Proposal

This document describes my proposal for Springboard's capstone 1 project (CP1). Specifically, this document summarizes the:

- Business problem
- Data
- Data science problem
- Deliverables

Business problem: What drives the price of residential homes?

For consumers, the price of their residential home has a major impact on their life. This price massively impacts the consumer's net worth, income, and cash flow. To buy and sell a home intelligently, consumers should understand what drives the price of residential homes.

With a better understanding of these price drivers, consumers might:

- Make better buying and selling decisions
- Have more-informed conversations with real estate agents around the price
- Make smarter decisions regarding what sort of investments to make in their homes

Data

I will use data from the Kaggle competition, House Prices: Advanced Regression Techniques. This competition uses the Ames Housing data set. The Ames Housing data:

- Focuses on the sale of about 3,000 residential properties in Ames, Iowa between 2006 and 2010
- Includes the sales price of each property in dollars
- Includes about 80 variables describing different aspects of the home, such as the number of rooms, number of bathrooms, year built, the size of the home, etc.
- Is already divided into a training and test set, each of which includes about 1,500 observations (aka, a 50/50 split)
- Is considered to be the modernized and expanded version of the Boston Housing data set, which we work in some Springboard exercises.

Data science problem

My initial data science problem is to:

- Identify a few features that have a strong, independent impact on the price
- Predict prices based on a few key variables using linear regressions
- Predict price based on key variables using an extended model, such as a random forest

Deliverables

The deliverables for this project are:

- A notebook focused on the initial data assessment as required by mini project 5.4.5
- A notebook focused on the exploratory analysis as required by mini project 7.2
- A notebook focused on inferential statistics as required by mini project 8.5
- A notebook focused on linear regression, with hyperparameter tuning (regularization parameter), L1 and L2 regularization when applicable
- The milestone report on as per unit 8.5
- A game plan for the extending the model
- The first draft of the CP1 Final Report
- The final deliverables - CP1 Final Report, Presentation Slide Deck, and cleaned-up code