
Video Game Sales (2013-2016) Exploratory Analysis

Louis Bailey

▼ Introduction

The Data: This dataset, called Video Game Sales can be found on Kaggle. It contains information on video games with sales greater than 100,000 copies. Columns include

- Rank - Ranking of overall sales
- Name - The games name
- Platform - Platform of the games release (i.e. PC,PS4, etc.)
- Year_of_Release - 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.
- User_score - user scores from vgcharts.com
- Rating - Rating of the Game (e.g. E, T, M)

Project Summary: A video game developer is interested in what the market looks like. They want to make a game that is likely to be profitable. The goal will be to analyze this data to find what gamers want, where demands aren't being met, and how these demands vary in different places of the world.

The data that will be analyzed dates from 2013-2016. Currently this information is out-of-date. For example no one is

```
In [1]: ► import pandas as pd
import numpy as np

import sqlite3
import seaborn as sns
import matplotlib.pyplot as plt
```

▼ > Querying Database <

▼ connecting to database

```
In [2]: ► conn = sqlite3.connect('video_game_database.db')
c = conn.cursor()
```

```
In [3]: ► ▼ def is_opened(conn):
▼     try:
        conn.execute("SELECT * FROM sales_and_ratings LIMIT 1")
        return True
▼     except sqlite3.ProgrammingError as e:
        print("Connection closed {}".format(e))
        return False

is_opened(conn)
```

Out[3]: True

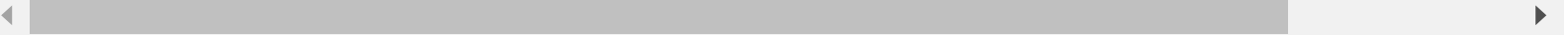
▼ selecting desired columns

```
In [4]: df = pd.read_sql_query(
    ...     SELECT Name, Platform, Year_of_Release, Genre, Publisher, NA_Sales, EU_Sales,
    ...           JP_Sales, Other_Sales, Global_Sales, User_Score, Rating
    ...     FROM sales_and_ratings
    ...     WHERE Year_of_Release IS NOT NULL AND Rating IS NOT NULL AND User_Score IS NOT NULL
    ... ,
    conn
)
```

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

```
Out[5]:
```

	Name	Platform	Year_of_Release	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
0	Wii Sports	Wii	2006.0	Sports	Nintendo	41.36	28.96	3.77	8.45	82.53
1	Mario Kart Wii	Wii	2008.0	Racing	Nintendo	15.68	12.76	3.79	3.29	35.52
2	Wii Sports Resort	Wii	2009.0	Sports	Nintendo	15.61	10.93	3.28	2.95	32.77
3	New Super Mario Bros.	DS	2006.0	Platform	Nintendo	11.28	9.14	6.50	2.88	29.80
4	Wii Play	Wii	2006.0	Misc	Nintendo	13.96	9.18	2.93	2.84	28.92



> Data Prep <

Clean

```
In [6]: ▶ print(f"Shape: {df.shape}\n{'-'*40}")
print(f"Missing:\n{df.isna().sum()}\n{'-'*40}")
print(f"Duplicates: {df.duplicated().sum()}")
```

Shape: (9707, 12)

Missing:

Name	0
Platform	0
Year_of_Release	0
Genre	0
Publisher	2
NA_Sales	0
EU_Sales	0
JP_Sales	0
Other_Sales	0
Global_Sales	0
User_Score	0
Rating	0
dtype: int64	

Duplicates: 0

▼ Filter Data 2013-2016

```
In [7]: ▶ df_modern = df[(df.Year_of_Release>=2013) & (df.Year_of_Release!=2020)]  
df_modern.shape
```

```
Out[7]: (1245, 12)
```

▼ Datatypes

```
In [8]: ▶ df_modern.dtypes
```

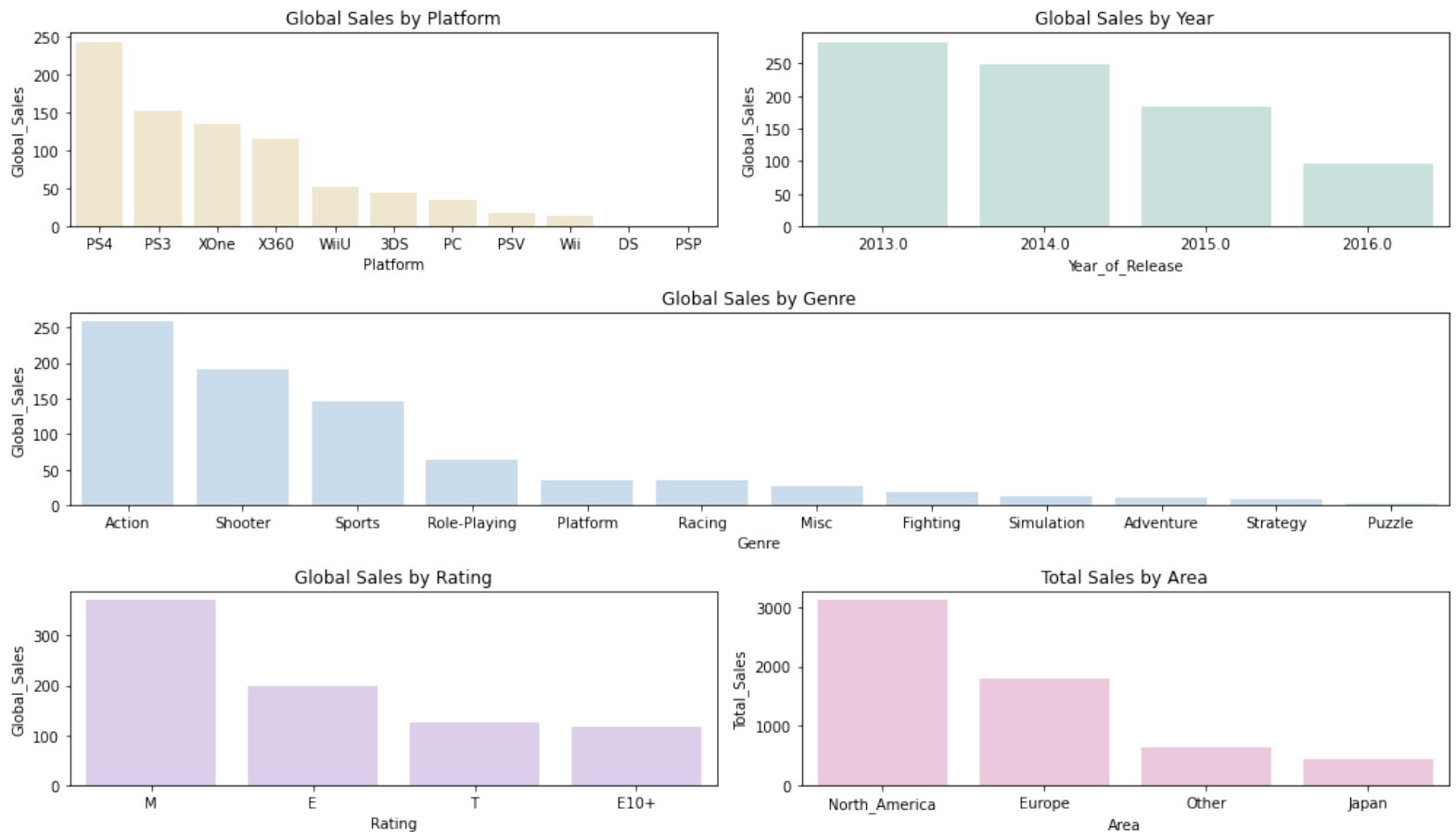
```
Out[8]: Name          object  
Platform          object  
Year_of_Release   float64  
Genre             object  
Publisher          object  
NA_Sales           float64  
EU_Sales           float64  
JP_Sales           float64  
Other_Sales        float64  
Global_Sales       float64  
User_Score         object  
Rating             object  
dtype: object
```

```
In [9]: ▶ ◀ #change User_Score to float  
df_modern.loc[:, 'User_Score'] = pd.to_numeric(df_modern['User_Score'], errors='coerce')
```

▼ **> Big Picture on Sales <**

In [10]: ▶ #↔

In [11]: ▶ #↔



Play Station is the best selling platform followed by Xbox and then Wii.

Global sales appear to be steadily declining from 2013 to 2016.

Best selling genres appear to be action, shooters, and then sports.

M rated games appear to sell way more than other ratings.

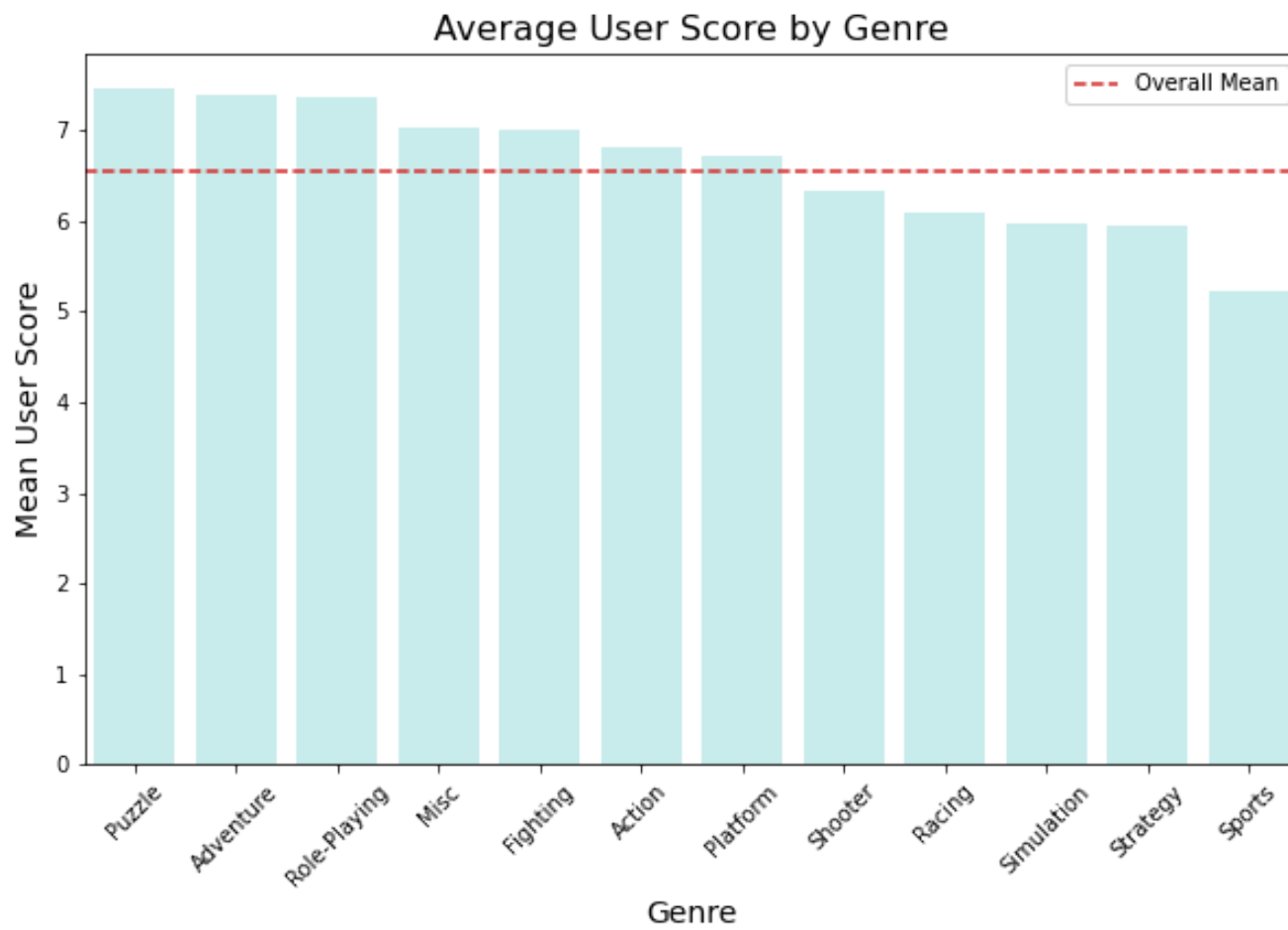
North America appears to buy about as many games as the rest of the world combined.



> Average User Score by Genre <

In [12]:

▶ #↔



It appears demands are not being met regarding both shooters and sports games. This is noteworthy because these genres are in the top 3 of most bought games.

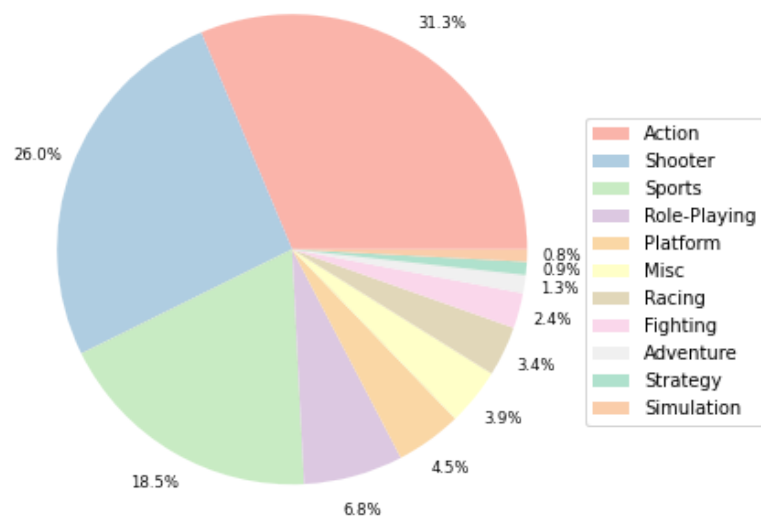
> Percent of Sales by Genre and Area <

percentages > 0.6%

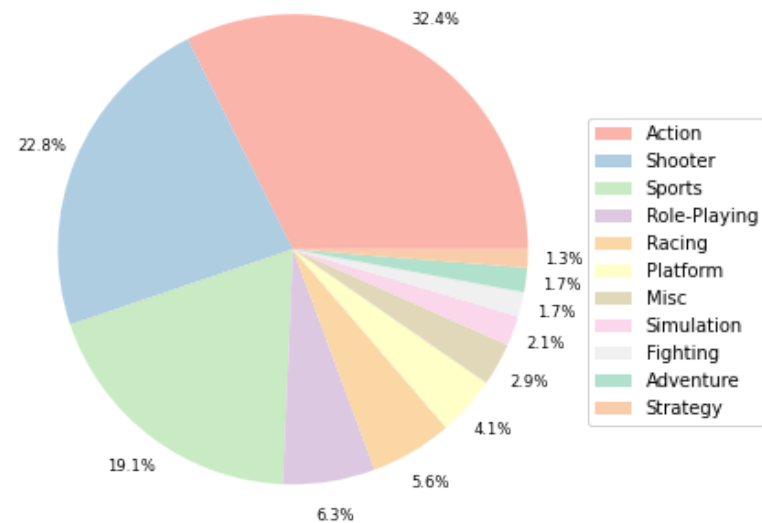
In [13]: ▶ #↔

In [14]: ▶ #↔

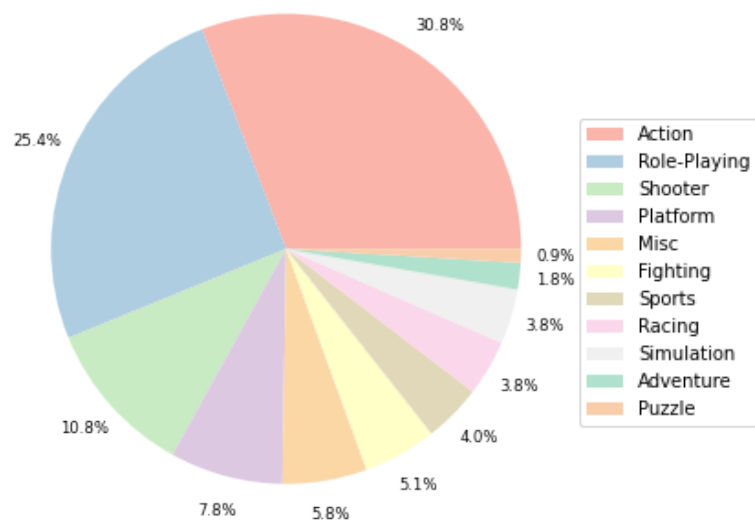
Percent of Sales by Genre North America



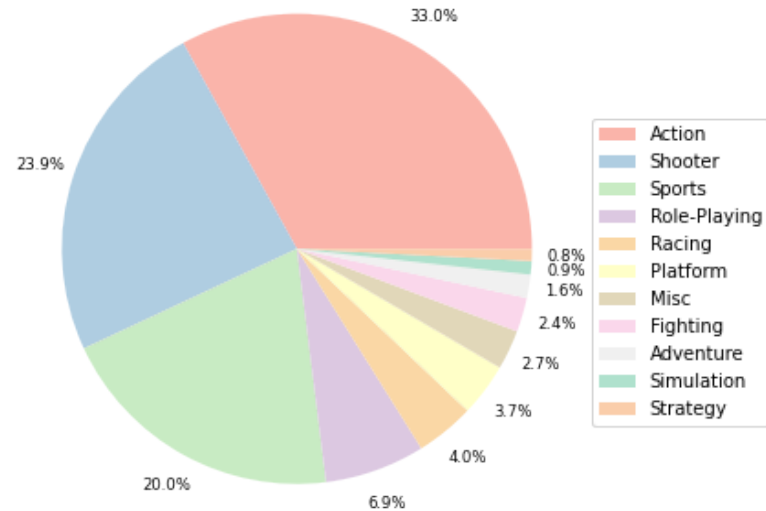
Percent of Sales by Genre Europe



Percent of Sales by Genre Japan



Percent of Sales by Genre Other



In all of the areas except Japan, the highest selling genres are action, shooter, and then sports. In Japan they are action, role-playing, and then shooters.

▼ > Percent of Sales by Platform and Area <

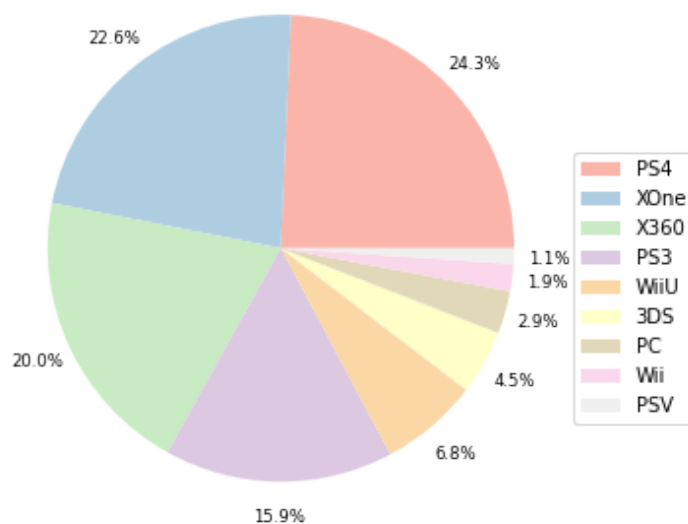
Percentages > 0.3%

In [15]:

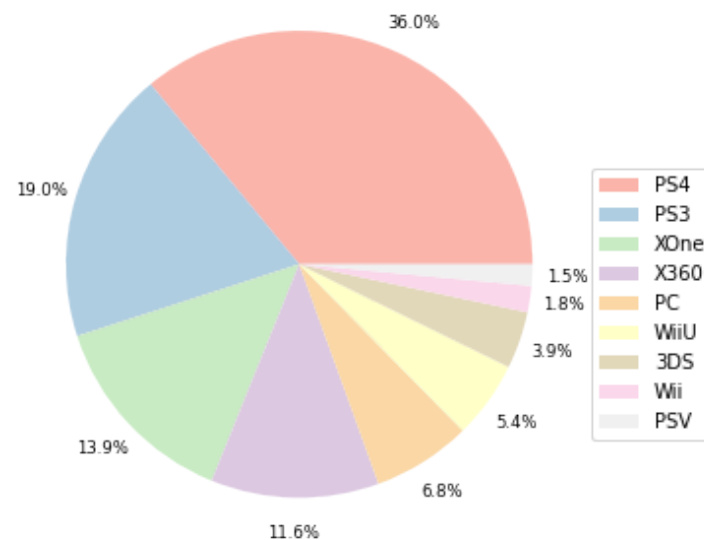
▶ #↔

In [16]: ▶ #↔

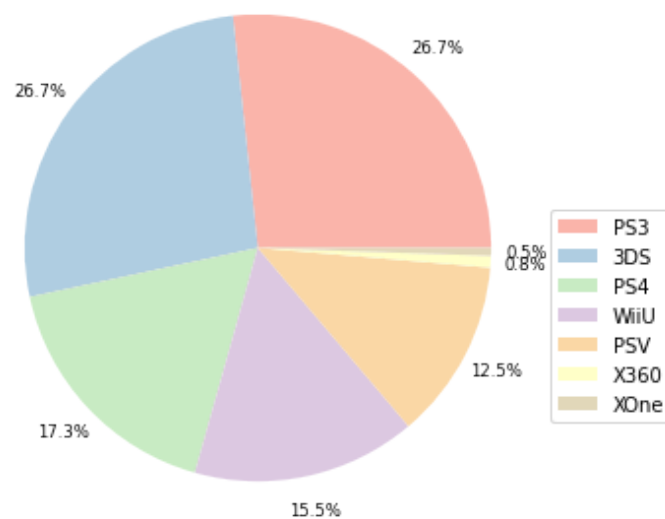
Percent of Sales by Genre North America



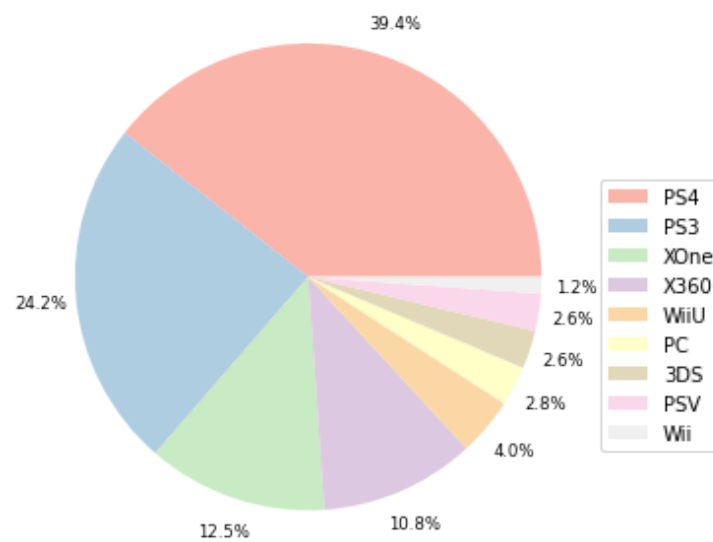
Percent of Sales by Genre Europe



Percent of Sales by Genre Japan



Percent of Sales by Genre Other



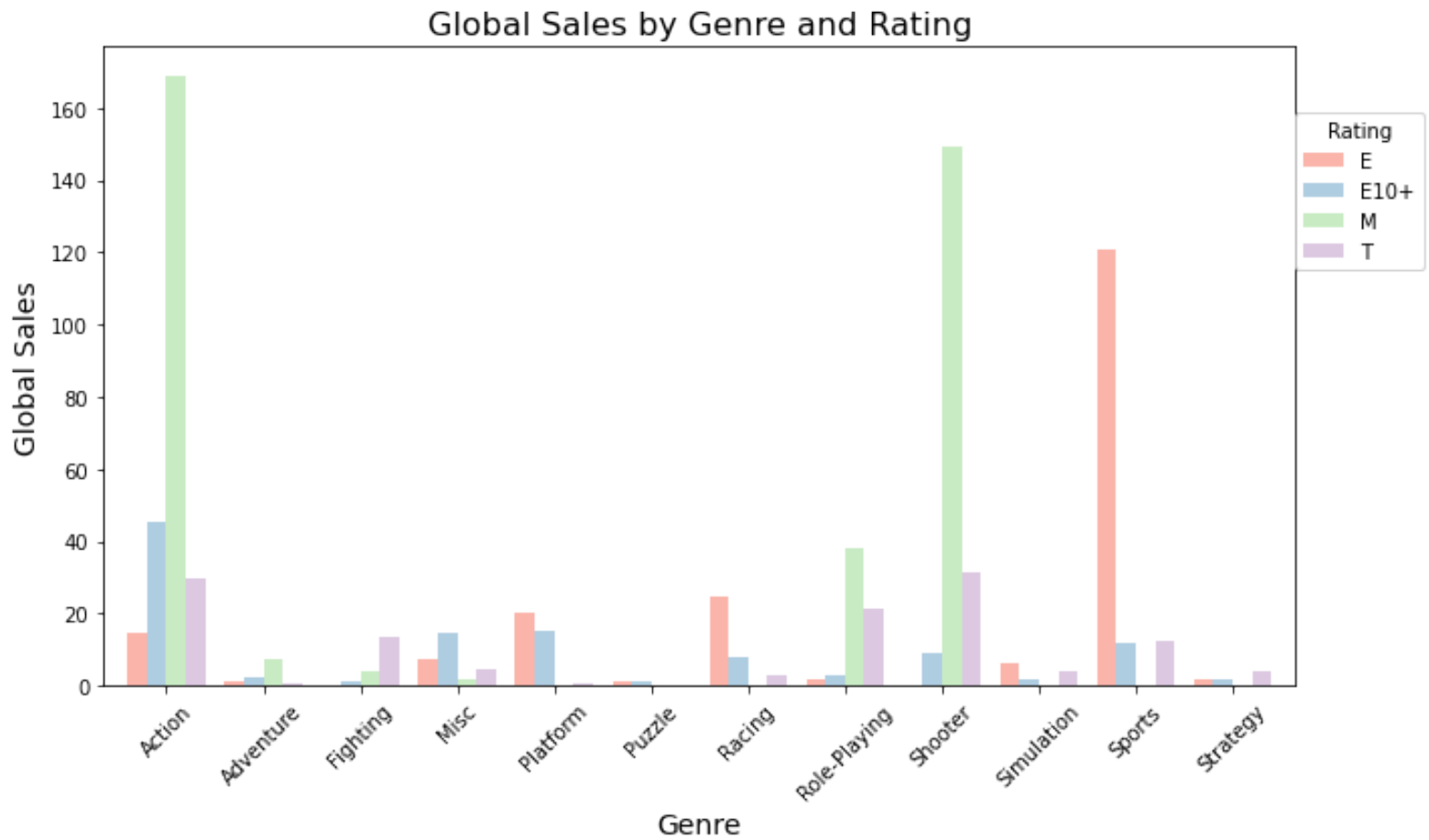
In Europe and Other, most sales are for PS4, PS3, Xbox One, and then Xbox 360.

In North America, most of the sales are for PS4, Xbox One, Xbox 360, and then PS3.

In Japan, most of the sales are for PS3, 3DS, PS4, and then WiiU.

▼ > Global Sales by Genre and Rating <

In [17]: ▶ #↔



For action the highest selling rating is M.

For shooters the highest selling rating is M.

For sports the highest selling rating is E.

For role-playing the highest selling rating is M.

For platform the highest selling rating is E.

▼ Closing Connection to Database

```
In [18]: ► conn.close()  
is_opened_status = is_opened(conn)  
print("Is the connection open after closing? ", is_opened_status)
```

Connection closed Cannot operate on a closed database.
Is the connection open after closing? False

▼ > Conclusion <

- My conclusion from analyzing this data is that making a shooter or sports game would be a good idea. Shooters and sports games are the 2nd and 3rd most bought globally, and from the mean user scores it appears that demands are not being met for these genres.
- These genres are also the 2nd and third most bought in every area in this dataset except for Japan (where shooters are 3rd and sports are 7th).

- The game should definitely be made available for both Xbox and PlayStation as these are the most popular platforms (except for Japan where the most popular are Playstation, 3DS, and Wii).
- If it is a shooter it should be rated M, and if it is a sports game it should be rated E as this is what most people buy for these genres.