

HOUSING: PRICE PREDICTION

Submitted by:

D.Mary Shereesha

**ACKNOWLEDGMENT**

This includes mentioning of all the references, research papers, data sources, professionals and other resources that helped you and guided you in completion of the project.

**INTRODUCTION**

* **Business Problem Framing**

Understand the parameters which influence the housing price in a new market, which the company has not played its role still

What to understand how these features would influence the pricing, how to manage the investment for good ROI.

If a machine Learning model can predict on what actual price the property will be sold by Local people, The company can buy the property, based on its margin and expected ROI(Return on investment)margins.

* **Conceptual Background of the Domain Problem**

Successful real estate investors always say that money is made when the property is purchased, not when it’s sold. That’s because it’s much easier to add value to a rental property than to try to turn a profit on a house that was overpriced.

Based on the above statement a machine learning model can help us to predict the Return on our investment and avoid falling into overpriced problems.

* **Review of Literature**

Some of the key factors to consider when analyzing a real estate market include

Property types with the greatest demand

Most active agents and investors

Who the local home wholesalers are

Percentage of renter-occupied households

Housing inventory stock

Where the biggest employers are located

* **Motivation for the Problem Undertaken**

It is seen that 63% of the population owns their own house

property by 2019.

Motivation this would be industry which would help the most of people dreams of having own house

**Analytical Problem Framing**

* **Mathematical/ Analytical Modeling of the Problem**

As the Housing Price is influenced by many features and it multi-dimensional, Principal component Analysis was chosen to reduce multi dimensionality in the problem

* **Data Sources and their formats**

The data sources were in csv format and imported to Jupiter notebook.

* **Data Preprocessing Done**

There were missing data, and the non-availability of a feature was indicated as blanks

The categorical features were encoded, and missing values were imputed

* **Data Inputs- Logic- Output Relationships**

The complete dataset was plotted against the sales price and features influence on sales Price were determined this was a good exercise, As in some of the features **average quality is preferred mor**e and were sold at higher price than excellent quality.

**Model/s Development and Evaluation**

* **Identification of possible problem-solving approaches (methods)**

This problem is a multidimensional influenced problem label, where the price is influenced my many features with little correlation of all of them to the Label

* **Testing of Identified Approaches (Algorithms)**

The Algorithms used were Random Forest regressor

Gradient Boost Regressor

AdaBoost Bagging Regressor

* **Run and evaluate selected models**

The model was tested with K-Fold Cross validation techniques which divides the training data set into different combinations of train-test (independently)- the accuracy at this point will help us know, if the accuracy of my model is overfitting or is just a mere chance or its real.

* **Key Metrics for success in solving problem under consideration**

The Key Metrics used is accuracy score from sklearn metrics

* **Visualizations**

**The Visualization explored for categorical variables are**

strip plots with Sales Price-To understand the relationship between both

Count plots to understand how these data are, balanced/imbalanced,

**The Visualization explored for Numeric variables are**

Distribution plots-To understand the distribution of the dataset

scatterplot to understand how these variables are influencing sales price

* **Interpretation of the Results**

reduction of the dimensions and feeding into the model help to predict the sales price with a accuracy of 83%

**CONCLUSION**

* **Key Findings and Conclusions of the Study**

**## The features which add sales Price significantly are-make the sale Price High- Can Invest on these properties having these features**

* Residential Low Density
* Paved Street
* All public Utilities (E,G,W,& S)
* having all Normal Conditions
* Single-family Detached- have high Sales
* Gable, Hip, Roofs have high price
* Fair Basement (70-79 inches), Average garage Quality- start high sales price
* SBrkr-Standard Circuit Breakers & Romex Electricity is expected to have high sales
* Overall quality- attracts more price
* Floor square feet- positively influence the price
* Average Overall condition show more price
* New Houses have slightly more price
* Average basement set is expected more
* TotRmsAbvGrd: Total rooms above grade-Good Grade is expected to give more price
* GrLivArea: Above grade (ground) living area square feet- Give more price
* Size of garage in car capacity- higher the better

**## The features which do not have sale impact- Can inore these faetures when investing**

* alley
* Neighborhood
* LotConfig
* Exterior1st: Exterior covering on house, Exterior covering on house (if more than one material)
* MasVnrType
* BsmtExposure: Refers to walkout or garden level walls
* GarageFinish: Interior finish of the garage
* SaleType: Type of sale
* SaleCondition: Condition of sale
* Month Sold
* Yr sold
* OpenPorchSF area
* EnclosedPorch area
* Wood deck area do not influence the price
* Total rooms above grade
* **Learning Outcomes of the Study in respect of Data Science**

Learning the dataset with machine learning will give a overview on investing our time and effort on things which are important in business

* **Limitations of this work and Scope for Future Work**

Learn from Local markets of real time challenges.