A picture containing text

Description automatically generated  
**Muma College of Business**

School of Information Systems and Management

**Data Visualization**

**ISM 6419 Spring 2023**

Prof. Johannes Han Reichgelt, PhD

**Final Project Report**

**Video Game Sales**

By

Prasanna Kumar Tummala

U92076820

Business Analytics and Information Systems

**Introduction:**

Over the past few decades, video games have gained popularity as a form of entertainment, and the industry now brings in billions of dollars annually. The development of new technologies and platforms has significantly increased the video game market, making a wide variety of games accessible to players all over the world. Utilizing data visualization techniques, this project seeks to investigate the sales of video games across various platforms, genres, and geographical locations.

We can learn a lot about the gaming industry, the most popular platforms and genres, and the regional differences in sales by examining data on video game sales. With this knowledge, gamers can better understand the gaming market and their personal preferences while game developers and publishers can decide which kinds of games to produce and sell.

Video game sales data analysis and visualization are at the heart of this project's background. Over the past few decades, the video game industry has grown significantly, Since the advent of mobile gaming, virtual reality, and esports, the video game industry has experienced rapid expansion.

This project's goal is to examine global video game sales statistics from various areas in order to learn more about the most popular titles, studios, genres, and platforms. We can determine market trends and patterns and design strategic game development, marketing, and sales plans by evaluating this data.

To provide background and insights on the video game industry, external sources including market research papers, blogs from the gaming industry, and news items have been cited.

The project on video game sales can be categorized as something ambitious. Analysis of a large data set comprising data on video games, including information on their sales, publishers, ratings, genres, platforms, and developers, is required.

The project aims to answer several questions, such as

1. How much sales are going on each region from the past years.

2. How the publisher is doing and how many sales are done by the top 10 publishers.

3. What are the Top games that users are enjoy playing them. What is the rating they gave for each game?

4. What are the Top genre of the games that people are playing. And on which platform users are playing in.

5. Who are the game developers around the world and when they have established and from which country they are from.

While these questions are not overly complex, the project requires a significant amount of data cleaning, processing, and visualization, as well as the use of statistical and analytical tools to gain meaningful insights. Therefore, the project can be considered moderately ambitious, as it requires a certain level of technical expertise and time commitment to complete successfully.

**Methodology:**

For the Video Game Sales project, the primary data source used is a public dataset obtained from Kaggle via <https://www.kaggle.com/datasets/rush4ratio/video-game-sales-with-ratings>, which includes information on global video game sales from 1980 to 2016. The dataset contains various attributes, such as the game's title, platform, publisher, release year, genre, sales in different regions, and user ratings. In addition to the primary dataset, I have also used other external data sources from data world platform accessible via <https://data.world/mhoangvslev/steam-games-dataset> which includes other data tables of developer, publisher and platform.

The datasets had been analysed as part of the project to ensure reliability and uniformity because they had a few discrepancies that needed to be plugged in. To accomplish this, I cleaned the data and lined in every value that was missing using both Tableau and Excel respectively.

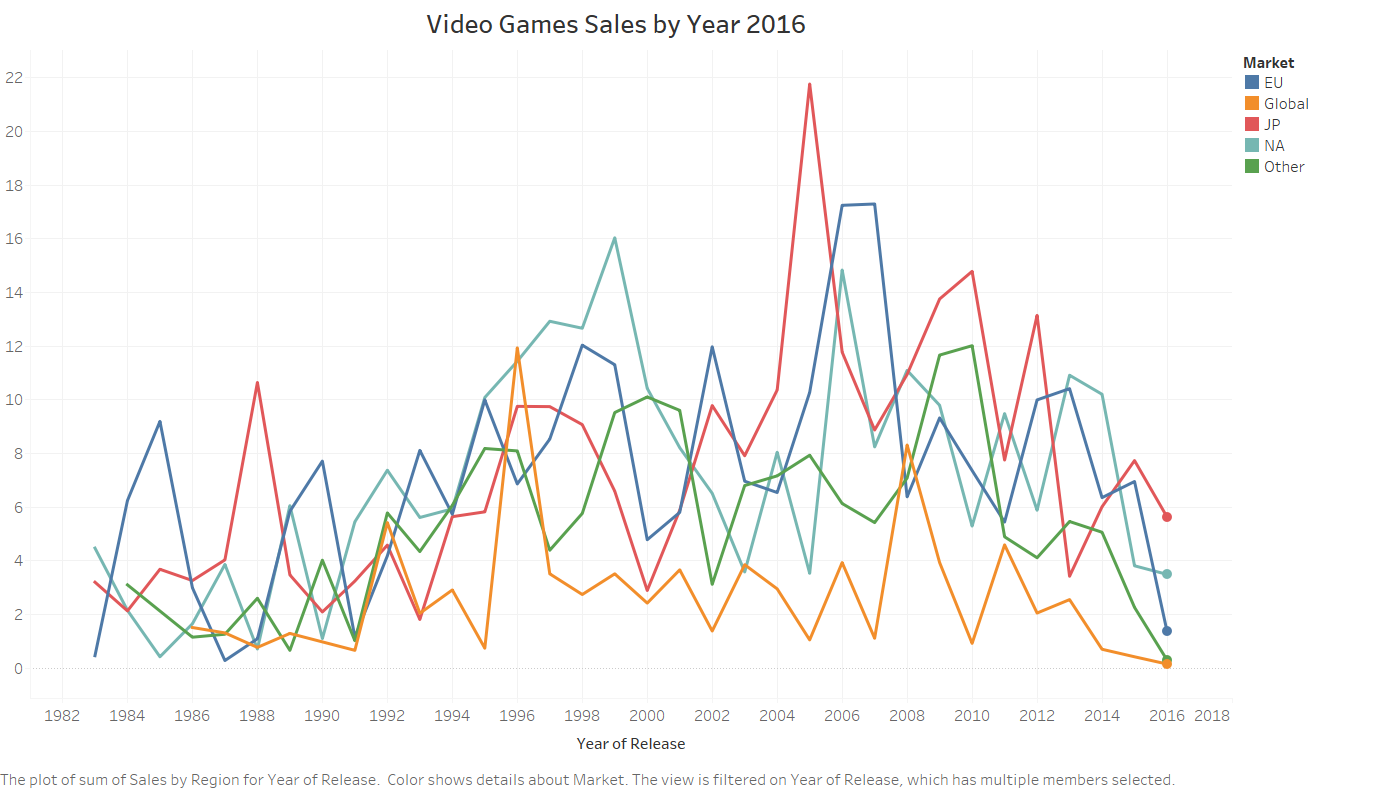
The Developer data consists of various attributes such as, Developer, Active, City, Country, Est., Game, Notes. The Publisher data consists of attributes, Publisher, Headquarters, Est., Games published, Notes, Active. The Platform data consists of attributes, Initials, Platform, manufacturer, gen.

The primary dataset from the Kaggle consists of attributes, game's title, platform, publisher, release year, genre, sales in different regions, and user ratings. Initially the dataset is with null values and missing values and data is uncleaned. So, I have cleansed the dataset with the following changes:

1. I eliminated every null value from all 4 data sets.
2. I have created the column market which consists of region category (NA, EU, JP, Global)
3. Later I have split the market with the help of delimitator and removed \_sales from the market column in dataset.
4. The same was done with the user\_score and critic\_score; I transposed them into rows, divided the text, and removed the \_score column.

**Analysis:**

**1. How much sales are going on each region from the past years.**

****

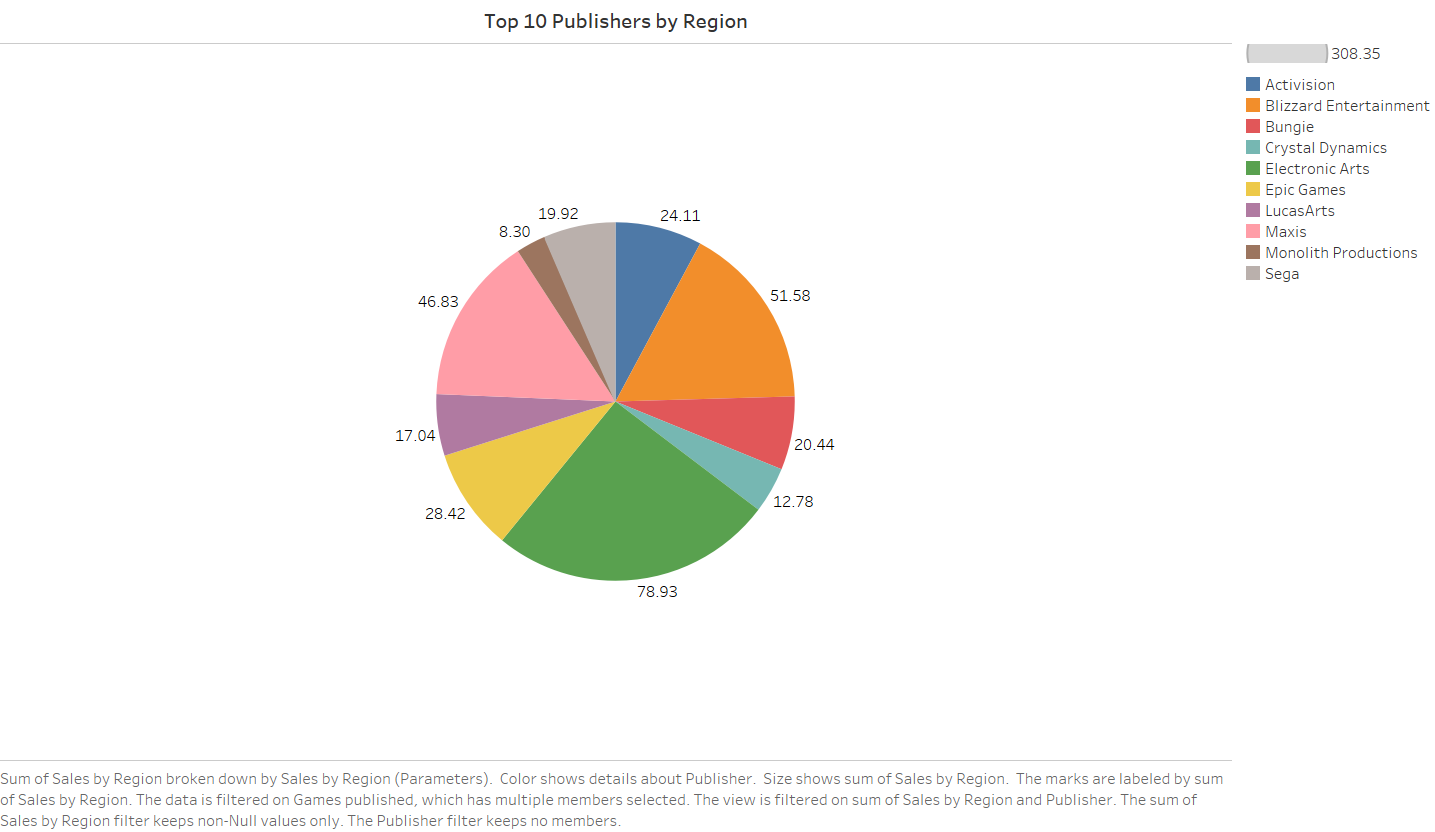
From the above visualization,

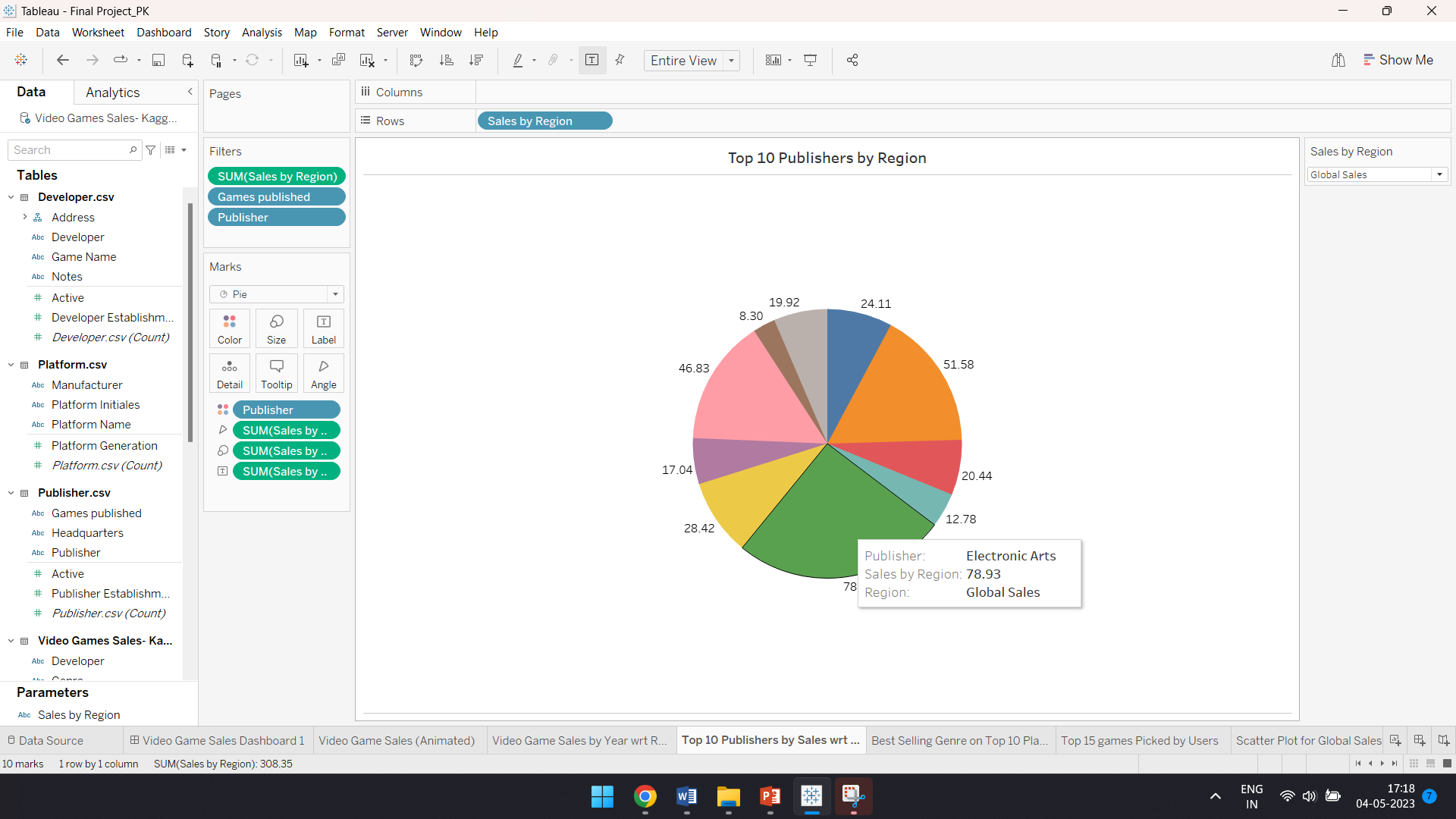
To visualize the sales of video games in each region, a line graph was created with the x-axis showing the year of release and the y-axis representing the total sales in millions of dollars. The chart was further divided into four sub-charts, each representing a specific region, namely North America, Europe, Japan, and other regions.

The Japan region consistently had the highest sales over the past years, followed by Europe and North America, while other regions had comparatively lower sales. The bar chart showed that there was a general increase in sales from 2004 to 2006, with a slight dip in 2011, followed by a continuous decrease in sales from 2012 to 2016.

This information provides insights into the popularity of video games in different regions and how it has evolved over the years. It can help developers and publishers in making strategic decisions on which regions to focus on for their game releases and marketing efforts.

**2. How the publisher is doing and how many sales are done by the top 10 publishers.**

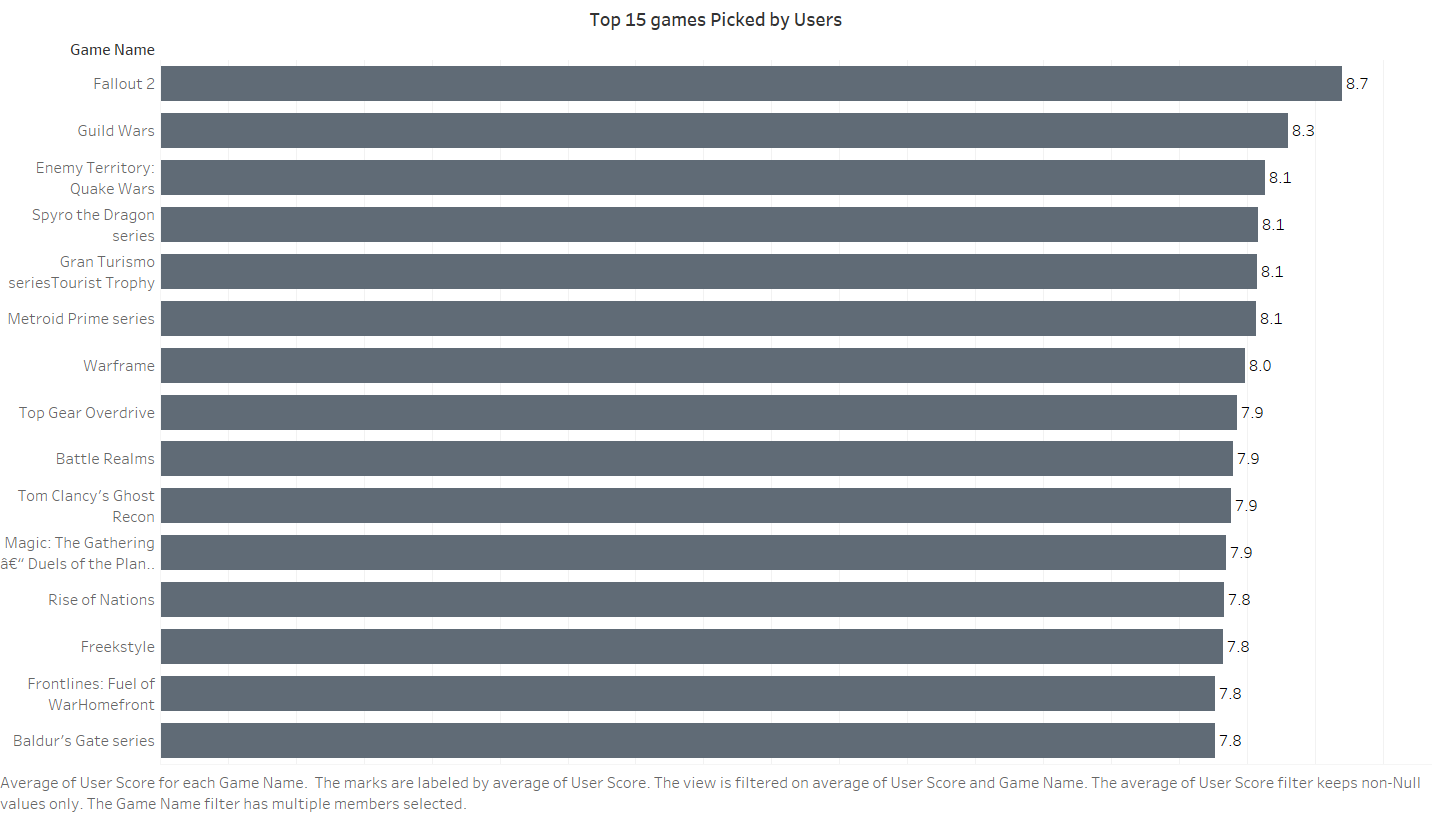




The above visualization represents the top 10 publishers from each region.

This pie chart shows the distribution of video game sales by the top 10 publishers in the industry. As we can see, Electronic Arts (EA) has the largest market share, with 78.9% of all sales, followed by Blizzard Entertainment (51.8%) and Maxis (46.83%). The remaining publishers, including Monolith Productions, Crystal Dynamics, and Sega, each account for less than 20% of the market share. It's interesting to note that the top 10 publishers combined represent about 308% of all video game sales, indicating a highly concentrated market. Overall, this pie chart provides a clear and concise visualization of the dominance of certain publishers in the video game industry.

**3. What are the Top games that users are enjoy playing them. What is the rating they gave for each game?**

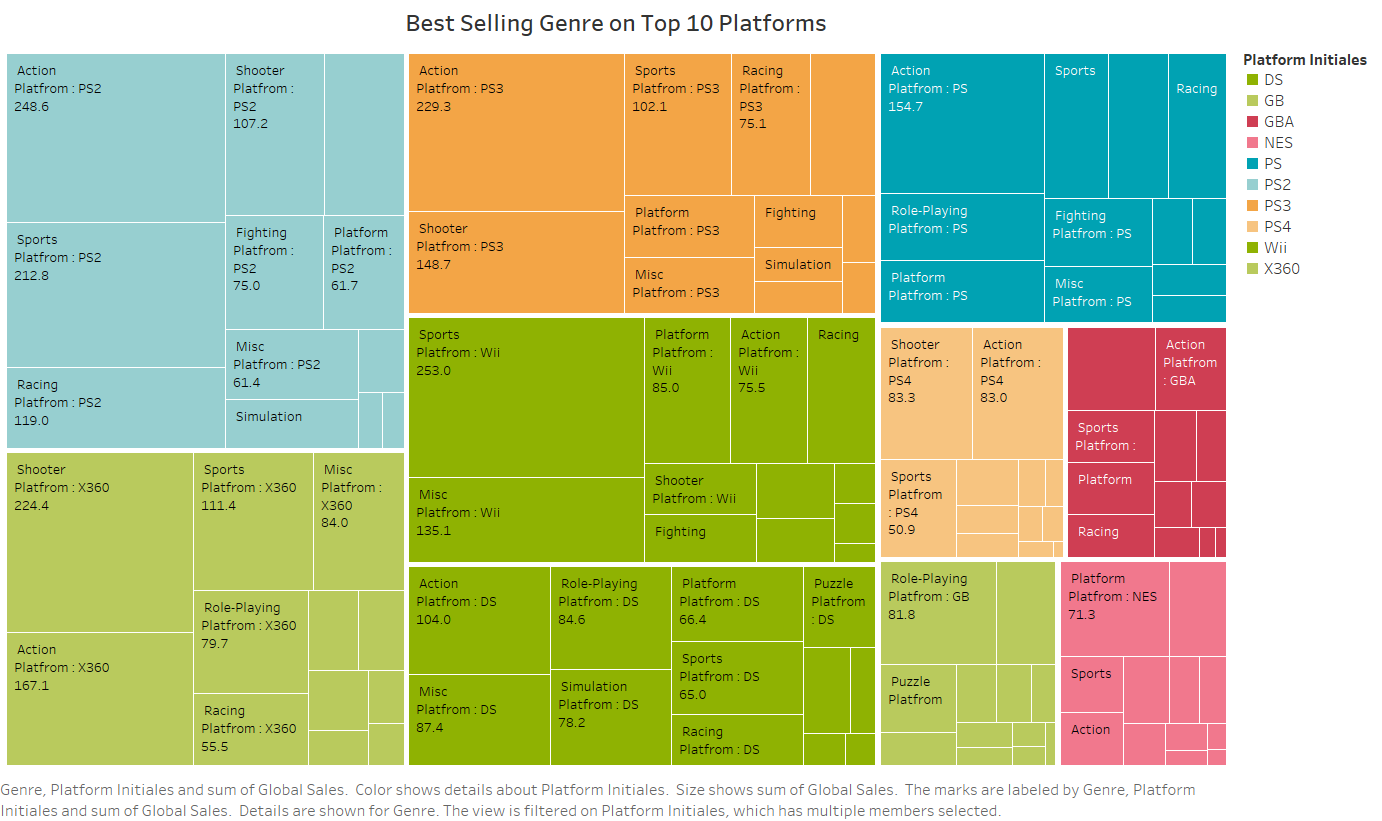
****

To answer this question, I used a bar chart to display the top 15 games based on global sales, along with their ratings. The ratings are displayed as a color-coded bar within the main bar, providing an easy-to-read visualization of both sales and ratings.

From the visualization, it is clear that the top game is "Fallout 2" with an average rating of 8.7 out of 10. Other top games include "Guild Wars", "Enemy Territory", "Quake Wars", and "Spyros the Dragon series".

Overall, the top games have high average ratings, with most falling between 7 and 9 out of 10. The ratings for the top games may be an indicator of their popularity, as users tend to rate games they enjoy playing more highly.

**4. What are the Top genre of the games that people are playing. And on which platform users are playing in.**



To address this research question, I used a combination of a bar chart and a stacked bar chart. The bar chart displays the top genres of games that people are playing, while the stacked bar chart shows the distribution of platforms that users are playing these games on.

The bar chart shows that the Action genre is the most popular, followed by the Sports and Shooter genres. On the other hand, the Strategy and Puzzle genres are the least popular among users.

The stacked bar chart displays the distribution of platforms that users are playing these games on. It shows that the majority of users are playing games on the PS2 platform, followed by the Xbox360, PS3 and Wii platforms. The chart also shows that the Nintendo Switch platform has a significant market share as well.

Overall, these visualizations provide a clear understanding of the top genres of games that users are playing and the platforms they prefer to play them on. This information can be valuable to game developers and publishers who want to tailor their games to specific platforms and genres that are popular among users.

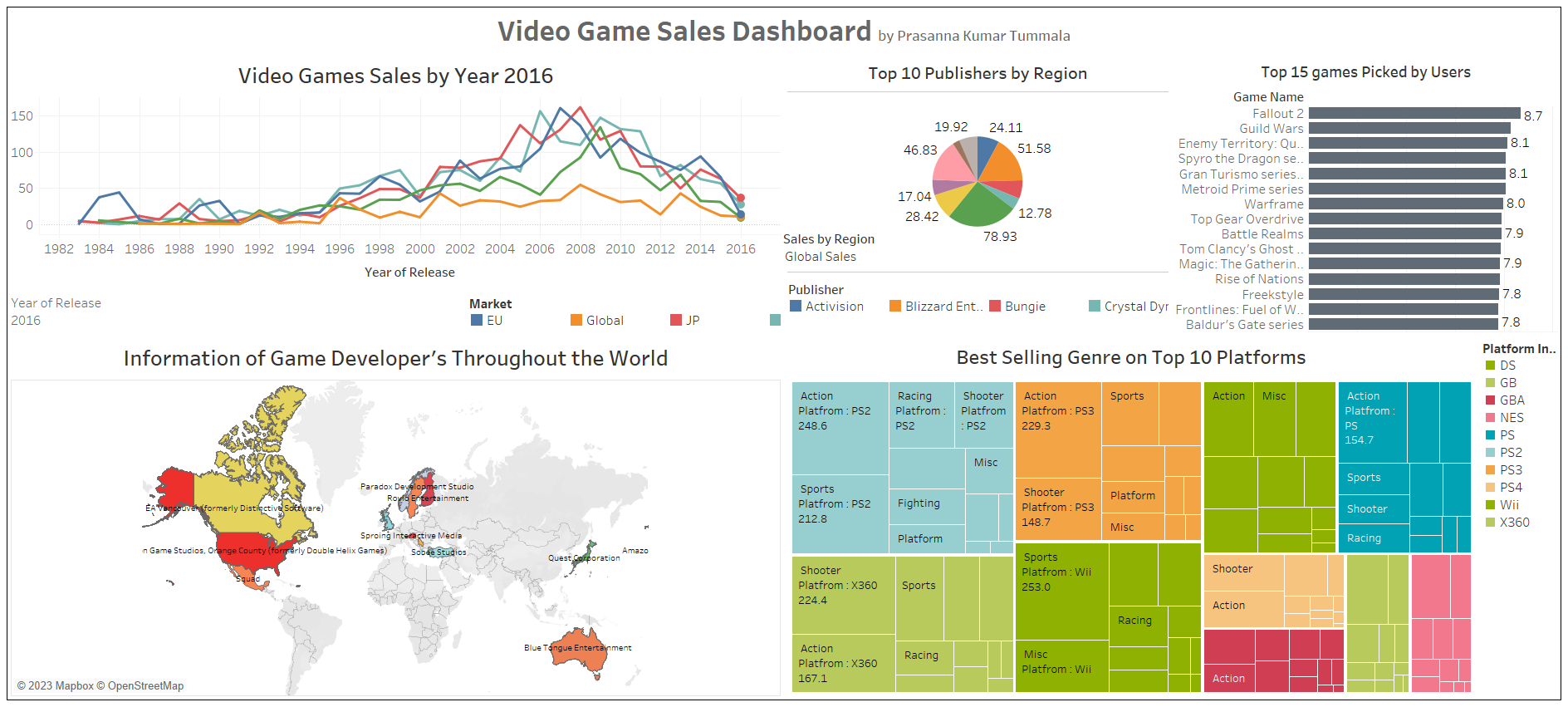
**5. Who are the game developers around the world and when they have established and from which country they are from.**



For this question, I have used the geographical Symbol Map. a symbol map could be used to display the location of game developers around the world. Each developer would be represented by a symbol on the map, with the size and colour of the symbol indicating the year the company was established and the country it is located in, respectively.

By using a symbol map, the user could quickly and easily identify where game developers are located around the world and get an idea of the distribution of companies by country and year of establishment. The visualization could also be interactive, allowing the user to hover over individual symbols to see more detailed information about each game developer.

**Video Game Sales Dashboard:**



The above dashboard has been created to summarize the key insights from the data visualizations. The dashboard includes visualizations for the following research questions:

1. Sales by Region: A stacked bar chart showing the total sales in millions of dollars for each region (North America, Europe, Japan, and Others) over the past years.

2. Top 10 Publishers: A pie chart showing the percentage of sales contributed by each of the top 10 publishers.

3. Top 15 Games: A bar chart showing the top 15 games based on their sales in millions of dollars, along with their user rating.

4. Top Genres and Platforms: A treemap showing the top genres of games, with the size of each tile representing the sales of that genre. Additionally, a horizontal stacked bar chart shows the breakdown of sales by platform for each genre.

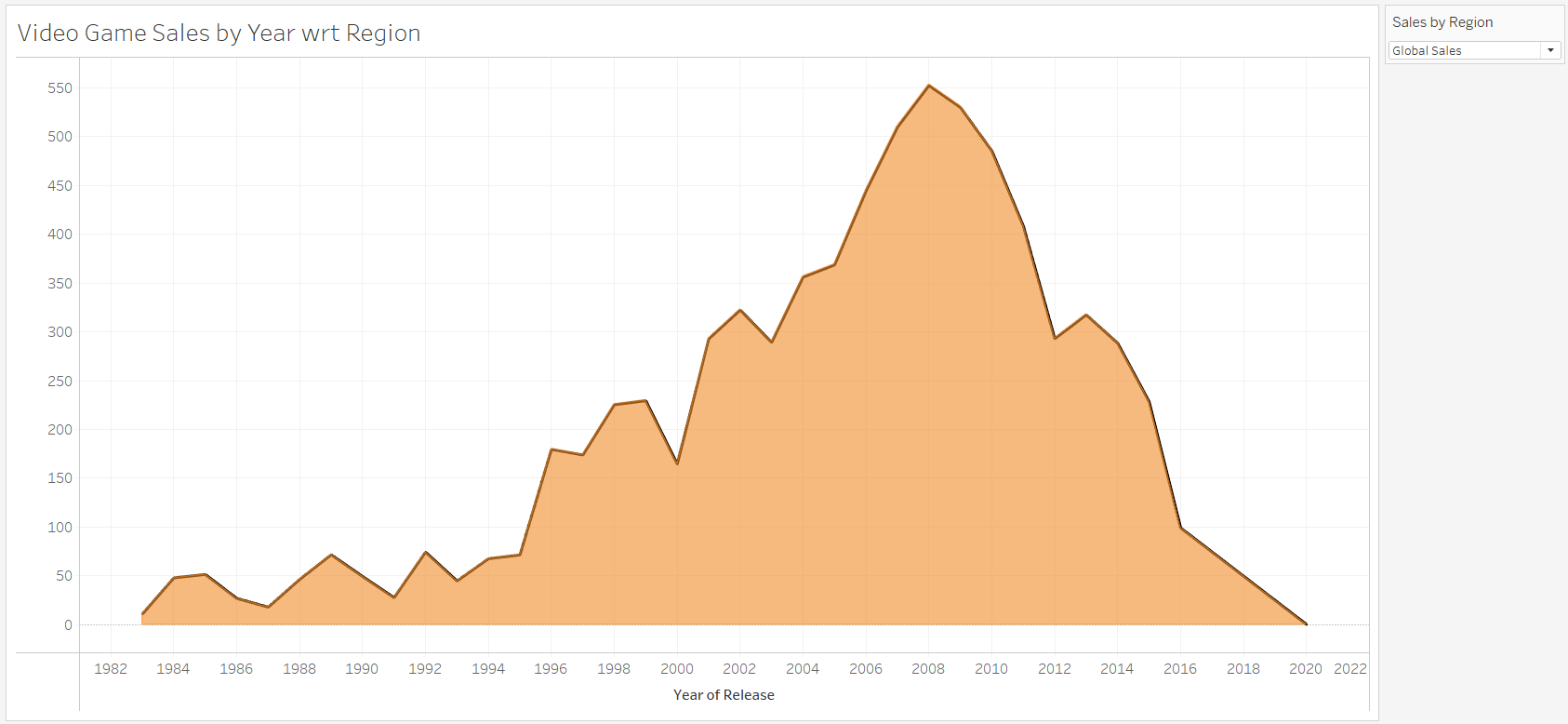
5. Game Developers around the World: A symbol map showing the location of game developers around the world, with the size of the symbol representing the total sales in millions of dollars by developers in that region.

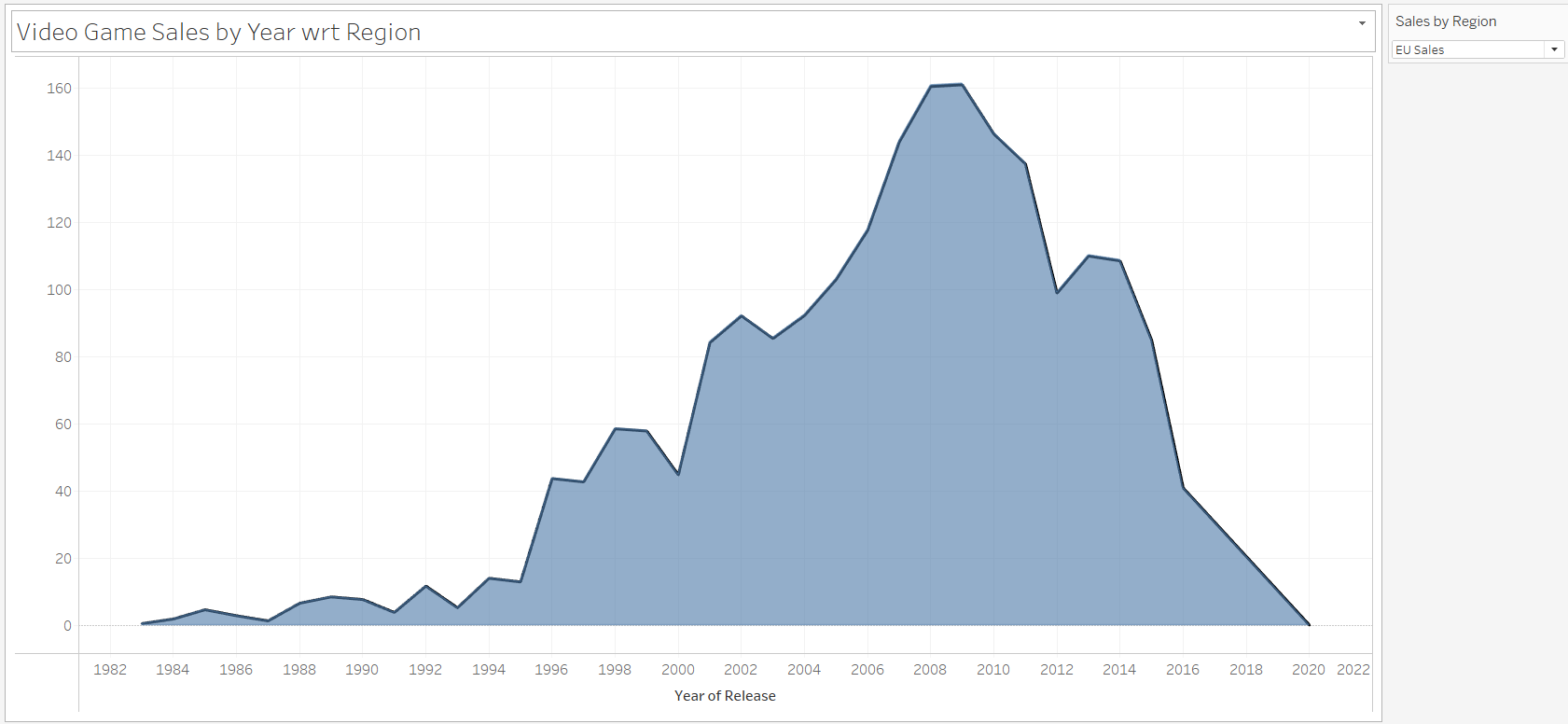
The dashboard provides a comprehensive view of the video game sales data, allowing users to easily explore the insights and trends related to sales, publishers, games, genres, platforms, and developers. The visualizations are interactive, enabling users to filter and drill down to specific regions, publishers, games, genres, and platforms. The dashboard is designed to be user-friendly and accessible to both technical and non-technical users, providing a clear and concise summary of the key findings from the data visualizations.

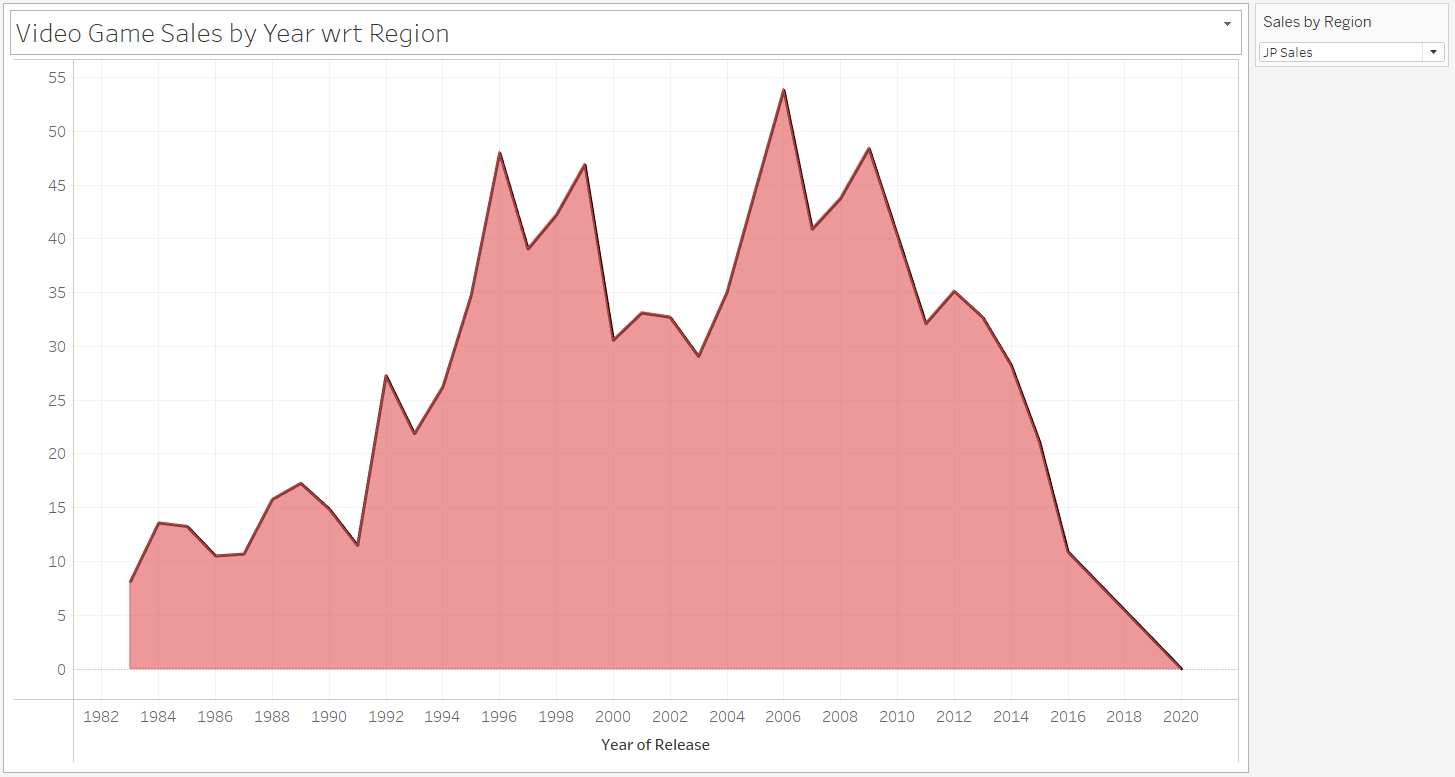
**Additional Visualizations:**

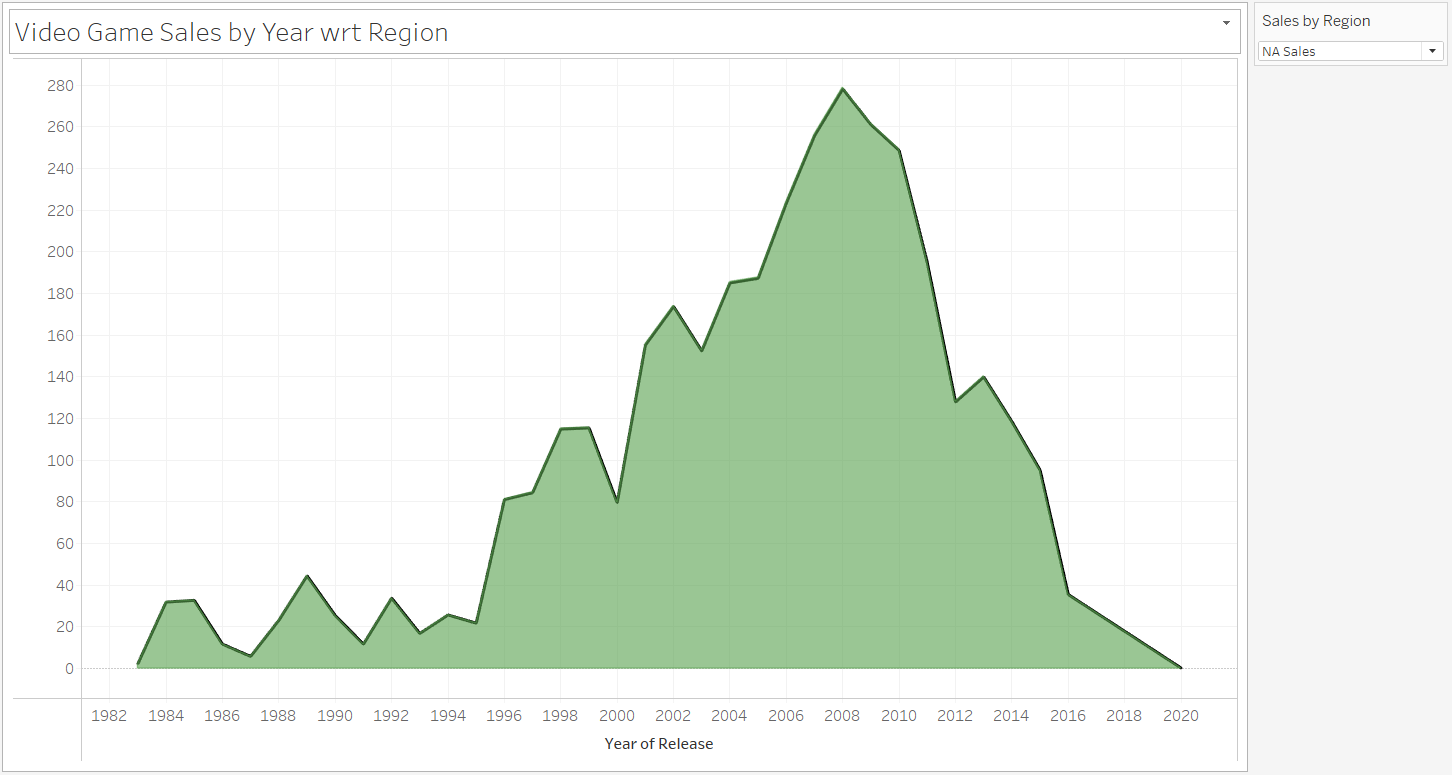
Additionally, I have created few more visualizations:

1. **Video game sales wrt region:**



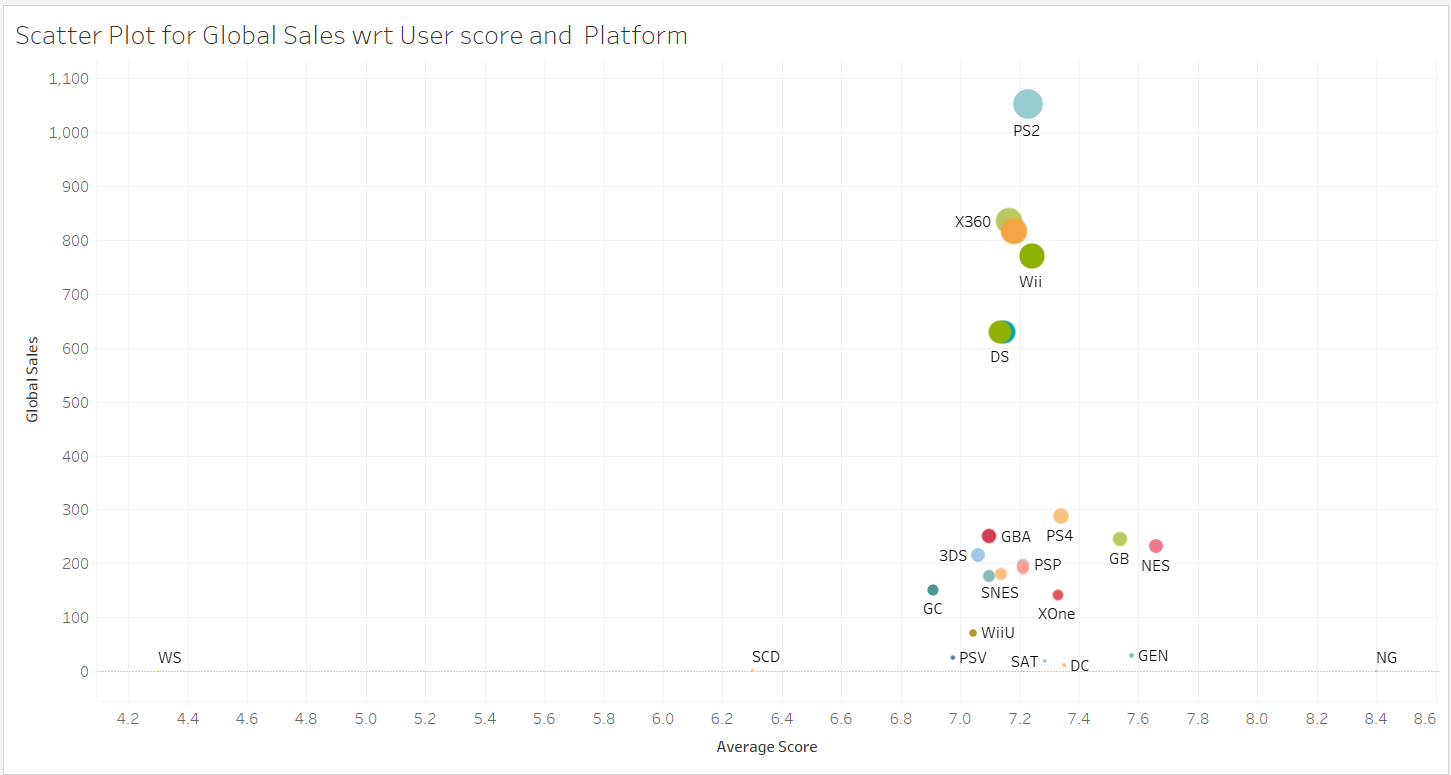






The above visualizations show the video game sales with respective to the parameter (sales by Region). This visualization is combination of line graph and Area graph. I have used the dual axis for this visualization. a line graph was created with the x-axis showing the year of release and the y-axis representing the total sales in millions of dollars. The chart was further divided into four sub-charts, each representing a specific region, namely North America, Europe, Japan, and other regions. So, the graph changes with respective to the parameter.

1. **Scatter plot for Global sales wrt User score and Platform**

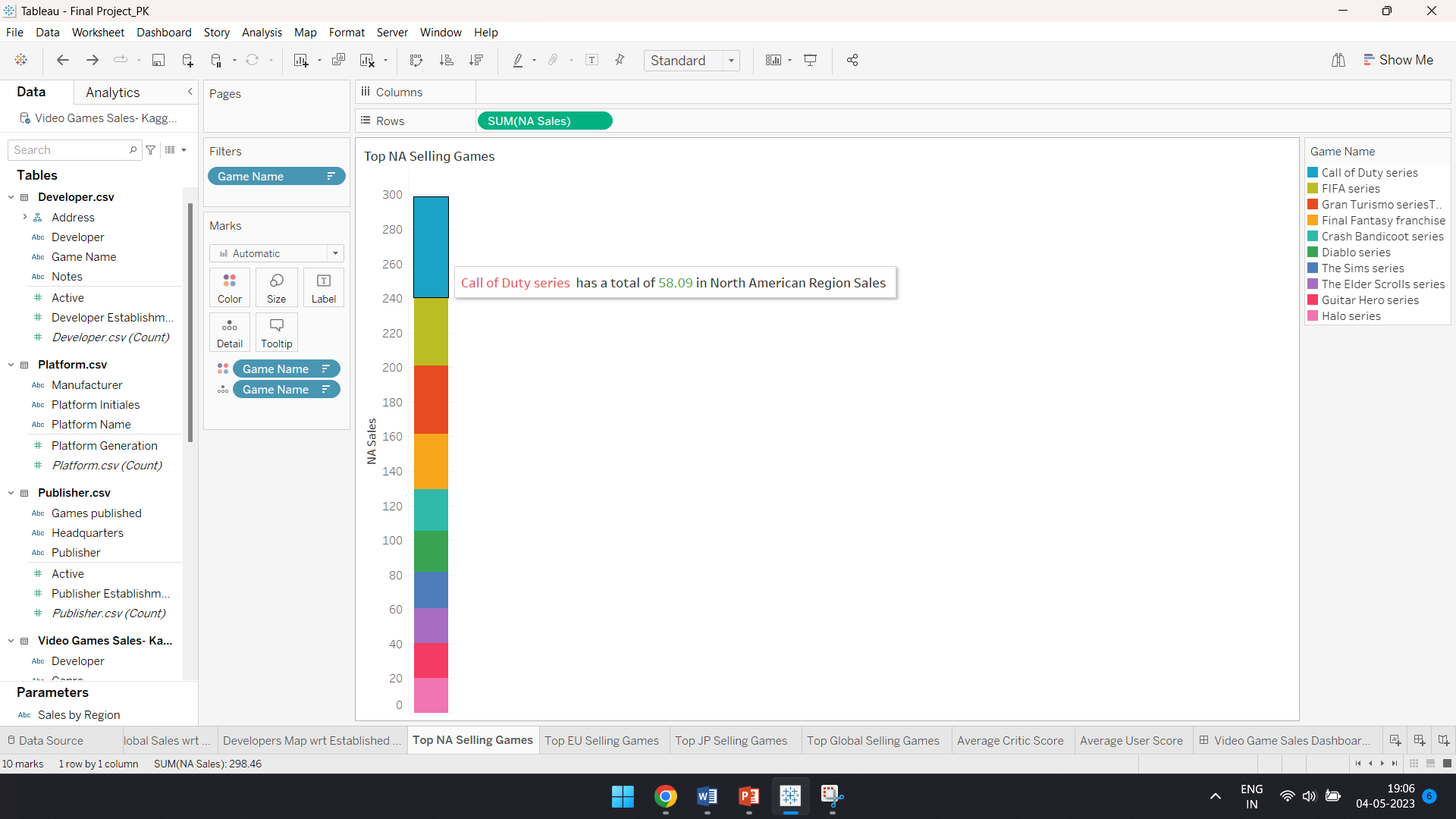


The scatter plot for global sales with respect to user score and platform is a visualization that showcases the relationship between the total global sales of video games and the user scores for each platform. Each point on the plot represents a specific platform and its corresponding global sales and user score. The X-axis represents the user score, while the Y-axis represents the global sales.

The plot shows that there is a positive correlation between user score and global sales, as platforms with higher user scores tend to have higher global sales. It also reveals that certain platforms, such as the PlayStation 2 and Xbox360, have much higher global sales than other platforms, indicating that they are more popular among gamers.

The plot also highlights the variation in user scores and global sales across different platforms, as seen by the varying cluster sizes and dispersion of points. This visualization is particularly useful for identifying which platforms are most popular among gamers and which ones have the highest user satisfaction rates.

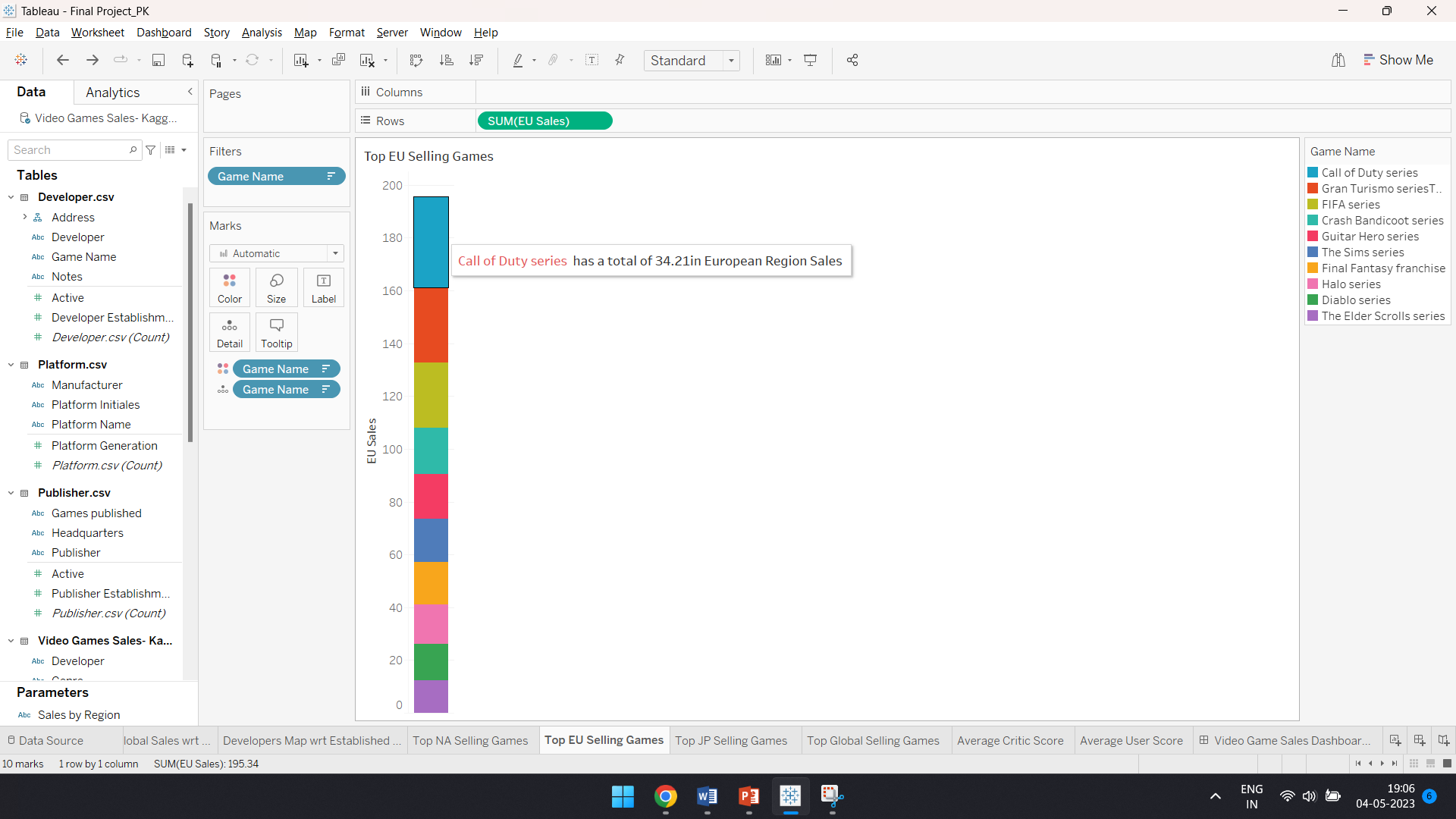
1. **Top Selling Games from North America Region**



This visualization provides a clear picture of the top-performing games in the North American market and the scale of their success in terms of units sold. It could be useful for game developers and publishers to understand the trends in the North American market and the types of games that are resonating with consumers.

From the bar graph, Call of Duty Series has the highest sales with 58.09 million dollars and Halo Series has the least sales with 20.08 million dollars in the North America Region

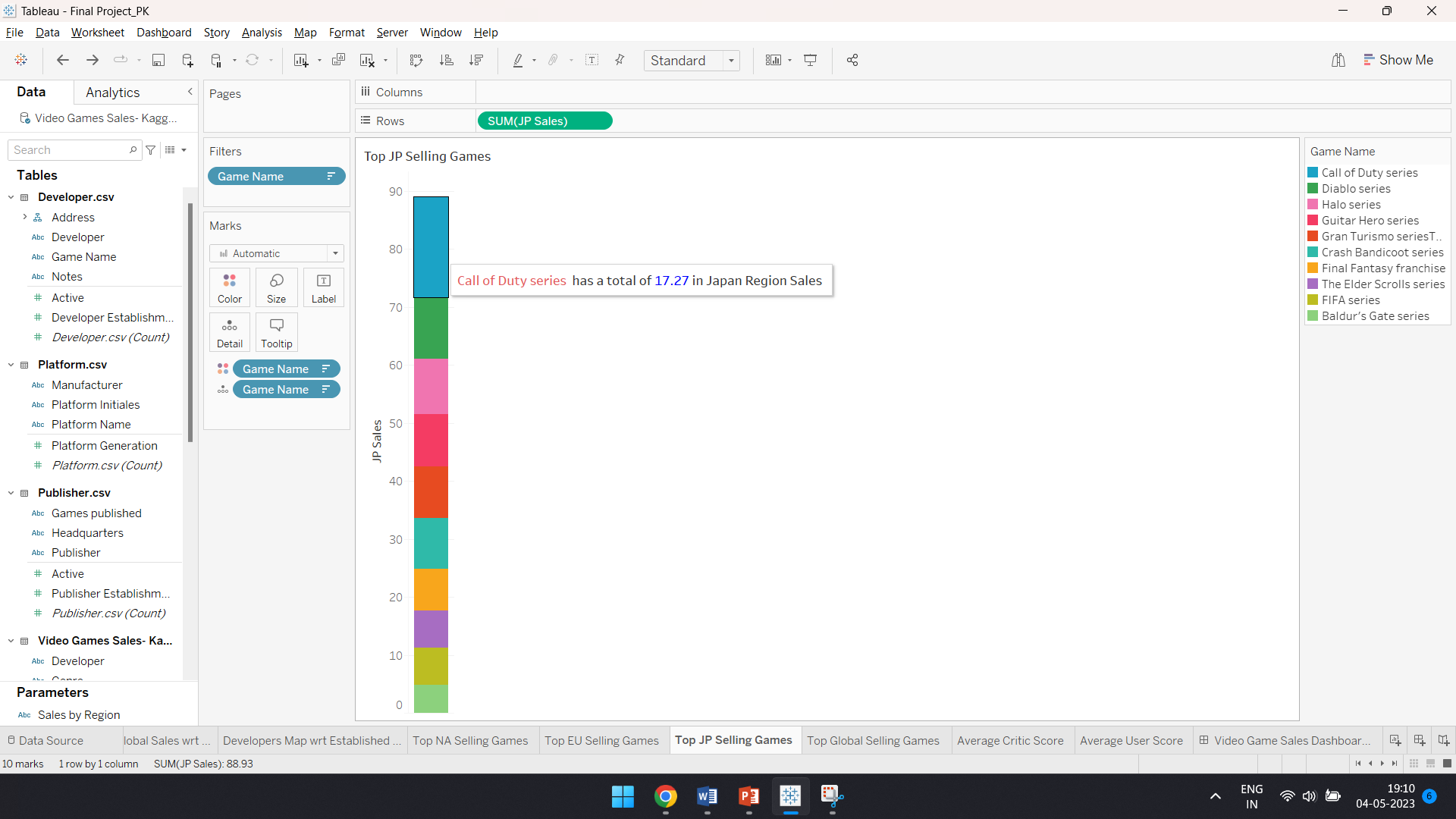
1. **Top Selling Game from Europe Region**



This visualization provides a clear picture of the top-performing games in the European market and the scale of their success in terms of units sold. It could be useful for game developers and publishers to understand the trends in the Europe market and the types of games that are resonating with consumers

From the above visualization, Call of Duty Series has the highest sales with the total sales of 34.21 million dollars and The Elder’s Scroll Series with least sales 12.29 million dollars in the Europe Region

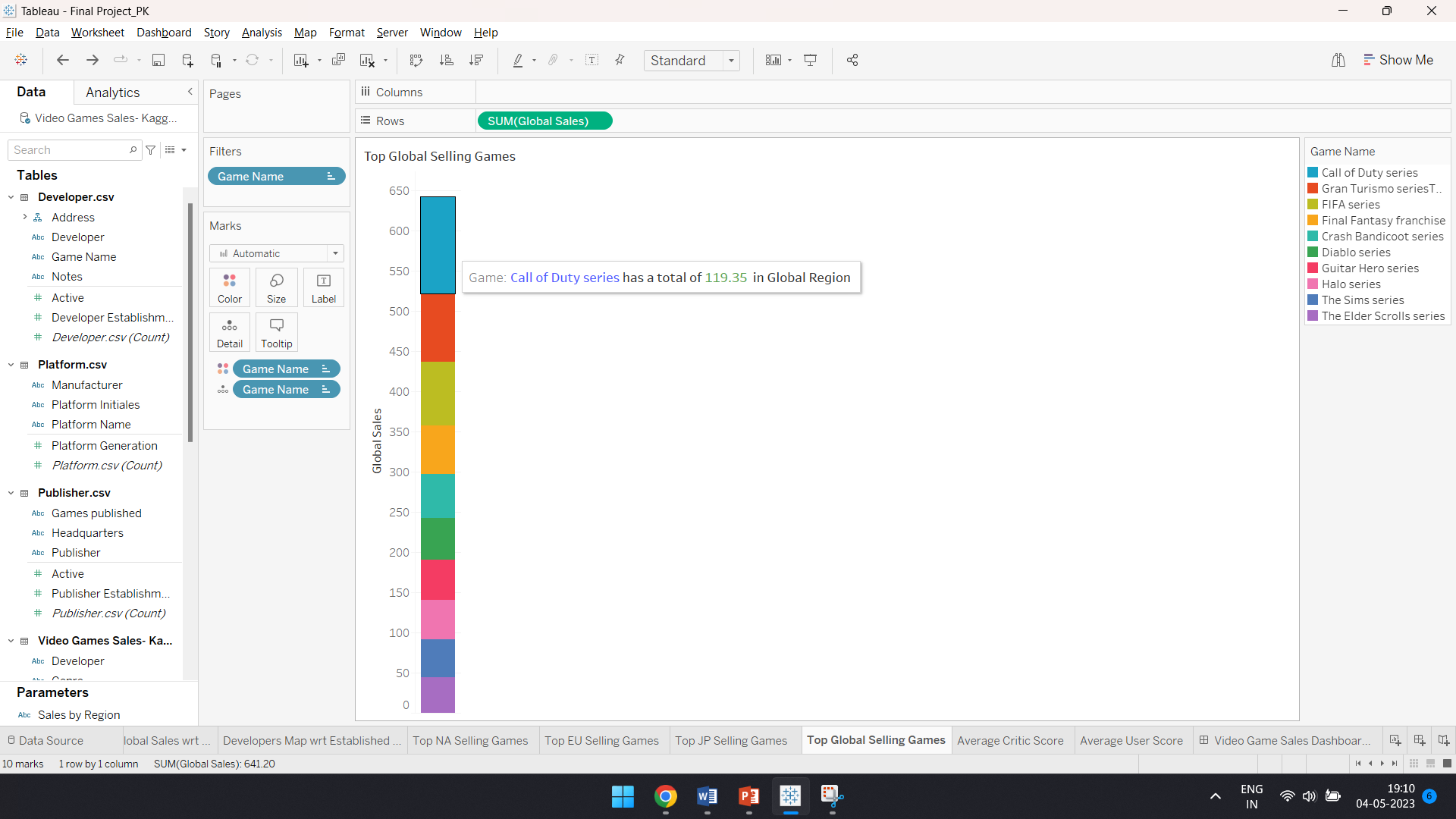
1. **Top Selling Game from Japan Region**



This visualization provides a clear picture of the top-performing games in the Japan market and the scale of their success in terms of units sold. It could be useful for game developers and publishers to understand the trends in the Japan market and the types of games that are resonating with consumers

From the above visualization, Call of Duty Series has the highest with the total sales of 17.27million dollars and Baldur’s Gate Series has the least sales with 4.28 million dollars in the Japan Region.

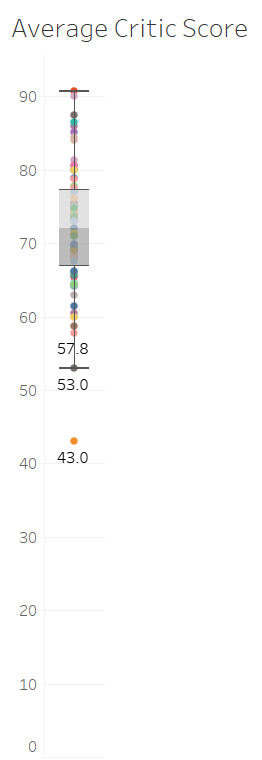
1. **Top Selling Game from Global Region**



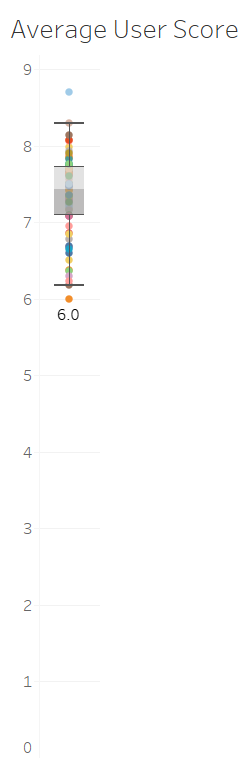
This visualization provides a clear picture of the top-performing games in the Global market and the scale of their success in terms of units sold. It could be useful for game developers and publishers to understand the trends in the Global market and the types of games that are resonating with consumers

From the above visualization, Call of Duty Series has the highest sales in the Global Region as well with the total sales of 119.35 million dollars and the Elder Scrolls Series has the least sales with 44.23 million dollars globally.

1. **Average Critic Score for overall games**



1. **Average User Score of overall games**

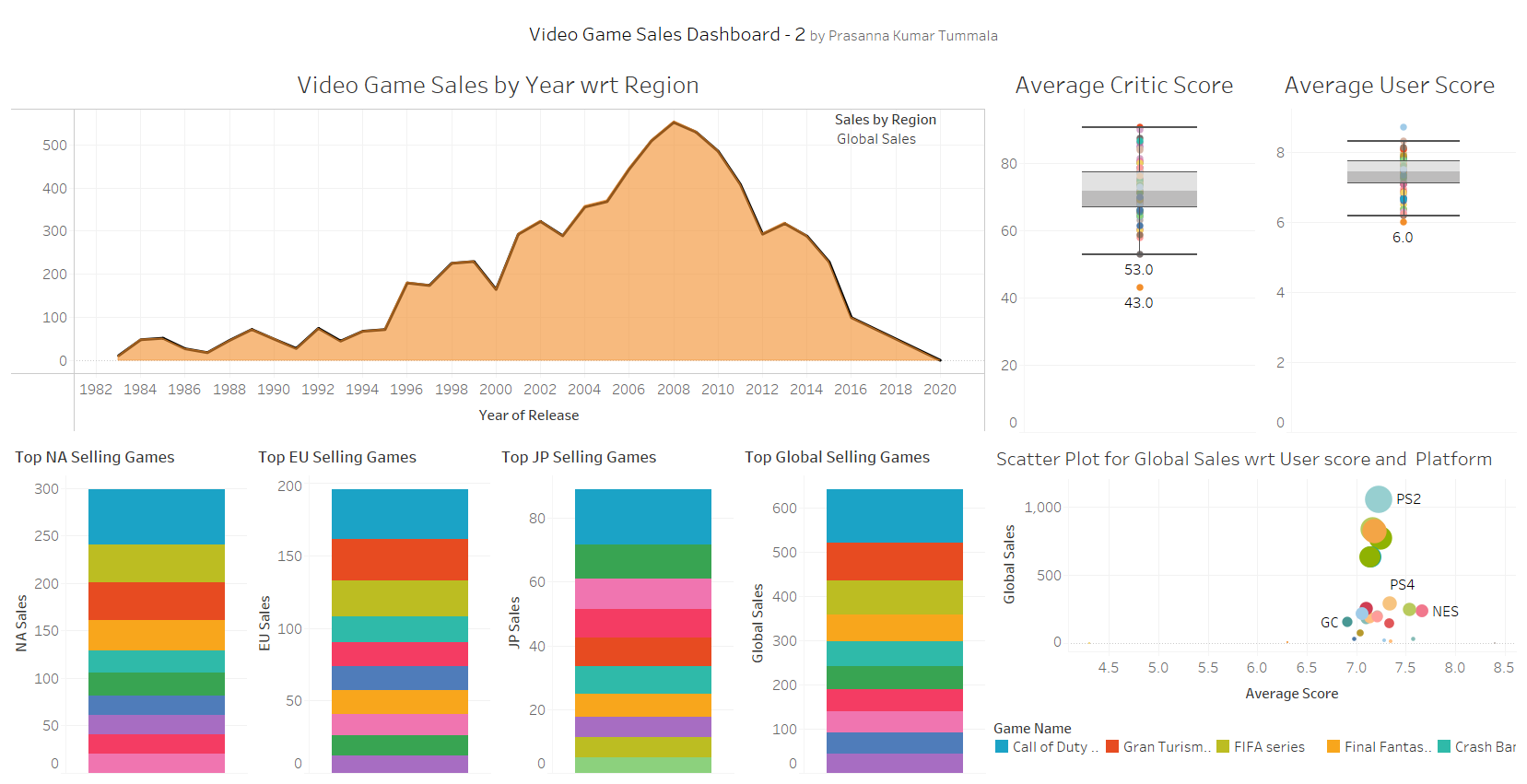


For the Average critic score of overall games and Average User score, I have used the box and whisper plot. The plots display the median (middle value of the data), the upper and lower quartiles (the values that separate the top and bottom 25% of the data), the minimum and maximum values, and any outliers.

The median critic score for all games is shown as a horizontal line inside the box. The box itself represents the middle 50% of the data, with the lower edge of the box being the lower quartile and the upper edge of the box being the upper quartile. The whiskers extending from the box represent the range of the data, excluding any outliers. Outliers are shown as individual data points outside the whiskers.

This visualization provides a summary of the distribution of average critic scores for all games, allowing us to easily identify any outliers or extreme values, as well as the spread and central tendency of the data.

**Video Game Sales Dashboard – 2:**



The dashboard provides an overview of various aspects of video game sales. The first visualization is a bar chart showing the video game sales in different regions over the years. The second visualization is a scatter plot showing the relationship between global sales and user score for different platforms. The third, fourth, fifth, and sixth visualizations show the top selling games in North America, Europe, Japan, and globally respectively. The seventh visualization is a box and whisker plot showing the distribution of the average critic score for overall games. The last visualization is a box and whisker plot showing the distribution of the average user score for overall games.

The dashboard provides insights into which regions have the highest video game sales and the top-selling games in different regions. The scatter plot shows the correlation between global sales and user score for different platforms. The box and whisker plots show the distribution of average critic score and average user score for overall games. These visualizations can help game developers and publishers to make informed decisions about the development and marketing of their games.

**Conclusion:**

In conclusion, the video game sales project explored various aspects of the gaming industry using a dataset from Kaggle and Data World. The project analyzed sales trends across different regions, top publishers, genres, platforms, and game developers around the world. The project also examined the relationship between user score, platform, and global sales.

The project findings show that North America is the region with the highest video game sales, followed by Europe and Japan. Electronic Arts, Blizzard Entertainment and Maxis are the top three game publishers, while action and sports genres are the most popular among gamers. The PlayStation 2 platform is the most widely used for playing games, followed by the Xbox 360.

The project's visualizations, including scatter plots, pie charts, and maps, effectively portrayed the data and allowed for easy interpretation of the findings. The project also provided insights into the gaming industry's current state and future directions.

Overall, the video game sales project provides a comprehensive analysis of the gaming industry and its trends. The project findings can be used by game publishers and developers to make informed decisions about the types of games to create and the platforms to release them on. Additionally, gamers can use the project findings to gain insights into the most popular games and make more informed purchasing decisions.

**Additional Research Questions:**

1. Impact of Covid-19

2. Cross Cultural Difference

3. Emerging Technology

4. Mobile Gaming

5. Game Development

**References:**

1. <https://www.kaggle.com/datasets/rush4ratio/video-game-sales-with-ratings>
2. <https://data.world/mhoangvslev/steam-games-dataset>
3. <https://en.wikipedia.org/wiki/List_of_best-selling_video_games>
4. <https://truelist.co/blog/gaming-statistics/>
5. <https://www.wepc.com/news/video-game-statistics/>