



# Brunel University London

---

**CS5803 Data Visualization**  
**Coursework for 2022/23**

**Student ID: 2242297**  
**Academic Year: 2022-23**

## 1. Introduction

The dataset 'Video Game Sales with Ratings' is used for this coursework. It is taken from Kaggle.com and is available at <https://www.kaggle.com/datasets/rush4ratio/video-game-sales-with-ratings>. This dataset is a comprehensive collection of information about video games released from 1986 to 2016. It consists of video game sales, ratings, and various other features such as the platform, genre, publisher, developer, and release year. With over 16,500 records, this dataset provides valuable insights into the gaming industry and offers an excellent opportunity for researchers and data analysts to study the trends and patterns of the video game market. With over 16,500 records, this dataset provides valuable insights into the gaming industry and offers an excellent opportunity for researchers and data analysts to study the trends and patterns of the video game market.

### 1.1 Data

The original dataset consists of 16719 rows with 16 columns and the size of this dataset is 1.62MB. The description of the data, column names along with their domain are mentioned below. For the 'Ratings' column, the values of it are also described in bold letters.

S.NO	Column Name	Description	Domain
1	Name	The name of the video game	Nominal
2	Platform	The platform on which the game is running	Nominal
3	Year_of_Release	Year of game released	Integer
4	Genre	The type of the game	Nominal
5	Publisher	The company that published the game	Nominal
6	NA_Sales	Sales in North America (in millions)	Float[0-100]
7	EU_Sales	Sales in Europe (in millions)	Float[0-100]
8	JP_Sales	Sales in Japan (in millions)	Float[0-100]
9	Other_Sales	Sales in the rest of the world	Float[0-100]
10	Global_Sales	Total worldwide sales	Float[0-100]
11	Critic_Score	Aggregate score compiled by Metacritic staff	Float[0-100]
12	Critic_Count	The number of critics used in coming up with the critic score	Integer
13	User_Score	Score by Metacritic's subscribers	Float[0-100]
14	User_Count	Number of users who gave the user_score	Integer
15	Developer	The party responsible for creating the game	Nominal
16	Rating	The ESB (Entertainment Software Rating Board) Ratings: <b>EC</b> : Early Childhood <b>E</b> : Everyone <b>E10+</b> : Everyone 10 and above <b>T</b> : Teenage	Nominal

		<b>M:</b> Mature <b>AO:</b> Adults Only <b>RP:</b> Rating pending (not yet assigned a final rating)	
--	--	---	--

## 1.2 Persona and Questions

Video games have become a lucrative investment for banks and financial institutions. With the steady growth of the gaming industry, many banks have recognized the potential for high returns on investment in gaming industries. ABC Bank Investment manager, who checks the portfolio of the bank, intends to broaden its portfolio by venturing into the gaming industry. To execute this, the bank's research team collected data and solicited the expertise of its Data Analytics team to provide insights on the below two questions.

1. What is the year-over-year (YOY) growth rate of video game sales in the three markets (NA, EU, JP) launched between 1996 and 2016 and how have the top five publishers performed during this period?
2. What is the strength of the correlation between user count and sales for video games launched between 1996 and 2016 and is there a variation in sales among different genres in the gaming industry and what is the total sales value for each genre?

**Addressing question 1** would enable the investment bank to determine which market has the most substantial growth rate and assess the sales value of the top five publishers during the period analyzed.

**Addressing question 2** would facilitate the bank in evaluating the extent to which user count influences game sales and in identifying whether there is variability in the impact of user count on sales among different gaming genres.

## 1.3 Requirements:

**Q1.** What is the year-over-year (YOY) growth rate of sales in the three largest markets for major games launched between 1996 and 2016, and how have the top five publishers performed during this period?

**R1:** To answer this question, the user needs to create a line chart with lines showing the YOY growth of each market (Europe, Japan, and North America).

**R2:** In addition to this, the user also needs to create an area chart to show the sales value of the top 5 publishers over the period of 30 years.

**Q2.** What is the strength of the correlation between user count and sales for video games launched between 1996 and 2016, and is there a variation in sales among different genres in the gaming industry and what is the total sales value for each genre?

**R1:** To answer this question the user must create a scatter plot with the user count on X-axis and global sales value on the Y-axis.

**R2:** Additionally, the user should also create a bar graph showing sales value by genre. This graph needs to be interactive so that clicking a particular bar and scatter plot will reflect and show the relationship between user count and sales for that genre.

**R3:** The user must create a global filter for genre and platform which will help the bank to see the scenario in more detail.

**R4:** All views should fit in the dashboard without the need for scrolling. The dashboard should be viewed on a desktop, as the mobile view will change the orientation to optimize the view, which will require scrolling.

**R5:** The user should have a basic understanding to visualize bar graph, area chart, line chart, and a scatter plot, also have a decent statistical understanding to understand correlation and R-squared.

**R6:** Should have a basic understanding of presentation mode in Tableau and Power BI

**R7:** The user should be aware of chart interactivity as it enables the users to select different filters and variables which will help to gain deeper insights into the data.

## **2. Design**

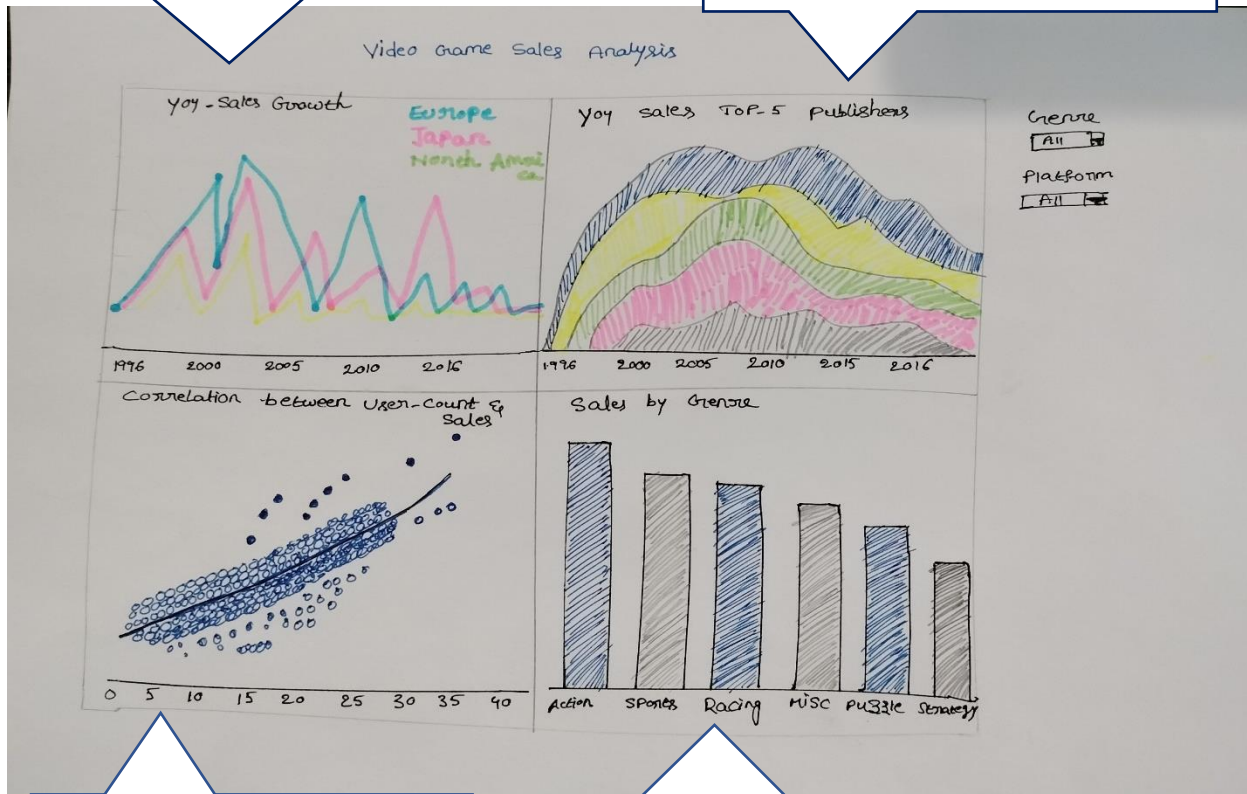
Before I start the design plan of the tableau dashboard, would discuss my journey in developing the current dashboard of video game sales analysis. Initially, I started analyzing a different dataset called 'stroke prediction' which is available on Kaggle.com and is available at <https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset>. It was a fascinating dataset containing information about various demographics, lifestyles, and health factors that can potentially contribute to strokes. Using Tableau, I visualized the distribution of stroke cases across different age groups, gender, and regions. I also explored the correlation between stroke and other health indicators such as hypertension, heart disease, and smoking habits. Through these visualizations, I gained valuable insights into the risk factors and patterns of stroke occurrence. However, there was no scope for creating multiple views on this dataset. With the supervio's advice, different views were tried; however, it did not give effective solutions to the persona and research questions.

Thus, I turned my attention to the 'Video Game Sales' with the rating dataset available on Kaggle. The reason for choosing this dataset is because of the richness of the data, the scope for comparative analysis, data quality, and relevance. Also, in this project, I aimed to showcase my skills in data visualization and storytelling by presenting the key points and insights on the research questions.

According to Bertin's theory, the significance of using suitable visual variables, including shape, size, color, and position, is to encode various elements of data. On the other hand, Mackinlay's guidelines, such as practical tips and techniques for designing an effective dashboard. Designed a paper landscape and the final dashboard.

**Q1. R1:** year-over-year (YOY) sales growth for each market (Europe, Japan, North America)

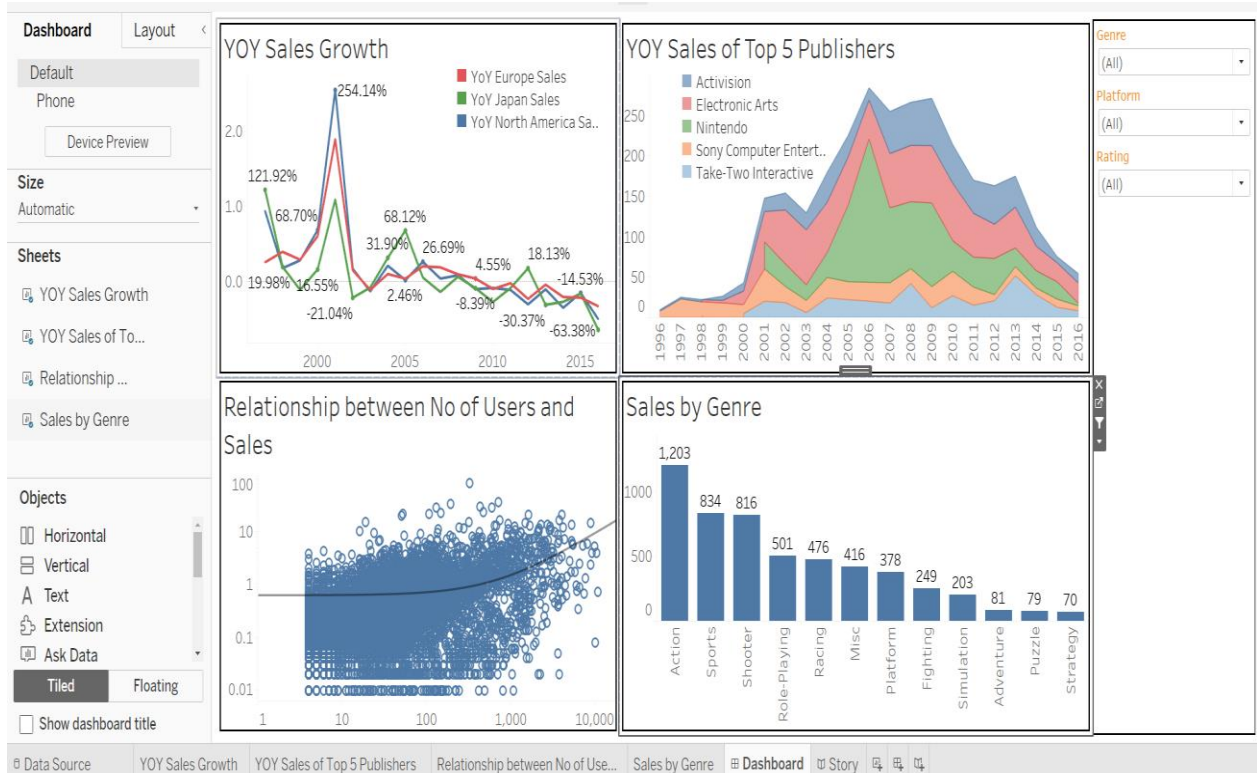
**Q1. R2:** Sales of the top five publishers launched between 1996 and 2016.



**Q2. R1:** Correlation between user\_count and global\_sales

**Q2. R2:** Variation between sales among different genre and the value of sales of each genre

## Final\_Dashboard



The important difference between the paper landscape and the final dashboard is that adding the filter based on 'Action' (genre) makes the two reports more interactive both on the reports and dashboard. Performing this step, not only answers the research question but also makes the user understand the data in deeper. Also, an additional global filter 'Rating' is added in the dashboard to give the user more details about the sales of all the age groups.

### 3. Implementation

The implementation of this coursework involves three steps.

1. Cleaning the dataset using python
2. Creating the Tableau Dashboard
3. Creating the same dashboard using Power BI

#### 3.1 Data Cleaning

The original dataset consists of 16719 rows and 16 columns. However, there were many missing values. So, cleaned them using Python. After cleaning the dataset, it is left with 6825 rows in total. The steps for cleaning the dataset in Python are given below.

**Import Libraries:** Import the required libraries (Numpy and Panda). Numpy and Pandas provide built-in functions to identify the missing data, such as 'Nan' values.

**Read Data:** The function `pd.read_csv()` is used to read the CSV file.

**Detect Missing Values:** The code '`print (video games sales.isna())`' analyses the dataset for missing values (NaN), then prints the findings. The Pandas `isna()` method is used to check whether or not each element in the dataset contains a missing value.

```
#lets check if missing data
#we can do this by checking individual values in dataset
print(video_games_sales.isna())
#isna method is to check each individual value for missingness
```

**Remove Missing Values:** The `dropna()` method is used to remove all missing values.

```
#from the plot the last columns has so many missing values because Metacritics only covers a subset of the platforms.
# Also, a game may not have all the the fields recorded
video_games_cleaned=video_games_sales.dropna()
```

**Save Clean Dataset:** Save the cleaned dataset.

```
# To save this file
video_games_cleaned.to_csv('video_games_cleaned.csv')
```

#### 3.2 Tableau Dashboard Implementation

Import the cleaned dataset 'video\_games\_cleaned.csv' into Tableau. As a first step, check all the datatypes of the columns to ensure correct data representation and prevent unexpected errors during implementation. Notice that, 'year-of-release' is displayed in number(decimal) format. Convert this into a string and create a new calculated column with the name of 'Date'.

Date

```
DATE([Year_of_Release] + "12-31")
```

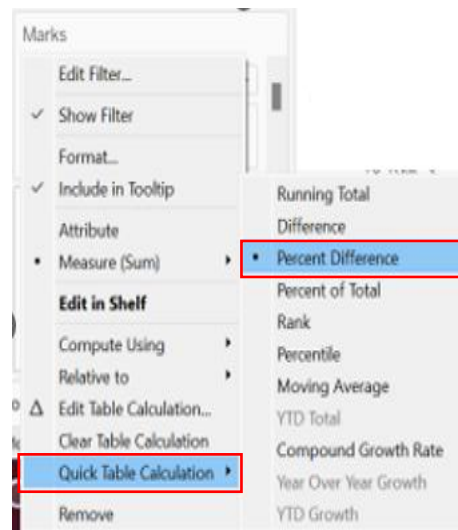
This code will add the 31st of each year to the Year\_of\_Release field and convert the resulting string into a date format and show the right format in the reports.



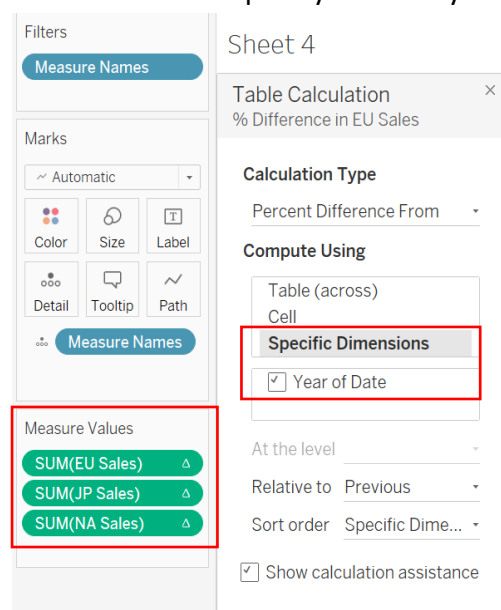
Rename EU\_Sales, JP\_Sales, and NA\_Sales into Europe\_Sales, Japan\_Sales, and North\_America\_Sales respectively to improve readability and simplification. To answer the research questions mentioned above, the user needs can develop four reports. The first two reports are for research question-1, and the next two reports are for answering research question-2. The development of all four reports is explained step-by-step below.

### 3.2.1 YoY (year-over-year) Sales Growth:

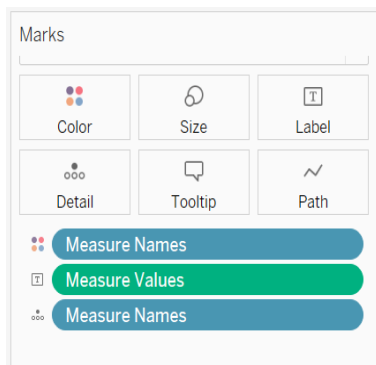
1. Drag and drop the date column into the column field, and measure values into the row field.
2. In the measure values pane, clear all the measure values other than Europe, Japan (JP), and North America (NA) sales.
3. For each of these three columns, select Quick table calculation >> Percent Difference



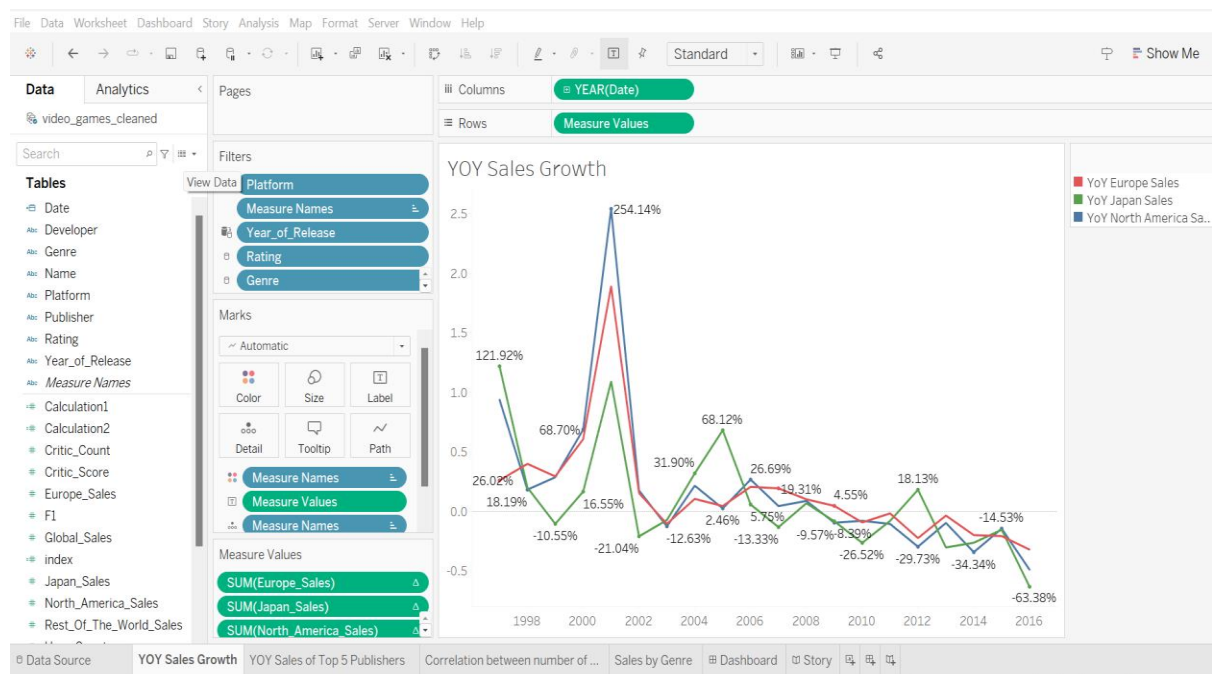
4. For the same columns, click on Edit table calculation and choose specific dimensions 'year-of-date'. The reason to specify the calculation to the 'year' dimension is that, without it, tableau will compute year-over-year growth for the entire dataset.



5. Add Measure names into the color panel and measure values into a label.



6. Add 'Year\_of\_Release' in the filters section and choose years from 1996 to 2016. This step needs to be performed for all the reports as per the research questions.
7. Finally, change the measure names to YoY Europe Sales, YoY Japan Sales, and YoY North America Sales to make more sense.

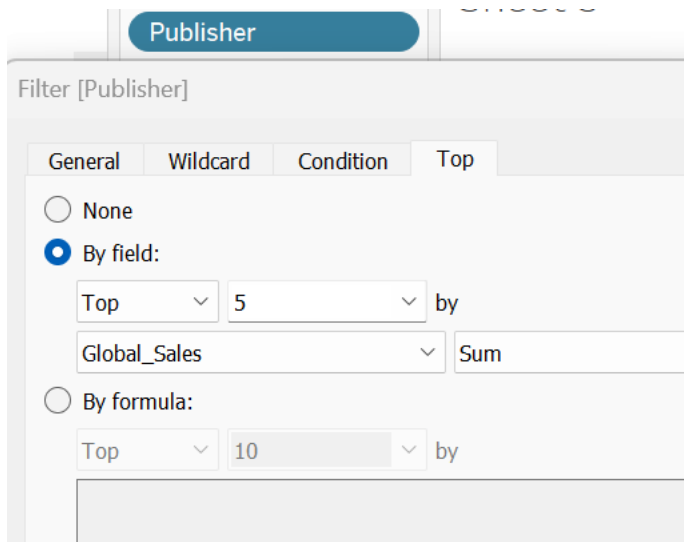


**Observations:** Overall, all three markets show a downward trend in the given period. NA sales went to a peak (254.14%) in 2001. But all three markets have seen negative growth.

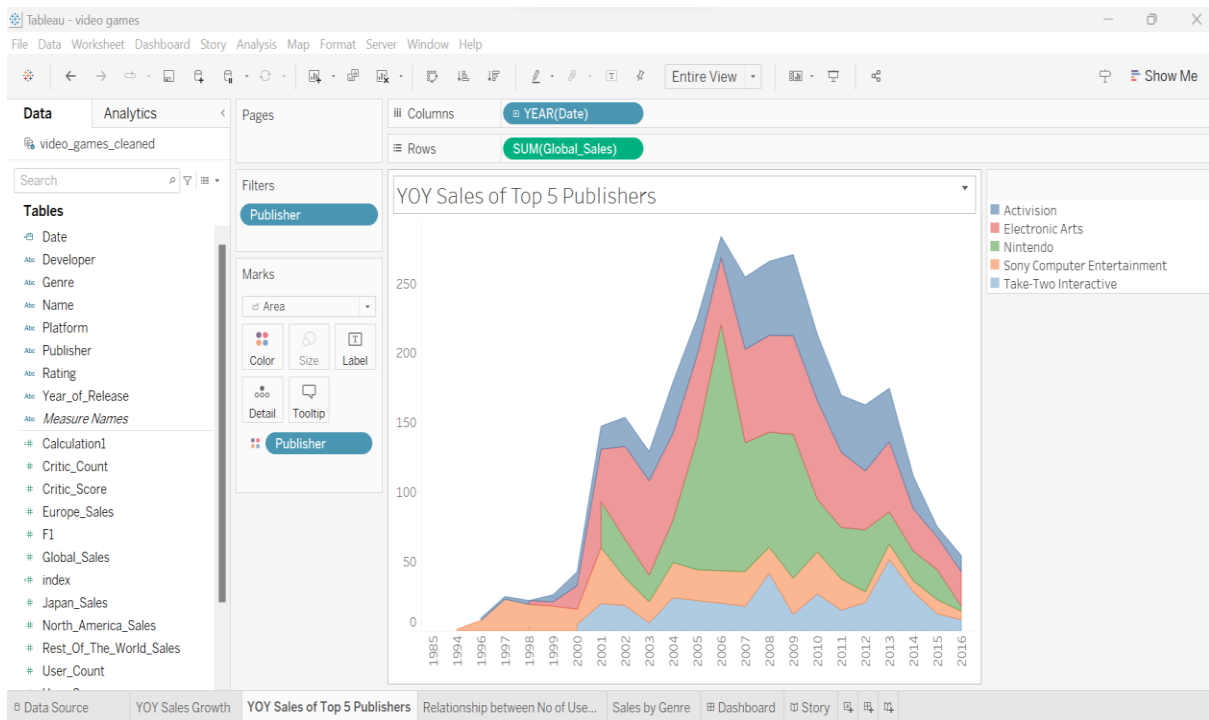
### 3.2.2 Year-Over-Year Sales of Top 5 Publishers:

1. Drag the 'date' column into the column and Global\_Sales into the rows fields.
2. Add the 'publisher' field to the color field, in order to separate the sales data by publishers.
3. To concentrate on the performance of the most important publishers in the dataset, choose the top-5 publishers in the filter pane based on sales.





4. Select an area chart, which gives a visual depiction of the sales patterns for top-5 publishers.



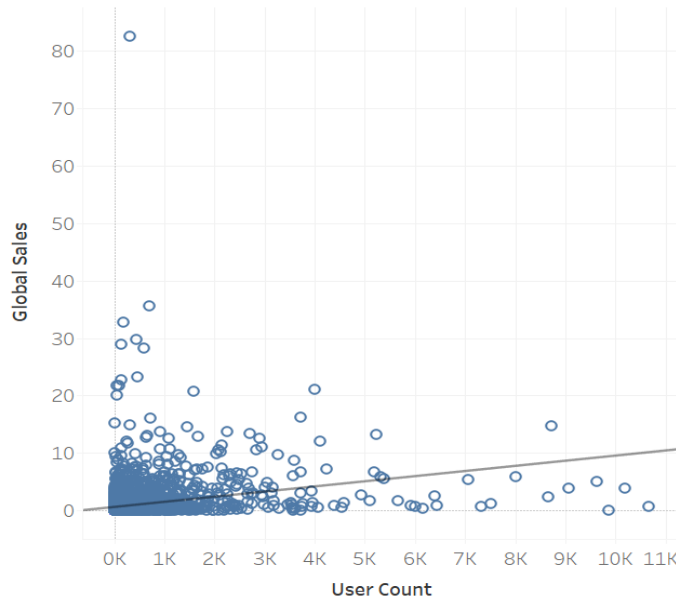
**Observations:** The area chart makes it simple for end users to examine the top-5 publishers' performance in terms of global sales. To understand each publisher's relative performance, clearly notice the trends, patterns, and changes in sales for each publisher. Also, observe that Nintendo and Electronic Arts have dominated the space for two decades.

### 3.2.3 Correlation between the Number of Users and Sales

1. Drag and drop the 'User\_Count' into the column shelf, and 'Global\_Sales' into the rows shelf.
2. Click on "Analysis" on the menu bar and select 'Aggregate Measure' to aggregate all the measure values. This step is very important here, without this step, tableau would not know how to aggregate the measures and create a scatter plot.
3. From the Analytics tab (Just beside the Data tab), under the menu, drag the 'trend line' to the scatter plot to identify the relationship and patterns between user\_count and sales.

Columns	User Count
Rows	Global Sales

Sheet 5



- To show the plot in a more appropriate and clear visual manner, click on the edit X-axis, and choose 'Logarithmic' under the scale option. Perform the same for Y-axis.

Edit Axis [User Count]

General

Tick Marks

Range

☒ Automatic
☒ Include zero

☐ Uniform axis range for all rows or columns

☐ Independent axis ranges for each row or column

☐ Fixed

Automatic

Automatic

1

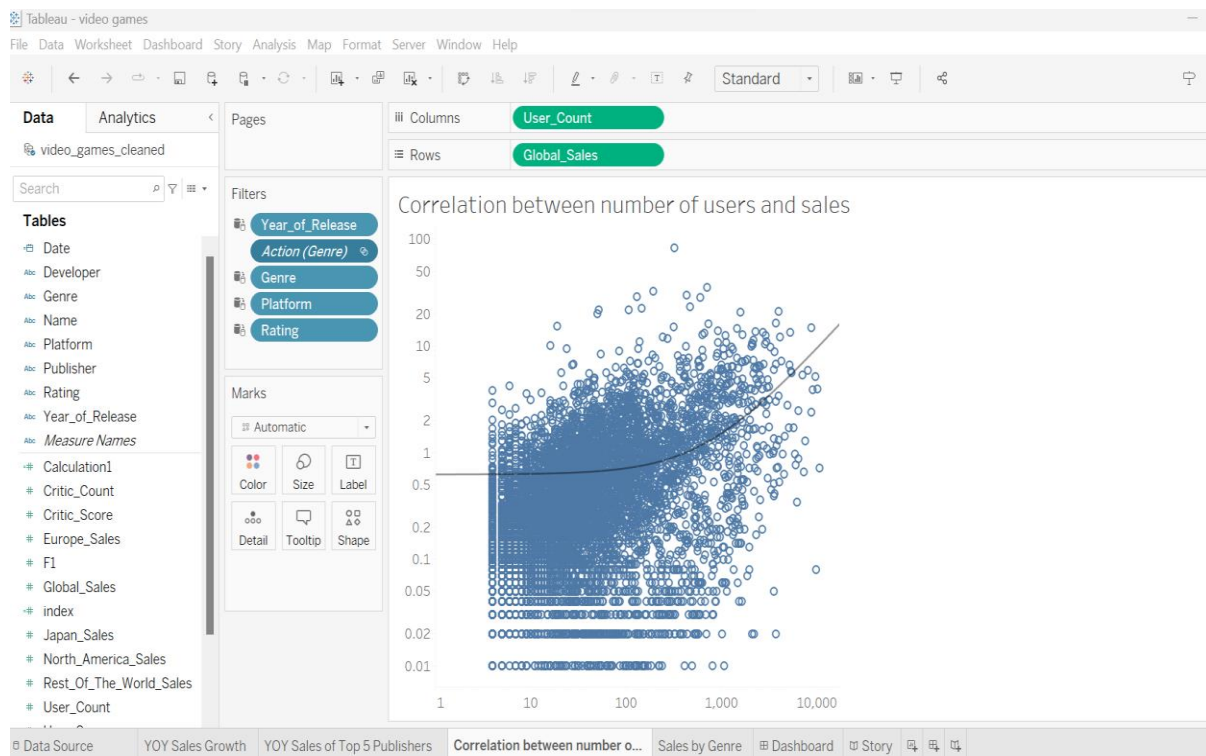
17,121.0123671

Scale

☐ Reversed
☒ Logarithmic

☒ Positive
☐ Symmetric

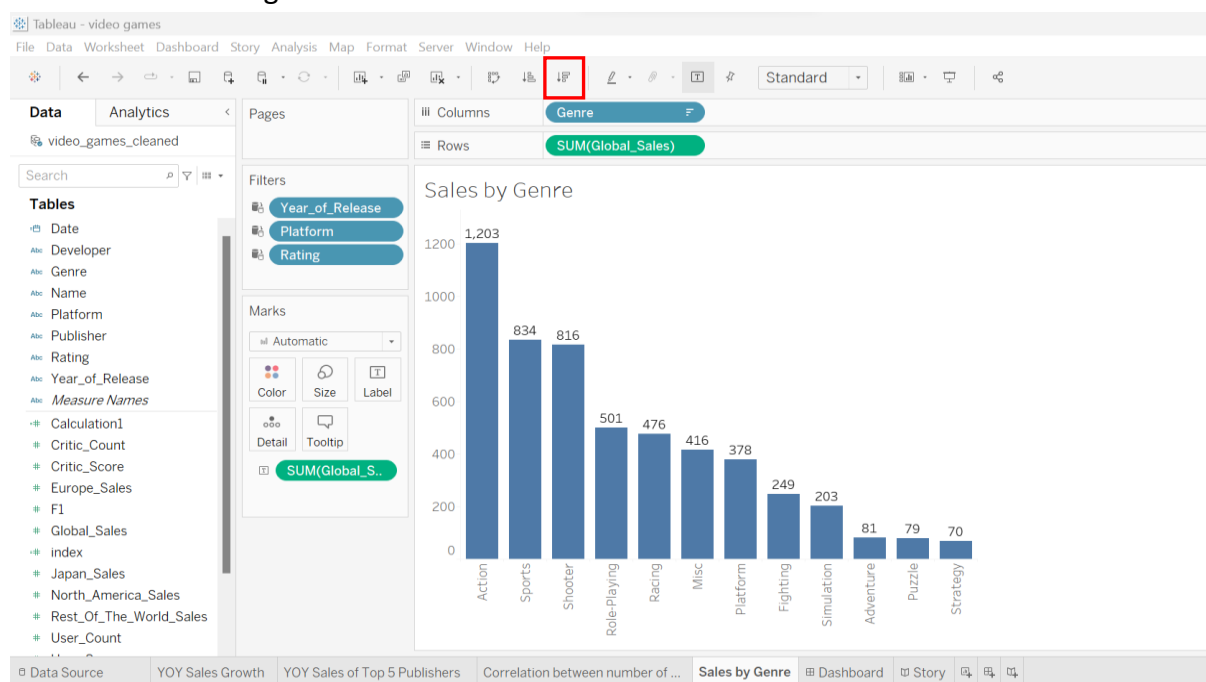
- After changing the scale, the plot is shown below.



**Observation:** The scatter plot of user\_count and global\_sales shows a positive correlation between the two measures. As user\_count increases, there appears to be a corresponding increase in global\_sales.

### 3.2.4 Sales by Genre:

1. Place the 'Genre' column into the column shelf, which creates a column for each genre in the bar chart. Place the 'Global Sales' into rows shelf, this will represent a sales value for each value on the y-axis of the bar chart.
3. It is always recommended to use the sort option either in ascending or descending order based on the requirement which helps the end user to check values easily. In this case, sorted values in descending order.

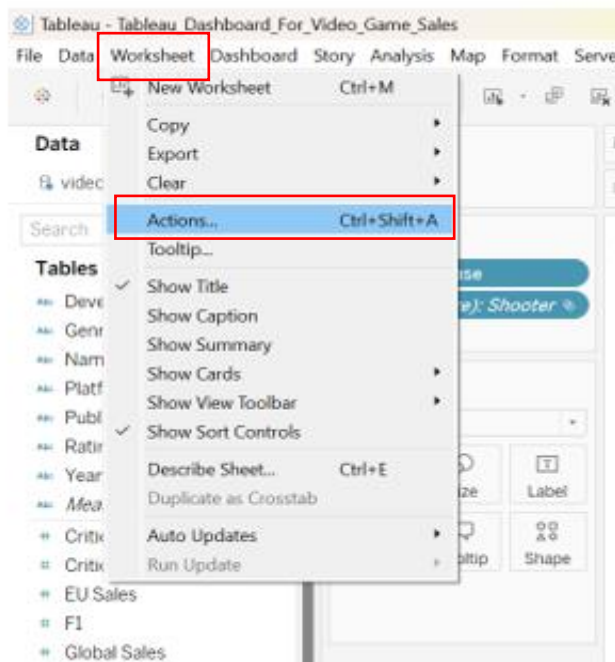


**Observations:** The bar chart depicts the variation in the sales among different genres in the gaming industry and the total sales value of each genre. The highest sales were observed in the 'Action' genre whereas the lowest sales were observed in 'Strategy'.

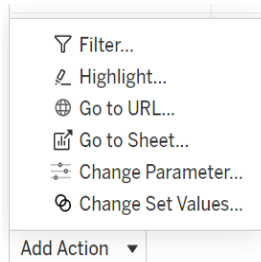
### 3.2.5 Additional 'Action' based on Filter:

This is an additional step to create an interaction between the bar chart and scatter plot. For example, clicking on a particular bar should show the related scatter plot for that specific genre.

1. Click on the Worksheet on the menu bar and select Actions.



2. Click on add action and choose Filter



3. In the Add Filter action window, choose the source sheet as 'Sales by Genre' and 'Select' the radio button. Choose the target sheet as 'Correlation between user\_count and 'global\_sales' and 'show all values' option.
4. While clicking on the 'Ok' button, the user might face issues due to the screen size resolution display. I faced difficulty here, was not able to see the 'Ok' option due to screen size. The solution for this is to 'change the resolution of the display' in your system to 125 scale instead of 150.
5. Perform the same interaction for the dashboard as well. While doing this, the user needs to choose the source and target as the dashboard and choose the source and target reports appropriately.

Edit Filter Action

Name

Action (Genre)
Insert

Source Sheets

video\_games\_cleaned

Run action on

☐ Correlation between number of users and s

☒ Sales by Genre

☐ YOY Sales Growth

☐ YOY Sales of Top 5 Publishers

☐ Hover

☒ Select

☐ Menu

☐ Single-select only

Target Sheets

Correlation between number of users...

Clearing the selection will

☐ Keep filtered values

☒ Show all values

☐ Exclude all values

Filter

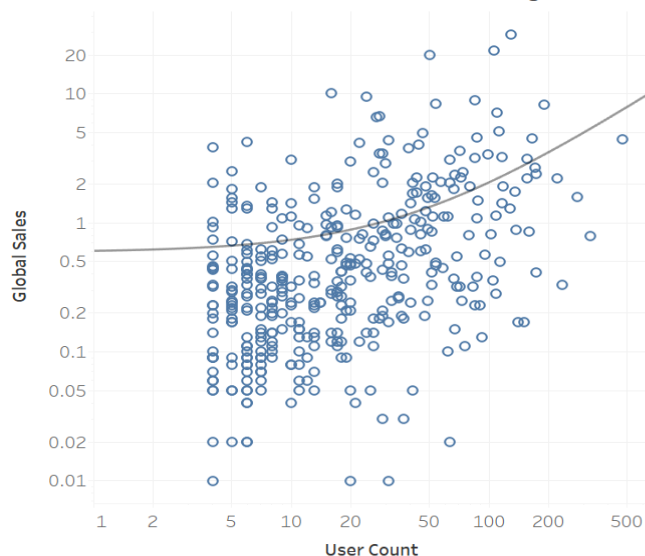
☒ All fields

☐ Selected fields

<input type="checkbox"/>	Source Field	Target Data Source	Target Field
<input type="checkbox"/>	Click to add		

**Observation:** Clicking on any of the bars gets reflected in the scatter plot and the report automatically goes from bar chart to scatter. For example, clicking on the 'Misc' genre on the bar chart shows the below scatter plot.

Correlation between user\_count and global sales

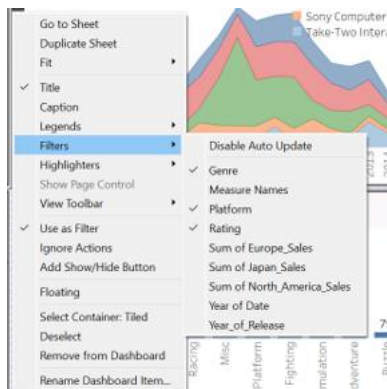


### 3.2.6 Dashboard Development

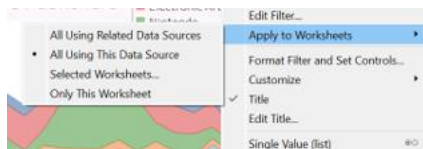
1. Create a new dashboard, place all four reports, and adjust them.
2. Change the size of the dashboard from 'Fixed' to 'Automatic'. If you choose to maintain the default configuration of using a 'fixed size' for dashboards, make sure to build your

visualization at the size it will be seen at. Additionally, if you set 'Automatic', Tableau will automatically adjust a visualization's overall dimensions according to screen size.

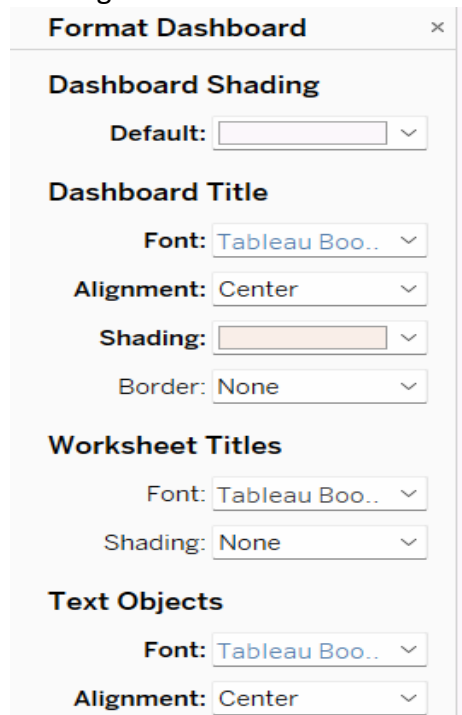
3. Add the global filters for Genre, Platform, and Rating.



4. Apply the filters for all the reports using choosing 'All Using This Data Source'.

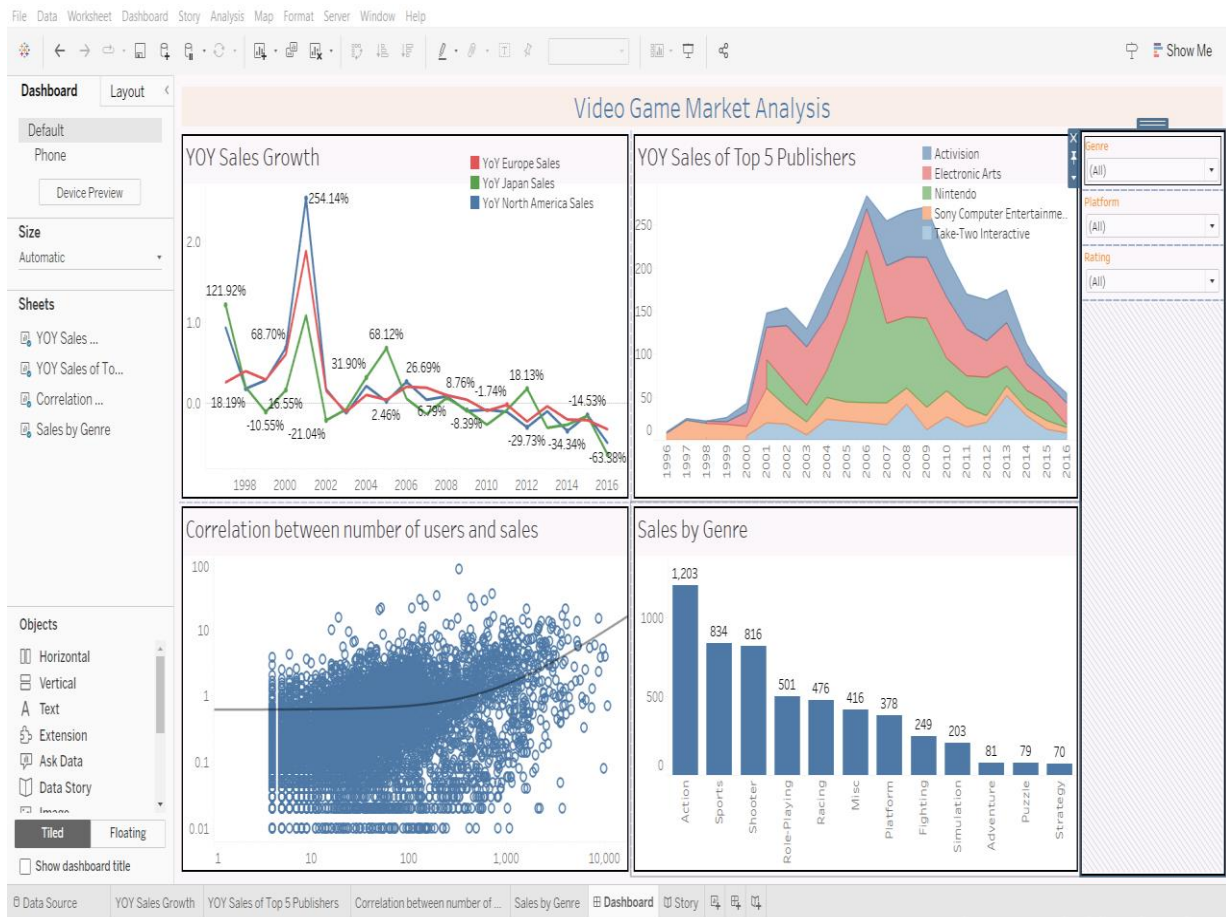


5. In the Dashboard Layout, the user can select the borders for good visuals.
6. Clicking on the show title gives the option to add the title. Users can edit the title names and format it.
7. In the format dashboard, the user can change the font color, font size, alignment, and shading of the title.



8. The final dashboard is created that meets all the requirements specified in the coursework. Here is the screenshot for the same.





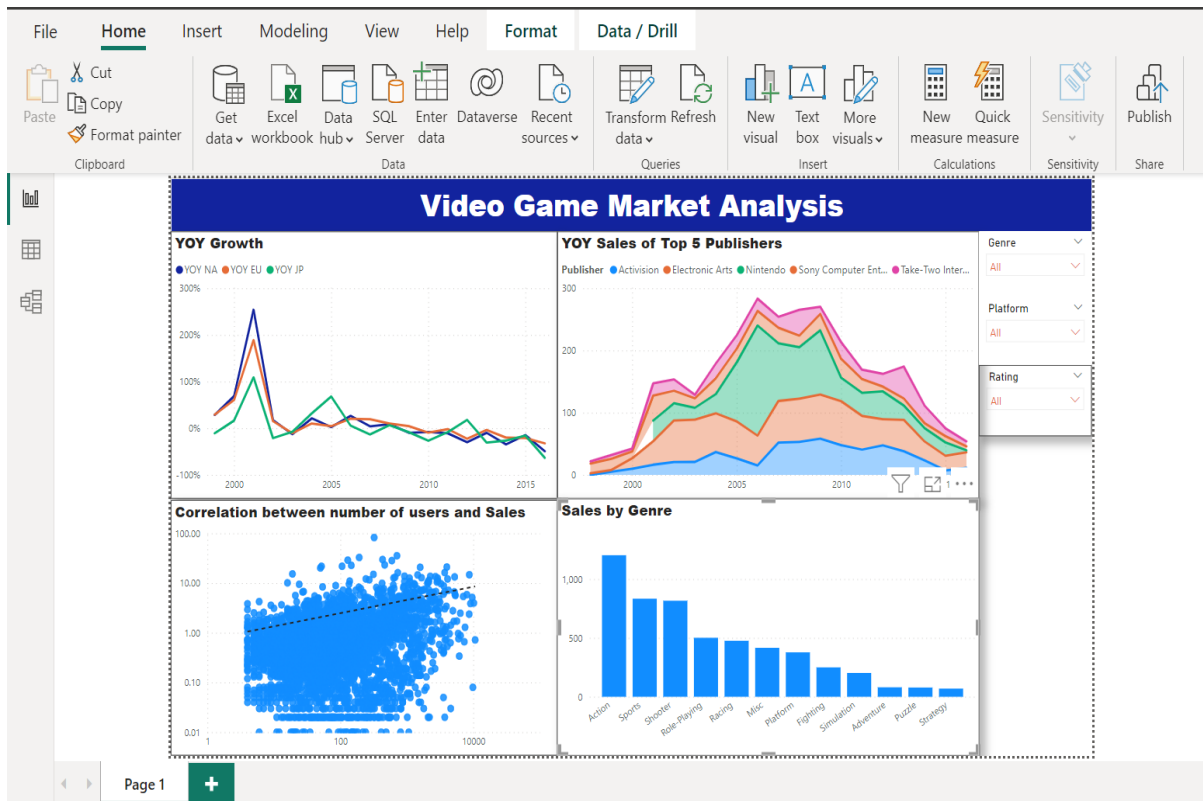
### 3.3 PowerBI Dashboard Development

While developing the powerBI dashboard, noticed some of the differences between Tableau and PowerBI.

**YOY Growth:** Tableau has an inbuilt function to derive YoY growth with its GUI whereas a complex DAX function needs to be built to achieve the same in Power BI.

**User Filter:** By default, all the charts in Tableau are independent of each other and the user must manually enable the filters while in Power BI all the charts by default are connected and the user must manually go and disable this if needed. Depending on the use case both these capabilities have their advantages and disadvantages.

**Dashboard Formatting:** Individual charts are created in separate worksheets and collated as the dashboard in Tableau. In Power BI, however, the user can create multiple charts in a single sheet thereby making formatting easier.



#### 4. Walkthrough

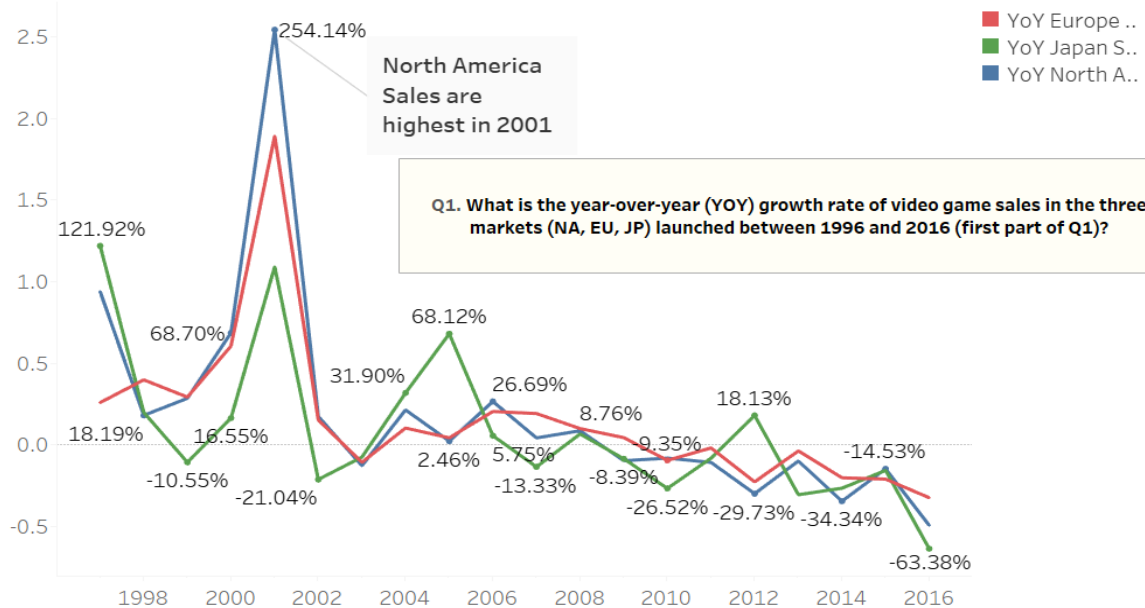
##### Answering Question-1

1. Users need to make use of all the filters of three markets (North America, Japan, and Europe) and see the trend lines for each market.

#### Video Game Sales Story

Q1. Overall, the sales declined over the given period. The sales went to peak in 2001. Also, all 3-markets are seeing negative growth since around 2010 with an exception of Japan in 2012. Thus, suggesting that the industry is on a decline.

Click on the Arrows to go to Next report



2. Users need to make use of global filters on the dashboard to get more insights into the data.
3. Users need to hover over the cursor to see the YoY sales of top-5
4. Users can also go and choose the top-3 or any other value. The Top-5 has been selected in this case to answer the research question.

Filter [Publisher]

General
Wildcard
Condition
Top

☐ None

☒ By field:

Top
5
by
Global\_Sales
Sum

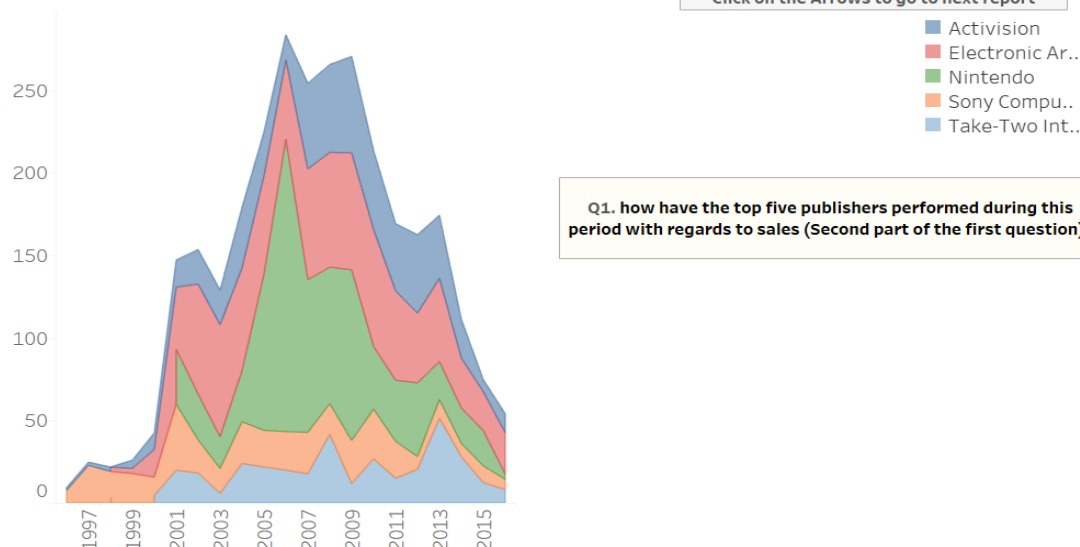
☐ By formula:

Top
10
by

## Video Game Sales Story

Q1. As seen in the first report, overall the sales got declined for top-5 publishers. Among all, Nintendo and Electronic Arts are dominating the space for two decades. Nintendo sales started at 33.3 in 2001 and it went to peak around 177 in 2006 and fell down to 3.2. Activision and Electronic Arts performed good. However

Click on the Arrows to go to next report

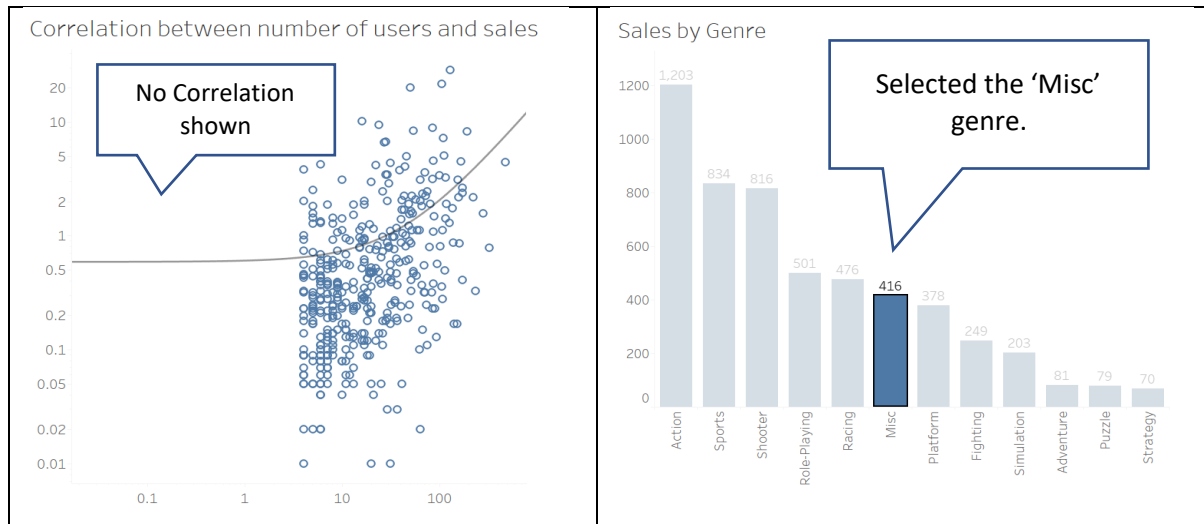


Q1. how have the top five publishers performed during this period with regards to sales (Second part of the first question)?

**Observation-1:** Overall, the sales declined over the given period. The sales went to peak in 2001. Additionally, all 3-markets are seeing negative growth since around 2010 except Japan in 2012. A similar trend has been seen in the area chart. Thus, suggesting that the video game industry is on a decline.

## Answering\_Question\_2

Users need to check each bar to see the sales value of the genre and click on each bar to see the correlation between user\_count and sales specific to the selected genre over the dashboard/report. For example, select the 'Misc' genre bar and see the relationship.

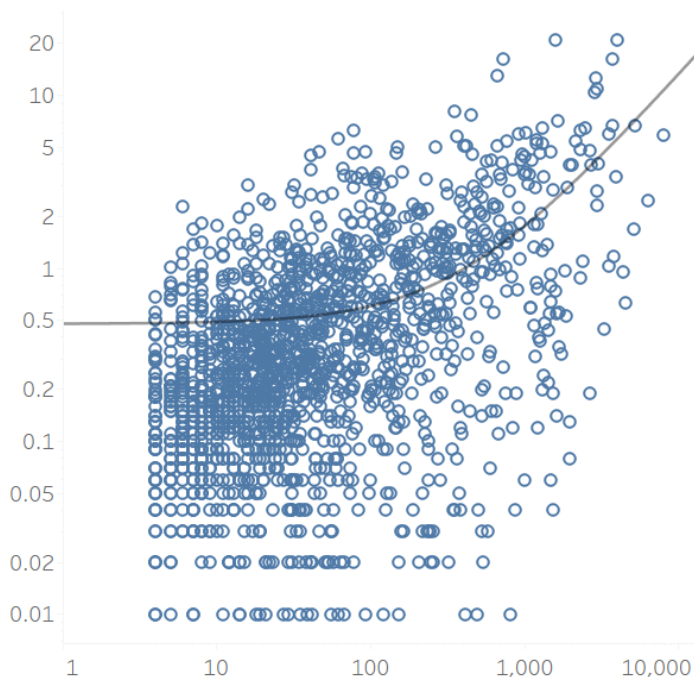


**Observation-2:** Overall, there is a correlation between user\_count and global\_sales. However, it's not the same for all genres. Some of the games do not have any relationship with the number of users and sales as shown above.

## Video Game Sales Story

Q2. Overall, there is correlation between user\_count and global\_sales. However, it's not the same for each genre. Some of the games are not having any relationship between the user\_count sales. Meaning, the relationship depends on the type of game. Please use the arrows to go to next report.

Click on the Arrows to go to next report

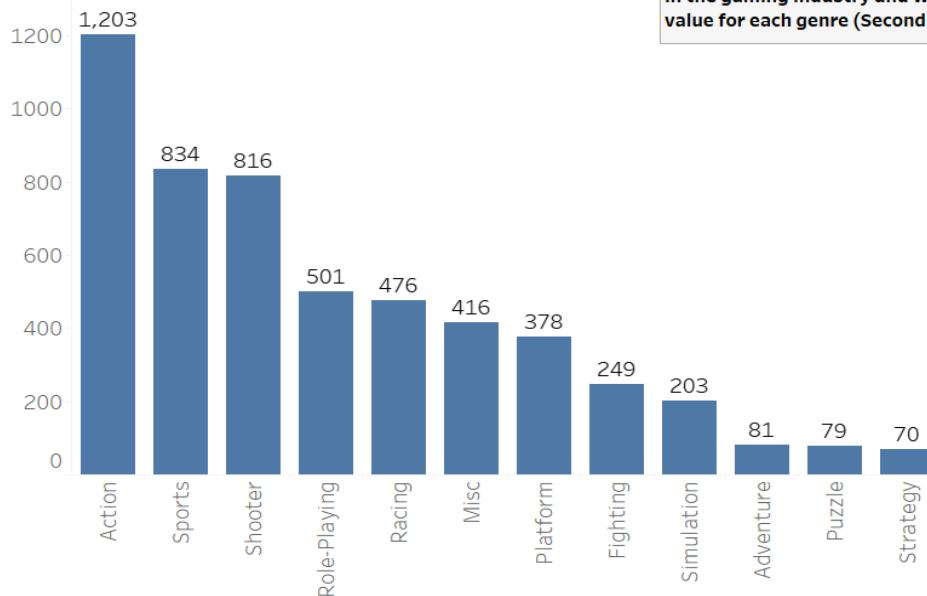


Q2. What is the strength of the correlation between user count and sales for video games launched between 1996 and 2016 (first part of Q2)?

## Video Game Sales Story

Click on the Arrows to go next report

Q2. The sales are sorted in descending order here. Only the games in Action, Sports, and Shooter genre are contributing to high sales. All the other games are not performing good. Thus, the video game industry is not a good option for investment for the bank. Please use the arrows to go to next report.



Q2. Is there a variation in sales among different genres in the gaming industry and what is the total sales value for each genre (Second part of Q2)?

### 5. Conclusion

Both the research questions are answered effectively with visualizations along with the observations. With these observations, we can clearly say that the video game industry is not a good choice of investment for the bank. For answering question-1, The analysis showed a downward trend in the sales growth rate over the years, with respect to countries as well as top-performing publishers. While answering question-2, the strength of the correlation between user count and sales for video games was found to be strong for some games and weak for others, with Action and Sports being the top-selling genres. Overall, this project used visualizations to provide insights into the performance of the video game industry in the past two decades.

### 6. Reflective Discussion

Data cleaning is inevitable for any data-related project. The learning curve is substantially steeper in Tableau. On the other hand, Power BI is only available for Windows, making it difficult for Mac users to use it. Indeed, I have a Mac system, but couldn't use it for this coursework, had to use a secondary laptop. The implementation of 'Action' based on the filter has been a mere static visualization. Overall, the module has provided fantastic insights into the many data visualization concepts, such as the necessity of optimizing the data into percentage differences, maintaining consistency between views, making the best use of various charts, etc. The ability to build visualizations utilizing various data sources, as well as from many sources at once, will be considered a personal learning target.

## 7. References

Playfair Data. (2017). 3 Ways to Make Lovely Line Graphs in Tableau. [online] Available at: <https://playfairdata.com/3-ways-to-make-lovely-line-graphs-in-tableau/> [Accessed 8 Apr. 2023].

www.youtube.com. (n.d.). Tableau - Sort. [online] Available at: <https://www.youtube.com/watch?v=LfhPo6PdRYk&list=PLWPirh4EWFpGXTBu8ldLZGJCueTMBpJFK&index=11> [Accessed 8 Apr. 2023].

www.youtube.com. (n.d.). Tableau - Actions Filter. [online] Available at: <https://www.youtube.com/watch?v=BnYCwgX7hZQ>.

community.powerbi.com. (2022). Need to calculate Year-over-Year Growth for 'year' data. [online] Available at: <https://community.powerbi.com/t5/Desktop/Need-to-calculate-Year-over-Year-Growth-for-quot-year-quot-data/td-p/2264711> [Accessed 8 Apr. 2023].

www.pluralsight.com. (n.d.). Build Scatter Chart in Power BI | Pluralsight. [online] Available at: <https://www.pluralsight.com/guides/build-scatter-chart-in-power-bi> [Accessed 8 Apr. 2023].

maggiesMSFT (n.d.). Add a filter to a report in Power BI - Power BI. [online] learn.microsoft.com. Available at: <https://learn.microsoft.com/en-us/power-bi/create-reports/power-bi-report-add-filter?tabs=powerbi-desktop>.

www.linkedin.com. (n.d.). Different Types of Charts in Power BI And How To Use Them. [online] Available at: <https://www.linkedin.com/pulse/different-types-charts-power-bi-how-use-them-jennifer-balsom/> [Accessed 8 Apr. 2023].

maggiesMSFT (n.d.). Sort one column by another column in Power BI - Power BI. [online] learn.microsoft.com. Available at: <https://learn.microsoft.com/en-us/power-bi/create-reports/desktop-sort-by-column?tabs=powerbi-desktop> [Accessed 17 Apr. 2023].

www.youtube.com. (n.d.). How to Create Power BI TOP N Report (Power BI Top 10 / Top N Filter). [online] Available at: <https://www.youtube.com/watch?v=4N-KocuYyNU> [Accessed 8 Apr. 2023].

www.youtube.com. (n.d.). How to Create Power BI TOP N Report (Variable TOP N Filter). [online] Available at: <https://www.youtube.com/watch?v=wKfPS1yZfEO> [Accessed 8 Apr. 2023].

## 8. Appendix

As part of the coursework, an archive file has been attached containing all supporting project documents, which includes a Tableau file, Power BI file, original and cleaned dataset. The archive also contains the python code file (Jupyter notebook) used for the data cleaning process. This comprehensive list of documents is intended to support and enhance the project's overall quality, allowing for a thorough analysis and evaluation of the project's findings.