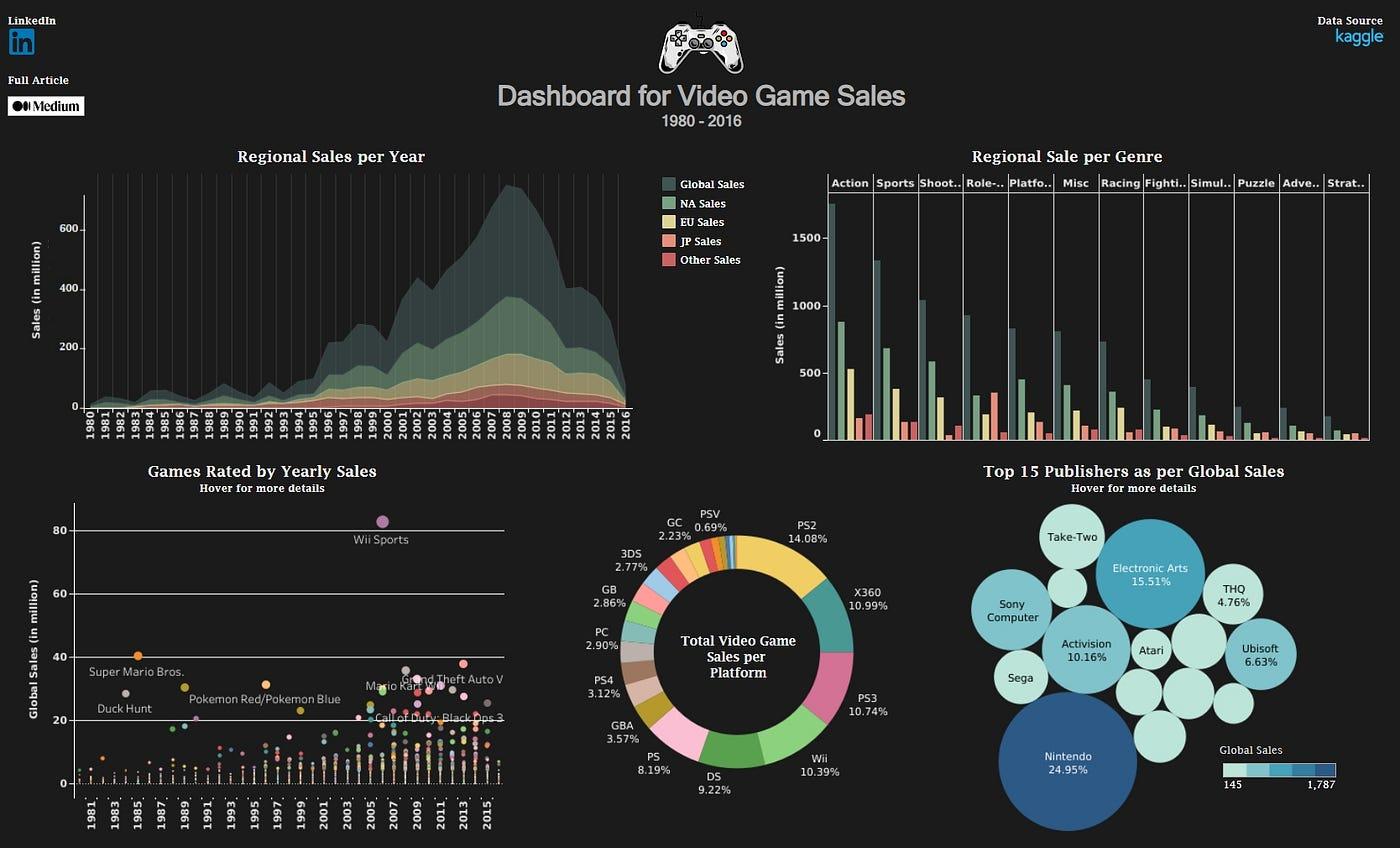
**REPORT ON**

**VIDEO GAME SALES PREDICTION USING DATA ANALYSIS AND VISUALIZATION**



# Abstract

The aim of this analysis is to understand what kind of games have been sold worldwide, trends in popularity and to understand how different regions have their difference in trends and popularity.This report will try to ﬁnd what factors can help determine whether a video game release will be successful in terms of the number of units it sells worldwide, and which of these factors have the strongest impact on sales. This analysis uses a public dataset that gathers data for video games that have sold more than 100,000 copies worldwide from 1976 to 2016, and includes the video game’s development and release information, as well as information on the public reception of the video game.Through data exploration, we will dissect factors influencing video game success. The analysis will consider:

* **Sales Performance:** We'll investigate total sales figures, identifying top-selling games, platforms, and genres across various regions.
* **Genre Preferences:** By analyzing genre popularity over time and across regions, we can uncover player preferences and evolving trends.
* **Platform Dominance:** The project will examine which platforms reign supreme in terms of sales, potentially revealing correlations with genre preferences and target audiences.
* **Regional Variations:** We will explore how video game sales differ geographically, uncovering regional market variations and potential cultural influences.

Data visualization will be a cornerstone of this project. Charts, graphs, and interactive dashboards will be employed to effectively communicate insights. These visualizations will:

* Clearly illustrate sales trends and patterns.
* Facilitate comparisons between genres, platforms, and regions.
* Allow viewers to explore the data interactively, fostering deeper understanding.

# Introduction

The video game industry is a global phenomenon, captivating players of all ages and backgrounds. But have you ever wondered what drives a game's success? What genres dominate the market? How do regional preferences differ? This project embarks on a data-driven exploration to answer these questions and delve deeper into the fascinating world of video game sales.

We'll be wielding the power of data analysis and visualization to uncover hidden trends and illuminate key insights. Prepare to discover:

* **The Blockbusters:** We'll identify the all-time top sellers, the genre giants, and the platforms that reign supreme.
* **Evolving Tastes:** By analyzing trends over time, we'll see how player preferences shift, with genres rising and falling in popularity.

**Global Appeal vs. Local Cravings:** We'll explore how video game sales differ across regions, potentially revealing cultural influences and market variations.

Throughout this project, captivating data visualizations will bring the insights to life. Charts, graphs, and potentially even interactive dashboards will be our tools to:

* Clearly showcase sales trends and patterns.
* Effectively compare genres, platforms, and regional performances.
* Empower you to explore the data interactively, fostering a deeper understanding.

By the end of this project, you'll be equipped with a comprehensive understanding of the forces shaping video game sales. We'll uncover hidden patterns, identify key demographics, and illuminate the trends that drive this dynamic and ever-changing industry.

# Problem Statement

To find if the newly developed game would be a success or not based on the trend of genres over the years, life span of platform and genre popularity in a region.

# Methodologies Used

## Bar plot

A barplot (or barchart) is one of the most common types of graphics. It shows the relationship between a numeric and a categorical variable. Each entity of the categoric variable is represented as a bar. The size of the bar represents its numeric value.

## Pairplot

A pair plot, also known as a scatterplot matrix, is a matrix of graphs that enables the visualization of the relationship between each pair of variables in a dataset. It combines both histogram and scatter plots, providing a unique overview of the dataset's distributions and correlations.

## Pie chart

A pie chart helps organize and show data as a percentage of a whole. True to the name, this kind of visualization uses a circle to represent the whole, and slices of that circle, or “pie”, to represent the specific categories that compose the whole. This type of chart helps the user compare the relationship between different dimensions within a specific context.

## Heatmap

Heatmaps are used to show relationships between two variables, one plotted on each axis. By observing how cell colors change across each axis, you can observe if there are any patterns in value for one or both variables.

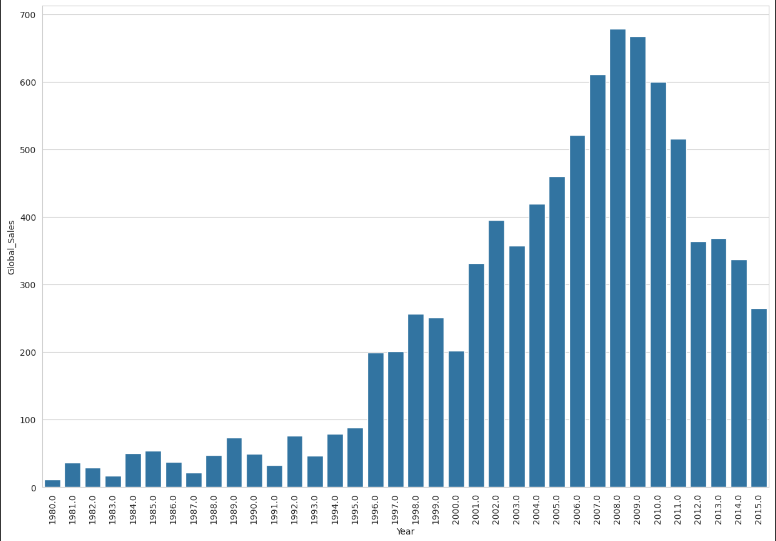
## Distplot

It is used basically for a univariant set of observations and visualizes it through a histogram i.e. only one observation and hence we choose one particular column of the dataset.

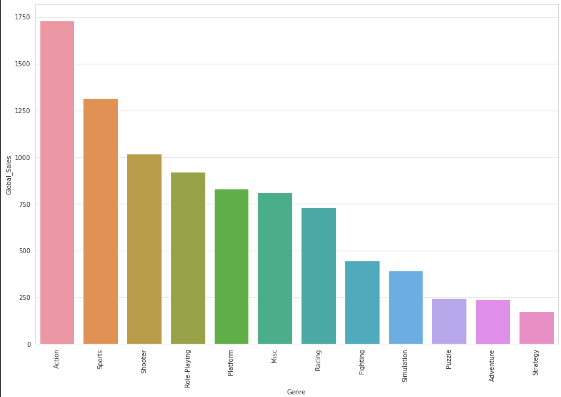
## Countplot

## The countplot is used to represent the occurrence(counts) of the observation present in the categorical variable.It uses the concept of a bar chart for the visual depiction.

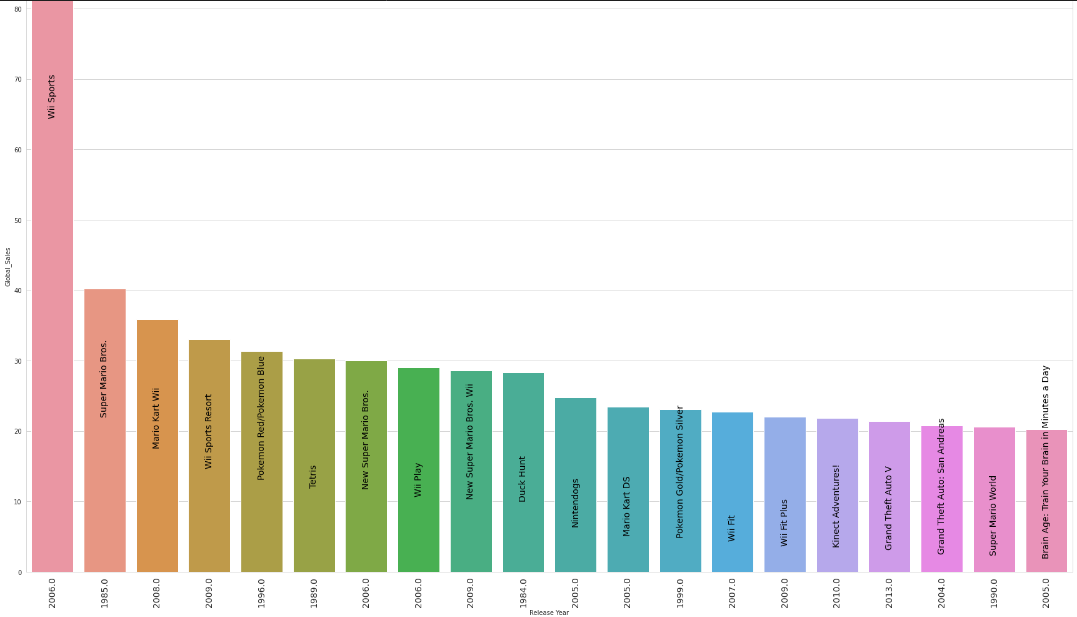
# Results and Insights



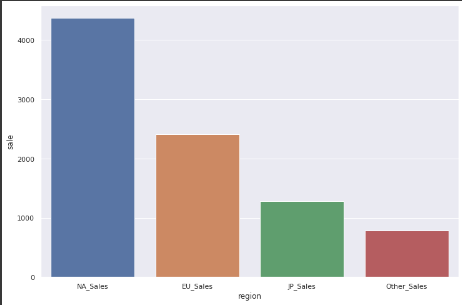
The count plot displays the number of game releases per year, revealing peaks in certain years, indicating periods of high game production, and a general trend over time in the gaming industry.



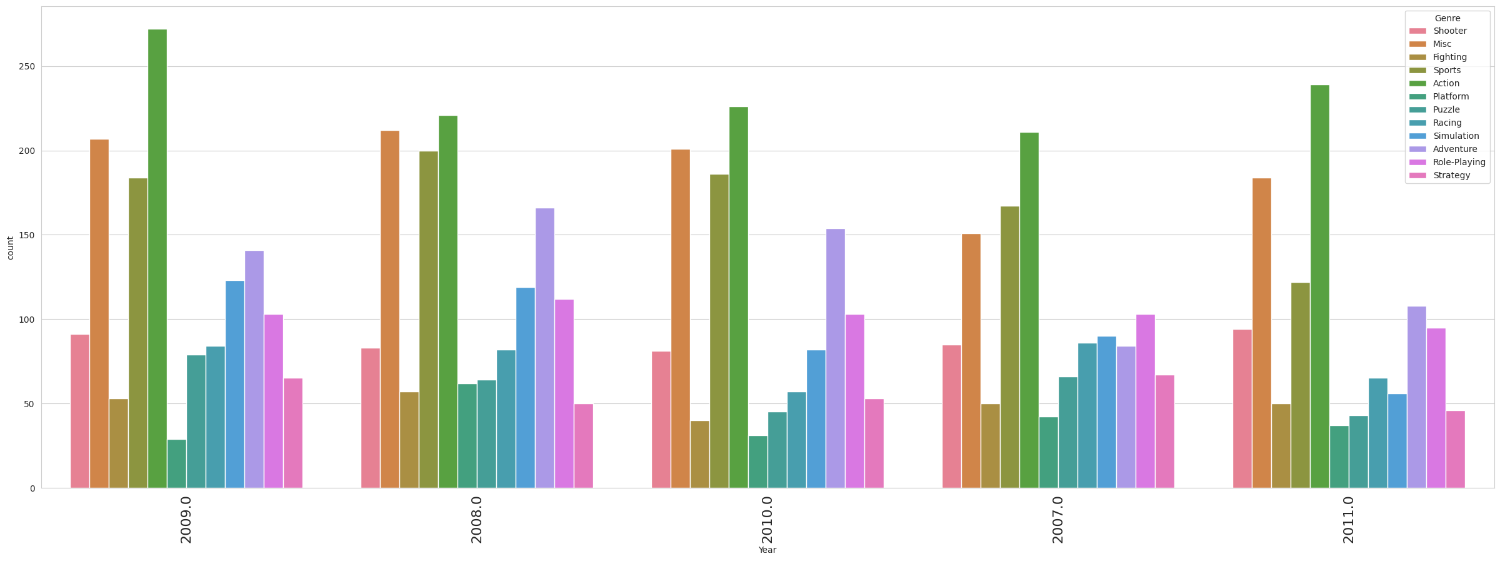
This bar plot provides a clear and concise visual summary of the global sales performance of various video game genres, enabling stakeholders to make informed decisions based on genre popularity and market demand



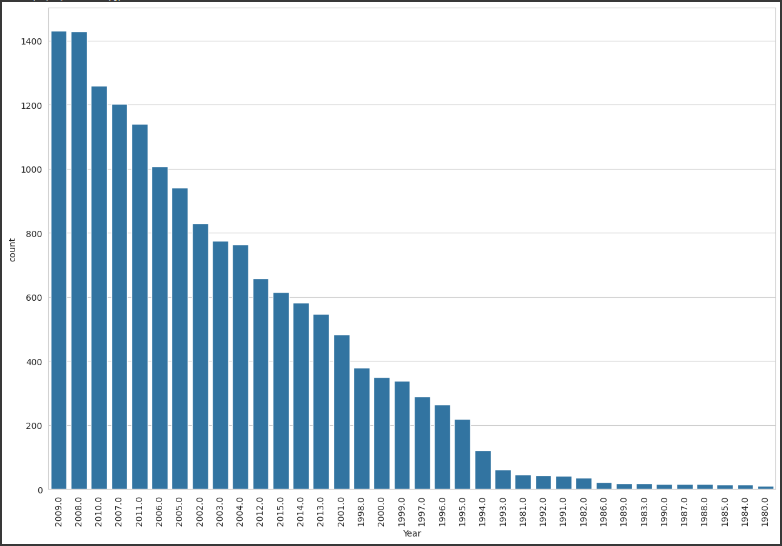
This visualization is useful for understanding the performance of top-selling games in terms of global sales and their release timeline.



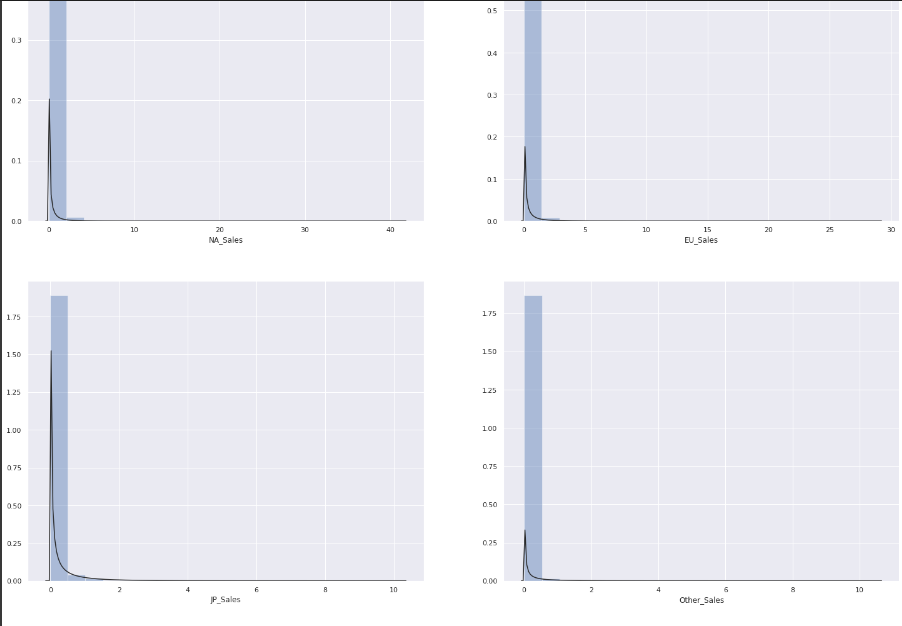
The bar plot shows the distribution of total sales by region, highlighting the highest sales in North America, followed by Europe, Japan, and other regions



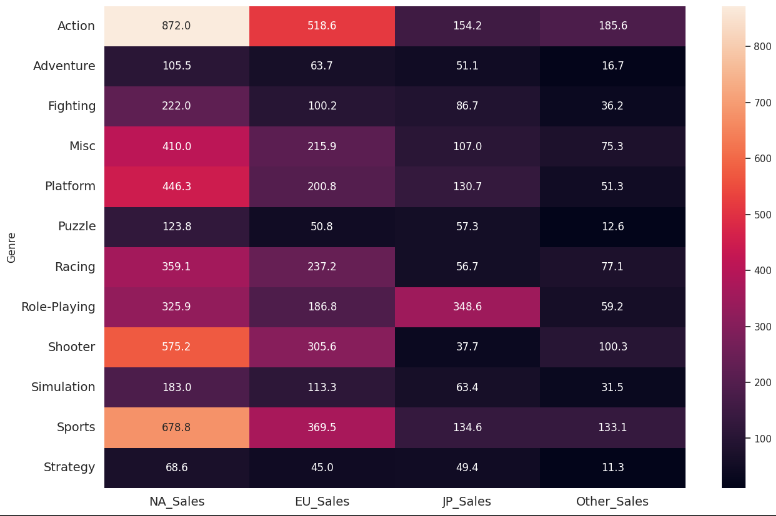
The count plot illustrates the distribution of game releases by genre over the top 5 most frequent years, showing trends and genre popularity across these years.



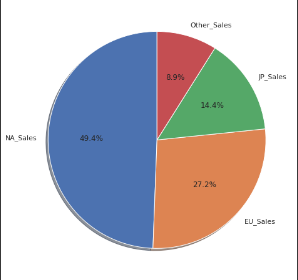
The count plot illustrates the yearly distribution of game releases, showing fluctuations and trends over time, with some years having notably higher numbers of game releases



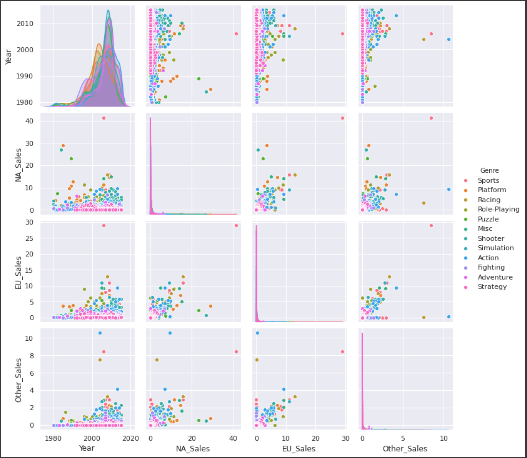
The distribution plots indicate that sales in each region and globally follow a skewed distribution, with most games having low sales and a few games achieving very high sales.



The heatmap reveals significant regional sales variations across game genres. Notably, North America and Europe exhibit higher sales for most genres, while Japan shows distinct preferences with comparatively lower sales in other regions.



The pie chart illustrates the sales distribution across regions, showing North America as the largest market, followed by Europe, Japan, and other regions, each contributing distinct proportions to global sales.



The pair plot reveals relationships between sales in different regions, highlighting genre-specific trends. Notably, some genres show clustered patterns, indicating distinct sales behaviors across years, platforms, and regions.

# Conclusion

Here are the top 3 observations made from our analysis:

1. **North America** is the **biggest market** globally for video games, making up **49.24%** of the total sales**.**
2. **Action** games are the **most popular** followed sports and shooters. **Wii sports**, the **highest selling game** of the given data was a sports game with total sales of 82.74 million.
3. **Nintendo** was the **most popular publisher**, making **24.95%** of global sales with **1787 million** in sales.