

In [1]:

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

In [2]:

```
data=pd.read_csv('IMDB-Movie-Data.csv')
```

In [4]:

data.head (10)

Out[4]:

	Rank	Title	Genre	Description	Director	Actors	Year
0	1	Guardians of the Galaxy	Action,Adventure,Sci-Fi	A group of intergalactic criminals are forced ...	James Gunn	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	20
1	2	Prometheus	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fa...	20
2	3	Split	Horror,Thriller	Three girls are kidnapped by a man with a diag...	M. Night Shyamalan	James McAvoy, Anya Taylor-Joy, Haley Lu Richar...	20
3	4	Sing	Animation,Comedy,Family	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet	Matthew McConaughey,Reese Witherspoon, Seth Ma...	20
4	5	Suicide Squad	Action,Adventure,Fantasy	A secret government agency recruits some of th...	David Ayer	Will Smith, Jared Leto, Margot Robbie, Viola D...	20
5	6	The Great Wall	Action,Adventure,Fantasy	European mercenaries searching for black powde...	Yimou Zhang	Matt Damon, Tian Jing, Willem Dafoe, Andy Lau	20
6	7	La La Land	Comedy,Drama,Music	A jazz pianist falls for an aspiring actress i...	Damien Chazelle	Ryan Gosling, Emma Stone, Rosemarie DeWitt, J....	20
7	8	Mindhorn	Comedy	A has-been actor best known for playing the ti...	Sean Foley	Essie Davis, Andrea Riseborough, Julian Barrat...	20
8	9	The Lost City of Z	Action,Adventure,Biography	A true-life drama, centering on British explor...	James Gray	Charlie Hunnam, Robert Pattinson, Sienna Mille...	20
9	10	Passengers	Adventure,Drama,Romance	A spacecraft traveling to a distant colony pla...	Morten Tyldum	Jennifer Lawrence, Chris Pratt, Michael Sheen,...	20

In [6]:

data.tail(10)

Out[6]:

Rank		Title	Genre	Description	Director	Actors	Year
990	991	Underworld: Rise of the Lycans	Action,Adventure,Fantasy	An origins story centered on the centuries-old...	Patrick Tatopoulos	Rhona Mitra, Michael Sheen, Bill Nighy, Steven...	2009
991	992	Taare Zameen Par	Drama,Family,Music	An eight-year-old boy is thought to be a lazy ...	Aamir Khan	Darsheel Safary, Aamir Khan, Tanay Chheda, Sac...	2007
992	993	Take Me Home Tonight	Comedy,Drama,Romance	Four years after graduation, an awkward high s...	Michael Dowse	Topher Grace, Anna Faris, Dan Fogler, Teresa P...	2011
993	994	Resident Evil: Afterlife	Action,Adventure,Horror	While still out to destroy the evil Umbrella C...	Paul W.S. Anderson	Milla Jovovich, Ali Larter, Wentworth Miller,K...	2010
994	995	Project X	Comedy	3 high school seniors throw a birthday party t...	Nima Nourizadeh	Thomas Mann, Oliver Cooper, Jonathan Daniel Br...	2012
995	996	Secret in Their Eyes	Crime,Drama,Mystery	A tight-knit team of rising investigators, alo...	Billy Ray	Chiwetel Ejiofor, Nicole Kidman, Julia Roberts...	2015
996	997	Hostel: Part II	Horror	Three American college students studying abroa...	Eli Roth	Lauren German, Heather Matarazzo, Bijou Philli...	2007
997	998	Step Up 2: The Streets	Drama,Music,Romance	Romantic sparks occur between two dance studen...	Jon M. Chu	Robert Hoffman, Briana Evigan, Cassie Ventura,...	2008
998	999	Search Party	Adventure,Comedy	A pair of friends embark on a mission to reuni...	Scot Armstrong	Adam Pally, T.J. Miller, Thomas Middleditch,Sh...	2014
999	1000	Nine Lives	Comedy,Family,Fantasy	A stuffy businessman finds himself trapped ins...	Barry Sonnenfeld	Kevin Spacey, Jennifer Garner, Robbie Amell,Ch...	2016

In [7]:

```
data.shape
```

Out[7]:

```
(1000, 12)
```

In [8]:

```
print("Number of Rows",data.shape [0])  
print("Number of Columns",data.shape[1])
```

```
Number of Rows 1000  
Number of Columns 12
```

In [9]:

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 1000 entries, 0 to 999  
Data columns (total 12 columns):  
#   Column                Non-Null Count  Dtype    
---  ---                  
0   Rank                  1000 non-null  int64    
1   Title                 1000 non-null  object   
2   Genre                 1000 non-null  object   
3   Description           1000 non-null  object   
4   Director              1000 non-null  object   
5   Actors                1000 non-null  object   
6   Year                  1000 non-null  int64    
7   Runtime (Minutes)    1000 non-null  int64    
8   Rating                1000 non-null  float64  
9   Votes                 1000 non-null  int64    
10  Revenue (Millions)   872 non-null   float64  
11  Metascore             936 non-null   float64  
dtypes: float64(3), int64(4), object(5)  
memory usage: 93.9+ KB
```

In [10]:

```
print("Any missing value?",data.isnull().values.any())
```

```
Any missing value? True
```

In [12]:

```
data.isnull().sum()
```

Out[12]:

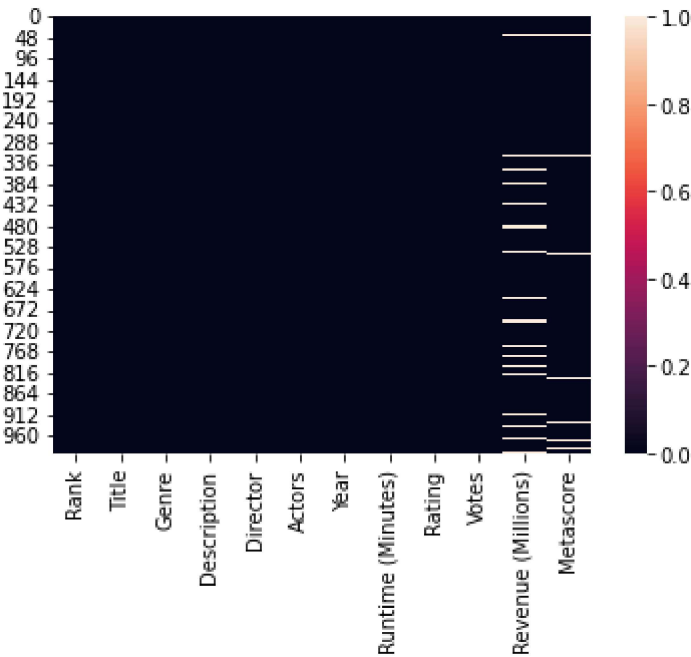
```
Rank          0
Title         0
Genre         0
Description   0
Director      0
Actors        0
Year          0
Runtime (Minutes)  0
Rating        0
Votes         0
Revenue (Millions) 128
Metascore     64
dtype: int64
```

In [13]:

```
persns.heatmap(data.isnull())
```

Out[13]:

<AxesSubplot:>



In [14]:

```
per_missing=data.isnull().sum() *100/len(data)
per_missing
```

Out[14]:

```
Rank          0.0
Title         0.0
Genre         0.0
Description   0.0
Director      0.0
Actors        0.0
Year          0.0
Runtime (Minutes) 0.0
Rating        0.0
Votes         0.0
Revenue (Millions) 12.8
Metascore     6.4
dtype: float64
```

In [15]:

```
data.dropna(axis=0)
```

Out[15]:

Rank		Title	Genre	Description	Director	Actors
0	1	Guardians of the Galaxy	Action,Adventure,Sci-Fi	A group of intergalactic criminals are forced ...	James Gunn	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...
1	2	Prometheus	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fa...
2	3	Split	Horror,Thriller	Three girls are kidnapped by a man with a diag...	M. Night Shyamalan	James McAvoy, Anya Taylor-Joy, Haley Lu Richar...
3	4	Sing	Animation,Comedy,Family	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet	Matthew McConaughey,Reese Witherspoon, Seth Ma...
4	5	Suicide Squad	Action,Adventure,Fantasy	A secret government agency recruits some of th...	David Ayer	Will Smith, Jared Leto, Margot Robbie, Viola D...
...	...	...	...	...	...	...
993	994	Resident Evil: Afterlife	Action,Adventure,Horror	While still out to destroy the evil Umbrella C...	Paul W.S. Anderson	Milla Jovovich, Ali Larter, Wentworth Miller,K...
994	995	Project X	Comedy	3 high school seniors throw a birthday party t...	Nima Nourizadeh	Thomas Mann, Oliver Cooper, Jonathan Daniel Br...
996	997	Hostel: Part II	Horror	Three American college students studying abroa...	Eli Roth	Lauren German, Heather Matarazzo, Bijou Philli...
997	998	Step Up 2: The Streets	Drama,Music,Romance	Romantic sparks occur between two dance studen...	Jon M. Chu	Robert Hoffman, Briana Evigan, Cassie Ventura,...
999	1000	Nine Lives	Comedy,Family,Fantasy	A stuffy businessman finds himself trapped ins...	Barry Sonnenfeld	Kevin Spacey, Jennifer Garner, Robbie Amell,Ch...

838 rows × 12 columns



In [16]:

```
dup_data=data.duplicated().any()
```

In [17]:

```
print("Are there any duplicated values?",dup_data)
```

Are there any duplicated values? False



In [18]:

```
data=data.drop_duplicates()  
data
```

Out[18]:

Rank		Title	Genre	Description	Director	Actors	
0	1	Guardians of the Galaxy	Action,Adventure,Sci-Fi	A group of intergalactic criminals are forced ...	James Gunn	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	2
1	2	Prometheus	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fa...	2
2	3	Split	Horror,Thriller	Three girls are kidnapped by a man with a diag...	M. Night Shyamalan	James McAvoy, Anya Taylor-Joy, Haley Lu Richar...	2
3	4	Sing	Animation,Comedy,Family	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet	Matthew McConaughey,Reese Witherspoon, Seth Ma...	2
4	5	Suicide Squad	Action,Adventure,Fantasy	A secret government agency recruits some of th...	David Ayer	Will Smith, Jared Leto, Margot Robbie, Viola D...	2
...	...	...	...	...	...	...	...
995	996	Secret in Their Eyes	Crime,Drama,Mystery	A tight-knit team of rising investigators, alo...	Billy Ray	Chiwetel Ejiofor, Nicole Kidman, Julia Roberts...	2
996	997	Hostel: Part II	Horror	Three American college students studying abroa...	Eli Roth	Lauren German, Heather Matarazzo, Bijou Philli...	2
997	998	Step Up 2: The Streets	Drama,Music,Romance	Romantic sparks occur between two dance studen...	Jon M. Chu	Robert Hoffman, Briana Evigan, Cassie Ventura,...	2
998	999	Search Party	Adventure,Comedy	A pair of friends embark on a mission to reuni...	Scot Armstrong	Adam Pally, T.J. Miller, Thomas Middleditch,Sh...	2
999	1000	Nine Lives	Comedy,Family,Fantasy	A stuffy businessman finds himself trapped ins...	Barry Sonnenfeld	Kevin Spacey, Jennifer Garner, Robbie Amell,Ch...	2

1000 rows × 12 columns

In [21]:

```
data.describe(include='all')
```

Out[21]:

	Rank	Title	Genre	Description	Director	Actors	Year
<b>count</b>	1000.000000	1000	1000	1000	1000	1000	1000.000000
<b>unique</b>	NaN	999	207	1000	644	996	NaN
<b>top</b>	NaN	The Host	Action,Adventure,Sci-Fi	A group of intergalactic criminals are forced ...	Ridley Scott	Jennifer Lawrence, Josh Hutcherson, Liam Hemsw...	NaN
<b>freq</b>	NaN	2	50	1	8	2	NaN
<b>mean</b>	500.500000	NaN	NaN	NaN	NaN	NaN	2012.783000
<b>std</b>	288.819436	NaN	NaN	NaN	NaN	NaN	3.205962
<b>min</b>	1.000000	NaN	NaN	NaN	NaN	NaN	2006.000000
<b>25%</b>	250.750000	NaN	NaN	NaN	NaN	NaN	2010.000000
<b>50%</b>	500.500000	NaN	NaN	NaN	NaN	NaN	2014.000000
<b>75%</b>	750.250000	NaN	NaN	NaN	NaN	NaN	2016.000000
<b>max</b>	1000.000000	NaN	NaN	NaN	NaN	NaN	2016.000000

In [22]:

```
data.columns
```

Out[22]:

```
Index(['Rank', 'Title', 'Genre', 'Description', 'Director', 'Actors', 'Year',
      'Runtime (Minutes)', 'Rating', 'Votes', 'Revenue (Millions)',
      'Metascore'],
      dtype='object')
```

In [26]:

```
data[data['Runtime (Minutes)']>=180]['Title']
```

Out[26]:

```
82      The Wolf of Wall Street
88      The Hateful Eight
311     La vie d'Adèle
828     Grindhouse
965     Inland Empire
Name: Title, dtype: object
```

In [27]:

```
data.columns
```

Out[27]:

```
Index(['Rank', 'Title', 'Genre', 'Description', 'Director', 'Actors', 'Year',
      'Runtime (Minutes)', 'Rating', 'Votes', 'Revenue (Millions)',
      'Metascore'],
      dtype='object')
```

In [29]:

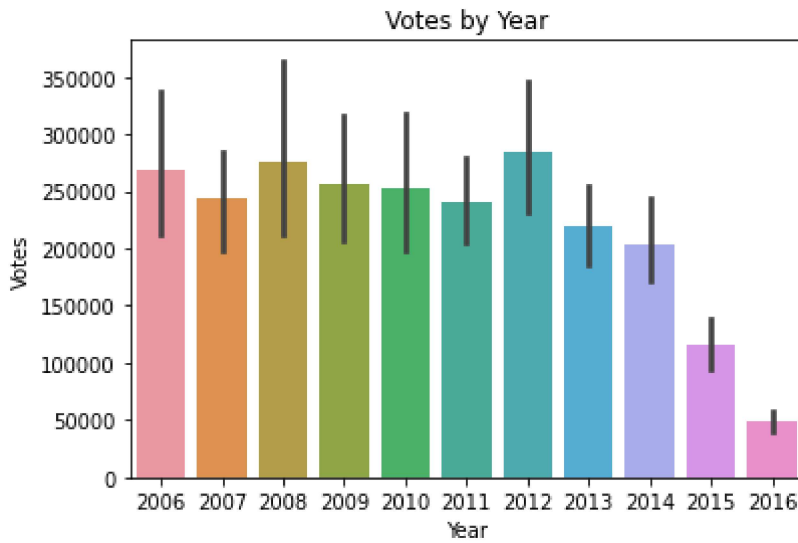
```
data.groupby('Year')['Votes'].mean().sort_values(ascending=False)
```

Out[29]:

```
Year
2012    285226.093750
2008    275505.384615
2006    269289.954545
2009    255780.647059
2010    252782.316667
2007    244331.037736
2011    240790.301587
2013    219049.648352
2014    203930.224490
2015    115726.220472
2016     48591.754209
Name: Votes, dtype: float64
```

In [35]:

```
sns.barplot(x='Year',y='Votes',data=data)
plt.title("Votes by Year")
plt.show()
```



In [36]:

```
data.columns
```

Out[36]:

```
Index(['Rank', 'Title', 'Genre', 'Description', 'Director', 'Actors', 'Year',
      'Runtime (Minutes)', 'Rating', 'Votes', 'Revenue (Millions)',
      'Metascore'],
      dtype='object')
```

In [37]:

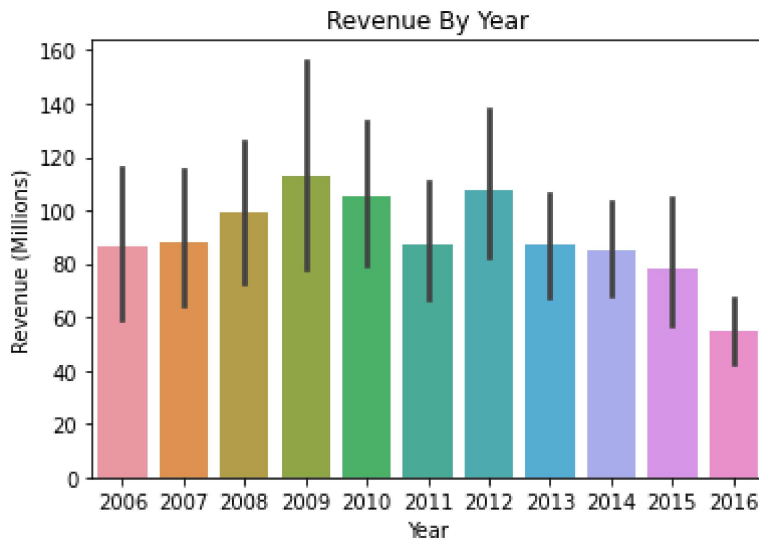
```
data.groupby('Year')['Revenue (Millions)'].mean().sort_values(ascending=False)
```

Out[37]:

```
Year
2009    112.601277
2012    107.973281
2010    105.081579
2008     99.082745
2007     87.882245
2011     87.612258
2013     87.121818
2006     86.296667
2014     85.078723
2015     78.355044
2016     54.690976
Name: Revenue (Millions), dtype: float64
```

In [42]:

```
sns.barplot(x='Year',y='Revenue (Millions)',data=data)
plt.title("Revenue By Year")
plt.show()
```



In [43]:

```
data.columns
```

Out[43]:

```
Index(['Rank', 'Title', 'Genre', 'Description', 'Director', 'Actors', 'Year',
      'Runtime (Minutes)', 'Rating', 'Votes', 'Revenue (Millions)',
      'Metascore'],
      dtype='object')
```

In [45]:

```
data.groupby('Director')['Rating'].mean().sort_values(ascending=False)
```

Out[45]:

```
Director
Nitesh Tiwari      8.80
Christopher Nolan  8.68
Olivier Nakache    8.60
Makoto Shinkai     8.60
Aamir Khan         8.50
...
Micheal Bafaro     3.50
Jonathan Holbrook  3.20
Shawn Burkett      2.70
James Wong         2.70
Jason Friedberg    1.90
Name: Rating, Length: 644, dtype: float64
```

In [46]:

```
data.columns
```

Out[46]:

```
Index(['Rank', 'Title', 'Genre', 'Description', 'Director', 'Actors', 'Year',  
      'Runtime (Minutes)', 'Rating', 'Votes', 'Revenue (Millions)',  
      'Metascore'],  
      dtype='object')
```

In [55]:

```
data.nlargest(10, 'Runtime (Minutes)')[['Title', 'Runtime (Minutes)']]\  
.set_index('Title')
```

Out[55]:

Runtime (Minutes)	
Title	
Grindhouse	191
The Hateful Eight	187
The Wolf of Wall Street	180
La vie d'Adèle	180
Inland Empire	180
Cloud Atlas	172
3 Idiots	170
Interstellar	169
Pirates of the Caribbean: At World's End	169
The Hobbit: An Unexpected Journey	169

In [70]:

```
data.columns
```

Out[70]:

```
Index(['Rank', 'Title', 'Genre', 'Description', 'Director', 'Actors', 'Year',  
      'Runtime (Minutes)', 'Rating', 'Votes', 'Revenue (Millions)',  
      'Metascore'],  
      dtype='object')
```

In [71]:

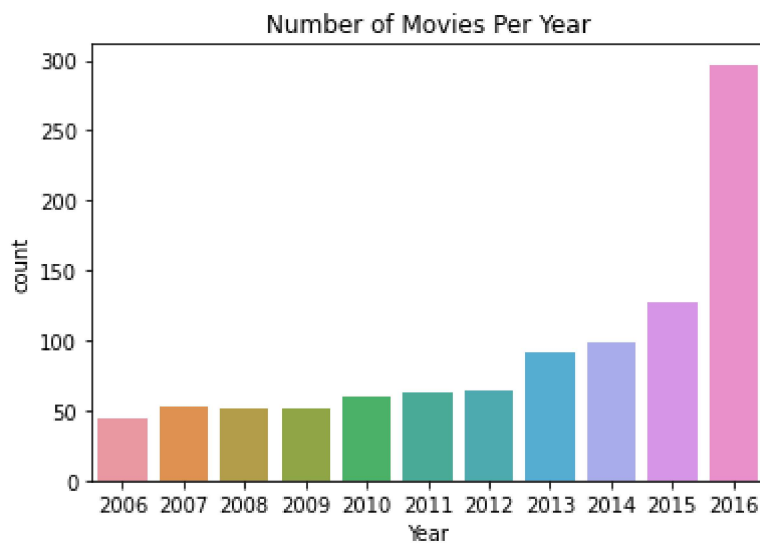
```
data[ 'Year' ].value_counts()
```

Out[71]:

```
2016    297
2015    127
2014     98
2013     91
2012     64
2011     63
2010     60
2007     53
2008     52
2009     51
2006     44
Name: Year, dtype: int64
```

In [74]:

```
sns.countplot(x='Year',data=data)
plt.title("Number of Movies Per Year")
plt.show()
```



In [75]:

```
data.columns
```

Out[75]:

```
Index(['Rank', 'Title', 'Genre', 'Description', 'Director', 'Actors', 'Year',
      'Runtime (Minutes)', 'Rating', 'Votes', 'Revenue (Millions)',
      'Metascore'],
      dtype='object')
```

In [79]:

```
data[data['Revenue (Millions)'].max()==data['Revenue (Millions)']]
```

Out[79]:

	Rank	Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)
50	51	Star Wars: Episode VII - The Force Awakens	Action,Adventure,Fantasy	Three decades after the defeat of the Galactic...	J.J. Abrams	Daisy Ridley, John Boyega, Oscar Isaac, Domhna...	2015	136

In [80]:

```
data.columns
```

Out[80]:

```
Index(['Rank', 'Title', 'Genre', 'Description', 'Director', 'Actors', 'Year',
      'Runtime (Minutes)', 'Rating', 'Votes', 'Revenue (Millions)',
      'Metascore'],
      dtype='object')
```

In [83]:

```
data.groupby('Year')['Rating'].mean().sort_values(ascending=False)
```

Out[83]:

```
Year
2007    7.133962
2006    7.125000
2009    6.960784
2012    6.925000
2011    6.838095
2014    6.837755
2010    6.826667
2013    6.812088
2008    6.784615
2015    6.602362
2016    6.436700
Name: Rating, dtype: float64
```

In [84]:

```
data.columns
```

Out[84]:

```
Index(['Rank', 'Title', 'Genre', 'Description', 'Director', 'Actors', 'Year',
      'Runtime (Minutes)', 'Rating', 'Votes', 'Revenue (Millions)',
      'Metascore'],
      dtype='object')
```

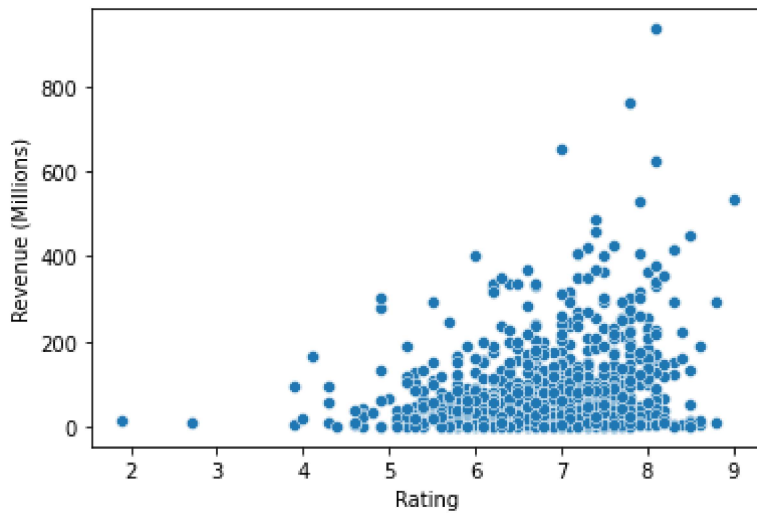


In [86]:

```
sns.scatterplot(x='Rating',y='Revenue (Millions)',data=data)
```

Out[86]:

```
<AxesSubplot:xlabel='Rating', ylabel='Revenue (Millions)'>
```



In [87]:

```
data.columns
```

Out[87]:

```
Index(['Rank', 'Title', 'Genre', 'Description', 'Director', 'Actors', 'Year',
      'Runtime (Minutes)', 'Rating', 'Votes', 'Revenue (Millions)',
      'Metascore'],
      dtype='object')
```

In [88]:

```
data['Genre']
```

Out[88]:

```
0      Action,Adventure,Sci-Fi
1      Adventure,Mystery,Sci-Fi
2      Horror,Thriller
3      Animation,Comedy,Family
4      Action,Adventure,Fantasy
...
995     Crime,Drama,Mystery
996     Horror
997     Drama,Music,Romance
998     Adventure,Comedy
999     Comedy,Family,Fantasy
Name: Genre, Length: 1000, dtype: object
```

In [91]:

```
list1=[]
for value in data['Genre']:
    list1.append(value.split(','))
```

In [92]:

```
list1
```

Out[92]:

```
[['Action', 'Adventure', 'Sci-Fi'],  
 ['Adventure', 'Mystery', 'Sci-Fi'],  
 ['Horror', 'Thriller'],  
 ['Animation', 'Comedy', 'Family'],  
 ['Action', 'Adventure', 'Fantasy'],  
 ['Action', 'Adventure', 'Fantasy'],  
 ['Comedy', 'Drama', 'Music'],  
 ['Comedy'],  
 ['Action', 'Adventure', 'Biography'],  
 ['Adventure', 'Drama', 'Romance'],  
 ['Adventure', 'Family', 'Fantasy'],  
 ['Biography', 'Drama', 'History'],  
 ['Action', 'Adventure', 'Sci-Fi'],  
 ['Animation', 'Adventure', 'Comedy'],  
 ['Action', 'Comedy', 'Drama'],  
 ['Animation', 'Adventure', 'Comedy'],  
 ['Biography', 'Drama', 'History'],  
 ['Action', 'Thriller']]
```

In [ ]:

In [ ]: