

Bellabeat Case Study

Ched Chaves

2025-07-01

Contents

1 SCENARIO:	1
2 ASK	1
2.1 The Business Task:	1
2.2 The Stakeholders:	1
2.3 The Deliverables:	2
3 PREPARE:	2
4 PROCESS	2
4.1 Data Sources	2
4.2 Data Cleaning and Manipulation	3
5 ANALYZE	4
6 SHARE	7
7 ACT	7

1 SCENARIO:

I am a junior data analyst working on the marketing analyst team at Bellabeat, a high-tech manufacturer of health-focused products for women. Bellabeat has the potential to become a large 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.

I have been asked to focus on one of Bellabeat's products and analyze smart device data to gain meaningful insight into how consumers are using their smart devices. The insights that I will discover will then help guide the marketing strategy for the company. Towards the end of the case study, I will present my analysis and recommendation/s to the Bellabeat Executive Team.

2 ASK

2.1 The Business Task:

Analyze smart device usage data to gain meaningful insights on trends that can guide Bellabeat's marketing strategy and improve users' engagement with their wellness products.

2.2 The Stakeholders:

- Urška Sršen – Bellabeat Cofounder and Chief Creative Officer

- Sando Mur: Mathematician and Bellabeat's cofounder; key member of the Bellabeat executive team
- Marketing Analytics Team – The team responsible for collecting, analyzing and reporting data that helps guide Bellabeat's marketing strategy.

2.3 The Deliverables:

- A clear summary of the business task.
- A description of all data sources used.
- Documentation of any cleaning or manipulation of data.
- A summary of the analysis.
- Visualizations and key findings.
- Recommendations based on the analysis.

3 PREPARE:

Fitbit Fitness Tracker Data is a dataset that was generated from a distributed survey conducted via Amazon Mechanical Turk between 03.12.2016-05.12.2016. Thirty (30) consenting Fitbit users provided personal fitness tracker data, which includes minute-level records of physical activity, heart rate, and sleep patterns. Individual data can be identified by export session ID (column A) or timestamp (column B). Variation between output represents use of different types of Fitbit trackers and individual tracking behaviors and/or preferences.

The shareholder informs of some limitations with the data set and encourages adding another data to help address such limitation.

The raw data files are stored on a secure local directory on my laptop. All data processing and analysis were conducted using R with files imported directly from this folder.

The dataset contains various data types such as numeric, integer, character, date and date/time. Its data is organized in wide format, where each row represents a day of activity per user, with columns for various measurements like steps, calories, and sleep duration.

The dataset is publicly available under a CC0 license and thirty (30) users consented to the survey which is a small sample size so it has limitations. Also, the data was collected in 2016 and there was no mention of demographics which may not fully represent user behaviors.

While there are limitations to this dataset, it can still be used for analyzing trends for smart device usage since it has detailed and time-stamped fitness and health data which aligns with Bellabeat's focus on wellness tracking.

4 PROCESS

I chose R, which has tidyverse packages, to analyze the Fitbit dataset. Using some R commands guided me to ensure the data's integrity. In this process, I reviewed the data types and its structure, checked for duplicate data and missing values.

I performed the following steps in order to come up with a clean data: - Standardized column names - Ensured dates have appropriate data type - Removed incomplete records

4.1 Data Sources

- Fitbit data from Kaggle: <https://www.kaggle.com/arashnic/fitbit>
- Files used:
 - `dailyActivity_merged.csv` (daily activity data)
 - `sleepDay_merged.csv` (daily sleep data)

4.2 Data Cleaning and Manipulation

```
# Install packages
library(tidyverse)
library(janitor)
library(lubridate)
library(skimr)

theme_set(theme_minimal())

# Load and clean daily activity data
daily_activity <- read_csv("dailyActivity_merged.csv") %>%
  clean_names() %>%
  distinct() %>%
  mutate(activity_date = mdy(activity_date)) # Don't leave a dangling pipe here

# Load and clean sleep data
sleep_day <- read_csv("sleepDay_merged.csv", col_types = cols(SleepDay = col_character())) %>%
  clean_names() %>%
  distinct() %>%
  mutate(
    sleep_day = str_sub(sleep_day, 1, 10),
    sleep_day = mdy(sleep_day)
  ) %>%
  filter(!is.na(sleep_day))

# Merge the two datasets
combined_data <- merge(
  sleep_day, daily_activity,
  by.x = c("id", "sleep_day"),
  by.y = c("id", "activity_date")
)

# Preview structure
glimpse(combined_data)

## Rows: 295
## Columns: 18
## $ id <dbl> 1503960366, 1503960366, 1503960366, 1503960-
## $ sleep_day <date> 2016-04-12, 2016-04-13, 2016-04-15, 2016-0-
## $ total_sleep_records <dbl> 1, 2, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1-
## $ total_minutes_asleep <dbl> 327, 384, 412, 340, 700, 304, 360, 325, 361-
## $ total_time_in_bed <dbl> 346, 407, 442, 367, 712, 320, 377, 364, 384-
## $ total_steps <dbl> 13162, 10735, 9762, 12669, 9705, 15506, 105-
## $ total_distance <dbl> 8.50, 6.97, 6.28, 8.16, 6.48, 9.88, 6.68, 6-
## $ tracker_distance <dbl> 8.50, 6.97, 6.28, 8.16, 6.48, 9.88, 6.68, 6-
## $ logged_activities_distance <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0-
## $ very_active_distance <dbl> 1.88, 1.57, 2.14, 2.71, 3.19, 3.53, 1.96, 1-
## $ moderately_active_distance <dbl> 0.55, 0.69, 1.26, 0.41, 0.78, 1.32, 0.48, 0-
## $ light_active_distance <dbl> 6.06, 4.71, 2.83, 5.04, 2.51, 5.03, 4.24, 4-
## $ sedentary_active_distance <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0-
## $ very_active_minutes <dbl> 25, 21, 29, 36, 38, 50, 28, 19, 41, 39, 73, ~
## $ fairly_active_minutes <dbl> 13, 19, 34, 10, 20, 31, 12, 8, 21, 5, 14, 2-
## $ lightly_active_minutes <dbl> 328, 217, 209, 221, 164, 264, 205, 211, 262-
```

```
## $ sedentary_minutes      <dbl> 728, 776, 726, 773, 539, 775, 818, 838, 732~
## $ calories               <dbl> 1985, 1797, 1745, 1863, 1728, 2035, 1786, 1~
```

5 ANALYZE

```
# Exploratory Data Analysis
# Number of unique participants
n_participants <- n_distinct(combined_data$id)
cat("Number of participants in combined data:", n_participants, "\n")
```

```
## Number of participants in combined data: 24
```

```
# Summary statistics for key variables
combined_data %>%
  select(total_steps, total_distance, sedentary_minutes,
         total_minutes_asleep, total_time_in_bed) %>%
  skim()
```

Table 1: Data summary

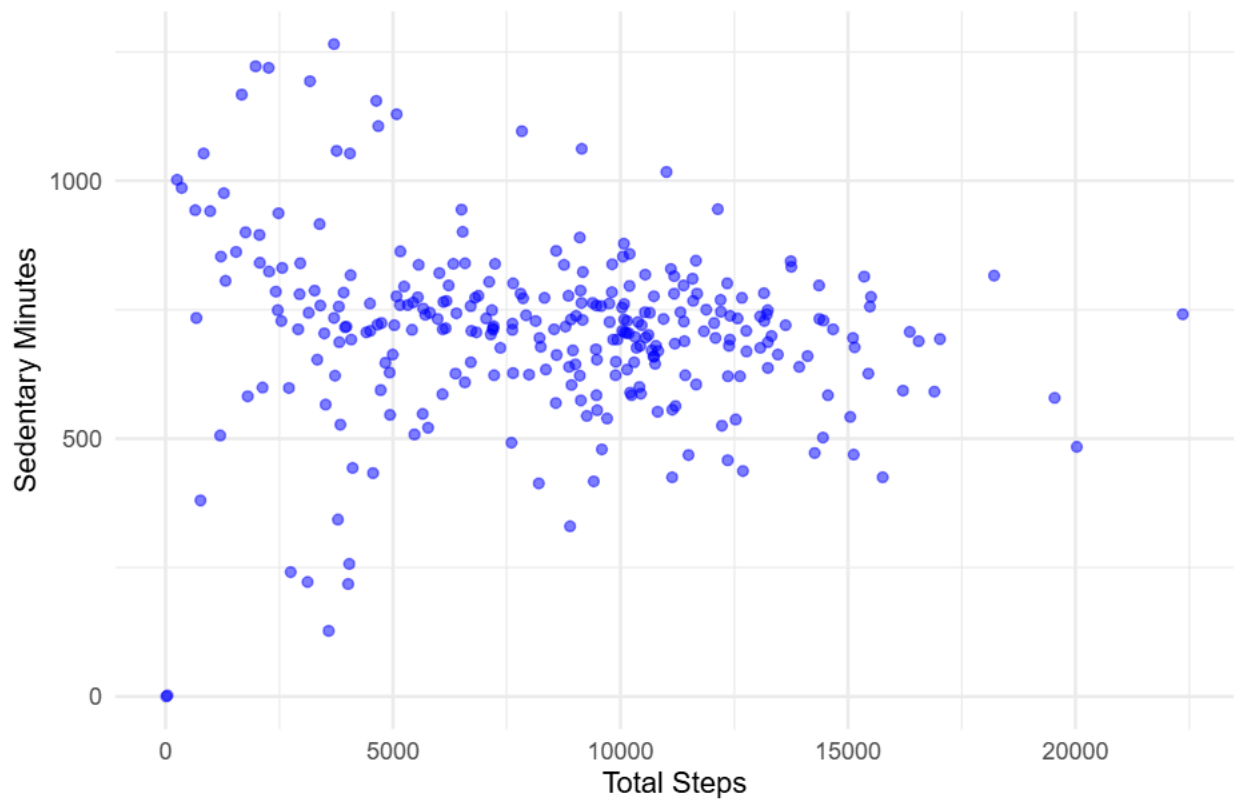
Name	Piped data
Number of rows	295
Number of columns	5
Column type frequency:	
numeric	5
Group variables	None

Variable type: numeric

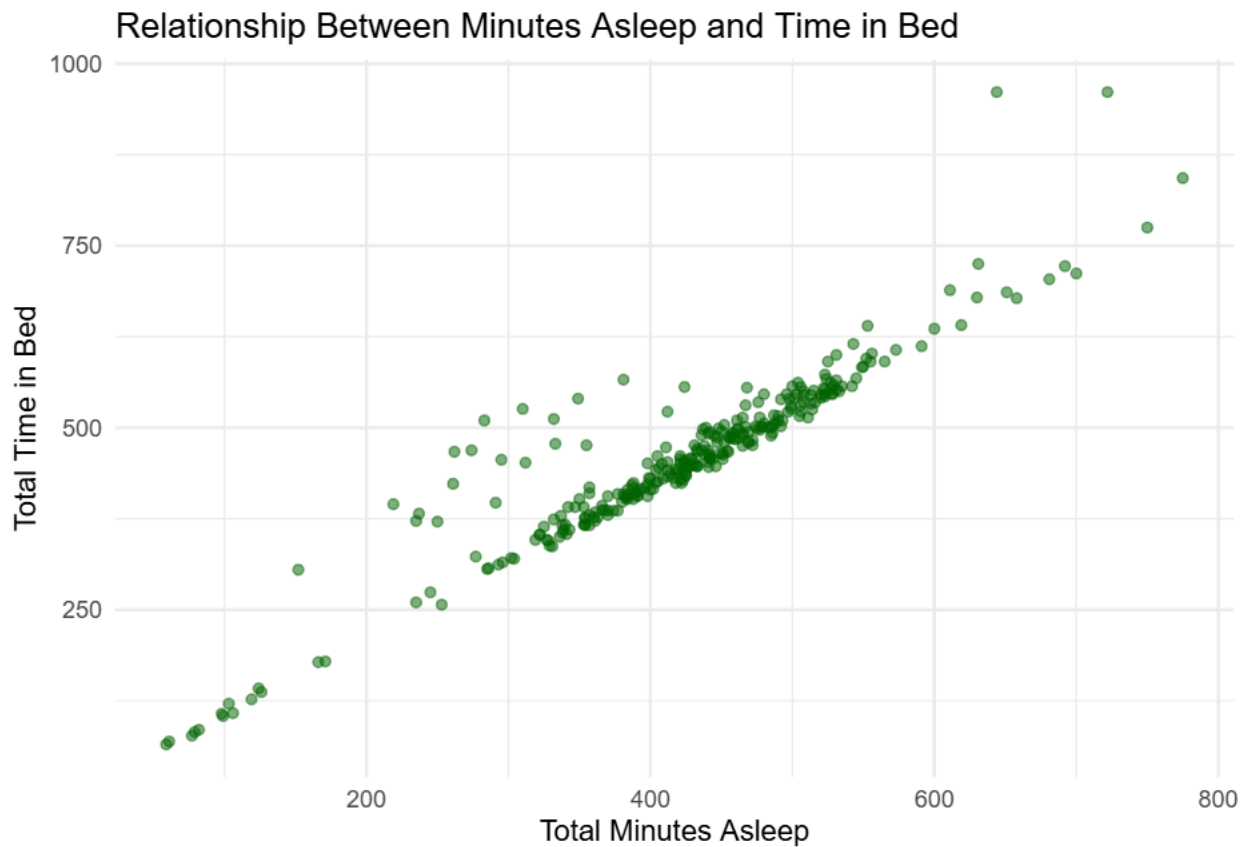
skim_variable	n_missing	complete	rat	mean	sd	p0	p25	p50	p75	p100	hist
total_steps	0	1		8427.12	4199.52	17.00	5053.0	8925.00	11187.00	22359.00	
total_distance	0	1		5.96	3.07	0.01	3.5	6.18	7.94	17.19	
sedentary_minutes	0	1		709.92	169.95	0.00	635.5	720.00	780.50	1265.00	
total_minutes_asleep	0	1		420.46	115.19	59.00	362.0	432.00	486.50	775.00	
total_time_in_bed	0	1		458.38	122.24	65.00	406.0	462.00	522.00	961.00	

```
# Visualization and Key Findings
# Scatter plot: Total steps vs Sedentary minutes
ggplot(combined_data, aes(x = total_steps, y = sedentary_minutes)) +
  geom_point(alpha = 0.5, color = "blue") +
  labs(
    title = "Relationship Between Total Steps and Sedentary Minutes",
    x = "Total Steps",
    y = "Sedentary Minutes"
  )
```

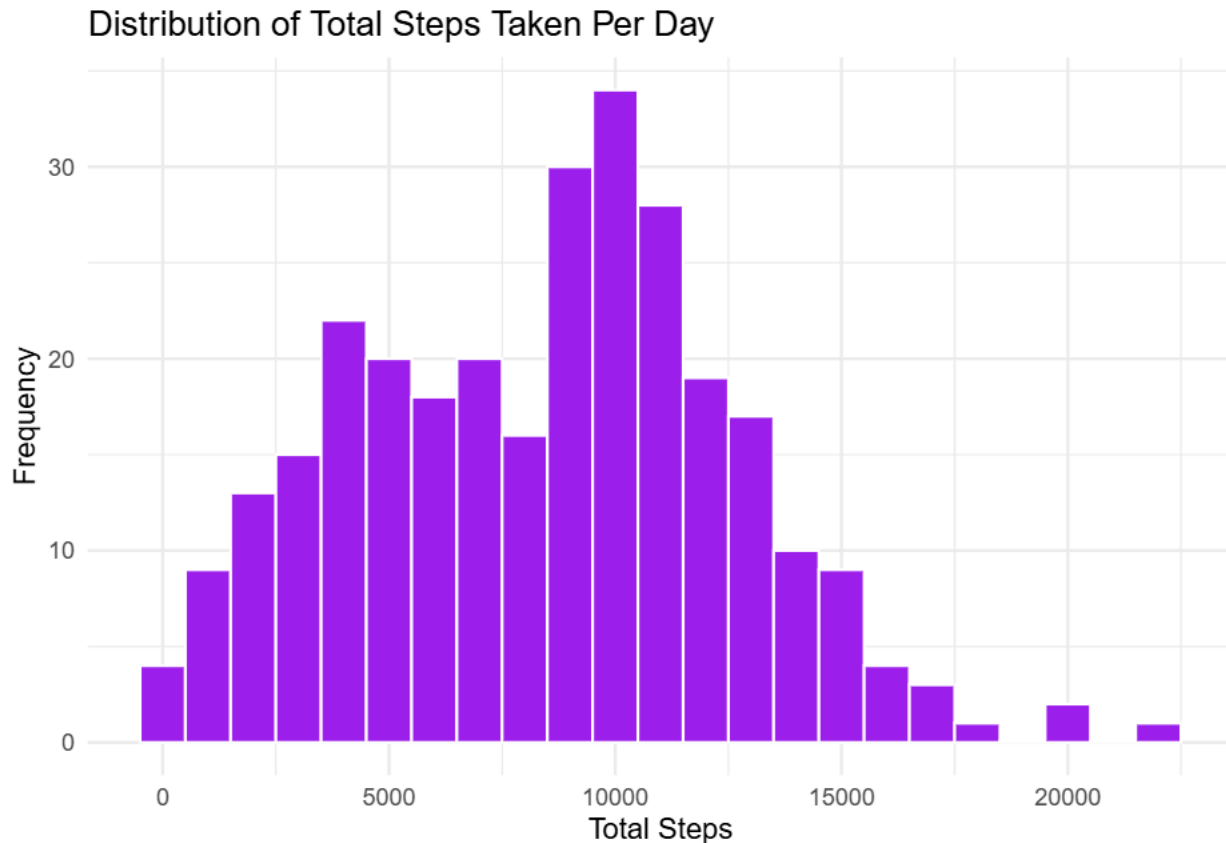
Relationship Between Total Steps and Sedentary Minutes



```
# Scatter plot: Minutes asleep vs Time in bed
ggplot(combined_data, aes(x = total_minutes_asleep, y = total_time_in_bed)) +
  geom_point(alpha = 0.5, color = "darkgreen") +
  labs(
    title = "Relationship Between Minutes Asleep and Time in Bed",
    x = "Total Minutes Asleep",
    y = "Total Time in Bed"
  )
```



```
# Histogram: Distribution of Total Steps
ggplot(combined_data, aes(x = total_steps)) +
  geom_histogram(binwidth = 1000, fill = "purple", color = "white") +
  labs(
    title = "Distribution of Total Steps Taken Per Day",
    x = "Total Steps",
    y = "Frequency"
  )
```



6 SHARE

After merging the dataset that contains information for both activity and sleep patterns, the data shows only twenty four (24) users. The following trends were observed for this dataset:

- A slight negative relationship between total steps and sedentary minutes.
- Spending more time in bed did not consistently result in longer sleep durations which may suggest potential issues with sleep quality.
- The majority of the users manifest light to moderate levels of physical activity.

These trends can help segment Bellabeat users for targeted campaigns (e.g., beginner vs. advanced fitness users).

7 ACT

Based on the observations and analysis of Fitbit user activity and sleep data, the following recommendations are proposed to guide Bellabeat's marketing strategy:

7.0.1 1. Segment and Target Users by Activity Level

The data reveals a wide variation in total steps and sedentary time among users, indicating different levels of physical activity. To address this, Bellabeat should engage less active users with beginner-friendly fitness challenges and daily movement reminders, while also encouraging consistent moderate activity across the user base through step-based rewards and progress tracking features.

7.0.2 2. Highlight Sleep Insights as a Unique Value Proposition

A big number of users spend more time in bed than actually sleeping, which may indicate potential issues with sleep quality. Bellabeat can market its sleep tracking feature as a tool for improving sleep efficiency.

Also, introduce personalized sleep tips or insights via the app to increase engagement.

7.0.3 3. Promote Holistic Wellness, Not Just Fitness

The overlap between physical activity and sleep patterns shows that users are not just fitness-focused. Bellabeat can position its product as a comprehensive wellness partner. I would also recommend to launch campaigns focused on stress reduction and rest together with physical activity.

7.0.4 4. Encourage Daily App Engagement

To improve retention and engage more on product usage, Bellabeat should send daily wellness tips based on recent user behavior. The company can also offer rewards or weekly summaries to incentivize and encourage continued use of the app and device.