In [1]:

```
import pandas as pd
```

In [2]:

```
import numpy as np
```

In [3]:

```
import warnings
```

In [4]:

```
warnings.filterwarnings("ignore")
```

In [5]:

```
columns_name = ["user_id","item_id","rating","timestamp"]

dataset = pd.read_csv("ml-100k/u.data",sep="\t"  ,names=columns_name)
```

In [6]:

```
# User_id is nothing but a person and item_id is movie name.

dataset.head()
```

Out[6]:

	user_id	item_id	rating	timestamp
0	196	242	3	881250949
1	186	302	3	891717742
2	22	377	1	878887116
3	244	51	2	880606923
4	166	346	1	886397596

In [7]:

```
dataset.shape
```

Out[7]:

```
(100000, 4)
```

In [ ]:

```

```

In [8]:

```
movie_titles = pd.read_csv("ml-100k/u.item",sep="\|",header=None)
```

In [9]:

```
movie_titles.head()
```

Out[9]:

	0	1	2	3	4	5	6	7	8	9	...	14	15	16	17	18	19	20	21	22	23
0	1	Toy Story (1995)	01-Jan-1995	NaN	http://us.imdb.com/M/title-exact?Toy%20Story%2...	0	0	0	1	1	...	0	0	0	0	0	0	0	0	0	0
1	2	GoldenEye (1995)	01-Jan-1995	NaN	http://us.imdb.com/M/title-exact?GoldenEye%20(...	0	1	1	0	0	...	0	0	0	0	0	0	0	1	0	0
2	3	Four Rooms (1995)	01-Jan-1995	NaN	http://us.imdb.com/M/title-exact?Four%20Rooms%...	0	0	0	0	0	...	0	0	0	0	0	0	0	1	0	0
3	4	Get Shorty (1995)	01-Jan-1995	NaN	http://us.imdb.com/M/title-exact?Get%20Shorty%...	0	1	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	5	Copycat (1995)	01-Jan-1995	NaN	http://us.imdb.com/M/title-exact?Copycat%20(1995)	0	0	0	0	0	...	0	0	0	0	0	0	0	1	0	0

5 rows × 24 columns

In [10]:

```
movie_titles.shape
```

Out[10]:

```
(1682, 24)
```

In [11]:

```
movie titles = movie titles[[0,1]]
```

In [12]:

```
movie_titles.head()
```

Out[12]:

	0	1
0	1	Toy Story (1995)
1	2	GoldenEye (1995)
2	3	Four Rooms (1995)
3	4	Get Shorty (1995)
4	5	Copycat (1995)

In [13]:

```
movie_titles.columns=["item_id","title"]
```

In [14]:

```
movie_titles
```

Out[14]:

	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)
...	...	...
1677	1678	Mat' i syn (1997)
1678	1679	B. Monkey (1998)
1679	1680	Sliding Doors (1998)

	item_id	title
1680	1681	You So Crazy (1994)
1681	1682	Scream of Stone (Schrei aus Stein) (1991)

1682 rows × 2 columns

In [15]:

```
final_dataset = pd.merge(dataset,movie_titles,on="item_id")
```

In [16]:

```
final_dataset.tail()
```

Out[16]:

	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 [17]:

```
final_dataset.shape
```

Out[17]:

```
(100000, 5)
```

In [18]:

```
final_dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 100000 entries, 0 to 99999
Data columns (total 5 columns):
 #   Column      Non-Null Count   Dtype
---  ------      --------------   -----
```

```
 0   user_id    100000 non-null  int64
 1   item_id    100000 non-null  int64
 2   rating     100000 non-null  int64
 3   timestamp  100000 non-null  int64
 4   title      100000 non-null  object
dtypes: int64(4), object(1)
memory usage: 4.6+ MB
```

In [19]:

```
final_dataset.isnull().sum().sum()
```

Out[19]:

```
0
```

Exploratory Data Analysis

In [20]:

```
import matplotlib.pyplot as plt
import seaborn as sns
sns.set_style("white")
```

In [21]:

```
final_dataset.groupby("title").mean()["rating"].sort_values(ascending=False)
```

Out[21]:

```
title
Marlene Dietrich: Shadow and Light (1996)       5.0
Prefontaine (1997)                              5.0
Santa with Muscles (1996)                       5.0
Star Kid (1997)                                 5.0
Someone Else's America (1995)                   5.0
                                               ...
Touki Bouki (Journey of the Hyena) (1973)       1.0
JLG/JLG - autoportrait de décembre (1994)       1.0
Daens (1992)                                    1.0
Butterfly Kiss (1995)                           1.0
Eye of Vichy, The (Oeil de Vichy, L') (1993)    1.0
Name: rating, Length: 1664, dtype: float64
```

In [22]:

```
# How many people give revue a particular movie

final_dataset.groupby("title").count()["rating"].sort_values(ascending=False)
```

Out[22]:

```
title
Star Wars (1977)                             583
Contact (1997)                               509
Fargo (1996)                                 508
Return of the Jedi (1983)                    507
Liar Liar (1997)                             485
                                            ...
Man from Down Under, The (1943)                1
Marlene Dietrich: Shadow and Light (1996)      1
Mat' i syn (1997)                              1
Mille bolle blu (1993)                         1
Á köldum klaka (Cold Fever) (1994)             1
Name: rating, Length: 1664, dtype: int64
```

In [23]:

```
final_dataset.groupby("title")
```

Out[23]:

```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x0000028FE6DC8508>
```

In [24]:

```
rating =  pd.DataFrame(final_dataset.groupby("title").mean()["rating"])
```

In [25]:

```
rating.head()
```

Out[25]:

	rating
title	
'Til There Was You (1997)	2.333333

title	rating
1-900 (1994)	2.600000
101 Dalmatians (1996)	2.908257
12 Angry Men (1957)	4.344000
187 (1997)	3.024390

In [26]:

```
rating["no. of rating "] =  final_dataset.groupby("title").count()["rating"]
```

In [27]:

```
rating.head()
```

Out[27]:

title	rating	no. of rating
'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 [28]:

```
rating.sort_values(by="rating",ascending=False)
```

Out[28]:

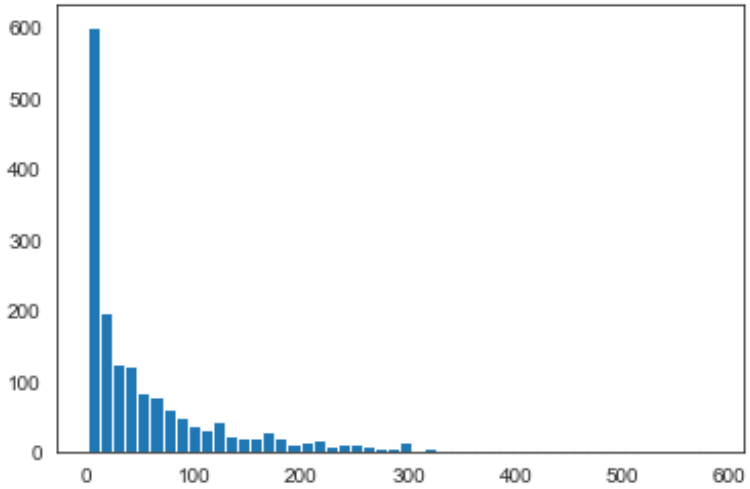
title	rating	no. of rating
They Made Me a Criminal (1939)	5.0	1

title	rating	no. of rating
Marlene Dietrich: Shadow and Light (1996)	5.0	1
Saint of Fort Washington, The (1993)	5.0	2
Someone Else's America (1995)	5.0	1
Star Kid (1997)	5.0	3
...	...	...
Eye of Vichy, The (Oeil de Vichy, L') (1993)	1.0	1
King of New York (1990)	1.0	1
Touki Bouki (Journey of the Hyena) (1973)	1.0	1
Bloody Child, The (1996)	1.0	1
Crude Oasis, The (1995)	1.0	1

1664 rows × 2 columns

In [29]:

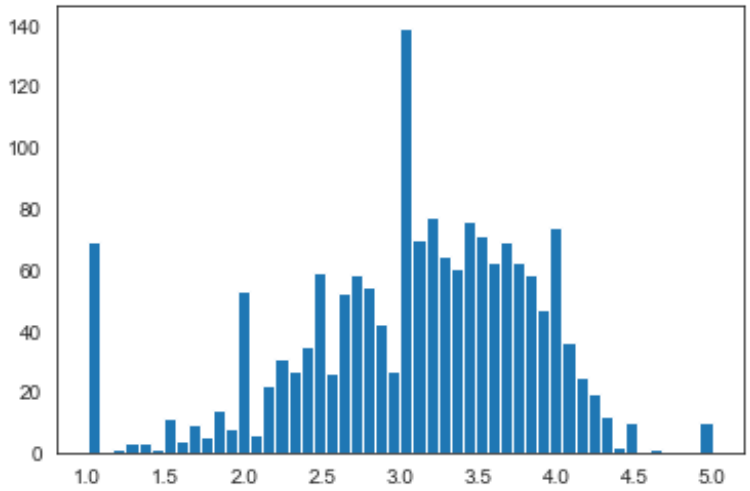
```
plt.hist(rating["no. of rating "],bins=50)
plt.show()
```



In [30]:

```
# Most of the people rated the movie is threes.
```

```
plt.hist(rating["rating"],bins=50)
plt.show()
```

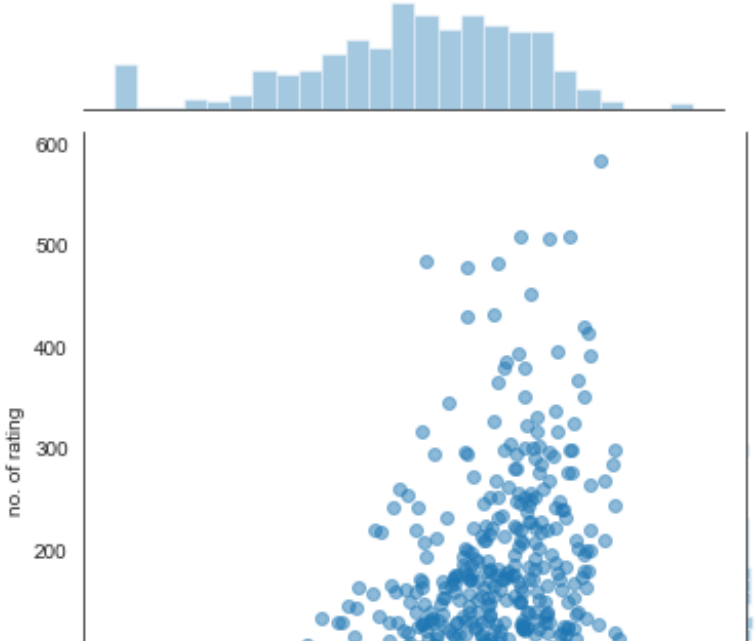


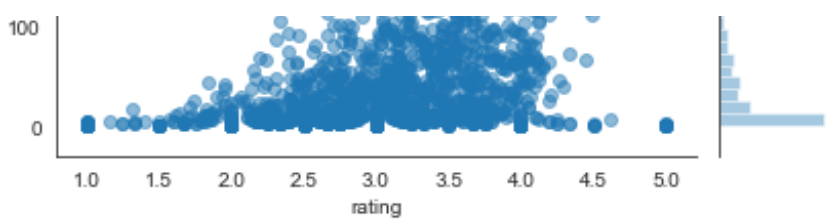
In [31]:

```
sns.jointplot(x="rating" , y= "no. of rating " , data=rating,alpha=0.5)
```

Out[31]:

```
<seaborn.axisgrid.JointGrid at 0x28ff068a888>
```





Create Movie Recommandation System

In [32]:

```
final_dataset.head()
```

Out[32]:

	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 [33]:

```
movie_matrix = final_dataset.pivot_table(index='user_id',columns="title",values="rating")
```

In [34]:

```
movie_matrix.head()
```

Out[34]:

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)	Young Guns (1988)	Young Guns II (1990)	Young Poisoner's Handbook, The (1995)

user_id	'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)	Young Guns (1988)	Young Guns II (1990)	Young Poisoner's Handbook, The (1995)
1	NaN	NaN	2.0	5.0	NaN	NaN	3.0	4.0	NaN	NaN	...	NaN	NaN	NaN	5.0	3.0	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN	2.0	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5	NaN	NaN	2.0	NaN	NaN	NaN	NaN	4.0	NaN	NaN	...	NaN	NaN	NaN	4.0	NaN	NaN	NaN

5 rows × 1664 columns

In [35]:

```
rating.sort_values(by="no. of rating ",ascending=False).head()
```

Out[35]:

title	rating	no. of rating
Star Wars (1977)	4.358491	583
Contact (1997)	3.803536	509
Fargo (1996)	4.155512	508
Return of the Jedi (1983)	4.007890	507
Liar Liar (1997)	3.156701	485

In [36]:

```
star_wars_movie_rating = movie_matrix["Star Wars (1977)"]
```

In [37]:

```
star_wars_movie_rating.head()
```

Out[37]:

```
user_id
1    5.0
```

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

In [38]:

```
# Co-Relation between Star_Wars and others novies
```

In [39]:

```
similar_to_starwars = movie_matrix.corrwith(star_wars_movie_rating)
```

In [40]:

```
similar_to_starwars
```

Out[40]:

```
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 [41]:

```
corr_with_starwars = pd.DataFrame(similar_to_starwars,columns=["corelation"])
```

In [42]:

```
corr_with_starwars
```

Out[42]:

	corelation
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

1664 rows × 1 columns

In [43]:

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

In [44]:

```
corr_with_starwars
```

Out[44]:

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

	corelation
187 (1997)	0.627698
title	...
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 [45]:

```
corr_with_starwars.sort_values(by="corelation",ascending=False).head(10)
```

Out[45]:

	corelation
title	
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 [46]:

```
rating
```

Out[46]:

title	rating	no. of rating
'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
...	...	...
Young Guns II (1990)	2.772727	44
Young Poisoner's Handbook, The (1995)	3.341463	41
Zeus and Roxanne (1997)	2.166667	6
unknown	3.444444	9
Á köldum klaka (Cold Fever) (1994)	3.000000	1

1664 rows × 2 columns

In [47]:

```
corr_with_starwars = corr_with_starwars.join(rating).drop(columns="rating")
```

In [48]:

```
corr_with_starwars.head()
```

Out[48]:

title	corelation	no. of rating

title	corelation	no. of rating
'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 [54]:

```
corr_with_starwars[corr_with_starwars["no. of rating "]>100].sort_values("corelation",ascending=False)
```

Out[54]:

title	corelation	no. of rating
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

Predicted Movies

In [61]:

```
def predicted_movies(movie_name):
    movie_user_rating = movie_matrix[movie_name]
    similar_to_movie = movie_matrix.corrwith(movie_user_rating)

    corr_movie =  pd.DataFrame(similar_to_movie,columns=["corelation"])
    corr_movie.dropna(inplace=True)

    corr_movie = corr_movie.join(rating["no. of rating "])
    prediction =  corr_movie[corr_movie["no. of rating "]>100].sort_values("corelation",ascending=False)
    return prediction
```

In [66]:

```
predictions = predicted_movies("Titanic (1997)")
```

In [68]:

```
predictions.head()
```

Out[68]:

title	corelation	no. of rating
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 [ ]:

```

```