

Housing: Price Prediction

Submitted by:

Pramodh.P

**ACKNOWLEDGMENT**

References used that helped in completing the project Online Research Gate , Google , Medium,Kaggle

Data ,Goals and Problem statement given by FlipRobo

**INTRODUCTION**

Introduction The real estate sector is an important industry with many stakeholders ranging from regulatory bodies to private companies and investors, there is a high demand for a better understanding of the industry operational mechanism and driving factors. Today there is a large amount of data available on relevant statistics as well as on additional contextual factors, and it is natural to try to make use of these in order to improve our understanding of the industry. Notably, this has been done in Zillow’s Zestimate [4] and Kaggle’s competitions on housing prices [2]. In some cases, non-traditional variables have proved to be useful predictors of real estate trends. For example, This project can be considered as a further step towards more evidence-based decision making for the benefit of these stakeholders. The project focused on assessment value for residential properties for a US-based housing company named **Surprise Housing**

* Business Problem Framing

Predict the price of the house with the provided variables

And understand how the prices are affected with the variable based on which we can come up with strategy and target areas that will yield high return.Using the model we can understand the market and come with the required strategy

**Analytical Problem Framing**

The main steps in our research were the following.

• Exploratory Data Analysis (EDA). By conducting explanatory data analysis, we obtain a better understanding of our data. This yields insights that can be helpful later when building a model, as well as insights that are independently interesting.

• Feature Selection

• Modeling We apply Liner regression for prediction of the housing prices.

**A pseudo code of understanding of the steps :**

Step 1: Import the data

Step2: Find the missing values in the data set as the cleaning process is done and separate the data

Step3: separate the data to numerical and categorical

Step4:”Visualization “

selection of feature by heatmap or matrix correlation.

Step5: splitting of the dataset into two part training data and the testing data

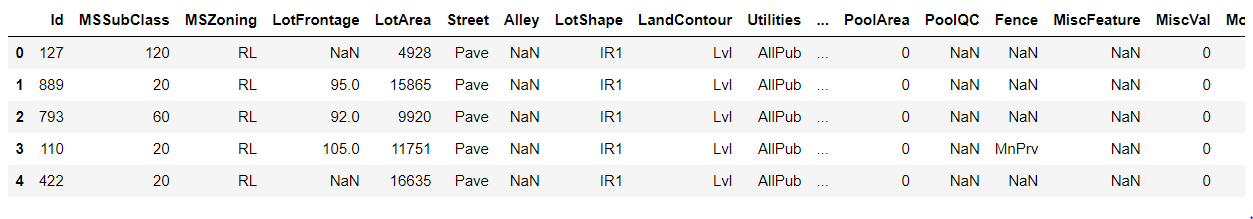
Step6: apply or fit the regression technique on training data and test it with testing data

Step7: Compare the accuracy result.

**Data Set**

**Importing data sets provided**

Example of the data



**Splitting the data**

We are splitting the features into two

1; Numerical Data 2. Categorical Data

**Cleaning The data**

Removing or Dropping unnecessary columns

Finding Nan

**Separating the Target Feature**

In our case the SALE PRICE

**Data Visualization**

Data visualization is the graphical representation of information and data. By

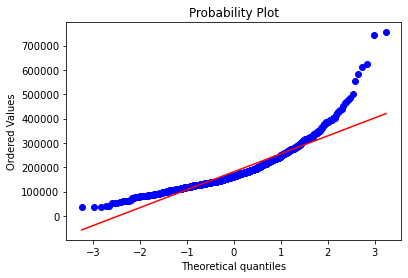
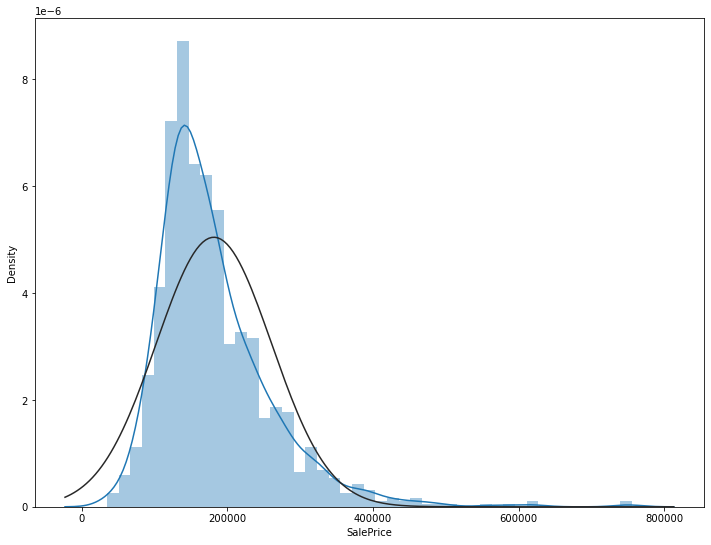
using visual elements like charts, graphs, and maps, data visualization tools provide an

accessible way to see and understand trends, outliers, and patterns in data. In the

world of Big Data, data visualization tools and technologies are essential to analyse

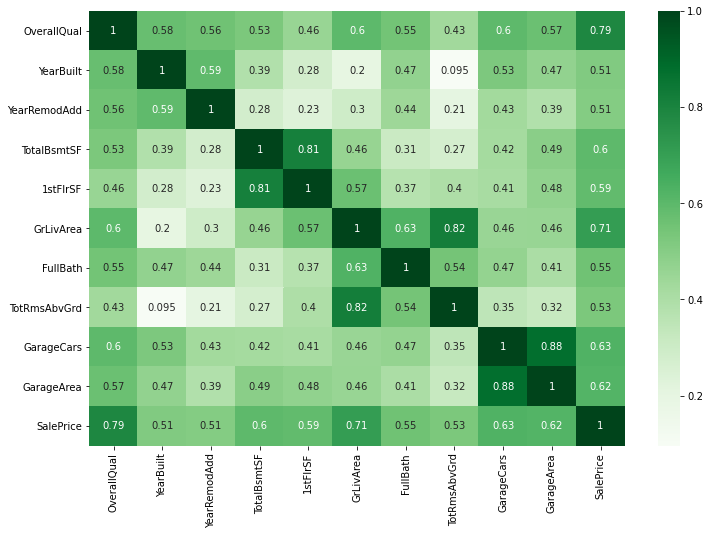
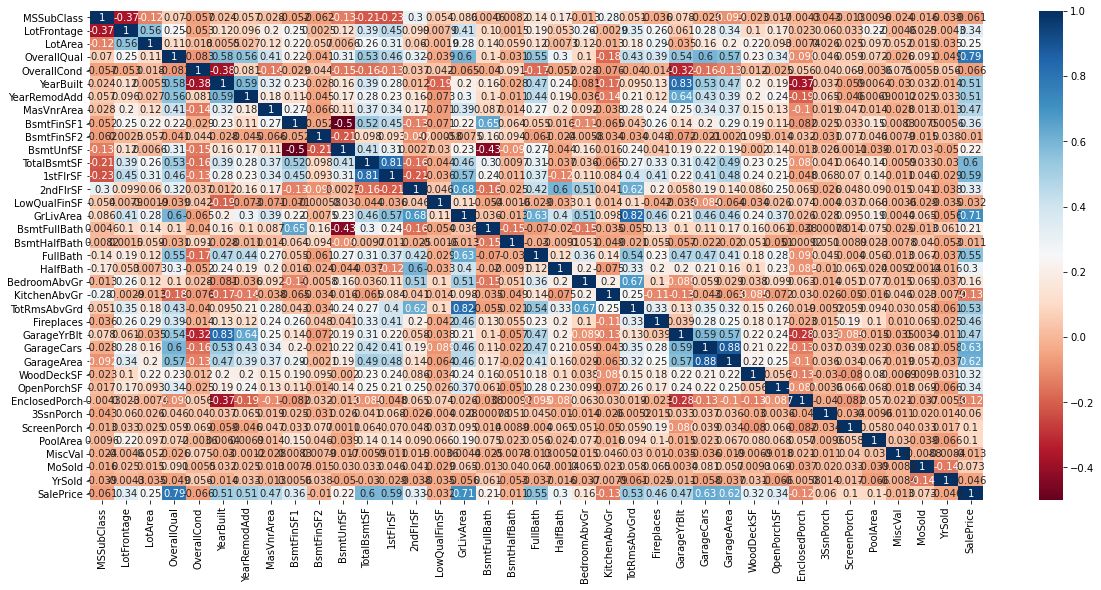
massive amounts of information and make data-driven decisions

**Distribution Of target variable - Sale Price using Seaborn distplot and Probability Plot**



**We will use heatmap to see the Correlation between variables**

**Find the highest and plot using a heatmap**



Mathematical/ Analytical Modeling of the Problem

Describe the mathematical, statistical and analytics modelling done during this project along with the proper justification.

* Data Sources and their formats

What are the data sources, their origins, their formats and other details that you find necessary? They can be described here. Provide a proper data description. You can also add a snapshot of the data.

* Data Preprocessing Done

What were the steps followed for the cleaning of the data? What were the assumptions done and what were the next actions steps over that?

* Data Inputs- Logic- Output Relationships

Describe the relationship behind the data input, its format, the logic in between and the output. Describe how the input affects the output.

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

Here, you can describe any presumptions taken by you.

* Hardware and Software Requirements and Tools Used

Listing down the hardware and software requirements along with the tools, libraries and packages used. Describe all the software tools used along with a detailed description of tasks done with those tools.

**Model/s Development and Evaluation**

* Identification of possible problem-solving approaches (methods)

Describe the approaches you followed, both statistical and analytical, for solving of this problem.

* Testing of Identified Approaches (Algorithms)

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

* 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.

* Key Metrics for success in solving problem under consideration

What were the key metrics used along with justification for using it? You may also include statistical metrics used if any.

* Visualizations

Mention all the plots made along with their pictures and what were the inferences and observations obtained from those. Describe them in detail.

If different platforms were used, mention that as well.

* Interpretation of the Results

Give a summary of what results were interpreted from the visualizations, preprocessing and modelling.

**CONCLUSION**

* Key Findings and Conclusions of the Study

Describe the key findings, inferences, observations from the whole problem.

* Learning Outcomes of the Study in respect of Data Science

List down your learnings obtained about the power of visualization, data cleaning and various algorithms used. You can describe which algorithm works best in which situation and what challenges you faced while working on this project and how did you overcome that.

* Limitations of this work and Scope for Future Work

What are the limitations of this solution provided, the future scope? What all steps/techniques can be followed to further extend this study and improve the results.