Predicting House Prices in King County Washington

By: Daniel Blake, Andrew Calkins, Ross Richesin

Summary

We used a linear regression model to predict house prices in King County Washington. This model allows a user to enter information about the house they are trying to sell and predicts a price based on these factors.

Outline

- Data

- Methods

- Results

- Conclusion

Data & Methods

The dataset we used contained 21 predictors and 21000+ data points.

Preliminary data cleaning included filling in missing values, translating predictors such as date of sale and zip code into a form useable for the model, and adjusting data types.

After the data was cleaned a new version of the data set was created a distributed to the project members.

Results

Grade of Home vs Price of Home

Price of Home (y-axis), Grade of Home (x-axis)
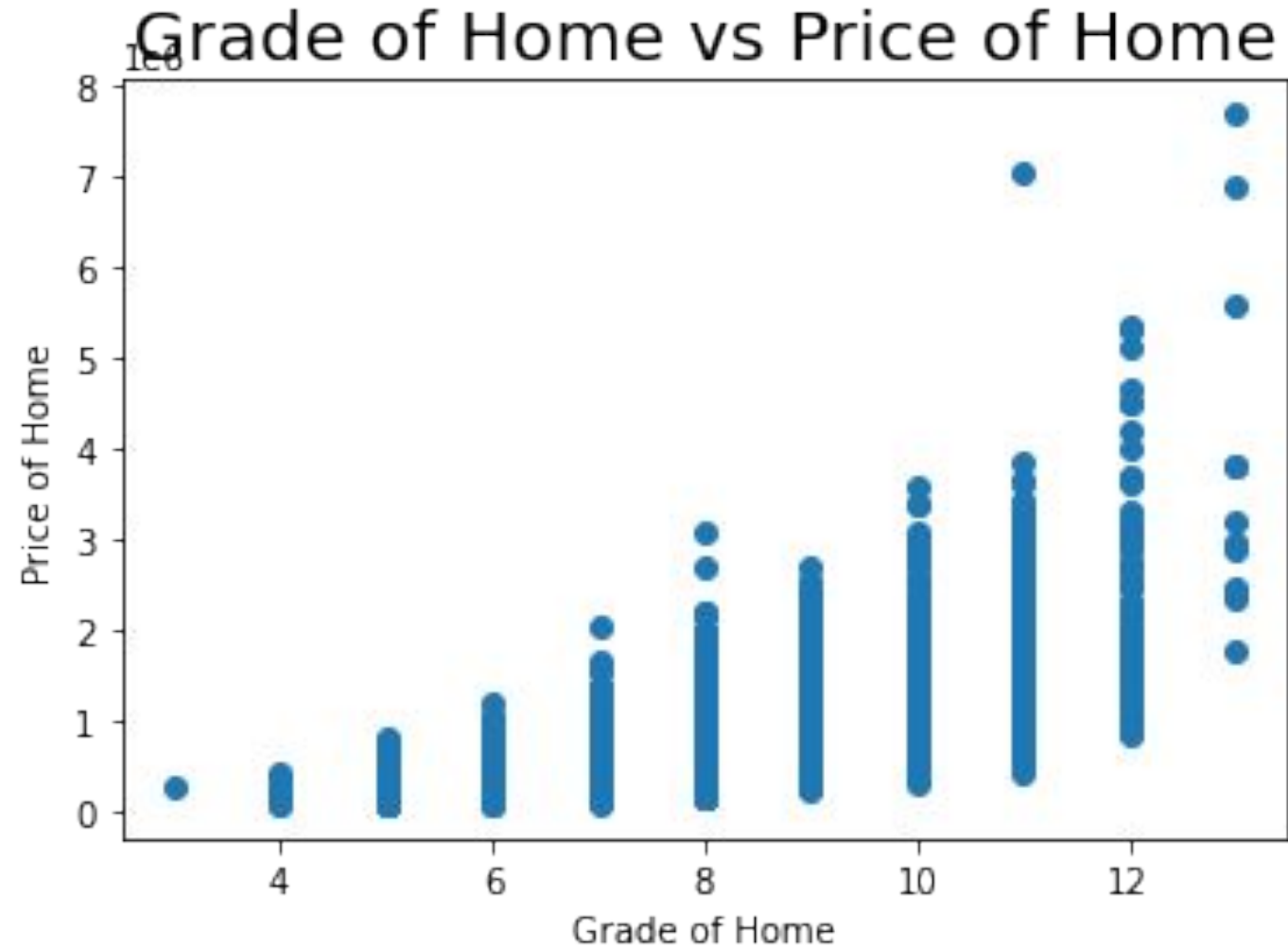
price is in $100,000

-Initially attempted to split data as per your representatives recommendation.

-split based on grade, one of the strongest individual predictor of price

-reevaluated after running into accuracy issues with split models.

-High split explained 74% of the variance, low split only 56%

Baseline Model

To get a baseline for our model to improve upon, we simply fed our entire cleaned dataset into a linear regression model. It showed decent correlation of around 74% .

-on the left you see some of the strongest predictors in our data, square feet living area, the King county area grading scale, number of bathrooms, and the sqft of the nearest 15 neighbors.

Results
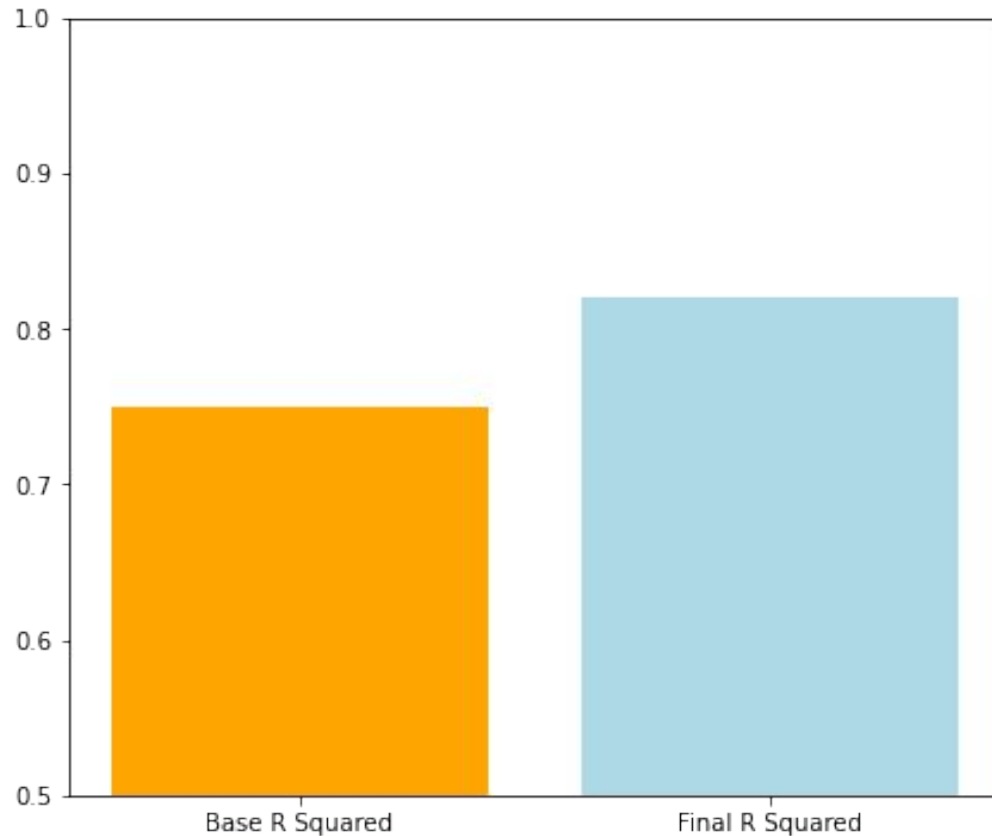
-To improve our model we performed some standardization to make the data more uniform. In addition we also normalized the data, dropped unnecessary features, and performed some engineering for increased correlation

-The final model explain around 82% of the data's variance or in other words, around 82% of the variability in Y(price) can be explained using X (Independent Variables).

Conclusion

-Splitting data resulted in model not having enough data to become properly trained, likely what caused the decreased accuracy.

-Main predictors of price are about what you'd expect, grade, sqft and the area, seen by the correlation of sqft_living15

Thank you!

Daniel Blake: dablake@bsc.edu

Andrew Calkins: awcalkin@bsc.edu

Ross Richesin: rrriches@bsc.edu