CS 4824 Project Proposal

1. There is a competition on Kaggle that asks you to predict a home price based on several features. Normally when you think of a home it's about generic details; "the kitchen needs to be big", "I like vaulted ceilings", "I want a garage". This dataset provides the general information with specific details. For example, you probably don't think about the height of your basement when describing your dream house. Well this data does and takes that into account.
   Since this asks for a single numerical output based on the data I will use a supervised model with regression. I plan to test the data on multiple types of regression and compare the results to find the best model.
2. The training data consists of 1460 home samples with 79 data points for each home and the testing data consists of 1459 home samples with 79 data points for each home. The data points consist of numerical and categorical figures.
   URL: https://www.kaggle.com/c/house-prices-advanced-regression-techniques/overview
3. Looking around at different types of regression I plan to use "Random Forest" since it is said to handle categorical, numerical, and binary features without scaling or normalization therefore cutting down on computation time. Luckily this training set is pretty small so it shouldn't take long to run through the model.
4. Milestones/Dates:
    a. **2021-11-12:** Submit this proposal
       Short gap here to finish the final problem set.
    b. **2021-11-23:** Finish sanitizing the data for duplicates and errors
        i. I could do this manually but I plan on building a function to do it automatically
    c. **2021-11-24:** Feed in the data to several graphing functions to check for similarities and get a first look into what I will be working with
        i. I'll most likely use some sort of bar graph to check the average prices and I will compare each of the features to see which ones are included in the higher/lower priced homes.
    d. **2021-11-25:** Confirm the desire to use Random Forest and begin coding
        i. Confirmation is just a step to research more on Random Forest/Ensemble Models
    e. **2021-11-29:** Have at least the pseudo-code written
        i. Creating a skeleton of the project is important to get a scale of things and a larger picture
    f. **2021-11-30:** Write interim report
    g. **2021-12-01:** Finish and submit interim report
    h. **2021-12-02:** Finish implementation of Random Forest
        i. I hope to have this finished before a vacation on December 3
    i. **2021-12-06:** Begin writing final report and finishing touches on code
    j. **2021-12-08:** Finish writing final report and submit to Canvas