

Exploratory Data Analysis

By: Brian Aragon
Video Game Sales 1984-2018

Video games have become a major and evolving industry since it's popularity during the 1980's. With the integration of technology into society, the video game industry has become a champion of integrating arts with graphical technicalities and game mechanics. It has and will continue to grow as new technologies and creative artistic minds continue to produce content, making the industry that is highly profitable across the world.

This report showcases an analysis on video game sales from 1984-2018 for the purpose of gaining valuable insight to manager and stakeholder for decision making. By using exploratory data techniques with the Python programming language and data visualization with Tableau. The data extracted is in an excel format containing roughly 16,600 rows of data from 11 distinguishable features (columns).

Dataset sources used in this analysis can be found on kaggle.com and is also easily accessible through my github page.

Dataset source: <https://www.kaggle.com/gregorut/videogamesales>

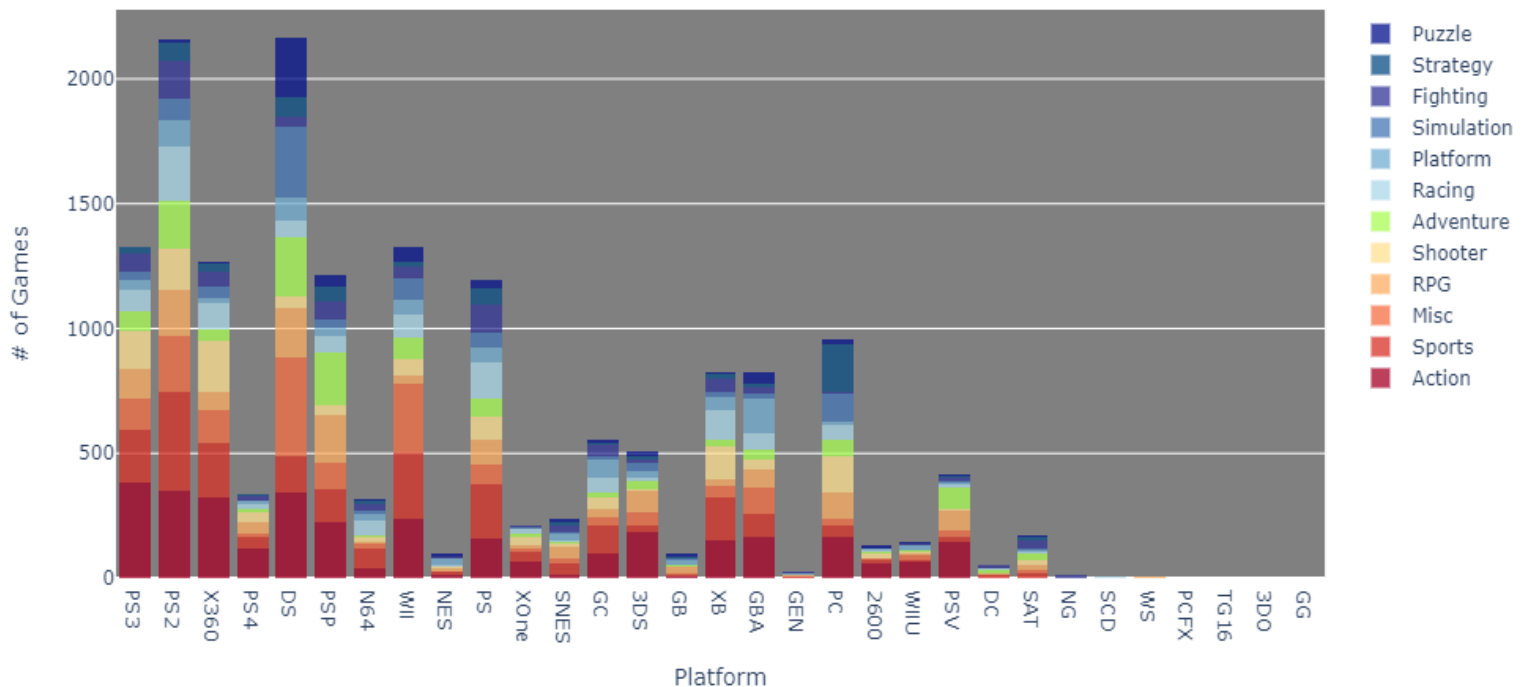
Github Repository: <https://github.com/BA150030-utd/Video-Game-Sales-EDA-Dashboard>

The report will include insight and graphs for:

- Genres and Platforms
- Global Sales According to Platform/Genre
- Number of Games per Genre and Platform
- Genre and Global Sales
- Game Publishers
- Sales of Top 100 Games by Region

Univariate Analysis: Relationships between Genres and Platforms

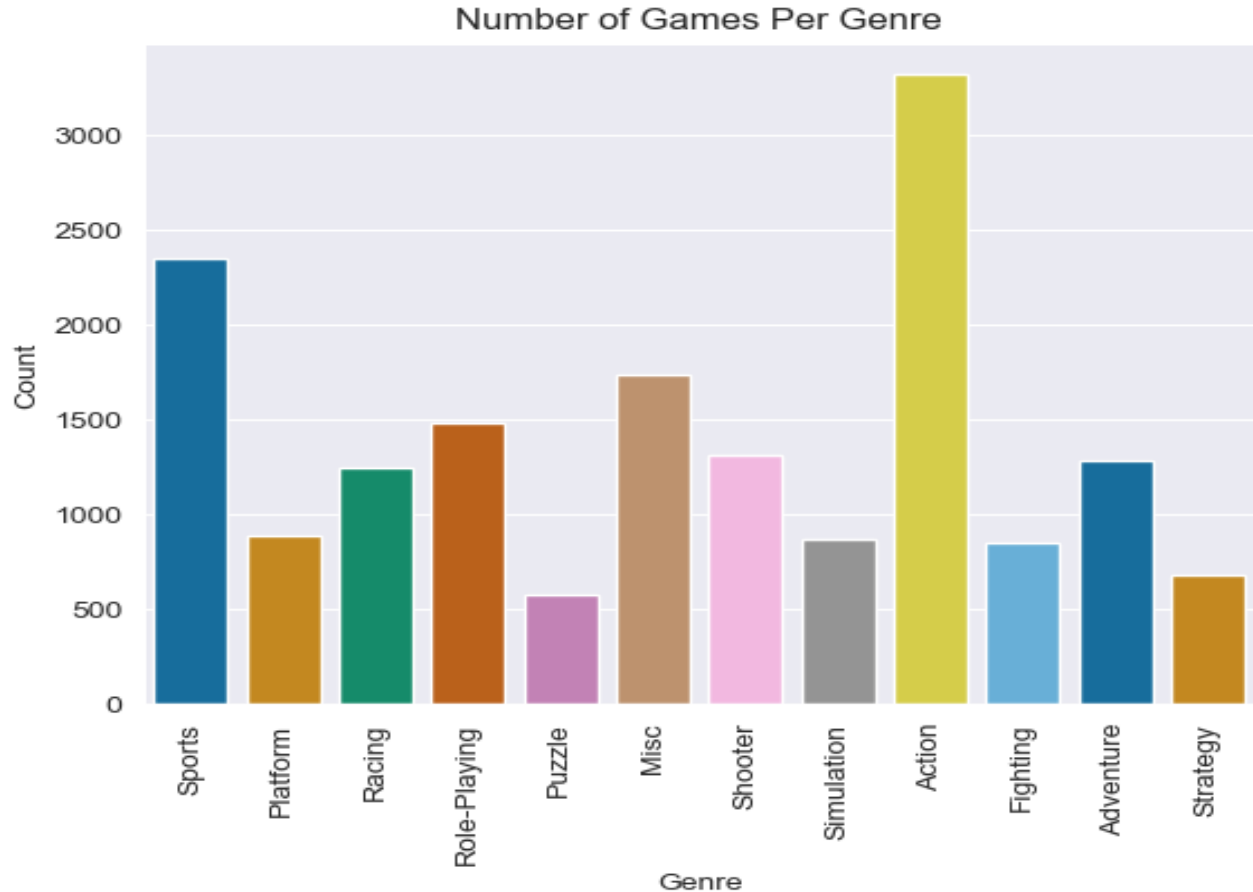
Genre/Platform Relationship



This displays the number of games that particular platform has. From this graph, we can conclude that Nintendo DS and PS2 had the most games available overall. However the histogram also shows that the PS3 and PS2 had the most games in the Action genre compared to the other platforms. PC has the most games in the Strategy Genre then any other platform.

What does this mean?

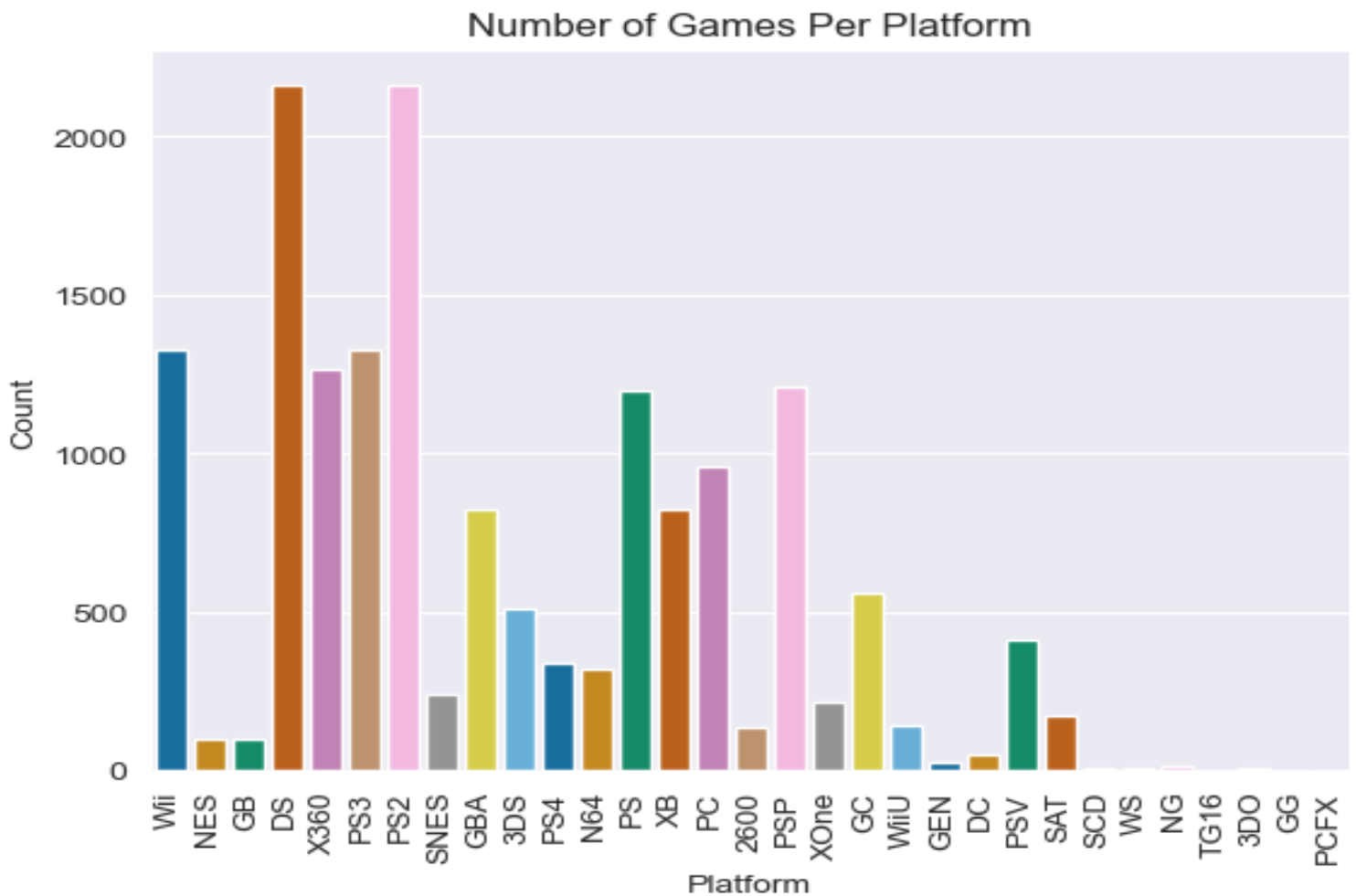
We can make conclusions like these in regards to any other genre on the dataset and compare them to other variables. Using a histogram will be useful because we are dealing with numerical values that can be measured on an interval scale. We can also use histograms to detect any potential outliers.



The highest number of available games by genre would be Action. It has the most games out of any other genre. Sports would fall second, and Misc(indie-games) would come third. This simple bar chart can help the decision of what genre is the most profitable. If a particular genre is popular, we can infer that the demand and demographic are relatively high and managers can make a decision on what type of game they want to invest in.

What does this mean?

This graph displays the number of games each platform had. The PS2 and DS platforms offer the highest variety of games out of any other console. Followed by PS3 and xbox. Out of all of the platforms, I can speculate that the highest variety of games for a platform would generally fall under the Playstation, Xbox, and Nintendo owned platforms.

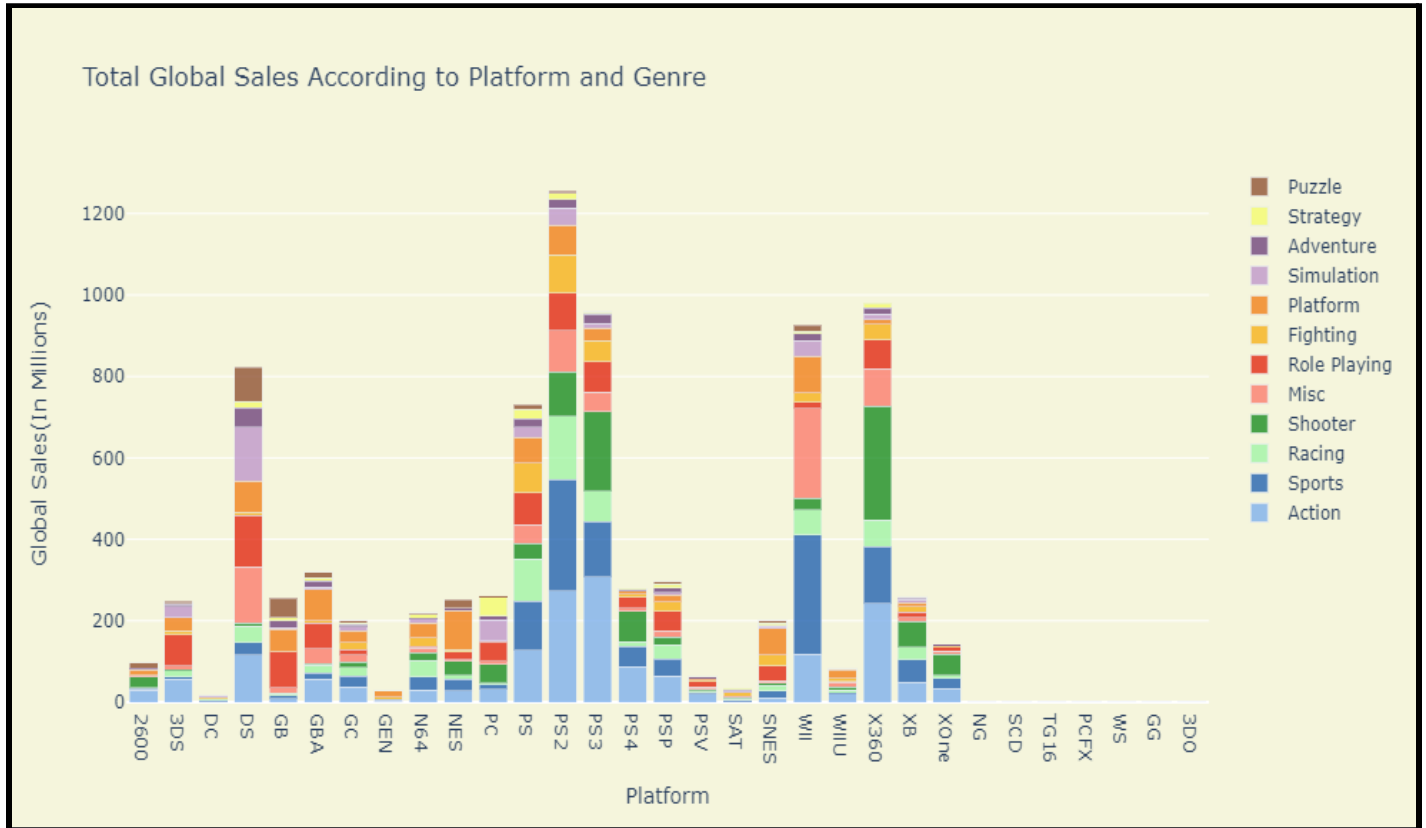


We can see that early platforms did not have much to offer regarding the variety of games. This is because the industry was relatively new and new technology was currently in its infant stage. As a result of technology growing at a rapid rate, video game platforms became obsolete and competition from better hardware took over. However, there are a few ways that companies are able to find a way to get around this issue of outdated hardware.

What does this mean?

Playstation, Microsoft, and Nintendo have created a continued lineup of generational consoles in order to stay ahead of. Sony created playstation, Microsoft created xbox, and Nintendo created handheld and console devices. Having the knowledge of what genre gamers play according to certain demographics helps continue their growth in the video game industry.

Genres and Global Sales

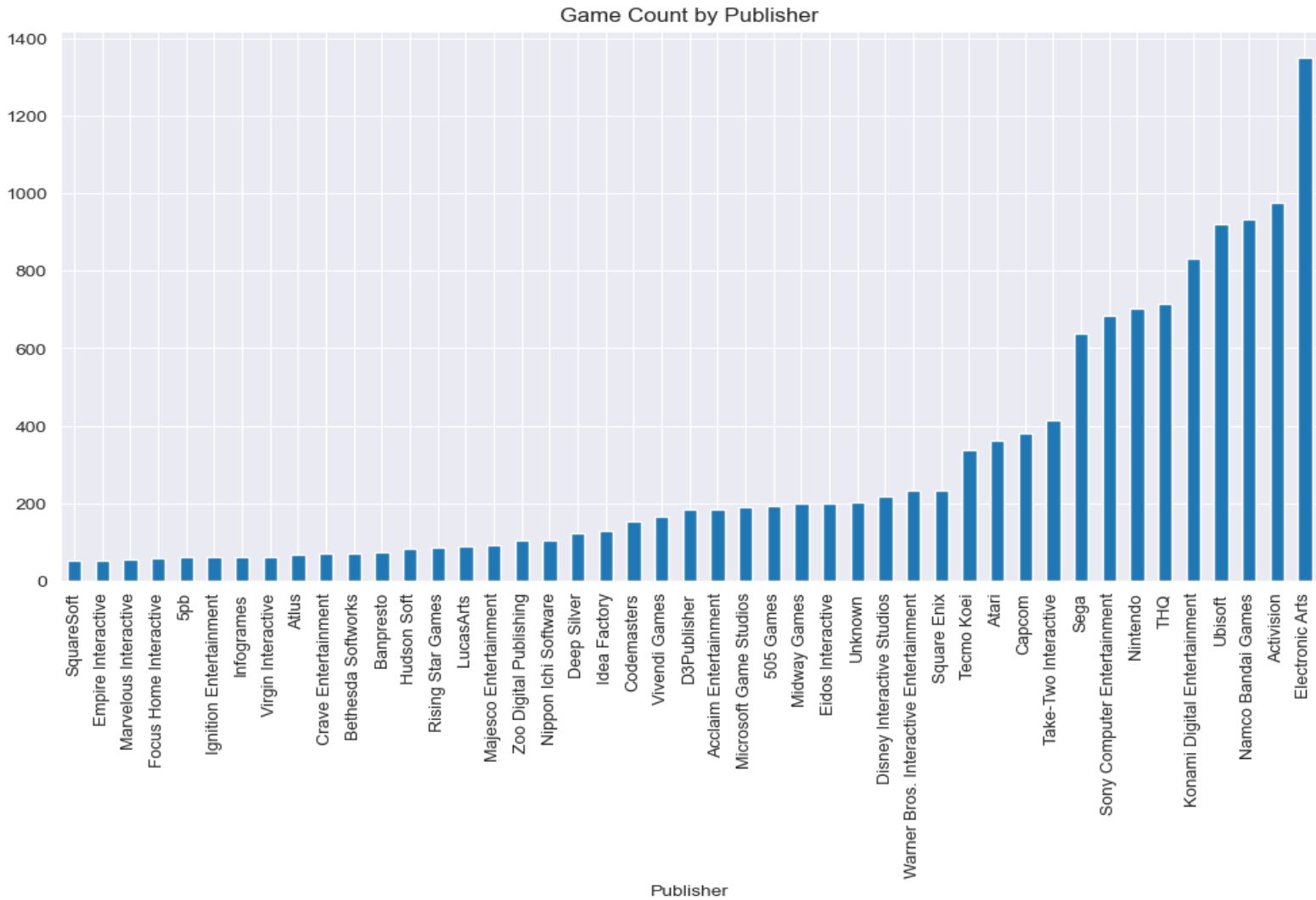


From the graph we can infer that sports and action would result in higher global sales. We can also infer that the genres that sell more on a global scale have been on the playstation, xbox, and Nintendo Wii platforms. Improved generations of these 3 platforms are sold in a couple of years.

What does this mean?

Game development companies can use this information to decide what platform would best suit their game in regards to sales. Companies that create gaming consoles can also use this information to decide which game development company would most compliment overall video game sales. Additionally, companies can design there console to cater towards a specific genre

Game Publishers

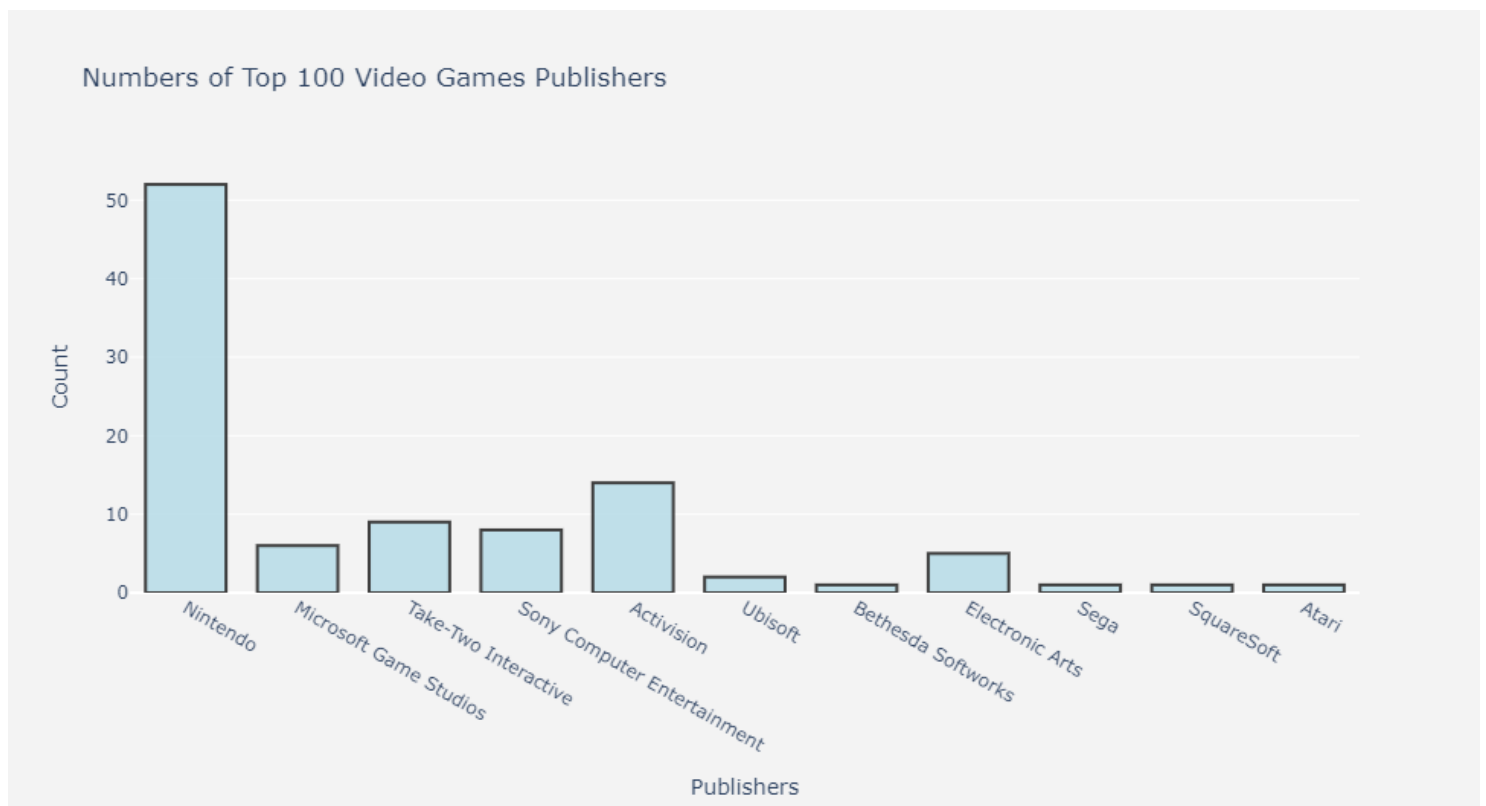


The graph above displays the number of games by publisher. We can see that the top ten publishers in order are:

1. Electronic Arts
2. Activision
3. Namco Bandai Games
4. Ubisoft
5. Konami Digital Entertainment
6. THQ
7. Nintendo
8. Sony
9. Sega

What does this mean?

There is a direct relationship between the number of games a company publishes and game sales. The general rule to consider is the more games a company can publish, the more potential video game sales will increase. A good thing to keep in mind is that publishers influence how a game will turn out as an end product. It is a collaboration between publisher and game developer that results in the launch product of the game. As a result, game developers are more willing to pitch their developers and projects towards more successful publishers such as the top ten i have listed. Return of investment towards game development is relatively higher when working with the top ten publishers. However, this all depends on project management, budgeting and pricing strategies. I will discuss this further in the following data visualization.



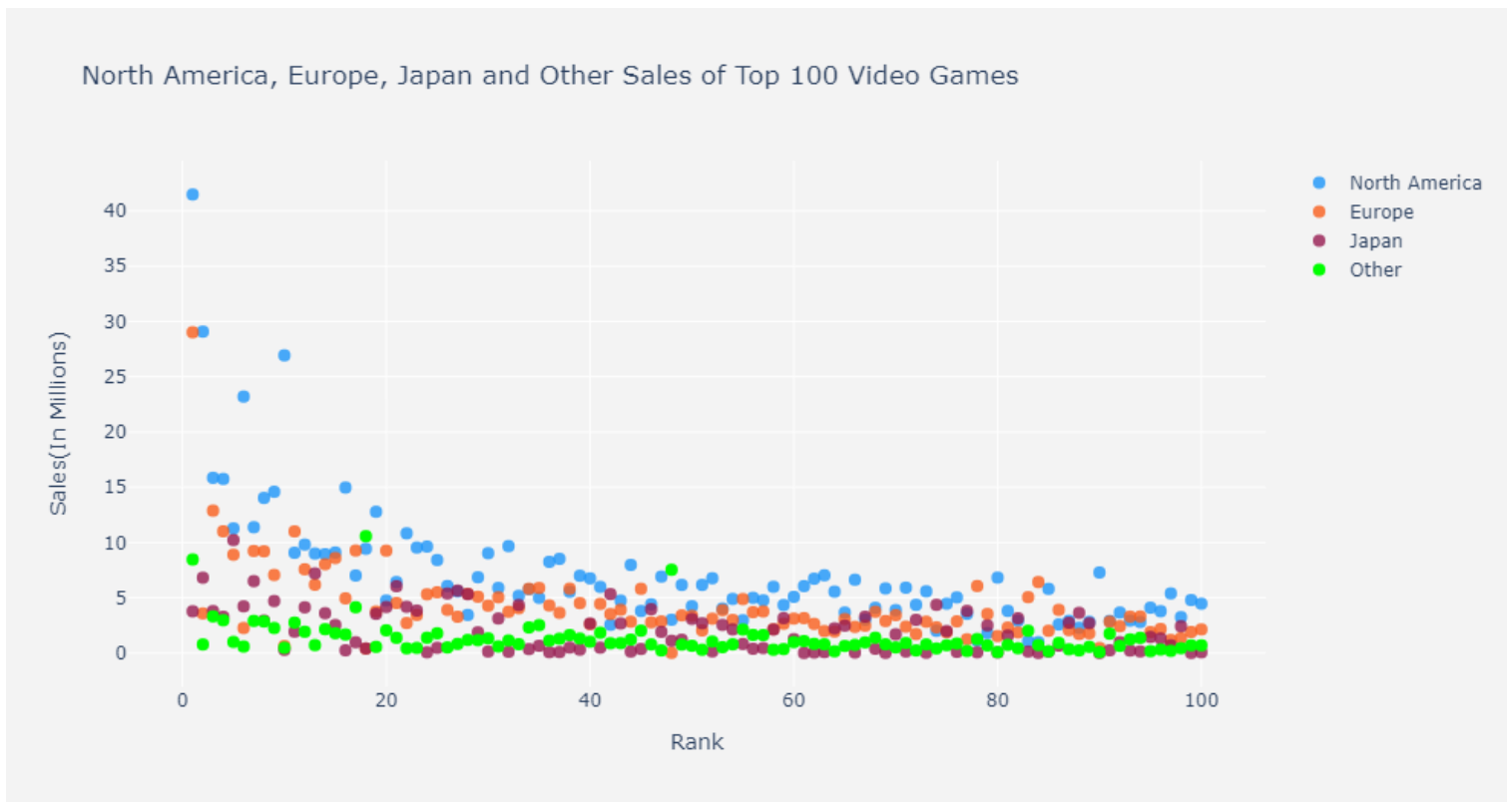
Previously I mentioned that in general, the more a company publishes the more sales a game can potentially get. I also mentioned that there may be some expansion to this. Each game goes through development cycles. This involves different teams, budgets to create the game, marketing, and game development talent. Above is a simple bar chart showcasing the top publishers whose games made it to the top 100 best selling games list. We see that Nintendo dominates, occupying 52% of the games that were best selling. Activision comes second and EA Inc comes 3rd.

What does this mean?

From this information I speculate a three things:

1. This can be due to a consistent pricing strategy.
2. Business focus during development process
3. Quality of game

From the two graphs above, companies like Electronics Arts and Activision published the most games. However, Nintendo was able to generate more sales per game than Electronic Arts and Activision. We can infer that EA and Activision pushed a lot of their efforts in publishing games, but each game they have published and sold were not nearly as successful to Nintendo in terms of video game profits.

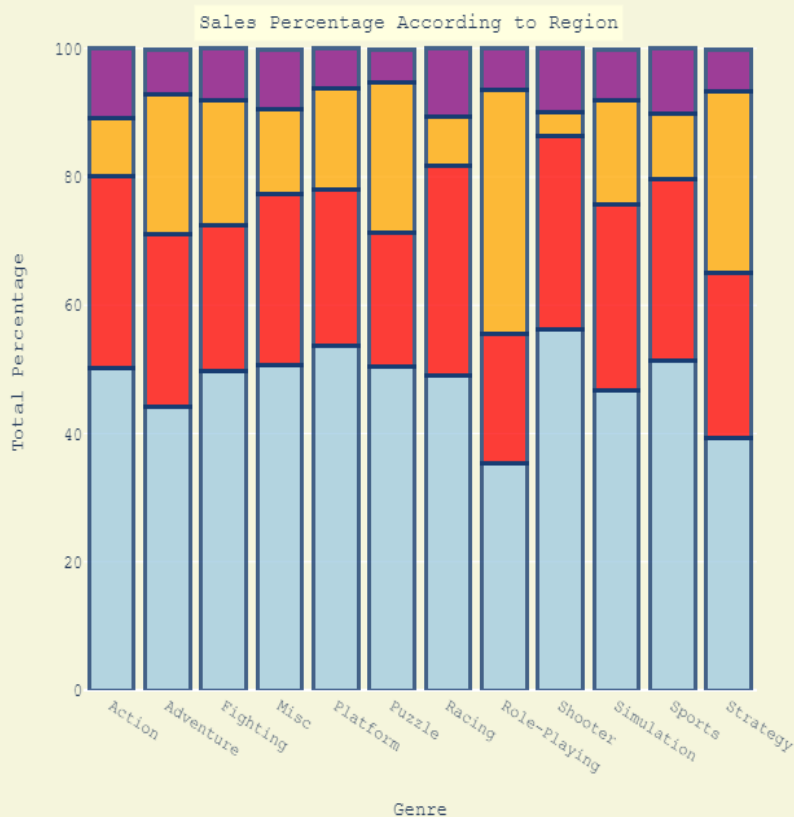


We now take the top 100 games and evaluate how well they sold in each region. From the graph we can infer that North America generally had the highest sales in regards to the top 100 games sold. Europe being second, Japan third, and other regions last in fourth.

What does this mean?

If we are talking about a positive return of investment, we can say that North America would be our best choice. Selling games in North America will create more profit in relation to other regions. Additionally, we previously inferred that action games were very popular in this region. Knowing the genre of game people to develop and understanding the sales demographic can help game developers and publishers grow their sales.

North America, Europe, Japan and Other Sales Percentage and Amounts According to Genre



Total Sales(In Millions)

Genre	America	Europe	Japan	Other Sales	Total
Action	878	525	160	187	1751
Adventure	106	64	52	17	239
Fighting	224	101	87	37	449
Misc	410	216	108	75	810
Platform	447	202	131	52	831
Puzzle	124	51	57	13	245
Racing	359	238	57	77	732
Role-Playi	327	188	352	60	927
Shooter	583	313	38	103	1037
Simulation	183	113	64	32	392
Sports	683	377	135	135	1331
Strategy	69	45	49	11	175

Other Sales Japan Europe America

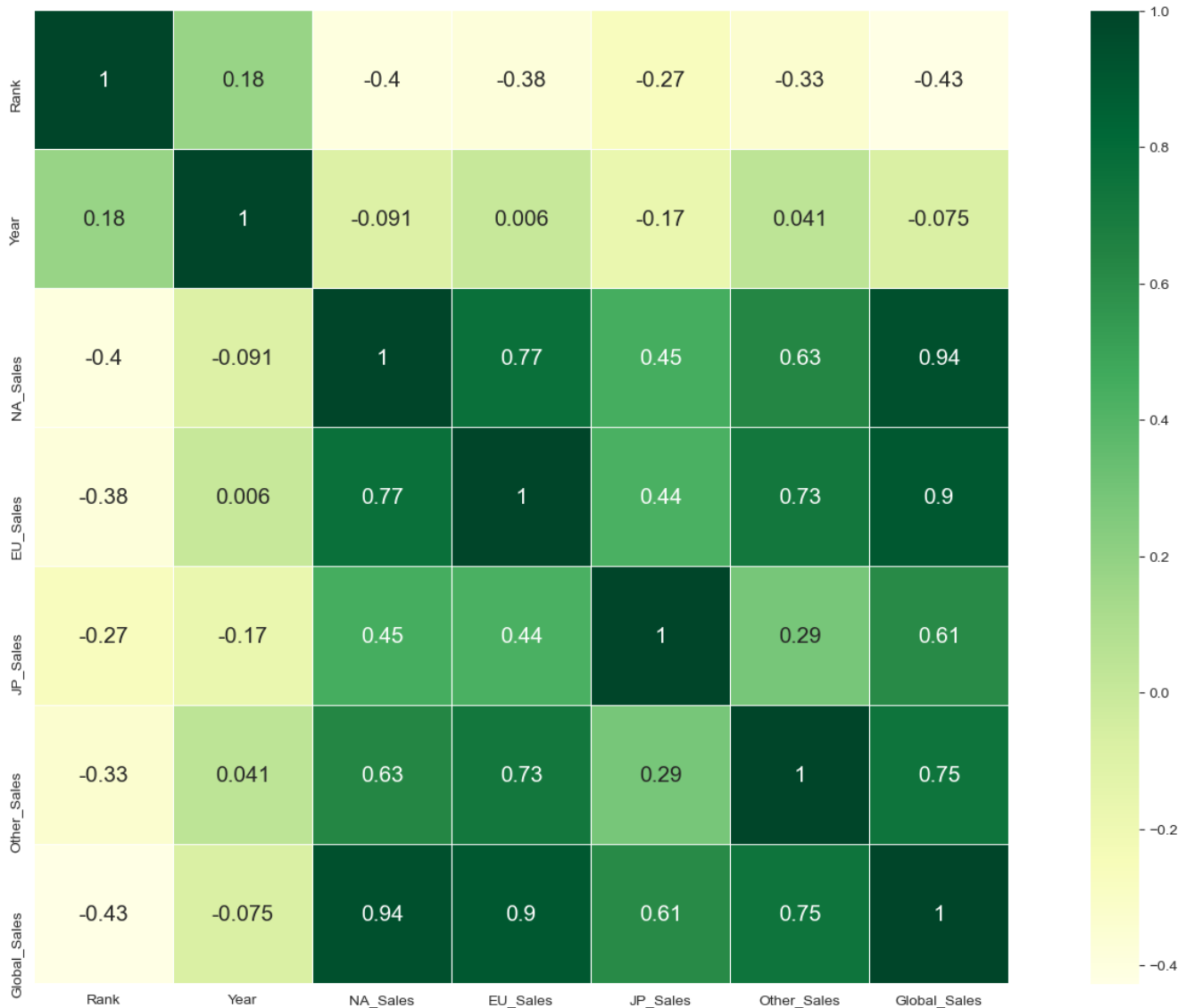
This displays the sales percentage of each genre according to region. This supports my previous claim about North America being the best place to market new games. We can also forecast the type of game genre that could maximize a return of investment. In this case, it is still in the action and shooter genres.

What does this mean?

North America is generally the best market regardless of the type of genre the game is. We see that across all genres, North America does better in relation to other regions. In terms of creating the most profit however, action and shooters will be the most promising. A good thing to note is that some of these genres are not strictly exclusive to one another. Game development is a creative industry, so you will tend to see hybrids of two or more genres, e.g action-shooter or RPG-shooter. Hybrids of genres reach out to a larger audience who may have interests in a variety of genres.

Multivariate Data Visualization

Correlation of Features



The closest relationship I was able to find was the correlation between Global Sales and the Sales of each region. To simply this matrix, the sales of each region will affect the global sales (common sense). We see correlation of more than 0.6 and higher between these features. All other features have low correlation scores, for instance the Sales between America and Japan

What does this mean?

There is not much correlation between the two. This could potentially mean that the market for video games could mostly be unique from one another. What games to market and how to market in North America could be less or more successful and profitable in Japan. Vice Versa.

