A logo of a naval officer

Description automatically generated

Analysing mobile usage patterns via exploratory data analysis

Assignment 02

Statistical Computing with R – cM2062

Faculty of Computing

Department of Computational Mathematics

# **Table of Contents**

[1. Introduction 2](#_Toc197024396)

[2. Dataset Overview 2](#_Toc197024397)

[3. Exploratory Data Analysis EDA 3](#_Toc197024398)

[3.1. Summary Statistics 3](#_Toc197024399)

[Interpretation: 3](#_Toc197024400)

[3.2. Gender and User Behavior Class Distribution 4](#_Toc197024401)

[Interpretation: 4](#_Toc197024402)

[3.3. Correlation Between Variables 5](#_Toc197024403)

[Interpretation: 5](#_Toc197024404)

[3.4. Boxplot of Age by Gender 6](#_Toc197024405)

[Interpretation: 6](#_Toc197024406)

[3.5. Scatterplots and Heat Maps 7](#_Toc197024407)

[Interpretation: 7](#_Toc197024408)

[3.6. Battery Drain and Screen Time Analysis 9](#_Toc197024409)

[Interpretation: 9](#_Toc197024410)

[4. User Behavior Classification 11](#_Toc197024411)

[Interpretation: 11](#_Toc197024412)

[5. Key Insights and Findings 11](#_Toc197024413)

[Interpretation: 11](#_Toc197024414)

[6. Recommendations 12](#_Toc197024415)

[Interpretation: 12](#_Toc197024416)

[7. Conclusion 12](#_Toc197024417)

[8. Appendix: R Code and Visualizations 13](#_Toc197024418)

# Introduction

In today’s digital era, mobile devices are deeply woven into the fabric of everyday life, serving as essential tools for communication, productivity, entertainment, and social interaction. Understanding how people use their smartphones-how much time they spend, what drains their battery, and how these patterns differ by demographic-is crucial for developers, manufacturers and service providers seeking to optimize user experience and resource allocation. This report presents a comprehensive analysis of mobile device usage among 700 users, leveraging a robust dataset to uncover behavioral trends and actionable insights.

# Dataset Overview

The dataset comprises 700 users, each described by:

* **User ID**: Unique identifier
* **Device Model**: (e.g., Google Pixel 5, iPhone 12
* **Operating System**: Android or iOS
* **App Usage Time**: Minutes per day
* **Screen On Time**: Hours per day
* **Battery Drain**: mAh per day
* **Number of Apps Installed**
* **Data Usage**: MB per day
* **Age**
* **Gender**: Male or Female
* **User Behavior Class**: 1 (light) to 5 (heavy)

All variables were checked for missing values and consistency, ensuring the reliability of insights drawn from the analysis.

# Exploratory Data Analysis EDA

## Summary Statistics

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Minutes** | **1st Quartile** | **Median** | **Mean** | **3rd Quartile** | **Max** |
| App Usage Time (min/day) | 30 | 113.2 | 227.5 | 271.1 | 434.2 | 598 |
| Screen On Time  (hrs/day) | 1 | 2.5 | 4.9 | 5.27 | 7.4 | 12 |
| Battery Drain  (mAh/day) | 302 | 722.2 | 1502.5 | 1525.2 | 2229.5 | 2993 |
| Number of Apps  Installed | 10 | 26 | 49 | 50.68 | 74 | 99 |
| Data Usage  (MB/day) | 102 | 373 | 823.5 | 929.7 | 1341 | 2497 |
| Age | 18 | 28 | 38 | 38.48 | 49 | 59 |
| User Behavior  Class | 1 | 2 | 3 | 2.99 | 4 | 5 |

### Interpretation:

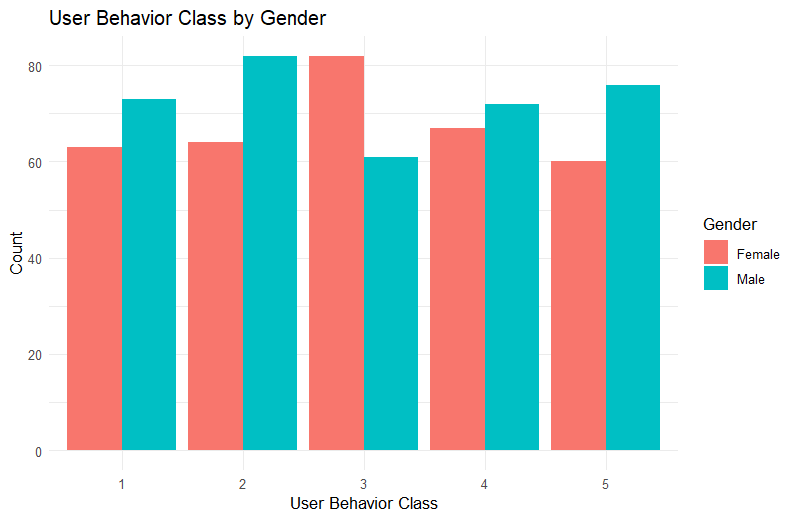
These statistics reveal a highly diverse user base, both in terms of engagement and demographic profile. The wide range and right-skewed distributions in app usage, screen time, battery drain, and data usage indicate a significant segment of "power users" who drive up the averages. For example, while the median app usage is about 3.8 hours/day, some users spend up to 10 hours daily. This highlights the importance of focusing optimization efforts on this influential group, as their behavior can disproportionately impact network resources, battery life, and app performance. The diversity in age and number of apps installed further reflects the broad appeal and varied use cases of modern smartphones.

## Gender and User Behavior Class Distribution

The distribution of user behavior classes by gender shows both males and females are well represented across all usage classes. Males are slightly more prevalent in lighter usage classes, while females are marginally more frequent in the heaviest usage class.

### Interpretation:

This pattern suggests that gender is not a strong determinant of mobile usage intensity in this sample. The near parity indicates that mobile engagement is increasingly gender-neutral, likely due to the universal adoption of smartphones for a wide range of activities. The slight overrepresentation of females among heavy users could reflect broader trends in digital engagement, but the difference is not pronounced. For marketers and developers, this finding means that targeting by gender alone is less effective than targeting by usage behavior or needs.



## Correlation Between Variables

The correlation matrix reveals strong positive relationships between app usage time, screen on time, battery drain, and data usage. Users who spend more time on apps also keep their screens on longer, which leads to greater battery and data consumption. The number of apps installed is moderately correlated with battery drain and data usage.

### Interpretation:

Resource consumption is primarily driven by user engagement rather than background activity or the sheer number of installed apps. This means that interventions to reduce battery or data usage should focus on the most engaged users, not just those with many apps. For developers and service providers, this supports prioritizing optimization for high-engagement scenarios and considering features like adaptive battery management or personalized data-saving recommendations for heavy users.

A screenshot of a computer

AI-generated content may be incorrect.

## Boxplot of Age by Gender

The boxplot shows that the age distribution is quite similar across genders, with both male and female users having a median age around 38 years and overlapping interquartile ranges.

### Interpretation:

Mobile device usage is a cross-generational phenomenon. There are no significant age or gender differences in the dataset, which enhances the generalizability of the findings and supports broad-based strategies for user engagement and optimization. This also suggests that product and service improvements will benefit a wide demographic, rather than being limited to any particular age or gender group.

A graph showing a couple of colored squares

AI-generated content may be incorrect.

## Scatterplots and Heat Maps

Several scatterplots and heat maps show the relationships between screen time, app usage, and battery consumption.

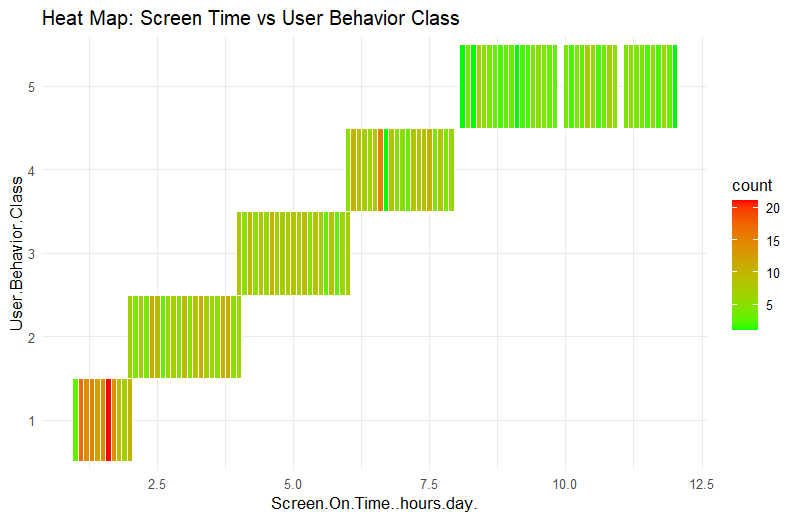
 **App Usage vs**. **Screen Time**: A strong linear relationship confirms that app engagement is the main driver of screen time.

 **Screen Time vs**. **User Behavior Class** **Heat Map**): Higher user behavior classes are concentrated at higher screen time levels.

 **Number of Apps Installed vs**. **Battery Drain**: More apps can contribute to higher battery consumption, but the relationship is not as strong as with screen time or app usage.

### Interpretation:

These visualizations provide compelling evidence that the most engaged users Class 5) are also the ones who consume the most resources. The heat map, in particular, makes it easy to identify user segments that may benefit from targeted interventions, such as battery optimization or personalized data plans. The weaker relationship between number of apps and battery drain suggests that simply having more apps does not necessarily lead to higher resource use-it's the intensity of usage that matters most.



A graph showing a line of blue dots

AI-generated content may be incorrect.

A graph showing a number of apps

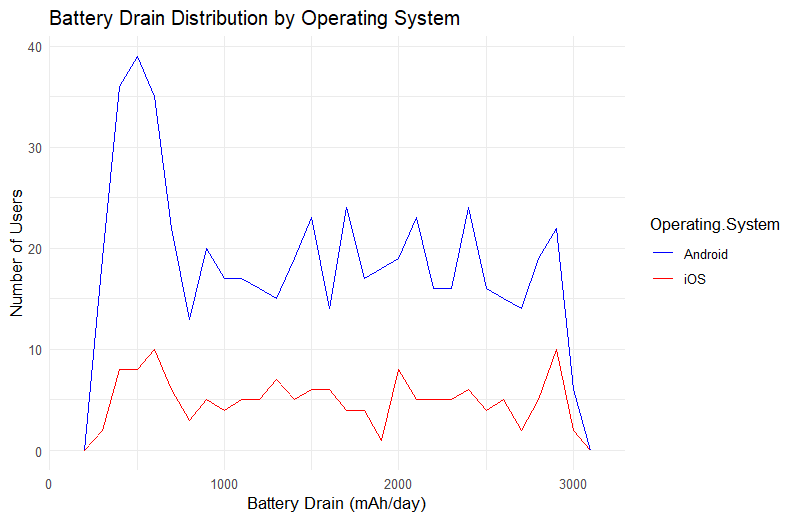
AI-generated content may be incorrect.

## Battery Drain and Screen Time Analysis

The scatterplot reveals that battery drain and screen time are highly correlated. Users who spend more time on their devices experience significantly higher battery consumption. Android users are more common among the highest battery drainers, and the pie chart of phone models shows a diverse mix of devices.

### Interpretation:

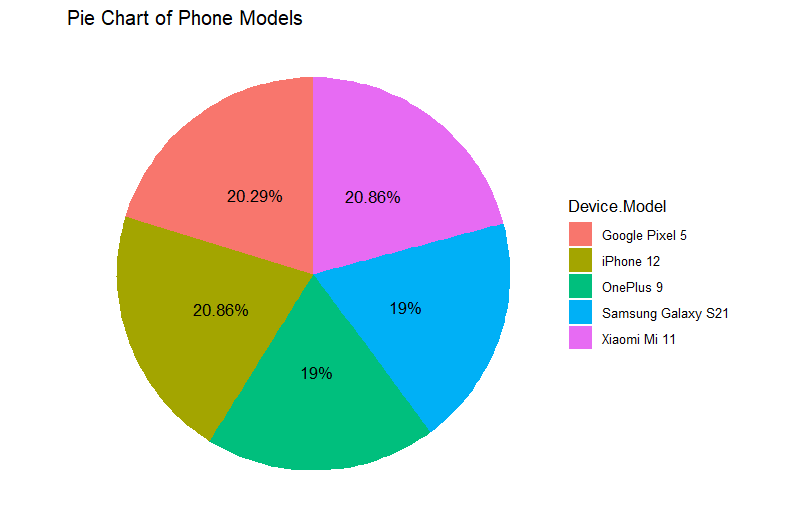
These findings highlight the direct impact of user engagement on battery life and reinforce the need for device manufacturers and app developers to focus on optimizing battery consumption for heavy users. The operating system differences may warrant further investigation, as they could indicate underlying platform-specific inefficiencies or user preferences. The diversity of device models supports the robustness of the analysis and reduces the risk of device-specific bias.



A graph of a chart

AI-generated content may be incorrect.A graph showing a number of apps

AI-generated content may be incorrect.



# User Behavior Classification

User behavior classification segments users based on their screen time and app usage. Class 1 users are light users, while Class 5 users are highly engaged, spending more time on apps and their screens.

## Interpretation:

This classification enables targeted analysis and recommendations. Heavy users, though a minority, account for a disproportionately large share of total resource consumption.

Understanding their characteristics is essential for effective optimization and engagement strategies. For example, personalized notifications, battery-saving tips, or exclusive features could be offered to these users to enhance satisfaction and loyalty.

# Key Insights and Findings

* **Heavy Users**: A small segment of heavy users Class 5) consumes a disproportionate amount of battery and data. These users are critical for network planning and app optimization, as their behavior can drive overall resource demand.
* **Android Dominance**: The dataset is heavily skewed towards Android users, consistent with global market trends. This suggests that improvements in Android efficiency could have a broad impact.
* **Age Distribution**: Users from various age groups engage with mobile devices, with no single demographic dominating the dataset. This points to the universal nature of mobile engagement.
* **Screen Time vs**. **Battery Drain**: Longer screen time leads to significantly higher battery consumption and data usage, underscoring the importance of battery-efficient design and data management features.
* **Gender**: Minimal difference in mobile usage patterns across genders, with slight variations in heavy user behavior. This reinforces the idea that behavior-based segmentation is more effective than demographic targeting.

## Interpretation:

These insights suggest that technical and commercial interventions should be tailored to usage patterns, not demographics. Heavy users have an outsized impact on network and device resources, making them a key focus for optimization. The findings also highlight opportunities for service providers to create high-value plans and for developers to design features that cater specifically to the needs of these power users.

# Recommendations

* **App Developers**: Focus on battery optimization for heavy app users, especially in Class 5.
* Consider features like adaptive brightness, background process management, and personalized usage insights.
* **Device Manufacturers**: Improve battery efficiency for Android devices, as they show a wide range of battery consumption. Innovations in hardware and software integration can help extend device longevity for the most active users.
* **Service Providers**: Offer data plans targeting high data users, with flexible options and rewards for loyalty. Usage-based pricing or "power user" bundles could better meet the needs of this segment.
* **Marketing**: Develop behavior-targeted campaigns, as usage class is a stronger predictor of engagement than age or gender. For example, offer exclusive features or rewards to heavy users to increase loyalty and engagement.

## Interpretation:

These recommendations are actionable and directly informed by the data. By aligning strategies with actual user behavior, stakeholders can improve satisfaction, reduce churn, and optimize resource allocation. Continuous monitoring and segmentation will also help adapt to evolving user patterns over time.

# Conclusion

This report provides deep insights into mobile device usage behavior, showing how screen time, app usage, and battery consumption are closely linked. With Android users dominating the dataset, the findings highlight the need for battery-efficient apps and targeted marketing based on user behavior. By focusing on the needs of heavy users and adopting behavior-based strategies, stakeholders can drive both user satisfaction and business success.

# Appendix: R Code and Visualizations

library(ggplot2)

library(dplyr)

library(readxl)

library(corrplot)

library(reshape2)

library(scales)

data <- read\_excel("C:/Users/ASUS VIVOBOOK/Downloads/user\_behavior\_dataset.xlsx", sheet = "user\_behavior\_dataset")

# View structure and summary

str(data)

summary(data)

sapply(data, function(x) sum(is.na(x)))

# Convert relevant columns to factors

data$Gender <- as.factor(data$Gender)

data$Operating System <- as.factor(data$Operating System)

data$Device Model <- as.factor(data$Device Model)

data$User Behavior Class <- as.factor(data$User Behavior Class)

num\_cols <- names(data)[sapply(data, is.numeric)]

# Create histograms using tidy evaluation

for (col in num\_cols) {

p <- ggplot(data, aes(x = !!sym(col))) +

geom\_histogram(fill = "steelblue", color = "black", bins = 30) +

ggtitle(paste("Histogram of", col)) +

xlab(col) +

theme\_minimal()

print(p)

}

ggplot(data, aes(x = Screen On Time (hours/day))) +

geom\_histogram(fill = "green", color = "black", bins = 30, alpha = 0.9) +

labs(

title = "Screen Time Distribution",

x = "Screen On Time (hours/day)",

y = "Number of Users"

) +

theme\_minimal()

## #1)

ggplot(data, aes(x = Gender, fill = Gender)) +

geom\_bar(color = "black") +

scale\_fill\_manual(values = c("Female" = "pink", "Male" = "blue")) +

ggtitle("Histogram of Gender") +

ylab("Count") +

theme\_minimal()

## #2)

ggplot(data, aes(x = Operating System)) +

geom\_bar(fill = "lightgreen", color = "black") +

ggtitle("Bar Chart of Operating System") +

ylab("Count") +

theme\_minimal()

## #3)

# Count and calculate percentage

phone\_model\_counts <- data %>%

count(Device Model) %>%

mutate(percentage = round(n / sum(n) \* 100, 2),

label = paste0(percentage, "%"))

# Pie chart with only percentage labels

ggplot(phone\_model\_counts, aes(x = "", y = n, fill = Device Model)) +

geom\_bar(width = 1, stat = "identity") +

coord\_polar("y", start = 0) +

theme\_void() +

geom\_text(aes(label = label),

position = position\_stack(vjust = 0.5), size = 4) +

ggtitle("Pie Chart of Phone Models")

## #4)

ggplot(data, aes(x = Number of Apps Installed, y = Battery Drain (mAh/day))) +

geom\_point(color = "hotpink") +

ggtitle("Scatter Plot: Number of Apps vs Battery Drain") +

theme\_minimal()

## #5)

ggplot(data, aes(x = App Usage Time (min/day), y = Screen On Time (hours/day))) +

geom\_point(color = "navyblue") +

ggtitle("Scatter Plot: App Usage vs Screen Time") +

theme\_minimal()

## #6)

Box plot of Battery Drain by Screen Time Group

data <- data %>%

mutate(ScreenTimeGroup = cut(Screen On Time (hours/day),

breaks = c(0, 2, 4, 6, 8, Inf),

labels = c("0–2", "2–4", "4–6", "6–8", "8+")))

ggplot(data, aes(x = ScreenTimeGroup, y = Battery Drain (mAh/day))) +

geom\_boxplot(fill = "orange") +

ggtitle("Box Plot: Screen Time Groups vs Battery Drain") +

xlab("Screen Time Group (hours/day)") +

ylab("Battery Drain (mAh/day)") +

theme\_minimal()

## #7)

heat\_data <- data %>%

group\_by(User Behavior Class, Screen On Time (hours/day)) %>%

summarise(count = n(), .groups = "drop")

ggplot(heat\_data, aes(x = Screen On Time (hours/day), y = User Behavior Class, fill = count)) +

geom\_tile(color = "white") +

scale\_fill\_gradient(low = "green", high = "red") +

ggtitle("Heat Map: Screen Time vs User Behavior Class") +

theme\_minimal()

## #8)

ggplot(data, aes(x = Battery Drain (mAh/day), color = Operating System)) +

geom\_freqpoly(binwidth = 100) +

ggtitle("Battery Drain Distribution by Operating System") +

xlab("Battery Drain (mAh/day)") +

ylab("Number of Users") +

scale\_color\_manual(values = c("Android" = "blue", "iOS" = "red")) +

theme\_minimal()

**# Scatterplot: Gender vs User Behavior with OS as color**

ggplot(data, aes(x = Gender, y = User Behavior Class, color = Operating System)) +

geom\_jitter(width = 0.2, height = 0.2, size = 2) + # Jitter to avoid overlapping points

scale\_color\_manual(values = c("Android" = "green", "iOS" = "blue")) +

ggtitle("Scatterplot: Gender vs User Behavior by OS") +

xlab("Gender") +

ylab("User Behavior Class") +

theme\_minimal()

# **Boxplots by Gender**

num\_cols <- names(data)[sapply(data, is.numeric)]

for (col in num\_cols) {

p <- ggplot(data, aes(x = Gender, y = .data[[col]], fill = Gender)) +

geom\_boxplot() +

ggtitle(paste("Boxplot of", col, "by Gender")) +

xlab("Gender") +

ylab(col) +

theme\_minimal()

print(p)

}

# **Correlation matrix (numeric only)**

num\_data <- data %>% select(where(is.numeric)) # Updated select\_if to select()

cor\_matrix <- cor(num\_data, use = "complete.obs") # Handle any potential NA values

corrplot(cor\_matrix, method = "color", addCoef.col = "black", number.cex = 0.7)

# **Bar plot: User Behavior Class by Gender**

ggplot(data, aes(x = User Behavior Class, fill = Gender)) +

geom\_bar(position = "dodge") +

ggtitle("User Behavior Class by Gender") +

theme\_minimal() +

xlab("User Behavior Class") +

ylab("Count")

**# Mean usage stats per User Behavior Class**

data %>%

group\_by(User Behavior Class) %>%

summarise(across(where(is.numeric), ~ mean(.x, na.rm = TRUE))) %>%

  print()

# Group Members Details

|  |  |  |
| --- | --- | --- |
| **No** | **Index Number** | **Name** |
| 01 | D/DBA/24/0002 | Methara Hewavitharana |
| 02 | D/DBA/24/0006 | Akithma Thewanjee |
| 03 | D/DBA/24/0017 | Binara Gunathilake |
| 04 | D/DBA/24/0032 | Dushan Liyange |