Bellabeat Case Study: Data Analysis for Marketing Strategy

Mohammad Mohidul Islam | Junior Data Analyst

2025-07-14

Introduction

Urška Sršen and Sando Mur founded Bellabeat, a high-tech company that manufactures health-focused smart products. Sršen used her background as an artist to develop beautifully designed technology that informs and inspires women around the world. Collecting data on activity, sleep, stress, and reproductive health has allowed Bellabeat to empower women with knowledge about their own health and habits. Since it was founded in 2013, Bellabeat has grown rapidly and quickly positioned itself as a tech-driven wellness company for women. Bellabeat is a successful small company, but they have the potential to become a larger player in the global smart device market. Urška Sršen, cofounder and Chief Creative Officer of Bellabeat, believes that analyzing smart device fitness data could help unlock new growth opportunities for the company.

Products

- ***Bellabeat app:*** The Bellabeat app provides users with health data related to their activity, sleep, stress, menstrual cycle, and mindfulness habits.
- ***Leaf:*** Bellabeat's classic wellness tracker can be worn as a bracelet, necklace, or clip.
- ***Time:*** This wellness watch combines the timeless look of a classic timepiece with smart technology to track user activity, sleep, and stress.
- ***Spring:*** This is a water bottle that tracks daily water intake using smart technology to ensure that you are appropriately hydrated throughout the day.
- ***Bellabeat membership:*** Bellabeat also offers a subscription-based membership program for users.

Summary of the Business Task: As a junior data analyst working on the marketing analyst team at Bellabeat, I have been asked to focus on one of Bellabeat's products and analyze smart device data to gain insight into how consumers are using their smart devices. The insights I discover will then help guide the marketing strategy for the company. I will present my analysis to the Bellabeat executive team along with my high-level recommendations for Bellabeat's marketing strategy.

Primary objective: My primary objective is to help Bellabeat to analyze smart device usage data to uncover trends and provide actionable insights for marketing strategy.

Scope: The analysis will leverage available smart device data to identify patterns in usage, preferences, and engagement. Key areas of focus will include activity tracking, user habits, and feature adoption. These insights will then be evaluated to determine their potential impact on Bellabeat's marketing initiatives, including customer segmentation, personalized messaging, and product positioning.

Deliverables:

- ***Data-Driven Insights:*** Identification of trends in smart device usage relevant to Bellabeat's target audience.
- ***Strategic Recommendations:*** Actionable marketing strategies based on user behavior analysis to improve engagement and retention.
- ***Presentation of Findings:*** A clear, data-supported report outlining opportunities for Bellabeat to optimize its marketing approach.

Stage 1: Ask (Define the Problem)

Key Questions: According to the business task, below questions will guide my analysis:

- What key trends can be identified in smart device usage patterns?
- In what ways might these trends be relevant to Bellabeat's customer base?
- How can these insights be leveraged to shape and enhance Bellabeat's marketing strategy?

Key stakeholders:

- Urška Sršen (Co-founder & Chief Creative Officer)
- Sando Mur (Mathematician & Co-founder)
- Bellabeat Marketing Team

Step 2: Prepare (Data Collection & Selection)

Dataset Used: For this project I used FitBit Fitness Tracker Data (Kaggle). which is publicly available and License: CCO (Publice domain). I choose the product Leaf: Bellabeat's classic wellness tracker can be worn as a bracelet, necklace, or clip.

- Contains daily activity, sleep, heart rate, and steps from 30 users.

- Files to focus on: There were 18 datasets. After reviewing all of them, I have decided to move forward with the followings as in my opnion those can serve to fulfill my objectives.

  - dailyActivity_merged.csv (steps, calories, activity levels)
  - sleepDay_merged.csv (sleep patterns)
  - hourlySteps_merged.csv (activity trends by hour)

Dataset Review Process: After initial review with Excel, Spreadsheets & RStudio I found the data is being organized in long format and somewhere id repeatative. Moreover, I used ROCCC methods to verify the data as follows,

- Reliable: Not reliable, because the sample size is too small.
- Original: Third party data survey via Amazon Mechanical Turk.
- Comprehensive:No demographic consideration of participants of the survey.
- Current: Data published on 2016 and not updated later on.
- Cited: Datasets are originally comes from Link.

Limitations:

- Small sample size (only 30 users).
- Data is from 2016 (may be outdated).

Step 3: Process (Data Cleaning & Transformation)

Setting up my environment: Setting up my environment by installing and loading packages of 'tidyverse', 'readr', and 'ggplot2'. Moreover, loading all the CSV data file to RStudio.

Tools Used:
RStudio (for analysis)

Excel: for quick checks
PowerPoint: for final presentation

Timeline:
Day 1-2: Data cleaning & merging

Day 3-4: Analysis & visualizations

Day 5: Report & recommendations

Loading Databases in R Environment:

```r
# Load raw data
daily_activity_merged <- read_csv("dailyActivity_merged.csv")
hourly_steps_merged <- read_csv("hourlySteps_merged.csv")
sleep_day_merged <- read_csv("sleepDay_merged.csv")
```

Check for Missing values: Here, I have checked with the following functions of data for missing and duplicate check, and therefore no missing found. sum(is.na(daily_activity_merged))
sum(is.na(hourly_steps_merged))
sum(is.na(sleep_day_merged))

To remove duplicate: Normally, If I found any missing value, either I remove that data or fill with mean or median, based on sample.

```r
# Remove duplicates
daily_activity <- distinct(daily_activity_merged)
hourly_steps <- distinct(hourly_steps_merged)
sleep_day <- distinct(sleep_day_merged)
```

Format Dates-Times:

```r
# Date formatting
daily_activity <- daily_activity %>%
  mutate(ActivityDate = as.Date(ActivityDate, format= "%m/%d/%Y"))

sleep_day <- sleep_day %>%
  mutate(SleepDay = as.Date(SleepDay, format = "%m/%d/%Y"))

hourly_steps <- hourly_steps %>%
  mutate(ActivityHour = as.POSIXct(ActivityHour, format = "%m/%d/%Y %I:%M:%S %p", tz = "UTC"))
```

Data Merging: Here I combine daily_activity and sleep_day using one key column i.e. the User ID. This merged data help me to find the correlation between activity and sleep.

```r
# Merge data
merged_data <- inner_join(daily_activity, sleep_day,
                          by = c("Id" = "Id", "ActivityDate" = "SleepDay"))
```

Adding 2 useful Columns:

```r
# Add calculated columns
merged_data <- merged_data %>%
  mutate(TotalActiveMinutes = VeryActiveMinutes + FairlyActiveMinutes + LightlyActiveMinutes,
         sleepEfficiency = TotalMinutesAsleep / TotalTimeInBed)
```

```r
# Export cleaned data
write_csv(merged_data, "bellabeat_merged_clean.csv")
```

Step 4: Analyze (Find Trends & Patterns)

```r
# Average daily steps and calories burned
mean(merged_data$TotalSteps)
```

```
## [1] 8514.91
```

```r
mean(merged_data$Calories)
```

```
## [1] 2389.295
```

```r
# Compare to recommended 10,000 steps
daily_activity %>%
  summarise(PercentMeetingGoal = mean(TotalSteps >= 10000)*100)
```

```
## # A tibble: 1 x 1
##   PercentMeetingGoal
##                <dbl>
## 1               32.2
```

```r
# Average sleep time
mean(sleep_day$TotalMinutesAsleep)/60
```

```
## [1] 6.98622
```

```r
# Activity analysis
activity_summary <- merged_data %>%
  summarise(
    AvgTotalSteps = mean(TotalSteps),
    AvgActiveMinutes = mean(TotalActiveMinutes),
    UserMeetActiveGoal = mean(TotalSteps >= 10000)*100
  )

head(activity_summary)
```

```
## # A tibble: 1 x 3
##   AvgTotalSteps AvgActiveMinutes UserMeetActiveGoal
##           <dbl>            <dbl>              <dbl>
## 1         8515.             260.               40.2
```

```r
# Sleep analysis
sleep_summary <- merged_data %>%
  summarise(
    AvgSleepHours = mean(TotalMinutesAsleep)/60,
    AvgSleepEfficiency = mean(sleepEfficiency)*100
  )

head(sleep_summary)
```

```
## # A tibble: 1 x 2
##   AvgSleepHours AvgSleepEfficiency
##           <dbl>              <dbl>
## 1          6.99               91.6
```

Visualization

1. Activity vs Sleep visualization:
To Identify the correlation between active minutes per day and Minutes asleep per day.

```
# Visualization
ggplot(data = merged_data) +
  geom_point(mapping = aes(x=TotalActiveMinutes, y=TotalMinutesAsleep)) +
  geom_smooth(mapping = aes(x=TotalActiveMinutes, y=TotalMinutesAsleep)) +
  labs(title = "Activity vs Sleep",
       x = "Active Minutes per day",
       y = "Minutes Asleep per day") +
  theme_minimal()
```
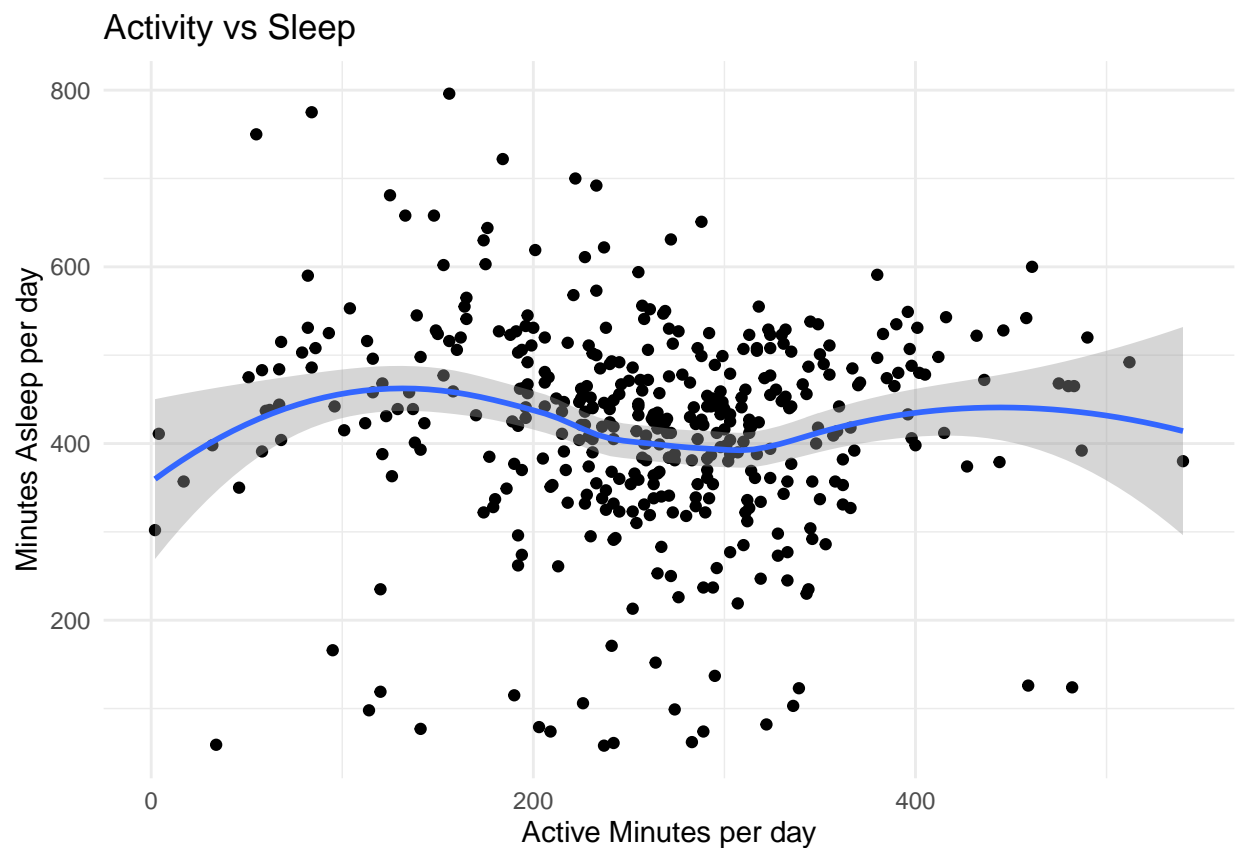

Marketing Insights: There is a positive correlation found, Bellabeat can prompt a message that "More daily activity can improve sleep duration or quality

2. Sleep Efficiency distribution: To understand how efficient users' sleep

```
# Sleep efficiency
ggplot(merged_data, aes(x = sleepEfficiency)) +
  geom_histogram(fill = "lightblue", alpha = 0.7) +
  labs(title = "Distribution of Sleep Efficiency",
       x = "Sleep Efficiency",
       y = "Frequency") +
  theme_minimal()
```

Distribution of Sleep Efficiency
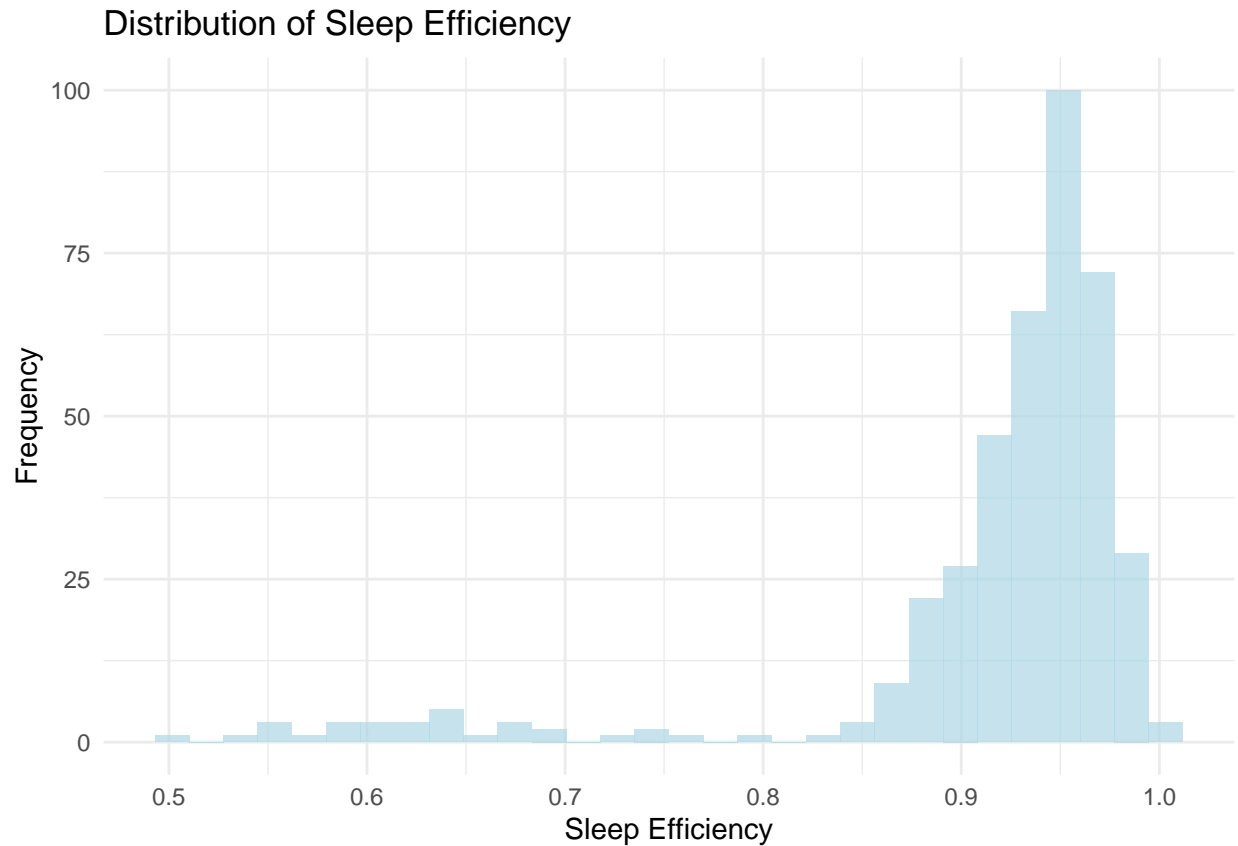
Marketing Insights: If many users has low sleep efficiency, Bellabeat can support for better sleep through guided medication for better sleep.

3. Active Minutes vs Calories Burned: To demonstrate that more active minutes will lead the users get more calories burnt.

```
# Calories burn
ggplot(merged_data, aes(x = TotalActiveMinutes, y = Calories)) +
  geom_point(color = "purple") +
  geom_smooth(method = "lm") +
  labs(title = "Active Minutes vs Calories Burned",
       x = "Total Active Minutes",
       y = "Calories Burned") +
  theme_minimal()
```

Active Minutes vs Calories Burned
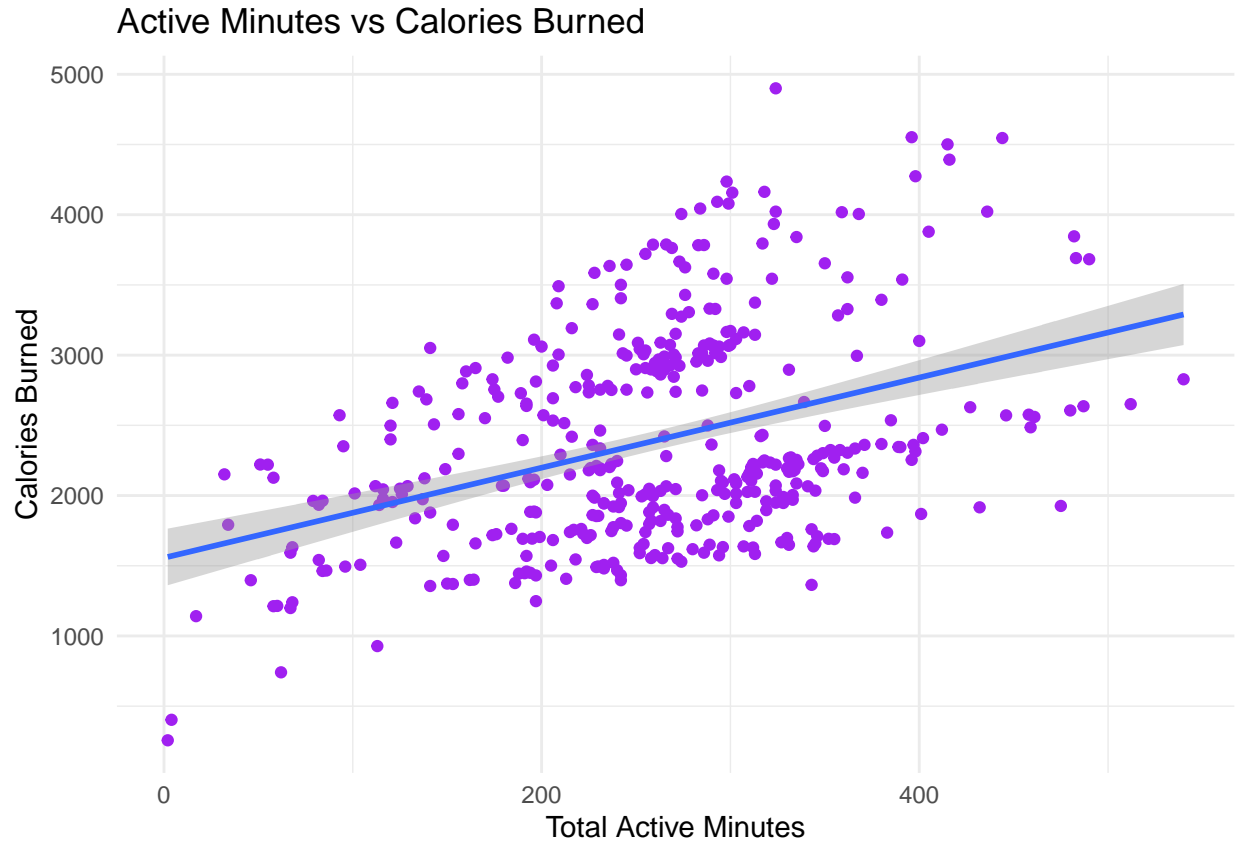
Marketing Insights: this can motivate the user that with activity challenge in the app, users can burn more calories with activity. Marketers can engage the users with activity challenge.

Step 5: Share & Act (Key Insights and Recommendations)

Findings:

- Users average around 8,500 steps/day, burning approximately 2,300 calories/day. More steps correspond to more calories burned

- Only 32% Users sleep around 7 hours on average. Longer and more consistent sleep is often seen in users with higher daily step counts.

- Users are most active at 6-7 PM.

- There is a positive but modest correlation between daily steps and sleep duration, suggesting active users tend to sleep slightly better.

Recommendations:

Based on the data analysis of smart device usage, I recommend the marketing team focus on positioning the Bellabeat Leaf as a holistic wellness tracker that connects daily activity with improved sleep quality.

Key opportunities include:

- Promoting morning and evening routines when users are most active.

- Introducing weekend challenges to boost engagement during lower activity periods.

- Using personalized notifications to encourage daily step goals that contribute to better sleep.

- Creating data-backed educational content highlighting the link between movement, sleep, and overall wellness.

These strategies can help Bellabeat strengthen its brand as a health companion for women, not just a fitness tracker.