

minor-project

August 4, 2024

```
[6]: import numpy as np
import pandas as pd
import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
```

```
[7]: import warnings
warnings.filterwarnings('ignore')
```

```
[8]: data = pd.read_csv('IMDB-Movie-Data.csv')
```

1 1. Display Top 10 Rows of The Dataset

```
[9]: data.head(10)
```

```
[9]:
```

	Rank	Title	Genre \
0	1	Guardians of the Galaxy	Action,Adventure,Sci-Fi
1	2	Prometheus	Adventure,Mystery,Sci-Fi
2	3	Split	Horror,Thriller
3	4	Sing	Animation,Comedy,Family
4	5	Suicide Squad	Action,Adventure,Fantasy
5	6	The Great Wall	Action,Adventure,Fantasy
6	7	La La Land	Comedy,Drama,Music
7	8	Mindhorn	Comedy
8	9	The Lost City of Z	Action,Adventure,Biography
9	10	Passengers	Adventure,Drama,Romance

	Description	Director \
0	A group of intergalactic criminals are forced ...	James Gunn
1	Following clues to the origin of mankind, a te...	Ridley Scott
2	Three girls are kidnapped by a man with a diag...	M. Night Shyamalan
3	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet
4	A secret government agency recruits some of th...	David Ayer
5	European mercenaries searching for black powde...	Yimou Zhang
6	A jazz pianist falls for an aspiring actress i...	Damien Chazelle

7	A has-been actor best known for playing the ti...	Sean Foley
8	A true-life drama, centering on British explor...	James Gray
9	A spacecraft traveling to a distant colony pla...	Morten Tyldum

	Actors	Year	Runtime (Minutes)	\
0	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	2014	121	
1	Noomi Rapace, Logan Marshall-Green, Michael Fa...	2012	124	
2	James McAvoy, Anya Taylor-Joy, Haley Lu Richar...	2016	117	
3	Matthew McConaughey, Reese Witherspoon, Seth Ma...	2016	108	
4	Will Smith, Jared Leto, Margot Robbie, Viola D...	2016	123	
5	Matt Damon, Tian Jing, Willem Dafoe, Andy Lau	2016	103	
6	Ryan Gosling, Emma Stone, Rosemarie DeWitt, J...	2016	128	
7	Essie Davis, Andrea Riseborough, Julian Barrat...	2016	89	
8	Charlie Hunnam, Robert Pattinson, Sienna Mille...	2016	141	
9	Jennifer Lawrence, Chris Pratt, Michael Sheen,...	2016	116	

	Rating	Votes	Revenue (Millions)	Metascore
0	8.1	757074	333.13	76.0
1	7.0	485820	126.46	65.0
2	7.3	157606	138.12	62.0
3	7.2	60545	270.32	59.0
4	6.2	393727	325.02	40.0
5	6.1	56036	45.13	42.0
6	8.3	258682	151.06	93.0
7	6.4	2490	NaN	71.0
8	7.1	7188	8.01	78.0
9	7.0	192177	100.01	41.0

2. Check Last 10 Rows of The Dataset

```
[10]: data.tail(10)
```

[10]:	Rank	Title	Genre	\
990	991	Underworld: Rise of the Lycans	Action,Adventure,Fantasy	
991	992	Taare Zameen Par	Drama,Family,Music	
992	993	Take Me Home Tonight	Comedy,Drama,Romance	
993	994	Resident Evil: Afterlife	Action,Adventure,Horror	
994	995	Project X	Comedy	
995	996	Secret in Their Eyes	Crime,Drama,Mystery	
996	997	Hostel: Part II	Horror	
997	998	Step Up 2: The Streets	Drama,Music,Romance	
998	999	Search Party	Adventure,Comedy	
999	1000	Nine Lives	Comedy,Family,Fantasy	

	Description	Director	\
990	An origins story centered on the centuries-old...	Patrick Tatopoulos	

991	An eight-year-old boy is thought to be a lazy ...	Aamir Khan
992	Four years after graduation, an awkward high s...	Michael Dowse
993	While still out to destroy the evil Umbrella C...	Paul W.S. Anderson
994	3 high school seniors throw a birthday party t...	Nima Nourizadeh
995	A tight-knit team of rising investigators, alo...	Billy Ray
996	Three American college students studying abroa...	Eli Roth
997	Romantic sparks occur between two dance studen...	Jon M. Chu
998	A pair of friends embark on a mission to reuni...	Scot Armstrong
999	A stuffy businessman finds himself trapped ins...	Barry Sonnenfeld

	Actors	Year	\
990	Rhona Mitra, Michael Sheen, Bill Nighy, Steven...	2009	
991	Darsheel Safary, Aamir Khan, Tanay Chheda, Sac...	2007	
992	Topher Grace, Anna Faris, Dan Fogler, Teresa P...	2011	
993	Milla Jovovich, Ali Larter, Wentworth Miller,K...	2010	
994	Thomas Mann, Oliver Cooper, Jonathan Daniel Br...	2012	
995	Chiwetel Ejiofor, Nicole Kidman, Julia Roberts...	2015	
996	Lauren German, Heather Matarazzo, Bijou Philli...	2007	
997	Robert Hoffman, Briana Evigan, Cassie Ventura,...	2008	
998	Adam Pally, T.J. Miller, Thomas Middleditch,Sh...	2014	
999	Kevin Spacey, Jennifer Garner, Robbie Amell,Ch...	2016	

	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
990	92	6.6	129708	45.80	44.0
991	165	8.5	102697	1.20	42.0
992	97	6.3	45419	6.92	NaN
993	97	5.9	140900	60.13	37.0
994	88	6.7	164088	54.72	48.0
995	111	6.2	27585	NaN	45.0
996	94	5.5	73152	17.54	46.0
997	98	6.2	70699	58.01	50.0
998	93	5.6	4881	NaN	22.0
999	87	5.3	12435	19.64	11.0

3 3. Find Shape of Our Dataset (Number of Rows And Number of Columns)

```
[11]: data.shape
```

```
[11]: (1000, 12)
```

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

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

4 4. Getting Information About Our Dataset Like Total Number Rows, Total Number of Columns, Datatypes of Each Column And Memory Requirement

```
[13]: 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
```

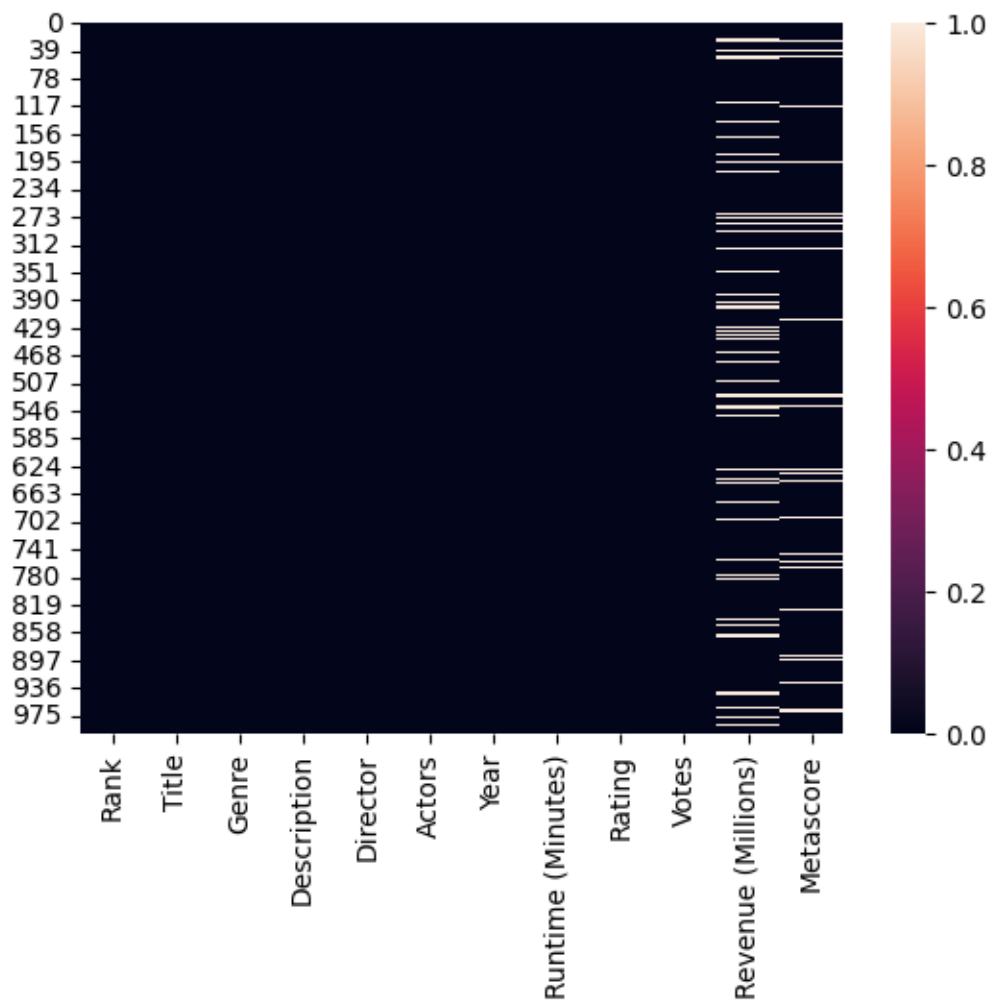
5 5. Check Null Values In The Dataset

```
[14]: data.isnull().sum()
```

```
[14]: 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
```

```
[15]: import matplotlib.pyplot as plt
      import seaborn as sns
```

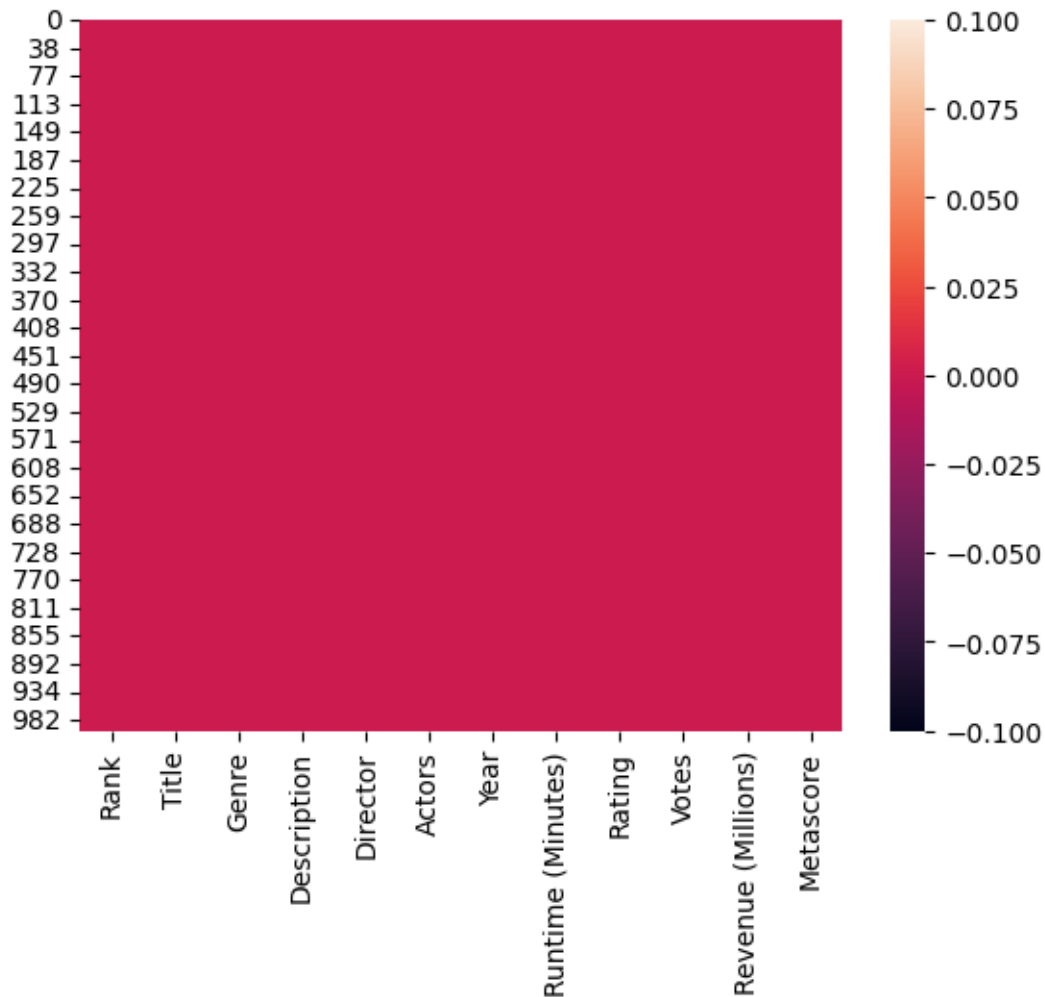
```
sns.heatmap(data.isnull())
plt.show()
```



6 6. Drop All The Missing Values

```
[16]: data = data.dropna(axis=0)
```

```
[17]: sns.heatmap(data.isnull())
plt.show()
```



7 7. Check For Duplicate Data

```
[18]: dup_data=data.duplicated().any()
      print("Are there any duplicated values in data?",dup_data)
```

Are there any duplicated values in data? False

8 8. Get Overall Statistics About The DataFrame

```
[19]: data.describe()
```

```
[19]:
```

	Rank	Year	Runtime (Minutes)	Rating	Votes \
count	838.000000	838.00000	838.000000	838.000000	8.380000e+02
mean	485.247017	2012.50716	114.638425	6.814320	1.932303e+05

std	286.572065	3.17236	18.470922	0.877754	1.930990e+05
min	1.000000	2006.00000	66.000000	1.900000	1.780000e+02
25%	238.250000	2010.00000	101.000000	6.300000	6.127650e+04
50%	475.500000	2013.00000	112.000000	6.900000	1.368795e+05
75%	729.750000	2015.00000	124.000000	7.500000	2.710830e+05
max	1000.000000	2016.00000	187.000000	9.000000	1.791916e+06

	Revenue (Millions)	Metascore
count	838.000000	838.000000
mean	84.564558	59.575179
std	104.520227	16.952416
min	0.000000	11.000000
25%	13.967500	47.000000
50%	48.150000	60.000000
75%	116.800000	72.000000
max	936.630000	100.000000

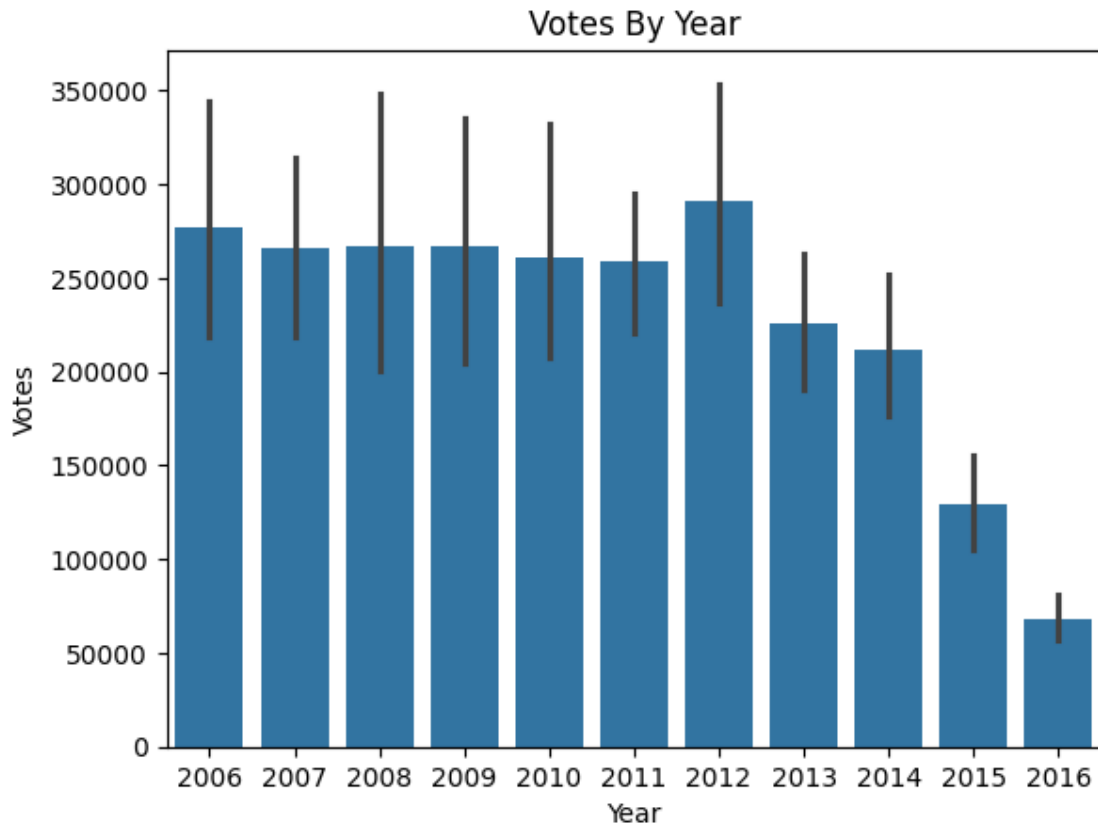
9 9. Display Title of The Movie Having Runtime \geq 180 Minutes

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

```
[20]: 82      The Wolf of Wall Street
      88      The Hateful Eight
      311     La vie d'Adèle
      Name: Title, dtype: object
```

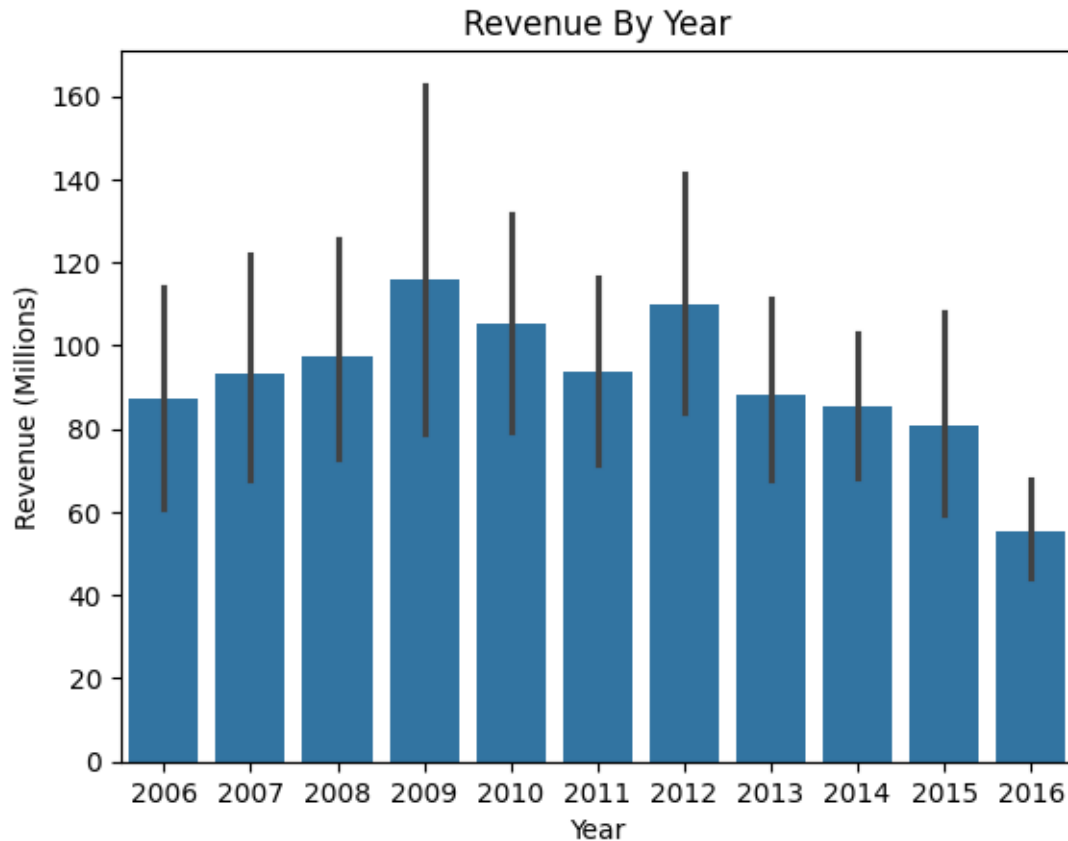
10 10. In Which Year There Was The Highest Voting?

```
[21]: sns.barplot(x='Year',y='Votes',data=data)
      plt.title("Votes By Year")
      plt.show()
```



11. In Which Year There Was The Highest Revenue?

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

12. Find The Average Rating For Each Director

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

```
[23]: Director
Christopher Nolan      8.68
Olivier Nakache        8.60
Makoto Shinkai         8.60
Florian Henckel von Donnersmarck  8.50
Aamir Khan             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
```

13. Display Top 10 Lengthy Movies Title

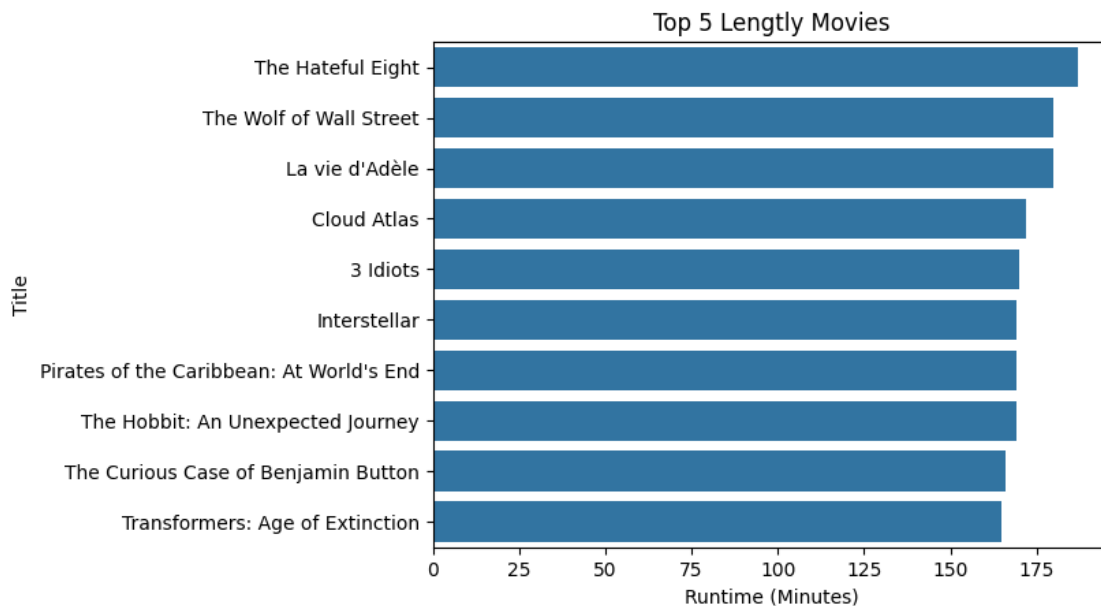
```
[24]: top10_len=data.nlargest(10,'Runtime (Minutes)')[['Title','Runtime (Minutes)']].\
      ↪\
      set_index('Title')
      top10_len
```

```
[24]:
```

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

```
[25]: sns.barplot(x='Runtime (Minutes)',y=top10_len.index,data=top10_len)
      plt.title('Top 5 Lengtly Movies')
```

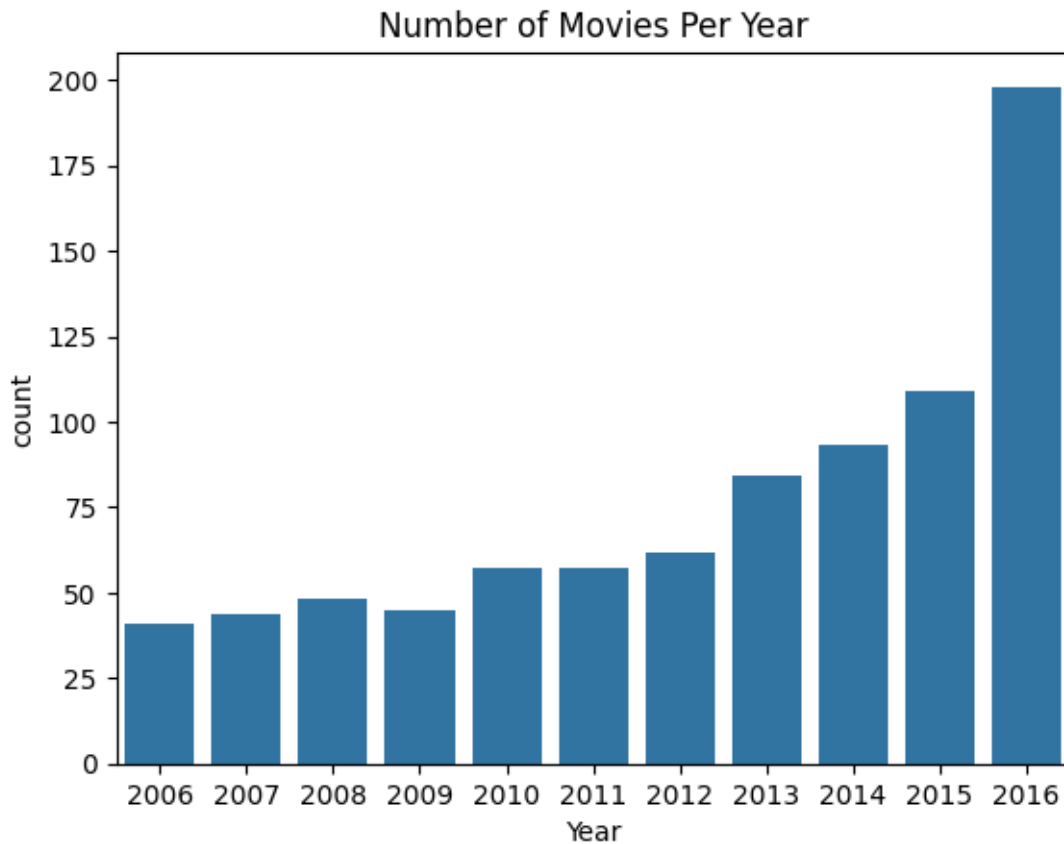
```
[25]: Text(0.5, 1.0, 'Top 5 Lengtly Movies')
```



13 14. Display Number of Movies Per Year

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

```
[26]: Text(0.5, 1.0, 'Number of Movies Per Year')
```



14 15. Find Most Popular Movie Title (Higest Revenue)

```
[27]: data.columns
```

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

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

```
[28]: 50    Star Wars: Episode VII - The Force Awakens
      Name: Title, dtype: object
```

15 16. Display Top 10 Highest Rated Movie Titles And its Directors

```
[29]: top_10=data.nlargest(10,'Rating')[['Title','Rating','Director']].
      ↪set_index('Title')
```

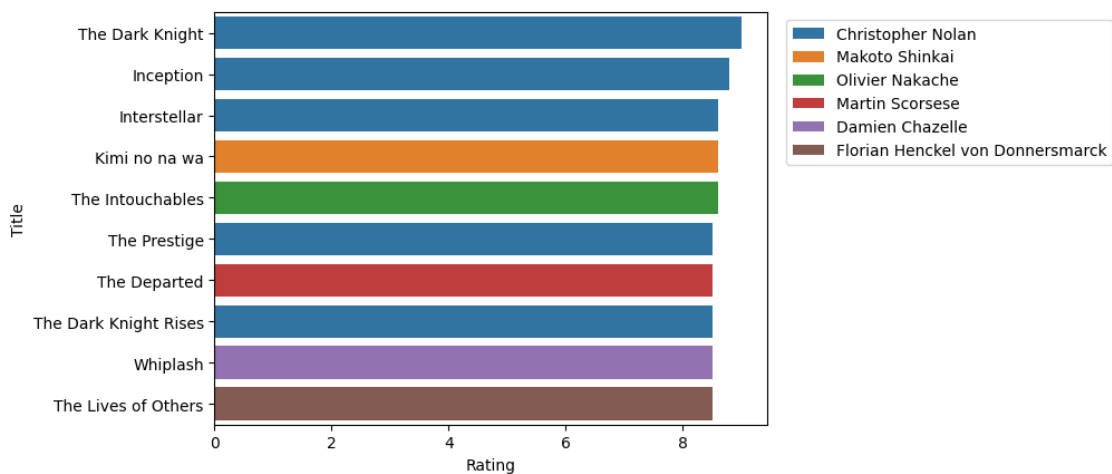
```
[30]: top_10
```

```
[30]:
```

Title	Rating	Director
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

```
[31]: sns.barplot(x='Rating',y=top_10.index,data=top_10,hue='Director',dodge=False)
      plt.legend(bbox_to_anchor=(1.02,1),loc=2)
```

```
[31]: <matplotlib.legend.Legend at 0x7c00f6fcfbe0>
```



16 17. Display Top 10 Highest Revenue Movie Titles

```
[32]: data.columns
```

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

```
[33]: data.sort_values(by='Revenue (Millions)',ascending=False).head(10)
```

```
[33]:
```

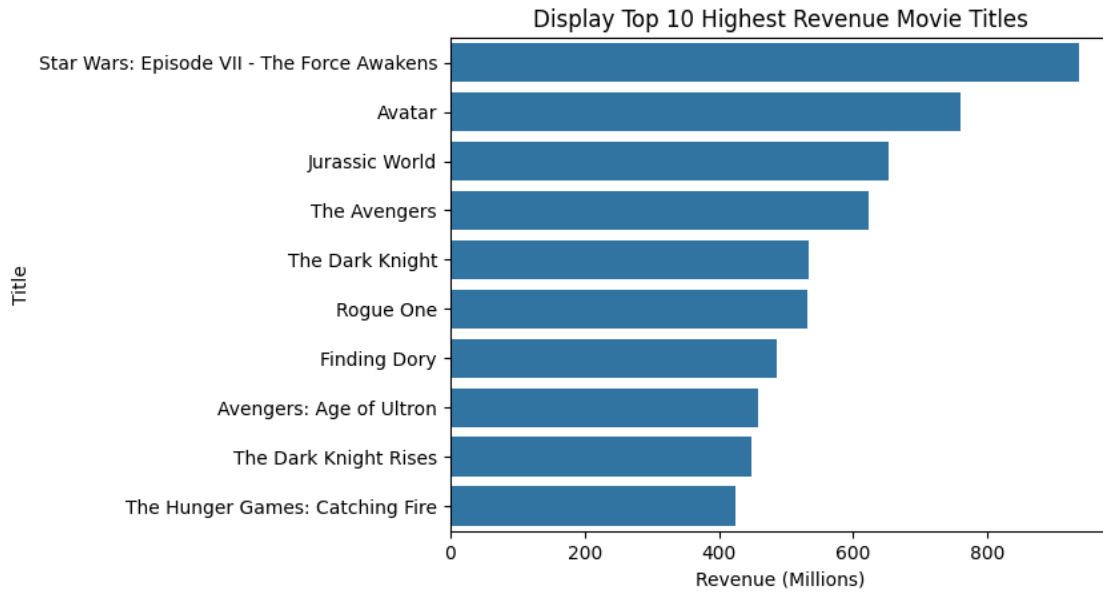
	Rank	Title \	Genre \	Description	Director \
50	51	Star Wars: Episode VII - The Force Awakens	Action,Adventure,Fantasy	Three decades after the defeat of the Galactic...	J.J. Abrams
87	88	Avatar	Action,Adventure,Fantasy	A paraplegic marine dispatched to the moon Pan...	James Cameron
85	86	Jurassic World	Action,Adventure,Sci-Fi	A new theme park, built on the original site o...	Colin Trevorrow
76	77	The Avengers	Action,Sci-Fi	Earth's mightiest heroes must come together an...	Joss Whedon
54	55	The Dark Knight	Action,Crime,Drama	When the menace known as the Joker wreaks havo...	Christopher Nolan
12	13	Rogue One	Action,Adventure,Sci-Fi	The Rebel Alliance makes a risky move to steal...	Gareth Edwards
119	120	Finding Dory	Animation,Adventure,Comedy	The friendly but forgetful blue tang fish, Dor...	Andrew Stanton
94	95	Avengers: Age of Ultron	Action,Adventure,Sci-Fi	When Tony Stark and Bruce Banner try to jump-s...	Joss Whedon
124	125	The Dark Knight Rises	Action,Thriller	Eight years after the Joker's reign of anarchy...	Christopher Nolan
578	579	The Hunger Games: Catching Fire	Action,Adventure,Mystery	Katniss Everdeen and Peeta Mellark become targ...	Francis Lawrence

	Actors	Year	\
50	Daisy Ridley, John Boyega, Oscar Isaac, Domhnall	2015	
87	Sam Worthington, Zoe Saldana, Sigourney Weaver...	2009	
85	Chris Pratt, Bryce Dallas Howard, Ty Simpkins,...	2015	
76	Robert Downey Jr., Chris Evans, Scarlett Johan...	2012	
54	Christian Bale, Heath Ledger, Aaron Eckhart, Mi...	2008	
12	Felicity Jones, Diego Luna, Alan Tudyk, Donnie...	2016	
119	Ellen DeGeneres, Albert Brooks, Ed O'Neill, Kai...	2016	
94	Robert Downey Jr., Chris Evans, Mark Ruffalo, ...	2015	
124	Christian Bale, Tom Hardy, Anne Hathaway, Gary ...	2012	
578	Jennifer Lawrence, Josh Hutcherson, Liam Hemsw...	2013	

	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
50	136	8.1	661608	936.63	81.0
87	162	7.8	935408	760.51	83.0
85	124	7.0	455169	652.18	59.0
76	143	8.1	1045588	623.28	69.0
54	152	9.0	1791916	533.32	82.0
12	133	7.9	323118	532.17	65.0
119	97	7.4	157026	486.29	77.0
94	141	7.4	516895	458.99	66.0
124	164	8.5	1222645	448.13	78.0
578	146	7.6	525646	424.65	76.0

```
[34]: top_10 = data.nlargest(10, 'Revenue (Millions)')[['Title', 'Director', 'Revenue (Millions)']].set_index('Title')
```

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



17 18. Find Average Rating of Movies Year-wise

```
[36]: data.columns
```

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

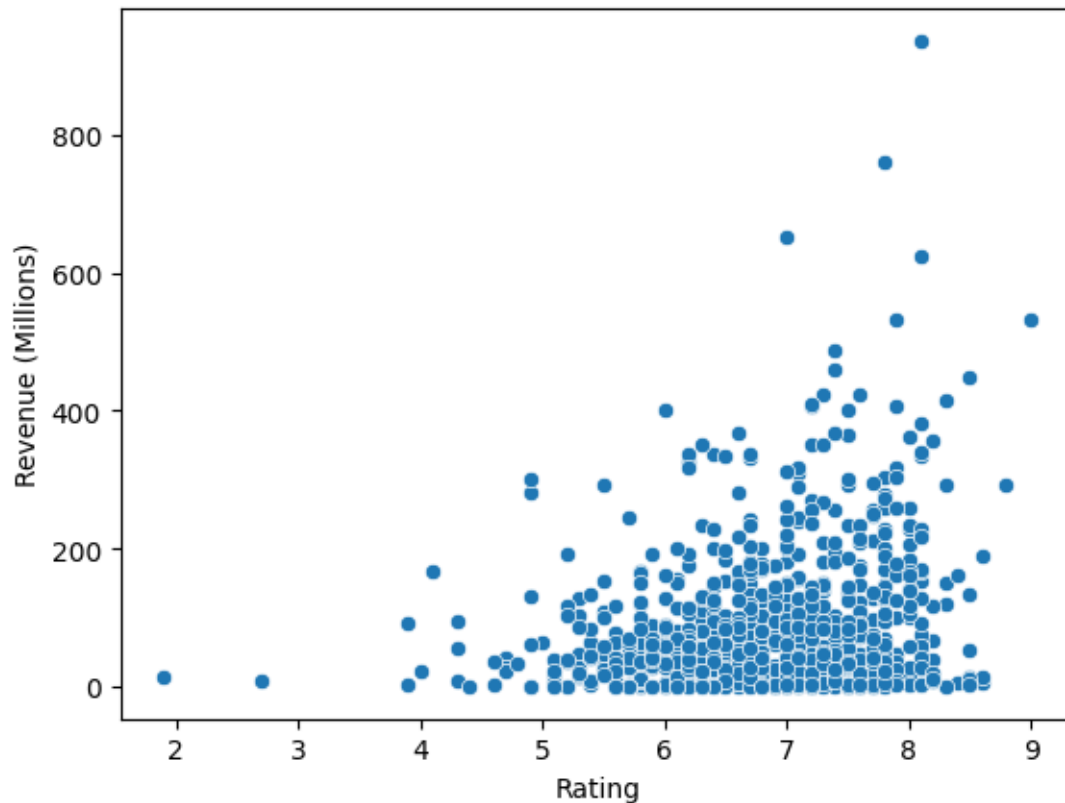
```
[37]: data.groupby('Year')['Rating'].mean().sort_values(ascending=False)
```

```
[37]: Year
2006    7.143902
2007    7.140909
2011    6.945614
2012    6.933871
2009    6.911111
2010    6.894737
2013    6.832143
2014    6.822581
2008    6.708333
2015    6.674312
2016    6.644444
Name: Rating, dtype: float64
```

18 19. Does Rating Affect The Revenue?

```
[38]: sns.scatterplot(x='Rating',y='Revenue (Millions)',data=data)
```

```
[38]: <Axes: xlabel='Rating', ylabel='Revenue (Millions)'>
```



19 Answer : Yes

20 20. Classify Movies Based on Ratings [Good,Better and Best]

```
[39]: data.columns
```

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

```
[40]: def rating(rating):  
       if rating>=7.0:  
           return 'Excellent'
```



```

elif rating>=6.0:
    return 'Good'
else:
    return 'Average'

```

```
[41]: data['rating_cat']=data['Rating'].apply(rating)
```

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

```
[42]:
```

	Rank	Title	Genre \
0	1	Guardians of the Galaxy	Action,Adventure,Sci-Fi
1	2	Prometheus	Adventure,Mystery,Sci-Fi
2	3	Split	Horror,Thriller
3	4	Sing	Animation,Comedy,Family
4	5	Suicide Squad	Action,Adventure,Fantasy

	Description	Director \
0	A group of intergalactic criminals are forced ...	James Gunn
1	Following clues to the origin of mankind, a te...	Ridley Scott
2	Three girls are kidnapped by a man with a diag...	M. Night Shyamalan
3	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet
4	A secret government agency recruits some of th...	David Ayer

	Actors	Year	Runtime (Minutes) \
0	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	2014	121
1	Noomi Rapace, Logan Marshall-Green, Michael Fa...	2012	124
2	James McAvoy, Anya Taylor-Joy, Haley Lu Richar...	2016	117
3	Matthew McConaughey,Reese Witherspoon, Seth Ma...	2016	108
4	Will Smith, Jared Leto, Margot Robbie, Viola D...	2016	123

	Rating	Votes	Revenue (Millions)	Metascore	rating_cat
0	8.1	757074	333.13	76.0	Excellent
1	7.0	485820	126.46	65.0	Excellent
2	7.3	157606	138.12	62.0	Excellent
3	7.2	60545	270.32	59.0	Excellent
4	6.2	393727	325.02	40.0	Good

21 21. Count Number of Action Movies

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

```
[44]: data['temp']=list1
```

```
[45]: genre=input("Enter Genre you want to count : ").title()
      count=0
      for value in data['temp']:
          if genre in value:
              count=count+1
      print("Total Count is",count)
```

Enter Genre you want to count : 15
Total Count is 0

22 OR

```
[46]: len(data[data['Genre'].str.contains('action',case=False)])
```

[46]: 277

[46]: