Features (independent Variables):

LSTAT: This is the percentage lower status of the population

PTRATIO: This is the pupil-teacher ratio by town

RM: This is the average number of rooms per dwelling

Target variable (Dependent Variable):

MEDV: This is the median value of owner-occupied homes in $1000s

The **Exploratory data analysis** is where I looked at the data and found some insights, for example the mean of the prices, max price, min price etc. In addition to this I also look at how the features (‘RM’, ‘LSTAT’, ‘PTRATIO’) correlate with the dependent variable Price (‘MEDV’). In order to do this, I built a heatmap as well as a pair plot.

The R2 score to evaluate our different regression model. The R2 score ranges from -1 to 1, Where -1 is the worst and 1 is the ideal model. I decided to use 5 model and compared the scores of each of our models to select the best one. I have created a separate python notebook to evaluate each of these models.

Multiple Linear Regression:

R2-score: 0.6574622113312862

Polynomial Regression:

R2 Score: 0.7752997053051813

Random Forest Regression:

R2 score: 0.7958582987098185

Support Vector Regression:

R2 score: 0.787488569711881

Decision Tree Regression:

0.6714474055340578

From all the models we can see that Random Forest Regression has the highest R2 score which shows us that it is the best model for Boston housing predictive pricing project.