Initial model evaluation has not suggested any overwhelming issues within the data though model transformation may be necessary. The data was scaled and separated into a training set and test set. We then acquired an initial Mean Squared Error which we will seek to minimize through proper model selection. MSE for the full model measured at 0.3685. Initial analysis of the residuals suggests at least a slight skew in the data as well as several outlying observations. Please see Figure 1 to the right where the red horizontal lines represent 3 standard deviations from the center. Additionally, the normality assumption is clearly not met suggesting a possible need for transformation. The plot seems to suggest a positive skew. Refer to the Quantile-Quantile plot in Figure 2. This is not entirely surprising and this pattern can also be seen in Figure 1 where we see a multitude of points greater than 3 standard deviations from the mean of the residuals. The maximum price of a house tends to be further from the mean price than the minimum.

Figure 2. QQ Plot

Figure 1. Residual Plot

Also measured was multicollinearity which doesn’t appear to excessively impact variability of the predictors. The highest inflation factor appeared to be sqft\_above (square footage above the ground floor) at 4.82. A factor of almost 5 warrants some investigation, particularly how it relates to sqft\_lot (square footage on ground level).