Predicting USA_Housing Prices

Minimum Viable Product (MVP)

By

Fatimah Naif Al-Rashed

To Instructor:

Dr. Essam Al Daoud

❖ USA_Housing Dataset Description

USA_Housing dataset contains 5000 rows. The dataset has 6 columns, and the prices column is the target value. There is non-null in USA_Housing dataset. Also, there is no duplicated rows in USA_Housing dataset as shown in Figure 1 below.

```
data.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5000 entries, 0 to 4999
Data columns (total 7 columns):
 #   Column                        Non-Null Count  Dtype  
---  ------                        --------------  -----  
 0   Avg. Area Income              5000 non-null   float64
 1   Avg. Area House Age           5000 non-null   float64
 2   Avg. Area Number of Rooms     5000 non-null   float64
 3   Avg. Area Number of Bedrooms  5000 non-null   float64
 4   Area Population               5000 non-null   float64
 5   Price                         5000 non-null   float64
 6   Address                       5000 non-null   object 
dtypes: float64(6), object(1)
memory usage: 273.6+ KB
```

Figure 1: USA_Housing dataset info

❖ Exploratory Data Analysis (EDA)

We used Seaborn and Matplotlib for data visualization. We used the seaborn heatmap to represents the collinearity of the multiple features in the dataset. We used data.corr() to show the correlation between the features. We used these methods to find the answer to our questions

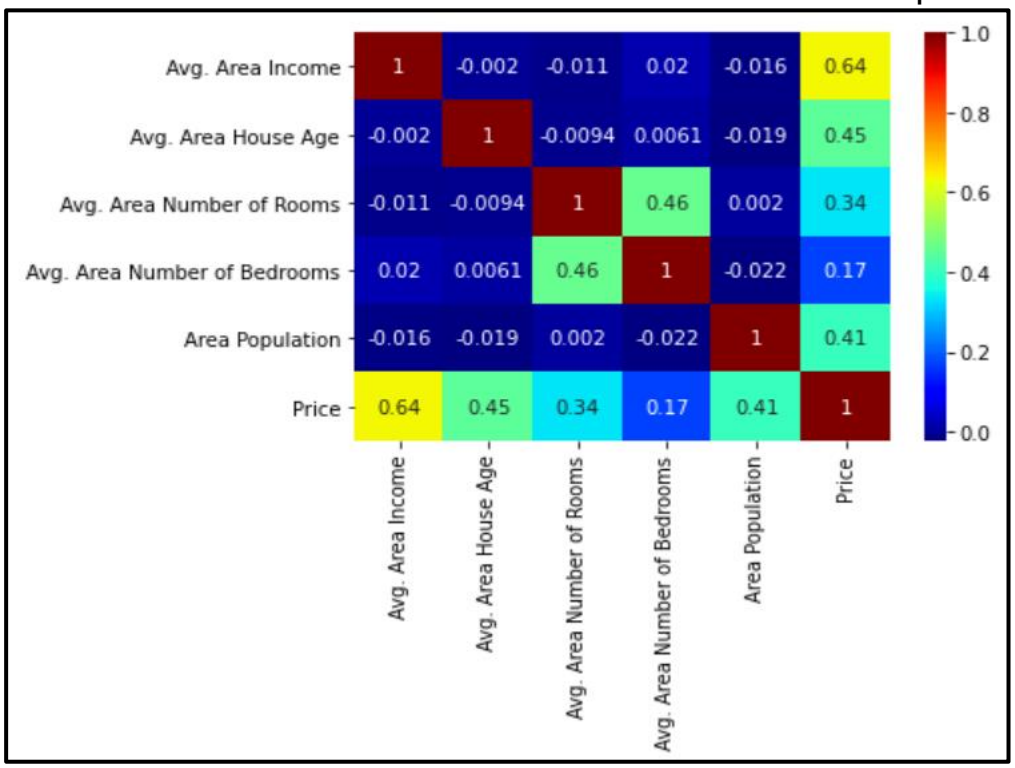

Figure 2: USA_Housing dataset heatmap

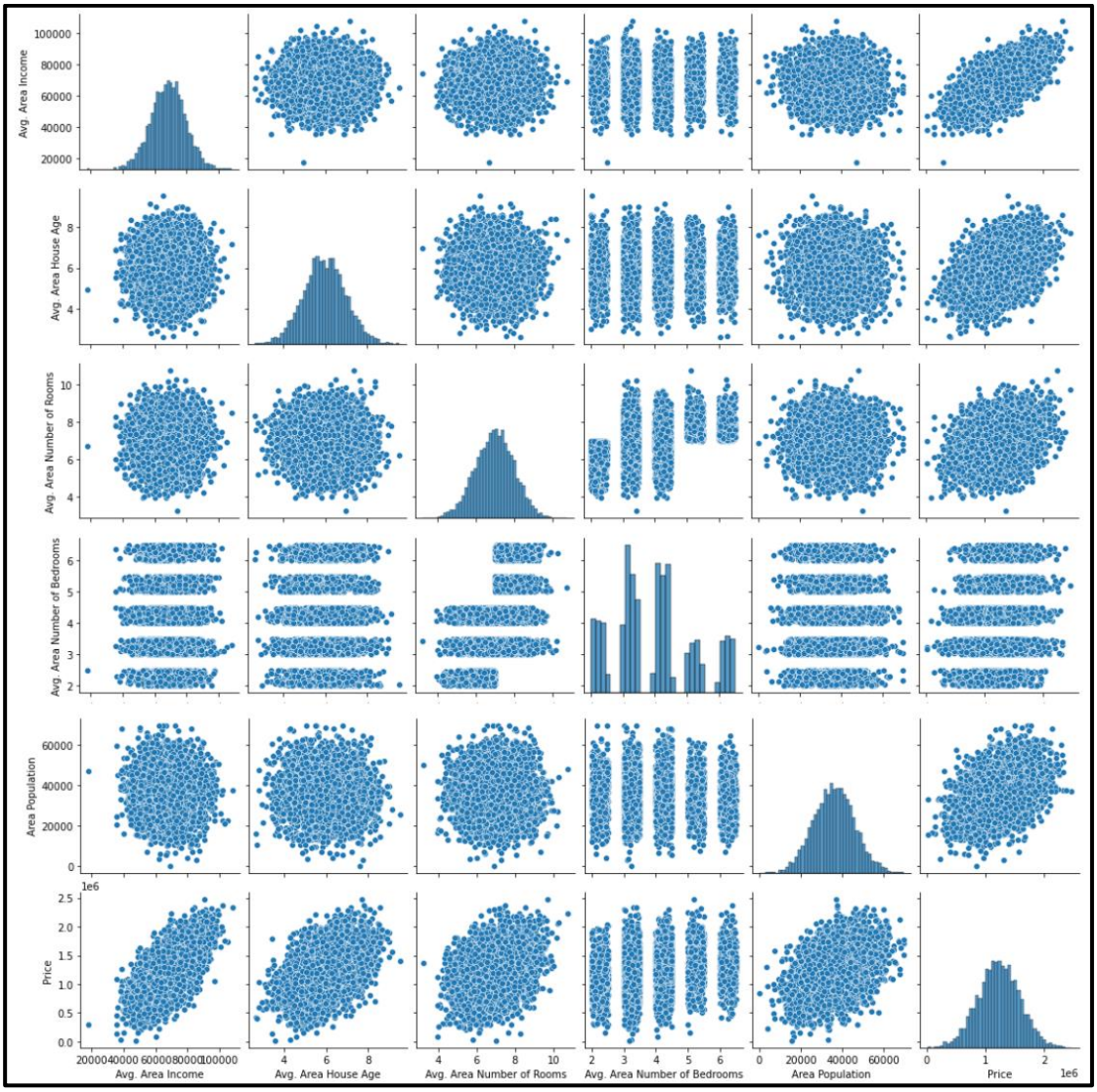

Figure 3: : USA_Housing dataset pair plot

From In Figure 2 and 3, we can see there is no clear relation between 'Avg. Area Number of Bedrooms' and the target 'Price'. The highest correlation among all features is between 'Avg. Area Income' feature and the target 'Price'. Also, from the graph we find the three features 'Avg. Area House Age' 'Avg. Area Number of Rooms' and 'Area Population' had medium correlation with the target 'Price'.

Figure 4, illustrates a strong relationship between 'Avg. Area Income' and 'price'. It shows that the higher in come of individuals the higher the prices of houses are, and vice versa.

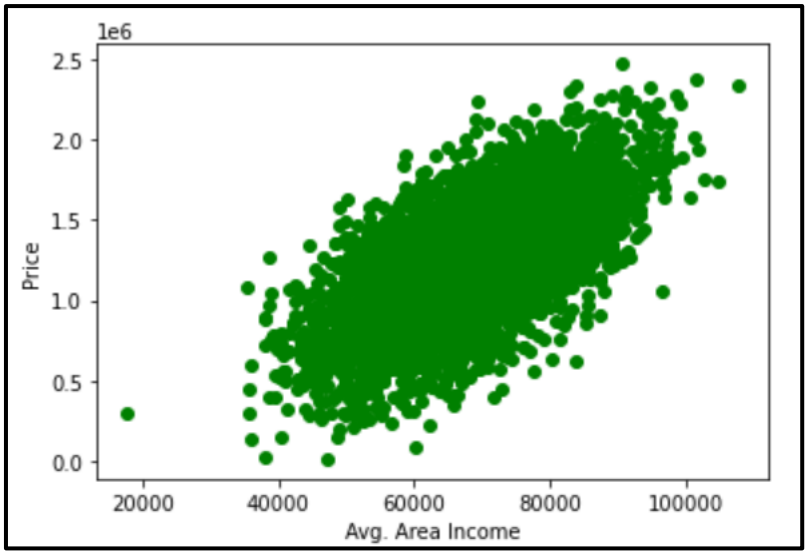

Figure 4: Scatter plots between 'Avg. Area Income' and price

The price has normal distribution as shown in Figure 5 below.

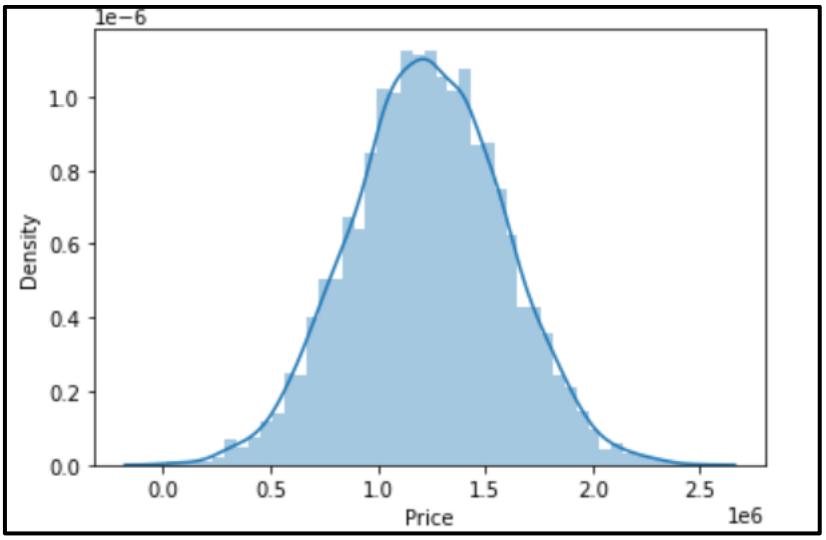

Figure 5: Price distribution plot

❖ Implementation:

We used Python and Jupyter notebooks for implement MVP - Predicting USA_Housing prices. The link for Python notebook is

https://github.com/FatimahNainf/Project/blob/main/MVP%20-%20Predicting%20USA_Housing%20prices.ipynb