
# HOUSE PRICE PREDICTION PROJECT

Submitted by:

LOKESH BISEN

# ACKNOWLEDGMENT

Estimating the sale prices of houses is one of the basic projects to have in our Data Science.Here we will be able to predict continuous variables using various types of regression algorithms.

# INTRODUCTION

- Business Problem Framing

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. The data is provided in the CSV file below. The company is looking at prospective properties to buy houses to enter the market. You are 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. For this company wants to know:

- Which variables are important to predict the price of variable?
- How do these variables describe the price of the house?

- Motivation for the Problem Undertaken

AbstractHouse price forecasting is an important topic of real estate. The literature attempts toderive useful knowledge from historical data of property markets. Machine learningtechniques are applied to analyze historical property transactions in India to discoveruseful models for house buyers and sellers. Revealed is the high discrepancy

between house prices in the most expensive and most affordable suburbs in the Australian market.

- Review of Literature

As I said before, we are going to work with the house price data-set that contains various features and information about the house and its sale price. And sale price is continuous variables hence the problems seems regression. Data set contains 81 columns and 1168 rows. There are many columns which has
Categorical data which we have to convert into numeric in order to feed to our ML model. As per statistical reviews of data set, I found there are many null values and outliers present.

# Analytical Problem Framing

- Data Preprocessing Done

Data exploration and preprocessing is the first step in data analysis and typically involves summarizing the main characteristics of a data set, including its size, accuracy, initial patterns in the data and other attributes. It is commonly conducted by data analysts using visual analytics tools, but it can also be done in more advanced statistical software, Python. Before it can conduct analysis on data collected by multiple data sources and stored in data warehouses, an organization must know how many cases are in a data set, what variables are included, how many missing values there are and what general hypotheses the data is likely to support. An initialexploration of the data set can help answer these questions by familiarizing analysts with thedata with which they are working. In addition In the Data preprocessing we have checked null values using pandas library and in order to get better insights from data set used visualization technique using matplotlib and seaborne library. Using all such method we found outliers and skweness in the data set which we have removed using power transformation technique because we have training data in less quantity.

- Hardware and Software Requirements and Tools Used

In order to build a good model it is necessary to have good computational power hardware and software's. The used hardware for this project 8GB RAM laptop with I5 processor.And we have used python and its respective libraries.

## Model/s Development and Evaluation

- Identification of possible problem-solving approaches (methods) and Testing , Evaluations of model scores

After exploratory data analysis and preprocessing the main steps comes up that is model building and evaluation is the technique where we make assumptions that which model giving the best accuracy. Moreover , we have used hyper parameter technique to check the chances in increment of accuracy.

## CONCLUSION

- Key Findings , Conclusions of the Study and Learning Outcomes of the Study in respect of Data Science

Data analysis is a proven way for any organizations and enterprises to gain the information they need to make better decisions, serve their customers, and increase productivity and revenue. The benefits of data analysis are almost too numerous to count, and some of the most rewarding benefits include getting the right information for business, getting more value out of it. And it creates more effective marketing campaigns, gaining a better understanding of customers, and so on. Furthermore, there are many points

like available service in respective city , decision time for first shopping and most important preferred payment Options such factors contributing most. However, there are few reasons also where customers selected the products and left as it is in the cart. As per the analysis we found various reasons like lack of trust , promo code not applicable or they got better alternate offer could be offline shopping mode and frequent change in price so on. Return and replacement policy of the re-tailer is important for purchase decision. I analysis we found highest ratio of male and female those are strongly agree with replacement policy. And different type of loyalty program's give benefit's and attract customers for shopping online. Although, we can say that such data in-sights gives enterprises the ability to listen to customer interactions, learn from behavior and contextual information, create more effective actionable insights, and execute more intelligently on insights in order to optimize and engage targets and improve business practices.