Assignment 4

Author: Rohit Pendse

Roll No: 33358 ¶

Problem Statement

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

1. Create data subsets
2. Merge Data
3. Sort Data
4. Transposing Data
5. Shape and reshape Data

Importing required libraries

In [2]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
sns.set()
%matplotlib inline
```

In [13]:
```
# Reading the dataset
dataset = pd.read_csv(r'D:\ROHIT\TE Assignments SEM II\DSBDAL\33358_Rohit\Assignment4\Fa
dataset.head()
```

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	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	

In [14]:
```
# Shape of the dataset
dataset.shape
```

Out[14]:
```
(500, 19)
```

In [15]:
```
dataset.describe(include="all")
```

Out[15]:

	Page total likes	Type	Category	Post Month	Post Weekday	Post Hour	Paid	Lifetime Post Total Reach	L
count	500.000000	500	500.000000	500.000000	500.000000	500.000000	499.000000	500.00000	5
unique	NaN	4	NaN	NaN	NaN	NaN	NaN	NaN	
top	NaN	Photo	NaN	NaN	NaN	NaN	NaN	NaN	
freq	NaN	426	NaN	NaN	NaN	NaN	NaN	NaN	
mean	123194.176000	NaN	1.880000	7.038000	4.150000	7.840000	0.278557	13903.36000	2
std	16272.813214	NaN	0.852675	3.307936	2.030701	4.368589	0.448739	22740.78789	7
min	81370.000000	NaN	1.000000	1.000000	1.000000	1.000000	0.000000	238.00000	5
25%	112676.000000	NaN	1.000000	4.000000	2.000000	3.000000	0.000000	3315.00000	5
50%	129600.000000	NaN	2.000000	7.000000	4.000000	9.000000	0.000000	5281.00000	9
75%	136393.000000	NaN	3.000000	10.000000	6.000000	11.000000	1.000000	13168.00000	2
max	139441.000000	NaN	3.000000	12.000000	7.000000	23.000000	1.000000	180480.00000	1

In [16]:
```
dataset.dtypes
```

Out[16]:
```
Page total likes                                                    int64
Type                                                               object
Category                                                            int64
Post Month                                                          int64
Post Weekday                                                        int64
Post Hour                                                           int64
Paid                                                              float64
Lifetime Post Total Reach                                           int64
Lifetime Post Total Impressions                                     int64
Lifetime Engaged Users                                              int64
Lifetime Post Consumers                                             int64
Lifetime Post Consumptions                                          int64
Lifetime Post Impressions by people who have liked your Page        int64
Lifetime Post reach by people who like your Page                    int64
Lifetime People who have liked your Page and engaged with your post int64
comment                                                             int64
like                                                              float64
share                                                             float64
Total Interactions                                                  int64
dtype: object
```

Preprocessing

In [17]:
```
# checking for null values
dataset.isnull().sum()
```

Out[17]:
```
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
```

In [18]:
```
# Dropping rows with null values
dataset = dataset.dropna()
dataset.shape
```

Out[18]:
```
(495, 19)
```

In [19]:
```
# Testing data for null values
dataset.isnull().sum()
```

Out[19]:
```
Page total likes                                                    0
Type                                                                0
Category                                                            0
Post Month                                                          0
Post Weekday                                                        0
Post Hour                                                           0
Paid                                                                0
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                                                                0
share                                                               0
Total Interactions                                                  0
dtype: int64
```

Generating subsets on the basis of type

In [20]:
```
# identifying unique values for column "Type"
dataset["Type"].unique()
```

Out[20]:
```
array(['Photo', 'Status', 'Link', 'Video'], dtype=object)
```

Generating subsets

In [21]:
```
photo_subset = dataset[dataset["Type"] == "Photo"]
status_subset = dataset[dataset["Type"] == "Status"]
link_subset = dataset[dataset["Type"] == "Link"]
video_subset = dataset[dataset["Type"] == "Video"]
```

Shape of subsets

In [22]:
```
print("Photo Subset shape  : ", photo_subset.shape)
print("Status Subset shape : ", status_subset.shape)
print("Link Subset shape   : ", link_subset.shape)
print("Video Subset shape  : ", video_subset.shape)
```

```
Photo Subset shape  :  (421, 19)
Status Subset shape :  (45, 19)
Link Subset shape   :  (22, 19)
Video Subset shape  :  (7, 19)
```

Analysing Photos subset

In [23]:
```
# Statistical description of numerical subset
photo_subset.describe(include="all")
```

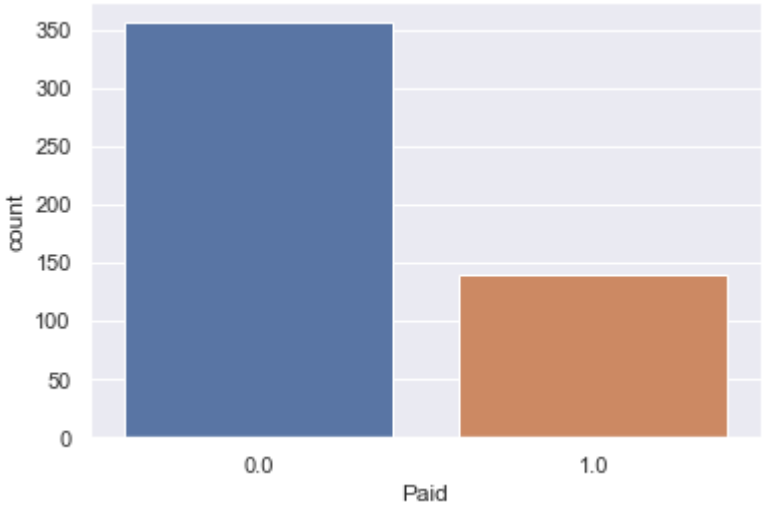
Out[23]:

	Page total likes	Type	Category	Post Month	Post Weekday	Post Hour	Paid	Lifetime Post Total Reach
count	421.000000	421	421.000000	421.000000	421.000000	421.000000	421.000000	421.000000
unique	NaN	1	NaN	NaN	NaN	NaN	NaN	NaN
top	NaN	Photo	NaN	NaN	NaN	NaN	NaN	NaN
freq	NaN	421	NaN	NaN	NaN	NaN	NaN	NaN
mean	122319.612827	NaN	1.926366	6.790974	4.087886	8.004751	0.282660	13275.389549
std	16242.669134	NaN	0.884681	3.228447	2.056203	4.432561	0.450828	22977.950816
min	81370.000000	NaN	1.000000	1.000000	1.000000	1.000000	0.000000	238.000000
25%	109670.000000	NaN	1.000000	4.000000	2.000000	3.000000	0.000000	3110.000000
50%	128032.000000	NaN	2.000000	7.000000	4.000000	9.000000	0.000000	4708.000000
75%	136013.000000	NaN	3.000000	10.000000	6.000000	11.000000	1.000000	10844.000000
max	139441.000000	NaN	3.000000	12.000000	7.000000	23.000000	1.000000	180480.000000

In [24]:
```
photo_subset["Paid"].unique()
```

Out[24]:
```
array([0., 1.])
```

In [25]:
```
# Counting number of paid and unpaid posts
sns.countplot(x=dataset["Paid"])
plt.show()
```



Transpose of data

In [26]:
```
# Shape of data before transposing
print("Shape of Video subset : ", video_subset.shape)
```

```
Shape of Video subset :  (7, 19)
```

In [27]:
```
# Transposing data
video_subset_transpose = video_subset.transpose()
```

In [28]:
```
# Shape of data after transposing
print("Shape of Video subset transpose: ", video_subset_transpose.shape)
```

```
Shape of Video subset transpose:  (19, 7)
```

In [29]:
```
video_subset_transpose
```

Out[29]:

	29	55	71	74	183	243	277
Page total likes	138895	138329	137893	137893	134879	130791	126424
Type	Video	Video	Video	Video	Video	Video	Video
Category	1	1	1	1	1	1	1
Post Month	12	11	11	11	9	7	6
Post Weekday	4	6	5	3	2	3	2
Post Hour	11	2	3	11	10	11	13
Paid	1.0	1.0	1.0	0.0	0.0	1.0	0.0
Lifetime Post Total Reach	36208	16416	100768	13544	30624	21872	139008
Lifetime Post Total Impressions	61262	31950	220447	30235	56950	40413	277100
Lifetime Engaged Users	1141	459	2101	517	2080	3872	1779
Lifetime Post Consumers	1068	411	1735	458	1956	3822	1643
Lifetime Post Consumptions	1728	539	2331	667	3253	7327	2356
Lifetime Post Impressions by people who have liked your Page	30131	21436	59658	26622	32033	24667	107502
Lifetime Post reach by people who like your Page	14112	9568	18880	11760	15744	12920	38720
Lifetime People who have liked your Page and engaged with your post	559	363	885	447	1376	2218	1008
comment	18	2	17	2	6	18	23
like	143.0	65.0	449.0	99.0	345.0	315.0	204.0
share	13.0	14.0	84.0	13.0	121.0	76.0	44.0
Total Interactions	174	81	550	114	472	409	271

Merging data

In [30]:
```
print("Shape of photo subset : ", photo_subset.shape)
print("Shape of video subset : ", video_subset.shape)
```

```
Shape of photo subset :  (421, 19)
Shape of video subset :  (7, 19)
```

In [31]:
```
# Checking columns of both data subsets
print("Columns of photo subset : ", photo_subset.columns)
print("Columns of video subset : ", video_subset.columns)
```

```
Columns of photo subset :  Index(['Page total likes', 'Type', 'Category', 'Post Mont
h', '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'],
      dtype='object')
Columns of video subset :  Index(['Page total likes', 'Type', 'Category', 'Post Mont
h', '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'],
      dtype='object')
```

In [32]:
```
# Merging the 2 subsets (DataFrames)
photo_video_merged = pd.merge(
    left=photo_subset,
    right=video_subset,
    on="Paid"
)
```

In [33]:
```
photo_video_merged.head()
```

Out[33]:

	Page total likes_x	Type_x	Category_x	Post Month_x	Post Weekday_x	Post Hour_x	Paid	Lifetime Post Total Reach_x	Lifetime Post Total Impressions_x	Lifetime Engaged Users_x	...
0	139441	Photo	2	12	4	3	0.0	2752	5091	178	...
1	139441	Photo	2	12	4	3	0.0	2752	5091	178	...
2	139441	Photo	2	12	4	3	0.0	2752	5091	178	...
3	139441	Photo	3	12	3	3	0.0	2413	4373	177	...
4	139441	Photo	3	12	3	3	0.0	2413	4373	177	...

5 rows × 37 columns

In [34]:
```
photo_video_merged.shape
```

Out[34]:
```
(1382, 37)
```

Sorting data

In [38]:
```
# Sorting the data on the basis of number of likes
likes_sorted_data = dataset.sort_values(by="Page total likes")
```

In [39]:
```
# Displaying the top 5 liked records
likes_sorted_data.head()
```

Out[39]:

	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
498	81370	Photo	3	1	4	11	0.0	4156	7564	626	574	
497	81370	Photo	1	1	5	2	0.0	3778	7216	625	572	
496	81370	Photo	2	1	5	8	0.0	3480	6229	537	508	
493	85093	Photo	3	1	1	2	0.0	8412	13960	1179	1111	
495	85093	Photo	3	1	7	2	0.0	4684	7536	733	708	

In [40]:
```
# Displaying the bottom 10 liked records
likes_sorted_data.tail(10)
```

Out[40]:

	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
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
12	139441	Photo	2	12	5	10	0.0	2847	5133	193	115
8	139441	Status	2	12	7	3	0.0	11844	22538	1530	1407
9	139441	Photo	3	12	6	10	0.0	4694	8668	280	183
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
13	139441	Photo	2	12	5	3	0.0	2549	4896	249	134
7	139441	Photo	3	12	7	9	1.0	13720	24137	537	232
0	139441	Photo	2	12	4	3	0.0	2752	5091	178	109

Reshaping the data

Melting

In [41]:
```
# Melting the data on the value variables as type and category
melting_result = pd.melt(
    frame=dataset,
    id_vars="Page total likes",
    value_vars=["Type", "Category"]
)
```

In [42]:
```
melting_result.head()
```

Out[42]:

	Page total likes	variable	value
0	139441	Type	Photo
1	139441	Type	Status
2	139441	Type	Photo
3	139441	Type	Photo
4	139441	Type	Photo

In [43]:
```
melting_result.tail()
```

Out[43]:

	Page total likes	variable	value
985	85093	Category	3
986	85093	Category	3
987	81370	Category	2
988	81370	Category	1
989	81370	Category	3

In [44]:
```
# Checking shape of melted data
melting_result.shape
```

Out[44]: (990, 3)