

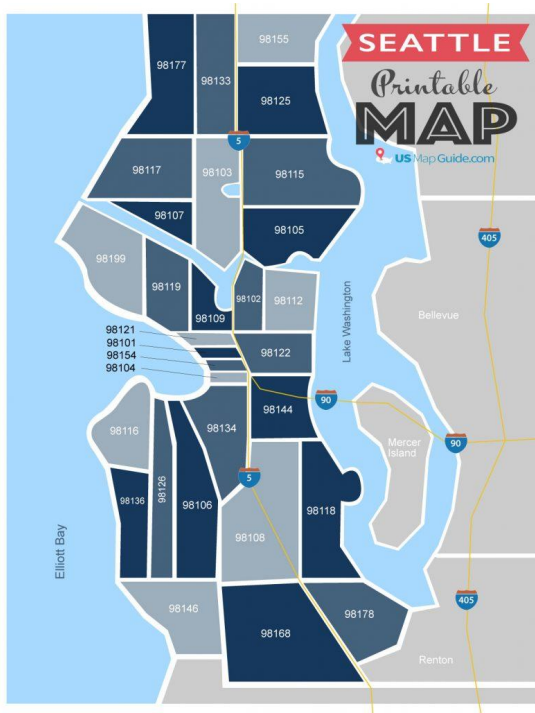
**King County Housing  
Consultants  
(Market overview &  
price prediction)**

**Ravi Dahiya & Marc Inizan**

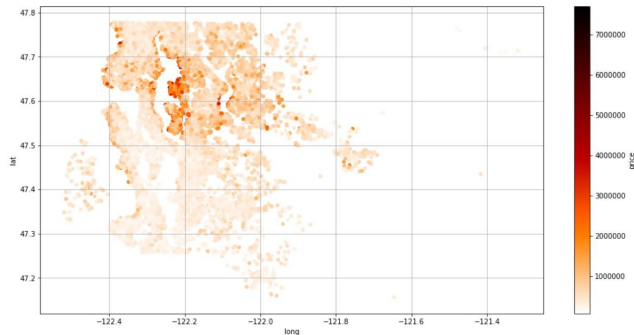
# Objectives

- \* Analyze the geographical data to provide in-depth information about the general market scenario.
- \* Find out the features that have the highest impact on the price of the house.
- \* Predict the price of a house based on its most important features.

# The market

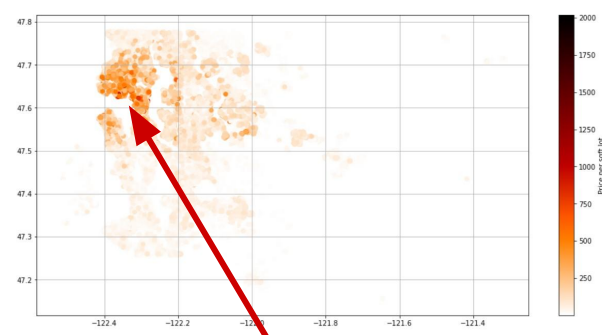


Price as per location



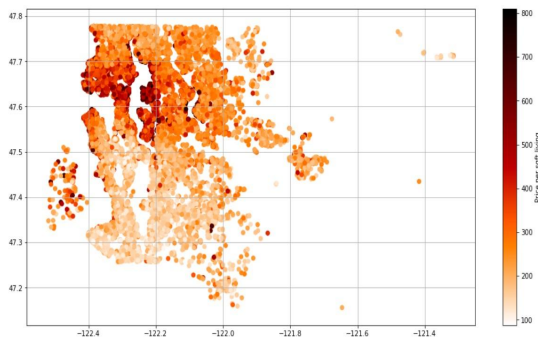
\*\*As we can see that the prices are highest around (-47.7, -122.2)\*\*

Price as per sqft lot as per location



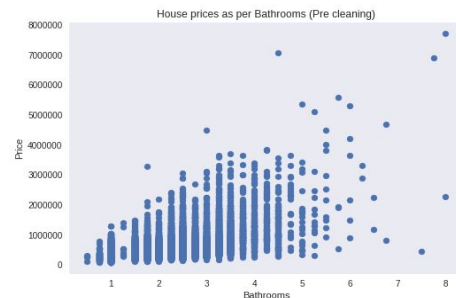
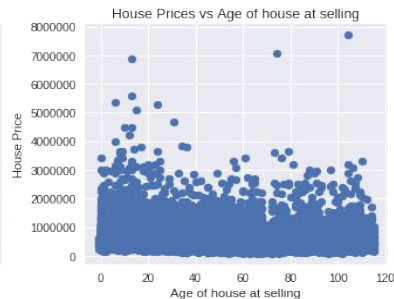
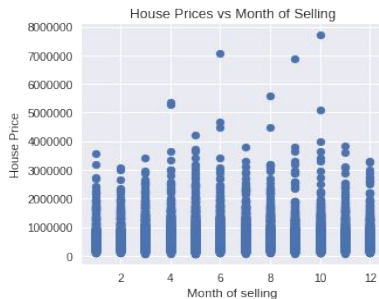
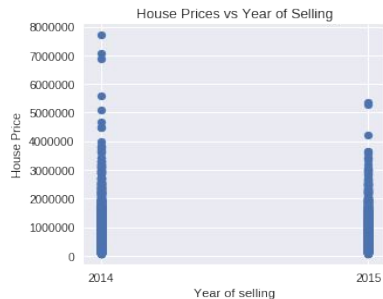
\*\*Areas in the darker shades will attract higher prices per sqft of land\*\*

Price per sqft living as per location



As we can see that even though the houses with the highest prices are scattered over a large area, the per sqft lot price is highest in the North West of Seattle's King County

# Data Cleaning & Data Transformation



Merging the Columns

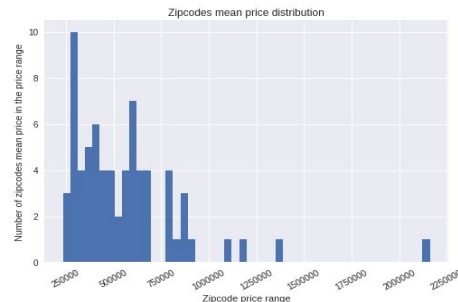
Binning



Log Transformation

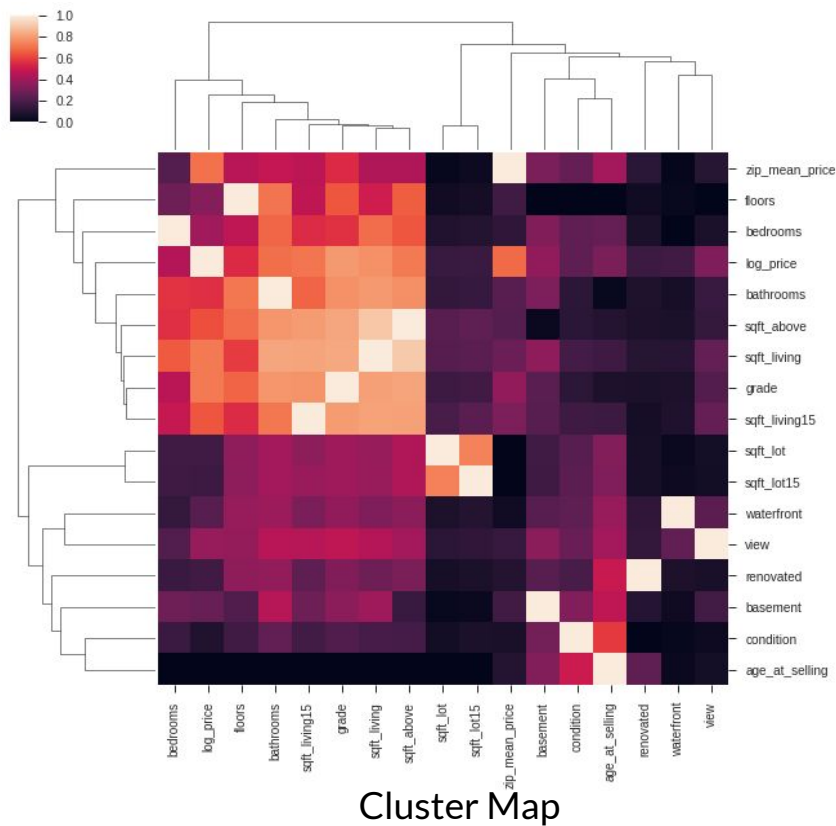


Yes/No



Zipcode Bins

# Feature Selection



Top 5 features correlated with Log Price:

- Grade
- Square footage of the living area
- Zipcode
- Square footage of the living area of the 15 closest houses
- Square footage of house apart from basement

Those last 2 features are highly correlated with the square footage of the living area.

# Prediction Model

<b>Dep. Variable:</b>	log_price	<b>R-squared:</b>	0.776
<b>Model:</b>	OLS	<b>Adj. R-squared:</b>	0.776
<b>Method:</b>	Least Squares	<b>F-statistic:</b>	2.491e+04
<b>Date:</b>	Tue, 03 Dec 2019	<b>Prob (F-statistic):</b>	0.00
<b>Time:</b>	12:55:04	<b>Log-Likelihood:</b>	-640.13
<b>No. Observations:</b>	21594	<b>AIC:</b>	1288.
<b>Df Residuals:</b>	21590	<b>BIC:</b>	1320.
<b>Df Model:</b>	3		
<b>Covariance Type:</b>	nonrobust		
	<b>coef</b>	<b>std err</b>	<b>t</b> <b>P&gt; t </b> <b>[0.025</b> <b>0.975]</b>
<b>const</b>	11.8141	0.005	2415.595 0.000 11.804 11.824
<b>grade</b>	0.1061	0.002	44.570 0.000 0.101 0.111
<b>sqft_living</b>	0.0002	2.79e-06	83.541 0.000 0.000 0.000
<b>zip_mean_price</b>	0.3027	0.002	145.806 0.000 0.299 0.307
<b>Omnibus:</b>	707.520	<b>Durbin-Watson:</b>	1.975
<b>Prob(Omnibus):</b>	0.000	<b>Jarque-Bera (JB):</b>	1588.247
<b>Skew:</b>	0.190	<b>Prob(JB):</b>	0.00
<b>Kurtosis:</b>	4.273	<b>Cond. No.</b>	6.69e+03

With 3 featured: R-squared value of 0.776, which means that 77.6 percent variance in the dependent variable can be explained by independent variables.

## Model in the working

```
1 predict(1200, 98117, 7)
```

404865.0

A 1200 square feet house located in the 98117 zipcode with a grade of 7 would be worth 405K\$

# Recommendations

1. Should include features such as custom design, high quality cabinet work, wood trim ,marble, bigger entries
2. As the size of lot has small impact on the price, focus should be on high living area. Also, basement doesn't add much value to the house.
3. Houses in zipcodes with high average prices will be priced higher