

Academia to Board Game Brilliance

A Data-Driven Strategy for Academia The Board Game



June 2024

Academia to Board Game Brilliance - A Data-Driven Strategy for 'Academia: The Board Game

Paulo Henrique da Silva Mota

Summary

Data Preparation and Description	3
Data Source	3
Data Preparation	6
Performance Measures	6
Key Variables	7
Data Summary	7
Data Analysis	9
Relationships with Performance Measures	9
Model Evaluation	10
Causality Discussion	10
Results Presentation	10
Strategy Recommendation	10
Limitations	11
Additional Data and Analysis	11
Conclusion	11
References	11

Data Preparation and Description

Data Source

The data source pertains to BoardGameGeek Reviews and is composed of two datasets. The first dataset contains details about various board games, and the second dataset includes reviews from users for these games, linked by an ID. All reviews are from users, and their comments are included.

Details Dataset:

Number of observations: 21631

Number of columns: 23

The Board detail dataset contains information about many board games, where each observation represents a board game.

Table 1: Board game details dataset

variable	class	description
num	double	Game number
id	double	Game ID
primary	character	Primary name
description	character	Description of game
yearpublished	double	Year published
minplayers	double	Min n of players
maxplayers	double	Max n of players
playingtime	double	Playing time in minutes
minplaytime	double	Min play time
maxplaytime	double	Max play time
minage	double	minimum age
boardgamecategory	character	Category
boardgamemechanic	character	Mechanic
boardgamefamily	character	Board game family
boardgameexpansion	character	Expansion
boardgameimplementation	character	Implementation
boardgamedesigner	character	Designer
boardgameartist	character	Artist
boardgamepublisher	character	Publisher

variable	class	description
owned	double	Num owned
trading	double	Num trading
wanting	double	Num wanting
wishing	double	Num wishing

Ratings Dataset:

Number of rows: 21831

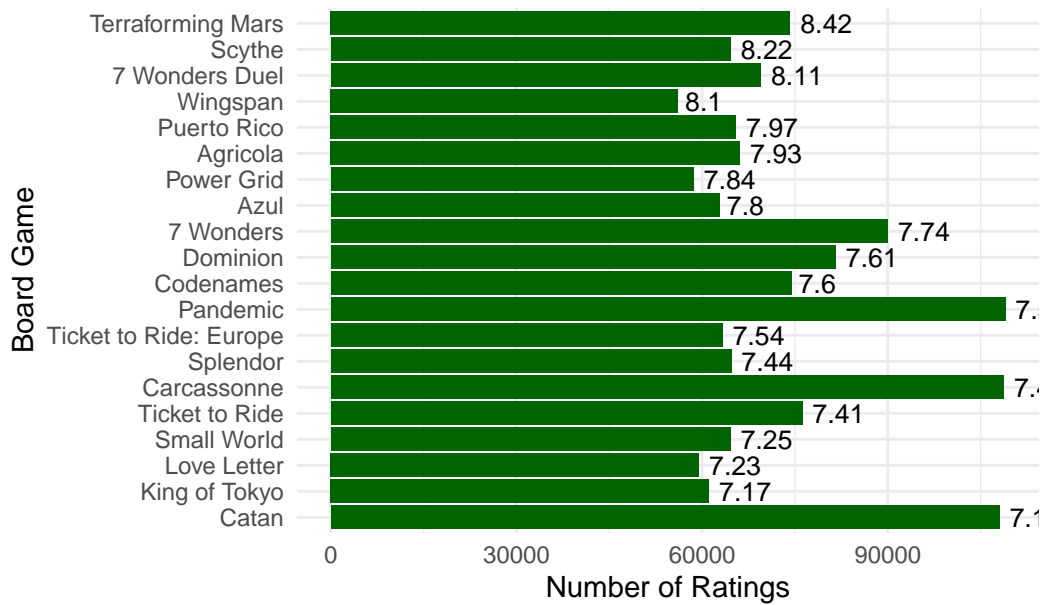
Number of columns: 10

Table 2: Ratings dataset

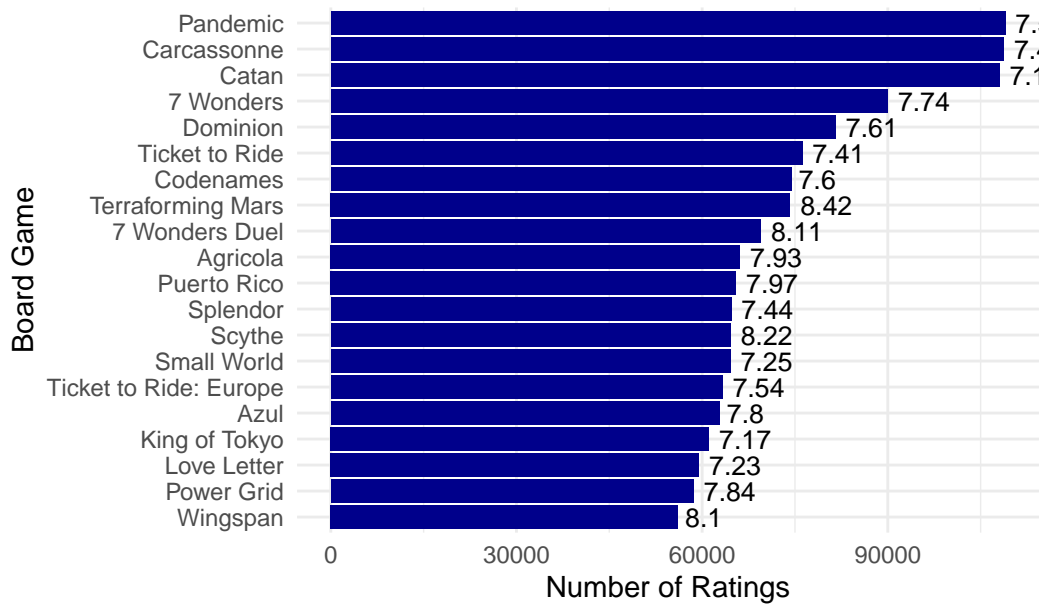
variable	class	description
num	double	Game number
id	double	Game ID
name	character	Game name
year	double	Game year
rank	double	Game rank
average	double	Average rating
bayes_average	double	Bayes average rating
users Rated	double	Users rated
url	character	Game url
thumbnail	character	Game thumbnail

Here are some trends about the boarding games:

Top 20 Board Games by Average Rating



Top 20 Board Games by Number of Ratings



Data Preparation

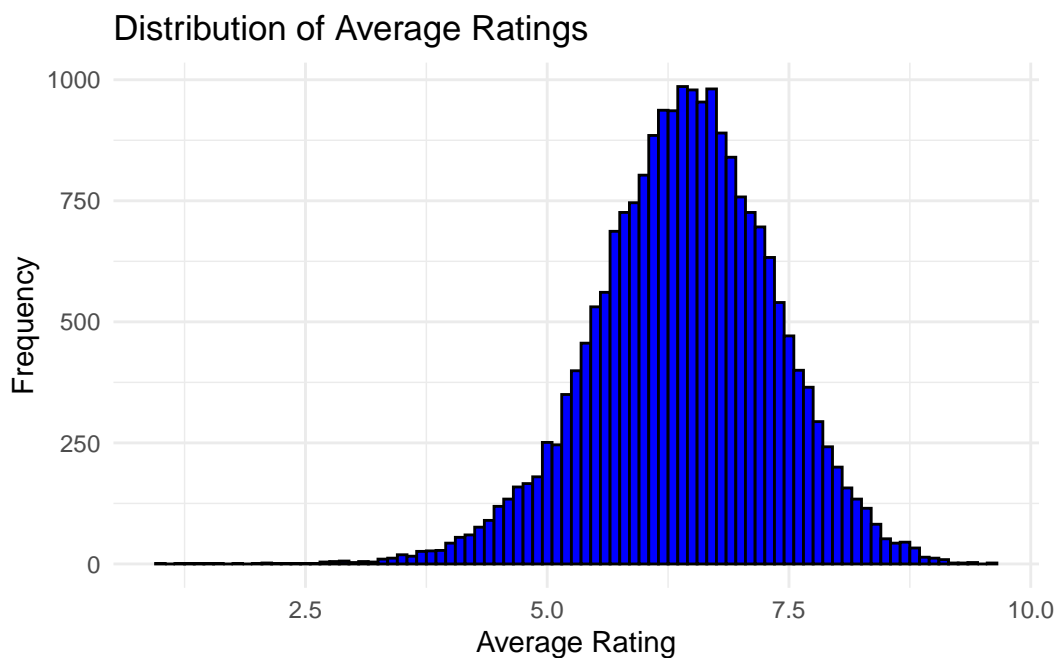
Detail the steps taken to clean and preprocess the data in R. Include code snippets where appropriate.

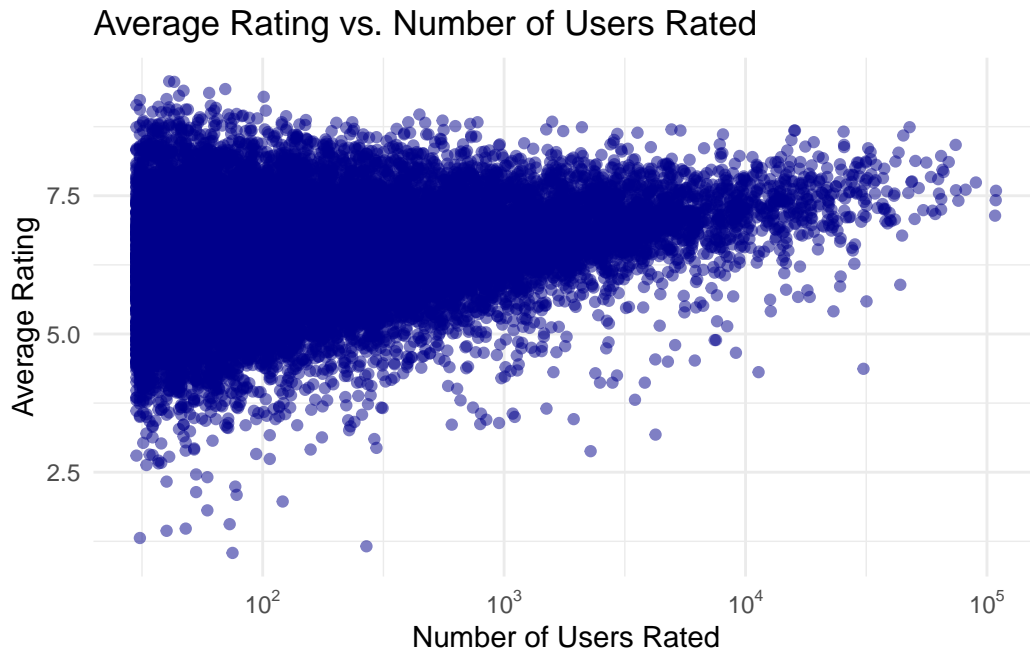
In order to have a better dataset to make our analysis, first I merged details with ratings by the column ID, following the instructions in the dataset documentation, with my dataset merged I will remove all games with no ratings information. After that I plot the correlation matrix to identify possibly correlation with my variables.

Performance Measures

In the context of profitability within a board game, two critical performance metrics stand out: the average rating, denoted by the 'average' column, as higher ratings typically signify superior market reception; and the number of reviews, captured by the 'users_rated' variable, as a greater volume of reviews tends to correlate with increased sales and heightened popularity.

The charts below illustrate the distribution of the performance metrics:





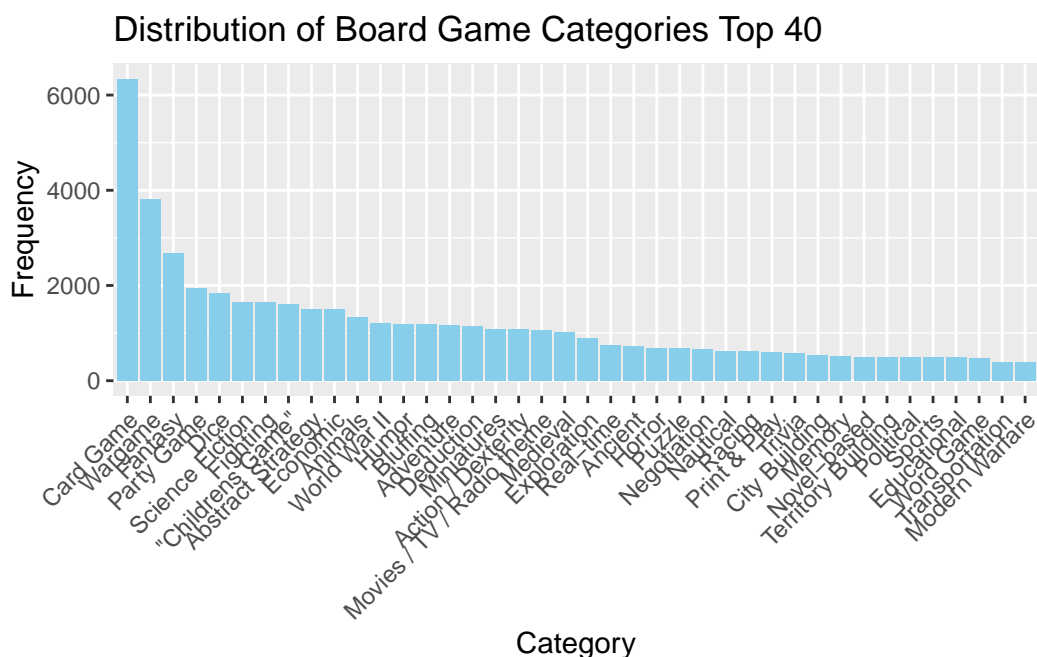
Key Variables

The variables influencing game design decisions within our dataset are closely intertwined with the intrinsic traits of board games. These variables encompass Category and Mechanic where specific genres may demonstrate stronger financial performance, and Play Time, which profoundly shape a game's attractiveness to potential audiences. Moreover, MaxPlayers act as indicators of the potential player base, potentially aligning with increased review counts.

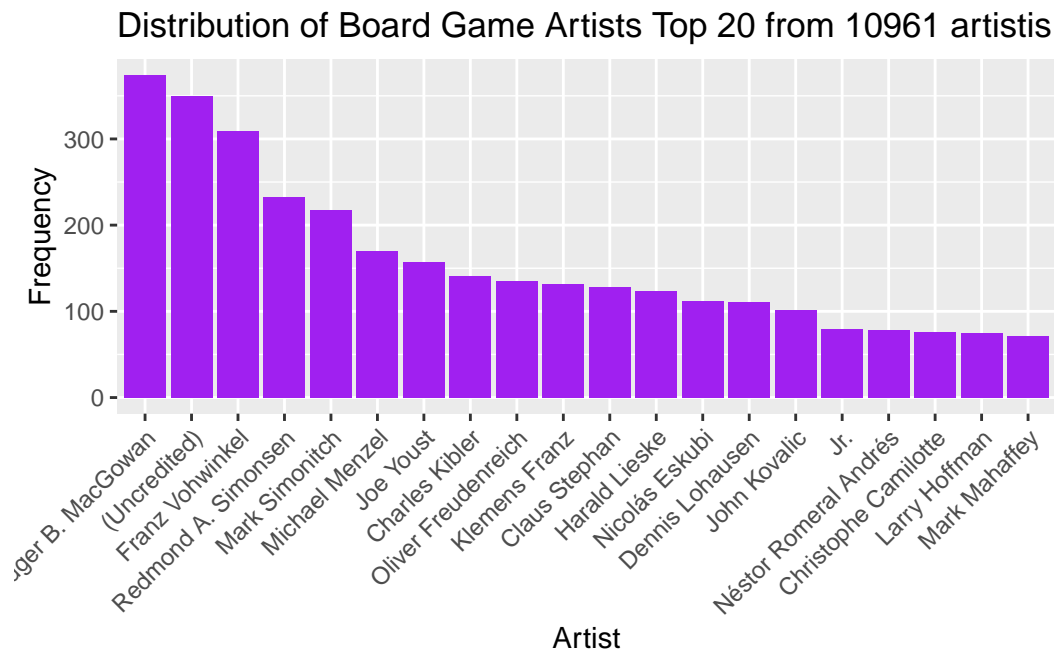
In addition to these variables, it's crucial to consider certain control variables to ensure a comprehensive analysis. Control variables such as Year of Release, denoted by "Year Published," are vital as newer games may experience different market dynamics compared to older ones. Moreover, factors like Artist, Designer, and Publisher reputation play significant roles and could impact the outcomes of success, warranting their inclusion in the analysis for a more nuanced understanding of performance metrics within the board game industry.

Data Summary

The chart below you can see the distribution form the categorical values and realize that the top categories may have the best probability to be more successful, for being high trended in the data. The same reason is applied to the mechanic.



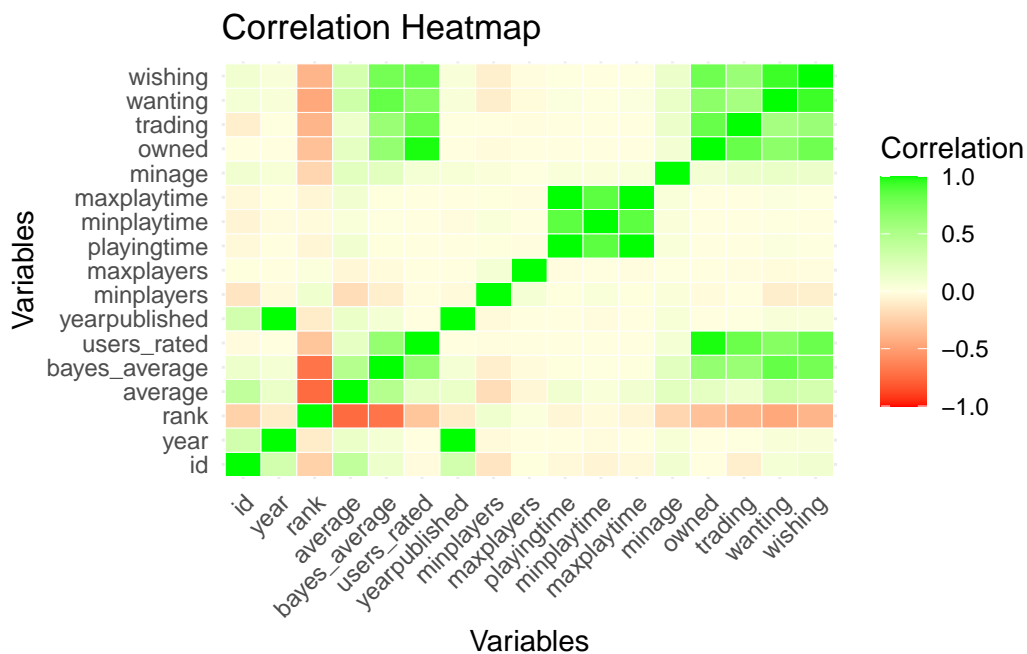
About the artist, design and publishers, it is more about reputation, some new fame with the right names could end up successful just because the best experience that those users had with a previous game. These people and companies become well known between the community with more credibility.



Data Analysis

Relationships with Performance Measures

Analyzing the correlation was my initial step in identifying potential variables to include in my linear regression model. This analysis helps to determine the relationships between the measured variables. We observed that the variables “owned,” “wishing,” and “wanted” are highly correlated with our measured variables. This suggests that by increasing the average rating and the number of user reviews, we can potentially enhance the profitability of a board game.



Call:

```
lm(formula = average ~ users_rated + maxplayers + playingtime +  
    yearpublished + boardgamecategory + boardgamemechanic + boardgameartist +  
    boardgamepublisher + boardgamedesigner, data = clean_data)
```

Residuals:

Min	1Q	Median	3Q	Max
-35.133	-0.692	-0.075	0.735	6.633

Coefficients:

Estimate	Std. Error	t value	Pr(> t)
----------	------------	---------	----------

(Intercept)	3.666e+00	4.818e-03	760.99	<2e-16 ***
users Rated	2.698e-05	4.176e-08	646.02	<2e-16 ***
maxplayers	-2.180e-02	9.551e-05	-228.28	<2e-16 ***
playingtime	5.684e-04	2.150e-06	264.36	<2e-16 ***
yearpublished	1.050e-03	2.223e-06	472.21	<2e-16 ***
boardgamecategory	4.384e-03	2.436e-05	179.95	<2e-16 ***
boardgamemechanic	-2.080e-03	1.074e-05	-193.65	<2e-16 ***
boardgameartist	-1.327e-05	1.725e-07	-76.93	<2e-16 ***
boardgamepublisher	-3.973e-05	2.862e-07	-138.81	<2e-16 ***
boardgamedesigner	2.109e-04	1.858e-07	1135.03	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.002 on 3383229 degrees of freedom

(128350 observations deleted due to missingness)

Multiple R-squared: 0.507, Adjusted R-squared: 0.507

F-statistic: 3.865e+05 on 9 and 3383229 DF, p-value: < 2.2e-16

Model Evaluation

Discuss the appropriateness of the model based on R output. Explore other methods to test the relationships (e.g., ANOVA, correlation analysis).

Causality Discussion

Examine the potential for causal inference in the dataset. Discuss limitations and assumptions.

Results Presentation

Provide a single result table and a figure to present the analysis findings.

Strategy Recommendation

Strategic Insights Build on the analysis to provide strategic recommendations for the game design.

Limitations

Discuss the limitations of the analysis and recommendations.

Additional Data and Analysis

Suggest additional data that could further inform the strategy and potential analyses that could be conducted with it.

Conclusion

Summarize the key findings and strategic recommendations. Reinforce the goal of creating a successful and profitable board game.

References

List all sources and references used in the report.