

## My Notebook

```
In [ ]: # Importing all the relevant libraries for my analysis
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
```

### Importing my Dataset

```
In [ ]: VideoGames=pd.read_csv("VideoGameSales.csv")
print(VideoGames)
```

```
Rank          Name Platform Year   Genre \
0      1    Wii Sports     Wii 2006  Sports
1      2 Super Mario Bros.   NES 1985 Platform
2      3    Mario Kart Wii     Wii 2008 Racing
3      4    Wii Sports Resort     Wii 2009 Sports
4      5  Pokemon Red/Pokemon Blue     GB 1996 Role-Playing
..      ...
194     220        FIFA 15     PS3 2014  Sports
195     221  Crash Team Racing     PS 1999 Racing
196     222        FIFA 17     PS4 2016  Sports
197     223    Batman: Arkham City    X360 2011 Action
198     224       Driver 2     PS 2000 Action

Publisher  NorthAmerica_Sales \
0           Nintendo      41.49
1           Nintendo      29.08
2           Nintendo       NaN
3           Nintendo      15.75
4           Nintendo      11.27
..           ...
194        Electronic Arts     0.57
195  Sony Computer Entertainment     2.57
196        Electronic Arts     0.28
197 Warner Bros. Interactive Entertainment     2.99
198                 NaN      2.36

EuropeanUnion_Sales  Japan_Sales  Other_Sales  Global_Sales
0            29.02      3.77      8.46      82.74
1            3.58      6.81      0.77      40.24
2           12.88      3.79      NaN      35.82
3           11.01      NaN      2.96      33.00
4            8.89     10.22      1.00      31.37
..           ...
194            3.14      0.04      1.07      4.82
195            1.57      0.44      0.21      4.79
196            3.75      0.06      0.69      4.77
197            1.31      0.04      0.41      4.76
198            2.10      0.02      0.25      NaN
```

[199 rows x 11 columns]

Show the number of rows and columns in my dataset

```
In [ ]: rows=VideoGames.shape[0]
columns=VideoGames.shape[1]
print(rows)
print(columns)
```

```
199
11
```

Dataset column names

```
In [ ]: colnames=list(VideoGames)
print(colnames)

['Rank', 'Name', 'Platform', 'Year', 'Genre', 'Publisher', 'NorthAmerica_Sales',
'EuropeanUnion_Sales', 'Japan_Sales', 'Other_Sales', 'Global_Sales']
```

Show the data types in my dataset

```
In [ ]: print(VideoGames.info())

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 199 entries, 0 to 198
Data columns (total 11 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Rank             199 non-null    int64  
 1   Name             199 non-null    object  
 2   Platform         199 non-null    object  
 3   Year             199 non-null    int64  
 4   Genre            199 non-null    object  
 5   Publisher        198 non-null    object  
 6   NorthAmerica_Sales 185 non-null    float64
 7   EuropeanUnion_Sales 190 non-null    float64
 8   Japan_Sales      189 non-null    float64
 9   Other_Sales       192 non-null    float64
 10  Global_Sales     187 non-null    float64
dtypes: float64(5), int64(2), object(4)
memory usage: 17.2+ KB
None
```

Column description

Rank => Rank of the game

Name => Name of the game

Platform => Platform on which the game can run

Year => Release year

Genre => Genre of game

Publisher => Publishing company

NorthAmerica\_Sales => Sales in North America

EuropeanUnion\_Sales => Sales in European Union

Japan\_Sales => Sales in Japan

Other\_Sales => Sales in other countries

Global\_Sales => Total Global sales

Data types (Quantitative/Categorical)

Year => Quantitative

Genre => Categorical

Publisher => Categorical

NorthAmerica\_Sales => Quantitative

EuropeanUnion\_Sales => Quantitative

Japan\_Sales => Quantitative

Other\_Sales => Quantitative

Global\_Sales => Quantitative

Frequency Distribution, Relative Frequency, and Percentage Frequency

```
In [ ]: #Frequency Distribution

f1= VideoGames['Rank'].value_counts(bins=4)

f2=VideoGames['Year'].value_counts(bins=4)

f3=VideoGames['NorthAmerica_Sales'].value_counts(bins=4)

f4=VideoGames['EurpeanUnion_Sales'].value_counts(bins=4)

f5=VideoGames['Japan_Sales'].value_counts(bins=4)

f6=VideoGames['Other_Sales'].value_counts(bins=4)

f7=VideoGames['Global_Sales'].value_counts(bins=4)

f8=VideoGames['Name'].value_counts()

f9=VideoGames['Platform'].value_counts()

f10=VideoGames['Genre'].value_counts()

f11=VideoGames['Publisher'].value_counts()

print(f1)
print(f2)
print(f3)
print(f4)
print(f5)
print(f6)
print(f7)
print(f8)
print(f9)
print(f10)
print(f11)
```

```
(56.75, 112.5]      53
(0.776, 56.75]      51
(112.5, 168.25]     49
(168.25, 224.0]     46
Name: Rank, dtype: int64
(2007.5, 2016.0]     99
(1999.0, 2007.5]     48
(1990.5, 1999.0]     40
(1981.965, 1990.5]   12
Name: Year, dtype: int64
(-0.0425, 10.372]    174
(10.372, 20.745]     7
(20.745, 31.118]     3
(31.118, 41.49]      1
Name: NorthAmerica_Sales, dtype: int64
(-0.0300000000000002, 7.255]    178
(7.255, 14.51]        11
(21.765, 29.02]       1
(14.51, 21.765]       0
Name: EuropeanUnion_Sales, dtype: int64
(-0.0112000000000002, 2.555]    152
(2.555, 5.11]          28
(5.11, 7.665]          8
(7.665, 10.22]         1
Name: Japan_Sales, dtype: int64
(-0.00946, 2.115]      181
(2.115, 4.23]          10
(6.345, 8.46]          1
(4.23, 6.345]          0
Name: Other_Sales, dtype: int64
(4.681, 24.255]        177
(24.255, 43.75]        9
(63.245, 82.74]        1
(43.75, 63.245]        0
Name: Global_Sales, dtype: int64
Grand Theft Auto V      4
Call of Duty: World at War 2
Dr. Mario                2
Call of Duty: Black Ops 3 2
Call of Duty: Black Ops II 2
..
Wii Party                1
Halo 2                   1
Mario Party 8             1
Pokemon Black 2/Pokemon White 2 1
Driver 2                  1
Name: Name, Length: 167, dtype: int64
PS3          31
X360         30
Wii          21
DS           20
PS           20
GB           11
PS4          9
3DS          8
N64          8
GBA          7
NES          7
SNES         6
```

PC	5
PSP	4
XOne	3
GC	3
XB	2
WiiU	2
2600	1
GEN	1
Name: Platform, dtype: int64	
Action	37
Shooter	34
Platform	30
Role-Playing	28
Sports	15
Racing	14
Misc	13
Fighting	8
Puzzle	7
Simulation	7
Adventure	3
Strategy	3
Name: Genre, dtype: int64	
Nintendo	83
Activision	23
Sony Computer Entertainment	17
Microsoft Game Studios	15
Electronic Arts	12
Take-Two Interactive	10
Ubisoft	9
Bethesda Softworks	4
Capcom	4
Sega	3
Konami Digital Entertainment	2
Warner Bros. Interactive Entertainment	2
Virgin Interactive	2
LucasArts	2
Eidos Interactive	2
505 Games	2
SquareSoft	2
Sony Computer Entertainment Europe	1
GT Interactive	1
Square Enix	1
Atari	1
Name: Publisher, dtype: int64	

```
In [ ]: # Relative Frequency
r1= VideoGames['Rank'].value_counts(bins=4, normalize=True)

r2=VideoGames['Year'].value_counts(bins=4, normalize=True)

r3=VideoGames['NorthAmerica_Sales'].value_counts(bins=4, normalize=True)

r4=VideoGames['EuropeanUnion_Sales'].value_counts(bins=4, normalize= True)

r5=VideoGames['Japan_Sales'].value_counts(bins=4, normalize=True)

r6=VideoGames['Other_Sales'].value_counts(bins=4, normalize=True)

r7=VideoGames['Global_Sales'].value_counts(bins=4, normalize=True)

r8=VideoGames['Name'].value_counts(normalize=True)

r9=VideoGames['Platform'].value_counts(normalize=True)

r10=VideoGames['Genre'].value_counts(normalize=True)

r11=VideoGames['Publisher'].value_counts(normalize=True)

print(r1)
print(r2)
print(r3)
print(r4)
print(r5)
print(r6)
print(r7)
print(r8)
print(r9)
print(r10)
print(r11)
```

(56.75, 112.5] 0.266332  
(0.776, 56.75] 0.256281  
(112.5, 168.25] 0.246231  
(168.25, 224.0] 0.231156  
Name: Rank, dtype: float64  
(2007.5, 2016.0] 0.497487  
(1999.0, 2007.5] 0.241206  
(1990.5, 1999.0] 0.201005  
(1981.965, 1990.5] 0.060302  
Name: Year, dtype: float64  
(-0.0425, 10.372] 0.874372  
(10.372, 20.745] 0.035176  
(20.745, 31.118] 0.015075  
(31.118, 41.49] 0.005025  
Name: NorthAmerica\_Sales, dtype: float64  
(-0.0300000000000002, 7.255] 0.894472  
(7.255, 14.51] 0.055276  
(21.765, 29.02] 0.005025  
(14.51, 21.765] 0.000000  
Name: EuropeanUnion\_Sales, dtype: float64  
(-0.0112000000000002, 2.555] 0.763819  
(2.555, 5.11] 0.140704  
(5.11, 7.665] 0.040201  
(7.665, 10.22] 0.005025  
Name: Japan\_Sales, dtype: float64  
(-0.00946, 2.115] 0.909548  
(2.115, 4.23] 0.050251  
(6.345, 8.46] 0.005025  
(4.23, 6.345] 0.000000  
Name: Other\_Sales, dtype: float64  
(4.681, 24.255] 0.889447  
(24.255, 43.75] 0.045226  
(63.245, 82.74] 0.005025  
(43.75, 63.245] 0.000000  
Name: Global\_Sales, dtype: float64  
Grand Theft Auto V 0.020101  
Call of Duty: World at War 0.010050  
Dr. Mario 0.010050  
Call of Duty: Black Ops 3 0.010050  
Call of Duty: Black Ops II 0.010050  
...  
Wii Party 0.005025  
Halo 2 0.005025  
Mario Party 8 0.005025  
Pokemon Black 2/Pokemon White 2 0.005025  
Driver 2 0.005025  
Name: Name, Length: 167, dtype: float64  
PS3 0.155779  
X360 0.150754  
Wii 0.105528  
DS 0.100503  
PS 0.100503  
GB 0.055276  
PS4 0.045226  
3DS 0.040201  
N64 0.040201  
GBA 0.035176  
NES 0.035176  
SNES 0.030151

PC	0.025126
PSP	0.020101
XOne	0.015075
GC	0.015075
XB	0.010050
WiiU	0.010050
2600	0.005025
GEN	0.005025
Name: Platform, dtype: float64	
Action	0.185930
Shooter	0.170854
Platform	0.150754
Role-Playing	0.140704
Sports	0.075377
Racing	0.070352
Misc	0.065327
Fighting	0.040201
Puzzle	0.035176
Simulation	0.035176
Adventure	0.015075
Strategy	0.015075
Name: Genre, dtype: float64	
Nintendo	0.419192
Activision	0.116162
Sony Computer Entertainment	0.085859
Microsoft Game Studios	0.075758
Electronic Arts	0.060606
Take-Two Interactive	0.050505
Ubisoft	0.045455
Bethesda Softworks	0.020202
Capcom	0.020202
Sega	0.015152
Konami Digital Entertainment	0.010101
Warner Bros. Interactive Entertainment	0.010101
Virgin Interactive	0.010101
LucasArts	0.010101
Eidos Interactive	0.010101
505 Games	0.010101
SquareSoft	0.010101
Sony Computer Entertainment Europe	0.005051
GT Interactive	0.005051
Square Enix	0.005051
Atari	0.005051
Name: Publisher, dtype: float64	

```
In [ ]: #Percentage Frequencies  
pf1=r1*100  
  
pf2=r2*100  
  
pf3=r3*100  
  
pf4=r4*100  
  
pf5=r5*100  
  
pf6=r6*100  
  
pf7=r7*100  
  
pf8=r8*100  
  
pf9=r9*100  
  
pf10=r10*100  
  
pf11=r11*100  
  
print(pf1)  
print(pf2)  
print(pf3)  
print(pf4)  
print(pf5)  
print(pf6)  
print(pf7)  
print(pf8)  
print(pf9)  
print(pf10)  
print(pf11)
```

```
(56.75, 112.5]      26.633166
(0.776, 56.75]     25.628141
(112.5, 168.25]    24.623116
(168.25, 224.0]    23.115578
Name: Rank, dtype: float64
(2007.5, 2016.0]    49.748744
(1999.0, 2007.5]    24.120603
(1990.5, 1999.0]   20.100503
(1981.965, 1990.5]  6.030151
Name: Year, dtype: float64
(-0.0425, 10.372]   87.437186
(10.372, 20.745]    3.517588
(20.745, 31.118]    1.507538
(31.118, 41.49]     0.502513
Name: NorthAmerica_Sales, dtype: float64
(-0.0300000000000002, 7.255]    89.447236
(7.255, 14.51]        5.527638
(21.765, 29.02]       0.502513
(14.51, 21.765]      0.000000
Name: EuropeanUnion_Sales, dtype: float64
(-0.0112000000000002, 2.555]    76.381910
(2.555, 5.11]          14.070352
(5.11, 7.665]          4.020101
(7.665, 10.22]         0.502513
Name: Japan_Sales, dtype: float64
(-0.00946, 2.115]     90.954774
(2.115, 4.23]          5.025126
(6.345, 8.46]          0.502513
(4.23, 6.345]          0.000000
Name: Other_Sales, dtype: float64
(4.681, 24.255]       88.944724
(24.255, 43.75]       4.522613
(63.245, 82.74]       0.502513
(43.75, 63.245]       0.000000
Name: Global_Sales, dtype: float64
Grand Theft Auto V           2.010050
Call of Duty: World at War  1.005025
Dr. Mario                      1.005025
Call of Duty: Black Ops 3    1.005025
Call of Duty: Black Ops II   1.005025
                                ...
Wii Party                     0.502513
Halo 2                        0.502513
Mario Party 8                 0.502513
Pokemon Black 2/Pokemon White 2 0.502513
Driver 2                       0.502513
Name: Name, Length: 167, dtype: float64
PS3      15.577889
X360     15.075377
Wii      10.552764
DS       10.050251
PS       10.050251
GB       5.527638
PS4      4.522613
3DS      4.020101
N64      4.020101
GBA      3.517588
NES      3.517588
SNES     3.015075
```

```

PC      2.512563
PSP     2.010050
XOne    1.507538
GC      1.507538
XB      1.005025
WiiU    1.005025
2600    0.502513
GEN     0.502513
Name: Platform, dtype: float64
Action    18.592965
Shooter   17.085427
Platform  15.075377
Role-Playing 14.070352
Sports    7.537688
Racing    7.035176
Misc      6.532663
Fighting  4.020101
Puzzle    3.517588
Simulation 3.517588
Adventure 1.507538
Strategy   1.507538
Name: Genre, dtype: float64
Nintendo 41.919192
Activision 11.616162
Sony Computer Entertainment 8.585859
Microsoft Game Studios 7.575758
Electronic Arts 6.060606
Take-Two Interactive 5.050505
Ubisoft    4.545455
Bethesda Softworks 2.020202
Capcom    2.020202
Sega      1.515152
Konami Digital Entertainment 1.010101
Warner Bros. Interactive Entertainment 1.010101
Virgin Interactive 1.010101
LucasArts 1.010101
Eidos Interactive 1.010101
505 Games 1.010101
SquareSoft 1.010101
Sony Computer Entertainment Europe 0.505051
GT Interactive 0.505051
Square Enix 0.505051
Atari     0.505051
Name: Publisher, dtype: float64

```

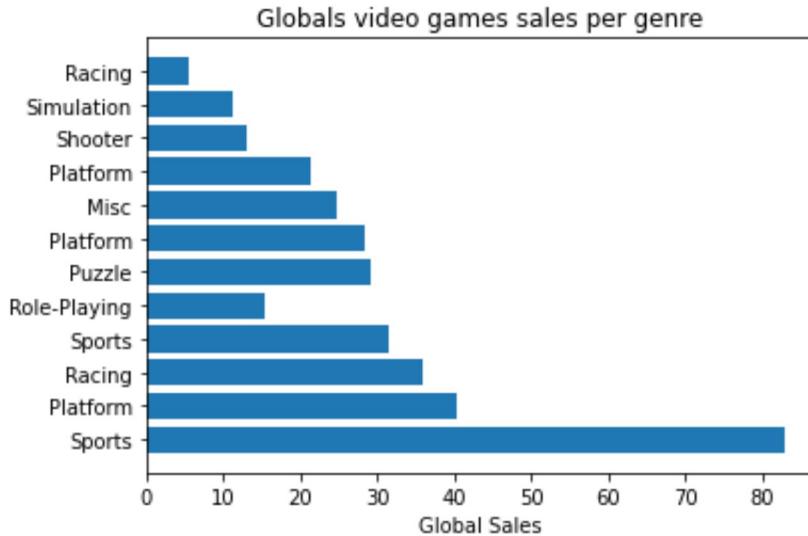
## Data Visualization

```

In [ ]: #A barplot showing the global sales for different video game genres
%matplotlib inline
xpos=np.arange(len(VideoGames['Genre']))
plt.yticks(xpos,VideoGames['Genre'])
plt.xlabel("Global Sales")
plt.title('Global video games sales per genre')
plt.barh(VideoGames['Genre'],VideoGames['Global_Sales'])

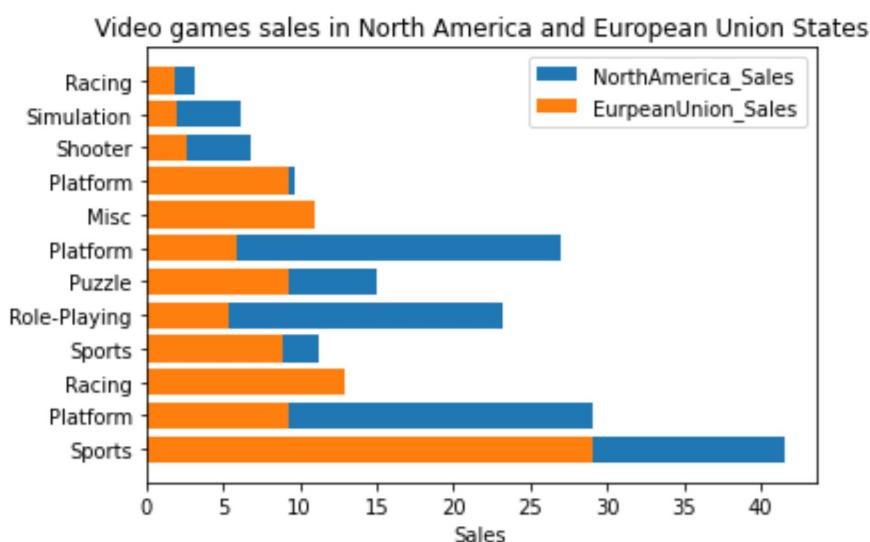
```

Out[ ]: <BarContainer object of 199 artists>



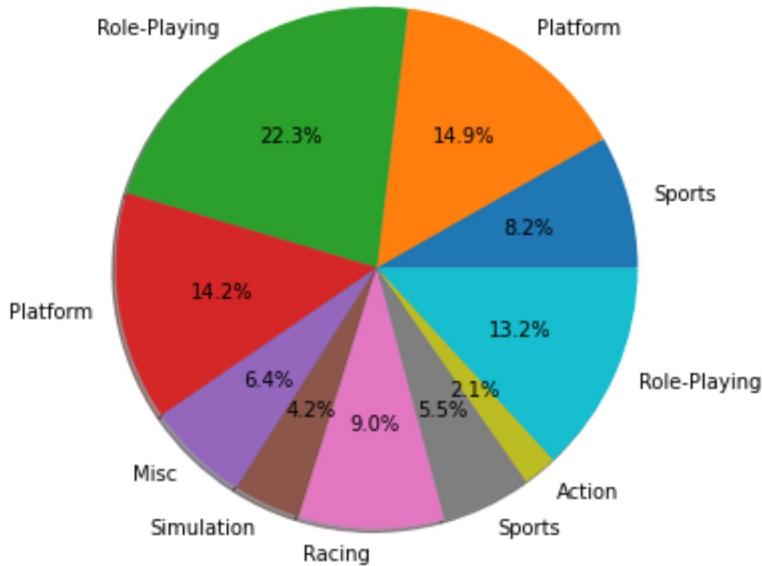
```
In [ ]: # Horizontal bar graphs showing video games sales in North America and Europe
%matplotlib inline
xpos=np.arange(len(VideoGames['Genre']))
plt.yticks(xpos,VideoGames['Genre'])
plt.xlabel("Sales")
plt.title('Video games sales in North America and European Union States')
plt.barh(VideoGames['Genre'],VideoGames['NorthAmerica_Sales'], label="NorthA")
plt.barh(VideoGames['Genre'],VideoGames['EuropeanUnion_Sales'], label="Europea")
plt.legend()
```

```
Out[ ]: <matplotlib.legend.Legend at 0x289edf00f70>
```



```
In [ ]: VideoGames.dropna(axis=0, inplace=True)
```

```
In [ ]: # A pie chart showing global distribution of sales using different platforms
sales_value=VideoGames['Japan_Sales'].head(10)
genre=VideoGames['Genre'].head(10)
plt.axis("equal")
plt.pie(sales_value, labels=genre, shadow=True, autopct='%.1f%%', radius=1.5
plt.show()
```



Calculating the mean, median, mode(s), variance and standard deviation for quantitative attributes

```
In [ ]: import statistics as stats
x=VideoGames['NorthAmerica_Sales']
print(np.mean(x))
print(np.median(x))
print(stats.mode(x))
print(np.var(x))
print(np.std(x))
```

4.26473333333334  
3.195000000000003  
3.66  
19.35709026222223  
4.399669335554915

```
In [ ]: import statistics as stats
x=VideoGames['EuropeanUnion_Sales']
print(np.mean(x))
print(np.median(x))
print(stats.mode(x))
print(np.var(x))
print(np.std(x))
```

2.921  
2.15  
1.24  
8.190599666666667  
2.86192237257873

```
In [ ]: import statistics as stats
x=VideoGames['Japan_Sales']
print(np.mean(x))
print(np.median(x))
print(stats.mode(x))
print(np.var(x))
print(np.std(x))
```

```
1.1839333333333335  
0.44  
0.0  
2.6529905288888886  
1.628800334261044
```

```
In [ ]: import statistics as stats  
x=VideoGames['Other_Sales']  
print(np.mean(x))  
print(np.median(x))  
print(stats.mode(x))  
print(np.var(x))  
print(np.std(x))
```

```
0.8249333333333334  
0.635  
0.2  
0.7955529955555556  
0.8919377756074442
```

```
In [ ]: import statistics as stats  
x=VideoGames['Global_Sales']  
print(np.mean(x))  
print(np.median(x))  
print(stats.mode(x))  
print(np.var(x))  
print(np.std(x))
```

```
9.193933333333332  
6.725  
8.49  
67.0909878622222  
8.190908854469216
```

Instances in the 1st, 2nd, and 3rd Standard Deviations for the quantitative columns

```
In [ ]: Mean=np.mean(VideoGames['NorthAmerica_Sales'])  
SD=np.std(VideoGames['NorthAmerica_Sales'])  
  
x= [Mean - SD,Mean + SD] # 1st instances  
y= [Mean - 2*SD,Mean + 2*SD] # 2nd instances  
z= [Mean - 3*SD,Mean + 3*SD] # 3rd instances  
  
print(x)  
print(y)  
print(z)
```

```
[-0.13493600222158086, 8.664402668888249]  
[-4.534605337776496, 13.064072004443164]  
[-8.93427467333141, 17.46374133999808]
```

```
In [ ]: Mean=np.mean(VideoGames['EuropeanUnion_Sales'])
SD=np.std(VideoGames['EuropeanUnion_Sales'])

x= [Mean - SD,Mean + SD] # 1st instances
y= [Mean - 2*SD,Mean + 2*SD] # 2nd instances
z= [Mean - 3*SD,Mean + 3*SD] # 3rd instances

print(x)
print(y)
print(z)

[0.059077627421269874, 5.78292237257873]
[-2.80284474515746, 8.64484474515746]
[-5.66476711773619, 11.506767117736189]
```

```
In [ ]: Mean=np.mean(VideoGames['Japan_Sales'])
SD=np.std(VideoGames['Japan_Sales'])

x= [Mean - SD,Mean + SD] # 1st instances
y= [Mean - 2*SD,Mean + 2*SD] # 2nd instances
z= [Mean - 3*SD,Mean + 3*SD] # 3rd instances

print(x)
print(y)
print(z)

[-0.4448670009277105, 2.8127336675943777]
[-2.0736673351887545, 4.4415340018554215]
[-3.7024676694497987, 6.070334336116465]
```

```
In [ ]: Mean=np.mean(VideoGames['Other_Sales'])
SD=np.std(VideoGames['Other_Sales'])

x= [Mean - SD,Mean + SD] # 1st instances
y= [Mean - 2*SD,Mean + 2*SD] # 2nd instances
z= [Mean - 3*SD,Mean + 3*SD] # 3rd instances

print(x)
print(y)
print(z)

[-0.06700444227411084, 1.7168711089407775]
[-0.9589422178815551, 2.608808884548222]
[-1.850879993488999, 3.500746660155666]
```

```
In [ ]: Mean=np.mean(VideoGames['Global_Sales'])
SD=np.std(VideoGames['Global_Sales'])

x= [Mean - SD,Mean + SD] # 1st instances
y= [Mean - 2*SD,Mean + 2*SD] # 2nd instances
z= [Mean - 3*SD,Mean + 3*SD] # 3rd instances

print(x)
print(y)
print(z)

[1.0030244788641163, 17.384842187802548]
[-7.187884375605099, 25.57575104227176]
[-15.378793230074317, 33.76665989674098]
```

## Modes for the categorical attribute Outcome

```
In [ ]: print(stats.mode(VideoGames['Genre']))
print(stats.mode(VideoGames['Publisher']))
print(stats.mode(VideoGames['Platform']))
print(stats.mode(VideoGames['Name']))
```

```
Action
Nintendo
PS3
Grand Theft Auto V
```

These are the insights obtained from the analysis

In the sales of video games, action genre was the most sold.

The publisher that did many sales was the Nintendo Publishers.

PS3 was the frequently used platform for playing the games,

The name that appeared most frequently was the Grand Theft Auto V

Sports genre was the most sold video games in North America and European union.