

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

```
In [2]: title=['user_id','item_id','rating','timestamp']
file1= pd.read_csv("D:\\r\\Movie-Recommender-in-python-master\\u.data", sep='\t', names=title)
file1.head()
```

Out[2]:

	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

```
In [3]: file2=pd.read_csv("D:\\r\\Movie-Recommender-in-python-master\\Movie")
file2.head()
```

Out[3]:

	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 [4]: df=pd.merge(file1,file2,on='item_id')
df.head()
```

Out[4]:

	user_id	item_id	rating	timestamp	title
0	0	50	5	881250949	Star Wars (1977)
1	290	50	5	880473582	Star Wars (1977)
2	79	50	4	891271545	Star Wars (1977)
3	2	50	5	888552084	Star Wars (1977)
4	8	50	5	879362124	Star Wars (1977)

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

Out[5]: (100003, 5)

```
In [6]: df.isnull().sum()
```

Out[6]:

user_id	0
item_id	0
rating	0
timestamp	0
title	0

dtype: int64

```
In [7]: df.groupby('title')['rating'].mean()
```

Out[7]:

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
...	...
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

Name: rating, Length: 1664, dtype: float64

```
In [8]: df.groupby('title')['rating'].count()
```

Out[8]:

title	
'Til There Was You (1997)	9
1-900 (1994)	5
101 Dalmatians (1996)	109
12 Angry Men (1957)	125
187 (1997)	41
...	...
Young Guns II (1990)	44
Young Poisoner's Handbook, The (1995)	41
Zeus and Roxanne (1997)	6
unknown	9
Á köldum klaka (Cold Fever) (1994)	1

Name: rating, Length: 1664, dtype: int64

```
In [9]: ratings=pd.DataFrame(df.groupby('title')['rating'].mean())
ratings.head()
```

Out[9]:

	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 [10]: ratings['count']=pd.DataFrame(df.groupby('title')['rating'].count())
ratings.head()
```

Out[10]:

	rating	count
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

```
In [11]: recommendation= df.pivot_table(index='user_id',columns='title',values='rating')
recommendation.head()
```

Out[11]:

	'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)	F
user_id															
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	
1	NaN	NaN	2.0	5.0	NaN	NaN	3.0	4.0	NaN	NaN	...	NaN	NaN	NaN	
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0	NaN	...	NaN	NaN	NaN	
3	NaN	NaN	NaN	NaN	2.0	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	

5 rows × 1664 columns

```
In [12]: moviename=recommendation['12 Angry Men (1957)']
moviename.head()
```

Out[12]:

user_id	
0	NaN
1	5.0
2	NaN
3	NaN
4	NaN

Name: 12 Angry Men (1957), dtype: float64

```
In [13]: next=recommendation.corrwith(moviename)

c:\python38\lib\site-packages\numpy\lib\function_base.py:2526: RuntimeWarning: Degrees of freedom <= 0 for slice
  c = cov(x, y, rowvar)
c:\python38\lib\site-packages\numpy\lib\function_base.py:2455: RuntimeWarning: divide by zero encountered in true_divide
  c *= np.true_divide(1, fact)
```

```
In [14]: corr=pd.DataFrame(next,columns=['Correlation'])
corr.dropna(inplace=True)
corr.head()
```

Out[14]:

	Correlation
title	
'Til There Was You (1997)	-0.500000
101 Dalmatians (1996)	-0.049890
12 Angry Men (1957)	1.000000
187 (1997)	0.666667
2 Days in the Valley (1996)	0.256625

```
In [15]: corr=corr.join(ratings['count'])
corr.head()
```

Out[15]:

	Correlation	count
title		
'Til There Was You (1997)	-0.500000	9
101 Dalmatians (1996)	-0.049890	109
12 Angry Men (1957)	1.000000	125
187 (1997)	0.666667	41
2 Days in the Valley (1996)	0.256625	93

```
In [16]: corr[corr['count']>100].sort_values('Correlation',ascending=False).head()
```

Out[16]:

	Correlation	count
title		
12 Angry Men (1957)	1.000000	125
Ulee's Gold (1997)	0.619544	184
Rear Window (1954)	0.570513	209
Seven Years in Tibet (1997)	0.549939	155
Clerks (1994)	0.528173	148