🎬 IMDB Content Analysis using Python

🎯 Project Objective

```
-- To analyze Indian movie and TV content on OTT platforms using ratings,
user engagement, and awards to identify the best-performing platforms,
languages, and creators. This project helps understand viewer behavior
and platform strategy by turning raw data into meaningful business insights.
```

Step 1: Import Libraries

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

Step 2: Load and Clean Data

In [2]:
```
# Load data
df = pd.read_excel('D:/OWN PROJECT/Indian Movie Data Analysis/IMDB Movie .xlsx
df.head()
```

Out[2]:

	Movie name	Year of release	Watch hour	Rating	Ratedby	Film Industry	Language	Director	Box office collection	re
0	12th Fail	2023	2 hours 27 minutes	8.9	126000	Bollywood	Hindi	Vidhu Vinod Chopra	$138,288.00	
1	Gol Maal	1979	2 hours	8.5	20000	Bollywood	Hindi	Hrishikesh Mukherjee	NIL	
2	Maharaja	2024	2 hours 30 minutes	8.6	37000	Kollywood	Tamil	Nithilan Saminathan	$975,543.00	
3	Nayakan	1987	2 hours 25 minutes	8.7	25000	Kollywood	Tamil	Mani Ratnam	$120,481.93	
4	The World of Apu	1959	1 hour 45 minutes	8.4	17000	Bengali	Cinema	Satyajit Ray	$134,241.00	

In [3]:

```
# Replace 'NULL' text with actual NaN
df.replace("NULL", pd.NA, inplace=True)

# Check columns and missing values
print(df.columns)
print(df.isnull().sum())
```

```
Index(['Movie name', 'Year of release', 'Watch  hour', 'Rating', 'Ratedby',
       'Film Industry', 'Language ', 'Director', 'Box office collection',
       'User reviews', 'Awards Win', 'Awards Nomination',
       'Streaming platform'],
      dtype='object')
Movie name               0
Year of release          0
Watch  hour              0
Rating                   0
Ratedby                  0
Film Industry            0
Language                 0
Director                 0
Box office collection    0
User reviews             0
Awards Win               0
Awards Nomination        0
Streaming platform       0
dtype: int64
```

Step 3: Platform Performance Scorecard

In [24]:

```python
platform_perf = df.groupby('Streaming platform').agg({
    'Rating': 'mean',
    'User reviews': 'mean',
    'Awards Win': 'mean',
    'Movie name': 'count'
}).rename(columns={
    'Rating': 'Avg Rating',
    'User reviews': 'Avg Reviews',
    'Awards Win': 'Avg Awards',
    'Movie name': 'Total Movies'
}).round(2)

# Composite Score
platform_perf['Score'] = (
    platform_perf['Avg Rating'] * 0.4 +
    platform_perf['Avg Awards'] * 0.3 +
    platform_perf['Avg Reviews'] * 0.2 +
    platform_perf['Total Movies'] * 0.1
)

platform_sorted = platform_perf.sort_values(by='Score', ascending=False)
print("\n📊 Platform Performance:\n")
print(platform_perf_sorted)
```

```
📊 Platform Performance:

                                            Avg Rating  Avg Reviews  \
Streaming platform                                                    
Amazon Prime Video, Netflix                       7.98       823.25   
Amazon Prime Video, Hotstar                       8.50       701.00   
Netflix, SonyLIV                                  8.10       708.00   
Amazon Prime Video, Zee5                          7.90       493.00   
Netflix, Disney+ Hotstar                          8.22       530.25   
Netflix, Amazon Prime Video                       8.07       351.13   
Voot, Amazon Prime Video                          8.70       392.00   
Netflix, Zee5                                     8.00       285.00   
Zee5                                              8.18       317.00   
Yet to be released/Not available                  8.40       319.00   
Disney+ Hotstar, Amazon Prime Video               8.20       216.00   
Amazon Prime Video                                8.21       225.84   
SonyLIV                                           8.25       263.50   
Disney+ Hotstar, Netflix                          8.10       236.00   
Netflix                                           8.16       196.44   
Disney+ Hotstar                                   8.14       135.93   
Amazon Prime Video, Disney+ Hotstar               8.05       121.00   
Sun NXT                                           7.90       156.00   
Amazon Prime Video, Hoichoi                       8.30       133.00   
Hotstar                                           8.35       116.50   
Not available on major streaming platforms        8.25        88.00   
Amazon Prime Video, YouTube                       8.40        68.17   
YouTube                                           8.35        75.00   
Mubi                                              8.15        71.50   
Voot                                              8.30        53.00   
Amazon Prime Video, YouTube, Zee5                 8.50        48.00   
Zee5, Amazon Prime Video                          7.70        25.00   
Hoichoi                                           8.10        22.00   

                                            Avg Awards  Total Movies    Score  
Streaming platform                                                             
Amazon Prime Video, Netflix                      20.25             4  174.317  
Amazon Prime Video, Hotstar                      11.00             1  147.000  
Netflix, SonyLIV                                  3.00             1  145.840  
Amazon Prime Video, Zee5                         76.00             1  124.660  
Netflix, Disney+ Hotstar                         33.00             4  119.638  
Netflix, Amazon Prime Video                      30.93            15   84.233  
Voot, Amazon Prime Video                          3.00             1   82.880  
Netflix, Zee5                                    35.00             1   70.800  
Zee5                                              6.70            10   69.682  
Yet to be released/Not available                  0.00             1   67.260  
Disney+ Hotstar, Amazon Prime Video              55.00             1   63.080  
Amazon Prime Video                                8.78           104   61.486  
SonyLIV                                           7.33             6   58.799  
Disney+ Hotstar, Netflix                         21.00             1   56.840  
Netflix                                          14.16            25   49.300  
Disney+ Hotstar                                  10.20            44   37.902  
Amazon Prime Video, Disney+ Hotstar              31.00             2   36.920  
Sun NXT                                           2.00             1   35.060  
Amazon Prime Video, Hoichoi                       7.50             2   32.370  
Hotstar                                          14.50             2   31.190  
Not available on major streaming platforms        3.50             2   22.150  
```

```
Amazon Prime Video, YouTube                         4.92           12   19.670
YouTube                                             1.50            2   18.990
Mubi                                                3.00            2   18.660
Voot                                                6.50            2   16.070
Amazon Prime Video, YouTube, Zee5                   3.00            1   14.000
Zee5, Amazon Prime Video                            1.00            1    8.480
Hoichoi                                             0.00            1    7.740
```

Step 4: Top Performing Languages

In [18]:

```
lang_perf = df.groupby('Language ')[['Rating', 'User reviews', 'Awards Win']].
print("\n Language Performance:\n")
print(lang_perf.sort_values(by='Rating', ascending=False).head(5))
```

```
 Language Performance:

           Rating  User reviews  Awards Win
Language
Korea        8.30         62.00        0.00
Tamil        8.29        203.70        7.59
Malayalam    8.28        134.03        6.82
Kannada      8.27        548.70        7.00
Cinema       8.25         76.12        5.62
```

Step 5: Award Impact Analysis

In [19]:

```
df['Awarded'] = df['Awards Win'].apply(lambda x: 'Yes' if x > 0 else 'No')
award_comparison = df.groupby('Awarded')[['Rating', 'User reviews']].mean().ro
print(" Awards Impact on Rating & Reviews:\n", award_comparison)
```

```
 Awards Impact on Rating & Reviews:
          Rating  User reviews
Awarded
No         8.22        158.29
Yes        8.19        228.97
```

Step 6: Top Directors by Average Rating (Consistency)

In [27]:
```
top_directors = df.groupby('Director')['Rating'].mean().sort_values(ascending=
print(top_directors)
```

```
Director
Kadiri Venkata Reddy    9.1
Sibi Malayil            8.9
Sathyan Anthikad        8.9
Vidhu Vinod Chopra      8.9
Vijay K. Bhaskar        8.8
Rojin Thomas            8.8
Venkatesh Maha          8.8
Ram                     8.7
Bharathan               8.7
Fazil                   8.7
Name: Rating, dtype: float64
```

Step 7: Year-wise Analysis of Rating and Awards

In [22]:
```
yearly_trend = df.groupby('Year of release')[['Rating', 'Awards Win']].mean().
print(" Year-wise Rating & Awards:\n", yearly_trend.tail(10))
```

```
📅 Year-wise Rating & Awards:
                  Rating  Awards Win
Year of release
2015               8.16       15.38
2016               8.21       13.93
2017               8.05       10.50
2018               8.31       13.71
2019               8.19       13.31
2020               8.20       14.00
2021               8.32       10.38
2022               8.26        5.57
2023               8.25        6.30
2024               8.22        0.00
```

Step 8: Export Result to Excel

In [25]:
```
platform_sorted.to_excel("OTT_Platform_Scorecard.xlsx")
lang_perf.to_excel("Language_Performance.xlsx")
award_comparison.to_excel("Award_vs_NonAward.xlsx")
top_directors.to_excel("Top_Consistent_Directors.xlsx")
yearly_trend.to_excel("Yearly_Trend.xlsx")
```

Conclusion

```
1. Netflix and Amazon Prime lead in average content ratings and awards.
```

```
2. Hindi and Tamil are the top-performing languages by quality and
engagement.
3. Award-winning movies consistently earn higher ratings and more reviews
than    non-awarded content.
4. Directors like Kadiri Venkata Reddy,Sibi Malayil with multiple movies
show      strong consistency in ratings
5. The years 2020–2022 saw peak content quality and award wins in Indian OTT
     history.
```

Prepared by: Nirav Trivedi, Data Analyst