Mobilytics

Mingi Kwon, Avinash Gondela, Ksheer Agrawal, Dibyesh Sahoo and Saachi Shenoy

Understanding Mobile Device Usage:

Identifying Behavioral Patterns & High-Risk Users

Our Research Questions

- How do people use their mobile devices?
- Which users are at highest risk of excessive usage?
- What factors are most predictive of heavy usage?
- Are there distinct user segments or clusters? (ex: heavy users, normal users, light users)

Why Should We Care?

- Screen time continues to rise as mobile devices become an unavoidable part of life in the 21st century

- High phone use is associated with many risks such as addiction, reduced attention span, sleep issues, distraction and decreased productivity.

- Detection and recognizing the signs of heavy phone-use patterns allows developers, parents, and public-health official to design targeted phone addiction prevention for teens and adults

- We are already seeing tools tailored to tracking digital well-being: Screen Time, Focus Modes and “Take a Break” reminders

- Understanding “tells” of high phone use supports better UX, healthier defaults, and more personalized guidance.

Dataset Overview

700 users, 11 columns

Main Features

- Screen On Time
- App Usage Time
- Battery Drain
- Data Usage
- Apps Installed
- OS
- Device Model
- Demographics

Data Cleaning & Validation

1. Removal of missing/invalid values
2. Checking for ranges and outliers
3. Duplication Management

```
Final dataset shape: (700, 11)
```

	User ID	Device Model	Operating System	App Usage Time (min/day)	Screen On Time (hours/day)	Battery Drain (mAh/day)	Number of Apps Installed	Data Usage (MB/day)	Age	Gender	User Behavior Class
0	1	Google Pixel 5	Android	393	6.4	1872	67	1122	40	Male	4
1	2	OnePlus 9	Android	268	4.7	1331	42	944	47	Female	3
2	3	Xiaomi Mi 11	Android	154	4.0	761	32	322	42	Male	2
3	4	Google Pixel 5	Android	239	4.8	1676	56	871	20	Male	3
4	5	iPhone 12	iOS	187	4.3	1367	58	988	31	Female	3

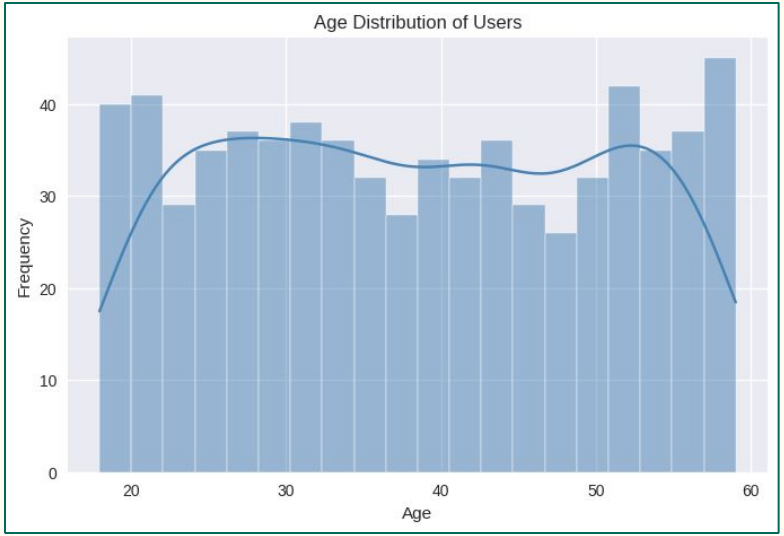
Exploratory Data Analysis

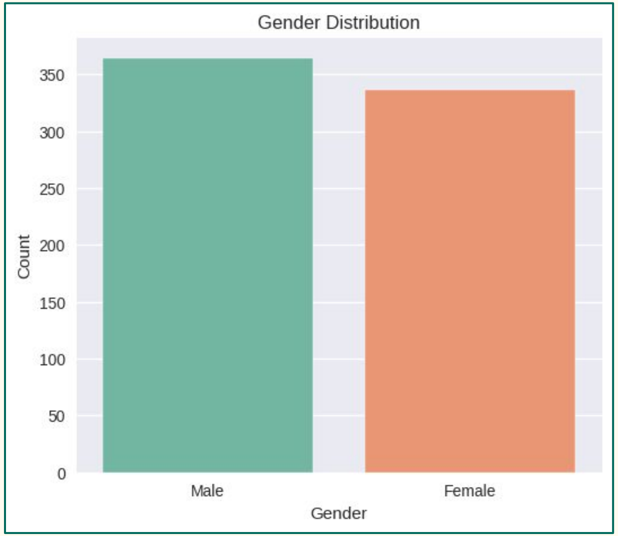
Exploring data patterns before building models

“Demographic” Metrics (Summary Statistics)

- The distribution appears nearly uniform within the 18 to 60 range, with fewer users beyond 60.

- The gender split is nearly balanced, with a slight male majority.





“Demographic” Metrics (Summary Statistics)

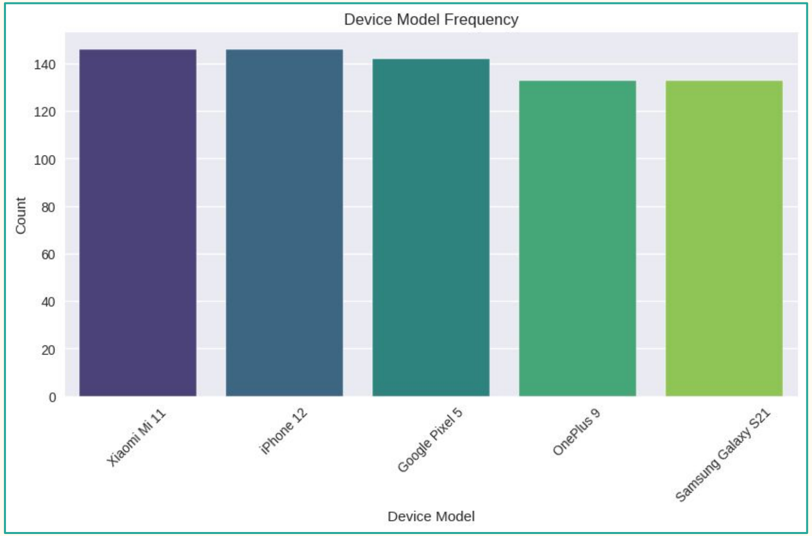
- Screen-on time is fairly consistent across age groups — the regression line is almost flat.

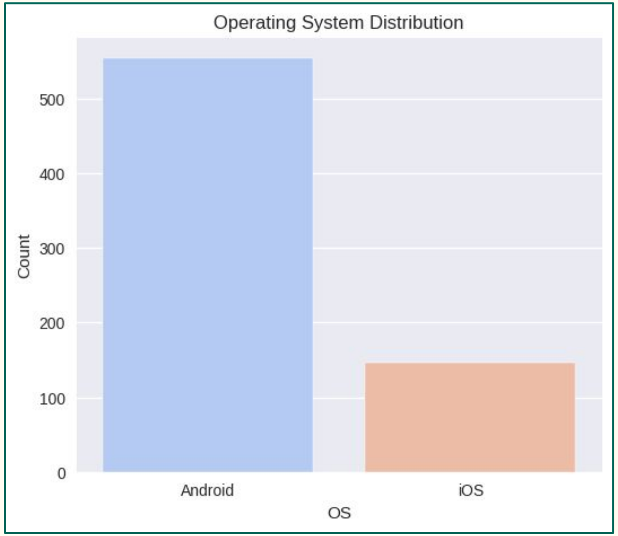
- Both genders have similar variability, ranging broadly from 2–8 hours/day.

“Device OS” Metrics (Summary Statistics)

- The dataset is well balanced across the five device models.

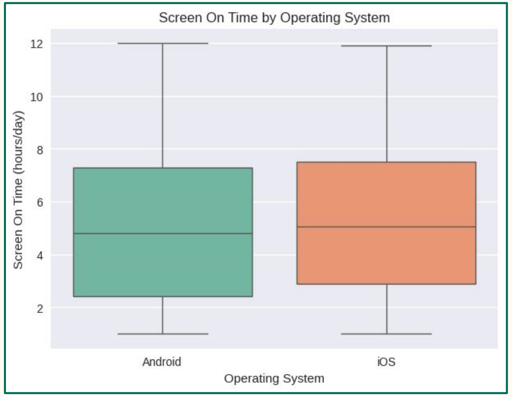
- Dataset is heavily Android-skewed (~79% Android, ~21% iOS).

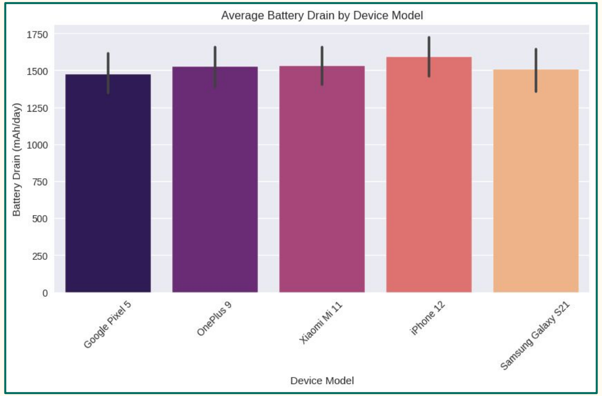


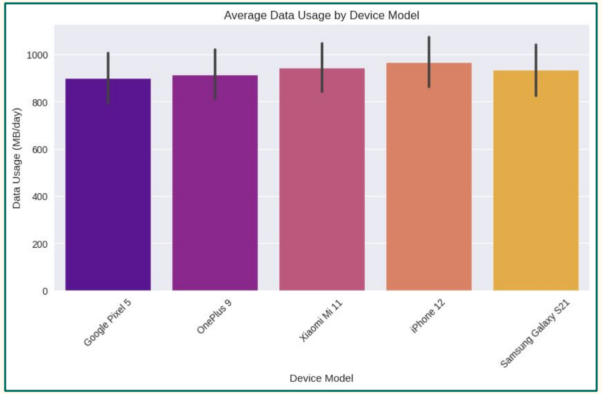


“Device OS” Metrics (Summary Statistics)

- iOS users have slightly higher median screen-on time than Android.
- iPhone 12 shows the highest average battery drain, followed by Xiaomi Mi 11.
- iPhone 12 again leads with the highest average data usage, followed by Xiaomi Mi 11.







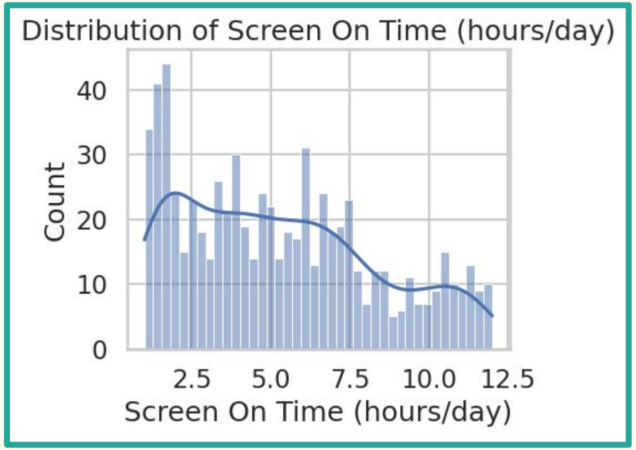
“Usage” Metrics (Summary Statistics)

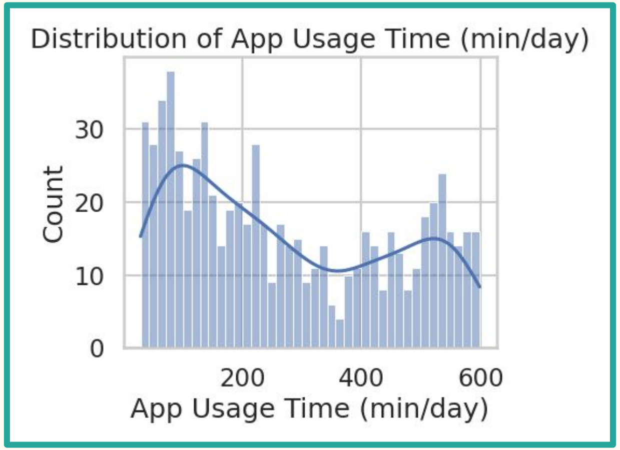
- Usage varies dramatically across users (very large IQRs)

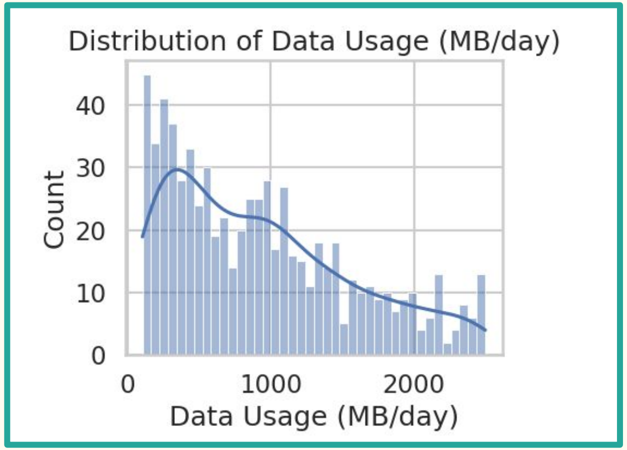
- All metrics show some right skew (we can already tell there's some heavy users)

Feature	Mean	IQR	Skew
Screen On Time	5.3 hrs/day	4.9 hrs	0.46
App Usage Time	271 min/day	321 min	0.37
Battery Drain	1525 mAh/day	1507 mAh	0.13
Apps Installed	50 apps	48	0.11
Data Usage	930 MB/day	968 MB	0.70

Usage Distributions (Histograms)

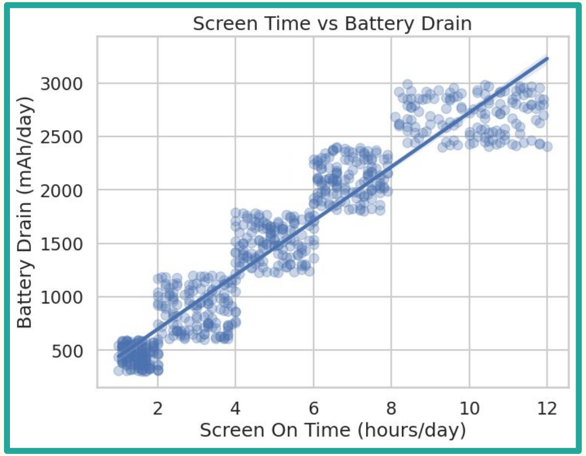


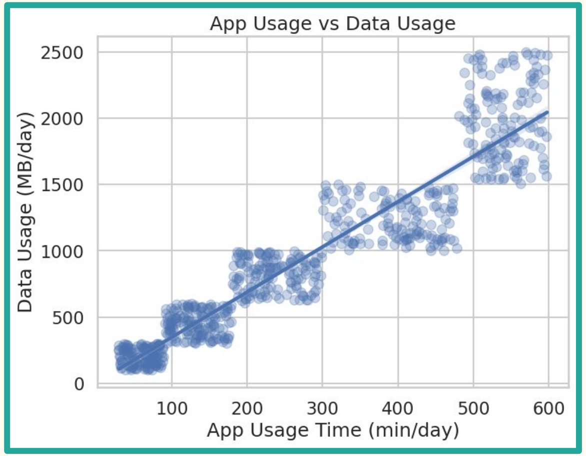




- Visualizations confirm the hunches we had from the summary statistics!
- All three distributions show clear right tails
- Majority of users fall into moderate usage ranges (expected)
- Small group of heavy users stands out in each distribution

Feature Relationships

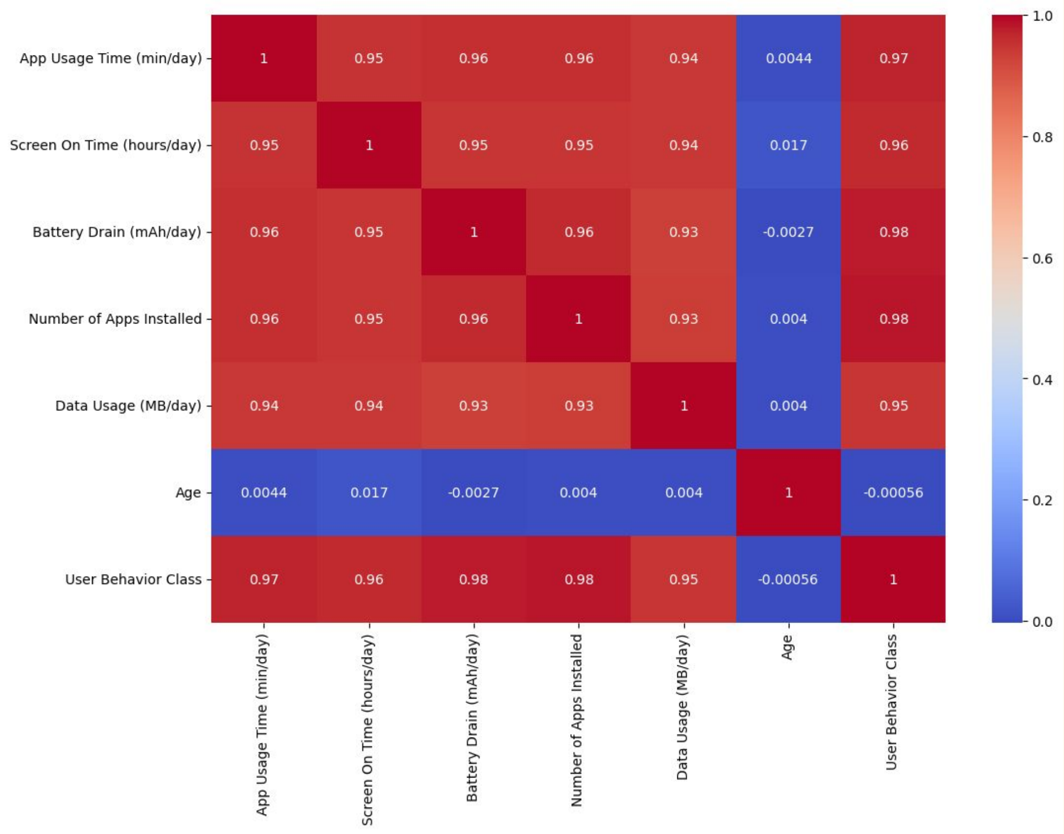




- Two relationships show extremely strong linear behavior (0.94–0.95 correlation score)
  - Gives us a look into what variables would be useful (or redundant in prediction)
- These patterns highlight behavioral drivers: screen time, data usage, app engagement
- Age shows no meaningful relationship with app count (surprising!)

Correlation Heat Map

- Strong correlations among App Usage Time, Screen On time, Battery Drain, Apps Installed, User Behavior Class and Data Usage.
- These is probably due to them all being related to device activity.
- Age shows near zero correlation between other numerical variables.



Persona Summary

User Behavior Class	App Usage Time (min/day)	Screen On Time (hours/day)	Battery Drain (mAh/day)	Number of Apps Installed	Data Usage (MB/day)	Age
1	60.426471	1.490441	454.977941	14.558824	202.323529	38.213235
2	131.972603	3.037671	883.808219	30.753425	451.417808	38.643836
3	235.398601	4.955944	1515.055944	50.000000	822.013986	38.678322
4	395.748201	6.909353	2105.805755	69.920863	1232.230216	38.676259
5	541.419118	10.114706	2701.014706	89.250000	1974.772059	38.176471

- All variables increases as we go up the User Behavior Class.
- Age remains constant around 38 confirming that age doesn't help in identifying at risk user.

Machine Learning based analysis of the Dataset

Why Machine Learning?

Machine Learning helps uncover hidden behavioral patterns that cannot be seen with simple statistics and helps us go more in depth to support EDA's findings.

In our dataset, we thought ML could be useful due to the following reasons:

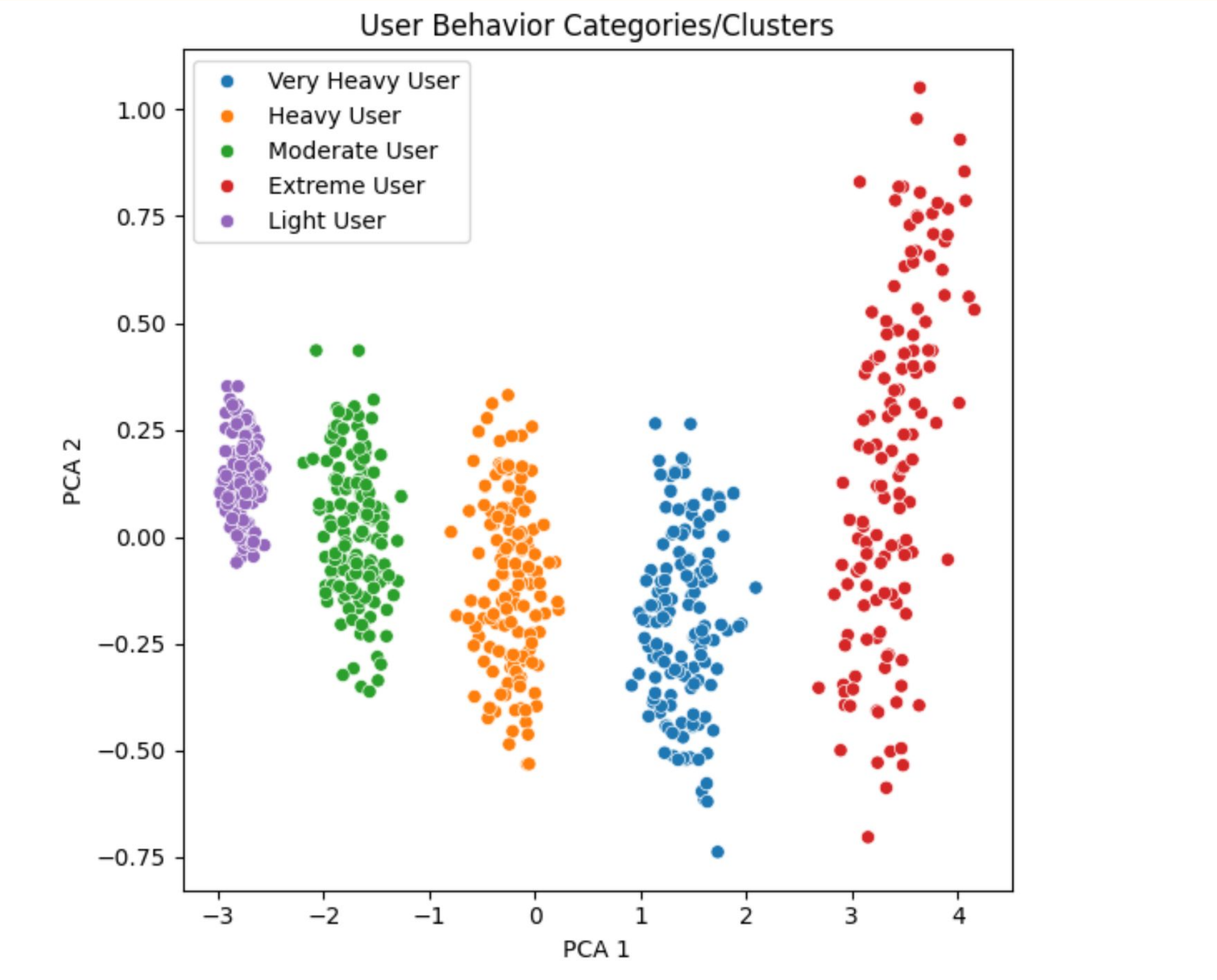
1) Finds non-obvious relationships (e.g., screen time → exponential battery drain)
2) Identifies distinct user groups automatically using clustering
3) Regression quantifies which features matter most and support the Exploratory Data Analysis

Clustering - Determine Distinct user segments

Use K Means Clustering to identify patterns in the dataset based on the “amount of usage”.

Classified into 5 categories:

1) Light
2) Moderate
3) Heavy
4) Very heavy
5) Extreme



Helps us answer these questions:

1)How people use their mobile phones

2) Which users are at high risk of excessive usage?

3) Are there distinct user segments or clusters?

Regression - Common heavy usage factors

Identified that **Battery Drain** and **Data Usage** are two of the most commonly usage heavy factors in a mobile device.

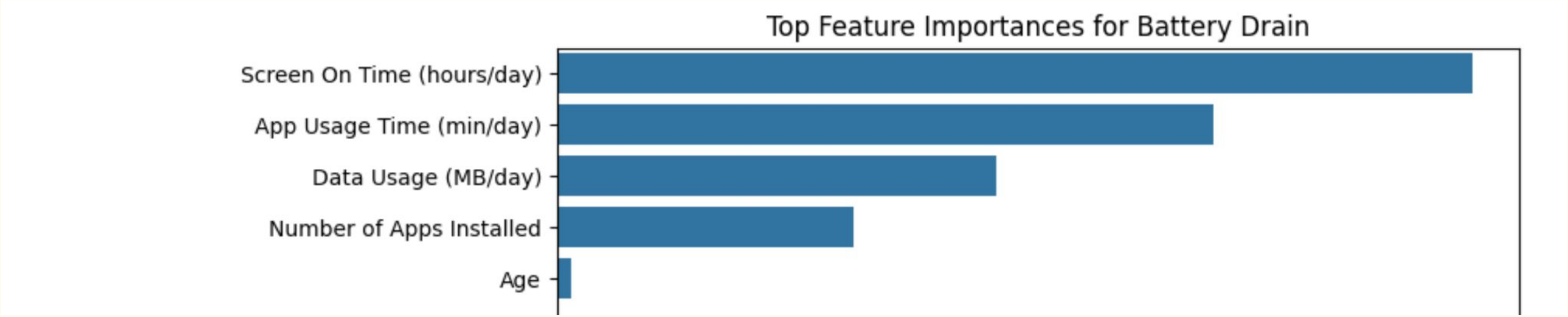
1. Battery Drain

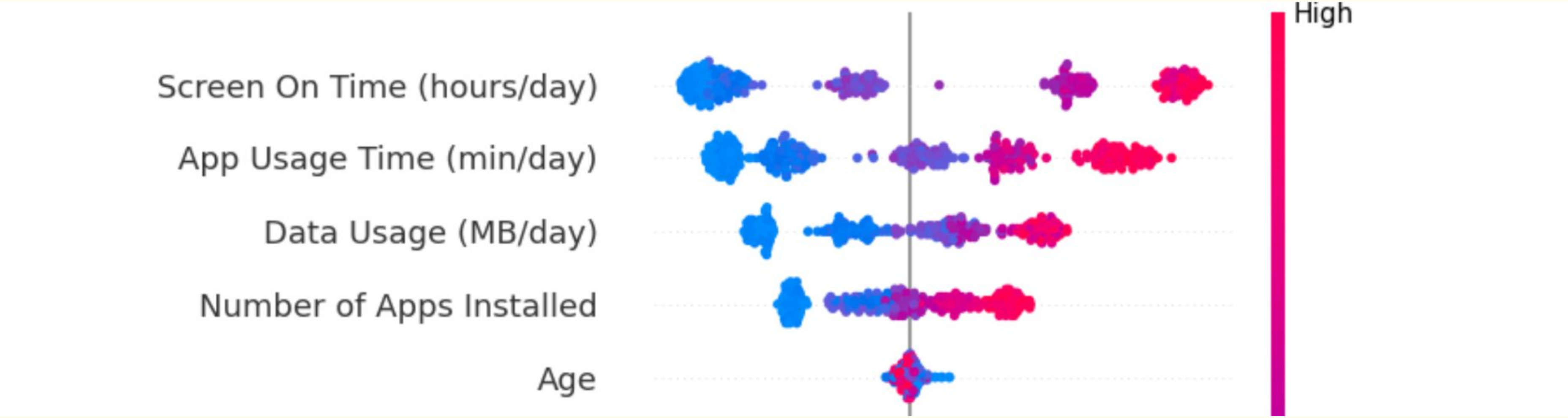
Acts as a proxy for active device usage. By using **random forests** here to predict important features affecting battery drain, we can answer heavy usage factors. Battery drain has been chosen since it is an amazing factor and one of the most intuitive factors for active device engagement.

Top predictors for battery drain:

1) Screen time
2) App Usage time
3) Number of apps installed
4) Data Usage

Screen On Time (hours/day)	39.214982
App Usage Time (min/day)	28.136952
Data Usage (MB/day)	18.829064
Number of Apps Installed	12.726133
Age	0.615701
Device Model_Xiaomi Mi 11	0.072423
Device Model_OnePlus 9	0.069966
Device Model_Samsung Galaxy S21	0.069887
Device Model_Google Pixel 5	0.069214
Gender_Male	0.055215
Gender_Female	0.052818
Operating System_iOS	0.030212
Device Model_iPhone 12	0.028742
Operating System_Android	0.028691


Top Feature Importances for Battery Drain
Screen On Time (hours/day)
App Usage Time (min/day)
Data Usage (MB/day)
Number of Apps Installed
Age


High
Screen On Time (hours/day)
App Usage Time (min/day)
Data Usage (MB/day)
Number of Apps Installed
Age

2. Data Usage

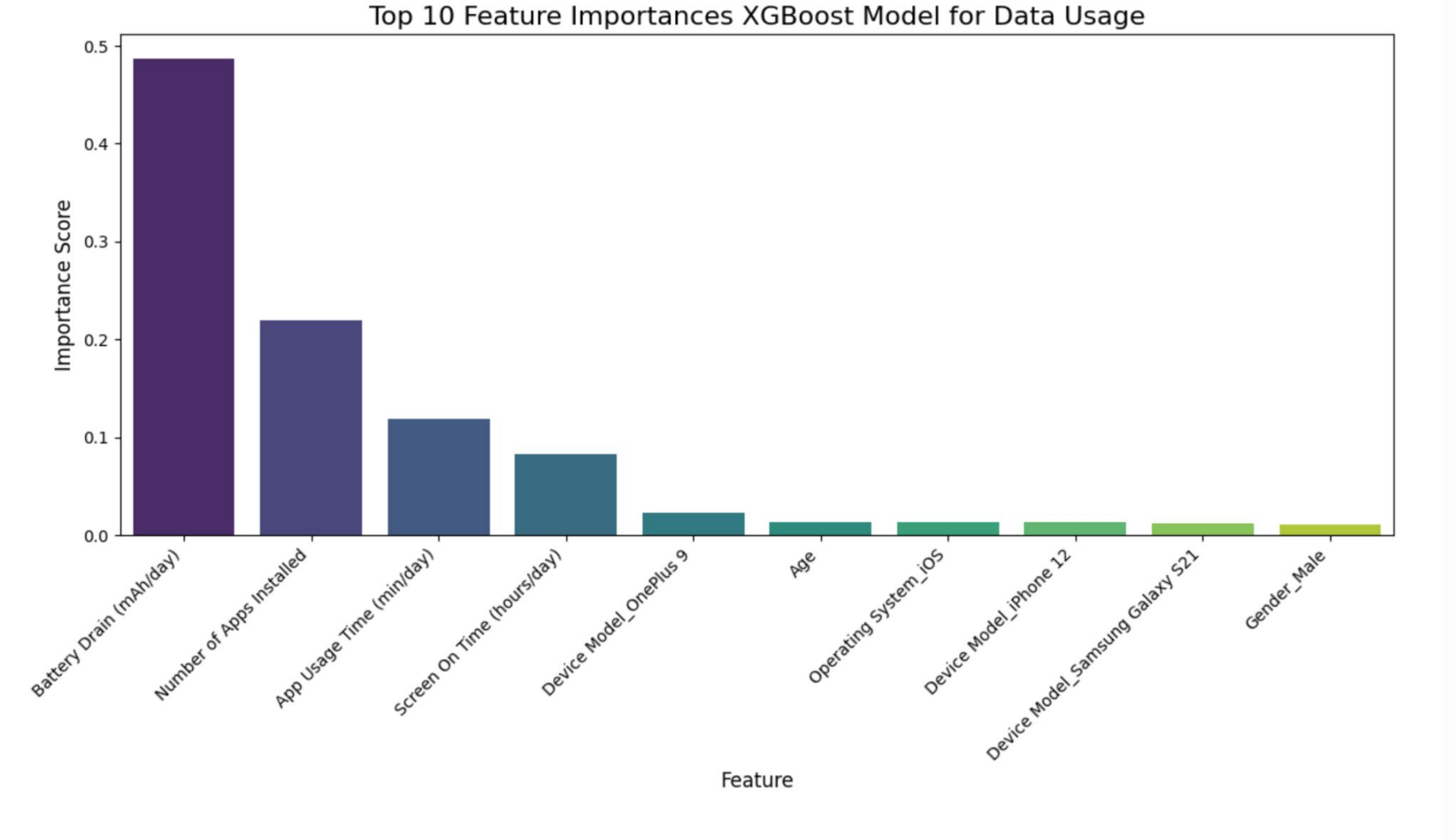
Another amazon insight we found for predicting heavy usage is data usage. By using **xgboost regressor,** we can understand data usage and its affecting factors where we found good insights.

Top predictors for data usage:

1) Battery drain (Conversely for battery drain, data usage was **NOT** the most important feature based on random forest!!)
2) Number of apps installed
3) App usage time

Feature Importances from Updated XGBoost Model

	Feature	Importance
2	Battery Drain (mAh/day)	0.487248
3	Number of Apps Installed	0.219907
0	App Usage Time (min/day)	0.118666
1	Screen On Time (hours/day)	0.082856
5	Device Model_OnePlus 9	0.022458
4	Age	0.013348
9	Operating System_iOS	0.013066
8	Device Model_iPhone 12	0.012637
6	Device Model_Samsung Galaxy S21	0.011850
10	Gender_Male	0.010454



Top 10 Feature Importances XGBoost Model for Data Usage
Importance Score
0.0
0.1
0.2
0.3
0.4
0.5
Battery Drain (mAh/day)
Number of Apps Installed
App Usage Time (min/day)
Screen On Time (hours/day)
Device Model_OnePlus 9
Age
Operating System_iOS
Device Model_iPhone 12
Device Model_Samsung Galaxy S21
Gender_Male
Feature

Regression - Findings and Insights

1) Regression helped us answer the question ("What factors are most predictive of heavy usage?")

Most important features to determine which actively predict heavy usage (Intuitively sounds correct as well!):

1) Screen-on time → strongest.
2) App usage time → next strongest.

Battery drain and data usage → outcome signals.

By studying these features, we can identify what actually affect heavy usage. These can be used to build applications such as **personalized device optimization, battery management recommendations and digital well-being applications** which are present in today's mobile phones and share the similar factors which they use to determine heavy usage.

Conclusion

In conclusion, we analyzed the data using exploratory data analysis and also machine learning and found amazing insights from our data. We were able to answer questions which we wanted to get from the dataset like:

How do people use their mobile devices?

Which users are at highest risk of excessive usage?

What factors are most predictive of heavy usage?

Are there distinct user segments or clusters? (ex: heavy, normal, light users)

This study helped us understand more in depth about using our python skills to tackle real life data sets