
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 Currently this information is out-of-date.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 buying PS3s or Xbox 360s anymore. However, for this project, this data will be treated as if it is up-to-date and relevant.

```
In [11]: ▶ 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 [12]: ▶ conn = sqlite3.connect('video_game_database.db')
c = conn.cursor()
```

```
In [13]: ▶ ▼ 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[13]: True

selecting desired columns

```
In [14]: 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 AND Year_of_Release >= 2013
              AND Year_of_Release != 2020
    ''',
    conn
)
```

```
In [15]: df.head()
```

Out[15]:

	Name	Platform	Year_of_Release	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales	User_Score	Rating
0	Grand Theft Auto V	PS3	2013.00	Action	Take-Two Interactive	7.02	9.09	0.98	3.96	21.04	8.2	M
1	Grand Theft Auto V	X360	2013.00	Action	Take-Two Interactive	9.66	5.14	0.06	1.41	16.27	8.1	M
2	Grand Theft Auto V	PS4	2014.00	Action	Take-Two Interactive	3.96	6.31	0.38	1.97	12.61	8.3	M
3	Call of Duty: Ghosts	X360	2013.00	Shooter	Activision	6.73	2.56	0.04	0.91	10.25	2.6	M
4	Call of Duty: Ghosts	PS3	2013.00	Shooter	Activision	4.10	3.63	0.38	1.25	9.36	2.6	M

> Data Prep <

▼ Clean

```
In [16]: ▶ print(f"Shape: {df.shape}\n{'-'*40}")
          print(f"Missing:\n{df.isna().sum()}\n{'-'*40}")
          print(f"Duplicates: {df.duplicated().sum()}")
```

Shape: (1245, 12)

Missing:

Name	0
Platform	0
Year_of_Release	0
Genre	0
Publisher	0
NA_Sales	0
EU_Sales	0
JP_Sales	0
Other_Sales	0
Global_Sales	0
User_Score	0
Rating	0

dtype: int64

Duplicates: 0

▼ Datatypes

```
In [17]: ▶ df.dtypes
```

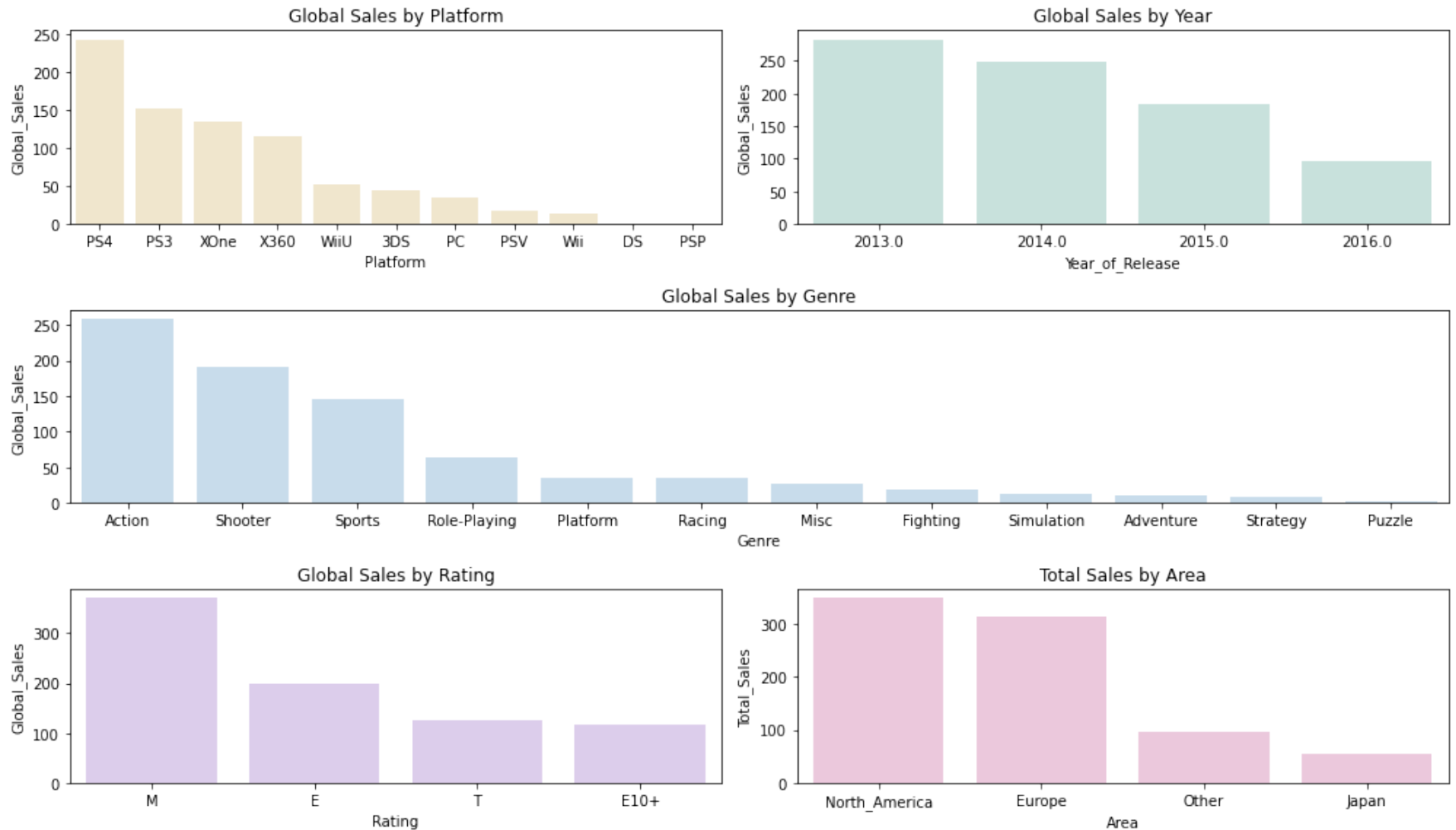
```
Out[17]: Name          object
          Platform      object
          Year_of_Release float64
          Genre         object
          Publisher      object
          NA_Sales       float64
          EU_Sales       float64
          JP_Sales       float64
          Other_Sales    float64
          Global_Sales   float64
          User_Score     object
          Rating         object
          dtype: object
```

```
In [18]: #change User_Score to float
df.loc[:, 'User_Score'] = pd.to_numeric(df['User_Score'], errors='coerce')
```

> Big Picture on Sales <

```
In [19]: #↔
```

```
In [20]: #↔
```



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

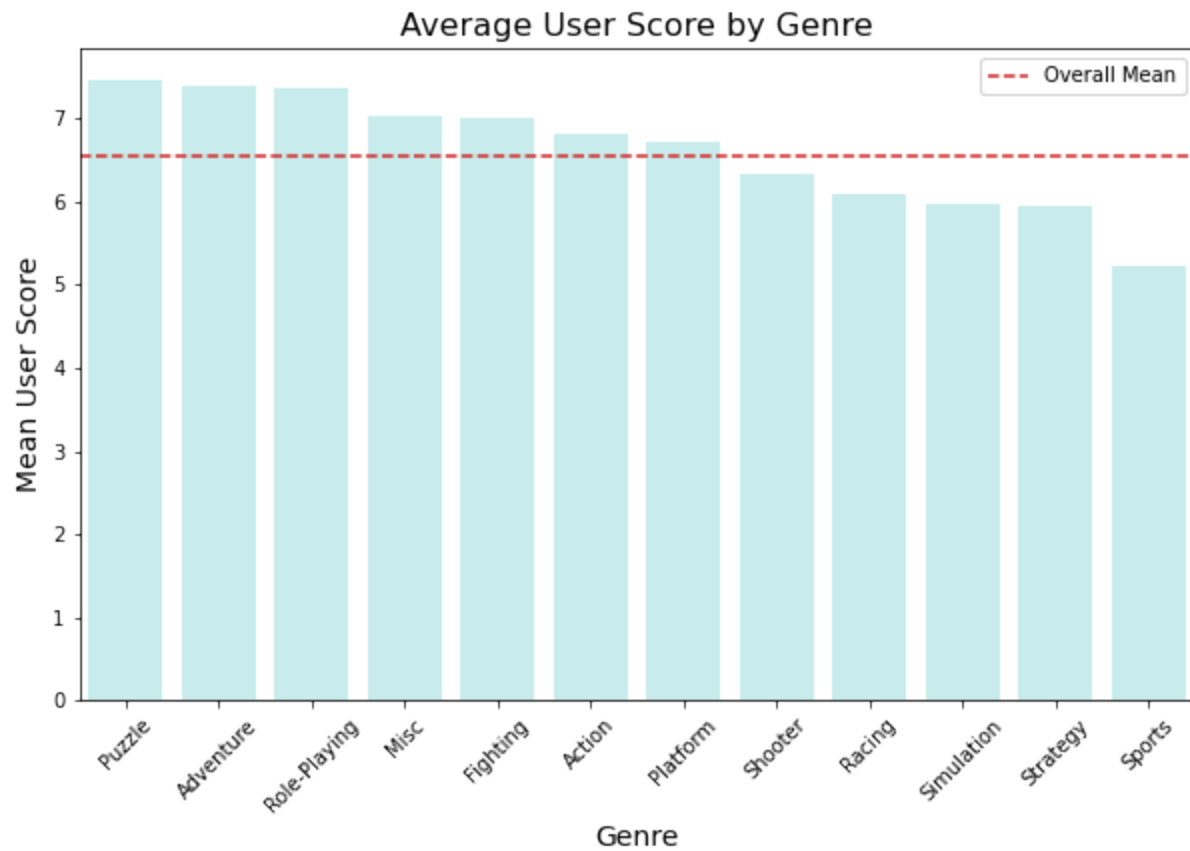
Global sales appear to be steadily declining from 2013 to 2016.

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 <



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.



> Pectent of Sales by Genre and Area <

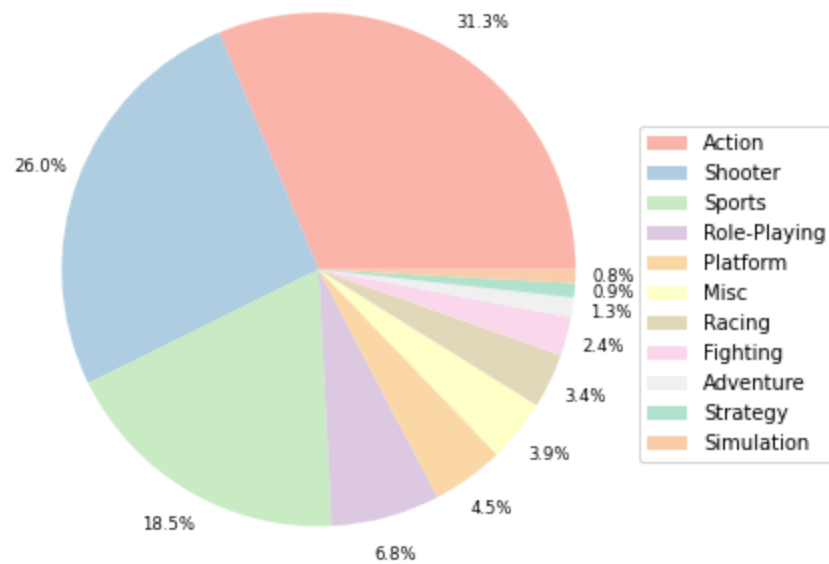
percentages > 0.6%

In [22]:

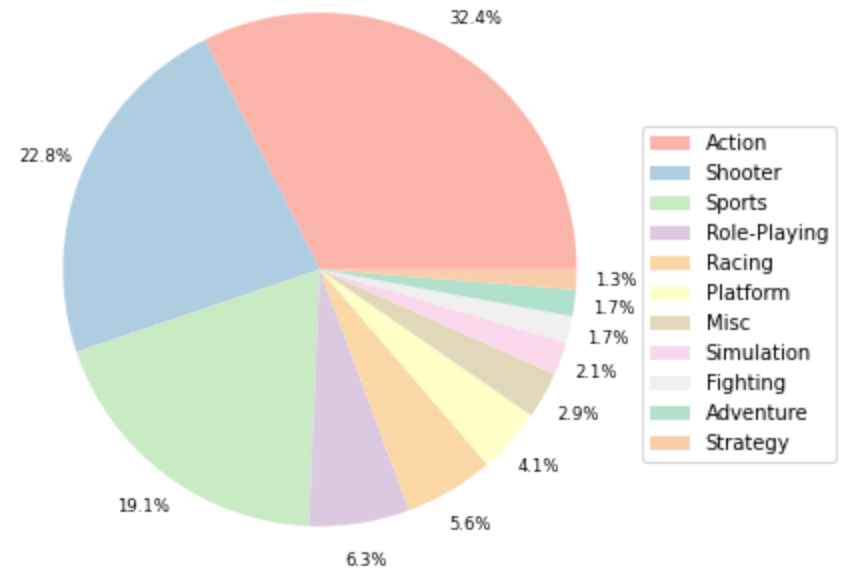


#↔

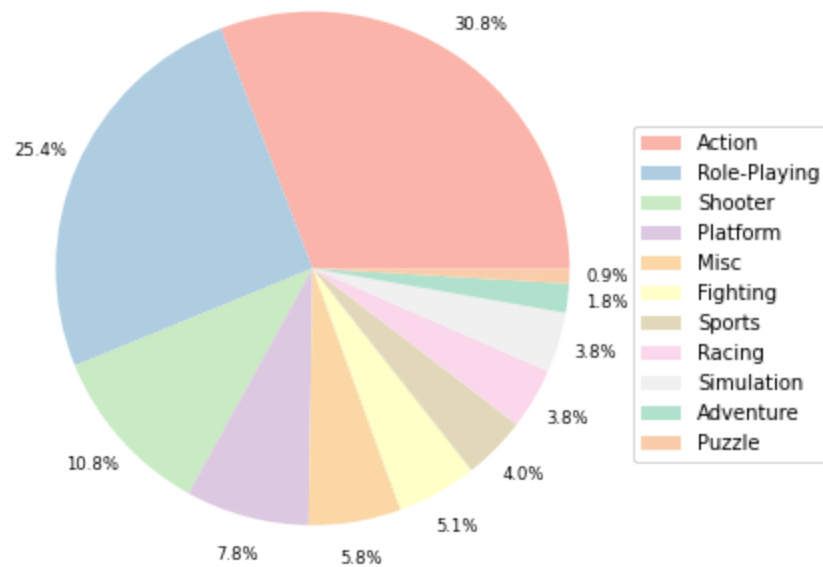
Percent of Sales by Genre North America



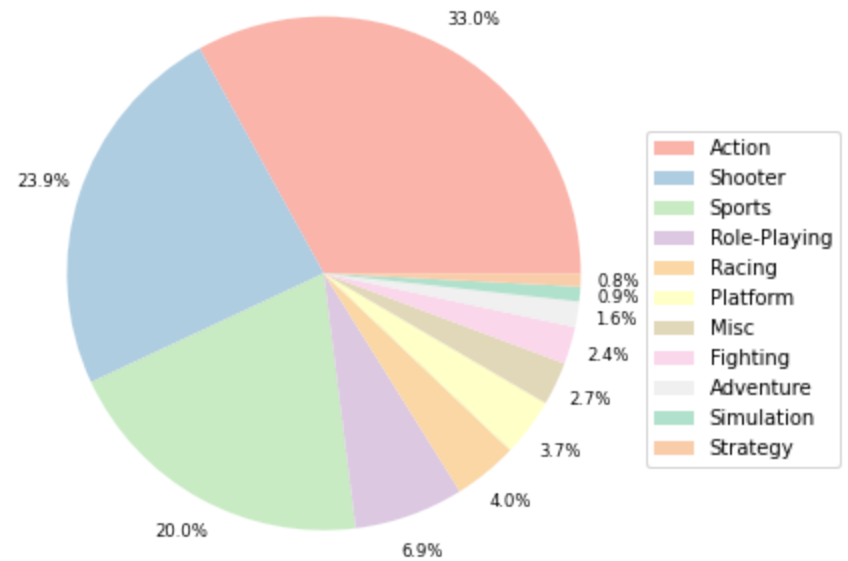
Percent of Sales by Genre Europe



Percent of Sales by Genre Japan



Percent of Sales by Genre Other



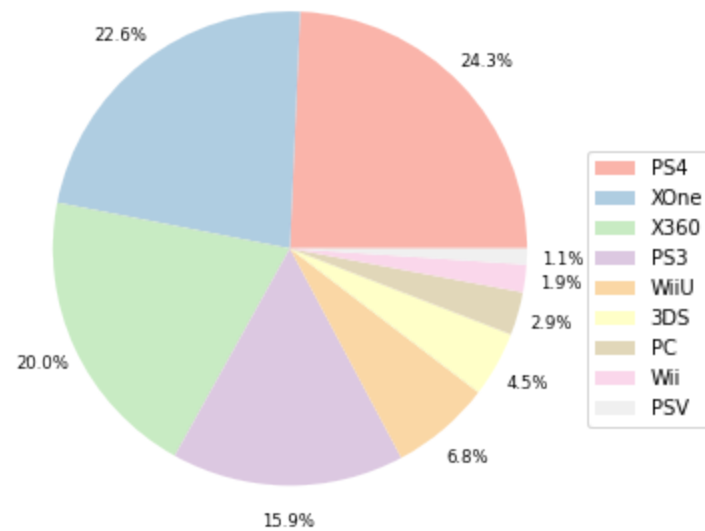
In all of the areas except Japan, the highest selling genres are action, shooter, and then sports. In Japan they are action, role-playing, and then shooters.

> Percent of Sales by Platform and Area <

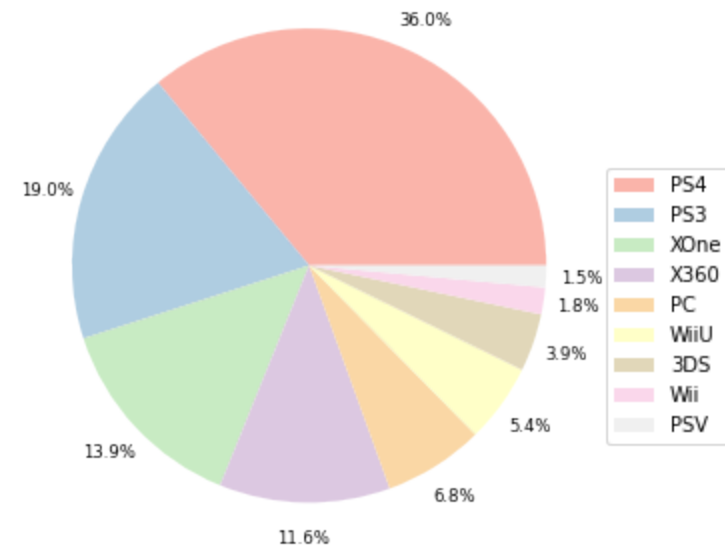
Percentages > 0.3%

In [25]: ▶ #↔

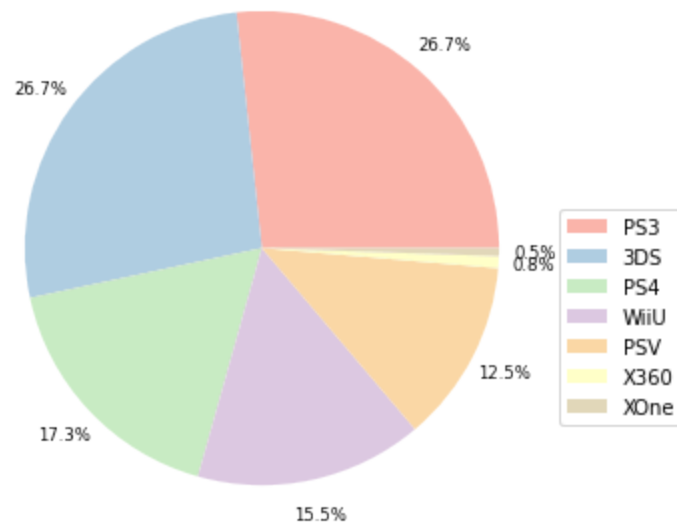
Percent of Sales by Platform North America



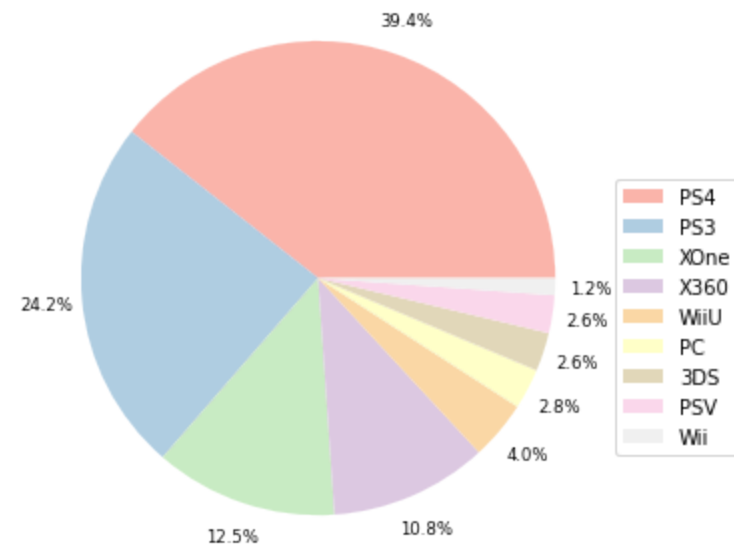
Percent of Sales by Platform Europe



Percent of Sales by Platform Japan



Percent of Sales by Platform Other



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

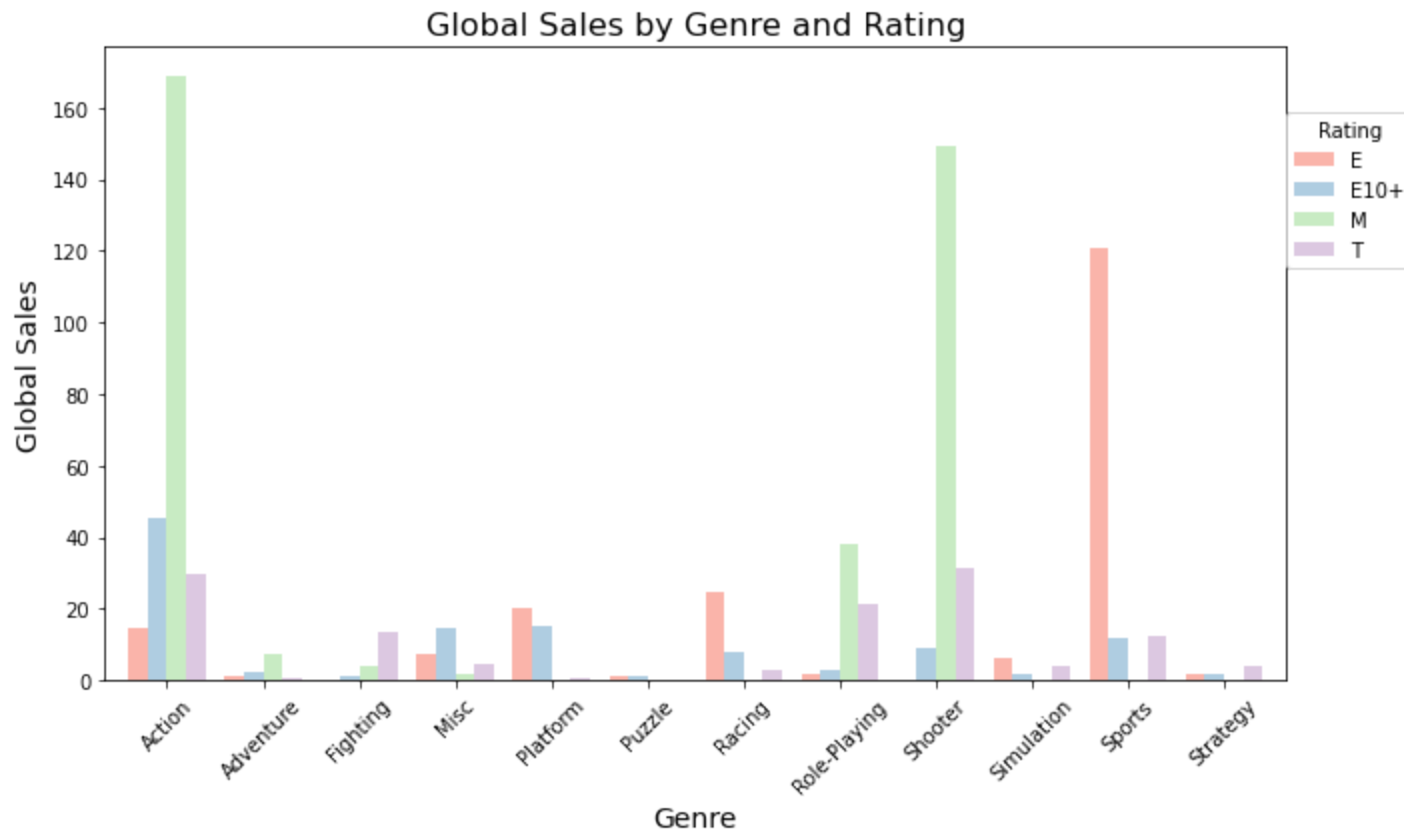
In North America, most of the sales are for PS4, Xbox One, Xbox 360, and then PS3.

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

> Global Sales by Genre and Rating <

In [28]:

#↔



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 [29]: ► 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.

