```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

internet_usage = pd.read_csv('C:\Users\KNOT\Desktop\Assignments\internet_session.csv', parse_dates=['start_time'
])
internet_usage
internet_usage.shape
internet_usage.columns
internet_usage.columns = internet_usage.columns.str.lower()
internet_usage.columns
#Now let's check the data type of the columns
internet_usage.dtypes
#check the null values and drop them if necessary
internet_usage.isna().sum()
#column "seession_break_reason" has some null values but since it's a very low amount, we can safely delete
internet_usage = internet_usage.dropna().copy()
internet_usage.isna().sum()
# dataset contains duplicates and drop them if that's the case
internet_usage.duplicated().sum()
#tart converting the right columns from strings to numeric
internet_usage['usage_time'] = internet_usage['usage_time'].str.replace('00:', '', 1)
internet_usage['usage_time'] = pd.to_datetime(internet_usage['usage_time'])

internet_usage['upload'] = internet_usage['upload'].str.extract('(\d+)', expand=False)
internet_usage.upload = internet_usage.upload.astype(float)

internet_usage['download'] = internet_usage['download'].str.extract('(\d+)', expand=False)
internet_usage.download = internet_usage.download.astype(float)

device = []
basename = 'device'
mac = internet_usage['mac'][0]
device_number = 1
for i in internet_usage['mac']:
    if i == mac:
        device.append(basename + str(device_number))
    else:
        device_number += 1
        device.append(basename + str(device_number))
        mac = i
internet_usage['device'] = device

internet_usage.dtypes

#Exploratory Data Analysis
internet_usage.describe(include='all', datetime_is_numeric=True)
#users
internet_usage.name.value_counts()

plt.figure(figsize=(18, 9))
ax = sns.countplot(x='name', data=internet_usage)
```

```
ax.bar_label(ax.containers[0])
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())

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

print('The minimum usage time per user:')
usage_time_minimum = internet_usage.groupby('name').usage_time.min()
usage_time_minimum

plt.figure(figsize=(18, 9))
usage_time_minimum.plot(kind='bar', logy=True)
plt.title("The minimum usage time per user")
plt.show()
plt.clf()

print('The maximum usage time per user:')
usage_time_maximum = internet_usage.groupby('name').usage_time.max()
usage_time_maximum

plt.figure(figsize=(18, 9))
usage_time_maximum.plot(kind='bar', logy=True)
plt.title("The maximum usage time per user")
plt.show()
plt.clf()

print('The average usage time per user:')
usage_time_average = internet_usage.groupby('name').usage_time.mean()
usage_time_average

plt.figure(figsize=(18, 9))
usage_time_average.plot(kind='bar', logy=True)
plt.title("The average usage time per user")
plt.show()
plt.clf()

# User 8 has the most usage with 4 hours 3 minutes and 14 seconds, while User 5 has the least amount of average ti
me usage with one hour 20 minutes and 11 seconds

internet_usage.ip.value_counts()
#The most used IP Adress is 10:55:0:89

internet_usage.device.value_counts()
#The most used device is device1206 with 194 times
```

```
print('The minimum upload is: ' + str(internet_usage.upload.min()) + 'Kb')
print('The maximum upload is: ' + str(internet_usage.upload.max()) + 'Kb')
print('The average upload is: ' + str(round(internet_usage.upload.mean(), 2)) + 'Kb')

#The minimum upload is: 2.0Kb
#The maximum upload is: 2841640.0Kb
#The average upload is: 33787.02Kb

# minimum, maximum and average upload per user

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

plt.figure(figsize=(18, 9))
ax = sns.barplot(x='name', y='upload' , data=internet_usage, ci=None, estimator=np.min)
ax.bar_label(ax.containers[0])
plt.title("Minimum upload per user")
plt.show()
plt.clf()
```

user 2 has the lowest minimum upload with 2Kb while user5 has the highest minimum upload with 382Kb

```
#Now we check the maximum upload per user

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

plt.figure(figsize=(18, 9))
ax = sns.barplot(x='name', y='upload' , data=internet_usage, ci=None, estimator=np.max)
ax.bar_label(ax.containers[0])
plt.title("Maximum upload per user")
plt.show()
plt.clf()
```

User 5 has the highest maximum upload with 2841640Kb with user 2 having the lowest of maximum upload with 37 9955Kb

```
#We will check the average upload per user

print('The average upload per user:')
round(internet_usage.groupby('name').upload.mean(), 2)
```

alculating minimum, maximum and average total, and after that per user

```
print('The minimum download is: ' + str(internet_usage.download.min()) + 'Kb')
print('The maximum download is: ' + str(internet_usage.download.max()) + 'Kb')
print('The average download is: ' + str(round(internet_usage.download.mean(), 2)) + 'Kb')
```