King County House

Sales Analysis

Understanding House Prices in King

County (2014-2015)

Presented by:

1. MERCY SILALI
2. WAFULA SIMIYU
3. JEFFREY KANYI
4. MICHAEL ARITA
5. ALEX MININGWA

OUTLINE

Quick Reminders

01

Our Project focuses on use of use regression modelling to analyse house sales in a northwestern county. (Kings County: 2014-2015)

02

This presentation will take around 5 minutes.

03

Feel free to ask questions during the Q&A session that follows the presentation.

INTRODUCTION

Housing is a major issue that seems to increase with the ever growing population. People not only search for secure but comfortable housing with various factors to consider to ensure proper living. However, prices in the housing sector fluctuate over time, being influenced by various factors such as size, location, number of bedrooms, year of renovation, grade, and number of flows, just to mention a few. Having the ability to predict the prices of houses basing on various factors can be vital to real estate agencies and other stakeholders in the real estate industry.

PROBLEM STATEMENT

House prices are influenced by multiple factors, making it challenging to predict prices accurately.
Accurate predictions can assist in better pricing strategies and investment decisions.

OBJECTIVE

To understand the factors influencing house prices in King County and create a tool to predict these prices.
The project aims to analyse King County House Sales between the year of 2014 and 2015. The analysis is to investigate how different factors such as the size of the house, the condition, the grade and other factors affect the house prices.
By building a regression model, we aim to investigate the determinants of house prices in order to potentially create a predictive tool for use real estate professionals and home owners.

Project Goals

❖ Explore the house sales data to see what it includes.

❖ Clean and organize the data for accurate analysis.

❖ Find patterns and trends in house prices.

❖ Build a model to predict house prices.

❖ Understand how different factors like house size and condition affect prices.

❖ Provide insights and tools to help stakeholders make informed decisions

Methodology

- ❖ **Understanding Stakeholder Needs**: Identifying how different stakeholders will benefit from the analysis.

- ❖ **Exploring the Data**: Reviewing what data is available and relevant.

- ❖ **Cleaning the Data**: Removing inaccuracies and filling in missing values.

- ❖ **Analysing the Data:** Looking for trends and patterns.

- ❖ **Building a Model:** Creating a tool to predict house prices.

- ❖ **Evaluating the Model:** Ensuring the tool provides accurate predictions

Data Overview

Dataset: Information on houses sold in King County during 2014-2015.Key Features: House size (sqft_living, sqft_above), condition, grade, bathrooms, year built, location, and price.

Price vs. Square Footage of Living Space
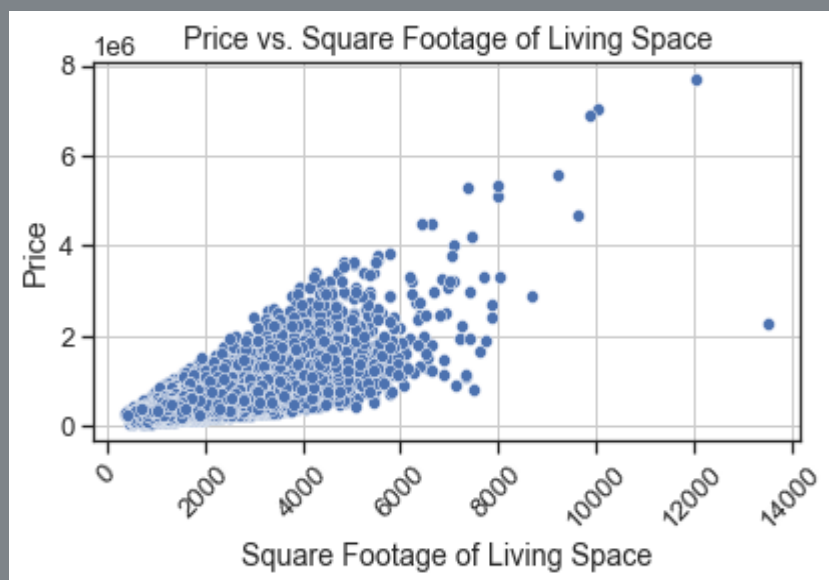
The scatter plot above shows a positive correlation between the price and the sqft_living indicating a general upward trend.
The larger the sqft_living the higher the price. However there is a also a significant scatter indicating that other factors might also affect the prices.
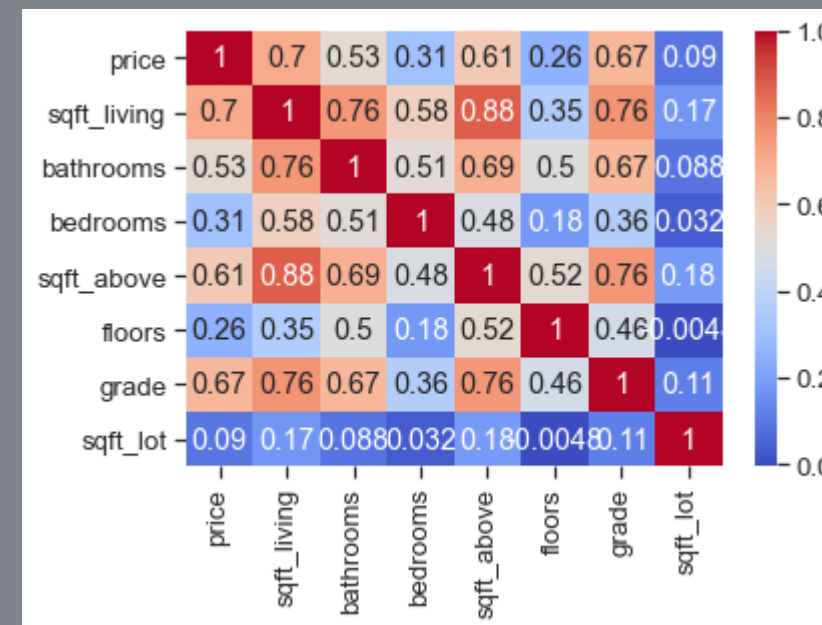
HEATMAP: Values closer to 1: Represent a strong positive correlation, meaning as one variable increases, the other tends to increase as well.
Values closer to -1: Indicate a strong negative correlation, where an increase in one variable is associated with a decrease in the other.
Values close to 0: Suggest little to no linear relationship between the variables.
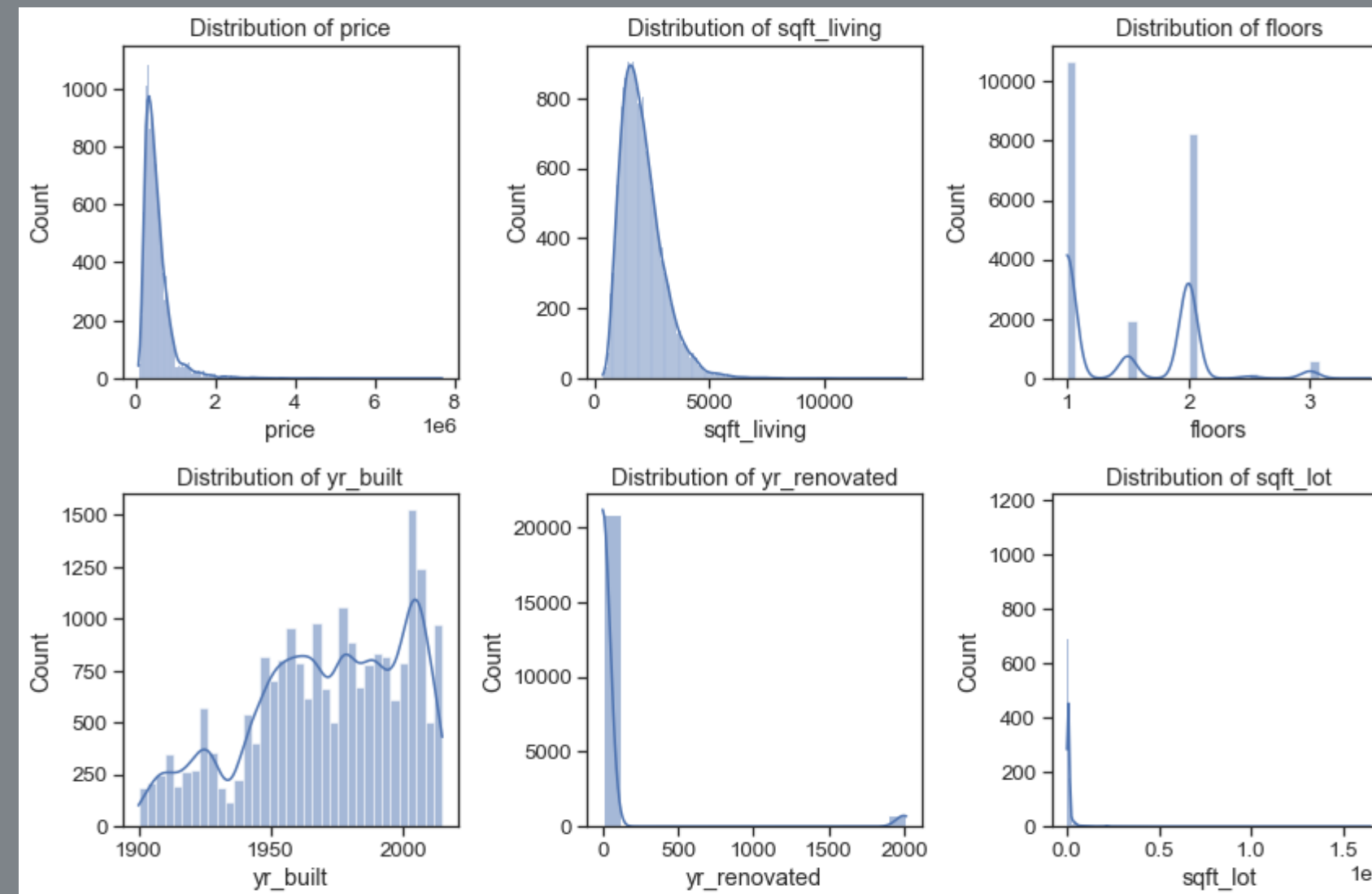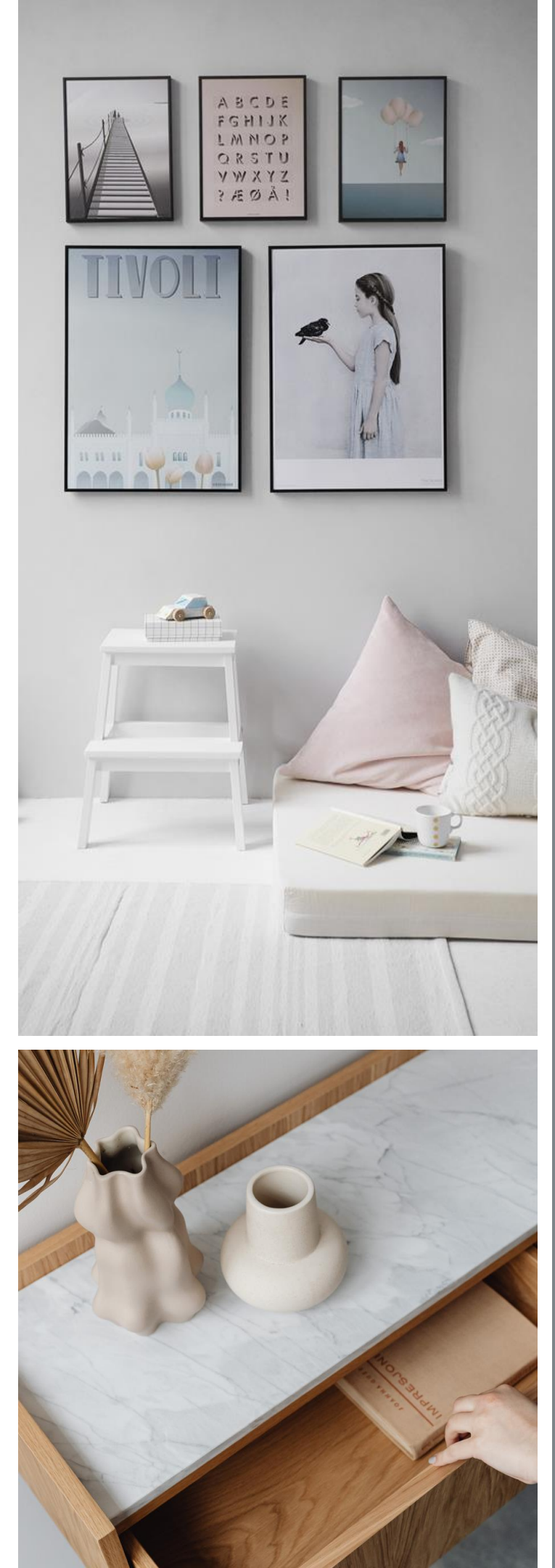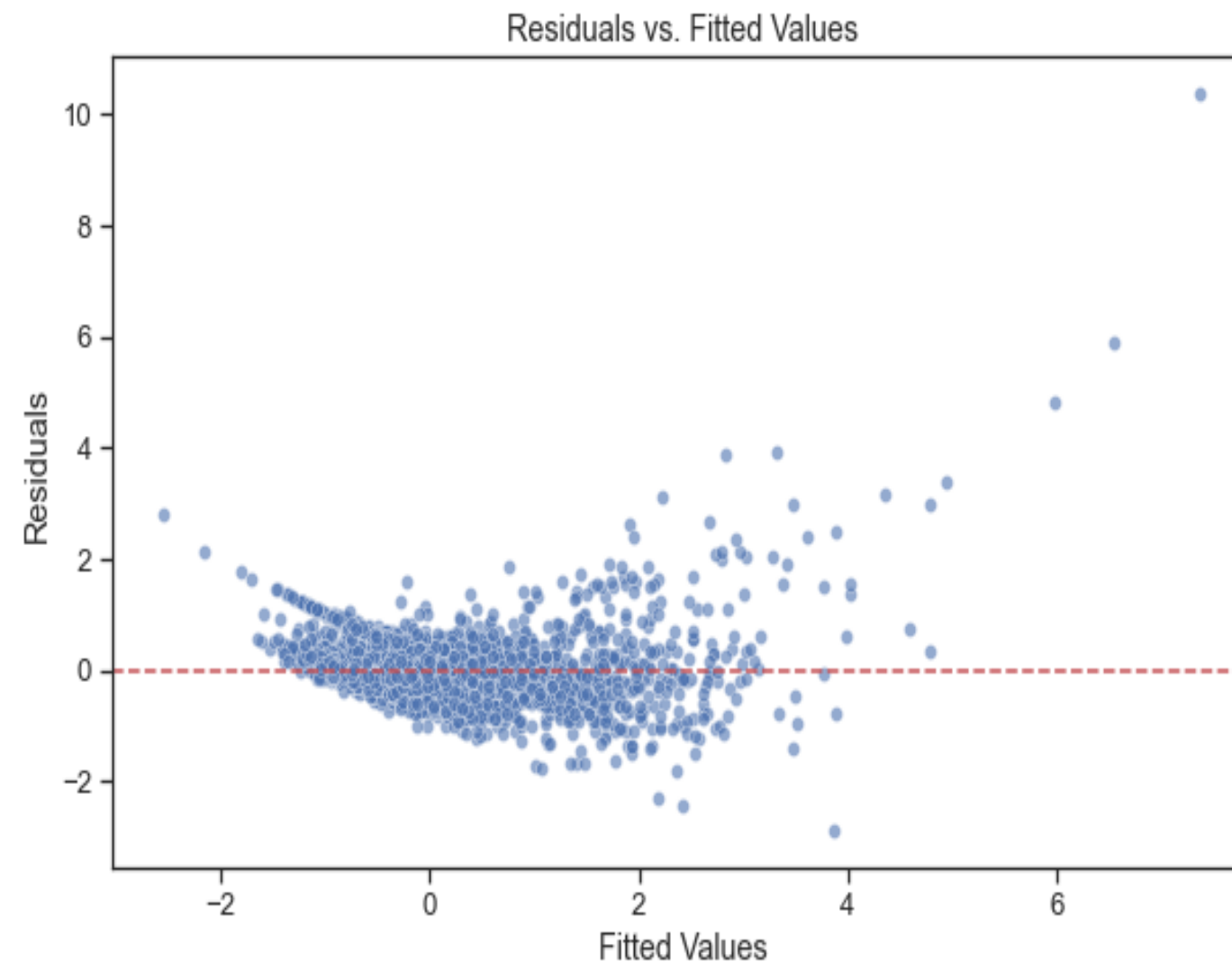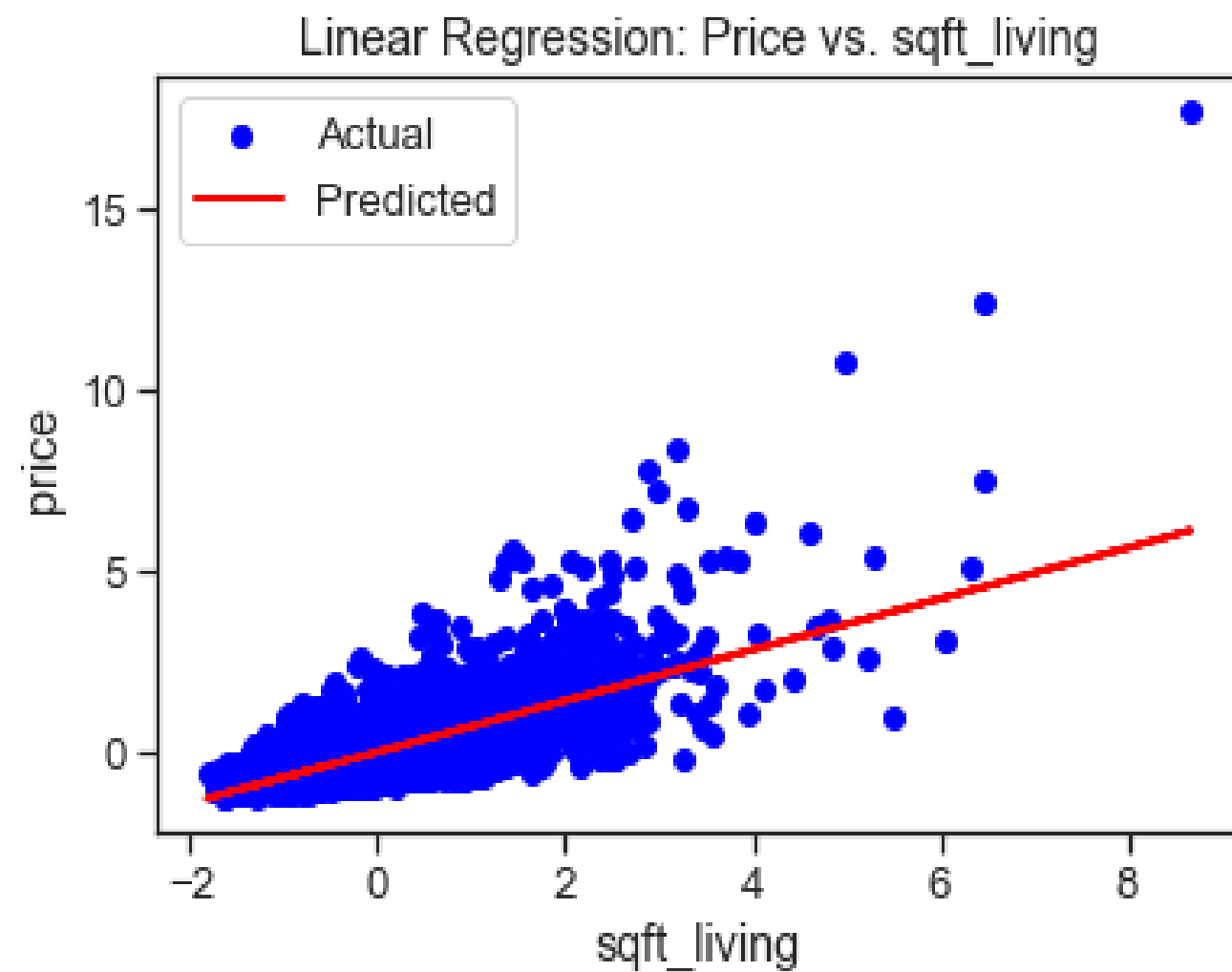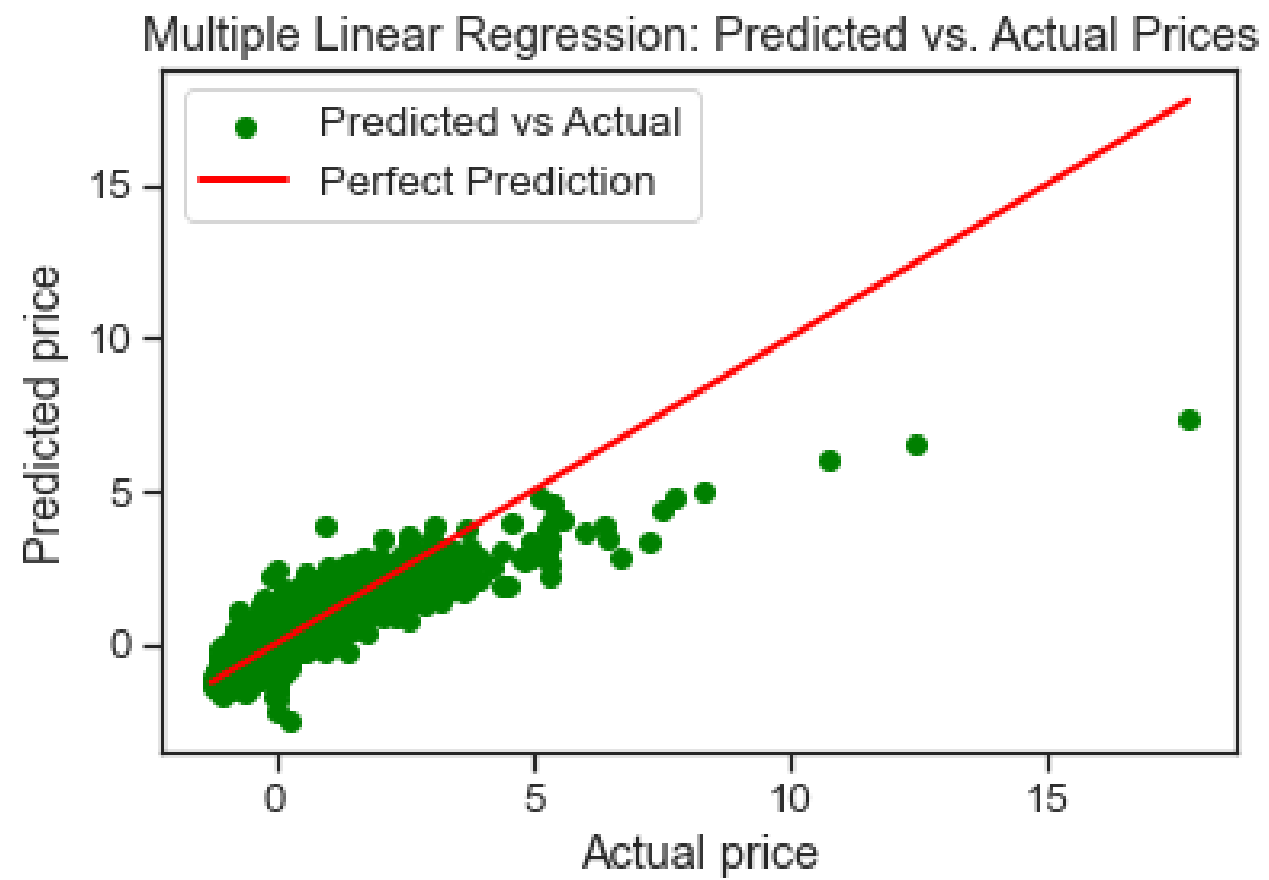

❖ Larger houses generally have higher prices.
❖ Newer houses tend to have higher prices .
❖ Year of renovation and grades influenced the price of houses.

Modelling.

- ❖ Simple linear regression

- ❖ Multiple linear regression

- ❖ Homoscedasticity Check

Modelling.

❖ Simple linear regression

❖ Multiple linear regression

❖ Homoscedasticity Check

Simple Linear Regression: From this model, the R_squared score was 0.49 meaning that the sqft|_living affected the price by 49%. This supports that the 51% is affected by other factors.

Multiple linear Regression: All factors were included and the model's r_squared score increased to 0.73 (73%) meaning that other factors also affect the prices.
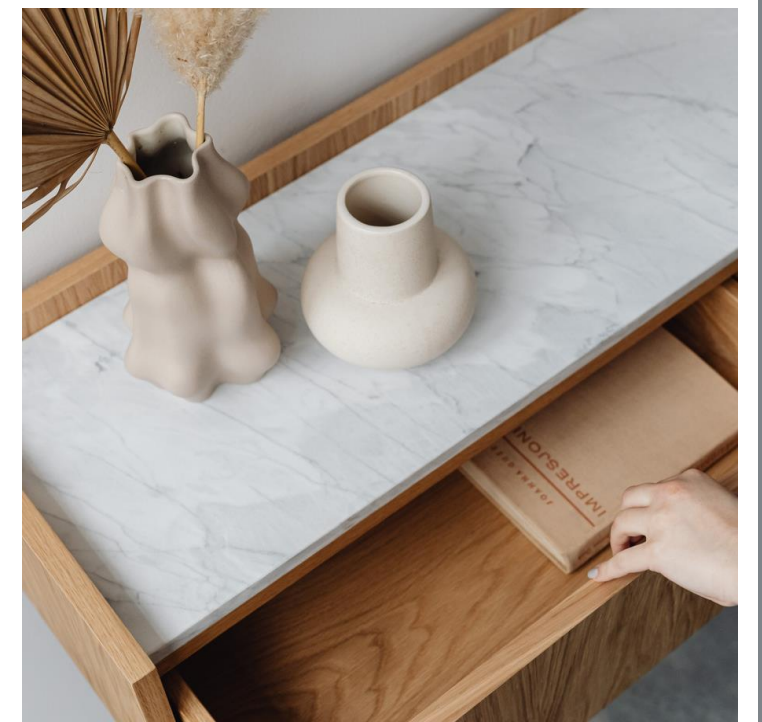
Results

- Based on the simple linear regression model, there is a relationship between the price and the sqft_living. However, some prices deviate further away from the line of best fit indicating that there are other factors that affect the prices.
- A multiple regression model gives further insights on factors that affect the prices of the houses since it analyses the relationship between price and other features such as the floors bathrooms and grade.
- The spread of residuals is random signifying homoscedasticity and roughly the same for all values of the independent variables.

Benefits for Stakeholders
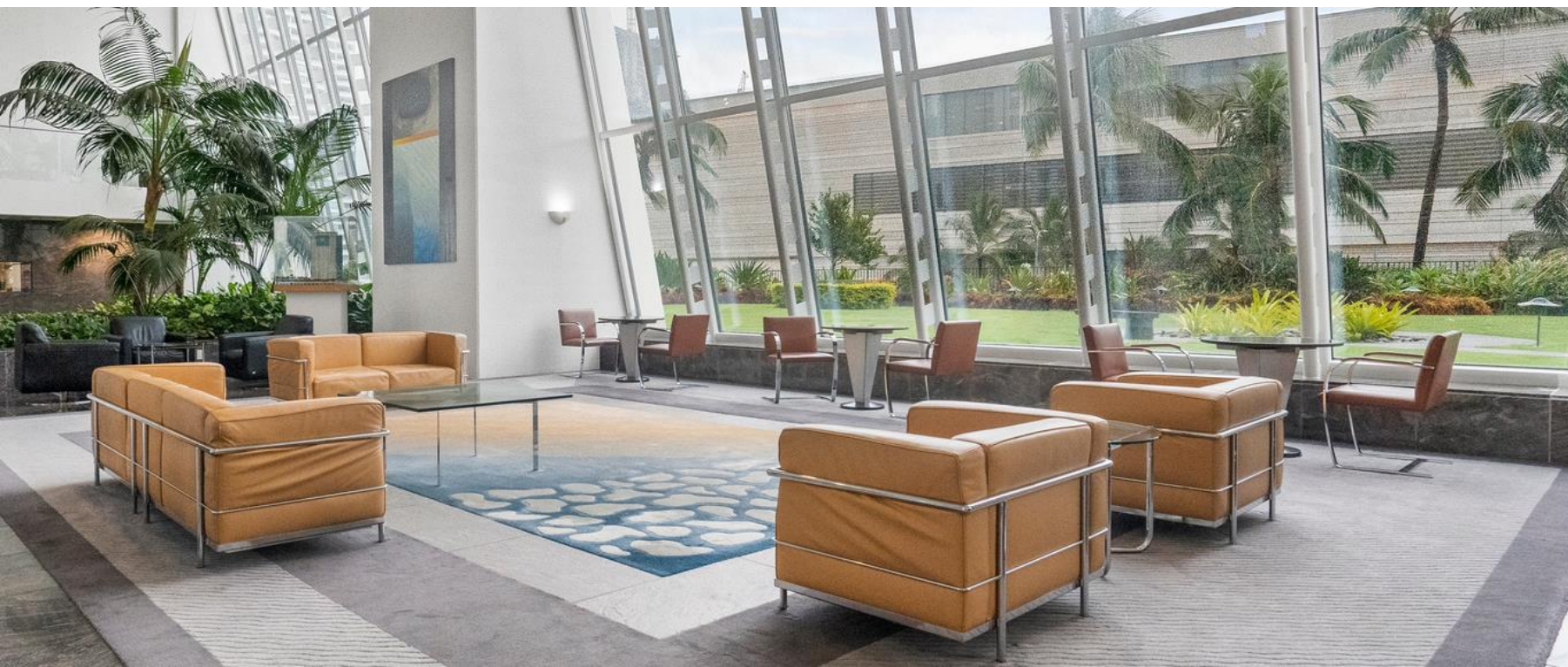
- **Real Estate Companies:** Set accurate prices and improve market strategies.
- **Homeowners**: Estimate property values before selling.
- **Buyers:** Determine fair market prices and make informed purchase decisions.
- **Policy Makers:** Develop data-driven housing policies to address market needs

Conclusion:

Summary: Our analysis and predictive models provide valuable insights into the King County housing market.

Impact: The findings and tools help stakeholders make more informed decisions regarding house sales and purchases..

Recommendations

❖ Collect more detailed data on house features like interior renovations and landscaping to enhance the model's ability to capture the true value of unique house characteristics.

❖ Integrate economic factors such as interest rates, employment rates, and income levels since they significantly impact housing markets, and their inclusion can improve model accuracy.

❖ Ensure the model uses the most recent data available by regularly updating the. Dataset which will help the model remain relevant and accurate in a dynamic housing market.

❖ Incorporate data from surrounding areas and regions to capture a broader market perspective which can help the model account for regional price differences and market trends.

❖ Engage with Stakeholders

❖ Include external factors such as school quality, crime rates, and proximity to amenities since they play a crucial role in pricing.

❖ Regular testing and validation help identify weaknesses and areas for improvement, ensuring the model's reliability.

Thank You
Q&A Session

Thank you for listening.