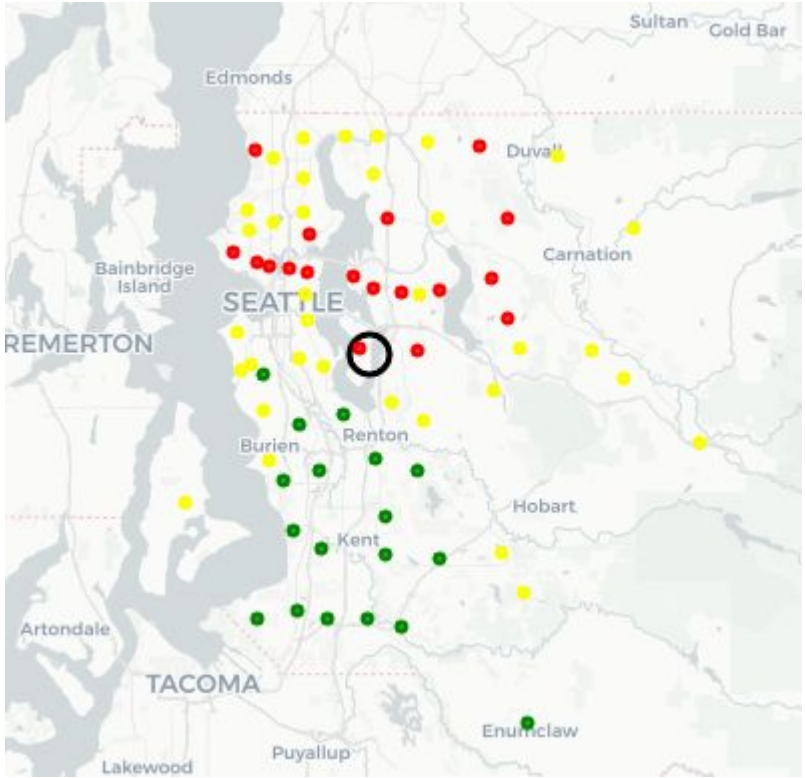
Data Housing Project

Mod 1 Project

1. Data Cleaning

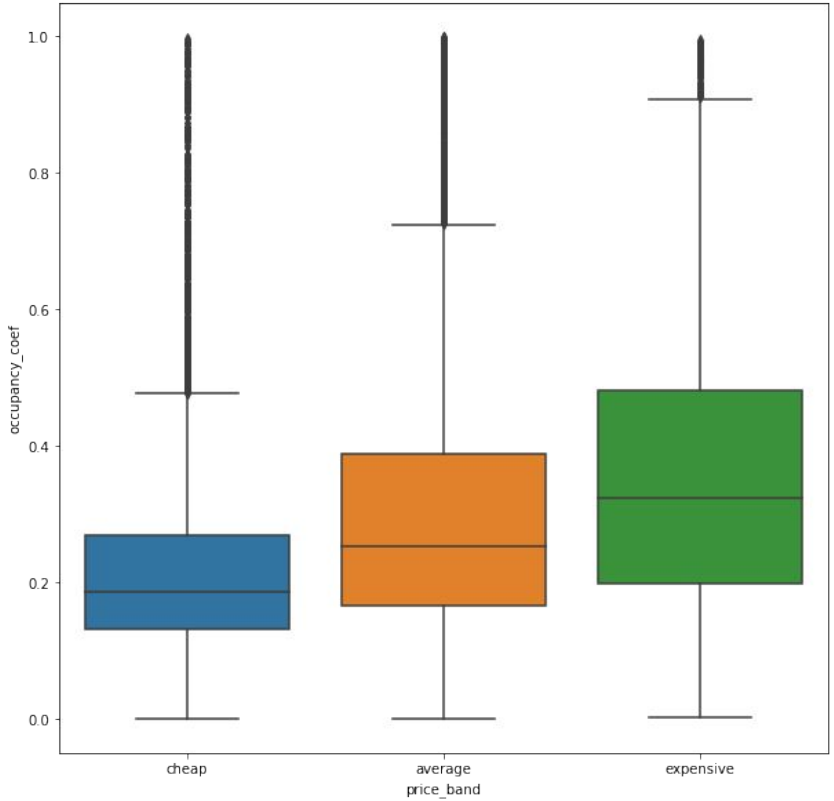
a. Data import
b. Checking data types
c. Resolving missing values
d. Removing outliers

2. Exploratory Data Analysis



a. How does location have an impact on price?

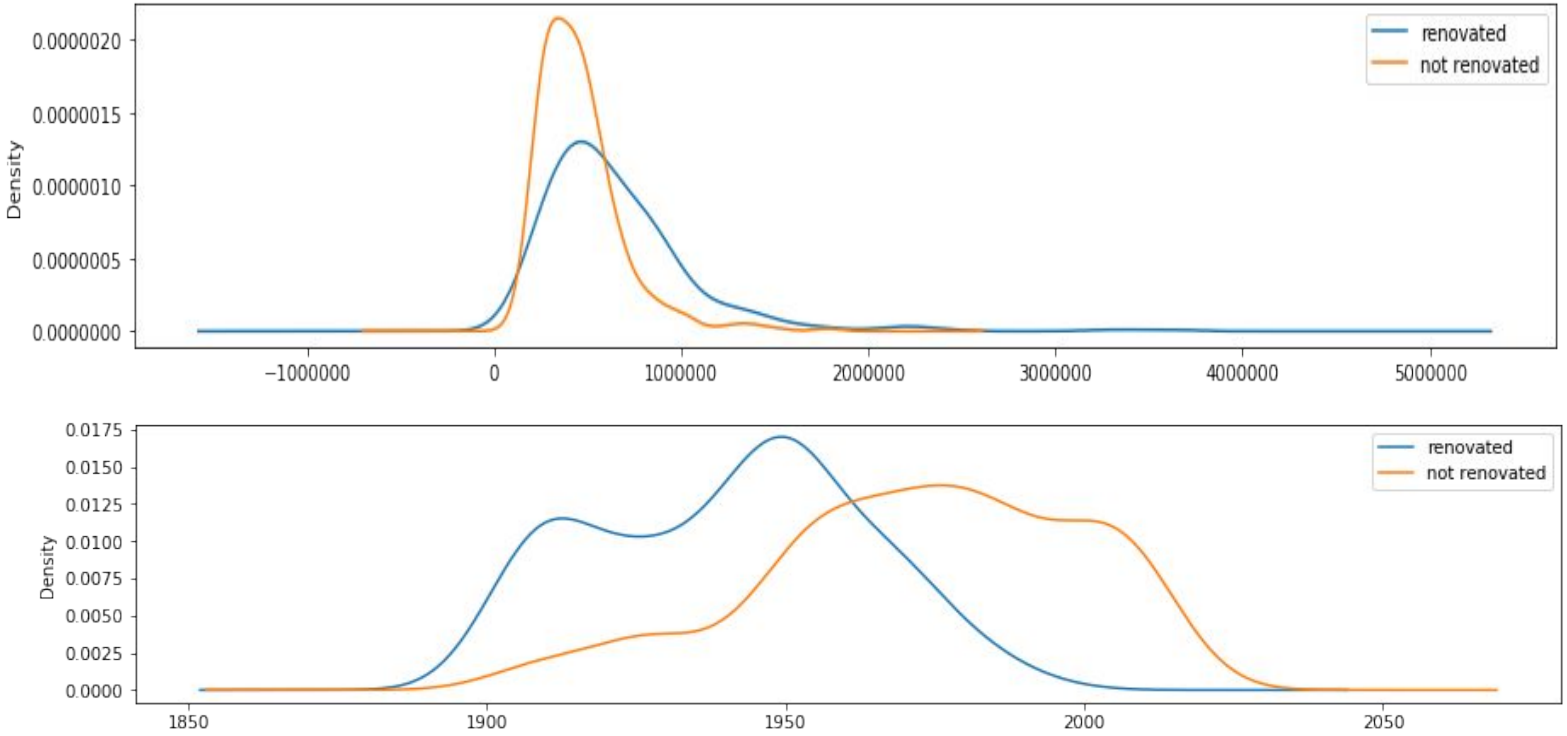
2. Exploratory Data Analysis



b. How does living density have an impact on price?

2. Exploratory Data Analysis

c. How does renovation have an impact on price?



3. Modeling

	correlation
price	1.000000
sqft_living	0.701554
grade	0.668262
sqft_above	0.605510
sqft_living15	0.585597
bathrooms	0.524823
view	0.395640
sqft_basement	0.319199
bedrooms	0.315193
lat	0.308032

a. Adding predictors to the model

3. Modeling

	correlation
sqft_living	1.000000
sqft_above	0.945416
sqft_living_per_bed	0.786544
grade	0.782506
bathrooms	0.777182
sqft_living15	0.772939

b. Checking for collinearity between predictors

3. Modeling

OLS Regression Results

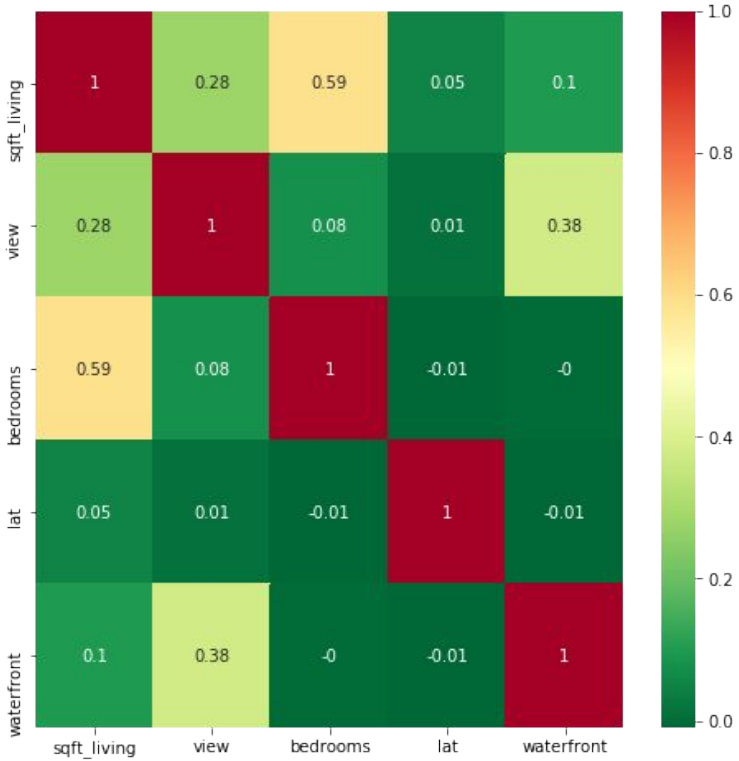
Dep. Variable:	price	**R-squared (uncentered):**	0.839
Model:	OLS	**Adj. R-squared (uncentered):**	0.839
Method:	Least Squares	**F-statistic:**	1.126e+05
Date:	Tue, 22 Oct 2019	**Prob (F-statistic):**	0.00
Time:	09:56:38	**Log-Likelihood:**	-2.9910e+05
No. Observations:	21529	**AIC:**	5.982e+05
Df Residuals:	21528	**BIC:**	5.982e+05
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
sqft_living	262.9891	0.784	335.489	0.000	261.453	264.526

Omnibus:	15748.539	**Durbin-Watson:**	1.979
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	650098.249
Skew:	3.075	**Prob(JB):**	0.00
Kurtosis:	29.209	**Cond. No.**	1.00

c. Evaluating the coefficients

3. Modeling



d. Reducing collinearity using feature engineering

4. Possible extensions

a. Using feature scaling to scale model coefficients

b. Splitting the data into training and test sets

c. Checking for normality of predictors