Predicting Sale Price for Houses in King County
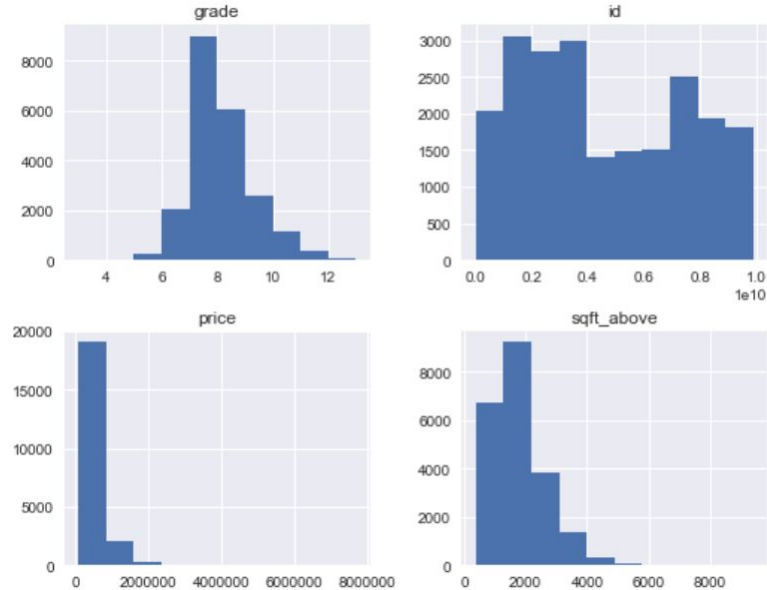
David Cuervo

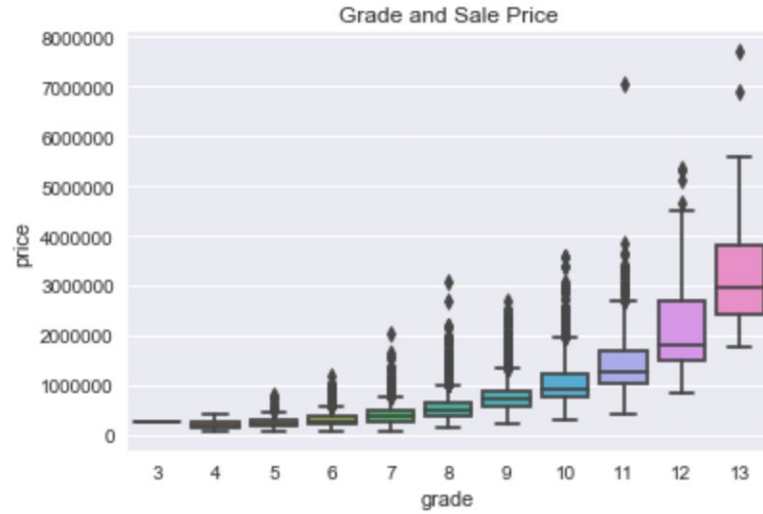# Overview

I. Methodology
  A. Understanding the data
  B. Exploring through the data
  C. Developing a regression model

II. Results
  A. How can we use this information?

Understanding the Data

Which Variables Would Best Predict Sales Price?

Building a Regression Model

The final model includes:

- Number of bathrooms
- The number of times the house has been viewed
- The overall grade given, based on the King County grading system
- Waterfront status
- Square footage of the nearest 15 neighbors

Out[45]:

OLS Regression Results

Dep. Variable:	price	R-squared:	0.541
Model:	OLS	Adj. R-squared:	0.540
Method:	Least Squares	F-statistic:	5080.
Date:	Tue, 04 Feb 2020	Prob (F-statistic):	0.00
Time:	17:45:46	Log-Likelihood:	-2.9899e+05
No. Observations:	21597	AIC:	5.980e+05
Df Residuals:	21591	BIC:	5.980e+05
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-7.889e+05	1.24e+04	-63.436	0.000	-8.13e+05	-7.65e+05
bathrooms	5.288e+04	3003.468	17.607	0.000	4.7e+04	5.88e+04
view	8.028e+04	2493.208	32.198	0.000	7.54e+04	8.52e+04
waterfront	6.501e+05	2.24e+04	29.057	0.000	6.06e+05	6.94e+05
grade	1.34e+05	2315.076	57.870	0.000	1.29e+05	1.39e+05
sqft_living15	84.6912	3.631	23.324	0.000	77.574	91.809

Omnibus:	18507.122	Durbin-Watson:	1.963
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1851789.543
Skew:	3.617	Prob(JB):	0.00
Kurtosis:	47.783	Cond. No.	2.78e+04

Results

- ## R-Squared of 0.541
  - ### 54.1% of the variance in sales price can be explained by the features in the model
- ## Coefficients
  - ### Grade: For each increase in grade point, the price of a house increases by $134,000 on average
  - ### Waterfront: If the house is a waterfront property, the price increases by $650,100 on average

Thank you

Questions?