# Predicting King County Housing Prices

PREPARED BY;

Jeremy Nguyo



## Overview

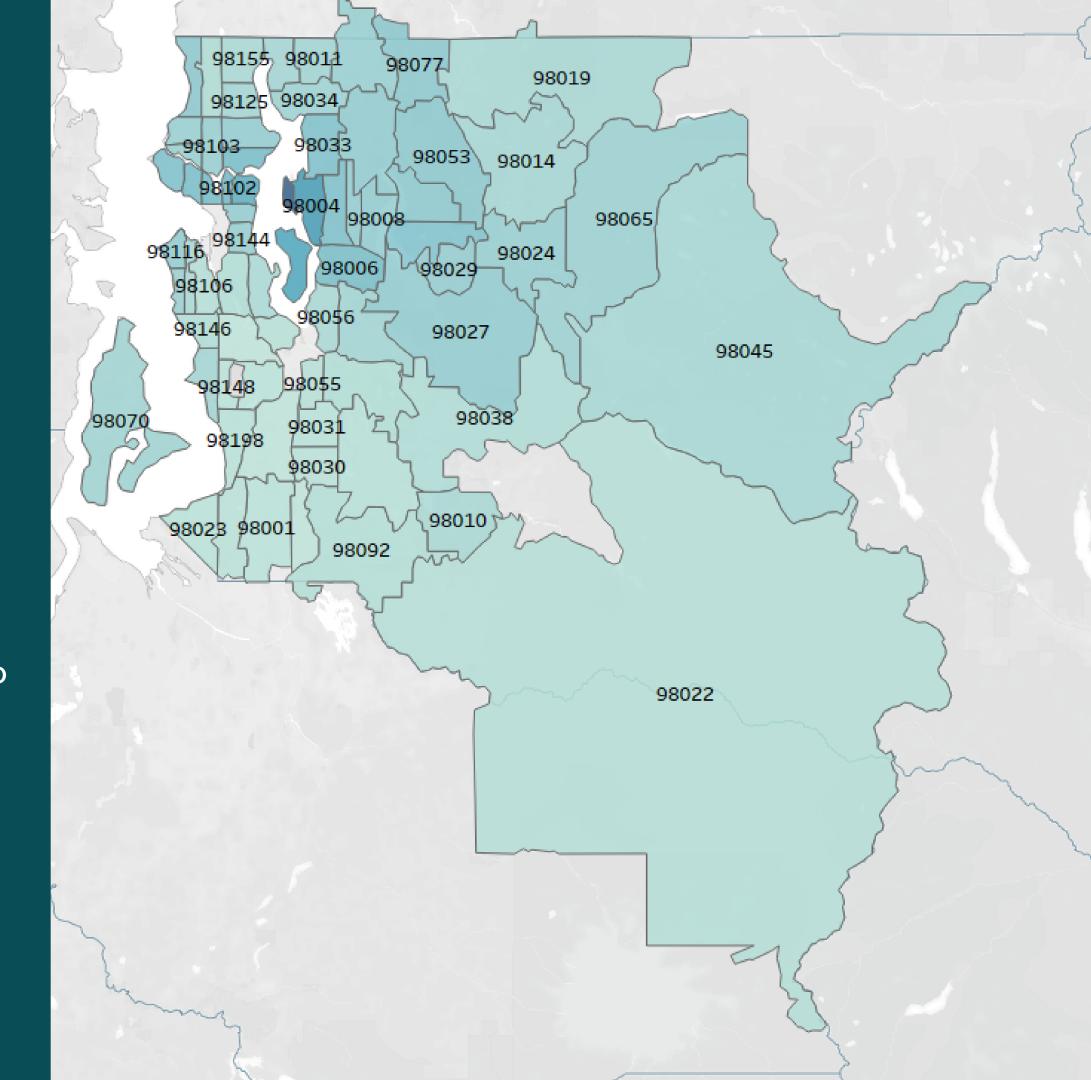
The King County Housing Data Set contains information about the size, location, condition, and other features of houses in Washington's King County.

#### The Data:

- 21,000 house sales
- 20 predictors

The goal is to explain some of the top features used to predict the sale price of homes and ultimately lead to higher sale price.

A Linear Regression Model was adopted for analysis.



## Linear Regression Analysis Results

The model accounts for 54.7% of the variance in price. A p-value of less than 0.05 means we can reject the null hypothesis that there is no relationship between price and the predictor variables.

The predictor variables are:

- a. Sqft\_Living
- b. View
- c. Sqft\_Living15

P < 0.05 $R^2 = .547$ 

PRICE ~ SQFT\_LIVING + VIEW + SQFT\_LIVING15

## Question 1:

Which features are most highly correlated with price?

The features are listed below in descending order:

- Square Footage of the Living Space
- Square Footage of the House apart from the Basement
- Square footage of Interior Housing Living Space for the Nearest 15 Neighbors
- Bathrooms

#### Variable Correlations

price -	1	0.31	0.53	0.7	0.09	0.26	0.26	0.61	0.32	0.12	0.053	0.31	0.022	0.59	0.083	0.0099	0.0037	0.054
bedrooms -	0.31	1	0.51	0.58	0.032	0.18	0.0021	0.48	0.3	0.018	0.15	0.01	0.13	0.39	0.031	0.001	0.0099	0.16
bathrooms -	0.53	0.51	1	0.76	0.088	0.5	0.064	0.69	0.28	0.047	0.2	0.024	0.22	0.57	0.088	0.0075	0.027	0.51
sqft_living -	0.7	0.58	0.76	1	0.17	0.35	0.1	0.88	0.43	0.051	0.2	0.052	0.24	0.76	0.18	0.012	0.029	0.32
sqft_lot -	0.09	0.032	0.088	0.17	1	0.0048	0.021	0.18	0.015	0.005	0.13	0.086	0.23	0.14	0.72	0.0026	0.0056	0.053
floors -	0.26	0.18	0.5	0.35	0.0048	1	0.021	0.52	0.24	0.0038	0.06	0.049	0.13	0.28	0.011	0.014	0.022	0.49
waterfront -	0.26	0.0021	0.064	0.1	0.021	0.021	1	0.072	0.083	0.074	0.029	0.012	0.038	0.084	0.031	0.0082	0.005	0.024
sqft_above -	0.61	0.48	0.69	0.88	0.18	0.52	0.072	1	0.051	0.021	0.26	0.0012	0.34	0.73	0.2	0.01	0.024	0.42
sqft_basement -	0.32	0.3	0.28	0.43	0.015	0.24	0.083	0.051	1	0.065	0.073	0.11	0.14	0.2	0.016	0.0071	0.015	0.13
yr_renovated -	0.12	0.018	0.047	0.051	0.005	0.0038	0.074	0.021	0.065	1	0.062	0.028	0.065	0.00068	0.0043	0.0076	0.02	0.2
zipcode -	0.053	0.15	0.2	0.2	0.13	0.06	0.029	0.26	0.073	0.062	1	0.27	0.56	0.28	0.15	0.00014	0.0013	0.35
lat -	0.31	0.01	0.024	0.052	0.086	0.049	0.012	0.0012	0.11	0.028	0.27	1	0.14	0.049	0.086	0.015	0.029	0.15
long -	0.022	0.13	0.22	0.24	0.23	0.13	0.038	0.34	0.14	0.065	0.56	0.14	1	0.34	0.26	0.0084	0.0003	0.41
sqft_living15 -	0.59	0.39	0.57	0.76	0.14	0.28	0.084	0.73	0.2	0.00068	0.28	0.049	0.34	1	0.18	0.0025	0.022	0.33
sqft_lot15 -	0.083	0.031	0.088	0.18	0.72	0.011	0.031	0.2	0.016	0.0043	0.15	0.086	0.26	0.18	1	0.0032	0.00016	0.071
month_sold -	0.0099	0.001	0.0075	0.012	0.0026	0.014	0.0082	0.01	0.0071	0.0076	0.00014	0.015	0.0084	0.0025	0.0032	1	0.78	0.0062
yr_sold -	0.0037	0.0099	0.027	0.029	0.0056	0.022	0.005	0.024	0.015	0.02	0.0013	0.029	0.0003	0.022	0.00016	0.78	1	0.0036
age -	0.054	0.16	0.51	0.32	0.053	0.49	0.024	0.42	0.13	0.2	0.35	0.15	0.41	0.33	0.071	0.0062	0.0036	1
	price -	bedrooms -	bathrooms -	sqft_living -	sqft_lot -	floors -	waterfront -	sqft_above -	sqft_basement -	yr_renovated -	zipcode -	lat -	- buol	sqft_living15 -	sqft_lot15 -	month_sold -	yr_sold -	- aŭe

- 0.8

- 0.6

- 0.4

- 0.2

## Question 2:

Which features have the strongest correlations with other predictor variables?

The features are listed below in descending order:

- Square Footage of the Living Space
- Square Footage of the House apart from the Basement
- Square footage of the basement

#### Variable Correlations

price -	1	0.31	0.53	0.7	0.09	0.26	0.26	0.61	0.32	0.12	0.053	0.31	0.022	0.59	0.083	0.0099	0.0037	0.054
bedrooms -	0.31	1	0.51	0.58	0.032	0.18	0.0021	0.48	0.3	0.018	0.15	0.01	0.13	0.39	0.031	0.001	0.0099	0.16
bathrooms -	0.53	0.51	1	0.76	0.088	0.5	0.064	0.69	0.28	0.047	0.2	0.024	0.22	0.57	0.088	0.0075	0.027	0.51
sqft_living -	0.7	0.58	0.76	1	0.17	0.35	0.1	0.88	0.43	0.051	0.2	0.052	0.24	0.76	0.18	0.012	0.029	0.32
sqft_lot -	0.09	0.032	0.088	0.17	1	0.0048	0.021	0.18	0.015	0.005	0.13	0.086	0.23	0.14	0.72	0.0026	0.0056	0.053
floors -	0.26	0.18	0.5	0.35	0.0048	1	0.021	0.52	0.24	0.0038	0.06	0.049	0.13	0.28	0.011	0.014	0.022	0.49
waterfront -	0.26	0.0021	0.064	0.1	0.021	0.021	1	0.072	0.083	0.074	0.029	0.012	0.038	0.084	0.031	0.0082	0.005	0.024
sqft_above -	0.61	0.48	0.69	0.88	0.18	0.52	0.072	1	0.051	0.021	0.26	0.0012	0.34	0.73	0.2	0.01	0.024	0.42
sqft_basement -	0.32	0.3	0.28	0.43	0.015	0.24	0.083	0.051	1	0.065	0.073	0.11	0.14	0.2	0.016	0.0071	0.015	0.13
yr_renovated -	0.12	0.018	0.047	0.051	0.005	0.0038	0.074	0.021	0.065	1	0.062	0.028	0.065	0.00068	0.0043	0.0076	0.02	0.2
zipcode -	0.053	0.15	0.2	0.2	0.13	0.06	0.029	0.26	0.073	0.062	1	0.27	0.56	0.28	0.15	0.00014	0.0013	0.35
lat -	0.31	0.01	0.024	0.052	0.086	0.049	0.012	0.0012	0.11	0.028	0.27	1	0.14	0.049	0.086	0.015	0.029	0.15
long -	0.022	0.13	0.22	0.24	0.23	0.13	0.038	0.34	0.14	0.065	0.56	0.14	1	0.34	0.26	0.0084	0.0003	0.41
sqft_living15 -	0.59	0.39	0.57	0.76	0.14	0.28	0.084	0.73	0.2	0.00068	0.28	0.049	0.34	1	0.18	0.0025	0.022	0.33
sqft_lot15 ·	0.083	0.031	0.088	0.18	0.72	0.011	0.031	0.2	0.016	0.0043	0.15	0.086	0.26	0.18	1	0.0032	0.00016	0.071
month_sold -	0.0099	0.001	0.0075	0.012	0.0026	0.014	0.0082	0.01	0.0071	0.0076	0.00014	0.015	0.0084	0.0025	0.0032	1	0.78	0.0062
yr_sold -	0.0037	0.0099	0.027	0.029	0.0056	0.022	0.005	0.024	0.015	0.02	0.0013	0.029	0.0003	0.022	0.00016	0.78	1	0.0036
age -	0.054	0.16	0.51	0.32	0.053	0.49	0.024	0.42	0.13	0.2	0.35	0.15	0.41	0.33	0.071	0.0062	0.0036	1
	price -	bedrooms -	bathrooms -	sqft_living -	sqft_lot -	floors -	waterfront -	sqft_above -	sqft_basement -	yr_renovated -	zipcode -	lat -	- buol	sqft_living15 -	sqft_lot15 -	month_sold -	yr_sold -	age -

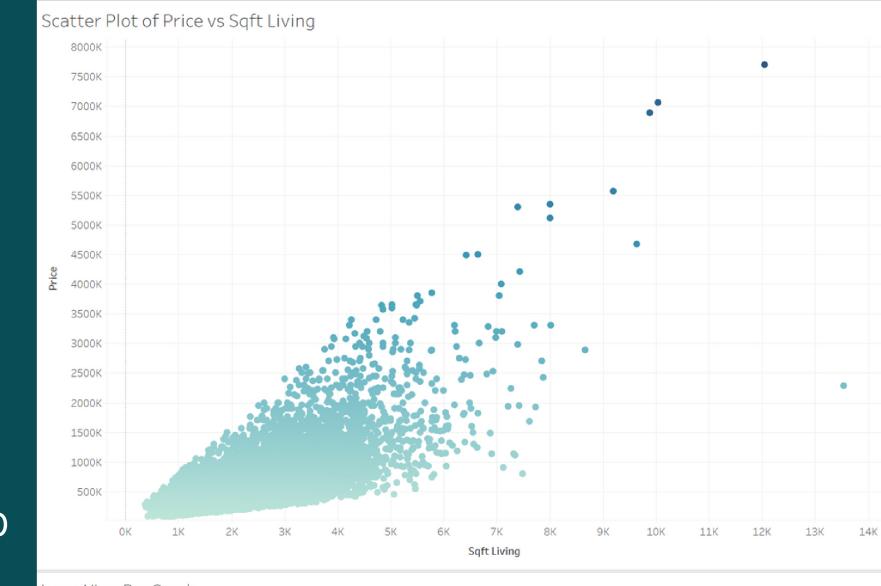
- 0.8 - 0.6

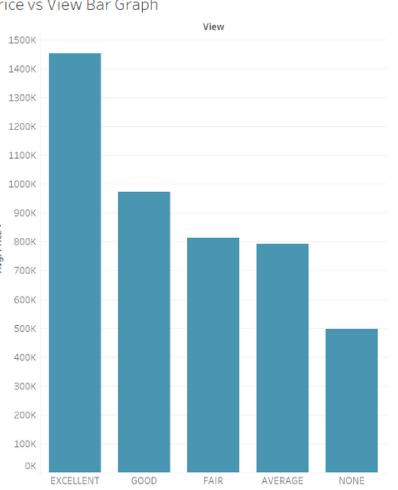
## Question 3:

What combination of features is the best fit, in terms of predictive power, for a multiple regression model to predict house prices?

The features are listed below in descending order:

- Square Footage of the Living Space. For each additional square footage of the house, the price is projected to increase by USD 200
- Square footage of Interior Housing Living Space for the Nearest 15 Neighbors. The price of a house is projected to increase by USD 52 for every increment of square footage of the nearest 15 neighbors.
- The View. House with the best views will always go for a very high price. Below is a cost comparison to houses with no view:
  - a. One with an excellent view will cost USD600k more
  - b. One with a good view will cost USD 200k more
  - c. One with fair view will cost USD 160k more
  - d. One with an average view will cost USD 119k more





### Conclusion

- 1. The houses with the best views always fetch a high price
- 2. A larger living space will increase the price of a house
- 3. The size of the neighboring houses affects the cost of a house. The larger the houses, the higher the price.

