# 1.0 Introduction:

The data that is going to be analysed consists of information regarding the total number of copies of video games that have been sold all over the world between the years 2000 and 2020. There are 16599 rows of data to help organise the information, and there are 15 columns to help with that. These pieces of information are contained inside the columns:

|  |  |
| --- | --- |
| Columns | Description |
| Rank | Ranking of overall sales. |
| Name | The name of the game. |
| Platform | The platform on which the game was released (i.e. PC, PS4, etc.) |
| Year | The Year The Game was Released. |
| Genre | The Genre of the Game. |
| Publisher | Publisher of the Game. |
| NA Sales | Sales in North America. |
| EU Sales | Sales in Europe. |
| JP Sales | Sales in Japan. |
| Other Sales | Sales in the rest of the world. |
| Global Sales | Total sales from all over the world. |
| Critic score - | Total rating compiled by members of the Metacritic crew |
| Critic count | Number of reviewers utilised to calculate Critic score |
| User score | Score determined by the users who subscribe to Metacritic. |
| User count | The total number of users who contributed to the user score. |

Excel was used to remove all NULL values, duplicates, and gaps in the data (such as when just part of a name or a year was provided), as well as correct certain incorrect dates associated with games. This action was taken to improve the reliability of the data.

# 2.0 Data Analysis and Visualisation:

## Declining Trend in the video game market.

## 1.Total Sales by Year:

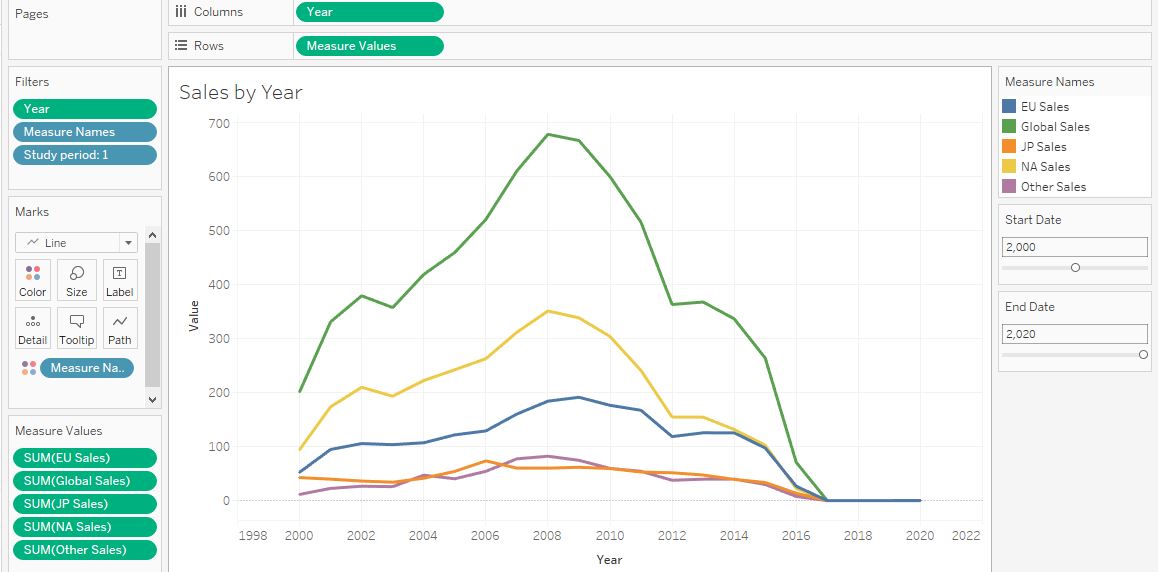


Figure 1. Video Game Sales by Year

The above graph is demonstrating an overall trend in the video games sale from the years 2000 to 2020 in each region. Global sales are the sum of all the regions depicted above in my data set and the zones in the graph above are depicted by the different colours.   
From the above it can be noted that in the early 2000’s to 2008 an increasing trend can be noted in all regions however, after 2008 a declining trend can be observed right up to 2020. Overall, from my data set it can be concluded that globally sales have declined post 2008.

However, I started to wonder whether it is true that the market for video games is contracting, as well as what the possible factors are that can lead to the declining trend. As a result, I did some additional secondary research on top of the dataset.

According to the WePC report titled "Video Game Statistics 2021," the data trend of the value of the video games industry has been positive over the course of the past decade. In addition, the increase in the number of people who play video games has been a favourable, with 35% of gamers coming from the Asia-Pacific region however, there has been a minor reduction in the number of players in the North American region.

I drew several plausible explanations for the downward trend in global video game sales. These theories include the following:

* Dataset marketplaces are incomplete. In 2016, the Asia-Pacific market accounted for over 50% of the worldwide video game industry's value, however Newzoo's dataset only includes Japan. Since 2016, China has become the world's largest game market, surpassing the US and doubling the Japanese market, explaining the graph's inaccuracy.
* Dataset platforms are scarce. Mobile is missing although the mobile game market has grown dramatically over the previous decade. In 2020, mobile games accounted up 51% of global gaming revenue, while PC and console games split the rest.
* It's possible that the late 2000s saw the most video game releases overall. The number of games produced every year can have a big impact on the graph's readability, as it represents the cumulative sales of games.

## 2. Top 5 Video Games Sales by All Markets.



Figure 2. Top 5 Video Game Sales

As video games are ranked according to how many copies, are sold it can be noted from the above that between the years 2000 and 2020, the title of "most successful" goes to Wii Sports, "Grand Theft Auto V," and "Mario Kart Wii." However, it can be noted that the trend is a bit different in JP as Grand Theft is the third most popular game which is not similar in Europe, North America and other regions.

## Top 10 Platform Global Sales

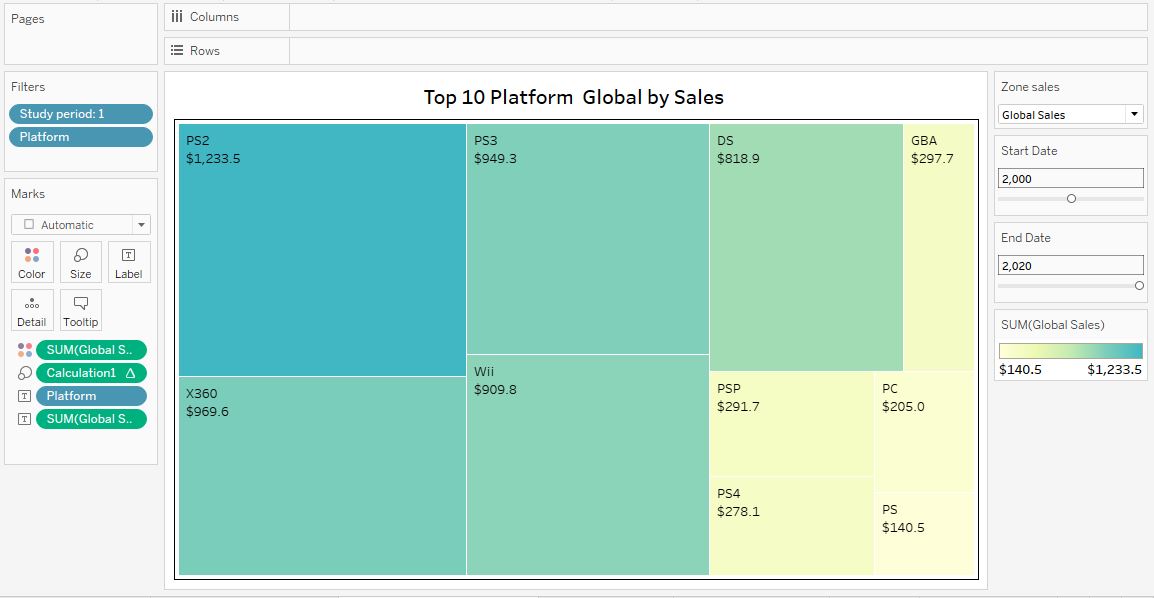


Figure 3. Top 10 Platform Global Sales

The ranking of these products in terms of sales can be seen in the chart that can be found above; the colour and size of the squares are used to classify the popularity of the platforms. Since darker blue indicates the biggest sales, it is evident that PS2 is the most successful platform with X360 being the second most popular platform. It can be observed that they have been the most successful gaming consoles from the year 2000 through the year 2020, followed by PS3 and Wii.

## 4. In terms of sales, action games dominated.

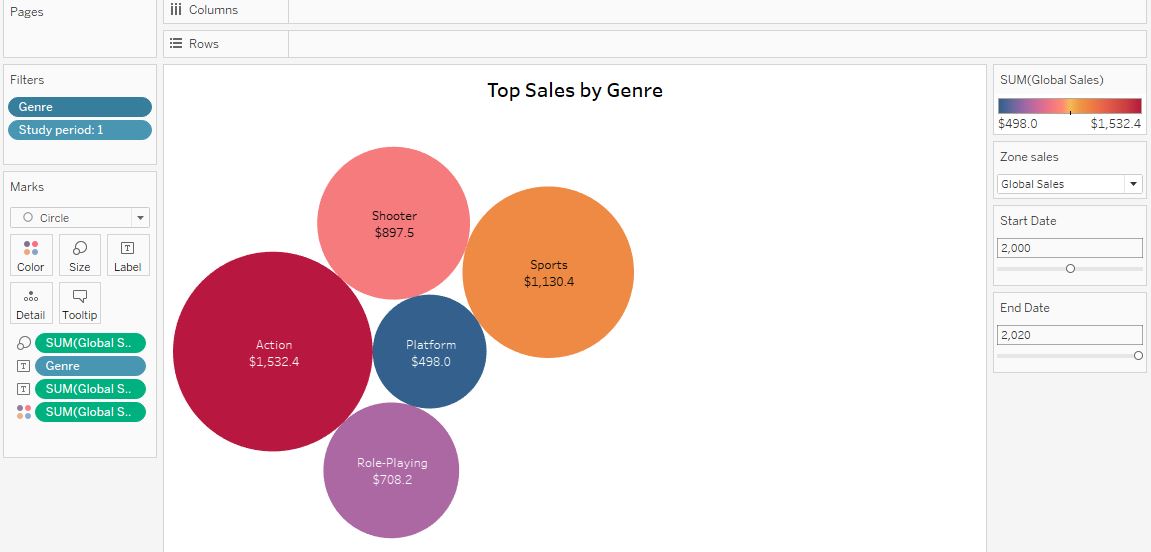


Figure 4. Top 5 Genre Global Sales

In the image that can be seen above, the number of global sales is represented by the size of the circles, and the colours are used to differentiate between the various genres. When the genres are ranked according to the size of their circles, it becomes abundantly evident that the genres with the highest global sales are Action, Sports, Shooter, Role-Playing, and Platform.

## 5 Top 10 Publisher by Sales

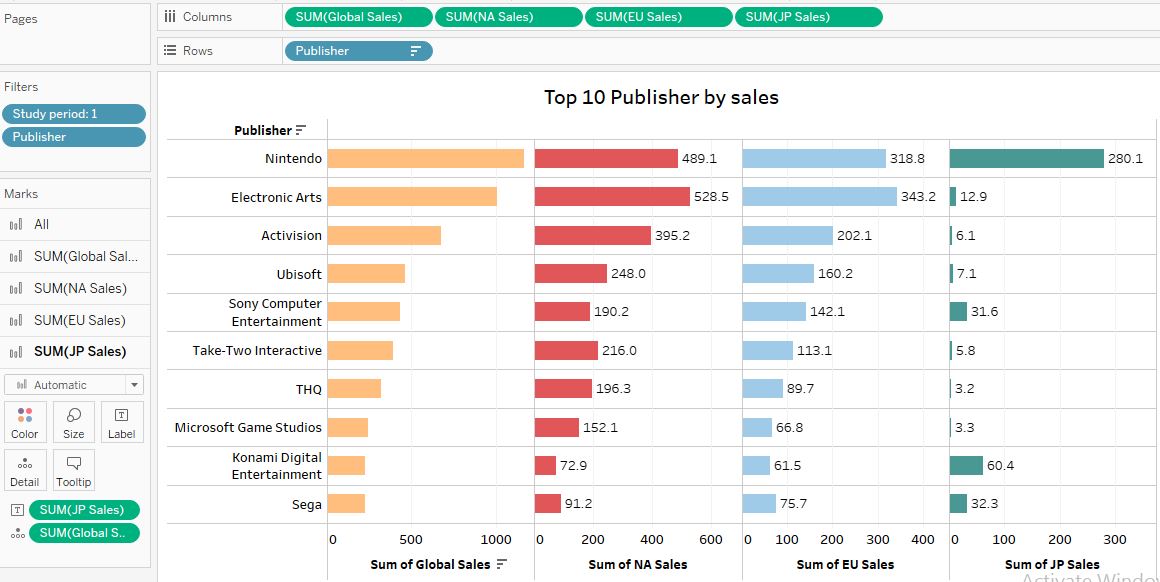


Figure 5. Top 10 Publisher by Region Sales

The top 10 publishers in terms of total sales are displayed in the graph that can be found above. The graph employs colours to define regions and displays sales to slightly differentiate publications so that the information can be read more easily. While Electronic Arts holds the biggest market share in the Europe, Nintendo reclaims first position owing to its domination in Japan. North America has the largest share of the worldwide video game business and Japan has the smallest share.

## 

## 6 Best-selling games in top 5 Genre and Platforms

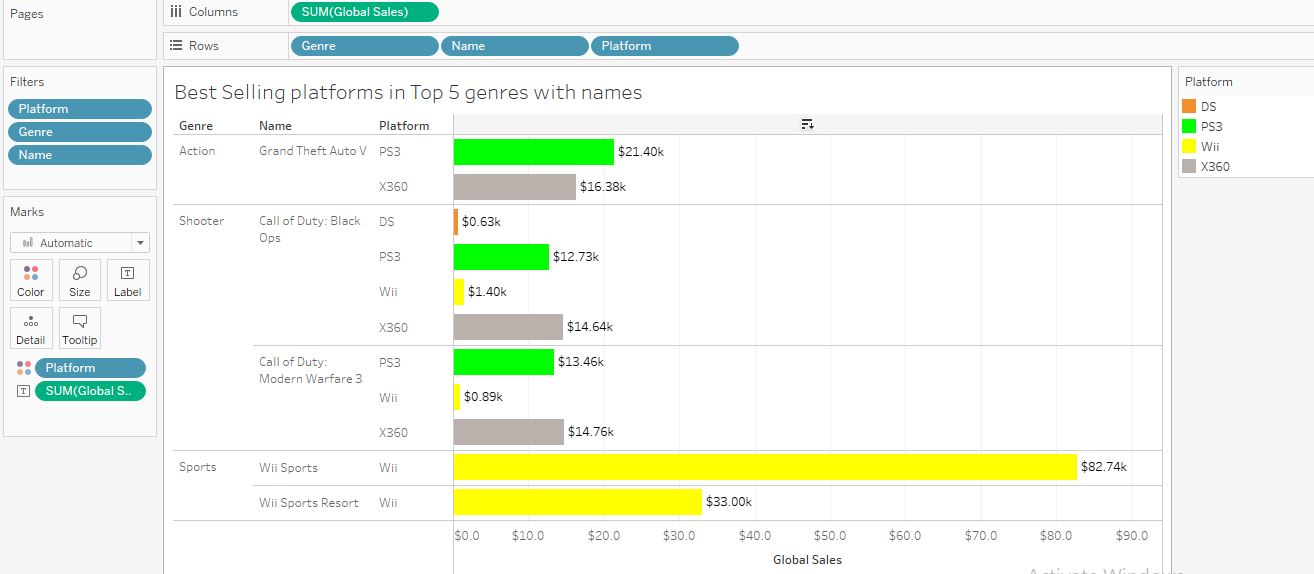


Figure 6. Best-selling games in top 5 Genre and Platforms

The best-selling games in top 5 genres and platforms is denoted by the colours against each of the platforms. It can be noted that for action the best-selling game is Grand Theft on the PS3 console. for Shooter the Call of duty is the best-selling on the X360 platform and for sports Wii sports is the best-selling on the Wii platform.

## 7. Yearly Sales of NA and JP Market.

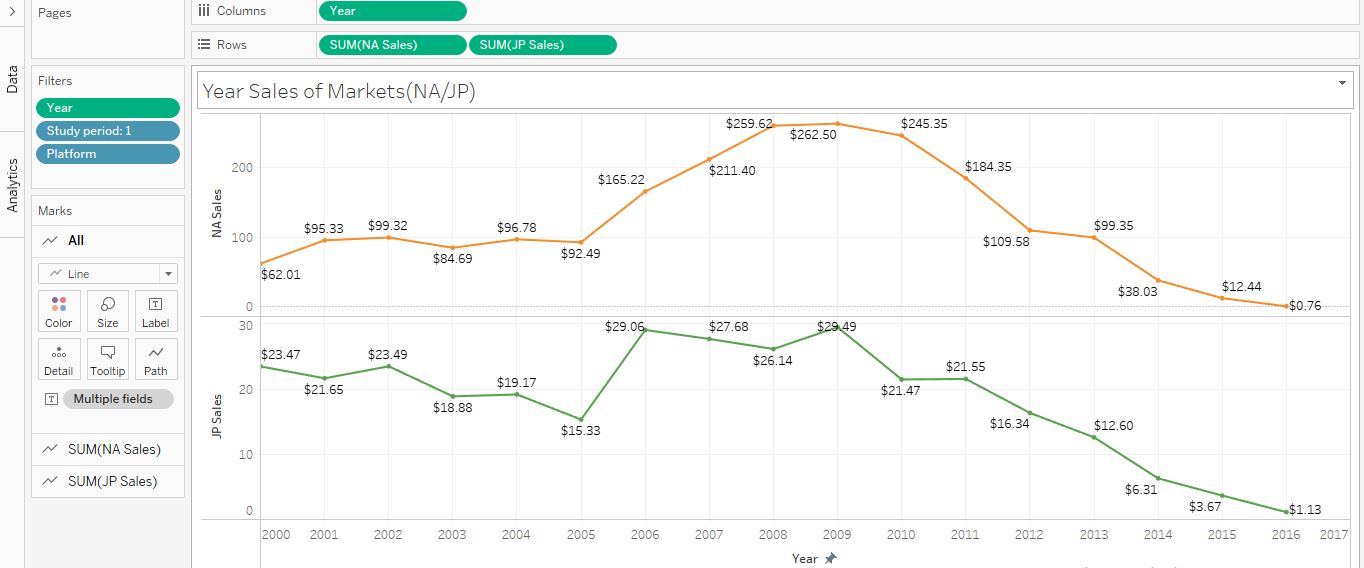


Figure 7. Yearly Sales of NA and JP Market

While reviewing the dataset I noted that yearly sales for each region where quite high except for Japan. The above graph depicts North America sales when compared to Japan. It was noted that an overall downward trend is noted in the Japan for the years 2000 to 2005 which is different to North America. The sales in North America have overall increased from $62.01 to $92.49 whereas in Japan the sales have decreased from $23.47m to $15.33m. A similar trend is noted for the years 2005 to 2010 where North America sales have increased to $245.35m from $92.49m whereas in Japan the sales increased to $29.49m in 2009 from $15.33 but then decreased to $16.34 in 2012 demonstrating an overall increase of only $1.01m. Further, it can also be noted that sales for years 2010 have decreased in both North America and Japan. Moreover, the decrease in North America is 40% compared to 15% in Japan.

## 8 Comparing European and Japanese Publisher Top 10 Sales.

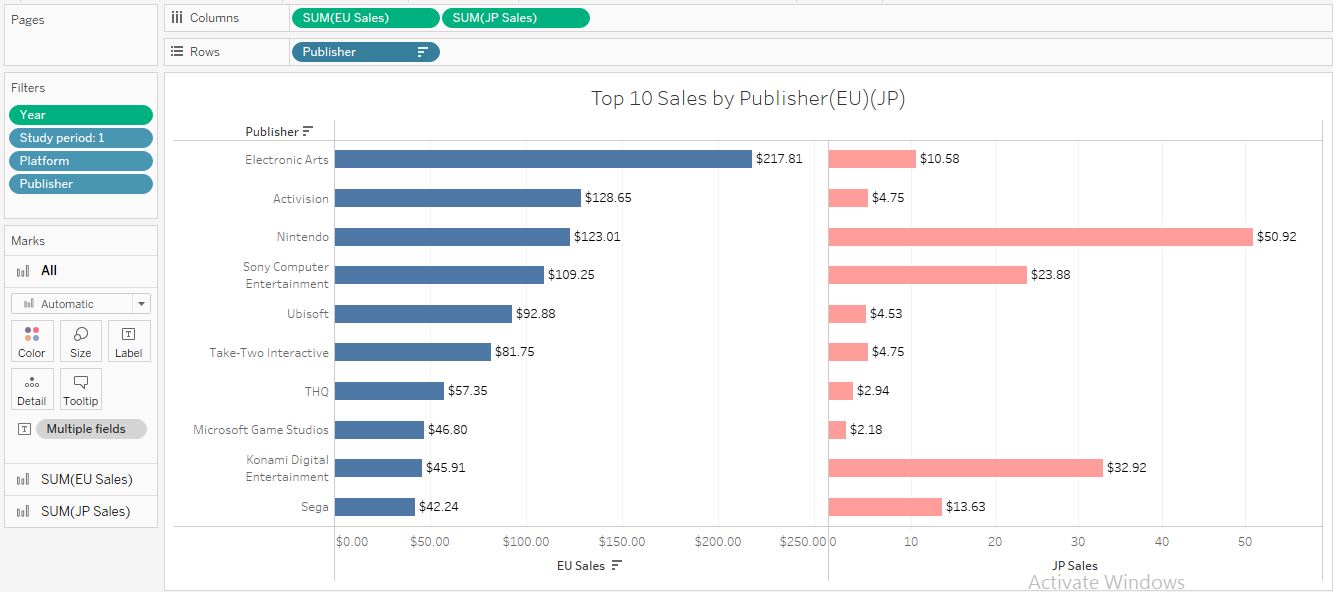


Figure no:8 Comparing European and Japanese Publisher Top 10 Sales.

The graph above is depicting the top 10 sales by publishers of Europe and Japan. The sales per publisher in Europe and Japan are quite different. The top three publishers in Europe are electronic arts, Activision, Nintendo, Sony opposed to Nintendo, Konami Digital, Sony Computer Entertainment.

It can be noted electronic arts are the top publisher in Europe with a sale of $217.81 whereas in Japan the top publisher is Nintendo with a sale of $50.92. In Europe Sega has the least sales amounting to 42.24 whereas in Japan the least sales of $2.18m can be noted in Microsoft Games Studio. A striking contrast is also depicted in Activision sales in Europe amounting to $128.65m whereas in Japan sales are only $4.75m.

## 9. Comparing Top 5 Genre Sales of EU AND JP sales.

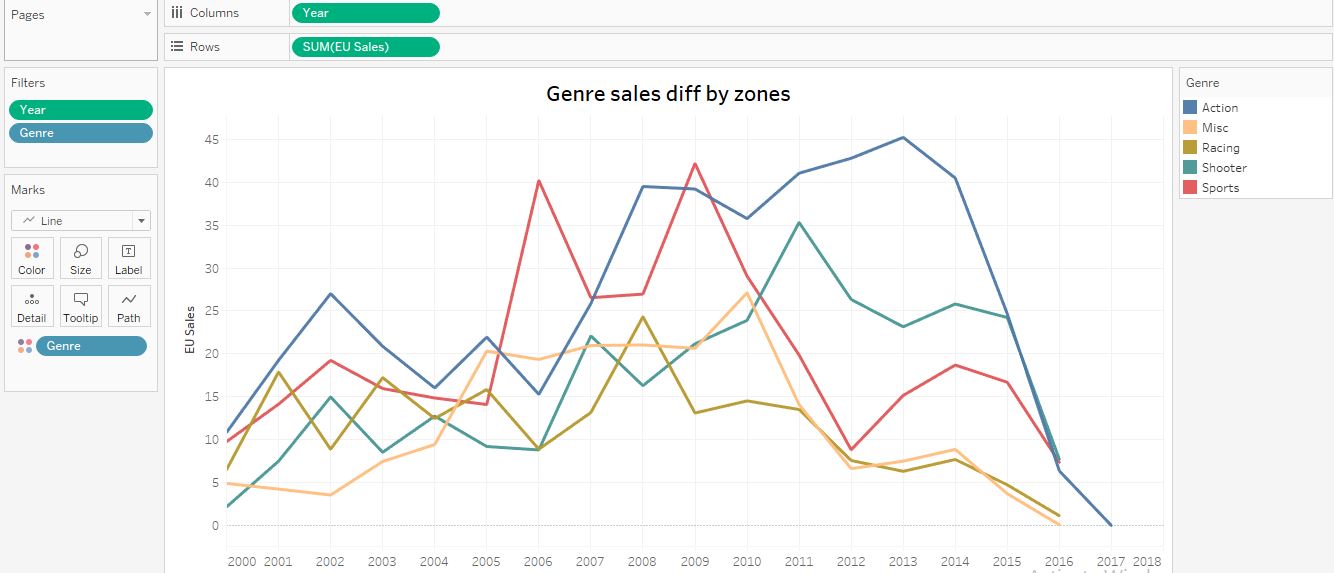


Figure no:9 Top 5 Genre Sales of EU

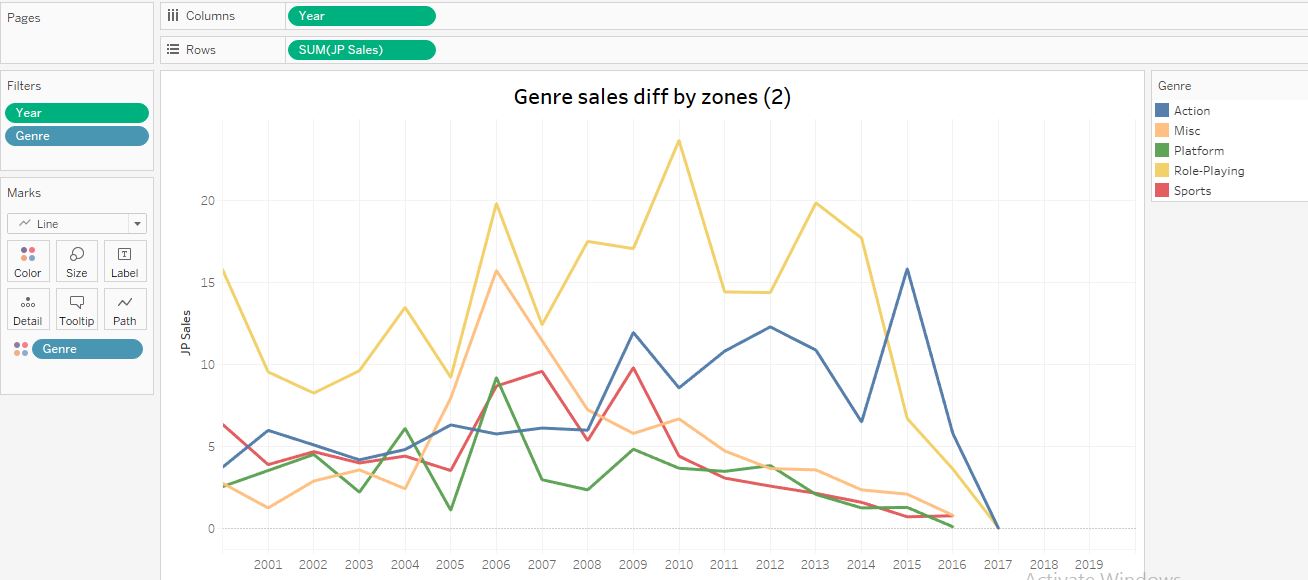


Figure no:9.1 Top 5 Genre Sales of JP

The graphs above show the top five genres in terms of sales in both Europe and Japan. In the Europe it is observed that action was the most chosen genre, evidenced by an increase in sales in 2002, 2008, and 2013 totalling $26m, $40m, and $45m correspondingly. On the preceding graph for Japan, it is possible to see that role playing was the most popular genre, with its peak sales in 2010 totalling $30m.

3.0 Selection of Data Mining Algorithm and Data Pre-processing

I have decided to utilise a Time Series Algorithm as it works well with the dataset selected.

Analysing a set of data points gathered over a period using a technique called time series analysis is a specialised form of data analysis. Instead of recording the data points sporadically time series captures them at regular intervals over a predetermined amount of time.

This algorithm is used for non-stationary data, which refers to items that are constantly fluctuating over time or are impacted by time, and that is evident in my dataset. Thus, it works well for my dataset because it is utilised for non-stationary data. On the other hand, because sales are continually changing, the most effective method is to apply time series analysis.

While we are about improvements I have made to my dataset, I should mention that I've included four new columns. These columns help me comprehend the upcoming forecasts as they provide a clear view of the critic scores given by critics to the games and the user score and counts. These factors have a direct impact on popularity and sales of video games, therefore, are important to predict future trends. Below are the columns that were added.

|  |  |
| --- | --- |
| Critic score - | Total rating compiled by members of the Metacritic crew |
| Critic count | Number of reviewers utilised to calculate Critic score |
| User score | Score determined by the users who subscribe to Metacritic. |
| User count | The total number of users who contributed to the user score. |

## Data Mining

## 3.1.1 Trend Lines

## 1. Global Market sales.

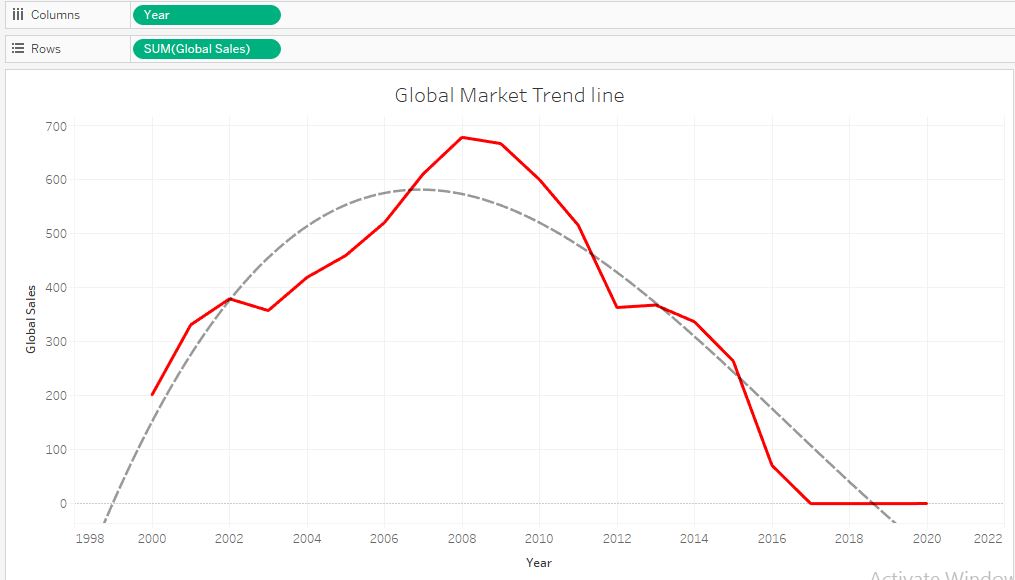


Fig no:1 Global Market Sales trend.

The graph above depicts the trend in total global sales of video games over the period 2000 to 2020. It can be noted from the above that the R value is 0.862 and a Polynomial trend line best suits this data set. Overall, a declining trend can be seen in the global sales of video games.

## Sales Trend by Each Market

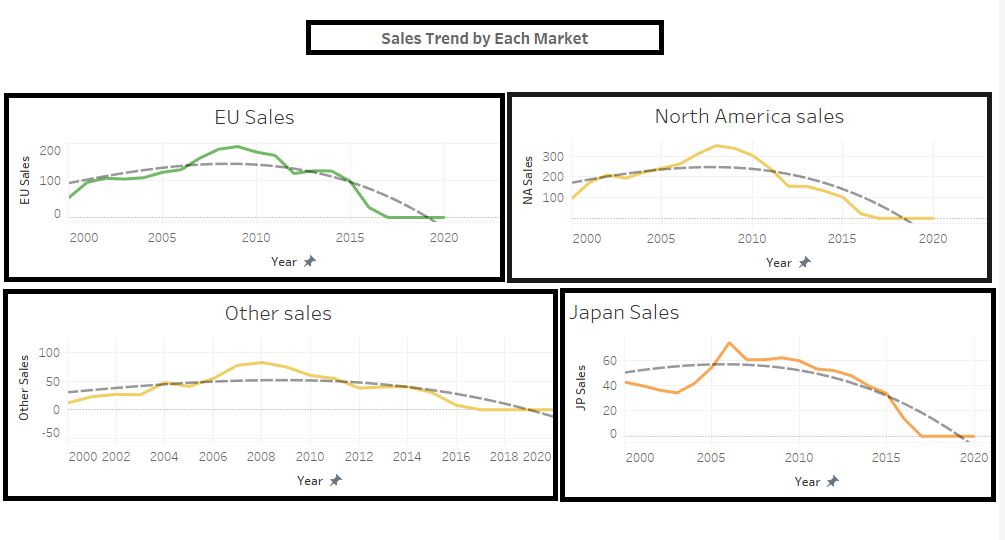


Fig no: 2 Trend by All Market Sales.

The graph above depicts the sales of video games in all four markets zone across the period 2000 to 2020. The following was noted:

A minor uptick in sales of video games can be noticed through the year 2009 in Europe, but a subsequent declining trend can be traced all the way through 2017. This set of data is best described by a polynomial curve.

In Japan the sale of video games experiences a moderate decline up until the year 2009, after which point, they continue on a downward trajectory until the year 2020. The R score of 0.819 indicates that the polynomial trend line provides the best fit for this data set.

In North America there is a rise in the sales of video games up until the year 2008, but after that, a declining tendency is visible all the way up until the year 2020. The R value for this data set is 0.8149, which indicates that the polynomial curve provides the best fit.

For other regions the sales of video games rose steadily up until 2008, and then declined steadily through 2020. With an R-value of 0.765, the polynomial curve provides the best match for these numbers.

## Year platform SALES of Market (NA/JP)

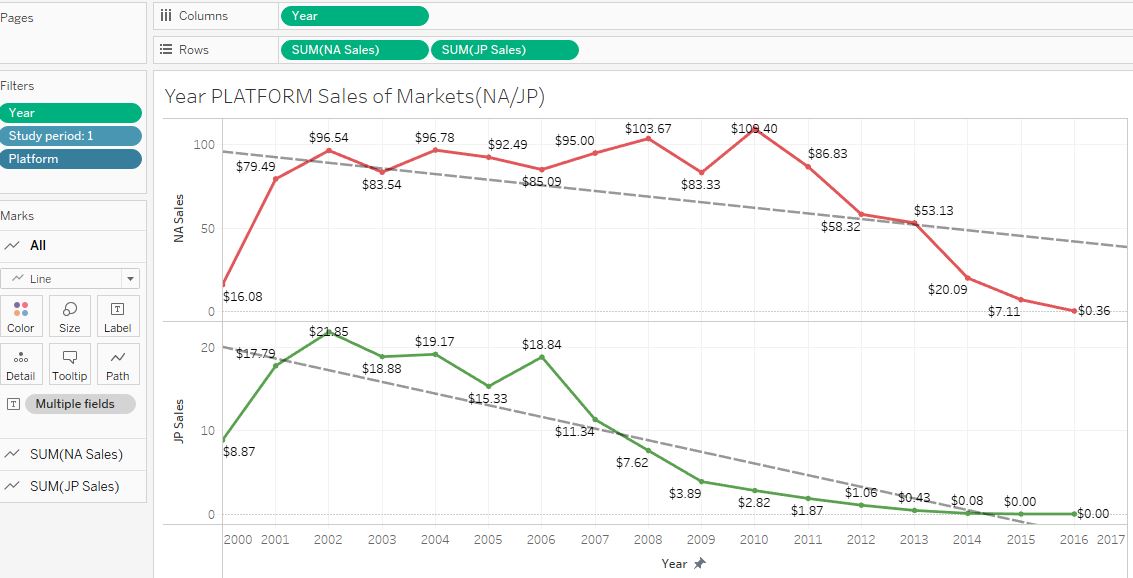


Fig no:3 Platform sales in Markets of Japan and North America Sales trend.

The above graph depicts the sales as per platform in North America and Japan. Although there are slight increases in sales for both regions an overall downward trend is noted in sales to the year ended 2017. Due to the continuous downward trend a linear regression best fits the data set.

## Sales by year and genre of Japan and European Market

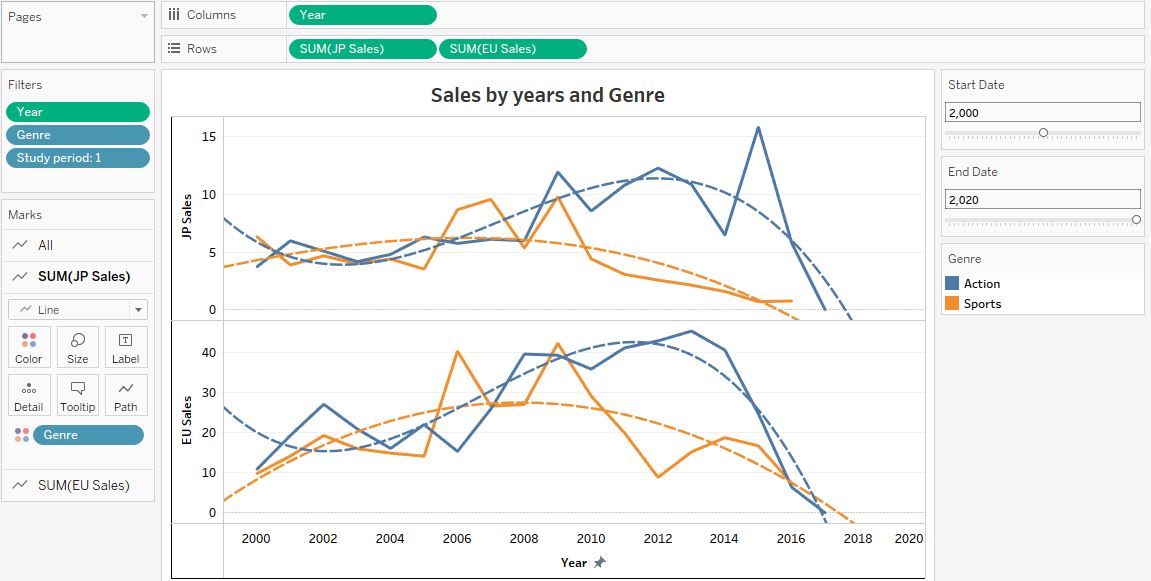


Fig no: 4 Sales by year and genre of JP and EU

The above chart shows sales broken down by the top two genres in both the Japanese and European markets. In Japan sales declined steadily until 2003, only to start climbing again in 2012. Additionally, sales have been declined through 2018. Comparatively, sales in the sports industry increased until 2005, and then declined. Based on the information above, it appears that a polynomial curve best fits the data for both Action and Sports, with a R value of 0.5838 and 0.5148, respectively.

In Europe, sales of action decreased until 2003, and then increased to 2012. Additionally, a decline is seen continuing through to 2018. Sales of sporting goods peaked in 2006 and have been on the slide ever since. The data are best fit by a polynomial curve (R = 0.8139 for Action and 0.450 for Sports).

# 3.1.2 Forecast

## Global sales Forecast

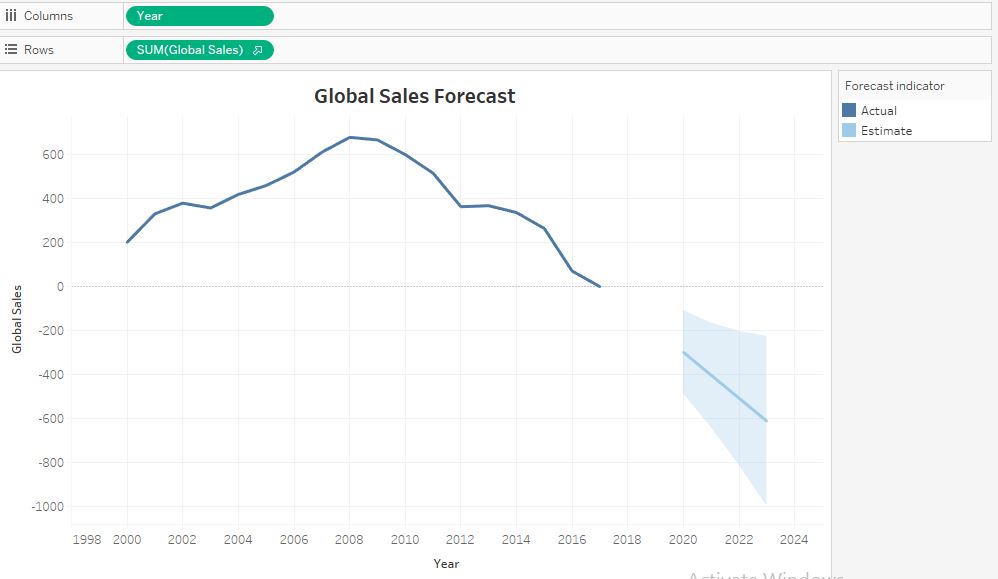


Fig no: 5 Global sales Forecast

The above graph depicts the overall trend in video game sales globally. Initially an inclining trend is observed however, after 2008 a declining trend is noted. Based on the actual declining trend a negative trend is further estimated up to the years 2023. From the above we can summarize that the focus from video games is slowly diminishing in the years to come.

## Yearly sales Forecast for all Markets:

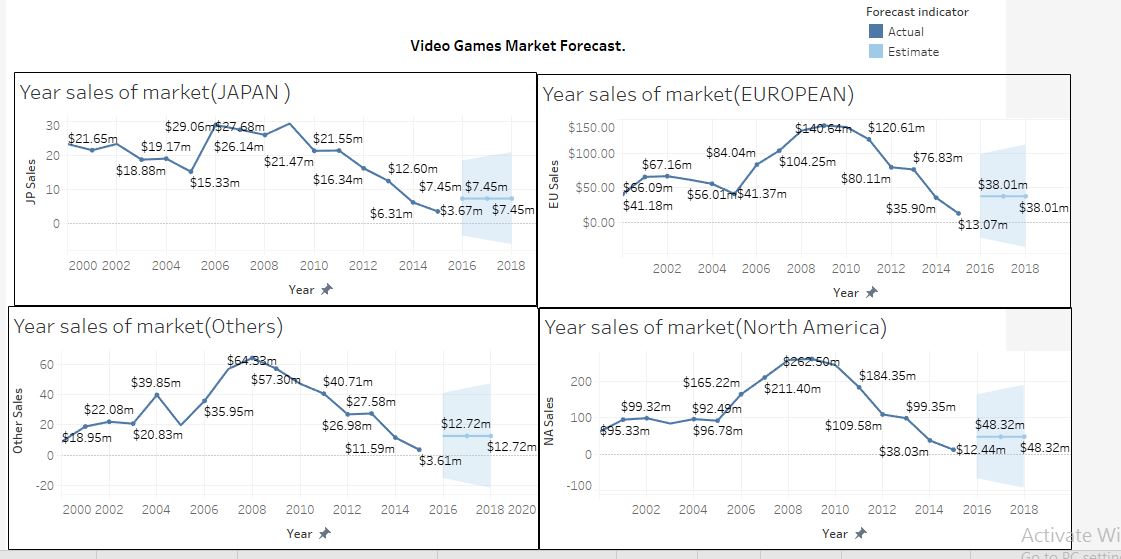


Fig no: 6 Forecasting of All 4 Markets

Sales projections for video games across Europe, Japan, North America, and other regions are shown above. Based on historical sales there an annual growth in sales is forecasted to $38m in Europe, $7.45m in Japan, $12.72m in other regions and $48.32m in North America.

## Top 2 Games Publisher with Critic Score.

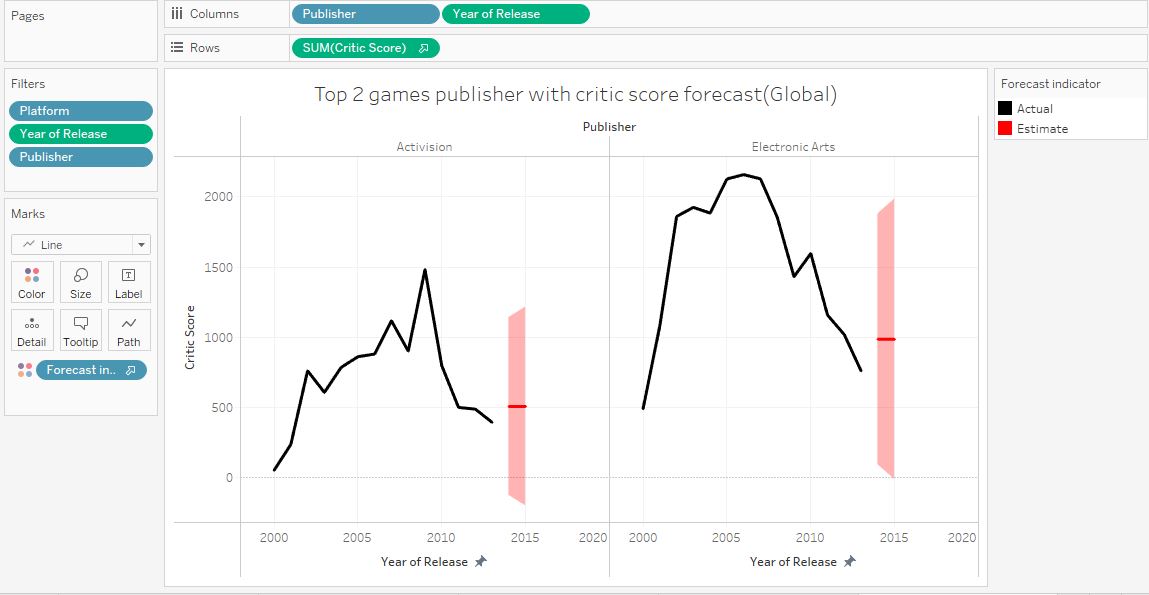


Fig no: 7 Best 2 Games Publisher with Critic Score.

The graph illustrates the critic scores against year of release for the top 2 games publishers. It can be noted that Activision and Electronic arts are the top 2 publishers. Critic score is a key dependent in assessing popularity of games and publishers as feedback from critics will help develop the market. It can be seen from the above that critic score is estimated to 500 and 1000 for Activision and Electronic arts respectively.

## Comparing Between European and Japan Yearly sales of Genres.

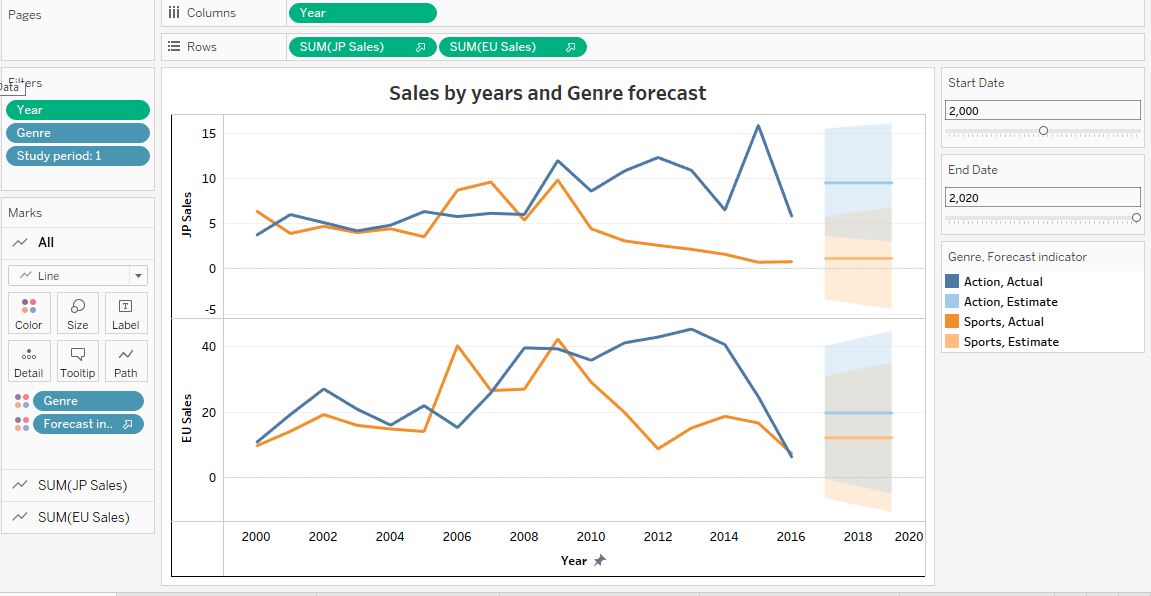


Fig no: 8 Comparing Between EU and JP Market Yearly sales of Genres.

The graph that you just looked at illustrates how the actual sales in the Europe and Japan compare to the anticipated sales for the top two genres, which are action and sports. The action and sports categories are expected to bring in $10 million and $1 million in sales, respectively, in Japan.

The sales forecast for the European Union is approximately $20 million for the action category and approximately $16 million for the sports category.

# 4.0 Data Ethics

Ethical, professional, and legal considerations are the three key areas that Data Ethics encompasses. So, I will elaborate on all three perspectives

## Ethical Consideration

Analysing data include acquiring, processing, and evaluating data. Ethical considerations in data analysis are important for preventing harm. Data analysis ethics involve thinking about things like: Privacy, Bias, Data accuracy and Data security.

Privacy: Data analysis ethics revolve around privacy. Individual data may include sensitive information like health, finances, and views. Data analysts must keep this information private so not to hurt others.

## Professional Considerations

Data analysis calls for a wide variety of professional considerations, including experience and competence, ethical behaviour, constant learning, communication skills, project management abilities, and technological skills.

## Legal consideration

Data protection, intellectual property, data retention, consumer protection, data breach notification must be considered when conducting data analysis. Data analysts are legally obligated to act in a responsible and ethical manner when analysing data.

While review my data I have taken the following into consideration:

Privacy: I ensured that no personal data was being used in my analysis. I removed such information before commencing my analysis.

Data Security: I ensured that player data is stored securely. This includes implementing appropriate security measures, such as encryption and access controls, and regularly monitoring and updating security protocols on my laptop.

Bias: Video game datasets may be susceptible to bias, such as racial or gender bias. While reviewing the dataset I took steps to ensure that there was no biasness in the data set.

Data Protection: As the dataset includes sensitive information, such as location data or biometric data I ensured that this data is protected and used only for its intended purposes, and that it is not used for discriminatory practices.

# 5.0 Conclusion:

In conclusion, we noted several significant observations. Sales reached their peak in the late 2000’s and have been declining since.

Historically, action games have established themselves as the premier category of video games; however, it is unknown if this pattern will continue into the indefinite future. Japan, which is one of the regions with the highest single-year income, has a gaming culture that is based primarily on role-playing games and has a significant amount of anime influence. It is likely that this will continue to have an impact on the popularity of specific video game titles, platforms, genres, and other categories that are geared more towards this sector of the video game industry.

Several discoveries with annual sales figures and findings are included below.

* The region with the greatest average sales is North America, amounting to $264,667,430.
* The Microsoft X360 is the most popular gaming system in North America, with $601.05 million in sales.
* The PlayStation 3 (Sony) has amassed $343.71 million in sales, making it the most popular system in Europe.
* Consumers in Japan have spent a whopping $175.57 million on DS (Nintendo) consoles.
* Sony's PS2 is the most popular console worldwide, with sales of $193.44 million.
* The most popular game of all time is Wii Sports making a total of $82.74 million.
* The most popular game in North America, Europe, and several other places is Wii Sports.
* The most popular video game in Japan is the Pokémon Red/Blue duo.
* Released in 1985, Super Mario Bros. is the top ancient game still producing $40.24 million in global sales. With Tetris, Pokémon Red and Pokémon Blue to follow.
* In terms of worldwide revenue, action games bring in $1751.18 million.
* The Best publisher in North America ($816.77M), Europe ($418.74M), and Japan ($455.42M) is Nintendo.
* Electronic Arts has made $129.77 million as the top publisher for all other territories combined.

As there is a trend towards fluctuation in sales, the time series model has proven to be a good fit with the data set that I have. The polynomial regression curve has been shown to be the best fit for the data set when considering the trends that have emerged throughout the years.

Tableau enables a clear visual representation of patterns and trends resulting in widespread adoption by major corporations. This information might be utilized by senior management of organizations to analyses trends associated with their products, which would be of assistance to them in the process of formulating their strategy. The same information is utilized to make decisions about day-to-day operations at the managerial level. Example daily sales can help managers decide consumer preferences, while at a senior management level, it can be used to make decisions about introducing more and broadening the production line.

# 

# References:

* <https://www.studysmarter.co.uk/explanations/english/research-and-composition/ethical-issues-in-data-collection\> [Accessed on: 03/28/2023]
* <https://help.tableau.com/current/pro/desktop/en-us/trendlines_add.htm>

[Accessed on: 04/01/2023]

* <https://www.tableau.com/en-gb/learn/articles/data-visualization> [Accessed on: 03/10/2023]
* Entertainment Software Rating Board <https://www.esrb.org/> [Accessed on: 04/08/2023]
* <https://www.bluestacks.com/blog/news/global-gaming-market-reaches-99-6-billion-in-2016-mobile-generating-37.html> [Accessed on: 04/09/2023]
* <https://www.kaggle.com/datasets/rush4ratio/video-game-sales-with-ratings> [Accessed on: 02/10/2023]
* <https://www.wepc.com/statistics/console-gaming/> [Accessed on: 04/10/2023]
* <https://www.wepc.com/news/video-game-statistics/> [Accessed on: 04/12/2023]
* Robert H. Shumway, David S.Stoffer. (2017). Time Series Analysis and its Applications 4th edition. America. Springer.