In [8]:
```
import pandas as pd

r_cols = ['user_id', 'movie_id', 'rating']
ratings = pd.read_csv(r'C:\Users\SAKTHI\Downloads\milestone\github\u.data', sep='\t', na

m_cols = ['movie_id', 'title']
movies = pd.read_csv(r'C:\Users\SAKTHI\Downloads\milestone\github\u.item', sep='|', name

ratings = pd.merge(movies, ratings)
```

In [9]:
```
ratings.head()
```

Out[9]:

	movie_id	title	user_id	rating
0	1	Toy Story (1995)	308	4
1	1	Toy Story (1995)	287	5
2	1	Toy Story (1995)	148	4
3	1	Toy Story (1995)	280	4
4	1	Toy Story (1995)	66	3

In [10]:
```
movieRatings = ratings.pivot_table(index=['user_id'],columns=['title'],values='rating')
movieRatings.head()
```

Out[10]:

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)	Y of Ho (19
user_id													
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	I
1	NaN	NaN	2.0	5.0	NaN	NaN	3.0	4.0	NaN	NaN	...	NaN	I
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0	NaN	...	NaN	I
3	NaN	NaN	NaN	NaN	2.0	NaN	NaN	NaN	NaN	NaN	...	NaN	I
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	I

5 rows × 1664 columns

In [11]:
```
starWarsRatings = movieRatings['Star Wars (1977)']
starWarsRatings.head()
```

Out[11]:
```
user_id
0    5.0
1    5.0
2    5.0
3    NaN
4    5.0
Name: Star Wars (1977), dtype: float64
```

In [12]:
```
similarMovies = movieRatings.corrwith(starWarsRatings)
similarMovies = similarMovies.dropna()
df = pd.DataFrame(similarMovies)
df.head(10)
```

```
C:\Users\SAKTHI\anaconda3\Lib\site-packages\numpy\lib\function_base.py:2846: RuntimeWarn
ing: Degrees of freedom <= 0 for slice
  c = cov(x, y, rowvar, dtype=dtype)
C:\Users\SAKTHI\anaconda3\Lib\site-packages\numpy\lib\function_base.py:2705: RuntimeWarn
ing: divide by zero encountered in divide
  c *= np.true_divide(1, fact)
```

Out[12]:

	0
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
2 Days in the Valley (1996)	0.066654
20,000 Leagues Under the Sea (1954)	0.289768
2001: A Space Odyssey (1968)	0.230884
39 Steps, The (1935)	0.106453
8 1/2 (1963)	-0.142977

In [13]:
```
similarMovies.sort_values(ascending=False)
```

Out[13]:
```
title
Hollow Reed (1996)            1.0
Commandments (1997)           1.0
Cosi (1996)                   1.0
No Escape (1994)              1.0
Stripes (1981)                1.0
                             ...
For Ever Mozart (1996)       -1.0
Frankie Starlight (1995)     -1.0
I Like It Like That (1994)   -1.0
American Dream (1990)        -1.0
Theodore Rex (1995)          -1.0
Length: 1410, dtype: float64
```

In [14]:
```
import numpy as np
movieStats = ratings.groupby('title').agg({'rating': [np.size, np.mean]})
movieStats.head()
```

Out[14]:

	rating	
	size	**mean**
title		
'Til There Was You (1997)	9	2.333333
1-900 (1994)	5	2.600000
101 Dalmatians (1996)	109	2.908257
12 Angry Men (1957)	125	4.344000
187 (1997)	41	3.024390

In [15]:
```
popularMovies = movieStats['rating']['size'] >= 100
movieStats[popularMovies].sort_values([('rating', 'mean')], ascending=False)[:15]
```

Out[15]:

title	rating size	rating mean
Close Shave, A (1995)	112	4.491071
Schindler's List (1993)	298	4.466443
Wrong Trousers, The (1993)	118	4.466102
Casablanca (1942)	243	4.456790
Shawshank Redemption, The (1994)	283	4.445230
Rear Window (1954)	209	4.387560
Usual Suspects, The (1995)	267	4.385768
Star Wars (1977)	584	4.359589
12 Angry Men (1957)	125	4.344000
Citizen Kane (1941)	198	4.292929
To Kill a Mockingbird (1962)	219	4.292237
One Flew Over the Cuckoo's Nest (1975)	264	4.291667
Silence of the Lambs, The (1991)	390	4.289744
North by Northwest (1959)	179	4.284916
Godfather, The (1972)	413	4.283293

In [21]:
```
df.head()
```

Out[21]:

title	0
'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 [ ]:

In [ ]: