Day_41_071223

January 23, 2024

1 Correlation

1.1 If two values are correlated

1.1.1 Positively Correlated - If one value increase other value also increase

1.1.2 Negative Correlated - If one increase other descrease vice versa

For example - If you take more calories the weight will also increase So Calories and Weight are positively Correlated

[37]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
data = pd.read_csv("final_vg.csv")
```

[38]:
```
data.head()
```

[38]:

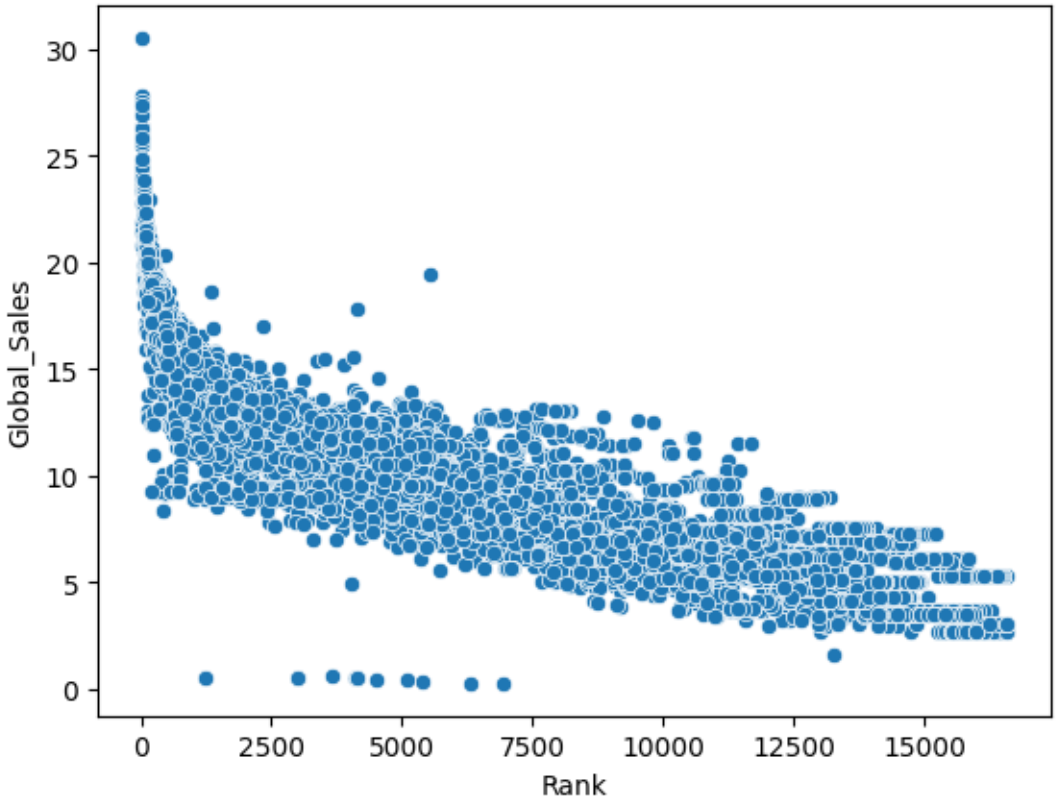
	Unnamed: 0	Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
0	0	2061	1942	NES	1985.0	Shooter	Capcom	4.569217	3.033887	3.439352	1.991671	12.802935
1	1	9137	¡Shin Chan Flipa en colores!	DS	2007.0	Platform	505 Games	2.076955	1.493442	3.033887	0.394830	7.034163
2	2	14279	.hack: Sekai no Mukou ni + Versus	PS3	2012.0	Action	Namco Bandai Games	1.145709	1.762339	1.493442	0.408693	4.982552
3	3	8359	.hack//G.U. Vol.1//Rebirth	PS2	2006.0	Role-Playing	Namco Bandai Games	2.031986	1.389856	3.228043	0.394830	7.226880
4	4	7109	.hack//G.U. Vol.2//Reminisce	PS2	2006.0	Role-Playing	Namco Bandai Games	2.792725	2.592054	1.440483	1.493442	8.363113

2 Suppose we have find relation between two data points

3 Scatter Plot using Seaborn

[39]:
```
sns.scatterplot(data,x='Rank',y='Global_Sales')
```

[39]:
```
<Axes: xlabel='Rank', ylabel='Global_Sales'>
```



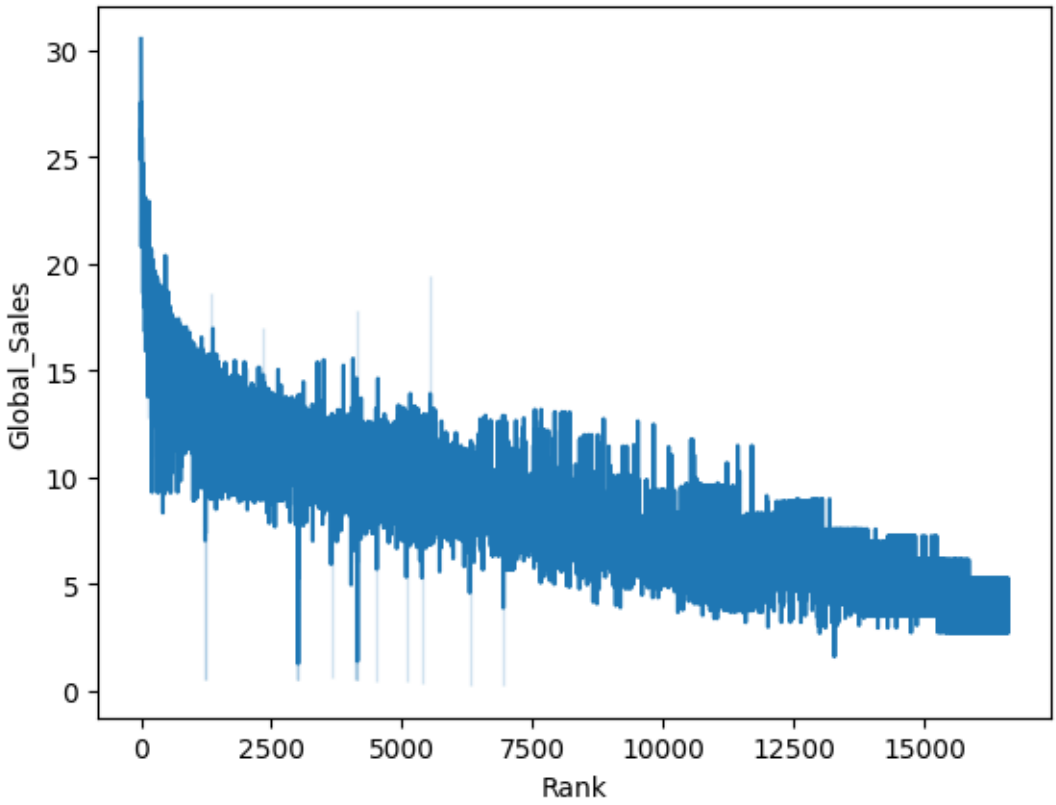
- In this scatter plot we can observe that when there is higher sales the rank is less and vice-versa
- So it is negatively correlated

3.0.1 Same can be visualized using line plot

4 Line plot using Seabornm

[40]:
```
sns.lineplot(data,x='Rank',y='Global_Sales')
```

[40]:
```
<Axes: xlabel='Rank', ylabel='Global_Sales'>
```



- But it is not that effective So we consider scatter plot

Upto now we looked into continous - continous data

5 Categorical - Categorical data

[41]:
```
data.head(3)
```

[41]:
```
   Unnamed: 0   Rank                                 Name Platform    Year  \
0           0   2061                                 1942      NES  1985.0
1           1   9137       ¡Shin Chan Flipa en colores!       DS  2007.0
2           2  14279  .hack: Sekai no Mukou ni + Versus      PS3  2012.0

      Genre           Publisher  NA_Sales  EU_Sales  JP_Sales  Other_Sales  \
0   Shooter             Capcom  4.569217  3.033887  3.439352     1.991671
1  Platform          505 Games  2.076955  1.493442  3.033887     0.394830
2    Action  Namco Bandai Games  1.145709  1.762339  1.493442     0.408693

   Global_Sales
0     12.802935
```

```
1      7.034163
2      4.982552
```

[42]:
```
top3_pub = data['Publisher'].value_counts().index[:3]
top3_gen = data['Genre'].value_counts().index[:3]
top3_plat = data['Platform'].value_counts().index[:3]
```

[43]:
```
top3_pub
```

[43]:
```
Index(['Electronic Arts', 'Activision', 'Namco Bandai Games'], dtype='object',
name='Publisher')
```

[44]:
```
top3_gen
```

[44]:
```
Index(['Action', 'Sports', 'Misc'], dtype='object', name='Genre')
```

[45]:
```
top3_plat
```

[45]:
```
Index(['DS', 'PS2', 'PS3'], dtype='object', name='Platform')
```

[46]:
```
top3_data = data.loc[(data['Publisher'].isin(top3_pub)) & (data['Genre'].
↪isin(top3_gen)) & (data['Platform'].isin(top3_plat))]
```

[47]:
```
top3_data
```

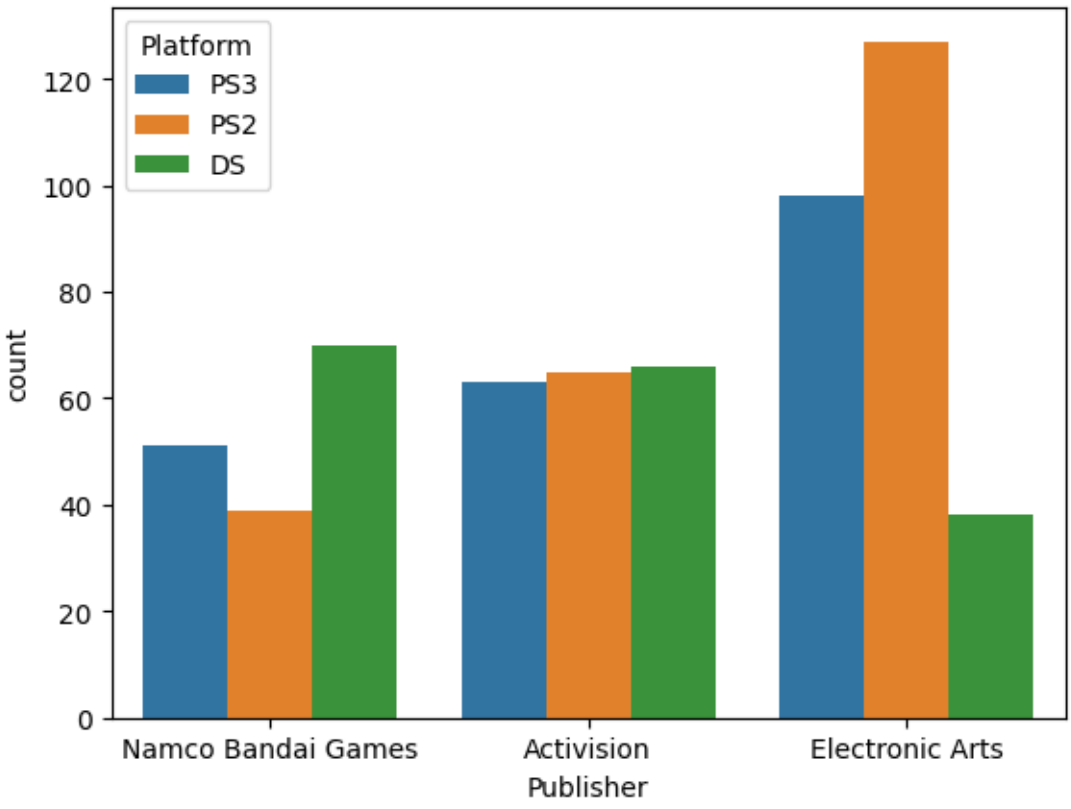
[47]:
```
       Unnamed: 0   Rank                                                Name  \
2               2  14279                 .hack: Sekai no Mukou ni + Versus
13             13   2742                                       [Prototype 2]
16             16   1604                                         [Prototype]
19             19   1741                           007: Quantum of Solace
21             21   4501                           007: Quantum of Solace
…             …    …                                                   …
16438       16438  14938  Yes! Precure 5 Go Go Zenin Shu Go! Dream Festival
16479       16479  10979                           Young Justice: Legacy
16601       16601  11802                      ZhuZhu Pets: Quest for Zhu
16636       16636   9196                        Zoobles! Spring to Life!
16640       16640   9816                                            Zubo

      Platform    Year   Genre           Publisher  NA_Sales  EU_Sales  \
2          PS3  2012.0  Action  Namco Bandai Games  1.145709  1.762339
13         PS3  2012.0  Action         Activision  3.978349  3.727034
16         PS3  2009.0  Action         Activision  4.569217  4.108402
19         PS3  2008.0  Action         Activision  4.156030  4.346074
21         PS2  2008.0  Action         Activision  3.228043  2.738800
…        …      …     …                      …       …       …
16438       DS  2008.0  Action  Namco Bandai Games  1.087977  0.592445
16479      PS3  2013.0  Action  Namco Bandai Games  2.186589  1.087977
16601       DS  2011.0    Misc         Activision  2.340740  1.525543
```

```
16636       DS  2011.0    Misc          Activision  2.697415  1.087977
16640       DS  2008.0    Misc     Electronic Arts  2.592054  1.493442

       JP_Sales  Other_Sales  Global_Sales
2      1.493442     0.408693      4.982552
13     0.848807     2.792725     11.447989
16     1.187272     3.339269     13.181205
19     1.087977     3.390562     12.980643
21     2.585598     3.652926     11.780257
…           …            …            …
16438  1.087977     0.394830      3.509168
16479  3.409089     0.394830      7.359902
16601  3.103825     0.394830      7.372592
16636  2.760718     0.394830      6.915540
16640  1.493442     0.394830      5.969572

[617 rows x 12 columns]
```

6 Dodged Bar chart

[48]:
```
sns.countplot(x='Publisher',data=top3_data,hue='Platform')
```

[48]:
```
<Axes: xlabel='Publisher', ylabel='count'>
```



7 Multivariate data

8 Heatmap - It shows the correlation between numerical columns

[49]:
```
top3_multi = top3_data.drop(['Name','Unnamed:␣
↪0','Platform','Genre','Publisher'],axis=1)
```

[52]:
```
sns.heatmap(top3_multi.corr(),cmap='Blues',annot=True)
```

[52]: <Axes: >