
# Multiple Linear Regression

Laura Miller

# Summary

Multiple linear regression analysis of housing data for King County, Washington to predict home value and identify how renovations may improve home value.

# Outline

- Business Problem
- Data
- Methods
- Data Understanding
- Results
- Conclusions

# Business Problem

King Realty needs a model to help homeowners buy/sell homes. The model should help clients learn the following:

- Estimated price of current home
- Estimated price after renovation

# Data

From KingCounty.gov

- King County House Sales
  - 21158 entries, sale years 2014 & 2015
- Cities & Zipcodes
- Cities GIS data

# Methods

## Obtain data

Sales data was provided. Obtained zipcodes + cities info and city geo data from KingCounty.gov.

## Scrub

Imputed missing values with median for a feature, save for “waterfront,”. Removal of outliers.

## Explore

Examined distributions + relationships between individual features and price.

## Feature Engineering

Included month sold, years since renovation (at time of sale), bed-bath ratio, cities.

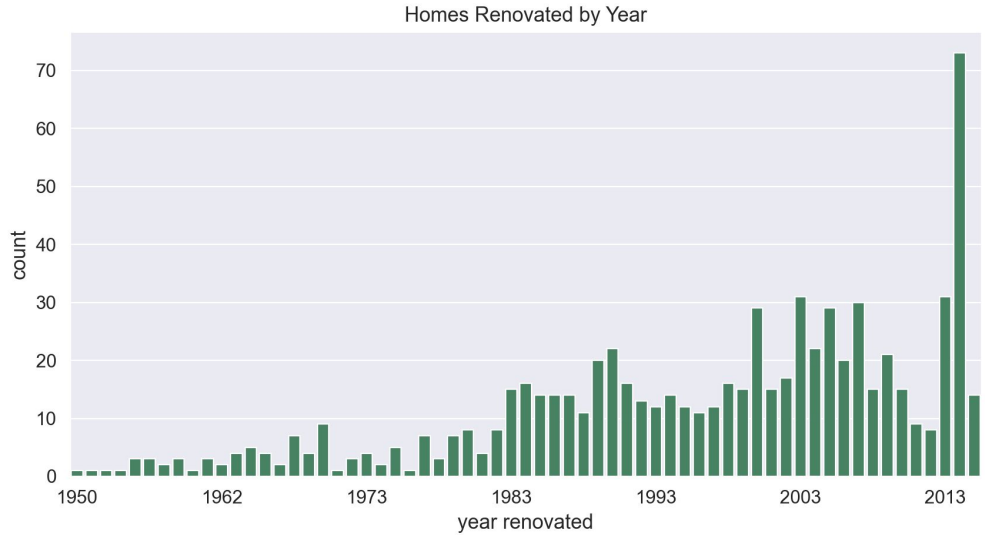
## Model

Built and tuned multiple regression model. Made dummy variables, Log-transformed continuous data and target (price).

## Interpret

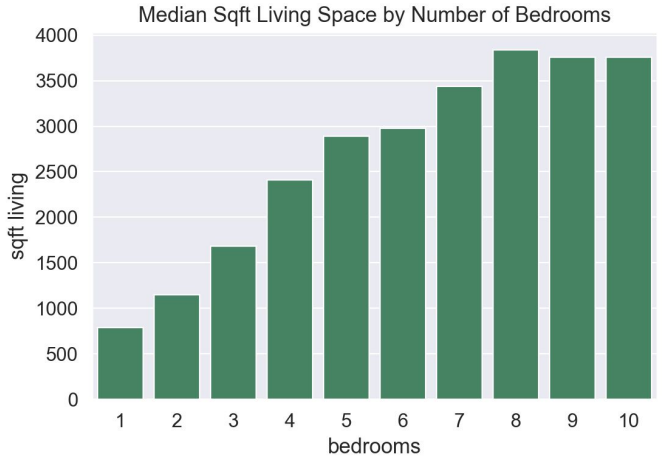
Evaluate model fit and summary statistics.

# Data Understanding

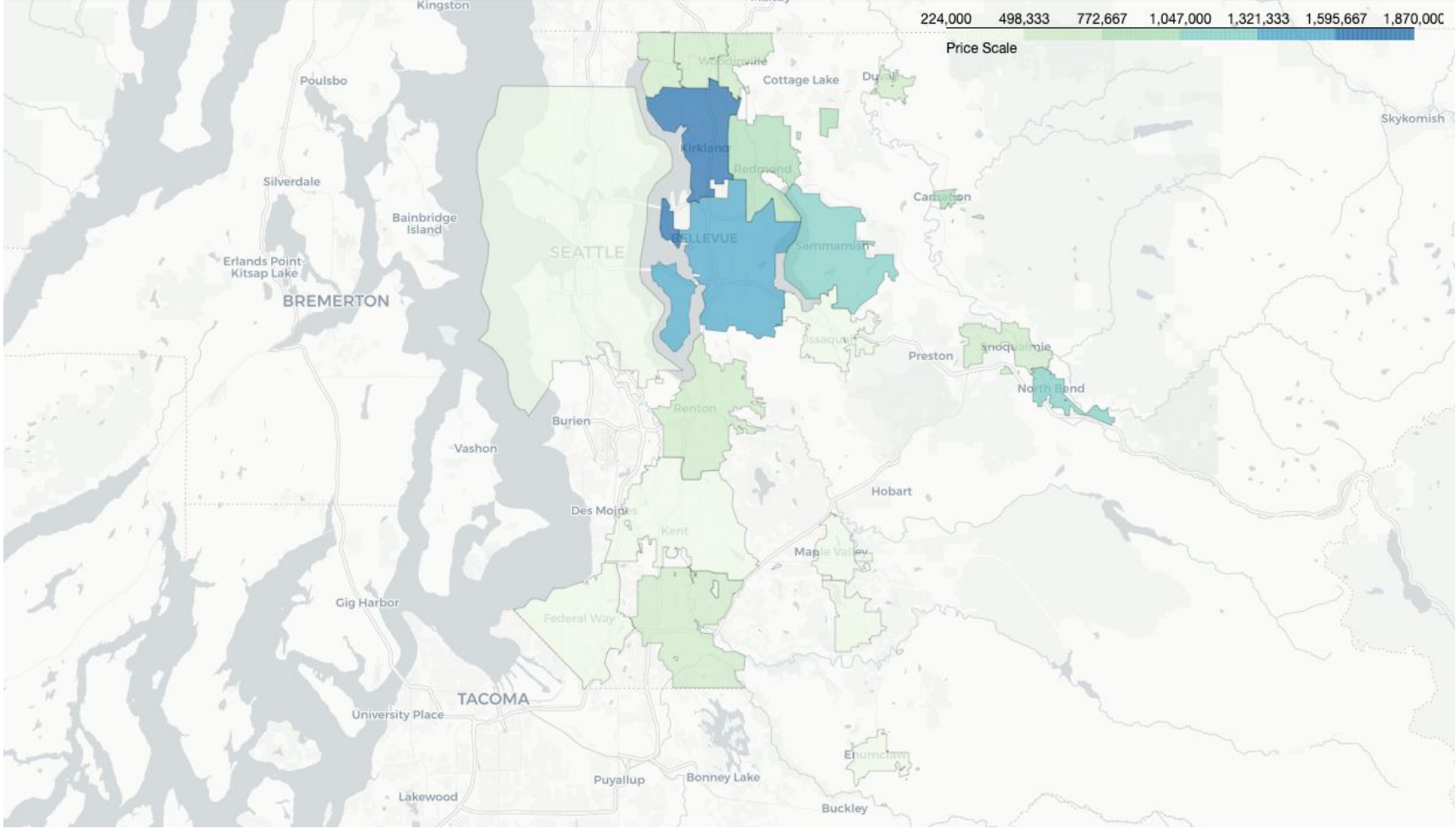



- Home renovation has been increasingly important over the years

# Data Understanding


- Price decreases around 5 bedrooms, while sqft living tapers off




Home prices of King County cities

# Results: Model

log(**price**) = - 128.3740

+ 0.3702 * log(**sqft_living**)
+ 0.1871 * log(**sqft_living15**)
+ 0.0664 * **floors**
+ 0.3237 * **waterfront**
- 0.0030 * **yr_built**
+ 1.2755 * **lat**
- 0.6724 * **long**
- 0.0055 * **month**
- 2.106e-05 * **yrs_since_reno**
- 0.0486 * **bed_bath_ratio**
+ 0.0694 * **grade_6**
+ ...
+ 0.1005 * **view_1_2**
+ ...
+ 0.0907 * **condition_5**
+ ...
+ 0.3636 * **city_Bellevue**
+ ...
+ error

# Results: Model Interpretation

**sqft_living**

Coeff = 0.3702

A 1% increase in square foot living space is associated with a 0.37% increase in price.

**bed_bath_ratio**

Coeff = -0.0486

A 1 unit increase in bed-bath ratio is associated with a 4.9% decrease in price.

**grade_13**

Coeff = 1.2094

Compared to grade 6 homes, grade 13 homes sell for ~121% more.

# Results: Model Diagnostics

- R-squared & adjusted R-squared: 0.819
- All p-values under 0.05
- Root Mean Squared Error (RMSE)
  - Want train/test split to have close & “small” values
    - Train data (75%): $158,178.97
    - Test data (25%): $156,898.73

# Results: Model Test Case

Predicted price changes of 23 Renton homes when bed-bath ratio is reduced from 4 to 2

- Homes had 1 bathroom
- Median price increase: ~6.96%
  - ~3.48% increase in price per unit decrease in bed-bath ratio
  - Coefficient from model: -0.0486

# Conclusion

Achieved model with R-squared above 0.6 (R-squared = 0.817) and low p-values but prediction accuracy can improve

- Next steps:
  - Improve normality of residuals (lower JB value)
    - Remove more outliers
  - Remove additional features
    - Bathrooms: borderline multicolinear with sqft_living
  - Split more features
  - Feature scaling

# Thank You!

Email: laura@warrahdesign.com
Github: @NariMo91