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

Our collaboration with a local real estate agency in King County, WA, aims to uncover key factors influencing property values using extensive house sales data. Using multiple linear regression models, we overcome challenges like economic downturns and data scarcity to provide actionable insights for informed homeowner decisions

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 **ProjectSolution**

We aim to empower homeowners, investors, and agents with insights to assess property values accurately, identify investment opportunities, and advise on pricing strategies 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

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

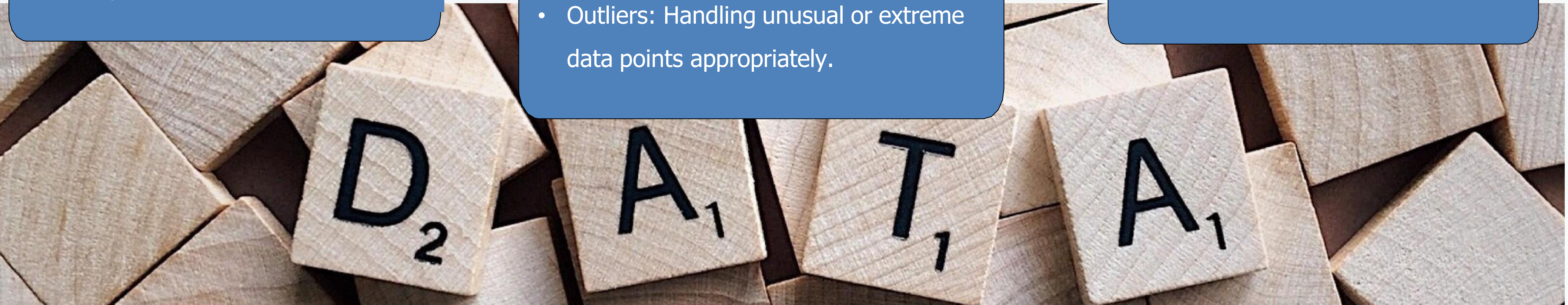
In this process, we decide how to address:

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

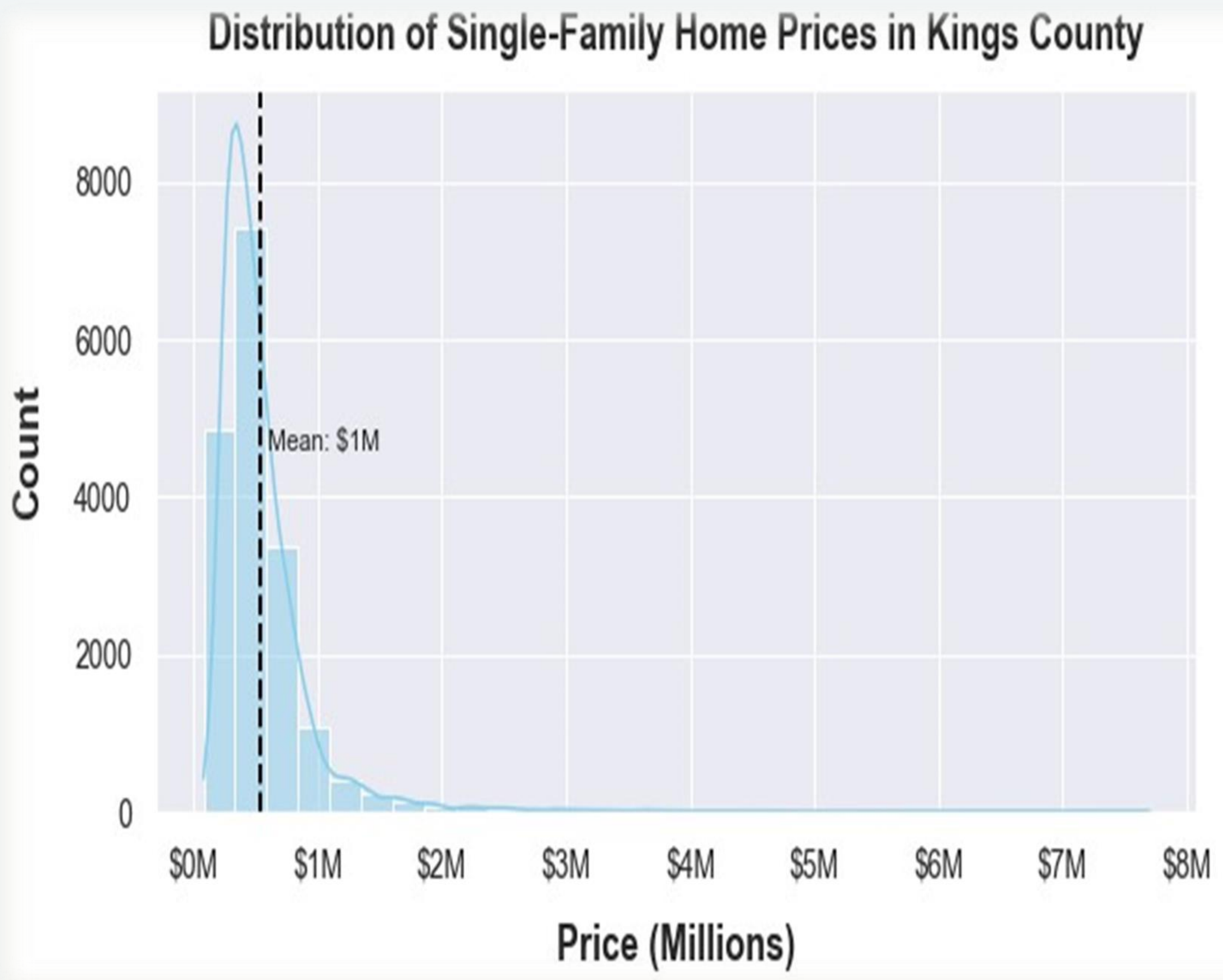
- Duplicated values: Identifying and resolving any repeated entries in the dataset.
- Wrong data types: Correcting any inconsistencies in the data formats.
- Outliers: Handling unusual or extreme data points appropriately.

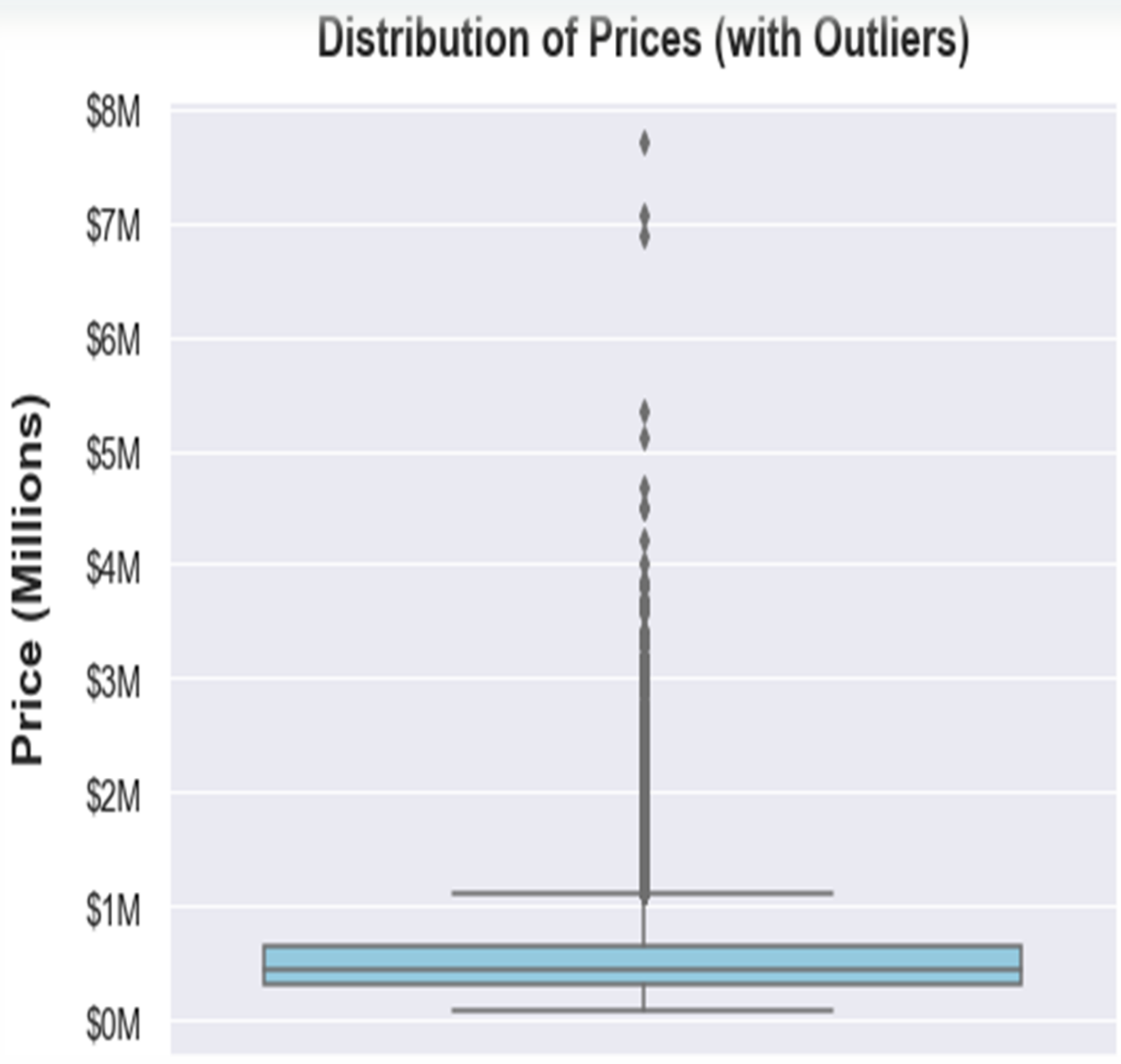
Data Visualization

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



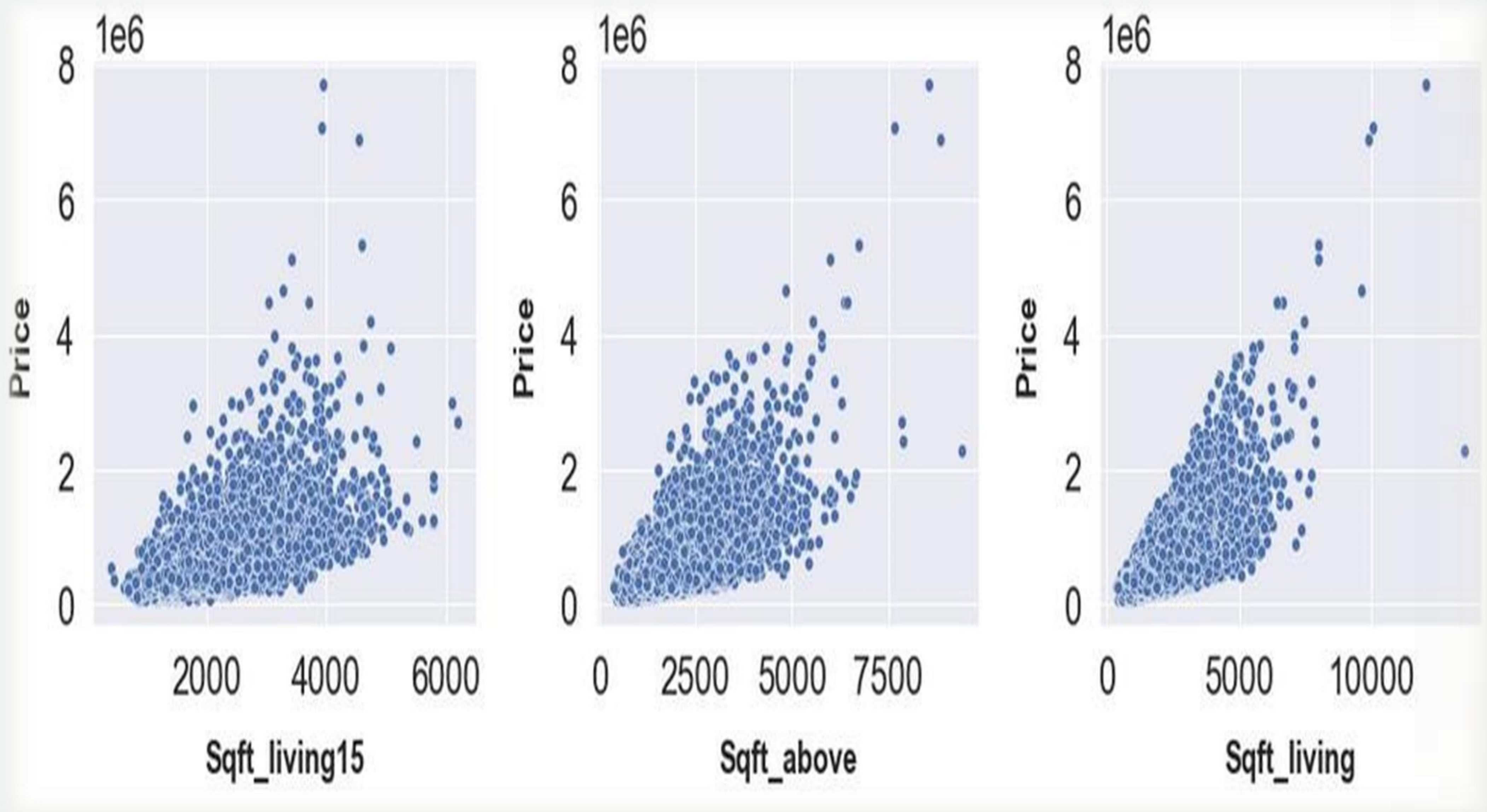
UNIVARIATE ANALYSIS

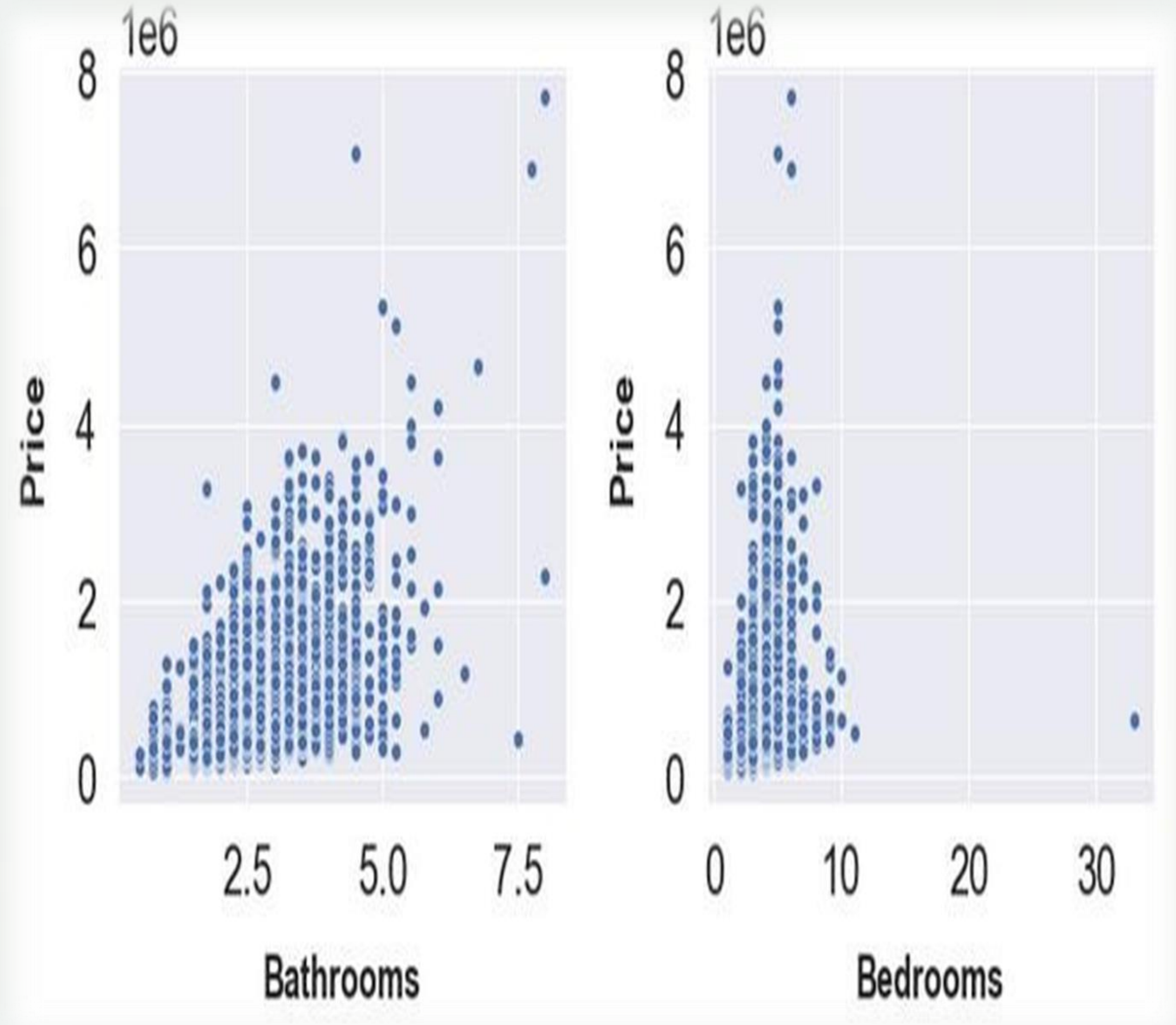




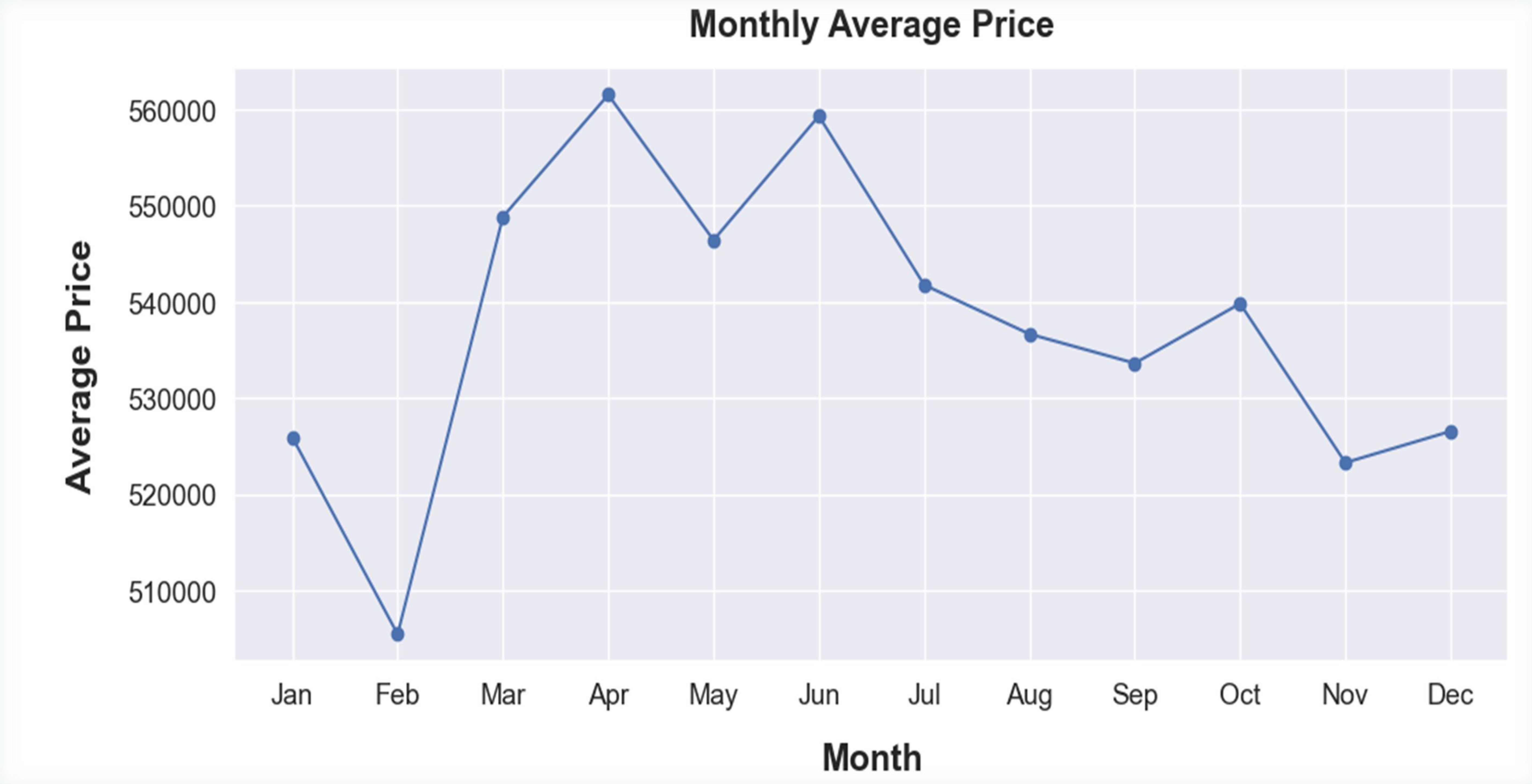
The histogram and box plot reveal skewed prices, predominantly below average, with a surplus of outliers contributing to peakedness.

BIVARIATE ANALYSIS

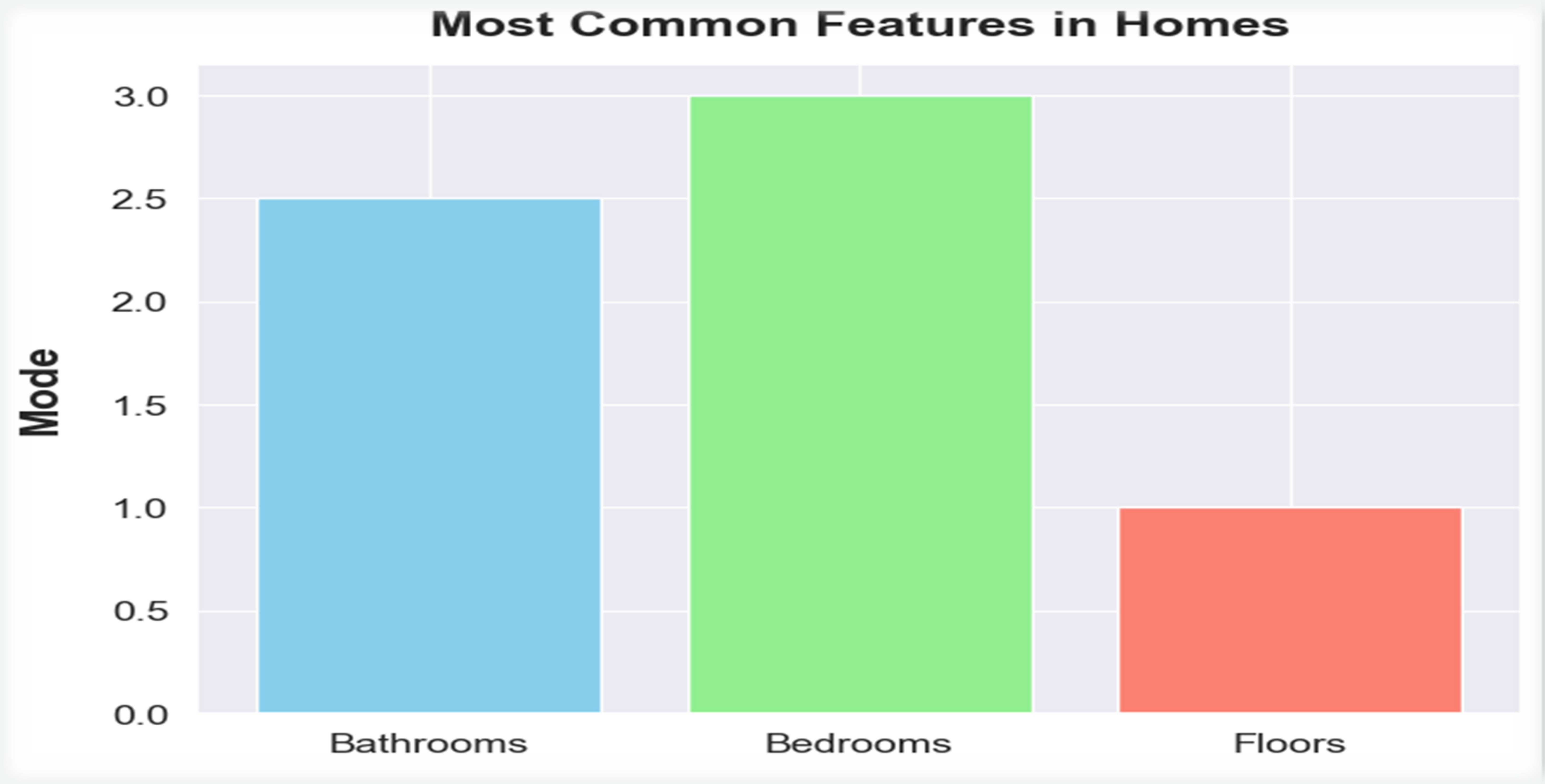




The plots show strong positive links between house prices and features like living space and bathrooms, with bedrooms having a moderate impact.



House prices fluctuate seasonally, offering strategic insights.



Analysis shows most homes have 3 bedrooms, 2.5 bathrooms, and 1 floor, aligning with buyer preferences.

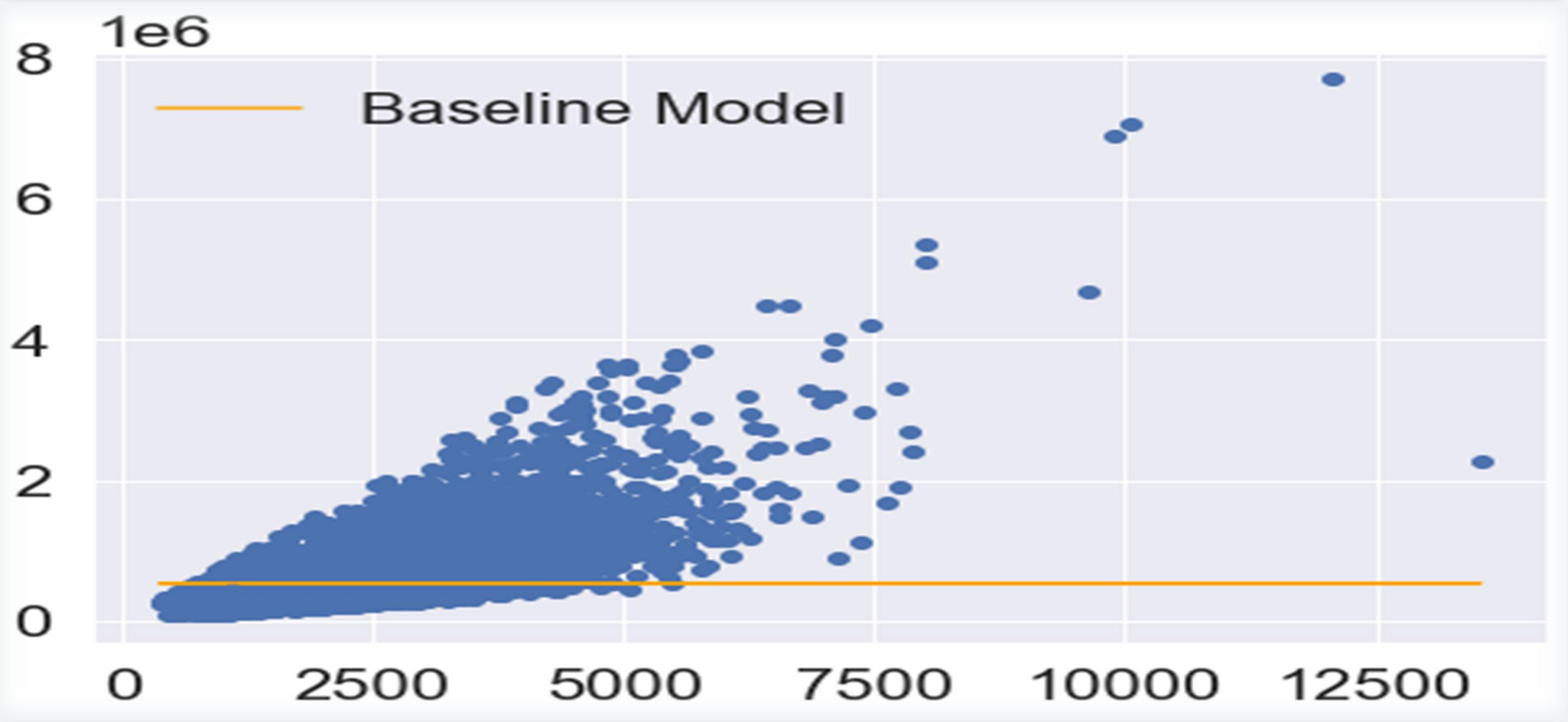
MODELLING

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

80%

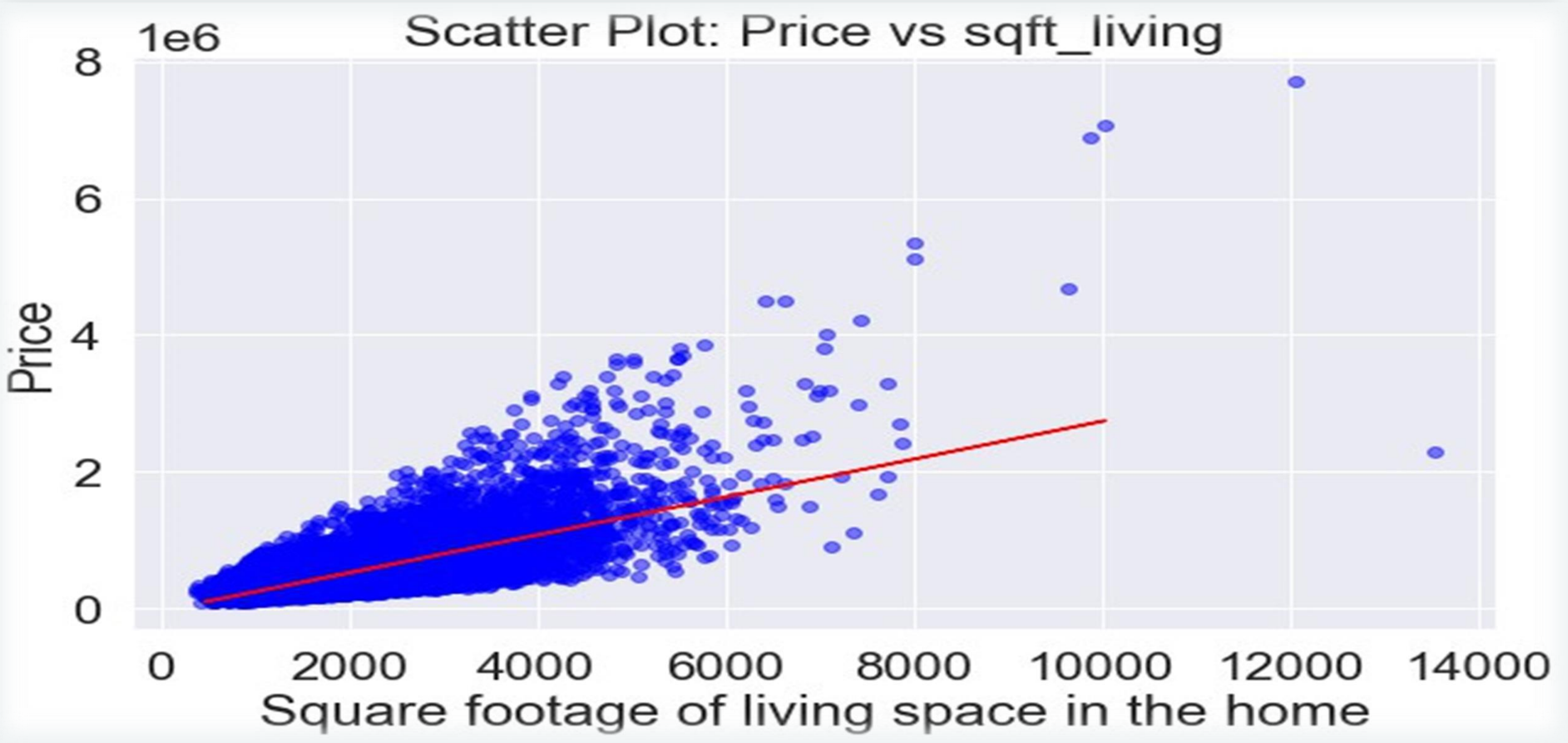
20%

RESULTS



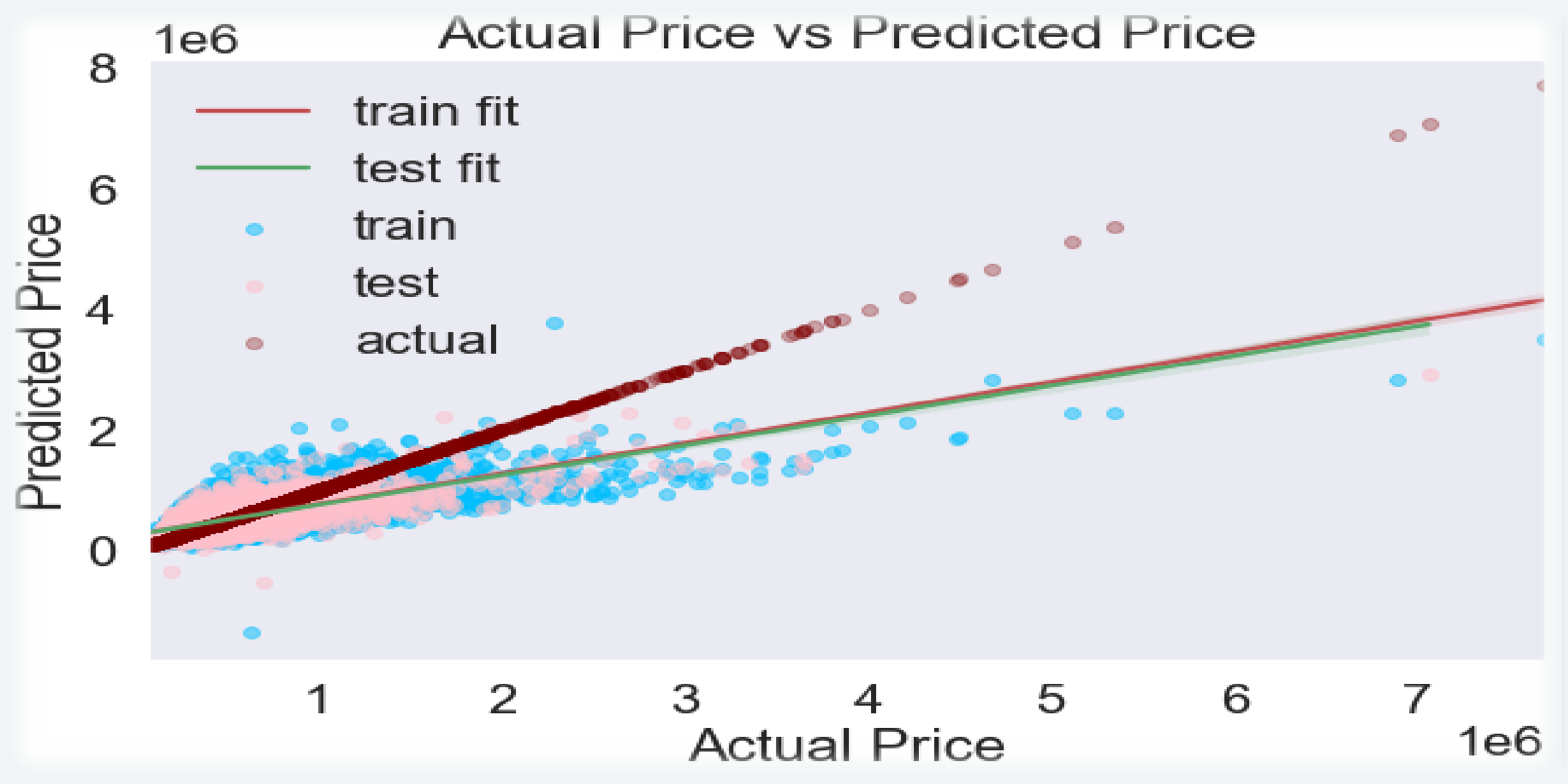
With a baseline Mean Absolute Error (MAE) of $233,684.32, the model predicts house prices with an average deviation from actual prices of less than this amount.

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 correlate with higher prices.

Living Area and Lot Size: Larger living areas increase price, while larger lots decrease it.

Floors: More floors lead to higher prices.

Condition and Grade: Better condition and grade mean higher prices.

RECOMMENDATIONS

Age and Renovation: Newer and renovated homes are priced higher.

Waterfront View: Properties with waterfront views command higher prices.

Season: Spring sales typically yield higher prices than fall sales.

NEXT STEPS

Collect More Variables: Incorporate neighborhood amenities, economic indicators, and housing market trends.

Improve Data Coverage: Obtain recent data for validation; expand geographic and housing type coverage.

Explore Advanced Modeling Techniques: Utilize advanced algorithms; rigorously evaluate and validate the model.

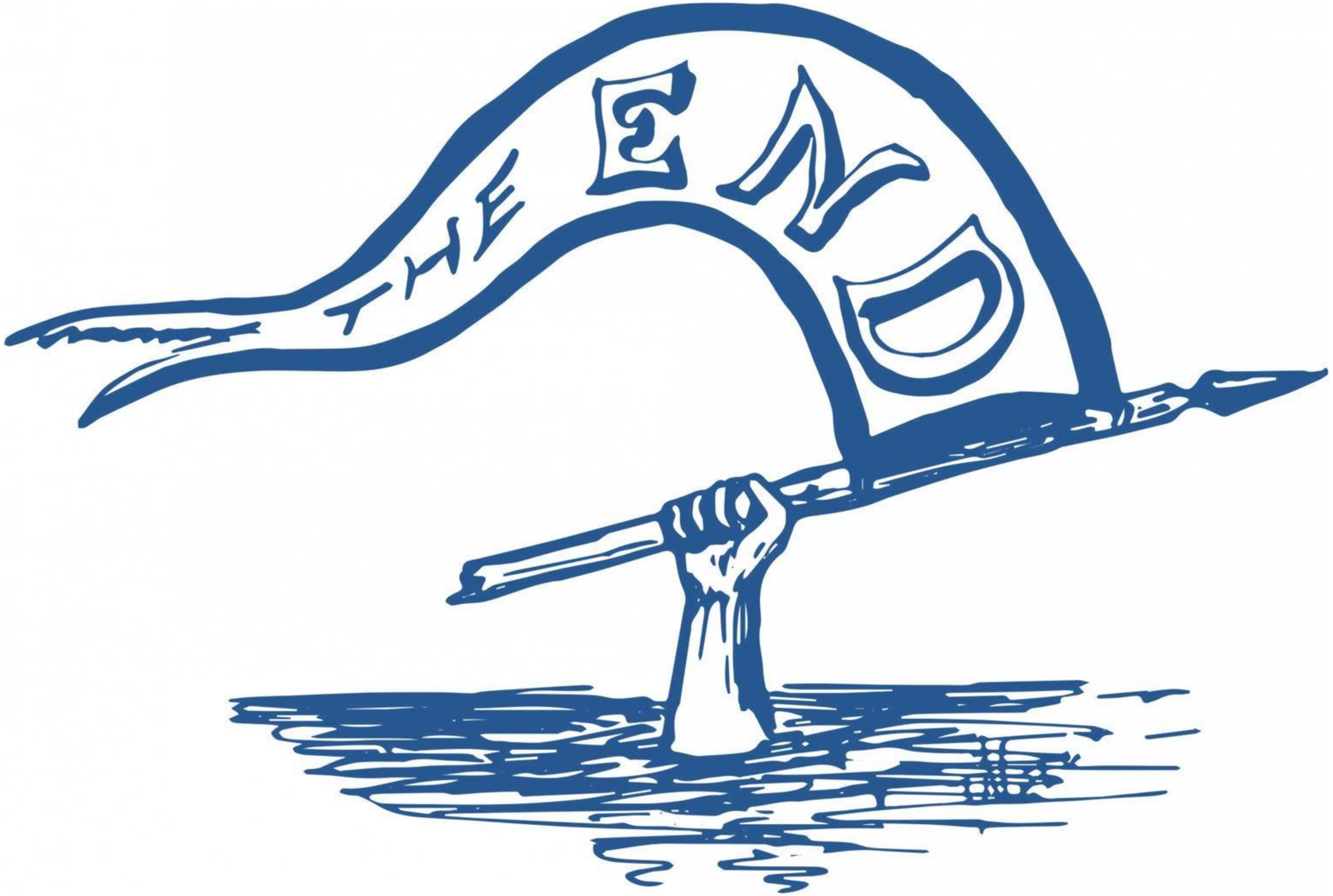
CONCLUSION

In summary, this endeavor effectively constructed a regression algorithm aimed at forecasting house values, leveraging diverse attributes including bedroom and bathroom counts, square footage, age, and location. The model demonstrated commendable precision in estimating house prices and shed light on the influential factors affecting housing costs. Such insights can prove invaluable to both real estate professionals and prospective buyers seeking precise assessments of property values based on their specifications.

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

Any Questions….


THE END