

**NAME OF THE PROJECT**

HOUSING: PRICE PREDICTION

Submitted by:

Homa Firdaus

**ACKNOWLEDGMENT**

I would like to convey my heartfelt gratitude to Mr. [**Shwetank Mishra**](https://www.flipnwork.com/index.php/team_members/view/2046) for his tremendous support and assistance in the completion of my project. I would also like to thank whole team of Flip Robo for providing me with this wonderful opportunity to work on a project with the topic HOUSING: PRICE PREDICTION. The completion of the project would not have been possible without their help and insights.

**INTRODUCTION**

* Business Problem Framing
* Required to build a model using Machine Learning in order to predict the actual value of the prospective properties and decide whether to invest in them or not.
* Conceptual Background of the Domain Problem
* Housing and real estate market is one of the markets which is one of the major contributors in the world’s economy.
* A US-based housing company named Surprise Housing has decided to enter the Australian market. The company uses data analytics to purchase houses at a price below their actual values and flip them at a higher price. For the same purpose, the company has collected a data set from the sale of houses in Australia. Predictive modelling, Market mix modelling, recommendation systems are some of the machine learning techniques used for achieving the business goals for housing companies. Our problem is related to one such housing company.

Review of Literature

* This is a comprehensive summary of the research done on the housing project. The review should enumerate, describe, summarize, evaluate and clarify the research done.
* Motivation for the Problem Undertaken

This project is done to build a model to prediction weather to invest on particular Housing project or not. Different feature have been given on the basis of which decision need to be taken.

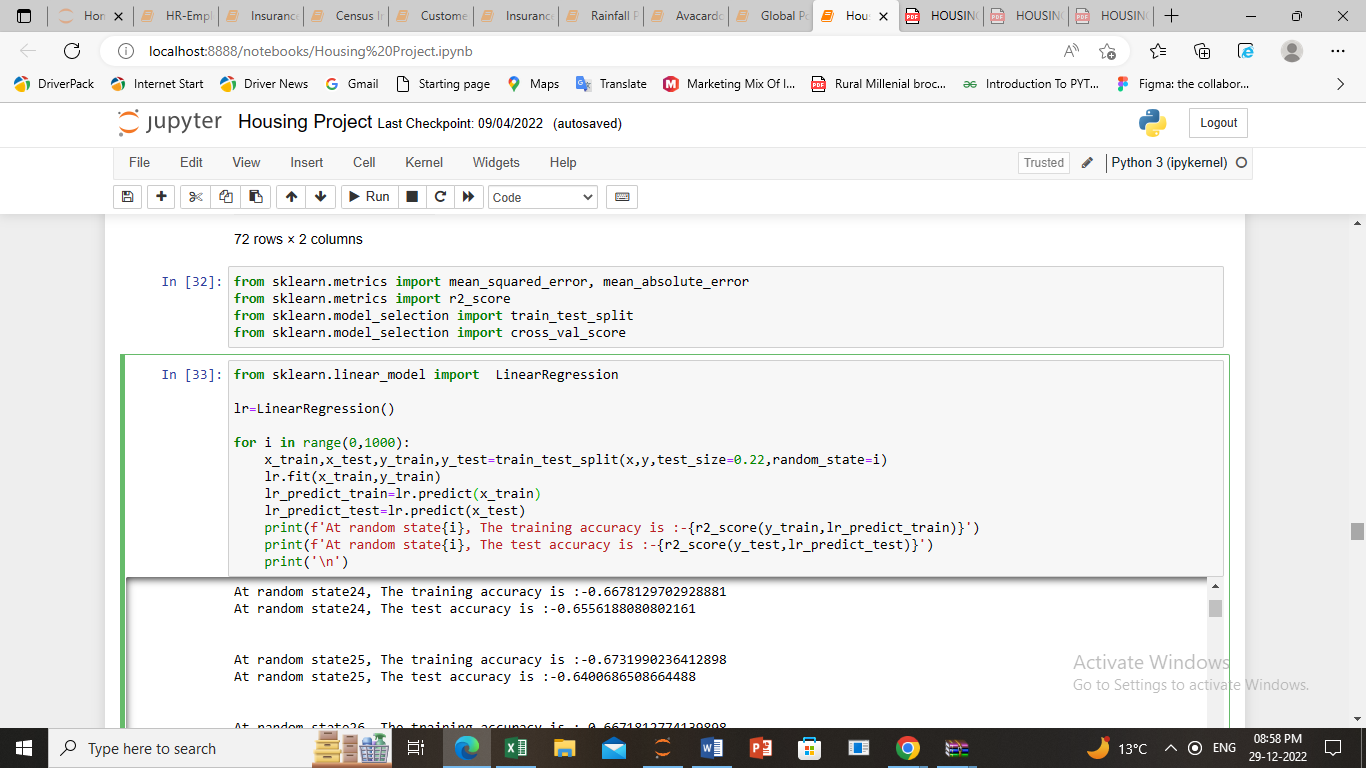
Analytical Problem Framing

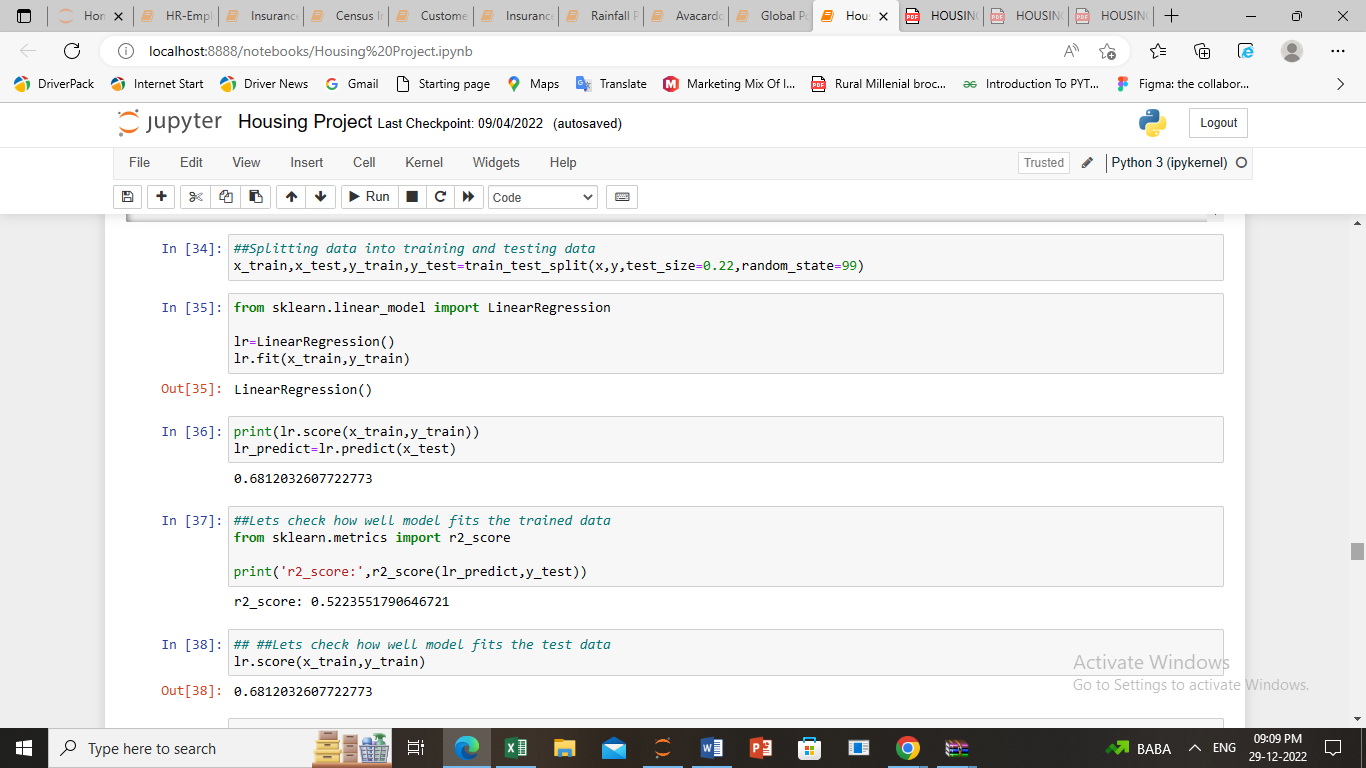
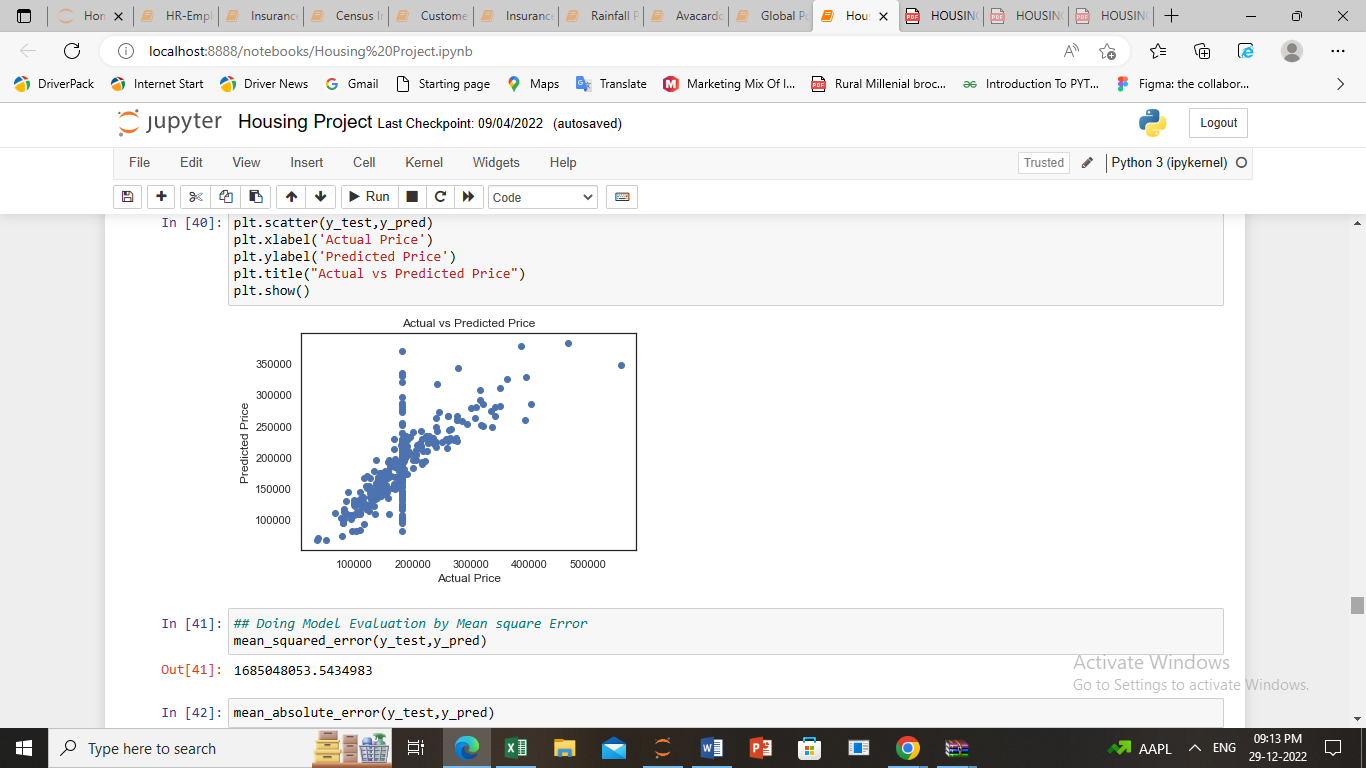
Mathematical/ Analytical Modelling of the Problem

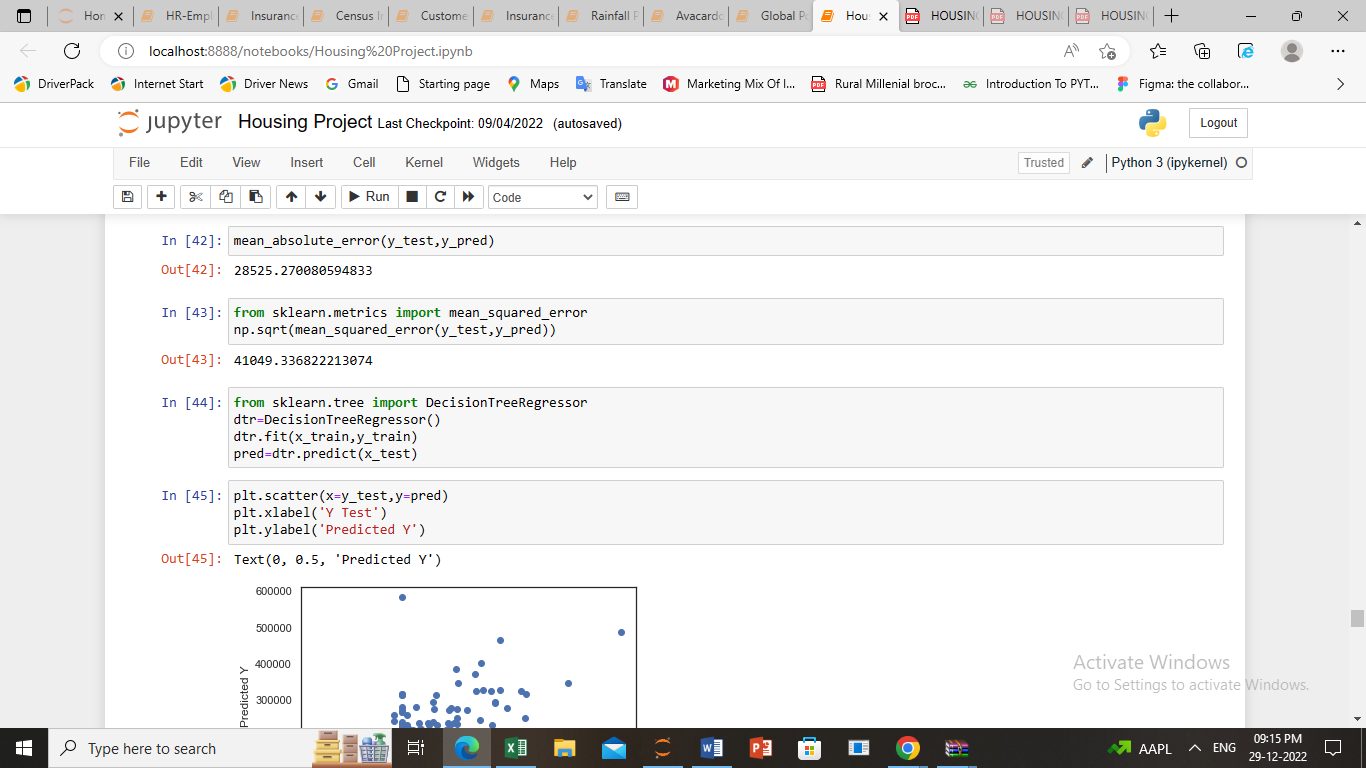
In the data set given I there were null values available, data was skewed and outliers were also available. I tried to treat the data as per the requirement.

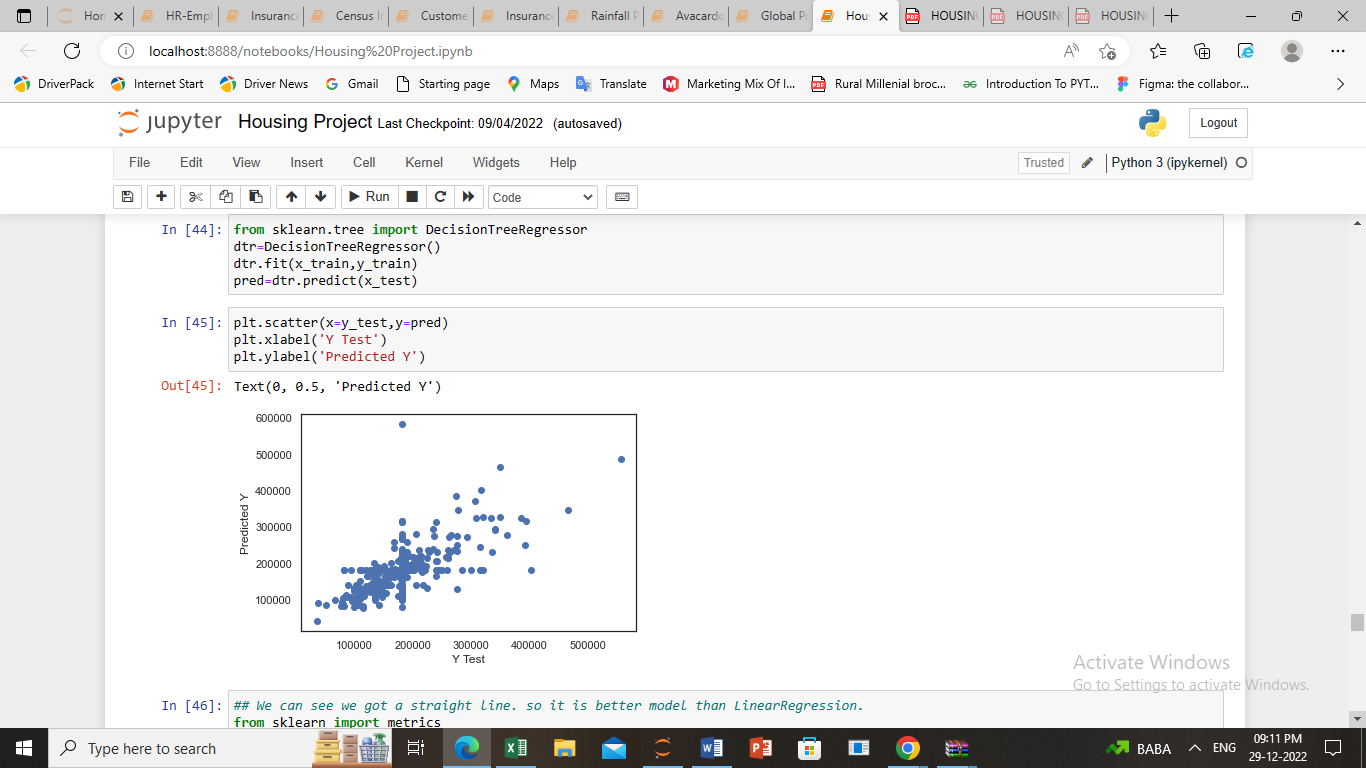
* Data Sources and their formats Data are available in csv format in the form of test and train data.
* Data Pre-processing Done
* Acquire the dataset
* Import all the crucial libraries
* Import the dataset
* Identifying and handling the missing values
* Removing skew ness and Outliers
* Encoding the categorical data
* Splitting the dataset
* Feature scaling
* Data Inputs- Logic- Output Relationships
* In this model we have seen correlation of different feature with target variable as well as seen importance of different feature for Sales prediction.
* **Model/s Development and Evaluation**
* I have used pd. concat method to merge test and train data.
* Df.drop is used to delete those columns which have null values more than 50%.
* Treated null data by filling it using mean and mode.
* I have used sqrt transformation method to remove skewness.
* Worked with sklearn. Pre-processing import Ordinal Encoder.
* **Testing of Identified Approaches (Algorithms)**

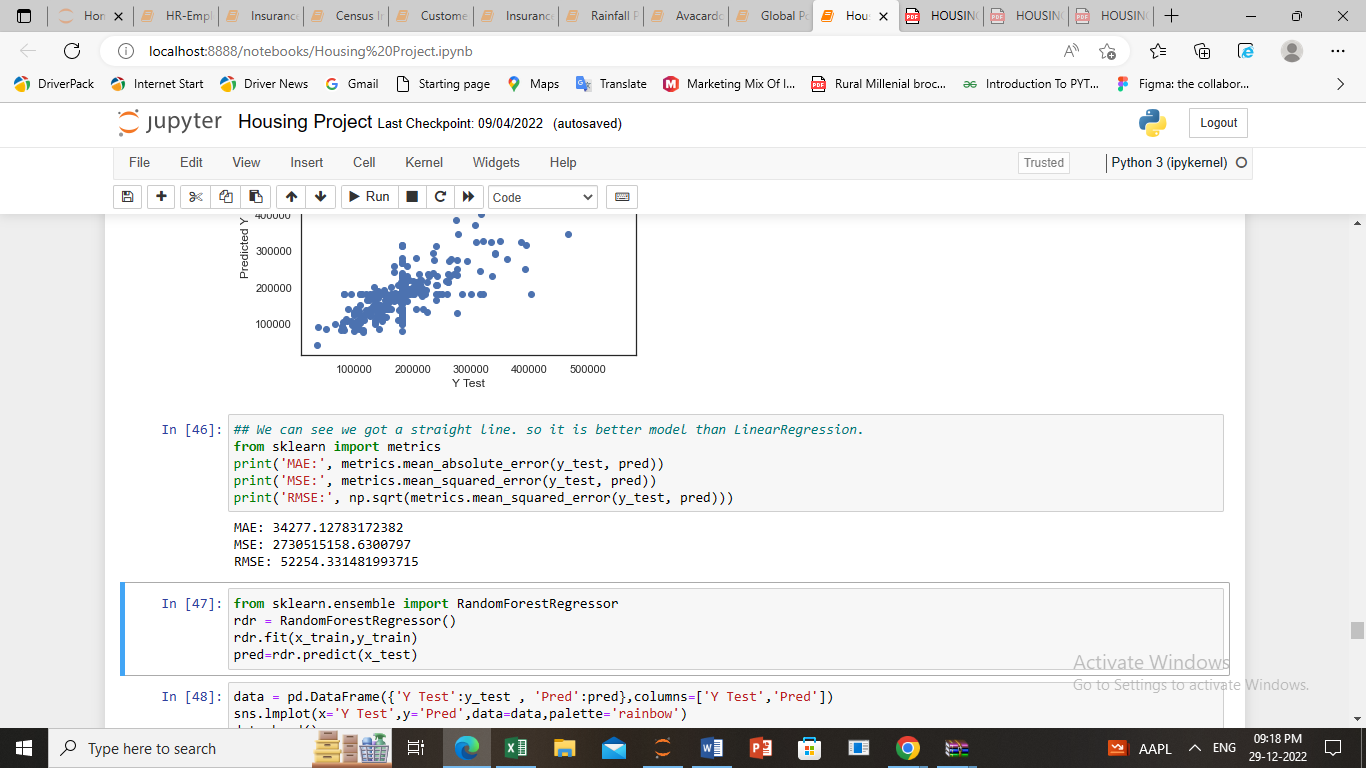
* Used train-test-split method to divide data for training model and testing it

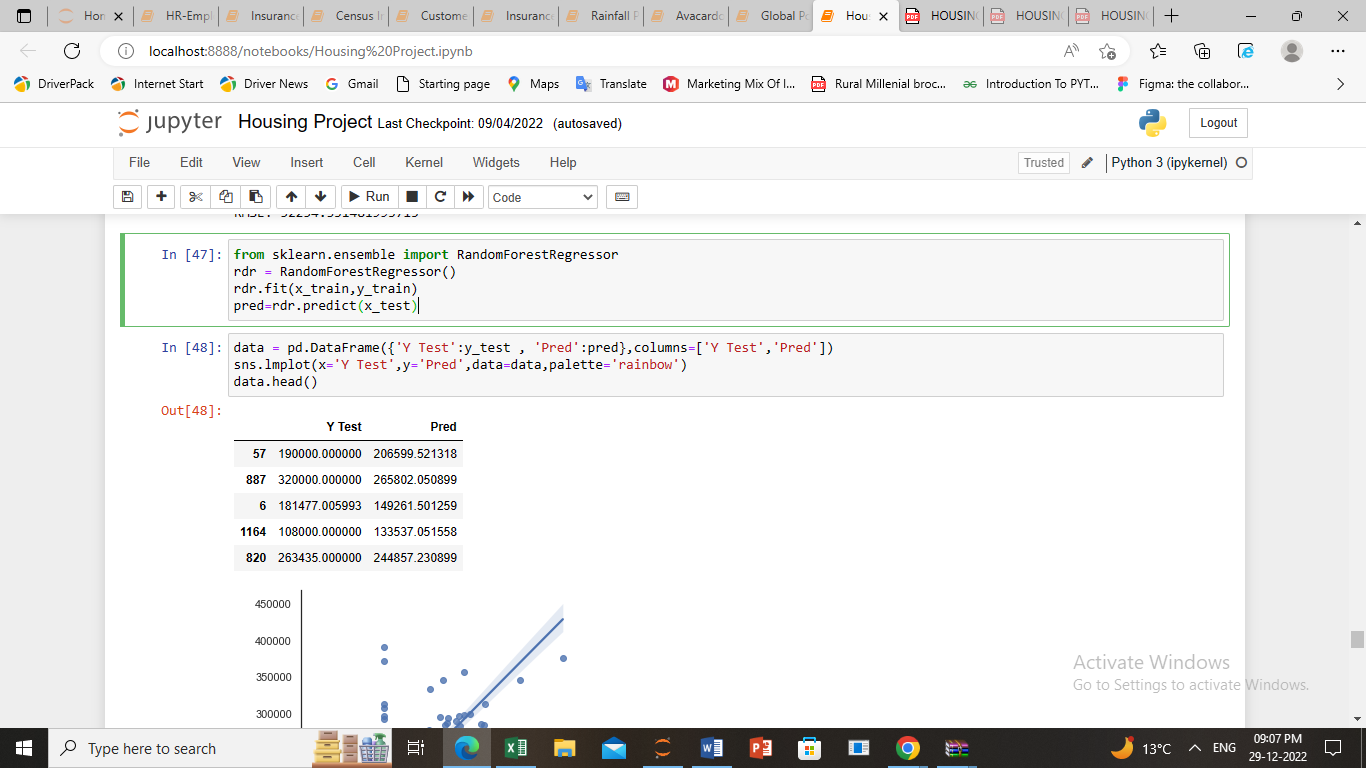
. 

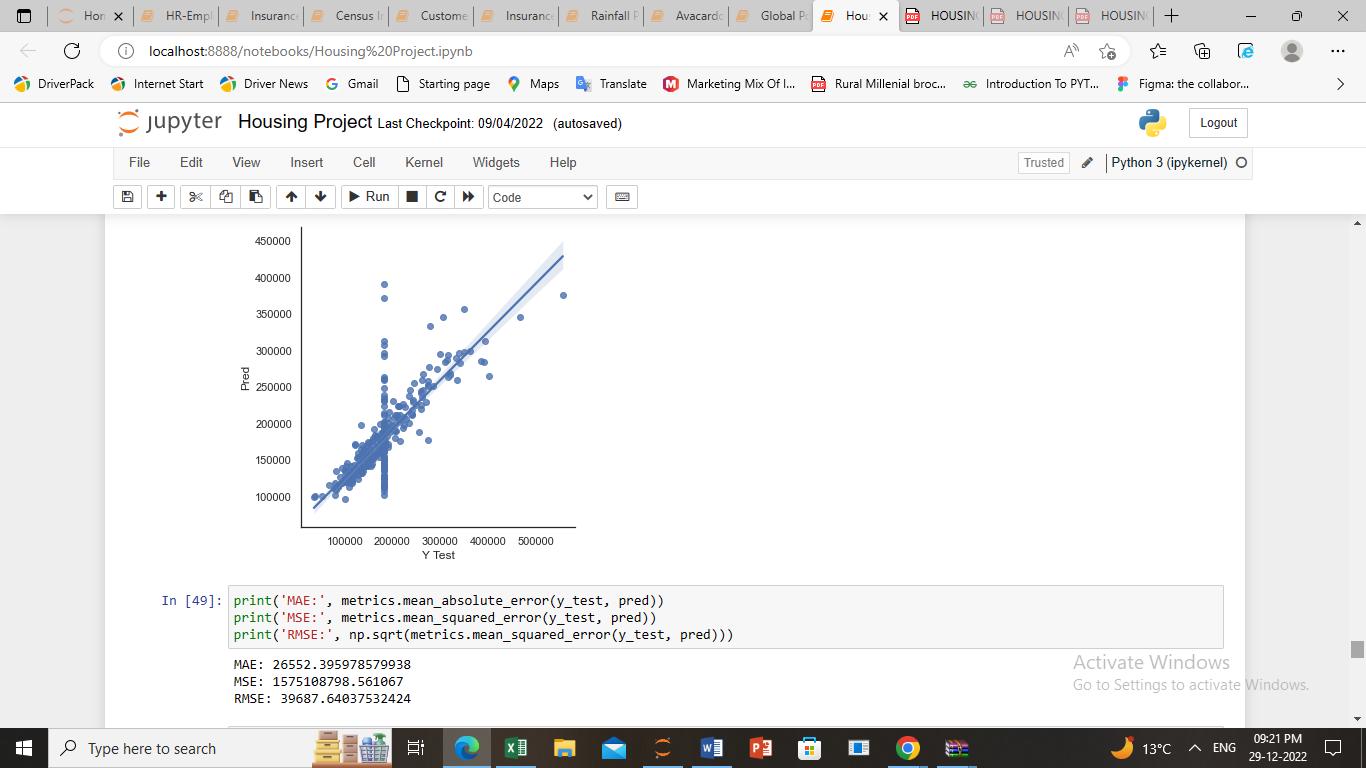
* Run and Evaluate selected models
* 
* 











* Visualizations
* I have done scatterplot visualization to see relation of different features with Target variable, to visualize skewness I have done sub-plotting and box-plotting to visualize outliers.
* **Interpretation of the Results**
* I came to know through analysis which model will be work with better accuracy with the help of low residual and RMSE scores.
* In this project we can see that min RMSE is of Random Forest hence this can be a good model option

**CONCLUSION**

* Key Findings and Conclusions of the Study
* On the basis of above project we can see Random Forest has highest Accuracy value.
* Also get to know where to use Linear, Decision Tree and other applicable and required models to fine tune the predictions.
* Learning Outcomes of the Study in respect of Data Science
* I have done EDA using Pandas and other plotting libraries.
* Also made use of packages like matplotlib, plotly and seaborn to develop better insights about the data.
* I also learnt how to create new features which will in turn help us to better predict the survival.
* I get to know what are the features on which price is highly positively and negatively correlated with.
* I came to know through analysis which model will be work with better accuracy with the help of low residual and RMSE scores.
* In this project we can see that min RMSE is of Random Forest hence this can be a good model option
* This project helped me to understand which model to choose first and go step by step to attain results with good accuracy.
* In this project we have also seen importance of each features in ascending order.