INFO 5709 FALL 2021

DATA VISUALIZATION PROJECT REPORT

Video Games Sales Over Years Dataset

University of North Texas

Fall 2021

Date Submitted: 11/30/2021

Introduction:

There are many people interest in video games, and there are many popular video games that has billions of fans. Video games' publishers has gained billions of profits from their games. This project focused on analyzing profits of video games over different countries. People can get different information from this dataset such as the year of the game's release, the Genre of the game, and the game's world sales.

About the data set:

This dataset contains a record of video games with more than 100,000 sales. A scrape of vgchartz.com collects this dataset. The source of this dataset is Kaggle which is website has many different datasets. The script to scrape the data is available at https://github.com/GregorUT/vgchartzScrape. It is based on BeautifulSoup using Python. There are 16,598 registers for this dataset. However, two records were dropped because their information weren't complete. I chose to work on this dataset because it has different types of attributes that helps me to test my analytical and visualization skills

The attributes of video games dataset:

1. Rank - Ranking of overall sales
2. Name - The games name
3. Platform - Platform of the games release (i.e. PC,PS4, etc.)
4. Year - Year of the game's release
5. Genre - Genre of the game
6. Publisher - Publisher of the game
7. NA_Sales - Sales in North America (in millions)
8. EU_Sales - Sales in Europe (in millions)
9. JP_Sales - Sales in Japan (in millions)
10. Other_Sales - Sales in the rest of the world (in millions)
11. Global_Sales - Total worldwide sales.

Cleaning The Dataset:

I tried to clean my dataset by excel. First, I looked for blanks, and it shows me that there are no blanks in my dataset. I also looked for errors by "Go to Special" options, and there are no errors.

Tools Used:

- Tableau
- Excel
- Python

Exploratory Dataset Analysis:

After cleaning the dataset, I have used different tools to get different information by visualization and analyzing the data.

First, I used Excel to analyze and visualize some data:

- Pivot table, and Pivot chart:
- The pivot table shows total of global Sales and EU_Sales depend on genre
- The conditional formatting fills the sales that equal or larger than sales median

Genre	Sum of Global_Sales	Sum of EU_Sales
Action	1751.18	525
Adventure	239.04	64.13
Fighting	448.91	101.32
Misc	809.96	215.98
Platform	831.37	201.63
Puzzle	244.95	50.78
Racing	732.04	238.39
Role-Playing	927.37	188.06
Shooter	1037.37	313.27
Simulation	392.2	113.38
Sports	1330.93	376.85
Strategy	175.12	45.34
Grand Total	**8920.44**	**2434.13**

Median of Global Sales
771
Median of EU_Sales
194.845

The pivot tabels shows total of global Sales and EU_Sales depend on genre
The condentional formating fill the sales that equal or larger than sales median

Figure 1

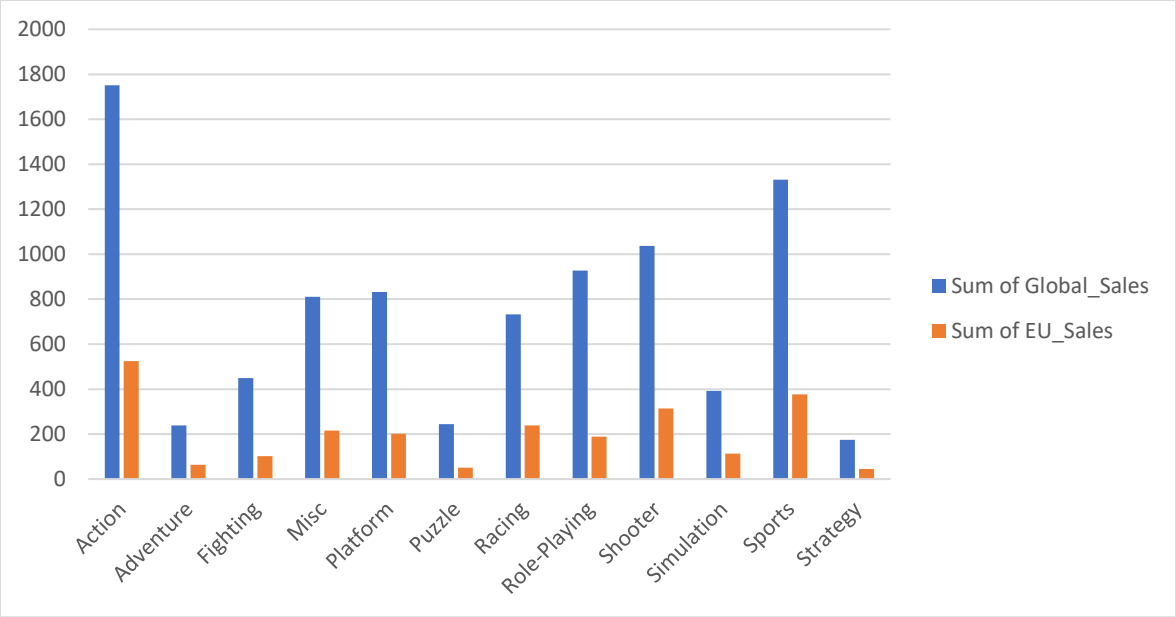


Figure 2

- Pivot Chart shows global_sales and EU_Sales total of all genre.

Another analyzing using Pivot table and chart:

- Pivot Chart:

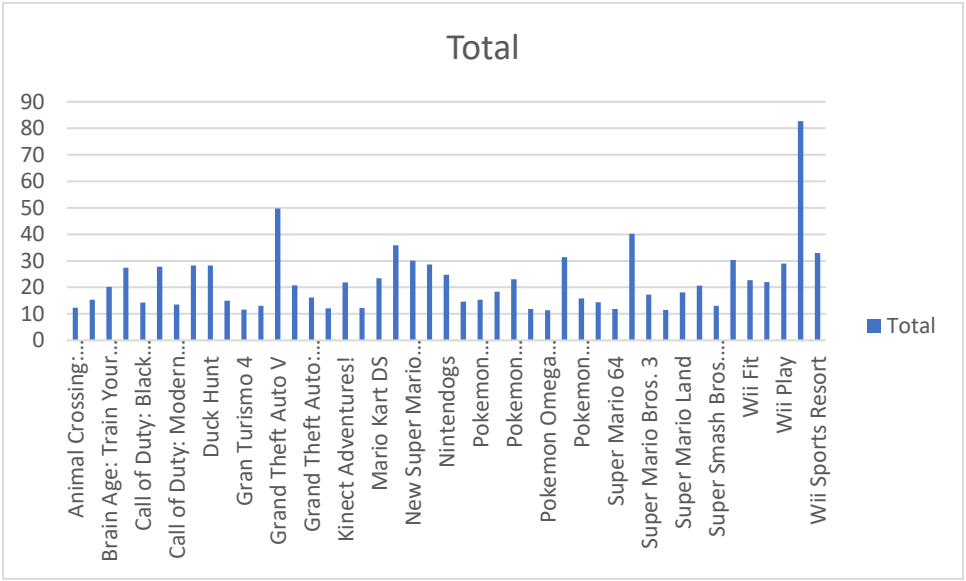


Figure 3

- The pivot table:

Rank	(Multiple Items)
Video Game Name	**Sum of Global_Sales**
Animal Crossing: Wild World	12.27
Brain Age 2: More Training in Minutes a Day	15.3
Brain Age: Train Your Brain in Minutes a Day	20.22
Call of Duty: Black Ops	27.37
Call of Duty: Black Ops 3	14.24
Call of Duty: Black Ops II	27.76
Call of Duty: Modern Warfare 2	13.51
Call of Duty: Modern Warfare 3	28.22
Duck Hunt	28.31
Gran Turismo 3: A-Spec	14.98
Gran Turismo 4	11.66
Grand Theft Auto III	13.1
Grand Theft Auto V	49.76
Grand Theft Auto: San Andreas	20.81
Grand Theft Auto: Vice City	16.15
Halo 3	12.14
Kinect Adventures!	21.82
Mario Kart 7	12.21
Mario Kart DS	23.42
Mario Kart Wii	35.82
New Super Mario Bros.	30.01
New Super Mario Bros. Wii	28.62
Nintendogs	24.76
PokÃ©mon Yellow: Special Pikachu Edition	14.64
Pokemon Black/Pokemon White	15.32
Pokemon Diamond/Pokemon Pearl	18.36
Pokemon Gold/Pokemon Silver	23.1
Pokemon HeartGold/Pokemon SoulSilver	11.9
Pokemon Omega Ruby/Pokemon Alpha Sapphire	11.33
Pokemon Red/Pokemon Blue	31.37
Pokemon Red/Pokemon Blue	31.37
Pokemon Ruby/Pokemon Sapphire	15.85
Pokemon X/Pokemon Y	14.35
Super Mario 64	11.89
Super Mario Bros.	40.24
Super Mario Bros. 3	17.28
Super Mario Galaxy	11.52
Super Mario Land	18.14
Super Mario World	20.61
Super Smash Bros. Brawl	13.04
Tetris	30.26
Wii Fit	22.72
Wii Fit Plus	22
Wii Play	29.02
Wii Sports	82.74
Wii Sports Resort	33
Grand Total	**1011.14**

- The pivot table and chart show the total global sales of top 50 video games
- Conditional Format (fill in yellow) shows all Global Sales larger than average sales of top 50.

Second: Using Python in analyzing and visualization data:

- Uploading data by pandas.

```
In [24]: #using pandas to read cvs dataset
         import pandas as pd
```

```
In [5]: VideoGames.head()
```

Out[5]:

	Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
0	1	Wii Sports	Wii	2006.0	Sports	Nintendo	41.49	29.02	3.77	8.46	82.74
1	2	Super Mario Bros.	NES	1985.0	Platform	Nintendo	29.08	3.58	6.81	0.77	40.24
2	3	Mario Kart Wii	Wii	2008.0	Racing	Nintendo	15.85	12.88	3.79	3.31	35.82
3	4	Wii Sports Resort	Wii	2009.0	Sports	Nintendo	15.75	11.01	3.28	2.96	33.00
4	5	Pokemon Red/Pokemon Blue	GB	1996.0	Role-Playing	Nintendo	11.27	8.89	10.22	1.00	31.37

```
In [6]: #presenting numbers of rows and columns
        VideoGames.shape
```

```
Out[6]: (16598, 11)
```

Figure 4

- Shape helps me understand how many columns and rows the data have: 11 column and 16598 rows.
- Finding how many missing values in publisher column.
- Calculating the global sales mean.

```
In [6]: #presenting numbers of rows and columns
        VideoGames.shape
```

```
Out[6]: (16598, 11)
```

```
In [15]: #calculate_missing_value
         missing_values=VideoGames['Publisher'].isnull()
         print(sum(missing_values))
```

```
58
```

```
In [19]: #calculating_mean_global_sales
         Mean_Global=VideoGames['Global_Sales'].mean()

         print(Mean_Global)
```

```
0.5374406555006628
```

Figure 5

- Visualization data by python:
  - presenting scatter plot of global sales and publishing dates.

```
In [23]: import matplotlib.pyplot as plt
         %matplotlib inline
         Global_Sales=VideoGames['Global_Sales']
         Date=VideoGames['Year']
         #visualization part
         plt.scatter(Global_Sales,Date,color='green',s=10) #the visualization by scatter

Out[23]: <matplotlib.collections.PathCollection at 0x21d57029a90>
```

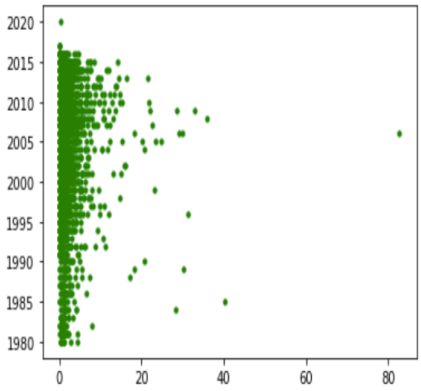


Figure 6

- Japanies Sales of different games based on Rank:

```
In [43]: #Line chart for Jpanies Sales
         Rank=VideoGames['Rank']
         Platform=VideoGames["JP_Sales"]
         plt.plot(Rank,Platform)
         plt.show()
```

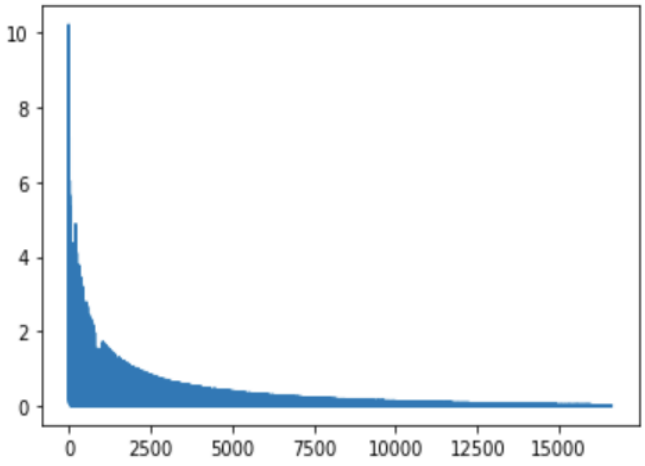


Figure 7

Third: Using Tableau for visualization data:

- What is the top 10 video games' genres?

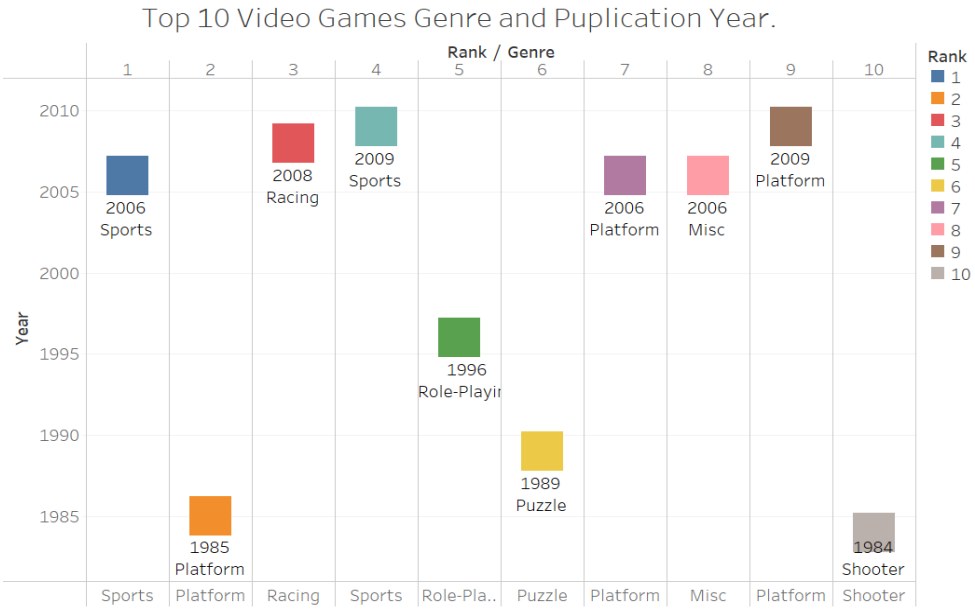


Figure 8

In this gantt chart, I filttered video games to the top 10 because I want to represent their genre and year. The genre differe over years. I notice that top 10 list doen't have any video games were published in last 10 years. However, some 80s' video games are still in the top ten.

- **To be more specific, I wanted to know if video games's fans in last 10 years has different favorite genre or it's simillar to other fans over last years.**

- What is the 10 bottom video games' genres?

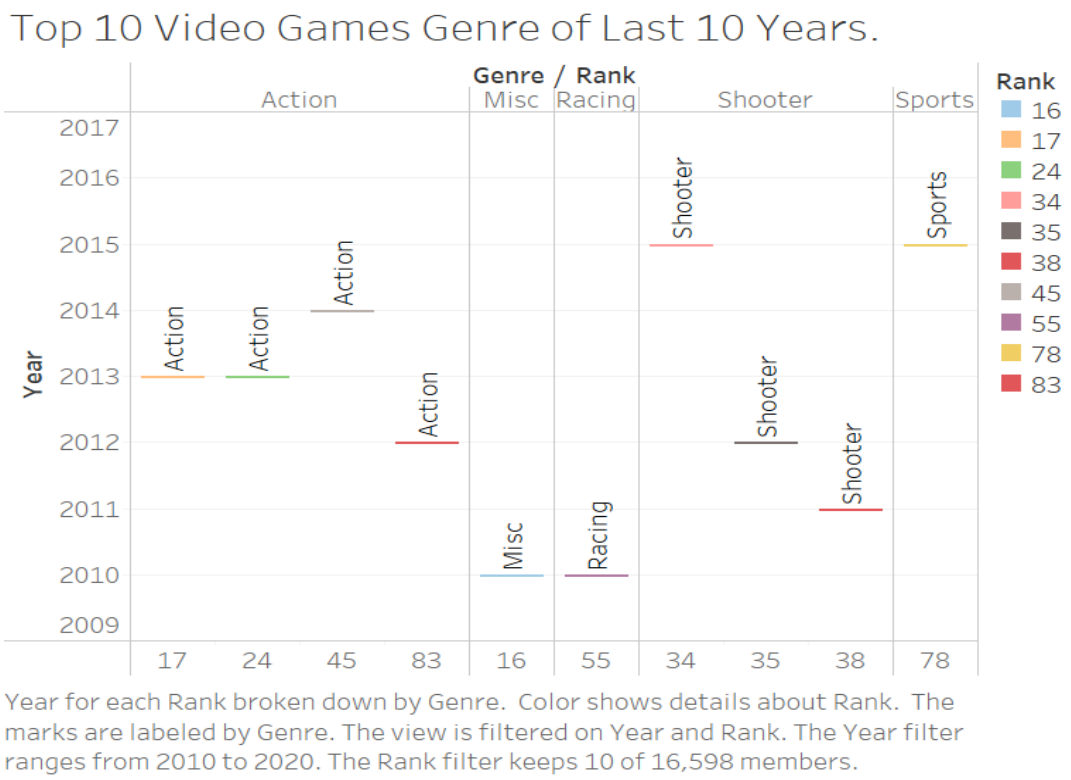


Figure 9

- **What is last bottom genre in the last 10 years?**

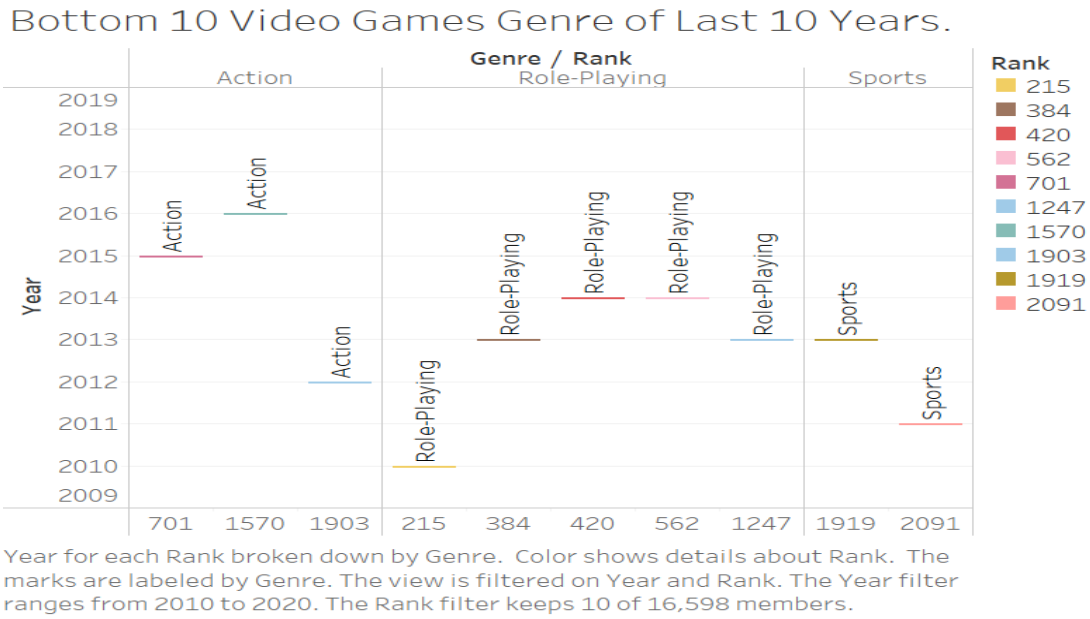


Figure 10

I filtered the years to last 10 years. This Gantt chart shows that action, sports, and role-playing are the bottom genre in last 10 years.

- **Video games publishing is a good business, and in every good business there will be leaders of this business. For example, google is the leader of search engines. I want to figure out the top 20 leaders of video games publishers. Who are they the top 20 video games' publishers?**

Top 20 Video Games' Publishers

Publisher	Rank	
Activision	34	2015
	35	2012
	38	2011
	138	2004
Electronic Arts	78	2015
	84	2009
Nintendo	1	2006
	3	2008
	4	2009
	5	1996
	7	2006
	8	2006
	9	2009
	11	2005
	12	2005
	13	1999
	14	2007
	15	2009
	20	2005
Take-Two Intera..	17	2013

Year broken down by Publisher and Rank. The view is filtered on Rank, which keeps 20 of 16,598 members.

Figure 11

- **In contrast, I need to know the publishers who make lowest profit.**
- **Who is the bottom 20 publishers?**

Bottom 20 Video Games's Publishers

Publisher	Rank	
Atari	470	2007
Capcom	215	2010
	384	2013
	612	1986
Electronic Arts	441	2007
Enix	313	1988
Corporation	427	1995
	447	1990
	532	1992
	603	1987
Konami Digital	266	2004
Entertainment	349	2007
Level 5	562	2014
Nintendo	339	2009
	420	2014
	574	1986
Square Enix	148	2006
SquareSoft	390	1994
	630	1992
Vivendi Games	282	1997

Year broken down by Publisher and Rank. The view is filtered on Rank, which keeps 20 of 16,598 members.

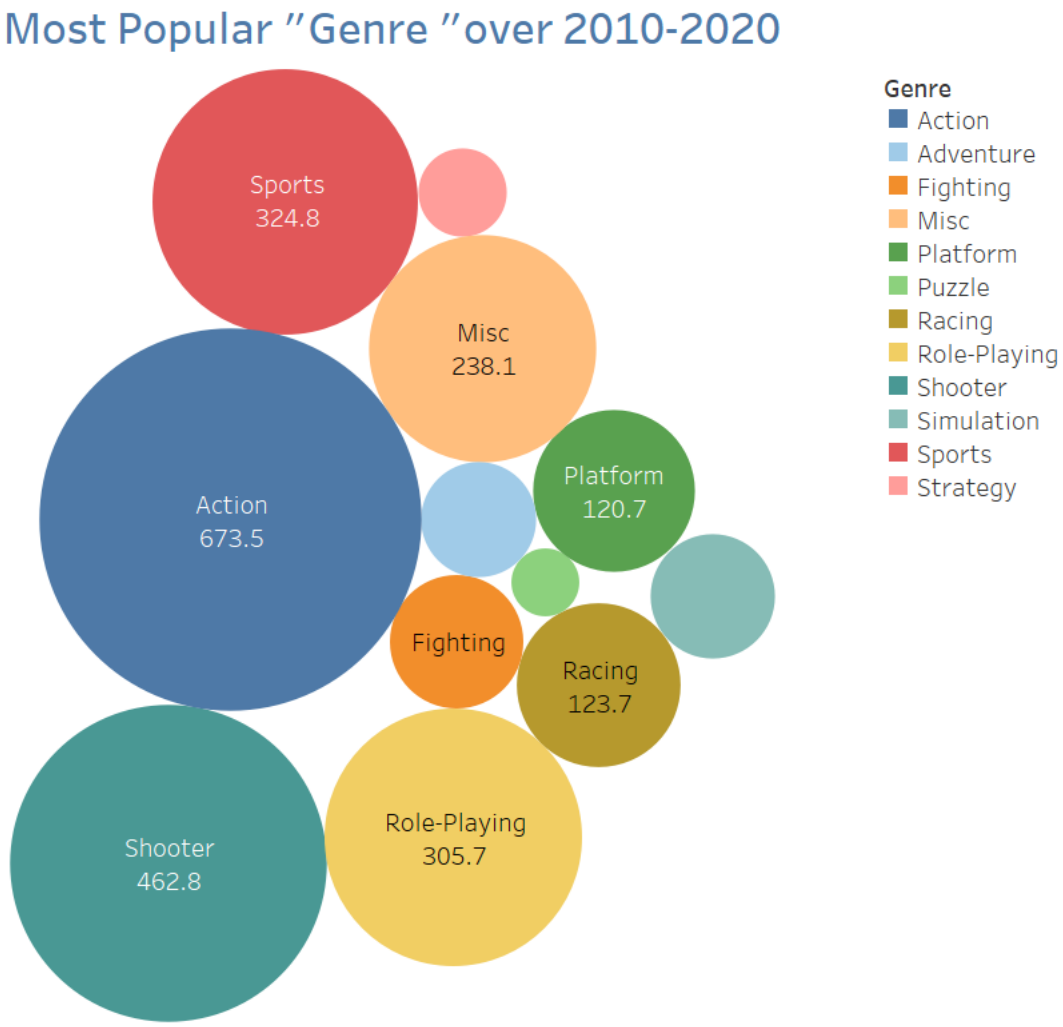
Figure 12

- **Hypothesis:**

After finding the genre of top 20 video games 's, I need to find the top 20 genre. In last analysis, I found that many video games' genres such as platforms, shooters, and action has been in the top 10.

Does that mean these are the top genre or it may be different?

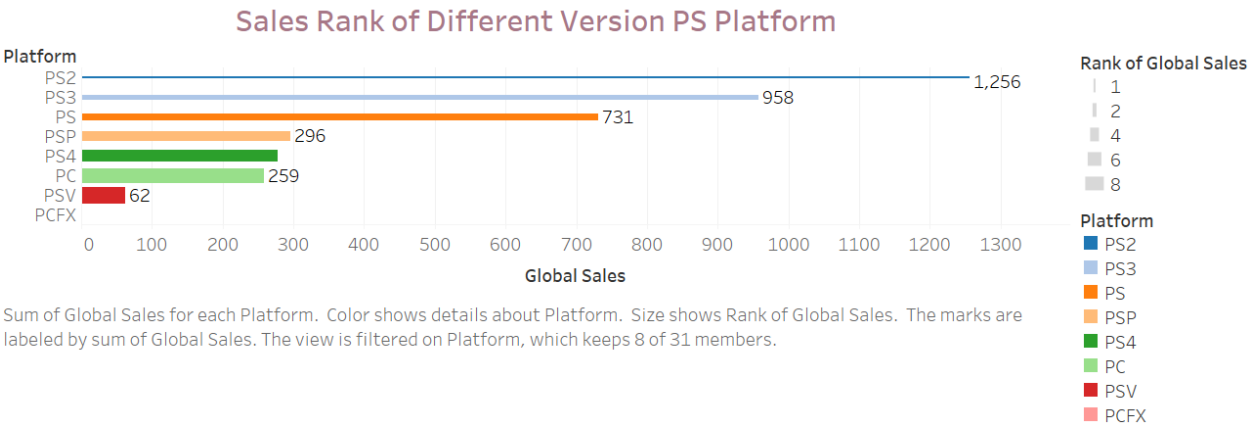
- **What is the top 20 genre in last 10 years?**



Genre and sum of Global Sales. Color shows details about Genre. Size shows sum of Global Sales. The marks are labeled by Genre and sum of Global Sales. The data is filtered on Year, which ranges from 2010 to 2020.

- **What is PS platform different versions sales' rank?**

I have analyzed different ranks based on video games, genre, and publishers. PS platform is popular platform that many users prefer to play video games by it. Sales of different versions of PS platform differs, so it is important to know which version has the highest sales. After finding sales of different versions of PS platform, I ranked it by "Quick Table Calculation" in the size of sales. The size of bars in the bar chart depends on the rank.



- **List publishers who got higher than 400 in global sales.**

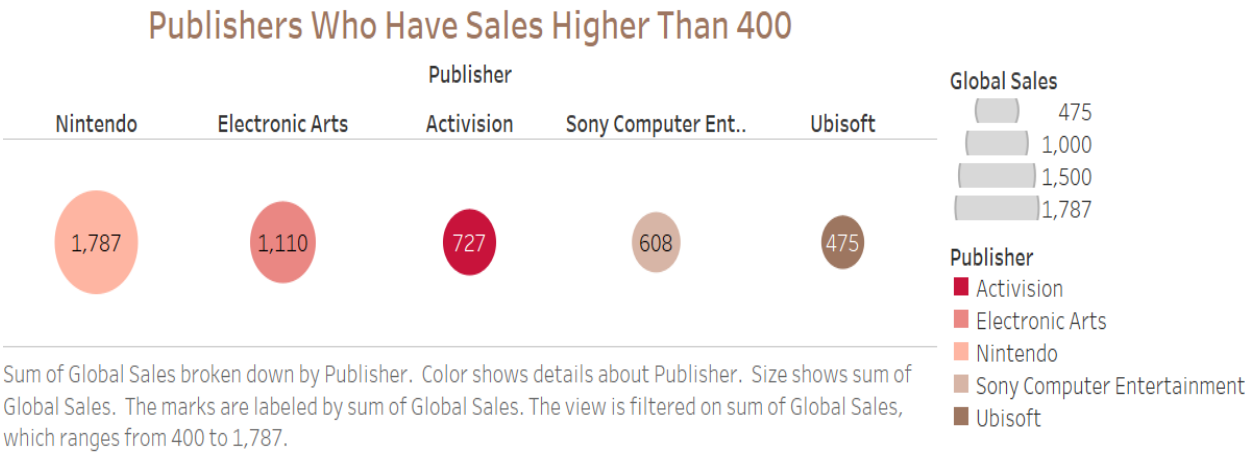
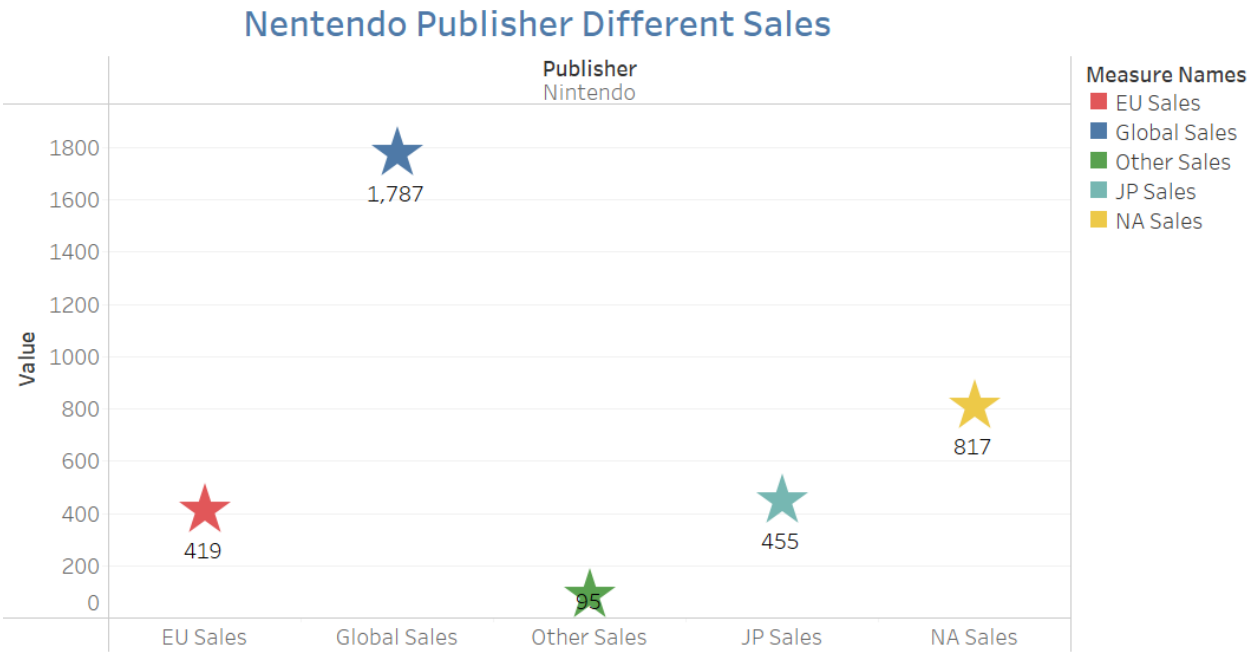


Figure 13

- **Last visualization shows that Nentendo has the highest sales in global sales.**
- **How much did Nentendo achieve in the different other sales?**



EU Sales, Global Sales, Other Sales, JP Sales and NA Sales for each Publisher. Color shows details about EU Sales, Global Sales, Other Sales, JP Sales and NA Sales. The marks are labeled by EU Sales, Global Sales, Other Sales, JP Sales and NA Sales. The view is filtered on Publisher, which keeps Nintendo.

- **Dashboard of Different Visualizations:**
- **Can we figure out different information sales (Global, EU, JP) sales of the Different Platforms?**

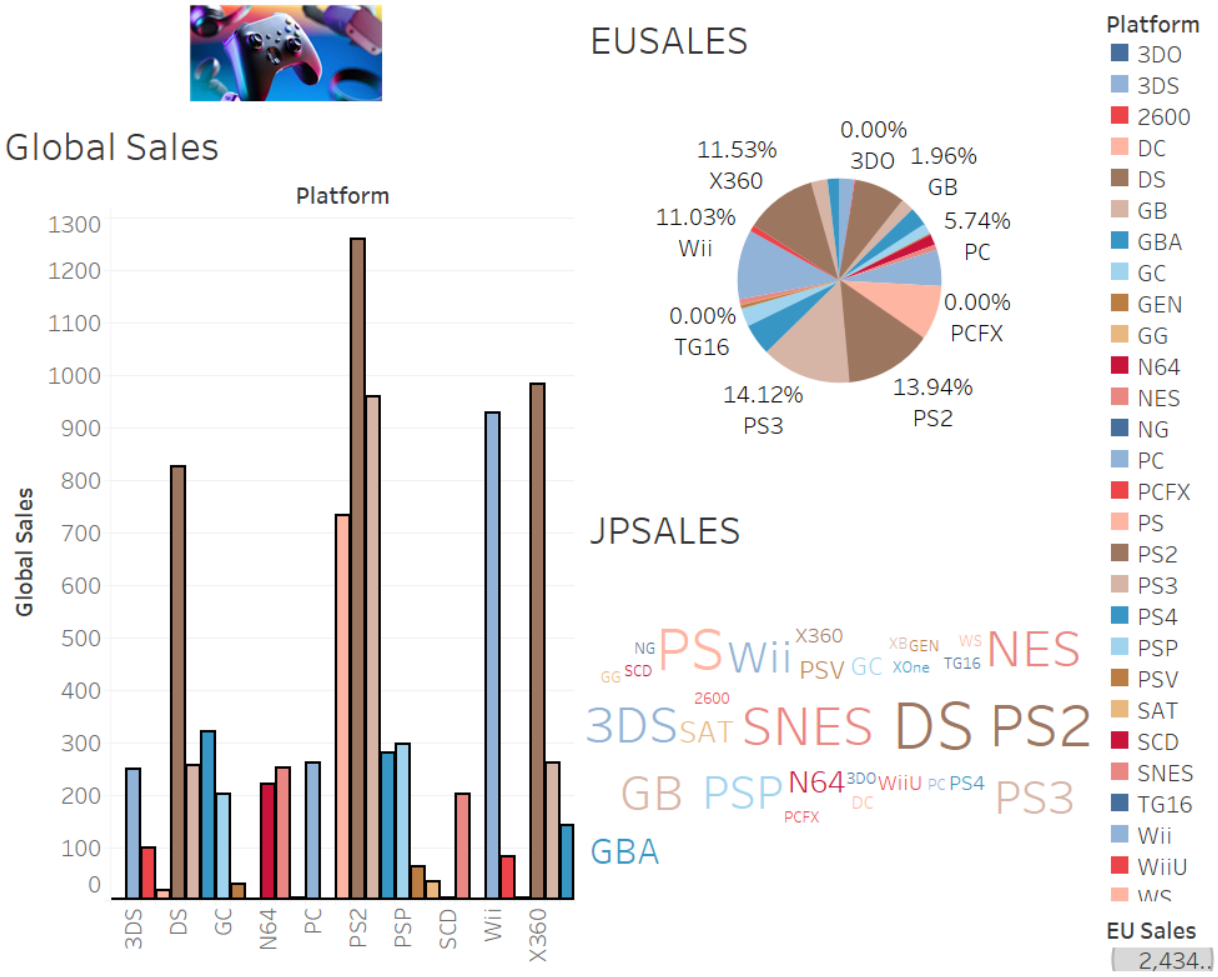
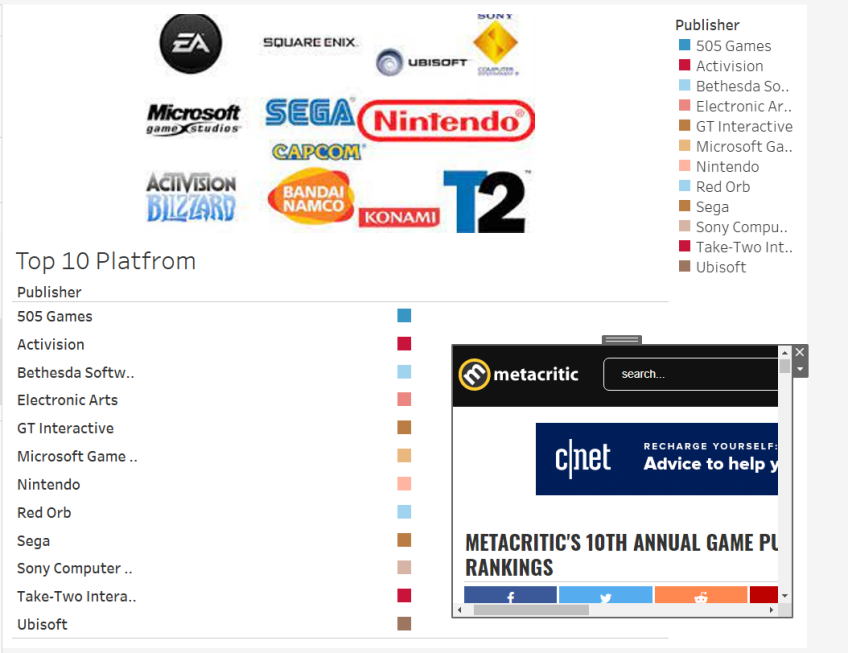


Figure 14

Yes, because of Dashboard feature we are able to figure out different sales of different platforms. We also can insert images as shown in figure 6.

- **Users need to get information about what they see in the visualizations.**
- **Can users get information about to 100 video games platforms?**



By using dashboard, I have added link to “Metacritic” website that has article about to 10 Platforms in the last 20 years. I also have added a Top 10 platform sheet. Also, in the top of the dashboard you can see an image of famous platforms.

References

S. (2020, March 9). *Goodreads-books*. Kaggle. https://www.kaggle.com/jealousleopard/goodreadsbooks