19CSE305

MACHINE LEARNING

Roll_nos:

1. CB.EN.U4CSE21123
2. CB.EN.U4CSE21156
3. CB.EN.U4CSE21159
4. CB.EN.U4CSE21169

Url of the Dataset: https://www.kaggle.com/datasets/ibriiee/video-games-sales-dataset-2022-updated-extra-feat

Dataset description (attribute details):

```
In [1]: import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt
        import seaborn as sns

        df=pd.read_csv('Video_Games.csv')
        df
```

Out[1]:

	Name	Platform	Year_of_Release	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales	Critic_Score	Critic_Count	User_Score	User_Count	Developer	Rating
0	Wii Sports	Wii	2006.0	Sports	Nintendo	41.36	28.96	3.77	8.45	82.53	76.0	51.0	8	322.0	Nintendo	E
1	Super Mario Bros.	NES	1985.0	Platform	Nintendo	29.08	3.58	6.81	0.77	40.24	NaN	NaN	NaN	NaN	NaN	NaN
2	Mario Kart Wii	Wii	2008.0	Racing	Nintendo	15.68	12.76	3.79	3.29	35.52	82.0	73.0	8.3	709.0	Nintendo	E
3	Wii Sports Resort	Wii	2009.0	Sports	Nintendo	15.61	10.93	3.28	2.95	32.77	80.0	73.0	8	192.0	Nintendo	E
4	Pokemon Red/Pokemon Blue	GB	1996.0	Role-Playing	Nintendo	11.27	8.89	10.22	1.00	31.37	NaN	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
16714	Samurai Warriors: Sanada Maru	PS3	2016.0	Action	Tecmo Koei	0.00	0.00	0.01	0.00	0.01	NaN	NaN	NaN	NaN	NaN	NaN
16715	LMA Manager 2007	X360	2006.0	Sports	Codemasters	0.00	0.01	0.00	0.00	0.01	NaN	NaN	NaN	NaN	NaN	NaN
16716	Haitaka no Psychedelica	PSV	2016.0	Adventure	Idea Factory	0.00	0.00	0.01	0.00	0.01	NaN	NaN	NaN	NaN	NaN	NaN
16717	Spirits & Spells	GBA	2003.0	Platform	Wanadoo	0.01	0.00	0.00	0.00	0.01	NaN	NaN	NaN	NaN	NaN	NaN
16718	Winning Post 8 2016	PSV	2016.0	Simulation	Tecmo Koei	0.00	0.00	0.01	0.00	0.01	NaN	NaN	NaN	NaN	NaN	NaN

16719 rows × 16 columns

```
In [17]: df.shape
Out[17]: (16719, 16)
```

```
In [2]: df.describe(include='all')
```

Out[2]:

	Name	Platform	Year_of_Release	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales	Critic_Score	Critic_Count	User_Score	User_Count	Developer	Rating
count	16717	16719	16450.000000	16717	16665	16719.000000	16719.000000	16719.000000	16719.000000	16719.000000	8137.000000	8137.000000	10015	7590.000000	10096	9950
unique	11562	31	NaN	12	581	NaN	NaN	NaN	NaN	NaN	NaN	NaN	96	NaN	1696	8
top	Need for Speed: Most Wanted	PS2	NaN	Action	Electronic Arts	NaN	NaN	NaN	NaN	NaN	NaN	NaN	tbd	NaN	Ubisoft	E
freq	12	2161	NaN	3370	1356	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2425	NaN	204	3991
mean	NaN	NaN	2006.487356	NaN	NaN	0.263330	0.145025	0.077602	0.047332	0.533543	68.967679	26.360821	NaN	162.229908	NaN	NaN
std	NaN	NaN	5.878995	NaN	NaN	0.813514	0.503283	0.308818	0.186710	1.547935	13.938165	18.980495	NaN	561.282326	NaN	NaN
min	NaN	NaN	1980.000000	NaN	NaN	0.000000	0.000000	0.000000	0.000000	0.010000	13.000000	3.000000	NaN	4.000000	NaN	NaN
25%	NaN	NaN	2003.000000	NaN	NaN	0.000000	0.000000	0.000000	0.000000	0.060000	60.000000	12.000000	NaN	10.000000	NaN	NaN
50%	NaN	NaN	2007.000000	NaN	NaN	0.080000	0.020000	0.000000	0.010000	0.170000	71.000000	21.000000	NaN	24.000000	NaN	NaN
75%	NaN	NaN	2010.000000	NaN	NaN	0.240000	0.110000	0.040000	0.030000	0.470000	79.000000	36.000000	NaN	81.000000	NaN	NaN
max	NaN	NaN	2020.000000	NaN	NaN	41.360000	28.960000	10.220000	10.570000	82.530000	98.000000	113.000000	NaN	10665.000000	NaN	NaN

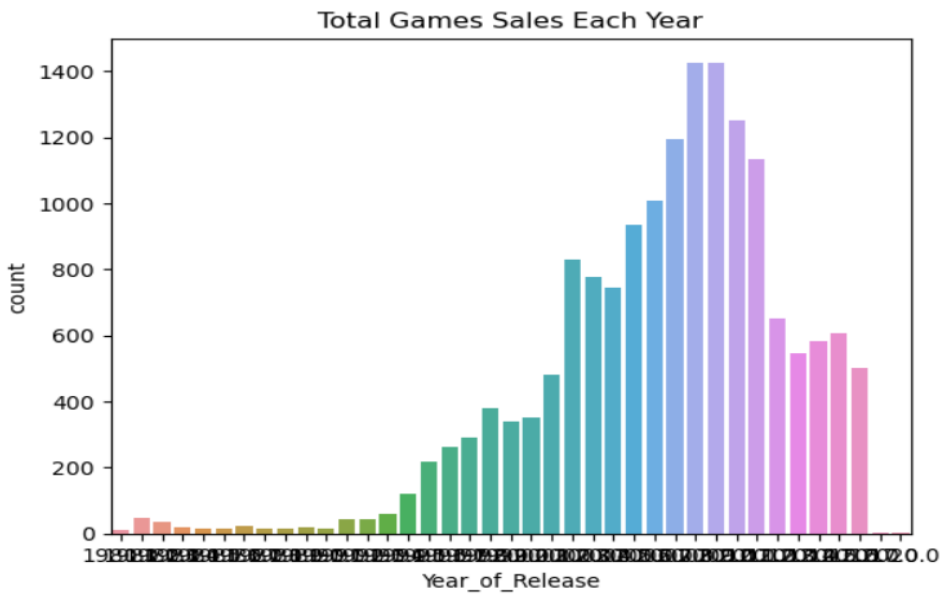
```
In [3]: df.info
```

```
Out[3]: <bound method DataFrame.info of                                      Name Platform  Year_of_Release         Genre  \
0                           Wii Sports      Wii           2006.0        Sports
1                    Super Mario Bros.      NES           1985.0      Platform
2                      Mario Kart Wii      Wii           2008.0        Racing
3                   Wii Sports Resort      Wii           2009.0        Sports
4             Pokemon Red/Pokemon Blue       GB           1996.0  Role-Playing
...                                ...      ...              ...           ...
16714  Samurai Warriors: Sanada Maru      PS3           2016.0        Action
16715               LMA Manager 2007     X360           2006.0        Sports
16716        Haitaka no Psychedelica      PSV           2016.0     Adventure
16717                Spirits & Spells      GBA           2003.0      Platform
16718            Winning Post 8 2016      PSV           2016.0    Simulation

          Publisher  NA_Sales  EU_Sales  JP_Sales  Other_Sales  Global_Sales  \
0          Nintendo     41.36     28.96      3.77         8.45         82.53
1          Nintendo     29.08      3.58      6.81         0.77         40.24
2          Nintendo     15.68     12.76      3.79         3.29         35.52
3          Nintendo     15.61     10.93      3.28         2.95         32.77
4          Nintendo     11.27      8.89     10.22         1.00         31.37
...             ...       ...       ...       ...          ...           ...
16714    Tecmo Koei      0.00      0.00      0.01         0.00          0.01
16715   Codemasters      0.00      0.01      0.00         0.00          0.01
16716  Idea Factory      0.00      0.00      0.01         0.00          0.01
16717       Wanadoo      0.01      0.00      0.00         0.00          0.01
16718    Tecmo Koei      0.00      0.00      0.01         0.00          0.01

       Critic_Score  Critic_Count User_Score  User_Count Developer Rating
0              76.0          51.0          8       322.0  Nintendo      E
1               NaN           NaN        NaN         NaN       NaN    NaN
2              82.0          73.0        8.3       709.0  Nintendo      E
3              80.0          73.0          8       192.0  Nintendo      E
4               NaN           NaN        NaN         NaN       NaN    NaN
...             ...           ...        ...         ...       ...    ...
16714           NaN           NaN        NaN         NaN       NaN    NaN
16715           NaN           NaN        NaN         NaN       NaN    NaN
16716           NaN           NaN        NaN         NaN       NaN    NaN
16717           NaN           NaN        NaN         NaN       NaN    NaN
16718           NaN           NaN        NaN         NaN       NaN    NaN

[16719 rows x 16 columns]>
```

Visualization of dataset - Inference of each graph:

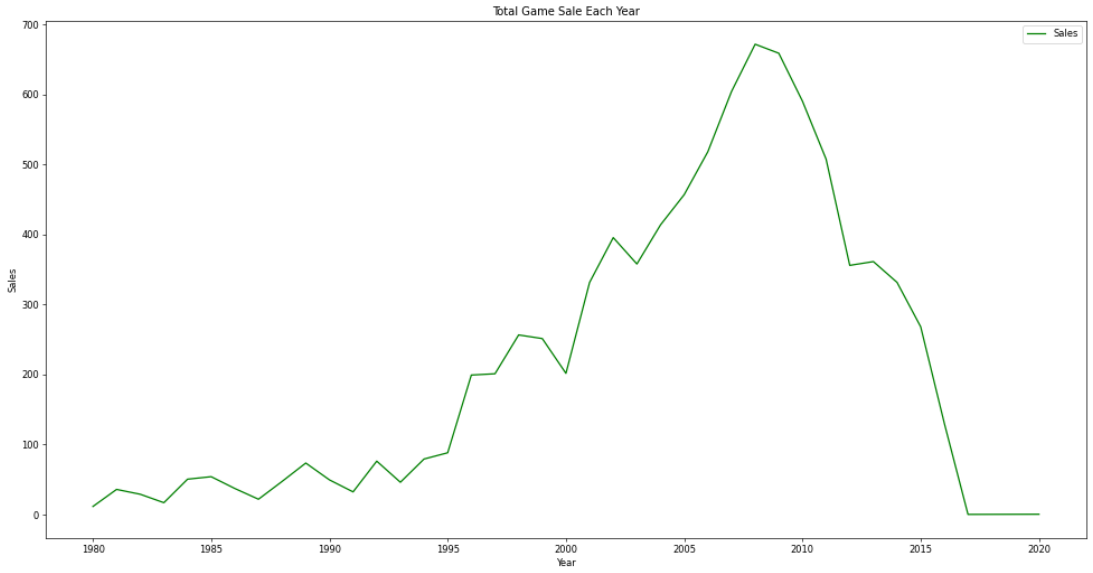
1. Total Sales Each Year

```
In [6]: sns.countplot(x= df['Year_of_Release'])
        plt.title('Total Games Sales Each Year')
        plt.show()
```



```
In [7]: sales_df=df.groupby('Year_of_Release',as_index=False).sum()

        x_axis=sales_df['Year_of_Release']
        y_axis=sales_df['Global_Sales']

        plt.figure(figsize=(20,10),dpi=60)
        plt.plot(x_axis,y_axis,label='Sales',color='green')
        plt.xlabel('Year')
        plt.ylabel('Sales')
        plt.title('Total Game Sale Each Year')
        plt.legend()
        plt.show()
```

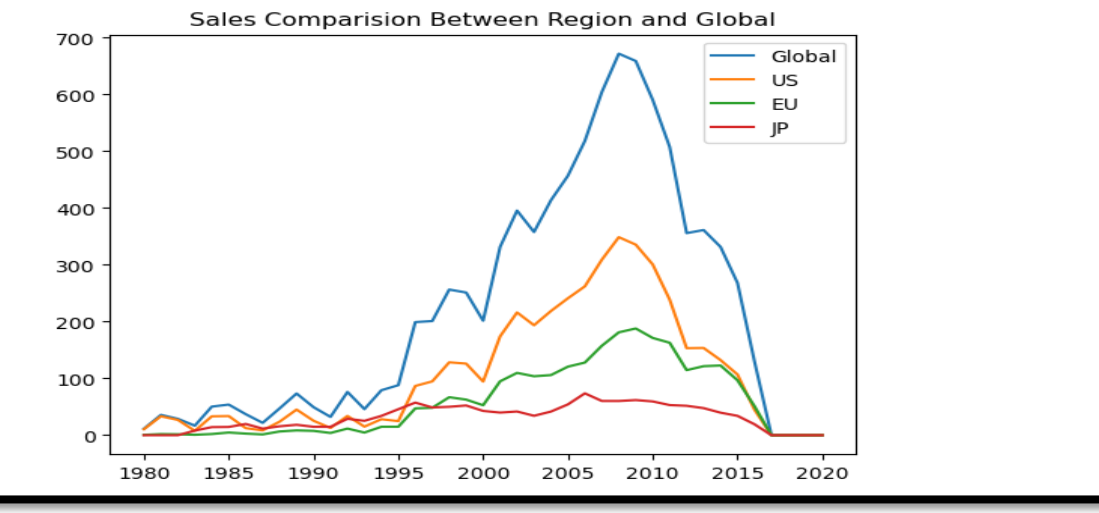
```
C:\Users\H Akshaya\AppData\Local\Temp\ipykernel_6308\2282229153.py:1: FutureWarning: The default value of numeric_only in DataFrameGroupBy.su
only or select only columns which should be valid for the function.
  sales_df=df.groupby('Year_of_Release',as_index=False).sum()
```



2. Total Sales Comparision Between Region Area

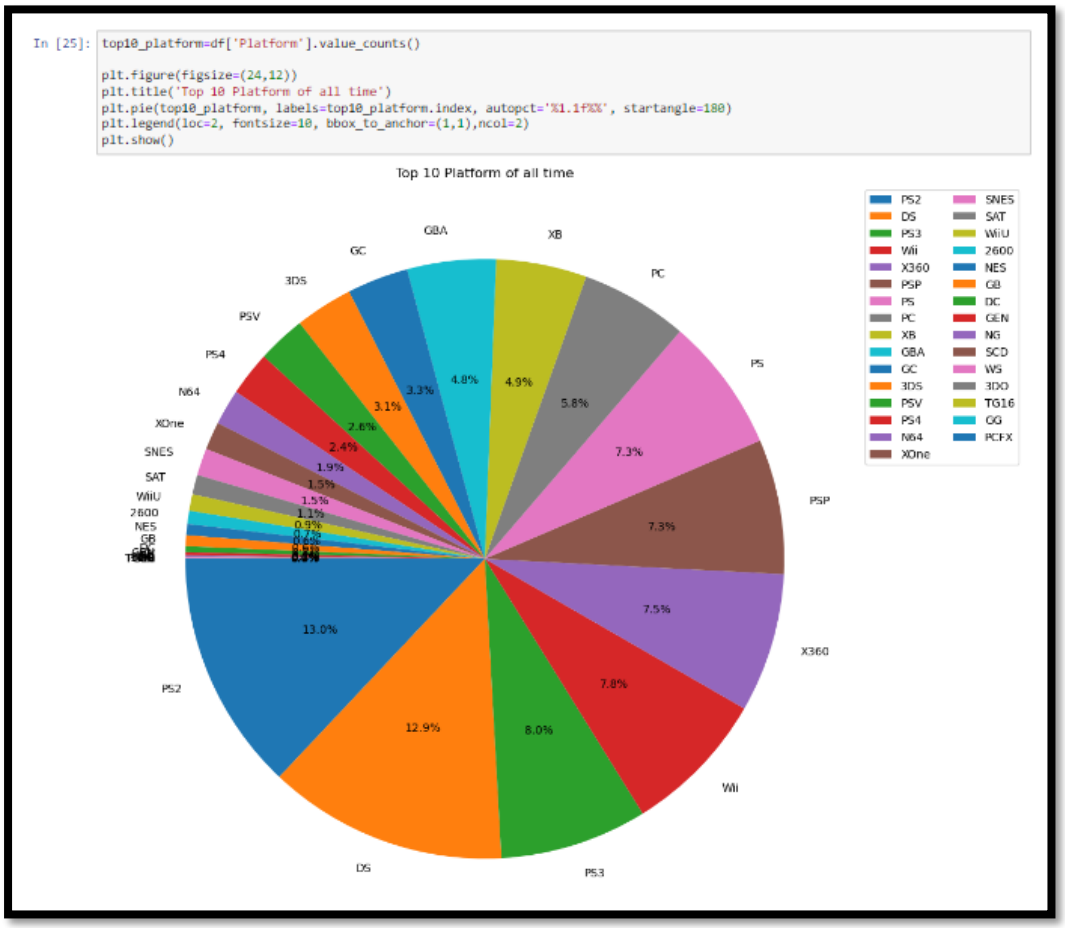
```
In [8]: na=sales_df['NA_Sales']
        eu=sales_df['EU_Sales']
        jp=sales_df['JP_Sales']
        total=sales_df['Global_Sales']


        plt.plot(x_axis, total, label='Global')
        plt.plot(x_axis, na, label='US')
        plt.plot(x_axis, eu, label='EU')
        plt.plot(x_axis, jp, label='JP')
        plt.title('Sales Comparision Between Region and Global')
        plt.legend(bbox_to_anchor=(1,1))
```

```
Out[8]: <matplotlib.legend.Legend at 0x15fa60f3390>
```



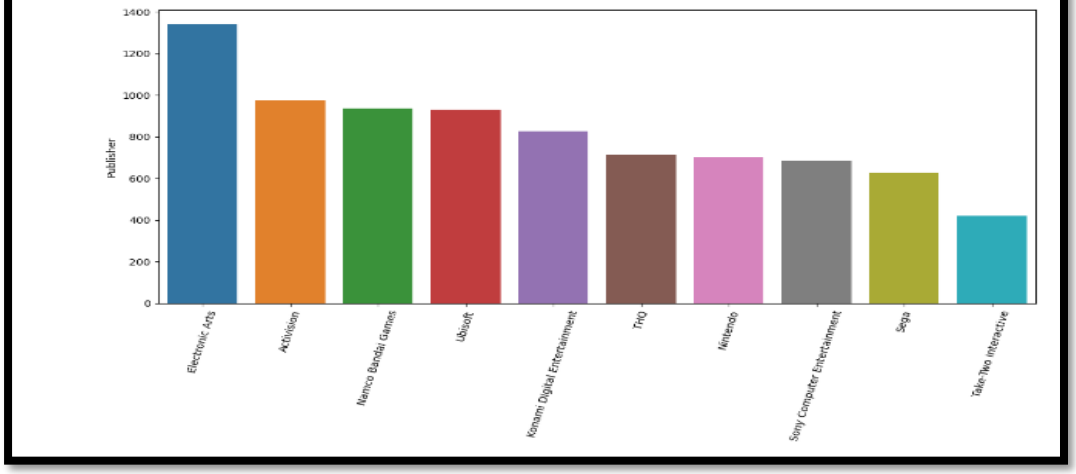
3. Top Platform Overall

```
In [25]: top10_platform=df['Platform'].value_counts()

         plt.figure(figsize=(24,12))
         plt.title('Top 10 Platform of all time')
         plt.pie(top10_platform, labels=top10_platform.index, autopct='%1.1f%%', startangle=180)
         plt.legend(loc=2, fontsize=10, bbox_to_anchor=(1,1),ncol=2)
         plt.show()
```



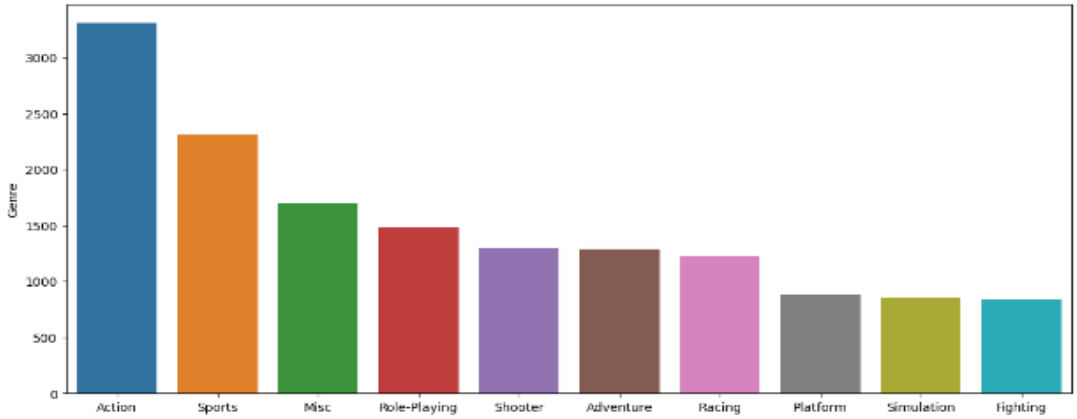
4. Top 10 Publishers

```
In [27]: top10_publisher=df['Publisher'].value_counts().head(10)
         top10_publisher

Out[27]: Electronic Arts                1344
         Activision                      976
         Namco Bandai Games              935
         Ubisoft                         929
         Konami Digital Entertainment    825
         THQ                             712
         Nintendo                        700
         Sony Computer Entertainment     686
         Sega                            629
         Take-Two Interactive            421
         Name: Publisher, dtype: int64

In [28]: plt.figure(figsize=(14,6))
         plt.xticks(rotation=75)
         sns.barplot(x=top10_publisher.index, y=top10_publisher)

Out[28]: <Axes: ylabel='Publisher'>
```

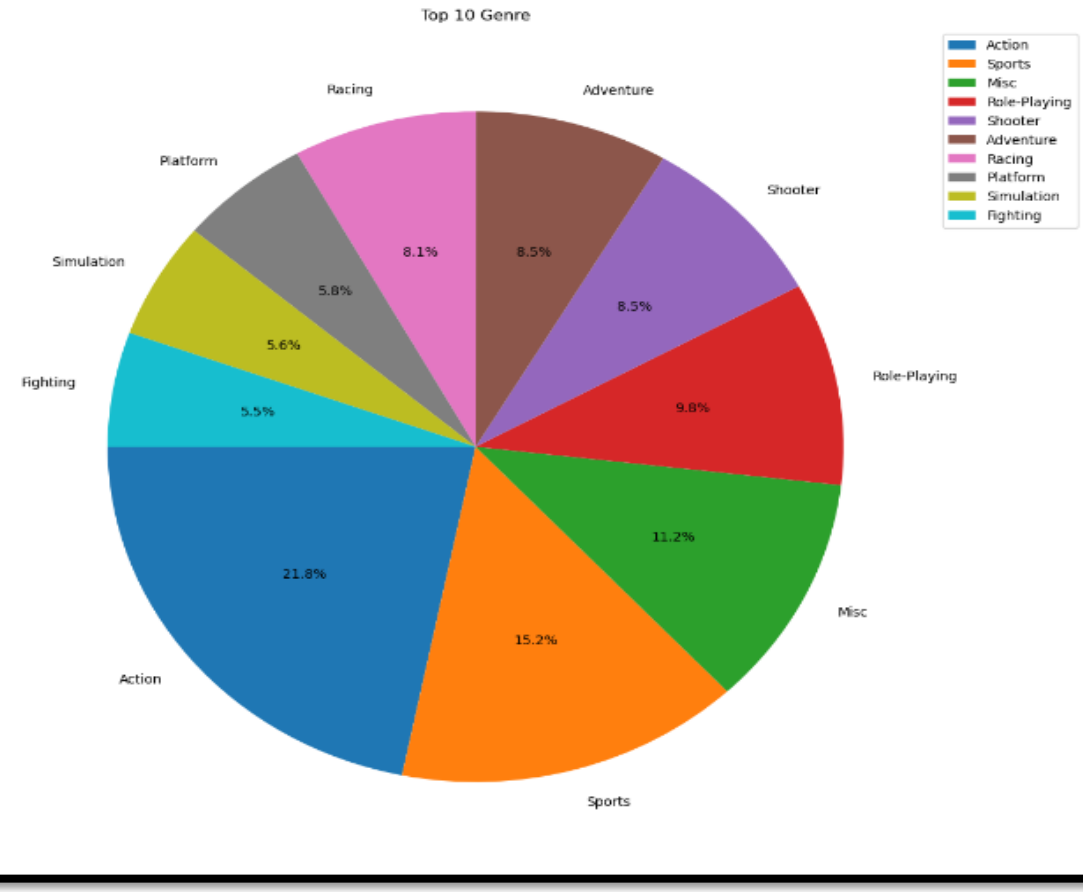


5. Top Genre

```
In [10]: top10_genre=df['Genre'].value_counts().head(10)

         plt.figure(figsize=(14,6))
         sns.barplot(x=top10_genre.index, y=top10_genre)
         plt.show()
```

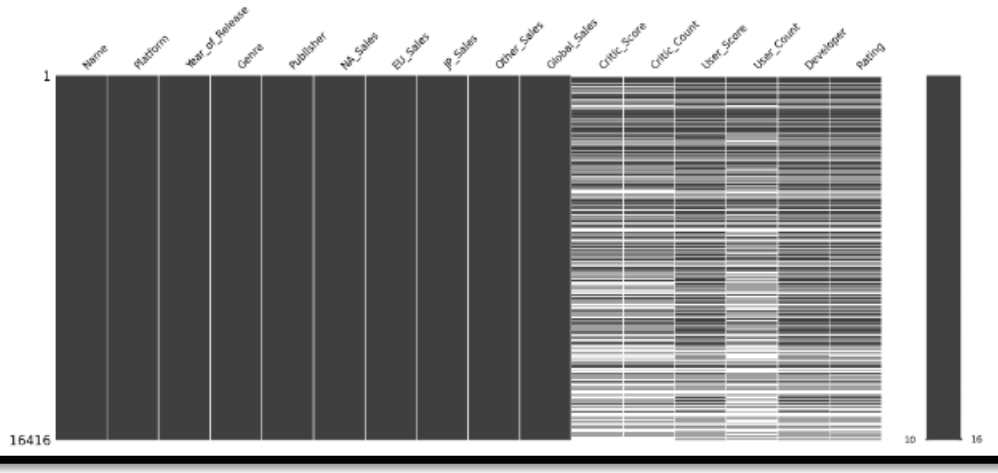



```
In [11]: plt.figure(figsize=(24,12))
         plt.title('Top 10 Genre')
         plt.pie(top10_genre, labels=top10_genre.index, autopct='%1.1f%%', startangle=180)
         plt.legend(loc=2,fontsize=10, bbox_to_anchor=(1,1))
         plt.show()
```



Quality of data (noise,missing values):

```
In [12]: df.isna().sum()

Out[12]: 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
         Critic_Score     8434
         Critic_Count     8434
         User_Score       6579
         User_Count       8955
         Developer        6512
         Rating           6649
         dtype: int64

In [14]: import missingno as msno

         msno.matrix(df)

Out[14]: <Axes: >
```



Handling missing values:

```
In [18]: df['Critic_Score']=df['Critic_Score'].fillna('U')

In [19]: df['Critic_Count']=df['Critic_Count'].fillna('U')

In [20]: df['User_Score']=df['User_Score'].fillna('U')

In [21]: df['User_Count']=df['User_Count'].fillna('U')

In [22]: df['Developer']=df['Developer'].fillna('U')

In [23]: df['Rating']=df['Rating'].fillna('U')

In [24]: df.isna().sum()

Out[24]: 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
         Critic_Score     0
         Critic_Count     0
         User_Score       0
         User_Count       0
         Developer        0
         Rating           0
         dtype: int64
```

In [25]:
```
df
```

Out[25]:

	Name	Platform	Year_of_Release	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales	Critic_Score	Critic_Count	User_Score	User_Count	Developer	Rating
0	Wii Sports	Wii	2006.0	Sports	Nintendo	41.36	28.96	3.77	8.45	82.53	76.0	51.0	8	322.0	Nintendo	E
1	Super Mario Bros.	NES	1985.0	Platform	Nintendo	29.08	3.58	6.81	0.77	40.24	U	U	U	U	U	U
2	Mario Kart Wii	Wii	2008.0	Racing	Nintendo	15.68	12.76	3.79	3.29	35.52	82.0	73.0	8.3	709.0	Nintendo	E
3	Wii Sports Resort	Wii	2009.0	Sports	Nintendo	15.61	10.93	3.28	2.95	32.77	80.0	73.0	8	192.0	Nintendo	E
4	Pokemon Red/Pokemon Blue	GB	1996.0	Role-Playing	Nintendo	11.27	8.89	10.22	1.00	31.37	U	U	U	U	U	U
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
16714	Samurai Warriors: Sanada Maru	PS3	2016.0	Action	Tecmo Koei	0.00	0.00	0.01	0.00	0.01	U	U	U	U	U	U
16715	LMA Manager 2007	X360	2006.0	Sports	Codemasters	0.00	0.01	0.00	0.00	0.01	U	U	U	U	U	U
16716	Haitaka no Psychedelica	PSV	2016.0	Adventure	Idea Factory	0.00	0.00	0.01	0.00	0.01	U	U	U	U	U	U
16717	Spirits & Spells	GBA	2003.0	Platform	Wanadoo	0.01	0.00	0.00	0.00	0.01	U	U	U	U	U	U
16718	Winning Post 8 2016	PSV	2016.0	Simulation	Tecmo Koei	0.00	0.00	0.01	0.00	0.01	U	U	U	U	U	U

16416 rows × 16 columns