Project Submission C2

By: Trey Hoopchuk, Ryan Torres, Eli Martin, Garrett Pritchard

We hope to understand what factors of a Video Game have the most impact on the success a game measured in sales. For this analysis we will be focusing on the Developer of the game, the Critic review scores, Genre, and Platform to try and determine their potential impact on our success metric of total shipped units. While general you would expect a positive correlation between review score and Units shipped of a game, this may only hold true for Critic scores. User scores have been known to be much more volatile and may have been influenced by things outside of the game’s attributes, such as review bombing due to the publisher having a PR scandal. It is likely that the distribution of the Critic scores variable will follow a normal distribution given the nature of our dataset.

Chart, pie chart

Description automatically generated

This graph is a representation of the top 50 companies in terms of total units shipped across all games a publisher has published. This data isn’t entirely up to date as it only goes to 2018, but given Nintendo’s storied history and consistent performance in the gaming industry it’s no wonder they hold more than 50% of the total sales of all games published in this data set.

Graphical user interface

Description automatically generated with medium confidence

While hard to read, this is our graph showing the relationship between publisher and the average value of total shipped units of a game that the top 50 publishers published. Essentially what publisher has the best average sales across their entire publishing history. Some companies, such as Nintendo, fall a little down in the rankings due to the sheer number of games published bringing their average units shipped down somewhat despite still having a respectable rank. An opposite case also showed up where PUBG Corporation ended up with the highest average units shipped, however only due to the fact they have only published one game.

Chart, histogram

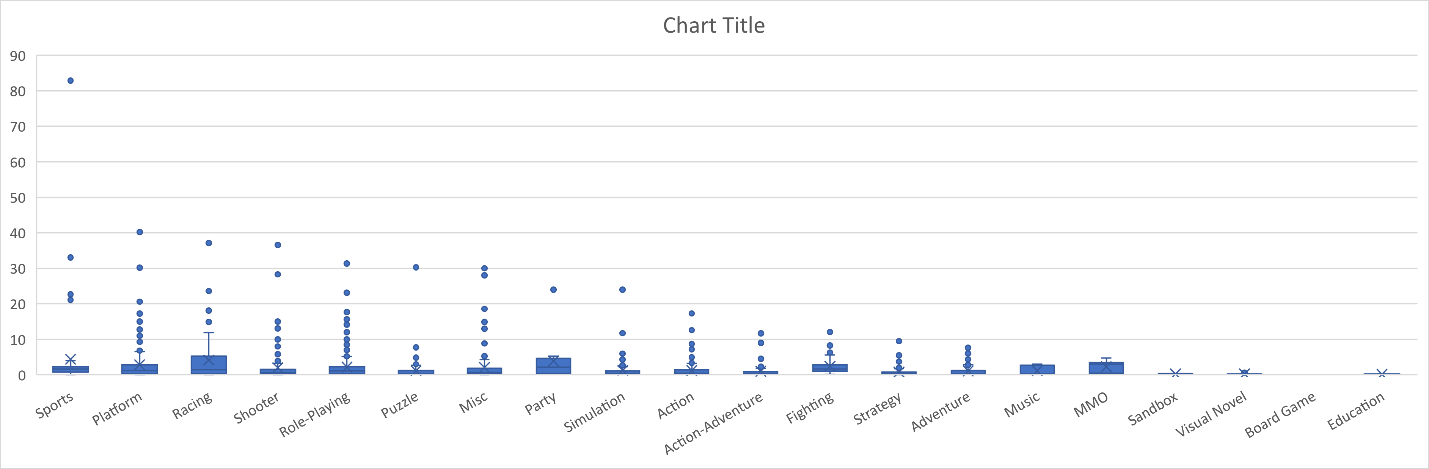
Description automatically generated

This graph shows the distribution of Critic scores across all games. While not entirely normalized, it’s very easy to see the pattern of a normal distribution within this box plot. This holds with our initial hypothesis that Critic scores would generally follow normal distribution given the nature of our dataset.

Chart, scatter chart

Description automatically generated

This graph shows the relation between the total number of units shipped of a game against that game’s respective critic review scores. There is in fact a positive correlation between the total sales of a game and its critic review score, and while it may seem slight, keep in mind the total units shipped is in units of a million (ie: 10 = 10,000,000). This is a very legitimate factor in determining overall sales of a game and on average a difference in a score by 1 or 2 points can potentially change our expected units shipped value by over 1 million.



This is our box plot graph showing the relationship between genre and total sales of a given game. This graph somewhat surprised us as we did not expect the dominant genre in terms of IQR to be so heavily favored from Racing and Party games.

Table

Description automatically generated

This shows the contingency table we had for the top 50 publishers in terms of average units sold. Much of the information isn’t fully shown as it breaks down where the values come from. It’s very interesting to look at this because there are a fair number of publishers with a somewhat large library and high average units shipped, but a generally low average critic score.

Table

Description automatically generated

These are our relevant statistics for our critic score variable. Given the full range of scores range from 1-10, having a mean at 7.2 goes to show that critics seem apprehensive towards giving low scores in general. Having an average score be 7.2 when generally 5 is thought of as the metric for an average game. It’s skewed towards the right side of values. In terms of the range of values we have a somewhat moderate to low variance and standard deviation, showing similar results that Critics stay close to a “safe” score around 7. Our median is greater than our mean here showing that the distribution is negatively skewed which is understandable given how high our mean and median are as values in the range 1-10. IQR shows similar information about this distribution.

Overall, much of the data fell within expectations in terms of how it was distributed, however some data points such as critic score, and genre held slightly different conclusions than initially thought. With our assumption being that critic score wouldn’t show as much of an effect to overall units shipped, and genre having different relationships to the units shipped than expected given our experiences with the games. Genres like platformers were very low compared to genres like sports and racing for instance, which wasn’t what we initially thought.

In terms of biases this dataset doesn’t combine the unit’s sold metric for games that were always available on multiple platforms and never one specific platform. The units shipped metric is entirely limited to the platform that was picked as a games “original” platform which in some cases may be an arbitrary choice or have them as separate sales metrics instead of one combined metric which can change how we perceive how successful a game was. We also didn’t consider the differences between PC and console market and separate the data so that may well have introduced other forms of bias due to the communities at large being somewhat separated and different. There are many factors at play here so these are just two of possibly many that weren’t fully considered but the data and analysis extracted was mostly satisfactory for our purposes of this project.