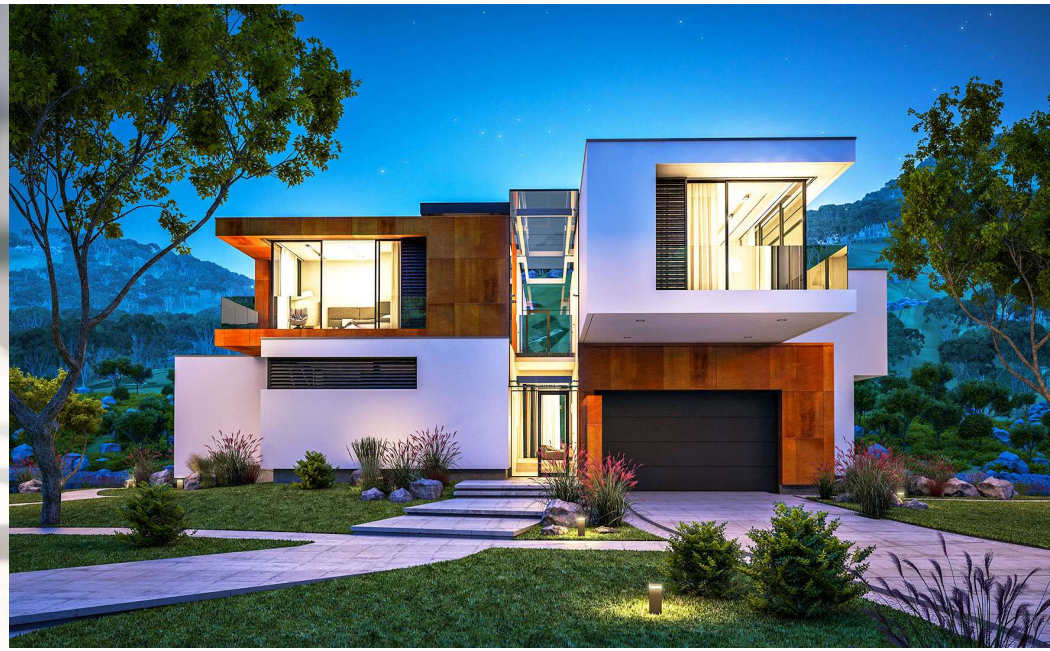
Real Estate Price Predication ML Model Project Report



Problem Statement

In the competitive real estate market, accurately predicting property prices is crucial for buyers, sellers, and investors. However, property prices are influenced by a variety of factors such as location, size, number of bedrooms and bathrooms. The challenge is to build a data-driven model that can analyze these factors and accurately estimate the market price of a property.

The objective of this project is to develop a **Machine Learning model** that can predict the **selling price of a real estate property** based on historical housing data. The model should consider key features such as:

By providing accurate price estimates, this model aims to assist real estate agents, property buyers, and sellers in making informed decisions and negotiating effectively.

Dataset Description

Bengaluru House Price Data

This dataset contains information about residential properties in Bengaluru, India, and is used to develop a real estate price prediction model. It includes **13,320 records with 9 features** describing various aspects of the properties.

Dataset Features:

- Area type
- Availability
- Location
- Area
- society
- Bath
- Bhk
- Balcony
- Price

Task

1. Data Collection
2. Data Preprocessing
3. Exploratory Data Analysis (EDA)
4. Model Selection
5. Model Evaluation
6. Challenges & Future Scope
7. Conclusion

Data Collection

🎯 Objective

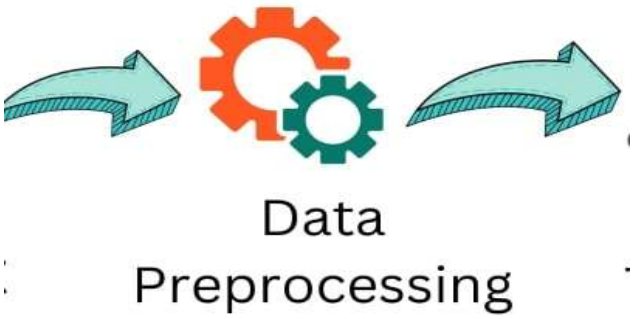
- Collect relevant real estate data that can be used to train a machine learning model to predict house prices accurately.

- **Kaggle Datasets** (e.g., “Bengaluru House Price Data”)

Data Preprocessing

Task: Prepare data for modeling

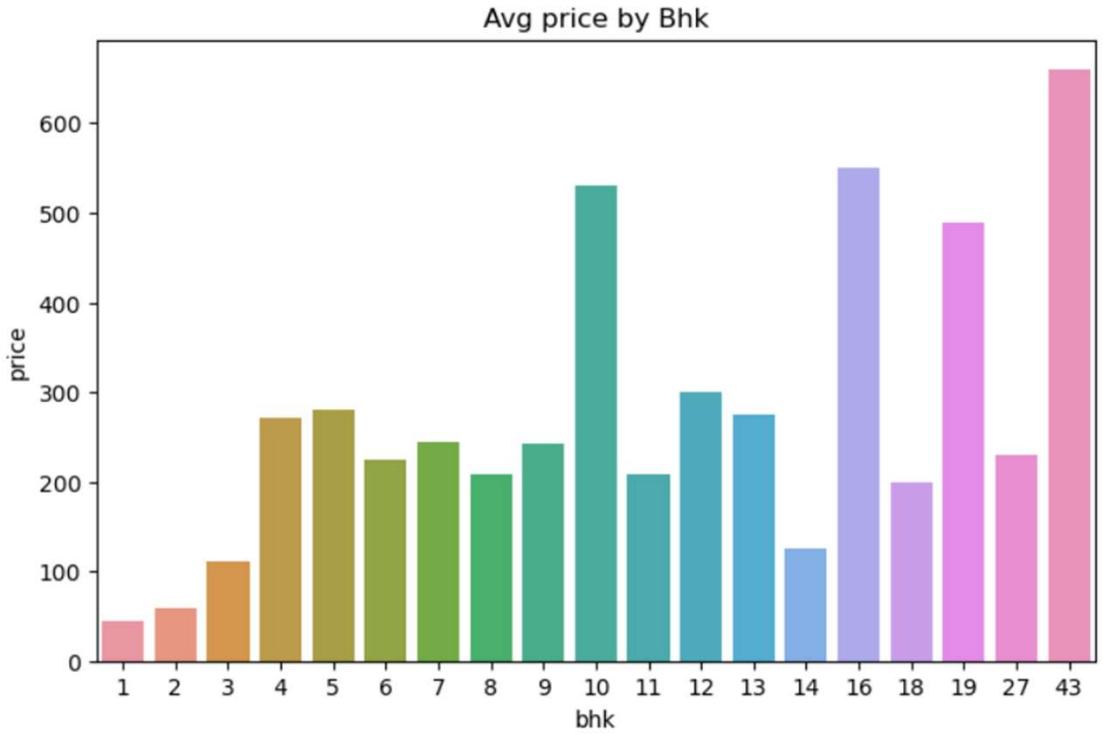
- Talk about handling missing values.
- Feature engineering: creating new features (e.g., price per square foot).
- Encoding categorical variables (like location).



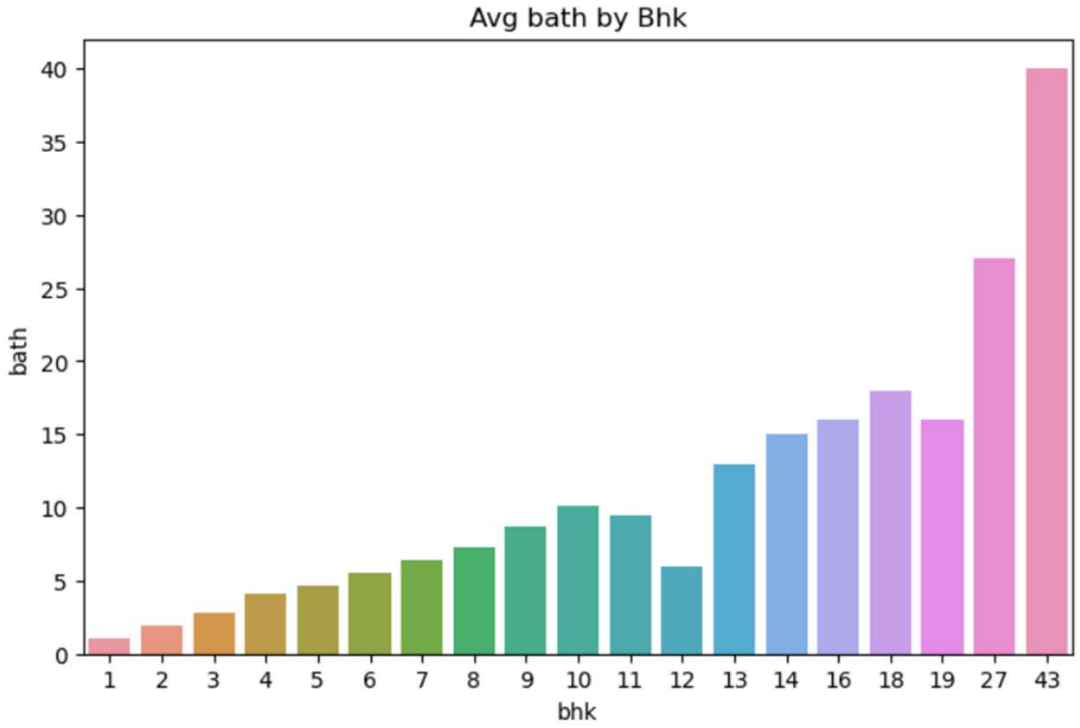
Exploratory Data Analysis

1. Average Price by BHK
2. Average Number of Bathrooms by BHK
3. Top 10 Locations by Avg Price per Square Foot
4. Correlation between variables

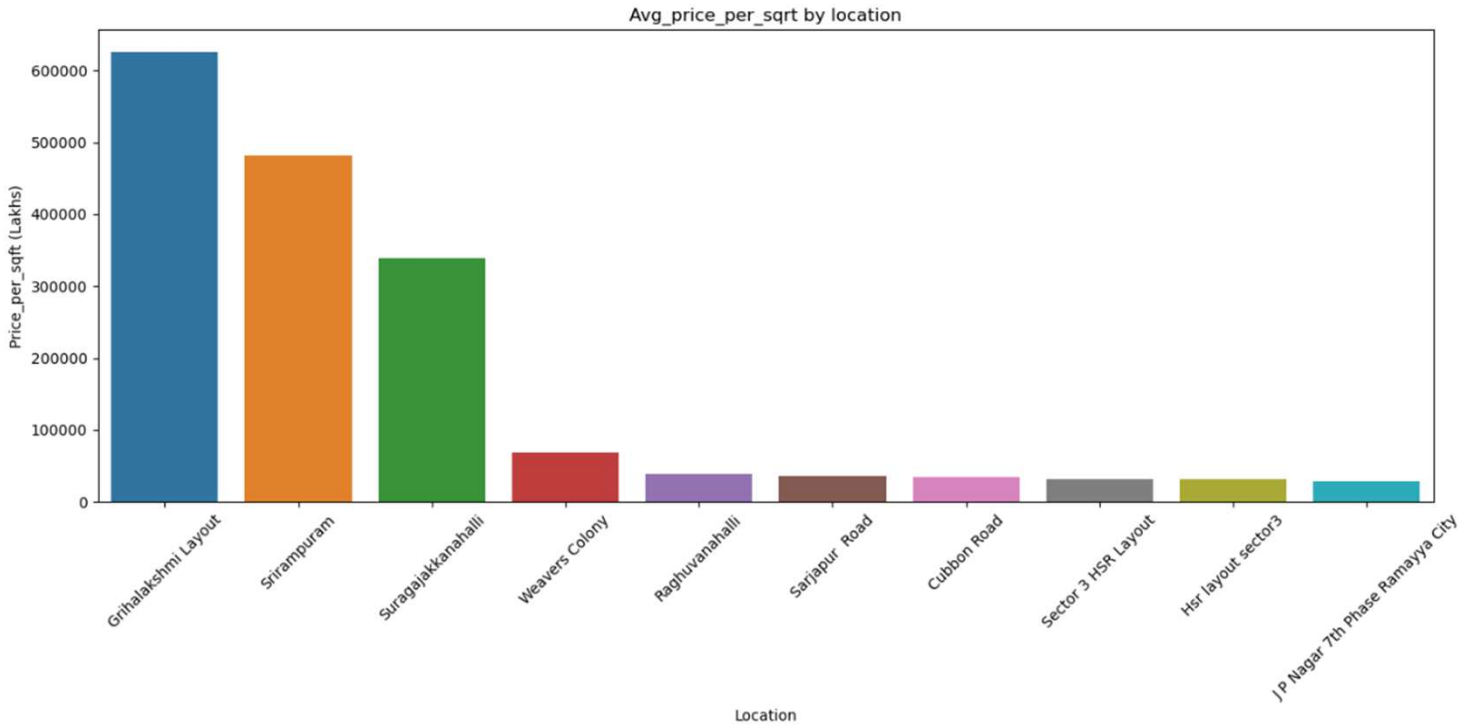
Average Price by BHK:



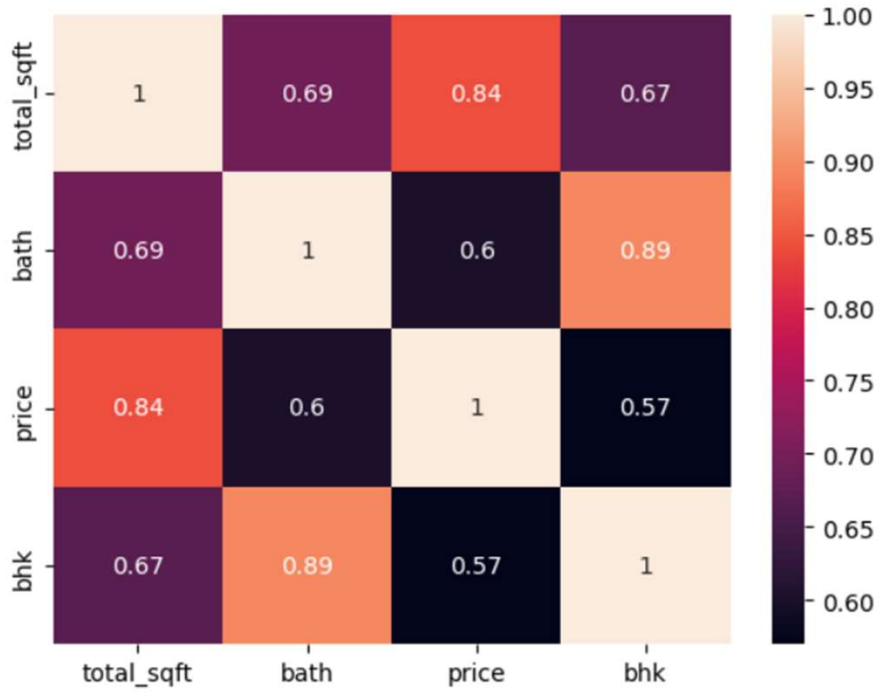
Average Number of Bathrooms by BHK:



Top 10 Locations by Avg Price per Square Foot:



Correlation between variables:



Model Selection

Objective:

To identify the most accurate and generalizable model for predicting real estate prices using **cross-validation**.

- Used **cross-validation score** from scikit-learn
- Ensures model performance is **consistent across different data splits**
- Helps avoid overfitting or underfitting issues

	model	Score
0	Linear_Regression	81.835408
1	Decision_tress	71.888199
2	Lasso_Regression	68.742866
3	SVR	55.857905

✅ **Linear Regression performed best overall** with the highest average score.

🏆 **Final Model Selected: Linear Regression**

- Simple yet effective for this dataset
- Performed consistently well in all folds
- Easy to interpret and deploy
- Lesser overfitting compared to tree-based models or SVR

Model Evaluation

🎯 **Objective**

- To assess the final model's performance using evaluation metrics and ensure it generalizes well to unseen data.

Training Set - 85% Accuracy

Testing Set – 84% Accuracy

🔍 **Insight**: Very small gap between training and testing scores.
✅ Indicates a well-balanced model with minimal overfitting.

Challenges & Future Scope

Project Challenges:

- Data inconsistencies and missing values
- Outliers in price and size (e.g., unrealistic listings)
- High-cardinality categorical feature: location
- Feature engineering for better prediction (e.g., location)
- Choosing the right model with good generalization

Future Scope:

- Add more features (amenities, property age, distance to city center)
- Deploy model as a web app (Flask/Streamlit)

Project Conclusion

- Built a **real estate price prediction model** using **Linear Regression.**
- Achieved **85% accuracy on training** and **84% on testing**, showing strong generalization.
- Performed **EDA** to extract key **insights** like average price per BHK and top locations.
- Used **cross-validation** to compare models and select the **best one**.
- **The model can assist buyers, sellers, and agents in making data-driven decisions.**

✅ **A simple yet powerful ML approach to bring transparency and predictability to the real estate market.**