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

- <u>Rank</u> - Ranking of overall sales
- <u>Name</u> - The games name
- <u>Platform</u> - Platform of the games release (i.e. PC,PS4, etc.)
- <u>Year_of_Release</u> - Year of the game's release
- <u>Genre</u> - Genre of the game
- <u>Publisher</u> - Publisher of the game
- <u>NA_Sales</u> - Sales in North America (in millions)
- <u>EU_Sales</u> - Sales in Europe (in millions)
- <u>JP_Sales</u> - Sales in Japan (in millions)
- <u>Other_Sales</u> - Sales in the rest of the world (in millions)
- <u>Global_Sales</u> - Total worldwide sales.
- <u>User_score</u> - user scores from vgcharts.com
- <u>Rating</u> - 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.

In [1]:
```python
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]:
```python
conn = sqlite3.connect('video_game_database.db')
c = conn.cursor()
```

In [3]:
```python
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	User_
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
```
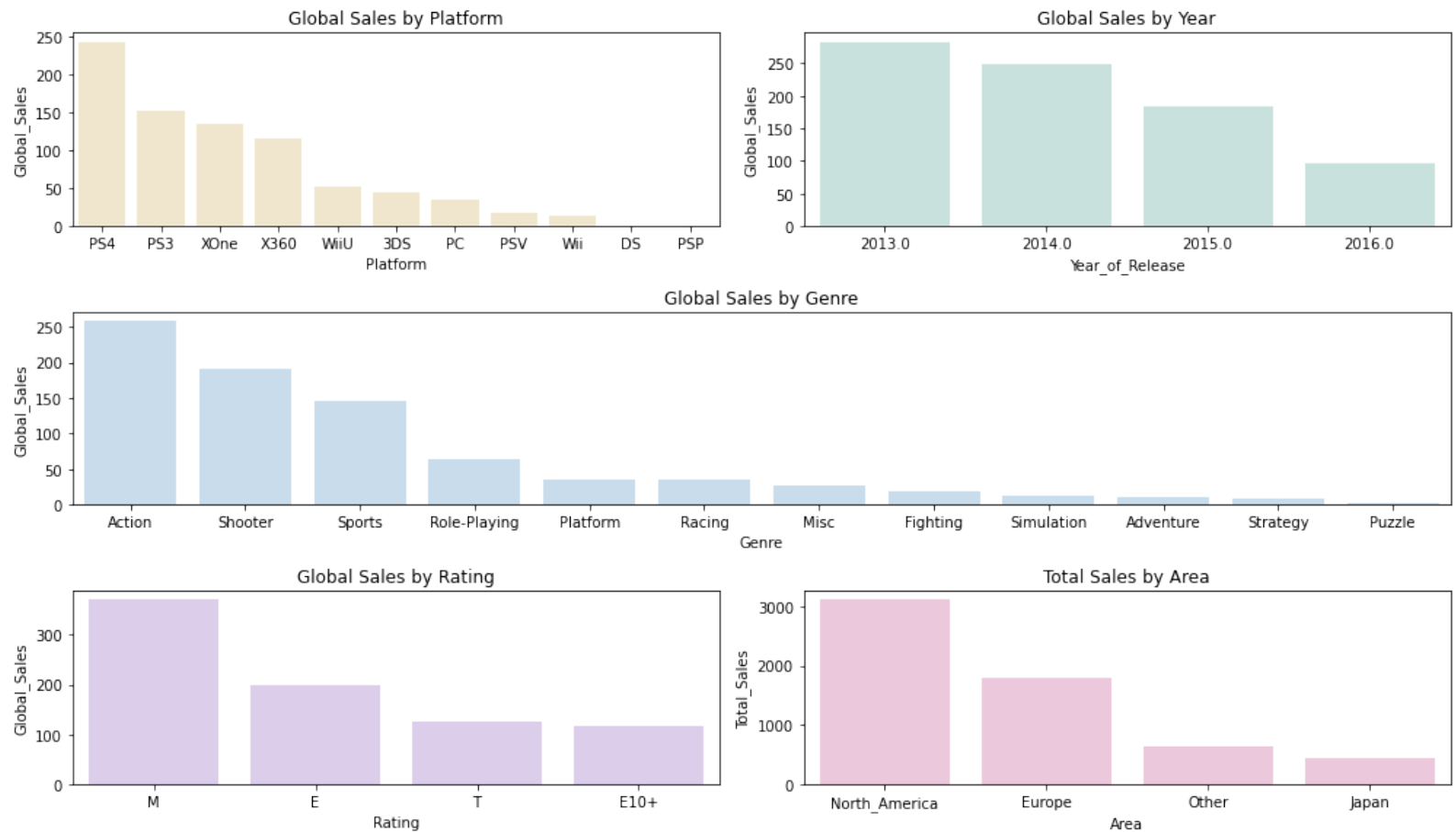
```
In [9]:  #change User_Score to float
         df_modern.loc[:, 'User_Score'] = pd.to_numeric(df_modern['User_Score'], errors='coerce')
```

> Big Picture on Sales <

```
In [10]:  #↔
```

In [11]:   ▶| ▶   #↔

Global Sales by Platform

Global Sales by Year

Global Sales by Genre

Global Sales by Rating

Total Sales by Area

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

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

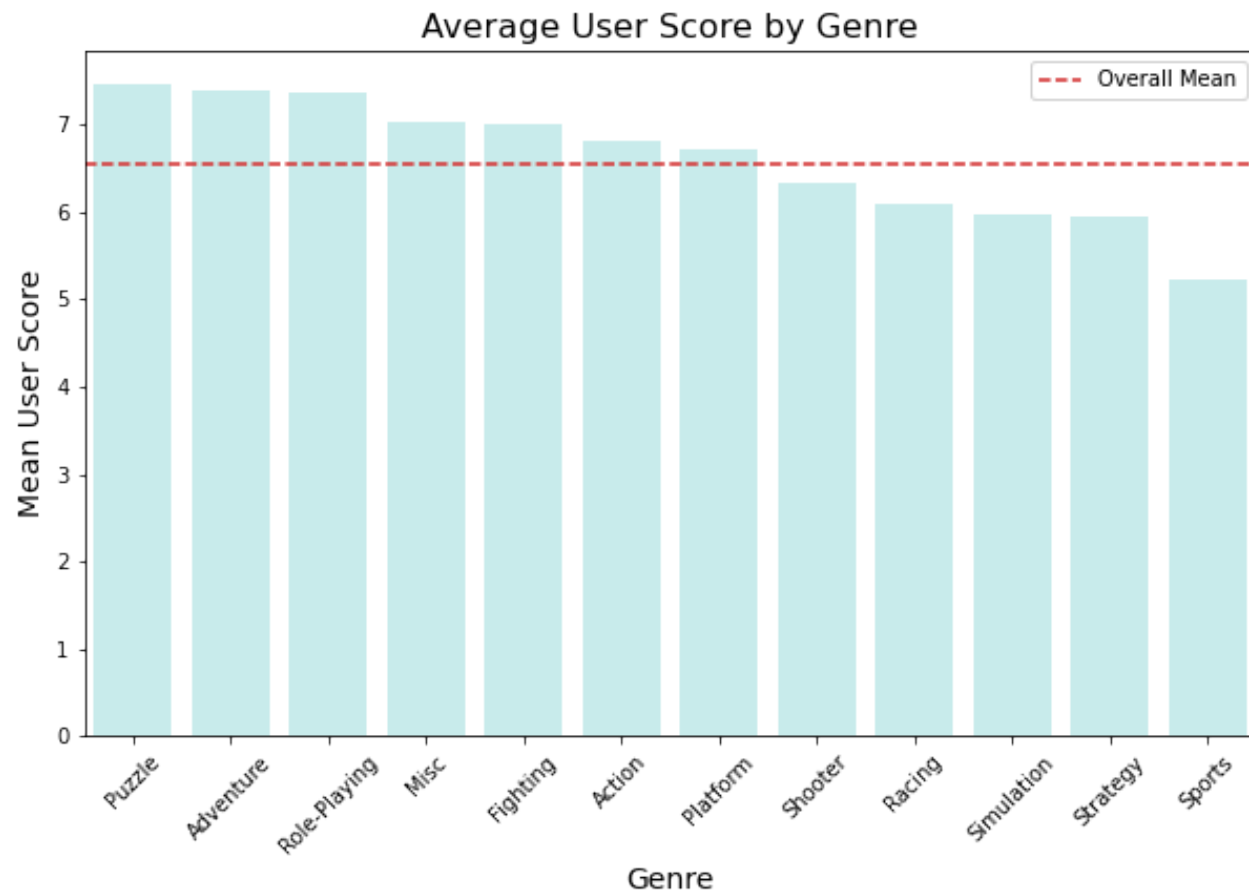Best selling genres appear to be action, shooters, and then sports.

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 <

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.

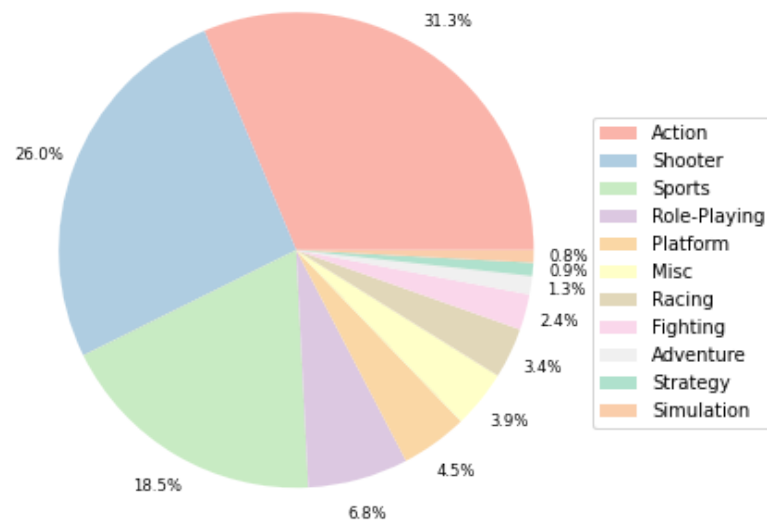# > Pecent of Sales by Genre and Area <
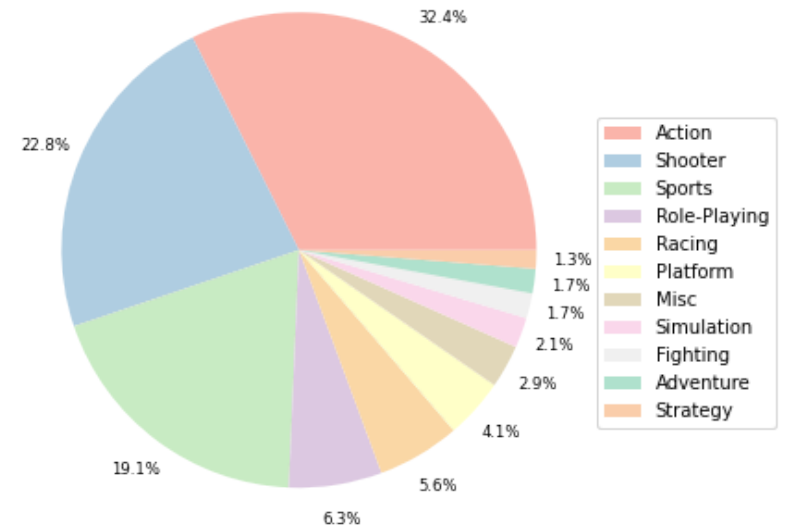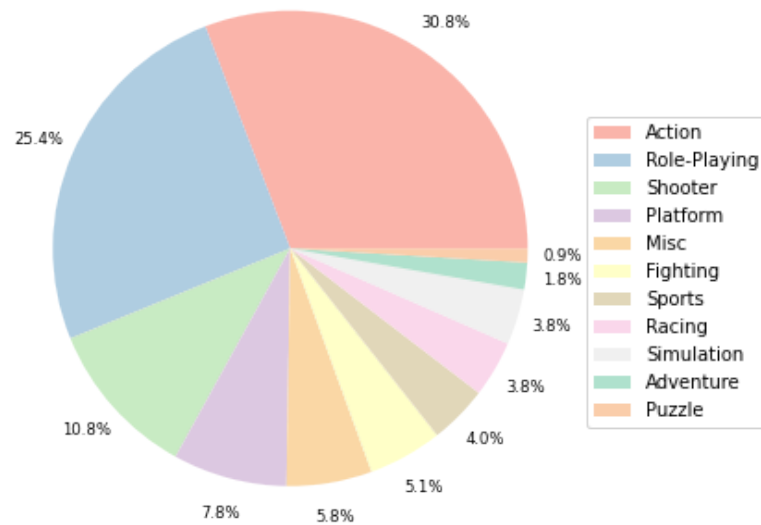
percentages > 0.6%

In [13]: #↔

In [14]: ▶ #↔
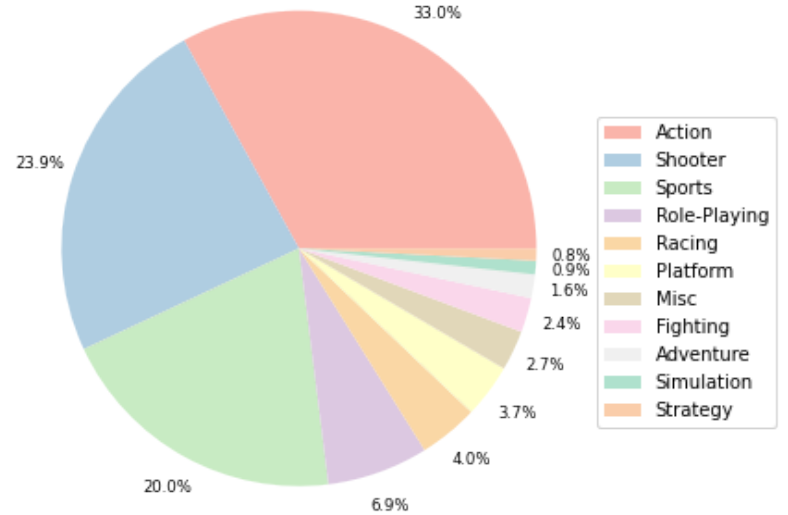
Percent of Sales by Genre North America

31.3%

26.0%

0.8%
0.9%
1.3%
2.4%

3.4%

3.9%

18.5%

4.5%

6.8%

	Action
	Shooter
	Sports
	Role-Playing
	Platform
	Misc
	Racing
	Fighting
	Adventure
	Strategy
	Simulation

Percent of Sales by Genre Europe

32.4%

22.8%

1.3%
1.7%
1.7%

2.1%

2.9%

19.1%

4.1%

5.6%

6.3%

	Action
	Shooter
	Sports
	Role-Playing
	Racing
	Platform
	Misc
	Simulation
	Fighting
	Adventure
	Strategy

Percent of Sales by Genre Japan

30.8%

25.4%

0.9%
1.8%

3.8%

3.8%

10.8%

4.0%

7.8%

5.1%

5.8%

	Action
	Role-Playing
	Shooter
	Platform
	Misc
	Fighting
	Sports
	Racing
	Simulation
	Adventure
	Puzzle

Percent of Sales by Genre Other

33.0%

23.9%

0.8%
0.9%
1.6%

2.4%

2.7%

20.0%

3.7%

6.9%

4.0%

	Action
	Shooter
	Sports
	Role-Playing
	Racing
	Platform
	Misc
	Fighting
	Adventure
	Simulation
	Strategy

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.
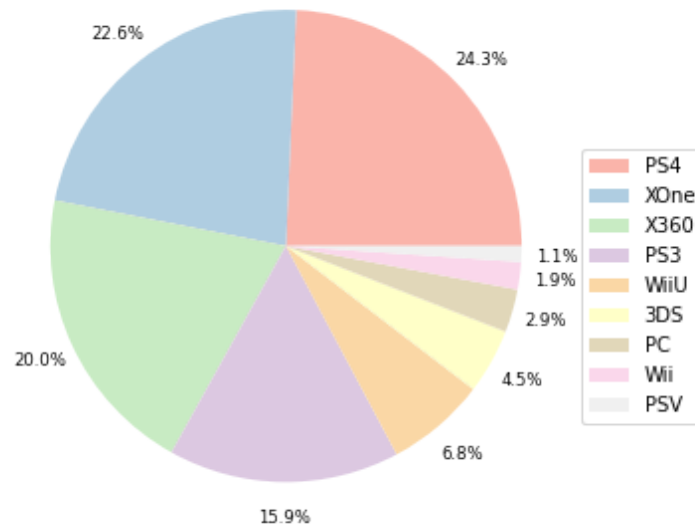
> Percent of Sales by Platform and Area <
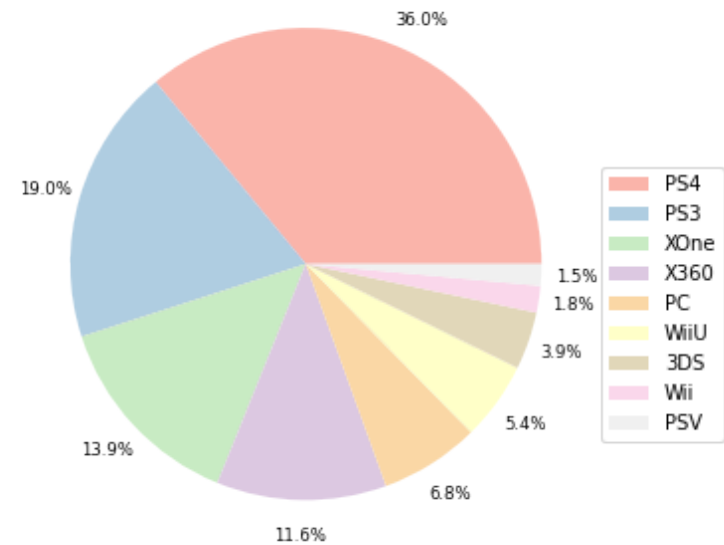
Percentages > 0.3%

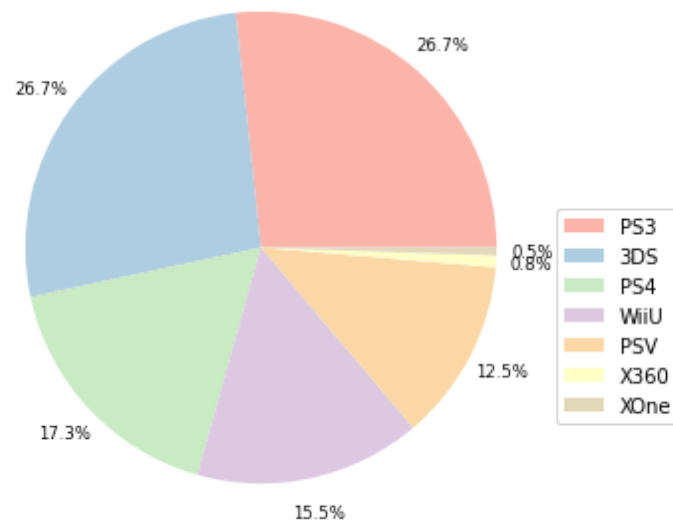In [15]: ▶| ▸ `#↔`

Percent of Sales by Genre North America

PS4	
XOne	
X360	
PS3	
WiiU	
3DS	
PC	
Wii	
PSV	

24.3%
22.6%
20.0%
15.9%
6.8%
4.5%
2.9%
1.9%
1.1%

Percent of Sales by Genre Europe

PS4	
PS3	
XOne	
X360	
PC	
WiiU	
3DS	
Wii	
PSV	

36.0%
19.0%
13.9%
11.6%
6.8%
5.4%
3.9%
1.8%
1.5%

Percent of Sales by Genre Japan

PS3	
3DS	
PS4	
WiiU	
PSV	
X360	
XOne	

26.7%
26.7%
17.3%
15.5%
12.5%
0.8%
0.5%

Percent of Sales by Genre Other

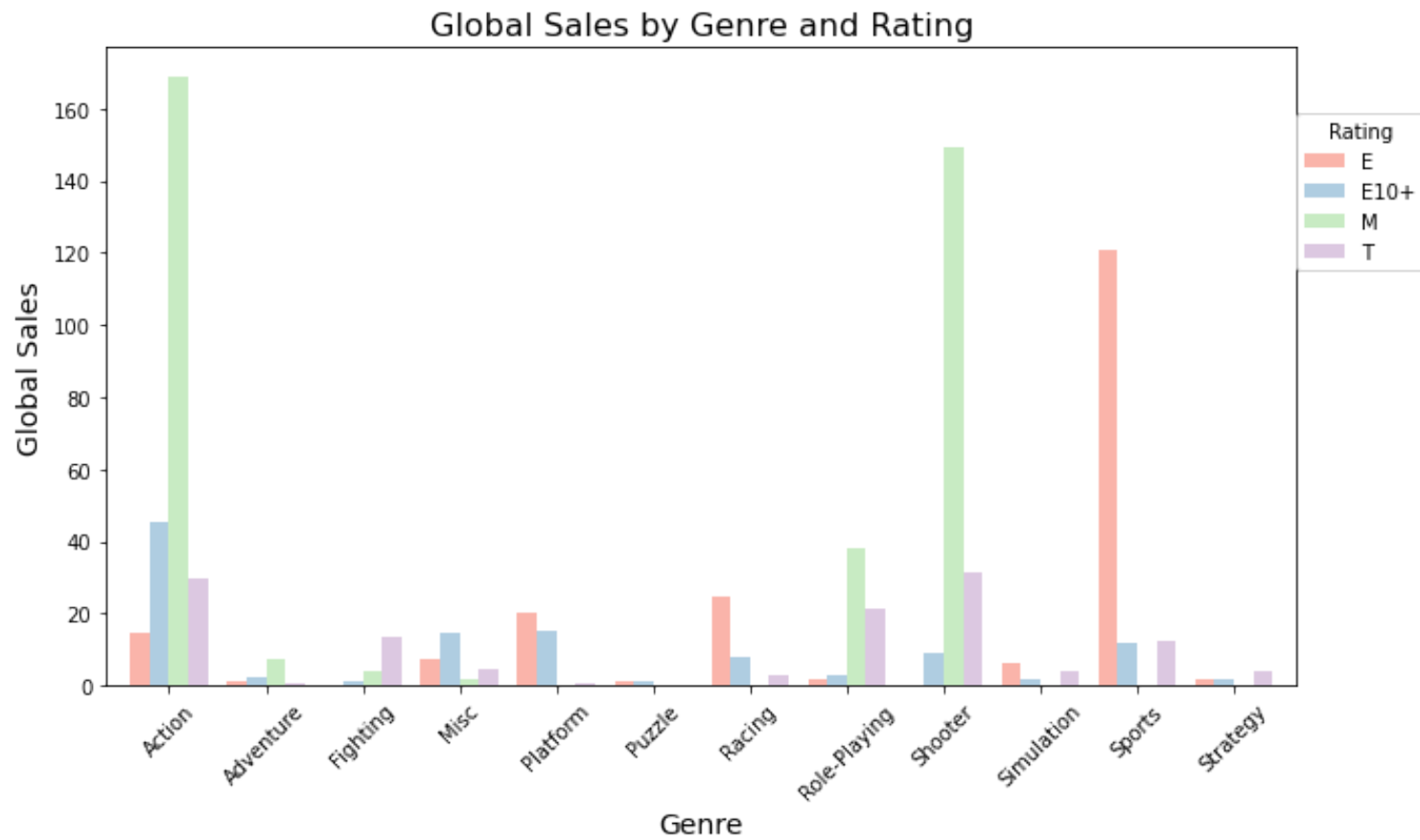|	
PS4	
PS3	
XOne	
X360	
WiiU	
PC	
3DS	
PSV	
Wii	

39.4%
24.2%
12.5%
10.8%
4.0%
2.8%
2.6%
2.6%
1.2%
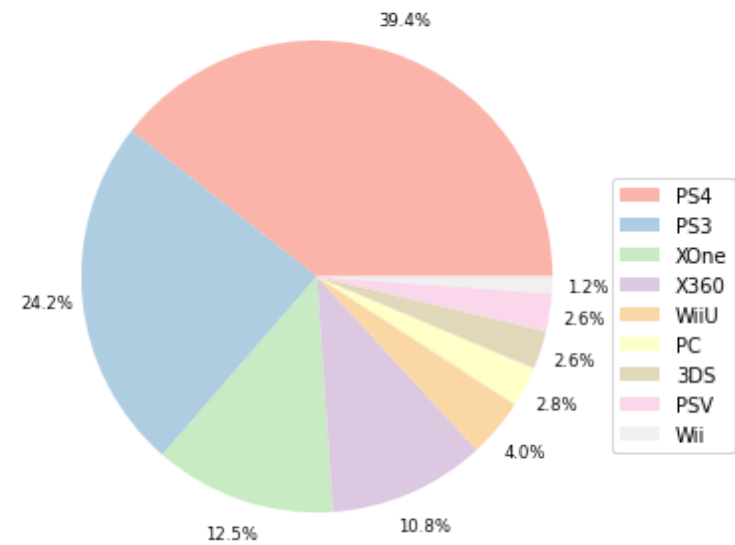
In Europe and Other, most sales are for PS4, PS3, Xbox One, and then Xbox 360.

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 <**

Global Sales by Genre and Rating

For action the highest selling rating is M.

For shooters the highest selling rating is M.

For sports the highest selling reting 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]: ▶
```python
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 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.