

PROBLEM STATEMENT

- This notebook implements a movie recommender system.
- Recommender systems are used to suggest movies or songs to users based on their interest or usage history.
- For example, Netflix recommends movies to watch based on the previous movies you've watched.
- Using Item-based Collaborative Filter
- Dataset MovieLens: <https://grouplens.org/datasets/movielens/100k/>



Importing Libraries

```
In [55]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

Importing Dataset

```
In [56]: # Two datasets are available, let's load the first one:
```

```
movie_titles_df = pd.read_csv("Movie_Id_Titles")
movie_titles_df.head(20)
```

Out[56]:

	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)
5	6	Shanghai Triad (Yao a yao yao dao waipo qiao) ...
6	7	Twelve Monkeys (1995)
7	8	Babe (1995)
8	9	Dead Man Walking (1995)
9	10	Richard III (1995)
10	11	Seven (Se7en) (1995)
11	12	Usual Suspects, The (1995)
12	13	Mighty Aphrodite (1995)
13	14	Postino, Il (1994)
14	15	Mr. Holland's Opus (1995)
15	16	French Twist (Gazon maudit) (1995)
16	17	From Dusk Till Dawn (1996)
17	18	White Balloon, The (1995)
18	19	Antonia's Line (1995)
19	20	Angels and Insects (1995)

```
In [57]: # Let's load the second one!
movies_rating_df = pd.read_csv('u.data', sep='\t', names=['user_id', 'i
```

```
tem_id', 'rating', 'timestamp']])
```

```
In [58]: movies_rating_df.head(10)
```

```
Out[58]:
```

	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
5	22	377	1	878887116
6	244	51	2	880606923
7	166	346	1	886397596
8	298	474	4	884182806
9	115	265	2	881171488

```
In [59]: movies_rating_df.tail()
```

```
Out[59]:
```

	user_id	item_id	rating	timestamp
99998	880	476	3	880175444
99999	716	204	5	879795543
100000	276	1090	1	874795795
100001	13	225	2	882399156
100002	12	203	3	879959583

Dropping the timestamp

```
In [60]: # Let's drop the timestamp
movies_rating_df.drop(['timestamp'], axis = 1, inplace = True)
```

```
In [61]: movies_rating_df
```

Out[61]:

	user_id	item_id	rating
0	0	50	5
1	0	172	5
2	0	133	1
3	196	242	3
4	186	302	3
5	22	377	1
6	244	51	2
7	166	346	1
8	298	474	4
9	115	265	2
10	253	465	5
11	305	451	3
12	6	86	3
13	62	257	2
14	286	1014	5
15	200	222	5
16	210	40	3
17	224	29	3
18	303	785	3
19	122	387	5
20	194	274	2

	user_id	item_id	rating
21	291	1042	4
22	234	1184	2
23	119	392	4
24	167	486	4
25	299	144	4
26	291	118	2
27	308	1	4
28	95	546	2
29	38	95	5
...
99973	449	120	1
99974	661	762	2
99975	721	874	3
99976	821	151	4
99977	764	596	3
99978	537	443	3
99979	618	628	2
99980	487	291	3
99981	113	975	5
99982	943	391	2
99983	864	685	4
99984	750	323	3
99985	279	64	1
99986	646	750	3
99987	654	370	2

	user_id	item_id	rating
99988	617	582	4
99989	913	690	3
99990	660	229	2
99991	421	498	4
99992	495	1091	4
99993	806	421	4
99994	676	538	4
99995	721	262	3
99996	913	209	2
99997	378	78	3
99998	880	476	3
99999	716	204	5
100000	276	1090	1
100001	13	225	2
100002	12	203	3

100003 rows × 3 columns

In [62]: `movies_rating_df.describe()`

Out[62]:

	user_id	item_id	rating
count	100003.000000	100003.000000	100003.000000
mean	462.470876	425.520914	3.529864
std	266.622454	330.797791	1.125704
min	0.000000	1.000000	1.000000
25%	254.000000	175.000000	3.000000

	user_id	item_id	rating
50%	447.000000	322.000000	4.000000
75%	682.000000	631.000000	4.000000
max	943.000000	1682.000000	5.000000

In [63]: `movies_rating_df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100003 entries, 0 to 100002
Data columns (total 3 columns):
user_id      100003 non-null int64
item_id      100003 non-null int64
rating       100003 non-null int64
dtypes: int64(3)
memory usage: 2.3 MB
```

In [64]: *# Let's merge both dataframes together so we can have ID with the movie name*
`movies_rating_df = pd.merge(movies_rating_df, movie_titles_df, on = 'item_id')`

In [65]: `movies_rating_df`

Out[65]:

	user_id	item_id	rating	title
0	0	50	5	Star Wars (1977)
1	290	50	5	Star Wars (1977)
2	79	50	4	Star Wars (1977)
3	2	50	5	Star Wars (1977)
4	8	50	5	Star Wars (1977)
5	274	50	5	Star Wars (1977)
6	227	50	4	Star Wars (1977)

	user_id	item_id	rating	title
7	99	50	5	Star Wars (1977)
8	305	50	5	Star Wars (1977)
9	108	50	4	Star Wars (1977)
10	63	50	4	Star Wars (1977)
11	234	50	4	Star Wars (1977)
12	97	50	5	Star Wars (1977)
13	117	50	5	Star Wars (1977)
14	70	50	4	Star Wars (1977)
15	318	50	2	Star Wars (1977)
16	145	50	5	Star Wars (1977)
17	124	50	3	Star Wars (1977)
18	253	50	4	Star Wars (1977)
19	271	50	5	Star Wars (1977)
20	254	50	5	Star Wars (1977)
21	326	50	5	Star Wars (1977)
22	213	50	5	Star Wars (1977)
23	80	50	3	Star Wars (1977)
24	322	50	5	Star Wars (1977)
25	28	50	4	Star Wars (1977)
26	267	50	5	Star Wars (1977)
27	130	50	5	Star Wars (1977)
28	262	50	2	Star Wars (1977)
29	230	50	5	Star Wars (1977)
...
99973	405	1564	1	To Cross the Rubicon (1991)

	user_id	item_id	rating	title
99974	456	1604	4	He Walked by Night (1948)
99975	405	1574	1	Pharaoh's Army (1995)
99976	883	1448	5	My Favorite Season (1993)
99977	655	1448	3	My Favorite Season (1993)
99978	234	1448	3	My Favorite Season (1993)
99979	787	1671	1	Further Gesture, A (1996)
99980	587	1625	4	Nightwatch (1997)
99981	782	1665	2	Brother's Kiss, A (1997)
99982	181	1325	1	August (1996)
99983	782	1670	3	Tainted (1998)
99984	181	1341	1	Hedd Wyn (1992)
99985	181	1352	1	Shadow of Angels (Schatten der Engel) (1976)
99986	655	1636	4	Brothers in Trouble (1995)
99987	896	1681	3	You So Crazy (1994)
99988	655	1638	3	Normal Life (1996)
99989	782	1667	3	Next Step, The (1995)
99990	851	1675	3	Sunchaser, The (1996)
99991	234	1460	3	Sleepover (1995)
99992	532	1594	4	Everest (1998)
99993	416	1594	5	Everest (1998)
99994	648	1626	1	Nobody Loves Me (Keiner liebt mich) (1994)
99995	655	1645	4	Butcher Boy, The (1998)
99996	747	1659	1	Getting Away With Murder (1996)
99997	916	1682	3	Scream of Stone (Schrei aus Stein) (1991)
99998	840	1674	4	Mamma Roma (1962)

	user_id	item_id	rating	title	
	99999	655	1640	3	Eighth Day, The (1996)
	100000	655	1637	3	Girls Town (1996)
	100001	655	1630	3	Silence of the Palace, The (Saimt el Qusur) (1996)
	100002	655	1641	3	Dadetown (1995)

100003 rows × 4 columns

In [66]: `movies_rating_df.shape`

Out[66]: (100003, 4)

In [67]: `df=movies_rating_df.groupby('title')`
`df.head()`

Out[67]:

	user_id	item_id	rating	title
0	0	50	5	Star Wars (1977)
1	290	50	5	Star Wars (1977)
2	79	50	4	Star Wars (1977)
3	2	50	5	Star Wars (1977)
4	8	50	5	Star Wars (1977)
584	0	172	5	Empire Strikes Back, The (1980)
585	213	172	5	Empire Strikes Back, The (1980)
586	92	172	4	Empire Strikes Back, The (1980)
587	77	172	3	Empire Strikes Back, The (1980)
588	194	172	3	Empire Strikes Back, The (1980)
952	0	133	1	Gone with the Wind (1939)
953	297	133	4	Gone with the Wind (1939)

user_id item_id rating

title

	user_id	item_id	rating	title
954	213	133	3	Gone with the Wind (1939)
955	312	133	5	Gone with the Wind (1939)
956	286	133	4	Gone with the Wind (1939)
1124	196	242	3	Kolya (1996)
1125	63	242	3	Kolya (1996)
1126	226	242	5	Kolya (1996)
1127	154	242	3	Kolya (1996)
1128	306	242	5	Kolya (1996)
1241	186	302	3	L.A. Confidential (1997)
1242	191	302	4	L.A. Confidential (1997)
1243	49	302	4	L.A. Confidential (1997)
1244	54	302	4	L.A. Confidential (1997)
1245	62	302	3	L.A. Confidential (1997)
1538	22	377	1	Heavyweights (1994)
1539	5	377	1	Heavyweights (1994)
1540	314	377	3	Heavyweights (1994)
1541	13	377	1	Heavyweights (1994)
1542	222	377	1	Heavyweights (1994)
...
99973	405	1564	1	To Cross the Rubicon (1991)
99974	456	1604	4	He Walked by Night (1948)
99975	405	1574	1	Pharaoh's Army (1995)
99976	883	1448	5	My Favorite Season (1993)
99977	655	1448	3	My Favorite Season (1993)
99978	234	1448	3	My Favorite Season (1993)

user_id item_id rating

title

	user_id	item_id	rating	title
99979	787	1671	1	Further Gesture, A (1996)
99980	587	1625	4	Nightwatch (1997)
99981	782	1665	2	Brother's Kiss, A (1997)
99982	181	1325	1	August (1996)
99983	782	1670	3	Tainted (1998)
99984	181	1341	1	Hedd Wyn (1992)
99985	181	1352	1	Shadow of Angels (Schatten der Engel) (1976)
99986	655	1636	4	Brothers in Trouble (1995)
99987	896	1681	3	You So Crazy (1994)
99988	655	1638	3	Normal Life (1996)
99989	782	1667	3	Next Step, The (1995)
99990	851	1675	3	Sunchaser, The (1996)
99991	234	1460	3	Sleepover (1995)
99992	532	1594	4	Everest (1998)
99993	416	1594	5	Everest (1998)
99994	648	1626	1	Nobody Loves Me (Keiner liebt mich) (1994)
99995	655	1645	4	Butcher Boy, The (1998)
99996	747	1659	1	Getting Away With Murder (1996)
99997	916	1682	3	Scream of Stone (Schrei aus Stein) (1991)
99998	840	1674	4	Mamma Roma (1962)
99999	655	1640	3	Eighth Day, The (1996)
100000	655	1637	3	Girls Town (1996)
100001	655	1630	3	Silence of the Palace, The (Saimt el Qusur) (1...
100002	655	1641	3	Dadetown (1995)

7404 rows x 4 columns

In [68]: `movies_rating_df.groupby('title')['rating'].describe()`

Out[68]:

	count	mean	std	min	25%	50%	75%	max
title								
'Til There Was You (1997)	9.0	2.333333	1.000000	1.0	2.00	2.0	3.00	4.0
1-900 (1994)	5.0	2.600000	1.516575	1.0	1.00	3.0	4.00	4.0
101 Dalmatians (1996)	109.0	2.908257	1.076184	1.0	2.00	3.0	4.00	5.0
12 Angry Men (1957)	125.0	4.344000	0.719588	2.0	4.00	4.0	5.00	5.0
187 (1997)	41.0	3.024390	1.172344	1.0	2.00	3.0	4.00	5.0
2 Days in the Valley (1996)	93.0	3.225806	0.957000	1.0	3.00	3.0	4.00	5.0
20,000 Leagues Under the Sea (1954)	72.0	3.500000	0.731581	2.0	3.00	4.0	4.00	5.0
2001: A Space Odyssey (1968)	259.0	3.969112	1.026307	1.0	3.00	4.0	5.00	5.0
3 Ninjas: High Noon At Mega Mountain (1998)	5.0	1.000000	0.000000	1.0	1.00	1.0	1.00	1.0
39 Steps, The (1935)	59.0	4.050847	0.859505	2.0	4.00	4.0	5.00	5.0
8 1/2 (1963)	38.0	3.815789	1.182194	1.0	3.00	4.0	5.00	5.0
8 Heads in a Duffel Bag (1997)	4.0	3.250000	1.500000	1.0	3.25	4.0	4.00	4.0
8 Seconds (1994)	4.0	3.750000	1.258306	2.0	3.50	4.0	4.25	5.0
A Chef in Love (1996)	8.0	4.125000	0.834523	3.0	3.75	4.0	5.00	5.0
Above the Rim (1994)	5.0	3.000000	0.707107	2.0	3.00	3.0	3.00	4.0
Absolute Power (1997)	127.0	3.370079	0.949546	1.0	3.00	3.0	4.00	5.0
Abyss, The (1989)	151.0	3.589404	0.911201	1.0	3.00	4.0	4.00	5.0
Ace Ventura: Pet Detective (1994)	103.0	3.048544	1.215735	1.0	2.00	3.0	4.00	5.0
Ace Ventura: When Nature Calls (1995)	37.0	2.675676	1.106899	1.0	2.00	3.0	3.00	5.0
Across the Sea of Time (1995)	4.0	2.750000	1.258306	1.0	2.50	3.0	3.25	4.0

	count	mean	std	min	25%	50%	75%	max
title								
Addams Family Values (1993)	87.0	2.816092	0.946582	1.0	2.00	3.0	3.00	5.0
Addicted to Love (1997)	54.0	3.166667	1.004706	1.0	3.00	3.0	4.00	5.0
Addiction, The (1995)	11.0	2.181818	1.328020	1.0	1.00	2.0	3.50	4.0
Adventures of Pinocchio, The (1996)	39.0	3.051282	1.145902	1.0	2.00	3.0	4.00	5.0
Adventures of Priscilla, Queen of the Desert, The (1994)	111.0	3.594595	1.030255	1.0	3.00	4.0	4.00	5.0
Adventures of Robin Hood, The (1938)	67.0	3.791045	0.844542	2.0	3.00	4.0	4.00	5.0
Affair to Remember, An (1957)	26.0	4.192308	0.895287	2.0	4.00	4.0	5.00	5.0
African Queen, The (1951)	152.0	4.184211	0.758362	2.0	4.00	4.0	5.00	5.0
Afterglow (1997)	18.0	3.111111	1.022620	1.0	2.25	3.0	4.00	5.0
Age of Innocence, The (1993)	65.0	3.384615	1.010759	1.0	3.00	3.0	4.00	5.0
...
Window to Paris (1994)	1.0	4.000000	NaN	4.0	4.00	4.0	4.00	4.0
Wings of Courage (1995)	1.0	4.000000	NaN	4.0	4.00	4.0	4.00	4.0
Wings of Desire (1987)	57.0	4.000000	1.118034	1.0	3.00	4.0	5.00	5.0
Wings of the Dove, The (1997)	75.0	3.680000	0.902848	1.0	3.00	4.0	4.00	5.0
Winnie the Pooh and the Blustery Day (1968)	75.0	3.800000	1.090500	1.0	3.00	4.0	5.00	5.0
Winter Guest, The (1997)	9.0	3.444444	1.424001	1.0	3.00	3.0	5.00	5.0
Wishmaster (1997)	27.0	2.444444	1.395965	1.0	1.00	2.0	3.00	5.0
With Honors (1994)	46.0	3.065217	1.143137	1.0	2.00	3.0	4.00	5.0
Withnail and I (1987)	13.0	3.230769	1.300887	1.0	2.00	3.0	4.00	5.0
Witness (1985)	1.0	4.000000	NaN	4.0	4.00	4.0	4.00	4.0
Wizard of Oz, The (1939)	246.0	4.077236	0.959448	1.0	3.00	4.0	5.00	5.0
Wolf (1994)	67.0	2.701493	0.953700	1.0	2.00	3.0	3.00	5.0

	count	mean	std	min	25%	50%	75%	max
title								
Woman in Question, The (1950)	1.0	1.000000	NaN	1.0	1.00	1.0	1.00	1.0
Women, The (1939)	15.0	3.666667	1.175139	1.0	3.00	4.0	4.50	5.0
Wonderful, Horrible Life of Leni Riefenstahl, The (1993)	10.0	4.000000	0.666667	3.0	4.00	4.0	4.00	5.0
Wonderland (1997)	10.0	3.200000	1.475730	1.0	2.25	3.5	4.00	5.0
Wooden Man's Bride, The (Wu Kui) (1994)	3.0	2.666667	1.527525	1.0	2.00	3.0	3.50	4.0
World of Apu, The (Apu Sansar) (1959)	6.0	4.000000	1.673320	1.0	3.50	5.0	5.00	5.0
Wrong Trousers, The (1993)	118.0	4.466102	0.823607	1.0	4.00	5.0	5.00	5.0
Wyatt Earp (1994)	50.0	3.100000	0.974156	1.0	3.00	3.0	4.00	5.0
Yankee Zulu (1994)	1.0	1.000000	NaN	1.0	1.00	1.0	1.00	1.0
Year of the Horse (1997)	7.0	3.285714	1.889822	1.0	1.50	4.0	5.00	5.0
You So Crazy (1994)	1.0	3.000000	NaN	3.0	3.00	3.0	3.00	3.0
Young Frankenstein (1974)	200.0	3.945000	0.903358	2.0	3.00	4.0	5.00	5.0
Young Guns (1988)	101.0	3.207921	0.886756	1.0	3.00	3.0	4.00	5.0
Young Guns II (1990)	44.0	2.772727	1.008421	1.0	2.00	3.0	3.00	5.0
Young Poisoner's Handbook, The (1995)	41.0	3.341463	1.237129	1.0	3.00	4.0	4.00	5.0
Zeus and Roxanne (1997)	6.0	2.166667	0.983192	1.0	1.25	2.5	3.00	3.0
unknown	9.0	3.444444	1.130388	1.0	3.00	4.0	4.00	5.0
Á köldum klaka (Cold Fever) (1994)	1.0	3.000000	NaN	3.0	3.00	3.0	3.00	3.0

1664 rows × 8 columns

```
In [69]: ratings_df_mean = movies_rating_df.groupby('title')['rating'].describe
          ()['mean']
          ratings_df_mean.head(10)
```

```
Out[69]: title
1711  The 13th Year (1987)      3.333333
```



```
'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
2 Days in the Valley (1996)       3.225806
20,000 Leagues Under the Sea (1954) 3.500000
2001: A Space Odyssey (1968)      3.969112
3 Ninjas: High Noon At Mega Mountain (1998) 1.000000
39 Steps, The (1935)              4.050847
Name: mean, dtype: float64
```

```
In [70]: ratings_df_count = movies_rating_df.groupby('title')['rating'].describe
         ()['count']
```

```
In [71]: ratings_df_count.head(10)
```

```
Out[71]: title
'Til There Was You (1997)          9.0
1-900 (1994)                      5.0
101 Dalmatians (1996)             109.0
12 Angry Men (1957)              125.0
187 (1997)                       41.0
2 Days in the Valley (1996)       93.0
20,000 Leagues Under the Sea (1954) 72.0
2001: A Space Odyssey (1968)     259.0
3 Ninjas: High Noon At Mega Mountain (1998) 5.0
39 Steps, The (1935)             59.0
Name: count, dtype: float64
```

```
In [72]: ratings_mean_count_df = pd.concat([ratings_df_count, ratings_df_mean],
         axis = 1)
```

```
In [73]: ratings_mean_count_df.reset_index()
```

```
Out[73]:
```

	title	count	mean
--	-------	-------	------

	title	count	mean
0	'Til There Was You (1997)	9.0	2.333333
1	1-900 (1994)	5.0	2.600000
2	101 Dalmatians (1996)	109.0	2.908257
3	12 Angry Men (1957)	125.0	4.344000
4	187 (1997)	41.0	3.024390
5	2 Days in the Valley (1996)	93.0	3.225806
6	20,000 Leagues Under the Sea (1954)	72.0	3.500000
7	2001: A Space Odyssey (1968)	259.0	3.969112
8	3 Ninjas: High Noon At Mega Mountain (1998)	5.0	1.000000
9	39 Steps, The (1935)	59.0	4.050847
10	8 1/2 (1963)	38.0	3.815789
11	8 Heads in a Duffel Bag (1997)	4.0	3.250000
12	8 Seconds (1994)	4.0	3.750000
13	A Chef in Love (1996)	8.0	4.125000
14	Above the Rim (1994)	5.0	3.000000
15	Absolute Power (1997)	127.0	3.370079
16	Abyss, The (1989)	151.0	3.589404
17	Ace Ventura: Pet Detective (1994)	103.0	3.048544
18	Ace Ventura: When Nature Calls (1995)	37.0	2.675676
19	Across the Sea of Time (1995)	4.0	2.750000
20	Addams Family Values (1993)	87.0	2.816092
21	Addicted to Love (1997)	54.0	3.166667
22	Addiction, The (1995)	11.0	2.181818
23	Adventures of Pinocchio, The (1996)	39.0	3.051282
24	Adventures of Priscilla, Queen of the Desert, ...	111.0	3.594595

	title	count	mean
25	Adventures of Robin Hood, The (1938)	67.0	3.791045
26	Affair to Remember, An (1957)	26.0	4.192308
27	African Queen, The (1951)	152.0	4.184211
28	Afterglow (1997)	18.0	3.111111
29	Age of Innocence, The (1993)	65.0	3.384615
...
1634	Window to Paris (1994)	1.0	4.000000
1635	Wings of Courage (1995)	1.0	4.000000
1636	Wings of Desire (1987)	57.0	4.000000
1637	Wings of the Dove, The (1997)	75.0	3.680000
1638	Winnie the Pooh and the Blustery Day (1968)	75.0	3.800000
1639	Winter Guest, The (1997)	9.0	3.444444
1640	Wishmaster (1997)	27.0	2.444444
1641	With Honors (1994)	46.0	3.065217
1642	Withnail and I (1987)	13.0	3.230769
1643	Witness (1985)	1.0	4.000000
1644	Wizard of Oz, The (1939)	246.0	4.077236
1645	Wolf (1994)	67.0	2.701493
1646	Woman in Question, The (1950)	1.0	1.000000
1647	Women, The (1939)	15.0	3.666667
1648	Wonderful, Horrible Life of Leni Riefenstahl, ...	10.0	4.000000
1649	Wonderland (1997)	10.0	3.200000
1650	Wooden Man's Bride, The (Wu Kui) (1994)	3.0	2.666667
1651	World of Apu, The (Apu Sansar) (1959)	6.0	4.000000
1652	Wrong Trousers, The (1993)	118.0	4.466102

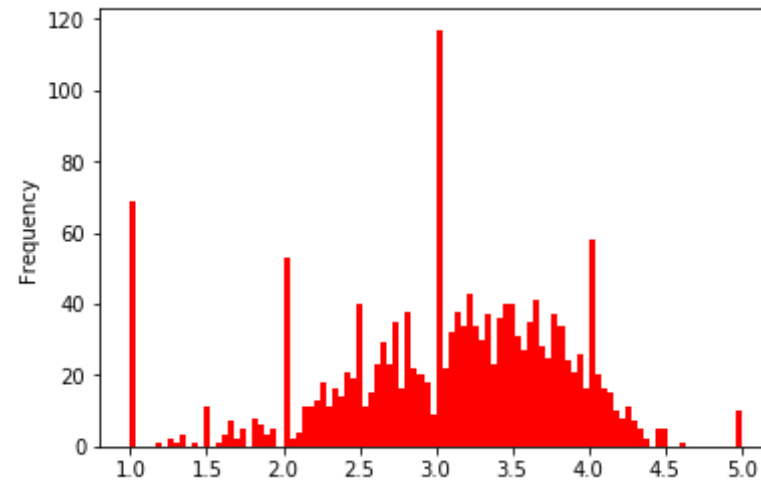
	title	count	mean
1653	Wyatt Earp (1994)	50.0	3.100000
1654	Yankee Zulu (1994)	1.0	1.000000
1655	Year of the Horse (1997)	7.0	3.285714
1656	You So Crazy (1994)	1.0	3.000000
1657	Young Frankenstein (1974)	200.0	3.945000
1658	Young Guns (1988)	101.0	3.207921
1659	Young Guns II (1990)	44.0	2.772727
1660	Young Poisoner's Handbook, The (1995)	41.0	3.341463
1661	Zeus and Roxanne (1997)	6.0	2.166667
1662	unknown	9.0	3.444444
1663	Á köldum klaka (Cold Fever) (1994)	1.0	3.000000

1664 rows × 3 columns

Visualization Mean Vs Count

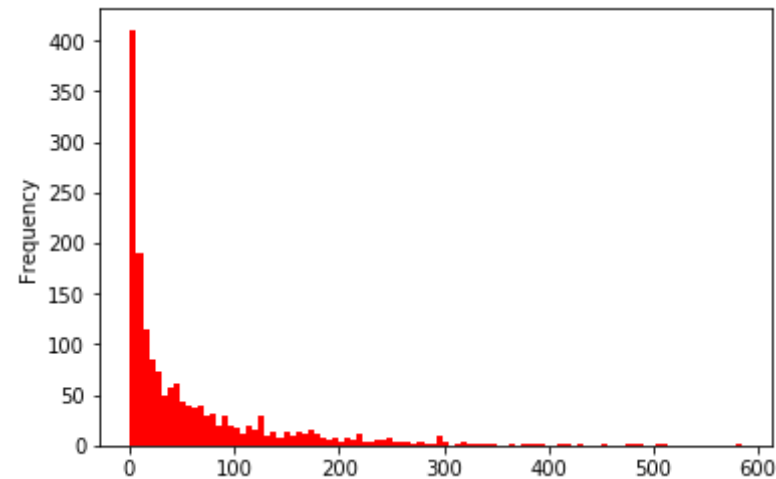
```
In [76]: ratings_mean_count_df['mean'].plot(bins=100, kind='hist', color = 'r')
```

```
Out[76]: <matplotlib.axes._subplots.AxesSubplot at 0x19171938be0>
```



```
In [77]: ratings_mean_count_df['count'].plot(bins=100, kind='hist', color = 'r')
```

```
Out[77]: <matplotlib.axes._subplots.AxesSubplot at 0x19171bf2780>
```



```
In [78]: # Let's see the highest rated movies!  
# Apparently these movies does not have many reviews (i.e.: small numbe
```

```
r of ratings)
ratings_mean_count_df[ratings_mean_count_df['mean'] == 5]
```

Out[78]:

	count	mean
title		
Aiqing wansui (1994)	1.0	5.0
Entertaining Angels: The Dorothy Day Story (1996)	1.0	5.0
Great Day in Harlem, A (1994)	1.0	5.0
Marlene Dietrich: Shadow and Light (1996)	1.0	5.0
Prefontaine (1997)	3.0	5.0
Saint of Fort Washington, The (1993)	2.0	5.0
Santa with Muscles (1996)	2.0	5.0
Someone Else's America (1995)	1.0	5.0
Star Kid (1997)	3.0	5.0
They Made Me a Criminal (1939)	1.0	5.0

```
In [80]: # List all the movies that are most rated

ratings_mean_count_df.sort_values('count', ascending = False).head(100)
```

Out[80]:

	count	mean
title		
Star Wars (1977)	584.0	4.359589
Contact (1997)	509.0	3.803536
Fargo (1996)	508.0	4.155512
Return of the Jedi (1983)	507.0	4.007890
Liar Liar (1997)	485.0	3.156701

	count	mean
title		
English Patient, The (1996)	481.0	3.656965
Scream (1996)	478.0	3.441423
Toy Story (1995)	452.0	3.878319
Air Force One (1997)	431.0	3.631090
Independence Day (ID4) (1996)	429.0	3.438228
Raiders of the Lost Ark (1981)	420.0	4.252381
Godfather, The (1972)	413.0	4.283293
Pulp Fiction (1994)	394.0	4.060914
Twelve Monkeys (1995)	392.0	3.798469
Silence of the Lambs, The (1991)	390.0	4.289744
Jerry Maguire (1996)	384.0	3.710938
Chasing Amy (1997)	379.0	3.839050
Rock, The (1996)	378.0	3.693122
Empire Strikes Back, The (1980)	368.0	4.206522
Star Trek: First Contact (1996)	365.0	3.660274
Titanic (1997)	350.0	4.245714
Back to the Future (1985)	350.0	3.834286
Mission: Impossible (1996)	344.0	3.313953
Fugitive, The (1993)	336.0	4.044643
Indiana Jones and the Last Crusade (1989)	331.0	3.930514
Willy Wonka and the Chocolate Factory (1971)	326.0	3.631902
Princess Bride, The (1987)	324.0	4.172840
Forrest Gump (1994)	321.0	3.853583
Saint, The (1997)	316.0	3.123418

	count	mean
title		
Monty Python and the Holy Grail (1974)	316.0	4.066456
...
Wizard of Oz, The (1939)	246.0	4.077236
Star Trek: The Wrath of Khan (1982)	244.0	3.815574
Phenomenon (1996)	244.0	3.557377
Casablanca (1942)	243.0	4.456790
Die Hard (1988)	243.0	3.872428
Sting, The (1973)	241.0	4.058091
Devil's Own, The (1997)	240.0	3.108333
Dante's Peak (1997)	240.0	2.933333
Graduate, The (1967)	239.0	4.104603
Psycho (1960)	239.0	4.100418
Seven (Se7en) (1995)	236.0	3.847458
Time to Kill, A (1996)	232.0	3.685345
It's a Wonderful Life (1946)	231.0	4.121212
In & Out (1997)	230.0	3.304348
Speed (1994)	230.0	3.647826
Hunt for Red October, The (1990)	227.0	3.863436
Stand by Me (1986)	227.0	3.881057
GoodFellas (1990)	226.0	3.951327
Heat (1995)	223.0	3.569507
Sound of Music, The (1965)	222.0	3.765766
Clockwork Orange, A (1971)	221.0	3.909502
Courage Under Fire (1996)	221.0	3.610860

	count	mean
title		
Apocalypse Now (1979)	221.0	4.045249
Top Gun (1986)	220.0	3.481818
Lion King, The (1994)	220.0	3.781818
Aladdin (1992)	219.0	3.812785
Babe (1995)	219.0	3.995434
Volcano (1997)	219.0	2.808219
To Kill a Mockingbird (1962)	219.0	4.292237
Murder at 1600 (1997)	218.0	3.087156

100 rows × 2 columns

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

Out[24]:

	user_id	item_id	rating	title
0	0	50	5	Star Wars (1977)
1	290	50	5	Star Wars (1977)
2	79	50	4	Star Wars (1977)
3	2	50	5	Star Wars (1977)
4	8	50	5	Star Wars (1977)

Performing Item based collaborative filtering on one Movie sample

In [81]: `userid_movietitle_matrix = movies_rating_df.pivot_table(index = 'user_id', columns = 'title', values = 'rating')`

In [82]: `userid_movietitle_matrix`

Out[82]:

	'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)
user_id										
0	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
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0	NaN
3	NaN	NaN	NaN	NaN	2.0	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5	NaN	NaN	2.0	NaN	NaN	NaN	NaN	4.0	NaN	NaN
6	NaN	NaN	NaN	4.0	NaN	NaN	NaN	5.0	NaN	NaN
7	NaN	NaN	NaN	4.0	NaN	NaN	5.0	5.0	NaN	4.0
8	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
9	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	4.0
10	NaN	NaN	NaN	5.0	NaN	NaN	NaN	5.0	NaN	4.0
11	NaN	NaN	NaN	NaN	NaN	NaN	NaN	4.0	NaN	NaN
12	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
13	NaN	NaN	2.0	4.0	NaN	NaN	2.0	5.0	1.0	4.0
14	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
15	NaN	NaN	3.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
16	NaN	NaN	NaN	5.0	NaN	NaN	NaN	4.0	NaN	NaN
17	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
18	NaN	NaN	NaN	3.0	NaN	NaN	NaN	3.0	NaN	NaN

	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)
user_id											
19	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
20	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
21	NaN	NaN	NaN	NaN	NaN	4.0	NaN	NaN	NaN	NaN	NaN
22	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
23	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
24	NaN	NaN	NaN	NaN	5.0	NaN	NaN	NaN	NaN	NaN	NaN
25	NaN	NaN	NaN	NaN	NaN	NaN	NaN	4.0	3.0	NaN	5.0
26	NaN	NaN	NaN	NaN	NaN	NaN	3.0	NaN	NaN	NaN	NaN
27	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
28	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
29	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...
914	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
915	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
916	NaN	NaN	NaN	NaN	NaN	NaN	4.0	NaN	4.0	NaN	NaN
917	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
918	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0	NaN	NaN
919	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
920	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
921	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
922	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2.0	NaN	NaN

	'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)
user_id										
923	NaN	NaN	NaN	NaN	NaN	4.0	NaN	NaN	NaN	NaN
924	NaN	NaN	NaN	5.0	NaN	3.0	NaN	NaN	NaN	NaN
925	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
926	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
927	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
928	NaN	NaN	NaN	NaN	NaN	NaN	NaN	4.0	NaN	NaN
929	NaN	NaN	NaN	NaN	NaN	NaN	NaN	5.0	NaN	NaN
930	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
931	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
932	NaN	NaN	2.0	5.0	NaN	NaN	4.0	5.0	NaN	5.0
933	NaN	NaN	NaN	NaN	NaN	NaN	NaN	4.0	NaN	NaN
934	NaN	NaN	2.0	NaN	NaN	NaN	NaN	4.0	NaN	NaN
935	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
936	NaN	NaN	NaN	NaN	NaN	4.0	NaN	NaN	NaN	NaN
937	NaN	NaN	2.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
938	NaN	NaN	4.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
939	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
940	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
941	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
942	NaN	NaN	NaN	NaN	NaN	NaN	NaN	3.0	NaN	3.0
943	NaN	NaN	NaN	NaN	NaN	2.0	NaN	NaN	NaN	NaN

944 rows × 1664 columns

```
In [83]: titanic = userid_movietitle_matrix['Titanic (1997)']
```

```
In [84]: titanic
```

```
Out[84]: user_id
0      NaN
1      NaN
2      5.0
3      NaN
4      NaN
5      NaN
6      NaN
7      NaN
8      NaN
9      NaN
10     NaN
11     NaN
12     NaN
13     4.0
14     2.0
15     NaN
16     NaN
17     NaN
18     NaN
19     2.0
20     NaN
21     NaN
22     NaN
23     NaN
24     NaN
25     NaN
26     5.0
27     NaN
28     NaN
29     NaN
...
```

```
914    3.0
915    4.0
916    NaN
917    NaN
918    NaN
919    5.0
920    5.0
921    5.0
922    NaN
923    NaN
924    4.0
925    NaN
926    3.0
927    NaN
928    NaN
929    NaN
930    NaN
931    4.0
932    NaN
933    NaN
934    3.0
935    5.0
936    4.0
937    NaN
938    5.0
939    NaN
940    5.0
941    NaN
942    3.0
943    NaN
Name: Titanic (1997), Length: 944, dtype: float64
```

```
In [85]: # Let's calculate the correlations
titanic_correlations = pd.DataFrame(userid_movietitle_matrix.corrwith(t
itanic), columns=['Correlation'])
titanic_correlations = titanic_correlations.join(ratings_mean_count_df[
'count'])
```

```
C:\Users\Soumyansh\Anaconda3\envs\tensorflow\lib\site-packages\numpy\li
b\function base.py:2522: RuntimeWarning: Degrees of freedom <= 0 for sl
```

```

ice
c = cov(x, y, rowvar)
C:\Users\Soumyansh\Anaconda3\envs\tensorflow\lib\site-packages\numpy\li
b\function_base.py:2451: RuntimeWarning: divide by zero encountered in
true_divide
c *= np.true_divide(1, fact)

```

In [86]: `titanic_correlations.sort_values('Correlation',ascending=False)`

Out[86]:

	Correlation	count
title		
Nadja (1994)	1.0	8.0
Pest, The (1997)	1.0	8.0
Savage Nights (Nuits fauves, Les) (1992)	1.0	3.0
For Ever Mozart (1996)	1.0	3.0
Jerky Boys, The (1994)	1.0	3.0
Newton Boys, The (1998)	1.0	4.0
Hearts and Minds (1996)	1.0	5.0
Simple Wish, A (1997)	1.0	3.0
Purple Noon (1960)	1.0	7.0
Year of the Horse (1997)	1.0	7.0
Titanic (1997)	1.0	350.0
Suture (1993)	1.0	4.0
Squeeze (1996)	1.0	3.0
Loaded (1994)	1.0	5.0
Albino Alligator (1996)	1.0	6.0
Wedding Gift, The (1994)	1.0	3.0

	Correlation	count
title		
Solo (1996)	1.0	12.0
To Have, or Not (1995)	1.0	2.0
Dangerous Ground (1997)	1.0	8.0
Getting Even with Dad (1994)	1.0	5.0
Bad Moon (1996)	1.0	6.0
Ed's Next Move (1996)	1.0	3.0
Sudden Manhattan (1996)	1.0	2.0
Underworld (1997)	1.0	4.0
C'est arrivé près de chez vous (1992)	1.0	4.0
Vermin (1998)	1.0	2.0
Bonheur, Le (1965)	1.0	4.0
Pie in the Sky (1995)	1.0	4.0
Scarlet Letter, The (1995)	1.0	5.0
Crossfire (1947)	1.0	4.0
...
Tough and Deadly (1995)	NaN	2.0
Touki Bouki (Journey of the Hyena) (1973)	NaN	1.0
Truth or Consequences, N.M. (1997)	NaN	3.0
Turning, The (1992)	NaN	2.0
Twisted (1996)	NaN	6.0
Two Deaths (1995)	NaN	4.0
Two Friends (1986)	NaN	1.0
Two or Three Things I Know About Her (1966)	NaN	4.0
Underground (1995)	NaN	5.0

	Correlation	count
title		
Underneath, The (1995)	NaN	4.0
Venice/Venice (1992)	NaN	2.0
Vermont Is For Lovers (1992)	NaN	1.0
Very Natural Thing, A (1974)	NaN	1.0
Vie est belle, La (Life is Rosey) (1987)	NaN	1.0
Visitors, The (Visiteurs, Les) (1993)	NaN	2.0
Walk in the Sun, A (1945)	NaN	1.0
War at Home, The (1996)	NaN	1.0
Wedding Bell Blues (1996)	NaN	1.0
Wend Kuuni (God's Gift) (1982)	NaN	1.0
Wife, The (1995)	NaN	1.0
Window to Paris (1994)	NaN	1.0
Wings of Courage (1995)	NaN	1.0
Witness (1985)	NaN	1.0
Woman in Question, The (1950)	NaN	1.0
Wonderful, Horrible Life of Leni Riefenstahl, The (1993)	NaN	10.0
Yankee Zulu (1994)	NaN	1.0
You So Crazy (1994)	NaN	1.0
Zeus and Roxanne (1997)	NaN	6.0
unknown	NaN	9.0
Á köldum klaka (Cold Fever) (1994)	NaN	1.0

1664 rows × 2 columns

In [87]: `titanic_correlations.dropna(inplace=True)`

```
titanic_correlations
```

Out[87]:

	Correlation	count
title		
'Til There Was You (1997)	-0.062017	9.0
101 Dalmatians (1996)	0.120113	109.0
12 Angry Men (1957)	0.077700	125.0
187 (1997)	0.315654	41.0
2 Days in the Valley (1996)	0.017295	93.0
20,000 Leagues Under the Sea (1954)	0.246718	72.0
2001: A Space Odyssey (1968)	-0.042024	259.0
39 Steps, The (1935)	0.155752	59.0
8 1/2 (1963)	-0.039301	38.0
8 Seconds (1994)	-1.000000	4.0
A Chef in Love (1996)	0.000000	8.0
Above the Rim (1994)	-1.000000	5.0
Absolute Power (1997)	0.327028	127.0
Abyss, The (1989)	0.472103	151.0
Ace Ventura: Pet Detective (1994)	0.104435	103.0
Ace Ventura: When Nature Calls (1995)	0.165004	37.0
Across the Sea of Time (1995)	0.755929	4.0
Addams Family Values (1993)	-0.018609	87.0
Addicted to Love (1997)	0.261187	54.0
Addiction, The (1995)	-0.080064	11.0
Adventures of Pinocchio, The (1996)	-0.662122	39.0
Adventures of Priscilla, Queen of the Desert, The (1994)	0.224003	111.0

	Correlation	count
title		
Adventures of Robin Hood, The (1938)	0.597532	67.0
Affair to Remember, An (1957)	0.289193	26.0
African Queen, The (1951)	0.099746	152.0
Afterglow (1997)	-0.228896	18.0
Age of Innocence, The (1993)	0.012832	65.0
Air Bud (1997)	-0.186933	43.0
Air Force One (1997)	0.383730	431.0
Air Up There, The (1994)	0.645497	16.0
...
White Squall (1996)	0.028843	85.0
Whole Wide World, The (1996)	-0.577350	6.0
Widows' Peak (1994)	-0.250801	19.0
Wild America (1997)	0.158114	9.0
Wild Bill (1995)	0.068041	12.0
Wild Bunch, The (1969)	-0.142731	43.0
Wild Reeds (1994)	-0.062017	14.0
Wild Things (1998)	0.628655	11.0
William Shakespeare's Romeo and Juliet (1996)	0.430243	106.0
Willy Wonka and the Chocolate Factory (1971)	-0.115819	326.0
Wings of Desire (1987)	0.043284	57.0
Wings of the Dove, The (1997)	-0.209230	75.0
Winnie the Pooh and the Blustery Day (1968)	0.138524	75.0
Winter Guest, The (1997)	0.000000	9.0
Wishmaster (1997)	0.205902	27.0

	Correlation	count
title		
With Honors (1994)	0.393241	46.0
Withnail and I (1987)	-0.721688	13.0
Wizard of Oz, The (1939)	0.155928	246.0
Wolf (1994)	0.230693	67.0
Women, The (1939)	0.394055	15.0
Wonderland (1997)	-0.845154	10.0
Wooden Man's Bride, The (Wu Kui) (1994)	1.000000	3.0
World of Apu, The (Apu Sansar) (1959)	-1.000000	6.0
Wrong Trousers, The (1993)	0.227949	118.0
Wyatt Earp (1994)	0.357506	50.0
Year of the Horse (1997)	1.000000	7.0
Young Frankenstein (1974)	0.107666	200.0
Young Guns (1988)	0.199931	101.0
Young Guns II (1990)	0.317274	44.0
Young Poisoner's Handbook, The (1995)	0.356783	41.0

1356 rows × 2 columns

```
In [88]: # Let's sort the correlations vector
titanic_correlations.sort_values('Correlation', ascending=False)
```

Out[88]:

	Correlation	count
title		
Nadja (1994)	1.0	8.0
Pest, The (1997)	1.0	8.0

	Correlation	count
title		
Savage Nights (Nuits fauves, Les) (1992)	1.0	3.0
For Ever Mozart (1996)	1.0	3.0
Jerky Boys, The (1994)	1.0	3.0
Newton Boys, The (1998)	1.0	4.0
Hearts and Minds (1996)	1.0	5.0
Simple Wish, A (1997)	1.0	3.0
Purple Noon (1960)	1.0	7.0
Year of the Horse (1997)	1.0	7.0
Titanic (1997)	1.0	350.0
Suture (1993)	1.0	4.0
Squeeze (1996)	1.0	3.0
Loaded (1994)	1.0	5.0
Albino Alligator (1996)	1.0	6.0
Wedding Gift, The (1994)	1.0	3.0
Solo (1996)	1.0	12.0
To Have, or Not (1995)	1.0	2.0
Dangerous Ground (1997)	1.0	8.0
Getting Even with Dad (1994)	1.0	5.0
Bad Moon (1996)	1.0	6.0
Ed's Next Move (1996)	1.0	3.0
Sudden Manhattan (1996)	1.0	2.0
Underworld (1997)	1.0	4.0
C'est arrivé près de chez vous (1992)	1.0	4.0
Vermin (1998)	1.0	2.0

	Correlation	count
title		
Bonheur, Le (1965)	1.0	4.0
Pie in the Sky (1995)	1.0	4.0
Scarlet Letter, The (1995)	1.0	5.0
Crossfire (1947)	1.0	4.0
...
Anna (1996)	-1.0	2.0
Maya Lin: A Strong Clear Vision (1994)	-1.0	4.0
Metisse (Café au Lait) (1993)	-1.0	6.0
Fille seule, La (A Single Girl) (1995)	-1.0	4.0
Rendezvous in Paris (Rendez-vous de Paris, Les) (1995)	-1.0	3.0
Even Cowgirls Get the Blues (1993)	-1.0	5.0
Nelly & Monsieur Arnaud (1995)	-1.0	3.0
Stripes (1981)	-1.0	5.0
Substance of Fire, The (1996)	-1.0	4.0
Angel and the Badman (1947)	-1.0	6.0
Some Mother's Son (1996)	-1.0	2.0
Secret Adventures of Tom Thumb, The (1993)	-1.0	5.0
Madame Butterfly (1995)	-1.0	7.0
Alphaville (1965)	-1.0	12.0
Good Man in Africa, A (1994)	-1.0	2.0
Mark of Zorro, The (1940)	-1.0	13.0
The Innocent (1994)	-1.0	4.0
Ciao, Professore! (1993)	-1.0	4.0
Caro Diario (Dear Diary) (1994)	-1.0	4.0

	Correlation	count
title		
Stars Fell on Henrietta, The (1995)	-1.0	3.0
Germinal (1993)	-1.0	4.0
Above the Rim (1994)	-1.0	5.0
My Favorite Season (1993)	-1.0	3.0
Lamerica (1994)	-1.0	4.0
Children of the Revolution (1996)	-1.0	5.0
Pather Panchali (1955)	-1.0	8.0
Angel Baby (1995)	-1.0	4.0
Blood Beach (1981)	-1.0	6.0
Two Bits (1995)	-1.0	5.0
Faces (1968)	-1.0	4.0

1356 rows × 2 columns

```
In [89]: titanic_correlations[titanic_correlations['count']>80].sort_values('Correlation',ascending=False)
```

Out[89]:

	Correlation	count
title		
Titanic (1997)	1.000000	350.0
River Wild, The (1994)	0.497600	146.0
Abyss, The (1989)	0.472103	151.0
Bram Stoker's Dracula (1992)	0.443560	120.0
True Lies (1994)	0.435104	208.0
William Shakespeare's Romeo and Juliet (1996)	0.430243	106.0

	Correlation	count
title		
Last of the Mohicans, The (1992)	0.427239	128.0
Outbreak (1995)	0.426237	104.0
Demolition Man (1993)	0.411489	92.0
Die Hard (1988)	0.403124	243.0
Frighteners, The (1996)	0.398516	115.0
Somewhere in Time (1980)	0.394700	82.0
Air Force One (1997)	0.383730	431.0
Jurassic Park (1993)	0.382785	261.0
Star Trek: Generations (1994)	0.382300	116.0
Hamlet (1996)	0.381213	90.0
Father of the Bride Part II (1995)	0.381120	128.0
Waterworld (1995)	0.376446	102.0
Evil Dead II (1987)	0.374610	89.0
E.T. the Extra-Terrestrial (1982)	0.373201	300.0
Pretty Woman (1990)	0.372027	164.0
Eraser (1996)	0.371140	206.0
Nixon (1995)	0.365371	90.0
Twister (1996)	0.362297	293.0
Under Siege (1992)	0.360960	124.0
Maverick (1994)	0.359468	128.0
Dante's Peak (1997)	0.357226	240.0
Braveheart (1995)	0.356765	297.0
Firm, The (1993)	0.353753	151.0
GoldenEye (1995)	0.350824	131.0

	Correlation	count
title		
...
Henry V (1989)	-0.091609	124.0
Hudsucker Proxy, The (1994)	-0.094045	110.0
Up Close and Personal (1996)	-0.109985	85.0
Willy Wonka and the Chocolate Factory (1971)	-0.115819	326.0
Postino, Il (1994)	-0.116010	183.0
Annie Hall (1977)	-0.120213	180.0
Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb (1963)	-0.121776	194.0
Philadelphia Story, The (1940)	-0.122325	104.0
Reservoir Dogs (1992)	-0.125392	148.0
Dead Man Walking (1995)	-0.129370	299.0
Crying Game, The (1992)	-0.129470	119.0
Maltese Falcon, The (1941)	-0.139893	138.0
Lone Star (1996)	-0.157035	187.0
Big Night (1996)	-0.157361	171.0
My Life as a Dog (Mitt liv som hund) (1985)	-0.161814	93.0
Duck Soup (1933)	-0.182252	93.0
My Fellow Americans (1996)	-0.188786	86.0
Spawn (1997)	-0.191710	143.0
Emma (1996)	-0.195052	177.0
Breakfast at Tiffany's (1961)	-0.198325	95.0
Mighty Aphrodite (1995)	-0.199766	184.0
Rosencrantz and Guildenstern Are Dead (1990)	-0.208944	90.0
Raging Bull (1980)	-0.223660	116.0

	Correlation	count
title		
Escape from New York (1981)	-0.225129	91.0
Jackie Brown (1997)	-0.236841	126.0
Brazil (1985)	-0.243532	208.0
Richard III (1995)	-0.275451	89.0
Cold Comfort Farm (1995)	-0.307150	125.0
Unbearable Lightness of Being, The (1988)	-0.314476	92.0
James and the Giant Peach (1996)	-0.370248	126.0

416 rows × 2 columns

Creating Collaborative filter on entire dataset

In [91]: `userid_movietitle_matrix`

Out[91]:

	'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)
user_id										
0	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
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0	NaN
3	NaN	NaN	NaN	NaN	2.0	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5	NaN	NaN	2.0	NaN	NaN	NaN	NaN	4.0	NaN	NaN

	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)
user_id											
6	NaN	NaN	NaN	NaN	4.0	NaN	NaN	NaN	5.0	NaN	NaN
7	NaN	NaN	NaN	NaN	4.0	NaN	NaN	5.0	5.0	NaN	4.0
8	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
9	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	4.0
10	NaN	NaN	NaN	NaN	5.0	NaN	NaN	NaN	5.0	NaN	4.0
11	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	4.0	NaN	NaN
12	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
13	NaN	NaN	2.0	4.0	NaN	NaN	2.0	5.0	1.0	4.0	
14	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
15	NaN	NaN	3.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
16	NaN	NaN	NaN	5.0	NaN	NaN	NaN	4.0	NaN	NaN	
17	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
18	NaN	NaN	NaN	3.0	NaN	NaN	NaN	3.0	NaN	NaN	
19	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
20	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
21	NaN	NaN	NaN	NaN	4.0	NaN	NaN	NaN	NaN	NaN	NaN
22	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
23	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
24	NaN	NaN	NaN	5.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
25	NaN	NaN	NaN	NaN	NaN	NaN	4.0	3.0	NaN	5.0	
26	NaN	NaN	NaN	NaN	NaN	3.0	NaN	NaN	NaN	NaN	NaN

	'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)
user_id										
27	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
28	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
29	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...
914	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
915	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
916	NaN	NaN	NaN	NaN	NaN	4.0	NaN	4.0	NaN	NaN
917	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
918	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0	NaN	NaN
919	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
920	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
921	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
922	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2.0	NaN	NaN
923	NaN	NaN	NaN	NaN	NaN	4.0	NaN	NaN	NaN	NaN
924	NaN	NaN	NaN	5.0	NaN	3.0	NaN	NaN	NaN	NaN
925	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
926	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
927	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
928	NaN	NaN	NaN	NaN	NaN	NaN	NaN	4.0	NaN	NaN
929	NaN	NaN	NaN	NaN	NaN	NaN	NaN	5.0	NaN	NaN
930	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

	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)
user_id											
931	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
932	NaN	NaN	2.0	5.0	NaN	NaN	4.0	5.0	NaN	5.0	
933	NaN	NaN	NaN	NaN	NaN	NaN	NaN	4.0	NaN	NaN	
934	NaN	NaN	2.0	NaN	NaN	NaN	NaN	4.0	NaN	NaN	
935	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
936	NaN	NaN	NaN	NaN	NaN	4.0	NaN	NaN	NaN	NaN	
937	NaN	NaN	2.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
938	NaN	NaN	4.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
939	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
940	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
941	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
942	NaN	NaN	NaN	NaN	NaN	NaN	NaN	3.0	NaN	3.0	
943	NaN	NaN	NaN	NaN	NaN	2.0	NaN	NaN	NaN	NaN	

944 rows × 1664 columns



```
In [92]: movie_correlations = userid_movietitle_matrix.corr(method = 'pearson',
# pearson : standard correlation coefficient
# Obtain the correlations between all movies in the dataframe
```

```
In [93]: movie_correlations
```

Out[93]:

	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)
	title									
	'Til There Was You (1997)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	1-900 (1994)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	101 Dalmatians (1996)	NaN	NaN	1.0	NaN	NaN	NaN	NaN	NaN	NaN
	12 Angry Men (1957)	NaN	NaN	NaN	1.000000	NaN	NaN	NaN	0.178848	NaN
	187 (1997)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	2 Days in the Valley (1996)	NaN	NaN	NaN	NaN	NaN	1.0	NaN	NaN	NaN
	20,000 Leagues Under the Sea (1954)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	2001: A Space Odyssey (1968)	NaN	NaN	NaN	0.178848	NaN	NaN	NaN	1.000000	NaN
	3 Ninjas: High Noon At Mega Mountain (1998)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	39 Steps, The (1935)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	8 1/2 (1963)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

	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)
	title									
	8 Heads in a Duffel Bag (1997)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	8 Seconds (1994)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	A Chef in Love (1996)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Above the Rim (1994)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Absolute Power (1997)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Abyss, The (1989)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.089206	NaN
	Ace Ventura: Pet Detective (1994)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Ace Ventura: When Nature Calls (1995)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Across the Sea of Time (1995)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Addams Family Values (1993)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Addicted to Love (1997)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

	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)
	title									
	Addiction, The (1995)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Adventures of Pinocchio, The (1996)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Adventures of Priscilla, Queen of the Desert, The (1994)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Adventures of Robin Hood, The (1938)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Affair to Remember, An (1957)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	African Queen, The (1951)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.152433	NaN
	Afterglow (1997)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Age of Innocence, The (1993)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

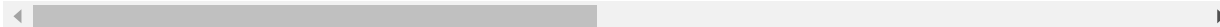
	Window to Paris (1994)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

	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)
	title									
	Wings of Courage (1995)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Wings of Desire (1987)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Wings of the Dove, The (1997)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Winnie the Pooh and the Blustery Day (1968)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Winter Guest, The (1997)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Wishmaster (1997)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	With Honors (1994)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Withnail and I (1987)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Witness (1985)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Wizard of Oz, The (1939)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.236830	NaN
	Wolf (1994)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

	'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)
title									
Woman in Question, The (1950)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Women, The (1939)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Wonderful, Horrible Life of Leni Riefenstahl, The (1993)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Wonderland (1997)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Wooden Man's Bride, The (Wu Kui) (1994)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
World of Apu, The (Apu Sansar) (1959)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Wrong Trousers, The (1993)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Wyatt Earp (1994)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Yankee Zulu (1994)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Year of the Horse (1997)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

	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)
	title									
	You So Crazy (1994)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Young Frankenstein (1974)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	-0.001307	NaN
	Young Guns (1988)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Young Guns II (1990)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Young Poisoner's Handbook, The (1995)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Zeus and Roxanne (1997)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	unknown	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Á köldum klaka (Cold Fever) (1994)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

1664 rows × 1664 columns



```
In [96]: ## Creating Data frame of own ratings
myRatings = pd.read_csv("My_Ratings.csv")
#myRatings.reset_index
```

```
In [97]: myRatings
```

Out[97]:

	Movie Name	Ratings
0	Liar Liar (1997)	5
1	Star Wars (1977)	1

In [98]: `len(myRatings.index)`

Out[98]: 2

In [99]: `myRatings['Movie Name'][0]`

Out[99]: 'Liar Liar (1997)'

```
In [101]: similar_movies_list = pd.Series()
          for i in range(0, 2):
              similar_movie = movie_correlations[myRatings['Movie Name'][i]].drop
              na() # Get same movies with same ratings
              similar_movie = similar_movie.map(lambda x: x * myRatings['Ratings'
              ][i]) # Scale the similarity by given ratings
              similar_movies_list = similar_movies_list.append(similar_movie)
```

```
In [102]: similar_movies_list.sort_values(inplace = True, ascending = False)
          print (similar_movies_list.head(10))
```

```
Liar Liar (1997)          5.000000
Con Air (1997)           2.349141
Pretty Woman (1990)      2.348951
Michael (1996)           2.210110
Indiana Jones and the Last Crusade (1989) 2.072136
Top Gun (1986)           2.028602
G.I. Jane (1997)         1.989656
Multiplicity (1996)       1.984302
Grumpier Old Men (1995)   1.953494
Ghost and the Darkness, The (1996) 1.895376
dtype: float64
```

In []:

