



# KING COUNTY HOUSE PROJECT





# CONTENT

**01**

Business  
Understanding

**02**

Problem  
Statement

**03**

Objectives

**04**

Data Analysis

**05**

Modeling

**06**

Conclusion

**07**

Participants

# BUSINESS UNDERSTANDING

The real estate agency serves homeowners looking to buy and sell homes. For homeowners looking to sell their properties, understanding how specific home renovations might impact the estimated value of their homes is crucial. Therefore, the company needs data-driven insights to value properties accurately and competitively price listings.



# **PROBLEM STATEMENT**

Homeowners looking to sell their properties need to understand how specific renovations will impact the estimated market value of their home and make decisions when upgrading their homes in preparation for sale. Solving this will enable homeowners to strategically choose renovations that maximize return on investment and yield the greatest price premiums when listing their properties.



# OBJECTIVES



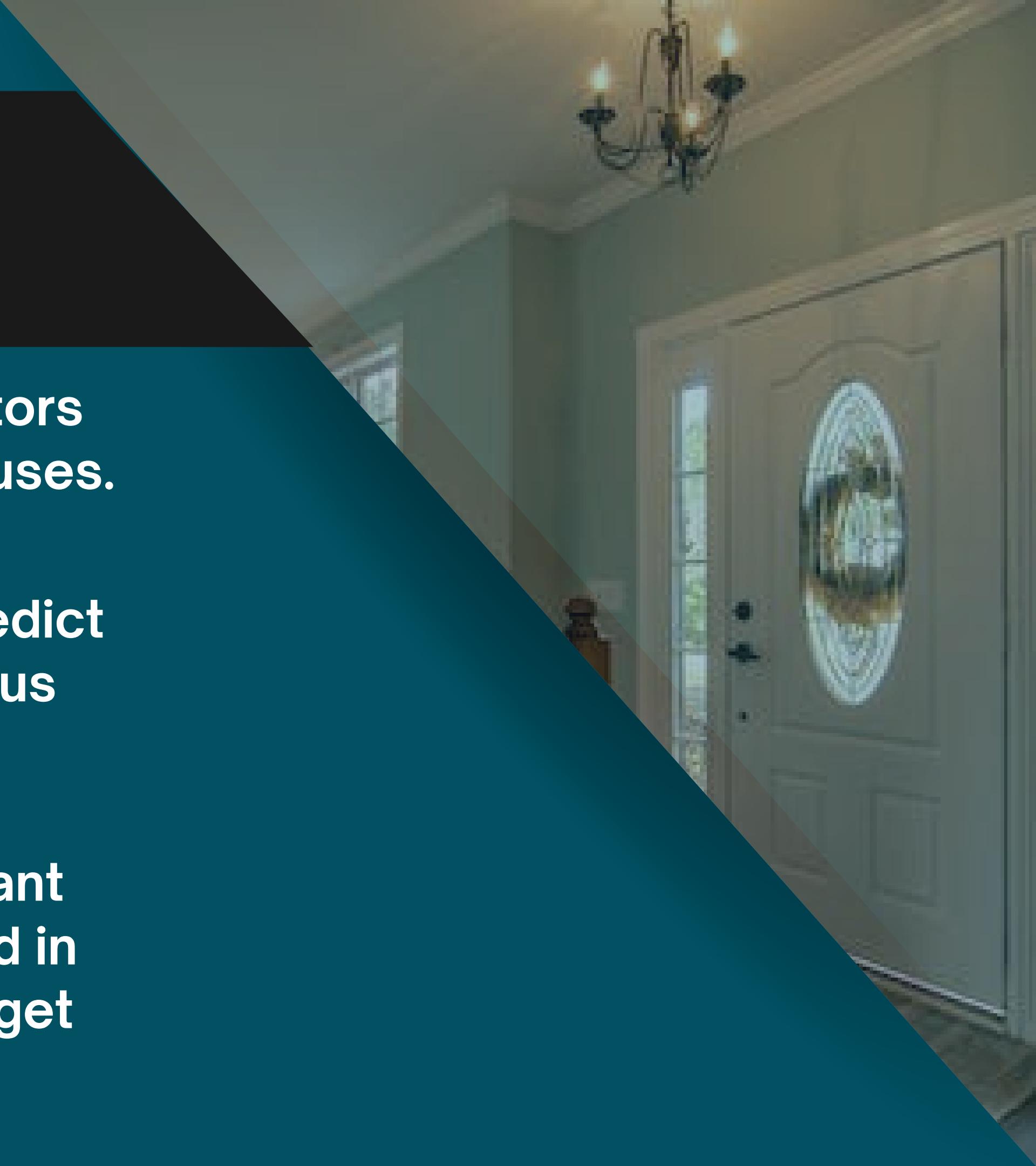
To understand the top four factors that influence the prices of a houses.

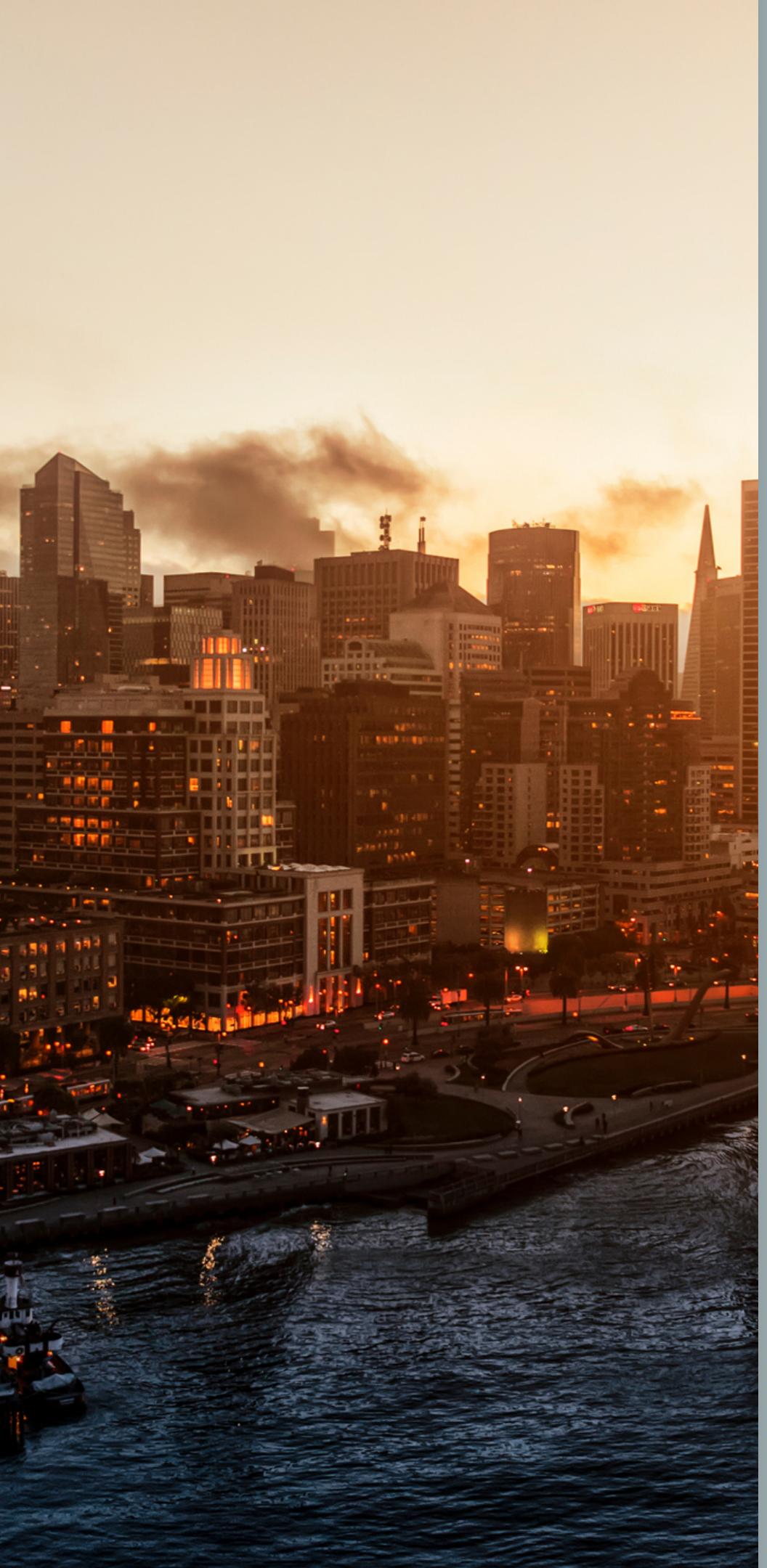


To develop a model that can predict housing prices based on various features.

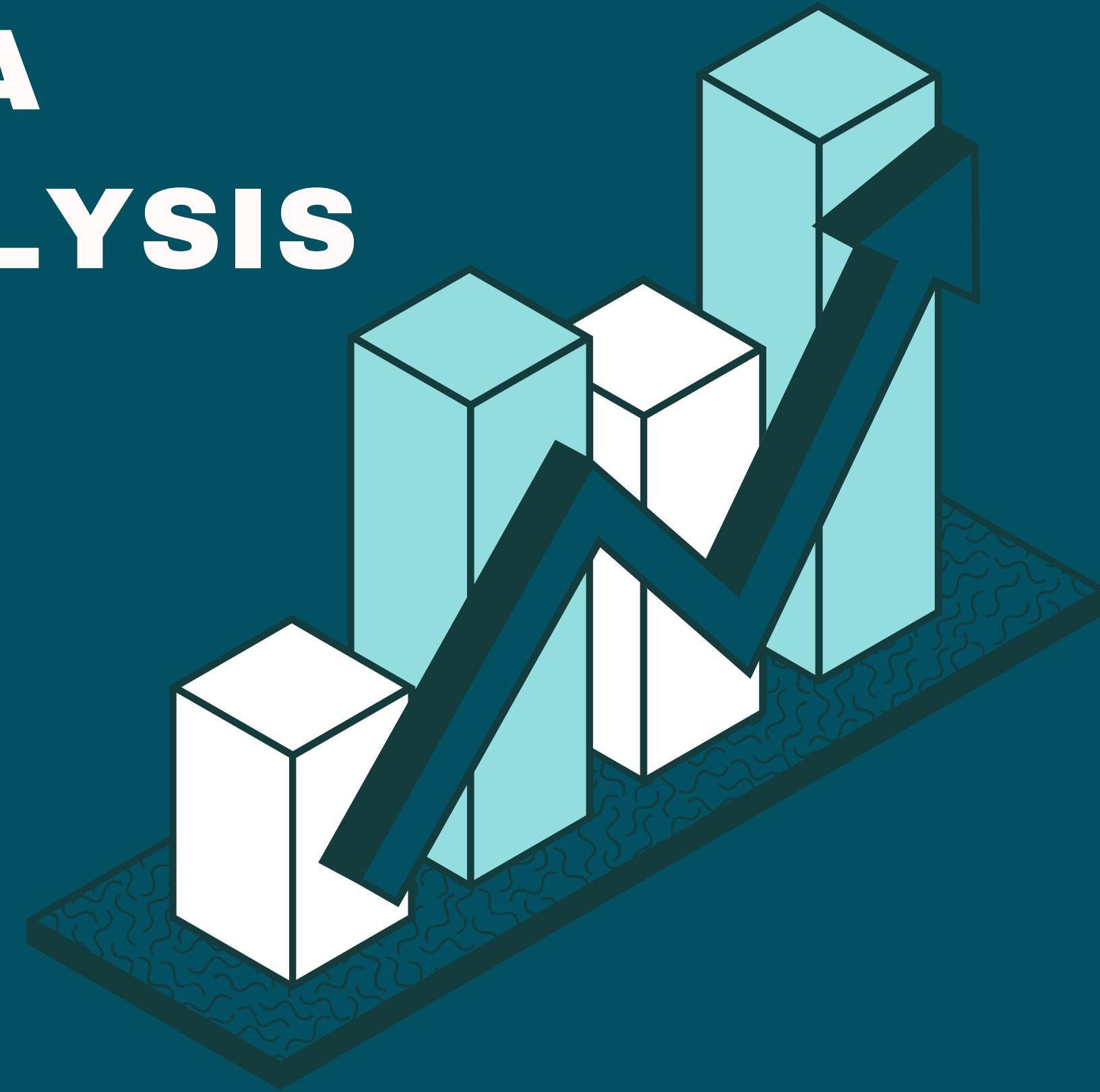


To investigate how the important factors affecting price obtained in the first objective vary with target variable.





# DATA ANALYSIS



# DATA IMPLEMENTED

- We looked at over 21,000 home sales records in King County provided by the real estate company
- The analysis was conducted based on two broad categories: the categorical variables and the continuous data.
- This data had sale price and many details for each home i.e Location, construction year, quality (grading) and space in square feet.
- We focused our analysis on features that seemed most predictive of the sale price.

# DATA PREPARATION

Handling  
missing and  
extreme  
values



Removing  
duplicate  
entries



Standardizing  
data formats



# DATA PREPARATION

Feature  
Engineering



Data  
Mapping



Correlation  
Matrix

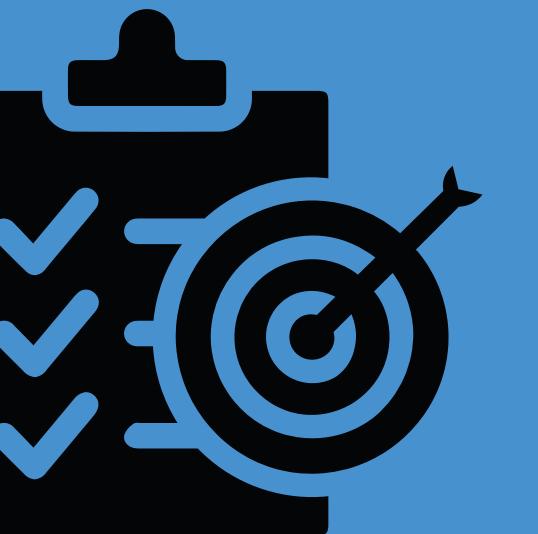


# DATA PREPARATION

Data  
Mapping



Univariate  
and Bi  
variate  
Analysis

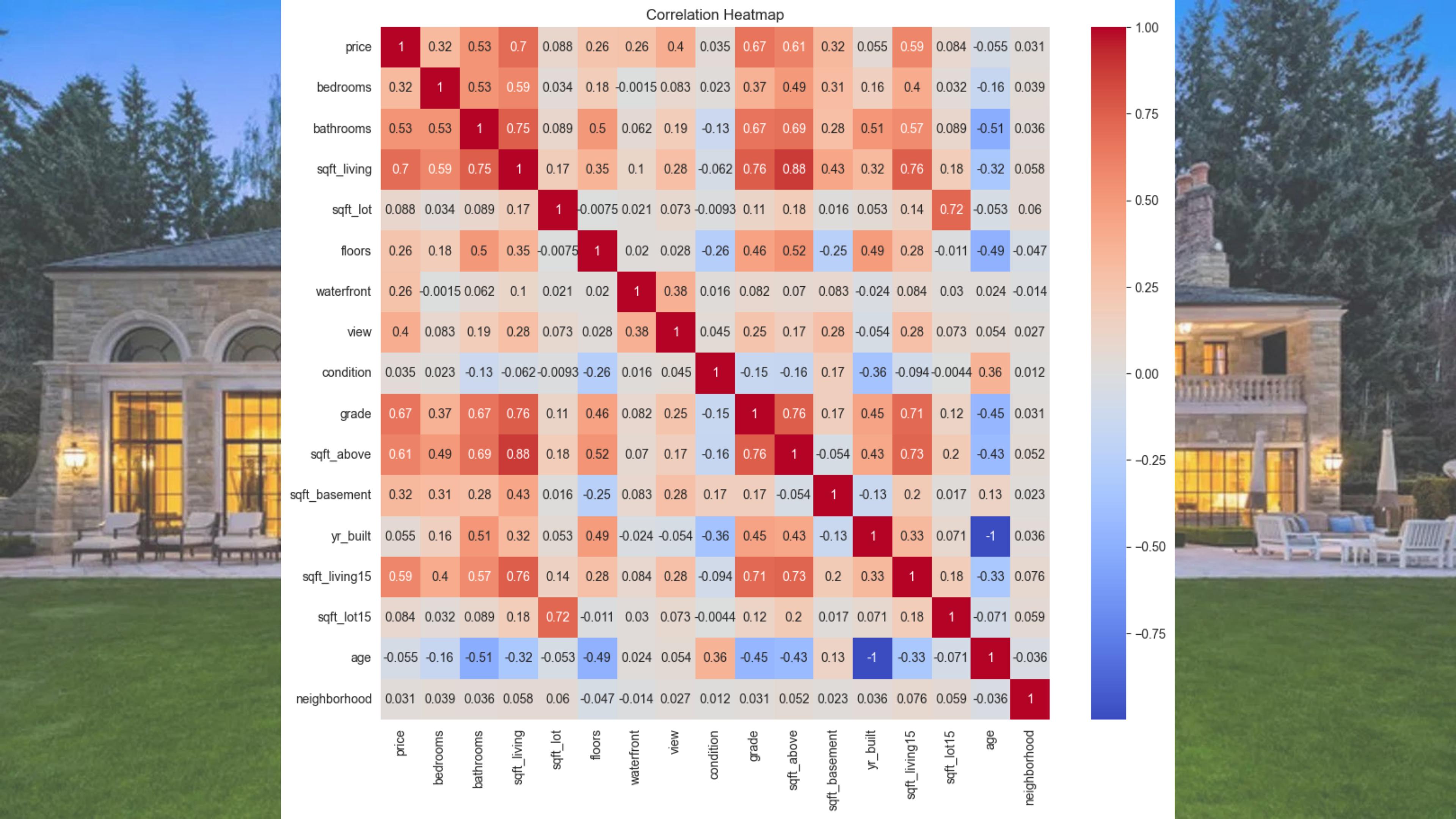


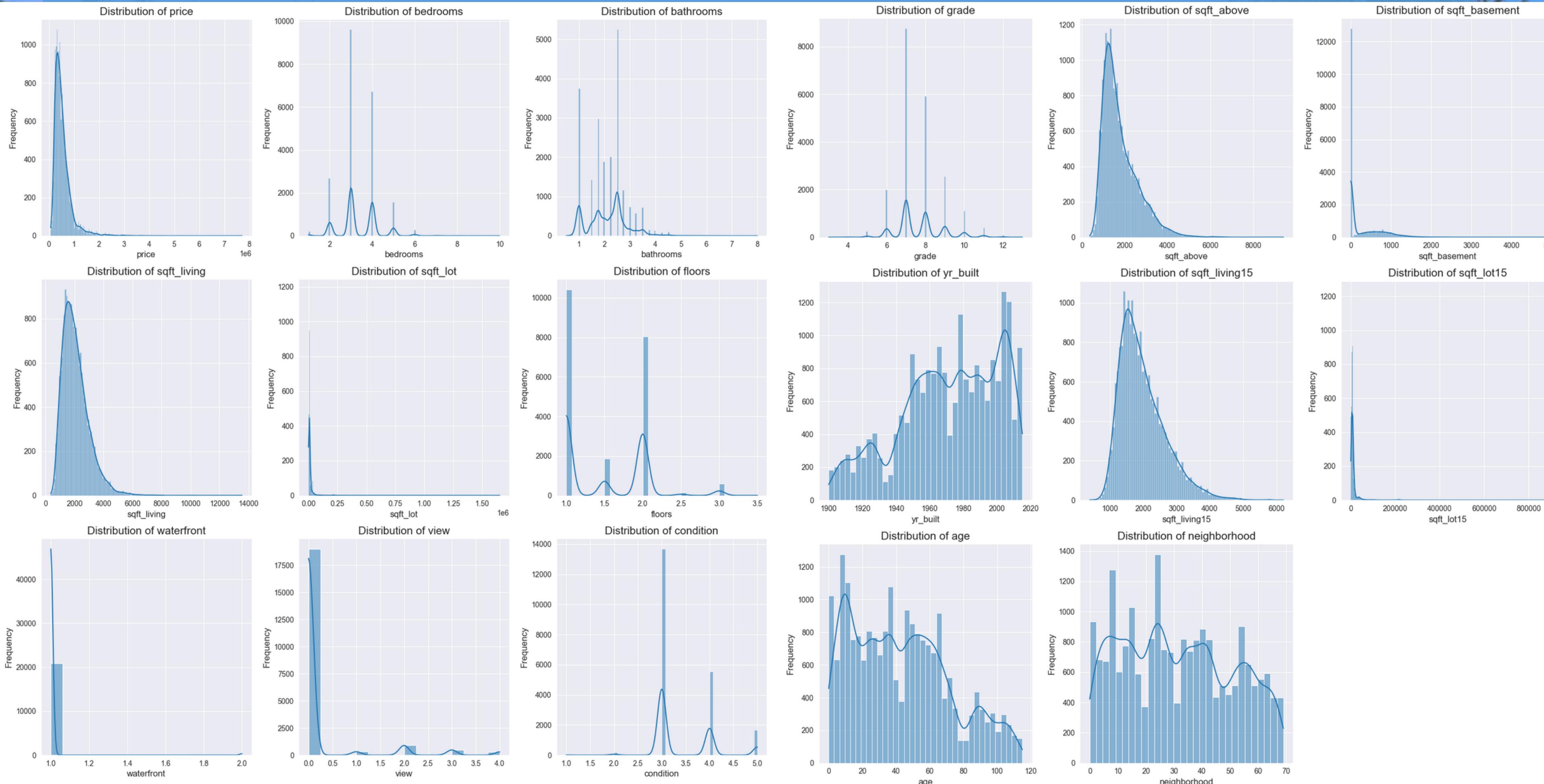
Correlation  
Matrix



# CORRELATIONS

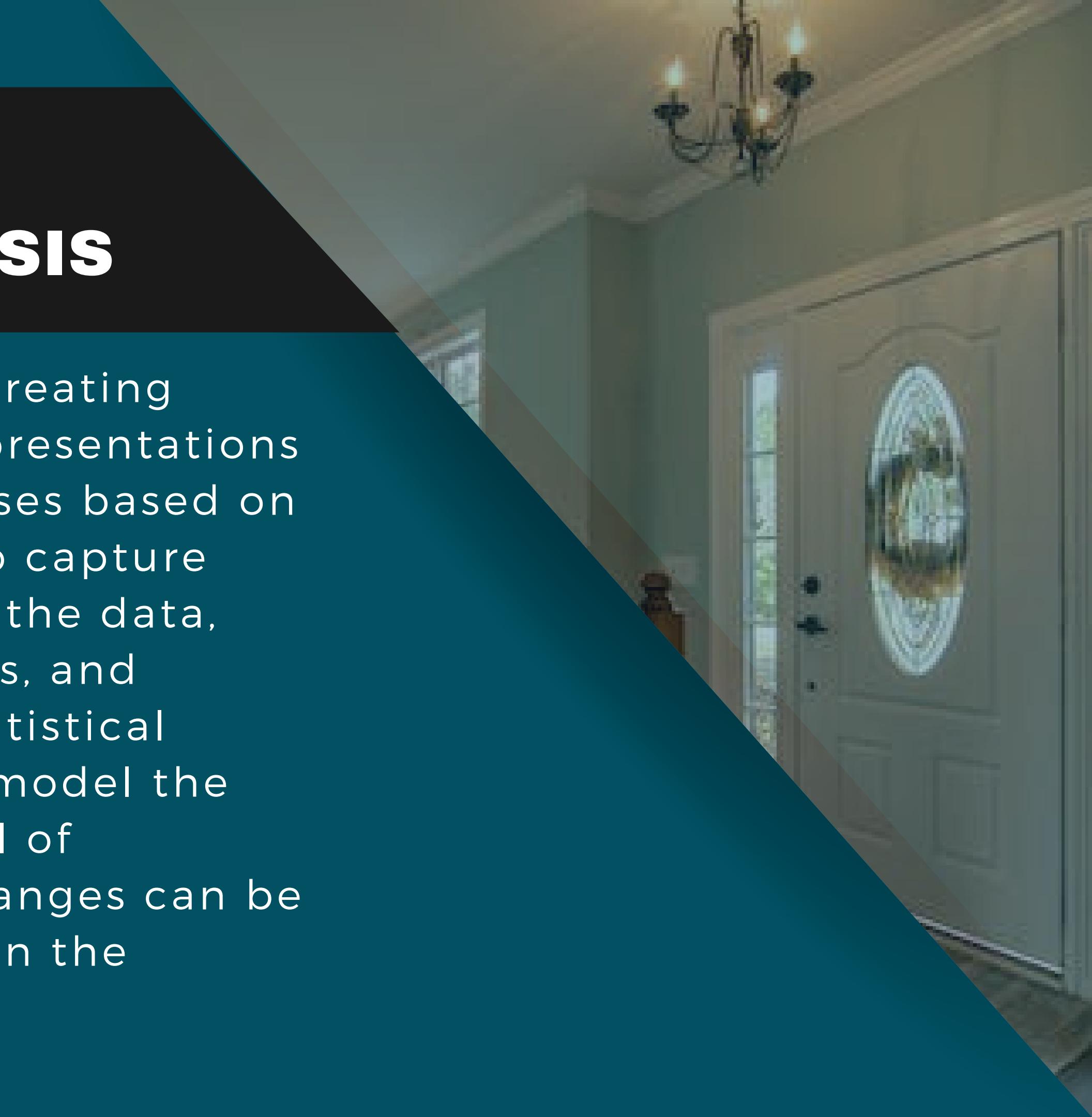
The analysis was conducted based on two broad categories: the categorical variables and the continuous data. We therefore created a heat map to visually represent the correlations between the numerical features to understand the relationships between the numerical features in our dataset. Then we looked at the relationships of the prices and the data.





# **MODELING AND REGRESSION ANALYSIS**

Modeling in data analysis involves creating mathematical or computational representations of real-world phenomena or processes based on observed data. These models aim to capture relationships, patterns, or trends in the data, allowing for predictions, simulations, and insights. Regression analysis is a statistical technique used in data analysis to model the relationship from the data. The goal of regression is to understand how changes can be affected by the relationship between the categories.

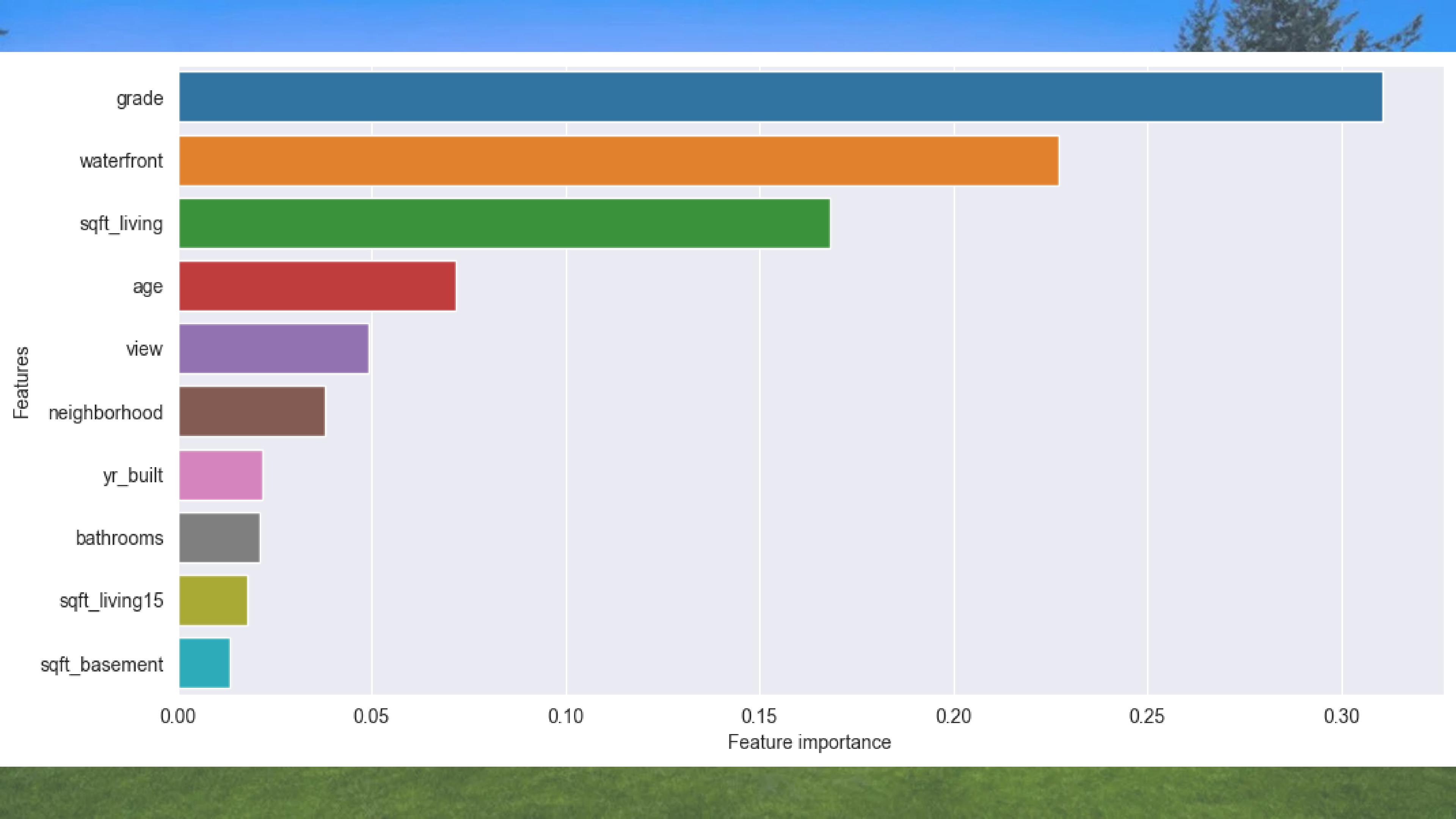


# **LIGHT GBM MODEL**

After testing out different modeling techniques, we found that a method called LightGBM worked best for predicting home prices accurately in this case.

LightGBM is a sophisticated modeling approach that combines many decision trees to learn complex relationships in data.

This model was able to predict prices with 87% accuracy on average when tested on thousands of real home sales it hadn't seen before. That level of accuracy gave us confidence in the model's capabilities.



# **BENEFITS OF USING LIGHT GBM MODEL**

1. It is able to run very efficiently on large datasets, even with limited computing resources.
2. It can take advantage of parallel computing,
3. It is very efficient in handling both large and small datasets.

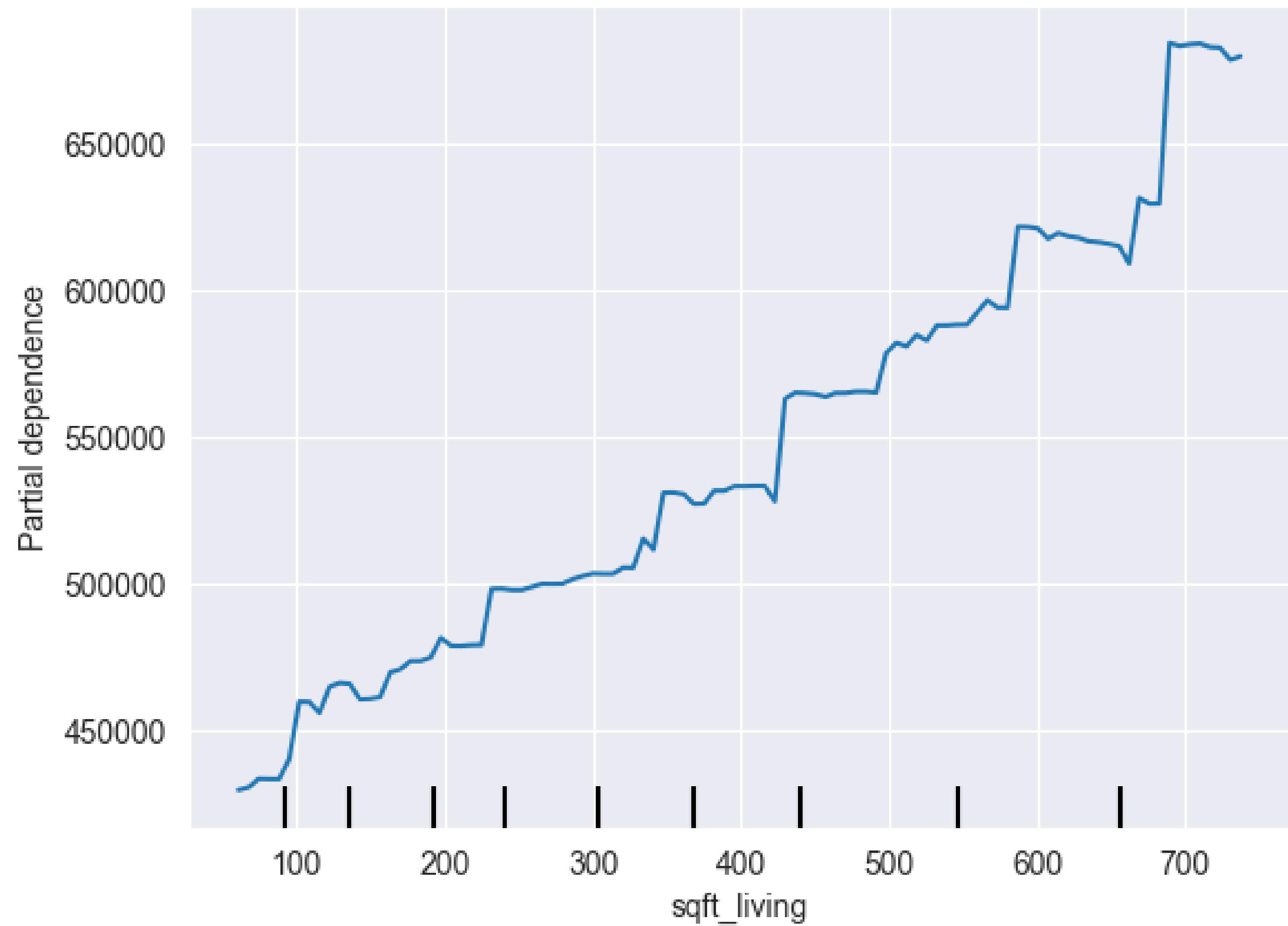
For a real estate company analyzing millions of home transactions across counties, LightGBM's speed and efficiency would allow generating useful price insights at scale while controlling computing costs.



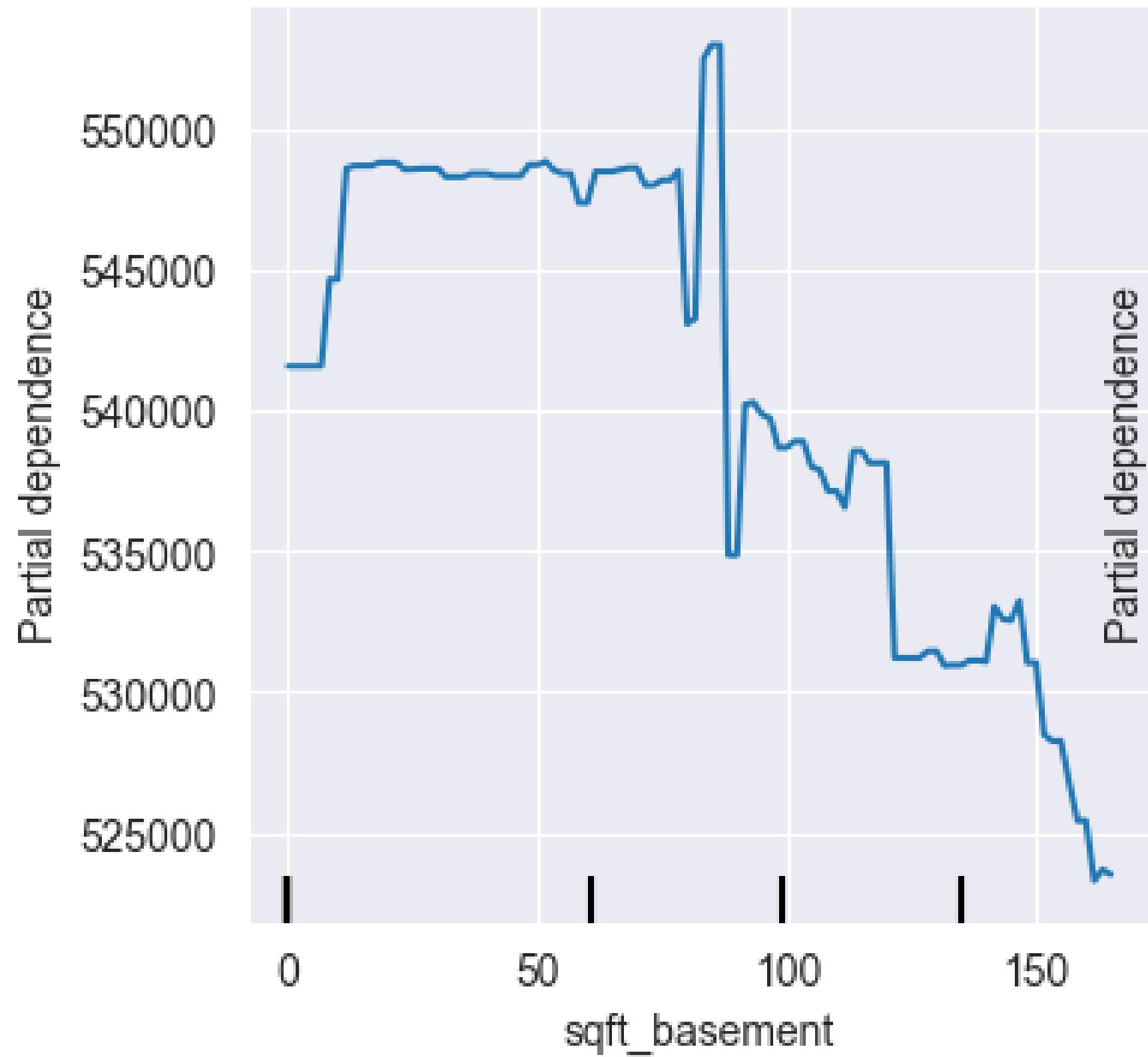
# PARTIAL DEPENDENCE PLOTS

These are plots which offer more insights on how feature values impact on the prediction scores. They help to answer investor and stakeholders questions on how increasing a feature may impact on the prediction scores. The plots come in handy and are the best approach to investigating the impact of increasing or decreasing any of the feature values on predictions.

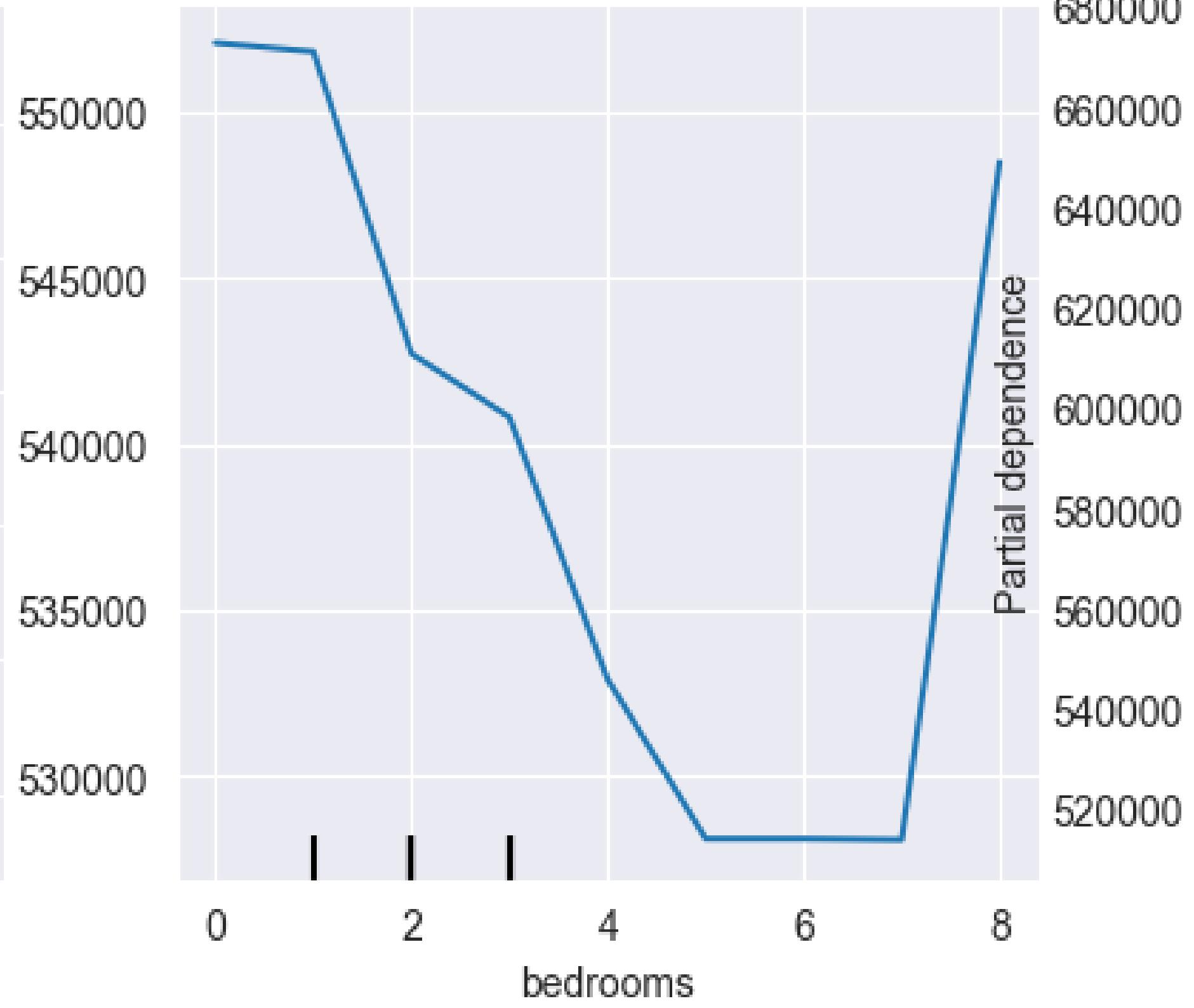
# Partial Depedence plot for ['sqft\_living']



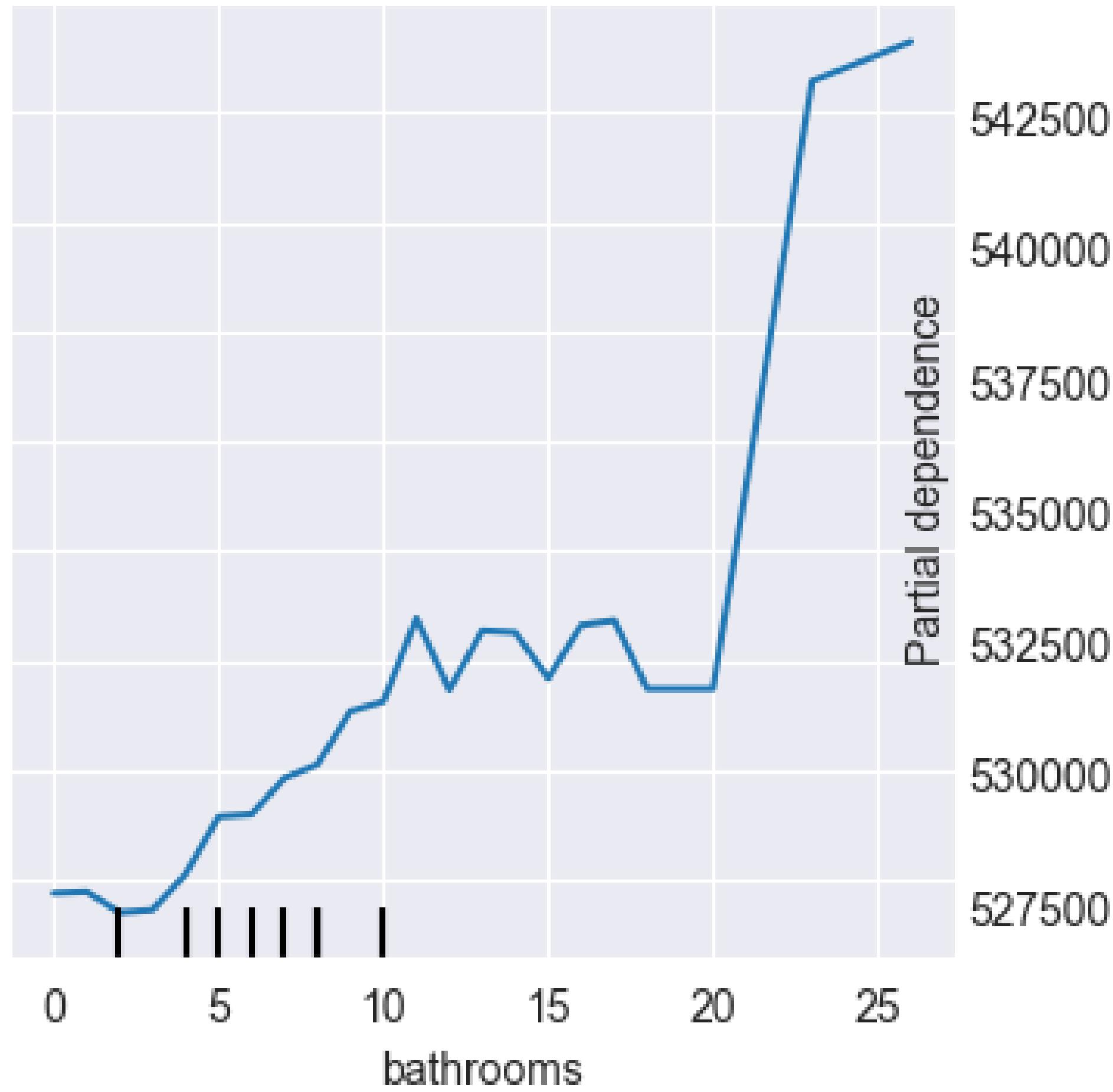
Partial Depedence plot for sqft\_basement



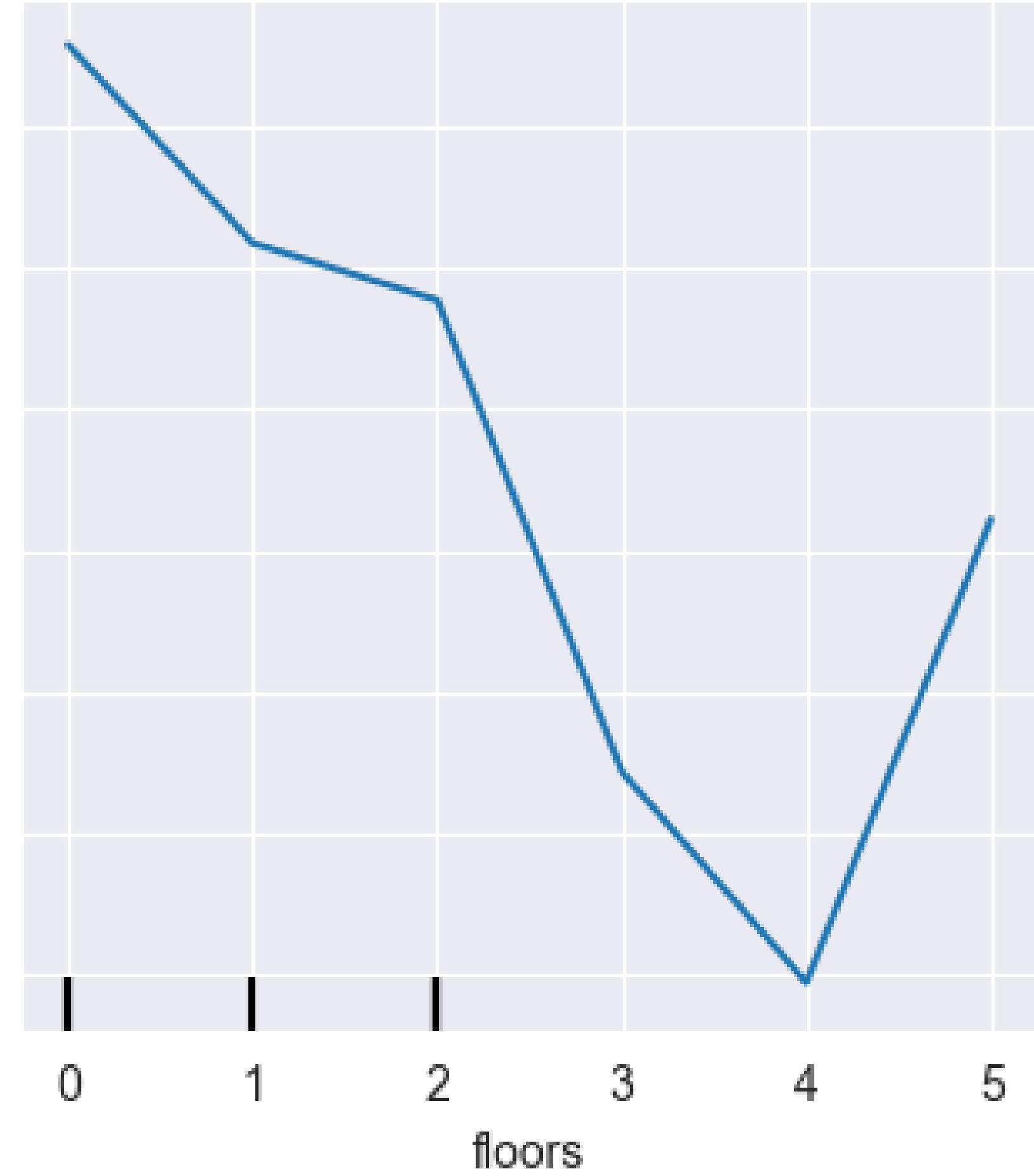
Partial Depedence plot for bedrooms



Partial Depedence plot for bathrooms



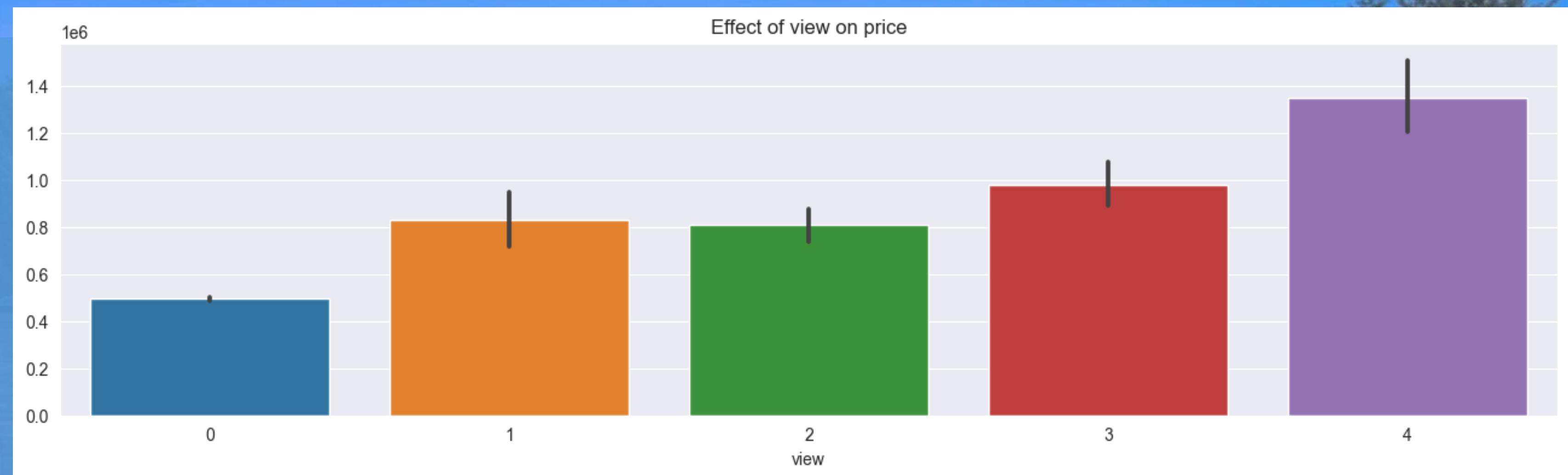
Partial Depedence plot for floors

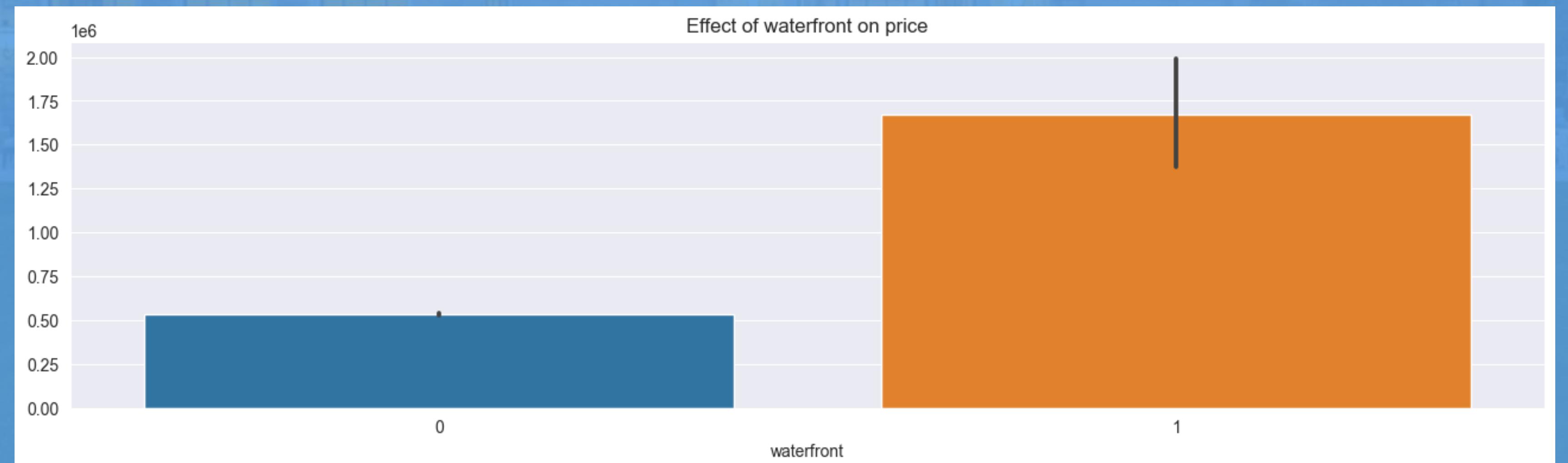
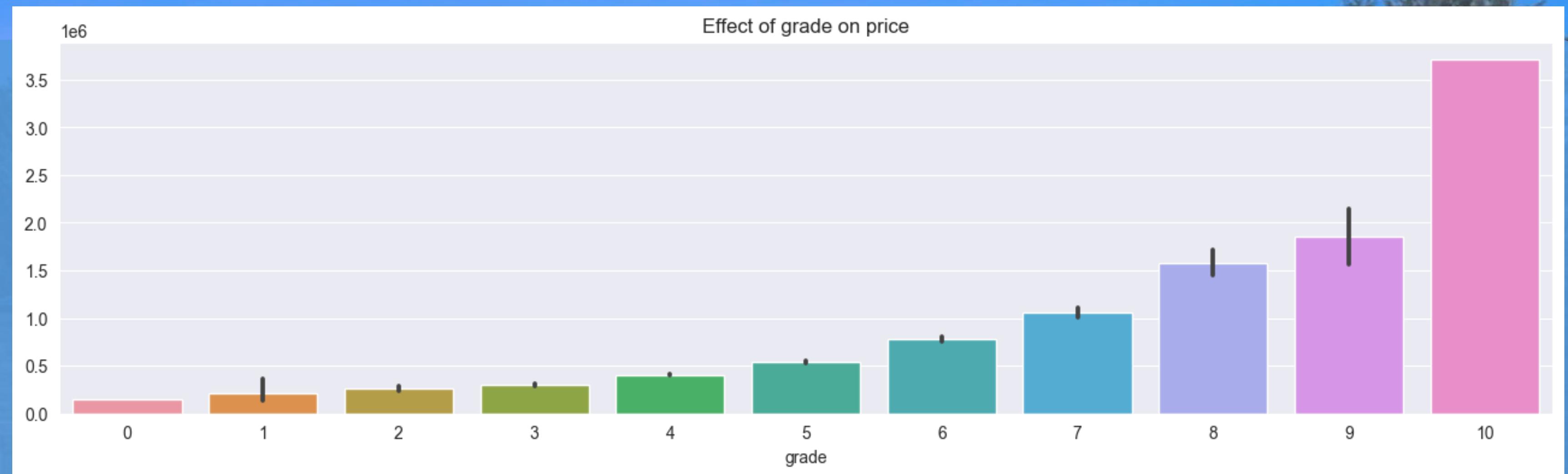


# PARTIAL DEPENDENCE PLOTS (CATEGORIC VARIABLES)

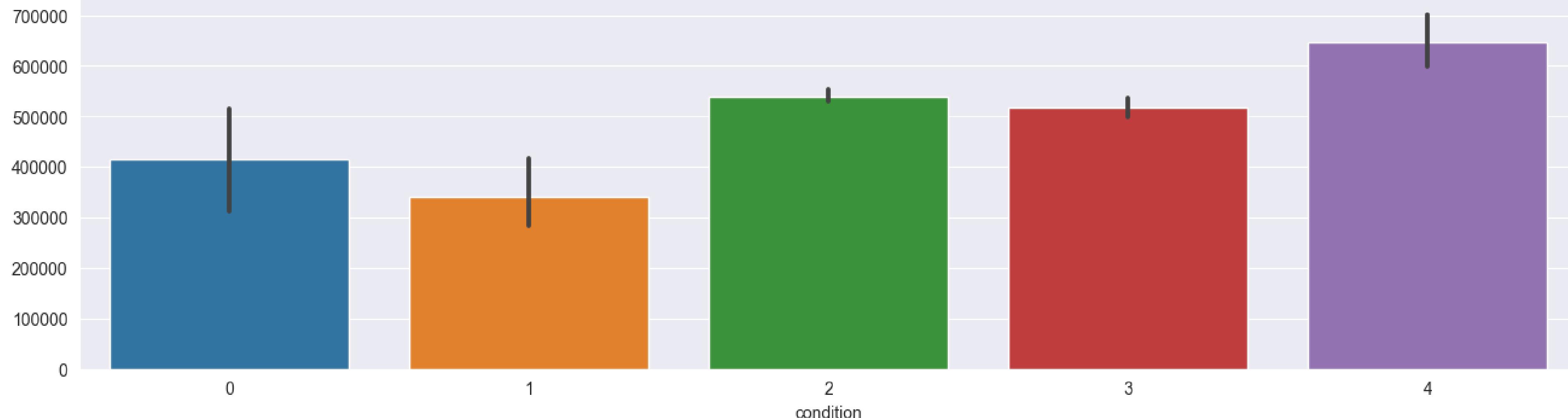
Visuals for categorical variables are best viewed using the bar graphs which give a clear distinction of how the specific variables affect price

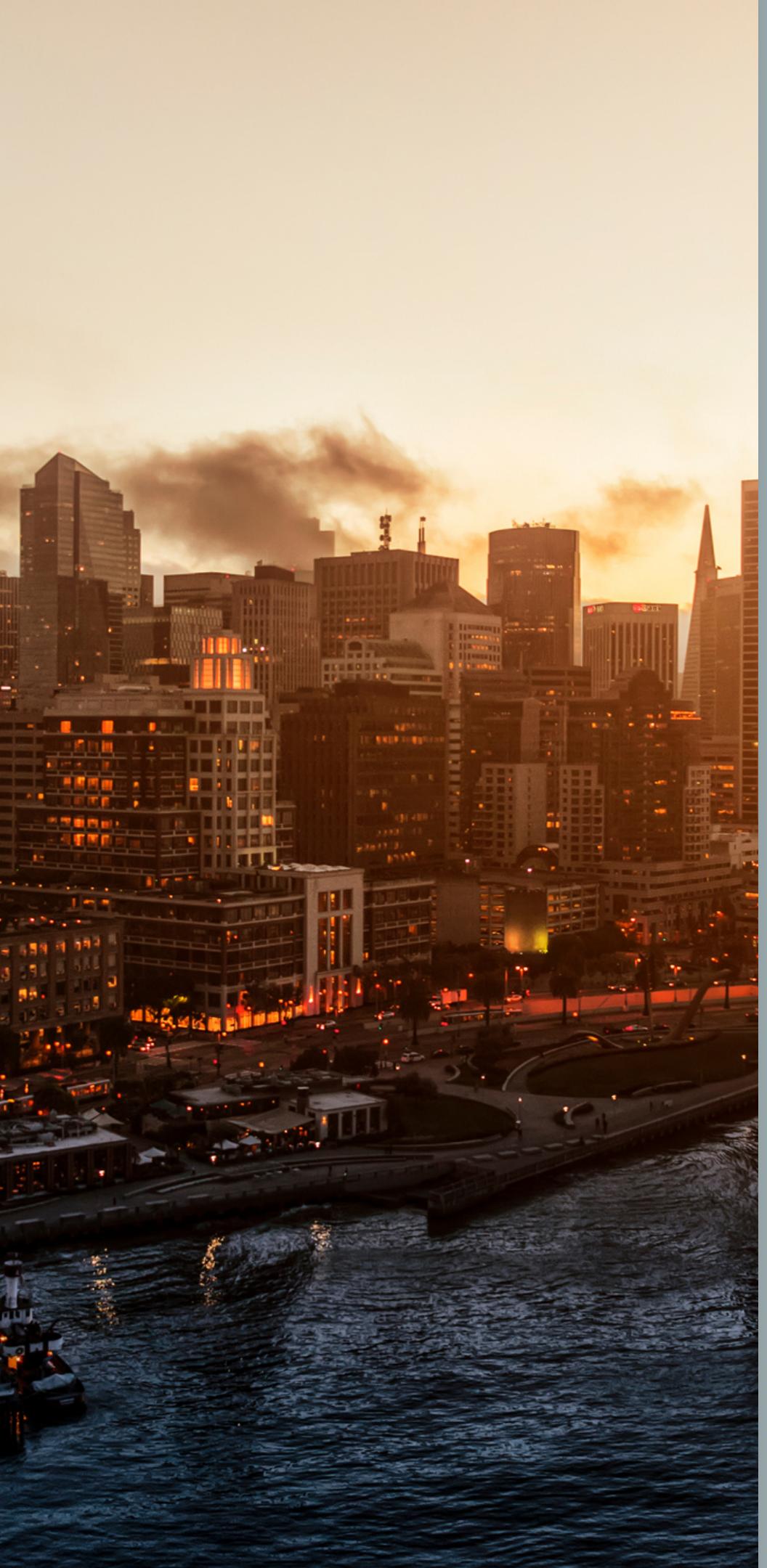




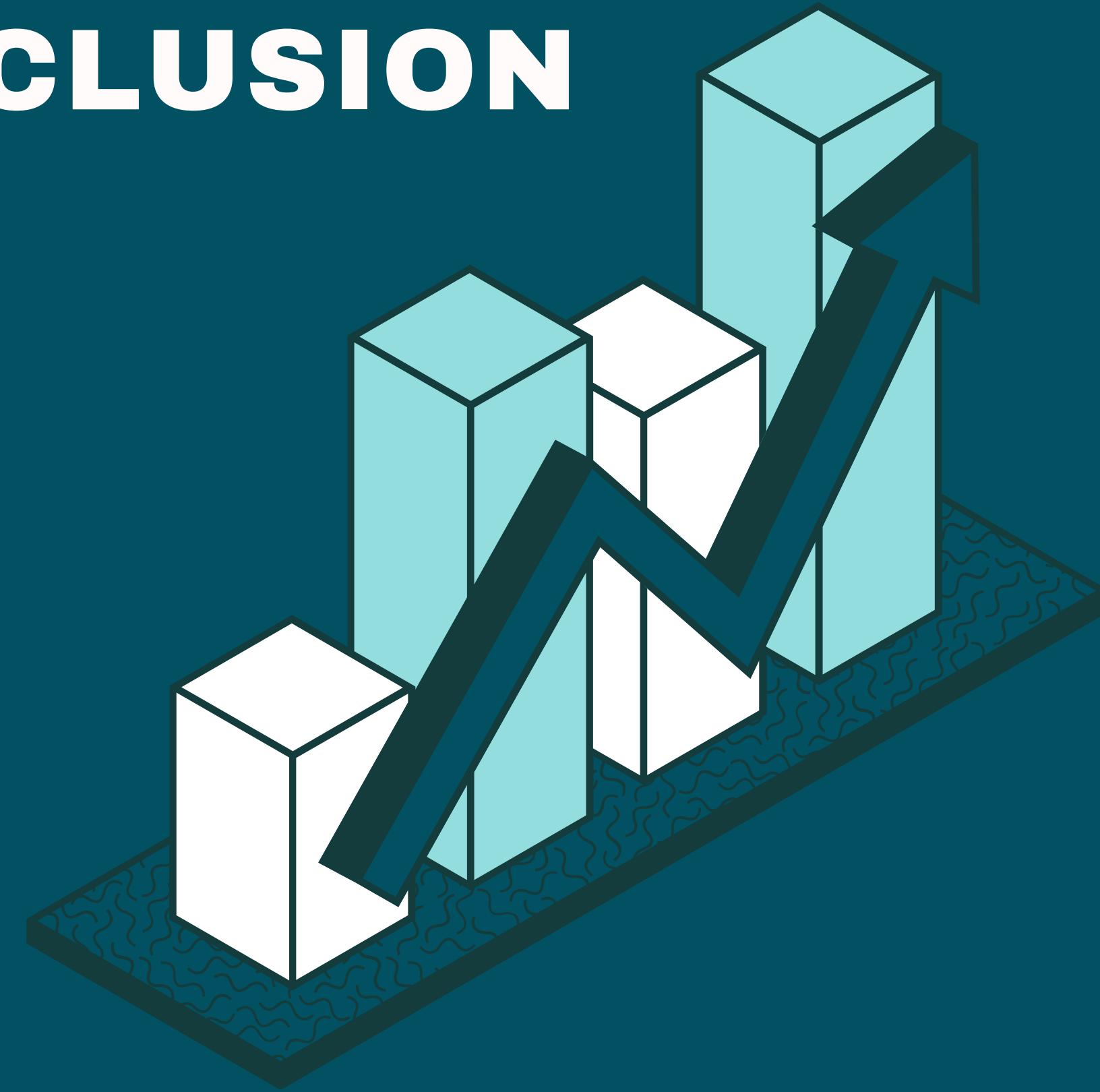


Effect of condition on price



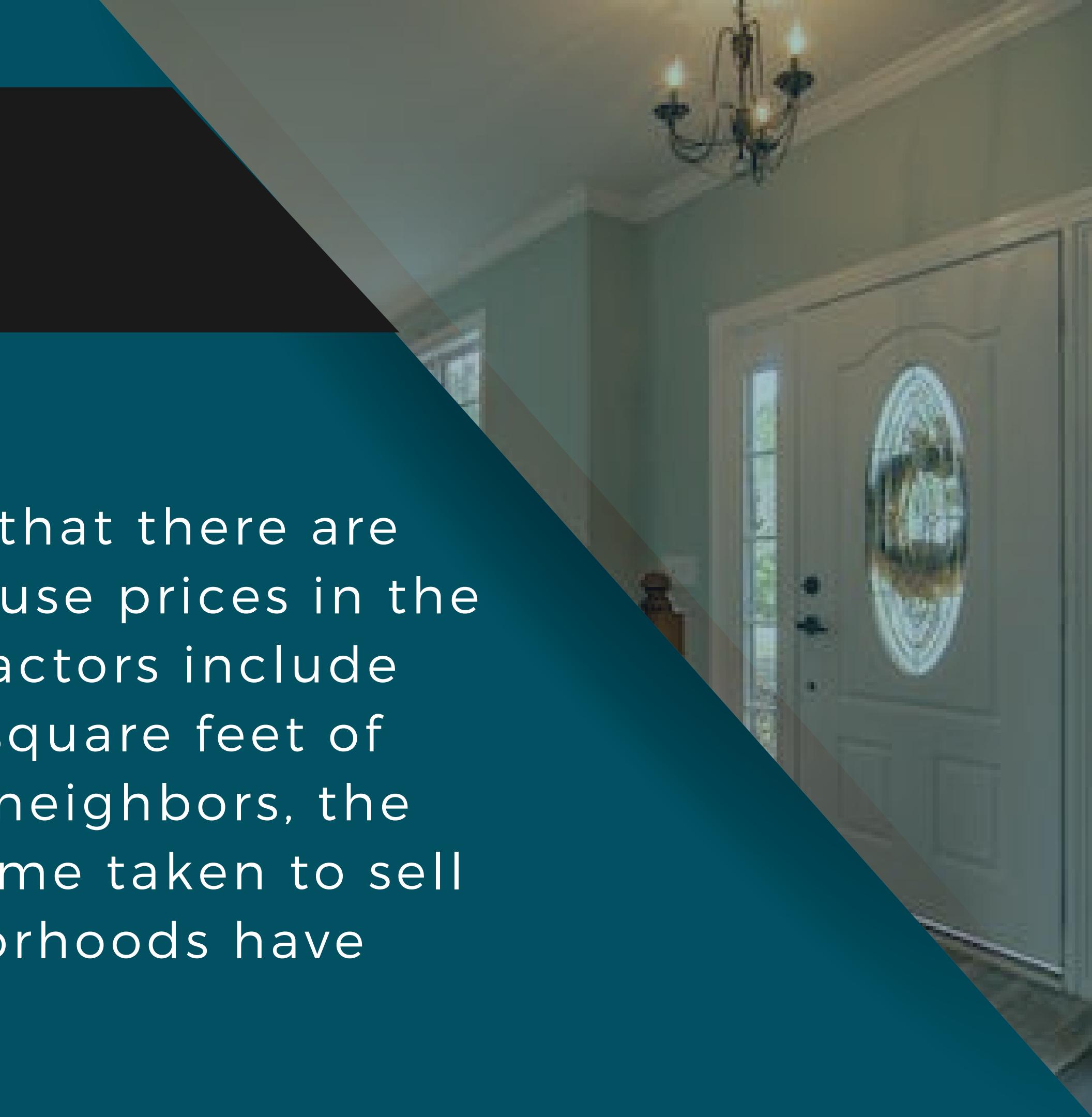


# CONCLUSION



## 1.CONCLUSION

We came to the conclusion that there are four main factors driving house prices in the King County region. These factors include the location of houses, the square feet of the living room and that of neighbors, the house view condition, the time taken to sell the house, and that neighborhoods have different prices.



## **2.CONCLUSION**

Improving the house grading right from construction with better and unique designs and investing in high-end neighborhoods seem to be the go-to renovations to realize a good return on investment



### **3.CONCLUSION**

We also came to the conclusion that the neighborhoods in the southern direction of the King County region have the most low-priced houses and homes.

# RECOMENDATIONS

- 01** Improving house grades right from construction with better designs
  
- 02** investing in high-end neighborhoods to realize a good return on investment.

## FUTURE PROPOSAL

We would propose to be provided with more data from other real estate agencies in order to provide a more accurate model and polish our analysis.

# GROUP MEMBERS

01

Chris Kamau

02

Kamande Karigi

03

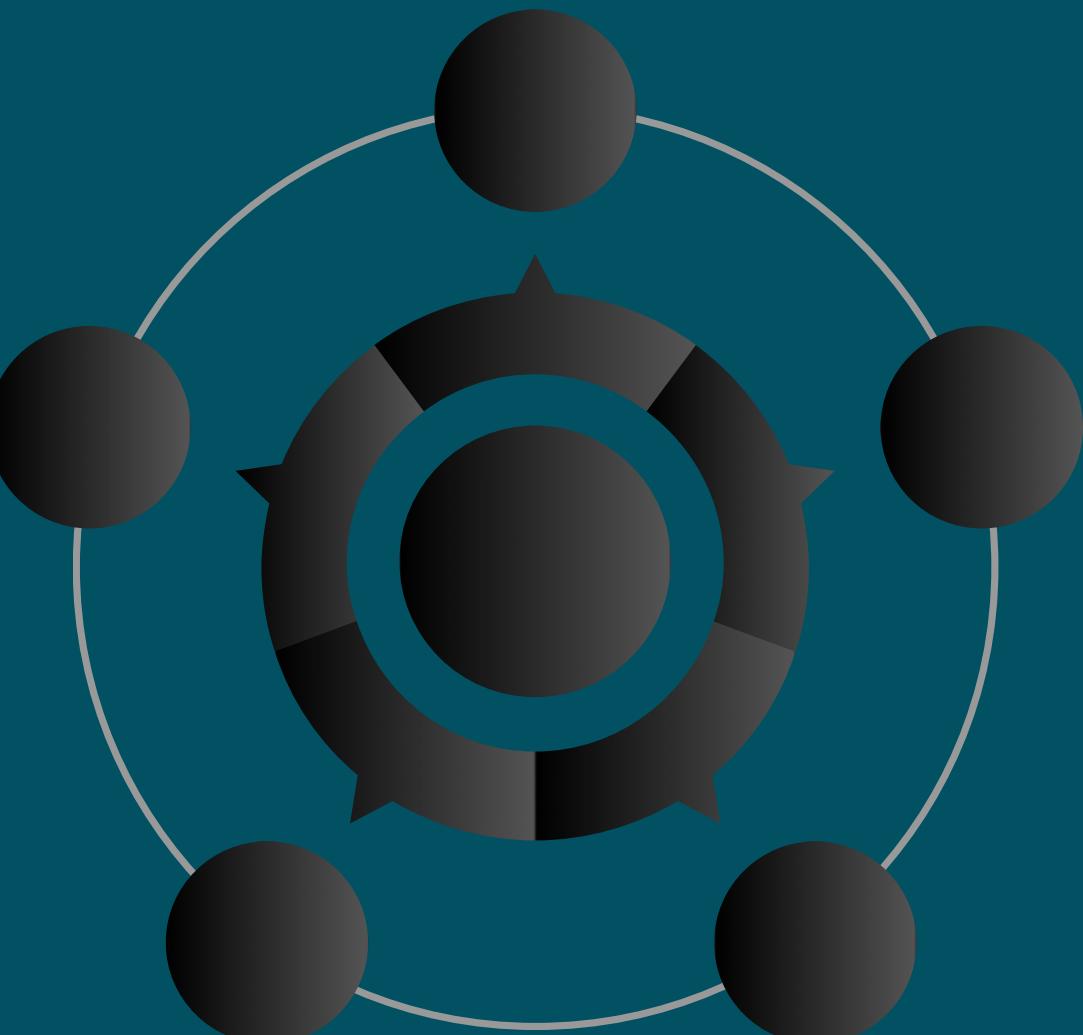
Cynthia Wanyeki

04

Isaack Odera

05

Gichogu Macharia



# KING COUNTY HOUSE



# THANK YOU

