

COVER PAGE

PRACTICAL ASSIGNMENT

Subject	UCCC3073 Data Science
Trimester	MAY 2021
Lecturer	Dr. Tong Dong Ling (tongdl@utar.edu.my)
Submission Date	(Monday, 16 August 2021, 5 p.m.)

Individual Information:

Name	Student ID	Programme
1) LEE WAI HIN	18ACB02736	CS
2) LEW JUN XIAN	18ACB04184	CS
3) LEE JAC SERN	18ACB05162	CS

Course Learning Outcomes Assessed:

CLO2	Assess and manipulate data sources from various online sources.
------	---

Assignment Criteria	Total Marks	Given Marks
Part 1	25	
Part 2	20	
Part 3	25	
Total	70	
Final	20%	

Task 1: Data Summary

1.1 Problems / Challenges

A video game is an electronic game that can be enjoyed and played on a computing device, such as a personal computer, gaming console or mobile phone. Video games are a medium of entertainment that comes in a variety of different genres. The video games industry has grown quite considerably since the first games emerged in the 1970s. [It was estimated that the global gaming market pulled in a revenue of 175.8 billion USD in 2021, and this figure is projected to increase to 268.8 billion USD annually in 2025.](#)

The ongoing global pandemic of COVID–19 has caused the entire global population considerable amount of stress and anxiety. Gaming has become a way of escaping our arduous reality and de – stress. As such, the dataset that has been chosen as the basis of our study is “Video Game Sales”. The reason behind the selection of this dataset is that we would like to study the factors that affect the sales of various video game titles. In addition to that, the exploration and analysis of this dataset would allow us to gain relevant insights that may be used to help business executives of companies in the industry to boost their sales and increase revenue.

In short, the challenges that require solving regarding this dataset is the exploration of which region of the world has the highest video game sales as the population of different regions differ in spending habits. In addition to that, we would also like to understand how the genre of video games affect its sales across different regions of the world as it gives us a better understanding on what titles should be marketed more heavily in different regions to increase their sales.

In conclusion, we hope to gain enough relevant insight and knowledge from the exploration of this dataset, which could be used as a foundation to construct reliable recommendations that business executives in the industry could refer to.

1.2 Nature of Dataset

This dataset was published on Kaggle by Gregory Smith. The dataset contains a total of 11 attributes and 16, 598 records. There are a total of 5 categorical variables and 6 numerical variables. The categorical variables can be sub – divided into 4 nominal variables and 1 ordinal variable, while the numerical variables can be sub – divided into 5 ratio variables and 1 interval variable.

Attribute Name	Description	Example	Data Type	No. of Missing Values
Rank	The overall ranking of a title based on the total amount of sales .	1	Ordinal	–
Name	The title of the video game .	Tetris	Nominal	–
Platform	<p>The platform that the video game was released on.</p> <p>Notable platforms include:</p> <ul style="list-style-type: none"> • PS3: PlayStation 3 – a home video game console developed by Sony to serve as the successor to the PlayStation 2. • GBA: Game Boy Advance – a 32 – bit handheld console that was developed by Nintendo as a successor to the Game Boy Colour. • DS: Nintendo Dual Screen – a handheld console that features 2 screens which work together in tandem. • X360: Xbox 360 – a home video game console developed by Microsoft as the successor to the original Xbox. • NES: Nintendo Entertainment System – an 8 – bit video game console produced by Nintendo. 	PS3	Nominal	–

Year	The release year of the video game.	2009	Interval	271
Genre	<p>The genre of the video game.</p> <p>Notable genres include:</p> <ul style="list-style-type: none"> • Action: Games with a heavy emphasis on physical challenges that require hand – eye coordination and motor skill to overcome. • Shooter: Games with a heavy emphasis on ranged gameplay involving ranged weapons. • Sports: Games that emulate the feeling of traditional physical sports. • Platform: Games with gameplay primarily centered around jumping and climbing to navigate a level. • Fighting: Games that are centered around close – range combat typically one – on – one fights. 	Puzzle	Nominal	–
Publisher	The publishing company that brought the game to market .	Nintendo	Nominal	58
NA_Sales	Units sold in North America (in millions).	14.03	Ratio	–
EU_Sales	Units sold in Europe (in millions).	9.20	Ratio	–
JP_Sales	Units sold in Japan (in millions).	2.93	Ratio	–
Other_Sales	Units sold in the rest of the world (in millions).	2.85	Ratio	–
Global_Sales	Total units sold worldwide (in millions).	29.02	Ratio	–

Table 1.2 Properties of the Dataset

1.3 Questions to be Answered about this Dataset

1. Which platform has the most games released for it?
2. Which publisher has the highest sales worldwide?
3. Which genre of games are the most well received?
4. Which region generated the highest number of sales?
5. What genre of games do each region in the world favour?
6. Which platform's titles generated the highest number of sales?
7. Which year had the highest number of game releases?
8. Which year had the highest sales worldwide?

(490 words)

Part 1: Marking Rubric (25 marks)

Criteria	5 – 4 marks Excellent	3 marks Good	2 – 0 marks Unsatisfactory	Total
Data set (5 marks)	The selected data set has met all the requirements outlined in the assignment.	The selected data set has met most of the requirements outlined in the assignment but has some minor flaws.	The selected data set is not appropriate for the assignment.	
Description (5 marks)	The data set is clearly and thoroughly explained.	The data set is partially explained, leaving some important information out (ambiguity)	The description is mainly brief and leaving doubts to reader.	
Content (5 marks)	All questions are meaningful and are answerable from the data set.	At least two meaningful questions have been outlined and they are answerable from the data set.	Only one meaningful question has been outlined.	
Writing (5 marks)	A well-written and well-structured report with proper use of grammar, punctuation and spelling.	A relatively well-written report with some minor flaws on grammar, punctuation and spelling errors that do not impede overall understanding.	A fair-written report with major flaws on grammar, punctuation and spelling errors that may disrupt reader's understanding.	
Overall Understanding (5 marks)	Evidence of some extensive independent study to understand the data set.	Some good understanding on the chosen data set evident with a relatively detailed explanation on the data set.	Some evidence of an understanding on the chosen data set.	
TOTAL				

Task 2: Initial Analysis

2.1 Lead Sentence

Relax and have fun with beloved games from the past. What titles did people enjoy and buy across the globe? What genres appealed to the widest range of individuals?

2.2 Summary of Lead Sentence

Video games have been a source of immense joy for multiple generations and will continue to be an integral part of life for many people worldwide. The video games industry has been elevated to a whole new level due to various advancements in technology, artistic creativity, and development tools. Developers can now create video games that can also serve as modern art pieces. Games have become a canvas that developers are able to populate with detailed environments, vivid characters, and great music. Therefore, it would come to no surprise that various publishing companies are competing with one another to secure the rights to up – and – coming titles that seem interesting and has the potential to captivate gamers.

An interesting finding that we have discovered from this dataset about video game sales is that the games that have been published by Nintendo have sold the most units in total worldwide over the years. This could be due to the fact that Nintendo is a family – oriented company whose games' primary focus is fun across all ages. Parents would be more inclined to purchase titles that are from Nintendo for their children if it is family – friendly and does not contain any mature themes, such as violence, and sex. These games allow parents to bond with their children over something fun and engaging. Thus, it is recommended for publishing companies to secure the rights and invest in the development of fun games that appeals to every demographic and is easily accessible. This would help companies increase their yearly revenue as family – friendly titles would bring in more customers.

In addition to that, it has also been discovered through the analysis of this dataset that publishers with the rights to well established series tend to rank very high in terms of sales, such as Electronic Arts, Activision and Take – Two Interactive. These established series usually have a very large following. Fans of these series are willing to purchase the latest instalment in the series whenever it is released. As such, publishing companies could invest more heavily into the development of games in established universes, such as Grand Theft Auto, Call of Duty, and FIFA, instead of original properties that do not have an existing fanbase.

Besides that, it has also been discovered that titles in the ‘Action’ genre tends to perform the best when it comes to sales worldwide. Games in this genre places a heavy emphasis on physical challenges that require hand – eye coordination and motor skill to overcome. These challenges give players a sense of pride and accomplishment when it is solved or conquered. Thus, it would be profitable to development studios and publishing companies to direct more resources into the development of interesting games in the ‘action’ genre with novel challenges and concepts.

In conclusion, the exploration and analysis of this dataset allowed us to gain enough relevant insight and knowledge into the factors affecting the sales of a video game title. This insight and knowledge allow us to make educated predictions and recommendations that might be of interest of executives in the video games industry.

2.3 Data Visualisation

2.3.1 Visualization 1

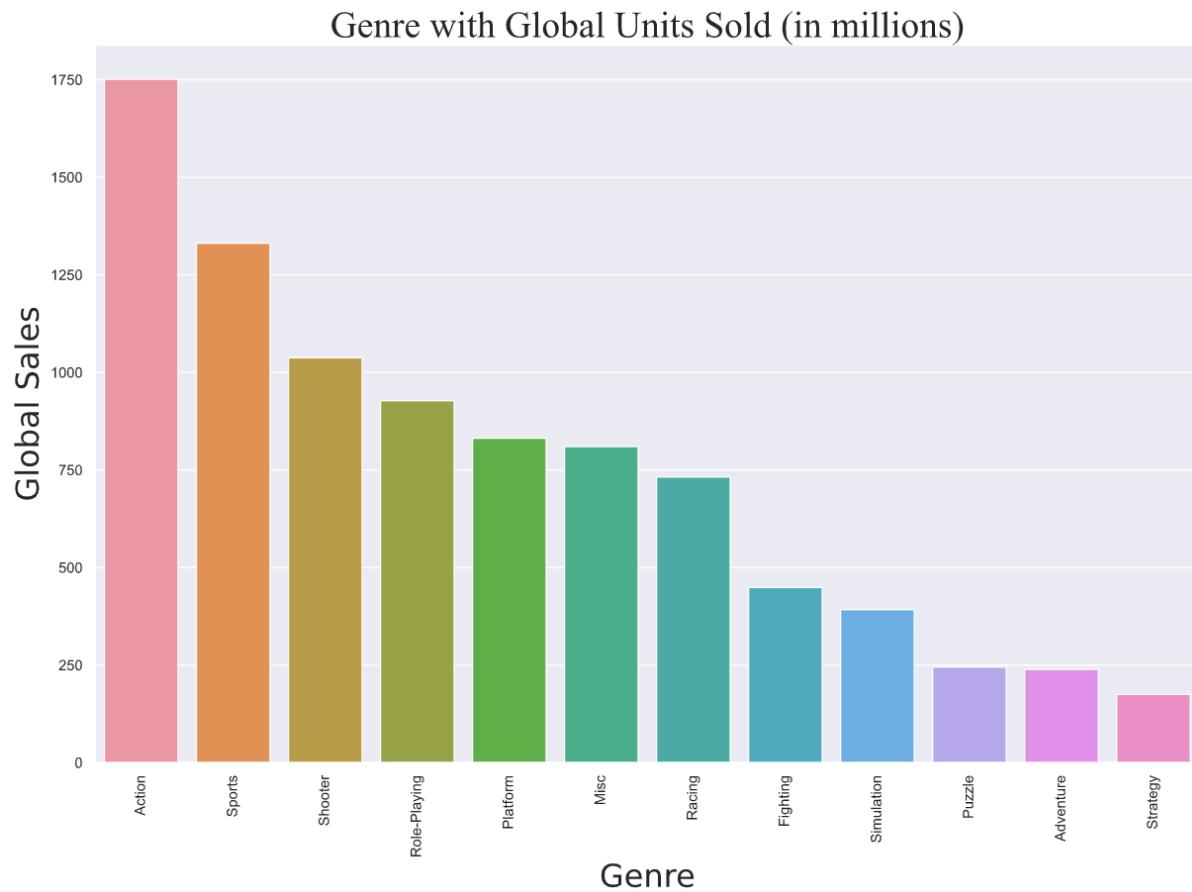


Figure 2.3.1 Genre with Global Units Sold (in millions)

The figure above illustrates the 10 publishers with the highest number of video game copies sold worldwide. Titles published by Nintendo sold the greatest number of copies over the years, with a total of 1.79 billion copies. Moreover, publishers with the rights to well established series tend to rank very high in terms of sales, such as Electronic Arts at 2nd place with 1.11 billion copies sold, Activision at 3rd place with 727 million copies sold, and Take – Two Interactive at 6th place with 399 million copies sold. Sequels to established series have the potential to bring in a lot of revenue.

Key Takeaway: Consider releasing video games in established universes that already have an existing following to ensure a return on your investment.

2.3.2 Visualization 2

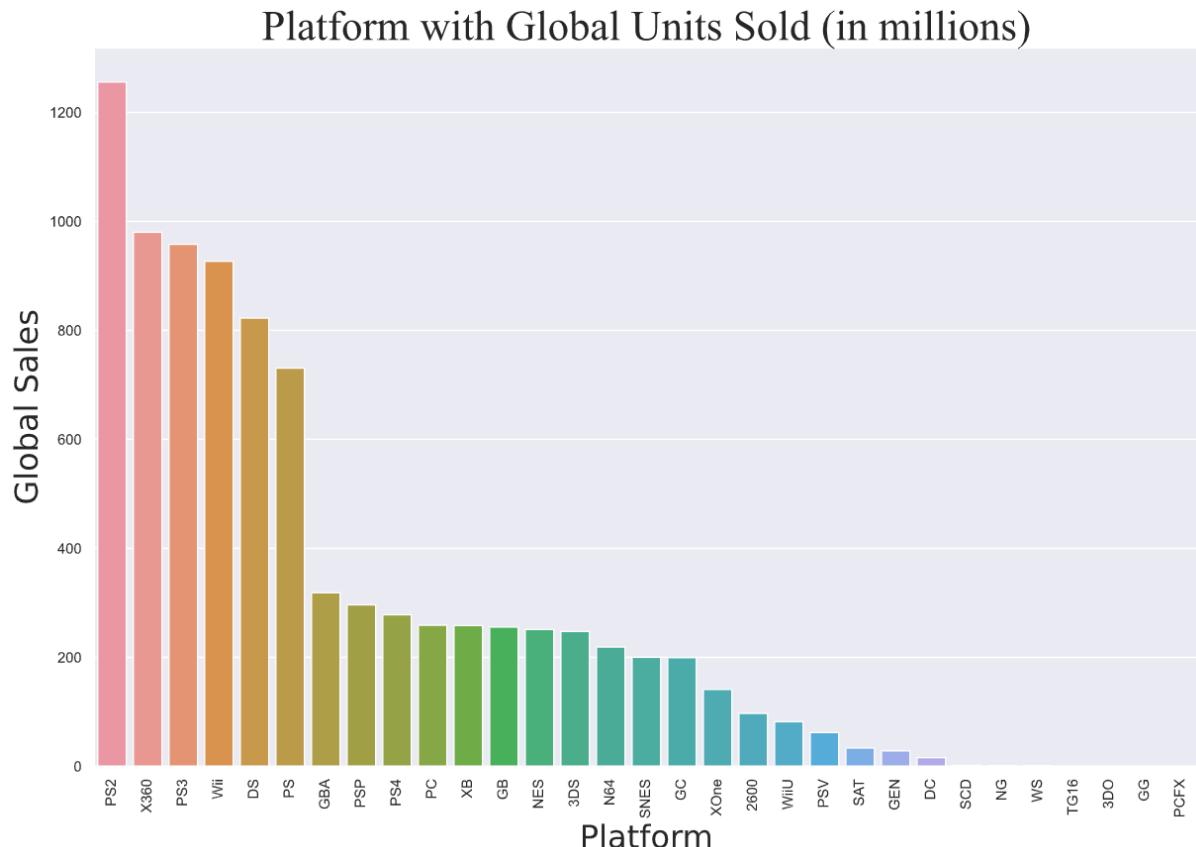


Figure 2.3.2 Platform with Highest Global Sales (units sold in millions)

The figure above illustrates the total number of units sold worldwide across all titles that have been released have been released for a particular console platform. It comes to no surprise that the titles that have been released for the PS2 sold the most copies, with a total of roughly 1.26 billion units. The 2nd and 3rd being the X360 and PS3 with a total of roughly 980 million and 958 million copies, respectively. The PS2 was the preferred console during its generation, while X360 and PS3 were on relatively equal footing during that generation. This situation is what led to PS2 titles selling the most copies worldwide, while sales for X360 and PS3 titles were on relatively equal footing.

Key Takeaway: Consider releasing video games for consoles that are more popular to increase the number of video game copies sold.

2.3.3 Visualization 3

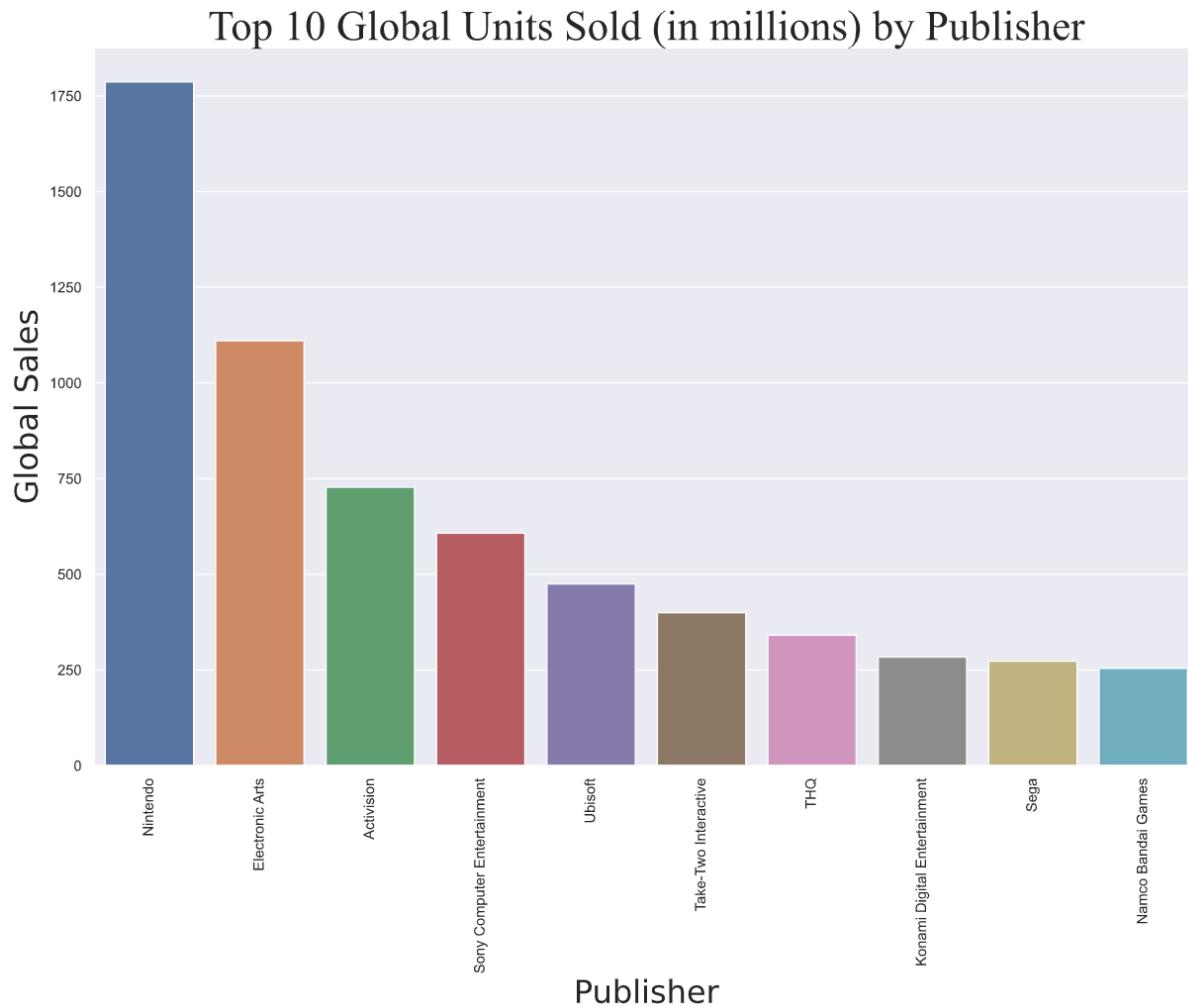


Figure 2.3.3 Top 10 Global Units sold (in millions) by Publisher

The figure above illustrates the total number of units sold for each genre of video games worldwide. The genre of video games that has the highest number of units sold worldwide is ‘action’, with 1.75 billion units. The 2nd highest being ‘sports’ with 1.33 billion units, 3rd being ‘shooter’ with 1.04 billion units and the last place being ‘strategy’ with 175 million units sold. If a company wishes to release a successful game that is popular, they should release games in popular genres as more people are interested in these genres and are more willing to buy a game.

Key Takeaway: Consider releasing video games in popular genres as more people would be interested the game and it is more likely sell.

2.3.4 Visualization 4

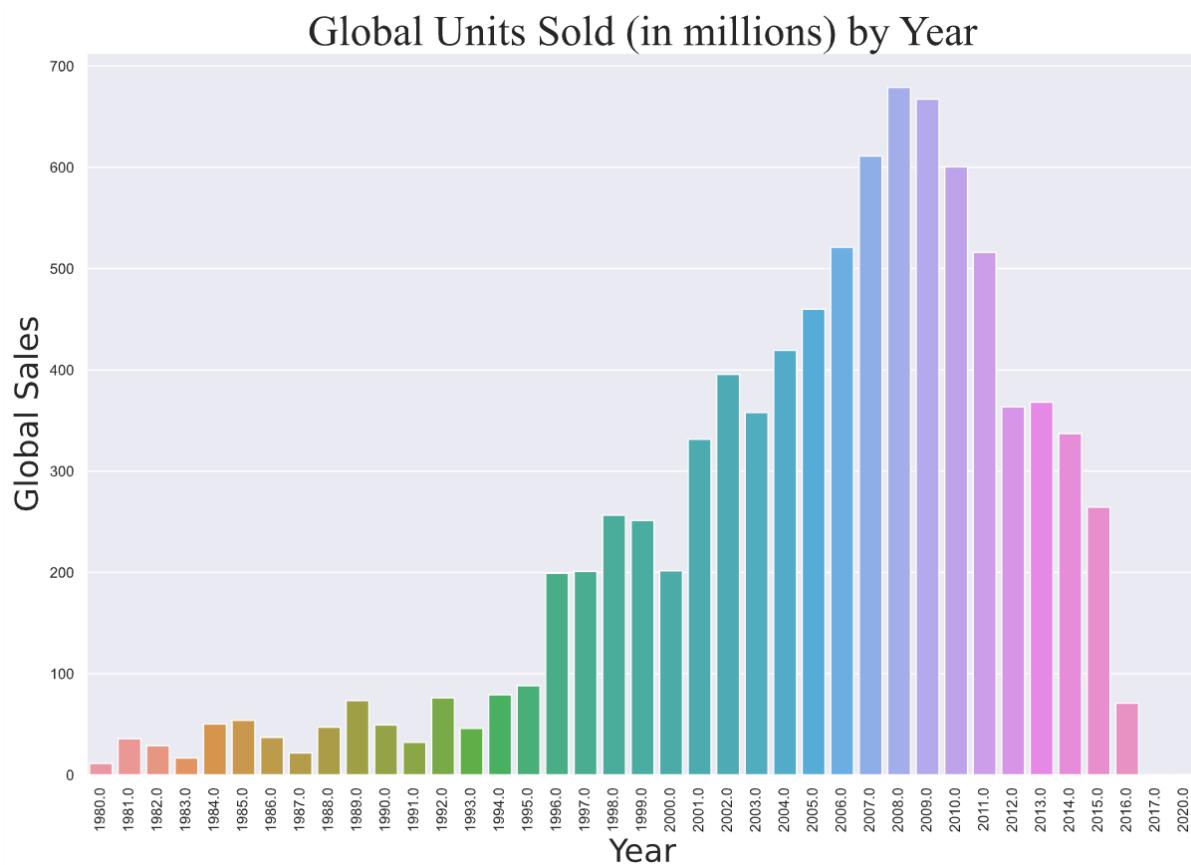


Figure 2.3.4 Global Units Sold (in millions) by Year

The figure above illustrates the total amount of units sold worldwide for any given year from 1980 – 2020. From the figure, it could be seen that the peak year for video game sales worldwide is 2008, which peaked at 679 million units. The primary factor that led to a peak in video game sales worldwide is the release of new consoles by Sony, Microsoft, and Nintendo. Customers who purchased the latest console would be looking to fully experience the capabilities of the console and purchase games specifically made for the console.

Key Takeaway: Consider releasing video games for the latest console platforms as players would be looking to purchase and experience video games on the latest console platform.

2.3.5 Visualization 5

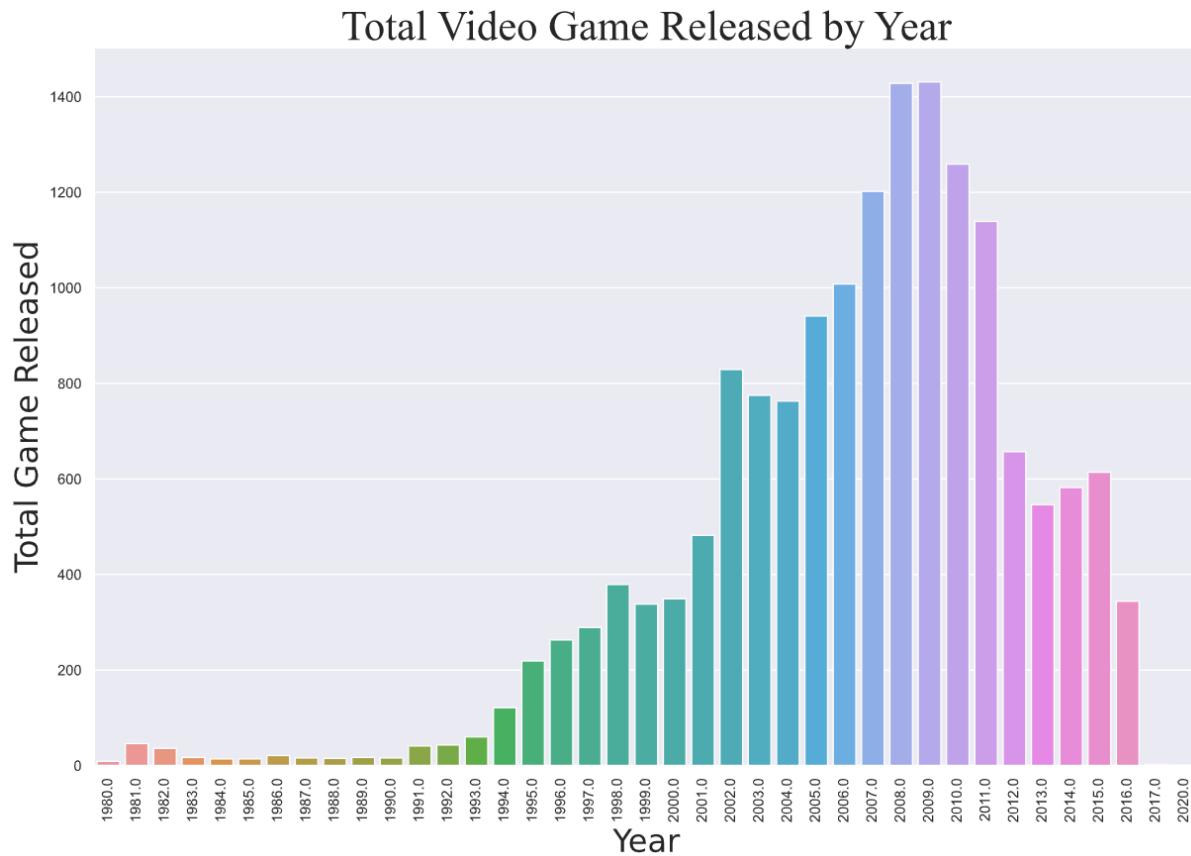


Figure 2.3.5 Total Video Game Released by Year

The figure above illustrates the total number of titles released for any given year from 1980 – 2020. From the figure, it could be seen that 2009 is the year with the highest number of titles released, with 2008 being a close second. After the release of new consoles 3 – 4 years prior, companies are incentivized to release new video games for these consoles as an attempt at capturing the market demand for console specific video games. Competition would be very intense during this time, where each company attempts to innovate and introduce novel concepts in order to stand out in the sea of new video games.

Key Takeaway: Consider releasing video games after the console platform has matured and competition between different video game releases has died down.

(1256 words)

Part 2: Marking Rubric (20 marks)

Criteria	5 – 4 marks Excellent	3 marks Good	2 – 0 marks Unsatisfactory	Total
Lead Sentence (5 marks)	The lead sentence clearly and concisely illustrates the most interesting thing in the data set in a way that makes the reader want to learn more.	The lead sentence illustrates something interesting in the data set but is not written as clearly and concisely as it could be.	The lead sentence does not do a good job of identifying something interesting in the data set and/or does not make the reader want to know more.	
Statistical summary (5 marks)	The summary clearly shows key insights of the data set.	The summary shows some insights of the data set.	The summary does not show key insights from the data set.	
Visualisations (10 marks, maximum 5 marks can be awarded if only one chart is provided)	All visualisations clearly illustrate key insights of the data set and contain a strong “takeaway” message. All visualisations are easy to read and understand.	The visualisations illustrate some insights of the data set and are easy to read and understand.	The visualisations do not illustrate key insights of the data set and/or contain major flaws.	
TOTAL				

Task 3: Data Memo

3.1 Intended Audience

The intended audience for this data memo is personnel that have a vested interest in the video game industry, such as development studios, publishing companies, and distributors. The goal of this data memo is to present our findings and recommendations to the relevant parties who are interested in understanding the factors that affect the performance of a video game, in terms of sales and making changes in their company to increase revenue.

3.2 Findings & Recommendations

3.2.1 North America has the highest market potential for the video game industry

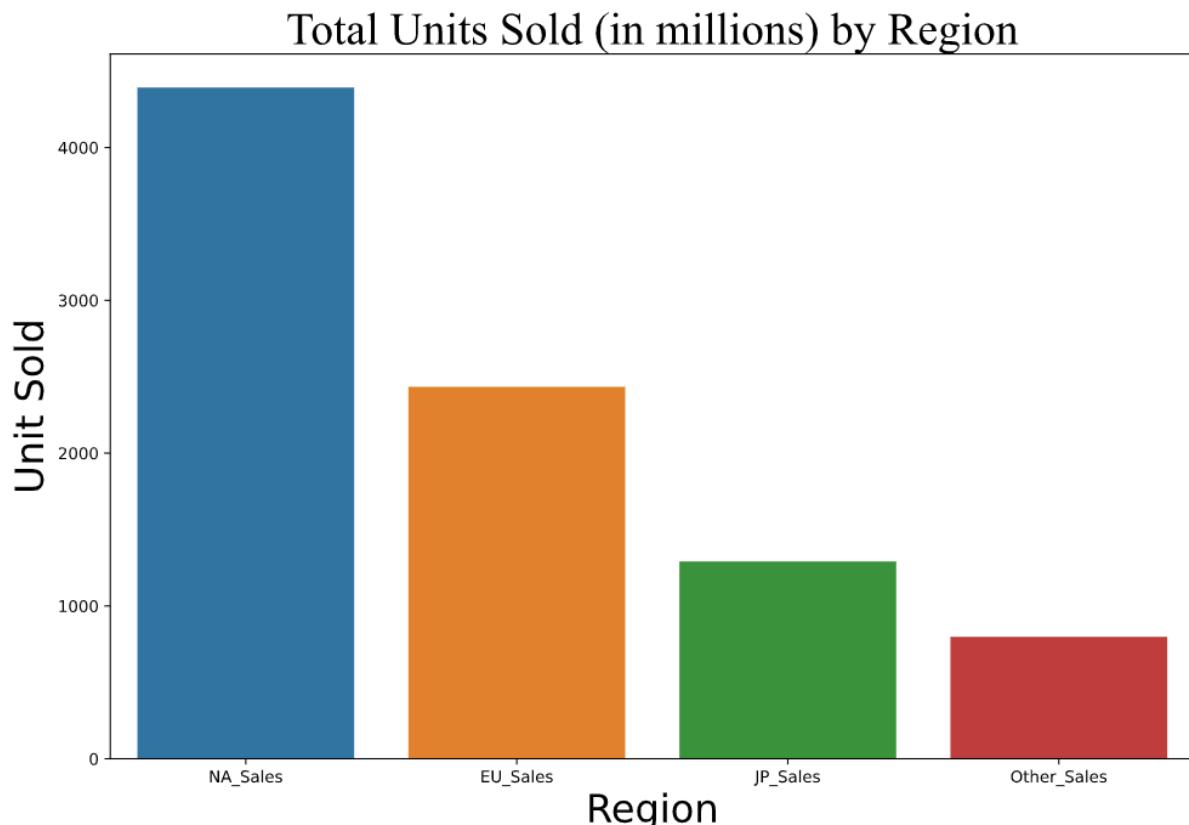


Figure 3.2.1 Total Units Sold (in millions) by Region

3.2.1.1 Findings with supporting evidence

As shown in the figure above, the region that has the highest number of units sold across all titles that have been released is North America, which clocks in at 4.37 billion units over the years. This fact alone tells us that North America has the highest market potential for the video game industry.

According to Research and Markets, [the North America gaming market is valued at 42.83 billion USD in 2020, and that number is projected to hit 83.73 billion USD by 2026.](#) The population's consumption capacity has been gradually increasing due to a higher standard of living and the strong economic presence of North America worldwide, which in turn drives the growth of the market. Breakthroughs, such as virtual and augmented reality that emerged because of the continuously evolving landscape in the industry has also played a key role in the growth of the gaming industry in North America. In addition to that, the COVID-19 pandemic has led to a 45% increase in the time spent playing video games in the region. Moreover, e – sports have gained further traction in the region as there was an increase of 14% in the first – time download of Twitch, an English – dominated live streaming platform that is based in the United States.

3.2.1.2 Recommendation

In lieu of our findings above, it is recommended for executives in the gaming industry to heavily market their company's titles in North America. The North American population have a tremendous amount of purchasing power and would likely hold the highest market share in the gaming industry. Further proving this point, video game spending across hardware, software, and accessories reached 1.6 billion USD amid the COVID-19 pandemic as of March 2020. As such, companies that are not based in the United States should set aside a portion of their budget for marketing campaigns of their titles in the United States. This allows companies to grow and hit their annual target revenue more easily.

3.2.2 The most popular genre of video games across the world is ‘Action’

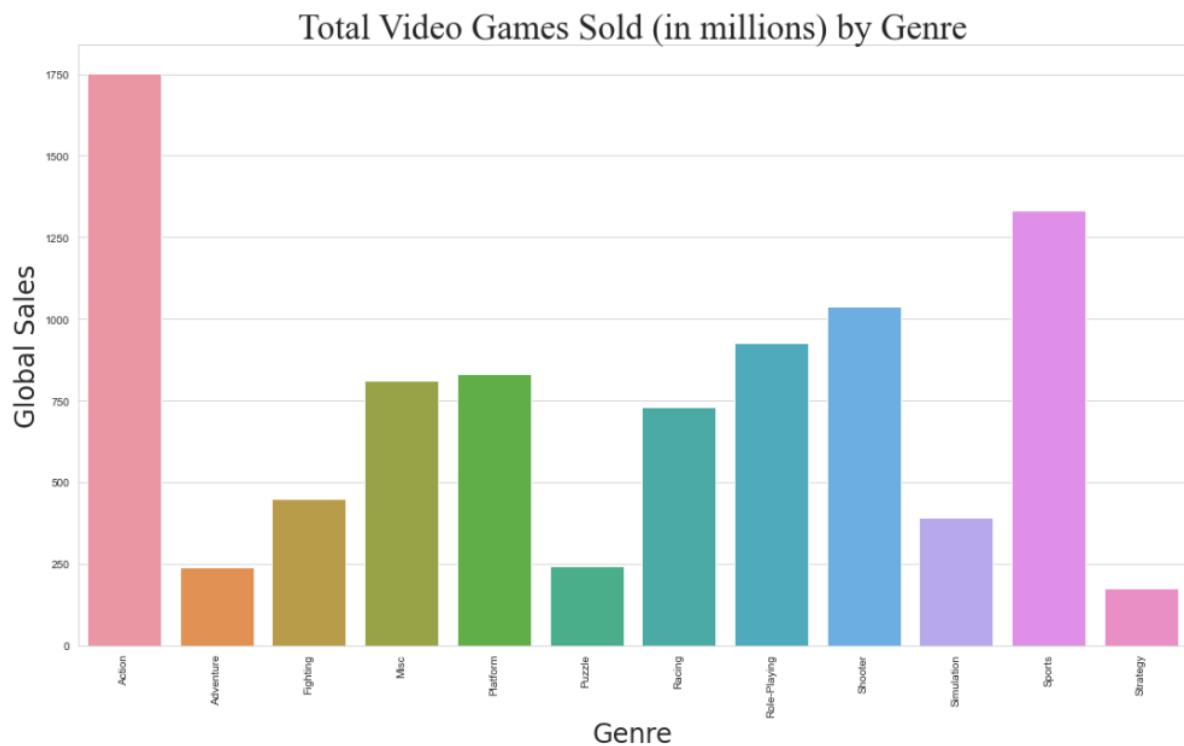


Figure 3.2.2 Total Video Games Sold (in millions) by Genre

3.2.2.1 Findings with supporting evidence

The figure above illustrates the total number of units sold for each genre of video games worldwide. The genre of video games that has the highest number of units sold worldwide is ‘action’, with 1.75 billion units. With this, we can safely assume that the most popular genre of games across the world is ‘action’.

The ‘action’ genre is often associated with violence and adrenaline – inducing gameplay. People who play this genre of games would often experience an adrenaline rush during tense moments in the game. Violence – induced conflict would often trigger a stress response from the players, as a cocktail of hormones – noradrenaline, testosterone, and cortisol is pumped into their bloodstream. The presence of these hormones in the bloodstream would often trigger a fight – or – flight response, which prepares the body to either stay and deal with the threat or to run away to safety. This heightened sense of stress may be very fun and appealing to players, which could be a factor why the most popular genre of video games is ‘action’.

3.2.2.2 Recommendation

It has been well established that the most popular genre of video games is ‘Action’. Therefore, it would be beneficial for development studios and publishing companies to invest more resources into the development of titles in the ‘action’ genre. Titles in this genre perform well in terms of sales and the companies involved would often see a return on their investment.

3.2.3 Different regions heavily favour different genres of video games

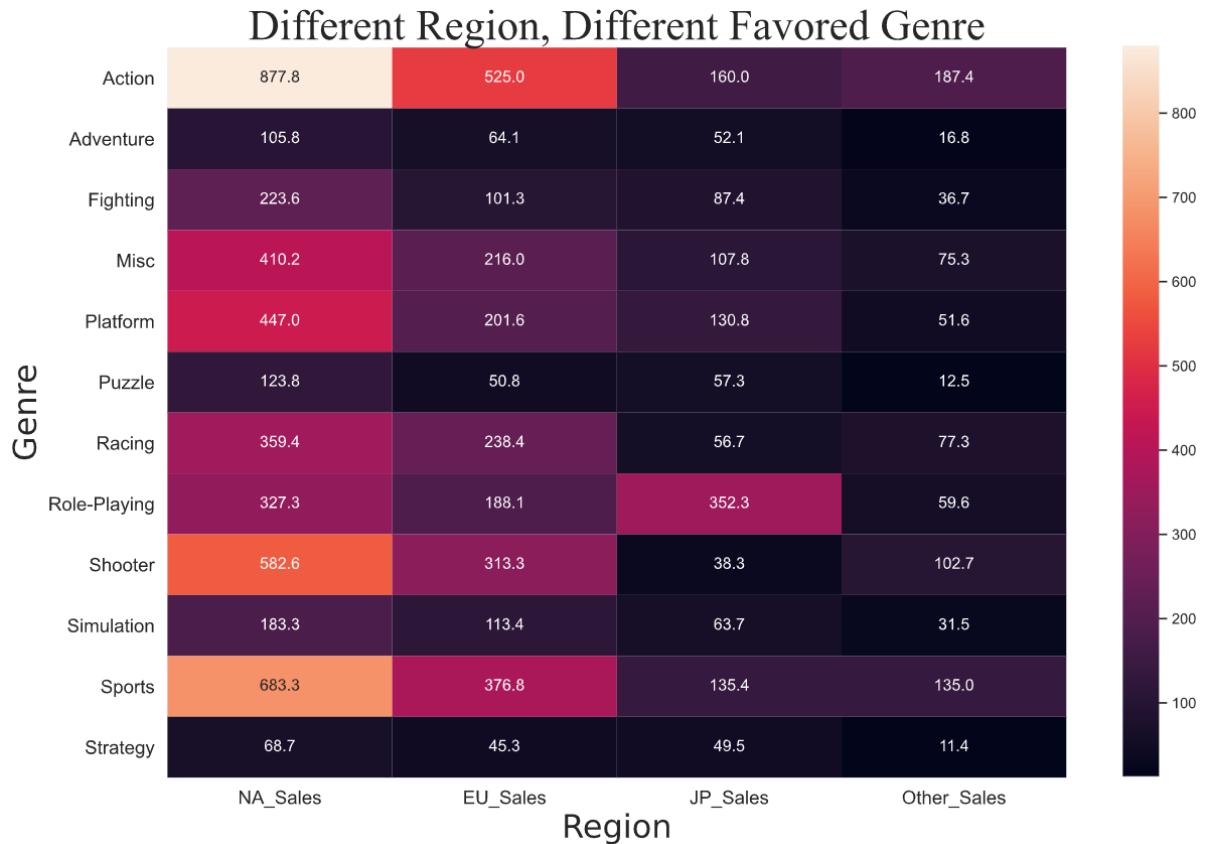


Figure 3.2.3 Different Region, Different Favoured Genre

3.2.3.1 Findings with supporting evidence

However, the dataset tells a different story when the total number of units sold for each genre of video games is further sub – divided to individual regions, which is illustrated in the heatmap in the above figure. It is clearly shown in the heatmap that different regions of the world heavily favour different genres of video games. The most popular genre of video games in North America, Europe and other regions is still ‘action’, while the most popular genre in Japan is ‘role – playing’.

[There are two main reasons why ‘role – playing’ games are more popular than ‘action’ games in Japan](#). The 1st being that Japan has a well – established lineage of ‘role – playing’ games, which includes the ‘Dragon Quest’ series and the ‘Final Fantasy’ series. Besides that, the 2nd reason why ‘role – playing’ games are more popular in Japan is that Japan was completely enamoured by various ‘role – playing’ games that have been developed for home consoles during the time when ‘action’ games involving violence were developed mainly for personal computers by American developers in the 1990s.

3.2.3.2 Recommendation

The genre that is heavily favoured by the people varies highly across different regions. As such, it is highly recommended for companies to perform their individual market research on what games people would be most interested in playing before investing resources into the development of a new title. This dataset is not extensive enough as it only contains the sales figure in three major regions and simply groups every other country into the ‘other’ category.

3.2.4 The PS2 is the most favoured video game console, 2nd and 3rd being the X360 and PS3

Platform with Global Units Sold (in millions)

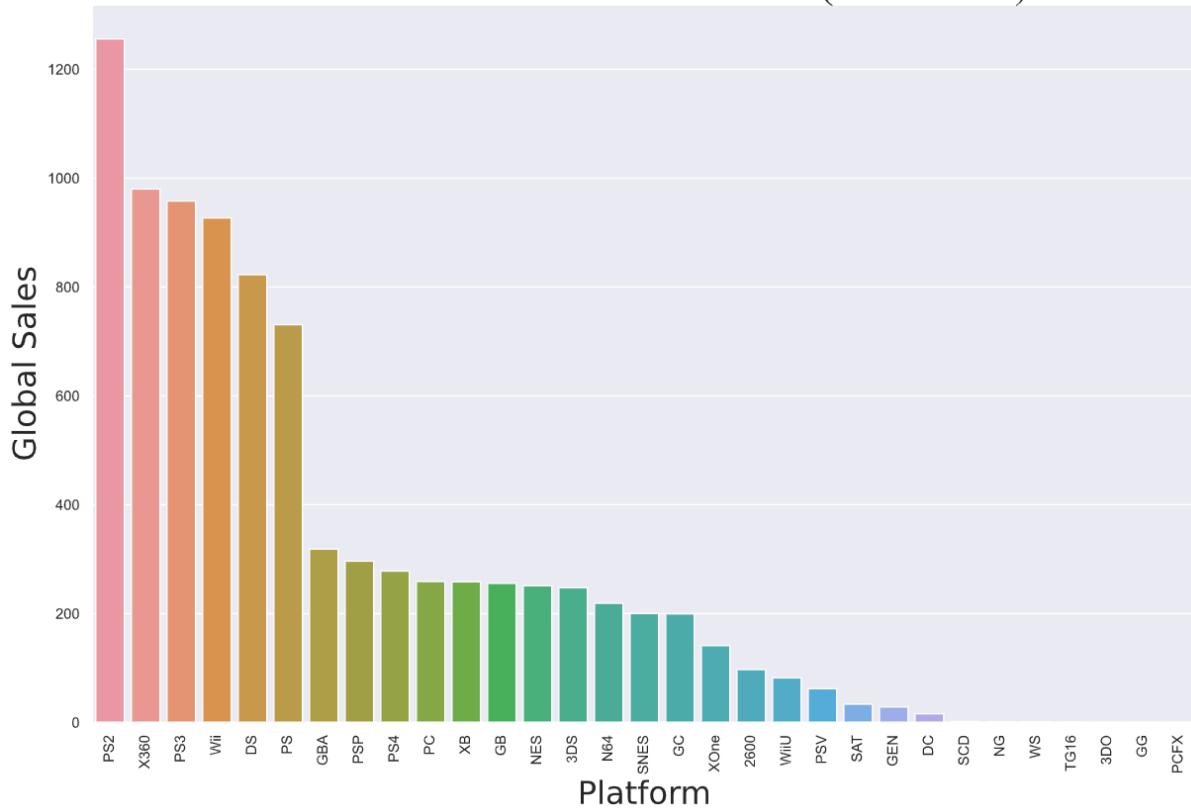


Figure 3.2.4 Platform with Global Units Sold (in millions)

3.2.4.1 Findings with supporting evidence

The figure above illustrates the total number of units sold worldwide across all titles that have been released have been released for a particular console platform. It comes to no surprise that the titles that have been released for the PS2 sold the most copies, with a total of roughly 1.26 billion units. The 2nd and 3rd being the X360 and PS3 with a total of roughly 980 million and 958 million copies, respectively. The dominance of the PS2 over its competitors is the primary factor that led to PS2 titles having the highest number of units sold worldwide.

[There are various reasons which contributed to the dominance of PS2 during its generation, which in turn inflated the total number of units sold across all PS2 titles.](#) One of the selling points of the PS2 is its backwards compatibility with its predecessor, the PS1. This feature effectively expanded the PS2's game library and is one of the factors that pushed customers to choose the PS2 over its competitors. In addition to that, the PS2's stranglehold on the console market during its generation was partly due to strong third – party support. Sony collaborated with Rockstar Games to make Grand Theft Auto III a timed exclusive on the PS2.

3.2.4.2 Recommendation

In lieu of the findings that have been presented above, it is clear that the console with a wider range of titles available to it and a more extensive feature set would be the go to choice between different consoles in its generation. The choice to release more titles on the more popular console would also prove to be beneficial. As such, hardware developers should conduct surveys as an attempt to understand what features customers would like to see in the next – generation consoles. Besides that, development studios should develop and release games for the console platform that is more popular. These recommendations could be integrated into the business plans of these companies to increase their revenue and bring more enjoyment to customers.

3.2.5 2008 is the peak year for video game sales worldwide

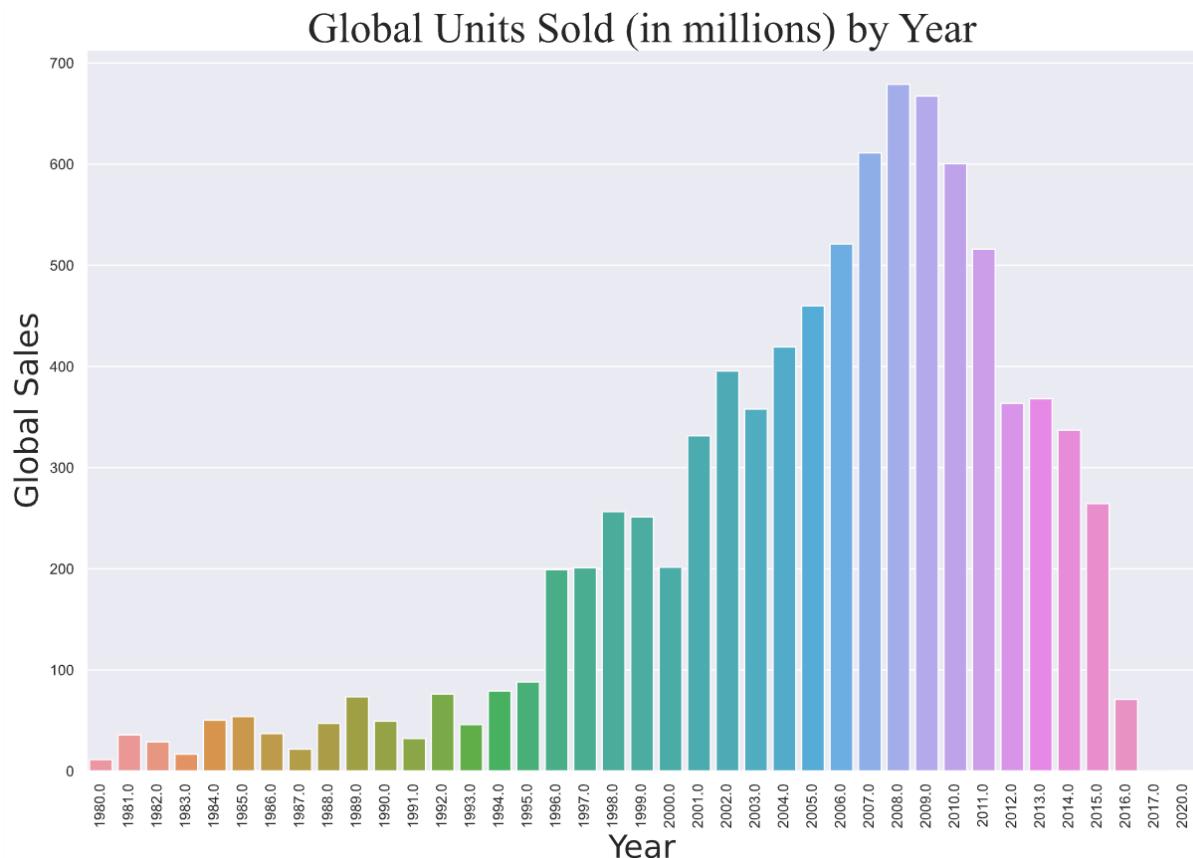


Figure 3.2.5 Global Units Sold (in millions) by Year

3.2.5 Findings with supporting evidence

The figure above illustrates the total amount of units sold worldwide for any given year from 1980 – 2020. From the figure, it could be seen that the peak year for video game sales worldwide is 2008, which peaked at 679 million units.

The primary factor that led to a peak in video game sales worldwide in 2008 is the release of new consoles by Sony, Microsoft, and Nintendo. These three companies all released a new generation of consoles 1 or 2 years before 2008, which meant that there was a huge incentive for companies to release new video games for these consoles. [These consoles have an advantage over the traditional gaming experience on PC](#). One of them being consoles cost less than a PC with similar specifications. Consoles were also much simpler to set up compared to PC, which introduced the idea of gaming to the masses. As such, the introduction of these consoles led to the release of new titles for these platforms, which in turn increased the video game sales worldwide for the next few years.

3.2.5.2 Recommendation

The findings above indicates that the introduction of new consoles is accompanied by a boost in video game sales worldwide. Thus, it is beneficial for companies to heavily invest in the development of new titles for the new generation of consoles instead of older hardware. This is because people who bought a new console would want to experience games on new hardware as it is more powerful and have more features. Thus, heeding this recommendation would increase the revenue of companies who develop and publish games.

3.2.6 Games that have been published by Nintendo are loved all around the world

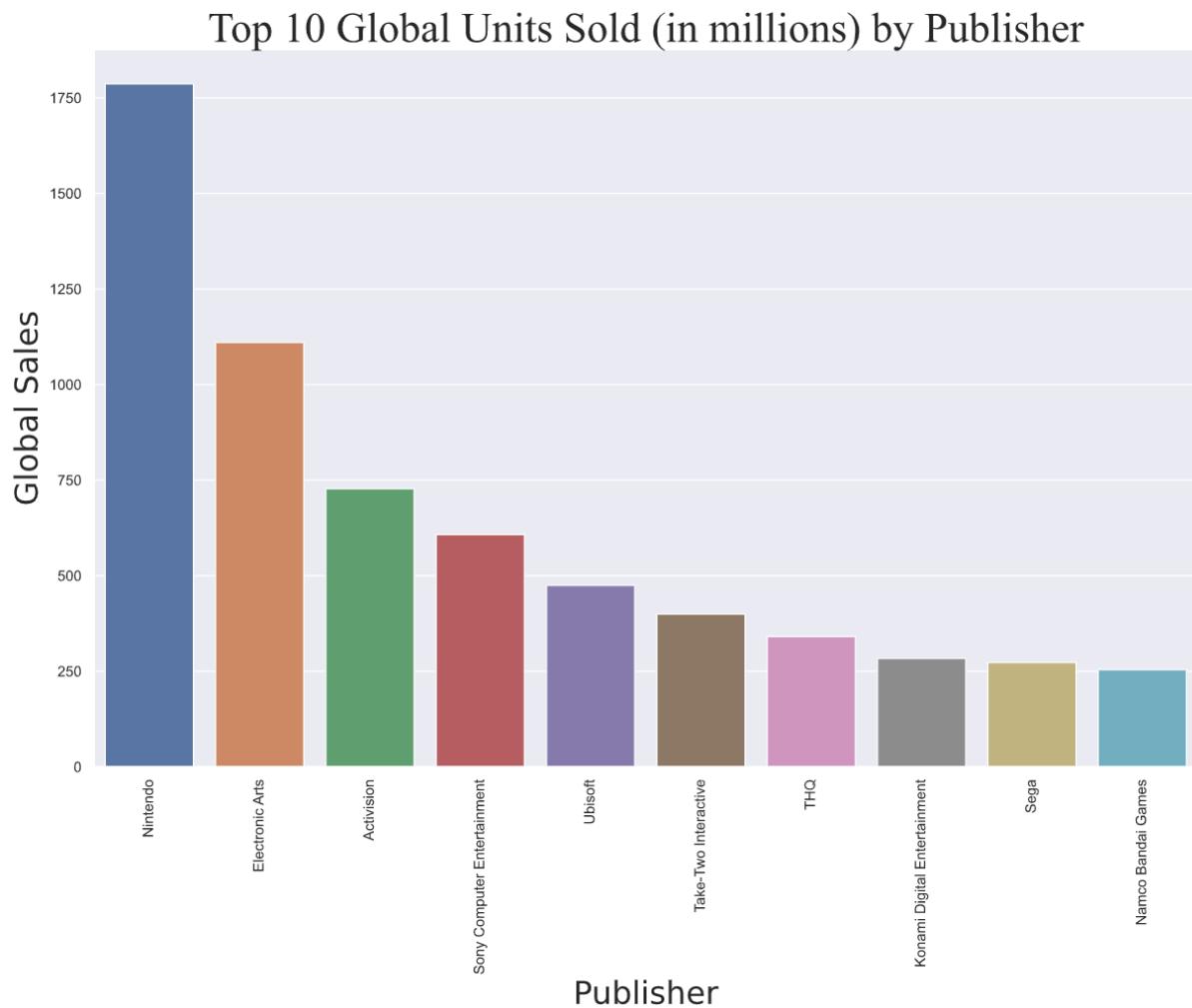


Figure 3.2.6 Top 10 Global Units Sold (in millions) by Publisher

3.2.6.1 Findings with supporting evidence

The figure above illustrates the 10 publishers with the highest number of units sold worldwide. It has been shown that games that have been published by Nintendo are very well liked all around the world and have sold the greatest number of copies over the years, with a total of 1.79 billion copies.

[Nintendo is a family – oriented company whose games' primary focus is fun across all ages.](#) Nintendo designs and publishes a plethora of games for younger audiences. These games have such a wide range of appeal and can be enjoyed by anybody, regardless of age. Furthermore, Nintendo is not afraid of experimentation and is constantly pushing the envelope. Nintendo is willing to innovate and present novel concepts to the market if they think it is fun and enjoyable to the general public. Last but not least, Nintendo owns the rights to franchises that defined the childhood of a lot of adults today, such as Super Mario, and Donkey Kong.

3.2.6.2 Recommendation

In lieu of the findings presented above, development studios and publishing companies could look into the possibility of creating fun and interesting games that can be enjoyed by anybody, regardless of age. These games usually have an emphasis on sociability among players, which promotes the formation of healthy relationships with family and friends. In addition to that, companies should also invest in the development of sequels to well established series, instead of original properties that do not have an existing following. Existing fans of these series are usually willing to purchase the latest instalment whenever it is released.

(1926 words)

Part 3 Marking Rubric (25 marks)

Criteria	5 – 4 marks Excellent	3 marks Good	2 – 0 marks Unsatisfactory	Total
Length Requirements (5 marks)	The report is in-between 1,500 and 2,000 words and has at least six charts.	The report is too short or too long but only by a few hundred words and/or there are four or five charts in the report.	The report is significantly longer or shorter than it should be and/or it falls well short of the number of required charts.	
Writing (5 marks)	A well-written and well-structured report with proper use of grammar, punctuation and spelling.	A relatively well-written report with some minor flaws on grammar, punctuation and spelling errors that do not impede overall understanding.	A fair-written report with major flaws on grammar, punctuation and spelling errors that may disrupt reader's understanding.	
Ideas and Insight (5 marks)	Some clear, compelling points have been derived from the data analysis and an appropriate recommendation has been proposed.	Some clear, compelling points have been derived from the data analysis, but some key insights have been overlooked.	The insights and ideas are not clearly presented.	
Visualisations (5 marks)	The charts clearly illustrated key insights of your findings. The charts are appropriately sized to be easily read within the report	The chart illustrates some insights of your findings.	The charts do not illustrate key insights of your findings.	
Outside Sources (5 marks)	At least six outside sources have been well-summarised and referenced in the report. The sources are meaningful and greatly improve the readability and content of the report.	The outside sources are adequately summarised but some of the sources does not contain meaningful information.	Little meaningful reference to published literature and mainly are “nice-to-know” information.	
TOTAL				