In [1]:
```
pip install pandas
```

```
Requirement already satisfied: pandas in c:\users\pc\anaconda3\lib\site-packages (1.
3.4)
Requirement already satisfied: pytz>=2017.3 in c:\users\pc\anaconda3\lib\site-packag
es (from pandas) (2021.3)
Requirement already satisfied: python-dateutil>=2.7.3 in c:\users\pc\anaconda3\lib\s
ite-packages (from pandas) (2.8.2)
Requirement already satisfied: numpy>=1.17.3 in c:\users\pc\anaconda3\lib\site-packa
ges (from pandas) (1.20.3)
Requirement already satisfied: six>=1.5 in c:\users\pc\anaconda3\lib\site-packages
(from python-dateutil>=2.7.3->pandas) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

In [2]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [3]:
```
import os
os.chdir('C:\\Users\\pc\\Desktop\\SCDE-JNTUH')
```

In [4]:
```
df=pd.read_csv('internet_session.csv')
```

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

Out[5]:
```
name                    object
start_time              object
usage_time              object
IP                      object
MAC                     object
upload                  object
download                object
total_transfer         float64
seession_break_reason   object
dtype: object
```

In [6]:
```
df['usage_time'] = df['usage_time'].str.replace('00:', '', 1)
df['usage_time'] = pd.to_datetime(df['usage_time'])
```

In [7]:
```
df
```

Out[7]:

	name	start_time	usage_time	IP	MAC	upload	download	total_trai
0	user1	10-05-2022 02:59	2022-12-20 00:36:28	10.55.14.222	48:E7:DA:58:22:E9	15861.76	333168.64	3490
1	user1	10-05-2022 18:53	2022-12-20 01:49:56	10.55.2.253	48:E7:DA:58:22:E9	16957.44	212152.32	2291(
2	user1	10-05-2022 21:20	2022-12-20 01:35:00	10.55.2.253	48:E7:DA:58:22:E9	14080	195153.92	2092

	name	start_time	usage_time	IP	MAC	upload	download	total_tra
3	user1	11-05-2022 00:37	2022-12-20 00:26:00	10.55.2.253	48:E7:DA:58:22:E9	5242.88	40806.4	460
4	user1	11-05-2022 02:59	2022-12-20 00:11:52	10.55.2.253	48:E7:DA:58:22:E9	22067.2	10772.48	328
...	...	...	...	...	...	...	...	
4707	user9	04-11-2022 01:11	2022-12-20 06:54:32	10.55.4.189	DA:2F:97:0E:B7:D0	107960.32	2390753.28	24956
4708	user9	04-11-2022 10:26	2022-12-20 00:23:49	10.55.4.59	DA:2F:97:0E:B7:D0	11407.36	209674.24	2210
4709	user9	04-11-2022 20:41	2022-12-20 01:24:13	10.55.15.186	DA:2F:97:0E:B7:D0	18995.2	373657.6	3926
4710	user9	05-11-2022 00:21	2022-12-20 08:49:43	10.55.4.159	DA:2F:97:0E:B7:D0	46602.24	593766.4	6403
4711	user9	05-11-2022 20:55	2022-12-20 01:06:20	10.55.2.33	DA:2F:97:0E:B7:D0	21237.76	298536.96	3197

4712 rows × 9 columns

In [8]:
```
device = []
basename = 'device'
mac = df['MAC'][0]
device_number = 1
for i in df['MAC']:
    if i == mac:
        device.append(basename + str(device_number))
    else:
        device_number += 1
        device.append(basename + str(device_number))
        mac = i
df['device'] = device
```

In [9]:
```
mac=df['MAC'][31]
```

In [10]:
```
mac
```

Out[10]:
```
'0C:F3:46:71:E2:27'
```

In [11]:
```
df.head(40)
```

Out[11]:

	name	start_time	usage_time	IP	MAC	upload	download	total_transf

	name	start_time	usage_time	IP	MAC	upload	download	total_transf
0	user1	10-05-2022 02:59	2022-12-20 00:36:28	10.55.14.222	48:E7:DA:58:22:E9	15861.76	333168.64	349030.4
1	user1	10-05-2022 18:53	2022-12-20 01:49:56	10.55.2.253	48:E7:DA:58:22:E9	16957.44	212152.32	229109.7
2	user1	10-05-2022 21:20	2022-12-20 01:35:00	10.55.2.253	48:E7:DA:58:22:E9	14080	195153.92	209233.9
3	user1	11-05-2022 00:37	2022-12-20 00:26:00	10.55.2.253	48:E7:DA:58:22:E9	5242.88	40806.4	46049.2
4	user1	11-05-2022 02:59	2022-12-20 00:11:52	10.55.2.253	48:E7:DA:58:22:E9	22067.2	10772.48	32839.6
5	user1	11-05-2022 17:17	2022-12-20 02:44:29	10.55.2.253	48:E7:DA:58:22:E9	39905.28	676167.68	716083.2
6	user1	11-05-2022 21:27	2022-12-20 00:51:03	10.55.2.253	48:E7:DA:58:22:E9	9472	166584.32	176056.3
7	user1	11-05-2022 23:10	2022-12-20 00:20:35	10.55.2.253	48:E7:DA:58:22:E9	11960.32	170977.28	182947.8
8	user1	12-05-2022 00:04	2022-12-20 00:03:12	10.55.2.253	48:E7:DA:58:22:E9	1146.88	3532.8	4679.6
9	user1	12-05-2022 12:48	2022-12-20 00:45:24	10.55.7.61	48:E7:DA:58:22:E9	12974.08	130764.8	143738.8
10	user1	12-05-2022 14:10	2022-12-20 01:14:50	10.55.7.61	48:E7:DA:58:22:E9	13445.12	168089.6	181534.7
11	user1	12-05-2022 18:10	2022-12-20 01:34:39	10.55.15.202	48:E7:DA:58:22:E9	25344	169369.6	194713.6
12	user1	12-05-2022 21:09	2022-12-20 03:02:17	10.55.15.202	48:E7:DA:58:22:E9	39976.96	571975.68	611942.4
13	user1	13-05-2022 13:20	2022-12-20 00:47:37	10.55.15.202	48:E7:DA:58:22:E9	13649.92	94228.48	107868.1
14	user1	13-05-2022 17:05	2022-12-20 06:52:23	10.55.15.202	48:E7:DA:58:22:E9	78960.64	1195376.64	1268776.9
15	user1	14-05-2022 11:08	2022-12-20 02:37:11	10.55.12.137	48:E7:DA:58:22:E9	14120.96	142612.48	156733.4

	name	start_time	usage_time	IP	MAC	upload	download	total_transf
16	user1	14-05-2022 14:16	2022-12-20 01:11:46	10.55.12.137	48:E7:DA:58:22:E9	17838.08	87091.2	104929.2
17	user1	14-05-2022 18:23	2022-12-20 00:13:54	10.55.12.137	48:E7:DA:58:22:E9	3440.64	11888.64	15319.(
18	user1	14-05-2022 18:50	2022-12-20 02:06:35	10.55.12.137	48:E7:DA:58:22:E9	15278.08	126566.4	141844.4
19	user1	14-05-2022 21:27	2022-12-20 04:25:16	10.55.12.137	48:E7:DA:58:22:E9	68567.04	939571.2	1008128.(
20	user1	15-05-2022 12:04	2022-12-20 01:43:36	10.55.12.137	48:E7:DA:58:22:E9	16435.2	141465.6	157900.8
21	user1	15-05-2022 14:13	2022-12-20 05:04:49	10.55.12.137	48:E7:DA:58:22:E9	116766.72	3030384.64	3145728.(
22	user1	15-05-2022 20:02	2022-12-20 00:45:10	10.55.12.137	48:E7:DA:58:22:E9	5713.92	15472.64	21186.5
23	user1	16-05-2022 01:48	2022-12-20 01:01:58	10.55.12.137	48:E7:DA:58:22:E9	12963.84	137687.04	150650.8
24	user1	16-05-2022 02:51	2022-12-20 00:37:05	10.55.12.137	48:E7:DA:58:22:E9	5222.4	55398.4	60610.5
25	user1	16-05-2022 08:12	2022-12-20 00:50:04	10.55.12.137	48:E7:DA:58:22:E9	7219.2	20899.84	28119.(
26	user1	16-05-2022 10:38	2022-12-20 00:28:14	10.55.12.137	48:E7:DA:58:22:E9	6625.28	86568.96	93194.2
27	user1	16-05-2022 11:46	2022-12-20 01:13:30	10.55.12.137	48:E7:DA:58:22:E9	10240	22845.44	33095.6
28	user1	16-05-2022 13:23	2022-12-20 00:07:35	10.55.7.89	0C:F3:46:71:E2:27	1873.92	21463.04	23326.7
29	user1	16-05-2022 13:52	2022-12-20 02:04:02	10.55.12.137	48:E7:DA:58:22:E9	11130.88	49643.52	60774.4
30	user1	16-05-2022 16:10	2022-12-20 01:02:36	10.55.12.137	48:E7:DA:58:22:E9	8048.64	36106.24	44154.8
31	user1	16-05-2022 16:25	2022-12-20 00:03:10	10.55.11.75	0C:F3:46:71:E2:27	105.72	751.23	856.9

	name	start_time	usage_time	IP	MAC	upload	download	total_transf
32	user1	16-05-2022 19:23	2022-12-20 00:12:04	10.55.12.137	48:E7:DA:58:22:E9	7188.48	13578.24	20766.7
33	user1	16-05-2022 19:50	2022-12-20 00:13:51	10.55.12.137	48:E7:DA:58:22:E9	691.88	4741.12	5437.4
34	user1	16-05-2022 22:10	2022-12-20 00:11:24	10.55.12.137	48:E7:DA:58:22:E9	5294.08	12052.48	17346.5
35	user1	16-05-2022 22:31	2022-12-20 00:02:58	10.55.12.137	48:E7:DA:58:22:E9	688.7	3706.88	4392.9
36	user1	16-05-2022 22:44	2022-12-20 00:09:33	10.55.6.12	0C:F3:46:71:E2:27	209.67	651.48	861.1
37	user1	16-05-2022 23:56	2022-12-20 00:01:40	10.55.12.137	48:E7:DA:58:22:E9	392.22	2365.44	2754.5
38	user1	17-05-2022 01:22	2022-12-20 00:17:25	10.55.9.101	48:E7:DA:58:22:E9	4229.12	52285.44	56504.3
39	user1	17-05-2022 08:36	2022-12-20 00:15:28	10.55.9.101	48:E7:DA:58:22:E9	11632.64	99758.08	111390.7

In [12]:
```
## HOW Often IP Changed##
```

In [13]:
```
base_ip = '10.55.14.222'
ip_count = 0
for i in range(1, df.shape[0]):
    if df.iloc[i]['IP'] != base_ip:
        ip_count +=1
        base_ip = df.iloc[i]['IP']

print('The IP Adress changed ' + str(ip_count) + ' times')
```

```
The IP Adress changed 2308 times
```

In [14]:
```
## How often device changed##
```

In [15]:
```
base_device='48:E7:DA:58:22:E9'
device_count=0
for i in range(1,df.shape[0]):
    if df.iloc[i]['MAC']!=base_device:
        device_count +=1
        base_device=df.iloc[i]['MAC']


print('The device changed '+ str(device_count) +' Times' )
```

```
The device changed 1226 Times
```

In [16]:
```
df['start_time']=pd.to_datetime(df['start_time'])
```

In [17]:
```
df.dtypes
```

Out[17]:
```
name                          object
start_time            datetime64[ns]
usage_time            datetime64[ns]
IP                            object
MAC                           object
upload                        object
download                      object
total_transfer               float64
seession_break_reason         object
device                        object
dtype: object
```

In [24]:
```
df['hour']=df['start_time'].dt.hour
df['day']=df['start_time'].dt.day
df['month']=df['start_time'].dt.month
```

In [25]:
```
df.dtypes
```

Out[25]:
```
name                          object
start_time            datetime64[ns]
usage_time            datetime64[ns]
IP                            object
MAC                           object
upload                        object
download                      object
total_transfer               float64
seession_break_reason         object
device                        object
day                            int64
month                          int64
hour                           int64
dtype: object
```

In [26]:
```
df.head()
```

Out[26]:

	name	start_time	usage_time	IP	MAC	upload	download	total_transfer
0	user1	2022-10-05 02:59:00	2022-12-20 00:36:28	10.55.14.222	48:E7:DA:58:22:E9	15861.76	333168.64	349030.40
1	user1	2022-10-05 18:53:00	2022-12-20 01:49:56	10.55.2.253	48:E7:DA:58:22:E9	16957.44	212152.32	229109.76
2	user1	2022-10-05 21:20:00	2022-12-20 01:35:00	10.55.2.253	48:E7:DA:58:22:E9	14080	195153.92	209233.92
3	user1	2022-11-05 00:37:00	2022-12-20 00:26:00	10.55.2.253	48:E7:DA:58:22:E9	5242.88	40806.4	46049.28

	name	start_time	usage_time	IP	MAC	upload	download	total_transfer
4	user1	2022-11-05 02:59:00	2022-12-20 00:11:52	10.55.2.253	48:E7:DA:58:22:E9	22067.2	10772.48	32839.68

In [27]:
```
hourly_average= df.groupby('hour').total_transfer.mean()
print('Hourly Average '+ str(hourly_average))
```

```
Hourly Average hour
0     464530.443023
1     530880.856788
2     431576.112743
3     345303.341176
4     359809.443333
5     275960.910769
6     468959.586757
7     292886.830164
8     366681.918762
9     377480.638954
10    393259.119955
11    309492.445992
12    310137.981415
13    335270.579648
14    472403.712765
15    517005.111506
16    403919.401872
17    525423.692116
18    665414.452500
19    390839.426364
20    355740.055442
21    471461.399116
22    449600.499185
23    407785.083903
Name: total_transfer, dtype: float64
```

In [ ]:
```
## Average usage per day ##
```

In [29]:
```
day_average=df.groupby('day').total_transfer.mean()
print('Day Average '+ str(day_average))
```

```
Day Average day
5     289375.902346
6     323132.999760
7     418583.993765
8     684726.478718
9     571484.142727
10    503412.695784
11    399675.450813
13    501906.696839
14    352701.095843
15    521520.508537
16    426719.389162
17    475795.713377
18    337490.929720
19    301941.319205
20    365130.117857
21    462211.693202
22    486595.372151
```

```
23    383153.932258
24    320598.936839
25    443689.468743
26    463432.019073
27    324318.123558
28    494576.340769
29    363645.605755
30    361418.880000
31    369118.009524
Name: total_transfer, dtype: float64
```

In [ ]:

```
## Average usage per month##
```

In [31]:

```
month_average=df.groupby('month').total_transfer.mean()
print('Month Average use  '+ str(month_average))
```

```
Month Average use  month
1     396705.036138
2     494496.481046
3     445865.633099
4     676332.032787
5     373137.832164
6     355747.948782
7     402259.891473
8     393170.178707
9     419855.952000
10    521525.979053
11    729857.654151
12    346695.953810
Name: total_transfer, dtype: float64
```

In [ ]: