The Effect of Genre and Publisher on Global Sales

1 Introduction

Once a videogame has completed development, it handed off to a publishing company to handle selling it. The company picked to publish a game is a decision that could have a significant impact on how well the game sells on a global scale. Additionally, due to the immense variety of videogames that exists it would be helpful for developers to understand the interaction between these two aspects so that they can hand the game off to a publishing company that is best for their specific game. Given the data of many videogames and their sales, genre, and publisher I aim to answer the following questions.

* + - How large of an effect does a publisher have on the global sales of a game?
    - How large of an effect does a genre have on the global sales of a game?
    - Does the interaction between genre and publisher produce a noticeably different effect than they do individually?

If a publisher is given a game to market that they have experience with they will be able to more effectively influence the global sales of that videogame. I base this belief on how a company with a lot of experience marketing action games should theoretically be better at it than a company that has mostly marketed sports games. From the data it seems that there exists some kind of association between these two variables and the global sales.

2 Data

The data for this analysis was originally compiled together by user Gregory Smith and made publicly available on Kaggle[[1]](#footnote-1). User Rush Kirubi found this set and was inspired by it, choosing to expand upon the data by adding additional variables and republished this expanded data set to his Kaggle[[2]](#footnote-2). The unit of observation for this dataset is the videogame name and Smith initially created the dataset through non-probability sampling, only adding games that sold 100,000 copies or more to his dataset. Kirubi expanded upon this initial collection by conducting a web scrape of the website Metacritic, a popular media review site that has an extensive section for video games. Due to the differences in games listed on each site the limitations of this data show itself in how there are many missing observations in the expanded data set. Of the 16,719 entries Kirubi states that there are about 6,900 entries that have an observation for every variable. From this 6,700 we are only going to be analyzing those that sell under 2 million units globally, come from the 5 most popular publishing companies, and fall under the 4 most popular genres; Action, Sports, RPGs, and Miscellaneous. Ideally, with this data set we should be able to identify A screenshot of a cell phone

Description automatically generatedstatistically significant relationships to predict the success of games developed in the future.

2.1 Variables

For this analysis the variables I will be looking at the genre and publishing company of a video game and determining whether these have an effect on the global sales of a game.

2.1.1 Genre

Genre is a variable that is similar to those used by movies and TV shows. For videogames it is used as a way to categorize media based on its gameplay challenges and is not defined by its visual style or narrative. From our univariate table, the genres that make up the largest proportions are the Action genre (37%) and the Sports genre (36.5%). Meanwhile, games that fall under the “Miscellaneous” genre make up 16.2% and RPGs make up 10.3%.

A picture containing sitting, black, computer

Description automatically generated2.1.2 Publisher

A publishing company is in charge of financing the development of a videogame, securing contracts for distribution, marketing, and handling the localization of these games to regions outside of the one it was created in. Table 1 shows us that there are publishing companies that favor games of a certain genre as opposed to others. For instance, the publishing company Activision has taken on 2 times more action games than sports and over 8 times more than RPGs. Company genre preference can be seen in the other popular publishing companies as well, with Electronic Arts publishing 507 sports titles while only publishing only 175 action ones.

A close up of a logo

Description automatically generated2.1.3 Global Sales

The final variable involved in this analysis is the global sales variable. The global sales variable is a measurement of units sold by a videogame in all regions. This metric is also measured in millions of units meaning that 1.2 would equal 1,200,000 copies sold. From the summary statistic table for this variable we can see that there exists standard deviation of around 396,000 copies and a mean of 367,000 copies. Interestingly, the top selling 25% of videogames make up the range of 500,000-1,990,000 copies sold. This means that the lower 75% of videogames encompass about 25% of the overall range of sales numbers.

3 Analysis

A close up of a map

Description automatically generated Prior to conducting our ANOVA analysis, we do a preliminary inspection and check the 6 assumptions of an ANOVA analysis. The preliminary inspection will give us an idea if there is a potential interaction to analyze and the assumptions must be checked before.

3.1 Preliminary Inspections

A screenshot of a cell phone

Description automatically generatedThrough the figures generate for the preliminary inspection we can conclude that interactions may exist between these three variables. First, by analyzing Figure 1 we can see that genre seems to affect the distribution of global sales in a videogame. For instance, sports games tend to sell better than the other genres based how the density of the other genres is higher on the lower end and lower than sports throughout the rest of the distribution. Figure 2 strongly shows us that there may be a connection between publisher and global sales. This is due to how Electronic Art has a mean of around .4, the highest among the companies. Figure 3 looks for significant interactions between the genre and publisher. What we found is that there is evidence to suggest that there is a significant interaction due to how there are multiple instances of lines crossing each other. Looking at the yellow, purple, and green lines we can see A close up of a map

Description automatically generatedthat there is no interaction between action oriented videogames and those labelled as miscellaneous. With this we can move on to checking the assumptions before beginning the ANOVA analysis.

3.2 Assumptions

Before we began the ANOVA analysis, we checked the data against the 6 assumptions for ANOVA analysis. First, using the information provided by Rush Kirubi, the person who created the dataset, we can determine that each datapoint is independent of the others as each game was recorded individually. Second, the global sales variable is ratio numeric due to how it has a defined value at 0.0. Third, by filtering out the extreme outliers we prevent them from biasing our analysis. Following that, the homogeneity of variance is investigated through Levene’s test and the p-values calculated for each variable and their interaction show that this dataset violates the homogeneity of variance. Finally, as it can be seen in Table 1 the group sizes are not equal or similar in size, with the largest difference being 29 samples to 507 samples and thus violates the assumption of equal groups. While the data violates multiple assumptions of an ANOVA analysis we will continue with our analysis.

A screenshot of a cell phone

Description automatically generated3.3 Analysis

A picture containing game

Description automatically generatedFor this ANOVA analysis, we have chosen 0.05 to be our alpha level. Looking at the test for the main effect of genre we can see that the p-value is less than the alpha level, allowing us to conclude that genre is a significant predictor of global sales. Moving on to the test for the main effect of publisher we see that the p-value is less than the alpha level. This lets us come a similar conclusion in how publisher is a significant predictor of global sales. Then, the f-test for interaction resulted in a p-value of 0 and this shows that the interaction between publisher and genre is significant. From these tests we can reach the conclusion that genre and publisher are predictors of global sales individually and when interacting with each other. Moving on from the ANOVA analysis, we can check the last assumption; the normality of residuals. Referring to Figure 4, we can see that the residuals of our dataset are not normally distributed. This can be seen in how the residuals are positively skewed. From the pairwise comparisons we find that the difference in genre is around 80k copies sold. This is fairly large as the mean number of copies sold globally for a video game is around 370k so a difference of this level is significant. Next, by calculating the r-squared value we conclude that the overall model of genre and publisher explain 12.49% of the variance in global sales and at this level it means that the combination of genre and publisher influences sales to a large degree. After calculating the Cohen’s f value we find that the power of this analysis leans more to the medium degree as the value of .378 comes closer to the .5 medium value rather than the .2 small value. Taking all of those post-hoc values into consideration we can conclude that the difference in global sales by genre and publisher is significant and substantive.

3.4 Conclusion

Overall, the conclusion that we reached is that at the selected alpha level of .05 our findings are statistically and substantively significant and as such we are able to reject the null hypothesis that our two variables have no effect on global sales figures. I believe that the dataset was large enough that we can generalize these findings to the videogames of these genres and use this information to better market and sell these types of games.

Ideally, in a future analysis we would use a dataset that had much more samples with more variety in genre so that we can see if this trend holds true across the videogame industry as a whole or if publisher only matters for certain genres. To fix this, we would conduct web scrapes on a variety of videogame review sites, not just VGChartz and Metacritic.

1. <https://www.kaggle.com/gregorut/videogamesales> [↑](#footnote-ref-1)
2. <https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings> [↑](#footnote-ref-2)