REGRESSION ON HOUSE SALES AND PARAMETRS

*

1st Ketaki Ransing
EXTC Department
Sardar Patel Institute of Technology
Mumbai,India
ketaki.ransing@spit.ac.in

2nd Apurva Lingayat
EXTC Department
Sardar Patel Institute of Technology
Mumbai,India
apurva.lingayat@spit.ac.in

3rd Ashutosh Pandya
EXTC Department
Sardar Patel Institute of Technology
Mumbai,India
ashutosh.pandya@spit.ac.in

***Abstract*—At onset of modernization new houses and buildings are built every day, so there is a need for a system to predict which features of the houses are important for the customers.The factors that influence the price of a house includes physical conditions, concept and location.The aim of this report is to predict the house sales in King County, Washington State, USA using Multiple Linear Regression (MLR). We have compared the Linear Regression and Multiple regression.The dataset consisted of historic data of houses sold between May 2014 to May 2015.By analysis, we have predicted which factors have affected the house prices. We performed Linear Regression, Multiple Regression and interpret the better from the above.[1]**

***Index Terms*—Linear Regression, Multiple Regression.**

I. Introduction

A. LINEAR REGRESSION

In statistics, linear regression is a linear approach to modeling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). The case of one explanatory variable is called simple linear regression. Linear Regression refers to a group of techniques for fitting and studying the straight-line relationship between two variables. In our project we have plotted the linear regression and calculated the the intercept in section III.[3]

B. MULTIPLE LINEAR REGRESSION

A statistical tool that allows you to examine how multiple independent variables are related to a dependent variable. Once you have identified how these multiple variables relate to your dependent variable, you can take information about all of the independent variables and use it to make much more powerful and accurate predictions about why things are the way they are. This latter process is called "Multiple Regression".[3]

C. CORRELATION MATRIX

A correlation matrix is a table showing correlation coefficients between variables. Each cell in the table shows the correlation between two variables. A correlation matrix is used to summarize data, as an input into a more advanced analysis, and as a diagnostic for advanced analyses. There are three broad reasons for computing a correlation matrix:
1.To summarize a large amount of data where the goal is to see patterns. In our example above, the observable pattern is that all the variables highly correlate with each other.
2.To input into other analyses. For example, people commonly use correlation matrixes as inputs for exploratory factor analysis, confirmatory factor analysis, structural equation models, and linear regression when excluding missing values pairwise.
3.As a diagnostic when checking other analyses. For example, with linear regression a high amount of correlations suggests that the linear regression's estimates will be unreliable.[8]

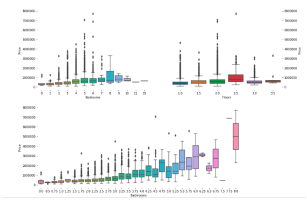
Identify applicable funding agency here. If none, delete this.

II. Methodology

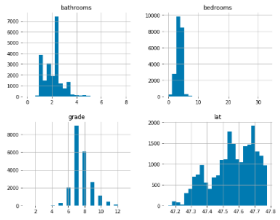
We evaluated how different parameters affect the price of the houses. We imported the dataset for the house sales in King County, Washington State, USA.We read the columns of the data and information about the columns which are the parameters for house sales.
With the help of histogram and boxplot we visualized the given data for prediction.By Pearson correlation matrix we found out the most crucial parameters for the prediction.We created the linear regression model on crucial parameters.
We created multiple linear regression model on various features and predicted the house sales. We compared both the models and analyzed the better of the model.

III. Results

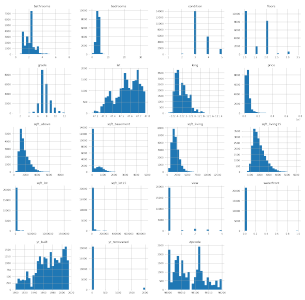
A. Boxplot



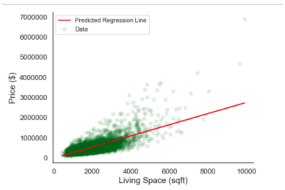
B. Histogram



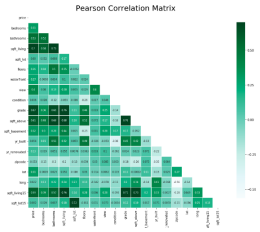
C. Histograms of all parameters



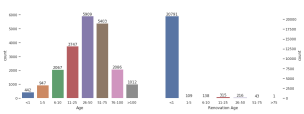
D. Linear Regression



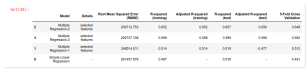
E. Correlation



F. Histogram For Age and Renovation Age



G. Comparison



IV. Result Analysis

From section III(A),III(B)and III(C) we can analyze how different parameters affect the price of house.It shows how locations, the physical attributes of the house can help the builders in planning.

Section III(D),shows the relationship between price and living space.We interpret that most of houses have living space upto 4000 sq. feet with a price range upto 1cr.The red line in the graph shows the linear relation between living area and price.

From section III(E),shows that few parameters like living area and price are highly related.While basement area and numbers of floors are least related.

From section III(F),shows us the age and renovation age of the houses.

Section III(G),multiple regression with more features gives better analysis with less error compared to simple linear regression and multiple regression with less features. We observed that multiple regression works better than linear regression.

V. Conclusion

The user can find out how price of houses depends on various parameters.Thus user can find the house which fits his requirement. Pearson correlation matrix is a crucial tool which helps us to find the dependency of one parameters with all other parameters.

We analyzed which parameters affect the price most and least. By comparing simple and multiple linear regression ,we analyzed the better model for prediction.

As we can see from section III(A) bedroom houses are most commonly sold followed by 4 bedroom.For a builder having this data , He can make a new building with more 3 and 4 bedroom's to attract more buyers.

References

[1] Adyan Nur Alfiyatin,Hilman Taufiq,Ruth Ema Febrita and Wayan Firdaus Mahmudy, "Wayan Firdaus Mahmudy" (IJACSA) International Journal of Advanced Computer Science and Applications,Vol. 8, No. 10, 2017

[2] Kavitha S , Varuna S and Ramya R on A Comparative Analysis on Linear Regression and Support Vector Regression,2016 Online International Conference on Green Engineering and Technologies (IC-GET)

[3] https://en.wikipedia.orgwikiLinear-regression

[4] https://github.com/Shreyas3108/house-price-prediction/blob/master/housesales.ipynb

[5] https://towardsdatascience.com/regression-using-sklearn-on-kc-housing-dataset-1ac80ca3d6d4

[6] https://medium.com/@datalesdatales/predicting-house-prices-with-linear-regression-595422992c48

[7] https://www.kaggle.com/erick5/predicting-house-prices-with-machine-learning

[8] https://www.displayr.com/what-is-a-correlation-matrix