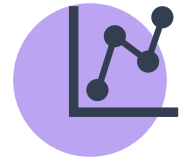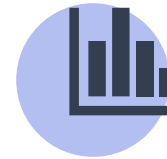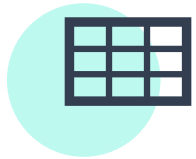# VIDEO GAME SALES ANALYZATION

*KAGGLE GAMING SALES PROJECT*

*CHRISTINA LU JIN | November 2020*

# PROJECT OUTLINE

## Dataset Overview

Source of Data

Data Brief Digestion

Entities and Attributes

## Dataset Cleaning

Data Loading

Libraries Loading

Data Cleaning in R

## Descriptive Analysis

Global Sales by Game

Game Sales by Genre

Game Sales per Platform

Statistical Information

Business Insights

## Exploration Analysis

World's Best-selling Games

Global Sales by Region

Global Sales per Year

Inferential Analysis

# RESOURCES OF DATA



The video game sales dataset

file (in CSV format) was

retrieved from online open

data platforms:

- Kaggle

- VGChartz

# DATA OVERVIEW

11 COLUMNS | 16600 RECORDS

A 5 STRINGS    0.0 5 DECIMALS    # 1 INTEGER

**FIELDS INCLLUDE :**

- Name - The games name

- Platform - Platform of the games release (PC,PS4, etc.)

- Year - Year of the game's release

- Genre - Genre of the game

- Publisher - Publisher of the game

- NA_Sales - Sales in North America (in millions)

- EU_Sales - Sales in Europe (in millions)

- JP_Sales - Sales in Japan (in millions)

- Other_Sales - Sales in the rest of the world (in millions)

- Global_Sales - Total worldwide sales.

# DATA LOADING AND CLEANING

❑ 2 records were omitted because of incomplete information, since ranking goes up to 16600 but there are only 16598 records.

| | Rank | Name | Platform | Year | Genre | Publisher | NA_Sales (M) | EU_Sales (M) | JP_Sales (M) | Other_Sales (M) | Global_Sales (M) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **16598** | 16600 | Spirits & Spells | GBA | 2003 | Platform | Wanadoo | 0.01 | 0.00 | 0.00 | 0 | 0.01 |
| **16597** | 16599 | Know How 2 | DS | 2010 | Puzzle | 7G//AMES | 0.00 | 0.01 | 0.00 | 0 | 0.01 |
| **16596** | 16598 | SCORE International Baja 1000: The Official Game | PS2 | 2008 | Racing | Activision | 0.00 | 0.00 | 0.00 | 0 | 0.01 |

❑ In order to clean up the dataset, as shown on the right, I have omitted all the records that have invalid values in columns 'Year' and 'Publisher'. This helped removing a total of 307 records with incomplete information.

**INITIAL DATASET**

```
> summary(vgfull)
     Rank            Name              Platform             Year
 Min.   :     1   Length:16598       Length:16598       Length:16598
 1st Qu.: 4151    Class :character   Class :character   Class :character
 Median : 8300    Mode  :character   Mode  :character   Mode  :character
 Mean   : 8301
 3rd Qu.:12450
 Max.   :16600
```

**DATA CLEANING**

```
# Data cleaning: filter records with Year and Publisher = N/A
vgsales <- vgfull[vgfull$Year != "N/A" & vgfull$Publisher !="N/A", ]
summary(vgsales)
```

**CLEAN DATASET**

```
> summary(vgsales)
     Rank            Name              Platform             Year
 Min.   :     1   Length:16291       Length:16291       Length:16291
 1st Qu.: 4132    Class :character   Class :character   Class :character
 Median : 8292    Mode  :character   Mode  :character   Mode  :character
 Mean   : 8290
 3rd Qu.:12440
 Max.   :16600
```

# DESCRIPTIVE ANALYSIS | GLOBAL GAMING SALES

## FINDINGS ON GAMING SALES OF THE GLOBAL MARKET

By looking at the descriptive data to the right, it is showing a very high density towards the lower end of the sales with 3rd quartile being not even half of a million sales, and most of the records (mode) are as low as 0.02 million. But there are some extremely high value, with the maximum sales of the single game being almost 83 million, Which is why the mean value has been brought to a little more than half of the million.

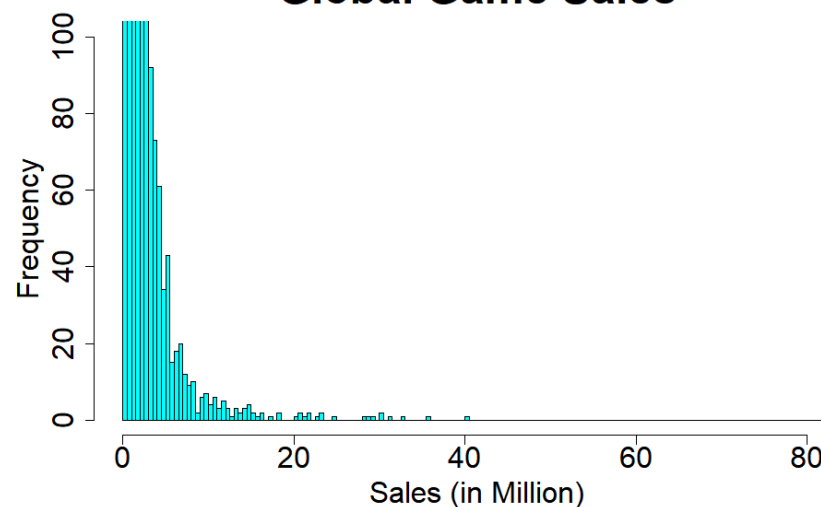| | Global |
|---|---|
| Min. | 0.0100000 |
| Mode | 0.0200000 |
| Q1 | 0.0600000 |
| Median | 0.1700000 |
| Mean | 0.5409103 |
| Q3 | 0.4800000 |
| Max. | 82.7400000 |

Similar information are showing on the histograms as well, where global sales are extremely right skewed and most of the data are clustered around 1 million or less.

These descriptive data conveys that there are very little key players in the gaming market, and it is these players who are driving the entire market and almost like dominating the industry.
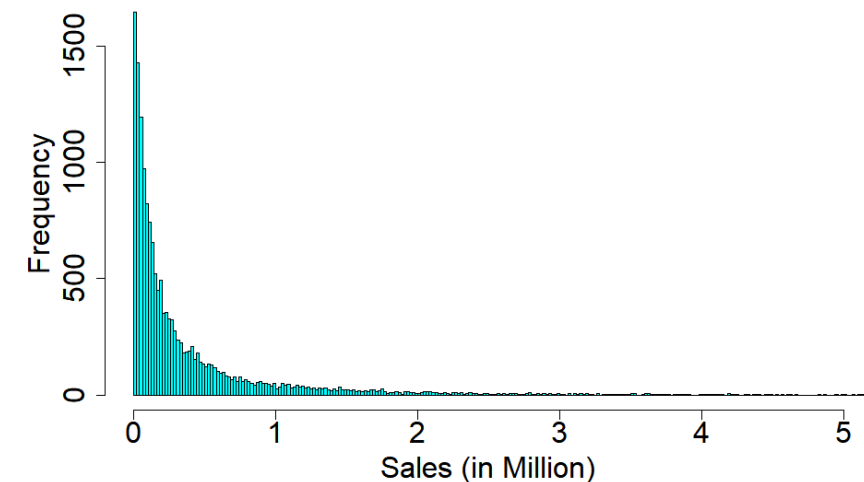


### Global Game Sales
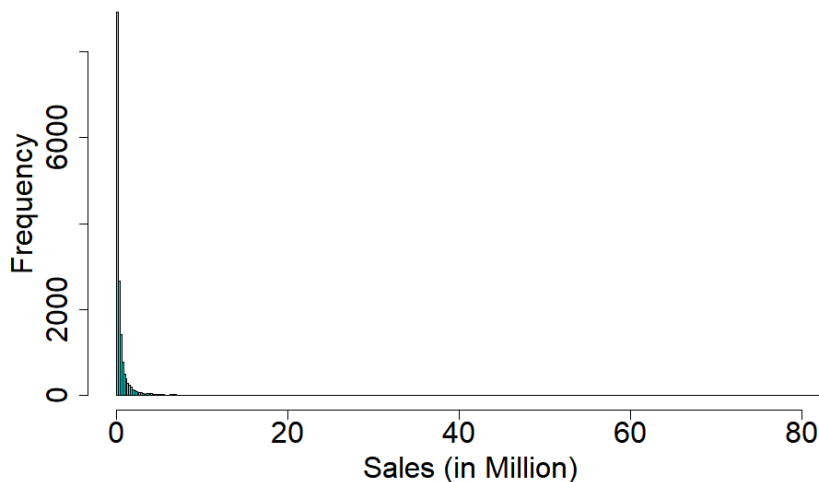
### Global Game Sales

### Global Game Sales

OVERVIEW OF THE ENTIRE GLOBAL SALES ENTITY

ZOOMED IN AT RECORDS THAT ARE WITH A LOWER FREQUENCY THAN 100

ZOOMED IN FOR THE ONES THAT HAVE A SALES VALUE LESS THAN 5 MILLION

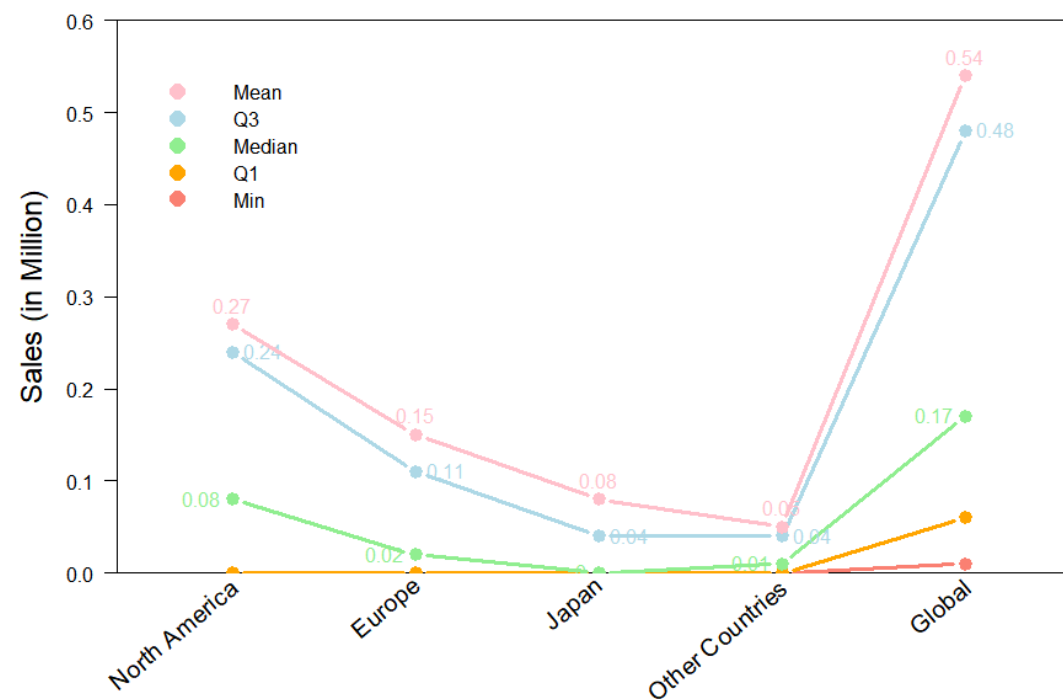# DESCRIPTIVE ANALYSIS | REGIONAL SALES BREAKDOWN

## REGIONAL SALES DESCRIPTIVE STATISTICS BREAKDOWN

In this section, I have pulled all central tendency measurements along with some other descriptive datapoints for each of the four major regions – North America, Europe, Japan, and the rest of the World. Comparing to the global market, these four regions have the same issue as well, all regions are highly positive skewed just like what we mentioned in the last slide.

As we can see from the graph to the right, North America has a higher variation in sales then the rest of the regions, and it made a huge affect to the global gaming market. The chart also conveyed that for 75% of the total game sales in each region, none of the four regions have excessed 0.24 million. Which could be also interpreted as the top 25% percent of the games are contributing more than



Regional Sales Statistics

| | North America | Europe | Japan | Other Countries | Global |
|---|---|---|---|---|---|
| **Min.** | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 |
| **Mode** | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 |
| **Q1** | 0.00 | 0.00 | 0.00 | 0.00 | 0.06 |
| **Median** | 0.08 | 0.02 | 0.00 | 0.01 | 0.17 |
| **Mean** | 0.27 | 0.15 | 0.08 | 0.05 | 0.54 |
| **Q3** | 0.24 | 0.11 | 0.04 | 0.04 | 0.48 |
| **Max.** | 41.49 | 29.02 | 10.22 | 10.57 | 82.74 |

# DESCRIPTIVE ANALYSIS | REGIONAL MARKET PERCENTAGE

## REGIONAL MARKET SALES PERCENTAGE OVERVIEW

Moreover we are going to look at how much are each regional market weighted comparing to the global market.

According to both table and chart, it is quite obvious that North America is the major market worldwide, which took place of almost half of the entire global sales. Europe comes in as the second largest market, provides a quarter of the global sales.

|  | Total.Sales | Percent |
|---|---|---|
| North America | 4327.65 | 49.1% |
| Europe | 2406.69 | 27.3% |
| Japan | 1284.27 | 14.6% |
| Other Countries | 788.91 | 9% |
| Global | 8811.97 | 100% |

### Regional Sales Percentage

Other Countries 9.0%

Japan 14.6%

Europe 27.3%

North America 49.1%
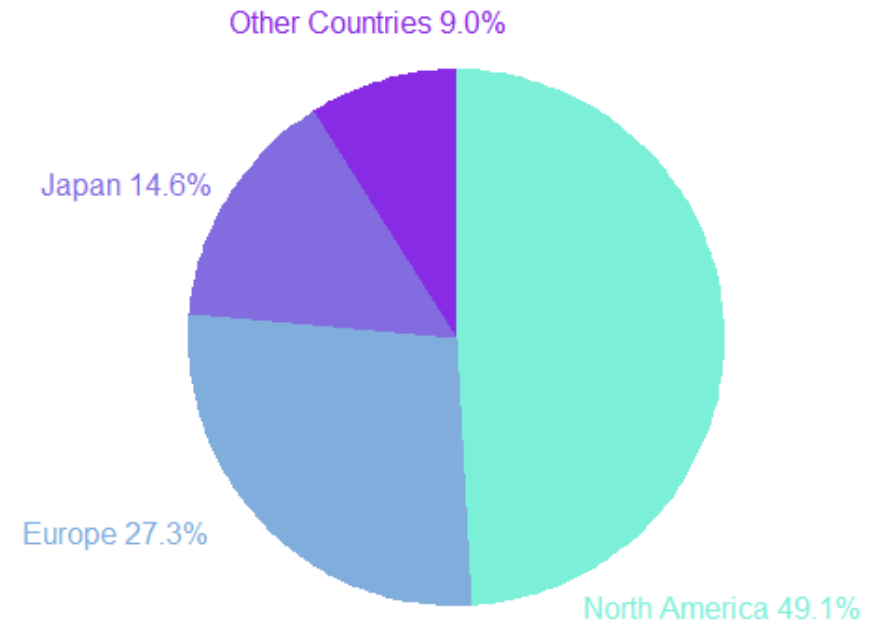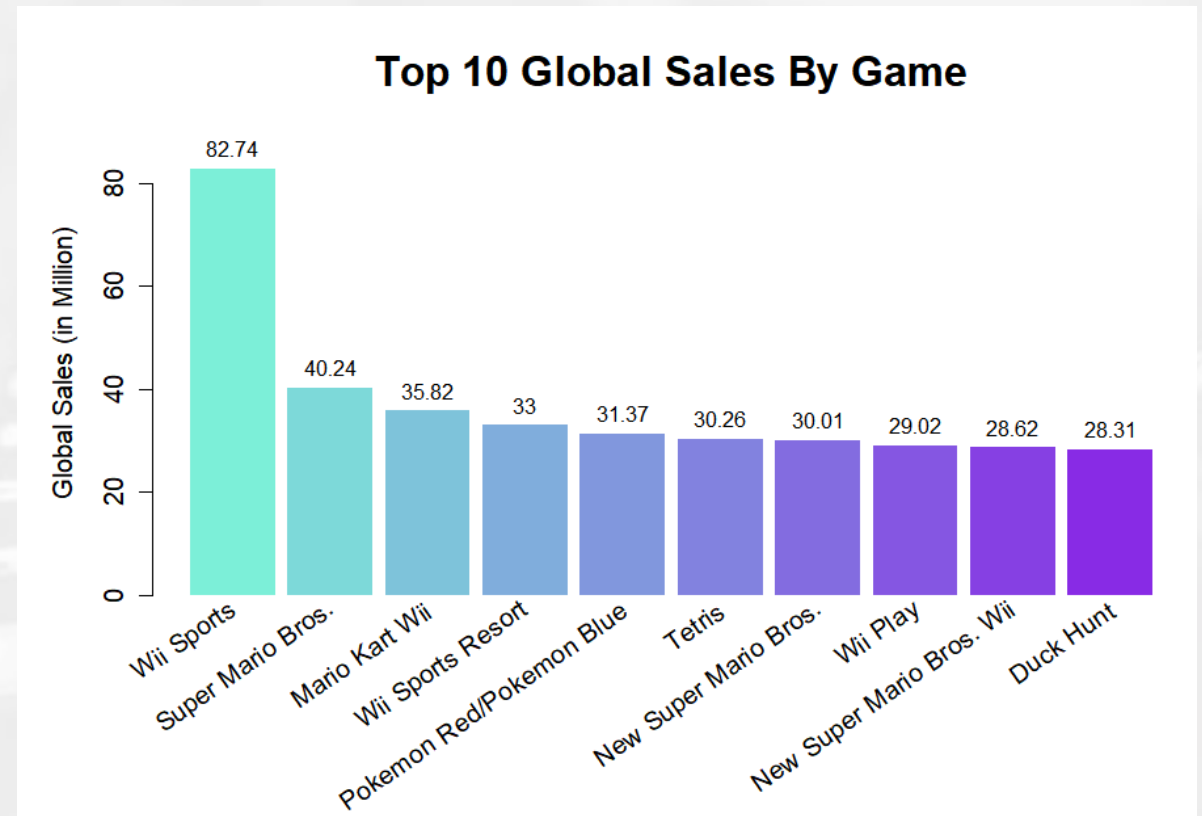
# DESCRIPTIVE ANALYSIS | GLOBAL BEST-SELLING GAMES

## BUSINESS QUESTION: WHAT ARE SOME OF THE BEST-SELLING GAMES IN GLOBAL MARKET?

| Name | Platform | Year | Genre | Publisher | Global_Sales |
|---|---|---|---|---|---|
| Wii Sports | Wii | 2006 | Sports | Nintendo | 82.74 |
| Super Mario Bros. | NES | 1985 | Platform | Nintendo | 40.24 |
| Mario Kart Wii | Wii | 2008 | Racing | Nintendo | 35.82 |
| Wii Sports Resort | Wii | 2009 | Sports | Nintendo | 33.00 |
| Pokemon Red/Pokemon Blue | GB | 1996 | Role-Playing | Nintendo | 31.37 |
| Tetris | GB | 1989 | Puzzle | Nintendo | 30.26 |
| New Super Mario Bros. | DS | 2006 | Platform | Nintendo | 30.01 |
| Wii Play | Wii | 2006 | Misc | Nintendo | 29.02 |
| New Super Mario Bros. Wii | Wii | 2009 | Platform | Nintendo | 28.62 |
| Duck Hunt | NES | 1984 | Shooter | Nintendo | 28.31 |

This analysis answer the business question above by showing, and plotting a bar chat of, the top 10 games that made the most sales worldwide. "Wii Sports" being the top selling game, had a surprisingly well performance. The sales has doubled the game in the second place (Super Mario Bros.)

But it has brought to my attention that maybe there are some games that have multiple records, since there might be couple version release on different year or through different platform. Hence, an in-depth analysis, regarding this specific issue, will be provided on the next slide



**Top 10 Global Sales By Game**

## BUSINESS QUESTION: WHAT ARE SOME OF THE BEST-SELLING GAMES IN GLOBAL MARKET?

Taking the issue that I've mentioned in the last slide, I did a little further investigation on the given data. After pulling the number of games with unique names, result shows a total number of 11325, far less than the 16291 records of dataset that we are working with. This implies that my concern was correct, that are actually a decent amount of games that came in as different records, but they are actually the same game that are just released on different years or published n different platforms.

```
> count(distinct(vgsales,vgsales$Name))
      n
1 11325
```

TI did a little more data cleaning by aggerating the records of the same game together, so now we are dealing with this even cleaner and more effective data. From there, I was able to another subset of the top 10 worldwide best-selling games, as shown below.. Which does look quite different than the one that we have previously pulled, both the game rankings and the global sales. And, from the table below, it is quite self-explanatory that how each records are aggregated and when and where did each version been release. A new bar graph of this updated data subset has generated and displayed on the following page.

| | Game_Name | Platform | Year | Genre | Publisher | Global_Sales |
|---|---|---|---|---|---|---|
| 1 | Wii Sports | Wii | 2006 | Sports | Nintendo | 82.74 |
| 2 | Grand Theft Auto V | PS3/X360/PS4/XOne/PC | 2013/2013/2014/2014/2015 | Action/Action/Action/Action/Action | Take-Two Interactive/Take-Two Interactive/Take-Two Interact... | 55.92 |
| 3 | Super Mario Bros. | NES/GB | 1985/1999 | Platform/Platform | Nintendo/Nintendo | 45.31 |
| 4 | Tetris | GB/NES | 1989/1988 | Puzzle/Puzzle | Nintendo/Nintendo | 35.84 |
| 5 | Mario Kart Wii | Wii | 2008 | Racing | Nintendo | 35.82 |
| 6 | Wii Sports Resort | Wii | 2009 | Sports | Nintendo | 33.00 |
| 7 | Pokemon Red/Pokemon Blue | GB | 1996 | Role-Playing | Nintendo | 31.37 |
| 8 | Call of Duty: Modern Warfare 3 | X360/PS3/PC/Wii | 2011/2011/2011/2011 | Shooter/Shooter/Shooter/Shooter | Activision/Activision/Activision/Activision | 30.83 |
| 9 | New Super Mario Bros. | DS | 2006 | Platform | Nintendo | 30.01 |
| 10 | Call of Duty: Black Ops II | PS3/X360/PC/WiiU | 2012/2012/2012/2012 | Shooter/Shooter/Shooter/Shooter | Activision/Activision/Activision/Activision | 29.72 |

## BUSINESS QUESTION: WHAT ARE SOME OF THE BEST-SELLING GAMES IN GLOBAL MARKET?

As shown on this updated bar chart to the right, the global sales distribution curve is slightly different than the one that we analyzed in the previous slide. But if given a closer look, the game ranking has been changed a lot. Although "Wii Sports" still maintains the first place, but the second best-selling game is now "Grand Theft Auto V", instead of the previous "Super Mario Bros." (which is now at the 3rd place). Interestingly this new 2nd place was not even in the previous top 10 list at all. This implies that this game might have a lot of versions and each version probably got a decent amount of hits.

The two tables below provides a quick overlook at the both lists, before data aggregation and after.

### BEFORE AGGREGATION

| Year | Platform | Name |
|------|----------|------|
| 2006 | Wii | Wii Sports |
| 1985 | NES | Super Mario Bros. |
| 2008 | Wii | Mario Kart Wii |
| 2009 | Wii | Wii Sports Resort |
| 1996 | GB | Pokemon Red/Pokemon Blue |
| 1989 | GB | Tetris |
| 2006 | DS | New Super Mario Bros. |
| 2006 | Wii | Wii Play |
| 2009 | Wii | New Super Mario Bros. Wii |
| 1984 | NES | Duck Hunt |

### AFTER AGGREGATION

| | Game_Name | Platform | Year |
|----|-----------|----------|------|
| 1 | Wii Sports | Wii | 2006 |
| 2 | Grand Theft Auto V | PS3/X360/PS4/XOne/PC | 2013/2013/2014/2014/2015 |
| 3 | Super Mario Bros. | NES/GB | 1985/1999 |
| 4 | Tetris | GB/NES | 1989/1988 |
| 5 | Mario Kart Wii | Wii | 2008 |
| 6 | Wii Sports Resort | Wii | 2009 |
| 7 | Pokemon Red/Pokemon Blue | GB | 1996 |
| 8 | Call of Duty: Modern Warfare 3 | X360/PS3/PC/Wii | 2011/2011/2011/2011 |
| 9 | New Super Mario Bros. | DS | 2006 |
| 10 | Call of Duty: Black Ops II | PS3/X360/PC/WiiU | 2012/2012/2012/2012 |



Top 10 Global Sales By Game (Updated)
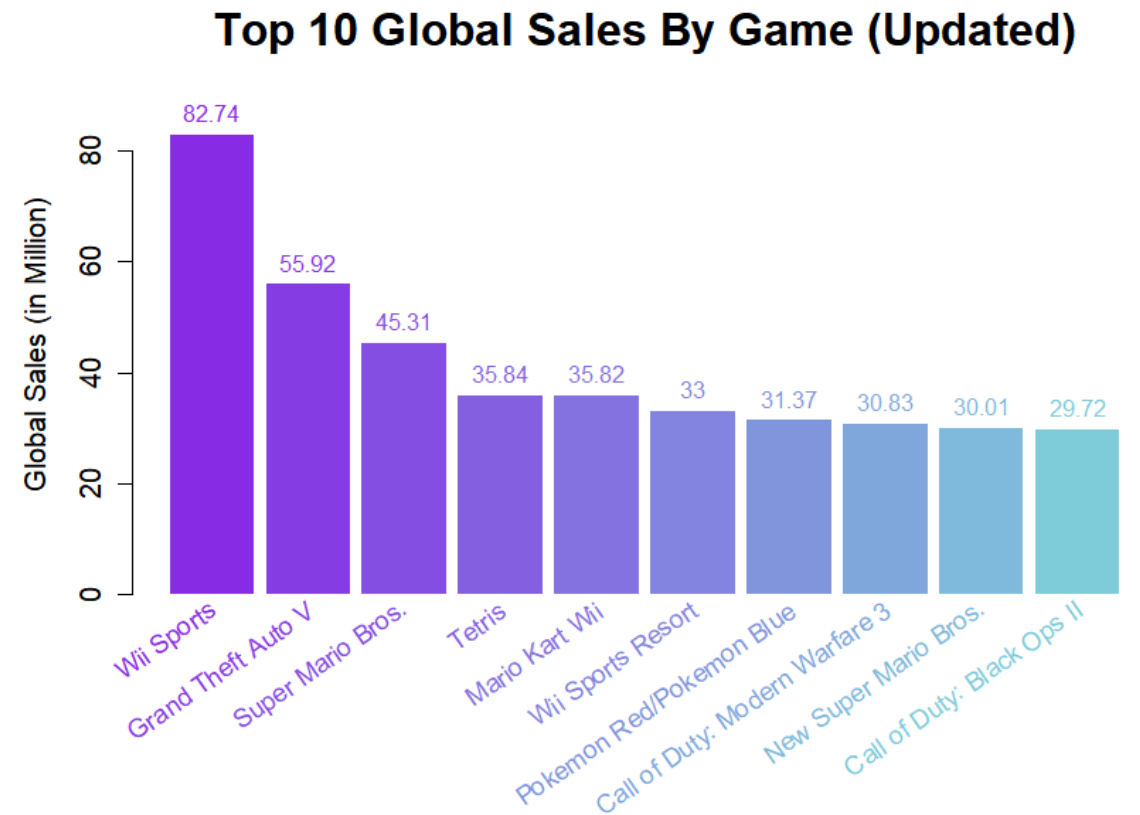
# EXPLORATION ANALYSIS | POPULAR GENRE OF GAMES

## BUSINESS QUESTION: WHAT ARE SOME POPULAR GENRES OF GAME AROUND THE WORLD?

Within this section, we are going to explore a little more with the data as well as doing some inferential analysis, see what else this dataset tells us. And how can we further improve the gamming industry.

Firstly, I'm curious whether people have specific preferences or easier get attracted to a certain types of games. So I started off listing out all the genres appeared in the dataset. This attribute is rather easy to manage since there are only 12.

Evidently, people do have a quite strong preference over action and sports type of games, and a slight preference on role-playing, shooting and adventure sort of games, there are still quite a few playing the rest types of games but definitely not the major customer.

So it would be a safe guess that if there are more actions games come out later on, they will still be quite popular.

|  | Count | Percent | CumCount | CumPercent |
|---|---|---|---|---|
| **Action** | 3251 | 19.96 | 3251 | 19.96% |
| **Sports** | 2304 | 14.14 | 5555 | 34.1% |
| **Misc** | 1686 | 10.35 | 7241 | 44.45% |
| **Role-Playing** | 1470 | 9.02 | 8711 | 53.47% |
| **Shooter** | 1282 | 7.87 | 9993 | 61.34% |
| **Adventure** | 1274 | 7.82 | 11267 | 69.16% |
| **Racing** | 1225 | 7.52 | 12492 | 76.68% |
| **Platform** | 875 | 5.37 | 13367 | 82.05% |
| **Simulation** | 848 | 5.21 | 14215 | 87.26% |
| **Fighting** | 836 | 5.13 | 15051 | 92.39% |
| **Strategy** | 670 | 4.11 | 15721 | 96.5% |
| **Puzzle** | 570 | 3.50 | 16291 | 100% |



Gerne Popularity Distribution
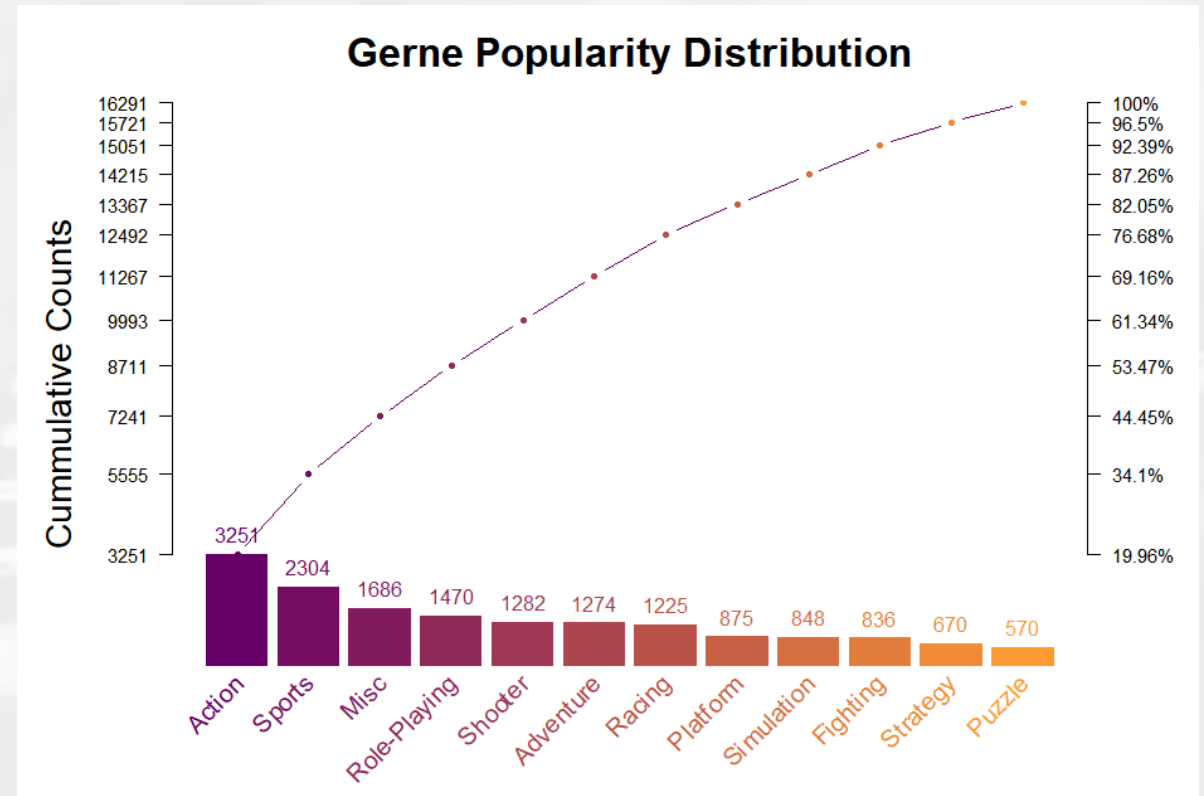
**BUSINESS QUESTION: WHAT ARE SOME POPULAR GAMING PLATFORMS AROUND THE WORLD?**

| | Count | Percent | CumCount | CumPercent |
|---|---|---|---|---|
| DS | 2131 | 13.08 | 2131 | 13.08% |
| PS2 | 2127 | 13.06 | 4258 | 26.14% |
| PS3 | 1304 | 8.00 | 5562 | 34.14% |
| Wii | 1290 | 7.92 | 6852 | 42.06% |
| X360 | 1234 | 7.57 | 8086 | 49.63% |
| PSP | 1197 | 7.35 | 9283 | 56.98% |
| PS | 1189 | 7.30 | 10472 | 64.28% |
| PC | 938 | 5.76 | 11410 | 70.04% |
| XB | 803 | 4.93 | 12213 | 74.97% |
| GBA | 786 | 4.82 | 12999 | 79.79% |

None like genre of games, there are way more gaming platforms, thus we are only going to only analyze the top 10 platforms that has the most released games. And these 10 platforms have actually held 80% of the games on the market.

DS and PS2 are performing similarly , they came quite close in terms of the number of games they own. But they definitely not dominating the industry, each owns about 13% of the total games. I believe the following couple of platforms like PS3, Wii, X360, and PSP, would be able to catch up rather soon.



**Plaform Popularity Distribution**
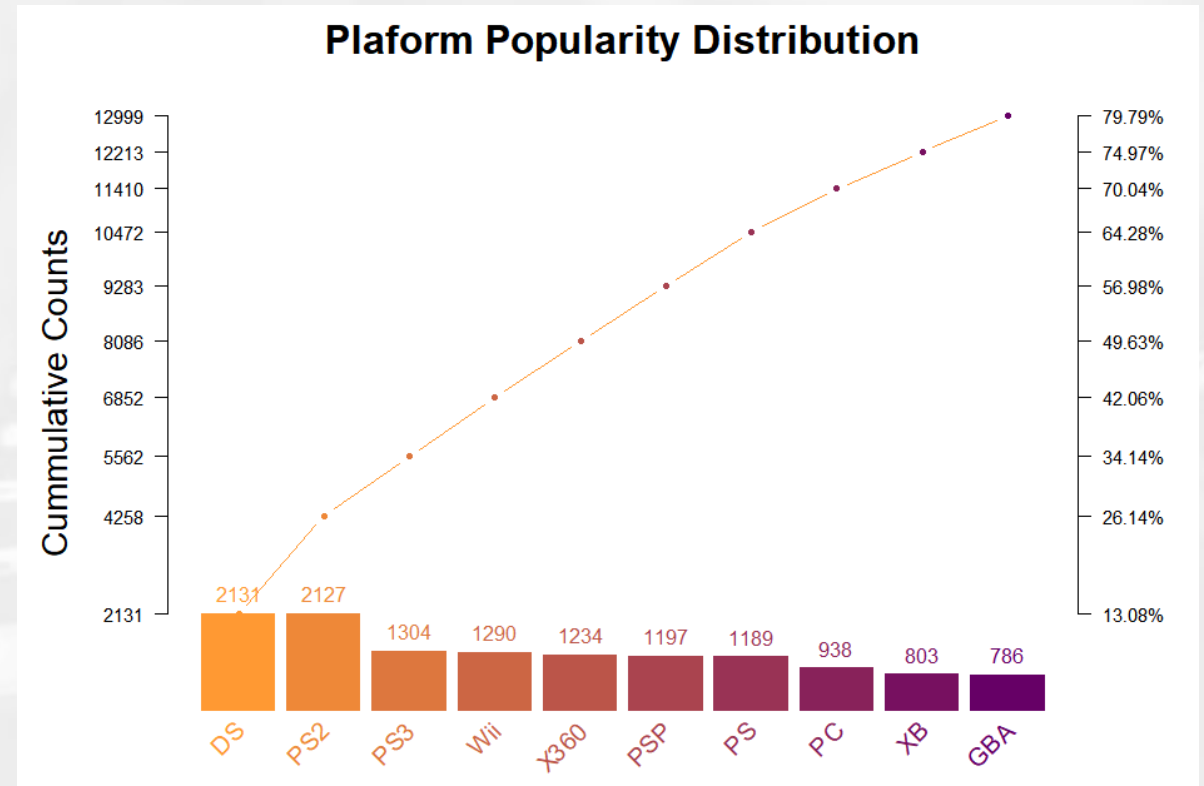
# EXPLORATION ANALYSIS | POPULAR GAME PUBLISHERS

## BUSINESS QUESTION: WHAT ARE SOME POPULAR GAME PUBLISHERS AROUND THE WORLD?

| | Count | Percent | CumCount | CumPercent |
|---|---|---|---|---|
| Electronic Arts | 1339 | 8.22 | 1339 | 8.22% |
| Activision | 966 | 5.93 | 2305 | 14.15% |
| Namco Bandai Games | 928 | 5.70 | 3233 | 19.85% |
| Ubisoft | 918 | 5.64 | 4151 | 25.49% |
| Konami Digital Entertainment | 823 | 5.05 | 4974 | 30.54% |
| THQ | 712 | 4.37 | 5686 | 34.91% |
| Nintendo | 696 | 4.27 | 6382 | 39.18% |
| Sony Computer Entertainment | 682 | 4.19 | 7064 | 43.37% |
| Sega | 632 | 3.88 | 7696 | 47.25% |
| Take-Two Interactive | 412 | 2.53 | 8108 | 49.78% |

The final analysis will solve the business question of which are the most popular publishers. Again, we will look at the top 10 for now, these 10 publishers have published almost 50% of the total games around the world which is pretty impressive.

Even though Electronic Arts is ranked at top 1 for the most frequent release of games. I do realize that Nintendo looks like a dark horse with huge potentials, since 7 out of the top 10 best-selling games were from them.



Publisher Popularity Distribution
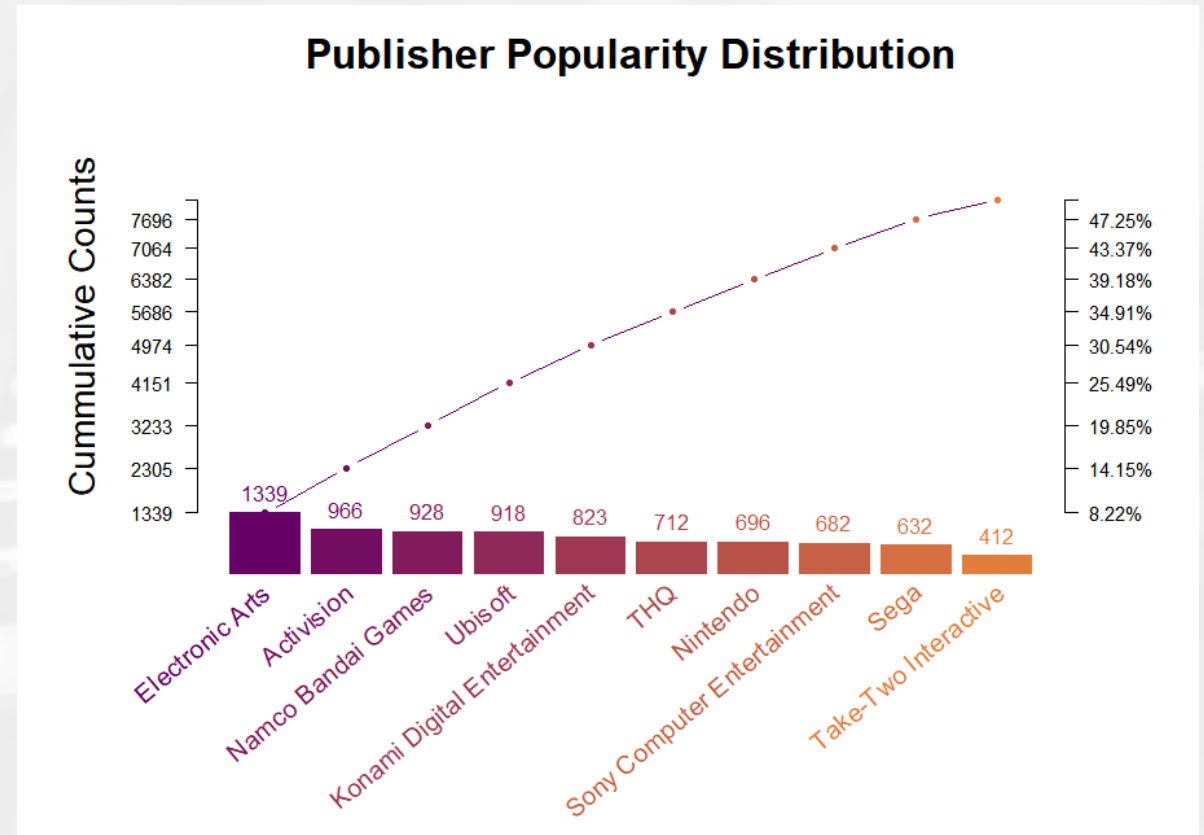
# REFLECTIONS & INSIGHTS

I found this case study to be quite interesting. Especially at this pandemic moment, there are more people staying at home, video games are becoming a rather popular activity to do either by oneself or online team up with other friends. So it was quite delightful analysis.

Although more researches could be made toward this case study of the gaming industry, we did get quite a lot from the analyzation. Moving forward, while maintaining the current customers, I suggest to focus more on developing action, sports and adventurous type of games.

Also Nintendo, Wii, and X360 are some of the great platform and publishers to put more attention on.

# References

Smith, G. Video Game Sales Data. Retrieved from https://www.kaggle.com/gregorut/videogamesales

VGChartz. https://www.vgchartz.com/#

Kabacoff, R.K. (2015). R in Action 2nd Edition. Manning. Shelter Island, NY.

Bluman, A.B. (2018). Elementary Statistics: A Step by Step Approach 10th Edition, McGraw-Hill. New York, NY.

R: Basics of R objects; entering and manipulating data. https://egret.psychol.cam.ac.uk/statistics/R/enteringdata.html

combine rows in r, summing values and concatenating characters; combining factor levels. https://stackoverflow.com/questions/56682598/combine-rows-in-r-summing-values-and-concatenating-characters-combining-factor

David, D. (Nov. 12, 2019). How to Make Beautiful Tables in R. https://rfortherestofus.com/2019/11/how-to-make-beautiful-tables-in-r/

Plot multiple lines (data series) each with unique color in R. https://stackoverflow.com/questions/14860078/plot-multiple-lines-data-series-each-with-unique-color-in-r

How do I flip rows and columns in R. https://stackoverflow.com/questions/33643181/how-do-i-flip-rows-and-columns-in-r