# REPORT: TRENDS IN VIDEO GAME SALES

**By: Amélie MEYER**

*Date last edit : 02/05/2022*

## INTRODUCTION:

This report is going to give a **highly detailed explanation** of the step-by-step process of my analysis.
In this project, only an exploratory analysis will be done, with the objective to understand how the data is distributed and generate insight for future data-driven decision-making.
Moreover, this analysis aims to explore as much as possible the data in a simple and informative way.

*Note:* My analysis will be completed in **Microsoft Excel, Microsoft SQL Server Management Studio 18** and **Tableau Public.**

## PHASE 1: ASK

The purpose of this analysis is to discover the trends in the video games industry from **1981 to 2017**, observing **games**, their **genre**, their **publisher**, their **platform** and **year** of release and so determine the most successful companies, games, gaming genres, and platforms.

## PHASE 2: PREPARE

In this analysis, the public data set "Video Game Sales" ([Video Game Sales | Kaggle](#)) will be used.
The dataset contains around 40 years of data regarding more than 16500 games,their genre, their publisher, their platform and year of release.

- **Bias and Credibility**

The **ROCCC system can be used in order to determine** the credibility and extent to which the 'Video Game Sales' data is biassed.

- <u>Reliable:</u> This dataset **is not** reliable as no information is given on the source of the dataset.

- <u>Original:</u> This dataset **is not** original as no information is given on the source of the dataset.

- <u>Comprehensive:</u> This dataset **is** comprehensive as each game is linked to their genre, their publisher, their platform and year of release; we have enough information to understand the data.

- <u>Current:</u> The dataset **is not** current, the last year reported is 2020.

- <u>Cited:</u> This data set **is not** cited.

The **data integrity and credibility** is insufficient to make the insight of this analysis reliable, as the data might be wrong or biassed. Nevertheless, the amount of data, and the fact that it is correctly formatted and that around 800 people submitted an analysis of this data, I made the decision to still use it.

## PHASE 3: PROCESS

Before analysing the data, the data's integrity has to be checked and the data has to be cleaned. To do so, the data will be cleaned in **Microsoft Excel.**

**-        Summary of the step taken:**

-Formatting the cells to the right data type.

-Removing duplicates: using the action *remove duplicate.* No duplicates were found.

-Treating blank and  N/A values: No blank values were found, but 271 N/A values were found in the *year* column (using the =COUNTIF function). The video games with N/A in the year column representing 1.6% of the overall database the decision to keep those entries for the analysis was made.

The data was already clean and ready for analysis, but it is always important to check before proceeding to the analysis.

Once on **Microsoft SQL Server** the column 'Rank' was removed and the name column 'Name' changed to 'Game' using this code:

```
--Deleting rank column

ALTER TABLE vg_sales

DROP COLUMN Rank;



--Renaming Name column

EXEC sp_RENAME 'vg_sales.Name', 'Game';
```

## PHASE 4: ANALYSE

## -      Sales share on the market

```
--Sales share on the markets

SELECT SUM(NA_Sales)/SUM(Global_Sales)*100 AS Per_Sales_NA,

       SUM(EU_Sales)/SUM(Global_Sales)*100 AS Per_Sales_EU,

       SUM(JP_Sales)/SUM(Global_Sales)*100 AS Per_Sales_JP,

       SUM(Other_Sales)/SUM(Global_Sales)*100 AS Per_Sales_Other

FROM vg_sales
```

### Sales share on the market

| Region | North America | Europe | Japan | Other Region |
|---|---|---|---|---|
| Percentage of market | 49.25 | 27.29 | 14.47 | 8.94 |

## Observations:

- The market in North America represents almost 50% of the global sales.
- The market in Japan alone represents 14% of the global sales.


## -      Games sales worldwide over the years

```
-- Game sales worldwide over the years

SELECT year, SUM(Global_Sales) AS Sales

FROM vg_sales

WHERE Year IS NOT NULL

GROUP BY Year

ORDER BY Sales DESC;
```

Number of sales worldwide over the years
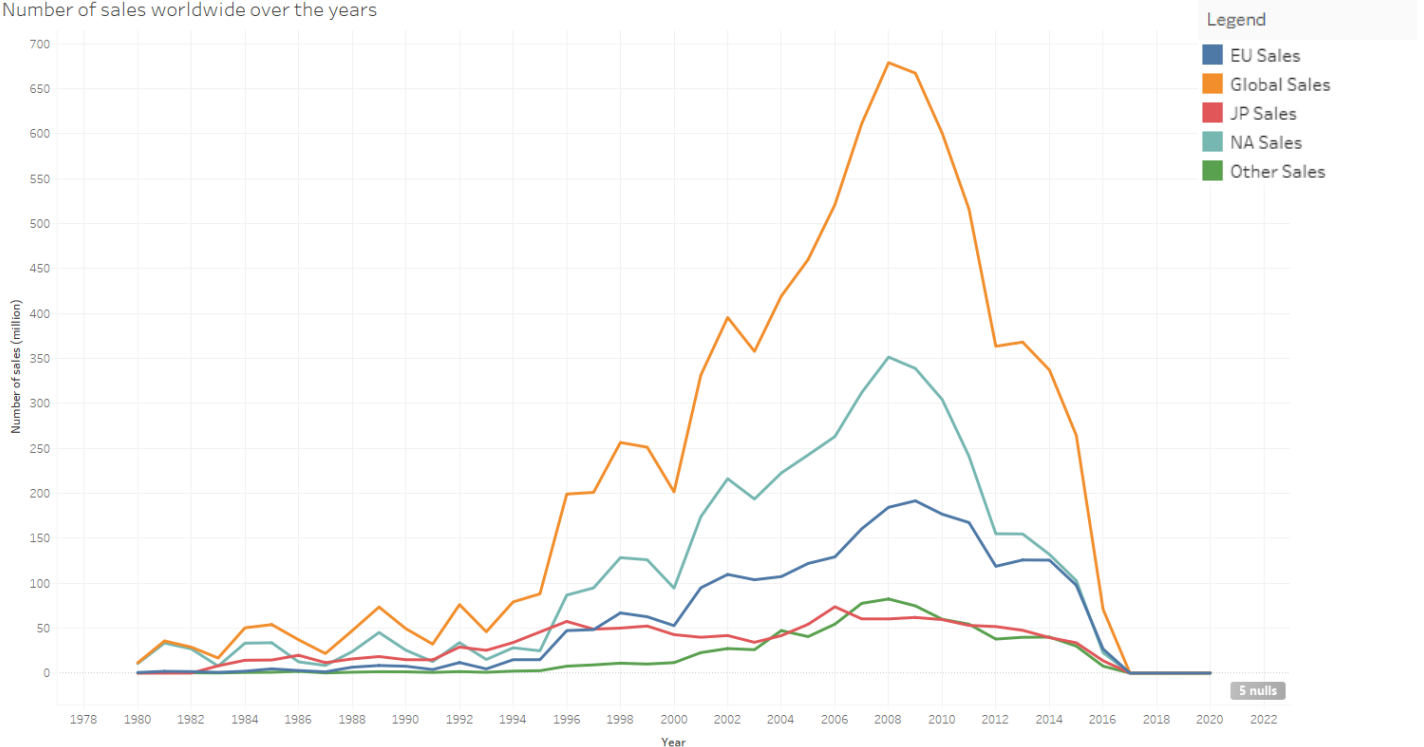
## Observations:

- The years 2006, 2007, 2008, 2010 were the years where the most games were sold.
- The number of sales globally and for each market has been increasing constantly, having a huge spike starting in 2000, until 2010 where the numbers are decreasing rapidly.
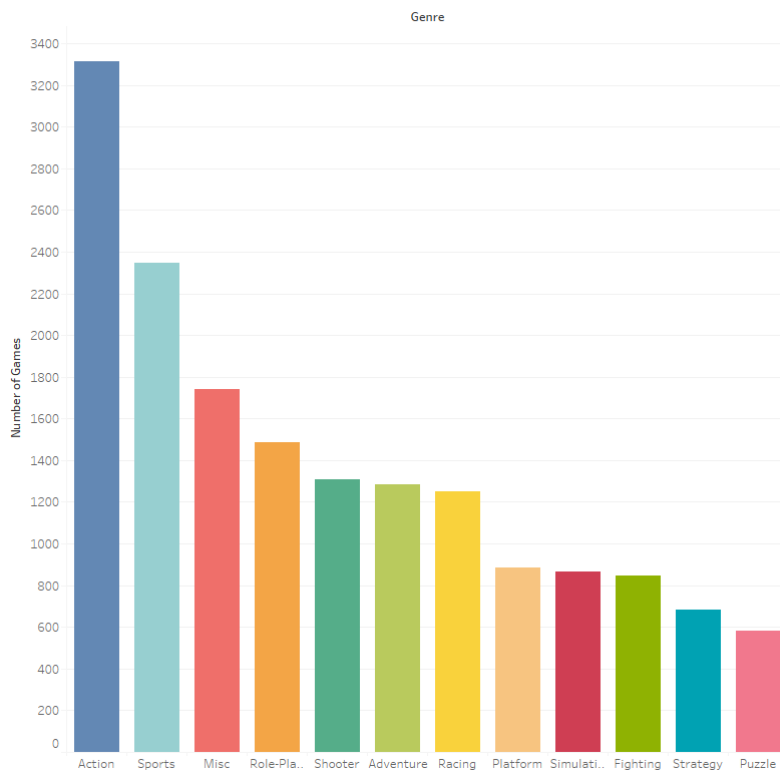
- **Analysis of the 'Genre' Category**

```
--Analysis of Genre category

    --Number of games released for each genre

SELECT Genre, COUNT('Genre') AS Number_Games

FROM vg_sales

GROUP BY Genre

ORDER BY Genre;
```

Number of games realsed by genre



```
    --Number of sales by genre

WITH Total_GB AS

    (SELECT SUM(Global_Sales) AS Total_GB

    FROM vg_sales)

SELECT Genre, AVG(Global_Sales)*100 AS Avg_Sales_Thousand, SUM(Global_Sales)/Total_GB*100 AS Percentage

FROM vg_sales, Total_GB

GROUP BY Genre, Total_GB

ORDER BY Avg_Sales_Thousand DESC;

--ORDER BY Percentage DESC;
```

Representation of the average number of sales for each genre in thousand
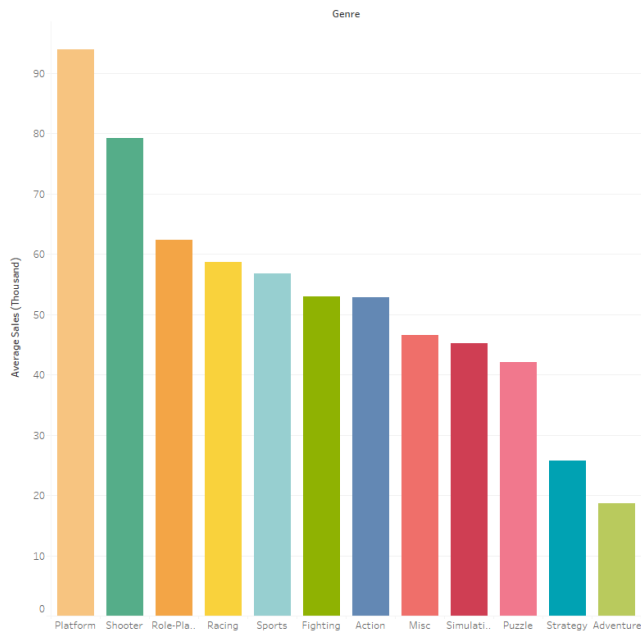


Percentage of the sales of each genre in the video games market
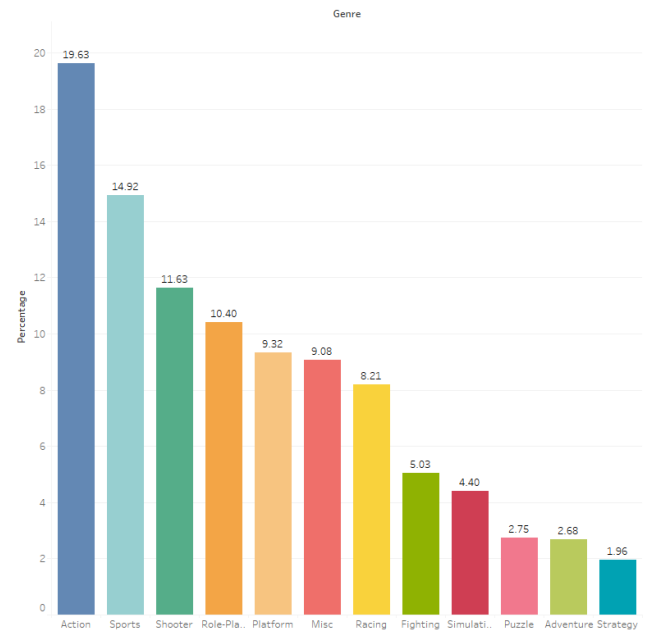
## Observations:

- The genres: "Action", "Misc", and "Role-Playing" are the genres of most games.
- The genres: "Platform", "Shooter" and "Role-Playing" are the genres that on average sell for most money
- The genres: "Action", "Sports" and "Shooter" together represent around 35% of the video game's sales market.
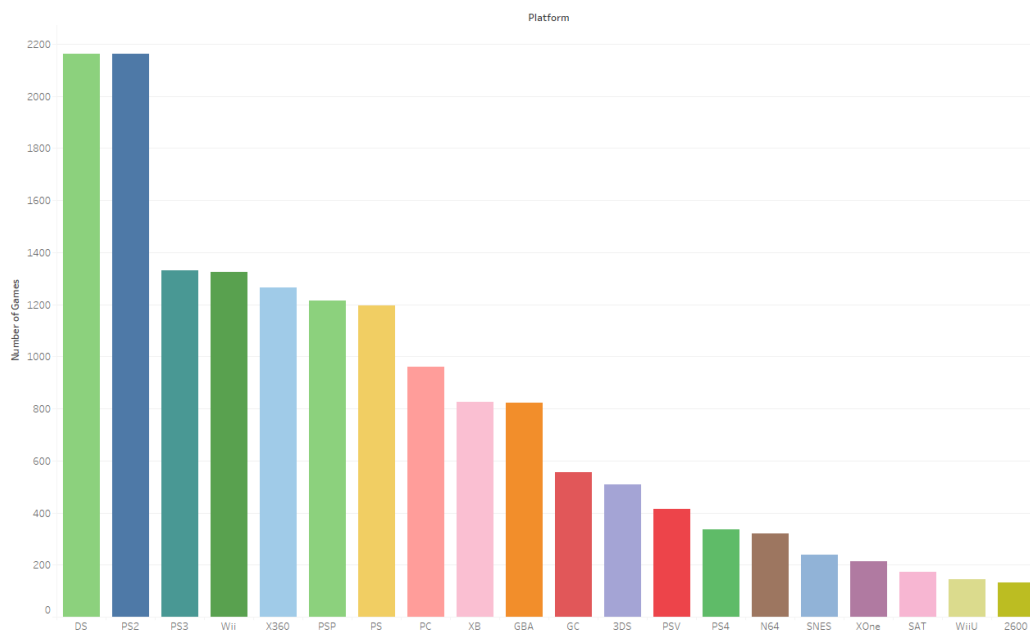
## -       Analysis of the 'Platform' Category

```
--Analysis of Platform category

    --Number of games released on each platform

SELECT Platform, COUNT(Platform) AS Game_Platform

FROM vg_sales

GROUP BY Platform

ORDER BY Game_Platform DESC;
```

Number of games realsed by platform

```
--Number of sales per platform

WITH Total_Sales AS

        (SELECT SUM(Global_Sales) AS Total_Sales

        FROM vg_sales)

SELECT Platform,  AVG(Global_Sales)*100 AS Avg_Sales_Thousand, SUM(Global_Sales) / Total_Sales *100 AS
Percentage

FROM vg_sales, Total_Sales

GROUP BY Platform, Total_Sales

ORDER BY Percentage DESC;

--ORDER BY Avg_Sales_Thousand DESC;
```
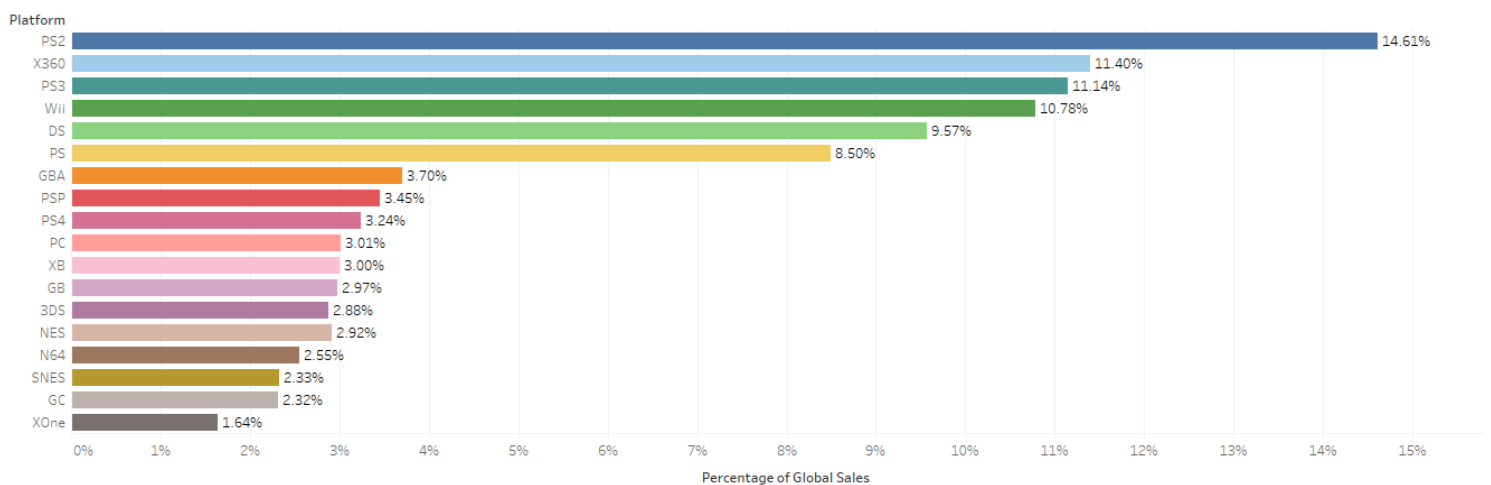
Percentage of global sales for each platform



## Observations:

- The three platforms which released the most games were: "DS", "PS2" and "PS3"
- The platforms "PS2", "X360", and "PS3" sold the most games.

## -        Analysis of the 'Game' Category

```
--Number of games in the dataset

SELECT COUNT(Game)

FROM vg_sales
```

| NUMBER OF GAMES IN THE DATASET | 16598 |
| --- | --- |

```
    --Number of sales percentage

WITH Total_Sales AS

        (SELECT SUM(Global_Sales) AS Total_Sales

        FROM vg_sales)

SELECT DISTINCT Game, Publisher, Global_Sales, SUM(Global_Sales) / Total_Sales *100 AS Percentage
```
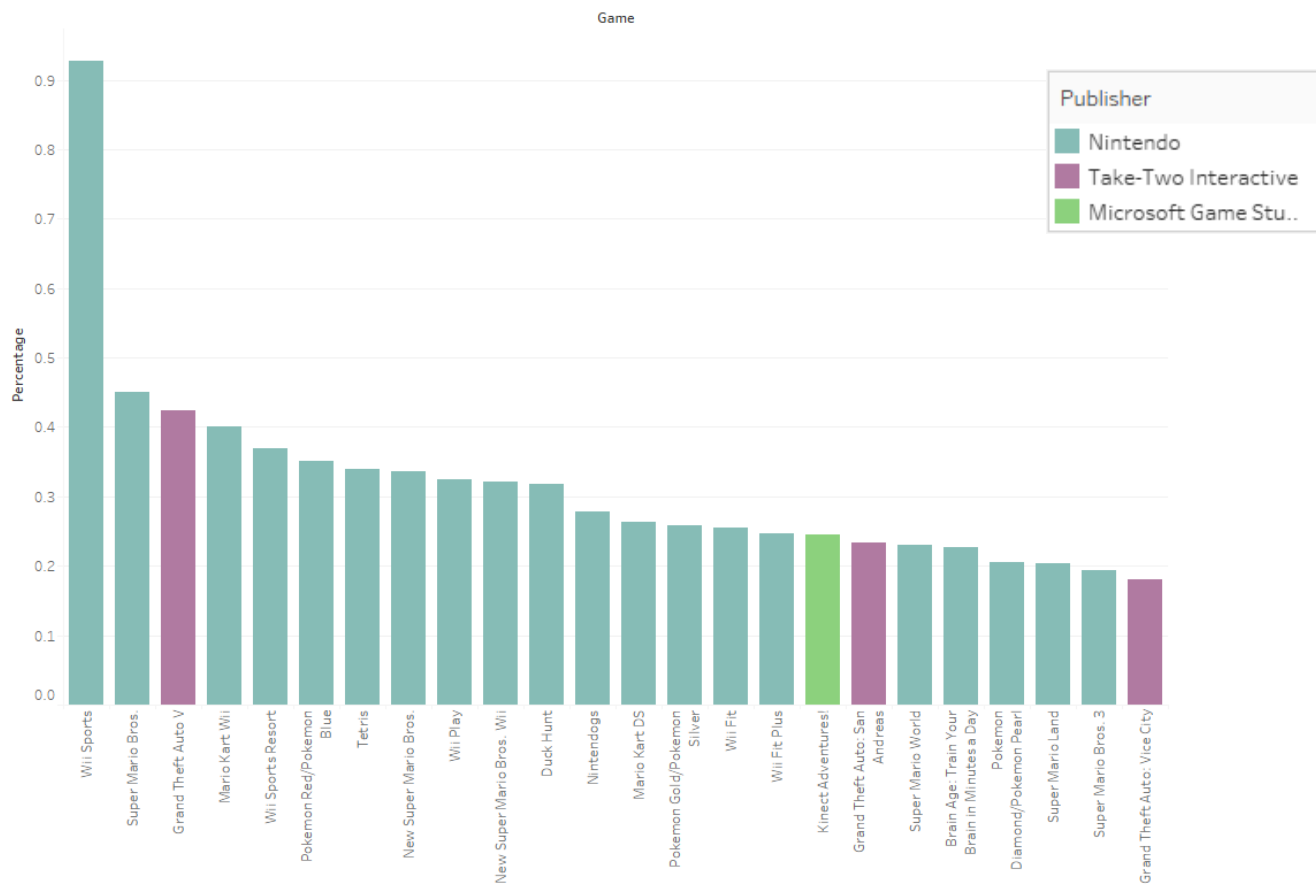
```
FROM vg_sales, Total_Sales

GROUP BY Game, Total_Sales, Global_Sales, Publisher

ORDER BY Percentage DESC;
```

Percentage of games sold globaly organized by publisher



**Observations:**

- There are 16598 games in the dataset.
- Out of the top 24 most sold games globally, the publisher 'Nintendo' created 20, representing 83%.
- The game that sold the best globally is 'Wii Sports', representing almost 1% of the total of games ever sold, in other word, one out of 100 games ever sold is 'Wii Sports'.


- **Analysis of the 'Publisher' Category**

```
    --Number of game released by each publisher

SELECT Publisher, COUNT(Publisher) AS Game_Publisher

FROM vg_sales

GROUP BY Publisher

ORDER BY Game_Publisher  DESC;
```
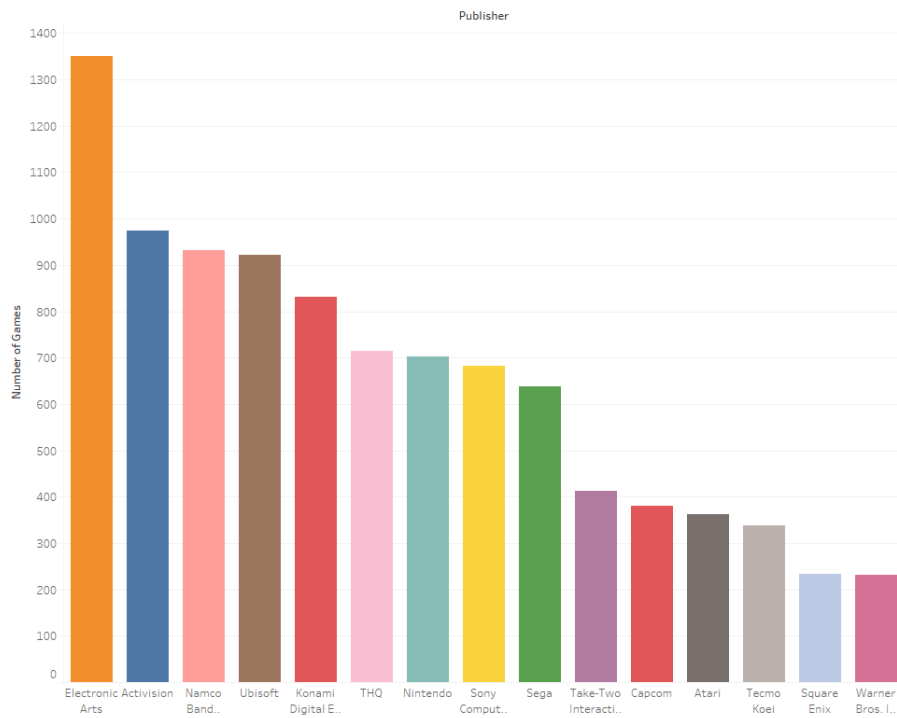
## Number of games realsed by publisher



--Number of sales percentage

```
WITH Total_Sales AS

    (SELECT SUM(Global_Sales) AS Total_Sales

    FROM vg_sales)

SELECT Publisher, AVG(Global_Sales) AS Avg_Sales, SUM(Global_Sales) / Total_Sales *100 AS Percentage

FROM vg_sales, Total_Sales

GROUP BY Publisher, Total_Sales

ORDER BY Percentage DESC;
```
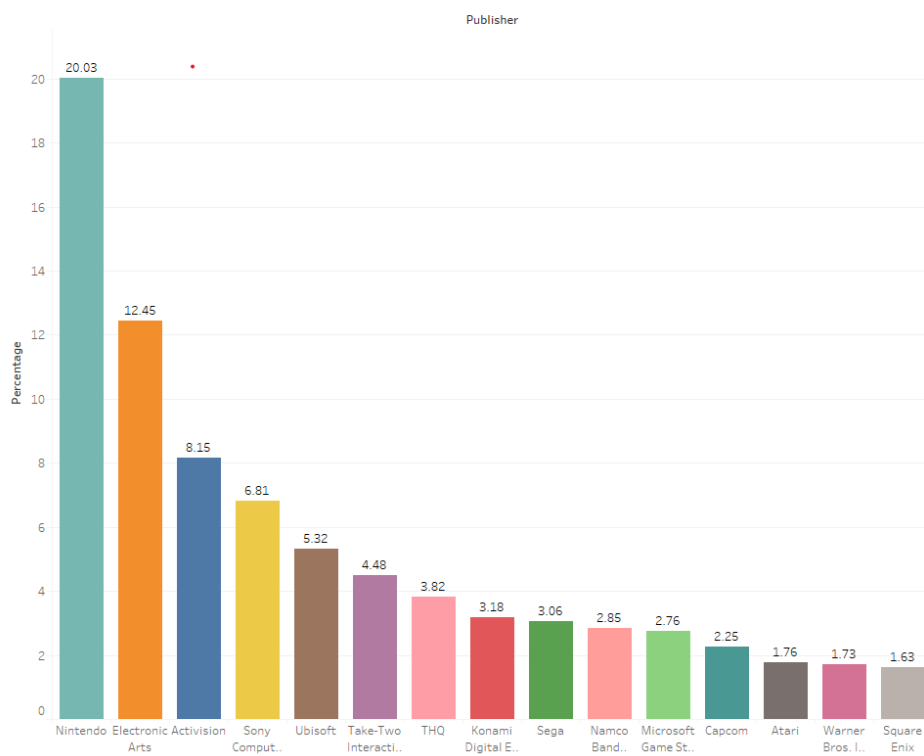
Number of games realsed by publisher
## Percentage of global sales for each publisher

## Observations:

●      'Electronic Art's', 'Activision' and 'Namco Bandai Games' are the publishers that released the highest number of games.

●      'Nintendo', 'Electronic Arts' and 'Activision' are the publishers that sold the most games.

●      'Nintendo' and 'Electronic Arts' sales together represent more than 30% of the whole video games market.

-       **Analysis of Nintendo and Electronic Arts' success over time**

```
--Nintendo's Success

SELECT Game, Year, Global_Sales

FROM vg_sales

WHERE Publisher = 'Nintendo' AND Year IS NOT NULL

GROUP BY Year, Game, Global_Sales

ORDER BY Global_Sales DESC, Game, Year;



--Electronic Arts' Success

SELECT Game, Year, Global_Sales

FROM vg_sales

WHERE Publisher = 'Electronic Arts' AND Year IS NOT NULL

GROUP BY Year, Game, Global_Sales

ORDER BY Global_Sales DESC, Year DESC, Game
```

Evolution of Nintendo and Electronic Arts' sales over time
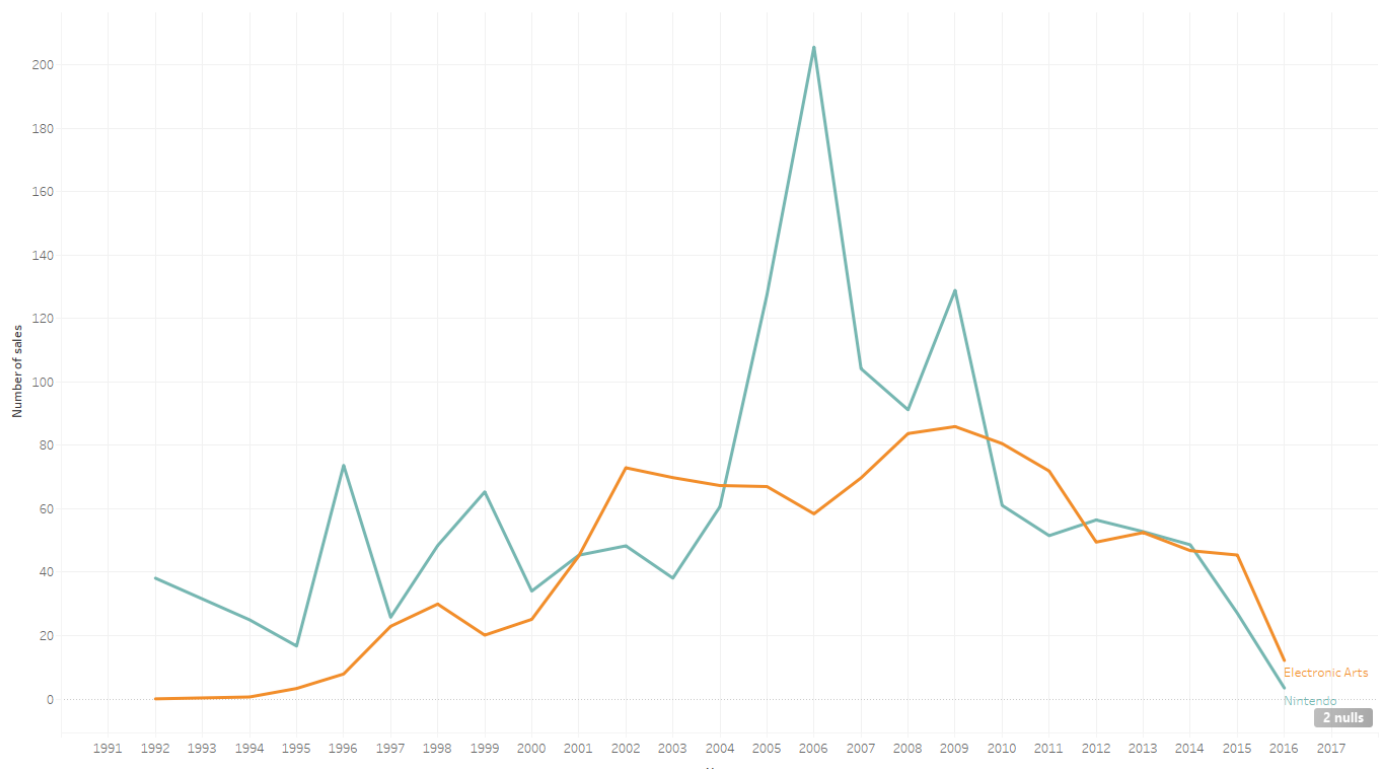
# CONCLUSION OF MY ANALYSIS

It was a very interesting exercise to work with SQL and Tableau in a more collaborative way. This project helped me deepen my skills on both software and enabled me to discover trends in the video games industry. I always knew Nintendo held a big part in the market, but I wouldn't have guessed that it is such a massive part of it.

*Note:* The dashboards and interactive visualisations can be found here: [Dashboards & Visualisations | Tableau Public](#)