BEHAVIORAL SEGMENTATION OF FITNESS APP USERS IN THE U.S. USING R

1. Background

A U.S.-based fitness technology startup operating a mobile app for personalized workout plans and calorie tracking wanted to improve content personalization and reduce drop-offs. Despite collecting rich behavioral data, the company lacked actionable segmentation. Our objective was to segment users using behavioral clustering in R, focusing on engagement patterns, fitness goals, and app usage frequency.

2. Objective

- To segment app users into distinct behavioral profiles using unsupervised learning in R

- To help personalize fitness plans and notification flows for each group

- To generate input for future recommendation system improvements

3. Data Summary

- **Source:** App database export (PostgreSQL to CSV)

- **Volume:** 22,478 user records from Jan–Oct 2023

- **Variables:**

    o Demographics: Age, Gender, City, Fitness Goal (e.g., weight loss, strength)

    o Usage: App Opens per Week, Sessions per Day, Avg Workout Time (min)

    o Engagement: Calories Logged, Days Active per Week, Completion Rate

    o Subscription Info: Plan Type (Free, Premium), Renewal Count

4. Methodology

4.1 Data Preparation

- Cleaned missing records using tidyr::drop_na()

- Normalized usage variables using scale() to prepare for PCA

- Converted categorical variables using one-hot encoding (caret::dummyVars)

- Removed outliers (extreme logins or 0 activity) using IQR filtering

4.2 Feature Engineering

- Created derived metrics:

    o   Consistency Score = Active Days / Total Weeks

    o   Workout Intensity Index = Avg Workout Time × Completion Rate

    o   Engagement Level = Composite of Opens, Sessions, and Calories Logged

4.3 Dimensionality Reduction and Clustering

- Applied Principal Component Analysis to reduce 12 features to 3 PCs

- Identified optimal cluster count as **4** using the Elbow and Silhouette methods

- Performed clustering using kmeans() and validated results with cluster::silhouette and factoextra::fviz_cluster

5. Segment Profiles

Cluster	Share	Behavior Summary	Recommended Action
C1	34%	High usage, strong goal alignment, high consistency	Promote premium upgrades and coaching
C2	26%	New users, low engagement, inconsistent activity	Trigger onboarding reminders and gamify
C3	21%	Weight loss-focused, logs meals but skips workouts	Push short workout routines and rewards
C4	19%	Free plan users, browse often, never commit	Offer trial extension or limited-time plan

6. Results

- Conversion from free to premium increased by **9.5%** in Cluster C1

- Drop-off in week 2 of onboarding reduced by **13.2%** in Cluster C2 after journey personalization

- Personalized workout suggestions improved plan adherence by **18%** for C3

- Re-engagement email CTR rose by **22%** for C4 using personalized copy

7. Deliverables

- Final dataset with cluster labels and user engagement score

- Interactive segmentation dashboard using R Shiny (hosted on internal server)

- Comprehensive R Markdown report documenting all stages of the analysis

- Exportable CSVs for marketing automation integration with HubSpot

- Documentation for rerunning the clustering workflow on updated data

8. Tools and Packages Used

- dplyr, tidyr, ggplot2, cluster, factoextra, shiny, caret, psych

9. Future Scope

- Integration of real-time clustering into app backend using plumber API in R

- Overlay clustering with sentiment from in-app feedback and reviews

- Develop an ML-based recommendation engine based on cluster behavior