Assignment 2a - DSBDA

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

- Create data subsets
- Merge Data
- Sort Data
- Transposing Data
- Shape and Reshape Data

Import python libraries and reading the data

In [3]:
```
import pandas as pd
import numpy as nm
df = pd.read_csv("dataset_Facebook.csv")
df
```

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	Lifetime Post Consumptions	Lifetime Post Impressions by people who have liked your Page	Lifetime Post reach by people who like your Page	l	e
0	139441	Photo	2	12	4	3	0.0	2752	5091	178	109	159	3078	1640		
1	139441	Status	2	12	3	10	0.0	10460	19057	1457	1361	1674	11710	6112		
2	139441	Photo	3	12	3	3	0.0	2413	4373	177	113	154	2812	1503		
3	139441	Photo	2	12	2	10	1.0	50128	87991	2211	790	1119	61027	32048		
4	139441	Photo	2	12	2	3	0.0	7244	13594	671	410	580	6228	3200		
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...		
495	85093	Photo	3	1	7	2	0.0	4684	7536	733	708	985	4750	2876		
496	81370	Photo	2	1	5	8	0.0	3480	6229	537	508	687	3961	2104		
497	81370	Photo	1	1	5	2	0.0	3778	7216	625	572	795	4742	2388		
498	81370	Photo	3	1	4	11	0.0	4156	7564	626	574	832	4534	2452		
499	81370	Photo	2	1	4	4	NaN	4188	7292	564	524	743	3861	2200		

500 rows × 19 columns

In [4]:
```
df.describe()
```

Out[4]:

	Page total likes	Category	Post Month	Post Weekday	Post Hour	Paid	Lifetime Post Total Reach	Lifetime Post Total Impressions	Lifetime Engaged Users	Lifetime Post Consumers	Lifet Consu
count	500.000000	500.000000	500.000000	500.000000	500.000000	499.000000	500.00000	5.000000e+02	500.000000	500.000000	50
mean	123194.176000	1.880000	7.038000	4.150000	7.840000	0.278557	13903.36000	2.958595e+04	920.344000	798.772000	141
std	16272.813214	0.852675	3.307936	2.030701	4.368589	0.448739	22740.78789	7.680325e+04	985.016636	882.505013	200
min	81370.000000	1.000000	1.000000	1.000000	1.000000	0.000000	238.00000	5.700000e+02	9.000000	9.000000	
25%	112676.000000	1.000000	4.000000	2.000000	3.000000	0.000000	3315.00000	5.694750e+03	393.750000	332.500000	50
50%	129600.000000	2.000000	7.000000	4.000000	9.000000	0.000000	5281.00000	9.051000e+03	625.500000	551.500000	85
75%	136393.000000	3.000000	10.000000	6.000000	11.000000	1.000000	13168.00000	2.208550e+04	1062.000000	955.500000	146
max	139441.000000	3.000000	12.000000	7.000000	23.000000	1.000000	180480.00000	1.110282e+06	11452.000000	11328.000000	1977

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

Out[5]:
```
(500, 19)
```

In [20]:
```
df.isnull().sum()
```

Out[20]:
```
Page total likes                                                  0
Type                                                              0
Category                                                          0
Post Month                                                        0
Post Weekday                                                      0
Post Hour                                                         0
Paid                                                              1
Lifetime Post Total Reach                                         0
Lifetime Post Total Impressions                                   0
Lifetime Engaged Users                                            0
Lifetime Post Consumers                                           0
Lifetime Post Consumptions                                        0
Lifetime Post Impressions by people who have liked your Page      0
Lifetime Post reach by people who like your Page                  0
Lifetime People who have liked your Page and engaged with your post    0
comment                                                           0
like                                                              1
share                                                             4
Total Interactions                                                0
dtype: int64
```

Creating Subset

In [6]:
```
df1 = df[['Page total likes', 'Category', 'Post Month', 'Post Weekday']].loc[0:15]
df1
```

Out[6]:

	Page total likes	Category	Post Month	Post Weekday
0	139441	2	12	4
1	139441	2	12	3
2	139441	3	12	3
3	139441	2	12	2
4	139441	2	12	2
5	139441	2	12	1
6	139441	3	12	1
7	139441	3	12	7
8	139441	2	12	7
9	139441	3	12	6
10	139441	2	12	5
11	139441	2	12	5
12	139441	2	12	5
13	139441	2	12	5
14	138414	2	12	4
15	138414	2	12	3

In [7]:
```
df2 = df[['Page total likes', 'Category', 'Post Month', 'Post Weekday']].loc[16:30]
df2
```

Out[7]:

	Page total likes	Category	Post Month	Post Weekday
16	138414	3	12	3
17	138414	1	12	2
18	138414	3	12	2
19	138414	3	12	1
20	138414	2	12	1
21	138414	1	12	7
22	138414	1	12	7
23	138414	3	12	7
24	138414	2	12	6
25	138458	2	12	6
26	138458	2	12	5
27	138458	3	12	5
28	138895	2	12	5
29	138895	1	12	4
30	138895	2	12	4

In [8]:
```
df3 = df[['Page total likes', 'Category', 'Post Month', 'Post Weekday']].loc[31:50]
df3
```

Out[8]:

	Page total likes	Category	Post Month	Post Weekday
31	138895	2	12	3
32	138895	3	12	3
33	138895	3	12	2
34	138895	1	12	2
35	138895	2	12	1
36	138895	3	12	1
37	138895	1	12	7
38	138895	2	12	7
39	138895	1	12	7
40	138895	2	12	6
41	138895	1	12	6
42	138353	1	12	5
43	138353	1	12	5
44	138353	1	12	4
45	138353	1	12	4
46	138353	1	12	3
47	138353	1	12	3
48	138353	1	12	2
49	138353	1	12	2
50	138353	2	11	1

Merge Data

In [9]:
```
merging = pd.concat([df1, df2, df3])
merging
```

Out[9]:

	Page total likes	Category	Post Month	Post Weekday
0	139441	2	12	4
1	139441	2	12	3
2	139441	3	12	3
3	139441	2	12	2
4	139441	2	12	2
5	139441	2	12	1
6	139441	3	12	1
7	139441	3	12	7
8	139441	2	12	7
9	139441	3	12	6
10	139441	2	12	5
11	139441	2	12	5
12	139441	2	12	5
13	139441	2	12	5
14	138414	2	12	4
15	138414	2	12	3
16	138414	3	12	3
17	138414	1	12	2
18	138414	3	12	2
19	138414	3	12	1
20	138414	2	12	1
21	138414	1	12	7
22	138414	1	12	7
23	138414	3	12	7
24	138414	2	12	6
25	138458	2	12	6

	Page total likes	Category	Post Month	Post Weekday
26	138458	2	12	5
27	138458	3	12	5
28	138895	2	12	5
29	138895	1	12	4
30	138895	2	12	4
31	138895	2	12	3
32	138895	3	12	3
33	138895	3	12	2
34	138895	1	12	2
35	138895	2	12	1
36	138895	3	12	1
37	138895	1	12	7
38	138895	2	12	7
39	138895	1	12	7
40	138895	2	12	6
41	138895	1	12	6
42	138353	1	12	5
43	138353	1	12	5
44	138353	1	12	4
45	138353	1	12	4
46	138353	1	12	3
47	138353	1	12	3
48	138353	1	12	2
49	138353	1	12	2
50	138353	2	11	1

Sort Data

Sort the 'Page total likes' and 'Lifetime Post Total Reach'

In [10]:
```
sort_values = df.sort_values('Page total likes', ascending=False)
sort_values
```

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	Lifetime Post Consumptions	Lifetime Post Impressions by people who have liked your Page	Lifetime Post reach by people who like your Page	l	e
0	139441	Photo	2	12	4	3	0.0	2752	5091	178	109	159	3078	1640		
8	139441	Status	2	12	7	3	0.0	11844	22538	1530	1407	1692	15220	7912		
1	139441	Status	2	12	3	10	0.0	10460	19057	1457	1361	1674	11710	6112		
12	139441	Photo	2	12	5	10	0.0	2847	5133	193	115	133	3779	2072		
11	139441	Photo	2	12	5	10	0.0	3112	5590	208	127	145	3887	2174		
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...		
495	85093	Photo	3	1	7	2	0.0	4684	7536	733	708	985	4750	2876		
496	81370	Photo	2	1	5	8	0.0	3480	6229	537	508	687	3961	2104		
497	81370	Photo	1	1	5	2	0.0	3778	7216	625	572	795	4742	2388		
498	81370	Photo	3	1	4	11	0.0	4156	7564	626	574	832	4534	2452		
499	81370	Photo	2	1	4	4	NaN	4188	7292	564	524	743	3861	2200		

500 rows × 19 columns

In [19]:
```
sort_values = df.sort_values('Lifetime Post Total Reach', ascending=True)
sort_values
```

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 Post Consumers	Lifetime Post Consumptions	Lifetime Post Impressions by people who have liked your Page	Lifetime Post reach by people who like your Page	L
422	102112	Photo	1	3	1	19	0.0	238	570	143	142	834	567	236	
426	100732	Photo	1	3	7	18	0.0	391	746	131	130	766	723	380	
432	100732	Photo	1	3	6	17	0.0	452	726	186	184	889	721	450	
120	136393	Photo	1	10	7	9	0.0	584	1029	273	271	308	943	511	
123	136393	Photo	1	10	7	7	0.0	617	1071	229	223	265	935	521	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
380	111620	Photo	1	4	7	14	0.0	128064	251269	1539	1408	2506	84046	32384	
277	126424	Video	1	6	2	13	0.0	139008	277100	1779	1643	2356	107502	38720	
463	92186	Photo	3	2	7	2	1.0	153536	497910	1713	1633	2493	8907	5696	
464	92079	Photo	1	2	6	13	0.0	158208	453213	2482	2319	3412	81938	20608	
244	130791	Photo	2	7	3	5	1.0	180480	319133	8072	4010	6242	108752	51456	

500 rows × 19 columns

Transpose Data

In [11]:
```
df.transpose()
```

Out[11]:

	0	1	2	3	4	5	6	7	8	9	...	490	491	492	493	494	495
Page total likes	139441	139441	139441	139441	139441	139441	139441	139441	139441	139441	...	85979	85979	85979	85093	85093	85093
Type	Photo	Status	Photo	Photo	Photo	Status	Photo	Photo	Status	Photo	...	Photo	Photo	Link	Photo	Photo	Photo
Category	2	2	3	2	2	2	3	3	2	3	...	3	3	1	3	3	3
Post Month	12	12	12	12	12	12	12	12	12	12	...	1	1	1	1	1	1
Post Weekday	4	3	3	2	2	1	1	7	7	6	...	6	6	5	1	7	7
Post Hour	3	10	3	10	3	9	3	9	3	10	...	11	3	11	2	10	2
Paid	0.0	0.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	...	0.0	1.0	0.0	0.0	0.0	0.0
Lifetime Post Total Reach	2752	10460	2413	50128	7244	10472	11692	13720	11844	4694	...	5280	6184	45920	8412	5400	4684
Lifetime Post Total Impressions	5091	19057	4373	87991	13594	20849	19479	24137	22538	8668	...	8703	10228	5808	13960	9218	7536
Lifetime Engaged Users	178	1457	177	2211	671	1191	481	537	1530	280	...	951	956	753	1179	810	733
Lifetime Post Consumers	109	1361	113	790	410	1073	265	232	1407	183	...	911	901	655	1111	756	708
Lifetime Post Consumptions	159	1674	154	1119	580	1389	364	305	1692	250	...	1237	1140	763	1632	1003	985
Lifetime Post Impressions by people who have liked your Page	3078	11710	2812	61027	6228	16034	15432	19728	15220	4309	...	5757	6085	15766	8632	5654	4750
Lifetime Post reach by people who like your Page	1640	6112	1503	32048	3200	7852	9328	11056	7912	2324	...	3300	3502	10720	5348	3230	2876
Lifetime People who have liked your Page and engaged with your post	119	1108	132	1386	396	1016	379	422	1250	199	...	431	437	220	699	422	392

	0	1	2	3	4	5	6	7	8	9	...	490	491	492	493	494	495
comment	4	5	0	58	19	1	3	0	0	3	...	1	1	0	17	10	5
like	79.0	130.0	66.0	1572.0	325.0	152.0	249.0	325.0	161.0	113.0	...	79.0	105.0	128.0	185.0	125.0	53.0
share	17.0	29.0	14.0	147.0	49.0	33.0	27.0	14.0	31.0	26.0	...	30.0	46.0	9.0	55.0	41.0	26.0
Total Interactions	100	164	80	1777	393	186	279	339	192	142	...	110	152	137	257	176	84

19 rows × 500 columns

Shape and Reshape Data

In [12]:
```
shaping = df.shape
shaping
```

Out[12]:
```
(500, 19)
```

In [13]:
```
pivot_table = pd.pivot_table(df, index = ['Type', 'Category'], values = 'comment')
print(pivot_table)
```

```
                 comment
Type   Category
Link   1        2.900000
       2        2.000000
       3        2.000000
Photo  1        5.897297
       2       11.692308
       3        6.913333
Status 1        4.333333
       2        9.921053
       3        2.750000
Video  1       12.285714
```

In [14]:
```
reshaping_arr = nm.array([1,2,3,4,5,6])
reshaping_arr.reshape(3,2)
```

Out[14]:
```
array([[1, 2],
       [3, 4],
       [5, 6]])
```