

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

```
In [30]:
data1= pd.read_csv(r"C:\\Users\\Sanjay\\Desktop\\csv\\archive3\\BollywoodM
data1
```

0	tt0118578	Albela	2001	20 Apr 2001	Romance	Honey Irani (screenplay)   Honey Irani (story)...	Govinda Aishwarya Ra Bachchan   Jackie Shroff
1	tt0169102	Lagaan: Once Upon a Time in India	2001	08 May 2002	Adventure   Drama   Musical	Ashutosh Gowariker (story)   Ashutosh Gowariker...	Aamir Khan Gracy Singh Rache Shelley Pa...
2	tt0187279	Meri Biwi Ka Jawab Nahin	2004	02 Jul 2004	Action   Comedy	NaN	Akshay Kumar Sridevi Gulshar Grover Laxm...
3	tt0222024	Hum Tumhare Hain Sanam	2002	24 May 2002	Drama   Romance	K.S. Adiyaman   Arun Kumar (assistant dialogue...	Shaf Rukt Khan Madhur Dixit Salmar Khan J...
4	tt0227194	One 2 Ka 4	2001	30 Mar 2001	Action   Comedy   Drama	Sanjay Chhel   Raaj Kumar Dahima (screenplay) ...	Shaf Rukt Khan Juh Chawla Jackie Shroff J...
...	...	...	...	...	...	...	...
1279	tt4121522	Shuruaat Ka Interval	2014	15 Aug 2014	NaN	Aarti S. Bagdi   Amrit Raj Gupta   Krishan Hoo...	Ki Ch Raj Avid
1280	tt4187650	Trip to Bhangarh	2014	28 Aug 2014	Horror	NaN	Chandra Prakash Dwivedi (screen play and dialo...
1281	tt4190220	Zed Plus	2014	28 Nov 2014	NaN	NaN	...
1282	tt4219300	Gollu aur Pappu	2014	21 Nov 2014	Comedy	NaN	...
1283	tt4228746	Zid	2014	28 Nov 2014	Thriller	Vivek Agnihotri   Rohit Malhotra	...

1284 rows x 10 columns

1284 rows × 10 columns

```
In [34]:
data1.head()
```

	imdbId	title	releaseYear	releaseDate	genre	writers	actors
0	tt0118578	Albela	2001	20 Apr 2001	Romance	Honey Irani (screenplay)   Honey Irani (story)...	Govinda Aishwarya Ra Bachchan   Jackie Shroff
1	tt0169102	Lagaan: Once Upon a Time in India	2001	08 May 2002	Adventure   Drama   Musical	Ashutosh Gowariker (story)   Ashutosh Gowariker...	Aamir Khan Gracy Singh Rache Shelley Pa...
2	tt0187279	Meri Biwi Ka Jawab Nahin	2004	02 Jul 2004	Action   Comedy	NaN	Akshay Kumar Sridevi Gulshan Grover Laxm...
3	tt0222024	Hum Tumhare Hain Sanam	2002	24 May 2002	Drama   Romance	K.S. Adiyaman   Arun Kumar (assistant dialogue...	Shaf Rukt Khan Madhur Dixit Salmar Khan J...
4	tt0227194	One 2 Ka 4	2001	30 Mar 2001	Action   Comedy   Drama	Sanjay Chhel   Raaj Kumar Dahima (screenplay) ...	Shaf Rukt Khan Juh Chawla Jackie Shroff J...

```
In [33]:
data1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1284 entries, 0 to 1283
Data columns (total 10 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   imdbId          1284 non-null  object
1   title           1284 non-null  object
2   releaseYear     1284 non-null  int64
3   releaseDate     1231 non-null  object
4   genre           1282 non-null  object
5   writers         1165 non-null  object
6   directors       1280 non-null  object
7   sequel          1281 non-null  float64
8   hitFlop         1284 non-null  int64
dtypes: float64(1), int64(2), object(7)
memory usage: 100.4+ KB
```

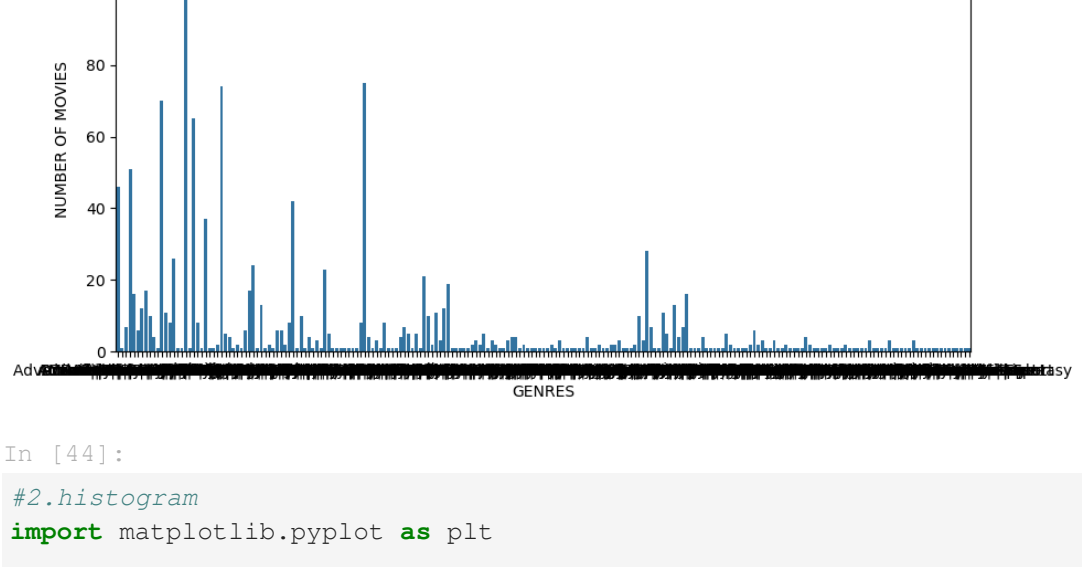
```
In [36]:
# genres analysis
Genres=data1["genre"].value_counts()
print(Genres)

genre
Drama          113
Comedy | Drama  75
Comedy | Drama  74
Action | Crime | Drama  70
Comedy          65
...
Fantasy | Family  1
Fantasy | Sci-Fi  1
Romance | Action | Adventure  1
Family | Musical | Romance  1
Horror | Romance  1
Name: count, Length: 215, dtype: int64
```

```
In [37]:
# year_of_release
Year=data1["releaseYear"].value_counts()
print(Year)

releaseYear
2010    116
2011    112
2014    110
2005    106
2013    102
2012     99
2008     98
2003     95
2009     91
2004     88
2002     79
2007     66
2001     62
2006     60
Name: count, dtype: int64
```

```
In [39]:
plt.figure(figsize=(10,5))
sns.countplot(x='genre',data=data1)
plt.title('NUMBER OF MOVIES BY GENRES')
plt.xlabel('GENRES')
plt.ylabel('NUMBER OF MOVIES')
plt.show()
```



```
In [44]:
#2.histogram
import matplotlib.pyplot as plt
```

```
# Convert the categorical genre data into numerical codes
data1['genre_codes'] = data1['genre'].astype('category').cat.codes

# Create a histogram using the numerical codes
plt.figure(figsize=(10,5))
plt.hist(data1['genre_codes'], bins=len(data1['genre'].unique()), edgecolor='r')
plt.title('Genre Distribution')
plt.xlabel('Genre')
plt.ylabel('Count')
plt.xticks(ticks=range(len(data1['genre'].unique())), labels=data1['genre'].unique())
plt.show()
```



```
In [ ]:
```