Data Analysis

Video Game Sales Analysis

James Bond 12/16/2019

In [9]:

```
import pandas as pd
```

In [10]:

```
print(pd)
```

```
<module 'pandas' from '/Users/jamesbond/.local/share/virtualenvs/401-Yo
pfmMgc/lib/python3.7/site-packages/pandas/__init__.py'>
```

In [11]:

```
import numpy as np
```

In [12]:

```
print(np)
```

```
<module 'numpy' from '/Users/jamesbond/.local/share/virtualenvs/401-Yop
fmMgc/lib/python3.7/site-packages/numpy/__init__.py'>
```

In [15]:

```
df=pd.read_csv('vgsales.csv')
```

Which company is the most common video game publisher?

In [16]:

```
df['Publisher'].mode()
```

Out[16]:

```
0    Electronic Arts
dtype: object
```

What's the most common platform?

In [78]:

```
df['Platform'].mode()
```

Out[78]:

```
0    DS
dtype: object
```

What about the most common genre?

In [79]:

```
df['Genre'].mode()
```

Out[79]:

```
0    Action
dtype: object
```

What are the top 20 highest grossing games?

In [39]:

```
df[['Name', 'Global_Sales']].sort_values('Global_Sales', ascending=False).head(20)
```

Out[39]:

	Name	Global_Sales
0	Wii Sports	82.74
1	Super Mario Bros.	40.24
2	Mario Kart Wii	35.82
3	Wii Sports Resort	33.00
4	Pokemon Red/Pokemon Blue	31.37
5	Tetris	30.26
6	New Super Mario Bros.	30.01
7	Wii Play	29.02
8	New Super Mario Bros. Wii	28.62
9	Duck Hunt	28.31
10	Nintendogs	24.76
11	Mario Kart DS	23.42
12	Pokemon Gold/Pokemon Silver	23.10
13	Wii Fit	22.72
14	Wii Fit Plus	22.00
15	Kinect Adventures!	21.82
16	Grand Theft Auto V	21.40
17	Grand Theft Auto: San Andreas	20.81
18	Super Mario World	20.61
19	Brain Age: Train Your Brain in Minutes a Day	20.22

For North American video game sales, what's the median?

In [99]:

```
df['NA_Sales'].median()
```

Out[99]:

```
0.08
```

ten games surrounding the median sales output:

In [19]:

```
x = df['NA_Sales'].between(0.07999999999999,0.08000000000001, inclusive=True)
df[x].head(10)
```

Out[19]:

	Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sal
446	447	Dragon Warrior IV	NES	1990.0	Role-Playing	Enix Corporation	0.08	0.00	3.(
497	498	World Soccer Winning Eleven 7 International	PS2	2003.0	Sports	Konami Digital Entertainment	0.08	1.24	1.1
1617	1619	Farming Simulator 2015	PC	2014.0	Simulation	Focus Home Interactive	0.08	1.02	0.(
1926	1928	Pro Evolution Soccer 2008	X360	2007.0	Sports	Konami Digital Entertainment	0.08	0.90	0.(
2067	2069	Winning Eleven: Pro Evolution Soccer 2007 (All...	X360	2006.0	Sports	Konami Digital Entertainment	0.08	0.90	0.(
2373	2375	Phantasy Star Portable 2	PSP	2009.0	Role-Playing	Sega	0.08	0.11	0.(
2579	2581	The Sims 2: Castaway	PSP	2007.0	Simulation	Electronic Arts	0.08	0.46	0.(
3186	3188	SingStar Queen	PS2	2009.0	Misc	Sony Computer Entertainment	0.08	0.12	0.(
3503	3505	Top Spin 3	PS3	2008.0	Action	Take-Two Interactive	0.08	0.37	0.(
3703	3705	Sonic & All-Stars Racing Transformed	PS3	2012.0	Racing	Sega	0.08	0.33	0.(

For the top-selling game of all time, how many standard deviations above/below the mean are its sales for North America?

In [20]:

```
(df['NA_Sales'].max()-df['NA_Sales'].mean())/df['NA_Sales'].std()
```

Out[20]:

```
50.47898767479108
```

How does Wii's average number of sales compare with all of the other platforms?

In [74]:

```
df['Global_Sales'].groupby(df['Platform']).mean().sort_values(ascending=False)
```

Out[74]:

```
Platform
GB      2.606633
NES     2.561939
GEN     1.050370
SNES    0.837029
PS4     0.827679
X360    0.774672
2600    0.729925
PS3     0.720722
Wii     0.699404
N64     0.686144
XOne    0.662254
PS      0.610920
PS2     0.581046
WiiU    0.572448
3DS     0.486169
GBA     0.387470
DS      0.380254
GC      0.358561
XB      0.313422
SCD     0.311667
DC      0.307115
PC      0.269604
PSP     0.244254
WS      0.236667
SAT     0.194162
PSV     0.149952
NG      0.120000
TG16    0.080000
GG      0.040000
3DO     0.033333
PCFX    0.030000
Name: Global_Sales, dtype: float64
```

What are total Global Sales of top 5 Publishers?

In [88]:

```
df['Global_Sales'].groupby(df['Publisher']).sum().sort_values(ascending=False).head(5)
```

Out[88]:

```
Publisher
Nintendo                        1786.56
Electronic Arts                 1110.32
Activision                       727.46
Sony Computer Entertainment      607.50
Ubisoft                          474.72
Name: Global_Sales, dtype: float64
```

What are the number games of top 10 Publishers?

In [92]:

```
df['Name'].groupby(df['Publisher']).count().sort_values(ascending=False).head(10)
```

Out[92]:

```
Publisher
Electronic Arts                 1351
Activision                       975
Namco Bandai Games               932
Ubisoft                          921
Konami Digital Entertainment     832
THQ                              715
Nintendo                         703
Sony Computer Entertainment      683
Sega                             639
Take-Two Interactive             413
Name: Name, dtype: int64
```

What are the number games of top 10 Platforms?

In [94]:

```
df['Name'].groupby(df['Platform']).count().sort_values(ascending=False).head(10)
```

Out[94]:

```
Platform
DS      2163
PS2     2161
PS3     1329
Wii     1325
X360    1265
PSP     1213
PS      1196
PC       960
XB       824
GBA      822
Name: Name, dtype: int64
```

What is the difference of sales amount first two top Platforms?

In [97]:

```
(df['Global_Sales'].groupby(df['Platform']).sum().sort_values(ascending=False).head
()[0])-(df['Global_Sales'].groupby(df['Platform']).sum().sort_values(ascending=Fals
e).head()[1])
```

Out[97]:

```
275.67999999998756
```

What is the difference of sales amount first two top Publishers?

In [98]:

```
(df['Global_Sales'].groupby(df['Publisher']).sum().sort_values(ascending=False).hea
d()[0])-(df['Global_Sales'].groupby(df['Publisher']).sum().sort_values(ascending=Fa
lse).head()[1])
```

Out[98]:

```
676.2400000000066
```

In [ ]:

```

```