King County Housing Analysis
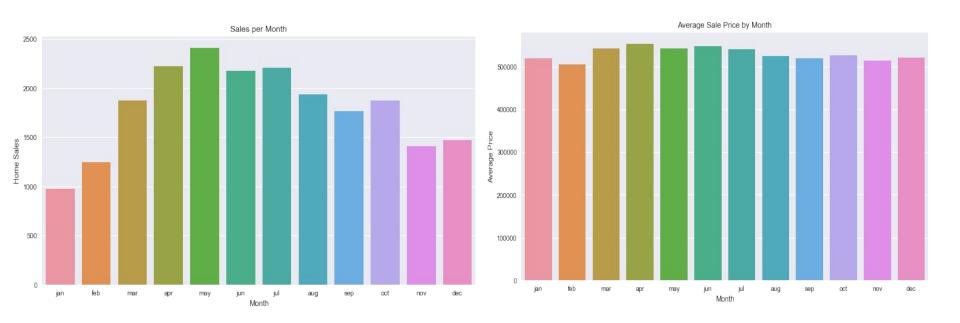
by Jonathan Bardey

Data Science Process

OSEMN model

Source: A Taxonomy of Data Science

- Data: Information about 21,597 houses sold in King County, Washington.
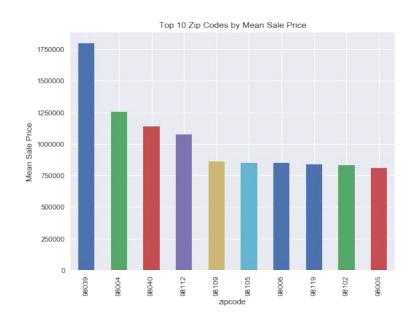- Model: Multivariate Linear Regression

Effect of Date on House Sales

Sales per Month

Average Sale Price by Month

- Sales range from May 2, 2014 to May 27, 2015.
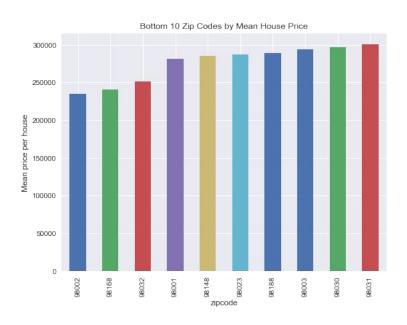- Avoid the winter months

Bathrooms or Bedrooms?

Bathrooms & Bedrooms vs. Price

- Additional bathrooms provide greater increase to sale price than bedrooms.

Location: Mean Sale Price by Zip Code

Top 10 Zip Codes by Mean Sale Price

Bottom 10 Zip Codes by Mean House Price

- Big difference between top 10 (> $750K) and bottom 10 (< $300)
- Explore cost of land as well using sqft_lot

Most Important Feature: Size

Square Footage of Living Space vs. Price

- Square footage (sqft_living) is the most highly correlated feature to Price and one of the strongest predictors (coeffcients) in our model.

Conclusion

- Our model explains 85% of the value of housing sales price with variables of square footage, zip code, bedrooms, bathrooms, etc.

- Future work could include clustering houses by location (zip, long, lat) and modelling within those clusters instead of over the region as a whole.

Thank You!