# ****NAME- Piyush Sanjay More****

# ****TOPIC-BOSTON HOUSE PREDICTION****

# ****Introduction****

In this project, we will develop and evaluate the performance and the predictive power of a model trained and tested on data collected from houses in Boston’s suburbs. Once we get a good fit, we will use this model to predict the monetary value of a house located at the Boston’s area. A model like this would be very valuable for a real state agent who could make use of the information provided in a daily basis.

Dataset Description

This data represents aggregate information about 14 features of homes from various suburbs located in Boston.

The features can be summarized as follows:

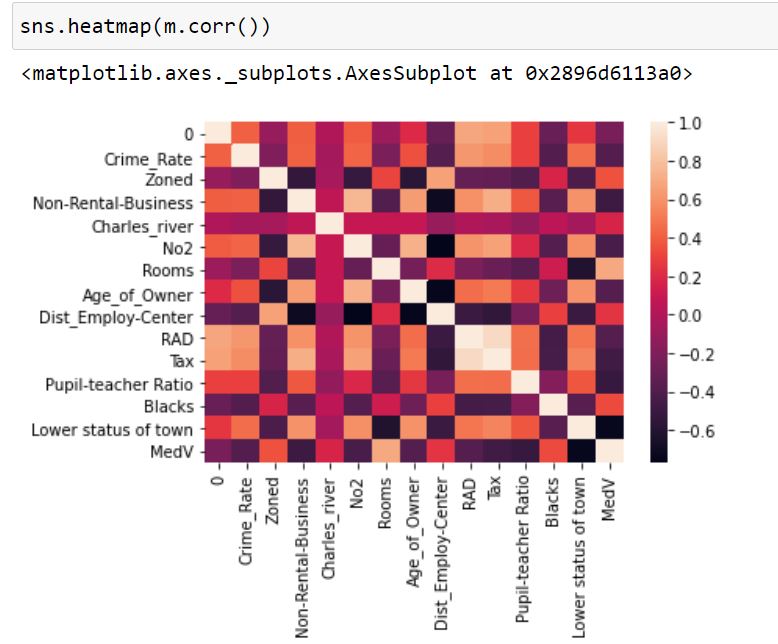
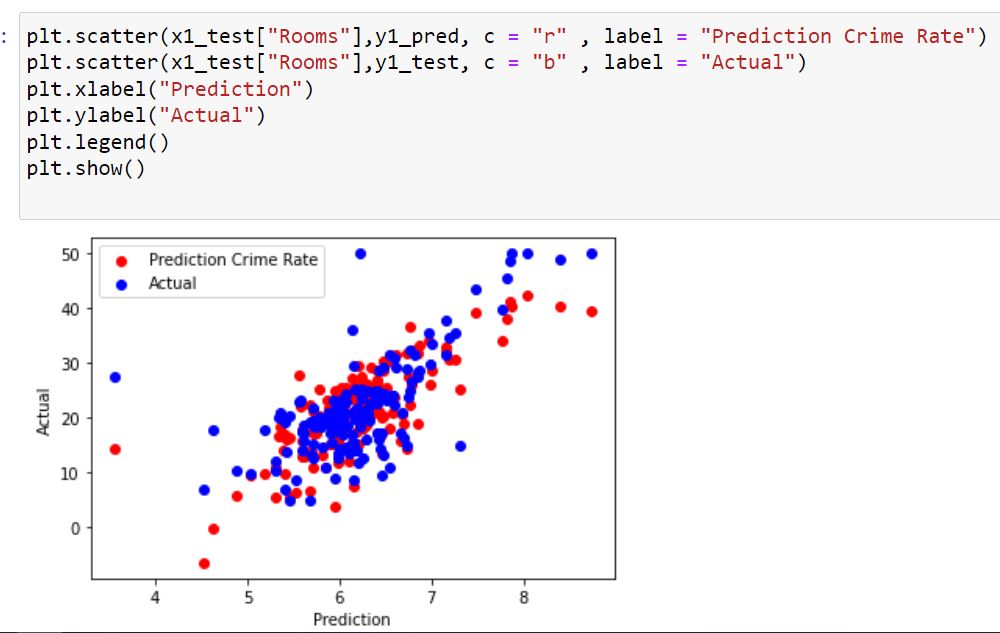
* CRIM: This is the per capita crime rate by town
* ZN: This is the proportion of residential land zoned for lots larger than 25,000 sq.ft.
* INDUS: This is the proportion of non-retail business acres per town.
* CHAS: This is the Charles River dummy variable (this is equal to 1 if tract bounds river; 0 otherwise)
* NOX: This is the nitric oxides concentration (parts per 10 million)
* RM: This is the average number of rooms per dwelling
* AGE: This is the proportion of owner-occupied units built prior to 1940
* DIS: This is the weighted distances to five Boston employment centers
* RAD: This is the index of accessibility to radial highways
* TAX: This is the full-value property-tax rate per $10,000
* PTRATIO: This is the pupil-teacher ratio by town
* B: This is calculated as 1000(Bk — 0.63)², where Bk is the proportion of people of African American descent by town
* LSTAT: This is the percentage lower status of the population
* MEDV: This is the median value of owner-occupied homes in $1000s

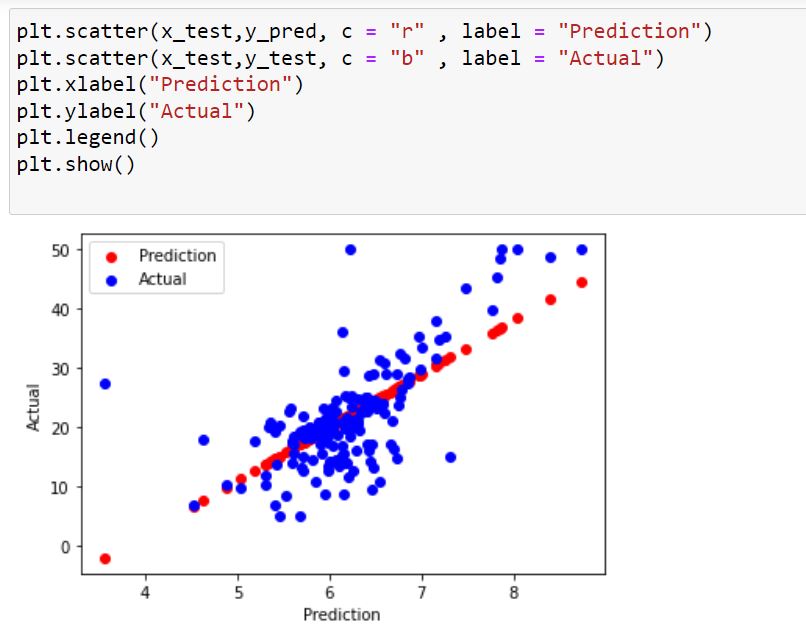
Approach

We started proceeding with reading with data and we found that various factors that affected the Boston house cost were rad, ptratio, lstat .So our target variable was mdev . Then we started with processing the data in which we preformed data cleaning, data description and found that the data was having no null values The maximum cost was 6377000 .We started visualizing 3 factors .that is ptratio ,istat ,mdev and found that highest house charges . We converted categories into numerical form using label encoder.

We plotted heat map using seaborn for the data core relation and found that prtatio and Mdev were highly co-related with 0.67 co-relation .We were having train and test dataset separate, so there was no need to split the data .Then we stated training models multiple linear regression , linear regression ,polynomial regression.

**Visualization**

****



Algorithm

The major aim of in this project is to predict the house prices based on the features using some of the regression techniques and algorithms.

**1.** **Linear Regression**

**2.** **Multiple Linear Regression**

**3. Polynomial Linear Regression**

**Result and Discussion**

An important strength of this study is that we combined both a systematic literature review and a head-to-head empirical evaluation of different supervised learning methods reported in the literature. An additional strength is that we evaluated state-of-the-art supervised learning methods not previously evaluated in the literature for cost on cost prediction in Boston’s House.

**Conclusion**

Throughout this I made a machine learning regression project from end-to-end and I learned and obtained several insights about regression models and how they are developed.

**Difficulties faced**

Difficulty faced during choosing a model, analysing various factors .

Reference

Website

* https://scikit- learn.org/stable/modules/generated/sklearn.model\_selection.train\_test\_split.html
* <https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html>