

William Dew's Capstone 1 Proposal

Board Game Geek Blog

Board Game Geek (BGG) is the biggest board game forum for hobbyist and contains a database of reviews, images, and ratings. People or companies may join BGG for free and join in conversations about different games, post questions, or rate their favorite games. There is a lot of data on the site about each game.

Problem

With hundreds of new games coming out each year it is hard to not have a game get lost in all the noise and fanfare. How can a fan find games that they will enjoy but haven't received the hype and publicity that other games have gotten? Can you predict a games popularity by the mechanics, theme, and player count?

Data

The data that will be used is from Kaggle. (<https://www.kaggle.com/mrpantherson/board-game-data>). It was collected from BGG in March of 2017. The data contains values like minimum and maximum player count, minimum and maximum play time, year board game came out, the number of votes the game received from the community, recommended age, game category, designer, year the game come out, among others.

One problem with the data is some categories are user given such as weight, and rating. Others will be on the board game box but could be entered wrong. These are things that must be taken into account when exploring the data. It isn't full of hard facts but we must assume most categories are correct.

Method

Data needs to be cleaned up because of missing values and columns that will not help with the analysis.

Find the columns with the greatest correlations.

Using models to train and test algorithms to find the model for the best prediction.