

## Project Report - intern id 380124

### Predicting the Asset price using Data Analysis and Machine Learning

#### Summary of Project : House Price Prediction

##### Motivation:

The motivation behind this internship was to gain practical experience in the field of data science and machine learning while working on a project with real-world applications. Choosing the "House Price Prediction" project allowed me to explore the significance of predictive modeling in the real estate domain, particularly in Bengaluru city, where housing prices are crucial for buyers, sellers, and investors.

##### Scope:

The scope of the project encompassed utilizing real estate data from Bengaluru city spanning four years. The dataset contained various features related to houses, including their square footage, the number of bedrooms and bathrooms, location, amenities, and more. The primary objective was to develop a predictive model that could estimate house prices based on these features. Additionally, the project involved building a user-friendly web application that enabled users to input house information and receive accurate price predictions.

##### Objectives:

1.Data Collection and Cleaning: The initial objective involved collecting relevant real estate data from Kaggle and ensuring its integrity by performing data cleaning and preprocessing. This step was crucial to ensuring the quality of the subsequent analysis.

Before processing:

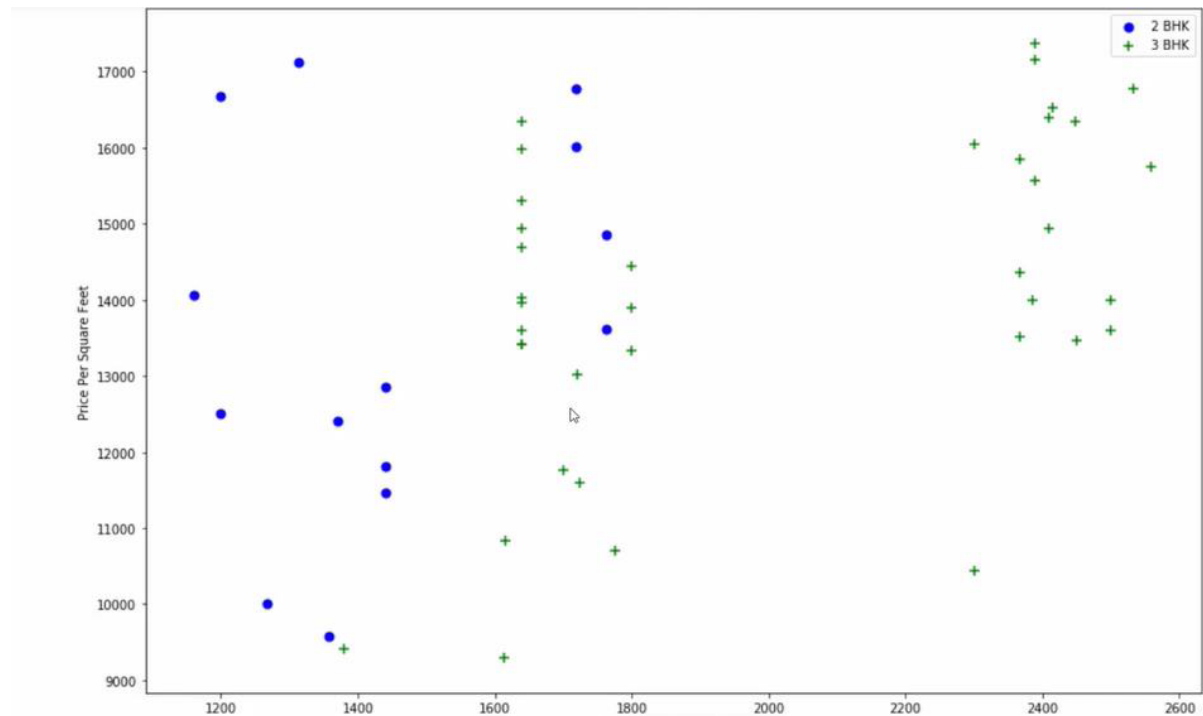
```
In [8]: df1.shape
Out[8]: (13320, 9)
```

After processing:

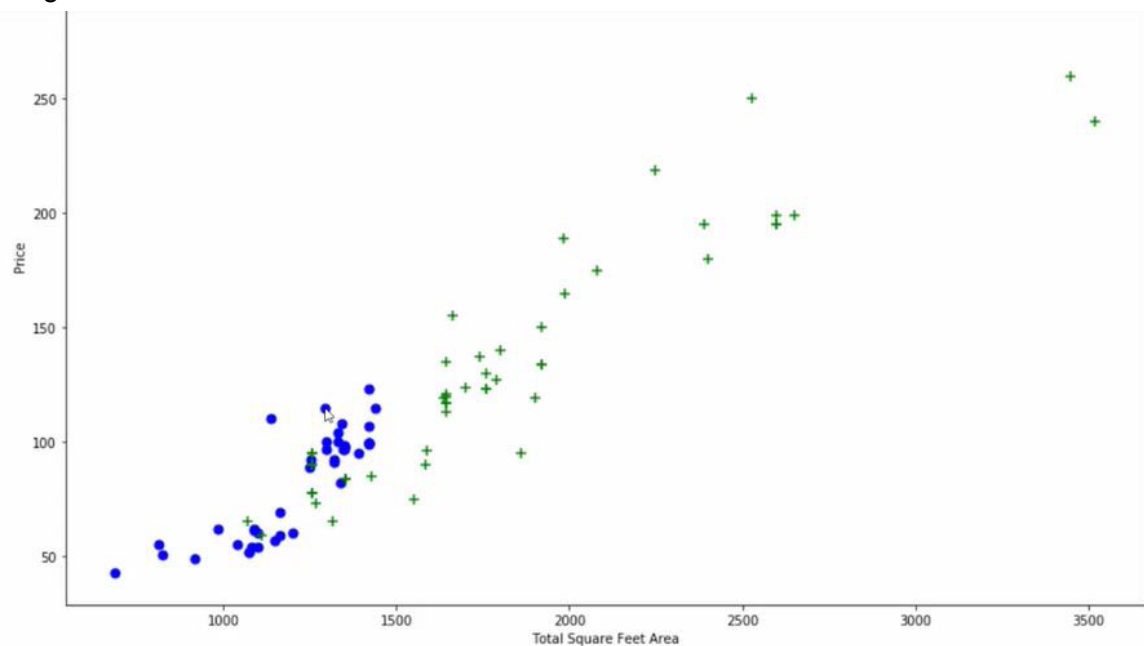
```
In [63]: df9 = df8[df8.bath<df8
          df9.shape
Out[63]: (7251, 7)
```

2.Feature Engineering: To enhance the predictive model's performance, the next objective was to engineer new features and transform existing ones. This process aimed to extract valuable insights and patterns from the raw data specific real estate market.

New feature :



Existing feature :



3. Model Selection and Training: The main focus was on selecting an appropriate machine learning algorithm for the Bengaluru housing data, with linear regression being chosen for this project. The model was then trained using the preprocessed data.

	model	best_score	best_params
0	linear_regression	0.818354	{'normalize': False}
1	lasso	0.687468	{'alpha': 2, 'selection': 'random'}
2	decision_tree	0.723130	{'criterion': 'mse', 'splitter': 'best'}

4. Model Testing: To assess the model's accuracy and effectiveness in the context of Bengaluru's real estate market, the project aimed to test its performance.

```
In [80]: def predict_price(location,sqft,bath,bhk):
          loc_index = np.where(X.columns==location)[0][0]

          x = np.zeros(len(X.columns))
          x[0] = sqft
          x[1] = bath
          x[2] = bhk
          if loc_index >= 0:
              x[loc_index] = 1

          return lr_clf.predict([x])[0]

In [81]: predict_price('1st Phase JP Nagar',1000, 2, 2)
Out[81]: 83.49904677167721
```

5. Web Application Development: The final objective was to create an intuitive web application using Flask, a web framework in Python. This application integrated the trained model, enabling users to input house features and obtain reliable price predictions tailored to Bengaluru's housing market.

