

Assignment no. - 1

1. 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

Importing the required libraries

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

Reading the csv file

```
In [2]: data = pd.read_csv("dataset_Facebook.csv")
```

```
In [3]: data
```

```
Out[3]:
```

	Page total likes	Type	Category	Post Month	Post Weekday	Post Hour	Paid	Lifetime Post Total Reach	Lifetime Post Total Impressions	Lifetime Engaged Users	Lifetime Post Consumers	C
0	139441	Photo		2	12	4	3	0.0	2752	5091	178	109
1	139441	Status		2	12	3	10	0.0	10460	19057	1457	1361
2	139441	Photo		3	12	3	3	0.0	2413	4373	177	113
3	139441	Photo		2	12	2	10	1.0	50128	87991	2211	790
4	139441	Photo		2	12	2	3	0.0	7244	13594	671	410
...
495	85093	Photo		3	1	7	2	0.0	4684	7536	733	708
496	81370	Photo		2	1	5	8	0.0	3480	6229	537	508
497	81370	Photo		1	1	5	2	0.0	3778	7216	625	572
498	81370	Photo		3	1	4	11	0.0	4156	7564	626	574
499	81370	Photo		2	1	4	4	NaN	4188	7292	564	524

500 rows × 19 columns



Removing null valued rows

```
In [4]: data = data.dropna()  
data
```

Out[4]:

	Page total likes	Type	Category	Post Month	Post Weekday	Post Hour	Paid	Lifetime Post Total Reach	Lifetime Post Total Impressions	Lifetime Engaged Users	Lifetime Post Consumers	C
0	139441	Photo		2	12	4	3	0.0	2752	5091	178	109
1	139441	Status		2	12	3	10	0.0	10460	19057	1457	1361
2	139441	Photo		3	12	3	3	0.0	2413	4373	177	113
3	139441	Photo		2	12	2	10	1.0	50128	87991	2211	790
4	139441	Photo		2	12	2	3	0.0	7244	13594	671	410
...
494	85093	Photo		3	1	7	10	0.0	5400	9218	810	756
495	85093	Photo		3	1	7	2	0.0	4684	7536	733	708
496	81370	Photo		2	1	5	8	0.0	3480	6229	537	508
497	81370	Photo		1	1	5	2	0.0	3778	7216	625	572
498	81370	Photo		3	1	4	11	0.0	4156	7564	626	574

495 rows × 19 columns

Shape, Head and Tail of data

```
In [5]: data.shape
```

Out[5]: (495, 19)

In [6]: `data.head()`

Out[6]:

	Page total likes	Type	Category	Post Month	Post Weekday	Post Hour	Paid	Lifetime Post Total Reach	Lifetime Post Total Impressions	Lifetime Engaged Users	Lifetime Post Consumers	Lif Con
0	139441	Photo		2	12	4	3	0.0	2752	5091	178	109
1	139441	Status		2	12	3	10	0.0	10460	19057	1457	1361
2	139441	Photo		3	12	3	3	0.0	2413	4373	177	113
3	139441	Photo		2	12	2	10	1.0	50128	87991	2211	790
4	139441	Photo		2	12	2	3	0.0	7244	13594	671	410

In [7]: `data.tail()`

Out[7]:

	Page total likes	Type	Category	Post Month	Post Weekday	Post Hour	Paid	Lifetime Post Total Reach	Lifetime Post Total Impressions	Lifetime Engaged Users	Lifetime Post Consumers	Li Cor
494	85093	Photo		3	1	7	10	0.0	5400	9218	810	756
495	85093	Photo		3	1	7	2	0.0	4684	7536	733	708
496	81370	Photo		2	1	5	8	0.0	3480	6229	537	508
497	81370	Photo		1	1	5	2	0.0	3778	7216	625	572
498	81370	Photo		3	1	4	11	0.0	4156	7564	626	574

Creating the first column subset

```
In [8]: colm_sub1 = data[["Type", "Category", "Post Weekday", "Post Hour"]]
        colm_sub1
```

Out[8]:

	Type	Category	Post Weekday	Post Hour
0	Photo	2	4	3
1	Status	2	3	10
2	Photo	3	3	3
3	Photo	2	2	10
4	Photo	2	2	3
...
494	Photo	3	7	10
495	Photo	3	7	2
496	Photo	2	5	8
497	Photo	1	5	2
498	Photo	3	4	11

495 rows × 4 columns

Creating the second column subset

```
In [9]: colm_sub2 = data[["Post Hour", "Lifetime Post Consumers"]]
        colm_sub2
```

Out[9]:

	Post Hour	Lifetime Post Consumers
0	3	109
1	10	1361
2	3	113
3	10	790
4	3	410
...
494	10	756
495	2	708
496	8	508
497	2	572
498	11	574

495 rows × 2 columns

Creating the first row subset

```
In [10]: row_sub1 = data.iloc[[1,3,5,7,9,11,12,14,16]]
row_sub1
```

Out[10]:

	Page total likes	Type	Category	Post Month	Post Weekday	Post Hour	Paid	Lifetime Post Total Reach	Lifetime Post Total Impressions	Lifetime Engaged Users	Lifetime Post Consumers	L Co
1	139441	Status	2	12	3	10	0.0	10460	19057	1457	1361	
3	139441	Photo	2	12	2	10	1.0	50128	87991	2211	790	
5	139441	Status	2	12	1	9	0.0	10472	20849	1191	1073	
7	139441	Photo	3	12	7	9	1.0	13720	24137	537	232	
9	139441	Photo	3	12	6	10	0.0	4694	8668	280	183	
11	139441	Photo	2	12	5	10	0.0	3112	5590	208	127	
12	139441	Photo	2	12	5	10	0.0	2847	5133	193	115	
14	138414	Photo	2	12	4	5	1.0	22784	39941	887	337	
16	138414	Photo	3	12	3	3	0.0	1722	2981	163	123	

Creating the second row subset

```
In [11]: row_sub2 = data.iloc[[0,2,4,6,8,10,12,14,16]]
row_sub2
```

Out[11]:

	Page total likes	Type	Category	Post Month	Post Weekday	Post Hour	Paid	Lifetime Post Total Reach	Lifetime Post Total Impressions	Lifetime Engaged Users	Lifetime Post Consumers	L Co
0	139441	Photo	2	12	4	3	0.0	2752	5091	178	109	
2	139441	Photo	3	12	3	3	0.0	2413	4373	177	113	
4	139441	Photo	2	12	2	3	0.0	7244	13594	671	410	
6	139441	Photo	3	12	1	3	1.0	11692	19479	481	265	
8	139441	Status	2	12	7	3	0.0	11844	22538	1530	1407	
10	139441	Status	2	12	5	10	0.0	21744	42334	4258	4100	
12	139441	Photo	2	12	5	10	0.0	2847	5133	193	115	
14	138414	Photo	2	12	4	5	1.0	22784	39941	887	337	
16	138414	Photo	3	12	3	3	0.0	1722	2981	163	123	

Merging two subset

```
In [12]: merg = row_sub1.merge(row_sub2)
merg
```

Out[12]:

	Page total likes	Type	Category	Post Month	Post Weekday	Post Hour	Paid	Lifetime Post Total Reach	Lifetime Post Total Impressions	Lifetime Engaged Users	Lifetime Post Consumers	Lif Cons
0	139441	Photo	2	12	5	10	0.0	2847	5133	193	115	
1	138414	Photo	2	12	4	5	1.0	22784	39941	887	337	
2	138414	Photo	3	12	3	3	0.0	1722	2981	163	123	

Concatenating two subset

In [13]:

```
conc = pd.concat([row_sub1, row_sub2])
conc
```

Out[13]:

	Page total likes	Type	Category	Post Month	Post Weekday	Post Hour	Paid	Lifetime Post Total Reach	Lifetime Post Total Impressions	Lifetime Engaged Users	Lifetime Post Consumers	L Co
1	139441	Status	2	12	3	10	0.0	10460	19057	1457	1361	
3	139441	Photo	2	12	2	10	1.0	50128	87991	2211	790	
5	139441	Status	2	12	1	9	0.0	10472	20849	1191	1073	
7	139441	Photo	3	12	7	9	1.0	13720	24137	537	232	
9	139441	Photo	3	12	6	10	0.0	4694	8668	280	183	
11	139441	Photo	2	12	5	10	0.0	3112	5590	208	127	
12	139441	Photo	2	12	5	10	0.0	2847	5133	193	115	
14	138414	Photo	2	12	4	5	1.0	22784	39941	887	337	
16	138414	Photo	3	12	3	3	0.0	1722	2981	163	123	
0	139441	Photo	2	12	4	3	0.0	2752	5091	178	109	
2	139441	Photo	3	12	3	3	0.0	2413	4373	177	113	
4	139441	Photo	2	12	2	3	0.0	7244	13594	671	410	
6	139441	Photo	3	12	1	3	1.0	11692	19479	481	265	
8	139441	Status	2	12	7	3	0.0	11844	22538	1530	1407	
10	139441	Status	2	12	5	10	0.0	21744	42334	4258	4100	
12	139441	Photo	2	12	5	10	0.0	2847	5133	193	115	
14	138414	Photo	2	12	4	5	1.0	22784	39941	887	337	
16	138414	Photo	3	12	3	3	0.0	1722	2981	163	123	

Sorting the subset

```
In [14]: st = conc.sort_values(by="Category")
st
```

Out[14]:

	Page total likes	Type	Category	Post Month	Post Weekday	Post Hour	Paid	Lifetime Post Total Reach	Lifetime Post Total Impressions	Lifetime Engaged Users	Lifetime Post Consumers	L Co
1	139441	Status	2	12	3	10	0.0	10460	19057	1457	1361	
3	139441	Photo	2	12	2	10	1.0	50128	87991	2211	790	
5	139441	Status	2	12	1	9	0.0	10472	20849	1191	1073	
12	139441	Photo	2	12	5	10	0.0	2847	5133	193	115	
10	139441	Status	2	12	5	10	0.0	21744	42334	4258	4100	
11	139441	Photo	2	12	5	10	0.0	3112	5590	208	127	
12	139441	Photo	2	12	5	10	0.0	2847	5133	193	115	
14	138414	Photo	2	12	4	5	1.0	22784	39941	887	337	
14	138414	Photo	2	12	4	5	1.0	22784	39941	887	337	
0	139441	Photo	2	12	4	3	0.0	2752	5091	178	109	
8	139441	Status	2	12	7	3	0.0	11844	22538	1530	1407	
4	139441	Photo	2	12	2	3	0.0	7244	13594	671	410	
16	138414	Photo	3	12	3	3	0.0	1722	2981	163	123	
2	139441	Photo	3	12	3	3	0.0	2413	4373	177	113	
9	139441	Photo	3	12	6	10	0.0	4694	8668	280	183	
7	139441	Photo	3	12	7	9	1.0	13720	24137	537	232	
6	139441	Photo	3	12	1	3	1.0	11692	19479	481	265	
16	138414	Photo	3	12	3	3	0.0	1722	2981	163	123	

Transposing Data

```
In [15]: trans = st.transpose()
trans
```

Out[15]:

	1	3	5	12	10	11	12	14	14	0	8	
Page total likes	139441	139441	139441	139441	139441	139441	139441	138414	138414	139441	139441	1
Type	Status	Photo	Status	Photo	Status	Photo	Photo	Photo	Photo	Photo	Status	
Category	2	2	2	2	2	2	2	2	2	2	2	
Post Month	12	12	12	12	12	12	12	12	12	12	12	
Post Weekday	3	2	1	5	5	5	5	4	4	4	7	
Post Hour	10	10	9	10	10	10	10	5	5	3	3	
Paid	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	
Lifetime Post Total Reach	10460	50128	10472	2847	21744	3112	2847	22784	22784	2752	11844	
Lifetime Post Total Impressions	19057	87991	20849	5133	42334	5590	5133	39941	39941	5091	22538	
Lifetime Engaged Users	1457	2211	1191	193	4258	208	193	887	887	178	1530	
Lifetime Post Consumers	1361	790	1073	115	4100	127	115	337	337	109	1407	
Lifetime Post Consumptions	1674	1119	1389	133	4540	145	133	417	417	159	1692	
Lifetime Post Impressions by people who have liked your Page	11710	61027	16034	3779	37849	3887	3779	34415	34415	3078	15220	
Lifetime Post reach by people who like your Page	6112	32048	7852	2072	18952	2174	2072	19312	19312	1640	7912	
Lifetime People who have liked your Page and engaged with your post	1108	1386	1016	152	3798	165	152	684	684	119	1250	
comment	5	58	1	0	0	0	0	2	2	4	0	
like	130.0	1572.0	152.0	90.0	233.0	88.0	90.0	577.0	577.0	79.0	161.0	
share	29.0	147.0	33.0	14.0	19.0	18.0	14.0	20.0	20.0	17.0	31.0	
Total Interactions	164	1777	186	104	252	106	104	599	599	100	192	

Reshaping the Data

```
In [17]: reshape = st.melt(id_vars="Type", value_vars=["Category", "Post Hour"])  
reshape
```

Out[17]:

	Type	variable	value
0	Status	Category	2
1	Photo	Category	2
2	Status	Category	2
3	Photo	Category	2
4	Status	Category	2
5	Photo	Category	2
6	Photo	Category	2
7	Photo	Category	2
8	Photo	Category	2
9	Photo	Category	2
10	Status	Category	2
11	Photo	Category	2
12	Photo	Category	3
13	Photo	Category	3
14	Photo	Category	3
15	Photo	Category	3
16	Photo	Category	3
17	Photo	Category	3
18	Status	Post Hour	10
19	Photo	Post Hour	10
20	Status	Post Hour	9
21	Photo	Post Hour	10
22	Status	Post Hour	10
23	Photo	Post Hour	10
24	Photo	Post Hour	10
25	Photo	Post Hour	5
26	Photo	Post Hour	5
27	Photo	Post Hour	3
28	Status	Post Hour	3
29	Photo	Post Hour	3
30	Photo	Post Hour	3
31	Photo	Post Hour	3
32	Photo	Post Hour	10
33	Photo	Post Hour	9
34	Photo	Post Hour	3
35	Photo	Post Hour	3

