

```
In [1]: import pandas as pd

In [2]: import numpy as np

In [8]: import matplotlib.pyplot as plt

In [5]: import seaborn as sns

In [6]: df = pd.read_csv(r'/home/nineleaps/Documents/IMDB-Movie-Data.csv')

In [7]: df.head()
```

	Rank	Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore	Unnamed: 12
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	757074.0	333.13	76.0	NaN
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	485820.0	126.46	65.0	NaN
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	157606.0	138.12	62.0	NaN
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	60545.0	270.32	59.0	NaN
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	393727.0	325.02	40.0	NaN

```
In [10]: df.shape

Out[10]: (1000, 13)
```

```
In [11]: print('Number of rows:', df.shape[0])
print('Number of columns:', df.shape[1])

Number of rows: 1000
Number of columns: 13
```

```
In [13]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 13 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  object
7   Runtime (Minutes)     1000 non-null  int64
8   Rating                1000 non-null  float64
9   Votes                 1000 non-null  float64
10  Revenue (Millions)    873 non-null   float64
11  Metascore             936 non-null   float64
12  Unnamed: 12           2 non-null     float64
dtypes: float64(5), int64(2), object(6)
memory usage: 101.7+ KB
```

```
In [20]: print('Any missing values?', df.isnull().values.any())

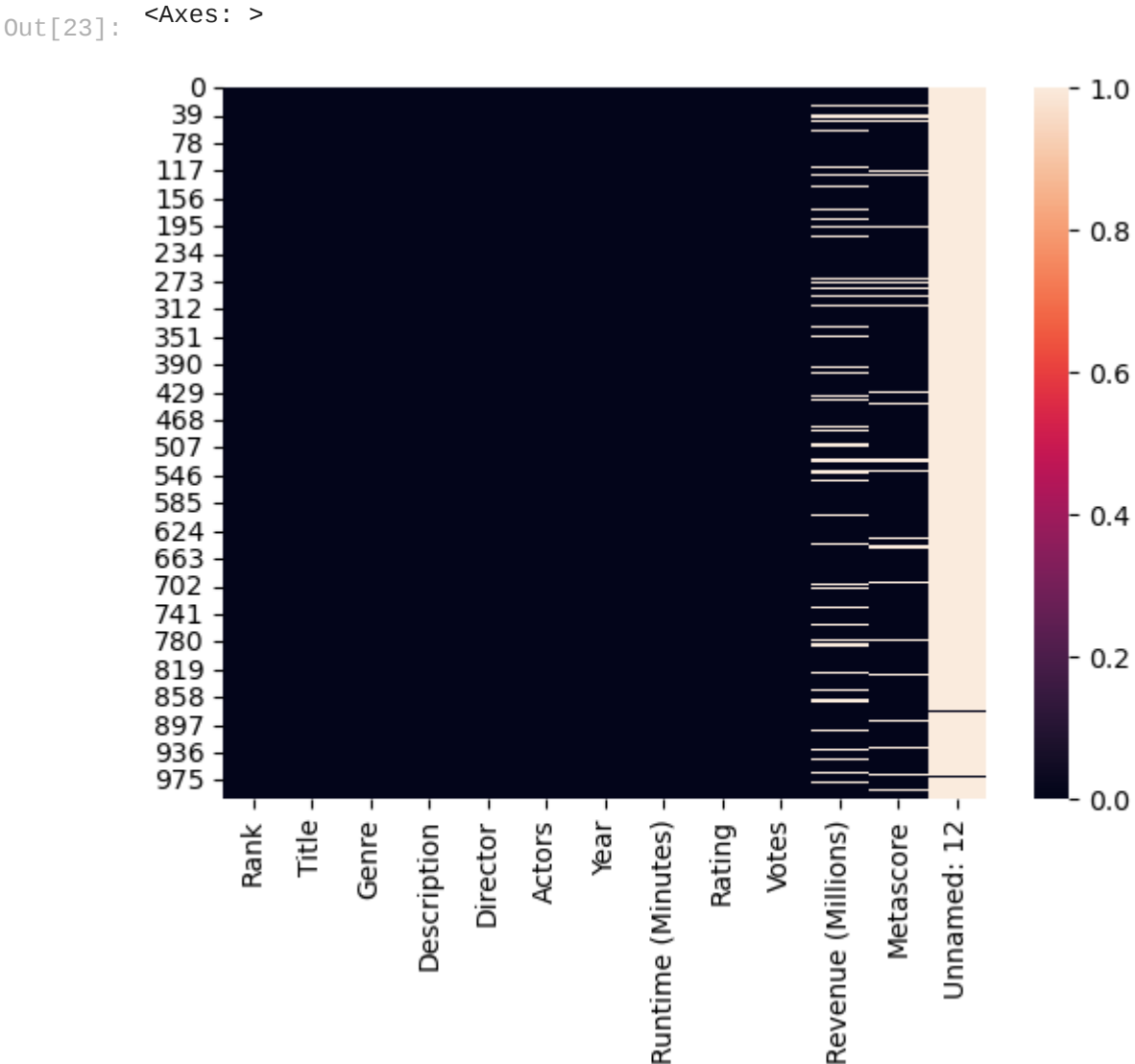
Any missing values? True
```

```
In [22]: df.isnull().sum()
```

Rank	0
Title	0
Genre	0
Description	0
Director	0
Actors	0
Year	0
Runtime (Minutes)	0
Rating	0
Votes	0
Revenue (Millions)	127
Metascore	64
Unnamed: 12	998
dtype:	int64

Missing values by Heatmap

```
In [23]: sns.heatmap(df.isnull())
```



Dropping a unnamed column which has maximum null values

```
In [27]: df1= df.drop(df.columns[12], axis=1)

In [28]: df1.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	757074.0	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	485820.0	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	157606.0	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	60545.0	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	393727.0	325.02	40.0

```
In [31]: percentage_missing = df1.isnull().sum()*100/len(df1)
percentage_missing
```

Rank	0.0
Title	0.0
Genre	0.0
Description	0.0
Director	0.0
Actors	0.0
Year	0.0
Runtime (Minutes)	0.0
Rating	0.0
Votes	0.0
Revenue (Millions)	12.7
Metascore	6.4
dtype:	float64

Dropping the rows of null values

```
In [33]: df1.dropna(axis=0, inplace= True)

In [36]: percentage_missing_new = df1.isnull().sum()*100/len(df1)
percentage_missing_new
```

Rank	0.0
Title	0.0
Genre	0.0
Description	0.0
Director	0.0
Actors	0.0
Year	0.0
Runtime (Minutes)	0.0
Rating	0.0
Votes	0.0
Revenue (Millions)	0.0
Metascore	0.0
dtype:	float64

```
In [37]: df1.shape

Out[37]: (838, 12)
```

Check for duplicate values

```
In [41]: dup_value = df1.duplicated().any()

In [42]: print('Are there any dupliacted value in dataset?', dup_value)

Are there any dupliacted value in dataset? False

In [ ]: # If there is any duplicated value then
# df1.drop_duplicates()
```