

NAME OF THE PROJECT

**Project Housing Prediction**

Submitted by:

**Tabassum Khan**

**ACKNOWLEDGMENT**

I have taken reference from different platforms from the Google and have learn some of the code ,learned about the data and then I have done the analyzation about the data.

**INTRODUCTION**

* Business Problem Framing

Predicting the home prices and features of the home .Every one in the real world have different dreams about their home but all that also depend on the budget and how much they can afford to buy their dream house.

* Review of Literature

The goal of this project was to use EDA, visualization, data cleaning, pre processing, and linear models to predict home prices given the features of the home, and interpret your linear models to find out what features add value to a home.

* Motivation for the Problem Undertaken

For improving skills to analyze, predict, and convey data-driven facts from large datasets.

**Analytical Problem Framing**

* Mathematical/ Analytical Modeling of the Problem

For accuracy improvement I have used hyperparameter tunning,cross validation,xgboost to get the better accuracy from our data.

* Data Sources and their formats

The variables most correlated with 'SalePrice':

1.'OverallQual', 'GrLivArea' and 'TotalBsmtSF' are strongly correlated with 'SalePrice'.

2.'GarageCars' and 'GarageArea' are also some of the most strongly correlated variables.

3.'GarageCars' and 'GarageArea' are alike. we will never be able to distinguish them. Therefore, we just need one of these variables in our analysis and we can keep 'GarageCars' since its correlation with 'SalePrice' is higher.

4.'TotalBsmtSF' and '1stFloor' also seem to be same.

* Data Preprocessing Done

Firstly, I have found all the null values from the data and the variables which contained more than 50% of the null values dump all that and which contained only 10% to 15% of null values to that variables I have filled with help of median or mode

* Hardware and Software Requirements and Tools Used-

Jupyter notebook, Python, Github.

**Model/s Development and Evaluation**

* Identification of possible problem-solving approaches (methods)

This project required a lot of data cleaning, feature selection. Extratreesregressor is used for selecting numerical attributes. Extratressclassifier is used for selecting categorical attributes. Linear Support Vector Regressor is used to predict the values.

* Testing of Identified Approaches (Algorithms)

hyperparameter tunning, cross validation, xgboost

* Run and Evaluate selected models

Describe all the algorithms used along with the snapshot of their code and what were the results observed over different evaluation metrics.

**CONCLUSION**

By using various range of analytical and graphical model and tools we were able to find out various range of prices in this House price prediction data set.

By using the model we were able to find out the best suitable prices of the house and the different variations which has been see while analysing the dataset.