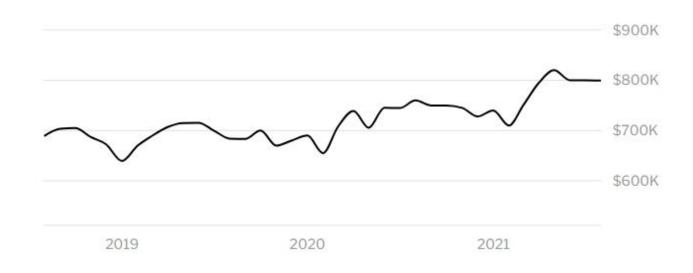
How much would a house sell?

Predicting House Sales Price

Introduction

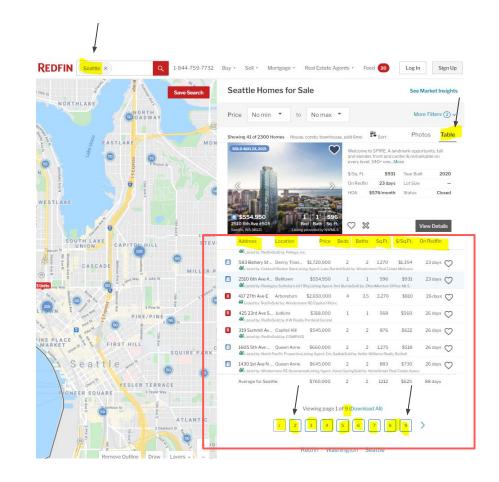
Seattle housing market is HOT, houses are selling fast, and at a higher price.





Data

- Web-scraped from Redin.com search result
- Combined four regions from great seattle area
- House sold within last 6 months
- Single family house, townhouse and condo
- 1334 sold houses



Framework

- Outcome: predict the sales price for a house based on existing features
- Features:
 - o Bed
 - Bath
 - Sqft
 - Location
 - Region of Great Seattle Area

Methodology

Baseline Model

+ Feature engineer

+ Regularization

R² 63.3%

Not overfitting

Model: Linear Regression

Input: Numeric feature only

R² 73.1%

Overfitting

Input:

1. Add interaction terms:

bed/bath

2. Bucket location data, add as dummy variables ★

3. Add deviation feature: sqft

in region 🜟

R² 73.3%

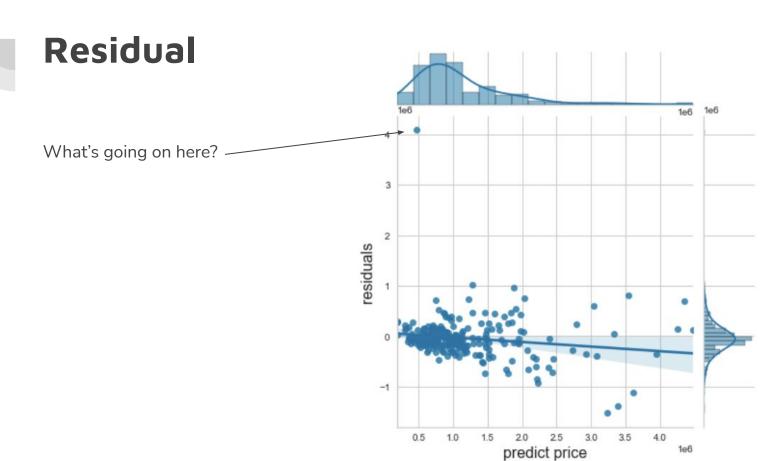
Reduce overfitting

Model: Lasso Regression

Result

R² 73.3%



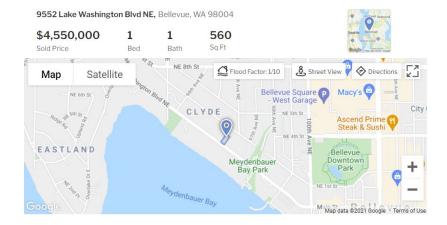


Worst Prediction

- 1. Predict price is 1/10 of actual price.
- 2. Poor Data Quality:
- Wrong bed, bath number, sqft
- Land miscategorized as SFH
- 3. Great view: not captured in data
- 4. Great location: close to downtown
- 5. Top school district



Listing provided courtesy of COMPASS, Sold by Realogics Sotheby's Int'l Rity,



Good prediction

- Predict price: \$624,912 (-0.6%)
- Townhouse
- Newly developed area
- Recently sold property



Listing provided courtesy of Windermere Real Estate/PSR Inc. Sold by ZL Realty, LLC.

13331 NE 137th PI, Kirkland, WA 98034

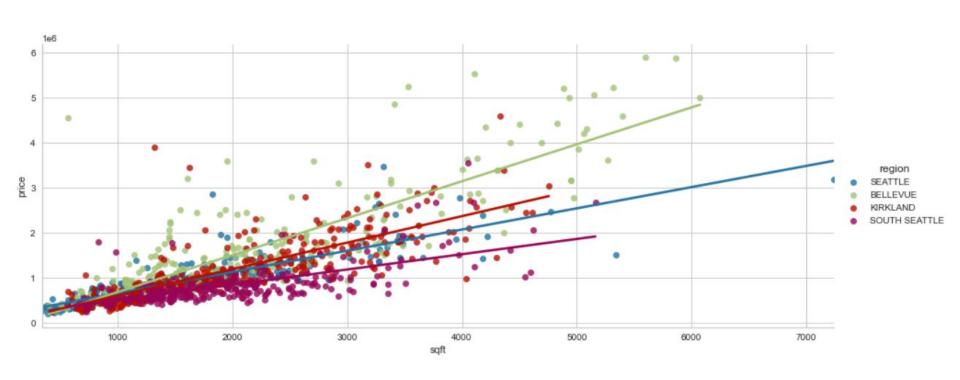
\$629,000 2 2 1,140 Sold Price Beds Baths Sq Ft



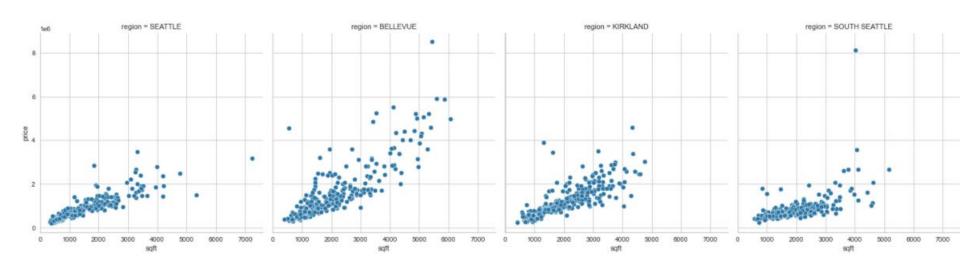
Future work

- Improve data quality
- Increase volume: expand web scraping strategy to collect more sold house data
- Explore other features like has_view, type of property, HOA fee, school district

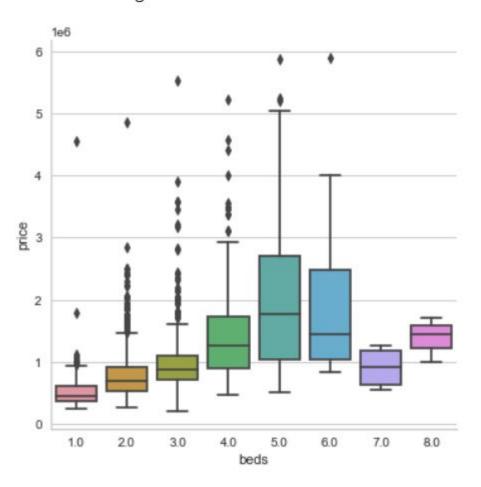
Appendix - Price Trend Per Region







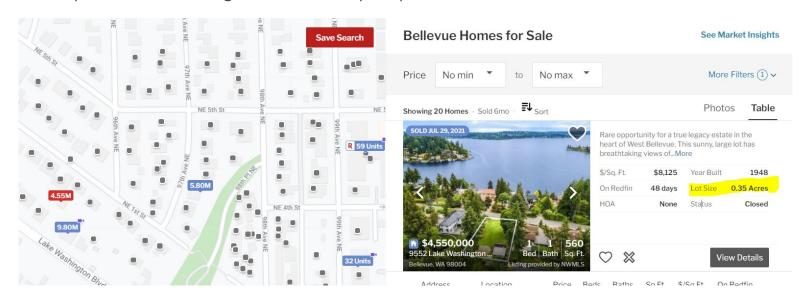
Appendix Price vs Beds



Likely A Data Entry Error

Square foot is wrong: should be 15246 sq ft.

All key features are wrong, which leads to poor prediction.



Appendix - Pair

