



# King County Houses Sales Data



# Agenda

What this report covers

---

**01** Overview

**02** Business and Data Understanding

**03** Modelling

**04** Regression Results

**05** Recommendations

**06** Thank you

# Overview



## King County

King county is a county in Washington state with the highest population of 2,269,675 according to the 2020 census.

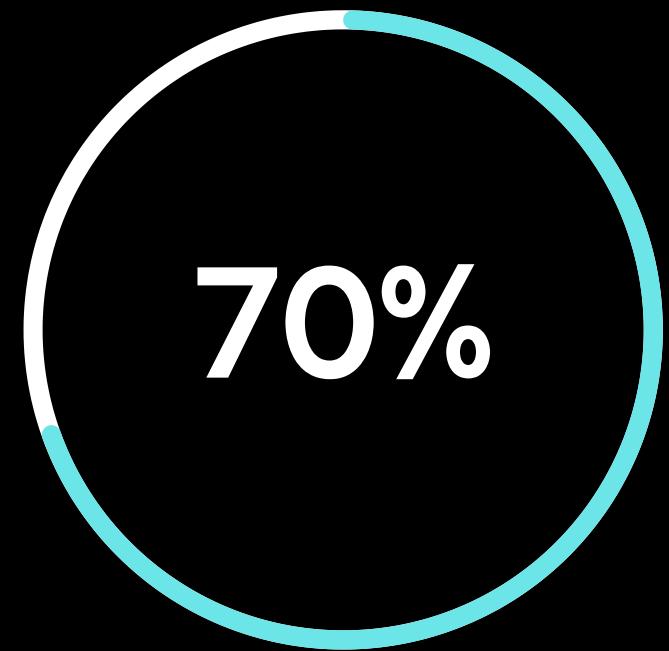
## Role as Data Scientist

The role as a data scientist is to give a report on the King County Real estate Industry and come up with price predicting models.

## Business and Data Understanding

---

In our analysis we used 70%  
of the data collected.



---

Data Used: Date, Price, Sqft\_lot, Sqft\_living, Sqft\_above,  
Bedrooms, Bathrooms, Condition, Year Built, Waterfront, View



# Data Understanding

## Date

Date a house was sold

The data house prices was collected in 2014 and 2015.

## Price

Sales Price

The average selling price of a house is \$ 540, 297.

## Sqft\_Living/Lot/Above

Square Footage

Square footage of house part



## Bedrooms

Number of Bedrooms

The number of bedrooms range from 1 to 6 bedrooms.

## Bathrooms

Number of Bathrooms

The number of bathrooms range from 1 to 5 bathrooms

## Condition

Overall Condition of the House

This includes Average, Very Good, Good, Poor and Fair

# Year

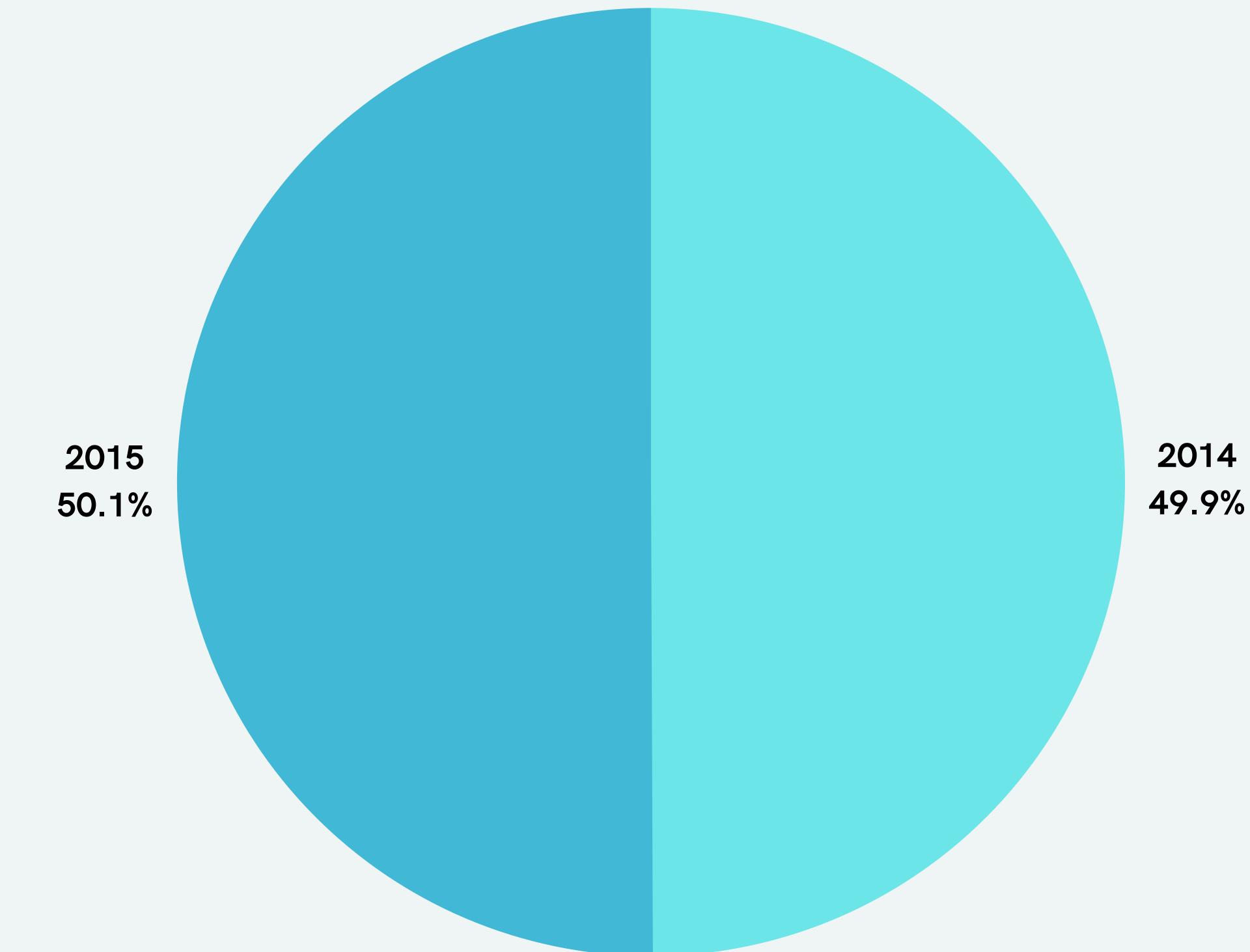
Per Year

---

## Prices Per Year

The ratio of the data collected was 67% from 2014 and 33% from 2015.

However, the prices in 2015 were higher than the prices in 2014.



# Exploratory Data Analysis

## Price

**\$1,640,000**

**Maximum Price**

From our analysis this was the maximum price of the houses in King County.

**\$78,000**

**Minimum Price**

From our analysis this was the minimum price of the houses in King County.

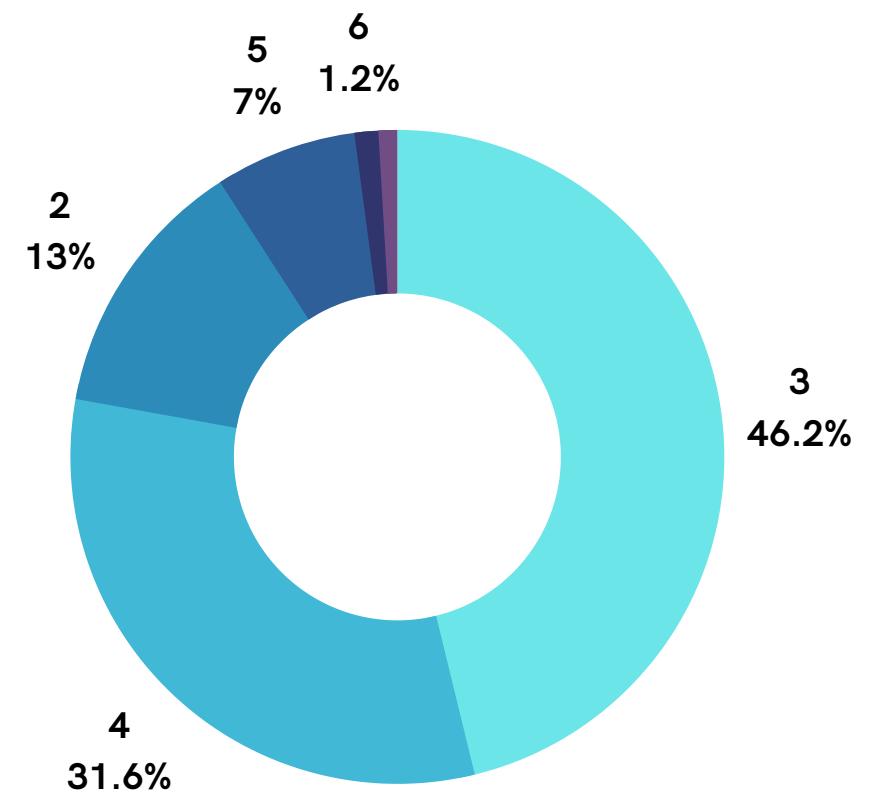
**\$505,219**

**Average Price**

From our analysis this was the average price of the houses in King County.



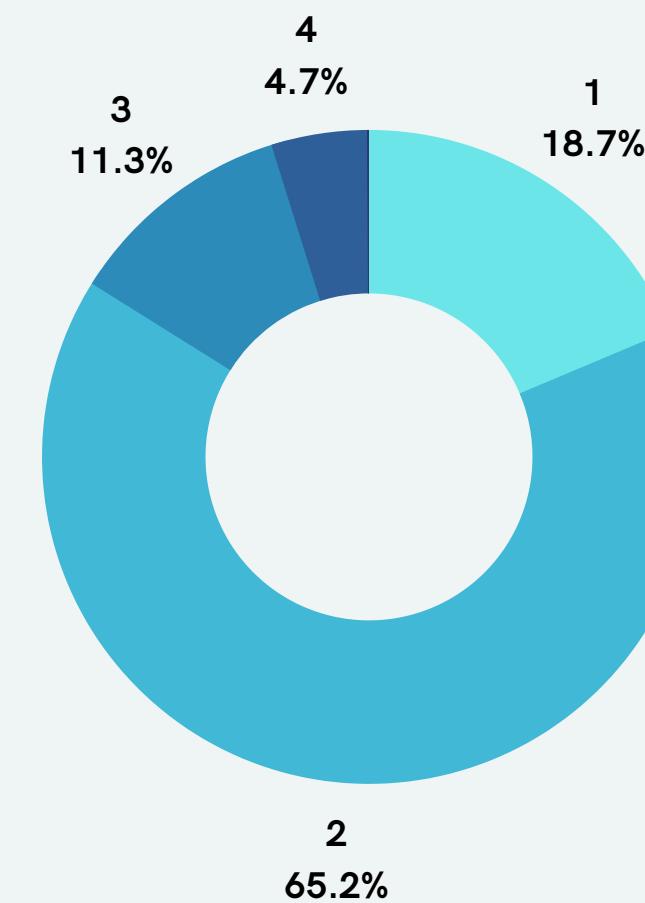
# Bedrooms and Bathrooms



6

Bedrooms

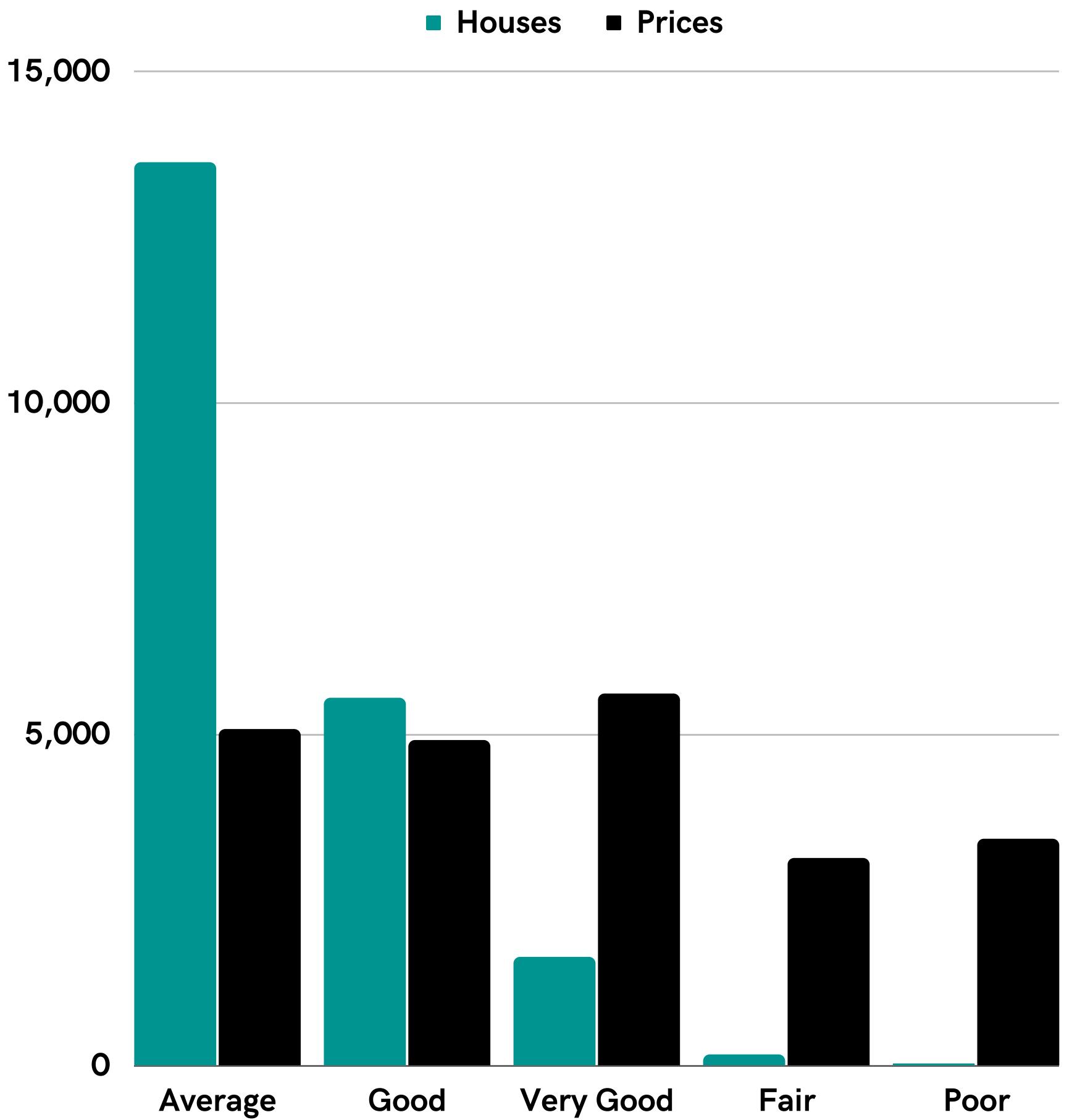
King County Houses



5

Bathrooms

King County Houses



# Condition

Per House

## King County Houses Condition

Most houses have an average condition however the ones rated as very good have the highest prices.

# Square Footage

---

**20,201sqft**

**Living Room Space**

Square footage of living space in  
the home

**12,448 sqft**

**Lot**

Square footage of the lot

**1,742 sqft**

**Above**

Square footage of house apart  
from basement

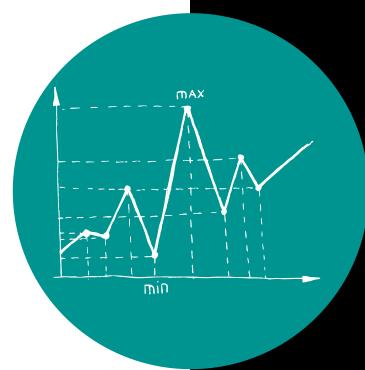
**The average price per sqft is \$ 200 to \$300**

# Modeling



## Simple Linear Regression

We created a model to predict price using Sqft\_living



## Simple Linear Regression

We created a model to predict price using Sqft\_above



## Multiple Linear Regression

We created a model to predict price using bedrooms + bathrooms + sqft\_living + sqft\_lot + waterfront

# Regression Results

---

## Price vs Sqft\_Living Regression Results

The r-squared value, 0.438, shows that the model can account for about 43.8% of the variability of price around its mean.

The null hypothesis for the simple linear regression is that there is no relationship between the chosen variable and the response variable.

All of the p-values round to 0, which means we can reject the null hypothesis.

## Price vs Sqft\_Above Regression Results

The r-squared value, 0.341, shows that the model can account for about 34.1% of the variability of price around its mean.

The null hypothesis for the simple linear regression is that there is no relationship between the chosen variable and the response variable.

All of the p-values round to 0, which means we can reject the null hypothesis.

## Multiple Linear Regression Regression Results

The r-squared value, 0.461, shows that the model can account for about 46% of the variability of price around its mean.

The null hypothesis for multiple regression is that there is no relationship between the chosen variables and the response variable.

All of the p-values round to 0, which means we can reject the null hypothesis.



King County

# Conclusion

## Regression Model Conclusion

The multiple linear regression model is better than the two simple regression models as it had a higher R-squared value. This shows that categorical variables had a positive effect on the prices as illustrated in the univariate analysis.

### Sqft\_living and Above

The real estate agency should review the sqft of the living space/ the ground above the basement when setting their house prices as there is a linear relationship between these variables.

### Bathrooms, Bedrooms, waterfront and view

These factors have a positive effect on the house prices.

### Inflation

The house prices seemed to have increased in 2015 they should therefore consider inflation when setting house prices in the future.

# Thank you!

Feel free to reach out to us if you have any questions.

---

**Phone Number**

+254 770 200 601

**Email Address**

julia.karanja@student.moringaschool.com

