Videogames

October 10, 2018

This dataset contains a list of video games with sales greater than 100,000 copies. It was generated by a scrape of vgchartz.com.

Fields include

- Rank - Ranking of overall sales
- Name - The games name
- Platform - Platform of the games release (i.e. PC,PS4, etc.)
- Year - 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.

```
In [46]: import pandas as pd
         import numpy as np
         import matplotlib.pyplot as plt
```

```
In [47]: path = '/Users/manuelgijonagudo/Documents/Programación/GIT/Data/vgsales.csv'
         data = pd.read_csv(path)
```

0.0.1 Exploring the data set:

```
In [48]: data.head()
```

Out[48]:

	Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
0	1	Wii Sports	Wii	2006.0	Sports	Nintendo	41.49	29.02	3.77	8.46	82.74
1	2	Super Mario Bros.	NES	1985.0	Platform	Nintendo	29.08	3.58	6.81	0.77	40.24
2	3	Mario Kart Wii	Wii	2008.0	Racing	Nintendo	15.85	12.88	3.79	3.31	35.82
3	4	Wii Sports Resort	Wii	2009.0	Sports	Nintendo	15.75	11.01	3.28	2.96	33.00
4	5	Pokemon Red/Pokemon Blue	GB	1996.0	Role-Playing	Nintendo	11.27	8.89	10.22	1.00	31.37

In [49]: `data.tail()`

Out[49]:

	Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
16593	16596	Woody Woodpecker in Crazy Castle 5	GBA	2002.0	Platform	Kemco	0.01	0.00	0.0	0.0	0.01
16594	16597	Men in Black II: Alien Escape	GC	2003.0	Shooter	Infogrames	0.01	0.00	0.0	0.0	0.01
16595	16598	SCORE International Baja 1000: The Official Game	PS2	2008.0	Racing	Activision	0.00	0.00	0.0	0.0	0.01
16596	16599	Know How 2	DS	2010.0	Puzzle	7G//AMES	0.00	0.01	0.0	0.0	0.01
16597	16600	Spirits & Spells	GBA	2003.0	Platform	Wanadoo	0.01	0.00	0.0	0.0	0.01

In [50]: `data.dtypes`

Out[50]:

Rank	int64
Name	object
Platform	object
Year	float64
Genre	object
Publisher	object
NA_Sales	float64
EU_Sales	float64
JP_Sales	float64
Other_Sales	float64

```
Global_Sales    float64
dtype: object
```

```
In [51]: data.index
```

```
Out[51]: RangeIndex(start=0, stop=16598, step=1)
```

```
In [52]: data.columns
```

```
Out[52]: Index(['Rank', 'Name', 'Platform', 'Year', 'Genre', 'Publisher', 'NA_Sales',
                'EU_Sales', 'JP_Sales', 'Other_Sales', 'Global_Sales'],
               dtype='object')
```

```
In [53]: data.values
```

```
Out[53]: array([[1, 'Wii Sports', 'Wii', ..., 3.77, 8.46, 82.74],
                [2, 'Super Mario Bros.', 'NES', ..., 6.81, 0.77, 40.24],
                [3, 'Mario Kart Wii', 'Wii', ..., 3.79, 3.31, 35.82],
                ...,
                [16598, 'SCORE International Baja 1000: The Official Game', 'PS2',
                 ..., 0.0, 0.0, 0.01],
                [16599, 'Know How 2', 'DS', ..., 0.0, 0.0, 0.01],
                [16600, 'Spirits & Spells', 'GBA', ..., 0.0, 0.0, 0.01]], dtype=object)
```

```
In [54]: data.describe()
```

```
Out[54]:                Rank          Year      NA_Sales      EU_Sales      JP_Sales  \
         count  16598.000000  16327.000000  16598.000000  16598.000000  16598.000000
         mean    8300.605254   2006.406443      0.264667      0.146652      0.077782
         std     4791.853933      5.828981      0.816683      0.505351      0.309291
         min        1.000000   1980.000000      0.000000      0.000000      0.000000
         25%     4151.250000   2003.000000      0.000000      0.000000      0.000000
         50%     8300.500000   2007.000000      0.080000      0.020000      0.000000
         75%    12449.750000   2010.000000      0.240000      0.110000      0.040000
         max    16600.000000   2020.000000     41.490000     29.020000     10.220000

                Other_Sales  Global_Sales
         count  16598.000000  16598.000000
         mean       0.048063      0.537441
         std        0.188588      1.555028
         min        0.000000      0.010000
         25%        0.000000      0.060000
         50%        0.010000      0.170000
         75%        0.040000      0.470000
         max       10.570000     82.740000
```

```
In [55]: data.T # transpose the data
```

Out[55]:

	0	1	2	\
Rank	1	2	3	
Name	Wii Sports	Super Mario Bros.	Mario Kart Wii	
Platform	Wii	NES	Wii	
Year	2006	1985	2008	
Genre	Sports	Platform	Racing	
Publisher	Nintendo	Nintendo	Nintendo	
NA_Sales	41.49	29.08	15.85	
EU_Sales	29.02	3.58	12.88	
JP_Sales	3.77	6.81	3.79	
Other_Sales	8.46	0.77	3.31	
Global_Sales	82.74	40.24	35.82	

	3	4	5	\
Rank	4	5	6	
Name	Wii Sports Resort	Pokemon Red/Pokemon Blue	Tetris	
Platform	Wii	GB	GB	
Year	2009	1996	1989	
Genre	Sports	Role-Playing	Puzzle	
Publisher	Nintendo	Nintendo	Nintendo	
NA_Sales	15.75	11.27	23.2	
EU_Sales	11.01	8.89	2.26	
JP_Sales	3.28	10.22	4.22	
Other_Sales	2.96	1	0.58	
Global_Sales	33	31.37	30.26	

	6	7	8	\
Rank	7	8	9	
Name	New Super Mario Bros.	Wii Play	New Super Mario Bros. Wii	
Platform	DS	Wii	Wii	
Year	2006	2006	2009	
Genre	Platform	Misc	Platform	
Publisher	Nintendo	Nintendo	Nintendo	
NA_Sales	11.38	14.03	14.59	
EU_Sales	9.23	9.2	7.06	
JP_Sales	6.5	2.93	4.7	
Other_Sales	2.9	2.85	2.26	
Global_Sales	30.01	29.02	28.62	

	9	...	16588	\
Rank	10	...	16591	
Name	Duck Hunt	...	Mega Brain Boost	
Platform	NES	...	DS	
Year	1984	...	2008	
Genre	Shooter	...	Puzzle	
Publisher	Nintendo	...	Majesco Entertainment	
NA_Sales	26.93	...	0.01	
EU_Sales	0.63	...	0	

```
JP_Sales           0.28        ...                            0
Other_Sales        0.47        ...                            0
Global_Sales      28.31        ...                         0.01

                                                          16589  \
Rank                                                      16592
Name          Chou Ezaru wa Akai Hana: Koi wa Tsuki ni Shiru...
Platform                                                    PSV
Year                                                       2016
Genre                                                    Action
Publisher                                        dramatic create
NA_Sales                                                    0
EU_Sales                                                    0
JP_Sales                                                 0.01
Other_Sales                                                 0
Global_Sales                                             0.01

                                                          16590  \
Rank                                                      16593
Name          Eiyuu Densetsu: Sora no Kiseki Material Collec...
Platform                                                    PSP
Year                                                       2007
Genre                                                Role-Playing
Publisher                                       Falcom Corporation
NA_Sales                                                    0
EU_Sales                                                    0
JP_Sales                                                 0.01
Other_Sales                                                 0
Global_Sales                                             0.01

                           16591       16592  \
Rank                       16594       16595
Name          Myst IV: Revelation    Plushees
Platform                       PC          DS
Year                         2004        2008
Genre                   Adventure  Simulation
Publisher                 Ubisoft   Destineer
NA_Sales                     0.01        0.01
EU_Sales                        0           0
JP_Sales                        0           0
Other_Sales                     0           0
Global_Sales                 0.01        0.01

                                        16593  \
Rank                                    16596
Name          Woody Woodpecker in Crazy Castle 5
Platform                                   GBA
Year                                      2002
```

Genre					Platform
Publisher					Kemco
NA_Sales					0.01
EU_Sales					0
JP_Sales					0
Other_Sales					0
Global_Sales					0.01

					16594 \
Rank					16597
Name	Men in Black II: Alien Escape				
Platform					GC
Year					2003
Genre					Shooter
Publisher					Infogrames
NA_Sales					0.01
EU_Sales					0
JP_Sales					0
Other_Sales					0
Global_Sales					0.01

					16595	16596 \
Rank					16598	16599
Name	SCORE International Baja 1000: The Official Game					Know How 2
Platform					PS2	DS
Year					2008	2010
Genre					Racing	Puzzle
Publisher					Activision	7G//AMES
NA_Sales					0	0
EU_Sales					0	0.01
JP_Sales					0	0
Other_Sales					0	0
Global_Sales					0.01	0.01

					16597
Rank					16600
Name	Spirits & Spells				
Platform					GBA
Year					2003
Genre					Platform
Publisher					Wanadoo
NA_Sales					0.01
EU_Sales					0
JP_Sales					0
Other_Sales					0
Global_Sales					0.01

[11 rows x 16598 columns]

```
In [56]: data.sort_index(axis = 1, ascending = False) # sorting by axis
```

Out[56]:

	Year	Rank	Publisher	Platform	Other_Sales
0	2006.0	1	Nintendo	Wii	8.46
1	1985.0	2	Nintendo	NES	0.77
2	2008.0	3	Nintendo	Wii	3.31
3	2009.0	4	Nintendo	Wii	2.96
4	1996.0	5	Nintendo	GB	1.00
5	1989.0	6	Nintendo	GB	0.58
6	2006.0	7	Nintendo	DS	2.90
7	2006.0	8	Nintendo	Wii	2.85
8	2009.0	9	Nintendo	Wii	2.26
9	1984.0	10	Nintendo	NES	0.47
10	2005.0	11	Nintendo	DS	2.75
11	2005.0	12	Nintendo	DS	1.92
12	1999.0	13	Nintendo	GB	0.71
13	2007.0	14	Nintendo	Wii	2.15
14	2009.0	15	Nintendo	Wii	1.79
15	2010.0	16	Microsoft Game Studios	X360	1.67
16	2013.0	17	Take-Two Interactive	PS3	4.14
17	2004.0	18	Take-Two Interactive	PS2	10.57
18	1990.0	19	Nintendo	SNES	0.55
19	2005.0	20	Nintendo	DS	2.05
20	2006.0	21	Nintendo	DS	1.37
21	1989.0	22	Nintendo	GB	0.42
22	1988.0	23	Nintendo	NES	0.46
23	2013.0	24	Take-Two Interactive	X360	1.38
24	2002.0	25	Take-Two Interactive	PS2	1.78
25	2002.0	26	Nintendo	GBA	0.50
26	2010.0	27	Nintendo	DS	0.82
27	2005.0	28	Nintendo	DS	1.18
28	2001.0	29	Sony Computer Entertainment	PS2	1.16
29	2011.0	30	Activision	X360	1.32
...	...	...	...	...	...
16568	2006.0	16571	Sony Computer Entertainment	PSP	0.00
16569	2016.0	16572	Capcom	XOne	0.00
16570	2016.0	16573	UIG Entertainment	PS4	0.00
16571	2013.0	16574	Prototype	PSP	0.00
16572	2008.0	16575	Kadokawa Shoten	PS2	0.00
16573	2007.0	16576	Popcorn Arcade	Wii	0.00
16574	2008.0	16577	Takara Tomy	Wii	0.00
16575	2008.0	16578	Tecmo Koei	PSP	0.00
16576	2016.0	16579	Alternative Software	XOne	0.00
16577	2009.0	16580	Codemasters	PC	0.00
16578	2011.0	16581	Mastiff	3DS	0.00
16579	2000.0	16582	Infogrames	N64	0.00
16580	2008.0	16583	Kadokawa Shoten	PS2	0.00
16581	2011.0	16584	Unknown	Wii	0.00

16582	2001.0	16585	Titus	GBA	0.00
16583	1999.0	16586	Virgin Interactive	N64	0.00
16584	2003.0	16587	Ubisoft	GC	0.00
16585	2011.0	16588	Destineer	PC	0.00
16586	2009.0	16589	Deep Silver	DS	0.00
16587	2009.0	16590	Hudson Soft	DS	0.00
16588	2008.0	16591	Majesco Entertainment	DS	0.00
16589	2016.0	16592	dramatic create	PSV	0.00
16590	2007.0	16593	Falcom Corporation	PSP	0.00
16591	2004.0	16594	Ubisoft	PC	0.00
16592	2008.0	16595	Destineer	DS	0.00
16593	2002.0	16596	Kemco	GBA	0.00
16594	2003.0	16597	Infogrames	GC	0.00
16595	2008.0	16598	Activision	PS2	0.00
16596	2010.0	16599	7G//AMES	DS	0.00
16597	2003.0	16600	Wanadoo	GBA	0.00

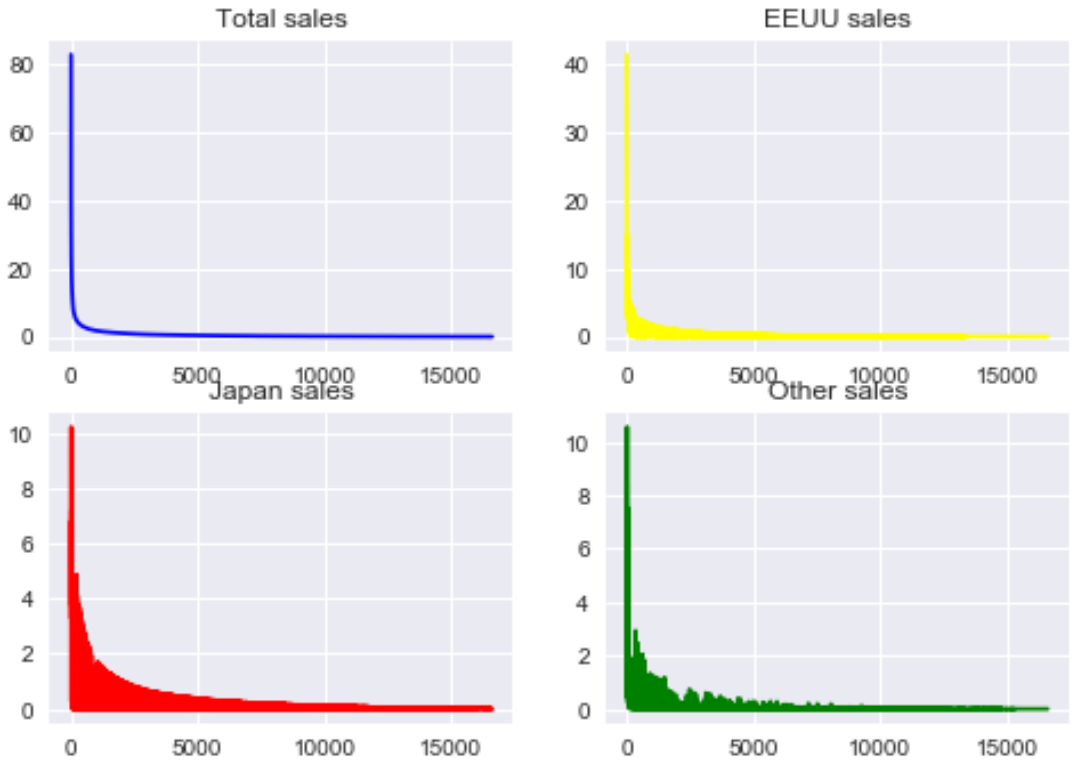
	Name	NA_Sales	JP_Sales	\
0	Wii Sports	41.49	3.77	
1	Super Mario Bros.	29.08	6.81	
2	Mario Kart Wii	15.85	3.79	
3	Wii Sports Resort	15.75	3.28	
4	Pokemon Red/Pokemon Blue	11.27	10.22	
5	Tetris	23.20	4.22	
6	New Super Mario Bros.	11.38	6.50	
7	Wii Play	14.03	2.93	
8	New Super Mario Bros. Wii	14.59	4.70	
9	Duck Hunt	26.93	0.28	
10	Nintendogs	9.07	1.93	
11	Mario Kart DS	9.81	4.13	
12	Pokemon Gold/Pokemon Silver	9.00	7.20	
13	Wii Fit	8.94	3.60	
14	Wii Fit Plus	9.09	2.53	
15	Kinect Adventures!	14.97	0.24	
16	Grand Theft Auto V	7.01	0.97	
17	Grand Theft Auto: San Andreas	9.43	0.41	
18	Super Mario World	12.78	3.54	
19	Brain Age: Train Your Brain in Minutes a Day	4.75	4.16	
20	Pokemon Diamond/Pokemon Pearl	6.42	6.04	
21	Super Mario Land	10.83	4.18	
22	Super Mario Bros. 3	9.54	3.84	
23	Grand Theft Auto V	9.63	0.06	
24	Grand Theft Auto: Vice City	8.41	0.47	
25	Pokemon Ruby/Pokemon Sapphire	6.06	5.38	
26	Pokemon Black/Pokemon White	5.57	5.65	
27	Brain Age 2: More Training in Minutes a Day	3.44	5.32	
28	Gran Turismo 3: A-Spec	6.85	1.87	
29	Call of Duty: Modern Warfare 3	9.03	0.13	

```
...                                                ...      ...      ...
16568                                       XI Coliseum     0.00     0.01
16569                                Resident Evil 4 HD     0.01     0.00
16570                      Farming 2017 - The Simulation     0.00     0.00
16571         Grisaia no Kajitsu: La Fruit de la Grisaia     0.00     0.01
16572                  Scarlett: Nichijou no Kyoukaisen     0.00     0.01
16573                              Mini Desktop Racing      0.01     0.00
16574  Yattaman Wii: BikkuriDokkiri Machine de Mou Ra...     0.00     0.01
16575                             Neo Angelique Special     0.00     0.01
16576                                 Rugby Challenge 3     0.00     0.00
16577                                         Damnation     0.00     0.00
16578                      Outdoors Unleashed: Africa 3D     0.01     0.00
16579                                 PGA European Tour      0.01     0.00
16580                                         Real Rode     0.00     0.01
16581                                         Fit & Fun     0.00     0.00
16582                                   Planet Monsters     0.01     0.00
16583                                    Carmageddon 64     0.01     0.00
16584                                  Bust-A-Move 3000     0.01     0.00
16585                                           Breach     0.01     0.00
16586                    Secret Files 2: Puritas Cordis     0.00     0.00
16587                         Mezase!! Tsuri Master DS      0.00     0.01
16588                                  Mega Brain Boost     0.01     0.00
16589  Chou Ezaru wa Akai Hana: Koi wa Tsuki ni Shiru...     0.00     0.01
16590  Eiyuu Densetsu: Sora no Kiseki Material Collec...     0.00     0.01
16591                               Myst IV: Revelation      0.01     0.00
16592                                           Plushees     0.01     0.00
16593                 Woody Woodpecker in Crazy Castle 5     0.01     0.00
16594                      Men in Black II: Alien Escape     0.01     0.00
16595   SCORE International Baja 1000: The Official Game     0.00     0.00
16596                                         Know How 2     0.00     0.00
16597                                   Spirits & Spells     0.01     0.00

       Global_Sales         Genre  EU_Sales
0             82.74        Sports     29.02
1             40.24      Platform      3.58
2             35.82        Racing     12.88
3             33.00        Sports     11.01
4             31.37  Role-Playing      8.89
5             30.26        Puzzle      2.26
6             30.01      Platform      9.23
7             29.02          Misc      9.20
8             28.62      Platform      7.06
9             28.31       Shooter      0.63
10            24.76    Simulation     11.00
11            23.42        Racing      7.57
12            23.10  Role-Playing      6.18
13            22.72        Sports      8.03
14            22.00        Sports      8.59
```

15	21.82	Misc	4.94
16	21.40	Action	9.27
17	20.81	Action	0.40
18	20.61	Platform	3.75
19	20.22	Misc	9.26
20	18.36	Role-Playing	4.52
21	18.14	Platform	2.71
22	17.28	Platform	3.44
23	16.38	Action	5.31
24	16.15	Action	5.49
25	15.85	Role-Playing	3.90
26	15.32	Role-Playing	3.28
27	15.30	Puzzle	5.36
28	14.98	Racing	5.09
29	14.76	Shooter	4.28
...	...	...	...
16568	0.01	Puzzle	0.00
16569	0.01	Shooter	0.00
16570	0.01	Simulation	0.01
16571	0.01	Adventure	0.00
16572	0.01	Adventure	0.00
16573	0.01	Racing	0.00
16574	0.01	Racing	0.00
16575	0.01	Adventure	0.00
16576	0.01	Sports	0.01
16577	0.01	Shooter	0.01
16578	0.01	Sports	0.00
16579	0.01	Sports	0.00
16580	0.01	Adventure	0.00
16581	0.01	Sports	0.01
16582	0.01	Action	0.00
16583	0.01	Action	0.00
16584	0.01	Puzzle	0.00
16585	0.01	Shooter	0.00
16586	0.01	Adventure	0.01
16587	0.01	Sports	0.00
16588	0.01	Puzzle	0.00
16589	0.01	Action	0.00
16590	0.01	Role-Playing	0.00
16591	0.01	Adventure	0.00
16592	0.01	Simulation	0.00
16593	0.01	Platform	0.00
16594	0.01	Shooter	0.00
16595	0.01	Racing	0.00
16596	0.01	Puzzle	0.01
16597	0.01	Platform	0.00

[16598 rows x 11 columns]

```
In [57]: data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16598 entries, 0 to 16597
Data columns (total 11 columns):
Rank            16598 non-null int64
Name            16598 non-null object
Platform        16598 non-null object
Year            16327 non-null float64
Genre           16598 non-null object
Publisher       16540 non-null object
NA_Sales        16598 non-null float64
EU_Sales        16598 non-null float64
JP_Sales        16598 non-null float64
Other_Sales     16598 non-null float64
Global_Sales    16598 non-null float64
dtypes: float64(6), int64(1), object(4)
memory usage: 1.4+ MB


In [58]: ventas_totales = data['Global_Sales']
         ventas_EEUU = data['NA_Sales']
         ventas_Japon = data['JP_Sales']
         ventas_otros = data['Other_Sales']

In [59]: import matplotlib.pyplot as plt

         plt.figure(1)

         plt.subplot(221)
         plt.title('Total sales')
         plt.plot(ventas_totales, color = 'blue')

         plt.subplot(222)
         plt.title('EEUU sales')
         plt.plot(ventas_EEUU, color = 'yellow')

         plt.subplot(223)
         plt.title('Japan sales')
         plt.plot(ventas_Japon, color = 'red')

         plt.subplot(224)
         plt.title('Other sales')
         plt.plot(ventas_otros, color = 'green')

         plt.show()
```

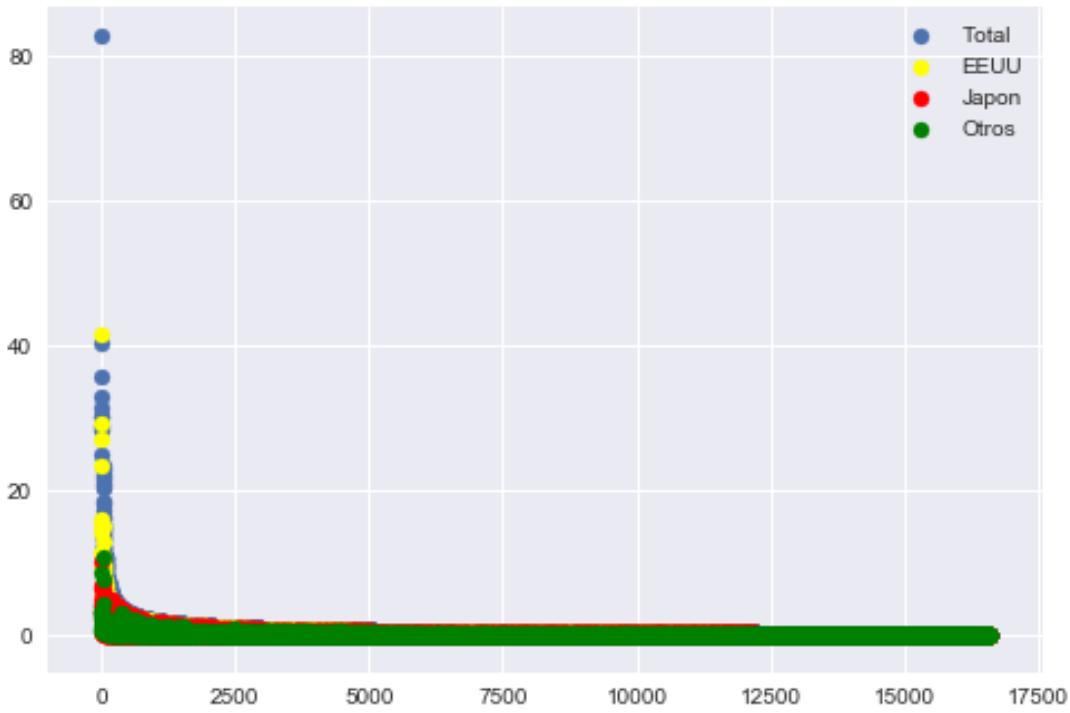



```
In [60]: print(ventas_totales.describe())
         print('\n')
         print(ventas_EEUU.describe())
         print('\n')
         print(ventas_Japon.describe())
         print('\n')
         print(ventas_otros.describe())
         print('\n')
```

```
count    16598.000000
mean         0.537441
std          1.555028
min          0.010000
25%          0.060000
50%          0.170000
75%          0.470000
max         82.740000
Name: Global_Sales, dtype: float64


count    16598.000000
mean         0.264667
```

```
std          0.816683
min          0.000000
25%          0.000000
50%          0.080000
75%          0.240000
max         41.490000
Name: NA_Sales, dtype: float64


count    16598.000000
mean         0.077782
std          0.309291
min          0.000000
25%          0.000000
50%          0.000000
75%          0.040000
max         10.220000
Name: JP_Sales, dtype: float64


count    16598.000000
mean         0.048063
std          0.188588
min          0.000000
25%          0.000000
50%          0.010000
75%          0.040000
max         10.570000
Name: Other_Sales, dtype: float64
```

```
In [61]: plt.scatter([x for x in range(0, len(ventas_totales))], ventas_totales, label = 'Total
         plt.scatter([x for x in range(0, len(ventas_EEUU))], ventas_EEUU, label = 'EEUU', colo
         plt.scatter([x for x in range(0, len(ventas_Japon))], ventas_Japon, label = 'Japon', c
         plt.scatter([x for x in range(0, len(ventas_otros))], ventas_otros, label = 'Otros', c

         plt.legend()
         plt.show()
```




```
In [62]: years = data['Year']

In [63]: years.describe()

Out[63]: count    16327.000000
         mean      2006.406443
         std          5.828981
         min       1980.000000
         25%       2003.000000
         50%       2007.000000
         75%       2010.000000
         max       2020.000000
         Name: Year, dtype: float64

In [64]: years.max()  # ????!!!!

Out[64]: 2020.0

In [65]: data[data['Year'] >= 2017]

Out[65]:         Rank                                                Name Platform  \
         5957    5959                             Imagine: Makeup Artist       DS
         14390  14393  Phantasy Star Online 2 Episode 4: Deluxe Package      PS4
         16241  16244  Phantasy Star Online 2 Episode 4: Deluxe Package      PSV
```

```
16438  16441                  Brothers Conflict: Precious Baby      PSV

         Year         Genre     Publisher  NA_Sales  EU_Sales  JP_Sales  \
5957   2020.0    Simulation       Ubisoft      0.27       0.0      0.00
14390  2017.0  Role-Playing          Sega      0.00       0.0      0.03
16241  2017.0  Role-Playing          Sega      0.00       0.0      0.01
16438  2017.0        Action  Idea Factory      0.00       0.0      0.01

       Other_Sales  Global_Sales
5957          0.02          0.29
14390         0.00          0.03
16241         0.00          0.01
16438         0.00          0.01
```

In [66]:
```
# limpio los datos antes de continuar

data = data[data['Year'] < 2017]
```

In [67]:
```
years = data['Year']
years.describe()
```

Out[67]:
```
count    16323.000000
mean      2006.403664
std          5.826954
min       1980.000000
25%       2003.000000
50%       2007.000000
75%       2010.000000
max       2016.000000
Name: Year, dtype: float64
```

In [68]:
```
plt.title('Distribution by year')
plt.hist(years, bins = 30, color = 'red')
plt.xlabel('Year')
plt.ylabel('Number of games in the ranking')

plt.show()
```

```
---------------------------------------------------------------------------

TypeError                                 Traceback (most recent call last)

<ipython-input-68-614a393422b7> in <module>()
  1 plt.title('Distribution by year')
  2 plt.hist(years, bins = 30, color = 'red')
----> 3 plt.xlabel('Year')
  4 plt.ylabel('Number of games in the ranking')
  5
```

```
TypeError: 'str' object is not callable
```

0.1 Action Videogames

```
In [ ]: action = data[data['Genre'] == 'Action']
        action.head()
```

```
In [ ]: action = action.reset_index(level = 0, drop = True)
        action.head()
```

```
In [ ]: # no tenemos ya necesidad de la variable género, la eliminaremos del data_frame:
        # axis = 0 para filas
        # axis = 1 para columnas
        # inplace = True para sobreescribir el data frame, no tenemos que reasignarlo
        action.drop('Genre', axis = 1, inplace = True)
```

```
In [ ]: action.head()
```

```
In [ ]: plataformas = list( set( action['Platform'] ) )
        desarrollador = list( set( action['Publisher']))
```

```
In [ ]: action['Year'].describe()
```

```
In [ ]: plt.title('Distribution by year')
        plt.hist(action['Year'], bins = 30, color = 'yellow')
        plt.xlabel('Year')
        plt.ylabel('Number of games in the ranking')

        plt.show()
```

```
In [ ]: # comparamos con el total

        plt.figure(1)

        plt.subplot(121)
        plt.title('Distribution by year')
        plt.hist(years, bins = 30, color = 'red')
        plt.xlabel('Year')
        plt.ylabel('Number of games in the ranking')

        plt.subplot(122)
        plt.title('Action Videogames')
        plt.hist(action['Year'], bins = 30, color = 'yellow')


        plt.show()
```

0.2 By Platform

In [ ]:
```
# https://www.kaggle.com/garfieldliang/video-games-analysis

plataforma_genero = pd.crosstab(data.Platform, data.Genre)  # (index, values)
plataforma_genero_total = plataforma_genero.sum(axis = 1).sort_values(ascending = Fals
plataforma_genero.head()
```

In [ ]:
```
import seaborn as sns

plt.figure(figsize = (8, 6))
sns.barplot(y = plataforma_genero_total.index, x = plataforma_genero_total.values, ori
plt.ylabel = 'Plataforma'
plt.xlabel = 'Cantidad de juegos'
plt.show()
```

In [ ]:
```
plataforma_genero.head()
```

In [ ]:
```
# así podremos hacernos una idea del periodo de servicio de cada plataforma
plataforma_year = pd.crosstab(data.Year, data.Platform)
```

In [69]:
```
plataforma_year.head()
```

Out[69]:

Platform	2600	3DO	3DS	DC	DS	GB	GBA	GC	GEN	GG	...	SAT	SCD	SNES	\
Year											...				
1980.0	9	0	0	0	0	0	0	0	0	0	...	0	0	0	
1981.0	46	0	0	0	0	0	0	0	0	0	...	0	0	0	
1982.0	36	0	0	0	0	0	0	0	0	0	...	0	0	0	
1983.0	11	0	0	0	0	0	0	0	0	0	...	0	0	0	
1984.0	1	0	0	0	0	0	0	0	0	0	...	0	0	0	

Platform	TG16	WS	Wii	WiiU	X360	XB	XOne
Year							
1980.0	0	0	0	0	0	0	0
1981.0	0	0	0	0	0	0	0
1982.0	0	0	0	0	0	0	0
1983.0	0	0	0	0	0	0	0
1984.0	0	0	0	0	0	0	0

[5 rows x 31 columns]

In [70]:
```
sns.heatmap(plataforma_year)

plt.show()
```

In [71]:
```
# (index, values)

plataforma_Global = pd.crosstab(data.Platform, data.Global_Sales)
plataforma_Global.head()
```

Out[71]:

Global_Sales	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	\
Platform										
2600	0	0	0	0	0	0	1	1	0	
3DO	0	2	0	0	0	1	0	0	0	
3DS	23	25	39	24	16	23	28	20	14	
DC	0	1	0	3	3	4	2	3	2	
DS	88	158	128	110	121	110	96	84	67	

Global_Sales	0.10	...	28.31	28.62	29.02	30.01	30.26	31.37	33.00	\
Platform		...								
2600	0	...	0	0	0	0	0	0	0	
3DO	0	...	0	0	0	0	0	0	0	
3DS	12	...	0	0	0	0	0	0	0	
DC	4	...	0	0	0	0	0	0	0	
DS	79	...	0	0	0	1	0	0	0	

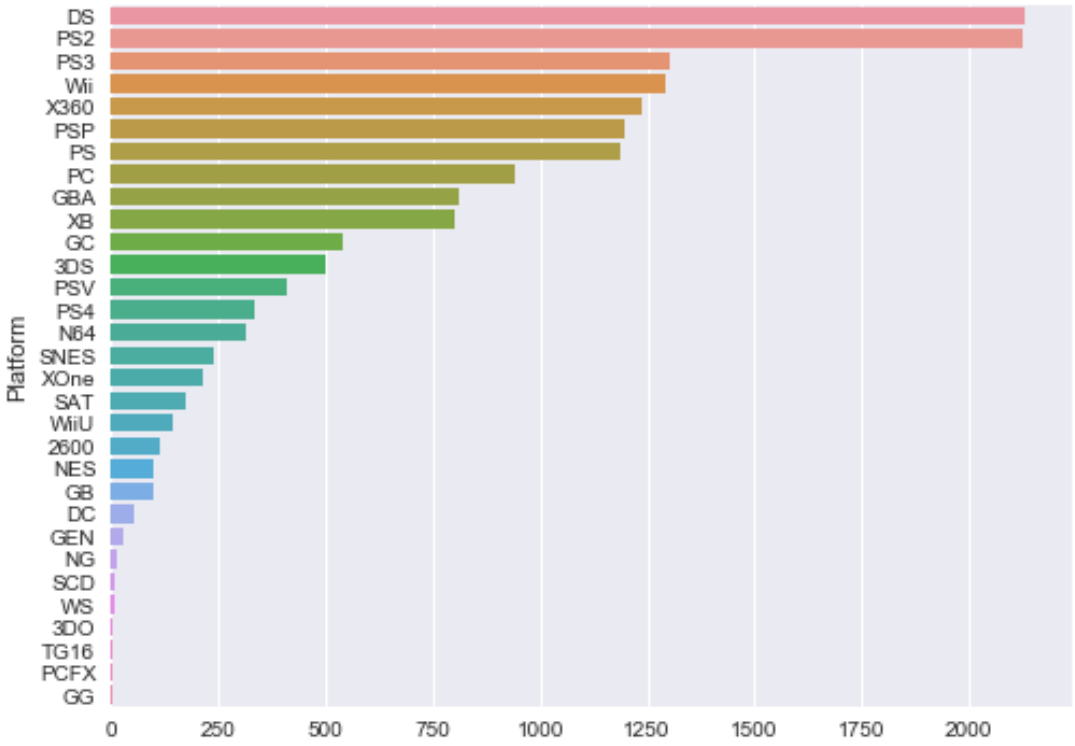
```
Global_Sales  35.82  40.24  82.74
Platform
2600              0      0      0
3DO               0      0      0
3DS               0      0      0
DC                0      0      0
DS                0      0      0

[5 rows x 621 columns]
```

In [72]:
```
plataforma_Global_total = plataforma_genero.sum(axis = 1).sort_values(ascending = Fals
plataforma_Global_total
```

Out[72]:
```
Platform
DS      2132
PS2     2127
PS3     1304
Wii     1290
X360    1235
PSP     1197
PS      1189
PC       943
GBA      811
XB       803
GC       542
3DS      500
PSV      410
PS4      335
N64      316
SNES     239
XOne     213
SAT      173
WiiU     143
2600     116
NES       98
GB        97
DC        52
GEN       27
NG        12
SCD        6
WS         6
3DO        3
TG16       2
PCFX       1
GG         1
dtype: int64
```

In [73]:
```
plt.figure(figsize = (8, 6))
```

```
sns.barplot(y = plataforma_Global_total.index, x = plataforma_Global_total.values, ori
plt.ylabel = 'Plataforma'
plt.xlabel = 'Ventas totales'
plt.show()
# coincide con el gráfico de cantidad de juegos por género y plataforma
'''
LAS VENTAS NO ESTÁN EN UNIDADES MONETARIAS, SINO EN UNIDADES VENDIDAS !!!!
'''
```




```
Out[73]: '\nLAS VENTAS NO ESTÁN EN UNIDADES MONETARIAS, SINO EN UNIDADES VENDIDAS !!!!\n'
```

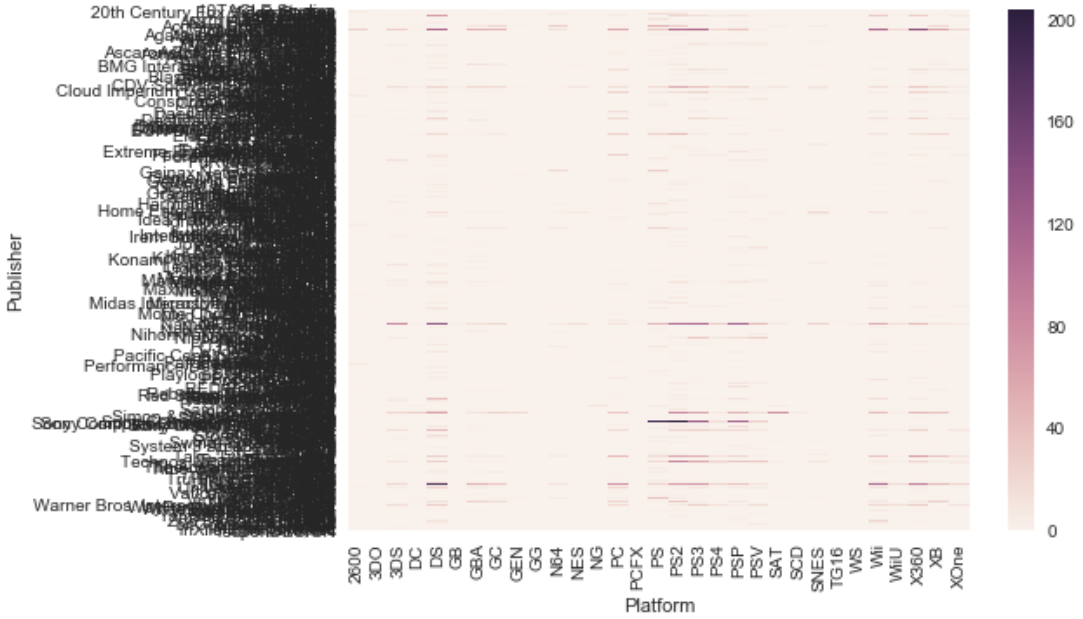
0.3 By Publisher

```
In [74]: publisher_platform = pd.crosstab(data.Publisher, data.Platform)
         publisher_platform.head()
```

```
Out[74]: Platform                    2600  3DO  3DS  DC  DS  GB  GBA  GC  GEN  GG  \
         Publisher
         10TACLE Studios                0    0    0   0   2   0    0   0    0   0
         1C Company                     0    0    0   0   0   0    0   0    0   0
         20th Century Fox Video Games   5    0    0   0   0   0    0   0    0   0
         2D Boy                         0    0    0   0   0   0    0   0    0   0
         3DO                            0    0    0   0   0   0    0   1    0   0
```

```
Platform                      ...   SAT  SCD  SNES  TG16  WS  Wii  WiiU  X360  \
Publisher                     ...
10TACLE Studios               ...     0    0     0     0   0    0     0     0
1C Company                    ...     0    0     0     0   0    0     0     0
20th Century Fox Video Games  ...     0    0     0     0   0    0     0     0
2D Boy                        ...     0    0     0     0   0    0     0     0
3DO                           ...     0    0     0     0   0    0     0     0

Platform                      XB  XOne
Publisher
10TACLE Studios                0     0
1C Company                     0     0
20th Century Fox Video Games   0     0
2D Boy                         0     0
3DO                            1     0

[5 rows x 31 columns]
```

In [77]:
```
sns.heatmap(publisher_platform)

plt.show()
```



In [75]:
```
publisher_genre = pd.crosstab(data.Publisher, data.Genre)
publisher_genre.head()
```

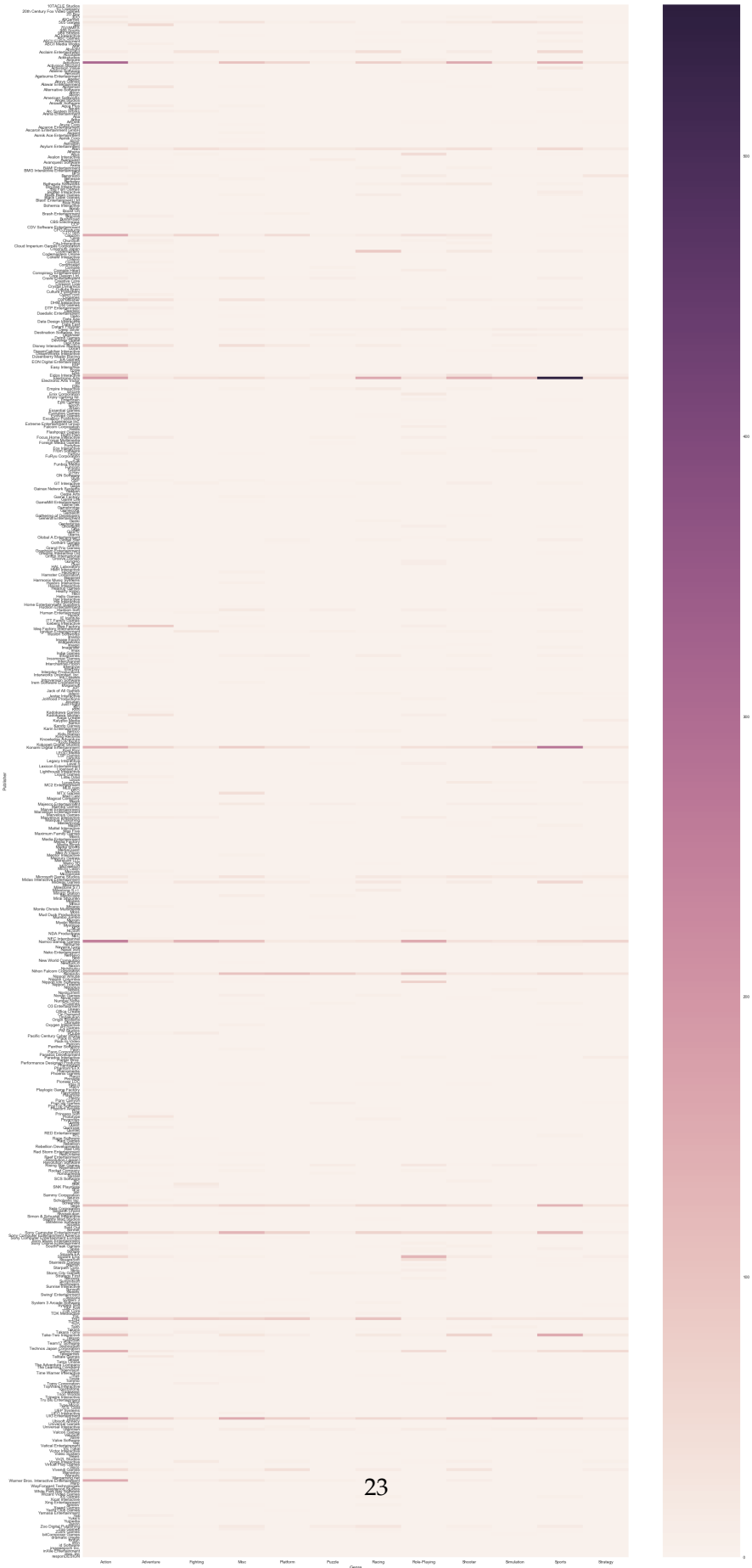
Out[75]:

Genre	Action	Adventure	Fighting	Misc	Platform
Publisher					
10TACLE Studios	0	1	0	0	0
1C Company	0	0	0	0	0
20th Century Fox Video Games	4	0	0	0	0
2D Boy	0	0	0	0	0
3DO	17	3	1	0	1

Genre	Puzzle	Racing	Role-Playing	Shooter
Publisher				
10TACLE Studios	1	0	0	0
1C Company	0	1	1	0
20th Century Fox Video Games	0	0	0	1
2D Boy	1	0	0	0
3DO	1	0	1	5

Genre	Simulation	Sports	Strategy
Publisher			
10TACLE Studios	0	0	1
1C Company	0	0	1
20th Century Fox Video Games	0	0	0
2D Boy	0	0	0
3DO	0	6	1

In [93]:
```
plt.figure(figsize = (30, 70))
sns.heatmap(publisher_genre)

plt.show()
```


Publisher
Action
Adventure
Fighting
Misc
Platform
Puzzle
Racing
Role-Playing
Shooter
Simulation
Sports
Strategy
Genre
400
300
200
100
0

0.4 By Genre

```
In [76]: genre_publisher = pd.crosstab(data.Genre, data.Publisher)
         genre_publisher.head()
```

Out[76]:

Publisher	10TACLE Studios	1C Company	20th Century Fox Video Games	2D Boy	\
Genre					
Action	0	0	4	0	
Adventure	1	0	0	0	
Fighting	0	0	0	0	
Misc	0	0	0	0	
Platform	0	0	0	0	

Publisher	3DO	49Games	505 Games	5pb	7G//AMES	989 Sports	...	\
Genre							...	
Action	17	0	21	5	0	0	...	
Adventure	3	0	7	49	0	0	...	
Fighting	1	0	11	1	0	0	...	
Misc	0	0	31	0	1	0	...	
Platform	1	0	6	0	0	0	...	

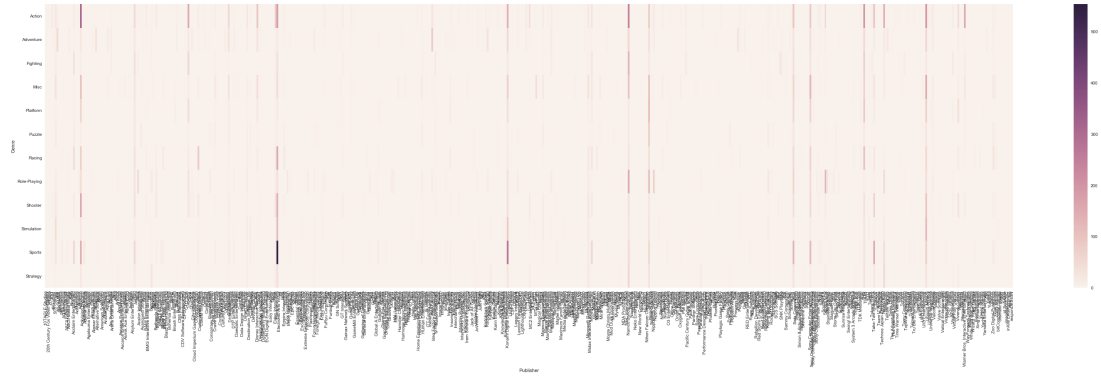
Publisher	Zushi Games	bitComposer Games	dramatic create	fonfun	iWin	\
Genre						
Action	1	0	1	0	0	
Adventure	1	0	4	0	0	
Fighting	0	0	0	0	0	
Misc	1	0	0	0	0	
Platform	0	0	0	0	0	

Publisher	id Software	imageepoch Inc.	inXile Entertainment	mixi, Inc	\
Genre					
Action	0	0	0	1	
Adventure	0	1	0	0	
Fighting	0	0	0	0	
Misc	0	0	0	0	
Platform	0	0	0	0	

Publisher	responDESIGN
Genre	
Action	0
Adventure	0
Fighting	0
Misc	0
Platform	0

```
[5 rows x 576 columns]
```

```
In [94]: plt.figure(figsize = (50, 12))
         sns.heatmap(genre_publisher)

         plt.show()
```

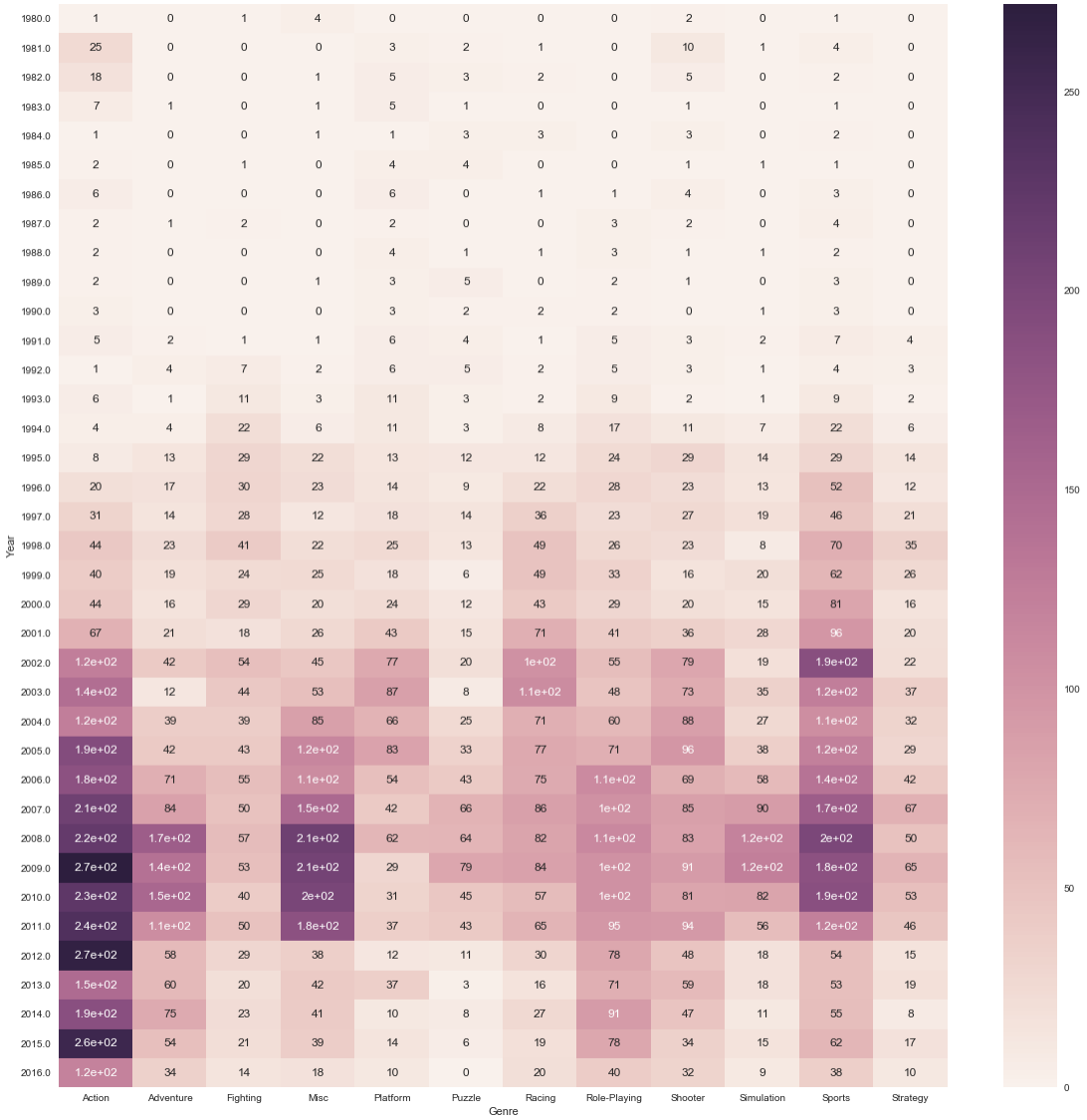


0.5 By Year

```
In [81]: year_genre = pd.crosstab(data.Year, data.Genre)
         year_genre.head()
```

```
Out[81]: Genre   Action  Adventure  Fighting  Misc  Platform  Puzzle  Racing  \
         Year
         1980.0       1          0         1     4         0       0       0
         1981.0      25          0         0     0         3       2       1
         1982.0      18          0         0     1         5       3       2
         1983.0       7          1         0     1         5       1       0
         1984.0       1          0         0     1         1       3       3

         Genre   Role-Playing  Shooter  Simulation  Sports  Strategy
         Year
         1980.0             0        2           0       1         0
         1981.0             0       10           1       4         0
         1982.0             0        5           0       2         0
         1983.0             0        1           0       1         0
         1984.0             0        3           0       2         0
```

```
In [90]: plt.figure(figsize = (20, 20))
         sns.heatmap(year_genre, annot = True)

         plt.show()
```

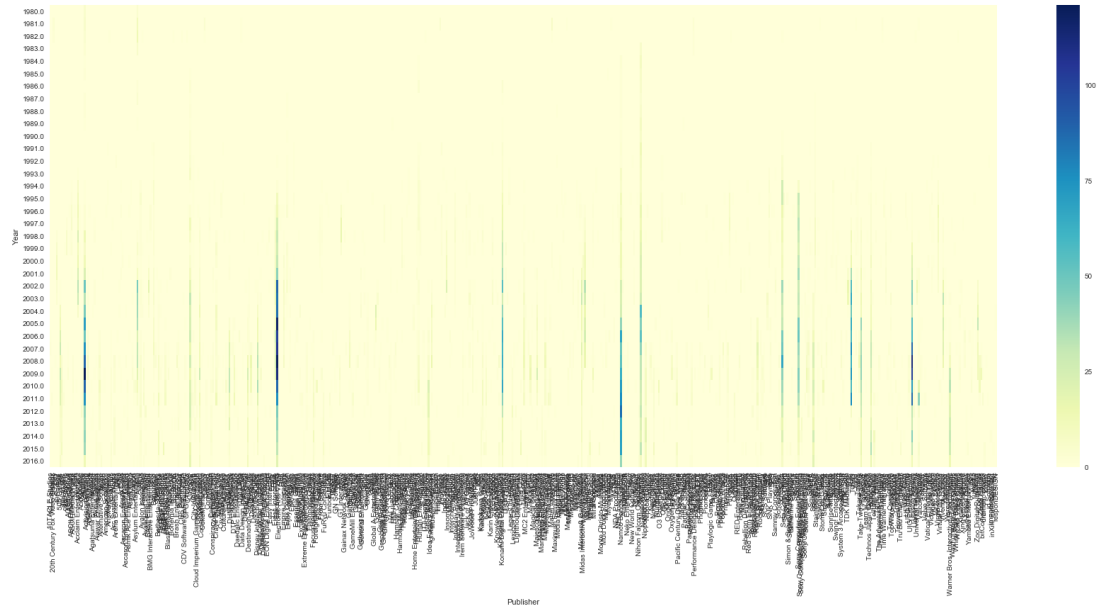



```
In [84]: year_publisher = pd.crosstab(data.Year, data.Publisher)
         year_publisher.head()
```

```
Out[84]: Publisher  10TACLE Studios  1C Company  20th Century Fox Video Games  2D Boy  \
         Year
         1980.0                   0           0                             0       0
         1981.0                   0           0                             3       0
         1982.0                   0           0                             2       0
         1983.0                   0           0                             0       0
         1984.0                   0           0                             0       0

         Publisher  3DO  49Games  505 Games  5pb  7G//AMES  989 Sports      ...       \
```

```
Year                                                                ...
1980.0       0        0          0    0        0           0      ...
1981.0       0        0          0    0        0           0      ...
1982.0       0        0          0    0        0           0      ...
1983.0       0        0          0    0        0           0      ...
1984.0       0        0          0    0        0           0      ...

Publisher  Zushi Games  bitComposer Games  dramatic create  fonfun  iWin  \
Year
1980.0               0                  0                0       0     0
1981.0               0                  0                0       0     0
1982.0               0                  0                0       0     0
1983.0               0                  0                0       0     0
1984.0               0                  0                0       0     0

Publisher  id Software  imageepoch Inc.  inXile Entertainment  mixi, Inc  \
Year
1980.0               0                0                     0          0
1981.0               0                0                     0          0
1982.0               0                0                     0          0
1983.0               0                0                     0          0
1984.0               0                0                     0          0

Publisher  responDESIGN
Year
1980.0                0
1981.0                0
1982.0                0
1983.0                0
1984.0                0

[5 rows x 576 columns]
```

In [89]:
```
plt.figure(figsize = (30, 12))
sns.heatmap(year_publisher, cmap="YlGnBu")

plt.show()
```


Year
Publisher
100
75
50
25
0