King County Real Estate Consulting

Jake Sherman

Overview

- Using the King County data set, I create a linear regression to provide local homeowners advice on how best to renovate their home to optimize returns when they are ready to sell.

- Additional square foot adds $96 to home value.

- Each unit of condition adds about $41,000 to home value.

Business Value

- Linear Regression Model inform clients how units changes in certain independent variables will affect sales price.

|	date	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	condition	grade	sqft_above	sqft_basement
0	2014	221900.0	3	1.00	1180	5650	1.0	0	3	7	1180	0.0
1	2014	538000.0	3	2.25	2570	7242	2.0	0	3	7	2170	400.0
2	2015	180000.0	2	1.00	770	10000	1.0	0	3	6	770	0.0
3	2014	604000.0	4	3.00	1960	5000	1.0	0	5	7	1050	910.0
4	2015	510000.0	3	2.00	1680	8080	1.0	0	3	8	1680	0.0

Methodology

SOURCED DATA FROM KC HOUSING DATASET

PERFORMED LINEAR REGRESSION ON THE DATA SET

ITERATED PARAMETERS TO OPTIMIZE RMSE

Correlation Heatmap

Heatmap of Correlation Between Attributes
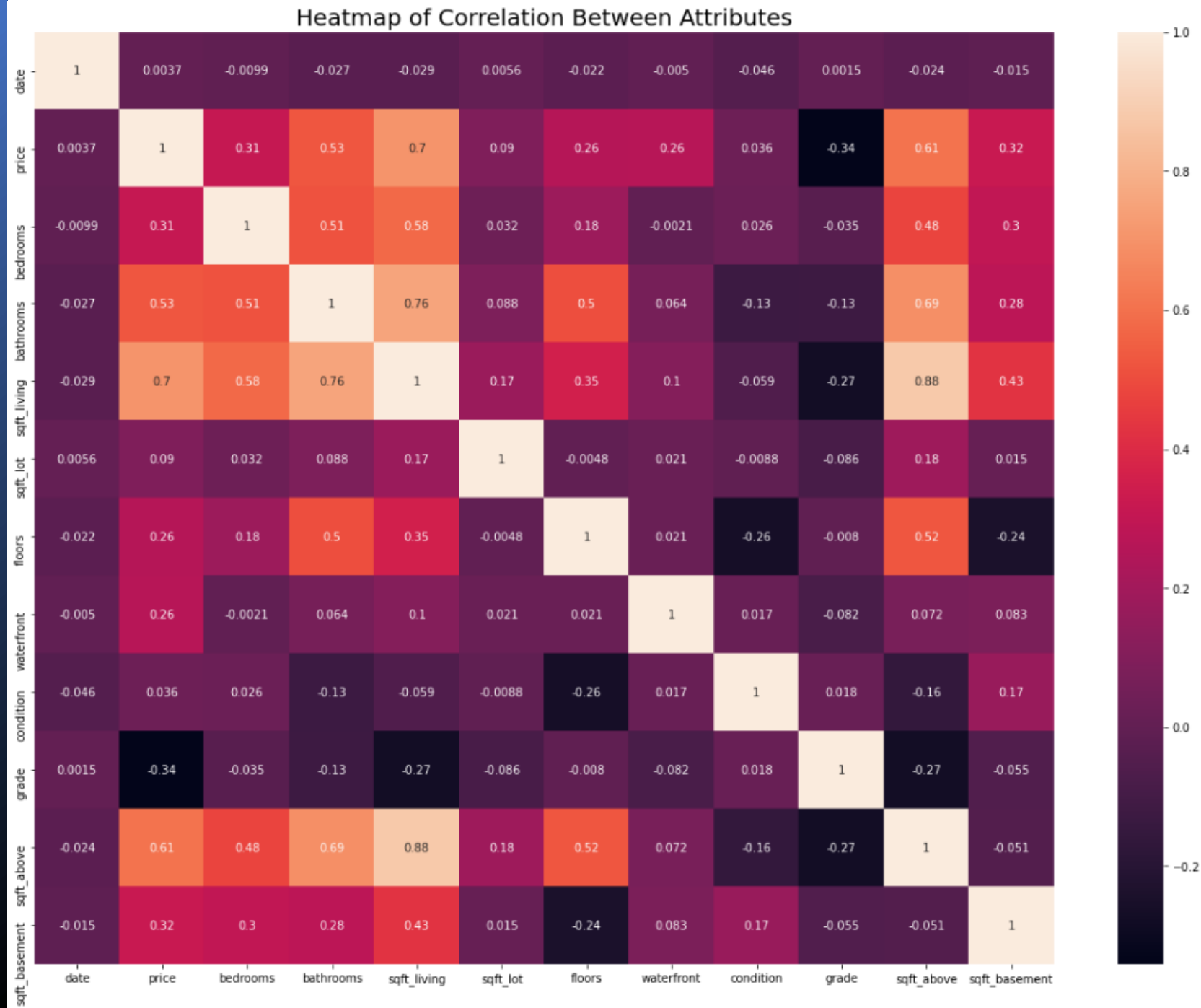
Correlation Breakdown

NOT A LOT OF STRONG
CORRELATIONS

SQUARE FOOTAGE OF LIVING SPACE
WAS THE STRONGEST AT .71

Initial Regression

OLS Regression Results

Dep. Variable:	price	R-squared:	0.570
Model:	OLS	Adj. R-squared:	0.570
Method:	Least Squares	F-statistic:	2599.
Date:	Sat, 03 Sep 2022	Prob (F-statistic):	0.00
Time:	18:44:58	Log-Likelihood:	-2.9828e+05
No. Observations:	21597	AIC:	5.966e+05
Df Residuals:	21585	BIC:	5.967e+05
Df Model:	11		
Covariance Type:	nonrobust		

|	coef	std err	t	P>|t|	[0.025	0.975]
const	-4.748e+07	7.08e+06	-6.707	0.000	-6.14e+07	-3.36e+07
date	2.361e+04	3514.386	6.717	0.000	1.67e+04	3.05e+04
bedrooms	-4.731e+04	2246.541	-21.058	0.000	-5.17e+04	-4.29e+04
bathrooms	1.557e+04	3618.897	4.304	0.000	8480.624	2.27e+04
sqft_living	244.9618	21.530	11.378	0.000	202.761	287.163
sqft_lot	-0.3602	0.041	-8.843	0.000	-0.440	-0.280
floors	3.758e+04	4123.840	9.112	0.000	2.95e+04	4.57e+04
waterfront	7.869e+05	2.02e+04	38.884	0.000	7.47e+05	8.27e+05
condition	5.187e+04	2646.959	19.596	0.000	4.67e+04	5.71e+04
grade	-2.825e+04	920.754	-30.686	0.000	-3.01e+04	-2.64e+04
sqft_above	21.2562	21.531	0.987	0.324	-20.946	63.458
sqft_basement	53.3135	21.437	2.487	0.013	11.296	95.331

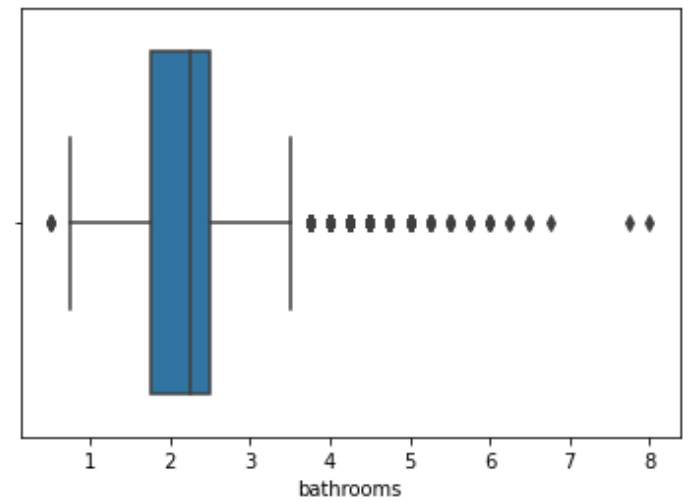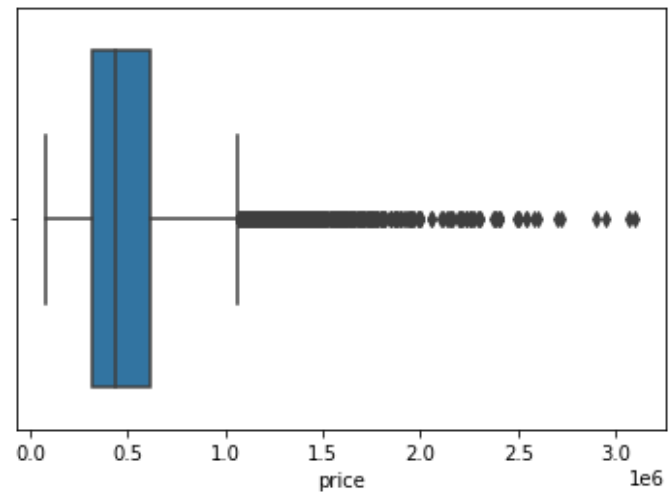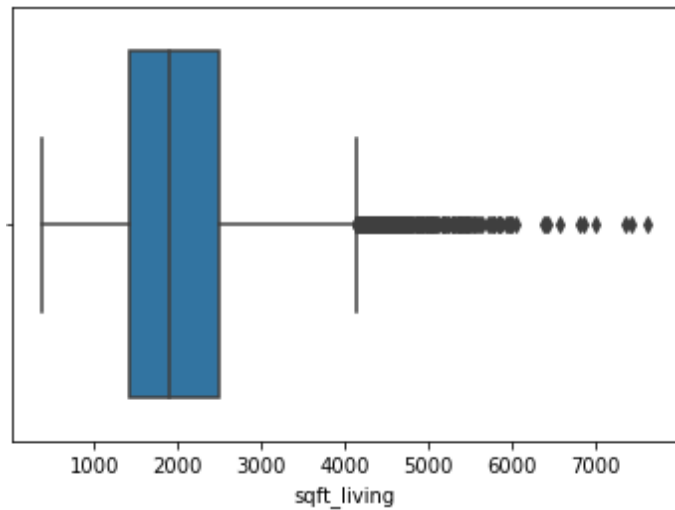| Omnibus:	12932.261	Durbin-Watson:	1.987
Prob(Omnibus):	0.000	Jarque-Bera (JB):	417616.686
Skew:	2.339	Prob(JB):	0.00
Kurtosis:	24.028	Cond. No.	1.90e+08

Baseline
Root Mean Squared Error and Linearity

- Initial RMSE: $241,000

- Average Price of home: $540,300

linearity check
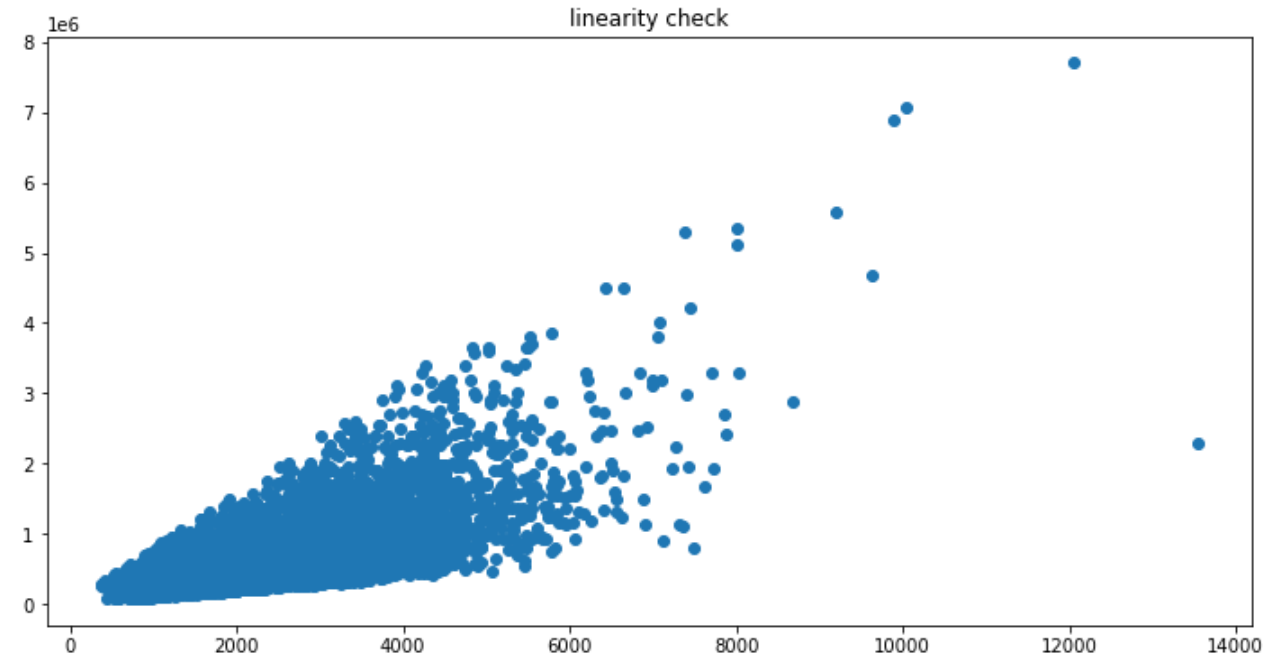
Baseline Assumptions
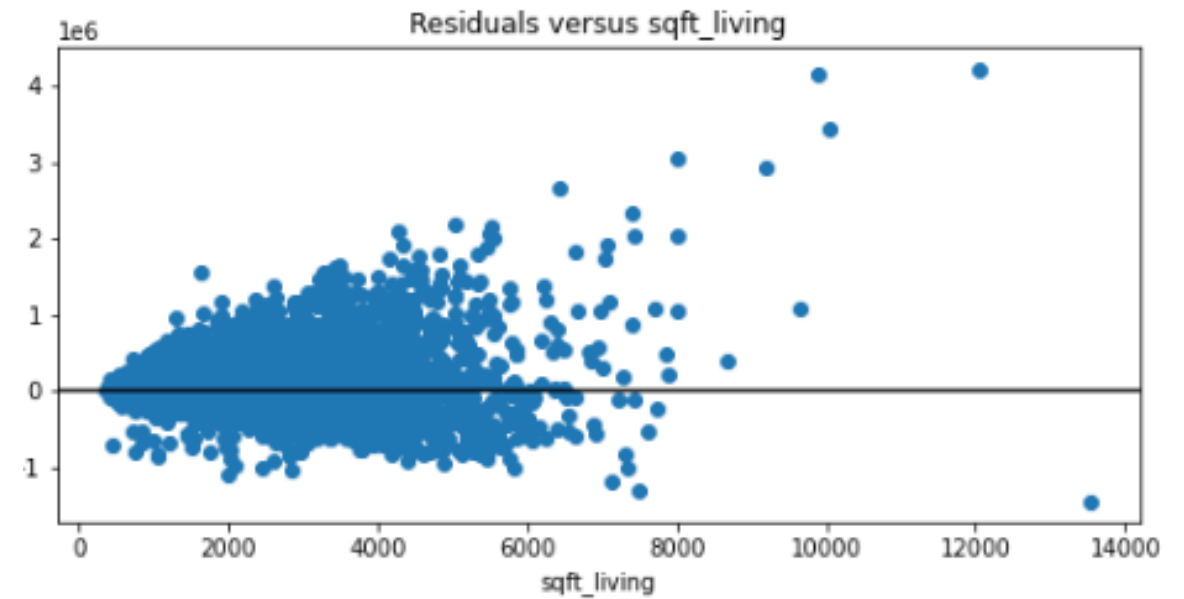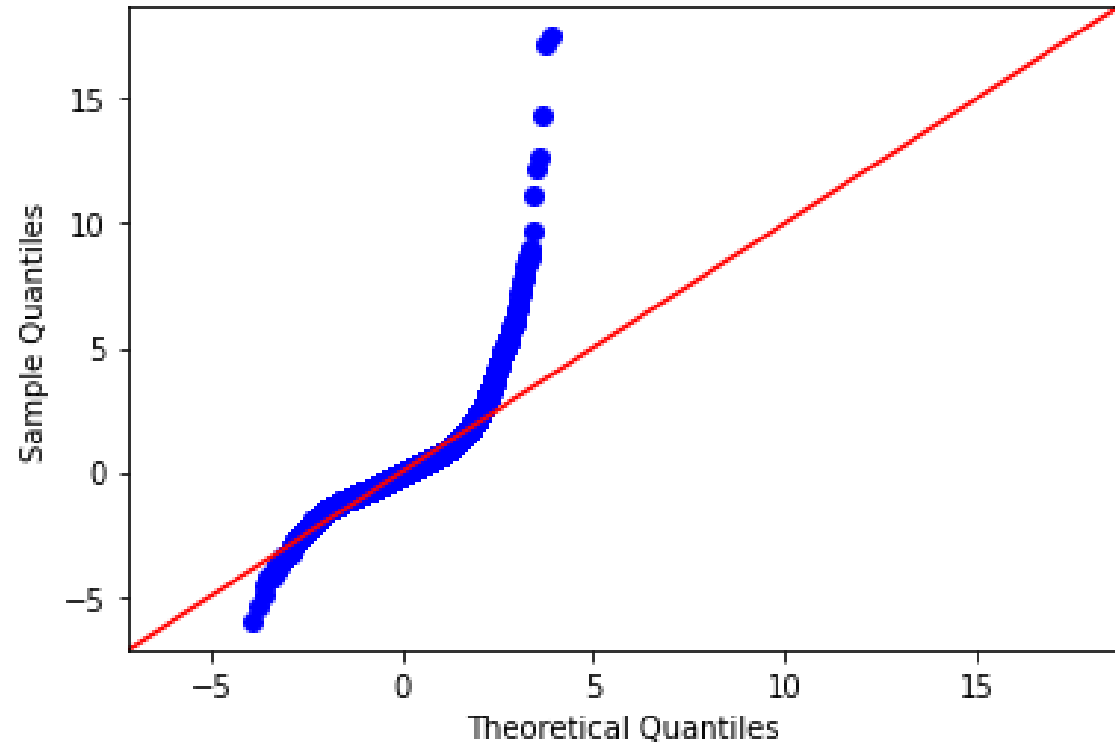
Iteration 1: Drop Outliers

Iteration 2: Create Categorical Values

Improved R2 score but too many high p-values

floors_1.5	8.993e+04	3754.784	23.950	0.000	8.26e+04	9.73e+04
floors_2.0	9327.3288	3416.321	2.730	0.006	2631.044	1.6e+04
floors_2.5	6.647e+04	1.43e+04	4.661	0.000	3.85e+04	9.44e+04
floors_3.0	6.101e+04	6777.159	9.003	0.000	4.77e+04	7.43e+04
floors_3.5	7.162e+04	5.76e+04	1.243	0.214	-4.14e+04	1.85e+05
waterfront_1	1.932e+05	2.16e+04	8.930	0.000	1.51e+05	2.36e+05
grade_3	-4.357e+05	1.42e+05	-3.060	0.002	-7.15e+05	-1.57e+05
grade_4	-4.216e+05	2.97e+04	-14.215	0.000	-4.8e+05	-3.63e+05
grade_5	-4.207e+05	1.2e+04	-35.067	0.000	-4.44e+05	-3.97e+05
grade_6	-3.782e+05	7924.213	-47.722	0.000	-3.94e+05	-3.63e+05
grade_7	-3.011e+05	6875.988	-43.795	0.000	-3.15e+05	-2.88e+05
grade_8	-2.057e+05	6426.845	-32.013	0.000	-2.18e+05	-1.93e+05
grade_9	-9.122e+04	6380.173	-14.298	0.000	-1.04e+05	-7.87e+04
condition_2	-2.273e+04	2.89e+04	-0.787	0.432	-7.94e+04	3.39e+04
condition_3	2178.0411	2.69e+04	0.081	0.935	-5.05e+04	5.48e+04
condition_4	3.327e+04	2.69e+04	1.237	0.216	-1.94e+04	8.6e+04
condition_5	8.193e+04	2.71e+04	3.025	0.002	2.88e+04	1.35e+05
bedrooms_2	-2892.6767	1.09e+04	-0.265	0.791	-2.43e+04	1.85e+04
bedrooms_3	-5.2e+04	1.09e+04	-4.763	0.000	-7.34e+04	-3.06e+04
bedrooms_4	-6.072e+04	1.12e+04	-5.422	0.000	-8.27e+04	-3.88e+04
bedrooms_5	-7.247e+04	1.2e+04	-6.027	0.000	-9.6e+04	-4.89e+04

Iteration 3: Feature Scaling
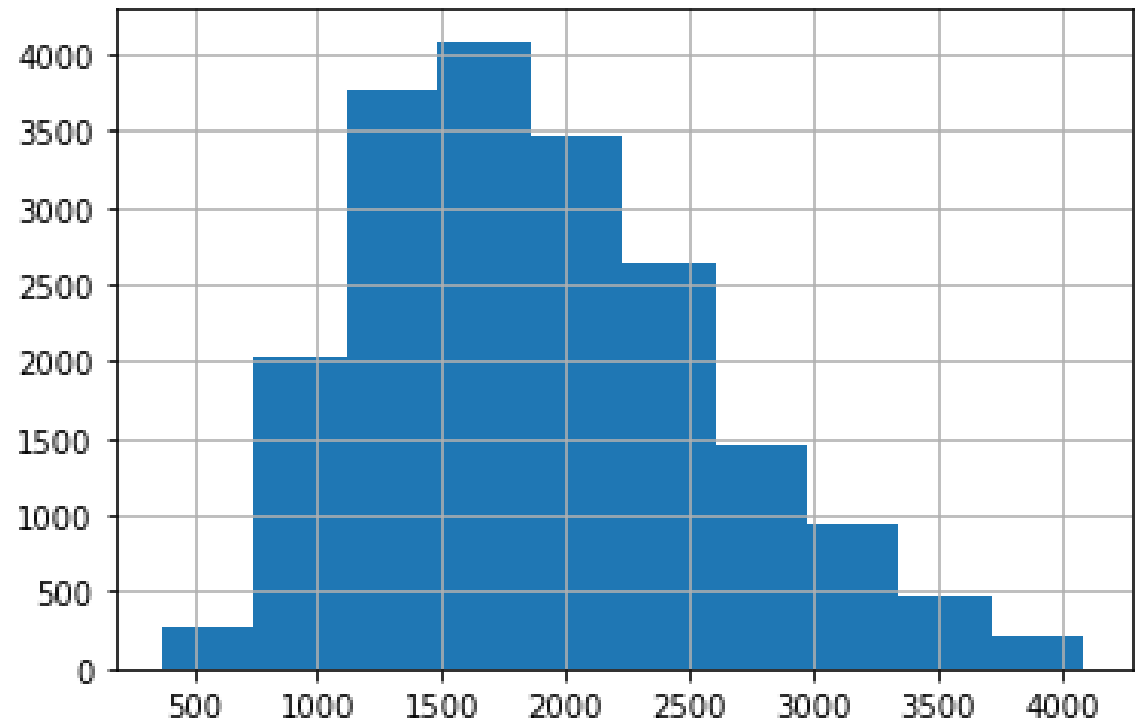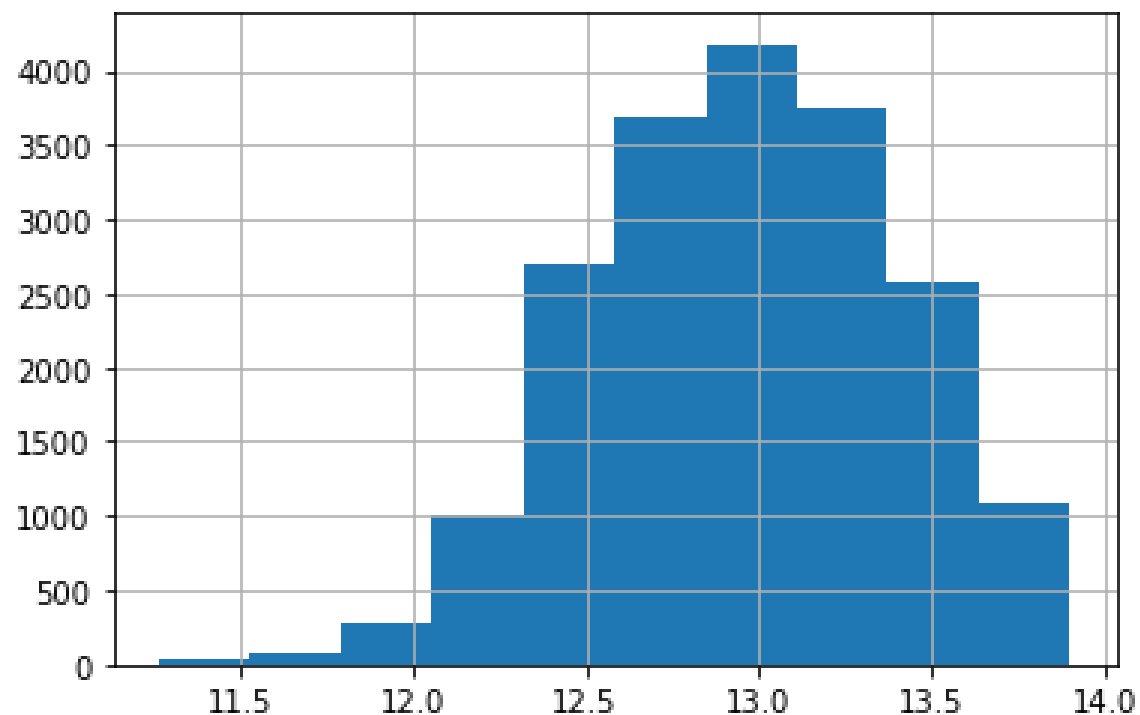
Results

- Improved linearity, normality, and homoscedasticity

- No high p-values (above .05)

- RMSE is now down $144,361 from $241,000
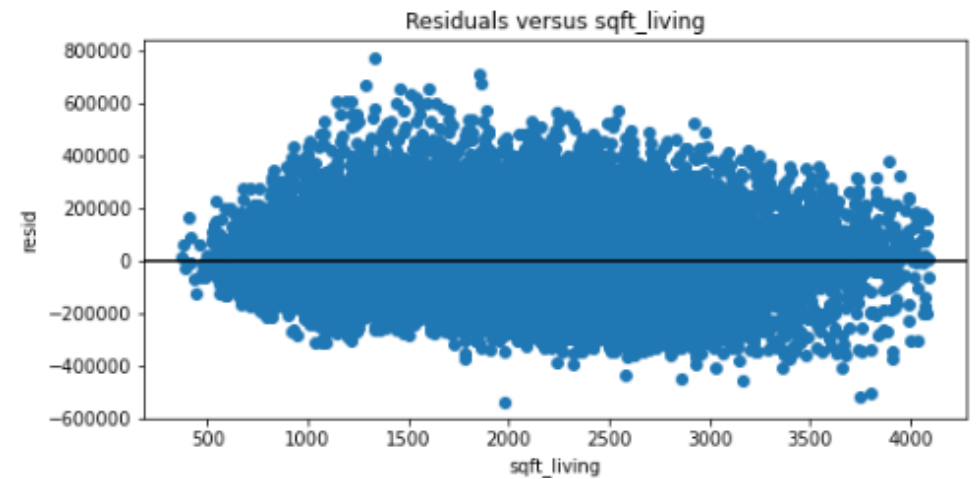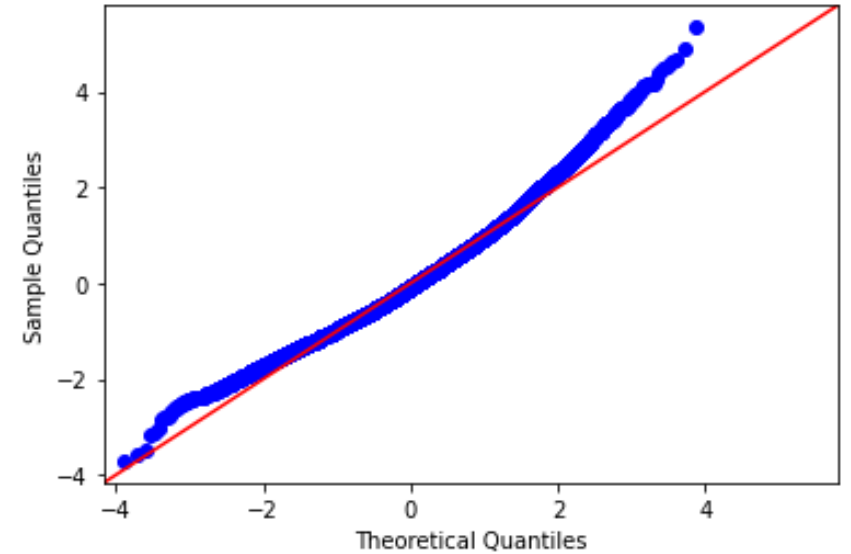
Conclusion

- Owner can add $96 of home value for each additional square foot added of living space (finish attic, add onto house, etc).

- For every unit of increased condition as defined by Kingcounty.gov, a homeowner can expect to a gain of $40,910. For a home with a condition level 1, doing a full renovation to a level 5 will add an estimated $163,640.

Limitations

- Data did not have an ideal linearity which affected goodness of fit and predictability- different zip codes have a large effect on home value and weren't accounted for.

- Data such as view, condition, and grade are subjective.

- Going forward would like to include complete set of variables in analysis.

Thank You!

Github.com/jsherman918