```
In [ ]:   #1. Import Libraries

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

In [ ]:   #2. Load Dataset

In [3]:   df = pd.read_csv(r"C:\Users\Dell\Downloads\IMDB-Movie-Data.csv")

In [9]:   df.head(10)
```

`Out[9]:`

	Rank	Title	Genre	Description	Director	
0	1	Guardians of the Galaxy	Action,Adventure,Sci-Fi	A group of intergalactic criminals are forced ...	James Gunn	Chris Dies Coope
1	2	Prometheus	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott	Noor Logan Green, M
2	3	Split	Horror,Thriller	Three girls are kidnapped by a man with a diag...	M. Night Shyamalan	James McA Taylor-Joy
3	4	Sing	Animation,Comedy,Family	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet	McConaug Withersp
4	5	Suicide Squad	Action,Adventure,Fantasy	A secret government agency recruits some of th...	David Ayer	Will Sr Leto, Marg
5	6	The Great Wall	Action,Adventure,Fantasy	European mercenaries searching for black powde...	Yimou Zhang	Matt Da Jing, Will
6	7	La La Land	Comedy,Drama,Music	A jazz pianist falls for an aspiring actress i...	Damien Chazelle	Ryan Gosli Stone, D
7	8	Mindhorn	Comedy	A has-been actor best known for playing the ti...	Sean Foley	Essie Dav Riseboro
8	9	The Lost City of Z	Action,Adventure,Biography	A true-life drama, centering on British explor...	James Gray	Charlie Robert Sier
9	10	Passengers	Adventure,Drama,Romance	A spacecraft traveling to a distant	Morten Tyldum	Jennifer Chris Pra

Rank	Title	Genre	Description	Director
			colony pla...	

```
In [11]:  df.tail(10)
```

	Rank	Title	Genre	Description	Director	
990	991	Underworld: Rise of the Lycans	Action,Adventure,Fantasy	An origins story centered on the centuries-old...	Patrick Tatopoulos	Rhon Michael Bil S
991	992	Taare Zameen Par	Drama,Family,Music	An eight-year-old boy is thought to be a lazy ...	Aamir Khan	Darsheel Aam Tanay C
992	993	Take Me Home Tonight	Comedy,Drama,Romance	Four years after graduation, an awkward high s...	Michael Dowse	Topher Anna Fa Fogler
993	994	Resident Evil: Afterlife	Action,Adventure,Horror	While still out to destroy the evil Umbrella C...	Paul W.S. Anderson	Milla Jc Al We M
994	995	Project X	Comedy	3 high school seniors throw a birthday party t...	Nima Nourizadeh	Thoma Oliver Jonathan
995	996	Secret in Their Eyes	Crime,Drama,Mystery	A tight-knit team of rising investigators, alo...	Billy Ray	Chiwetel Nicole I Julia Rc
996	997	Hostel: Part II	Horror	Three American college students studying abroa...	Eli Roth	Lauren C I Matarazz
997	998	Step Up 2: The Streets	Drama,Music,Romance	Romantic sparks occur between two dance studen...	Jon M. Chu	Robert H Briana Ve
998	999	Search Party	Adventure,Comedy	A pair of friends embark on a mission to reuni...	Scot Armstrong	Adam F Miller, Middledit
999	1000	Nine Lives	Comedy,Family,Fantasy	A stuffy businessman finds himself	Barry Sonnenfeld	Kevin Jennifer

	Rank	Title	Genre	Description	Director	
				trapped ins...		Am

```
In [15]: df.shape
```

```
Out[15]: (1000, 12)
```

```
In [17]: df.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
```

```
In [19]: df.isnull().sum()
```

```
Out[19]: 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
```

```
In [21]: df.dtypes
```

```
Out[21]:  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
```

In [23]:
```python
df['Revenue (Millions)'] = df['Revenue (Millions)'].fillna(df['Revenue (Millio
df['Metascore'] = df['Metascore'].fillna(df['Metascore'].mean())
```

In [25]:
```python
df.isnull().sum()
```

```
Out[25]:  Rank                0
          Title               0
          Genre               0
          Description         0
          Director            0
          Actors              0
          Year                0
          Runtime (Minutes)   0
          Rating              0
          Votes               0
          Revenue (Millions)  0
          Metascore           0
          dtype: int64
```

In [27]:
```python
df.to_csv("IMDB-Movie-Data-cleaned.csv", index=False)
```

In [29]:
```python
df
```

`Out[29]:`

	Rank	Title	Genre	Description	Director	
0	1	Guardians of the Galaxy	Action,Adventure,Sci-Fi	A group of intergalactic criminals are forced ...	James Gunn	Ch Die Coo
1	2	Prometheus	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott	No Log Green,
2	3	Split	Horror,Thriller	Three girls are kidnapped by a man with a diag...	M. Night Shyamalan	James M Taylor-J
3	4	Sing	Animation,Comedy,Family	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet	McConau Wither
4	5	Suicide Squad	Action,Adventure,Fantasy	A secret government agency recruits some of th...	David Ayer	Will Leto, Ma
...	...	...	...	...	...	
995	996	Secret in Their Eyes	Crime,Drama,Mystery	A tight-knit team of rising investigators, alo...	Billy Ray	Chiv Nicole K
996	997	Hostel: Part II	Horror	Three American college students studying abroa...	Eli Roth	Laur Heather
997	998	Step Up 2: The Streets	Drama,Music,Romance	Romantic sparks occur between two dance studen...	Jon M. Chu	Robe Br Cassi
998	999	Search Party	Adventure,Comedy	A pair of friends embark on a mission to reuni...	Scot Armstrong	Ada Mi Midd
999	1000	Nine Lives	Comedy,Family,Fantasy	A stuffy	Barry	Ke

	Rank	Title	Genre	Description	Director
				businessman finds himself trapped ins...	Sonnenfeld

Jenr
Robbie

1000 rows × 12 columns

In [31]:
```python
#Check For Duplicate Data
dup_data=df.duplicated().any()
print("Are there any duplicated values in data?",dup_data)
```

Are there any duplicated values in data? False

In [33]:
```python
#Overall Statistics About The DataFrame
df.describe()
```

Out[33]:

	Rank	Year	Runtime (Minutes)	Rating	Votes	Re (Mil
count	1000.000000	1000.000000	1000.000000	1000.000000	1.000000e+03	1000.0
mean	500.500000	2012.783000	113.172000	6.723200	1.698083e+05	82.9
std	288.819436	3.205962	18.810908	0.945429	1.887626e+05	96.4
min	1.000000	2006.000000	66.000000	1.900000	6.100000e+01	0.0
25%	250.750000	2010.000000	100.000000	6.200000	3.630900e+04	17.4
50%	500.500000	2014.000000	111.000000	6.800000	1.107990e+05	60.3
75%	750.250000	2016.000000	123.000000	7.400000	2.399098e+05	99.1
max	1000.000000	2016.000000	191.000000	9.000000	1.791916e+06	936.6

In [35]:
```python
#Calculates mean, median, mode for ratings

# Summary Statistics for Ratings
rating_mean = df['Rating'].mean()
rating_median = df['Rating'].median()
rating_mode = df['Rating'].mode()[0]

print("\nRating Summary:")
print(f"Mean: {rating_mean:.2f}")
print(f"Median: {rating_median}")
print(f"Mode: {rating_mode}")
```
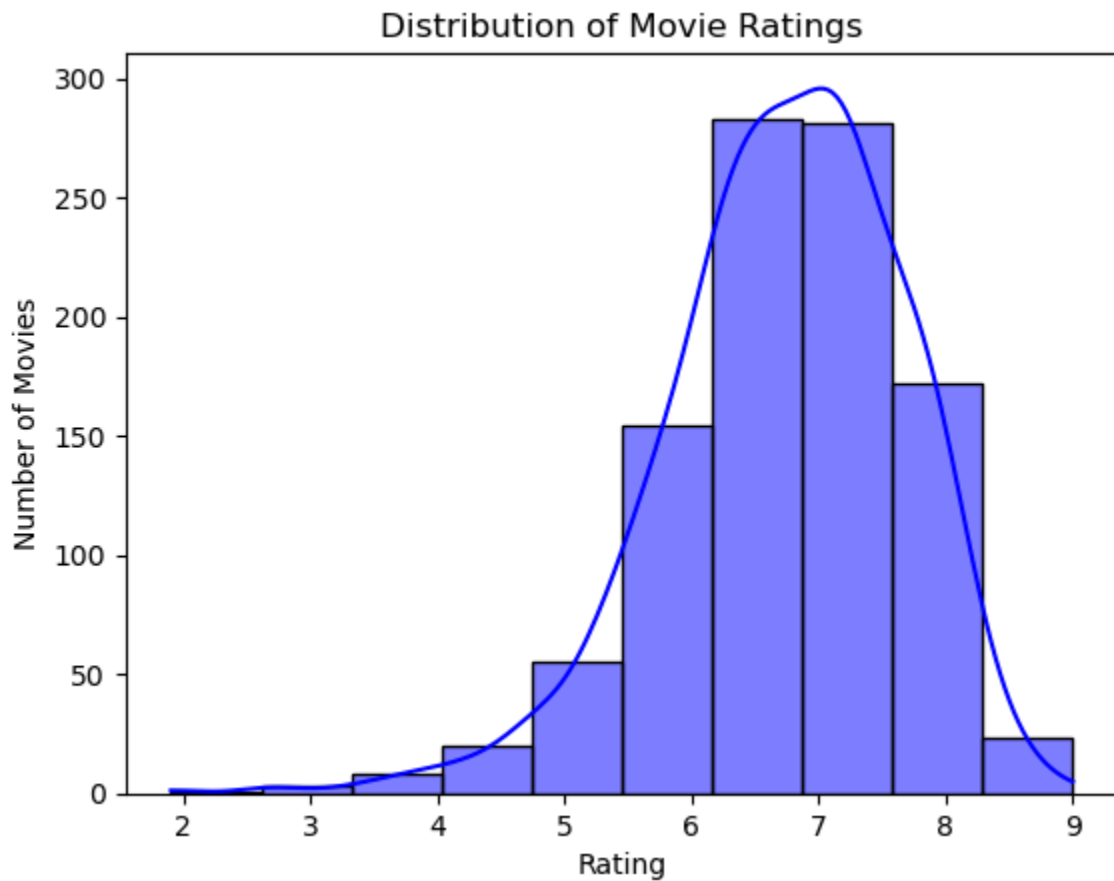
Rating Summary:
Mean: 6.72
Median: 6.8
Mode: 7.1

In [43]:
```python
# 5. Visualize Rating Distribution
```

```
sns.histplot(df['Rating'], bins=10, kde=True, color='blue')
plt.title("Distribution of Movie Ratings")
plt.xlabel("Rating")
plt.ylabel("Number of Movies")
plt.show()
```


Distribution of Movie Ratings

```
plt.figure(figsize=(6,4))
sns.boxplot(x=df['Rating'], color='lightgreen')
plt.title("Boxplot of Movie Ratings")
plt.show()
```

Boxplot of Movie Ratings
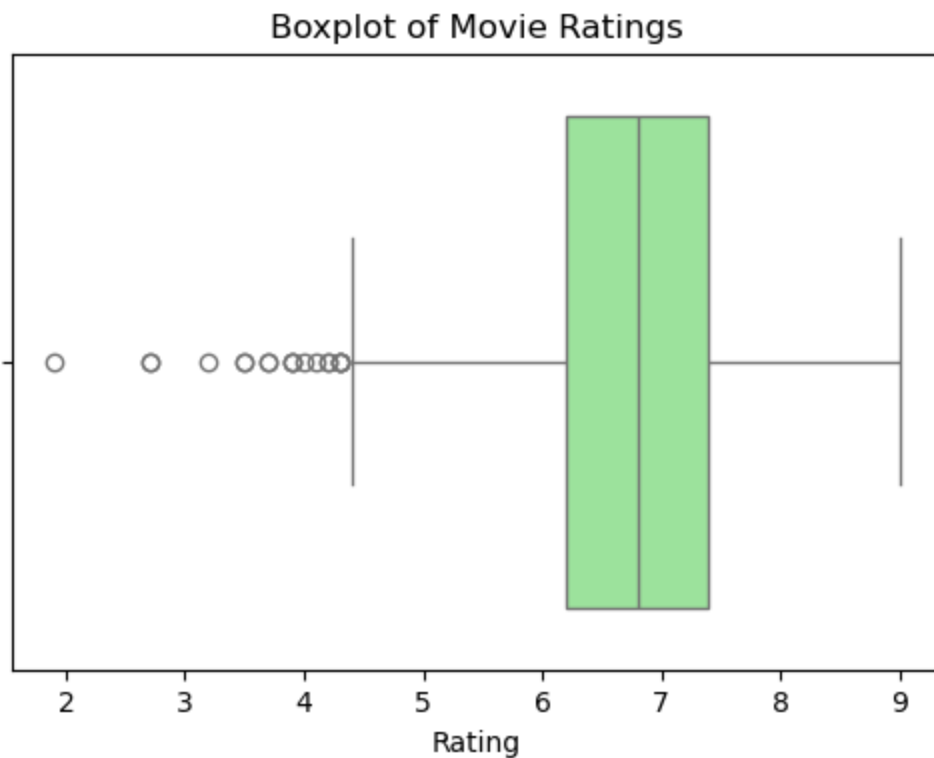
In [47]:
```python
#Top 10 Movies by Rating
top_movies = df.sort_values(by='Rating', ascending=False).head(10)[['Title', '
top_movies
```

Out[47]:

	Title	Rating	Genre
54	The Dark Knight	9.0	Action,Crime,Drama
80	Inception	8.8	Action,Adventure,Sci-Fi
117	Dangal	8.8	Action,Biography,Drama
36	Interstellar	8.6	Adventure,Drama,Sci-Fi
96	Kimi no na wa	8.6	Animation,Drama,Fantasy
249	The Intouchables	8.6	Biography,Comedy,Drama
133	Whiplash	8.5	Drama,Music
64	The Prestige	8.5	Drama,Mystery,Sci-Fi
99	The Departed	8.5	Crime,Drama,Thriller
991	Taare Zameen Par	8.5	Drama,Family,Music

In [49]:
```python
df_genres = df.assign(Genre=df['Genre'].str.split(',')).explode('Genre')
df_genres['Genre'] = df_genres['Genre'].str.strip()
avg_genre_rating = df_genres.groupby('Genre')['Rating'].mean().sort_values(asc
avg_genre_rating
```

```
Out[49]: Genre
         War          7.353846
         Animation    7.324490
         Biography    7.290123
         History      7.127586
         Music        7.075000
         Sport        7.011111
         Drama        6.953801
         Musical      6.940000
         Mystery      6.886792
         Crime        6.786667
         Name: Rating, dtype: float64
```

In [51]:
```python
sns.set(style="whitegrid")
```

In [53]:
```python
# Plot Top 10 Movies

plt.figure(figsize=(10, 6))
sns.barplot(x='Rating', y='Title', data=top_movies, palette='viridis')
plt.title("Top 10 Rated Movies", fontsize=16)
plt.xlabel("Rating")
plt.ylabel("Movie Title")
plt.show()
```

```
C:\Users\Dell\AppData\Local\Temp\ipykernel_21616\1088078136.py:2: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same e
ffect.

  sns.barplot(x='Rating', y='Title', data=top_movies, palette='viridis')
```
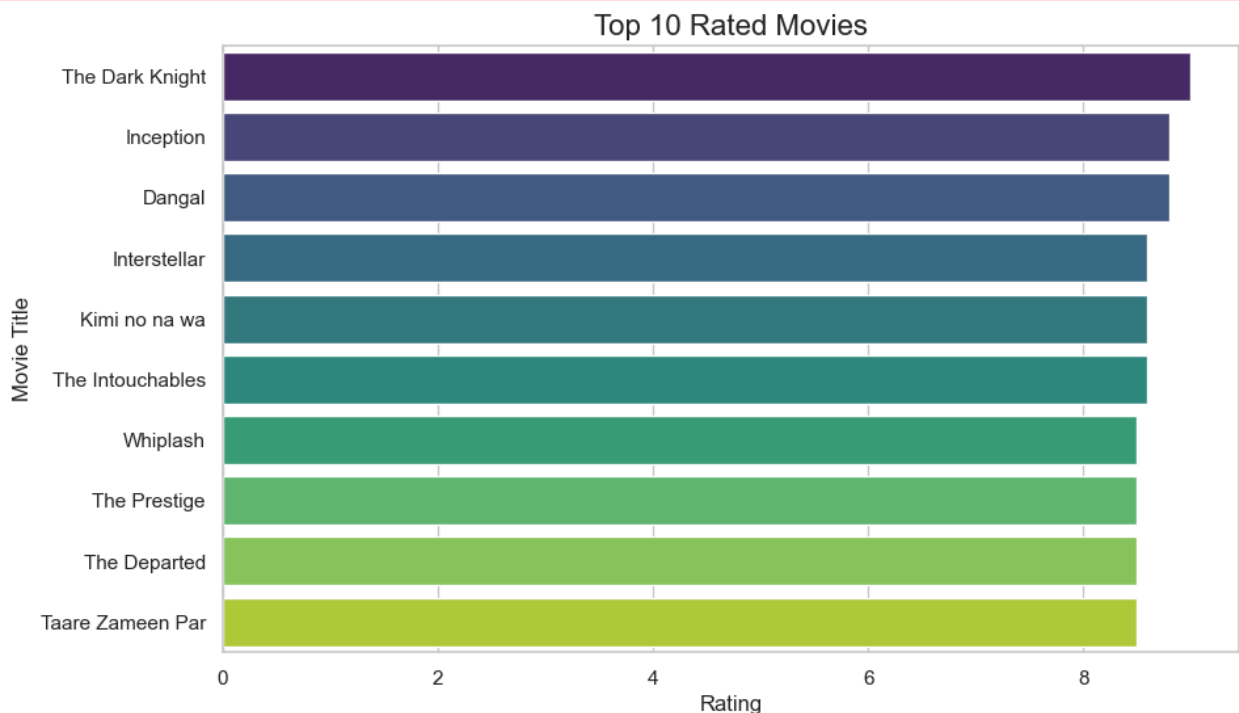
```python
# Plot Top 10 Genres
plt.figure(figsize=(10, 6))
sns.barplot(x=avg_genre_rating.values, y=avg_genre_rating.index, palette='magm
plt.title("Top 10 Genres by Average Rating", fontsize=16)
plt.xlabel("Average Rating")
plt.ylabel("Genre")
plt.show()
```

C:\Users\Dell\AppData\Local\Temp\ipykernel_21616\3240967074.py:3: FutureWarnin
g:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same e
ffect.

  sns.barplot(x=avg_genre_rating.values, y=avg_genre_rating.index, palette='mag
ma')
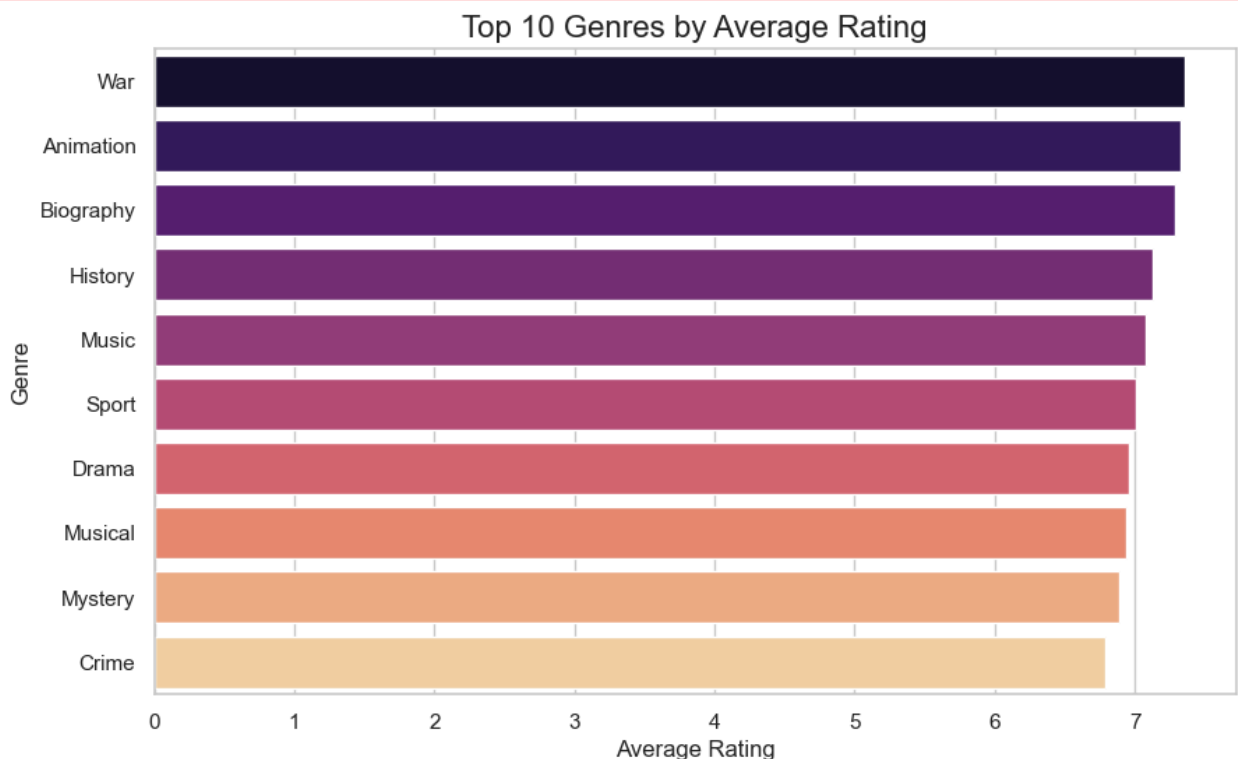
```python
df.to_csv("IMDB-Movie-Data-cleaned.csv", index=False)
print("\nCleaned data saved as 'IMDB-Movie-Data-cleaned.csv'")
```

Cleaned data saved as 'IMDB-Movie-Data-cleaned.csv'

In [ ]: