PHASE TWO

KING COUNTY HOUSE PRICING PROJECT

STUDENT NAMES:

JANE MWANGI

YVONNE KAMARI

WILFRED KIVINDA

ERIC NJERU

SEPT 2023

TM: MARYANN MWIKALI

Table of Contents

PROJECT OVERVIEW

This project uses data analysis to help King County real estate firms understand what drives house prices, enabling informed decisions and improved market strategies.

BUSINESS PROBLEM

- The King County housing market exhibits various patterns and factors influencing house prices. However, there is a need to understand these factors more comprehensively to aid real estate stakeholders in making informed decisions.

- The primary business problem is to develop a robust pricing model that accurately predicts house prices based on various features.
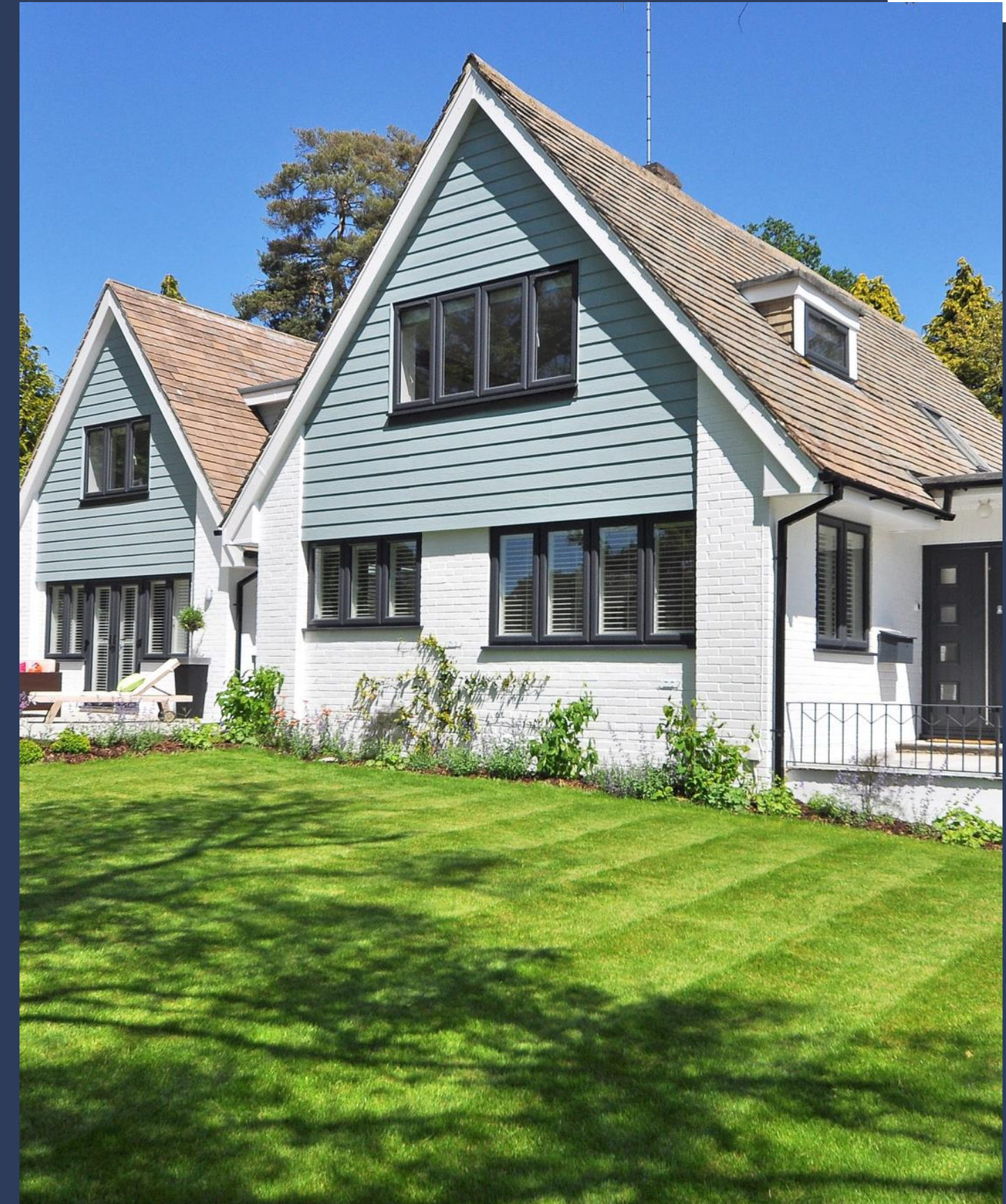
PROJECT OBJECTIVES

- The project aims to create a predictive model for King County, Washington, house prices using various data features such as bedrooms, square footage, condition, and location.

- It seeks to identify key price-influencing factors and enhance pricing accuracy in real estate transactions.

DATA UNDERSTANDING

- We utilized the King County House Sales dataset, which comprises various house features including bedrooms, bathrooms, square footage, views, and ZIP codes, as influential factors on house prices.

- We analyzed these features to predict house prices and provide valuable insights for sale price predictions. These insights can be beneficial for stakeholders such as real estate agents in advising their clients.

METHODOLOGY

01 Data Preprocessing
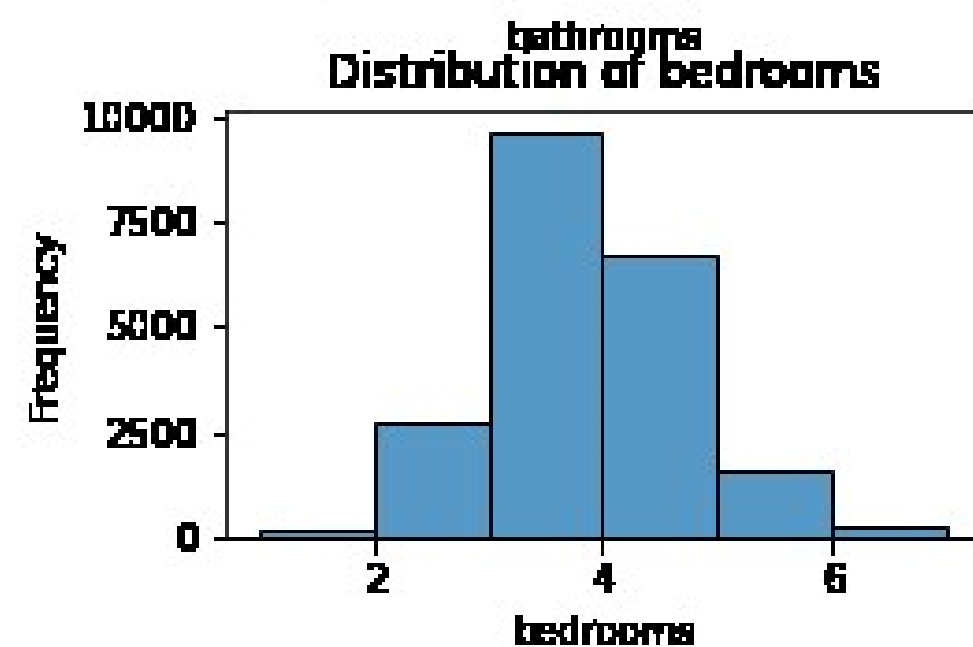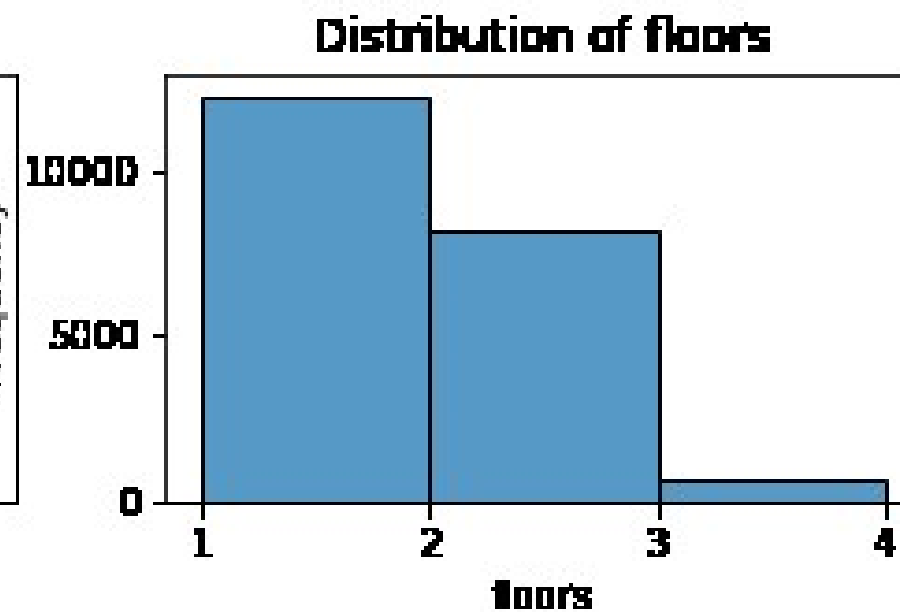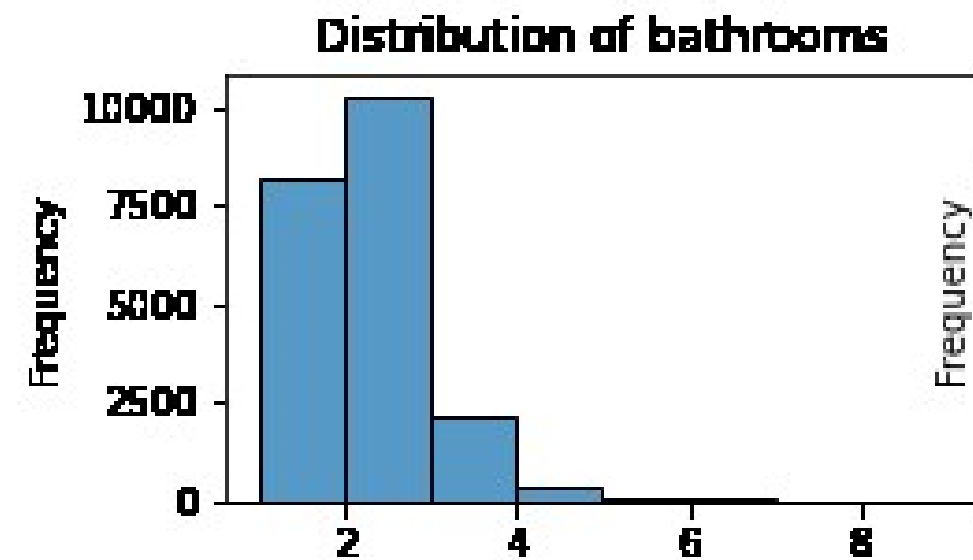
Exploratory Data Analysis **02**

03 Feature Selection

Model Building **04**

05 Model Evaluation

Model Optimization **06**

EDA RESULTS

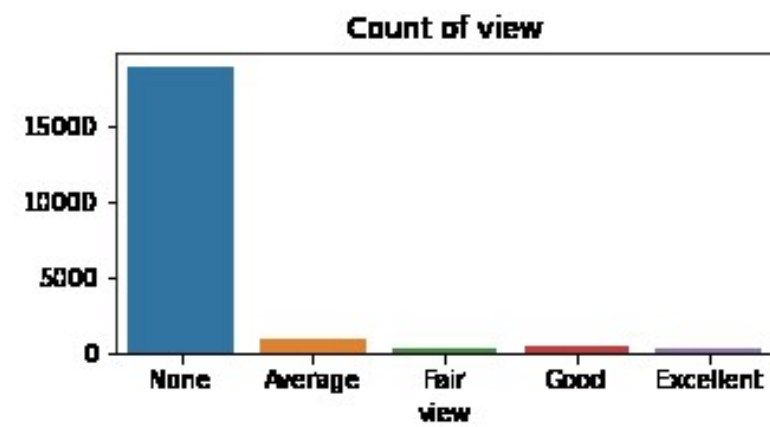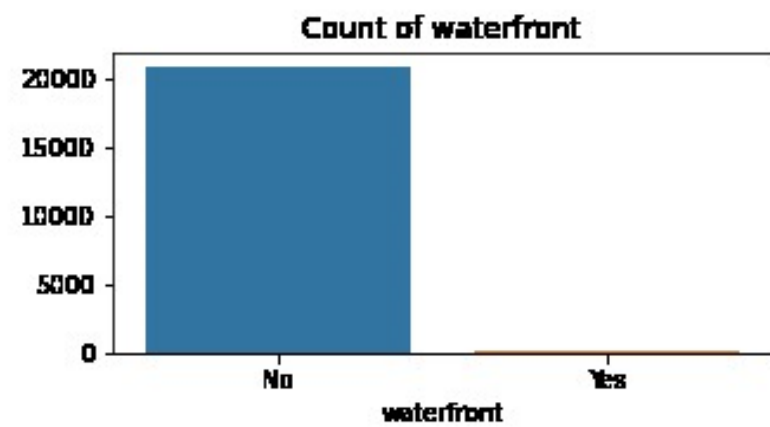The majority of homes in our dataset have between 2 and 5 bedrooms, 0.5 to 3 bathrooms and 1 and 2 floors.

EDA RESULTS

- The dataset is predominantly composed of houses that are not on a waterfront and lack scenic views.

- Most of these houses are in moderate condition, with grades typically ranging from average to good, and renovations are infrequently observed.
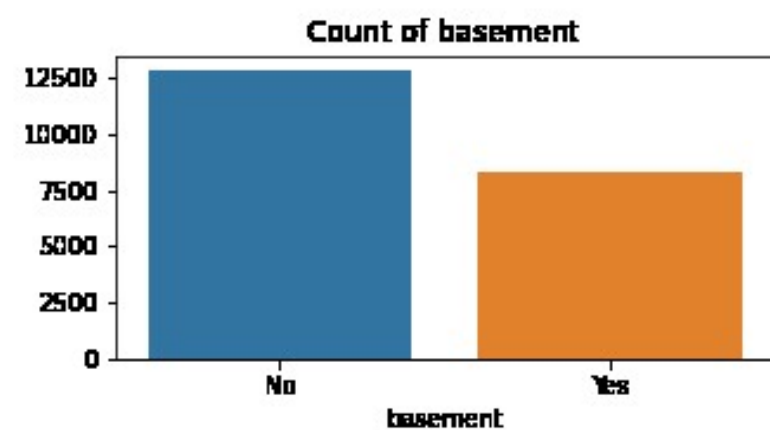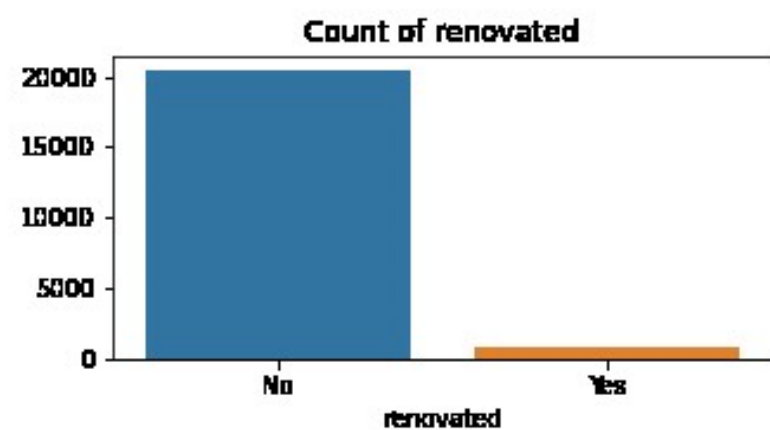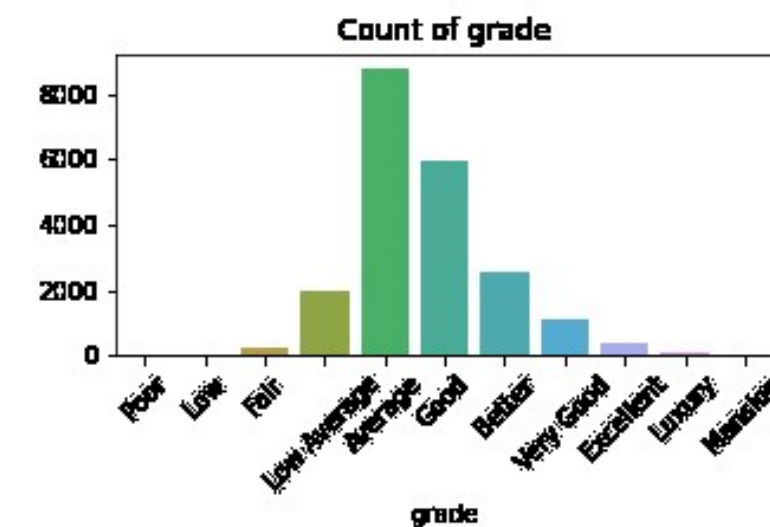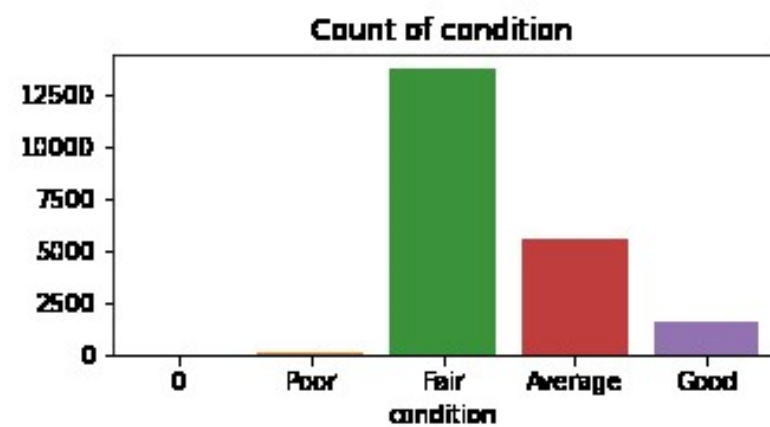
EDA RESULTS
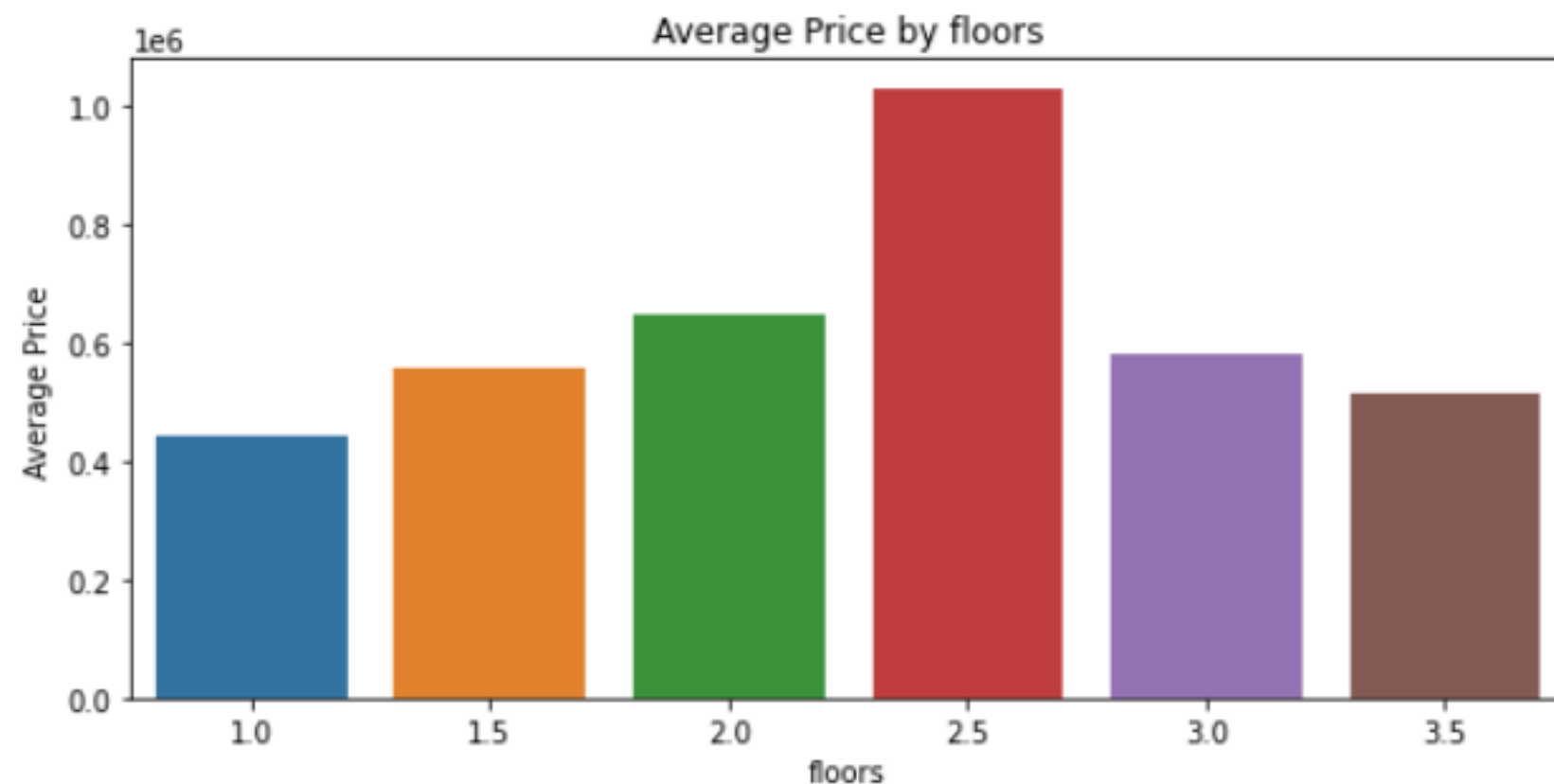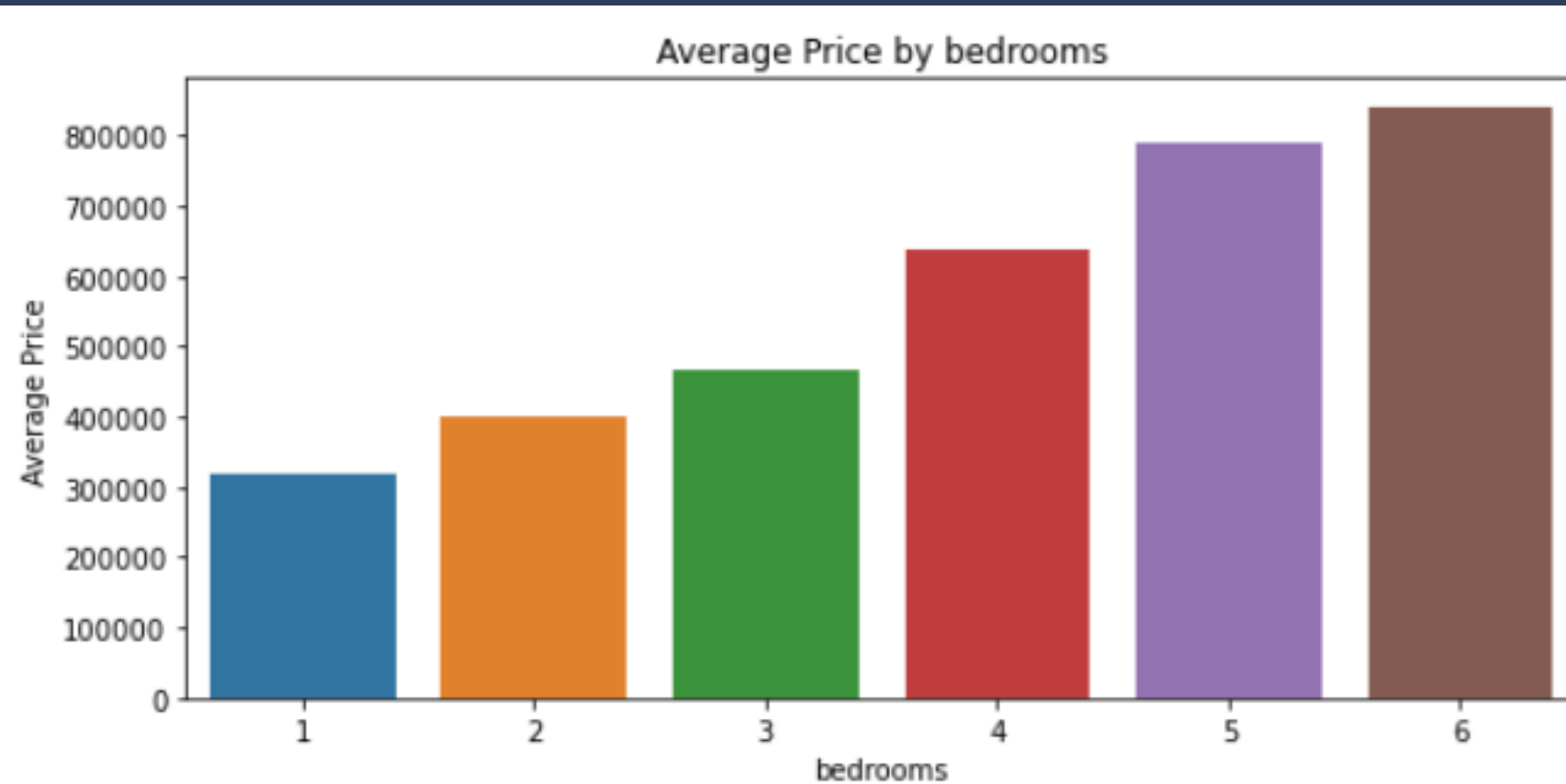
- The number of bedrooms and bathrooms directly affects house prices.

- Houses with 2.5 floors tend to be the most expensive on average.

Distribution of Houses Sold per Season
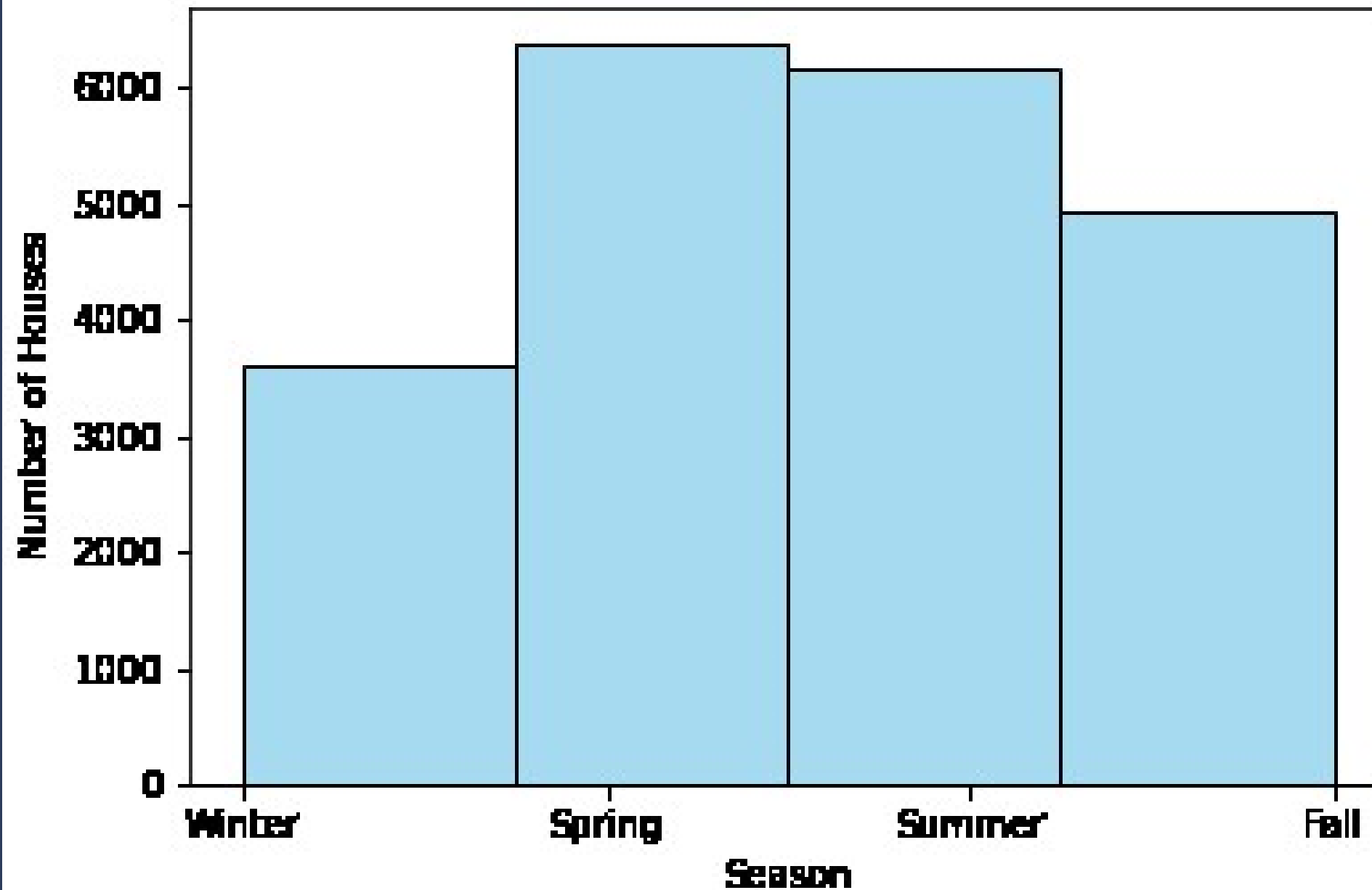
EDA RESULTS

While seasonality does not significantly impact prices, sellers may still benefit from listing their properties during the spring and summer seasons when demand tends to be higher.

EDA RESULTS

The analysis highlights the importance of considering location, as specific zipcodes exhibit variations in average property prices.

Average Price per Zipcode

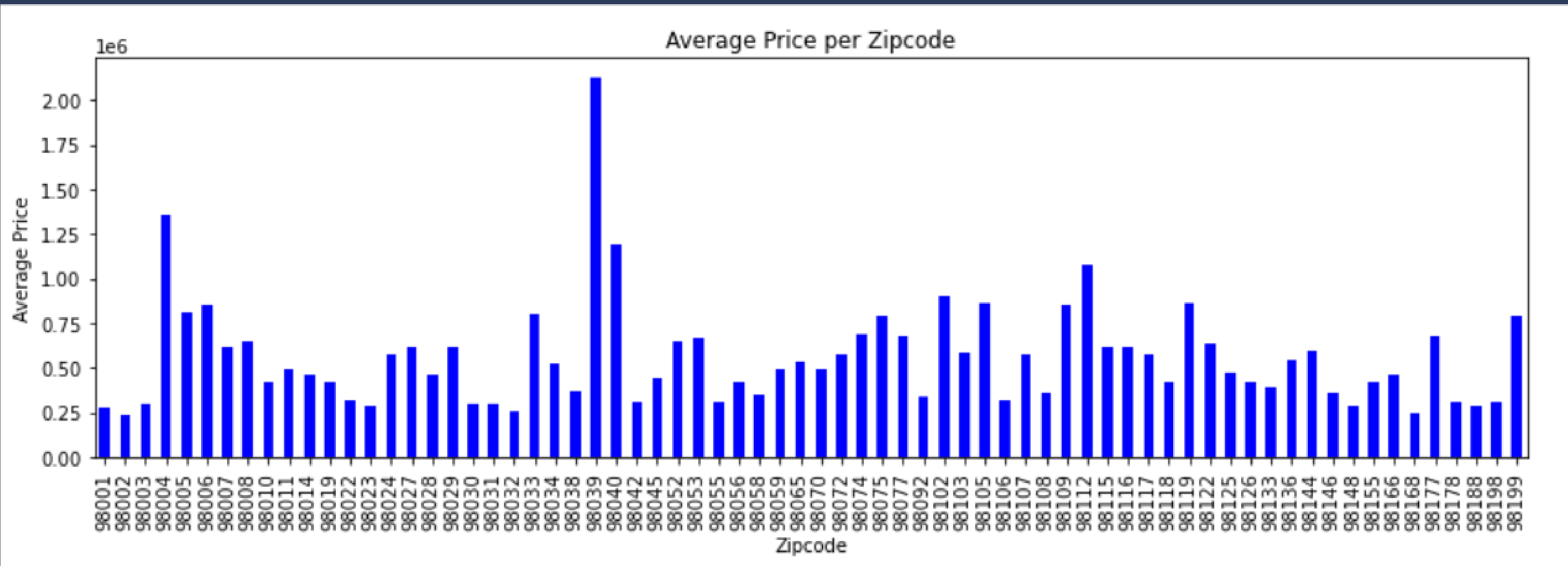# SIMPLE LINEAR REGRESSION RESULTS

Simple Linear Regression: Price vs. Sqft Living
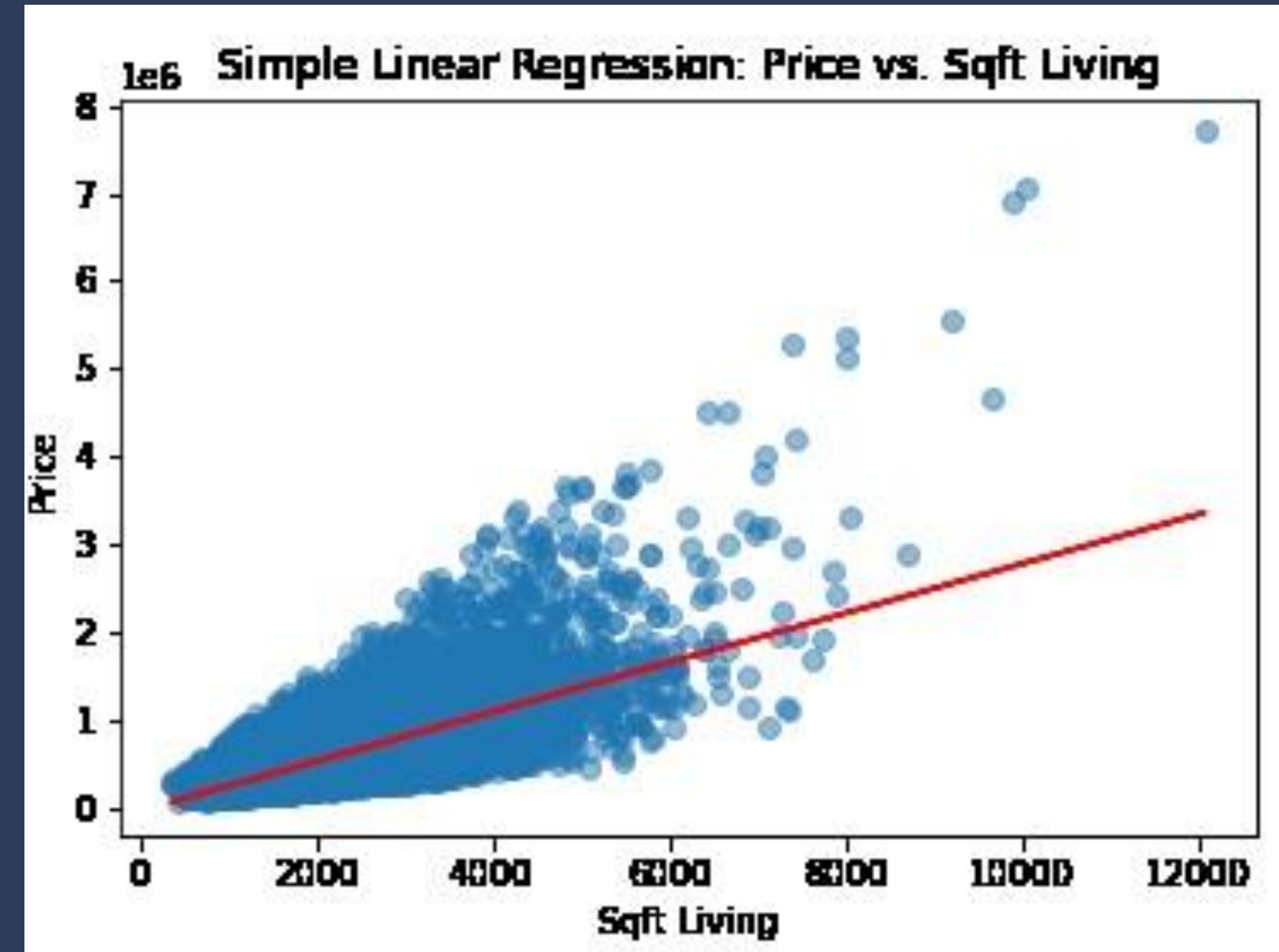
There is a strong positive correlation between price and sqft_living.

For each additional square foot of living space, the price increases by about $282.18.

MULTIPLE LINEAR REGRESSION RESULTS

Actual vs. Predicted Values
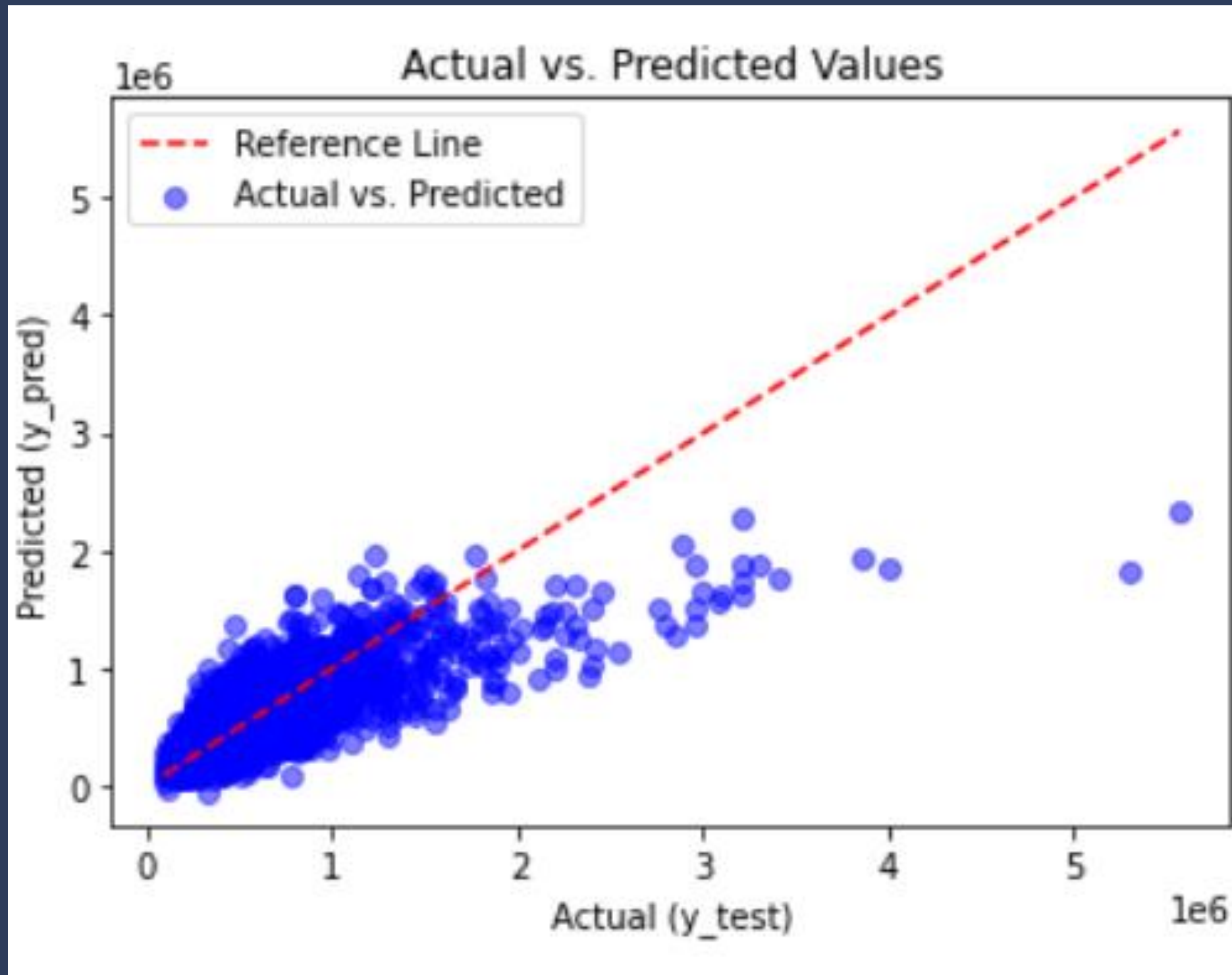
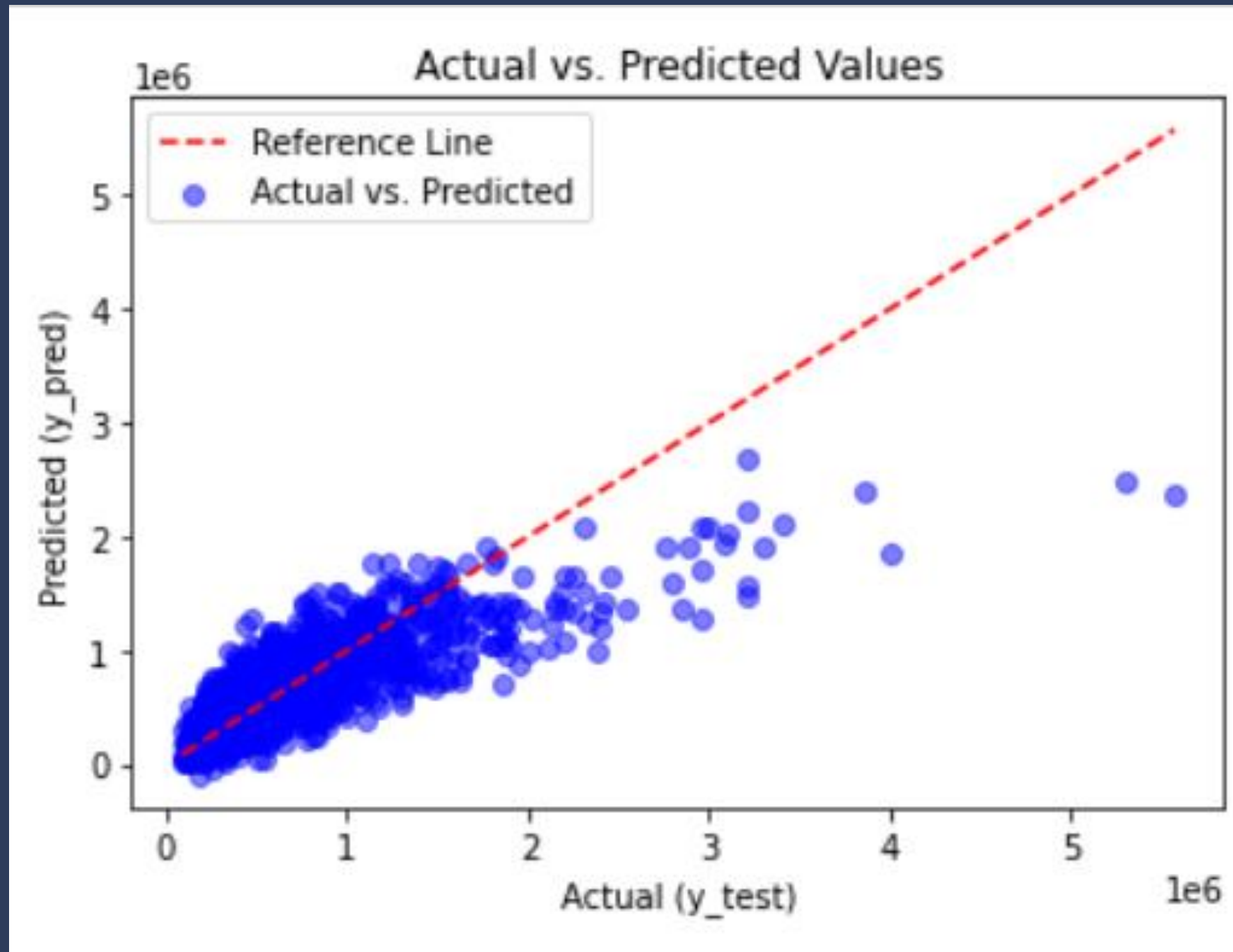- The first model focused on highly correlated variables with price; 'bathrooms', 'grade', 'sqft_living15', 'sqft_living', 'view', 'bedrooms'.

- The model explains 0.57 of the variance in price.

MULTIPLE LINEAR REGRESSION RESULTS

Actual vs. Predicted Values

- The final model utilized our previous findings and domain knowledge.

- The independent variables included; 'sqft_lot', 'floors', 'waterfront', 'view', 'condition', 'grade', 'yr_built', 'sqft_living15', 'sqft_lot15', 'renovated', sqft_living', 'basement', 'bedrooms'.

- The model explains 0.65 of the variance in price.

STUDY LIMITATIONS

- Our analysis recognizes the limitations stemming from the absence of several critical factors in our dataset, which can substantially influence house prices. Consequently, our model's predictive accuracy was constrained, achieving a rate of 65%.

- Notable missing data include Geospatial Data, Location, Market Trends, Amenities, Local Infrastructure, Property Attributes, Market Dynamics, Property History, Zoning Regulations among others.

CONCLUSIONS

- Spring and summer listing may benefit sellers due to increased demand.

- Zipcodes' findings underline the importance of location in pricing.

- Scenic settings, such as waterfront views, command a substantial premium.

- Take caution on the multicollinearity of features as it can affect pricing model accuracy.

- Possible predictive model limitations due to missing factors e.g., market trends, amenities, economic factors, and more, all influencing house prices.

RECOMMENDATIONS

- Develop a Data-Driven Pricing Strategy.

- Promote Unique Features, such as waterfront views, through targeted marketing to attract premium-paying buyers.

- Location-Based Marketing.

- Adapt sales and marketing strategies to align with seasonal trends.

- Data-Driven Decision Making.