In [98]:
```
import pandas as pd
```

In [99]:
```
df=pd.read_csv("internet_session.csv")
df['seession_break_reason'].value_counts()
```

Out[99]:
```
Idle-Timeout    4350
Lost-Carrier     162
Lost-Service     124
User-Request      65
NAS-Reboot         2
Name: seession_break_reason, dtype: int64
```

In [100]:
```
df
```

Out[100]:

	name	start_time	usage_time	IP	MAC	upload	download	total_transfer	seession_break_reason
0	user1	2022-05-10 02:59:32	00:00:36:28	10.55.14.222	48:E7:DA:58:22:E9	15861.76	333168.64	349030.40	Idle-Timeout
1	user1	2022-05-10 18:53:27	00:01:49:56	10.55.2.253	48:E7:DA:58:22:E9	16957.44	212152.32	229109.76	Idle-Timeout
2	user1	2022-05-10 21:20:44	00:01:35:00	10.55.2.253	48:E7:DA:58:22:E9	14080.0	195153.92	209233.92	Idle-Timeout
3	user1	2022-05-11 00:37:42	00:00:26:00	10.55.2.253	48:E7:DA:58:22:E9	5242.88	40806.4	46049.28	Idle-Timeout
4	user1	2022-05-11 02:59:38	00:00:11:52	10.55.2.253	48:E7:DA:58:22:E9	22067.2	10772.48	32839.68	Idle-Timeout
...	...	...	...	...	...	...	...	...	...
4707	user9	2022-11-04 01:11:34	00:06:54:32	10.55.4.189	DA:2F:97:0E:B7:D0	107960.32	2390753.28	2495610.88	Idle-Timeout
4708	user9	2022-11-04 10:26:09	00:00:23:49	10.55.4.59	DA:2F:97:0E:B7:D0	11407.36	209674.24	221081.60	Idle-Timeout
4709	user9	2022-11-04 20:41:42	00:01:24:13	10.55.15.186	DA:2F:97:0E:B7:D0	18995.2	373657.6	392652.80	Idle-Timeout
4710	user9	2022-11-05 00:21:06	00:08:49:43	10.55.4.159	DA:2F:97:0E:B7:D0	46602.24	593766.4	640368.64	Idle-Timeout
4711	user9	2022-11-05 20:55:37	00:01:06:20	10.55.2.33	DA:2F:97:0E:B7:D0	21237.76	298536.96	319774.72	NaN

4712 rows × 9 columns

In [101]:
```
df['Date']=pd.to_datetime(df['start_time']).dt.date
df[['start_time','Date']]
```

Out[101]:

	start_time	Date
0	2022-05-10 02:59:32	2022-05-10
1	2022-05-10 18:53:27	2022-05-10
2	2022-05-10 21:20:44	2022-05-10
3	2022-05-11 00:37:42	2022-05-11
4	2022-05-11 02:59:38	2022-05-11
...	...	...
4707	2022-11-04 01:11:34	2022-11-04
4708	2022-11-04 10:26:09	2022-11-04
4709	2022-11-04 20:41:42	2022-11-04
4710	2022-11-05 00:21:06	2022-11-05
4711	2022-11-05 20:55:37	2022-11-05

4712 rows × 2 columns

In [102]:
```
df['Time']=pd.to_datetime(df['start_time']).dt.time
df['Time']
```

Out[102]:
```
0       02:59:32
1       18:53:27
2       21:20:44
3       00:37:42
4       02:59:38
          ...
4707    01:11:34
4708    10:26:09
4709    20:41:42
4710    00:21:06
4711    20:55:37
Name: Time, Length: 4712, dtype: object
```

In [103]:
```
# 1.What is the most frequent internet activity time of the day ?
dfg=df.groupby('Date')['Time'].agg(pd.Series.mode)
dfg
```

Out[103]:
```
Date
2022-05-09                                                 22:52:41
2022-05-10                          [02:59:32, 18:53:27, 21:20:44]
2022-05-11    [00:37:42, 02:59:38, 17:17:21, 21:27:01, 23:10...
2022-05-12    [00:04:25, 12:48:13, 14:10:50, 18:10:10, 21:09...
2022-05-13                                [13:20:21, 17:05:29]
                                    ...
2022-11-01             [10:40:46, 14:05:58, 19:29:12, 22:59:43]
2022-11-02             [11:44:36, 15:15:36, 18:59:19, 21:55:29]
2022-11-03                                [19:36:08, 21:36:59]
2022-11-04                                [06:45:15, 22:08:51]
2022-11-05    [10:51:49, 12:42:49, 13:47:30, 16:46:38, 18:41...
Name: Time, Length: 154, dtype: object
```

In [104]:
```
#2. How often the ip changes
dfg_ip=df.groupby('name')['IP'].agg(pd.Series.nunique)
dfg_ip
```

Out[104]:
```
name
user1    130
user2    193
user3    239
user4    211
user5    188
user6    130
user7    132
user8    117
user9    289
Name: IP, dtype: int64
```

In [105]:
```
#3. How often the MAC changes
dfg_mac=df.groupby('name')['MAC'].agg(pd.Series.nunique)
dfg_mac
```

Out[105]:
```
name
user1    5
user2    6
user3    2
user4    4
user5    5
user6    5
user7    8
user8    2
user9    3
Name: MAC, dtype: int64
```

In [106]:
```
# 4.Monthly average usage
df['Month']=pd.to_datetime(df['start_time']).dt.strftime('%b')
dfg_month=df.groupby('Month')['total_transfer'].mean()
dfg_month
```

Out[106]:
```
Month
Aug    479042.438202
Jul    418583.993765
Jun    338418.082988
May    311177.156960
Nov    399675.450813
Oct    549467.626233
Sep    482955.522841
Name: total_transfer, dtype: float64
```

In [107]:
```
# 5.Day average usage
df['Day']=pd.to_datetime(df['start_time']).dt.strftime('%d')
dfg_day=df.groupby('Day')['total_transfer'].mean()
dfg_day
```

Out[107]:
```
Day
01    396705.036138
02    494496.481046
03    445865.633099
04    676332.032787
05    634564.047308
06    396261.748364
07    402259.891473
08    301859.567305
09    393521.968614
10    350665.024785
11    729857.654151
12    346695.953810
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 [108]:
```
# 6.Per hour average usage
df['Hour']=df['usage_time'].str.split(":").str[1]
tot=(df['total_transfer']/1024).sum()
df['Hour']=pd.to_numeric(df['Hour'],errors='coerce')
toth=df['Hour'].sum()
tott=(df['total_transfer']).sum()
print("Average usage per hour in KB is", tott/toth)
```

```
Average usage per hour in KB is 251509.84170904127
```