TI35_DSBDA_4th

February 5, 2025

[5]:
```
import pandas as pd
```

[6]:
```
df=pd.read_csv("/home/dpl11/facebook.csv")
```

[7]:
```
df.head()
```

[7]:

	Unnamed: 0	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
0	0	139441	Photo	2	12	4	3	0.0	2752	5091	178	109	159	3078
1	1	139441	Status	2	12	3	10	0.0	10460	19057	1457	1361	1674	11710
2	2	139441	Photo	3	12	3	3	0.0	2413	4373	177	113	154	
3	3	139441	Photo	2	12	2	10	1.0	50128	87991	2211	790	1119	
4	4	139441	Photo	2	12	2	3	0.0	7244	13594	671	410	580	

```
2                                                 2812
3                                                61027
4                                                 6228

   Lifetime Post reach by people who like your Page  \
0                                                1640
1                                                6112
2                                                1503
3                                               32048
4                                                3200

   Lifetime People who have liked your Page and engaged with your post  \
0                                                119
1                                               1108
2                                                132
3                                               1386
4                                                396

   comment    like  share  Total Interactions
0        4    79.0   17.0                 100
1        5   130.0   29.0                 164
2        0    66.0   14.0                  80
3       58  1572.0  147.0                1777
4       19   325.0   49.0                 393
```

[8]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 20 columns):
 #   Column                                                                  Non-
Null Count  Dtype
---  ------
--------------  -----
 0   Unnamed: 0                                                              500
non-null    int64
 1   Page total likes                                                        500
non-null    int64
 2   Type                                                                    500
non-null    object
 3   Category                                                                500
non-null    int64
 4   Post Month                                                              500
non-null    int64
 5   Post Weekday                                                            500
non-null    int64
 6   Post Hour                                                               500
```

```
non-null    int64
 7   Paid                                                        499
non-null    float64
 8   Lifetime Post Total Reach                                   500
non-null    int64
 9   Lifetime Post Total Impressions                             500
non-null    int64
 10  Lifetime Engaged Users                                      500
non-null    int64
 11  Lifetime Post Consumers                                     500
non-null    int64
 12  Lifetime Post Consumptions                                  500
non-null    int64
 13  Lifetime Post Impressions by people who have liked your Page  500
non-null    int64
 14  Lifetime Post reach by people who like your Page            500
non-null    int64
 15  Lifetime People who have liked your Page and engaged with your post  500
non-null    int64
 16  comment                                                     500
non-null    int64
 17  like                                                        499
non-null    float64
 18  share                                                       496
non-null    float64
 19  Total Interactions                                          500
non-null    int64
dtypes: float64(3), int64(16), object(1)
memory usage: 78.2+ KB
```

[9]:
```
df.shape
```

[9]:
```
(500, 20)
```

[10]:
```
df.dtypes
```

[10]:
```
Unnamed: 0                                                  int64
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
```

[11]:
```
df.isnull().sum()
```

[11]:
```
Unnamed: 0                                                               0
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
```

[12]:
```
#Creating the subset of data
df_subset1 = df[['like','share']]
```

[13]:
```
df_subset1
```

[13]:

	like	share
0	79.0	17.0
1	130.0	29.0
2	66.0	14.0
3	1572.0	147.0
4	325.0	49.0
..	…	…

```
495    53.0   26.0
496    53.0   22.0
497    93.0   18.0
498    91.0   38.0
499    91.0   28.0

[500 rows x 2 columns]
```

[14]:
```
df_subset2 = df[['comment','Type']]
```

[15]:
```
df_subset2
```

[15]:
```
     comment    Type
0          4   Photo
1          5  Status
2          0   Photo
3         58   Photo
4         19   Photo
..       …     …
495        5   Photo
496        0   Photo
497        4   Photo
498        7   Photo
499        0   Photo

[500 rows x 2 columns]
```

[16]:
```
#Merging the DataFrames
merged_data = pd.merge(df_subset2, df_subset1, left_on='comment',␣
 ↪right_on='like')
merged_data
```

[16]:
```
      comment    Type  like  share
0           4   Photo   4.0    2.0
1           4   Photo   4.0    1.0
2           4   Photo   4.0    0.0
3           4   Photo   4.0    1.0
4           4  Status   4.0    2.0
…       …      …   …    …
1462       56   Photo  56.0   17.0
1463       56   Photo  56.0    8.0
1464       56   Photo  56.0   12.0
1465       56   Photo  56.0    9.0
1466       56   Photo  56.0   25.0

[1467 rows x 4 columns]
```

[17]:
```
# Sort by 'Likes' in descending order
sorted_df = df.sort_values(by='like', ascending=False)
print("\nSorted Data by Likes (Descending):")
print(sorted_df)
```

```
Sorted Data by Likes (Descending):
     Unnamed: 0  Page total likes   Type  Category  Post Month  Post Weekday  \
244         244            130791  Photo         2           7             3
379         379            111620  Photo         3           4             1
349         349            117764  Photo         3           5             5
168         168            135428  Photo         1           9             3
3             3            139441  Photo         2          12             2
..          …                 …    …       …          …            …
21           21            138414  Photo         1          12             7
100         100            137020  Photo         1          10             4
441         441             98195  Photo         1           3             5
417         417            104070  Photo         1           3             3
111         111            136736  Photo         1          10             6

     Post Hour  Paid  Lifetime Post Total Reach  \
244          5   1.0                     180480
379         14   1.0                     105632
349         13   0.0                      81856
168         10   0.0                      41984
3           10   1.0                      50128
..         …   …                        …
21          10   0.0                       1384
100          9   1.0                       1357
441          4   1.0                       1845
417         10   0.0                       1874
111          8   0.0                       1261

     Lifetime Post Total Impressions  Lifetime Engaged Users  \
244                           319133                    8072
379                           147918                    3984
349                           124753                    3000
168                            68290                    3370
3                              87991                    2211
..                               …                       …
21                              2467                      15
100                             2453                      37
441                             2670                       9
417                             2474                      25
111                             2158                      37

     Lifetime Post Consumers  Lifetime Post Consumptions  \
```

244	4010	6242
379	2254	3391
349	1637	2718
168	2420	4074
3	790	1119
..	…	…
21	15	20
100	37	55
441	9	9
417	25	31
111	37	49

	Lifetime Post Impressions by people who have liked your Page
244	108752
379	48575
349	52477
168	34802
3	61027
..	…
21	2196
100	2154
441	1614
417	1483
111	1911

	Lifetime Post reach by people who like your Page
244	51456
379	27328
349	27392
168	20928
3	32048
..	…
21	1172
100	1120
441	1008
417	1062
111	1077

	Lifetime People who have liked your Page and engaged with your post
244	3316
379	1936
349	1756
168	2126
3	1386
..	…
21	15
100	32
441	9

```
417                                                  15
111                                                  33

     comment    like  share  Total Interactions
244      372  5172.0  790.0                6334
379       51  1998.0  128.0                2177
349       45  1639.0  122.0                1806
168      144  1622.0  208.0                1974
3         58  1572.0  147.0                1777
..       …     …    …                    …
21         0     0.0    0.0                   0
100        0     0.0    0.0                   0
441        0     0.0    0.0                   0
417        0     0.0    0.0                   0
111        0     NaN    NaN                   0

[500 rows x 20 columns]
```

[18]:
```
# Transpose the DataFrame
merged_data.transpose()
```

[18]:
```
           0      1      2      3       4       5       6       7       8     \
comment      4      4      4      4       4       4       4       4       4
Type     Photo  Photo  Photo  Photo  Status  Status  Status  Status  Status
like       4.0    4.0    4.0    4.0     4.0     4.0     4.0     4.0     4.0
share      2.0    1.0    0.0    1.0     2.0     1.0     0.0     1.0     2.0

            9     …   1457   1458   1459   1460   1461   1462   1463   1464  \
comment       4  …     51     51     51    146    146     56     56     56
Type     Status  …  Photo  Photo  Photo  Photo  Photo  Photo  Photo  Photo
like        4.0  …   51.0   51.0   51.0  146.0  146.0   56.0   56.0   56.0
share       1.0  …   11.0    6.0    6.0    9.0   15.0   17.0    8.0   12.0

          1465   1466
comment     56     56
Type     Photo  Photo
like      56.0   56.0
share      9.0   25.0

[4 rows x 1467 columns]
```

[19]:
```
merged_data.T
```

[19]:
```
           0      1      2      3       4       5       6       7       8     \
comment      4      4      4      4       4       4       4       4       4
Type     Photo  Photo  Photo  Photo  Status  Status  Status  Status  Status
like       4.0    4.0    4.0    4.0     4.0     4.0     4.0     4.0     4.0
```

```
share      2.0    1.0    0.0    1.0     2.0     1.0     0.0     1.0     2.0

           9     …   1457   1458   1459   1460   1461   1462   1463   1464  \
comment      4   …     51     51     51    146    146     56     56     56
Type    Status   …  Photo  Photo  Photo  Photo  Photo  Photo  Photo  Photo
like       4.0   …   51.0   51.0   51.0  146.0  146.0   56.0   56.0   56.0
share      1.0   …   11.0    6.0    6.0    9.0   15.0   17.0    8.0   12.0

          1465   1466
comment     56     56
Type    Photo  Photo
like     56.0   56.0
share     9.0   25.0

[4 rows x 1467 columns]
```

[20]:
```
#Shape
df.shape
```

[20]:
```
(500, 20)
```

[21]:
```
# Reshape
df_temp = pd.DataFrame({'foo': ['one', 'one', 'one', 'two', 'two', 'two'],

'bar': ['A', 'B', 'C', 'A', 'B', 'C'],
'baz': [1, 2, 3, 4, 5, 6],
'zoo': ['x', 'y', 'z', 'q', 'w', 't']})
```

[22]:
```
df_temp
```

[22]:
```
   foo bar  baz zoo
0  one   A    1   x
1  one   B    2   y
2  one   C    3   z
3  two   A    4   q
4  two   B    5   w
5  two   C    6   t
```

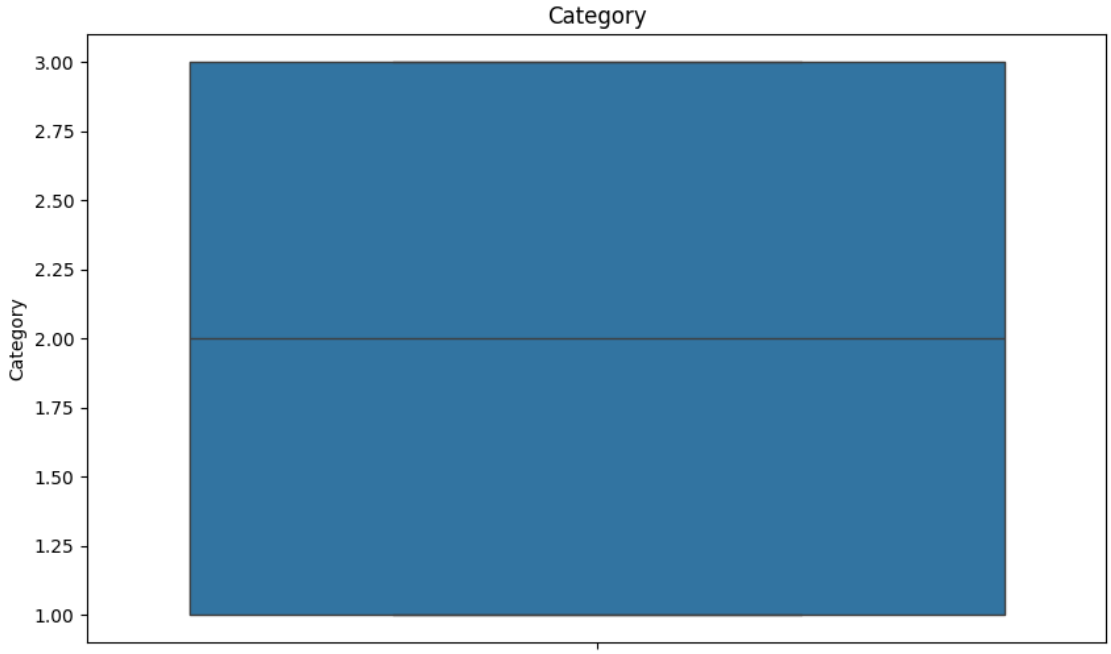
[23]:
```
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import accuracy_score,confusion_matrix
from sklearn.linear_model import LogisticRegression
import seaborn as sns
import matplotlib.pyplot as plt
# import pandas library
import numpy as np
```

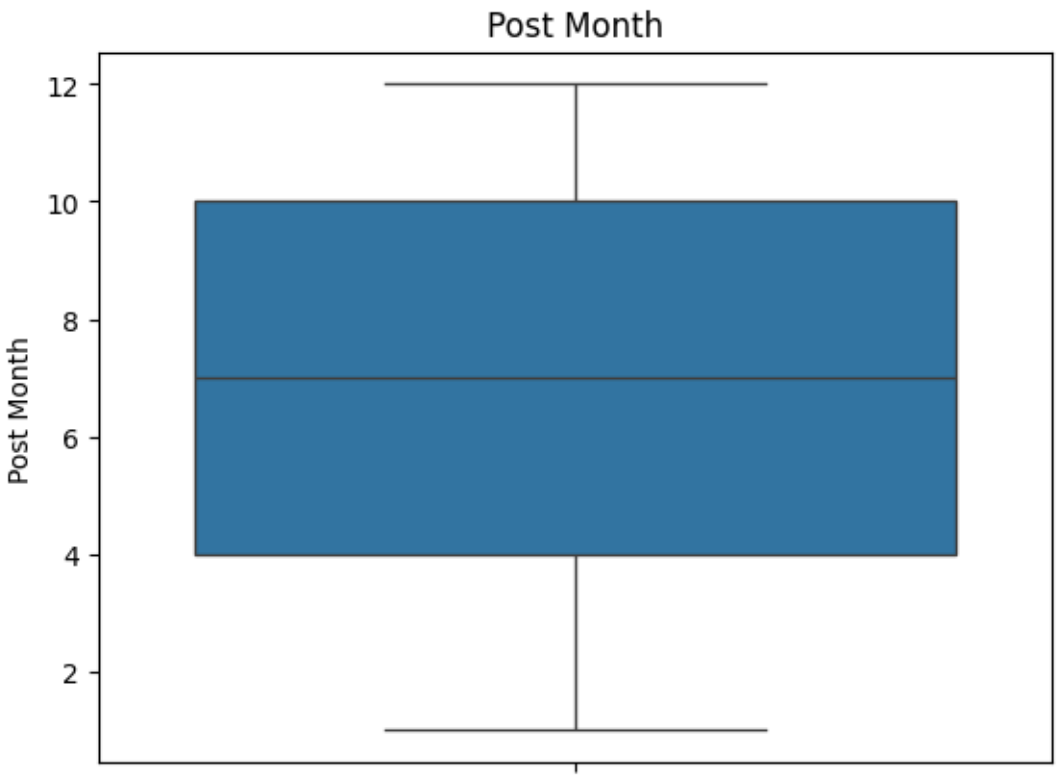
[25]:
```
def remove_outliers(column):
  Q1 = column.quantile(0.25)
  Q3 = column.quantile(0.75)
  IQR = Q3 - Q1
  threshold = 1.5 * IQR
  outlier_mask = (column < Q1 - threshold) | (column > Q3 + threshold)
  return column[~outlier_mask]
```

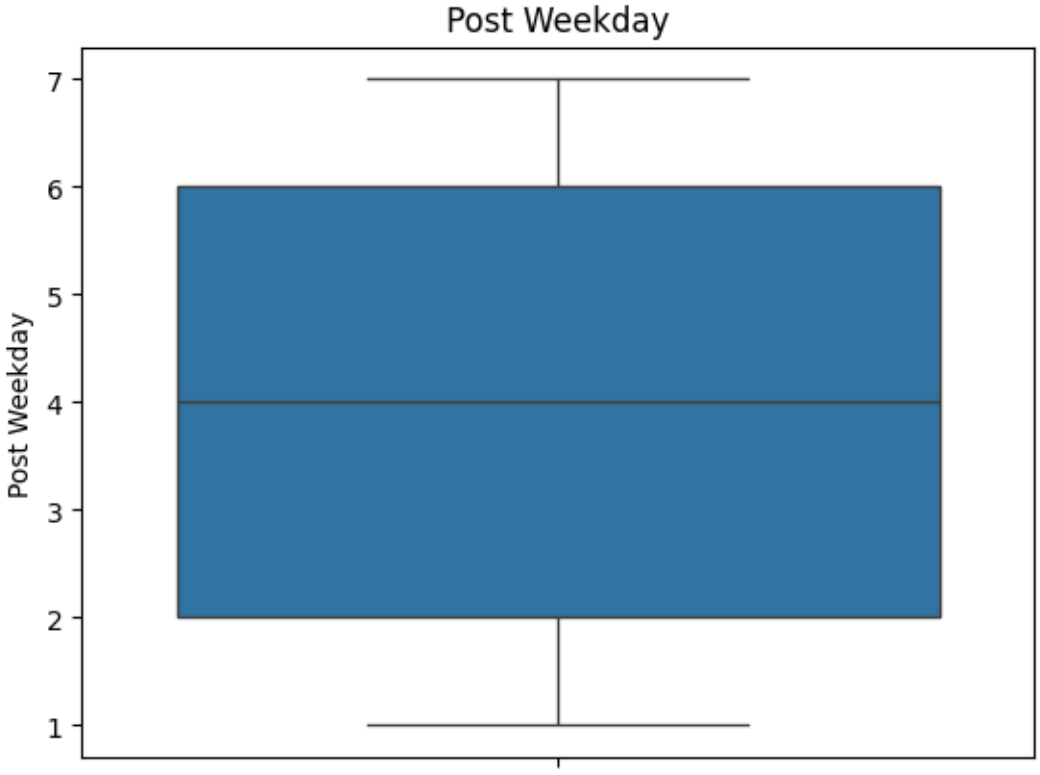
[27]:
```
df = df.drop_duplicates()
```

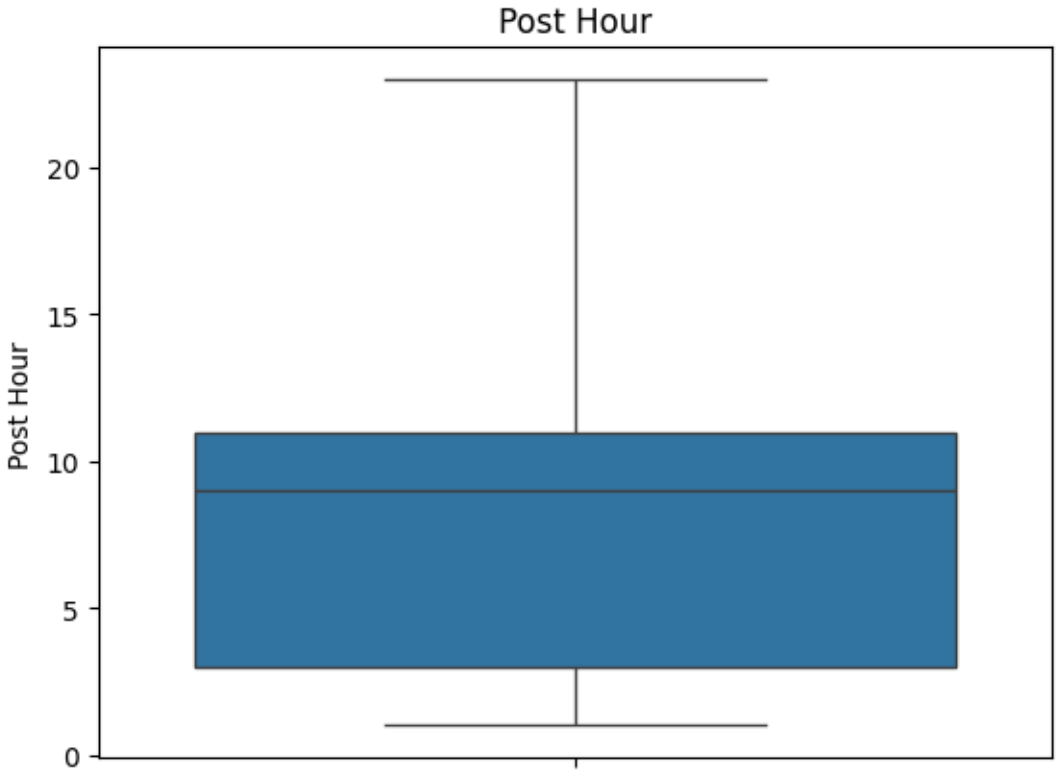
[29]:
```
col_name = ['Category','Post Month','Post Weekday','Post Hour','Paid']
for col in col_name:
  df[col] = remove_outliers(df[col])
```

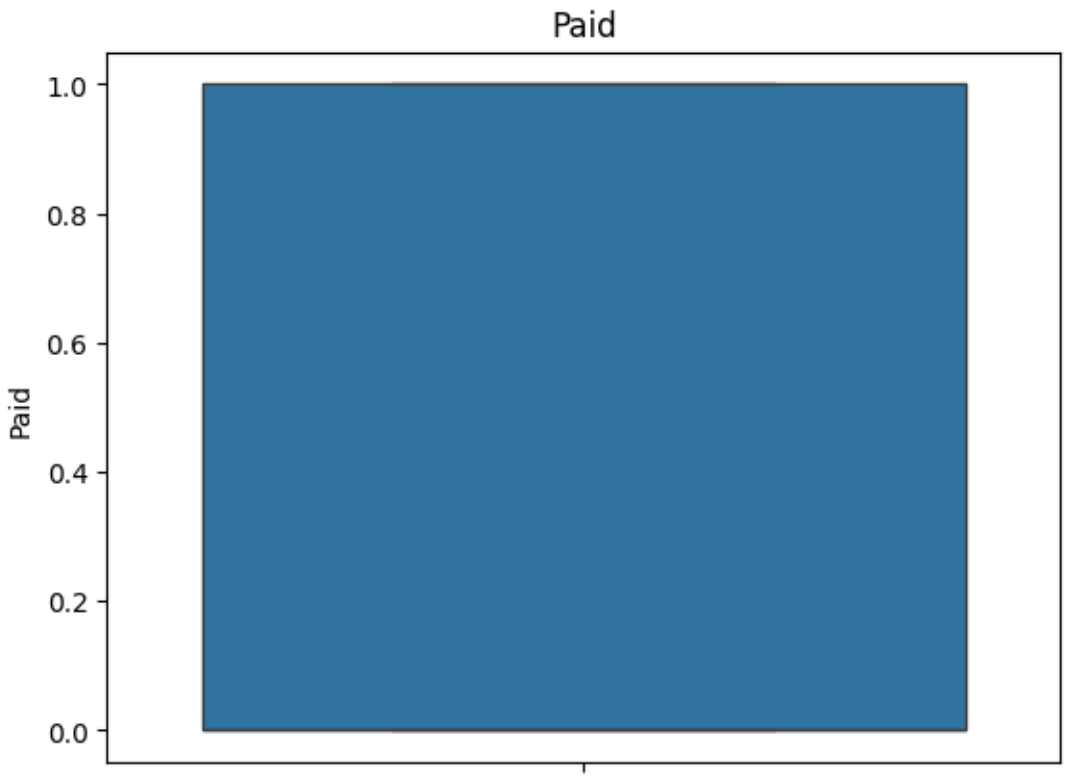
[30]:
```
plt.figure(figsize=(10, 6)) # Adjust the figure size if needed
for col in col_name:
  sns.boxplot(data=df[col])
  plt.title(col)
  plt.show()
```




Post Month
Post Month
12
10
8
6
4
2


Post Weekday
Post Weekday
7
6
5
4
3
2
1


Post Hour
Post Hour
20
15
10
5
0


Paid
1.0
0.8
0.6
0.4
0.2
0.0
Paid

[ ]: