Term Project (Step 3)

Kinnick Fox

2023-02-18

Introduction

Board gaming has had a resurgence in the last couple of decades leading to more games being developed by the day. Many developers are players first, and believe they have an idea that could make for a successful board game. This project is designed to distinguish which attributes of a board game could lead to it's successfulness.

Problem Statement

What variables are most commonly found in successful board games?

Addressing the problem statement

A dataset that contains fields for mechanics, themes, player count, user rating, game weight, and year published was curated from Kaggle (link below). Success will be determined by whether or not a game resides in the top 25% based on rating. Bayes average rating will be used to adjust for board games that have a lower numbers of ratings.

Board Game Database from BoardGameGeek - https://www.kaggle.com/datasets/threnjen/board-games-database-from-boardgamegeek?select=artists_reduced.csv

Analysis

The following tools will be required for this analysis.

```
library(caTools)
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(purrr)
library(knitr)
```

Each CSV containing pertinent data will need to be imported. Any data points that have no input for category (found in games.csv), mechanics, sub-category, or theme will need to be dropped. Not dropping these data points results in every variable appearing to be influential to a board game's rating because any board game that lacks sufficient data is unpopular and suffers a poor rating. The third quadrant of the rating summery show the cutoff for the 75th percentile to be a rating of 5.68.

```
games <- read.csv("games.csv", header = TRUE)
summary(games[5])
```

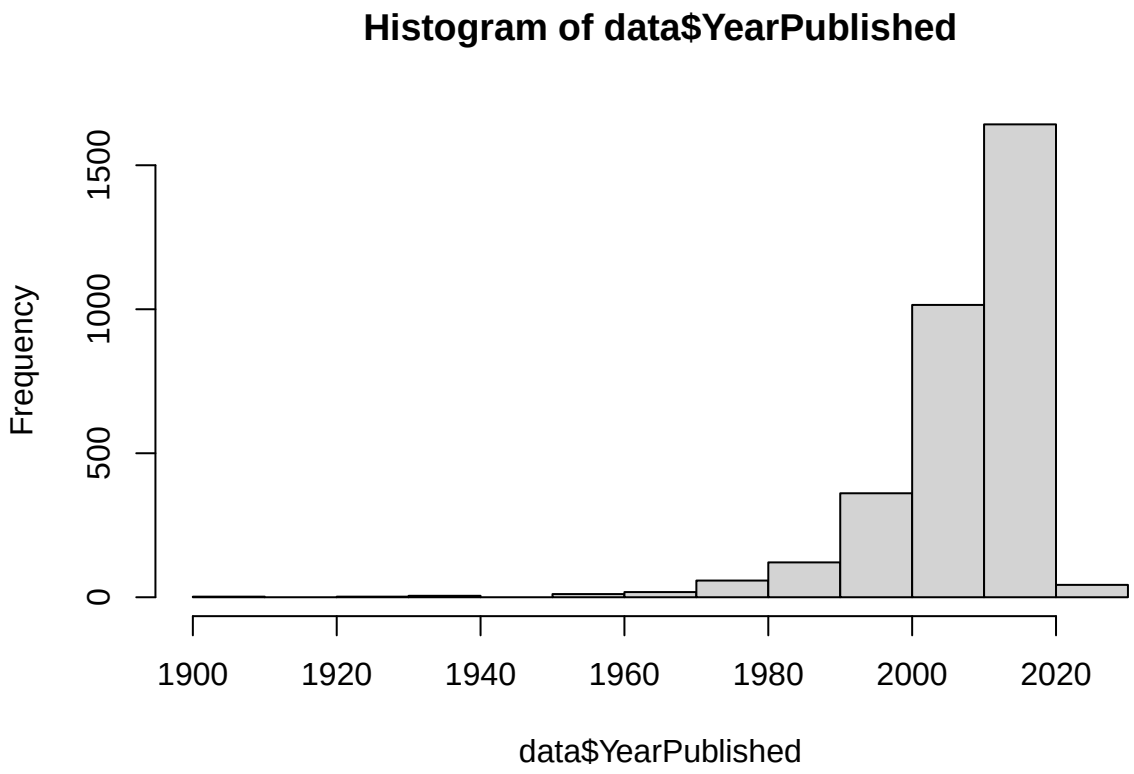
```
##  BayesAvgRating
##  Min.   :3.575
##  1st Qu.:5.510
##  Median :5.547
##  Mean   :5.686
##  3rd Qu.:5.680
##  Max.   :8.515
```

```
games <- games[rowSums(games[,-1:-10])>0,]
mech <- read.csv("mechanics.csv", header = TRUE)
mech <- mech[rowSums(mech[,-1])>0,]
subcat <- read.csv("subcategories.csv", header = TRUE)
subcat <- subcat[rowSums(subcat[,-1])>0,]
themes <- read.csv("themes.csv", header = TRUE)
themes <- themes[rowSums(themes[,-1])>0,]
```

The data from each CSV is then condensed down to a single dataset while dropping any rows that do not have sufficient data in all fields.

```
raw <- games %>% inner_join(mech,by="BGGId")
raw <- raw %>% inner_join(subcat,by="BGGId")
raw <- raw %>% inner_join(themes,by="BGGId")
```

```
data <- raw[raw$YearPublished >= 1900,]
hist(data$YearPublished)
```



All data from before 1970 was considered an outlier and removed.

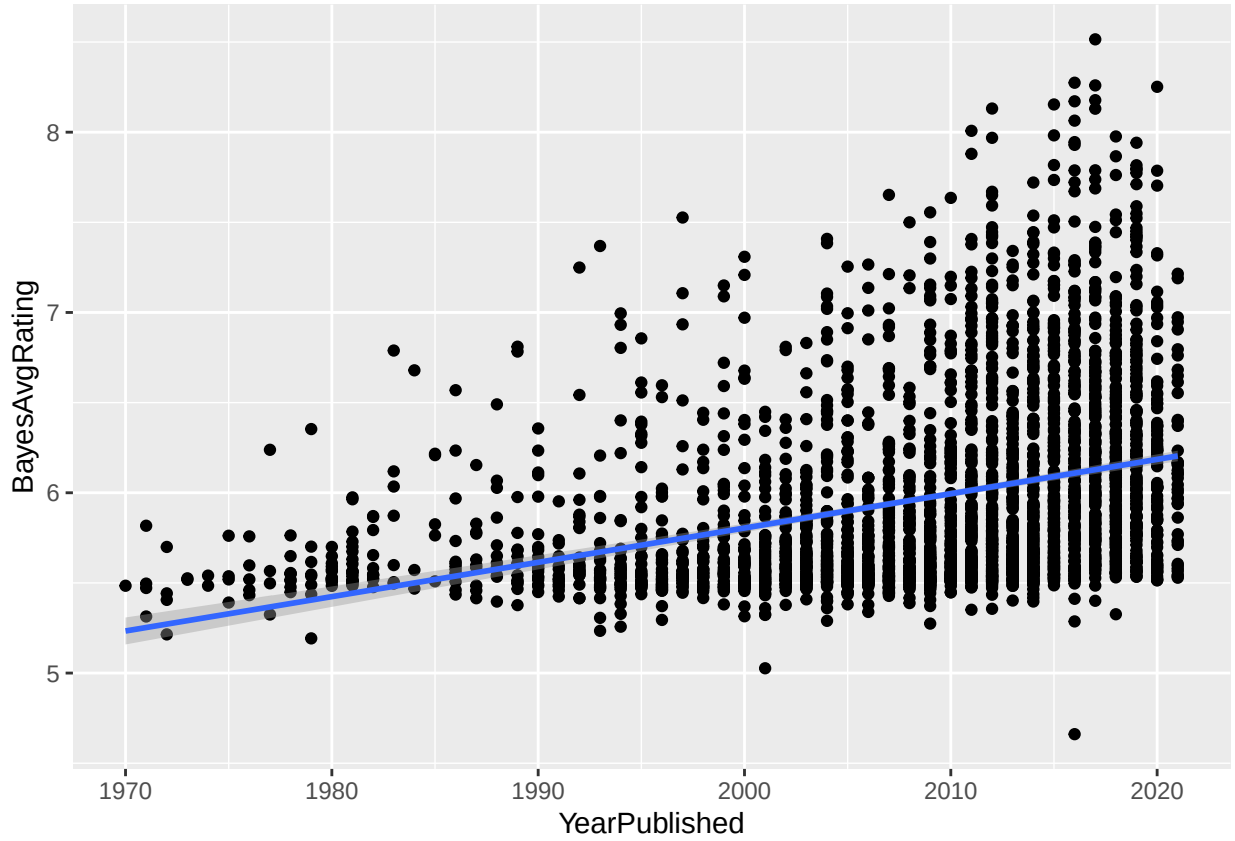
Any data the had zero values in player count or manufacturer recommendations was removed.

Any column that had less than 20 instances was removed.

```
data <- data[,-2]
data <- data[data$YearPublished >= 1970,]
data <- data[data$MinPlayers > 0,]
data <- data[data$MaxPlayers > 0,]
data <- data[data$MfgPlaytime > 0,]
data <- data[data$MfgAgeRec > 0,]
data <- data[data$GameWeight > 0,]
data <- data[, which(colSums(data[]) >= 20)]
```

```
model <- glm(BayesAvgRating>5.68 ~ . -BGGId, data = data)
ggplot(data, aes(x=YearPublished, y=BayesAvgRating)) + geom_point() + geom_smooth(method=lm)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



Variables with a high impact on board game successfulness.

```
coef <- summary(model)[12]
coef.sort <- as.data.frame(coef)
coef.sort <- coef.sort[-1,]
names(coef.sort) <- c("Estimate","SE","Tval","Pval")
knit_print(arrange(coef.sort[coef.sort$Pval < 0.001,],Pval))
```

```
##                                    Estimate          SE      Tval         Pval
## YearPublished                   0.007262679 0.001024916  7.086122 1.723017e-12
## Cat.War                        -0.175352183 0.035991874 -4.871994 1.163887e-06
## Take.That                       0.134152369 0.030359843  4.418744 1.029027e-05
## Deck..Bag..and.Pool.Building    0.141648853 0.032711486  4.330248 1.539719e-05
## Kickstarted                    -0.092881860 0.021779943 -4.264559 2.066742e-05
## Cat.Childrens                  -0.201075640 0.047467951 -4.236029 2.345652e-05
## Drafting                        0.090510257 0.022675874  3.991478 6.728409e-05
## Racing                         -0.226600494 0.059431269 -3.812816 1.402417e-04
## Simultaneous.Action.Selection   0.117038488 0.031071033  3.766804 1.686255e-04
## Variable.Player.Powers          0.079632721 0.021557221  3.694016 2.248002e-04
## Cat.Family                      0.096895303 0.026319743  3.681468 2.361055e-04
## Modern.Warfare                 -0.246903735 0.068657690 -3.596156 3.283228e-04
## GameWeight                      0.066225684 0.019032672  3.479579 5.095777e-04
```

Variables with a moderate impact on board game successfulness.

```
knit_print(arrange(coef.sort[coef.sort$Pval < 0.01 & coef.sort$Pva > 0.001,],Pval))
```

```
##                               Estimate         SE     Tval        Pval
## Hand.Management           0.06128566 0.01862533 3.290446 0.001012211
## Push.Your.Luck            0.11898877 0.03753324 3.170224 0.001539034
## Nautical                  0.13111805 0.04497849 2.915128 0.003582570
## Area.Majority...Influence 0.08356663 0.02922920 2.859012 0.004279859
## Deduction                 0.11740656 0.04219104 2.782737 0.005425022
## Trick.taking              0.17944468 0.06489564 2.765127 0.005725933
## Cooperative.Game          0.08934766 0.03279282 2.724611 0.006476218
## TableauBuilding           0.13814883 0.05104399 2.706466 0.006840034
## Layering                  0.23811820 0.09117877 2.611553 0.009059484
## Set.Collection            0.05904900 0.02261209 2.611391 0.009063773
```

Data shows that year published has the highest impact on board game successfulness. Twenty-three of the initial 402 variables have at least a moderate impact on board game successfulness. Take that, bag building, and drafting are all different mechanics and have the highest positive influence on a board game's success. War and childrens' games as well as games that were funded on Kickstarter have the most negative effect on a board game's success.

Implications.

Year published being the most influential variable implies that board games are becoming generally more competent in the eye of players as time goes on. War and childrens' games are considered too drawn out and too simple respectively, so it is no surprise that board games in these categories would not be found in the previously stated success range. Games funded on the crowed funding site Kickstarter are also unlikely to be considered successful by this project. This makes sense as game developers using this platform are often beginners who have never developed a board game before or dealt with the logistical challenges of such. The mechanics that most impact a board game's successfulness positively are take that, bag building, and drafting implying that a board game that involves one or all of these mechanics will be favored in by the board game community.

Limitations.

The limitations of the project lie with the dataset. All data found on Board Game Geek (the website scraped for this dataset) is community driven. This means that the more popular a game is, the more fact-checking and data entry occurs. This also means less popular games that have the potential to be considered successful by this project are not included in the analyzed dataset due to not containing enough data. This dataset is also completely dependent on users truthfully rating board games based on enjoyment of play without influence of quality of materials, setting played in, or quality of fellow players. This survey style dataset can provide many insights but also must contend with many caveats.

Concluding Remarks.

All and all, this project has highlighted some mechanics a developer might consider when developing a game as well as some "red flags" to steer clear from. Many variables exist when developing a board games and this analysis is in no way meant to discredit works that include elements that are considered "a mark if unsuccessfulness" by this analysis.