

```
In [13]: import pandas as pd  
import numpy as np
```

```
In [14]: df=pd.read_csv('../Desktop/DS/dataset_Facebook.csv')
```

```
In [15]: df.head()
```

```
Out[15]: 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;Lifetime Post Consumptions;Lifetime Post Impressions by people who have liked your Page;Lifetime Post reach by people who like your Page;Lifetime People who have liked your Page and engaged with your post;comment;like;share;Total Interactions
```

0	139441;Photo;2;12;4;3;0;2752;5091;178;109;159;...
1	139441;Status;2;12;3;10;0;10460;19057;1457;136...
2	139441;Photo;3;12;3;3;0;2413;4373;177;113;154;...
3	139441;Photo;2;12;2;10;1;50128;87991;2211;790;...
4	139441;Photo;2;12;2;3;0;7244;13594;671;410;580;...

```
In [17]: df=pd.read_csv('../Desktop/DS/dataset_Facebook.csv',sep=';')  
df
```

Out[17]:

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

500 rows × 19 columns

In [18]: df.head()	

Out[18]:

	Page total likes	Type	Category	Post Month	Post Weekday	Post Hour	Paid	Lifetime Post Total Reach	Lifetime Post Total Impressions	Lifetime Engaged Users	Lifetime P Consum
<b>0</b>	139441	Photo		2	12	4	3	0.0	2752	5091	178
<b>1</b>	139441	Status		2	12	3	10	0.0	10460	19057	1457
<b>2</b>	139441	Photo		3	12	3	3	0.0	2413	4373	177
<b>3</b>	139441	Photo		2	12	2	10	1.0	50128	87991	2211
<b>4</b>	139441	Photo		2	12	2	3	0.0	7244	13594	671

In [19]: df.tail()

Out[19]:

	Page total likes	Type	Category	Post Month	Post Weekday	Post Hour	Paid	Lifetime Post Total Reach	Lifetime Post Total Impressions	Lifetime Engaged Users	Lifetime Consum
<b>495</b>	85093	Photo		3	1	7	2	0.0	4684	7536	733
<b>496</b>	81370	Photo		2	1	5	8	0.0	3480	6229	537
<b>497</b>	81370	Photo		1	1	5	2	0.0	3778	7216	625
<b>498</b>	81370	Photo		3	1	4	11	0.0	4156	7564	626
<b>499</b>	81370	Photo		2	1	4	4	NaN	4188	7292	564

In [20]: df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 19 columns):
 #   Column           Non-Null Co
 0   Page total likes    500 non-nul
 1   int64
 1   Type             500 non-nul
 1   object
 2   Category         500 non-nul
 1   int64
 3   Post Month       500 non-nul
 1   int64
 4   Post Weekday     500 non-nul
 1   int64
 5   Post Hour        500 non-nul
 1   int64
 6   Paid             499 non-nul
 1   float64
 7   Lifetime Post Total Reach 500 non-nul
 1   int64
 8   Lifetime Post Total Impressions 500 non-nul
 1   int64
 9   Lifetime Engaged Users 500 non-nul
 1   int64
 10  Lifetime Post Consumers 500 non-nul
 1   int64
 11  Lifetime Post Consumptions 500 non-nul
 1   int64
 12  Lifetime Post Impressions by people who have liked your Page 500 non-nul
 1   int64
 13  Lifetime Post reach by people who like your Page 500 non-nul
 1   int64
 14  Lifetime People who have liked your Page and engaged with your post 500 non-nul
 1   int64
 15  comment          500 non-nul
 1   int64
 16  like              499 non-nul
 1   float64
 17  share             496 non-nul
 1   float64
 18  Total Interactions 500 non-nul
 1   int64
dtypes: float64(3), int64(15), object(1)
memory usage: 74.3+ KB
```

In [21]: df.describe()

Out[21]:

	Page total likes	Category	Post Month	Post Weekday	Post Hour	Paid	Lifetime Post Total Reach	Lifetime Imp
<b>count</b>	500.000000	500.000000	500.000000	500.000000	500.000000	499.000000	500.000000	5.000000
<b>mean</b>	123194.176000	1.880000	7.038000	4.150000	7.840000	0.278557	13903.360000	2.958000
<b>std</b>	16272.813214	0.852675	3.307936	2.030701	4.368589	0.448739	22740.78789	7.680000
<b>min</b>	81370.000000	1.000000	1.000000	1.000000	1.000000	0.000000	238.000000	5.700000
<b>25%</b>	112676.000000	1.000000	4.000000	2.000000	3.000000	0.000000	3315.000000	5.694000
<b>50%</b>	129600.000000	2.000000	7.000000	4.000000	9.000000	0.000000	5281.000000	9.051000
<b>75%</b>	136393.000000	3.000000	10.000000	6.000000	11.000000	1.000000	13168.000000	2.208000
<b>max</b>	139441.000000	3.000000	12.000000	7.000000	23.000000	1.000000	180480.000000	1.110000

In [22]: df.shape

Out[22]: (500, 19)

```
#Creating Subset----->
#Subset 1:
df1=df[['Page total likes','Post Month','Post Weekday']].loc[0:10]
df1
```

Out[27]:

	Page total likes	Post Month	Post Weekday
<b>0</b>	139441	12	4
<b>1</b>	139441	12	3
<b>2</b>	139441	12	3
<b>3</b>	139441	12	2
<b>4</b>	139441	12	2
<b>5</b>	139441	12	1
<b>6</b>	139441	12	1
<b>7</b>	139441	12	7
<b>8</b>	139441	12	7
<b>9</b>	139441	12	6
<b>10</b>	139441	12	5

In [28]: #Subset 2:

```
df2=df[['Page total likes','Category','Post Hour','Paid']].loc[16:26]
df2
```

Out[28]:

	Page total likes	Category	Post Hour	Paid
16	138414	3	3	0.0
17	138414	1	12	1.0
18	138414	3	3	0.0
19	138414	3	11	0.0
20	138414	2	3	0.0
21	138414	1	10	0.0
22	138414	1	10	0.0
23	138414	3	3	0.0
24	138414	2	10	0.0
25	138458	2	3	0.0
26	138458	2	11	0.0

In [29]:

```
#Subset:
df3=df[['Post Month','Lifetime Post Consumers','like']].loc[51:61]
df3
```

Out[29]:

	Post Month	Lifetime Post Consumers	like
51	11	467	285.0
52	11	181	202.0
53	11	182	64.0
54	11	213	66.0
55	11	411	65.0
56	11	204	164.0
57	11	131	40.0
58	11	173	76.0
59	11	304	139.0
60	11	335	101.0
61	11	513	144.0

In [31]:

```
#Merge Data----->
#Merge Subset1,2,3/all:
merging=pd.concat([df1,df2,df3])
merging
```

Out[31]:

	Page total likes	Post Month	Post Weekday	Category	Post Hour	Paid	Lifetime Post Consumers	like
0	139441.0	12.0	4.0	NaN	NaN	NaN	NaN	NaN
1	139441.0	12.0	3.0	NaN	NaN	NaN	NaN	NaN
2	139441.0	12.0	3.0	NaN	NaN	NaN	NaN	NaN
3	139441.0	12.0	2.0	NaN	NaN	NaN	NaN	NaN
4	139441.0	12.0	2.0	NaN	NaN	NaN	NaN	NaN
5	139441.0	12.0	1.0	NaN	NaN	NaN	NaN	NaN
6	139441.0	12.0	1.0	NaN	NaN	NaN	NaN	NaN
7	139441.0	12.0	7.0	NaN	NaN	NaN	NaN	NaN
8	139441.0	12.0	7.0	NaN	NaN	NaN	NaN	NaN
9	139441.0	12.0	6.0	NaN	NaN	NaN	NaN	NaN
10	139441.0	12.0	5.0	NaN	NaN	NaN	NaN	NaN
16	138414.0	NaN	NaN	3.0	3.0	0.0	NaN	NaN
17	138414.0	NaN	NaN	1.0	12.0	1.0	NaN	NaN
18	138414.0	NaN	NaN	3.0	3.0	0.0	NaN	NaN
19	138414.0	NaN	NaN	3.0	11.0	0.0	NaN	NaN
20	138414.0	NaN	NaN	2.0	3.0	0.0	NaN	NaN
21	138414.0	NaN	NaN	1.0	10.0	0.0	NaN	NaN
22	138414.0	NaN	NaN	1.0	10.0	0.0	NaN	NaN
23	138414.0	NaN	NaN	3.0	3.0	0.0	NaN	NaN
24	138414.0	NaN	NaN	2.0	10.0	0.0	NaN	NaN
25	138458.0	NaN	NaN	2.0	3.0	0.0	NaN	NaN
26	138458.0	NaN	NaN	2.0	11.0	0.0	NaN	NaN
51	NaN	11.0	NaN	NaN	NaN	NaN	467.0	285.0
52	NaN	11.0	NaN	NaN	NaN	NaN	181.0	202.0
53	NaN	11.0	NaN	NaN	NaN	NaN	182.0	64.0
54	NaN	11.0	NaN	NaN	NaN	NaN	213.0	66.0
55	NaN	11.0	NaN	NaN	NaN	NaN	411.0	65.0
56	NaN	11.0	NaN	NaN	NaN	NaN	204.0	164.0
57	NaN	11.0	NaN	NaN	NaN	NaN	131.0	40.0
58	NaN	11.0	NaN	NaN	NaN	NaN	173.0	76.0
59	NaN	11.0	NaN	NaN	NaN	NaN	304.0	139.0
60	NaN	11.0	NaN	NaN	NaN	NaN	335.0	101.0
61	NaN	11.0	NaN	NaN	NaN	NaN	513.0	144.0

```
In [33]: #Sort Data----->
#Sort 1:
Sort_Values1=df.sort_values('Page total likes',ascending=False)
Sort_Values1
```

Out[33]:

Page total likes	Type	Category	Post Month	Post Weekday	Post Hour	Paid	Lifetime Post Total Reach	Lifetime Post Total Impressions	Lifetime Engaged Users	Lif
0	139441	Photo	2	12	4	3	0.0	2752	5091	178
8	139441	Status	2	12	7	3	0.0	11844	22538	1530
1	139441	Status	2	12	3	10	0.0	10460	19057	1457
12	139441	Photo	2	12	5	10	0.0	2847	5133	193
11	139441	Photo	2	12	5	10	0.0	3112	5590	208
...	...	...	...	...	...	...	...	...	...	...
495	85093	Photo	3	1	7	2	0.0	4684	7536	733
496	81370	Photo	2	1	5	8	0.0	3480	6229	537
497	81370	Photo	1	1	5	2	0.0	3778	7216	625
498	81370	Photo	3	1	4	11	0.0	4156	7564	626
499	81370	Photo	2	1	4	4	NaN	4188	7292	564

500 rows × 19 columns

```
In [34]: #Sort 2;
Sort_Values2=df.sort_values(['Page total likes','Paid'],ascending=False)
Sort_Values2
```

Out[34]:

Page total likes	Type	Category	Post Month	Post Weekday	Post Hour	Paid	Lifetime Post Total Reach	Lifetime Post Total Impressions	Lifetime Engaged Users	Life
3 139441	Photo		2	12	2	10	1.0	50128	87991	2211
6 139441	Photo		3	12	1	3	1.0	11692	19479	481
7 139441	Photo		3	12	7	9	1.0	13720	24137	537
0 139441	Photo		2	12	4	3	0.0	2752	5091	178
1 139441	Status		2	12	3	10	0.0	10460	19057	1457
...	...	...	...	...	...	...	...	...	...	...
495 85093	Photo		3	1	7	2	0.0	4684	7536	733
496 81370	Photo		2	1	5	8	0.0	3480	6229	537
497 81370	Photo		1	1	5	2	0.0	3778	7216	625
498 81370	Photo		3	1	4	11	0.0	4156	7564	626
499 81370	Photo		2	1	4	4	NaN	4188	7292	564

500 rows × 19 columns



In [35]: #Transposing Data----->  
df.transpose()

Out[35]:

	0	1	2	3	4	5	6	7	8	9	...
<b>Page total likes</b>	139441	139441	139441	139441	139441	139441	139441	139441	139441	139441	...
<b>Type</b>	Photo	Status	Photo	Photo	Photo	Status	Photo	Photo	Status	Photo	...
<b>Category</b>	2	2	3	2	2	2	3	3	2	3	...
<b>Post Month</b>	12	12	12	12	12	12	12	12	12	12	...
<b>Post Weekday</b>	4	3	3	2	2	1	1	7	7	6	...
<b>Post Hour</b>	3	10	3	10	3	9	3	9	3	10	...
<b>Paid</b>	0.0	0.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	...
<b>Lifetime Post Total Reach</b>	2752	10460	2413	50128	7244	10472	11692	13720	11844	4694	...
<b>Lifetime Post Total Impressions</b>	5091	19057	4373	87991	13594	20849	19479	24137	22538	8668	...
<b>Lifetime Engaged Users</b>	178	1457	177	2211	671	1191	481	537	1530	280	...
<b>Lifetime Post Consumers</b>	109	1361	113	790	410	1073	265	232	1407	183	...
<b>Lifetime Post Consumptions</b>	159	1674	154	1119	580	1389	364	305	1692	250	...
<b>Lifetime Post Impressions by people who have liked your Page</b>	3078	11710	2812	61027	6228	16034	15432	19728	15220	4309	...
<b>Lifetime Post reach by people who like your Page</b>	1640	6112	1503	32048	3200	7852	9328	11056	7912	2324	...
<b>Lifetime People who have liked your Page and engaged with your post</b>	119	1108	132	1386	396	1016	379	422	1250	199	...
<b>comment</b>	4	5	0	58	19	1	3	0	0	3	...
<b>like</b>	79.0	130.0	66.0	1572.0	325.0	152.0	249.0	325.0	161.0	113.0	...
<b>share</b>	17.0	29.0	14.0	147.0	49.0	33.0	27.0	14.0	31.0	26.0	...
<b>Total Interactions</b>	100	164	80	1777	393	186	279	339	192	142	...

19 rows × 500 columns

```
In [36]: #Shape and Reshape Data----->
#Shaping:
Shaping=df.shape
Shaping
```

```
Out[36]: (500, 19)
```

```
In [37]: #Reshaping:
Pivot_table1=pd.pivot_table(df,index=['Type','Category'],values=['comment','like'])
Pivot_table1
```

```
Out[37]:
```

Type	Category	comment	like
<b>Link</b>	<b>1</b>	2.900000	75.650000
	<b>2</b>	2.000000	32.000000
	<b>3</b>	2.000000	68.000000
<b>Photo</b>	<b>1</b>	5.897297	126.000000
	<b>2</b>	11.692308	235.857143
	<b>3</b>	6.913333	219.753333
<b>Status</b>	<b>1</b>	4.333333	136.333333
	<b>2</b>	9.921053	182.552632
	<b>3</b>	2.750000	151.500000
<b>Video</b>	<b>1</b>	12.285714	231.428571

```
In [39]: #Extra Information----->
arr=np.array([1,2,3,4,5,6,7,8,9,10])
arr.reshape(5,2)
```

```
Out[39]: array([[ 1,  2],
 [ 3,  4],
 [ 5,  6],
 [ 7,  8],
 [ 9, 10]])
```

```
In [ ]:
```