What Makes a Popular Board Game?

...

Questions to answer:

What factors lead to a board game's popularity?

Can we use these factors to influence design decisions for our first board game?

Assumptions

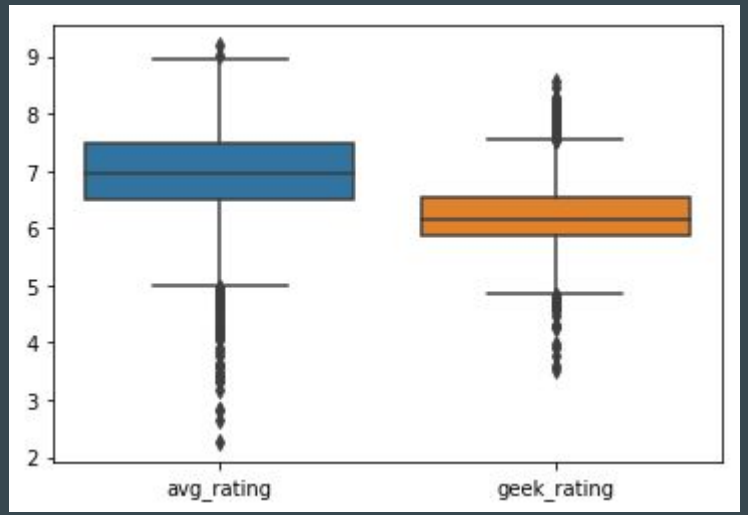
A board game's popularity on BoardGameGeek web site is correlated to success and revenue.

Information source:

bgg_db.csv dowloaded 6/4/21 from Kaggle:
https://www.kaggle.com/phizzuela/boardgamegeek2020

Model Decision Points

Use “Geek Rating” or “User Rating”?



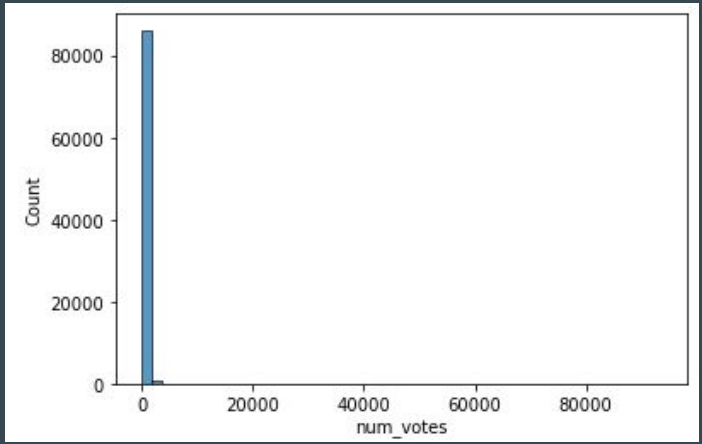
Both ratings are similar.

“Geek Rating” is Baysian average of user rating, so user rating will be studied here

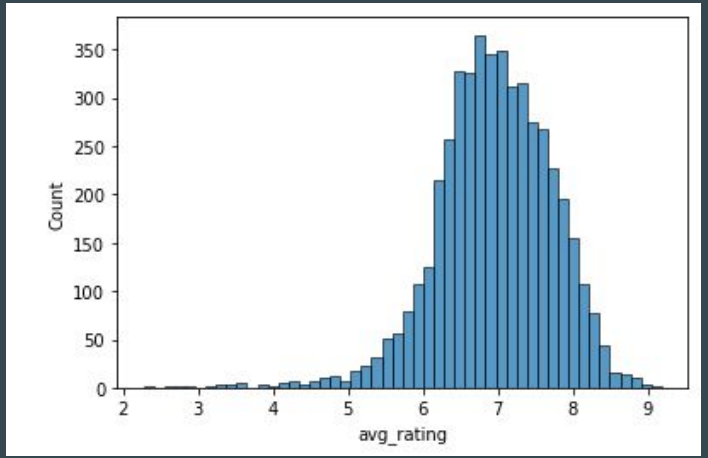
Which Games to Examine?

Original file had 116455 entries, but many of which held little information.

Many had a rating of zero and 75% had less than 41 total votes. Decision is to filter for only games with 500 or more votes. This left 4768.



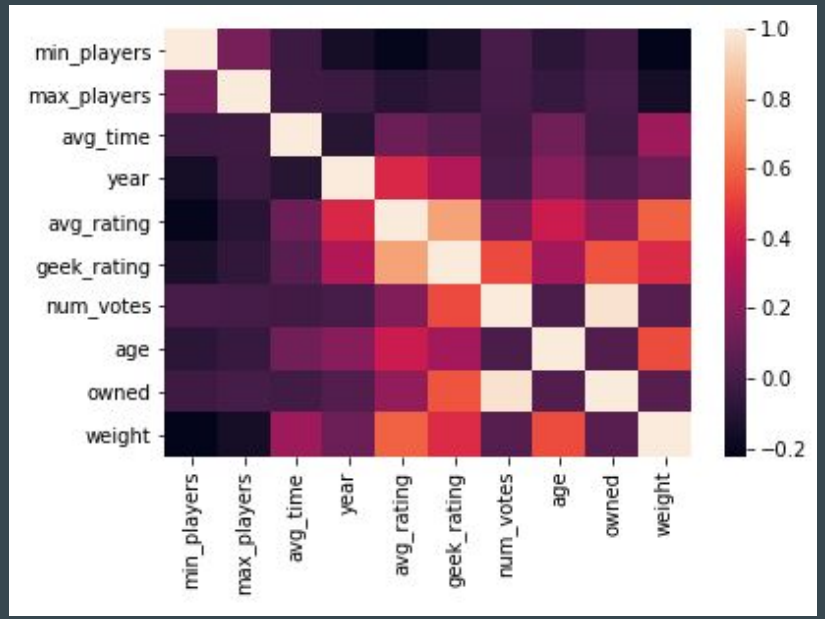
Number of votes clustered at low end



Ratings of games with >500 Votes

Which Factors to Study?

Many factors are related to each other and can be eliminated from the model:



Correlations between numerical factors

These factors were discarded as related to other factors:

Geek Rating
Number of Votes
Number of Users who Own

A Categorical Problem

- Initial aim of the study was to determine a “killer app” combination of mechanics that would lead to a popular game.
- Unfortunately, each game had a “category” column and a “mechanic” column that listed multiple words.
- The number of possible combinations was high
  - 84 possible categories and 182 possible mechanics
- Using only the first 5 values in each of the “category” and “mechanic” columns, “One Hot” encoding led to 262 sparsely encoded columns.
- Attempted Lasso Regression dropped every column but “weight” and “year”
- Solution: Frequency Encoding!

What's the Frequency?

- All of the categories and mechanics were counted and each was divided by the total
- This gives the frequency of an individual category/mechanic
- Because each game had multiple attributes, the frequency value for each was averaged to give one "category score" and "mechanic score" for every game

The final factors to be analyzed were:

Minimum number of players	Average user rating
Maximum number of players	Recommended age
Average play time	Weight
Year of release	Category Average
Mechanic Average	

Model selection

Four different regression models were created to predict User Rating.

Models were evaluated based on the R-Squared value and Mean Squared Error (MSE)

	Linear	Random Forest	Lasso	Ridge
R-Squared	0.497	0.482	**0.498**	0.497
MSE	0.239	0.246	**0.238**	0.239

Best model is the Lasso Regression

Evaluation

Two Lasso Regression Coefficients were an order of magnitude greater than the rest

- Weight of game: 0.359
- Year of release: 0.270
- All others: < 0.063

Random Forest Feature Importances tell the same story:

- Weight of game: 0.434
- Year of release: 0.245
- All others: < 0.089

The two most important factors in a popular game are Weight and Release Date

Conclusion

No magic bullet

No combination of factors that lead to a guaranteed popular game

Complication is king

The users of BoardGameGeek like more complicated (“weightier”) games

The new hotness

Newer games are more popular

Further questions

- Because “weight” was the strongest predictor of popularity, is BGG popularity correlated with financial success?
  - Need to examine sales numbers
  - BGG users may cause selection bias towards complication
- Examine data to determine if any specific mechanics are correlated to “weight”
  - May help answer the original question