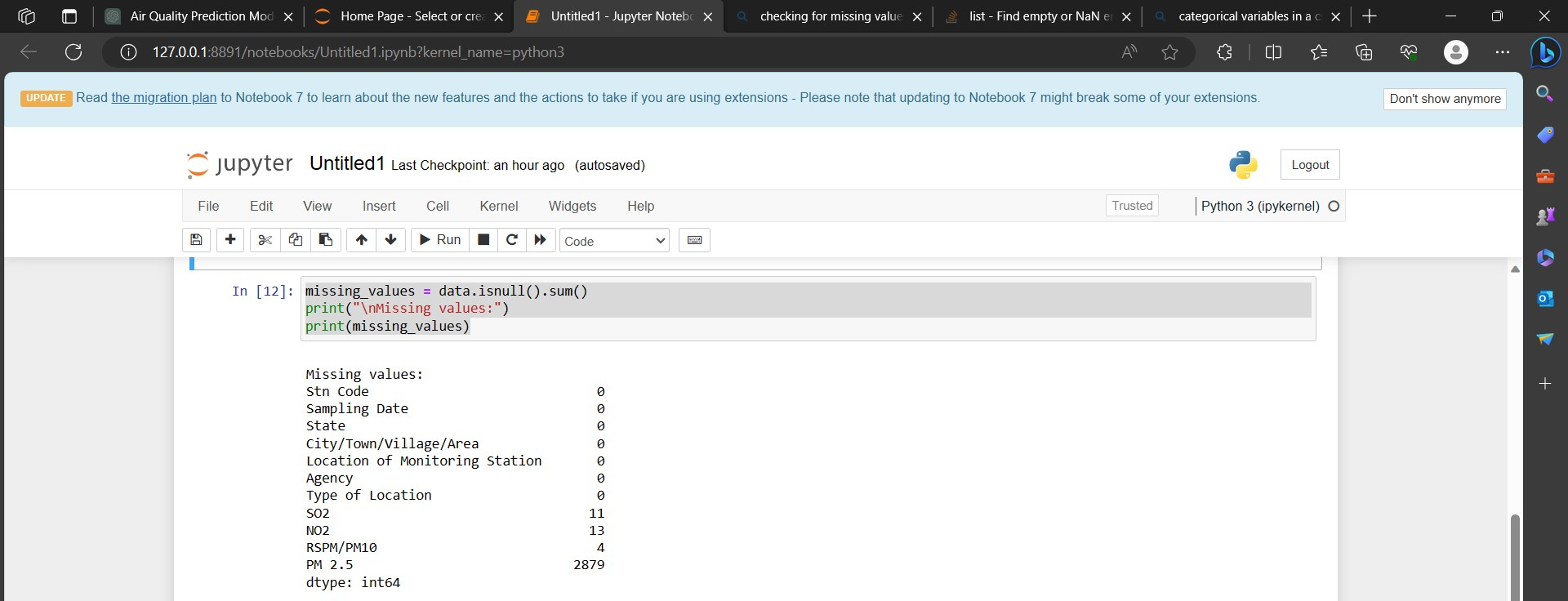
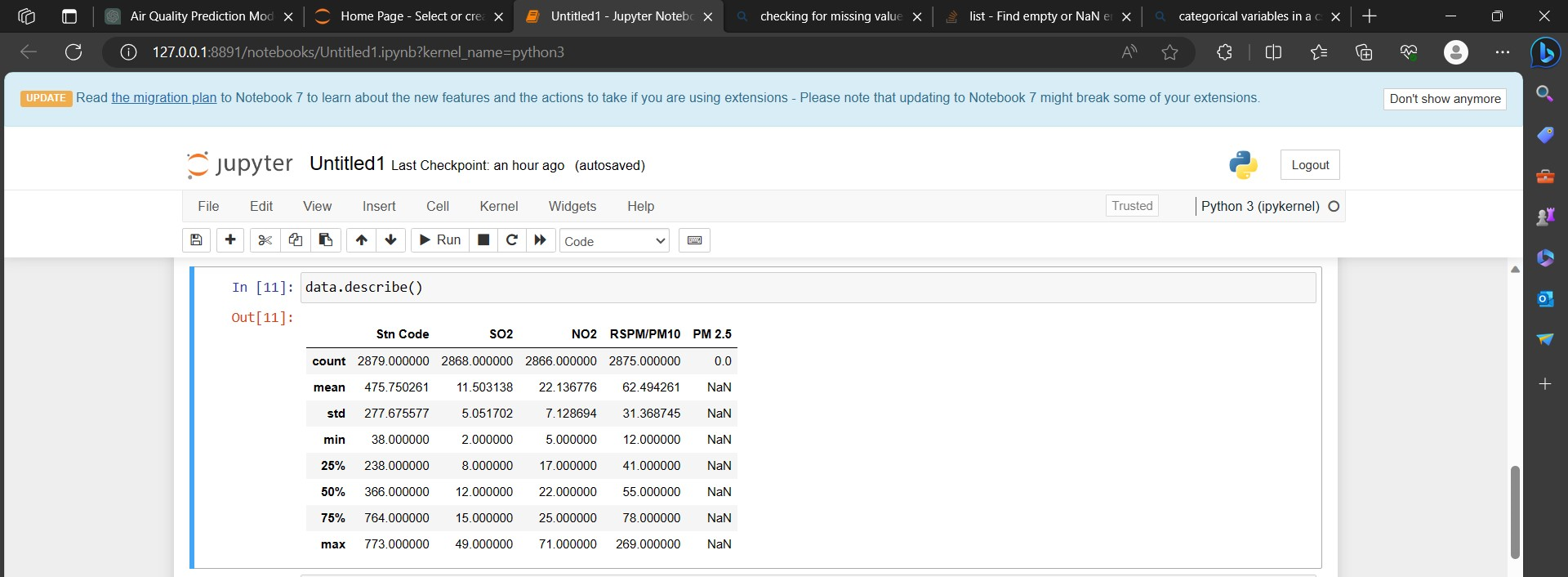


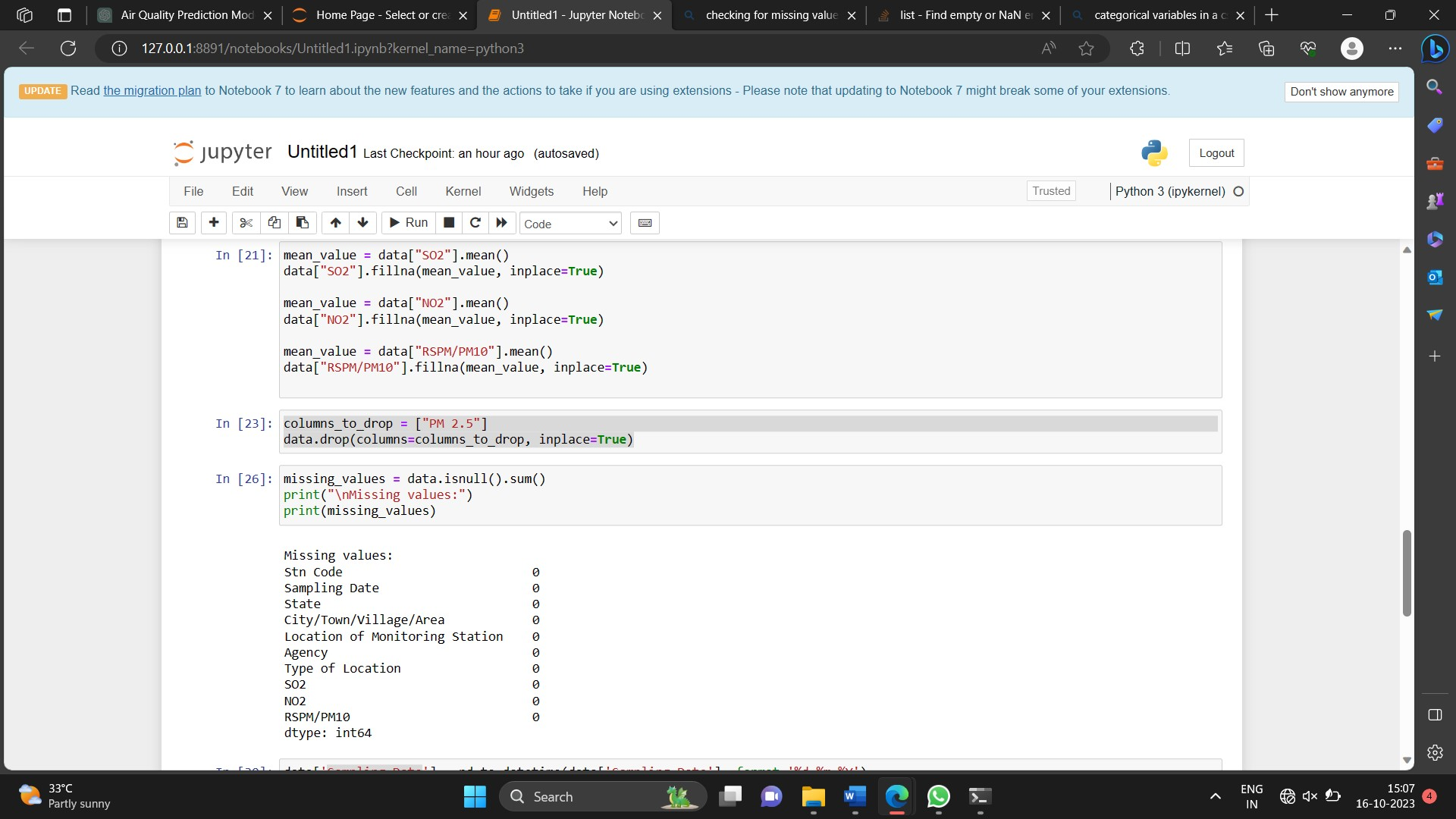
**Loading Of Dataset:**

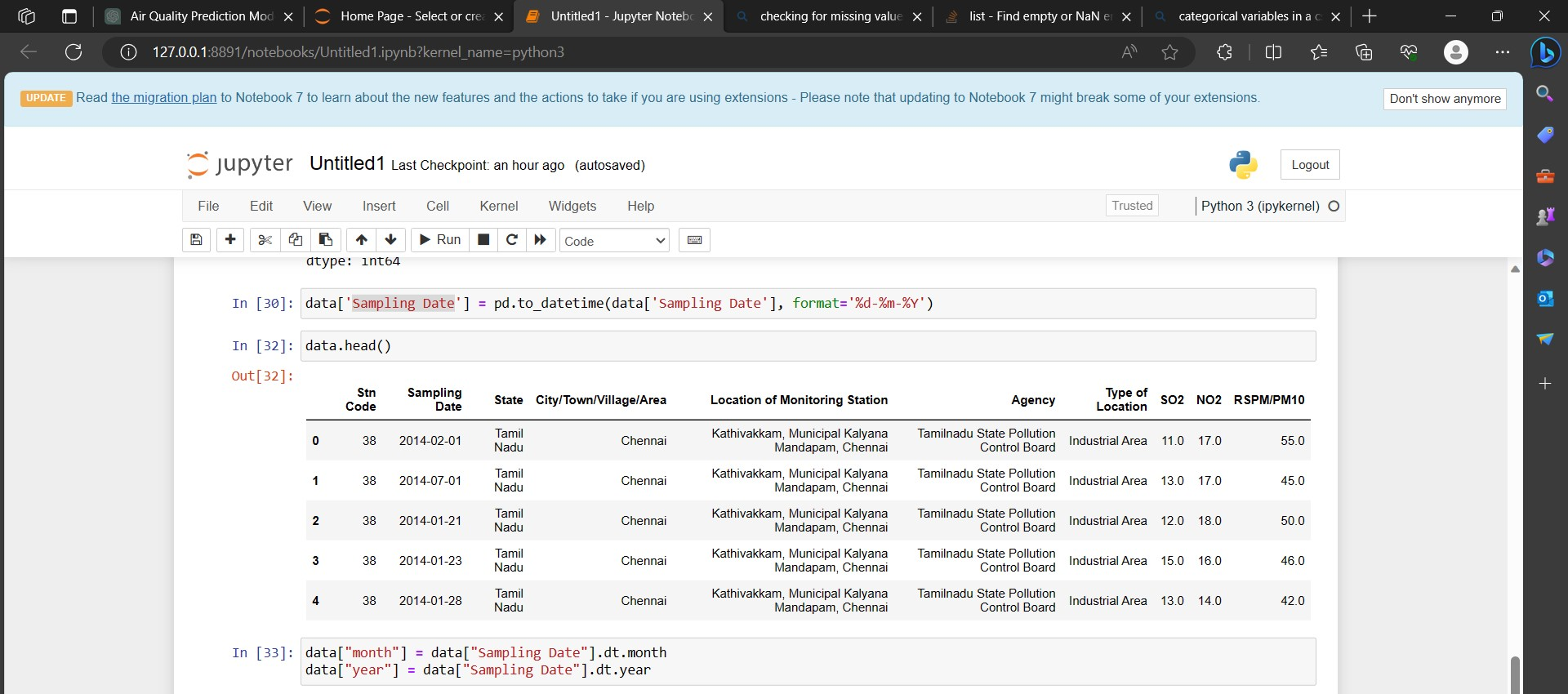


**Pre-Processing of Dataset:**











We first import the necessary libraries: Pandas for data manipulation, NumPy for numerical operations, and Matplotlib for data visualization.

We load the dataset using pd.read\_csv("your\_dataset.csv"). Replace "your\_dataset.csv" with the actual path to your dataset.

We display basic information about the dataset using data.info(). This provides details about column names, data types, and the presence of missing values.

We display the first few rows of the dataset using data.head(), allowing us to understand the structure of the data.

We check for missing values using data.isnull().sum(), and we handle missing values as needed. In this example, we replace missing values in a column with the mean value.

We drop irrelevant columns using data.drop(columns=columns\_to\_drop). Replace columns\_to\_drop with the names of the columns you want to remove.

We convert data types when necessary. In the example, we convert a column to datetime format.

We perform feature engineering, creating new features if necessary. Here, we extract the month and year from a date column.

If there are categorical variables, we encode them into numerical values using one-hot encoding.

We split the data into training and testing sets using train\_test\_split.

You can adapt this code to your specific dataset and preprocessing needs.