

HOUSE PRICE FORECASTING WITH REGRESSION MODELING



GROUP 8

OCTOBER,
2023

Table of Content

03

PROJECT OVERVIEW

04

BUSINESS
UNDERSTANDING

05

PROBLEM STATEMENT

06

OBJECTIVES

07

DATA UNDERSTANDING

08

VISUALIZATION

09

MODELLING AND
EVALUATION

10

CHALLENGES

10

CONCLUSION

11

RECOMMENDATIONS

• • • • •
• • • • •
• • • • •
• • • • •
• • • • •





PROJECT OVERVIEW

The project focuses on the creation of a machine learning project for house price forecasting for all stakeholders involved from investors to owners to buyers.

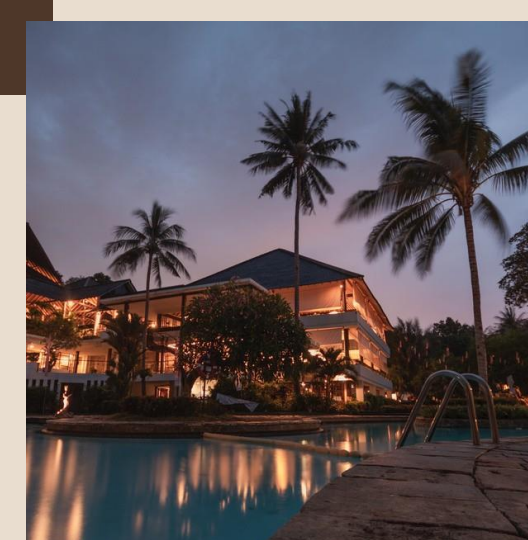
BUSINESS UNDERSTANDING

House price forecasting is a crucial task in the real estate industry. Accurate predictions assist homebuyers, sellers, and investors in making informed decisions regarding property transactions.



PROBLEM STATEMENT

Real estate developers encounter difficulties when assessing the precise influence of individual metrics and attributes on house pricing within the KC housing dataset. Their primary concern is the degree to which these factors interact to affect pricing outcomes. The current lack of clarity in pricing decisions can result in instances of both overpricing and underpricing of properties.



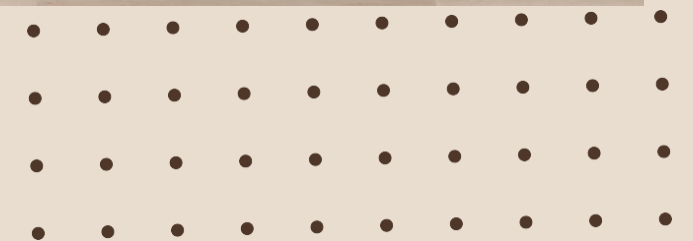
OBJECTIVES

General Objective

- To develop and implement a ML based solution that leverages data analytics and predictive models to give accurate insights and enable real estate developers to make informed decisions

Specific Objectives

- To identify and select the most relevant features to build the model
- To determine the peak season for house sales
- To develop a price prediction model
- To provide insightful recommendations based on our analysis





DATA UNDERSTANDING

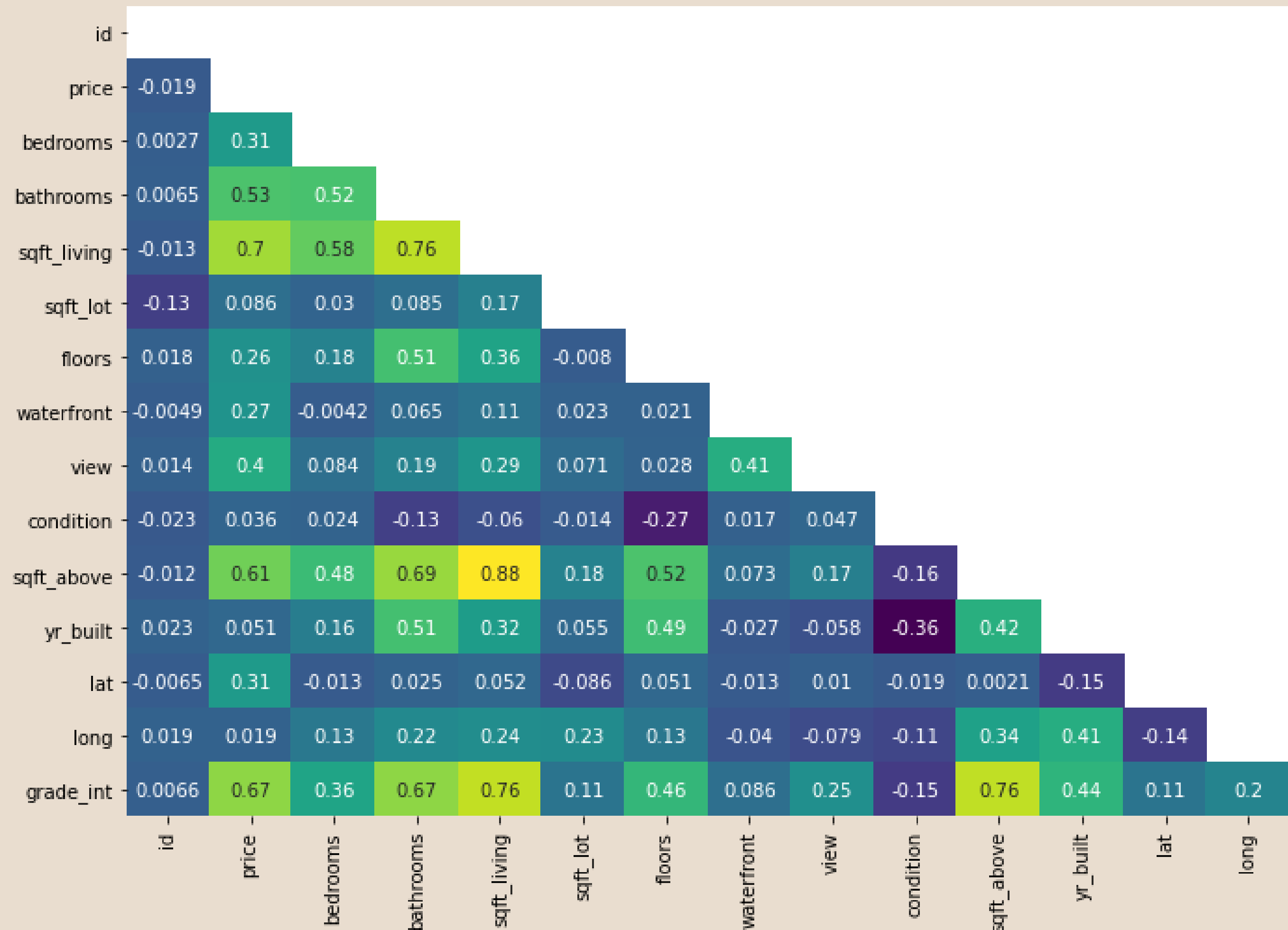
This project uses the King County House Sales dataset, which can be found in `kc_house_data.csv` in the data folder in this GitHub repository. The following was done:

- 01 performed a series of preprocessing steps to parse the dataset to obtain insights
feature engineering was done to come up
- 02 with new columns with more information for the data set
- 03 columns with null and no values were dropped to allow for seamless analysis

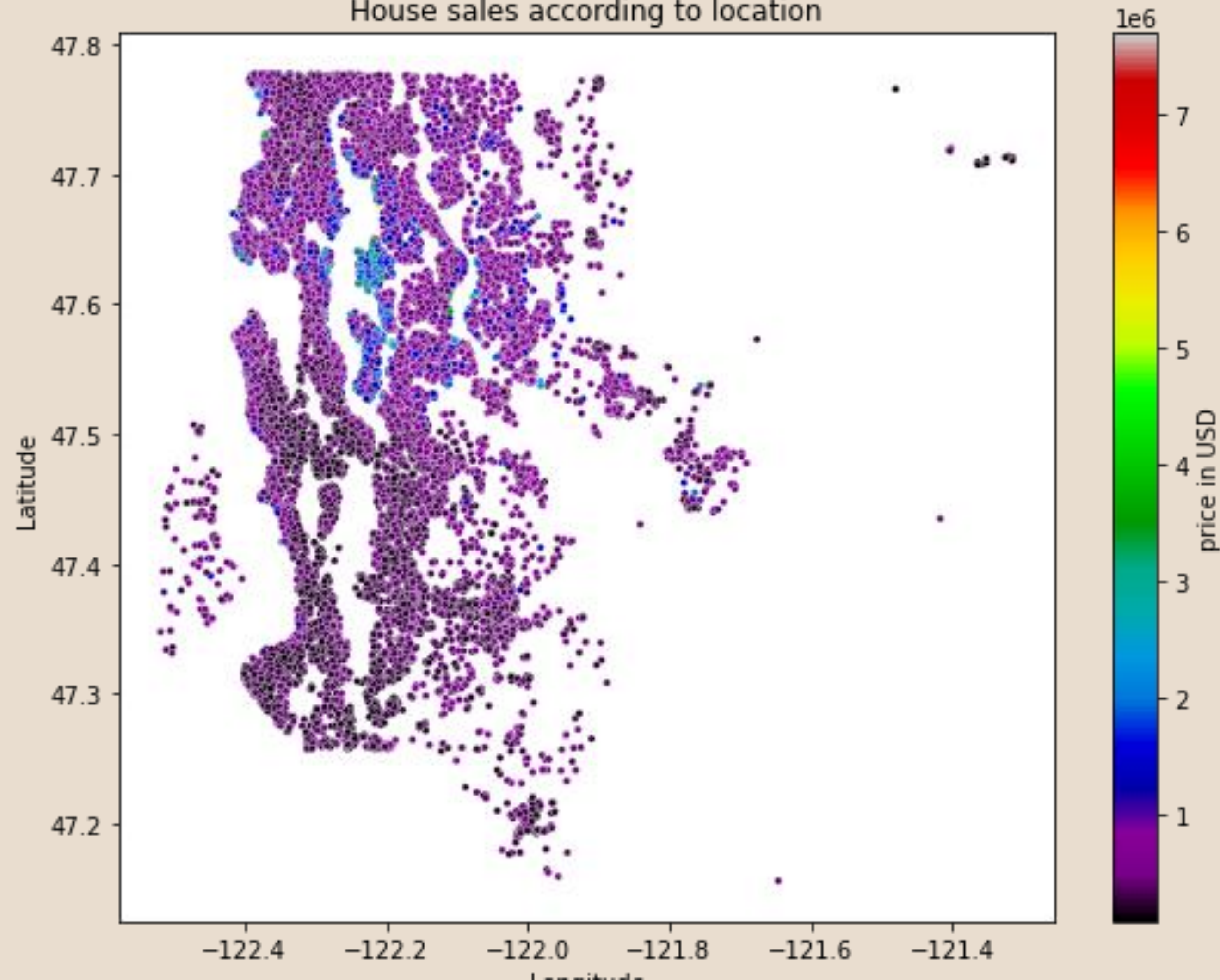


VISUALIZATION

For enhanced data analysis, the heatmap visualizes the relationships and correlations between various columns within our dataset. The darker the box the higher the correlation.

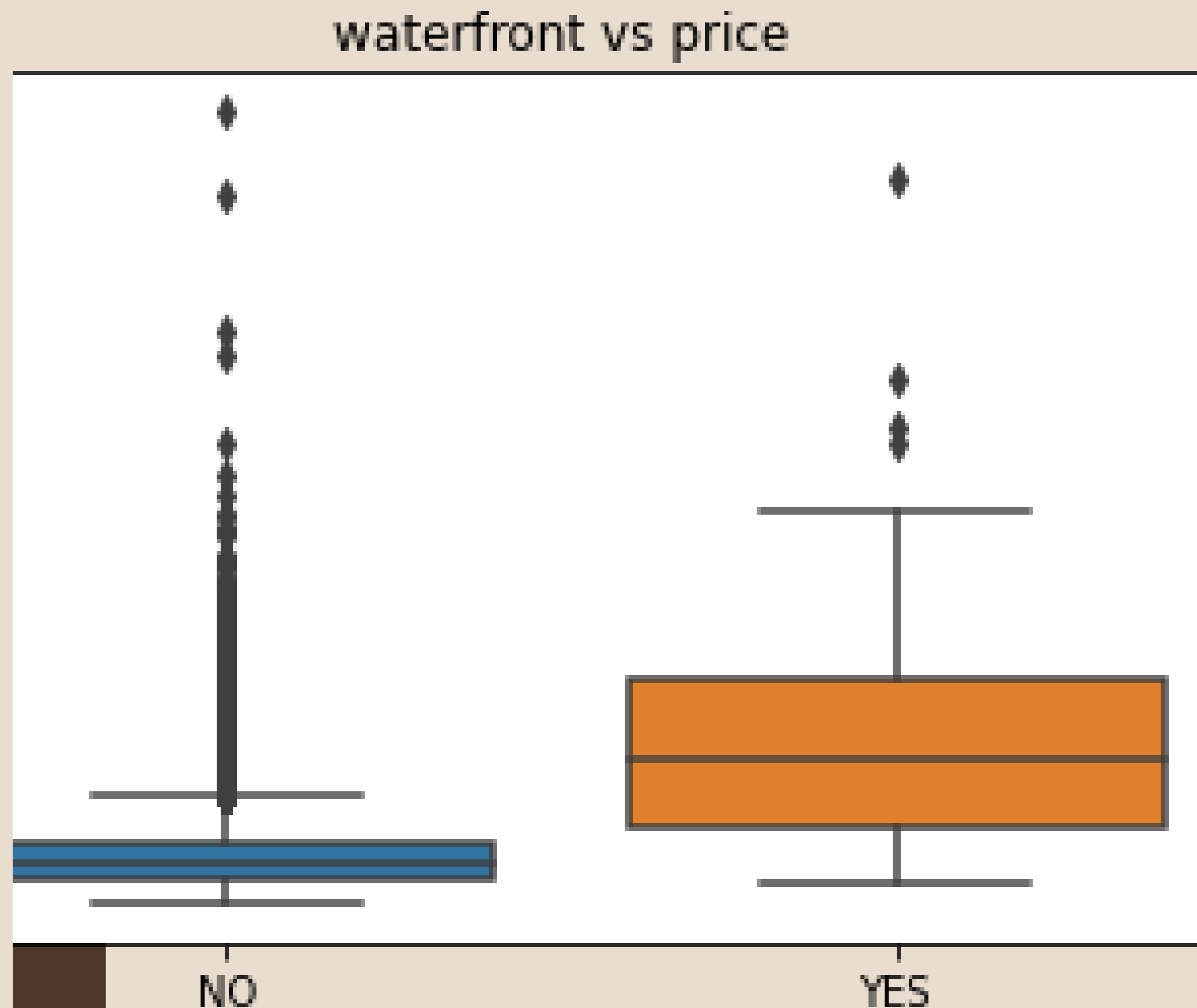


House sales according to location



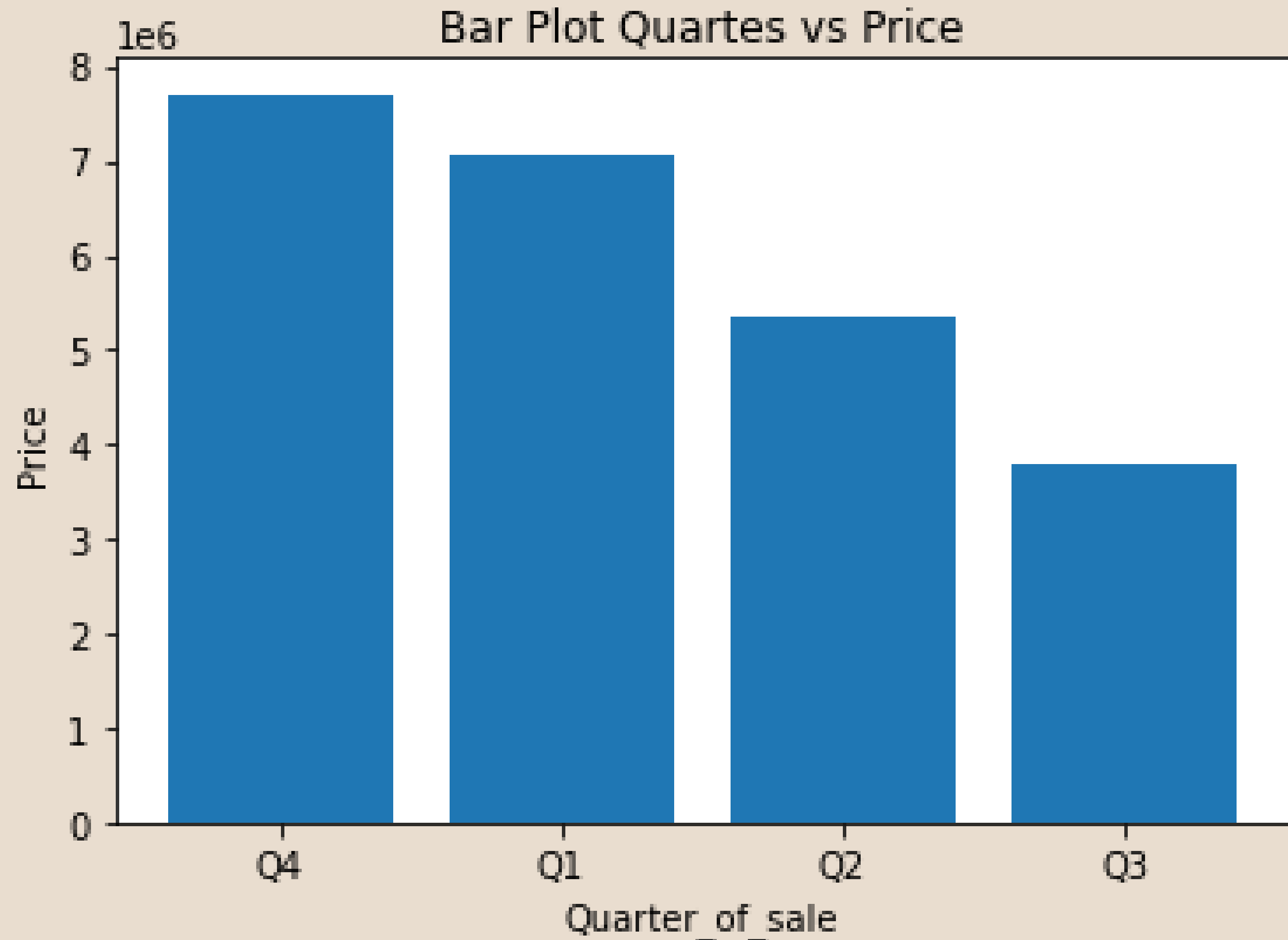
From the visualization, houses around latitude 47.6 to 47.7 and longitude -122.0 to -122.5 have some of the highest prices.

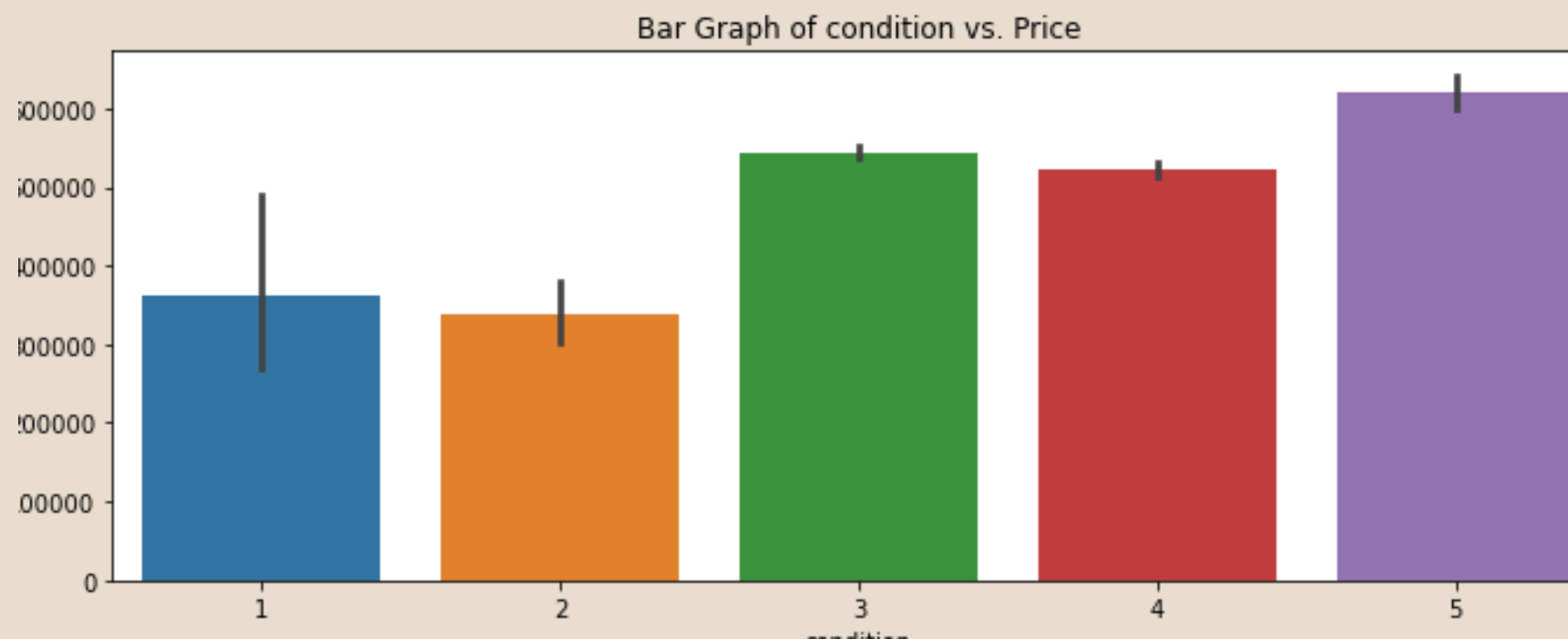
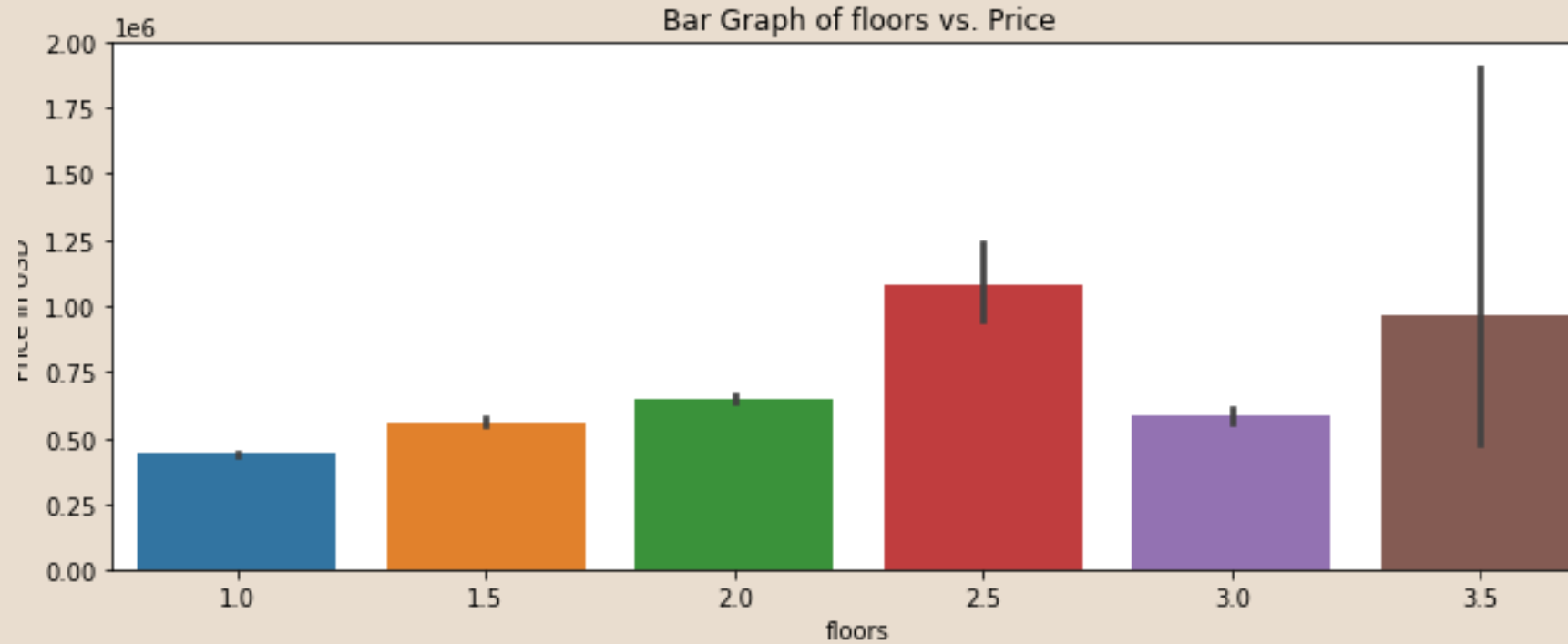
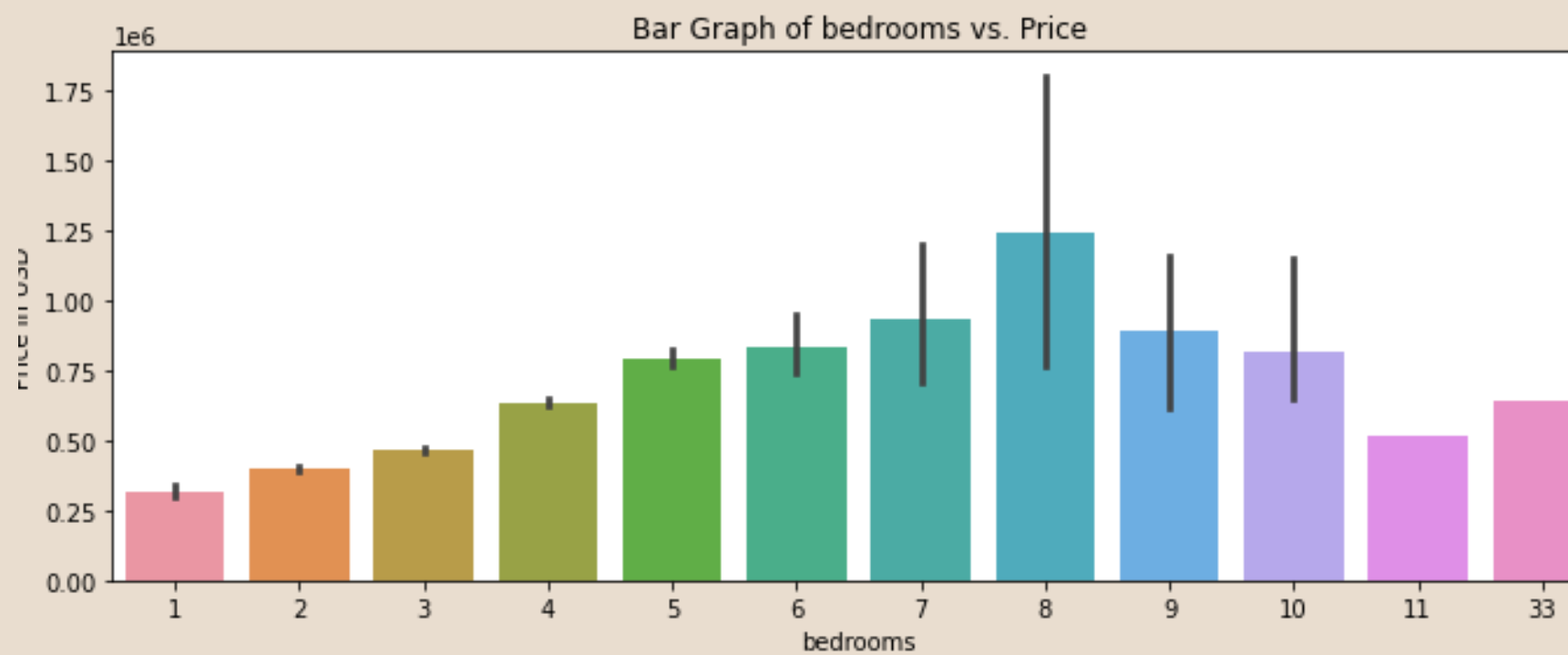
Houses near the waterfront are more expensive, hence developers should consider waterfront properties for high returns



More houses are sold during the fourth quarter of the year,

 GROUP 8





- It can be noticed that as 'bedrooms' increase, so does the house's selling price
- more 'floors', preferably up to 2.5 have a higher price
- House prices also depend on the condition of the house

MODELLING

GROUP 8

Simple linear regression R- squared of 0.49

this indicates that 49% of variation can be

explained by sqft living

Multiple Linear regression R-squared of

0.69 which implies that 69% of the

variation in price can be explained by the

independent variables

Random Forest Regressor R -squared of

0.87





EVALUATION

we evaluated several models and
settled on with the lowest rmse and
highest r^2 of 87

CHALLENGES

1. Historical Data: Tracking historical house price changes and their geographical variations is currently impossible
2. External Factors: The housing market is susceptible to external factors such as changing policies and economic fluctuations, impacting price trends.
3. Insufficient Data: Limited data hampers the development of a recommendation system. We need more data on local amenities, security, and how they vary across geographical regions.



CONCLUSION

This project represents a significant step toward revolutionizing property valuation by leveraging machine learning. By providing accurate, accessible, and interpretable house price forecasts, the project adds substantial value to the real estate industry. As the project evolves, incorporating user feedback and embracing technological advancements will be pivotal in ensuring its ongoing success and relevance in an ever-changing market landscape.

RECOMMENDATIONS

- Real estate developers should focus on developing houses near the waterfront as they are more expensive hence high returns
- These features have the most impact on prices hence they can focus on incorporating them when developing houses
- Strategize and plan their marketing during the peak season which is the fourth quarter of the year during fall as that is when the most sales occur.

ACKNOWLEDGEMENT

This project is a testament to the incredible synergy that is born when a group of talented individuals comes together with a shared vision. It is our ability to collaborate, communicate, and work together seamlessly that has led us to success.

- 01 PRISCILLAH WAIRIMU (Team Leader)
- 02 BRIAN CHACHA
- 03 STEPHEN BUTIYA
- 04 LUCY WARUGURU
- 05 LEWIS KAMINDU
- 06 MESHAEL AYANGA ODUOR



THE
END

 GROUP 8

Thank You