**VGSALES ANALYSIS**

**By Okwaraojimadu Kingsley Chisom**

# Tabe of Content

Data card

Aims of the objectives

Objectives

Methodology

Data Reconnaissance

Data Upload

Data Cleaning: Remove empty cells, Remove duplicates, Remove improper values (Max and Min)

Cleaned Data Exportation

Analysis on Excel

Data Processing

Data Visualization

Discussion

# Data card

This dataset contains a list of video games with sales greater than 100,000 copies.

Fields include:

* Rank - Ranking of overall sales
* Name - The games name
* Platform - Platform of the game’s release
* Year - Year of 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

## Aim of the objectives

Sales analysis for Nintendo sports games based on Year of Publish and Platform

## Objectives

1. Best Selling Nintendo sports games in the world for Wii
2. Best selling years for Nintendo sports games in the world for Wii
3. Best Selling Nintendo sports games in North America for Wii
4. Best selling years for Nintendo sports games in North America for Wii
5. Best Selling Nintendo sports games in the world for Gameboy(GB)
6. Best selling years for Nintendo sports games in the world for Dual Screen(DS)
7. Best Selling Nintendo sports games in North America for DS
8. Best selling years for Nintendo sports games in North America for DS

# Methodology

## Data Reconnaissance

I downloaded the dataset for kaggle and thoroughly went through its data card to get familiarized with the data. Next i opened the data on excel to make sure that what was mentioned in the datacard was correct.

## Data Upload

I set up my jupyter notebook and imported the necessary preliminary libraries before I imported the data.

## Data Cleaning

Next was to clean the data and I set up the following objectives to get it done;

1. Remove empty cells
2. Remove duplicates
3. Remove improper values (Max and Min)
4. Remove empty cells:

To do this i ran the *df.isna().sum()* code to find the empty spaces and it turned out that the data had 271empty cells under the Year column, 58 in the Publisher column and 2 under Global\_Sales column.

To remove the empty spaces, I ran the *df.dropna()* code then to verify my actions I rerun the *df.isna().sum()* code in the product dataset and it produced nils.

1. Remove duplicates:

The first step was to check if duplicates actually existed so I ran the *duplicated().sum()* code and it return nil. So therefore there were no duplicates to be treated.

1. Remove improper values (Max and Min):

To do this I had to run the following arithmetic calculations; mean, max and min. I did these using the following codes; *mean(), max(), min().* From my analysis of the resultant values, I could see that the values were not out of range and in proper alignment.

Note: I only did these cleaning steps due to the nature of the project. Further cleaning can be done if one deems it necessary.

## Cleaned Data Exportation

After cleaning was done. I exported the cleaned data as *“cleanvgsales.csv”*.

## Analysis on excel

I uploaded the cleaned data on excel for more analysis, data validation and possible further cleaning.

I used the filter option to check the string values for anomalies but didn’t find any that will be a hindrance to any of my objectives.

## Data Processing

I uploaded *“cleanvgsales.csv”* back to my notebook and was ready to start data processing based on my objectives.

Firstly I made sure all the important libraries were imported. Before copying the dataset into a new dataset which I named *Clean\_data.*

Next I filtered out all the rows that it’s *Publisher* column had *Nintendo* in it. As Nintendo was my desired Publisher. This was done with the help of *Clean\_data[Clean\_data['Publisher'].str.contains('Nintendo')]* code.

Next I removed the Rank column using the *drop(index=)* code as it was of no use to my analysis.

Going further, to secure my analysis so far I saved the current dataset as *nintendovgsales.csv* and as its name implies, it was a containing only data for the publisher named *nintendo.*

Next I imported *nintendovgsales.csv* as *tendo.df.*

Contiuning my processing, I filtered out all the rows that its *Genre* column contains *Sports.* I also repeated this process for *Racing* and *Action* before settling for the sports dataset and giving it the name *nin\_sports\_df.*

## Data processing and Visualization

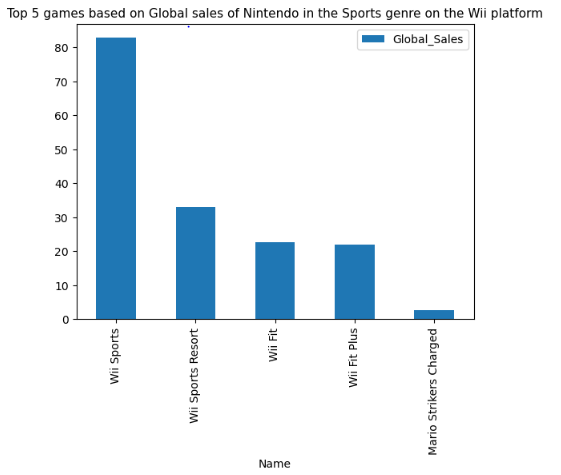
The only thing remaining was to filter out the particular platforms I wanted and visualize them.

### Wii:

From the *nin\_sports\_df* dataset, I filtered out all the rows that its *platform* column contains *Wii* and named it *nin\_sports\_wii\_df.*

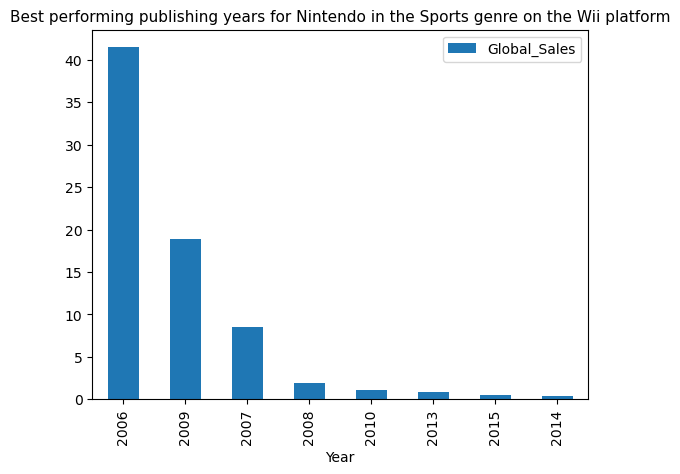
Next I sorted out the top 5 games in descending order based on global sales using the *sort\_values(by='Global\_Sales', ascending=False)* code.

For visualization, I used a bar chart with *name* on the x axis and *global\_sales* on the y axis.



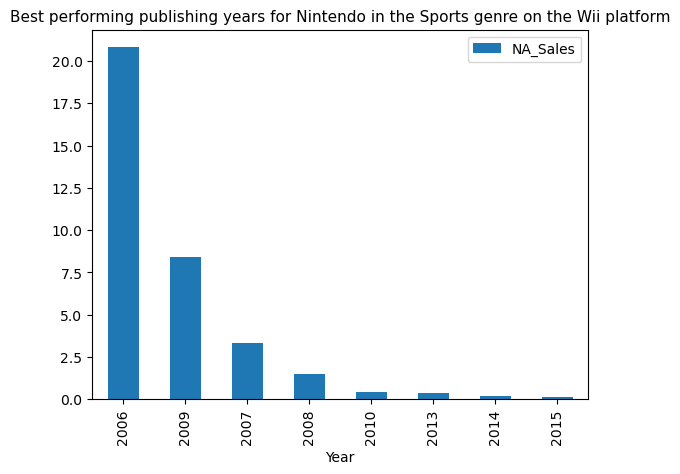
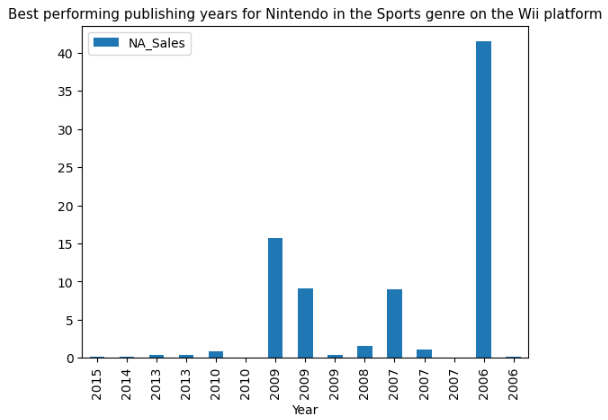
From the diagram, we can see Wii sports topping the charts with Mario strikers charged being the least.

Next, I filtered out the columns *Year* and *Global\_sales* and sorted it by year while visualizing.



From the chart, I could see that the sports games published by nintendo on wii in the year 2006 were doing fantastically well compared to other years. This was followed by 2009 and 2007.

I repeated the earlier procedures for *NA\_Sales* and got the following Charts:



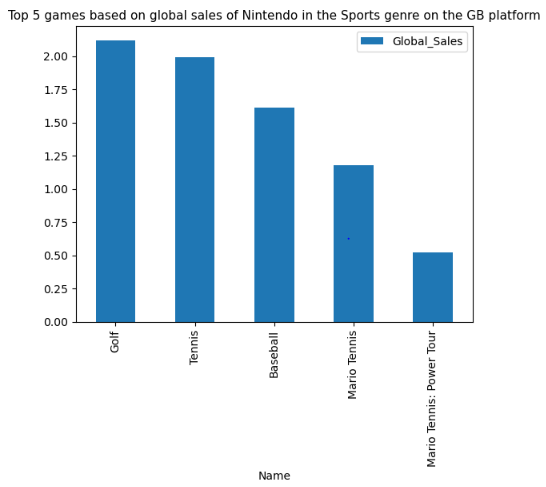
From the above charts, we can see that Wii exhibits the same characteristics for both global and north america sales.

### Gameboy, GB

Going forward, I filtered out all the rows that its *platform* column contains *GB* and named it *nin\_sports\_GB\_df,*

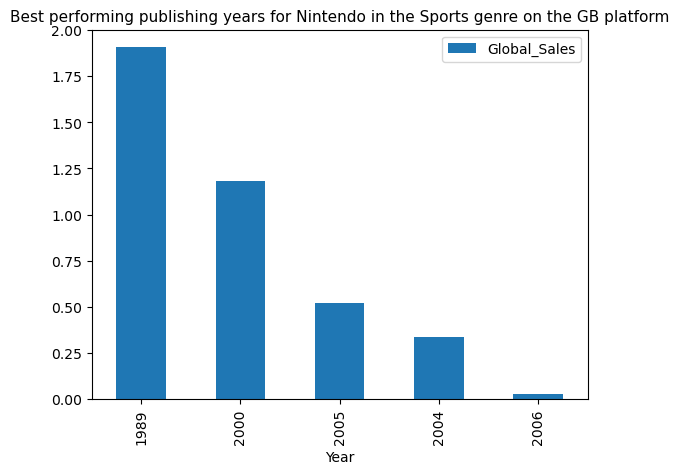
Next I sorted out the top 5 games in descending order based on global sales using the *sort\_values(by='Global\_Sales', ascending=False)* code.

For visualization, I used a bar chart with *name* on the x axis and *global\_sales* on the y axis.



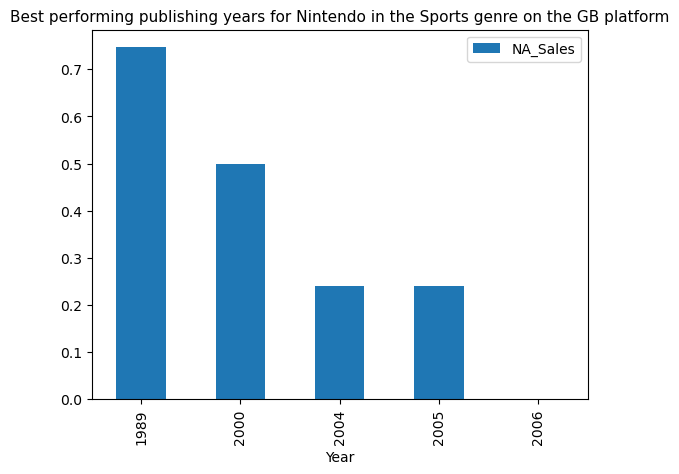
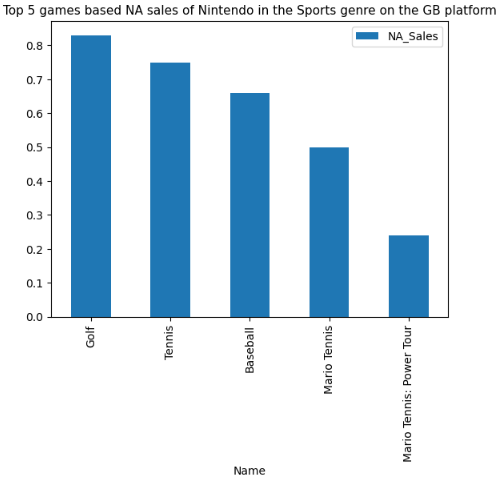
From the diagram, we can see Golf topping the charts with Mario Tennis: Power Tour being the least.

Next, I filtered out the columns *Year* and *Global\_sales* and sorted it by year while visualizing.



From the chart, I could see that the sports games published by nintendo on GB in the year 1989 were doing fantastically well compared to other years. This was followed by 2000 and 2005.

I repeated the earlier procedures for *NA\_Sales* and got the following Charts:



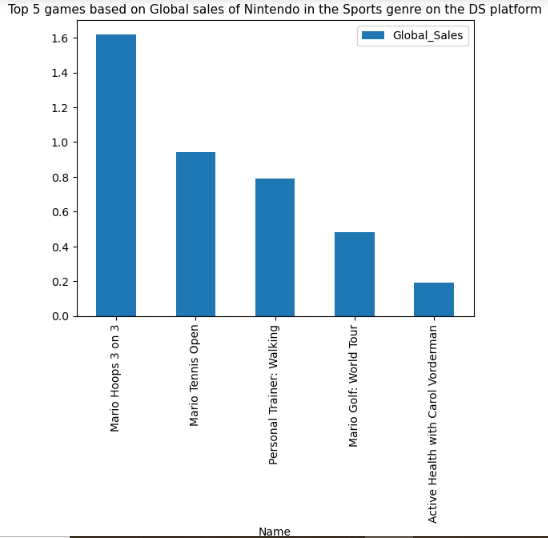
From the above charts we can see that GB exhibits the same characteristics for both global and North America sales.

### DS

To state, I filtered out all the rows that its *platform* column contains *DS* and named it *nin\_sports\_DS\_df,*

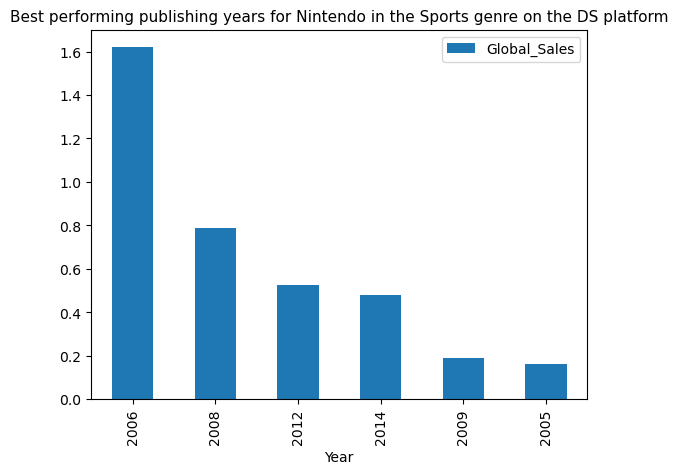
Next I sorted out the top 5 games in descending order based on global sales using the *sort\_values(by='Global\_Sales', ascending=False)* code.

For visualization, I used a bar chart with *name* on the x axis and *global\_sales* on the y axis.



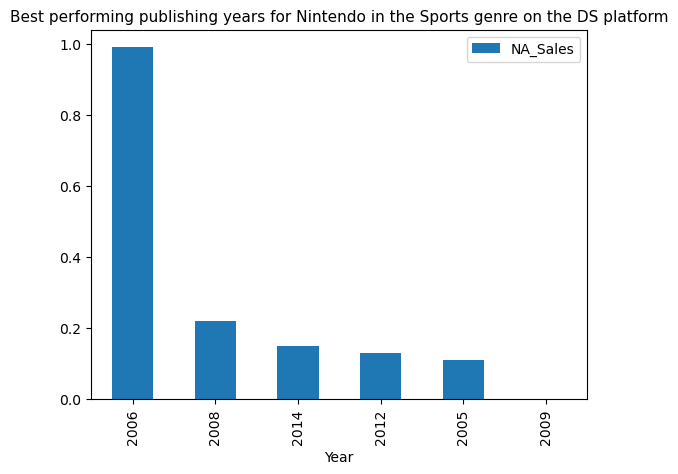
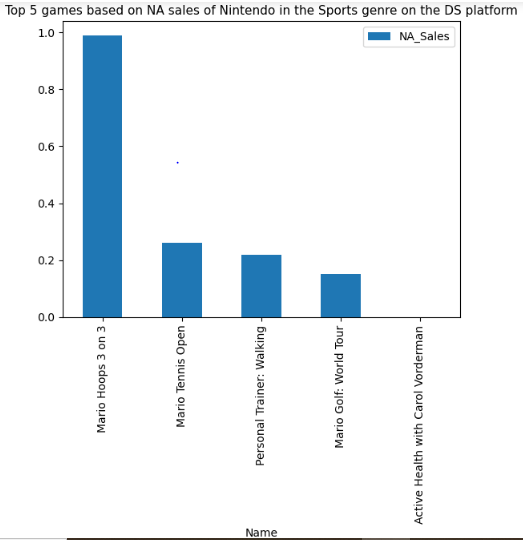
From the diagram, we can see Mario Hoops 3 on 3 topping the charts with Active Health with Carol Vorderman being the least.

Next, I filtered out the columns *Year* and *Global\_sales* and sorted it by year while visualizing.



From the chart, I could see that the sports games published by nintendo on DS in the year 2006 were doing fantastically well compared to other years. This was followed by 2012 and 2008.

I repeated the earlier procedures for *NA\_Sales* and got the following Charts:



From the above charts we can see that DS exhibits the same characteristics for both global and North America sales for the first chart but there were no games published on DS between 2021 and 2009. so therefore there is no bar for them. Even if any was published, it didn’t reach the minimum number of sales that was used for the dataset.

# Discussion

From the chart and data we can clearly see that most lucrative platform for Nintendo is Wii so I would advice for them to try to get more market share on other platforms to expand their reach and sales which will definitely result in more profit.

For the lower sales that I see for the more recent games, i can comfortable assume that this is because the have been on the market for lesser time than older games but nevertheless some earlier years exhibit more sales than earlier years which can be a good sign.

My primary solution for Nintendo is to both remodel and relaunch much more popular old games while recruiting better and creative talent to help create and market recent games.