



Understanding
King County
Housing Prices

STEPHEN LANIER

PROBLEM STATEMENT

- What factors affect the price of a house in King County?
- An **understandable** model

Project is focusing on inference (understandability) over prediction (how accurate can we get the model?) for the sake of increasing domain knowledge of the King County housing market.

BUSINESS VALUE

- Deeper understanding of housing market
 - Factors with highest affect on value? Lowest?
 - Factors with positive affect on value? Negative?
- A model for predicting a house's value

Want a model that clearly shows which factors have the biggest effects on a house's price, and are those effects positive or negative?

Additionally, it won't be the best model for raw predictive power, but it will be a pretty good model for predicting a house's price.

METHODOLOGY

1. Clean and reshape data
2. Find baseline model
3. Remove factors (but retain model accuracy)

1. Some feature engineering: convert basement square footage to 'has a basement (Y/N)'. Convert year renovated to "has been renovated (Y/N)". Etc.
2. A metric for improvement. Used cross-validated RMSE at every step, for every model for comparability across the process.
3. Fewer features to consider, easier to understand. Used the Lasso over Recursive Feature Elimination.

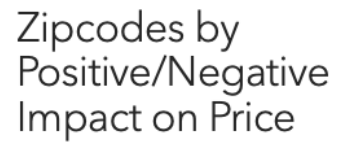
FINDINGS

- Highest impact factors:
 - Zip code
 - Square footage
 - Presence of a basement
 - Grade
 - Waterfront location
 - Renovation

'Grade' is a number grade received from King County Housing, 1 Poor, 13 Excellent.

'Waterfront Location' is Y/N.

'Renovation' is Y/N.



6

CONCLUSIONS

POSITIVE EFFECT

- Basement
- Renovation
- Grade of 9 or higher
- Waterfront location
- Zip codes near Seattle, Bellevue

NEGATIVE EFFECT

- Grade of 7 or lower
- Zip codes south of Seattle

SCALING EFFECT

- Square footage of house

See Appendix for coefficients the Lasso produced. A positive coefficient means that feature increased a house's value; a negative coefficient decreased a house's value.

Future Work

- Prediction-focused model
 - Higher accuracy, less interpretability
 - A “black box” model

Thank You!

Citations and Acknowledgments

- https://gis-kingcounty.opendata.arcgis.com/datasets/e6c555c6ae7542b2bdec92485892b6e6_113/data (List of King County Zipcodes)
- <https://github.com/OpenDataDE/State-zip-code-GeoJSON> (zipcode GeoJSON data)

Appendix 1: Positive Impact Factor Coefficients

feature	coefficient						
zipcode_98004	0.521011	zipcode_98102	0.255622	grade_11	0.103897	renovated	0.0463937
zipcode_98039	0.508023	zipcode_98115	0.25031	view_3.0	0.0839275	floors_1.5	0.0268665
zipcode_98112	0.420499	zipcode_98117	0.245765	zipcode_98136	0.0805921	view_1.0	0.0186502
zipcode_98119	0.355919	view_4.0	0.233761	grade_9	0.0721661	month_sold_4	0.0146636
zipcode_98105	0.324765	zipcode_98107	0.230771	zipcode_98052	0.0715896	zipcode_98053	0.0113463
zipcode_98109	0.307728	zipcode_98033	0.220859	zipcode_98008	0.0713831	zipcode_98007	0.00856
waterfront_1.0	0.292281	zipcode_98122	0.202291	zipcode_98006	0.0672796	grade_12	0.00804484
zipcode_98040	0.279702	zipcode_98116	0.17351	zipcode_98144	0.0671594	zipcode_98029	0.00686909
zipcode_98199	0.271987	grade_10	0.124267	condition_5	0.0507265	basement	0.00276861
zipcode_98103	0.256475	zipcode_98005	0.108034	view_2.0	0.0496268	zipcode_98177	0.00242109
						sqft_living	0.000221837
						sqft_living15	0.000114527

Coefficients produced by the Lasso.

Appendix 2: Negative Impact Factors

feature	coefficient					
condition_3	-0.0174655	61	grade_6	-0.185186	zipcode_98030	-0.277594
zipcode_98106	-0.0240696	55	zipcode_98178	-0.186794	zipcode_98032	-0.302324
zipcode_98148	-0.0244981	34	zipcode_98058	-0.197554	zipcode_98042	-0.30372
condition_2	-0.0359858	32	zipcode_98055	-0.1983	zipcode_98003	-0.311501
zipcode_98059	-0.0363614	26	zipcode_98038	-0.204068	zipcode_98092	-0.313628
grade_4	-0.0382272	19	zipcode_98022	-0.204984	zipcode_98002	-0.314816
zipcode_98056	-0.0386621	60	grade_5	-0.243892	zipcode_98023	-0.401641
zipcode_98146	-0.0532703	57	zipcode_98198	-0.254735		
grade_7	-0.0883592	23	zipcode_98031	-0.25562		
zipcode_98188	-0.18085	53	zipcode_98168	-0.263408		

Coefficients produced by the Lasso.