King County House Sales

Did Zhaleh get a deal on her house?

Predicting Housing Prices

- King County house sales: 2014 & 2015
- I bought a townhouse in King County in the year 2015.

Research & Questions

- What can this dataset tell me about my house? Did I pay too much or too little?

- use model to predict price of my house

    - How good is the model?
    - How good was my offer?

Business Impact

- Gage market & offer fair value
- Realtor’s acumen
- Clean dataset
- Find best features & update model with recent data

Methodology

- Histograms & Correlations
  - Investigate outliers & missing values
  - Probability distributions to bin missing values
- Domain knowledge
  - King County
  - Housing
- Stepwise Feature selection
- Multi-Linear Regression Analysis

Results

Model Name	R^2 - Train	R^2 - Test	RMSE - Train	RMSE - Train	My House Error
Model 1	70.11%	70.68%	$200,767	$192,274	$ -285,275
Model 2	71.03%	71.61%	$197,664	$189,197	$259,973
Model 3	**71.13%**	**71.89%**	**$197,328**	**$188,269**	**$234,248**
Model 4	71.05%	71.76%	$197,590	$188,696	$261,002

Model 3 Features: zip code, sale date, square feet of living for nearest 15 neighbors, square feet of the lot, num of bathrooms, bedrooms, quality of construction, age of renovation, waterfront

Location Location Location!!!

Top Features	Contribution
98004 (Bellevue - west of 405)	$852,887
Waterfront property	$1,033,744
98039 (where Bill Gates lives)	$1,706,286

There's a reason for the phrase Location, Location, Location - the top three features have to do with location. Number of bedrooms is 54th out of 60 features on the list.

Future Work

- Build separate model for luxury vs. single family home model
- Find more nuanced location features
  - School district
  - Neighborhood name
  - Lat/Long binning
- Try different combinations of features
- Use other machine learning algorithms that do not assume normal distribution of features, or can accommodate correlated features

Thank you!