KING COUNTY HOUSE PROJECT

CONTENT

01 Business Understanding

02 Problem Statement

03 Objectives

04 Data Analysis

05 Modeling

06 Conclusion

07 Participants

BUSINESS UNDERSTANDING

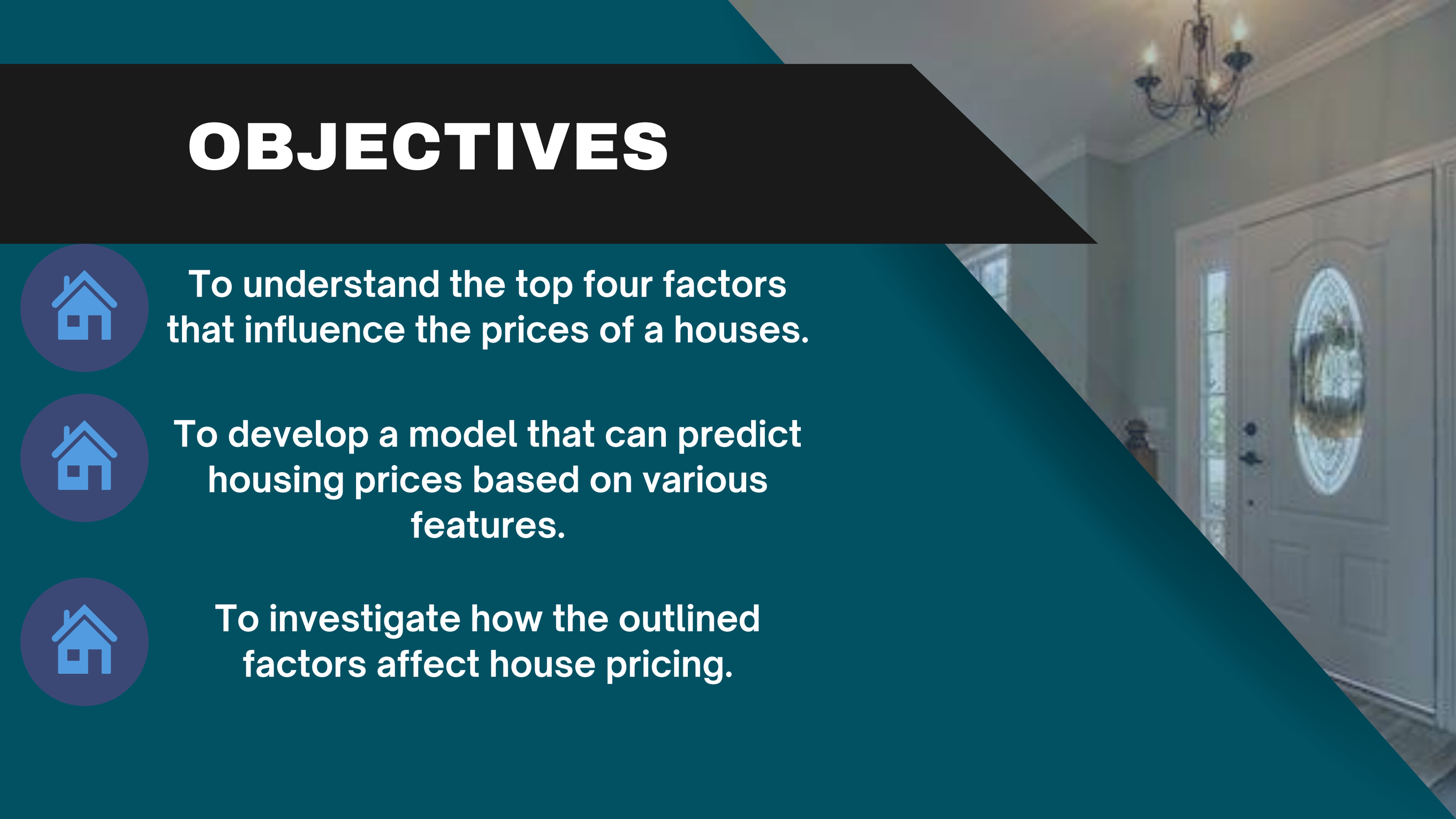
- The real estate agency serves homeowners looking to buy and sell homes.
- For homeowners looking to sell, understanding how specific renovations and expansions might impact the value of their homes is crucial.
- It is therefore necessary to extract insights that influence homes pricing.

PROBLEM STATEMENT

- Home owners need to understand how the value of their houses may be impacted by renovations.

- They also need to understand the bare-minimum housing conditions for various price ranges.

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 outlined factors affect house pricing.

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 the categorical and continuous data with price as the target.

- We focused our analysis on features that seemed to influence the sale pricing.

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 heat map came in handy in viewing interactions between the numerical features.

- We then sought to look at the interactions of house prices and the other features.

Correlation Heatmap

	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	view	condition	grade	sqft_above	sqft_basement	yr_built	sqft_living15	sqft_lot15	age	neighborhood
price	1	0.32	0.53	0.7	0.088	0.26	0.26	0.4	0.035	0.67	0.61	0.32	0.055	0.59	0.084	-0.055	0.031
bedrooms	0.32	1	0.53	0.59	0.034	0.18	-0.0015	0.083	0.023	0.37	0.49	0.31	0.16	0.4	0.032	-0.16	0.039
bathrooms	0.53	0.53	1	0.75	0.089	0.5	0.062	0.19	-0.13	0.67	0.69	0.28	0.51	0.57	0.089	-0.51	0.036
sqft_living	0.7	0.59	0.75	1	0.17	0.35	0.1	0.28	-0.062	0.76	0.88	0.43	0.32	0.76	0.18	-0.32	0.058
sqft_lot	0.088	0.034	0.089	0.17	1	-0.0075	0.021	0.073	-0.0093	0.11	0.18	0.016	0.053	0.14	0.72	-0.053	0.06
floors	0.26	0.18	0.5	0.35	-0.0075	1	0.02	0.028	-0.26	0.46	0.52	-0.25	0.49	0.28	-0.011	-0.49	-0.047
waterfront	0.26	-0.0015	0.062	0.1	0.021	0.02	1	0.38	0.016	0.082	0.07	0.083	-0.024	0.084	0.03	0.024	-0.014
view	0.4	0.083	0.19	0.28	0.073	0.028	0.38	1	0.045	0.25	0.17	0.28	-0.054	0.28	0.073	0.054	0.027
condition	0.035	0.023	-0.13	-0.062	-0.0093	-0.26	0.016	0.045	1	-0.15	-0.16	0.17	-0.36	-0.094	-0.0044	0.36	0.012
grade	0.67	0.37	0.67	0.76	0.11	0.46	0.082	0.25	-0.15	1	0.76	0.17	0.45	0.71	0.12	-0.45	0.031
sqft_above	0.61	0.49	0.69	0.88	0.18	0.52	0.07	0.17	-0.16	0.76	1	-0.054	0.43	0.73	0.2	-0.43	0.052
sqft_basement	0.32	0.31	0.28	0.43	0.016	-0.25	0.083	0.28	0.17	0.17	-0.054	1	-0.13	0.2	0.017	0.13	0.023
yr_built	0.055	0.16	0.51	0.32	0.053	0.49	-0.024	-0.054	-0.36	0.45	0.43	-0.13	1	0.33	0.071	-1	0.036
sqft_living15	0.59	0.4	0.57	0.76	0.14	0.28	0.084	0.28	-0.094	0.71	0.73	0.2	0.33	1	0.18	-0.33	0.076
sqft_lot15	0.084	0.032	0.089	0.18	0.72	-0.011	0.03	0.073	-0.0044	0.12	0.2	0.017	0.071	0.18	1	-0.071	0.059
age	-0.055	-0.16	-0.51	-0.32	-0.053	-0.49	0.024	0.054	0.36	-0.45	-0.43	0.13	-1	-0.33	-0.071	1	-0.036
neighborhood	0.031	0.039	0.036	0.058	0.06	-0.047	-0.014	0.027	0.012	0.031	0.052	0.023	0.036	0.076	0.059	-0.036	1

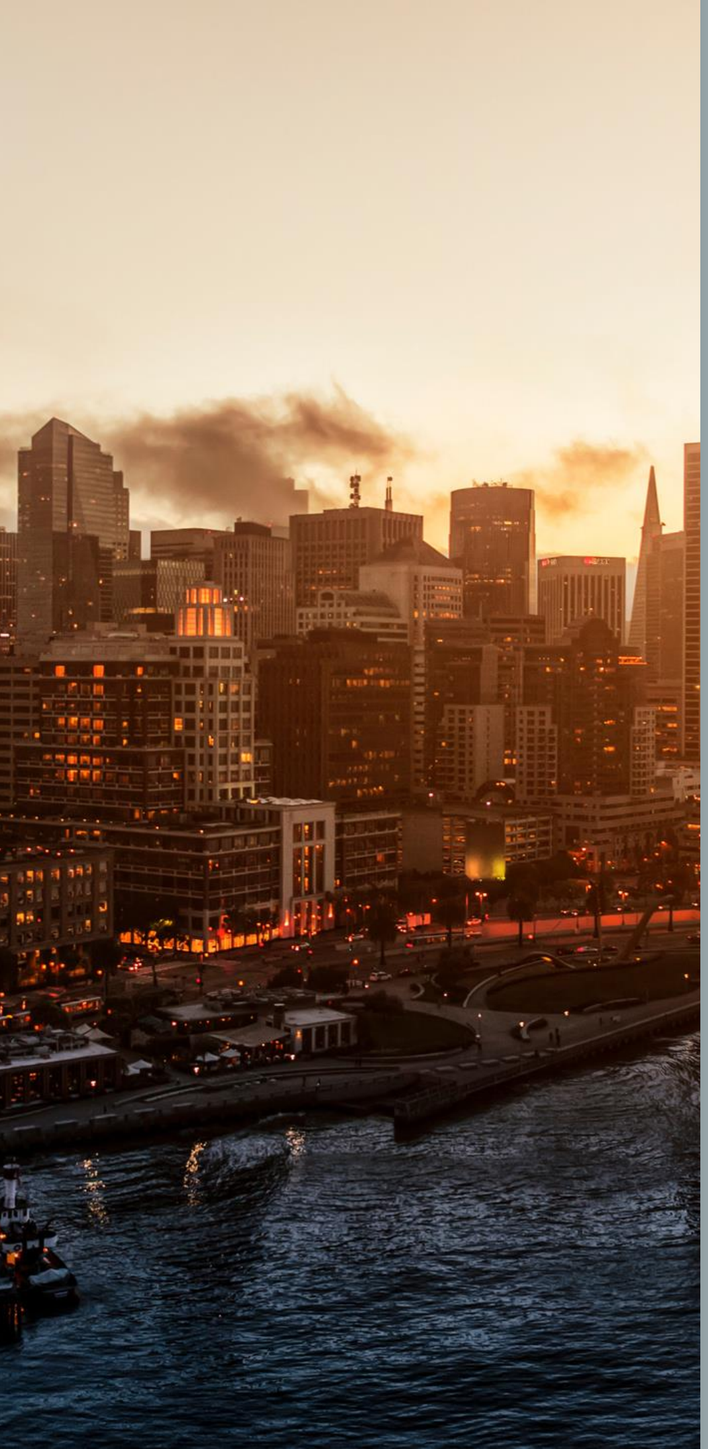
price bedrooms bathrooms sqft_living sqft_lot floors waterfront view condition grade sqft_above sqft_basement yr_built sqft_living15 sqft_lot15 age neighborhood

1.00 0.75 0.50 0.25 0.00 −0.25 −0.50 −0.75

MODELING AND REGRESSION ANALYSIS

We developed four regression models with the housing data and finally settled on a gradient boosting model.

The model had an accuracy score of 86% and an RMSE of 0.0172.

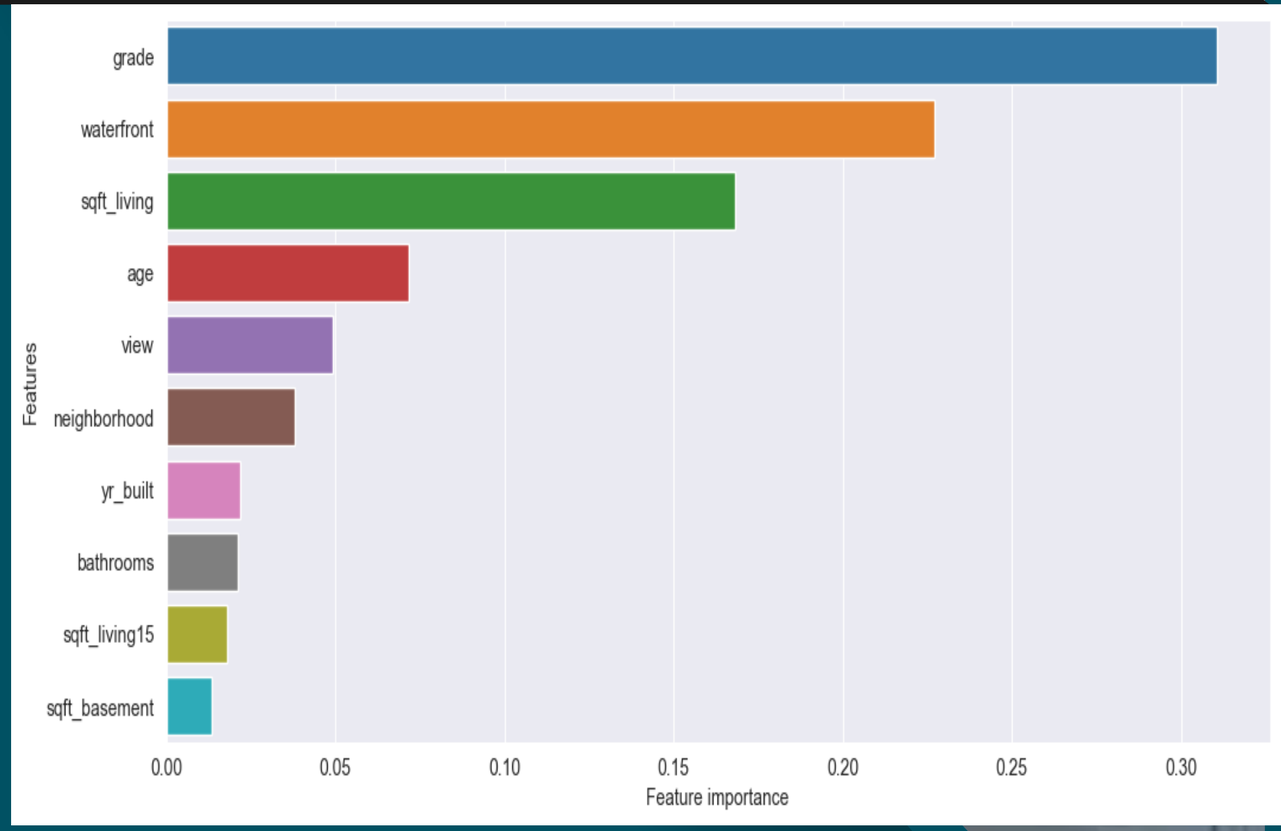


CONCLUSION

Factors influencing house pricing

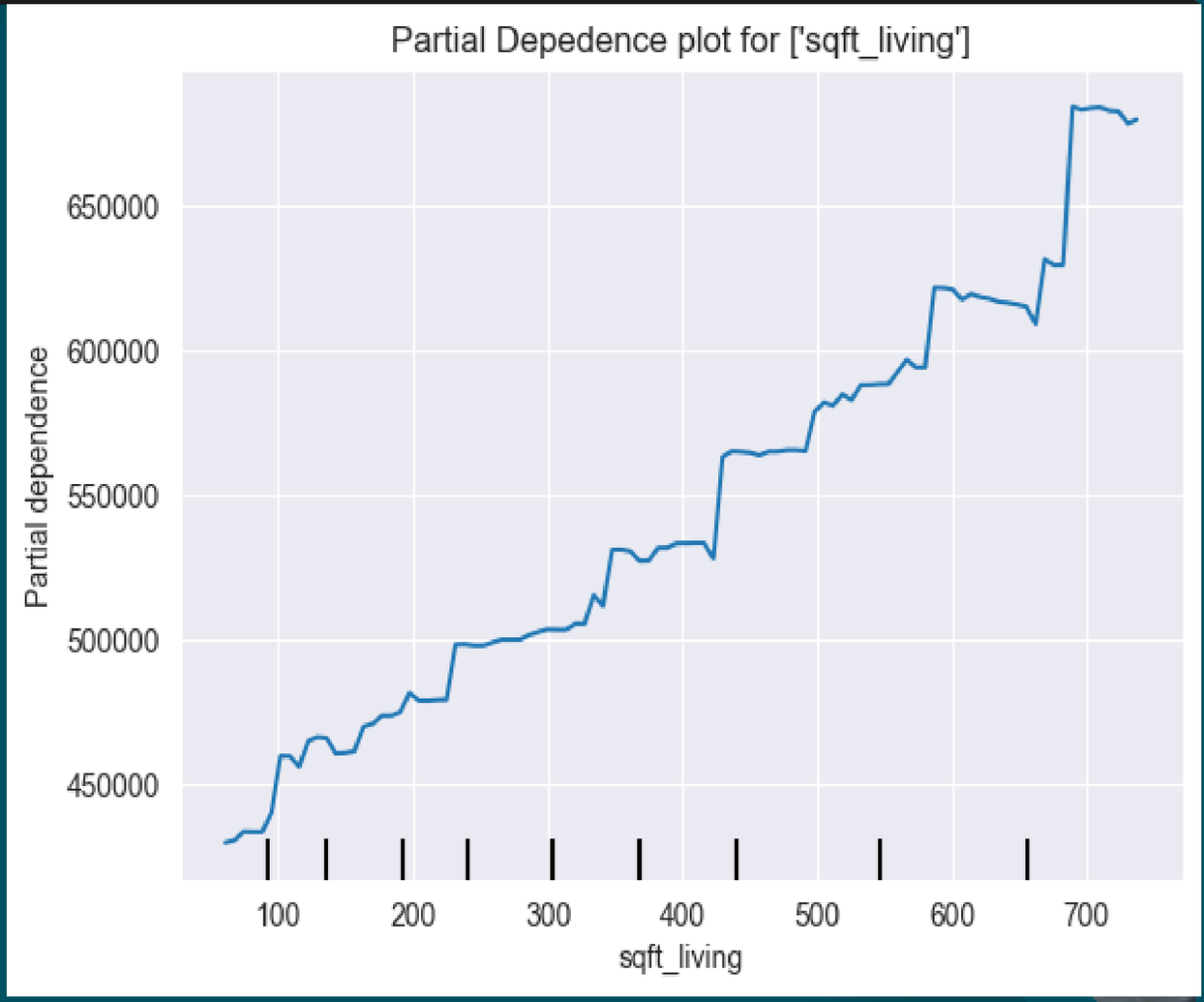
The following top four factors determine house pricing.

- The house grading.
- The presence or absence of waterfronts
- The living space.
- The age of the houses.

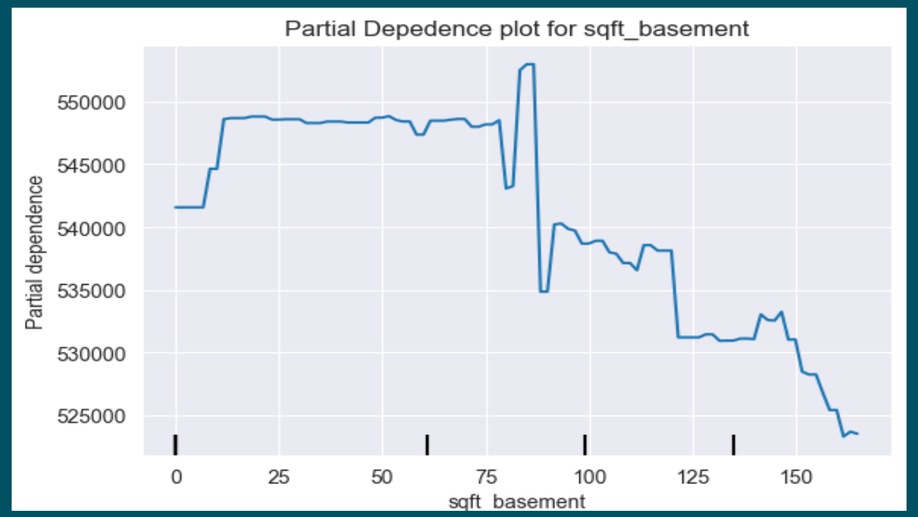


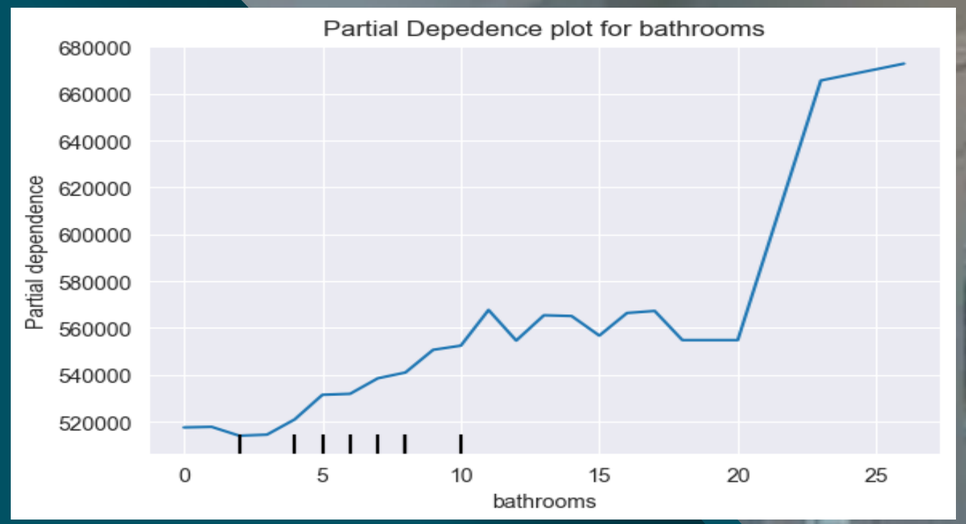
Factors influencing house pricing(Living Space)

- Any increase of the living space has a direct impact on the prices of homes.
- A living space between 100 and 695sqft guarantees a good investment.



Factors influencing house pricing(Basement & bathroom)

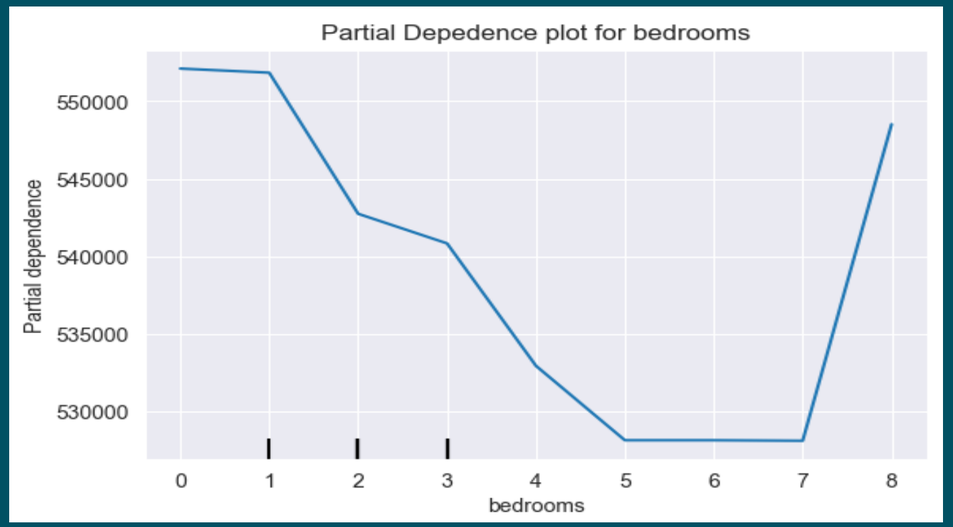


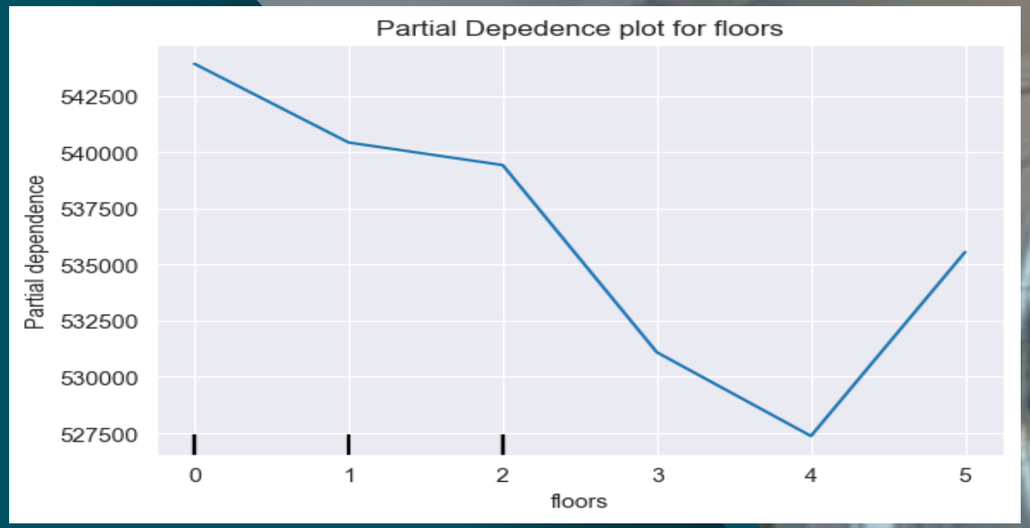


- A basement of 75-80 square foot has high prices.

- The more the bathrooms, the higher the price.

Factors influencing house pricing(Bedrooms & Floors)

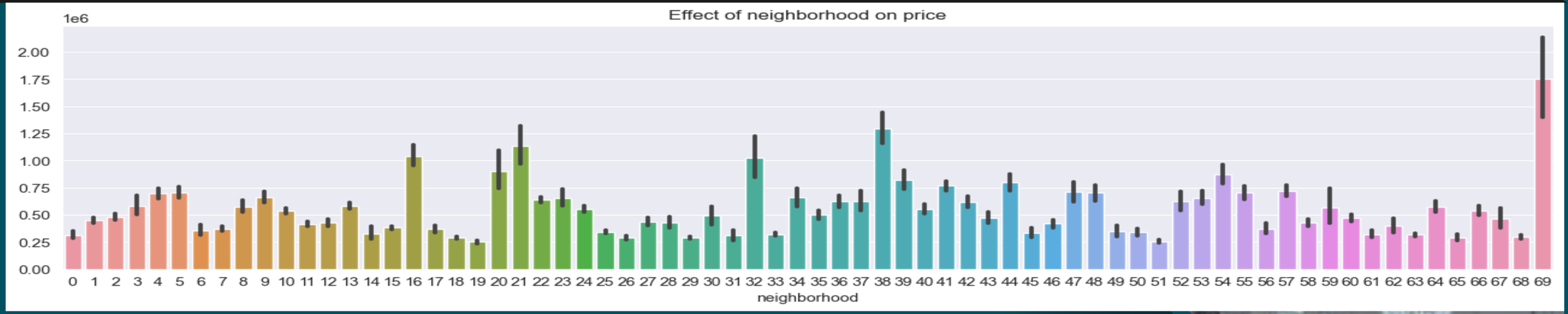




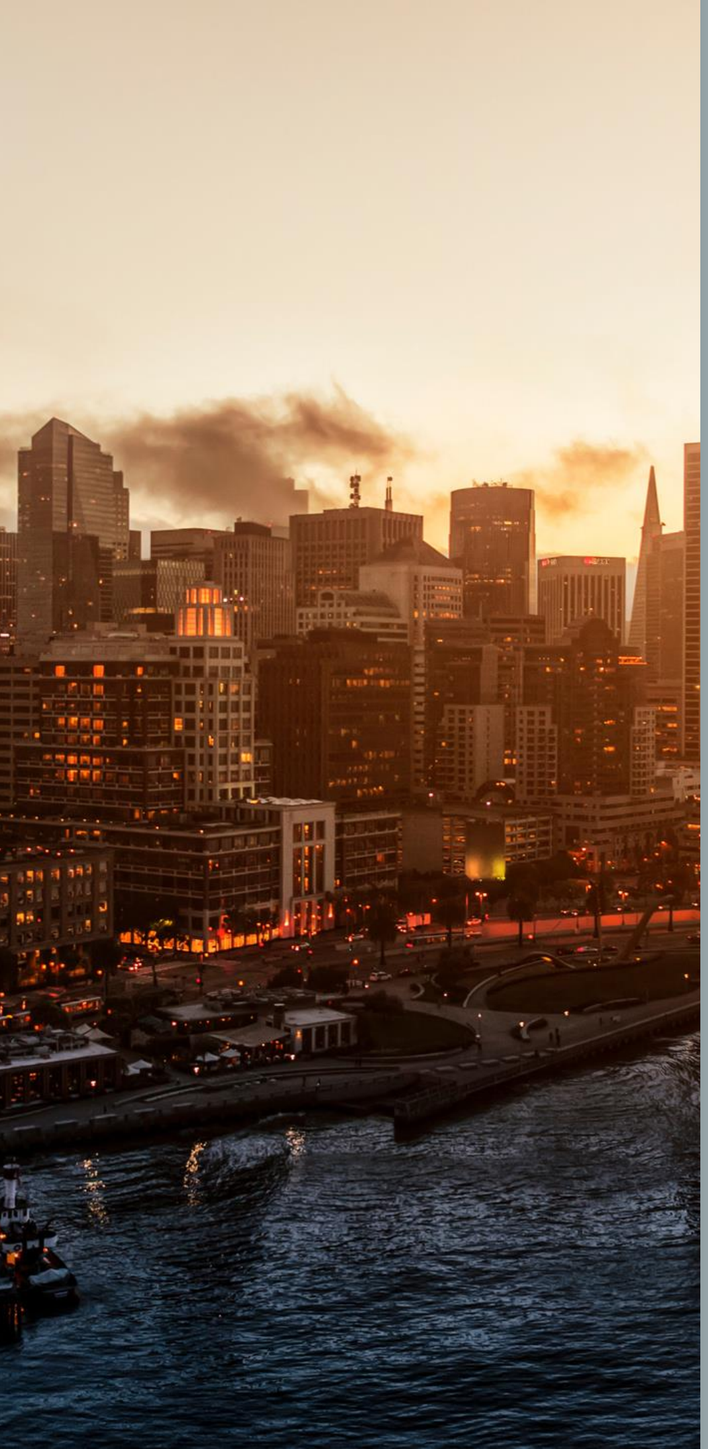
- The more the bedrooms, the higher the house pricing.

- More than four floors in houses attracts higher prices.

Factors influencing house pricing(Neighborhood)



- Investing in neighborhoods 69 and 38 guarantees a good investment.
- Investing in neighborhoods 19, 51 and 65 does not guarantee a good ROI.



RECOMMENDATIONS

RECOMMENDATIONS.1

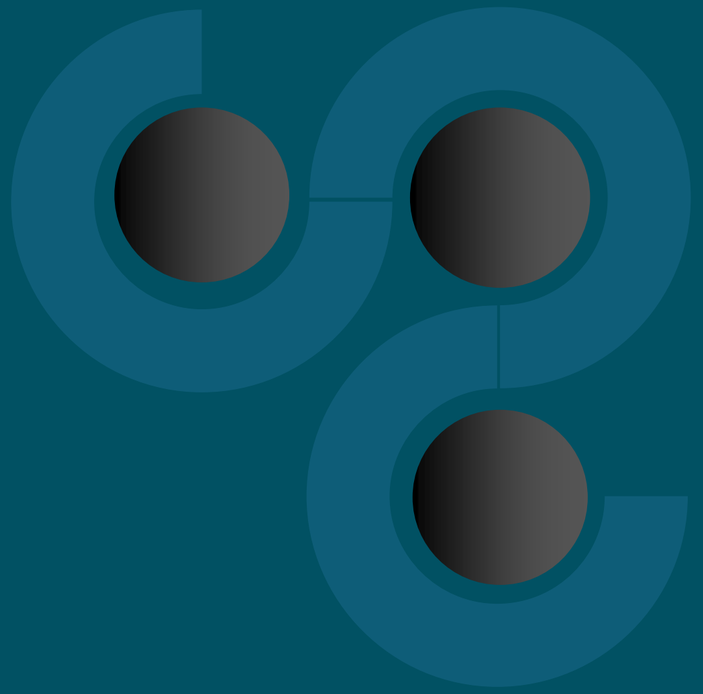
01 King County agencies should improve house grades and conditions to realize good ROI.

02 King County Real Estate investors should consider investing in high-end neighborhoods.

RECOMMENDATIONS.2

03 King County developers should construct houses with more than 5 bathrooms.

04 King County developers should sketch plans of houses with more than four floors.



RENOVATIONS & EXPANSION ECOMMENDATIONS.3(Home sellers)

- Increase the number of floors from 4 to 5 increase value from 527,500 to 535,000.
- Add more bathrooms to be more than 20 to increase value from 556,000 to 660,000.
- Increase number of bedrooms from 7 to 8 to have the house value increase from 528,000 to 550,000.
- Increase the size of living space from 680 to 690sqft to increase value from 625,000 to 680,000.

FUTURE PROPOSAL

In the future, more information about the following should be provided.

- Location of social amenities.
- Number of rooms for every house.
- Years of renovations.

GROUP MEMBERS

01

Chris Kamau

02

Kamande Karigi

03

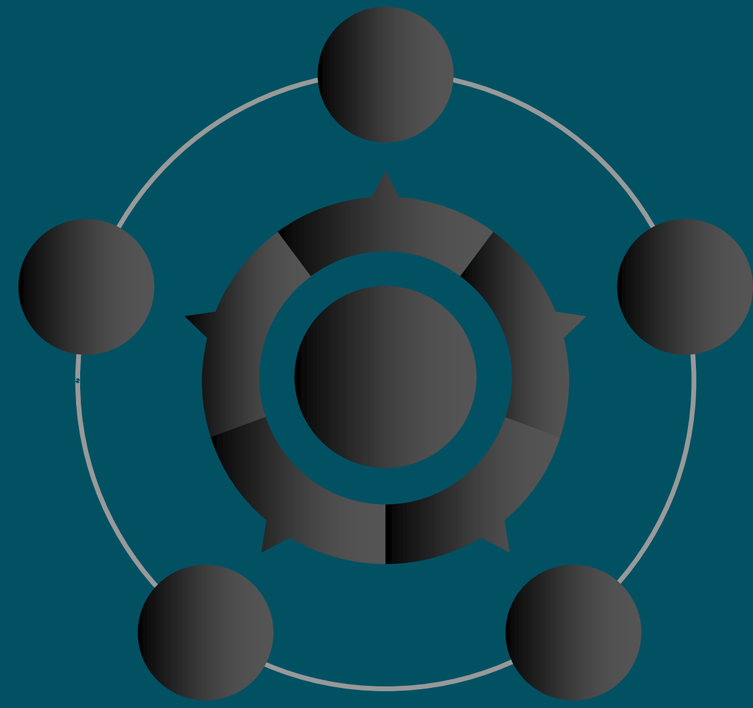
Cynthia Wanyeki

04

Isaack Odera

05

Gichogu Macharia



KING COUNTY HOUSE

THANK YOU