**Predicting Bangalore House Prices: A Data Science Journey**

When it comes to making one of the most significant investments of your life, buying a house, the stakes are high. Factors like location, size, and price play a pivotal role in making the right decision. In this data science project, I aimed to predict house prices in Bangalore using a dataset containing essential features such as location, size, total square footage, bedrooms, bathrooms, and, of course, the target variable - price. Along the way, I encountered the challenge of outliers and utilized feature engineering to enhance the model's performance. Let's explore this exciting journey!

**Project Overview**

Objective: To develop a machine learning model for predicting home prices based on input features such as location, no of bedrooms, no of bathrooms, square area feet and surge pricing.

Tools and Technologies: Python, Pandas, Scikit-Learn, Matplotlib

**Understanding the Dataset**

**Dataset Name:** banglore\_home\_price.csv

**Dataset Source:** Kaggle

Our journey begins with understanding the dataset. The data comprises several columns, each representing a crucial aspect of a house:

**Location:** The neighborhood or area where the house is located.

**Size:** The size of the house, typically indicated in terms of BHK (Bedroom, Hall, and Kitchen).

**Total Sq. Ft:** The total square footage of the property.

**Bedrooms:** The number of bedrooms in the house.

**Bathrooms:** The number of bathrooms in the house.

**Price:** The target variable, which is the price of the house.

**Data Preprocessing**

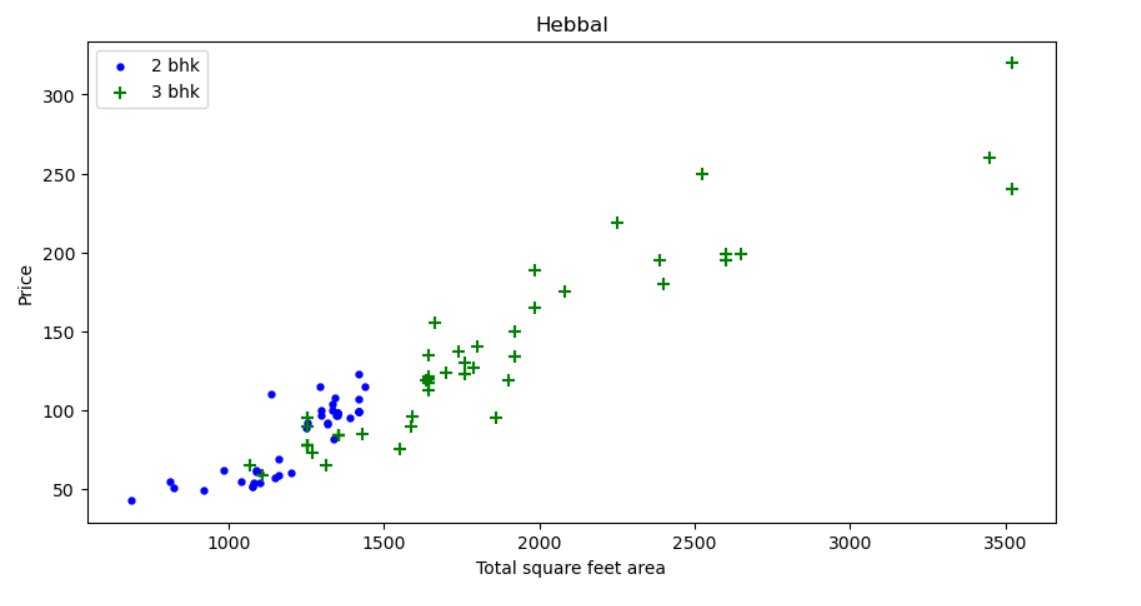
Before diving into modeling, we must ensure that our data is clean and well-preprocessed. Data preprocessing is a crucial step in any data science project. I meticulously handled missing values, encoded categorical variables (like 'Location') using techniques such as one-hot encoding, and scaled numerical features. This meticulous preparation ensures that our data is ready for model training.

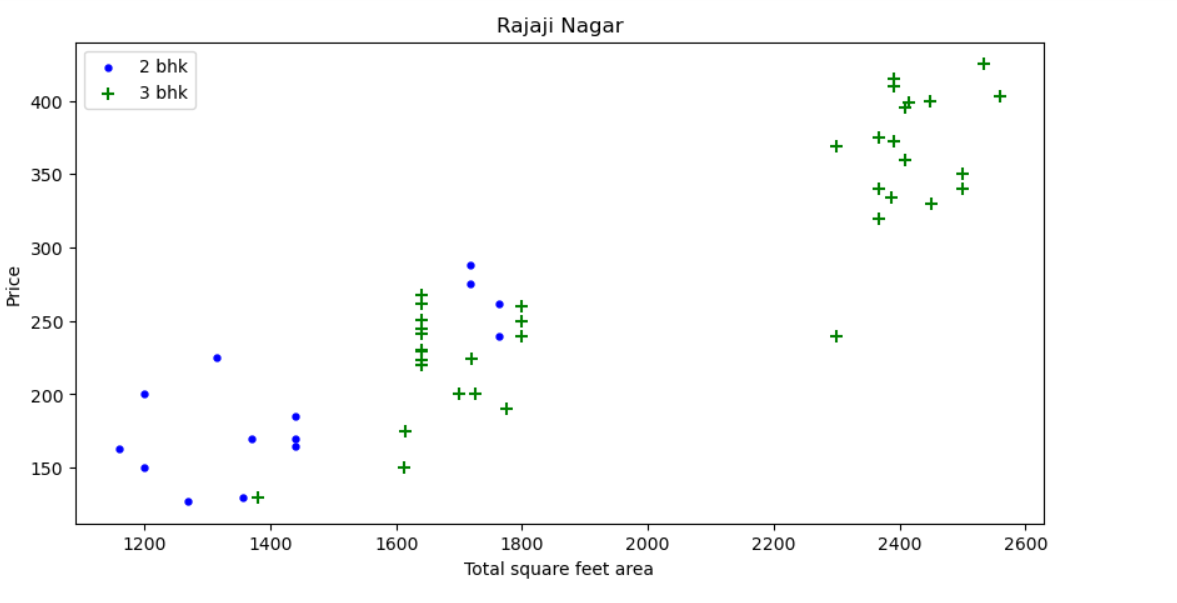
**Model Selection with Grid Search CV**

Choosing the right model is a pivotal decision in any data science project. To determine the best-fit model for this regression task, I employed Grid Search Cross-Validation (GridSearchCV). This technique systematically explores different hyper parameters of various regression models while evaluating their performance using cross-validation. After rigorous experimentation, the clear winner was the Linear Regression model, which consistently provided the highest accuracy for predicting house prices in Bangalore.

**Identifying Outliers**

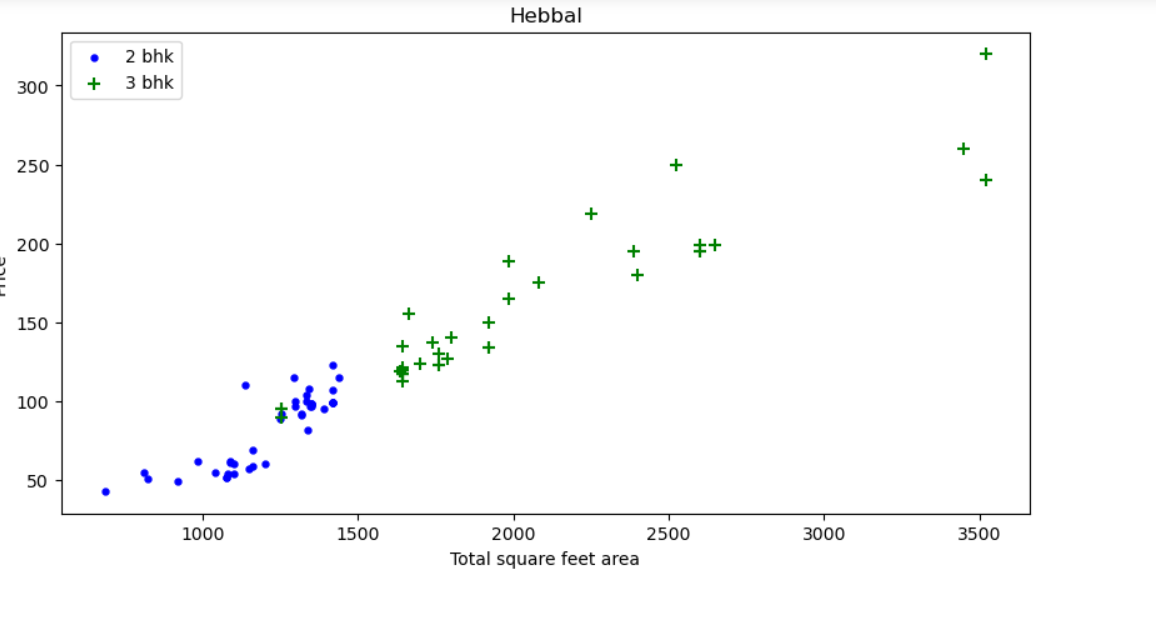
Before proceeding with feature engineering, it was crucial to identify these outliers. I employed various techniques such as box plots, scatter plots, and statistical methods to pinpoint the data points that deviated significantly from the norm. These outliers had the potential to distort the model's predictions and lead to inaccuracies.





**Handling Outliers with Feature Engineering**

One of the challenges I encountered during this project was dealing with outliers in the dataset. Outliers are data points that significantly deviate from the majority of the data, and if left unaddressed, they can adversely affect the model's performance.

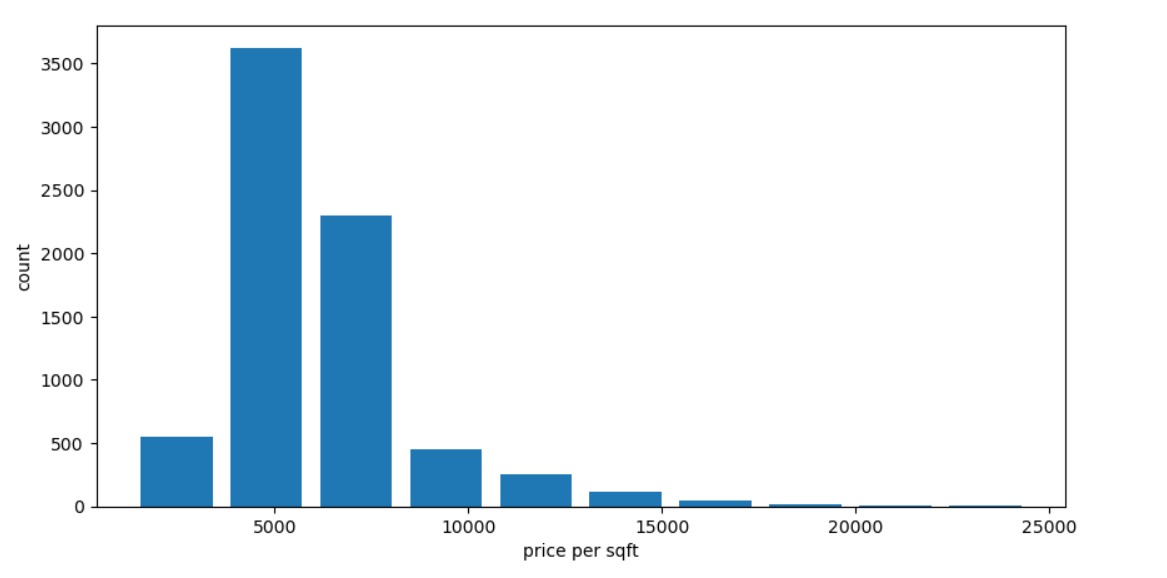
****

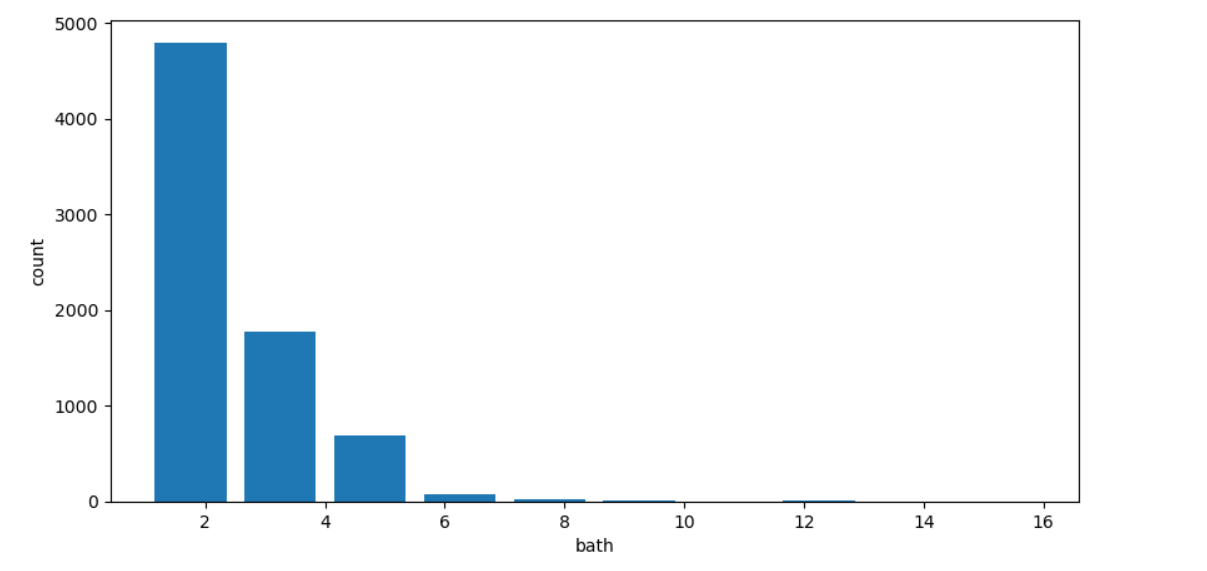
**The Impact on Model Performance**

After implementing these feature engineering techniques, I retrained the Linear Regression model and re-evaluated its performance. The results were remarkable. Removing or mitigating the influence of outliers led to a more stable model, resulting in more accurate and reliable house price predictions for Bangalore.

# Key Findings

While conducting data analysis and model development, I made several key findings:





**Model Evaluation**

With the model in place, I rigorously evaluated its performance. Metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R2) were employed to assess how well the model's predictions aligned with the actual house prices. The model exhibited impressive accuracy and showed its potential as a valuable tool for estimating house prices.

**Conclusion**

In this data science project, I embarked on a journey to predict house prices in Bangalore. Through meticulous data preprocessing, model selection via Grid Search CV, and innovative feature engineering techniques to handle outliers, I have crafted a powerful tool for real estate enthusiasts. Predicting house prices accurately is not just a data science endeavor but a valuable resource for anyone in the property market.

As I continue my data science journey, I remain inspired by the endless possibilities that data holds. Challenges like handling outliers only fuel my curiosity, propelling me to explore more advanced techniques and tackle even more complex projects in the future.

Thank you for joining me on this data-driven journey. I hope this blog post provides insight into the exciting world of data science and inspires your own data-driven adventures!

