Housing Market Analysis - King County

James Benasuli & Geoff Vogt

Agenda

1. Business Problem
2. Data Understanding
3. Exploratory Data Analysis
4. Baseline Results
5. Iteration to Final Model
6. Final Results
7. Recommendations and Next Steps

Business Problem -

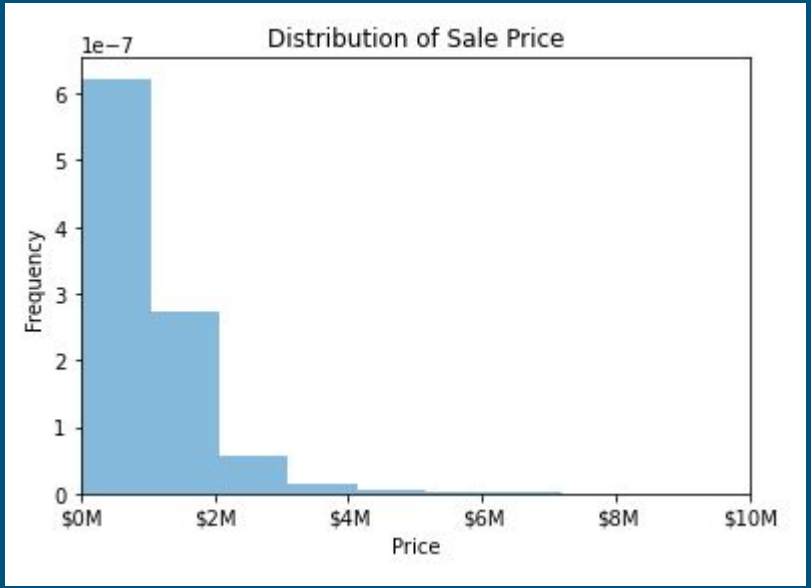
Imagine there was an easier way for Centurion Appraisal to estimate housing prices using an algorithm to assist in the process of property valuation. By analyzing the features of the property, the appraiser can estimate the expected market value of the property and use that information to produce a more accurate appraisal report.

Note that Centurion only appraises homes greater than $250k and less than $15M.

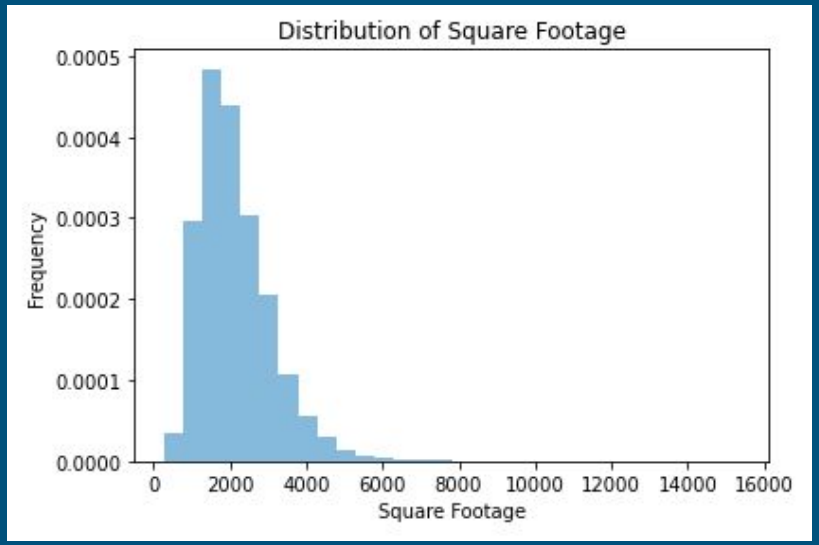
Data Understanding

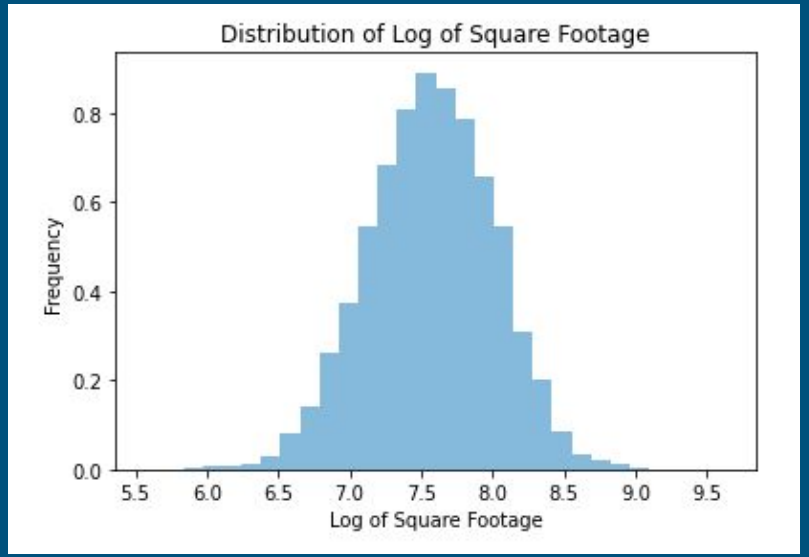
- King County house sales data 2021 - 2022
- Target variable: Price
- Washington State 2021-2022 Report Card Data
- Location data from Google Maps for Amazon, University of Washington, and Microsoft
- Features of home include square footage, floors, bathrooms and condition

Sale Price Distribution



Square Footage Distribution

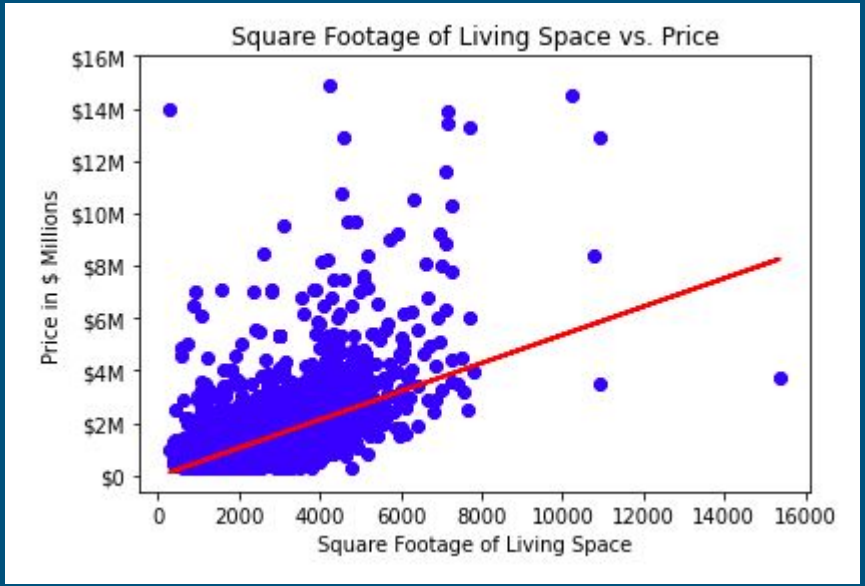




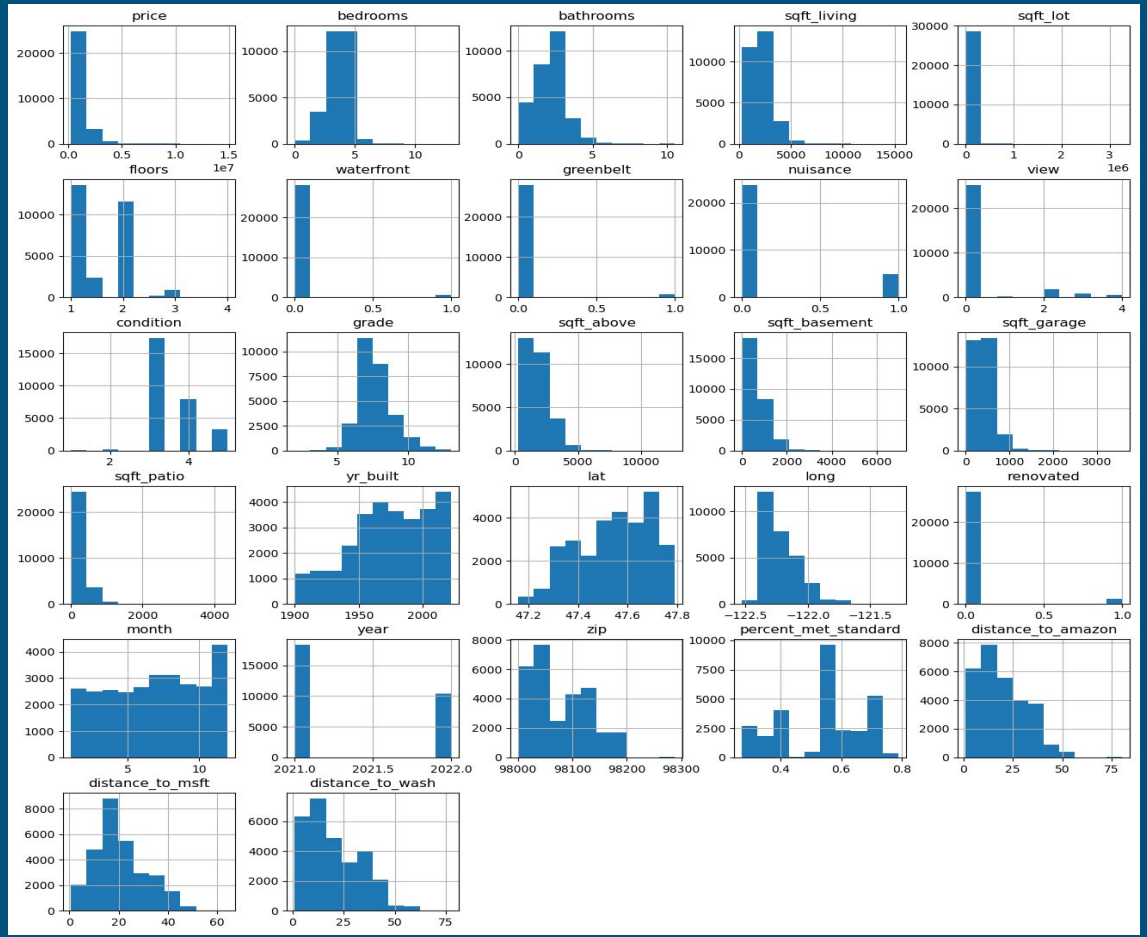
Baseline Results

Baseline model is square footage of living space due to highest correlation coefficient for predicting price of home

R-Squared: 0.3882



Feature Distributions Before Transformation



Iteration to Final Model

- Model #2: Looking at the relationship between all original numeric features vs. price including education data and distance to key landmarks
  - R squared: .555
- Model #3: Power transformed (log) values for price of home
  - R squared : .734
- Model #4: Power transformed all square footage variables
  - R squared: .735
- Model #5: Encoded all categorical variables including ZIP codes, heat and sewer data and standard scaling
  - R squared: .793

Final Model Results

- R-Squared of .793
- Low (nearly zero) p-values of for all variables besides sewer system, ZIP and nuisance (neighborhood noise) categories
- Mean squared error of 0.2018
- Coefficients with greatest impact:
  - Square footage of living space: every 1 increase in log of Square Footage (living) results in 0.2 increase in log of Home Price
  - Distance to University of Washington: every 1 standard deviation increase of Distance to UWash results in 0.63 decrease in log of Home Price
  - Grade of the home: every 1 standard deviation increase of Home Grade results in 0.144 increase in log of Home Price

Conclusion

Recommendations

- Home appraisals are significantly impacted by the size, status of home health and distance to UWash - be aware of these key factors!
- If there is a discrepancy between home valuation and one of these factors, you can highlight this home as needing more hands-on analysis for that home

Next Steps (issues we didn't have time to tackle)

- Localized crime data
- Look at home data beyond years 2021 - 2022
- Add visualizations for final model expected vs. actual
- Troubleshooting the log un-transform
- Look at distinction in statistical significance for homes > 10 miles away from our key landmarks


+6.31%
THAPK YOU!
+5.27%
+6.31%
+7.12%
-3.54%