



# King County House Sales

Methodology and Predictions



Made by George Bennett

## What I learned from the dataset.

I found the most useful pieces of information to make predictions with are the:

- Amount of square foot living space
- Zip Code
- View on a scale of 1 to 4
- Whether or not the house is on the waterfront

# Challenges with the data.

- The data contained many null values that had to be dealt with.
- Some of the data was unusable because It did not follow the requirements for linear regression.
- In the end these challenges were overcome and the model did good, having an r-squared value of 83.9%.

# My methodology for creating a model to predict house prices.

- I used the OSEMN process to create the model.
- This model is a multiple linear regression model.
- I was able to use noncontinuous data such as the zipcode to good effect.
- I reiterated through the project making sure the model was accurate as possible.

# Validating the model

- I validated the model using holdback validation.
- The model performed very well with little error difference between the train set and test set.
- The overall error was low, signaling that the model is doing its job well for making predictions.

Questions and Comments.