1

Donye Wakefield

**Video Games and their ratings**

Video games are arguably the most popular form of entertainment today, as a result, video games are a big source of entertainment for consumers and a huge lucrative source for businesses.  According to a report on Statista, as of 2020, the gaming industry ranks as the very top with global revenues reaching up to $145.7b compared to $42.5b for the box office (movies), and $20.2b for the music industry. (Richter, 2020) The statistical fact that the gaming industry is raking in so much revenue should showcase how huge video games are as a whole industry. There are millions of fans across the world that are always curious on whether or not a video game will do either well or poorly upon release, as this affects their gaming experience, as well as their wallets. Due to this, the idea for a project related to how gamers should interpret what may make a game successful was born. There may arguably be many metrics for what may make a game be interpreted as successful, however, for the purposes of this project, sales would be the success metric.

As a gamer attempt to determine if a game is good or not, they are likely to be met with ratings on the game of interest. There are generally going to be two ratings: critic ratings and user ratings. The general understanding seems to be that there is a distinction between how both critics and gamers rate their games. This is likely the case for gamers themselves, as the idea of rating anything is completely up to the person giving the rating. As a result, everyone will have separate opinions regarding how good or bad a game is or will be. A writer on IGN explains this

2

idea briefly by basically pointing out the idea of games being an art form, and like any other art form it is basically impossible to truly quantify its value through a scoring system. (IGN) Critics are present on platform such as IGN, so by them explaining the thought process behind why they may rate games a way that is not agreeable whether by users or fellow critics, it shows that even they know that the idea of having a score does not mean everything for a game.

The idea of this project, is to keep this notion of subjectivity in mind, but to also attempt in quantifying the differences between how users rate games and how critics rate games. By conducting this analysis, gamers can see whether a critics scoring is more objectively accurate regarding the success of a game, or if user scoring is more objectively accurate. This may help gamers in general carefully evaluate if a game is worth buying or not based on what score is quantifiably more accurate than the other. That being said, companies may have a different perspective on this as well. The purposes this project is to help make sure the gamer is informed on what game may be better based on ratings, but companies may also benefit from the analysis of this project as well. What makes this seem reasonable is due to what writer Alec Meer summarized in his article. There was a EEDAR (Electronic Entertainment Design and Research) survey experiment conducted, which evaluated gamers opinions pertaining to a game *Plants vs. Zombies*. Basically, the conclusion was that gamers opinion of the game was much higher if the game was already known and seemed well respected through its ratings, compared to their peers who weren’t told on how good the game was. (Meer, 2010) This idea could certainly be useful to companies as well in order for them to evaluate if a particular rating is having a certain relationship with their sales. So overall, an analysis of looking at critic score and user score and

3

seeing which one is more objectively accurate can be very useful for both gamers and companies.

**Description of the Data Sources**

The initial decision was made to choose 2 data sources for this project, one that is relatively small in terms of data collected, and another that is far more massive and expansive in terms of collected data. The larger data chosen consists of thousands of records of data collected (55,792 observations and 23 variables). This one was chosen to do an analysis based on global sales as the success metric. Being that this data set has a decent amount of variables, it would be best to use this one for that part of the research question. The only con of this data set is that it has a lot of missing values, especially for particular sections like sales for all the countries listed as variables. The con with this other data set is that it is smaller, so there are not as many variables to account for (16 variables), unlike the bigger data set. In addition, the observation number (or sample size) itself is not as large, however, for the purposes of this project it is certainly large enough, as this data set has 16,598 observations.

In the linked Jupyter notebook, the smaller data frame will be referred to as ‘vgs\_small’, and the bigger data frame as ‘vgs\_big’. For vgs\_small, the data was put together by the user named Rush Kirubi, and he seemingly just extended another data frame, which was scraped by another researcher named Gregory smith. Smith scraped data off of VGChartz, which is a video game sales tracking website, and Kirubi scraped data off of Metacritic, which is a website that accumulates reviewer scores of all types of entertainment, which includes video games. Each game has a pool of reviews done for them, and is then averaged for an overall rating for the

4

game. Kirubi makes mention that this data frame does indeed have missing values, and his reasoning for this was that Metacritic seems to only cover a subset of gaming platforms. That being said, for the purposes of this analysis, there is not much of a concern with this, as both data sets provide a good enough counter balance to this problem.

In terms of the vgs\_big, there does not seem as thorough of an explanation given by the user Abdulshaheed Alqunber pertaining to this data frame compared to Kirubi’s data frame. In terms of where the data comes from, it would seem that this data is scraped from the platform VGChartz, which is of course the same primary platform as the other data set. However, there is also the addition of reviewer scores, similar to that of the other data set, so it would seem that these scores likely also came from the same platform, which was Metacritic. It is important to note that this data frame is also a continuation of Gregory Smith’s data frame. However, when looking at Smith’s original data frame, there is still little of an explanation of the process of gathering this data. What is mentioned is the tool used to scrape the data off of VGChartz, a web scraper based off of python’s “beautiful soup” apparently, as well as the link to the GitHub for accessing this scraper. One last thing worth noting is the context of the games scraped off of VGChartz; Considering that Smith’s data frame is the original, this applies to both Alqunber’s and Kirubi’s data sets. Smith particularly chose games that had sales greater than 100,000 copies. This means that the games in the chosen data sets are already relatively popular and relatively successful compared to games not in this data frame. This is even better for the purposes of this project, considering that the purpose is to see what variable (critic score or user score) is more reliable in interpreting relative success of a game, and it may help if games are already structured in a more competitive data matrix. This competition aspect can help provide an analysis that is

5

far stronger than that of a data frame that has games with lower than 100,000 copies, considering that those games would clearly be less popular and successful.

**Analysis**

In terms of an analysis, there is much speculation on what seems most feasible with the current data given by these data sources. Ideas thought of for conducting this analysis would be the use of splitting up the data. If data is split, this would allow the analysis to go even deeper due to the data being more zoomed in on, rather than just having a big picture view of the data as a whole. Perhaps, the data can be split into two groups and named in accordance to their purposes. This split may be done in terms of the success metrics: ranking and global sales. So it would make sense to split them in terms of how better ranked they are and how many more sales games have. The main general data frames would probably still be useful for the analysis just because it may provide a good big picture look at the data as a whole just in case that would be needed.

In addition, there would likely be the need to evaluate relationships amongst variables. An example would probably have to be the relationship amongst critic score and global sales. This relationship test would likely be done for all relevant variables, and in order to best do that, it would be essential to use scatter plots. Another tool that may make most sense for evaluating relationships would probably be a histogram. This may give a good visual representation of each variable that would be worked with, and these representations could be compared with another in order to evaluate the difference between measures of central tendency and variability amongst the variables. In addition, this may also expose certain flaws with the data being worked with.

6

An example would be the lack of proper visualization comparison due to variables lacking equivalent number of observations compared to one another. Lastly it may expose the site of outliers, which if that was the case, it can be accompanied by a box plot to get an even better visualization of them.

Several libraries would likely be utilized in the analysis, such as *Numpy*, *Pandas*, *Matplotlib*, and *Seaborn*. There may also be a possibility of using Scikit-Learn and Stats Models. In terms of why these would all be used, it would seem reasonable to use Numpy for possible cleaning of the data, Pandas for cleaning and transformation of the data, and Matplotlib with Seaborn for actually visualizing the data. So it’s expected that all of these libraries will help give the analysis all the functionality needed to run to the fullest extent. However, if they do not, there may be the need to do some additional research to find libraries that do indeed assist these ideal methods, or simply just ignore a particular library all together.

**Critic Score and User Score Analysis**

***Data cleaning and management***

For conducting this study, all relevant libraries that pertained to the analysis that revolved around visualizing, transforming, and cleaning the data sets were imported. In addition, the directory for where my sources are located was also coded. Once the data is imported, the data must be cleaned properly before there can be any real analysis on it. First, a quick scan of the data was done to see if it was all tidy. All columns had a variable associated with it and all observations had its own unique row with no shared values with any other observation, unless it was unique to its own of course. The only issue pertaining to observations and columns was in

7

regards to a string value inside the ‘User\_Score’ variable column. This issue managed to be corrected by simply converting that variables contents all into numeric values. After that, the decision was made to eliminate all NA (Non-Available) values from the data frame. The reason for doing so was because of how robust the data would still be despite the removal of NAs, and even more importantly because the plan for the analysis, which revolves around splitting up the data would heavily rely on all observations initially being equal in terms of their amount.

Once at this point, there had to be some thought to decide on how exactly the exploration of the research question regarding the data sources given was to be done. As discussed before, a success metric was needed in order to properly evaluate the differences between critic scores and user scores. This success metric was global sales and it was to be utilized by the larger data set, which was going to be imported as vgs\_big. However, for the previously mentioned smaller data frame (vgs\_small), the decision was made to drop the whole data set altogether. This was because of the realization that having two different data sets pertaining to separate success metrics would seem quite redundant for the purposes of this particular analysis. This realization occurred literally during the actual analysis, so this was completely unexpected of a change, but it has occurred anyways, and it is important to note that. Regardless, the idea was now to focus on global sales as the success metric solely, which meant that vgs\_big was the only data set that was to be worked with for the entire analysis. The next step of this tidying up process was to narrow down on only variables that were most relevant. Most of the variables that were present in the data frame imported were of no use in terms of conducting the analysis on user and critic scoring. As a result, the decision was made to create a brand new data frame (vgs) that would

8

only have the variables that would be most relevant to the analysis. Those variables would be ‘Name’, ‘Global\_Sales’, Critic\_Score’, and ‘User\_Score’.

The next step was to zoom in on global sales to discover the average global sales as defined by the general data frame itself (vgs), which was done by looking at the summary stats of the whole data frame through the use of the describe method. This average was 0.777590, and this number was to be utilized in order to split the data. The data was split into 2 separate groups: one group being games above this average (above\_ave), and another group being games below that average (below\_ave). So essentially there would be two separate data frames derived from the main data frame. It is here where the actual analysis on both data frames can officially begin.

***Is there a significant difference on how users and critics rate games?***

The first question to be evaluated pertained to how distinct the distributions of both critic and user scores was for both the above average games and for the below average games. This was done by creating histograms to showcase the distribution of a single variable pertaining to games above average, and comparing that with the distribution of the same single variable but pertaining to games below average. This process was done first for critic score and second for user score.

***Is there a strong relationship between critic scoring and user scoring?***

For this question, the idea was to visualize and evaluate the direct relationship between critic scoring and user scoring. This is done so that it is known if one variable influences the other or if both variables are fairly independent of each other. The best way to observe this relationship is by utilizing a scatterplot. Critic score was plotted as the independent variable (x)

9

and then user score as the dependent variable (y). In addition, actual correlation numbers were shown right underneath the graph, to give a more exact estimate as well. By doing this, the correlation can be both given through visualization and as an actual number. Out of curiosity, as well as testing for accuracy, the decision was made to do this same method for when the roles of the variables are switched: as in having user score become the independent variable and critic score become the dependent variable.

***Which variable (critic score or user score) has a stronger relationship with global sales?***

Focusing back on the success metric for this analysis, the point was to be able to see if either critic score or user score has a stronger correlation with global sales. The scatter plot visualization was done to evaluated this relationship for the whole data frame (vgs), then the same idea was conducted for both sub-groups as well: visualizations were done to evaluate the relationship between the two different scores with global sales for games above average and for games below average.

**Post-Analysis Findings and Results**

***Is there a significant difference on how users and critics rate games? - Result***

When conducting the analysis on the question on whether or not there is a significant difference between how users and critics rate games, it would seem that there is indeed a significant difference. The histograms that overlap one another for both graphs showcase this difference very well. For critic score, there seems to be a glaring disconnect between the critic score distribution of games above average and for the games below average. As for user score, there seems to be a lot more overlap amongst the user score distribution of games above average

10

and games below average. This seems to suggest that critics rate games more accurately in terms of how high a games sales are compared to users, who by the looks of their heavily overlapping distribution, seems to rate games very much the same despite a game being above average or below average. As a result, it would seem that there is indeed a significant difference between how users would rate a game compared to how a critic rates a game. This makes sense considering that a critic is more likely going to look at a game in a different manner then that of a typical gamer. In general, critics are meant to be critical of whatever medium they are reviewing, which in this case is games, so in theory, they must rate games more objectively above everything else.

***Is there a strong relationship between critic scoring and user scoring? – Result***

By interpreting both scatter plots, it would seem that there is indeed a fairly strong correlation between user score and critic score. When critic score is independent and user score is dependent, it would seem that user score can be fairly predicted based on critic score, which could be represented even further by drawing an imaginary least squares line within the cloud of data, and this is also the case vice versa. As user score increases so does critic score, and when critic score increases so does user score. Additionally, this may showcase that critic score and user score are not that independent of each other, and that both are quite influenced by the other with an approximate correlation of 58%. To rehash, the point of this test was to see how influential either variable was towards the other. In theory, this would show if both variables are independent, and if that was the case, then it would showcase if critic scoring is intrinsically distinct from user scoring.

11

***Which variable (critic score or user score) has a stronger relationship with global sales? - Result***

For this part of the analysis, it would seem that all correlations represented by the scatter plots, and actual correlation numbers, are pretty weak. Correlation generally is much stronger the closer it is to the number 1, and it would seem that all correlations represented here are below even 50% (half of a literal direct correlation of 100%). However, even though all correlations are weak, there is an argument to be made about the relative correlation strength of critic score and global sales compared to that of user score and global sales. When looking at the general correlation pertaining to the original data frame (vgs), there is a significant difference between correlation strength of critic score and global sales versus that of user score and global sales. Critic score has an approximate correlation of 27.8% with global sales, while user score has only a 9.8% correlation with global sales. This shows that critic score has a significantly stronger relationship with global sales compared to user score. When zooming in even closer, which is in regard to the two sub-data frames (above\_ave and below\_ave) this argument still holds true. For games above average, critic score has a stronger correlation of approximately 17.7% compared to user score correlation of approximately 3.6%. For games below average, critic score once again has a stronger correlation of approximately 19.5% compared to user score correlation of approximately 9.8%. Overall, this test showcases that if there was to be a model done for the prediction of better games in terms of how successful they are (with global sales as the success metric, or response variable in this context), the decision would logically be to choose critic score as the predictor, or explanatory variable, over user score, due to how stronger of a correlation critic score has with global sales than that of user scores. Being that the correlation is

12

still fairly weak between critic score and global sales, there would likely have to be another predictor on top of critic score to have an even better model, but that’s a topic for another time. For the purposes of this analysis, it would seem that critic score is technically better.

**Conclusion and Final Thoughts**

In conclusion, it would seem that critic score is the more accurate variable in determining the success of a game compared to that of the user score. As a reminder, this analysis was done for the purposes of determining what type of rating would be more beneficial in terms of accurately rating a game for people who buy games. A gamer that looks at ratings should be informed on what rating would be more accurate in defining the game so that he/she makes the best decision on their purchase. Without this knowledge, gamers would either one: simply debate on which rating is more accurate without any statistical backing, or two: make relatively uninformed decisions on the games that they are thinking of buying. According to this whole analysis, it would seem that gamers may want to rely more on what critics say than what users say.

In terms of the conclusion, it would seem that gamers are probably not the best when it comes to rating the games they play. Based on the data, gamers seem to be less judgmental about the games they play likely due to them not being focused on reviewing them, and instead being more focused on just playing them for fun. Critics on the other hand are going to be generally more critical of a game due to them having their main focus on reviewing that game rather than simply just playing it for fun. So the roles of a casual gamer and a game critic are clearly different, and this difference seems to be displayed by the analysis.

13

**Work Cited**

**Academic Source #1**

Richter, Felix. “Infographic: Gaming: The Most Lucrative Entertainment Industry By Far.” *Statista Infographics*, 22 Sept. 2020. <https://www.statista.com/chart/22392/global-revenue-of-selected-entertainment-industry-sectors/>

**Academic Source #2**

“Game Review Scoring.” *IGN Entertainment*, corp.ign.com/review-practices.

**Academic Source #3**

2010, Alec Meer. Wednesday 7th July. “EEDAR Study: Review Scores Do Affect Sales.” *GamesIndustry.biz*, [www.gamesindustry.biz/articles/eedar-review-scores-do-affect-sales#:~:text=Gamers%20think%20more%20positively%20towards,to%20a%20new%20EEDAR%20survey.&text=However%2C%20EEDAR%20was%20convinced%20its,marketing%20to%20a%20greater%20extent](http://www.gamesindustry.biz/articles/eedar-review-scores-do-affect-sales#:~:text=Gamers%20think%20more%20positively%20towards,to%20a%20new%20EEDAR%20survey.&text=However%2C%20EEDAR%20was%20convinced%20its,marketing%20to%20a%20greater%20extent).

**Data Source #1**

<https://www.kaggle.com/ashaheedq/video-games-sales-2019>

**Data Source #2**

<https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings>