Task 5

GO_STP_5247

Pandas - Data Analysis of IMDB movies data

let's explore the fun and interesting 'IMDB-movies-dataset' and get our hands dirty by performing practical data analysis on real data.

```
[1]  import numpy as np
     import pandas as pd
     import seaborn as sns
     import matplotlib.pyplot as plt
```

1. Load the IMDb Dataset and read

```
[2]  df = pd.read_csv('/IMDB-Movie-Data.csv')
```

2. View the dataset

```
[3]  df.head()
```

	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

✓ 1s   completed at 11:34 PM

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

```
[4]  df.tail()
```

	Rank	Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
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. Understand some basic information about the dataset and Inspect the dataframe Inspect the dataframe's columns, shapes, variable types etc.

```
[5]  df.shape
```

```
    (1000, 12)
```

```
    df.ndim
```

```
    2
```

[7] `df.dtypes`

```
    Rank                  int64
    Title                object
    Genre                object
    Description          object
    Director             object
    Actors               object
    Year                  int64
    Runtime (Minutes)     int64
    Rating              float64
    Votes                 int64
    Revenue (Millions)  float64
    Metascore           float64
    dtype: object
```

[8] `df.columns`

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

[9] `df.info()`

```
    <class 'pandas.core.frame.DataFrame'>
    RangeIndex: 1000 entries, 0 to 999
    Data columns (total 12 columns):
     #   Column              Non-Null Count   Dtype
```

```
[9]    #    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
```

```
[10] df.nunique()

      Rank                1000
      Title               999
      Genre               207
      Description         1000
      Director            644
      Actors              996
      Year                11
      Runtime (Minutes)   94
      Rating              55
      Votes               997
      Revenue (Millions)  814
      Metascore           84
      dtype: int64
```

```
[11] df.Genre.value_counts()

      Action,Adventure,Sci-Fi     50
```

```
Drama                          48
Comedy,Drama,Romance           35
Comedy                         32
Drama,Romance                  31
                               ..
Sci-Fi,Thriller                 1
Comedy,Horror,Romance           1
Adventure,Drama                 1
Action,Thriller,War             1
Action,Horror                   1
Name: Genre, Length: 207, dtype: int64
```

[12] `df.Year.value_counts()`

```
2016    297
2015    127
2014     98
2013     91
2012     64
2011     63
2010     60
2007     53
2008     52
2009     51
2006     44
Name: Year, dtype: int64
```

[13] `df.describe()`

	Rank	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
count	1000.000000	1000.000000	1000.000000	1000.000000	1.000000e+03	872.000000	936.000000
mean	500.500000	2012.783000	113.172000	6.723200	1.698083e+05	82.956376	58.985043
std	288.819436	3.205962	18.810908	0.945429	1.887626e+05	103.253540	17.194757

min	1.000000	2006.000000	66.000000	1.900000	6.100000e+01	0.000000	11.000000
25%	250.750000	2010.000000	100.000000	6.200000	3.630900e+04	13.270000	47.000000
50%	500.500000	2014.000000	111.000000	6.800000	1.107990e+05	47.985000	59.500000
75%	750.250000	2016.000000	123.000000	7.400000	2.399098e+05	113.715000	72.000000
max	1000.000000	2016.000000	191.000000	9.000000	1.791916e+06	936.630000	100.000000

4. Data Selection – Indexing and Slicing data

```
[14] df['Rating']
```

```
0      8.1
1      7.0
2      7.3
3      7.2
4      6.2
      ...
995    6.2
996    5.5
997    6.2
998    5.6
999    5.3
Name: Rating, Length: 1000, dtype: float64
```

```
[15] df.loc[: ,'Rating']
```

```
0      8.1
1      7.0
2      7.3
3      7.2
4      6.2
      ...
995    6.2
```

```
[15] 995    6.2
     996    5.5
     997    6.2
     998    5.6
     999    5.3
     Name: Rating, Length: 1000, dtype: float64
```

```
[16] df.iloc[: ,8]
```

```
0      8.1
1      7.0
2      7.3
3      7.2
4      6.2
      ...
995    6.2
996    5.5
997    6.2
998    5.6
999    5.3
Name: Rating, Length: 1000, dtype: float64
```

```
[17] df.loc[:, "Rank" : "Director"]
```

	Rank	Title	Genre	Description	Director
0	1	Guardians of the Galaxy	Action,Adventure,Sci-Fi	A group of intergalactic criminals are forced ...	James Gunn
1	2	Prometheus	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott
2	3	Split	Horror,Thriller	Three girls are kidnapped by a man with a diag...	M. Night Shyamalan
3	4	Sing	Animation,Comedy,Family	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet
4	5	Suicide Squad	Action,Adventure,Fantasy	A secret government agency recruits some of th...	David Ayer

```
[17]
```

	...	...	...	...	...	...
995	996	Secret in Their Eyes	Crime,Drama,Mystery	A tight-knit team of rising investigators, alo...	Billy Ray	
996	997	Hostel: Part II	Horror	Three American college students studying abroa...	Eli Roth	
997	998	Step Up 2: The Streets	Drama,Music,Romance	Romantic sparks occur between two dance studen...	Jon M. Chu	
998	999	Search Party	Adventure,Comedy	A pair of friends embark on a mission to reuni...	Scot Armstrong	
999	1000	Nine Lives	Comedy,Family,Fantasy	A stuffy businessman finds himself trapped ins...	Barry Sonnenfeld	

1000 rows × 5 columns

```
[18] df.iloc[:, 0:5]
```

	Rank	Title	Genre	Description	Director
0	1	Guardians of the Galaxy	Action,Adventure,Sci-Fi	A group of intergalactic criminals are forced ...	James Gunn
1	2	Prometheus	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott
2	3	Split	Horror,Thriller	Three girls are kidnapped by a man with a diag...	M. Night Shyamalan
3	4	Sing	Animation,Comedy,Family	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet
4	5	Suicide Squad	Action,Adventure,Fantasy	A secret government agency recruits some of th...	David Ayer
...	...	...	...	...	...
995	996	Secret in Their Eyes	Crime,Drama,Mystery	A tight-knit team of rising investigators, alo...	Billy Ray
996	997	Hostel: Part II	Horror	Three American college students studying abroa...	Eli Roth
997	998	Step Up 2: The Streets	Drama,Music,Romance	Romantic sparks occur between two dance studen...	Jon M. Chu
998	999	Search Party	Adventure,Comedy	A pair of friends embark on a mission to reuni...	Scot Armstrong
999	1000	Nine Lives	Comedy,Family,Fantasy	A stuffy businessman finds himself trapped ins...	Barry Sonnenfeld

5. Data Selection – Based on Conditional filtering

```
[19] df.loc[(df["Rating"] > 7 ),['Title','Director', 'Actors', 'Year','Runtime (Minutes)', 'Rating']]
```

	Title	Director	Actors	Year	Runtime (Minutes)	Rating
0	Guardians of the Galaxy	James Gunn	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	2014	121	8.1
2	Split	M. Night Shyamalan	James McAvoy, Anya Taylor-Joy, Haley Lu Richar...	2016	117	7.3
3	Sing	Christophe Lourdelet	Matthew McConaughey,Reese Witherspoon, Seth Ma...	2016	108	7.2
6	La La Land	Damien Chazelle	Ryan Gosling, Emma Stone, Rosemarie DeWitt, J...	2016	128	8.3
8	The Lost City of Z	James Gray	Charlie Hunnam, Robert Pattinson, Sienna Mille...	2016	141	7.1
...	...	...	...	...	...	...
979	The Skin I Live In	Pedro Almodóvar	Antonio Banderas, Elena Anaya, Jan Cornet,Mari...	2011	120	7.6
982	Across the Universe	Julie Taymor	Evan Rachel Wood, Jim Sturgess, Joe Anderson, ...	2007	133	7.4
988	Martyrs	Pascal Laugier	Morjana Alaoui, Mylène Jampanoï, Catherine Bég...	2008	99	7.1
989	Selma	Ava DuVernay	David Oyelowo, Carmen Ejogo, Tim Roth, Lorrain...	2014	128	7.5
991	Taare Zameen Par	Aamir Khan	Darsheel Safary, Aamir Khan, Tanay Chheda, Sac...	2007	165	8.5

399 rows × 6 columns

```
[20] df.where(df.Rating > 7)
```

	Rank	Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
0	1.0	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...	2014.0	121.0	8.1	757074.0	333.13	76.0

```
[20]
```

1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	3.0	Split	Horror,Thriller	Three girls are kidnapped by a man with a diag...	M. Night Shyamalan	James McAvoy, Anya Taylor-Joy, Haley Lu Richar...	2016.0	117.0	7.3	157606.0	138.12	62.0
3	4.0	Sing	Animation,Comedy,Family	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet	Matthew McConaughey,Reese Witherspoon, Seth Ma...	2016.0	108.0	7.2	60545.0	270.32	59.0
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...
995	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
996	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
997	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
998	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
999	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

1000 rows × 12 columns

```
[21] df.query('Rating > 7')
```

	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
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
6	7	La La Land	Comedy,Drama,Music	A jazz pianist falls for an ...	Damien	Ryan Gosling, Emma Stone,	2016	128	8.3	258682	151.06	93.0

[21]	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
	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
	...	...	...	...	...	...	...	...	...	...	...	...	...
	979	980	The Skin I Live In	Drama,Thriller	A brilliant plastic surgeon, haunted by past t...	Pedro Almodóvar	Antonio Banderas, Elena Anaya, Jan Cornet,Mari...	2011	120	7.6	108772	3.19	70.0
	982	983	Across the Universe	Drama,Fantasy,Musical	The music of the Beatles and the Vietnam War f...	Julie Taymor	Evan Rachel Wood, Jim Sturgess, Joe Anderson, ...	2007	133	7.4	95172	24.34	56.0
	988	989	Martyrs	Horror	A young woman's quest for revenge against the ...	Pascal Laugier	Morjana Alaoui, Mylène Jampanoï, Catherine Bég...	2008	99	7.1	63785	NaN	89.0
	989	990	Selma	Biography,Drama,History	A chronicle of Martin Luther King's campaign t...	Ava DuVernay	David Oyelowo, Carmen Ejogo, Tim Roth, Lorrain...	2014	128	7.5	67637	52.07	NaN
	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

399 rows × 12 columns

6. Groupby operations

```
[22] df.groupby('Genre').sum()
```

Genre	Rank	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
Action	705	4032	176	12.3	17237	131.56	37.0
Action,Adventure	999	6033	373	19.2	1003377	671.22	172.0
Action,Adventure,Biography	568	4031	263	14.0	97560	33.00	125.0

Action,Adventure,Biography	568	4031	263	14.0	97560	33.00	125.0
Action,Adventure,Comedy	5700	28199	1535	92.3	2709033	1340.27	661.0
Action,Adventure,Crime	2379	12074	699	38.6	1201937	543.72	290.0
...	...	...	...	...	...	...	...
Romance,Sci-Fi,Thriller	707	2011	106	7.1	208632	62.45	0.0
Sci-Fi	801	4032	213	9.9	1453	20.76	53.0
Sci-Fi,Thriller	513	2008	91	5.0	170897	64.51	34.0
Thriller	6133	18133	857	53.6	305642	0.32	418.0
Thriller,War	161	2016	106	6.0	5926	0.00	40.0

207 rows × 7 columns

```
[23] df.groupby('Year').sum()
```

Year	Rank	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
2006	25894	5317	313.5	11848758	3624.46	2641.0
2007	30559	6446	378.1	12949545	4306.23	2967.0
2008	32533	5763	352.8	14326280	5053.22	2813.0
2009	29928	5922	355.0	13044813	5292.26	2799.0
2010	35239	6668	409.6	15166939	5989.65	3504.0
2011	37143	7220	430.8	15169789	5431.96	3580.0
2012	34104	7623	443.2	18254470	6910.29	3791.0

2015	62407	14541	838.5	14697230	8854.12	7016.0
2016	111804	31890	1911.7	14431751	11211.65	15620.0

```
[24] df.groupby(["Genre", "Year"])["Rating"].count()
```

```
Genre                      Year
Action                     2016   2
Action,Adventure           2007   1
                           2012   1
                           2014   1
Action,Adventure,Biography 2015   1
                             ..
Sci-Fi,Thriller            2008   1
Thriller                   2007   1
                           2014   1
                           2016   7
Thriller,War               2016   1
Name: Rating, Length: 583, dtype: int64
```

```
[25] df.groupby(["Genre", "Year"])["Rating"].sum()
```

```
Genre                      Year
Action                     2016   12.3
Action,Adventure           2007    6.2
                           2012    7.0
                           2014    6.0
Action,Adventure,Biography 2015    6.9
                            ...
Sci-Fi,Thriller            2008    5.0
Thriller                   2007    7.1
                           2014    6.7
                           2016   39.8
Thriller,War               2016    6.0
Name: Rating, Length: 583, dtype: float64
```

```
[26] df.groupby(["Genre", "Year"])["Rating"].mean()
```

```
Genre                      Year
Action                     2016    6.150000
Action,Adventure           2007    6.200000
                           2012    7.000000
                           2014    6.000000
Action,Adventure,Biography 2015    6.900000
                                      ...
Sci-Fi,Thriller            2008    5.000000
Thriller                   2007    7.100000
                           2014    6.700000
                           2016    5.685714
Thriller,War               2016    6.000000
Name: Rating, Length: 583, dtype: float64
```

7. Sorting operation

```
[27] df.sort_values('Year')
```

	Rank	Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
773	774	Lady in the Water	Drama,Fantasy,Mystery	Apartment building superintendent Cleveland He...	M. Night Shyamalan	Paul Giamatti, Bryce Dallas Howard, Jeffrey Wr...	2006	110	5.6	82701	42.27	36.0
445	446	Silent Hill	Adventure,Horror,Mystery	A woman, Rose, goes in search for her adopted ...	Christophe Gans	Radha Mitchell, Laurie Holden, Sean Bean,Debor...	2006	125	6.6	184152	46.98	NaN
99	100	The Departed	Crime,Drama,Thriller	An undercover cop and a mole in the police att...	Martin Scorsese	Leonardo DiCaprio, Matt Damon, Jack Nicholson,...	2006	151	8.5	937414	132.37	85.0
550	551	The Break-Up	Comedy,Drama,Romance	In a bid to keep their luxurious condo from th...	Peyton Reed	Jennifer Aniston, Vince Vaughn, Jon Favreau, J...	2006	106	5.8	106381	118.68	45.0

	...	...	...	...	...	...	...	...	...	...	...	...
247	248	20th Century Women	Comedy,Drama	The story of a teenage boy, his mother, and tw...	Mike Mills	Annette Bening, Elle Fanning, Greta Gerwig, Bi...	2016	119	7.4	14708	5.66	83.0
250	251	Bonjour Anne	Comedy,Drama,Romance	Anne is at a crossroads in her life. Long marr...	Eleanor Coppola	Diane Lane, Alec Baldwin, Arnaud Viard, Linda ...	2016	92	4.9	178	0.32	50.0
258	259	Lights Out	Horror	Rebecca must unravel the terror behind her litt...	David F. Sandberg	Teresa Palmer, Gabriel Bateman, Maria Bello,Bi...	2016	81	6.4	69823	67.24	58.0
237	238	The Disappointments Room	Drama,Horror,Thriller	A mother and her young son release unimaginabl...	D.J. Caruso	Kate Beckinsale, Mel Raido, Duncan Joiner, Luc...	2016	85	3.9	4895	2.41	31.0
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

1000 rows × 12 columns

`df.sort_index(axis = 1)`

	Actors	Description	Director	Genre	Metascore	Rank	Rating	Revenue (Millions)	Runtime (Minutes)	Title	Votes	Year
0	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	A group of intergalactic criminals are forced ...	James Gunn	Action,Adventure,Sci-Fi	76.0	1	8.1	333.13	121	Guardians of the Galaxy	757074	2014
1	Noomi Rapace, Logan Marshall-Green, Michael Fa...	Following clues to the origin of mankind, a te...	Ridley Scott	Adventure,Mystery,Sci-Fi	65.0	2	7.0	126.46	124	Prometheus	485820	2012
2	James McAvoy, Anya Taylor-Joy, Haley Lu Richar...	Three girls are kidnapped by a man with a diag...	M. Night Shyamalan	Horror,Thriller	62.0	3	7.3	138.12	117	Split	157606	2016
3	Matthew McConaughey,Reese Witherspoon, Seth Ma...	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet	Animation,Comedy,Family	59.0	4	7.2	270.32	108	Sing	60545	2016
4	Will Smith, Jared Leto, Margot ...	A secret government agency ...	David Ayer	Action,Adventure,Fantasy	40.0	5	6.2	325.02	123	Suicide Squad	393727	2016

	Actors	Description	Director	Genre						Title		
997	Robert Hoffman, Briana Evigan, Cassie Ventura,...	Romantic sparks occur between two dance studen...	Jon M. Chu	Drama,Music,Romance	50.0	998	6.2	58.01	98	Step Up 2: The Streets	70699	200
998	Adam Pally, T.J. Miller, Thomas Middleditch,Sh...	A pair of friends embark on a mission to reuni...	Scot Armstrong	Adventure,Comedy	22.0	999	5.6	NaN	93	Search Party	4881	201
999	Kevin Spacey, Jennifer Garner, Robbie Amell,Ch...	A stuffy businessman finds himself trapped ins...	Barry Sonnenfeld	Comedy,Family,Fantasy	11.0	1000	5.3	19.64	87	Nine Lives	12435	201

1000 rows × 12 columns

8. Dealing with missing values

```
[29] df.isna().sum()
```

```
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
```
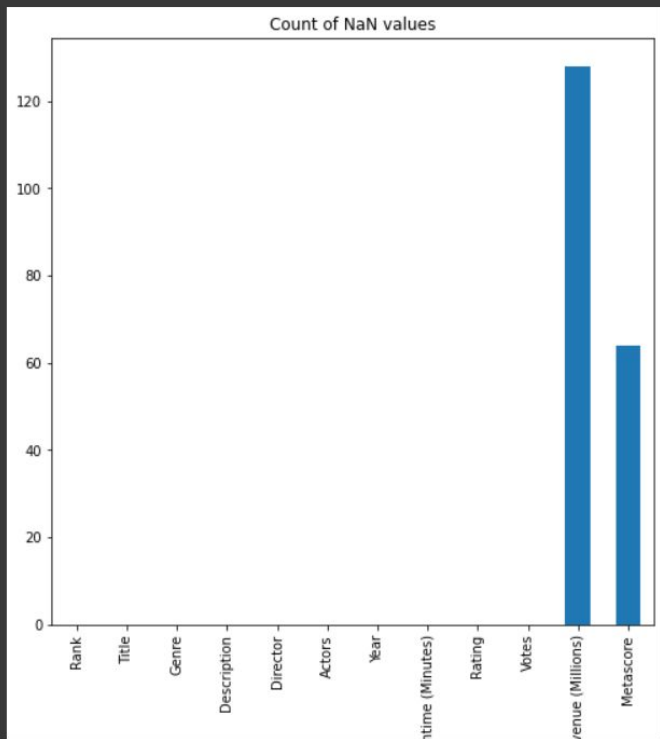
```
[30] import missingno as msno
     msno.matrix(df)
     plt.show()
```

```
plt.figure(figsize=(8,8))
df.isna().sum().plot.bar()
plt.title('Count of NaN values')
plt.show()
```


Count of NaN values

9. Dropping columns and null values

```
[32]  df.drop('Description',axis = 1,inplace=True)
      df.head()
```

	Rank	Title	Genre	Director	Actors	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
0	1	Guardians of the Galaxy	Action,Adventure,Sci-Fi	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	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fa...	2012	124	7.0	485820	126.46	65.0
2	3	Split	Horror,Thriller	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	Christophe Lourdelet	Matthew McConaughey,Reese Witherspoon, Seth Ma...	2016	108	7.2	60545	270.32	59.0
4	5	Suicide Squad	Action,Adventure,Fantasy	David Ayer	Will Smith, Jared Leto, Margot Robbie, Viola D...	2016	123	6.2	393727	325.02	40.0

```
[35]  df.dropna(inplace=True)
```

```
[36]  df.isna().sum()
```

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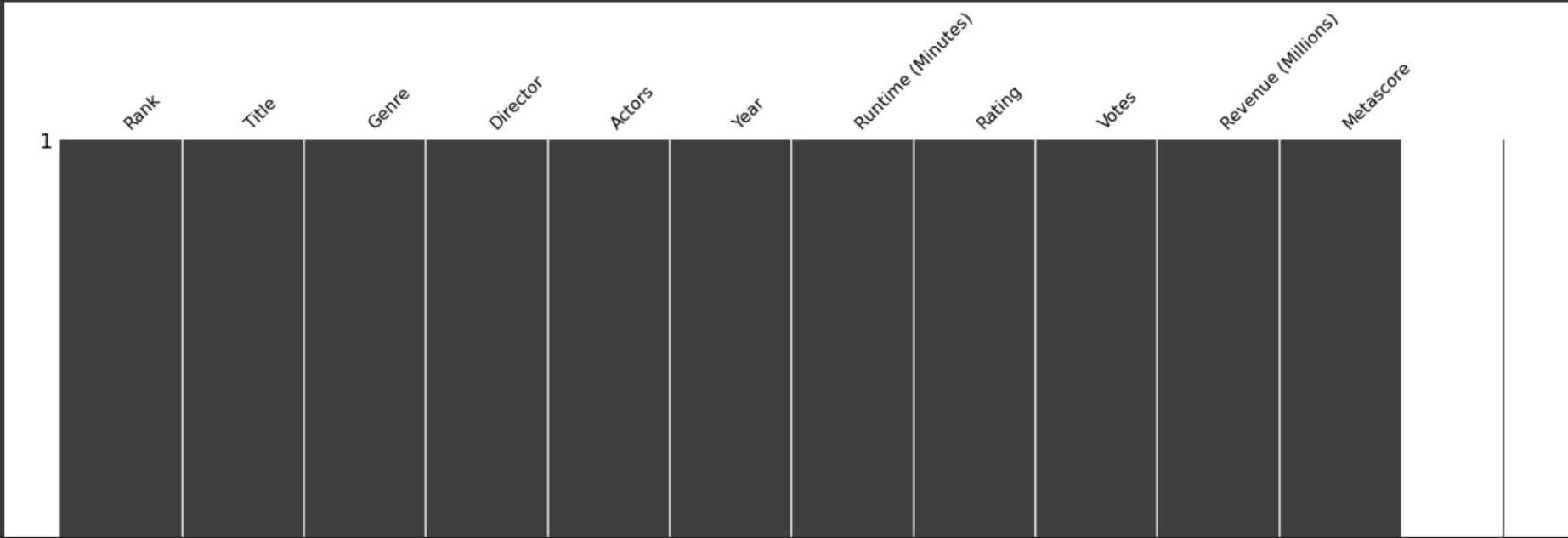
```
      Rating              0
[36]  Votes               0
      Revenue (Millions)   0
      Metascore           0
      dtype: int64
```

```
[37]  msno.matrix(df)
      plt.show()
```

10. Apply( ) functions

```
[38] def split_value(value):
         return value.split(",")
```

```
[39] def divide(number):
         return number / 2
```

```
[40] df.Genre.apply(split_value)
```

```
0           [Action, Adventure, Sci-Fi]
1           [Adventure, Mystery, Sci-Fi]
2                      [Horror, Thriller]
3           [Animation, Comedy, Family]
4           [Action, Adventure, Fantasy]
                     ...
993         [Action, Adventure, Horror]
994                            [Comedy]
996                            [Horror]
997           [Drama, Music, Romance]
999         [Comedy, Family, Fantasy]
Name: Genre, Length: 838, dtype: object
```

```
[41] df.Rating.apply(divide)
```

```
0        4.05
1        3.50
2        3.65
3        3.60
4        3.10
```

```
996    2.75
997    3.10
999    2.65
Name: Rating, Length: 838, dtype: float64
```

[42] `df[['Rating', 'Votes','Runtime (Minutes)','Revenue (Millions)', 'Metascore']].apply(divide)`

	Rating	Votes	Runtime (Minutes)	Revenue (Millions)	Metascore
0	4.05	378537.0	60.5	166.565	38.0
1	3.50	242910.0	62.0	63.230	32.5
2	3.65	78803.0	58.5	69.060	31.0
3	3.60	30272.5	54.0	135.160	29.5
4	3.10	196863.5	61.5	162.510	20.0
...	...	...	...	...	...
993	2.95	70450.0	48.5	30.065	18.5
994	3.35	82044.0	44.0	27.360	24.0
996	2.75	36576.0	47.0	8.770	23.0
997	3.10	35349.5	49.0	29.005	25.0
999	2.65	6217.5	43.5	9.820	5.5

838 rows × 5 columns

[43] `df.corr()`

	Rank	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
Rank	1.000000	0.212800	0.254783	0.243125	0.303284	0.273170	0.195000

838 rows × 5 columns

```
df.corr()
```

	Rank	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
Rank	1.000000	-0.312809	-0.254783	-0.243125	-0.303284	-0.273170	-0.195909
Year	-0.312809	1.000000	-0.101933	-0.145703	-0.362445	-0.129198	-0.062303
Runtime (Minutes)	-0.254783	-0.101933	1.000000	0.374566	0.399298	0.281721	0.221397
Rating	-0.243125	-0.145703	0.374566	1.000000	0.517452	0.217106	0.672731
Votes	-0.303284	-0.362445	0.399298	0.517452	1.000000	0.636833	0.332674
Revenue (Millions)	-0.273170	-0.129198	0.281721	0.217106	0.636833	1.000000	0.142397
Metascore	-0.195909	-0.062303	0.221397	0.672731	0.332674	0.142397	1.000000

```python
plt.figure(figsize=(12,12))
sns.heatmap(df.corr(), annot=True, fmt =".3f",linewidths= 2, square=True , cmap='plasma')
plt.title("Correlation",size = 15)
plt.show()
```

1.0

Correlation