

Project Reflection: Intro To Data Analytics

Background

The expectations of GameCo is that the sales across various geographical locations have stayed the same over time. To test this theory, I have looked into the data of the sales to determine if this is true.

The goal of this project is to conduct a descriptive analysis of a video game data set in order to have a better understanding of how GameCo's games would do in the market. To fulfill this goal, the first step is to identify what questions we need to answer with this data set:

- Which geographical region sold the most games
- Data trends in the past years up until 2016
- Game Genres and Publishers that sell the most in each geographical region.
- Games that were the most popular/sold the most in each geographical region
- How many games have decreased and/or increased in popularity over time

To find the answers to these questions, I was provided a dataset named “vgsales” which contained historical records of video game sales.

1. Data Exploration

To do this task, I had to look at the datasheet for GameCo that provided the ranks, names, platforms, year, genre, publisher, global sales, and geographical region sales. I identified observations and characteristics of the data set such as there were 16,599 rows including the header, and 11 columns. I also identified that the data set included data from between 1980 to 2020.

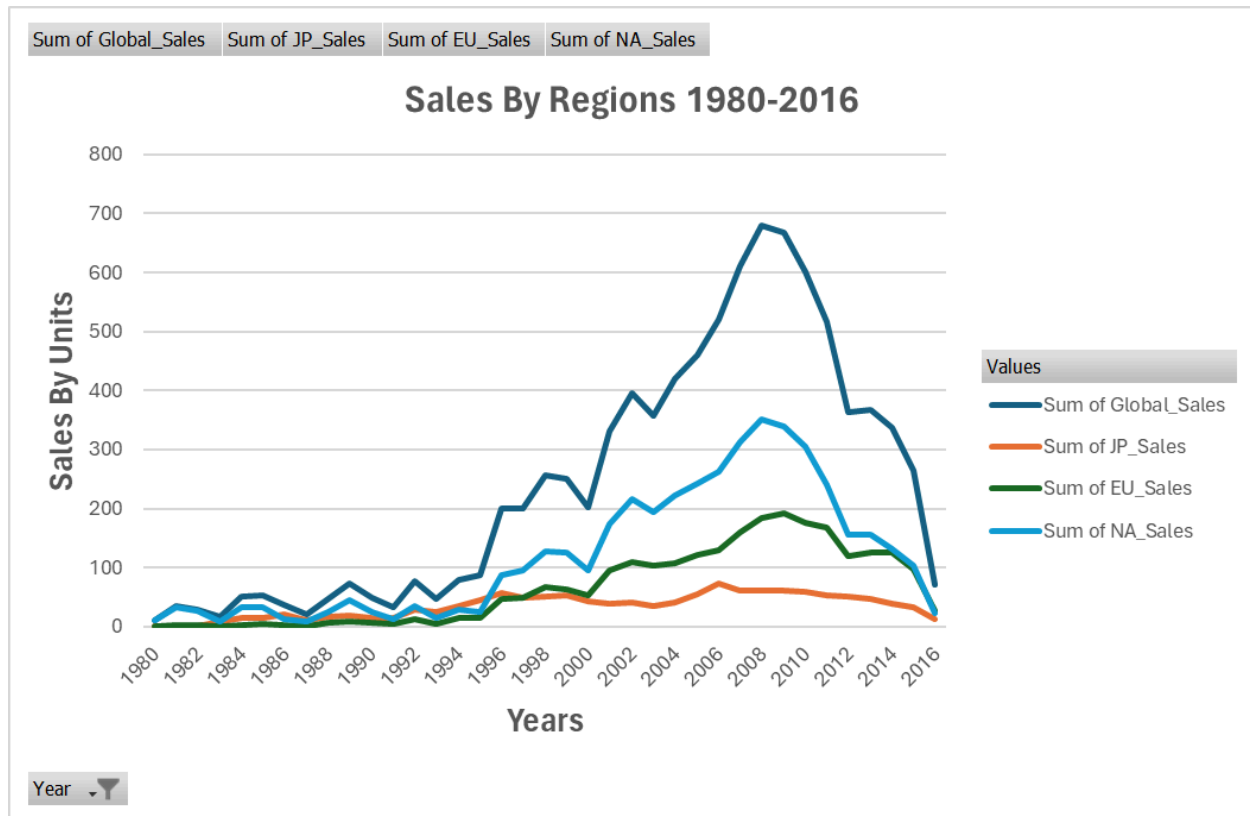
2. Data Cleaning

A lot of the data that was provided had missing data in many of the rows and columns. I filtered and cleaned 271 N/A values, and removed 16 rows since it was not useful to the data. After filtering and cleaning, I created pivot tables and graphs to show the sales for each geographical location.

3. Data Analysis

Many different analyses were conducted such as the following:

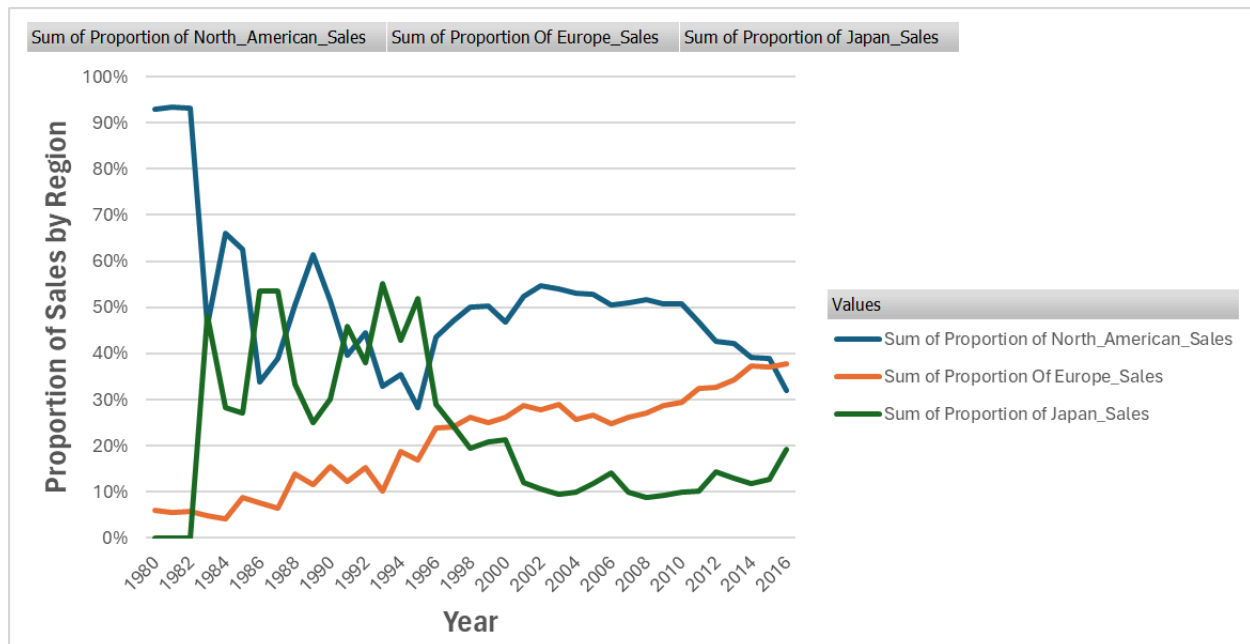
Line Chart of The Different Sales by Each Region 1980-2016



3a. Total Sales by Region

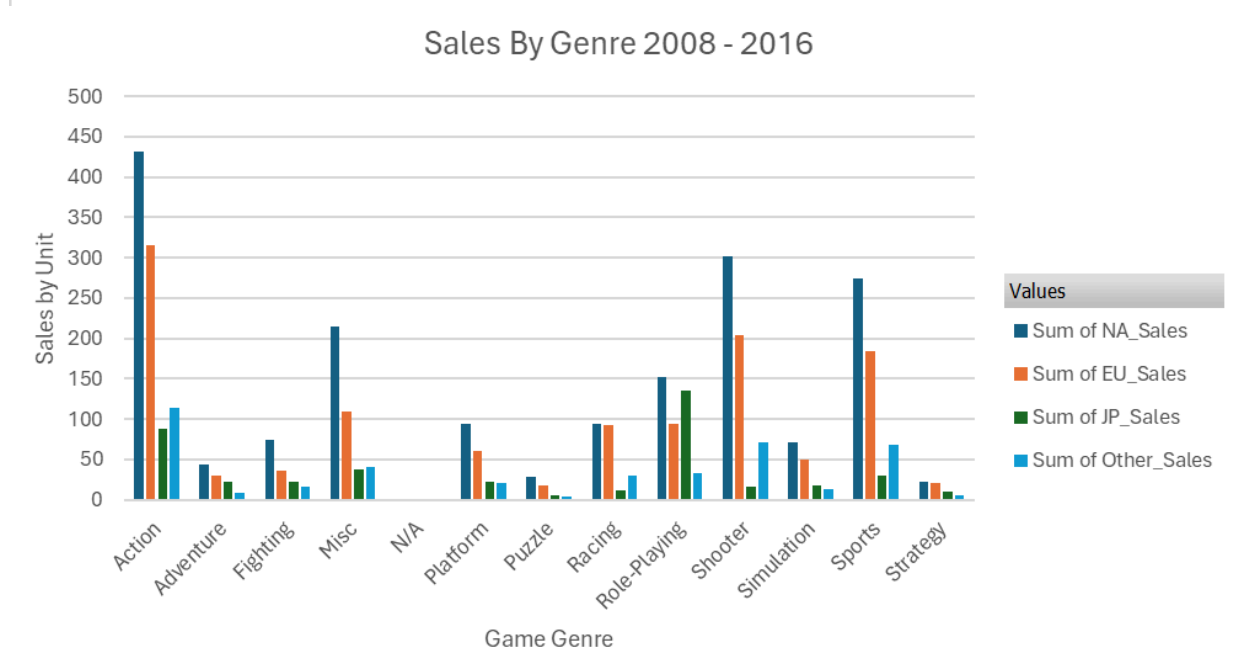
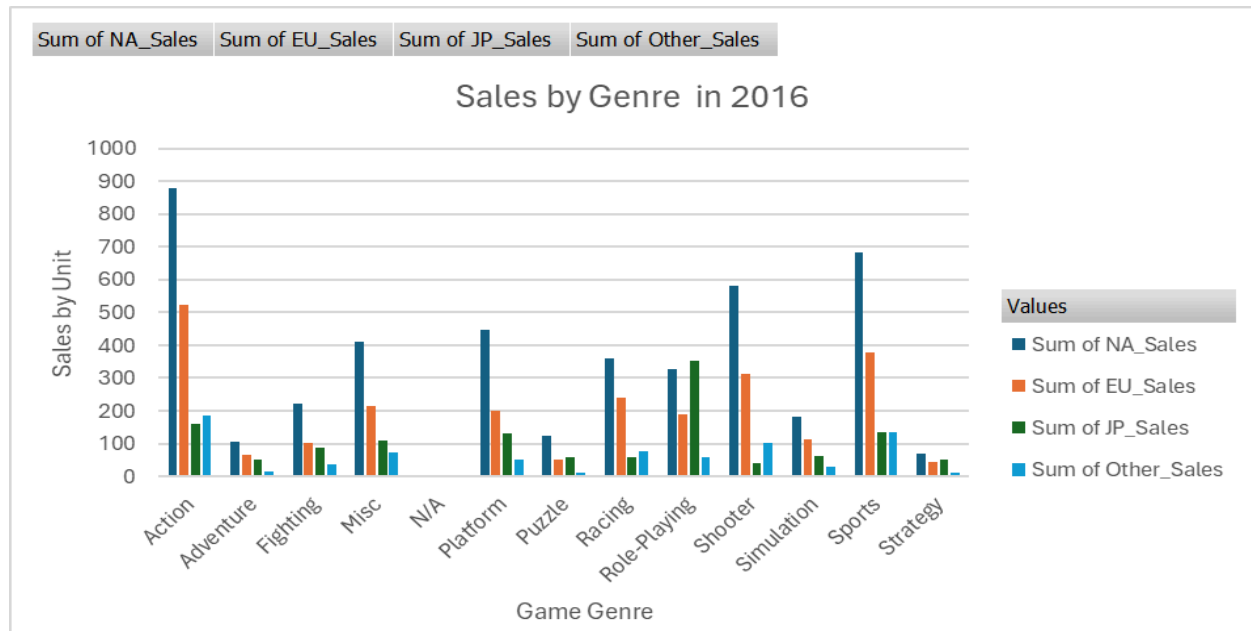
The data chart above shows the data trends of the sales by each region from the years 1980 to 2016. Looking at the data, I can see that there was a spike in sales in 2008, but those sales decreased in 2010 in all regions. The graph also shows that there was a familiar pattern in sales across all regions between 1982 and 1994, but in 1996 the sales in North America spiked, becoming higher than the other two regions. North America has the highest sales, Europe has the second highest, and Japan has the lowest sales. These trends, however, changed from 2014 to 2016 when North America sales became lower than European sales by 4.10 million units.

Line Chart of The Different Proportions for Every Sales From 1980 - 2016



3b. Proportion of Total Sales

The line chart above shows the sum of the proportion of sales in each region. This shows that the data shows an ongoing trend of deferred sales figures across different areas beginning from the year 1980. Given the previous year's pattern, it is realistic to expect that regional sales performance will continue to change from year to year in the future. Also, the data, with the exception of Europe, shows no consistent pattern in sales from year-to-year sales. For the majority of the years, North America continuously dominated in sales, but there has been a recent downturn. In contrast, both Europe and Japan have seen an increase in sales in recent years. Europe also has a consistent increase in sales, and surpasses Japanese and North American sales in 2016. This can show that in the years coming, Europe might continue to succeed in sales compared to North America and Japan.



3c. Sales by Genre

The bar graph of the different sales by each genre is shown above. This data shows that North America has dominated almost every genre, with the exception of role-playing, in which Japan had the most sales. This chart can provide a valuable insight in what genres might be selling the most in the years after 2016. Action games sell the most in North America by a lot compared to the other regions. There has also been an

upward trend of sales in some genres such as sports, sports has increased in sales since 2008.

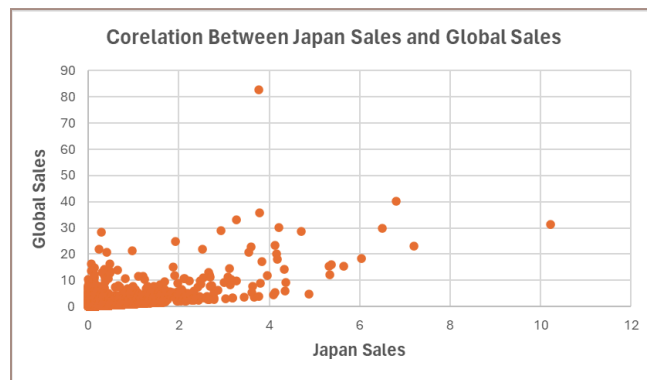
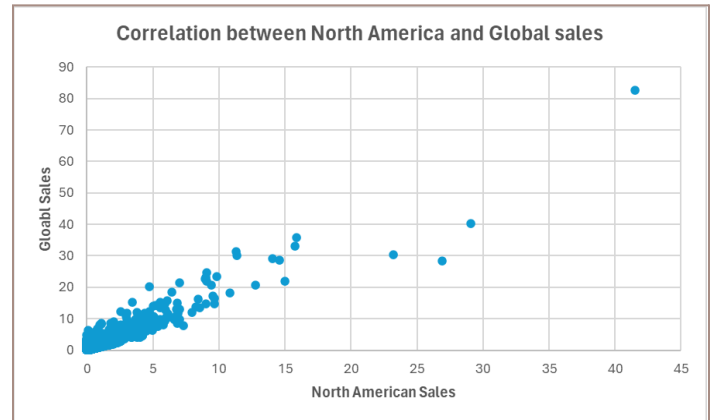
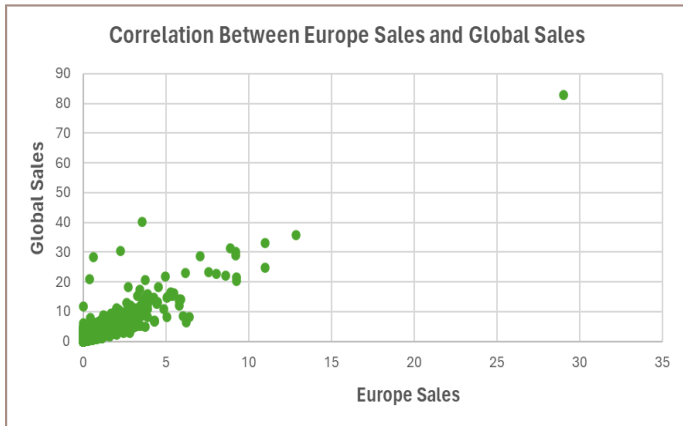
Top 10 Games in Every Geographical Region

North America	Europe	Japan	Other	Global
Wii Sports	Wii Sports	Pokemon Red/Pokemon Blue	Grand Theft Auto: San Andreas	Wii Sports
Super Mario Bros	Grand Theft Auto V	Pokemon Gold/Pokemon Silver	Wii Sports	Grand Theft Auto V
Duck Hunt	Mario Kart Wii	Super Mario Bros.	Grand Theft Auto V	Super Mario Bros.
Tetris	FIFA 15	New Super Mario Bros.	Gran Turismo 4	Tetris
Grand Theft Auto V	Call of Duty: Modern warfare 3	Pokemon Diamond/Pokemon Pearl	Call of Duty: Black Ops II	Mario Kart Wii
Call of Duty: Black Ops	FIFA 16	Tetris	FIFA Soccer 08	Wii Sports Resort
Super Mario world	FIFA 16	Pokemon Black/ Pokemon White	Pro Evolution Soccer 2008	Pokemon Red/Pokemon Blue
Mario Kart Wii	Call of Duty: Black Ops II	Dragon Quest VII: Warriors of Eden	Call of Duty: Black Ops 3	Call of Duty: Black Ops
Wii Sport Resort	Wii Sport Resort	Pokemon Ruby/Pokemon Sapphire	Call of Duty: Black Ops	Call of Duty: Modern Warfare 3
Call of Duty: Modern Warfare 3	Nintendogs	Animal Crossing: Wild World	Call of Duty: Modern Warfare 3	New Super Mario Bros.

3d. Popular Games in Every Geographical Region

The table chart above shows the top ten games across each geographical region from 2008 to 2016. The table is listed in ascending to descending order by how much each game sells. Looking at the table chart, we can see that Wii Sports is the top seller in almost all geographical regions except Japan and Other. Many of the region's top sellers are similar to other regions top sellers, for instance, Call of Duty: Modern Warfare 3 is a top seller in North America, Europe, Global, and Other regions. The data shows that the game sales vary in every geographical location since the popularity of each game is not the same. This table also provides insights on how well a game might do the following years regarding its popularity in each region. Wii sports will probably be a top seller in the upcoming years because it has been the number one seller in three geographical regions.

Scatter Plots of Geographical Region Sales and Global Sales



3e. Scatter Plots of Geographical region sales and global sales

Above in this chart I made a scatterplot that compared global sales to all three geographical regions (North America, Europe, and Japan). . These scatterplots showed a clear and significant positive relation between all three regions to global sales. Based on the data, when a region is doing well in sales, globally sales are also doing well. There are also a few outliers in each region as well, this means that all regions have a single value that is much higher than the average value.

4. Data Visualizations

I used a variety of images above that support show sales from multiple perspectives. I'm using the image below to show how sales changed over time. The data shown below also has three distinctive phases.

- Phase 1: Years 1980 - 1994, these show a steady but low rise in sales.
- Phase 2: Years 1996 - 2008, these show a steep rise in the increase of sales
- Phase 3: Years 2009 - 2016, these show a quick decline in the decrease of sales.

