# Introduction

In the filed of machine learning and data mining, the perdiction of house price is very common and chose to make a program that predict the house price based on dataset of a huge records.

The program we made was designed to be an html page that required some data from user describe the some features of the house he wants and predict the price of the house he wanted all based on dataset of 13320 record using data mining algorithms.

The model building

First data set downloaded from <https://www.kaggle.com/amitabhajoy/bengaluru-house-price-data> and is consisting of (['area\_type', 'availability', 'location', 'size', 'society', 'total\_sqft', 'bath', 'balcony', 'price'] columns and 13320 rows

Second data cleaning started with droping unnesesary columns as 'area\_type', 'availability',

'balcony', 'society'.

Then the 'total\_sqft' filed which contains non-specific values so we needed to get a mean average of the original value which was a rang of minimum and maximum values (e.g. 3067 – 8156), note that not all row are equal, but some have a single value and others have a rang of tow values as shown. And we builed a function to convert those rang values to a single average valu.

As the houses price is very dependent on price per the total square area so we needed this new feature insid the csv file, as the location filed contains lots of values so We need to apply dimensionality reduction technique here to reduce number of locations by declaring any location has less than 10 records to get a 240 from 1287 unique location.

After the data cleaning, the outlier removal using standard deviation of the new feature “price\_per\_sqft” to remove any values greater than mean minus standred deviation and less than mean plus standred deviation

Then removing any record whose bedroom area less than 300 ft as a standard policy very common in pricing houses market.

After finishing from the cleaning stage we can now build our model by using a common algorithm(GridSearchCV) who is going to compare between the score of linear\_regression, lasso and decision\_tree o and the result shows that linear regression was the best of them

Related works

## Housing Price Prediction App with AWS Deployment[[1]](#footnote-1) [link](1.%09https:/github.com/codebasics/py/blob/master/DataScience/BangloreHomePrices/model/banglore_home_prices_final.ipynb)

This is the main project we work on it and editing it in our way

### The problems with the code

1. Acurecy was just around 80 percent all the time
2. Data visualizing was not good enugh
3. Data cleaning was implemented before full understanding and visualizing the core dataset

## Bengaluru\_House\_Price\_Prediction[[2]](#footnote-2) [link](https://www.kaggle.com/code/nishigandhazanje/bengaluru-house-price-prediction)

this project is pretty good, she uses very simple code to do the job .however, most problem in prevous one still found.

## Bangaluru House Price Analysis[[3]](#footnote-3) link

This project was mainly focusing on data analysis and nearly clear of previused mentioned data understanding and visualizing problems but still it didn’t work on perdiction.

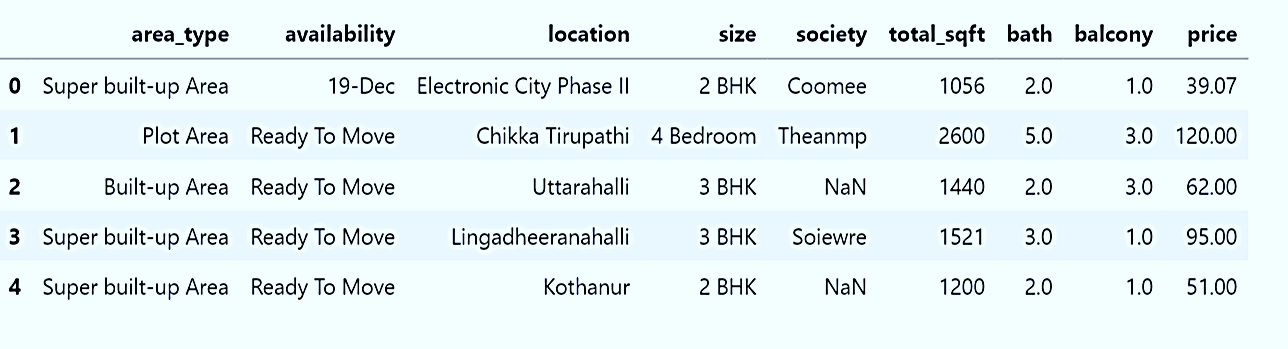
HousePricePrediction[[4]](#footnote-4) link

# Methods

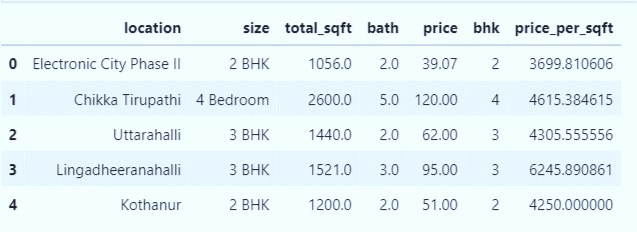
In this section we are going to explain every technology we use including datasets, Jubyter notebook, all requirement libraries, model algorithms and website requirments in both front-end and back-end.

## The dataset

Dataset was downloaded from <https://www.kaggle.com/amitabhajoy/bengaluru-house-price-data>

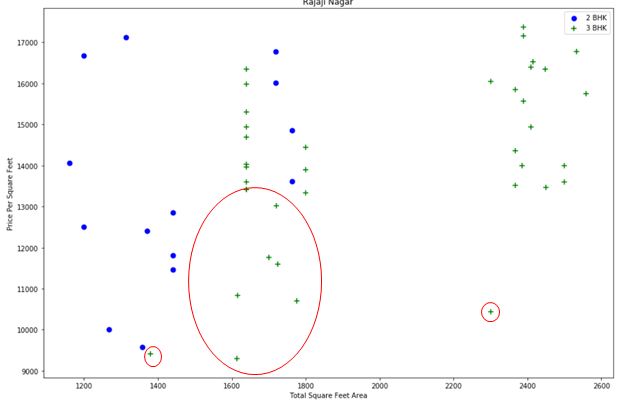
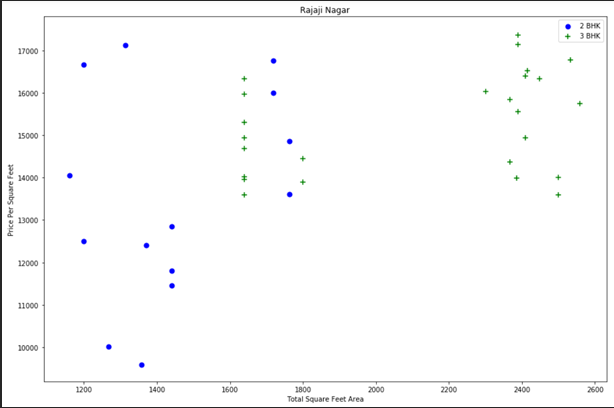


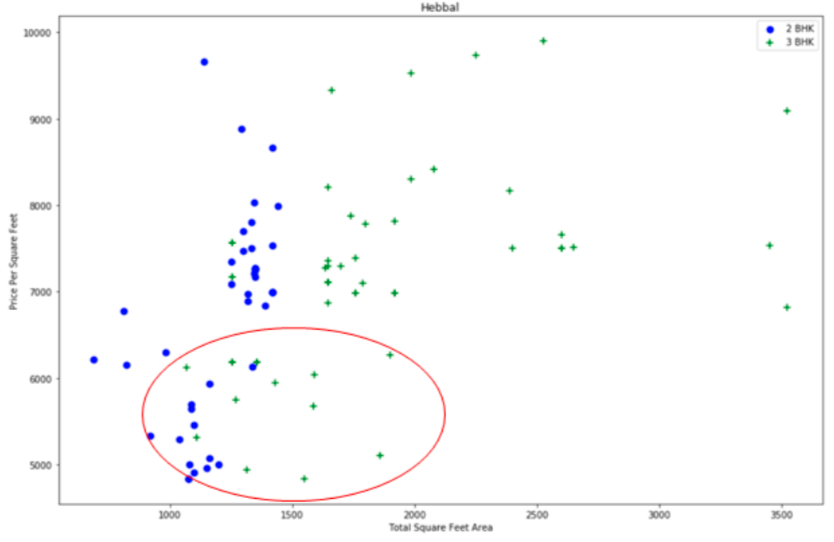
After data cleaning and feature engenering



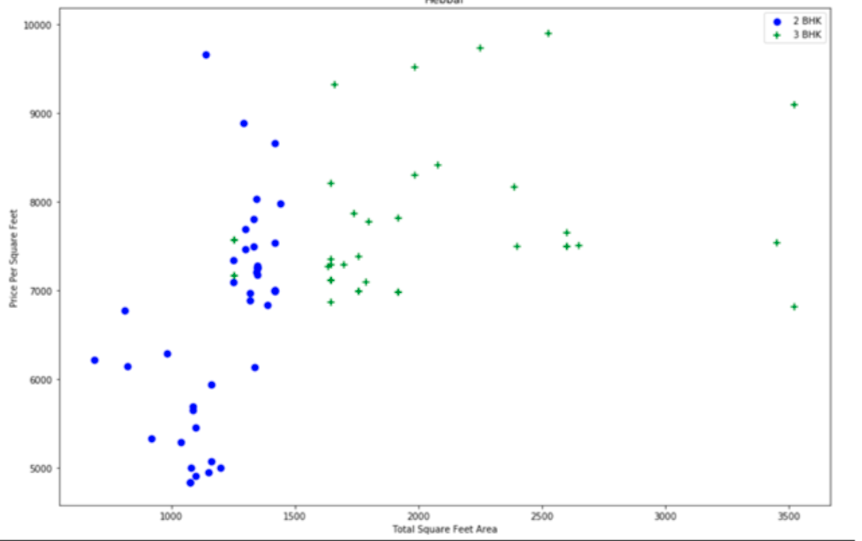
## Outlier Removal

For any house with 3 bedrooms with price less than once with 2 bedrooms



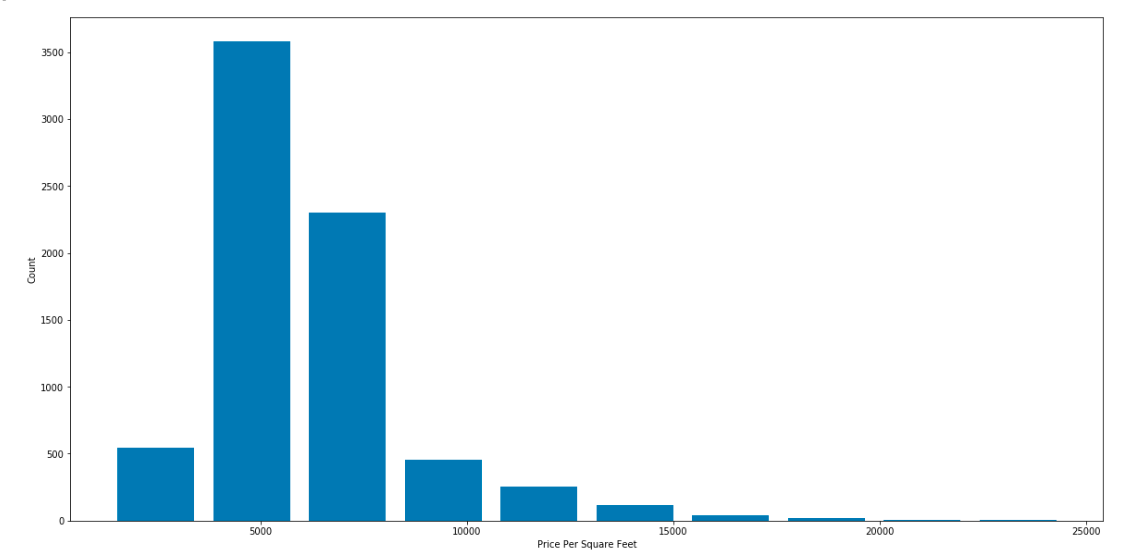
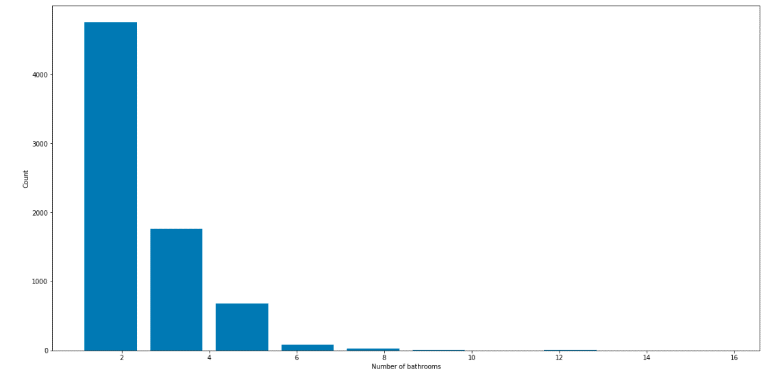


After procesing



Outlier Removal Using Bathrooms Feature

Before after



Programing languages and libraries

Model

1. Python language and jupyter notebook as an IDE
2. Libraries
   1. Numpy
   2. warnings
   3. Padas
   4. Matlib
   5. Skitlearn

Web-site

Front-end

1. Html, Css, javaScribte, and bootstrab

Back-end

1. Python language, and VScode as IDE
2. Python flask server

# Conclusion

# References

1. [ME2MLE](https://www.kaggle.com/samsonlo), Mechanical Engineer, Hong Kong, Hong Kong [↑](#footnote-ref-1)
2. [Nishigandha.Zanje](https://www.kaggle.com/nishigandhazanje), Student at Symbiosis Institute of Technology - Pune, Maharashtra, India [↑](#footnote-ref-2)
3. [Tanusha](https://www.kaggle.com/tanushagupta) Gupta, Nothing at Nowhere - Patna, Bihar, India [↑](#footnote-ref-3)
4. [Parul Rajwade](https://www.kaggle.com/parulrajwade), Student at Symbiosis Institute of Technology - Pune, Maharashtra, India [↑](#footnote-ref-4)