The Impact of Crowdfunding on Board Games

Analyzing the Relationship Between BoardGameGeek & Kickstarter

Background: Since 2010, board gaming has experienced unprecedented growth as a hobby market, with 86 of the top 100 games on BoardGameGeek – regarded as the definitive rating forum – released in this period. With access to new publishing channels, more than a quarter of the top 100 have been crowdfunded through Kickstarter, but few formal analyses exist investigating the relationship between these powerhouses of the market. If current trends continue, Kickstarter funded games may soon comprise more than half of the top 10% of rated games on BoardGameGeek, suggesting there are fundamental relationships yet to be explained between them.

Intent: No publicly available datasets combine BoardGameGeek game data with associated Kickstarter campaign data. By merging these into a novel dataset for all ranked games on BoardGameGeek, we sought to investigate both directions of this relationship – how game features might impact crowdfunding success and how that success might impact user response. By illuminating these connections, we could then explore their complex intersections. This exploration can potentially form a knowledge base for the development of analysis tackling acute questions about the current and future state of board gaming.

Motivation: Despite being a small fraction of the overall gaming market, board games have an outsized mindshare in the crowdfunding space – two of the top ten Kickstarter campaigns have been tabletop games, with a new record set by Frosthaven in 2020. The lack of joint data relative to the wealth of individual data points from Kickstarter and BoardGameGeek reflected a critical gap in the analysis of board gaming as a growing industry. By generating both this novel dataset and the accompanying analysis, we hope to spark increased interest in projects examining the wealth of rich data available relating to board games, a source underutilized relative to other hobbyist datasets.

Guiding Questions:

1. To what degree does the appearance of a given feature in a game correlate to crowdfunding success, as measured by Kickstarter campaign engagement and funding levels?
2. What relationships exist in the distribution between retail and crowdfunded games and how does the success of a campaign impact critical user response via Bayesian rating and user engagement?
3. Can subsets of campaigns engaging in potential platform manipulation be identified by their game and campaign features and, if so, how do they relate to the first two questions?

Prepared by
Nicholas Canu, Jonathan Ellis, & Adam Oldenkamp

Data Sources and Cleaning

Kickstarter: A global crowdfunding platform, Kickstarter (KS) helps creatives launch campaigns for their ideas in order to reach a wider audience and gain funding through pledges. Kickstarter does not provide public datasets about specific campaigns, but the site Web Robots provides scraped data. This includes campaign-level information such as the number of backers, amount pledged, whether it was funded, and more. After cleaning, this data set consisted of ~40,000 campaigns within the 'tabletop gaming' genre and was pulled on April 9th, 2022.

Data Scope: ~40,000 campaigns in a 240mb CSV available at: https://github.com/canunj/BGG_KS_Analysis/blob/main/KICK_Data.csv

File Origin: Data was scraped using a script on April 9th, 2022, that downloaded zip files stored at https://webrobots.io/kickstarter-datasets/. Web Robots itself scrapes Kickstarter every two weeks to build these files.

Data Cleaning: Data was spread across 80 zip files, each with 60 .csv files that needed to be combined into a single dataset. We loaded these files in a loop, performed cleaning operations, and then appended them to a final dataframe for export. Only a small percentage of the total campaigns were board game-related, so we used keyword filtering in the campaign categories to find a reasonable subset that reflected just board games. In our exploration of this dataset we learned that many campaigns would start just after a WebRobots scrape, gain funding, but then fail and end before the next scrape. These campaigns were unfortunately not included in our Kickstarter campaign dataset.

BoardGameGeek: The premier board gaming forum, BoardGameGeek (BGG) hosts a nearly comprehensive database of board games along with a robust user base that reviews, categorizes, and discusses releases. BGG does not make a public dataset available, but its onsite data can be scraped. This subset of data, post-cleaning, comprises ~18,000 ranked game item entries with associated features – such as themes, components and mechanics – as of April 4, 2022. Its core feature is the 'Bayesian Rating,' which underpins the BGG ranking system. To be ranked, a game must have 30 votes (on a 10-point scale), at which point BGG will impute dummy votes to weight against all other ranked items. We have limited our data to this subset of BGG-ranked games.

Data Scope: 18,512 game items in a 70mb CSV available at: https://github.com/canunj/BGG_KS_Analysis/blob/main/ranked_data.csv

File Origin: Scraped from BoardGameGeek.com in April 2022 by @mshepherd on GitLab for Recommend.Games. Data obtained via the BGG XML API and is available under CC BY-NC-SA 2.0.

Data Cleaning: Key game item features (game type, mechanics/categories, and release data) are encoded by BGG as lists of 'text':'id' dictionary pairs, but they scrape to a single string representing the same data. We extracted those text values in all list columns using a RegEx pattern, then reconstructed the data into a new list as individual items. We also encoded 'game_type', 'mechanics', and 'category' columns and the 'Crowdfunding: Kickstarter' family value using get_dummies as binary labels for each unique value in these features. Finally, we pruned extraneous columns such as min and max review score.

Data Manipulation - Identification and Key Creation

Data Identification:

In our efforts to join games from the BGG data with their corresponding Kickstarter campaigns, our first challenge came from the lack of shared identifiers between the BGG and Kickstarter data. To merge the data, all BGG game items with a Kickstarter tag needed a Kickstarter campaign id value associated with them. Our solution implemented fuzzy text matching on key features through the FuzzyWuzzy Python library paired with manual data review.

The implemented game item parser utilized a two pass system on Kickstarter campaign items, limited to the 'Tabletop' category. For each BGG game item:

- A primary parser performed a token sort ratio match between the 'name' field and all Kickstarter 'name' features.
- A secondary parser iterated through the following sequence attempting to narrow all possible results to a single item as a backup match. If at any point a single match was found, it was stored as an alternate result. If none were found, the string None was stored.
    - Search for the game item name as a RegEx pattern.
    - Perform a looser token set ratio match between names storing results above a 90 match ratio.
    - Perform a token sort ratio match against all alternate names for the game item.
    - Create a list of all designer and publisher names as a RegEx pattern, searching for any that appear in the 'blurb', 'creator', 'slug', or 'name' features. Then compare these results to the alternate name matches above.
- If the primary parser scored above 90, the 'ks_name', 'ks_id', 'name', and 'bgg_id' features are stored in a list of tuples. If the primary parser fails, the secondary parser stores its value.
- The overall function returns both a BGG dataframe with the associated ks_ids concatenated as a new column and the tuple id list.

Key Creation & Matching:

We achieved an average 80% first pass identification rate, then manually reviewed each item. For consistency, we matched each BGG game item that was initially funded via Kickstarter to its original campaign data. This meant several categories of games regularly tagged with crowdfunding on BGG were excluded; specifically deluxe/revised editions without their own game item and games reprinted through Kickstarter.

Using custom functions to track unmatched ids and search both the pre-cleaned datasets as well as Kickstarter, Kicktraq.com, and BGG, we successfully:

- Matched the remaining ranked items and disassociated false positives.
- Identified the 2.5% of campaigns with data missing in the Kickstarter scraped dataset.
- Disabled game items that did not follow BGG game conventions in the data, including roleplaying games and non-standalone expansions.

With this complete, we were able to merge the datasets on the 'ks_id' feature and complete the remaining manipulation steps on joined data.

Data Manipulation - Data Integrity & Feature Engineering

Data Integrity and Scope:

With merged data associated, we then pared back to our final analysis dataset by addressing potential problems in the data. This included:

- Removing game items with confirmed campaigns not captured in the WebRobots data source.
- Limiting the data to games published on or before December 31st, 2021.
- Segmenting a separate dataset where multiple items were associated with one 'ks_id' value (e.g., Kickstarter campaigns that funded more than one board game) and testing impact on aggregate measures and feature correlation.

Kickstarter Feature Engineering:

Imported features in the KS dataset are informative but limited. Though total funding pledged is clearly an indicator of a KS campaign's success, it can be skewed by campaigns that draw a large amount of money from a small number of backers, thus not achieving widespread appeal. To define a functional success variable we explored creating several new ones, including average pledge amount, percentage that each backer's pledge contributed to the total funding achieved, final funding over goal multiplier, and a year funded to publishing time delta.

The success feature 'ks_rank' was created through the weighted joint rank of descending total funding pledged and ascending per backer contribution. This weights the ranking toward campaigns that achieved the greatest funding while engaging the most backers, which aligns with our intent to explore cross-metric impact of crowdfunding as a developing release channel for board games. KS rank is the primary success metric for much of our analysis.

Continuing Challenges:

Despite these robust datasets, informative features remain inaccessible on the Kickstarter platform and limit the scope of analysis. Lacking data on pledge tiers offered, it is not possible to properly weight backing averages. Also, critical creator data – including frequency of post-campaign updates and previous funding history – are not encoded, although the latter could be approximately aggregated from extant data.

Additional analysis on BGG-unranked game items also fell outside the scope of this project but may contain valuable information on the relationship between Kickstarter and BGG. In a 500-item sample of BGG-unranked data, game items with confirmed successful Kickstarter campaigns ranged from projects with less than ten KS backers to projects funding upwards of 50,000 USD that have low engagement metrics on BGG. At the same time, more than half of BGG-unranked game items with a crowdfunding tag had unsuccessful campaigns, due to failure or cancellation, and were then published via retail channels or not published but remain the BGG database. Due to these challenges, this report focuses on BGG-ranked items, but a comprehensive analysis following extensive data cleaning would be warranted for future research.

Kickstarter's Influence on the Board Game Industry

Since its inception in 2009, Kickstarter has had a major effect on the board game industry, as our dataset shows. Our dataset includes all BGG-ranked games, a reasonable approximation for games that have found enough commercial success for 30 people to have rated them on BGG.

The number of BGG-ranked games published per year has increased most years this century. A drop-off since 2020 may be pandemic-related or reflect the lag time between publication and hitting the 30-rating threshold.

Kickstarter campaigns have funded a significant portion of recent games published (Fig. 1). Since 2014 an average of 311 KS-backed games have been published each year, compared to 656 other games. Kickstarter has been responsible for 26 percent of all BGG-ranked games published since 2009, and 16 percent of all BGG-ranked games *ever* published (BGG's data includes games as far back as Monopoly, etc.). Kickstarter's production of games has held steady since 2020 while other games have fallen off.

KS-backed and non-KS games follow roughly equal distributions in BGG's Bayesian rankings, with a mean of ~5.75 and σ of only ~0.42 on a 10-point scale. This is as designed. The highly ranked outliers, however, are of great interest, since they are some of the best-known, most successful games.

Indeed, Kickstarter has been responsible for an increasing share of the top-10% rated games year over year (Fig. 2). But the Bayesian ratings skew newer games toward the mean, so the Kickstarter effect is best seen in BGG's average rating metric (Fig. 3), which has no Bayesian adjustment.

Either way, the trend is clear: Kickstarter's share of the top 10% of games is growing, and Kickstarter now funds almost half of top-rated games.

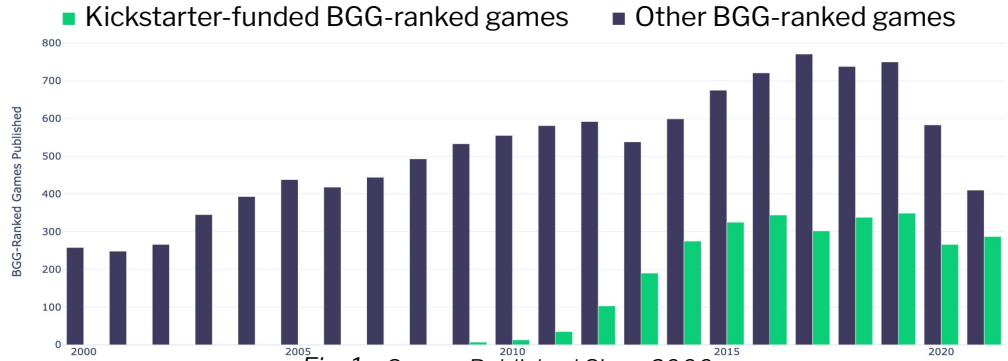

Fig. 1 - Games Published Since 2000

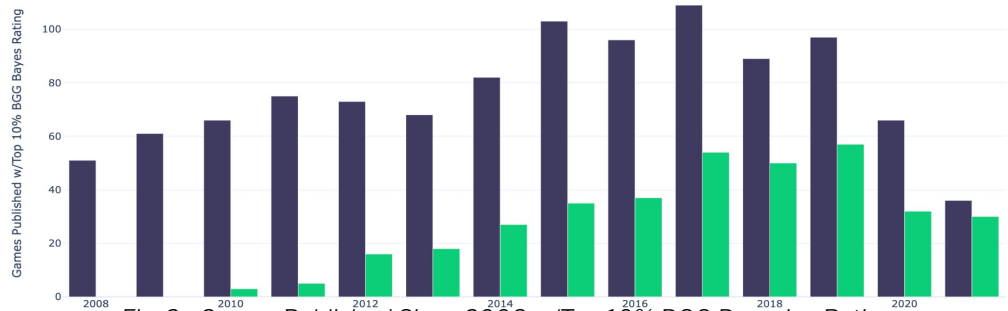

Fig. 2 - Games Published Since 2008 w/Top 10% BGG Bayesian Rating

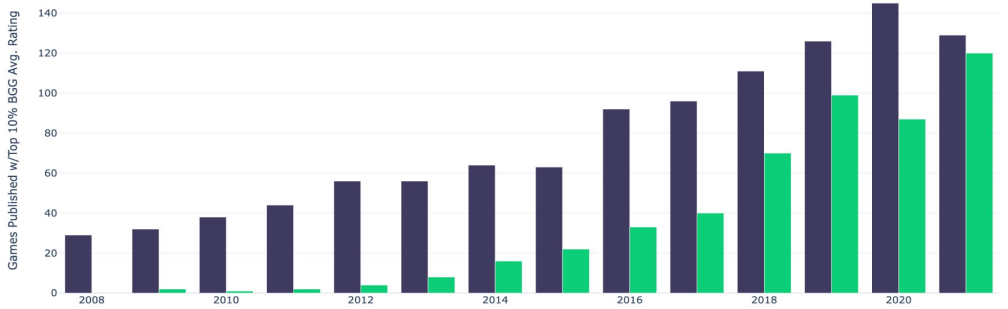

Fig. 3 - Games Published Since 2008 w/Top 10% BGG Average Rating

Do Game Features Influence Kickstarter Success? - BGG Game Type

Game Item Feature Structure:

In the BGG data, game items have a variety of descriptive feature tags in addition to release data. These include game type, mechanics, and categories. Taken together they represent both mechanical and thematic design attributes for each game. All features are non-exclusive and a game item may have multiple type, category, or family attributes present. We began with the most abstracted tag, game type, looking to identify both how broadly applied features might relate to funding success as well as areas to drill into for further analysis.

Game Type Analysis:

Looking at aggregate Kickstarter rank, we found clear stratification occurring between game types with the thematic tag (a category encompassing games with *mechanics dictated by theme*) at a mean rank of 788 consistently outperforming other types. When broken down to individual features, party-tagged games rise in both USD pledged and backer count despite a mean rank of 1112. Thematic and party have a narrow mean funding gap at 501,977 and 491,796 USD, but party takes a commanding lead in backer count at more than two times the next closest type (10712 versus 4180 for thematic). We did not observe this disparity reflected in per-backer funding contributions where thematic tagged games boast 0.1% per backer rate, meaning the presence of a thematic tag may indicate that a campaign will raise more money and do so while having higher backer engagement. This distribution of types becomes evident through quartile value spread. Heavily weighted outliers, such as Exploding Kittens, inflate party games across individual Kickstarter features. Despite the second highest release density, thematic games have the narrowest quartile spread across types with a high Q1 floor, indicating consistent Kickstarter performance versus the volatility of party games.

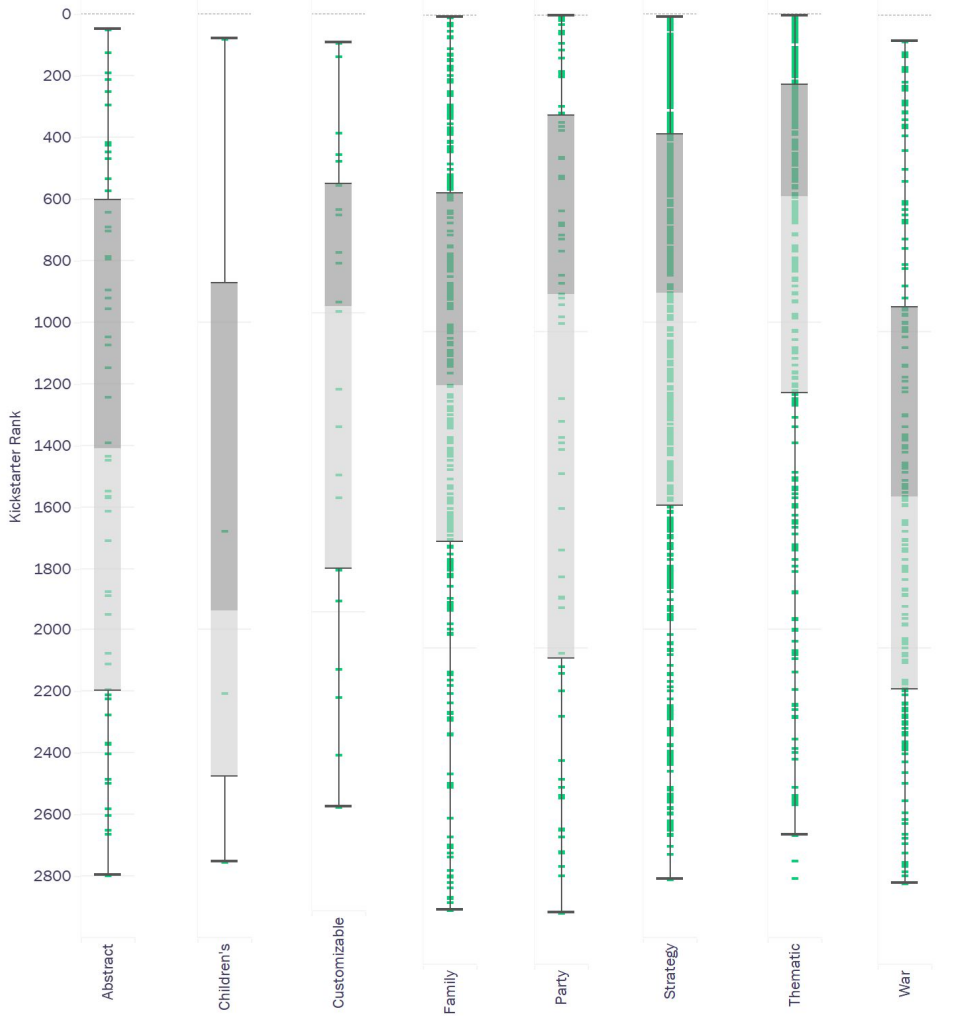

Fig. 4 - Kickstarter Rank Quartile Spread by Game Type

Do Game Features Influence Kickstarter Success? - Mechanics, Components, & Theme

Identifying Content by Category:

We have generally identified thematic-tagged games as the most popular type on Kickstarter. Can we more specifically identify common categorical features, including game theme and component content tags, linked to changes in crowdfunding engagement?

To explore this question, we examined performance across Kickstarter rank, USD pledged, backer count, and per-backer contribution for category tags appearing in 200 or more Kickstarter campaigns. Of note, the most common component tag, cards, appeared in 40.5% of campaigns while the most common theme tag, fantasy, appeared in 25%. These two saturate crowdfunding campaigns with the next closest category, fighting, only appearing in a relatively meager 13%. But for their popularity, do their associated games outperform the competition?

Overall Performance by Kickstarter Rank:

Of the three mechanic categories in this subset, only cards correlate with poorer performance across all categories. Specifically Kickstarter rank sees a weak negative correlation of -.15 while dice show no correlation at all. However, miniatures improve performance with a positive rank correlation of .26 suggesting we can expect their presence in a campaign to align with greater net engagement. For most common themes we observe weak positive correlations between .9 and .16. We may expect the inclusion of these themes to slightly improve net engagement if they have any impact at all. Adventure games show a negligible lead of .16 with fantasy at .14, both indicating that regularity of appearance may not be contributing strongly to the poor performance observed for cards. Only the humor theme, with a weak negative correlation of -.05, generally aligns with decreases in overall rank.

Diverging Performance Across Funding Features:

Across the considered funding features, miniatures remain a consistent performer. Not only do we observe moderate positive correlation to total USD funding of .32, it sits as the fourth highest funded category tag with a mean of ~570,000 USD, behind only categories such as comic strip and mature that are boosted by powerful outliers. However, when compared to backer count at a correlation of only .19, we find an interesting pattern emerging. Games with miniatures do have one of the lowest per-backer contribution rates of 0.11%, but sit in the lower portion of the upper quartile for backers at a mean of 3,770. We might consider that while the average backer contributes a small portion of high funding, a game with miniatures, which are expensive to produce, will generate more net funding per backer compared to other categories. We see this reflected in the highest average pledge of any category at 135 USD per backer.

In contrast, card games, our most common category, lag across these funding features but do so unevenly. With the worst performance of this subset in USD pledged, we see a negative correlation of -.22 but the change in backers correlates weakly at -.06. What does this tell us about card games as a category? Looking at the average and raw numbers reveals how this pattern is emerging. With a mean backer count of 1,981, it sits near the overall average of 2,047. But the average pledge is the 11th lowest of all categories at only 47 per backer. From these contrasting categories, we begin to see trends emerging across crowdfunding, where the most common categories are likely also the cheapest to fund and produce, generating less backer engagement but dominating the market space, where a more expensive product may find backers willing to contribute higher funding and push upwards across all features.

Can Niche Game Attributes Manipulate Kickstarter?

Identifying and Analyzing Potential Platform Manipulation:

A popular conception claims that some crowdfunding campaigns maximize up-front funding through targeting specific niches and deluxe editions that balloon costs and backer contributions. We explored this notion by seeing if we could identify suspicious hotspots between BGG game categories in and amongst the highest-funded Kickstarter campaigns. We built a category adjacency matrix from the top 10% of Kickstarter campaigns by pledge amount. Each row and column in the matrix counts the frequency of the given genre pair. For example, 5 games have both the 'Fantasy' and 'Animals' category tag, while no games had both the 'Exploration' and 'Animals' tag.

What we discovered was a clear group of heavily featured BGG categories; 'Fantasy', 'Fighting', 'Miniatures', 'Adventure', and 'Exploration'. The relative frequencies in this hotspot, compared to the out-of-group frequencies, suggested that these BGG categories were utilized in the preponderance of successful Kickstarter campaigns. This confirmed our suspicion that particular categories, such as 'Miniatures,' are indeed a potential lever for funding campaigns seeking to give themselves the best chance of hyper-successful funding.

While including these categories does seem to improve your chances of garnering a significant pledge amount, we were unable to identify any malicious implementation of this knowledge by board game publishers. As we will see in the subsequent slide, there seems to be a relationship between highly ranked Kickstarter campaigns and the BGG Bayesian rating. Perhaps KS backers are actually getting what they're paying for.

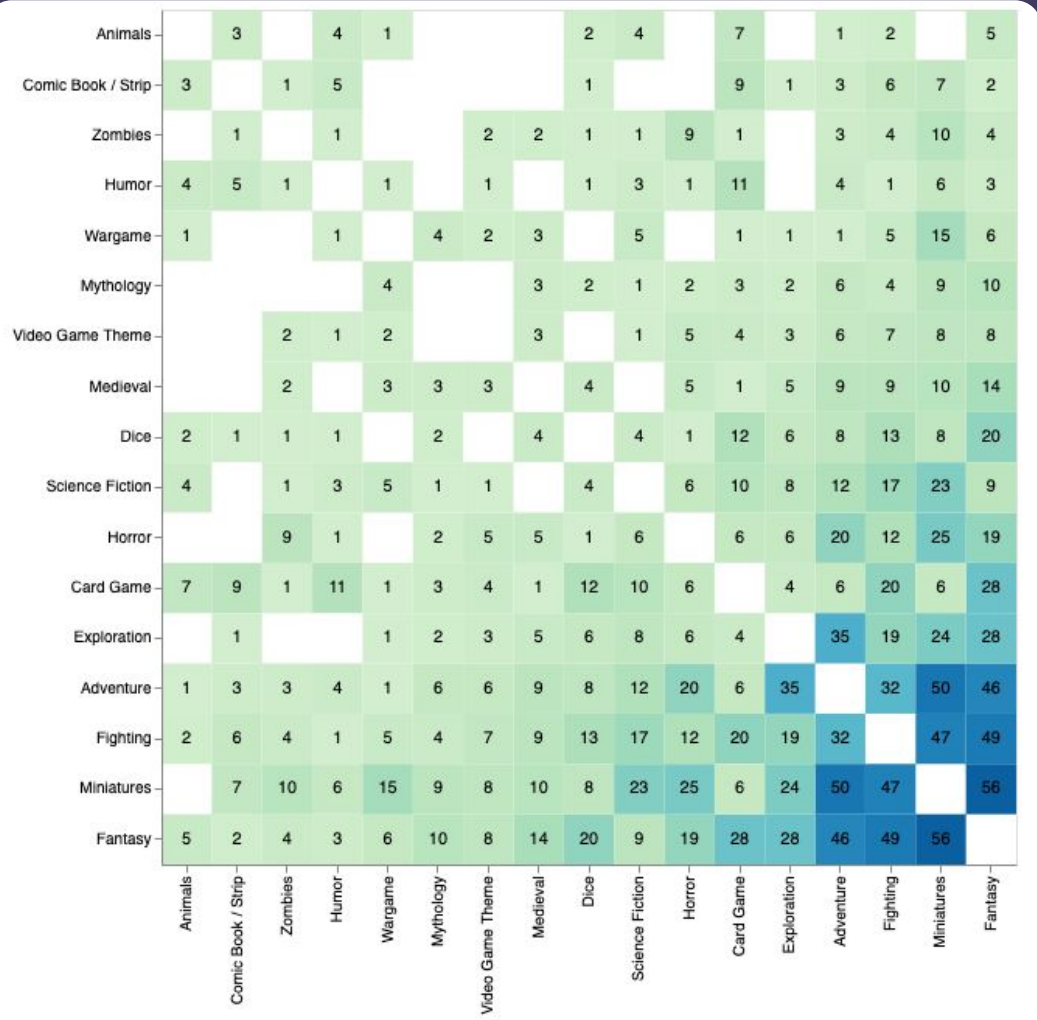

Fig. 5 - Frequency of Board Game Genre Pairs Amongst Top 10% of Kickstarter Campaigns.

Does Success on Kickstarter Correlate With Success on BoardGameGeek?

Crowdfunding Success in Critical User Response:

Comparing individual Kickstarter features to Bayesian rating, the user response variable on BGG dictating site ranking, we find little to no correlation; funding features fall into approximately normal distributions. However, the aggregate KS rank feature, weighing both funding and backer engagement, shows a monotonic non-linear positive relationship with a correlation of 0.56. In the chart below, we find lower ranks cluster tightly along a minimum net Bayesian ratings, but as rank increases beyond the median, ratings begin to steadily increase. This leads us to believe that individual Kickstarter features may not strongly impact critical response but taken collectively we would expect successful campaigns to increasingly produce games that achieve a higher Bayesian rating and BGG rank.

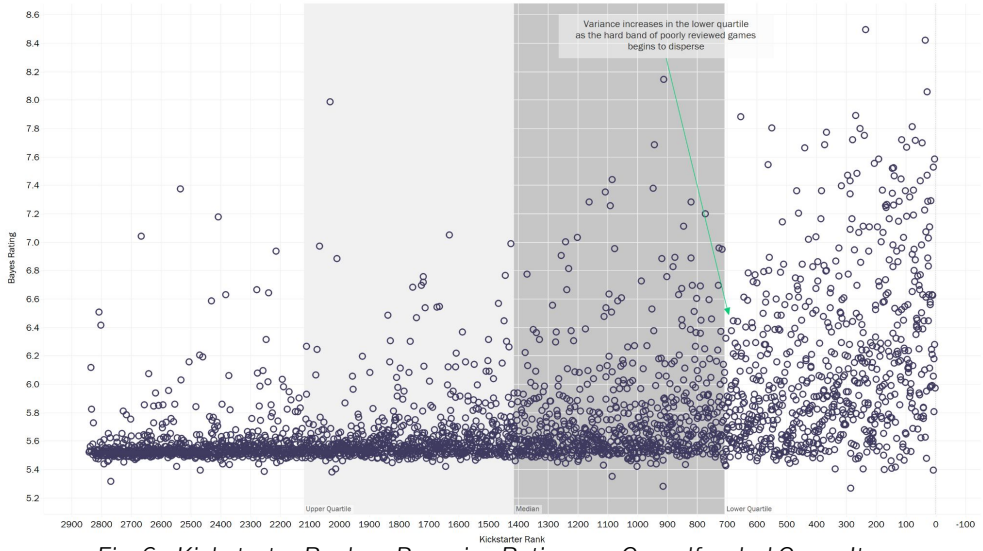

Fig. 6 - Kickstarter Rank vs Bayesian Rating per Crowdfunded Game Item

When Users Disagree:

We have examined a number of game features and their correlation with Kickstarter success. But how do they correlate with BGG ratings? Usually, a feature that positively correlates with KS rank also positively correlates with BGG rating, and vice versa. The set of features whose correlations with KS and BGG diverge – one positive, one negative – suggests the kinds of games that may be KS hits but poorly received on BGG, or vice versa. This set is shown here. For example, KS has a positive correlation with some "pop culture" game themes that may be frowned upon by the BGG community, such as movies and trivia. But KS has a negative correlation with various "historical" game themes that have fans on BGG, such as war games and certain historical periods. BGG has a positive correlation with strategic mechanics such as hand management, bluffing, deduction and auctions that are not as popular on KS. These disparities align with a common view/criticism of BGG: that its rankings reflect preferences of board game hobbyists and not necessarily the public at large.

Fig. 7 - Correlation with Kickstarter Rank vs. Bayesian Rating per BGG Game Feature

Report Summary & Ideas for Expansion

Summary:

Crowdfunded releases represent a growing proportion of published board games year over year, especially among the top-10% rated titles on BoardGameGeek. While exploring this evolving dynamic we found that categorical features, foremost among them miniatures, were indicative of increased funding and engagement levels on Kickstarter. We discovered certain feature combinations may be leveraged to maximizing funding potential. And finally, in analyzing how success on Kickstarter correlates to positive critical user response, we saw that for the top quarter of Kickstarter releases, while users on each site may not always agree, we expect to see positive trends emerging in their BoardGameGeek ratings. Beyond this initial analysis, we propose several avenues for future exploration of the complex intersection between these sites.

Ideas for Expansion:

1. **Predictive Feature Analysis:** While we found several strong potential relationships in this data, the next step is further analysis expanding on the predictive nature of features in either data direction. Applying supervised learning techniques may offer an avenue to either potentially predict the financial success of a campaign based on its game features or critical response based on a combination of game and funding features.
2. **Board Game Feature Deep Dive:** This analysis centered on BoardGameGeek features critical to crowdfunding and, by necessity, left many interesting features unanalyzed. Options for further consideration might include design/publisher performance over time, the intersection of feature tags, core game versus expansion response, or the impact of shared family qualities (a game feature that includes shared publishing lines, e.g. all Pandemic titles) on the success of follow-up titles.
3. **Subsequent Kickstarter Campaign Trends:** Many successful games have multiple Kickstarter campaigns, including games originally released through retail channels, which were not included in this analysis. This would consist of reprints, new editions that did not receive their own game item, deluxe editions, and expansion campaigns; exploring how these factors might impact user response and funding is a key next step in building a comprehensive view of crowdfunding in board gaming.
4. **User Sentiment in Review Text:** The raw BoardGameGeek dataset includes both user and review data which track all site users, their collections and activity, as well as all reviews assigned to each game item. Extracting and associating the text of these reviews would allow for a more granular natural language sentiment analysis on critical user response than is possible with aggregate feature data used in this analysis.

Statement of Work

Nicholas Canu

- Collating & cleaning BoardGameGeek data.
- Coordinating GitHub repository and project storage.
- Game item parser function creation and testing.
- Organizing game item to campaign manual review process.
- Kickstarter feature engineering and rank success feature development.
- BoardGameGeek feature impact on Kickstarter campaign success analysis and slide generation.

Jonathan Ellis

- Cleaning Kickstarter to BoardGameGeek fuzzy matching data join results.
- Creation of helper functions to assist with data join cleaning.
- Editing of project proposal and final presentation.
- Analysis of Kickstarter influence on the board game industry.
- Analysis of divergent correlation between game features and KS rank and BGG ratings.

Adam Oldenkamp

- Scraping and cleaning Kickstarter data from webrobots.io.
- Parsing Kickstarter to BoardGameGeek fuzzy matching data join results.
- Final Presentation template/graphic design/visuals.
- Identifying and Analyzing Potential Platform Manipulation analysis.

Sources:

1. Vitulskis, Tomas, and Paulius Jonaitis. “Kickstarter Datasets.” *Web Scraping Service*, 26 Apr. 2022, https://webrobots.io/kickstarter-datasets/.
2. Shepherd, Markus. “Board Game Scraper.” *Recommend.Games*, 08 Apr. 2022, https://gitlab.com/recommend.games/board-game-scraper

All Project Data & Work Files Available at GitHub