Relation of Home Prices to Home Features

Jarod, Sam and Calvin

3/6/20

Problem Domain

- Anyone looking to buy or sell a house is likely curious as to what factors really influence the price of a home they are looking to purchase or list for sale.

- Real estate investors who look to ‘flip’ houses after remodels are likely also curious as to what factors would allow them to maximize sale price.

- Typically, people have relied on real estate experts but what if there was evidence of certain home features actually impacting the price of a house?

Overview of Methodology

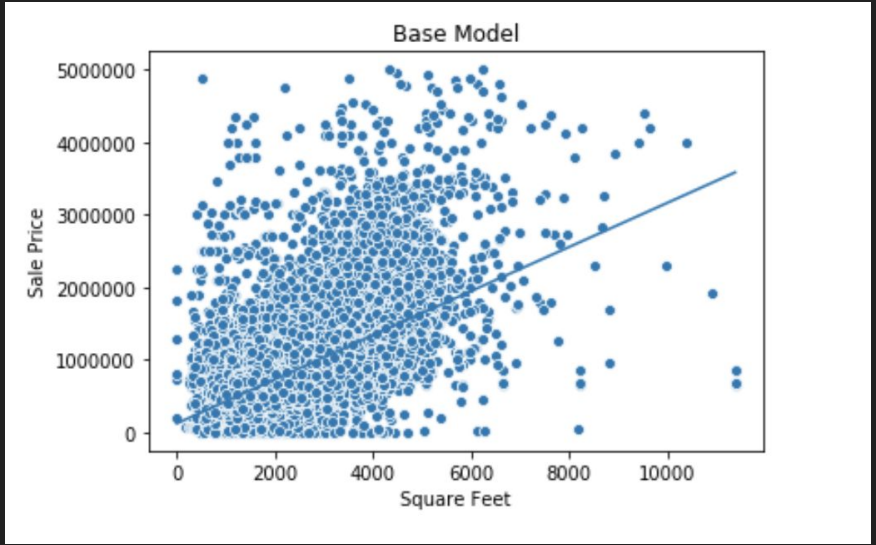
- The specific manner in which our group used to find a correlation between home feature(s) and sale price was by using a computer model.

- This computer model ‘takes in’ certain home features (square footage, bedrooms) and outputs a value (called R-squared) that shows the connection between these features and how likely they are to impact home sale price.

- This method is not without limitations. Certain home features are too closely related and can result in a false positive regarding their impact on sale price. This meant that our group had to account for multicollinearity.

Base Model

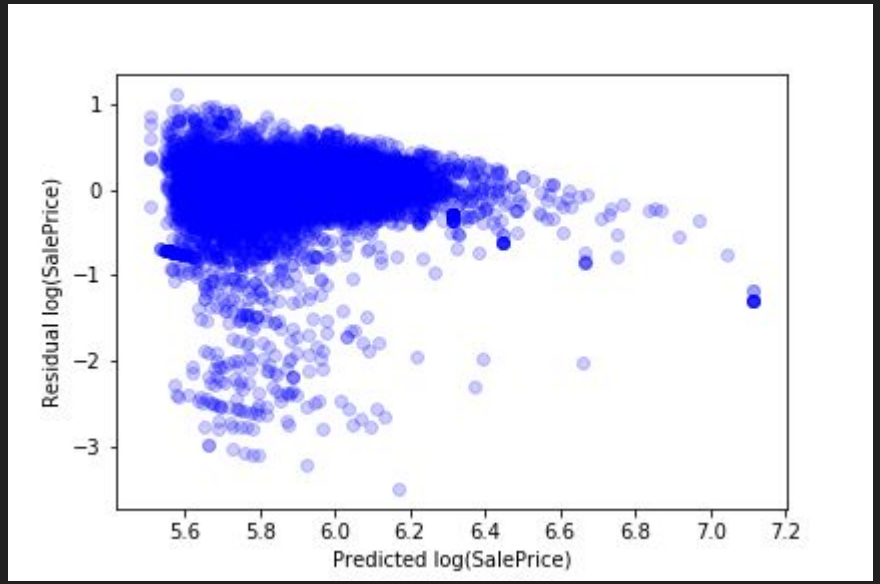
- Started by predicting sale price using only one feature: total sqft
- ‘Best Fit’ Line: Y = 303.82X + 124,030
- Sq ft of houses are not normally distributed
- R-squared value of: 0.343



Evaluation of our Model

- Linearity
  - Rainbow p: 0.013
- Residual Distribution
  - JB: 229982
- Homoscedasticity
  - F-stat: 15470

Plot of Residuals



Note that the logarithm of sale price is used. This makes the graph more easily visualized.

Model Optimization

- Features: SqFtTotLiving, TidelandShoreland, SqFtOpenPorch, SqFtDeck, Area, LakeSammamish, LakeWashington, SeattleSkyline, MtRainier, Olympics, Cascades, Territorial, SmallLakeRiverCreek

- Collinearity did not exceed our threshold score of 5

- Residuals are not evenly distributed

- R-squared value of: 0.422

Final Model

- Features: BldgGrade, TidelandShoreland, SqFtOpenPorch, SqFtDeck, Area, ViewScore, YrRenovated

- Collinearity did not exceed our threshold score of 5

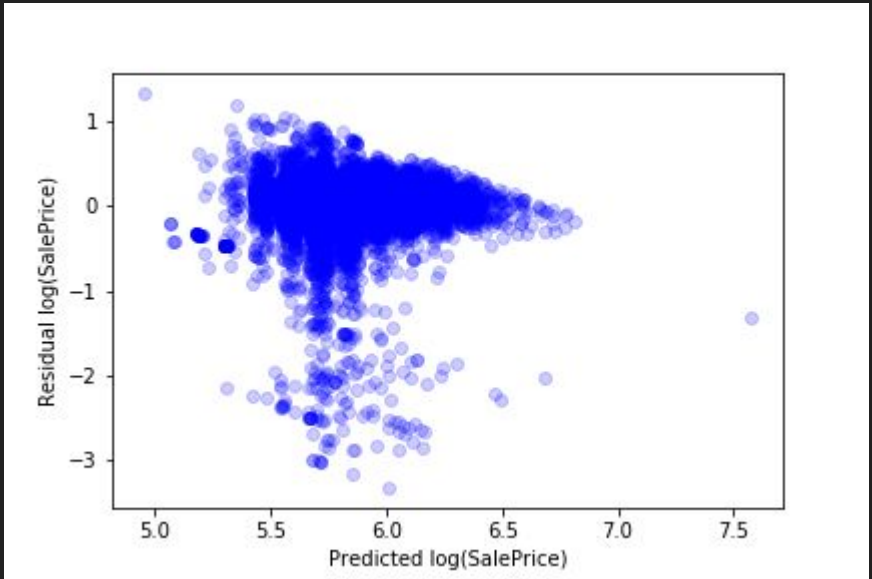
- Residuals still not evenly distributed, to an extent a result of our large amount of data

- R-squared value of: 0.435

Final Model Evaluation

- Linearity
  - Rainbow p: 0.354
  - Better than Base Model
- Residual Distribution
  - JB: 283186
  - Worse than base model
- Homoscedasticity
  - F-stat: 3258
  - Much better than base model
- Variance Inflation Factor
  - All under 5

Plot of Residuals



Note that the logarithm of sale price is used. This makes the graph more easily visualized.

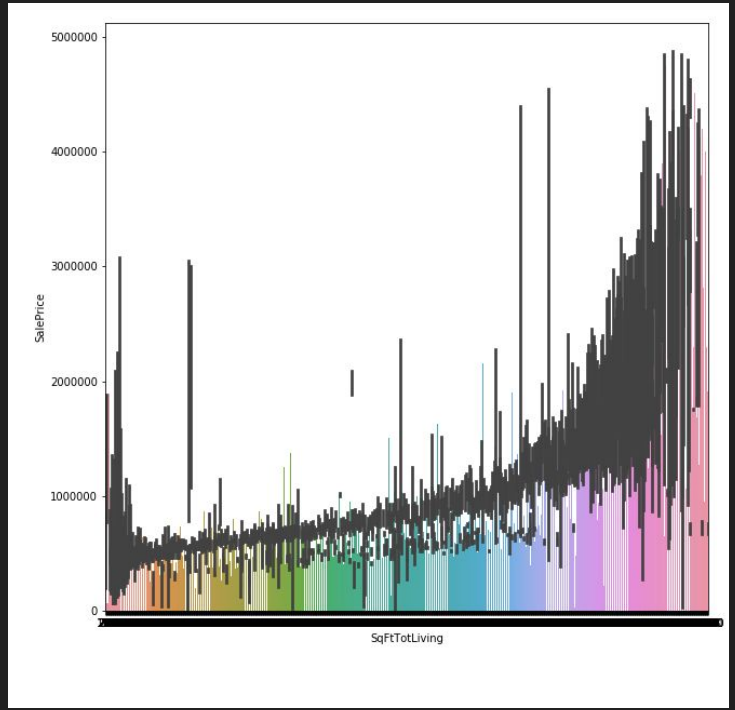
Specific Insights

- Null Hypothesis:
  - Higher square footage increases home sale price.

- Based on our model and data, we fail to reject the null hypothesis.

- Null Hypothesis:
  - Having a beachfront or lakefront increases home sale price.

- According to our model, we are unable to reject the null hypothesis.



Specific Insights Part 2

- Null Hypothesis:
  - The presence of a nuisance (power lines, traffic noise, airport noise) decreases home sale price.

- It is clear that there is some statistical impact (P-value < 0.00001) of power lines on sale price but the effect was minimal. Traffic noise had no substantial statistical impact on sale price.

- Therefore, we are reporting our results as non-finding for the null hypothesis.

Conclusions & Real World Recommendations

- Stakeholders should care about what we found because, we now have a working base model to use as a tool to guide and provide insight - regarding the correlation of home prices and specific home features.

- Our group's findings can help home buyers, sellers, real estate investors and home builders. Certain home features are highly sought after in today's real estate climate but not everything necessarily correlates to higher home prices.

- For example, our group found that the top 4 individual factors of home price are: Building grade, square footage, quality of view and waterfront (or not).