**Project 3 – Housing prices prediction for Bangaluru city**

**Name : R. Vaibhava lakshmi**

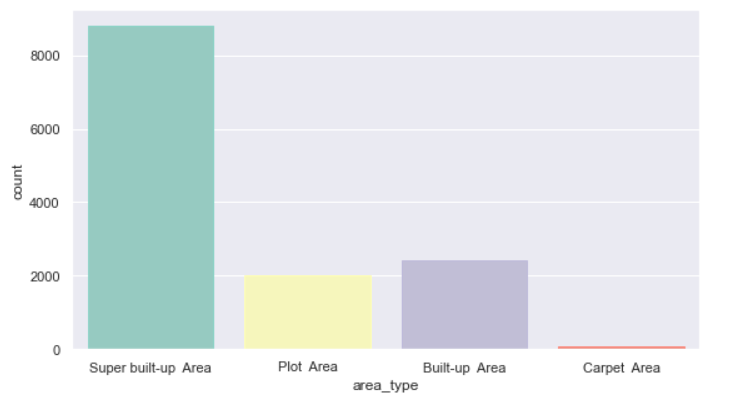
1.Introduction

Predicting house prices can help to determine the selling price of a house of a particular region and can help people to find the right time to buy a home. Our task is to use data from the Bangaluru census to create a machine learning model to predict house prices in the State. Ultimately, our machine learning model (XGB regression) should learn from this data and be able to predict the median house price in any neighbourhood, given all other metrics.The results obtained are fairly good , with a train score of 83.29% and test score of 71.63%. Apart from this , Exploratory analysis has also been done to find relationships between several determining factors.

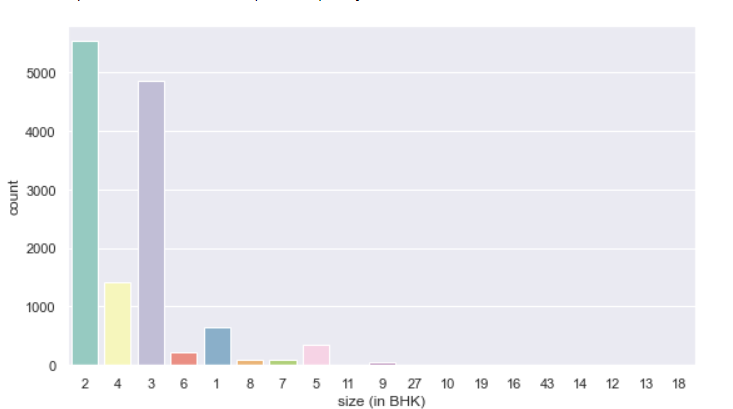
2.Problem definition and algorithm

In this project , the XGB regression algorithm being used to predict the housing prices in future. Ensemble learning involves training and combining individual models (known as base learners) to get a single prediction, and XGBoost is one of the ensemble learning methods. XGBoost expects to have the base learners which are uniformly bad at the remainder so that when all the predictions are combined, bad predictions cancels out and better one sums up to form final good predictions.

3.Experimental evaluation

BAR GRAPH DENOTING COUNT OF DIFFERENT AREA TYPES

BAR GRAPH DENOTING COUNT OF DIFFERENT HOUSE SIZES (IN BHK)



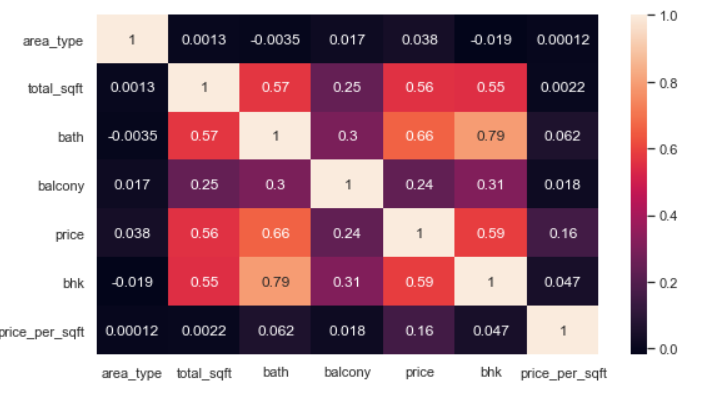
Relationship between price and total square feet before removing outlier



Relationship between price and total square feet after removing outlier



Heatmap denoting correlation between several independent variables



RESULTS OF LINEAR REGRESSION:

train\_score : 0.6408787149895181

test\_score : 0.6257205297112027

RESULTS OF RANDOM FOREST REGRESSION:

test\_score 0.6565422003575168

train\_score 0.9332400258135476

RESULTS OF XGBOOST REGRESSION:

test\_score 0.7163920207963567

train\_score 0.8329092848908546

Hence , XGBoost model was used for the prediction, due to its high test and train score compared to the other two models.

GUI was developed using tkinker package in Python. First GUI widget is to display the details of areas with top 10 price per square feet.

Second GUI widget is to input the CSV data and predict prices and display the TRAINING and TEST scores.

Related work and references:

<https://www.kaggle.com/sonukiller99/house-price-in-india>

<https://www.kaggle.com/nschaitanya/house-price-prediction-using-hybrid-ml-model-nfs>

<https://www.kaggle.com/malyajkaloni/house-price-prediction-bangalore>

The above mentioned links contain projects that have used similar methodologies and algorithms to predict housing prices using the same dataset.

Conclusion:

Various insights have been drawn using EDA , which were useful to remove the outliers in the data , after which the fitting of Machine learning model(XGBoost regression) was done, with a considerably good result.