



Seattle

Real Estate Analytics

SEATTLERS BANK

Executive Summary

- ▶ The objective of the analysis conducted is to gain an alternative understanding of the Seattle property market for the development of our real estate financing business. This is an experimental in-house analysis.
- ▶ Aim to complement the traditional micro level evaluation for real estate financing
- ▶ Focusing on following questions using regression analysis:
 - ▶ ***Will price increase provided we keep the living surface constant and increase the grading by 1?***
 - ▶ ***Will price increase provided we increase living surface and keep the grading constant?***
 - ▶ ***How do the models compare if we proceed with the removal of outlier values in our variables?***

Process Timeline

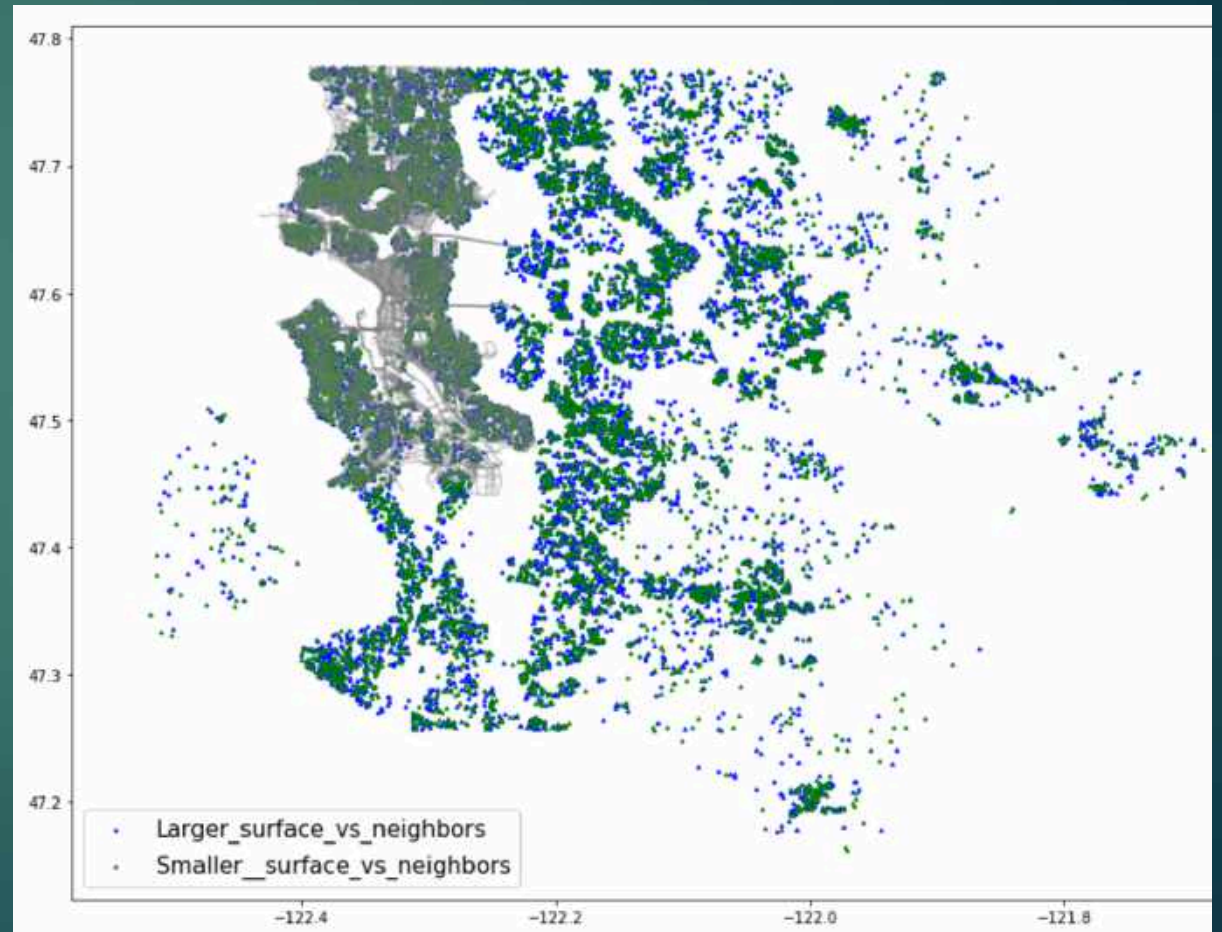
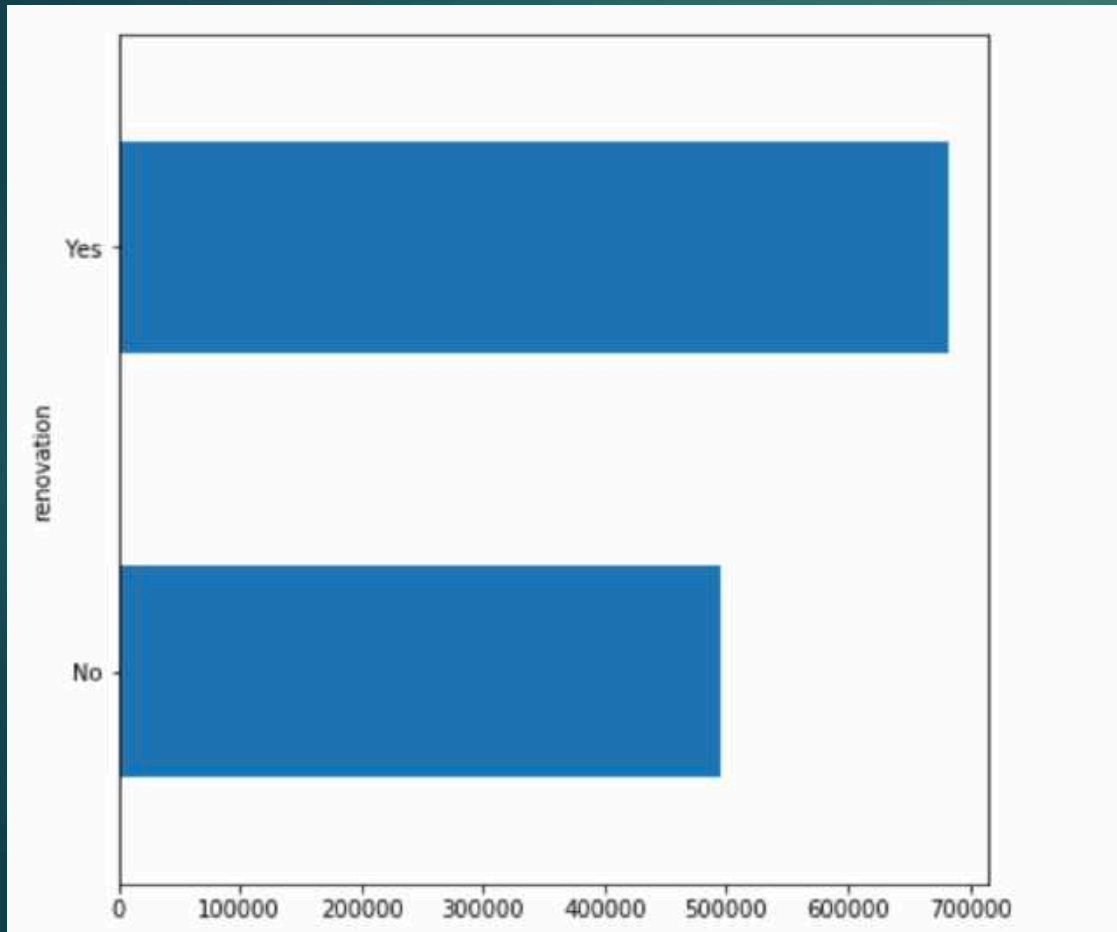
- ▶ Step 1: Load the dataset and identify null values
- ▶ Step 2: Treatment of null values and special characters
- ▶ Step 3: Exploratory Data Analysis / Data Observations
- ▶ Step 4: Graph relationship between all relevant variables and price
- ▶ Step 5: Model Approach
- ▶ Step 6: Choose variables for regression models
- ▶ Step 7: Run regression model
- ▶ Step 8: interpretation of models
- ▶ Step 9: follows ups, correction and questions

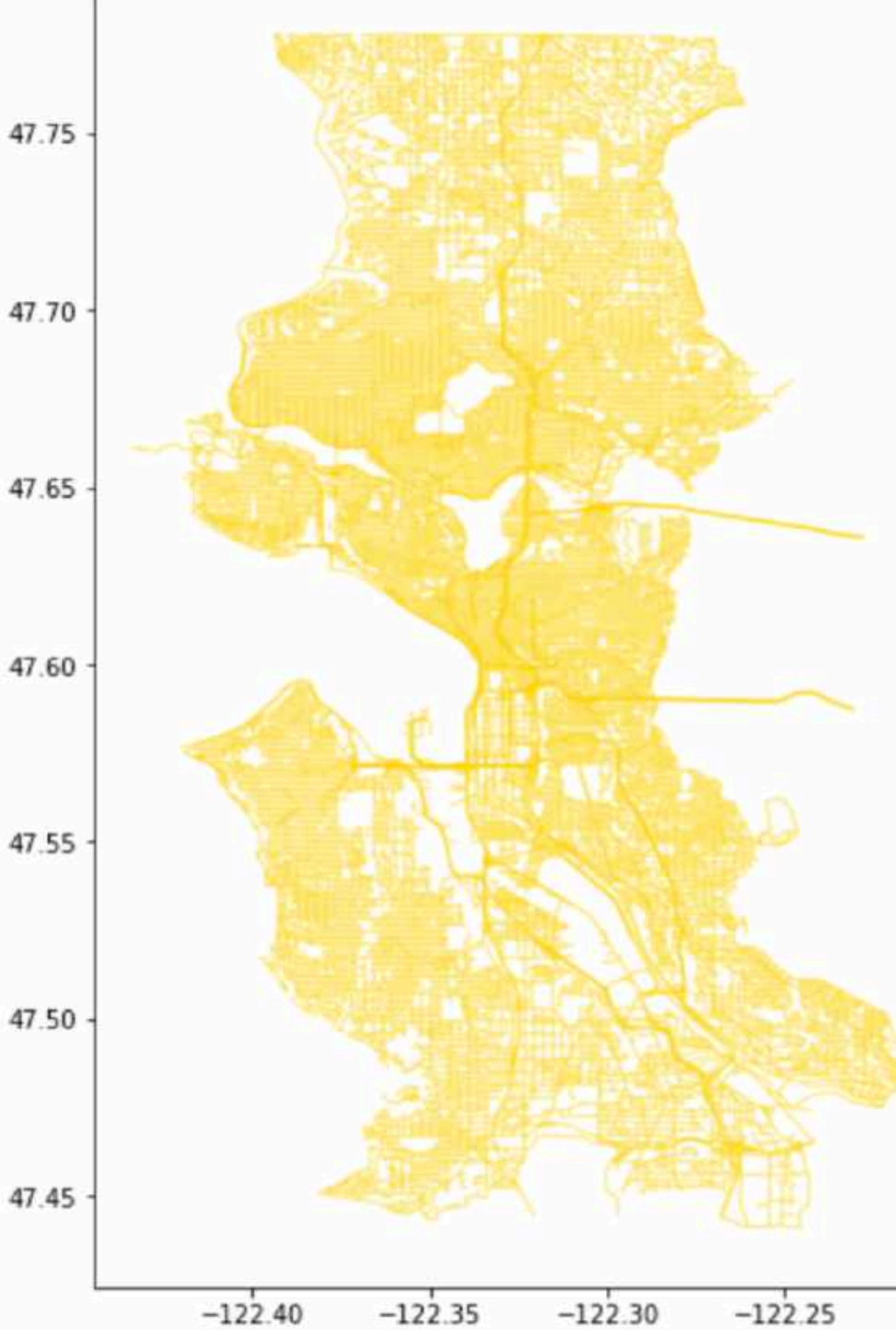
Step 1 & 2: Load the dataset + identify & treat null values

- ▶ Pandas data frame was used to load and gain an understanding of the Seattle housing market dataset
- ▶ Three approaches to treat null values in *waterfront*, *view*, *year renovated*
- ▶ Column removal to simplify set
- ▶ Treatment of nulls varied on the different columns

Step 3: Data Observations

Here are some of the observations and visual representations of the data we made during the EDA phase

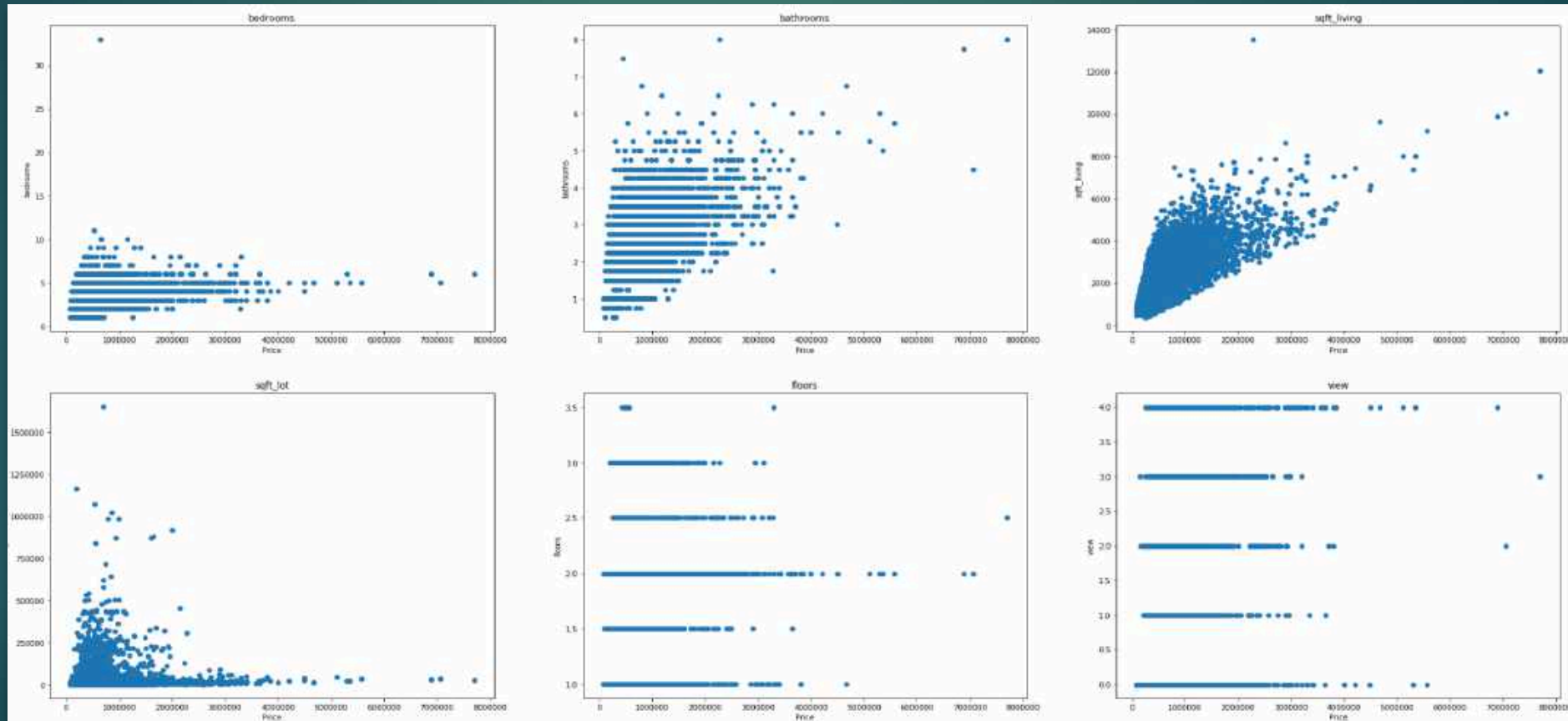




Step 3: Data Observations (continued)

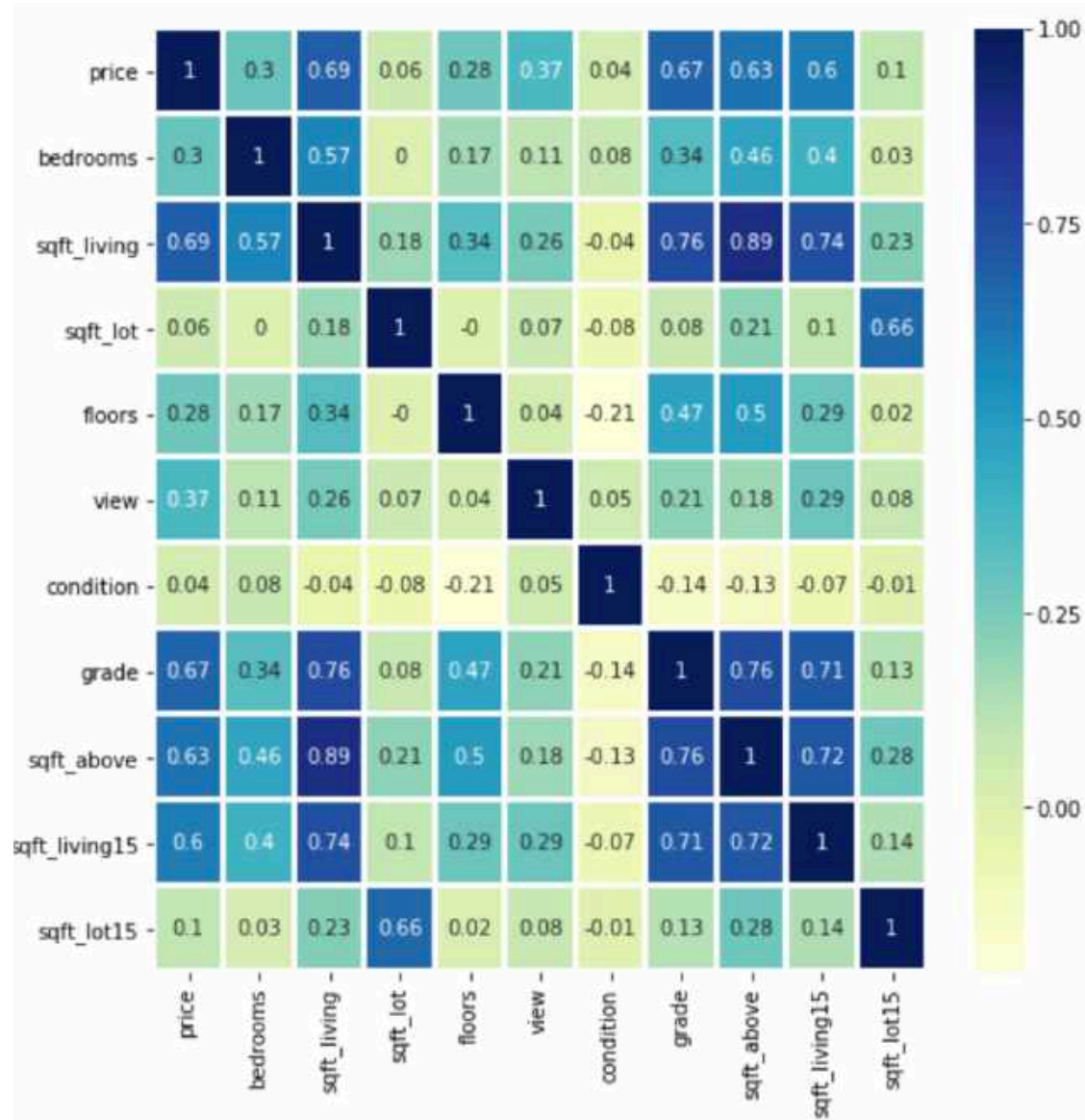
Step 4: Graph relationship between all relevant variables and price

- We tested most variables to establish their relationship with price and identify the relevant ones for model testing



Step 5: Model Approach

- ▶ To answer our questions - two models with slight alterations
- ▶ Model 1: Run OLS model on price and two independent variables without removing data outliers before selecting our sample
- ▶ Model 2: Run OLS model on price and two independent variables with outliers removed using standard deviation before selecting our sample
- ▶ Used Sampling approach of 1000 houses with <15 houses selected at random per zip code.



Step 6: Build correlations table and matrix to get an overview variables & elect variables

A CORRELATION TABLE AND MATRIX TO DETERMINE WHICH VARIABLES WE WOULD ELECT.

ELECTED SQFT_LIVING / GRADE AS THE TWO VARIABLES TO TEST.

Step 7: Run regression model and interpret the results

Model 1 – Outliers Present in the Data

$$P = -710,300 + 185 (\text{sqft} = +1) + 146,000 (\text{grade} = +1)$$

Model 2 – Outliers Removed from the Data

$$P = -661,000 + 134 (\text{sqft} = +1) + 120,700 (\text{grade} = +1)$$

Dep. Variable:	price	R-squared:	0.535	
Model:	OLS	Adj. R-squared:	0.534	
Method:	Least Squares	F-statistic:	601.3	
Date:	Tue, 21 Jan 2020	Prob (F-statistic):	1.29e-174	
Time:	12:49:32	Log-Likelihood:	-14615.	
No. Observations:	1050	AIC:	2.924e+04	
Df Residuals:	1047	BIC:	2.925e+04	
Df Model:	2			
Covariance Type:	nonrobust			
	coef	std err	t P> t [0.025 0.975]	
const	-7.103e+05	6.31e+04	-11.252 0.000	-8.34e+05 -5.86e+05
sqft_living	185.0059	13.825	13.382 0.000	157.877 212.134
grade	1.146e+05	1.06e+04	10.796 0.000	9.38e+04 1.35e+05
Omnibus:	585.573	Durbin-Watson:	1.073	
Prob(Omnibus):	0.000	Jarque-Bera (JB):	6786.525	
Skew:	2.329	Prob(JB):	0.00	
Kurtosis:	14.551	Cond. No.	1.74e+04	

Dep. Variable:	price	R-squared:	0.414	
Model:	OLS	Adj. R-squared:	0.413	
Method:	Least Squares	F-statistic:	369.6	
Date:	Tue, 21 Jan 2020	Prob (F-statistic):	3.51e-122	
Time:	12:49:44	Log-Likelihood:	-14385.	
No. Observations:	1050	AIC:	2.878e+04	
Df Residuals:	1047	BIC:	2.879e+04	
Df Model:	2			
Covariance Type:	nonrobust			
	coef	std err	t P> t [0.025 0.975]	
const	-6.618e+05	6.44e+04	-10.275 0.000	-7.88e+05 -5.35e+05
sqft_living	134.3940	12.342	10.889 0.000	110.177 158.611
grade	1.207e+05	1.03e+04	11.672 0.000	1e+05 1.41e+05
Omnibus:	501.034	Durbin-Watson:	0.904	
Prob(Omnibus):	0.000	Jarque-Bera (JB):	3650.157	
Skew:	2.066	Prob(JB):	0.00	
Kurtosis:	11.146	Cond. No.	2.05e+04	

Step 8: interpretation, follows ups and correction of model

- ▶ The two models provide us limited conclusions to be drawn.
- ▶ Price does increase with both square foot living and grading but the models don't have the best linear fit.
- ▶ R-Squared is still relatively low and we may have an issue of multicollinearity as both variables are correlated (0.7)
- ▶ Outlier removal causes loss of model precision

Step 9: Future work, correction and questions

- ▶ Review variable choices – Grade / Square foot living
- ▶ Apply correct treatment to categorical variable
- ▶ Develop more specific business questions

Thank you for listening! We will do our best to answer any questions you have.