

Predicting Sales Prices for Houses in King County

Predictions and Methodology

By George Bennett

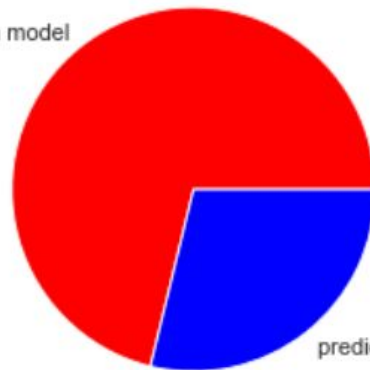


Problem: Maximize profit for home sellers by creating a model to predict housing prices in King County

King County House Sales dataset

- The dataset has 20 Predictors, all but 6 needed to be dropped
- The data set had missing values

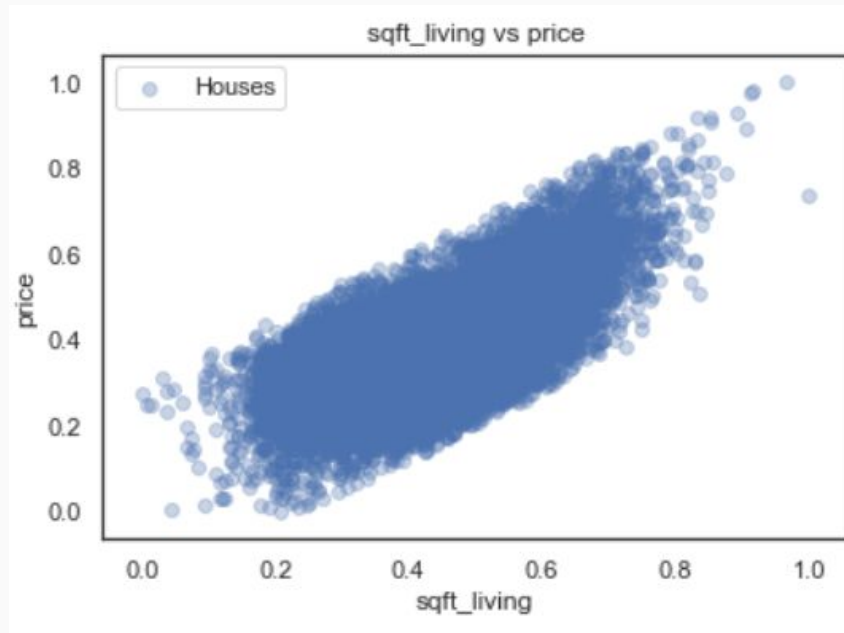
predictors not used in model



predictors used in model

The top five predictors for estimating price

- Square foot living area
- Number of floors
- View on a scale of 1 to 4
- Condition on a scale of 1 to 5
- Zip Code
- Whether or not the home is on the waterfront



The effects of the predictors

- Price increases with sq ft living space, view and condition.
- Price increases dramatically if the house is by the waterfront
- Some zip codes have a positive effect on the price and some zip codes have a negative effect on price.

Why choose this model?

- The R-squared is high and the overall error is low, meaning the model is highly accurate
- This model has a low difference in error between the test set and train set meaning that the model is going to work well on new data

Conclusion

- The problem has been handled as the model can predict house prices for those who are selling in order to maximize profit

Questions and Comments