

NAME OF THE PROJECT

Submitted by:

SUDHAN T

**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

Buying a house Is one of the tough task on current situation, it is hardly to find is the place are developing place is there a good water resources is there any nearby educational sector, we do see in the real estate business we do see a genuine person but there are some people who are here to cheat the people, so finding the best place is the hard for the people now a days. But when we have data in our hand we can easily predict where we can buy the house or land

* Conceptual Background of the Domain Problem

The things I say which will be more useful to understand the domain problems majorly the people look into is there a good water resources is that the place is developing place is there transport facility available, if e have the data’s for the which will be more correlated to the Target and it will be more useful to predict easily

* Motivation for the Problem Undertaken

I would say the Motivation behind this given project , I have a dream to buy a home also I do have dream how it should be which made me to involve in this project to analyse more and learn deeply

**Analytical Problem Framing**

* Mathematical/ Analytical Modeling of the Problem

The scaling and sampling has not been done since it is a regression type model also the Encoding has been done since given data set are Categorical value

* Data Sources and their formats

The given data everything are important but mainly I have focused on some of the data set which is more important they are Year of sold , Fence , Sales condition , Sales types , Overall Quality , lot area year of built, Garage area , external quality which are some of the important data need to be focused while buying a new houses

* Data Pre-processing Done

I have not removed any data except the ID because each single information add the value when we buy a home which give added advantage to every home

* Data Inputs- Logic- Output Relationships

If we are buying a new home we will spend or invest by seeing the future like if it is profit or loss else this city or town get Devloped or not in the case each and every given data’s are important according to me specially while buying the home the price get vary because of the single advantage we see in one home so all the data set are parallel important to the output

* Hardware and Software Requirements and Tools Used

The software I used is Jupyter notebook

**The libraries I used are**

Pandas

Numpy

Seaborn

matplotlib

**Model/s Development and Evaluation**

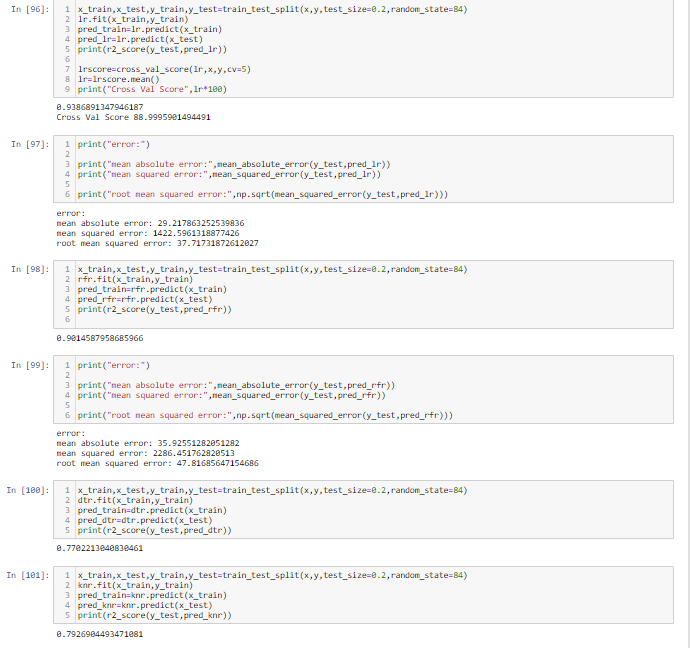
* Testing of Identified Approaches (Algorithms)

The used algorithms are

Linear Regression, Knearest neighbour, Random forest regressor,

Decision tree Regressor, Support vector Regressor

* Run and Evaluate selected models



* Visualizations

For Visualization I used the Library seaborn and the plot like Histogram , Swarm plot , Pie chart, Count plot , Cat plot

**CONCLUSION**

* Learning Outcomes of the Study in respect of Data Science

The best learning for me in this assignment which is working with the visualization and converting the Categorical into Numerical I have face little difficulties while encoding the Data’s but finally I can able to manage the to get the Model with the score of about 93% in linear regression method

* Limitations of this work and Scope for Future Work

Since the given all the data’s placed the major roles of there are dataset like The future development of the city how it would be after a decade if there are dataset provided like that which will be more useful