Isaiah Sarria and Randy Koliha

April 26th, 2023

Video Game Sales Since the 80s

An International Analysis

Abstract

The video game market is an entertainment industry that is ready to make large amounts of revenue for investors involved. Using Python data analysis and visualizations, our goal is to find methods in the future to capitalize from this business by determining what are the global trends regarding video game sales. The data we are analyzing was discovered from a Kaggle posting which sourced the data from VGChartz. The dataset provides a vast number of sales regarding regions and platforms of almost 17 thousand rows. The dataset tells us information about sales in North America, Europe, and Japan specifically, but it also gives insight on all other global sales. Though some imperfections within this large dataset prevents us from using more predictive methods such as linear regression accurately. The approach is to use tables, graphs, and interpretations to answer many other smaller questions. These smaller queries will aid our investigation to eventually discover the global trends of consumers in this specific field of technology and software. Essentially, the graphs indicated that North America and the European Union regions share very similar tastes and preferences when it comes to the platform, genre, and the developer of the game. The most latter point is especially applied to whether the video game developer is from the western vs eastern hemisphere. The North America and the European Union liking some of the Japanese titles, whereas the Japanese market almost exclusively consumes Japanese content.

1. Introduction
   1. Motivation

Chart, bar chart

Description automatically generatedAccording to [Insider Intelligence](https://www.insiderintelligence.com/insights/us-gaming-industry-ecosystem/), two-thirds of the US population under eighteen years old play video games online. The video game industry is not new, but the pandemic of COVID-19 caused the average rates of time spent gaming increased by 16.5% between 2019 and 2020 from 12.7 to 14.8 hours per week, according

Figure 1: US Consumer Spending on Video Game Products 2019-2022

to The NPD group and has barely slowed down as it was at 16.5 hours per week in 2021. The video game business is quickly growing and very popular as it is estimated that in 2022, more than half (at 54.2%) of the population do some sort of digital gaming. Figure 1 reflects these findings as it shows a net increase of about 10 million more USD spent by consumers in the US market from 2019 to 2022. If there is growth and realized potential in video game production, then it is an important market for investors to pay attention to and they already are paying attention to it as ad revenues grow to $6.28 billion in 2023 with a forecasted continued growth rate of 8-10% within the next few years. The competitive Esports scene for video games has an ad revenue with a similar rate of growth “and will pass a quarter of a billion dollars in 2023.” Looking at the data, it was observed that video games won’t go anywhere, and we should monitor its trends to better cater to consumers in the market to find out what gamers want and where those customers are located, which greatly helps game developers choose profitable projects.

* 1. Report Description

This market research report focuses on sales over time in different regions with a time series analysis to make some predictions in the future about the patterns of video game sales among different regions. Furthermore, the outcome of the market research determines what platforms, titles, and genres certain regions prefer to buy. Broadly speaking, the aim is to see what certain groups of consumers prefer and to see if the market is growing. The big question is: “What are the patterns of regional video game sales?”

1. Data
   1. About the Data

The dataset we are analyzing is coming from a [Kaggle posting](https://www.kaggle.com/datasets/gregorut/videogamesales) where they got their information from [VGCharts](https://www.vgchartz.com) as the post acknowledged. VGChartz is a video game industry research firm that hosts a game database system that features current and legacy video game sales data among providing other content such as news, sales (discounts), reviews, articles, and a social network supported with a community forum. The data dictionary of the information will be given in Figure 2.

|  |  |
| --- | --- |
| Rank | The higher the number of copies were sold the higher the rank |
| Name | Name of the game |
| Platform | The platform it was sold for |
| Year | The year the game was released |
| Genre | The genre of the game |
| Publisher | The publisher of the game |
| NA\_Sales | The copies sold in North America by the millions |
| EU\_Sales | The copies sold in Europe by the millions |
| JP\_Sales | The copies sold in Japan by the millions |
| Other\_Sales | The copies sold in other regions that are not North America, Europe, or Japan in the millions |
| Global\_Sales | The total number of copies sold |

Figure 2: Notation of what each column represents.

* 1. Limitations and Future Potential Study

Unfortunately, some data is missing. For example, we think that we don’t have access to the latest data because 2018 and 2019 are completely missing. Furthermore, 2017 and 2020 reported too few games at three and one respectively, which is not possible as more games than that have released during those years. Considering missing data and years, the data reflects a significant dip in video game sales which may be attributed mostly to this missing data. If proper data was provided with this dataset in the future, it would have been very possible to provide an accurate linear regression model to forecast future patterns. However, the current data only enables us to make simple inferences and an inaccurate, but suitable linear regression model by ignoring data from 2017 and beyond.

A screenshot of a computer

Description automatically generated with medium confidence

* 1. Summary Statistics

In Figure 3, the first row counts the number of not null values in each column. Therefore, the reason why the Year column does not have the same number values as the other columns is because some entries in the Year column are null,“NA.” The next row calculates the means of all the values. Some means aren’t as useful as others such as the average year or rank of the games in the dataset because it doesn’t provide an advantageous insight for any kind of strategy or planning. Though, the average video game sales per region may help to establish a good benchmark for some video game developers to aim for. For example, the average sales of a video game in North America is about 264,5667 copies sold. This number won’t satisfy the goals of well-established companies with high budget production teams, but for smaller scale indie developers it could suffice. Because this number is the mean of all video game sales in North America since the 80s, so some considerations such as the general popularity of video games at the time must also be considered along with other factors (company size or name recognition). If you reference the table’s 25%, 50%, and 75% rows for the rank column, it may be confusing. For example, the value that corresponds with the 25th percentile is not supposed to be as ideal of a position compared to the 75th percentile because it would mean that the value is lower than 75% of the other values. Now that is true, and but this is not the case because a higher value in ranking is inferior to a lower value rank. Consider the fact that the number 1 rank is better than the number 3 rank. That is the reason why interpreting this way, which would otherwise be incorrect, is actually the correct line of thought.

Figure 3: Summary Statistics of the Data frame.

The min and max row simply state the minimum and maximum number of sales a video game achieved per region. In the instance of North American sales, the maximum number of copies that a video game sold there was 41.49 million, and that was Wii Sports for the Nintendo Wii which was released in the year 2006.

1. Methods
   1. Software Selection

It was decided to conduct the analysis via Python because of the Pandas and Matplotlib libraries, access to Jupyter Notebook for any needed notation, to produce a report fit for a resume or portfolio using Python, and Python can be branched out into other complex data analysis tasks. Python’s feature of Jupyter Notebook helps to better organize the workflow of the raw code as well as provide the ability to run code one line at a time if needed for testing. Another reason Python was chosen is to practice, hone, and showcase what can be done in Python to then add this project to a portfolio for future employers. The practice and exposure was also greatly needed to be better prepared for real-world cases in related fields to better address any issues, errors, or other hinderances we may experience in the future.

* 1. Analytical Approach

The Pandas Python library provides with the means of data analysis and manipulation. In combination with the Matplotlib library, we can use the Pandas “groupby” function to organize data in a way so we can then use Matplotlib library to create a large variety of graphs depending on what is appropriate for the insight we wish to share.

In the approach, first we took time to understand the general nature of the data. This time to study the data included checking the unique values in each column to prevent repeating instances later in analysis, discovering the data types being utilized, and creating a summary-statistics table was done to find counts, means, standard deviations, minimum and maximum values, and percentiles (50th is median) of all the columns as seen in Figure 3.

Then, using what is available, some questions were conducted to help guide what graphs should be created to be able to answer the supporting questions: What is the proportion of game sales globally? What platforms are preferred for each region? Over time, what are the sales per region? What are popular genres for every region? Is there a preference for publishers? How many games are made per year? Does that explain the patterns with sales?

* 1. Python v. Other Methods

The final reason why Python was selected over R and Excel is because Python specializes in data analysis for big data. Excel simply cannot handle data analysis on larger scales, which is exactly what was needed to work for this data set. In fact, Excel wasn’t even able to open and save the CSV file of the data set as stated in a pop-up prompt in Excel. In comparison to R, R is more of a specialized programming for statistical analysis which is not in the scope of this analysis. Though, in R the methodology would be the same in terms of data analysis, Python is a newer and more powerful language which indicates more longevity than R for future expansions on this study.

1. Results

Research Q1: What is the best-selling games were in each region?

A blue and white rectangle

Description automatically generated with low confidence

Figure 4: Top Selling Games in NA All Time.

Icon

Description automatically generated with low confidence

Figure 5: Top Selling Games in JP All Time.

Icon

Description automatically generated

Figure 7: Top Selling Games in Other All Time.

Figure 6: Top Selling Games in EU All Time.

Icon

Description automatically generated

Figure 7: Top Selling Games in EU All Time.

Bar chart

Description automatically generated

Figure 8: Top Selling Games in Global All Time.

In these figures we found that the Wii Sports title was the most present. We believe this is because the Wii Sports was included with the Wii console, making Wii Sports seem like a popular seller when it was just an inclusion. We also see here that there are a lot of Nintendo titles as well as Grand Theft Auto V in the figures. Something to highlight here is that in the Japan figure (figure 5), we find that the titles that were top sellers are all Japanese developed titles. This shows a particular taste in the Japanese market for Japanese made games.

Chart, bar chart, histogram

Description automatically generatedResearch Q2: How do regional markets compare by platform? More specifically, how does the world differ from Asian markets?

Figure 9: JP sales by platform.

Chart

Description automatically generated

Figure 10: NA sales by platform.

Chart, histogram

Description automatically generated

Figure 11: EU sales by platform.

Chart, bar chart

Description automatically generated

Figure 12: All regions compared sales by platform.

From the figures we observed that the Japanese market (figure 9) differs from the western figures 10 and 11 which shows North American and European Union trends. The Japanese preference for platforms is heavy leaning more on Japanese made consoles. In the Japanese market almost no Xbox 360 games were sold compared to the western market. In the western market, Japanese consoles are almost as preferred as United States consoles. As seen in figure 10, the North American market had almost balanced sales for PlayStation and Xbox console games. However, it seems that Xbox is slightly more preferred in North America. In the EU (figure 11), it seems that Japanese consoles are slightly more preferred than US consoles but are very close. This seems to show an openness to outside game development from other countries in the western world whereas the Japanese market seems to be more closed.

Research Q3: Which region has the most video games sales?

Chart, pie chart

Description automatically generated

Figure 13: Proportion of games sales globally

Research Q4: What does each regions games sales look like over time?

Chart, line chart

Description automatically generated

Figure 14: Regions timeseries of sales.

As seen in figure 14, sales seem to increase over time until around 2010 where there is a sharp decline. During the investigation, we found that the dataset has missing values for years and the data can be skewed due to the Wii Sports games sales (As mentioned in this paper Wii Sports was bundled with the Wii consoles, so we believe that its sales are not representational because there were not actually purchased). So, if Wii Sports was taken out of the dataset and we had access to the later years sales data, then we believe the trend would have less dramatic descend in later years and an increase after that decline. Possible reasons for less video game spending around the 2008 years could be due to the 2008 economic downturn, however in the data we should have seen an increase in video games sales in the years after the economic recovery.

Research Q5: How do preference of genres differ in each region?

Chart, bar chart

Description automatically generated

Figure 15: NA game sales by genre

Chart, bar chart

Description automatically generated

Figure 16: Other game sales by genre

Chart, bar chart

Description automatically generated

Figure 17: EU game sales by genre

Chart, bar chart

Description automatically generated

Figure 18: JP game sales by genre

Chart, bar chart

Description automatically generated

Figure 19: Combined game sales by genre

As seen in the above figures 15-19 there is a little bit of a theme when it comes to the tastes of genre in the western world compared to the Japanese market. In the European Union and North America, the popular genres gravitate around Action, Shooter, and Sports. This is very different in Japan; in the Japanese market the most popular genre is Role Playing games. Although Action and Sports are still somewhat popular in Japan, they are not nearly as much as Role Playing games.

Research Q6: How do regions differ in terms of publisher preference.

Icon

Description automatically generated

Chart, bar chart

Description automatically generated

Figure 20: EU sales by publisher

Figure 21: Global sales by publisher

Icon, bar chart

Description automatically generated with medium confidence

Figure 22: Other sales by publisher

Chart, bar chart

Description automatically generated

Figure 23: NA sales by publisher

Chart

Description automatically generated with medium confidence

Figure 24: JP sales by publisher

As seen in the above figures, in the western regions of the European Union and North America the taste in video game publisher is the same. In the European Union and North America, the taste in publisher varies by taste in region. In Japan the taste in publisher is solely Japanese publishers. In the Japanese market they prefer Japanese publishers. This is to be expected because we know from figure 5 that the titles preferred by Japanese are Japanese made titles. This tells us in specific which Japanese publishers are preferred. So, this can be used as a more zoomed in lens for Japanese preferences.

Research Q7: How many games were made each year globally.

Chart, line chart

Description automatically generated

Figure 24: Total number of titles made each year.

As seen in the time series graph, video games made over time has sharply increased and then sharply decreased. Like what was mentioned the sharp decline could be in part due to missing data as well as titles included with consoles which skew the data like Wii Sports.

1. Conclusion

In conclusion, the patterns of regional video games sales can be broken up into seven sub-research questions of the dataset: Q1) What are the best-selling games in each region?,  Q2) How do regional markets compare by platform? More specifically how does the world differ from Asian markets?, Q3) Which region has the most video games sales?, Q4) What does each regions games sales look like over time?, Q5) How do preference of genres differ in each region?, Q6) How do regions differ in terms of publisher preference?, Q7) How many games were made each year globally? We found that the best-selling games were varied in the western world, meaning they had preferences of games made in and out of their respective region. However, when it came to the Japanese market, they preferred Japanese made titles. Regarding platform, similarly to the last research sub-question, western markets varied in the preferences in platform from different regions while the Japanese market preferred Japanese made platforms. We also found that North America had the most video game sales compared to every other region. For regions game sales over time, we found that they all have a sharp incline and a steep decline, we believe this is in part due to missing data and game sales that were inclusions rather than actual real game sales. Additionally, we think we can predict the time series more accurately with a linear regression in the future. For preferences of genre in each region we found that the western world follows the same preference which are: action, shooter, sports games. In the Japanese market this was very different. In Japan the overwhelming preference was for Role Playing games, action was a popular genre in Japan but not nearly as popular as Role Playing. Looking at publisher preference we found again that the western market followed similar taste while Japan had specific Japanese taste in game publishers. Just like game sales over time the global games made over time followed the similar trend and the rationale for its trend is the same as the sales over time trend. In summary, when selling games to the Asian market more specifically the Japanese market, focus on role playing games by Japanese publishers with Japanese platforms. For the western market when selling games stick to action or shooter games from a wide range of publishers and platforms. Now, if you want to make the most sales of video games the North American market might be the best choice due to the high consumption. Lastly, the game market with adjustments to the time series should be a steady incline with dips, so the current model should be adjusted for future reference.

References

Intelligence, Insider. “US Video Gaming Industry in 2021: Gaming Devices & Gaming Video Content Viewership Trends.” Insider Intelligence, 21 Jan. 2022, www.insiderintelligence.com/insights/us-gaming-industry-ecosystem/.

“Video Game Charts, Game Sales, Top Sellers, Game Data - VGChartz.” Www.vgchartz.com, [www.vgchartz.com/](http://www.vgchartz.com/).

“Video Game Sales.” Www.kaggle.com, [www.kaggle.com/datasets/gregorut/videogamesales](http://www.kaggle.com/datasets/gregorut/videogamesales).