Name- Ninad Kasar.

Session 4.

Certificate course.

Roll_no- 33.

Problem Statement:

Perform the following operations using Python on the Facebook metrics data sets

a. Create data subsets , b. Merge Data, c. Sort Data, d. Transposing Data, e. Shape and reshape Data

```
import pandas as pd
import numpy as np

df=pd.read_csv('pseudo_facebook.csv')
df
```

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

[99003 rows x 15 columns]
```

```
#Getting basic info about our dataset.

df.head()
```

```
    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
```

```
df.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
```

```
#Creation of data subsets.
#Using the loc function to create subsets.
```

```
df_1= df.loc[0:15,['userid','age','gender','tenure']]
df_1
```

```
     userid  age  gender  tenure
0   2094382   14    male   266.0
1   1192601   14  female     6.0
2   2083884   14    male    13.0
3   1203168   14  female    93.0
4   1733186   14    male    82.0
5   1524765   14    male    15.0
6   1136133   13    male    12.0
7   1680361   13  female     0.0
8   1365174   13    male    81.0
9   1712567   13    male   171.0
10  1612453   13    male    98.0
```

```
11  2104073   13    male    55.0
12  1918584   13    male   106.0
13  1704433   13    male    61.0
14  1932519   13  female     0.0
15  1751722   13  female    16.0
```

```
df_3= df.loc[0:15,['userid','friend_count']]
df_3.head()
```

```
    userid  friend_count
0  2094382             0
1  1192601             0
2  2083884             0
3  1203168             0
4  1733186             0
```

```
#Merging data.
#Merged data using the userid column.
```

```
df_merge= df_1.merge(df_3,how="right")
df_merge
```

```
     userid  age  gender  tenure  friend_count
0   2094382   14    male   266.0             0
1   1192601   14  female     6.0             0
2   2083884   14    male    13.0             0
3   1203168   14  female    93.0             0
4   1733186   14    male    82.0             0
5   1524765   14    male    15.0             0
6   1136133   13    male    12.0             0
7   1680361   13  female     0.0             0
8   1365174   13    male    81.0             0
9   1712567   13    male   171.0             0
10  1612453   13    male    98.0             0
11  2104073   13    male    55.0             0
12  1918584   13    male   106.0             0
13  1704433   13    male    61.0             0
14  1932519   13  female     0.0             0
15  1751722   13  female    16.0             0
```

```
#Sorting data using userid.
```

```
df_1.sort_values(by=['userid'])
```

```
     userid  age  gender  tenure
6   1136133   13    male    12.0
1   1192601   14  female     6.0
3   1203168   14  female    93.0
8   1365174   13    male    81.0
5   1524765   14    male    15.0
10  1612453   13    male    98.0
```

```
7   1680361   13  female     0.0
13  1704433   13    male    61.0
9   1712567   13    male   171.0
4   1733186   14    male    82.0
15  1751722   13  female    16.0
12  1918584   13    male   106.0
14  1932519   13  female     0.0
2   2083884   14    male    13.0
0   2094382   14    male   266.0
11  2104073   13    male    55.0
```

```
#Transpose data.

df_1_transpose= df_1.T
df_1_transpose
```

```
             0        1        2        3        4        5        6
\
userid  2094382  1192601  2083884  1203168  1733186  1524765  1136133

age          14       14       14       14       14       14       13

gender     male   female     male   female     male     male     male

tenure    266.0      6.0     13.0     93.0     82.0     15.0     12.0


             7        8        9        10       11       12       13
\
userid  1680361  1365174  1712567  1612453  2104073  1918584  1704433

age          13       13       13       13       13       13       13

gender   female     male     male     male     male     male     male

tenure      0.0     81.0    171.0     98.0     55.0    106.0     61.0


             14       15
userid  1932519  1751722
age          13       13
gender   female   female
tenure      0.0     16.0
```

```
#Shape data.
df.shape
```

```
(99003, 15)
```

```
df_1.shape
```

```
(16, 4)
```

```
#Reshaping data.

df_1_numpy= df_1.to_numpy()
df_1_numpy
```

```
array([[2094382, 14, 'male', 266.0],
       [1192601, 14, 'female', 6.0],
       [2083884, 14, 'male', 13.0],
       [1203168, 14, 'female', 93.0],
       [1733186, 14, 'male', 82.0],
       [1524765, 14, 'male', 15.0],
       [1136133, 13, 'male', 12.0],
       [1680361, 13, 'female', 0.0],
       [1365174, 13, 'male', 81.0],
       [1712567, 13, 'male', 171.0],
       [1612453, 13, 'male', 98.0],
       [2104073, 13, 'male', 55.0],
       [1918584, 13, 'male', 106.0],
       [1704433, 13, 'male', 61.0],
       [1932519, 13, 'female', 0.0],
       [1751722, 13, 'female', 16.0]], dtype=object)
```

```
np.reshape(df_1_numpy, (-2,2))
```

```
array([[2094382, 14],
       ['male', 266.0],
       [1192601, 14],
       ['female', 6.0],
       [2083884, 14],
       ['male', 13.0],
       [1203168, 14],
       ['female', 93.0],
       [1733186, 14],
       ['male', 82.0],
       [1524765, 14],
       ['male', 15.0],
       [1136133, 13],
       ['male', 12.0],
       [1680361, 13],
       ['female', 0.0],
       [1365174, 13],
       ['male', 81.0],
       [1712567, 13],
       ['male', 171.0],
       [1612453, 13],
       ['male', 98.0],
       [2104073, 13],
       ['male', 55.0],
       [1918584, 13],
```

```
['male', 106.0],
[1704433, 13],
['male', 61.0],
[1932519, 13],
['female', 0.0],
[1751722, 13],
['female', 16.0]], dtype=object)
```