Importing Libraries

In [1]:
```
1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 %matplotlib inline
5 import seaborn as sns
```

In [2]:
```
1 import warnings
2 warnings.filterwarnings("ignore")
```

In [3]:
```
1 data=pd.read_csv('telcom_data.csv')
```

In [4]:
```
1 data
```

Out[4]:

	Bearer Id	Start	Start ms	End	End ms	Dur. (ms)	IMSI	MSISDN/Numl
0	1.311448e+19	4/4/2019 12:01	770.0	4/25/2019 14:35	662.0	1823652.0	2.082014e+14	3.366496e+
1	1.311448e+19	4/9/2019 13:04	235.0	4/25/2019 8:15	606.0	1365104.0	2.082019e+14	3.368185e+
2	1.311448e+19	4/9/2019 17:42	1.0	4/25/2019 11:58	652.0	1361762.0	2.082003e+14	3.376063e+
3	1.311448e+19	4/10/2019 0:31	486.0	4/25/2019 7:36	171.0	1321509.0	2.082014e+14	3.375034e+
4	1.311448e+19	4/12/2019 20:10	565.0	4/25/2019 10:40	954.0	1089009.0	2.082014e+14	3.369980e+
...	...	...	...	...	...	...	...	
149996	7.277826e+18	4/29/2019 7:28	451.0	4/30/2019 6:02	214.0	81230.0	2.082022e+14	3.365069e+
149997	7.349883e+18	4/29/2019 7:28	483.0	4/30/2019 10:41	187.0	97970.0	2.082019e+14	3.366345e+
149998	1.311448e+19	4/29/2019 7:28	283.0	4/30/2019 10:46	810.0	98249.0	2.082017e+14	3.362189e+
149999	1.311448e+19	4/29/2019 7:28	696.0	4/30/2019 10:40	327.0	97910.0	2.082021e+14	3.361962e+
150000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N

150001 rows × 55 columns

In [5]:
```
1 data["Bearer Id"].unique
```

Out[5]:
```
<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>
```

In [6]:
```
1 data.head()
```

Out[6]:

	Bearer Id	Start	Start ms	End	End ms	Dur. (ms)	IMSI	MSISDN/Number	
0	1.311448e+19	4/4/2019 12:01	770.0	4/25/2019 14:35	662.0	1823652.0	2.082014e+14	3.366496e+10	3
1	1.311448e+19	4/9/2019 13:04	235.0	4/25/2019 8:15	606.0	1365104.0	2.082019e+14	3.368185e+10	3
2	1.311448e+19	4/9/2019 17:42	1.0	4/25/2019 11:58	652.0	1361762.0	2.082003e+14	3.376063e+10	3
3	1.311448e+19	4/10/2019 0:31	486.0	4/25/2019 7:36	171.0	1321509.0	2.082014e+14	3.375034e+10	3
4	1.311448e+19	4/12/2019 20:10	565.0	4/25/2019 10:40	954.0	1089009.0	2.082014e+14	3.369980e+10	3

5 rows × 55 columns

In [7]:
```
1 data.tail()
```

Out[7]:

	Bearer Id	Start	Start ms	End	End ms	Dur. (ms)	IMSI	MSISDN/Numbe
149996	7.277826e+18	4/29/2019 7:28	451.0	4/30/2019 6:02	214.0	81230.0	2.082022e+14	3.365069e+1(
149997	7.349883e+18	4/29/2019 7:28	483.0	4/30/2019 10:41	187.0	97970.0	2.082019e+14	3.366345e+1(
149998	1.311448e+19	4/29/2019 7:28	283.0	4/30/2019 10:46	810.0	98249.0	2.082017e+14	3.362189e+1(
149999	1.311448e+19	4/29/2019 7:28	696.0	4/30/2019 10:40	327.0	97910.0	2.082021e+14	3.361962e+1(
150000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

5 rows × 55 columns

In [8]:
```
1 data.shape
```

Out[8]:
```
(150001, 55)
```

In [9]:

```
1 data.dtypes
```

Out[9]:
```
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
```

In [10]:
```
1 data.describe()
```

Out[10]:

	Bearer Id	Start ms	End ms	Dur. (ms)	IMSI	MSISDN/Numbe
count	1.490100e+05	150000.000000	150000.000000	1.500000e+05	1.494310e+05	1.489350e+0
mean	1.013887e+19	499.188200	498.800880	1.046086e+05	2.082016e+14	4.188282e+1
std	2.893173e+18	288.611834	288.097653	8.103762e+04	2.148809e+10	2.447443e+1
min	6.917538e+18	0.000000	0.000000	7.142000e+03	2.040471e+14	3.360100e+1
25%	7.349883e+18	250.000000	251.000000	5.744050e+04	2.082014e+14	3.365130e+1
50%	7.349883e+18	499.000000	500.000000	8.639900e+04	2.082015e+14	3.366371e+1
75%	1.304243e+19	749.000000	750.000000	1.324302e+05	2.082018e+14	3.368349e+1
max	1.318654e+19	999.000000	999.000000	1.859336e+06	2.140743e+14	8.823971e+1

8 rows × 50 columns

In [11]:

```
1 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
```

In [12]:
```
1 data.duplicated()
```

Out[12]:
```
0         False
1         False
2         False
3         False
4         False
          ...
149996    False
149997    False
149998    False
149999    False
150000    False
Length: 150001, dtype: bool
```

In [13]:
```
1 null_pct= data.isnull().sum() / len(data) * 100
```

In [14]:

```
1 null_pct
```

Out[14]:

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

In [15]:
```
cols_to_drop = null_pct[null_pct > 50].index.tolist()
cols_to_drop
```

Out[15]:
```
['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']
```

In [16]:
```
data1=data.drop(cols_to_drop,axis=1)
data1
```

Out[16]:

	Bearer Id	Start	Start ms	End	End ms	Dur. (ms)	IMSI	MSISDN/Numl
0	1.311448e+19	4/4/2019 12:01	770.0	4/25/2019 14:35	662.0	1823652.0	2.082014e+14	3.366496e+
1	1.311448e+19	4/9/2019 13:04	235.0	4/25/2019 8:15	606.0	1365104.0	2.082019e+14	3.368185e+
2	1.311448e+19	4/9/2019 17:42	1.0	4/25/2019 11:58	652.0	1361762.0	2.082003e+14	3.376063e+
3	1.311448e+19	4/10/2019 0:31	486.0	4/25/2019 7:36	171.0	1321509.0	2.082014e+14	3.375034e+
4	1.311448e+19	4/12/2019 20:10	565.0	4/25/2019 10:40	954.0	1089009.0	2.082014e+14	3.369980e+
...	...	...	...	...	...	...	...	
149996	7.277826e+18	4/29/2019 7:28	451.0	4/30/2019 6:02	214.0	81230.0	2.082022e+14	3.365069e+
149997	7.349883e+18	4/29/2019 7:28	483.0	4/30/2019 10:41	187.0	97970.0	2.082019e+14	3.366345e+
149998	1.311448e+19	4/29/2019 7:28	283.0	4/30/2019 10:46	810.0	98249.0	2.082017e+14	3.362189e+
149999	1.311448e+19	4/29/2019 7:28	696.0	4/30/2019 10:40	327.0	97910.0	2.082021e+14	3.361962e+
150000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N

150001 rows × 45 columns

In [17]:
```
1 data1.isnull().sum()
```

Out[17]:

```
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
```

In [18]:
```
data1=data1.dropna(how='any',axis=0)
data1
```

Out[18]:

	Bearer Id	Start	Start ms	End	End ms	Dur. (ms)	IMSI	MSISDN/Numl
0	1.311448e+19	4/4/2019 12:01	770.0	4/25/2019 14:35	662.0	1823652.0	2.082014e+14	3.366496e+
1	1.311448e+19	4/9/2019 13:04	235.0	4/25/2019 8:15	606.0	1365104.0	2.082019e+14	3.368185e+
6	1.311448e+19	4/13/2019 8:41	612.0	4/25/2019 8:16	168.0	1035261.0	2.082014e+14	3.366537e+
7	1.304243e+19	4/14/2019 2:11	592.0	4/25/2019 2:26	512.0	951292.0	2.082010e+14	3.376349e+
9	1.304243e+19	4/15/2019 0:32	0.0	4/25/2019 0:40	284.0	864482.0	2.082003e+14	3.365922e+
...	...	...	...	...	...	...	...	
149995	1.304243e+19	4/29/2019 7:28	615.0	4/30/2019 0:01	407.0	59587.0	2.082014e+14	3.366865e+
149996	7.277826e+18	4/29/2019 7:28	451.0	4/30/2019 6:02	214.0	81230.0	2.082022e+14	3.365069e+
149997	7.349883e+18	4/29/2019 7:28	483.0	4/30/2019 10:41	187.0	97970.0	2.082019e+14	3.366345e+
149998	1.311448e+19	4/29/2019 7:28	283.0	4/30/2019 10:46	810.0	98249.0	2.082017e+14	3.362189e+
149999	1.311448e+19	4/29/2019 7:28	696.0	4/30/2019 10:40	327.0	97910.0	2.082021e+14	3.361962e+

120739 rows × 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

In [19]:
```
top_10_handsets = data1["Handset Type"].value_counts(ascending=False).hea
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
```

In [20]:
```
top_3_Manufacturer = data["Handset Manufacturer"].value_counts(ascending=
print(top_3_Manufacturer )
```

```
Apple      59565
Samsung    40839
Huawei     34423
Name: Handset Manufacturer, dtype: int64
```

In [21]:
```
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().hea
    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)    9419
Apple iPhone 6 (A1586)     9023
Apple iPhone 7 (A1778)     6326
Apple iPhone Se (A1723)    5187
Apple iPhone 8 (A1905)     4993
Name: Handset Type, dtype: int64
Manufacturer: {Manufacturer}
Samsung Galaxy S8 (Sm-G950F)    4520
Samsung Galaxy A5 Sm-A520F      3724
Samsung Galaxy J5 (Sm-J530)     3696
Samsung Galaxy J3 (Sm-J330)     3484
Samsung Galaxy S7 (Sm-G930X)    3199
Name: Handset Type, dtype: int64
Manufacturer: {Manufacturer}
Huawei B528S-23A                  19752
Huawei E5180                       2079
Huawei P20 Lite Huawei Nova 3E     2021
Huawei P20                         1480
Huawei Y6 2018                      997
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

In [22]:
```
categorical = [var for var in data.columns if data[var].dtypes=='O']
numerical = [var for var in data.columns if data[var].dtypes!='O']
```

In [23]:
```
categorical
```

Out[23]:
```
['Start', 'End', 'Last Location Name', 'Handset Manufacturer', 'Handset Type']
```

In [24]:
```
1 numerical
```

Out[24]:
```
['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)',
 'TCP DL Retrans. Vol (Bytes)',
 'TCP UL Retrans. Vol (Bytes)',
 '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 (%)',
 'HTTP DL (Bytes)',
 'HTTP UL (Bytes)',
 'Activity Duration DL (ms)',
 'Activity Duration UL (ms)',
 'Dur. (ms).1',
 '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',
 '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)']
```

In [25]:
```
1 data.columns
```

Out[25]:
```
Index(['Bearer Id', 'Start', 'Start ms', 'End', 'End ms', 'Dur. (ms)', 'IMS
I',
       'MSISDN/Number', 'IMEI', 'Last Location Name', 'Avg RTT DL (ms)',
       'Avg RTT UL (ms)', 'Avg Bearer TP DL (kbps)', 'Avg Bearer TP UL (kbp
s)',
       'TCP DL Retrans. Vol (Bytes)', 'TCP UL Retrans. Vol (Bytes)',
       '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 (%)',
       'HTTP DL (Bytes)', 'HTTP UL (Bytes)', 'Activity Duration DL (ms)',
       'Activity Duration UL (ms)', 'Dur. (ms).1', 'Handset Manufacturer',
       'Handset Type', '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',
       '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')
```

In [26]:

```
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)
TCP DL Retrans. Vol (Bytes)
TCP UL Retrans. Vol (Bytes)
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 (%)
HTTP DL (Bytes)
HTTP UL (Bytes)
Activity Duration DL (ms)
Activity Duration UL (ms)
Dur. (ms).1
Handset Manufacturer
Handset Type
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
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)

In [27]:

```
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	21489.0	390.0	
2	6.917538e+18	27786.0	401941.0	
3	6.917538e+18	15635.0	73347.0	
4	6.917538e+18	24264.0	117340.0	
...	...	...	...	
134703	1.318654e+19	80024.0	2512362.0	
134704	1.318654e+19	145291.0	2067.0	
134705	1.318654e+19	86399.0	3968131.0	
134706	1.318654e+19	86399.0	1689999.0	
134707	1.318654e+19	103113.0	0.0	

	Activity Duration UL (ms)	Social Media DL (Bytes)	\
0	101470.0	2404741.0	
1	1459.0	2478607.0	
2	399092.0	944612.0	
3	81378.0	1817239.0	
4	347852.0	1867318.0	
...	...	...	
134703	2437668.0	3240226.0	
134704	45217.0	3062671.0	
134705	3537154.0	720996.0	
134706	1513764.0	2492460.0	
134707	30367.0	1314234.0	

	Social Media UL (Bytes)	Google DL (Bytes)	Google UL (Bytes)	\
0	2410.0	5791591.0	2871336.0	
1	11936.0	3605446.0	2825198.0	
2	2827.0	10373157.0	56392.0	
3	19827.0	269988.0	3696393.0	
4	18928.0	1689296.0	195216.0	
...	...	...	...	
134703	38284.0	2036152.0	2271168.0	
134704	48953.0	9363661.0	4001970.0	
134705	42836.0	1541915.0	2100839.0	
134706	39905.0	11318188.0	466218.0	
134707	27938.0	6969652.0	3756009.0	

	Email DL (Bytes)	Email UL (Bytes)	Youtube DL (Bytes)	\
0	782388.0	806920.0	6139644.0	
1	446376.0	525108.0	10281221.0	
2	128003.0	34038.0	5385159.0	
3	3191192.0	896670.0	12347020.0	
4	740633.0	590043.0	15231815.0	
...	...	...	...	
134703	2410615.0	387548.0	12404964.0	
134704	2192057.0	866373.0	22147919.0	
134705	2315638.0	839789.0	11879062.0	
134706	2612190.0	618629.0	22163800.0	
134707	3317462.0	408257.0	12099319.0	

	Youtube UL (Bytes)	Netflix DL (Bytes)	Netflix UL (Bytes)	\
0	2071526.0	19494278.0	14668354.0	
1	18119976.0	19455048.0	10631652.0	
2	4295851.0	15755839.0	1300571.0	
3	11089528.0	2859358.0	1738176.0	

```
4                8401567.0          21563985.0           2817981.0
...                    ...                 ...                 ...
134703           4343114.0          11108134.0          21649273.0
134704           2152449.0          21468525.0           8603105.0
134705           1290963.0          22596930.0          11943452.0
134706          16763435.0           9522397.0           8346624.0
134707           1636122.0           9992219.0          17624886.0

        Gaming DL (Bytes)  Gaming UL (Bytes)  Other DL (Bytes)  \
0             466109357.0          5333340.0       670751043.0
1             673282567.0          2670856.0       501608458.0
2             821879090.0          8521398.0       472846860.0
3             805301713.0         16257481.0        24303797.0
4             583864716.0          6992868.0       685122214.0
...                   ...                ...               ...
134703        114093049.0          2834548.0       695881178.0
134704        328766801.0          7569327.0       371261255.0
134705        833634251.0         10607174.0       697260277.0
134706        338246033.0          1845068.0        17385489.0
134707        293519955.0         16295588.0       440290470.0

        Other UL (Bytes)
0              15950724.0
1               3908870.0
2               1337849.0
3              15907613.0
4               2966860.0
...                   ...
134703          3888729.0
134704         12947410.0
134705          9094407.0
134706         12797797.0
134707          6398758.0

[134708 rows x 18 columns]
```

In [28]:

```
1 data=pd.DataFrame(user_behaviour)
```

In [29]:
```
1 data
```

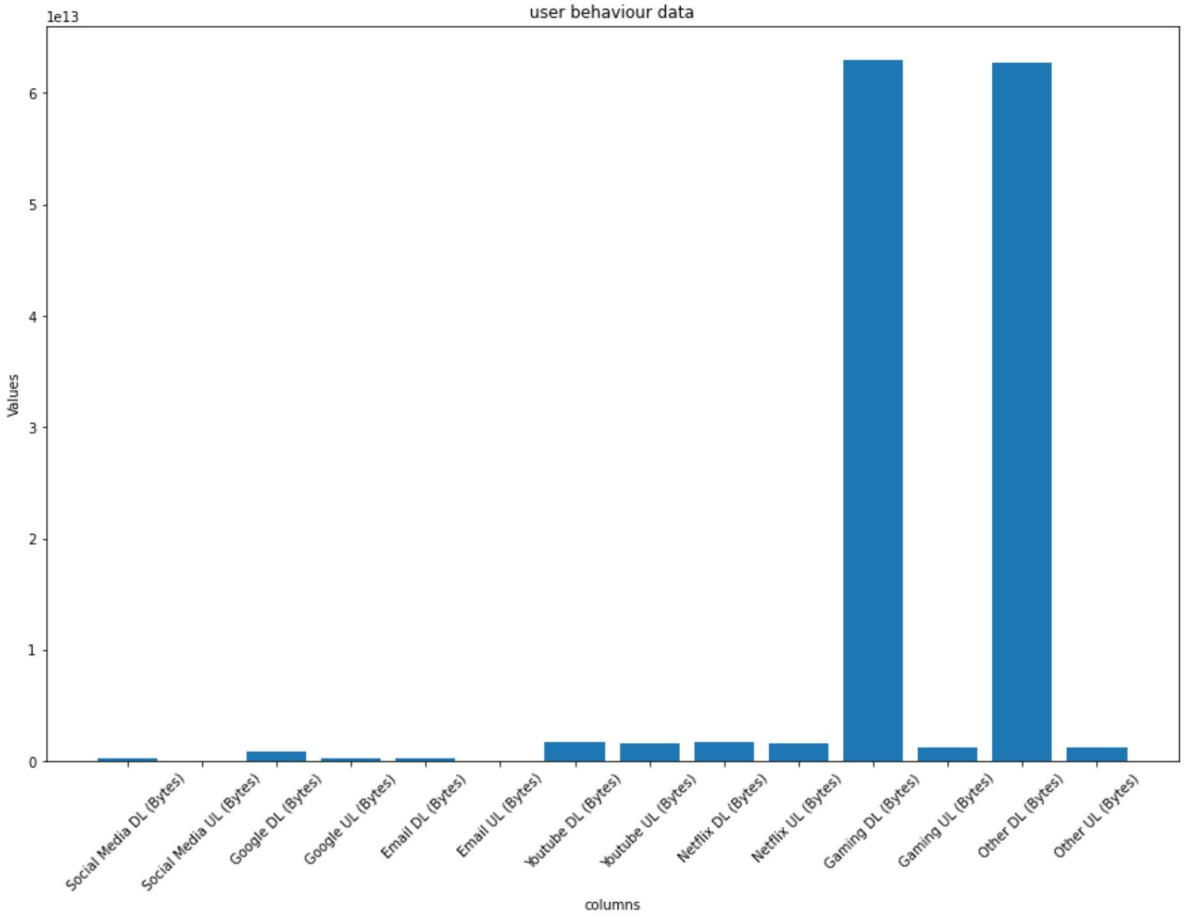
Out[29]:

	Bearer Id	Dur. (ms)	Activity Duration DL (ms)	Activity Duration UL (ms)	Social Media DL (Bytes)	Social Media UL (Bytes)	Google DL (Bytes)	Googl UI (Bytes
0	6.917538e+18	24534.0	131798.0	101470.0	2404741.0	2410.0	5791591.0	2871336.(
1	6.917538e+18	21489.0	390.0	1459.0	2478607.0	11936.0	3605446.0	2825198.(
2	6.917538e+18	27786.0	401941.0	399092.0	944612.0	2827.0	10373157.0	56392.(
3	6.917538e+18	15635.0	73347.0	81378.0	1817239.0	19827.0	269988.0	3696393.(
4	6.917538e+18	24264.0	117340.0	347852.0	1867318.0	18928.0	1689296.0	195216.(
...	...	...	...	...	...	...	...	..
134703	1.318654e+19	80024.0	2512362.0	2437668.0	3240226.0	38284.0	2036152.0	2271168.(
134704	1.318654e+19	145291.0	2067.0	45217.0	3062671.0	48953.0	9363661.0	4001970.(
134705	1.318654e+19	86399.0	3968131.0	3537154.0	720996.0	42836.0	1541915.0	2100839.(
134706	1.318654e+19	86399.0	1689999.0	1513764.0	2492460.0	39905.0	11318188.0	466218.(
134707	1.318654e+19	103113.0	0.0	30367.0	1314234.0	27938.0	6969652.0	3756009.(

134708 rows × 18 columns

In [30]:

```
# Plot the bar chart using seaborn
plt.figure(figsize=(15,10))
Bar_columns = [ 'Social Media DL (Bytes)','Social Media UL (Bytes)','Goog
                'Email DL (Bytes)','Email UL (Bytes)','Youtube DL (Bytes)'
                'Netflix UL (Bytes)','Gaming DL (Bytes)','Gaming 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()
```



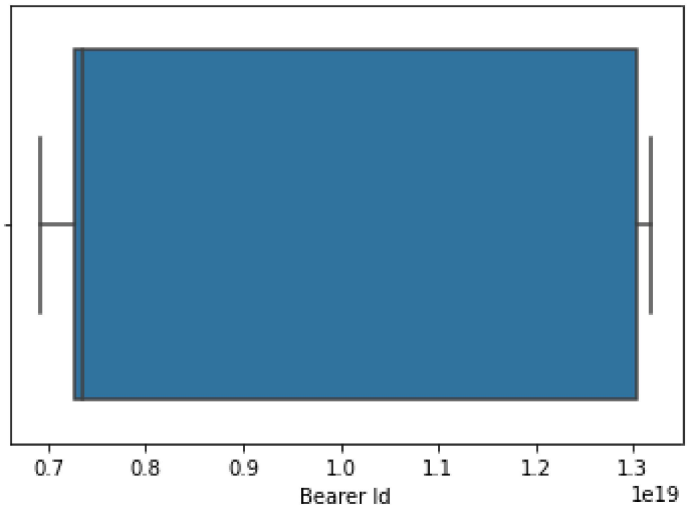
In [31]:

```
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)
```

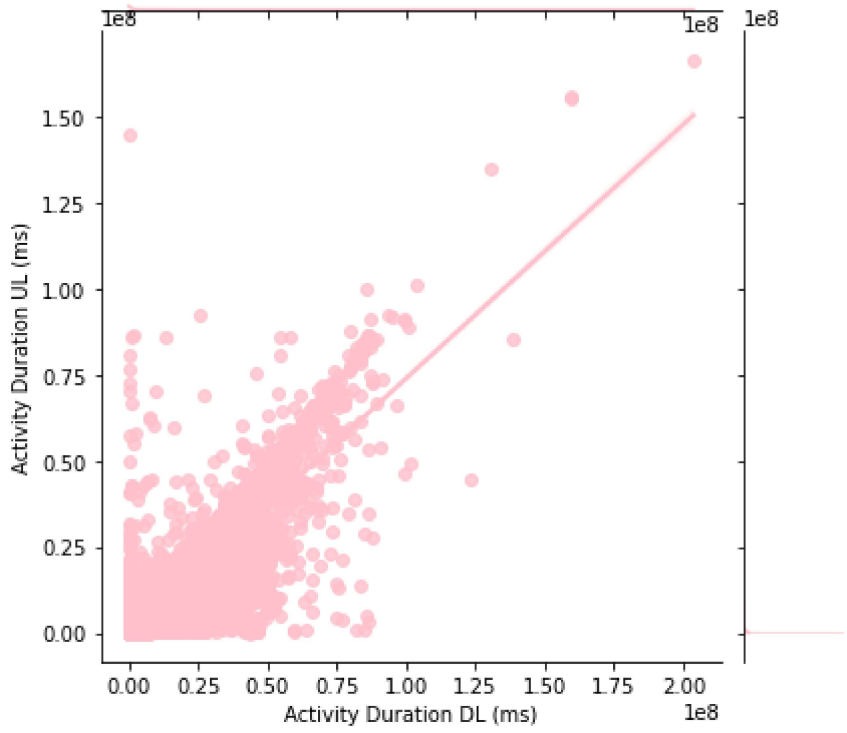
In [32]:

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



In [33]:
```
plt.figure(figsize=(15,5))
sns.jointplot(data['Activity Duration DL (ms)'],
              data['Activity Duration UL (ms)'],color='pink',kind="reg")
plt.xlabel('Activity Duration DL (ms)')
plt.ylabel('Activity Duration UL (ms)')
plt.show()
```

```
<Figure size 1080x360 with 0 Axes>
```



In [34]:
```
user_behaviour.skew()
```

Out[34]:
```
Bearer Id                     0.182443
Dur. (ms)                     7.402203
Activity Duration DL (ms)     6.159777
Activity Duration UL (ms)     7.538985
Social Media DL (Bytes)       1.231431
Social Media UL (Bytes)       1.180260
Google DL (Bytes)             1.211919
Google UL (Bytes)             1.248976
Email DL (Bytes)              1.224819
Email UL (Bytes)              1.248869
Youtube DL (Bytes)            1.223591
Youtube UL (Bytes)            1.245732
Netflix DL (Bytes)            1.145590
Netflix UL (Bytes)            1.272487
Gaming DL (Bytes)             1.163896
Gaming UL (Bytes)             1.153490
Other DL (Bytes)              1.270813
Other UL (Bytes)              1.222591
dtype: float64
```