User Analytics in Telecommunication Industry

Abubakar Alaro

Statistical Summary of xDR Session, Session duration, total download and upload

Measures	Xdr Session	Session Duration (ms)	Total DL & UL
Minimum	1.00	7,142.00	28,956,107.00
Maximum	1083.00	81,238,323.00	955,984,776.00
Average	9.66	798,652.26	495,764,636.37
1st Quartile	1.00	86,310.75	284,475,924.00
2nd Quartile	1.00	139,854.50	496,861,101.50
3rd Quartile	2.00	253,312.00	706,505,882.25

Table 1.0

Statistical Summary for all Applications

Measures	Email Session	Gaming Session	Social Media Session	YouTube Session	Netflix Session	Google Session
Minimum	8,359.00	306,358.00	1,563.00	78,903.00	98,432.00	40,330.00
Maximum	4,518,036.00	859,202,784.00	3,650,861.00	45,190,078.00	45,198,153.00	15,528,782.00
Average	2,259,102.31	430,333,100.71	1,828,250.21	22,643,482.64	22,628,606.54	7,807,294.55
1st Quartile	1,359,343.00	218,727,371.00	932,213.25	15,998,460.50	15,979,437.25	4,943,592.00
2nd Quartile	2,263,574.00	431,615,155.50	1,826,442.00	22,661,814.00	22,635,588.50	7,812,852.50
3rd Quartile	3,159,824.75	641,416,114.00	2,727,489.00	29,292,658.00	29,290,604.25	15,528,782.00

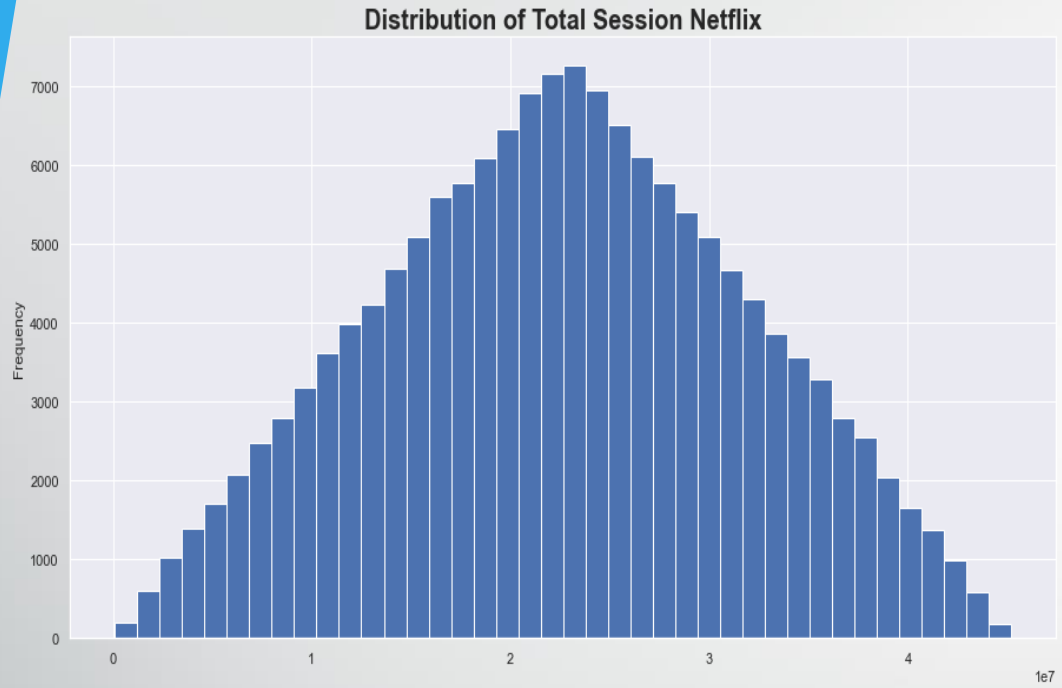
Applications used by customers and their statistical summary

Table 2.0

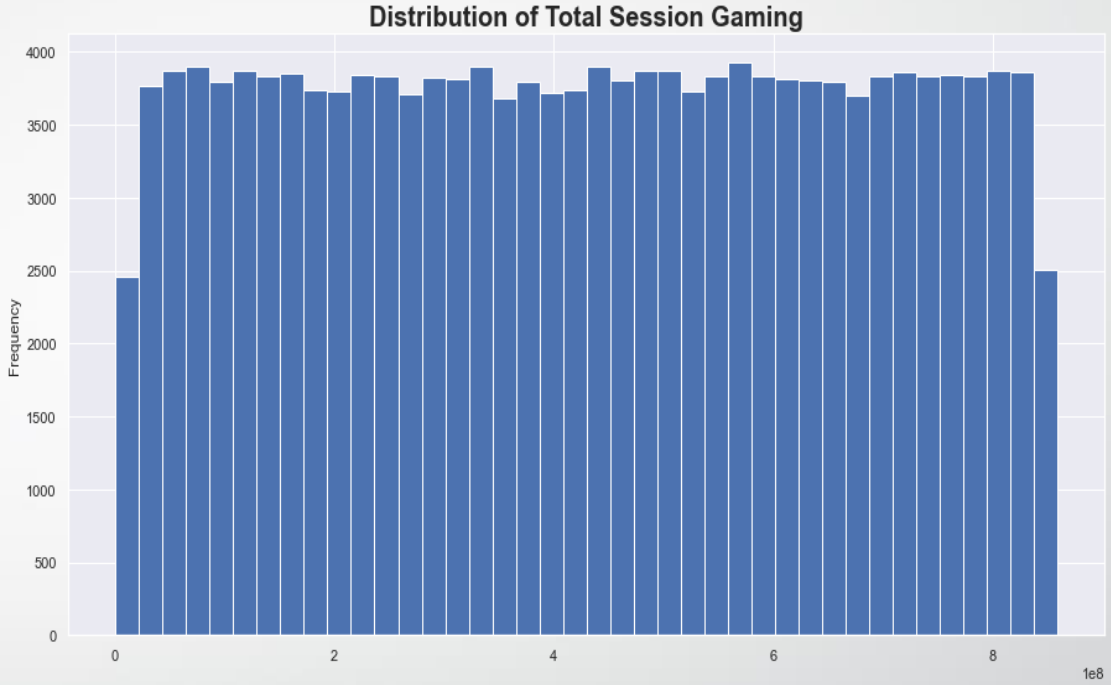
INTERPRETATION

- From table 1.0, the highest data session by a single customer was 1083 and on a average customers will have at least 9 data sessions. This illustrates that data consumption rate by customers is high and if purchase it could lead to an increase in profit.
- The least session duration by a customer was 7,142 MS while the maximum was 81,238,323 Ms. On average, customers spends 798,652.26 Ms. The average download and upload was 495,764,636.37 bytes which is equivalent to 495.765 Megabytes. This means that for each session on average, customers spends a total of 495.765Mb per session. About 25% of all customers consume more than 706.501Mb per session and only 25% consume below 284.475Mb. From this, given the number of customers, data volume consumption is high and only a small fraction of customer consume less than 200 Mb.
- From table 2.0, Gaming Application Session has the highest data volume consumption with the maximum being 859,202,784.00 bytes and it minimum 306,358.00 bytes. Netflix and YouTube as similar statistical values as the difference between both averages was less than 20,000 bytes which is 0.02 Mb. This is no surprise as the 2 applications are both used for streaming videos.

Graphical Distribution of Application Session

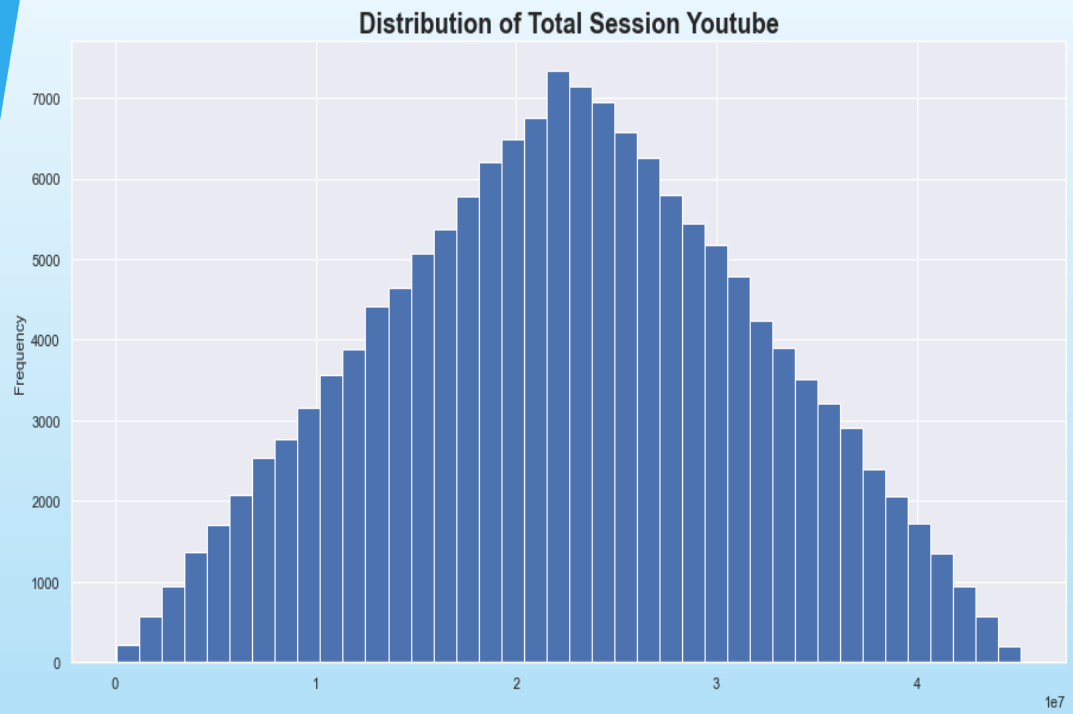


Bell Shaped distribution of Netflix Session

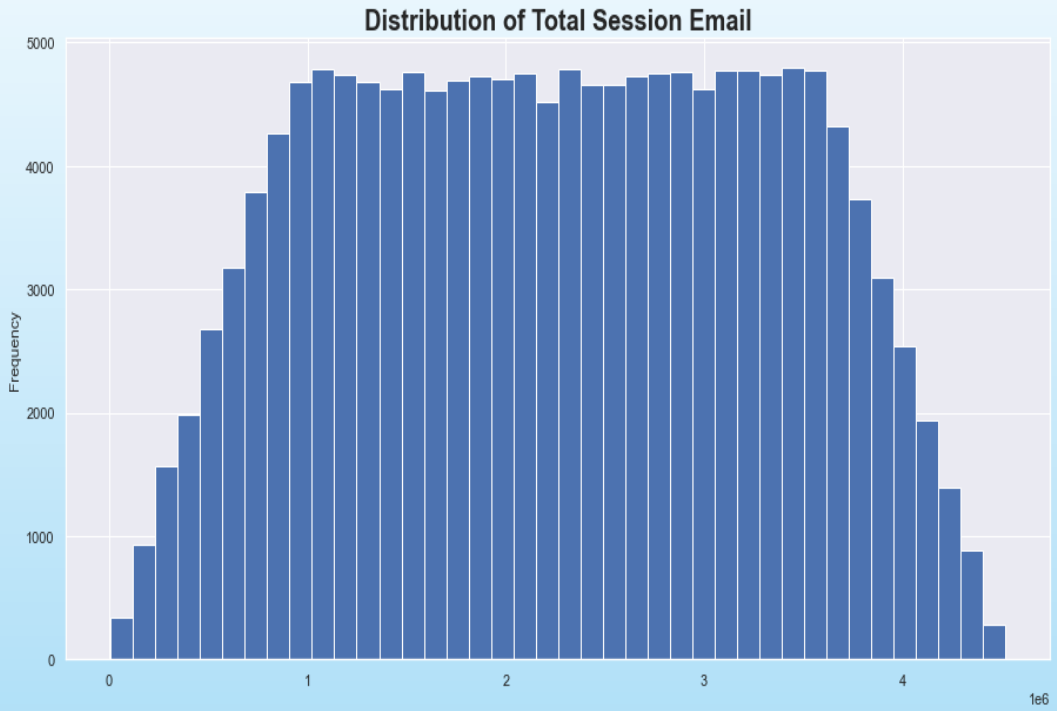


Normal distribution of Gaming Session

Graphical Distribution of Application Session

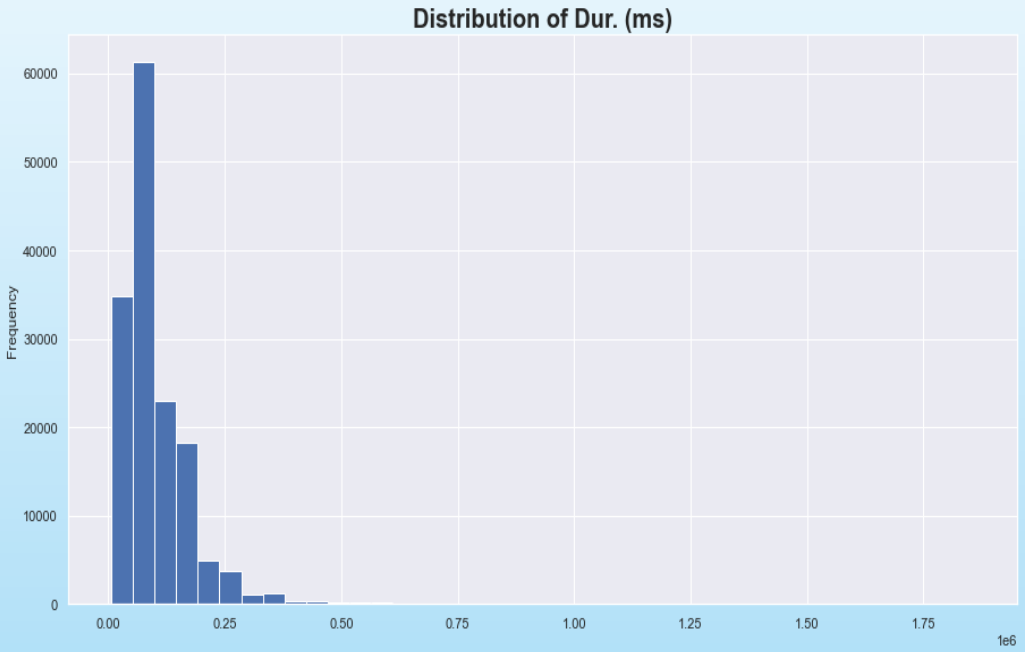


Bell Shaped chart of YouTube Session – Data values are normally distributed

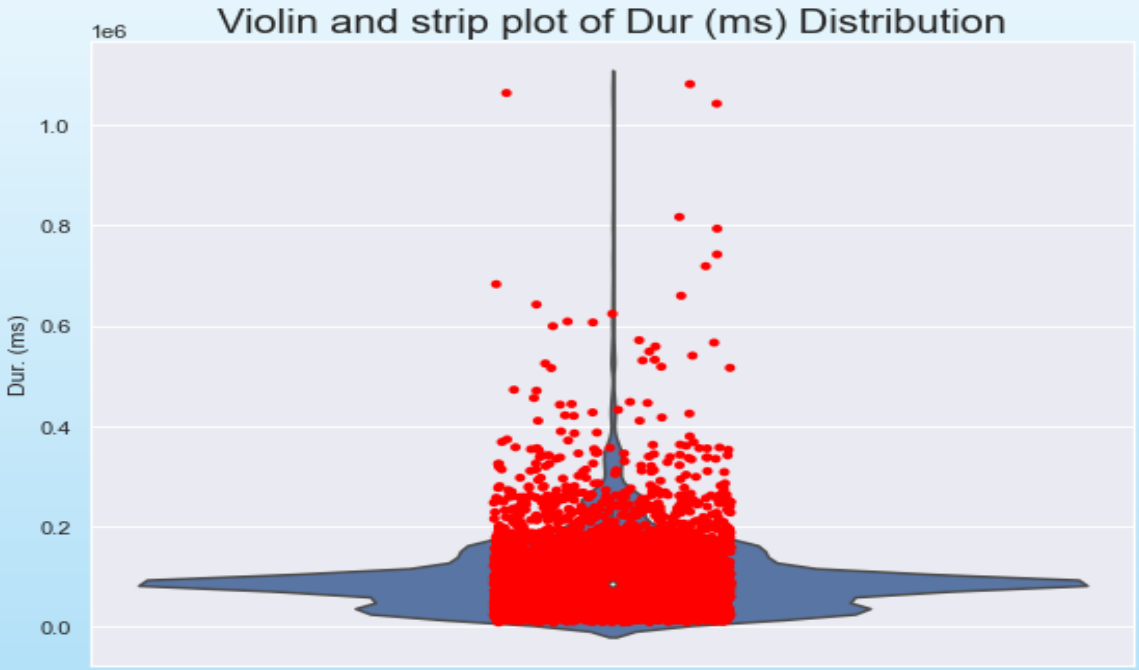


Histogram chart of Email Session – Data values are normally distributed

Graphical Distribution of Application Session

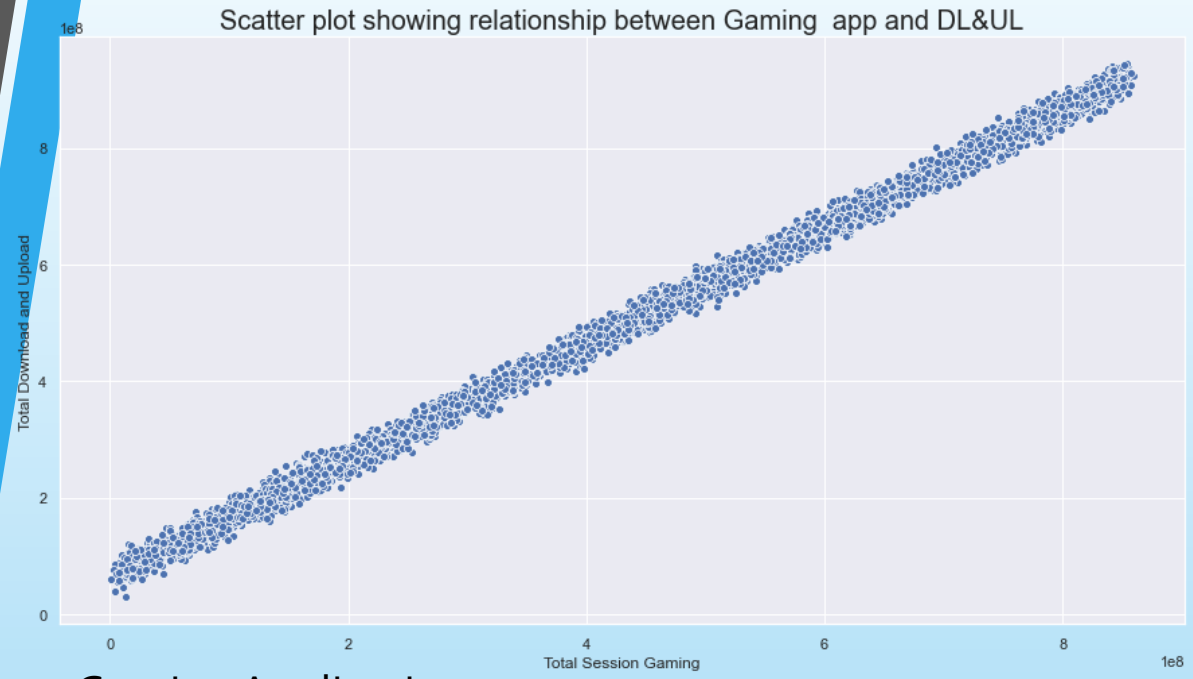


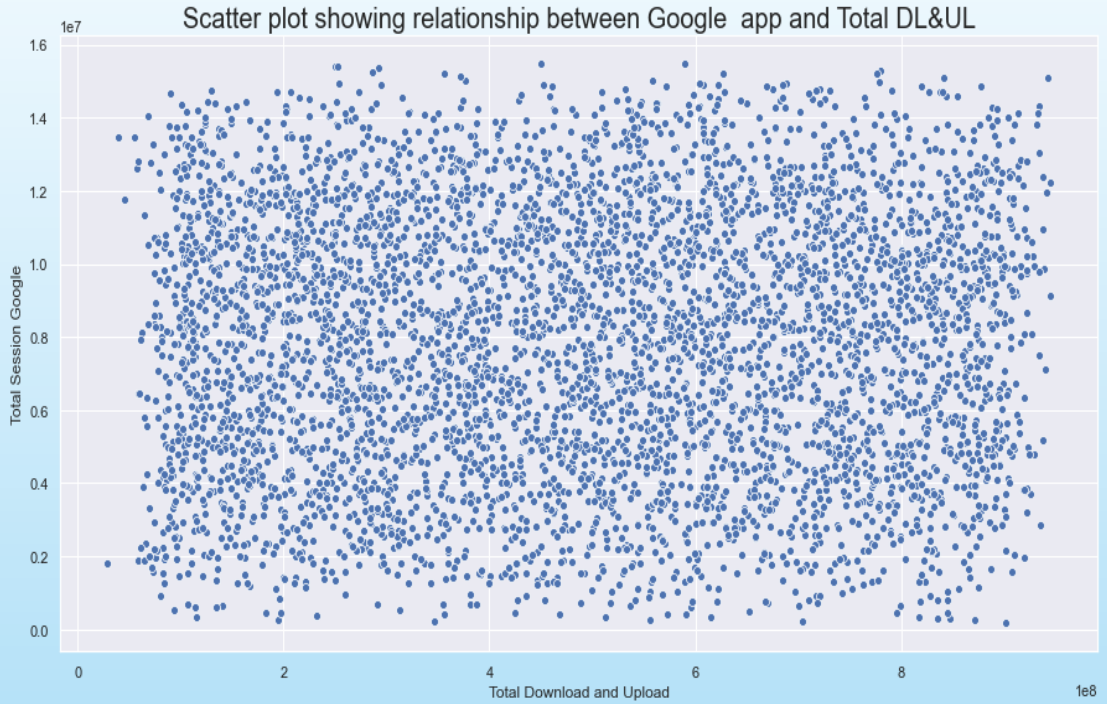
Left skewed distribution of Dur. (ms) – Data are not normally distributed



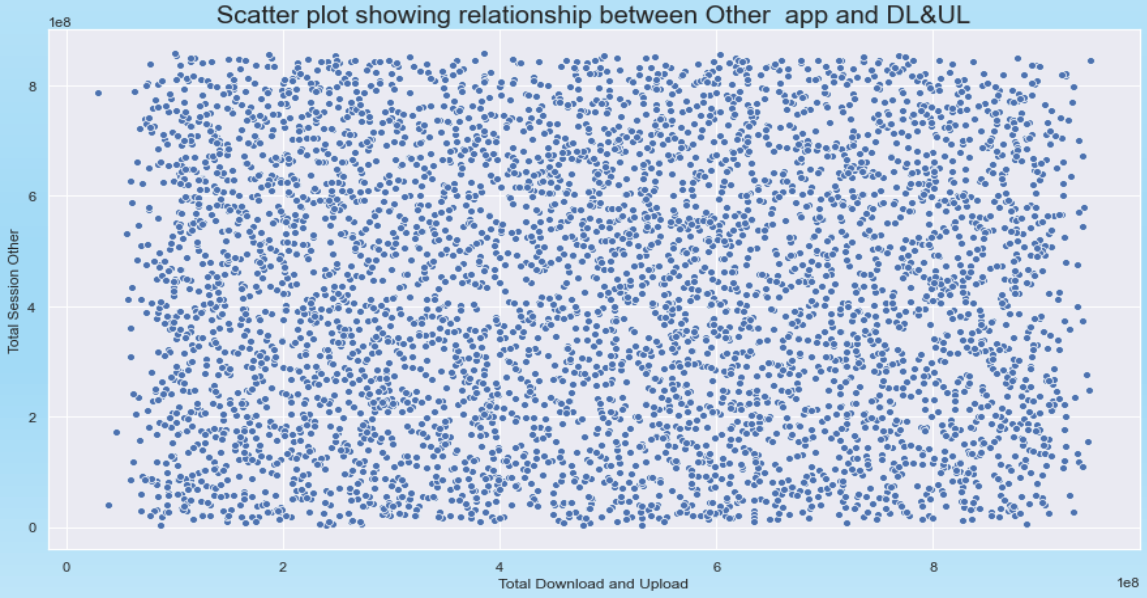
Majority of the data point are small values

Relationship between total data consumption and applications

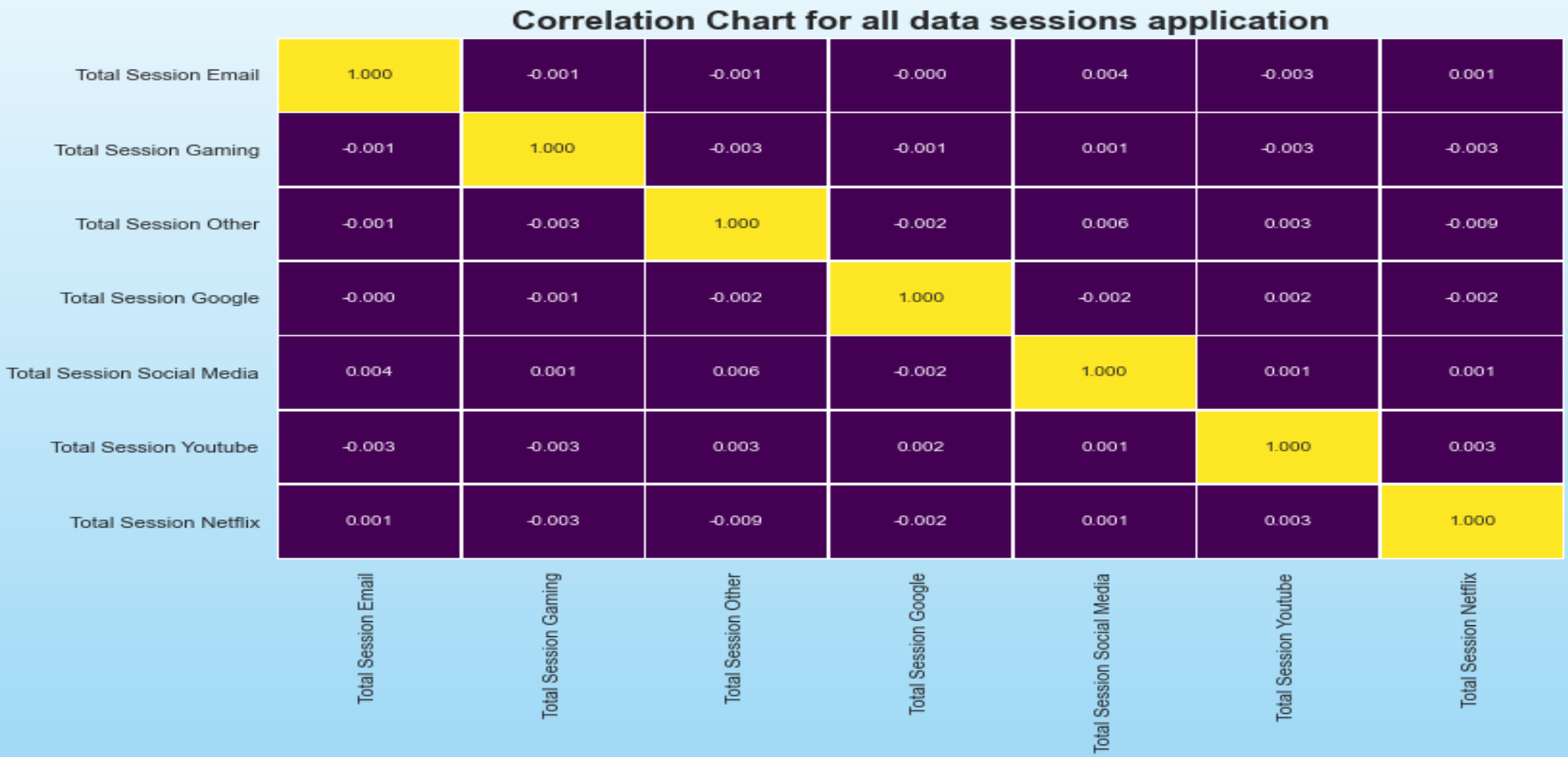




Gaming Application Session is positively correlated with the total DL & UL



Correlation Matrices



No Application has a higher correlation between them.

User Engagement Analysis

	MSISDN/Number	xDR Session
13526	3.362632e+10	1172889
13180	3.362578e+10	289
6437	3.361489e+10	289
37052	3.365973e+10	256
92923	3.376054e+10	225
76363	3.367588e+10	225
65118	3.366716e+10	169
13994	3.362708e+10	144
1279	3.360452e+10	144
92577	3.376041e+10	144

	MSISDN/Number	Total Download and Upload
13526	3.362632e+10	5.167961e+11
92923	3.376054e+10	8.514774e+09
76363	3.367588e+10	7.891111e+09
13180	3.362578e+10	7.708153e+09
6437	3.361489e+10	7.601873e+09
37052	3.365973e+10	6.932144e+09
63028	3.366646e+10	6.530860e+09
57241	3.366471e+10	6.472725e+09
86455	3.369879e+10	6.149693e+09
30715	3.365836e+10	6.051459e+09

	MSISDN/Number	session duration
13526	3.362632e+10	8.798110e+10
13180	3.362578e+10	3.154138e+08
6437	3.361489e+10	1.694373e+08
92923	3.376054e+10	1.391915e+08
65118	3.366716e+10	1.136839e+08
666	3.360313e+10	7.545313e+07
76363	3.367588e+10	7.298920e+07
37052	3.365973e+10	6.456685e+07
92577	3.376041e+10	6.386000e+07
1279	3.360452e+10	6.249588e+07

Top 10 Customer Per Engagement Metrics

Statistics for customer clusters Engagement metrics

Minimum Aggregate per Cluster

Segment K-means	Total Download and Upload	xDR Session	session duration
0	457354533.0	1	7142.0
1	28956107.0	1	7189.0
2	64615748.0	1083	81238323.0

Maximum Aggregate per Cluster

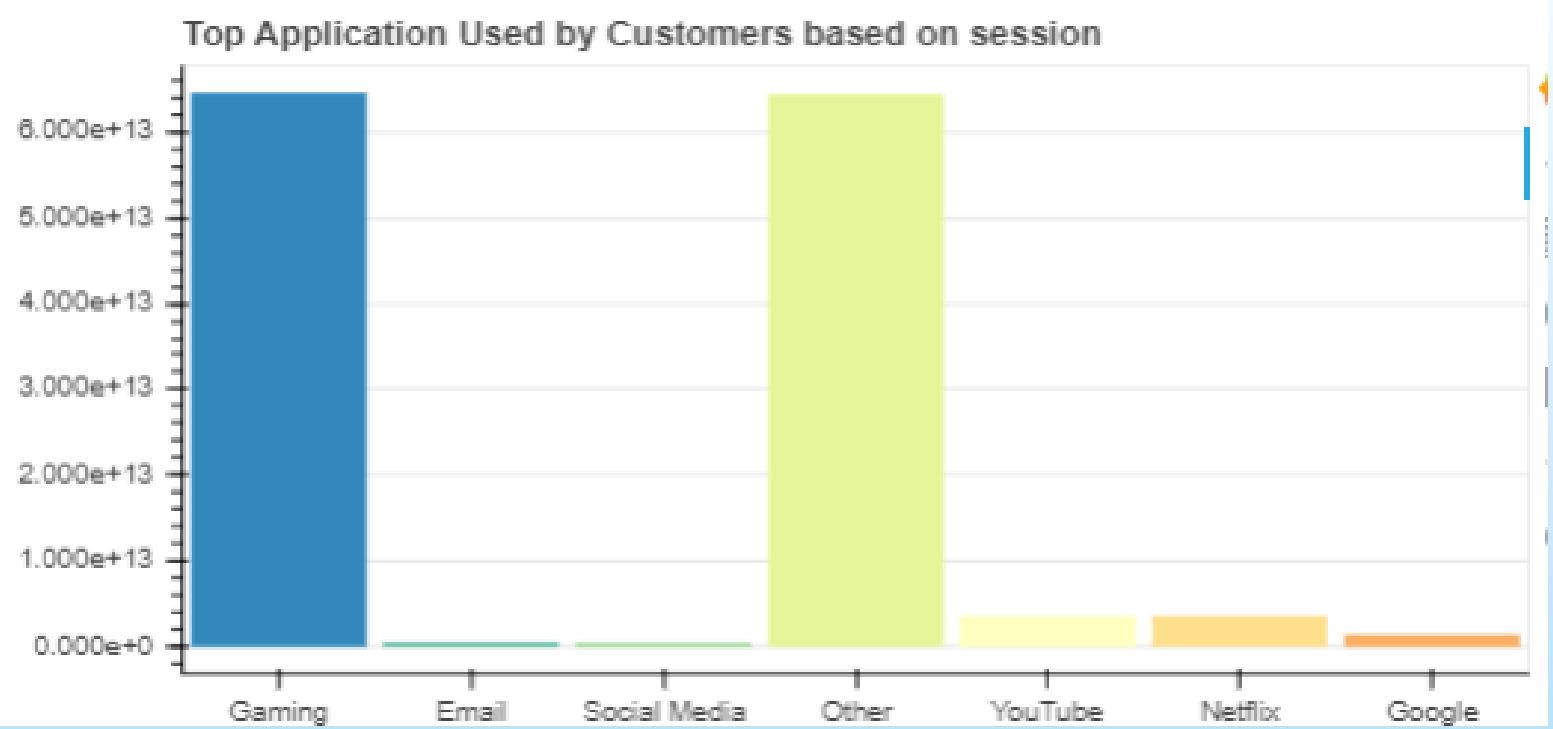
Segment K-means	Total Download and Upload	xDR Session	session duration
0	875313739.0	17	18553754.0
1	457348129.0	17	18553754.0
2	873407674.0	1083	81238323.0

Mean Aggregate per Cluster

Segment K-means	Total Download and Upload	xDR Session	session duration
0	6.508386e+08	1.854172	2.136989e+05
1	2.643515e+08	1.862516	2.136001e+05
2	4.771894e+08	1083.000000	8.123832e+07

Sum Aggregate per Cluster

Segment K-means	Total Download and Upload	xDR Session	session duration
0	5.298542e+13	150950	1.739744e+10
1	1.784532e+13	125731	1.441929e+10
2	5.167961e+11	1172889	8.798110e+10

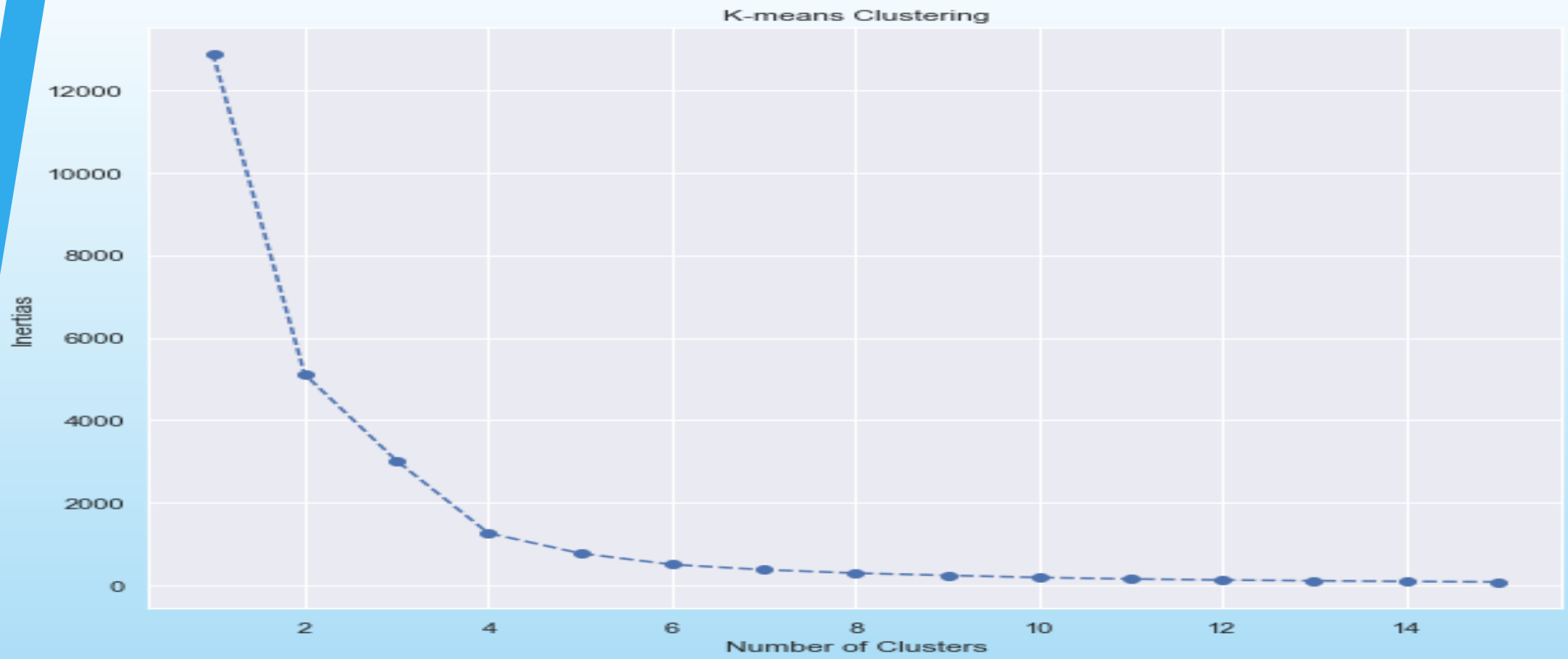


	MSISDN/Number	Average Other Session per Customer
13210	3.362583e+10	859520934.0
10491	3.362142e+10	859285331.0
48681	3.366247e+10	859053653.0
106510	3.378784e+10	858926613.0
53473	3.366366e+10	858787718.0
36291	3.365955e+10	858555595.0
47198	3.366213e+10	858495391.0
102309	3.376981e+10	858469987.0
68414	3.366822e+10	858467471.0
78515	3.368022e+10	858427905.0

	MSISDN/Number	Average Netflix Session per Customer
9082	3.361921e+10	45188499.0
51459	3.366314e+10	45054355.0
17668	3.363433e+10	45010959.0
82620	3.368845e+10	44983646.0
11379	3.362286e+10	44935983.0
59360	3.366527e+10	44930670.0
86101	3.369869e+10	44915543.0
104643	3.378231e+10	44881803.0
35174	3.365930e+10	44853893.0
51214	3.366308e+10	44843988.0

	MSISDN/Number	Average Gaming Session per Customer
26537	3.365084e+10	859202784.0
102592	3.377040e+10	859010933.0
16346	3.363145e+10	858735183.0
25671	3.365064e+10	858652760.0
70186	3.366881e+10	858643269.0
10405	3.362127e+10	858562729.0
39790	3.366037e+10	858506410.0
37624	3.365986e+10	858201303.0
22549	3.364925e+10	858143650.0
103489	3.377864e+10	858023375.0

Top 3 Most Used Application based on Average Session



1. Optimal cluster is 4
2. On Average, Cluster 1 has the highest number of Total Download and Upload but low session duration
3. Cluster 0 and 2 has the least number of Download and Upload
4. Cluster 3 are special customers, they have the highest number of session duration and high values for Downloads and Upload

User Experience Analysis

Top 10 Values for Through Put , TCP Retransmission and Round Trip Time

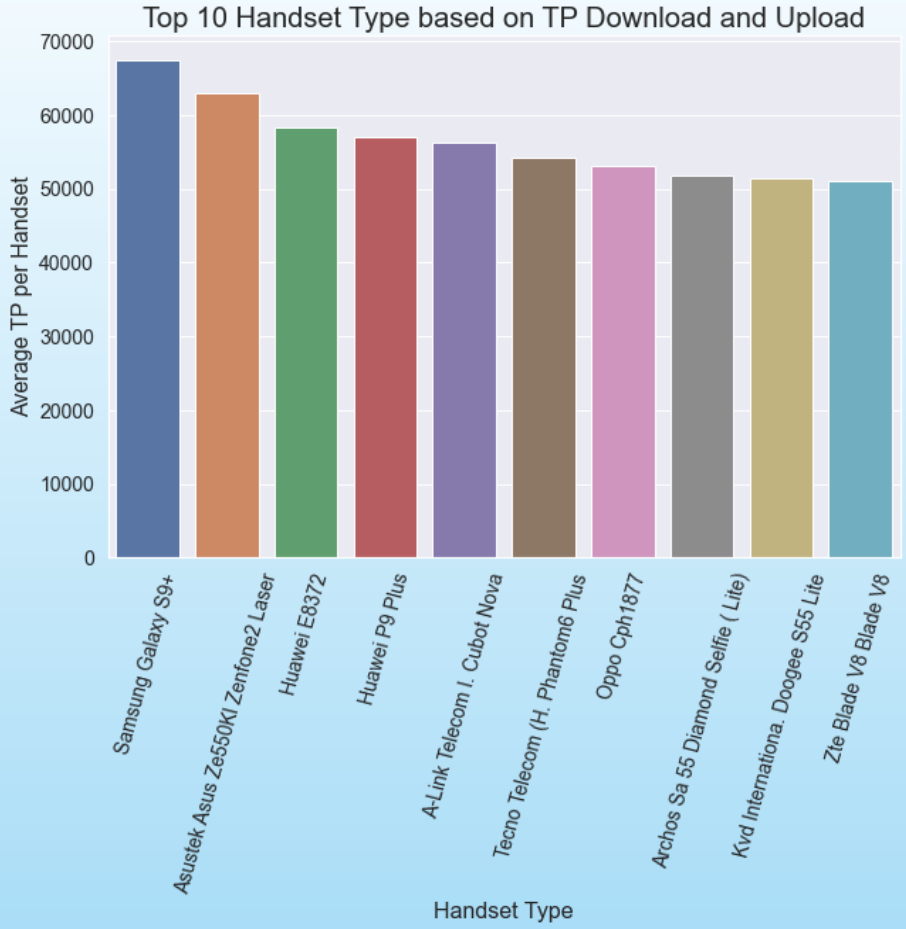
	MSISDN/Number	Average Throughput
89296	3.369967e+10	0.072416
94398	3.376118e+10	0.071794
58043	3.366491e+10	0.071766
79451	3.368207e+10	0.071635
21722	3.364688e+10	0.071537
71240	3.366916e+10	0.071409
86837	3.369891e+10	0.071130
24228	3.365031e+10	0.070999
16793	3.363254e+10	0.070450
36115	3.365951e+10	0.070344

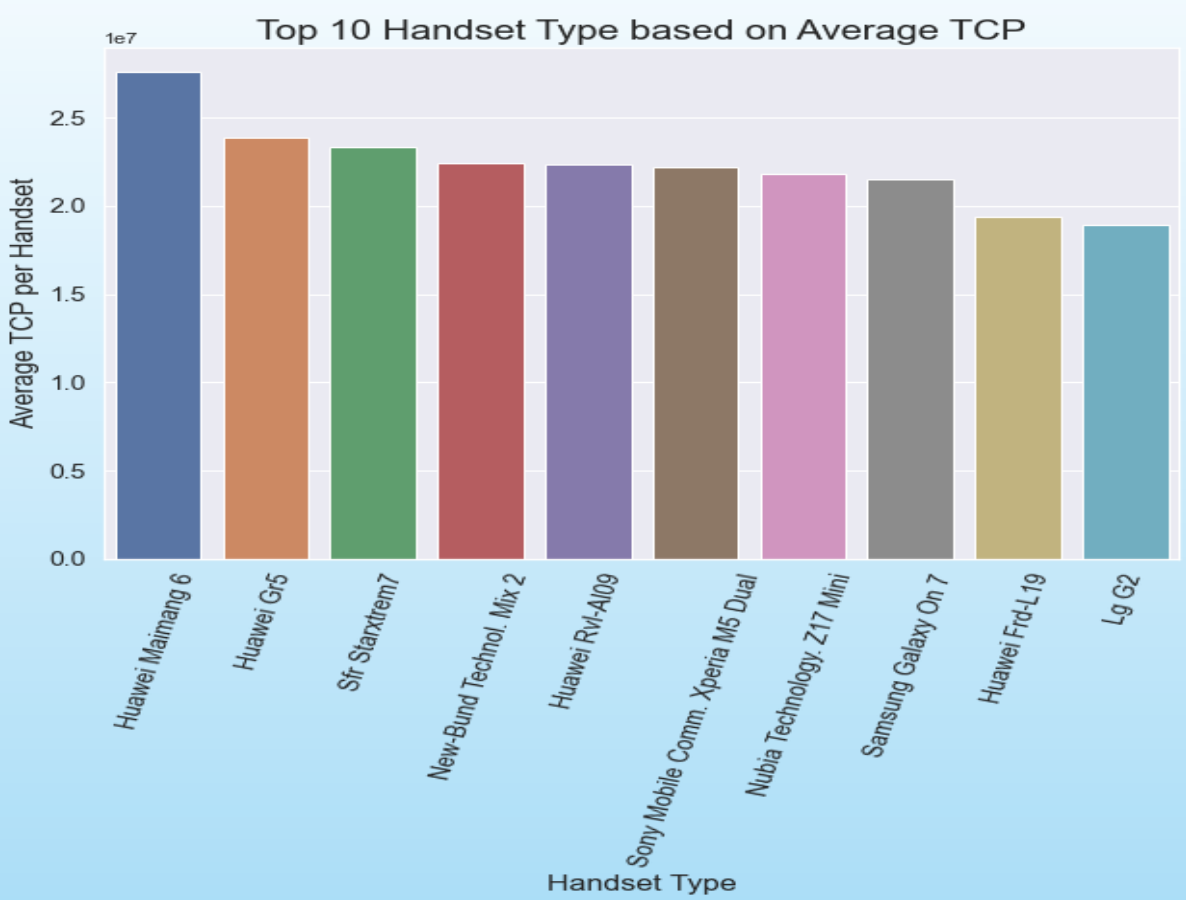
	MSISDN/Number	Average TCP Retransmission
23511	3.365014e+10	31.663032
106738	3.378929e+10	31.658308
49021	3.366255e+10	31.657498
46432	3.366196e+10	31.642448
35093	3.365929e+10	31.641094
57531	3.366479e+10	31.637910
71076	3.366911e+10	31.633212
86209	3.369872e+10	31.632878
34498	3.365916e+10	31.626984
57166	3.366470e+10	31.616370

	MSISDN/Number	Average RT Time
61786	3.366606e+10	318.0
82410	3.368795e+10	317.0
43442	3.366125e+10	315.0
97724	3.376243e+10	314.0
96914	3.376204e+10	312.0
40719	3.366060e+10	311.0
37520	3.365983e+10	309.0
35692	3.365942e+10	308.0
47286	3.366215e+10	306.0
12415	3.362454e+10	306.0

Top 10 Handset with the highest Average Through Put

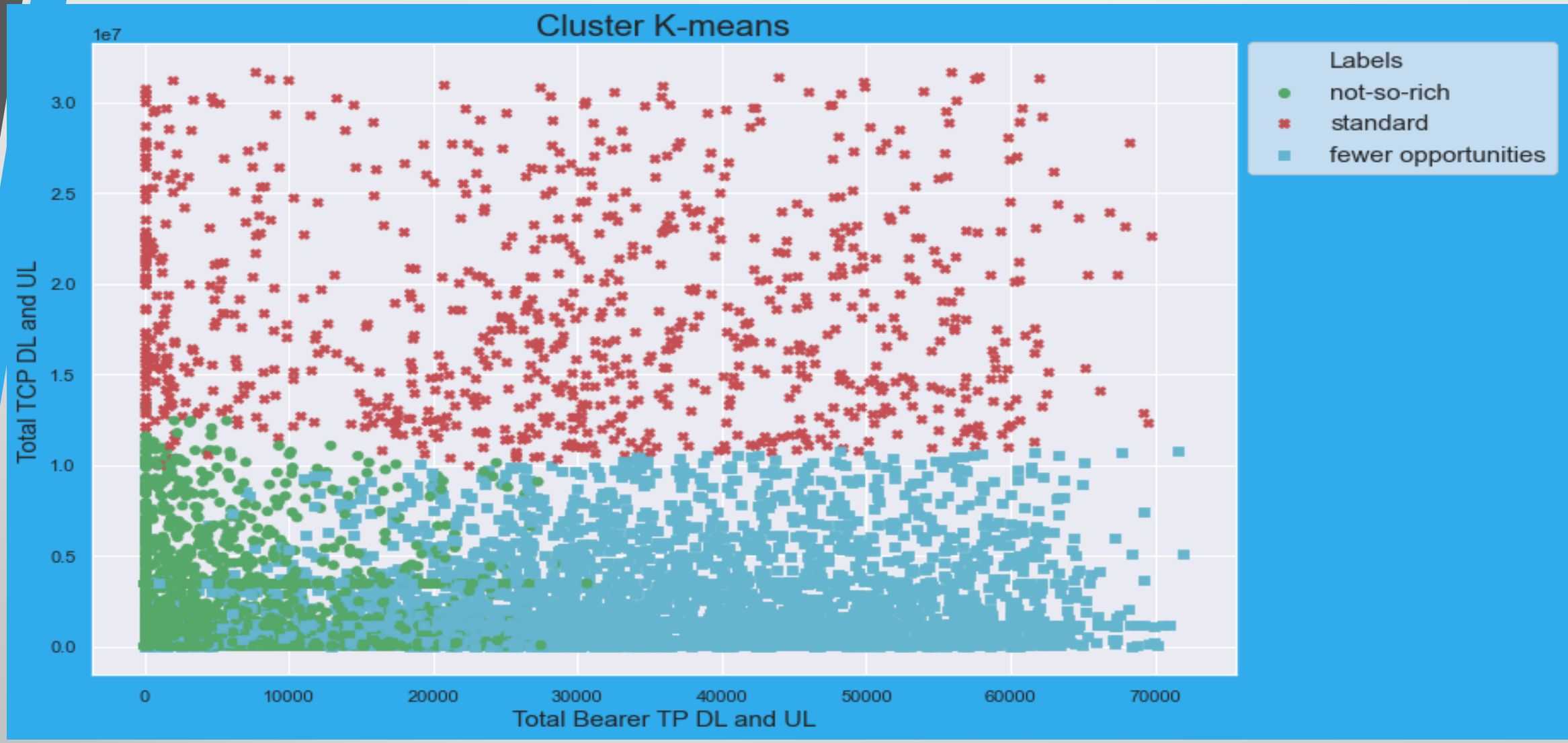
	Handset Type	Average TP per Handset
966	Samsung Galaxy S9+	67383.0
124	Asustek Asus Ze550Kl Zenfone2 Laser	62989.0
349	Huawei E8372	58190.0
423	Huawei P9 Plus	56918.0
3	A-Link Telecom I. Cubot Nova	56217.0
1219	Tecno Telecom (H. Phantom6 Plus	54197.0
770	Oppo Cph1877	53058.0
95	Archos Sa 55 Diamond Selfie (Lite)	51797.0
489	Kvd Internationa. Doogee S55 Lite	51389.0
1384	Zte Blade V8 Blade V8	50950.0





Samsung Galaxy S9+ leads the chart with close to 70000 kbps. Handsets with higher TP generally are from China that is the Manufacturers of this Handset are based or originates from China. Many different handset type from Huawei has higher value of average TP when compared with other handset, which means they have strong connections to internet

Huawei Miamong 6 top the chart of handsets with higher average retransmission. It means that the connection on these phones has bad connections to internet hence the need to always retransmit packets. This influence bad user experience

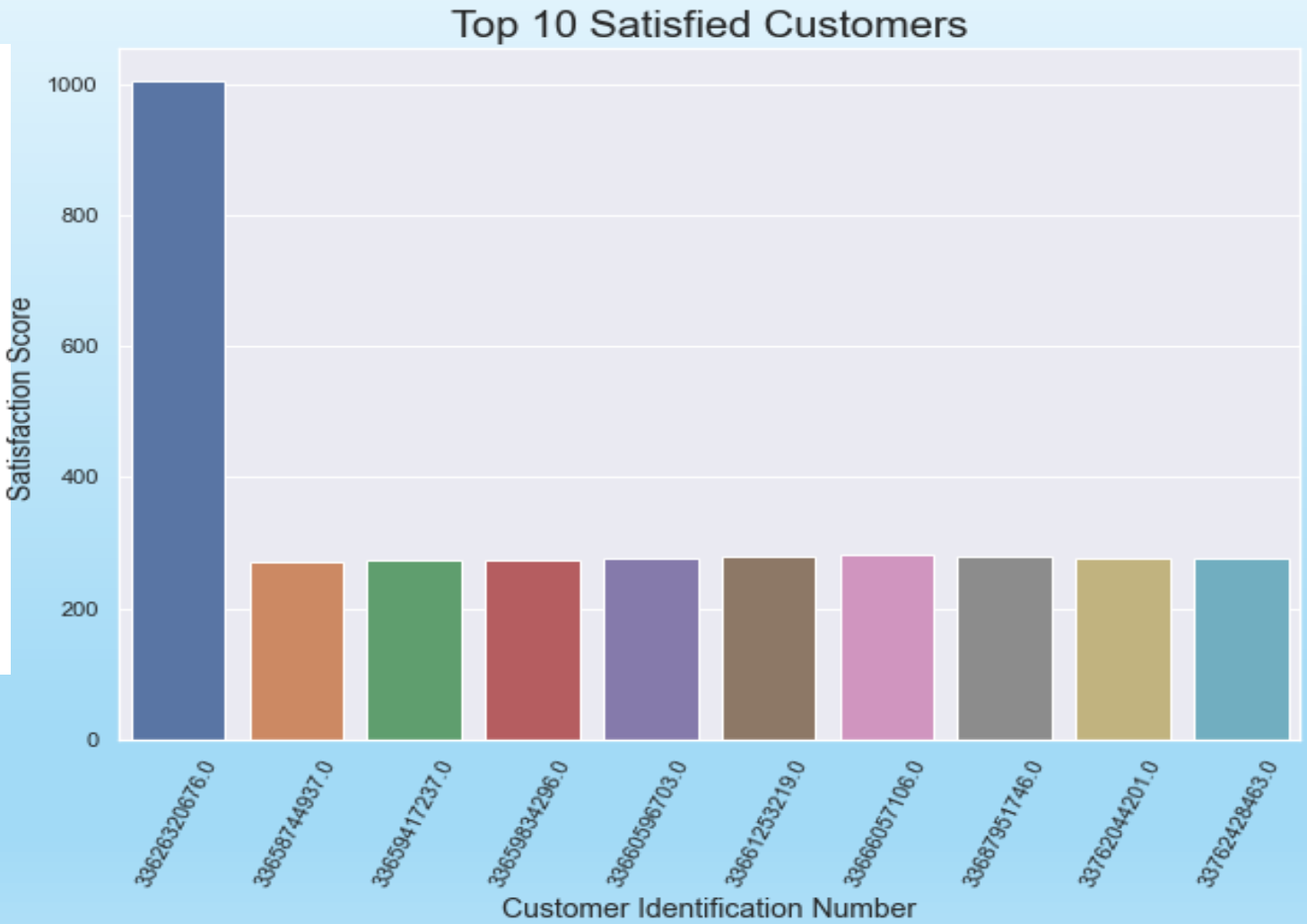


The green cluster (not so rich) sends little amount of data based on Through Put and have less retransmission when compared to the fewer opportunities cluster who sends huge amount of data and has less retransmission of packets. The red cluster (Standard) sends huge amount of data and has higher rate of retransmission packet.

Satisfaction Analysis

Top 10 Satisfied Customer and their Satisfaction Score

	MSISDN/Number	Satisfaction Score
13526	3.362632e+10	1004.982338
61786	3.366606e+10	281.085047
82410	3.368795e+10	280.219022
43442	3.366125e+10	278.486971
97724	3.376243e+10	277.620946
96914	3.376204e+10	275.888895
40719	3.366060e+10	275.022870
37520	3.365983e+10	273.290819
35692	3.365942e+10	272.424794
32478	3.365874e+10	270.692743



Cluster	Experience Score	Engagement Score
0	113.058285	10.466982
1	140.799788	1869.164888

```
# convert Dataframe to SQL
df_sql.to_sql('user_score', con, if_exists='replace', index=False)
print('Dataframe converted to sql and saved Succesfully...!!!')
```

```
c.execute("""
        SELECT * FROM user_score
        """)
# for row in c.fetchall():
#     print(row)
```

```
<sqlite3.Cursor at 0x158ee70a490>
```

```
# convert the sql back to dataframe
sql_df = pd.DataFrame(c.fetchall(), columns=df_sql.columns)
sql_df.head(6)
```

	MSISDN/Number	Engagement Score	Experience Score	Satisfaction Score
0	3.366496e+10	0.945494	0.538447	0.741970
1	3.368185e+10	0.441621	0.540743	0.491182
2	3.376063e+10	1.010053	0.543225	0.776639
3	3.375034e+10	0.434334	0.536521	0.485427
4	3.369980e+10	0.544386	0.543225	0.543806
5	3.366819e+10	0.404953	0.533225	0.469089

Sql Dataframe conversion

Recommendation

1. More focus should be given to the Gaming Application as it has more traffic than other applications

2. Collaboration with Samsung Handset Manufacturers is encouraged as it has the highest Through Put. This is deemed to increase customers thus leading to rapid marginal return for both parties.

3. More data should be generated with respect to customer's behavior and response to utilities and services of the company

Limitations

1. Data given is not enough to recommend a purchase of the company because financial details of the company was not provided
2. Data about the company competitors were not provided so we cannot grade their performance with respect to other telecommunication industries
3. More data is required to understand the satisfaction level of customers

Reference

- 1. https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.pipe.html
- 2. http://www.unitconversion.org/time/milliseconds-to-seconds-conversion.html
- 3. https://accedian.com/blog/network-packet-loss-retransmissions-and-duplicate-acknowledgements/
- 4. https://www.collinsdictionary.com/dictionary/english/put-through