



# Video Game Analytics

- Pragathi S P

The given dataset has 16598 entries and 11 features.

- Rank
- Name
- Platform
- Year
- Genre
- Publisher
- NA\_Sales
- EU\_Sales
- JP\_Sales
- Other\_Sales
- Global\_Sales

It contains both categorical and numerical data

Numerical	Categorical
<ul style="list-style-type: none"><li>● Rank</li><li>● Year</li><li>● NA_Sales</li><li>● EU_Sales</li><li>● JP_Sales</li><li>● Other_Sales</li><li>● Global_Sales</li></ul>	<ul style="list-style-type: none"><li>● Name</li><li>● Platform</li><li>● Genre</li><li>● Publisher</li></ul>

- There are around 11325 unique names.
- 12 unique genre like : Sports, Platform', Racing, Role-Playing, Puzzle, Misc, Shooter, Simulation, Action, Fighting, Adventure and Strategy.
- 39 unique year starting from 1980 to year 2020
- 11325 unique publisher

## Null values:

The data consists of null values and since the null values are less than 5% of total data, we can drop the values.

## Insights

The details of top 5 companies with maximum global sales are:

Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
1	Wii Sports	Wii	2006.0	Sports	Nintendo	41.49	29.02	3.77	8.46	82.74
2	Super Mario Bros.	NES	1985.0	Platform	Nintendo	29.08	3.58	6.81	0.77	40.24
3	Mario Kart Wii	Wii	2008.0	Racing	Nintendo	15.85	12.88	3.79	3.31	35.82
4	Wii Sports Resort	Wii	2009.0	Sports	Nintendo	15.75	11.01	3.28	2.96	33.00
5	Pokemon Red/Pokemon Blue	GB	1996.0	Role-Playing	Nintendo	11.27	8.89	10.22	1.00	31.37
6	Tetris	GB	1989.0	Puzzle	Nintendo	23.20	2.26	4.22	0.58	30.26
7	New Super Mario Bros.	DS	2006.0	Platform	Nintendo	11.38	9.23	6.50	2.90	30.01
8	Wii Play	Wii	2006.0	Misc	Nintendo	14.03	9.20	2.93	2.85	29.02
9	New Super Mario Bros. Wii	Wii	2009.0	Platform	Nintendo	14.59	7.06	4.70	2.26	28.62
10	Duck Hunt	NES	1984.0	Shooter	Nintendo	26.93	0.63	0.28	0.47	28.31

The details of bottom 5 companies with minimum sales are:

Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
16591	Mega Brain Boost	DS	2008.0	Puzzle	Majesco Entertainment	0.01	0.00	0.00	0.0	0.01
16592	Chou Ezaru wa Akai Hana: Koi wa Tsuki ni Shiru...	PSV	2016.0	Action	dramatic create	0.00	0.00	0.01	0.0	0.01
16593	Eiyuu Densetsu: Sora no Kiseki Material Collec...	PSP	2007.0	Role-Playing	Falcom Corporation	0.00	0.00	0.01	0.0	0.01
16594	Myst IV: Revelation	PC	2004.0	Adventure	Ubisoft	0.01	0.00	0.00	0.0	0.01

16595	Plushees	DS	2008.0	Simulation	Destineer	0.01	0.00	0.00	0.0	0.01
16596	Woody Woodpecker in Crazy Castle 5	GBA	2002.0	Platform	Kemco	0.01	0.00	0.00	0.0	0.01
16597	Men in Black II: Alien Escape	GC	2003.0	Shooter	Infogrames	0.01	0.00	0.00	0.0	0.01
16598	SCORE International Baja 1000: The Official Game	PS2	2008.0	Racing	Activision	0.00	0.00	0.00	0.0	0.01
16599	Know How 2	DS	2010.0	Puzzle	7G//AMES	0.00	0.01	0.00	0.0	0.01
16600	Spirits & Spells	GBA	2003.0	Platform	Wanadoo	0.01	0.00	0.00	0.0	0.01

The sales statistics with respect to year

Top 10 year with maximum global sales			The top 10 year with maximum sales in the region of North America																																																		
<table><tr><th colspan="2">Global_Sales</th></tr><tr><th>Year</th><th></th></tr><tr><td>2008.0</td><td>678.90</td></tr><tr><td>2009.0</td><td>667.30</td></tr><tr><td>2007.0</td><td>609.92</td></tr><tr><td>2010.0</td><td>600.29</td></tr><tr><td>2006.0</td><td>521.04</td></tr><tr><td>2011.0</td><td>515.80</td></tr><tr><td>2005.0</td><td>458.51</td></tr><tr><td>2004.0</td><td>414.01</td></tr><tr><td>2002.0</td><td>395.52</td></tr><tr><td>2013.0</td><td>368.11</td></tr></table>			Global_Sales		Year		2008.0	678.90	2009.0	667.30	2007.0	609.92	2010.0	600.29	2006.0	521.04	2011.0	515.80	2005.0	458.51	2004.0	414.01	2002.0	395.52	2013.0	368.11	<table><tr><th colspan="2">NA_Sales</th></tr><tr><th>Year</th><th></th></tr><tr><td>2008.0</td><td>351.44</td></tr><tr><td>2009.0</td><td>338.85</td></tr><tr><td>2007.0</td><td>311.18</td></tr><tr><td>2010.0</td><td>304.24</td></tr><tr><td>2006.0</td><td>263.12</td></tr><tr><td>2005.0</td><td>241.58</td></tr><tr><td>2011.0</td><td>241.00</td></tr><tr><td>2004.0</td><td>218.80</td></tr><tr><td>2002.0</td><td>216.19</td></tr><tr><td>2003.0</td><td>193.59</td></tr></table>			NA_Sales		Year		2008.0	351.44	2009.0	338.85	2007.0	311.18	2010.0	304.24	2006.0	263.12	2005.0	241.58	2011.0	241.00	2004.0	218.80	2002.0	216.19	2003.0	193.59
Global_Sales																																																					
Year																																																					
2008.0	678.90																																																				
2009.0	667.30																																																				
2007.0	609.92																																																				
2010.0	600.29																																																				
2006.0	521.04																																																				
2011.0	515.80																																																				
2005.0	458.51																																																				
2004.0	414.01																																																				
2002.0	395.52																																																				
2013.0	368.11																																																				
NA_Sales																																																					
Year																																																					
2008.0	351.44																																																				
2009.0	338.85																																																				
2007.0	311.18																																																				
2010.0	304.24																																																				
2006.0	263.12																																																				
2005.0	241.58																																																				
2011.0	241.00																																																				
2004.0	218.80																																																				
2002.0	216.19																																																				
2003.0	193.59																																																				
The top 10 year with maximum sales in the region of Europe			The top 10 year with maximum sales in the region of Japan																																																		
<table><tr><th colspan="2">EU_Sales</th></tr><tr><th>Year</th><th></th></tr><tr><td>2009.0</td><td>191.59</td></tr><tr><td>2008.0</td><td>184.40</td></tr><tr><td>2010.0</td><td>176.57</td></tr><tr><td>2011.0</td><td>167.31</td></tr><tr><td>2007.0</td><td>160.18</td></tr><tr><td>2006.0</td><td>129.24</td></tr><tr><td>2013.0</td><td>125.80</td></tr><tr><td>2014.0</td><td>125.63</td></tr></table>			EU_Sales		Year		2009.0	191.59	2008.0	184.40	2010.0	176.57	2011.0	167.31	2007.0	160.18	2006.0	129.24	2013.0	125.80	2014.0	125.63	<table><tr><th colspan="2">JP_Sales</th></tr><tr><th>Year</th><th></th></tr><tr><td>2009.0</td><td>61.89</td></tr><tr><td>2008.0</td><td>60.26</td></tr><tr><td>2010.0</td><td>59.49</td></tr><tr><td>2011.0</td><td>53.04</td></tr><tr><td>2007.0</td><td>60.29</td></tr><tr><td>2006.0</td><td>73.73</td></tr><tr><td>2013.0</td><td>47.59</td></tr><tr><td>2014.0</td><td>39.46</td></tr></table>			JP_Sales		Year		2009.0	61.89	2008.0	60.26	2010.0	59.49	2011.0	53.04	2007.0	60.29	2006.0	73.73	2013.0	47.59	2014.0	39.46								
EU_Sales																																																					
Year																																																					
2009.0	191.59																																																				
2008.0	184.40																																																				
2010.0	176.57																																																				
2011.0	167.31																																																				
2007.0	160.18																																																				
2006.0	129.24																																																				
2013.0	125.80																																																				
2014.0	125.63																																																				
JP_Sales																																																					
Year																																																					
2009.0	61.89																																																				
2008.0	60.26																																																				
2010.0	59.49																																																				
2011.0	53.04																																																				
2007.0	60.29																																																				
2006.0	73.73																																																				
2013.0	47.59																																																				
2014.0	39.46																																																				

2005.0	121.56	2005.0	54.28
2012.0	118.76	2012.0	51.74

*The top publishers with the leading global sales are*

Global_Sales	
Publisher	
Nintendo	1784.43
Electronic Arts	1093.39
Activision	721.41
Sony Computer Entertainment	607.28
Ubisoft	473.54
Take-Two Interactive	399.30
THQ	340.44
Konami Digital Entertainment	278.56
Sega	270.70
Namco Bandai Games	253.65

*Top 10 publishers with highest global sales are*

Global_Sales	
Platform	
PS2	1233.46
X360	969.60
PS3	949.35
Wii	909.81
DS	818.91
PS	727.39
GBA	305.62
PSP	291.71
3DS	270.10

PS4	278.10
PC	254.70

The top 5 platforms contributing to maximum sales in region of north America , Japan and Europe are

NA_Sales	
Platform	
X360	594.33
PS2	572.92
Wii	497.37
PS3	388.90
DS	388.53

EU_Sales	
Platform	
PS3	340.47
PS2	332.63
X360	278.00
Wii	264.35
PS	212.38

EU_Sales	
Platform	
PS3	340.47
PS2	332.63
X360	278.00
Wii	264.35
PS	212.38

*The ranking of Genre based on the global sales is:*

Global_Sales	
Genre	

Action	1722.84
Sports	1309.24
Shooter	1026.20
Role-Playing	923.83
Platform	829.13
Misc	789.87
Racing	726.76
Fighting	444.05
Simulation	389.98
Puzzle	242.21
Adventure	234.59
Strategy	173.27

The top 5 genre contributing to maximum sales in region of North America , Japan and Europe are

NA_Sales	
Genre	
Action	861.77
Sports	670.09
Shooter	575.16
Platform	445.99
Misc	396.92

EU_Sales	
Genre	
Action	516.48
Sports	371.34
Shooter	310.45
Racing	236.31
Misc	211.77

**JP\_Sales****Genre**

<b>Role-Playing</b>	350.29
---------------------	--------

<b>Action</b>	158.65
---------------	--------

<b>Sports</b>	134.76
---------------	--------

<b>Platform</b>	130.65
-----------------	--------

<b>Misc</b>	106.67
-------------	--------

Code

```
In [16]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
In [2]: data=pd.read_csv(r'C:\Users\Lenovo\Downloads\vg-sales.csv')
```

```
In [138]: data.head(10)
```

Out[138]:

	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
5	6	Tetris	GB	1989.0	Puzzle	Nintendo	23.20	2.26	4.22	0.58	30.26
6	7	New Super Mario Bros.	DS	2006.0	Platform	Nintendo	11.38	9.23	6.50	2.90	30.01
7	8	Wii Play	Wii	2006.0	Misc	Nintendo	14.03	9.20	2.93	2.85	29.02
8	9	New Super Mario Bros. Wii	Wii	2009.0	Platform	Nintendo	14.59	7.06	4.70	2.26	28.62
9	10	Duck Hunt	NES	1984.0	Shooter	Nintendo	26.93	0.63	0.28	0.47	28.31

```
In [139]: data.tail(10)
```

Out[139]:

	Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
16588	16591	Mega Brain Boost	DS	2008.0	Puzzle	Majesco Entertainment	0.01	0.00	0.00	0.0	0.01
16589	16592	Chou Ezaru wa Akai Hana: Koi wa Tsuki ni Shiru...	PSV	2016.0	Action	dramatic create	0.00	0.00	0.01	0.0	0.01
16590	16593	Eiyuu Densetsu: Sora no Kiseki Material Collec...	PSP	2007.0	Role-Playing	Falcom Corporation	0.00	0.00	0.01	0.0	0.01
16591	16594	Myst IV: Revelation	PC	2004.0	Adventure	Ubisoft	0.01	0.00	0.00	0.0	0.01
16592	16595	Plushees	DS	2008.0	Simulation	Destineer	0.01	0.00	0.00	0.0	0.01
16593	16596	Woody Woodpecker in Crazy Castle 5	GBA	2002.0	Platform	Kemco	0.01	0.00	0.00	0.0	0.01
16594	16597	Men in Black II: Alien Escape	GC	2003.0	Shooter	Infogrames	0.01	0.00	0.00	0.0	0.01
16595	16598	SCORE International Baja 1000: The Official Game	PS2	2008.0	Racing	Activision	0.00	0.00	0.00	0.0	0.01
16596	16599	Know How 2	DS	2010.0	Puzzle	7G//AMES	0.00	0.01	0.00	0.0	0.01
16597	16600	Spirits & Spells	GBA	2003.0	Platform	Wanadoo	0.01	0.00	0.00	0.0	0.01

```
In [7]: data.shape
```

```
Out[7]: (16598, 11)
```

```
In [10]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16598 entries, 0 to 16597
Data columns (total 11 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Rank            16598 non-null  int64
1   Name            16598 non-null  object
2   Platform        16598 non-null  object
3   Year            16327 non-null  float64
4   Genre           16598 non-null  object
5   Publisher       16540 non-null  object
6   NA_Sales        16598 non-null  float64
7   EU_Sales        16598 non-null  float64
8   JP_Sales        16598 non-null  float64
9   Other_Sales     16598 non-null  float64
10  Global_Sales    16598 non-null  float64
dtypes: float64(6), int64(1), object(4)
memory usage: 1.4+ MB
```

Dealing with Null Values



```
In [20]: data.isna().sum()
```

```
Out[20]: Rank          0
Name          0
Platform      0
Year         271
Genre         0
Publisher     58
NA_Sales      0
EU_Sales      0
JP_Sales      0
Other_Sales   0
Global_Sales  0
dtype: int64
```

Since the null values are less than 5% of total data, we can drop the values

```
In [22]: data=data.dropna()
```

```
In [23]: data.isna().sum()
```

```
Out[23]: Rank          0
Name          0
Platform      0
Year          0
Genre         0
Publisher     0
NA_Sales      0
EU_Sales      0
JP_Sales      0
Other_Sales   0
Global_Sales  0
dtype: int64
```

```
In [24]: data.describe()
```

```
Out[24]:
```

	Rank	Year	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
count	16291.000000	16291.000000	16291.000000	16291.000000	16291.000000	16291.000000	16291.000000
mean	8290.190228	2006.405561	0.265647	0.147731	0.078833	0.048426	0.540910
std	4792.654450	5.832412	0.822432	0.509303	0.311879	0.190083	1.567345
min	1.000000	1980.000000	0.000000	0.000000	0.000000	0.000000	0.010000
25%	4132.500000	2003.000000	0.000000	0.000000	0.000000	0.000000	0.060000
50%	8292.000000	2007.000000	0.080000	0.020000	0.000000	0.010000	0.170000
75%	12439.500000	2010.000000	0.240000	0.110000	0.040000	0.040000	0.480000
max	16600.000000	2020.000000	41.490000	29.020000	10.220000	10.570000	82.740000

```
In [ ]: sns.set(font_scale=1.2)
sns.catplot(x='Publisher', y='Global_Sales', palette='rainbow', kind='box', data=data)
plt.gcf().set_size_inches(15,10)
```

```
In [26]: data.columns
```

```
Out[26]: Index(['Rank', 'Name', 'Platform', 'Year', 'Genre', 'Publisher', 'NA_Sales',
              'EU_Sales', 'JP_Sales', 'Other_Sales', 'Global_Sales'],
              dtype='object')
```

```
In [87]: cat_col=data.select_dtypes(exclude='number').columns
cat_col
```

```
Out[87]: Index(['Name', 'Platform', 'Genre', 'Publisher'], dtype='object')
```

```
In [95]: unique_Platform=[]
for i in data['Platform'].unique():
    unique_Platform.append(i)
print(unique_Platform)
len(unique_Platform)
```

```
['Wii', 'NES', 'GB', 'DS', 'X360', 'PS3', 'PS2', 'SNES', 'GBA', '3DS', 'PS4', 'N64', 'PS', 'XB', 'PC', '2600', 'PSP', 'XOne',
'GC', 'WiiU', 'GEN', 'DC', 'PSV', 'SAT', 'SCD', 'WS', 'NG', 'TG16', '3DO', 'GG', 'PCFX']
```

```
Out[95]: 31
```

```
In [96]: unique_Genre=[]
for i in data['Genre'].unique():
    unique_Genre.append(i)
print(unique_Genre)
len(unique_Genre)
```

['Sports', 'Platform', 'Racing', 'Role-Playing', 'Puzzle', 'Misc', 'Shooter', 'Simulation', 'Action', 'Fighting', 'Adventure', 'Strategy']

Out[96]: 12

```
In [101]: unique_year=[]
for i in data['Year'].unique():
    unique_year.append(i)
print(unique_year)
print(len(unique_year),min(unique_year), max(unique_year))
```

[2006.0, 1985.0, 2008.0, 2009.0, 1996.0, 1989.0, 1984.0, 2005.0, 1999.0, 2007.0, 2010.0, 2013.0, 2004.0, 1990.0, 1988.0, 2002.0, 2001.0, 2011.0, 1998.0, 2015.0, 2012.0, 2014.0, 1992.0, 1997.0, 1993.0, 1994.0, 1982.0, 2003.0, 1986.0, 2000.0, 1995.0, 2016.0, 1991.0, 1981.0, 1987.0, 1980.0, 1983.0, 2020.0, 2017.0]  
39 1980.0 2020.0

```
In [103]: unique_publisher=[]
for i in data['Publisher'].unique():
    unique_publisher.append(i)
print(unique_publisher)
len(unique_publisher)
```

ft', 'vap', 'Misawa', 'Tradewest', 'Team1/ Software', 'yeti', 'Pow', 'Navarre Corp', 'MediaQuest', 'Max Five', 'Comfort', 'Monte Christo Multimedia', 'Pony Canyon', 'Riverhillsoft', 'Summitsoft', 'Milestone S.r.l', 'Playmore', 'MLB.com', 'Kool Kizz', 'Flashpoint Games', '49Games', 'Legacy Interactive', 'Alawar Entertainment', 'CyberFront', 'Cloud Imperium Games Corporation', 'Societa', 'Virtual Play Games', 'Interchannel', 'Sonnet', 'Experience Inc.', 'Zenrin', 'Iceberg Interactive', 'Ivolgamus', '2D Boy', 'MC2 Entertainment', 'Kando Games', 'Just Flight', 'Office Create', 'Mamba Games', 'Fields', 'Princess Soft', 'Maximum Family Games', 'Berkeley', 'Fuji', 'Dusenberry Martin Racing', 'imageepoch Inc.', 'Big Fish Games', 'Her Interactive', 'Kamui', 'ASK', 'TopWare Interactive', 'Headup Games', 'KSS', 'Cygames', 'KID', 'Quinrose', 'Sunflowers', 'dramatic create', 'TGL', 'Encore', 'Extreme Entertainment Group', 'Intergrow', 'G.Rev', 'Sweets', 'Kokopeli Digital Studios', 'Number None', 'Nexon', 'id Software', 'BushiRoad', 'Tryfirst', 'Strategy First', '7G//AMES', 'GN Software', 'Yuke's', 'Easy Interactive', 'Licensed 4U', 'FuRyu Corporation', 'Lexicon Entertainment', 'Paon Corporation', 'Kids Station', 'GOA', 'Graphsim Entertainment', 'King Records', 'Introversion Software', 'Minato Station', 'Devolver Digital', 'Blue Byte', 'Gaga', 'Yamasa Entertainment', 'Plenty', 'Views', 'fonfun', 'NetRevo', 'Codemasters Online', 'Quintet', 'Phoenix Games', 'Dorart', 'Marvelous Games', 'Focus Multimedia', 'Imageworks', 'Karin Entertainment', 'Aerosoft', 'Technos Japan Corporation', 'Gakken', 'Mirai Shounen', 'Datam Polystar', 'Saurus', 'Hunex', 'Revolution (Japan)', 'Giza10', 'Visco', 'Alvion', 'Mycom', 'Giga', 'Warashi', 'System Soft', 'Sold Out', 'Lighthouse Interactive', 'Masque Publishing', 'RED Entertainment', 'Michaelsoft', 'Media Entertainment', 'New World Computing', 'Genterprise', 'Interworks Unlimited, Inc.', 'Boost On', 'Stainless Games', 'EON Digital Entertainment', 'Epic Games', 'Naxat Soft', 'Ascaron Entertainment', 'Piacchi', 'Nitroplus', 'Paradox Development', 'Otomate', 'Ongakukan', 'Commseed', 'Inti Creates', 'Takuyo', 'Interchannel-Holon', 'Rain Games', 'UIG Entertainment']

Out[103]: 576

```
In [98]: unique_name=[]
for i in data['Name'].unique():
    unique_name.append(i)
print(unique_name)
len(unique_name)
```

py Dance Collection', 'Space Hulk', 'Fuun Shinsengumi Bakumatsuden Portable', 'Musou Tourou', 'Really? Really! DS', 'Gakuen Hetalia', 'Daiya no Kuni no Alice: Wonderful Mirror World', 'Touch Detective 3: Does Funghi Dream of Bananas?', 'Period Cube: Torikago no Amadeus', 'Tringo', 'Jam Sessions: Sing and Play Guitar (jp sales)', 'Binary Star', 'RalliSport Challenge(JP sales)', 'Dragon Quest X: All in One Package', 'Wedding Planner: Dream Weddings Guaranteed', 'Element Hunter', 'Ultraman Fighting Evolution 3', 'Rondo of Swords (jp sales)', 'The Eye of Judgment: Legends', 'Nakayoshi All-Stars: Mezase Gakuen Idol', 'Shinsieiki Evangelion: Koutetsu no Girlfriend Tokubetsu-Hen Portable', 'Fullmetal Alchemist: Prince of the Dawn', 'Command & Conquer Renegade', 'H2O+: Footprints in the Sand', 'Fashionable Puppy: Oshare na Koinu DS', 'Sekirei: Mirai Kara no Okurimono', 'Cities: Skylines Snowfall', 'Dinotopia: The Sunstone Odyssey', 'Idol Janshi Suchie-Pai IV', 'Dungeons & Dragons: Neverwinter Nights - Complete', 'Natsuzora no Monologue', 'Kenka Bancho Portable', 'World Snooker Challenge 2005', 'G1 Jockey 4 2008', 'Icewind Dale II', 'The Amazing Spider-Man (DS & Mobile Versions)', 'Ishin Renka: Ryouma Gaiden', 'Shirahana no Ori: Hiroy no Kakera 4 - Shiki no Uta', 'Act of Aggression', 'Fallout Anthology', 'Tropico 3: Gold Edition', 'Survivor', 'Wand of Fortune', 'Gobliins 4', 'S.Y.K: Shinsetsu Saiyuuiki Portable', 'Meru Purana', 'Cross Edge Dash', 'Fullmetal Alchemist: Trading Card Game', 'Luxor: Pharaoh's Challenge', 'Grand Prix Legends', 'Fate/kaleid liner Prisma Illya', 'Hoshigami: Ruining Blue Earth Remix', 'Akatsuki no Amaneka to Aoi Kyojin', 'Pilot Academy', 'Poker for Dummies', 'Darkened Skye', 'Wand of Fortune: Mirai e no Prologue', 'Nova Usagi no Game de Ryuugaku!? DS', 'LMA Manager 2007', 'San-X: Chara Sagashi Land', 'Heart no Kuni no Alice', 'Memories Off', 'Tenchu San Portable', 'Vitamin X', 'The Wild West', 'Saint Seiya Omega: Ultimate Cosmo', 'Element Girl: Love, Fashion and Friends', 'Sakigake!! Otokojuku - Nihon yo, Kore ga Otoko Dearu!', 'Teenage Mutant Ninja Turtles (3DS)', 'Time Leap', 'thinkSMART: Chess for Kids', 'World Heroes Anthology', 'think Logic Trainer', 'Tennis no Oji-Sama: Doubles no Oji-Sama - Boys, Be Glorious!', 'Routes PE', 'Uchida Yasuo Mystery: Meitantei Senken Mitsuhiro Series: Fukutoshin Renzoku Satsujin Jiken', 'Railfan', 'SoniPro', 'Real Crimes: The Unicorn Killer', 'Akatsuki no Goei Trinity', 'Aladdin Magic Racer', 'Goat Sim

```
In [33]: num_col=data.select_dtypes(include='number').columns
num_col
```

```
Out[33]: Index(['Rank', 'Year', 'NA_Sales', 'EU_Sales', 'JP_Sales', 'Other_Sales',
              'Global_Sales'],
              dtype='object')
```

The top 10 publishers with highest gobal sales

```
In [42]: Gsales_Publisr=data.groupby(["Publisher"]).sum().sort_values("Global_Sales", ascending=False)
```

```
In [46]: Gsales_Publisr
```

Out[46]:

	Rank	Year	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
Publisher							
Nintendo	2646098	1394666.0	815.75	418.30	454.99	95.19	1784.43
Electronic Arts	7102442	2686562.0	584.22	367.38	13.98	127.63	1093.39
Activision	6324706	1938979.0	426.01	213.72	6.54	74.79	721.41
Sony Computer Entertainment	4098421	1367034.0	265.22	187.55	74.10	80.40	607.28
Ubisoft	6945872	1843194.0	252.81	163.03	7.33	50.16	473.54
...	...	...	...	...	...	...	...
Epic Games	16263	2016.0	0.01	0.00	0.00	0.00	0.01
Paradox Development	16347	2016.0	0.00	0.01	0.00	0.00	0.01
Rain Games	16568	2015.0	0.00	0.01	0.00	0.00	0.01
Ascaron Entertainment	16292	2012.0	0.00	0.01	0.00	0.00	0.01
Commseed	16460	2008.0	0.00	0.00	0.01	0.00	0.01

576 rows × 7 columns

```
In [48]: Gsales_Publisr[["Global_Sales']].head(10)
```

Out[48]:

	Global_Sales
Publisher	
Nintendo	1784.43
Electronic Arts	1093.39
Activision	721.41
Sony Computer Entertainment	607.28
Ubisoft	473.54
Take-Two Interactive	399.30
THQ	340.44
Konami Digital Entertainment	278.56
Sega	270.70
Namco Bandai Games	253.65

```
In [ ]: data.groupby(["Publisher"]).sum().sort_values("Global_Sales", ascending=False)
```

Which Genre sold the most

```
In [55]: Genre_gsales=data.groupby(['Genre']).sum().sort_values('Global_Sales', ascending=False)
Genre_gsales[['Global_Sales']]
```

Out[55]:

Global_Sales	
Genre	
Action	1722.84
Sports	1309.24
Shooter	1026.20
Role-Playing	923.83
Platform	829.13
Misc	789.87
Racing	726.76
Fighting	444.05
Simulation	389.98
Puzzle	242.21
Adventure	234.59
Strategy	173.27

```
In [60]: g1=data.groupby(["Publisher","Genre"]).sum()
total_gsales=g1[['Global_Sales']].groupby(level=0,group_keys=False)
total_gsales.nlargest(5)
```

Out[60]:

Publisher	Genre	
10TACLE Studios	Strategy	0.06
	Puzzle	0.03
	Adventure	0.02
	Strategy	0.05
1C Company	Racing	0.04
	...	
imageepoch Inc.	Role-Playing	0.03
	Adventure	0.01
inXile Entertainment	Role-Playing	0.10
mixi, Inc	Action	0.86
responDESIGN	Sports	0.13
Name: Global_Sales, Length: 1424, dtype: float64		

Type casting numerical year type into categorical

```
In [111]: datap=data['Year'].astype('category', copy=False)
```

```
In [116]: datap=pd.DataFrame(datap)
datap
```

Out[116]:

Year	
0	2006.0
1	1985.0
2	2008.0
3	2009.0
4	1996.0
...	...
16593	2002.0
16594	2003.0
16595	2008.0
16596	2010.0
16597	2003.0

16291 rows × 1 columns

```
In [117]: data['Year']=datap['Year']

C:\Users\Lenovo\AppData\Local\Temp\ipykernel_20248\977412872.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy
data['Year']=datap['Year']
```

```
In [118]: data.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 16291 entries, 0 to 16597
Data columns (total 11 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   Rank            16291 non-null  int64
1   Name            16291 non-null  object
2   Platform        16291 non-null  object
3   Year            16291 non-null  category
4   Genre           16291 non-null  object
5   Publisher       16291 non-null  object
6   NA_Sales        16291 non-null  float64
7   EU_Sales        16291 non-null  float64
8   JP_Sales        16291 non-null  float64
9   Other_Sales     16291 non-null  float64
10  Global_Sales    16291 non-null  float64
dtypes: category(1), float64(5), int64(1), object(4)
memory usage: 1.4+ MB
```

```
In [130]: Year_Gsales=data.groupby('Year').sum().sort_values('Global_Sales',ascending=False)
Year_Gsales[['Global_Sales']].head(10)
```

Out[130]:

Global_Sales	
Year	
2008.0	678.90
2009.0	667.30
2007.0	609.92
2010.0	600.29
2006.0	521.04
2011.0	515.80
2005.0	458.51
2004.0	414.01
2002.0	395.52
2013.0	368.11

```
In [134]: Year_NAsales=data.groupby('Year').sum().sort_values('NA_Sales',ascending=False)
Year_NAsales[['NA_Sales']].head(10)
```

Out[134]:

NA_Sales	
Year	
2008.0	351.44
2009.0	338.85
2007.0	311.18
2010.0	304.24
2006.0	263.12
2005.0	241.58
2011.0	241.00
2004.0	218.80
2002.0	216.19
2003.0	193.59

```
In [135]: Year_EUsales=data.groupby('Year').sum().sort_values('EU_Sales',ascending=False)
Year_EUsales[['EU_Sales']].head(10)
```

Out[135]:

EU_Sales	
Year	
2009.0	191.59
2008.0	184.40
2010.0	176.57
2011.0	167.31
2007.0	160.18
2006.0	129.24
2013.0	125.80
2014.0	125.63
2005.0	121.56
2012.0	118.76

```
In [136]: Year_JPsales=data.groupby('Year').sum().sort_values('JP_Sales',ascending=False)
Year_EUsales[['JP_Sales']].head(10)
```

Out[136]:

JP_Sales	
Year	
2009.0	61.89
2008.0	60.26
2010.0	59.49
2011.0	53.04
2007.0	60.29
2006.0	73.73
2013.0	47.59
2014.0	39.46
2005.0	54.28
2012.0	51.74

*Platform analysis*

```
In [143]: Platform_sales=data.groupby('Platform').sum().sort_values('Global_Sales',ascending=False)
Platform_sales[['Global_Sales']].head(10)
```

Out[143]:

Global_Sales	
Platform	
PS2	1233.46
X360	969.60
PS3	949.35
Wii	909.81
DS	818.91
PS	727.39
GBA	305.62
PSP	291.71
PS4	278.10
PC	254.70

```
In [145]: Platform_sales_NA=data.groupby('Platform').sum().sort_values('NA_Sales',ascending=False)
Platform_sales_NA[['NA_Sales']].head(5)
```

Out[145]:

NA_Sales	
Platform	
X360	594.33
PS2	572.92
Wii	497.37
PS3	388.90
DS	388.53

```
In [156]: Platform_sales_EU=data.groupby('Platform').sum().sort_values('EU_Sales',ascending=False)
Platform_sales_EU[['EU_Sales']].head(5)
```

Out[156]:

EU_Sales	
Platform	
PS3	340.47
PS2	332.63
X360	278.00
Wii	264.35
PS	212.38

```
In [157]: Platform_sales_JP=data.groupby('Platform').sum().sort_values('JP_Sales',ascending=False)
Platform_sales_JP[['JP_Sales']].head(5)
```

Out[157]:

JP_Sales	
Platform	
DS	175.02
PS	139.78
PS2	137.54
SNES	116.55
NES	98.65

## Genre

```
In [162]: Genre_sales=data.groupby('Genre').sum().sort_values('Global_Sales',ascending=False)
Genre_sales[['Global_Sales']]
```

Out[162]:

Global_Sales	
Genre	
Action	1722.84
Sports	1309.24
Shooter	1026.20
Role-Playing	923.83
Platform	829.13
Misc	789.87
Racing	726.76
Fighting	444.05
Simulation	389.98
Puzzle	242.21
Adventure	234.59
Strategy	173.27

```
In [165]: Genre_sales_NA=data.groupby('Genre').sum().sort_values('NA_Sales',ascending=False)
Genre_sales_NA[['NA_Sales']].head(5)
```

Out[165]:

NA_Sales	
Genre	
Action	861.77
Sports	670.09
Shooter	575.16
Platform	445.99
Misc	396.92

```
In [167]: Genre_sales_EU=data.groupby('Genre').sum().sort_values('EU_Sales',ascending=False)
Genre_sales_EU[['EU_Sales']].head(5)
```

Out[167]:

EU_Sales	
Genre	
Action	516.48
Sports	371.34
Shooter	310.45
Racing	236.31
Misc	211.77

```
In [168]: Genre_sales_EU=data.groupby('Genre').sum().sort_values('JP_Sales',ascending=False)
Genre_sales_EU[['JP_Sales']].head(5)
```

Out[168]:

JP_Sales	
Genre	
Role-Playing	350.29
Action	158.65
Sports	134.76
Platform	130.65
Misc	106.67