

```
In [1]: import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd
import numpy as np

In [2]: data = pd.read_excel(r"D:\python data\facebook_user_data1.xlsx")

In [3]: data.shape

Out[3]: (99003, 15)

In [4]: data.head()

Out[4]:   userid  age  dob_day  dob_year  dob_month  gender  tenure  friend_count  friendships_initiated  likes  likes_received  mobile_likes  mobile_likes_received  www_likes  www_likes_received
0  2094382   14     19    1999         11  male    266.0         0           0  0  0  0  0  0  0
1  1192601   14      2    1999         11 female     6.0         0           0  0  0  0  0  0  0
2  2083884   14    16    1999         11  male    13.0         0           0  0  0  0  0  0  0
3  1203168   14    25    1999         12 female    93.0         0           0  0  0  0  0  0  0
4  1733186   14      4    1999         12  male    82.0         0           0  0  0  0  0  0  0

In [5]: data.isnull().sum()

Out[5]: 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

In [6]: data['gender'] = data['gender'].fillna(data['gender'].mode()[0])

In [9]: data.isnull().sum()

Out[9]: userid          0
age          0
dob_day      0
dob_year     0
dob_month    0
gender        0
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

In [11]: data['tenure'] = data['tenure'].fillna(data['tenure'].median())

In [12]: data.isnull().sum()

Out[12]: 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

In [13]: data=np.random.rand(20,15)
data

sns.heatmap(data)

plt.figure(figsize=(20,15))
sns.heatmap(data, annot=True)

Out[13]: <AxesSubplot:~>

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14
1 2 3 4 5 6 7 8 9 10 11 12 13 14
2 3 4 5 6 7 8 9 10 11 12 13 14
3 4 5 6 7 8 9 10 11 12 13 14
4 5 6 7 8 9 10 11 12 13 14
5 6 7 8 9 10 11 12 13 14
6 7 8 9 10 11 12 13 14
7 8 9 10 11 12 13 14
8 9 10 11 12 13 14
9 10 11 12 13 14
10 11 12 13 14
11 12 13 14
12 13 14
13 14
14
15
16
17
18
19
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14

0.41 0.23 0.62 0.74 0.18 0.11 0.22 0.5 0.56 0.2 0.38 0.082 0.1 0.4 0.7
0.71 0.8 0.88 0.0056 0.098 0.98 0.87 0.1 0.0079 0.48 0.21 0.72 0.46 0.34 0.91
0.8 0.57 0.085 0.38 0.82 0.36 0.15 0.95 0.56 0.98 0.8 0.018 0.1 0.11 0.78
0.43 0.019 0.15 0.43 0.13 0.32 0.51 0.72 0.69 0.068 0.88 0.43 0.7 0.72 0.59
0.38 0.56 0.18 0.41 0.82 0.99 0.86 0.44 0.74 0.042 0.34 0.33 0.12 0.51 0.55
0.98 0.38 0.77 0.86 0.38 0.33 0.24 0.29 0.93 0.48 0.51 0.38 0.61 0.21 0.81
0.11 0.43 0.98 0.42 0.23 0.51 0.11 0.65 0.43 0.48 0.45 0.28 0.62 0.62 0.069
0.19 0.0008 0.55 0.63 0.38 0.83 0.79 0.32 0.19 0.58 0.2 0.18 0.6 0.12 0.26
0.076 0.86 0.7 0.24 0.92 0.26 0.69 0.67 0.8 0.62 0.51 0.44 0.68 0.59 0.57
0.4 0.86 0.21 0.1 0.84 0.045 0.38 0.38 0.71 0.47 0.44 0.069 0.78 0.24 0.0041
0.63 0.18 1 0.77 0.53 0.92 0.82 0.62 0.58 0.82 0.63 0.76 0.67 0.53 0.98
0.053 0.28 0.42 0.26 0.82 0.058 0.8 0.66 0.18 0.21 0.69 0.72 0.68 0.53 0.95
0.65 0.043 0.96 0.52 0.86 0.094 0.6 0.72 0.76 0.16 0.077 0.68 0.4 0.49 0.97
0.045 0.72 0.19 0.32 0.91 0.63 0.72 0.34 0.23 0.76 0.91 0.95 0.58 0.44 0.84
0.77 0.25 0.46 0.61 0.97 0.99 0.75 0.86 0.35 0.11 0.18 0.33 0.15 0.4 0.041
0.72 0.81 0.78 0.72 0.92 0.15 0.84 0.72 0.63 0.59 0.58 0.94 0.0063 1 0.65
0.96 0.54 0.81 0.88 0.53 0.62 0.96 0.69 0.69 0.66 0.79 0.21 0.29 0.017 0.49
0.86 0.92 0.29 0.16 0.25 0.14 0.00056 0.74 0.4 0.23 0.89 0.46 0.56 0.84 0.84
0.94 0.65 0.98 0.86 0.19 0.92 0.84 0.31 0.077 0.48 0.24 0.36 0.52 0.52 0.99
0.31 0.53 0.012 0.58 0.053 0.38 0.06 0.056 0.85 0.27 0.77 0.076 0.63 0.4 0.05

In [15]: data = pd.read_excel(r"D:\python data\facebook_user_data.xlsx")
data.groupby(['gender']).size().reset_index(name='count')

Out[15]:   gender  count
0  female  40254
1   male   58574

In [8]: data.groupby('gender').agg({'friend_count':['sum']})

Out[8]:   friend_count
sum
gender
female  9740258
male    9699059

In [16]: data.groupby('gender').agg({'friendships_initiated':['sum']})

Out[16]:   friendships_initiated
sum
gender
female  4584894
male    6037023

In [23]: data.groupby('tenure').gender.hist()

Out[23]: tenure
0.0 AxesSubplot(0.125,0.125;0.775x0.755)
1.0 AxesSubplot(0.125,0.125;0.775x0.755)
2.0 AxesSubplot(0.125,0.125;0.775x0.755)
3.0 AxesSubplot(0.125,0.125;0.775x0.755)
4.0 AxesSubplot(0.125,0.125;0.775x0.755)
...
3019.0 AxesSubplot(0.125,0.125;0.775x0.755)
3101.0 AxesSubplot(0.125,0.125;0.775x0.755)
3128.0 AxesSubplot(0.125,0.125;0.775x0.755)
3129.0 AxesSubplot(0.125,0.125;0.775x0.755)
3139.0 AxesSubplot(0.125,0.125;0.775x0.755)
Name: gender, Length: 2426, dtype: object

120
100
80
60
40
20
0
female male

In [18]: data[data.friend_count < 1].shape[0]
len(data[data.friend_count < 1])
data[data.friend_count < 1]

Out[18]:   userid  age  dob_day  dob_year  dob_month  gender  tenure  friend_count  friendships_initiated  likes  likes_received  mobile_likes  mobile_likes_received  www_likes  www_likes_received
0  2094382   14     19    1999         11  male    266.0         0           0  0  0  0  0  0  0
1  1192601   14      2    1999         11 female     6.0         0           0  0  0  0  0  0  0
2  2083884   14    16    1999         11  male    13.0         0           0  0  0  0  0  0  0
3  1203168   14    25    1999         12 female    93.0         0           0  0  0  0  0  0  0
4  1733186   14      4    1999         12  male    82.0         0           0  0  0  0  0  0  0
...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...
1957 1395843   90      4    1923         7  male     0.0         0           0  95  78  95  56  0  22
1958 1306409   23     24    1990         1  male    48.0         0           0  164  0  164  0  0  0
1959 1838179   18    29    1995        10 female     0.0         0           0  246  100  0  9  246  91
1960 1594192   63      1    1950         1  female    57.0         0           0  271  1  271  0  0  1
1961 1609579   20    23    1993         1  female   257.0         0           0  848  1  848  0  0  1

1962 rows x 15 columns

In [19]: data[data.likes < 1].shape[0]
len(data[data.likes < 1])
data[data.likes < 1].count()

Out[19]: userid          22308
age          22308
dob_day      22308
dob_year     22308
dob_month    22308
gender       22295
tenure       22308
friend_count  22308
friendships_initiated  22308
likes        22308
likes_received  22308
mobile_likes  22308
mobile_likes_received  22308
www_likes    22308
www_likes_received  22308
dtype: int64

In [20]: data[data.likes_received < 1].shape[0]
len(data[data.likes_received < 1])
data[data.likes_received < 1]

Out[20]:   userid  age  dob_day  dob_year  dob_month  gender  tenure  friend_count  friendships_initiated  likes  likes_received  mobile_likes  mobile_likes_received  www_likes  www_likes_received
0  2094382   14     19    1999         11  male    266.0         0           0  0  0  0  0  0  0
1  1192601   14      2    1999         11 female     6.0         0           0  0  0  0  0  0  0
2  2083884   14    16    1999         11  male    13.0         0           0  0  0  0  0  0  0
3  1203168   14    25    1999         12 female    93.0         0           0  0  0  0  0  0  0
4  1733186   14      4    1999         12  male    82.0         0           0  0  0  0  0  0  0
...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...
98005 1883270   18     14    1995         11  female   341.0        2050         467  0  0  0  0  0  0
98016 1993020   21     8    1992         1  male   1262.0       1954        1050  0  0  0  0  0  0
98061 1680023   17    13    1996         10 female   736.0       2415        1795  3  0  0  0  3  0
98081 2022131   16    14    1997         8  male    743.0       2294        1157  6  0  6  0  0  0
98082 1867596   41    14    1972         3  male    412.0       1936        1488  6  0  6  0  0  0

24428 rows x 15 columns

In [21]: data.groupby("gender")[["mobile_likes", "www_likes"]].mean()

Out[21]:   mobile_likes  www_likes
gender
female    172.912928  87.138297
male      60.261328  24.416550

In [22]: data.groupby("gender")[["mobile_likes_received", "www_likes_received"]].mean()

Out[22]:   mobile_likes_received  www_likes_received
gender
female    147.100884    104.334451
male       40.833015     27.078533

In [ ]: data.plot(x="gender", y=["tenure"], kind="bar")

In [ ]:
```