Assignment 8

This project is about the Data of internet usage [in kb] by graduate student at an indian university.

Answer for the following questions usign Data Analysis.

What is the most frequent internet activity time of the day ? How often the ip changes ? How often the device changed. What is the average usage per hour , per day and per month ?

Hint: internet_session.csv provided to students.

In [72]:
```
import pandas as pd
import numpy as np

df = pd.read_csv('Documents/Data science Material JNTU/Assignments/Assignment 8/internet_session.csv')
df.head()
```

Out[72]:

	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

In [73]:
```
df.shape
```

Out[73]:
```
(4712, 9)
```

In [74]:
```
df.columns
```

Out[74]:
```
Index(['name', 'start_time', 'usage_time', 'IP', 'MAC', 'upload', 'download',
       'total_transfer', 'seession_break_reason'],
      dtype='object')
```

In [75]:
```
df.dtypes
```

Out[75]:
```
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 [76]:
```
df.isna().sum()
```

Out[76]:
```
name                    0
start_time              0
usage_time              0
IP                      0
MAC                     0
upload                  0
download                0
total_transfer          0
seession_break_reason   9
dtype: int64
```

In [77]:
```
df = df.dropna().copy()
df.isna().sum()
```

Out[77]:
```
name                    0
start_time              0
usage_time              0
IP                      0
MAC                     0
upload                  0
download                0
total_transfer          0
seession_break_reason   0
dtype: int64
```

In [78]:
```
df.duplicated().sum()
```

Out[78]:
```
0
```

In [79]:
```
df['usage_time'] = df['usage_time'].str.replace('00:', '', 1)
df['usage_time'] = pd.to_datetime(df['usage_time'])
```

In [80]:
```
df['upload'] = df['upload'].str.extract('(\d+)', expand=False)
df.upload = df.upload.astype(float)
```

In [81]:
```
df['download'] = df['download'].str.extract('(\d+)', expand=False)
df.download = df.download.astype(float)
```

In [82]:
```
df.describe(include='all', datetime_is_numeric=True)
```

Out[82]:

	name	start_time	usage_time	IP	MAC	upload	download	total_transfer	seession_break_reason
count	4703	4703	4703	4703	4703	4.703000e+03	4.703000e+03	4.703000e+03	4703
unique	9	4029	NaN	1299	33	NaN	NaN	NaN	5
top	user4	2022-05-10 02:59:32	NaN	10.55.0.89	48:E7:DA:58:22:E9	NaN	NaN	NaN	Idle-Timeout
freq	725	2	NaN	80	1235	NaN	NaN	NaN	4350
mean	NaN	NaN	2022-12-23 02:10:05.038486272	NaN	NaN	3.378702e+04	3.966645e+05	4.304372e+05	NaN
min	NaN	NaN	2022-12-23 00:00:01	NaN	NaN	2.000000e+00	9.000000e+00	1.120000e+00	NaN
25%	NaN	NaN	2022-12-23 00:31:42	NaN	NaN	6.082000e+03	5.199800e+04	6.187008e+04	NaN
50%	NaN	NaN	2022-12-23 01:19:40	NaN	NaN	1.531900e+04	1.782680e+05	2.027930e+05	NaN
75%	NaN	NaN	2022-12-23 02:49:02	NaN	NaN	3.399600e+04	4.593660e+05	4.993997e+05	NaN
max	NaN	NaN	2022-12-23 22:00:07	NaN	NaN	2.841640e+06	2.790261e+07	2.855272e+07	NaN
std	NaN	NaN	NaN	NaN	NaN	9.493243e+04	9.657778e+05	9.960848e+05	NaN

In [83]:
```
df.name.value_counts()
```

Out[83]:
```
user4    725
user6    674
user1    673
user9    571
user7    526
user3    518
user2    456
user5    335
user8    225
Name: name, dtype: int64
```

What is the most frequent internet activity time of the day ?

In [84]:
```
#What is the most frequent internet activity time of the day ?
df['hour'] = pd.to_datetime(df['start_time']).dt.hour
frequent_time = dict(df['hour'].value_counts())
for i in frequent_time:
    if frequent_time[i] == df['hour'].value_counts().max():
        print("most frequent internet activity time of the day is ", i)
        break;
```

```
most frequent internet activity time of the day is  22
```

How often the ip changes ?

In [86]:
```
prev_ip=df.iloc[0]['IP']
ip_changed=0
for i in range(1, df.shape[0]):
    if df.iloc[i]['IP'] != prev_ip:
        ip_changed +=1
        prev_ip = df.iloc[i]['IP']

print('How often IP Adress changed ',ip_changed)
```

```
How often IP Adress changed  2303
```

How often the device changed?

In [87]:
```
device_count = 0
prev_mac=df.iloc[0]['MAC']

for i in range(1, df.shape[0]):
    if df.iloc[i]['MAC'] != prev_mac:
        device_count +=1
        prev_mac = df.iloc[i]['MAC']

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

```
The device changed 1223 times
```

What is the average usage per hour , per day and per month ?

In [88]:
```
df['day'] = pd.to_datetime(df['start_time']).dt.day
df['month'] = pd.to_datetime(df['start_time']).dt.month

hourly_average = df.groupby('hour').total_transfer.mean()
print('The Average usage per hour is:\n ' +  str(round(hourly_average, 2)))

daily_average = df.groupby('day').total_transfer.mean()
print('The Average usage per day is:\n ' +  str(round(daily_average, 2)))

monthly_average = df.groupby('month').total_transfer.mean()
print('The Average usage per month is:\n ' +  str(round(monthly_average, 2)))
```

```
The Average usage per hour is:
 hour
0     464530.44
1     530880.86
2     431576.11
3     345303.34
4     359809.44
5     275960.91
6     468959.59
7     292886.83
8     366681.92
9     377480.64
10    393259.12
11    309492.45
12    310137.98
13    335270.58
14    472403.71
15    517005.11
16    403919.40
17    525423.69
18    666590.76
19    389841.79
20    355862.80
21    474038.34
22    449600.50
23    407785.08
Name: total_transfer, dtype: float64
The Average usage per day is:
```

```
 day
1     396705.04
2     494496.48
3     445865.63
4     676332.03
5     652195.66
6     396261.75
7     402259.89
8     301859.57
9     393521.97
10    350665.02
11    729857.65
12    346695.95
13    501906.70
14    352701.10
15    521520.51
16    426719.39
17    475795.71
18    337490.93
19    301941.32
20    365130.12
21    462211.69
22    486595.37
23    383153.93
24    320598.94
25    443689.47
26    463432.02
27    324318.12
28    494576.34
29    363645.61
30    361418.88
31    369118.01
Name: total_transfer, dtype: float64
The Average usage per month is:
 month
5     311177.16
6     338418.08
7     418583.99
8     479042.44
9     482955.52
10    549467.63
11    399804.11
Name: total_transfer, dtype: float64
```