

Name :- Sakshi Kadam

Roll no :- ET2-55

PRN :- 202401070169

Dataset Link: <https://www.kaggle.com/datasets/PromptCloudHQ/imdb-data>

```
import warnings
```

```
warnings.filterwarnings('ignore')
```

```
import pandas as pd
```

```
data = pd.read_csv('/kaggle/input/imdb-data/IMDB-Movie-Data.csv')
```

1. Display Top 10 Rows of The Dataset :-

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

1. Display Top 10 Rows of The Dataset

```
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

2. Check Last 10 Rows of The Dataset :-

data.tail(10)

2. Check Last 10 Rows of The Dataset

+ Code + Markdown

```
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

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

data.shape

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

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

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

```
22]: data.shape
```

```
22]: (1000, 12)
```

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

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

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

data.info()

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

```
16]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 1000 entries, 0 to 999  
Data columns (total 12 columns):  
#   Column                Non-Null Count  Dtype    
---  ---                  
0   Rank                  1000 non-null  int64    
1   Title                 1000 non-null  object    
2   Genre                 1000 non-null  object    
3   Description            1000 non-null  object    
4   Director              1000 non-null  object    
5   Actors                1000 non-null  object    
6   Year                  1000 non-null  int64    
7   Runtime (Minutes)     1000 non-null  int64    
8   Rating                1000 non-null  float64   
9   Votes                 1000 non-null  int64    
10  Revenue (Millions)    872 non-null   float64   
11  Metascore             936 non-null   float64   
dtypes: float64(3), int64(4), object(5)  
memory usage: 93.9+ KB
```

5. Check Null Values In The Dataset :-

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

```
import matplotlib.pyplot as plt
```

```
import seaborn as sns
```

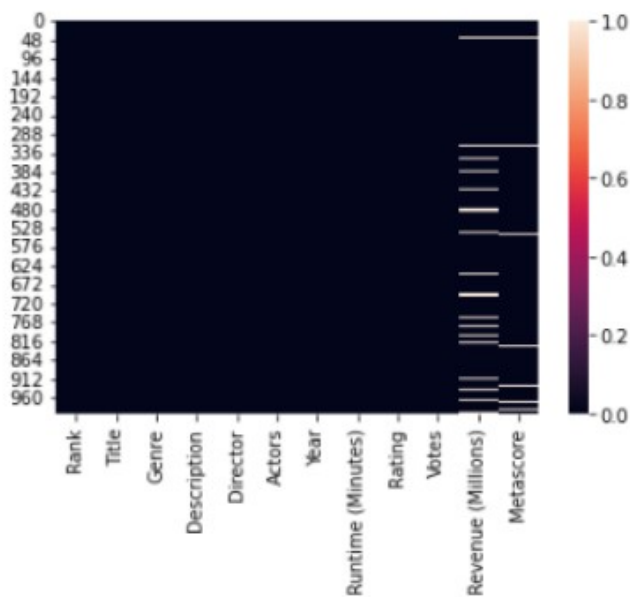
```
sns.heatmap(data.isnull())
```

```
plt.show()
```

```
[24]: data.isnull().sum()
```

```
[24]: Rank          0
      Title         0
      Genre         0
      Description   0
      Director      0
      Actors        0
      Year          0
      Runtime (Minutes)  0
      Rating        0
      Votes         0
      Revenue (Millions) 128
      Metascore     64
      dtype: int64
```

```
[25]: import matplotlib.pyplot as plt
      import seaborn as sns
      sns.heatmap(data.isnull())
      plt.show()
```



6. Drop All The Missing Values :-

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

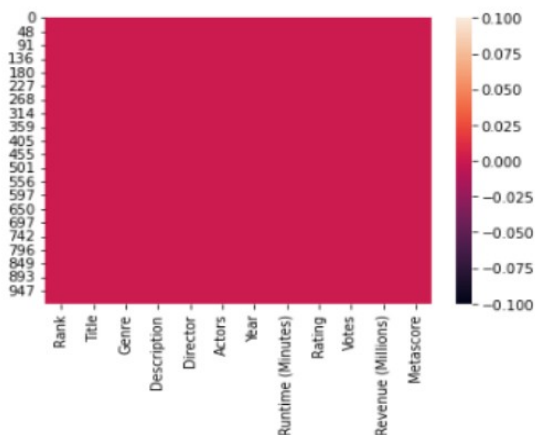
```
sns.heatmap(data.isnull())
```

```
plt.show()
```

6. Drop All The Missing Values

```
[28]: data = data.dropna(axis=0)
```

```
[29]: sns.heatmap(data.isnull())  
plt.show()
```



7. Check For Duplicate Data :-

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

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

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

```
Are there any duplicated values in data? False
```

8. Get Overall Statistics About The DataFrame :-

data.describe()

```
[31]: data.describe()
```

```
[31]:
```

	Rank	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
count	838.000000	838.00000	838.000000	838.000000	8.380000e+02	838.000000	838.000000
mean	485.247017	2012.50716	114.638425	6.814320	1.932303e+05	84.564558	59.575179
std	286.572065	3.17236	18.470922	0.877754	1.930990e+05	104.520227	16.952416
min	1.000000	2006.00000	66.000000	1.900000	1.780000e+02	0.000000	11.000000
25%	238.250000	2010.00000	101.000000	6.300000	6.127650e+04	13.967500	47.000000
50%	475.500000	2013.00000	112.000000	6.900000	1.368795e+05	48.150000	60.000000
75%	729.750000	2015.00000	124.000000	7.500000	2.710830e+05	116.800000	72.000000
max	1000.000000	2016.00000	187.000000	9.000000	1.791916e+06	936.630000	100.000000

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

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

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

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

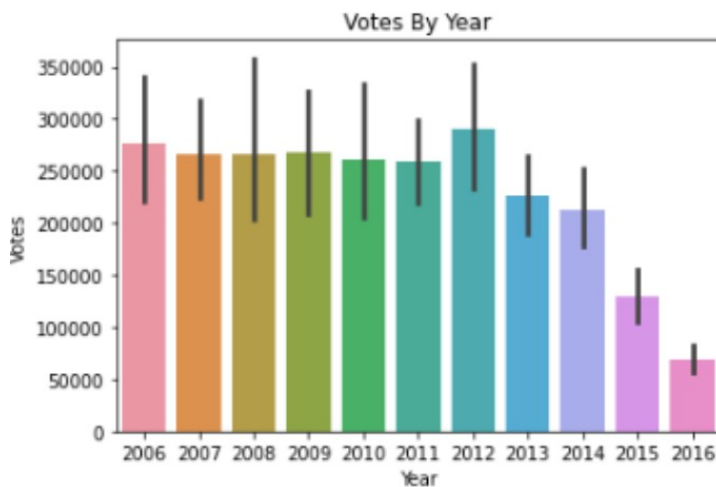
10. In Which Year There Was The Highest Voting?

sns.barplot(x='Year',y='Votes',data=data)

plt.title("Votes By Year")

plt.show()

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

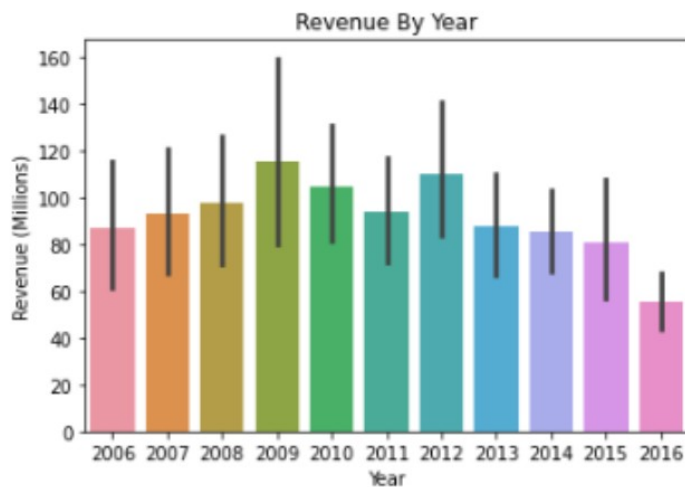


11. In Which Year There Was The Highest Revenue?

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



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



12. Find The Average Rating For Each Director :-

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

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

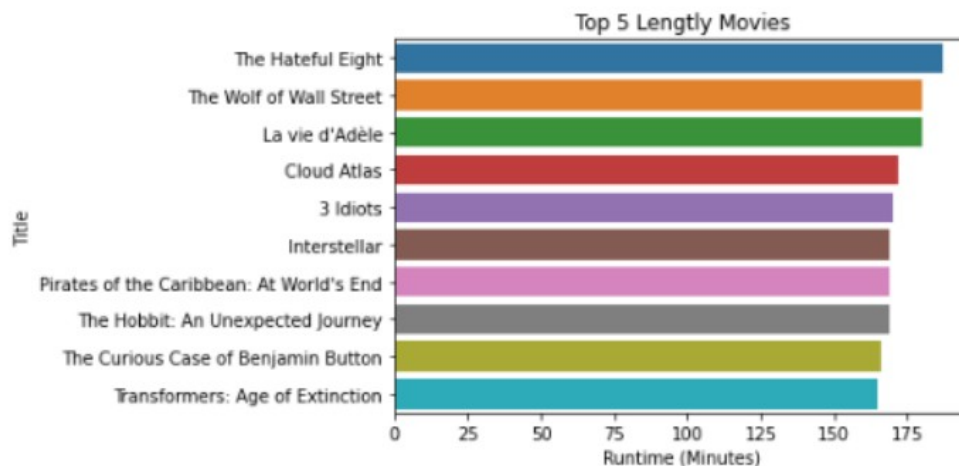
```
Director
Christopher Nolan      8.68
Olivier Nakache        8.60
Makoto Shinkai         8.60
Florian Henckel von Donnersmarck  8.50
Aamir Khan             8.50
...
Sam Taylor-Johnson     4.10
Joey Curtis            4.00
George Nolfi           3.90
James Wong             2.70
Jason Friedberg        1.90
Name: Rating, Length: 524, dtype: float64
```

13. Display Top 10 Lengthy Movies Title :-

```
le =data.nlargest(10,'Runtime (Minutes)')[['Title','Runtime
(Minutes)']]. \
set_index('Title')
sns.barplot(le['Runtime (Minutes)'],y=le.index)
plt.title('Top 5 Lengthly Movies')
plt.show()
```

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

```
sns.barplot(le['Runtime (Minutes)'],y=le.index)
plt.title('Top 5 Lengthly Movies')
plt.show()
```

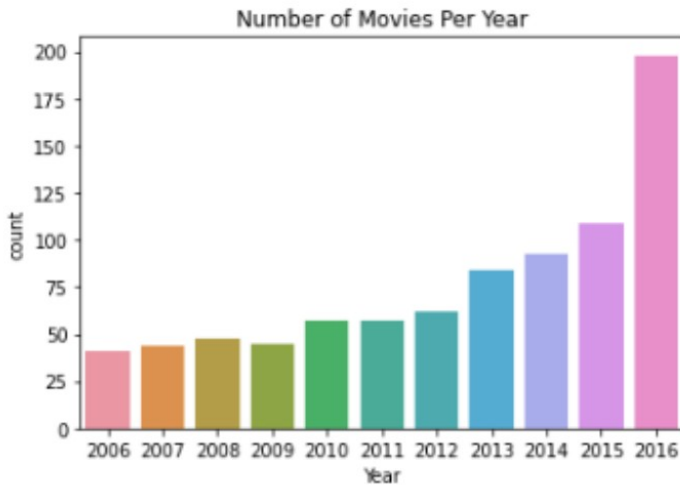


14. Display Number of Movies Per Year :-

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

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

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



15. Find Most Popular Movie Title (Highest Revenue) :-

data.columns

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

```
data.columns
```

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

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

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

16. Display Top 10 Highest Rated Movie Titles And its

Directors :-

```
top_10=data.nlargest(10,'Rating')
[['Title','Rating','Director']].set_index('Title')
top_10
sns.barplot(top_10['Rating'],top_10.index)
plt.title("Display Top 10 Highest Rated Movie Titles")
```

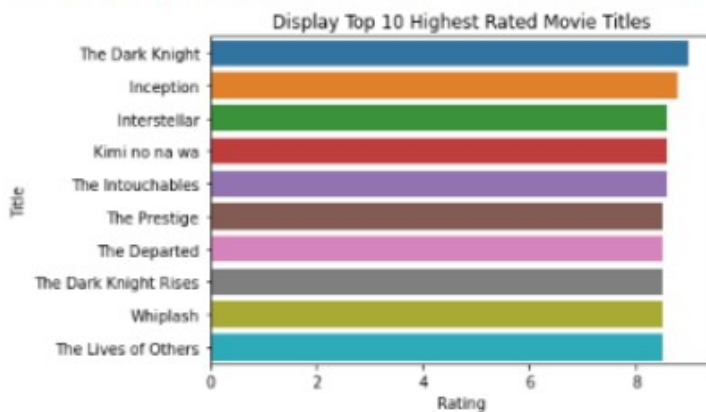
```
top_10=data.nlargest(10,'Rating')[['Title','Rating','Director']].set_index('Title')
```

```
top_10
```

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

```
sns.barplot(top_10['Rating'],top_10.index)
plt.title("Display Top 10 Highest Rated Movie Titles")
```

```
Text(0.5, 1.0, 'Display Top 10 Highest Rated Movie Titles')
```



17. Display Top 10 Highest Revenue Movie Titles :-

```
data.columns
```

```
data.sort_values(by='Revenue')
```

```
(Millions)',ascending=False).head(10)
```

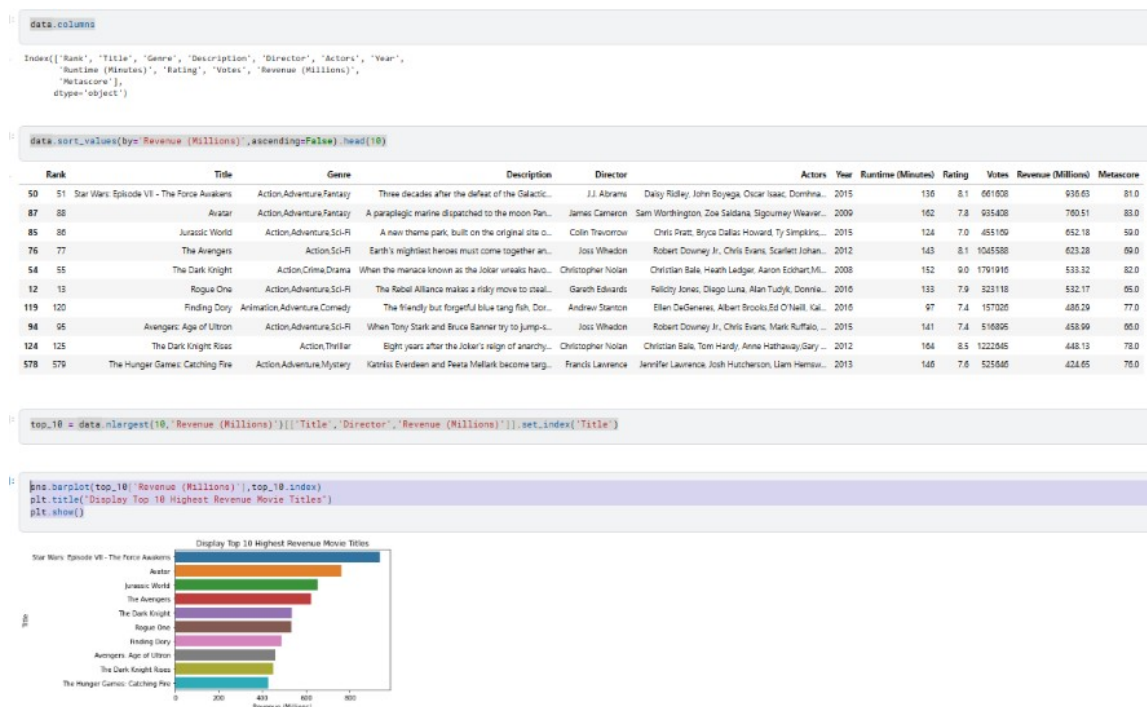
```
top_10 = data.nlargest(10,'Revenue (Millions)')
```

```
[['Title','Director','Revenue (Millions)']].set_index('Title')
```

```
sns.barplot(top_10['Revenue (Millions)'],top_10.index)
```

```
plt.title("Display Top 10 Highest Revenue Movie Titles")
```

```
plt.show()
```



18. Find Average Rating of Movies Year-wise :-

```
data.columns
```

```
data1=data.groupby('Year')['Year','Rating'].mean().\
```

```
sort_values(by='Rating',ascending=False).set_index('Year')
```

```
data1
```

```
plt.figure(figsize=(10,5))
sns.barplot(data1.index,data1['Rating'])
plt.show()
```



19. Classify Movies Based on Ratings [Good,Better and Best] :-

```
data.columns
```

```
def rating(rating):
```

```
    if rating>=7.0:
```

```
        return 'Excellent'
```

```
    elif rating>=6.0:
```

```
        return 'Good'
```

```
    else:
```

```
        return 'Average'
```

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

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

```
data.columns
```

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

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

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

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

	Rank	Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore	rating_cat
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	Excellent

20. Count Number of Action Movies :-

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



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

```
: 277
```

