Capstone Project 1 Proposal - Dan Feldman

This project seeks to answer the question of whether or not it is possible to determine the factors that lead to the success of a board game, given its various attributes (game type, number of players, play time, complexity, etc.). In recent years, there has been a resurgence in the number of board games being created by companies big and small, even by individual developers. Having knowledge of what makes a successful game, as well as identifying underserved niches in the community would help guide developers when making game-building decisions.

As such, my client would be a game developer or developers who would want to identify if there are trends that make for popular games, especially if there are some popular niches that are underserved in the current market. With my analysis, they would build their new board games to serve the community and therefore maximize their likelihood of having a successful game (and by extension, profits). Similarly, board game sellers could use this knowledge to predict the success of a new board game, deciding beforehand if it's a good or bad idea to offer a new board game for sale.

In order to accomplish this analysis, I will use the board game statistical dataset at boardgamegeek.com, a website dedicated to the niche of board gaming. They have an XML API that allows for download of the statistical data and reviews of all board games in their archives. I also aim to download review data from Amazon to clean and combine with the boardgamegeek data for additional robustness of the sample.

Once I have the data from both sources, I will create a scheme for combining their reviews into one dataset (one source is on a 1-10 scale, while the other a 1-5 scale), and then correlate these data to the board game metadata, which includes game time, genre, complexity, number of players, and more. Correlating the various together should help elucidate trends between successful and unsuccessful games. I can also separate out the data from each source to see if there is a difference between the markets. More specifically, Amazon will capture reviews by players not interested in board games enough to use a website specifically designed for hardcore gamers — are there different trends among different player audiences? This is something we also wish to address.

My deliverables for the project will be the code, an archive of the available data for reproducibility, and a paper describing our findings. These will all be available on my Github in its own repository.