King County Housing: Pricing Insights

Data Cleaning and Modeling by:
Aaron Chen and Dirk Van Curan

Business Questions

As a company in Bellevue, can you purchase affordable housing for your C-Level employees while keeping their commute under and hour?

Is waterfront property worth the investment?
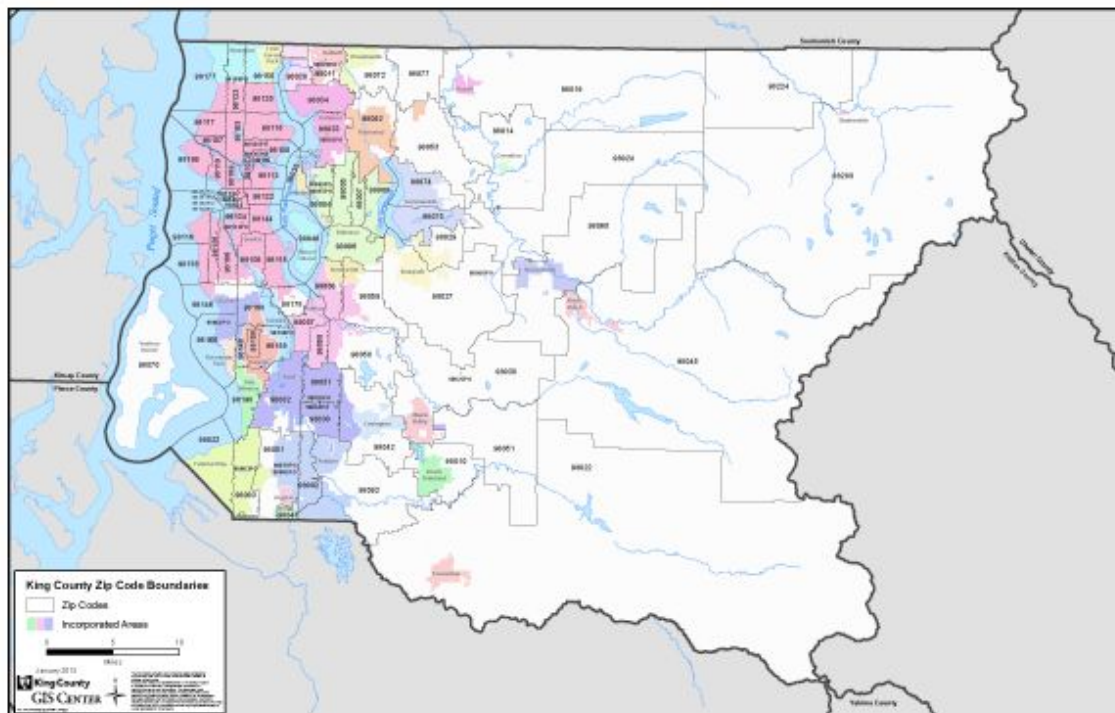
Does the King County assessor's grade affect price?
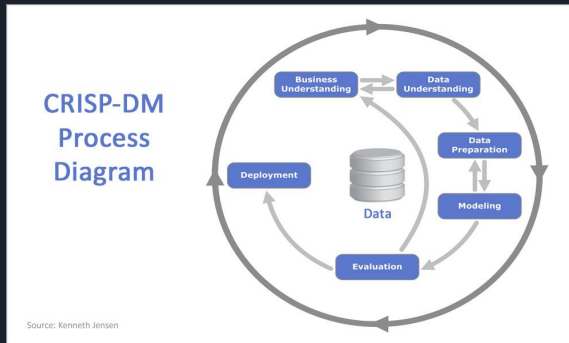
Data Set Used

King County Housing Sales

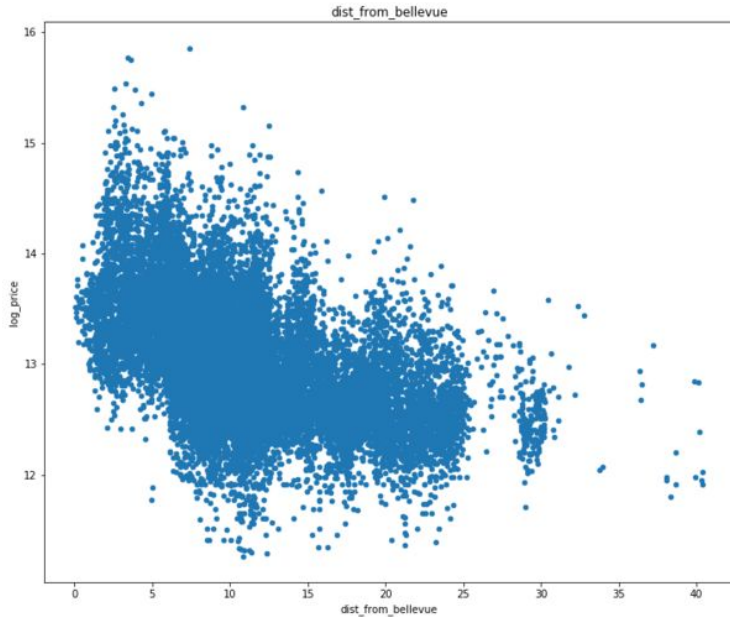Sale information for houses sold between May 2014 and May 2015

Model Approach

1. Clean data - eliminate errors and remove columns that don't add value to solve business problems.
2. Transform columns and generate new columns to fit model.
3. Use Linear Regression and Lasso function to help reduce error.
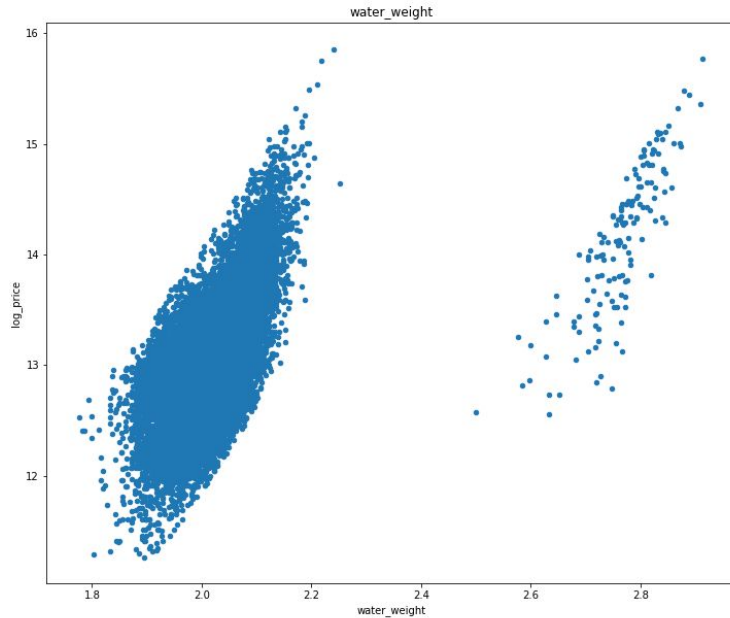4. Refit model for accuracy

CRISP-DM Process Diagram

Business Understanding

Data Understanding

Data Preparation

Deployment

Data

Modeling

Evaluation

Source: Kenneth Jensen

	0	1
id	7129300520	6414100192
date	10/13/2014	12/9/2014
price	221900	538000
bedrooms	3	3
bathrooms	1	2.25
sqft_living	1180	2570
sqft_lot	5650	7242
floors	1	2
waterfront	NaN	0
view	0	0
condition	3	3
grade	7	7
sqft_above	1180	2170
sqft_basement	0.0	400.0
yr_built	1955	1951
yr_renovated	0	1991
zipcode	98178	98125
lat	47.5112	47.721
long	-122.257	-122.319
sqft_living15	1340	1690
sqft_lot15	5650	7639

How does Distance from Bellevue affect Price?

- Distance from the heart Bellevue does see a steady decrease in home prices.
- As the model uses linear distances the prices take a significant step down after approximately 7.1 miles, the distance from Bellevue to the Puget Sound.
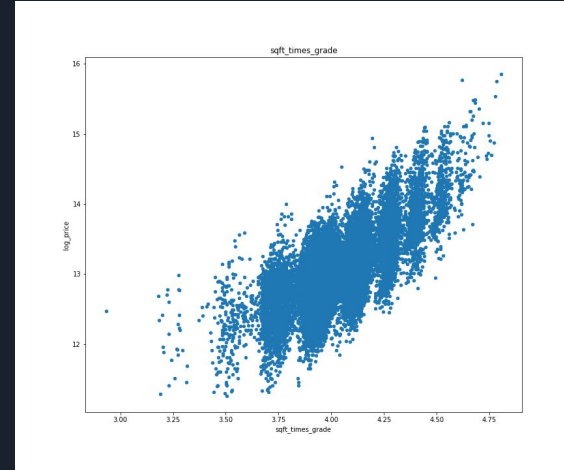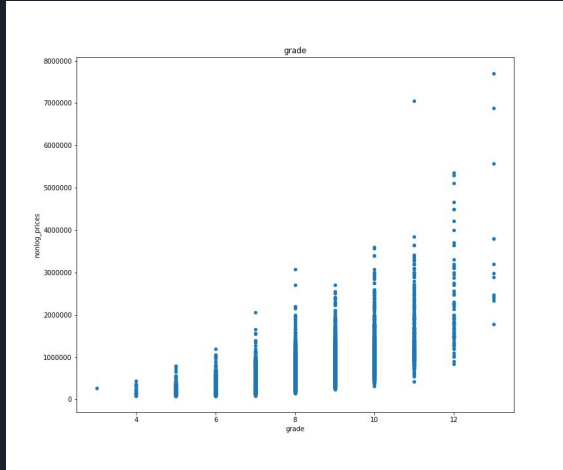
Waterfront property

Is it worth it?
1. Yes.. but why?
2. Price certainly has increased when using square footage as the multiplying factor to waterfront.
3. Waterfront provides a baseline higher price, as indicated, than any property without.

What does Grade Tell Us about Price?

Grade, though important as a predictor, was difficult to quantify by itself as it was an arbitrary value given by the assessor of the property at time of sale. By using its predictive power as a multiplier for other values, however gave better insight on the importance of grade in relation to sale price.

Next Steps

- Improve model by seeing how data from more years worth of sales affects the predictive power of our variables.
- Develop a better scale for particular variables to better predict price (i.e. waterfront).