Flatiron Data Science Mod 2 Final Project

Regression Model - Predicting Home Sale Prices

John J. Cho
February 24, 2020

Dataset: 21,597 Properties, 21 Features

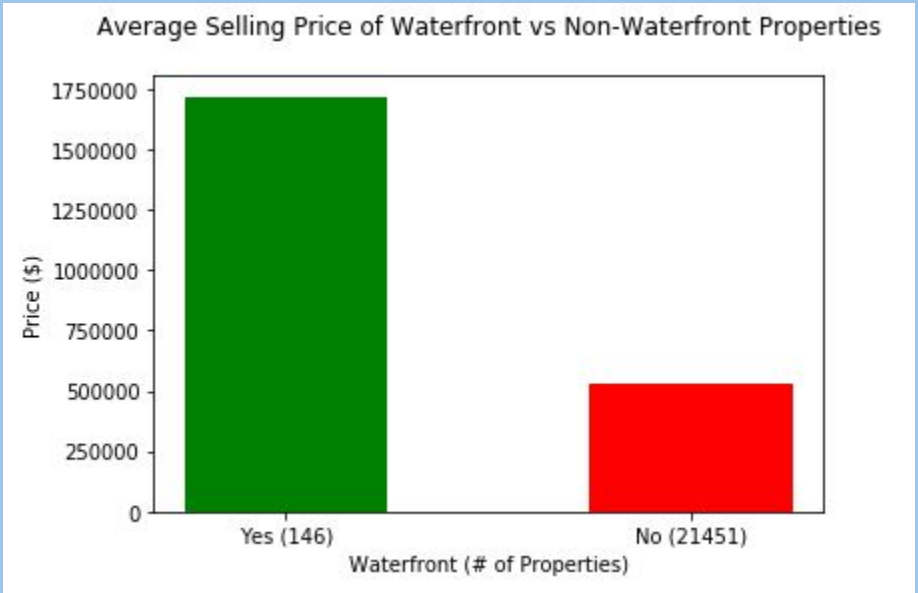
- Location: King's County, WA (Seattle)
- Sell dates: May 2014 - May 2015
- Age: Built in 1900 - 2015
- Price range: 78k - 7.7M
- Square Footage:
  - Living: 370 - 13,540 sq ft
  - Lot: 520 - 1.6M sq ft

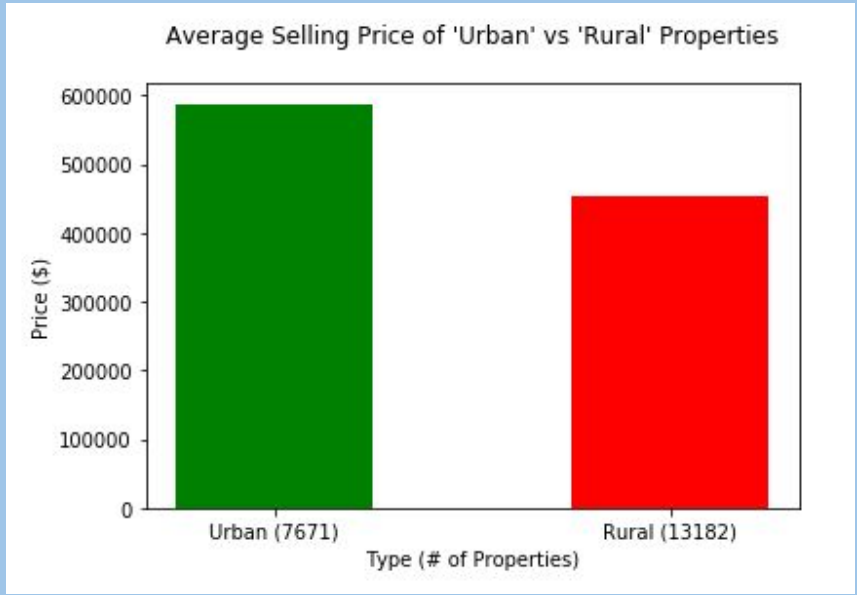
Price Prediction Modeling:
Getting Started

- DATA: the more, the merrier!
- Homogenous property datasets yield higher prediction accuracy: urban, rural, waterfront
- 2 methods:
  - Individual - mapping housing price changes over time
  - Market Snapshot - single reference point comparing property prices with differing features (this project!)

Exploring Our Data

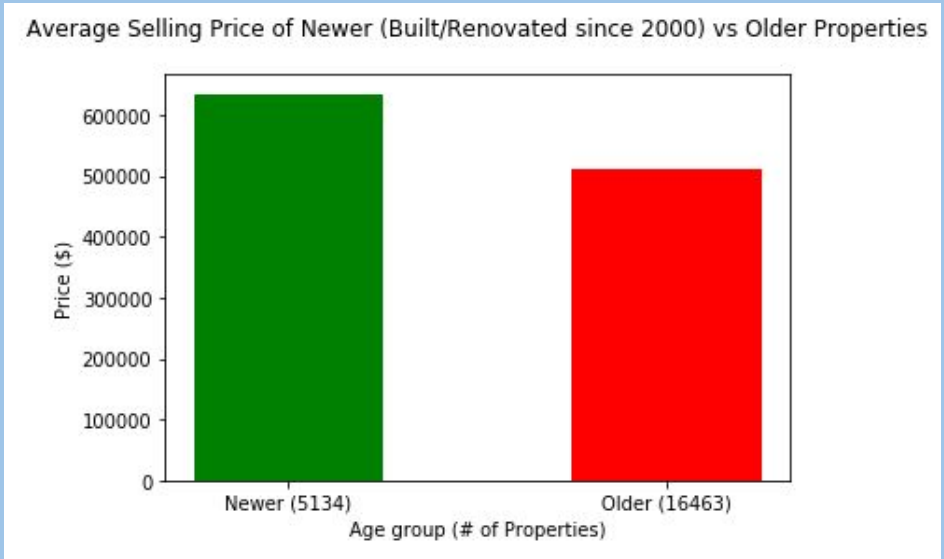
- Significant average selling price differences between waterfront, urban and rural property types

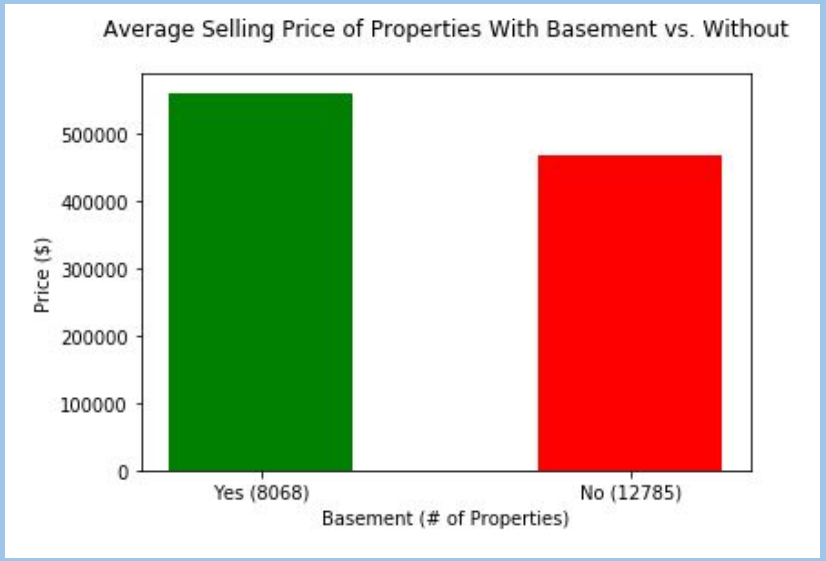




Exploring Our Data

- Newer (built/renovated since 2000) sells for ~$125k more, homes with basements go for ~$95k more





Exploring Our Data

➢ Notable Findings:

★ **Waterfront** properties: **+$1.1M** higher sales price

★ **Urban (vs Rural)**: **$135k** higher price and ~14 years younger

★ Is **newer** better? **+$125k** higher sales price*

★ **Basement** included? **+$95k** higher.. but ⅔ did not have one

★ **Hot** housing market? (May 2014 - May 2015)

- 176 properties were sold twice, 1 sold 3 times!
- Average profit: **$135k**
- Most common renovation year: 2014 (hello house flippers!)

*Built/renovated in the last 15 years (since 2000)

Final Regression Model Results

How did our model fare predicting prices based on a given set of features?

Original Model: Homes up to $1.6M

- Average error of $170k with accuracy of 89.5%

Updated Model: Homes up to $1M

- Average error of $130k with 91.6% accuracy

Recommendations:

More Data + Features = More Accuracy!