Video Game Sales Analysis

Problem Statement:

The global video game industry is a powerhouse in the entertainment sector, generating billions in revenue annually. However, predicting what drives game sales remains a challenge. Factors such as platform, genre, and regional preferences play a significant role but are often misunderstood or misanalyzed. Stakeholders, including publishers, marketers, and developers, struggle to optimize strategies to capitalize on these trends.

This project leverages the **Video Game Sales** dataset to uncover actionable insights into global and regional sales trends, platform success, and genre popularity. By better understanding these dynamics, industry stakeholders can craft strategies to maximize revenue, reach target audiences, and align game development with market demand.

Refer to the following notebook to check further analysis and Python code: EDA ipynb

---

Business Impact:

The analysis of this dataset has the potential to drive strategic decision-making across the video game industry:

- **Identify high-performing genres and platforms**: By understanding which genres and platforms dominate sales, publishers can prioritize development and marketing efforts.
- **Tailor marketing campaigns regionally**: Insights into regional preferences can enable marketers to design campaigns that resonate with local audiences.
- **Optimize partnerships and exclusivity**: Platform owners can use sales data to guide decisions about exclusive partnerships and game availability.
- **Strategize for future releases**: Developers can analyze trends to align game features and designs with evolving consumer demands.
- **Assist investors**: Investors can use these insights to identify profitable sectors and emerging opportunities within the gaming market.

This analysis equips stakeholders with the tools and knowledge needed to stay competitive in an evolving industry.

General Dataset Information:

File Name:	vgsales.csv
Description:	Comprehensive data on video game sales globally
Dataset Details:	Original: 16598 Rows and 11 Columns. After Curation: 16291 Rows and 18 Columns.
Size:	390 KB
Source:	Kaggle

Data Profiling:

- Checking first and last five rows in the Dataframe for general insight.
- Getting the total number of rows and columns (16598 rows & 11 Columns).
- Getting general info about the dataset using the .info() method.(Data Type, Number of cols & rows and null values if present).
- Investigate the number of unique values in each column using the .nunique() method.
- Checking the total number of null values in each column using the .isnull() & .sum() methods.
- Getting summary statistics for both 1)Numerical 2)Categorical columns. Using the method .describe()

Observations:

- 2 out of 11 columns include null values. One of the columns 'Year' can be filled with the median, column 'Publisher' can be filled with most frequent. I'll opt to remove all rows with missing values, since they're an insignificant amount of the dataset and filling 'Year' with median can distort timeline analysis and introduce bias.

Data Wrangling:

- Dealing with null values: (Total number of rows = 16291)
- Feature engineering new features:
    - 'Game_Age'
    - 'NA Sales Percentage', 'EU Sales Percentage', 'JP Sales Percentage' 'Other Sales Percentage'
    - 'Sales Category' low, medium, high
    - 'Publisher Global Sales'
    - 'Genre Global Sales' total sales for each genre

Column Name:
Name
Platform
Year
Genre
Publisher

NA_Sales
EU_Sales
JP_Sales
JP_Sales
Other_Sales
Global_Sales
Game_Age
NA Sales Percentage
EU Sales Percentage
JP Sales Percentage
Other Sales Percentage
Sales Category
Publisher Global Sales
Genre Global Sales

- Dropped one column ‘Rank’.
- Added 8 new columns by feature extraction for future predictive modeling.
- 16291 rows and 18 columns

Target Features:
The following columns will be used to simplify and analyze the problem statement:

- **Name**: Game title, useful for identifying specific high-performing games.
- **Platform**: Console or platform where the game is available.
- **Year**: Release year, critical for identifying sales trends over time.
- **Genre**: Game category, used to evaluate consumer preferences.
- **Publisher**: The company responsible for publishing the game.
- **NA_Sales, EU_Sales, JP_Sales, Other_Sales**: Regional sales data, providing insights into geographical market performance.
- **Global_Sales**: Total worldwide sales, indicating the overall success of a game.
- **Game_Age**: The number of years since the game's release.
- **Sales Category**: Categorization of games (e.g., High or Low sales performance).

- **Publisher Global Sales**: Cumulative global sales for each publisher.
- **Genre Global Sales**: Cumulative global sales for each genre.
- **Sales Percentages (NA, EU, JP, Other)**: Percentages of global sales from different regions

---

Analysis Section: Video Game Sales

The analysis is broken into key areas aligned with the problem statement. Each section highlights a specific aspect of the dataset and provides insights supported by visualizations.

1. Global and Regional Sales Trends

Objective:

To identify how sales are distributed across regions and their contribution to global sales.

Findings:

- **North America** contributes the largest share of global sales, accounting for approximately **49.14%** of total sales.
- **Europe** follows with **27.33%** of sales, while **Japan** has a significant share despite being a smaller market.
- **Other regions** collectively make up a modest proportion of sales, indicating room for growth in emerging markets.

Visualizations:

Chart 1: A pie chart illustrating the sales contribution of each region to the global total.

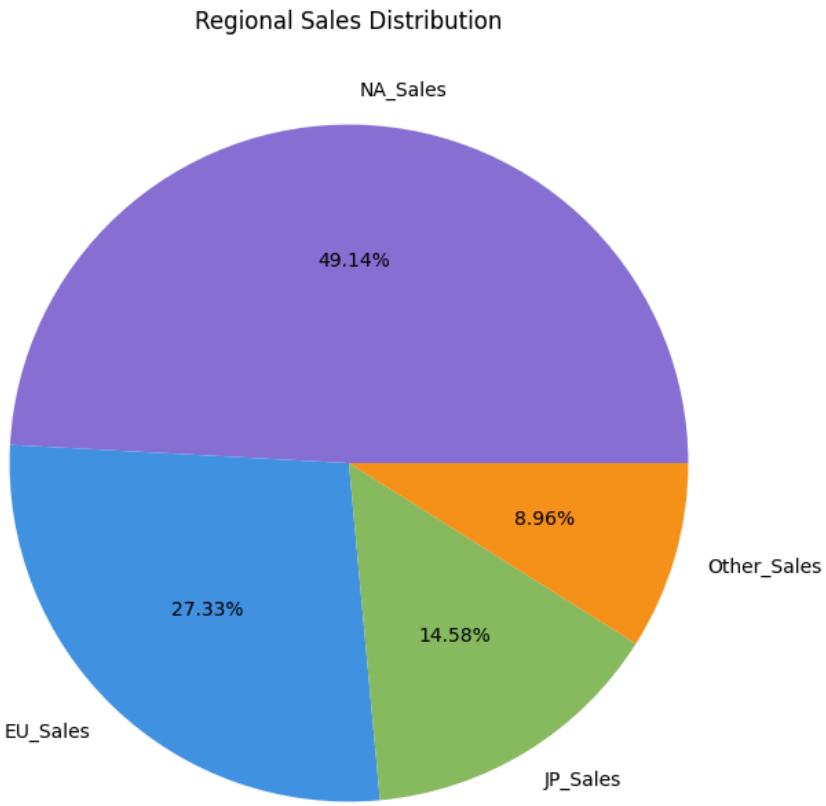
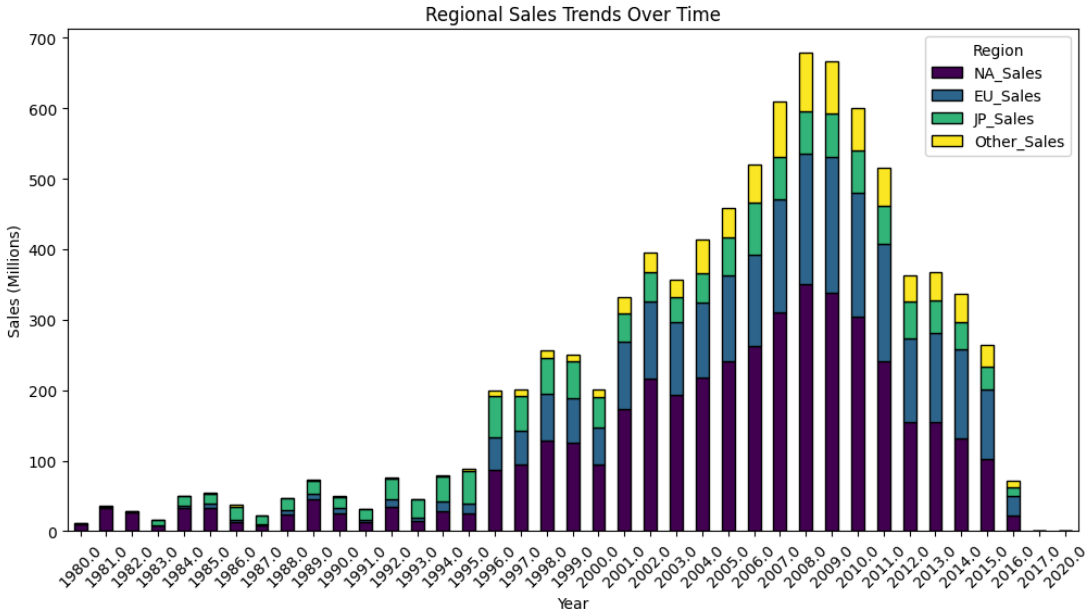


Chart 2: A stacked bar chart showing sales trends across regions over time.



2. Genre Popularity

Objective:

To uncover which game genres dominate global sales and analyze their regional performance.

Findings:

- **Sports**, **Action**, and **Shooter** games lead in global sales, with cumulative sales significantly outpacing other genres.
- Role-playing games are particularly popular in Japan, contributing a disproportionate share of sales in the region.
- Action dominating all regions except for Japan.

Visualizations:

Chart 1: A bar plot for total global sales by genre.

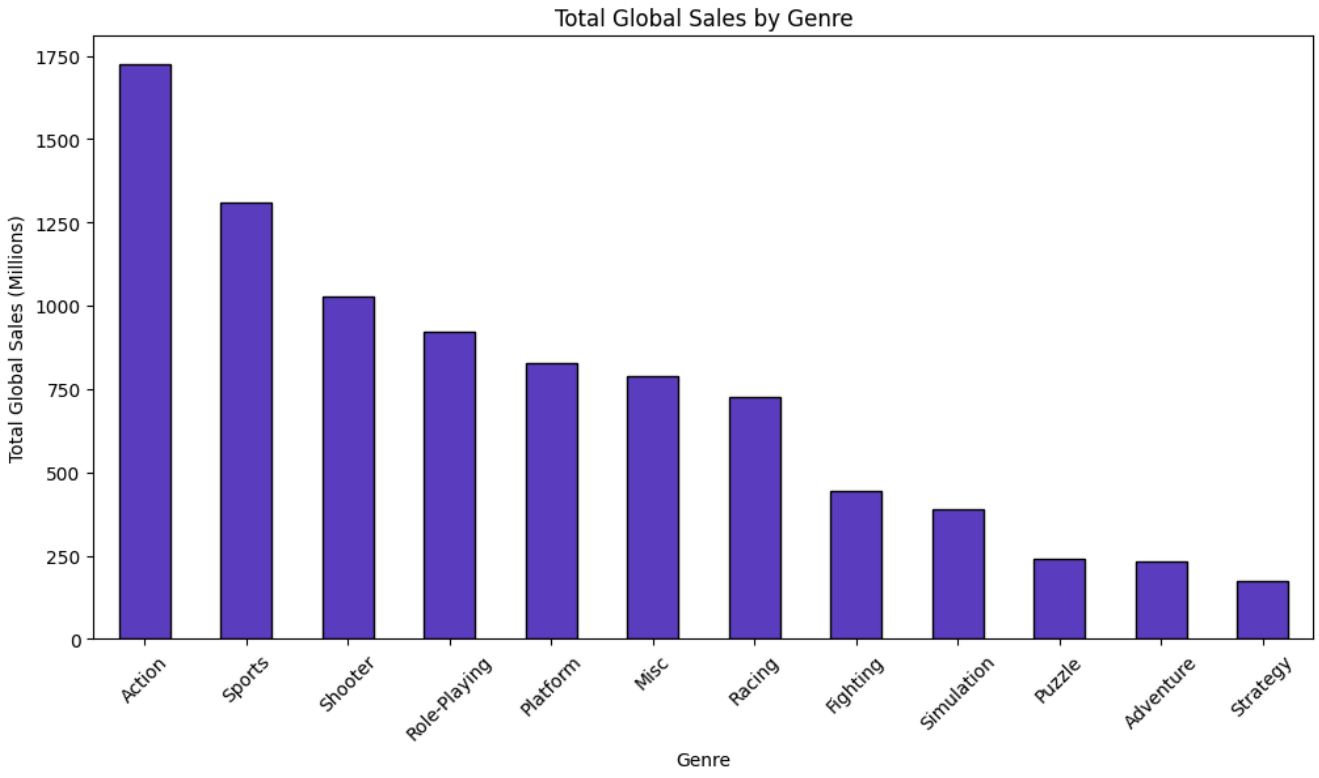
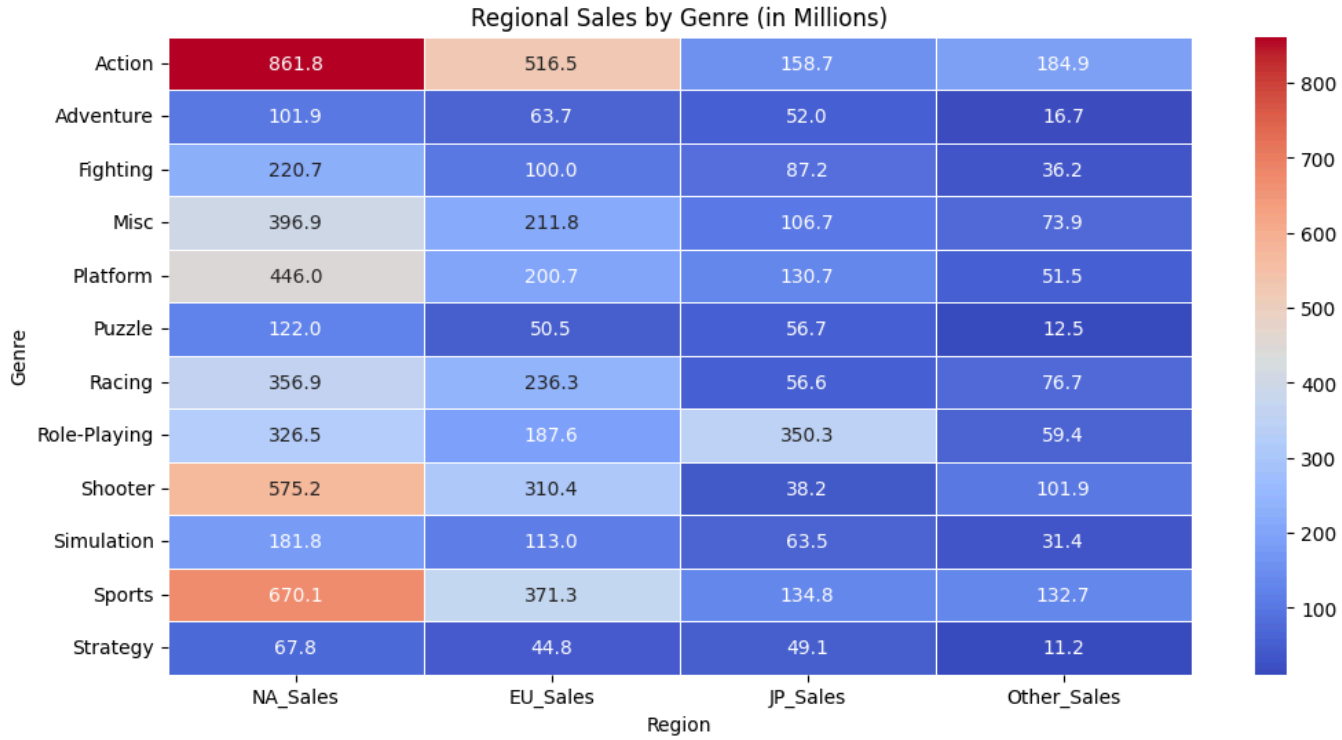


Chart 2: Heatmap that shows how much each region contributes to the total sales for each genre.



3. Platform Success

Objective:

To evaluate the performance of different gaming platforms and their evolution over time.

Findings:

- Platforms such as the **Wii**, **PS2**, **PS3**, and **Xbox 360** dominate global sales, hosting a large number of successful games.
- Platform popularity has shifted significantly over the years, with newer platforms like **PS4** gaining traction.

Visualizations:

Chart 1: Horizontal bar showing total global sales by platform.

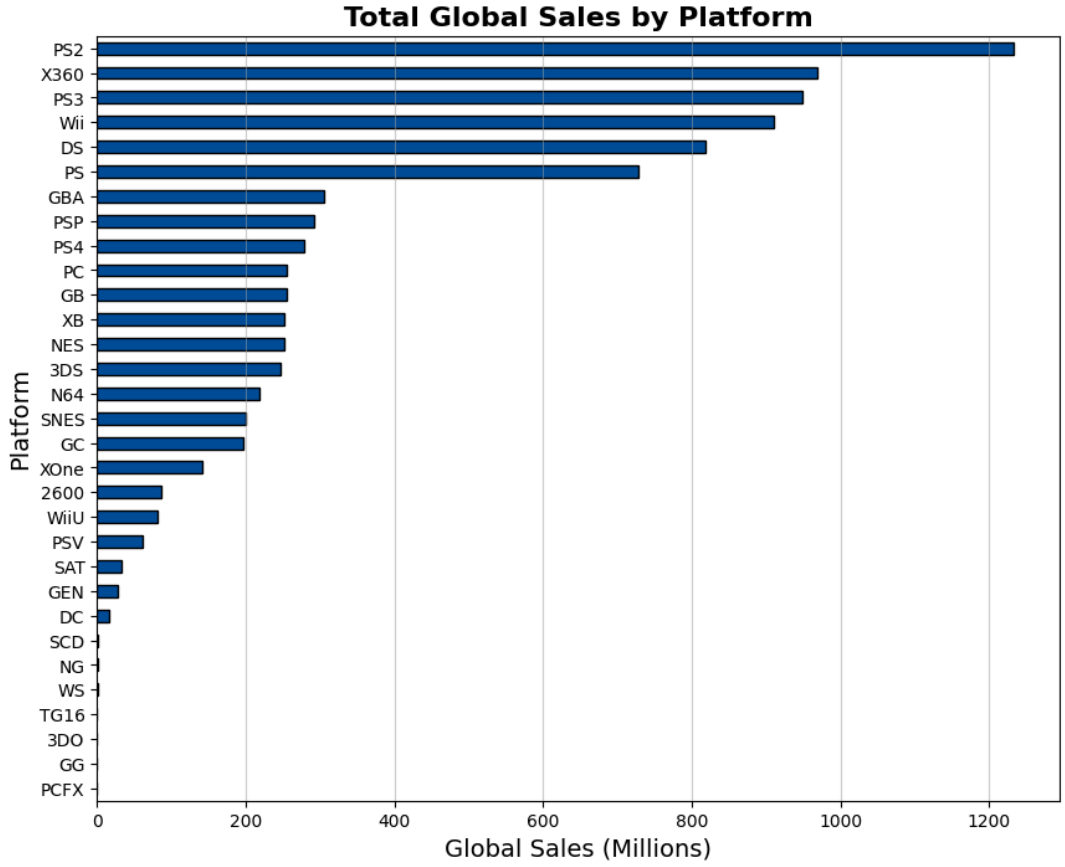
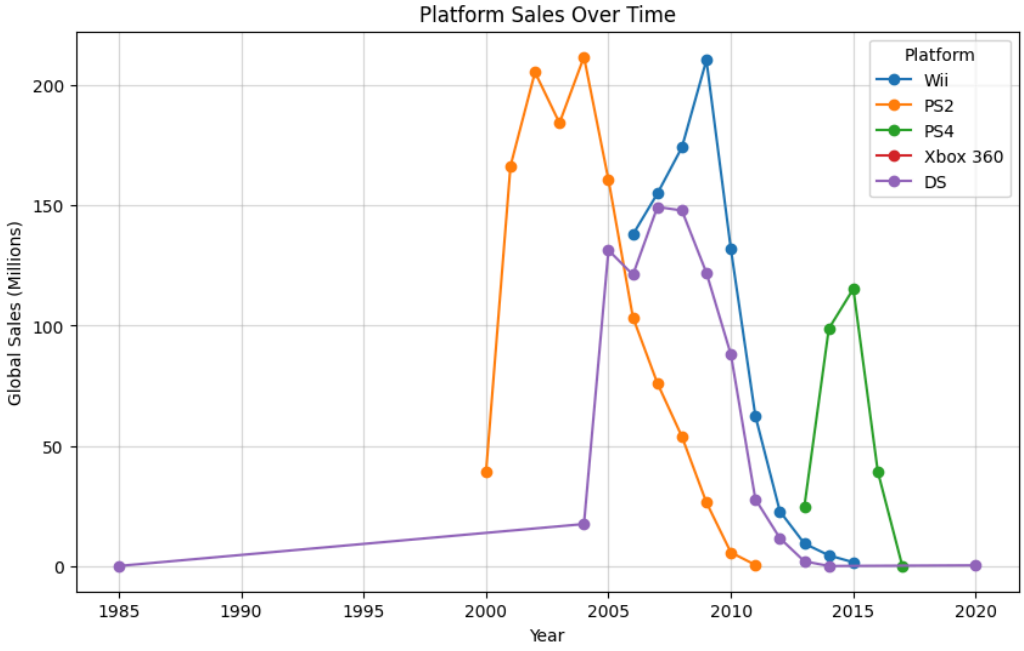


Chart 2: Line chart of platforms over time to reveal the rise & fall of key(Top 5) platforms.



4. Publisher Performance

Objective:

To analyze which publishers consistently perform well and identify strategies contributing to their success.

Findings:

- **Nintendo** leads global sales by a significant margin, followed by **Electronic Arts**.
- Most publishers experienced steady growth over time, while Nintendo stood out with significant bursts in sales during key periods, followed by notable declines.

Visualizations:

Chart 1: Horizontal bar chart showing the top 10 publishers by total global sales.

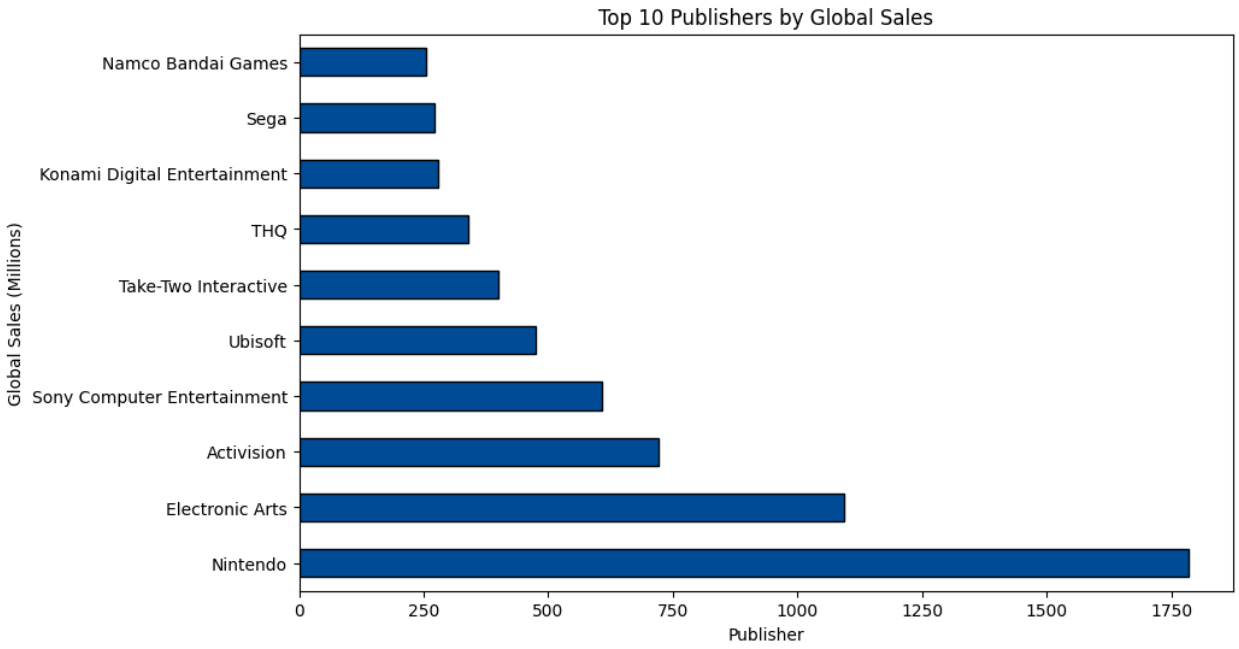
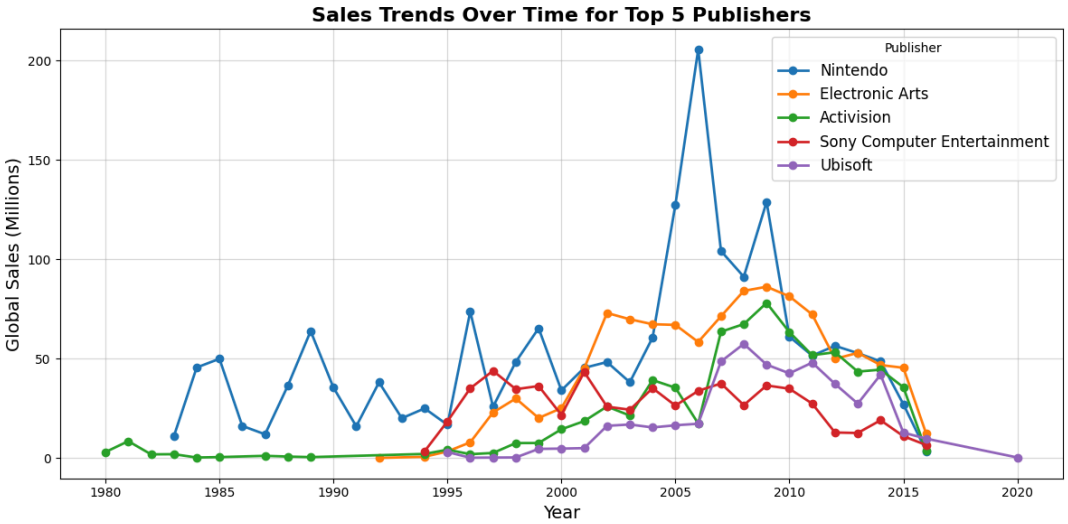


Chart 2: Line chart showing sales trends over time for the top 5 publishers.



5. Correlation Analysis

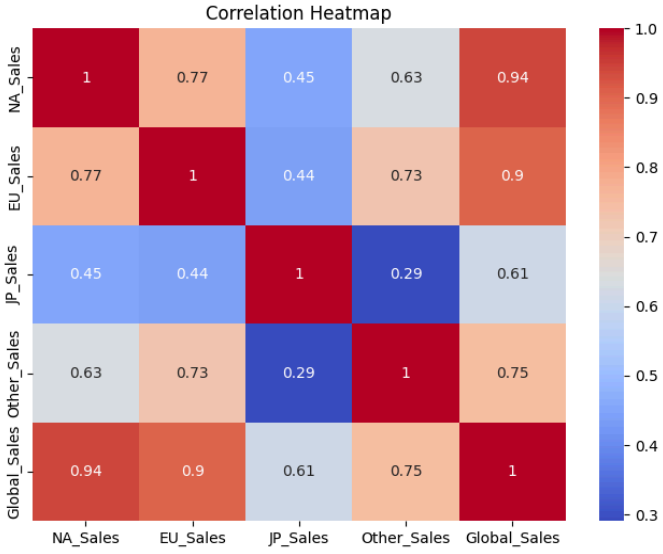
Objective:

To understand the relationships between numerical sales metrics and validate hypotheses.

Findings:

- **North American sales** show the strongest correlation with global sales, followed by European sales.
- Regional sales are moderately correlated, suggesting overlapping but distinct preferences across regions.

Visualizations:



Challenges & Solutions

During the analysis, several challenges emerged that required adjustments in methodology and data processing.

1. Missing Data in Key Columns

- **Problem:** The dataset had missing values in critical columns like **Year** and **Publisher**, which could impact sales trend analysis.
- **Solution:**
    - Games without a **Year & Publisher values** were dropped to maintain consistency in time-series trends.
    - Created several new columns for future predictive modelling using feature extraction techniques.

2. Identifying Meaningful Insights from EDA

- **Problem:** The dataset contained **a lot of numerical and categorical information**, making it challenging to determine which aspects were most relevant.
- **Solution:** Instead of analyzing all features at once, the approach focused on the **most impactful factors**:
    - **Publisher success over time**
    - **Regional sales dominance**
    - **Genre performance in different markets**
    - **Correlation between publisher strategies and sales trends**

Dashboard

An interactive dashboard was developed to allow users to explore sales trends dynamically.

Purpose & Key Features:

- **Filterable Views:** Users can select specific **years, platforms, or regions** to view sales distributions.
- **Genre Comparison:** Displays a **bar chart and donut chart** comparing the most successful game genres in different regions.
- **Region Dominance:** Interactive time-series charts showing **how top publishers' sales evolved over time**.

Conclusion

The analysis provides critical insights into the evolving landscape of the video game industry, highlighting **platform dominance, genre preferences, and publisher performance** over time.

Historically, platforms like the **Wii and PS2** exhibited remarkable success, dominating global sales during their peak years. Their appeal stemmed from hardware innovation, extensive game libraries, and strong brand loyalty. As gaming technology advanced, newer platforms like the **PS4 and Xbox One** gained traction, reflecting the industry's transition towards higher-performance consoles and digital gaming ecosystems.

From a genre perspective, **Sports, Action, and Shooter games** consistently outperformed other categories in global sales. These genres cater to broad demographics, benefiting from franchise-driven releases (e.g., FIFA, Call of Duty, and Grand Theft Auto). Notably, Role-Playing Games (RPGs) excel in Japan, indicating regional market distinctions that publishers leverage for strategic positioning.

In terms of publisher performance, **Nintendo and Electronic Arts (EA)** emerged as industry leaders, with Nintendo exhibiting substantial sales spikes during pivotal years in the mid-2000s. This period saw the success of titles like **Wii Sports, Mario Kart, and The Legend of Zelda**, reinforcing Nintendo's franchise-driven sales model. EA, on the other hand, capitalized on annualized sports releases and multi-platform accessibility, ensuring steady revenue streams.

However, a **noticeable decline in video game sales post-2010** signals an industry shift. Several factors contribute to this trend:

- The rise of **digital distribution platforms** (e.g., Steam, PlayStation Network) reducing physical game sales.
- Increasing popularity of **mobile gaming**, diverting market share from traditional consoles.
- The emergence of **subscription services and cloud gaming** (e.g., Xbox Game Pass, PlayStation Plus), altering consumer spending patterns.

These evolving dynamics indicate that future success in the gaming industry will depend on adaptability to new distribution models, emerging technologies, and shifting consumer behavior. Publishers and developers must continuously innovate, leveraging data-driven insights, player engagement analytics, and evolving monetization strategies to remain competitive in this rapidly changing landscape.

Refer to the following notebook to check further analysis and Python code: EDA ipynb