Assigment-8

1. Internet Usage Analysis

This project is about the data of internet usage [in KB] by graduate student at an indian university

Answer for the following questions using 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?

```
for google.doc import drive
drive.mount('/content/drive')
        Mounted at/content/drive

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
path = '/content/internet_session.csv'
internet_usage = pd.read_csv(path)
# Dataset is now stored in a pandas Data frame.
internet_usage.describe.
internet_usage.describe of name  Start_time  usage_time  IP Mac
<bound method NDFrame.describe of
0  User1  10-05-2022  02:59  00:00:36.28  10.55.14.222  48:E7:DA:58
1  User1  10-05-2022  02:51  00:00:49:56  10:55:2:253  48:E7:DA:58:22
2  User3  10-05-2022  02:59  00:00:35:00  10:55:2:253  48:E7:DA:58:22:E9
3  User4  11-05-2022  02:59  00:00:26.00  10.55.2:253  48:E7:DA:58:22:E9
4  User5  11-05-2022  02:59  00:00:11.52  10:55:2:253  48:E7:DA:58:22:E9
```

4707	User9	04-11-2022	00:06:54:32	10.55.4.189	DA:2F:97:0E:B3:0
4708	User9	04-11-2022	00:00:23:49	10.55.4.59	DA:2F:97:0E:B7:0
4709	User9	04-11-2022	00:01:24:13	10.55.15.186	DA:2F:97:0E:B7:0
4710	User9	05-11-2022	00:08:49:43	10.55.4.159	DA:2F:97:0E:B3:0
4711	User9	05-11-2022	00:01:06:20	10.55.2.33	DA:2F:97:0E:B7:D

	Upload	download	total_transfer	session_break_reason	Unamed:9
0	15861.76	333168.64	349030.40	Idle_Timeout	NaN
1	16957.44	212152.32	229109.76	Idle_Timeout	NaN
2	14080	195153.92	209233.92	Idle_Timeout	NaN
3	5242.88	40806.4	46049.28	Idle_Timeout	NaN
4	22067.2	10772.48	32839.68	Idle_Timeout	NaN

4707	107960.32	2390753.28	2495610.88	Idle_Timeout	NaN
4708	11407.36	209674.24	221081.60	Idle_Timeout	NaN
4709	18995.2	373657.6	392652.80	Idle_Timeout	NaN
4710	46602.24	593766.4	640368.64	Idle_Timeout	NaN
4711	21237.76	298536.96	319774.72	Idle_Timeout	NaN

	Unnamed:10
0	NaN
1	NaN
2	NaN
3	NaN
4	NaN

4707	NaN
4708	NaN
4709	NaN
4710	NaN
4711	NaN

[4712 rows x 11 columns]>

```
internet_usage.shape
    (4712,11)

internet_usage.columns
    Index(['name','start_time','usage_time','IP','MAC','upload','download',
           'total_transfer','session_break_reason','Unnamed:9','Unnamed:10'],
          dtype='object')

internet_usage.columns = internet_usage.columns.str.lower()

internet_usage.columns
    Index(['name','start_time','usage_time','ip','mac','upload','download',
           'total_transfer','session_break_reason','unnamed:9','unnamed:10'],
          dtype='object')

internet_usage.dtypes
    name                  object
    start_time            object
    usage_time            object
    ip                    object
    mac                   object
    upload                object
    download              object
    total_transfer        object
    session_break_reason  object
    unnamed:9             object
    unnamed:10            object
    dtype: object

internet_usage.isna().sum()
    name          0
    start_time    0
    usage_time    0
    ip            0
```

```
mac                          0
upload                       0
dowload                      0
total_tranger                0
Session_breall_reason     4712
Unnamed:9                 4712
Unnamed:10                4712
dtype: int64
```

```
internet_usage = internet_usage.dropna().copy()
Internet_usage.isna().sum()
```

```
name                    0.0
start_time              0.0
usage_time              0.0
qp                      0.0
mac                     0.0
upload                  0.0
dowload                 0.0
total_tranger           0.0
Session_break_reason    0.0
Unnamed:9               0.0
Unnamed:10              0.0
dtype: float64
```

```
Internet_Usage.duplicated().sum()
```

```
0
```

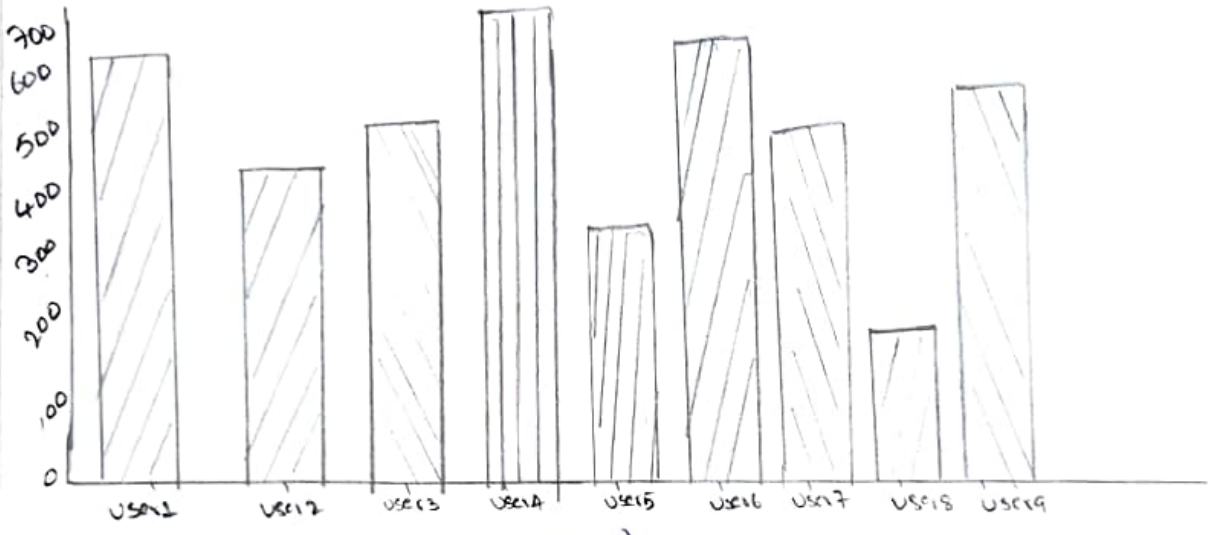
```
Internet_Usage.describe(include='all', datatime_is_numeric=True)
```

	name	start_time	usage_time	qp	mac	upload	dowload	Total_tranger	Session_breall reason	Unnamed:9	Unnamed:10
count	0	0	0	0	0	0.0	0.0	0.0	0	0.0	0
unique	0	0	NaN	0	0	NaN	NaN	NaN	0	0.0	NaN
top	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

```
Internet_Usage.name.value_counts()
```

```
Series([], Name: name, dtype: int64)
min      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN
```

```
plt.figure(figsize=(18,9))
ax = sns.countplot(x='name', data=internet_usage)
plt.title("User count")
plt.show()
plt.clf()
```




```
print("The earliest start time is :")
print('internet_usage.start_time.min())
print('The latest start time is :')
print('internet_usage.start_time.max())
```

```
The earliest start time is :
01-06-2022 00:30
The latest start time is :
31-10-2022 22:46
```

```
print('The minimum usage time is :')
print(internet_usage.usage_time.min())
print('The maximum Usage time is :')
print('internet_usage.usage_time.max())
```

```
The minimum usage time is :
00:00:00:01
The maximum usage time is :
```

```
print ('The minimum usage time per user:')
Usage_time_minimum = internet_usage.groupby('name').Usage_time.min()
Usage_time_minimum
```

The minimum usage time per user:

name	
User 1	00:00:00:18
User 2	00:00:00:08
User 3	00:00:00:01
User 4	00:00:00:65
User 5	00:00:00:07
User 6	00:00:00:18
User 7	00:00:00:20
User 8	00:00:00:20
User 9	00:00:00:09

Name: Usage_time, dtype: Object

```
print ('The maximum usage time per user:')
Usage_time_maximum = internet_usage.groupby('name').Usage_time.max()
Usage_time_maximum
```

The maximum usage time per user:

name	
User 1	00:19:35:11
User 2	00:20:39:52
User 3	00:17:01:28
User 4	01:00:21:07
User 5	00:06:36:11
User 6	00:19:35:11
User 7	00:21:00:07
User 8	00:17:24:26
User 9	00:19:16:09

Name: Usage_time, dtype: Object

internet_usage.Ip.value_counts()

10.55.0.89	80
10.55.14.148	64
10.55.15.221	55
10.55.1.50	48
10.55.10.46	44
...	...
10.55.14.67	1
10.55.12.190	1
10.56.2.33	1

Name:
Name:IP, length: 1302, dtype: int64

```
print('The minimum upload is: ' + str(internet_usage.upload.min()) + 'Kb')
print('The maximum upload is: ' + str(internet_usage.upload.max()) + 'Kb')
```

The minimum upload is: 10004.48 Kb

The maximum upload is: 9994.24 Kb

```
print('The minimum upload per user:')
internet_usage.groupby('name').upload.min()
```

The minimum upload per user:

name	
User 1	1006.74
User 2	10117.12
User 3	10055.68
User 4	10147.84
User 5	10076.16
User 6	1006.74
User 7	10035.2
User 8	10004.48
User 9	100720.64

Name: upload, dtype: object

```
print('The maximum upload per user:')
Internet_Usage.groupby('name').upload.max()
```

The maximum upload per user:

name	
User1	9932.8
User2	9994.24
User3	99860.48
User4	9963.52
User5	9973.76
User6	9932.8
User7	99860.48
User8	9922.56
User9	9963.52

Name: upload, dtype: object

```
print('The minimum download is:'+str(Internet_Usage.download.min())+'KB')
print('The maximum download is:'+str(Internet_Usage.download.max())+'KB')
```

The minimum download is: 100014.08 KB

The maximum download is: 999393.28 KB

```
print('The minimum download per user:')
Internet_Usage.groupby('name').download.min()
```

The minimum download per user

name	
User1	100167.68
User2	10035.2
User3	100044.8
User4	100055.04
User5	10076.16
User6	100167.68
User7	100014.08
User8	101672.96
User9	1001216.

Name: download, dtype: object

```
print('The maximum download per user:')
Internet_usage.groupby('name').download.max()
```

The maximum download per user:

```
name
User 1    999393.28
User2     99676.16
User3     996352
User4     9953.28
User5     990812.16
User7     999393.28
User8     99938.45
User9     9908.06
Name: download, dtype: object
```

```
print('The minimum total transfer per user:')
Internet_usage.groupby('name').total_transfer.min()
```

The minimum total transfer per user

```
name
User2     75.34
User2     73.13
User3     1.12
User4     13.45
User5     924.40
User6     75.34
User7     15.25
User8     18.08
User1     102.64
Name: total_transfer, dtype: float64
```

```
print('The maximum total transfer per user:')
Internet_usage.groupby('name').total_transfer.max()
```

The maximum total transfer per user:

name.

User1 24389877.76

User2 22051553.28

User3 3282042.88

User4 8524922.88

User5 5158993.92

User6 24389877.76

User7 28552724.48

User8 3166699.52

User9 6155141.12

Name: total_transfer, dtype: float64

```
internet_usage.session_break_reason.value_counts()
```

Idle_Timeout 4350

Lost_carrier 162

Lost_service 124

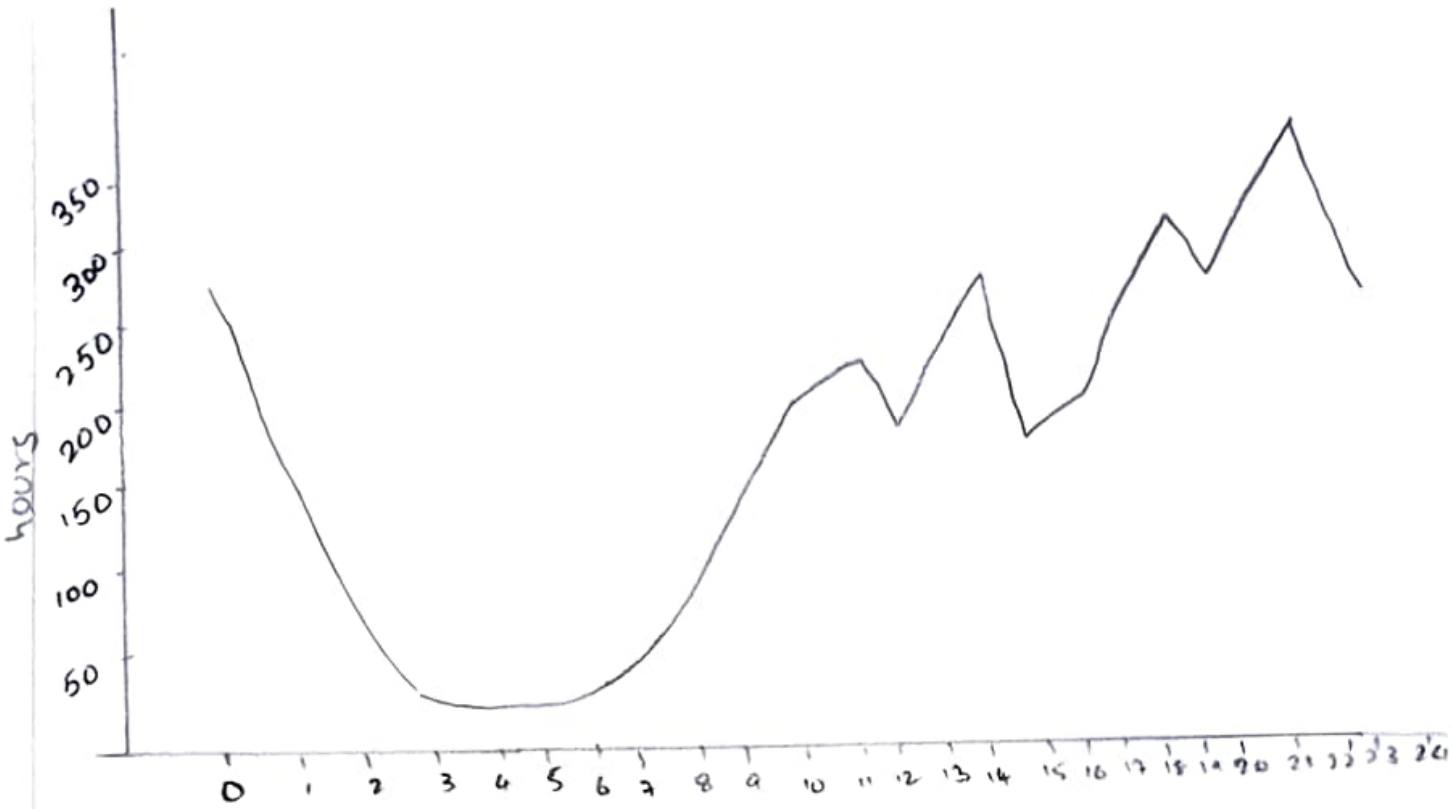
User_Request 65

Nas_Reboot 2

Name: session_break_reason, dtype: int64

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

```
internet_usage['hour'] = pd.to_datetime(internet_usage['Start_time']).dt.hour
frequent_activity_time_of_day = internet_usage['hour'].value_counts().sort_index()
plt.figure(fig_size=(18,9))
sns.lineplot(data=frequent_activity_time_of_day)
plt.xticks(np.linspace(start=0, stop=24, num=25))
plt.show()
plt.clf()
```



How often the ip changes?

```
base_Ip = '48:0E7:0A:58:22:E9'
Ip_count = 0
for i in range(1, internet_usage.shape[0]):
    if internet_usage.iloc[i]['Ip'] != base_Ip:
        Ip_count += 1
        base_Ip = internet_usage.iloc[i]['Ip']
print('The Ip Adress changed ' + str(Ip_count) + 'time')
```

The Ip Adress changed 2308 times

The dataset contain a user that used 1224 difference devices to connect to the internet while uploading 2841640.Kb and downloading 27902607.0Kb with a total tranfer of 28552724.48Kb during a period of 7 months.

The most frequent internet activity time of the day is 22hrs or 10pm

The Ip Adress changed 2303 times while the device used changed 1223 times.

The highest average usage per hour was 666590.76 KB around 18h or 6pm. the highest average usage per day was 729857.65 KB around the 11th day of the month, while the highest usage per month was during the month of October with 549467.63 KB total tranfer of data.