TELECOM INDUSTRY OVERVIEW

### Table of Contents

* [Libraries](https://github.com/ada-k/Telecommunication---UserAnalytics--/blob/dd95e5e68f964a35c8c900101bce0cabef6bf105//" \l "Libraries)
* [Dataset](https://github.com/ada-k/Telecommunication---UserAnalytics--/blob/dd95e5e68f964a35c8c900101bce0cabef6bf105//" \l "Dataset)
* [Null Values](https://github.com/ada-k/Telecommunication---UserAnalytics--/blob/dd95e5e68f964a35c8c900101bce0cabef6bf105//" \l "Null-Values)
* [Preprocessing](https://github.com/ada-k/Telecommunication---UserAnalytics--/blob/dd95e5e68f964a35c8c900101bce0cabef6bf105//" \l "PreProcessing)
* [1. User Overiew Analysis](https://github.com/ada-k/Telecommunication---UserAnalytics--/blob/dd95e5e68f964a35c8c900101bce0cabef6bf105//" \l "1.-User-Overiew-Analysis)
  + [User Aggregates](https://github.com/ada-k/Telecommunication---UserAnalytics--/blob/dd95e5e68f964a35c8c900101bce0cabef6bf105//" \l "Aggregate-per-user-the-following-information-in-the-column:)
  + [Non-Graphical Univariate Analysis](https://github.com/ada-k/Telecommunication---UserAnalytics--/blob/dd95e5e68f964a35c8c900101bce0cabef6bf105//" \l "Non-Graphical-Univariate-Analysis)
  + [Graphical Univariate Analysis.](https://github.com/ada-k/Telecommunication---UserAnalytics--/blob/dd95e5e68f964a35c8c900101bce0cabef6bf105//" \l "Graphical-Univariate-Analysis.)
  + [Bivariate Analysis](https://github.com/ada-k/Telecommunication---UserAnalytics--/blob/dd95e5e68f964a35c8c900101bce0cabef6bf105//" \l "Bivariate-Analysis)
  + [Variable Transformation](https://github.com/ada-k/Telecommunication---UserAnalytics--/blob/dd95e5e68f964a35c8c900101bce0cabef6bf105//" \l "Variable-transformations)
  + [Correlation Analysis](https://github.com/ada-k/Telecommunication---UserAnalytics--/blob/dd95e5e68f964a35c8c900101bce0cabef6bf105//" \l "Correlation-Analysis)
  + [Dimensionality Reduction](https://github.com/ada-k/Telecommunication---UserAnalytics--/blob/dd95e5e68f964a35c8c900101bce0cabef6bf105//" \l "Dimensionality-Reduction)
* [2. User Engagement Analysis](https://github.com/ada-k/Telecommunication---UserAnalytics--/blob/dd95e5e68f964a35c8c900101bce0cabef6bf105//" \l "2.-User-Engagement-analysis)
  + [Top 10 customers per metric.](https://github.com/ada-k/Telecommunication---UserAnalytics--/blob/dd95e5e68f964a35c8c900101bce0cabef6bf105//" \l "Top-10-users-with-high-sessions-frequency)
  + [Normalization and Kmeans](https://github.com/ada-k/Telecommunication---UserAnalytics--/blob/dd95e5e68f964a35c8c900101bce0cabef6bf105//" \l "Normalize-each-engagement-metric-and-run-a-k-means-(k=3)-to-classify-customers-in-three-groups-of-engagement.)
  + [Most engaged users per app](https://github.com/ada-k/Telecommunication---UserAnalytics--/blob/dd95e5e68f964a35c8c900101bce0cabef6bf105//" \l "*-Aggregate-user-total-traffic-per-application-and-derive-the-top-10-most-engaged-users-per-application)
  + [Most used apps](https://github.com/ada-k/Telecommunication---UserAnalytics--/blob/dd95e5e68f964a35c8c900101bce0cabef6bf105//" \l "*-Plot-the-top-3-most-used-applications.)
  + [KMeans](https://github.com/ada-k/Telecommunication---UserAnalytics--/blob/dd95e5e68f964a35c8c900101bce0cabef6bf105//" \l "*-Using-k-means-clustering-algorithm,-group-users-in-k-engagement-clusters-based-on-the-engagement-metrics:)
* [3. User Experience Analysis](https://github.com/ada-k/Telecommunication---UserAnalytics--/blob/dd95e5e68f964a35c8c900101bce0cabef6bf105//" \l "3.-Experience-Analytics)
  + [Top 10, Bottom 10 and Frequent 10](https://github.com/ada-k/Telecommunication---UserAnalytics--/blob/dd95e5e68f964a35c8c900101bce0cabef6bf105//" \l "Compute-&-list-10-of-the-top,-bottom-and-most-frequent:)
  + [Throughput Distribution and Average TCP](https://github.com/ada-k/Telecommunication---UserAnalytics--/blob/dd95e5e68f964a35c8c900101bce0cabef6bf105//" \l "Compute-&-report:)
  + [KMeans](https://github.com/ada-k/Telecommunication---UserAnalytics--/blob/dd95e5e68f964a35c8c900101bce0cabef6bf105//" \l "Using-the-experience-metrics-above,-perform-a-k-means-clustering-(where-k-=-3)-to-segment-users-into-groups-of-experiences-and-provide-a-brief-description-of-each-cluster.)

**INTER-PRETATIONS FROM OVERVIEW:**

Apple, Huawei and Samsung are the leading manufacturers. The top 10 handsets also are manufactured by these companies. Some of the bottom 10 handsets also belong to these manufacturers. Thus the volume of the most common handsets should be increased to replace the least common sets.

Products from the bottom 3 manufacturers should no longer be stocked. Their extremely small volume matches small profit percentage in comparison to other manufacturers.

**NON - GRAPHICAL UNIVARIATE ANALYSIS:**

Summary of some of the variables:

****Bearer Id**** ~ This is the session identifier, thus not quantitative. It has 134709 instances thus the recorded number of xDR sessions are 134709.

****Dur. (ms).1**** ~ This is the total duration of an xDR session in micro seconds. Across the 134709 sessions, it has the following measures:

* 1. Average: 1.046091e+08
  2. Standard Deviation: 8.103734e+07
  3. Minimum Value: 7.142988e+06
  4. Lower Quartile 5.744206e+07
  5. Median 8.639998e+07
  6. Upper Quartile 1.324307e+08
  7. Maximum Value: 1.859336e+09

****Total UL (Bytes)**** ~ Data volume in bytes sent during this session. Across the 134709 sessions, it has the following measures:

* 1. Average: 4.112121e+07
  2. Standard Deviation: 1.127635e+07
  3. Minimum Value: 2.866892e+06
  4. Lower Quartile 3.322203e+07
  5. Median 4.558409e+08
  6. Upper Quartile 4.903424e+07
  7. Maximum Value: 7.833131e+07

****Total DL (Bytes)**** ~ Data volume in bytes received during this session. Across the 134709 sessions, it has the following measures:

* 1. Average: 4.546434e+085
  2. Standard Deviation: 2.441421e+08
  3. Minimum Value: 7.114041e+06
  4. Lower Quartile 2.431072e+08
  5. Median 1.826471e+06
  6. Upper Quartile 6.657051e+08
  7. Maximum Value: 9.029696e+08

****social\_media**** ~ Total social media data volume sent and received by the MS during this session. Across the 134709 sessions, it has the following measures:

* 1. Average: 1.828250e+06
  2. Standard Deviation: 1.035646e+06
  3. Minimum Value: 1.563000e+03
  4. Lower Quartile 9.322180e+05
  5. Median 8.639900e+04
  6. Upper Quartile 2.727487e+06
  7. Maximum Value: 3.650861e+06

****netflix**** ~ Total netflix data volume sent and received by the MS during this session. Across the 134709 sessions, it has the following measures:

* 1. Average: 2.262861e+07
  2. Standard Deviation: 9.260820e+06
  3. Minimum Value: 9.843200e+04
  4. Lower Quartile 1.597946e+07
  5. Median 2.263554e+07
  6. Upper Quartile 2.929044e+07
  7. Maximum Value: 4.519815e+07

****gaming**** ~ Total gaming data volume sent and received by the MS during this session. Across the 134709 sessions, it has the following measures:

* 1. Average: 4.303331e+08
  2. Standard Deviation: 2.440199e+08
  3. Minimum Value: 3.063580e+05
  4. Lower Quartile 2.187279e+08
  5. Median 4.316150e+08
  6. Upper Quartile 6.414159e+08
  7. Maximum Value: 8.592028e+08

****youtube)**** ~ Total youtube data volume sent and received by the MS during this session. Across the 134709 sessions, it has the following measures:

* 1. Average: 2.264348e+07
  2. Standard Deviation: 9.246800e+06
  3. Minimum Value: 7.890300e+04
  4. Lower Quartile 1.599846e+07
  5. Median 2.266177e+07
  6. Upper Quartile 2.929260e+07
  7. Maximum Value: 4.519008e+07

****email**** ~ Total email data volume sent and received by the MS during this session. Across the 134709 sessions, it has the following measures:

* 1. Average: 2.259102e+06
  2. Standard Deviation: 1.071109e+06
  3. Minimum Value: 8.359000e+03
  4. Lower Quartile 1.359344e+06
  5. Median 2.263567e+06
  6. Upper Quartile 3.159818e+06
  7. Maximum Value: 4.518036e+06

****google**** ~ Total google data volume sent and received by the MS during this session. Across the 134709 sessions, it has the following measures:

* 1. Average: 7.807295e+06
  2. Standard Deviation: 3.516420e+06
  3. Minimum Value: 4.033000e+04
  4. Lower Quartile 4.943599e+06
  5. Median 7.812835e+06
  6. Upper Quartile 1.068280e+07
  7. Maximum Value: 1.552878e+07

****other**** ~ Total Other data volume sent and received by the MS during this session. Across the 134709 sessions, it has the following measures:

* 1. Average: 4.293653e+08
  2. Standard Deviation: 2.432681e+08
  3. Minimum Value: 1.490450e+05
  4. Lower Quartile 2.185534e+08
  5. Median 4.299865e+08
  6. Upper Quartile 6.399275e+08
  7. Maximum Value: 8.595209e+08