

Housing Price Prediction

Submitted by:

YOUR NAME

**ACKNOWLEDGMENT**

This is price prediction project. I got data from Flip Robo in two form Train and Test Csv file. I take help Miss Gulshana Chaudhary .she help me at every time when i am stuck in any phase.

**INTRODUCTION**

* Business Problem Framing

We all have experienced a time when we have to look up for a new house to buy. But then the journey begins with a lot of frauds, negotiating deals, researching the local areas and so on.

* Conceptual Background of the Domain Problem

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

* Review of Literature

Here i cover some topic which is use full for understanding the data

* Data understanding and exploration
* Data cleaning
* Data preparation
* Model building and evaluation
* Observation and inference
* Motivation for the Problem Undertaken

I believe this model could be optimized and tuned more to add accuracy either by adding new features or engineering new features. This model can be used to predict the house prices in any geographic location by just slightly fine tuning the features and parameters.

**Analytical Problem Framing**

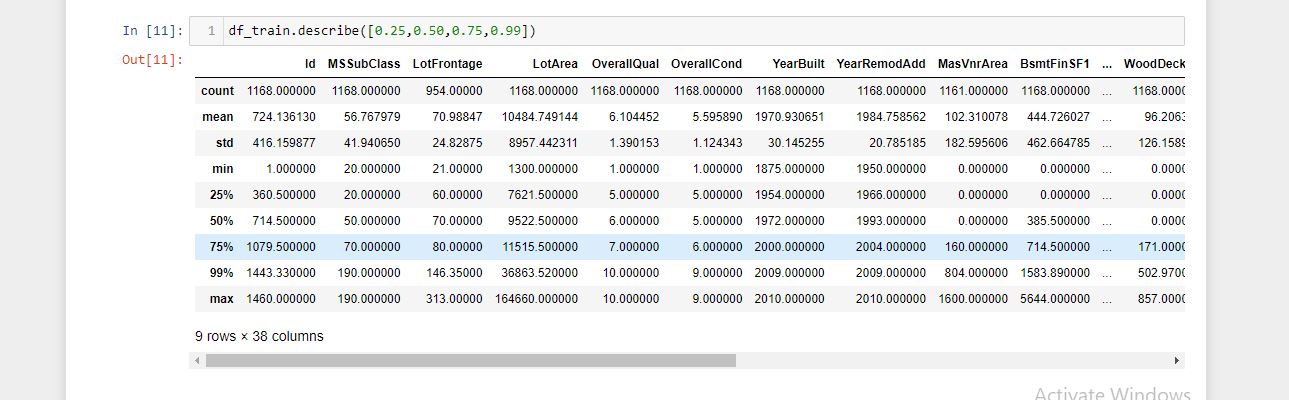
* Mathematical/ Analytical Modeling of the Proble

Mathematical modelling---here i check the shape of both data i.e Train and Test.checking the missing values by considering the 10% or >10%.

Statistical modelling—we apply the EDA step by ploting the map.diffrent type plot gives correlation b/w target and features..

* Data Sources and their formats

Here data sorce is CSV file which is Zip format ,firstly import all the necessary liberaries after that pd.read-csv file.here some data have missing values, it will fill by fillna method.df.info() gives all the detail about dataset like type, rows , column.



* Data Preprocessing Done

1. first check the how many null values are present, then it should fill appliying fillna method.

2. then drop the unwanted column which is not useful at any stage.

3.Apply Encoding Technique to spread all data in same manner like integer.

After that we apply the EDA step..

* Data Inputs- Logic- Output Relationships

This is Regression Problem because my target column has continoues value.here i have to predict the sale price of house.this is dependent Features, and independent features has two type column one has continous and other is categorical features.

Some features does’t affect the target but some has value to affect this like “overall quality” .

* State the set of assumptions (if any) related to the problem under consideration

No any assumpution i put.

* Hardware and Software Requirements and Tools Used

Hardware---64 bit, windows 10 and other advanced version should use for implement this.

Software---firstly install anaconda to find the jupitor notebook.this is tool for optimization the result.

Sklearn tool use for importing the liabraries.here pandas and numpy are use to import the dataset.

**Model/s Development and Evaluation**

* Identification of possible problem-solving approaches (methods)

I use EDA technique to find the relation b/w target and features.find skewness of the data .find outliers and remove them for normalized data.heat map and line plot show the relation .

* Testing of Identified Approaches (Algorithms)

Listing down all the algorithms used for the training and testing.

Linear Regression Model

Ada Boost Regressor

Decision Tree Regressor

* Run and Evaluate selected models

Linear regression model-

RMSE Score is: 37648.687373240515

R2 Score is: 79.00415420824845

Cross Validation Score: 83.4802599461583

R2 Score - Cross Validation Score is -4.4761057379098474

Ada Boost Regressior—

RMSE Score is: 38182.53301059012

R2 Score is: 78.40450475848

Cross Validation Score: 79.07776918647077

R2 Score - Cross Validation Score is -0.6732644279907731

Decision Tree Regressior

RMSE Score is: 42387.89159892613

R2 Score is: 73.38556038423256

Cross Validation Score: 68.72185658843557

R2 Score - Cross Validation Score is 4.663703795796991

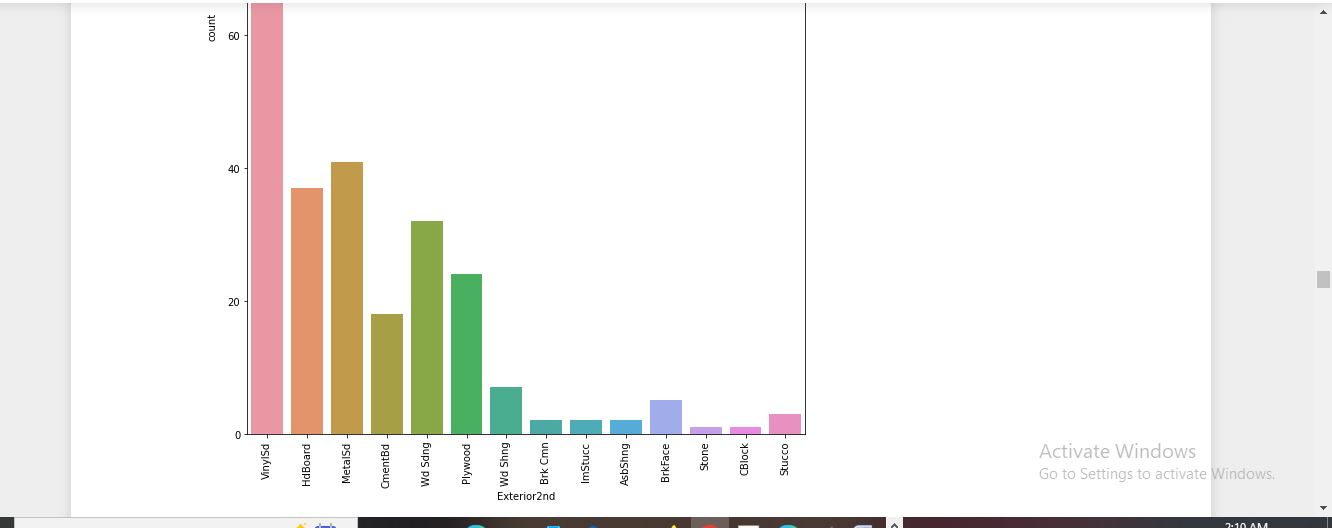
* Key Metrics for success in solving problem under consideration

1...accuracy\_score

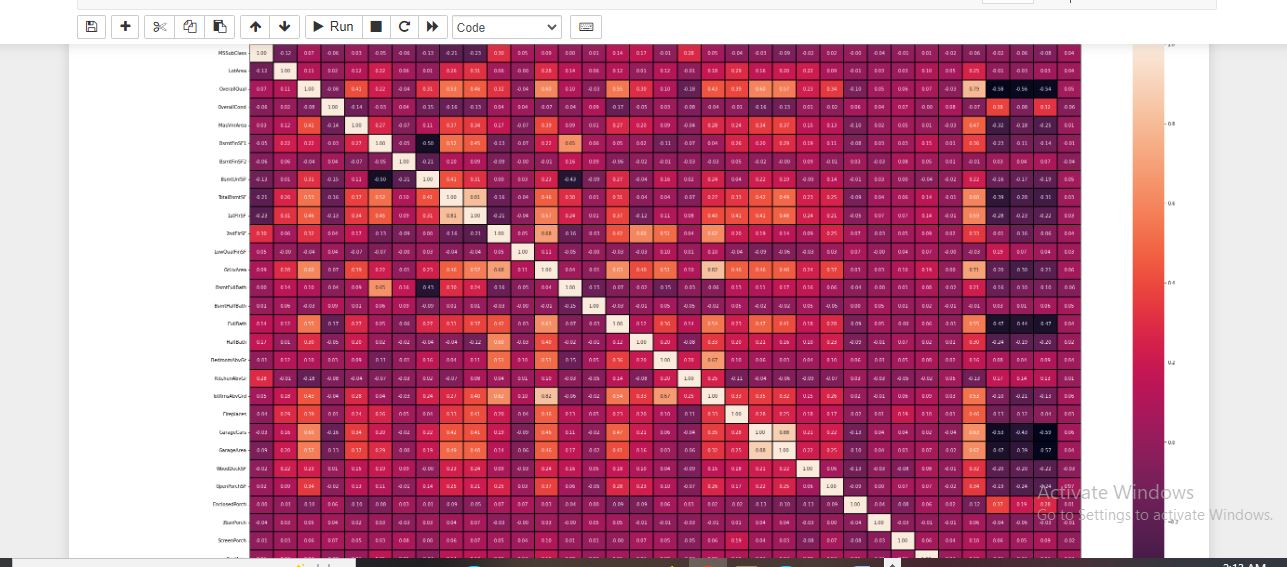
2...r2\_score,mean\_squared\_error

* Visualizations

1....I use count plot to visualize the distribution



2...heat map show coreelation



* Interpretation of the Results

Linear regression gives high accuracy for this prediction.

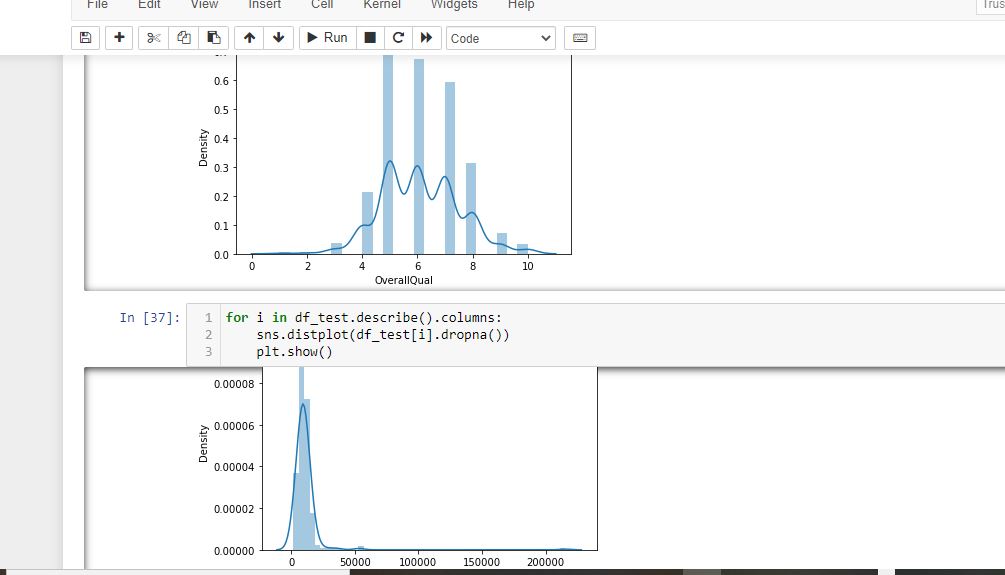
**CONCLUSION**

* Key Findings and Conclusions of the Study

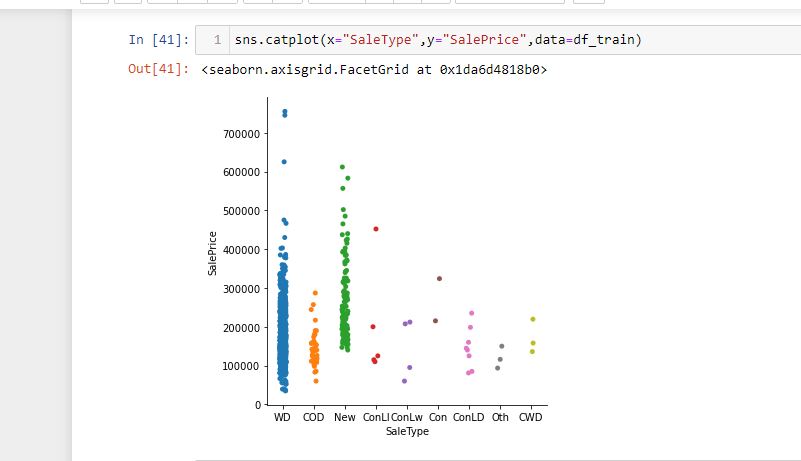
In this data i observe that price will be high when house is too old but if it’s new it give less price .

* Learning Outcomes of the Study in respect of Data Science

Here i apply some visualization technique to find the distribution and relation..



Distplot to find the distribution normalized or not..



Catplot to find how target depend the feature..



Scatter plot to linear relation b/w target and features.

In all algorithm i find Linear regression is best to give good accuracy.

The END