In [2]:

```
import numpy as pd
```

In [3]:

```
import pandas as pd
```

In [4]:

```
column_names=['user_id','item_id','rating','timestamp']
```

In [5]:

```
df=pd.read_csv('u.data',sep='\t',names=column_names)
```

In [6]:

```
df.head()
```

Out[6]:

	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 [7]:

```
movie_titles=pd.read_csv('Movie_Id_Titles')
```

In [9]:

```
movie_titles.head()
```

Out[9]:

	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 [10]:

```
df=pd.merge(df,movie_titles,on='item_id')
```

In [13]:

```
df.head()
```

Out[13]:

	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 [14]:

```
import matplotlib.pyplot as plt
```

In [15]:

```
import seaborn as sns
```

In [16]:

```
sns.set_style('white')
```

In [17]:

```
%matplotlib inline
```

In [20]:

```
df.groupby('title')['rating'].mean().sort_values(ascending=False).head()
```

Out[20]:

```
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
Name: rating, dtype: float64
```

In [21]:

```
df.groupby('title')['rating'].count().sort_values(ascending=False).head()
```

Out[21]:

```
title
Star Wars (1977)             584
Contact (1997)               509
Fargo (1996)                 508
Return of the Jedi (1983)    507
Liar Liar (1997)             485
Name: rating, dtype: int64
```

In [22]:

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

In [23]:

```
ratings.head()
```

Out[23]:

	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 [24]:

```
ratings['number of ratings']=pd.DataFrame(df.groupby('title')['rating'].count())
```

In [25]:

```
ratings.head()
```

Out[25]:

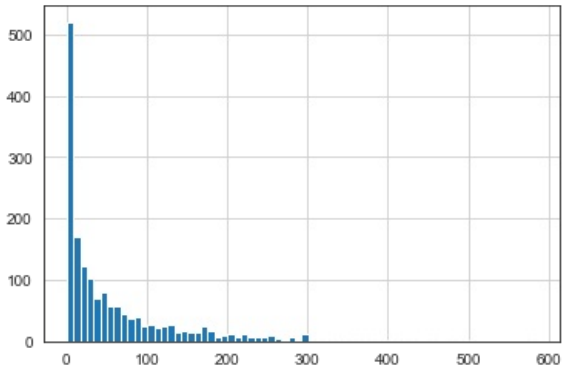
title	rating	number of ratings
'Til There Was You (1997)	2.333333	9
1-900 (1994)	2.600000	5
101 Dalmatians (1996)	2.908257	109
12 Angry Men (1957)	4.344000	125
187 (1997)	3.024390	41

In [29]:

```
ratings['number of ratings'].hist(bins=70)
```

Out[29]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x21c6a4bf308>
```

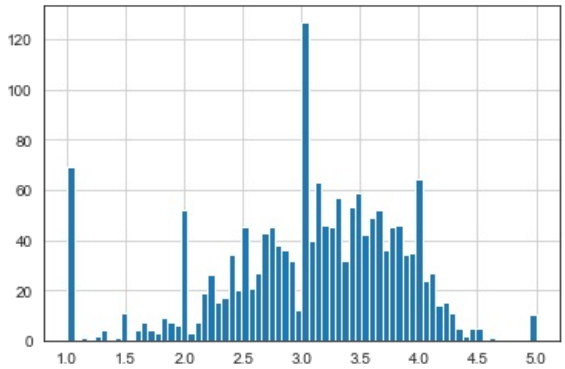


In [31]:

```
ratings['rating'].hist(bins=70)
```

Out[31]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x21c6a5cea48>
```

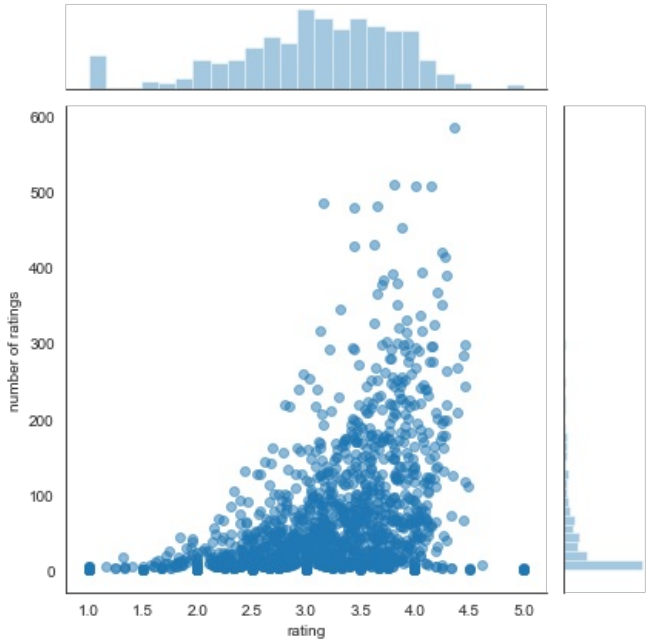


In [32]:

```
sns.jointplot(x='rating',y='number of ratings',data=ratings,alpha=0.5)
```

Out[32]:

```
<seaborn.axisgrid.JointGrid at 0x21c6a6f6848>
```



In [33]:

```
moviemat=df.pivot_table(index='user_id',columns='title',values='rating')
```

In [34]:

```
moviemat.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)	Yo Frankens (1
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 [35]:

```
ratings.sort_values('number of ratings',ascending=False).head(10)
```

Out[35]:

title	rating	number of ratings
Star Wars (1977)	4.359589	584
Contact (1997)	3.803536	509
Fargo (1996)	4.155512	508
Return of the Jedi (1983)	4.007890	507
Liar Liar (1997)	3.156701	485
English Patient, The (1996)	3.656965	481
Scream (1996)	3.441423	478
Toy Story (1995)	3.878319	452
Air Force One (1997)	3.631090	431
Independence Day (ID4) (1996)	3.438228	429

In [36]:

```
starwars_user_ratings=moviemat['Star Wars (1977)']
liarliar_user_ratings=moviemat['Liar Liar (1997)']
```

In [37]:

```
starwars_user_ratings.head(10)
```

Out[37]:

```
user_id
0    5.0
1    5.0
2    5.0
3    NaN
4    5.0
5    4.0
6    4.0
7    5.0
8    5.0
9    5.0
Name: Star Wars (1977), dtype: float64
```

In [39]:

```
similar_to_starwars=moviemat.corrwith(starwars_user_ratings)
```

In [40]:

```
similar_to_liarliar=moviemat.corrwith(liarliar_user_ratings)
```

```
C:\Users\hp\anaconda3\lib\site-packages\numpy\lib\function_base.py:2526: RuntimeWarning: Degrees of
freedom <= 0 for slice
  c = cov(x, y, rowvar)
C:\Users\hp\anaconda3\lib\site-packages\numpy\lib\function_base.py:2455: RuntimeWarning: divide by z
ero encountered in true_divide
  c *= np.true_divide(1, fact)
```

In [42]:

```
corr_starwars=pd.DataFrame(similar_to_starwars,columns=['Correlation'])
corr_starwars.dropna(inplace=True)
```

In [43]:

```
corr_starwars.head()
```

Out[43]:

title	Correlation
'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 [47]:

```
corr_starwars.sort_values('Correlation',ascending=False).head(10)
```

Out[47]:

title	Correlation	number of ratings
Commandments (1997)	1.0	3
Cosi (1996)	1.0	4
No Escape (1994)	1.0	5
Stripes (1981)	1.0	5
Man of the Year (1995)	1.0	9
Hollow Reed (1996)	1.0	6
Beans of Egypt, Maine, The (1994)	1.0	2
Good Man in Africa, A (1994)	1.0	2
Old Lady Who Walked in the Sea, The (Vieille qui marchait dans la mer, La) (1991)	1.0	5
Outlaw, The (1943)	1.0	2

In [51]:

```
corr_starwars[corr_starwars['number of ratings']>100].sort_values('Correlation',ascending=False).head(10)
```

Out[51]:

title	Correlation	number of ratings
Star Wars (1977)	1.000000	584
Empire Strikes Back, The (1980)	0.748353	368
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
Sting, The (1973)	0.367538	241
Indiana Jones and the Last Crusade (1989)	0.350107	331
Pinocchio (1940)	0.347868	101
Frighteners, The (1996)	0.332729	115
L.A. Confidential (1997)	0.319065	297

In [52]:

```
corr_liarliar=pd.DataFrame(similar_to_liarliar,columns=['Correlation'])
```

In [55]:

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

In [56]:

```
corr_liarliar.head(10)
```

Out[56]:

title	Correlation
'Til There Was You (1997)	0.118913
101 Dalmatians (1996)	0.469765
12 Angry Men (1957)	0.066272
187 (1997)	0.175145
2 Days in the Valley (1996)	0.040739
20,000 Leagues Under the Sea (1954)	-0.027932
2001: A Space Odyssey (1968)	-0.057864
39 Steps, The (1935)	0.400918
8 1/2 (1963)	0.178064
A Chef in Love (1996)	0.000000

In [57]:

```
corr_liarliar =corr_liarliar.join(ratings['number of ratings'])
```

In [58]:

```
corr_liarliar.head()
```

Out[58]:

title	Correlation	number of ratings
'Til There Was You (1997)	0.118913	9
101 Dalmatians (1996)	0.469765	109
12 Angry Men (1957)	0.066272	125
187 (1997)	0.175145	41
2 Days in the Valley (1996)	0.040739	93

In [59]:

```
corr_liarliar[corr_liarliar['number of ratings']>100].sort_values('Correlation',ascending=False).head(10)
```

Out[59]:

title	Correlation	number of ratings
Liar Liar (1997)	1.000000	485
Batman Forever (1995)	0.516968	114
Mask, The (1994)	0.484650	129
Down Periscope (1996)	0.472681	101
Con Air (1997)	0.469828	137
Pretty Woman (1990)	0.469790	164
101 Dalmatians (1996)	0.469765	109
Michael (1996)	0.442022	119
Waterworld (1995)	0.438405	102
Indiana Jones and the Last Crusade (1989)	0.414427	331

In [ ]:

```

```