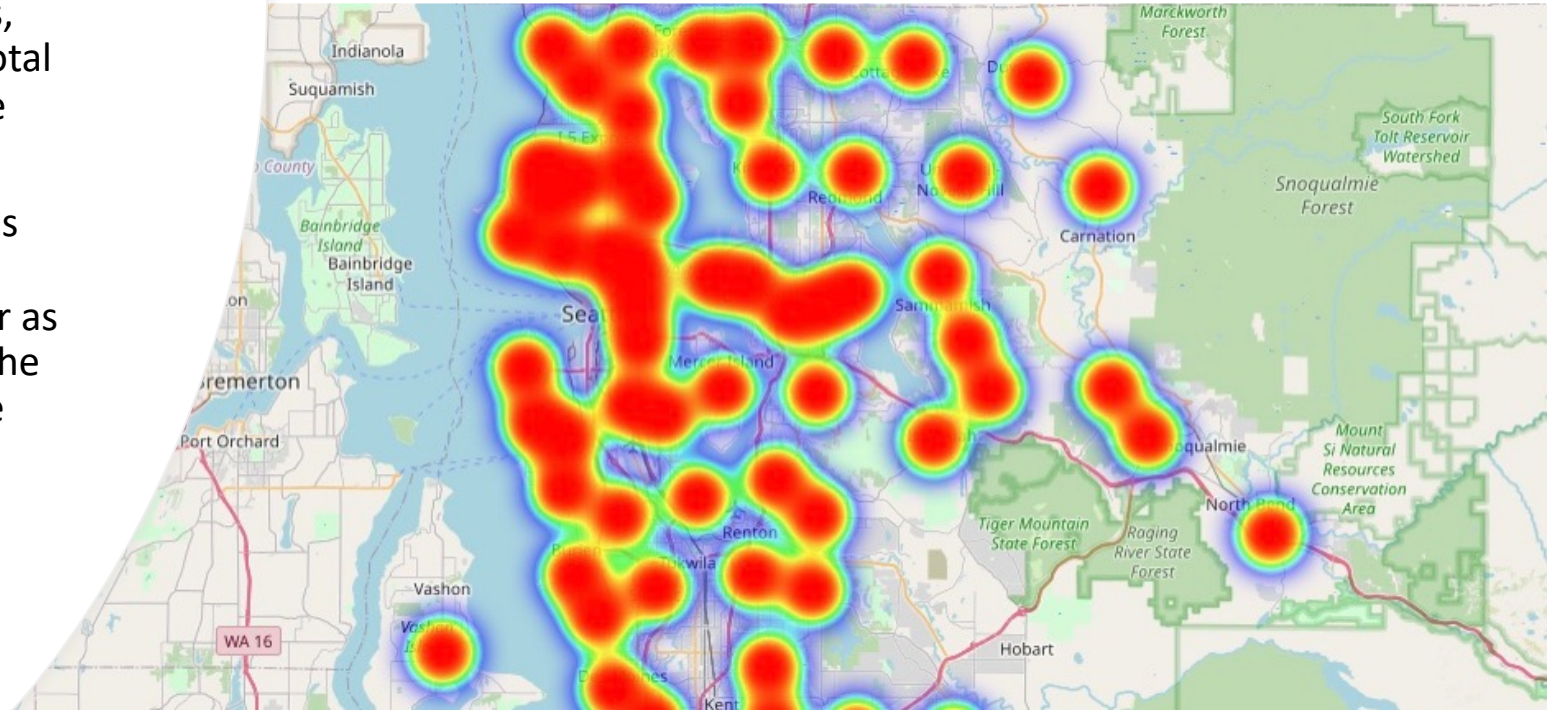
King County Realty Investment

RockBlack Investment Management Co.

Summary

- The aim of this experiment in linear regression is to assess the fluctuation of prices of homes in King County, Seattle for a private equity firm focusing on real estate acquisitions. The manipulation of the homes' variables such as square feet, waterfronts, and other architectural features affect the over total worth of the home and it is the intention of these linear models to explain the value of the home.
- The data from collected for the experiment ranges from 2014 to 2015 and it is important to acknowledge the limitations of that specific factor as home prices were in a steep heel which marked the 2010's housing market from its recovery from the previous decade.



Outline

Business Problem

Linear Models

Results

Conclusion

Future Work

Business Problem

Increase of Price Per Square Feet

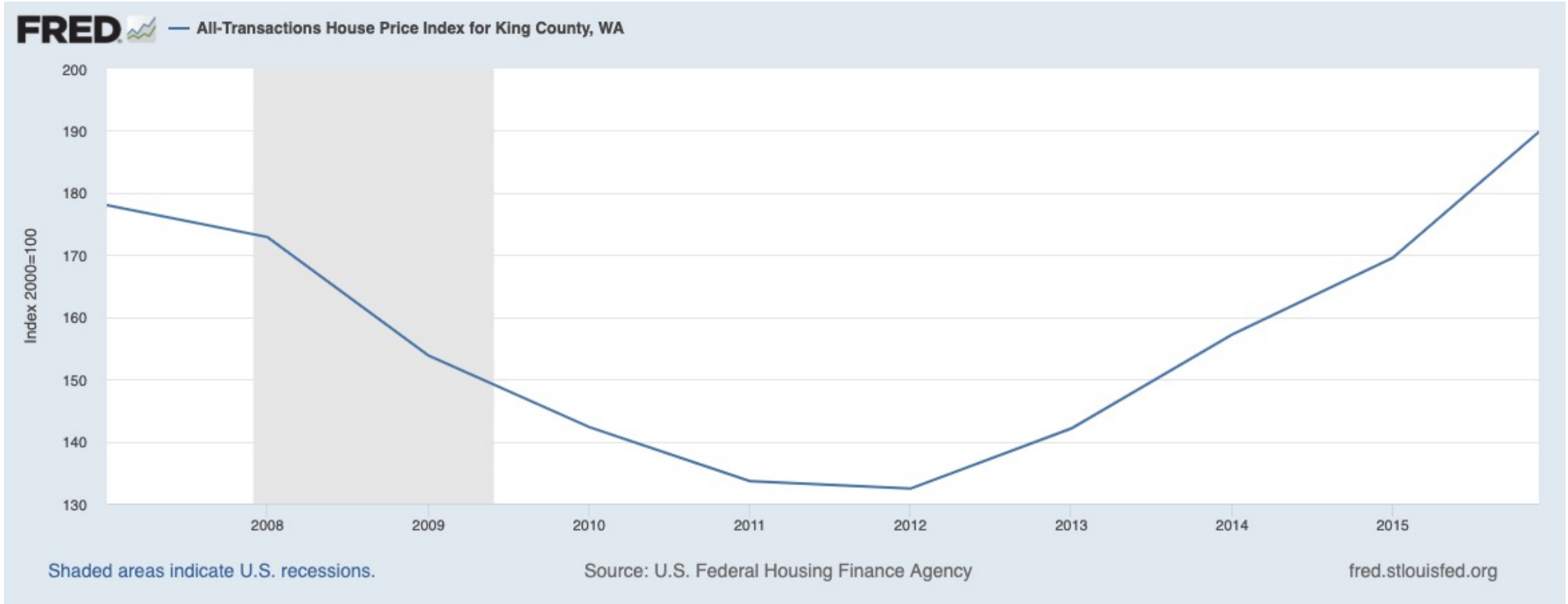
- Limited to King County, Seattle, WA

Increase of Price Per Bedroom

- Niche Market
  - Homes under 2M
  - Homes up to six bedrooms

Business Problem

- The aim of this linear regression model is to assess the fluctuation of prices of homes in King County, Seattle for the real estate acquisitions division at RockBlack. The manipulation of the homes' variables such as square feet, waterfronts, and other architectural features affect the over total worth of the home and it is the intention of these models to interpret the value of homes under 2M USD and six bedrooms.
- The data from collected for the experiment ranges from homes sold from 2014 to 2015 and it is important to acknowledge the limitations of this specific housing market as home prices were in a steep heel which marked the 2010's housing market from its recovery from the previous decade.



Data & Methods

The data used in this linear regression is housing data from Kings' County, Seattle, Washington. The data included the following columns; these are their names and descriptions:

- **id** - unique identified for a house
- **dateDate** - house was sold
- **pricePrice** - is prediction target
- **bedroomsNumber** - of Bedrooms/House
- **bathroomsNumber** - of bathrooms/bedrooms
- **sqft_livingsquare** - footage of the home
- **sqft_lotsquare** - footage of the lot
- **floorsTotal** - floors (levels) in house
- **waterfront** - House which has a view to a waterfront
- **view** - Has been viewed
- **condition** - How good the condition is (Overall)
- **grade** - overall grade given to the housing unit, based on King County grading system
- **sqft_above** - square footage of house apart from basement
- **sqft_basement** - square footage of the basement
- **yr_built** - Built Year
- **yr_renovated** - Year when house was renovated
- **zipcode** - zip
- **lat** - Latitude coordinate
- **long** - Longitude coordinate
- **sqft_living15** - The square footage of interior housing living space for the nearest 15 neighbors
- **sqft_lot15** - The square footage of the land lots of the nearest 15 neighbors

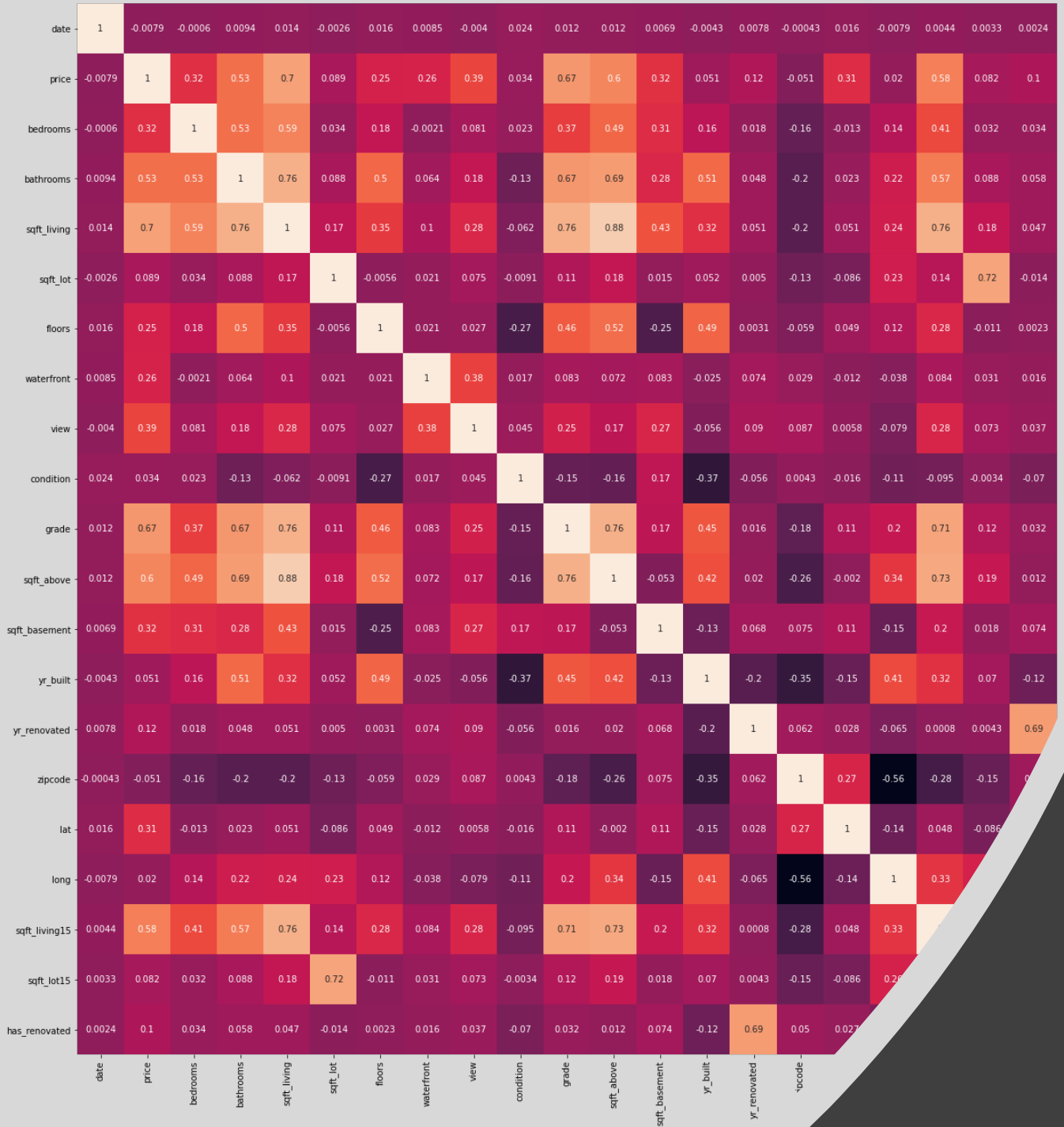
Of the columns provided the following where use for all three models:

- Continious Variables:
  - sqft_living, sqft_lot,sqft_basement,lat,long
- Categorical Variables:
  - bedroom,bathroom,floor,waterfront,view,condition,grade,yr_built,renovated

The baseline model was carried out using raw data

The subsequent model used a logarithmic transformation of the price of the homes to make statistical predictions more valid

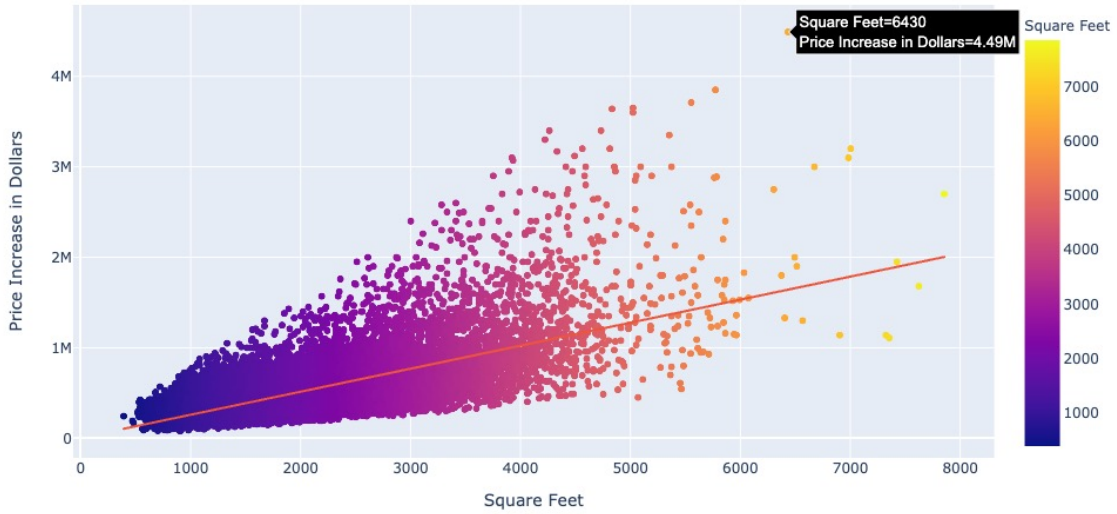
The third and final model focused on homes with a price under two-million dollars and under six bedrooms honning in on the niche market targeted by the private equity firm

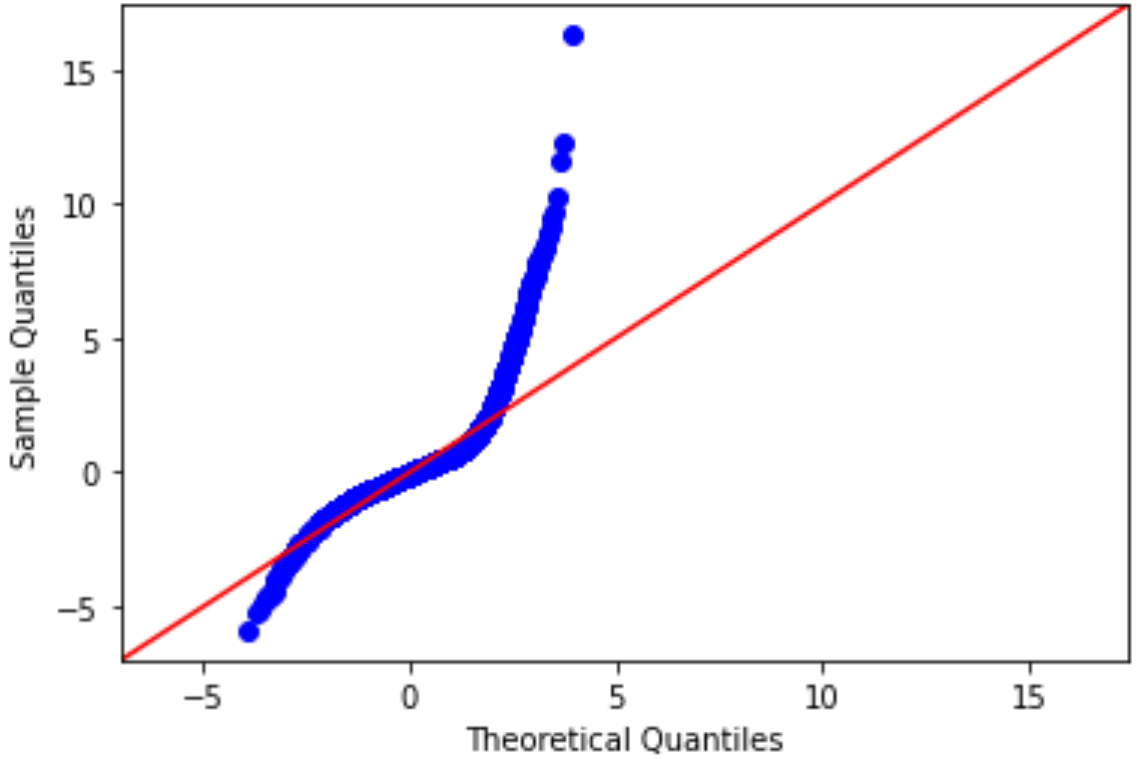

date 1 -0.0079 -0.0006 0.0094 0.014 -0.0026 0.016 0.0085 -0.004 0.024 0.012 0.012 0.0069 -0.0043 0.0078 -0.00043 0.016 -0.0079 0.0044 0.0033 0.0024
price -0.0079 1 0.32 0.53 0.7 0.089 0.25 0.26 0.39 0.034 0.67 0.6 0.32 0.051 0.12 -0.051 0.31 0.02 0.58 0.082 0.1
bedrooms -0.0006 0.32 1 0.53 0.59 0.034 0.18 -0.0021 0.081 0.023 0.37 0.49 0.31 0.16 0.018 -0.16 -0.013 0.14 0.41 0.032 0.034
bathrooms 0.0094 0.53 0.53 1 0.76 0.088 0.5 0.064 0.18 -0.13 0.67 0.69 0.28 0.51 0.048 -0.2 0.023 0.22 0.57 0.088 0.058
sqft_living 0.014 0.7 0.59 0.76 1 0.17 0.35 0.1 0.28 -0.062 0.76 0.88 0.43 0.32 0.051 -0.2 0.051 0.24 0.76 0.18 0.047
sqft_lot -0.0026 0.089 0.034 0.088 0.17 1 -0.0056 0.021 0.075 -0.0091 0.11 0.18 0.015 0.052 0.005 -0.13 -0.086 0.23 0.14 0.72 -0.014
floors 0.016 0.25 0.18 0.5 0.35 -0.0056 1 0.021 0.027 -0.27 0.46 0.52 -0.25 0.49 0.0031 -0.059 0.049 0.12 0.28 -0.011 0.0023
waterfront 0.0085 0.26 -0.0021 0.064 0.1 0.021 0.021 1 0.38 0.017 0.083 0.072 0.083 -0.025 0.074 0.029 -0.012 -0.038 0.084 0.031 0.016
view -0.004 0.39 0.081 0.18 0.28 0.075 0.027 0.38 1 0.045 0.25 0.17 0.27 -0.056 0.09 0.087 0.0058 -0.079 0.28 0.073 0.037
condition 0.024 0.034 0.023 -0.13 -0.062 -0.0091 -0.27 0.017 0.045 1 -0.15 -0.16 0.17 -0.37 -0.056 0.0043 -0.016 -0.11 -0.095 -0.0034 -0.07
grade 0.012 0.67 0.37 0.67 0.76 0.11 0.46 0.083 0.25 -0.15 1 0.76 0.17 0.45 0.016 -0.18 0.11 0.2 0.71 0.12 0.032
sqft_above 0.012 0.6 0.49 0.69 0.88 0.18 0.52 0.072 0.17 -0.16 0.76 1 -0.053 0.42 0.02 -0.26 -0.002 0.34 0.73 0.19 0.012
sqft_basement 0.0069 0.32 0.31 0.28 0.43 0.015 -0.25 0.083 0.27 0.17 0.17 -0.053 1 -0.13 0.068 0.075 0.11 -0.15 0.2 0.018 0.074
yr_built -0.0043 0.051 0.16 0.51 0.32 0.052 0.49 -0.025 -0.056 -0.37 0.45 0.42 -0.13 1 -0.2 -0.35 -0.15 0.41 0.32 0.07 -0.12
yr_renovated 0.0078 0.12 0.018 0.048 0.051 0.005 0.0031 0.074 0.09 -0.056 0.016 0.02 0.068 -0.2 1 0.062 0.028 -0.065 0.0008 0.0043 0.69
zipcode -0.00043 -0.051 -0.16 -0.2 -0.2 -0.13 -0.059 0.029 0.087 0.0043 -0.18 -0.26 0.075 -0.35 0.062 1 0.27 -0.56 -0.28 -0.15
lat 0.016 0.31 -0.013 0.023 0.051 -0.086 0.049 -0.012 0.0058 -0.016 0.11 -0.002 0.11 -0.15 0.028 0.27 1 -0.14 0.048 -0.086
long -0.0079 0.02 0.14 0.22 0.24 0.23 0.12 -0.038 -0.079 -0.11 0.2 0.34 -0.15 0.41 -0.065 -0.56 -0.14 1 0.33
sqft_living15 0.0044 0.58 0.41 0.57 0.76 0.14 0.28 0.084 0.28 -0.095 0.71 0.73 0.2 0.32 0.0008 -0.28 0.048 0.33
sqft_lot15 0.0033 0.082 0.032 0.088 0.18 0.72 -0.011 0.031 0.073 -0.0034 0.12 0.19 0.018 0.07 0.0043 -0.15 -0.086
has_renovated 0.0024 0.1 0.034 0.058 0.047 -0.014 0.0023 0.016 0.037 -0.07 0.032 0.012 0.074 -0.12 0.69 0.05
date price bedrooms bathrooms sqft_living sqft_lot floors waterfront view condition grade sqft_above sqft_basement yr_built yr_renovated

Models

Model 1

- The baseline model included data from all the homes sold across Kings County, Seattle.
- The following graphic depicts the early correlation between price and square feet
- It is when we look at the linearity of the model that we appreciate the short comings of its predictive power

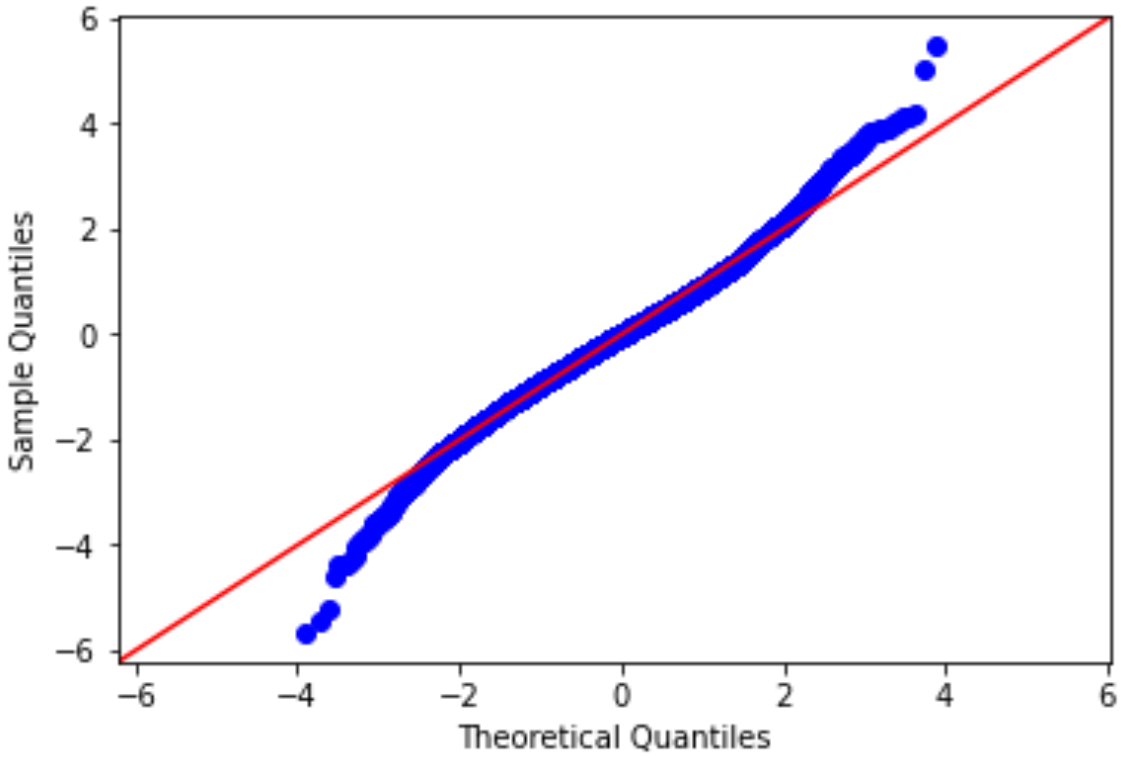




Model 2

- The second model was carried out using data after a logarithmic or Log-Level Regression was made on the price value as it is the dependent variable or determining outcome we are after.
- The 'R-squared' value depicts the percentage of uncertainty that the data used reflects the outcomes of this model The error below depicts the errors made while calculating the values from the same data set split into 80/20 percent between the "train" data and the "test" data
    - Train: 173545.4282194569
      Test: 184910.2649419146
- This number reflects the possible error from the actual home price of the underlying asset
- The Price to variable relationship did, however, considerably increase

Dep. Variable:	log_price	**R-squared:**	0.752
Model:	OLS	**Adj. R-squared:**	0.751
Method:	Least Squares	**F-statistic:**	1566.
Date:	Thu, 23 Sep 2021	**Prob (F-statistic):**	0.00
Time:	16:42:59	**Log-Likelihood:**	-815.23
No. Observations:	20747	**AIC:**	1712.
Df Residuals:	20706	**BIC:**	2038.
Df Model:	40		
Covariance Type:	nonrobust		



Model 3

Dep. Variable:	log_price	**R-squared:**	0.739
Model:	OLS	**Adj. R-squared:**	0.739
Method:	Least Squares	**F-statistic:**	1457.
Date:	Tue, 28 Sep 2021	**Prob (F-statistic):**	0.00
Time:	11:16:52	**Log-Likelihood:**	-576.07
No. Observations:	20610	**AIC:**	1234.
Df Residuals:	20569	**BIC:**	1559.
Df Model:	40		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-51.9952	1.957	-26.566	0.000	-55.831	-48.159
sqft_living	0.0002	4.48e-06	53.520	0.000	0.000	0.000
sqft_lot	4.573e-07	4.64e-08	9.859	0.000	3.66e-07	5.48e-07
sqft_basement	-4.495e-05	5.85e-06	-7.685	0.000	-5.64e-05	-3.35e-05
lat	1.3867	0.013	104.070	0.000	1.361	1.413
long	0.0133	0.015	0.879	0.380	-0.016	0.043
bedrooms_2	0.0438	0.021	2.057	0.040	0.002	0.086
bedrooms_3	0.0327	0.021	1.535	0.125	-0.009	0.074
bedrooms_4	0.0192	0.022	0.886	0.376	-0.023	0.062
bedrooms_5	-0.0006	0.023	-0.025	0.980	-0.045	0.044
bedrooms_6	-0.0491	0.028	-1.753	0.080	-0.104	0.006
bathrooms_1_5	0.0285	0.008	3.417	0.001	0.012	0.045
bathrooms_1_75	0.0621	0.007	8.701	0.000	0.048	0.076
bathrooms_2	0.0673	0.008	8.454	0.000	0.052	0.083
bathrooms_2_25	0.0920	0.009	10.343	0.000	0.075	0.109
bathrooms_2_5	0.0882	0.009	10.252	0.000	0.071	0.105
bathrooms_2_75	0.1100	0.011	10.250	0.000	0.089	0.131
bathrooms_3	0.1046	0.012	8.405	0.000	0.080	0.129
bathrooms_3_25	0.1478	0.014	10.304	0.000	0.120	0.176
bathrooms_3_5	0.1561	0.014	11.215	0.000	0.129	0.183
bathrooms_3_75	0.1839	0.024	7.651	0.000	0.137	0.231
bathrooms_4	0.1263	0.027	4.711	0.000	0.074	0.179
bathrooms_4_25	0.1225	0.036	3.437	0.001	0.053	0.192
floors_1_5	0.0120	0.007	1.675	0.094	-0.002	0.026
floors_2	0.0046	0.006	0.744	0.457	-0.008	0.017
floors_3	-0.0029	0.013	-0.231	0.817	-0.027	0.022
waterfront_1	0.3711	0.031	12.087	0.000	0.311	0.431
view_1	0.1920	0.015	13.202	0.000	0.163	0.220
view_2	0.1483	0.009	16.875	0.000	0.131	0.166
view_3	0.2184	0.012	17.841	0.000	0.194	0.242
view_4	0.3009	0.020	15.298	0.000	0.262	0.340
condition_4	0.0760	0.004	17.187	0.000	0.067	0.085
condition_5	0.1238	0.007	17.545	0.000	0.110	0.138
grade_fair	0.2114	0.007	30.419	0.000	0.198	0.225
grade_good	0.4211	0.008	50.925	0.000	0.405	0.437
grade_excellent	0.6347	0.012	50.882	0.000	0.610	0.659
yr_built_early_century	-0.1336	0.007	-17.938	0.000	-0.148	-0.119
yr_built_mid_century	-0.2605	0.008	-34.220	0.000	-0.275	-0.246
yr_built_modern	-0.3118	0.008	-36.944	0.000	-0.328	-0.295
yr_built_post_modern	-0.2789	0.009	-29.641	0.000	-0.297	-0.261
has_renovated_1	0.1380	0.014	9.546	0.000	0.110	0.166

The third model was performed under the same cirmustances as the second model with the exception that home prices were kept under two-million dollars and the number of bedrooms per home was six or less.

Train: 157,669.94
Test: 153,145.46

The final model depicts a complex relationship between price and the underlying features of the home. The most correlated asset was the of square foot per home with the a positive correlation of 2% increase in the value of the home per square foot.

The R-squared value lets us know that the results we infer from this analysis have a 74% chance of being account for in this model.

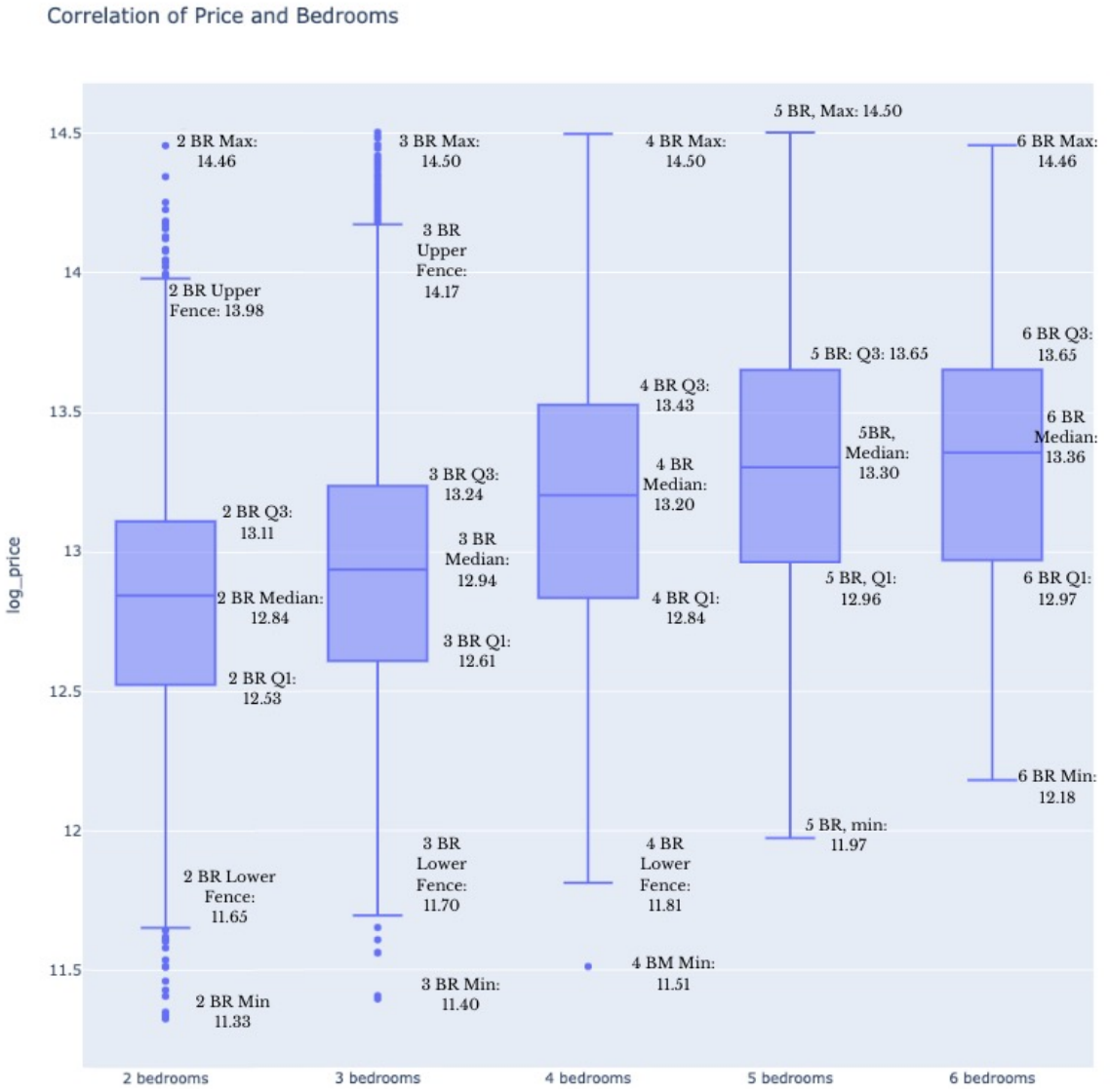
Model 3

This is a model where the dependent variable is logged but the independent variable is not:

$$\ln(Y) = a + bX + e$$

This is known as a log-level model and the interpretation is that a unit increase in X results in a 100*b% increase in Y (we multiply by 100 because b is a percentage).

This would mean that a year increase in bedrooms is associated with a roughly 100*b% increase in price (log_price)



Conclusion

The price of homes and square feet have a positive collinearity. However, it is imperative to note that the overall price of the underlying asset is greatly influenced by many factors such as number of rooms, if the home has a waterfront and many more features denoted in the in the regression model.

Future Work

FURTHER ANALYSIS OF HOMES BEYOND NICHE MARKET

FURTHER ANALYSIS OF HOMES IN CURRENT MARKET CONDITIONS