In [26]:
```
import pandas as pd
import numpy as np
```

In [20]:
```
gender = pd.read_csv("C:/Users/Flora Huang/Desktop/Personalization/Assignment/
Homework 2/Data/gender.txt", sep=",", header=None)
gender.columns = ["ID", "Gender"]
```

In [12]:
```
ratings = pd.read_csv("C:/cygwin64/home/Flora Huang/ratings.dat", sep=",",head
er=None)
ratings.columns = ["UserID","ProfileID","Rating"]
```

In [19]:
```
gender.head(5)
```

Out[19]:

	ID	Gender
0	1	F
1	2	F
2	3	U
3	4	F
4	5	F

In [31]:
```
ratings.head(5)
```

Out[31]:

	UserID	ProfileID	Rating	UserGender	ProfileGender
0	1	133	8	F	M
1	1	720	6	F	F
2	1	971	10	F	M
3	1	1095	7	F	M
4	1	1616	10	F	M

In [15]:
```
ratings['UserGender'] = ratings['UserID'].map(gender.set_index('ID')['Gender']
.drop_duplicates())
```

In [30]:
```
ratings['ProfileGender'] = ratings['ProfileID'].map(gender.set_index('ID')['Ge
nder'])
```

In [32]:
```
gender.loc[132]
```

Out[32]:
```
ID        133
Gender      M
Name: 132, dtype: object
```

In [44]:
```
test  = ratings.groupby(['UserID','ProfileGender'])['ProfileGender'].count().to_frame(name = 'count').reset_index()
```

In [146]:
```
test
```

Out[146]:

	UserID	ProfileGender	count	UserGender
0	1	F	24	F
1	1	M	260	F
2	1	U	61	F
3	2	F	7	F
4	2	M	74	F
5	2	U	16	F
6	3	M	18	U
7	3	U	2	U
8	4	F	4	F
9	4	M	74	F
10	4	U	23	F
11	5	F	4	F
12	5	M	79	F
13	5	U	22	F
14	6	F	16	F
15	6	M	57	F
16	6	U	23	F
17	7	F	7	F
18	7	M	48	F
19	7	U	26	F
20	8	F	15	M
21	8	M	4	M
22	8	U	2	M
23	9	F	2458	M
24	9	M	423	M
25	9	U	640	M
26	10	F	10	M
27	10	M	5	M
28	10	U	7	M
29	11	F	13	F
...	...	...	...	...
402574	135350	F	43	F

	UserID	ProfileGender	count	UserGender
402575	135350	M	248	F
402576	135350	U	76	F
402577	135351	F	50	M
402578	135351	M	17	M
402579	135351	U	19	M
402580	135352	F	19	M
402581	135352	M	1	M
402582	135352	U	1	M
402583	135353	F	66	M
402584	135353	M	16	M
402585	135353	U	9	M
402586	135354	F	15	F
402587	135354	M	84	F
402588	135354	U	36	F
402589	135355	F	53	M
402590	135355	M	7	M
402591	135355	U	9	M
402592	135356	F	52	M
402593	135356	M	5	M
402594	135356	U	19	M
402595	135357	F	24	F
402596	135357	M	246	F
402597	135357	U	66	F
402598	135358	F	33	M
402599	135358	M	7	M
402600	135358	U	24	M
402601	135359	F	18	F
402602	135359	M	179	F
402603	135359	U	49	F

402604 rows × 4 columns

In [46]:
```
test['UserGender'] = test['UserID'].map(gender.set_index('ID')['Gender'])
```

In [127]:
```
female = test[(test['UserGender']=="F")]
male = test[(test['UserGender']=="M")]
```

In [129]:
```
summary_ffy = female[((female['ProfileGender']=="F")&(female['count']!=0))]
summary_mfy = male[((male['ProfileGender']=="F")&(male['count']!=0))]
```

In [130]:
```
summary_fmy = female[((female['ProfileGender']=="M")&(female['count']!=0))]
summary_mmy = male[((male['ProfileGender']=="M")&(male['count']!=0))]
```

In [131]:
```
summary_ffn = female[((female['ProfileGender']=="F")&(female['count']==0))]
summary_mfn = male[((male['ProfileGender']=="F")&(male['count']==0))]
```

In [132]:
```
summary_fmn = female[((female['ProfileGender']=="M")&(female['count']==0))]
summary_mmn = male[((male['ProfileGender']=="M")&(male['count']==0))]
```

In [133]:
```
straight_f = len(np.intersect1d(summary_ffn.UserID,summary_fmy.UserID))
straight_f
```

Out[133]: 0

In [134]:
```
bi_f = len(np.intersect1d(summary_ffy.UserID,summary_fmy.UserID))
bi_f
```

Out[134]: 59258

In [135]:
```
les_f = len(np.intersect1d(summary_ffy.UserID,summary_fmn.UserID))
les_f
```

Out[135]: 0

In [136]:
```
straight_m = len(np.intersect1d(summary_mfy.UserID,summary_mmn.UserID))
straight_m
```

Out[136]: 0

In [137]:
```
bi_m = len(np.intersect1d(summary_mfy.UserID,summary_mmy.UserID))
bi_m
```

Out[137]: 58498

In [138]:
```
gay_f = len(np.intersect1d(summary_ffn.UserID,summary_fmy.UserID))
gay_f
```

Out[138]: 0

In [145]:
```
female[((female['ProfileGender']=="F")&(female['count']==0))].shape[0]
```

Out[145]: 0