

# Flatiron Project 1

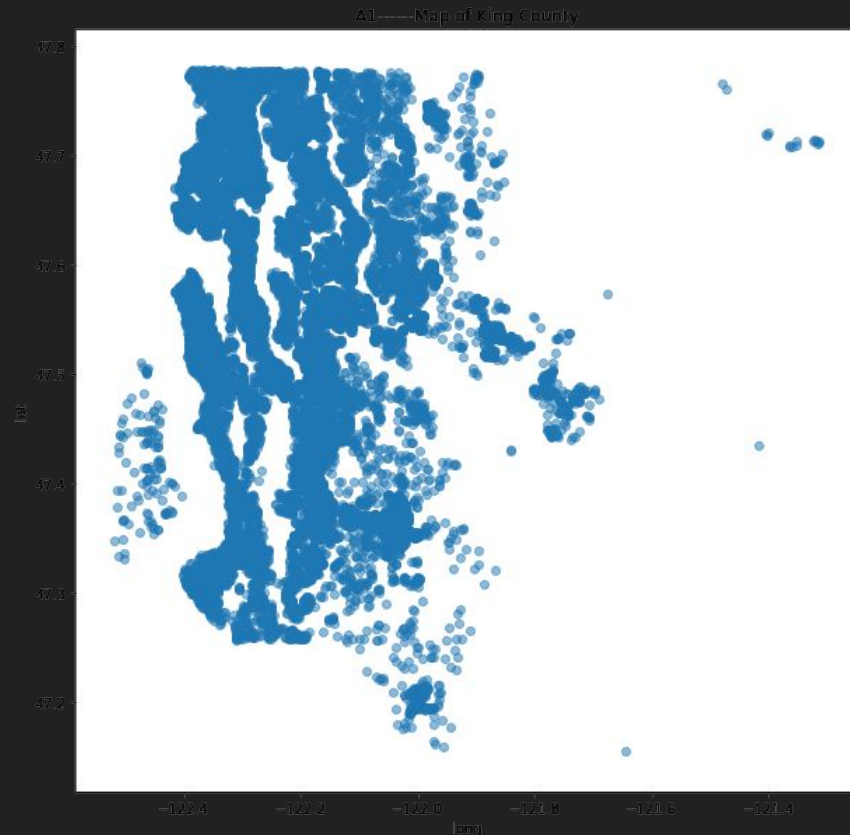
King County House Prices

# Data Overview

Data contains features such as:  
#bedrooms,#bathrooms,sqftage,and  
zipcode

Goal is to use these features to predict the  
target variable price

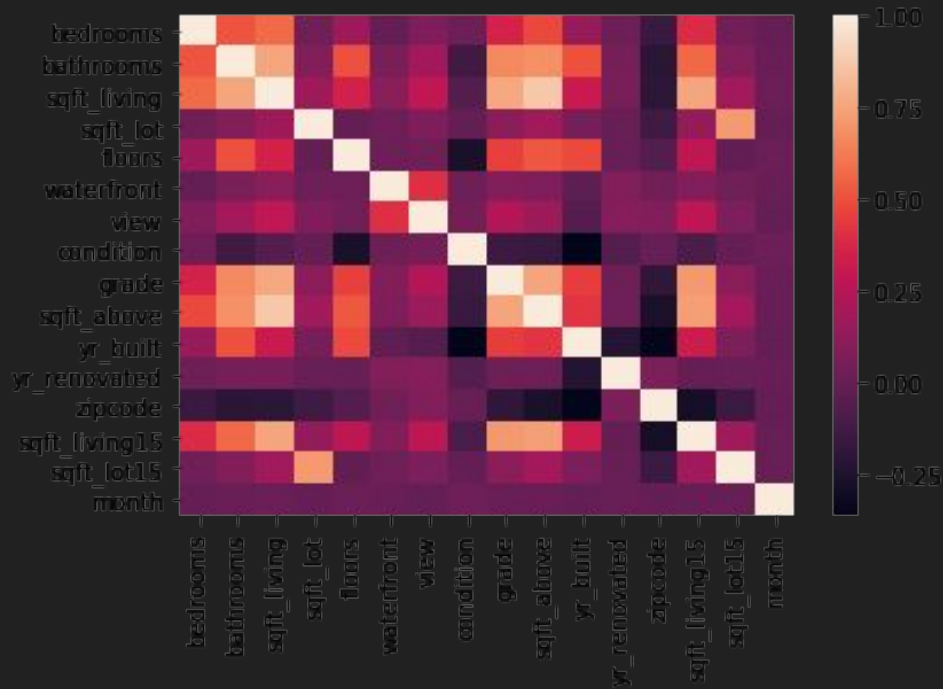
Methodology will be to use a stepwise  
selection algorithm



# Map of King County

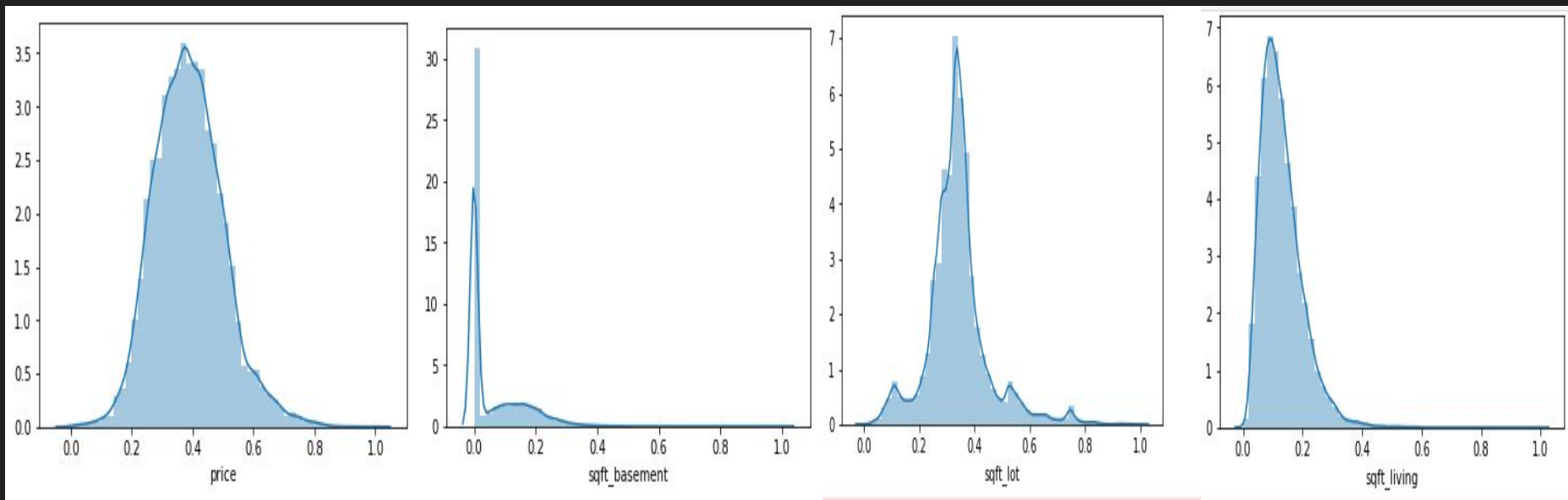
Several predictors are not useful:

latitude, longitude, #bathrooms, and neighbors square footage are all dropped.



# Data Preprocessing

Most of the features are re-formatted as categories. The remaining features (sqft lot, sqft basement and sqft living) are normalized to be between 0 and 1. The same normalization is applied to the price target.



# Stepwise selection

The features are selected to be included as long as their p-value is less than .01, and will be removed if it goes over .05. The r-squared value of .87 is very strong.

|                          |                  |                            |            |
|--------------------------|------------------|----------------------------|------------|
| <b>Dep. Variable:</b>    | price            | <b>R-squared:</b>          | 0.869      |
| <b>Model:</b>            | OLS              | <b>Adj. R-squared:</b>     | 0.868      |
| <b>Method:</b>           | Least Squares    | <b>F-statistic:</b>        | 1427.      |
| <b>Date:</b>             | Wed, 06 Nov 2019 | <b>Prob (F-statistic):</b> | 0.00       |
| <b>Time:</b>             | 17:38:40         | <b>Log-Likelihood:</b>     | 38082.     |
| <b>No. Observations:</b> | 21597            | <b>AIC:</b>                | -7.596e+04 |
| <b>Df Residuals:</b>     | 21496            | <b>BIC:</b>                | -7.516e+04 |
| <b>Df Model:</b>         | 100              |                            |            |
| <b>Covariance Type:</b>  | nonrobust        |                            |            |

# Most vital parameters

Sqft living is the most important parameter with a coefficient of .8651. The rest are particular zipcodes.

|               |          |
|---------------|----------|
| zipcode_98103 | 0.201662 |
| zipcode_98122 | 0.204733 |
| zipcode_98107 | 0.208007 |
| zipcode_98199 | 0.215560 |
| zipcode_98040 | 0.219342 |
| zipcode_98105 | 0.235598 |
| zipcode_98119 | 0.244725 |
| zipcode_98102 | 0.247027 |
| zipcode_98109 | 0.247763 |
| zipcode_98112 | 0.264377 |
| zipcode_98004 | 0.267209 |
| zipcode_98039 | 0.297640 |
| sqft_living   | 0.855234 |