First of all, we decided to review the 1000 movies that have become famous in the digital world and here we will examine the priorities that make the movies high.

these are the important variables

missing data detected

For the right movie, we determine how the audience attaches importance to the longest-running movies.

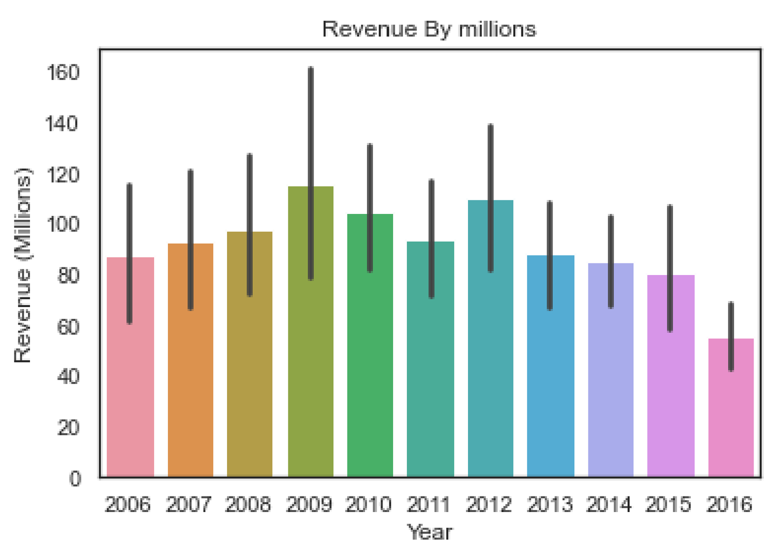
In [46]:
```
data[data['Runtime (Minutes)']>=180]['Title']
```
Out[46]:
```
82     The Wolf of Wall Street
88           The Hateful Eight
311               La vie d'Adèle
Name: Title, dtype: object
```

The participation rate of the audience according to the years will show us the order of importance of the film industry.

In [48]:
```
data.groupby('Year')['Votes'].mean().sort_values(ascending=False)
```
Out[48]:
```
Year
2012    290861.483871
2006    277232.219512
2009    267180.577778
2008    266580.145833
2007    266530.704545
2010    261082.929825
2011    259254.736842
2013    225531.892857
2014    211926.881720
2015    129512.651376
2016     68437.823232
Name: Votes, dtype: float64
```

The income increase over the years shows us how the film industry is in an economic race.

In [56]:
```
sns.barplot(x='Year',y='Revenue (Millions)',data=data)
plt.title('Revenue By millions')
plt.show()
```



The list of directors with the most votes was determined.

In [58]:
```
data.groupby('Director')['Rating'].mean().sort_values(ascending=False)
```
Out[58]:
```
Director
Christopher Nolan                   8.68
Olivier Nakache                     8.60
Makoto Shinkai                      8.60
Aamir Khan                          8.50
Florian Henckel von Donnersmarck    8.50
                                    ...
Sam Taylor-Johnson                  4.10
Joey Curtis                         4.00
George Nolfi                        3.90
James Wong                          2.70
Jason Friedberg                     1.90
Name: Rating, Length: 524, dtype: float64
```

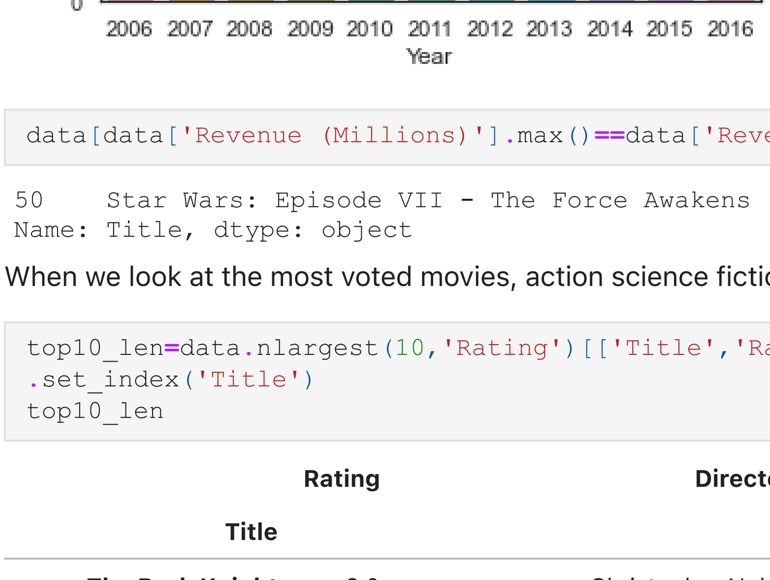
Of course, the running time of the films is one of the most important factors.

In [63]:
```
top10_len=data.nlargest(10,'Runtime (Minutes)')[['Title','Runtime (Minutes)']]
top10_len
```
Out[63]:

	Title	Runtime (Minutes)
88	The Hateful Eight	187
82	The Wolf of Wall Street	180
311	La vie d'Adèle	180
267	Cloud Atlas	172
430	3 Idiots	170
36	Interstellar	169
75	Pirates of the Caribbean: At World's End	169
271	The Hobbit: An Unexpected Journey	169
425	The Curious Case of Benjamin Button	166
126	Transformers: Age of Extinction	165

In [70]:
```
data['Year'].value_counts()
```
Out[70]:
```
2016    198
2015    109
2014     93
2013     84
2012     62
2011     57
2010     57
2008     48
2009     45
2007     44
2006     41
Name: Year, dtype: int64
```

In [72]:
```
sns.countplot(x='Year' ,data=data)
plt.title('Number of Movies Per Year')
plt.show()
```



In [77]:
```
data[data['Revenue (Millions)'].max()==data['Revenue (Millions)']]['Title']
```
Out[77]:
```
50    Star Wars: Episode VII - The Force Awakens
Name: Title, dtype: object
```

When we look at the most voted movies, action science fiction movies are at the forefront.

In [86]:
```
top10_len=data.nlargest(10,'Rating')[['Title','Rating','Director']]\
.set_index('Title')
top10_len
```
Out[86]:

	Rating	Director
Title		
The Dark Knight	9.0	Christopher Nolan
Inception	8.8	Christopher Nolan
Interstellar	8.6	Christopher Nolan
Kimi no na wa	8.6	Makoto Shinkai
The Intouchables	8.6	Olivier Nakache
The Prestige	8.5	Christopher Nolan
The Departed	8.5	Martin Scorsese
The Dark Knight Rises	8.5	Christopher Nolan
Whiplash	8.5	Damien Chazelle
The Lives of Others	8.5	Florian Henckel von Donnersmarck

Christopher Nolan movies grab our attention

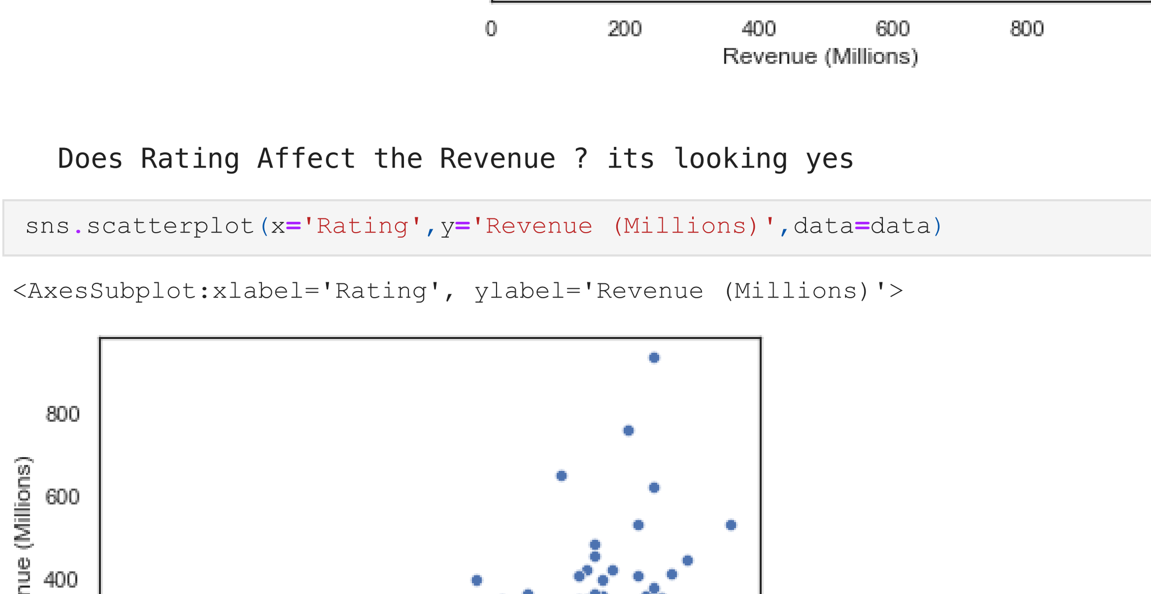
In [89]:
```
sns.barplot(x='Rating',y=top10_len.index,data=top10_len,hue="Director",dodge=False)
plt.legend(bbox_to_anchor=(1.05,1),loc=2)
```
Out[89]:
```
<matplotlib.legend.Legend at 0x7fc90ee3d910>
```

In [91]:
```
data.nlargest(10,'Revenue (Millions)')['Title']
```
Out[91]:
```
50     Star Wars: Episode VII - The Force Awakens
87                                         Avatar
85                                 Jurassic World
76                                   The Avengers
54                                The Dark Knight
12                                      Rogue One
119                                  Finding Dory
94                        Avengers: Age of Ultron
124                         The Dark Knight Rises
578               The Hunger Games: Catching Fire
Name: Title, dtype: object
```

In [93]:
```
top_10=data.nlargest(10,'Revenue (Millions)')[['Title','Revenue (Millions)']].\
set_index('Title')
top_10
```
Out[93]:

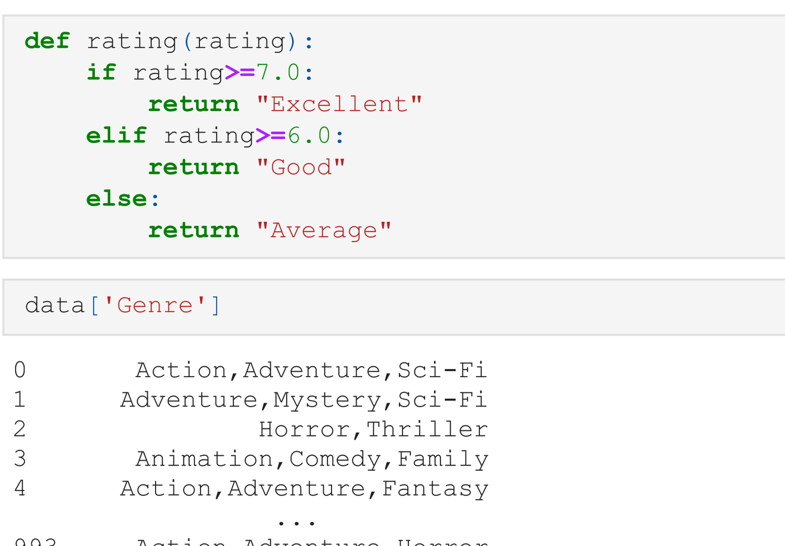
	Revenue (Millions)
Title	
Star Wars: Episode VII - The Force Awakens	936.63
Avatar	760.51
Jurassic World	652.18
The Avengers	623.28
The Dark Knight	533.32
Rogue One	532.17
Finding Dory	486.29
Avengers: Age of Ultron	458.99
The Dark Knight Rises	448.13
The Hunger Games: Catching Fire	424.65

In [95]:
```
sns.barplot(x='Revenue (Millions)',y=top_10.index,data=top_10)
plt.title('Top 10 Highest Revenue Movie Titles')
plt.show()
```




```
Does Rating Affect the Revenue ? its looking yes
```

In [98]:
```
sns.scatterplot(x='Rating',y='Revenue (Millions)',data=data)
```
Out[98]:
```
<AxesSubplot:xlabel='Rating', ylabel='Revenue (Millions)'>
```



In [99]:
```
def rating(rating):
    if rating>=7.0:
        return "Excellent"
    elif rating>=6.0:
        return "Good"
    else:
        return "Average"
```

In [106…
```
data['Genre']
```
Out[106…
```
0        Action,Adventure,Sci-Fi
1       Adventure,Mystery,Sci-Fi
2                Horror,Thriller
3        Animation,Comedy,Family
4       Action,Adventure,Fantasy
                 ...
993      Action,Adventure,Horror
994                       Comedy
996                       Horror
997          Drama,Music,Romance
999        Comedy,Family,Fantasy
Name: Genre, Length: 838, dtype: object
```

In [117…
```
from collections import Counter
Counter(one_d)
```
Out[117…
```
Counter({'Action': 277,
         'Adventure': 244,
         'Sci-Fi': 107,
         'Mystery': 86,
         'Horror': 87,
         'Thriller': 148,
         'Animation': 45,
         'Comedy': 250,
         'Family': 48,
         'Fantasy': 92,
         'Drama': 419,
         'Music': 15,
         'Biography': 67,
         'Romance': 120,
         'History': 25,
         'Western': 4,
         'Crime': 126,
         'War': 10,
         'Musical': 5,
         'Sport': 15})
```

If we consider the data, it would be right for us to invest in genre science fiction movies, when the runtime is 180.

In [ ]: