

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

```
In [2]: # import Gaming dataset
data = '/Users/hambolu/Downloads/Video_Games_Sales_as_at_22_Dec_2016.csv'
df = pd.read_csv(data)
df.head()
```

```
Out[2]:
```

	Name	Platform	Year_of_Release	Genre	Publisher	NA_Sales	EU_Sales	JP_Sal
0	Wii Sports	Wii	2006.0	Sports	Nintendo	41.36	28.96	3
1	Super Mario Bros.	NES	1985.0	Platform	Nintendo	29.08	3.58	6
2	Mario Kart Wii	Wii	2008.0	Racing	Nintendo	15.68	12.76	3
3	Wii Sports Resort	Wii	2009.0	Sports	Nintendo	15.61	10.93	3.
4	Pokemon Red/Pokemon Blue	GB	1996.0	Role-Playing	Nintendo	11.27	8.89	10.

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

```
Out[3]: (16719, 16)
```

```
In [4]: df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16719 entries, 0 to 16718
Data columns (total 16 columns):
 #   Column                Non-Null Count  Dtype  
---  -
 0   Name                  16717 non-null  object 
 1   Platform              16719 non-null  object 
 2   Year_of_Release       16450 non-null  float64
 3   Genre                 16717 non-null  object 
 4   Publisher             16665 non-null  object 
 5   NA_Sales              16719 non-null  float64
 6   EU_Sales              16719 non-null  float64
 7   JP_Sales              16719 non-null  float64
 8   Other_Sales           16719 non-null  float64
 9   Global_Sales          16719 non-null  float64
10   Critic_Score          8137 non-null   float64
11   Critic_Count          8137 non-null   float64
12   User_Score            10015 non-null  object 
13   User_Count            7590 non-null   float64
14   Developer             10096 non-null  object 
15   Rating                9950 non-null   object 
dtypes: float64(9), object(7)
memory usage: 2.0+ MB

```

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

```

Out[5]: Name                11562
Platform                  31
Year_of_Release           39
Genre                     12
Publisher                 581
NA_Sales                  402
EU_Sales                  307
JP_Sales                  244
Other_Sales               155
Global_Sales              629
Critic_Score              82
Critic_Count              106
User_Score                 96
User_Count                888
Developer                 1696
Rating                    8
dtype: int64

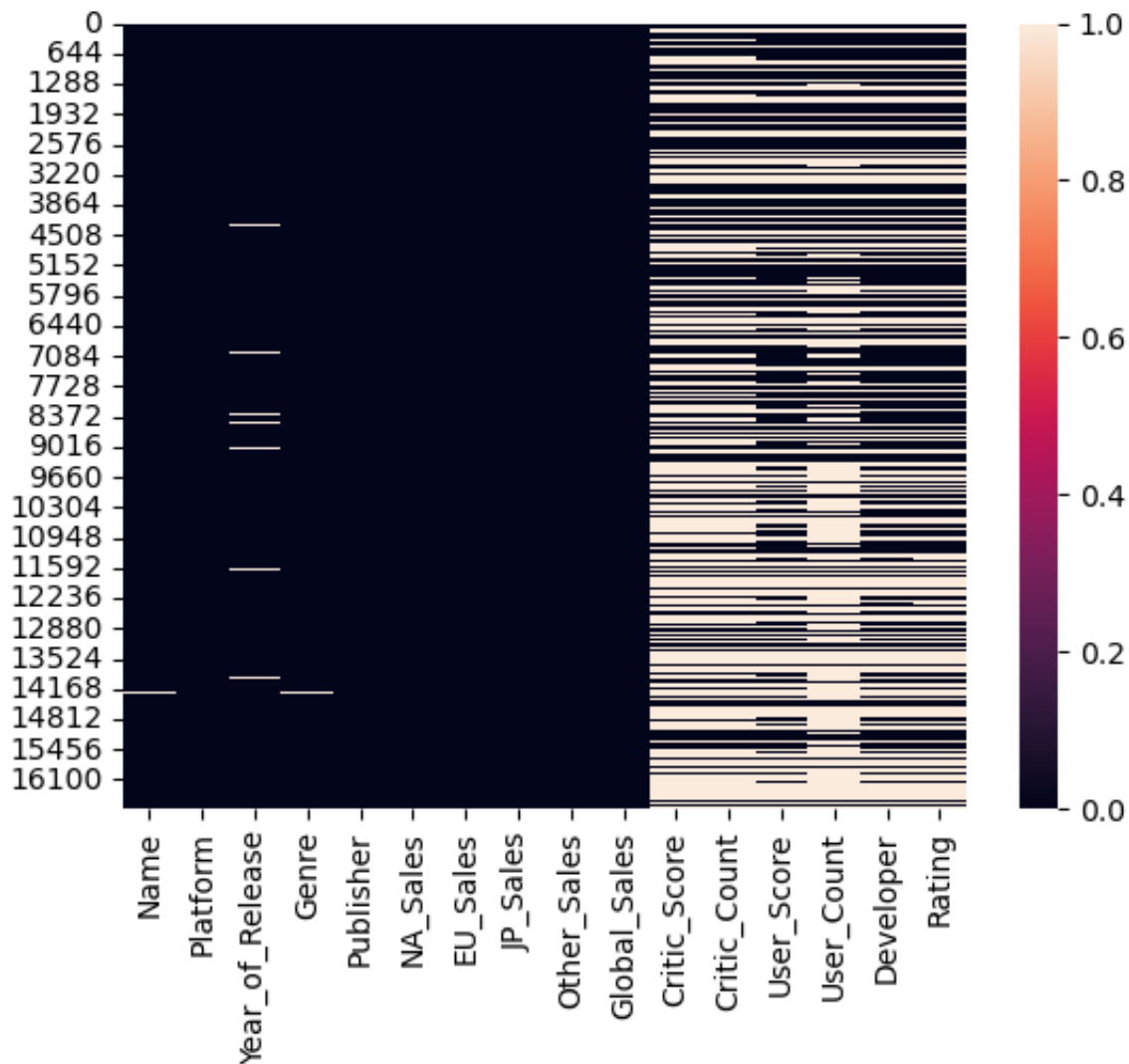
```

```
In [6]: df.isna().any()
```

```
Out[6]: Name                True
        Platform            False
        Year_of_Release      True
        Genre                True
        Publisher            True
        NA_Sales             False
        EU_Sales             False
        JP_Sales             False
        Other_Sales          False
        Global_Sales         False
        Critic_Score         True
        Critic_Count         True
        User_Score           True
        User_Count           True
        Developer            True
        Rating               True
        dtype: bool
```

```
In [7]: sns.heatmap(df.isnull())
```

```
Out[7]: <AxesSubplot:>
```



In [8]:

```
Out[8]: Year_of_Release  Year_of_Release    1.000000
NA_Sales      NA_Sales      1.000000
Critic_Count   Critic_Count  1.000000
Critic_Score   Critic_Score  1.000000
Global_Sales   Global_Sales  1.000000

...
Year_of_Release  Other_Sales    0.037700
                Critic_Score    0.011411
Critic_Score     Year_of_Release 0.011411
Year_of_Release  EU_Sales       0.003842
EU_Sales         Year_of_Release 0.003842
Length: 81, dtype: float64
```

In [9]: `df.isnull()`

Out[9]:

	Name	Platform	Year_of_Release	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False
...
16714	False	False	False	False	False	False	False	False
16715	False	False	False	False	False	False	False	False
16716	False	False	False	False	False	False	False	False
16717	False	False	False	False	False	False	False	False
16718	False	False	False	False	False	False	False	False

16719 rows × 16 columns

In [10]:

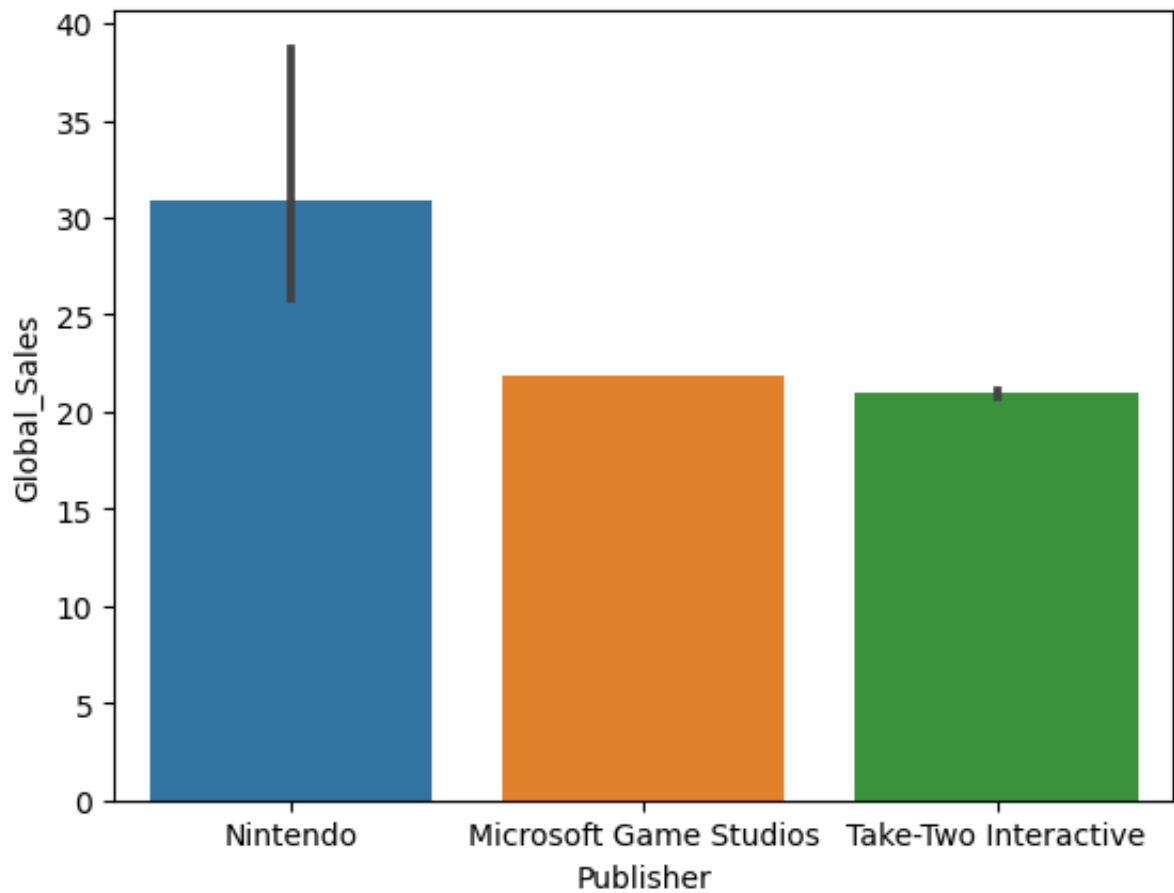
```
# Most popular gaming publishers
Publisher = df['Publisher'].value_counts().sort_values(ascending = False)
print(Publisher)
```

```
Electronic Arts          1356
Activision               985
Namco Bandai Games      939
Ubisoft                 933
Konami Digital Entertainment 834
...
KSS                      1
Giza10                   1
Palcom                   1
EON Digital Entertainment 1
Red Flagship             1
Name: Publisher, Length: 581, dtype: int64
```

In [11]:

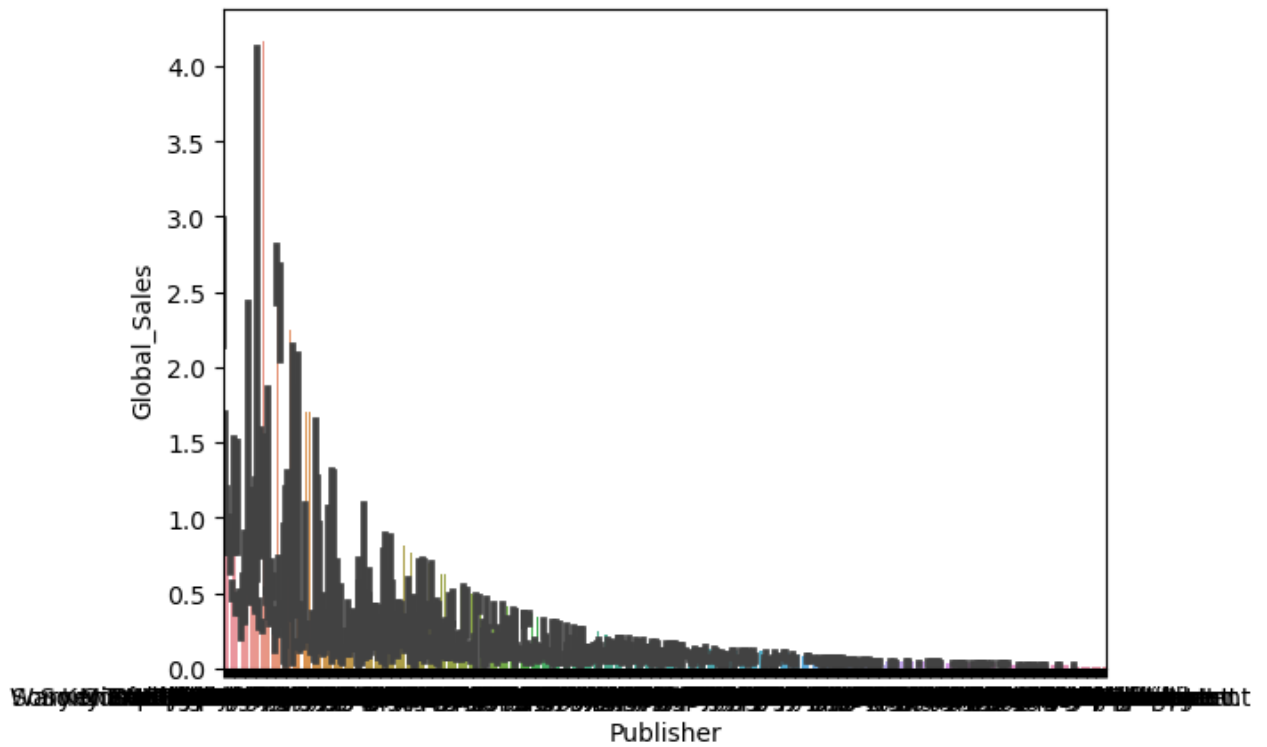
```
#Top 3 publihsers grouped
top_3 = df.nlargest(20, 'Global_Sales')
sns.barplot(data = top_3, x = "Publisher", y = "Global_Sales")
```

Out[11]: <AxesSubplot:xlabel='Publisher', ylabel='Global_Sales'>

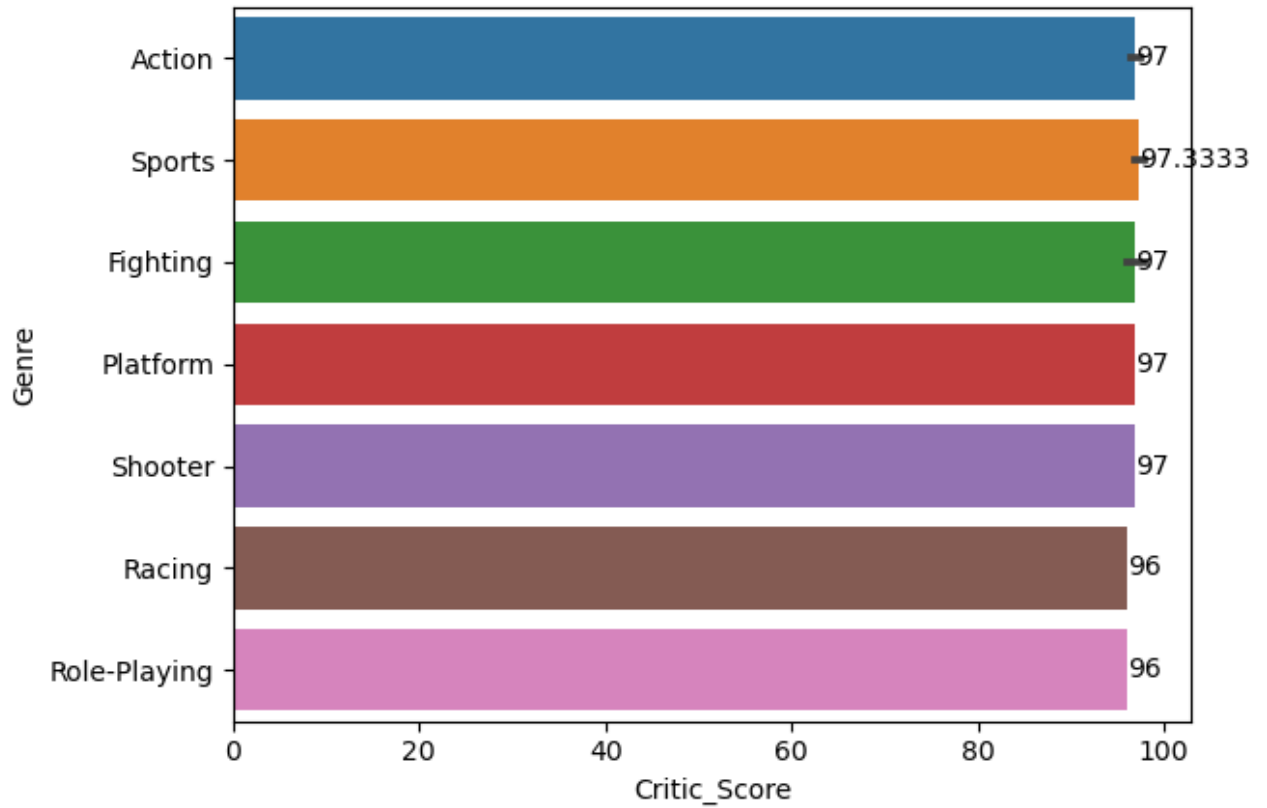


```
In [12]: sns.barplot(data = df, x = "Publisher", y = "Global_Sales")
```

```
Out[12]: <AxesSubplot:xlabel='Publisher', ylabel='Global_Sales'>
```

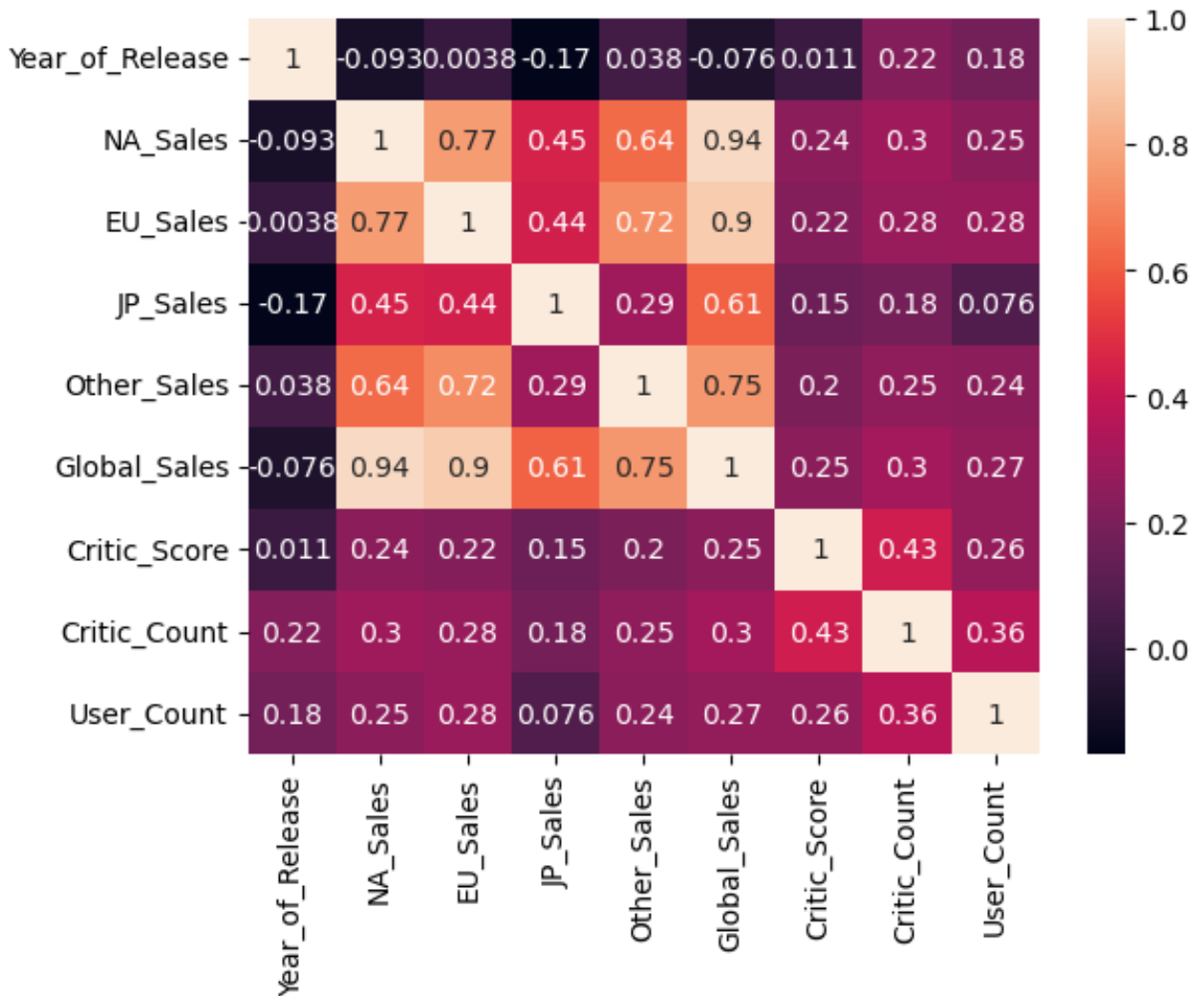


```
In [13]: #top critic scores by genre
Critic = df.nlargest(20, 'Critic_Score')
C1 = sns.barplot(data = Critic, x = 'Critic_Score', y = "Genre")
for i in C1.containers:
    C1.bar_label(i,)
```



```
In [15]: #Correlation between columns  
plt.figure  
sns.heatmap(df.corr(), annot=True)
```

```
Out[15]: <AxesSubplot:>
```

```
In [16]: # Checking for correlation by unstacking data
corr = df.corr()
c1 = corr.abs().unstack()
c1.sort_values(ascending = False)
```

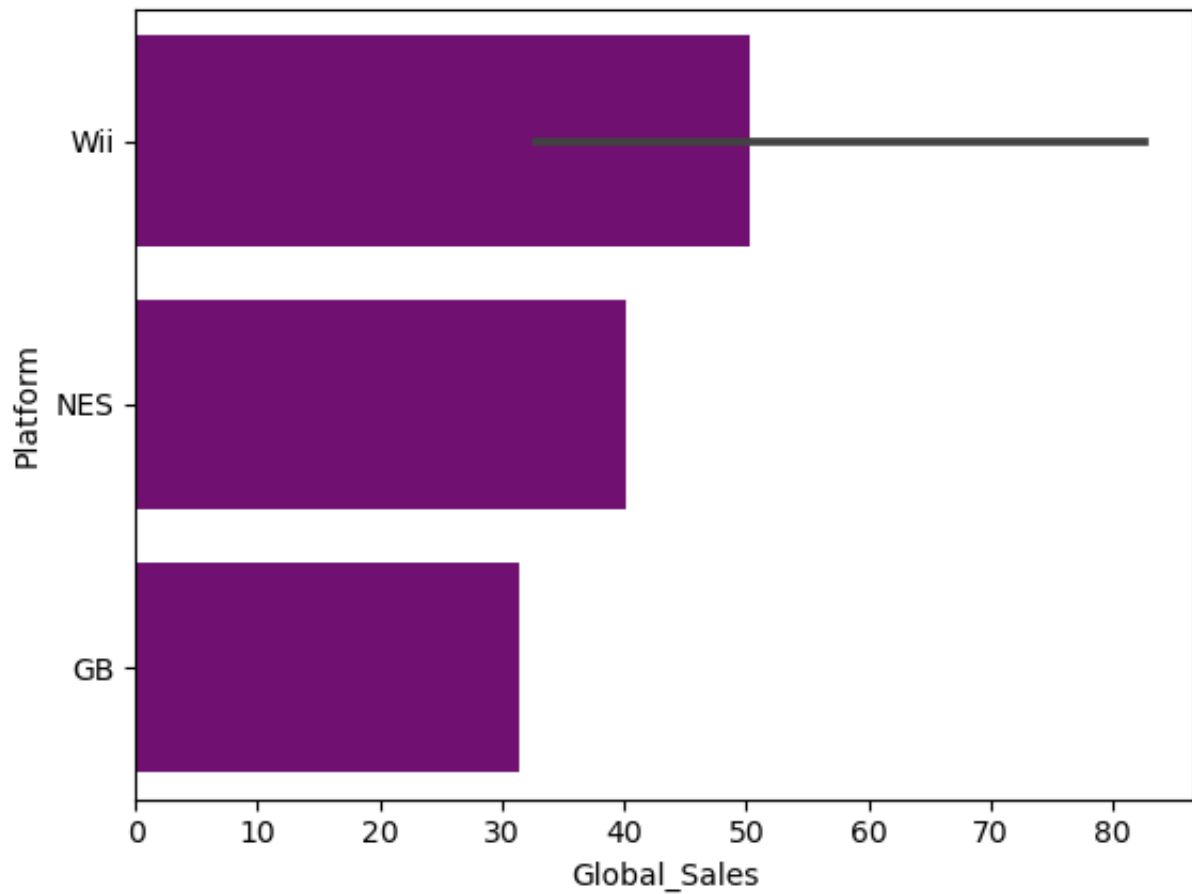
```
Out[16]: Year_of_Release  Year_of_Release    1.000000
NA_Sales      NA_Sales      1.000000
Critic_Count   Critic_Count  1.000000
Critic_Score   Critic_Score  1.000000
Global_Sales   Global_Sales  1.000000
...
Year_of_Release  Other_Sales    0.037700
                Critic_Score    0.011411
Critic_Score     Year_of_Release  0.011411
Year_of_Release  EU_Sales       0.003842
EU_Sales         Year_of_Release  0.003842
Length: 81, dtype: float64
```

```
In [64]: #Most popular year for games releases
variety = df['Year_of_Release'].value_counts(ascending = False)
print(variety)
```

```
2008.0    1427
2009.0    1426
2010.0    1255
2007.0    1197
2011.0    1136
2006.0    1006
2005.0     939
2002.0     829
2003.0     775
2004.0     762
2012.0     653
2015.0     606
2014.0     581
2013.0     544
2016.0     502
2001.0     482
1998.0     379
2000.0     350
1999.0     338
1997.0     289
1996.0     263
1995.0     219
1994.0     121
1993.0      62
1981.0      46
1992.0      43
1991.0      41
1982.0      36
1986.0      21
1989.0      17
1983.0      17
1990.0      16
1987.0      16
1988.0      15
1985.0      14
1984.0      14
1980.0       9
2017.0       3
2020.0       1
Name: Year_of_Release, dtype: int64
```

```
In [79]: #Top 3 major sales by platform
Biggest_platform = df.nlargest(5, 'Global_Sales')
sns.barplot(data = Biggest_platform, x = 'Global_Sales', y = 'Platform', col
```

```
Out[79]: <AxesSubplot:xlabel='Global_Sales', ylabel='Platform'>
```



```
In [80]: #Top 5 major sales by genre  
Largest_genre = df.nlargest(6, 'Global_Sales')  
sns.barplot(data = Largest_genre, x = 'Global_Sales', y = 'Genre')
```

```
Out[80]: <AxesSubplot:xlabel='Global_Sales', ylabel='Genre'>
```

