ASSESSMENT OF PRICE DRIVERS IN KING COUNTY

WHY WE'RE HERE

Purpose of the project: *to find the best factors, from a large collection of factors, to help us predict the prices of homes in the King County area.*

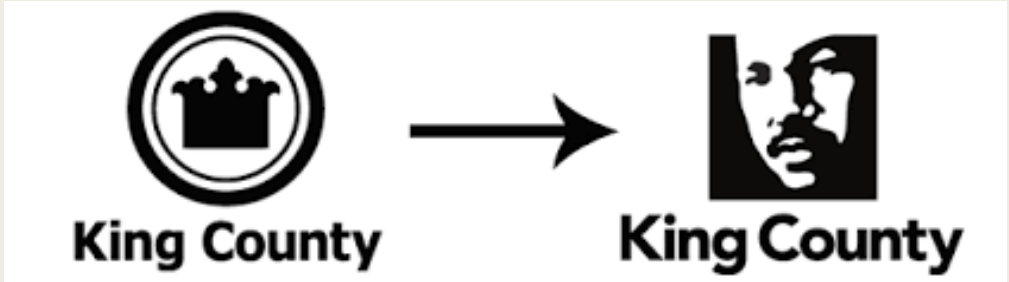
Source: A Taxonomy of Data Science

METHODOLOGY

- We're using the OSEMN model of data analytics; very common among data scientists.
- The process is not quite finished, however.
- A model that hasn't been explained isn't a complete model.
- You will be able to take the findings of the analysis and implement them as you see fit.

KING COUTY AT A GLANCE

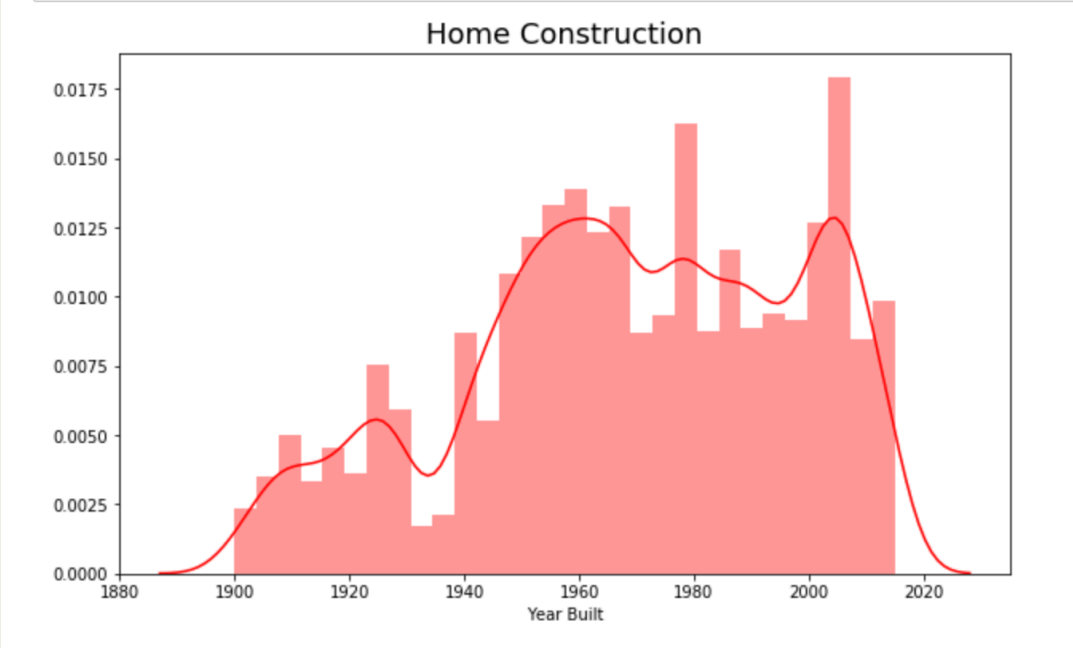
- Founded – 1852
- Renamed after MLK, Jr. – 1986
- Est. Population. – 2,233,163 *(2018)*



- Median Household Income - $78,800
- Largest Industry – Information/Technology
- Census Median Home Value
  - 2000 – $235,000
  - 2016 – $407,400
- 2018 Net New Residential Units – 14,970
  - Single Family – 2,610
  - Multifamily – 12,360

OUR DATA SET AT A GLANCE

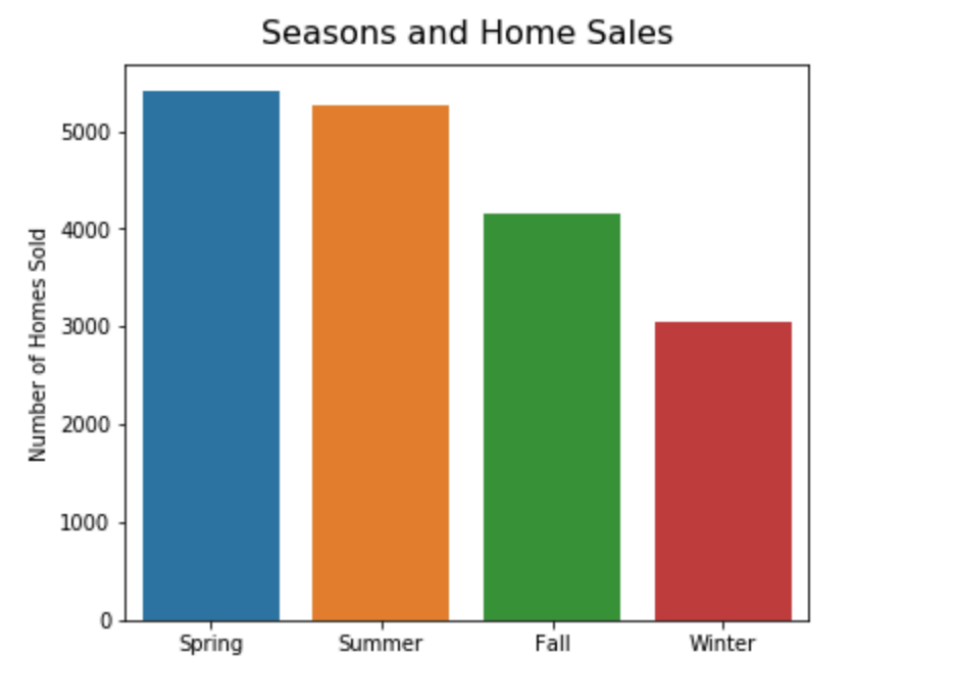
- Over 20,000 observations—a very healthy sample size
- Median Home Value: $450,000
- Median Home Square Footage: 1,910
- Some features included:
  - *Square footage*
  - *Year home was built*
  - *Grade given by K. County*
  - *# of beds & baths*
  - *Zip code*
  - *Latitude & longitude*



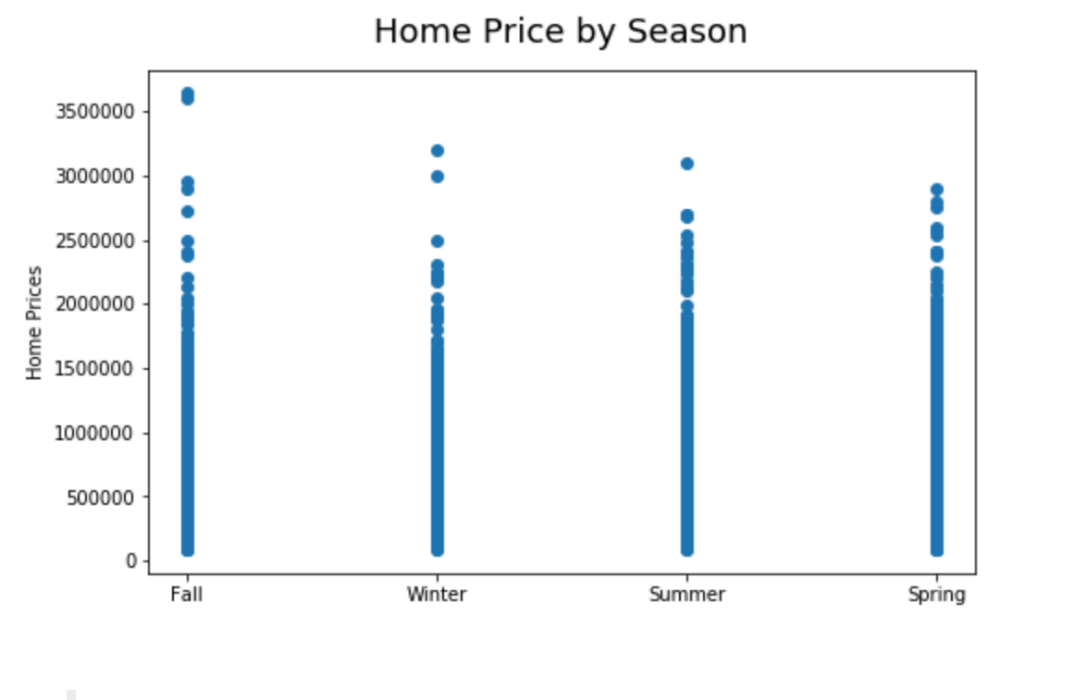
Home construction ranges from 1900 – 2015, with consistent growth from the mid 50's to mid 60's, and brief spikes in construction around 1980 and 2010.

A NOTE ON SEASONS

Common sense would tell us that the time of year would have a substantial effect on our data set. And it does. But maybe not in all the ways we would expect it to. Let's investigate below.



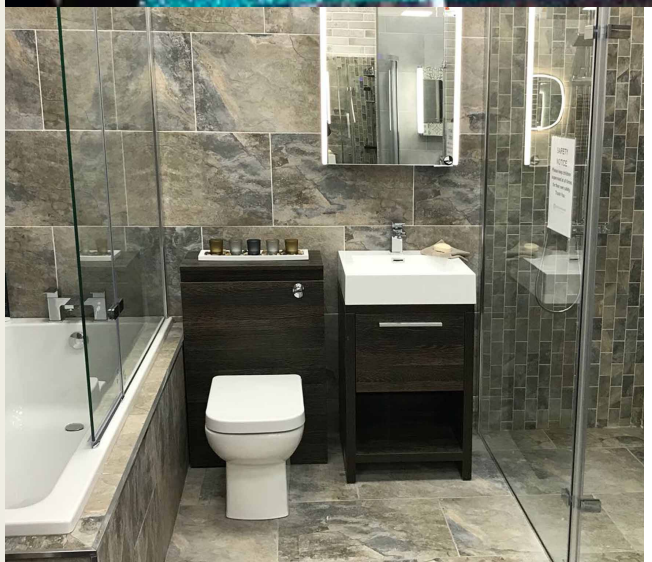
There is clearly a higher number of homes sold in the warmer months.

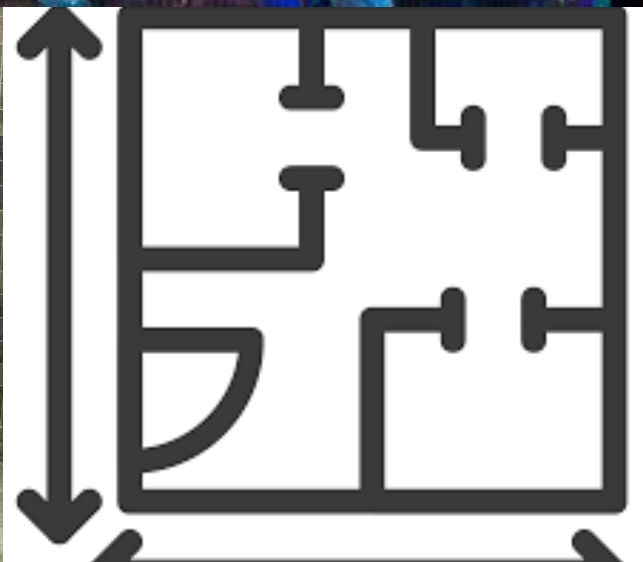


But with the exception of a few outliers, the time of year doesn't seem to have a large effect on price.

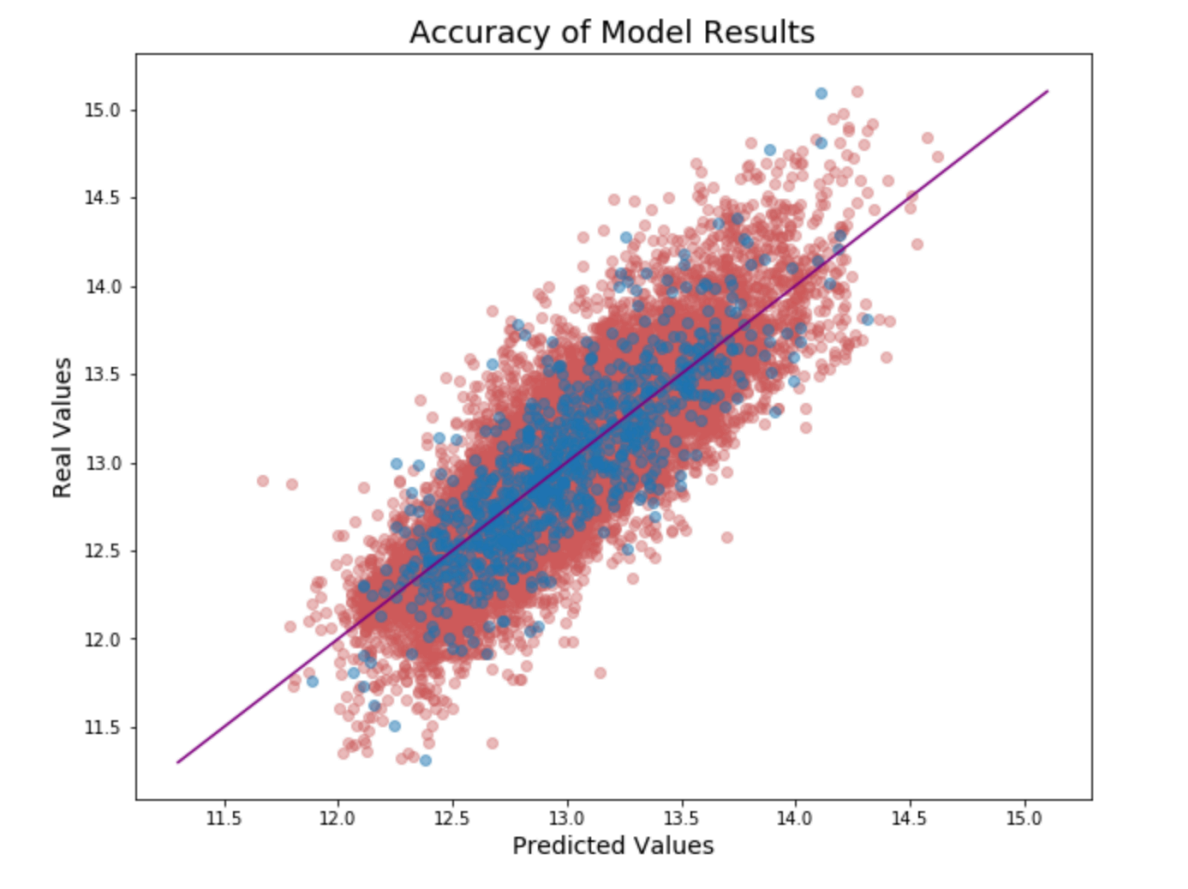
RESULTS — FEATURES

- Most Significant Features:
  - *Grade*
  - *Bathrooms*
  - *Square Footage*
- When we crunched the numbers we got a “R^2 score” of .71. This means our model can explain ~ 71% of the cost of the homes





RESULTS — ACCURACY



One way of checking the soundness of your model is to plot your actual values against your predicted values. The closer to a 45 degree angle they are, the better.

In this example, the red dots represent the data we trained our model on, and the blue are the data we tested our model on.

The fact that they both follow a nice 45 degree trend, as well as aligning with each other, bodes well for the accuracy of our model.

THANK YOU!

- Questions?
- Comments?

- Parker Lumkes, data scientist in-training
- plumkes@gmail.com