Importing Libraries

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

import warnings
warnings.filterwarnings("ignore")

data=pd.read_csv('telcom_data.csv')

data
```

```
           Bearer Id            Start  Start ms              End  End
ms  \
0       1.311448e+19   4/4/2019 12:01     770.0  4/25/2019 14:35
662.0
1       1.311448e+19   4/9/2019 13:04     235.0   4/25/2019 8:15
606.0
2       1.311448e+19   4/9/2019 17:42       1.0  4/25/2019 11:58
652.0
3       1.311448e+19   4/10/2019 0:31     486.0   4/25/2019 7:36
171.0
4       1.311448e+19  4/12/2019 20:10     565.0  4/25/2019 10:40
954.0
...              ...              ...       ...              ...     .
..
149996  7.277826e+18   4/29/2019 7:28     451.0   4/30/2019 6:02
214.0
149997  7.349883e+18   4/29/2019 7:28     483.0  4/30/2019 10:41
187.0
149998  1.311448e+19   4/29/2019 7:28     283.0  4/30/2019 10:46
810.0
149999  1.311448e+19   4/29/2019 7:28     696.0  4/30/2019 10:40
327.0
150000           NaN              NaN       NaN              NaN
NaN

        Dur. (ms)          IMSI  MSISDN/Number          IMEI  \
0       1823652.0  2.082014e+14   3.366496e+10  3.552121e+13
1       1365104.0  2.082019e+14   3.368185e+10  3.579401e+13
2       1361762.0  2.082003e+14   3.376063e+10  3.528151e+13
3       1321509.0  2.082014e+14   3.375034e+10  3.535661e+13
4       1089009.0  2.082014e+14   3.369980e+10  3.540701e+13
...           ...           ...            ...           ...
149996    81230.0  2.082022e+14   3.365069e+10  3.548311e+13
149997    97970.0  2.082019e+14   3.366345e+10  3.566051e+13
149998    98249.0  2.082017e+14   3.362189e+10  3.572121e+13
```

```
149999    97910.0  2.082021e+14   3.361962e+10  8.618620e+13
150000        NaN           NaN            NaN           NaN

           Last Location Name  ...  Youtube DL (Bytes)  Youtube UL
(Bytes)  \
0       9.16456699548519E+015  ...        1.585461e+07
2.501332e+06
1                      L77566A  ...        2.024740e+07
1.911173e+07
2                      D42335A  ...        1.972566e+07
1.469958e+07
3                      T21824A  ...        2.138812e+07
1.514664e+07
4                      D88865A  ...        1.525938e+07
1.896287e+07
...                        ...  ...                 ...
...
149996                 D20434A  ...        1.619167e+07
1.176343e+07
149997                 D10223C  ...        1.387723e+07
8.288284e+06
149998                 T51102A  ...        2.266051e+07
1.855903e+06
149999                 L88342B  ...        8.817106e+06
8.305402e+06
150000                     NaN  ...        1.163407e+07
1.100941e+07

        Netflix DL (Bytes)  Netflix UL (Bytes)  Gaming DL (Bytes)  \
0             8.198936e+06        9.656251e+06       2.780823e+08
1             1.833841e+07        1.722713e+07       6.087501e+08
2             1.758779e+07        6.163408e+06       2.295846e+08
3             1.399465e+07        1.097942e+06       7.995382e+08
4             1.712458e+07        4.152180e+05       5.277072e+08
...                    ...                 ...                ...
149996        1.788370e+07        1.967816e+07       5.266097e+08
149997        1.935015e+07        2.129315e+07       6.268931e+08
149998        9.963942e+06        5.065760e+06       5.535395e+08
149999        3.322253e+06        1.317259e+07       3.525370e+08
150000        1.162685e+07        1.100175e+07       4.220447e+08

        Gaming UL (Bytes)  Other DL (Bytes)  Other UL (Bytes)  \
0            1.434415e+07      1.717444e+08      8.814393e+06
1            1.170709e+06      5.269042e+08      1.505514e+07
2            3.956300e+05      4.106926e+08      4.215763e+06
3            1.084972e+07      7.490399e+08      1.279728e+07
4            3.529801e+06      5.507095e+08      1.391032e+07
...                   ...               ...               ...
149996       9.197207e+06      3.264510e+06      1.348742e+07
149997       4.735033e+06      7.121804e+08      2.457758e+06
```

```
149998       1.339432e+07      1.211009e+08      1.131473e+07
149999       2.529475e+06      8.147131e+08      1.406930e+06
150000       8.288398e+06      4.211005e+08      8.264799e+06

        Total UL (Bytes)  Total DL (Bytes)
0             36749741.0       308879636.0
1             53800391.0       653384965.0
2             27883638.0       279807335.0
3             43324218.0       846028530.0
4             38542814.0       569138589.0
...                  ...               ...
149996        57628851.0       574175259.0
149997        39135081.0       666648844.0
149998        34912224.0       592786405.0
149999        29626096.0       371895920.0
150000               NaN               NaN

[150001 rows x 55 columns]
```

```
data["Bearer Id"].unique
```

```
<bound method Series.unique of 0         1.311448e+19
1         1.311448e+19
2         1.311448e+19
3         1.311448e+19
4         1.311448e+19
             ...
149996    7.277826e+18
149997    7.349883e+18
149998    1.311448e+19
149999    1.311448e+19
150000             NaN
Name: Bearer Id, Length: 150001, dtype: float64>
```

```
data.head()
```

```
      Bearer Id            Start  Start ms              End  End ms  \
0  1.311448e+19   4/4/2019 12:01     770.0  4/25/2019 14:35   662.0
1  1.311448e+19   4/9/2019 13:04     235.0   4/25/2019 8:15   606.0
2  1.311448e+19   4/9/2019 17:42       1.0  4/25/2019 11:58   652.0
3  1.311448e+19   4/10/2019 0:31     486.0   4/25/2019 7:36   171.0
4  1.311448e+19  4/12/2019 20:10     565.0  4/25/2019 10:40   954.0

   Dur. (ms)          IMSI  MSISDN/Number          IMEI  \
0  1823652.0  2.082014e+14   3.366496e+10  3.552121e+13
1  1365104.0  2.082019e+14   3.368185e+10  3.579401e+13
2  1361762.0  2.082003e+14   3.376063e+10  3.528151e+13
3  1321509.0  2.082014e+14   3.375034e+10  3.535661e+13
4  1089009.0  2.082014e+14   3.369980e+10  3.540701e+13

      Last Location Name  ...  Youtube DL (Bytes)  Youtube UL (Bytes)
```

```
\
0  9.16456699548519E+015  ...          15854611.0           2501332.0

1                L77566A  ...          20247395.0          19111729.0

2                D42335A  ...          19725661.0          14699576.0

3                T21824A  ...          21388122.0          15146643.0

4                D88865A  ...          15259380.0          18962873.0


   Netflix DL (Bytes)  Netflix UL (Bytes)  Gaming DL (Bytes)  \
0           8198936.0           9656251.0        278082303.0
1          18338413.0          17227132.0        608750074.0
2          17587794.0           6163408.0        229584621.0
3          13994646.0           1097942.0        799538153.0
4          17124581.0            415218.0        527707248.0

   Gaming UL (Bytes)  Other DL (Bytes)  Other UL (Bytes)  Total UL
(Bytes)  \
0         14344150.0       171744450.0         8814393.0
36749741.0
1          1170709.0       526904238.0        15055145.0
53800391.0
2           395630.0       410692588.0         4215763.0
27883638.0
3         10849722.0       749039933.0        12797283.0
43324218.0
4          3529801.0       550709500.0        13910322.0
38542814.0

   Total DL (Bytes)
0       308879636.0
1       653384965.0
2       279807335.0
3       846028530.0
4       569138589.0

[5 rows x 55 columns]

data.tail()

           Bearer Id           Start  Start ms              End  End
ms  \
149996  7.277826e+18  4/29/2019 7:28     451.0   4/30/2019 6:02
214.0
149997  7.349883e+18  4/29/2019 7:28     483.0  4/30/2019 10:41
187.0
149998  1.311448e+19  4/29/2019 7:28     283.0  4/30/2019 10:46
```

```
810.0
149999  1.311448e+19  4/29/2019 7:28     696.0  4/30/2019 10:40
327.0
150000           NaN             NaN       NaN              NaN
NaN

        Dur. (ms)          IMSI  MSISDN/Number          IMEI  \
149996    81230.0  2.082022e+14   3.365069e+10  3.548311e+13
149997    97970.0  2.082019e+14   3.366345e+10  3.566051e+13
149998    98249.0  2.082017e+14   3.362189e+10  3.572121e+13
149999    97910.0  2.082021e+14   3.361962e+10  8.618620e+13
150000        NaN           NaN            NaN           NaN

       Last Location Name  ...  Youtube DL (Bytes)  Youtube UL (Bytes)
\
149996            D20434A  ...        1.619167e+07        1.176343e+07

149997            D10223C  ...        1.387723e+07        8.288284e+06

149998            T51102A  ...        2.266051e+07        1.855903e+06

149999            L88342B  ...        8.817106e+06        8.305402e+06

150000                NaN  ...        1.163407e+07        1.100941e+07


        Netflix DL (Bytes)  Netflix UL (Bytes)  Gaming DL (Bytes)  \
149996        1.788370e+07        1.967816e+07       5.266097e+08
149997        1.935015e+07        2.129315e+07       6.268931e+08
149998        9.963942e+06        5.065760e+06       5.535395e+08
149999        3.322253e+06        1.317259e+07       3.525370e+08
150000        1.162685e+07        1.100175e+07       4.220447e+08

        Gaming UL (Bytes)  Other DL (Bytes)  Other UL (Bytes)  \
149996       9.197207e+06      3.264510e+06      1.348742e+07
149997       4.735033e+06      7.121804e+08      2.457758e+06
149998       1.339432e+07      1.211009e+08      1.131473e+07
149999       2.529475e+06      8.147131e+08      1.406930e+06
150000       8.288398e+06      4.211005e+08      8.264799e+06

        Total UL (Bytes)  Total DL (Bytes)
149996        57628851.0       574175259.0
149997        39135081.0       666648844.0
149998        34912224.0       592786405.0
149999        29626096.0       371895920.0
150000               NaN               NaN

[5 rows x 55 columns]
```

```
data.shape
```

```
(150001, 55)
```

```
data.dtypes
```

```
Bearer Id                                    float64
Start                                         object
Start ms                                     float64
End                                           object
End ms                                       float64
Dur. (ms)                                    float64
IMSI                                         float64
MSISDN/Number                                float64
IMEI                                         float64
Last Location Name                            object
Avg RTT DL (ms)                              float64
Avg RTT UL (ms)                              float64
Avg Bearer TP DL (kbps)                      float64
Avg Bearer TP UL (kbps)                      float64
TCP DL Retrans. Vol (Bytes)                  float64
TCP UL Retrans. Vol (Bytes)                  float64
DL TP < 50 Kbps (%)                          float64
50 Kbps < DL TP < 250 Kbps (%)               float64
250 Kbps < DL TP < 1 Mbps (%)                float64
DL TP > 1 Mbps (%)                           float64
UL TP < 10 Kbps (%)                          float64
10 Kbps < UL TP < 50 Kbps (%)                float64
50 Kbps < UL TP < 300 Kbps (%)               float64
UL TP > 300 Kbps (%)                         float64
HTTP DL (Bytes)                              float64
HTTP UL (Bytes)                              float64
Activity Duration DL (ms)                    float64
Activity Duration UL (ms)                    float64
Dur. (ms).1                                  float64
Handset Manufacturer                          object
Handset Type                                  object
Nb of sec with 125000B < Vol DL              float64
Nb of sec with 1250B < Vol UL < 6250B        float64
Nb of sec with 31250B < Vol DL < 125000B     float64
Nb of sec with 37500B < Vol UL               float64
Nb of sec with 6250B < Vol DL < 31250B       float64
Nb of sec with 6250B < Vol UL < 37500B       float64
Nb of sec with Vol DL < 6250B                float64
Nb of sec with Vol UL < 1250B                float64
Social Media DL (Bytes)                      float64
Social Media UL (Bytes)                      float64
Google DL (Bytes)                            float64
Google UL (Bytes)                            float64
Email DL (Bytes)                             float64
Email UL (Bytes)                             float64
```

```
Youtube DL (Bytes)                          float64
Youtube UL (Bytes)                          float64
Netflix DL (Bytes)                          float64
Netflix UL (Bytes)                          float64
Gaming DL (Bytes)                           float64
Gaming UL (Bytes)                           float64
Other DL (Bytes)                            float64
Other UL (Bytes)                            float64
Total UL (Bytes)                            float64
Total DL (Bytes)                            float64
dtype: object
```

```
data.describe()
```

```
          Bearer Id       Start ms         End ms     Dur. (ms)
IMSI  \
count  1.490100e+05  150000.000000  150000.000000  1.500000e+05
1.494310e+05
mean   1.013887e+19     499.188200     498.800880  1.046086e+05
2.082016e+14
std    2.893173e+18     288.611834     288.097653  8.103762e+04
2.148809e+10
min    6.917538e+18       0.000000       0.000000  7.142000e+03
2.040471e+14
25%    7.349883e+18     250.000000     251.000000  5.744050e+04
2.082014e+14
50%    7.349883e+18     499.000000     500.000000  8.639900e+04
2.082015e+14
75%    1.304243e+19     749.000000     750.000000  1.324302e+05
2.082018e+14
max    1.318654e+19     999.000000     999.000000  1.859336e+06
2.140743e+14

       MSISDN/Number          IMEI  Avg RTT DL (ms)  Avg RTT UL
(ms)  \
count   1.489350e+05  1.494290e+05    122172.000000    122189.000000

mean    4.188282e+10  4.847455e+13       109.795706        17.662883

std     2.447443e+12  2.241637e+13       619.782739        84.793524

min     3.360100e+10  4.400152e+11         0.000000         0.000000

25%     3.365130e+10  3.546071e+13        32.000000         2.000000

50%     3.366371e+10  3.572201e+13        45.000000         5.000000

75%     3.368349e+10  8.611970e+13        70.000000        15.000000

max     8.823971e+14  9.900120e+13     96923.000000      7120.000000
```

```
       Avg Bearer TP DL (kbps)  ...  Youtube DL (Bytes)  Youtube UL
(Bytes)  \
count            150000.000000  ...        1.500010e+05
1.500010e+05
mean              13300.045927  ...        1.163407e+07
1.100941e+07
std               23971.878541  ...        6.710569e+06
6.345423e+06
min                   0.000000  ...        5.300000e+01
1.050000e+02
25%                  43.000000  ...        5.833501e+06
5.517965e+06
50%                  63.000000  ...        1.161602e+07
1.101345e+07
75%               19710.750000  ...        1.744852e+07
1.651556e+07
max              378160.000000  ...        2.325910e+07
2.201196e+07

       Netflix DL (Bytes)  Netflix UL (Bytes)  Gaming DL (Bytes)  \
count        1.500010e+05        1.500010e+05       1.500010e+05
mean         1.162685e+07        1.100175e+07       4.220447e+08
std          6.725218e+06        6.359490e+06       2.439675e+08
min          4.200000e+01        3.500000e+01       2.516000e+03
25%          5.777156e+06        5.475981e+06       2.104733e+08
50%          1.164222e+07        1.099638e+07       4.234081e+08
75%          1.747048e+07        1.650727e+07       6.331742e+08
max          2.325919e+07        2.201196e+07       8.434419e+08

       Gaming UL (Bytes)  Other DL (Bytes)  Other UL (Bytes)  \
count       1.500010e+05      1.500010e+05      1.500010e+05
mean        8.288398e+06      4.211005e+08      8.264799e+06
std         4.782700e+06      2.432050e+08      4.769004e+06
min         5.900000e+01      3.290000e+03      1.480000e+02
25%         4.128476e+06      2.101869e+08      4.145943e+06
50%         8.291208e+06      4.218030e+08      8.267071e+06
75%         1.243162e+07      6.316918e+08      1.238415e+07
max         1.655879e+07      8.434425e+08      1.655882e+07

       Total UL (Bytes)  Total DL (Bytes)
count      1.500000e+05      1.500000e+05
mean       4.112121e+07      4.546434e+08
std        1.127639e+07      2.441429e+08
min        2.866892e+06      7.114041e+06
25%        3.322201e+07      2.431068e+08
50%        4.114331e+07      4.558411e+08
75%        4.903424e+07      6.657055e+08
max        7.833131e+07      9.029696e+08
```

```
[8 rows x 50 columns]

data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150001 entries, 0 to 150000
Data columns (total 55 columns):
 #   Column                                    Non-Null Count   Dtype

---  ------                                    --------------   -----

 0   Bearer Id                                 149010 non-null
float64
 1   Start                                     150000 non-null  object

 2   Start ms                                  150000 non-null
float64
 3   End                                       150000 non-null  object

 4   End ms                                    150000 non-null
float64
 5   Dur. (ms)                                 150000 non-null
float64
 6   IMSI                                      149431 non-null
float64
 7   MSISDN/Number                             148935 non-null
float64
 8   IMEI                                      149429 non-null
float64
 9   Last Location Name                        148848 non-null  object

 10  Avg RTT DL (ms)                           122172 non-null
float64
 11  Avg RTT UL (ms)                           122189 non-null
float64
 12  Avg Bearer TP DL (kbps)                   150000 non-null
float64
 13  Avg Bearer TP UL (kbps)                   150000 non-null
float64
 14  TCP DL Retrans. Vol (Bytes)               61855 non-null
float64
 15  TCP UL Retrans. Vol (Bytes)               53352 non-null
float64
 16  DL TP < 50 Kbps (%)                       149247 non-null
float64
 17  50 Kbps < DL TP < 250 Kbps (%)            149247 non-null
float64
 18  250 Kbps < DL TP < 1 Mbps (%)             149247 non-null
float64
```

```
 19  DL TP > 1 Mbps (%)                          149247 non-null
float64
 20  UL TP < 10 Kbps (%)                         149209 non-null
float64
 21  10 Kbps < UL TP < 50 Kbps (%)               149209 non-null
float64
 22  50 Kbps < UL TP < 300 Kbps (%)              149209 non-null
float64
 23  UL TP > 300 Kbps (%)                        149209 non-null
float64
 24  HTTP DL (Bytes)                             68527 non-null
float64
 25  HTTP UL (Bytes)                             68191 non-null
float64
 26  Activity Duration DL (ms)                   150000 non-null
float64
 27  Activity Duration UL (ms)                   150000 non-null
float64
 28  Dur. (ms).1                                 150000 non-null
float64
 29  Handset Manufacturer                        149429 non-null  object

 30  Handset Type                                149429 non-null  object

 31  Nb of sec with 125000B < Vol DL             52463 non-null
float64
 32  Nb of sec with 1250B < Vol UL < 6250B       57107 non-null
float64
 33  Nb of sec with 31250B < Vol DL < 125000B    56415 non-null
float64
 34  Nb of sec with 37500B < Vol UL              19747 non-null
float64
 35  Nb of sec with 6250B < Vol DL < 31250B      61684 non-null
float64
 36  Nb of sec with 6250B < Vol UL < 37500B      38158 non-null
float64
 37  Nb of sec with Vol DL < 6250B               149246 non-null
float64
 38  Nb of sec with Vol UL < 1250B               149208 non-null
float64
 39  Social Media DL (Bytes)                     150001 non-null
float64
 40  Social Media UL (Bytes)                     150001 non-null
float64
 41  Google DL (Bytes)                           150001 non-null
float64
 42  Google UL (Bytes)                           150001 non-null
float64
 43  Email DL (Bytes)                            150001 non-null
float64
```

```
 44  Email UL (Bytes)                          150001 non-null
float64
 45  Youtube DL (Bytes)                        150001 non-null
float64
 46  Youtube UL (Bytes)                        150001 non-null
float64
 47  Netflix DL (Bytes)                        150001 non-null
float64
 48  Netflix UL (Bytes)                        150001 non-null
float64
 49  Gaming DL (Bytes)                         150001 non-null
float64
 50  Gaming UL (Bytes)                         150001 non-null
float64
 51  Other DL (Bytes)                          150001 non-null
float64
 52  Other UL (Bytes)                          150001 non-null
float64
 53  Total UL (Bytes)                          150000 non-null
float64
 54  Total DL (Bytes)                          150000 non-null
float64
dtypes: float64(50), object(5)
memory usage: 62.9+ MB
```

```
data.duplicated()
```

```
0         False
1         False
2         False
3         False
4         False
          ...
149996    False
149997    False
149998    False
149999    False
150000    False
Length: 150001, dtype: bool
```

```
null_pct= data.isnull().sum() / len(data) * 100
```

```
null_pct
```

```
Bearer Id                                    0.660662
Start                                        0.000667
Start ms                                     0.000667
End                                          0.000667
End ms                                       0.000667
Dur. (ms)                                    0.000667
IMSI                                         0.379997
```

```
MSISDN/Number                                  0.710662
IMEI                                           0.381331
Last Location Name                             0.768662
Avg RTT DL (ms)                               18.552543
Avg RTT UL (ms)                               18.541210
Avg Bearer TP DL (kbps)                        0.000667
Avg Bearer TP UL (kbps)                        0.000667
TCP DL Retrans. Vol (Bytes)                   58.763608
TCP UL Retrans. Vol (Bytes)                   64.432237
DL TP < 50 Kbps (%)                            0.502663
50 Kbps < DL TP < 250 Kbps (%)                 0.502663
250 Kbps < DL TP < 1 Mbps (%)                  0.502663
DL TP > 1 Mbps (%)                             0.502663
UL TP < 10 Kbps (%)                            0.527996
10 Kbps < UL TP < 50 Kbps (%)                  0.527996
50 Kbps < UL TP < 300 Kbps (%)                 0.527996
UL TP > 300 Kbps (%)                           0.527996
HTTP DL (Bytes)                               54.315638
HTTP UL (Bytes)                               54.539636
Activity Duration DL (ms)                      0.000667
Activity Duration UL (ms)                      0.000667
Dur. (ms).1                                    0.000667
Handset Manufacturer                           0.381331
Handset Type                                   0.381331
Nb of sec with 125000B < Vol DL               65.024900
Nb of sec with 1250B < Vol UL < 6250B         61.928920
Nb of sec with 31250B < Vol DL < 125000B      62.390251
Nb of sec with 37500B < Vol UL                86.835421
Nb of sec with 6250B < Vol DL < 31250B        58.877607
Nb of sec with 6250B < Vol UL < 37500B        74.561503
Nb of sec with Vol DL < 6250B                  0.503330
Nb of sec with Vol UL < 1250B                  0.528663
Social Media DL (Bytes)                        0.000000
Social Media UL (Bytes)                        0.000000
Google DL (Bytes)                              0.000000
Google UL (Bytes)                              0.000000
Email DL (Bytes)                               0.000000
Email UL (Bytes)                               0.000000
Youtube DL (Bytes)                             0.000000
Youtube UL (Bytes)                             0.000000
Netflix DL (Bytes)                             0.000000
Netflix UL (Bytes)                             0.000000
Gaming DL (Bytes)                              0.000000
Gaming UL (Bytes)                              0.000000
Other DL (Bytes)                               0.000000
Other UL (Bytes)                               0.000000
Total UL (Bytes)                               0.000667
Total DL (Bytes)                               0.000667
dtype: float64
```

```
cols_to_drop = null_pct[null_pct > 50].index.tolist()
cols_to_drop
```

```
['TCP DL Retrans. Vol (Bytes)',
 'TCP UL Retrans. Vol (Bytes)',
 'HTTP DL (Bytes)',
 'HTTP UL (Bytes)',
 'Nb of sec with 125000B < Vol DL',
 'Nb of sec with 1250B < Vol UL < 6250B',
 'Nb of sec with 31250B < Vol DL < 125000B',
 'Nb of sec with 37500B < Vol UL',
 'Nb of sec with 6250B < Vol DL < 31250B',
 'Nb of sec with 6250B < Vol UL < 37500B']
```

```
data=data.drop(cols_to_drop,axis=1)
data
```

```
           Bearer Id            Start  Start ms              End  End
ms  \
0       1.311448e+19   4/4/2019 12:01     770.0  4/25/2019 14:35
662.0
1       1.311448e+19   4/9/2019 13:04     235.0   4/25/2019 8:15
606.0
2       1.311448e+19   4/9/2019 17:42       1.0  4/25/2019 11:58
652.0
3       1.311448e+19   4/10/2019 0:31     486.0   4/25/2019 7:36
171.0
4       1.311448e+19  4/12/2019 20:10     565.0  4/25/2019 10:40
954.0
...              ...              ...       ...              ...     .
..
149996  7.277826e+18   4/29/2019 7:28     451.0   4/30/2019 6:02
214.0
149997  7.349883e+18   4/29/2019 7:28     483.0  4/30/2019 10:41
187.0
149998  1.311448e+19   4/29/2019 7:28     283.0  4/30/2019 10:46
810.0
149999  1.311448e+19   4/29/2019 7:28     696.0  4/30/2019 10:40
327.0
150000           NaN              NaN       NaN              NaN
NaN

        Dur. (ms)          IMSI  MSISDN/Number          IMEI  \
0       1823652.0  2.082014e+14   3.366496e+10  3.552121e+13
1       1365104.0  2.082019e+14   3.368185e+10  3.579401e+13
2       1361762.0  2.082003e+14   3.376063e+10  3.528151e+13
3       1321509.0  2.082014e+14   3.375034e+10  3.535661e+13
4       1089009.0  2.082014e+14   3.369980e+10  3.540701e+13
...           ...           ...            ...           ...
149996    81230.0  2.082022e+14   3.365069e+10  3.548311e+13
149997    97970.0  2.082019e+14   3.366345e+10  3.566051e+13
```

```
149998    98249.0  2.082017e+14   3.362189e+10  3.572121e+13
149999    97910.0  2.082021e+14   3.361962e+10  8.618620e+13
150000        NaN           NaN            NaN           NaN

           Last Location Name  ...  Youtube DL (Bytes)  Youtube UL
(Bytes)  \
0       9.16456699548519E+015  ...        1.585461e+07
2.501332e+06
1                   L77566A  ...        2.024740e+07
1.911173e+07
2                   D42335A  ...        1.972566e+07
1.469958e+07
3                   T21824A  ...        2.138812e+07
1.514664e+07
4                   D88865A  ...        1.525938e+07
1.896287e+07
...                     ...  ...                 ...
...
149996              D20434A  ...        1.619167e+07
1.176343e+07
149997              D10223C  ...        1.387723e+07
8.288284e+06
149998              T51102A  ...        2.266051e+07
1.855903e+06
149999              L88342B  ...        8.817106e+06
8.305402e+06
150000                  NaN  ...        1.163407e+07
1.100941e+07

        Netflix DL (Bytes)  Netflix UL (Bytes)  Gaming DL (Bytes)  \
0             8.198936e+06        9.656251e+06       2.780823e+08
1             1.833841e+07        1.722713e+07       6.087501e+08
2             1.758779e+07        6.163408e+06       2.295846e+08
3             1.399465e+07        1.097942e+06       7.995382e+08
4             1.712458e+07        4.152180e+05       5.277072e+08
...                    ...                 ...                ...
149996        1.788370e+07        1.967816e+07       5.266097e+08
149997        1.935015e+07        2.129315e+07       6.268931e+08
149998        9.963942e+06        5.065760e+06       5.535395e+08
149999        3.322253e+06        1.317259e+07       3.525370e+08
150000        1.162685e+07        1.100175e+07       4.220447e+08

        Gaming UL (Bytes)  Other DL (Bytes)  Other UL (Bytes)  \
0            1.434415e+07      1.717444e+08      8.814393e+06
1            1.170709e+06      5.269042e+08      1.505514e+07
2            3.956300e+05      4.106926e+08      4.215763e+06
3            1.084972e+07      7.490399e+08      1.279728e+07
4            3.529801e+06      5.507095e+08      1.391032e+07
...                   ...               ...               ...
149996       9.197207e+06      3.264510e+06      1.348742e+07
```

```
149997       4.735033e+06      7.121804e+08      2.457758e+06
149998       1.339432e+07      1.211009e+08      1.131473e+07
149999       2.529475e+06      8.147131e+08      1.406930e+06
150000       8.288398e+06      4.211005e+08      8.264799e+06

        Total UL (Bytes)  Total DL (Bytes)
0             36749741.0       308879636.0
1             53800391.0       653384965.0
2             27883638.0       279807335.0
3             43324218.0       846028530.0
4             38542814.0       569138589.0
...                  ...               ...
149996        57628851.0       574175259.0
149997        39135081.0       666648844.0
149998        34912224.0       592786405.0
149999        29626096.0       371895920.0
150000               NaN               NaN

[150001 rows x 45 columns]
```

```
data.isnull().sum()
```

```
Bearer Id                           991
Start                                 1
Start ms                              1
End                                   1
End ms                                1
Dur. (ms)                             1
IMSI                                570
MSISDN/Number                      1066
IMEI                                572
Last Location Name                 1153
Avg RTT DL (ms)                   27829
Avg RTT UL (ms)                   27812
Avg Bearer TP DL (kbps)               1
Avg Bearer TP UL (kbps)               1
DL TP < 50 Kbps (%)                 754
50 Kbps < DL TP < 250 Kbps (%)      754
250 Kbps < DL TP < 1 Mbps (%)       754
DL TP > 1 Mbps (%)                  754
UL TP < 10 Kbps (%)                 792
10 Kbps < UL TP < 50 Kbps (%)       792
50 Kbps < UL TP < 300 Kbps (%)      792
UL TP > 300 Kbps (%)                792
Activity Duration DL (ms)             1
Activity Duration UL (ms)             1
Dur. (ms).1                           1
Handset Manufacturer                572
Handset Type                        572
Nb of sec with Vol DL < 6250B       755
```

```
Nb of sec with Vol UL < 1250B       793
Social Media DL (Bytes)               0
Social Media UL (Bytes)               0
Google DL (Bytes)                     0
Google UL (Bytes)                     0
Email DL (Bytes)                      0
Email UL (Bytes)                      0
Youtube DL (Bytes)                    0
Youtube UL (Bytes)                    0
Netflix DL (Bytes)                    0
Netflix UL (Bytes)                    0
Gaming DL (Bytes)                     0
Gaming UL (Bytes)                     0
Other DL (Bytes)                      0
Other UL (Bytes)                      0
Total UL (Bytes)                      1
Total DL (Bytes)                      1
dtype: int64
```

```
data=data.dropna(how='any',axis=0)
data
```

```
           Bearer Id           Start  Start ms              End  End
ms  \
0       1.311448e+19  4/4/2019 12:01     770.0  4/25/2019 14:35
662.0
1       1.311448e+19  4/9/2019 13:04     235.0   4/25/2019 8:15
606.0
6       1.311448e+19  4/13/2019 8:41     612.0   4/25/2019 8:16
168.0
7       1.304243e+19  4/14/2019 2:11     592.0   4/25/2019 2:26
512.0
9       1.304243e+19  4/15/2019 0:32       0.0   4/25/2019 0:40
284.0
...              ...             ...       ...              ...     ..
.
149995  1.304243e+19  4/29/2019 7:28     615.0   4/30/2019 0:01
407.0
149996  7.277826e+18  4/29/2019 7:28     451.0   4/30/2019 6:02
214.0
149997  7.349883e+18  4/29/2019 7:28     483.0  4/30/2019 10:41
187.0
149998  1.311448e+19  4/29/2019 7:28     283.0  4/30/2019 10:46
810.0
149999  1.311448e+19  4/29/2019 7:28     696.0  4/30/2019 10:40
327.0

        Dur. (ms)          IMSI  MSISDN/Number          IMEI  \
0       1823652.0  2.082014e+14   3.366496e+10  3.552121e+13
1       1365104.0  2.082019e+14   3.368185e+10  3.579401e+13
```

```
6       1035261.0  2.082014e+14   3.366537e+10  8.676270e+13
7        951292.0  2.082010e+14   3.376349e+10  8.654640e+13
9        864482.0  2.082003e+14   3.365922e+10  3.557311e+13
...           ...           ...            ...           ...
149995    59587.0  2.082014e+14   3.366865e+10  3.533251e+13
149996    81230.0  2.082022e+14   3.365069e+10  3.548311e+13
149997    97970.0  2.082019e+14   3.366345e+10  3.566051e+13
149998    98249.0  2.082017e+14   3.362189e+10  3.572121e+13
149999    97910.0  2.082021e+14   3.361962e+10  8.618620e+13

           Last Location Name  ...  Youtube DL (Bytes)  Youtube UL
(Bytes)  \
0       9.16456699548519E+015  ...           15854611.0
2501332.0
1                      L77566A  ...           20247395.0
19111729.0
6       9.16456701058919E+015  ...            7234465.0
1886295.0
7        CELL_208_20_520025_1  ...           12189786.0
19735183.0
9                      L20434C  ...           16338345.0
6573194.0
...                        ...  ...                  ...
...
149995                 T85721A  ...            6550499.0
18003146.0
149996                 D20434A  ...           16191667.0
11763428.0
149997                 D10223C  ...           13877234.0
8288284.0
149998                 T51102A  ...           22660510.0
1855903.0
149999                 L88342B  ...            8817106.0
8305402.0

        Netflix DL (Bytes)  Netflix UL (Bytes)  Gaming DL (Bytes)  \
0                8198936.0           9656251.0        278082303.0
1               18338413.0          17227132.0        608750074.0
6               13231006.0          17981760.0         44796223.0
7                6858496.0          12075206.0        748906189.0
9               12211149.0            974201.0        811355188.0
...                    ...                 ...                ...
149995          22468983.0           7149728.0        784435351.0
149996          17883703.0          19678161.0        526609673.0
149997          19350146.0          21293148.0        626893062.0
149998           9963942.0           5065760.0        553539484.0
149999           3322253.0          13172589.0        352536971.0

        Gaming UL (Bytes)  Other DL (Bytes)  Other UL (Bytes)  \
0              14344150.0       171744450.0         8814393.0
```

```
1               1170709.0       526904238.0        15055145.0
6              14084934.0       803653557.0        10649768.0
7              16149622.0       537601820.0         7714657.0
9               1591237.0        64856042.0        12815208.0
...                   ...              ...               ...
149995         12724751.0       321383162.0        14890486.0
149996          9197207.0         3264510.0        13487416.0
149997          4735033.0       712180387.0         2457758.0
149998         13394316.0       121100856.0        11314729.0
149999          2529475.0       814713113.0         1406930.0

        Total UL (Bytes)  Total DL (Bytes)
0             36749741.0       308879636.0
1             53800391.0       653384965.0
6             47925246.0        70562047.0
7             58813016.0       775350343.0
9             22417975.0       850570347.0
...                  ...              ...
149995        53745392.0       819016468.0
149996        57628851.0       574175259.0
149997        39135081.0       666648844.0
149998        34912224.0       592786405.0
149999        29626096.0       371895920.0

[120739 rows x 45 columns]
```

Task 1 - User Overview Analysis

• Start by identifying the top 10 handsets used by the customers.

• Then, identify the top 3 handset manufacturers

• Next, identify the top 5 handsets per top 3 handset manufacturer

• Make a short interpretation and recommendation to marketing teams

```
top_10_handsets = data["Handset
Type"].value_counts(ascending=False).head(10)
print(top_10_handsets)
```

```
Huawei B528S-23A                19310
Apple iPhone 6S (A1688)          9244
Apple iPhone 6 (A1586)           8786
Apple iPhone 7 (A1778)           6134
Apple iPhone Se (A1723)          5062
Apple iPhone 8 (A1905)           4900
undefined                        4816
Apple iPhone Xr (A2105)          4490
Apple iPhone X (A1901)           3746
```

```
Apple iPhone 8 Plus (A1897)     2968
Name: Handset Type, dtype: int64
```

```
top_3_Manufacturer = data["Handset
Manufacturer"].value_counts(ascending=False).head(3)
print(top_3_Manufacturer )
```

```
Apple      58190
Huawei     33306
Samsung    19167
Name: Handset Manufacturer, dtype: int64
```

```
top_5_handsets_per_Manufacturer = {}
for Manufacturer in top_3_Manufacturer .index:
    Manufacturer_data = data[data['Handset Manufacturer'] ==
Manufacturer]
    top_5_handsets = Manufacturer_data['Handset
Type'].value_counts().head(5)
    top_5_handsets_per_Manufacturer[ Manufacturer] = top_5_handsets
print("top_5_handsets_per_Manufacturer:")
for Manufacturer, top_5_handsets in
top_5_handsets_per_Manufacturer.items():
    print("Manufacturer: {Manufacturer}")
    print(top_5_handsets)
```

```
top_5_handsets_per_Manufacturer:
Manufacturer: {Manufacturer}
Apple iPhone 6S (A1688)    9244
Apple iPhone 6 (A1586)     8786
Apple iPhone 7 (A1778)     6134
Apple iPhone Se (A1723)    5062
Apple iPhone 8 (A1905)     4900
Name: Handset Type, dtype: int64
Manufacturer: {Manufacturer}
Huawei B528S-23A                   19310
Huawei E5180                        2054
Huawei P20 Lite Huawei Nova 3E      1980
Huawei P20                          1394
Huawei Y6 2018                       975
Name: Handset Type, dtype: int64
Manufacturer: {Manufacturer}
Samsung Galaxy S8 (Sm-G950F)    1820
Samsung Galaxy J5 (Sm-J530)     1813
Samsung Galaxy J3 (Sm-J330)     1547
Samsung Galaxy A5 Sm-A520F      1488
Samsung Galaxy S7 (Sm-G930X)    1328
Name: Handset Type, dtype: int64
```

Task 1.1 - Your employer wants to have an overview of the users' behaviour on those applications.

• Aggregate per user the following information in the column

• number of xDR sessions

• Session duration

• the total download (DL) and upload (UL) data

• the total data volume (in Bytes) during this session for each application

```
categorical = [var for var in data.columns if data[var].dtypes=='O']
numerical = [var for var in data.columns if data[var].dtypes!='O']

categorical

['Start', 'End', 'Last Location Name', 'Handset Manufacturer',
'Handset Type']

numerical

['Bearer Id',
 'Start ms',
 'End ms',
 'Dur. (ms)',
 'IMSI',
 'MSISDN/Number',
 'IMEI',
 'Avg RTT DL (ms)',
 'Avg RTT UL (ms)',
 'Avg Bearer TP DL (kbps)',
 'Avg Bearer TP UL (kbps)',
 'DL TP < 50 Kbps (%)',
 '50 Kbps < DL TP < 250 Kbps (%)',
 '250 Kbps < DL TP < 1 Mbps (%)',
 'DL TP > 1 Mbps (%)',
 'UL TP < 10 Kbps (%)',
 '10 Kbps < UL TP < 50 Kbps (%)',
 '50 Kbps < UL TP < 300 Kbps (%)',
 'UL TP > 300 Kbps (%)',
 'Activity Duration DL (ms)',
 'Activity Duration UL (ms)',
 'Dur. (ms).1',
 'Nb of sec with Vol DL < 6250B',
 'Nb of sec with Vol UL < 1250B',
 'Social Media DL (Bytes)',
 'Social Media UL (Bytes)',
 'Google DL (Bytes)',
 'Google UL (Bytes)',
```

```
 'Email DL (Bytes)',
 'Email UL (Bytes)',
 'Youtube DL (Bytes)',
 'Youtube UL (Bytes)',
 'Netflix DL (Bytes)',
 'Netflix UL (Bytes)',
 'Gaming DL (Bytes)',
 'Gaming UL (Bytes)',
 'Other DL (Bytes)',
 'Other UL (Bytes)',
 'Total UL (Bytes)',
 'Total DL (Bytes)']
```

```
data.columns
```

```
Index(['Bearer Id', 'Start', 'Start ms', 'End', 'End ms', 'Dur. (ms)',
'IMSI',
       'MSISDN/Number', 'IMEI', 'Last Location Name', 'Avg RTT DL
(ms)',
       'Avg RTT UL (ms)', 'Avg Bearer TP DL (kbps)', 'Avg Bearer TP UL
(kbps)',
       'DL TP < 50 Kbps (%)', '50 Kbps < DL TP < 250 Kbps (%)',
       '250 Kbps < DL TP < 1 Mbps (%)', 'DL TP > 1 Mbps (%)',
       'UL TP < 10 Kbps (%)', '10 Kbps < UL TP < 50 Kbps (%)',
       '50 Kbps < UL TP < 300 Kbps (%)', 'UL TP > 300 Kbps (%)',
       'Activity Duration DL (ms)', 'Activity Duration UL (ms)', 'Dur.
(ms).1',
       'Handset Manufacturer', 'Handset Type', 'Nb of sec with Vol DL
< 6250B',
       'Nb of sec with Vol UL < 1250B', 'Social Media DL (Bytes)',
       'Social Media UL (Bytes)', 'Google DL (Bytes)', 'Google UL
(Bytes)',
       'Email DL (Bytes)', 'Email UL (Bytes)', 'Youtube DL (Bytes)',
       'Youtube UL (Bytes)', 'Netflix DL (Bytes)', 'Netflix UL
(Bytes)',
       'Gaming DL (Bytes)', 'Gaming UL (Bytes)', 'Other DL (Bytes)',
       'Other UL (Bytes)', 'Total UL (Bytes)', 'Total DL (Bytes)'],
      dtype='object')
```

```
for i in data.columns:
    print(i)
```

```
Bearer Id
Start
Start ms
End
End ms
Dur. (ms)
IMSI
MSISDN/Number
IMEI
```

```
Last Location Name
Avg RTT DL (ms)
Avg RTT UL (ms)
Avg Bearer TP DL (kbps)
Avg Bearer TP UL (kbps)
DL TP < 50 Kbps (%)
50 Kbps < DL TP < 250 Kbps (%)
250 Kbps < DL TP < 1 Mbps (%)
DL TP > 1 Mbps (%)
UL TP < 10 Kbps (%)
10 Kbps < UL TP < 50 Kbps (%)
50 Kbps < UL TP < 300 Kbps (%)
UL TP > 300 Kbps (%)
Activity Duration DL (ms)
Activity Duration UL (ms)
Dur. (ms).1
Handset Manufacturer
Handset Type
Nb of sec with Vol DL < 6250B
Nb of sec with Vol UL < 1250B
Social Media DL (Bytes)
Social Media UL (Bytes)
Google DL (Bytes)
Google UL (Bytes)
Email DL (Bytes)
Email UL (Bytes)
Youtube DL (Bytes)
Youtube UL (Bytes)
Netflix DL (Bytes)
Netflix UL (Bytes)
Gaming DL (Bytes)
Gaming UL (Bytes)
Other DL (Bytes)
Other UL (Bytes)
Total UL (Bytes)
Total DL (Bytes)
```

```
user_behaviour=data.groupby('Bearer Id').agg({
                    'Dur. (ms)':'sum',
                      'Activity Duration DL (ms)':'sum',
                      'Activity Duration UL (ms)':'sum',
                      'Social Media DL (Bytes)':'sum',
                      'Social Media UL (Bytes)':'sum',
                      'Google DL (Bytes)':'sum',
                      'Google DL (Bytes)' : 'sum',

                      'Google UL (Bytes)' : 'sum',

                      'Email DL (Bytes)' : 'sum',
```

```
                    'Email UL (Bytes)' : 'sum',
                    'Youtube DL (Bytes)'  : 'sum',

                    'Youtube UL (Bytes)' : 'sum',

                    'Netflix DL (Bytes)'  : 'sum',

                    'Netflix UL (Bytes)' : 'sum',

                    'Gaming DL (Bytes)' : 'sum',

                    'Gaming UL (Bytes)' : 'sum',

                    'Other DL (Bytes)'  : 'sum',

                    'Other UL (Bytes)' :'sum',
}).reset_index()
print(user_behaviour)

           Bearer Id  Dur. (ms)  Activity Duration DL (ms)  \
0       6.917538e+18    24534.0                   131798.0
1       6.917538e+18    27786.0                   401941.0
2       6.917538e+18    15635.0                    73347.0
3       6.917538e+18    24264.0                   117340.0
4       6.917538e+18    15548.0                    76969.0
...              ...        ...                        ...
110427  1.318654e+19    80024.0                  2512362.0
110428  1.318654e+19   145291.0                     2067.0
110429  1.318654e+19    86399.0                  3968131.0
110430  1.318654e+19    86399.0                  1689999.0
110431  1.318654e+19   103113.0                        0.0

        Activity Duration UL (ms)  Social Media DL (Bytes)  \
0                        101470.0                2404741.0
1                        399092.0                 944612.0
2                         81378.0                1817239.0
3                        347852.0                1867318.0
4                         80241.0                1502346.0
...                           ...                      ...
110427                  2437668.0                3240226.0
110428                    45217.0                3062671.0
110429                  3537154.0                 720996.0
110430                  1513764.0                2492460.0
110431                    30367.0                1314234.0

        Social Media UL (Bytes)  Google DL (Bytes)  Google UL (Bytes)
\
0                        2410.0          5791591.0          2871336.0
```

```
1                     2827.0         10373157.0            56392.0
2                    19827.0           269988.0          3696393.0
3                    18928.0          1689296.0           195216.0
4                    37348.0          8917833.0           701876.0
...                      ...                ...                ...
110427               38284.0          2036152.0          2271168.0
110428               48953.0          9363661.0          4001970.0
110429               42836.0          1541915.0          2100839.0
110430               39905.0         11318188.0           466218.0
110431               27938.0          6969652.0          3756009.0
```

```
        Email DL (Bytes)  Email UL (Bytes)  Youtube DL (Bytes)  \
0               782388.0          806920.0           6139644.0
1               128003.0           34038.0           5385159.0
2              3191192.0          896670.0          12347020.0
3               740633.0          590043.0          15231815.0
4               332813.0          537902.0          14602694.0
...                  ...               ...                ...
110427         2410615.0          387548.0          12404964.0
110428         2192057.0          866373.0          22147919.0
110429         2315638.0          839789.0          11879062.0
110430         2612190.0          618629.0          22163800.0
110431         3317462.0          408257.0          12099319.0

        Youtube UL (Bytes)  Netflix DL (Bytes)  Netflix UL (Bytes)  \
0                2071526.0          19494278.0          14668354.0
1                4295851.0          15755839.0           1300571.0
2               11089528.0           2859358.0           1738176.0
3                8401567.0          21563985.0           2817981.0
4                1237356.0           4506463.0          17006591.0
...                    ...                 ...                ...
110427           4343114.0          11108134.0          21649273.0
110428           2152449.0          21468525.0           8603105.0
110429           1290963.0          22596930.0          11943452.0
110430          16763435.0           9522397.0           8346624.0
110431           1636122.0           9992219.0          17624886.0

        Gaming DL (Bytes)  Gaming UL (Bytes)  Other DL (Bytes)  \
0             466109357.0          5333340.0       670751043.0
```

```
1             821879090.0          8521398.0       472846860.0
2             805301713.0         16257481.0        24303797.0
3             583864716.0          6992868.0       685122214.0
4             542337372.0         15820659.0       463698729.0
...                   ...                ...               ...
110427        114093049.0          2834548.0       695881178.0
110428        328766801.0          7569327.0       371261255.0
110429        833634251.0         10607174.0       697260277.0
110430        338246033.0          1845068.0        17385489.0
110431        293519955.0         16295588.0       440290470.0

        Other UL (Bytes)
0             15950724.0
1              1337849.0
2             15907613.0
3              2966860.0
4             11939642.0
...                  ...
110427         3888729.0
110428        12947410.0
110429         9094407.0
110430        12797797.0
110431         6398758.0

[110432 rows x 18 columns]
```

```
data1=pd.DataFrame(user_behaviour)
```

```
data1
```

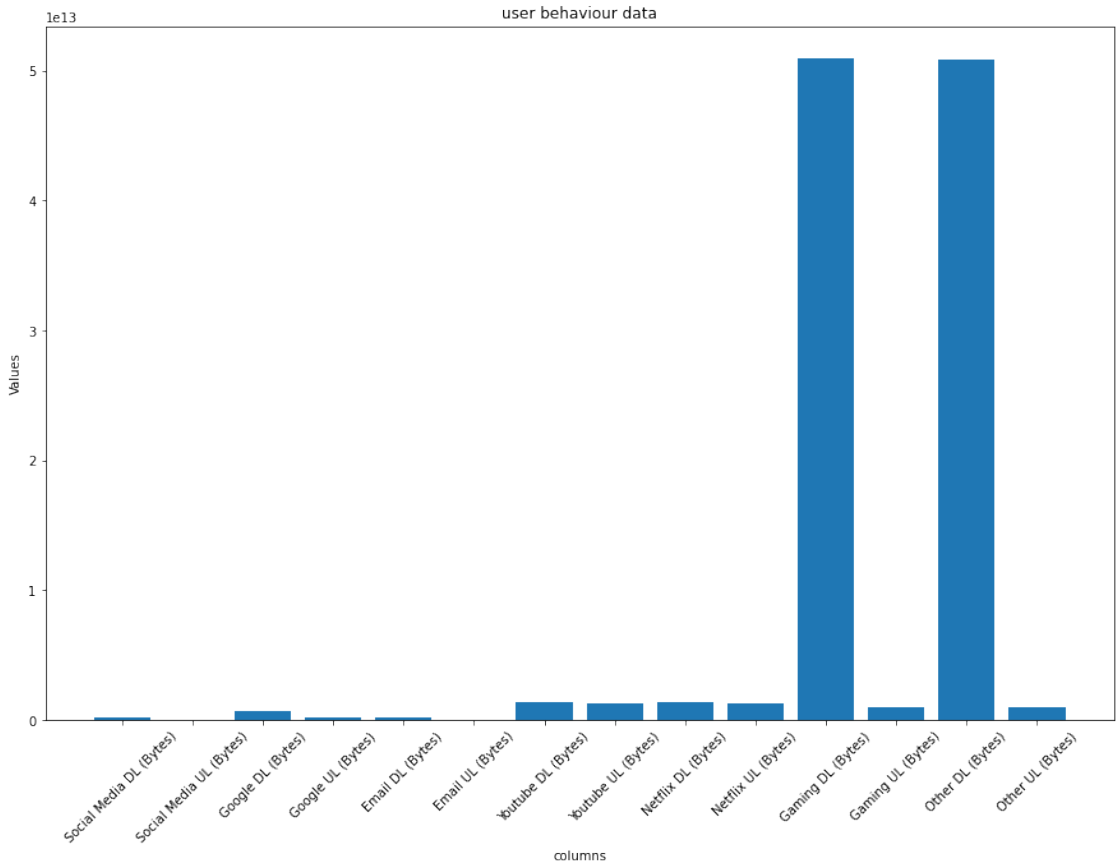
```
           Bearer Id  Dur. (ms)  Activity Duration DL (ms)  \
0       6.917538e+18    24534.0                   131798.0
1       6.917538e+18    27786.0                   401941.0
2       6.917538e+18    15635.0                    73347.0
3       6.917538e+18    24264.0                   117340.0
4       6.917538e+18    15548.0                    76969.0
...              ...        ...                        ...
110427  1.318654e+19    80024.0                  2512362.0
110428  1.318654e+19   145291.0                     2067.0
110429  1.318654e+19    86399.0                  3968131.0
110430  1.318654e+19    86399.0                  1689999.0
110431  1.318654e+19   103113.0                        0.0

        Activity Duration UL (ms)  Social Media DL (Bytes)  \
0                        101470.0                2404741.0
1                        399092.0                 944612.0
2                         81378.0                1817239.0
3                        347852.0                1867318.0
4                         80241.0                1502346.0
...                           ...                      ...
```

```
110427                  2437668.0               3240226.0
110428                    45217.0               3062671.0
110429                  3537154.0                720996.0
110430                  1513764.0               2492460.0
110431                    30367.0               1314234.0

        Social Media UL (Bytes)  Google DL (Bytes)  Google UL (Bytes)
\
0                        2410.0          5791591.0          2871336.0

1                        2827.0         10373157.0            56392.0

2                       19827.0           269988.0          3696393.0

3                       18928.0          1689296.0           195216.0

4                       37348.0          8917833.0           701876.0

...                         ...                ...                ...

110427                  38284.0          2036152.0          2271168.0

110428                  48953.0          9363661.0          4001970.0

110429                  42836.0          1541915.0          2100839.0

110430                  39905.0         11318188.0           466218.0

110431                  27938.0          6969652.0          3756009.0


        Email DL (Bytes)  Email UL (Bytes)  Youtube DL (Bytes)  \
0               782388.0          806920.0           6139644.0
1               128003.0           34038.0           5385159.0
2              3191192.0          896670.0          12347020.0
3               740633.0          590043.0          15231815.0
4               332813.0          537902.0          14602694.0
...                  ...               ...                ...
110427         2410615.0          387548.0          12404964.0
110428         2192057.0          866373.0          22147919.0
110429         2315638.0          839789.0          11879062.0
110430         2612190.0          618629.0          22163800.0
110431         3317462.0          408257.0          12099319.0

        Youtube UL (Bytes)  Netflix DL (Bytes)  Netflix UL (Bytes)  \
0                2071526.0          19494278.0          14668354.0
1                4295851.0          15755839.0           1300571.0
2               11089528.0           2859358.0           1738176.0
3                8401567.0          21563985.0           2817981.0
```

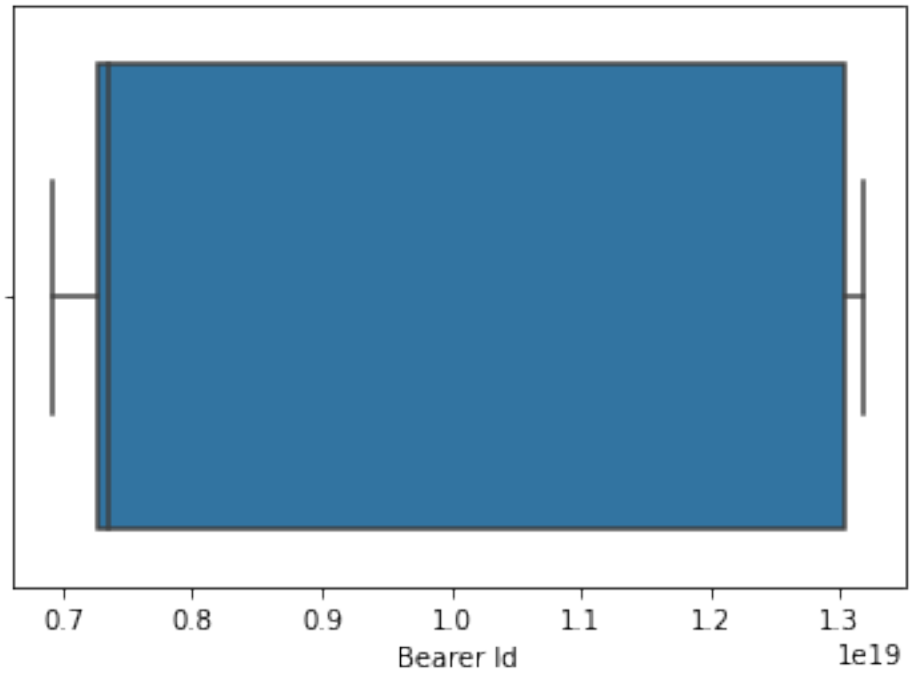
```
4                1237356.0           4506463.0          17006591.0
...                    ...                 ...                ...
110427           4343114.0          11108134.0          21649273.0
110428           2152449.0          21468525.0           8603105.0
110429           1290963.0          22596930.0          11943452.0
110430          16763435.0           9522397.0           8346624.0
110431           1636122.0           9992219.0          17624886.0

        Gaming DL (Bytes)  Gaming UL (Bytes)  Other DL (Bytes)  \
0             466109357.0          5333340.0       670751043.0
1             821879090.0          8521398.0       472846860.0
2             805301713.0         16257481.0        24303797.0
3             583864716.0          6992868.0       685122214.0
4             542337372.0         15820659.0       463698729.0
...                   ...                ...               ...
110427        114093049.0          2834548.0       695881178.0
110428        328766801.0          7569327.0       371261255.0
110429        833634251.0         10607174.0       697260277.0
110430        338246033.0          1845068.0        17385489.0
110431        293519955.0         16295588.0       440290470.0

        Other UL (Bytes)
0              15950724.0
1               1337849.0
2              15907613.0
3               2966860.0
4              11939642.0
...                   ...
110427          3888729.0
110428         12947410.0
110429          9094407.0
110430         12797797.0
110431          6398758.0

[110432 rows x 18 columns]
```

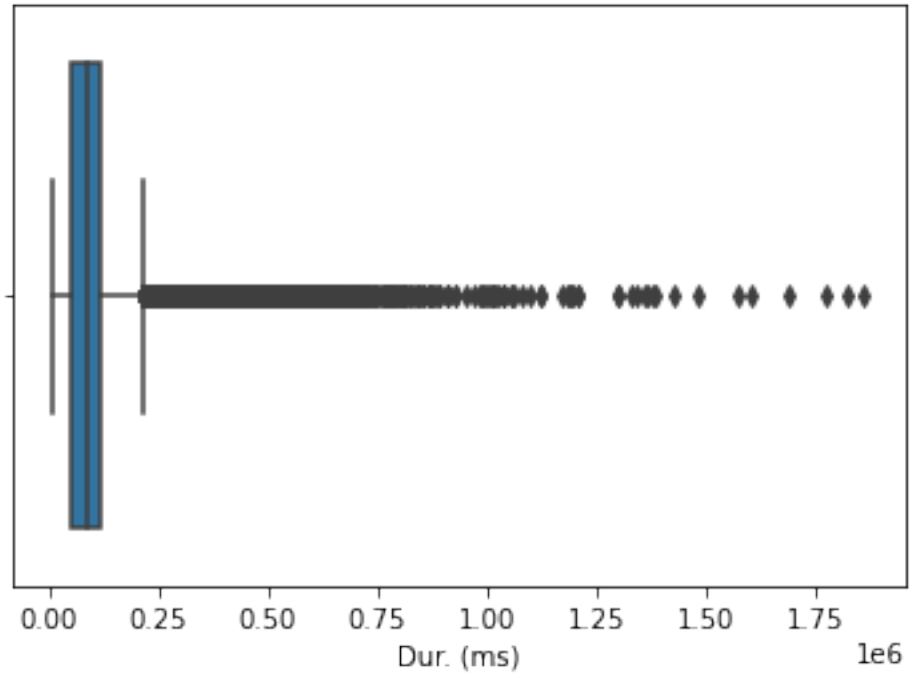
```
# Plot the bar chart using seaborn
plt.figure(figsize=(15,10))
Bar_columns = [ 'Social Media DL (Bytes)','Social Media UL
(Bytes)','Google DL (Bytes)','Google DL (Bytes)','Google UL (Bytes)',
               'Email DL (Bytes)','Email UL (Bytes)','Youtube DL
(Bytes)','Youtube UL (Bytes)','Netflix DL (Bytes)',
               'Netflix UL (Bytes)','Gaming DL (Bytes)','Gaming UL
(Bytes)','Other DL (Bytes)','Other UL (Bytes)']
values = user_behaviour[Bar_columns].sum()
plt.bar(Bar_columns, values)
plt.title('user behaviour data')
plt.xlabel('columns')
plt.ylabel('Values')
```

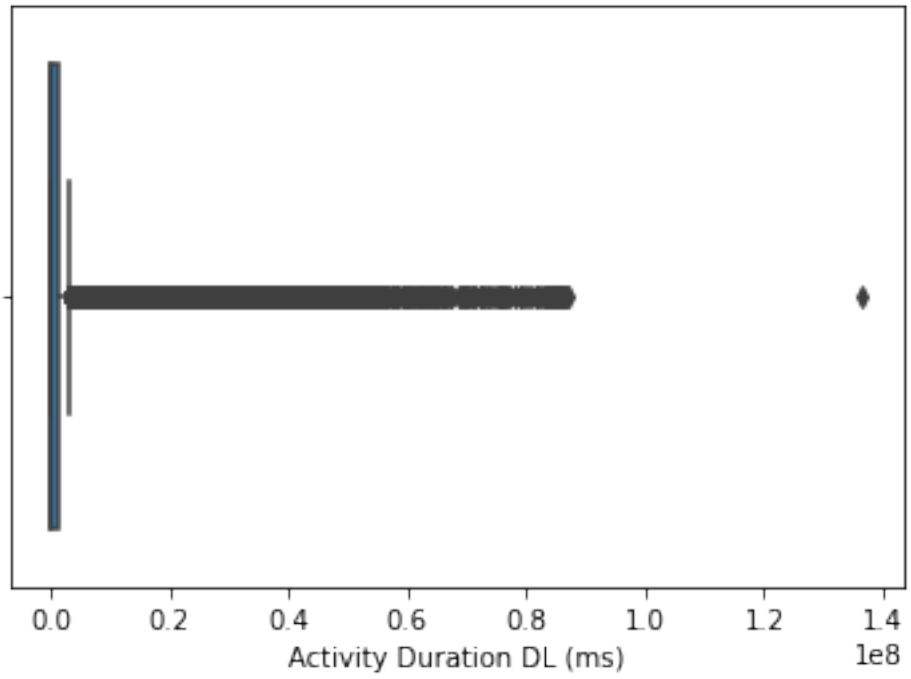
```
plt.xticks(rotation=45)
plt.show()
```

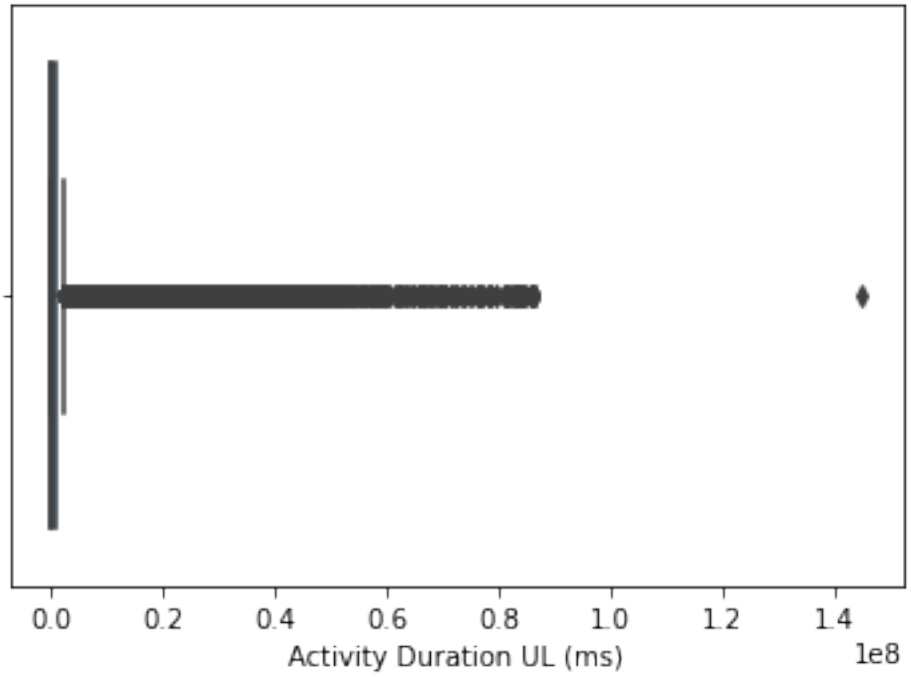


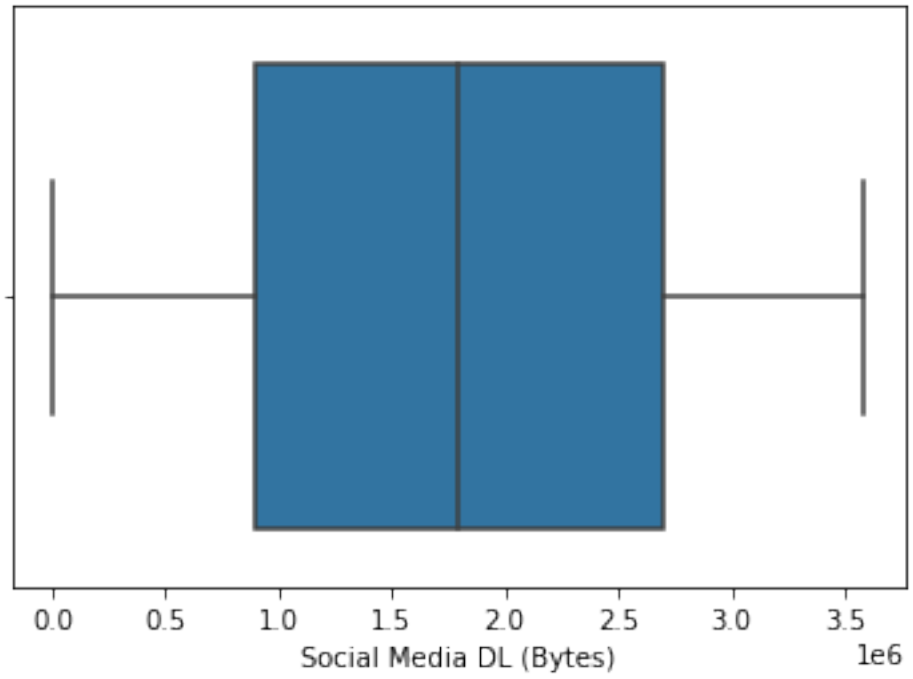

```
for i in data.columns:

    if data[i].dtype == 'object':
        mode_value = data[i].mode()[0]
        data[i].fillna(mode_value, inplace=True)
    else:
        mean_value = data[i].mean()
        data[i].fillna(mean_value, inplace=True)

for i in data1.columns:
    sns.boxplot(x=i,data=data)
    plt.show()
```

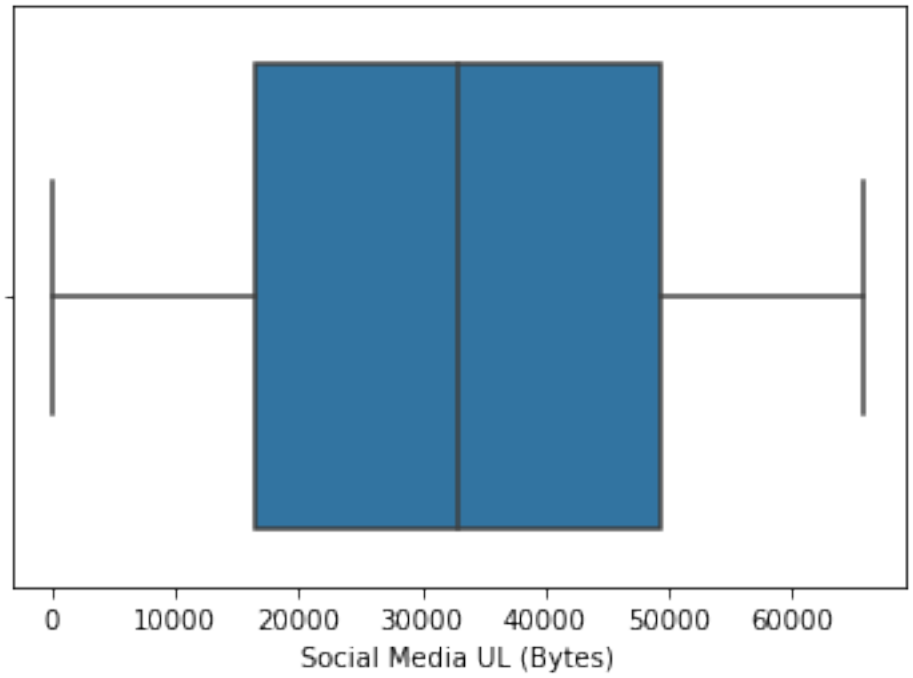

0.7
0.8
0.9
1.0
1.1
1.2
1.3
Bearer Id
1e19

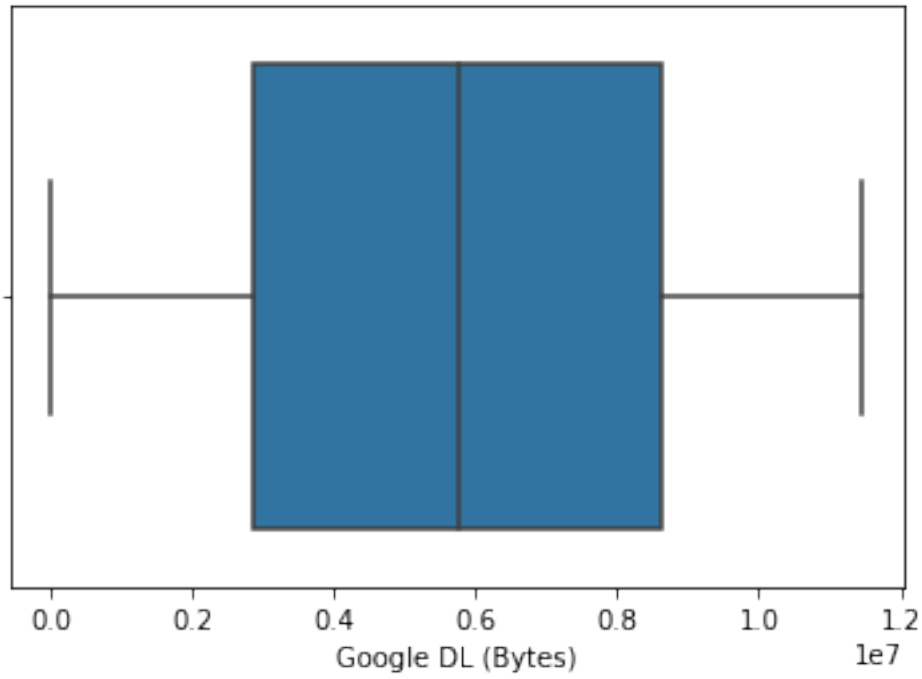

0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
Dur. (ms)
1e6

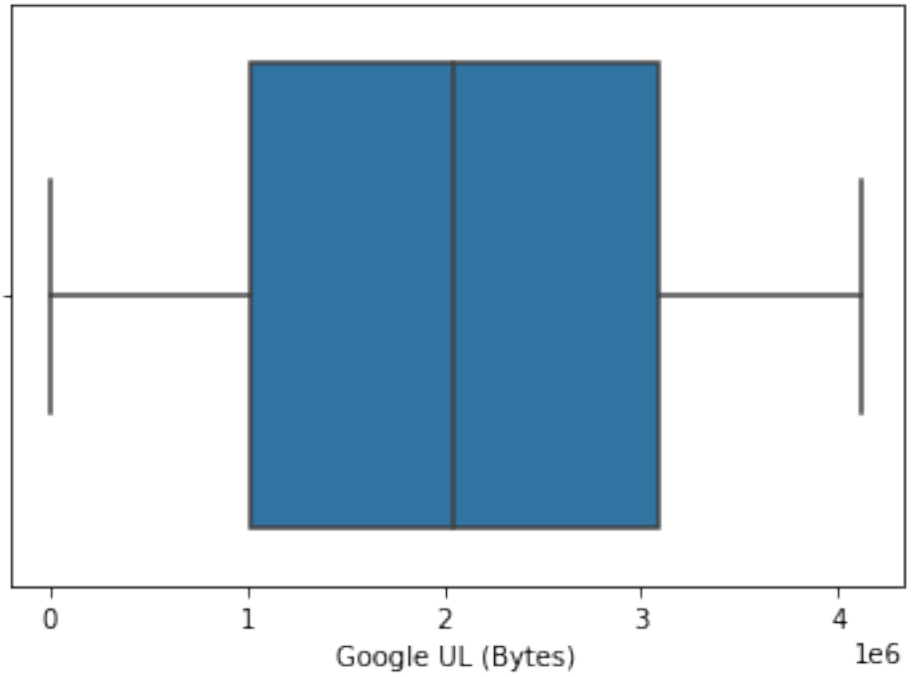

0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
Activity Duration DL (ms)
1e8

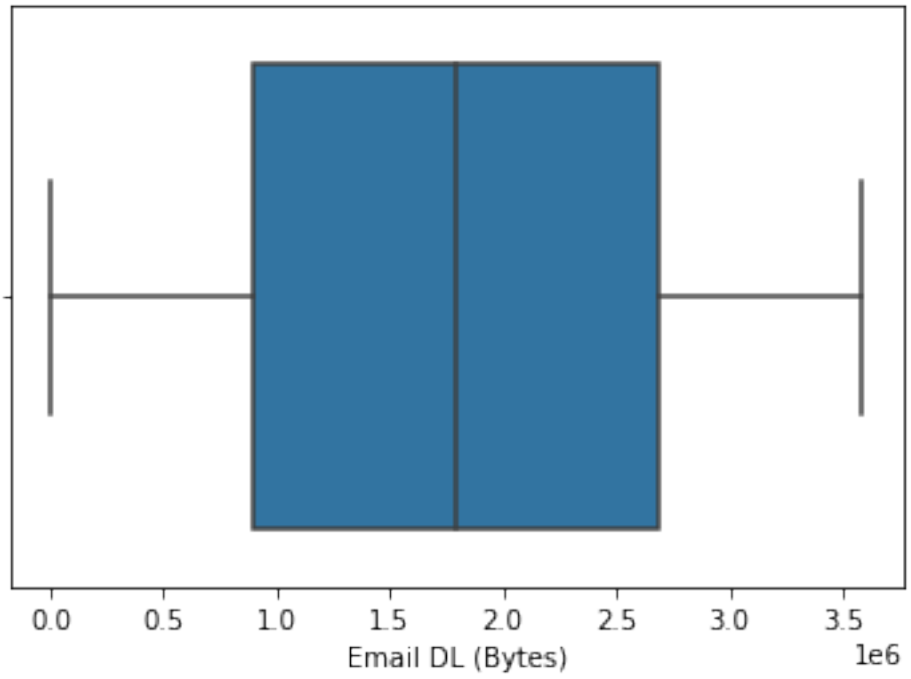

0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
Activity Duration UL (ms)
1e8

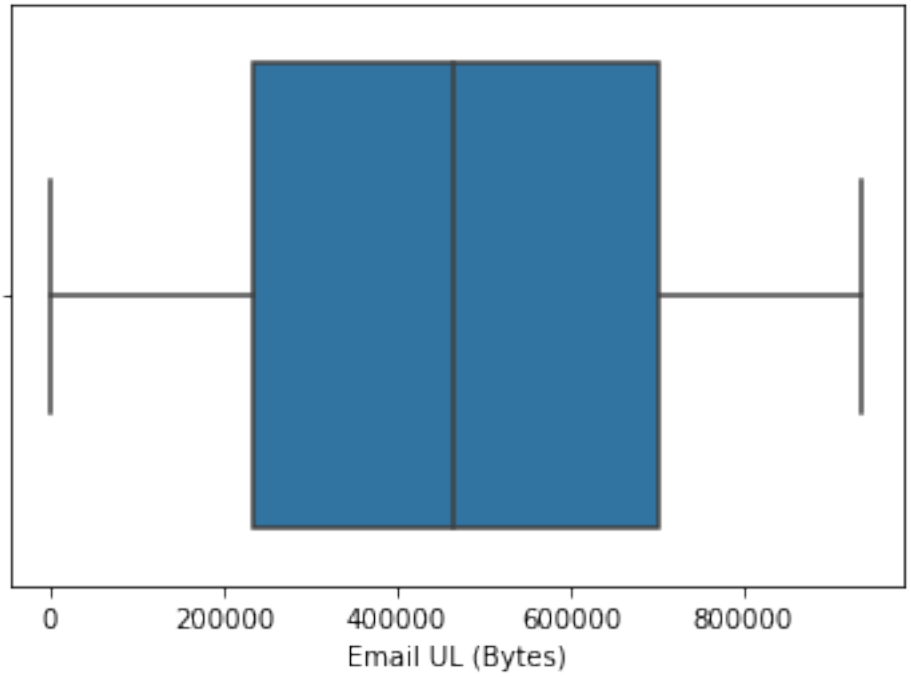

0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
1e6
Social Media DL (Bytes)

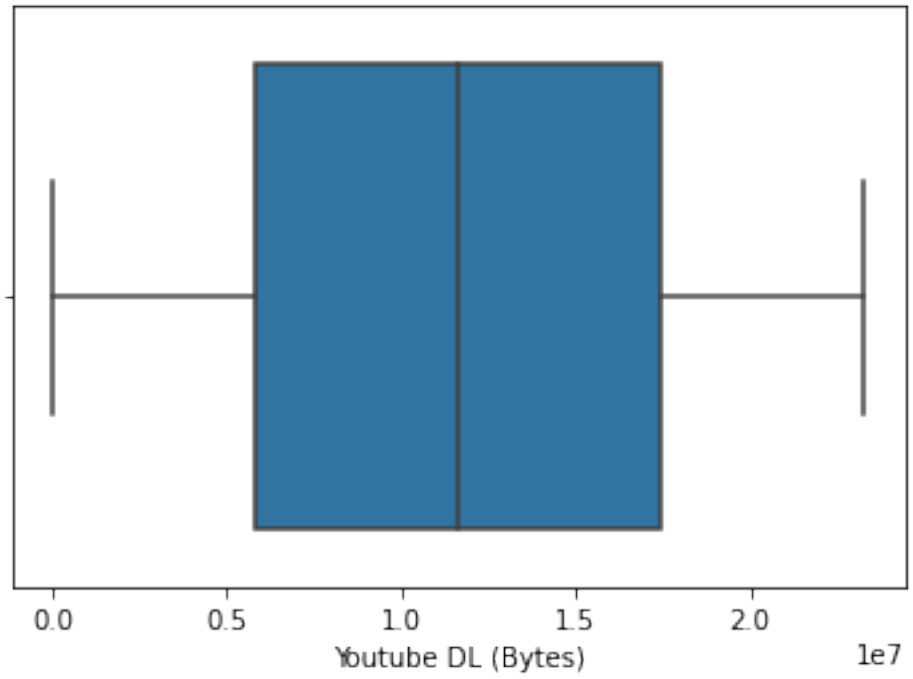

0
10000
20000
30000
40000
50000
60000
Social Media UL (Bytes)

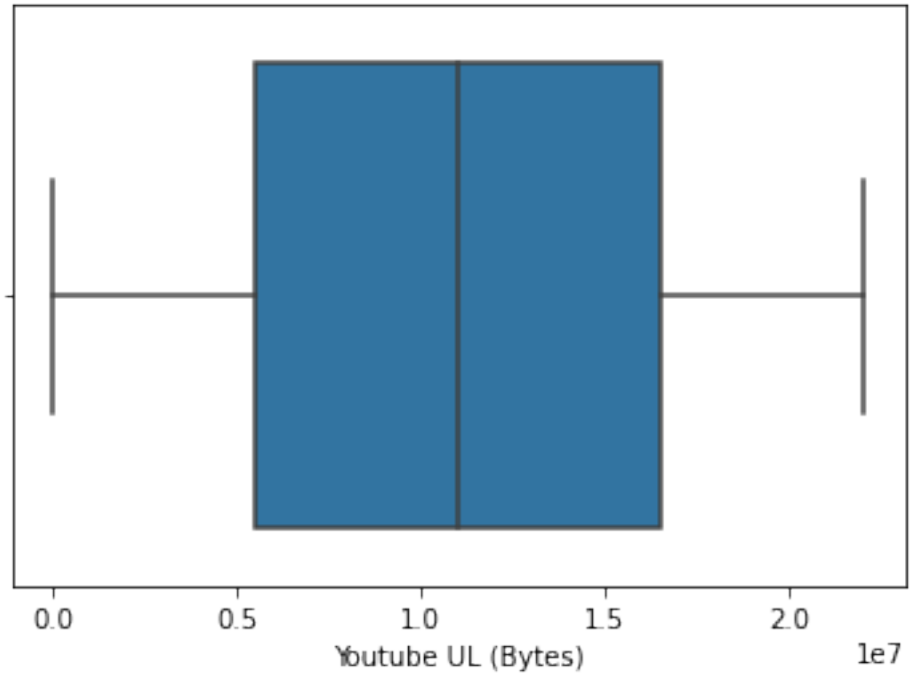

0.0
0.2
0.4
0.6
0.8
1.0
1.2
Google DL (Bytes)
1e7

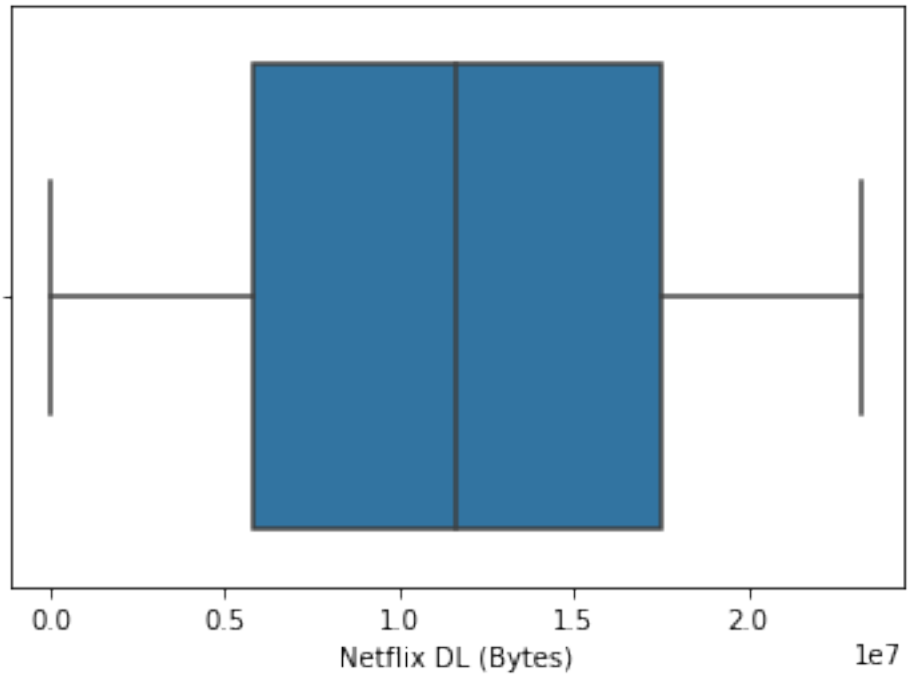

0
1
2
3
4
Google UL (Bytes)
1e6

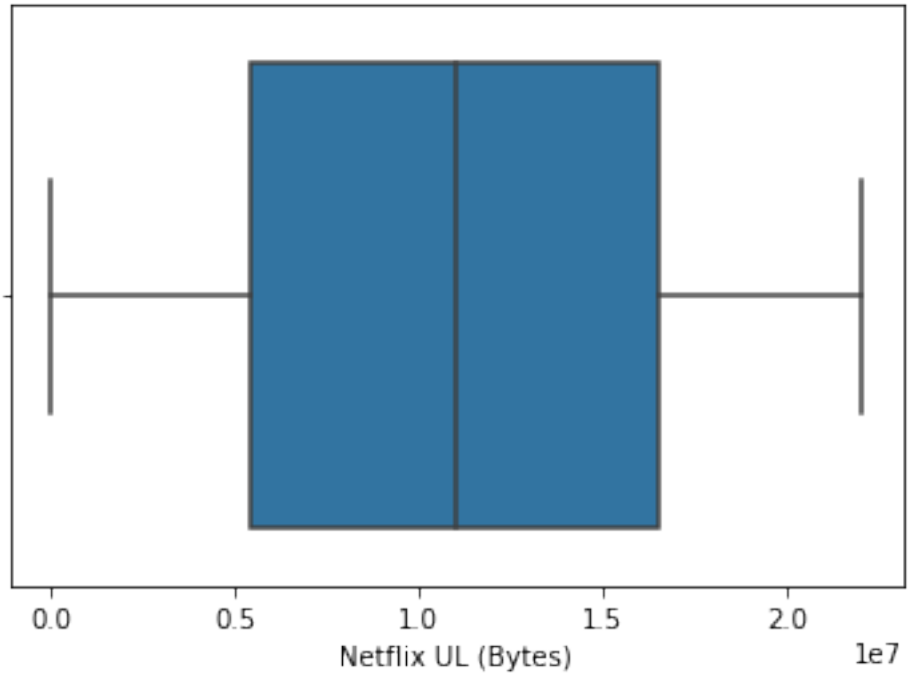

0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
Email DL (Bytes)
1e6

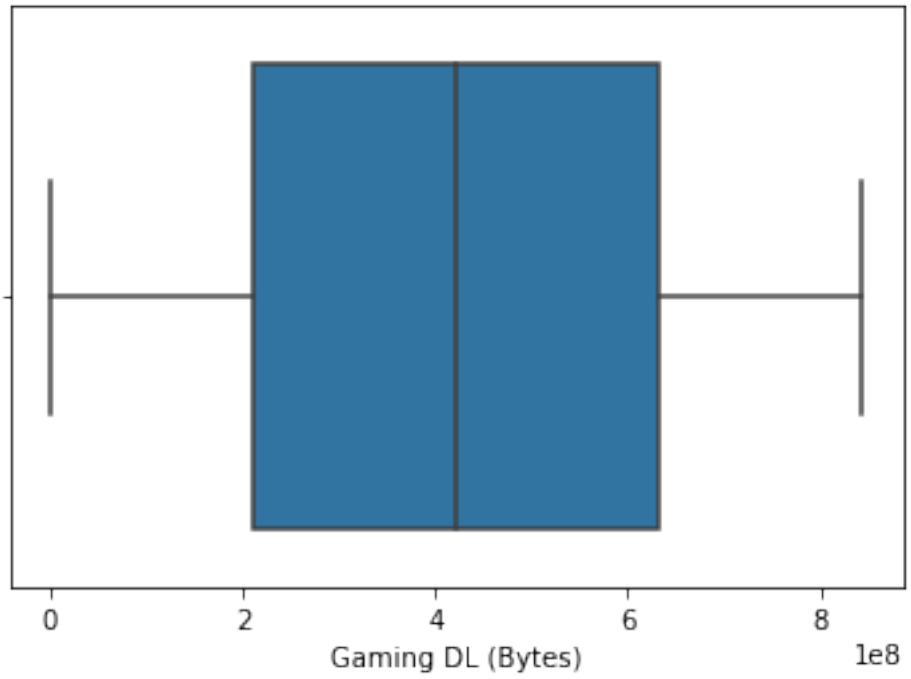

0
200000
400000
600000
800000
Email UL (Bytes)

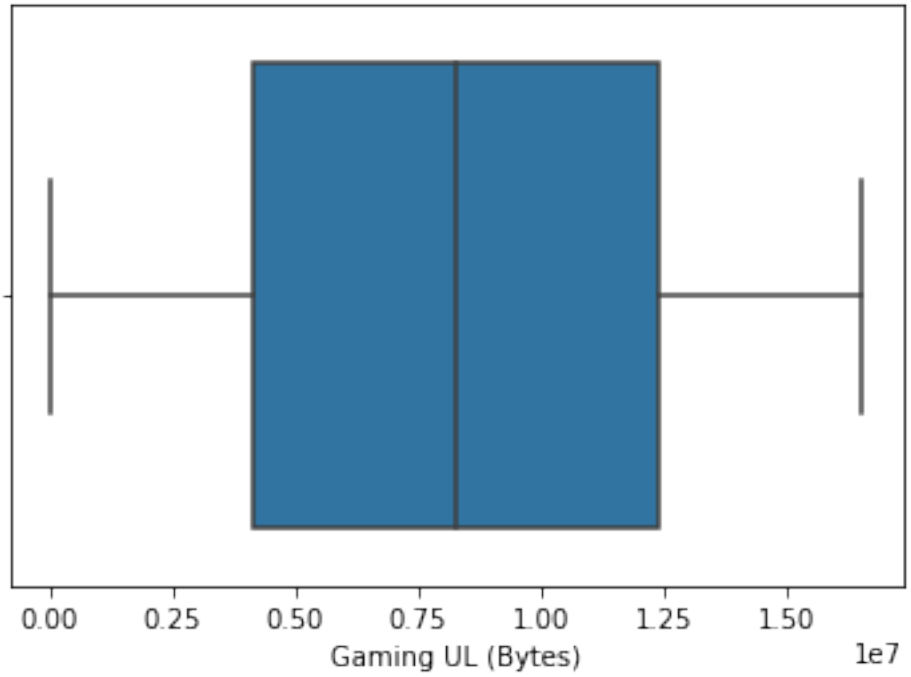

0.0
0.5
1.0
1.5
2.0
Youtube DL (Bytes)
1e7

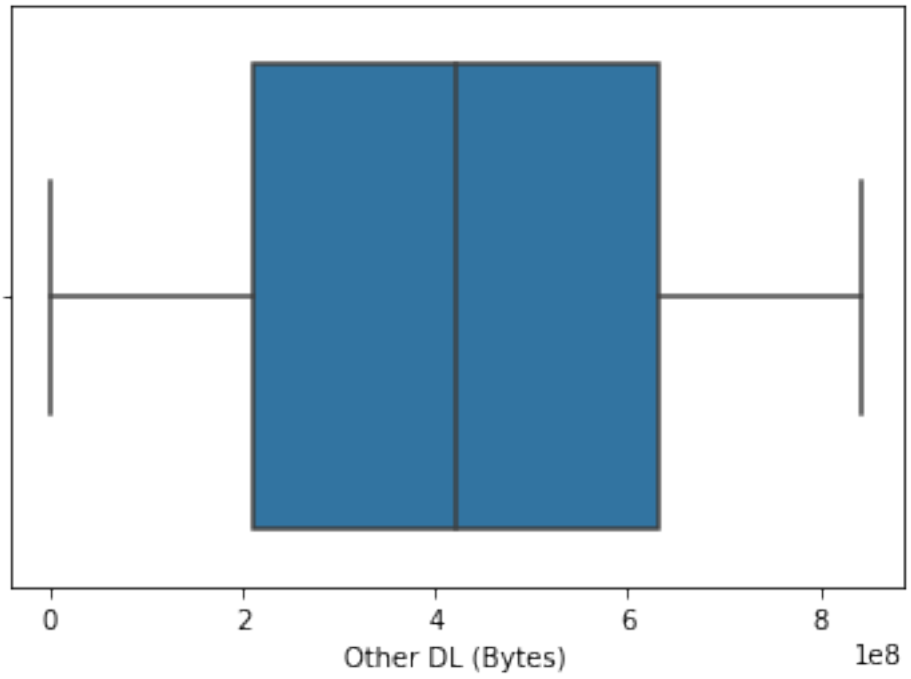

0.0
0.5
1.0
1.5
2.0
Youtube UL (Bytes)
1e7

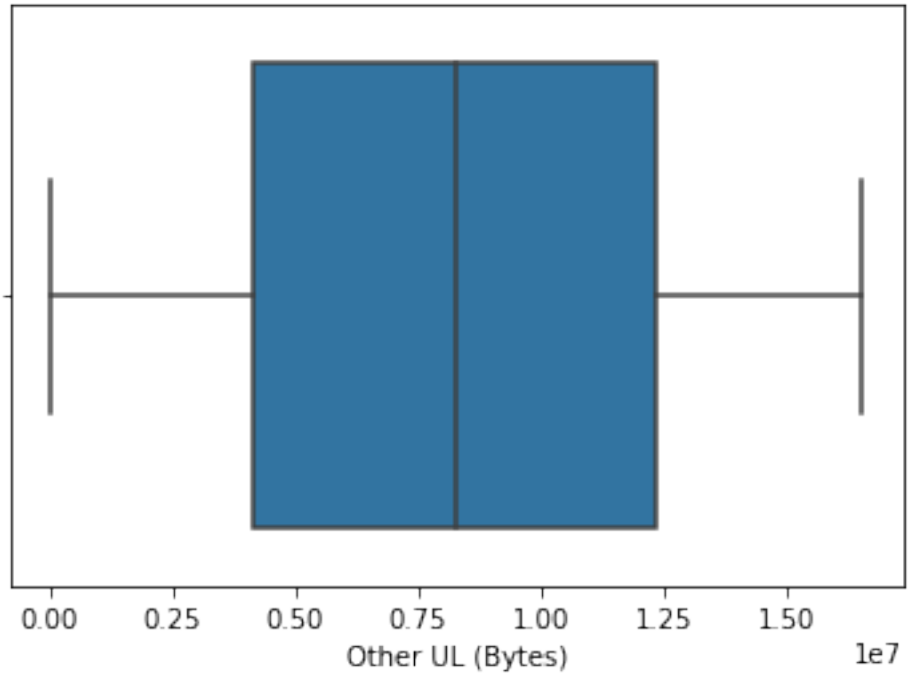

0.0
0.5
1.0
1.5
2.0
Netflix DL (Bytes)
1e7


0.0
0.5
1.0
1.5
2.0
Netflix UL (Bytes)
1e7


0
2
4
6
8
Gaming DL (Bytes)
1e8


0.00
0.25
0.50
0.75
1.00
1.25
1.50
Gaming UL (Bytes)
1e7






```
user_behaviour.skew()
```

```
Bearer Id                   0.181440
Dur. (ms)                   8.478570
Activity Duration DL (ms)   5.583385
Activity Duration UL (ms)   6.859796
Social Media DL (Bytes)     1.199202
Social Media UL (Bytes)     1.118963
```

```
Google DL (Bytes)            1.190602
Google UL (Bytes)            1.263205
Email DL (Bytes)             1.207259
Email UL (Bytes)             1.203918
Youtube DL (Bytes)           1.184562
Youtube UL (Bytes)           1.203035
Netflix DL (Bytes)           1.096394
Netflix UL (Bytes)           1.238333
Gaming DL (Bytes)            1.119912
Gaming UL (Bytes)            1.118836
Other DL (Bytes)             1.243104
Other UL (Bytes)             1.174505
dtype: float64
```

Task 1.2 - Conduct exploratory data analysis on those data & communicate useful insights. Ensure that you identify and treat all missing values and outliers in the dataset by replacing them with the mean of the corresponding column.

You're expected to report about the following using Python script and slide :

• Describe all relevant variables and associated data types (slide).

• Analyze the basic metrics (mean, median, etc) in the Dataset (explain) & their importance for the global objective.

• Conduct a Non-Graphical Univariate Analysis by computing dispersion parameters for each quantitative variable and provide useful interpretation.

• Conduct a Graphical Univariate Analysis by identifying the most suitable plotting options for each variable and interpreting your findings.

• Bivariate Analysis – explore the relationship between each application & the total DL+UL data using appropriate methods and interpret your findings.

• Variable transformations – segment the users into the top five decile classes based on the total duration for all sessions and compute the total data (DL+UL) per decile class.

• Correlation Analysis – compute a correlation matrix for the following variables and interpret your findings: Social Media data, Google data, Email data, Youtube data, Netflix data, Gaming data, and Other data

• Dimensionality Reduction – perform a principal component analysis to reduce the dimensions of your data and provide a useful interpretation of the results (Provide your interpretation in four (4) bullet points maximum).

```
data1
```

```
           Bearer Id  Dur. (ms)  Activity Duration DL (ms)  \
0       6.917538e+18    24534.0                   131798.0
1       6.917538e+18    27786.0                   401941.0
2       6.917538e+18    15635.0                    73347.0
3       6.917538e+18    24264.0                   117340.0
4       6.917538e+18    15548.0                    76969.0
...              ...        ...                        ...
110427  1.318654e+19    80024.0                  2512362.0
110428  1.318654e+19   145291.0                     2067.0
110429  1.318654e+19    86399.0                  3968131.0
110430  1.318654e+19    86399.0                  1689999.0
110431  1.318654e+19   103113.0                        0.0

        Activity Duration UL (ms)  Social Media DL (Bytes)  \
0                        101470.0                2404741.0
1                        399092.0                 944612.0
2                         81378.0                1817239.0
3                        347852.0                1867318.0
4                         80241.0                1502346.0
...                           ...                      ...
110427                  2437668.0                3240226.0
110428                    45217.0                3062671.0
110429                  3537154.0                 720996.0
110430                  1513764.0                2492460.0
110431                    30367.0                1314234.0

        Social Media UL (Bytes)  Google DL (Bytes)  Google UL (Bytes)
\
0                        2410.0          5791591.0          2871336.0

1                        2827.0         10373157.0            56392.0

2                       19827.0           269988.0          3696393.0

3                       18928.0          1689296.0           195216.0

4                       37348.0          8917833.0           701876.0

...                         ...                ...                ...

110427                  38284.0          2036152.0          2271168.0

110428                  48953.0          9363661.0          4001970.0

110429                  42836.0          1541915.0          2100839.0

110430                  39905.0         11318188.0           466218.0

110431                  27938.0          6969652.0          3756009.0
```

	Email DL (Bytes)	Email UL (Bytes)	Youtube DL (Bytes)	\
0	782388.0	806920.0	6139644.0	
1	128003.0	34038.0	5385159.0	
2	3191192.0	896670.0	12347020.0	
3	740633.0	590043.0	15231815.0	
4	332813.0	537902.0	14602694.0	
...	...	...	...	
110427	2410615.0	387548.0	12404964.0	
110428	2192057.0	866373.0	22147919.0	
110429	2315638.0	839789.0	11879062.0	
110430	2612190.0	618629.0	22163800.0	
110431	3317462.0	408257.0	12099319.0	

	Youtube UL (Bytes)	Netflix DL (Bytes)	Netflix UL (Bytes)	\
0	2071526.0	19494278.0	14668354.0	
1	4295851.0	15755839.0	1300571.0	
2	11089528.0	2859358.0	1738176.0	
3	8401567.0	21563985.0	2817981.0	
4	1237356.0	4506463.0	17006591.0	
...	...	...	...	
110427	4343114.0	11108134.0	21649273.0	
110428	2152449.0	21468525.0	8603105.0	
110429	1290963.0	22596930.0	11943452.0	
110430	16763435.0	9522397.0	8346624.0	
110431	1636122.0	9992219.0	17624886.0	

	Gaming DL (Bytes)	Gaming UL (Bytes)	Other DL (Bytes)	\
0	466109357.0	5333340.0	670751043.0	
1	821879090.0	8521398.0	472846860.0	
2	805301713.0	16257481.0	24303797.0	
3	583864716.0	6992868.0	685122214.0	
4	542337372.0	15820659.0	463698729.0	
...	...	...	...	
110427	114093049.0	2834548.0	695881178.0	
110428	328766801.0	7569327.0	371261255.0	
110429	833634251.0	10607174.0	697260277.0	
110430	338246033.0	1845068.0	17385489.0	
110431	293519955.0	16295588.0	440290470.0	

	Other UL (Bytes)
0	15950724.0
1	1337849.0
2	15907613.0
3	2966860.0
4	11939642.0
...	...
110427	3888729.0
110428	12947410.0

```
110429         9094407.0
110430        12797797.0
110431         6398758.0

[110432 rows x 18 columns]
```

```
data1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 110432 entries, 0 to 110431
Data columns (total 18 columns):
 #   Column                      Non-Null Count   Dtype  
---  ------                      --------------   -----  
 0   Bearer Id                   110432 non-null  float64
 1   Dur. (ms)                   110432 non-null  float64
 2   Activity Duration DL (ms)   110432 non-null  float64
 3   Activity Duration UL (ms)   110432 non-null  float64
 4   Social Media DL (Bytes)     110432 non-null  float64
 5   Social Media UL (Bytes)     110432 non-null  float64
 6   Google DL (Bytes)           110432 non-null  float64
 7   Google UL (Bytes)           110432 non-null  float64
 8   Email DL (Bytes)            110432 non-null  float64
 9   Email UL (Bytes)            110432 non-null  float64
 10  Youtube DL (Bytes)          110432 non-null  float64
 11  Youtube UL (Bytes)          110432 non-null  float64
 12  Netflix DL (Bytes)          110432 non-null  float64
 13  Netflix UL (Bytes)          110432 non-null  float64
 14  Gaming DL (Bytes)           110432 non-null  float64
 15  Gaming UL (Bytes)           110432 non-null  float64
 16  Other DL (Bytes)            110432 non-null  float64
 17  Other UL (Bytes)            110432 non-null  float64
dtypes: float64(18)
memory usage: 15.2 MB
```

```
data1.head()
```

```
      Bearer Id  Dur. (ms)  Activity Duration DL (ms)  \
0  6.917538e+18    24534.0                   131798.0   
1  6.917538e+18    27786.0                   401941.0   
2  6.917538e+18    15635.0                    73347.0   
3  6.917538e+18    24264.0                   117340.0   
4  6.917538e+18    15548.0                    76969.0   

   Activity Duration UL (ms)  Social Media DL (Bytes)  \
0                   101470.0                2404741.0   
1                   399092.0                 944612.0   
2                    81378.0                1817239.0   
3                   347852.0                1867318.0   
4                    80241.0                1502346.0   

   Social Media UL (Bytes)  Google DL (Bytes)  Google UL (Bytes)  \
```

```
0                   2410.0          5791591.0          2871336.0
1                   2827.0         10373157.0            56392.0
2                  19827.0           269988.0          3696393.0
3                  18928.0          1689296.0           195216.0
4                  37348.0          8917833.0           701876.0

   Email DL (Bytes)  Email UL (Bytes)  Youtube DL (Bytes)  Youtube UL
(Bytes)  \
0          782388.0          806920.0           6139644.0
2071526.0
1          128003.0           34038.0           5385159.0
4295851.0
2         3191192.0          896670.0          12347020.0
11089528.0
3          740633.0          590043.0          15231815.0
8401567.0
4          332813.0          537902.0          14602694.0
1237356.0

   Netflix DL (Bytes)  Netflix UL (Bytes)  Gaming DL (Bytes)  \
0          19494278.0          14668354.0        466109357.0
1          15755839.0           1300571.0        821879090.0
2           2859358.0           1738176.0        805301713.0
3          21563985.0           2817981.0        583864716.0
4           4506463.0          17006591.0        542337372.0

   Gaming UL (Bytes)  Other DL (Bytes)  Other UL (Bytes)
0          5333340.0       670751043.0        15950724.0
1          8521398.0       472846860.0         1337849.0
2         16257481.0        24303797.0        15907613.0
3          6992868.0       685122214.0         2966860.0
4         15820659.0       463698729.0        11939642.0

data1.describe()

          Bearer Id     Dur. (ms)  Activity Duration DL (ms)  \
count  1.104320e+05  1.104320e+05               1.104320e+05
mean   9.907303e+18  1.051041e+05               2.464303e+06
std    2.883127e+18  9.217577e+04               7.058841e+06
min    6.917538e+18  7.142000e+03               0.000000e+00
25%    7.277826e+18  5.291850e+04               2.131075e+04
50%    7.349883e+18  8.639900e+04               6.175800e+04
75%    1.304243e+19  1.357822e+05               1.219347e+06
max    1.318654e+19  4.838428e+06               2.035294e+08

       Activity Duration UL (ms)  Social Media DL (Bytes)  \
count               1.104320e+05             1.104320e+06
mean                1.893064e+06             1.962976e+06
std                 5.711516e+06             1.264421e+06
min                 0.000000e+00             1.200000e+01
```

25%	2.755000e+04	9.625455e+05
50%	6.907150e+04	1.910490e+06
75%	1.040554e+06	2.836990e+06
max	1.666611e+08	2.618898e+07

	Social Media UL (Bytes)	Google DL (Bytes)	Google UL (Bytes)
count	110432.000000	1.104320e+05	1.104320e+05
mean	35980.638031	6.288150e+06	2.247206e+06
std	23141.331972	4.048132e+06	1.456667e+06
min	0.000000	2.970000e+02	3.000000e+00
25%	17606.750000	3.090234e+06	1.097197e+06
50%	35040.000000	6.122278e+06	2.181378e+06
75%	51920.000000	9.070152e+06	3.253726e+06
max	478166.000000	9.556064e+07	3.861748e+07

	Email DL (Bytes)	Email UL (Bytes)	Youtube DL (Bytes)
count	1.104320e+05	1.104320e+05	1.104320e+05
mean	1.958303e+06	5.107500e+05	1.272504e+07
std	1.263999e+06	3.300887e+05	8.183277e+06
min	1.400000e+01	2.000000e+00	5.300000e+01
25%	9.572890e+05	2.499518e+05	6.278096e+06
50%	1.904660e+06	4.949325e+05	1.235056e+07
75%	2.830856e+06	7.386615e+05	1.839123e+07
max	3.379861e+07	6.744930e+06	1.796289e+08

	Youtube UL (Bytes)	Netflix DL (Bytes)	Netflix UL (Bytes)
count	1.104320e+05	1.104320e+05	1.104320e+05
mean	1.203117e+07	1.271781e+07	1.202267e+07
std	7.756424e+06	8.171737e+06	7.759512e+06
min	1.050000e+02	4.200000e+01	3.500000e+01
25%	5.896152e+06	6.219047e+06	5.876785e+06
50%	1.169876e+07	1.235935e+07	1.170558e+07
75%	1.737368e+07	1.840726e+07	1.738361e+07
max	1.879087e+08	1.452646e+08	2.277824e+08

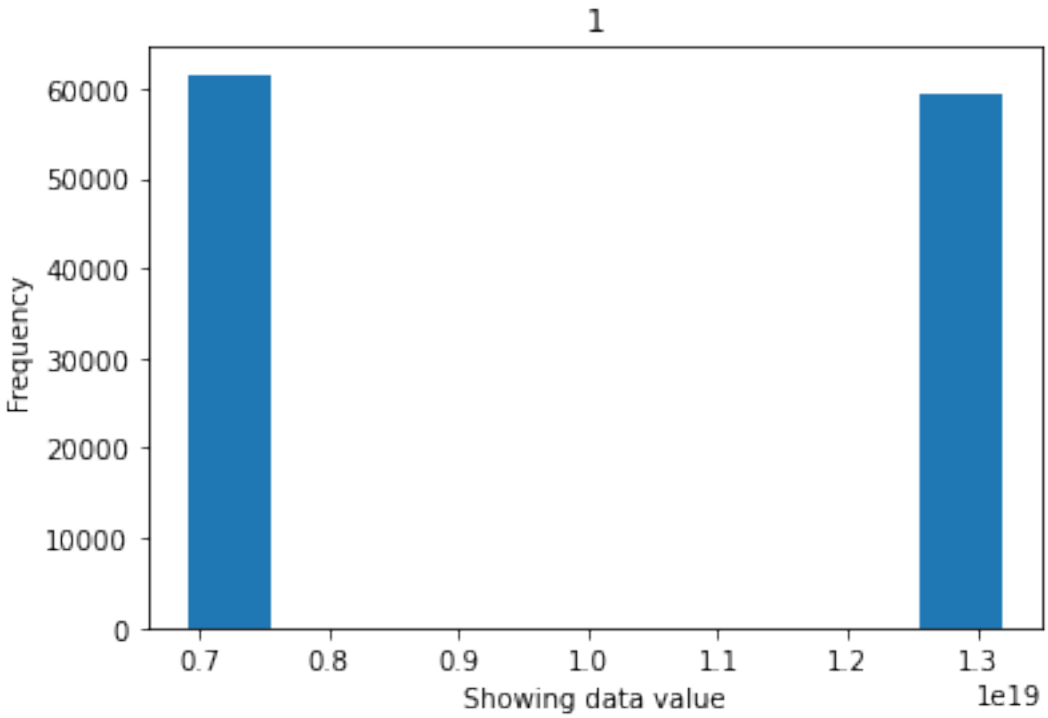
	Gaming DL (Bytes)	Gaming UL (Bytes)	Other DL (Bytes)
count	1.104320e+05	1.104320e+05	1.104320e+05
mean	4.608041e+08	9.066019e+06	4.607876e+08
std	2.972070e+08	5.831056e+06	2.966063e+08
min	1.251100e+04	5.900000e+01	4.626900e+04

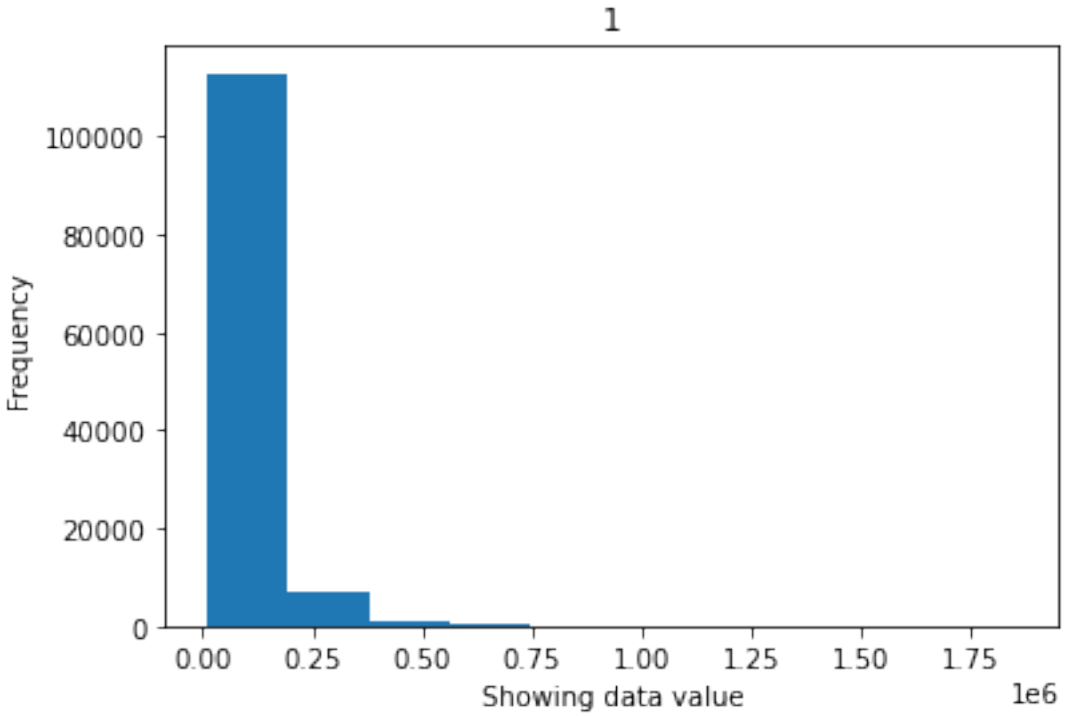
```
25%         2.254630e+08       4.442423e+06      2.262967e+08
50%         4.489663e+08       8.820961e+06      4.482374e+08
75%         6.660761e+08       1.305845e+07      6.657577e+08
max         5.808028e+09       1.118985e+08      8.051625e+09

       Other UL (Bytes)
count      1.104320e+05
mean       9.024526e+06
std        5.824731e+06
min        1.480000e+02
25%        4.445894e+06
50%        8.754301e+06
75%        1.302956e+07
max        1.287153e+08
```

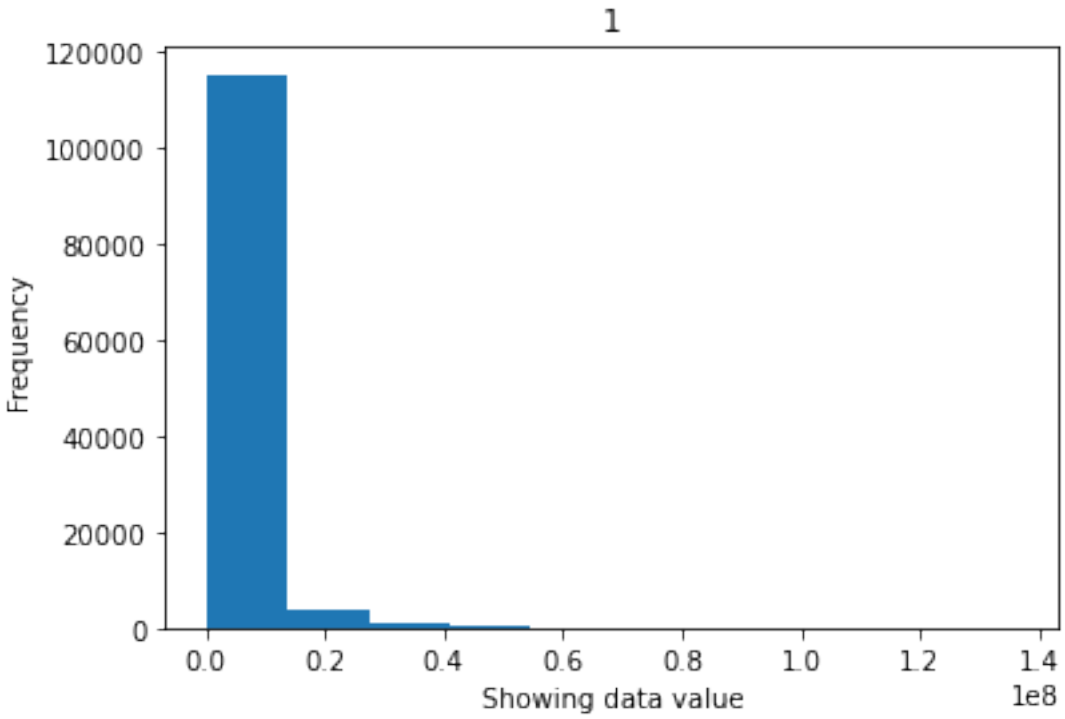
```
data1.columns
```

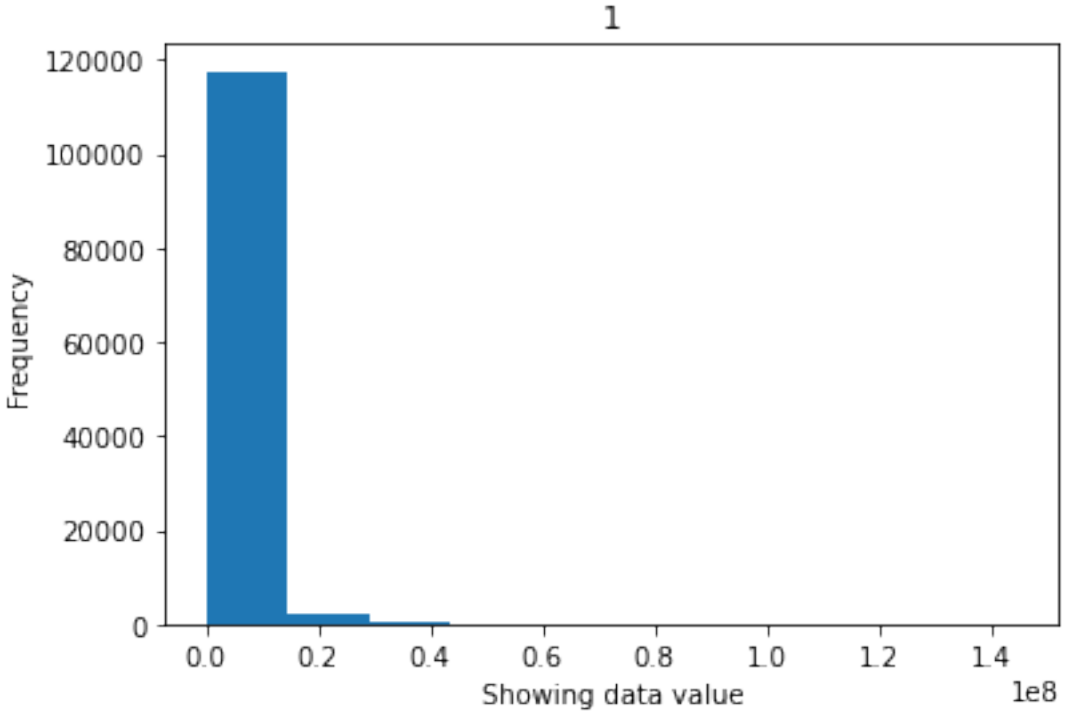
```
Index(['Bearer Id', 'Dur. (ms)', 'Activity Duration DL (ms)',
       'Activity Duration UL (ms)', 'Social Media DL (Bytes)',
       'Social Media UL (Bytes)', 'Google DL (Bytes)', 'Google UL
(Bytes)',
       'Email DL (Bytes)', 'Email UL (Bytes)', 'Youtube DL (Bytes)',
       'Youtube UL (Bytes)', 'Netflix DL (Bytes)', 'Netflix UL
(Bytes)',
       'Gaming DL (Bytes)', 'Gaming UL (Bytes)', 'Other DL (Bytes)',
       'Other UL (Bytes)'],
      dtype='object')
```

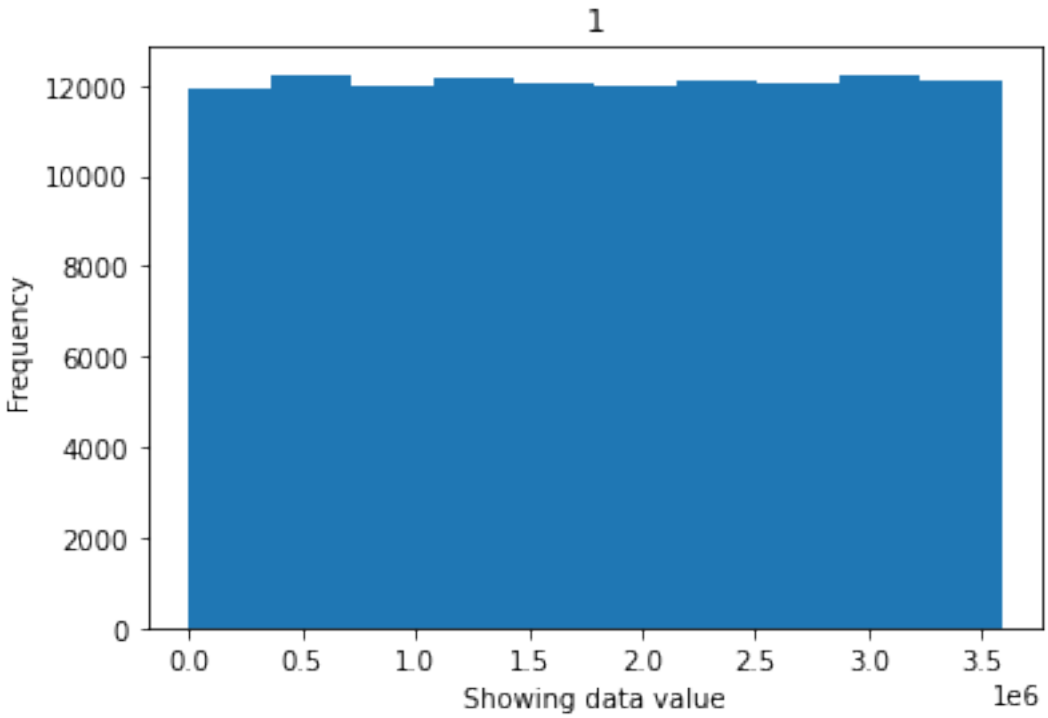
```
for i in data1.columns:
    plt.hist(data[i],bins=10)
    plt.title(1)
    plt.xlabel("Showing data value")
    plt.ylabel("Frequency")
    plt.savefig(f"{i}_histogram.png")
    plt.show()
```

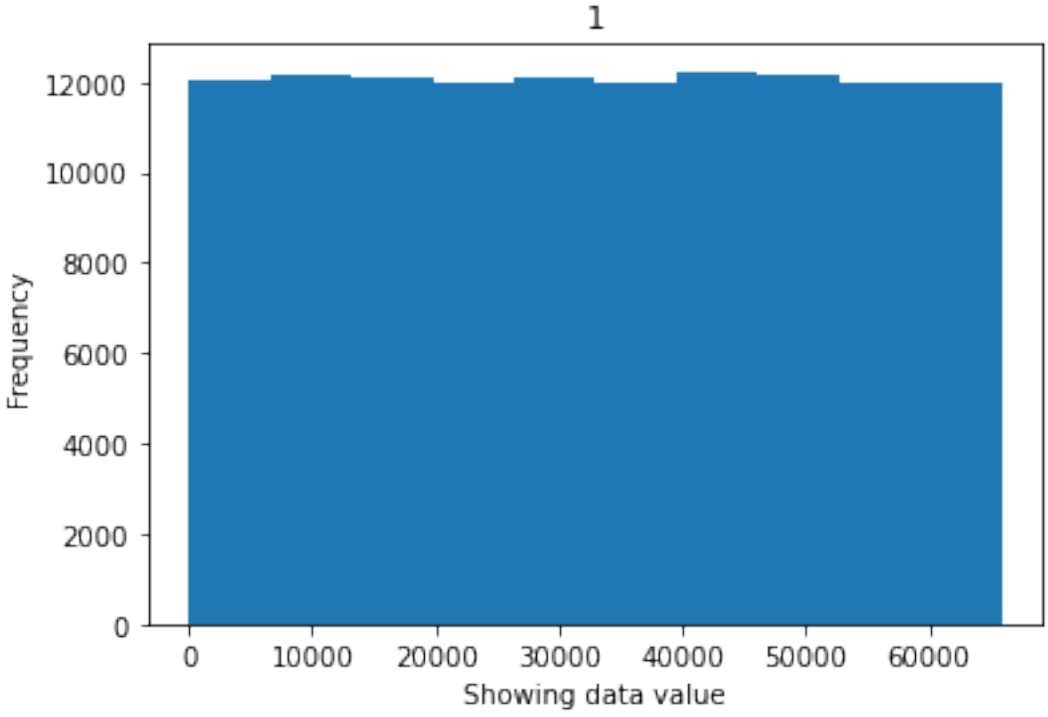

1
60000
50000
40000
30000
20000
10000
0
Frequency
0.7
0.8
0.9
1.0
1.1
1.2
1.3
Showing data value
1e19

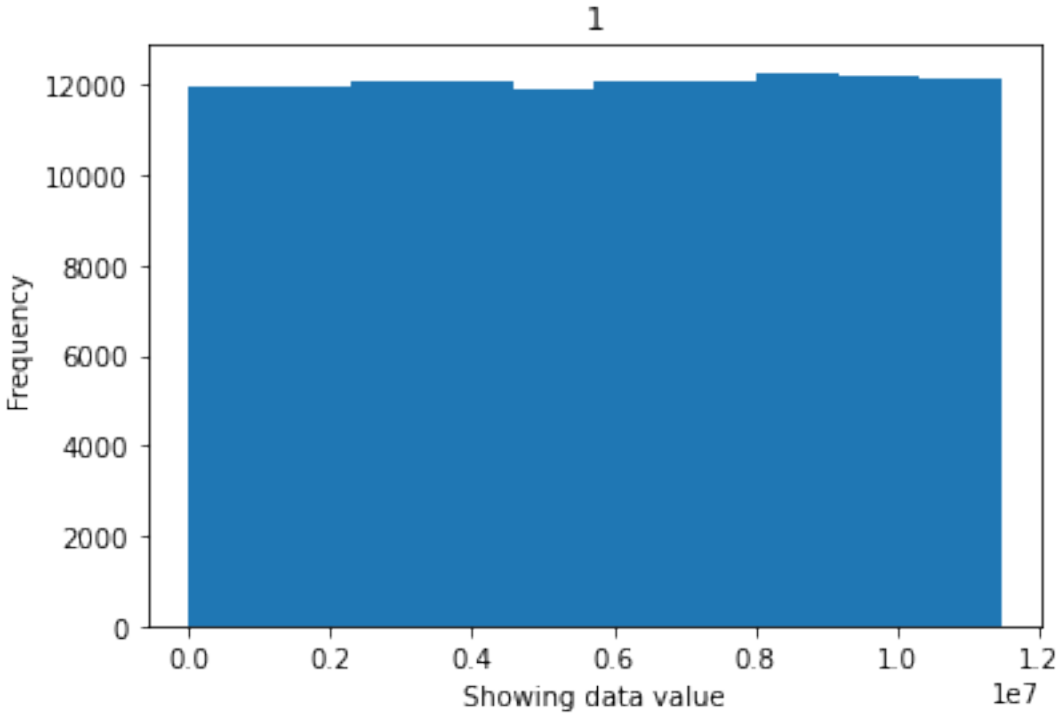

1
100000
80000
60000
40000
20000
0
Frequency
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
Showing data value
1e6

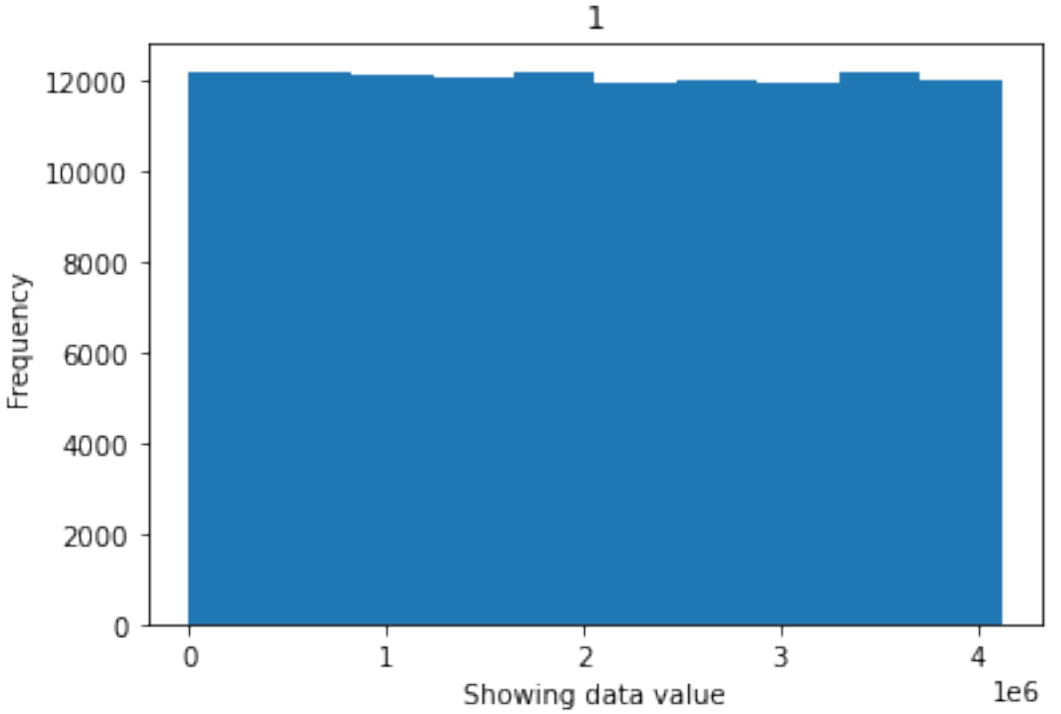

1
120000
100000
80000
60000
40000
20000
0
Frequency
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
Showing data value
1e8

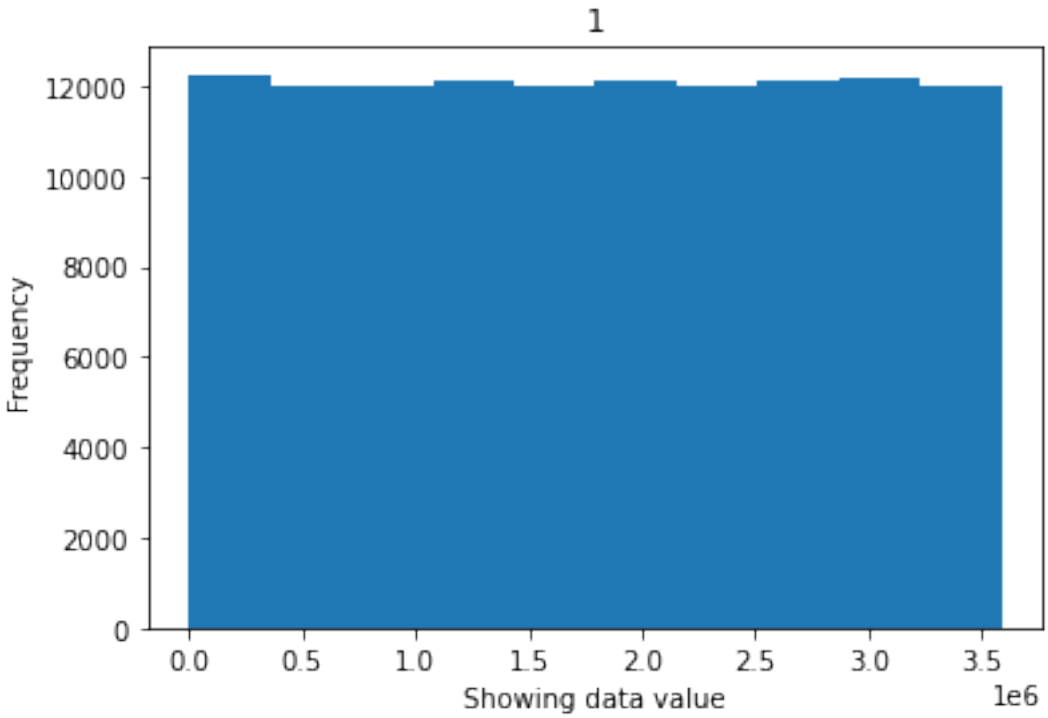

1
120000
100000
80000
60000
40000
20000
0
Frequency
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
Showing data value
1e8

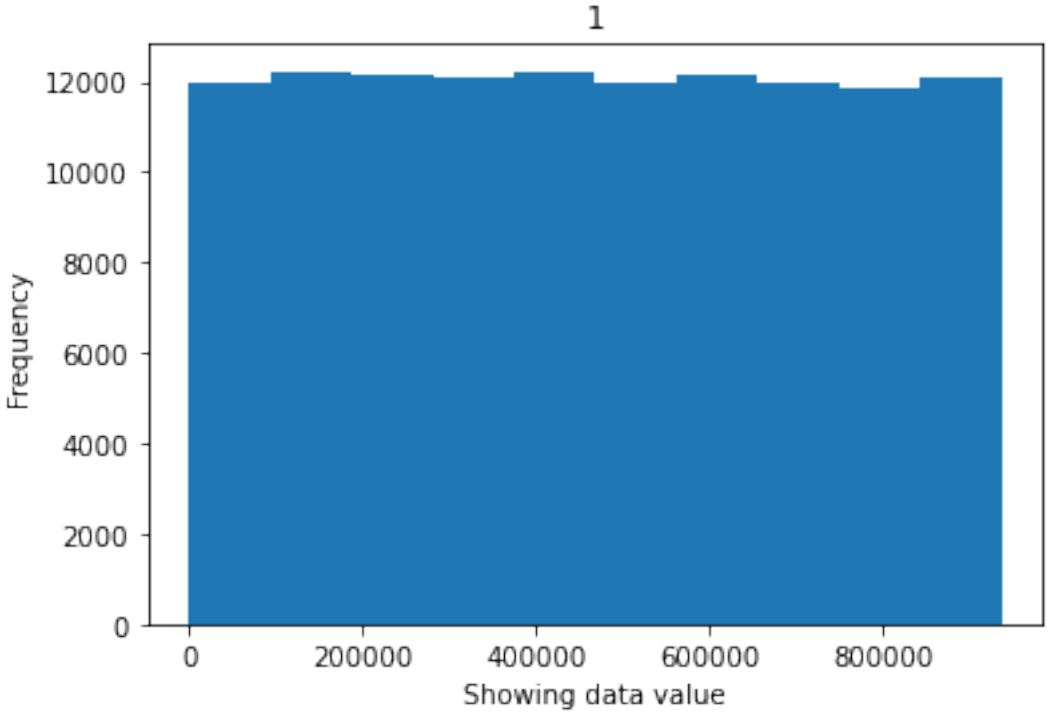

1
12000
10000
8000
6000
4000
2000
0
Frequency
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
1e6
Showing data value

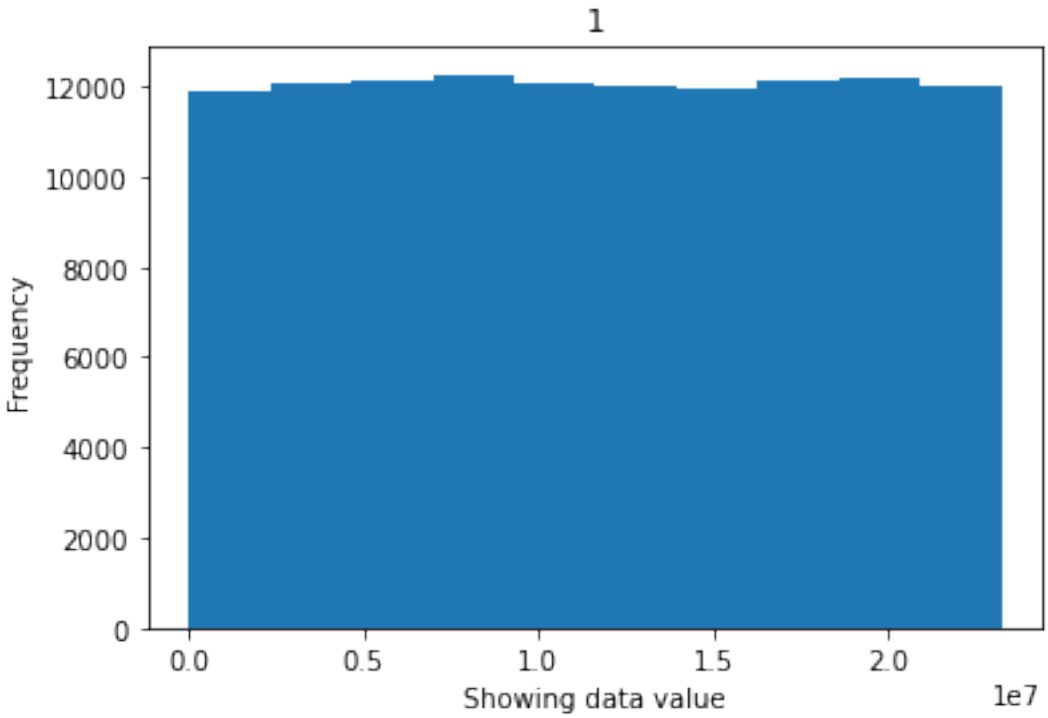

1
12000
10000
8000
6000
4000
2000
0
Frequency
0
10000
20000
30000
40000
50000
60000
Showing data value

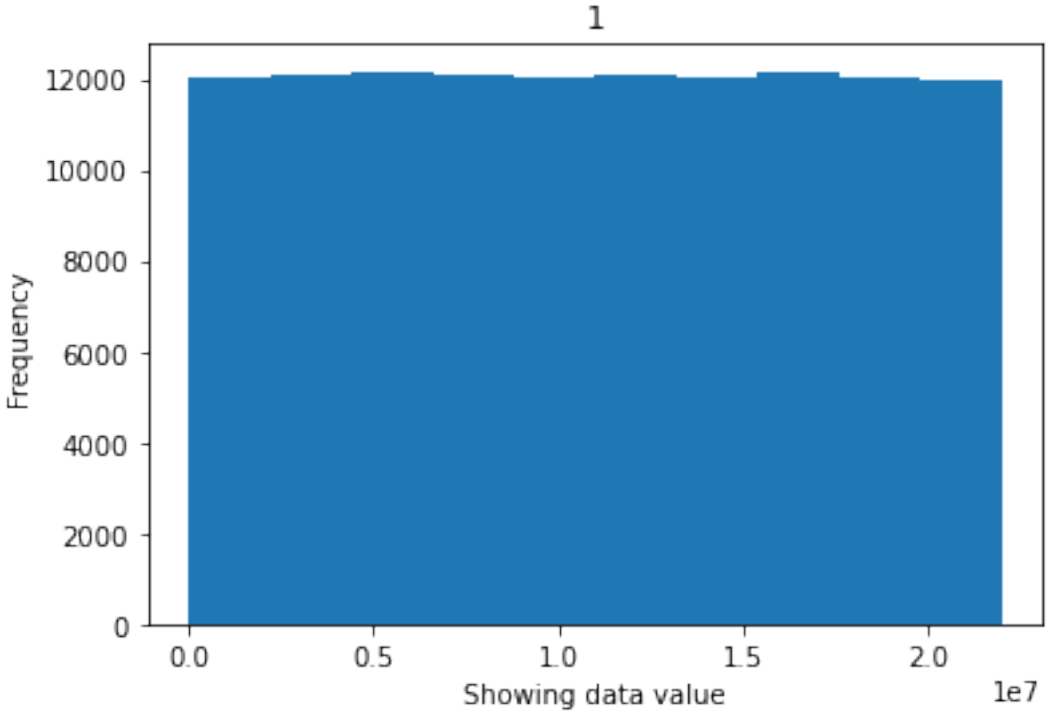

1
12000
10000
8000
6000
4000
2000
0
Frequency
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Showing data value
1e7

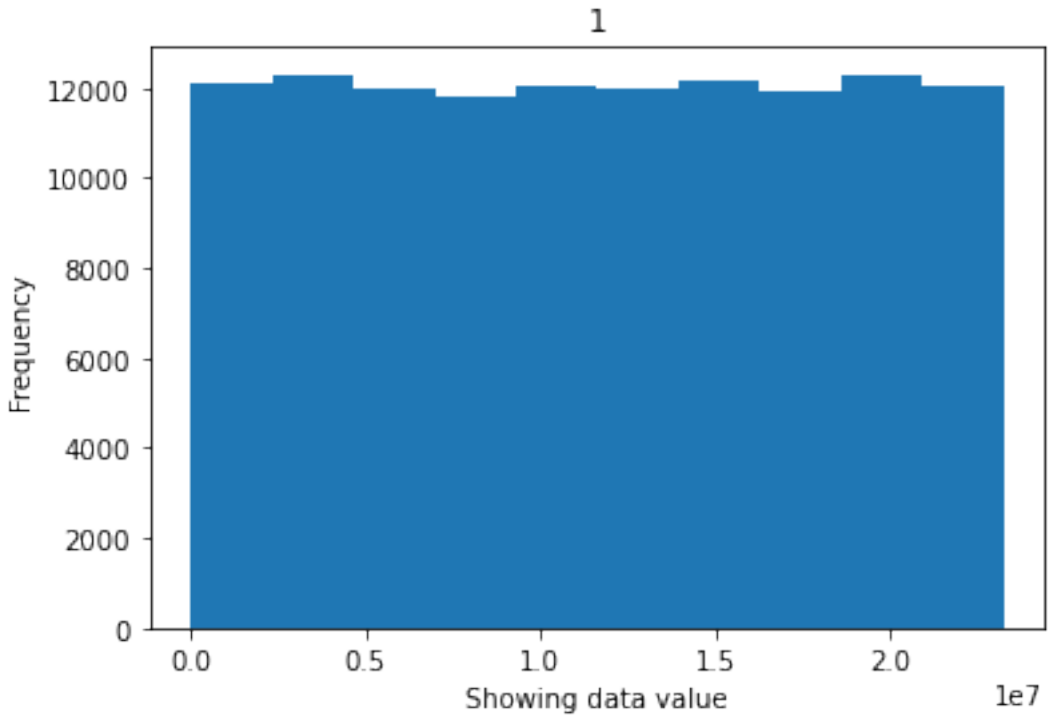

1
12000
10000
8000
6000
4000
2000
0
Frequency
0
1
2
3
4
Showing data value
1e6

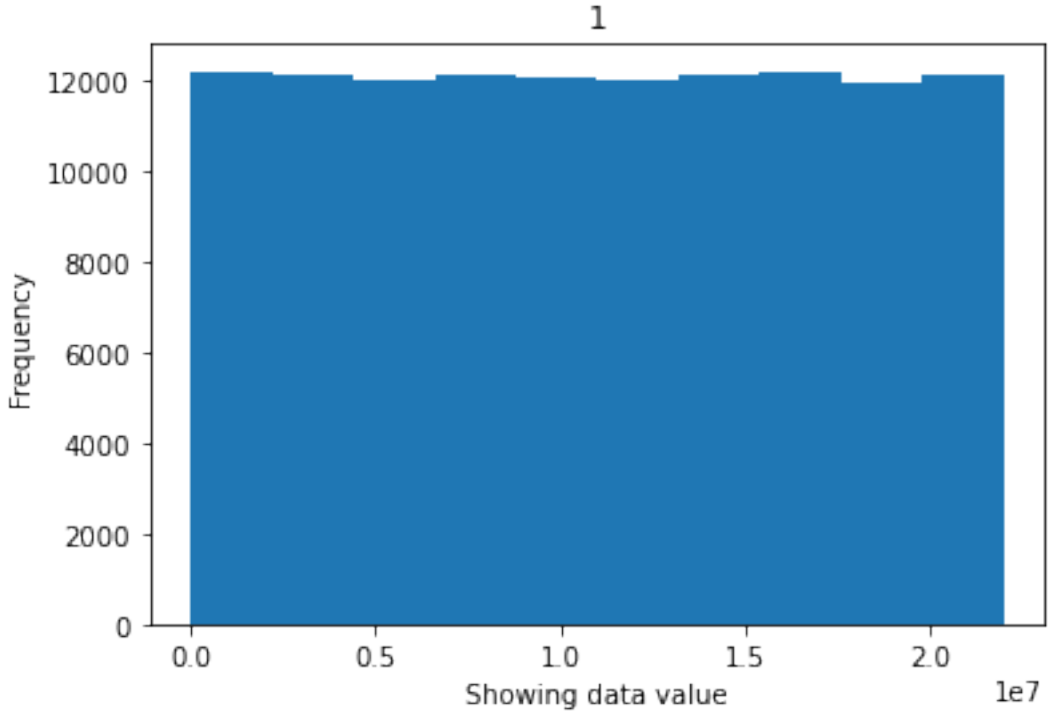

1
12000
10000
8000
6000
4000
2000
0
Frequency
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
1e6
Showing data value

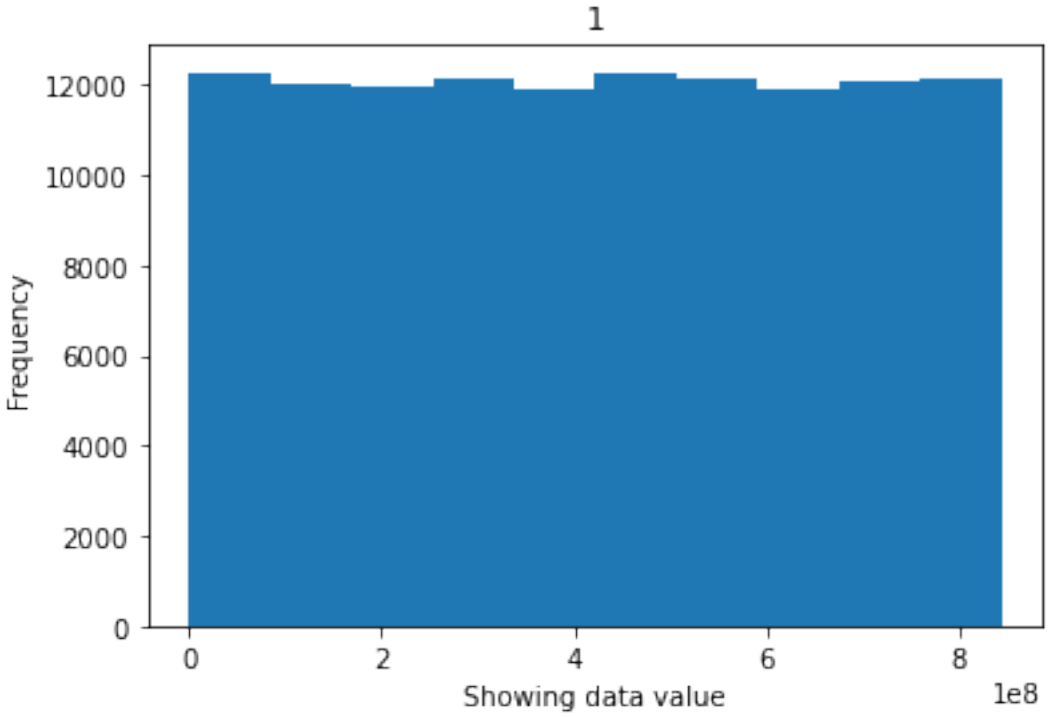

1
12000
10000
8000
6000
4000
2000
0
Frequency
0
200000
400000
600000
800000
Showing data value

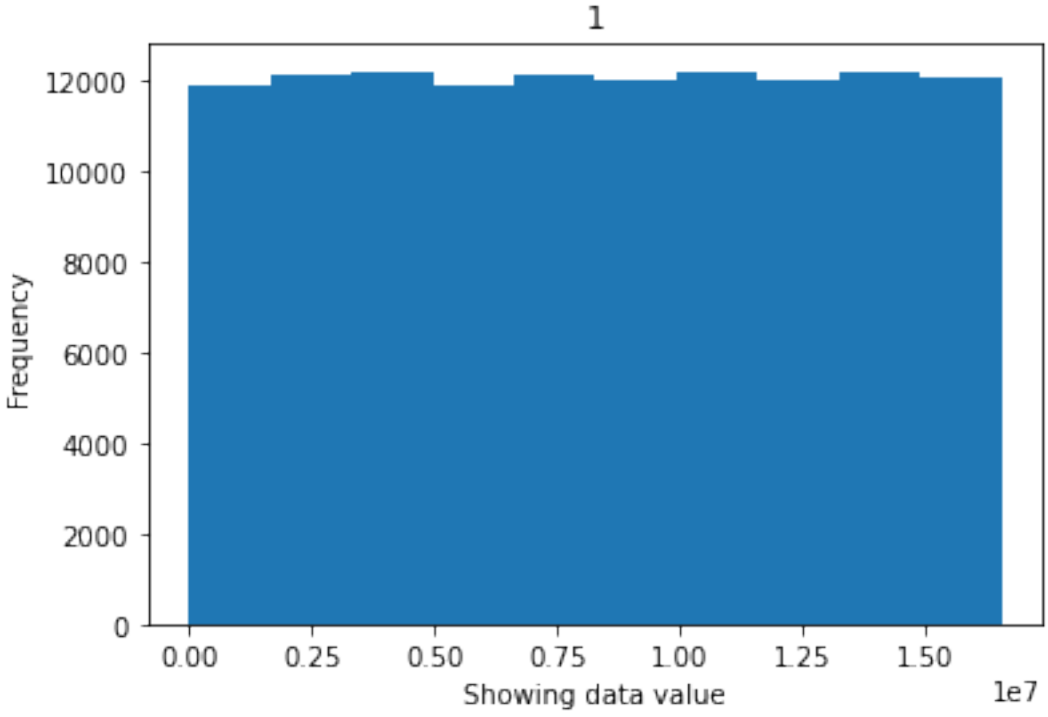

1
12000
10000
8000
6000
4000
2000
0
Frequency
0.0
0.5
1.0
1.5
2.0
Showing data value
1e7

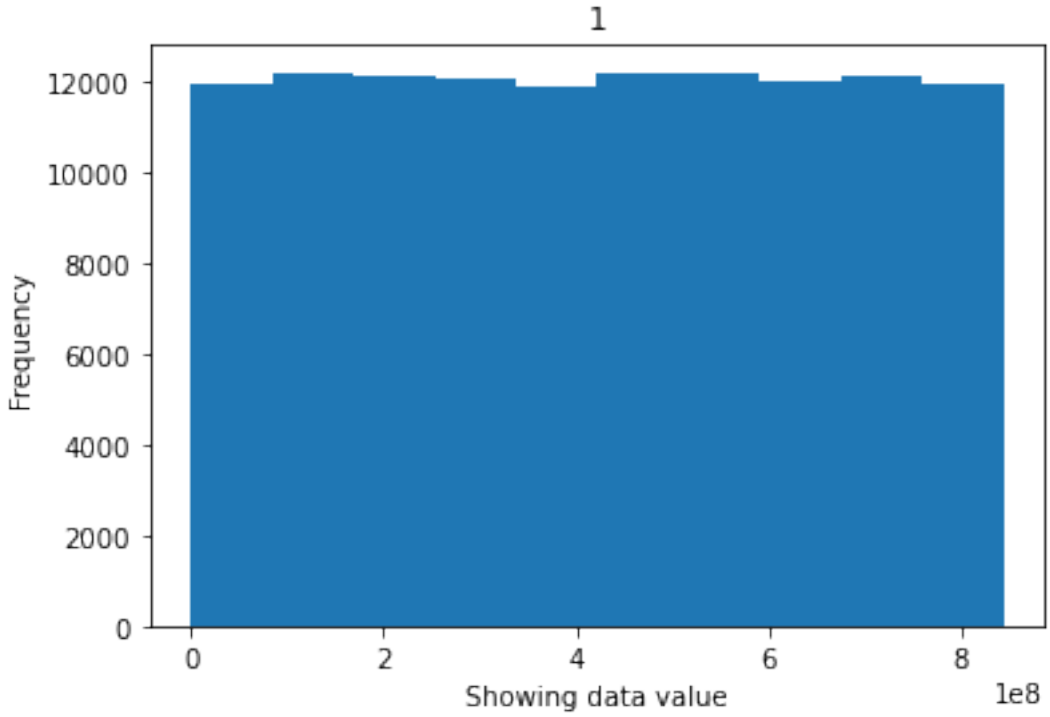

1
12000
10000
8000
6000
4000
2000
0
Frequency
0.0
0.5
1.0
1.5
2.0
Showing data value
1e7

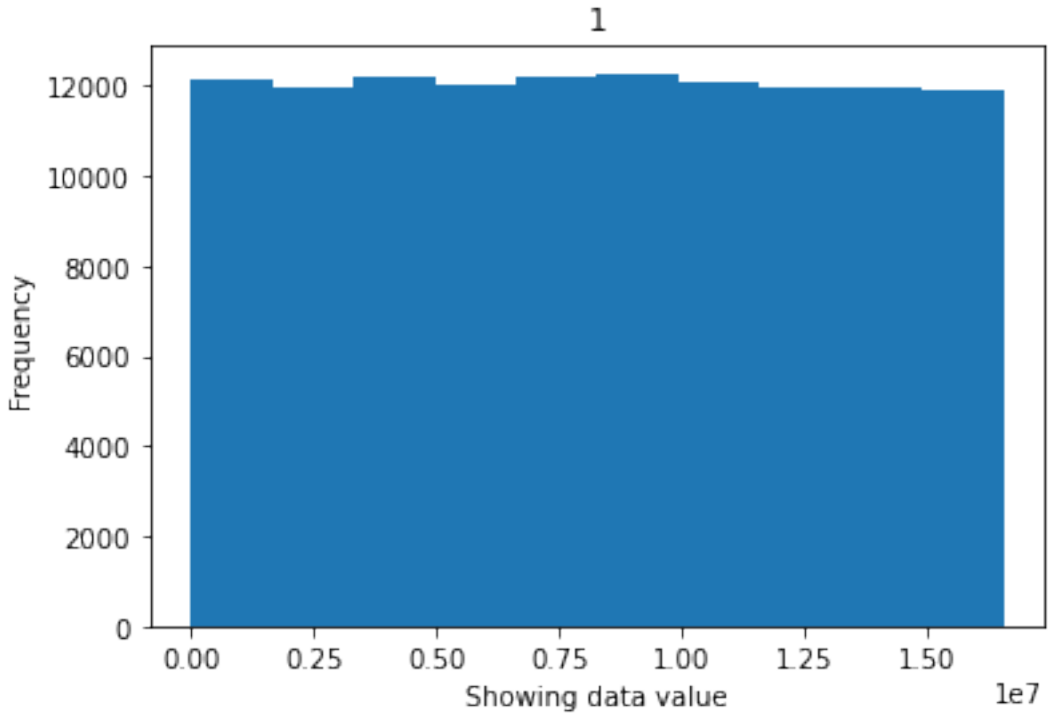

1
12000
10000
8000
6000
4000
2000
0
Frequency
0.0
0.5
1.0
1.5
2.0
Showing data value
1e7


1
12000
10000
8000
6000
4000
2000
0
Frequency
0.0
0.5
1.0
1.5
2.0
Showing data value
1e7


1
12000
10000
8000
6000
4000
2000
0
Frequency
0
2
4
6
8
Showing data value
1e8


1
12000
10000
8000
6000
4000
2000
0
Frequency
0.00
0.25
0.50
0.75
1.00
1.25
1.50
Showing data value
1e7



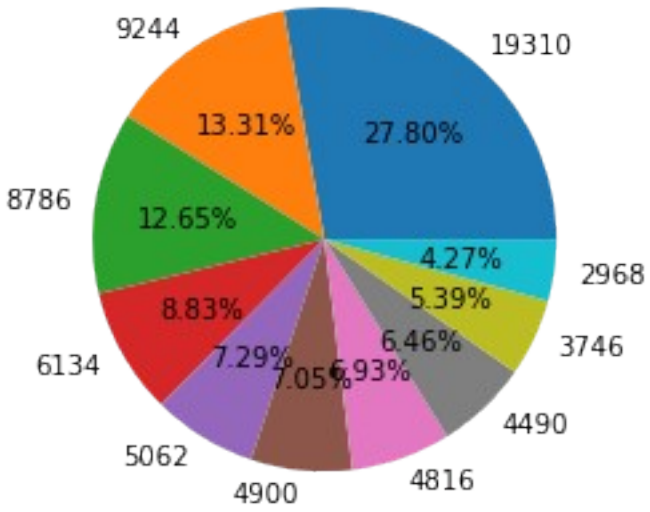



```
top_10_handsets = data["Handset
Type"].value_counts(ascending=False).head(10)
print(top_10_handsets)
```

```
Huawei B528S-23A                19310
Apple iPhone 6S (A1688)          9244
```

```
Apple iPhone 6 (A1586)          8786
Apple iPhone 7 (A1778)          6134
Apple iPhone Se (A1723)         5062
Apple iPhone 8 (A1905)          4900
undefined                       4816
Apple iPhone Xr (A2105)         4490
Apple iPhone X (A1901)          3746
Apple iPhone 8 Plus (A1897)     2968
Name: Handset Type, dtype: int64
```

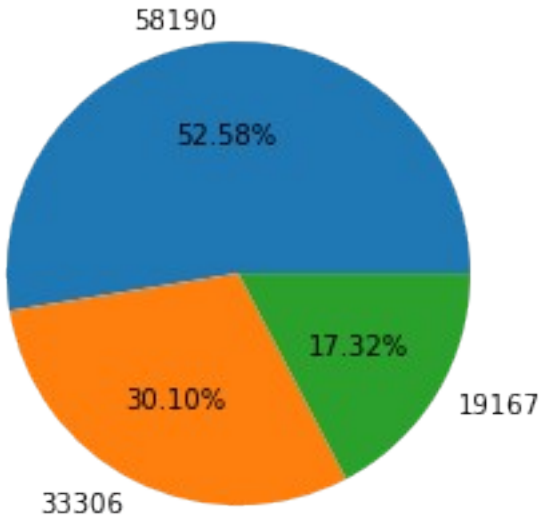
```
data['top_10_handsets']=top_10_handsets

plt.pie(x=top_10_handsets,labels=top_10_handsets,autopct ='%1.2f%%')
plt.show()
```



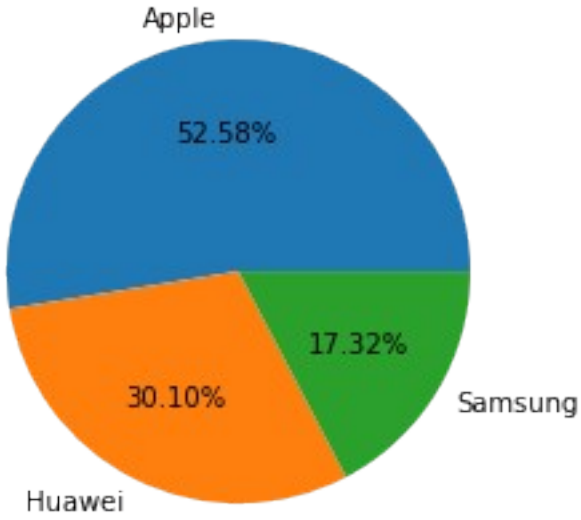

```
top_3_Manufacturer = data["Handset
Manufacturer"].value_counts(ascending=False).head(3)
print(top_3_Manufacturer )
```

```
Apple      58190
Huawei     33306
Samsung    19167
Name: Handset Manufacturer, dtype: int64
```

```
data['top_10_handsets']=top_10_handsets

plt.pie(x=top_3_Manufacturer,labels=top_3_Manufacturer,autopct ='%1.2f
%%')
plt.show()
```




```
plt.pie(labels=top_3_Manufacturer.index,x=top_3_Manufacturer,autopct =
"%1.2f%%")
plt.show();
```




```
data1.columns
```

```
Index(['Bearer Id', 'Dur. (ms)', 'Activity Duration DL (ms)',
       'Activity Duration UL (ms)', 'Social Media DL (Bytes)',
       'Social Media UL (Bytes)', 'Google DL (Bytes)', 'Google UL
(Bytes)',
       'Email DL (Bytes)', 'Email UL (Bytes)', 'Youtube DL (Bytes)',
       'Youtube UL (Bytes)', 'Netflix DL (Bytes)', 'Netflix UL
(Bytes)',
```

```
       'Gaming DL (Bytes)', 'Gaming UL (Bytes)', 'Other DL (Bytes)',
       'Other UL (Bytes)'],
      dtype='object')
```

```
data1.skew
```

```
<bound method NDFrame._add_numeric_operations.<locals>.skew of
Bearer Id  Dur. (ms)  Activity Duration DL (ms)  \
0       6.917538e+18    24534.0                   131798.0
1       6.917538e+18    27786.0                   401941.0
2       6.917538e+18    15635.0                    73347.0
3       6.917538e+18    24264.0                   117340.0
4       6.917538e+18    15548.0                    76969.0
...              ...        ...                        ...
110427  1.318654e+19    80024.0                  2512362.0
110428  1.318654e+19   145291.0                     2067.0
110429  1.318654e+19    86399.0                  3968131.0
110430  1.318654e+19    86399.0                  1689999.0
110431  1.318654e+19   103113.0                        0.0

        Activity Duration UL (ms)  Social Media DL (Bytes)  \
0                        101470.0                2404741.0
1                        399092.0                 944612.0
2                         81378.0                1817239.0
3                        347852.0                1867318.0
4                         80241.0                1502346.0
...                           ...                      ...
110427                  2437668.0                3240226.0
110428                    45217.0                3062671.0
110429                  3537154.0                 720996.0
110430                  1513764.0                2492460.0
110431                    30367.0                1314234.0

        Social Media UL (Bytes)  Google DL (Bytes)  Google UL (Bytes)
\
0                        2410.0          5791591.0          2871336.0

1                        2827.0         10373157.0            56392.0

2                       19827.0           269988.0          3696393.0

3                       18928.0          1689296.0           195216.0

4                       37348.0          8917833.0           701876.0

...                         ...                ...                ...

110427                  38284.0          2036152.0          2271168.0

110428                  48953.0          9363661.0          4001970.0
```

```
110429                  42836.0          1541915.0          2100839.0

110430                  39905.0         11318188.0           466218.0

110431                  27938.0          6969652.0          3756009.0


        Email DL (Bytes)  Email UL (Bytes)  Youtube DL (Bytes)  \
0               782388.0          806920.0           6139644.0
1               128003.0           34038.0           5385159.0
2              3191192.0          896670.0          12347020.0
3               740633.0          590043.0          15231815.0
4               332813.0          537902.0          14602694.0
...                  ...               ...                ...
110427         2410615.0          387548.0          12404964.0
110428         2192057.0          866373.0          22147919.0
110429         2315638.0          839789.0          11879062.0
110430         2612190.0          618629.0          22163800.0
110431         3317462.0          408257.0          12099319.0

        Youtube UL (Bytes)  Netflix DL (Bytes)  Netflix UL (Bytes)  \
0                2071526.0          19494278.0          14668354.0
1                4295851.0          15755839.0           1300571.0
2               11089528.0           2859358.0           1738176.0
3                8401567.0          21563985.0           2817981.0
4                1237356.0           4506463.0          17006591.0
...                    ...                ...                ...
110427           4343114.0          11108134.0          21649273.0
110428           2152449.0          21468525.0           8603105.0
110429           1290963.0          22596930.0          11943452.0
110430          16763435.0           9522397.0           8346624.0
110431           1636122.0           9992219.0          17624886.0

        Gaming DL (Bytes)  Gaming UL (Bytes)  Other DL (Bytes)  \
0            466109357.0          5333340.0       670751043.0
1            821879090.0          8521398.0       472846860.0
2            805301713.0         16257481.0        24303797.0
3            583864716.0          6992868.0       685122214.0
4            542337372.0         15820659.0       463698729.0
...                  ...                ...                ...
110427       114093049.0          2834548.0       695881178.0
110428       328766801.0          7569327.0       371261255.0
110429       833634251.0         10607174.0       697260277.0
110430       338246033.0          1845068.0        17385489.0
110431       293519955.0         16295588.0       440290470.0

        Other UL (Bytes)
0             15950724.0
```

```
1              1337849.0
2             15907613.0
3              2966860.0
4             11939642.0
...                  ...
110427         3888729.0
110428        12947410.0
110429         9094407.0
110430        12797797.0
110431         6398758.0

[110432 rows x 18 columns]>
```

```
for i in data.columns:
    if data[i].dtypes == "object":
        mode_value = data[i].mode()[0]
        data[i].fillna(mode_value, inplace = True)
    elif data[i].dtypes != "object":
        mean_value = data[i].mean()
        data[i].fillna(mean_value, inplace = True)
```

```
data.isnull().sum()
```

```
Bearer Id                                0
Start                                    0
Start ms                                 0
End                                      0
End ms                                   0
Dur. (ms)                                0
IMSI                                     0
MSISDN/Number                            0
IMEI                                     0
Last Location Name                       0
Avg RTT DL (ms)                          0
Avg RTT UL (ms)                          0
Avg Bearer TP DL (kbps)                  0
Avg Bearer TP UL (kbps)                  0
DL TP < 50 Kbps (%)                      0
50 Kbps < DL TP < 250 Kbps (%)           0
250 Kbps < DL TP < 1 Mbps (%)            0
DL TP > 1 Mbps (%)                       0
UL TP < 10 Kbps (%)                      0
10 Kbps < UL TP < 50 Kbps (%)            0
50 Kbps < UL TP < 300 Kbps (%)           0
UL TP > 300 Kbps (%)                     0
Activity Duration DL (ms)                0
Activity Duration UL (ms)                0
Dur. (ms).1                              0
Handset Manufacturer                     0
Handset Type                             0
```

```
Nb of sec with Vol DL < 6250B          0
Nb of sec with Vol UL < 1250B          0
Social Media DL (Bytes)                0
Social Media UL (Bytes)                0
Google DL (Bytes)                      0
Google UL (Bytes)                      0
Email DL (Bytes)                       0
Email UL (Bytes)                       0
Youtube DL (Bytes)                     0
Youtube UL (Bytes)                     0
Netflix DL (Bytes)                     0
Netflix UL (Bytes)                     0
Gaming DL (Bytes)                      0
Gaming UL (Bytes)                      0
Other DL (Bytes)                       0
Other UL (Bytes)                       0
Total UL (Bytes)                       0
Total DL (Bytes)                       0
top_10_handsets                   120739
dtype: int64
```

```
IQR = 'Quartile3 – Quartile1'
```

```
Q1 = np.percentile(data1,25)
Q2 = np.percentile(data1,50)
Q3 = np.percentile(data1,75)
print(Q1,Q2,Q3)
```

```
541061.0 3936649.5 16100163.25
```

```
IQR=Q3-Q1
print(IQR)
```

```
15559102.25
```

```
upper = 'Q3 +1.5*IQR'

lower = 'Q1 – 1.5*IQR'
```

```
upper=Q3+1.5*IQR
upper_array=np.array(user_behaviour['Dur. (ms)']>=upper)
print("Upper Bound:",upper)
print(upper_array.sum())

#Below Lower bound
lower=Q1-1.5*IQR
lower_array=np.array(user_behaviour['Dur. (ms)']<=lower)
print("Lower Bound:",lower)
print(lower_array.sum())
```

```
Upper Bound: 39438816.625
0
```

```
Lower Bound: -22797592.375
0
```

```
upperrange=(Q3+IQR)*1.5
print('upperrange=',(Q3+IQR)*1.5)
lowerrange=(Q1-IQR)*1.5
print('lowerrange=',(Q1-IQR)*1.5)
```

```
upperrange= 47488898.25
lowerrange= -22527061.875
```

```
def find_outlier(data1,threshold=1.5):
    outliers=pd.DataFrame()
    for i in data1.columns:
        if data1[i].dtype !='object':

            i_outlier=data1[(data1[i]<lowerrange)|
(data1[i]>upperrange)]
            outlier=pd.DataFrame()
            outlier=pd.concat([outlier,i_outlier])
            data1[i]=np.where((data1[i]<lowerrange)|
(data1[i]>upperrange),data1[i].mean(),data1[i])
    return outlier
```

```
find_outlier(data1)
```

```
          Bearer Id  Dur. (ms)  Activity Duration DL (ms)  \
7880    9.907303e+18  1063952.0               6.601000e+05
29394   9.907303e+18  1247317.0               8.702510e+05
67253   9.907303e+18   106584.0               1.315224e+06
68806   9.907303e+18    67198.0               8.002700e+04
68924   9.907303e+18  1507998.0               5.890606e+06
69801   9.907303e+18   518394.0               2.464303e+06
69835   9.907303e+18   340788.0               1.571891e+07
69905   9.907303e+18   431996.0               2.464303e+06
70298   9.907303e+18   431997.0               2.021032e+07
70336   9.907303e+18  1468804.0               2.464303e+06
70424   9.907303e+18  1050221.0               8.532424e+06
70447   9.907303e+18   431996.0               1.974265e+07
70602   9.907303e+18   775239.0               2.581390e+07
70818   9.907303e+18   864001.0               4.044378e+07
71112   9.907303e+18   325535.0               4.041298e+07
71133   9.907303e+18   431996.0               3.730937e+07
71675   9.907303e+18   417975.0               2.464303e+06
72634   9.907303e+18  1960914.0               1.596210e+07
72645   9.907303e+18  1475653.0               2.331734e+07
72903   9.907303e+18   864002.0               7.510138e+06
73175   9.907303e+18  1117511.0               1.185512e+07
73256   9.907303e+18  1319645.0               2.694907e+07
73414   9.907303e+18   431135.0               2.464303e+06
73484   9.907303e+18  2418727.0               2.091366e+07
```

73730	9.907303e+18	1296005.0	4.287512e+06
74419	9.907303e+18	1734912.0	2.464303e+06
74670	9.907303e+18	1131647.0	4.433025e+06
74695	9.907303e+18	928854.0	1.037289e+07
74717	9.907303e+18	909952.0	2.187496e+07
74772	9.907303e+18	1556434.0	2.066912e+07
75671	9.907303e+18	880384.0	7.693302e+06
76321	9.907303e+18	1501420.0	7.032942e+06
76756	9.907303e+18	784483.0	2.464303e+06
77151	9.907303e+18	2377940.0	1.762254e+06
77259	9.907303e+18	1262072.0	2.626484e+07
77502	9.907303e+18	622221.0	2.385087e+07
80052	9.907303e+18	1534218.0	7.112424e+06
81674	9.907303e+18	830722.0	9.724392e+06
84259	9.907303e+18	2170150.0	9.523994e+06
84739	9.907303e+18	431996.0	2.464303e+06
84963	9.907303e+18	343221.0	1.361760e+07
85249	9.907303e+18	372584.0	1.900233e+07
85461	9.907303e+18	2484980.0	3.943613e+07
85795	9.907303e+18	2052703.0	4.238755e+06
87354	9.907303e+18	465565.0	1.790148e+07
87681	9.907303e+18	1317282.0	2.060609e+06
88288	9.907303e+18	2423726.0	8.160787e+06
88555	9.907303e+18	3990246.0	2.022999e+07
89327	9.907303e+18	518394.0	3.029330e+07
89341	9.907303e+18	1329280.0	2.464303e+06
89441	9.907303e+18	1042956.0	2.758022e+06
89680	9.907303e+18	2604267.0	2.585884e+06
89933	9.907303e+18	1390780.0	3.873719e+07
97244	9.907303e+18	1572096.0	1.132443e+06
100587	9.907303e+18	4120953.0	1.277930e+07
100662	9.907303e+18	2524675.0	2.023511e+06

	Activity Duration UL (ms)	Social Media DL (Bytes)	\
7880	6.271550e+05	12065608.0	
29394	8.813150e+05	5318154.0	
67253	1.169613e+06	5376189.0	
68806	1.039640e+05	8414716.0	
68924	5.281474e+06	9889718.0	
69801	1.893064e+06	13054280.0	
69835	1.556476e+07	7516441.0	
69905	1.893064e+06	6649767.0	
70298	4.395977e+07	10828963.0	
70336	2.302037e+07	10858194.0	
70424	7.510434e+06	9554325.0	
70447	1.300915e+07	11152831.0	
70602	2.286255e+07	10807105.0	
70818	1.358318e+07	3422143.0	
71112	2.560577e+07	11123354.0	
71133	2.443059e+07	5137157.0	

71675	1.893064e+06	7537420.0
72634	8.794107e+06	11238224.0
72645	2.194194e+07	8886660.0
72903	6.393964e+06	3867243.0
73175	1.051952e+07	5496294.0
73256	1.801094e+07	7400944.0
73414	1.893064e+06	8244196.0
73484	1.265915e+07	13445179.0
73730	3.297345e+06	9579071.0
74419	2.960062e+07	15691444.0
74670	1.479171e+06	9779800.0
74695	9.395395e+06	10950957.0
74717	1.531425e+07	11091536.0
74772	1.742055e+07	11486692.0
75671	6.323335e+06	12154285.0
76321	6.072272e+06	8724177.0
76756	3.086231e+07	9069813.0
77151	1.404338e+06	12565025.0
77259	2.507784e+07	10358833.0
77502	1.273152e+07	7105536.0
80052	5.217562e+06	9727711.0
81674	5.112691e+06	5942916.0
84259	8.136190e+06	12985618.0
84739	3.224348e+07	5035665.0
84963	1.278085e+07	8762084.0
85249	1.949376e+07	9338505.0
85461	2.910492e+07	17685208.0
85795	3.823250e+06	12387634.0
87354	1.282601e+07	10209251.0
87681	2.089353e+06	8333925.0
88288	7.424225e+06	17392073.0
88555	2.255607e+07	13127153.0
89327	2.026302e+07	11545789.0
89341	1.893064e+06	26188977.0
89441	1.848506e+06	12118788.0
89680	6.576600e+05	18205049.0
89933	3.077506e+07	11500407.0
97244	1.032171e+06	6485564.0
100587	1.117550e+07	10697509.0
100662	1.319980e+06	17325759.0

	Social Media UL (Bytes)	Google DL (Bytes)	Google UL (Bytes)
7880	192836.0	1.997650e+07	15961483.0
29394	154867.0	4.114084e+07	11369457.0
67253	102915.0	1.798342e+07	3839007.0
68806	116568.0	1.593428e+07	13860606.0

68924	138392.0	3.130054e+07	8320695.0
69801	191178.0	2.240049e+07	14137153.0
69835	73022.0	1.238805e+07	11990054.0
69905	224580.0	2.305591e+07	8820964.0
70298	146130.0	2.235237e+07	15170823.0
70336	251973.0	3.702023e+07	11627562.0
70424	170575.0	3.147998e+07	10324851.0
70447	70188.0	2.441095e+07	14271845.0
70602	148757.0	3.432890e+07	13367484.0
70818	61927.0	2.175851e+07	6683615.0
71112	162275.0	1.011205e+07	6098489.0
71133	209370.0	1.509313e+07	14685789.0
71675	161070.0	3.290942e+07	10970558.0
72634	158257.0	4.309459e+07	9152556.0
72645	195111.0	2.520216e+07	8523422.0
72903	163226.0	2.287799e+07	6568526.0
73175	109965.0	2.105290e+07	11014607.0
73256	170013.0	2.795737e+07	8393857.0
73414	174485.0	1.994129e+07	7234134.0
73484	247287.0	4.601201e+07	12333099.0
73730	226963.0	3.300066e+07	11295256.0
74419	314047.0	6.288150e+06	24423584.0
74670	238435.0	3.901363e+07	14067047.0
74695	63483.0	3.233953e+07	4269581.0

74717	86373.0	4.662386e+07	12251325.0
74772	180416.0	4.166681e+07	10748639.0
75671	171529.0	3.471458e+07	12284439.0
76321	195794.0	2.392670e+07	11751436.0
76756	108306.0	2.154879e+07	9557138.0
77151	335767.0	6.288150e+06	15865041.0
77259	151199.0	3.118401e+07	7666551.0
77502	138161.0	3.103030e+07	7151987.0
80052	234882.0	3.917658e+07	9308525.0
81674	158739.0	3.017979e+07	16186767.0
84259	114504.0	4.721672e+07	11206798.0
84739	211437.0	2.336369e+07	11918900.0
84963	180096.0	1.662024e+07	10393266.0
85249	136809.0	3.157844e+07	7690469.0
85461	206719.0	6.288150e+06	17813083.0
85795	131985.0	3.624851e+07	11220485.0
87354	123753.0	2.146870e+07	8217477.0
87681	228510.0	2.818442e+07	8181989.0
88288	261469.0	3.390403e+07	13650023.0
88555	165881.0	4.727747e+07	11731439.0
89327	132766.0	4.209783e+07	13554999.0
89341	478166.0	6.288150e+06	38617483.0
89441	249117.0	2.540394e+07	14660911.0
89680	289172.0	4.669842e+07	17138263.0

89933	258236.0	3.413013e+07	11168318.0
97244	186380.0	3.076997e+07	13146865.0
100587	260729.0	3.782076e+07	7857451.0
100662	257241.0	4.354167e+07	18335688.0

	Email DL (Bytes)	Email UL (Bytes)	Youtube DL (Bytes)	\
7880	11231575.0	1972235.0	2.769808e+07	
29394	5183039.0	3010617.0	1.272504e+07	
67253	10881218.0	2066085.0	1.272504e+07	
68806	8290179.0	3827418.0	1.272504e+07	
68924	7655312.0	1697073.0	1.272504e+07	
69801	5145847.0	2480733.0	1.272504e+07	
69835	3269195.0	1532864.0	4.209157e+07	
69905	11689441.0	2173748.0	1.272504e+07	
70298	10241272.0	1365449.0	4.520454e+07	
70336	16704195.0	4254396.0	4.366018e+07	
70424	10319386.0	3747944.0	1.272504e+07	
70447	6872814.0	2075624.0	4.505911e+07	
70602	12771512.0	2403445.0	1.272504e+07	
70818	8087758.0	1060245.0	3.987437e+07	
71112	6644610.0	1499829.0	1.272504e+07	
71133	6360247.0	3240041.0	1.272504e+07	
71675	9144172.0	2738173.0	2.774418e+07	
72634	15734320.0	2826234.0	1.272504e+07	
72645	9882214.0	1882167.0	1.272504e+07	
72903	9289515.0	1644461.0	1.272504e+07	
73175	7893657.0	1431100.0	1.272504e+07	
73256	8925766.0	1056722.0	3.097584e+07	
73414	9359438.0	2099315.0	4.488028e+07	
73484	15539104.0	3771618.0	1.272504e+07	
73730	10919494.0	2114602.0	1.272504e+07	
74419	12657446.0	4737415.0	1.272504e+07	
74670	16654353.0	4274289.0	1.272504e+07	
74695	6917289.0	2531472.0	4.444422e+07	
74717	11105577.0	2421490.0	1.272504e+07	
74772	9962993.0	3068303.0	1.272504e+07	
75671	9988060.0	1112004.0	3.947088e+07	
76321	11104268.0	2219763.0	1.272504e+07	
76756	12798077.0	2604676.0	1.272504e+07	
77151	10517911.0	3109426.0	1.272504e+07	
77259	8284082.0	2765985.0	1.272504e+07	
77502	7791103.0	1936662.0	1.272504e+07	
80052	12415544.0	920393.0	1.272504e+07	
81674	10084278.0	3547208.0	1.272504e+07	
84259	9066036.0	3757376.0	1.272504e+07	

84739	7725244.0	2639373.0	1.272504e+07
84963	7028221.0	2041557.0	3.606245e+07
85249	8472045.0	844196.0	1.272504e+07
85461	18155310.0	3064384.0	1.272504e+07
85795	9127323.0	2497270.0	1.272504e+07
87354	4561436.0	2176805.0	1.272504e+07
87681	8986004.0	1885918.0	1.272504e+07
88288	13329755.0	4640566.0	1.272504e+07
88555	11434548.0	2305977.0	1.272504e+07
89327	12297738.0	2251547.0	1.272504e+07
89341	33798609.0	6744930.0	1.272504e+07
89441	10620792.0	3279302.0	1.272504e+07
89680	12511511.0	3711363.0	1.272504e+07
89933	9822222.0	3082898.0	1.272504e+07
97244	7654421.0	1663475.0	1.272504e+07
100587	9225038.0	3264818.0	1.272504e+07
100662	10220519.0	2880932.0	1.272504e+07

	Youtube UL (Bytes)	Netflix DL (Bytes)	Netflix UL (Bytes)	\
7880	1.203117e+07	1.271781e+07	1.202267e+07	
29394	1.203117e+07	1.271781e+07	1.202267e+07	
67253	3.646618e+07	1.271781e+07	4.547011e+07	
68806	1.203117e+07	1.271781e+07	1.202267e+07	
68924	1.203117e+07	1.271781e+07	2.708597e+07	
69801	3.953070e+07	1.271781e+07	1.202267e+07	
69835	1.203117e+07	4.643770e+07	1.202267e+07	
69905	3.916596e+07	4.160956e+07	1.202267e+07	
70298	1.203117e+07	1.271781e+07	1.202267e+07	
70336	1.203117e+07	1.271781e+07	1.202267e+07	
70424	1.203117e+07	1.271781e+07	1.202267e+07	
70447	1.203117e+07	1.271781e+07	4.187393e+07	
70602	1.203117e+07	1.271781e+07	4.696836e+07	
70818	3.667662e+07	1.271781e+07	1.202267e+07	
71112	2.932434e+07	1.271781e+07	2.393736e+07	
71133	3.791663e+07	1.271781e+07	1.202267e+07	
71675	1.203117e+07	1.271781e+07	1.202267e+07	
72634	4.251627e+07	1.271781e+07	1.202267e+07	
72645	4.355817e+07	1.271781e+07	1.202267e+07	
72903	4.025200e+07	3.572694e+07	1.202267e+07	
73175	1.203117e+07	1.271781e+07	1.202267e+07	
73256	3.530932e+07	1.271781e+07	1.202267e+07	
73414	4.326230e+07	1.271781e+07	1.202267e+07	
73484	1.203117e+07	1.271781e+07	1.202267e+07	
73730	1.203117e+07	4.677622e+07	4.353004e+07	
74419	1.203117e+07	1.271781e+07	1.202267e+07	
74670	1.203117e+07	1.271781e+07	3.938948e+07	
74695	3.848046e+07	4.517117e+07	1.202267e+07	
74717	4.321166e+07	1.271781e+07	3.418807e+07	
74772	1.203117e+07	1.271781e+07	1.202267e+07	
75671	3.593694e+07	1.271781e+07	1.202267e+07	

76321	4.135406e+07	3.994501e+07	3.679697e+07
76756	1.203117e+07	3.197276e+07	1.202267e+07
77151	1.203117e+07	1.271781e+07	1.202267e+07
77259	1.203117e+07	1.271781e+07	2.453050e+07
77502	1.203117e+07	3.074802e+07	3.633603e+07
80052	1.203117e+07	1.271781e+07	1.202267e+07
81674	1.203117e+07	1.271781e+07	1.202267e+07
84259	1.203117e+07	1.271781e+07	1.202267e+07
84739	4.305971e+07	1.271781e+07	4.290793e+07
84963	2.284456e+07	3.765715e+07	2.565069e+07
85249	4.453732e+07	2.801131e+07	1.202267e+07
85461	1.203117e+07	1.271781e+07	1.202267e+07
85795	1.203117e+07	1.271781e+07	1.202267e+07
87354	1.203117e+07	1.271781e+07	1.202267e+07
87681	1.203117e+07	3.789540e+07	1.202267e+07
88288	1.203117e+07	1.271781e+07	1.202267e+07
88555	1.203117e+07	1.271781e+07	1.202267e+07
89327	1.203117e+07	1.271781e+07	1.202267e+07
89341	1.203117e+07	1.271781e+07	1.202267e+07
89441	1.788242e+07	1.271781e+07	1.202267e+07
89680	1.203117e+07	1.271781e+07	1.202267e+07
89933	1.203117e+07	1.271781e+07	1.202267e+07
97244	1.203117e+07	3.534024e+07	1.202267e+07
100587	1.203117e+07	1.271781e+07	1.202267e+07
100662	1.203117e+07	1.271781e+07	1.202267e+07

	Gaming DL (Bytes)	Gaming UL (Bytes)	Other DL (Bytes)	\
7880	4.608041e+08	3.926920e+07	4.607876e+08	
29394	4.608041e+08	4.199390e+07	4.607876e+08	
67253	4.608041e+08	2.129808e+07	4.607876e+08	
68806	4.608041e+08	3.072244e+07	4.607876e+08	
68924	4.608041e+08	3.453025e+07	4.607876e+08	
69801	4.608041e+08	9.066019e+06	4.607876e+08	
69835	4.608041e+08	3.276557e+07	4.607876e+08	
69905	4.608041e+08	9.066019e+06	4.607876e+08	
70298	4.608041e+08	9.066019e+06	4.607876e+08	
70336	4.608041e+08	9.066019e+06	4.607876e+08	
70424	4.608041e+08	3.215753e+07	4.607876e+08	
70447	4.608041e+08	2.912187e+07	4.607876e+08	
70602	4.608041e+08	9.066019e+06	4.607876e+08	
70818	4.608041e+08	2.254635e+07	4.607876e+08	
71112	4.608041e+08	4.603988e+07	4.607876e+08	
71133	4.608041e+08	9.066019e+06	4.607876e+08	
71675	4.608041e+08	9.066019e+06	4.607876e+08	
72634	4.608041e+08	4.367436e+07	4.607876e+08	
72645	4.608041e+08	3.311396e+07	4.607876e+08	
72903	4.608041e+08	3.598716e+07	4.607876e+08	
73175	4.608041e+08	9.066019e+06	4.607876e+08	
73256	4.608041e+08	9.066019e+06	4.607876e+08	
73414	4.608041e+08	9.066019e+06	4.607876e+08	

73484	4.608041e+08	9.066019e+06	4.607876e+08
73730	4.608041e+08	3.547314e+07	4.607876e+08
74419	4.608041e+08	9.066019e+06	4.607876e+08
74670	4.608041e+08	9.066019e+06	4.607876e+08
74695	4.608041e+08	3.187746e+07	4.607876e+08
74717	4.608041e+08	4.712999e+07	4.607876e+08
74772	4.608041e+08	4.504584e+07	4.607876e+08
75671	4.608041e+08	4.505545e+07	4.607876e+08
76321	4.608041e+08	4.424680e+07	4.607876e+08
76756	4.608041e+08	4.295631e+07	4.607876e+08
77151	4.608041e+08	9.066019e+06	4.607876e+08
77259	4.608041e+08	4.703218e+07	4.607876e+08
77502	4.608041e+08	3.756312e+07	4.607876e+08
80052	4.608041e+08	4.673146e+07	4.607876e+08
81674	4.608041e+08	9.066019e+06	4.607876e+08
84259	4.608041e+08	9.066019e+06	4.607876e+08
84739	4.608041e+08	9.066019e+06	4.607876e+08
84963	4.608041e+08	3.011592e+07	4.607876e+08
85249	4.608041e+08	2.644042e+07	4.607876e+08
85461	4.608041e+08	9.066019e+06	4.607876e+08
85795	4.608041e+08	2.898376e+07	4.607876e+08
87354	4.608041e+08	4.299542e+07	4.607876e+08
87681	4.608041e+08	3.743657e+07	4.607876e+08
88288	4.608041e+08	9.066019e+06	4.607876e+08
88555	4.608041e+08	9.066019e+06	4.607876e+08
89327	4.608041e+08	4.017621e+07	4.607876e+08
89341	4.608041e+08	9.066019e+06	4.607876e+08
89441	4.608041e+08	4.278891e+07	4.607876e+08
89680	4.608041e+08	4.252727e+07	4.607876e+08
89933	4.608041e+08	9.066019e+06	4.607876e+08
97244	4.608041e+08	2.269613e+07	4.607876e+08
100587	4.608041e+08	3.516079e+07	4.607876e+08
100662	4.608041e+08	2.893398e+07	4.607876e+08

	Other UL (Bytes)
7880	48706279.0
29394	60651254.0
67253	54043072.0
68806	51538897.0
68924	49706447.0
69801	66599994.0
69835	50753734.0
69905	60561728.0
70298	51209440.0
70336	70192623.0
70424	64474883.0
70447	54217928.0
70602	59246099.0
70818	48154317.0
71112	50913448.0

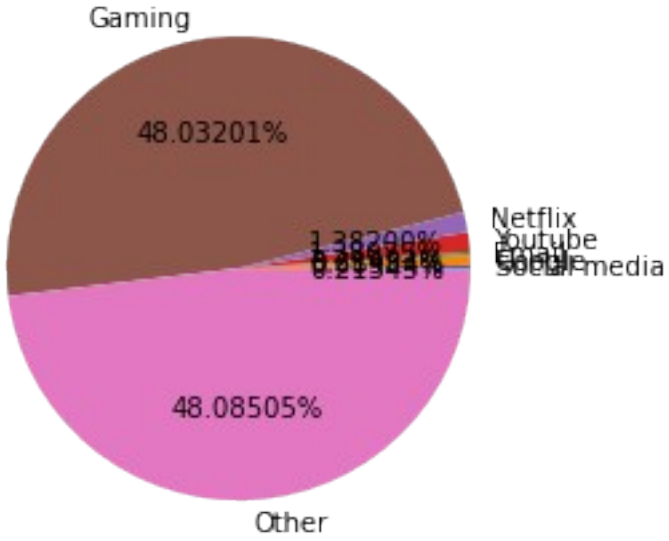
```
71133         47863045.0
71675         54478686.0
72634         61754055.0
72645         48141734.0
72903         49965988.0
73175         48023370.0
73256         48930687.0
73414         52685390.0
73484         53006405.0
73730         50671233.0
74419         79388455.0
74670         48177714.0
74695         48571116.0
74717         61767145.0
74772         52886243.0
75671         49063721.0
76321         58654734.0
76756         47534564.0
77151         54218645.0
77259         56062208.0
77502         47622908.0
80052         76549883.0
81674         49592823.0
84259         68100816.0
84739         52141513.0
84963         49591592.0
85249         48175673.0
85461         76617173.0
85795         56716734.0
87354         57891300.0
87681         50197623.0
88288         73736223.0
88555         68115759.0
89327         48669809.0
89341        128715313.0
89441         59836041.0
89680         75443538.0
89933         54963856.0
97244         56139536.0
100587        64857892.0
100662        50327587.0
```

```
data1.columns
```

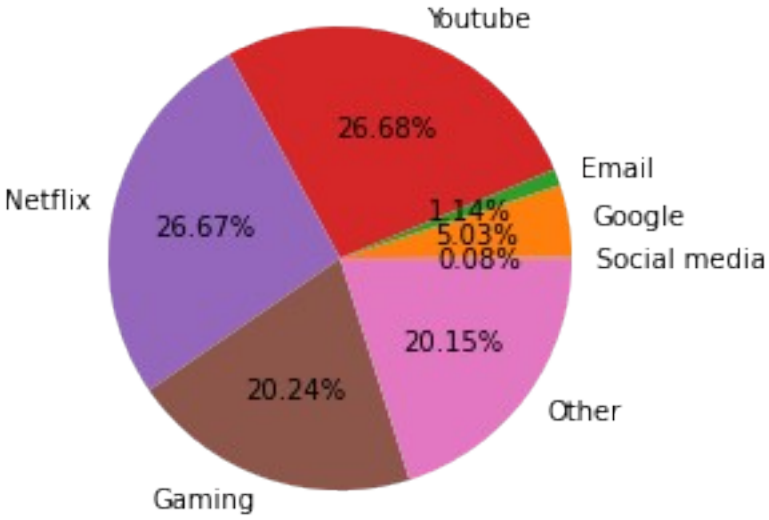
```
Index(['Bearer Id', 'Dur. (ms)', 'Activity Duration DL (ms)',
       'Activity Duration UL (ms)', 'Social Media DL (Bytes)',
       'Social Media UL (Bytes)', 'Google DL (Bytes)', 'Google UL
(Bytes)',
       'Email DL (Bytes)', 'Email UL (Bytes)', 'Youtube DL (Bytes)',
       'Youtube UL (Bytes)', 'Netflix DL (Bytes)', 'Netflix UL
(Bytes)',
```

```
       'Gaming DL (Bytes)', 'Gaming UL (Bytes)', 'Other DL (Bytes)',
       'Other UL (Bytes)'],
      dtype='object')
```

```
# total DL data for each categrory
dl_data=[
       data1['Social Media DL (Bytes)'].sum(),
       data1['Google DL (Bytes)'].sum(),
       data1['Email DL (Bytes)'].sum(),
       data1['Youtube DL (Bytes)'].sum(),
       data1['Netflix DL (Bytes)'].sum(),
       data1['Gaming DL (Bytes)'].sum(),
       data1['Other DL (Bytes)'].sum()
]
# total DL data for each categrory
ul_data=[
       data1['Social Media UL (Bytes)'].sum(),
       data1['Google UL (Bytes)'].sum(),
       data1['Email UL (Bytes)'].sum(),
       data1['Youtube UL (Bytes)'].sum(),
       data1['Netflix UL (Bytes)'].sum(),
       data1['Gaming UL (Bytes)'].sum(),
       data1['Other UL (Bytes)'].sum()
]

labels=["Social
media","Google","Email","Youtube","Netflix","Gaming","Other"]

plt.pie(labels=labels,
        x=dl_data,autopct ='%1.5f%%')
plt.show()
```




```
plt.pie(labels=labels,
        x=ul_data,autopct ='%1.2f%%')
plt.show()
```




```
data1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 110432 entries, 0 to 110431
Data columns (total 18 columns):
 #   Column                    Non-Null Count   Dtype
---  ------                    --------------   -----
 0   Bearer Id                 110432 non-null  float64
 1   Dur. (ms)                 110432 non-null  float64
```

```
 2   Activity Duration DL (ms)  110432 non-null  float64
 3   Activity Duration UL (ms)  110432 non-null  float64
 4   Social Media DL (Bytes)    110432 non-null  float64
 5   Social Media UL (Bytes)    110432 non-null  float64
 6   Google DL (Bytes)          110432 non-null  float64
 7   Google UL (Bytes)          110432 non-null  float64
 8   Email DL (Bytes)           110432 non-null  float64
 9   Email UL (Bytes)           110432 non-null  float64
 10  Youtube DL (Bytes)         110432 non-null  float64
 11  Youtube UL (Bytes)         110432 non-null  float64
 12  Netflix DL (Bytes)         110432 non-null  float64
 13  Netflix UL (Bytes)         110432 non-null  float64
 14  Gaming DL (Bytes)          110432 non-null  float64
 15  Gaming UL (Bytes)          110432 non-null  float64
 16  Other DL (Bytes)           110432 non-null  float64
 17  Other UL (Bytes)           110432 non-null  float64
dtypes: float64(18)
memory usage: 15.2 MB
```

```
#Basic col Analysis
columns=['Dur. (ms)','Start ms','End ms','Total DL (Bytes)','Total UL
(Bytes)']
stat=data[columns].describe()
print(columns)
```

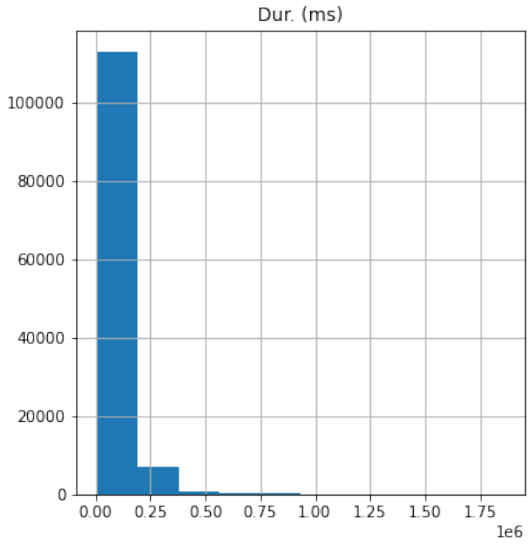
```
['Dur. (ms)', 'Start ms', 'End ms', 'Total DL (Bytes)', 'Total UL
(Bytes)']
```

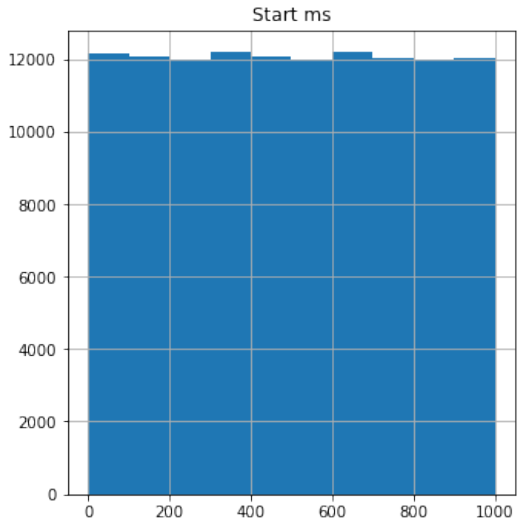
```
# Non Graphical Univariate Analysis
dispersion = data[columns].std()
print(dispersion)
```

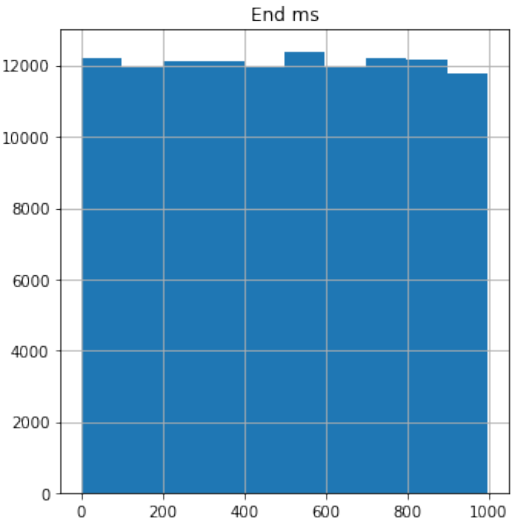
```
Dur. (ms)           7.335495e+04
Start ms            2.887598e+02
End ms              2.879436e+02
Total DL (Bytes)    2.440239e+08
Total UL (Bytes)    1.128297e+07
dtype: float64
```

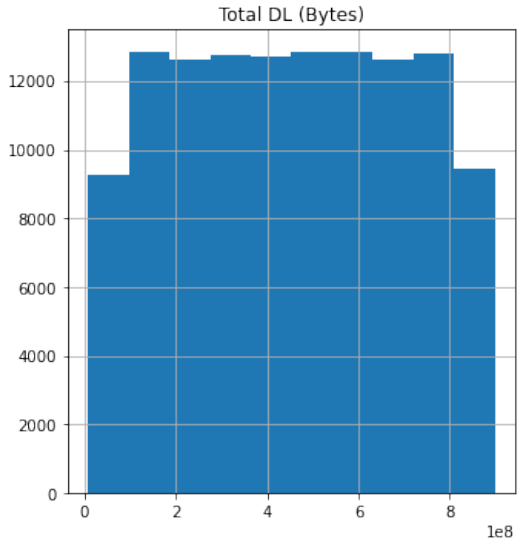
```
#Graphical Univariate Analysis
data.hist(column=columns,bins=10,figsize=(12,20))
```

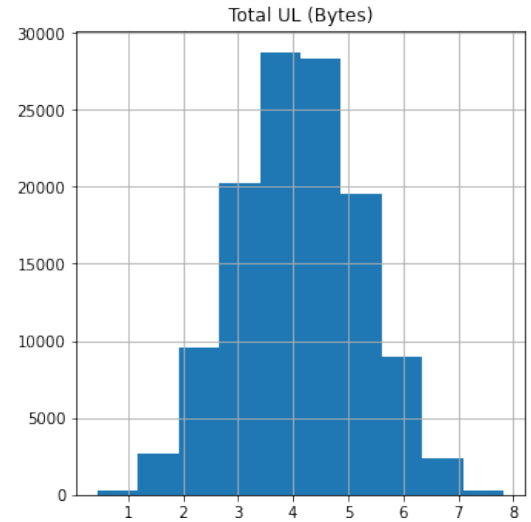
```
array([[<AxesSubplot:title={'center':'Dur. (ms)'}>,
        <AxesSubplot:title={'center':'Start ms'}>],
       [<AxesSubplot:title={'center':'End ms'}>,
        <AxesSubplot:title={'center':'Total DL (Bytes)'}>],
       [<AxesSubplot:title={'center':'Total UL (Bytes)'}>,
        <AxesSubplot:>]], dtype=object)
```


Dur. (ms)
100000
80000
60000
40000
20000
0
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
1e6


Start ms
12000
10000
8000
6000
4000
2000
0
0
200
400
600
800
1000


End ms
12000
10000
8000
6000
4000
2000
0
0
200
400
600
800
1000


Total DL (Bytes)
12000
10000
8000
6000
4000
2000
0
0
2
4
6
8
1e8


Total UL (Bytes)
30000
25000
20000
15000
10000
5000
0
1
2
3
4
5
6
7
8

```
corr_matrix = data[['Social Media DL (Bytes)', 'Social Media UL
(Bytes)',
       'Google DL (Bytes)', 'Google UL (Bytes)', 'Email DL (Bytes)',
       'Email UL (Bytes)', 'Youtube DL (Bytes)', 'Youtube UL (Bytes)',
       'Netflix DL (Bytes)', 'Netflix UL (Bytes)', 'Gaming DL
(Bytes)',
       'Gaming UL (Bytes)', 'Other DL (Bytes)', 'Other UL (Bytes)',
       'Total UL (Bytes)', 'Total DL (Bytes)']]

print(corr_matrix)
```

```
        Social Media DL (Bytes)  Social Media UL (Bytes)  Google DL
(Bytes)  \
0                     1545765.0                  24420.0
1634479.0
1                     1926113.0                   7165.0
3493924.0
6                     1623509.0                   1482.0
1837897.0
7                     3103940.0                  57463.0
1304524.0
9                     1541259.0                  44154.0
8601949.0
...                         ...                      ...
...
149995                 962097.0                  31078.0
1139573.0
149996                3464974.0                  52091.0
9967603.0
149997                2344568.0                   7613.0
2229420.0
149998                1245845.0                  14394.0
3850890.0
149999                 801547.0                  21562.0
4189773.0

        Google UL (Bytes)  Email DL (Bytes)  Email UL (Bytes)  \
0               1271433.0         3563542.0          137762.0
1                920172.0          629046.0          308339.0
6               3068975.0         1838947.0          252032.0
7               2568189.0         2987408.0          512696.0
9                176392.0          522457.0          243589.0
...                   ...               ...               ...
149995           126061.0         3459965.0          820142.0
149996          2817311.0           57639.0          633237.0
149997          2185941.0         1954414.0          167304.0
149998          2734579.0         1525734.0          532543.0
149999          3567494.0         2228270.0          622644.0

        Youtube DL (Bytes)  Youtube UL (Bytes)  Netflix DL (Bytes)  \
```

```
0               15854611.0          2501332.0          8198936.0
1               20247395.0         19111729.0         18338413.0
6                7234465.0          1886295.0         13231006.0
7               12189786.0         19735183.0          6858496.0
9               16338345.0          6573194.0         12211149.0
...                    ...                ...                ...
149995           6550499.0         18003146.0         22468983.0
149996          16191667.0         11763428.0         17883703.0
149997          13877234.0          8288284.0         19350146.0
149998          22660510.0          1855903.0          9963942.0
149999           8817106.0          8305402.0          3322253.0

        Netflix UL (Bytes)  Gaming DL (Bytes)  Gaming UL (Bytes)  \
0                9656251.0        278082303.0         14344150.0
1               17227132.0        608750074.0          1170709.0
6               17981760.0         44796223.0         14084934.0
7               12075206.0        748906189.0         16149622.0
9                 974201.0        811355188.0          1591237.0
...                    ...                ...                ...
149995           7149728.0        784435351.0         12724751.0
149996          19678161.0        526609673.0          9197207.0
149997          21293148.0        626893062.0          4735033.0
149998           5065760.0        553539484.0         13394316.0
149999          13172589.0        352536971.0          2529475.0

        Other DL (Bytes)  Other UL (Bytes)  Total UL (Bytes)  Total DL
(Bytes)
0            171744450.0         8814393.0        36749741.0
308879636.0
1            526904238.0        15055145.0        53800391.0
653384965.0
6            803653557.0        10649768.0        47925246.0
70562047.0
7            537601820.0         7714657.0        58813016.0
775350343.0
9             64856042.0        12815208.0        22417975.0
850570347.0
...                  ...               ...               ...
...
149995       321383162.0        14890486.0        53745392.0
819016468.0
149996         3264510.0        13487416.0        57628851.0
574175259.0
149997       712180387.0         2457758.0        39135081.0
666648844.0
149998       121100856.0        11314729.0        34912224.0
592786405.0
149999       814713113.0         1406930.0        29626096.0
371895920.0
```

```
[120739 rows x 16 columns]
```

```
corr=data1.corr() # correlation in numerical feature
fig=plt.figure(figsize=(15,7))
sns.heatmap(corr,annot=True)
```

```
<AxesSubplot:>
```