

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

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

1. Top 10 Rows of The Dataset

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

	Rank	Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)
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...	2014	121	8.1	757074	333.13
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...	2012	124	7.0	485820	126.46
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...	2016	117	7.3	157606	138.12
3	4	Sing	Animation,Comedy,Family	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet	Matthew McConaughey,Reese Witherspoon, Seth Ma...	2016	108	7.2	60545	270.32
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...	2016	123	6.2	393727	325.02
5	6	The Great Wall	Action,Adventure,Fantasy	European mercenaries searching for black powde...	Yimou Zhang	Matt Damon, Tian Jing, Willem Dafoe, Andy Lau	2016	103	6.1	56036	45.13
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....	2016	128	8.3	258682	151.06

2. Top 10 Rows of The Dataset

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

	Rank	Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
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	92	6.6	129708	45.80	
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	165	8.5	102697	1.20	
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	97	6.3	45419	6.92	
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	97	5.9	140900	60.13	
994	995	Project X	Comedy	3 high school seniors throw a birthday party t...	Nima Nourizadeh	Thomas Mann, Oliver Cooper, Jonathan Daniel Br...	2012	88	6.7	164088	54.72	
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	111	6.2	27585	NaN	
996	997	Hostel: Part II	Horror	Three American college students studvina	Eli Roth	Lauren German, Heather Matarazzo, Biiou Philli...	2007	94	5.5	73152	17.54	

3. Number of Rows And 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. Info about Our Dataset

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 Missing Values in The Dataset

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

↔ Any missing valus? True

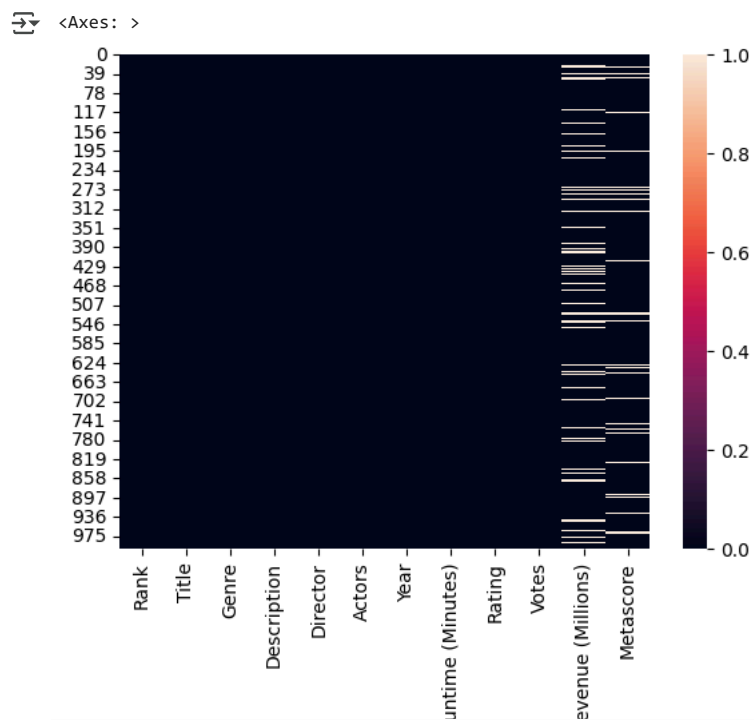
Start coding or [generate](#) with AI.

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

↔

	0
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

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



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

	0
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

6. Drop All The Missing Values

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

	Rank	Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)
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...	2014	121	8.1	757074	333.13
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...	2012	124	7.0	485820	126.46
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...	2016	117	7.3	157606	138.12
3	4	Sing	Animation,Comedy,Family	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet	Matthew McConaughey,Reese Witherspoon, Seth Ma...	2016	108	7.2	60545	270.32
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...	2016	123	6.2	393727	325.02
...
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	97	5.9	140900	60.13
994	995	Project X	Comedy	3 high school seniors throw a birthday party t...	Nima Nourizadeh	Thomas Mann, Oliver Cooper, Jonathan Daniel Br...	2012	88	6.7	164088	54.72
				Three American							

7. Check For Duplicate Data

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

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

Are there any duplicate values? False

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

	Rank	Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)
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...	2014	121	8.1	757074	333.13
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...	2012	124	7.0	485820	126.46
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...	2016	117	7.3	157606	138.12
3	4	Sing	Animation,Comedy,Family	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet	Matthew McConaughey,Reese Witherspoon, Seth Ma...	2016	108	7.2	60545	270.32
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...	2016	123	6.2	393727	325.02
...
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	111	6.2	27585	NaN
996	997	Hostel: Part II	Horror	Three American college students studying	Eli Roth	Lauren German, Heather Matarazzo, Bijou Philli...	2007	94	5.5	73152	17.54

8.Get Overall Statistics About The DataFrame

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

	Rank	Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)	Rating	Votes
count	1000.000000	1000	1000	1000	1000	1000	1000.000000	1000.000000	1000.000000	1.000000e+03
unique	NaN	999	207	1000	644	996	NaN	NaN	NaN	NaN
top	NaN	The Host	Action,Adventure,Sci-Fi	A stuffy businessman finds himself trapped ins...	Ridley Scott	Jennifer Lawrence, Josh Hutcherson, Liam Hemsw...	NaN	NaN	NaN	NaN
freq	NaN	2	50	1	8	2	NaN	NaN	NaN	NaN
mean	500.500000	NaN	NaN	NaN	NaN	NaN	2012.783000	113.172000	6.723200	1.698083e+05
std	288.819436	NaN	NaN	NaN	NaN	NaN	3.205962	18.810908	0.945429	1.887626e+05
min	1.000000	NaN	NaN	NaN	NaN	NaN	2006.000000	66.000000	1.900000	6.100000e+01
25%	250.750000	NaN	NaN	NaN	NaN	NaN	2010.000000	100.000000	6.200000	3.630900e+04
50%	500.500000	NaN	NaN	NaN	NaN	NaN	2014.000000	111.000000	6.800000	1.107990e+05

9. Title of The Movie Having Runtime>=180 Minutes

```
data.columns
```

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

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

```

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

10. In Which Year There Was The Highest Voting

```
data.columns
```

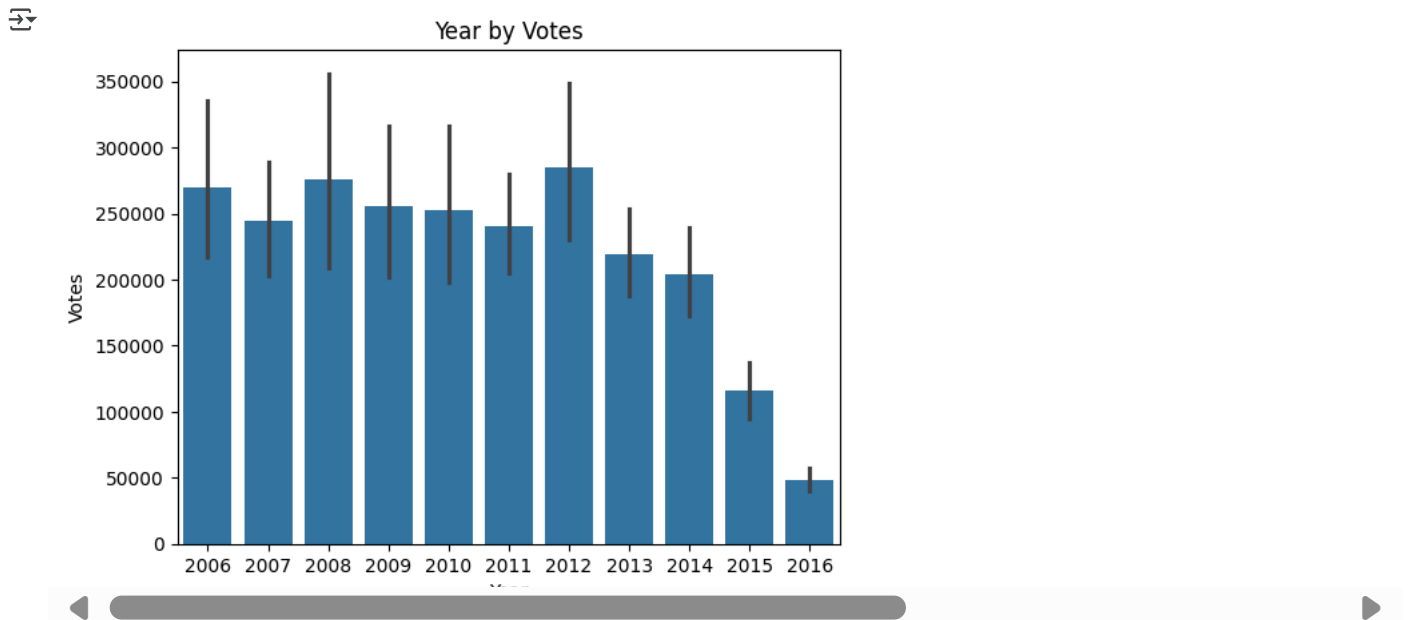
```
Index(['Rank', 'Title', 'Genre', 'Description', 'Director', 'Actors', 'Year',
      'Runtime (Minutes)', 'Rating', 'Votes', 'Revenue (Millions)',
      'Metascore'],
      dtype='object')
```

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

```

      Votes
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
```

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



11. In Which Year There Was The Highest Average Revenue

```
data.columns
```

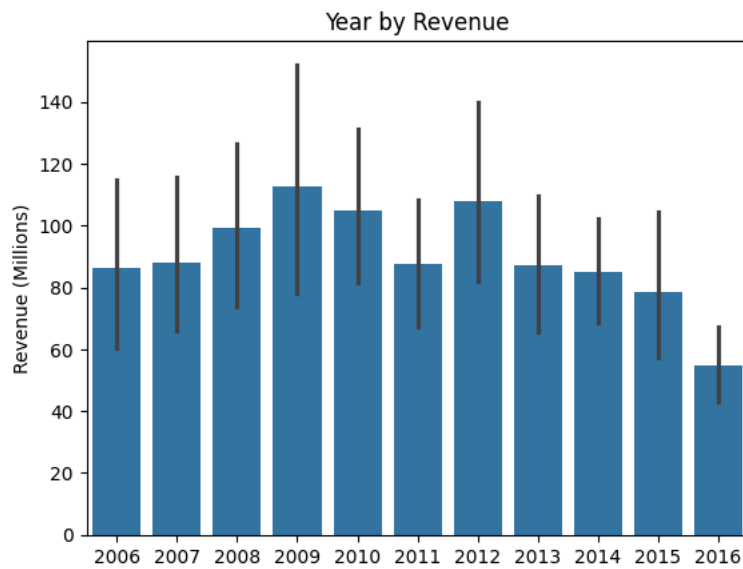
```
Index(['Rank', 'Title', 'Genre', 'Description', 'Director', 'Actors', 'Year',  
      'Runtime (Minutes)', 'Rating', 'Votes', 'Revenue (Millions)',  
      'Metascore'],  
      dtype='object')
```

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

```
Revenue (Millions)
```

Year	Revenue (Millions)
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

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



Average Rating For Each Director

data.columns



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

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



Rating	
Director	
Nitesh Tiwari	8.80
Christopher Nolan	8.68
Olivier Nakache	8.60
Makoto Shinkai	8.60
Florian Henckel von Donnersmarck	8.50
...	...
Micheal Bafaro	3.50
Jonathan Holbrook	3.20
Shawn Burkett	2.70
James Wong	2.70
Jason Friedberg	1.90

644 rows × 1 columns

13. Top 10 Lengthy Movie Title and Runtime

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')
```

top10_len

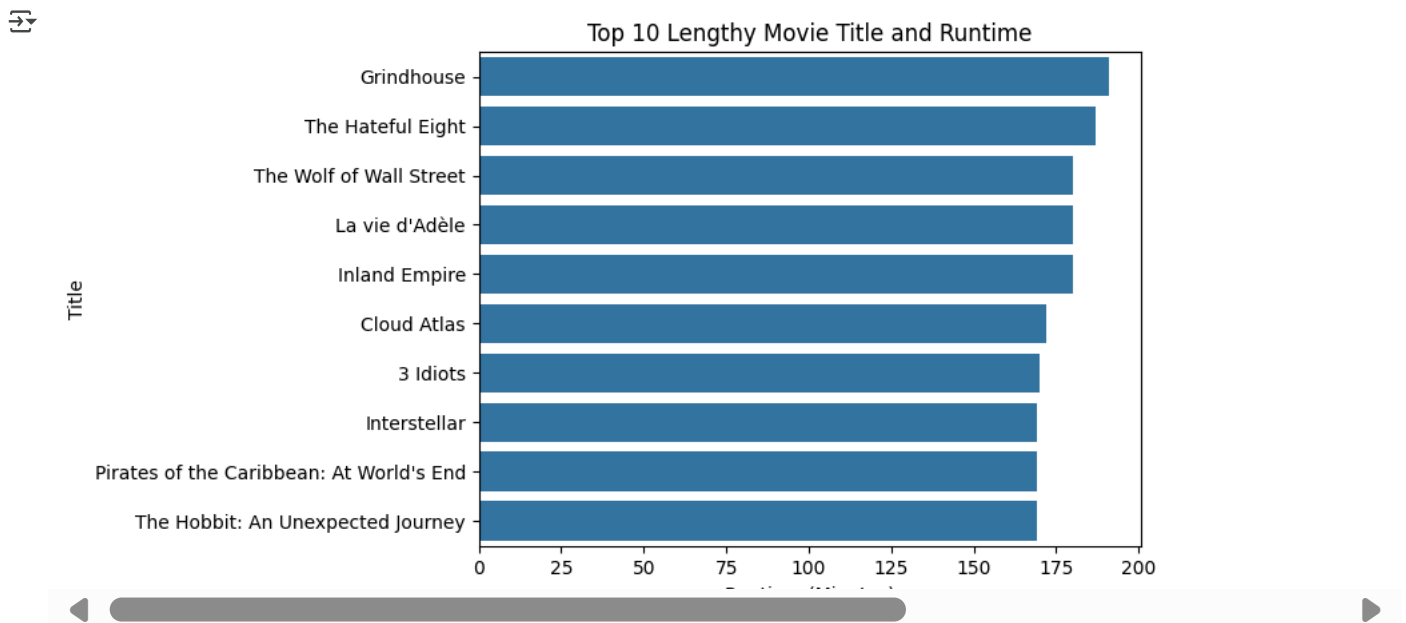
↔

	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	

◀ ▶

Next steps: [Generate code with top10_len](#) [View recommended plots](#) [New interactive sheet](#)

```
sns.barplot(x='Runtime (Minutes)',y=top10_len.index,data=top10_len)
plt.title('Top 10 Lengthy Movie Title and Runtime')
plt.show()
```



Double-click (or enter) to edit

14. Number of Movie Pre Year

```
data.columns
```

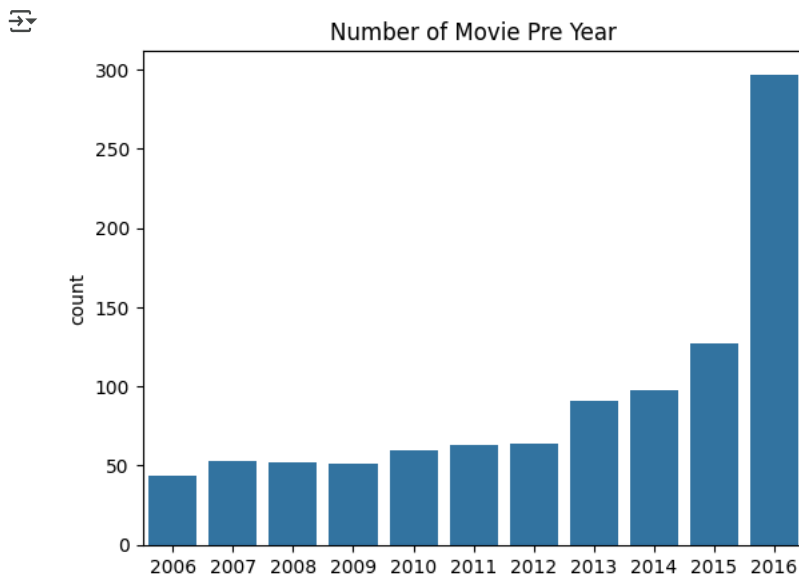
↔

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

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

	count
Year	
2016	297
2015	127
2014	98
2013	91
2012	64
2011	63
2010	60
2007	53
2008	52
2009	51
2006	44

```
sns.countplot(x='Year',data=data)
plt.title('Number of Movie Pre Year')
plt.show()
```



15. Find Most Popular Movie

```
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']
```

	Title
50	Star Wars: Episode VII - The Force Awakens

16. Top 10 Highest Rated Movie and Director

```
data.columns
```

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

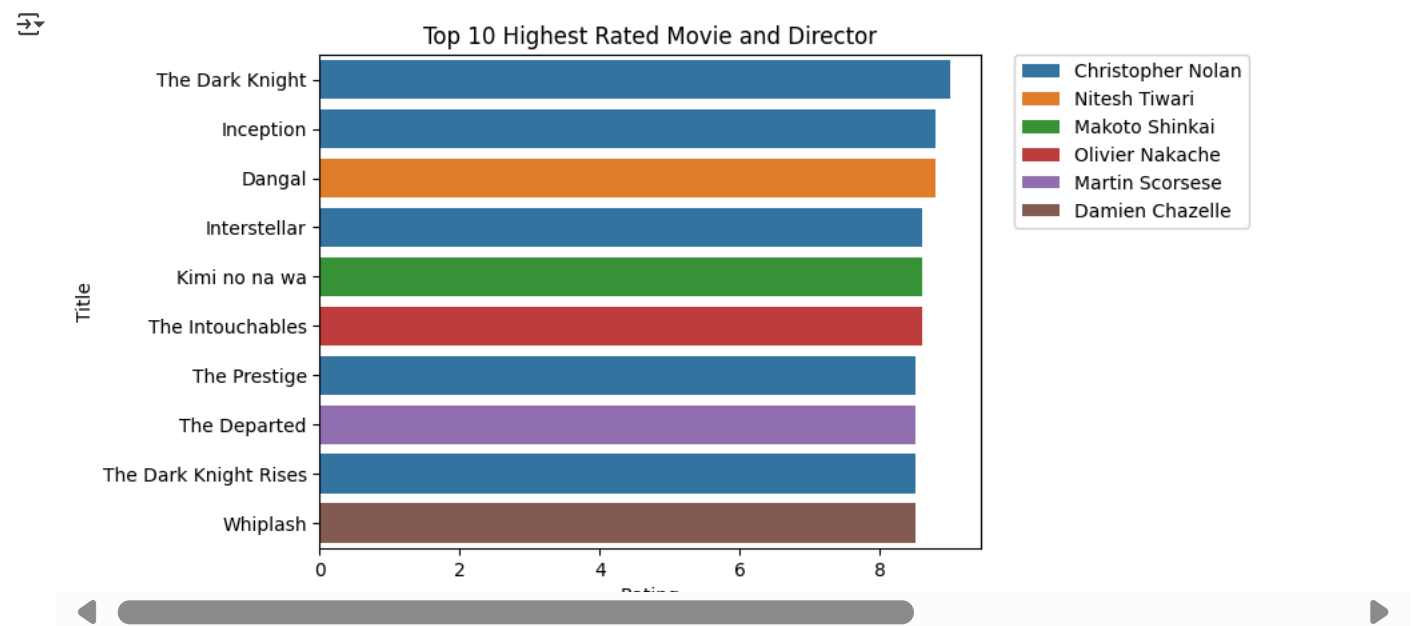
```
top10_len=data.nlargest(10,'Rating')[['Title','Rating','Director']]\
.set_index('Title')
```

top10_len

	Rating	Director
Title		
The Dark Knight	9.0	Christopher Nolan
Inception	8.8	Christopher Nolan
Dangal	8.8	Nitesh Tiwari
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

Next steps: [Generate code with top10_len](#) [View recommended plots](#) [New interactive sheet](#)

```
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, borderaxespad=0.)
plt.title('Top 10 Highest Rated Movie and Director')
plt.show()
```



17. 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')
```

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

top_10

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

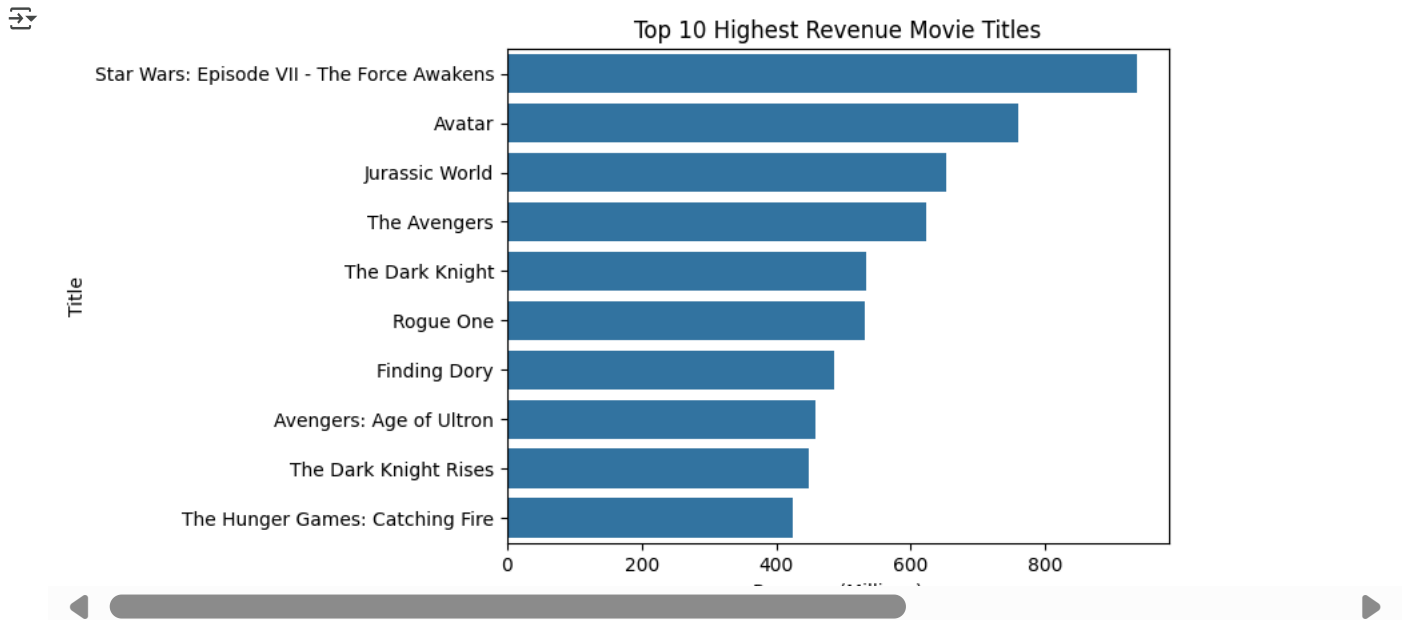
458.99

The Dark Knight Rises

448.13

Next steps: [Generate code with top_10](#) [View recommended plots](#) [New interactive sheet](#)

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



18. 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')
```

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



Rating

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

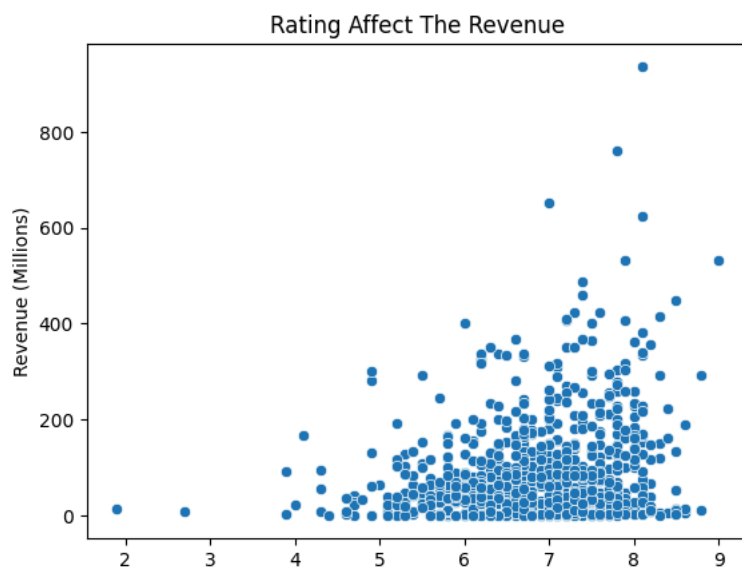
19. Does Rating Affect The Revenue?

```
data.columns
```



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

```
sns.scatterplot(x='Rating',y='Revenue (Millions)',data=data)  
plt.title('Rating Affect The Revenue')  
plt.show()
```



20. Movies Based on Ratings

```
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 'Excelient'  
    elif rating>=6.0:  
        return 'Good'
```

```

else:
    return 'Average'

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

data.head()

```

	Rank	Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	M
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...	2014	121	8.1	757074	333.13	
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...	2012	124	7.0	485820	126.46	
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...	2016	117	7.3	157606	138.12	
3	4	Sing	Animation,Comedy,Family	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet	Matthew McConaughey,Reese Witherspoon, Seth Ma...	2016	108	7.2	60545	270.32	

Next steps: [Generate code with data](#) [View recommended plots](#) [New interactive sheet](#)

21. Count Number of Action Movies

```

data.columns

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

data['Genre'].dtype

dtype('O')

len(data[data['Genre'].str.contains('Action',case=False)])

303

```

Unique Values From Genera

```

data.columns

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

data['Genre']

```



Genre

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

1000 rows × 1 columns



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

list1



```
[ 'Drama', 'Music', 'Romance'],  
['Adventure', 'Comedy'],  
['Comedy', 'Family', 'Fantasy']]
```

```
one_d=[]  
for item in list1:  
    for item1 in item:  
        one_d.append(item1)
```

one_d

```
↩ [ 'Action',  
    'Adventure']
```