Introduction

TellCo, a mobile service provider in the Republic of Pefkakia, is being considered for acquisition.

Objectives:

- Understand customer behavior and engagement to identify opportunities for growth and profitability.

Data Overview:

- Description of the dataset: xDR sessions data, customer metrics, application usage.
- Key variables: Session frequency, duration, total data usage, and application-specific usage.

Analysis Tasks: Conduct user overview analysis and user engagement analysis.

Dataset Description: xDR sessions, customer metrics, application usage data.

Key Variables:

- Session frequency, session duration, download/upload data (DL/UL).
- Application-specific data (e.g., Social Media, YouTube, Gaming).

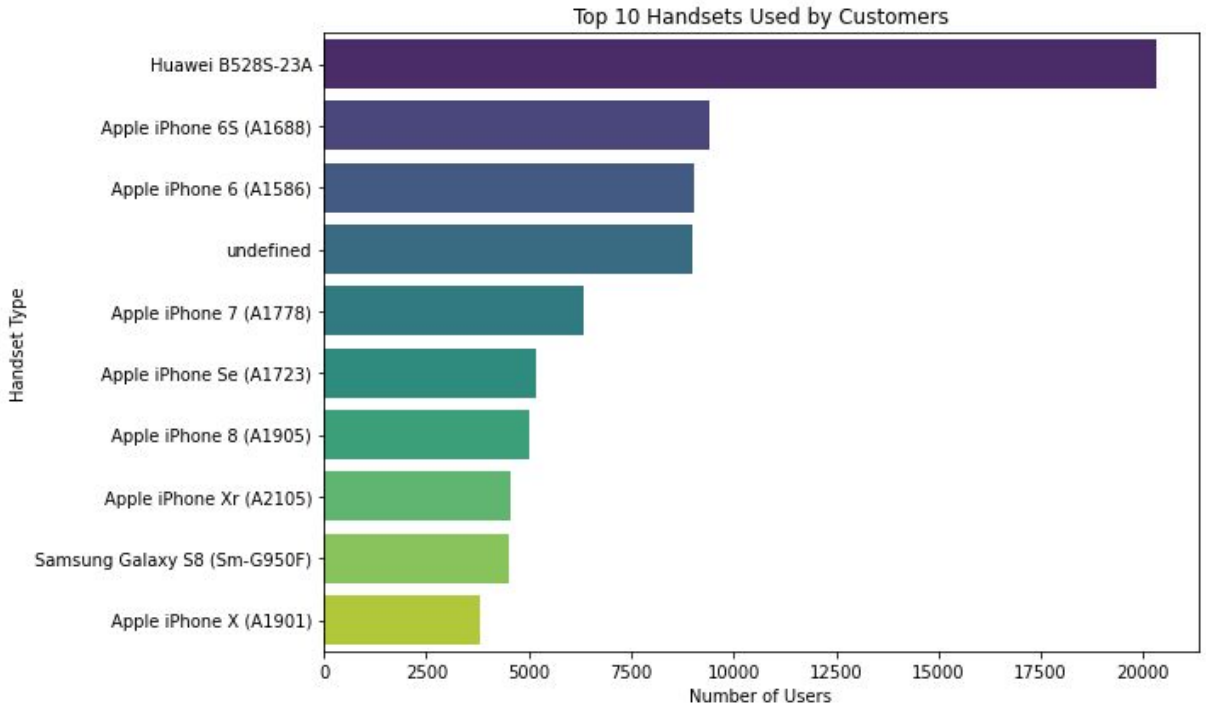
Top 10 Handsets Used by Customers

MOST POPULAR HANDSETS:

- Huawei B528S-23A, Apple iPhone 6S, Apple iPhone 6 among others.

Insight:

Focus on these popular models for targeted marketing or partnerships.



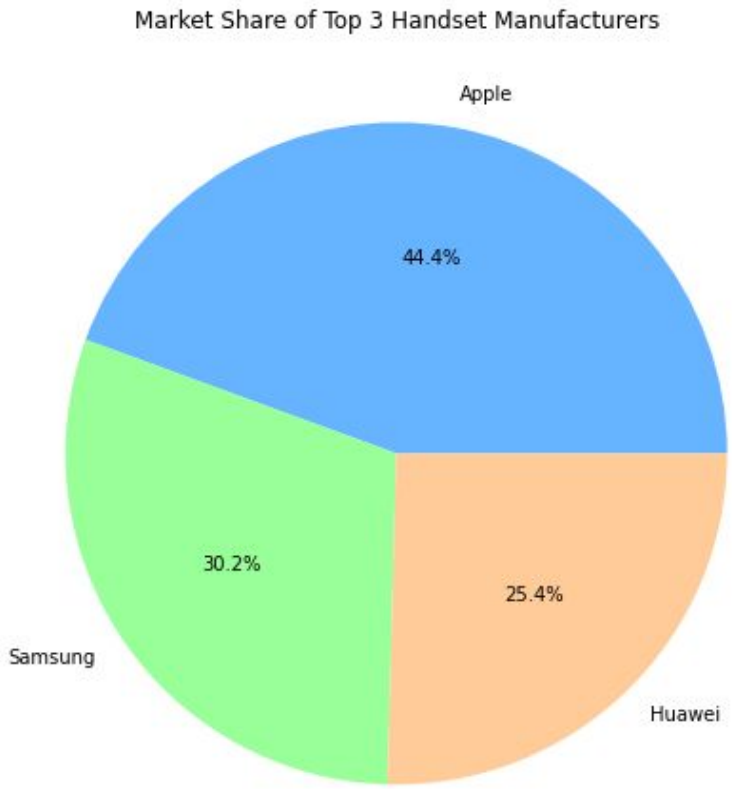
Top 3 Handset Manufacturers

MANUFACTURERS OVERVIEW:

- Apple (59,565), Samsung (40,839), Huawei (34,423).

Insight: Apple leads in handset usage, indicating strong customer loyalty.

- The Figure illustrates the dominance of key manufacturers in the customer base, which is critical for strategic alliances or marketing.

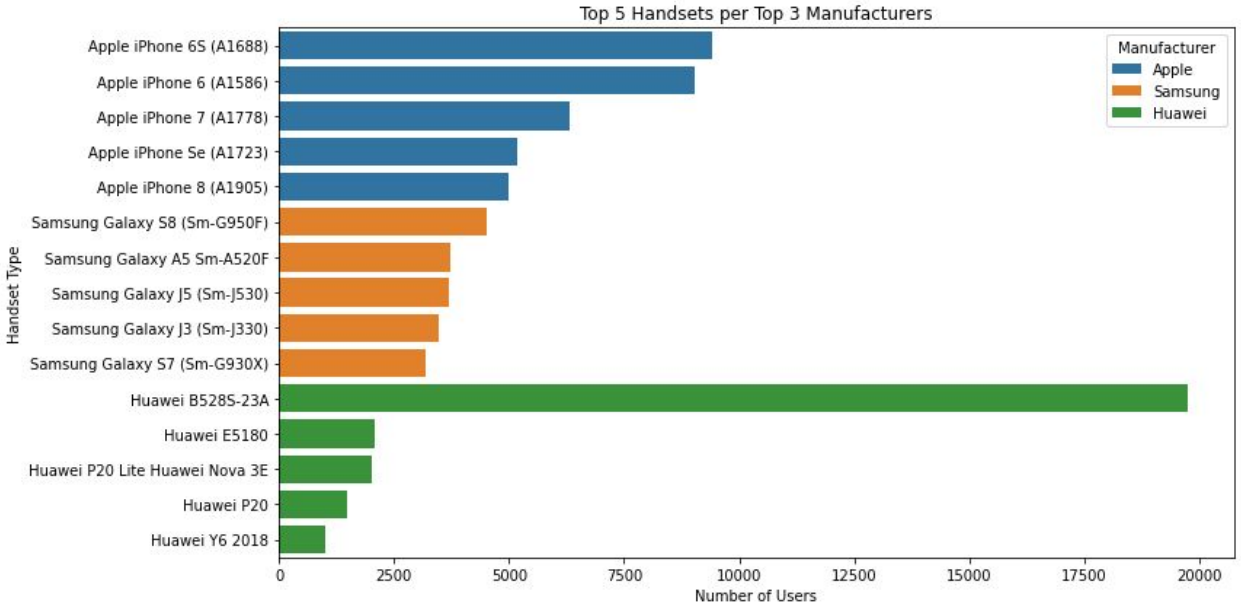


Top 5 Handsets per Top 3 Manufacturers

BREAKDOWN:

- Apple: iPhone 6S, iPhone 6, iPhone 7, iPhone SE, iPhone 8.
- Samsung: Galaxy S8, Galaxy A5, Galaxy J5, Galaxy J3, Galaxy S7.
- Huawei: B528S-23A, E5180, P20 Lite, P20, Y6 2018.

Insight: Specific handset popularity varies significantly across manufacturers.



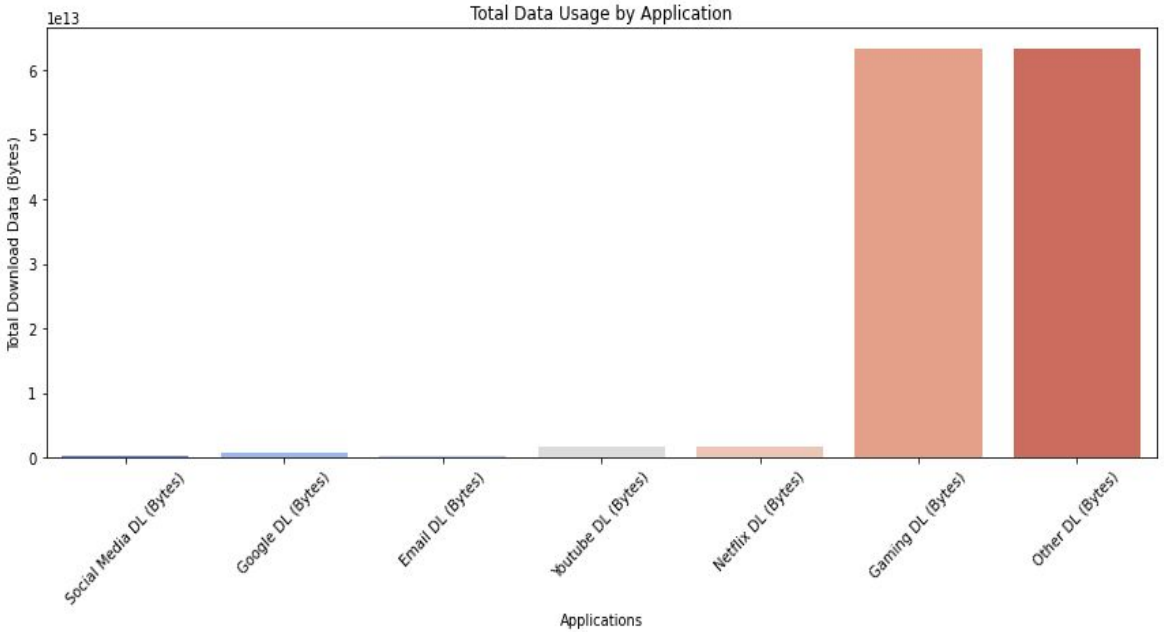
Overview of User Behavior on Applications

Aggregated Metrics:

- Number of xDR sessions, session duration, total DL/UL.
- Key applications: Social Media, Google, YouTube, Netflix, Gaming.

Top Applications by Total Traffic

1. Gaming: 64,088.92 GB
2. Other: 63,954.25 GB
3. YouTube: 3,372.20 GB
4. Netflix: 3,370.06 GB
5. Google: 1,162.85 GB

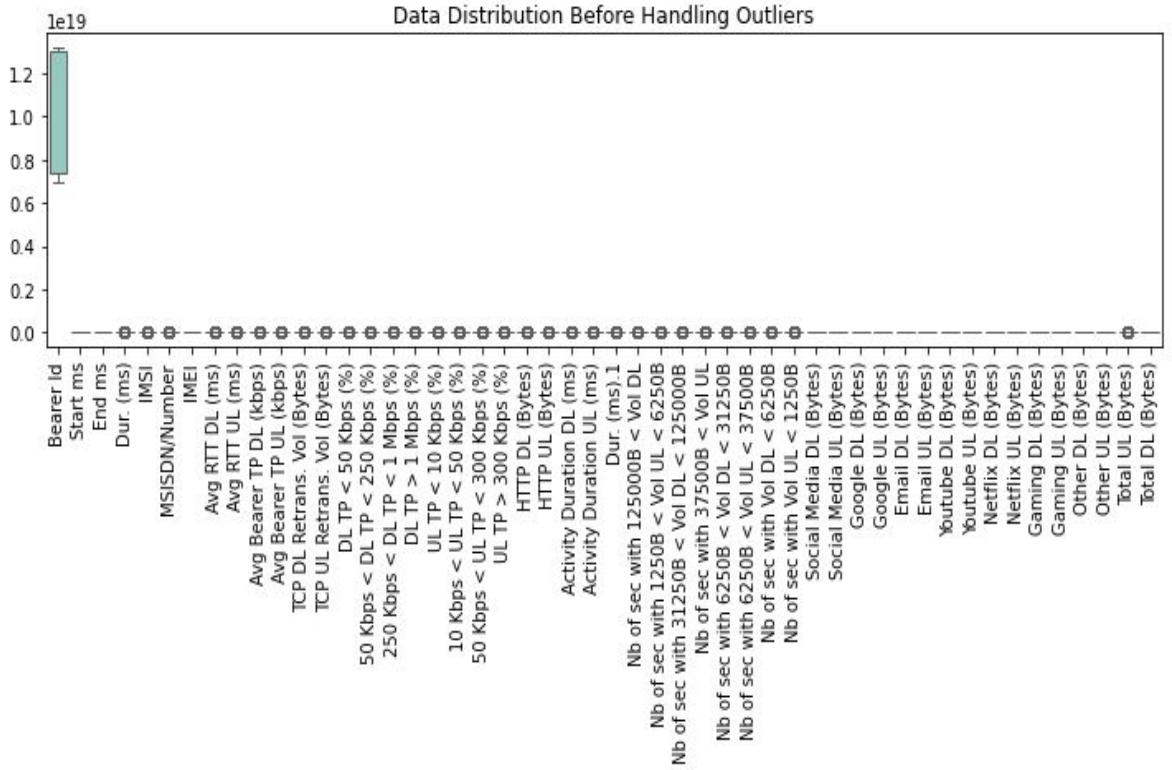


Data Preparation: Handling Missing Values and Outliers

Approach:

- Replaced missing values using mean or median.
- Detected outliers using z-scores and treated them.

Impact: Improved data quality for more accurate analysis.



Basic Metrics Analysis

Metric	Count	Mean	Std Dev	Min	25th Percentile	Median	75th Percentile	Max
Sessions	10	3	1.49	1	2	3	4	5
Duration (ms)	10	355,369	210,294	103,098	215,406	300,000	500,000	658,341
Traffic (GB)	10	1.53	0.77	0.46	1.04	1.50	2.13	2.50

Metrics Calculated: Mean, median, standard deviation.

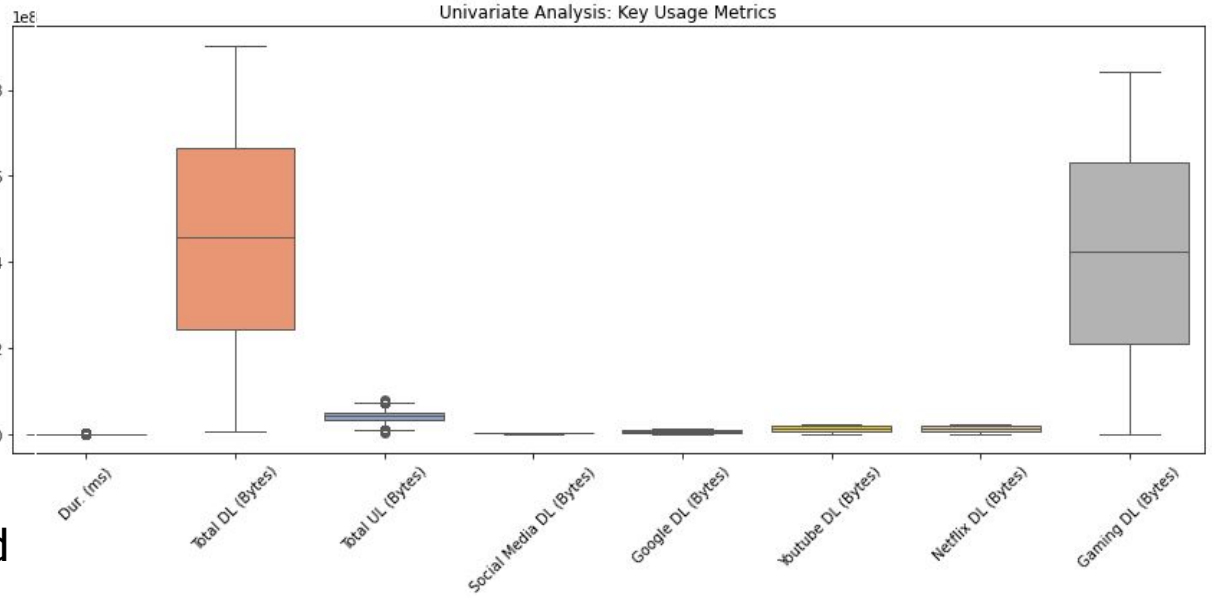
Insight: Provides an understanding of typical user behavior and data usage patterns.

Univariate Analysis: Key Metrics

Objective: Examine distribution and variability in user behavior.

Key Findings:

- **Total Duration**: Mostly moderate with some heavy users.
- **Total Data Usage**: Skewed toward low/moderate use; high outliers in streaming and gaming.
- **Application-Specific Data**: High variability in YouTube, Netflix, Gaming; less in Social Media, Email.

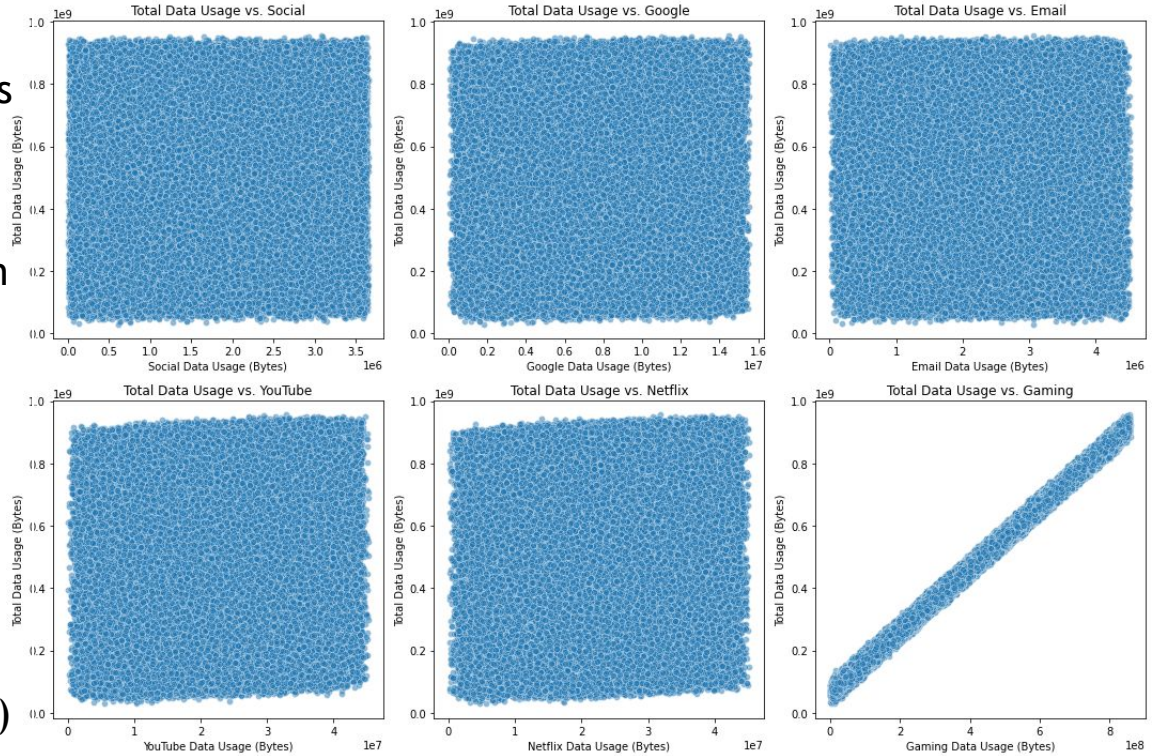


Bivariate Analysis: Applications vs. Total Data Usage

Objective: Understand the relationship between total data usage and specific application usage to identify key drivers of user engagement.

Key Findings:

- Strong correlations observed between total data usage and specific high-traffic applications such as YouTube, Netflix, and Gaming.
- Lower but notable correlations for applications like Social Media and Google.
- Insights indicate that multimedia content (video streaming and gaming) contributes significantly to total data consumption.



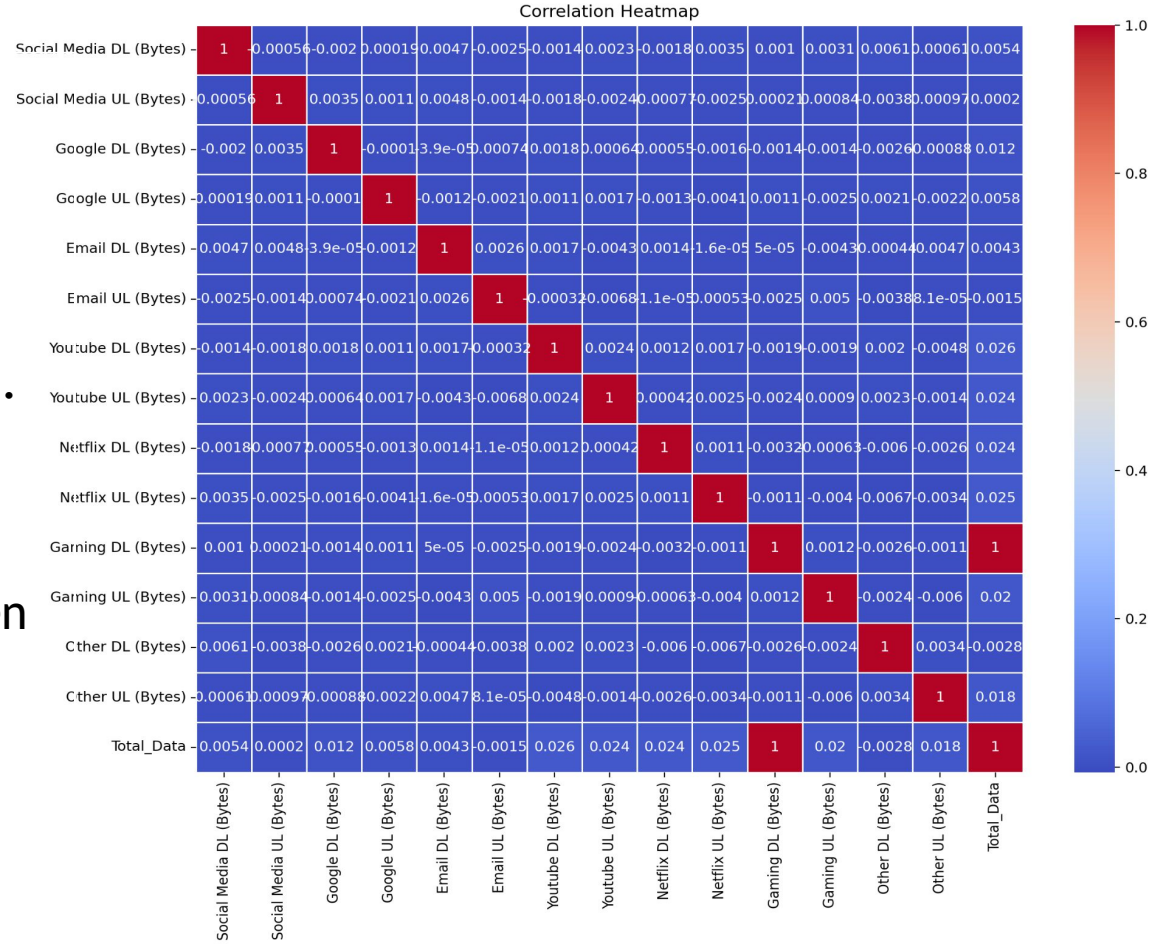
Correlation Matrix Analysis

Interpretation:

- Strong correlations identified among certain data usage variables (e.g., YouTube and Gaming data).

Insight:

- Helps understand interdependencies between different applications.

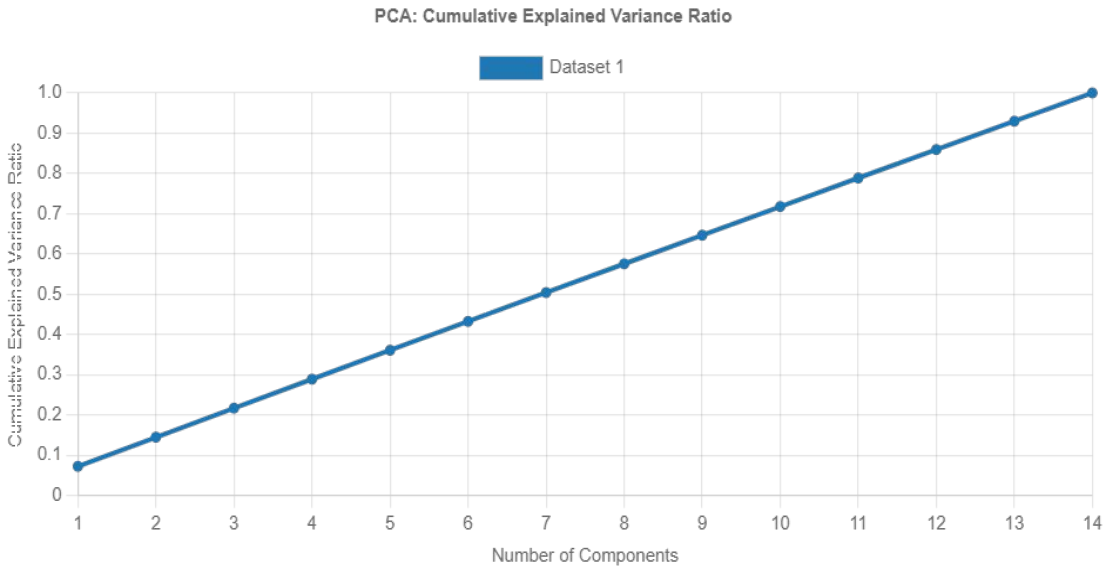


Dimensionality Reduction - PCA

PCA was performed to reduce the dimensionality of the data

Key Findings from PCA:

- The number of components needed to explain 80% of the variance is 12.
- The explained variance ratio of the first component is 0.07.
- The explained variance ratio of the second component is 0.07.
- The cumulative explained variance ratio of the first two components is 0.15.



These findings suggest:

- a relatively large number of components (12) are necessary to capture the majority (80%) of the variance,
- There is a complexity and high-dimensional nature of the dataset.

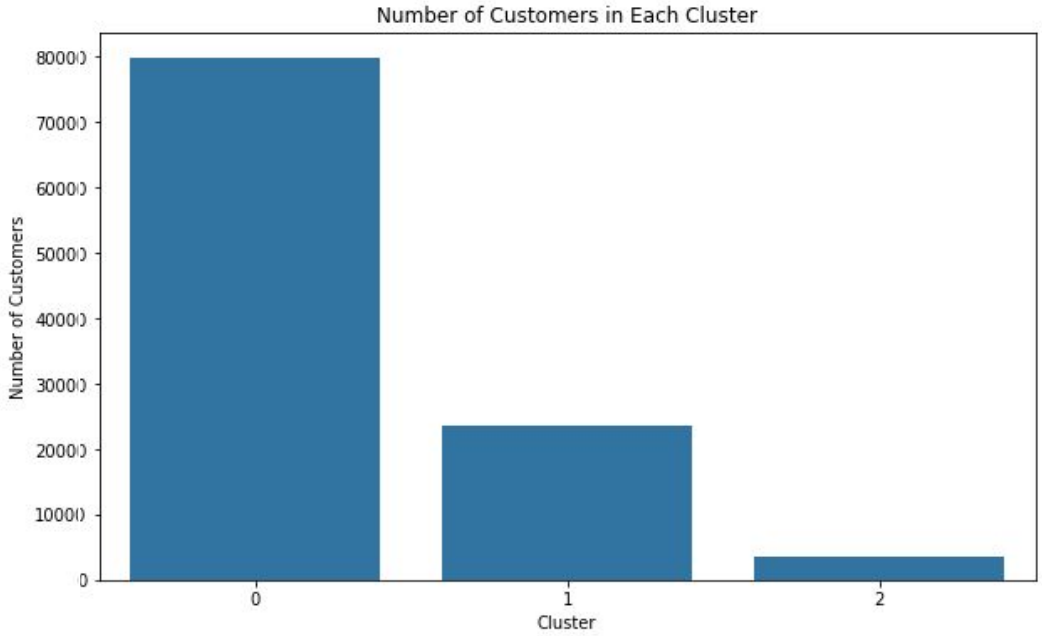
User Engagement Metrics and Clustering

Cluster Statistics:

A. **Cluster 0:**
   - Size: **23,437 customers**
   - Average Sessions: **2.18**
   - Average Duration: **215.41 seconds**
   - Average Traffic: **1.04 GB**

B. **Cluster 1:**
   - Size: **79,872 customers**
   - Average Sessions: **1.02**
   - Average Duration: **103.10 seconds**
   - Average Traffic: **0.46 GB**

C. **Cluster 2:**
   - Size: **3,547 customers**
   - Average Sessions: **4.41**
   - Average Duration: **658.34 seconds**
   - Average Traffic: **2.13 GB**



Interpretation:

- **Cluster 2** consists of highly engaged users with the most sessions, longest durations, and highest data traffic.
- **Cluster 1** represents users with the lowest engagement.
- **Cluster 0** contains moderately engaged users, with mid-level values across the metrics.

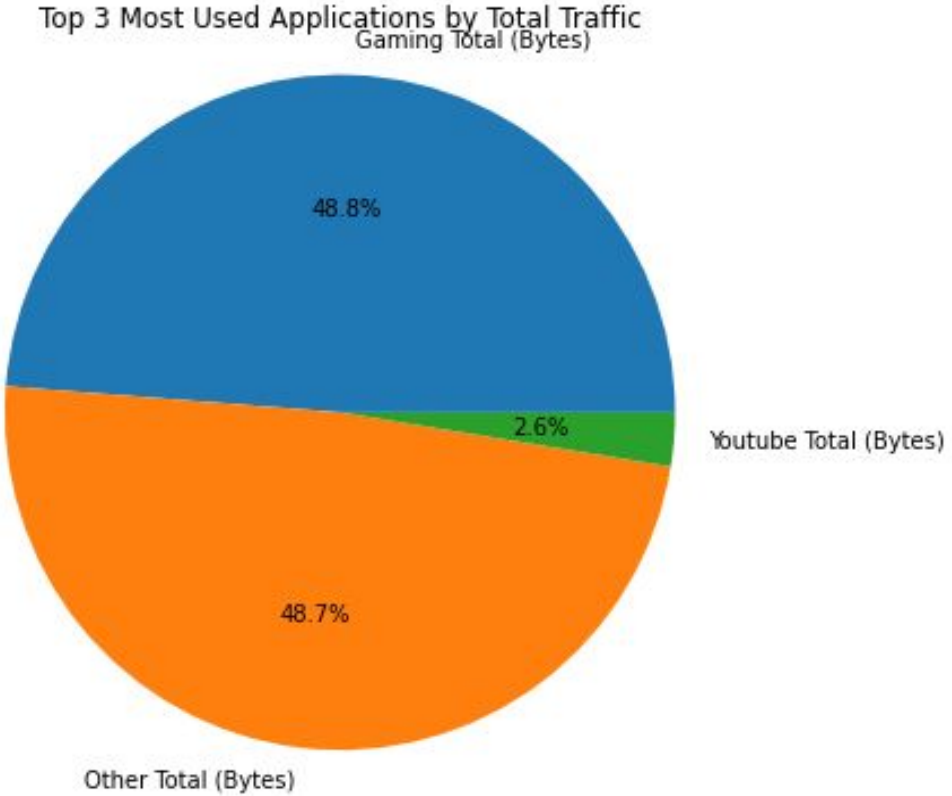
User Engagement Metrics and Clustering

User Engagement:

A. **User 33626320676**
   - Total Traffic: **7.42 GB**
   - Top Application: **Gaming** (6.41 GB)

B. **User 33760536639**
   - Total Traffic: **7.93 GB**
   - Top Application: **Gaming** (6.95 GB)

C. **User 33659725664**
   - Total Traffic: **7.18 GB**
   - Top Application: **Gaming** (6.26 GB)

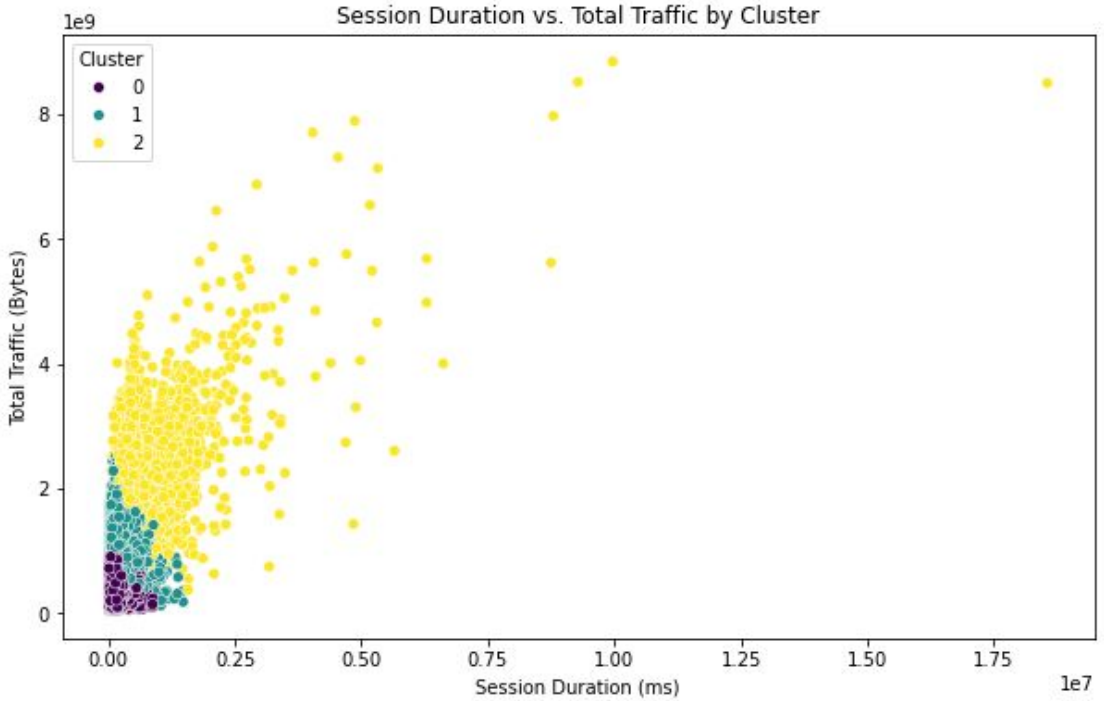


User Engagement Metrics and Clustering …Con'd

This scatter plot visualizes the relationship between session duration and total traffic for each cluster.

We can observe that:

- **Cluster 2 (high engagement):** has longer session durations and higher total traffic.
- **Cluster 1 (low engagement):** has shorter session durations and lower total traffic.
- **Cluster 0 (medium engagement):** falls between the other two clusters.



Recommendations and Next Steps

Based on the findings:

1. **Target High-Engagement Users (Cluster 2)**:
   - Develop personalized marketing campaigns or offers for Cluster 2 users
2. **Improve Service for Low-Engagement Users (Cluster 1)**:
   - Explore reasons for low engagement and address potential barriers (e.g., poor network quality or lack of appealing services).
3. **Promote Popular Applications**:
   - Focus on gaming and video streaming partnerships, as they are highly popular and drive significant data usage.
4. **Network Optimization**:
   - Allocate more resources to support high-traffic applications like Gaming and YouTube to enhance user experience and reduce churn.