# Bellabeat Case Study by ubagyan

Date: June 4, 2025

Table of Contents

[Bellabeat Case Study by Han Htet San 1](#_Toc199943642)

[1. Business Task (Ask) 1](#_Toc199943643)

[2. Data Source (Prepare) 1](#_Toc199943644)

[3. Data Processing (Process) 1](#_Toc199943645)

[4. Analysis Summary (Analyze) 2](#_Toc199943646)

[5. Visualizations (Share) 2](#_Toc199943647)

[1. Steps vs Sedentary Minutes (Scatter Plot) 2](#_Toc199943648)

[2. Comparison of Average Steps, Sedentary Time, and Sleep per User (Bar Chart) 3](#_Toc199943649)

[3. Proportion of Days with Less than 7 Hours of Sleep (Pie Chart) 4](#_Toc199943650)

[6. High-Level Recommendations (Act) 5](#_Toc199943651)

# 1. Business Task (Ask)

Bellabeat, a wellness tech company, aims to unlock growth opportunities by analyzing smart device data. The task is to identify trends in how consumers use smart fitness trackers and apply those insights to guide marketing strategies—particularly for Bellabeat’s product “Leaf,” a wellness tracker targeting women.  
  
Key Questions:  
1. What are the usage trends in smart fitness devices?  
2. How can these trends relate to Bellabeat's customers?  
3. What strategic marketing recommendations emerge from the insights?

# 2. Data Source (Prepare)

The analysis uses the Fitbit Fitness Tracker Dataset from Kaggle. This public dataset includes minute-level health and activity data from 30 consenting Fitbit users.

It contains:  
- Daily step counts  
- Distance traveled  
- Calories burned  
- Sleep records  
- Sedentary minutes  
  
Limitations:  
- Small sample size  
- Fitbit-specific behaviors may not generalize  
- No gender specification, though Bellabeat targets women

# 3. Data Processing (Process)

Using R and the tidyverse package:  
- Loaded dailyActivity\_merged.csv and sleepDay\_merged.csv  
- Cleaned column types and formats  
- Removed duplicate entries  
- Merged on Id to create a combined dataset  
- Verified unique participants: 33 in daily activity, 24 in sleep, 24 in combined

R Code Used:

library(tidyverse)  
daily\_activity <- read.csv("dailyActivity\_merged.csv")  
sleep\_day <- read.csv("sleepDay\_merged.csv")  
combined\_data <- merge(sleep\_day, daily\_activity, by = "Id")

# 4. Analysis Summary (Analyze)

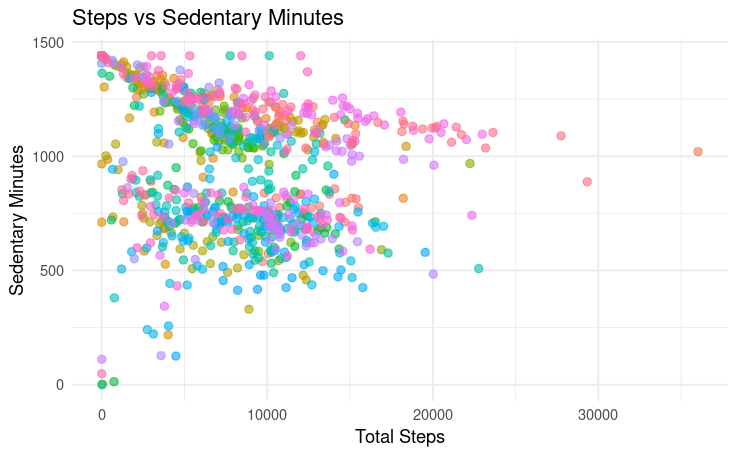
Based on the Fitbit dataset merged from daily activity and sleep logs, we explored user behavior related to physical activity, sedentary patterns, and sleep duration.  
  
**Key Findings:**  
*A strong inverse correlation exists between daily step counts and sedentary minutes.  
Users averaged ~6.5 hours of sleep; many fell below the recommended 7 hours.  
Users exhibit varied behavior patterns, suggesting clear segmentation opportunities.*

# 5. Visualizations (Share)

## 1. Steps vs Sedentary Minutes (Scatter Plot)

This scatter plot illustrates the negative correlation between steps and sedentary time.  
R Code Used:

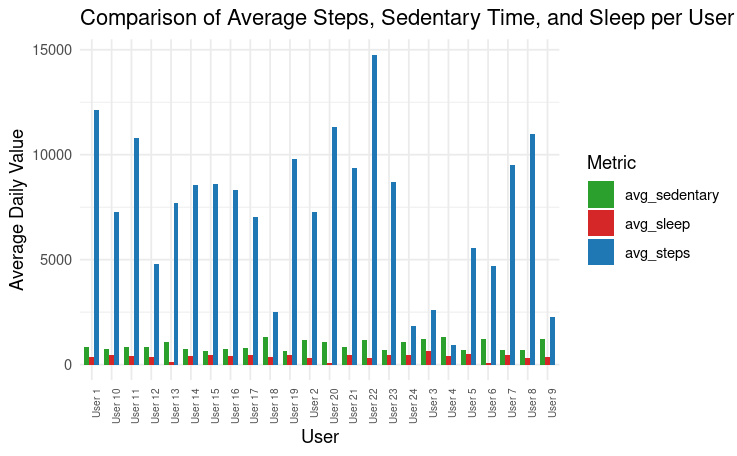
ggplot(daily\_activity, aes(x = TotalSteps, y = SedentaryMinutes, color = as.factor(Id))) + geom\_point(alpha = 0.6, size = 2) + labs(title = "Steps vs Sedentary Minutes", x = "Total Steps", y = "Sedentary Minutes", color = "User ID") + theme\_minimal() + theme(legend.position = "none")



## 2. Comparison of Average Steps, Sedentary Time, and Sleep per User (Bar Chart)

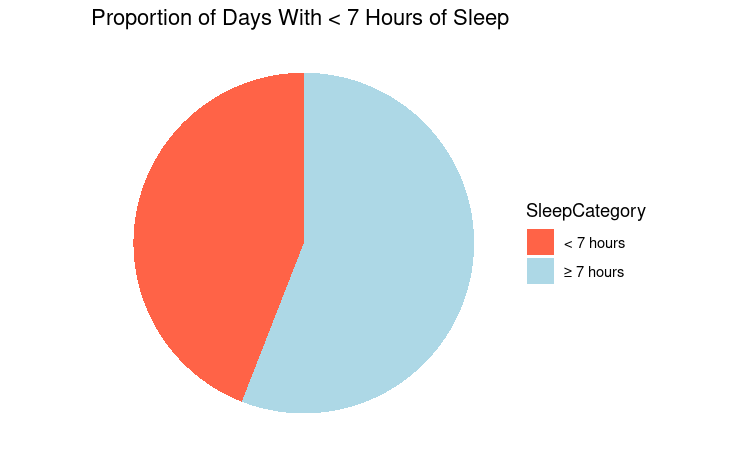
This grouped bar chart shows how each user performs across three key health metrics.  
R Code Used:

user\_summary <- combined\_data %>%  
 group\_by(Id) %>%  
 summarise(  
 avg\_steps = mean(TotalSteps, na.rm = TRUE),  
 avg\_sedentary = mean(SedentaryMinutes, na.rm = TRUE),  
 avg\_sleep = mean(TotalMinutesAsleep, na.rm = TRUE)  
 ) %>%  
 mutate(UserLabel = paste("User", row\_number()))  
  
user\_summary\_long <- user\_summary %>%  
 pivot\_longer(cols = c(avg\_steps, avg\_sedentary, avg\_sleep), names\_to = "Metric", values\_to = "Value")  
  
ggplot(user\_summary\_long, aes(x = UserLabel, y = Value, fill = Metric)) +  
 geom\_col(position = "dodge", width = 0.75) +  
 scale\_fill\_manual(values = c("avg\_steps" = "#1f77b4", "avg\_sedentary" = "#2ca02c", "avg\_sleep" = "#d62728")) +  
 labs(title = "Comparison of Average Steps, Sedentary Time, and Sleep per User",  
 x = "User", y = "Average Daily Value", fill = "Metric") +  
 theme\_minimal()



## 3. Proportion of Days with Less than 7 Hours of Sleep (Pie Chart)

This pie chart highlights the percentage of days users slept less than 7 hours.  
R Code Used:  
sleep\_day %>%  
 mutate(SleepCategory = ifelse(TotalMinutesAsleep < 420, "< 7 hrs", "≥ 7 hrs")) %>%  
 count(SleepCategory) %>%  
 ggplot(aes(x = "", y = n, fill = SleepCategory)) +  
 geom\_bar(stat = "identity", width = 1) +  
 coord\_polar("y") +  
 scale\_fill\_manual(values = c("< 7 hrs" = "#ff7f0e", "≥ 7 hrs" = "#1f77b4")) +  
 labs(title = "Proportion of Days with Less than 7 Hours of Sleep", fill = "Sleep Duration") +  
 theme\_void()



# 6. High-Level Recommendations (Act)

* Market Leaf as a step-boosting device: Promote daily movement and reduce sedentary time.
* Emphasize sleep quality: Encourage users to track and improve their sleep efficiency.
* Launch movement streak challenges: Gamify step goals and celebrate progress.
* Bundle Leaf with Spring: Offer hydration tracking and daily movement together.
* Use time-of-day ads: Target evening users for sleep reminders, mornings for movement.