In [ ]:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [ ]:

```
df = pd.read_csv("dataset_Facebook.csv", sep=";")
```

In [ ]:

```
print("Shape : ", df.shape)
```

```
Shape :  (500, 19)
```

In [ ]:

```
df.head()
```

Out[9]:

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

In [ ]:

```
print("Total Number of rows: ", df.shape[0])
print("Total numbe rof columsn : ", df.shape[1])
```

```
Total Number of rows:  500
Total numbe rof columsn :  19
```

In [ ]:

```
df.info()
```

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

In [ ]:

```
print("Columns : ")
for i in list(df.columns):
  print(i)
```

```
Columns :
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
```

In [ ]:

```
for i in df.columns:
  print(i , " : " ,(df[i].isnull().sum() / df.shape[0])*100)
```

```
Page total likes  :  0.0
Type  :  0.0
Category  :  0.0
Post Month  :  0.0
Post Weekday  :  0.0
Post Hour  :  0.0
Paid  :  0.2
Lifetime Post Total Reach  :  0.0
Lifetime Post Total Impressions  :  0.0
Lifetime Engaged Users  :  0.0
Lifetime Post Consumers  :  0.0
Lifetime Post Consumptions  :  0.0
Lifetime Post Impressions by people who have liked your Page  :  0.0
Lifetime Post reach by people who like your Page  :  0.0
Lifetime People who have liked your Page and engaged with your post  :  0.
0
comment  :  0.0
like  :  0.2
share  :  0.8
Total Interactions  :  0.0
```

In [ ]:

```
df.isnull().sum()
```

Out[19]:

```
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 [ ]:

```
df.describe(include=['O']).T
```

Out[23]:

	count	unique	top	freq
Type	500	4	Photo	426

In [ ]:

```
df['Type'].value_counts()
```

Out[24]:

```
Photo     426
Status     45
Link       22
Video       7
Name: Type, dtype: int64
```

Dropping the null value rows

In [ ]:

```
data = df.dropna()
data.shape
```

Out[25]:

```
(495, 19)
```

In [ ]:

```
data.shape
```

Out[32]:

```
(495, 19)
```

Segregating the data as per "Type"

Photo

In [ ]:

```
df_photo = data[data['Type']=="Photo"]
df_photo.shape
```

Out[33]:

```
(421, 19)
```

In [ ]:

```
df_photo.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 421 entries, 0 to 498
Data columns (total 19 columns):
 #   Column
Non-Null Count  Dtype
---  ------
--------------  -----
 0   Page total likes
421 non-null    int64
 1   Type
421 non-null    object
 2   Category
421 non-null    int64
 3   Post Month
421 non-null    int64
 4   Post Weekday
421 non-null    int64
 5   Post Hour
421 non-null    int64
 6   Paid
421 non-null    float64
 7   Lifetime Post Total Reach
421 non-null    int64
 8   Lifetime Post Total Impressions
421 non-null    int64
 9   Lifetime Engaged Users
421 non-null    int64
 10  Lifetime Post Consumers
421 non-null    int64
 11  Lifetime Post Consumptions
421 non-null    int64
 12  Lifetime Post Impressions by people who have liked your Page
421 non-null    int64
 13  Lifetime Post reach by people who like your Page
421 non-null    int64
 14  Lifetime People who have liked your Page and engaged with your post
421 non-null    int64
 15  comment
421 non-null    int64
 16  like
421 non-null    float64
 17  share
421 non-null    float64
 18  Total Interactions
421 non-null    int64
dtypes: float64(3), int64(15), object(1)
memory usage: 65.8+ KB
```

In [ ]:

```
df_photo.describe().T
```

Out[35]:

	count	mean	std	min	25%	50%	75%	
Page total likes	421.0	122319.612827	16242.669134	81370.0	109670.0	128032.0	136013.0	13
Category	421.0	1.926366	0.884681	1.0	1.0	2.0	3.0	
Post Month	421.0	6.790974	3.228447	1.0	4.0	7.0	10.0	
Post Weekday	421.0	4.087886	2.056203	1.0	2.0	4.0	6.0	
Post Hour	421.0	8.004751	4.432561	1.0	3.0	9.0	11.0	
Paid	421.0	0.282660	0.450828	0.0	0.0	0.0	1.0	
Lifetime Post Total Reach	421.0	13275.389549	22977.950816	238.0	3110.0	4708.0	10844.0	18
Lifetime Post Total Impressions	421.0	29306.147268	81387.055518	570.0	5439.0	8198.0	17442.0	111
Lifetime Engaged Users	421.0	825.368171	871.390487	9.0	421.0	612.0	951.0	1
Lifetime Post Consumers	421.0	695.475059	734.384315	9.0	337.0	539.0	837.0	1
Lifetime Post Consumptions	421.0	1310.334917	1974.750382	9.0	521.0	834.0	1336.0	1
Lifetime Post Impressions by people who have liked your Page	421.0	16594.650831	64864.680971	567.0	3779.0	5648.0	10987.0	110
Lifetime Post reach by people who like your Page	421.0	6118.539192	7845.941764	236.0	2110.0	3136.0	5696.0	5
Lifetime People who have liked your Page and engaged with your post	421.0	510.964371	400.796595	9.0	305.0	403.0	570.0	
comment	421.0	7.581948	22.641009	0.0	1.0	3.0	7.0	
like	421.0	184.066508	346.542549	0.0	58.0	101.0	187.0	
share	421.0	27.156770	44.973104	0.0	10.0	19.0	32.0	
Total Interactions	421.0	218.805226	407.568504	0.0	72.0	124.0	226.0	

In [ ]:

```
plt.figure(figsize=(15,15))

sns.heatmap(df_photo.corr(), annot=True, linewidths=1.1)
plt.show()
```

In [ ]:

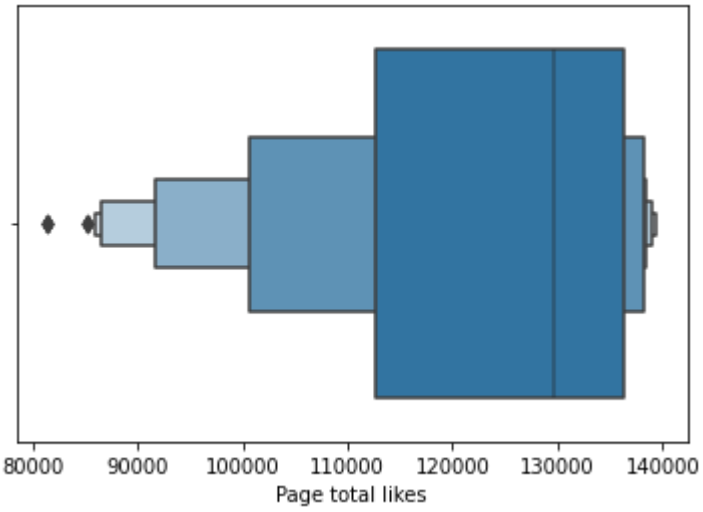
```
# df_photo.columns

cols = ['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', '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']

import warnings
warnings.filterwarnings("ignore")


for i in cols:
  print("\n\n",i," : ")
  sns.boxenplot(df[i])
  plt.show()
```

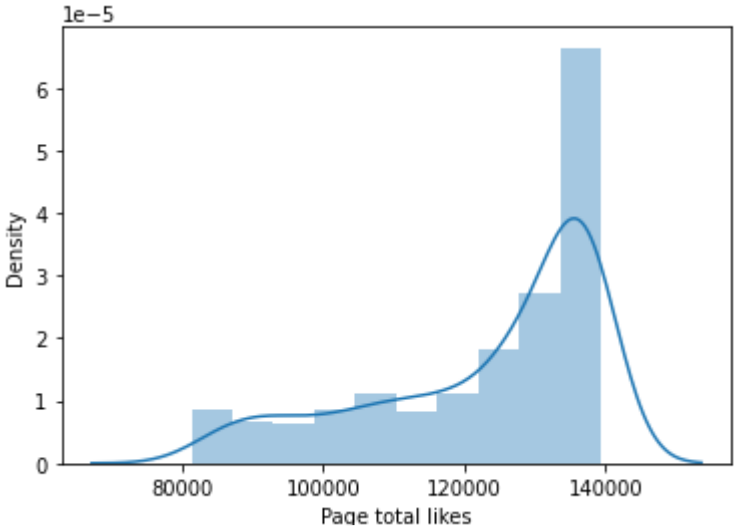
```
 Page total likes  :
```



In [ ]:

```
for i in cols:
  print("\n\n",i," : ")
  sns.distplot(df[i])
  plt.show()
```

Page total likes :

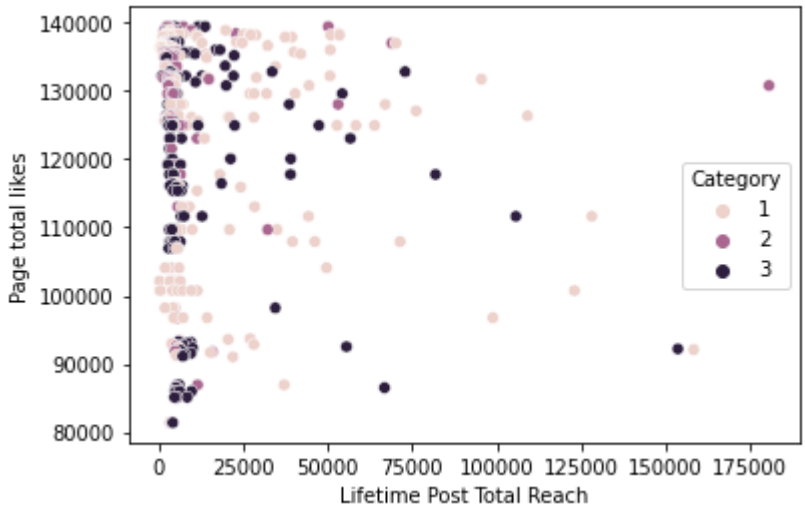


Bivariate analysis

In [ ]:

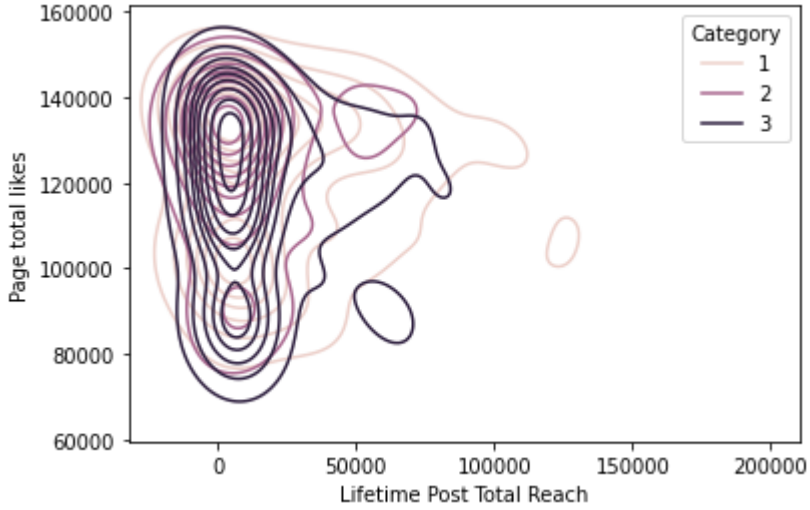
```
# Lifetime Post Total Reach , Page total likes
# plt.figure(figsize=(15,15))

sns.scatterplot(df_photo['Lifetime Post Total Reach'], df['Page total likes'], hue=df['C
plt.show()
```



In [ ]:

```
sns.kdeplot(df_photo['Lifetime Post Total Reach'], df['Page total likes'], hue=df["Categ
plt.show()
```



In [ ]:

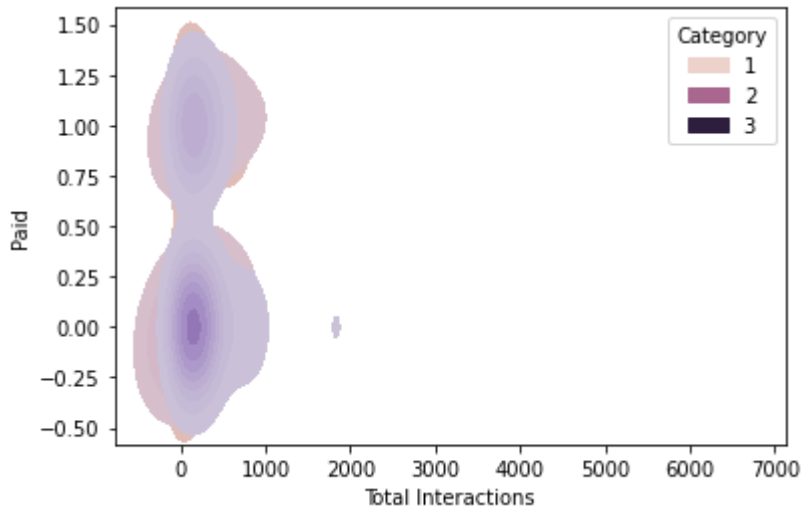
```
data.columns
```

Out[56]:

```
Index(['Page total likes', 'Type', 'Category', 'Post Month', 'Post Weekda
y',
       '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 pos
t',
       'comment', 'like', 'share', 'Total Interactions'],
      dtype='object')
```

In [ ]:

```
sns.kdeplot(data['Total Interactions'], data['Paid'], hue=df["Category"] ,fill=True)
plt.show()
```



Status

In [ ]:

```
df_status = data[data['Type']=="Status"]
print("Status : ",df_status.shape)

df_link= data[data['Type']=="Link"]
print("Link : ",df_link.shape)

df_video = data[data['Type']=="Video"]
print("Video : ",df_video.shape)
```

```
Status :  (45, 19)
Link :  (22, 19)
Video :  (7, 19)
```

#Merging any 2 subsets

In [ ]:

```
status_link_dataframe = pd.concat([df_status, df_link])
status_link_dataframe.shape
```

Out[48]:

```
(67, 19)
```

#Transpose

In [ ]:

```
data.transpose()
```

Out[68]:

	0	1	2	3	4	5	6	7	8	
Page total likes	139441	139441	139441	139441	139441	139441	139441	139441	139441	13
Type	Photo	Status	Photo	Photo	Photo	Status	Photo	Photo	Status	P
Category	2	2	3	2	2	2	3	3	2	
Post Month	12	12	12	12	12	12	12	12	12	
Post Weekday	4	3	3	2	2	1	1	7	7	
Post Hour	3	10	3	10	3	9	3	9	3	
Paid	0.0	0.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	
Lifetime Post Total Reach	2752	10460	2413	50128	7244	10472	11692	13720	11844	4
Lifetime Post Total Impressions	5091	19057	4373	87991	13594	20849	19479	24137	22538	8
Lifetime Engaged Users	178	1457	177	2211	671	1191	481	537	1530	
Lifetime Post Consumers	109	1361	113	790	410	1073	265	232	1407	
Lifetime Post Consumptions	159	1674	154	1119	580	1389	364	305	1692	
Lifetime Post Impressions by people who have liked your Page	3078	11710	2812	61027	6228	16034	15432	19728	15220	4
Lifetime Post reach by people who like your Page	1640	6112	1503	32048	3200	7852	9328	11056	7912	2
Lifetime People who have liked your Page and engaged with your post	119	1108	132	1386	396	1016	379	422	1250	
comment	4	5	0	58	19	1	3	0	0	
like	79.0	130.0	66.0	1572.0	325.0	152.0	249.0	325.0	161.0	1
share	17.0	29.0	14.0	147.0	49.0	33.0	27.0	14.0	31.0	
Total Interactions	100	164	80	1777	393	186	279	339	192	

19 rows × 495 columns

In [ ]:

```
data.sort_values(['Page total likes','Lifetime Post Total Reach'], ascending=[1,0])
```

Out[71]:

	Page total likes	Type	Category	Post Month	Post Weekday	Post Hour	Paid	Lifetime Post Total Reach	Lifetime Post Total Impressions	Lifetime Engaged Users	Cc
498	81370	Photo	3	1	4	11	0.0	4156	7564	626	
497	81370	Photo	1	1	5	2	0.0	3778	7216	625	
496	81370	Photo	2	1	5	8	0.0	3480	6229	537	
493	85093	Photo	3	1	1	2	0.0	8412	13960	1179	
494	85093	Photo	3	1	7	10	0.0	5400	9218	810	
...	...	...	...	...	...	...	...	...	...	...	
11	139441	Photo	2	12	5	10	0.0	3112	5590	208	
12	139441	Photo	2	12	5	10	0.0	2847	5133	193	
0	139441	Photo	2	12	4	3	0.0	2752	5091	178	
13	139441	Photo	2	12	5	3	0.0	2549	4896	249	
2	139441	Photo	3	12	3	3	0.0	2413	4373	177	

495 rows × 19 columns

In [ ]: