MOVIELENS_CASE_STUDY

In [70]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
```

In [2]:
```
movie_col_names=["MovieID","title","Genre"]
ratings_col_names=["UserID","MovieID","Rating","Timestamp"]
users_col_names=["UserID","Gender","Age","Occupation","Zip-code"]
```

READING DATA

In [3]:
```
movie_data=pd.read_csv('movies.dat',delimiter='::', names=movie_col_names, engine='python')
movie_data
```

Out[3]:

	MovieID	title	Genre
0	1	Toy Story (1995)	Animation|Children's|Comedy
1	2	Jumanji (1995)	Adventure|Children's|Fantasy
2	3	Grumpier Old Men (1995)	Comedy|Romance
3	4	Waiting to Exhale (1995)	Comedy|Drama
4	5	Father of the Bride Part II (1995)	Comedy
...	...	...	...
3878	3948	Meet the Parents (2000)	Comedy
3879	3949	Requiem for a Dream (2000)	Drama
3880	3950	Tigerland (2000)	Drama
3881	3951	Two Family House (2000)	Drama
3882	3952	Contender, The (2000)	Drama|Thriller

3883 rows × 3 columns

In [4]:
```
ratings_data=pd.read_csv('ratings.dat',delimiter='::', names=ratings_col_names, engine='python')
ratings_data
```

Out[4]:

	UserID	MovieID	Rating	Timestamp
0	1	1193	5	978300760
1	1	661	3	978302109
2	1	914	3	978301968
3	1	3408	4	978300275
4	1	2355	5	978824291
...	...	...	...	...
1000204	6040	1091	1	956716541
1000205	6040	1094	5	956704887
1000206	6040	562	5	956704746
1000207	6040	1096	4	956715648
1000208	6040	1097	4	956715569

1000209 rows × 4 columns

In [5]:
```
users_data=pd.read_csv('users.dat',delimiter='::', names=users_col_names, engine='python')
users_data
```

Out[5]:

	UserID	Gender	Age	Occupation	Zip-code
0	1	F	1	10	48067
1	2	M	56	16	70072
2	3	M	25	15	55117
3	4	M	45	7	02460
4	5	M	25	20	55455
...	...	...	...	...	...
6035	6036	F	25	15	32603

	UserID	Gender	Age	Occupation	Zip-code
6036	6037	F	45	1	76006
6037	6038	F	56	1	14706
6038	6039	F	45	0	01060
6039	6040	M	25	6	11106

6040 rows × 5 columns

In [6]:
```
movie_ratings_combo=pd.merge(movie_data, ratings_data,on='MovieID')
movie_ratings_combo
```

Out[6]:

	MovieID	title	Genre	UserID	Rating	Timestamp
0	1	Toy Story (1995)	Animation|Children's|Comedy	1	5	978824268
1	1	Toy Story (1995)	Animation|Children's|Comedy	6	4	978237008
2	1	Toy Story (1995)	Animation|Children's|Comedy	8	4	978233496
3	1	Toy Story (1995)	Animation|Children's|Comedy	9	5	978225952
4	1	Toy Story (1995)	Animation|Children's|Comedy	10	5	978226474
...	...	...	...	...	...	...
1000204	3952	Contender, The (2000)	Drama|Thriller	5812	4	992072099
1000205	3952	Contender, The (2000)	Drama|Thriller	5831	3	986223125
1000206	3952	Contender, The (2000)	Drama|Thriller	5837	4	1011902656
1000207	3952	Contender, The (2000)	Drama|Thriller	5927	1	979852537
1000208	3952	Contender, The (2000)	Drama|Thriller	5998	4	1001781044

1000209 rows × 6 columns

CONCAT ALL 3 DATASETS TO MASTER DATA

In [7]:
```
master_data=pd.merge(movie_ratings_combo, users_data, on='UserID')
master_data
```

Out[7]:

	MovieID	title	Genre	UserID	Rating	Timestamp	Gender	Age	Occupation	Zip-code
0	1	Toy Story (1995)	Animation|Children's|Comedy	1	5	978824268	F	1	10	48067
1	48	Pocahontas (1995)	Animation|Children's|Musical|Romance	1	5	978824351	F	1	10	48067
2	150	Apollo 13 (1995)	Drama	1	5	978301777	F	1	10	48067
3	260	Star Wars: Episode IV - A New Hope (1977)	Action|Adventure|Fantasy|Sci-Fi	1	4	978300760	F	1	10	48067
4	527	Schindler's List (1993)	Drama|War	1	5	978824195	F	1	10	48067
...	...	...	...	...	...	...	...	...	...	...
1000204	3513	Rules of Engagement (2000)	Drama|Thriller	5727	4	958489970	M	25	4	92843
1000205	3535	American Psycho (2000)	Comedy|Horror|Thriller	5727	2	958489970	M	25	4	92843
1000206	3536	Keeping the Faith (2000)	Comedy|Romance	5727	5	958489902	M	25	4	92843
1000207	3555	U-571 (2000)	Action|Thriller	5727	3	958490699	M	25	4	92843
1000208	3578	Gladiator (2000)	Action|Drama	5727	5	958490171	M	25	4	92843

1000209 rows × 10 columns

In [8]:
```
#master_data=master_data.drop(columns=['Timestamp','Zip-code','Genre'],axis=1)
master_data
```

Out[8]:

	MovieID	title	Genre	UserID	Rating	Timestamp	Gender	Age	Occupation	Zip-code
0	1	Toy Story (1995)	Animation|Children's|Comedy	1	5	978824268	F	1	10	48067
1	48	Pocahontas (1995)	Animation|Children's|Musical|Romance	1	5	978824351	F	1	10	48067
2	150	Apollo 13 (1995)	Drama	1	5	978301777	F	1	10	48067
3	260	Star Wars: Episode IV - A New Hope (1977)	Action|Adventure|Fantasy|Sci-Fi	1	4	978300760	F	1	10	48067

	MovieID	title	Genre	UserID	Rating	Timestamp	Gender	Age	Occupation	Zip-code
4	527	Schindler's List (1993)	Drama|War	1	5	978824195	F	1	10	48067
...	...	...	...	...	...	...	...	...	...	...
1000204	3513	Rules of Engagement (2000)	Drama|Thriller	5727	4	958489970	M	25	4	92843
1000205	3535	American Psycho (2000)	Comedy|Horror|Thriller	5727	2	958489970	M	25	4	92843
1000206	3536	Keeping the Faith (2000)	Comedy|Romance	5727	5	958489902	M	25	4	92843
1000207	3555	U-571 (2000)	Action|Thriller	5727	3	958490699	M	25	4	92843
1000208	3578	Gladiator (2000)	Action|Drama	5727	5	958490171	M	25	4	92843

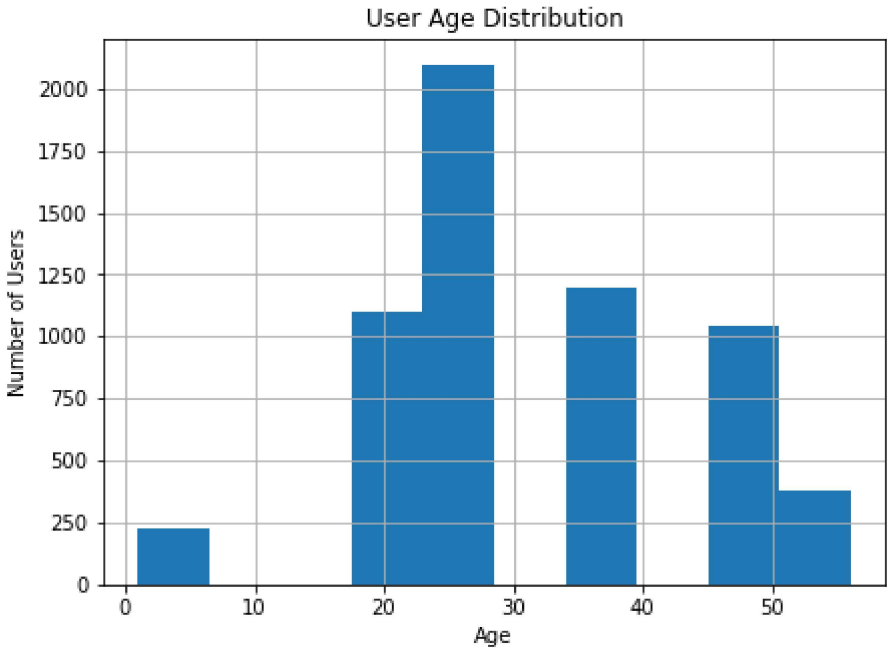
1000209 rows × 10 columns

In [9]:
```
#master_data[master_data['Occupation']==10].value_counts().plot(kind='bar')
```

VISUAL REPRESENTATION

In [10]:
```
# user age distribution
plt.figure(figsize=(7,5))
users_data.Age.hist()
plt.title('User Age Distribution')
plt.xlabel('Age')
plt.ylabel('Number of Users')
#plt.show()
```

Out[10]:
```
Text(0, 0.5, 'Number of Users')
```



In [11]:
```
#movies_ratings_group=movie_ratings_combo.groupby('title')
toy_story=movie_ratings_combo[movie_ratings_combo['title']==('Toy Story (1995)')]
```

In [12]:
```
toy_story[toy_story['Rating']>3]
```

Out[12]:

	MovieID	title	Genre	UserID	Rating	Timestamp
0	1	Toy Story (1995)	Animation|Children's|Comedy	1	5	978824268
1	1	Toy Story (1995)	Animation|Children's|Comedy	6	4	978237008
2	1	Toy Story (1995)	Animation|Children's|Comedy	8	4	978233496
3	1	Toy Story (1995)	Animation|Children's|Comedy	9	5	978225952
4	1	Toy Story (1995)	Animation|Children's|Comedy	10	5	978226474
...	...	...	...	...	...	...
2070	1	Toy Story (1995)	Animation|Children's|Comedy	6016	4	956778750
2072	1	Toy Story (1995)	Animation|Children's|Comedy	6022	5	956755763

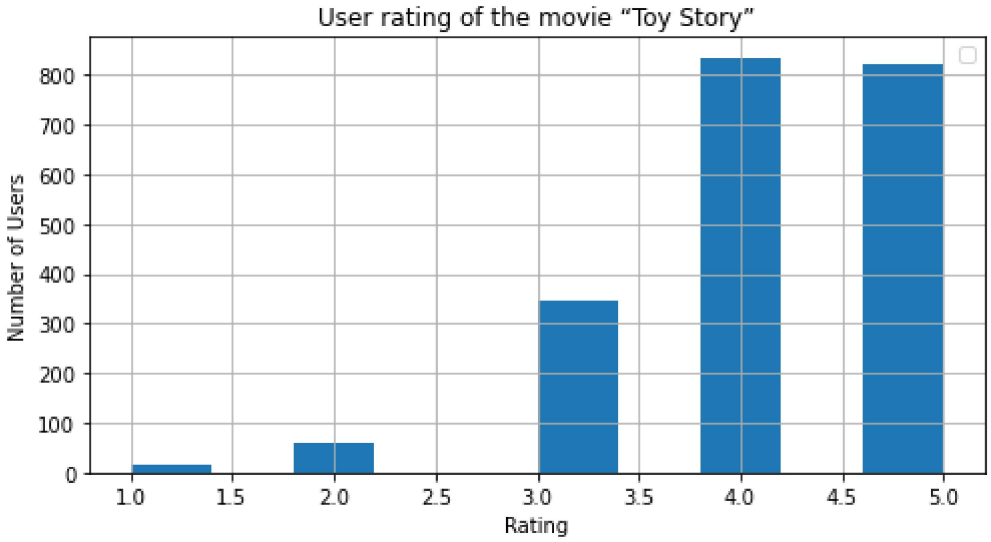
	MovieID	title	Genre	UserID	Rating	Timestamp
2073	1	Toy Story (1995)	Animation|Children's|Comedy	6025	5	956812867
2074	1	Toy Story (1995)	Animation|Children's|Comedy	6032	4	956718127
2075	1	Toy Story (1995)	Animation|Children's|Comedy	6035	4	956712849

1655 rows × 6 columns

In [13]:
```
plt.figure(figsize=(8,4))
toy_story['Rating'].hist()
plt.title('User rating of the movie “Toy Story”')
plt.xlabel('Rating')
plt.ylabel('Number of Users')
plt.legend()
```

```
No handles with labels found to put in legend.
```

Out[13]:
```
<matplotlib.legend.Legend at 0x213dc4a2d90>
```



NUMBER OF RATINGS PER MOVIE

In [14]:
```
rating_count = movie_ratings_combo.groupby('title')['Rating']
rating_count = rating_count.count().sort_values(ascending=False)
```

```
#rating_count.head()
rating_count[:25]
```

Out[14]:
```
title
American Beauty (1999)                                3428
Star Wars: Episode IV - A New Hope (1977)             2991
Star Wars: Episode V - The Empire Strikes Back (1980)  2990
Star Wars: Episode VI - Return of the Jedi (1983)     2883
Jurassic Park (1993)                                  2672
Saving Private Ryan (1998)                            2653
Terminator 2: Judgment Day (1991)                     2649
Matrix, The (1999)                                    2590
Back to the Future (1985)                             2583
Silence of the Lambs, The (1991)                      2578
Men in Black (1997)                                   2538
Raiders of the Lost Ark (1981)                        2514
Fargo (1996)                                          2513
Sixth Sense, The (1999)                               2459
Braveheart (1995)                                     2443
Shakespeare in Love (1998)                            2369
Princess Bride, The (1987)                            2318
Schindler's List (1993)                               2304
L.A. Confidential (1997)                              2288
Groundhog Day (1993)                                  2278
E.T. the Extra-Terrestrial (1982)                     2269
Star Wars: Episode I - The Phantom Menace (1999)      2250
Being John Malkovich (1999)                           2241
Shawshank Redemption, The (1994)                      2227
Godfather, The (1972)                                 2223
Name: Rating, dtype: int64
```

In [15]:
```
avg_rating=movie_ratings_combo.groupby('title')['Rating'].mean()
avg_rating=avg_rating.sort_values(ascending=False)
avg_rating.head()
```

Out[15]:
```
title
Ulysses (Ulisse) (1954)     5.0
Lured (1947)                5.0
Follow the Bitch (1998)     5.0
Bittersweet Motel (2000)    5.0
Song of Freedom (1936)      5.0
Name: Rating, dtype: float64
```

In [16]:
```
avg_rating_count = pd.DataFrame(data=avg_rating)
avg_rating_count['number_of_ratings'] = pd.DataFrame(rating_count)
avg_rating_count.head()
```

Out[16]:

title	Rating	number_of_ratings
Ulysses (Ulisse) (1954)	5.0	1
Lured (1947)	5.0	1
Follow the Bitch (1998)	5.0	1
Bittersweet Motel (2000)	5.0	1
Song of Freedom (1936)	5.0	1

In [17]:

```
grp=avg_rating_count.groupby('title')
grp.get_group('Lured (1947)')
```

Out[17]:

title	Rating	number_of_ratings
Lured (1947)	5.0	1

In [18]:

```
filter_data = avg_rating_count[avg_rating_count['number_of_ratings'] > 10]
filter_data['Rating'].sort_values(ascending=False)
```

Out[18]:

```
title
Sanjuro (1962)                                                  4.608696
Seven Samurai (The Magnificent Seven) (Shichinin no samurai) (1954)    4.560510
Shawshank Redemption, The (1994)                                4.554558
Godfather, The (1972)                                           4.524966
Close Shave, A (1995)                                           4.520548
                                                                   ...
Carnosaur 2 (1995)                                              1.461538
Amityville 3-D (1983)                                           1.372093
3 Ninjas: High Noon On Mega Mountain (1998)                     1.361702
Turbo: A Power Rangers Movie (1997)                             1.318182
Carnosaur 3: Primal Species (1996)                              1.058824
Name: Rating, Length: 3233, dtype: float64
```

In [19]:

```
filter_data[:25]
```

Out[19]:

title	Rating	number_of_ratings
Sanjuro (1962)	4.608696	69
Seven Samurai (The Magnificent Seven) (Shichinin no samurai) (1954)	4.560510	628
Shawshank Redemption, The (1994)	4.554558	2227
Godfather, The (1972)	4.524966	2223
Close Shave, A (1995)	4.520548	657
Usual Suspects, The (1995)	4.517106	1783
Schindler's List (1993)	4.510417	2304
Wrong Trousers, The (1993)	4.507937	882
Sunset Blvd. (a.k.a. Sunset Boulevard) (1950)	4.491489	470
Raiders of the Lost Ark (1981)	4.477725	2514
Rear Window (1954)	4.476190	1050
Paths of Glory (1957)	4.473913	230
Star Wars: Episode IV - A New Hope (1977)	4.453694	2991
Third Man, The (1949)	4.452083	480
Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb (1963)	4.449890	1367
For All Mankind (1989)	4.444444	27
Wallace & Gromit: The Best of Aardman Animation (1996)	4.426941	438
To Kill a Mockingbird (1962)	4.425647	928
Double Indemnity (1944)	4.415608	551
Casablanca (1942)	4.412822	1669
World of Apu, The (Apur Sansar) (1959)	4.410714	56
Sixth Sense, The (1999)	4.406263	2459
Yojimbo (1961)	4.404651	215
Pather Panchali (1955)	4.404255	47

	Rating	number_of_ratings
title		
Lawrence of Arabia (1962)	4.401925	831

In [20]:
```
user_2696 = movie_ratings_combo[movie_ratings_combo['UserID'] == 2696]
user_2696
```

Out[20]:

	MovieID	title	Genre	UserID	Rating	Timestamp
95261	350	Client, The (1994)	Drama|Mystery|Thriller	2696	3	973308886
200536	800	Lone Star (1996)	Drama|Mystery	2696	5	973308842
270342	1092	Basic Instinct (1992)	Mystery|Thriller	2696	4	973308886
274576	1097	E.T. the Extra-Terrestrial (1982)	Children's|Drama|Fantasy|Sci-Fi	2696	3	973308690
349333	1258	Shining, The (1980)	Horror	2696	4	973308710
360382	1270	Back to the Future (1985)	Comedy|Sci-Fi	2696	2	973308676
448293	1589	Cop Land (1997)	Crime|Drama|Mystery	2696	3	973308865
457193	1617	L.A. Confidential (1997)	Crime|Film-Noir|Mystery|Thriller	2696	4	973308842
459835	1625	Game, The (1997)	Mystery|Thriller	2696	4	973308842
464473	1644	I Know What You Did Last Summer (1997)	Horror|Mystery|Thriller	2696	2	973308920
465069	1645	Devil's Advocate, The (1997)	Crime|Horror|Mystery|Thriller	2696	4	973308904
480658	1711	Midnight in the Garden of Good and Evil (1997)	Comedy|Crime|Drama|Mystery	2696	4	973308904
493231	1783	Palmetto (1998)	Film-Noir|Mystery|Thriller	2696	4	973308865
496926	1805	Wild Things (1998)	Crime|Drama|Mystery|Thriller	2696	4	973308886
507001	1892	Perfect Murder, A (1998)	Mystery|Thriller	2696	4	973308904
631867	2338	I Still Know What You Did Last Summer (1998)	Horror|Mystery|Thriller	2696	2	973308920
645948	2389	Psycho (1998)	Crime|Horror|Thriller	2696	4	973308710
732119	2713	Lake Placid (1999)	Horror|Thriller	2696	1	973308710
851313	3176	Talented Mr. Ripley, The (1999)	Drama|Mystery|Thriller	2696	4	973308865

	MovieID	title	Genre	UserID	Rating	Timestamp
889623	3386	JFK (1991)	Drama|Mystery	2696	1	973308842

In [21]:
```
movie_ratings_combo
```

Out[21]:

	MovieID	title	Genre	UserID	Rating	Timestamp
0	1	Toy Story (1995)	Animation|Children's|Comedy	1	5	978824268
1	1	Toy Story (1995)	Animation|Children's|Comedy	6	4	978237008
2	1	Toy Story (1995)	Animation|Children's|Comedy	8	4	978233496
3	1	Toy Story (1995)	Animation|Children's|Comedy	9	5	978225952
4	1	Toy Story (1995)	Animation|Children's|Comedy	10	5	978226474
...	...	...	...	...	...	...
1000204	3952	Contender, The (2000)	Drama|Thriller	5812	4	992072099
1000205	3952	Contender, The (2000)	Drama|Thriller	5831	3	986223125
1000206	3952	Contender, The (2000)	Drama|Thriller	5837	4	1011902656
1000207	3952	Contender, The (2000)	Drama|Thriller	5927	1	979852537
1000208	3952	Contender, The (2000)	Drama|Thriller	5998	4	1001781044

1000209 rows × 6 columns

In [22]:
```
movie_ratings_combo['Genre'].value_counts()
```

Out[22]:
```
Comedy                          116883
Drama                           111423
Comedy|Romance                   42712
Comedy|Drama                     42245
Drama|Romance                    29170
                                 ...
Drama|Romance|Western               29
Children's|Fantasy                  27
Comedy|Film-Noir|Thriller            5
Film-Noir|Horror                     2
```

```
Fantasy                           1
Name: Genre, Length: 301, dtype: int64
```

Genre category with a one-hot encoding (1 and 0)

In [23]:
```
master_data
```

Out[23]:

	MovieID	title	Genre	UserID	Rating	Timestamp	Gender	Age	Occupation	Zip-code
0	1	Toy Story (1995)	Animation|Children's|Comedy	1	5	978824268	F	1	10	48067
1	48	Pocahontas (1995)	Animation|Children's|Musical|Romance	1	5	978824351	F	1	10	48067
2	150	Apollo 13 (1995)	Drama	1	5	978301777	F	1	10	48067
3	260	Star Wars: Episode IV - A New Hope (1977)	Action|Adventure|Fantasy|Sci-Fi	1	4	978300760	F	1	10	48067
4	527	Schindler's List (1993)	Drama|War	1	5	978824195	F	1	10	48067
...	...	...	...	...	...	...	...	...	...	...
1000204	3513	Rules of Engagement (2000)	Drama|Thriller	5727	4	958489970	M	25	4	92843
1000205	3535	American Psycho (2000)	Comedy|Horror|Thriller	5727	2	958489970	M	25	4	92843
1000206	3536	Keeping the Faith (2000)	Comedy|Romance	5727	5	958489902	M	25	4	92843
1000207	3555	U-571 (2000)	Action|Thriller	5727	3	958490699	M	25	4	92843
1000208	3578	Gladiator (2000)	Action|Drama	5727	5	958490171	M	25	4	92843

1000209 rows × 10 columns

In [24]:
```
movie_ratings_selected_features_df = master_data[[
    'Gender',
    'Age',
    'Occupation',
    'Rating',
    'Genre'
]]
```

In [25]:
```
movie_ratings_selected_features_df
```

Out[25]:

	Gender	Age	Occupation	Rating	Genre
0	F	1	10	5	Animation|Children's|Comedy
1	F	1	10	5	Animation|Children's|Musical|Romance
2	F	1	10	5	Drama
3	F	1	10	4	Action|Adventure|Fantasy|Sci-Fi
4	F	1	10	5	Drama|War
...	...	...	...	...	...
1000204	M	25	4	4	Drama|Thriller
1000205	M	25	4	2	Comedy|Horror|Thriller
1000206	M	25	4	5	Comedy|Romance
1000207	M	25	4	3	Action|Thriller
1000208	M	25	4	5	Action|Drama

1000209 rows × 5 columns

In [35]:
```
Genre=movie_ratings_selected_features_df['Genre']
Genre=Genre.str.get_dummies().add_prefix('Genre_')
```

In [36]:
```
movie_ratings_Genres_df = pd.concat([movie_ratings_selected_features_df,Genre],axis=1)
```

In [37]:
```
movie_ratings_Genres_df
```

Out[37]:

	Gender	Age	Occupation	Rating	Genre	Genre_Action	Genre_Adventure	Genre_Animation	Genre_Child
0	F	1	10	5	Animation|Children's|Comedy	0	0	1	
1	F	1	10	5	Animation|Children's|Musical|Romance	0	0	1	

	Gender	Age	Occupation	Rating	Genre	Genre_Action	Genre_Adventure	Genre_Animation	Genre_Child
2	F	1	10	5	Drama	0	0	0	
3	F	1	10	4	Action|Adventure|Fantasy|Sci-Fi	1	1	0	
4	F	1	10	5	Drama|War	0	0	0	
...	...	...	...	...	...	...	...	...	
1000204	M	25	4	4	Drama|Thriller	0	0	0	
1000205	M	25	4	2	Comedy|Horror|Thriller	0	0	0	
1000206	M	25	4	5	Comedy|Romance	0	0	0	
1000207	M	25	4	3	Action|Thriller	1	0	0	
1000208	M	25	4	5	Action|Drama	1	0	0	

1000209 rows × 23 columns

In [38]:
```
movie_ratings_Genres_df=pd.get_dummies(movie_ratings_Genres_df,columns=['Gender'])
movie_ratings_Genres_df
```

Out[38]:

	Age	Occupation	Rating	Genre	Genre_Action	Genre_Adventure	Genre_Animation	Genre_Children's	Ge
0	1	10	5	Animation|Children's|Comedy	0	0	1	1	
1	1	10	5	Animation|Children's|Musical|Romance	0	0	1	1	
2	1	10	5	Drama	0	0	0	0	
3	1	10	4	Action|Adventure|Fantasy|Sci-Fi	1	1	0	0	
4	1	10	5	Drama|War	0	0	0	0	
...	...	...	...	...	...	...	...	...	
1000204	25	4	4	Drama|Thriller	0	0	0	0	
1000205	25	4	2	Comedy|Horror|Thriller	0	0	0	0	
1000206	25	4	5	Comedy|Romance	0	0	0	0	

	Age	Occupation	Rating	Genre	Genre_Action	Genre_Adventure	Genre_Animation	Genre_Children's	Ge
1000207	25	4	3	Action|Thriller	1	0	0	0	
1000208	25	4	5	Action|Drama	1	0	0	0	

1000209 rows × 24 columns

In [39]:
```
movie_ratings_Genres_df.columns
```

Out[39]:
```
Index(['Age', 'Occupation', 'Rating', 'Genre', 'Genre_Action',
       'Genre_Adventure', 'Genre_Animation', 'Genre_Children's',
       'Genre_Comedy', 'Genre_Crime', 'Genre_Documentary', 'Genre_Drama',
       'Genre_Fantasy', 'Genre_Film-Noir', 'Genre_Horror', 'Genre_Musical',
       'Genre_Mystery', 'Genre_Romance', 'Genre_Sci-Fi', 'Genre_Thriller',
       'Genre_War', 'Genre_Western', 'Gender_F', 'Gender_M'],
      dtype='object')
```

In [41]:
```
movie_ratings_Genres_df.corr()
```

Out[41]:

	Age	Occupation	Rating	Genre_Action	Genre_Adventure	Genre_Animation	Genre_Children's	Genre_Comedy	Gen
Age	1.000000	0.078371	0.056869	-0.030975	-0.016730	-0.047020	-0.052858	-0.044046	-
Occupation	0.078371	1.000000	0.006753	0.018347	0.014309	-0.003834	-0.006906	-0.006149	
Rating	0.056869	0.006753	1.000000	-0.047633	-0.036718	0.019670	-0.039829	-0.039622	
Genre_Action	-0.030975	0.018347	-0.047633	1.000000	0.374961	-0.110294	-0.141314	-0.268092	
Genre_Adventure	-0.016730	0.014309	-0.036718	0.374961	1.000000	0.004732	0.098283	-0.124960	-
Genre_Animation	-0.047020	-0.003834	0.019670	-0.110294	0.004732	1.000000	0.576204	0.018544	-
Genre_Children's	-0.052858	-0.006906	-0.039829	-0.141314	0.098283	0.576204	1.000000	0.058711	-
Genre_Comedy	-0.044046	-0.006149	-0.039622	-0.268092	-0.124960	0.018544	0.058711	1.000000	-
Genre_Crime	-0.007931	0.002821	0.033446	0.088519	-0.045924	-0.062520	-0.081977	-0.078030	
Genre_Documentary	0.004407	-0.002689	0.028098	-0.052565	-0.035109	-0.018991	-0.024901	-0.040697	-

	Age	Occupation	Rating	Genre_Action	Genre_Adventure	Genre_Animation	Genre_Children's	Genre_Comedy	Gen
Genre_Drama	0.063856	-0.012326	0.122561	-0.202415	-0.194570	-0.154479	-0.135707	-0.249840	
Genre_Fantasy	-0.024222	0.001299	-0.023312	0.014551	0.227046	0.012025	0.263280	-0.006010	-
Genre_Film-Noir	0.033495	0.005246	0.060259	-0.080288	-0.014178	0.037013	-0.038033	-0.101425	
Genre_Horror	-0.023901	0.001439	-0.094353	-0.042733	-0.057256	-0.049730	-0.077099	-0.093064	-
Genre_Musical	0.005158	-0.007312	0.015643	-0.100432	-0.022327	0.335231	0.312567	0.030566	-
Genre_Mystery	0.024308	0.002421	0.015848	-0.054084	-0.043503	-0.042488	-0.052786	-0.105346	
Genre_Romance	0.017503	-0.014018	0.009644	-0.067830	-0.024389	-0.054540	-0.084550	0.112843	-
Genre_Sci-Fi	-0.010879	0.026250	-0.044487	0.319117	0.284190	-0.055526	-0.038844	-0.187079	-
Genre_Thriller	-0.014100	0.008981	-0.004806	0.202756	-0.038423	-0.085713	-0.132642	-0.299501	
Genre_War	0.038446	0.010264	0.075688	0.135872	0.016647	-0.046114	-0.066539	-0.127101	-
Genre_Western	0.038177	0.005924	0.007311	0.022242	-0.011964	-0.030908	-0.031269	0.007927	-
Gender_F	0.003189	-0.114974	0.019861	-0.094380	-0.038645	0.017719	0.031662	0.040758	-
Gender_M	-0.003189	0.114974	-0.019861	0.094380	0.038645	-0.017719	-0.031662	-0.040758	

23 rows × 23 columns

Features affecting the ratings of any particular movie.

In [46]:
```
movie_ratings_Genres_df=movie_ratings_Genres_df.drop(['Genre'], axis=1)
```

In [47]:
```
movie_ratings_Genres_df.dtypes
```

Out[47]:
```
Age                int64
Occupation         int64
Rating             int64
Genre_Action       int64
Genre_Adventure    int64
Genre_Animation    int64
Genre_Children's   int64
```

```
Genre_Comedy          int64
Genre_Crime           int64
Genre_Documentary     int64
Genre_Drama           int64
Genre_Fantasy         int64
Genre_Film-Noir       int64
Genre_Horror          int64
Genre_Musical         int64
Genre_Mystery         int64
Genre_Romance         int64
Genre_Sci-Fi          int64
Genre_Thriller        int64
Genre_War             int64
Genre_Western         int64
Gender_F              uint8
Gender_M              uint8
dtype: object
```

In [48]:

In [60]:
```
movie_ratings_Genres_df_sample=movie_ratings_Genres_df.sample(
    n=50000,
    random_state=0
)
```

In [61]:
```
movie_ratings_Genres_df_sample
```

Out[61]:

	Age	Occupation	Rating	Genre_Action	Genre_Adventure	Genre_Animation	Genre_Children's	Genre_Comedy	Genre_Crime	Genre_Do
324271	18	4	4	0	0	0	0	1	0	
818637	18	4	3	0	0	1	1	0	0	
148677	18	14	5	0	0	0	0	0	0	
778790	50	7	4	0	0	0	0	1	1	
525489	25	2	5	0	0	0	0	0	0	
...	...	...	...	...	...	...	...	...	...	...
106711	25	20	2	0	0	0	0	1	0	

	Age	Occupation	Rating	Genre_Action	Genre_Adventure	Genre_Animation	Genre_Children's	Genre_Comedy	Genre_Crime	Genre_Do
792958	25	14	4	1	0	0	0	0	0	
735824	56	1	4	0	0	0	0	0	0	
42804	18	14	4	0	0	0	0	1	0	
177795	35	1	3	1	1	0	0	0	0	

50000 rows × 23 columns

In [62]:
```
x=movie_ratings_Genres_df_sample.drop(['Rating'],axis=1)
y=movie_ratings_Genres_df_sample['Rating']
```

In [ ]:
```
```

MODEL TRAIN

In [63]:
```
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn import metrics

lr=LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None,
         normalize=False)
```

In [64]:
```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=0)
```

In [65]:
```
lr.fit(x_train,y_train)
```

Out[65]:
```
LinearRegression()
```

In [66]:
```
y_pred=lr.predict(x_test)
```

METRICS EVALUATION

In [67]:
```
MSE=metrics.mean_squared_error(y_test,y_pred)
MASE=metrics.mean_absolute_error(y_test,y_pred)
RMSE=np.sqrt(metrics.mean_squared_error(y_test,y_pred))
r2_score=metrics.r2_score(y_test,y_pred)
```

In [68]:
```
print("MSE:",MSE)
print("MASE:",MASE)
print("RMSE:",RMSE)
print("r2_score:",r2_score)
```

```
MSE: 1.1977731707567232
MASE: 0.8978299534841195
RMSE: 1.0944282391992282
r2_score: 0.03795269985311833
```

In [69]:
```
result=pd.DataFrame({'ACTUAL':y_test,'PREDICTED':y_pred})
result
```

Out[69]:

	ACTUAL	PREDICTED
187446	4	4.322363
69421	4	3.439548
941725	3	3.408593
841836	4	3.652663
869012	4	3.559433
...	...	...
178851	5	3.733527
374575	5	3.915716
96738	5	3.784465
727461	3	3.454760
28123	1	3.521620

10000 rows × 2 columns

In [ ]: