Portfolio - Recommendation System

June 10, 2018

```
In [1]: import numpy as np

In [2]: import pandas as pd

In [3]: columns_names= ['user_id','item_id','rating','timestamp']

In [4]: df = pd.read_csv('u.data',sep='\t',names=columns_names)

In [5]: df.head()
```

Out[5]:

	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 [6]: movie_titles = pd.read_csv('Movie_Id_Titles')

In [7]: movie_titles.head()
```

Out[7]:

	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 [8]: df = pd.merge(df,movie_titles,on='item_id')

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

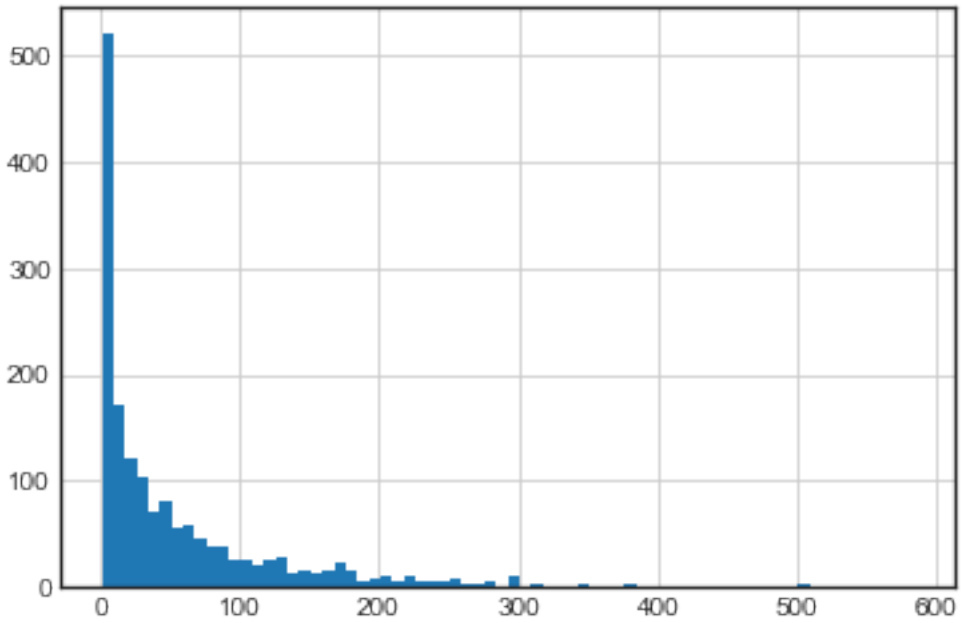
Out[9]:

	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 [10]: import matplotlib.pyplot as plt
         import seaborn as sns

In [11]: sns.set_style('white')

In [12]: %matplotlib inline

In [14]: df.groupby('title')['rating'].mean().sort_values(ascending=False).head()

Out[14]: 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 [15]: df.groupby('title')['rating'].count().sort_values(ascending=False).head()

Out[15]: 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 [16]: ratings = pd.DataFrame(df.groupby('title')['rating'].mean())

In [17]: ratings.head()

Out[17]:                             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 [19]: ratings['num of ratings'] = pd.DataFrame(df.groupby('title')['rating'].count())

In [20]: ratings.head()

Out[20]:                             rating  num of ratings
         title
         '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
```

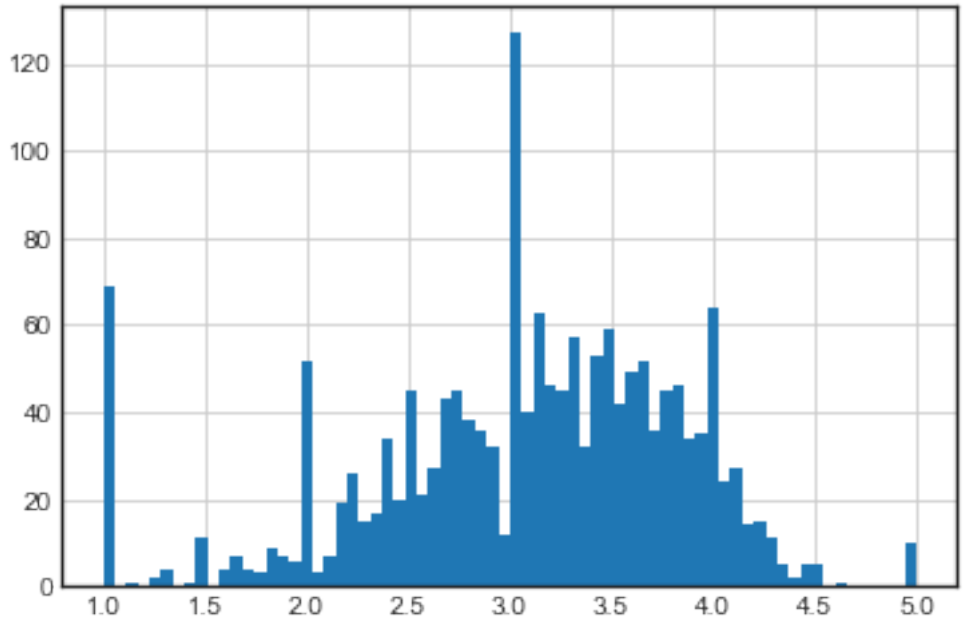
```
In [21]: ratings['num of ratings'].hist(bins=70)
```

```
Out[21]: <matplotlib.axes._subplots.AxesSubplot at 0x1a120db080>
```



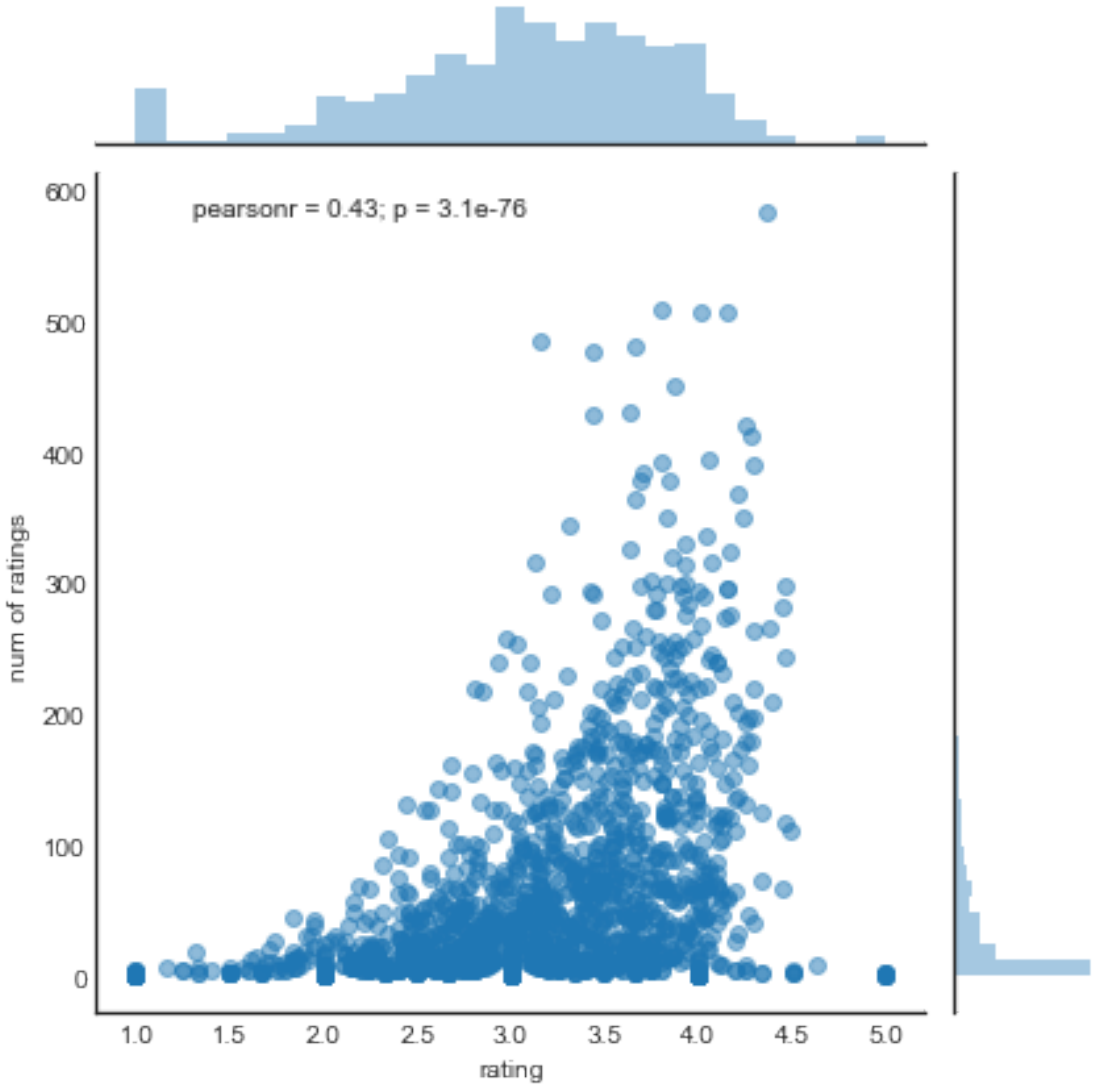

```
In [22]: ratings['rating'].hist(bins=70)
```

```
Out[22]: <matplotlib.axes._subplots.AxesSubplot at 0x1a172664e0>
```



In [24]: sns.jointplot(x='rating',y='num of ratings',data=ratings,alpha=0.5)

Out[24]: <seaborn.axisgrid.JointGrid at 0x1a1a545c18>



In [27]: moviemat = df.pivot_table(index='user_id',columns='title',values='rating')

In [28]: moviemat.head()

Out[28]:
```
title    'Til There Was You (1997)  1-900 (1994)  101 Dalmatians (1996)  \
user_id
0                              NaN           NaN                    NaN
```

```
1                            NaN           NaN                    2.0
2                            NaN           NaN                    NaN
3                            NaN           NaN                    NaN
4                            NaN           NaN                    NaN

title    12 Angry Men (1957)  187 (1997)  2 Days in the Valley (1996)  \
user_id
0                        NaN         NaN                          NaN
1                        5.0         NaN                          NaN
2                        NaN         NaN                          NaN
3                        NaN         2.0                          NaN
4                        NaN         NaN                          NaN

title    20,000 Leagues Under the Sea (1954)  2001: A Space Odyssey (1968)  \
user_id
0                                        NaN                           NaN
1                                        3.0                           4.0
2                                        NaN                           NaN
3                                        NaN                           NaN
4                                        NaN                           NaN

title    3 Ninjas: High Noon At Mega Mountain (1998)  39 Steps, The (1935)  \
user_id
0                                              NaN                   NaN
1                                              NaN                   NaN
2                                              1.0                   NaN
3                                              NaN                   NaN
4                                              NaN                   NaN

title                   ...                  Yankee Zulu (1994)  \
user_id                 ...
0                       ...                                 NaN
1                       ...                                 NaN
2                       ...                                 NaN
3                       ...                                 NaN
4                       ...                                 NaN

title    Year of the Horse (1997)  You So Crazy (1994)  \
user_id
0                            NaN                  NaN
1                            NaN                  NaN
2                            NaN                  NaN
3                            NaN                  NaN
4                            NaN                  NaN

title    Young Frankenstein (1974)  Young Guns (1988)  Young Guns II (1990)  \
user_id
0                              NaN                NaN                   NaN
```

```
1                                5.0                3.0                  NaN
2                                NaN                NaN                  NaN
3                                NaN                NaN                  NaN
4                                NaN                NaN                  NaN

title    Young Poisoner's Handbook, The (1995)  Zeus and Roxanne (1997)  \
user_id
0                                          NaN                      NaN
1                                          NaN                      NaN
2                                          NaN                      NaN
3                                          NaN                      NaN
4                                          NaN                      NaN

title    unknown  Á köldum klaka (Cold Fever) (1994)
user_id
0            NaN                                 NaN
1            4.0                                 NaN
2            NaN                                 NaN
3            NaN                                 NaN
4            NaN                                 NaN

[5 rows x 1664 columns]
```

In [29]:
```
ratings.sort_values('num of ratings',ascending=False).head(10)
```

Out[29]:
```
                                rating  num of ratings
title
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 [39]:
```
starwars_user_ratings = moviemat['Star Wars (1977)']
liarliar_user_ratings = moviemat['Liar Liar (1997)']
```

In [34]:
```
starwars_user_ratings.head()
```

Out[34]:
```
user_id
0    5.0
1    5.0
2    5.0
3    NaN
4    5.0
Name: Star Wars (1977), dtype: float64
```

```
In [36]: similar_to_starwars = moviemat.corrwith(starwars_user_ratings)

/Users/Momin/anaconda3/lib/python3.6/site-packages/numpy/lib/function_base.py:3154: RuntimeWarn
  c = cov(x, y, rowvar)
/Users/Momin/anaconda3/lib/python3.6/site-packages/numpy/lib/function_base.py:3088: RuntimeWarn
  c *= 1. / np.float64(fact)


In [40]: similar_to_liarliar = moviemat.corrwith(liarliar_user_ratings)

/Users/Momin/anaconda3/lib/python3.6/site-packages/numpy/lib/function_base.py:3154: RuntimeWarn
  c = cov(x, y, rowvar)
/Users/Momin/anaconda3/lib/python3.6/site-packages/numpy/lib/function_base.py:3088: RuntimeWarn
  c *= 1. / np.float64(fact)


In [41]: corr_starwars = pd.DataFrame(similar_to_starwars, columns=['Correlation'])
         corr_starwars.dropna(inplace=True)

In [42]: corr_starwars.head()

Out[42]:                          Correlation
         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

In [43]: corr_starwars.sort_values('Correlation',ascending=False).head(10)

Out[43]:                                                    Correlation
         title
         Hollow Reed (1996)                                         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
         Safe Passage (1994)                                        1.0
         Old Lady Who Walked in the Sea, The (Vieille qu...         1.0
         Outlaw, The (1943)                                         1.0
         Line King: Al Hirschfeld, The (1996)                       1.0
         Hurricane Streets (1998)                                   1.0

In [44]: corr_starwars = corr_starwars.join(ratings['num of ratings'])

In [45]: corr_starwars.head()
```

Out[45]:

title	Correlation	num of ratings
'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 [47]:
```
corr_starwars[corr_starwars['num of ratings']>100].sort_values('Correlation',
                                                               ascending=False).head()
```

Out[47]:

title	Correlation	num 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

In [48]:
```
corr_liarliar = pd.DataFrame(similar_to_liarliar,columns=['Correlation'])
```

In [49]:
```
corr_liarliar.dropna(inplace=True)
```

In [50]:
```
corr_liarliar = corr_liarliar.join(ratings['num of ratings'])
```

In [51]:
```
corr_liarliar[corr_liarliar['num of ratings']>100].sort_values('Correlation',
                                                               ascending=False).head()
```

Out[51]:

title	Correlation	num 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