


Predicting Housing Prices in King County, WA

NOVEMBER 2020

MARK, MATT, MEL

Context

M3 Consulting

Working with several non-profit groups in King County, WA:

1) What features help predict home price (under $500K)?
2) What features can be minimized to bring down home price?
3) Where in King County should new buyers look for affordable homes?

"Helping good people find affordable homes."

Understanding the Context

King County, WA:

- Population (2010): 1.9 million
- HH Income (2014): $75K
- Median home value (2014-2018): $494K
- Tech boom of the 2000 has created one of the most expensive housing markets in the country
- In a recent survey, the top feature buyers said they want most:
  - *“a home that is within my initial budget” (89%)*

Meet the Salazar family

Currently renting an apartment

Looking to buy their first home

HH Income of $75K

Assuming a $10K down payment, they can afford a **$316K** home*

**based on an affordability calculator*

“We want a home we can afford, and that will be a safe place for our kids to grow.”

Data

Exploring the Data

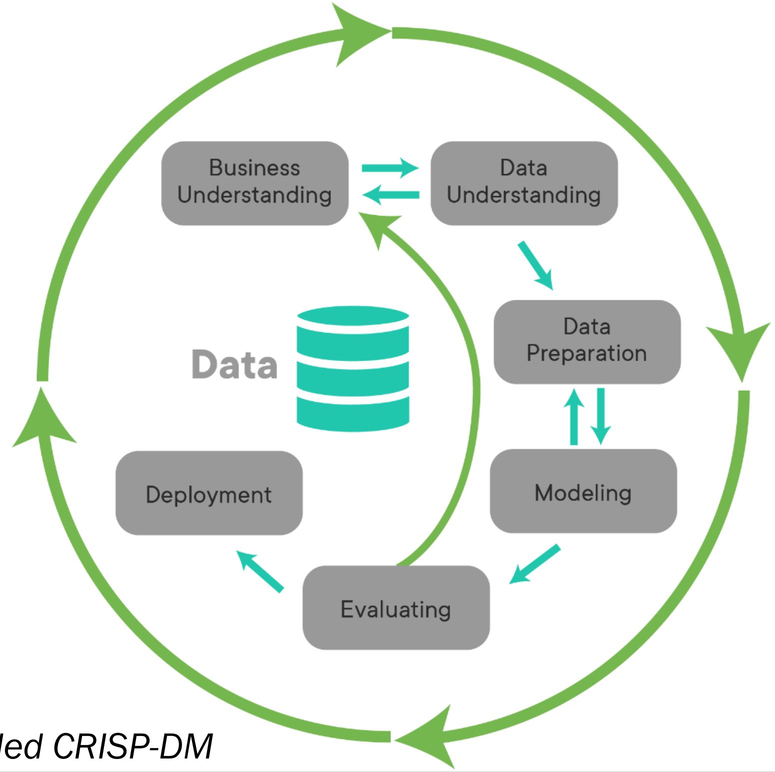
Housing Sales Data:

- Data provided by King County Assessors Office
- Home sales: May 2014 to May 2015
- ~ 21,000 records
- 21 variables

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.

Our Process*:



** This process is called CRISP-DM*

Our Data Journey

Researched features

Visualized and looked at descriptive statistics

Examined distributions

Looked for linear relationships

Formated features as numbers; null values turned into 0's

Removed un-used features
(zip code, date, longitude)

Checked for duplicates

BUILDING GRADE

Represents the construction quality of improvements. Grades run from grade 1 to 13. Generally defined as:

1-3 Falls short of minimum building standards. Normally cabin or inferior structure.

4 Generally older, low quality construction. Does not meet code.

5 Low construction costs and workmanship. Small, simple design.

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

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

8 Just above average in construction and design. Usually better materials in both the e

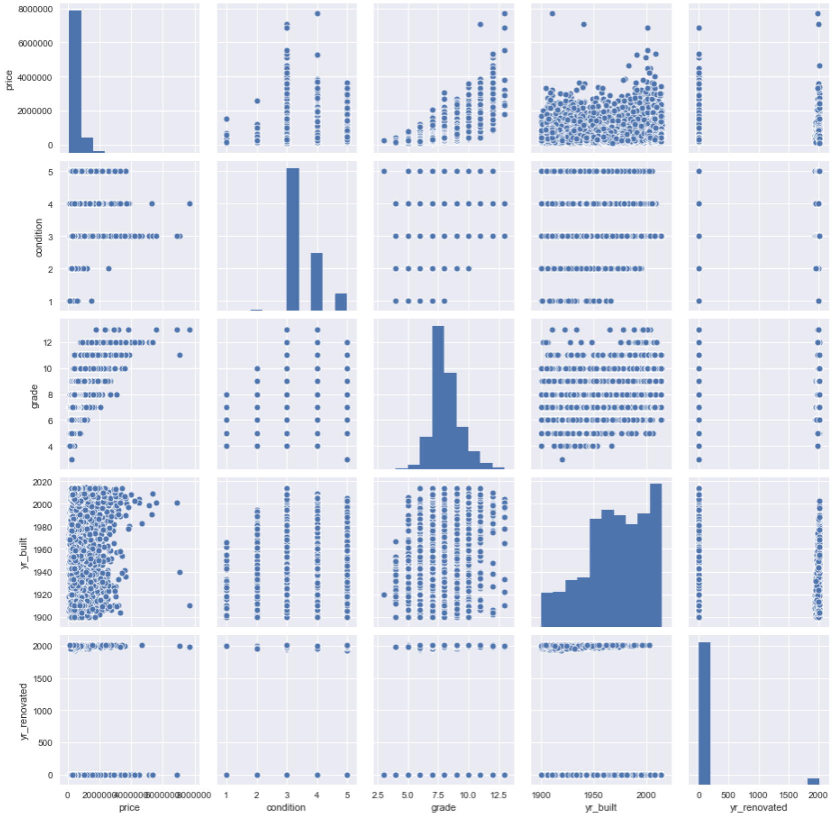
9 Better architectural design with extra interior and exterior design and quality.

10 Homes of this quality generally have high quality features. Finish work is better and footage.

11 Custom design and higher quality finish work with added amenities of solid woods,

12 Custom design and excellent builders. All materials are of the highest quality and al

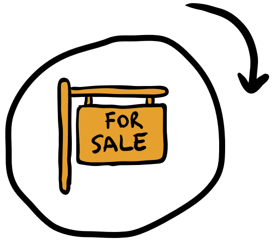
13 Generally custom designed and built. Mansion level. Large amount of highest qualit



	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	view	condition	grade	sqft
count	2.045200e+04	20452.000000	20452.000000	20452.000000	2.045200e+04	20452.000000	20452.000000	20452.000000	20452.000000	20452.000000	20452
mean	4.824466e+05	3.318942	2.034581	1971.131527	1.442586e+04	1.473988	0.002934	0.180227	3.409544	7.537209	1703
std	2.198012e+05	0.867106	0.674796	763.014605	3.907029e+04	0.536213	0.054085	0.658488	0.648611	1.041994	716
min	7.800000e+04	1.000000	0.500000	370.000000	5.200000e+02	1.000000	0.000000	0.000000	1.000000	3.000000	370
25%	3.150000e+05	3.000000	1.500000	1400.000000	5.000000e+03	1.000000	0.000000	0.000000	3.000000	7.000000	1170
50%	4.380000e+05	3.000000	2.000000	1860.000000	7.500000e+03	1.000000	0.000000	0.000000	3.000000	7.000000	1520
75%	6.050000e+05	4.000000	2.500000	2430.000000	1.034450e+04	2.000000	0.000000	0.000000	4.000000	8.000000	2100
max	1.270000e+06	11.000000	3.500000	7350.000000	1.651359e+06	3.500000	1.000000	4.000000	5.000000	12.000000	5450
+2_std	9.220490e+05	5.053153	3.384173	3497.160738	9.256644e+04	2.546414	0.111105	1.497203	4.706766	9.621196	3135
-2_std	4.284423e+04	1.584730	0.684989	445.102317	-6.371472e+04	0.401562	-0.105237	-1.136749	2.112322	5.453222	271

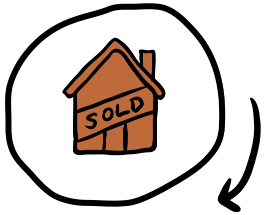
Modeling

Our Modeling Journey



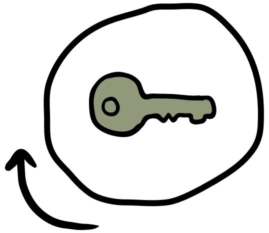
A) Mid-priced

Limited price:
$315K to $605K
(7 models; r2~0.10)



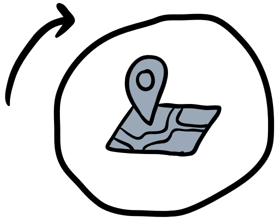
B) Low-priced

Limited price:
$154K to $315K
(8 models; r2~0.19)



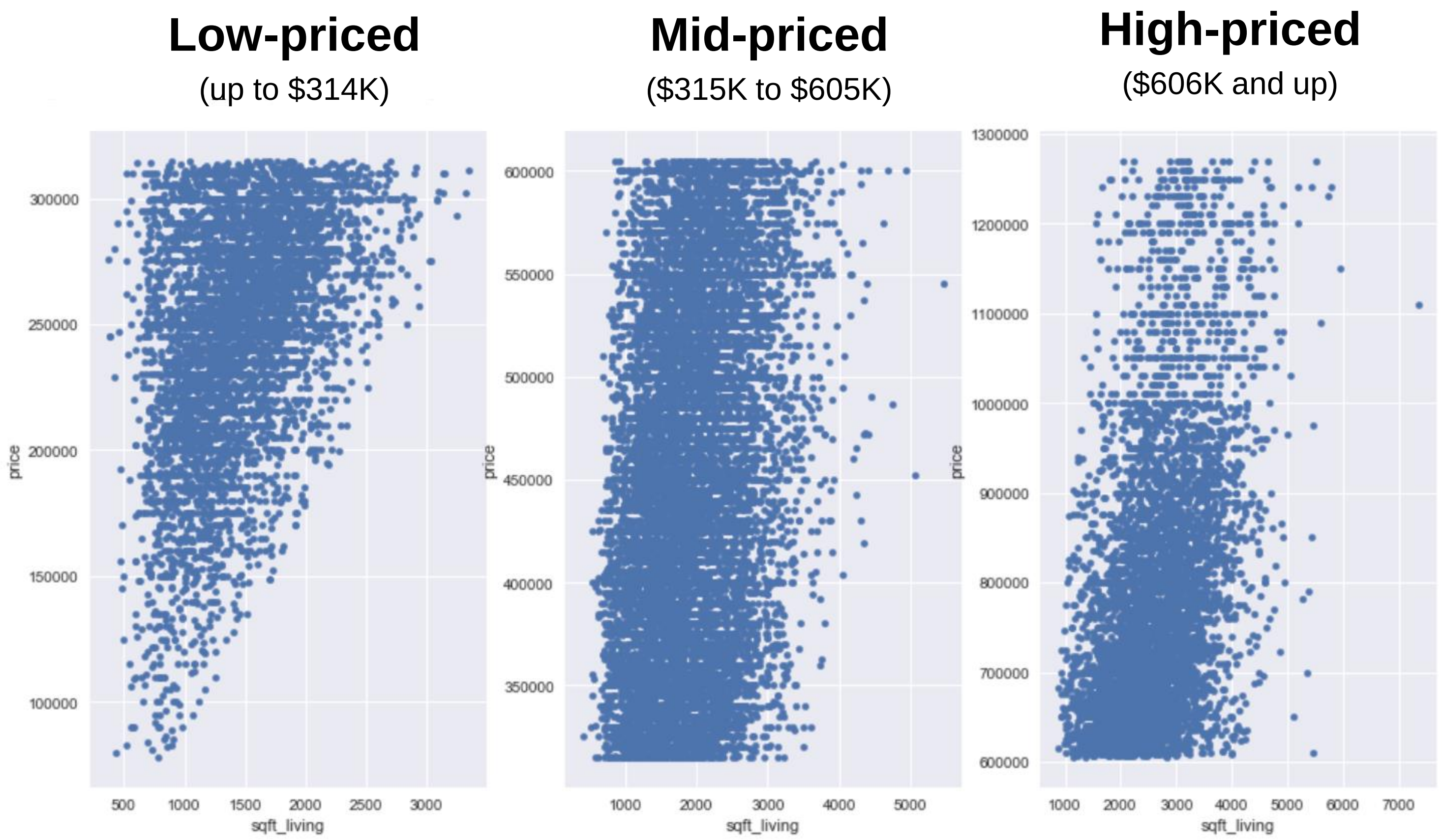
C) Adjust features

Tried adjusting scales of
most features (min-max)
and logged grade.
(3 models; r2~0.20)



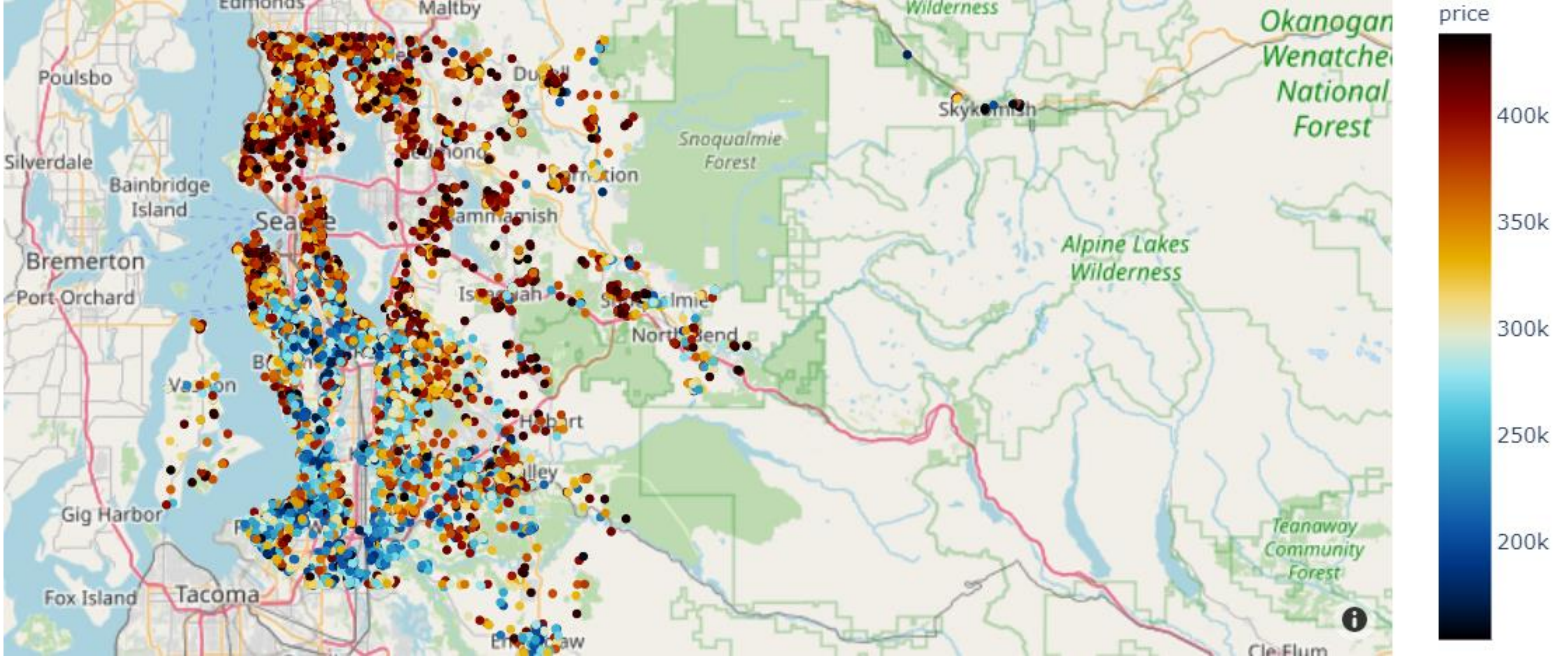
D) Add latitude

Added in latitude (bands)
and this helped.
Broadened price to $453K.
(4 models; r2~0.492)



Low-priced shows clearer linear relationship

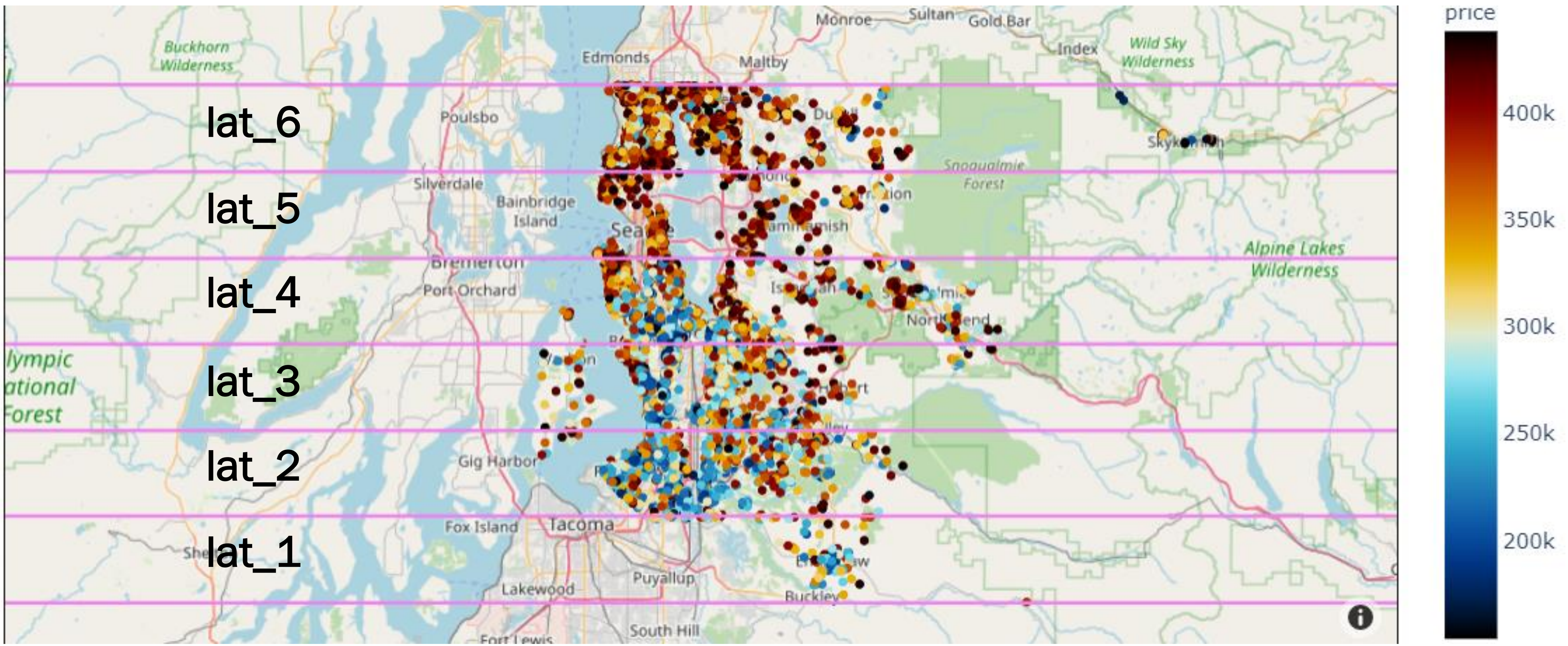
Re-examining scatter plots led us to shift from mid-priced homes to low-priced homes



Location, location, location…

Plotting home sales on a map revealed an important pattern in our data

(latitude more pronounced differences than longitude)



Created 6 bands for latitude

From the bands, created separate “dummy” features (lat_1 to lat_6)*.

** Note that lat_1 was “dropped” from the predictor set.*

Model Adjustments

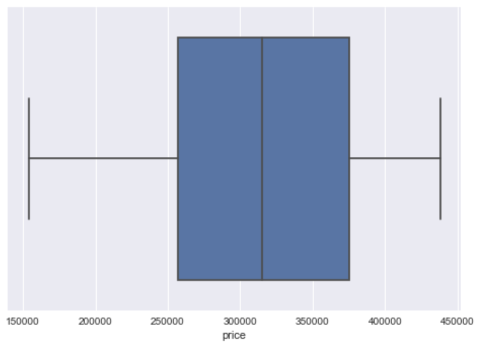
Removed outliers

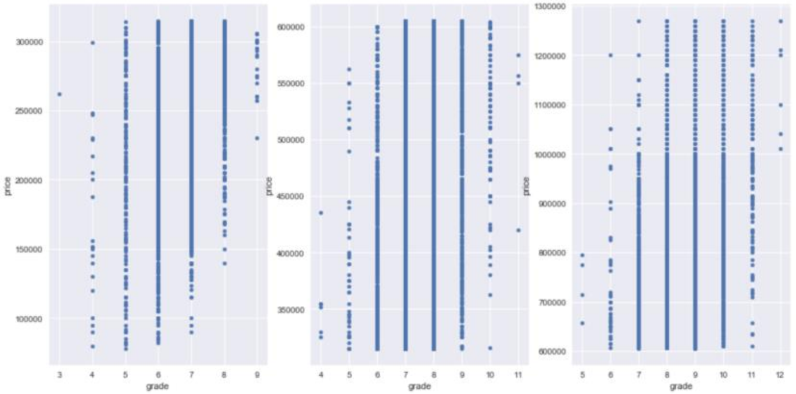
Created banding and "dummy" variables (latitude; year built)

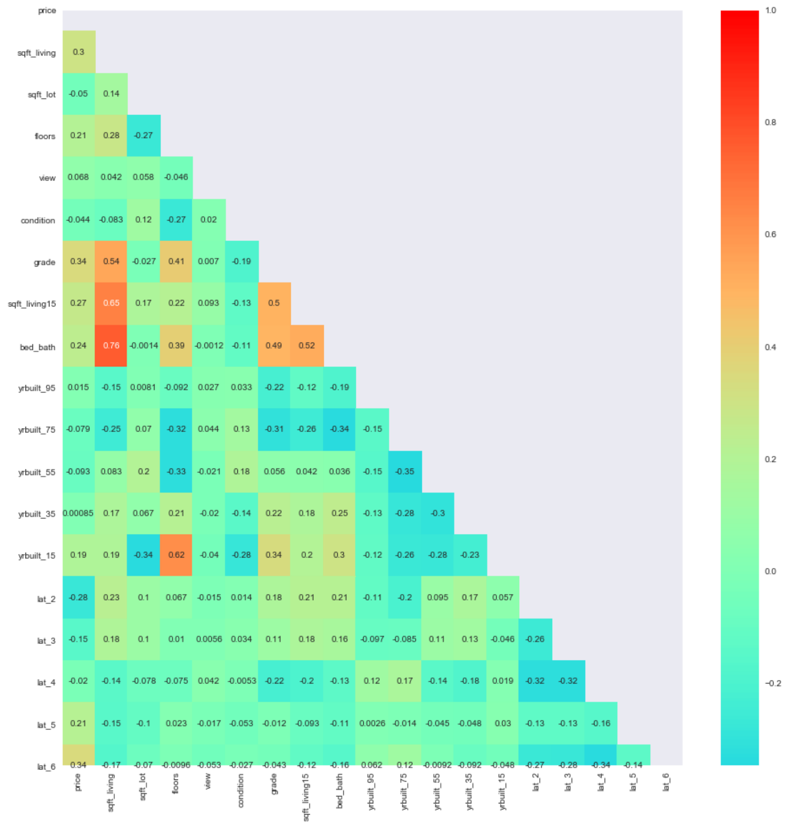
Created new variables (bed and bath combined)

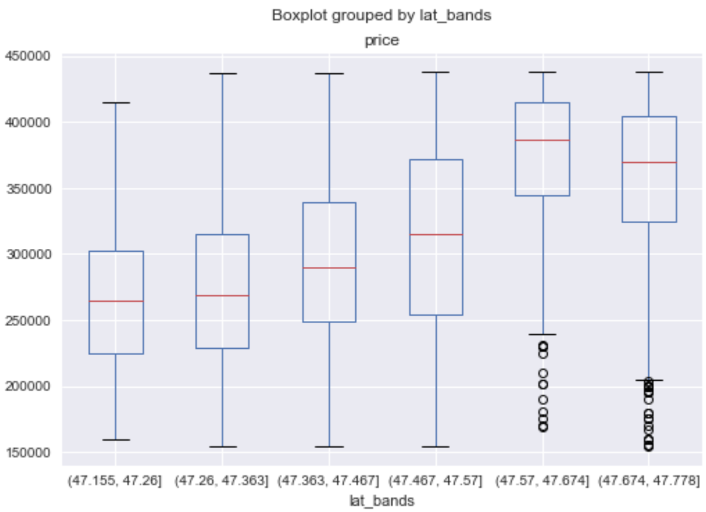
Didn't include highly related features

Removed insignificant features









Our Final Model

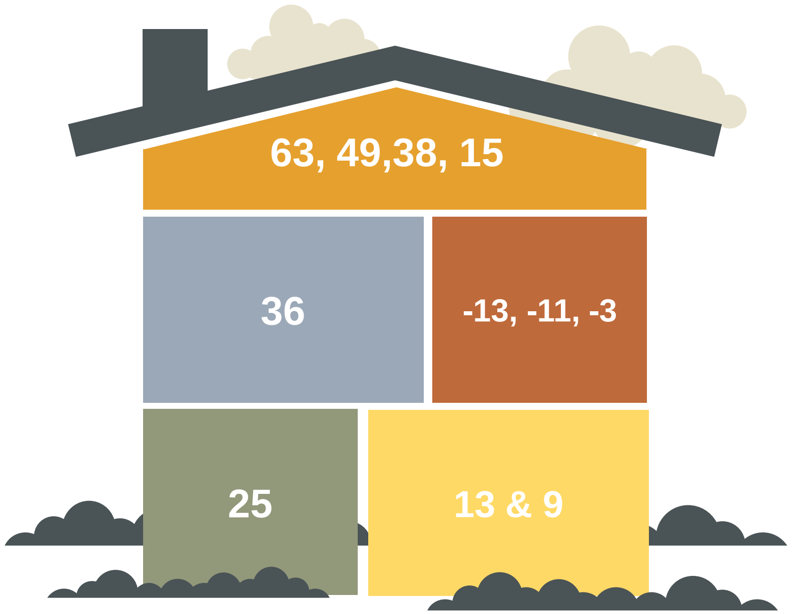
Top features contributing to the model (t-scores*):

Sqft Living Area

From 370 to 3,090 sqft

Grade

Construction quality: includes grades 5 to 9 (of 13).



Latitude

4 of the 6 latitude bands (3, 4, 5, 6)

Year built

3 of 6 bands (1940 to 2000)

Other

Condition (5 levels) and View (4 levels of quality)

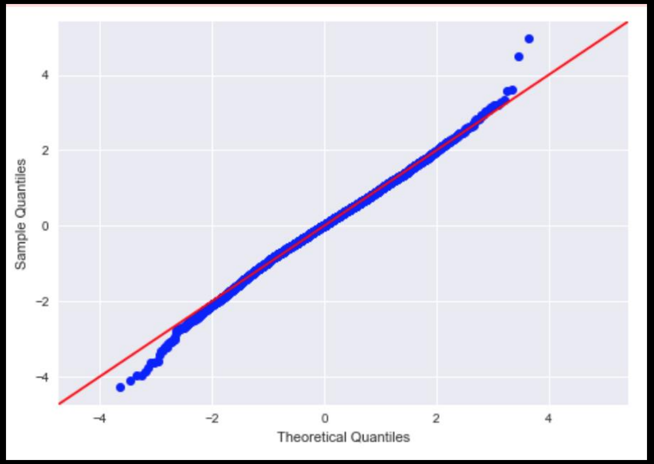
** Higher t-scores mean the feature contributes more to the model.*

Model Details:

- Price is target feature ($154K to $438K)
- 11 predictor features
- n = 7,212 (train set)

Success Criteria :

- r2 = **0.52**
- RMSE = **50,137** (test)
- Cross Validation (8-folds) RMSE = **49,932** (mean)



Predicting house price +

Example 1: size

- 1,800 sq. ft living area = $366K
- 1,000 sq. ft living area = $324K

Example 2: location

- Latitude 5 (1,200 sqft) = $365K
- Latitude 3 (1,200 sqft) = $256K

Note: +/- $50K (error in our model)

+ from the prediction model, with all other variables held constant (grade 7, condition 3, view 0, yr built 1960 – 80)

Predicting with the Model (The Salazar's $316K target price)

A) **Size:** Each additional sq. ft of living area will add $52 to the home price.*

B) **Location:** A home in latitude 5 (Seattle) will add ~$135K to the home price versus a home in latitude 1 (south).*

**** Based on coefficient values for the feature, when all other features are held constant.***

Conclusions

Insights

Key Learnings

- Location is a key factor
- Living space, quality also important
- Iterating is vital to modelling

Limitations of our Model / Approach

- Relatively low predictive power; high margin of error
- Limiting price range may have hurt model
- Data not fulfill all of model assumptions
- Over-reliance on categorical data?
- Linear regression may not be the best modeling approach for the data

Next Steps

More exploration of **location**:
- transit routes
- walkability
- proximity to grocery stores

Examine various **home types**:
- duplex
- town-homes
- condos

Investigate other dwelling / property **features**:
- floor plans
- private yard
- Parking

Explore impact of **Covid-19** on prices and market

Thanks

Many thanks to:

- Flatiron School
- Our fearless instructor: Yish Lim
- Fill-in instructors: Amber Y., Lindesy B., and Abhineet K.
- Our helpful classmates in cohort (onl-ft-092820)

Learn more at: https://github.com/melfriedman/KingHousing

Contact us:

Matthew: mattzhang989@gmail.com

Mark: markpatterson8@hotmail.com

Mel: melfriedman27@gmail.com