

**AUTHOR: ANITA
MAKORI**

HOUSE SALES PREDICTION



TABLE OF CONTENTS

- Introduction
- Problem Statement
- Main objective
- Specific Objective
- Notebook Structure
- Data Understanding
- Methodology
- Conclusions
- Recommendations

INTRODUCTION

The business problem at hand is the need to provide homeowners with accurate and reliable advice regarding home renovations and their potential impact on the estimated value of their homes. The stakeholder aims to assist homeowners in making informed decisions about renovating their properties, enabling them to maximize their investment returns while improving their living spaces.

To address this problem effectively, the stakeholder must gather relevant data and insights, analyze the factors that influence home value, and develop a comprehensive advisory framework. This framework should consider various renovation projects, their costs, and their potential impact on the estimated value of homes.

PROBLEM STATEMENT

The stakeholder has a critical business issue to solve: how to give homeowners reliable, individualized recommendations about house improvements and how such changes can affect the assessed worth of their properties. Currently, homeowners are unable to rely on trustworthy information about the extent to which and how much a certain restoration project will raise their property value.

The idea is to create a solution that makes use of data-driven insights and industry knowledge to provide thorough guidance to homeowners, aiding them to understand the potential value increase associated with various remodeling options and make decisions that are in line with their objectives and financial considerations.

MAIN OBJECTIVE

To give homeowners specific and reliable information on how and by how much home improvements can raise the **projected value** of their houses.

This goal strives to arm homeowners with the knowledge they need to choose renovation projects that maximize **return on investment** while raising the value of their homes.

SPECIFIC OBJECTIVES

To **identify** how to develop strategies that enhance the model's predictive power, considering additional predictor variables that may impact house prices, such as location-related factors or specific property features.

To **assess** properties with high market value. This can involve conducting thorough inspections and evaluations, considering factors like maintenance history, upgrades, and overall quality. As well as leveraging geospatial data or collaborating with real estate agents specializing in waterfront properties.

To **evaluate** the market trends and buyer preferences. Regularly monitor market trends, including changes in buyer preferences and demands, to adapt pricing strategies accordingly.

NOTEBOOK STRUCTURE

- Loading the Data
- Exploratory Data Analysis
- Modeling and Results
- Visualization of the Model
- Conclusions
- Recommendations

DATA UNDERSTANDING

This project will utilize datasets that consist of details regarding a house such as **view**, **condition**, **number of bedrooms** and **bathrooms**, and whether a house has a **lot** and a **basement**.

To gain a comprehensive understanding of the dataset, **explorative data analysis** techniques will be employed. These techniques involve addressing **missing values**, **verifying data types**, and **converting data types** to those suitable ones for creating a model.

METHODOLOGY

In this project, data analysis and visualization techniques in Python, utilizing libraries such as **pandas**, **numpy**, and **matplotlib**, will be employed.

The available dataset will be examined through **analysis and visualizations**. This process will help identify key relationships and trends, enabling the extraction of insights related to house sales prediction

EXPLORATORY DATA ANALYSIS

In this project, data visualizations were used to explore this data and aid in coming up with models that played a role in coming up with recommendations.

Such as: **histograms** to show **residuals** from the model, **heatmaps** to show **correlation**, **scatter plots** to show **actual vs predicted**, and others

MODELING AND RESULTS

We used **Ordinary Least Squares (OLS)** regression as the modeling technique to analyze the relationship between the **predictor** variables and the **target** variable (**house prices**). OLS is a widely used and fundamental method for linear regression analysis.

Why use OLS??

Find out in the next slide!!

MODELING AND RESULTS

Ordinary Least Squares (OLS) regression is suitable for modeling linear relationships and provides estimates for the **coefficients of predictor variables**, allowing for easy interpretation.

It also enables **statistical inference** by assessing the significance of predictors through p-values. OLS is an efficient method as it minimizes the sum of squared residuals, providing reliable estimates when the assumptions are met.

Overall, OLS facilitates **understanding, interpretation, and inference** in studying how **changes in predictors relate to house prices**.

RESULTS SUMMARY

The regression model indicates that the number of **bedrooms** has a **negative impact** on house prices, while the number of **bathrooms**, **square footage of living space**, and the **presence of a waterfront view** have positive effects.

Other factors like lot size, number of floors, condition, grade, and square footage above ground also influence prices.

The model is **statistically significant**, meaning it is reliable for predicting house prices based on these factors.

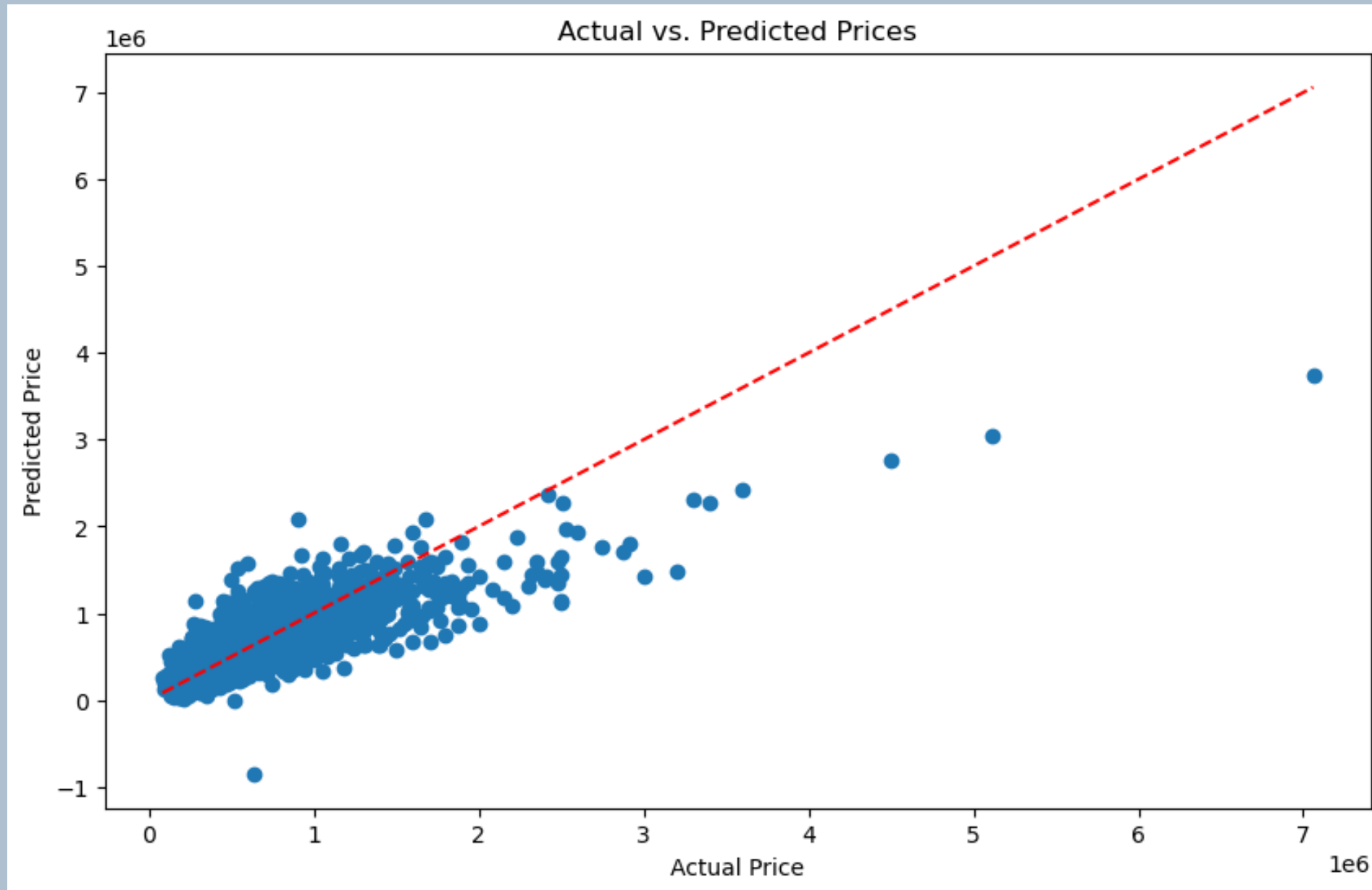
However, it's important to consider that these interpretations assume other variables remain constant and are based on statistical analysis.

DATA VISUALIZATIONS

A visualization of the model's results will be plotted so as to ensure a better understanding of the predictive analysis. A scatter plot of the target variable, a histogram of the residuals, and a barplot of the coefficients.

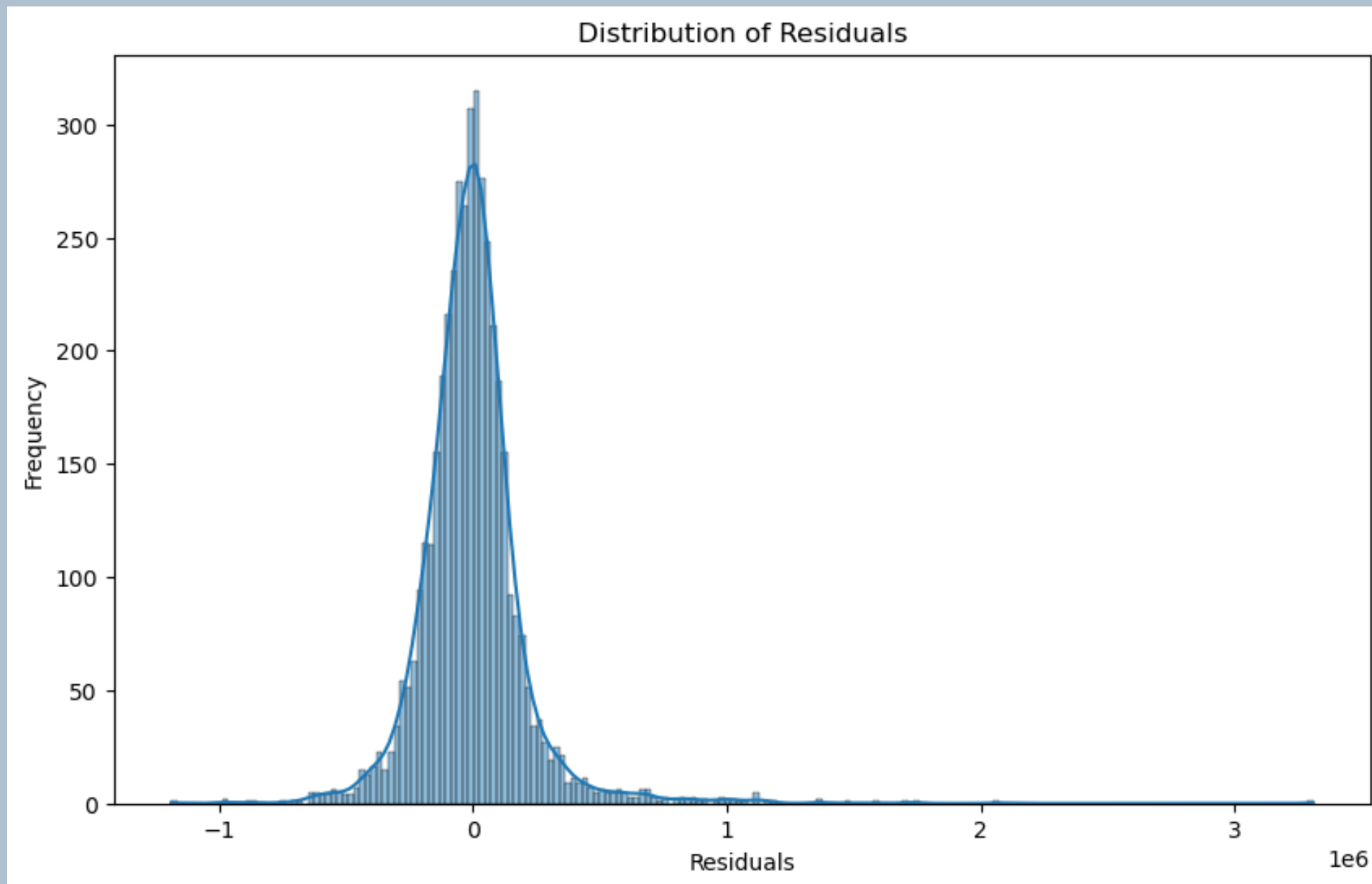
Some of them have been presented in the next slide.

ACTUAL VS PREDICTED PRICES



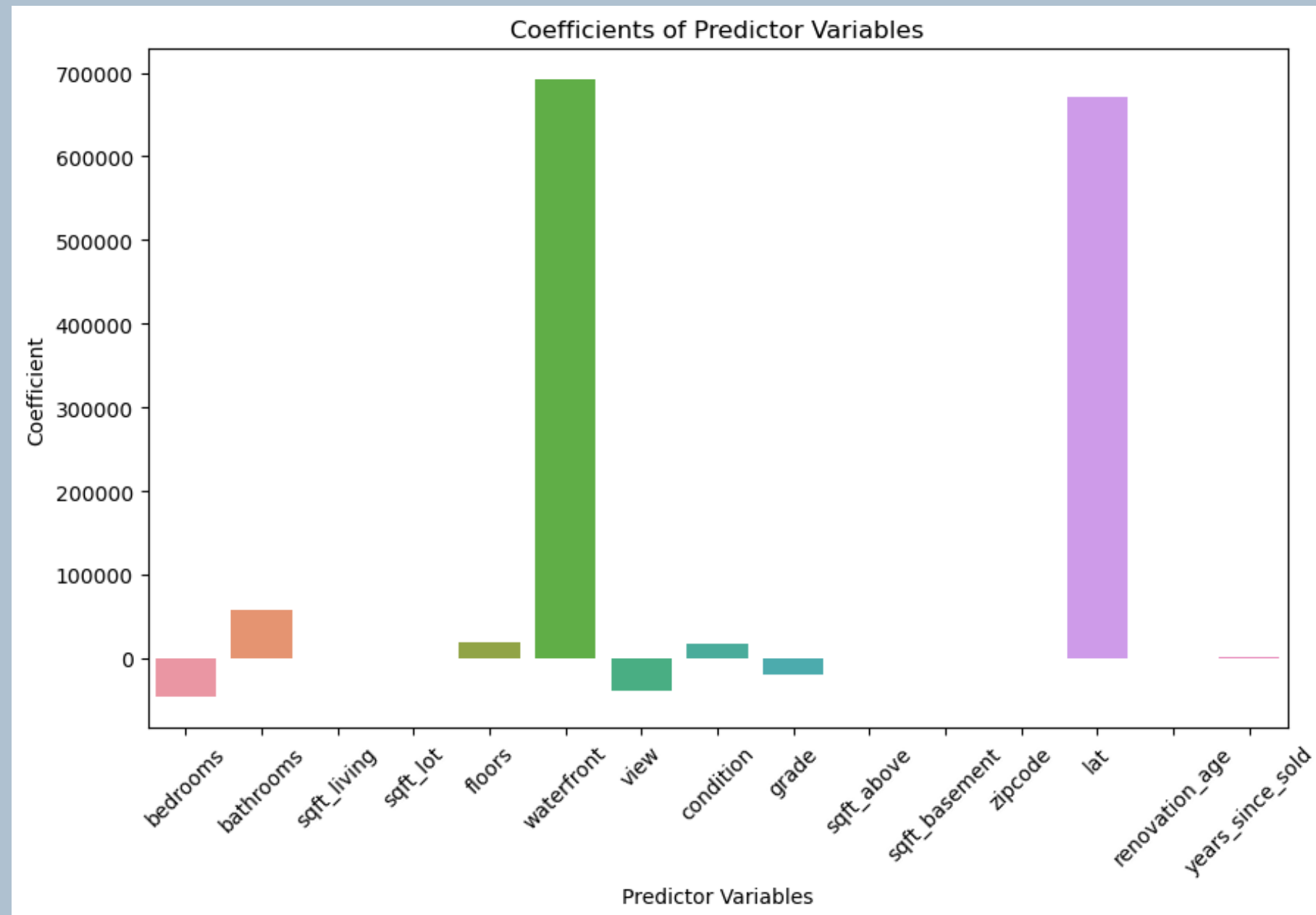
- The scatter plot of actual prices versus predicted prices shows how well the model's predictions align with the actual prices.
- Ideally, the data points should be close to the red dashed line, indicating a strong correlation between the predicted and actual prices.
- If the points deviate significantly from the line, it suggests that the model may have some limitations in accurately predicting prices

DISTRIBUTION OF RESIDUALS



- The histogram plot of residuals shows the distribution of the differences between the actual prices and the predicted prices.
 - If the residuals are normally distributed around zero, it indicates that the model's predictions are unbiased and have consistent errors.
 - Any deviations from a normal distribution or patterns in the residuals can suggest potential issues or biases in the model.

COEFFICIENTS OF VARIABLES



- The bar plot visualizes the coefficients of the predictor variables, representing the strength and direction of their relationship with the price.
- Positive coefficients indicate a positive relationship, meaning that an increase in the predictor variable leads to an increase in the price.
- Negative coefficients indicate a negative relationship, implying that an increase in the predictor variable leads to a decrease in the price.
- The height of each bar represents the magnitude of the coefficient, indicating the relative importance of the predictor variable in determining house prices.

CONCLUSIONS

1. The model has a statistically significant relationship with the target variable (price), as indicated by the overall model's F-statistic and individual predictor variable p-values.
2. The model explains approximately 66.5% of the variance in the target variable (R-squared value), suggesting that the selected predictor variables collectively have a moderate level of predictive power for determining house prices.
3. The coefficients of the predictor variables provide insights into their individual impacts on house prices. Some variables, such as the number of bedrooms, bathrooms, and square footage of living space, show expected relationships with the price, while others, such as the presence of a waterfront view or the property's condition and grade, have unique impacts on the price.

RECOMMENDATIONS

- Consider the number of bedrooms, bathrooms, and square footage of living space as key factors when determining house prices. Properties with more bedrooms and bathrooms, as well as larger living spaces, tend to command higher prices.
- Pay attention to the presence of a waterfront view, as it significantly affects house prices. Properties with waterfront views tend to have higher values, indicating their desirability and potential for premium pricing.
- Evaluate the condition and grade of the property. Both factors can influence the price, with better-conditioned properties and higher-grade ratings generally correlating with higher prices. Ensuring that properties are well-maintained and upgraded to higher-grade levels may lead to increased market value.