In [3]:

```
import numpy as np
import pandas as pd
import warnings
```

In [4]:

```
warnings.filterwarnings('ignore')
```

In [8]:

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

In [6]:

```
df.head()
```

Out[6]:

	196	242	3	881250949
0	186	302	3	891717742
1	22	377	1	878887116
2	244	51	2	880606923
3	166	346	1	886397596
4	298	474	4	884182806

In [7]:

```
df.shape
```

Out[7]:

```
(99999, 4)
```

In [9]:

```
df['user_id']
```

Out[9]:

```
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 [10]:

```
df['user_id'].nunique()
```

Out[10]:

```
943
```

In [11]:

```
df['item_id'].nunique()
```

Out[11]:

```
1682
```

In [18]:

```
movies_title=pd.read_csv('u.item',sep="\|",header=None,encoding="ISO-8859-1")
```

In [19]:

```
movies_title.shape
```

Out[19]:

```
(1682, 24)
```

In [23]:

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

Out[23]:

	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 [24]:

```
df=pd.merge(df,movies_title,on="item_id")
```

In [25]:

```
df
```

Out[25]:

	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 [26]:

```
df.tail()
```

Out[26]:

	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 [27]:

```
rating=pd.DataFrame(df.groupby('title').mean()['rating'])
```

In [29]:

```
rating.head()
```

Out[29]:

title	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

In [36]:

```
rating["number of ratings"]=pd.DataFrame(df.groupby("title").count()["rating"])
```

Creating the Recommendar System

In [31]:

```
df.head()
```

Out[31]:

	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 [38]:

```
moviemat=df.pivot_table(index="user_id",columns="title",values="rating")
```

In [39]:

```
moviemat.head()
```

Out[39]:

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)	Youn Gun (1988
user_id																
1	NaN	NaN	2.0	5.0	NaN	NaN	3.0	4.0	NaN	NaN	...	NaN	NaN	NaN	5.0	3.
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0	NaN	...	NaN	NaN	NaN	NaN	Na
3	NaN	NaN	NaN	NaN	2.0	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	Na
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	Na
5	NaN	NaN	2.0	NaN	NaN	NaN	NaN	4.0	NaN	NaN	...	NaN	NaN	NaN	4.0	Na

5 rows × 1664 columns

In [40]:

```
starwars_user_ratings=moviemat['Star Wars (1977)']
```

In [41]:

```
starwars_user_ratings.head()
```

Out[41]:

```
user_id
1    5.0
2    5.0
3    NaN
4    5.0
5    4.0
Name: Star Wars (1977), dtype: float64
```

In [42]:

```
similar_to_starwars=moviemat.corrwith(starwars_user_ratings)
```

In [44]:

```
similar_to_starwars
```

Out[44]:

```
title
'Til There Was You (1997)                0.872872
1-900 (1994)                            -0.645497
101 Dalmatians (1996)                    0.211132
12 Angry Men (1957)                      0.184289
187 (1997)                               0.027398
                                             ...
Young Guns II (1990)                     0.228615
Young Poisoner's Handbook, The (1995)   -0.007374
Zeus and Roxanne (1997)                  0.818182
unknown                                  0.723123
Á köldum klaka (Cold Fever) (1994)            NaN
Length: 1664, dtype: float64
```

In [45]:

```
corr_starwars=pd.DataFrame(similar_to_starwars,columns=["correlation"])
```

In [46]:

```
corr_starwars.dropna(inplace=True)
```

In [47]:

```
corr_starwars
```

Out[47]:

	correlation
title	
'Til There Was You (1997)	0.872872
1-900 (1994)	-0.645497
101 Dalmatians (1996)	0.211132
12 Angry Men (1957)	0.184289
187 (1997)	0.027398
...	...
Young Guns (1988)	0.186377
Young Guns II (1990)	0.228615
Young Poisoner's Handbook, The (1995)	-0.007374
Zeus and Roxanne (1997)	0.818182
unknown	0.723123

1410 rows × 1 columns

In [48]:

```
corr_starwars.head()
```

Out[48]:

title	correlation
'Til There Was You (1997)	0.872872
1-900 (1994)	-0.645497
101 Dalmatians (1996)	0.211132
12 Angry Men (1957)	0.184289
187 (1997)	0.027398

In [52]:

```
corr_starwars.sort_values("correlation",ascending=False).head(10)
```

Out[52]:

title	correlation
Hollow Reed (1996)	1.0
Commandments (1997)	1.0
Cosi (1996)	1.0
No Escape (1994)	1.0
Stripes (1981)	1.0
Star Wars (1977)	1.0
Man of the Year (1995)	1.0
Beans of Egypt, Maine, The (1994)	1.0
Old Lady Who Walked in the Sea, The (Vieille qui marchait dans la mer, La) (1991)	1.0
Outlaw, The (1943)	1.0

In [54]:

```
corr_starwars=corr_starwars.join(rating["number of ratings"])
```

In [55]:

```
corr_starwars
```

Out[55]:

title	correlation	number of ratings
'Til There Was You (1997)	0.872872	9
1-900 (1994)	-0.645497	5
101 Dalmatians (1996)	0.211132	109
12 Angry Men (1957)	0.184289	125
187 (1997)	0.027398	41
...	...	...
Young Guns (1988)	0.186377	101
Young Guns II (1990)	0.228615	44
Young Poisoner's Handbook, The (1995)	-0.007374	41
Zeus and Roxanne (1997)	0.818182	6
unknown	0.723123	9

1410 rows × 2 columns

In [56]:

```
corr_starwars.head()
```

```
corr_starwars.head()
```

Out[56]:

title	correlation	number of ratings
'Til There Was You (1997)	0.872872	9
1-900 (1994)	-0.645497	5
101 Dalmatians (1996)	0.211132	109
12 Angry Men (1957)	0.184289	125
187 (1997)	0.027398	41

In [57]:

```
corr_starwars[corr_starwars["number of ratings"]>100].sort_values("correlation",ascending=False)
```

Out[57]:

title	correlation	number of ratings
Star Wars (1977)	1.000000	583
Empire Strikes Back, The (1980)	0.747981	367
Return of the Jedi (1983)	0.672556	507
Raiders of the Lost Ark (1981)	0.536117	420
Austin Powers: International Man of Mystery (1997)	0.377433	130
...	...	...
Edge, The (1997)	-0.127167	113
As Good As It Gets (1997)	-0.130466	112
Crash (1996)	-0.148507	128
G.I. Jane (1997)	-0.176734	175
First Wives Club, The (1996)	-0.194496	160

334 rows × 2 columns

In [63]:

```
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(rating["number of ratings"])

    predictions=corr_movie[corr_movie["number of ratings"]>100].sort_values("correlation",ascending=False

    return predictions
```

In [64]:

```
predict_my_movie=predict_movies("Titanic (1997)")
```

In [65]:

```
predict_my_movie.head()
```

Out[65]:

title	correlation	number of ratings
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 [ ]: