Predicting House Prices

By Fatimah AL-Rashed

To Instructor:

Dr. Essam Al Daoud

Outline

- Problem Statement
- Data Set
- Exploratory Data Analysis
- Algorithms
- Experimental And Result
- Conclusion

Problem Statement

- One of the most Machine Learning problems is predicting house prices.

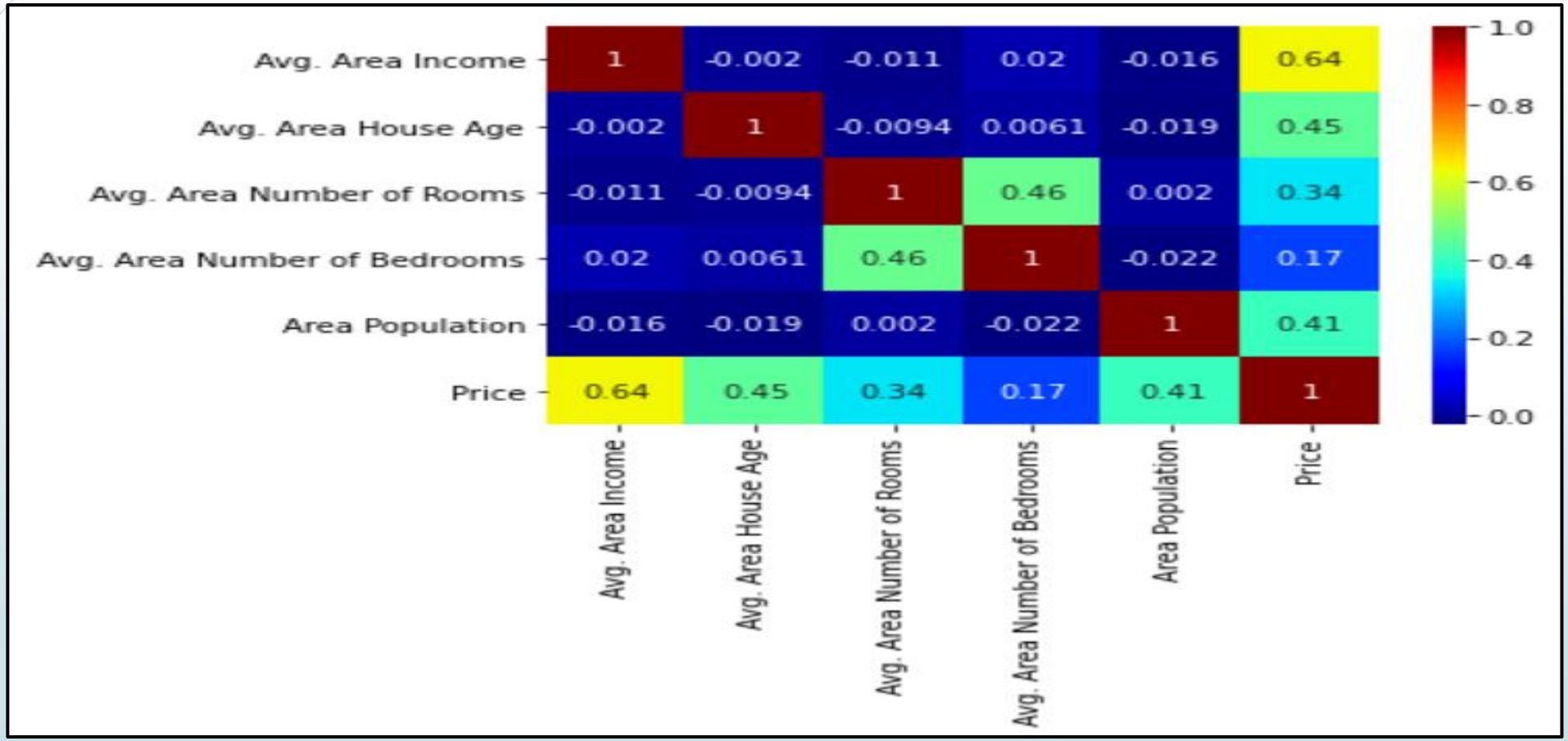
Data Set

We are going to use the USA_Housing dataset

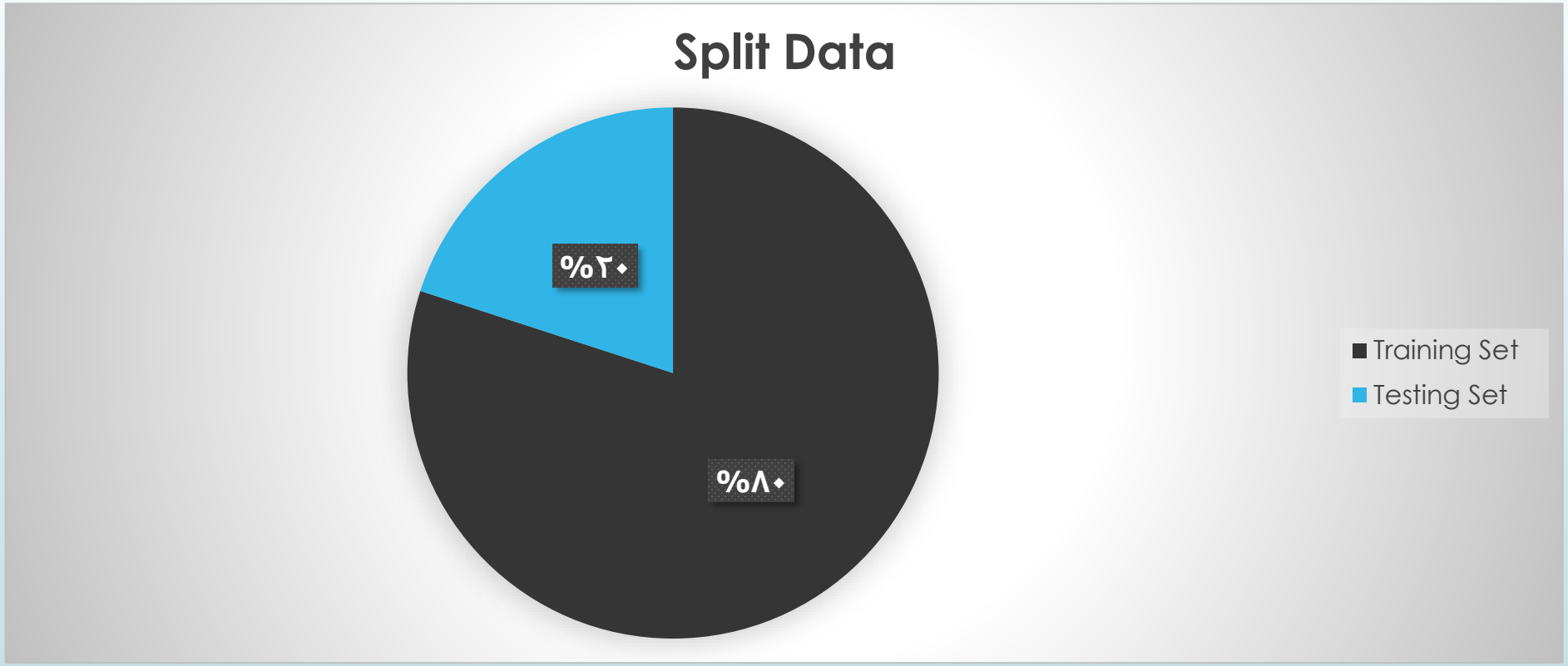
```
data.info()
```

```
<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
```

Exploratory Data Analysis



Data Set Split



Cross Validation

Algorithms	R2 Square
Linear Regression	0,9172
Ridge	0,9172
Lasso	0,9172

First Experiment

- We used all the features in the data set for X
- The features are Avg. Area Income, Avg. Area House Age ,Avg. Area Number of Rooms, Avg. Area Number of Bedrooms, and Area Population
- The Test Predicted Result :

Algorithms	R2 Square
Linear Regression	0,9179

Second Experiment

- We have excluded Avg. Area Number of Bedrooms from features
- The features are Avg. Area Income, Avg. Area House Age ,Avg. Area Number of Rooms, and Area Population
- The Test Predicted Result :

Algorithms	R2 Square
Linear Regression	0,9181

Third and Fourth Experiments

- In the third experiment, the feature is Avg. Area Income
- In the fourth experiment, the features are Avg. Area Income and Avg. Area House Age
- We got the same Test Predicted Result :

Algorithms	R2 Square
Linear Regression	0,3969

Conclusion

- We find the best performing model for Predicting House Prices are Linear Regression , Ridge and Lasso
- We can exclude Avg. Area Number of Bedrooms from features
- We should use all the feature exclude Avg. Area Number of Bedrooms from features to get more acuter data model

Any Questions