

Analyzing and Visualizing Movie Data

Movie Dataset Analysis Report

1. Introduction

- **Objective:**

To explore the dataset and derive insights into trends, correlations, and relationships in the movie industry.

- **Dataset Overview:**

- Number of records: 36.
- Key columns: Movie Title, Industry, Release Year, IMDB Rating, Studio, Budget, Revenue.

	movie_id	title	industry	language	release_year	imdb_rating	stu
0	101	K.G.F: Chapter 2	Bollywood	Kannada	2022	8.4	Hom F
1	102	Doctor Strange in the Multiverse of Madness	Hollywood	English	2022	7.0	Ma Stu
2	103	Thor: The Dark World	Hollywood	English	2013	6.8	Ma Stu
3	104	Thor: Ragnarok	Hollywood	English	2017	7.9	Ma Stu
4	105	Thor: Love and Thunder	Hollywood	English	2022	6.8	Ma Stu

Out[975...]

	movie_id	title	industry	language	release_year	imdb_rating
0	101	K.G.F: Chapter 2	Bollywood	Kannada	2022	8.4
1	102	Doctor Strange in the Multiverse of Madness	Hollywood	English	2022	7.0
2	103	Thor: The Dark World	Hollywood	English	2013	6.8
3	104	Thor: Ragnarok	Hollywood	English	2017	7.9
4	105	Thor: Love and Thunder	Hollywood	English	2022	6.8
5	111	The Shawshank Redemption	Hollywood	English	1994	9.3
6	113	Interstellar	Hollywood	English	2014	8.6
7	115	The Pursuit of Happyness	Hollywood	English	2006	8.0
8	116	Gladiator	Hollywood	English	2000	8.5
9	117	Titanic	Hollywood	English	1997	7.9
10	118	It's a Wonderful Life	Hollywood	English	1946	8.6
11	119	Avatar	Hollywood	English	2009	7.8
12	120	The Godfather	Hollywood	English	1972	9.2
13	121	The Dark Knight	Hollywood	English	2008	9.0
14	122	Schindler's List	Hollywood	English	1993	9.0
15	123	Jurassic Park	Hollywood	English	1993	8.2
16	124	Parasite	Hollywood	English	2019	8.5
17	125	Avengers: Endgame	Hollywood	English	2019	8.4
18	126	Avengers: Infinity War	Hollywood	English	2018	8.4

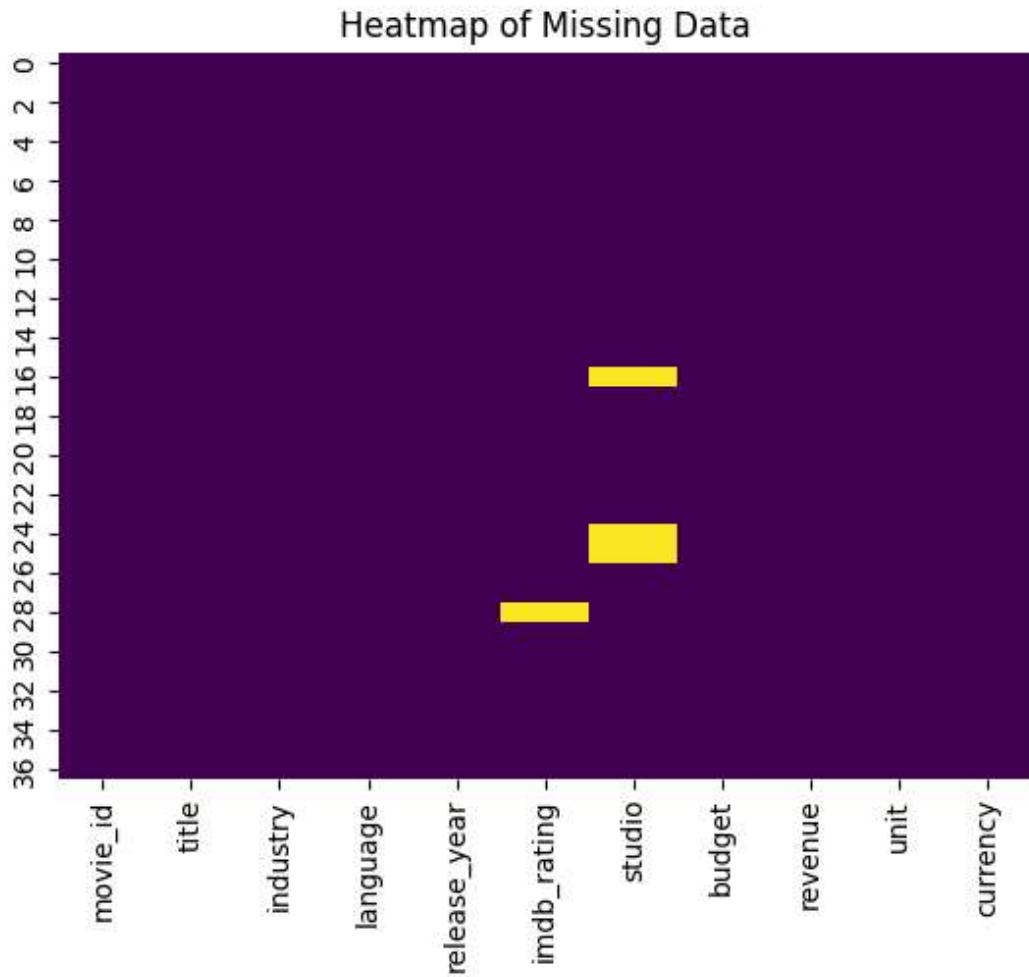
	movie_id	title	industry	language	release_year	imdb_rating
19	137	Captain America: The First Avenger	Hollywood	English	2011	6.9
20	138	Captain America: The Winter Soldier	Hollywood	English	2014	7.8
21	107	Dilwale Dulhania Le Jayenge	Bollywood	Hindi	1995	8.0
22	108	3 Idiots	Bollywood	Hindi	2009	8.4
23	109	Kabhi Khushi Kabhie Gham	Bollywood	Hindi	2001	7.4
24	110	Bajirao Mastani	Bollywood	Hindi	2015	7.2
25	128	Taare Zameen Par	Bollywood	Hindi	2007	8.3
26	129	Munna Bhai M.B.B.S.	Bollywood	Hindi	2003	8.1
27	130	PK	Bollywood	Hindi	2014	8.1
28	131	Sanju	Bollywood	Hindi	2018	NaN
29	135	The Kashmir Files	Bollywood	Hindi	2022	8.3
30	136	Bajrangi Bhaijaan	Bollywood	Hindi	2015	8.1
31	139	Race 3	Bollywood	Hindi	2018	1.9
32	140	Shershaah	Bollywood	Hindi	2021	8.4
33	127	Pather Panchali	Bollywood	Bengali	1955	8.3
34	132	Pushpa: The Rise - Part 1	Bollywood	Telugu	2021	7.6
35	133	RRR	Bollywood	Telugu	2022	8.0
36	134	Baahubali: The Beginning	Bollywood	Telugu	2015	8.0

2. Data Cleaning and Preprocessing

- Handled missing values in key columns (`IMDB Rating`, `Studio`):
 - Filled `Studio` with mode based on the `Industry`.
 - Filled `IMDB Rating` with the mean value of the column.
- Removed duplicates and standardized categorical columns like `Industry` and `Language`.

```
In [976]: import seaborn as sns
import matplotlib.pyplot as plt

# Visualize missing data
sns.heatmap(df.isnull(), cbar=False, cmap="viridis")
plt.title("Heatmap of Missing Data")
plt.show()
```



```
In [977]: # checking count of null values as column wise
print(df.isnull().sum())
```

```
movie_id      0
title         0
industry      0
language      0
release_year  0
imdb_rating   1
studio        3
budget        0
revenue       0
unit          0
currency      0
dtype: int64
```

```
In [978... # checking industry of null imdb rating
df[df.imdb_rating.isnull()]['industry']
```

```
Out[978... 28    Bollywood
Name: industry, dtype: object
```

```
In [979... df.replace('Vinod Chopra Productions', 'Vinod Chopra Films', inplace=True)
```

```
In [980... # fill null imdb rating with mean imdb rating of respective industry (i
df.fillna({
    'imdb_rating' : (df[df.industry == 'Bollywood']['imdb_rating'].mean
}))
```

Out[980...]

	movie_id	title	industry	language	release_year	imdb_rating
0	101	K.G.F: Chapter 2	Bollywood	Kannada	2022	8.40000
1	102	Doctor Strange in the Multiverse of Madness	Hollywood	English	2022	7.00000
2	103	Thor: The Dark World	Hollywood	English	2013	6.80000
3	104	Thor: Ragnarok	Hollywood	English	2017	7.90000
4	105	Thor: Love and Thunder	Hollywood	English	2022	6.80000
5	111	The Shawshank Redemption	Hollywood	English	1994	9.30000
6	113	Interstellar	Hollywood	English	2014	8.60000
7	115	The Pursuit of Happyness	Hollywood	English	2006	8.00000
8	116	Gladiator	Hollywood	English	2000	8.50000
9	117	Titanic	Hollywood	English	1997	7.90000
10	118	It's a Wonderful Life	Hollywood	English	1946	8.60000
11	119	Avatar	Hollywood	English	2009	7.80000
12	120	The Godfather	Hollywood	English	1972	9.20000
13	121	The Dark Knight	Hollywood	English	2008	9.00000
14	122	Schindler's List	Hollywood	English	1993	9.00000
15	123	Jurassic Park	Hollywood	English	1993	8.20000
16	124	Parasite	Hollywood	English	2019	8.50000
17	125	Avengers: Endgame	Hollywood	English	2019	8.40000
18	126	Avengers: Infinity War	Hollywood	English	2018	8.40000

	movie_id	title	industry	language	release_year	imdb_rating
19	137	Captain America: The First Avenger	Hollywood	English	2011	6.90000
20	138	Captain America: The Winter Soldier	Hollywood	English	2014	7.80000
21	107	Dilwale Dulhania Le Jayenge	Bollywood	Hindi	1995	8.00000
22	108	3 Idiots	Bollywood	Hindi	2009	8.40000
23	109	Kabhi Khushi Kabhie Gham	Bollywood	Hindi	2001	7.40000
24	110	Bajirao Mastani	Bollywood	Hindi	2015	7.20000
25	128	Taare Zameen Par	Bollywood	Hindi	2007	8.30000
26	129	Munna Bhai M.B.B.S.	Bollywood	Hindi	2003	8.10000
27	130	PK	Bollywood	Hindi	2014	8.10000
28	131	Sanju	Bollywood	Hindi	2018	7.65625
29	135	The Kashmir Files	Bollywood	Hindi	2022	8.30000
30	136	Bajrangi Bhaijan	Bollywood	Hindi	2015	8.10000
31	139	Race 3	Bollywood	Hindi	2018	1.90000
32	140	Shershaah	Bollywood	Hindi	2021	8.40000
33	127	Pather Panchali	Bollywood	Bengali	1955	8.30000
34	132	Pushpa: The Rise - Part 1	Bollywood	Telugu	2021	7.60000
35	133	RRR	Bollywood	Telugu	2022	8.00000
36	134	Baahubali: The Beginning	Bollywood	Telugu	2015	8.00000

```
In [981...]: df[df.studio.isnull()]      # find null studios
```

	movie_id	title	industry	language	release_year	imdb_rating	studio
16	124	Parasite	Hollywood	English	2019	8.5	NaN
24	110	Bajirao Mastani	Bollywood	Hindi	2015	7.2	NaN
25	128	Taare Zameen Par	Bollywood	Hindi	2007	8.3	NaN

```
In [982...]: # find industry wise mode(most frequent) for studio
```

```
print(df[df.industry == 'Bollywood'].studio.mode())
print(df[df.industry == 'Hollywood'].studio.mode())
```

```
0    Vinod Chopra Films
Name: studio, dtype: object
0    Marvel Studios
Name: studio, dtype: object
```

```
In [983...]: # fill missing studio with mode of respective industry
```

```
def fillNull(row):
    if row['industry'] == 'Bollywood':
        return 'Vinod Chopra Films'
    elif row['industry'] == 'Hollywood':
        return 'Marvel Studios'
    else:
        return row['studio']

df['studio'] = df.apply(lambda row: fillNull(row) if pd.isnull(row['studio']) else row['studio'])
```

```
In [984...]: # convert all units with standard million
```

```
def convert(row):
    if row['unit'] == 'Billions':
        row['unit'] = 'Millions'
        row['budget'] = row['budget'] * 1000
        row['revenue'] = row['revenue'] * 1000
    return row

    elif row['unit'] == 'Thousands':
        row['unit'] = 'Millions'
        row['budget'] = row['budget'] * 0.001
        row['revenue'] = row['revenue'] * 0.001
    return row

    else:
        return row

df = df.apply(lambda row: convert(row), axis=1)
```

```
In [985...]: # convert all units with standard USD with rate at 85 inr to usd
```

```
def convert(row):
    if row['currency'] == 'INR':
        row['currency'] = 'USD'
```

```
        row[ 'budget' ] = row[ 'budget' ] * 0.012
        row[ 'revenue' ] = row[ 'revenue' ] * 0.012
        return row
    else:
        return row

df = df.apply(lambda row: convert(row), axis=1)
```

```
In [986]: df # cleaned data with no null values
```

Out[986]:

	movie_id	title	industry	language	release_year	imdb_rating
0	101	K.G.F: Chapter 2	Bollywood	Kannada	2022	8.4
1	102	Doctor Strange in the Multiverse of Madness	Hollywood	English	2022	7.0
2	103	Thor: The Dark World	Hollywood	English	2013	6.8
3	104	Thor: Ragnarok	Hollywood	English	2017	7.9
4	105	Thor: Love and Thunder	Hollywood	English	2022	6.8
5	111	The Shawshank Redemption	Hollywood	English	1994	9.3
6	113	Interstellar	Hollywood	English	2014	8.6
7	115	The Pursuit of Happyness	Hollywood	English	2006	8.0
8	116	Gladiator	Hollywood	English	2000	8.5
9	117	Titanic	Hollywood	English	1997	7.9
10	118	It's a Wonderful Life	Hollywood	English	1946	8.6
11	119	Avatar	Hollywood	English	2009	7.8
12	120	The Godfather	Hollywood	English	1972	9.2
13	121	The Dark Knight	Hollywood	English	2008	9.0
14	122	Schindler's List	Hollywood	English	1993	9.0
15	123	Jurassic Park	Hollywood	English	1993	8.2
16	124	Parasite	Hollywood	English	2019	8.5
17	125	Avengers: Endgame	Hollywood	English	2019	8.4
18	126	Avengers: Infinity War	Hollywood	English	2018	8.4

	movie_id	title	industry	language	release_year	imdb_rating
19	137	Captain America: The First Avenger	Hollywood	English	2011	6.9
20	138	Captain America: The Winter Soldier	Hollywood	English	2014	7.8
21	107	Dilwale Dulhania Le Jayenge	Bollywood	Hindi	1995	8.0
22	108	3 Idiots	Bollywood	Hindi	2009	8.4
23	109	Kabhi Khushi Kabhie Gham	Bollywood	Hindi	2001	7.4
24	110	Bajirao Mastani	Bollywood	Hindi	2015	7.2
25	128	Taare Zameen Par	Bollywood	Hindi	2007	8.3
26	129	Munna Bhai M.B.B.S.	Bollywood	Hindi	2003	8.1
27	130	PK	Bollywood	Hindi	2014	8.1
28	131	Sanju	Bollywood	Hindi	2018	NaN
29	135	The Kashmir Files	Bollywood	Hindi	2022	8.3
30	136	Bajrangi Bhaijaan	Bollywood	Hindi	2015	8.1
31	139	Race 3	Bollywood	Hindi	2018	1.9
32	140	Shershaah	Bollywood	Hindi	2021	8.4
33	127	Pather Panchali	Bollywood	Bengali	1955	8.3
34	132	Pushpa: The Rise - Part 1	Bollywood	Telugu	2021	7.6
35	133	RRR	Bollywood	Telugu	2022	8.0
36	134	Baahubali: The Beginning	Bollywood	Telugu	2015	8.0

```
In [987...]: df['profit'] = df.revenue - df.budget  
df
```

Out[987]:

	movie_id	title	industry	language	release_year	imdb_rating
0	101	K.G.F: Chapter 2	Bollywood	Kannada	2022	8.4
1	102	Doctor Strange in the Multiverse of Madness	Hollywood	English	2022	7.0
2	103	Thor: The Dark World	Hollywood	English	2013	6.8
3	104	Thor: Ragnarok	Hollywood	English	2017	7.9
4	105	Thor: Love and Thunder	Hollywood	English	2022	6.8
5	111	The Shawshank Redemption	Hollywood	English	1994	9.3
6	113	Interstellar	Hollywood	English	2014	8.6
7	115	The Pursuit of Happyness	Hollywood	English	2006	8.0
8	116	Gladiator	Hollywood	English	2000	8.5
9	117	Titanic	Hollywood	English	1997	7.9
10	118	It's a Wonderful Life	Hollywood	English	1946	8.6
11	119	Avatar	Hollywood	English	2009	7.8
12	120	The Godfather	Hollywood	English	1972	9.2
13	121	The Dark Knight	Hollywood	English	2008	9.0
14	122	Schindler's List	Hollywood	English	1993	9.0
15	123	Jurassic Park	Hollywood	English	1993	8.2
16	124	Parasite	Hollywood	English	2019	8.5
17	125	Avengers: Endgame	Hollywood	English	2019	8.4
18	126	Avengers: Infinity War	Hollywood	English	2018	8.4

	movie_id	title	industry	language	release_year	imdb_rating
19	137	Captain America: The First Avenger	Hollywood	English	2011	6.9
20	138	Captain America: The Winter Soldier	Hollywood	English	2014	7.8
21	107	Dilwale Dulhania Le Jayenge	Bollywood	Hindi	1995	8.0
22	108	3 Idiots	Bollywood	Hindi	2009	8.4
23	109	Kabhi Khushi Kabhie Gham	Bollywood	Hindi	2001	7.4
24	110	Bajirao Mastani	Bollywood	Hindi	2015	7.2
25	128	Taare Zameen Par	Bollywood	Hindi	2007	8.3
26	129	Munna Bhai M.B.B.S.	Bollywood	Hindi	2003	8.1
27	130	PK	Bollywood	Hindi	2014	8.1
28	131	Sanju	Bollywood	Hindi	2018	NaN
29	135	The Kashmir Files	Bollywood	Hindi	2022	8.3
30	136	Bajrangi Bhaijaan	Bollywood	Hindi	2015	8.1
31	139	Race 3	Bollywood	Hindi	2018	1.9
32	140	Shershaah	Bollywood	Hindi	2021	8.4
33	127	Pather Panchali	Bollywood	Bengali	1955	8.3
34	132	Pushpa: The Rise - Part 1	Bollywood	Telugu	2021	7.6
35	133	RRR	Bollywood	Telugu	2022	8.0
36	134	Baahubali: The Beginning	Bollywood	Telugu	2015	8.0

3. Key Insights

3.1 Industry Analysis

- Most prominent industries in the dataset: **Hollywood**.
- Top-performing industry in terms of revenue: **Hollywood**.
- Average revenue by industry:
 - Hollywood: **929 Million USD**.
 - Bollywood: **57 Million USD**.

3.2 IMDB Ratings

- Average IMDB Rating:
 - Hollywood: **8.1**
 - Bollywood: **7.7**
- Movies with highest ratings:
 - Hollywood: **The Shawshank Redemption**
 - Bollywood: **K.G.F: Chapter 2**

3.3 Revenue and Budget Trends

- Average budget for movies: **89.0**
- Average revenue for movies: **528.35**
- Highest-grossing movie: **Avatar** with a revenue of **2847.0**

3.4 Studio Analysis

- Top 3 studios producing the highest revenue:
 1. **Marvel Studios**
 2. **20th Century Fox**
 3. **Paramount Pictures**

```
In [988...]: # General Analysis

# Industry wise Imdb rating
print('Bollywood:')
print('Minimun Imdb rating :', df[df.industry == 'Bollywood'].imdb_rating.min())
print('Maximum Imdb rating :', df[df.industry == 'Bollywood'].imdb_rating.max())
print('Average Imdb rating :', df[df.industry == 'Bollywood'].imdb_rating.mean())
print('Average Revenue :', df[df.industry == 'Bollywood'].revenue.mean())

print()
print('Hollywood:')
print('Minimun Imdb rating :', df[df.industry == 'Hollywood'].imdb_rating.min())
print('Maximum Imdb rating :', df[df.industry == 'Hollywood'].imdb_rating.max())
print('Average Imdb rating :', df[df.industry == 'Hollywood'].imdb_rating.mean())
print('Average Revenue :', df[df.industry == 'Hollywood'].revenue.mean())
```

Bollywood:

Minimun Imdb rating : 1.9
Maximum Imdb rating : 8.4
Average Imdb rating : 7.7
Average Revenue : 57.0

Hollywood:

Minimun Imdb rating : 6.8
Maximum Imdb rating : 9.3
Average Imdb rating : 8.1
Average Revenue : 929.0

In [989...]

```
# Top 5 movies of Bollywood by imdb rating
df[df.industry == 'Bollywood'].sort_values(by='imdb_rating', ascending=
```

Out[989...]

	movie_id	title	industry	language	release_year	imdb_rating	
0	101	K.G.F: Chapter 2	Bollywood	Kannada	2022	8.4	H
22	108	3 Idiots	Bollywood	Hindi	2009	8.4	
32	140	Shershaah	Bollywood	Hindi	2021	8.4	Proc
25	128	Taare Zameen Par	Bollywood	Hindi	2007	8.3	
29	135	The Kashmir Files	Bollywood	Hindi	2022	8.3	Zee

◀ ▶

In [990...]

```
# Top 5 movies of Hollywood by imdb rating
df[df.industry == 'Hollywood'].sort_values(by='imdb_rating', ascending=
```

Out[990...]

	movie_id	title	industry	language	release_year	imdb_rating	
5	111	The Shawshank Redemption	Hollywood	English	1994	9.3	Er
12	120	The Godfather	Hollywood	English	1972	9.2	
13	121	The Dark Knight	Hollywood	English	2008	9.0	
14	122	Schindler's List	Hollywood	English	1993	9.0	
6	113	Interstellar	Hollywood	English	2014	8.6	\

◀ ▶

In [991...]

```
# Average budget for movies
# Average revenue for movies
# Highest-grossing movie
```

```

print("Average budget for movies:", df.budget.mean().round(2))
print("Average revenue for movies:", df.revenue.mean().round(2))
print(f"Highest-grossing movie: {df[df.revenue == df.revenue.max()].titl

```

Average budget for movies: 89.0
 Average revenue for movies: 528.35
 Highest-grossing movie: ['Avatar'] with a revenue of 2847.0

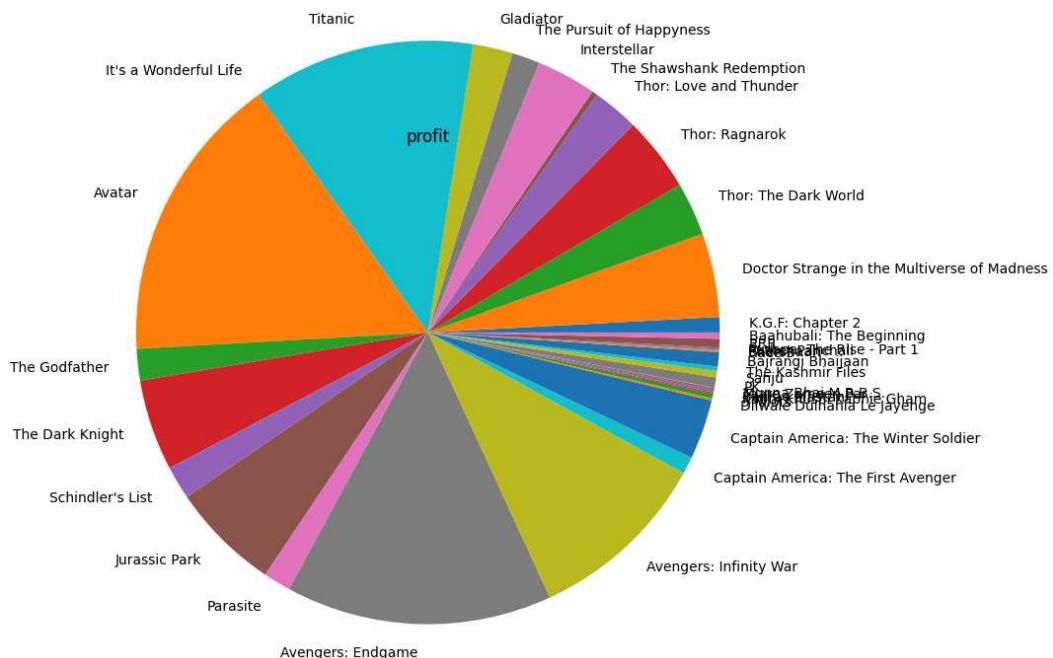
In [992...]

```

from matplotlib import pyplot

plt.pie(df['profit'], labels=df['title'], radius=2)
plt.title(label='profit')
plt.show()

```



In [993...]

```

# count of movies made by individual studios
df.groupby('studio')['title'].count()

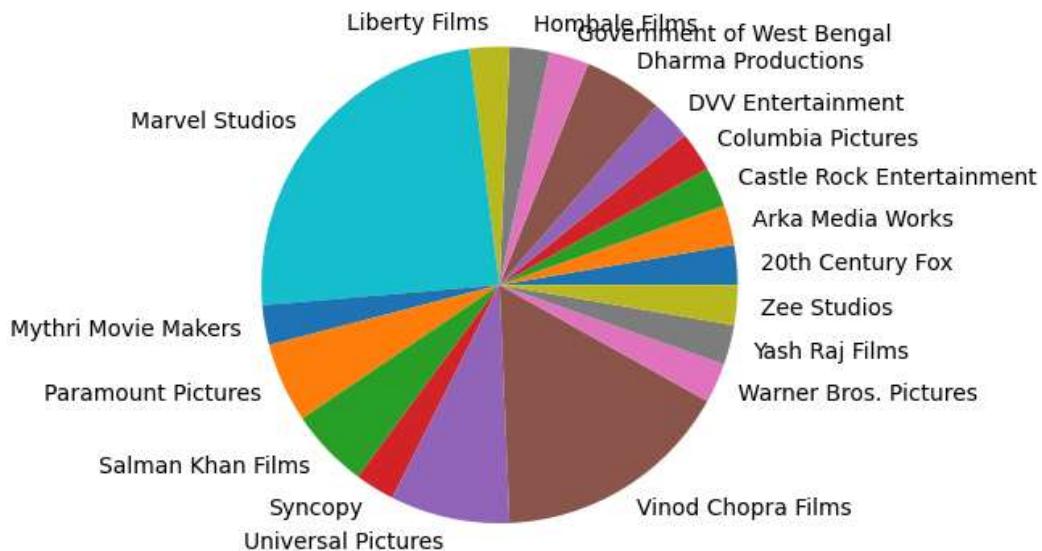
```

Out[993...]

studio	count
20th Century Fox	1
Arka Media Works	1
Castle Rock Entertainment	1
Columbia Pictures	1
DVV Entertainment	1
Dharma Productions	2
Government of West Bengal	1
Hombale Films	1
Liberty Films	1
Marvel Studios	9
Mythri Movie Makers	1
Paramount Pictures	2
Salman Khan Films	2
Syncopy	1
Universal Pictures	3
Vinod Chopra Films	6
Warner Bros. Pictures	1
Yash Raj Films	1
Zee Studios	1

Name: title, dtype: int64

```
In [994... pyplot.pie(df.groupby('studio')['title'].count(), labels=df.groupby('st
| plt.show()
```



```
In [995... # Top 5 Bollywood movies by revenue
df[df.industry == 'Bollywood'].sort_values(by='revenue', ascending=False)
```

	movie_id	title	industry	language	release_year	imdb_rating	
0	101	K.G.F: Chapter 2	Bollywood	Kannada	2022	8.4	
35	133	RRR	Bollywood	Telugu	2022	8.0	Ente
30	136	Bajrangi Bhaijaan	Bollywood	Hindi	2015	8.1	Sal
27	130	PK	Bollywood	Hindi	2014	8.1	Vinc
36	134	Baahubali: The Beginning	Bollywood	Telugu	2015	8.0	A

```
In [996... # Top 5 hollywood movies by revenue
df[df.industry == 'Hollywood'].sort_values(by='revenue', ascending=False)
```

Out[996...]

	movie_id	title	industry	language	release_year	imdb_rating	:
11	119	Avatar	Hollywood	English	2009	7.8	C
17	125	Avengers: Endgame	Hollywood	English	2019	8.4	I S
9	117	Titanic	Hollywood	English	1997	7.9	Para P
18	126	Avengers: Infinity War	Hollywood	English	2018	8.4	I S
15	123	Jurassic Park	Hollywood	English	1993	8.2	Un P

In [997...]

```
# Top 5 movies by revenue
df.sort_values(by='revenue', ascending=False).head(5)
```

Out[997...]

	movie_id	title	industry	language	release_year	imdb_rating	:
11	119	Avatar	Hollywood	English	2009	7.8	C
17	125	Avengers: Endgame	Hollywood	English	2019	8.4	I S
9	117	Titanic	Hollywood	English	1997	7.9	Para P
18	126	Avengers: Infinity War	Hollywood	English	2018	8.4	I S
15	123	Jurassic Park	Hollywood	English	1993	8.2	Un P

In [998...]

```
# Top 3 studio by revenue
df.groupby('studio')['revenue'].sum().sort_values(ascending=False).head
```

Out[998...]

```
studio
Marvel Studios      9317.7
20th Century Fox  2847.0
Paramount Pictures 2493.0
Name: revenue, dtype: float64
```

In [999...]

```
# Top 3 studio by revenue for Bollywood
df[df.industry == 'Bollywood'].groupby('studio')['revenue'].sum().sort_
```

Out[999...]

```
studio
Vinod Chopra Films  284.40
Salman Khan Films   177.48
Hombale Films       150.00
Name: revenue, dtype: float64
```

In [100... df.head()

	movie_id	title	industry	language	release_year	imdb_rating	stu
0	101	K.G.F: Chapter 2	Bollywood	Kannada	2022	8.4	Hom F
1	102	Doctor Strange in the Multiverse of Madness	Hollywood	English	2022	7.0	Ma Stu
2	103	Thor: The Dark World	Hollywood	English	2013	6.8	Ma Stu
3	104	Thor: Ragnarok	Hollywood	English	2017	7.9	Ma Stu
4	105	Thor: Love and Thunder	Hollywood	English	2022	6.8	Ma Stu

4. Visualizations

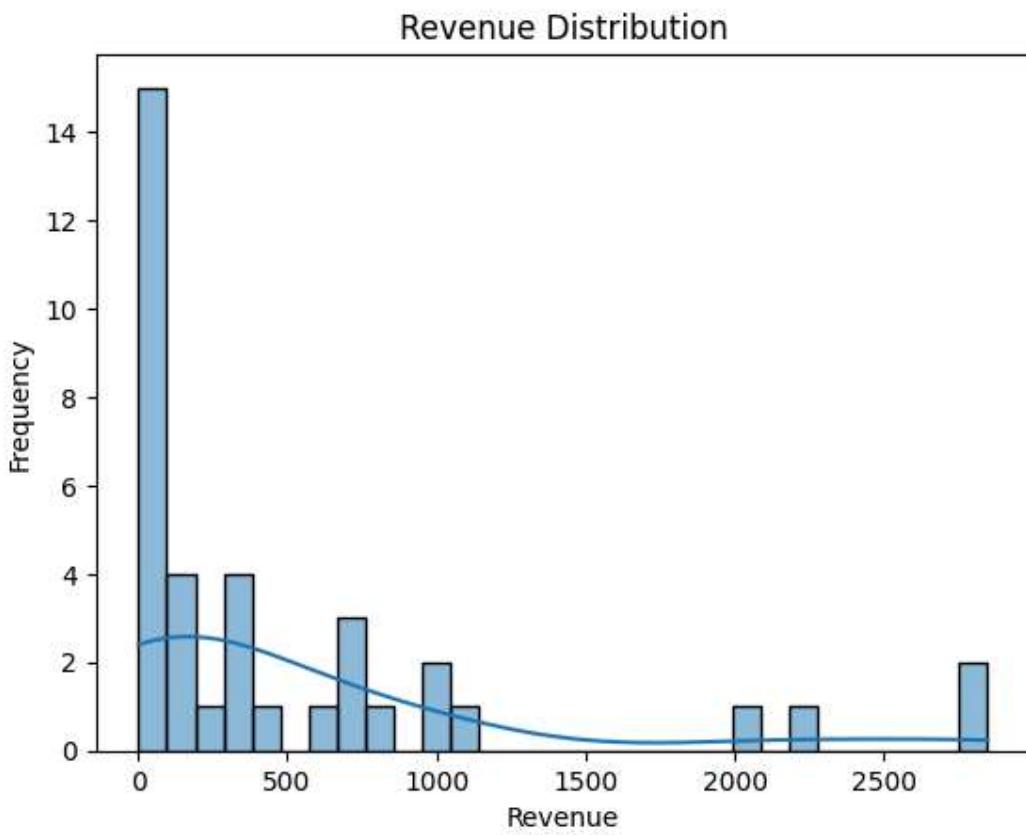
- **Revenue Distribution:** Histogram showing revenue spread across all movies.

In [100... # Revenue distribution

```

sns.histplot(df['revenue'], bins=30, kde=True)
plt.title('Revenue Distribution')
plt.xlabel('Revenue')
plt.ylabel('Frequency')
plt.show()

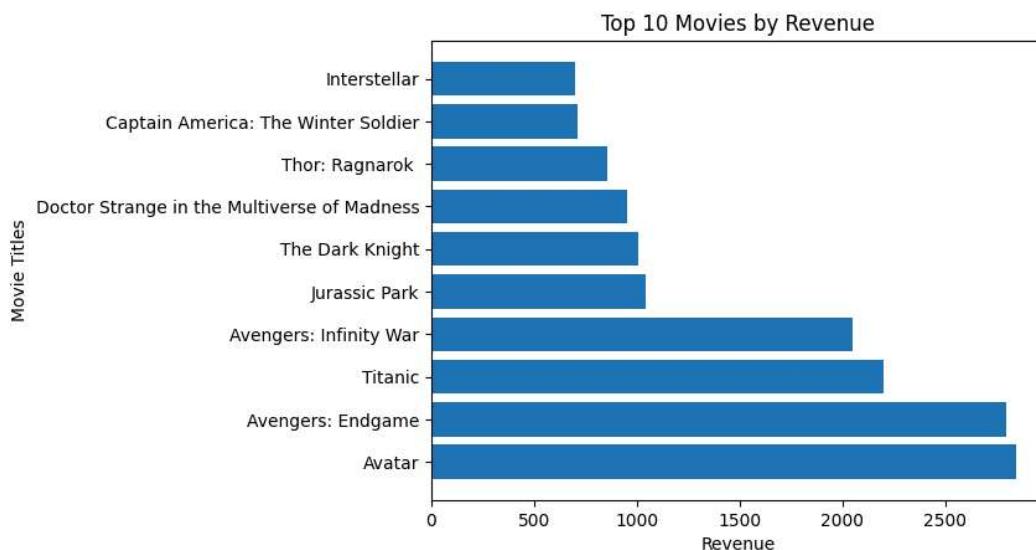
```



- **Top 10 Movies by Revenue:** Bar chart displaying the highest-grossing movies.

```
In [100]: # Top 10 movies by revenue
```

```
top_movies = df.nlargest(10, 'revenue')
plt.barh(top_movies['title'], top_movies['revenue'])
plt.title('Top 10 Movies by Revenue')
plt.xlabel('Revenue')
plt.ylabel('Movie Titles')
plt.show()
```



```
In [100]: df.head()
```

Out[100...]

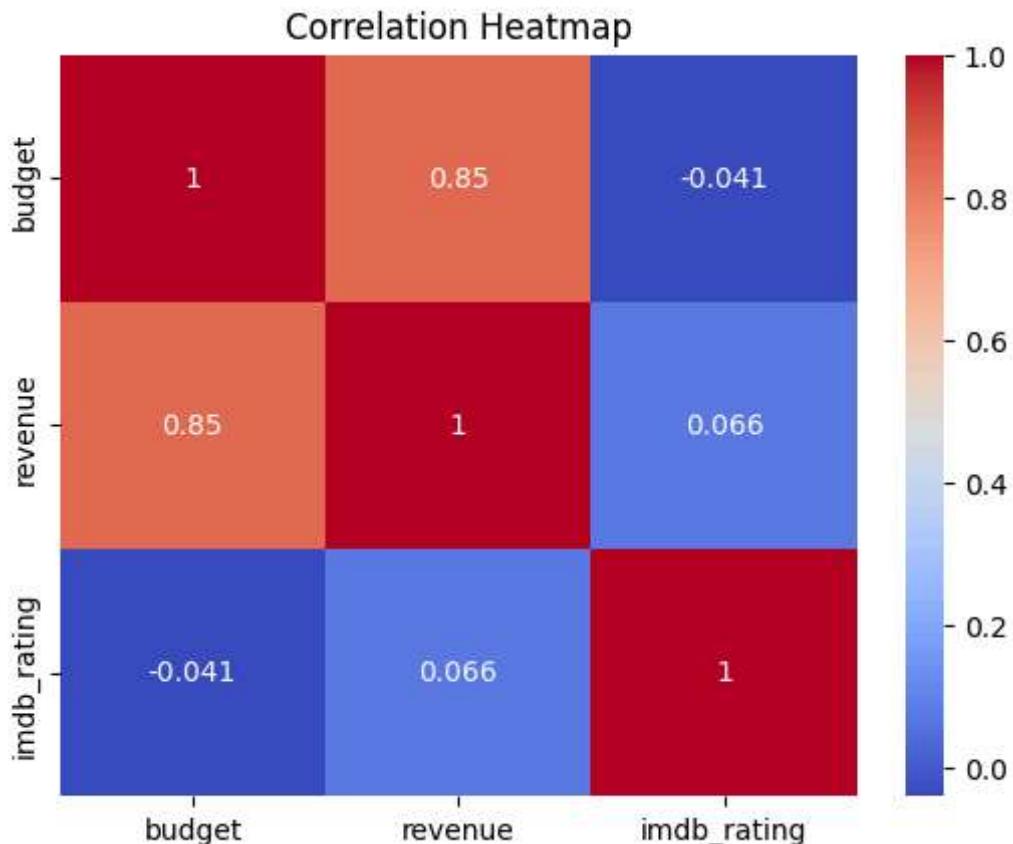
	movie_id	title	industry	language	release_year	imdb_rating	stt
0	101	K.G.F: Chapter 2	Bollywood	Kannada	2022	8.4	Hom F
1	102	Doctor Strange in the Multiverse of Madness	Hollywood	English	2022	7.0	Ma Stu
2	103	Thor: The Dark World	Hollywood	English	2013	6.8	Ma Stu
3	104	Thor: Ragnarok	Hollywood	English	2017	7.9	Ma Stu
4	105	Thor: Love and Thunder	Hollywood	English	2022	6.8	Ma Stu

◀ ▶

- **Correlation Heatmap:** Showcasing relationships between budget, revenue, IMDB Rating, and popularity.

In [100...]

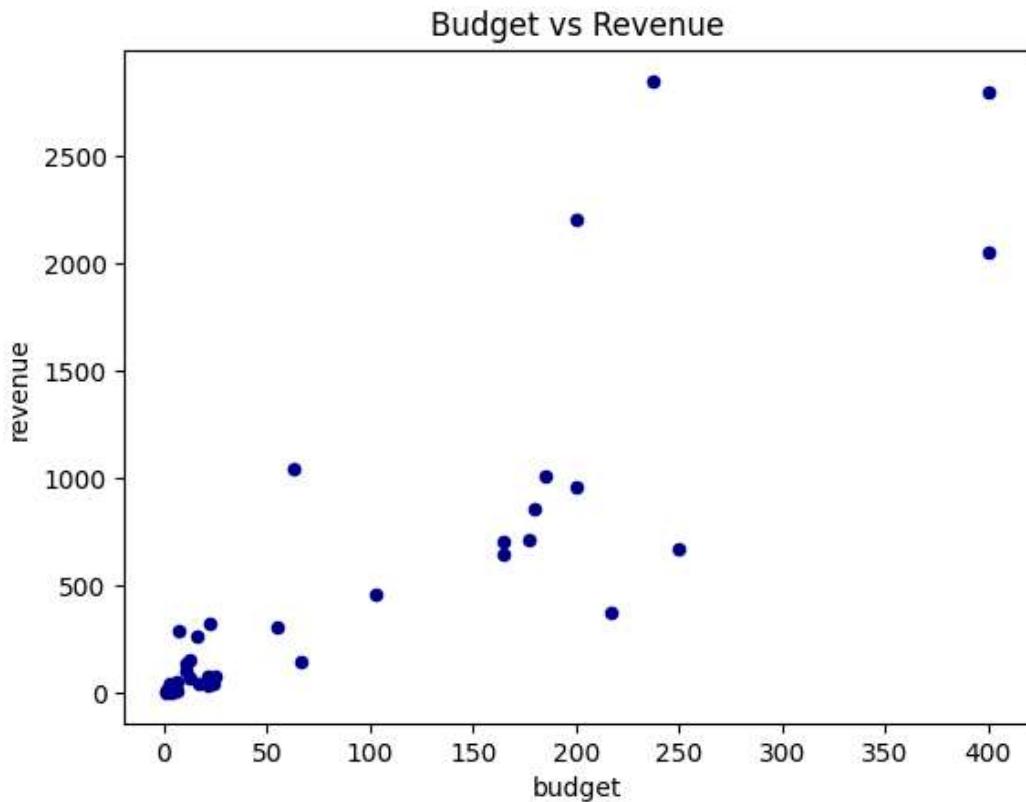
```
numerical_features = df[['budget', 'revenue', 'imdb_rating']]
sns.heatmap(numerical_features.corr(), annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```



- **scatterplot:** scatterplot for showing relation between budget and revenue.

```
In [100]: df.plot.scatter(x='budget', y='revenue', title='Budget vs Revenue', col
```

```
Out[100]: <Axes: title={'center': 'Budget vs Revenue'}, xlabel='budget', ylabel='revenue'>
```



5. Conclusion

- The dataset reveals clear trends in the dominance of **Hollywood**, with **Marvel Studios** producing the highest number of hits.
- Budget and revenue show a strong correlation, indicating higher budgets often lead to greater revenues.
- IMDB ratings provide an essential measure of audience reception, which does not always align with revenue.
- Marvel studio of Hollywood and Vinod Chopra Films of Bollywood produced maximum movies.
- 20th Century Fox of Hollywood and Hombale Films of Bollywood making highest revenue movie
- Marvel studio of Hollywood and Vinod Chopra Films of Bollywood has highest average revenue.