The Board Game Renaissance: What Makes A Great Board Game

By Sam Beardsley

CONTEXT

As of 2017, the worldwide tabletop gaming industry had a market size of **$7.2B**

which is expected to increase **40%** by 2023

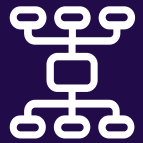
and, by 2025, the US market alone is predicted be **$5B**

Can we predict a game's popularity?

WHAT AFFECTS GAME POPULARITY?

Weight

- Weight represents the complexity of a game
- Games with a higher weight tend to rank higher
- Weight is the largest contributing factor to our model



Mechanics

- Mechanics define an element or type of gameplay
- Allowing for player choice and agency is important for success

Domain Expertise

- Fun is subjective
- A game can fail even with the right attributes
- The model's misclassifications can help guide us

WHO CARES?

Publishers

Established Developers

Start Ups & Crowd Funders

Consumer Endpoints

Big Box Stores

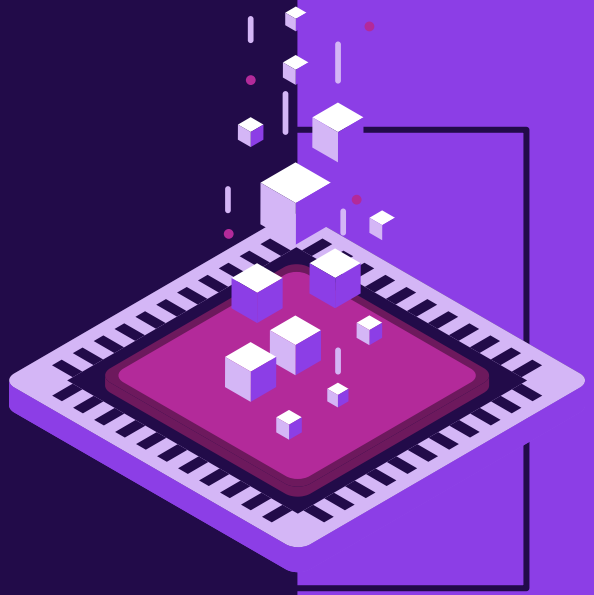
Local Establishments

- Mom & Pop Stores
- Board Game Cafes

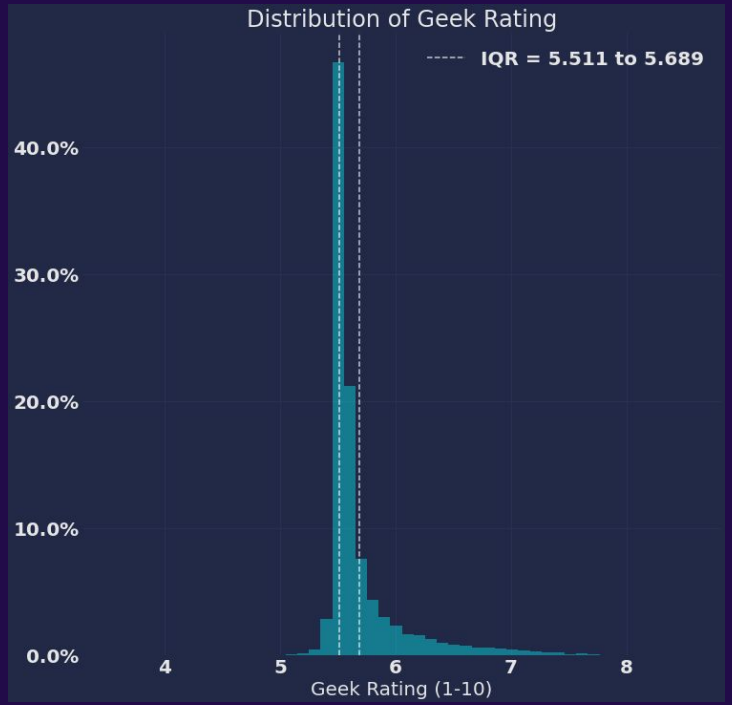
THE DATA

19,019 games scraped from BoardGameGeek.com including...

- Numerical features:
  - Age, Weight, Min/Max Players, Min/Max Playtime
- Categorical features:
  - category: a classification for the thematic type of board game
  - mechanic: an element or type of gameplay
  - family: an attempt to group a game into a broader set of descriptors

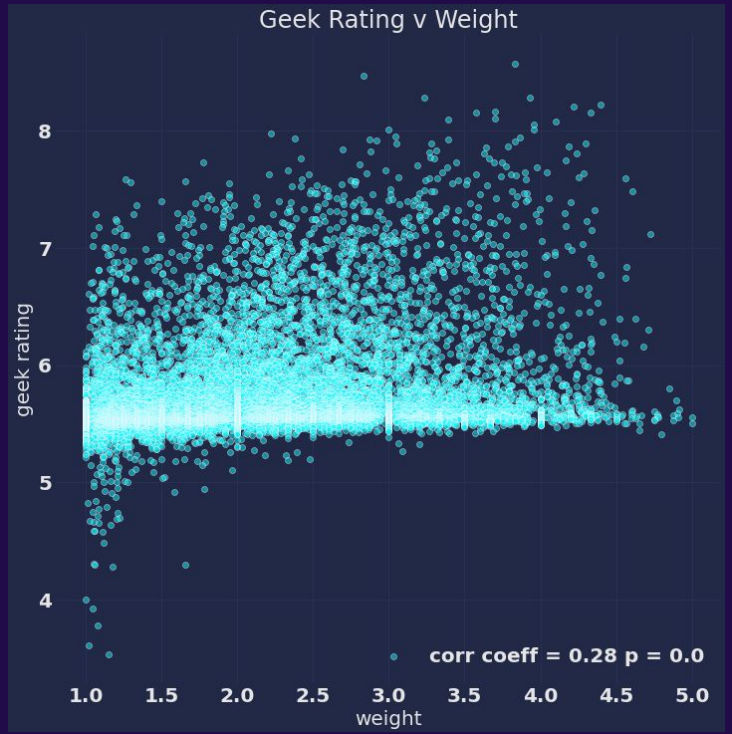


GEEK RATING



- Games are ranked by users on a 10 point scale
- BGG applies Bayesian averaging to the raw ranking to prevent small numbers of users from skewing the ranking
- Geek Rating determines a game's absolute rank
- A game must have 30 user rankings before a Geek Rating is given
- There are only 14 games with a Geek Rating > 8 (0.0074%), and 374 with a Geek Rating > 7 (1.966%)

WEIGHT



- Weight is a user generated number between 1 and 5 representing the complexity of a game
- The mean of weight is 2.04
- 75% of all games have a weight below 2.56
- There is a positive correlation between weight and Geek Rating with a 0.28 correlation coefficient and a small p-value

5 WORST RATED MECHANICS

	Count	Average Geek Rating
Rock-Paper-Scissors	151	5.615
Pattern Recognition	488	5.598
Zone of Control	115	5.597
Singing	47	5.542
Roll / Spin and Move	1,250	5.523

* see appendix for descriptions of these mechanics

5 BEST RATED MECHANICS

	Count	Average Geek Rating
Turn Order: Pass Order	7	7.633
Automatic Resource Growth	10	7.309
Turn Order: Role Order	6	7.149
Action Drafting	18	7.142
Force Commitment	7	7.103

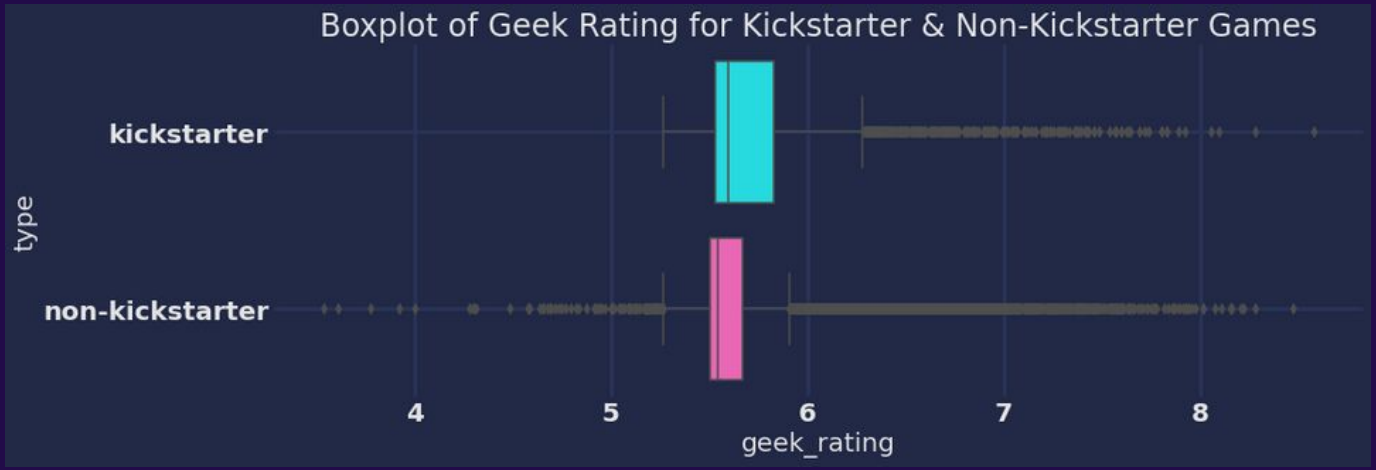
* see appendix for descriptions of these mechanics

MECHANICS: KEY TAKEAWAYS

1. Mechanics that involve chance are not well received

2. Mechanics that allow strategy and player choice are

KICKSTARTER



- 13% (2,502) of the games are in the Kickstarter family
- Kickstarter games have a mean Geek Rating of 5.78 while non-Kickstarter games have a mean Geek Rating of 5.68
- A t-test determined that this is statistically significant with an infinitesimally small p-value i.e. the difference in mean Geek Rating is not up to chance

FEATURE ENGINEERING & PREPROCESSING

01

Dummy variables created for each category and mechanic label which generated 269 features. Games missing labels were assigned “None”.

02

Missing values for numerical features were imputed with KNearestNeighbors and scaled to a range between 0 and 1.

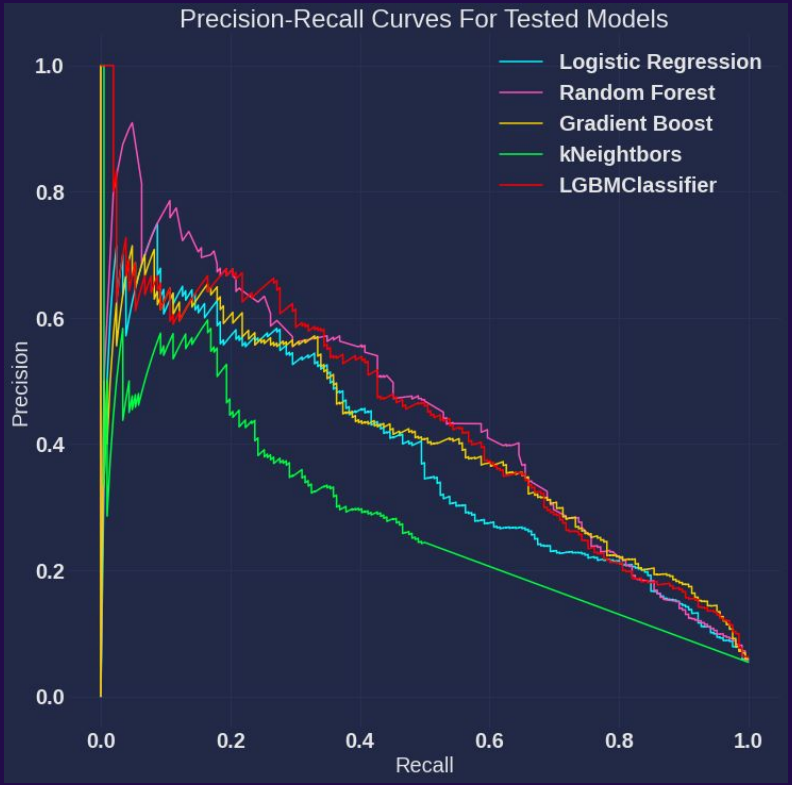
03

Binary column created to indicate if a game was Kickstarter or not. Family label was dropped due to high cardinality.

04

Created a target variable indicating if a game is in the Top 1,000.

MODEL SELECTION



Model	Accuracy	Precision Recall AUC	F1
Logistic Regression	0.836	0.379	0.339
Random Forest	0.950	0.450	0.203
Gradient Boost	0.949	0.406	0.358
k-Neighbors	0.947	0.281	0.197
LGBM Classifier	0.951	0.437	0.380

LGBMCLASSIFIER TUNING

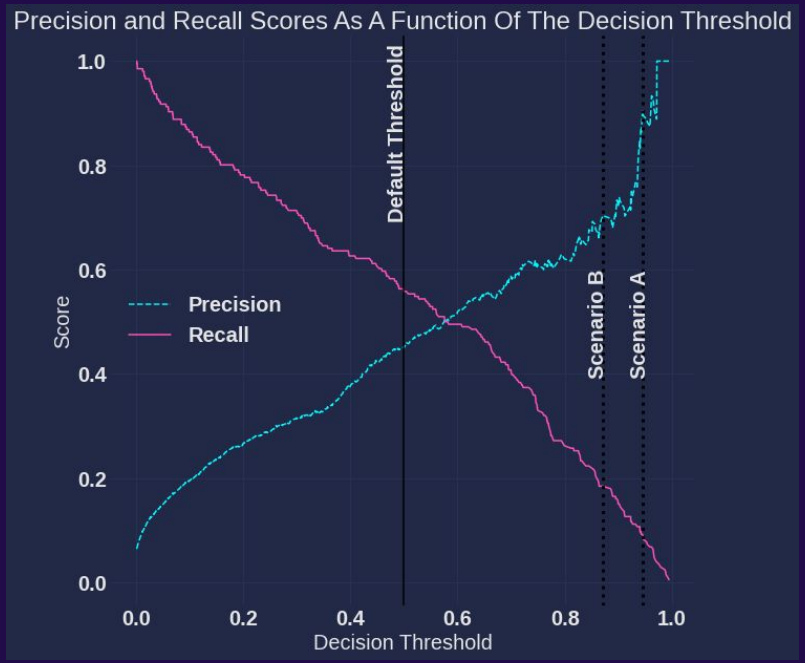
Model	Accuracy	Precision Recall AUC	F1
Base Model	0.951	0.437	0.380
MCA	0.951	0.423	0.295
OverSampled	0.948	0.469	0.472
Oversampled & Tuned	0.939	0.498	0.499

Steps taken to tune the model:

1. Tested MCA to reduce dimensionality
2. Oversampled minority class
3. Bayesian Hyperparameter tuning

Improved F1 Score by 0.119 (31%) over base model

DECISION THRESHOLDS

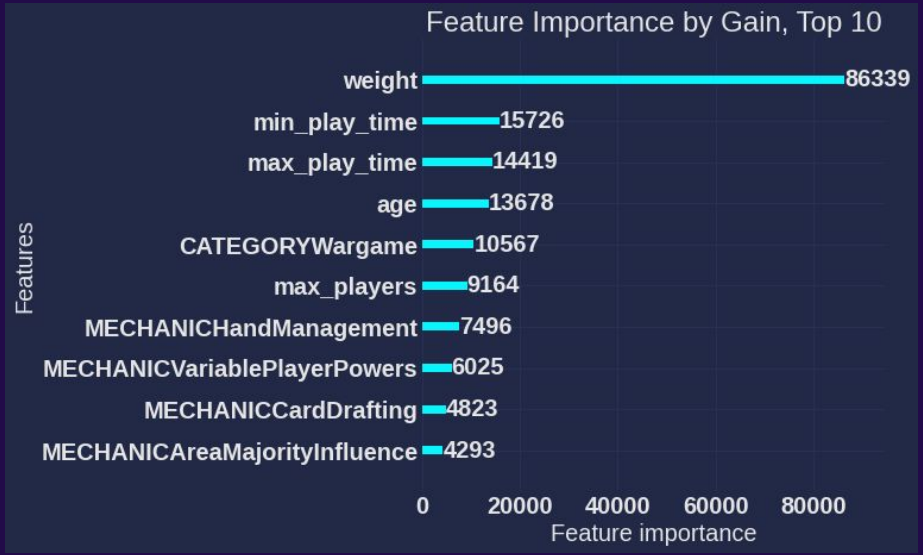


Scenario A: This publisher is new. Perhaps they are a startup or a hobbyist trying to break into the market with a Kickstarter campaign.

Scenario B: In this scenario the game publisher is an established business with a successful track history and experienced game developers on staff.

Business Case	Threshold	Precision
Scenario A	0.945	0.900
Scenario B	0.870	0.700

FEATURE IMPORTANCE



- Weight provides the most information by a significant margin
- All mechanics except Area Majority Influence are in the bottom 60 mechanics by Geek Rating (out of 283 total mechanics)
- The Wargame category is the 30th worst rated category out of 83 by Geek Rating

MISCLASSIFICATIONS

False Positive - Uchronia (Rank 2,628)

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.” - user Papa NinJa

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

MISCLASSIFICATIONS

False Positive - Magic: The Gathering – Heroes of Dominaria Board Game (Rank 5,351)

4/10 - "**Rather bland**. So much could be pulled from the magic universe, this is really **vanilla**." - user bneffer

3/10 - "**This game is the okay-est game I've played**. It is inoffensive, unless there is any art by Noah Bradley have not checked yet. If you can get me less than $10 than **it can stop your board games from falling over on your shelf**." - user The Game Piece

3/10 - "This is **competent** and just as exciting as something you would describe as competent." - user Goregrimm

MISCLASSIFICATIONS

False Negative - KLASK (Rank 289)

Use magnets to push a ball into the goal, but beware of falling in yourself!

Weight: 1.05
Category: Action / Dexterity & Real-time

8.5/10 - "Great foosball like fun with much more strategic variance than I typically expect in a dexterity game." - user neflight86

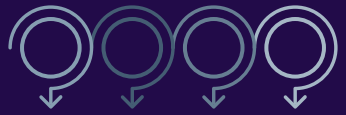
9/10 - "I HATE dexterity games. This should say enough about my rating." - user cuazzel

CONCLUSION

- The Light GBM Classifier is the best model with a 0.939 accuracy score, 0.498 precision-recall AUC and 0.499 F1 score after oversampling and hyperparameter tuning
- This model is useful for both publishers in game creation and small and large retailers for inventory management
- Great games...
  - Have a high weight
  - Allow for strategy and player agency
  - Are more than the sum of their features
    - Misclassifications help guide us

THANKS!

Do you have any questions?
samtbeardsley@gmail.com



APPENDIX

5 BEST RATED MECHANICS
(COUNT > 50)

	Count	Average Geek Rating
End Game Bonuses	54	6.727
Race	59	6.364
Communication Limits	70	6.242
Traitor Game	64	6.219
Solo / Solitaire Game	420	6.164

* see appendix for descriptions of these mechanics

MECHANIC DEFINITIONS

In **Rock-Paper-Scissors** there are three possible options, and they are cyclically superior (A beats B, B beats C, and C beats A). The name derives from the well-known children's game where Scissors cuts Paper, Paper covers Rock, Rock crushes Scissors, but can refer to any game with non-transitive mechanisms.

Roll/Spin and Move is defined as "games where players roll dice or spin spinners and move playing pieces in accordance with the roll. This term is often used derogatorily to imply that there is no thought involved. Roll and move games like Backgammon, however, contain tactical elements."

Zone of Control games are where "spaces adjacent to a unit impact the ability of opposing units to move or attack. This is a very common mechanism in wargames."

Force Commitment is a mechanic where players choose how many and how to use their units in battle.

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 T**urn 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 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 pay offs.

Games designed to be or have an option with rules on how to play by yourself are under the **Solo / Solitaire Game** mechanic.

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. Restricting communication keeps players immersed in the game, limits overplanning and over communication, and allows for unique and surprising situations to arise.

The **Race** mechanic is any sort of game when a player wins by accomplishing a fixed goal.

End Game Bonuses work off of a system where players only gain the points at the end of the game.

TERMINOLOGY

Accuracy: the percent of correct predictions. Not a great measure for model performance on a imbalanced dataset. Using spam detection as an example, if only 1% of emails are spam, a model can have a 99% accuracy by predicting that all emails are not be spam.

Precision: the number of correctly identified positive results divided by the number of all positive results.

Recall: the number of correctly identified positive results divided by the number of all samples that should have been identified as positive.

F1 score: A measure of a tests accuracy and is the harmonic mean of the precision and recall. The highest possible value of F1 is 1, indicating perfect precision and recall, and the lowest possible value is 0, if either the precision or the recall is zero.

Precision Recall AUC: A precision-recall curve (or PR Curve) is a plot of the precision (y-axis) and the recall (x-axis) for different probability thresholds. Precision Recall AUC summarizes the PR curve with a range of threshold values as a single score. The score (representing the area under the curve) can then be used as a point of comparison between different models on a binary classification problem where a score of 1.0 represents a model with perfect skill.

Decision (or Probability) Threshold: Classification algorithms calculate a probability that a data point is of a certain class. This threshold defaults to 0.5, but can be adjusted depending on use case.

T-test: A type of inferential statistic used to determine if there is a significant difference between the means of two groups.

P-value: The level of marginal significance within a statistical hypothesis test, representing the probability of the occurrence of a given event.