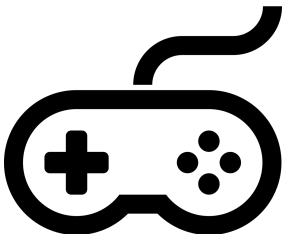
Video Game Trends

Sophia Leung



Introduction

The Video Game industry is one of the fastest growing industries as a multi-billion dollar industry. In 2017, 67% of American households reported that they own a device to play video games on with users the age 18 or older making up 72% of the gaming population making the average age of a gamer 35 years old. Not only is the gaming industry popular with people of all ages, it contributed $11.7 billion in value to the US GDP in 2016. In the U.S., there are more than 2,858 video game companies and 526 publishing companies. With such a growing industry, there are a lot of development studios wishing to gain the same notoriety that Super Mario or Legend of Zelda has gained in remaining as one of the timeless games that one should associate with when the word “video games” comes along. Video Game developers and publishers in this competitive market will benefit from knowing the most optimal time frame for a game release that will generate the most sales, and gauging popularity and potential funding through high Metacritic scores. This report will aim to provide insights on the best game release dates and realistic Metacritic scores for a successful game by analyzing 20 years of video game data.

Game Reviews by IGN and Metacritic’s Scoring System:

In order to be synonymous with video games, not only does the game have been well received by the gaming community but also receive high critic scores. Metacritic scores are widely criticized for how much they influence the success of a video game. Often, publishers will use a developer’s past Metacritic score as a way to determine to fund a video game in development or to accept/reject a pitch for a game. Developers will create and pitch to publishers components of their game that the developers believe may be well-received in order to receive higher ratings, thus resulting in more funding. The analysis shows that the IGN and Metacritic review scale for a critically acclaimed and well-received game is not a true scale for what is reflected in the data. IGN system

is a score within a score based on a 10 point system where each whole integer score is scored for whether it's a strong 7 or 9. IGN's average score is 7.011.717, indicating that scores 8.72 or higher is considered favorable in comparison to IGN's actual scale where a 7.0 or higher is considered favorable therefore the scale is actually higher than what IGN claims to be. Metacritic reviews range from 0-100 with scores 75 or higher are considered generally favorable. However, the data shows that Metacritics tend to give a score of 70.1713.90 indicating that 84.06 or higher is the threshold for a favorable review.

Game Release Dates:

When Sonic the Hedgehog was released in 1991, it became an instant hit and when it came time for the sequel, Sega wanted to have a release date that would be the first global release for any video game. The developers decided to release the sequel on a Tuesday and call it "Sonic 2sday". By announcing the date ahead of time, rather than sporadically releasing a game, built anticipation in the gaming community that was not seen before. Since then, it has become a standard in the industry to release video games on a Tuesday.

Video games now have become more complex with the technology as more genres and platforms have arisen. When deciding on a release date, developers need to take in consideration multiple factors like new platform releases and development, genre competitors, and time of the year where people are most likely to purchase video games. Therefore if release dates and reviews are important factors in the success or failure of a video game.

Data Acquisition and Cleaning

The dataset was combined from two datasets. The first dataset is from IGN and contains information about the Video Game's platform, genre, release year, release

month, release day, editor's choice, and score. The second dataset is from VGChartz that contains similar columns as the first dataset, however with the additional information on the game's publisher, sales in different regions, critic score, and user's score. From there, required information about a video game must include title, platform, release date(year/month/day), and genre so that games can be categorized based on these attributes. Basic information about a game is of interest, the two datasets were merged together on by their respected Name and Platform columns. Because the two datasets contain similar columns, the columns with similar information needed to be standardized. The platform and genre column had to be standardized as these columns had different abbreviations or spelling for the same value. For example, the platform column contains information about the game's platform where from one dataset the value is "PS3" and the other dataset has its value as "PlayStation 3 ". In order to standardize the columns, a unique count of all the column values was taken from both data frames. This was then copied to an excel sheet and was manually matched to the corresponding abbreviations to full platform name. The excel sheet was then loaded as a dictionary and mapped to the Platform column in the merged data frame. However, in the process of doing this, some of the information about the PC genre was lost significantly and may impact the data analysis later. In the genre column, there are row values that contain multi-genre games, e.g. "Action-Shooter" and had 81 unique genres whereas from one particular dataset only had 12 unique genres, so the latter genre column was used.

Also, data about the release year from the two datasets have some rows that do not match entirely. However, the rows of the release year from the two datasets were kept and renamed in order to make a distinction from which dataset it is from because one of the datasets were derived with the game's release year and global sales, it has to be kept. The two datasets have reviews on the game from different sources. One is from the IGN scoring system and the other is pulled from the Metacritic scoring system.

Missing Data and Outliers

The majority of the missing data are in columns associated with ratings of the video game and these missing values were kept while cleaning the data. Outliers in the global sales column were also kept because these outliers are considered to be widely successful games and need to be analyzed. Wii Sports generated the most sales globally at 82.53 million dollars. It's a best selling game because the game was included in the console. Also, the Wii console was the first of its kind with the hand-held controllers and interactive gameplay.

Exploratory Data Analysis

Sales by Genre:

The sum of the row values for the Global Sales column was grouped together by genre and genre sales for each region. Action, Sports, and Shooters are the most popular genres worldwide in terms of global sales as shown in Figure 1. And this trend is still consistent in the North American, Europe, and Other regions with the exception in Japan shown in Figure 2. With Action generating the most global sales at $1174.68 million in sales overall, then Sports at 972.03 and Shooting games at $814.12. Strategy had the least number of sales at $91.86. Genres that generate the most sales in Japan are Role-Playing, Action, and Platform games. Role-Playing sales in Japan are at $117.76 million. In the context of how much Japan loves role-playing games, Japan's regional sales for role-playing accounts in for 25% of the total role-playing sales globally.

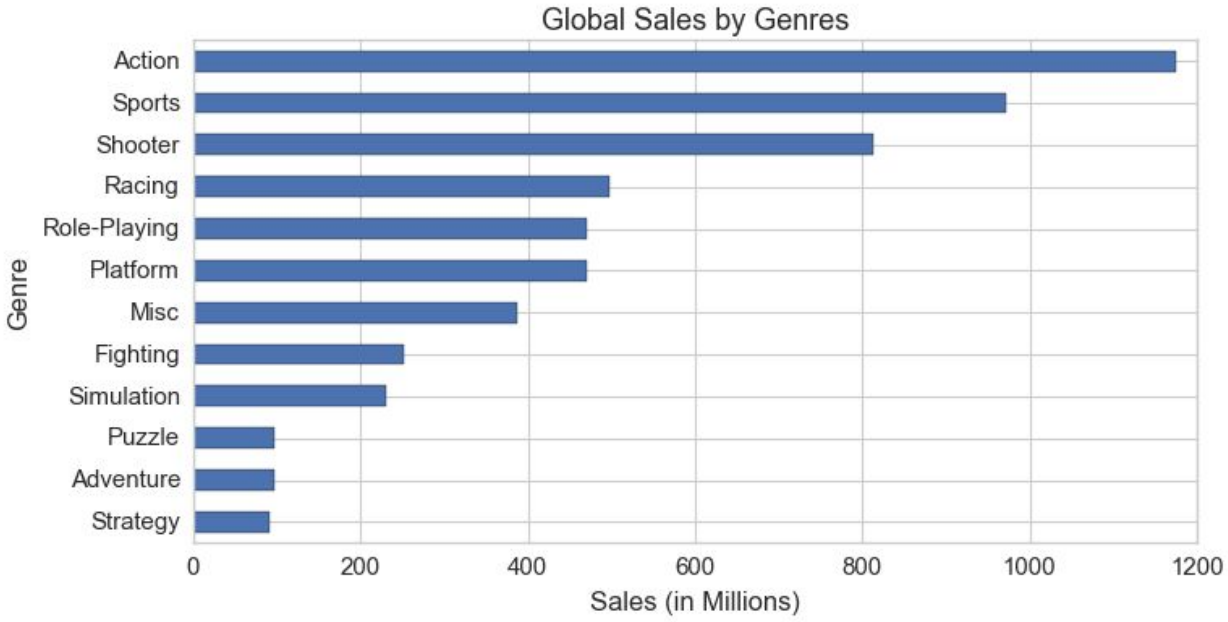


Figure 1: Global Sales by Genres where Action, Sports, and Shooter generated the most sales

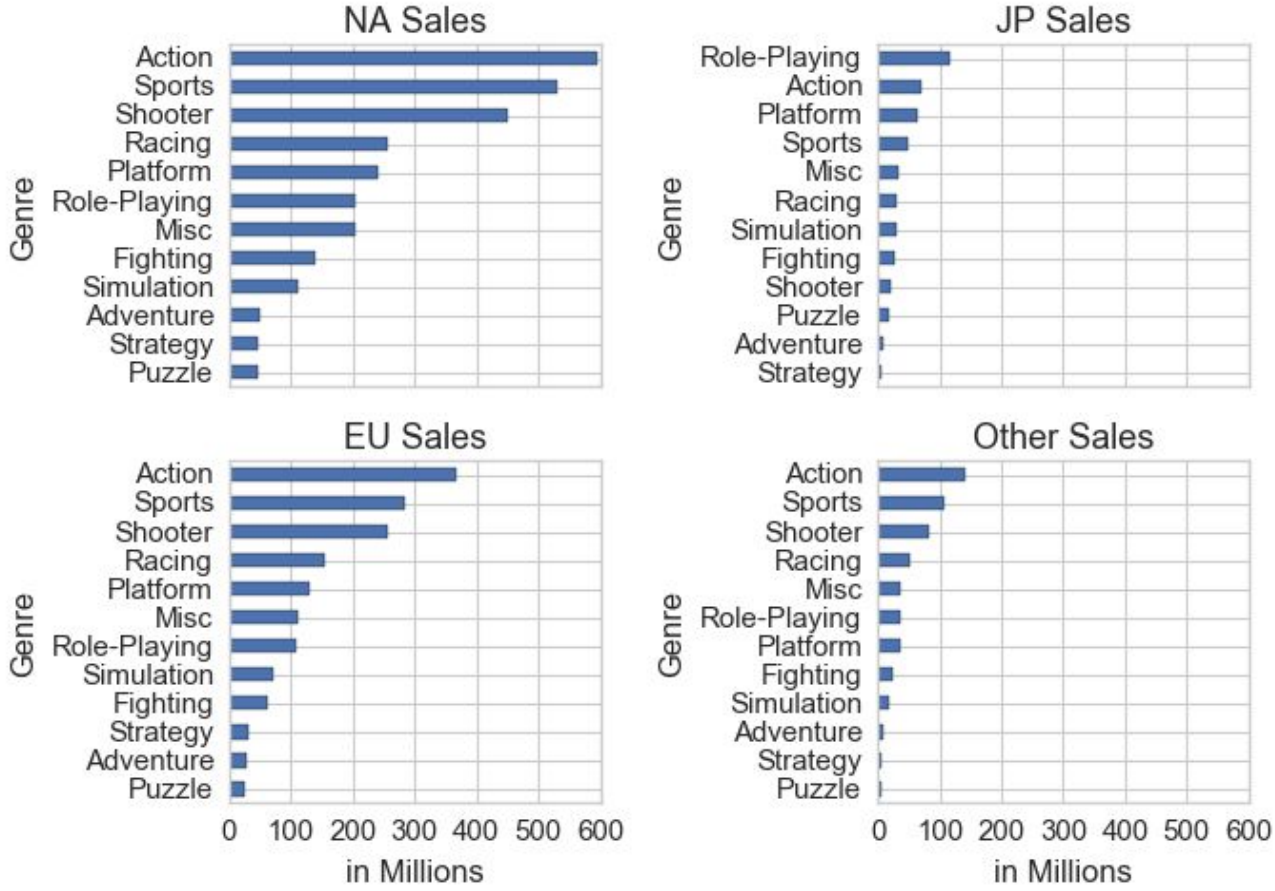


Figure 2: Genres by Regional Sales. Action had the most sales for all regions except Japan.

Platform Sales:

Each platform was grouped together by its sum of sales. Figure 3 shows that globally Playstation 2 is the best selling platform with $910.78 million in sales followed by the XBOX 360. With Playstation being the Xbox's arch nemesis in the gaming world, the data shows that for all generations of the Playstation and XBOX, Sony's Playstation outperforms Microsoft's Xbox in sales.

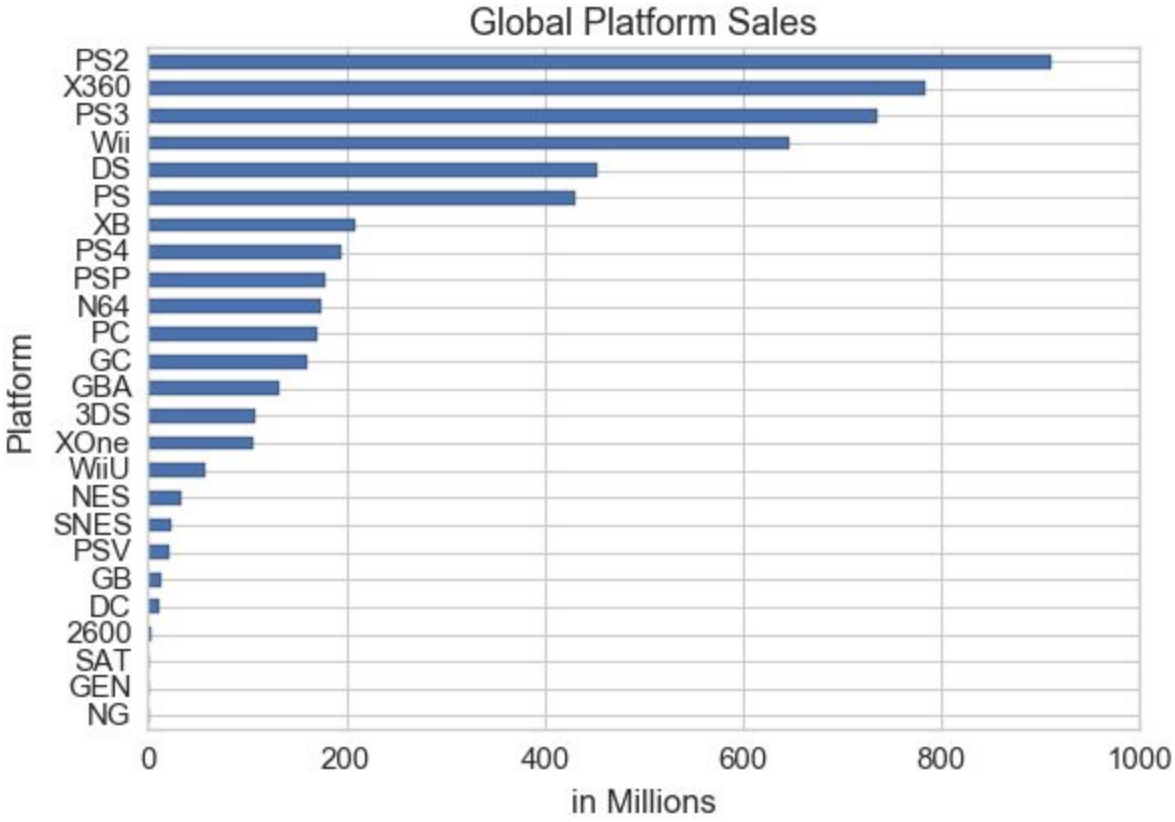


Figure 3: Platform Global Sales. Playstation 2 had the highest sales globally.

When the platform sales were grouped together by its respected region, the PlayStation 2 is in the two platforms that generated the most sales. However, in North America, the XBOX 360 is the most popular and Japan is the DS. The Nintendo DS had the highest number of sales followed by the PlayStation 2 and PlayStation. Any non-Japanese

platform tends to have fewer sales in Japan, like with the XBOX 360 and XBOX ONE. In addition, handheld platforms like the DS, 3DS, and PSP did considerably well in Japan.

The later console generations, like the XBOX ONE and PlayStation 3 did not do well in sales in comparison to their platform predecessors. This may be correlated with the rise in popularity in PC gaming. It's worth exploring at which point did PC gaming became more popular than console gaming. Also, PC gaming in Japan is practically non-existent with one of the least sales made.

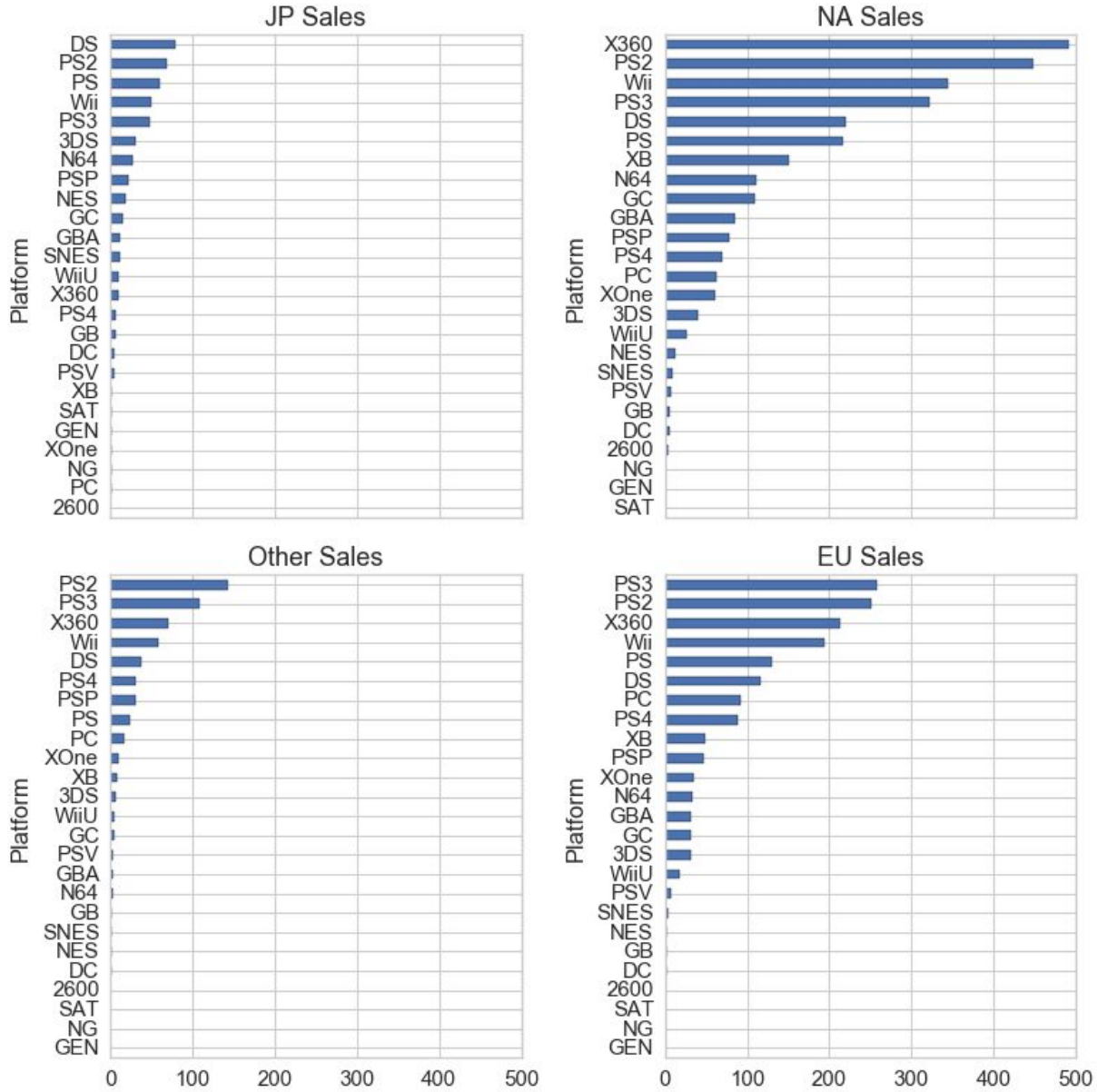


Figure 4: Platform Sales in various regions. Playstation consoles are the most popular.

Editor's Choice in Genres:

When a game is being reviewed by IGN, the editor will score the game based on their system and decide whether it's an editor's choice or not. IGN states their editors are well-rounded but can still weigh their score according to the genre and platform. Figure 5 shows that there are less games that are an editor's choice. Games that tend to be favored by the editors are Sports, Action, and Shooter genres. Games that are favored typically receive a Metacritic score of 84.66 on average. Games that do not receive an editor's choice, are Action, Sports, and Shooter genres. And these games have a Metacritic score of 65.75.

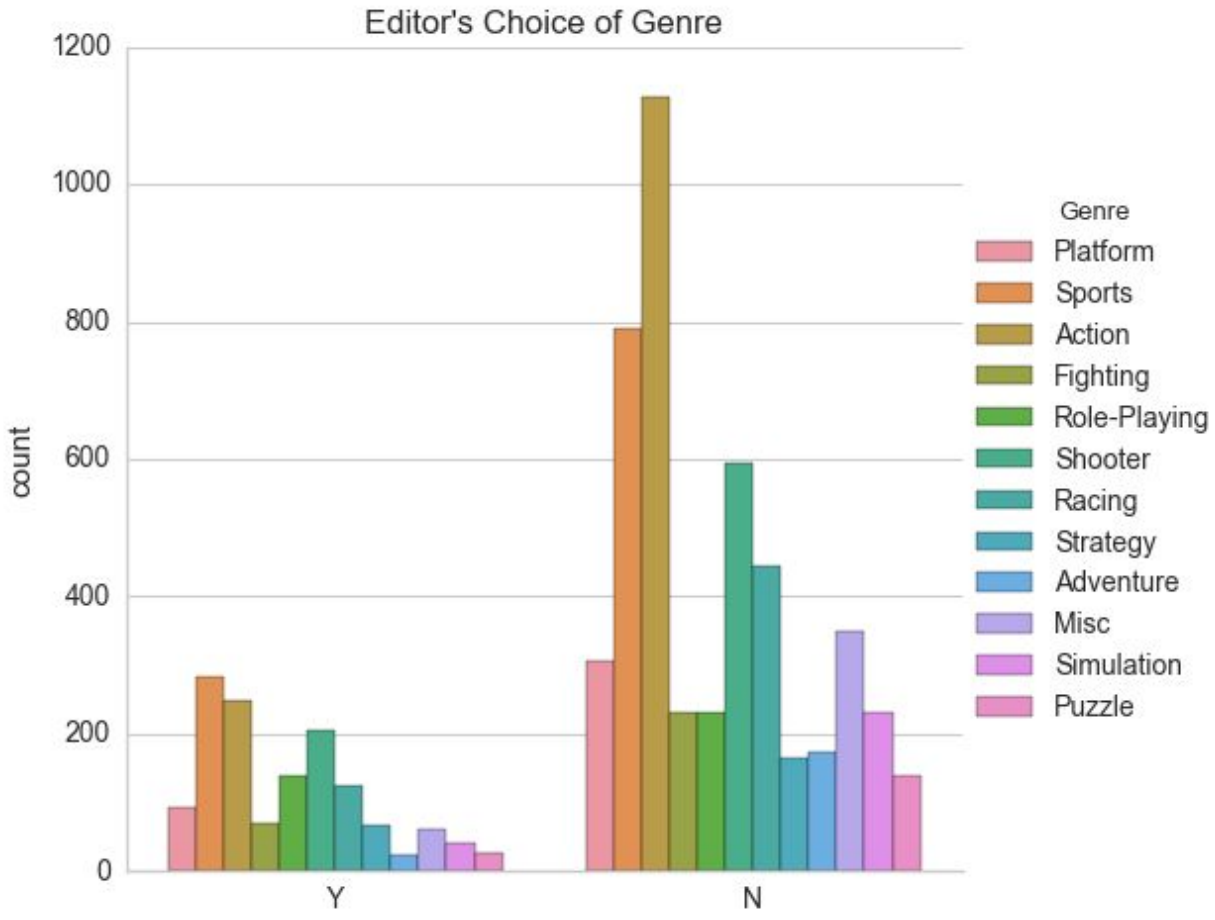


Figure 5: The counts of an Editor's Choice in different genres.

Video Game Release Dates Trends

Monthly Releases by Genre:

The most popular months to release any genre is in November followed by October. This trend seen in supplementary figure is the most consistent in the Action and Shooter genre, which are the most popular game genres by global sales as well. In the majority of the genres, there tend to be a heavy amount of games released around September through November. However, in the offseason, it appears from the data that March also has a decent amount of releases as well.

Sale Trends:

November generated the most sales globally generating $1,441.90 million in global sales alone shown in Figure 6. With October generating $914.78 million. If a video game developer wanted to release a game that generated the most sales, then Fall seems to be the ideal season to release a game (September through November). If developers cannot release their game in Fall when other competitors have oversaturated Fall with releases, then the March is the best month in the Spring followed by May. January and July are the least favorable months to release a game on. January generated the least global sales at $171.74 million and July at $172 million.

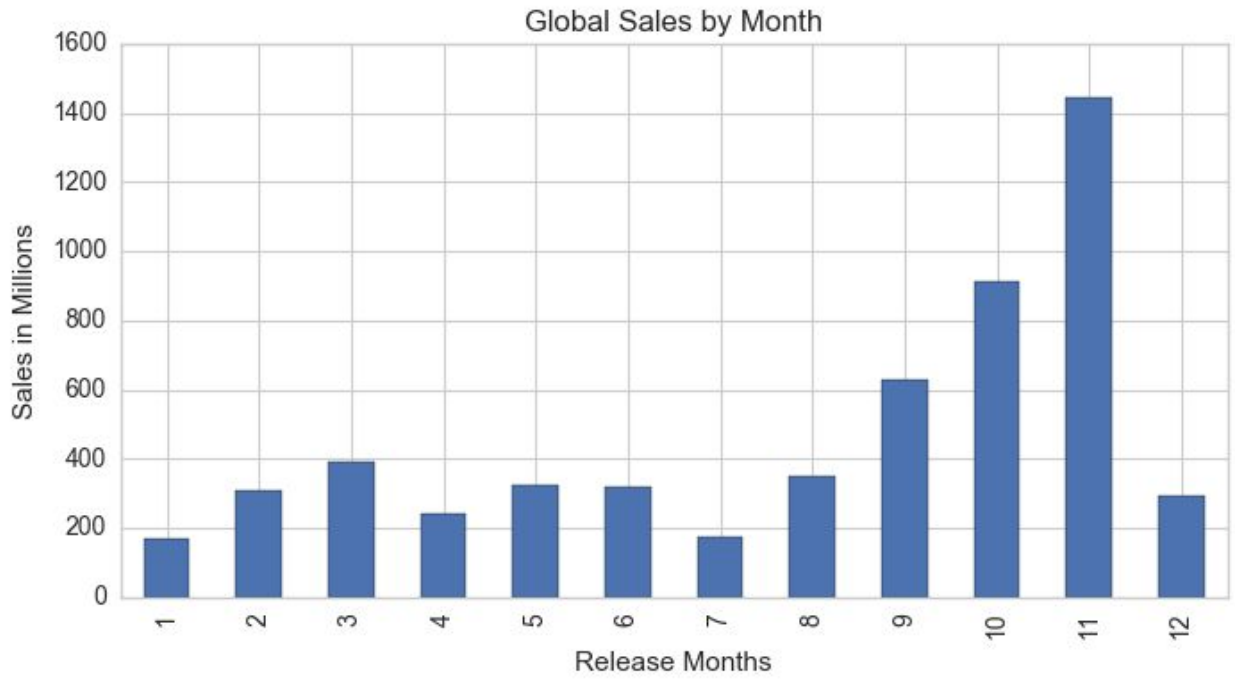


Figure 6: Global Sale Trends by Month where November has the highest sales for all months.

But let's see if November, despite its global sales generating value, if November have the most number of releases. The monthly release proportion is defined by the number of releases for that respected month divided by the total number of releases multiplied by 100 to be expressed in a percentage.

$$\frac{Number\ of\ Releases\ (month)}{Total\ Number\ of\ Releases} \times 100 = Monthly\ Release\ Proportion$$

November once, has the highest monthly release proportion at 19% as seen in Figure 7. So of the total amount of games released within the dataset, 19% of the games were released in November, followed by October (17.1%) and September (11.3%). January had the least releases at 2.1% and July at 4.1%.

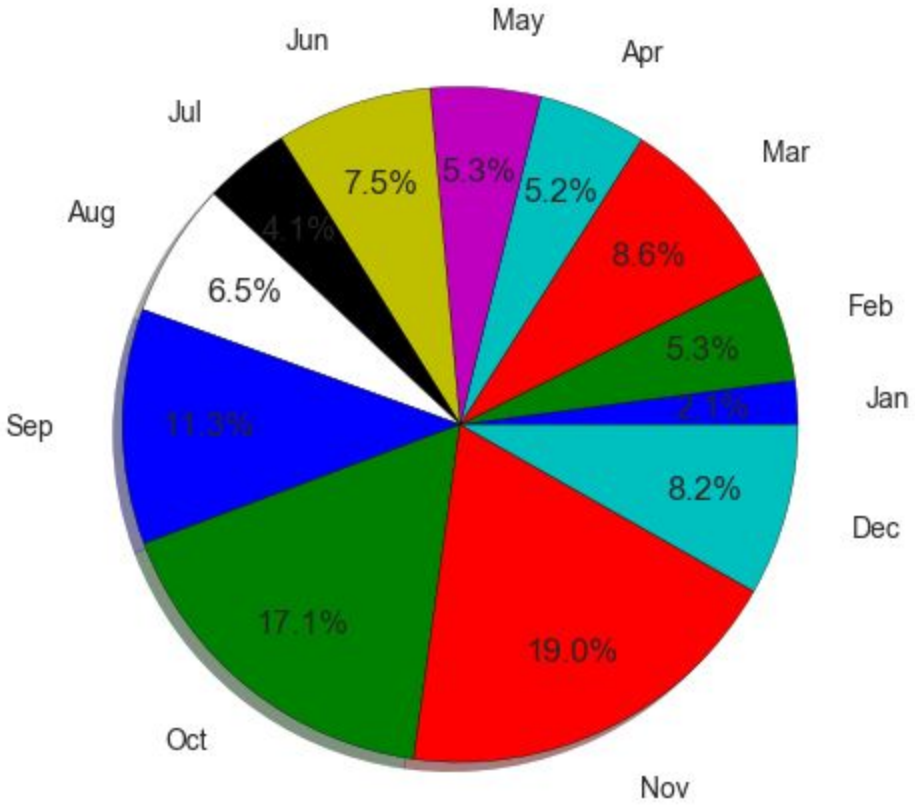


Figure 7: The release proportion by months. November and October having the most releases.

Let's once again, see how often each month releases a game per year and call that the average release rate for that month. Where the number of releases for a month is divided by the total releases for that respected year. This is the annual release rate for each month.

$$\frac{Number\ of\ Releases\ (month)}{Total\ Number\ of\ Releases\ of\ that\ year} \times 100 = Annual\ Release\ Rate$$

This will be used to calculate the overall release rate by taking the average of all the annual release rate for that respected month. Now, we can see how often a game is released by month as shown in Figure 8. On average, 16 games are released in

November out of any year. 15 games are released in October. September releases about 10. January and July have the lowest release rate by 4 and 5 releases respectively.

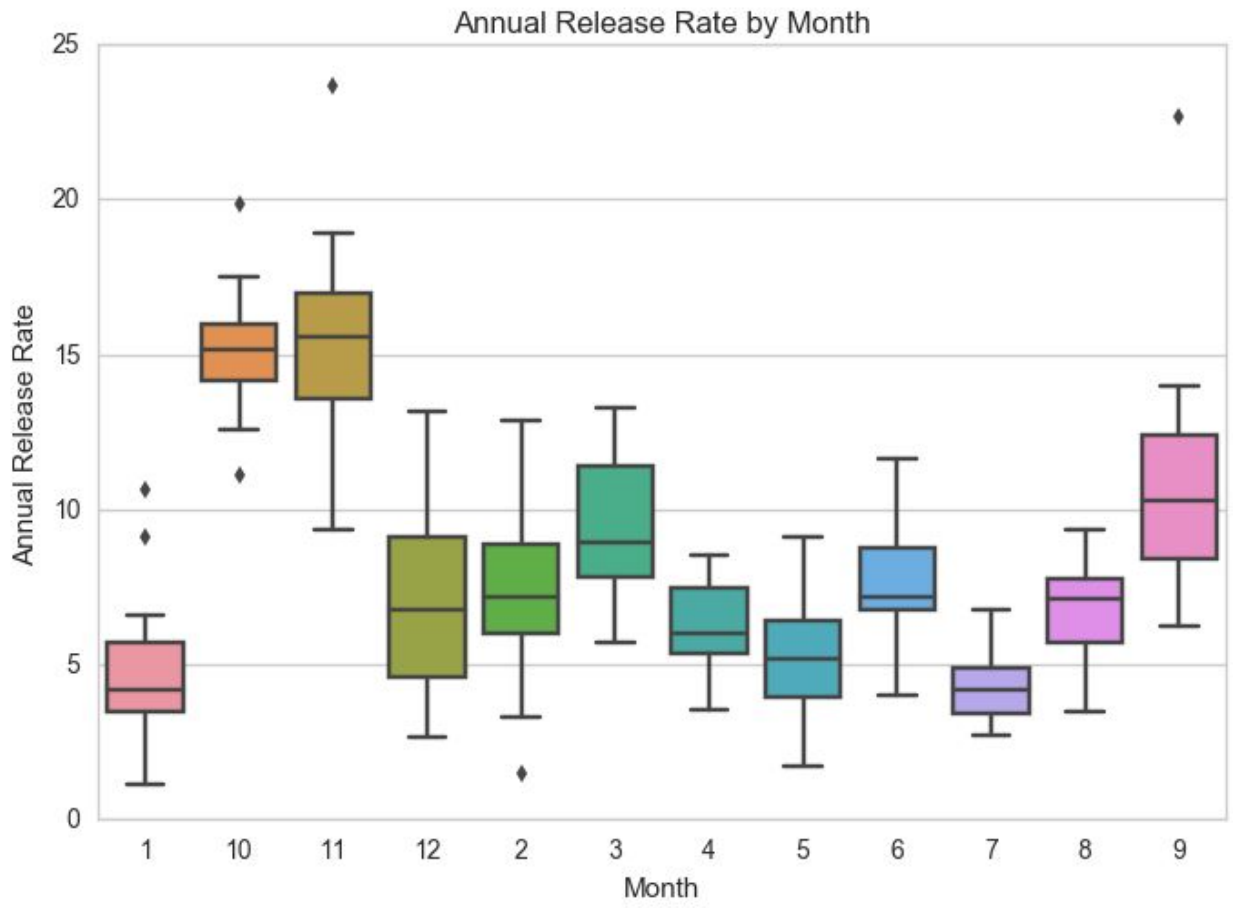


Figure 8: The annual release rate for any given month. On a given year, November and October have the highest number of released games.

The most popular days to release a video game in November are the 14th, 11th, and 3rd as seen in Figure 9. With the 23rd, 29th, and 28th with the lowest numbers of releases on that day of the month. In comparison to release days outside of the month

of November, the 14th, 11th, and 18th are popular days and the 31st, 30th, and 10th have the least numbers of releases. Tuesday and Friday have the highest number of releases in November for the days of the week as shown in Figure 9. In general, Tuesday has the highest number of releases at 1736 releases followed by Monday and Friday, with Saturday and Sunday having the least releases. Tuesday accounts for 19.12% of all releases.

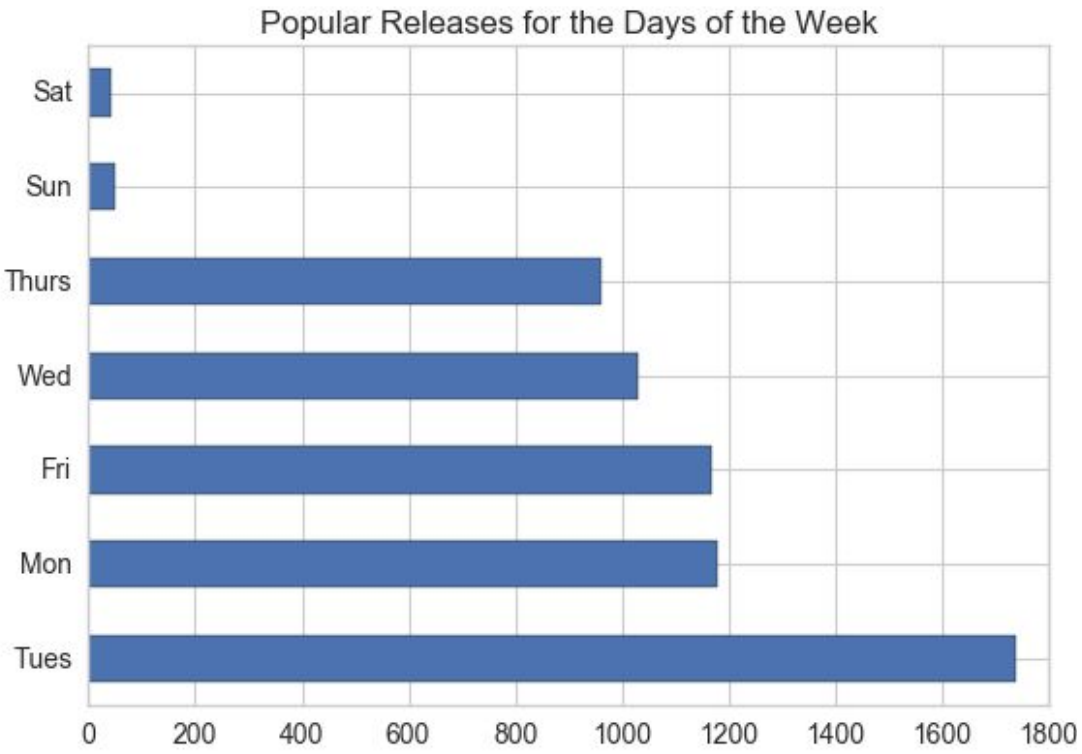


Figure 9: The proportion of releases by the days of the week. November has the highest number of releases and the weekend the least.

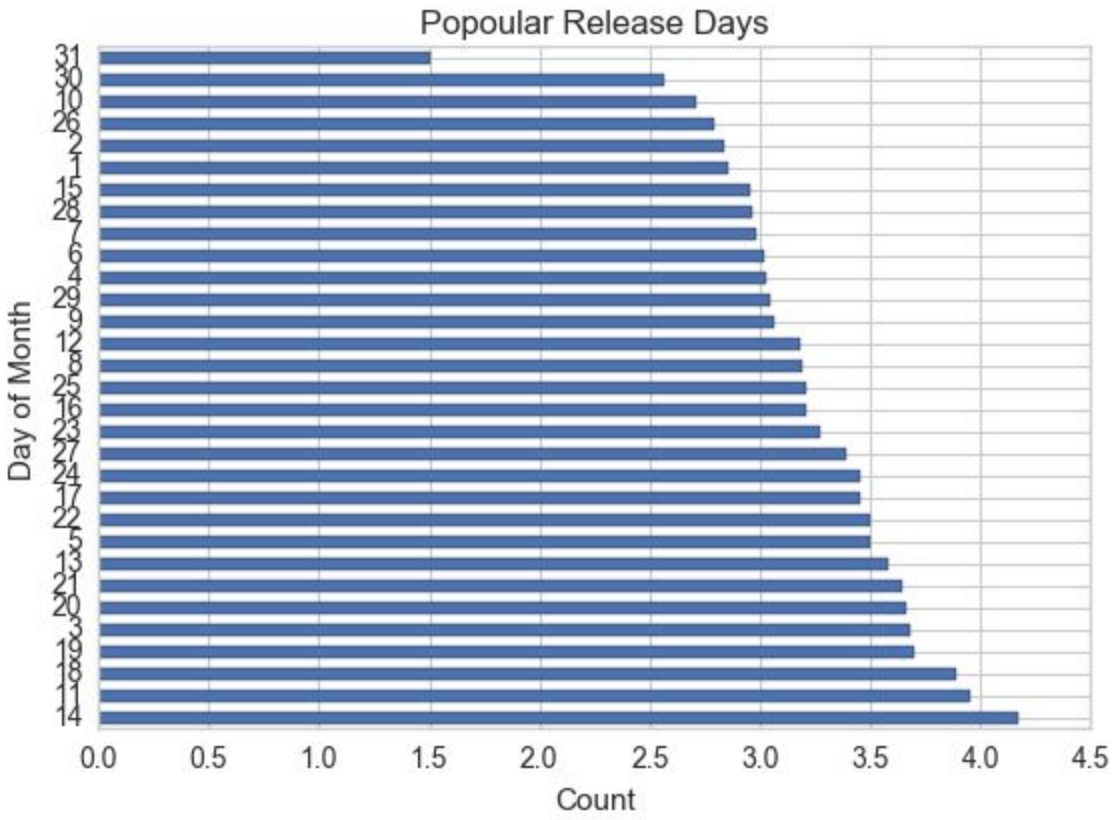


Figure 10: Popular days of the Month to release games. The 14th and 11th have the most reviews with the 30th and 31st with the least.

The Importance of Game Reviews

Distribution of IGN and Metacritic Scores:

The distribution of IGN scores is left-skewed where scores of >6 are more frequent seen in Figure 10. A score of 8 appears the most frequent, followed by a score of 9 and then 7. Scores less than 3 appears the least. According to the IGN scoring system, a score at 5 is considered mediocre, where anything less is considered bad and any score higher is considered good. There seem to be more good games than there

are bad. IGN on average gives a score of 7.01 ± 1.717. Based on the data, a score of 5.29 or less is considered bad and 8.72 or higher is considered successful.

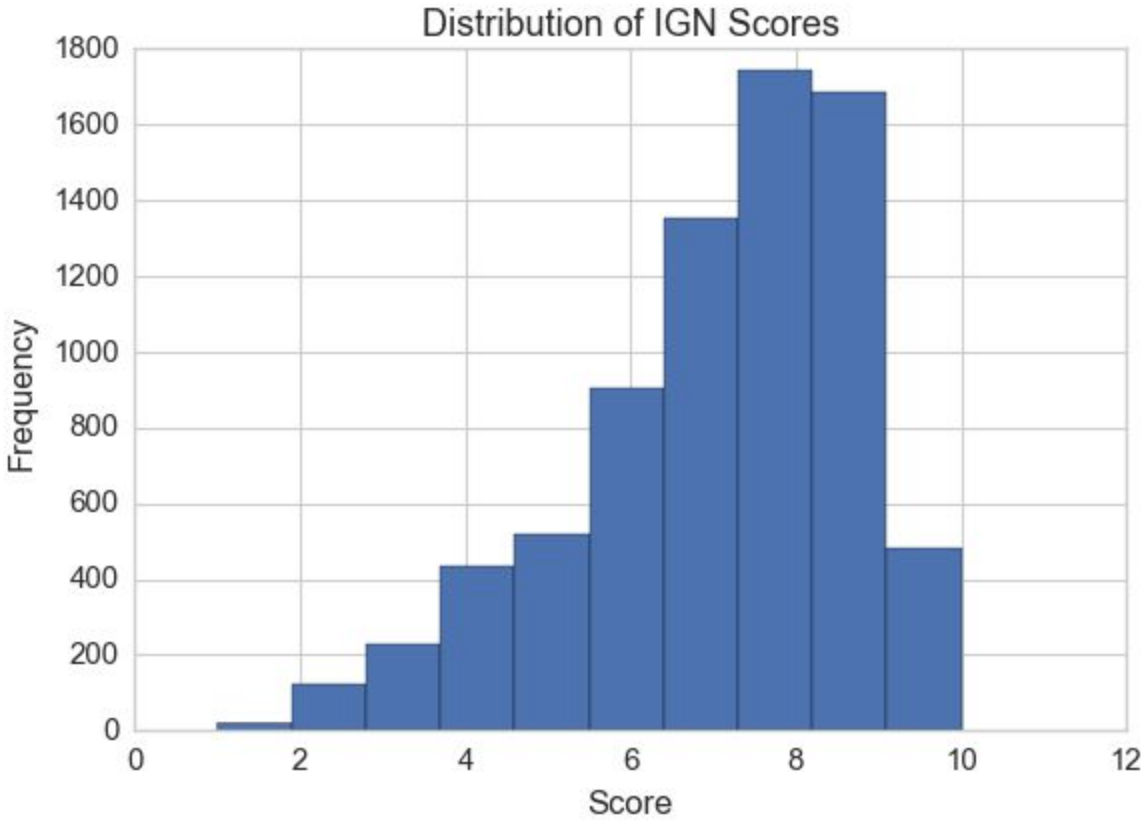


Figure 10: Distribution of IGN Scores shows it is left-skewed.

IGN considers a 10, “Masterpiece”, their highest recommendation for any game. IGN considers the game to be a classic where future developers can learn from. Perfect games include Legend of Zelda Ocarina of Time, Grand Theft Auto V, The Last of Us, Metal Gear Solid V, Super Mario Galaxy 2, Uncharted 3.

The distribution of the Metacritic scores is also left-skewed likewise with the IGN scores. A score of 75 is the most frequent score given. This histogram shows that critics tend to give scores higher than 60. According to the Metacritic scoring system, a 50-74 is considered average, where any score lower is considered unfavorable/disliked and

any score is generally favorable/acclaimed. From the data, Metacritic's average score for games is 70.17 ± 13.90. This differs from their scale, where from the data a score less than 56.27 is considered unfavorable and a score higher than 84.07 is considered favorable.

From the data, the highest Metacritic score given is 98 and this game is critically acclaimed. The only two games that received this score are Grand Theft Auto IV and Tony Hawk's Pro Skater 2. IGN also gave GTA IV a perfect score, but Tony Hawk's Pro Skater 2 did not receive a perfect 10, but a 9.6.

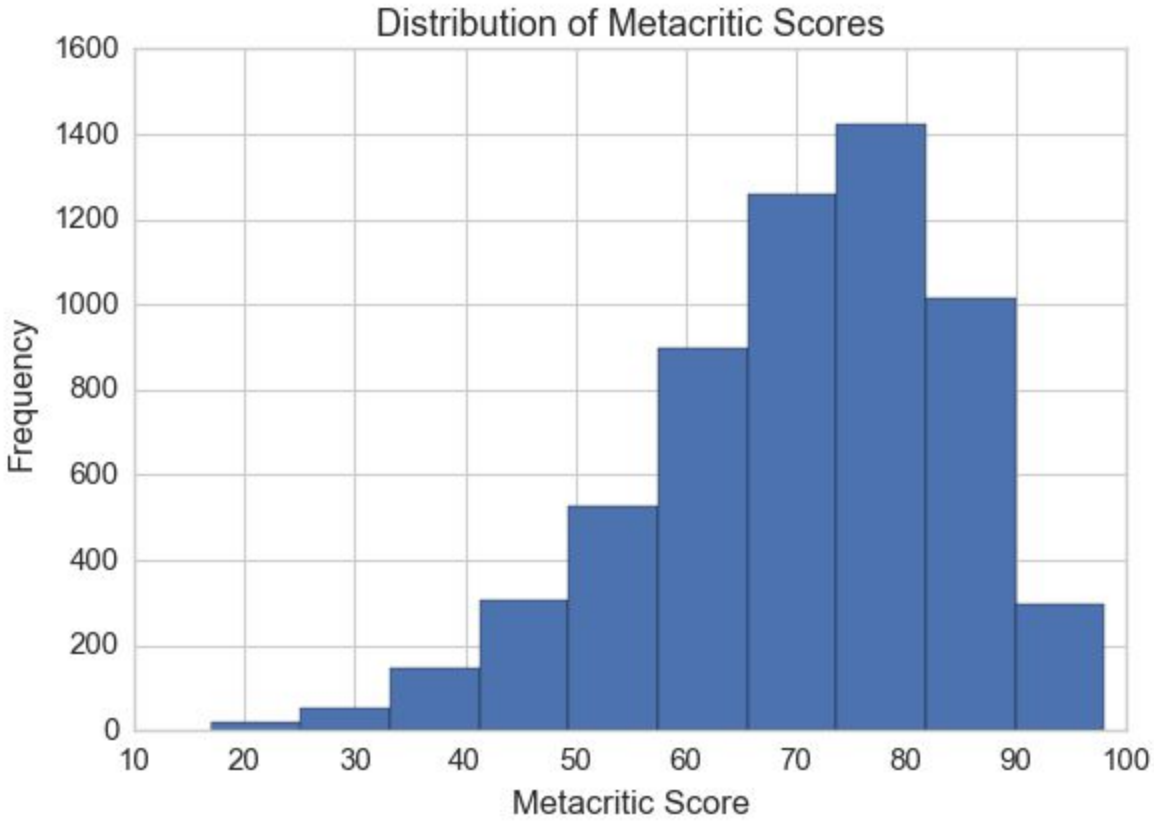


Figure 11: The left-skewed distribution of Metacritic Scores.

Comparing the distribution of the Metacritic and IGN scores show that the two distribution have the identical left-skewed shape. Because the Metacritic scores are on a 0-100 scale and IGN on 0-10, the IGN scores were multiplied by 100.

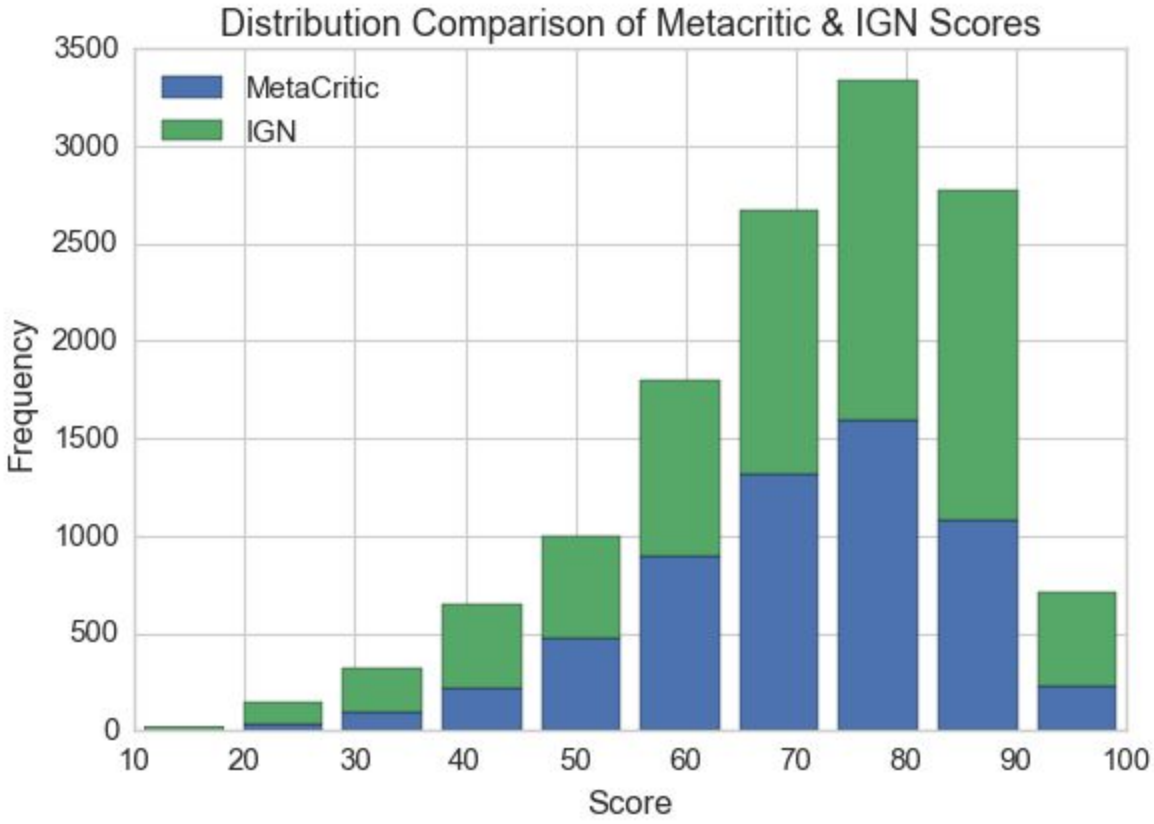


Figure 12: A comparison of the distributions between the Metacritic and IGN scores. The two distributions are left-skewed.

Do Metacritic Scores affect the sales of a Video Game?

In order to investigate if there is an association between a video game's Metacritic score and its sales, a linear regression model was used where the critic score is used as the explanatory variable and the global sales as the response variable. The first model used kept all the data, where outliers and influential points were not yet identified or removed resulted in an adjusted R2 = 0.064. In order to improve the model,

the log transformation of the global sales was taken, and outliers and influential points were removed. The model improved from an adjusted R2 of 0.064 to 0.164.

$$ln(Global\ Sales) = 0.0408 * Metacritic\ Score + 9.7340$$

By taking $(e^{0.0408} - 1)$ * 100, the percent change in global sales by one unit of the Metacritic score can be calculated. Where 16.4% of the global sales of a particular video game can be explained by the critic score and there is a 4.15% increase in global sales for every one unit increase of the Metacritic score.

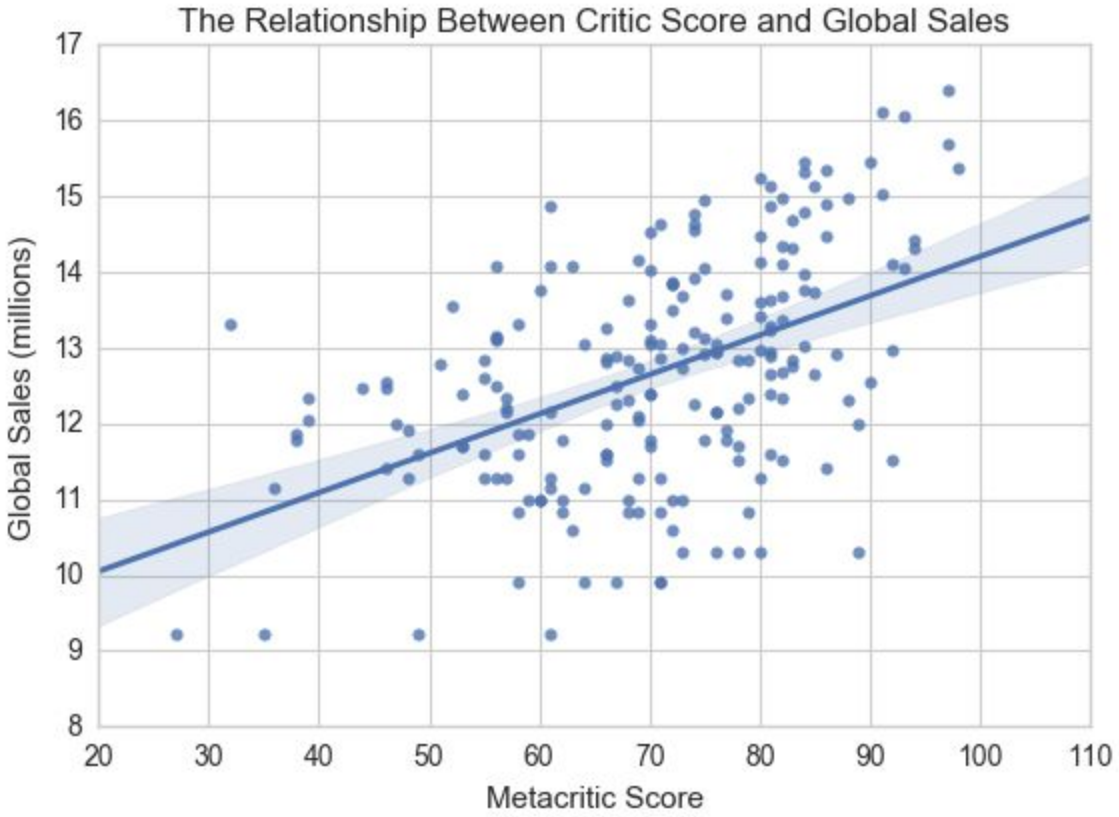


Figure 13: A randomized subset sample of the data was taken (n=200) to show the relationship between the Metacritic Score and Global Sales. There exist a strong and positive correlation of the two variables.

So what is a Metacritic score that a developer should aim for? A dataset without any outliers, where the global sales are more than 3 standard deviations, is considered as a typical game. Global sales will be used a metric of success. The average critic score for the typical game is 70.288 with a standard deviation of 13.90. With a such a large standard deviation doesn't help a developer much because of how spread the scores can be. For games that are widely successful, where the standard deviation of the global sales is more than 3, the average critic score is 87.05 with a standard deviation of 9.077. In contrast to the Metacritic scale, developers do not need to meet a score of 90 or above and a score of 87.05 can be considered successful.

Despite the Metacritic score, there are games that defy the model of higher scores means more money. Games like Just Dance have a Metacritic score of less than 50 and still generated 7.20 million dollars of sales. Games in the 90 percentile of global sales generate 1.7 million dollars in global sales. Despite its low Metacritic score, Just Dance generated almost 7 times more in sales than the threshold of global sales success. On the other hand, IL-2 Sturmovik scored a 91 and only generated $10000 in global sales.

More Releases More Sales?

In the figure shown below, where the number of releases and global sales of that year is plotted together over time. At the beginning portion of the graph by zero, the number of game releases is less than sales. But as time increased, there is an increased in popularity of video games such that there isn't a conservative number of releases being made anymore. This trend is consistent until 2011, where the number of game releases is lower than that of the sales line.

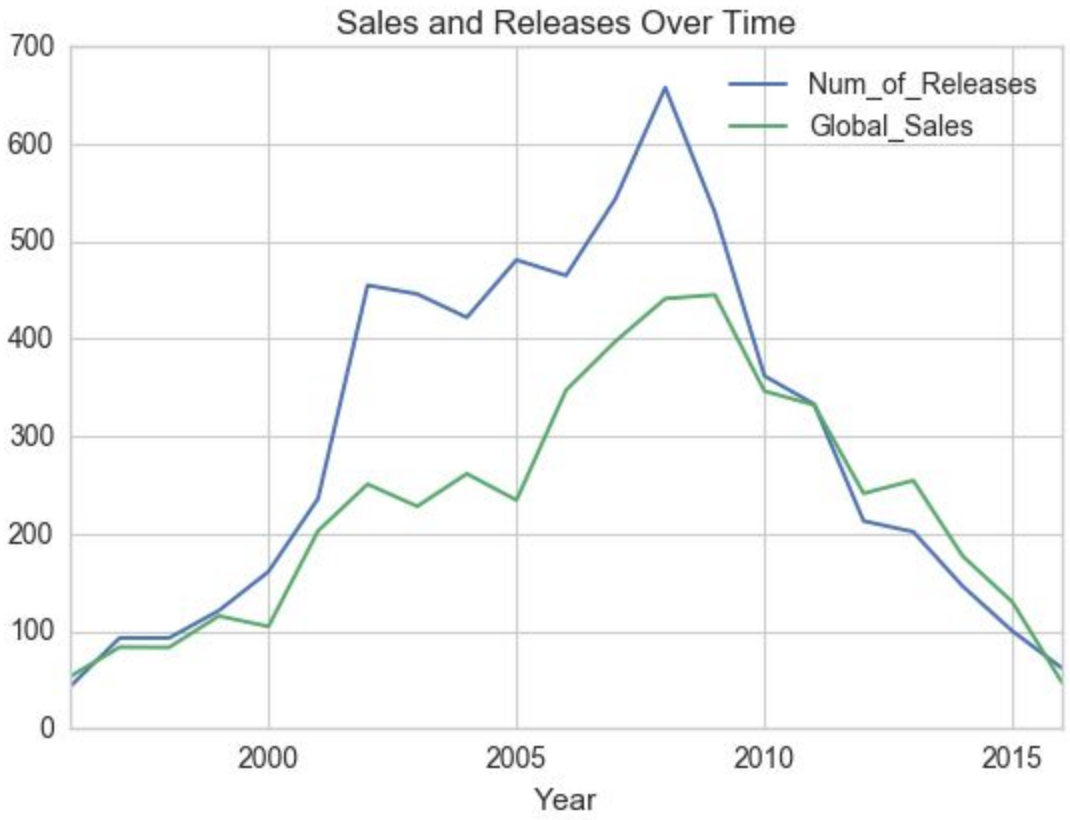


Figure 14: A comparison of the global sales and the number of releases over time. 2008 had the most releases and sales for any year.

At the height of the video game popularity, by its number of releases and global sales, 2008 appears to be the golden year for video games. In 2008, video games have made 485.61 million dollars in sales globally with more than 6166 releases. That year included major hits like Mario Kart Wii, Super Smash Bros. Brawl, GTA IV, Call of Duty World at War, Gears of War 2, Fallout 3 and many others.

In Figure below shows that there is a strong positive correlation between the number of releases and global sales. .An adjusted R-value of 0.726 shows that 72.6% of the number of games released per year can be used to explain the total global sales for that year. If the number of games released per year increased then the global sales will also increase by 0.31% million dollars for that respected year.

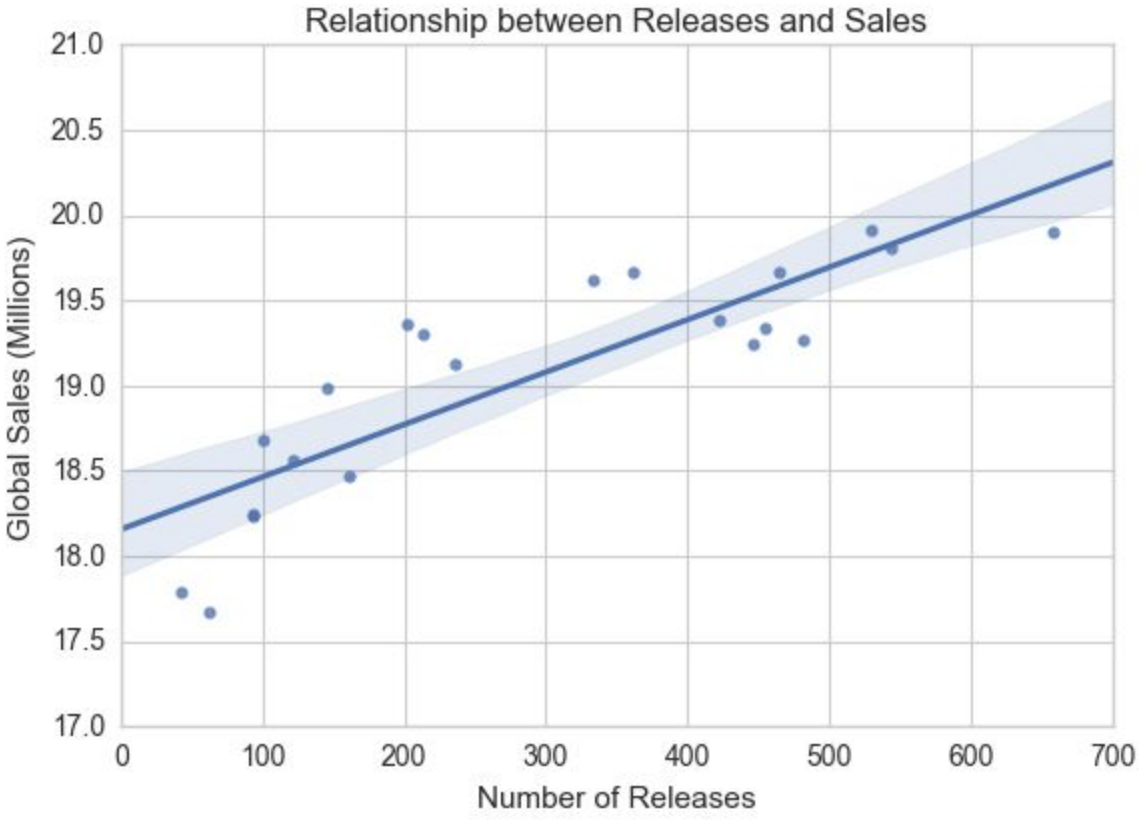


Figure 15: A strong and positive relationship between the two variables.

Limitations

1. When the two datasets were merged on the matching title and platform column, a sizable amount of data on PC games were lossed. Value counts for PC releases were the highest in the IGN dataset at 3370. The second dataset had 1197 value counts for PC releases, but were not the highest. Upon merging the two datasets, only 542 of the PC value counts remained. This may have skewed the data, where PC gaming should have been the most popular platform.

2. Mobile gaming and IPAD gaming also suffered some data loss where it may also impacted how popular the platform may be.

Conclusion

Release Date Recommendation:

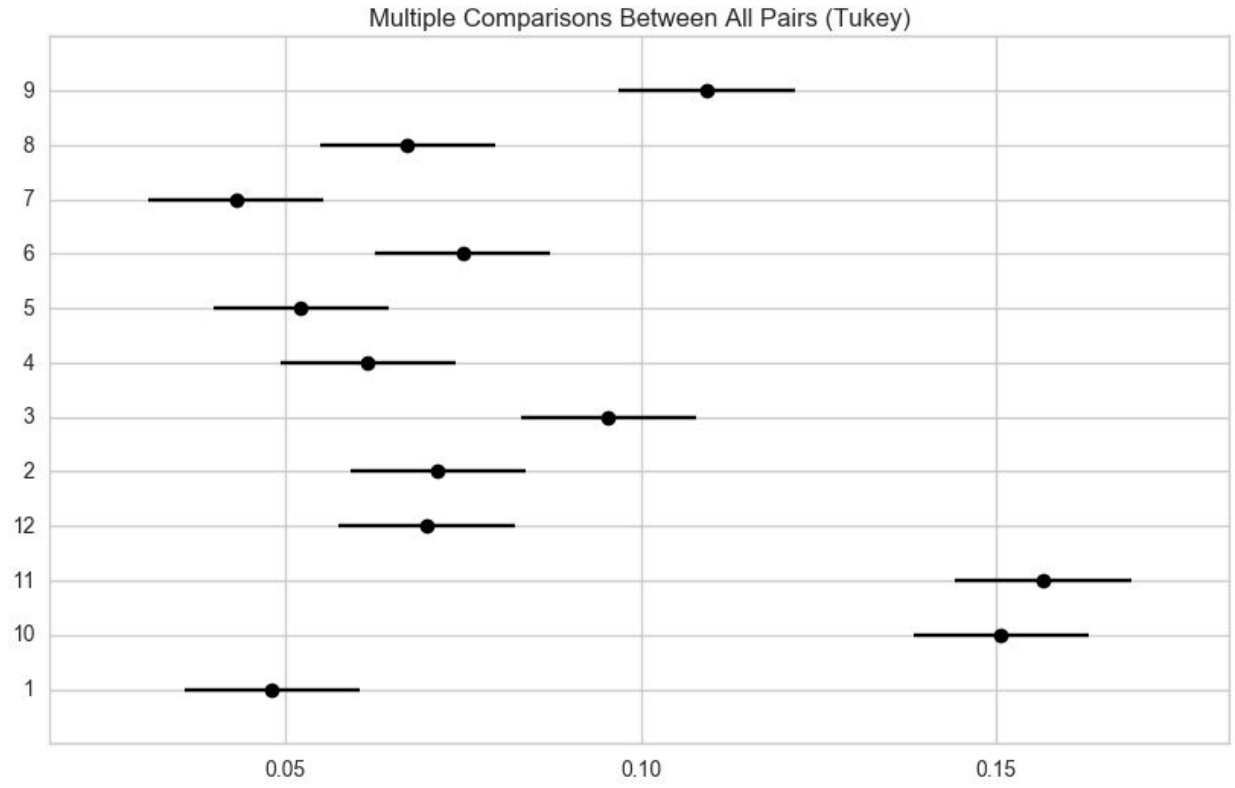
Developers should aim to release their game during the Fall season from the months of September through November. November is the best month for a game release. The month of March should be considered if developers cannot release a game during the fall. Tuesday is the best day of the week for a release despite any month and the weekend should be avoided. Popular release days of the month are the 13th, 11th, and 18th. The 31st and 30th should be avoided.

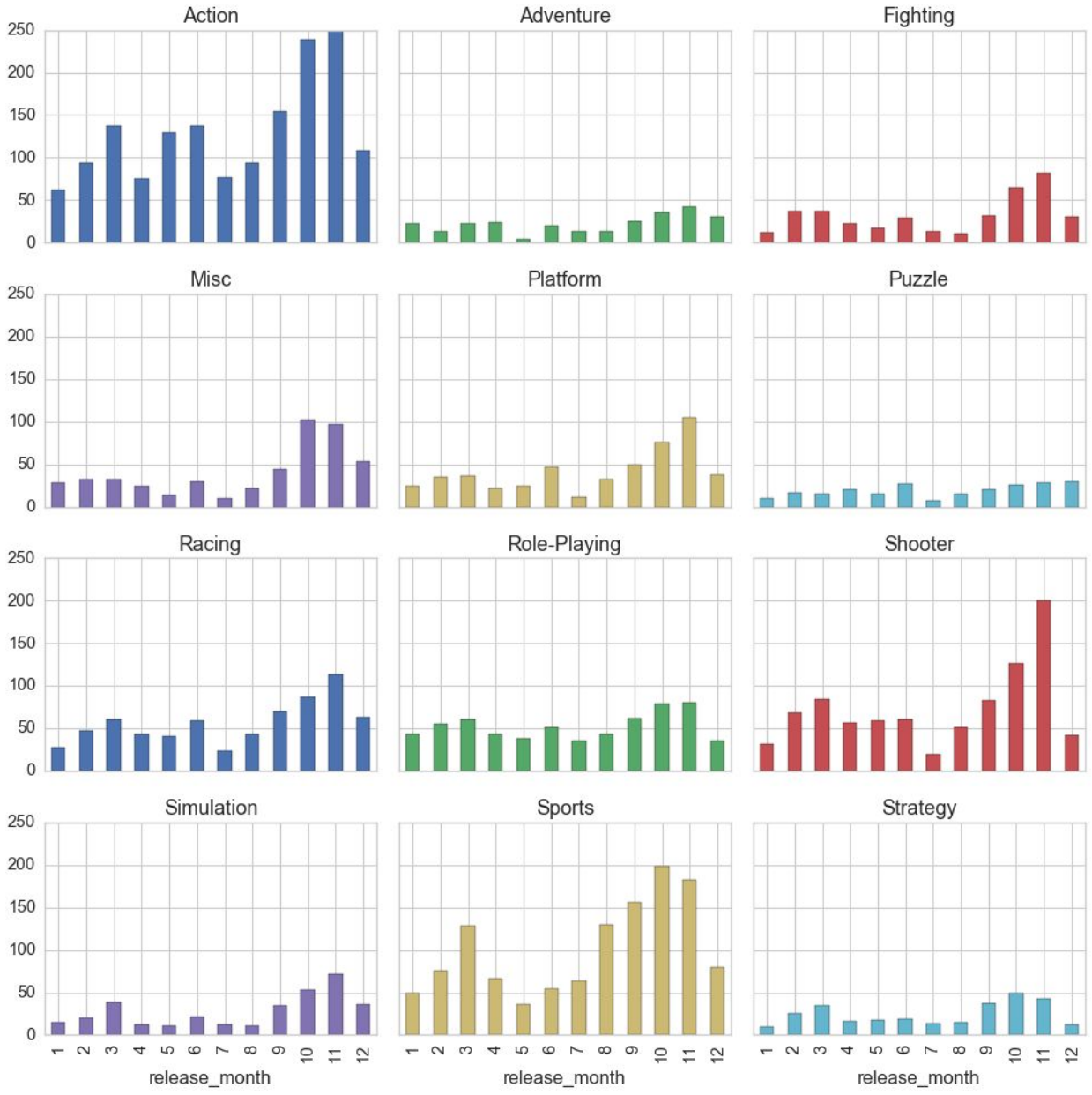
Critic Score Recommendation:

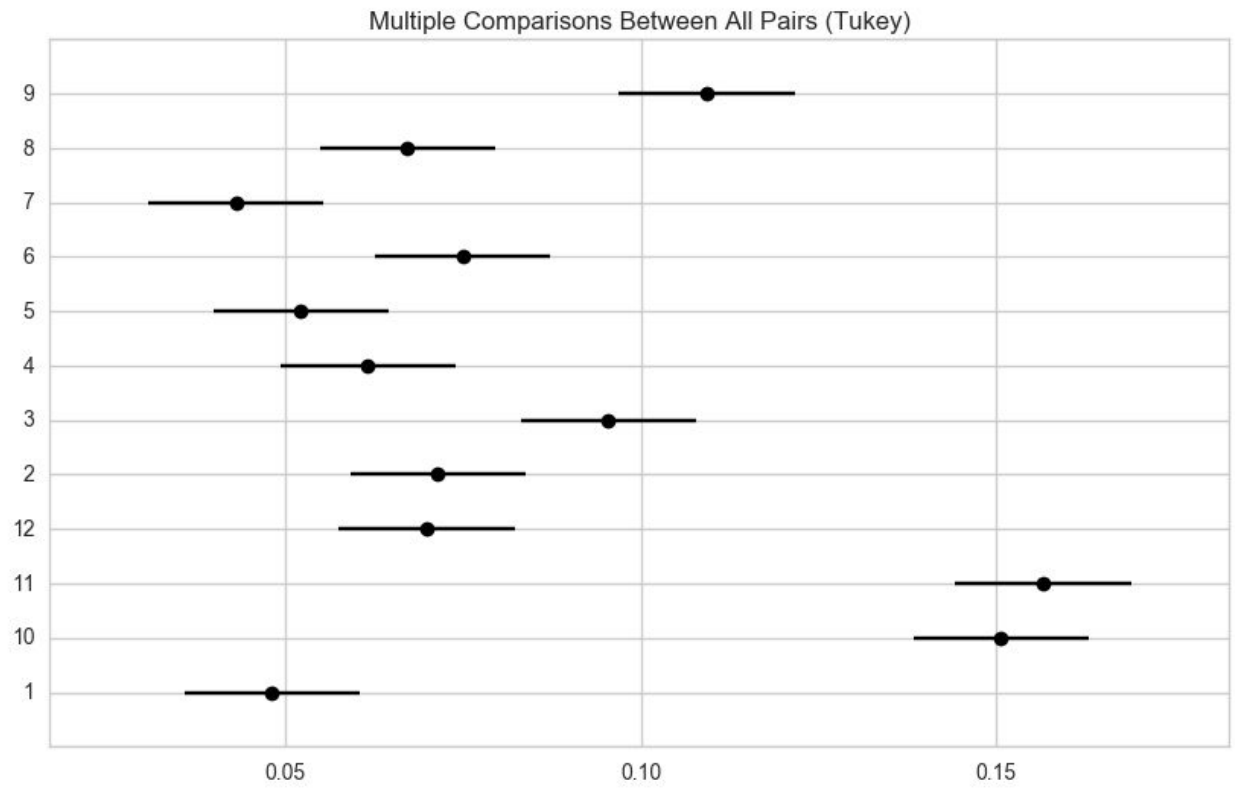
Developers should aim for a 8.72 or higher for IGN and 84.07 Metacritic score or higher in order to be considered widely successful. The scale from these two sources are actually scaled lower than what the data showed.

☆☆☆ ☆☆☆

SUPPLEMENT:


Multiple Comparisons Between All Pairs (Tukey)
9
8
7
6
5
4
3
2
12
11
10
1
0.05
0.10
0.15


Action
Adventure
Fighting
Misc
Platform
Puzzle
Racing
Role-Playing
Shooter
Simulation
Sports
Strategy
250
200
150
100
50
0
1 2 3 4 5 6 7 8 9 10 11 12
release_month


Multiple Comparisons Between All Pairs (Tukey)
9
8
7
6
5
4
3
2
12
11
10
1
0.05
0.10
0.15