House
Price
Predictions

HOUSE FOR SALE

Price From

Kes 9,000,000

+25411852860

HOME FEATURES

- 2 Living Room
- 4 Bedroom
- 6 Bathroom
- 1 Garage
- 3 Floors
- Swimming Pool
- 2 Balcony

Martin Dudi

martindudi6@gmail.com
+254711852860

Introduction

Welcome to our presentation on predicting house prices using a data-driven approach. The goal of this project is to analyze house sales data using multiple regression modelling techniques. By employing multiple regression, we aim to identify and quantify the relationships that exists between various factors or predictors and house sale prices. This analysis will provide insights into the key drivers of house prices and help shareholders make informed decisions related to real estate investment. In addition, by leveraging data analysis, we aim to provide valuable insights for buyers, sellers, and real estate professionals.

Problem Statement

This project addresses the challenge of identifying the key features that significantly impact house prices and developing a reliable model for accurate predictions. By analyzing the patterns of our dataset we aim to accurately estimate house prices. This is a complex task due to the multitude of factors that influence property values. Inconsistent pricing can lead to financial losses for both buyers and sellers, as well as a lack of transparency in the real estate market. Hence we aim to generate the best model for price estimation.

Main Objective

Develop a robust regression model to accurately predict house prices. This model will leverage data analysis techniques to provide accurate estimations, improving transparency and decision-making in the real estate market.

Specific Objectives

1. Conduct explanatory data analysis to gain insights on the relationships between different variables and target variable, assisting in selection of relevant variables for regression model.
2. Develop multiple regression model to predict house sale prices, considering the selected independent variables and their impact on the dependent variable. Validate the model assumptions, assess its goodness of fit and refine the model if necessary.
3. Interpret the coefficients of the independent variables in the model to determine their individual impact on house prices, identifying the most influential factors driving thr house sales prices and their respective effects.
4. Evaluate and validate the performnance of the model.
5. Provide actionable insights and recommendations based on the analysis to assist real estate investors, and policymakers in making informed decisions regarding property investment, market trends, and economic planning.

Notebook Structure

1. Reading Data.
2. Data Cleaning and Preprocessing.
3. Model Evaluation and Understanding.
4. Results, Presentations and Conclusions.
5. Recommendations.
6. References.

Data Understanding

The dataset used in this project contains information about the factors affecting the housing prices including variables such as date, sqft_above, view and sqft_basement.

Explanatory data analysis will be used to get a clear understanding of the dataset including the missing values, checking the data types, identifying outliers and also extracting relevant features for analysis.

Business Understanding

Real estate agents can benefit from the regression analysis by understanding the factors that influence house prices. They can utilize the regression model to provide more accurate price estimates to their clients, identify key features to highlight in property listings, and make informed recommendations to buyers and sellers. We could also consider Homeowners and sellers who can use the regression model to estimate the potential value of their properties.

By considering the influential features identified in the regression analysis, they can assess the impact of certain improvements or renovations on the property's price and make informed decisions regarding pricing and marketing strategies.

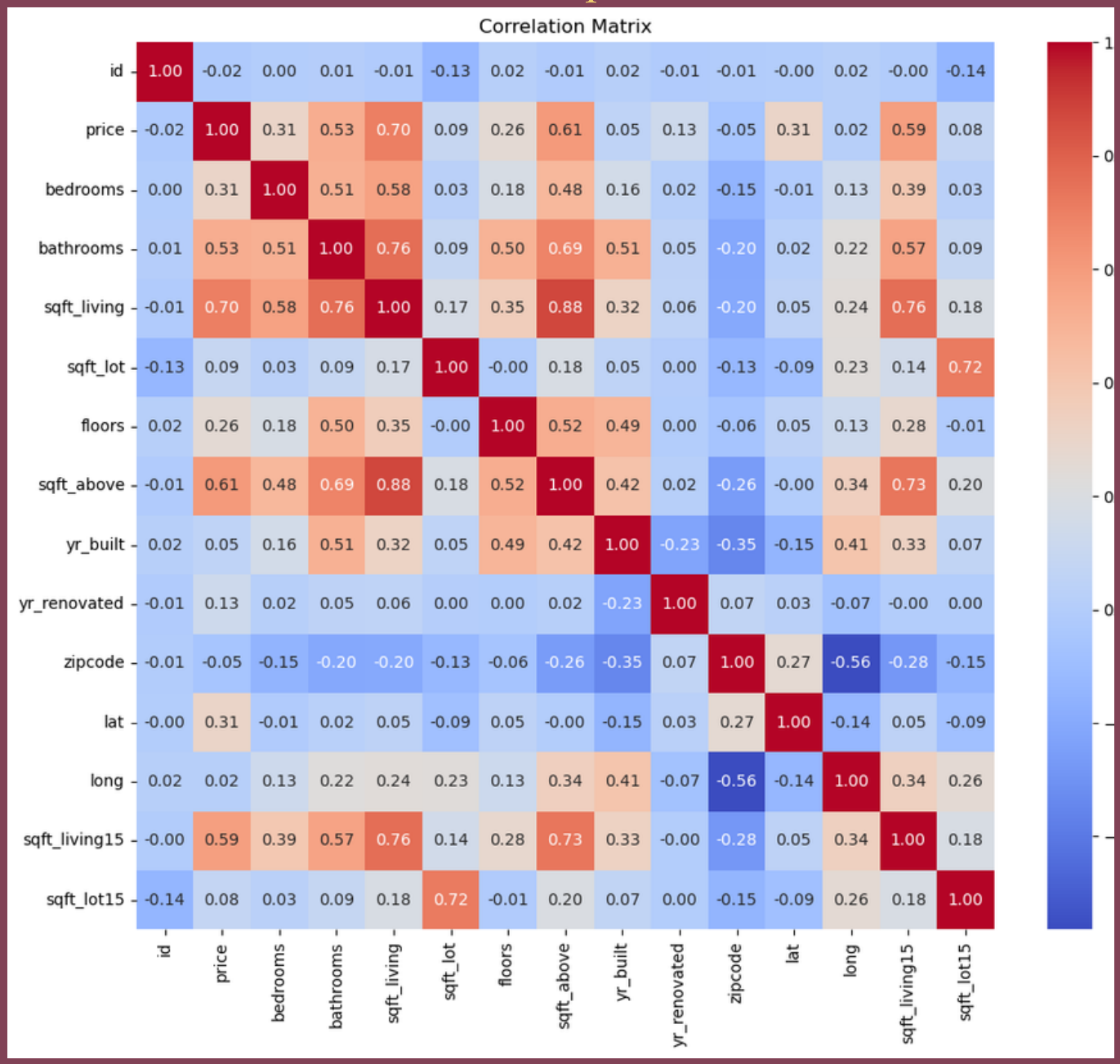
Methodology

The methodology provides a structured approach to conduct the analysis using multiple regression modelling. It ensures appropriate handling of data, selection of relevant variables, model development, and validation to achieve accurate and reliable insights into the factors impacting house sale prices.

Explanatory

Data Analysis

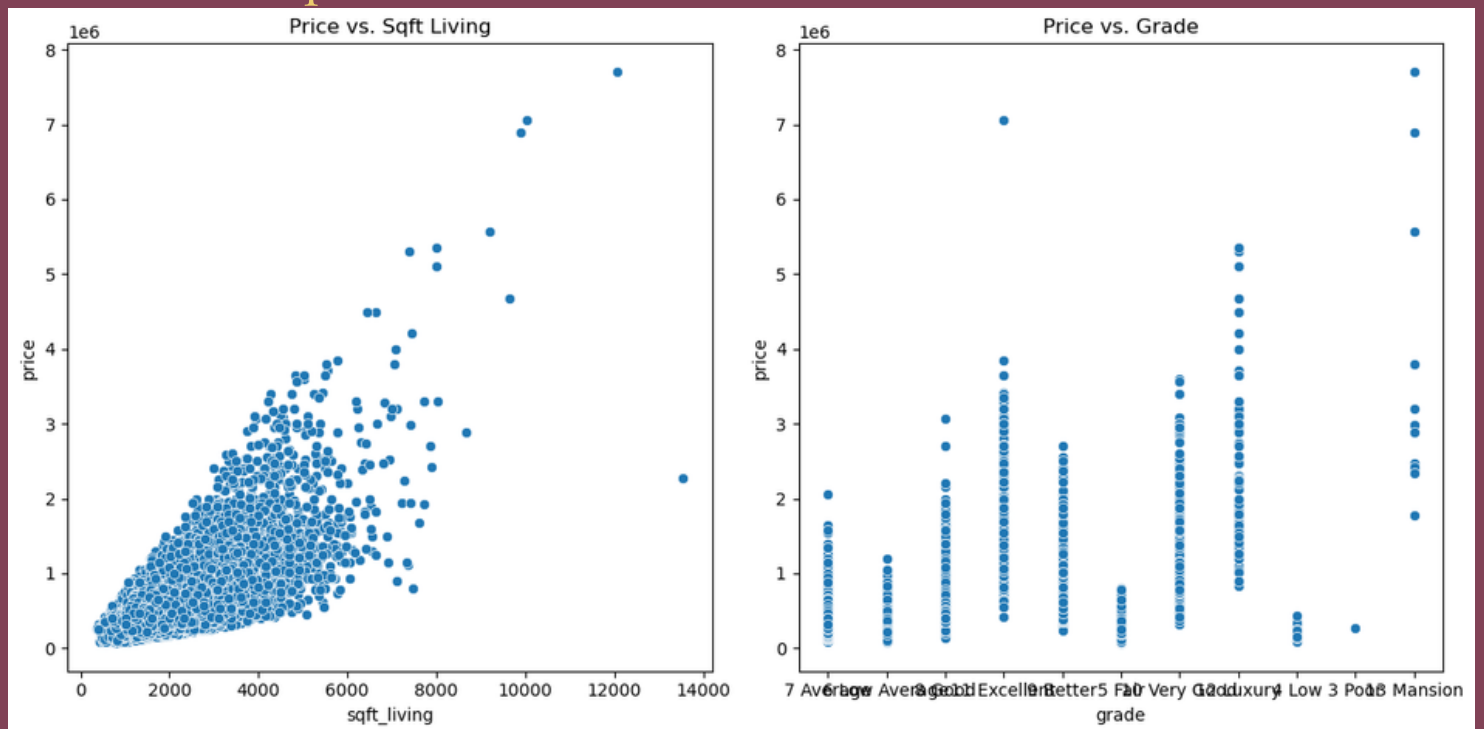
Correlation Matrix Heatmap



sqft_living sqft_above 0.876448
sqft_above sqft_living 0.876448
sqft_living sqft_living15 0.756402
sqft_living15 sqft_living 0.756402
bathrooms sqft_living 0.755758
sqft_living bathrooms 0.755758
These are the variables with the highest correslation between them

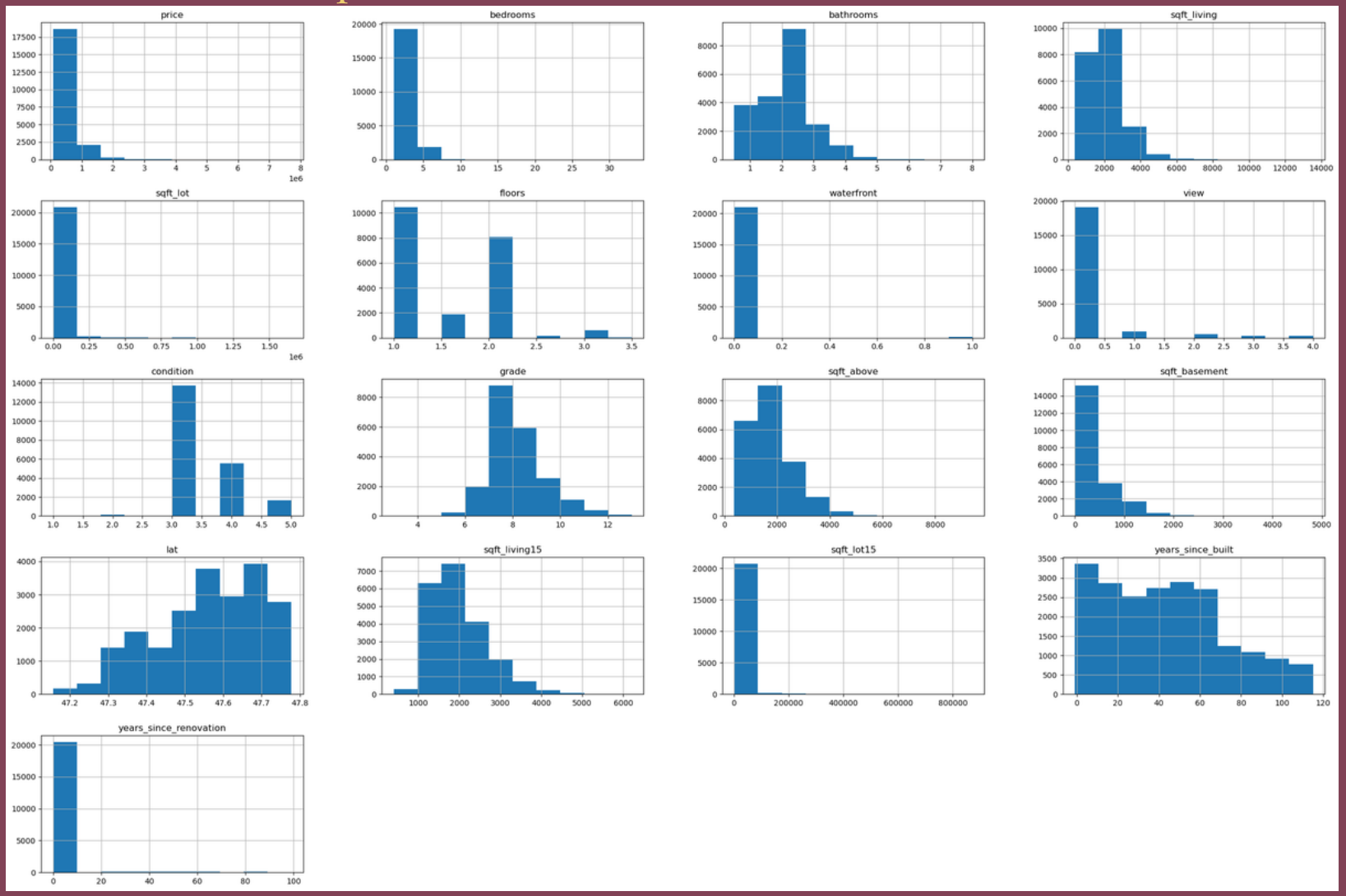
This will check for multicollinearity within independent variables. High multicollinearity might lead to poor performance of our model. In this model we dealt with collinearity greater than 0.75 as they could make us draw inaccurate conclusions.

Relationship between variables and the Price



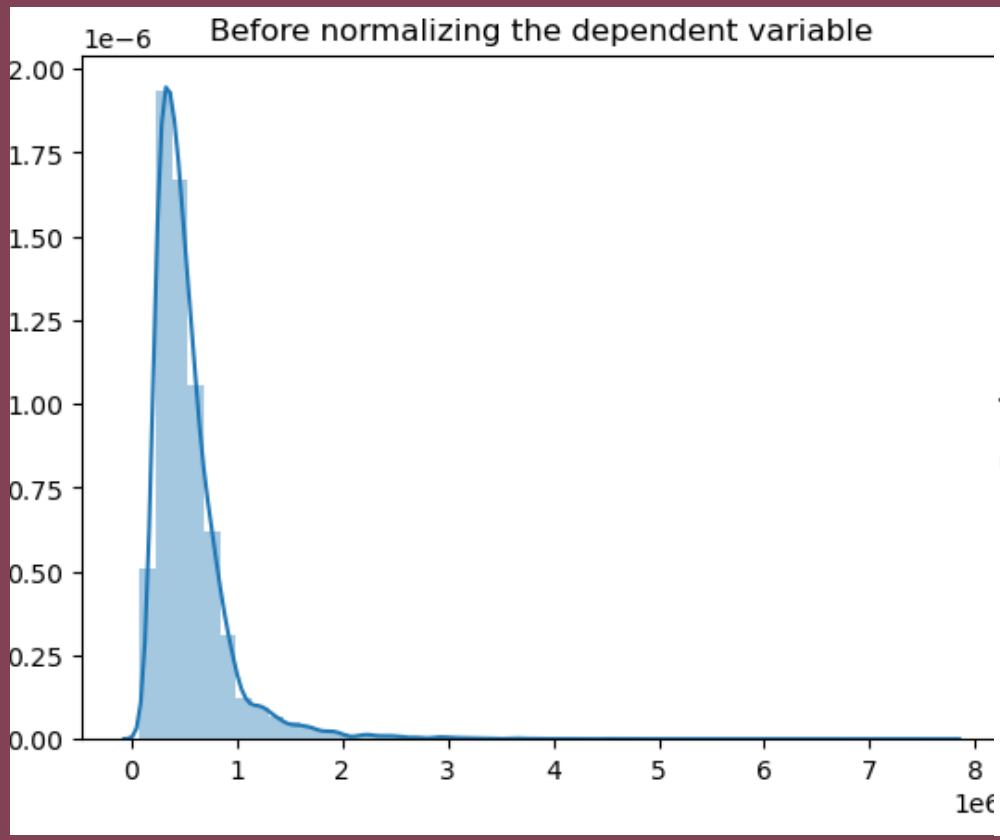
Here we visualize the distribution of predictor variables against our target variable which is the price of houses

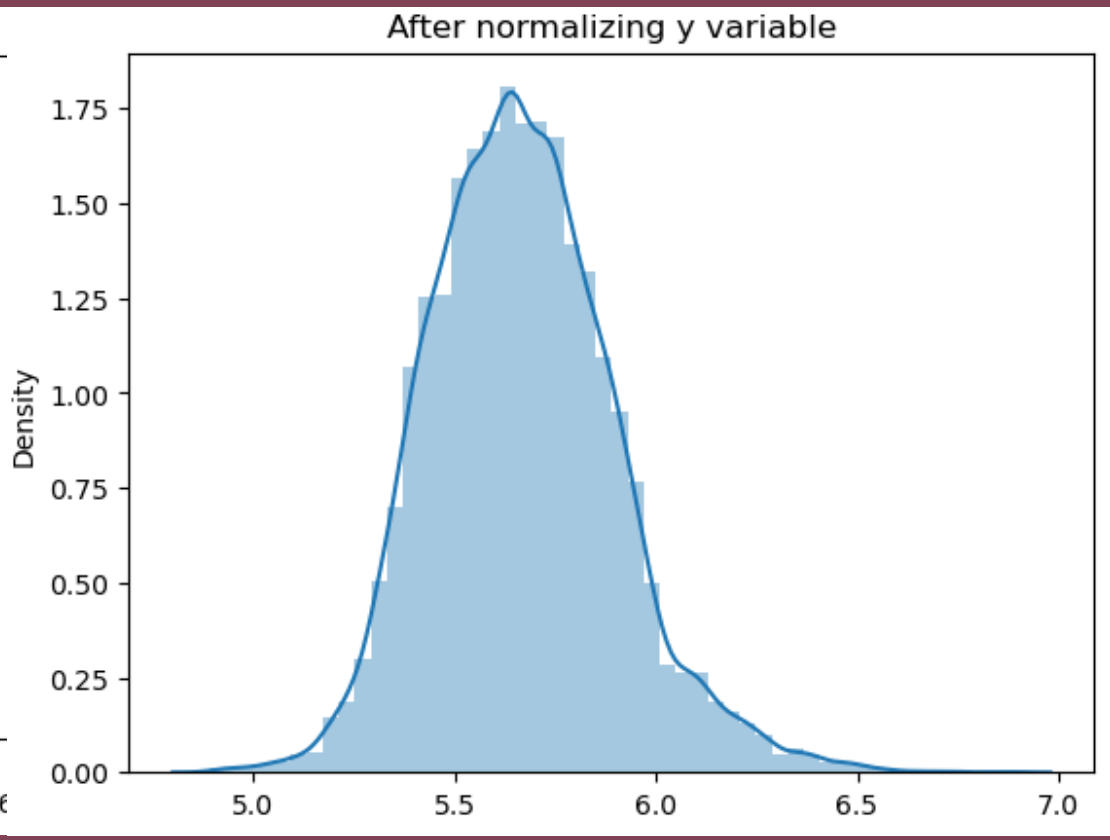
Distribution of predictor variables



We clearly see the how our predictor variables are distributed. Our desire is to generate normally distributed variables.

Visuals Of Price Before and After Normalization





Before normalization our price had a heavy positive tail.

After the process of normalization our distribution is bell-shaped suggesting a normal gaussian distribution.

Regression Model

House Price = β0 + β1 * bedrooms + β2 * bathrooms + β3 * sqft_living + β4 * sqft_lot + β5 * floors + β6 * grade + β7 * sqft_above + β8 * yr_built + β9 * lat + β10 * long

In this project, we developed a multiple linear regression model to predict house sale prices based on various independent variables. The model was trained using the scikit-learn library and applied to the house sales dataset.

The main independent variables included in the model were: sqft_living, bedrooms, bathrooms, location, amenities, and property grade. Before fitting the model, we performed data cleaning and preprocessing steps to handle missing values, outliers, and feature scaling.

Regression Results

The model achieved a train score of 0.769 and a test score of 0.762, indicating a good fit to the data and reasonable predictive performance. The mean absolute error (MAE) on the test set was 0.085, the root mean squared error (RMSE) was 0.110, and the mean squared error (MSE) was 0.012.
The regression coefficients of the independent variables in the model provide insights into their impact on house prices. The coefficient values can be interpreted as follows:

- Feature 1: -0.0040840740985575085 , Feature 2: 0.023083633186190923 Feature 3: 0.026759941440479003 , Feature 4: 0.008299695063583424 Feature 5: 0.018114213173503237, Feature 6: 0.01283514616127203 8 Feature 7: 0.020591366148559512 , Feature 8: 0.019519116795249102 Feature 9: 0.08327837466814307 , Feature 10: 0.021115008017844016 Feature 11: 0.016019043416200144, Feature 12: 0.08143985186453419 Feature 13: 0.03226939467353841, Feature 14: -0.002196191805659195 Feature 15: 0.04296425107804618 , Feature 16: 0.006934357433041017

- Feature 1: -0.004 - This indicates that a one-unit decrease in feature 1 results in a decrease of 0.004 in the predicted house price.
- Feature 2: 0.023 - A one-unit increase in feature 2 leads to an increase of 0.023 in the predicted house price.
- Feature 3: 0.027 - A one-unit increase in feature 3 contributes to a 0.027 increase in the predicted house price. ...

Recommendations & Next Steps

1. Feature Enhancement: Consider enhancing or upgrading the features that positively affect house prices. For example, increasing the square footage of the living area, improving the overall grade of the property, or adding more bathrooms can potentially increase the value of the house.
2. Feature Importance: Analyze the regression coefficients to identify the most influential features on house prices. Focus on the features with higher coefficients, such as 'sqft_living', 'grade', 'bathrooms', and 'sqft_above', as they have a stronger impact on the predicted prices.
3. Price Prediction: Utilize the regression model to predict house prices based on the given set of independent variables. This can be useful for estimating the selling price of houses or determining the potential value of a property.
4. Data Collection: Consider collecting additional relevant data that could improve the accuracy of the regression model. This may include variables such as location-specific factors, proximity to amenities, property age, or neighborhood characteristics.
5. Market Segmentation: Analyze the relationship between the independent variables and house prices to identify market segments or specific buyer preferences. For instance, if higher-grade houses tend to have higher prices, it may indicate a market segment of luxury or high-end properties.
6. Further Analysis: Perform additional exploratory data analysis and feature engineering to uncover additional patterns and insights. Consider examining correlations, visualizing distributions, or conducting hypothesis testing to deepen the understanding of the data.
7. Model Evaluation: Continuously evaluate and monitor the performance of the regression model. Assess the R-squared values, MAE, RMSE, and MSE to gauge the model's predictive accuracy and potential areas of improvement

THANKYOU