In [1]:
```
import numpy as np
import pandas as pd
import warnings
```

In [2]:
```
warnings.filterwarnings('ignore')
```

In [6]:
```
columns_name=['user_id','item_id','rating','timestamp']
df=pd.read_csv('u.data',sep="\t",names=columns_name)
```

In [7]:
```
df.head()
```

Out[7]:

	user_id	item_id	rating	timestamp
0	196	242	3	881250949
1	186	302	3	891717742
2	22	377	1	878887116
3	244	51	2	880606923
4	166	346	1	886397596

In [5]:
```
df.shape
```

Out[5]:
```
(99999, 4)
```

In [8]:
```
df['user_id']
```

Out[8]:
```
0        196
1        186
2         22
3        244
4        166
        ...
99995    880
99996    716
99997    276
99998     13
99999     12
Name: user_id, Length: 100000, dtype: int64
```

In [9]:
```
df['user_id'].nunique()
```

Out[9]:
```
943
```

In [10]:
```
df['item_id'].nunique()
```

Out[10]:
```
1682
```

In [11]:
```
movies_title=pd.read_csv('u.item',sep="\|",header=None)
```

In [12]:
```
movies_title.shape
```

Out[12]:
```
(1682, 24)
```

In [15]:
```
movies_titles=movies_title[[0,1]]
movies_titles.columns=["item_id","title"]
movies_titles.head()
```

Out[15]:

	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)

In [16]:
```
df=pd.merge(df,movies_titles,on="item_id")
```

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

Out[17]:

	user_id	item_id	rating	timestamp	title
0	196	242	3	881250949	Kolya (1996)
1	63	242	3	875747190	Kolya (1996)
2	226	242	5	883888671	Kolya (1996)
3	154	242	3	879138235	Kolya (1996)
4	306	242	5	876503793	Kolya (1996)
...	...	...	...	...	...
99995	840	1674	4	891211682	Mamma Roma (1962)
99996	655	1640	3	888474646	Eighth Day, The (1996)
99997	655	1637	3	888984255	Girls Town (1996)
99998	655	1630	3	887428735	Silence of the Palace, The (Saimt el Qusur) (1...
99999	655	1641	3	887427810	Dadetown (1995)

100000 rows × 5 columns

In [18]:
```
df.tail()
```

Out[18]:

	user_id	item_id	rating	timestamp	title
99995	840	1674	4	891211682	Mamma Roma (1962)
99996	655	1640	3	888474646	Eighth Day, The (1996)
99997	655	1637	3	888984255	Girls Town (1996)
99998	655	1630	3	887428735	Silence of the Palace, The (Saimt el Qusur) (1...
99999	655	1641	3	887427810	Dadetown (1995)

In [19]:
```
ratings=pd.DataFrame(df.groupby('title').mean()['rating'])
```

In [20]:
```
ratings.head()
```

Out[20]:

	rating
title	
'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

In [21]:
```
ratings['num of ratings']=pd.DataFrame(df.groupby('title').count()['rating'])
```

Now we will create the recommendar system

In [22]:
```
df.head()
```

Out[22]:

	user_id	item_id	rating	timestamp	title
0	196	242	3	881250949	Kolya (1996)
1	63	242	3	875747190	Kolya (1996)
2	226	242	5	883888671	Kolya (1996)
3	154	242	3	879138235	Kolya (1996)
4	306	242	5	876503793	Kolya (1996)

In [23]:
```
moviemat=df.pivot_table(index="user_id",columns="title",values="rating")
```

In [24]:
```
moviemat.head()
```

Out[24]:

title	'Til There Was You (1997)	1-900 (1994)	101 Dalmatians (1996)	12 Angry Men (1957)	187 (1997)	2 Days in the Valley (1996)	20,000 Leagues Under the Sea (1954)	2001: A Space Odyssey (1968)	3 Ninjas: High Noon At Mega Mountain (1998)	39 Steps, The (1935)	...	Yankee Zulu (1994)	Year of the Horse (1997)	You So Crazy (1994)	Young Frankenstein (1974)	Young Guns (1988)	Young Guns II (1990)	Young Poisoner's Handbook, The (1995)	Zeus and Roxanne (1997)	unknown	Á köldum klaka (Cold Fever) (1994)
user_id																					
1	NaN	NaN	2.0	5.0	NaN	NaN	3.0	4.0	NaN	NaN	...	NaN	NaN	NaN	5.0	3.0	NaN	NaN	NaN	4.0	NaN
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN	2.0	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5	NaN	NaN	2.0	NaN	NaN	NaN	NaN	4.0	NaN	NaN	...	NaN	NaN	NaN	4.0	NaN	NaN	NaN	NaN	4.0	NaN

5 rows × 1664 columns

In [30]:
```
ratings
```

Out[30]:

	rating	num of ratings
title		
'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

1664 rows × 2 columns

In [32]:
```
def predict_movies(movie_name):
    movie_user_ratings=moviemat[movie_name]
    similar_to_movie=moviemat.corrwith(movie_user_ratings)
    corr_movie=pd.DataFrame(similar_to_movie,columns=['correlation'])
    corr_movie.dropna(inplace=True)
    corr_movie=corr_movie.join(ratings['num of ratings'])
    predictions=corr_movie[corr_movie['num of ratings']>100].sort_values('correlation',ascending=False)
    return predictions
```

In [33]:
```
predict_my_movie=predict_movies("Titanic (1997)")
```

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

Out[34]:

	correlation	num of ratings
title		
Titanic (1997)	1.000000	350
River Wild, The (1994)	0.497600	146
Abyss, The (1989)	0.472103	151
Bram Stoker's Dracula (1992)	0.443560	120
True Lies (1994)	0.435104	208

In [ ]:
```
```