

```
# This Python 3 environment comes with many helpful analytics libraries installed
# It is defined by the kaggle/python Docker image: https://github.com/kaggle/docker-python
# For example, here's several helpful packages to load

import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

# Input data files are available in the read-only "../input/" directory
# For example, running this (by clicking run or pressing Shift+Enter) will list all files under the input c

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))

# You can write up to 20GB to the current directory (/kaggle/working/) that gets preserved as output when y
# You can also write temporary files to /kaggle/temp/, but they won't be saved outside of the current sessi
```

✓ 1. Display Top 10 Rows of The Dataset

```
data.head(10)
```

Rank		Title	Genre	Description	Director	Acto
0	1	Guardians of the Galaxy	Action,Adventure,Sci-Fi	A group of intergalactic criminals are forced ...	James Gunn	Chris Pratt, ' Diesel, Brad Cooper, Zoe S
	2	Prometheus	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott	Noomi Rapha Logan Marsh Green, Michael F
	3	Split	Horror,Thriller	Three girls are kidnapped by a man with a diag...	M. Night Shyamalan	James McAvoy, Ar Taylor-Joy, Haley Riche
	4	Sing	Animation,Comedy,Family	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet	Matth McConaughey,Ree Witherspoon, S M
	5	Suicide Squad	Action,Adventure,Fantasy	A secret government agency recruits some of th...	David Ayer	Will Smith, Jar Leto, Margot Robt Viola I
	6	The Great Wall	Action,Adventure,Fantasy	European mercenaries searching for black powde...	Yimou Zhang	Matt Damon, T Jing, Willem Daf Andy L
	7	La La Land	Comedy,Drama,Music	A jazz pianist falls for an aspiring actress i...	Damien Chazelle	Ryan Gosling, Em Stone, Rosema DeWitt, u
	8	Mindhorn	Comedy	A has-been actor best known for playing the ti...	Sean Foley	Essie Davis, Andi Riseborough, Jul Barra
	9	The Lost City of Z	Action,Adventure,Biography	A true-life drama, centering on British explor...	James Gray	Charlie Hunna Robert Pattins Sienna Mill
	10	Passengers	Adventure,Drama,Romance	A spacecraft traveling to a distant colony pla...	Morten Tyldum	Jennifer Lawren Chris Pratt, Mich Sheer

Next steps:

[Generate code with data](#)



[View recommended plots](#)

2. Check Last 10 Rows of The Dataset

```
data.tail(10)
```

Rank	Title		Genre	Description	Director	Actors
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...
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...
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...
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...
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...
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,...
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...
999	1000	Nine Lives	Comedy,Family,Fantasy	A stuffy businessman finds himself trapped ins...	Barry Sonnenfeld	Kevin Spacey, Jennifer Garner, Robbie Amell,Ch...

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

```
data.shape
```

```
(1000, 12)
```

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

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

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

```
data.info()
```

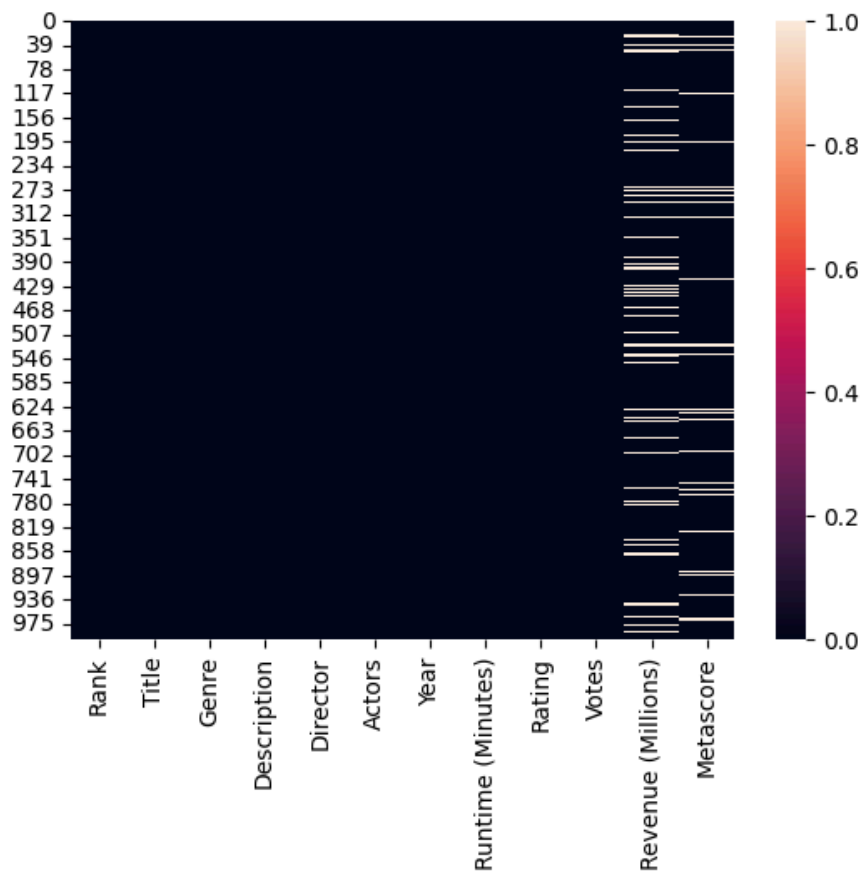
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 12 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Rank                  1000 non-null   int64
1   Title                 1000 non-null   object
2   Genre                 1000 non-null   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. Check Null Values In The Dataset

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

```
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
```

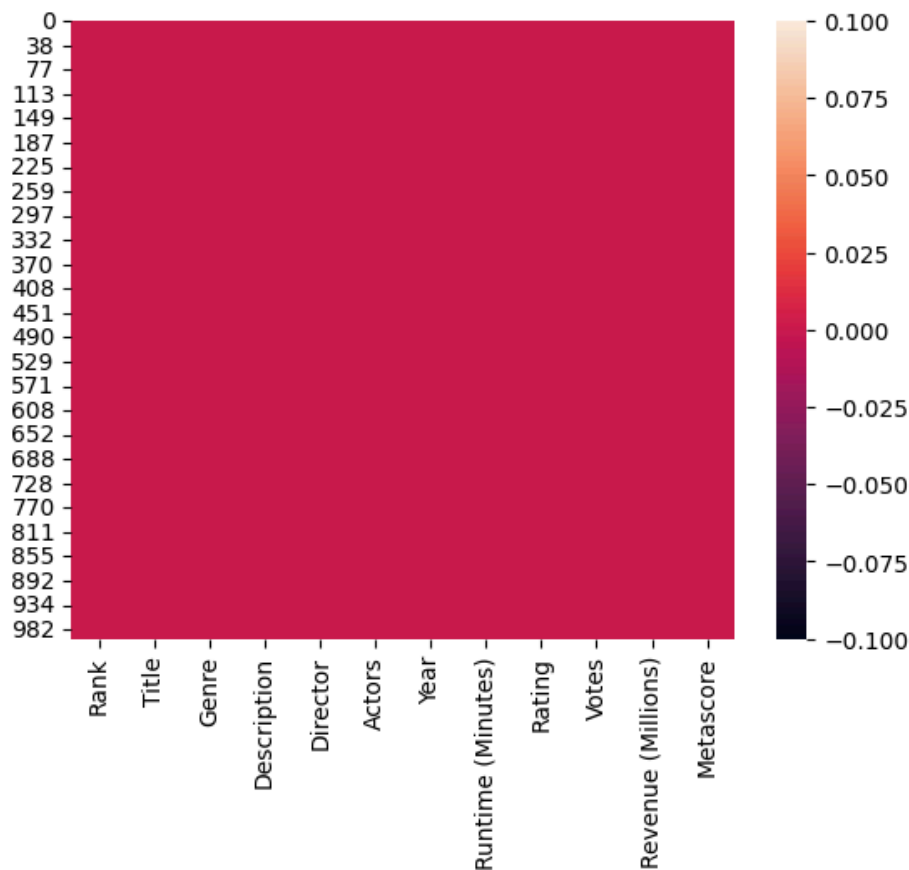
```
import matplotlib.pyplot as plt
import seaborn as sns
sns.heatmap(data.isnull())
plt.show()
```



✓ 6. Drop All The Missing Values

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

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



✓ 7. Check For Duplicate Data

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

Are there any duplicated values in data? False

✓ 8. Get Overall Statistics About The DataFrame

```
data.describe()
```

	Rank	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Meta
count	838.000000	838.00000	838.000000	838.000000	8.380000e+02	838.000000	838.0
mean	485.247017	2012.50716	114.638425	6.814320	1.932303e+05	84.564558	59.5
std	286.572065	3.17236	18.470922	0.877754	1.930990e+05	104.520227	16.9
min	1.000000	2006.00000	66.000000	1.900000	1.780000e+02	0.000000	11.0
25%	238.250000	2010.00000	101.000000	6.300000	6.127650e+04	13.967500	47.0
50%	475.500000	2013.00000	112.000000	6.900000	1.368795e+05	48.150000	60.0
75%	729.750000	2015.00000	124.000000	7.500000	2.710830e+05	116.800000	72.0
max	1000.000000	2016.00000	187.000000	9.000000	1.701016e+06	936.630000	100.0

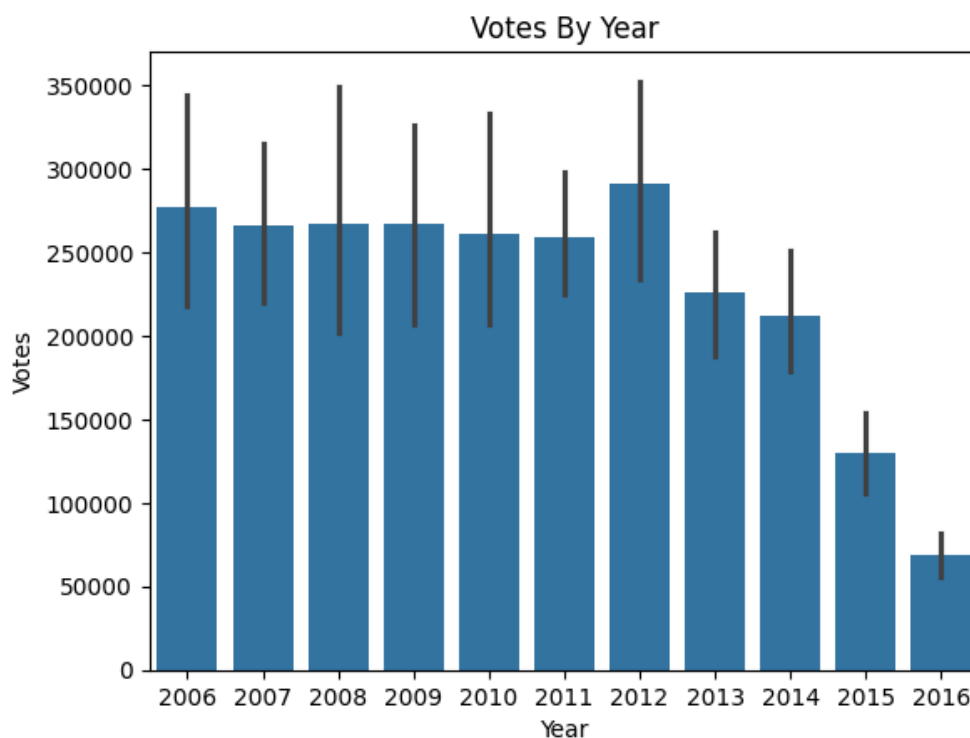
✓ 9. Display Title of The Movie Having Runtime ≥ 180 Minutes

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

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

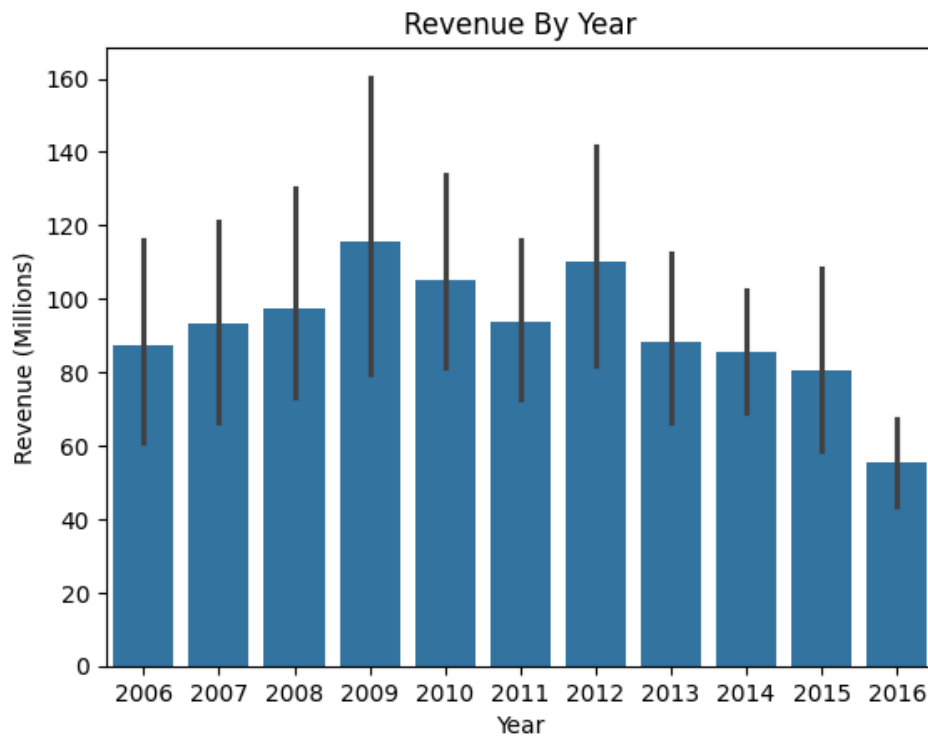
✓ 10. In Which Year There Was The Highest Voting?

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



✓ 11. In Which Year There Was The Highest Revenue?

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



✓ 12. Find The Average Rating For Each Director

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

```
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
```

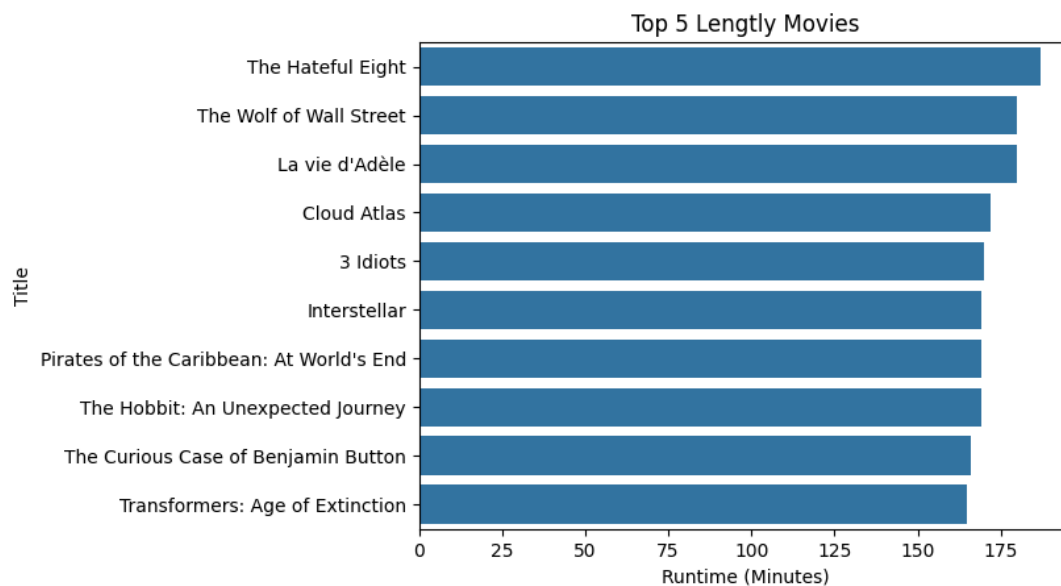
✓ 13. Display Top 10 Lengthy Movies Title

```
data.columns
```

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

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

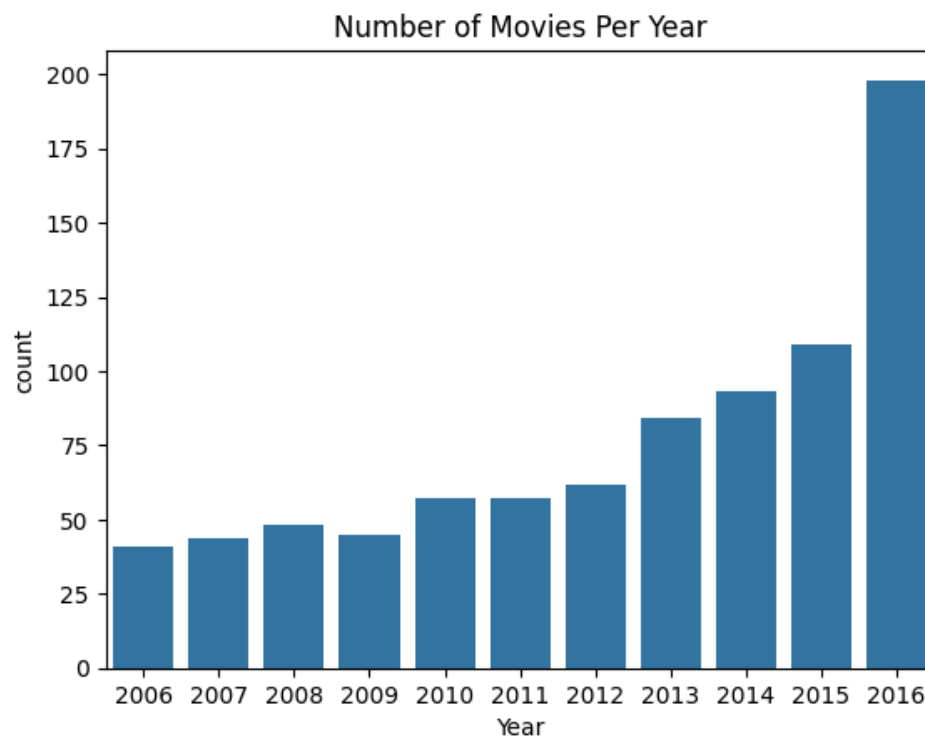
```
sns.barplot(x='Runtime (Minutes)',y=top10_len.index,data=top10_len)
plt.title('Top 5 Lengthy Movies')
plt.show()
```

✓ 14. Display Number of Movies Per Year

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

```
Text(0.5, 1.0, 'Number of Movies Per Year')
```



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

```
data.columns
```

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

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

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




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

```
data.columns
```

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

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

```
top_10
```

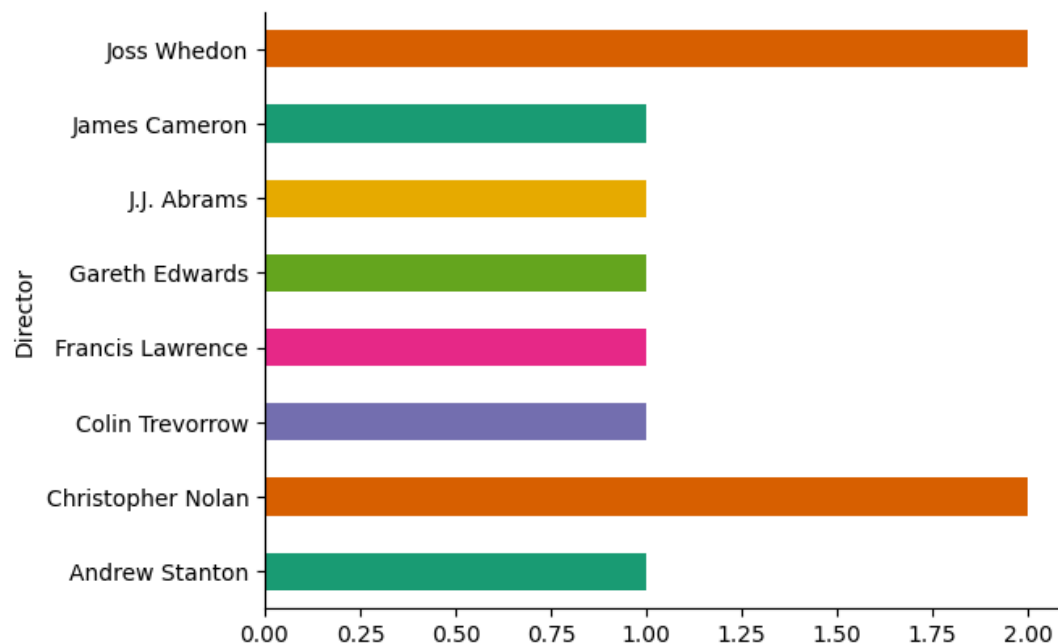
Rating		Director	  
Title			
The Dark Knight	9.0	Christopher Nolan	
Inception	8.8	Christopher Nolan	
Interstellar	8.6	Christopher Nolan	
Kimi no na wa	8.6	Makoto Shinkai	
The Intouchables	8.6	Olivier Nakache	
The Prestige	8.5	Christopher Nolan	
The Departed	8.5	Martin Scorsese	
The Dark Knight Rises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

Next steps:

Generate code with top_10

 View recommended plots

```
from matplotlib import pyplot as plt
import seaborn as sns
top_10.groupby('Director').size().plot(kind='barh', color=sns.palettes.mpl_palette('Dark2'))
plt.gca().spines[['top', 'right']].set_visible(False)
```



✓ 17. Display Top 10 Highest Revenue Movie Titles

```
data.columns
```

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

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

Rank		Title	Genre	Description	Director	Actors
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...
87	88	Avatar	Action,Adventure,Fantasy	A paraplegic marine dispatched to the moon Pan...	James Cameron	Sam Worthington, Zoe Saldana, Sigourney Weaver...
85	86	Jurassic World	Action,Adventure,Sci-Fi	A new theme park, built on the original site o...	Colin Trevorrow	Chris Pratt, Bryce Dallas Howard, Ty Simpkins,...
76	77	The Avengers	Action,Sci-Fi	Earth's mightiest heroes must come together an...	Joss Whedon	Robert Downey Jr., Chris Evans, Scarlett Johan...
54	55	The Dark Knight	Action,Crime,Drama	When the menace known as the Joker wreaks havo...	Christopher Nolan	Christian Bale, Heath Ledger, Aaron Eckhart,Mi...
12	13	Rogue One	Action,Adventure,Sci-Fi	The Rebel Alliance makes a risky move to steal...	Gareth Edwards	Felicity Jones, Diego Luna, Alan Tudyk, Donnie...
119	120	Finding Dory	Animation,Adventure,Comedy	The friendly but forgetful blue tang fish, Dor...	Andrew Stanton	Ellen DeGeneres, Albert Brooks,Ed O'Neill, Kai...
94	95	Avengers: Age of Ultron	Action,Adventure,Sci-Fi	When Tony Stark and Bruce Banner try to jump-s...	Joss Whedon	Robert Downey Jr., Chris Evans, Mark Ruffalo, ...
124	125	The Dark Knight Rises	Action,Thriller	Eight years after the Joker's reign of anarchy...	Christopher Nolan	Christian Bale, Tom Hardy, Anne Hathaway,Gary ...
578	579	The Hunger Games: Catching Fire	Action,Adventure,Mystery	Katniss Everdeen and Peeta Mellark become targ...	Francis Lawrence	Jennifer Lawrence, Josh Hutcherson, Liam Hemsw...

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

```
top_10
```

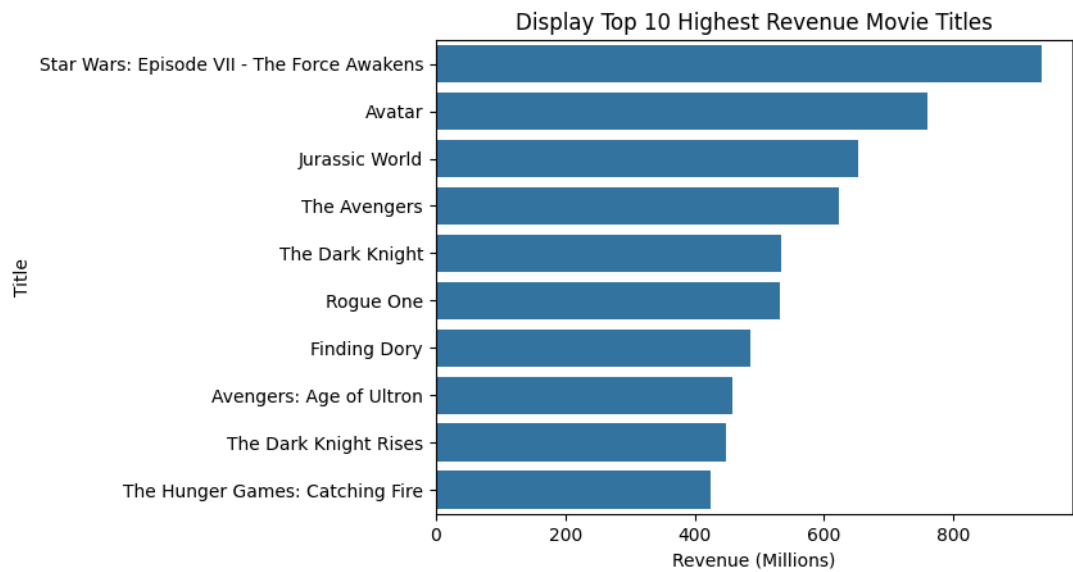
Title	Director	Revenue (Millions)	<div><div></div><div></div><div></div></div>
Star Wars: Episode VII - The Force Awakens	J.J. Abrams	936.63	<div></div> <div></div> <div></div>
Avatar	James Cameron	760.51	
Jurassic World	Colin Trevorrow	652.18	
The Avengers	Joss Whedon	623.28	
The Dark Knight	Christopher Nolan	533.32	
Rogue One	Gareth Edwards	532.17	
Finding Dory	Andrew Stanton	486.29	
Avengers: Age of Ultron	Joss Whedon	458.99	
The Dark Knight Rises	Christopher Nolan	448.13	
The Hunger Games: Catching Fire	Francis Lawrence	424.65	

Next steps:

Generate code with top_10

View recommended plots

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



✓ 18. Find Average Rating of Movies Year-wise

```
data.columns

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

```
data1=data.groupby('Year')['Year','Rating'].mean().\nsort_values(by='Rating',ascending=False).set_index('Year')
```

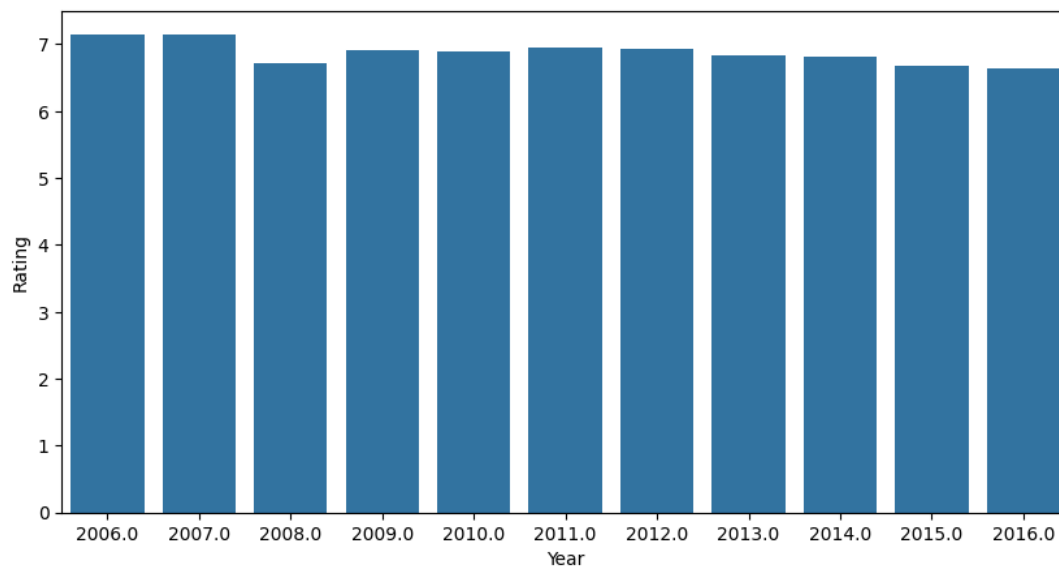
data1

	Rating	
Year		
2006.0	7.143902	
2007.0	7.140909	
2011.0	6.945614	
2012.0	6.933871	
2009.0	6.911111	
2010.0	6.894737	
2013.0	6.832143	
2014.0	6.822581	
2008.0	6.708333	
2015.0	6.674312	
2016.0	6.644444	

Next steps:

[Generate code with data1](#)[View recommended plots](#)

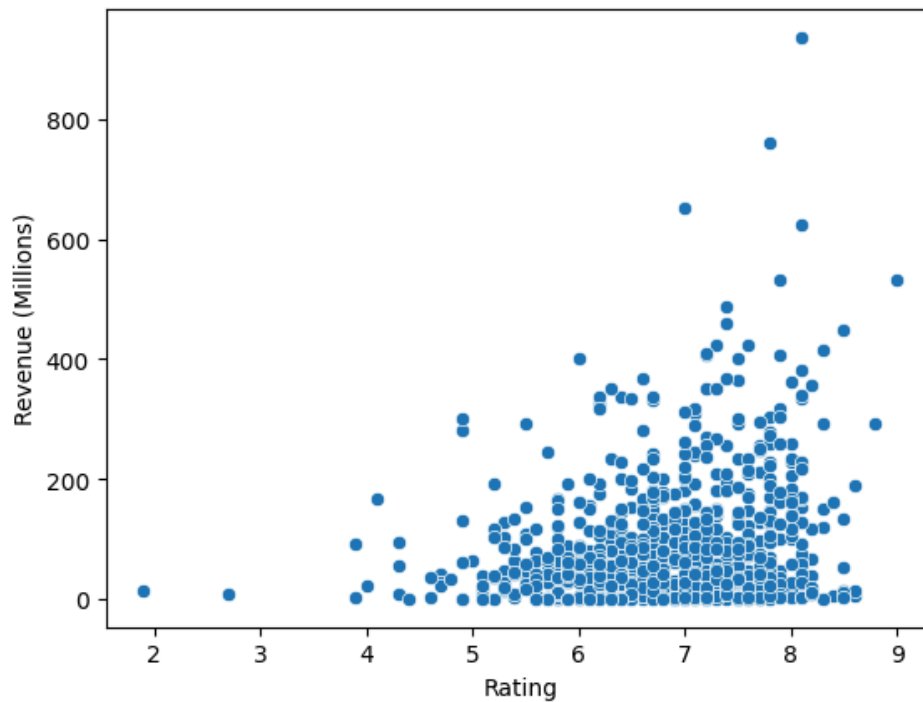
```
plt.figure(figsize=(10,5))\nsns.barplot(data1['Rating'])\nplt.show()
```



✓ 19. Does Rating Affect The Revenue?

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

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



Answer : Yes

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

```
data.columns
```

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

```
def rating(rating):  
    if rating>=7.0:  
        return 'Excellent'  
    elif rating>=6.0:  
        return 'Good'  
    else:  
        return 'Average'
```

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

```
data.head(1)
```

	Rank	Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)
Next		Generate code with data	<input checked="" type="checkbox"/> View recommended plots			Chris Pratt.		

```
from matplotlib import pyplot as plt
data['Runtime (Minutes)'].plot(kind='hist', bins=20, title='Runtime (Minutes)')
plt.gca().spines[['top', 'right',]].set_visible(False)
```

