



Predicting Global Video-Game Sales

Alice Yufa, Jonathan L. Yu, Henry Chan, Paul D. Berger

Bentley University Waltham, MA 02452 USA

Bentley University Waltham, MA 02452 USA

Bentley University Waltham, MA 02452 USA

Bentley University Waltham, MA 02452 USA

Corresponding Author: Alice Yufa

ABSTRACT: In this paper we examine video games, and consider selected independent variables and explore their relationship to global sales. Key variables that are identified include the number of critics that rate the video game, and the average score that the critics give to the game. We also find a non-intuitive result concerning user ratings of the game.

KEY WORDS: Video games, Ratings of expert critics, User ratings, Stepwise regression

Received 02 April, 2019; Accepted 09 April, 2019 © the Author(S) 2019.

Published With Open Access At www.Questjournals.Org

I. INTRODUCTION

It is forecast that 2.3 billion gamers across the globe will spend \$137.9 billion on games in 2018 (Wijman, 2018). The video game industry has grown at a tremendous rate over the past few decades. It is hard to argue about the financial impact successful video games have on a video-game company.

Our goal in this paper is to examine the predicting of future video-game sales based on past sales, based on data from about 8,000 different games from a variety of different platforms. These data comes from Metacritic, a website that aggregates reviews of media products: films, TV shows, music albums and video games. It was important to use a large range of games from all of the different video-game categories to truly get a broad view of the market.

Ideally, this study can be used by game developers to get a broad understanding of the video- game market to see what types of games are preferred and which variables impact sales. These publishers and developers are in the business of making money and generating sales, so they want to see which variables are most important for sales of their video games. Ideally, the results of our analyses will help developers to decide in which platform, genre, or country to release their future video games.

Video games are no longer a niche product with a small segment of consumers. Research has shown that video game consumers are a very diverse group and in all walks of life. The industry has influenced virtually a large consumer segment, in that most of the current population has grown up with video games and played them for entertainment or education in some way.

II. LITERATURE REVIEW

The video-game industry has evolved over the decades. From what began with arcade games and console games has grown to a massive market to include PC and mobile. For the largest companies, it has become increasingly important to stay competitive and publish successful games. The quote below by Alpert (2007) partially sums up the growth of the industry:

“Finally, consider how well games companies are doing financially. Many of the early game companies have shrunk or disappeared (e.g., Infocom, Broderbund games), but the largest modern games company is gigantic by any standard. In 1994 Electronic Arts (EA) was a company with \$500 million in sales and a stock market capitalization of \$1.8 billion (all dollar figures in this paper are US\$). By 2004, it had become a \$3 billion in sales company with a market capitalization of \$15 billion. This places an exclusively games company as the fourth largest capitalized software maker in the world, behind Microsoft, Oracle and SAP Corporation (Lowenstein 2004).” (Alpert, 2007)

The consolidation of the industry made it more important that each game released is successful.

The industry has continued to change with changes in technology. Although there is data analysis, market research, and literature now, this was not always the case. There is a limited amount of literature in regards to marketing, strategy, etc., available, considering the monetary size of the industry (Alpert, 2007). There is still limited literature that addresses the topic from a business marketing perspective.

With the growth of data analytics, data science, and related areas, market research groups that specialize in video games came about. One such group, Newzoo, works with large customers such as Microsoft, Google, Pokemon, and Sports Interactive to help them develop a better understanding of the market. Companies are leveraging the data and expertise of market research companies to capitalize on capturing market growth overall and into new segments. "Furthermore, live streaming and e-sports will be fully embraced by the console segment by 2021, further driving player engagement. The console segment is forecast to grow to \$39.0 billion in 2021 with a CAGR of +4.1%" (Wijman, 2018). A large percentage of data available is held by these [paid] marketing-research companies.

III. METHODOLOGY

We collected our data from the data website, Kaggle; the dataset was titled Video Game Sales with Ratings, and was released in 2016 and used data from the review website, Metacritic. We cleaned up the data by removing certain variables that we thought were insignificant to our research such as the year the video game was published and the specific name of the developer who released the game. We then ran a stepwise regression analysis, using SPSS statistical software, to identify key factors that contribute to the final model.

We used global sales as the dependent variable and had many independent variables. These independent variables were platform, genre, North American sales, European sales, Japanese sales, other sales, critic score, number of critics scoring, user score, number of users scoring, and rating. The variable platform was measured by re-coding 17 qualitative answers into numeric terms. These qualitative values are GameBoy Advanced, Wii, WiiU, Playstation, Playstation 2, Playstation 3, Playstation 4, Playstation Vita, Personal Computer, Xbox, Xbox One, Xbox 360, Game Cube, Nintendo DS, Nintendo 3DS, Dreamcast and Sony PSP. The platform variable is used to see which video-game platform users were most profit-generating; this can be important for developers to know, as they may not want to release their games on all platforms. Next we also re-coded the qualitative values for the genre category which included adventure, role playing, action, sports, simulation, puzzle, shooter, racing, strategy, playform, fighting, and miscellaneous. The genre category is another potentially important category for future game releases to decide which type of video game users want to play. The next four independent variables were sales in millions of units. These sales numbers were divided into four categories which were North American sales, European sales, Japanese sales and other sales (the rest of the world, excluding North America, Europe and Japan). The fifth sales variable was global sales but this is our dependent/output variable. These four sales independent variables may be helpful for publishing companies to decide which country or region they want to primarily put their marketing dollars in for release of future games. They may focus on all of these regions, but can decide if they wish to invest more money into one over the others.

The next two variables were critic score and critic count. The critic score was on a rating scale of 0-100 based off the opinions of expert video-game critics, and critic count is number of critics from the Metacritic website that provided a review, with some games having more critic reviews than others. We expected critic score to be an important variable, because many users decide on video games based on how good the reviews are, and the number of reviews indicates an inherent interest in the game. There are thousands, perhaps, tens of thousands, different games to choose from, so having good ratings, particularly from experts, would be expected to place your game above others in the mind of the consumers. Also, stores and websites may decide to display the game in a superior way (signage promotion, shelf-space, shelf-height, etc.), because of the higher ratings.

The next two variables are user score and user count. Similar to critic score, the user score is a 0-100 rating scale generated from users who have played the video games on their own; it is presumed that these scorers are not professional critics. The user count is also similar to critic count, being the number of users who gave their review of the game. It is presumed, *ceteris paribus*, that the more reviews the better. Similar to critic reviews, user reviews are possibly important for video games because they come directly from the people playing the game. If the users are not satisfied, they will not continue playing the game and not tell others to play the game. Presumably, the users need to be happy with the product for it to succeed in such a competitive market.

The final variable is rating. This was also re-coded from a qualitative rating to numeric variable and based on the ESRB (Entertainment Software Rating Board) guidelines. The ratings included E for Everyone, E10+ for Everyone-over-10, M for Mature, T for Teen and RP for Rating Pending. The rating is another decision developers should look at because it may be a major decision on how they market their games. Although it would be assumed that E for Everyone would be most inclusive of all users, so therefore would

bring in most sales, this may not necessarily be true. E for Everyone requires strict guidelines so that people of all ages can play - meaning no violence, nudity, swearing, etc, which some people may want in a video game. However, some users may not want to play a game with no graphic content, a game that is especially family friendly. This is the same concept of considering that adults may very quickly dismiss a movie with a rating of “G” or even “PG,” favoring a movie with a rating of at least “PG13.” Developers can receive only one ESRB rating and make a decision when creating the game how much graphic content to add which will likely affect their rating. They would want to know which rating scores generate the most profit, which is their ultimate end goal.

Before running any data through SPSS we decided upon a few different hypotheses concerning which variables would be very significant in generating higher sales. We considered that the critic and user reviews would be very significant because many video game players find new games by looking at how successful a game is in terms of these reviews. High ratings are what get games to be displayed in better store locations or higher on web pages. Many games come and go throughout the years, but those that are successful with users and critics are those that come back for new, “improved,” versions and continuous sales in the future. We also believed that North American sales would be an important variable as the U.S. is second highest top gaming country by revenue (after China) with \$25.43 billion in 2017 revenue, along with Canada at number 8 with \$1.97 billion and Mexico at number 12 with \$1.43 billion, according to Statista (Statista, 2017). These three countries together, which make up North America, are a huge part of total sales revenue for video games and one of the world’s largest markets for developers to focus on. We believed that if sales in North America were strong, it would reflect on global sales as well. We were not certain about the effect of genre. This is, in part, because genre is not well defined. A video game may have aspects of many different genres; for example, a game classified as action could also have shooting, adventure, puzzles and role playing. Video games have so many moving parts that it is sometimes difficult to classify them into one specific genre. We went into the analysis believing that users buy games based on their overall aspects and features, and not the genre classification.

Analysis And Discussion Of Results

Since the variables, NA_Sales, EU_Sales, JP_Sales and Other_Sales are already variables that represent sales, and are actually part of the dependent variable, we decided, in retrospect, to eliminate them from the analysis. So we ran a stepwise regression without any of the Sales Variables, and the output is as shown in Table 1.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-.076	.040		-1.897	.058
	Critic_Count	.029	.001	.289	25.313	.000
2	(Constant)	-.203	.056		-3.609	.000
	Critic_Count	.029	.001	.290	25.393	.000
	Genre	.022	.007	.037	3.214	.001
3	(Constant)	-1.470	.120		-12.207	.000
	Critic_Count	.024	.001	.233	18.927	.000
	Genre	.022	.007	.037	3.271	.001
	Critic_Score	.020	.002	.146	11.866	.000
4	(Constant)	-1.225	.133		-9.243	.000
	Critic_Count	.023	.001	.230	18.730	.000
	Genre	.021	.007	.035	3.140	.002
	Critic_Score	.025	.002	.182	12.301	.000
	User_Score	-.082	.019	-.061	-4.385	.000

a. Dependent Variable: Global_Sales

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.289 ^a	.084	.084	1.85748
2	.335 ^b	.112	.112	1.82821
3	.353 ^c	.125	.124	1.81558
4	.356 ^d	.127	.126	1.81357

- a. Predictors: (Constant), Critic_Count
- b. Predictors: (Constant), Critic_Count, User_Count
- c. Predictors: (Constant), Critic_Count, User_Count, Critic_Score
- d. Predictors: (Constant), Critic_Count, User_Count, Critic_Score, Genre

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2210.722	1	2210.722	640.748	.000 ^b
	Residual	24203.317	7015	3.450		
	Total	26414.039	7016			
2	Regression	2970.845	2	1485.422	444.425	.000 ^c
	Residual	23443.194	7014	3.342		
	Total	26414.039	7016			
3	Regression	3296.791	3	1098.930	333.379	.000 ^d
	Residual	23117.248	7013	3.296		
	Total	26414.039	7016			
4	Regression	3351.324	4	837.831	254.735	.000 ^e
	Residual	23062.715	7012	3.289		
	Total	26414.039	7016			

- a. Dependent Variable: Global_Sales
- b. Predictors: (Constant), Critic_Count
- c. Predictors: (Constant), Critic_Count, User_Count
- d. Predictors: (Constant), Critic_Count, User_Count, Critic_Score
- e. Predictors: (Constant), Critic_Count, User_Count, Critic_Score, Genre

Table 1: Stepwise regression results

We can see from the first section of Table 1 that the final model has four variables, Critic_Count, Genre, Critic_Score, and User_Score. The model is:

$$Y\text{-predicted} = \text{Global sales predicted} = -1.225 + .023 (\text{Critic_Count}) + .021(\text{Genre}) + .025(\text{Critic_score}) - .082(\text{User_Score})$$

It makes perfect sense that the coefficients of Critic_Count and Critic_Score are positive. We are not certain how to explain the negative coefficient of User-Score. As we suggested earlier, we would have predicted a positive coefficient. We cannot “blame” multi-colinearity, as the purpose of running a stepwise regression was to not have that be an issue. So, while it is suggested that, holding other variables in the model constant, global sales are higher when Critic_Count and Critic_Score are higher, it is suggested that global sales are lower when User_Score is higher. Perhaps potential game purchasers are somewhat skeptical of the ratings of other users, feeling that they are “different” (perhaps, superior??) and that the experts “know better.”

We had intended to code genre as a set of dummy variables of 10 variables for 11 categories, but clearly, something went wrong with the coding process, and genre, as a variable, was mistakenly coded as a single variable from 1 through 11; with no natural ordering to the different genres, we ascribe no useful meaning to that variable. However, this should not affect the results for other variables, for no other variables in our final model correlate highly with genre – as defined in the model - (or else, they would not both remain in the final model), and we do not envision the existence of variables not in the final model that are being “barred” from entering the final model due to a high correlation with the badly-formed genre variable.

The R^2 value for the final model was a modest .127 (adjusted $R^2 = .126$), indicating that our final model is estimated to explain about 12.7% of the variability in Global Sales. While this is not overly large in absolute terms, the model is highly significant. The results might indicate that there is a large amount of “personal taste” involved in the choice of video games that perhaps cannot be captured by the variables we included as eligible in our stepwise regression.

IV. LIMITATIONS

We had some limitations in our study. One limitation is that some of our data were missing. We removed the video games without a rating score as it would be hard to properly compare all the data with these data missing. We still had over 7,000 data points. There were a few variables not included that we thought might have been important to consider such as gender, as many males and females play different types of video games; it would be interesting to see if female-targeted games have different global sales numbers than male-targeted games, and to what extent.

It is clear that we could have done a better job in terms of handling the dummy variable coding, especially for genre, and we need to re-examine how the coding of platform took place in the analysis process. However, we still believe that what the model revealed was accurate, even allowing for the faulty recoding of genre. We cannot say what the effects of genre and platform are, and this may have contributed to the relatively low, albeit highly significant, value of R^2 .

REFERENCES

- [1]. Alpert, Frank (2007), “The Entertainment Software Industry: Suddenly Huge, Little Understood,” *Asia-Pacific Journal of Marketing and Logistics*, Vol. 19, No. 1, 87-100. Retrieved from <https://core.ac.uk/download/pdf/15035023.pdf>.
- [2]. “Top Gaming Countries by Revenue 2017 | Statista.” Statista, Dec. 2017, www.statista.com/statistics/308454/gaming-revenue-countries/.
- [3]. Wijman, Tom (2018), “Mobile Revenues Account for More Than 50% of the Global Games Market as It Reaches \$137.9 Billion in 2018,” *Newzoo*. Retrieved from <https://newzoo.com/insights/articles/global-games-market-reaches-137-9-billion-in-2018-mobile-games-take-half/>

Alice Yufa" Predicting Global Video-Game Sales" *Quest Journals Journal of Research in Business and Management*, vol. 07, no. 03, 2019, pp 60-64