In [21]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [22]:
```
from numpy.random import randn
np.random.seed(123)
```

In [23]:
```
m = ['movie_id', 'title', 'genres']
movies = pd.read_table('data/movies.dat', sep='::', header=None, names=m)
```

```
<ipython-input-23-6a95a2f964ea>:2: ParserWarning: Falling back to the 'pyth
on' engine because the 'c' engine does not support regex separators (separa
tors > 1 char and different from '\s+' are interpreted as regex); you can a
void this warning by specifying engine='python'.
  movies = pd.read_table('data/movies.dat', sep='::', header=None, names=m)
```

In [24]:
```
u = ['user_id', 'gender', 'age', 'occupation', 'zip']
users = pd.read_table('data/users.dat', sep='::', header=None, names=u)
```

```
<ipython-input-24-c02bf0d0c497>:2: ParserWarning: Falling back to the 'pyth
on' engine because the 'c' engine does not support regex separators (separa
tors > 1 char and different from '\s+' are interpreted as regex); you can a
void this warning by specifying engine='python'.
  users = pd.read_table('data/users.dat', sep='::', header=None, names=u)
```

In [25]:
```
r = ['user_id', 'movie_id', 'rating', 'timestamp']
ratings = pd.read_table('data/ratings.dat', sep='::', header=None, names=r)
```

```
<ipython-input-25-94ce83bf6a70>:2: ParserWarning: Falling back to the 'pyth
on' engine because the 'c' engine does not support regex separators (separa
tors > 1 char and different from '\s+' are interpreted as regex); you can a
void this warning by specifying engine='python'.
  ratings = pd.read_table('data/ratings.dat', sep='::', header=None, names=
r)
```

In [26]:
```
df = pd.merge(pd.merge(ratings, users), movies)
df
```

Out[26]:

	user_id	movie_id	rating	timestamp	gender	age	occupation	zip	title
0	1	1193	5	978300760	F	1	10	48067	One Flew Over the Cuckoo's Nest (1975)
1	2	1193	5	978298413	M	56	16	70072	One Flew Over the Cuckoo's Nest (1975)
2	12	1193	4	978220179	M	25	12	32793	One Flew Over the Cuckoo's Nest (1975)
3	15	1193	4	978199279	M	25	7	22903	One Flew Over the Cuckoo's Nest (1975)
4	17	1193	5	978158471	M	50	1	95350	One Flew Over the Cuckoo's Nest (1975)
...	...	...	...	...	...	...	...	...	...
1000204	5949	2198	5	958846401	M	18	17	47901	Modulations (1998)
1000205	5675	2703	3	976029116	M	35	14	30030	Broken Vessels (1998)
1000206	5780	2845	1	958153068	M	18	17	92886	White Boys (1999)
1000207	5851	3607	5	957756608	F	18	20	55410	One Little Indian (1973)
1000208	5938	2909	4	957273353	M	25	1	35401	Five Wives, Three Secretaries and Me (1998)

1000209 rows × 10 columns

In [27]:
```
df.iloc[0]
```

Out[27]:
```
user_id                                          1
movie_id                                      1193
rating                                           5
timestamp                                978300760
gender                                           F
age                                              1
occupation                                      10
zip                                          48067
title         One Flew Over the Cuckoo's Nest (1975)
genres                                       Drama
Name: 0, dtype: object
```

In [28]:
```
mr = df.pivot_table('rating', index='title', columns='gender', aggfunc='mean'
mr[:3]   # mr : mean_ratings
```

Out[28]:

gender	F	M
title		
$1,000,000 Duck (1971)	3.375000	2.761905
'Night Mother (1986)	3.388889	3.352941
'Til There Was You (1997)	2.675676	2.733333

In [29]:
```
rt = df.groupby('title').size()     # rt : ratings_by_title
rt[:10]
```

Out[29]:
```
title
$1,000,000 Duck (1971)                37
'Night Mother (1986)                  70
'Til There Was You (1997)             52
'burbs, The (1989)                   303
...And Justice for All (1979)        199
1-900 (1994)                           2
10 Things I Hate About You (1999)    700
101 Dalmatians (1961)                565
101 Dalmatians (1996)                364
12 Angry Men (1957)                  616
dtype: int64
```

In [30]:
```
at = rt.index[rt >= 250]   # at :active_titles
at
```

Out[30]:
```
Index(['''burbs, The (1989)', '10 Things I Hate About You (1999)',
       '101 Dalmatians (1961)', '101 Dalmatians (1996)', '12 Angry Men (195
7)',
       '13th Warrior, The (1999)', '2 Days in the Valley (1996)',
       '20,000 Leagues Under the Sea (1954)', '2001: A Space Odyssey (196
8)',
       '2010 (1984)',
       ...
       'X-Men (2000)', 'Year of Living Dangerously (1982)',
       'Yellow Submarine (1968)', 'You've Got Mail (1998)',
       'Young Frankenstein (1974)', 'Young Guns (1988)',
       'Young Guns II (1990)', 'Young Sherlock Holmes (1985)',
       'Zero Effect (1998)', 'eXistenZ (1999)'],
      dtype='object', name='title', length=1216)
```

In [31]:
```
mr = mr.loc[at]
mr
```

Out[31]:

gender	F	M
title		
'burbs, The (1989)	2.793478	2.962085
10 Things I Hate About You (1999)	3.646552	3.311966
101 Dalmatians (1961)	3.791444	3.500000
101 Dalmatians (1996)	3.240000	2.911215
12 Angry Men (1957)	4.184397	4.328421
...	...	...
Young Guns (1988)	3.371795	3.425620
Young Guns II (1990)	2.934783	2.904025
Young Sherlock Holmes (1985)	3.514706	3.363344
Zero Effect (1998)	3.864407	3.723140
eXistenZ (1999)	3.098592	3.289086

1216 rows × 2 columns

In [32]:
```
tm = mr.sort_values(by='M', ascending=False)
tm[:15]                                    # tm : top_male_ratings
```

Out[32]:

gender	F	M
title		
Godfather, The (1972)	4.314700	4.583333
Seven Samurai (The Magnificent Seven) (Shichinin no samurai) (1954)	4.481132	4.576628
Shawshank Redemption, The (1994)	4.539075	4.560625
Raiders of the Lost Ark (1981)	4.332168	4.520597
Usual Suspects, The (1995)	4.513317	4.518248
Star Wars: Episode IV - A New Hope (1977)	4.302937	4.495307
Schindler's List (1993)	4.562602	4.491415
Wrong Trousers, The (1993)	4.588235	4.478261
Close Shave, A (1995)	4.644444	4.473795
Rear Window (1954)	4.484536	4.472991
Double Indemnity (1944)	4.282051	4.468354
Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb (1963)	4.376623	4.464789
Sunset Blvd. (a.k.a. Sunset Boulevard) (1950)	4.572650	4.464589
Casablanca (1942)	4.300990	4.461340
Third Man, The (1949)	4.466019	4.448276

In [33]:
```
tf = mr.sort_values(by='F', ascending=False)
tf[:15]                                    # tf : top_female_ratings
```

Out[33]:

gender	F	M
title		
Close Shave, A (1995)	4.644444	4.473795
Wrong Trousers, The (1993)	4.588235	4.478261
Sunset Blvd. (a.k.a. Sunset Boulevard) (1950)	4.572650	4.464589
Wallace & Gromit: The Best of Aardman Animation (1996)	4.563107	4.385075
Schindler's List (1993)	4.562602	4.491415
Shawshank Redemption, The (1994)	4.539075	4.560625
Grand Day Out, A (1992)	4.537879	4.293255
To Kill a Mockingbird (1962)	4.536667	4.372611
Creature Comforts (1990)	4.513889	4.272277
Usual Suspects, The (1995)	4.513317	4.518248
It Happened One Night (1934)	4.500000	4.163934
Rear Window (1954)	4.484536	4.472991
Seven Samurai (The Magnificent Seven) (Shichinin no samurai) (1954)	4.481132	4.576628
Sixth Sense, The (1999)	4.477410	4.379944
Third Man, The (1949)	4.466019	4.448276

Measuring Rating Disagreement

In [34]:
```
mr['diff'] = mr['F'] - mr['M']
mr.head(2)
```

Out[34]:

gender	F	M	diff
title			
'burbs, The (1989)	2.793478	2.962085	-0.168607
10 Things I Hate About You (1999)	3.646552	3.311966	0.334586

In [35]:
```
sbd = mr.sort_values(by='diff')
sbd[:5]                                    # sbd : sorted_by_diff
```

Out[35]:

gender	F	M	diff
title			
Good, The Bad and The Ugly, The (1966)	3.494949	4.221300	-0.726351
Kentucky Fried Movie, The (1977)	2.878788	3.555147	-0.676359
Dumb & Dumber (1994)	2.697987	3.336595	-0.638608
Longest Day, The (1962)	3.411765	4.031447	-0.619682
Cable Guy, The (1996)	2.250000	2.863787	-0.613787

In [36]:
```
sbd[::-1][:5]
```

Out[36]:

gender	F	M	diff
title			
Dirty Dancing (1987)	3.790378	2.959596	0.830782
Jumpin' Jack Flash (1986)	3.254717	2.578358	0.676359
Grease (1978)	3.975265	3.367041	0.608224
Little Women (1994)	3.870588	3.321739	0.548849
Steel Magnolias (1989)	3.901734	3.365957	0.535777

دانشگاه شهید مدنی آذربایجان
برنامه نویسی پیشرفته با پایتون
امین گلزاری اسکوئی
۱۴۰۰-۱۴۰۱

Codes and Projects (click here) (https://github.com/Amin-Golzari-Oskouei/Python-Programming-Course-Advanced-2021) slides and videos (click here) (https://drive.google.com/drive/folders/1Dx3v7fD1QBWL-MNP2hd7iIxaRbeALkkA)