KC Real Estate

Predicting Home Prices

using linear regression

King County House Sales Data (2014-2015)

- 21,597 house sales

- includes 21 features from the King County Property Assessor

- description of features from the King County Assessor website here.

Modeling Technique:

- Linear/Multiple Regression

Model Success Metrics:

- R-Squared
- Root Mean Squared Error

Linear/Multiple Regression Modeling

- predictive framework for future house sales

- iterative approach to build an accurate model

Modeling Metrics:

- R-Squared: goodness-of-fit, range: 0-1 (bigger=better)
- Root Mean Squared Error: standard deviation of the model's error (smaller=better)

Model Progression:

Baseline Model:

- began with a 1-variable model: Square footage of the living space.

- Correlation: 70.2%

Model Metrics:

- R-Squared: 0.50

- RMSE: $269,556

Iterative Approaches:

- Increase variables
- Remove collinear variables
- One-Hot-Encoding
- Polynomial Regression

Regression Results:

Results:

- Increase in R-squared: 0.50 to 0.82
- Decrease in RMSE: $269,556 to $159,510

Final Model:

- 2nd degree polynomial regression model.

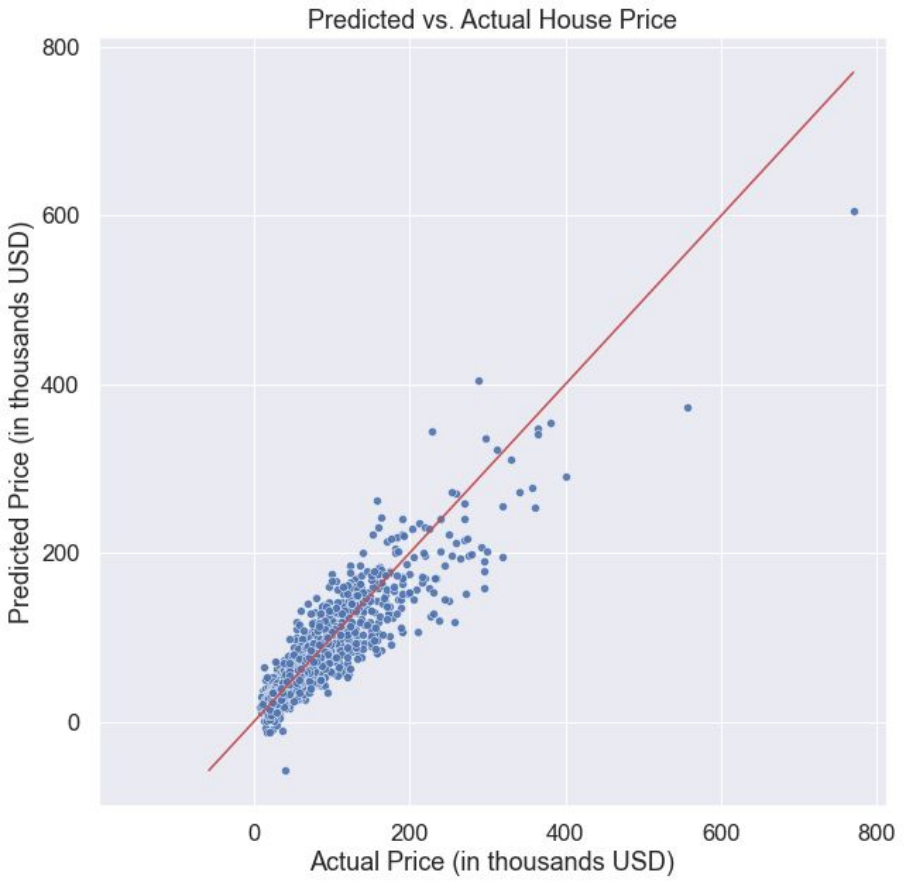
- variables: 171

Model Metrics:

- R-Squared: 0.82

- RMSE: $159,510

Model Validation/Recommendation:



Recommendation:

- The model is useful, but requires oversight. Especially for negative predictions and predictions over $200,000

Next Steps:

Follow-Up Analysis:

- Possible additional features
- Alternative modeling techniques
- Model updates as new data becomes available

Other Considerations:

- Market forces and impact
- Temporal changes

Questions?

Luke DiPerna

LinkedIn

GitHub