

Assignment on ML DL

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 ?

Hint: internet_session.csv provided to students.

```
from google.colab 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 Dataframe
internet_usage.describe
```

```
<bound method NDFrame.describe of
0 user1 10-05-2022 02:59 00:00:36:28 10.55.14.222 48:E7:DA:58:22:E9
49:56 10.55.2.253 48:E7:DA:58:22:E9
:35:00 10.55.2.253 48:E7:DA:58:22:E9
:26:00 10.55.2.253 48:E7:DA:58:22:E9
4 user1 11-05-2022 02:59 00:00:11:52 10.55.2.253 48:E7:DA:58:22:E9
...
4707 user9 04-11-2022 01:11 00:06:54:32 10.55.4.189 DA:2F:97:0E:B7:D0
4708 user9 04-11-2022 10:26 00:00:23:49 10.55.4.59 DA:2F:97:0E:B7:D0
4709 user9 04-11-2022 20:41 00:01:24:13 10.55.15.186 DA:2F:97:0E:B7:D0
4710 user9 05-11-2022 00:21 00:08:49:43 10.55.4.159 DA:2F:97:0E:B7:D0
4711 user9 05-11-2022 20:55 00:01:06:20 10.55.2.33 DA:2F:97:0E:B7:D0

upload download total_transfer seession_break_reason Unnamed: 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 NaN 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', 'seession_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', 'seession_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 float64
seession_break_reason object
unnamed: 9    float64
unnamed: 10   float64
dtype: object
```

```
internet_usage.isna().sum()
```

```
name          0
start_time    0
usage_time    0
ip            0
mac           0
upload        0
```

Saving...

```
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
ip            0.0
mac           0.0
upload        0.0
download      0.0
total_transfer 0.0
seession_break_reason 0.0
unnamed: 9    0.0
unnamed: 10   0.0
dtype: float64
```

```
internet_usage.duplicated().sum()
```

```
0
```

```
internet_usage.describe(include='all', datetime_is_numeric=True)
```

	name	start_time	usage_time	ip	mac	upload	download	total_transfer	seession_break_reason	unnamed: 9	unnamed: 10
count	0	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	NaN	NaN
top	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

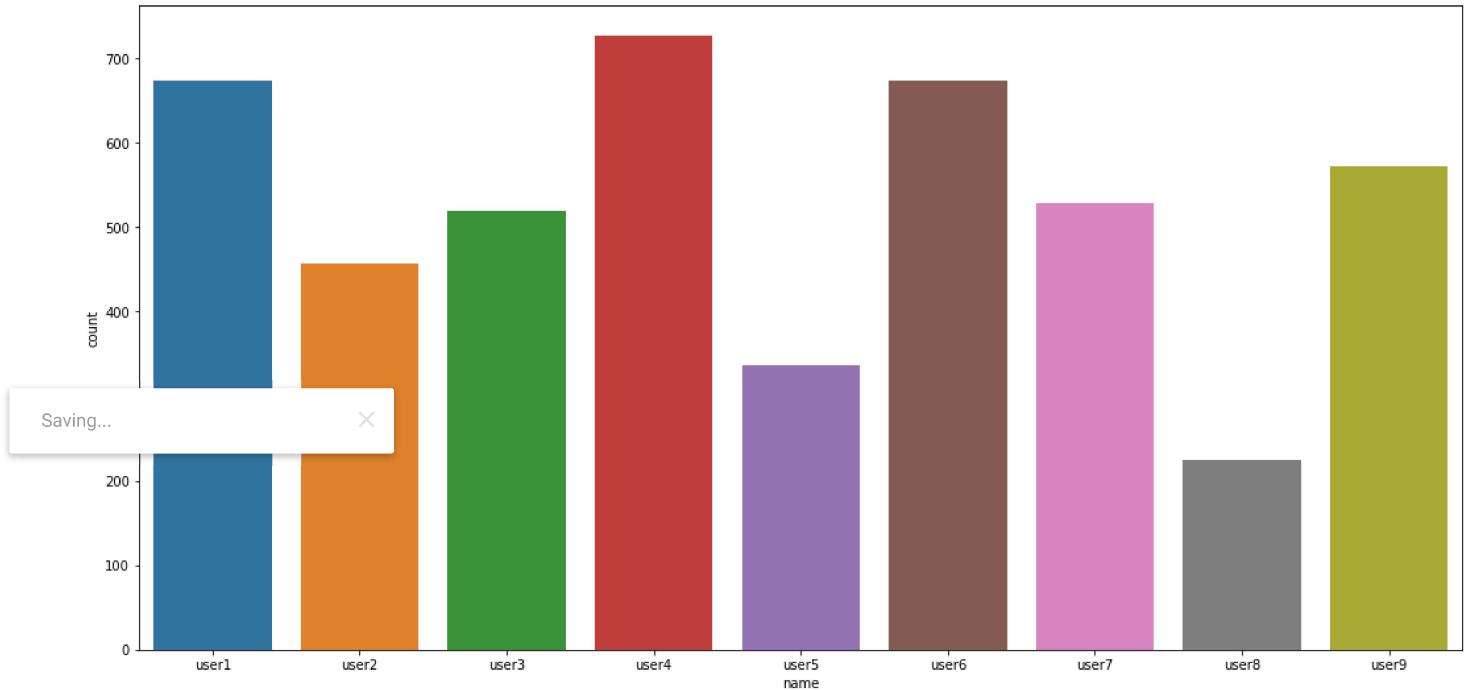
```
internet_usage.name.value_counts()
```

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

```
min NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN
```

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

User Count



<Figure size 432x288 with 0 Axes>

```
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:
01:00:21:07
```

```
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
user1 00:00:00:18
```

```

user2 00:00:00:08
user3 00:00:00:01
user4 00:00:00:45
user5 00:00:01:07
user6 00:00:00:18
user7 00:00:00:20
user8 00:00:00:20
user9 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
user1 00:19:35:11
user2 00:20:39:52
user3 00:17:01:28
user4 01:00:21:07
user5 00:06:36:11
user6 00:19:35:11
user7 00:22:00:07
user8 00:17:24:26
user9 00:19:26:09
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

```

```
10.55.14.47    1
```

Saving...

```

10.55.12.190  1
10.55.2.33    1

```

```
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: 10004.48Kb
The maximum upload is: 9994.24Kb

```

```

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

```

```

The minimum upload per user:
name
user1 1006.74
user2 10117.12
user3 10055.68
user4 10147.84
user5 10076.16
user6 1006.74
user7 10035.2
user8 10004.48
user9 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: 100014.08Kb
The maximum download is: 999393.28Kb
```

```
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
user1  999393.28
user2  99676.16
user3  99635.2
user4  9953.28
user5  990812.16
user6  999393.28
user7  991227.04
```

Saving...

```
Name: download, dtype: object
```

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

```
The minimum total transfer is: 1.12Kb
The maximum total transfer is: 28552724.48Kb
```

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

```
The minimum total transfer per user:
name
user1  75.34
user2  73.13
user3  1.12
user4  13.45
user5  924.40
user6  75.34
user7  15.25
user8  18.08
user9  102.64
Name: total_transfer, dtype: float64
```

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

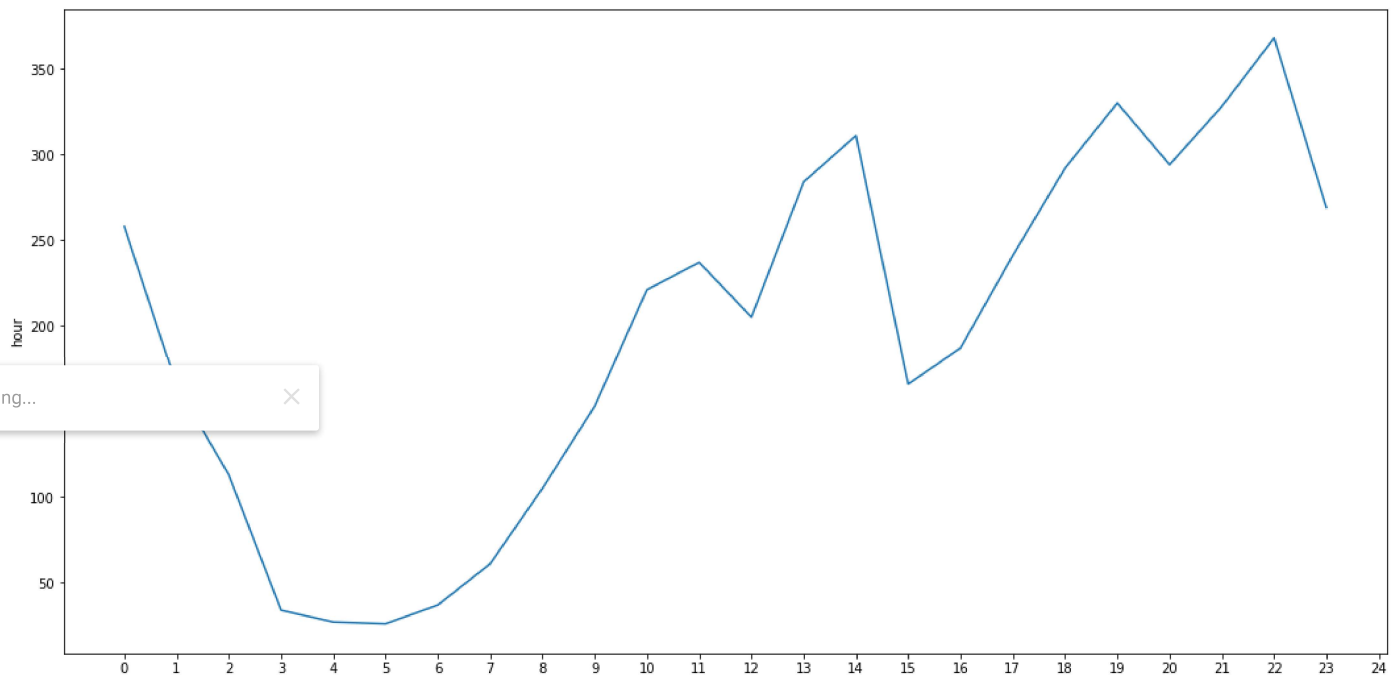
```
The maximum total transfer per user:
name
user1  24389877.76
user2  22051553.28
user3  3282042.88
user4  8524922.88
user5  5158993.92
user6  24389877.76
user7  28552724.48
user8  3166699.52
user9  6155141.12
Name: total_transfer, dtype: float64
```

```
internet_usage.seession_break_reason.value_counts()
```

```
Idle-Timeout    4350
Lost-Carrier    162
Lost-Service    124
User-Request    65
NAS-Reboot      2
Name: seession_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(figsize=(18, 9))
sns.lineplot(data=frequent_activity_time_of_day)
plt.xticks(np.linspace(start=0, stop=24, num=25))
plt.show()
plt.clf()
```



<Figure size 432x288 with 0 Axes>

How often the ip changes ?

```
base_IP = '48:E7:DA: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) + ' times')

The IP Adress changed 2308 times
```

The dataset contains 9 users that used 1224 difference devices to connect to the internet while uploading 2841640.0Kb and downloading 27902607.0Kb with a total transfer of 28552724.48Kb during a period of 7 months

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

The IP Adress changed 2303 times while the devices used changed 1223 times

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

 0s completed at 11:04 AM



Saving... 