The insights I discovered are:

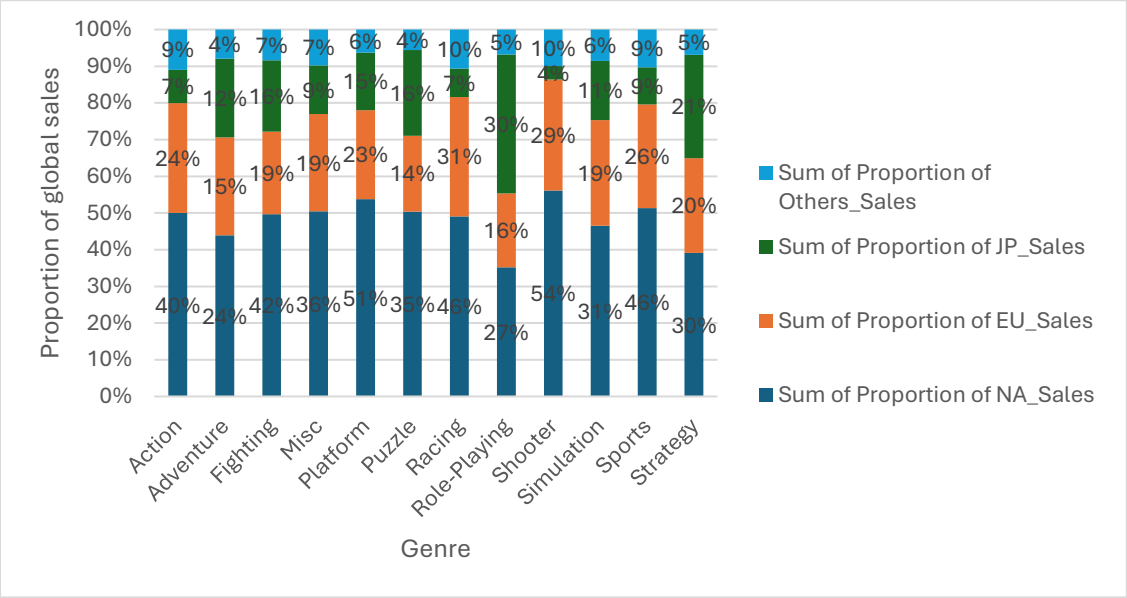
1. Action games have the most unit sale in North America, while shooter games have the highest proportion of North American sales.
2. The relationship between the North American sales and the Global sales has Positive Correlation.
3. Most sale values in North American exceed the maximum threshold of 0.6million units which makes the sales in NA to have more sale values as outliers.
4. The distribution of the NA_Sales variable is right skewed which has the mean to be higher than the median.
5. Whenever the North American sales fall below expectations, the proportion of sales for Action or Shooter games that year tends to be below average.

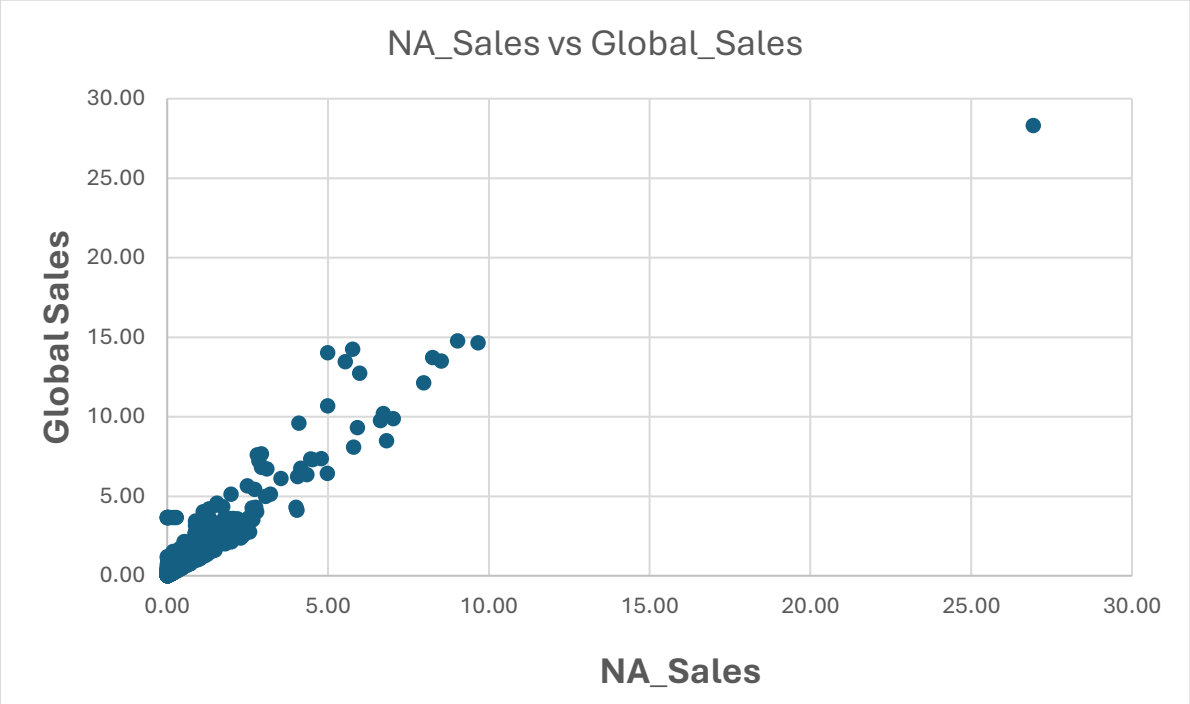
PROJECT REFLECTION

The analysis of GameCo Data began with uploading the dataset, which consists of 11 variables. I first focused on understanding the meaning and significance of each variable. Next, I cleaned the data by filtering out irrelevant entries, addressing missing values, and ensuring a complete and accurate dataset for analysis.

To explore relationships within the data, I utilized Pivot Tables to group variables and analyze their interactions. Additionally, I created new columns to calculate the proportion of sales in each region, allowing me to better understand the contribution of each game genre to global sales.

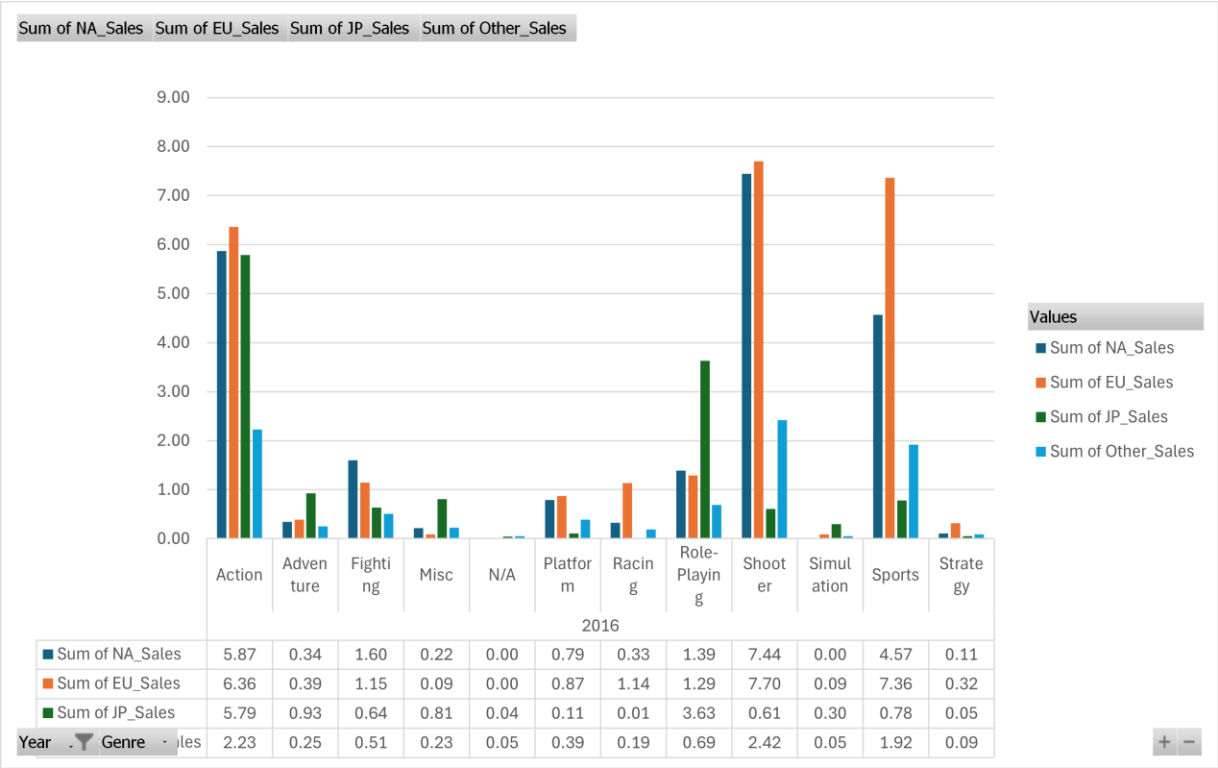
For visualization, I employed a 100% stacked column chart to illustrate the proportion of sales by genre across different regions, providing a clearer perspective on the global sales distribution. I also used a box-and-whisker plot to examine the distribution of sales in each region and identify outliers. Finally, I utilized a scatter plot to visualize and assess the correlation between the various sales variables, offering deeper insights into the relationships within the data.


Proportion of global sales
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Action
Adventure
Fighting
Misc
Platform
Puzzle
Racing
Role-Playing
Shooter
Simulation
Sports
Strategy
Genre
Sum of Proportion of Others_Sales
Sum of Proportion of JP_Sales
Sum of Proportion of EU_Sales
Sum of Proportion of NA_Sales


NA_Sales vs Global_Sales
Global Sales
30.00
25.00
20.00
15.00
10.00
5.00
0.00
0.00
5.00
10.00
15.00
20.00
25.00
30.00
NA_Sales

DIFFERENT REDIFINED CHART TO VISUALIZE THE GAMECO 2016 ANALYSIS.

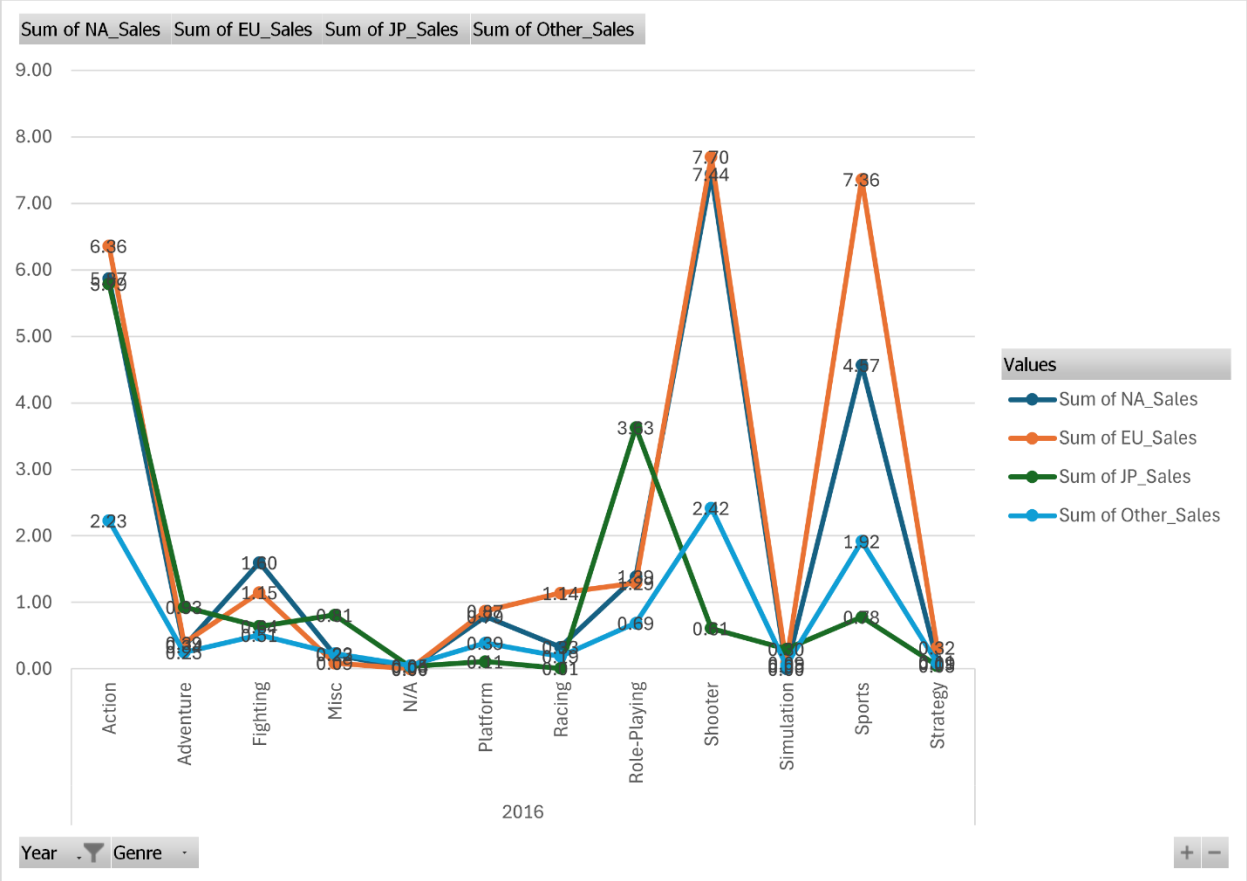
1. Clustered column chart for genre and the sum of sales in different regions in 2016.



Year	Genre	Action	Adventure	Fighting	Misc	N/A	Platform	Racing	Role-Playing	Shooter	Simulation	Sports	Strategy
2016	Sum of NA_Sales	5.87	0.34	1.60	0.22	0.00	0.79	0.33	1.39	7.44	0.00	4.57	0.11
	Sum of EU_Sales	6.36	0.39	1.15	0.09	0.00	0.87	1.14	1.29	7.70	0.09	7.36	0.32
	Sum of JP_Sales	5.79	0.93	0.64	0.81	0.04	0.11	0.01	3.63	0.61	0.30	0.78	0.05
	Sum of Other_Sales	2.23	0.25	0.51	0.23	0.05	0.39	0.19	0.69	2.42	0.05	1.92	0.09

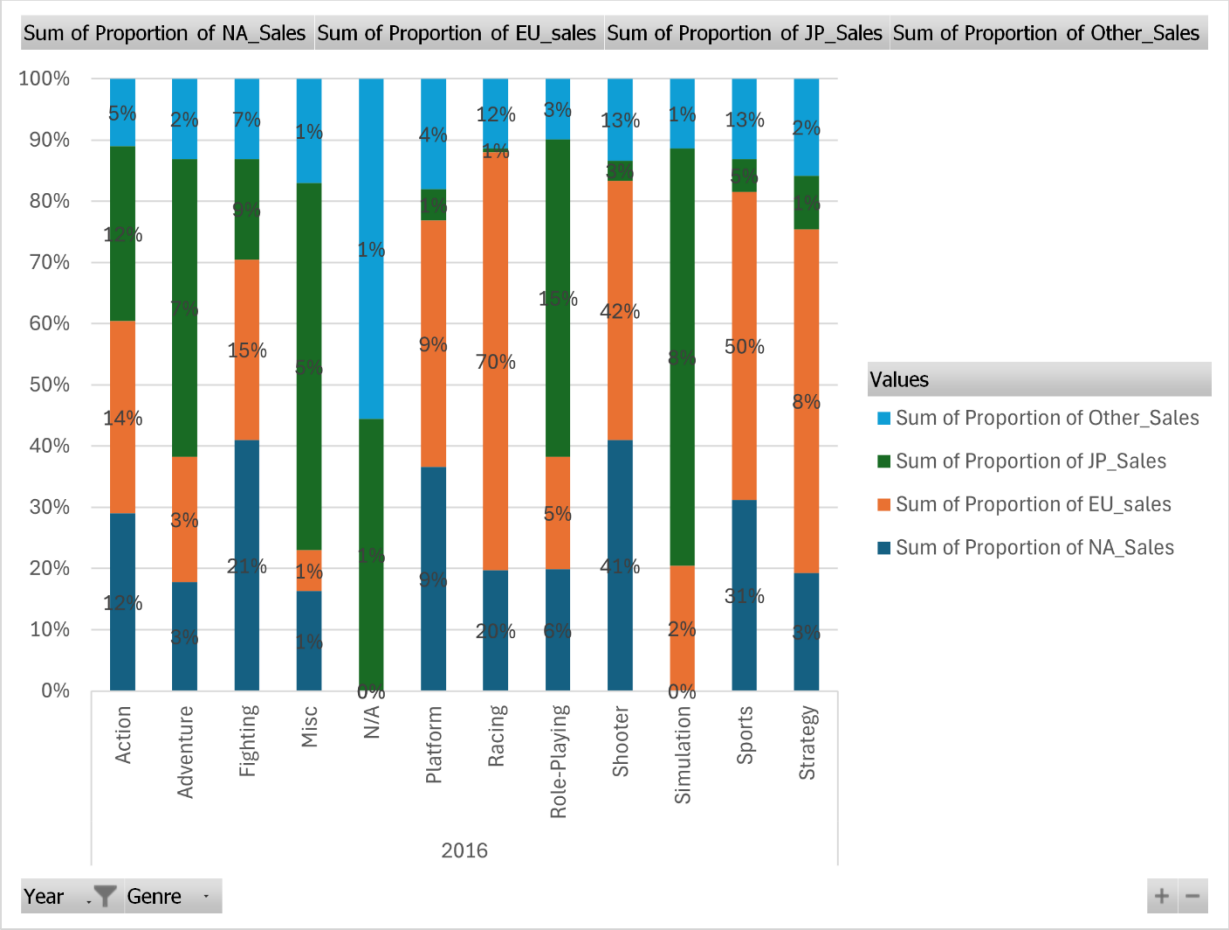
I chose the clustered column chart to represent the sum of sales in each region for 2016 because it enables an easy visual comparison of multiple data series within the same category. This chart provides a clear, side-by-side view of the values across different regions, making it especially effective for identifying trends and highlighting significant differences between categories and subcategories in the dataset.

2. Line chart for genre and sum of sales of different regions in 2016.



The line chart effectively visualizes how a variable changes over time, allowing for easy identification of trends, peaks, and valleys in data.

3. 100% stacked column chart.



100% stacked column helps me to visualize the proportional contribution of sales in each region to the global sales.

Also, this connects to the visualization I used in developing my analysis because the previous 100% stacked column chart comprises of all the years in the dataset while the one I will be using in my final presentation only analyse sales in year 2016.