I. Introduction

Board games are a popular past time, not only in the United States but also in other parts of the world as well. They allow us to meet new people or strengthen old relationships, increase our brain function, teach us how to set goals and be patient, reduce stress, and enhance creativity and self-confidence. Games can range from simple to complex, and can include as few as two people and as many as you can fit at your table. There are an abundance of themes that allow you to command an army trying to take over the world to being a cowboy in the wild west. Chances are very good that there is a board game for almost every interest out there, and if there's not, the innovation of larger, more established companies like Parker Brothers or Milton-Bradley, as well as smaller, independent companies will almost certainly produce them.

For this study, I utilized a dataset comprised of more than 10,000 board game titles listed on the website BoardGameGeek.com. The following are the variables I used for analysis that are native to the dataset, along with a brief description of them:

- YearPublished: the year in which a board game was published.
- MinPlayers: the minimum number of players required.
- MaxPlayers: the maximum number of players allowed.
- PlayTime: the estimated amount of time needed to play in minutes.
- MinAge: the minimum age requirement in years.
- RatingAverage: the average rating users of BoardGameGeek.com (BGG) gave the title on a scale of 0-10
- ComplexityAverage: the average complexity rating BGG gave a title on a scaled of 1-10
- Domain: the genre of a game

In addition to variables mentioned above, I also created two variables to allow for further analysis of the dataset:

- Period: the time period in which a game was published (1900-1919, 1920-1939, etc.)
- Popular: if a board game had an average rating of 7.5 or over, it is considered popular

I used frequency analysis to see the breakdown of the three categorical variables Domain, Period, and Popular, and generated the descriptives of the numerical variables MinPlayers, MaxPlayers, PlayTime, MinAge, RatingAverage, and ComplexityAverage. The linear relationship between Complexity, RatingAverage, and MinAge was investigated, as well as a cross table showing the dependence/independence of the variables Period and Domain. Further, I developed a multiple regression model to see if RatingAverage was affected by PlayTime, MinAge, and ComplexityRating, and I also developed a logistic regression model to see the impact of Year, MinPlayers, PlayTime, MinAge, and ComplexityAverage had on the binary variable Popular.

II. Frequency Analysis

For our categorical variables "Domain", "Period", and "Popular", frequency analysis allows us to see how each variable is split up among its categories. Table 1 displays the frequency analysis results, indicating that the proportion of board games that are not popular is vastly greater than that of popular board games, 85.7% vs 14.3%. Since the board game data set includes over 10,000 different games, it makes sense that only a small %age of games would be considered popular. In addition, over half of the games in this dataset were published in 2000 and beyond (69.2%), while only 0.1% of the games were produced between 1900-1919. 1980-1999 and 1960-1979 produced the second and third most board games respectively, with 1980-1999 producing 22% of the games and 1960-1979 producing 7.1% of the games.

Furthermore, 29.6% of board games are considered to be wargames, and the second highest %age of board games are Thematic. The lowest proportion of board game genres were Customizable, which sat at 2.9%, and the second lowest proportion were Party games at 5.9%. These results can be attributable to the increased popularity of of wargames in the in the 1970s and 1980s that continue to this day, while the rise in popularity of party games did not really happen until much later. Interestingly, there is no data here for the time period 1940-1959. I suspect that it is because of the effects of World War II.

Table 1: Frequency Analysis

	Variable	Frequency	Percent
Domain	Abstract	1005	10
	Children	821	8.2
	Customizable	296	2.9
	Family	1692	16.8
	Party	598	5.9
	Strategy	1553	15.4
	Thematic	1120	11.1
	Wargames	2982	29.6
	Total	10067	100
Period	Pre-1900	83	0.8
	1900-1919	10	0.1
	1920-1939	78	0.8
	1960-1979	715	7.1
	1980-1999	2213	22
	Post-2000	6968	69.2
	Total	10067	100
Popular	No	8625	85.7
	Yes	1442	14.3
	Total	10067	100

III. Descriptive Analysis

Table 2 displays descriptive statistics analysis for MinPlayers, MaxPlayers, PlayTime, MinAge, RatingAverage, and ComplexityAverage. The overall sample size for the study is 10067, with minimum and maximum scores for each variable. The mean score for MinPlayers is 1.95 0.622 with a minimum of 1 and a maximum of 10. The mean score for MaxPlayers is 4.82 12.238 with a minimum of 1 and a maximum of 999. The mean PlayTime score for more than 10,000 board games is 133.10 762.970, with a minimum of 0 and a high of 60000. With a maximum of 21 and a mean of 9.80 3.652, the lowest score for MinAge is similarly 0. RatingAverage and ComplexityAverage have respective means of 6.5908 0.8654 and 2.2964 0.84888, with minimum and maximum values of 1.43 and 9.34 for RatingAverage and 0.00 and 5.00 for ComplexityAverage, respectively. The variance of PlayTime is greatest around the mean, whereas the variance of MinPlayers is lowest near the mean. The mean scores for MinPlayers, MinAge, RatingAverage, and ComplexityAverage were above average.

Table 2: Descriptive Statistics

Variables	N	Minimum	Maximum	Mean	Std. Deviation
MinPlayers	10067	1	10	1.95	0.622
MaxPlayers	10067	1	999	4.82	12.238
PlayTime	10067	0	60000	133.10	762.970
MinAge	10067	0	21	9.80	3.652
RatingAverage	10067	1.43	9.34	6.5908	0.86549
ComplexityAverage	10067	0.00	5.00	2.2964	0.84888

IV. Correlation Analysis

I analyzed the relationship between ComplexityAverage, RatingAverage, and MinimumAge. Normality analysis was initially applied. The Kolmogorov-Smirnov Test (Table 3) indicated that not all variables are normally distributed; hence, a Spearman's Rank-Ordered correlation (non-parametric) analysis was conducted.

Table 3: Kolmogorov-Smirnov(a) Tests of Normality for MinAge, RatingAverage, and ComplexityAverage

	Statistic	**df**	**Sig.**
MinAge	0.204	10067	0.000
RatingAverage	0.026	10067	<.001
ComplexityAverage	0.061	10067	<.001

a. Lilliefors Significance Correction

The findings of Spearman's correlation analysis are displayed in Table 4. MinAge and RatingAverage have a positive, modest, and statistically significant association, according to the data (r=32.7%, p-value < 0.001). In addition, there exists a positive, moderate, and statistically significant link between MinAge and ComplexityAverage (r=49.9 %, p-value < 0.001) and between RatingAverage and ComplexityAverage (r=54.1%, p-value < 0.001).

Table 4: Spearman’s Correlation analysis (n = 10067))

Variables	**MinAge**	**RatingAverage**	**ComplexityAverage**
1) Minimum Age	1		
2) Rating Average	.327** W	1	
3) Complexity Average	.499** M	.541** M	1

***p-value <0.001*

V. Cross Tabular Analysis

The association between the period in which a board game was produced and the board game's genre is depicted in Table 5. 69.2% of 10,067 titles were produced after the year 2000, while just 0.1 % were produced between the years 1900 and 1919. It has also been stated that 29.6 % of the same sample consisted of Wargames. Comparatively, 2.9% are of the Customizable genre. Family games made up 16.8 % of the board games available. 55.4 % of titles in the Abstract genre were created before 1900, whereas 0.00 % of titles in the Thematic and Customizable genres were created during the same time period. According to the same research, 22.2% of Wargames were produced after 2000, yet only 3.00 % of these games were deemed Customizable.

51.8 % of Wargames were produced after 2000, whereas none were produced before 1919, based on a vertical examination of the findings. Similarly, 70,9 % of the 2,9 % of Customizable board games were produced after 2000 and none were produced before 1979. 82.4 % of Family-themed board game titles were created after the year 2000, while only 0.2 % were created between 1900 and 1919.

Chi-square techniques of analysis indicated a statistical association between the era of a board game and its genre. Chi-Square (2) = 1487,468 with df = 35 and p-Value 0.0001. Since the p-value is less than 0.05, the null hypothesis of independence is rejected in favor of the alternative hypothesis indicating that school type and socioeconomic position are not independent variables.

Table 5: Cross tabular analysis between experience domain and period

	Domain								
Period	Abstract	Children	Customizable	Family	Party	Strategy	Thematic	Wargames	Total
Pre-1900	46	15	0	10	9	2	0	1	83
	55.4%	18.1%	0.0%	12.0%	10.8%	2.4%	0.0%	1.2%	100.0%
	4.6%	1.8%	0.0%	0.6%	1.5%	0.1%	0.0%	0.0%	0.8%
1900-1919	1	3	0	4	1	0	0	1	10
	10.0%	30.0%	0.0%	40.0%	10.0%	0.0%	0.0%	10.0%	100.0%
	0.1%	0.4%	0.0%	0.2%	0.2%	0.0%	0.0%	0.0%	0.1%
1920-1939	17	23	0	23	2	3	4	6	78
	21.8%	29.5%	0.0%	29.5%	2.6%	3.8%	5.1%	7.7%	100.0%
	1.7%	2.8%	0.0%	1.4%	0.3%	0.2%	0.4%	0.2%	0.8%
1960-1979	124	73	0	57	9	28	41	383	715
	17.3%	10.2%	0.0%	8.0%	1.3%	3.9%	5.7%	53.6%	100.0%
	12.3%	8.9%	0.0%	3.4%	1.5%	1.8%	3.7%	12.8%	7.1%
1980-1999	233	237	86	204	91	139	176	1047	2213
	10.5%	10.7%	3.9%	9.2%	4.1%	6.3%	8.0%	47.3%	100.0%
	23.2%	28.9%	29.1%	12.1%	15.2%	9.0%	15.7%	35.1%	22.0%
Post-2000	584	470	210	1394	486	1381	899	1544	6968
	8.4%	6.7%	3.0%	20.0%	7.0%	19.8%	12.9%	22.2%	100.0%
	58.1%	57.2%	70.9%	82.4%	81.3%	88.9%	80.3%	51.8%	69.2%
Total	1005	821	296	1692	598	1553	1120	2982	10067
	10.0%	8.2%	2.9%	16.8%	5.9%	15.4%	11.1%	29.6%	100.0%
	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Note: Chi-square: 1487.468, d.f.:35, p-value: < 0.001

VI. Comparison Of Variable Between Groups

The average complexity rating of the various game genres is compared. This section of the study aims to determine whether the average ranking for complexity varies considerably between genres. While the observations were independent, the findings of the Shaprio-Wilk test demonstrated that complexity is not regularly distributed in at least one genre (Table 6). ableDue to the assumption of normality, the Kruskal-Wallis Evaluate was utilized to test the hypothesis.

Table 6: Kolmogorov-Smirnov and Shapiro-Wilk Tests of Normality for Complexity Average and Domain

	Kolmogorov-Smirnov(a)			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Abstract	0.088	1005	<.001	0.968	1005	<.001
Children	0.272	821	<.001	0.739	821	<.001
Customizable	0.049	296	0.090	0.992	296	0.127
Family	0.060	1692	<.001	0.977	1692	<.001
Party	0.127	598	<.001	0.881	598	<.001
Strategy	0.054	1553	<.001	0.980	1553	<.001
Thematic	0.041	1120	<.001	0.987	1120	<.001
Wargames	0.044	2982	<.001	0.995	2982	<.001

a. Lilliefors Significance Correction

A Kruskal-Wallis test revealed a statistically significant difference in the genres' complexity (x^2(7) - 5841.97, p-value - 0.000) with a mean rank complexity rating of 7111.3 for Wargames, 7020.4 for Strategy, 5700.39 for Thematic, 5550.6 for Customizable, 3864.19 for Abstract, 2907.54 for Family, 1676.76 for Party, and 946.86 for Children.

Table 7: Kurskal-Wallis Test for Complexity Among Domain

Domain	**N**	**Mean Rank**
Abstract	1005	3864.19
Children	821	946.86
Customizable	296	5550.60
Family	1692	2907.54
Party	598	1606.76
Strategy	1553	7020.40
Thematic	1120	5700.39
Wargames	2982	7111.30
Total	10067	

Chi-Square: 5841.97, d.f.: 7, p-value 0.00

VII. Multiple Regression

Multiple regression analysis was used to examine the influence of the playing duration, the player's minimum age, the complexity rating, on the average rating of board games. The VIF values of the independent variables varied from 1.002 to 1.172, which is much less than the cutoff value of 10. Consequently, there was no multicollinearity problem. The results demonstrated that the year a game was released, the length of time it takes to play, the minimum age of its players, and its average rating for complexity could accurately predict its average rating. Consequently, the model as a whole was significant (F-stat=1047.304, p-value<0.05). In

addition, YearPublished, PlayTime, MinAGe, and ComplexityAverage explained 29.4% of the variation in the rating average (Adj-R2=0.294).

The results demonstrated that the publication year of a game had a positive and statistically significant influence on its average rating (B=0.01, p-value 0.001). In addition, there is a positive and statistically significant association between the minimum age and the average rating (B=0.012, p0.001), as well as between the level of difficulty and the average rating (B=0.525, p0.001). In addition, PlayTime has a negative effect on the average rating, although this effect is not statistically significant (B=-4.202E-6, p>0.05).

Table 8: Regression analysis

Independent Variables	**Model** **Dependent Variable: RatingAverage**
YearPublished	4.142***
	(9.964)
PlayTime	-0.0000042
	(-0.434)
MinAge	0.012***
	(5.618)
Complexity Average	0.525***
	(56.782)
F-statistics	1047.304
Adj-R^2	0.294
N	10067

**p<0.05; **p<0.01; ***p<0.001*

VIII. Logistic Regression

Using binary logistic regression analysis, the influence of the year a board game was produced, the amount of time it takes to play, the minimum age to play, and its average complexity on its popularity was explored. The findings of the analysis are reported in Table 9.

If the rating average was higher than or equal to 7.5, the dependent variable Popular was coded as 1; otherwise, it was recorded as 0. YearPublished, PlayTime, Minimum Age, and ComplexityRating are all numeric variables.

In order to obtain the effect estimates of the variables, the odds ratio was utilized. The results revealed that can accurately predict if a game would be popular (-2LR stat=6264,738; Chi-Square=2006354; p0.001). Additionally, YearPublished, MinPlayers, PlayTime, MinAge, and ComplexityRating account for 32,3 percent of the variance in Popular.

The results indicated that the odds ratio for a popular game is predicted to grow by 0.045 for each additional year that a game was produced, indicating that each additional year was strongly connected with an increased possibility of a popular game. There is a large and positive relationship between YearPublished and Popularity. (p-value<0.001)

In addition, for every player added to the minimum number of required players, the odds ratio for a popular game is anticipated to reduce by 0.573, indicating that raising the minimum number of required players was strongly connected with a decreased chance of a popular game. There is a negative and statistically significant relationship between MinPlayers and Popular (p<0.001).

Additionally, the odds ratio for a popular game is not projected to rise or decrease for every 1 minute increase in play duration, i.e. there is no increased or decreased possibility for a popular game. There is a significant relationship between PlayTime and Popular, although neither a positive nor a negative relationship exists (p-value = 0.029).

In addition, the odds ratio for a popular game is anticipated to decline by 0.03 for every year that the minimum age increases, indicating that increasing the minimum age was not substantially connected with a

lower chance of a popular game. There is a negative correlation between MinAge and Popular that is not statistically significant (p-value = 0.750).

In conclusion, for every point rise in a game's complexity ratio, the odds ratio for a popular game is projected to increase by 1.470, indicating that increasing the game's complexity was strongly connected with an increased chance of popularity. There is a positive and statistically significant relationship between ComplexityAverage and Popular ($p<0.001$).

Table 9: Logistic regression predicting the popularity of a board game

Independent Variables	**Odds Ratio**
YearPublished	0.045***
	(220.726)
MinPlayers	-0.573***
	(93.055)
PlayTime	0.00*
	(4.751)
MinAge	-0.003
	(0.102)
ComplexityAverage	1.470***
	(1024.37)
n	10067
-2LogLikelihood	6264.738
Likelihood Ratio Test (chi-square/df)	2006.354***/5
Nagalkerke approximation of R-Square	0.323

**p<0.05; **p<0.01; ***p<0.001. Effect estimates are presented as odds ratios. Wald-chi-square statistics are in parentheses*

IX. Summary

As stated in the beginning, board games are a great method to reduce stress and develop connections. With the help of BoardGameGeek.com, I was able to examine over 10,000 board games. Frequency analysis was employed to determine the distribution of the three categorical variables Domain, Period, and Popular, and descriptive statistics were compiled for the numerical variables MinPlayers, MaxPlayers, PlayTime, MinAge, RatingAverage, and ComplexityAverage. In addition to examining the linear relationship between Complexity, RatingAverage, and MinAge, a cross table was created to illustrate the dependence/independence of the variables Period and Domain. I also created a multiple regression model to see if RatingAverage was impacted by PlayTime, MinAge, and ComplexityRating, and a logistic regression model to determine the influence of Year, MinPlayers, PlayTime, MinAge, and ComplexityAverage on the binary variable Popular.