

YOUTUBE LINK:

<https://www.youtube.com/watch?v=zlCCnh3YL9M>

PPT LINK:

https://github.com/donthula9908/Naveen_data606

GITHUB LINK:

https://github.com/donthula9908/Naveen_data606

REAL ESTATE PRICE PREDICTION

Subject: Data 606

Professor: Chaojie Wang

NAVEEN DONTULA (PP42893@umbc.edu)

SUMMER 2023.

Abstract:

People are cautious while attempting to purchase a new home using their budgets and market methods. The purpose of this study is to anticipate consistent house prices for non-houses holders depending on their financial preparations and their objectives. By examining the preceding products, fare ranges, and also Forecasted prices will be predicted to forewarn developments. The Predictions are made utilising several Regression approaches in this study. For example, multiple linear, ridge, LASSO, elastic net, and gradient Boost Regression and boosting House price forecast. All of the aforementioned tools were used to create the data set strategies to determine which one is the greatest. The goal of this study is to assist sellers in accurately estimating the selling price of a property and to assist individuals in predicting the precise time slap to accumulate a house. Physical conditions, idea, and location, among other things, were taken into account as cost-influencing elements.

Introduction:

This article refers together with latest Forecast on Research predictions considering trends to further plan their economics. The main motivation of the project FORECASTING VARIATIONS ON HOUSE PRICE was to make the best possible prediction of house prices by using appropriate algorithms and finding out which among them is best suitable for predicting the price with low error rate. This is an interesting problem because most of the people will eventually buy/sell a home. This problem allows us, as house price analysts, to learn more about the housing market and helps with making more informed decisions. The analysis that were done in this paper is mainly based on the dataset of Bangalore, KA. because of unexpected changes in price of houses in and around Bangalore due to emergence. In this paper ,we try to demonstrate all the possible Regression techniques which are suitable to our problem.

Problem Statement:

The aim of this project is to develop a reliable real estate price prediction model that can accurately estimate property prices based on various features. The dataset contains historical data of real estate properties, including information such as location, size, number of bedrooms, bathrooms, and other relevant attributes. You are required to build a linear regression model using the Scikit-learn (sklearn) library in Python and implement data science principles like outlier removal and feature engineering to improve the model's predictive performance.

Objective:

Our aim is to predict the price of houses located in a area by using the regression models. Aim is to accurate are the real estate price predictions provided by the developed model compared to the actual market prices, and what is the overall performance evaluation of the model in terms of mean squared error or R-squared value. The real estate price prediction website handle a high volume of concurrent user requests, and what is the average response time for serving predictions.vHow efficiently does the model perform with dimensionality reduction techniques, balancing predictive power and computational complexity? Used hyperparameter tuning using GridSearchCV in improving the model's performance and capable enough to use in the real estate.

Solution:

The research interests in this project include advanced machine learning models for real estate price prediction, novel feature engineering techniques, deep learning algorithms for market trend analysis, and the impact of external factors on real estate prices. The expected outcomes of this research are improved accuracy in predicting real estate prices, enhanced understanding of factors influencing prices, automated valuation models for financial institutions, and comprehensive market analysis tools for real estate professionals and policymakers. These research interests aim to drive informed decision-making, facilitate strategic investments, streamline property appraisal processes, and provide valuable insights into the dynamics of real estate markets.

Dataset:

Dataset has 13,320 rows and 9 columns. This dataset is from Kaggle created by amitabhajoy.

Data Load: Load banglore home prices into a dataframe

```
In [72]: df1 = pd.read_csv("bengaluru_house_prices.csv")
df1.head()
```

Out[72]:

	area_type	availability	location	size	society	total_sqft	bath	balcony	price
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Coomee	1056	2.0	1.0	39.07
1	Plot Area	Ready To Move	Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	3.0	120.00
2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	NaN	1440	2.0	3.0	62.00
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0	95.00
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	NaN	1200	2.0	1.0	51.00

IMPLEMENTATION:

This research makes use of jupyter IDE. It is an open-source web app that helps us to share as well create documents which have live code, visualizations, equations and text that narrates. It contains tools for data cleaning, transformation of data. Simulation of numeric values, modelling using statistics, visualization of data and machine learning tools. Here we collected house sales related data to estimate the house prices based on real world dataset. It is a public output dataset of that specified region in India. All the above mentioned regression techniques are implemented using the above specified tools. In order to find out the efficient regression technique for prediction, we require certain parameters to perform comparison among the techniques. The parameters chosen for the comparison are Scores of the algorithm, [MSE] Mean Square Error and [RMSE] Root Mean Square Error. The below Table 1 represents the resultant summary of the parameters, when above techniques are implemented practically.

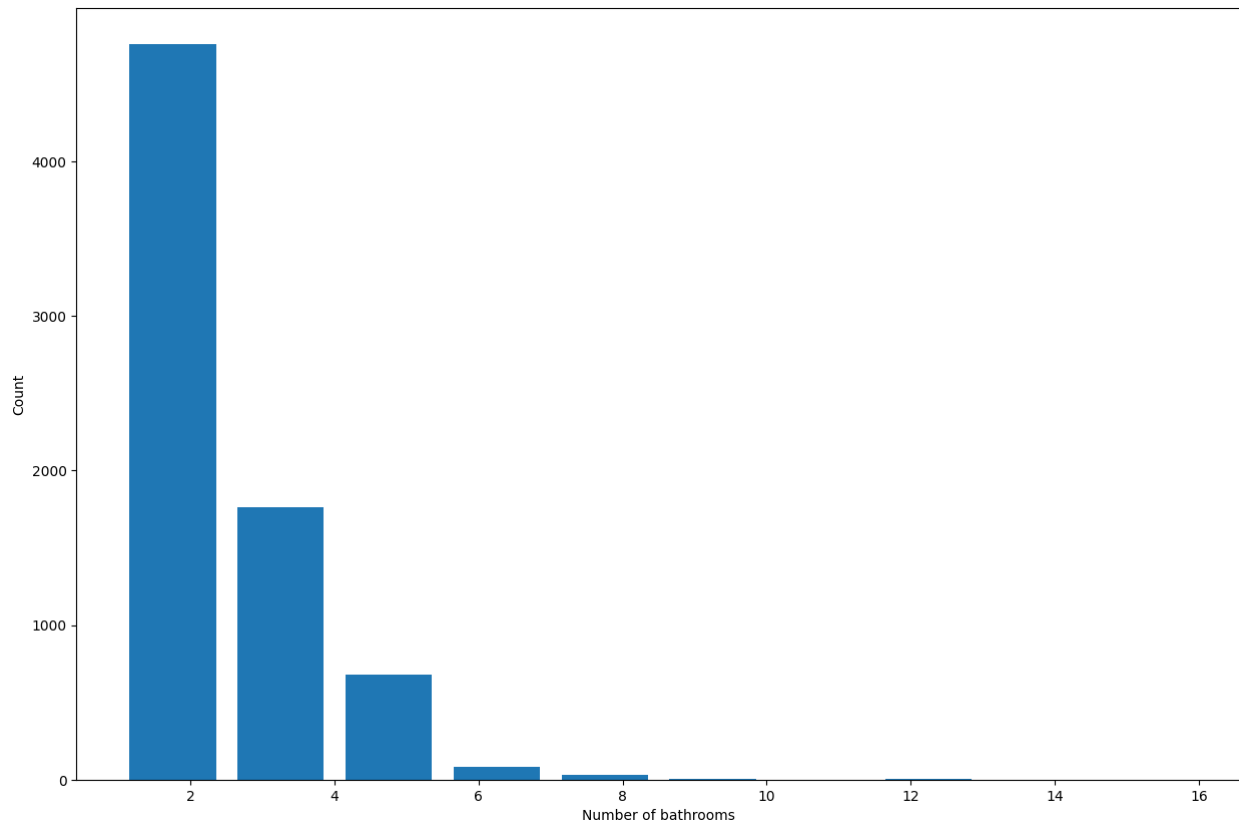
]:

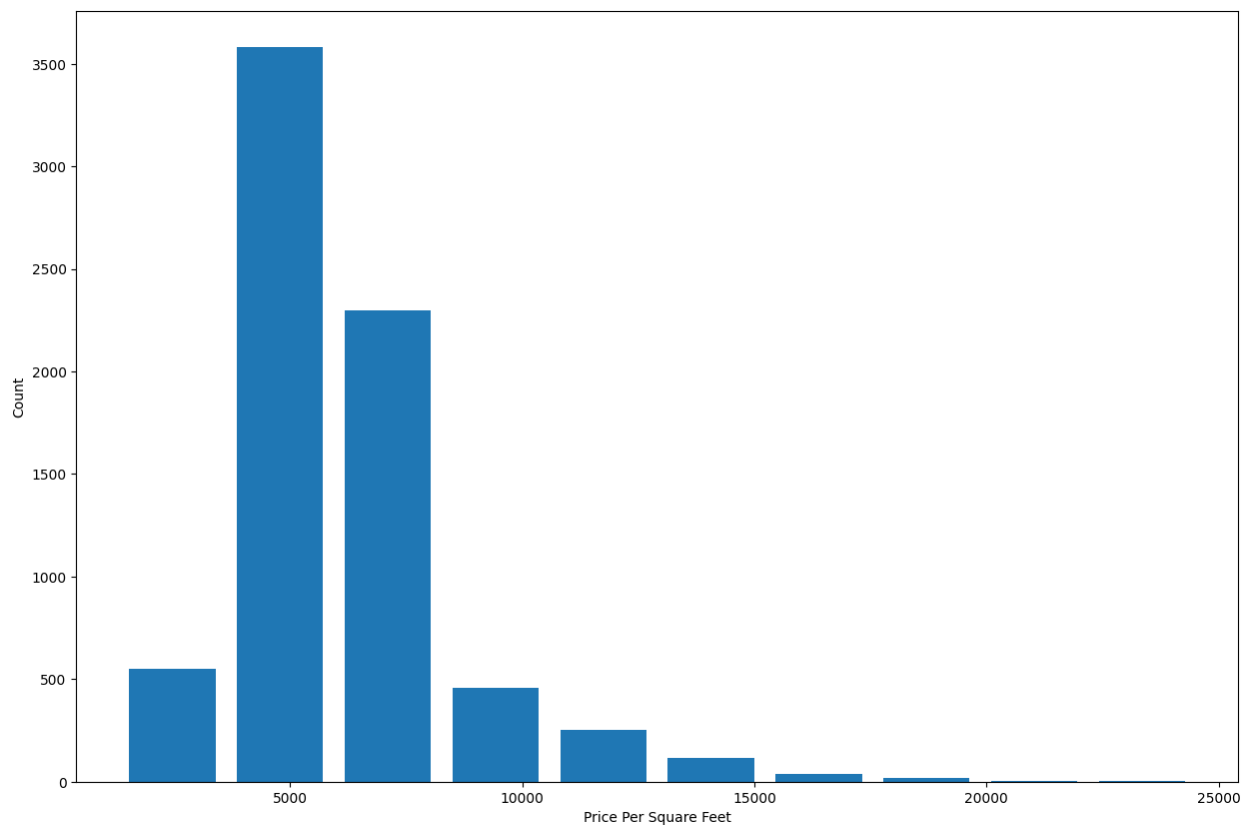
	model	best_score	best_params
0	linear_regression	0.847796	{'normalize': False}
1	lasso	0.726752	{'alpha': 2, 'selection': 'random'}
2	decision_tree	0.671597	{'criterion': 'friedman_mse', 'splitter': 'best'}

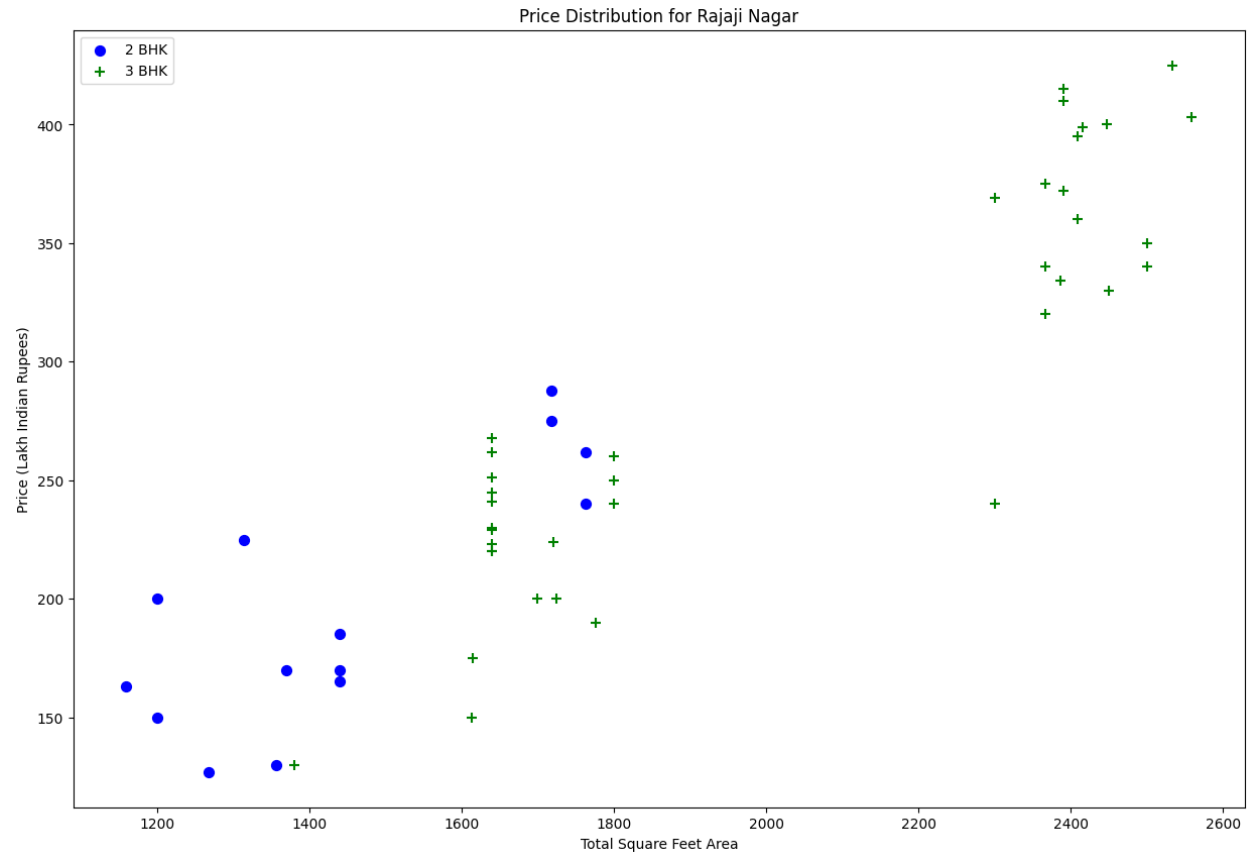
From the above table, we can easily perform comparison of different algorithms clearly to find the best among them. We can say that linear regression is having highest score so, I thought it will be better to use this model in the prediction.

Visualizations:

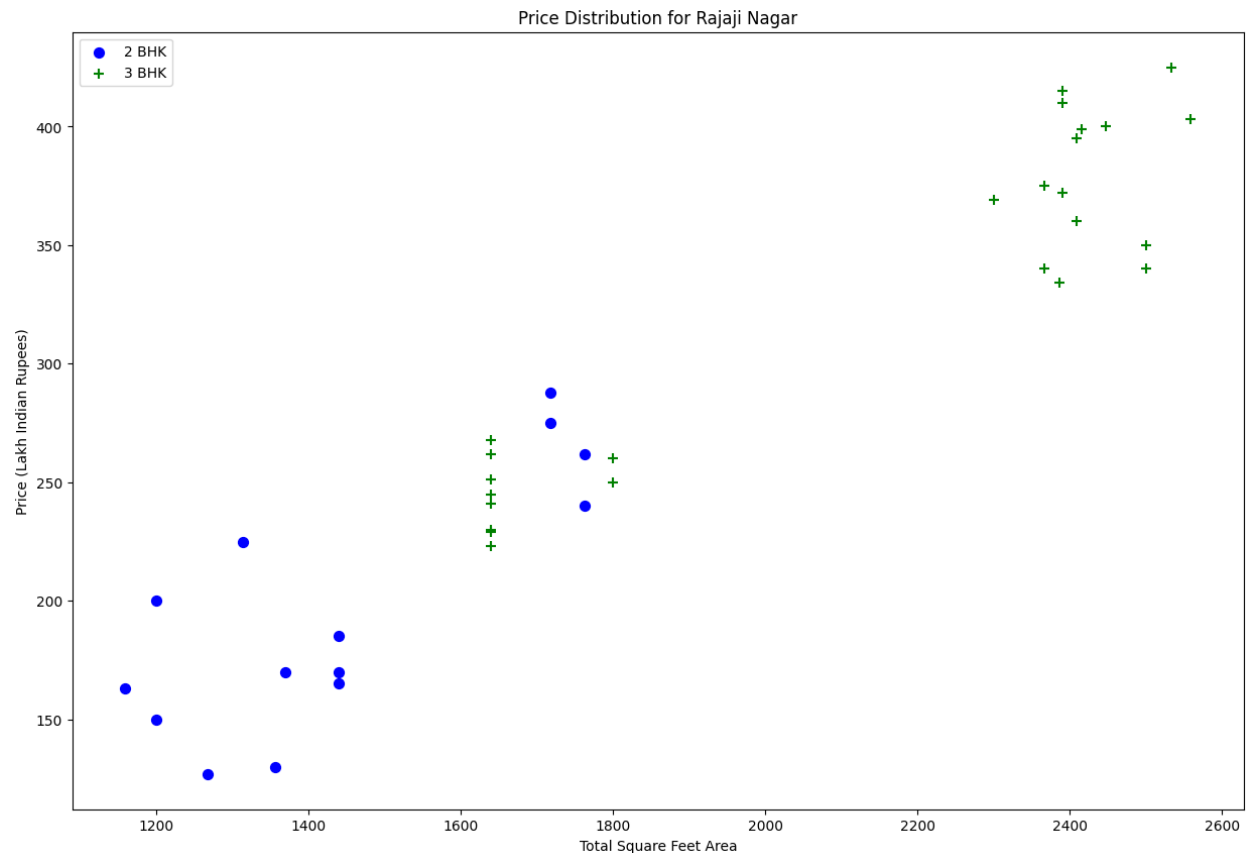
The below are some visualizations of number of bathrooms v/s count. The first one is before and second graph is after outline removal.







Plot same scatter chart again to visualize price_per_sqft for 2 BHK and 3 BHK properties



Model Evaluation:


The models' performances were evaluated using metrics such as R-squared and mean squared error. These evaluations shed light on the effectiveness of each model in capturing the variance in home prices.


Performed Gridsearchcv function that evaluates the performance of different regression models using hyperparameter tuning and cross-validation. It aims to find the best-performing model among Linear Regression, Lasso Regression, and Decision Tree Regression for predicting house prices based on provided features and target values.

Results:

```
[ ] predict_price('1st Phase JP Nagar',1000, 2, 2)
```

```
83.86570258311222
```

```
 predict_price('1st Phase JP Nagar',1000, 3, 3)
```

```
 86.08062284985995
```

```
[ ] predict_price('Indira Nagar',1000, 2, 2)
```

```
193.31197733179556
```

```
[ ] predict_price('Indira Nagar',1000, 3, 3)
```

```
195.52689759854331
```

The above are some predicted prices of houses in some areas where the price is in Indian rupees in lakhs.

Conclusion:

This article mainly concentrates on the comparison between different machine learning algorithms (Multiple Linear Regression, LASSO Regression, gradient boosting) about House price prediction Analysis. From the above experiment results, Linear regression algorithm has high accuracy value when compared to all the other algorithms regarding house price predictions. The paper can be extended by applying the above said algorithms to predict House resale value.

Future Work:

Further exploration could involve incorporating more advanced regression algorithms or exploring additional datasets to enhance the predictive accuracy of the models. Deployment of this model building into an application or a web server will be a good idea.

References:

- [1]. Aminah Md Yusof and Syuhaida Ismail ,Multiple Regressions in Analysing House Price Variations. IBIMA Publishing Communications of the IBIMA Vol. 2012 (2012), Article ID 383101, 9 pages DOI: 10.5171/2012.383101.
- [2]. Babyak, M. A. What you see may not be what you get: A brief, nontechnical introduction to over fitting regression-type models. Psychosomatic Medicine, 66(3), 411-421.
- [3]. Atharva chogle, priyanka khair, Akshata gaud, Jinal Jain .House Price Forecasting using Data Mining Techniques International Journal of Advanced Research in Computer and Communication Engineering ISO 3297:2007 Certified Vol. 6, Issue 12, December 2017
- [4]. Darshan Sangani, Kelby Erickson, and Mohammad Al Hasan, Predicting Zillow Estimation Error Using Linear Regression and Gradient Boosting, IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems (MASS), Page(s):530 - 534
- [5]. Model, Azme Bin Khamis, Nur Khalidah Khalilah Binti Kamarudin ,Comparative Study On Estimate House Price Using Statistical And Neural Network, International journal of scientific and technology, research volume 3, ISSUE 12, December 2014, Page(s):126-131.