

House Price Prediction Report

Problem Statement

Predicting house prices accurately is a critical task for real estate stakeholders, including buyers, sellers, and investors. The challenge lies in understanding the influence of multiple property features on pricing, such as size, number of bedrooms, bathrooms, and additional amenities. This project aims to develop a machine learning model capable of estimating house prices based on various property characteristics, thereby assisting stakeholders in making informed decisions.

Objectives

The main objectives of this project are to build a predictive model for house prices, analyze the impact of different property features on pricing, and provide a user-friendly application to estimate house prices in real time. Specifically, the project aims to:

1. Preprocess and explore a comprehensive house price dataset.
2. Train regression models to predict prices, including Linear Regression and Gradient Boosting.
3. Evaluate model performance using metrics like Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).
4. Deploy a web application with a modern and interactive interface for users to predict house prices.

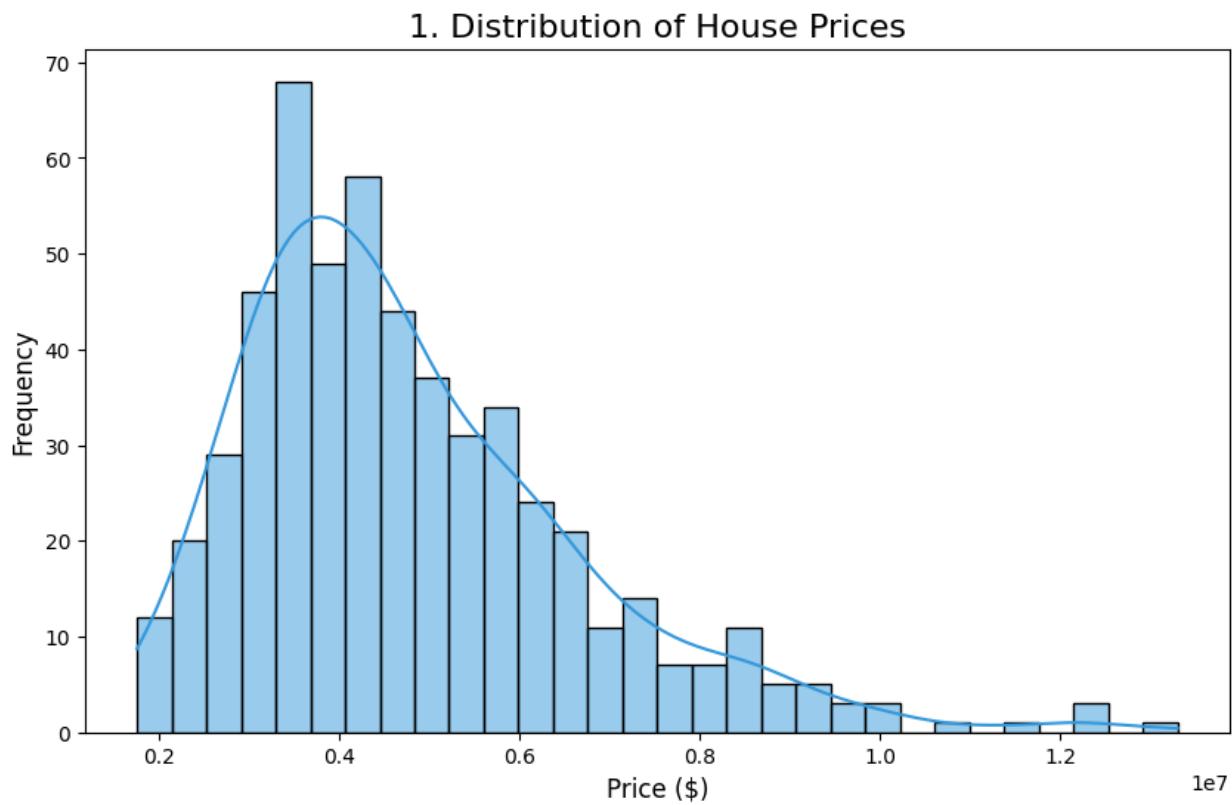
Dataset Description

The dataset used for this project was collected from Kaggle and contains detailed information for predicting house prices, with **13 key features**. The target variable is the house **Price**, while input features include **Area** (total house area in square feet), **Bedrooms**, **Bathrooms**, **Stories**, and additional binary features such as **Mainroad**, **Guestroom**, **Basement**, **Hot water heating**, **Airconditioning**, **Parking**, **Prefarea**, and **Furnishing status** (Fully Furnished, Semi-Furnished, Unfurnished). The dataset provides a diverse range of property types and configurations, enabling the model to capture complex relationships between features and house prices.

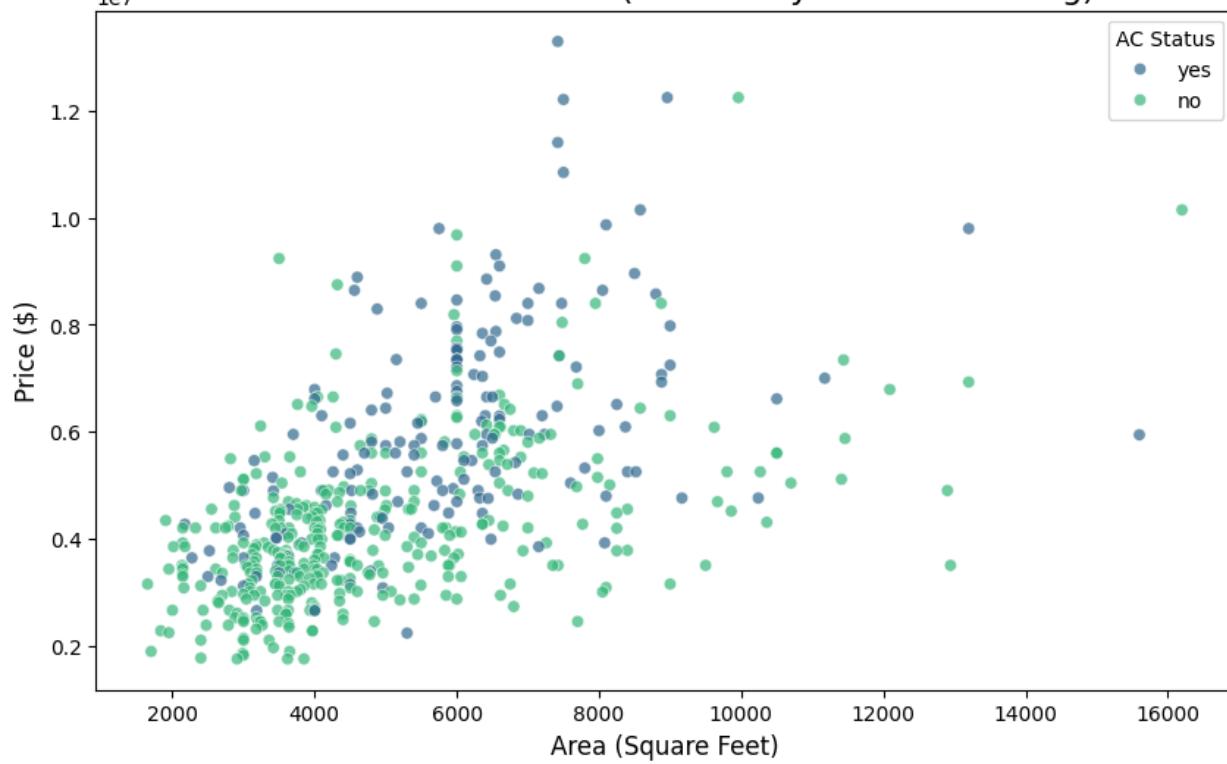
Methodology

The methodology of this project involved several steps. First, the dataset was preprocessed to handle missing values and scale numeric features such as area, bedrooms, bathrooms, stories, and parking. Categorical features like furnishing status were converted into one-hot encoded variables to ensure compatibility with regression models. Two machine learning models were trained: **Linear Regression** for baseline predictions and **Gradient Boosting Regressor** for improved performance. Model evaluation was performed using **MAE**, **RMSE**, and **R² score** to measure prediction accuracy. After model development, a **Flask web application** was created with a modern, aesthetic user interface, including sliders for numeric features, checkboxes for amenities, and radio buttons for furnishing status. The app allows users to input property characteristics and obtain instant price predictions.

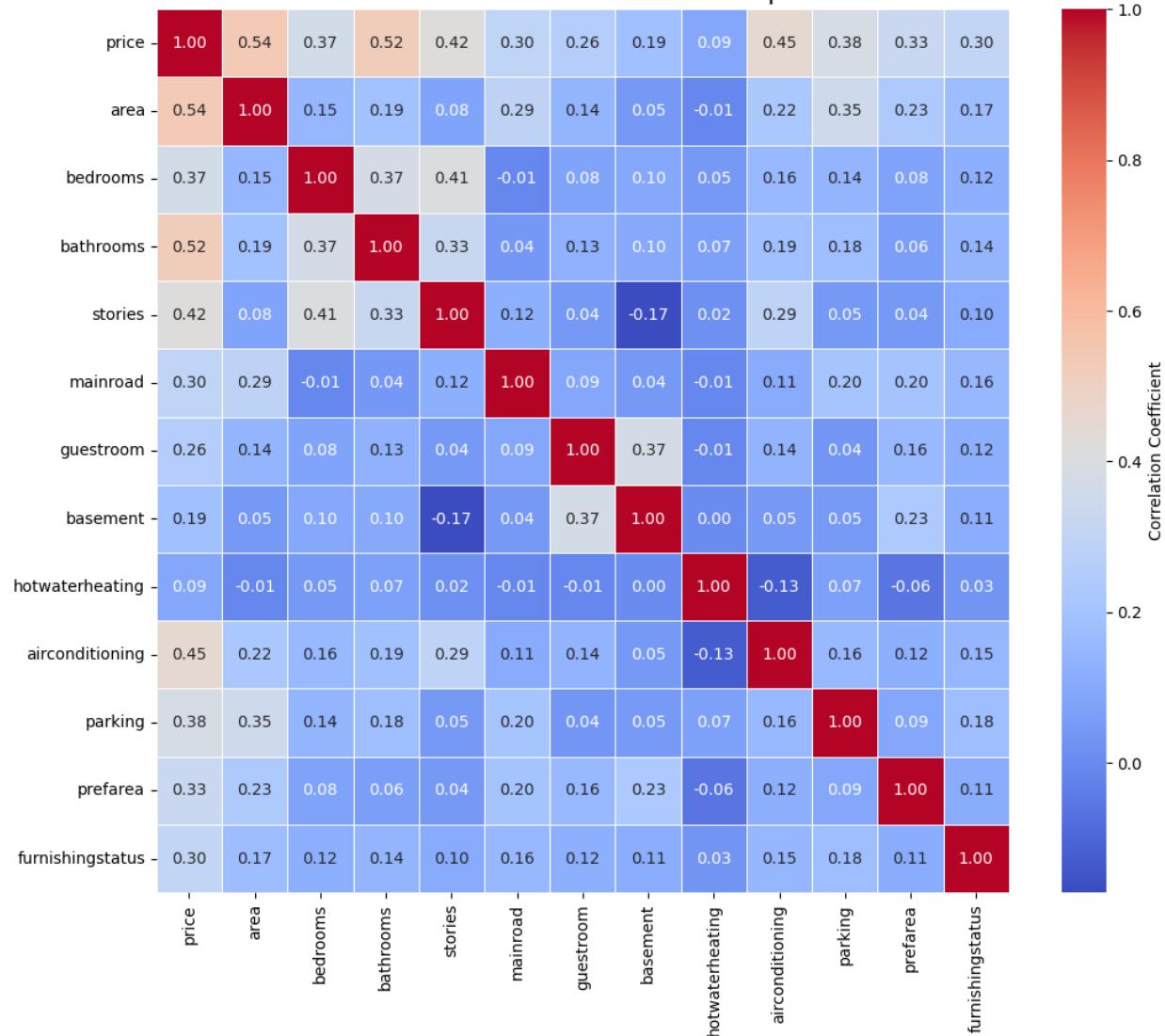
Results



2. House Area vs. Price (Colored by Air Conditioning)



3. Feature Correlation Heatmap



```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
import numpy as np

# 1. Initialize and train the model
lr_model = LinearRegression()
lr_model.fit(x_train, y_train)

# 2. Make predictions
y_pred_lr = lr_model.predict(x_test)

# 3. Evaluate the model
mae_lr = mean_absolute_error(y_test, y_pred_lr)
rmse_lr = np.sqrt(mean_squared_error(y_test, y_pred_lr))
r2_lr = r2_score(y_test, y_pred_lr)

print("Linear Regression Performance on Test Set:")
print(f"MAE: ${mae_lr:.2f}")
print(f"RMSE: ${rmse_lr:.2f}")
print(f"R^2: {r2_lr:.4f}")

Linear Regression Performance on Test Set:
MAE: $1,048,755.01
RMSE: $1,484,717.82
R^2: 0.5418
```

The trained models provided satisfactory predictions, with Linear Regression achieving an **R² score of approximately 0.54** and Gradient Boosting yielding improved accuracy. Sample predictions showed that the models could capture the trend of house prices effectively, although exact prediction values varied depending on feature combinations. Hyperparameter tuning and the inclusion of additional relevant features, such as location quality or year built, were suggested as future improvements to further enhance model performance.

House Price Predictor

Area (sqft): 500

Bedrooms: 2

Bathrooms: 3

Stories: 2

Parking: 1

Facilities:

- Main Road
- Guest Room
- Basement
- Hot Water Heating
- Air Conditioning
- Preferred Area

Furnishing Status:

- Unfurnished
- Semi-Furnished
- Furnished

Predict Price

House Price Predictor

Area (sqft): 500

Bedrooms: 1

Bathrooms: 1

Stories: 1

Parking: 0

Facilities:

- Main Road
- Guest Room
- Basement
- Hot Water Heating
- Air Conditioning
- Preferred Area

Furnishing Status:

- Unfurnished
- Semi-Furnished
- Furnished

Predict Price

Predicted House Price:
\$5048971.64

Conclusion

This project successfully developed a predictive model for house prices using a comprehensive dataset and deployed it as an interactive web application. By leveraging machine learning techniques and a modern UI design, the solution enables users to estimate house prices accurately and efficiently. The project demonstrates the importance of feature selection, preprocessing, and model evaluation in predictive analytics, while also highlighting potential future enhancements to improve prediction accuracy and expand applicability in real-world real estate scenarios.