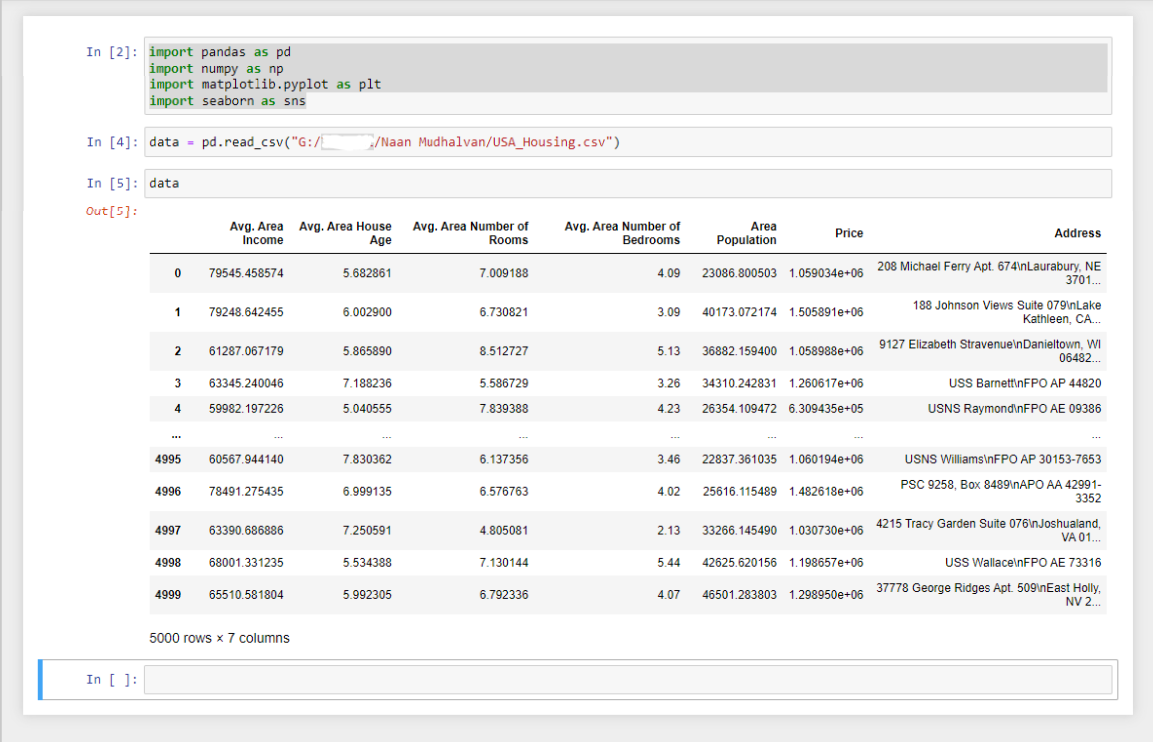
**Loading and Preprocessing the Dataset**

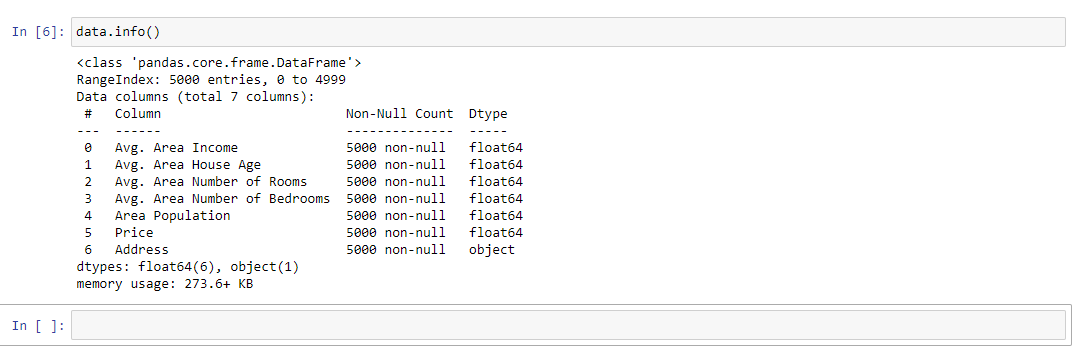
**Loading the Dataset**

* **Using pandas.read\_csv() function, we read the USA\_Housing dataset.**

****

**Check info for any null values**

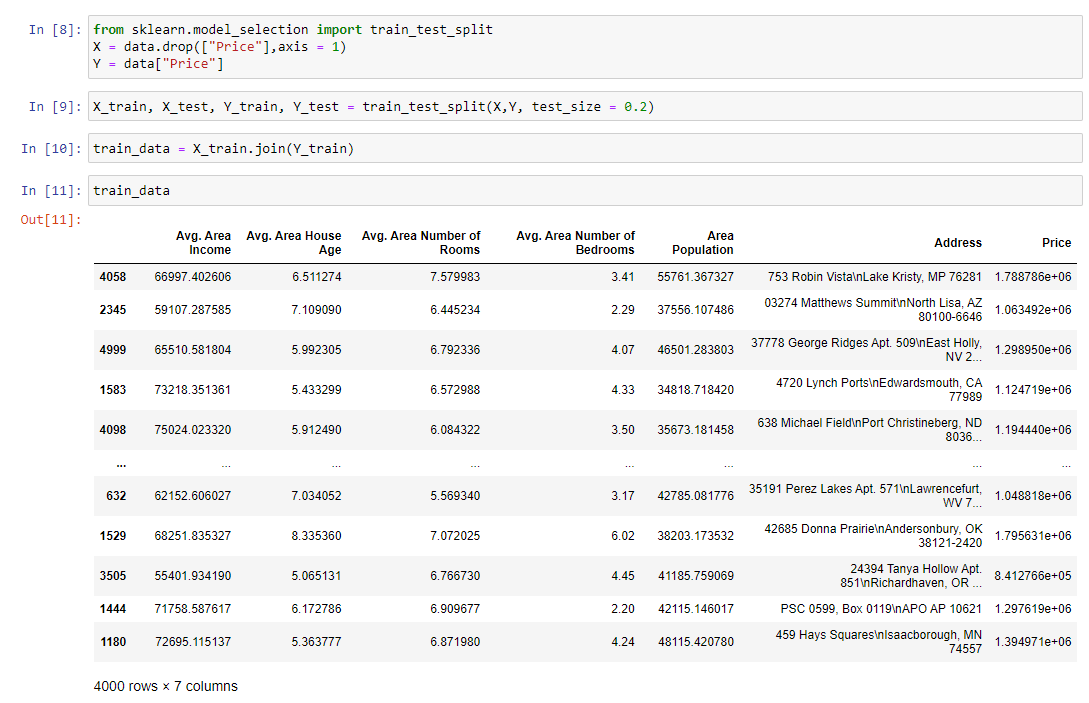
* **we use the info() function in pandas to check all data values**

****

* **Since the output showed no Null values we are free to proceed.**

**Splitting data into training and testing set**

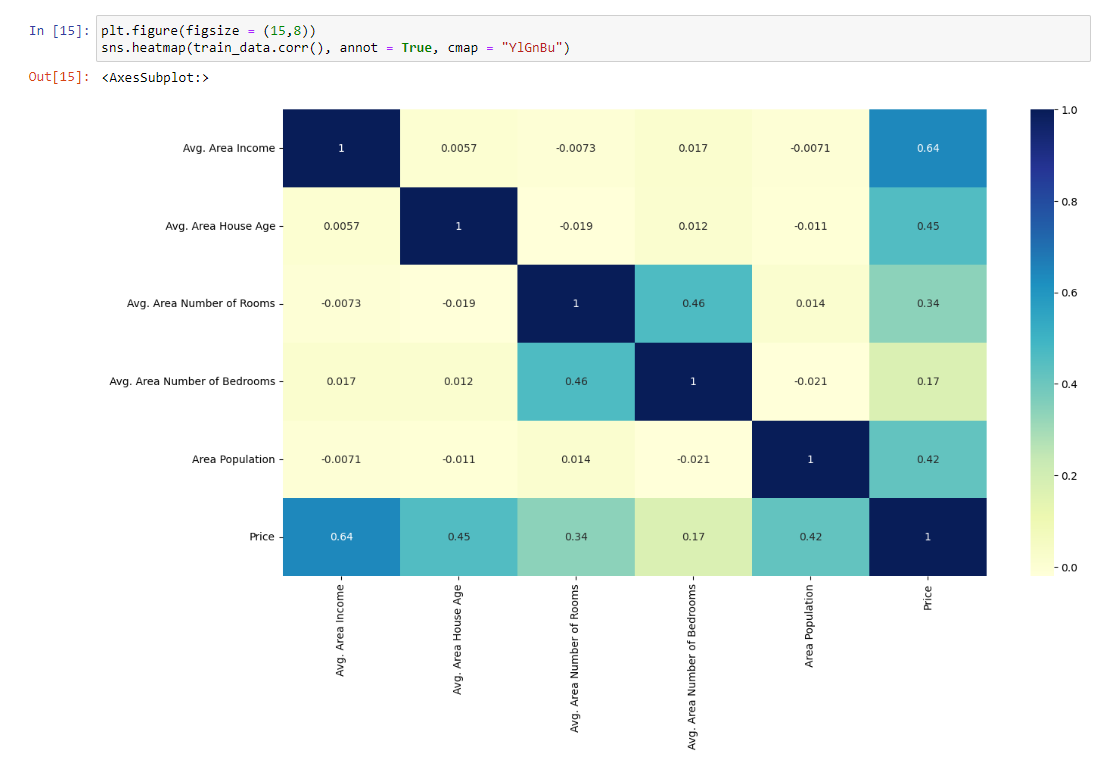
* **we drop the target variable from the dataset and set it to X and set Y with the target variable data. (Target variable: Price)**
* **then we split both X and Y into training and testing sets.**
* **Finally we join our training set for X and Y then store it in train\_data**

****

**Finding Correlation between all data with target variable (“Price”)**

* **We use the heatmap() function from seaborn to visualize the correlation between**

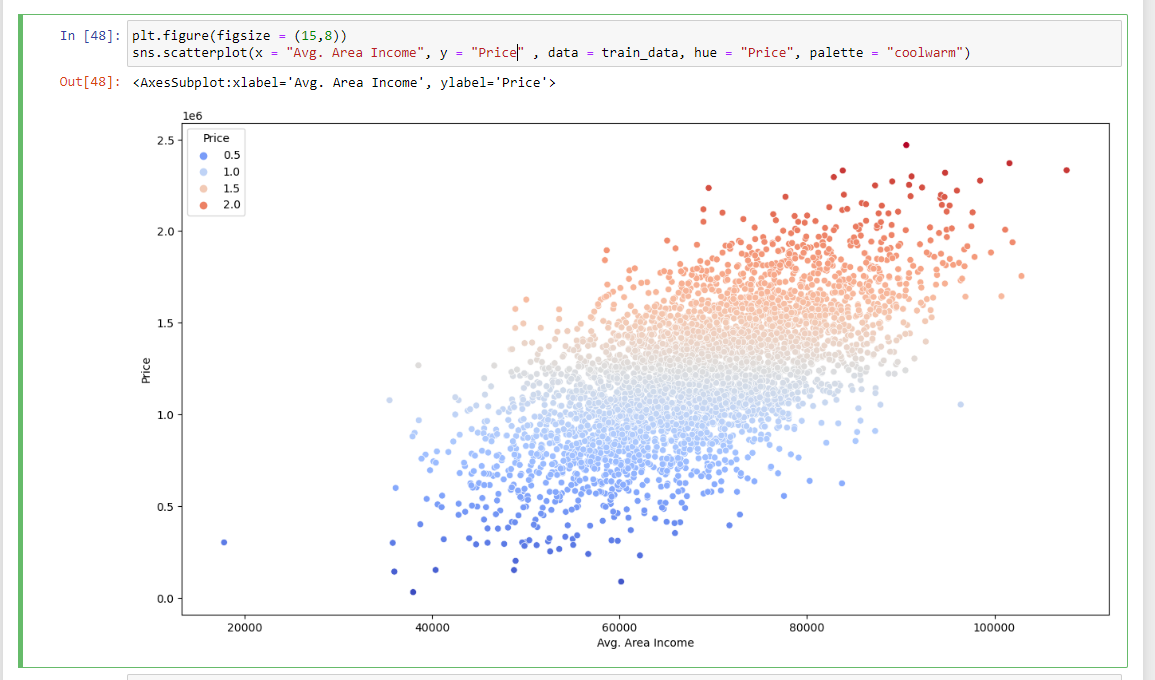
**the data and the target variable “Price”, we pass in the correlation matrix of train\_data as the parameter.**

****

* **we can see that Area population correlates highly with the target variable “Price”**

**Plotting the data to check correlation**

* **Using the scatterplot() function in seaborn we plot the data between Avg. Area Income and Price to find out its correlation.**

****

* **from the scatter plot we can see that the Avg. Area Income plays a huge role in the price value of a house.**

**Conclusion**

**Thus, the dataset was cleaned and preprocessed and the target variable was assessed. We split the dataset into training and testing sets and upon analysis, we found the high correlation between Avg. Area Income and the target variable “Price”. We also found variables that did not play much of a role in assessing the target variable “Price”**