The Board Game Renaissance: An Exploration of What Makes Board Games Popular

Context

Board games have long been a part of human culture with the historical records dating back as early as 2000 B.C.1[1] In recent years, board games have experienced a resurgence in popularity. Tabletop gaming is now a burgeoning industry which had an estimated market of approximately 7.2 billion USD in 2017 and is expected to increase by 4.8 billion USD by 2023[2]. The US card and board game market alone is predicted to increase to 5 billion USD by 2025[3]. With such growth we're clearly in a renaissance or arguably even a golden age of board games. Understanding what makes a game popular is important for game developers to determine what kinds of games they should focus on producing.

Data Wrangling

The first step is of course to get the data needed. I've scraped ranking data from the game ranking list on BoardGameGeek.com for 19,019 games which will be used in this analysis. As indicated in its name, BoardGameGeek is not just a database of board games and ratings, but is a place for everything board game related. The site has forums to discuss games and rules, resources such as translations and player aids, and even a board game aftermarket. This data is current as of 6/12/20.

I used the scrapy package to first crawl the rankings list for any game with a Geek Rating. From there I extracted the URLs for each game's page and created another crawler to scrape the data for the individual pages. This crawler was not able to extract all of the data I wanted and I ended up using selenium to gather the remaining data. After that, the data was cleaned, with manual corrections made for missing data or bad data points and finally merged into one data set.

We'll cover the majority of the features, but if you wish to explore it on your own or to see the full description of attributes please visit https://data.world/sambeardsley/boardgamegeek.

1 https://www.newyorker.com/culture/culture-desk/what-we-learn-from-one-of-the-worlds-oldest-board-games
2 https://en.wikipedia.org/wiki/Tabletop_game_industry
3 https://www.statista.com/statistics/1072762/us-card-and-board-games-market-value/

Exploratory Data Analysis

Geek Rating

This first chart shows the distribution of the geek rating. Geek Rating is what determines a game's placement on the Geek Charts (the scraped pages) and this is based on an altered average rating. As found in the FAQ, Bayesian averaging is used to prevent games with few votes from skewing the data. The exact algorithm is unknown, but it is believed by the community that every game receives approximately 100 dummy votes with a 5.5 value. Additionally, we can clearly see the Bayesian averaging in effect with the majority of games having a score between 5 and 6 as well as an incredible right skew. In terms of "all-stars" There are only 14 games with a geek rating > 8 (0.0074% of games), and 374 with a geek rating > 7 (1.966% of games).

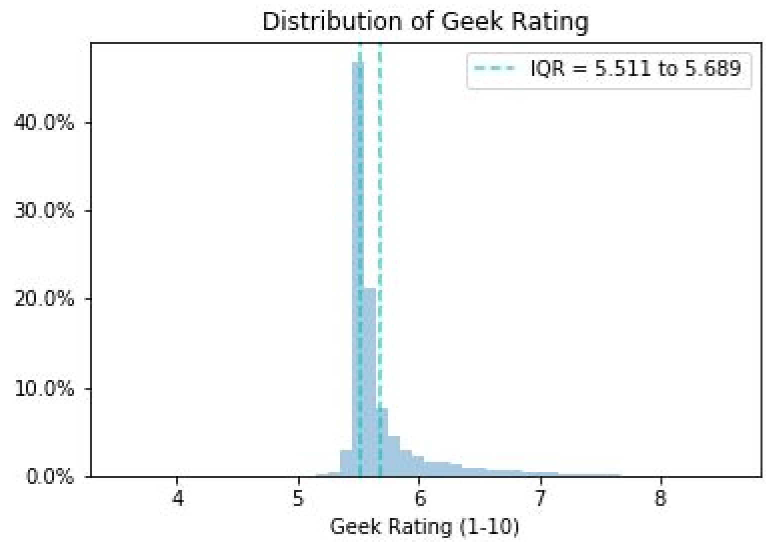

Figure 1: Histogram of Geek Rating

Average Rating

In comparison, average rating, representing raw user rating, shows more distribution among values. Avg rating does not show the same skew and peaks around a rating of 7. This tells us that users tend to rate games positively (>5). The driving factor behind the differences in the rating charts is the number of votes. A game requires 30 votes before it receives a Geek Rating and games without a Geek Rating were not considered in this data set. That said, the median number of votes in the data set is 130, which is barely more than the 100 dummy votes,

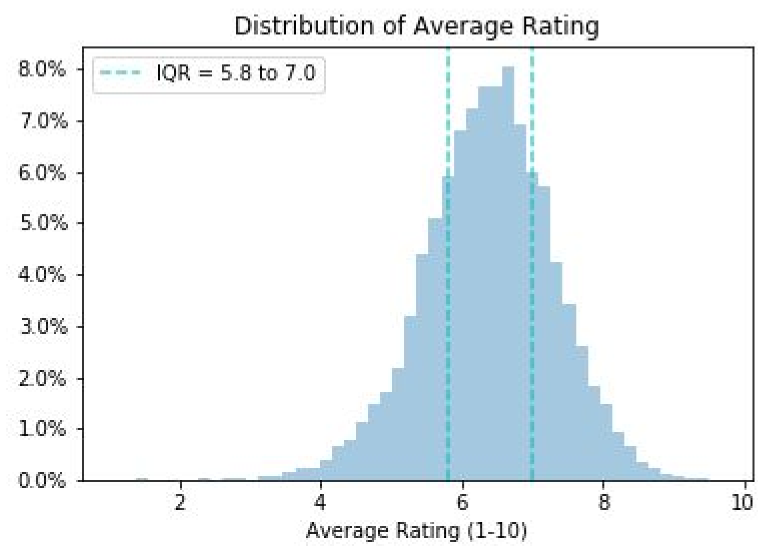

Figure 2: Histogram of Average Rating

resulting in the majority of games having a Geek Rating between 5.5 and 6 despite the distribution of average rating.

Correlation Heatmap

This chart shows the correlation between variables. We can see that the number of votes shows the highest correlation with Geek Rating (as opposed to avg rating which was expected). In essence, the Bayesian averaging represents a hurdle to overcome and a game needs to be popular by avg rating before it can have a good Geek Rating. There are also correlations between age and weight and max play time and min play time.

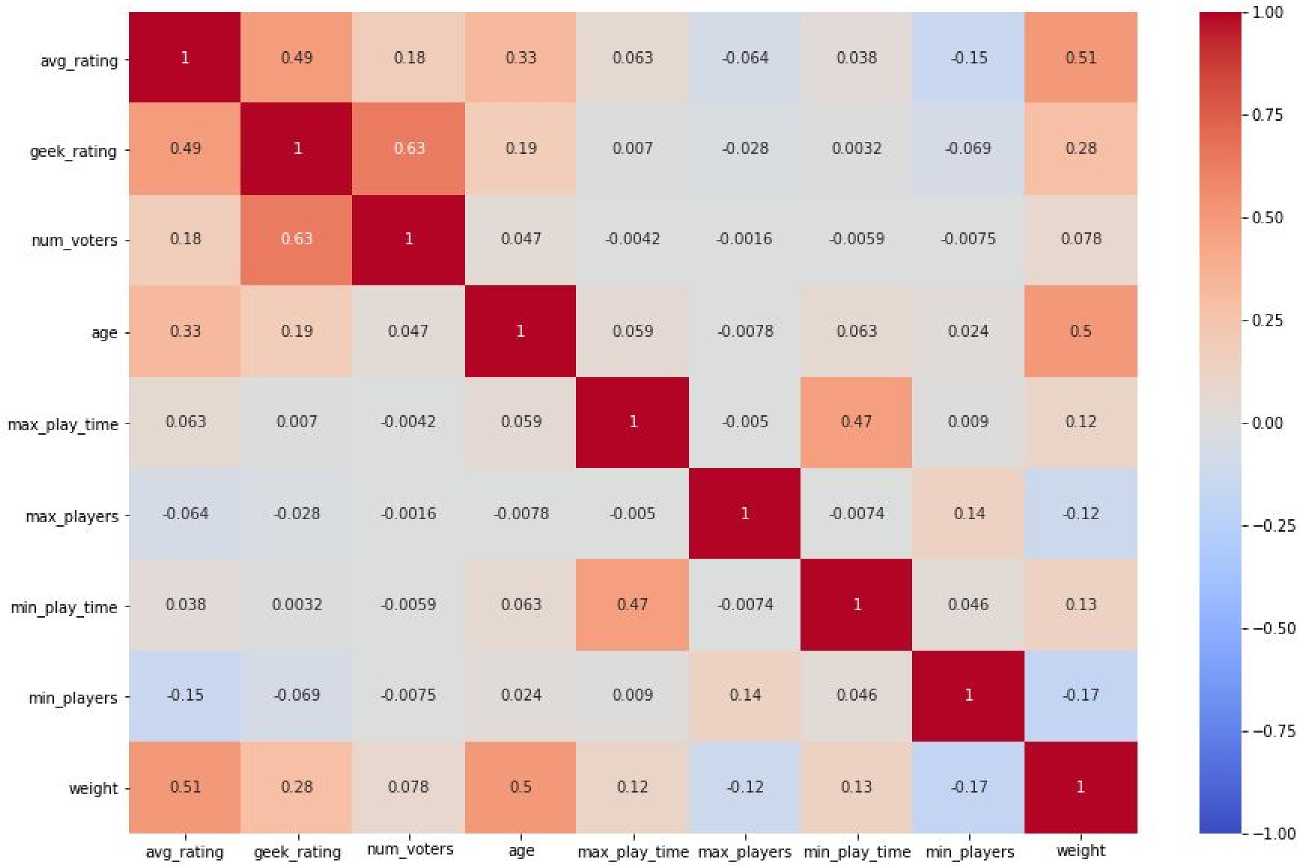

Figure 3: Correlation Heatmap

Weight

Here we see weight plotted against Geek Rating. Weight is a user generated number between 1 and 5 representing the complexity of a game. The mean of weight is 2.04 and 75% of all games are below 2.56. While it becomes more sparse at the higher end of weight, we can see a positive correlation between weight and Geek Rating with a 0.28 correlation coefficient and a small p-value.

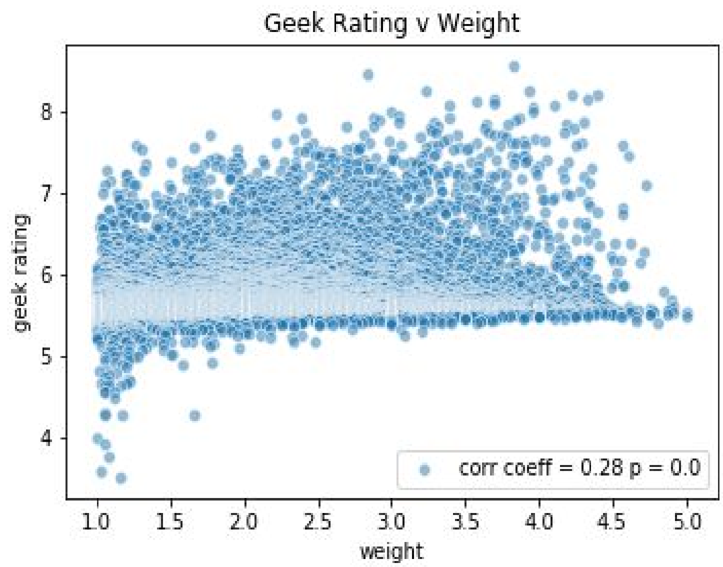

Figure 4: Scatterplot of Geek Rating vs Weight

Age

This plot shows age versus Geek Rating. The mean age is 10, but there are also distribution spikes around 8 and 12-13. Due to age being a discrete variable it is difficult to visualize, but there is correlation with Geek Rating. Geek Rating and age have a 0.19 correlation coefficient with a very small p value.

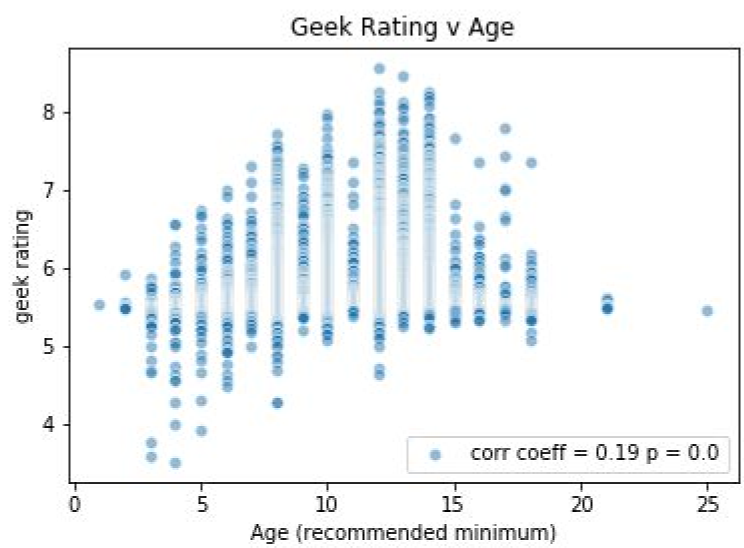

Figure 5: Scatterplot of Geek Rating vs Age

Labels

The scraped data also includes classifications for each game: Category, Mechanic, and Family. Games can be assigned multiple of each.

- Category is a classification for the thematic type of board game e.g. Fantasy or Children's Games. (There are 83 categories.)
- Mechanic is an element or type of game play. Examples include Dice Rolling, Role Playing, and Trading. (There are 186 mechanics.)

- Family attempts to group a game into a broader set of descriptors. There is a large variety in family including items such as Ancient: Greece, Crowdfunding: Kickstarter, and Video Game Theme: Pokémon. (There are 2,748 families.)

Category

Worst Rated Categories: These are the 5 worst rated categories with Trivia in last place. In particular, the Children's Games category being a low rating category aligns with the correlation between rating and age.

category	count	mean_geek_rating
'Math'	129	5.568543
'Electronic'	188	5.566516
'Memory'	477	5.552136
"Children's Game"	1427	5.529584
'Trivia'	515	5.528363

Figure 6: Worst Rated Categories

For fun, let's look at the 10 worst games in the data set:

	title	category	geek_rating	age	weight
19009	Operation	'Action / Dexterity', "Children's Game", 'Medical'	4.486	6.0	1.115
19010	LCR	'Dice'	4.311	5.0	1.050
19011	Monopoly	'Economic', 'Negotiation'	4.299	8.0	1.659
19012	Trouble	"Children's Game", 'Racing'	4.298	4.0	1.056
19013	The Game of Life	"Children's Game", 'Economic'	4.281	8.0	1.177
19014	War	'Card Game', "Children's Game"	4.004	4.0	1.000
19015	Bingo	'Movies / TV / Radio theme', 'Number', 'Party Game'	3.926	5.0	1.048
19016	Candy Land	"Children's Game", 'Racing'	3.779	3.0	1.082
19017	Chutes and Ladders	'Animals', "Children's Game", 'Dice', 'Movies / TV / Radio theme'	3.611	3.0	1.021
19018	Tic-Tac-Toe	'Abstract Strategy', "Children's Game"	3.537	4.0	1.154

Figure 7: The Ten Worst Rated Games

7 of the 10 worst games are in the Children's Games category except Bingo, Monopoly, and LCR (a game that is only about 'winning' dice rolls). We can also see that the weight of these games are

around 1, aside from Monopoly, which still only reaches a weight of 1.659. The majority of these games are determined purely by chance and involve little to no decision making.

A standout category however is Abstract Strategy under Tic-Tac-Toe. BoardGameGeek defines Abstract Strategy as:

- theme-less (without storyline)
- built on simple and/or straightforward design and mechanics
- perfect information games
- games that promote one player overtaking their opponent(s)
- little to no elements of luck, chance, or random occurrence

As a category this fits Tic-Tac-Toe, but there are only a handful of moves a player can take. Conversely, this category also includes games such as Chess (6.9 Geek Rating/3.71 weight) and Go (7.0 Geek Rating/3.99 weight) both of which have sophisticated strategies to them.

Top Rated Categories: And these are the 5 top rated categories. Notably, Farming, Industry/Manufacturing, City Building, and Civilization are all different flavors of a game mechanic commonly known as Resource Management. Game resources are defined as including "tokens, money, land, natural resources, human resources and game points... The game will have rules that determine how players can increase, spend, or exchange their various resources. The skillful management of resources under such rules allows players to influence the outcome of the game."

category	count	mean_geek_rating
'Civilization'	310	6.047019
'City Building'	473	5.993662
'Industry / Manufacturing'	260	5.978542
'Renaissance'	238	5.953256
'Farming'	213	5.931338

Figure 8: Best Rated Categories

Comparatively, we don't see the same consistency with the top 10 rated games as we did with the worst rated games and the Children's Game category. Half of the top rated games, Gloomhaven, Pandemic Legacy: Season 1, Star Wars: Rebellion, Twilight Struggle, and Great Western Trail, do not fall into the 5 top rated categories.

	title	category	geek_rating	age	weight
0	Gloomhaven	'Adventure', 'Exploration', 'Fantasy', 'Fighting', 'Miniatures'	8.573	12.0	3.832
1	Pandemic Legacy: Season 1	'Environmental', 'Medical'	8.472	13.0	2.830
2	Brass: Birmingham	'Economic', 'Industry / Manufacturing', 'Transportation'	8.279	14.0	3.930
3	Terraforming Mars	'Economic', 'Environmental', 'Industry / Manufacturing', 'Science Fiction', 'Space Exploration', 'Territory Building'	8.277	12.0	3.236
4	Through the Ages: A New Story of Civilization	'Card Game', 'Civilization', 'Economic'	8.219	14.0	4.395
5	Twilight Imperium (Fourth Edition)	'Civilization', 'Economic', 'Negotiation', 'Political', 'Science Fiction', 'Space Exploration', 'Wargame'	8.206	14.0	4.220
6	Star Wars: Rebellion	'Civil War', 'Fighting', 'Miniatures', 'Movies / TV / Radio theme', 'Science Fiction', 'Wargame'	8.158	14.0	3.701
7	Gaia Project	'Civilization', 'Economic', 'Science Fiction', 'Space Exploration', 'Territory Building'	8.153	12.0	4.332
8	Twilight Struggle	'Modern Warfare', 'Political', 'Wargame'	8.151	13.0	3.576
9	Great Western Trail	'American West', 'Animals'	8.104	12.0	3.700

Figure 9: The Ten Best Rated Games

There are both high and low rated games across the examined categories, so the takeaway is "not children's games bad, resource management good," but more so the importance of player agency. Games determined by dice rolls or random chance are generally not well received, but games that allow strategy and player choice are.

A perfect example of this is [Scythe](), the 11th highest ranked game, which is an "engine-building game set in an alternate-history 1920s period... Players conquer territory, enlist new recruits, reap resources, gain villagers, build structures, and activate monstrous mechs... Scythe gives players almost complete control over their fate". This was then adapted into a children's game, [My Little Scythe](), which is [My Little Pony]() themed and ranks at a respectable 641, where "players take turns choosing to Move, Seek, or Make. These actions will allow players to increase their friendship and pies, power up their actions, complete quests, learn magic spells, deliver gems and apples to Castle Everfree, and perhaps even engage in a pie fight."

Mechanic

Here are the top 5 rated mechanics which are indicative of a more complex rule set.

mechanic	count	mean_geek_rating
'Turn Order: Pass Order'	7	7.632571
'Automatic Resource Growth'	10	7.308900
'Turn Order: Role Order'	6	7.148500
'Action Drafting'	18	7.142444
'Force Commitment'	7	7.102714

Figure 10: Best Rated Mechanics

Force Commitment: Force Commitment is a mechanic where players choose how many and how to use their units in battle. Both Scythe and My Little Scythe use this mechanic.

Turn Order: Pass Order & Turn Order: Role Order: In games with the Turn Order: Pass Order mechanic, players can choose to take an action or pass their turn. This affects player order in subsequent rounds. While in Turn Order: Role Order games all players secretly and simultaneously make a decision on an action, role, or priority they wish, and this determines how the round is played.

Automatic Resource Growth & Action Drafting: Automatic Resource Growth increases your resources over time. This gives players options to consume their resources or save them to gain more. Likewise, Action Drafting is when "Players select from an assortment of Actions in a shared pool. The available Actions are limited in quantity, and once a player has chosen an Action it may not be chosen again." Both of these mechanics add layers of strategy where players must make determinations between immediate or future payoffs.

We can see Resource Management shows up again with Action Drafting and Automatic Resource Growth, but also game mechanics that allow players to be strategic in how or when they take their turns. The Turn Order mechanics create opportunities for players to play off of each other and make choices based on other players' tactics or between long and short term goals.

These top mechanics are spares, so let's also take a look if there are common mechanics that also have an above average Geek Rating.

mechanic	count	mean_geek_rating
'End Game Bonuses'	54	6.727352
'Race'	59	6.363559
'Communication Limits'	70	6.241571
'Traitor Game'	64	6.219469
'Solo / Solitaire Game'	420	6.164319

Figure 11: Best Rated Mechanics With >50 Games

Solo/Solitaire Game: Starting at the 5th top mechanic, games designed to be or have an option with rules on how to play by yourself are under the Solo/Solitaire Game mechanic. Scythe has this mechanic, but also the number 1 rated game Gloomhaven. Gloomhaven is a game of tactical combat, where players work together to explore and plunder dungeons and ruins in a world that is shaped by the actions and decisions players take during the game. This world persists between games and is designed to be played across many sessions.

Traitor Games & Communication Limits: Traitor Games are team or cooperative games with a betrayal mechanic. The traitor or traitors are unknown by other players, and is given an alternative condition to win the game by subversion. This also ties in well with the mechanic Communication Limits which prevents players from disseminating certain information to other players. Charades (rank 18,824) first comes to mind as an example, but is another mechanic found in Gloomhaven where players secretly make card selections from their hand that determines their actions during combat. Players are only allowed to speak in generalities about what they plan to do. For example you can say "I'm going to strike this monster pretty hard" but not "I'll do 8 points of damage and knock down the monster immobilizing him for the next round". Restricting communication in this way keeps players immersed in the game, limits overplanning and over communication, and allows for unique and surprising situations to arise.

Race: Not to be confused with NASCAR, the Race mechanic is any sort of game when a player wins by accomplishing a fixed goal. Catan (rank 382) uses this mechanic. In Catan players collect resources and build civilizations. Victory points are gained by performing certain actions such as building a city or creating the largest army and the first player to 10 victory points wins the game.

End Game Bonuses: End Game Bonuses work off of a system where players only gain the points at the end of the game. Terraforming Mars (rank 4) uses End Game Bonuses. In this game players take on the role of a corporation and "work together in the terraforming process [on Mars], but compete for getting victory points that are awarded not only for your contribution to the terraforming, but also for advancing human infrastructure throughout the solar system, and doing other commendable things." Once a certain game state is reached Mars is officially terraformed, the game ends, and players are awarded their victory points. Terraforming Mars is also a prime example of a Resource Management game.

Family

I'd be remiss if I did not mention Kickstarter in this analysis. The Crowdfunding: Kickstarter family label has 2,502 members representing 13% of games in the data set and is the most common family with over twice that of the next most common label. Kickstarter is responsible for some of the best rated games, but also some that are not so great.

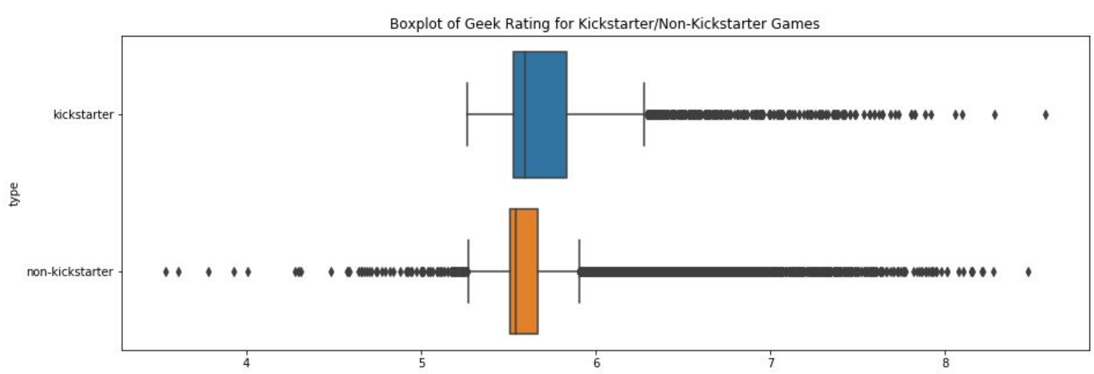

Figure 12: Boxplots of Geek Rating for Kickstarter and Non-Kickstarter Games

The mean Geek Rating for Kickstarter games is 5.78 while the mean Geek Rating for non-Kickstarter games is 5.68 and I ran a t-test to determine if this is statistically significant. The Geek Rating t-test statistic is 13.0 and the p-value is 0.0, so we can conclude that the difference is not up to chance and there is a difference between the mean Geek Ratings for Kickstarter and non-Kickstarter games.

Bias

At this point I want to take a moment to mention bias in the data. BoardGameGeek is for those who love board games and are invested in the hobby and this is creating a self-selection bias. In other words, the type of person invested enough to visit a board game website and rate each game in their collection may not be representative of the general population.

There may also be a similar selection bias with games that require a higher investment. For example, Monopoly can be bought for $20 at your local big box store, but the top 5 rated games all cost over $60. This barrier to entry for some games may inflate ratings as users who play certain games have made financial investments and preselected themselves via this investment.

Finally, returning to Kickstarter, crowdfunding itself is fraught with potentially confounding psychological factors (and more self-selection). Those that invest in a crowdfunding project are not simply motivated by the reward, but also become emotionally invested in the project and its success [Not just an ego-trip: Exploring backers' motivation for funding in incentive-based crowdfunding). We saw that it is no fluke that Kickstarter games have higher Geek Rating; could it be that Kickstarter games are not better, but the act itself of investing in a game creates an emotional connection to the game causing backers to have a more favorable impression of the game than they otherwise would have?

Clustering

In the last part of the analysis I used t-SNE to reduce the data to two dimensions and agglomerative clustering to look at the relationships between games.

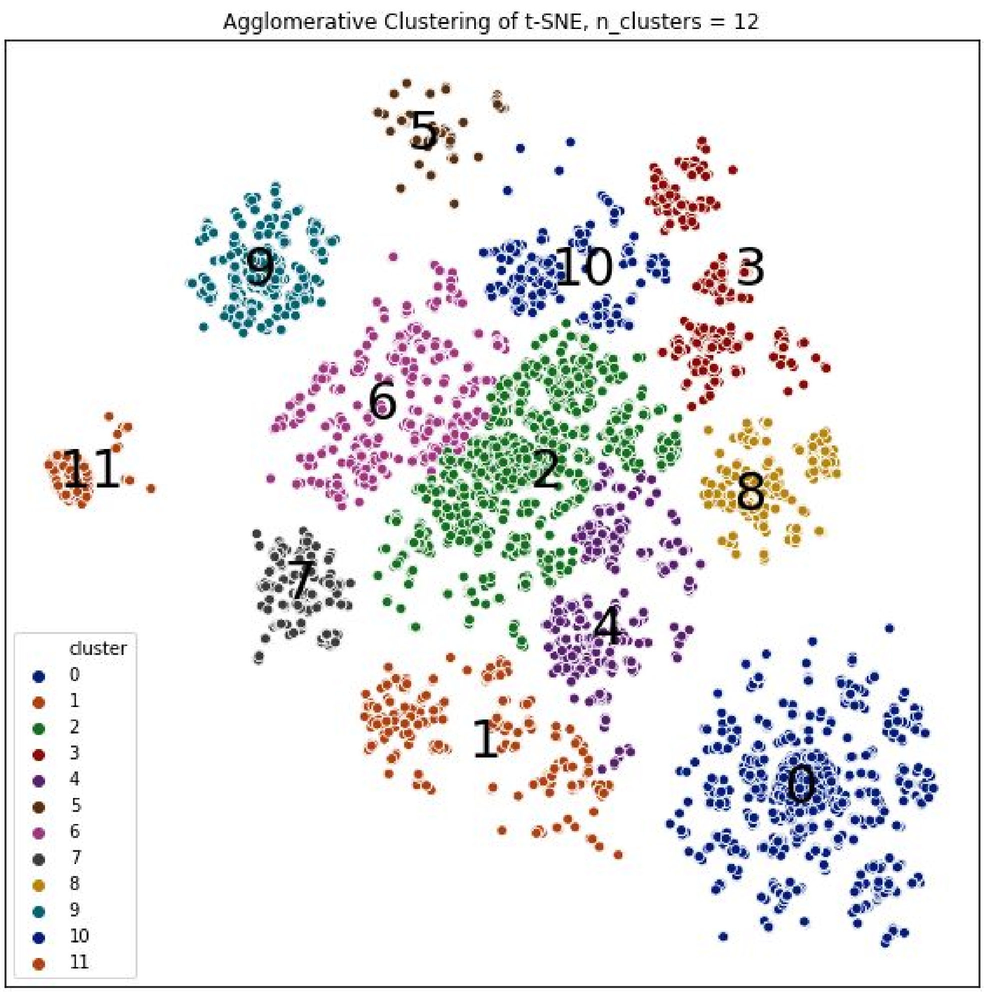

Figure 13: Agglomerative Clustering of t-SNE

I won't include all of the statistics for all 12 clusters, but there are some of note. Particularly, cluster 11 stands out as having the highest mean age, min/max playtime and weight. This is followed closely by cluster 5. On the other hand, cluster 8 has the lowest means for all variables.

	Mean Age	Mean Min Play Time	Mean Max Play Time	Mean Weight
Cluster 11	12	315	462	2.99
Cluster 5	12	160	332	2.84
Cluster 8	7	21	25	1.24
Dataset	10	69	100	2.04

Figure 14: Means of Age, Play Time, and Weight for Notable Clusters

This list outlines the categories and mechanics that dominate each cluster:

- Cluster 0: Card Games
- Cluster 1: Action, Adventure, Fantasy, Fighting & Miniatures
- Cluster 2: Adventure With Dice Rolling & Variable Player Powers
- Cluster 3: Animals, Bluffing & Card Games with Set Collection
- Cluster 4: Dice
- Cluster 5: Thematic Wargames
- Cluster 6: Sci-Fi: Fantasy, Fighting, Medieval, Miniatures, Science Fiction
- Cluster 7: Economic, Negotiation, Trains & Transportation
- Cluster 8: Children's Game & Action/Dexterity with Dice Rolling, Memory, & Pattern Recognition
- Cluster 9: Abstract Strategy With Pattern Building & Tile Placement
- Cluster 10: Ancient, Adventure, Card Games with Dice Rolling & Variable Player Powers
- Cluster 11: Simulation Wargames

The clusters dominated by the means of playing a game (clusters 0 and 4), dice and/or cards, are of interest. Of the 3,947 games in cluster 0 each one has the Card Game category. Including games such as Race for the Galaxy and the classic game of Go Fish. There are 5,504 total games in the data set with the Card Game category, so cluster 0 represents 72% of all Card Games. Likewise, cluster 4 contains 1,310 of the 1,617 games (81%) in the Dice category. While these are extreme examples, we can see that the clustering is not entirely predicated on labels.

Cluster 0

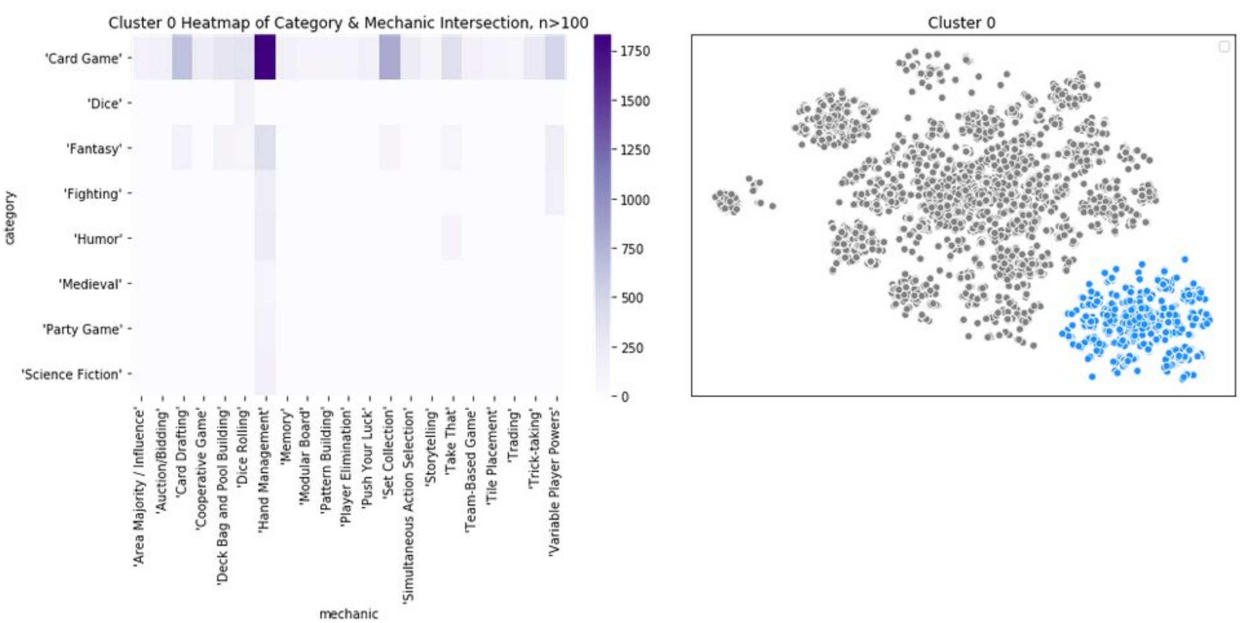

Figure 15: Category and Mechanic Heatmap for Cluster 0

Cluster 4

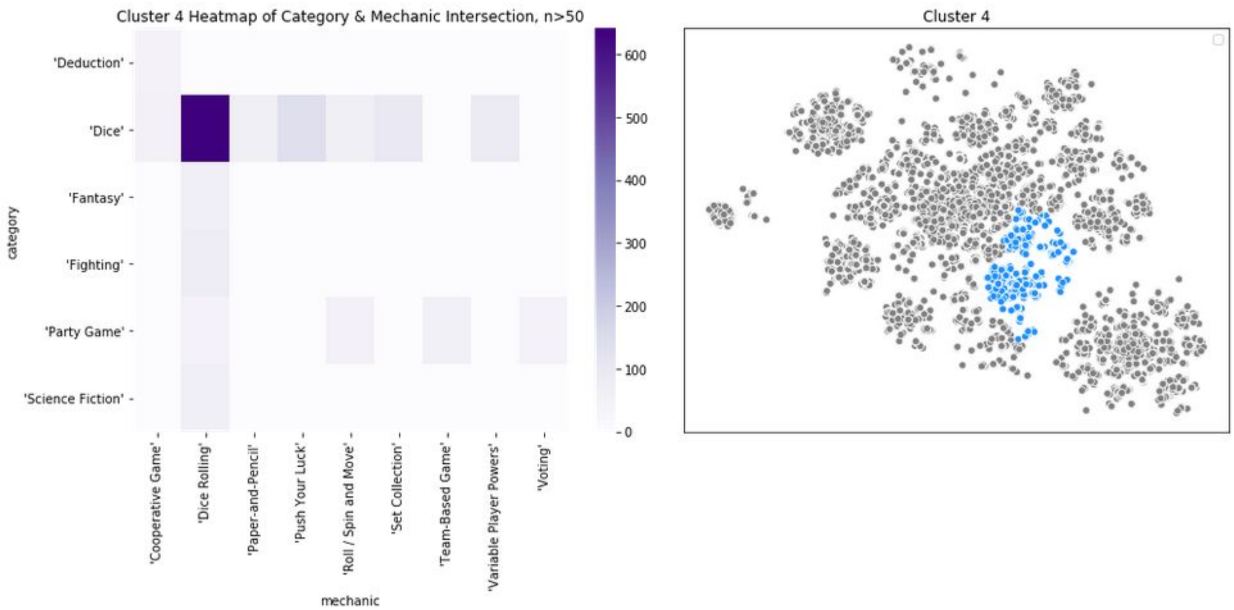

Figure 16: Category and Mechanic Heatmap for Cluster 4

The separation of clusters 5 and 11 are also of note. We saw that both these clusters have comparatively higher means for age, play times and weight. However, cluster 5 is dominated by more diverse categories such as Nautical, Napoleonic, Science Fiction, Medieval where cluster 11 is primarily World War II games and, also notably, with a high proportion of the Simulation i.e. "games that attempt to model actual events or situations." Examples of World War II/Simulation games are Combat Commander: Europe and Conflict of Heroes: Awakening the Bear! – Russia 1941-42 both involving "tactical infantry combat" in World War II scenarios.

Cluster 5

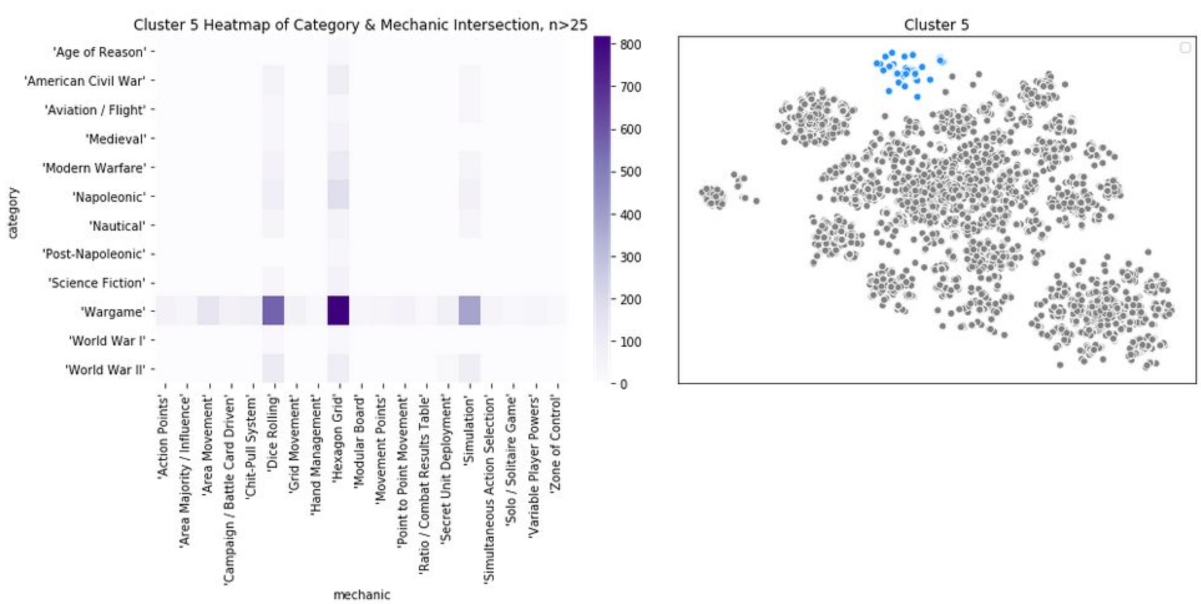

Figure 17: Category and Mechanic Heatmap for Cluster 5

Cluster 11

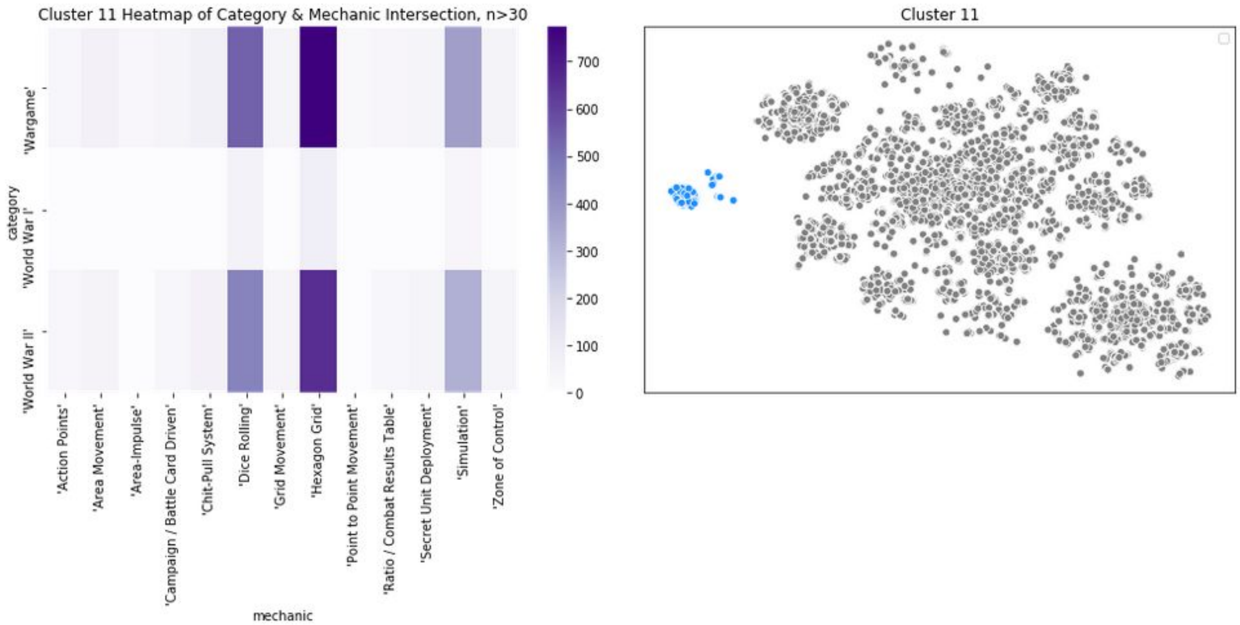

Figure 18: Category and Mechanic Heatmap for Cluster 11

Overall the dominant labels within each cluster can give you a good idea of what kind of game you might find in a given cluster. Except perhaps in cluster 3: Animals, Bluffing & Card Games. While it does contain 85 games (6% of the cluster) with both Bluffing and Animals labels there is a distinction with 873 Animals games (of the 1,132 total) and 551 Bluffing games (of the 1,105 total) in the cluster. Maybe some of the Animal games belong in the neighbouring cluster 8 with other Children's Games, but, on the other hand, the games you might think belong in the bluffing card game cluster, such as Poker or Blackjack, are in cluster 0. Clearly there is something in this cluster causing distinction from the others. Returning to the descriptive statistics we see that the means of cluster 3 are across the board higher than cluster 8 which is stratifying these clusters.

Cluster 3

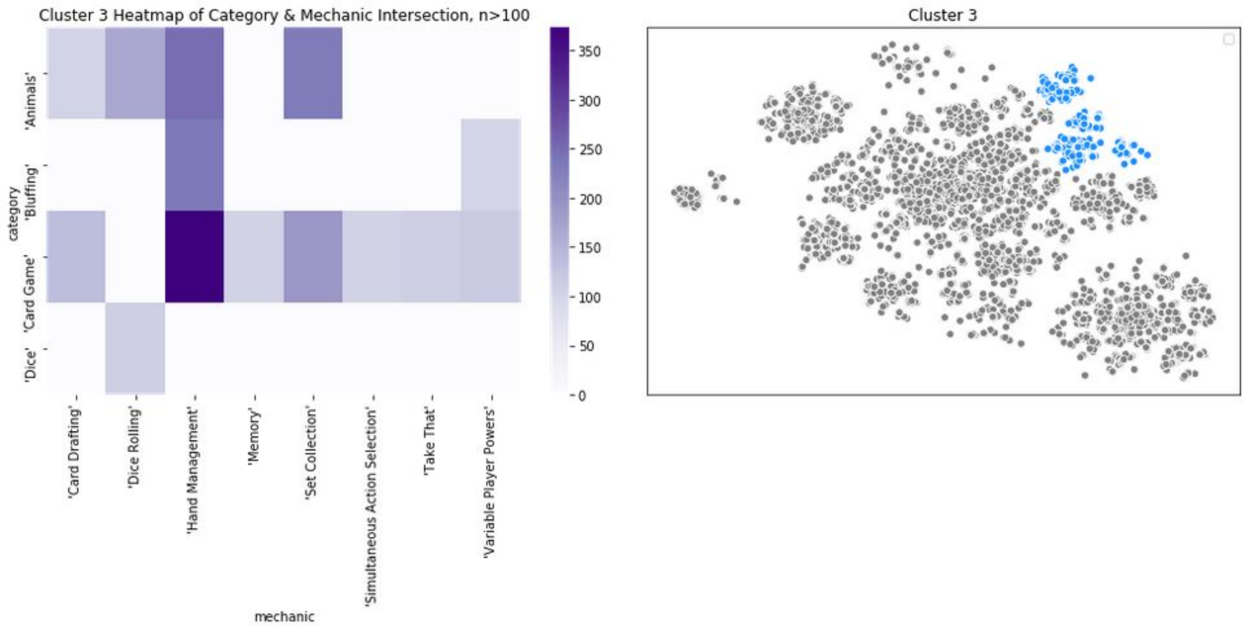

Figure 19: Category and Mechanic Heatmap for Cluster 3

Cluster 8

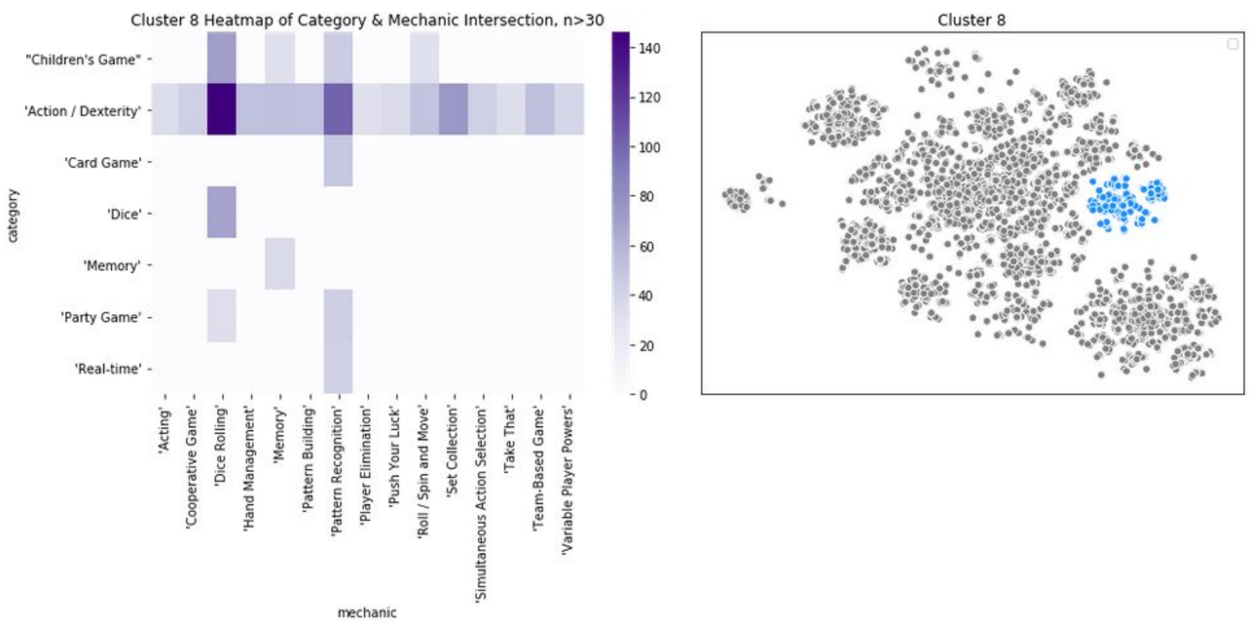

Figure 20: Category and Mechanic Heatmap for Cluster 8

Feature Engineering & Preprocessing

In feature engineering I completed two main steps. First I created dummy variables for each category and mechanic label which generated 269 features. Some games do not have labels for category or mechanic. This was recorded in the "None" columns.

As for the family label, it is sparse with 2,748 different families. Using this to generate features would increase dimensionality significantly. Many of the families are also unique or specific to a game or set of games, so they will not generalize well to new data. However, as we saw above, games in the Kickstarter family have a statistically higher average Geek Rating, so the second step was creating a column indicating if a game was or wasn't on Kickstarter. The remaining families we're dropped from the data set.

For preprocessing, numerical features (age, weight, min & max play time, min & max players) had missing values which were imputed using kNearestNeighbors and then scaled to a range between 0 and 1.

Lastly, I created a target variable indicating if a game was in the top 1000 of all games based on ranking which represents 5.256% of the data set.

Model Selection

To start out I tested five classifier models using GridSearch: Logistic Regression, Random Forest, Gradient Boost, kNeighbors, and LightGBM. Because the target variable I chose makes the data set imbalanced, the area under the precision recall curve will be the deciding factor. The Random Forest model resulted in the best score, so I chose to move forward with that model.

Initial Modeling Results

	roc_auc	precision-recall_auc
Logistic Regression	0.884259	0.379357
Random Forest	0.921491	0.508385
Gradient Boost	0.904461	0.416739
kNeighbors	0.839182	0.250997
LGBMClassifier	0.913805	0.456363

Figure 21: Initial Modeling GridSearch Results

Classifier	Parameter Space
Logistic Regression	**penalty:** l1, l2 **C:** 0.01, 0.1, 0.5, 1.0 **class_weight:** balanced **max_iter:** 100, 500, 1000, 2500, 5000
Random Forest	**n_estimators:** 50, 100, 250, 500, 1000 **criterion:** gini, entropy **max_depth:** 5, 25, 50, 100, None
Gradient Boost	**learning_rate:** 1, 0.5, 0.25, 0.1, 0.05, 0.01 **n_estimators:** 5, 50, 100, 250, 500 **max_depth:** 3, 5, 10, 15, 20, 25
KNeightbors	**n_neighbors:** 5, 25, 50, 75, 100, 125, 150, 175, 200 **weights:** uniform, distance
LGBM	**boosting_type:** gbdt, dart, goss **num_leaves:** 10, 25, 50, 100 **max_depth:** -1, 5, 10, 15, 20, 25 **learning_rate:** 1, 0.5, 0.25, 0.1, 0.05, 0.01 **n_estimators:** 5, 50, 100, 250, 500

Figure 22: GridSearch Parameter Space

Model Tuning

To improve the model I first tried Multiple Correspondence Analysis, however this did not improve results. From here I investigated feature importance in a further attempt to reduce the number of features. For each feature, I retrained the model minus that single feature and, using cross validation, compared the new model's precision recall AUC score to the model using all features. This identified 65 features to drop and gave a slight 0.001 increase in the precision recall AUC score.

Next, since the data set is imbalanced, I tested oversampling the minority class. Using cross validation again I tested oversampling at 10% increments, but no value provided an improvement.

Lastly, to generate the final model, I implemented Bayesian hyperparameter tuning over 100 trials to find a more optimal combination of model parameters. By tuning the model with the reduced feature space I was able to modestly increase the precision-recall AUC by 0.0149 over the base model.

Model Tuning Results

	roc_auc	precision-recall_auc
Random Forest	0.921491	0.508385
Random Forest MCA	0.882355	0.400760
Random Forest Reduced Features	0.921136	0.509506
Random Forest Reduced Features Tuned	0.920219	0.523304

Figure 23: Random Forest Tuning Results

Bayesian Hyperparameter Tuning Parameter Space	**N_ estimators:** range 50 - 2000, increment 50 **criterion:** gini, entropy **Max_features:** sqrt,'log2, None **min_samples_split:** range 2 - 20, increment 1 **min_samples_leaf:** range 1 - 20, increment 1 **bootstrap:** True, False
Best Parameters	**N_estimators:** 1050 **criterion:** entropy **max_features:** log2 **min_samples_split:** 19 **min_samples_leaf:** 1 **bootstrap:** True

Figure 24: Bayesian Hyperparameter Tuning Parameter Space and Best Parameters

Thresholds

Classifier models return the probability that an observation is of a particular class. The assumption is that if the probability is greater than 0.5 than its a Top 1000 game. However, this is not set in stone and is a decision point for us to make a choice about what's more important, precision or recall. Here I've plotted the precision and recall scores as a function of this threshold to find the appropriate thresholds for our uses.

Developing a board game is resource intensive in both time and cost, so precision is the primary metric to determine the threshold we want to use. There are two scenarios regarding the decision threshold to be considered depending on the situation of the game publisher.

Scenario A: This publisher is new. Perhaps they are a startup or a hobbyist trying to break into the market with a Kickstarter campaign. For this publisher a failure would be fatal, so they require high precision. At this threshold the model has trouble finding the games that fall into Top 1000, but when a top game is identified it is confident in the prediction. The threshold I recommend in this scenario is 0.385 representing a .92 precision score.

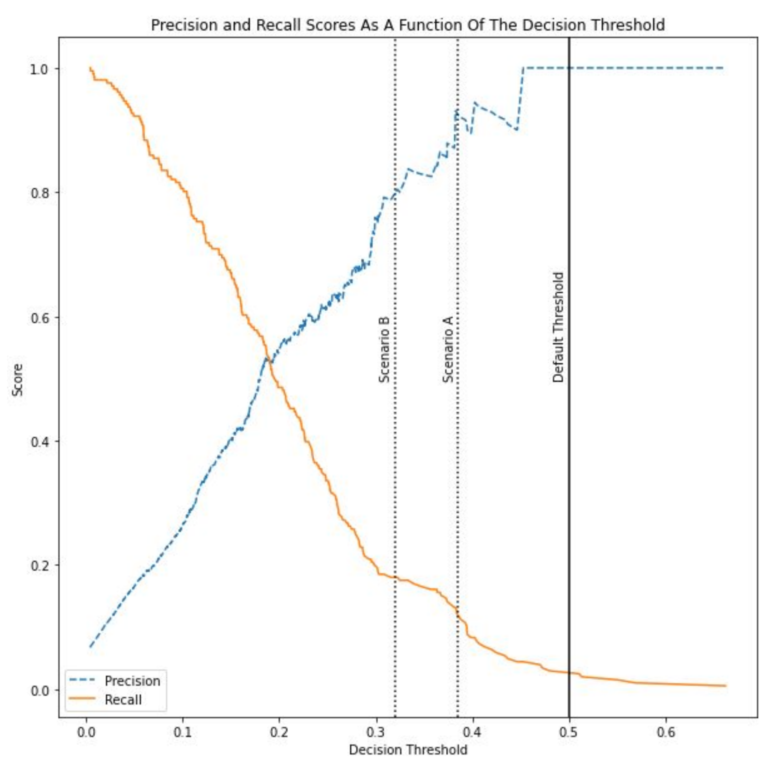

Figure 25: Precision and Recall as a Function of Decision Threshold

Scenario B: In this scenario the game publisher is an established business with a successful track history and has experienced game developers on staff. Due to their expertise in the industry they have more flexibility and can trade off some precision for a greater variety in the potential games they are looking to make. They might not get the best combination of features in their game, but with their vast domain knowledge they will still be able to make a successful game. The threshold for this scenario is 0.32 representing a 0.8 precision score.

Below are the confusion matrices for both scenarios.

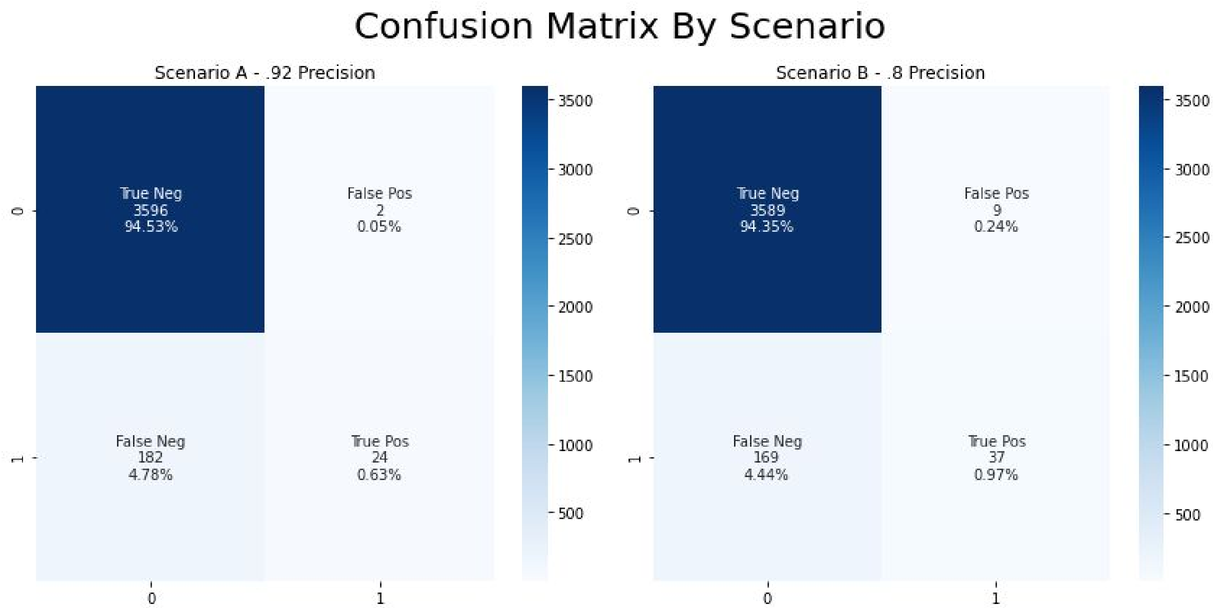

Figure 26: Confusion Matrix by Scenario

Further Analysis Of LGBM Model

The below chart shows the top to features by importance.

All numerical features (except age, which was previously dropped) are highly represented here. From EDA we saw a positive correlation (0.29) in Geek Rating and weight. There was a minimal correlation from min and max playtime (0.0032 and 0.007 respectively), and slightly negative correlations from min and max players (-0.069 and -0.028 respectively).

The categories are primarily those with many games. Economic, Wargame, Fighting, Card Game, and Fantasy are all within the top 10 categories by number of games having over 1,300 games each. The only exception is City Building, however this category (and Economic) is in the top ten by Geek Rating and has 473 games with this category.

We see a similar story with the mechanics. Hand Management, Dice Rolling, Set Collection, Variable Player Powers, Card Drafting, Area Majority Influence, Modular Board, and Grid Movement are in the top 20 mechanics by count with over 1,000 games each. The exception being Solo/Solitaire Game with only 420 games. However, it is the second best mechanic by Geek Rating.

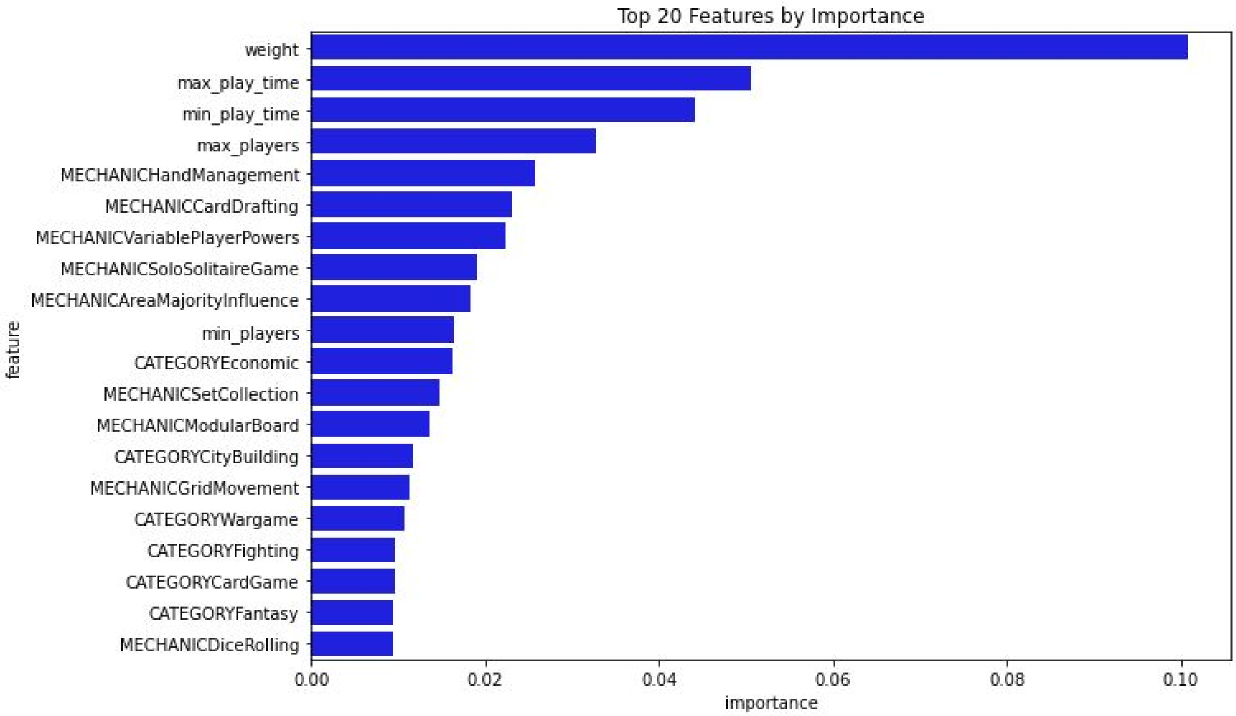

Figure 27: Features by Importance

Lastly, we should examine select false positives and false negatives in order to better understand how a game is more than just the sum of its features.

False Positives

First let's examine the false positives. (Note that I'm using the threshold from scenario B to examine misclassifications since the default threshold had perfect precision).

The false positive Uchronia (Rank 2628) is a great example. It has the right combination of mechanics, age and weight and by all accounts should be a top game. However, digging into the reviews, we can see where Uchronia fell short. User Papa Ninja summarizes these shortfalls in his review giving the game a 4/10:

"I have been wanting to play this one for a while since I really like Glory to Rome*. Personally I found this to be a far cry from GtR; in my opinion all of the tension and interesting card synergies are missing. The locations are all very lackluster and are just worth 1, 2 or 3 points with very generic powers. The worst part of the game is how poorly the theme is tied into the game. Dinosaurs, awesome... except none of the theme comes through with any of the cards. This game was a major disappointment and not something I would seek out to play again."

**Uchronia's gameplay is a based on the game mechanics of Glory to Rome (Rank 184)*

Another great example is Renegade (rank 1190). While still a well rated game, we can see what is holding it back from the top 1000 by looking at the negative reviews.

User bmarting gave Renegade a 4/10 saying "All the rules grit of Mage Knight (rank 24) with none of the payoff. It has really good ideas but they're brought down by some overly convoluted systems and poor conveyance."

We also have user decoseer who wrote a 658 word review giving the game a 4/10 and concludes by saying "I'm glad the designer has found success in this product, and that people seem to genuinely like it. But I'm at a loss to understand why. I really tried to enjoy this game, but the more I prodded at the mechanics, the more unbalanced they seemed. For such a technical theme, its strange that this game is so technically flawed."

Renegade has a high weight at 3.650, but what we can see from these reviews is that the publisher leaned too far into complexity.

False Negatives

To wrap up, we can move on to the other end of the spectrum by taking a deeper look at the games that the model gave the lowest probability of being a top 1000 game, but actually are.

What we can see here are games that are exceptions where game design trumps expectations. They generally have low weight, and have unpopular categories or mechanics such as action/dexterity, children's game, party game, roll spin and move, or memory.

For example, KLASK (rank 289), which can be described as a small scale version of foosball or air hockey, by all accounts should be a terrible game; action/dexterity is not a highly rated category, and both age and weight don't align with the trends we saw. But, this is an exception. User cuazzel summarizes this well with their 9/10 rating: "I HATE dexterity games. This should say enough about my rating". User neflight86 has a similar sentiment in their 8.5/10 review specifically mentioning the unexpected depth of KLASK. "Great foosball like fun with much more strategic variance than I typically expect in a dexterity game."

Conclusion

We learned that a game's weight is the largest factor in modeling and has the highest correlation (0.28) with Geek Rating. Additionally, mechanics that create player agency and strategic choices are well received. Our final Random Forest model ended up with a 0.523 precision recall AUC after feature selection and hyperparameter tuning which was improved over the base model's 0.508 AUC score.

Careful consideration while grouping or eliminating certain members of the category and mechanic features could show improvement.

As outlined in the Model Selection - Thresholds section, this model can be used by both small and large game publishers to predict the success of a game during the initial stages of development. This could be used during pre production to choose a game to produce from a set of ideas. Or, it could be used to evaluate the effect of changing a game’s features, for example, adding a new mechanic. Additionally, this model can be used by consumer facing retail establishments, where shelf space is limited, to determine if a game should be kept in stock.

However, we should remember that what makes a board game "fun" is a highly subjective matter. As we saw, Monopoly is one of the worst rated games, but also a household name and by and large a commercial success. We also see categories and mechanics representing games at both ends of the ranks. Other aspects such as artwork, story, and even the quality of the game components can be just as important in a game's success.

All in all, games are more than just the sum of their features. The model can provide a solid basis of where to start when building a game, but simply choosing the “best” mechanics and the “right” number of players, length of playtime and weight etc. won’t make a game great. Without careful implementation of these features the best concepts can still fall flat. Fortunately, the model can help us in this area. We can turn to the false positives for case studies on what to avoid or common mistakes, and we can look at the false negatives for inspiration on creating exceptional games.

Further Research

There are a few areas that have been top of mind for me throughout this project. To start, as mentioned at the top, BoardGameGeek is a site for enthusiasts. Considering this, another area of investigation is: how do the ratings from this enthusiast population compare to others? For example, Monopoly is one of the worst ranked games at 19,012 out of the 19,019 games and has a 4.299 out of 10 Geek Rating. However, on Amazon.com, Monopoly has a 4.6 out of 5 stars rating and over 12,000 reviews on Amazon. By using this data while developing a board game it is imperative to remember the audience that generated the data and that it might not translate to the general market.

Another consideration is to look at the effects of a game’s price. In the analysis I noted the higher costs of some of the top rated games, but I do not have the pricing data in the data set. Does price influence ratings via a type of sunk cost fallacy or self selection bias?

Lastly, crowdfunding via Kickstarter or other sites is a newer phenomenon. We saw it was not a fluke that Kickstarter games had a higher than average Geek Rating. But why? Deeper analysis between games that are and are not crowdfunded could provide valuable insights.