

King County House Prices Study

Purpose of the study:

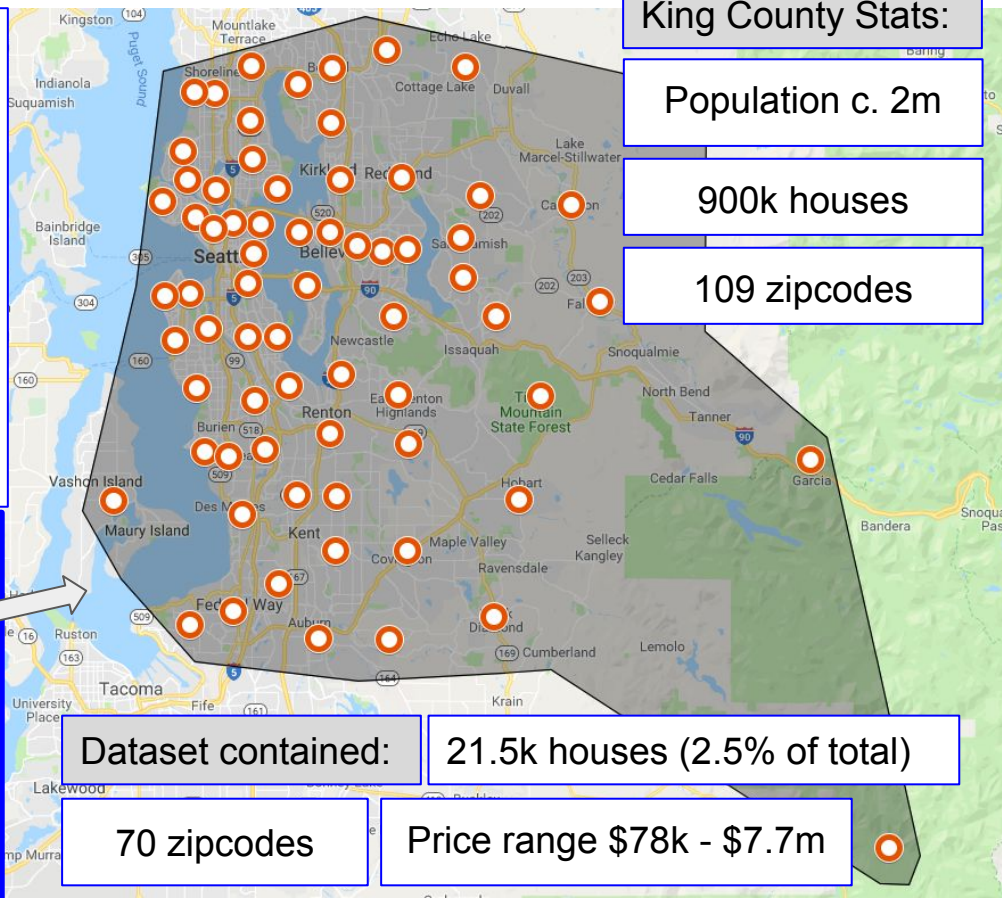
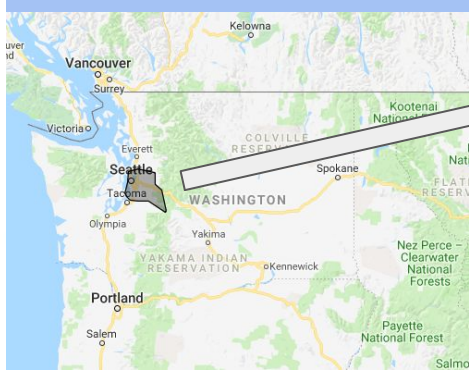
- To predict house prices as accurately as possible
- To highlight recommendations to homeowners looking to maximise the selling price of their home.

This study develops a model that predicts house-prices with over 70% accuracy.



King County

King County is the largest metropolitan county in the State of Washington in terms of population, number of cities and employment. It is the thirteenth most populous county in the United States.

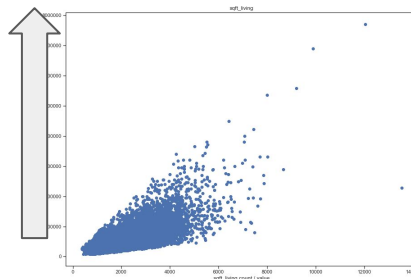


Interpreting the dataset

Each house had 21 'features'

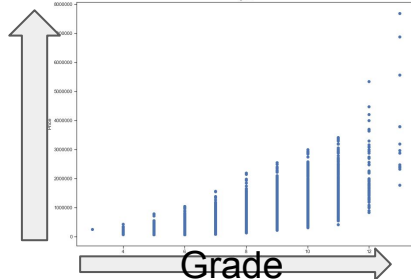
***id** - unique identified for a house
***dateDate** - house was sold
***bedroomsNumber** - of Bedrooms/House
***bathroomsNumber** - of bathrooms/bedrooms
***sqft_livingsquare** - footage of the home
***sqft_lotsquare** - footage of the lot
***floorsTotal** - floors (levels) in house
***waterfront** - House which has a view to a waterfront
***view** - Has been viewed
***condition** - How good the condition is (Overall)
***grade** - overall grade given to the housing unit, based on King County grading system
***sqft_above** - square footage of house apart from basement
***sqft_basement** - square footage of the basement
***yr_built** - Built Year
***yr_renovated** - Year when house was renovated
***sqft_living15** - Living room area in 2015(implies-- some renovations) This might or might not have affected the lotsize area
***sqft_lot15** - lotSize area in 2015(implies-- some renovations)
***zipcode** - zip
***lat** - Latitude coordinate
***long** - Longitude coordinate

Price



Square footage

Price



Grade

The two features **most relevant** to house price were identified as **sqft_living** and **grade**:

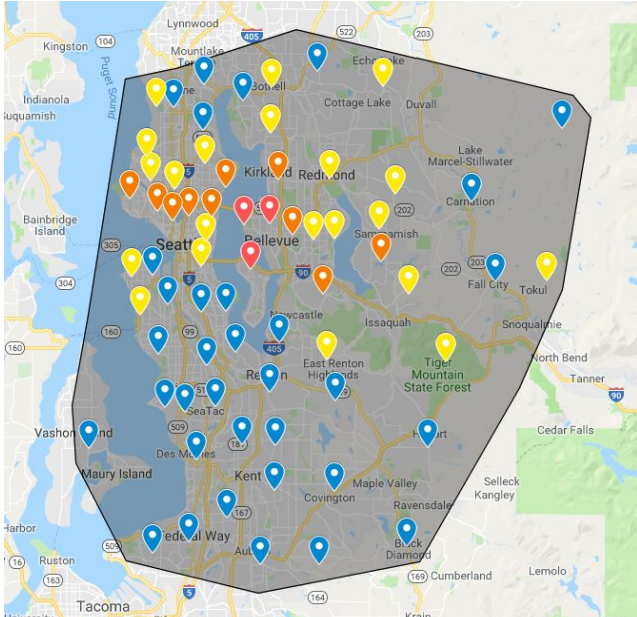
A number of features were identified as **not relevant** to the analysis because:





- There was not enough data provided to be useful to the analysis
- The data did not make sense according to its description.
- The feature did not show any connection at all to price.
- The feature too closely associated with sqft_living (e.g sqft_above)

?

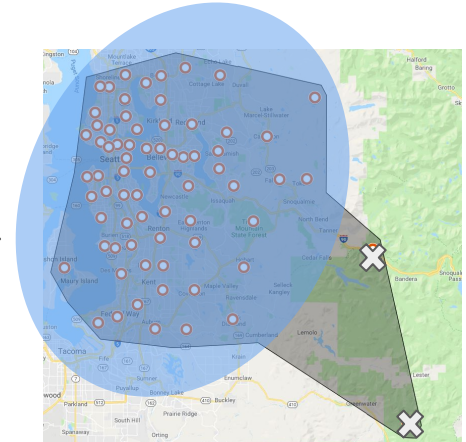
Intuitively location must play an important role in house price so...

IN ORDER TO MAKE USE OF LOCATION IN PREDICTING HOUSE PRICES ZIPCODE CATEGORIES WERE CREATED:



-  Zip Area A - average house price \$240k - 475k (8478 houses)
-  Zip Area B - average house price \$475k - 715k (6809 houses)
-  Zip Area C - average house price \$715k - 950k (2317 houses)
-  Zip Area D - average house price \$950k - \$1.4m (495 houses)

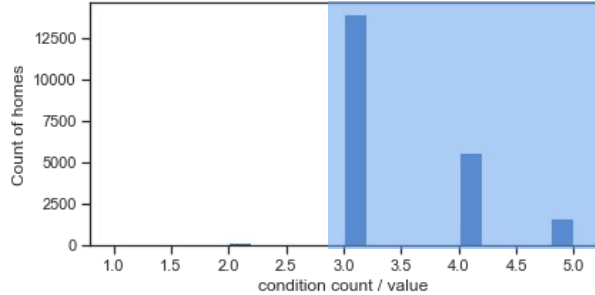
Note: two zipcodes were not included as they were regarded as geographical outliers:



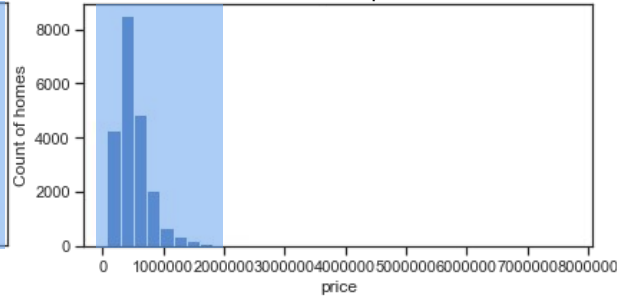
TRIMMING THE DATASET

The dataset was trimmed so that the model could apply to a broad but reasonably typical demographic of homeowners. Areas shaded in blue show the data that was selected for the analysis.

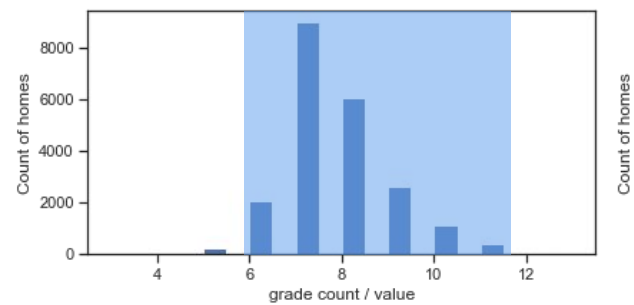
Condition: 3 - 5



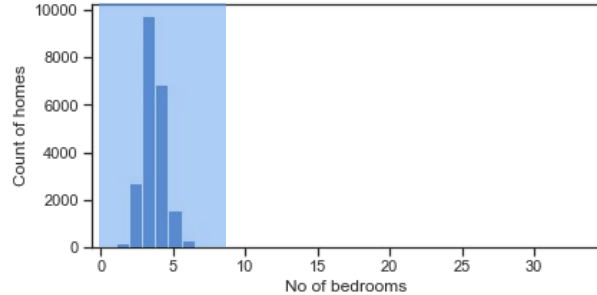
Price < \$2m



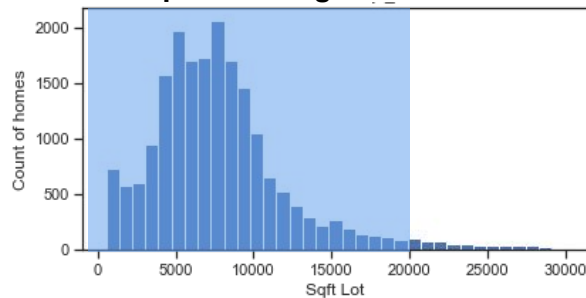
Grade 6 - 11



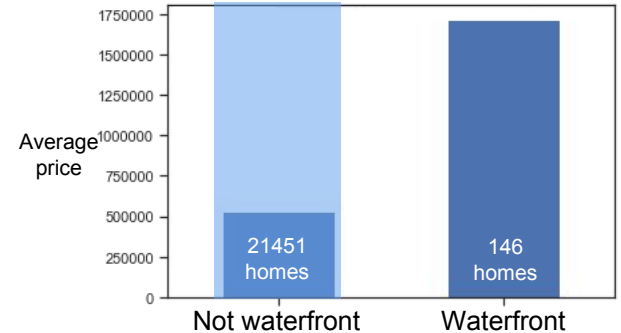
No of Bedrooms < 8



Square-footage of Lot < 2000



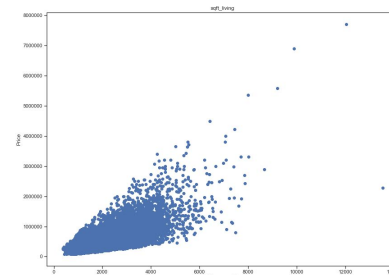
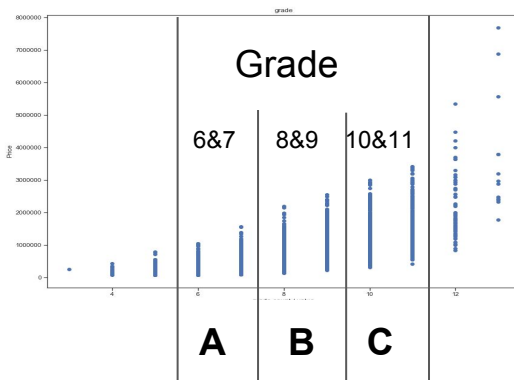
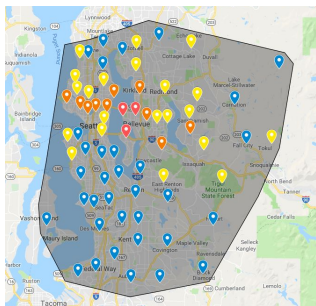
Waterfront homes



This removed around 16% of the dataset.
There were also some duplicate entries in the dataset which were removed.

Final model finds a house price based on zip category, grade category and square footage of living area.

Zipcode areas



Square footage →



**Price
Prediction**

9751 7391 956

(number of homes in each
grade category)

SOME SAMPLE PREDICTIONS

Predictions for each Grade Category:

sqft_living	zip_cat	grade_cat	pred_price_grade_A	pred_price_grade_B	pred_price_grade_C	sale_price	Actual / pred %
1340	A	A	259599	306811	472927	313000	82.93
1580	A	B	293160	340371	506488	375000	90.76
2820	A	B	466556	513767	679884	450000	114.17
2270	B	A	576464	623676	789792	645000	89.37
1990	B	C	537310	584522	750638	639000	117.47
3000	C	B	818161	865373	1031489	770000	112.38
1820	D	A	899804	947015	1113132	875000	102.83

Recommendations to homeowners

Aim to improve grade as it has a direct influence on price.

King County Grade Score descriptions:

6 Lowest grade currently meeting building code. Low quality materials and simple designs.

7 Average grade of construction and design. Commonly seen in plats and older sub-divisions.

8 Just above average in construction and design. Usually better materials in both the exterior and interior finish work.

9 Better architectural design with extra interior and exterior design and quality.

10 Homes of this quality generally have high quality features. Finish work is better and more design quality is seen in the floor plans. Generally have a larger square footage.

11 Custom design and higher quality finish work with added amenities of solid woods, bathroom fixtures and more luxurious options.

Use model to estimate value of house if a higher grade category is achieved through (for example):

Re-design of interior / exterior.

Loft conversion

Basement extension

Request an up-to-date grade score from the King County Assessors' office and / or visit local homes with a higher grade in the area.