

imdb

September 22, 2024

1 Data Loading and Preprocessing

```
[1]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
from datetime import datetime as dt
from wordcloud import WordCloud
```

```
[2]: imdb=pd.read_csv('imdb-movies-dataset.csv')
```

```
[3]: imdb.shape
```

```
[3]: (10000, 17)
```

```
[4]: imdb.head()
```

```
[4]:                                     Poster \
0  https://m.media-amazon.com/images/M/MV5BYWRkZj...
1  https://m.media-amazon.com/images/M/MV5BZGI4NT...
2  https://m.media-amazon.com/images/M/MV5BZjlyOT...
3  https://m.media-amazon.com/images/M/MV5BMjA5Zj...
4  https://m.media-amazon.com/images/M/MV5BNTk1MT...
```

	Title	Release Date	Budget	Revenue	\
0	The Idea of You	16-07-2022	42110861	4886739	
1	Kingdom of the Planet of the Apes	12-05-2023	14627858	7458541	
2	Unfrosted	20-04-2023	48805376	7814085	
3	The Fall Guy	07-10-2022	28147070	4951558	
4	Challengers	05-01-2023	49549901	6891688	

	Certificate	Duration (min)	Genre	Rating	Metascore	\
0	R	115.0	Comedy, Drama, Romance	6.4	67.0	
1	PG-13	145.0	Action, Adventure, Sci-Fi	7.3	66.0	
2	PG-13	97.0	Biography, Comedy, History	5.5	42.0	
3	PG-13	126.0	Action, Comedy, Drama	7.3	73.0	
4	R	131.0	Drama, Romance, Sport	7.7	82.0	

	Director	Cast \
0	Michael Showalter	Anne Hathaway, Nicholas Galitzine, Ella Rubin,...
1	Wes Ball	Owen Teague, Freya Allan, Kevin Durand, Peter ...
2	Jerry Seinfeld	Isaac Bae, Jerry Seinfeld, Chris Rickett, Rach...
3	David Leitch	Ryan Gosling, Emily Blunt, Aaron Taylor-Johnso...
4	Luca Guadagnino	Zendaya, Mike Faist, Josh O'Connor, Darnell Ap...

	Votes	Description	Review Count \
0	28,744	Solène, a 40-year-old single mom, begins an un...	166
1	22,248	Many years after the reign of Caesar, a young ...	183
2	18,401	In 1963 Michigan, business rivals Kellogg's an...	333
3	38,953	A down-and-out stuntman must find the missing ...	384
4	32,517	Tashi, a former tennis prodigy turned coach, t...	194

	Review Title \
0	Hypocrisy as an idea
1	A phenomenal start to another trilogy!
2	not funny
3	Everything you needed and more!
4	Watch "Match Point" instead

	Review
0	This film, as well as the reaction to it, is a...
1	I'm a big fan of all the planet of the apes, a...
2	Pretty much the worst criticism you can lay on...
3	Just got out of the Austin premier at SXSW and...
4	This is a tough one. I liked the concept and t...

```
[5]: imdb.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 17 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Poster          10000 non-null  object
1   Title           10000 non-null  object
2   Release Date    10000 non-null  object
3   Budget          10000 non-null  int64
4   Revenue         10000 non-null  int64
5   Certificate      7370 non-null   object
6   Duration (min)  9664 non-null   float64
7   Genre           9993 non-null   object
8   Rating          9596 non-null   float64
9   Metascore       7555 non-null   float64
10  Director        9995 non-null   object
11  Cast            9961 non-null   object
12  Votes           9596 non-null   object
```

```

13 Description      10000 non-null  object
14 Review Count    9999 non-null  object
15 Review Title     9483 non-null  object
16 Review          9484 non-null  object
dtypes: float64(3), int64(2), object(12)
memory usage: 1.3+ MB

```

```
[6]: imdb.isnull().sum()
```

```

[6]: Poster          0
    Title            0
    Release Date     0
    Budget           0
    Revenue          0
    Certificate      2630
    Duration (min)   336
    Genre            7
    Rating           404
    Metascore       2445
    Director         5
    Cast             39
    Votes            404
    Description      0
    Review Count     1
    Review Title     517
    Review           516
dtype: int64

```

```
[7]: imdb['Release Date']=pd.to_datetime(imdb['Release Date'],format="%d-%m-%Y")
```

```
[8]: imdb['Votes']=imdb['Votes'].str.replace(',','')
```

```
[9]: imdb['Votes']=imdb['Votes'].astype(float)
```

```

[10]: missing_values=imdb.isna().sum()
      missing_percentage=round((missing_values/len(imdb))*100,2)
      missing_data=pd.DataFrame({'Missing Values':missing_values,'Missing Percentage':
      ↪missing_percentage}).loc[missing_values>0]
      missing_data.reset_index().rename(columns={'index':'Columns_missing'})

```

```

[10]: Columns_missing  Missing Values  Missing Percentage
0      Certificate      2630          26.30
1  Duration (min)      336           3.36
2           Genre        7           0.07
3          Rating      404           4.04
4      Metascore     2445          24.45
5      Director        5           0.05
6           Cast       39           0.39

```

7	Votes	404	4.04
8	Review Count	1	0.01
9	Review Title	517	5.17
10	Review	516	5.16

```
[11]: cols_to_remove=[]
cols_null_values=[]
cols_to_replace=[]
for index, value in enumerate(missing_data['Missing Percentage']):
    if value>25:
        cols_to_remove.append(missing_data.index[index])
    elif value<1:
        cols_null_values.append(missing_data.index[index])
    else:
        cols_to_replace.append(missing_data.index[index])
```

```
[12]: cols_to_remove
```

```
[12]: ['Certificate']
```

```
[13]: for i in cols_to_remove:
        imdb.drop(columns=i,inplace=True)
```

```
[14]: for i in cols_null_values:
        imdb.dropna(subset=i,inplace=True)
```

```
[15]: cols_to_replace
```

```
[15]: ['Duration (min)', 'Rating', 'Metascore', 'Votes', 'Review Title', 'Review']
```

```
[16]: average_duration=imdb['Duration (min)'].mean()
average_duration=round(average_duration,0)
imdb['Duration (min)']=imdb['Duration (min)'].fillna(average_duration)
```

```
[17]: average_rating=imdb['Rating'].mean()
average_rating=round(average_rating,1)
imdb['Rating']=imdb['Rating'].fillna(average_rating)
```

```
[18]: average_votes=imdb['Votes'].mean()
average_votes=round(average_votes,0)
imdb['Votes']=imdb['Votes'].fillna(average_votes)
```

```
[19]: imdb['Review Title']=imdb['Review Title'].fillna('N/A')
imdb['Review']=imdb['Review'].fillna('N/A')
```

```
[20]: imdb['Metascore']=imdb['Metascore'].fillna(imdb['Metascore'].mean())
```

```
[21]: imdb.isnull().sum()
```

```
[21]: Poster          0
      Title           0
      Release Date    0
      Budget          0
      Revenue         0
      Duration (min)  0
      Genre           0
      Rating          0
      Metascore       0
      Director        0
      Cast            0
      Votes           0
      Description     0
      Review Count    0
      Review Title    0
      Review          0
      dtype: int64
```

```
[22]: imdb.shape
```

```
[22]: (9954, 16)
```

```
[23]: imdb.T
```

```
[23]:
      0 \
Poster      https://m.media-amazon.com/images/M/MV5BYWRkZj...
Title              The Idea of You
Release Date      2022-07-16 00:00:00
Budget              42110861
Revenue           4886739
Duration (min)    115.0
Genre              Comedy, Drama, Romance
Rating              6.4
Metascore          67.0
Director           Michael Showalter
Cast              Anne Hathaway, Nicholas Galitzine, Ella Rubin,...
Votes              28744.0
Description        Solène, a 40-year-old single mom, begins an un...
Review Count       166
Review Title       Hypocrisy as an idea
Review             This film, as well as the reaction to it, is a...

      1 \
Poster      https://m.media-amazon.com/images/M/MV5BZGI4NT...
Title              Kingdom of the Planet of the Apes
Release Date      2023-05-12 00:00:00
Budget              14627858
```

Revenue	7458541
Duration (min)	145.0
Genre	Action, Adventure, Sci-Fi
Rating	7.3
Metascore	66.0
Director	Wes Ball
Cast	Owen Teague, Freya Allan, Kevin Durand, Peter ...
Votes	22248.0
Description	Many years after the reign of Caesar, a young ...
Review Count	183
Review Title	A phenomenal start to another trilogy!
Review	I'm a big fan of all the planet of the apes, a...

2 \

Poster	https://m.media-amazon.com/images/M/MV5BZjIyOT...
Title	Unfrosted
Release Date	2023-04-20 00:00:00
Budget	48805376
Revenue	7814085
Duration (min)	97.0
Genre	Biography, Comedy, History
Rating	5.5
Metascore	42.0
Director	Jerry Seinfeld
Cast	Isaac Bae, Jerry Seinfeld, Chris Rickett, Rach...
Votes	18401.0
Description	In 1963 Michigan, business rivals Kellogg's an...
Review Count	333
Review Title	not funny
Review	Pretty much the worst criticism you can lay on...

3 \

Poster	https://m.media-amazon.com/images/M/MV5BMjA5Zj...
Title	The Fall Guy
Release Date	2022-10-07 00:00:00
Budget	28147070
Revenue	4951558
Duration (min)	126.0
Genre	Action, Comedy, Drama
Rating	7.3
Metascore	73.0
Director	David Leitch
Cast	Ryan Gosling, Emily Blunt, Aaron Taylor-Johnso...
Votes	38953.0
Description	A down-and-out stuntman must find the missing ...
Review Count	384
Review Title	Everything you needed and more!

Review	Just got out of the Austin premier at SXSW and...	4	\
Poster	https://m.media-amazon.com/images/M/MV5BNTk1MT...		
Title	Challengers		
Release Date	2023-01-05 00:00:00		
Budget	49549901		
Revenue	6891688		
Duration (min)	131.0		
Genre	Drama, Romance, Sport		
Rating	7.7		
Metascore	82.0		
Director	Luca Guadagnino		
Cast	Zendaya, Mike Faist, Josh O'Connor, Darnell Ap...		
Votes	32517.0		
Description	Tashi, a former tennis prodigy turned coach, t...		
Review Count	194		
Review Title	Watch "Match Point" instead		
Review	This is a tough one. I liked the concept and t...		
		5	\
Poster	https://m.media-amazon.com/images/M/MV5BY2JlYj...		
Title	Abigail		
Release Date	2023-02-25 00:00:00		
Budget	52068286		
Revenue	4253458		
Duration (min)	109.0		
Genre	Horror, Thriller		
Rating	6.8		
Metascore	62.0		
Director	Matt Bettinelli-Olpin		
Cast	Tyler Gillett, Melissa Barrera, Dan Stevens, A...		
Votes	27284.0		
Description	After a group of criminals kidnap the ballerin...		
Review Count	168		
Review Title	Underwhelming and Unimpressive		
Review	Everything in this movie is mediocre except th...		
		6	\
Poster	https://m.media-amazon.com/images/M/MV5BYTYyOD...		
Title	Civil War		
Release Date	2023-02-26 00:00:00		
Budget	46136436		
Revenue	4499539		
Duration (min)	109.0		
Genre	Action, Adventure, Thriller		
Rating	7.5		

Metascore	75.0
Director	Alex Garland
Cast	Kirsten Dunst, Wagner Moura, Cailee Spaeny, Ni...
Votes	64014.0
Description	A journey across a dystopian future America, f...
Review Count	610
Review Title	Meander: The Movie
Review	I am huge fan of writer/director Alex Garland...

7 \

Poster	https://m.media-amazon.com/images/M/MV5BZTc3Ym...
Title	Twisters
Release Date	2023-05-14 00:00:00
Budget	23150357
Revenue	4862696
Duration (min)	109.0
Genre	Action, Adventure, Thriller
Rating	6.4
Metascore	58.065537
Director	Lee Isaac Chung
Cast	Daisy Edgar-Jones, Glen Powell, David Corensw...
Votes	96422.0
Description	An update to the 1996 film 'Twister', which ce...
Review Count	0
Review Title	N/A
Review	N/A

8 \

Poster	https://m.media-amazon.com/images/M/MV5B0TVhZG...
Title	Anyone But You
Release Date	2022-06-11 00:00:00
Budget	14042912
Revenue	5006926
Duration (min)	103.0
Genre	Comedy, Romance
Rating	6.1
Metascore	52.0
Director	Will Gluck
Cast	Sydney Sweeney, Glen Powell, Alexandra Shipp, ...
Votes	82215.0
Description	After an amazing first date, Bea and Ben's fie...
Review Count	373
Review Title	Pure escapism featuring beautiful people, beau...
Review	This was like an R rated raunchy Hallmark movi...

9 ... \

Poster	https://m.media-amazon.com/images/M/MV5B0WI5YT... ...
--------	---

Title	The Ministry of Ungentlemanly Warfare	...
Release Date	2022-09-06 00:00:00	...
Budget	20939343	...
Revenue	6998497	...
Duration (min)	120.0	...
Genre	Action, Drama, War	...
Rating	7.0	...
Metascore	57.0	...
Director	Guy Ritchie	...
Cast	Henry Cavill, Alan Ritchson, Alex Pettyfer, Ei...	...
Votes	21084.0	...
Description	The British military recruits a small group of...	...
Review Count	117	...
Review Title	Fun Movie, fun cast!	...
Review	As a fan of both Henry Cavill and Guy Ritchie,...	...

9990 \

Poster	https://m.media-amazon.com/images/M/MV5BNjFjYm...
Title	The Masque of the Red Death
Release Date	2022-05-08 00:00:00
Budget	19612431
Revenue	5643421
Duration (min)	89.0
Genre	Drama, Horror
Rating	6.9
Metascore	77.0
Director	Roger Corman
Cast	Vincent Price, Hazel Court, Jane Asher, David ...
Votes	16302.0
Description	A European prince terrorizes the local peasant...
Review Count	109
Review Title	Price at his Most Wicked!
Review	Roger Corman has done an outstanding job with ...

9991 \

Poster	https://m.media-amazon.com/images/M/MV5BMjFkZG...
Title	Beyond the Law
Release Date	2022-09-20 00:00:00
Budget	54592464
Revenue	6940413
Duration (min)	89.0
Genre	Action, Crime, Thriller
Rating	3.7
Metascore	58.065537
Director	James Cullen Bressack
Cast	Johnny Messner, Steven Seagal, DMX, Bill Cobbs
Votes	3175.0

Description	A mobster trying to go straight and a dogged p...
Review Count	54
Review Title	Snoozefest
Review	Absolutely awful, Steven segal needs to stop m...
	9992 \
Poster	https://m.media-amazon.com/images/M/MV5BMTUzOT...
Title	Kiss of the Spider Woman
Release Date	2022-08-16 00:00:00
Budget	29288225
Revenue	7076615
Duration (min)	120.0
Genre	Drama
Rating	7.3
Metascore	80.0
Director	Hector Babenco
Cast	William Hurt, Raul Julia, Sonia Braga, José Le...
Votes	17378.0
Description	A gay man and a political prisoner are togethe...
Review Count	57
Review Title	"Film within a film" really works
Review	Reading the other reviews, I see that homophob...
	9993 \
Poster	https://m.media-amazon.com/images/M/MV5BMzRkNj...
Title	South Central
Release Date	2022-12-06 00:00:00
Budget	38200271
Revenue	7401048
Duration (min)	98.0
Genre	Crime, Drama
Rating	6.8
Metascore	64.0
Director	Stephen Milburn Anderson
Cast	Glenn Plummer, Byron Minns, Lexie Bigham, Vinc...
Votes	6730.0
Description	A former gangster fresh out of prison wants to...
Review Count	20
Review Title	"South Central was adapted from my novel "Sout...
Review	I would like to thank all of the fans and supp...
	9994 \
Poster	https://m.media-amazon.com/images/M/MV5BNmRhZW...
Title	Mutiny on the Bounty
Release Date	2022-06-07 00:00:00
Budget	15249672
Revenue	7429611

Duration (min)	132.0
Genre	Adventure, Biography, Drama
Rating	7.6
Metascore	87.0
Director	Frank Lloyd
Cast	Charles Laughton, Clark Gable, Franchot Tone, ...
Votes	24899.0
Description	First mate Fletcher Christian leads a revolt a...
Review Count	112
Review Title	Our Favorite "Mutiny": April 28, 1789
Review	Although the versions with Marlon Brando and T...

9995 \

Poster	https://m.media-amazon.com/images/M/MV5BMzg5MW...
Title	The Greatest Show on Earth
Release Date	2023-02-11 00:00:00
Budget	40892520
Revenue	5496436
Duration (min)	152.0
Genre	Drama, Family, Romance
Rating	6.5
Metascore	76.0
Director	Cecil B. DeMille
Cast	James Stewart, Charlton Heston, Betty Hutton, ...
Votes	16078.0
Description	The dramatic lives of trapeze artists, a clown...
Review Count	128
Review Title	Hey, doesn't anyone remember Last Emperor?
Review	It constantly amazes me that people carp that ...

9996 \

Poster	https://m.media-amazon.com/images/M/MV5BYzA0ZG...
Title	Berserk: Ougon Jidai-hen I - Haou no Tamago
Release Date	2022-08-19 00:00:00
Budget	29174237
Revenue	5128209
Duration (min)	76.0
Genre	Animation, Action, Adventure
Rating	7.5
Metascore	58.065537
Director	Toshiyuki Kubooka
Cast	Hiroaki Iwanaga, Carrie Keranen, Takahiro Saku...
Votes	14300.0
Description	A lone sellsword named Guts gets recruited int...
Review Count	12
Review Title	Masterfully directed climatic epic saga
Review	Few stories can capture your mind and soul in ...

	9997 \
Poster	https://m.media-amazon.com/images/M/MV5BM2U1Mj...
Title	Is-slottet
Release Date	2023-05-06 00:00:00
Budget	39411922
Revenue	6770819
Duration (min)	78.0
Genre	Mystery, Drama
Rating	6.5
Metascore	58.065537
Director	Per Blom
Cast	Line Storesund, Hilde Nyeggen Martinsen, Meret...
Votes	740.0
Description	A couple of twelve-year-old Norwegian girls st...
Review Count	4
Review Title	Beautiful Film
Review	This film might not be to everyone's taste, it...
	9998 \
Poster	https://m.media-amazon.com/images/M/MV5BMTAwOD...
Title	Loving Pablo
Release Date	2022-12-29 00:00:00
Budget	49927298
Revenue	5635966
Duration (min)	123.0
Genre	Biography, Crime, Drama
Rating	6.4
Metascore	42.0
Director	Fernando León de Aranoa
Cast	Javier Bardem, Penélope Cruz, Peter Sarsgaard,...
Votes	22447.0
Description	A journalist strikes up a romantic relationshi...
Review Count	84
Review Title	That film should be in Spanish
Review	Why anyone (the director?) made Spanish actors...
	9999
Poster	https://m.media-amazon.com/images/M/MV5BMmE4ZD...
Title	Un homme et une femme
Release Date	2022-05-22 00:00:00
Budget	35426947
Revenue	5634048
Duration (min)	102.0
Genre	Drama, Romance
Rating	7.5
Metascore	58.065537

```

Director          Claude Lelouch
Cast              Anouk Aimée, Jean-Louis Trintignant, Pierre Ba...
Votes              11826.0
Description       A widow and a widower find their relationship ...
Review Count      61
Review Title      Moved me tremendously.
Review            This is a movie that resonated to my core. I ...

```

[16 rows x 9954 columns]

2 Data Exploration

```
[24]: imdb.describe()
```

```

[24]:
count          Release Date          Budget          Revenue \
mean    2022-09-14 07:55:04.882459136  2.818803e+07  5.993786e+06
min      2016-10-02 00:00:00  1.004975e+06  4.001122e+06
25%      2022-07-02 00:00:00  1.440842e+07  4.996012e+06
50%      2022-10-10 00:00:00  2.801594e+07  6.010436e+06
75%      2023-01-20 00:00:00  4.202331e+07  6.983931e+06
max      2023-05-19 00:00:00  5.555545e+07  7.998826e+06
std              NaN  1.578142e+07  1.151750e+06

count    Duration (min)    Rating    Metascore    Votes
mean      108.803195      6.436940    58.065537  9.642243e+04
min        45.000000      1.300000     1.000000  6.000000e+00
25%        95.000000      5.900000    50.000000  1.205025e+04
50%       105.000000      6.500000    58.065537  3.978250e+04
75%       118.000000      7.200000    67.000000  9.686500e+04
max       746.000000      9.700000   100.000000  2.894940e+06
std        22.762499      1.031534    15.543383  1.786859e+05

```

```

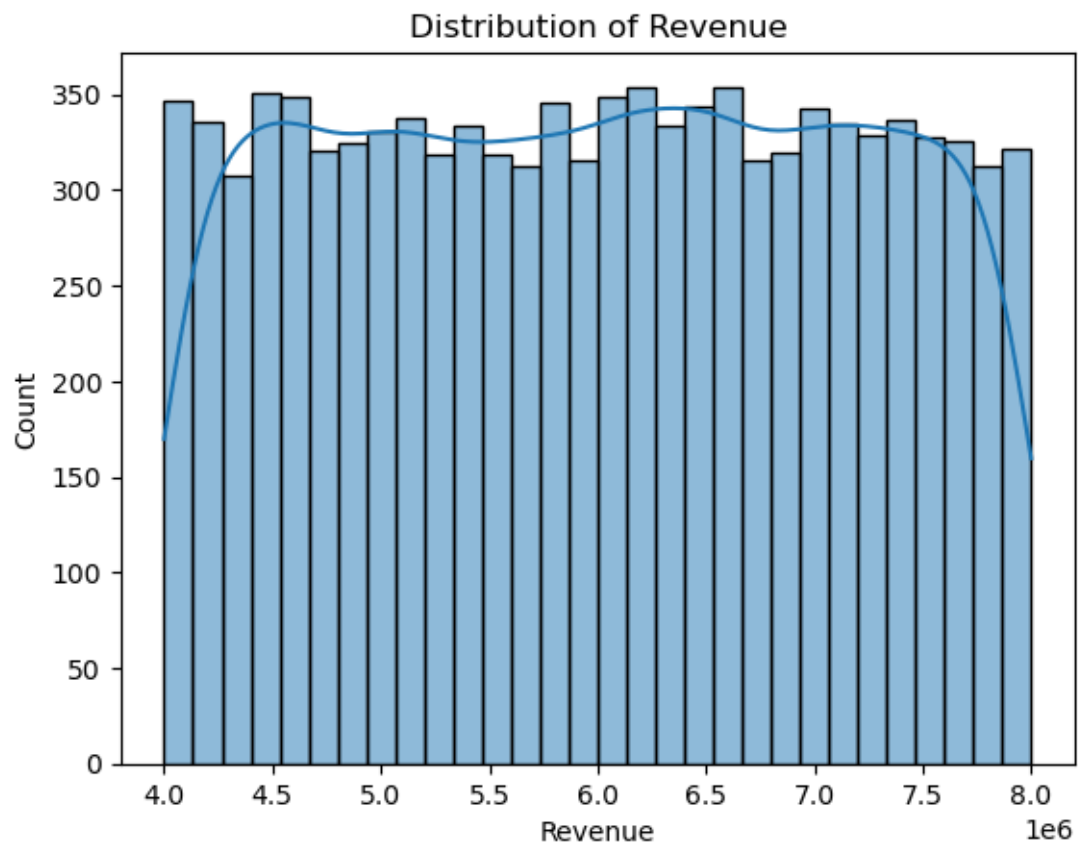
[25]: import warnings
warnings.simplefilter(action='ignore', category=FutureWarning)

```

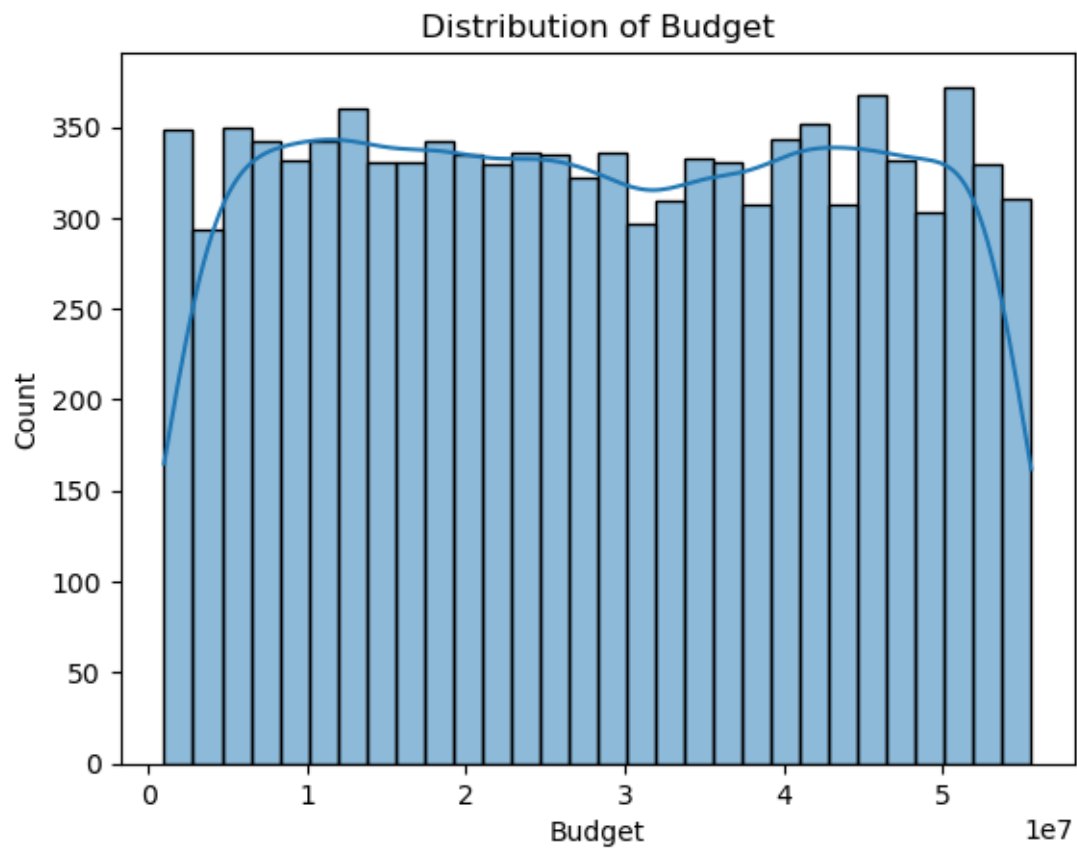
```

[26]: sns.histplot(imdb['Revenue'],bins=30,kde=True)
plt.title('Distribution of Revenue')
plt.show()

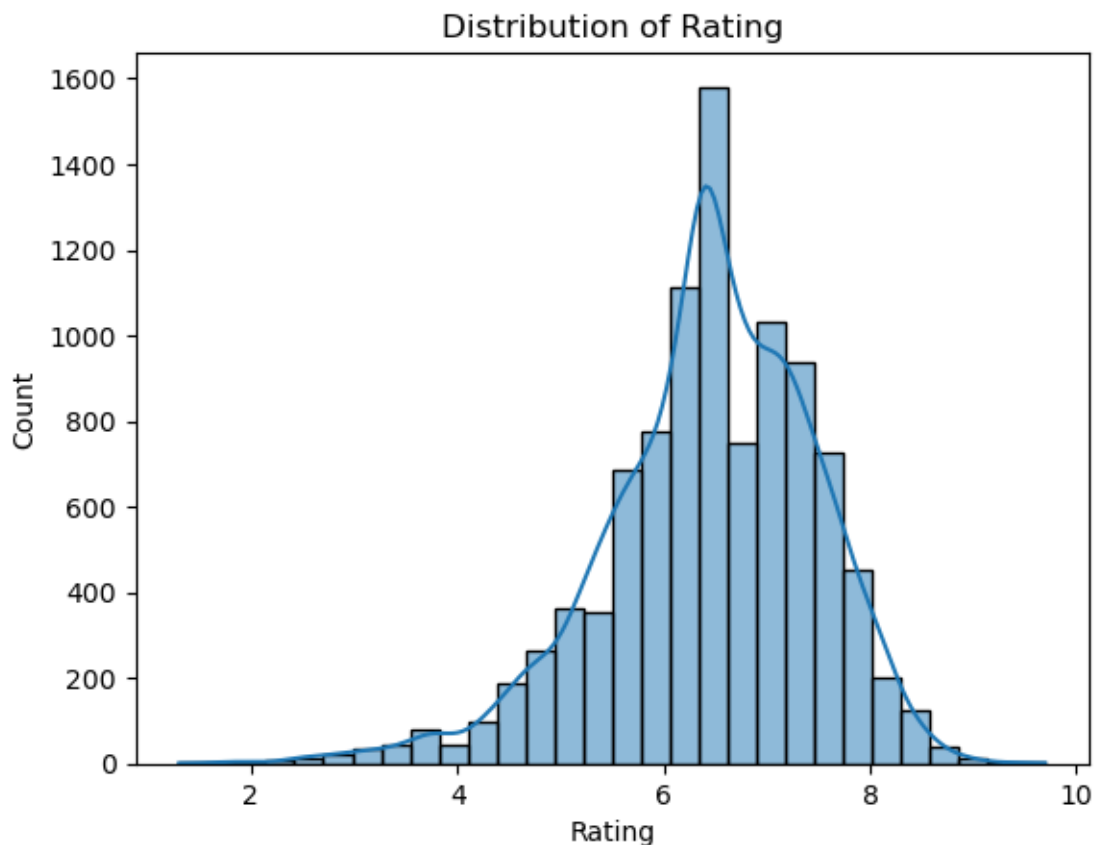
```



```
[27]: sns.histplot(imdb['Budget'],bins=30,kde=True)
plt.title('Distribution of Budget')
plt.show()
```



```
[28]: sns.histplot(imdb['Rating'],bins=30,kde=True)
plt.title('Distribution of Rating')
plt.show()
```



```
[29]: numerical_data=imdb.select_dtypes('number')
```

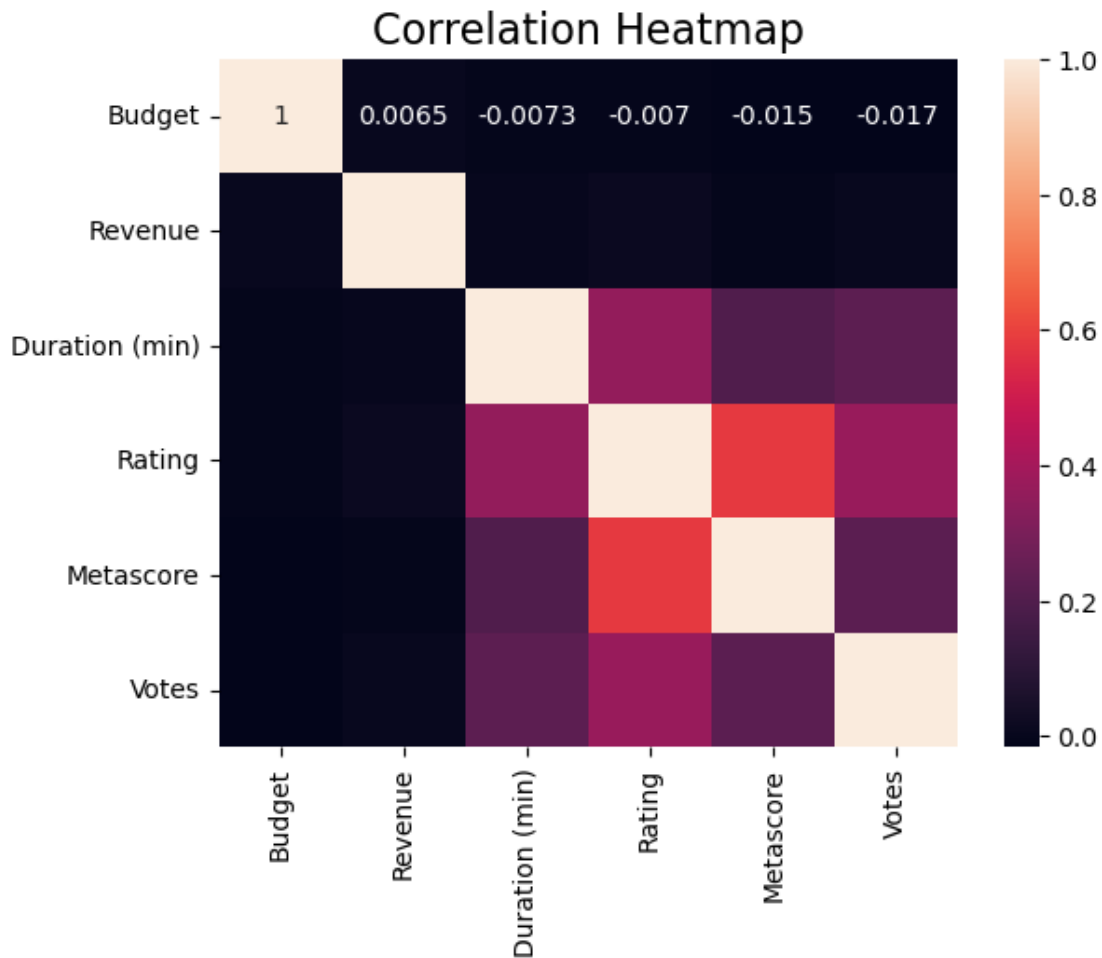
```
[30]: numerical_correlation=numerical_data.corr()
numerical_correlation
```

```
[30]:
```

	Budget	Revenue	Duration (min)	Rating	Metascore	\
Budget	1.000000	0.006483	-0.007314	-0.006967	-0.015496	
Revenue	0.006483	1.000000	0.003117	0.011716	-0.005948	
Duration (min)	-0.007314	0.003117	1.000000	0.360126	0.195864	
Rating	-0.006967	0.011716	0.360126	1.000000	0.579345	
Metascore	-0.015496	-0.005948	0.195864	0.579345	1.000000	
Votes	-0.016578	0.004242	0.229547	0.369326	0.223838	

	Votes
Budget	-0.016578
Revenue	0.004242
Duration (min)	0.229547
Rating	0.369326
Metascore	0.223838
Votes	1.000000


```
[31]: sns.heatmap(numerical_correlation,annot=True,)
plt.title('Correlation Heatmap', fontsize=16)
plt.show()
```



```
[32]: imdb['Genre1']=imdb['Genre'].str.split(',').str[0]
```

```
[33]: imdb['Genre1'].unique()
```

```
[33]: array(['Comedy', 'Action', 'Biography', 'Drama', 'Horror', 'Adventure',
        'Crime', 'Animation', 'Documentary', 'Thriller', 'Fantasy',
        'Mystery', 'Film-Noir', 'Romance', 'Music', 'Sci-Fi', 'Western',
        'Sport', 'War', 'Musical', 'History', 'Family'], dtype=object)
```

```
[34]: genre_revenue=imdb.groupby('Genre1')['Revenue'].
      .agg(Total_Revenue='sum',Average_Revenue='mean').reset_index()
pd.options.display.float_format = '{:,.1f}'.format
genre_revenue
```

```
[34]:
```

	Genre1	Total_Revenue	Average_Revenue
0	Action	13896236809	5992340.2
1	Adventure	3258550033	5924636.4
2	Animation	2882155978	6029615.0
3	Biography	3294244119	6000444.7
4	Comedy	13233813060	5999008.6
5	Crime	4165130365	6018974.5
6	Documentary	696141807	5949930.0
7	Drama	12742140461	5976613.7
8	Family	25397385	6349346.2
9	Fantasy	380465016	6039127.2
10	Film-Noir	30700824	6140164.8
11	History	11192281	5596140.5
12	Horror	3967725655	6020828.0
13	Music	10604287	5302143.5
14	Musical	43527294	7254549.0
15	Mystery	291235608	6196502.3
16	Romance	133296367	6058925.8
17	Sci-Fi	135185143	5877614.9
18	Sport	6923418	6923418.0
19	Thriller	349052486	6018146.3
20	War	4502911	4502911.0
21	Western	103924086	5773560.3

```
[35]: Top5Gross=imdb
Top5Gross.sort_values(by='Revenue',ascending=False).head(5)
```

```
[35]:
```

	Poster \
5932	https://m.media-amazon.com/images/M/MV5BZmZiYT...
8053	https://m.media-amazon.com/images/M/MV5BMzc4Nz...
2986	https://m.media-amazon.com/images/M/MV5BZDVkMW...
6741	https://m.media-amazon.com/images/M/MV5BMTg3OT...
2345	https://m.media-amazon.com/images/M/MV5BMTMyNT...

	Title	Release Date	Budget \
5932	The Stepford Wives	2023-03-18	26628571
8053	Spacehunter: Adventures in the Forbidden Zone	2022-04-03	29418808
2986	The Hate U Give	2023-03-03	48049780
6741	The Love Guru	2022-11-26	43585827
2345	Ocean's Thirteen	2023-02-04	40997329

	Revenue	Duration (min)	Genre	Rating	Metascore \
5932	7998826	115.0	Horror, Mystery, Sci-Fi	6.9	54.0
8053	7998531	90.0	Action, Adventure, Sci-Fi	5.4	58.1
2986	7997909	133.0	Crime, Drama	7.5	81.0
6741	7997697	87.0	Comedy, Romance, Sport	3.8	24.0
2345	7997178	122.0	Crime, Thriller	6.9	62.0

	Director	Cast \
5932	Bryan Forbes	Katharine Ross, Paula Prentiss, Peter Masterso...
8053	Lamont Johnson	Peter Strauss, Molly Ringwald, Ernie Hudson, A...
2986	George Tillman Jr.	Amandla Stenberg, Regina Hall, Russell Hornsby...
6741	Marco Schnabel	Mike Myers, Jessica Alba, Romany Malco, Jessic...
2345	Steven Soderbergh	George Clooney, Brad Pitt, Matt Damon, Michael...

	Votes	Description	Review Count \
5932	19605.0	Joanna Eberhart has come to the quaint little ...	130
8053	6637.0	On a distant planet inhabited by mutants, two ...	68
2986	40226.0	Starr witnesses the fatal shooting of her chil...	359
6741	55013.0	Pitka, an American raised outside of his count...	263
2345	367886.0	Danny Ocean rounds up the boys for a third hei...	268

	Review Title \
5932	"Like one of those robots in Disneyland..."
8053	Like Gold To A 12-Year-Old
2986	Just not a good film
6741	R.I.P. Mike Myers Career
2345	Really good

	Review	Genre1
5932	"The Stepford Wives" certainly isn't the great...	Horror
8053	Back when I was 12, renting this movie was lik...	Action
2986	Nothing in this film was new, it was just reha...	Crime
6741	I saw this film last night in a advance screen...	Comedy
2345	They are back and now there are 13...The movie...	Crime

The analysis of the top 5 highest-grossing movies reveals that a combination of genre diversity, directorial influence, audience ratings, and budget management significantly affects their box office performance. While higher budgets and star directors can attract viewers, the quality of storytelling and audience engagement ultimately play a crucial role in determining a movie's success.

```
[36]: imdb['Month']=imdb['Release Date'].dt.month
```

```
[37]: imdb['Quarter']=imdb['Release Date'].dt.to_period("Q")
```

```
[38]: imdb
```

```
[38]:
Poster \
0      https://m.media-amazon.com/images/M/MV5BYWRkZj...
1      https://m.media-amazon.com/images/M/MV5BZGI4NT...
2      https://m.media-amazon.com/images/M/MV5BZjlyOT...
3      https://m.media-amazon.com/images/M/MV5BMjA5Zj...
4      https://m.media-amazon.com/images/M/MV5BNTk1MT...
...
9995  https://m.media-amazon.com/images/M/MV5BMzg5MW...
```

9996 <https://m.media-amazon.com/images/M/MV5BYzA0ZG...>
 9997 <https://m.media-amazon.com/images/M/MV5BM2U1Mj...>
 9998 <https://m.media-amazon.com/images/M/MV5BMTAwOD...>
 9999 <https://m.media-amazon.com/images/M/MV5BMmE4ZD...>

	Title	Release Date	Budget \
0	The Idea of You	2022-07-16	42110861
1	Kingdom of the Planet of the Apes	2023-05-12	14627858
2	Unfrosted	2023-04-20	48805376
3	The Fall Guy	2022-10-07	28147070
4	Challengers	2023-01-05	49549901
...
9995	The Greatest Show on Earth	2023-02-11	40892520
9996	Berserk: Ougon Jidai-hen I - Haou no Tamago	2022-08-19	29174237
9997	Is-slottet	2023-05-06	39411922
9998	Loving Pablo	2022-12-29	49927298
9999	Un homme et une femme	2022-05-22	35426947

	Revenue	Duration (min)	Genre	Rating \
0	4886739	115.0	Comedy, Drama, Romance	6.4
1	7458541	145.0	Action, Adventure, Sci-Fi	7.3
2	7814085	97.0	Biography, Comedy, History	5.5
3	4951558	126.0	Action, Comedy, Drama	7.3
4	6891688	131.0	Drama, Romance, Sport	7.7
...
9995	5496436	152.0	Drama, Family, Romance	6.5
9996	5128209	76.0	Animation, Action, Adventure	7.5
9997	6770819	78.0	Mystery, Drama	6.5
9998	5635966	123.0	Biography, Crime, Drama	6.4
9999	5634048	102.0	Drama, Romance	7.5

	Metascore	Director \
0	67.0	Michael Showalter
1	66.0	Wes Ball
2	42.0	Jerry Seinfeld
3	73.0	David Leitch
4	82.0	Luca Guadagnino
...
9995	76.0	Cecil B. DeMille
9996	58.1	Toshiyuki Kubooka
9997	58.1	Per Blom
9998	42.0	Fernando León de Aranoa
9999	58.1	Claude Lelouch

	Cast	Votes \
0	Anne Hathaway, Nicholas Galitzine, Ella Rubin,...	28744.0
1	Owen Teague, Freya Allan, Kevin Durand, Peter ...	22248.0

2	Isaac Bae, Jerry Seinfeld, Chris Rickett, Rach...	18401.0
3	Ryan Gosling, Emily Blunt, Aaron Taylor-Johnso...	38953.0
4	Zendaya, Mike Faist, Josh O'Connor, Darnell Ap...	32517.0
...
9995	James Stewart, Charlton Heston, Betty Hutton, ...	16078.0
9996	Hiroaki Iwanaga, Carrie Keranen, Takahiro Saku...	14300.0
9997	Line Storesund, Hilde Nyeggen Martinsen, Meret...	740.0
9998	Javier Bardem, Penélope Cruz, Peter Sarsgaard,...	22447.0
9999	Anouk Aimée, Jean-Louis Trintignant, Pierre Ba...	11826.0

	Description	Review Count	\
0	Solène, a 40-year-old single mom, begins an un...	166	
1	Many years after the reign of Caesar, a young ...	183	
2	In 1963 Michigan, business rivals Kellogg's an...	333	
3	A down-and-out stuntman must find the missing ...	384	
4	Tashi, a former tennis prodigy turned coach, t...	194	
...	
9995	The dramatic lives of trapeze artists, a clown...	128	
9996	A lone sellword named Guts gets recruited int...	12	
9997	A couple of twelve-year-old Norwegian girls st...	4	
9998	A journalist strikes up a romantic relationshi...	84	
9999	A widow and a widower find their relationship ...	61	

	Review Title	\
0	Hypocrisy as an idea	
1	A phenomenal start to another trilogy!	
2	not funny	
3	Everything you needed and more!	
4	Watch "Match Point" instead	
...	...	
9995	Hey, doesn't anyone remember Last Emperor?	
9996	Masterfully directed climatic epic saga	
9997	Beautiful Film	
9998	That film should be in Spanish	
9999	Moved me tremendously.	

	Review	Genre1	Month	\
0	This film, as well as the reaction to it, is a...	Comedy	7	
1	I'm a big fan of all the planet of the apes, a...	Action	5	
2	Pretty much the worst criticism you can lay on...	Biography	4	
3	Just got out of the Austin premier at SXSW and...	Action	10	
4	This is a tough one. I liked the concept and t...	Drama	1	
...	
9995	It constantly amazes me that people carp that ...	Drama	2	
9996	Few stories can capture your mind and soul in ...	Animation	8	
9997	This film might not be to everyone's taste, it...	Mystery	5	
9998	Why anyone (the director?) made Spanish actors...	Biography	12	

9999 This is a movie that resonated to my core. I ... Drama 5

```
Quarter
0    2022Q3
1    2023Q2
2    2023Q2
3    2022Q4
4    2023Q1
...
9995 2023Q1
9996 2022Q3
9997 2023Q2
9998 2022Q4
9999 2022Q2
```

[9954 rows x 19 columns]

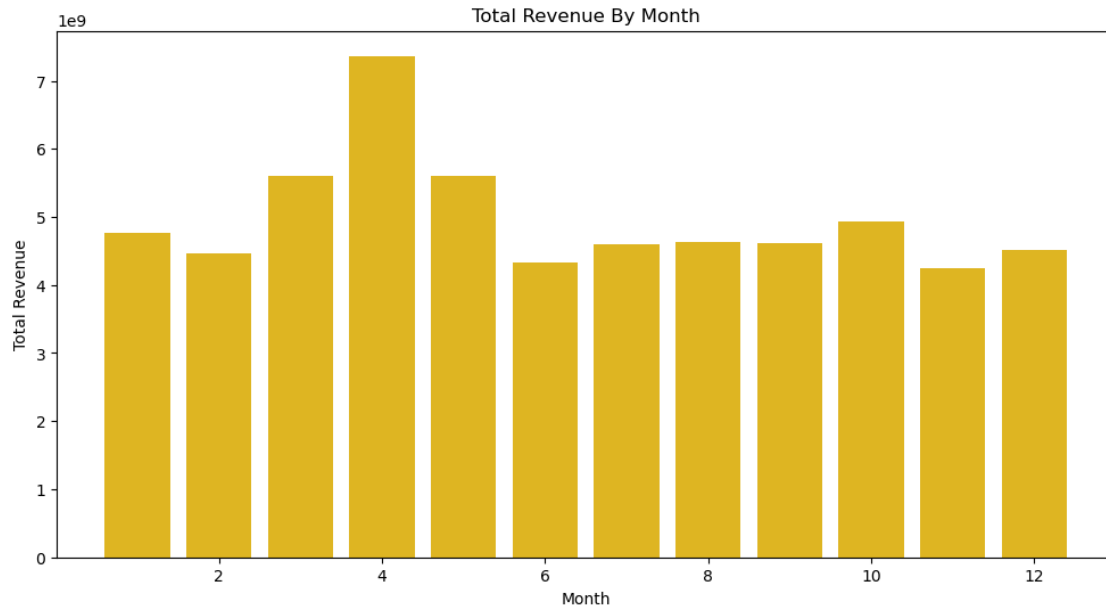
```
[39]: monthly_revenue=imdb.groupby('Month')['Revenue'].agg(Total_Revenue='sum').
      ↪reset_index()
      quarterly_revenue=imdb.groupby('Quarter')['Revenue'].agg(Total_Revenue='sum').
      ↪reset_index()
```

```
[40]: monthly_revenue
```

```
[40]:
```

	Month	Total_Revenue
0	1	4770402928
1	2	4468242785
2	3	5599387597
3	4	7364709198
4	5	5601369196
5	6	4327750496
6	7	4591785442
7	8	4623944096
8	9	4622801626
9	10	4928084858
10	11	4241873329
11	12	4521793842

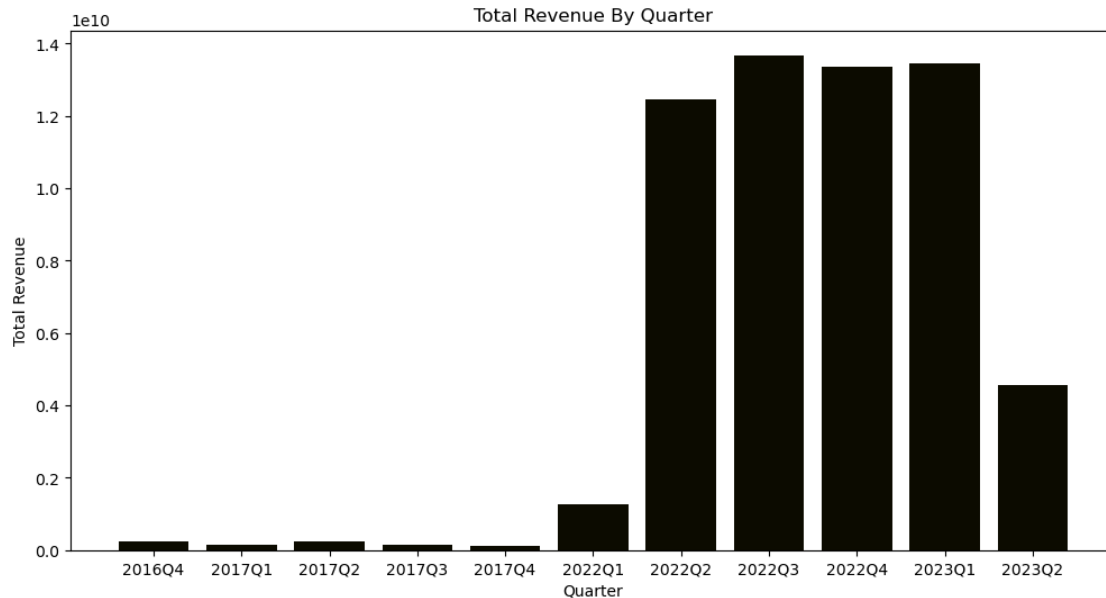
```
[41]: plt.figure(figsize=(12,6))
      plt.
      ↪bar(monthly_revenue['Month'],monthly_revenue['Total_Revenue'],color='#DEB522')
      plt.title('Total Revenue By Month')
      plt.xlabel('Month')
      plt.ylabel('Total Revenue')
      plt.show()
```



```
[42]: quarterly_revenue
```

```
[42]:   Quarter  Total_Revenue
0   2016Q4      228927332
1   2017Q1      136489557
2   2017Q2      254489223
3   2017Q3      158833692
4   2017Q4      102424066
5   2022Q1      1260204953
6   2022Q2      12466741869
7   2022Q3      13679697472
8   2022Q4      13360400631
9   2023Q1      13441338800
10  2023Q2       4572597798
```

```
[43]: plt.figure(figsize=(12,6))
plt.bar(quarterly_revenue['Quarter'].
        astype(str),quarterly_revenue['Total_Revenue'],color='#0c0b00')
plt.title('Total Revenue By Quarter')
plt.xlabel('Quarter')
plt.ylabel('Total Revenue')
plt.show()
```



Seasonality Trends:

Best Performance: Movies released in April, summer months (May to August), and the holiday season (November to December) tend to perform better. Q2 and Q3 of 2022 are standout periods, indicating that the summer blockbusters have a significant impact on overall box office revenue.

Recommendations: Release Strategy: Studios might consider scheduling major releases during peak months and quarters to maximize box office revenue. Market Analysis: Further investigation into specific genres and their performance during these peak times could provide additional insights for future releases.

Task 3.3: Analyze the performance of movies directed by top directors (those who directed the most movies). Do certain directors consistently produce high-grossing movies?

```
[44]: top=imdb.groupby('Director')['Director'].size().rename('count').reset_index()
top.sort_values(by='count',ascending=False).head(10)
```

```
[44]:
```

	Director	count
826	Clint Eastwood	35
4104	Steven Spielberg	33
3605	Ridley Scott	28
4103	Steven Soderbergh	26
3738	Ron Howard	26
2805	Martin Scorsese	25
167	Alfred Hitchcock	24
4481	Woody Allen	22
3691	Robert Zemeckis	21
4219	Tim Burton	20


```
[45]: top_director_movies = imdb[imdb['Director'].isin(top['Director'])]
director_performance = top_director_movies.groupby('Director').
    ↳agg(movie_count=('Title','size'),director_count=('Director','size'),average_revenue=('Revenue','mean'))
    ↳reset_index()
director_performance = director_performance.
    ↳sort_values(by='average_revenue',ascending=False).head(10)
```

```
[46]: director_performance
```

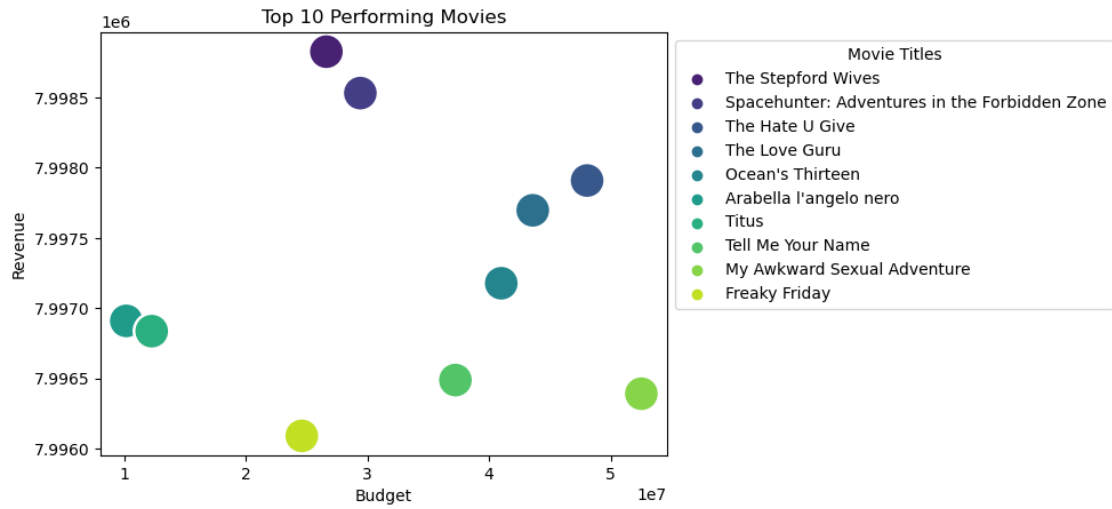
```
[46]:
```

	Director	movie_count	director_count	average_revenue	\
611	Bryan Forbes	1	1	7998826.0	
2514	Lamont Johnson	1	1	7998531.0	
2728	Marco Schnabel	1	1	7997697.0	
4032	Stelvio Massi	1	1	7996910.0	
1868	Jason DeVan	1	1	7996489.0	
3890	Sean Garrity	1	1	7996392.0	
2635	Louise Archambault	1	1	7995591.0	
1983	Jessica Bendinger	1	1	7992155.0	
2430	Kiah Roache-Turner	1	1	7990708.0	
3891	Sean Hoessli	1	1	7987419.0	

	max_revenue
611	7998826
2514	7998531
2728	7997697
4032	7996910
1868	7996489
3890	7996392
2635	7995591
1983	7992155
2430	7990708
3891	7987419

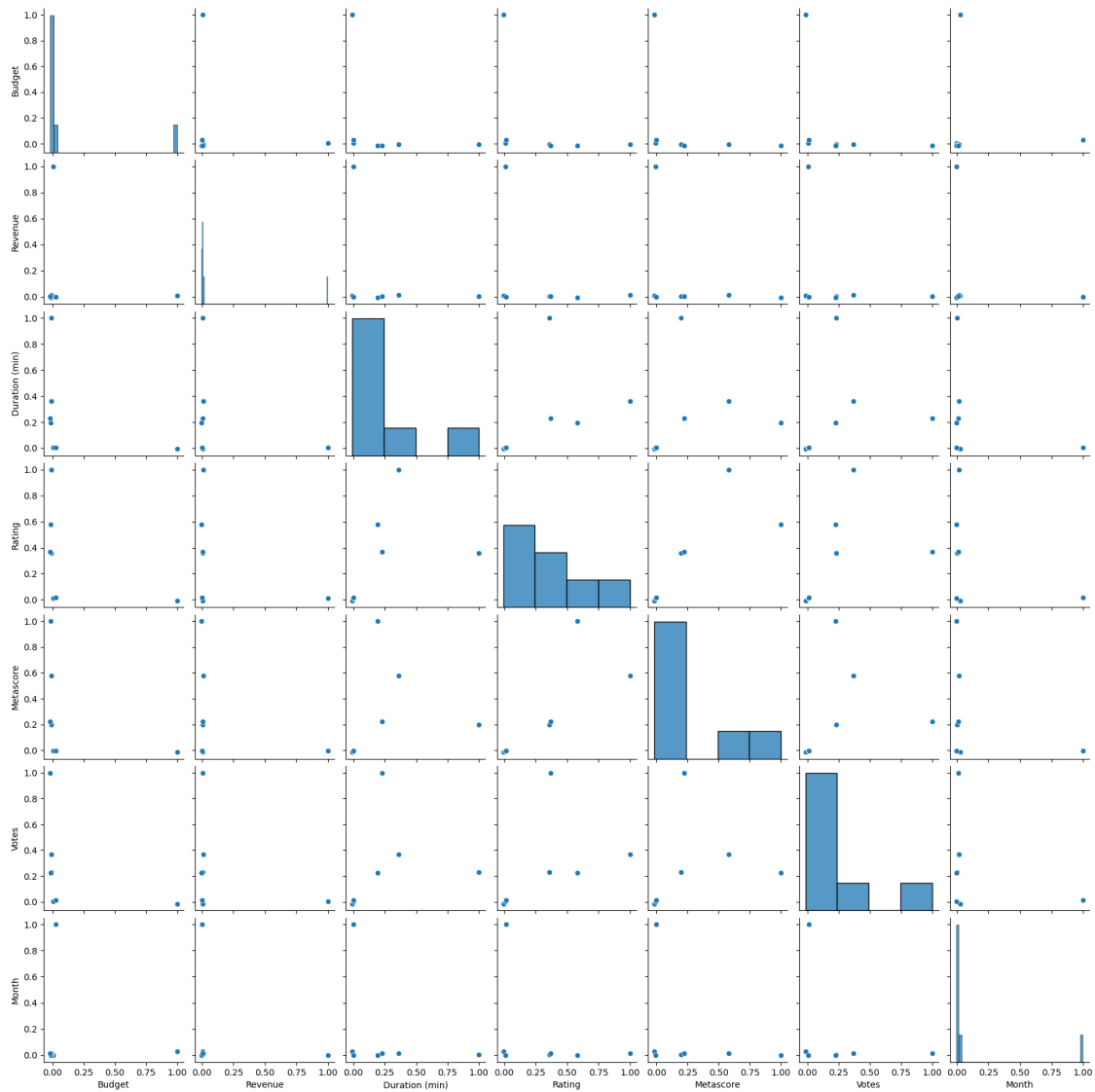
```
[47]: top_performers=imdb.nlargest(10, 'Revenue')
```

```
[48]: sns.scatterplot(data=top_performers, x='Budget', y='Revenue', hue='Title',
    ↳s=500, palette='viridis')
plt.title('Top 10 Performing Movies')
plt.xlabel('Budget')
plt.ylabel('Revenue')
plt.legend(title='Movie Titles', bbox_to_anchor=(1, 1), loc='upper left')
plt.show()
```



```
[49]: numerical_corr=imdb.select_dtypes('number').corr()
```

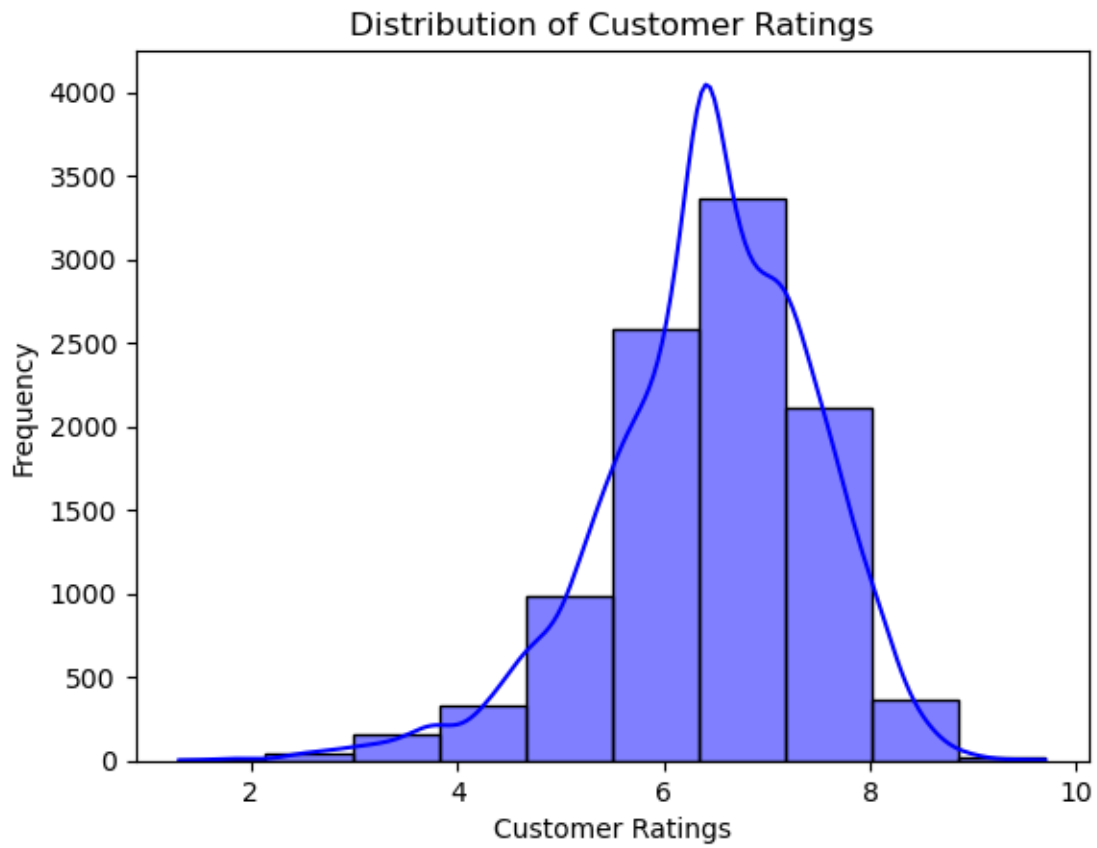
```
[50]: sns.pairplot(numerical_corr)
plt.show()
```



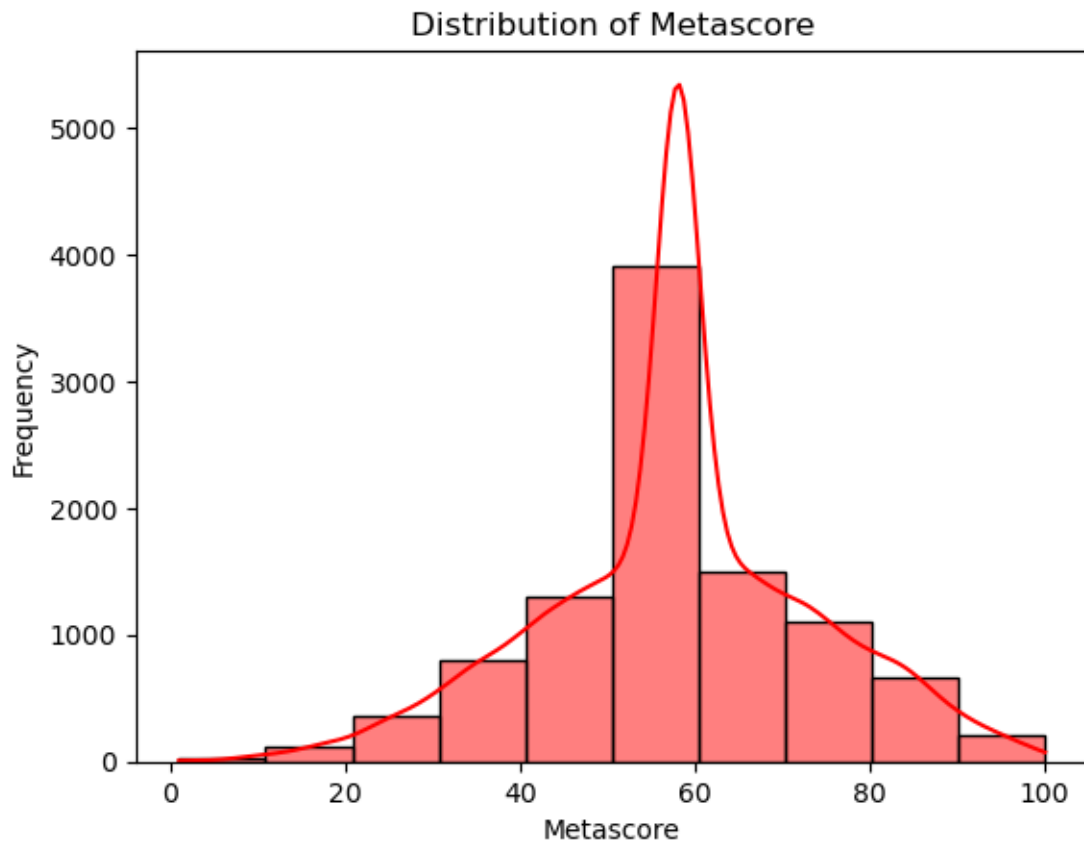
```
[51]: plt.figure(figsize=(15,6))
sns.boxplot(data=imdb, x='Genre1', y='Rating', palette='Set3')
plt.title('Comparing Movie Rating with different Genre')
plt.xlabel('Genre')
plt.ylabel('Rating')
plt.xticks(rotation=60)
plt.show()
```



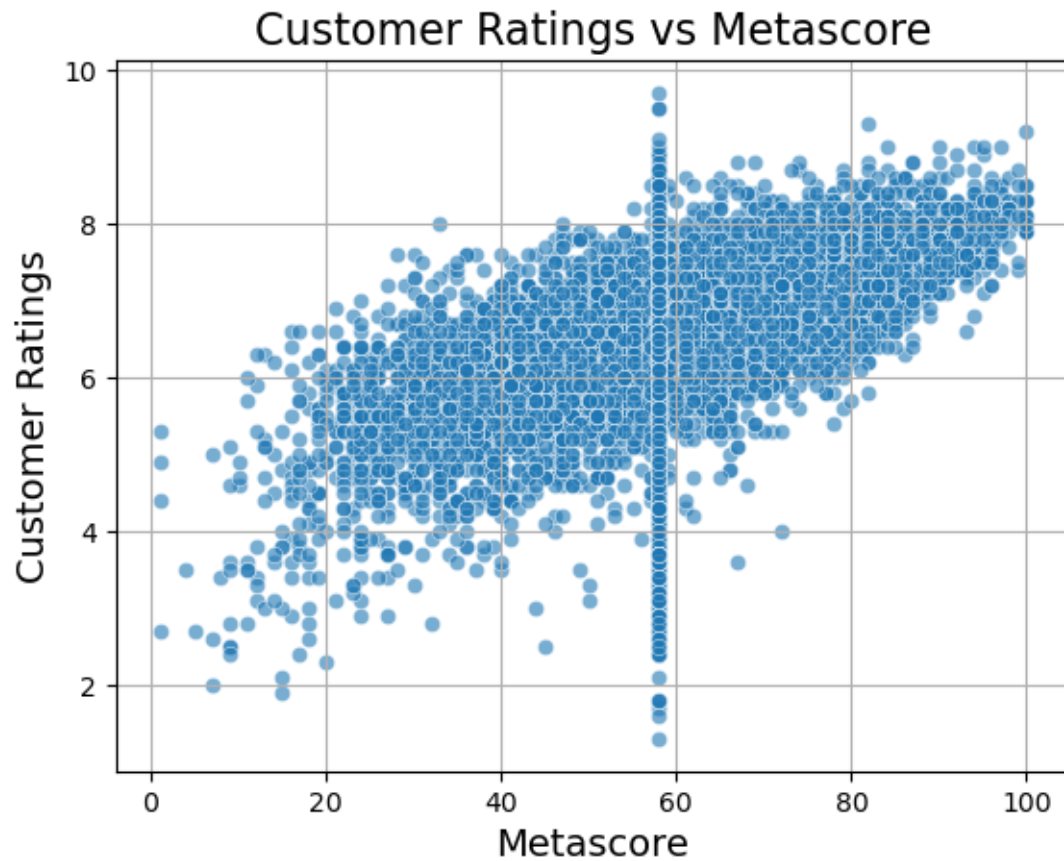
```
plt.show()
```



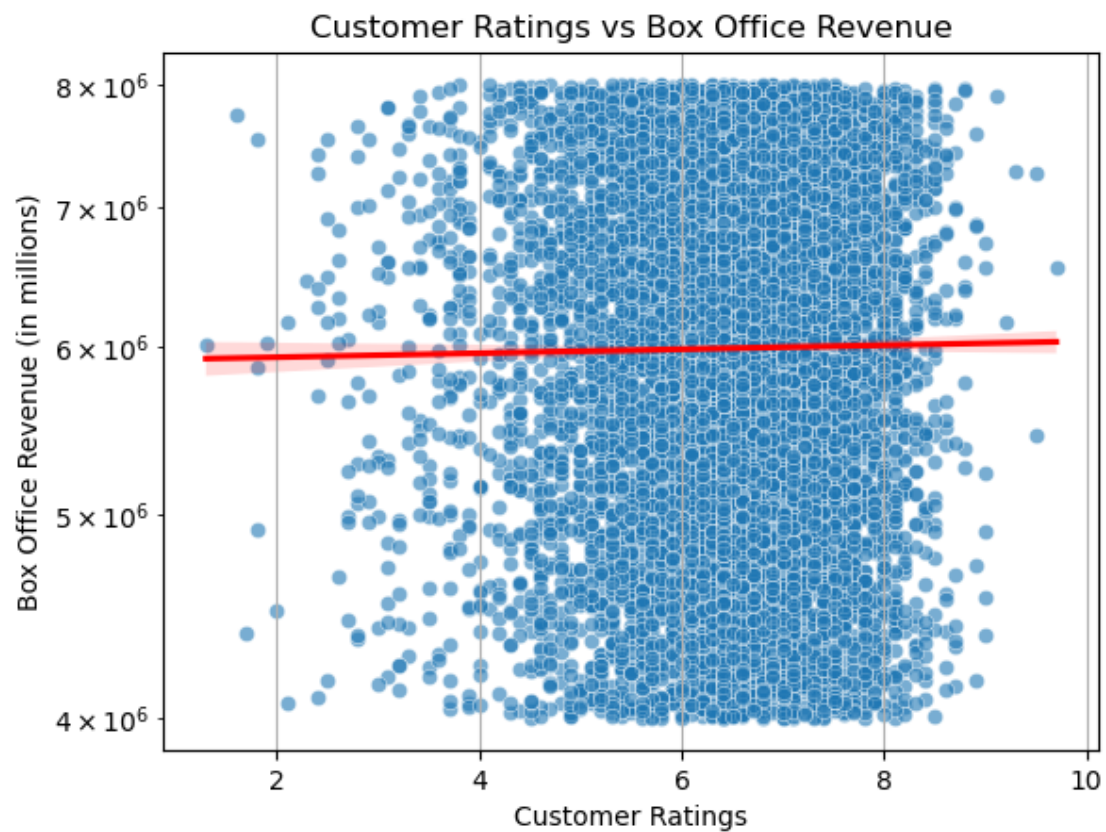
```
[54]: sns.histplot(imdb['Metascore'], bins=10, kde=True, color='red')
plt.title('Distribution of Metascore')
plt.xlabel('Metascore' )
plt.ylabel('Frequency' )
plt.show()
```



```
[55]: sns.scatterplot(data=imdb, x='Metascore', y='Rating', alpha=0.6)
plt.title('Customer Ratings vs Metascore', fontsize=16)
plt.xlabel('Metascore', fontsize=14)
plt.ylabel('Customer Ratings', fontsize=14)
plt.grid(True)
```



```
[56]: sns.scatterplot(data=imdb, x='Rating', y='Revenue', alpha=0.6)
sns.regplot(data=imdb, x='Rating', y='Revenue', scatter=False, color='red')
plt.title('Customer Ratings vs Box Office Revenue')
plt.xlabel('Customer Ratings')
plt.ylabel('Box Office Revenue (in millions)')
plt.xscale('linear')
plt.yscale('log')
plt.grid(True)
plt.show()
```



[]: