Project Proposal

1. I will be working alone.

2. Description of how your chose your project topic, why it is important to you, and how you got feedback about your idea from others

I will be using the Board Game Geek (BGG) API and database to build a model to predict the ratings of unrated and new boardgames.

One of the major issues with using BGG to find new games is that games have a weighted rating based on the number of votes. Games with less than 100 rating votes will be officially rated around 5.5. This is to prevent ‘brigading’ new games, or artificially increasing or decreasing the rating (typically using bots). This has the unintended side effect of keeping less popular games ‘underground’ with lower ratings than they may deserve.

Why not use the unweighted average? A game could have unbiased ratings from a small community, but the game might only be appealing to that small community. So, we can’t trust the rating to apply to all board game players.

If I can accurately predict a rating using metrics that are not community driven and/or provided by the developer, I can estimate the unweighted rating of games with few ratings, or even new games.

3. Description of your proposed project

* Research question of interest (what are you hoping to learn)
  + Can I accurately predict a new or ‘underground’ games’ rating?
* Where are you planning on getting the data
  + BGG XML API and the associated CSV
* The project description page lists 8 machine learning related tasks of which you are required to have at least 4 (C-level), 5 (B-level) or 6 (A-level) of these elements in your project.  While this is not set in stone, give me an idea of the elements you are planning on using
  + Feature Engineering – Web scrape number of files, dev participation, etc.
  + 4 supervised models – KNN, Linear Regression, SVM, Random Forest, etc.
  + Deep-learning model – Haven’t learned yet
  + Cross-validating model parameters – I will likely use a Random Search CV
  + SHAP feature importance – Haven’t learned
  + Apply Dimension-Reduction – Haven't learned
* Create a timeline for completing the project keeping in mind the following deadlines:
  + Data collection/read in by Oct 24
  + EDA complete by Oct 26
  + Finish majority of Modeling by Nov 9
  + Rough Draft complete by Nov 16
  + Final Draft complete by Dec 7