

The logo for IMDb, featuring the letters "IMDb" in a bold, black, sans-serif font. The letter "I" has a vertical yellow bar through its center, and the "b" has a vertical yellow bar through its left side.

# MOVIE RATING ANALYSIS



## **ABOUT IMDB**

**The Internet Movie Database (IMDb) is an online database containing information and statistics about movies, TV shows and video games as well as actors, directors and other film industry professionals.**



### **What are IMDb ratings?**

IMDb registered users can cast a vote (from 1 to 10) on every released title in the database.

Individual votes are then aggregated and summarized as a single IMDb rating, visible on the title's main page.

# OBJECTIVE

**Certainly! In this data analysis project, we will explore a movie dataset, aiming to uncover insights about the film industry. Our tasks include checking for missing values, calculating average ratings for directors, identifying temporal trends, and analyzing the relationship between ratings and revenue. Additionally, we'll delve into genres, count action movies, and determine the most popular film titles.**

**Let's embark on this cinematic journey!**

# 1. DISPLAY TOP 10 ROWS OF THE DATASET

## SOLUTION

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

Rank	Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
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	76.0
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	65.0
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	62.0
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	59.0
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	40.0
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	42.0
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	93.0
7	8 Mindhorn	Comedy	A has-been actor best known for playing the ti...	Sean Foley	Essie Davis, Andrea Riseborough, Julian Barrat... ...	2016	89	6.4	2490	NaN	71.0
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... ...	2016	141	7.1	7188	8.01	78.0
9	10 Passengers	Adventure, Drama, Romance	A spacecraft traveling to a distant colony pla...	Morten Tyldum	Jennifer Lawrence, Chris Pratt, Michael Sheen,... ...	2016	116	7.0	192177	100.01	41.0

1

## Showcase the First 10 Rows

The first step is to display the top 10 rows of the dataset. This will give us a high-level overview of the data and help us understand the structure and content of the information we're working with.

2

## Identify Key Columns and Data Types

By examining the top 10 rows, we can quickly identify the key columns in the dataset and the data types of the values in each column. This information will be crucial for our subsequent data analysis and manipulation steps.

3

## Spot Any Potential Issues

Reviewing the top 10 rows can also help us spot any potential issues or anomalies in the data, such as missing values, outliers, or inconsistencies. This will allow us to address these problems early on in the analysis process.

## 2. CHECK LAST 10 ROWS OF THE DATASET

### SOLUTION

```
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	44.0
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	42.0
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	NaN
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	37.0
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	48.0
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	45.0
996	997 Hostel: Part II	Horror	Three American college students studying abroa...	Eli Roth	Lauren German, Heather Matarazzo, Bijou Philli...	2007	94	5.5	73152	17.54	46.0
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	98	6.2	70699	58.01	50.0
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	93	5.6	4881	NaN	22.0
999	1000 Nine Lives	Comedy,Family,Fantasy	A stuffy businessman finds himself trapped ins...	Barry Sonnenfeld	Kevin Spacey, Jennifer Garner, Robbie Amell,Ch...	2016	87	5.3	12435	19.64	11.0

1

### Inspect the Last 10 Rows

Reviewing the last 10 rows of the dataset can provide valuable insights into the data structure and any potential anomalies or outliers that may be present.

2

### Identify Potential Issues

Checking the last 10 rows can also help you identify any potential issues, such as missing values, incorrect data types, or unexpected data points, that may need to be addressed.

3

### Validate Data Integrity

Reviewing the last 10 rows can also help you validate the integrity of the data and ensure that it is consistent with the rest of the dataset.

### 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
```

The dataset has 1,000 rows and 20 columns, providing a comprehensive view of the data.

# 4. GETTING INFORMATION ABOUT OUR DATASET

`data.info()`

Getting information about our dataset is a crucial step in understanding the data we're working with. This includes finding the total number of rows and columns, the data types of each column, and the memory requirements of the dataset.

```
<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 value? True
```

`data.isnull().sum()`

- Identify the columns with missing values in the dataset
- Determine the percentage of missing values in each column
- Visualize the missing values using a heatmap or other appropriate plot
- Decide on the appropriate method to handle the missing values, such as imputation or dropping the rows/columns
- Implement the chosen method to replace or remove the missing values
- Verify that the missing values have been successfully handled

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
<b>dtype:</b>	<b>int64</b>

# 6. DROP ALL THE MISSING VALUES

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



					C...								
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	48.0
996	997	Hostel: Part II		Horror	Three American college students studying abroa...	Eli Roth	Lauren German, Heather Matarazzo, Bijou Philli...	2007	94	5.5	73152	17.54	46.0
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	98	6.2	70699	58.01	50.0
999	1000	Nine Lives	Comedy,Family,Fantasy		A stuffy businessman finds himself trapped ins...	Barry Sonnenfeld	Kevin Spacey, Jennifer Garner, Robbie Amell,Ch...	2016	87	5.3	12435	19.64	11.0

838 rows × 12 columns

It is evident that the number of rows has been reduced from 1000 to 838, indicating that 162 rows containing missing values have been dropped.

# 7. CHECK FOR DUPLICATE DATA

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



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

```
Are there any duplicate values? False
```

## 1 Identify Duplicates

Examine the dataset to identify any duplicate rows or entries. This can be done by checking for rows with identical values across all columns.

## 2 Drop Duplicates

Once the duplicate rows have been identified, they can be removed from the dataset to ensure data integrity and accuracy.

## 3 Verify Uniqueness

After dropping the duplicate rows, double-check the dataset to ensure that all remaining rows are unique and there are no further duplicates.

# 8. GET OVERALL STATISTICS ABOUT THE DATAFRAME

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

	Rank	Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
count	1000.000000	1000		1000	1000	1000	1000.000000	1000.000000	1000.000000	1.000000e+03	872.000000	936.000000
unique		Nan	999		207	1000	644	996	Nan	Nan	Nan	Nan
top		Nan	The Host	Action, Adventure, Sci-Fi	A group of intergalactic criminals are forced ...	Ridley Scott	Jennifer Lawrence, Josh Hutcherson, Liam Hemsw...	Nan	Nan	Nan	Nan	Nan
freq		Nan	2		50	1	8	2	Nan	Nan	Nan	Nan
mean	500.500000	Nan			Nan	Nan	Nan	2012.783000	113.172000	6.723200	1.698083e+05	82.956376
std	288.819436	Nan			Nan	Nan	Nan	3.205962	18.810908	0.945429	1.887626e+05	103.253540
min	1.000000	Nan			Nan	Nan	Nan	2006.000000	66.000000	1.900000	6.100000e+01	0.000000
25%	250.750000	Nan			Nan	Nan	Nan	2010.000000	100.000000	6.200000	3.630900e+04	13.270000
50%	500.500000	Nan			Nan	Nan	Nan	2014.000000	111.000000	6.800000	1.107990e+05	47.985000
75%	750.250000	Nan			Nan	Nan	Nan	2016.000000	123.000000	7.400000	2.399098e+05	113.715000
max	1000.000000	Nan			Nan	Nan	Nan	2016.000000	191.000000	9.000000	1.791916e+06	936.630000



The `include='all'` parameter ensures that both numerical and categorical columns are included in the summary. This function is useful for gaining insights into the distribution, central tendency, and spread of data across all columns in the DataFrame.

## 9. DISPLAY TITLE OF THE MOVIE HAVING RUNTIME GREATER THAN OR EQUAL TO 180 MINUTES

### SOLUTION

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

- Analyzing the runtime of a movie is very important as it affects many other factors.
- Balancing runtime ensures viewer engagement. Too short, and the story feels rushed; too long, and it risks losing the audience's attention.
- Knowing the runtime helps viewers plan their movie-watching experience.
- Calculated the movie titles having runtime greater than or equal to 180 minutes that's equals to 3 hours or more.

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

# 10. IN WHICH YEAR THERE WAS THE HIGHEST AVERAGE VOTING?

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

## Analyze Voting Data

To determine the year with the highest average voting, we will need to examine the voting data across all movies and calculate the average for each year.

Year	Average Voting
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

## Understand Viewer Preferences

The year with the highest average voting can provide insights into the types of movies and content that resonated most with viewers during that time period.

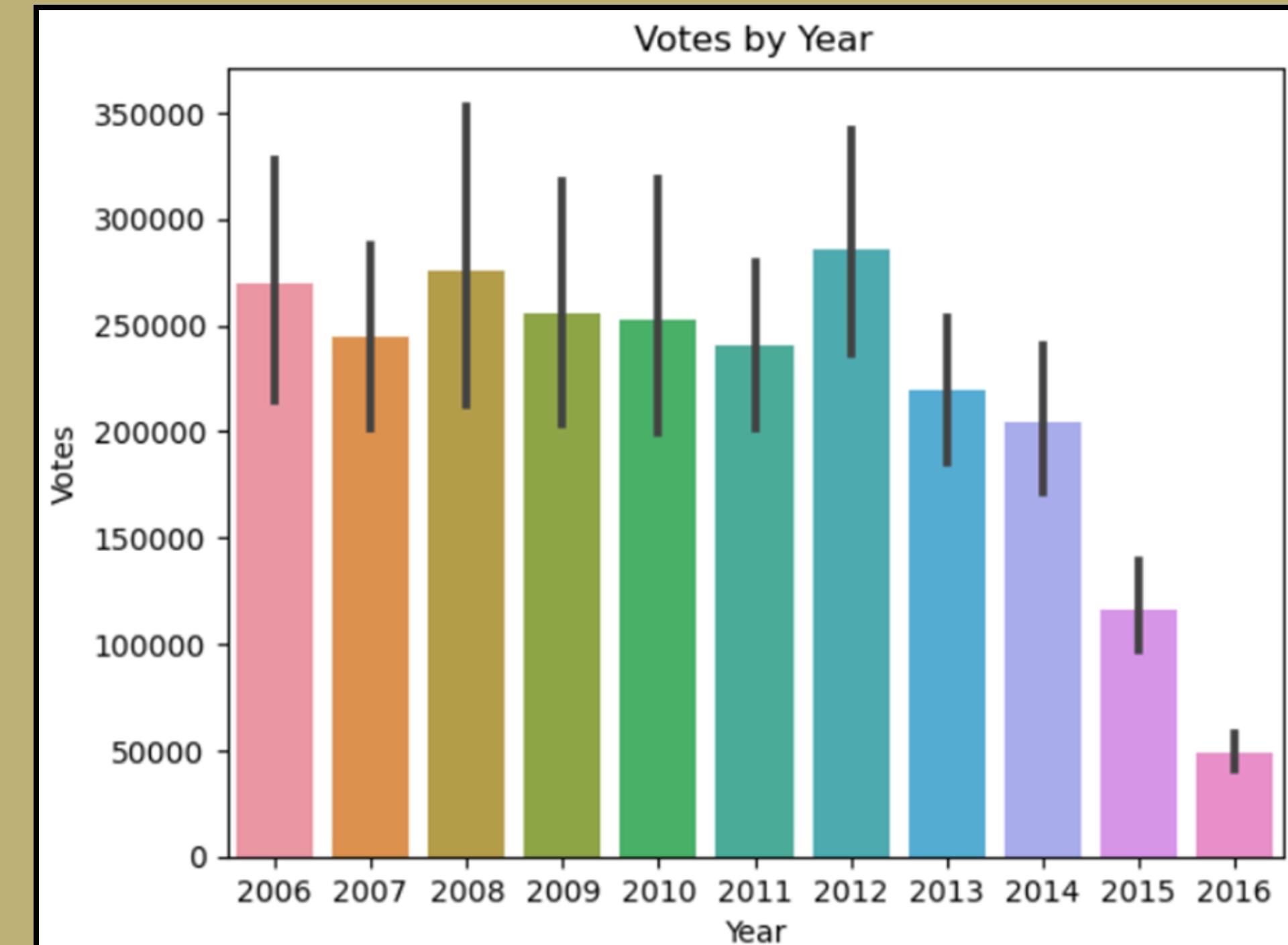
# 10. IN WHICH YEAR THERE WAS THE HIGHEST AVERAGE VOTING?

## VISUALIZATION

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

### Analyzing the Voting Trend

By visualizing the voting trends of movies with respect to year, the year 2012 seems to have the highest number of total votes cast according to the graph and as the year passing, the voting trend is going doing having least votes in 2016.



# 11. IN WHICH YEAR THERE WAS THE HIGHEST AVERAGE REVENUE?

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

Year	Average 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

## Analyze Revenue Data

To determine the year with the highest average revenue, we will need to carefully analyze the revenue data for each year in the dataset.

## Identify Peak Year

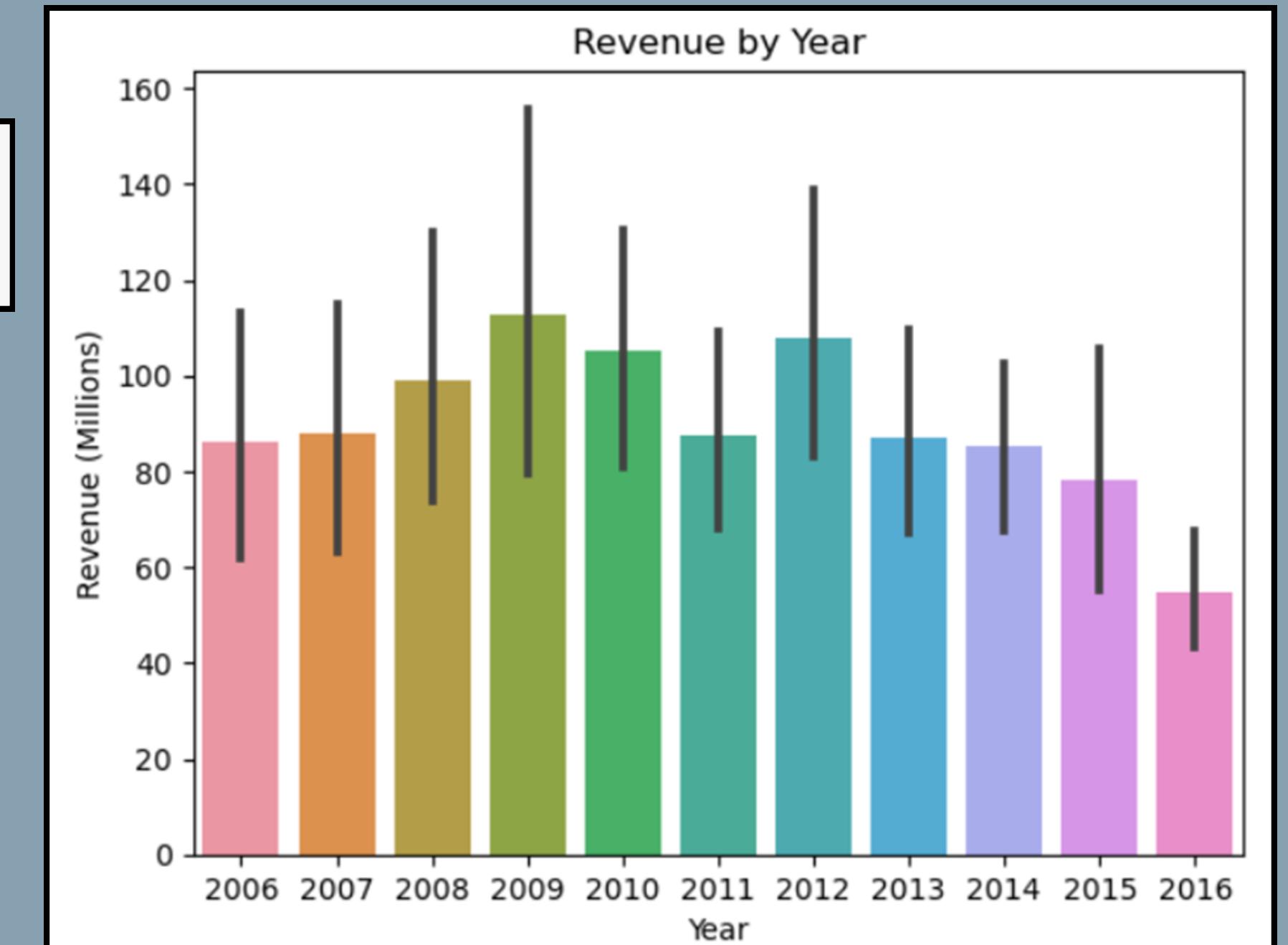
Once we have the average revenue for each year, we can identify the year with the highest average, which will be the answer to this question.

# 11. IN WHICH YEAR THERE WAS THE HIGHEST AVERAGE REVENUE?

## VISUALIZATION

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

The x-axis shows years, likely from 2006 to 2016, and the y-axis shows revenue in millions. The revenue appears to fluctuate somewhat over the years. In 2009 the revenue seems to be highest around 110 million, whereas in 2016 it appears to be closer to 50 million.



# 12. FIND THE AVERAGE RATING FOR EACH DIRECTOR

```
data.groupby("Director")["Rating"].mean().sort_values()
```

Ratings of directors serve several important purposes in the world of film and entertainment as a highly-rated director often indicates quality content, while lower ratings may signal potential issues.

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

We calculated the average ratings for each director and found that the director Nitesh Tiwari tops the rating by 8.80 and Jason Friedberg has the least ratings of 1.90.

# 13. DISPLAY TOP 10 LENGTHY MOVIES TITLE AND RUNTIME

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

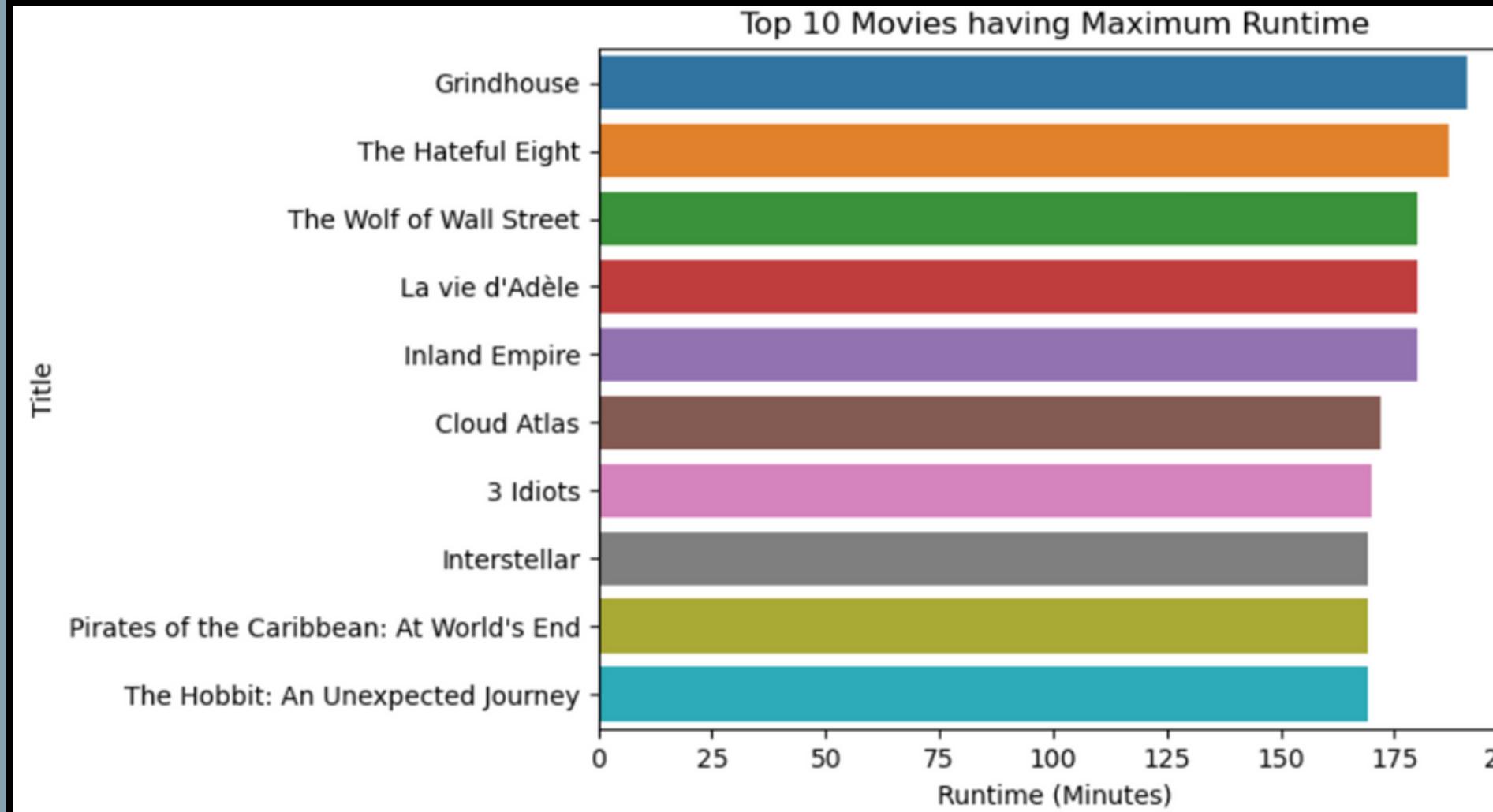
- It aims to display the titles and runtimes of the top 10 longest movies in the dataset.
- The length of a movie can be an important factor for some viewers, as longer films often indicate an epic scope or more in-depth storytelling.
- By highlighting the longest movies, we can provide useful information to help users make informed decisions about which films to watch based on their preferences for runtime.

Title	Runtime (Minutes)
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

# 13. DISPLAY TOP 10 LENGTHY MOVIES TITLE AND RUNTIME

## VISUALIZATION

```
sns.barplot(x="Runtime (Minutes)", y= top10_len.index, data=top10_len)
plt.title("Top 10 Movies having Maximum Runtime")
plt.show()
```



1. **Grindhouse** : Holds the record for the longest runtime, spanning 191 minutes.
2. **The Hateful Eight** : Second longest, with a runtime of 187 minutes.
3. **The Wolf of Wall Street** : Also lengthy, clocking in at 180 minutes.

# 14. DISPLAY NUMBER OF MOVIES PER YEAR

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

**Analyzing Movie Trends**  
By examining the number of movies produced each year, we can identify trends and patterns in the film industry.

**Identifying Peaks and Valleys**  
This information can help us understand factors that influence movie production, such as economic conditions, technological advancements, or changes in audience preferences.

1

2

3

**Visualizing the Data**  
Displaying the movie count per year allows us to see which years had the highest and lowest production volumes.

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

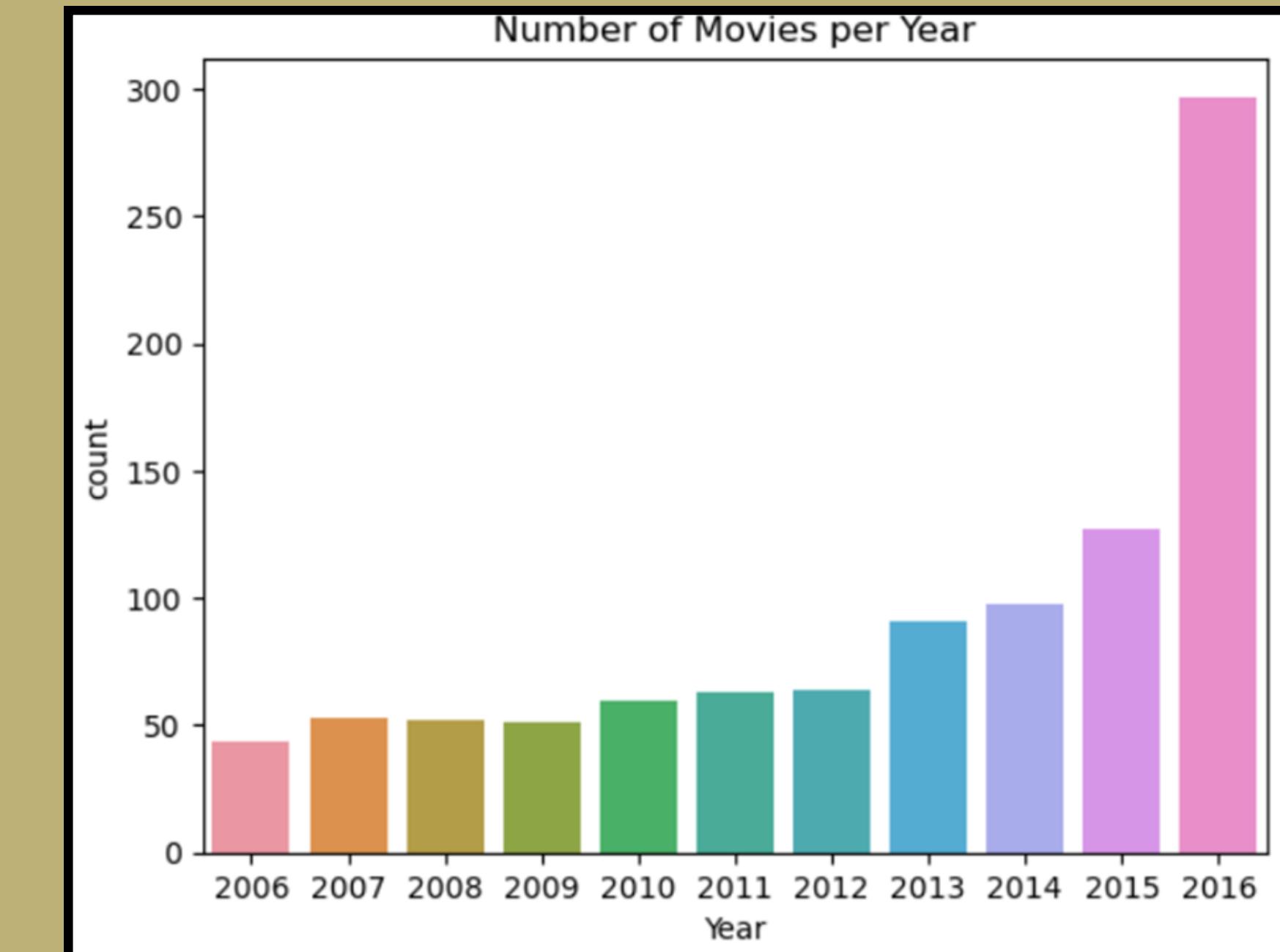
Name: count, dtype: int64

## 14. DISPLAY NUMBER OF MOVIES PER YEAR

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

### VISUALIZATION

The bar graph shows upward trend indicating that as the year passes the number of movies per year is increasing, which is maximum in 2016 total to 297.



## 15. FIND MOST POPULAR MOVIE TITLE (HIGHEST REVENUE)

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



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

“Star Wars: Episode VII - The Force Awakens” is the most popular movie having highest revenue

# 16. DISPLAY TOP 10 HIGHEST RATED MOVIE TITLES AND ITS DIRECTORS

```
top10_rating = data.nlargest(10,'Rating')[['Title','Rating','Director']].set_index("Title")
top10_rating
```

Title	Rating	Director
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
Whiplash	8.5	Damien Chazelle

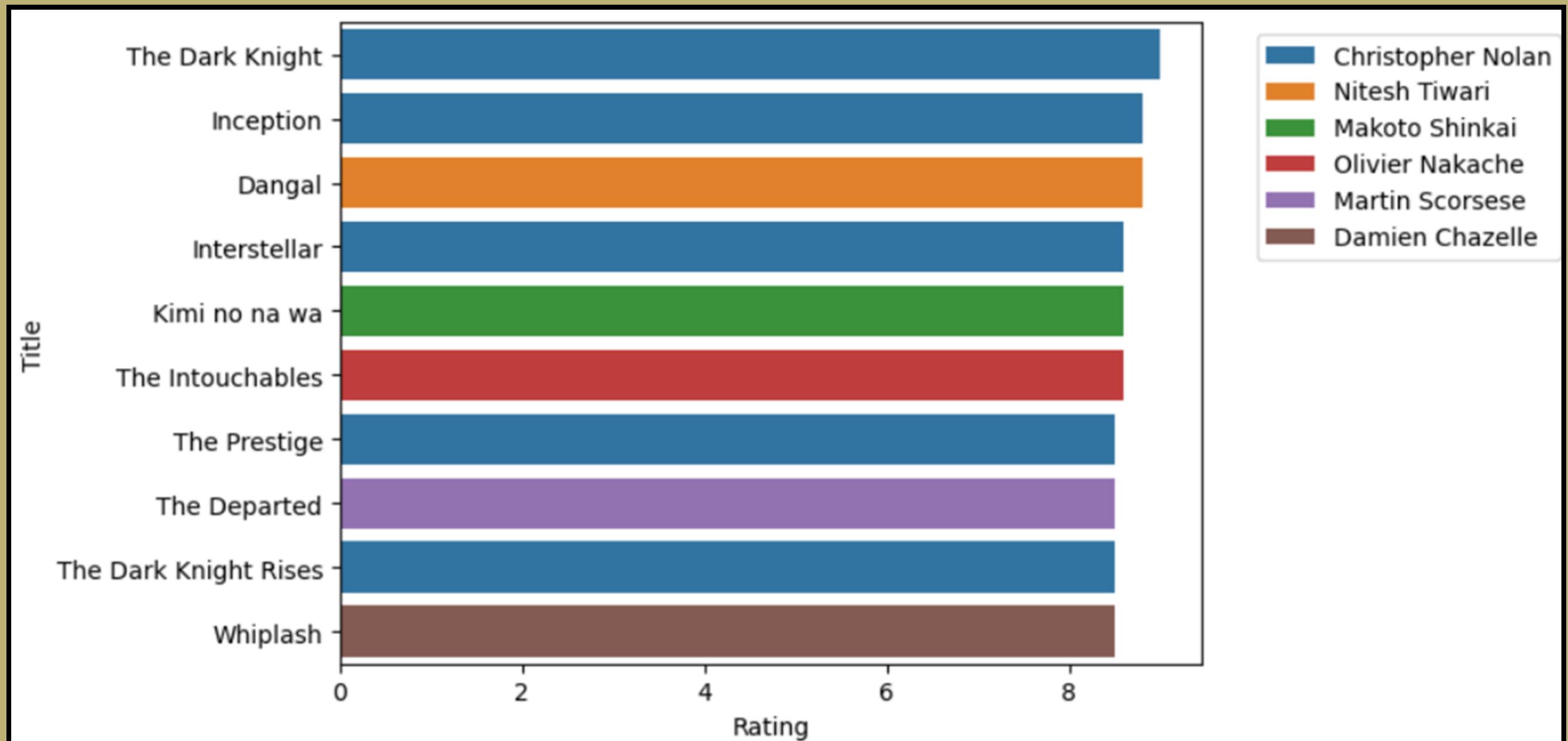
- Calculating the highest rated movie titles and their directors offers several valuable benefits for both film enthusiasts and the industry.
- Knowing the highest rated movies allows viewers to prioritize their watchlist.
- High ratings serve as a benchmark for filmmakers.
- High ratings increase the chances of winning awards and being featured at prestigious film festivals.

# 16. DISPLAY TOP 10 HIGHEST RATED MOVIE TITLES AND ITS DIRECTORS

## VISUALIZATION

It is quite evident that the director 'Christopher Nolan' is top rated having 5 of his movies in Top10 highest rated movies.

```
sns.barplot(x='Rating', y=top10_rating.index, data=top10_rating, hue="Director", dodge=False)  
plt.legend(bbox_to_anchor=(1.05,1),loc=2)  
plt.show()
```



# 17. DISPLAY TOP 10 HIGHEST REVENUE MOVIE TITLES

```
top10 = data.nlargest(10,'Revenue (Millions)')[['Title','Revenue (Millions)']].set_index("Title")
top10
```

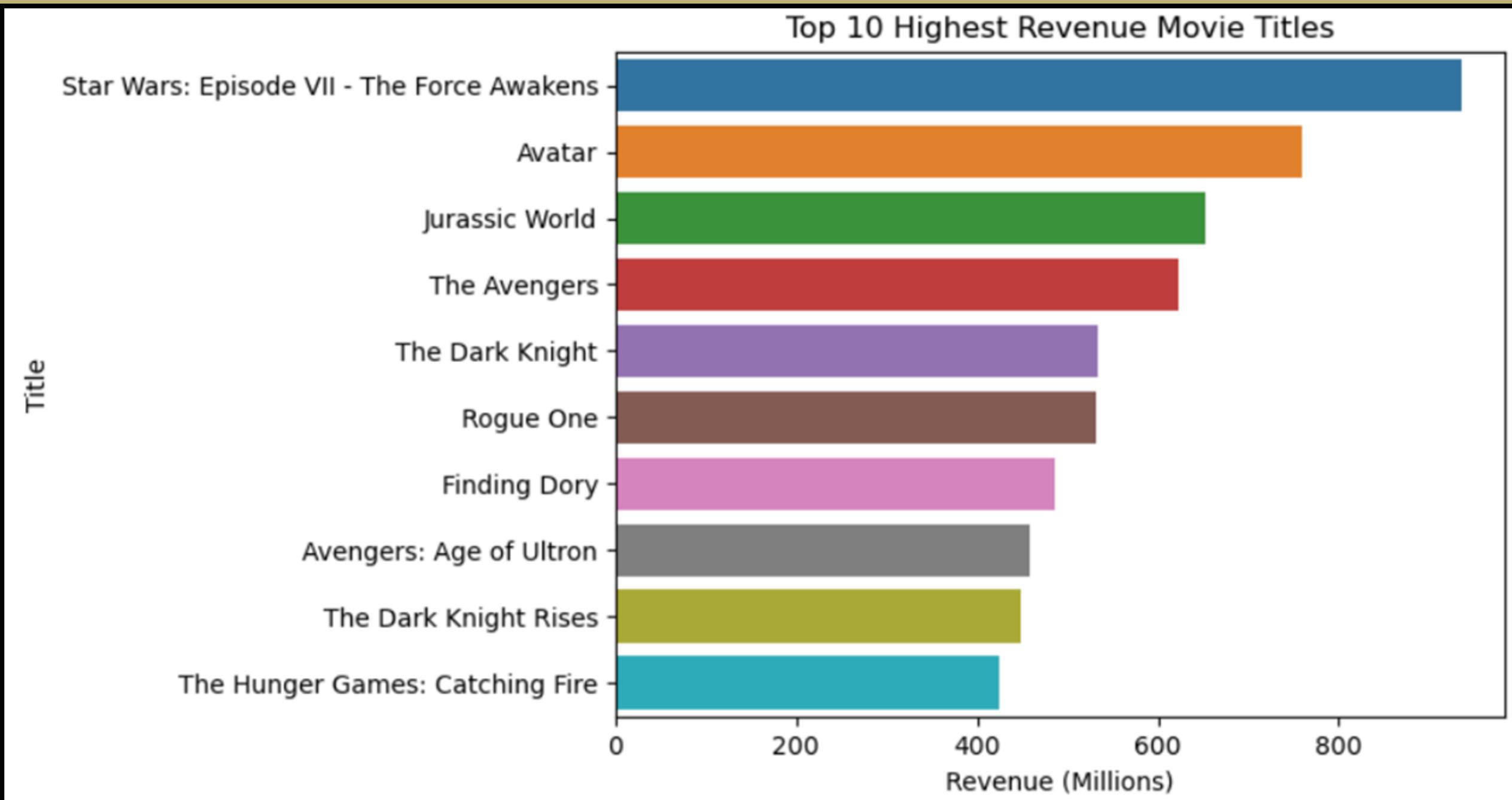
Title	Revenue (Millions)
Star Wars: Episode VII - The Force Awakens	936.63
Avatar	760.51
Jurassic World	652.18
The Avengers	623.28
The Dark Knight	533.32
Rogue One	532.17
Finding Dory	486.29
Avengers: Age of Ultron	458.99
The Dark Knight Rises	448.13
The Hunger Games: Catching Fire	424.65

- Analyzing the top 10 highest-grossing movie titles provides valuable insights for the film industry and enthusiasts.
- Examining the highest-grossing films reveals trends in audience preferences, marketing strategies, and global box office performance.
- Understanding what resonates with audiences helps in planning release dates, marketing campaigns, and distribution.

# 17. DISPLAY TOP 10 HIGHEST REVENUE MOVIE TITLES

```
sns.barplot(x='Revenue (Millions)', y= top10.index, data = top10)
plt.title("Top 10 Highest Revenue Movie Titles")
plt.show()
```

'Star Wars' has highest revenue whereas 'The Hunger Games' has the lowest revenue.



# 18. FIND AVERAGE RATING OF MOVIES YEAR WISE

```
data.groupby("Year")["Rating"].mean().sort_values()
```

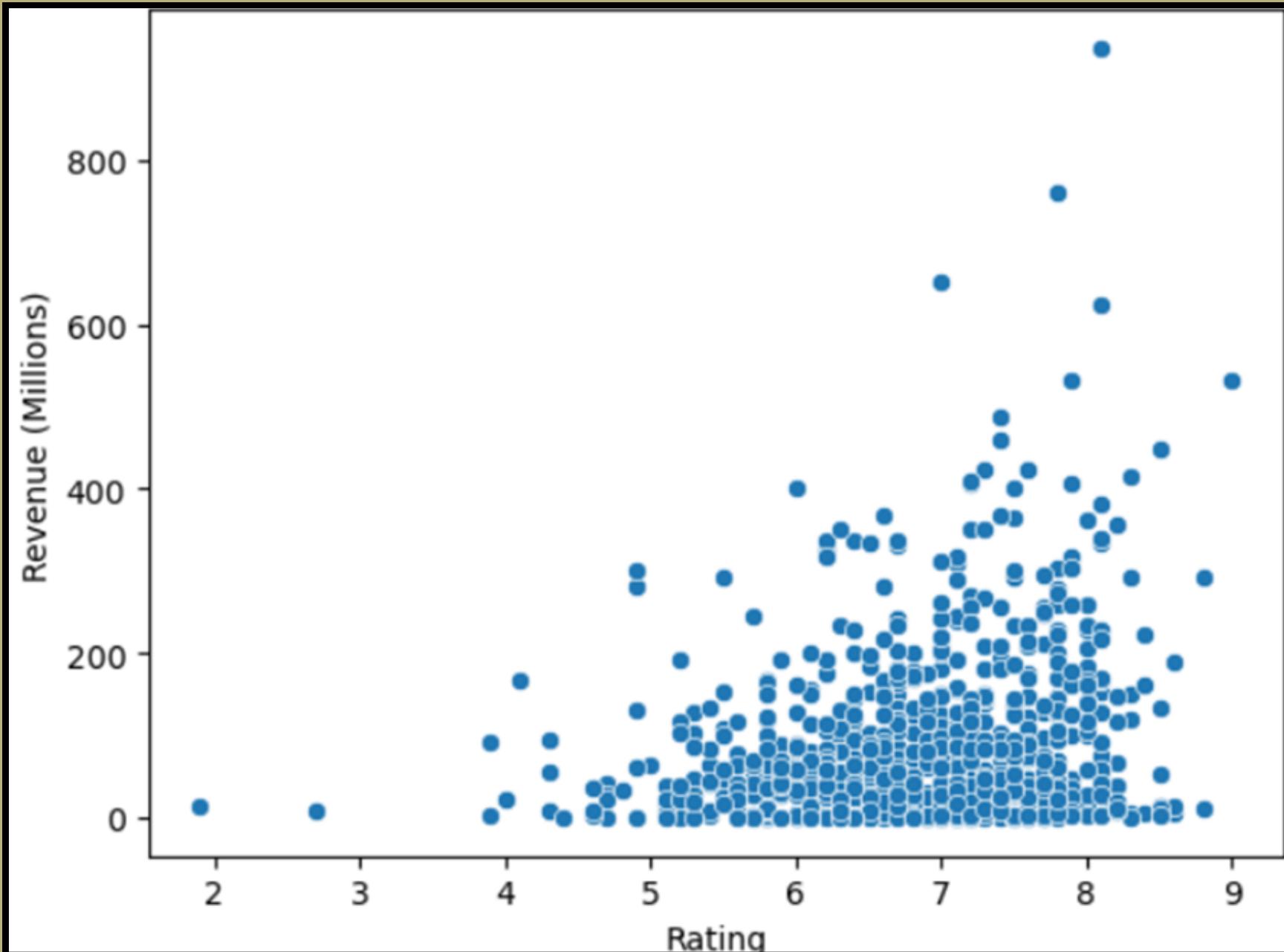
- Analyzing average movie ratings year-wise can provide valuable insights for both filmmakers and audiences.
- We can observe trends in audience preferences.
- Year-wise ratings allow us to assess the overall quality of movies released in a particular year.
- In this data, 2016 has lowest ratings whereas 2007 has highest, which indicates that movie ratings has decreased with time.

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

Name: Rating, dtype: float64

# 19. DOES RATING AFFECT THE REVENUE?

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



Based on the scatter plot in the image, there appears to be a positive correlation between movie ratings and revenue. Here are the key observations:

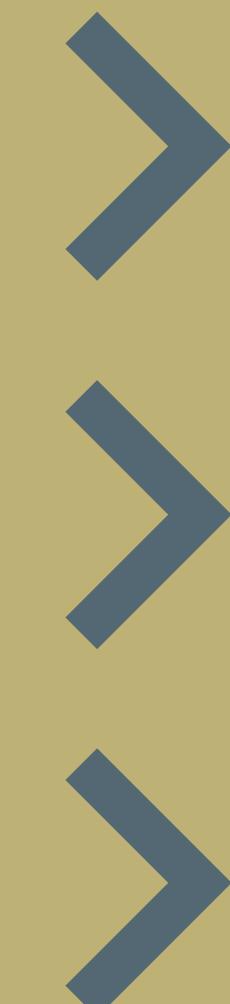
- **Positive Trend:** As the rating increases, the revenue tends to increase as well.
- **Concentration:** There's a concentration of dots around ratings of 5 to 8, with revenues between 100 and 200 million.
- **Outliers:** Some dots at higher ratings show significantly higher revenues.
- **Quality Matters:** Higher-rated movies tend to attract larger audiences, leading to higher box office revenues.

# 20. CLASSIFY MOVIES BASED ON RATINGS [EXCELLENT, GOOD, AND AVERAGE]

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



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



Description	Director	Actors	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore	Rating_Category
A group of intergalactic criminals are forced ...	James Gunn	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	2014	121	8.1	7074	333.13	76.0	Excellent
Following clues to the origin of mankind, a te...	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fa...	2012	124	7.0	5820	126.46	65.0	Excellent
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	7606	138.12	62.0	Excellent
In a city of humanoid animals, a hustling thea...	Christophe Lourdelet	Matthew McConaughey,Reese Witherspoon, Seth Ma...	2016	108	7.2	5045	270.32	59.0	Excellent
A secret government agency recruits some of th...	David Ayer	Will Smith, Jared Leto, Margot Robbie, Viola D...	2016	123	6.2	3727	325.02	40.0	Good
...	...	...	...	...	...	...	...	...	...
A tight-knit team of rising investigators, alo...	Billy Ray	Chiwetel Ejiofor, Nicole Kidman, Julia Roberts...	2015	111	6.2	7585	Nan	45.0	Good
Three American college students studying abroa...	Eli Roth	Lauren German, Heather Matarazzo, Bijou Phillips...	2007	94	5.5	3152	17.54	46.0	Average
Romantic sparks occur between two dance studen...	Jon M. Chu	Robert Hoffman, Briana Evigan, Cassie Ventura,...	2008	98	6.2	10699	58.01	50.0	Good

## 21. COUNT NUMBER OF ACTION MOVIES

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

303

Calculated the number of movies of one specific Genre  
i.e, Action which comes equals to 303

## 22. FIND UNIQUE VALUES FROM GENRE

STEP 1: Display the ‘Genre’ column

```
data['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
Name: Genre, Length: 1000, dtype: object
```

STEP 2: Splitting each category of  
Genre by commas

```
# splitting all by commas
list1 = []
for value in data['Genre']:
    list1.append(value.split(","))
```

```
list1
```

```
[['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'],
 ['Biography', 'Drama'],
```

# 22. FIND UNIQUE VALUES FROM GENRE

STEP 3: Converting 2D list into 1D

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

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

STEP 4: Extracted unique values

```
unique_list = []
for item in one_d:
    if item not in unique_list:
        unique_list.append(item)

unique_list
```

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

## UNIQUE VALUES

There are total 20 Genres in the  
IMBD Movie Data

# 23. HOW MANY FILMS OF EACH GENRE WERE MADE?

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

```
['Action',
 'Adventure',
 'Sci-Fi',
 'Adventure',
 'Mystery',
 'Sci-Fi',
 'Horror',
 'Thriller',
 'Animation',
 'Comedy',
 'Family',
 'Action',
 'Adventure',
 'Fantasy',
 'Action',
 'Adventure',
 'Fantasy',
 'Comedy',
 'Drama',
```

```
from collections import Counter
```

Counter(one\_d)

“DRAMA” Genre has maximum movies which is 513.  
“MUSICAL” Genre has minimum movies i.e, just 5.

```
Counter({'Drama': 513,
          'Action': 303,
          'Comedy': 279,
          'Adventure': 259,
          'Thriller': 195,
          'Crime': 150,
          'Romance': 141,
          'Sci-Fi': 120,
          'Horror': 119,
          'Mystery': 106,
          'Fantasy': 101,
          'Biography': 81,
          'Family': 51,
          'Animation': 49,
          'History': 29,
          'Sport': 18,
          'Music': 16,
          'War': 13,
          'Western': 7,
          'Musical': 5})
```