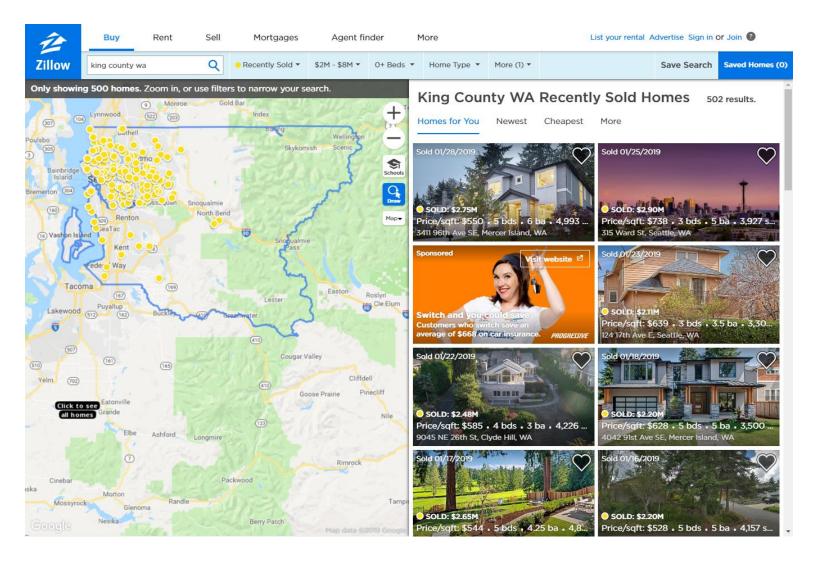
King County House Dataset

A Brief presentation for the analysis of the king count data set.

Feature Selection

- There where total 21 field in the dataset including price.
- Base on the R-squared values
- Made a filter of price up to 2 mil

Reason to select the range up to 2Mill

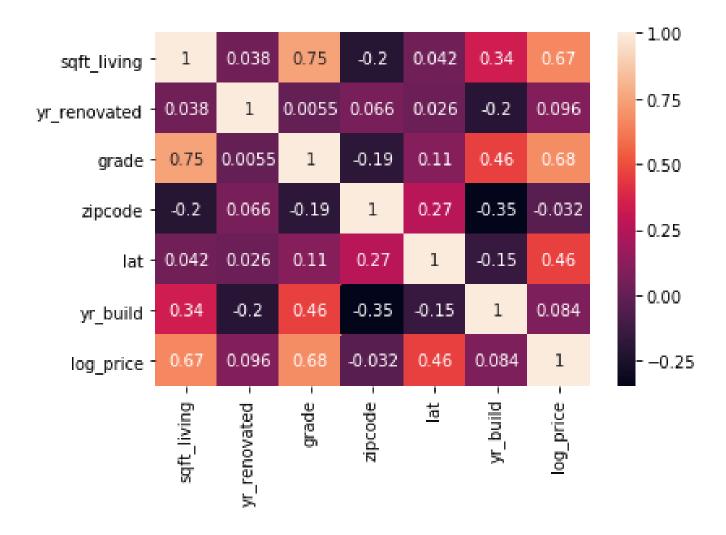


R Square value

OLS Regression Results

Dep. Variable:	log_price	R-squared:	0.716
Model:	OLS	Adj. R-squared:	0.716
Method:	Least Squares	F-statistic:	8974.
Date:	Wed, 30 Jan 2019	Prob (F-statistic):	0.00
Time:	10:06:00	Log-Likelihood:	-2096.7
No. Observations:	21399	AIC:	4207.
Df Residuals:	21392	BIC:	4263.
Df Model:	6		
Covariance Type:	nonrobust		

Heatmap



Train-test-Split

- Test # 1
 - Removing sqft_living
 - Trained MES
 - 0.08698825027764386
 - Test MES
 - 0.0903335895730355

- Test # 2
 - Removing grade
 - Trained MES
 - 0.08698825027764386
 - Test MES
 - 0.0903335895730355

Current Model Accuracy

• Base on the Current test and train MES this model has currently 71%.