King County Sales Assessment

Assembled by Harrison Carter

How can we determine high value investment areas?

Where to start?

Correlation Matrix for Features -0.017 0.0012 0.0052 -0.012 -0.13 0.019 0.022 -0.012 -0.0082 -0.0018 0.021 -0.0027 -0.14 0.31 0.09 0.26 0.054 0.31 0.083 price - -0.017 -0.053 0.022 bedrooms - 0.0012 0.032 0.18 0.018 -0.15 -0.01 0.39 0.031 0.16 bathrooms - 0.0052 0.76 0.088 0.69 0.051 0.024 0.088 sqft living - -0.012 0.35 0.32 0.056 -0.2 0.052 0.76 0.17 0.88 0.76 0.18 sqft lot - -0.13 0.09 0.032 0.088 0.17 1 -0.0048 0.053 0.0045 -0.13 -0.086 0.14 0.72 0.019 0.26 0.18 0.35 -0.0048 0.0035 -0.06 0.049 -0.011 0.28 sqft above - -0.011 0.88 0.18 0.022 -0.26 -0.0012 0.69 0.42 0.73 0.2 yr built - 0.022 0.16 0.32 0.053 -0.23 -0.35 -0.15 0.071 0.051 0.0045 0.0035 -0.230.032 -0.072 -0.00025 0.0039 yr renovated - -0.012 0.018 0.07 -0.56 -0.28 zipcode --0.0082 -0.053 -0.15 -0.2 -0.2 -0.13-0.06 -0.26 -0.35 0.07 0.27 -0.15 0.052 0.049 lat --0.0018 -0.01 0.024 -0.086 0.049 -0.0012 -0.15 0.032 -0.14 -0.086 0.022 0.23 0.13 0.34 -0.072 -0.56 -0.14 0.34 0.021 0.13 0.22 0.26 sqft living15 --0.0027 0.39 0.76 0.14 0.28 0.73 0.33 -0.00025 -0.28 0.049 0.18 0.72 -0.011 0.0039 -0.15 -0.086sqft lot15 -0.031 0.088 sqft_above sedrooms yr_renovated

- 1.0

- 0.8

- 0.6

- 0.4

- 0.2

- 0.0

-0.2

-0.4

Additional Predictors

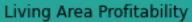
Living Space

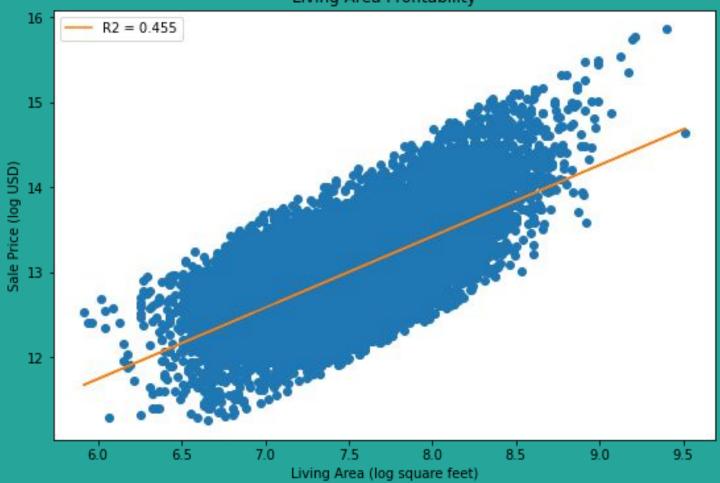
Waterfront Availability

What matters?

What do we have in our data that might help explain pricing?

What is not inherent to zip code?





Examine ZIP code

Rationale:

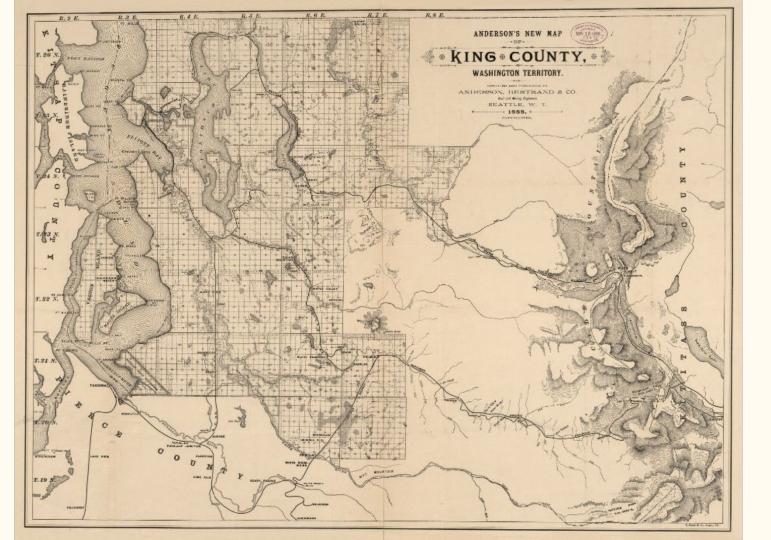
- Spatial
- While unintentional, ZIP may, in effect, include other predictors
- Geographical
- School districts
- Natural features, parks (waterfront)
- Shopping
- Transportation

Methodology:

- Categorical: nominal
- Incorporate as primary predictor
- Logarithmic transformation on price
- May contain other predictors, check heteroskedasticity upon multiple regression.

Baseline:

- Baseline model R2 = 0.533
- Skew = 0.594 (> 0.5)
- Kurtosis = 5.132 (<6)



Refining the model

Final Predictors:

ZIP code

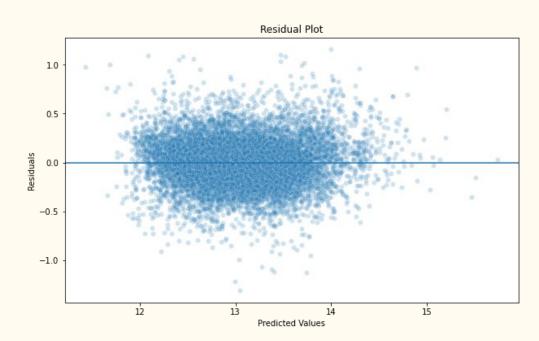
Living Space

Waterfront availability

*See Appendix A for equation

Waterfront availability incorporation

- Decreases Skew:
 - o 0.445 to 0.130
- Decreases Kurtosis:
 - o 5.920 to 4.956
- Increases R squared:
 - o 0.817 to 0.833



Further Inquiry



- Map geographic location with respect to landmarks (spatial reconstruction)
- Consider construction regulations and housing specifications (build quality, hazards)
- Consider school district (income inequality)
- Consider effects of tourism
- Consider non-residential development

Image credit: wta.org

Questions

References

- King County Assessor Website
 (https://info.kingcounty.gov/assessor/esales/Glossary.aspx?type=r)
- Library of Congress (https://www.loc.gov/resource/g4283k.la001374a/)
- Washington Trails Association
 (https://www.wta.org/go-hiking/hikes/snow-lake-1)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \dots \beta \Box x \Box$$

y =predicted price

 β_0 = predicted price when no waterfront and living space is zero

 β_1 = predicted change in price for each unit of change in living area square footage

 β_2 = predicted change in price between waterfront/non waterfront properties

 $\beta_3...\beta\Box$ = predicted change in price between ZIP codes, individual basis

 x_1 = Living area square footage

 x_2 = Waterfront availability, categorical

 $x_3...x \square = \text{ZIP codes}$, individual basis, categorical

*All y and beta values in equation are predicted