Video Game Sales Exploratory Data Analysis

July 10, 2021

[2]:
```
import pandas as pd
import os
import numpy as np
```

[3]:
```
import matplotlib.pyplot as plt
import seaborn as sns
```

[4]:
```
df = pd.read_csv('D:\Databases\Video_Games_Sales_as_at_22_Dec_2016.csv')
df
```

[4]:

	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	8.0	322.0	Nintendo	E
1	Super Mario Bros.	NES	1985.0	Platform	Nintendo	29.08	3.58	6.81	0.77	40.24						
2	Mario Kart Wii	Wii	2008.0	Racing	Nintendo	15.68	12.76	3.79	3.29	35.52						
3	Wii Sports Resort	Wii	2009.0	Sports	Nintendo	15.61	10.93	3.28	2.95	32.77						
4	Pokemon Red/Pokemon Blue	GB	1996.0	Role-Playing	Nintendo	11.27	8.89	10.22	1.00	31.37						
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
16714	Samurai Warriors: Sanada Maru	PS3	2016.0	Action	Tecmo Koei	0.00	0.00	0.01	0.00	0.01						
16715	LMA Manager 2007	X360	2006.0	Sports	Codemasters	0.00	0.01	0.00	0.00	0.01						
16716	Haitaka no Psychedelica	PSV	2016.0	Adventure	Idea Factory	0.00	0.00	0.01	0.00	0.01						
16717	Spirits & Spells	GBA	2003.0	Platform	Wanadoo	0.01	0.00	0.00	0.00	0.01						
16718	Winning Post 8 2016	PSV	2016.0	Simulation	Tecmo Koei	0.00	0.00	0.01	0.00	0.01						

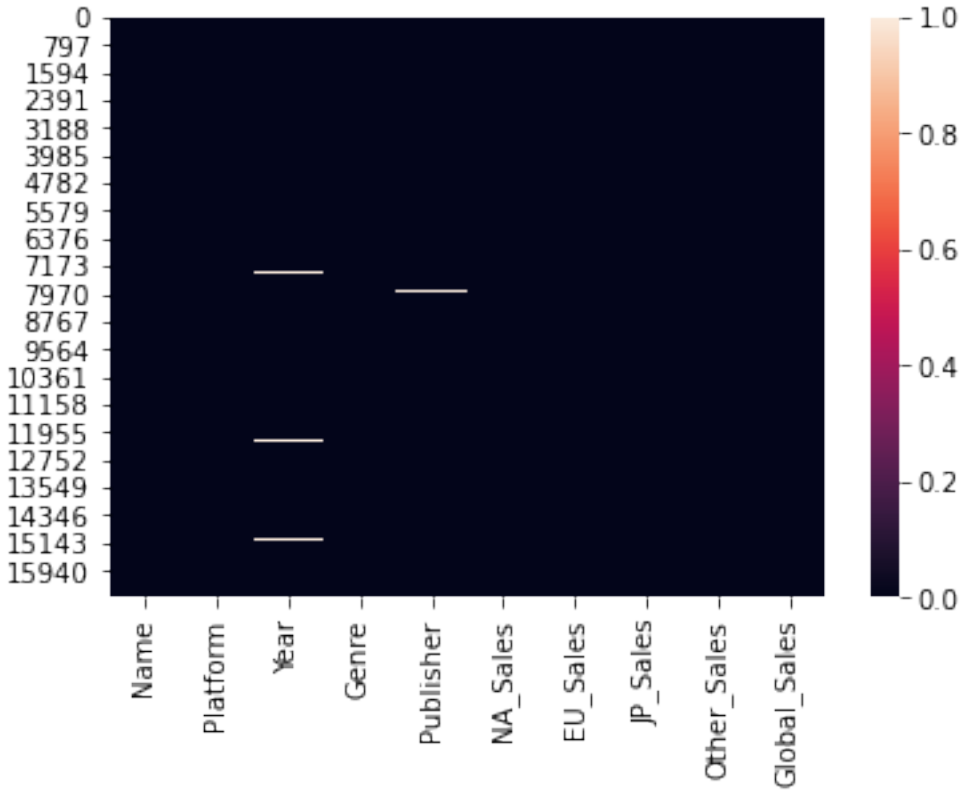
```
1               NaN           NaN         NaN          NaN       NaN    NaN
2              82.0          73.0         8.3        709.0  Nintendo      E
3              80.0          73.0         8.0        192.0  Nintendo      E
4               NaN           NaN         NaN          NaN       NaN    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]
```

[5]:
```
df = df.
 ↪drop(columns=['Critic_Score','Critic_Count','User_Score','User_Count','Developer','Rating'])
df = df.rename(columns={'Year_of_Release':'Year'})
df
```

[5]:
```
                              Name Platform    Year         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

[16719 rows x 10 columns]
```

[6]:
```
display(df.head())

sns.heatmap(df.isnull())
plt.show()

print(df.shape, "\n")

plt.show()
```

```
                      Name Platform    Year         Genre Publisher  \
0                Wii Sports      Wii  2006.0        Sports  Nintendo
1         Super Mario Bros.      NES  1985.0      Platform  Nintendo
2            Mario Kart Wii      Wii  2008.0        Racing  Nintendo
3         Wii Sports Resort      Wii  2009.0        Sports  Nintendo
4  Pokemon Red/Pokemon Blue       GB  1996.0  Role-Playing  Nintendo

   NA_Sales  EU_Sales  JP_Sales  Other_Sales  Global_Sales
0     41.36     28.96      3.77         8.45         82.53
1     29.08      3.58      6.81         0.77         40.24
2     15.68     12.76      3.79         3.29         35.52
3     15.61     10.93      3.28         2.95         32.77
4     11.27      8.89     10.22         1.00         31.37
```

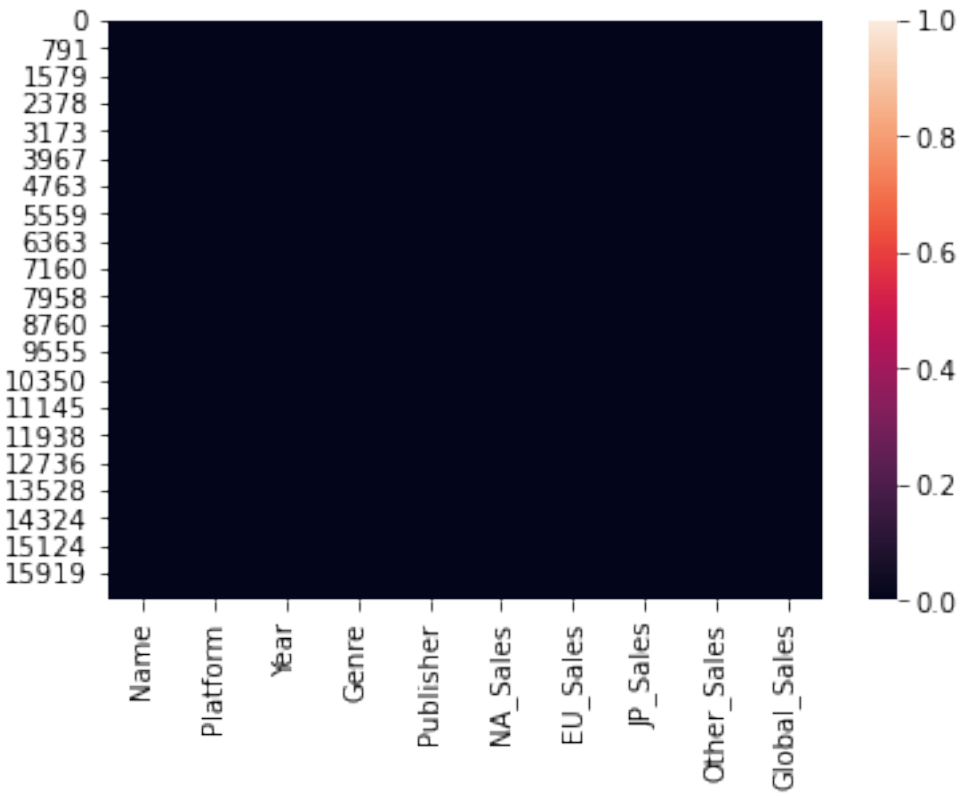


```
(16719, 10)
```

[7]:
```
df= df.dropna(subset=['Publisher','Year'])
df
```

[7]:
```
                              Name Platform    Year         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

[16418 rows x 10 columns]
```

[8]:
```
sns.heatmap(df.isnull())
plt.show()
print(df.shape, "\n")
```



(16418, 10)

[9]:
```
publishercol = df.drop(columns = ['Year', 'Genre', 'Name','Platform'])
publishercol
```

[9]:

	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

[16418 rows x 6 columns]

[10]:
```
publishersum = publishercol.groupby('Publisher').sum()
publishersum = publishersum.sort_values(by=['Global_Sales'], ascending=False).
 ↪head(10)
publishersum = publishersum.rename(index={'Electronic Arts':'EA',
                                    'Sony Computer Entertainment': 'Sony',
                                    'Take-Two Interactive': 'Take-Two',
                                    'Konami Digital Entertainment':'Konami',
                                    'Namco Bandai Games':'Namco'})
publishersum
```

[10]:

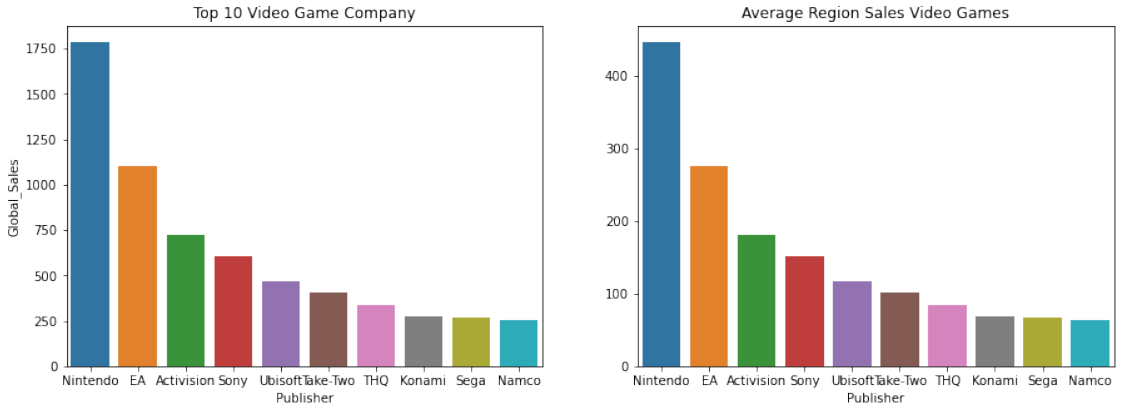
Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
Nintendo	815.86	418.58	457.74	94.55	1786.72
EA	588.65	370.02	14.29	126.82	1100.05
Activision	428.92	214.10	6.71	75.27	725.14
Sony	266.17	186.40	74.15	79.62	606.27
Ubisoft	252.12	161.72	7.35	49.09	470.45
Take-Two	222.92	119.06	5.93	55.68	403.58
THQ	207.55	93.66	5.01	31.87	338.11
Konami	88.67	67.93	91.03	29.54	277.36
Sega	107.99	80.08	56.22	23.59	268.11
Namco	69.76	42.14	127.62	14.55	254.19

[11]:
```
publishermean = publishersum.drop(columns = 'Global_Sales')
publishermean = publishermean.mean(axis=1)
publishermean
```

[11]:
```
Publisher
Nintendo      446.6825
EA            274.9450
Activision    181.2500
Sony          151.5850
Ubisoft       117.5700
Take-Two      100.8975
THQ            84.5225
Konami         69.2925
Sega           66.9700
Namco          63.5175
dtype: float64
```

[12]:
```
fig, ax = plt.subplots(1,2, figsize=(15,5))

sns.barplot(x=publishersum.index, y=publishersum['Global_Sales'], ax=ax[0])
# sns.set_style("whitegrid")
ax[0].set_title('Top 10 Video Game Company')

sns.barplot(x=publishermean.index, y=publishermean, ax=ax[1])
```

```
# sns.set_style("darkgrid")
ax[1].set_title('Average Region Sales Video Games')
plt.show()
```



[13]:
```
publishersum
```

[13]:

Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
Nintendo	815.86	418.58	457.74	94.55	1786.72
EA	588.65	370.02	14.29	126.82	1100.05
Activision	428.92	214.10	6.71	75.27	725.14
Sony	266.17	186.40	74.15	79.62	606.27
Ubisoft	252.12	161.72	7.35	49.09	470.45
Take-Two	222.92	119.06	5.93	55.68	403.58
THQ	207.55	93.66	5.01	31.87	338.11
Konami	88.67	67.93	91.03	29.54	277.36
Sega	107.99	80.08	56.22	23.59	268.11
Namco	69.76	42.14	127.62	14.55	254.19

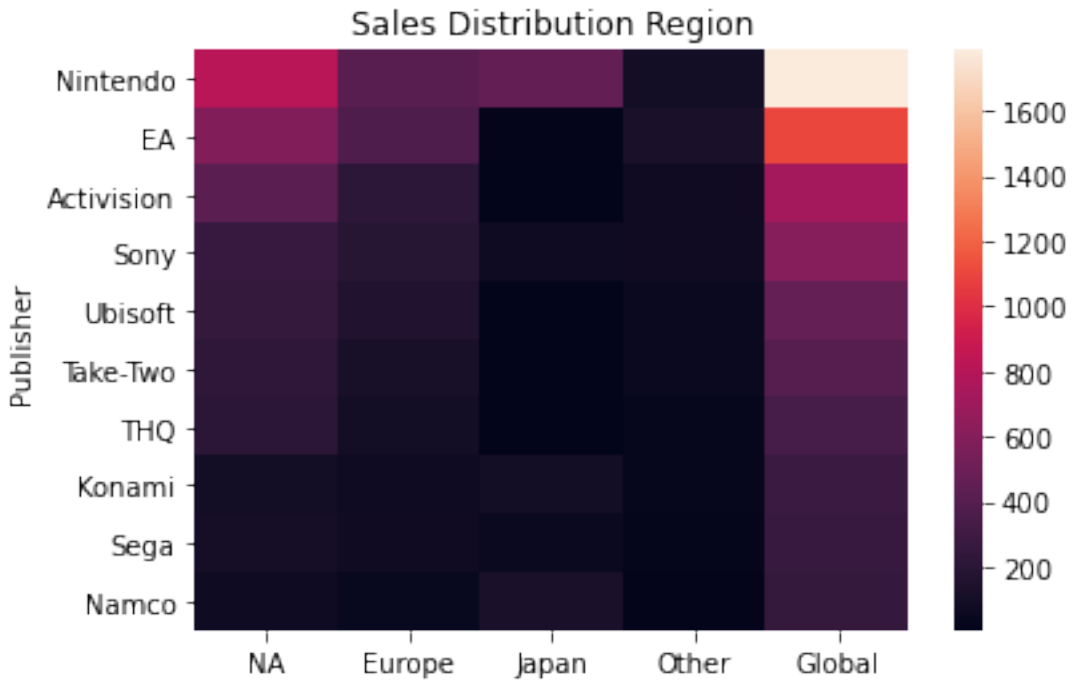
[14]:
```
publisherheatmap = publishersum.rename(columns={'NA_Sales':'NA',
                                    'EU_Sales': 'Europe',
                                    'JP_Sales': 'Japan',
                                    'Other_Sales':'Other',
                                    'Global_Sales':'Global'})
publisherheatmap
```

[14]:

Publisher	NA	Europe	Japan	Other	Global
Nintendo	815.86	418.58	457.74	94.55	1786.72
EA	588.65	370.02	14.29	126.82	1100.05
Activision	428.92	214.10	6.71	75.27	725.14

```
Sony          266.17  186.40   74.15   79.62   606.27
Ubisoft       252.12  161.72    7.35   49.09   470.45
Take-Two      222.92  119.06    5.93   55.68   403.58
THQ           207.55   93.66    5.01   31.87   338.11
Konami         88.67   67.93   91.03   29.54   277.36
Sega          107.99   80.08   56.22   23.59   268.11
Namco          69.76   42.14  127.62   14.55   254.19
```

[15]:
```
sns.heatmap(publisherheatmap)
plt.title('Sales Distribution Region')
plt.show()
```



[16]:
```
genre = df.drop(columns = ['Year','Name','Platform','Publisher'])
genre
```

[16]:

	Genre	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
0	Sports	41.36	28.96	3.77	8.45	82.53
1	Platform	29.08	3.58	6.81	0.77	40.24
2	Racing	15.68	12.76	3.79	3.29	35.52
3	Sports	15.61	10.93	3.28	2.95	32.77
4	Role-Playing	11.27	8.89	10.22	1.00	31.37
…	…	…	…	…	…	…
16714	Action	0.00	0.00	0.01	0.00	0.01
16715	Sports	0.00	0.01	0.00	0.00	0.01

```
16716     Adventure      0.00      0.00      0.01         0.00          0.01
16717      Platform      0.01      0.00      0.00         0.00          0.01
16718    Simulation      0.00      0.00      0.01         0.00          0.01

[16418 rows x 6 columns]
```

[17]:
```
df
```

[17]:
```
                                Name Platform    Year         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

[16418 rows x 10 columns]
```

[18]:
```
genrecounts = genre.groupby('Genre').count()
genrecounts
```

[18]:
```
              NA_Sales  EU_Sales  JP_Sales  Other_Sales  Global_Sales
Genre
Action            3307      3307      3307         3307          3307
Adventure         1291      1291      1291         1291          1291
Fighting           837       837       837          837           837
Misc              1697      1697      1697         1697          1697
Platform           878       878       878          878           878
Puzzle             569       569       569          569           569
```

```
Racing            1225      1225      1225          1225          1225
Role-Playing      1483      1483      1483          1483          1483
Shooter           1296      1296      1296          1296          1296
Simulation         855       855       855           855           855
Sports            2306      2306      2306          2306          2306
Strategy           672       672       672           672           672
```

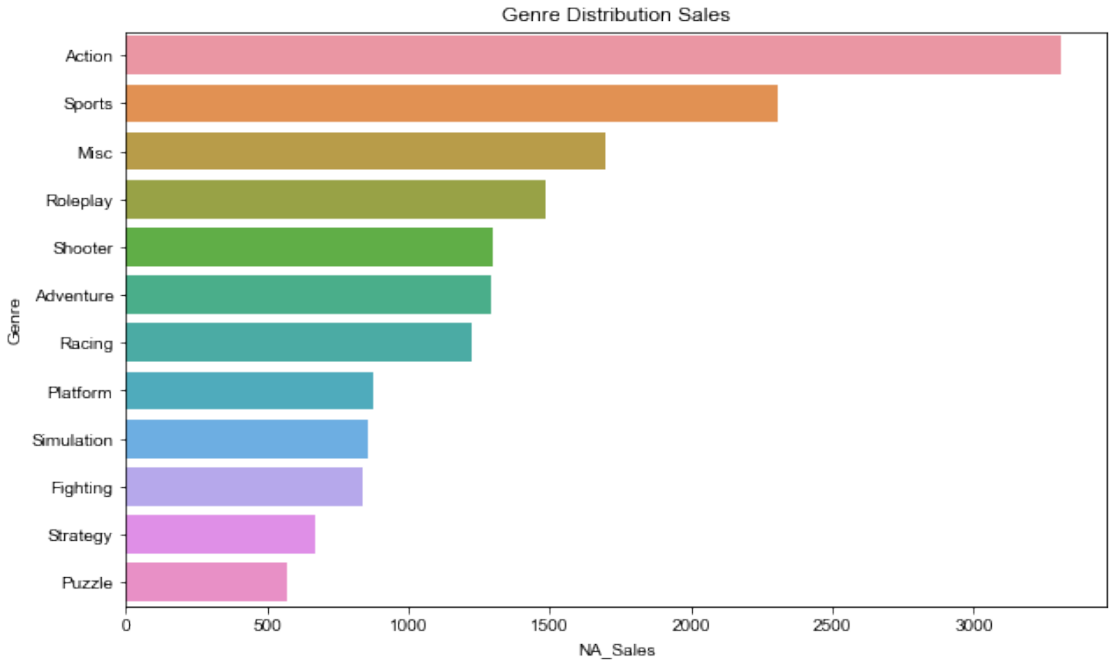
[19]:
```
genrecounts = genrecounts.rename(index={'Role-Playing':'Roleplay'})
genrecounts
```

[19]:
```
             NA_Sales  EU_Sales  JP_Sales  Other_Sales  Global_Sales
Genre
Action           3307      3307      3307         3307          3307
Adventure        1291      1291      1291         1291          1291
Fighting          837       837       837          837           837
Misc             1697      1697      1697         1697          1697
Platform          878       878       878          878           878
Puzzle            569       569       569          569           569
Racing           1225      1225      1225         1225          1225
Roleplay         1483      1483      1483         1483          1483
Shooter          1296      1296      1296         1296          1296
Simulation        855       855       855          855           855
Sports           2306      2306      2306         2306          2306
Strategy          672       672       672          672           672
```

[20]:
```
genrecounts = genrecounts.sort_values('NA_Sales', ascending=False)
genrecounts
```

[20]:
```
             NA_Sales  EU_Sales  JP_Sales  Other_Sales  Global_Sales
Genre
Action           3307      3307      3307         3307          3307
Sports           2306      2306      2306         2306          2306
Misc             1697      1697      1697         1697          1697
Roleplay         1483      1483      1483         1483          1483
Shooter          1296      1296      1296         1296          1296
Adventure        1291      1291      1291         1291          1291
Racing           1225      1225      1225         1225          1225
Platform          878       878       878          878           878
Simulation        855       855       855          855           855
Fighting          837       837       837          837           837
Strategy          672       672       672          672           672
Puzzle            569       569       569          569           569
```

[21]:
```
plt.figure(figsize=(10,6))
plt.title("Genre Distribution Sales")
sns.barplot(x=genrecounts['NA_Sales'], y=genrecounts.index)
sns.set_style("darkgrid")
```



[22]:
```
yeargrowth = df.drop(columns = ['Genre',␣
↪'Name','Platform','NA_Sales','EU_Sales', 'JP_Sales','NA_Sales',␣
↪'Other_Sales'])
yeargrowth
```

[22]:

	Year	Publisher	Global_Sales
0	2006.0	Nintendo	82.53
1	1985.0	Nintendo	40.24
2	2008.0	Nintendo	35.52
3	2009.0	Nintendo	32.77
4	1996.0	Nintendo	31.37
…	…	…	…
16714	2016.0	Tecmo Koei	0.01
16715	2006.0	Codemasters	0.01
16716	2016.0	Idea Factory	0.01
16717	2003.0	Wanadoo	0.01
16718	2016.0	Tecmo Koei	0.01

[16418 rows x 3 columns]

[23]:
```
nintendogrow = yeargrowth[yeargrowth['Publisher'] == 'Nintendo']
nintendogrow
```

[23]:

	Year	Publisher	Global_Sales
0	2006.0	Nintendo	82.53

```
1       1985.0  Nintendo         40.24
2       2008.0  Nintendo         35.52
3       2009.0  Nintendo         32.77
4       1996.0  Nintendo         31.37
…        …       …            …
16514   2015.0  Nintendo          0.01
16579   2008.0  Nintendo          0.01
16612   2015.0  Nintendo          0.01
16656   2016.0  Nintendo          0.01
16657   2011.0  Nintendo          0.01

[700 rows x 3 columns]
```

[24]:
```
nintendogrow = yeargrowth[yeargrowth['Publisher'] == 'Nintendo']
nintendogrow
```

[24]:
```
          Year Publisher  Global_Sales
0       2006.0  Nintendo         82.53
1       1985.0  Nintendo         40.24
2       2008.0  Nintendo         35.52
3       2009.0  Nintendo         32.77
4       1996.0  Nintendo         31.37
…        …       …            …
16514   2015.0  Nintendo          0.01
16579   2008.0  Nintendo          0.01
16612   2015.0  Nintendo          0.01
16656   2016.0  Nintendo          0.01
16657   2011.0  Nintendo          0.01

[700 rows x 3 columns]
```

[25]:
```
eagrow = yeargrowth[yeargrowth['Publisher'] == 'Electronic Arts']
eagrow
```

[25]:
```
          Year        Publisher  Global_Sales
77      2015.0  Electronic Arts          8.57
81      2012.0  Electronic Arts          8.16
85      2009.0  Electronic Arts          8.01
87      2015.0  Electronic Arts          7.98
94      2016.0  Electronic Arts          7.59
…        …               …            …
16380   2004.0  Electronic Arts          0.01
16489   2008.0  Electronic Arts          0.01
16491   2002.0  Electronic Arts          0.01
16510   2006.0  Electronic Arts          0.01
16670   1995.0  Electronic Arts          0.01
```

```
[1344 rows x 3 columns]
```

[26]:
```
activisiongrow = yeargrowth[yeargrowth['Publisher'] == 'Activision']
activisiongrow
```

[26]:
```
          Year   Publisher  Global_Sales
29      2011.0  Activision         14.73
31      2015.0  Activision         14.63
32      2010.0  Activision         14.61
34      2012.0  Activision         13.79
35      2012.0  Activision         13.67
…        …         …            …
16564   2012.0  Activision          0.01
16588   2012.0  Activision          0.01
16606   2010.0  Activision          0.01
16635   2014.0  Activision          0.01
16713   2008.0  Activision          0.01

[976 rows x 3 columns]
```

[27]:
```
sonygrow = yeargrowth[yeargrowth['Publisher'] == 'Sony Computer Entertainment']
sonygrow
```

[27]:
```
          Year                     Publisher  Global_Sales
28      2001.0  Sony Computer Entertainment         14.98
48      2004.0  Sony Computer Entertainment         11.66
52      1997.0  Sony Computer Entertainment         10.95
54      2010.0  Sony Computer Entertainment         10.70
65      1997.0  Sony Computer Entertainment          9.72
…        …                                …            …
16276   2013.0  Sony Computer Entertainment          0.01
16286   2009.0  Sony Computer Entertainment          0.01
16322   2011.0  Sony Computer Entertainment          0.01
16573   2010.0  Sony Computer Entertainment          0.01
16683   2006.0  Sony Computer Entertainment          0.01

[686 rows x 3 columns]
```

[28]:
```
fig, ax = plt.subplots(2,2,figsize=(20,12))

nintendogrow = nintendogrow.groupby('Year').sum()
eagrow = eagrow.groupby('Year').sum()
activisiongrow = activisiongrow.groupby('Year').sum()
sonygrow = sonygrow.groupby('Year').sum()

ax[0,0].plot(nintendogrow)
ax[0,0].set_title('Nintendo Sales Growth')
```

```
ax[0,1].plot(eagrow)
ax[0,1].set_title('EA Sales Growth')

ax[1,0].plot(activisiongrow)
ax[1,0].set_title('Activision Sales Growth')

ax[1,1].plot(sonygrow)
ax[1,1].set_title('Sony Sales Growth')
```

[28]:
```
Text(0.5, 1.0, 'Sony Sales Growth')
```

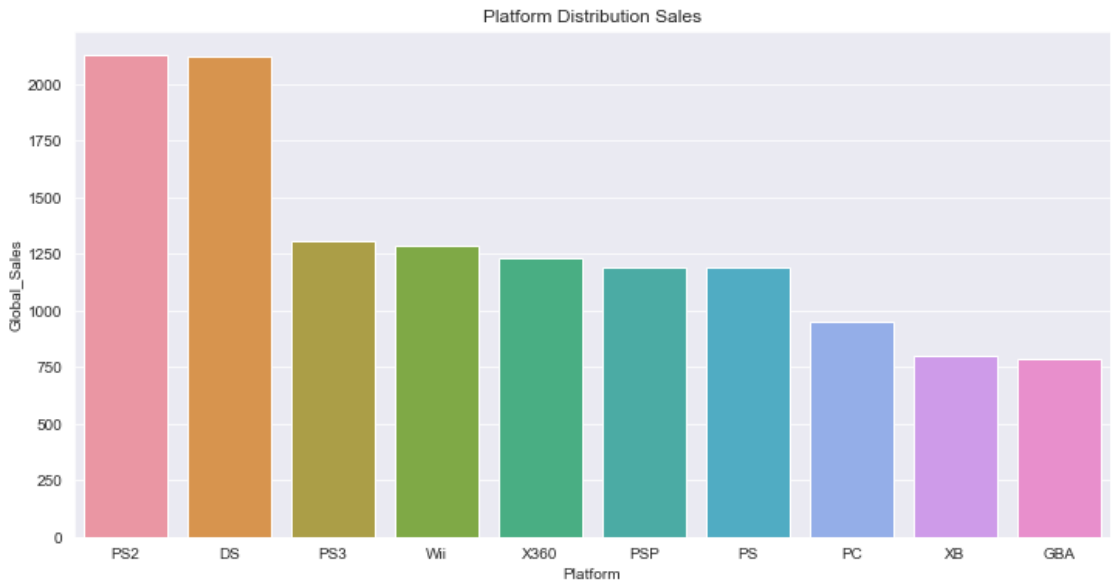
[29]:
```
platform = df.drop(columns = ['Genre',␣
↪'Name','Year','Publisher','NA_Sales','EU_Sales', 'JP_Sales','NA_Sales',␣
↪'Other_Sales'])
platform = platform.groupby('Platform').count()
platform = platform.sort_values('Global_Sales', ascending=False)
platform = platform.head(10)
# platform = platform['Platform'].value_counts()
# platform = platform.head(10)
platform
```

[29]:
```
          Global_Sales
Platform
PS2               2127
```

```
DS              2122
PS3             1306
Wii             1286
X360            1232
PSP             1193
PS              1190
PC               952
XB               803
GBA              786
```

[30]:
```
plt.figure(figsize=(12,6))
plt.title("Platform Distribution Sales")
diverging_colors = sns.color_palette("hls", 8)
sns.set_palette(diverging_colors)
sns.barplot(x=platform.index, y=platform['Global_Sales'])

# sns.set_style("darkgrid")
```

[30]:
```
<AxesSubplot:title={'center':'Platform Distribution Sales'}, xlabel='Platform',
ylabel='Global_Sales'>
```



[ ]:

[ ]: