GameStop Corporation’s Big Data Venture

Lawrence Wilson

MIS581 – Capstone - Business Intelligence and Data Analytics

Colorado State University – Global Campus

Dr. Orenthio Goodwin

October 6, 2019

**Table of Contents**

[GameStop Corporation’s Big Data Venture 2](#_Toc21289937)

[**Abstract** 5](#_Toc21289938)

[**VGsales Dataset** 6](#_Toc21289939)

[**Preliminary Data Visualizations** 8](#_Toc21289940)

[**Data Preprocessing** 11](#_Toc21289941)

[**Literature Review** 13](#_Toc21289942)

[**Developing SMART Business Questions and Research Hypotheses** 14](#_Toc21289943)

[**Ethical Big Data Project Concerns** 15](#_Toc21289944)

[**Research Methodology** 16](#_Toc21289945)

[**Limitations** 21](#_Toc21289946)

[**Descriptive Analytics Findings** 21](#_Toc21289947)

[**Predictive Analytics Findings** 33](#_Toc21289948)

[**Conclusion** 43](#_Toc21289949)

[**Recommendations** 44](#_Toc21289950)

[**Tools and Project Links** 45](#_Toc21289951)

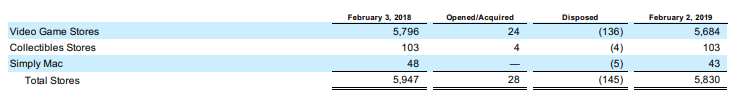
[References 46](#_Toc21289952)

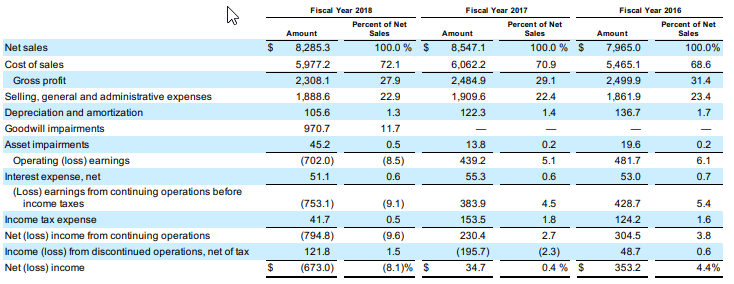
# GameStop Corporation’s Big Data Venture

Once an unrivaled and dominant video game retail giant, GameStop released a concerning 10-K statement in April 2019 that alarmed investors.

Table 1.

*GameStop Store Count and Consolidated Results of Operations.*





*Note*. This table lists GameStop’s store count and the consolidated operational results. Adapted from *Form 10-K 2019* by GameStop Corporation, 2019.

As depicted in the above table, GameStop Corporation (2019) posted a net loss revenue of $673 million for the fiscal year of 2018 and disposed of 145 stores to reduce its total store count from 5,947 in February of 2018 to 5,830 stores in February of 2019. In fact, GameStop Corporation (2019) is in the midst of a multiyear slump in which the company’s stock dropped 36% and lost an incredible two-thirds of its values in just two years. Although the company has historically paid out to dividends and has been known as relatively stable income stock, GameStop Corporation (2019) made a recent announcement that they would no longer pay out dividends due to financial limitations. However, despite these reductions, GameStop Corporation (2019) remains a brick-and-mortar retail giant with approximately 16,000 full-time salaried and hourly employees around the world, 23,000 to 45,000 part-time hourly employees globally, and a diverse range of stores across 14 countries. In fact, GameStop Corporation (2019) operates a loyalty rewards program known as PowerUp Rewards with a total member count of approximately 39.6 million members in the United States and 20.9 million members in other countries as of February 2019. However, of the nearly 40 million loyalty members in the United States with GameStop only actively marketing to 20.5 million members, only 16.2 million members have purchased at GameStop in the past year (GameStop Corporation, 2019).

In continuation, one may question how a company with a potentially unrivaled wealth of data on video game customers cannot gain a competitive advantage over those without nearly as much data. In the era of Big Data, technology has made it possible to collect large amounts of data. However, it is not the data itself that provides value but the act of discovering knowledge and generating information from such data that is valuable (Han, Kamber, & Pei, 2011). For example, the Royal Bank of Scotland was once the largest bank in the world but after its exposure to the subprime mortgage market during the 2008 housing crisis, the bank nearly collapsed and required the United Kingdom government to hold 84% of its shares to avert a total business failure (Marr, 2016). When it comes to banks, they inherently have more personal data on their customers just by virtue of the business nature of bank branches and the intimate knowledge gained from personal interaction and reviewing a customer’s financial transactions; they could track when customers got married, felt unwell, took vacations, and how they chose to spend their excess income just by reviewing such transactions (Marr, 2016). In comparison, companies like Amazon have very little data on customers when compared to banks; Amazon simply used the data that they have much better than banks such as the Royal Bank of Scotland (Marr, 2016). Similarly, just by virtue of the interpersonal business nature of its retail store locations and its treasure-trove of video game consumer data generated by the PowerUp Rewards program, GameStop has far more personal data on its customers than e-commerce rival Amazon. Some investors argue that brick-and-mortar businesses such as GameStop are akin to a record store operating in the age of iTunes and Spotify (Goldman, 2019). Yet, rather than rolling over in the face of e-commerce giant Amazon, a 12-month review in March of 2019 found that Wal Mart and Target shares were up 14.16% and 12.73% respectively while Amazon’s shares were only up 6.03% (Mourdoukoutas, 2019). Both Target and Wal Mart have strengthened their online presence and are now utilizing the vast treasure of customer data that they have in their disposal; brick-and-mortar retailers may not be as flexible as e-commerce retailers but they have more intimate knowledge of their customers that can be used to gain an unique competitive advantage (Mourdoukoutas, 2019). By combining their hidden gem of video game consumer data from their PowerUp Rewards program with external public video game datasets such as the vgsales-12-4-2019.csv, GameStop may be able to gain the valuable insights necessary in answering their SMART financial business questions and turning their business back towards the path of profitability.

# **Abstract**

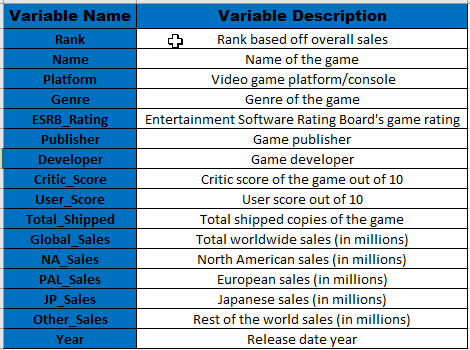
This research study analyzes a publicly available dataset of video game attributes to discover if any of the variables could predict the likelihood of a high shipment or high sales event; the purpose of the research study is to show the value of publicly available datasets and how it can complement the internal data of the brick-and-mortar retailer GameStop to develop a highly competitive recommendation engine for its customers. Since the dataset consists of a large amount of continuous and categorical variables, both correlation and binary logistic regression statistical techniques are utilized to predict the dependent high total shipment and high global and regional sales variables. The results demonstrate that the non-people factor such as the type of publisher, genre, and ESRB rating play a larger role in predicting high shipment and sales outcomes than the people-factor such as critic and user scores.

# **VGsales Dataset**

Furthermore, the vgsales-12-4-2019.csv dataset contains 16 variables, 55,792 records, and a data date good through 4-12-2019; the table below provides the original list of variables and their respective general descriptions.

Table 2

*VGsales Dataset Variables*

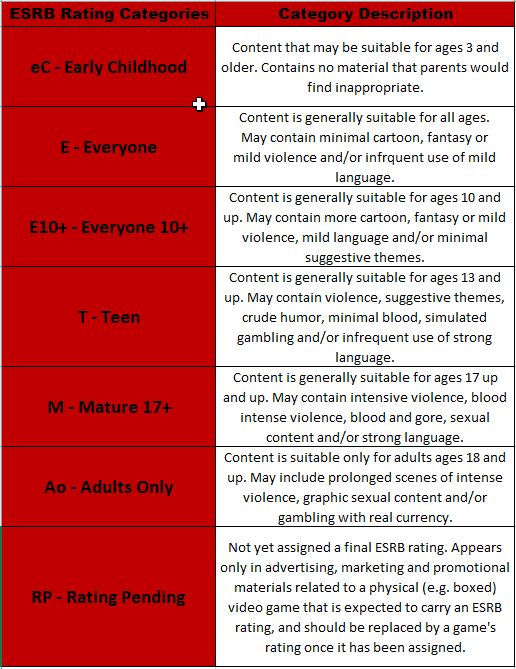


*Note*. This is a list and description of the variables found in the VGsales dataset. Adapted from Year on year sales & market share charts by VGChartz, 2019.

Also, the seven distinct ESRB rating categories along with their respective descriptions are shown in the table below.

Table 3

*ESRB Rating Categories and Descriptions*

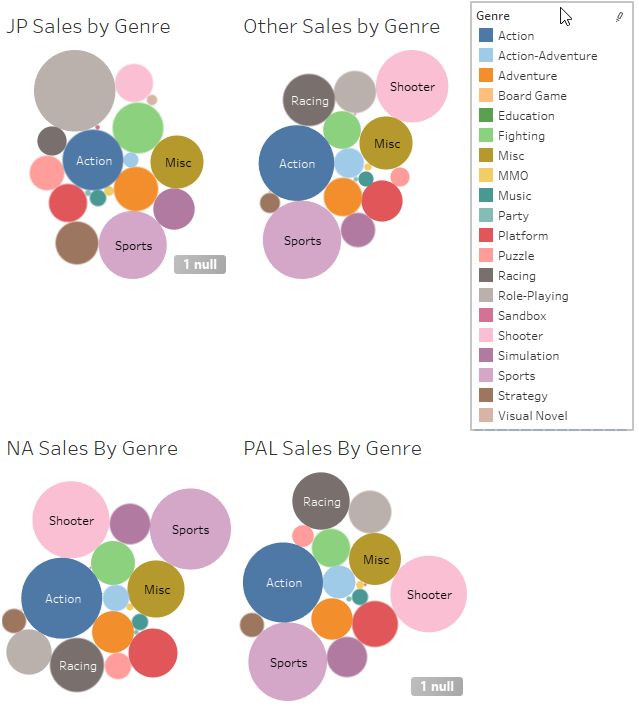


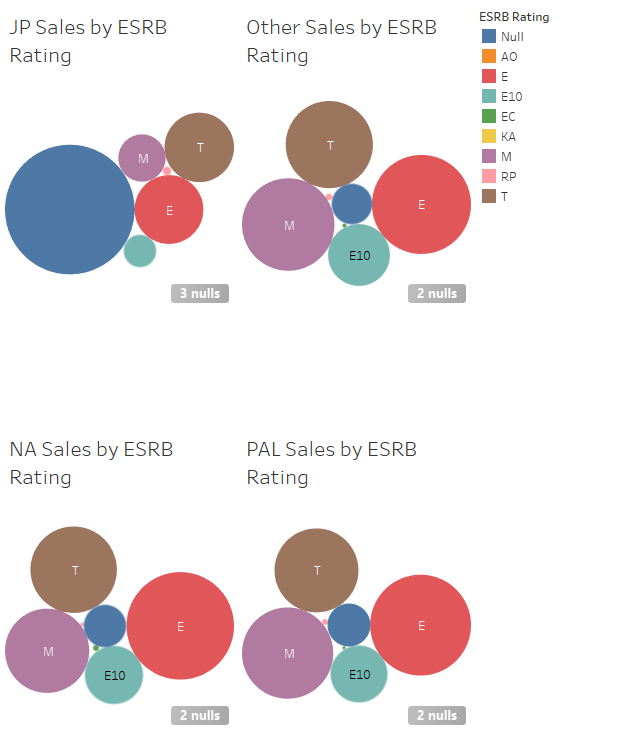
*Note*. This is a list and description of the ESRB rating categories. Adapted from Ratings guide by Entertainment Software Rating Board, 2019.

For the project, the Rank, Name, Platform, Genre, ESRB Rating, Publisher, and Developer were treated as qualitative variables; the Critic Score, User Score, Total Shipped, Global Sales, NA Sales, PAL Sales, JP Sales, Other Sales, and Year were treated as quantitative variables. In order to include the variables in correlation and binary logistic regression tests, descriptive analysis was utilized to review both what was happening in the dataset and to preprocess the data.

## **Preliminary Data Visualizations**

In order to gain a preliminary visual look into the VGsales dataset before any detailed exploration and preprocessing, Tableau was used to create packed-bubble data visualizations by genre and ESRB rating; this was important in identifying potential issues in the data distribution of the key discrete dimensions that would need to be addressed in data preprocessing.





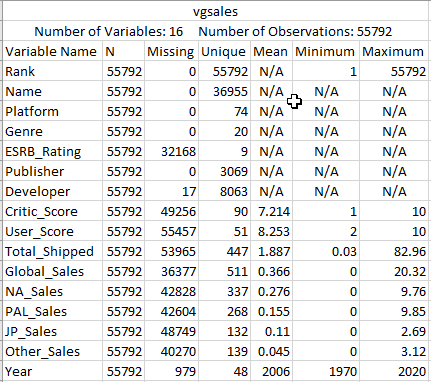
When reviewing the preliminary Tableau data visualizations of regional sales by genre, it could be seen that the sports, action, and shooter genres were well represented in most of the regions except Japan. Instead, the role-playing genre stood out with the highest sales in the Japanese market. Therefore, it was concluded that there was merit in conducting analysis by region as well as globally; different regions may have different preferences. In addition, this appeared to also be true when reviewing the Japanese market sales by ESRB rating. Although M, T, and E rated games dominated sales in the other regions, Japan stood out as a lone outlier with non-rated games taking up a majority of sales; null ESRB rating values would need to be addressed in data preprocessing. Also, there was an outlier rating of KA that was not identified in the most recent ESRB ratings category; more research would need to be conducted on the KA rating. Therefore, not only did preliminary Tableau data visualizations help identify the need to delve deeper into potential missing values and disguised missing values such as KA, it also helped identify key differences in markets that would need further statistical review.

## **Data Preprocessing**

In order to preprocess the data, some data exploration was conducted on the VGsales dataset by creating a basic data dictionary of the 16 variables as shown below.

Table 4

*VGsales Data Dictionary*



*Note*. This is a basic data dictionary of the VGsales dataset. Adapted from A practitioner’s guide to business analytics: Using data analysis tools to improve your organization’s decision making and strategy by R. Bartlett, 2013, p. 263.

When reviewing the data dictionary, there were 49,256 missing critic scores and 55,457 missing user scores. Data smoothing techniques were utilized in the preprocessing of the Critic Score and User Score variables. One technique that is commonly utilized to smooth over missing variables is use of a measure of central tendency such as the mean or median to fill in the missing value (Han et al., 2011). Any missing values of the Critic Score were filled in with the mean value of 7.2 while missing values of User Score were filled in with the mean value of 8.3. In addition, all missing values of the Total Shipping, Global Sales, NA Sales, PAL Sales, JP Sales, and Other Sales variables were smoothed over with their mean values as depicted in the data dictionary. Any missing ESRB ratings were assigned a rating of RP, rating pending, and any games with the older rating of KA were reclassified as E. Also, the new Publish Type variable was created by rolling up the various publishers into the higher classifications of first-party and third-party publishers. In addition, in order to conduct binary logistic regression, the dependent variable must be dichotomous. Therefore, the continuous total shipment, global sales, and regional sales variables were also turned into dichotomous categorical variables based on measures of central tendency; amounts greater than the median for each respective variable were classified as a 1 for a high transaction event. Furthermore, new binary variables were created specifically to be utilized in binary logistic regression; each distinct record within the genre, publisher type, and ESRB rating were converted into dichotomous variables. Thus, the VGsales dataset was preprocessed to accommodate the correlation and binary logistic regression tests conducted in the project; the data was smoothed over by utilizing the mean value while dichotomous variables were created form the original attributes.

# **Literature Review**

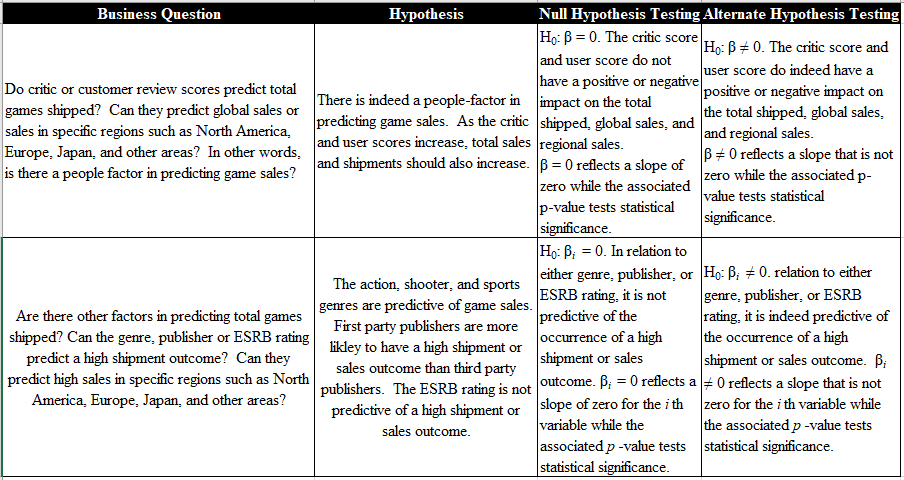
When it comes to video games, very little research was conducted to discover what actually constitutes a blockbuster game. In fact, much of the research that was completed on video games was based on the characteristics constituted a blockbuster movie. However, one major difference between a blockbuster movie and blockbuster game is that of the supposed correlation between the publisher and the “killer app.” In order for a customer to consume a high-profile game that is considered a “killer app,” he or she must also invest in the right console or technology that can be used to play the popular game. In many cases, the console maker also publishes the “killer app;” these games are known as first-party published games (Cox, 2014). In a research study conducted by Cox (2014), emphasis was placed on the console or platform in which the game was developed. Similar techniques were used as those utilized in this project such as obtaining a publicly available dataset from VGChartz and utilizing statistical techniques such as OLS regression and logistic regression. Cox (2014) defined a “killer app” as an application that allowed one console to dominate another; the game is an exclusive title that could only be found on the respective console. In the conclusion of his research, Cox (2014) discovered that three factors attribute to high video game sales; these factors include the game being released by a major publisher, was released on a popular home console, and is of higher quality based on critic response.

# **Developing SMART Business Questions and Research Hypotheses**

Without developing SMART financial questions, GameStop may lose any sense of direction and focus when dealing with a large volume of data. SMART is an acronym that stands for “Start with strategy,” “Measure metrics and data,” “Apply analytics,” “Report results,” and “Transform business” (Marr, 2015). Big Data giants such as Amazon, Wal Mart, and Target may have the technology and capability to utilize the majority of the data they have available both internally and externally. However, instead of getting lost in a random search exploratory search for association, GameStop may wish to first develop SMART financial questions and take a more targeted approach in the vast Big Data landscape; research questions may then be derived from these more focused business questions. Thus, the SMART business questions and respective research hypotheses are depicted in the table below.

Table 5

*GameStop’s Business Questions and Hypothesis Testing Statements*



*Note*. This table lists the proposed business questions and hypothesis testing statements. Adapted from SAS essentials: Mastering SAS for analytics (2nd ed.) by A.C. Elliot, & W.A. Woodward, 2015.

# **Ethical Big Data Project Concerns**

In the era of Big Data, social media users are considered data generators; people may not be aware that by giving their personal opinion when scoring a product out of a possible 10 points or engaging in everyday social media actions such as clicking “Likes” on Twitter and Facebook, they are acting as data generators; the data generated from such actions provide invaluable information for micro-targeted marketing, sentiment analysis, and even targeted political manipulation (Zwitter, 2014). In the past, such data may not have been in discussions regarding consumer privacy or Big Data ethics as the technology was insufficient to generate any valuable insights from such data (O’Leary, 2017). However, new capabilities such as sentiment analysis which allows organizations to extract opinions or sentiment from text data to determine whether such text is considered positive, neutral, or negative now requires organizations to rethink which consumer data needs to be protected (Marr, 2015). Therefore, data being generated when users and critics score games out of a possible 10 points may generate insights that may one day identify consumer preferences or act as characteristics that predict consumer behavior. Such ethical considerations were taken when reviewing the vgsales-12-4-2019.csv dataset. Much of the dataset is already deidentified with no way to trace user-generated scores back to specific businesses, customers, or critics. Furthermore, the results of the analysis were even further anonymized by rolling up the data to higher levels of summarization such as by third-party and first-party publishers, genre, and ESRB rating. Also, the critic scores and user scores are average scores for each respective game; they do not trace back to any specific customer or critic. Instead, they are only represented in analyses as aggregates in tables, horizontal bar charts, and other data visualizations. Thus, as required in any research project, care is taken to make sure that the researched are not being harmed; only the necessary collected data is utilized for the project (O’Leary, 2017).

# **Research Methodology**

In continuation, statistical analyses were conducted in SAS using the PROC CORR and PROC LOGISTIC statements. All of the data preprocessing techniques outlined above were conducted in the actual SAS code; the SAS code utilized for the project is pasted below.









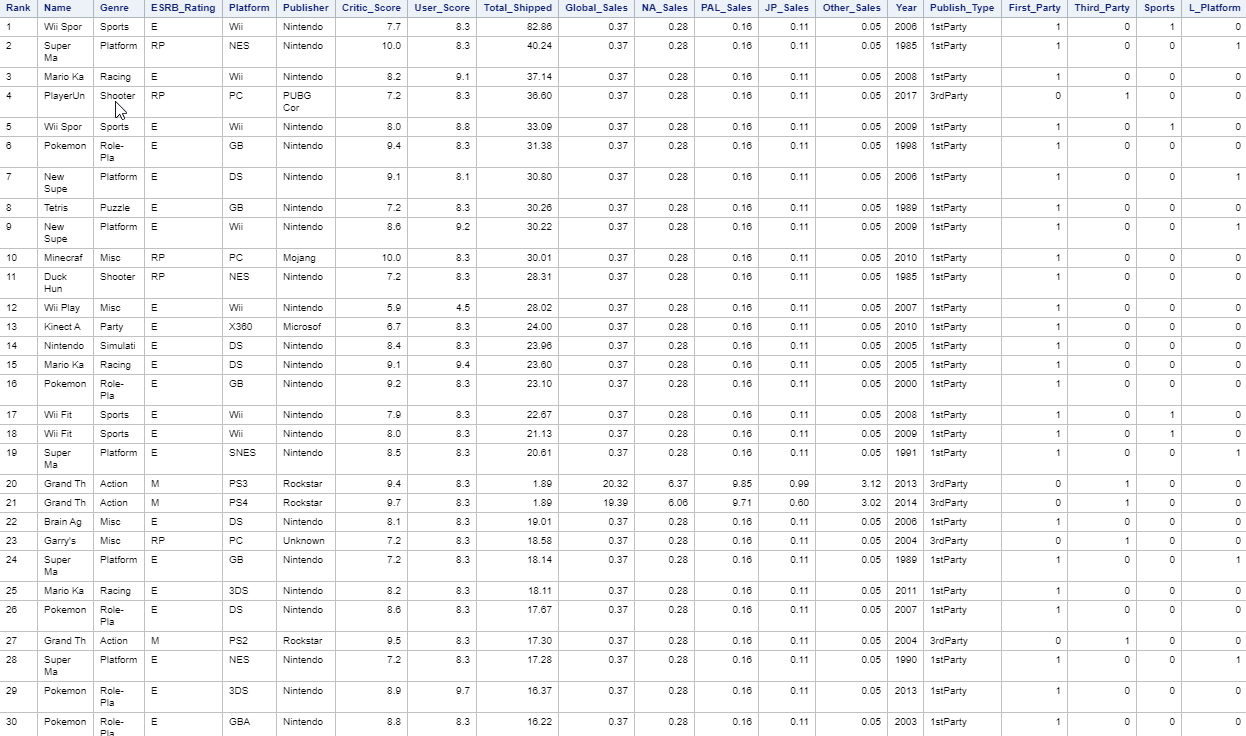


## **Limitations**

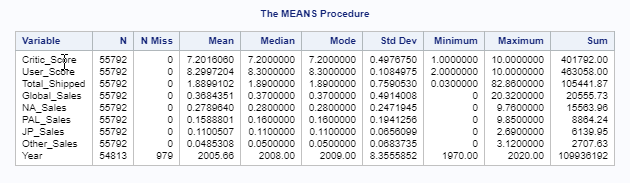
Furthermore, the project only analyzes publicly available data found in the vgsales-12-4-2019.csv dataset. In order for findings to be most effective for GameStop, insights from such publicly available data need to be combined with the internal data of their customers from the PowerUp Rewards program. Access to GameStop’s internal customers’ personally identifiable data was not available for the project. Therefore, the construct of a recommendation engine for GameStop’s customers could not be made without further analysis of the internal data collected from the PowerUp Rewards program. Although the project is limited in scope to only publicly available data, when such data is combined with internal customer data, data-driven decisions may be developed to answer SMART financial questions on how to maximize GameStop’s revenue. Another limitation of the dataset is that the data date is only good through 4-12-2019. Although the data is recent through the first quarter of the year, it is not the most current data available. Thus, the limitations of the project include the inability to utilize real-time data while only analyzing publicly available data.

# **Descriptive Analytics Findings**

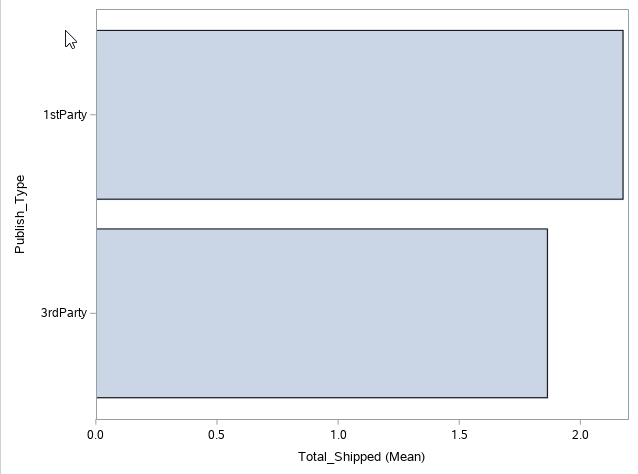
In order to generate key performance measurements and answer some of the “What?” questions so as to gain a sense of what was happening in the VGsales dataset, descriptive statistics procedures were used in SAS (Maisel & Cokins, 2014). Firstly, PROC PRINT was utilized to ensure that all of the new dichotomous variables were successfully created.

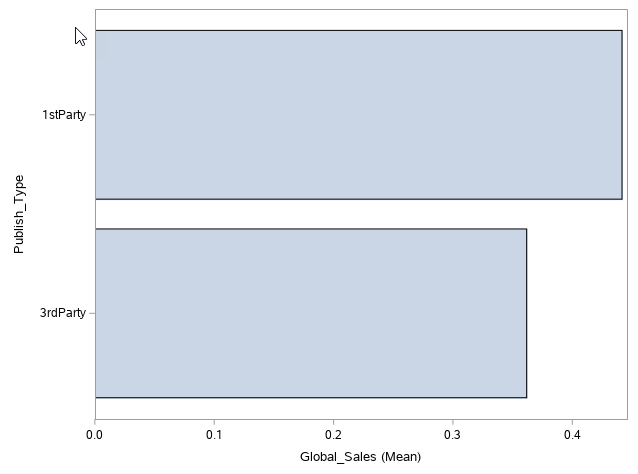


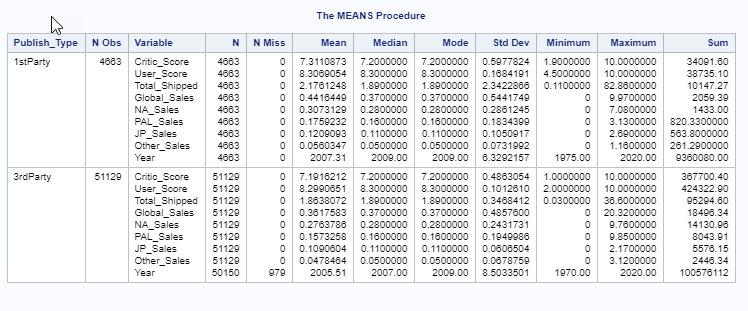
Secondly, PROC MEANS was used to generate a dashboard of basic descriptive measurements.



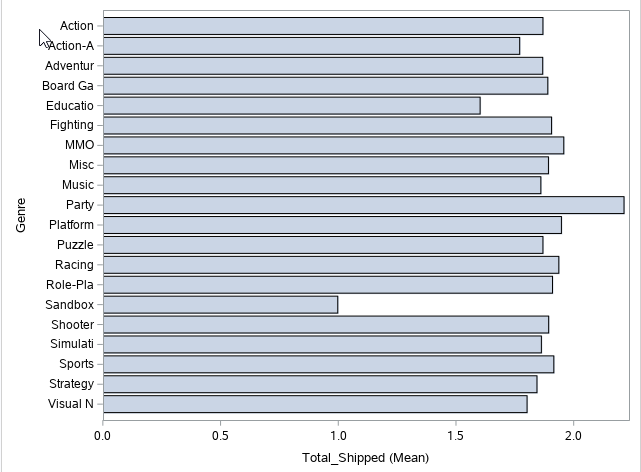
As shown above, the dashboard confirmed that the majority of the missing values were smoothed over with a measure of central tendency; since the year is not being used to answer any of the business questions, the missing values were not smoothed over and the variable was essentially ignored. Thirdly, PROC MEANS was used to create data visualizations in the form of horizontal bar charts with the publisher type as the rolled-up class.

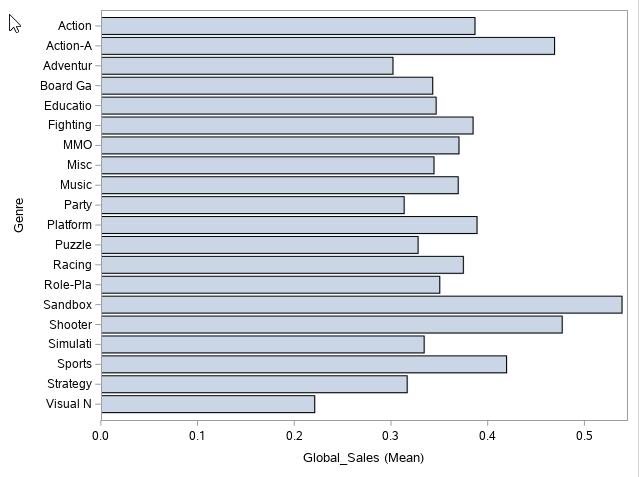


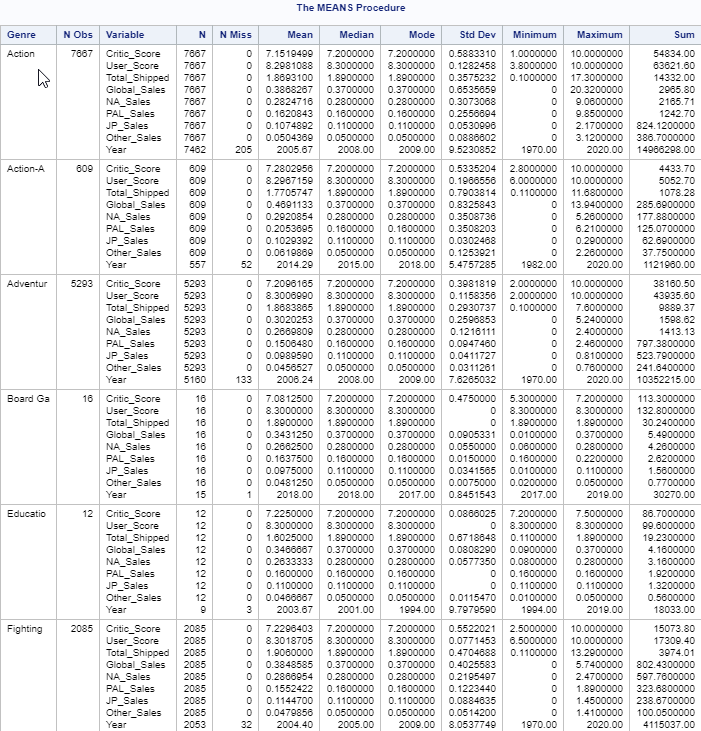


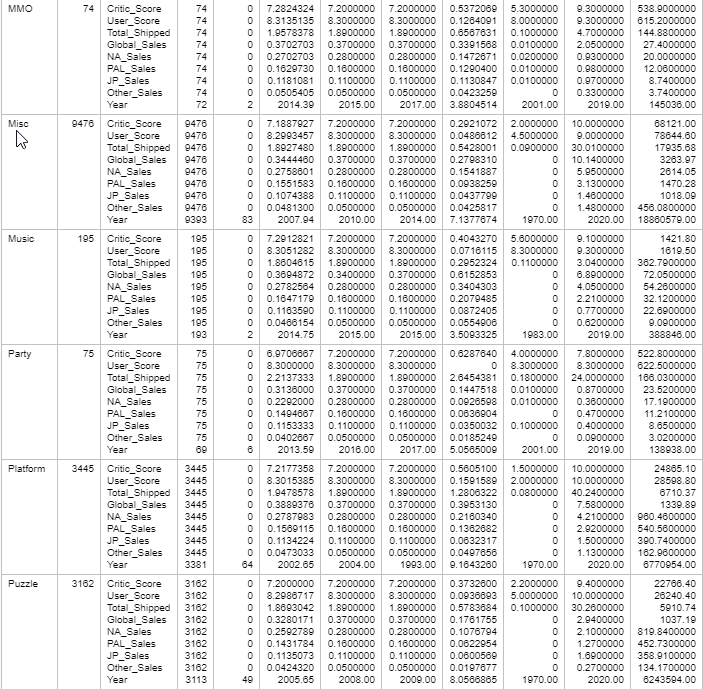


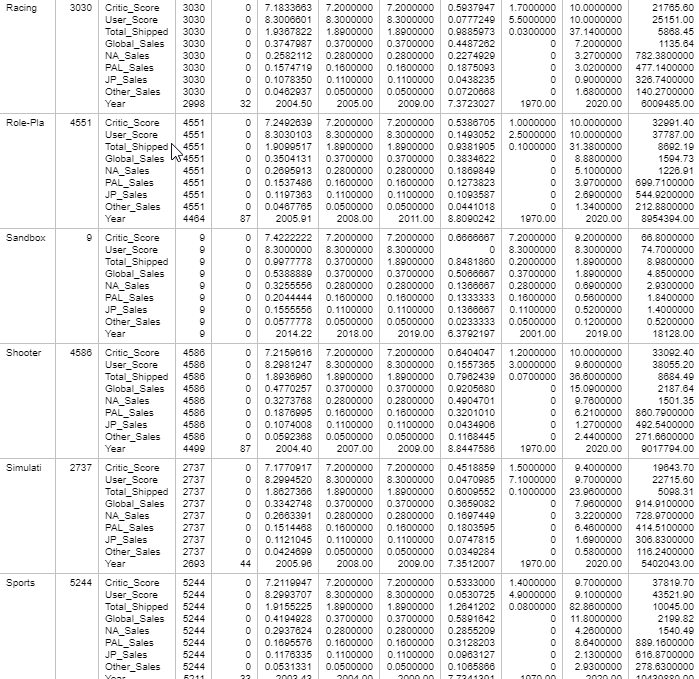
As shown above, first-party titles had higher total shipment and global sales means at 2.18 million and .44 million respectively; on average, first-party publishers ship more titles and generate more in sales per game than third-party titles. Fourthly, the same PROC MEANS procedure was generated with genre as the rolled-up class variable.

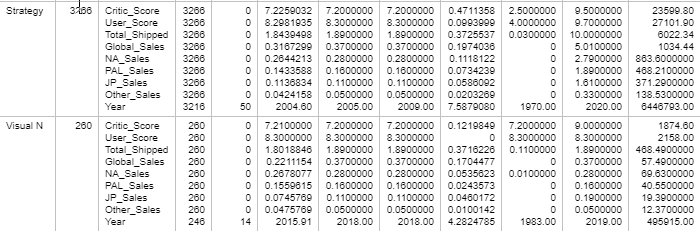




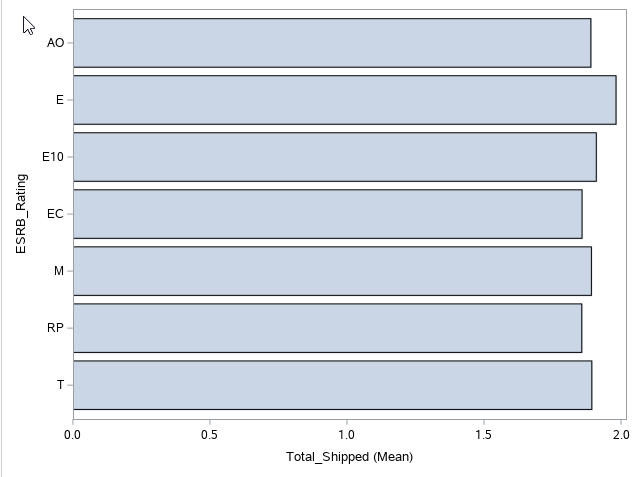


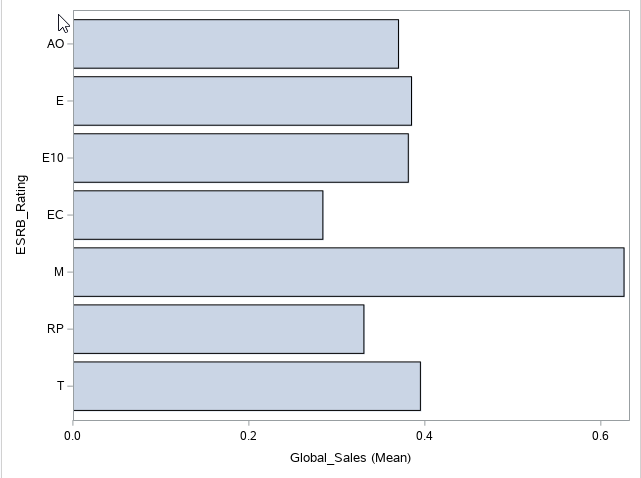


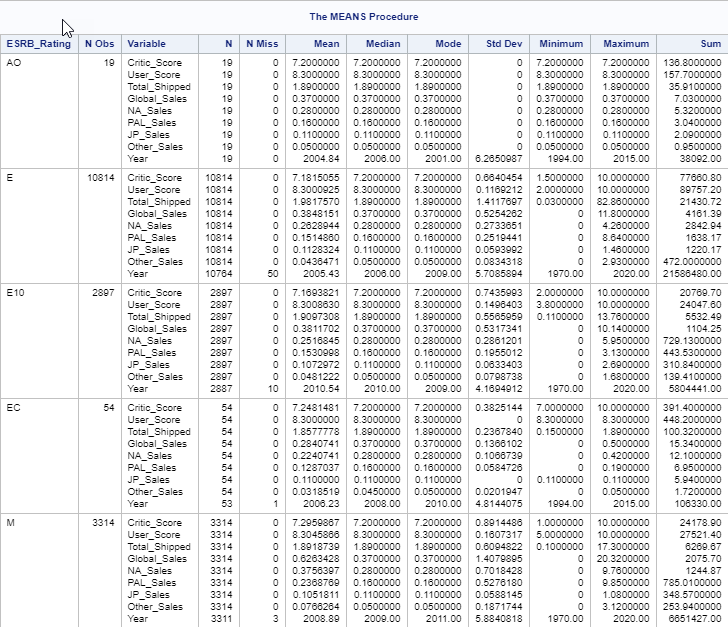


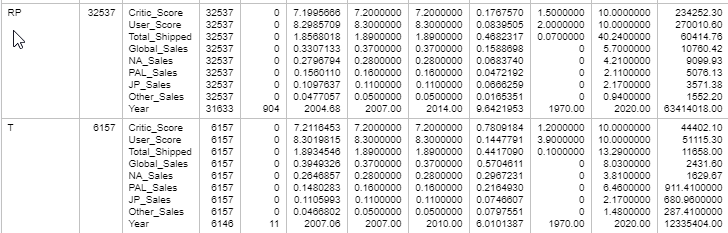


As depicted above, party games genre shipped the most on average at 2.214 million titles while sandbox games sold the most on average at .54 million. Next, PROC MEANS was utilized again while using ESRB rating as the higher rolled-up class variable.





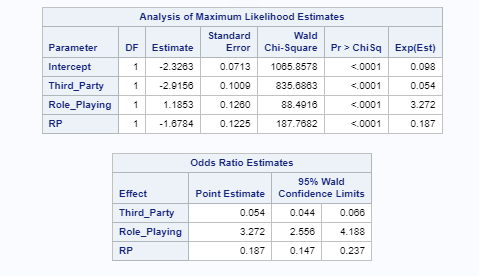




As shown above, E rated games barely nudged out the other ESRB ratings for total average shipped at 1.98 million while M rated games took a huge lead on average global sales per title at .63 million.

# **Predictive Analytics Findings**

In continuation, predictive analytics was utilized to answer the “Why?” questions by predicting how objects may behave in the future (Maisel & Cokins, 2014). Firstly, the results of the binary logistic regression analysis to predict outcomes of high total video game shipments are pasted below.



As shown above, third-party publisher, role playing game genre, and rating pending ESRB rating were all statistically significant at the .05 significance level with a *p*-value < .0001. In addition, the Odds Ratio Estimates shows that third-party publishers are .054 times more likely to have a high total shipment occurrence, role playing games are 3.272 times more likely to have a high shipment occurrence, and rating pending ESRB ratings are .187 times more likely to have a high shipment occurrence. Secondly, the results of the binary logistic regression analysis to predict outcomes of high global sales are pasted below.



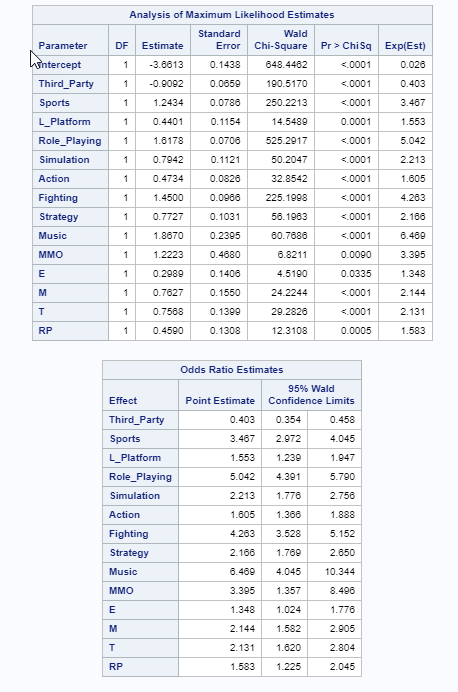
As shown above, first-party publisher, sports genre, racing genre, puzzling genre, fighting genre, strategy genre, music genre, E rating, M rating, and RP rating were all statistically significant at the .05 significance level with a *p*-value < .0001. In addition, the Odds Ratio Estimates shows that first-party publisher are 1.628 times more likely to have a high global sales occurrence, sports games are 2.248 times more likely to have a high global sales occurrence, racing games are 1.363 times more likely to have a high global sales occurrence, puzzle games are .388 times more likely to have a high global sales occurrence, fighting games are 1.397 times more likely to have a high global sales occurrence, strategy games are .386 times more likely to have a high global sales occurrence, music games are 2.162 times more likely to have a high global sales occurrence, E rated games are .703 times more likely to have a high global sales occurrence, M rated games are 1.458 times more likely to have a high global sales occurrence, and games that were not yet rated are .074 times more likely to have a high global sales occurrence. Thirdly, the results of the binary logistic regression analysis to predict outcomes of high North American Sales are pasted below.



As shown above, third-party publisher, sports genre, role playing game genre, puzzle genre, fighting genre, strategy genre, E rated games, M rated games, and RP rated games were all statistically significant at the .05 significance level; all categories had a p-value < .0001 with the exception of the fighting genre which had a p-value of .0009. In addition, the Odds Ratio Estimates reflects that third-party publishers were .784 times more likely to have a high sales occurrence in North America, sports games were 1.939 times more likely to have a high sales occurrence in North America, role playing games were .534 times more likely to have a high sales occurrence in North America, puzzle games were .290 times more likely to have a high sales occurrence in North America, fighting games were 1.322 times more likely to have a high sales occurrence in North America, strategy games were .206 times more likely to have a high sales occurrence in North America, E rated games were .782 times more likely to have a high sales occurrence in North America, M rated games were 1.435 times more likely to have a high sales occurrence in North America, and games within a rating pending were .043 times more likely to have a high sales occurrence in North America. Fourthly, the results of the binary logistic regression analysis to predict outcomes of high European sales are pasted below.



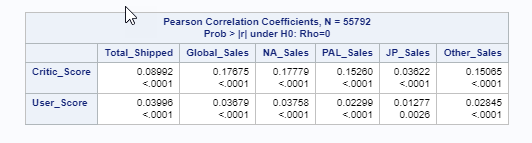
As shown above, first-party publisher, sports genre, racing genre, role playing game genre, puzzle genre, action-adventure genre, strategy genre, E rated games, and M rated games were all statistically significant at the .05 significance level with a *p*-value < .0001. In addition, the Odds Ratio Estimates demonstrates that first part publishers were 1.469 times more likely to have a high sales occurrence in Europe, sports games were 1.88 times more likely to have a high sales occurrence in Europe, racing games were 1.648 times more likely to have a high sales occurrence in Europe, role playing games were .526 times more likely to have a high sales occurrence in Europe, puzzle games were .329 times more likely to have a high sales occurrence in Europe, action-adventure games were 1.751 times more likely to have a high sales occurrence in Europe, strategy games were .272 times more likely to have a high sales occurrence in Europe, E rated games were .577 times more likely to have a high sales occurrence in Europe, M rated games were 1.779 times more likely to have a high sales occurrence in Europe, and games with a rating pending were .038 times more likely to have a high sales occurrence in Europe. Next, the results of the binary logistic regression analysis to predict outcomes of high Japanese sales are pasted below.



As shown above, third-party publisher, sports genre, platform genre, role playing game genre, simulation genre, action genre, fighting genre, strategy genre, music genre, MMO genre, E rated games, M rated games, T rated games, and RP rated games were statistically significant at the .05 significance level; all categories had a *p*-value < .0001 with the exceptions of platform, MMO, E, and RP which had p-values of .0001, .009, .0335, and .0005 respectively. In addition, the Odds Ratio Estimates shows that third-party games were .403 times more likely to have a high sales occurrence in Japan, sports games were 3.467 times more likely to have a high sales occurrence in Japan, platform games were 1.553 times more likely to have a high sales occurrence in Japan, role playing games were 5.042 times more likely to have a high sales occurrence in Japan, simulation games were 2.213 times more likely to have a high sales occurrence in Japan, action games were 1.605 times more likely to have a high sales occurrence in Japan, fighting games were 4.263 times more likely to have a high sales occurrence in Japan, strategy games were 2.166 times more likely to have a high sales occurrence in Japan, music games were 6.469 times more likely to have a high sales occurrence in Japan, MMO games were 3.395 times more likely to have a high sales occurrence in Japan, E rated games were 1.348 times more likely to have a high sales occurrence in Japan, M rated games were 2.144 times more likely to have a high sales occurrence in Japan, T rated games were 2.131 times more likely to have a high sales occurrence in Japan, and games with a rating pending were 1.583 times more likely to have a high sales occurrence in Japan. Furthermore, the results of the binary logistic regression analysis to predict outcomes of high other sales are pasted below.



As shown above, first-party publisher, sports genre, racing genre, role playing game genre, puzzle genre, miscellaneous genre, strategy genre, music genre, E rated games, M rated games, and RP rated games were statistically significant at the .05 significance level; all categories had a p-value < .0001 with the exceptions of miscellaneous and music which had a p-value of .0010 and .0001 respectively. In addition, the Odds Ratio Estimates reflects that first-party games were 1.517 times more likely to have a high sales occurrence in other regions, sports games were 2.642 times more likely to have a high sales occurrence in other regions, racing games were 1.745 times more likely to have a high sales occurrence in other regions, role playing games were .581 times more likely to have a high sales occurrence in other regions, puzzle games were .286 times more likely to have a high sales occurrence in other regions, miscellaneous games were 1.283 times more likely to have a high sales occurrence in other regions, strategy games were .217 times more likely to have a high sales occurrence in other regions, music games were 2.273 times more likely to have a high sales occurrence in other regions, E rated games were .458 times more likely to have a high sales occurrence in other regions, M rated games were 1.743 times more likely to have a high sales occurrence in other regions, and games with a pending rating were .028 times more likely to have a high sales occurrence in other regions. Finally, the results of the correlation analysis with critic scores and user scores as the dependent variables are shown below.



In relation to the critic score and rejecting the null hypothesis, the total shipped, global sales, North American sales, European sales, Japanese sales, and other sales were all statistically significant with a p-value < .0001; the null hypothesis could safely be rejected. However, there was only a slightly positive correlation between the variables with *p*-values of .08992, .17675, .17779, .1526, .03622, and .15065 respectively. In relation to the user score and rejecting the null hypothesis, the total shipped, global sales, North American sales, European sales, Japanese sales, and other sales were all statistically significant at p < .0001 with the exception of Japanese sales with a p-value of .0026. However, the positive correlations were even more slight than the critic scores with *p*-values of .03996, .03679, .03758, .02299, .01277, and .02845 respectively.

# **Conclusion**

After reviewing the results of the analyses, it was discovered that games that ship in higher numbers world-wide do not necessarily translate to high sales outcomes. For example, the only interesting observation from the Odds Ratio Estimates for a high shipment outcome was that role-playing games are 3.272 times more likely to have a high shipment occurrence. However, the descriptive statistics reflect that the mean global sales for role-playing games was not overly impressive when compared to other genres; the sandbox genre led the pack in global mean sales. This may be reflective of a supply issue in which retailers are ordering a larger supply of role-playing games than they can actually sell. Despite role-playing games having a higher likelihood of predicting a high sales event in regions such as Japan where they are 5.042 times more likely to lead to a high sales occurrence, they were not predictive of a high sales occurrence in global sales. In addition, despite being statistically significant, the people-factor in video game sales was almost non-existent; critic scores and user scores only had very slight positive correlations ranging from *p*-values of .01277 to .17779. Although critic scores did predict high shipment and sales better than user scores, especially in North America and Europe, the amount of correlation was not very interesting. In fact, many of the constant factors that predicted a blockbuster games were the publisher, genre, and ESRB rating. For example, M rated games, sports games, and first-party games were more likely to predict high transaction events in most of the analytical tests conducted. Thus, although there were several statistically significant discoveries, it was found that the non-people factor such as publisher, genre, and ESRB rating play a much larger role in predicting a blockbuster game than people-factor such as critic and user scores.

## **Recommendations**

In order best utilize their internal customer data from their PowerUp Rewards program, GameStop needs to know what constitutes a blockbuster game in each region in which it operates. One discovery is that games that shipped the most globally does not necessarily mean that those games sell well in every region. What GameStop may need to pay attention to are the publisher, ESRB rating, and genre when predicting high sales. Consumers are much more likely to purchase games in specific genres that they enjoy; this is evidenced by the varying genres that are more likely to predict high sales events in various regions. For example, in most regions such as North America, Europe, and Other, sports games dominate with a sports game being 1.939 times more likely, 1.88 times more likely, and 2.462 times more likely to predict a high sales event respectively. However, in Japan, music games are 6.469 times more likely to predict a high sales event; music games are far more likely to sell well than a sports game in Japan which is 3.467 times more likely. Furthermore, consumers are more likely to purchase M rated and first publisher games. If GameStop were to take such predictors of blockbuster games and tailor it to the internal data on its customers so as to develop recommendation engines based on genre, ESRB rating, and publisher that the customer is likely to purchase, GameStop would be able to predict games that customers would actually want to buy instead of trying to simply push games in which they believe would be the most profitable. In fact, predicting consumer behavior is the basic goal of many Big Data analytics projects. Thus, with up-close data on consumers that brick-and-mortar retailers such as GameStop collect every day, they have the data to derive invaluable insights; it is simply a matter of utilizing the data to the same degree as e-commerce giants such as Amazon.

## **Tools and Project Links**

Both Tableau and SAS were the two tools utilized for this project. In addition, a video of the PowerPoint presentation for this project can be found at https://youtu.be/ROzawlOLGt4. Furthermore, the project code and accompanying documents can be found at https://github.com/lawilson06/MIS581Project. Finally, the speech narration for the PowerPoint presentation was practiced using VoiceVibes and submitted to the instructor at https://app.myvoicevibes.com/join?code=CSU-Global-MIS581-FallB2019-Goodwin.

# References

Bartlett, R. (2013). A practitioner’s guide to business analytics: Using data analysis tools to improve your organization’s decision making and strategy. New York, NY: McGraw- Hill.

Cox, J. (2014). What Makes a Blockbuster Video Game? An Empirical Analysis of US Sales Data. *Managerial and Decision Economics*, *35*(3), 189–198. https://doi.org/10.1002/mde.2608

Davis, K. (2012). Ethics of big data. Sebastopol, CA: O'Reilly Media, Inc. ISBN: 9781449311797

Elliot, A.C., & Woodward, W.A. (2015) SAS essentials: Mastering SAS for analytics (2nd ed.). Hoboken, NJ: John Wiley and Sons.

Entertainment Software Rating Board. (2019). Ratings guide. Retrieved from https://www.esrb.org/ratings-guide/

GameStop Corporation. (2019). *Form 10-K 2019*. Retrieved from http://news.gamestop.com/sec- /sec-filing/10-k/

Goldman, D. (2019, June 05). The video game sales slump is killing GameStop. Retrieved from https://www.cnn.com/2019/06/05/tech/gamestop-sales/index.html

Han, J., Kamber, M., & Pei, J. (2011). Data mining: Concepts and techniques (3rd ed.). Waltham, MA: Morgan Kaufmann Publishers. ISBN 9780123814791

Machi, L., & MacEvoy, B. T. (2016). The literature review – Six steps to success (3rd ed.). Thousand Oak, CA: Corwin – Sage Publishing. ISBN: 9781506336244

Maisel, L. & Cokins, G. (2014). Predictive business analytics: Forward looking capabilities to improve business performance. Hoboken, New Jersey: Wiley Publishing.

Marr, B. (2015). Big data: Using SMART big data, analytics, and metrics to make better decisions and improve performance. Chichester, West Sussex, UK: John Wiley and Co.

Marr, B. (2016). Big data in practice: How 45 successful companies used big data, analytics, and metrics to make better decisions and improve performance. Chichester, West Sussex, UK: John Wiley and Co.

Mourdoukoutas, P. (2019, March 15). Walmart And Target Beat Amazon. Retrieved from https://www.forbes.com/sites/panosmourdoukoutas/2019/03/14/walmart-and-target-beat- amazon/#e74222f38a87

O’Leary, Z. (2017). The essential guide to doing your research project (3rd ed.). Thousand Oaks, CA: Sage Publishing. ISBN: 9781473952089

VGChartz. (2019). Year on year sales & market share charts. Retrieved from http://www.vgchartz.com/#

Zwitter, A. (2014). Big Data ethics. Big Data & Society, 1(2), 1–6. https://doi.org/10.1177/2053951714559253