

# facebook-data-easy-eda

October 30, 2022

## 1 1.Importing Libraries

```
[1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from pandas_profiling import ProfileReport
%matplotlib inline
```

## 2 2.Data Acquisition

```
[2]: fbdata= pd.read_csv("https://raw.githubusercontent.com/insaid2018/Term-1/master/
↳Data/Projects/facebook_data.csv")
fbdata.head()
```

```
[2]:
```

	userid	age	dob_day	dob_year	dob_month	gender	tenure	friend_count	\
0	2094382	14	19	1999	11	male	266.0	0	
1	1192601	14	2	1999	11	female	6.0	0	
2	2083884	14	16	1999	11	male	13.0	0	
3	1203168	14	25	1999	12	female	93.0	0	
4	1733186	14	4	1999	12	male	82.0	0	

	friendships_initiated	likes	likes_received	mobile_likes	\
0	0	0	0	0	
1	0	0	0	0	
2	0	0	0	0	
3	0	0	0	0	
4	0	0	0	0	

	mobile_likes_received	www_likes	www_likes_received
0	0	0	0
1	0	0	0
2	0	0	0
3	0	0	0
4	0	0	0

### 3 3.Describing the Data

```
[3]: fbdata.describe()
```

```
[3]:
```

	userid	age	dob_day	dob_year	dob_month \
count	9.900300e+04	99003.000000	99003.000000	99003.000000	99003.000000
mean	1.597045e+06	37.280224	14.530408	1975.719776	6.283365
std	3.440592e+05	22.589748	9.015606	22.589748	3.529672
min	1.000008e+06	13.000000	1.000000	1900.000000	1.000000
25%	1.298806e+06	20.000000	7.000000	1963.000000	3.000000
50%	1.596148e+06	28.000000	14.000000	1985.000000	6.000000
75%	1.895744e+06	50.000000	22.000000	1993.000000	9.000000
max	2.193542e+06	113.000000	31.000000	2000.000000	12.000000

	tenure	friend_count	friendships_initiated	likes \
count	99001.000000	99003.000000	99003.000000	99003.000000
mean	537.887375	196.350787	107.452471	156.078785
std	457.649874	387.304229	188.786951	572.280681
min	0.000000	0.000000	0.000000	0.000000
25%	226.000000	31.000000	17.000000	1.000000
50%	412.000000	82.000000	46.000000	11.000000
75%	675.000000	206.000000	117.000000	81.000000
max	3139.000000	4923.000000	4144.000000	25111.000000

	likes_received	mobile_likes	mobile_likes_received	www_likes \
count	99003.000000	99003.000000	99003.000000	99003.000000
mean	142.689363	106.116300	84.120491	49.962425
std	1387.919613	445.252985	839.889444	285.560152
min	0.000000	0.000000	0.000000	0.000000
25%	1.000000	0.000000	0.000000	0.000000
50%	8.000000	4.000000	4.000000	0.000000
75%	59.000000	46.000000	33.000000	7.000000
max	261197.000000	25111.000000	138561.000000	14865.000000

	www_likes_received
count	99003.000000
mean	58.568831
std	601.416348
min	0.000000
25%	0.000000
50%	2.000000
75%	20.000000
max	129953.000000

Age: Minimum-13, Maximum-113, Average age-37 DOB Year: Minimum-1900, Maximum-2000, Average age-37 Tenure: Minimum-0, Maximum-3139, Average age-537 Likes done are lower than likes received.

## 4 4.Extracting Info about the Data

```
[4]: fbdata.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 99003 entries, 0 to 99002
Data columns (total 15 columns):
#   Column                Non-Null Count  Dtype
---  -
0   userid                99003 non-null  int64
1   age                   99003 non-null  int64
2   dob_day               99003 non-null  int64
3   dob_year              99003 non-null  int64
4   dob_month             99003 non-null  int64
5   gender                98828 non-null  object
6   tenure                99001 non-null  float64
7   friend_count          99003 non-null  int64
8   friendships_initiated 99003 non-null  int64
9   likes                 99003 non-null  int64
10  likes_received         99003 non-null  int64
11  mobile_likes           99003 non-null  int64
12  mobile_likes_received 99003 non-null  int64
13  www_likes              99003 non-null  int64
14  www_likes_received    99003 non-null  int64
dtypes: float64(1), int64(13), object(1)
memory usage: 11.3+ MB
```

## 5 5.Data Pre - Profilling

```
[5]: profile = fbdata.profile_report(title="Pandas Pre-Profilling Report")
profile.to_file(output_file="pandas_pre_profilling.html")
```

```
Summarize dataset: 0%|          | 0/5 [00:00<?, ?it/s]
Generate report structure: 0%|          | 0/1 [00:00<?, ?it/s]
Render HTML: 0%|          | 0/1 [00:00<?, ?it/s]
Export report to file: 0%|          | 0/1 [00:00<?, ?it/s]
```

## 6 6.Data preprocessing

```
[6]: fbdata.isnull().sum()
```

```
[6]: userid                0
age                      0
dob_day                  0
dob_year                 0
```

```

dob_month          0
gender             175
tenure             2
friend_count       0
friendships_initiated 0
likes              0
likes_received     0
mobile_likes       0
mobile_likes_received 0
www_likes          0
www_likes_received 0
dtype: int64

```

We have 2 columns with missing values

### 6.0.1 Dropping the missing values in tenure

```

[7]: fbdata = fbdata[~fbdata.tenure.isnull()].copy()

fbdata.isnull().sum()

```

```

[7]: userid          0
age                0
dob_day           0
dob_year          0
dob_month         0
gender            175
tenure            0
friend_count       0
friendships_initiated 0
likes              0
likes_received     0
mobile_likes       0
mobile_likes_received 0
www_likes          0
www_likes_received 0
dtype: int64

```

### 6.0.2 Gender ratio between male and female users

```

[8]: x = fbdata['gender']=='male'
males = x.value_counts()[True]
non_male = x.value_counts()[False]

y = fbdata['gender']=='female'
females = y.value_counts()[True]

nan_count = non_male - females

```

```

total = males + non_male
print('males :',males)
print('females:',females)
print('nan count:', nan_count)
print('total:',total)
print('ratio - male:female :',males/females)

```

```

males : 58574
females: 40252
nan count: 175
total: 99001
ratio - male:female : 1.4551823511875186

```

### 6.0.3 Filling the nan values

```
[9]: fbdata.gender.fillna(value="not revealed",inplace=True)
```

```
[10]: fbdata.isnull().sum()
```

```

[10]: userid                0
      age                   0
      dob_day               0
      dob_year              0
      dob_month             0
      gender                0
      tenure                0
      friend_count          0
      friendships_initiated 0
      likes                 0
      likes_received         0
      mobile_likes          0
      mobile_likes_received 0
      www_likes             0
      www_likes_received    0
      dtype: int64

```

```

[11]: x = fbdata['gender']=='male'
      males = x.value_counts()[True]

      y = fbdata['gender']=='female'
      females = y.value_counts()[True]

      z = fbdata['gender']=='not revealed'
      not_revealed = z.value_counts()[True]

      total = males + females + not_revealed

      print('males :',males)

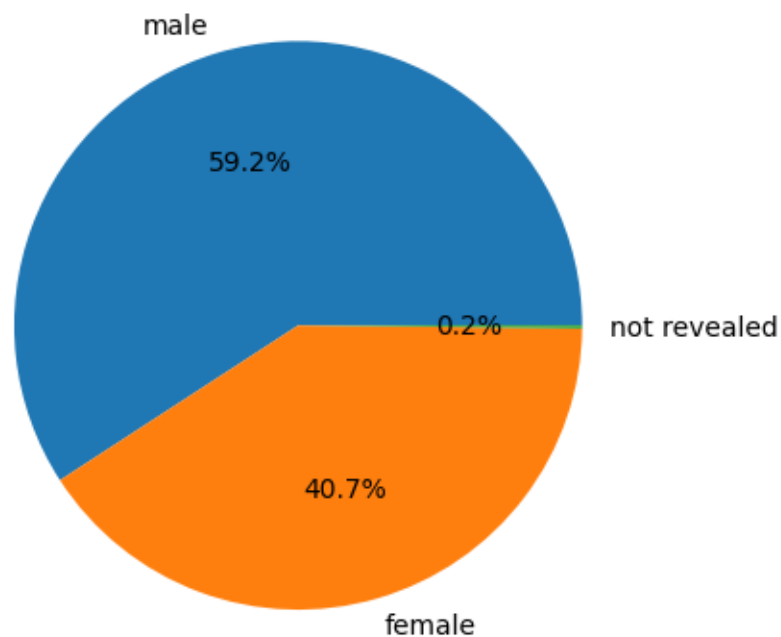
```

```
print('females:',females)
print('not_revealed:', not_revealed)
print('total:',total)
```

```
males : 58574
females: 40252
not_revealed: 175
total: 99001
```

## 6.1 Gender distribution through pie chart

```
[12]: plt.pie(fbdata.gender.value_counts(),labels=['male','female','not_
        revealed'],explode=(0,0,0),autopct='%1.1f%%')
plt.show()
```



## 7 7.Data post profiling

```
[13]: profile = fbdata.profile_report(title="Pandas Post-Profilling Report")
profile.to_file(output_file="pandas_post_profilling.html")
```

```
Summarize dataset: 0%|          | 0/5 [00:00<?, ?it/s]
```

```
Generate report structure: 0%|          | 0/1 [00:00<?, ?it/s]
```

Render HTML: 0% | 0/1 [00:00<?, ?it/s]

Export report to file: 0% | 0/1 [00:00<?, ?it/s]

## 8 Exploratory Data Analysis

### 8.1 Evaluating Distribution of age

```
[14]: sns.distplot(fbdata['age'])
```

C:\Users\hp\AppData\Local\Temp\ipykernel\_572\862979134.py:1: UserWarning:

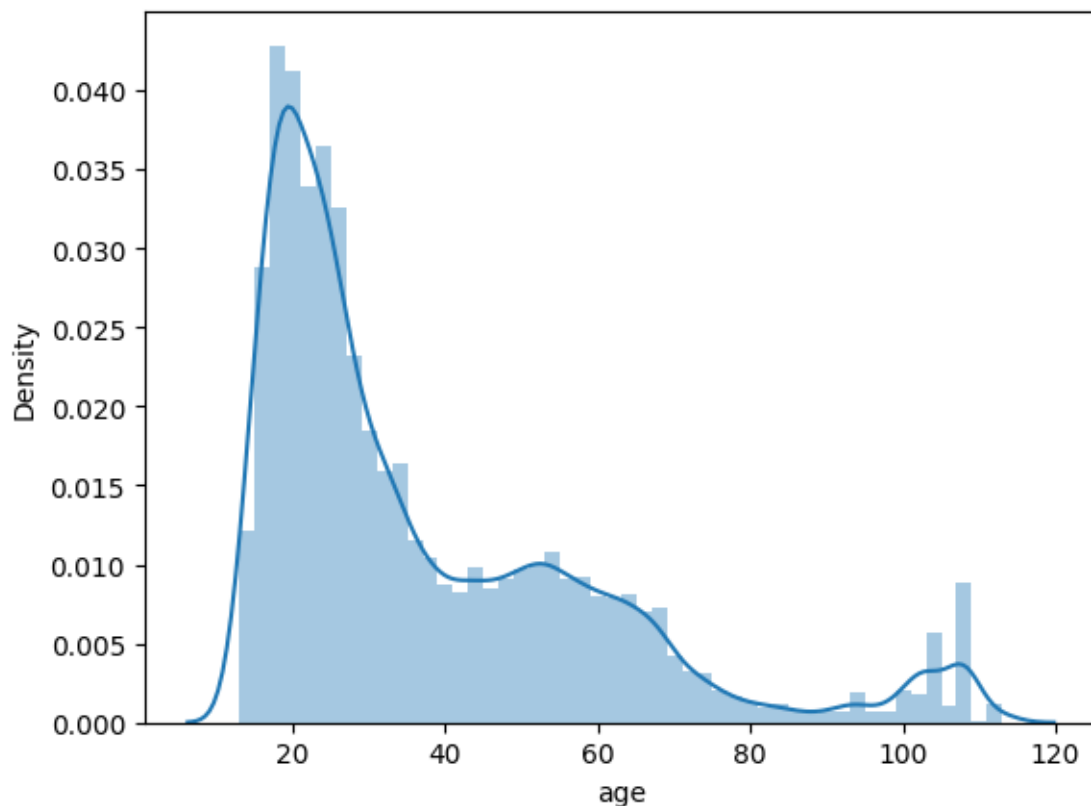
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see <https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

```
sns.distplot(fbdata['age'])
```

```
[14]: <AxesSubplot:xlabel='age', ylabel='Density'>
```



Majority of the People are of age near 20

## 8.2 Day, Month and Year wise distribution of DOB

```
[15]: sns.distplot(fbdata['dob_day'])
```

C:\Users\hp\AppData\Local\Temp\ipykernel\_572\1986788402.py:1: UserWarning:

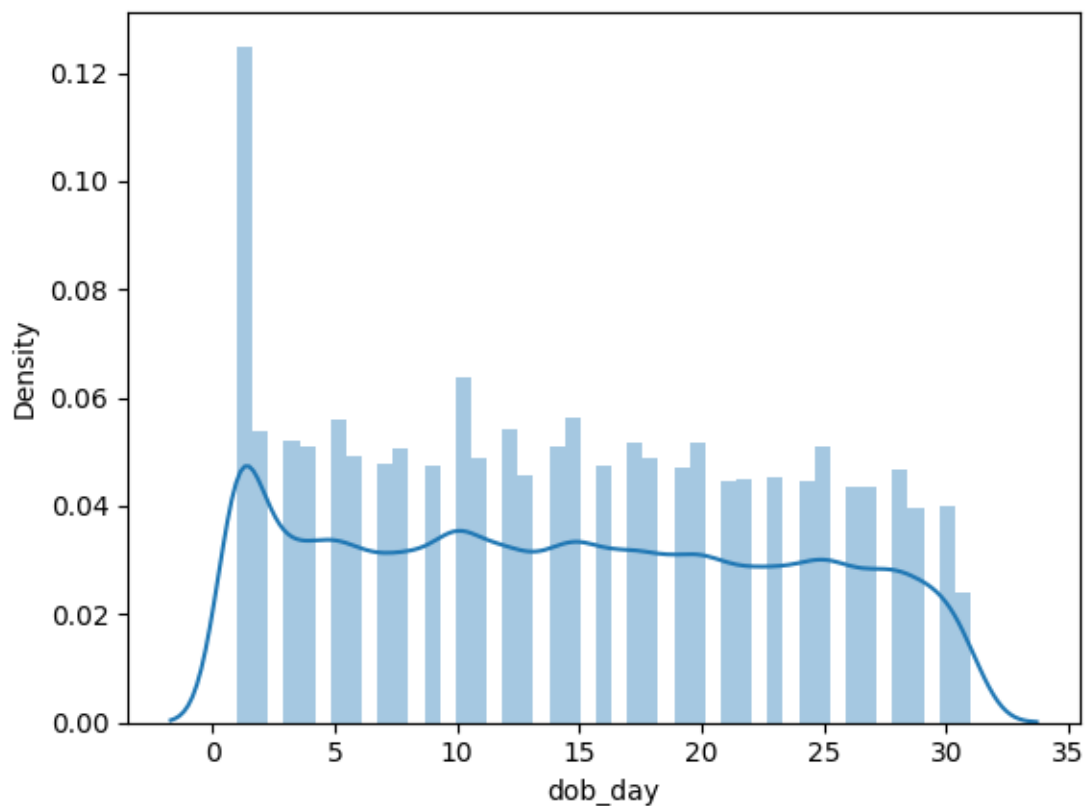
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see <https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

```
sns.distplot(fbdata['dob_day'])
```

```
[15]: <AxesSubplot:xlabel='dob_day', ylabel='Density'>
```



Day of birth is evenly spread across all the days with 1 having larger density



```
[16]: sns.distplot(fbdata['dob_month'])
```

C:\Users\hp\AppData\Local\Temp\ipykernel\_572\177632322.py:1: UserWarning:

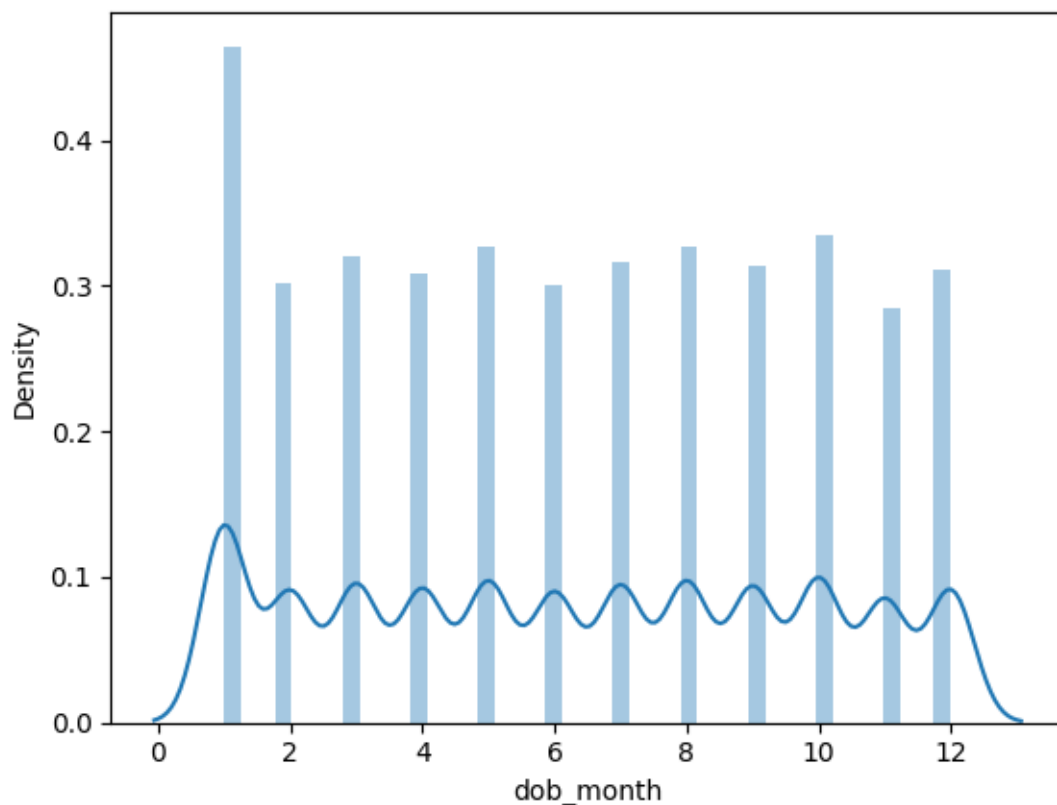
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see <https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

```
sns.distplot(fbdata['dob_month'])
```

```
[16]: <AxesSubplot:xlabel='dob_month', ylabel='Density'>
```



Month of born is evenly spread across all the months

```
[17]: sns.distplot(fbdata['dob_year'])
```

C:\Users\hp\AppData\Local\Temp\ipykernel\_572\933665262.py:1: UserWarning:

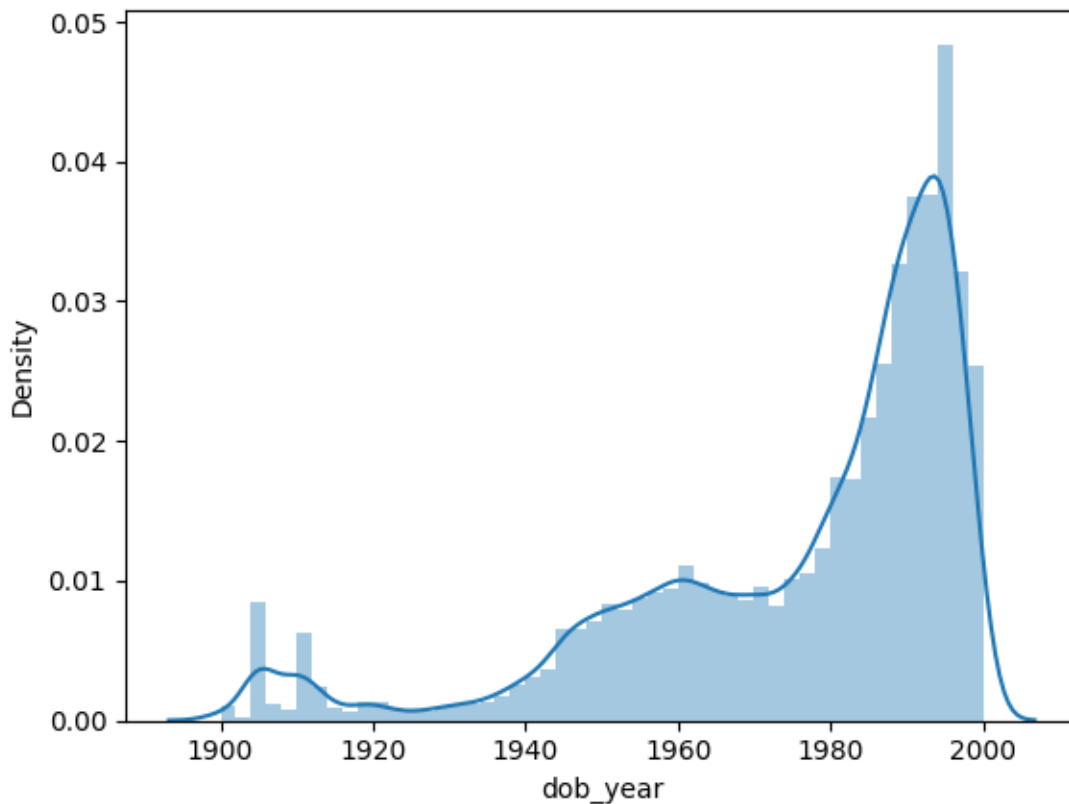
``distplot`` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either ``displot`` (a figure-level function with similar flexibility) or ``histplot`` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see <https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

```
sns.distplot(fbdata['dob_year'])
```

```
[17]: <AxesSubplot:xlabel='dob_year', ylabel='Density'>
```



Majority of the users are born after 1980

### 8.3 Dividing the dob\_year in groups of 10

```
[18]: labels=['1900-1910', '1910-1920', '1920-1930', '1930-1940', '1940-1950', '1950-1960', '1960-1970', '1970-1980', '1980-1990', '1990-2000']
fbdata['year_group'] = pd.cut(fbdata.dob_year, bins=np.
    ↳ arange(1900, 2001, 10), labels=labels, right=True)
fbdata.tail()
```

```
[18]:      userid  age  dob_day  dob_year  dob_month  gender  tenure  \
98998  1268299   68        4    1945         4  female   541.0
98999  1256153   18        12   1995         3  female    21.0
99000  1195943   15        10   1998         5  female   111.0
99001  1468023   23        11   1990         4  female   416.0
99002  1397896   39        15   1974         5  female   397.0

      friend_count  friendships_initiated  likes  likes_received  \
98998          2118                341   3996        18089
98999          1968                1720   4401        13412
99000          2002                1524  11959        12554
99001          2560                185   4506         6516
99002          2049                768   9410        12443

      mobile_likes  mobile_likes_received  www_likes  www_likes_received  \
98998          3505                11887        491        6202
98999          4399                10592         2        2820
99000          11959                11462         0        1092
99001          4506                5760         0         756
99002          9410                9530         0        2913

      year_group
98998  1940-1950
98999  1990-2000
99000  1990-2000
99001  1980-1990
99002  1970-1980
```

see last column for year group

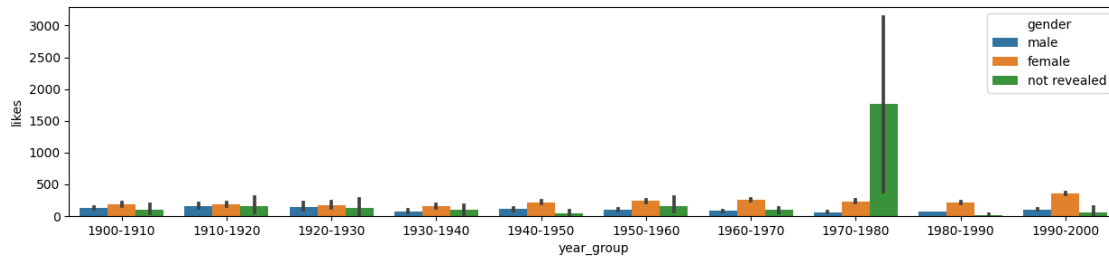
```
[19]: #Counting value in year groups
fbdata.year_group.value_counts()
```

```
[19]: 1990-2000    31455
1980-1990    25080
1970-1980    10990
1960-1970     9298
1950-1960     8921
1940-1950     5935
1900-1910     3140
1930-1940     1787
1910-1920     1435
1920-1930       758
Name: year_group, dtype: int64
```

### 8.3.1 Number of likes comparison between males and females based on year born range

```
[20]: plt.figure(figsize=(15,3))
sns.barplot(x=fbddata['year_group'],y=fbddata['likes'],hue=fbddata.gender)
```

```
[20]: <AxesSubplot:xlabel='year_group', ylabel='likes'>
```



Number of likes are greater for females as compared to males across all the year groups

### 8.4 Dividing the likes in groups of 1000

```
[21]: labels=['0-1000','1001-2000','2001-3000','3001-4000','4001-5000','5001-6000','6001-7000','7001-8000']
fbddata['likes_range'] = pd.cut(fbddata.likes,bins=np.
    ↳ arange(0,26001,1000),labels=labels,right=True)
fbddata.tail()
```

```
[21]:
```

	userid	age	dob_day	dob_year	dob_month	gender	tenure	\
98998	1268299	68	4	1945	4	female	541.0	
98999	1256153	18	12	1995	3	female	21.0	
99000	1195943	15	10	1998	5	female	111.0	
99001	1468023	23	11	1990	4	female	416.0	
99002	1397896	39	15	1974	5	female	397.0	

	friend_count	friendships_initiated	likes	likes_received	\
98998	2118	341	3996	18089	
98999	1968	1720	4401	13412	
99000	2002	1524	11959	12554	
99001	2560	185	4506	6516	
99002	2049	768	9410	12443	

	mobile_likes	mobile_likes_received	www_likes	www_likes_received	\
98998	3505	11887	491	6202	
98999	4399	10592	2	2820	
99000	11959	11462	0	1092	
99001	4506	5760	0	756	
99002	9410	9530	0	2913	

	year_group	likes_range
98998	1940-1950	3001-4000
98999	1990-2000	4001-5000
99000	1990-2000	11001-12000
99001	1980-1990	4001-5000
99002	1970-1980	9001-10000

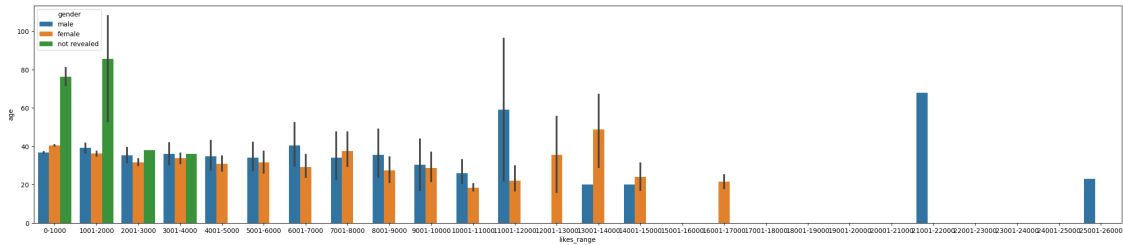
```
[22]: #Counting value in likes range
fbdata.likes_range.value_counts()
```

```
[22]: 0-1000          73278
      1001-2000      2059
      2001-3000       668
      3001-4000       293
      4001-5000       130
      5001-6000        82
      6001-7000        57
      7001-8000        34
      8001-9000        33
      9001-10000       16
     10001-11000       14
     11001-12000       12
     12001-13000        5
     14001-15000        4
     13001-14000        4
     16001-17000        2
     21001-22000        1
     25001-26000        1
     15001-16000        0
     17001-18000        0
     18001-19000        0
     19001-20000        0
     20001-21000        0
     22001-23000        0
     23001-24000        0
     24001-25000        0
      Name: likes_range, dtype: int64
```

#### 8.4.1 age comparison between males and females based on likes range

```
[23]: plt.figure(figsize=(30,6))
      sns.barplot(x=fbdata['likes_range'],y=fbdata['age'],hue=fbdata.gender)
```

```
[23]: <AxesSubplot:xlabel='likes_range', ylabel='age'>
```



Some males have very large number of likes

## 8.5 Counting people with some non - zero attributes

```
[24]: #No of people having non-zero likes
np.count_nonzero(fbdata.likes)
```

[24]: 76693

```
[25]: #No of people having non-zero tenure
np.count_nonzero(fbdata.tenure)
```

[25]: 98931

```
[26]: #No of people having non-zero friends
np.count_nonzero(fbdata.friend_count)
```

[26]: 97039

## 8.6 Evaluation on users with Zero friends

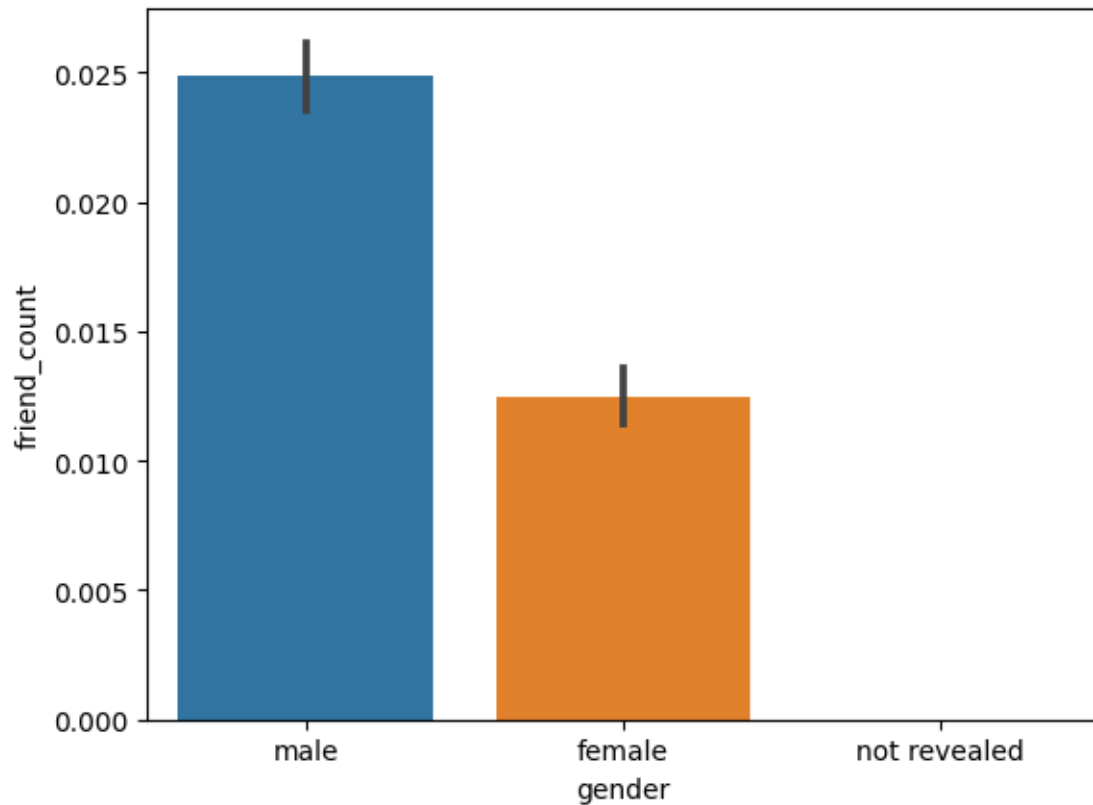
```
[27]: #All the people having zero friends
fc=fbdata.friend_count==0
fc.value_counts()
```

```
[27]: False    97039
      True     1962
      Name: friend_count, dtype: int64
```

**This infers that Males have more zero friends than females**

```
[28]: #plotting the gender vs zero friend count people
sns.barplot(y=fbdata.friend_count==0,x=fbdata.gender)
```

```
[28]: <AxesSubplot:xlabel='gender', ylabel='friend_count'>
```



## 8.7 Dividing tenure to year ranges

```
[29]: fbdata.tenure.interpolate(inplace=True)
tenlabel=['0-1 years','1-2 years','2-3 years','3-4 years','4-5 years','5-6_
↪years','6-7 years','7-8 years','8-9 years']
fbdata['tenure_range']=pd.cut(fbdata.tenure,bins=np.
↪arange(0,3300,365),labels=tenlabel,right=True)
```

```
[30]: fbdata.tail()
```

```
[30]:
```

	userid	age	dob_day	dob_year	dob_month	gender	tenure	\
98998	1268299	68	4	1945	4	female	541.0	
98999	1256153	18	12	1995	3	female	21.0	
99000	1195943	15	10	1998	5	female	111.0	
99001	1468023	23	11	1990	4	female	416.0	
99002	1397896	39	15	1974	5	female	397.0	

	friend_count	friendships_initiated	likes	likes_received	\
98998	2118	341	3996	18089	
98999	1968	1720	4401	13412	
99000	2002	1524	11959	12554	

99001	2560	185	4506	6516
99002	2049	768	9410	12443

	mobile_likes	mobile_likes_received	www_likes	www_likes_received	\
98998	3505	11887	491	6202	
98999	4399	10592	2	2820	
99000	11959	11462	0	1092	
99001	4506	5760	0	756	
99002	9410	9530	0	2913	

	year_group	likes_range	tenure_range
98998	1940-1950	3001-4000	1-2 years
98999	1990-2000	4001-5000	0-1 years
99000	1990-2000	11001-12000	0-1 years
99001	1980-1990	4001-5000	1-2 years
99002	1970-1980	9001-10000	1-2 years

```
[31]: fbdata.tenure_range.value_counts()
```

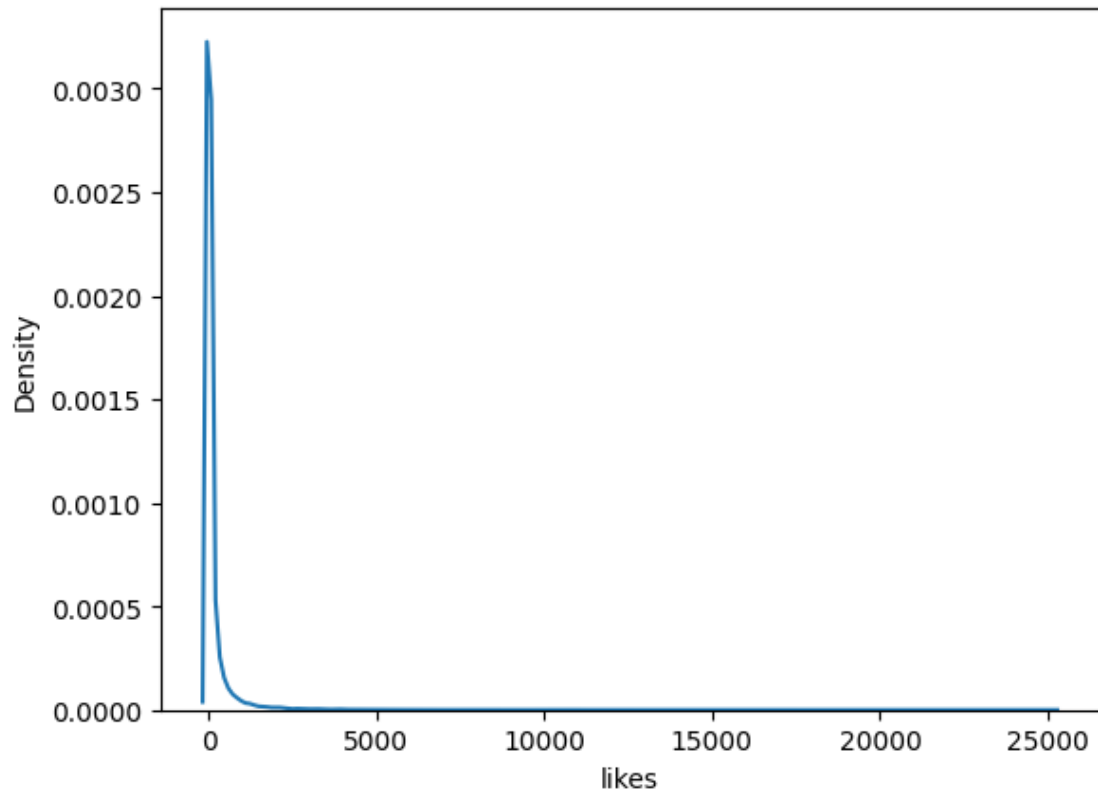
```
[31]: 0-1 years      43588
      1-2 years      33366
      2-3 years       9860
      3-4 years      5448
      4-5 years      4557
      5-6 years      1507
      6-7 years       581
      7-8 years        15
      8-9 years         9
      Name: tenure_range, dtype: int64
```

## 8.8 Evaluating LIKES

```
[32]: sns.kdeplot(fbdata['likes'])
```

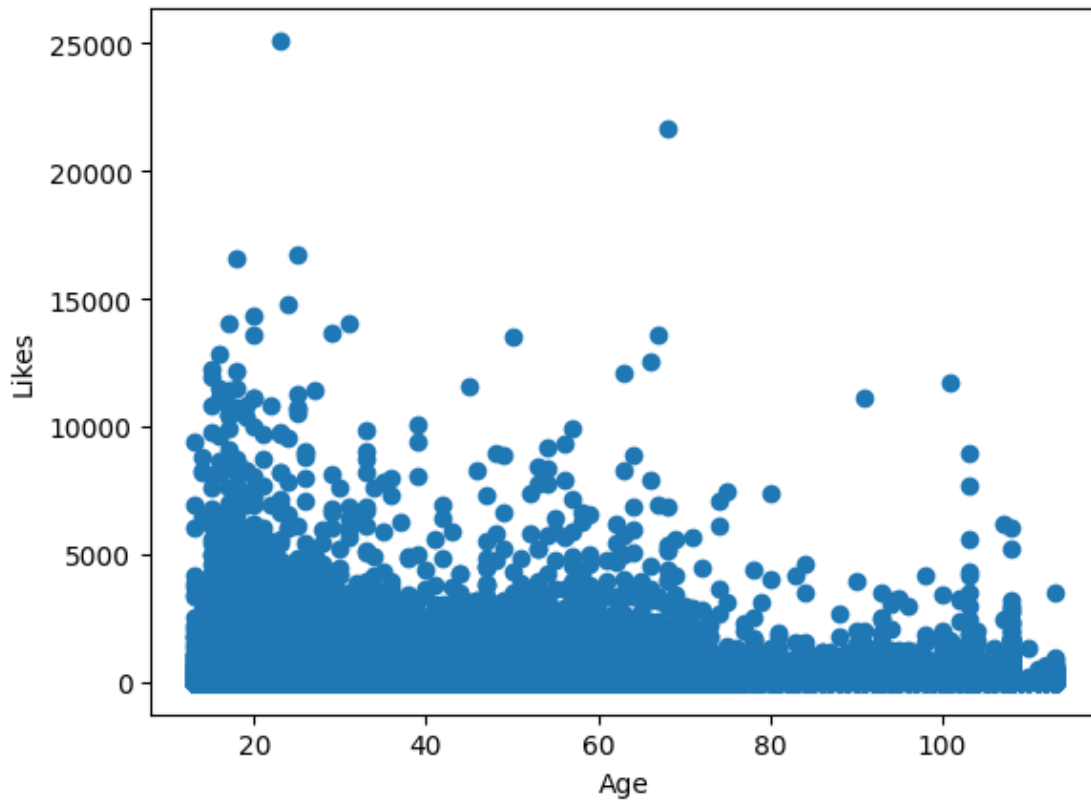
```
[32]: <AxesSubplot:xlabel='likes', ylabel='Density'>
```





### 8.8.1 Likes vs age

```
[33]: plt.scatter(fbdata.age,fbdata.likes)
plt.xlabel("Age")
plt.ylabel("Likes")
plt.show()
```



As age increases, the number of likes decreases

### 8.8.2 Most liked people

```
[34]: fbdata.sort_values(by='likes_received',ascending=False)[:10]
```

```
[34]:
```

	userid	age	dob_day	dob_year	dob_month	gender	tenure	\
94906	1674584	17	14	1996	8	female	401.0	
77121	1441676	20	5	1993	8	female	253.0	
98822	1715925	23	4	1990	9	female	705.0	
98994	2063006	20	4	1993	1	female	402.0	
98878	1053087	23	6	1990	6	male	596.0	
49230	1432020	20	12	1993	1	male	245.0	
98773	2042824	18	25	1995	1	male	51.0	
98937	1559908	20	4	1993	12	female	1334.0	
98936	1781243	17	1	1996	5	female	976.0	
98686	1015907	74	27	1939	11	male	832.0	

	friend_count	friendships_initiated	likes	likes_received	\
94906	818	395	1016	261197	
77121	230	73	2078	178166	

98822	4077	793	1877	152014
98994	1988	332	7351	106025
98878	4320	836	2996	82623
49230	79	50	477	53534
98773	4817	32	1346	52964
98937	4622	1819	4280	45633
98936	3683	755	10478	42449
98686	4630	831	966	39536

	mobile_likes	mobile_likes_received	www_likes	www_likes_received	\
94906	659	131244	357	129953	
77121	1982	138561	96	39605	
98822	80	89911	1797	62103	
98994	7248	73333	103	32692	
98878	179	43410	2817	39213	
49230	78	30387	399	23147	
98773	1342	18925	4	34039	
98937	472	30754	3808	14879	
98936	246	27353	10232	15096	
98686	941	10199	25	29337	

	year_group	likes_range	tenure_range
94906	1990-2000	1001-2000	1-2 years
77121	1990-2000	2001-3000	0-1 years
98822	1980-1990	1001-2000	1-2 years
98994	1990-2000	7001-8000	1-2 years
98878	1980-1990	2001-3000	1-2 years
49230	1990-2000	0-1000	0-1 years
98773	1990-2000	1001-2000	0-1 years
98937	1990-2000	4001-5000	3-4 years
98936	1990-2000	10001-11000	2-3 years
98686	1930-1940	0-1000	2-3 years

### 8.8.3 Calculating likes per day

```
[35]: fbdata['likes_per_day']=fbdata.likes_received/fbdata.tenure.where(fbdata.
      ↳tenure>0)
fbdata.tail()
```

```
[35]:
```

	userid	age	dob_day	dob_year	dob_month	gender	tenure	\
98998	1268299	68	4	1945	4	female	541.0	
98999	1256153	18	12	1995	3	female	21.0	
99000	1195943	15	10	1998	5	female	111.0	
99001	1468023	23	11	1990	4	female	416.0	
99002	1397896	39	15	1974	5	female	397.0	

	friend_count	friendships_initiated	likes	likes_received	\
--	--------------	-----------------------	-------	----------------	---

98998	2118	341	3996	18089
98999	1968	1720	4401	13412
99000	2002	1524	11959	12554
99001	2560	185	4506	6516
99002	2049	768	9410	12443

	mobile_likes	mobile_likes_received	www_likes	www_likes_received	\
98998	3505	11887	491	6202	
98999	4399	10592	2	2820	
99000	11959	11462	0	1092	
99001	4506	5760	0	756	
99002	9410	9530	0	2913	

	year_group	likes_range	tenure_range	likes_per_day
98998	1940-1950	3001-4000	1-2 years	33.436229
98999	1990-2000	4001-5000	0-1 years	638.666667
99000	1990-2000	11001-12000	0-1 years	113.099099
99001	1980-1990	4001-5000	1-2 years	15.663462
99002	1970-1980	9001-10000	1-2 years	31.342569

#### 8.8.4 Users with most likes per day

```
[36]: fbdata.sort_values(by='likes_per_day',ascending=False)[:10]
```

```
[36]:
```

	userid	age	dob_day	dob_year	dob_month	gender	tenure	\
94996	1149300	24	7	1989	3	female	2.0	
94057	2175941	18	25	1995	2	male	2.0	
98773	2042824	18	25	1995	1	male	51.0	
77121	1441676	20	5	1993	8	female	253.0	
94906	1674584	17	14	1996	8	female	401.0	
98999	1256153	18	12	1995	3	female	21.0	
61145	1409983	47	8	1966	10	female	4.0	
98994	2063006	20	4	1993	1	female	402.0	
76025	1494406	17	1	1996	1	female	2.0	
75126	1546811	39	11	1974	12	female	29.0	

	friend_count	friendships_initiated	likes	likes_received	\
94996	834	104	2169	5404	
94057	640	299	5640	2542	
98773	4817	32	1346	52964	
77121	230	73	2078	178166	
94906	818	395	1016	261197	
98999	1968	1720	4401	13412	
61145	123	61	3331	1200	
98994	1988	332	7351	106025	
76025	215	185	292	478	
75126	212	145	8091	6730	

	mobile_likes	mobile_likes_received	www_likes	www_likes_received	\
94996	2169	1520	0	3884	
94057	4251	1587	1389	955	
98773	1342	18925	4	34039	
77121	1982	138561	96	39605	
94906	659	131244	357	129953	
98999	4399	10592	2	2820	
61145	3331	468	0	732	
98994	7248	73333	103	32692	
76025	292	244	0	234	
75126	3181	2457	4910	4273	

	year_group	likes_range	tenure_range	likes_per_day
94996	1980-1990	2001-3000	0-1 years	2702.000000
94057	1990-2000	5001-6000	0-1 years	1271.000000
98773	1990-2000	1001-2000	0-1 years	1038.509804
77121	1990-2000	2001-3000	0-1 years	704.213439
94906	1990-2000	1001-2000	1-2 years	651.364090
98999	1990-2000	4001-5000	0-1 years	638.666667
61145	1960-1970	3001-4000	0-1 years	300.000000
98994	1990-2000	7001-8000	1-2 years	263.743781
76025	1990-2000	0-1000	0-1 years	239.000000
75126	1970-1980	8001-9000	0-1 years	232.068966

### 8.8.5 Extracting most famous people

```
[37]: famous=fbdata.sort_values(by='likes_per_day',ascending=False)[:10]
      famous.head()
```

```
[37]:
```

	userid	age	dob_day	dob_year	dob_month	gender	tenure	\
94996	1149300	24	7	1989	3	female	2.0	
94057	2175941	18	25	1995	2	male	2.0	
98773	2042824	18	25	1995	1	male	51.0	
77121	1441676	20	5	1993	8	female	253.0	
94906	1674584	17	14	1996	8	female	401.0	

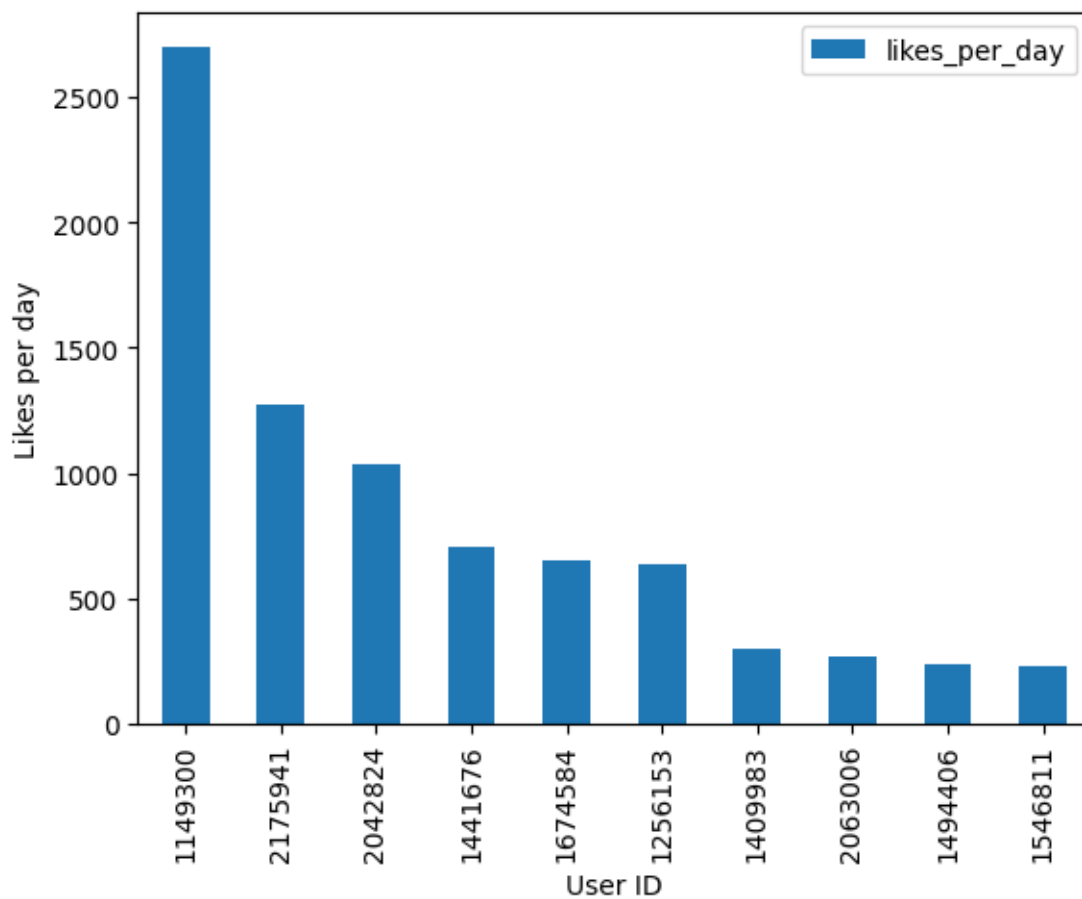
	friend_count	friendships_initiated	likes	likes_received	\
94996	834	104	2169	5404	
94057	640	299	5640	2542	
98773	4817	32	1346	52964	
77121	230	73	2078	178166	
94906	818	395	1016	261197	

	mobile_likes	mobile_likes_received	www_likes	www_likes_received	\
94996	2169	1520	0	3884	
94057	4251	1587	1389	955	

98773	1342	18925	4	34039
77121	1982	138561	96	39605
94906	659	131244	357	129953

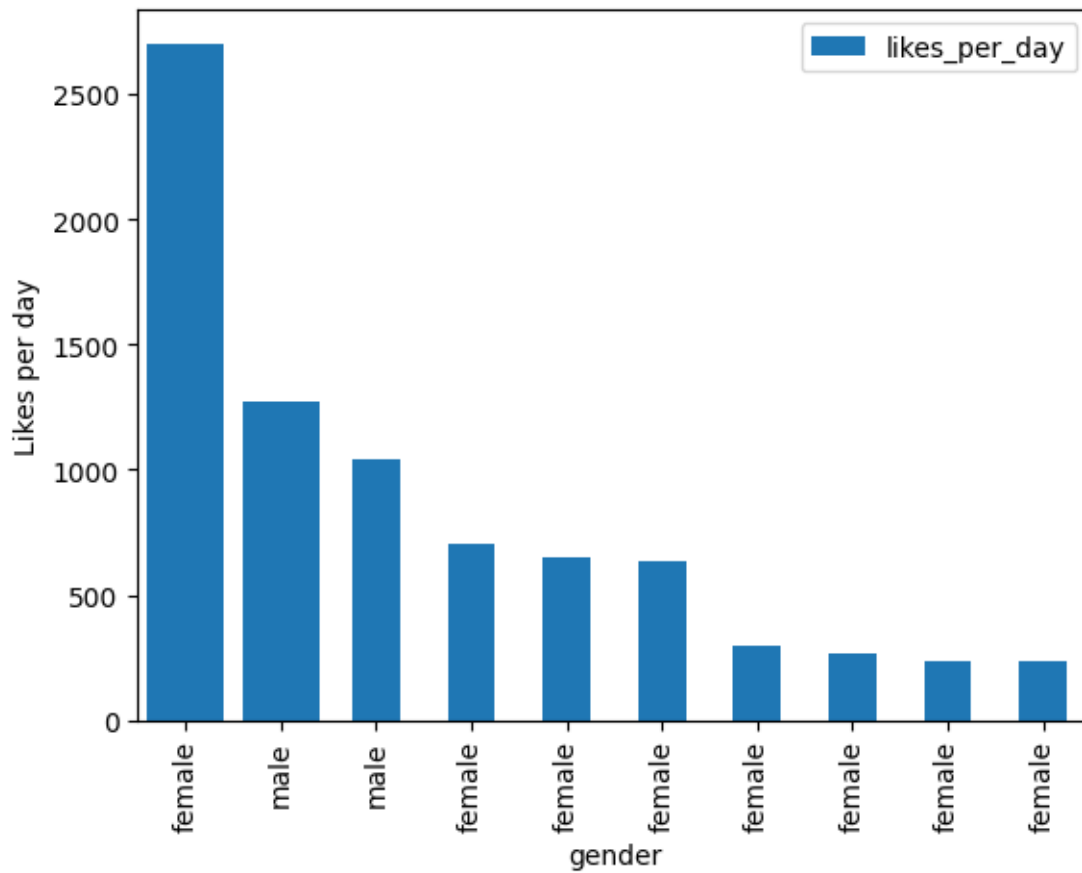
	year_group	likes_range	tenure_range	likes_per_day
94996	1980-1990	2001-3000	0-1 years	2702.000000
94057	1990-2000	5001-6000	0-1 years	1271.000000
98773	1990-2000	1001-2000	0-1 years	1038.509804
77121	1990-2000	2001-3000	0-1 years	704.213439
94906	1990-2000	1001-2000	1-2 years	651.364090

```
[38]: famous.plot(x='userid',y='likes_per_day',kind='bar')
plt.ylabel("Likes per day")
plt.xlabel("User ID")
plt.show()
```

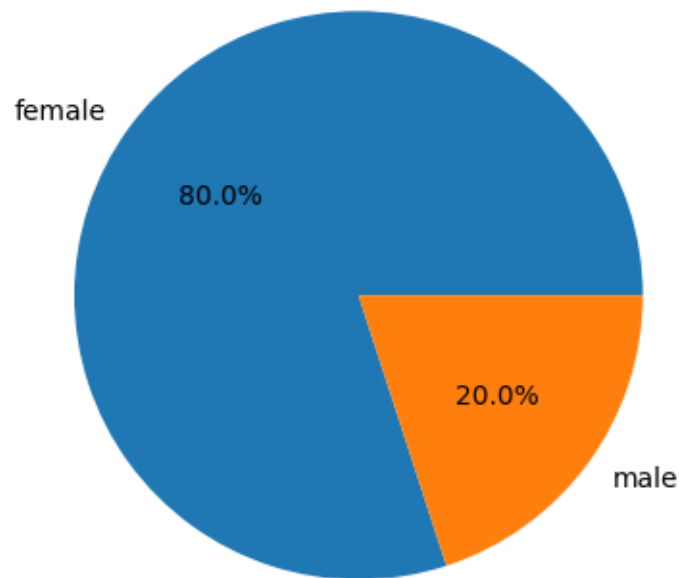


```
[39]: famous.plot(x='gender',y='likes_per_day',kind='bar')
plt.bar(famous.gender,famous.likes_per_day)
```

```
plt.ylabel("Likes per day")
plt.xlabel("gender")
plt.show()
```



```
[40]: plt.pie(famous.gender.
    ↪ value_counts(), labels=['female', 'male'], explode=(0,0), autopct='%1.1f%%')
plt.show()
```



Females have more likes per day as compared to males

## 8.9 Evaluating likes on mobile and www based on tenure range

```
[41]: pivot_table = fbdata.
      pivot_table(values=['mobile_likes_received', 'mobile_likes', 'www_likes_received', 'www_likes'],
      pivot_table
```

```
[41]:
```

	mobile_likes		mobile_likes_received \	
gender	female	male not revealed	female	
tenure_range				
0-1 years	161.021670	55.449295	50.937500	113.900098
1-2 years	171.062174	66.631718	NaN	161.524848
2-3 years	190.117290	60.030498	100.500000	183.920350
3-4 years	195.193971	64.467238	214.083333	186.717572
4-5 years	200.956416	62.295521	72.280000	178.627522
5-6 years	195.880923	55.479475	92.506667	159.533414
6-7 years	196.633758	42.675556	74.214286	170.789809
7-8 years	206.200000	26.250000	0.000000	226.100000
8-9 years	265.500000	15.666667	NaN	380.166667

	www_likes		\	
gender	male not revealed	female	male not revealed	

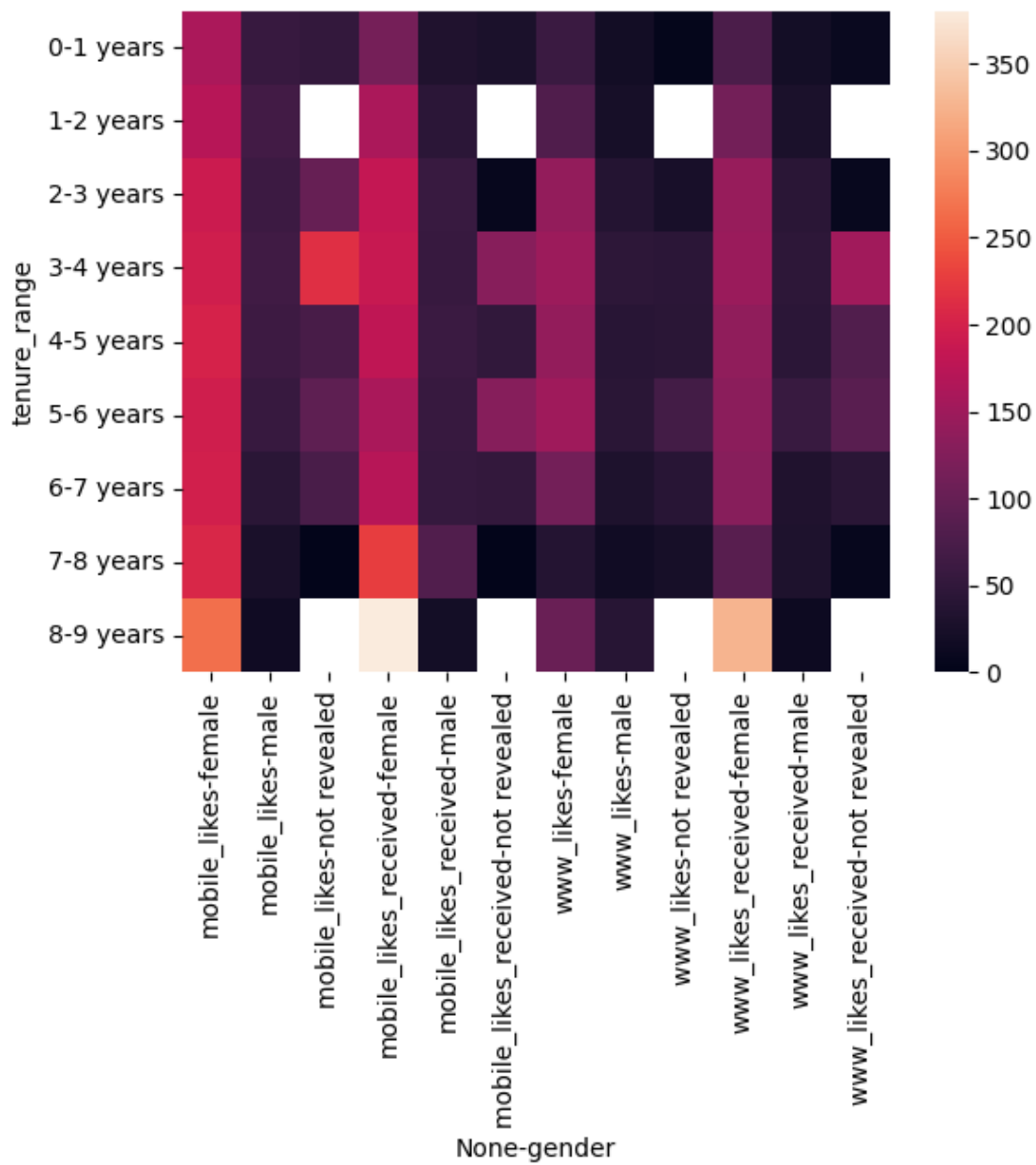


tenure_range					
0-1 years	31.413864	27.875000	59.344209	19.221508	3.562500
1-2 years	44.511169	NaN	79.169273	22.748620	NaN
2-3 years	56.742419	6.500000	139.919621	36.097072	25.000000
3-4 years	56.228130	129.583333	148.349068	45.720412	43.500000
4-5 years	58.890944	49.560000	140.661421	40.704479	42.880000
5-6 years	56.101806	128.173333	151.166464	42.881773	68.306667
6-7 years	54.613333	51.166667	111.831210	30.888889	40.309524
7-8 years	81.250000	1.000000	36.900000	16.500000	23.000000
8-9 years	20.000000	NaN	102.500000	39.666667	NaN

	www_likes_received		
gender	female	male not revealed	
tenure_range			
0-1 years	74.348800	19.734028	9.437500
1-2 years	112.287895	27.496936	NaN
2-3 years	143.989801	42.415824	8.750000
3-4 years	146.449425	44.799657	152.833333
4-5 years	137.153753	43.714703	80.320000
5-6 years	133.479951	56.655172	88.960000
6-7 years	129.557325	32.026667	43.000000
7-8 years	86.300000	31.000000	7.000000
8-9 years	326.333333	12.666667	NaN

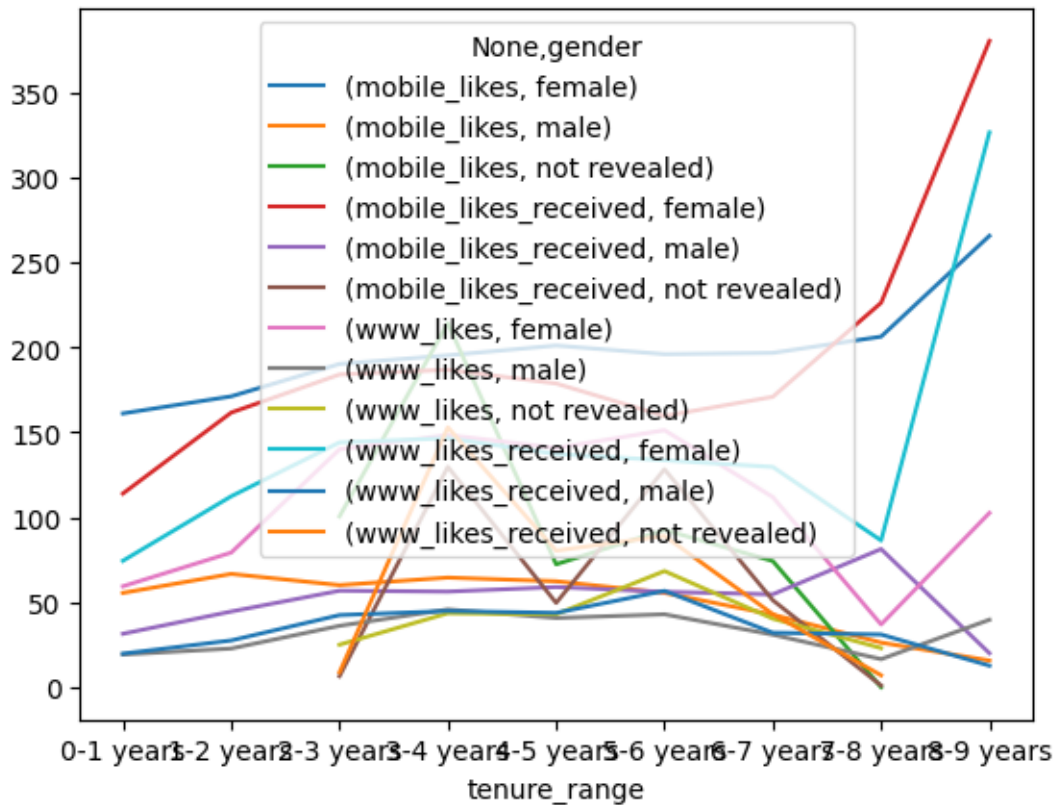
```
[42]: sns.heatmap(pivot_table)
```

```
[42]: <AxesSubplot:xlabel='None-gender', ylabel='tenure_range'>
```



```
[43]: pivot_table.plot()
```

```
[43]: <AxesSubplot:xlabel='tenure_range'>
```



## 8.10 Evaluating likes on mobile and www based on gender

```
[44]: pivot_table2 = fbdata.  
      pivot_table(values=['mobile_likes_received', 'mobile_likes', 'www_likes_received', 'www_likes'],  
                  pivot_table2
```

```
[44]:
```

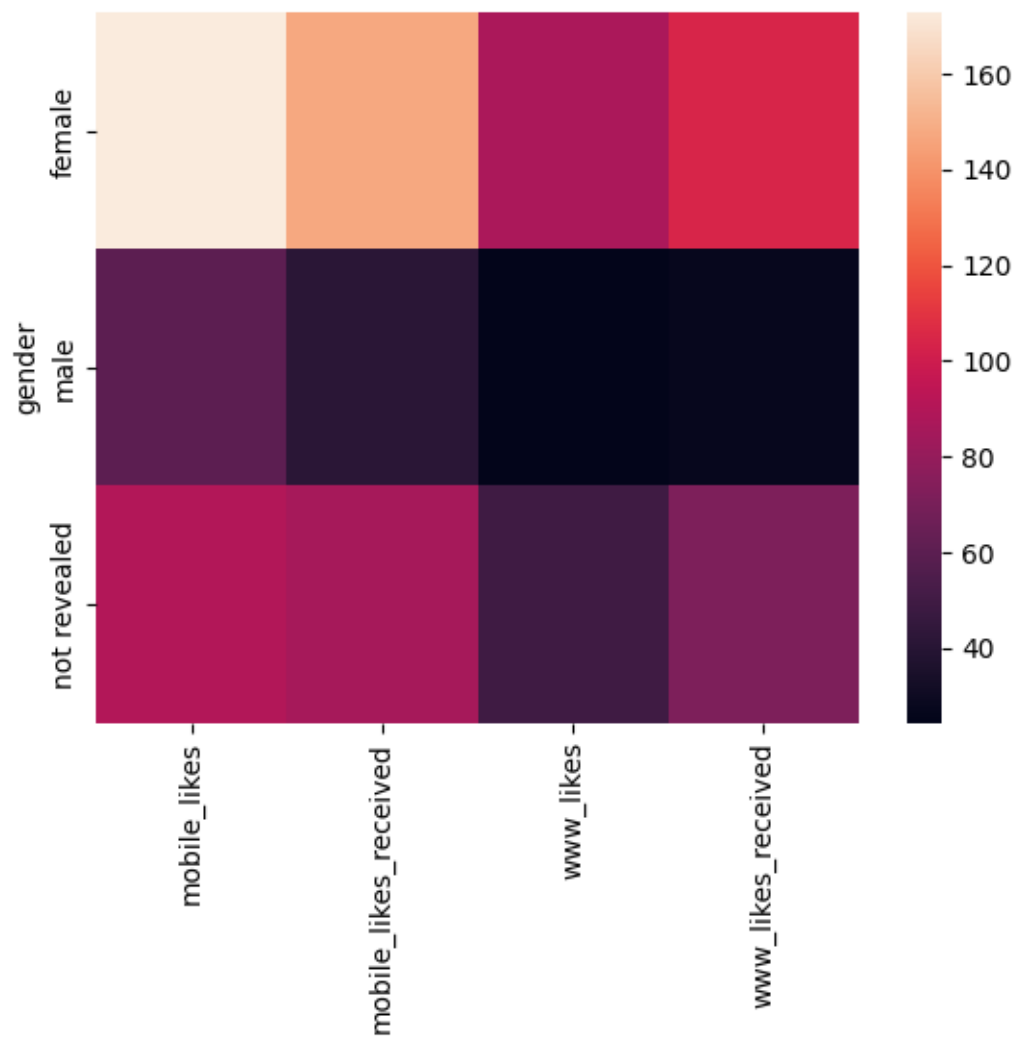
	mobile_likes	mobile_likes_received	www_likes	\
gender				
female	172.921097	147.107597	87.139869	
male	60.261328	40.833015	24.416550	
not revealed	89.417143	85.880000	49.085714	

	www_likes_received
gender	
female	104.338269
male	27.078533
not revealed	71.502857

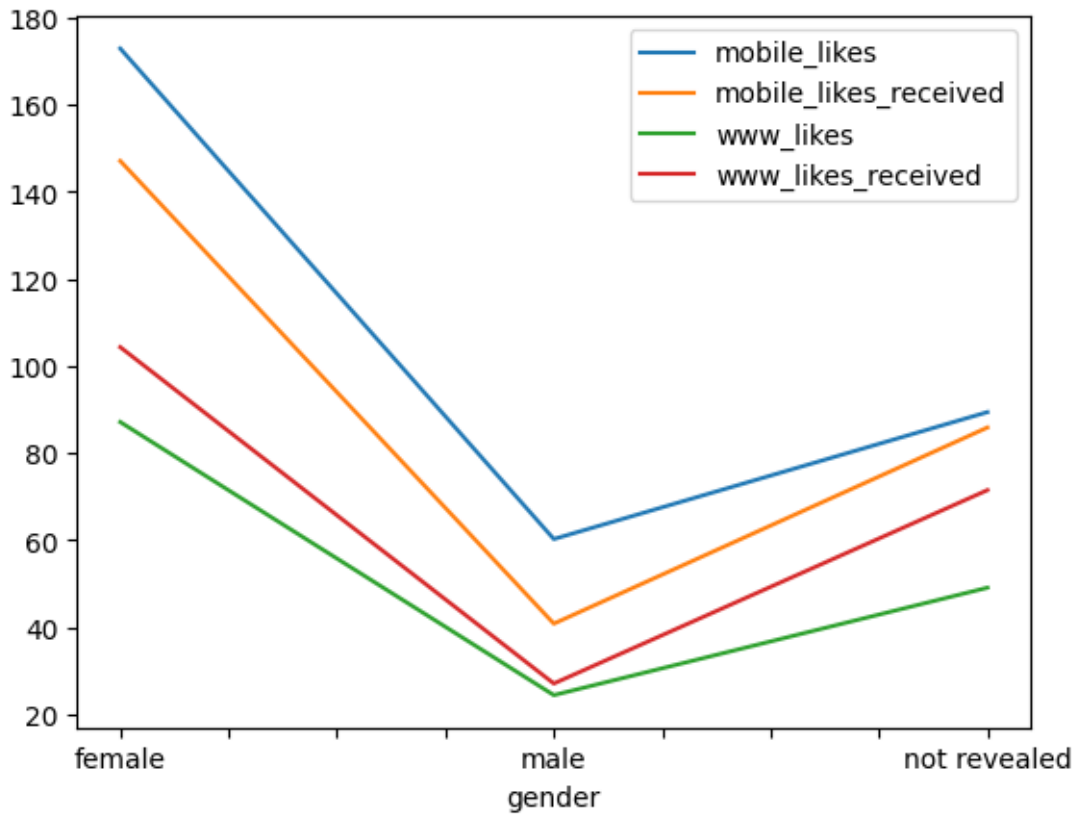
```
[45]: sns.heatmap(pivot_table2)
```

```
[45]: <AxesSubplot:ylabel='gender'>
```



```
[46]: pivot_table2.plot()
```

```
[46]: <AxesSubplot:xlabel='gender'>
```



### 8.11 People willing to initiate friendships

```
[47]: fbdata.sort_values(by='friendships_initiated',ascending=False)[:10]
```

```
[47]:
```

	userid	age	dob_day	dob_year	dob_month	gender	tenure	\
98993	1654565	19	15	1994	8	male	394.0	
98842	1052695	22	23	1991	9	female	874.0	
98675	1949247	19	9	1994	11	female	434.0	
98567	1205425	60	17	1953	6	female	1562.0	
98347	1403953	19	11	1994	11	male	519.0	
98960	1745067	17	1	1996	1	female	947.0	
98898	2010847	18	10	1995	2	female	1084.0	
98949	1103175	15	24	1998	8	female	487.0	
98685	1934087	19	19	1994	5	male	575.0	
98835	1075221	22	23	1991	5	male	907.0	

	friend_count	friendships_initiated	likes	likes_received	\
98993	4538	4144	4501	15088	
98842	4297	3654	1968	2006	
98675	4189	3594	927	2859	

98567	4794	3538	586	1318
98347	3693	3415	170	20
98960	4290	3238	3780	8185
98898	4509	3233	2672	2053
98949	3661	3086	6815	6177
98685	4516	3078	954	3075
98835	4693	3024	2028	948

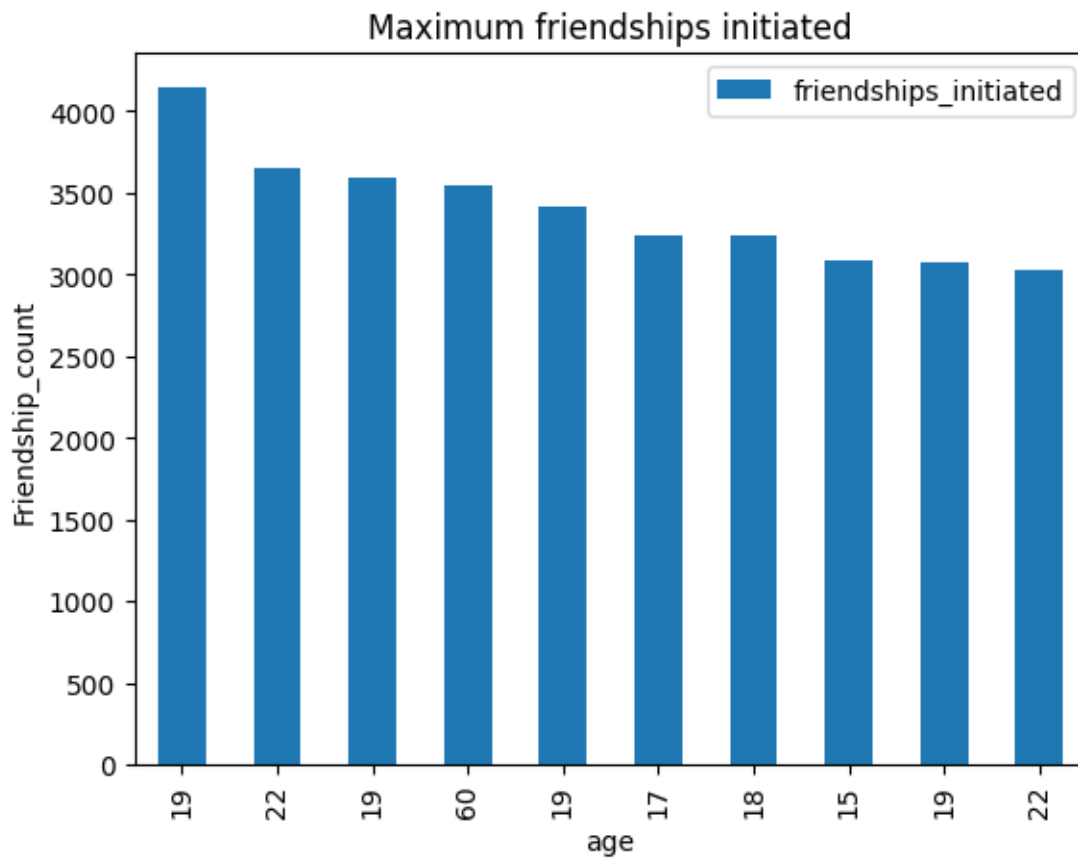
	mobile_likes	mobile_likes_received	www_likes	www_likes_received	\
98993	4435	5961	66	9127	
98842	1825	1632	143	374	
98675	528	1290	399	1569	
98567	560	960	26	358	
98347	170	19	0	1	
98960	1233	5380	2547	2805	
98898	2009	1074	663	979	
98949	1500	3990	5315	2187	
98685	834	2464	120	611	
98835	1990	517	38	431	

	year_group	likes_range	tenure_range	likes_per_day
98993	1990-2000	4001-5000	1-2 years	38.294416
98842	1990-2000	1001-2000	2-3 years	2.295195
98675	1990-2000	0-1000	1-2 years	6.587558
98567	1950-1960	0-1000	4-5 years	0.843790
98347	1990-2000	0-1000	1-2 years	0.038536
98960	1990-2000	3001-4000	2-3 years	8.643083
98898	1990-2000	2001-3000	2-3 years	1.893911
98949	1990-2000	6001-7000	1-2 years	12.683778
98685	1990-2000	0-1000	1-2 years	5.347826
98835	1990-2000	2001-3000	2-3 years	1.045204

```
[48]: interested = fbdata.sort_values(by='friendships_initiated',ascending=False)[:10]
```

### 8.11.1 age vs friendship count

```
[49]: interested.plot(x='age',y='friendships_initiated',kind='bar')
plt.ylabel("Friendship_count")
plt.xlabel("age")
plt.title("Maximum friendships initiated")
plt.show()
```



users with low age initiates more friendships

```
[50]: interested['fc_per_day']=interested.friendships_initiated / interested.tenure
interested
```

```
[50]:
```

	userid	age	dob_day	dob_year	dob_month	gender	tenure	\
98993	1654565	19	15	1994	8	male	394.0	
98842	1052695	22	23	1991	9	female	874.0	
98675	1949247	19	9	1994	11	female	434.0	
98567	1205425	60	17	1953	6	female	1562.0	
98347	1403953	19	11	1994	11	male	519.0	
98960	1745067	17	1	1996	1	female	947.0	
98898	2010847	18	10	1995	2	female	1084.0	
98949	1103175	15	24	1998	8	female	487.0	
98685	1934087	19	19	1994	5	male	575.0	
98835	1075221	22	23	1991	5	male	907.0	

	friend_count	friendships_initiated	likes	likes_received	\
98993	4538	4144	4501	15088	
98842	4297	3654	1968	2006	

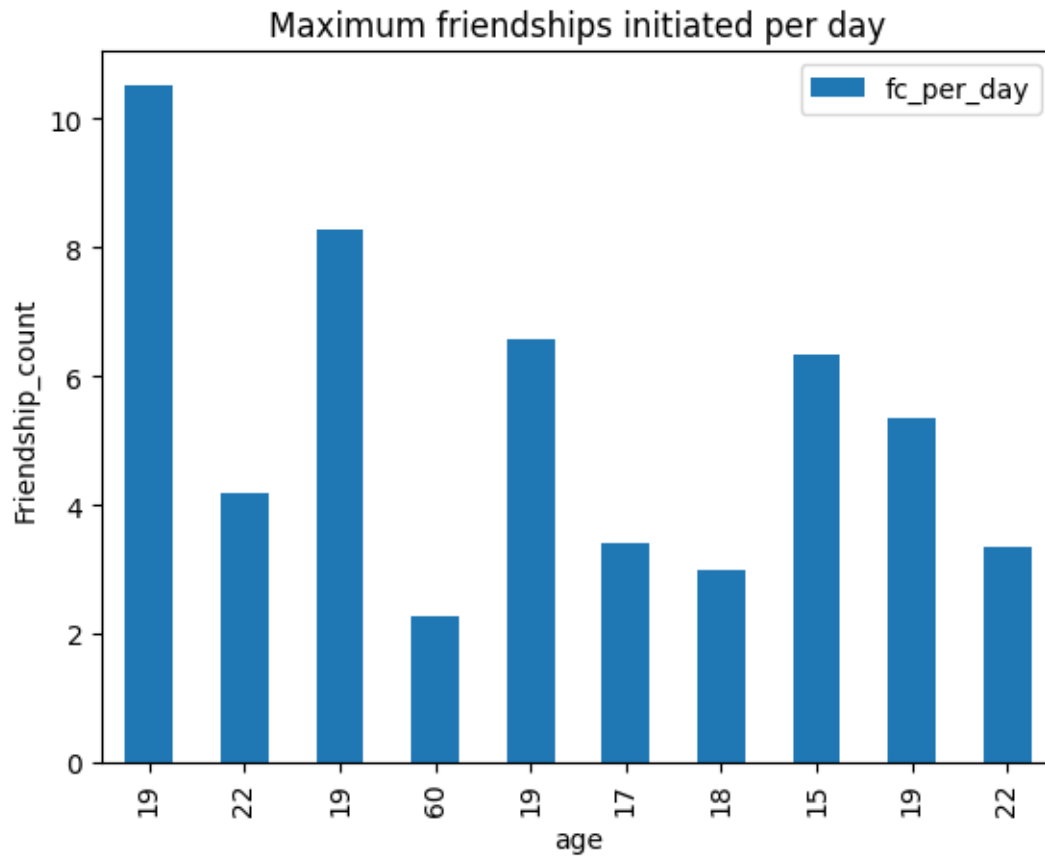
98675	4189	3594	927	2859
98567	4794	3538	586	1318
98347	3693	3415	170	20
98960	4290	3238	3780	8185
98898	4509	3233	2672	2053
98949	3661	3086	6815	6177
98685	4516	3078	954	3075
98835	4693	3024	2028	948

	mobile_likes	mobile_likes_received	www_likes	www_likes_received	\
98993	4435	5961	66		9127
98842	1825	1632	143		374
98675	528	1290	399		1569
98567	560	960	26		358
98347	170	19	0		1
98960	1233	5380	2547		2805
98898	2009	1074	663		979
98949	1500	3990	5315		2187
98685	834	2464	120		611
98835	1990	517	38		431

	year_group	likes_range	tenure_range	likes_per_day	fc_per_day
98993	1990-2000	4001-5000	1-2 years	38.294416	10.517766
98842	1990-2000	1001-2000	2-3 years	2.295195	4.180778
98675	1990-2000	0-1000	1-2 years	6.587558	8.281106
98567	1950-1960	0-1000	4-5 years	0.843790	2.265045
98347	1990-2000	0-1000	1-2 years	0.038536	6.579961
98960	1990-2000	3001-4000	2-3 years	8.643083	3.419219
98898	1990-2000	2001-3000	2-3 years	1.893911	2.982472
98949	1990-2000	6001-7000	1-2 years	12.683778	6.336756
98685	1990-2000	0-1000	1-2 years	5.347826	5.353043
98835	1990-2000	2001-3000	2-3 years	1.045204	3.334068

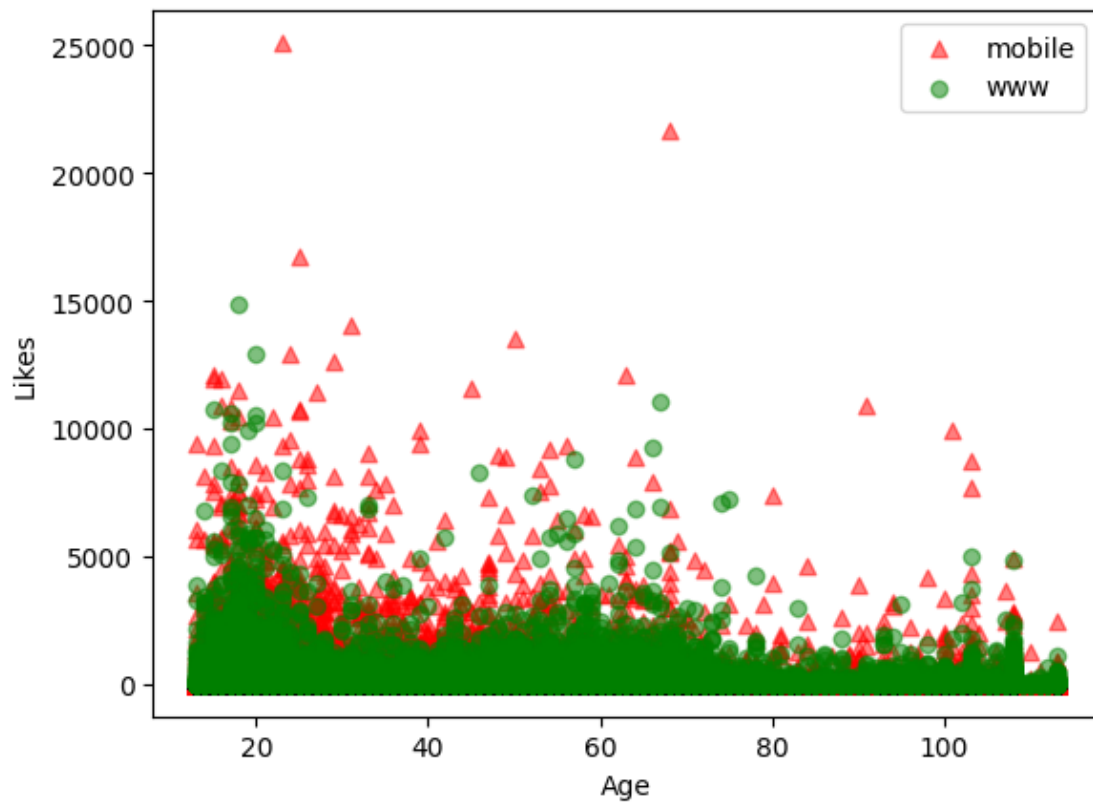
```
[51]: interested.plot(x='age',y='fc_per_day',kind='bar')
plt.ylabel("Friendship_count")
plt.xlabel("age")
plt.title('Maximum friendships initiated per day')
plt.show()
```





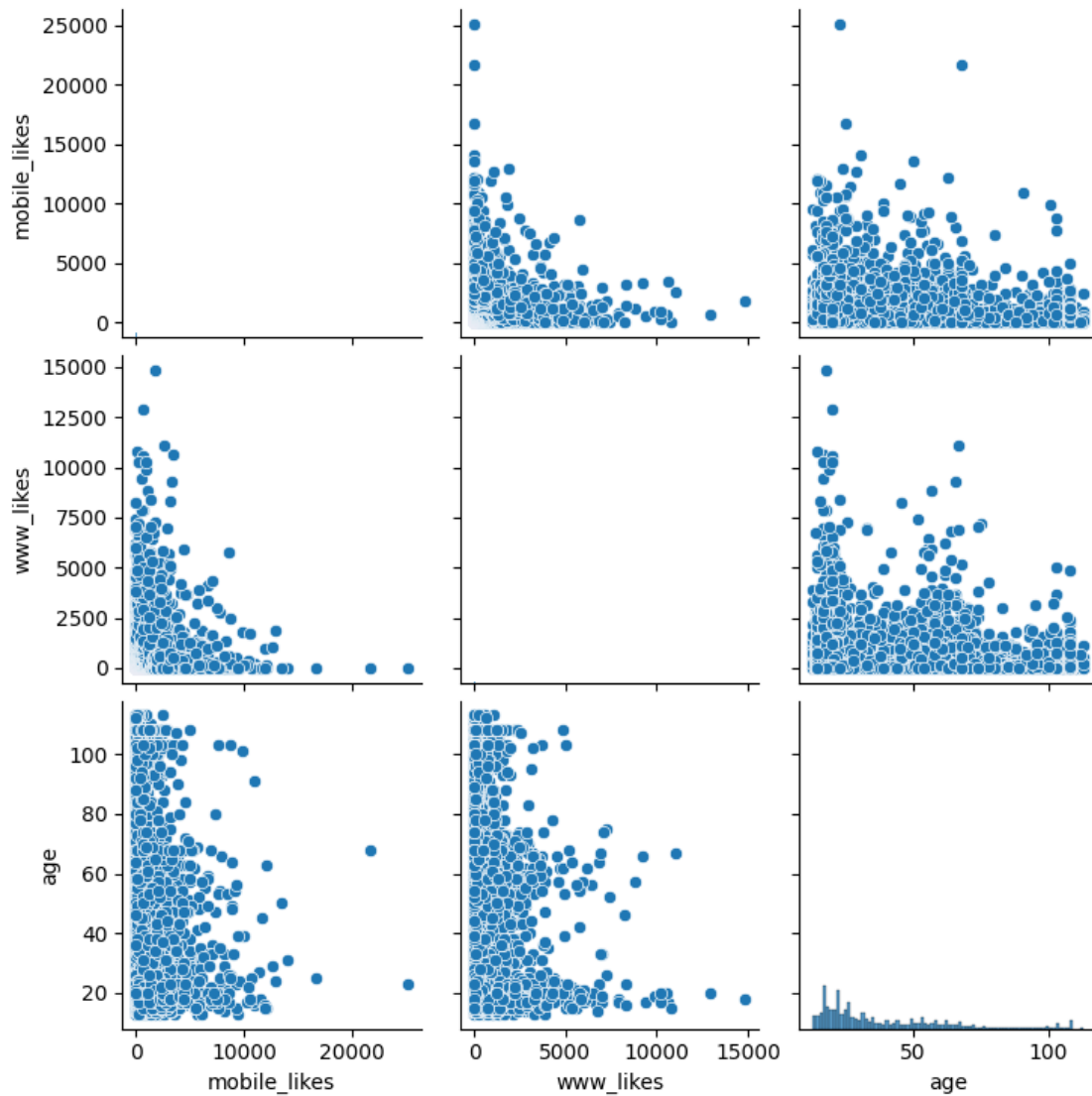
### 8.11.2 Comparing mobile likes, www likes and age

```
[52]: plt.scatter(fbdata.age,fbdata.
        ↳mobile_likes,color='red',label="mobile",marker='^',alpha=0.5)
plt.scatter(fbdata.age,fbdata.www_likes,color='green',label="www",alpha=0.5)
plt.xlabel("Age")
plt.ylabel("Likes")
plt.legend()
plt.show()
```



```
[53]: pairplot_data = fbdata[['mobile_likes', 'www_likes', 'age']]
pairplot_data.head()
sns.pairplot(pairplot_data)
```

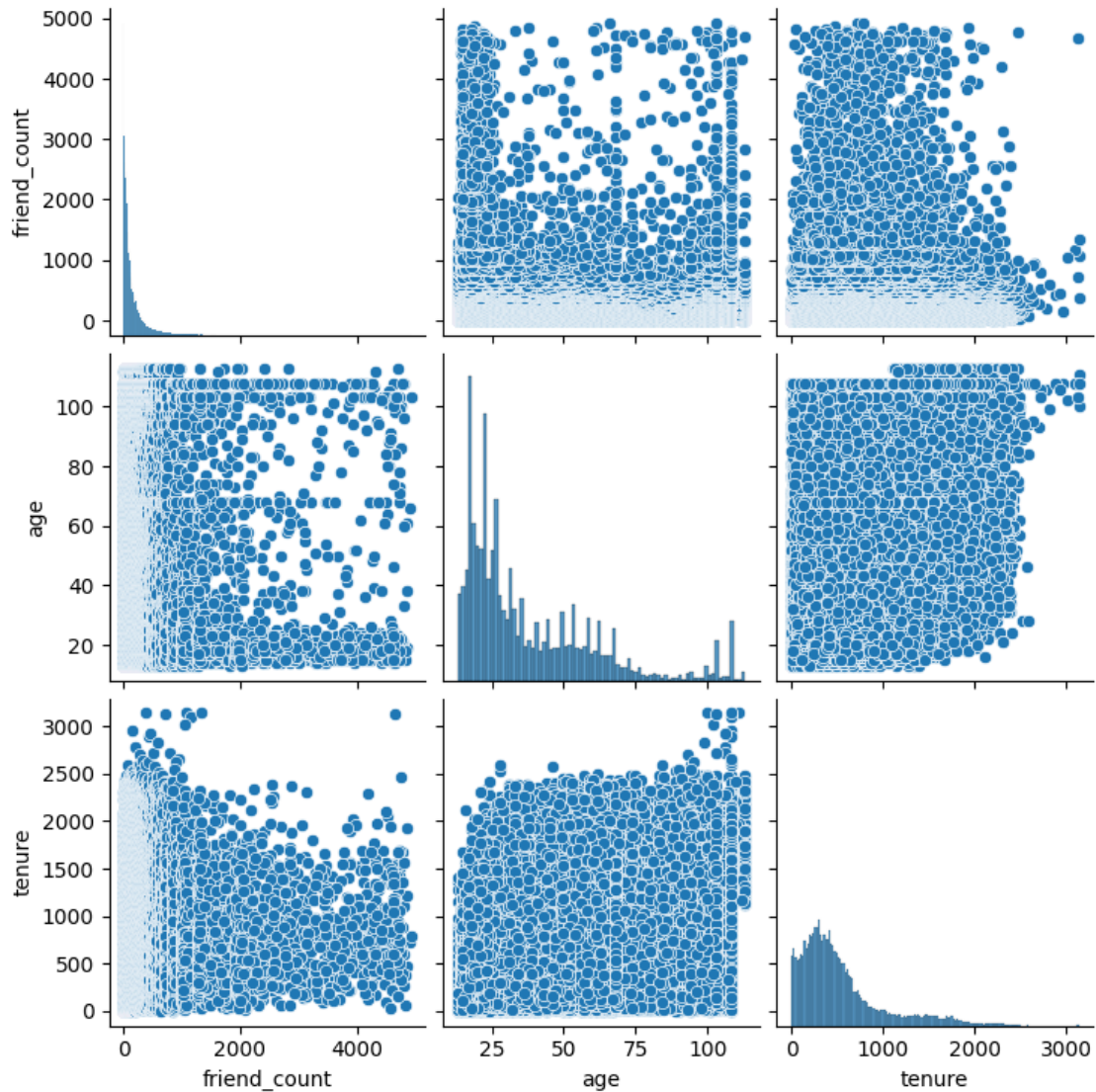
```
[53]: <seaborn.axisgrid.PairGrid at 0x1fe057aa380>
```



### 8.11.3 Comparing friend count, tenure and age

```
[54]: pairplot_data = fbdata[['friend_count', 'age', 'tenure']]
pairplot_data.head()
sns.pairplot(pairplot_data)
```

```
[54]: <seaborn.axisgrid.PairGrid at 0x1fe07aa0fa0>
```



## 8.12 Finding correlation between different attributes

```
[55]: fbdata_corr = fbdata.corr()
fbdata_corr
```

```
[55]:
```

	userid	age	dob_day	dob_year	dob_month	\
userid	1.000000	-0.007296	-0.000817	0.007296	0.002955	
age	-0.007296	1.000000	0.035084	-1.000000	0.025177	
dob_day	-0.000817	0.035084	1.000000	-0.035084	0.129426	
dob_year	0.007296	-1.000000	-0.035084	1.000000	-0.025177	
dob_month	0.002955	0.025177	0.129426	-0.025177	1.000000	
tenure	-0.003446	0.462742	0.041855	-0.462742	0.029446	

friend_count	-0.001314	-0.027396	0.021959	0.027396	0.019809
friendships_initiated	-0.001589	-0.058047	0.022994	0.058047	0.020077
likes	-0.002876	-0.013002	0.015979	0.013002	0.014150
likes_received	0.001526	-0.022568	0.001366	0.022568	0.006496
mobile_likes	-0.004866	-0.026706	0.014537	0.026706	0.010400
mobile_likes_received	0.001753	-0.024245	0.000496	0.024245	0.006436
www_likes	0.001824	0.015584	0.009356	-0.015584	0.012142
www_likes_received	0.001073	-0.018223	0.002460	0.018223	0.006004
likes_per_day	-0.001291	-0.021190	-0.002877	0.021190	-0.004306

	tenure	friend_count	friendships_initiated	\
userid	-0.003446	-0.001314	-0.001589	
age	0.462742	-0.027396	-0.058047	
dob_day	0.041855	0.021959	0.022994	
dob_year	-0.462742	0.027396	0.058047	
dob_month	0.029446	0.019809	0.020077	
tenure	1.000000	0.166256	0.133505	
friend_count	0.166256	1.000000	0.825850	
friendships_initiated	0.133505	0.825850	1.000000	
likes	0.057132	0.298016	0.285592	
likes_received	0.027745	0.236463	0.175131	
mobile_likes	0.028052	0.235656	0.229807	
mobile_likes_received	0.023971	0.232700	0.173803	
www_likes	0.070757	0.229803	0.214023	
www_likes_received	0.030553	0.220727	0.161439	
likes_per_day	-0.031023	0.079176	0.052640	

	likes	likes_received	mobile_likes	\
userid	-0.002876	0.001526	-0.004866	
age	-0.013002	-0.022568	-0.026706	
dob_day	0.015979	0.001366	0.014537	
dob_year	0.013002	0.022568	0.026706	
dob_month	0.014150	0.006496	0.010400	
tenure	0.057132	0.027745	0.028052	
friend_count	0.298016	0.236463	0.235656	
friendships_initiated	0.285592	0.175131	0.229807	
likes	1.000000	0.327376	0.871652	
likes_received	0.327376	1.000000	0.256997	
mobile_likes	0.871652	0.256997	1.000000	
mobile_likes_received	0.329258	0.973679	0.288513	
www_likes	0.644960	0.255365	0.187619	
www_likes_received	0.295687	0.947990	0.190173	
likes_per_day	0.152107	0.363036	0.137597	

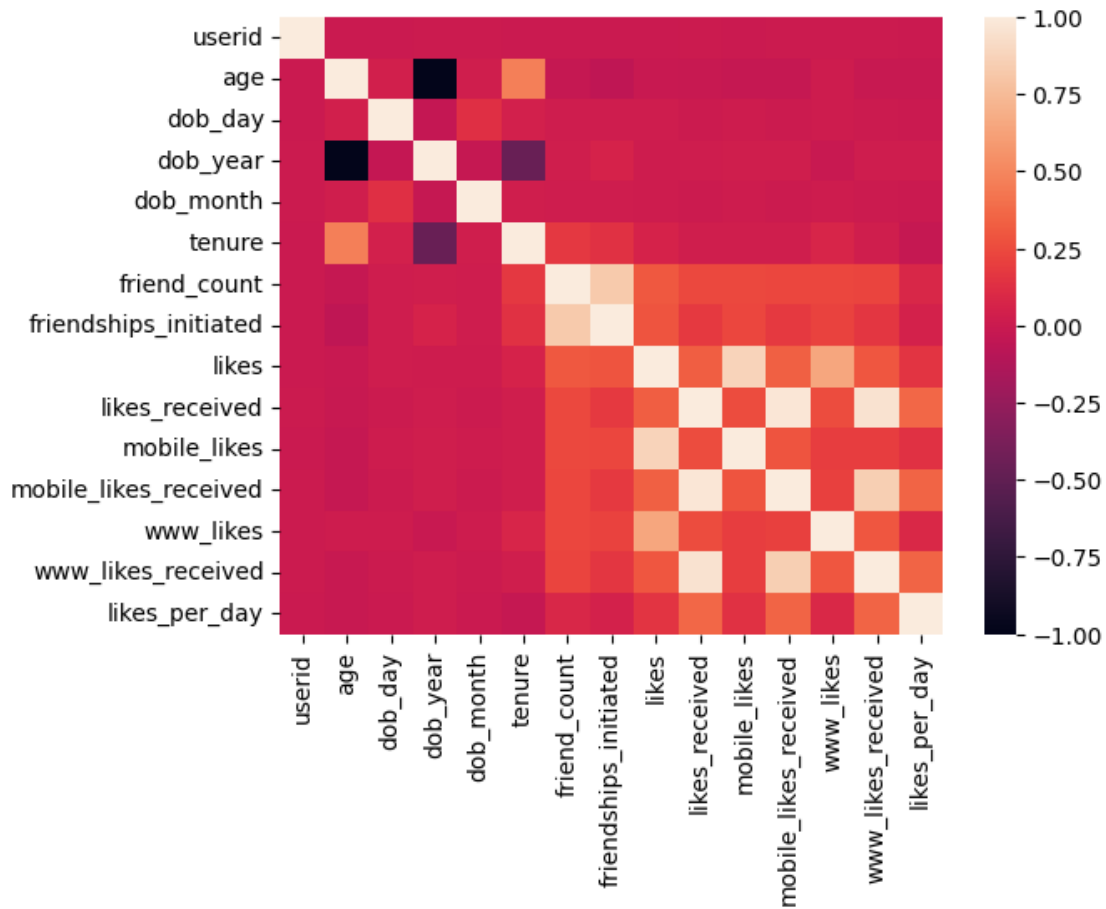
	mobile_likes_received	www_likes	www_likes_received	\
userid	0.001753	0.001824	0.001073	
age	-0.024245	0.015584	-0.018223	

dob_day	0.000496	0.009356	0.002460
dob_year	0.024245	-0.015584	0.018223
dob_month	0.006436	0.012142	0.006004
tenure	0.023971	0.070757	0.030553
friend_count	0.232700	0.229803	0.220727
friendships_initiated	0.173803	0.214023	0.161439
likes	0.329258	0.644960	0.295687
likes_received	0.973679	0.255365	0.947990
mobile_likes	0.288513	0.187619	0.190173
mobile_likes_received	1.000000	0.209998	0.850490
www_likes	0.209998	1.000000	0.296053
www_likes_received	0.850490	0.296053	1.000000
likes_per_day	0.349276	0.090286	0.350028

	likes_per_day
userid	-0.001291
age	-0.021190
dob_day	-0.002877
dob_year	0.021190
dob_month	-0.004306
tenure	-0.031023
friend_count	0.079176
friendships_initiated	0.052640
likes	0.152107
likes_received	0.363036
mobile_likes	0.137597
mobile_likes_received	0.349276
www_likes	0.090286
www_likes_received	0.350028
likes_per_day	1.000000

```
[56]: sns.heatmap(fbdata_corr)
```

```
[56]: <AxesSubplot:>
```



## 9 9.Summarisation

1. Most of the Users are of age near 20 years.
2. Day and Month of born are uniformly spread across the range.
3. Most of the users are born after 1980.
4. Number of likes are greater for females as compared to males across all the year groups.
5. Males have more number of zero friends than females.
6. As age increases, the number of likes decreases.
7. Females have more likes per day as compared to males.
8. Users with low age initiates more friendships.