NORTHWESTERN COUNTY REAL ESTATE PROJECT

OUTLINE

01 PROJECT OVERVIEW

02 BUSINESS UNDERSTANDING

03 DATA UNDERSTANDING

04 DATA MODELLING

05 RESULTS

06 RECOMMENDATION

07 NEXT STEPS

08 CONCLUSION

PROJECT OVERVIEW

In the dynamic real estate market of King County, Washington state, achieving success hinges on understanding key factors influencing property values. Our project, in collaboration with a local real estate agency, aims to uncover these factors using the extensive King County House Sales dataset.

Challenges such as economic downturns and data scarcity make precise forecasting difficult. To overcome these obstacles, we employ a blend of multiple linear regression models to analyze trends and provide actionable insights. Our ultimate goal is to develop a comprehensive advice system that empowers homeowners to make informed decisions about property renovations and understand their impact on property worth.

BUSINESS UNDERSTANDING

1

Business Problem

A King County real estate agency lacks a reliable system for data-driven insights on home prices, hindering stakeholders' ability to make informed decisions about property values.[2]

2

Project Solution

We aim to empower homeowners, investors, and real estate agents with valuable insights. Homeowners can accurately assess their property values, investors can spot discounted properties, and agents can advise clients on pricing strategies by predicting home prices effortlessly.

OBJECTIVES

Predicting Home Prices

Developing a model to estimate home value increases based on renovation factors

Identifying Important Features

Examining renovation variables to determine which ones have the greatest impact on increasing a home's estimated value.

Monitoring Market Trends

Analyzing regions with highest and lowest average sale prices and identifying most in-demand property types for market insights.

17704
17704

DATA UNDERSTANDING

This project is based on the dataset of a northwestern county. The dataset encompasses various features, including but not limited to:

- price
- bedrooms
- bathrooms
- sqft_living
- zipcode
- yr_built

Data time frame: 1900 - 2015

21,597 records

15 Numerical Columns

3 data types

6 categorical features

The methods used in handling the data set given include:

Data Preparation

This process entails cleaning, transforming, and organizing raw data to make it suitable for analysis and modeling. Through it, we uncovered insights such as:

- Characteristics of dataset columns.
- Types of data present.
- Shape of the dataset itself

Data Cleaning

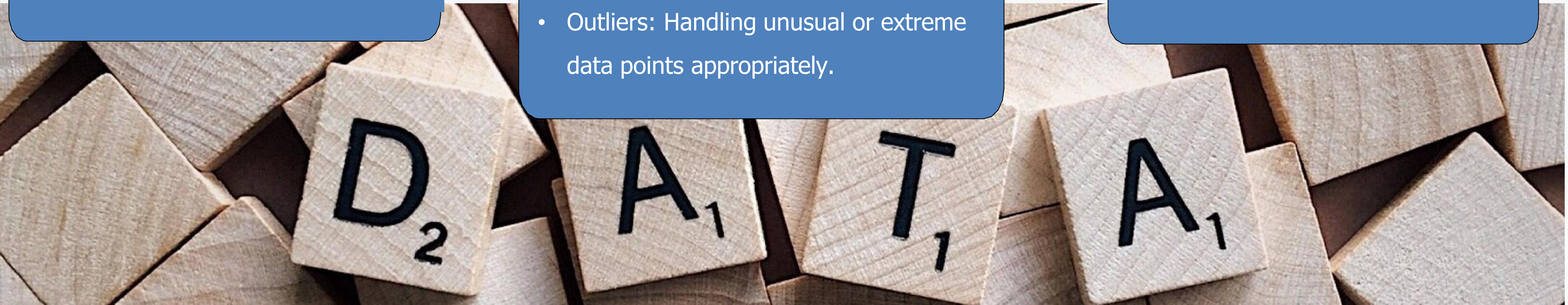
In this process, we decide how to address:

Missing values: Devising strategies to manage and fill in missing data.

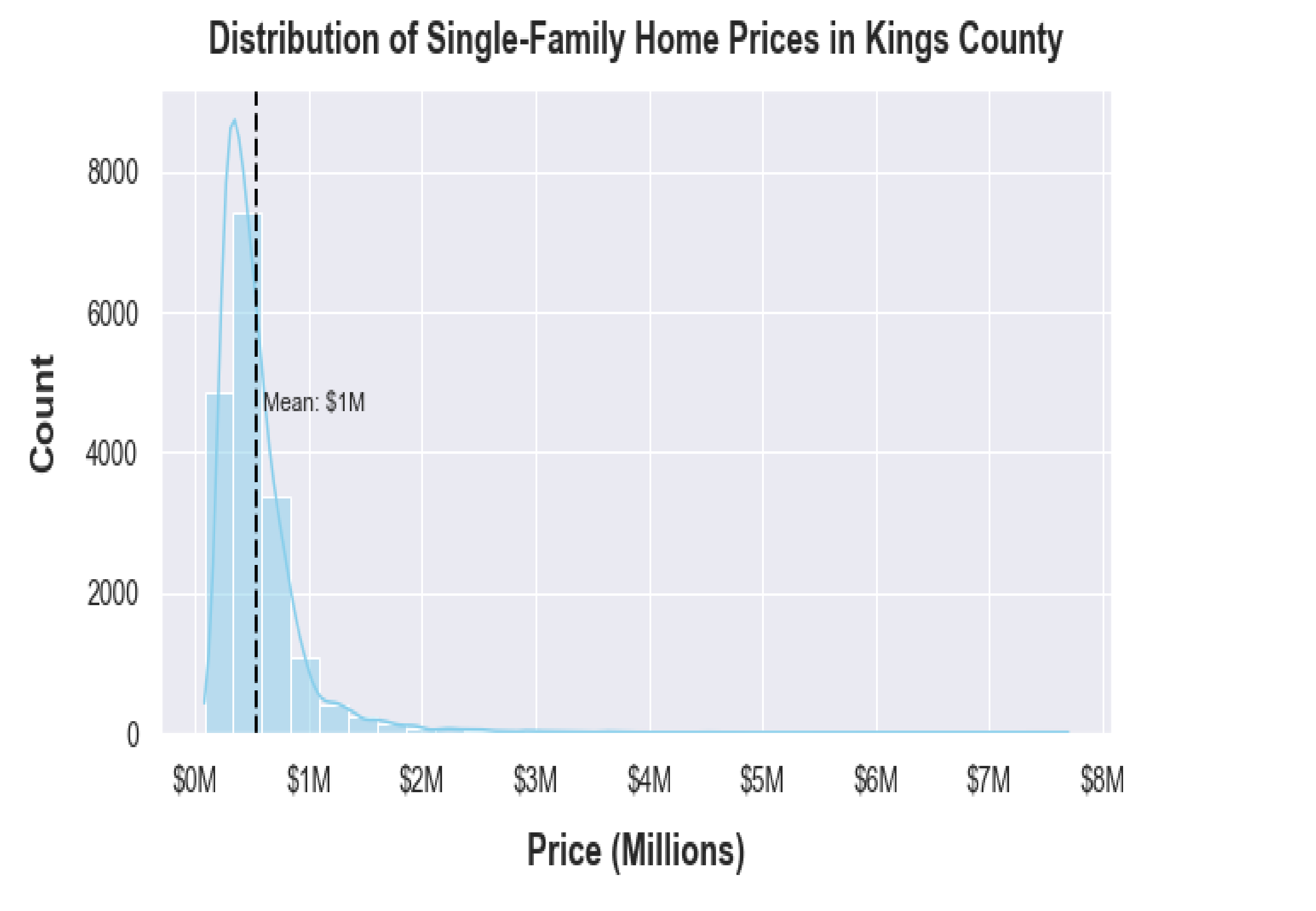
- Duplicated values: Identifying and resolving any repeated entries in the dataset.
- Wrong data types: Correcting any inconsistencies in the data formats.
- Outliers: Handling unusual or extreme data points appropriately.

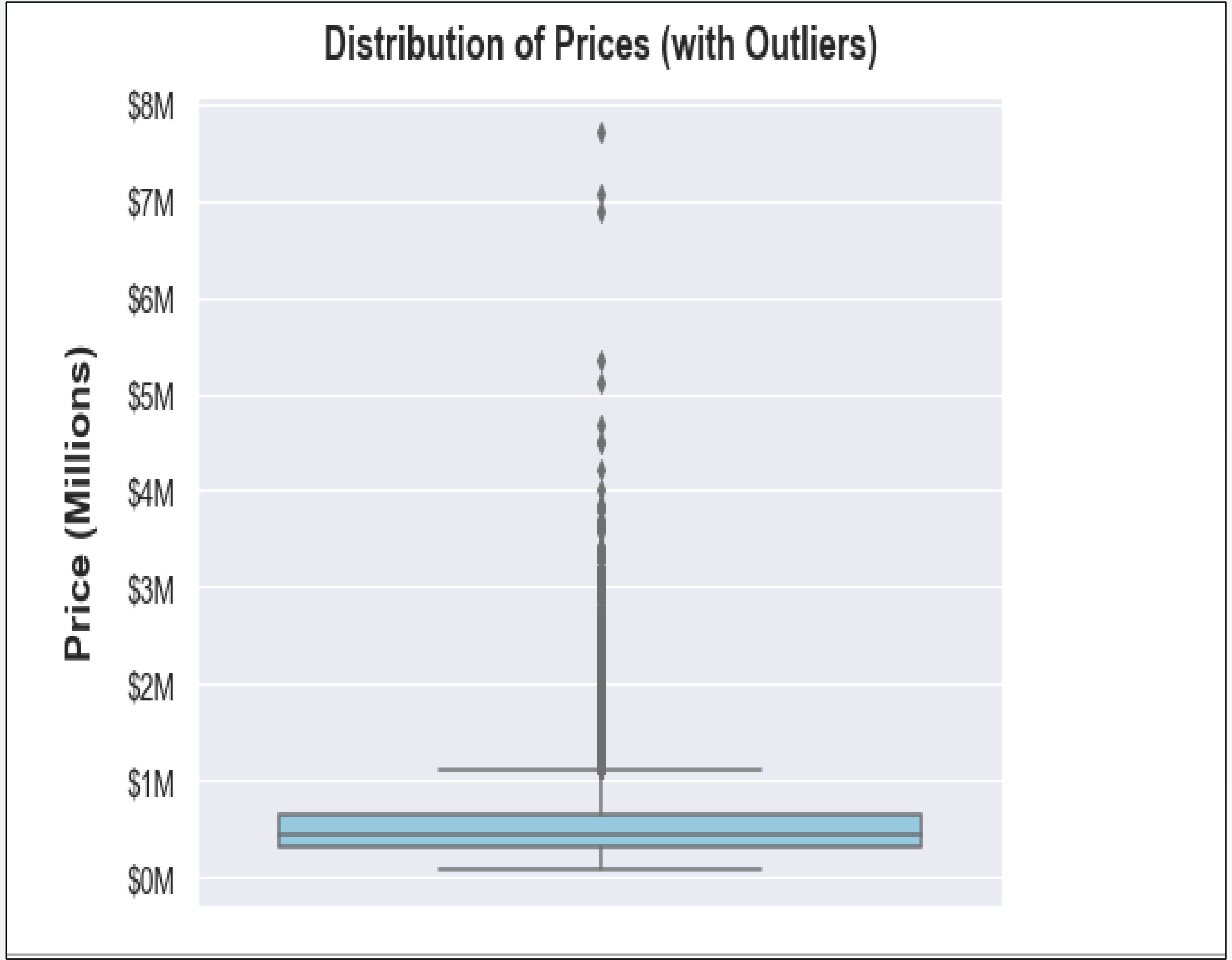
Data Visualization

This process allows us to visually represent complex data and trends in easy-to-understand charts and graphs, helping everyone grasp important information and make informed decisions



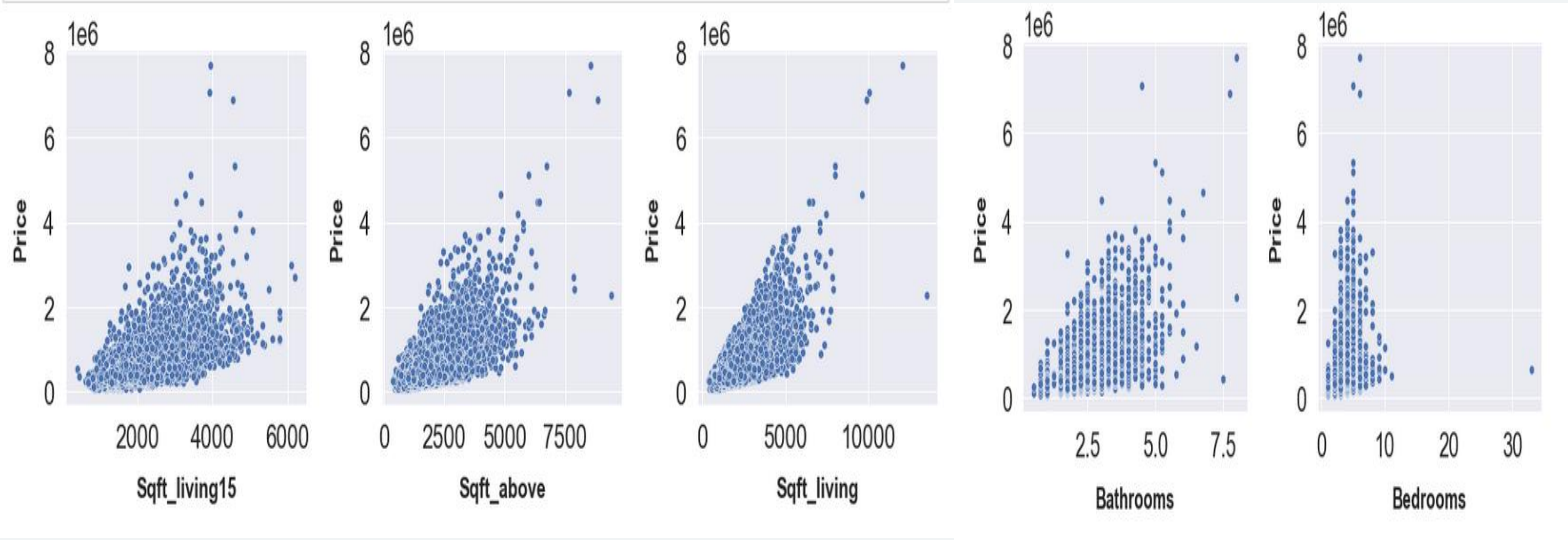
UNIVARIATE ANALYSIS



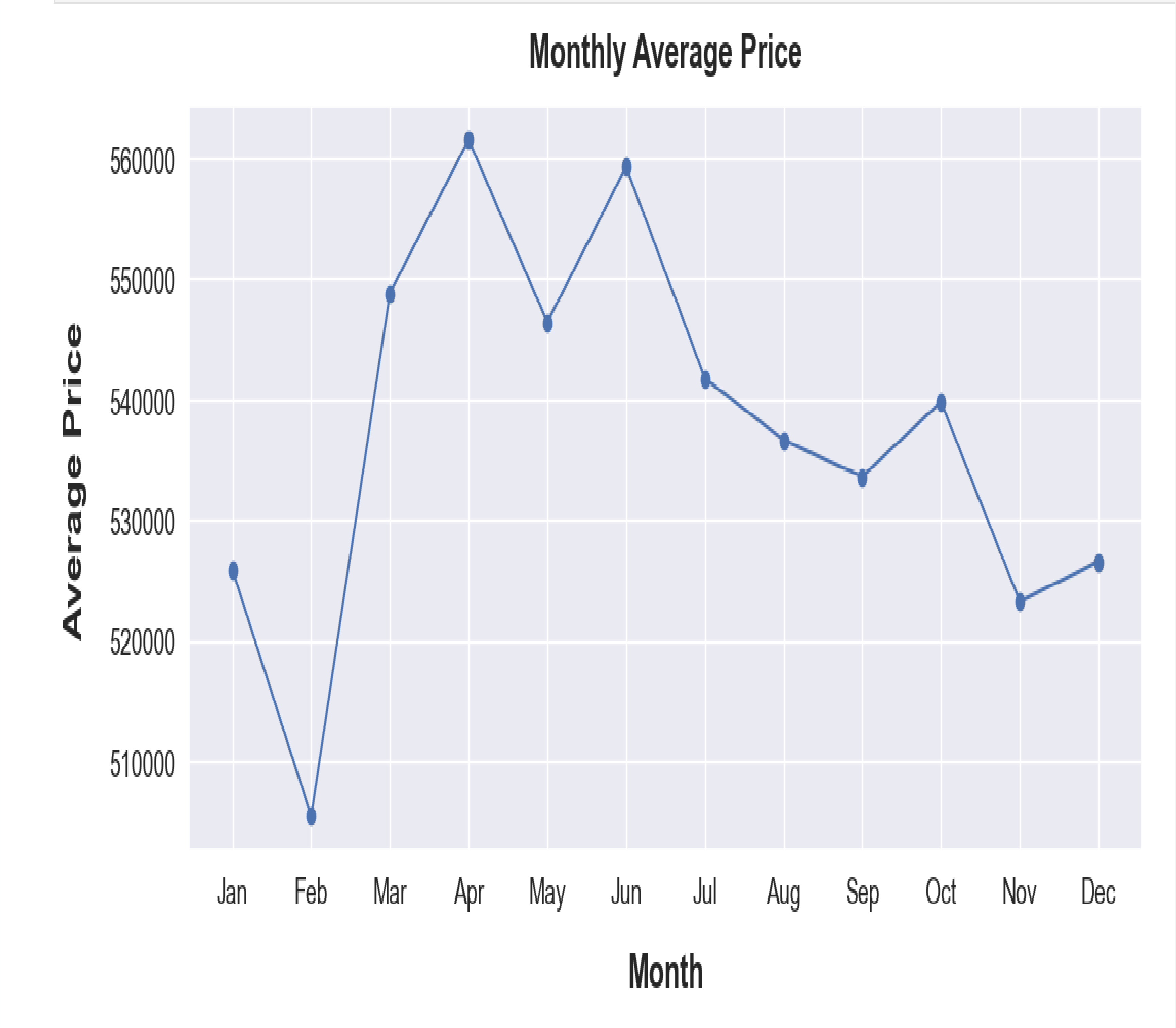


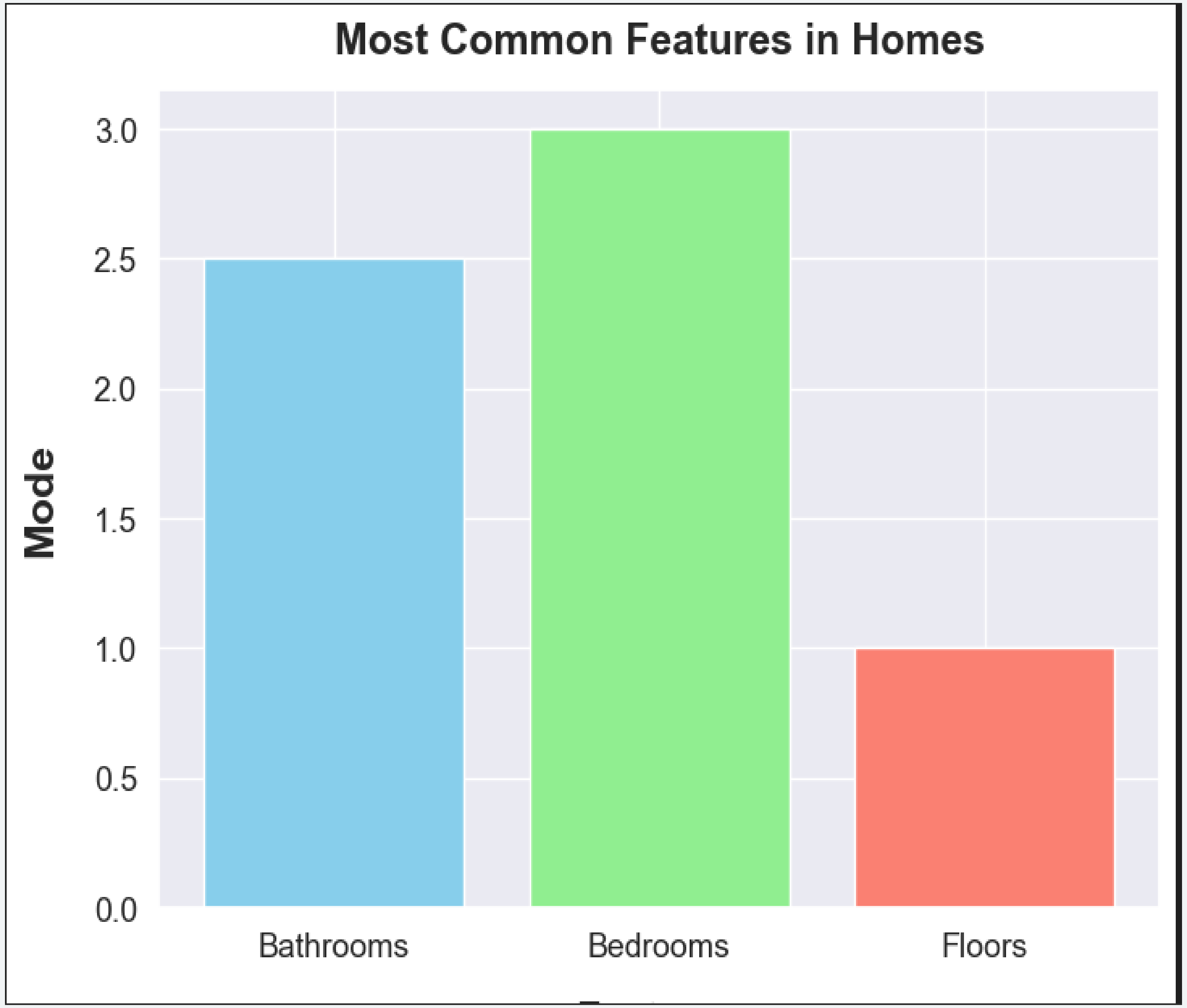
The histogram and box plot reveal skewed prices, predominantly below average, with a surplus of outliers contributing to peakedness.

BIVARIATE ANALYSIS



The plots show strong positive links between house prices and features like living space and bathrooms, with bedrooms having a moderate impact.





Monthly average house prices display seasonal fluctuations, with a notable decrease from January to February followed by a consistent increase until April, highlighting opportunities for strategic decision-making in the data science industry.

The analysis highlights that the majority of homes in the dataset typically feature at least three bedrooms, two bathrooms, and one floor, reflecting common buyer preferences in property attributes.

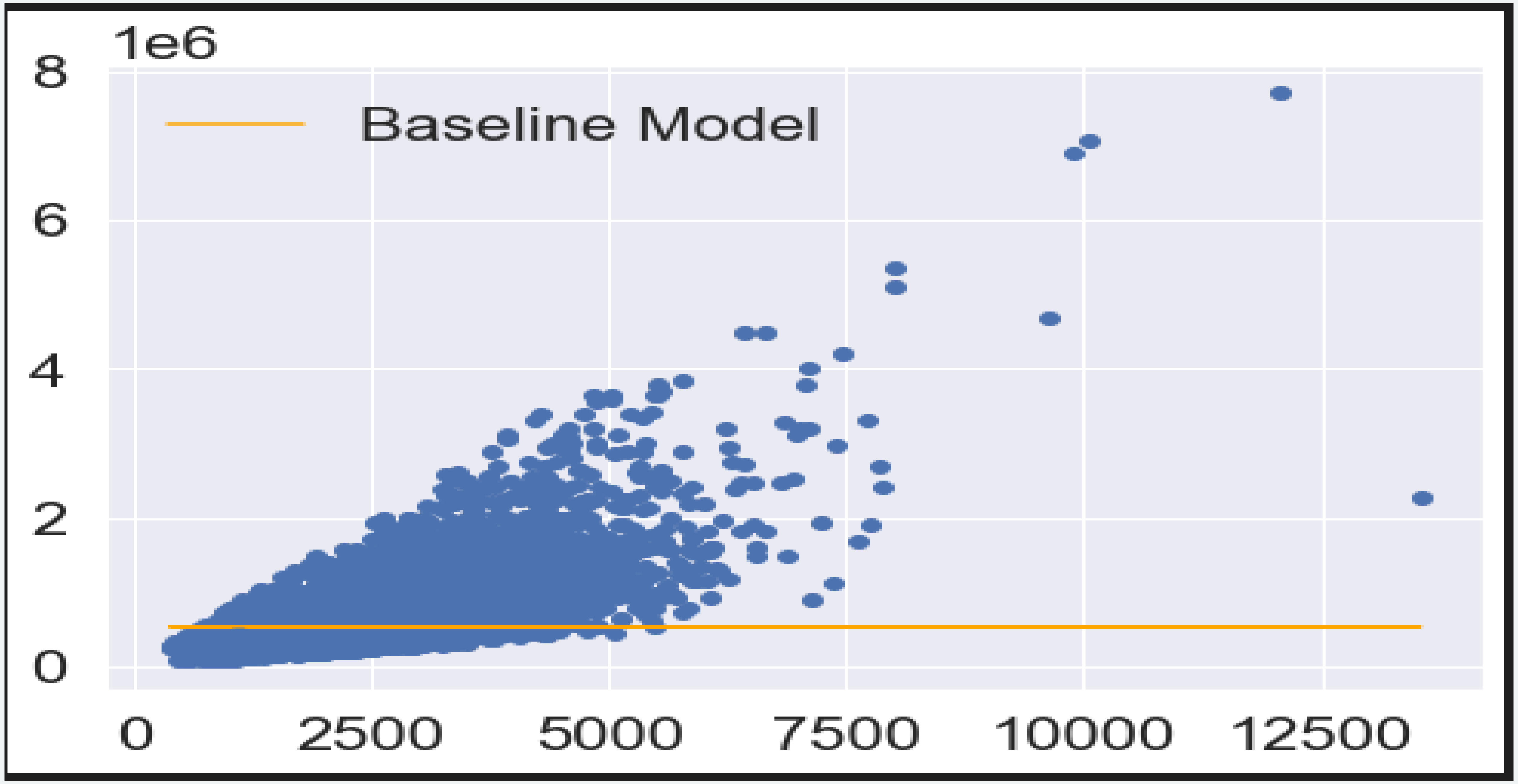
MODELLING

For the dataset provided, we are able to conduct regression modelling in order to draw conclusions and predictions from the data.

80%

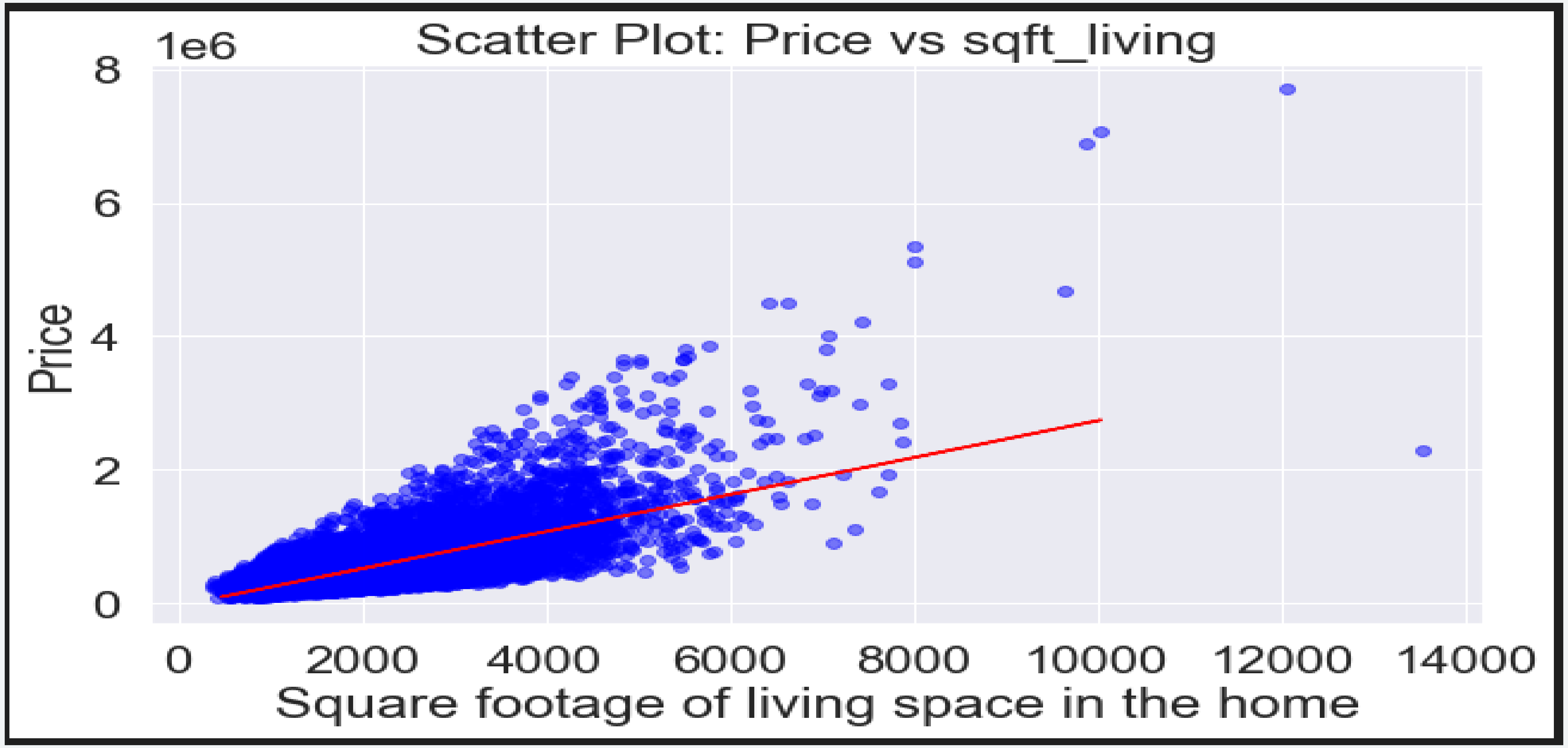
20%

RESULTS



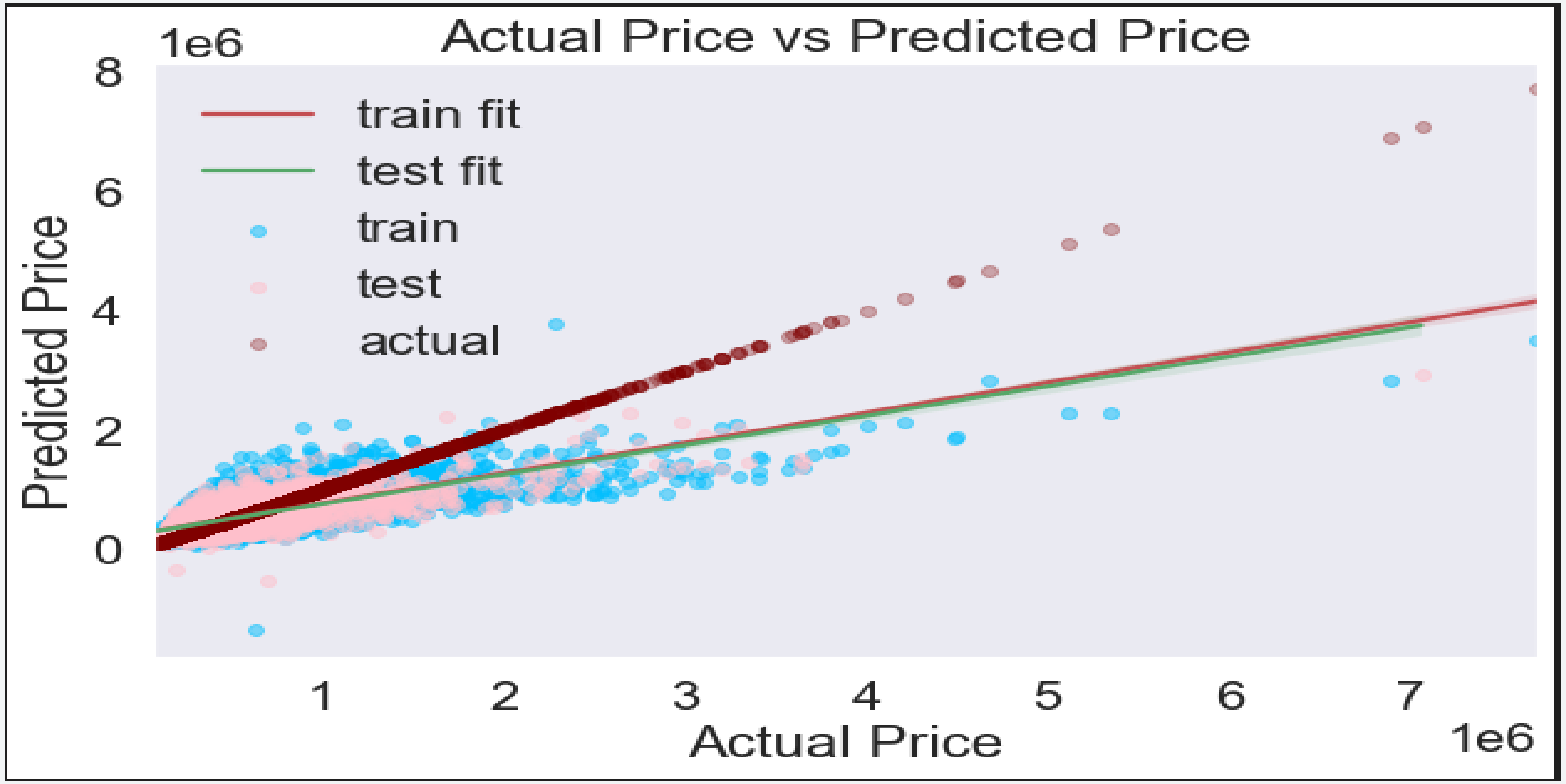
With a baseline Mean Absolute Error (MAE) of $233,684.32, the model predicts house prices with an average deviation from actual prices of less than this amount.

SIMPLE LINEAR MODEL



This Ordinary Least Squares (OLS) regression model explains 48.6% of the variance in house prices (R-squared = 0.486). A significant positive coefficient of 277.2332 suggests that for each square foot increase in living area, the price increases by $277.23.

MULTIPLE LINEAR MODEL



The model trained on the testing data explains 52.5% of the variance in house prices (R-squared = 0.525) with a Root Mean Square Error (RMSE) of approximately $256,868.

RECOMMENDATIONS

- **Bathrooms**: More bathrooms typically correlate with higher house prices, indicating greater appeal to buyers and suggesting the need for strategic pricing and marketing.

- **Living Area and Lot Size**: Larger living areas positively impact house prices, while larger lot sizes have a negative effect, necessitating emphasis on living space in marketing efforts.

- **Floors**: Homes with multiple floors tend to command higher prices, prompting consideration of this factor in pricing and marketing strategies for multi-story properties

NEXT STEPS

01

02

03

OUR TEAM

1. ABIGAEL NYABAGA
2. BEDAN CHEGE
3. BERYL AGAI
4. BRIAN MUTHAMA
5. FARHIYA JARSO
6. IAN KIPTOO
7. JACKSON MUNENE
8. KELSEY MAINA