Shubham Shashikant Bawane **Film Recommendation System** using numpy, pandas, seaborn

Language python

[1]:
```
import numpy as np
import pandas as pd
import seaborn as sns
```

[4]:
```
movie_titles = pd.read_csv("Movie_Id_Titles.data")
movie_titles.head()
```

[4]:

	item_id	title
0	1	Toy Story (1995)
1	2	GoldenEye (1995)
2	3	Four Rooms (1995)
3	4	Get Shorty (1995)
4	5	Copycat (1995)

[8]:
```
df = pd.merge(df, movie_titles, on= 'item_id')
df.head()
```

[8]:

	user_id	item_id	rating	timestamp	title
0	0	50	5	881250949	Runway 36 (1977)
1	0	172	5	881250949	Empire Strikes Back, The (1980)
2	0	133	1	881250949	Gone with the Wind (1939)
3	196	242	3	881250949	Kolya (1996)
4	186	302	3	891717742	L.A. Confidential (1997)

[12]:
```
df = df.drop('title_x', axis=1)
```

[13]:
```
df
```

[13]:

	user_id	item_id	rating	timestamp	title_y
0	0	50	5	881250949	Runway 36 (1977)
1	0	172	5	881250949	Empire Strikes Back, The (1980)
2	0	133	1	881250949	Gone with the Wind (1939)
3	196	242	3	881250949	Kolya (1996)
4	186	302	3	891717742	L.A. Confidential (1997)
...	...	...	...	...	...
99998	880	476	3	880175444	First Wives Club, The (1996)
99999	716	204	5	879795543	Back to the Future (1985)
100000	276	1090	1	874795795	Sliver (1993)
100001	13	225	2	882399156	101 Dalmatians (1996)
100002	12	203	3	879959583	Unforgiven (1992)

[2]:
```
columns_names = ['user_id', 'item_id', 'rating' ,'timestamp']
df  = pd.read_csv('u.data',sep = '\t', names= columns_names)
```

[3]:
```
df.head()
```

[3]:

	user_id	item_id	rating	timestamp
0	0	50	5	881250949
1	0	172	5	881250949
2	0	133	1	881250949
3	196	242	3	881250949
4	186	302	3	891717742

[14]:
```
import matplotlib.pyplot as plt
import seaborn as sns
sns.set_style('white')
%matplotlib inline
```

[15]:
```
df.groupby('title_y')['rating'].mean().sort_values(ascending=False).head()
```

[15]:
```
title_y
Aiqing wansui (1994)                              5.0
Entertaining Angels: The Dorothy Day Story (1996) 5.0
Saint of Fort Washington, The (1993)              5.0
They Made Me a Criminal (1939)                    5.0
Prefontaine (1997)                                5.0
Name: rating, dtype: float64
```

[16]:
```
df.groupby('title_y')['rating'].count().sort_values(ascending=False).head()
```

[16]:
```
title_y
Runway 36 (1977)            584
Contact (1997)              509
Fargo (1996)                508
Return of the Jedi (1983)   507
Liar Liar (1997)            485
Name: rating, dtype: int64
```

[18]:
```
ratings['num of ratings'] = pd.DataFrame(df.groupby('title_y')['rating'].count())
ratings
```

[18]:

	rating	num of ratings
title_y		
'Til There Was You (1997)	2.333333	9
1-900 (1994)	2.600000	5
101 Dalmatians (1996)	2.908257	109
12 Angry Men (1957)	4.344000	125
187 (1997)	3.024390	41
...	...	...
Young Guns II (1990)	2.772727	44
Young Poisoner's Handbook, The (1995)	3.341463	41
Zeus and Roxanne (1997)	2.166667	6
unknown	3.444444	9
Á köldum klaka (Cold Fever) (1994)	3.000000	1

[17]:
```
ratings = pd.DataFrame(df.groupby('title_y')['rating'].mean())
ratings
```

[17]:

title_y	rating
'Til There Was You (1997)	2.333333
1-900 (1994)	2.600000
101 Dalmatians (1996)	2.908257
12 Angry Men (1957)	4.344000
187 (1997)	3.024390
...	...
Young Guns II (1990)	2.772727
Young Poisoner's Handbook, The (1995)	3.341463
Zeus and Roxanne (1997)	2.166667
unknown	3.444444
Á köldum klaka (Cold Fever) (1994)	3.000000

[27]:
```
sns.scatterplot(data = ratings , x = "num of ratings", y = "rating")
```

[27]:
```
<Axes: xlabel='num of ratings', ylabel='rating'>
```