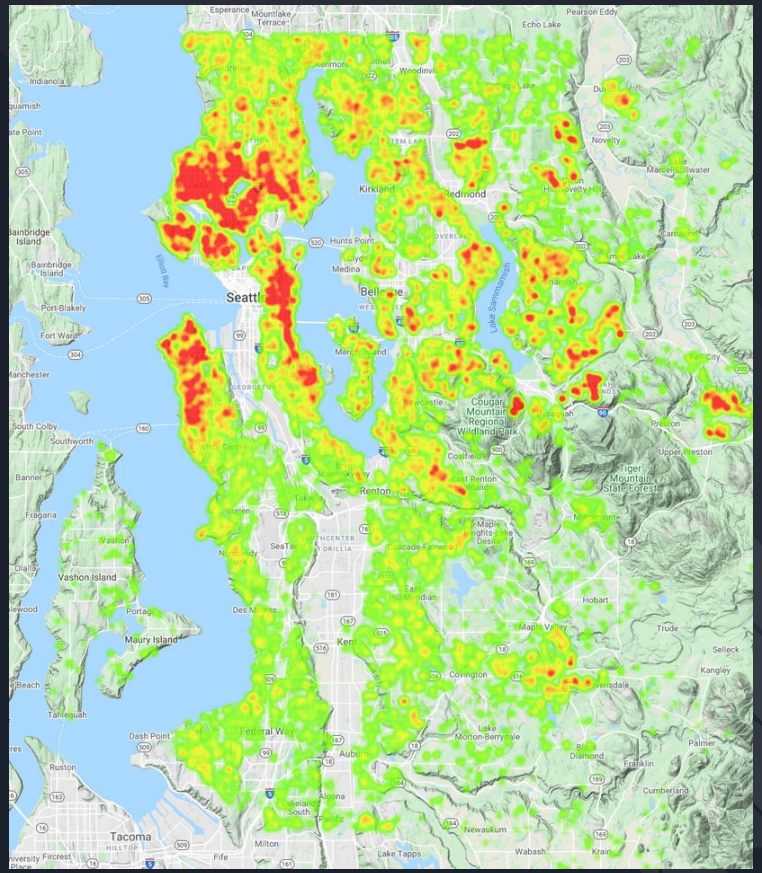
King County Real Estate Presentation

Flatiron School Module 2 Data Science Project

By: William Newton

Data & Methodology



Data & Methodology

- Data used for this project is from the King County Real Estate Dataset available on Kaggle
- Data covers home sales from May 2nd, 2014 to May 24th, 2015 in the metro-Seattle area
- Focus was on building a multiple linear regression model using home data under $1 million to predict housing prices for similar homes
- Used most effective predictors from the model for recommendations to home sellers
- Also focused on providing advice to prospective home sellers on what to do to increase their home's value and maximize selling price

Model Test Results

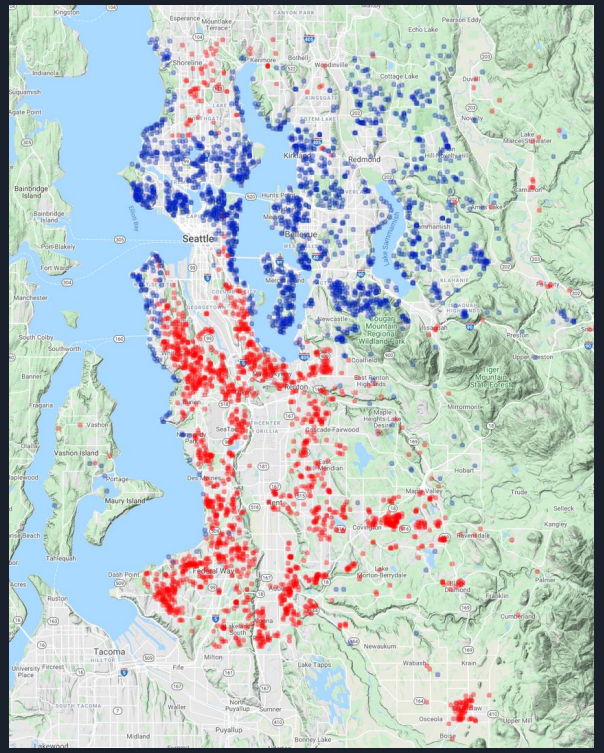
	Model #	Condition	R^2 Values
0	1	Baseline	0.528
0	2	Converted Grade and Condition to Categories	0.534
0	3	Normalized Numerical Values	0.527
0	4	Removed Floors and Bathrooms due to high p-value	0.527
0	5	Added Zipcode as Categorical Column	0.805
0	6	Removed Prices Greater than 1 Mil	0.804
0	7	Added has_basement and reno_new to model	0.805
0	8	Removed bed_bath_mult and added back Bedrooms ...	0.805
0	9	Condition and Grade back to numerical types	0.793
0	10	Removed Zipcodes with high P-values	0.793
0	11	One-Hot Encoded Grade and Condition	0.796

Final Variables Used in Model After Testing

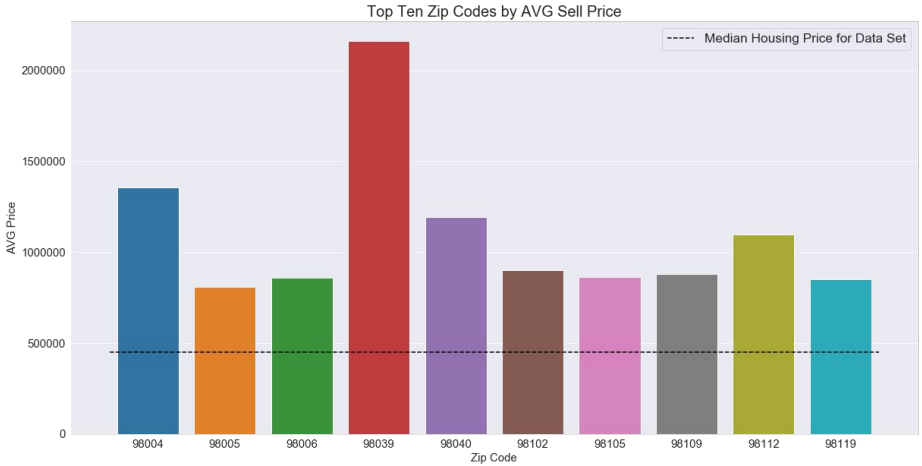
- Zip Code
- Bedrooms
- Bathrooms
- Living Space Sqft
- Has Basement: Yes or No?
- Has Been Renovated in Last 5 Years: Yes or No?
- Grade (Overall Grade of House)
- Condition (Overall Condition of Build or Renovation)

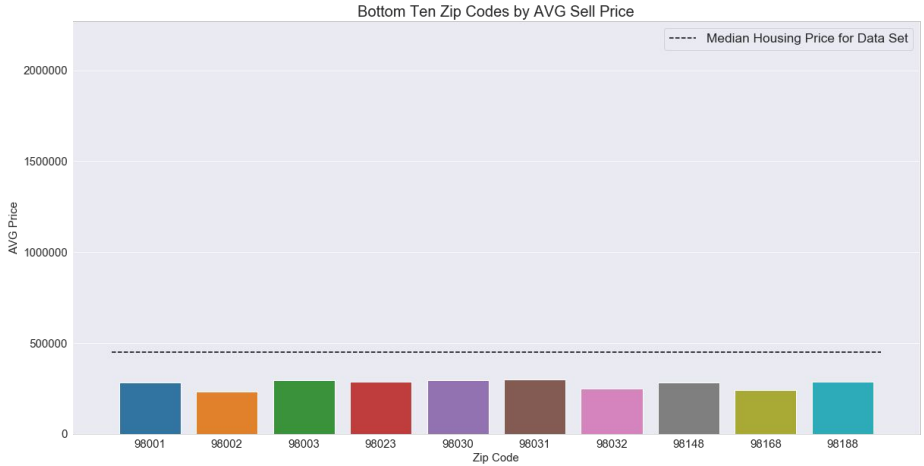
Location, Location, Location

- Zip Code and other location data was highly predictive for my model and should be used to set realistic expectations for home values.
- Top 10% Sell Price in Blue
- Bottom 10% Sell Price in Red



Location, Location, Location

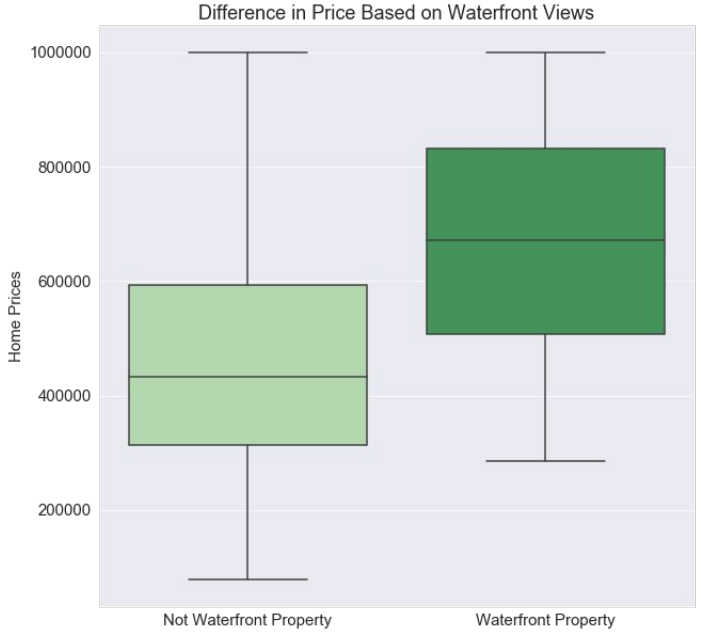


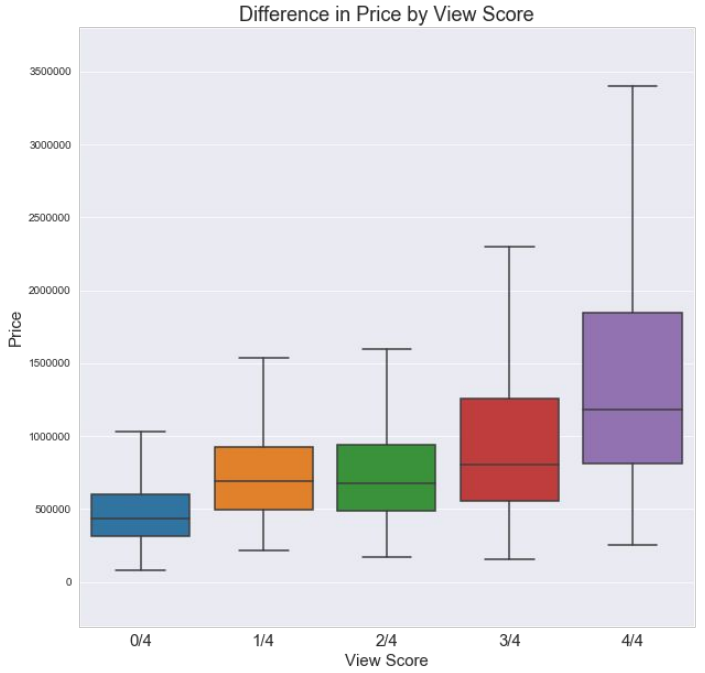


Median Home Price for King County: $450,000

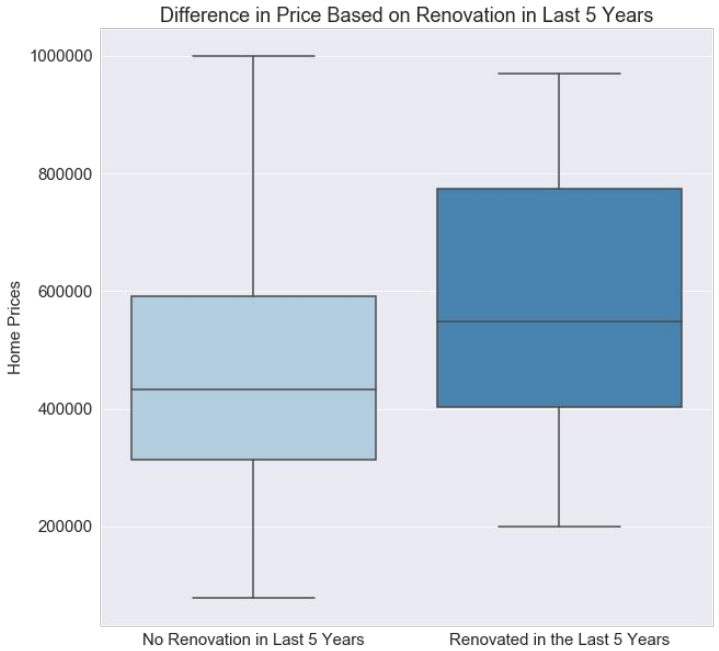
National Median Home Price: $226,800

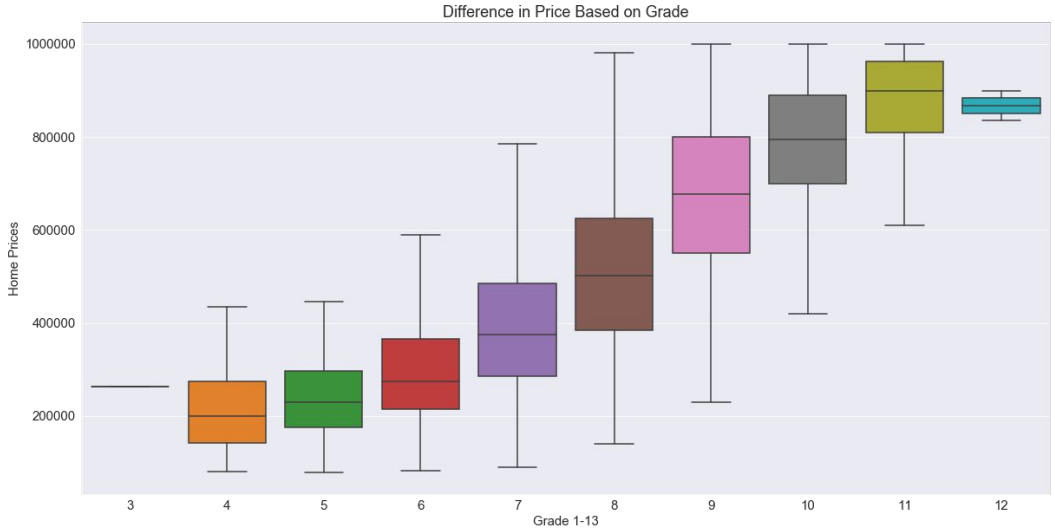
What a View!


Difference in Price Based on Waterfront Views
Home Prices
1000000
800000
600000
400000
200000
Not Waterfront Property
Waterfront Property

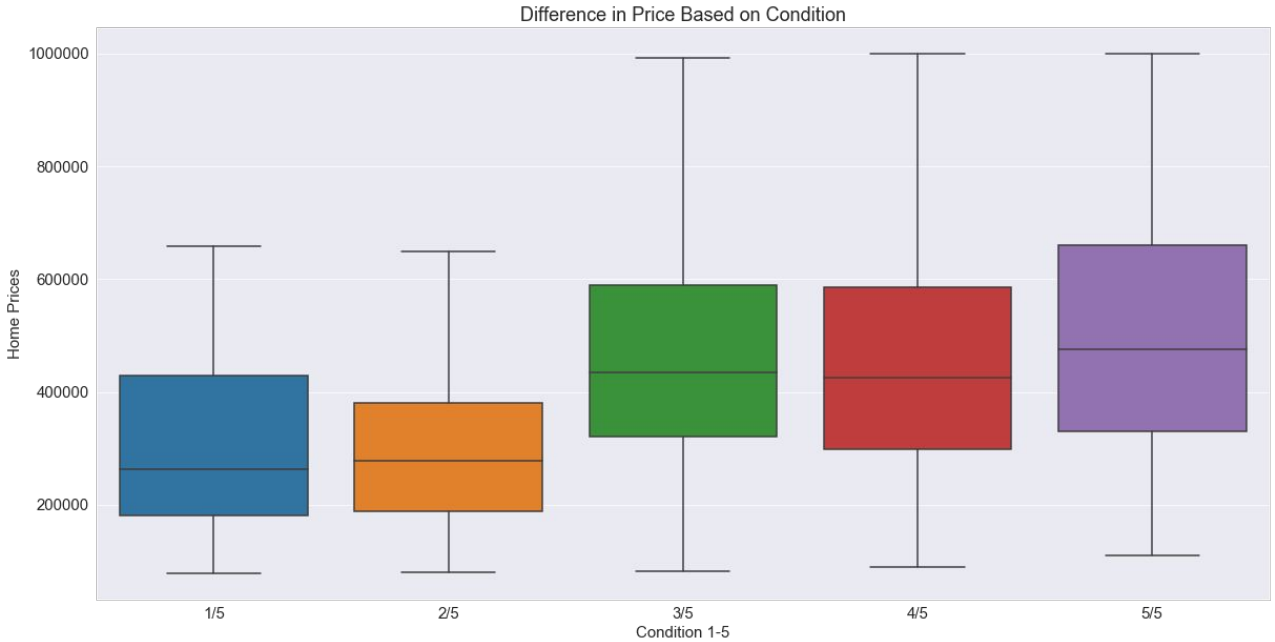

Difference in Price by View Score
Price
3500000
3000000
2500000
2000000
1500000
1000000
500000
0
0/4
1/4
2/4
3/4
4/4
View Score

What is within your control?

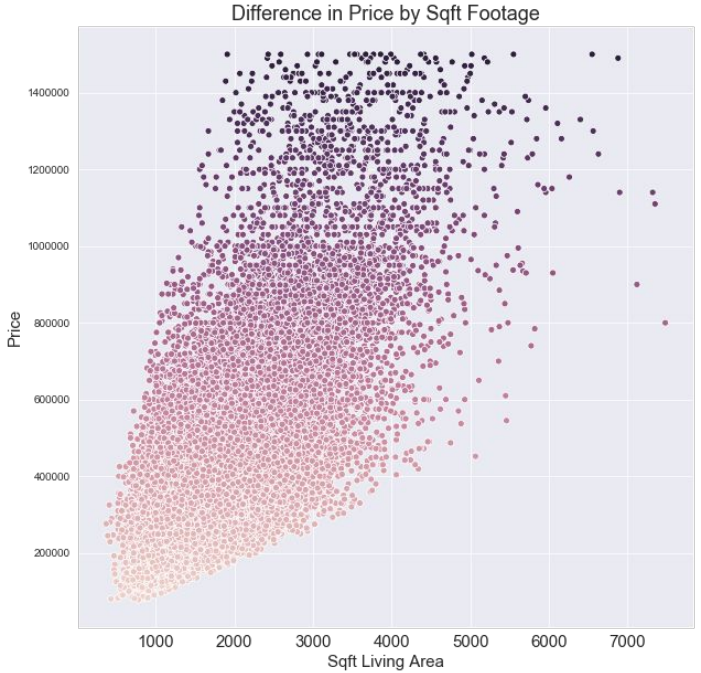




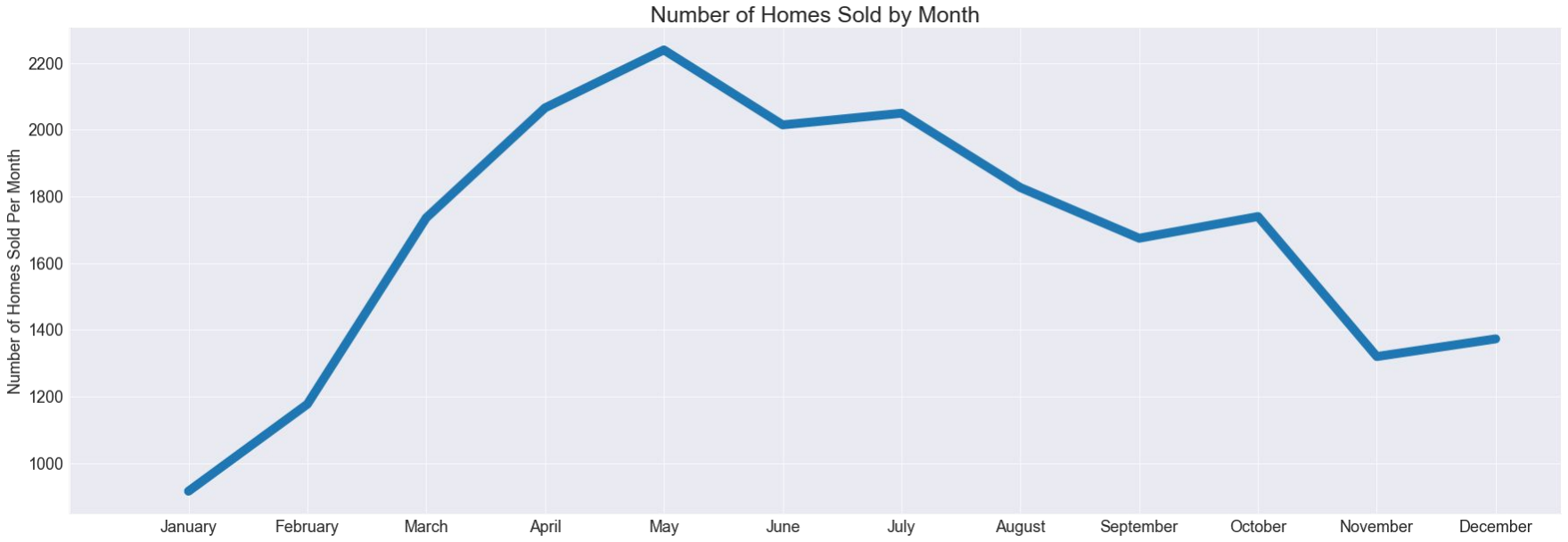
What is within your control?



Additions to your Home Can Increase Its' Value



Choose the Right Time to Put Your Home on the Market



Future Work

Given more time on this project and the data set, I would have loved to explore some additional features…

- Include demographic data to add further dimension to geographic information
- Get more data over a larger time frame to increase model fit. I have no way of knowing if this 2014 - 2015 data is an outlier if I looked at other year's data
- Further explore additional mapping libraries and functions. The gmaps library was easy to use but limited in its' functionality

Thanks and I look forward to working together soon!