

Project Reflections

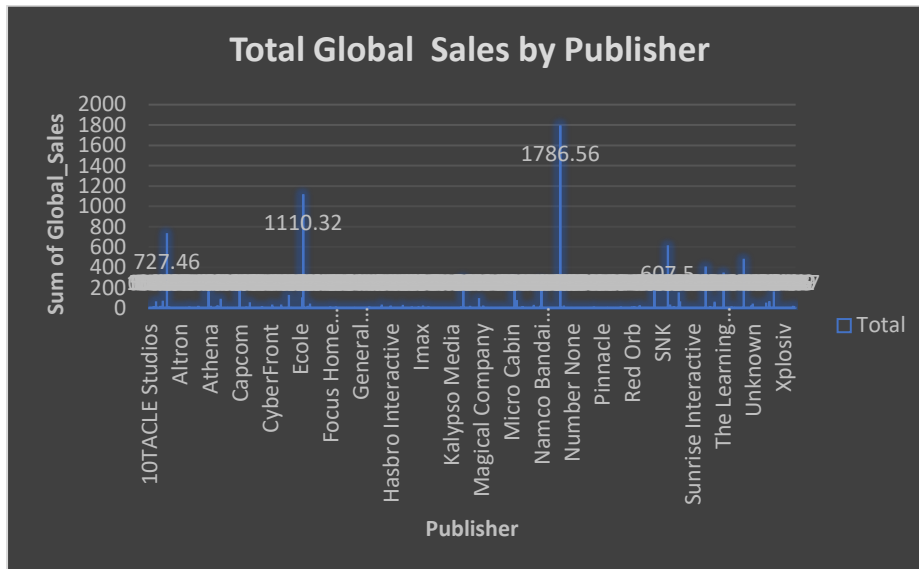
During the analysis of GameCo's data, one insight stood out as the most interesting and relevant for the executives. The insight relates to the sales performance of games in different regions over time. The data revealed a significant decline in-game sales, which prompted further investigation by the marketing team. The process that led to this insight involved grouping and summarizing the data and creating visualizations to uncover patterns and trends.

Data Snapshot

Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sal	Global_Sales
1	Wii Sports	Wii	2006	Sports	Nintendo	41.49	29.02	3.77	8.46	82.74
2	Super Mario Bros.	NES	1985	Platform	Nintendo	29.08	3.58	6.81	0.77	40.24
3	Mario Kart Wii	Wii	2008	Racing	Nintendo	15.85	12.88	3.79	3.31	35.82
4	Wii Sports Resort	Wii	2009	Sports	Nintendo	15.75	11.01	3.28	2.96	33
5	Pokemon Red/Pokemon Blue	GB	1996	Role-Playing	Nintendo	11.27	8.89	10.22	1	31.37
6	Tetris	GB	1989	Puzzle	Nintendo	23.2	2.26	4.22	0.58	30.26
7	New Super Mario Bros.	DS	2006	Platform	Nintendo	11.38	9.23	6.5	2.9	30.01
8	Wii Play	Wii	2006	Misc	Nintendo	14.03	9.2	2.93	2.85	29.02
9	New Super Mario Bros. Wii	Wii	2009	Platform	Nintendo	14.59	7.06	4.7	2.26	28.62
10	Duck Hunt	NES	1984	Shooter	Nintendo	26.93	0.63	0.28	0.47	28.31
11	Nintendogs	DS	2005	Simulation	Nintendo	9.07	11	1.93	2.75	24.76
12	Mario Kart DS	DS	2005	Racing	Nintendo	9.81	7.57	4.13	1.92	23.42
13	Pokemon Gold/Pokemon Silver	GB	1999	Role-Playing	Nintendo	9	6.18	7.2	0.71	23.1
14	Wii Fit	Wii	2007	Sports	Nintendo	8.94	8.03	3.6	2.15	22.72
15	Wii Fit Plus	Wii	2009	Sports	Nintendo	9.09	8.59	2.53	1.79	22
16	Kinect Adventures!	X360	2010	Misc	Microsoft Game Studios	14.97	4.94	0.24	1.67	21.82
17	Grand Theft Auto V	PS3	2013	Action	Take-Two Interactive	7.01	9.27	0.97	4.14	21.4
18	Grand Theft Auto: San Andreas	PS2	2004	Action	Take-Two Interactive	9.43	0.4	0.41	10.57	20.81
19	Super Mario World	SNES	1990	Platform	Nintendo	12.78	3.75	3.54	0.55	20.61

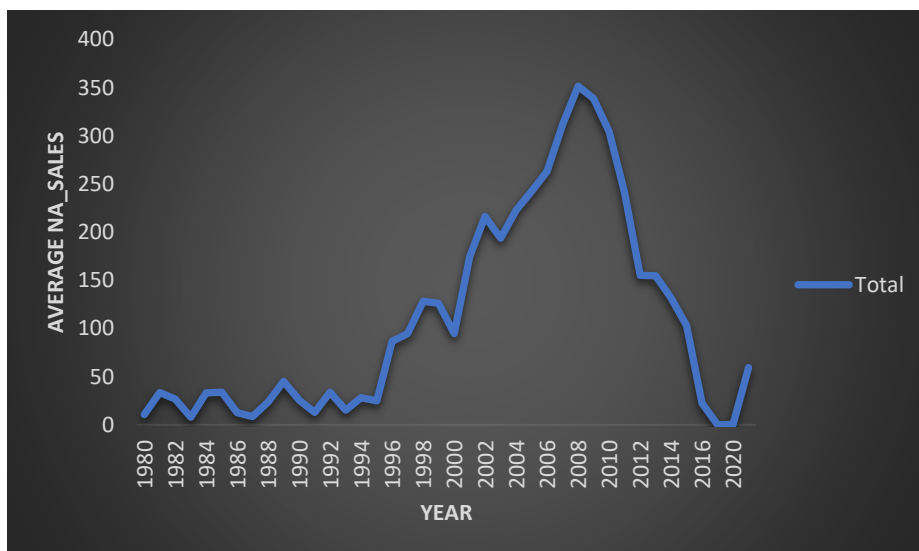
There are 16599-Rows, and 11- Columns. Columns are Characteristics. The column headers typically provide labels or names for each column in the data set. There are 271 missing data in Column D (year) and 58 missing data in Column F (publisher). The JP Sales were recorded as 0 for most of the year. Differences in distribution channels and marketing strategies can impact. Regions with robust distribution networks, effective marketing campaigns, and strong partnerships with retailers may experience higher sales compared to regions with limited distribution channels. The 0 values could be a result of missing or incomplete data for those specific sales. It is possible that the information for those sales was not recorded. Each observation could represent a different video game title. This would allow for tracking and analyzing various games individually. Each observation could represent a different video game title. In column D (year) and column F (Publish) there are many N/A records. NA Units are recorded as 0. The last record 16601 is blank. If N/A records are not cleaned from a data sample in an Excel sheet, it can have various consequences depending on how the data is being used and analyzed the data is used for creating charts or visualization, and the presence of N/A records can impact the accuracy and clarity of the visual representation. The unnecessary blank rows can contribute to increased file size, especially when working with large datasets. N/A records indicate that specific information is missing for those records. 1 duplicate data found and removed. The NA_Units column was deleted since it doesn't contain any useful information. Removed row 16601 since it was empty. Removed all the N/A records from column D (year) and column F(publish) and deleted the empty row (16601) and lastly deleted the NA_Units column.

Step 1: I first organized the sales data by region and year to the group and summarize the data. This allowed me to examine the sales performance of games in each region over time. I calculated the total sales for games in each region and created visualizations such as line charts and column charts to visualize the sales trends. By comparing the sales figures across different years and regions, I was able to identify the decline in sales for games



Graph-1 A Column Chart of total global sales by Publisher.

Nintendo – 1786.56 has the highest total global sales.



Graph-2 A line chart of average NA_Sales by Year.

To understand a trend over time. I created a line chart to visualize the trend over the years. The average sales in NA went through an upward trend in the early 1980s, although there were some big fluctuations in the mid-'80s when sales dropped drastically. The sales reached its peak in 1989, and thereafter it has been on a downward trend. Sales were consistently below 0.5 level from the mid-'90s and reached their lowest level in 2017.

We have 11 columns. Each column represents characteristics. I categorized each column as quantitative/qualitative, discrete/continuous, or binary/nominal/ordinal.

Category	Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_sales	Global_Sales
Quantitative	*			*			*	*	*	*	*
Qualitative		*	*		*	*					
Continuous											
Discrete	*			*			*	*	*	*	*
Nominal		*	*		*	*					
Binary											
Ordinal											

Step 2: Data Analysis.

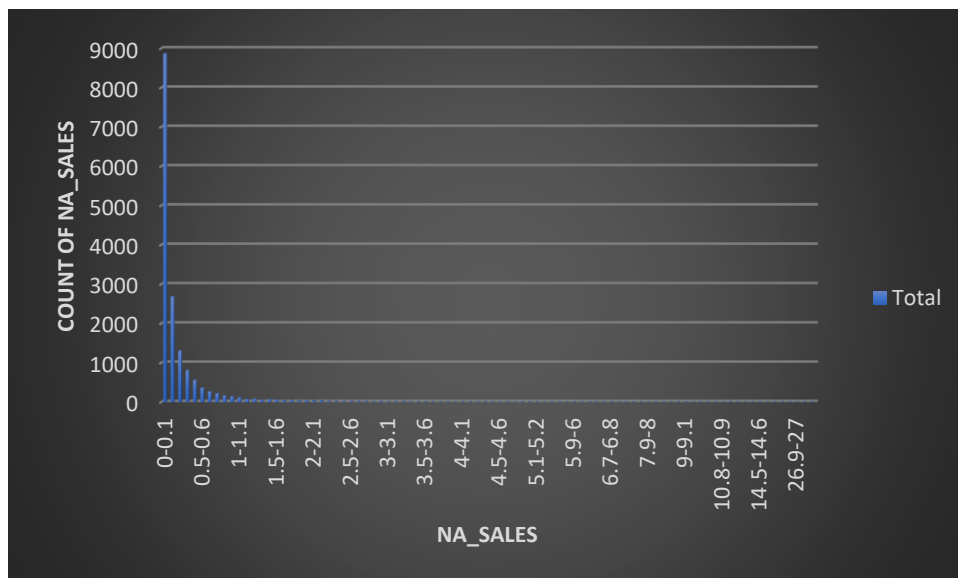
Exploratory Data Analysis (EDA) is a crucial step in data analysis that helps to understand and analyze the characteristics, patterns, and relationships within a dataset.

The Measure of Central Tendency Columns are NA_Sales, EU_Sales, JP_Sales, Other_Sales, and Global_Sales.

	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
Mean	0.26	0.15	0.08	0.05	0.54
Median	0.08	0.02	0	0.01	0.17
Mode	0	0	0	0	0.02

Table 1. Measures of Central Tendency.

As you can see from the above table the regions differ from each other. The mean is larger because the presence of outliers or extreme values in the right tail of the distribution can pull the mean towards higher values. The Distribution is Right Skewed since more data points are concentrated toward the left.



Graph – 3. A Histogram for NA_Sales with the values grouped by 0.1.

IQR measures the spread of the data in a way that's less susceptible to outliers as it measures the difference between the Q1 and Q3 percentiles of a data set.

To calculate IQR you need to know Quartiles.

Q1 – 0, Q2- 0.08, Q3 - 0.24

The IQR for NA_Sales - 0.24

The range for NA_Sales = MAX value - MIN value

41.49-0

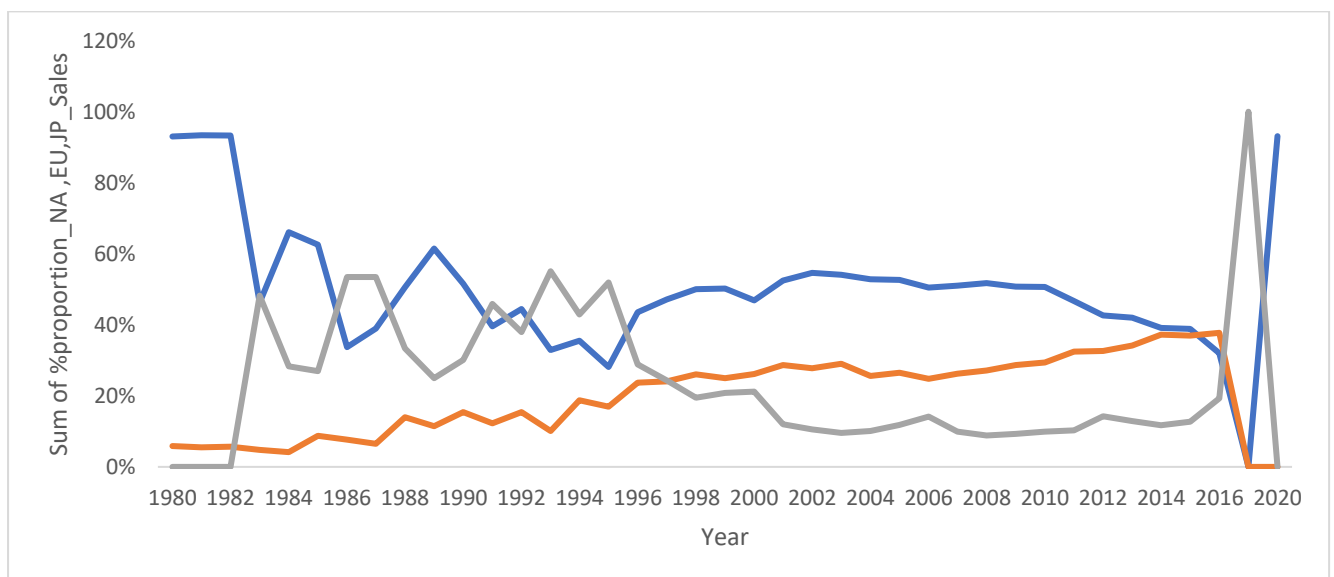
Range for NA_Sales = 41.49

There are 2 outlier values for NA_Sales -0.36 and 0.6

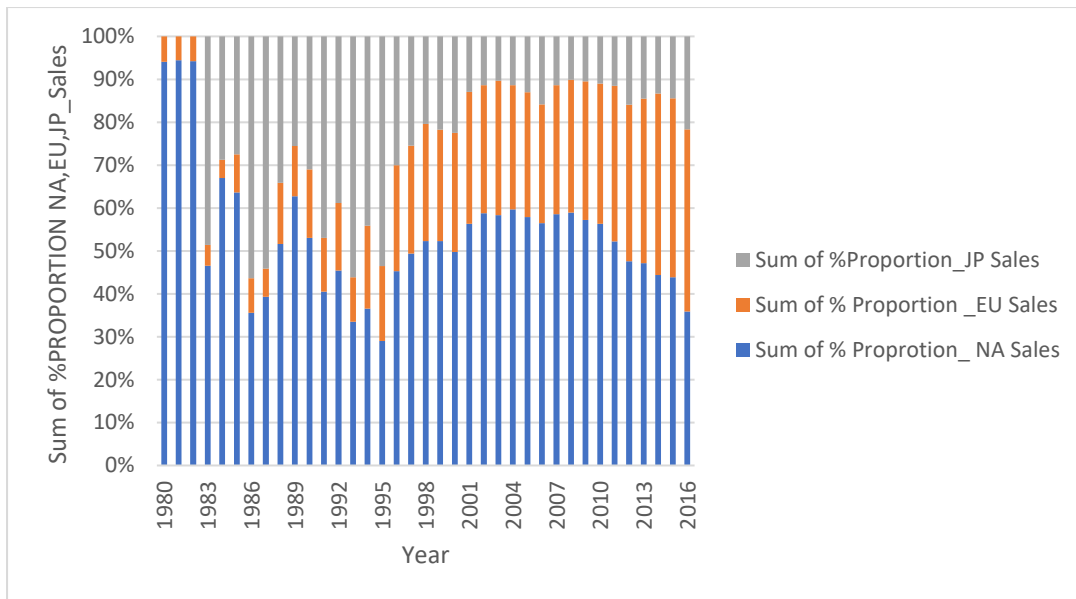
I would handle these observations by determining whether the outliers are due to measurement errors, data entry mistakes or represent genuine extreme values.

Step 3: Data Visualisation.

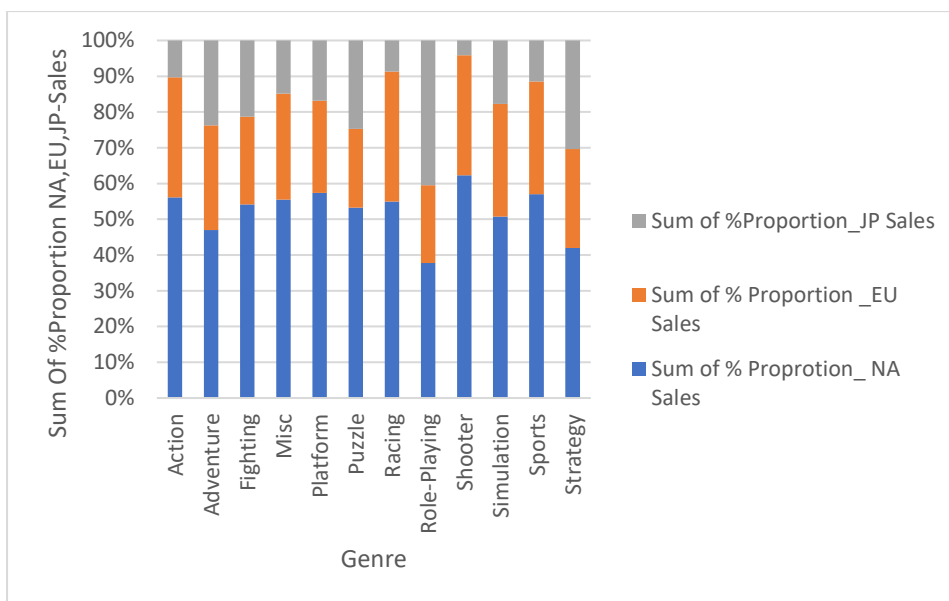
Three variables (NA, Europe, and Japan) represent the proportion of Global Sales for these regions.



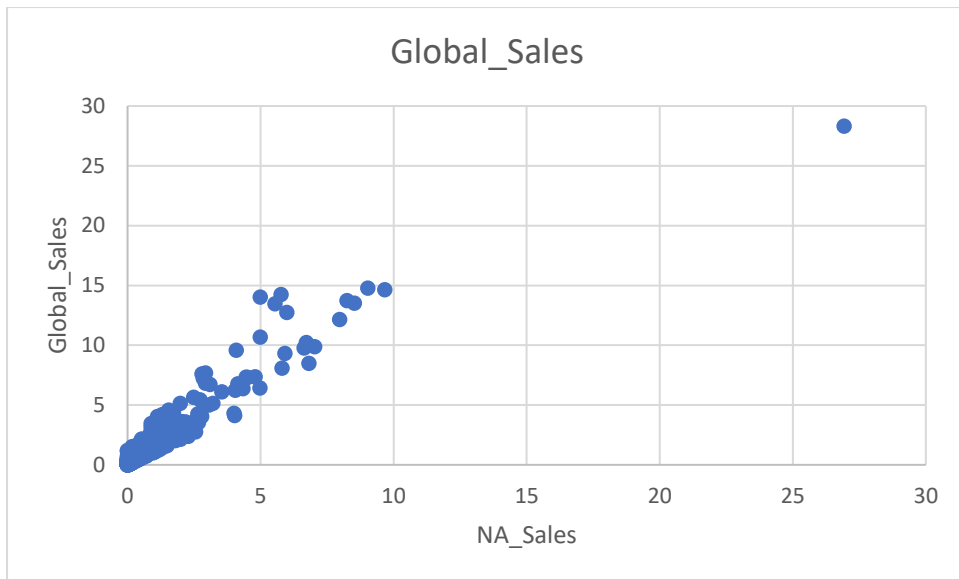
Graph 4. A line chart for three variables (NA, EU, JP_Sales) by Year (1980-2020)



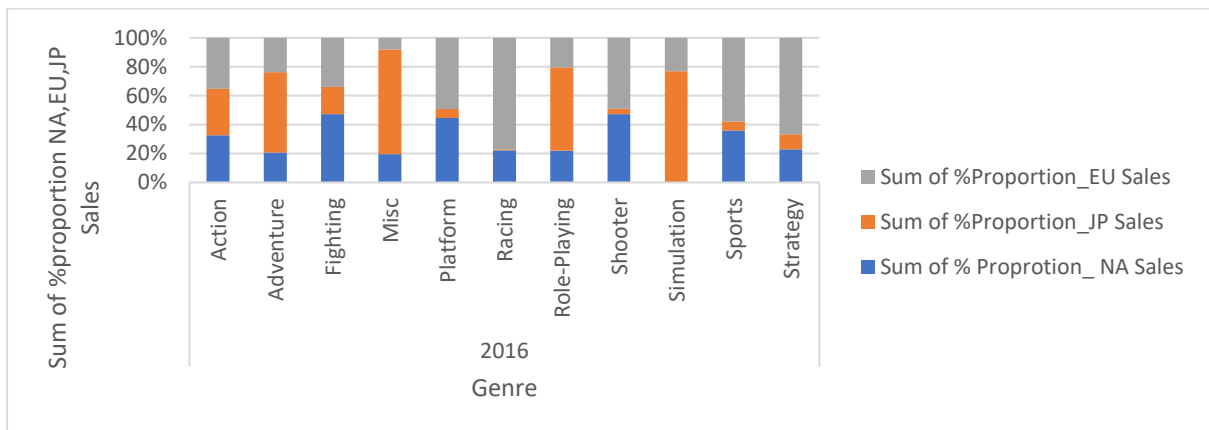
Graph 5. 100% Stacked bar chart with Year(1980-2016) & NA_Sales, EU_Sales, JP_Sales, making up the proportions.



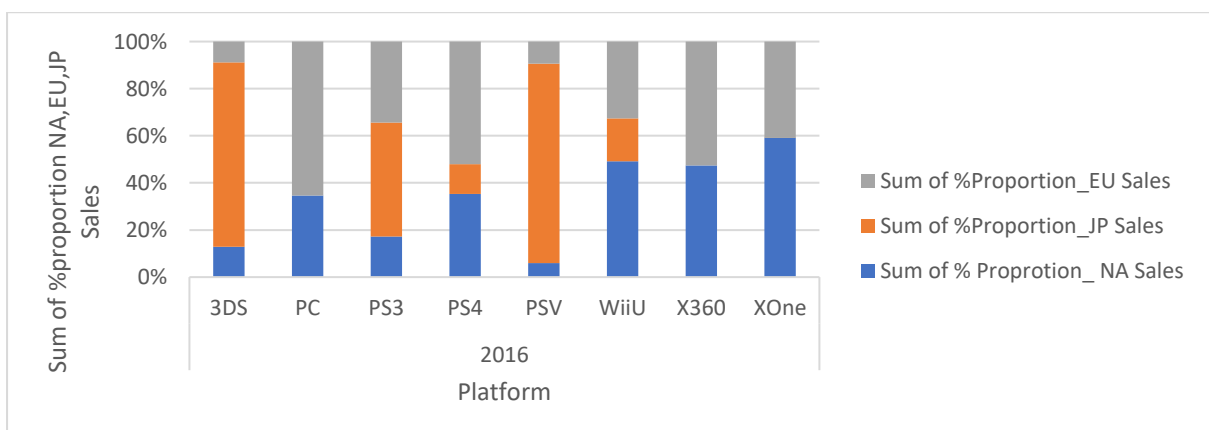
Graph 6. A 100% Stacked bar chart with Genre and NA, EU, JP_Sales.



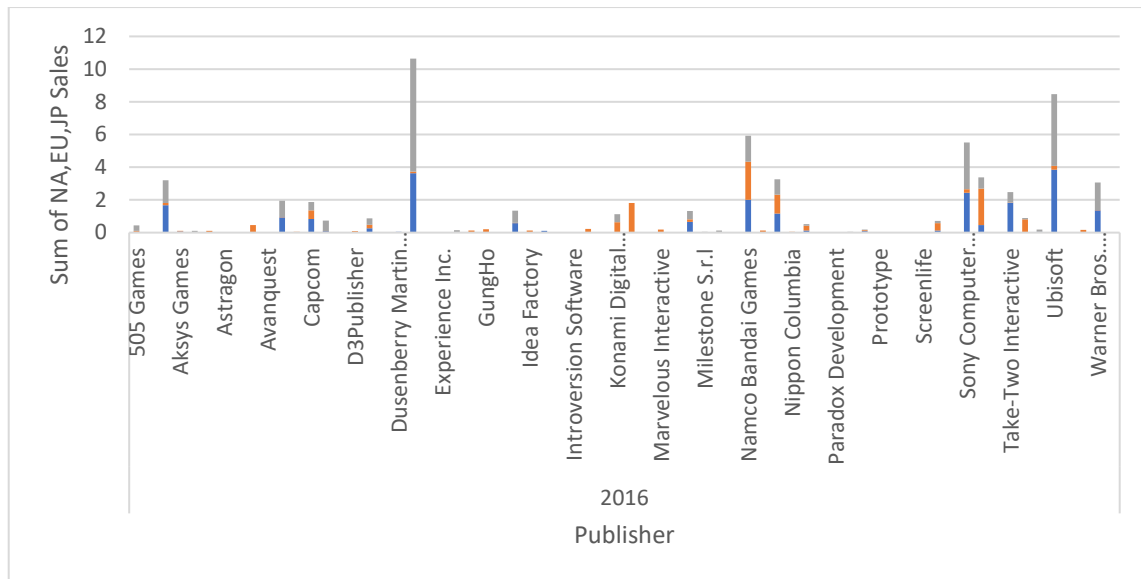
Graph 7. A Scatterplot of NA_Sales & Global_Sales.



Graph 8. A 100% Stacked bar chart of the Sum of %Proportion NA, EU, JP_Sales by Genre for the year 2016.



Graph 9. A 100% Stacked bar chart of the Sum of %Proportion NA, EU, JP_Sales by Platform for year.



Graph 10. A Stacked Column Chart of the Sum of NA, EU, and JP Sales by Publisher for the year 2016.

Step 4 : One of the visualizations that supported this insight was a line chart showing the sales trends of puzzle games in each region over time. The chart depicted a downward trend in sales for the current year, indicating a significant drop compared to previous years. Another visualization that provided valuable insights was a 100% Stacked bar graph showing the sales distribution of different game genres in each region. This allowed me to compare the performance of NA, EU, and JP_Sales with other genres and identify the specific decline in the game sales.

The regions differ from each other. Each region does not stay constant over the years as you see in the graph. NA_Sales have the highest sales. JP_Sales have the lowest sales. From the year 1995-2014, JP_Sales gradually decreased. To understand why a region behaves differently than expected, it is important to conduct a thorough analysis, review market data, gather feedback from the customer and stakeholders, and evaluate the various factors that could have influenced the outcome of the sale. This analysis can help identify areas for improvement, adjust strategies, and make more informed decisions for future planning. (Graph 4)

Role-playing 38% has the highest proportion of Japanese Sales. Shooter 56% has the highest proportion of NA_Sales. (Graph 6). JP_Sales (55%) had the highest sales in the year 1993, NA_Sales(93%) has the highest sales in 1982 and the lowest(28%) in the year 1995. (Graph 5)

The top 3 genres of NA_Sales for the year 2016 are Shooter, Sports, and Action. The top 3 genres of EU_Sales are Racing, Strategy, and Sports. The top 3 genres of JP_Sales are Simulation, Misc, and Role-playing. (Graph 8)

The process of grouping and summarizing the data, combined with visualizations, enabled me to uncover insight regarding the decline in game sales. This insight has significant implications for GameCo's marketing team, as it highlights the need to understand the factors contributing to the decline and develop strategies to address the issue. By focusing on this insight, GameCo's executives can prioritize their efforts to revive game sales and make informed decisions to optimize their product offerings and marketing campaigns.