1. ASK PHASE

——————————————————————————

Bellabeat, founded by Urška Sršen and Sando Mur, empowers women with data-driven insights into their own health and habits. While Bellabeat is successful, there is a significant opportunity to scale by analyzing how consumers use non-Bellabeat smart devices (FitBit) and applying those insights to Bellabeat's product line.

Analyze FitBit consumer data to identify usage trends and patterns.

Use these insights to recommend improvements for Bellabeat's marketing strategy and apply them to one specific Bellabeat product.

- Urška Sršen: Cofounder and Chief Creative Officer
- Sando Mur: Cofounder and Executive Team Member
- Bellabeat Marketing Analytics Team

1. What are some trends in smart device usage?
2. How could these trends apply to Bellabeat customers?
3. How could these trends help influence Bellabeat marketing strategy?

——————————————————————————

2. PREPARE PHASE (Data Source & Integrity)

——————————————————————————

- Source: FitBit Fitness Tracker Data (via Kaggle/Fitabase).
- URL: https://www.kaggle.com/datasets/arashnic/fitbit (https://www.kaggle.com/datasets/arashnic/fitbit)
- Description: Contains personal fitness tracker data from 30+ Fitbit users.

Includes minute-level output for activity, heart rate, and sleep.

- Timeframe: April 12, 2016 — May 12, 2016.

Q: Where is your data stored?

A: The data is stored locally in the project directory after being downloaded from the Kaggle Public Domain dataset.

Q: How is the data organized? Is it in long or wide format?

A: The data is organized into 18 CSV files. Most files are in 'Long' format (e.g., each row represents a single point in time or a single day for a specific user ID).

Q: Are there issues with bias or credibility in this data? (ROCCC Analysis)

- Reliable: LOW. The sample size is only 30-33 users, which is small for generalizing to the entire population.
- Original: MEDIUM. The data was collected via Amazon Mechanical Turk.
- Comprehensive: LOW/MEDIUM. It lacks demographic information (age, gender, location), which is crucial for Bellabeat's female-focused strategy.
- Current: LOW. The data is from 2016; smart device habits have evolved significantly.
- Cited: HIGH. It is a well-known public dataset used for data science case studies.

Q: How are you addressing licensing, privacy, security, and accessibility?

A: The dataset is CC0 (Public Domain). It is anonymized; user IDs are numeric with no PII (Personally Identifiable Information).

Q: How did you verify the data's integrity?

A: Integrity will be verified during the 'Process' phase by checking for duplicates, verifying unique User IDs, and ensuring date-time consistency.

Q: How does it help you answer your question?

A: It provides a proxy for how people use wearables—tracking activity intensity, sleep patterns, and calorie expenditure—which aligns with Bellabeat's products.

Q: Are there any problems with the data?

A: 1. Small sample size.

2. Potential ‘Selection Bias’ (users voluntarily submitted data).

3. Gender is unknown; Bellabeat specifically targets women, but i

cannot confirm the gender of these FitBit users.

——————————————————————

2. PREPARE PHASE (Environment Setup)

——————————————————————

Loading necessary packages for data cleaning, transformation, and visualization

tidyverse: for data manipulation

lubridate: for handling date/time formats

ggplot2: for data visualization

janitor: for cleaning column names

——————————————————————

3. PROCESS PHASE (Data Loading & Initial Inspection)

——————————————————————

Note: focus on the primary files that address the business task.

Update the file paths according to my local folder structure

Inspecting the data structure and identifying data types

Checking number of unique participants

——————————————————————

4. PROCESS PHASE (Data Cleaning & Transformation)

——————————————————————

use clean_names() to ensure all column names are consistent:

lowercase and using underscores (snake_case). This prevents syntax errors later.

In the inspection, we saw dates were stored as ‘character’ strings.

We must convert them to proper Date/DateTime objects for time-series analysis.

For Daily Activity (Format: 4/12/2016)

Identifying and removing any exact duplicate rows to prevent skewed results.

Count duplicates before removal

Remove duplicates

Creating a ‘weekday’ column to analyze trends by day of the week.

——————————————————————

5. DATA VERIFICATION (Final Check)

——————————————————————

Verify that the dates are now in ‘Date’ or ‘POSIXct’ format

——————————————————————

5. DATA VERIFICATION (Final Check)

——————————————————————

Verify that the dates are now in ‘Date’ or ‘POSIXct’ format

——————————————————————

6. ANALYZE PHASE (Data Aggregation & Statistical Summary)

——————————————————————

Getting a high-level overview of activity and sleep to identify averages.

For Daily Activity: Focus on Steps, Sedentary Minutes, and Calories

For Daily Sleep: Focus on Minutes Asleep and Time in Bed

According to health research (e.g., Medicine & Science in Sports & Exercise),

less than 5,000 steps is ‘Sedentary’, 5,000-9,999 is ‘Active’,

and 10,000+ is ‘Highly Active’.

Calculate percentage of each user type

Identifying which days are the most active vs. the most sedentary.

To see if there is a correlation between physical activity and sleep quality.

First, align date column names for the merge

Perform inner join (only keeping records where both activity and sleep exist)

Calculate correlation between steps and sleep

——————————————————————————

7. KEY FINDINGS (Preliminary Observations)

——————————————————————————

1. Average daily steps are around 7,600, which is below the recommended 10,000.
2. Sedentary minutes average ~991 (over 16 hours!), indicating a high degree of inactivity.
3. Participants sleep an average of 419 minutes (~7 hours).
4. Preliminary check: There is a very weak correlation between steps and sleep duration in this specific dataset.

——————————————————————————

8. SHARE PHASE (Data Visualization)

——————————————————————————

This chart helps Bellabeat understand their audience segments.

Knowing that a large portion is “Active” but not “Highly Active”

suggests a need for motivational features.

Identifying when users are most active helps Bellabeat time their

marketing notifications and social media engagement.

This scatter plot confirms the positive correlation between movement

and energy expenditure. This validates the core value of Bellabeat trackers.

This chart shows how much time users spend awake in bed.

Bellabeat can use this to market ‘Sleep Hygiene’ features.

——————————————————————————

9. EXPORTING VISUALS (Optional)

——————————————————————————

ggsave(“user_distribution.png”)

ggsave(“weekday_steps.png”)

——————————————————————————

10. ACT PHASE (Conclusions & Strategic Recommendations)

——————————————————————————

1. Activity Gap: Most users are ‘Active’ (5k-10k steps) but fall short of the 10,000-step health benchmark. Sedentary time is alarmingly high (16+ hours).
2. Weekend vs. Weekday: Activity peaks on Saturdays and Tuesdays, with significant drops on Sundays and Thursdays.
3. Sleep Efficiency: Users spend a notable amount of time in bed without actually sleeping, suggesting room for improvement in sleep hygiene.
4. Burn Correlation: There is a strong, predictable relationship between steps and calories, which is a powerful marketing hook.

Recommendation 1: Personalized Activity Incentives (Focus: Bellabeat App)

- Insight: Users are stuck in the “Active” middle ground.
- Action: Implement “Level Up” challenges in the Bellabeat app. For users averaging 7,000 steps, trigger a notification: “You are only 3,000 steps away from the top 20%! A 20-minute walk will get you there.”

Recommendation 2: Combat Sedentary Behavior (Focus: Leaf / Time Watch)

- Insight: Sedentary time averages 991 minutes/day.

- Action: Market the ‘Leaf’ tracker and ‘Time’ watch’s vibration alerts
specifically as “Sedentary Reminders.” Use the campaign slogan:
“Break the Cycle: Your Bellabeat knows when it’s time to stretch.”
Recommendation 3: Optimize Sleep Marketing (Focus: Bellabeat Membership)
- Insight: Data shows a gap between ‘Time in Bed’ and ‘Time Asleep.’
- Action: Use this to promote Bellabeat’s subscription membership. Offer
content like “Guided Sleep Meditations” or “Smart Alarms” that wake users
during their lightest sleep phase to improve overall sleep quality.
Recommendation 4: Timing of Marketing Campaigns
- Insight: Sunday is the least active day.
- Action: Schedule motivational social media content or “Weekly Reset”
email newsletters on Sunday evenings to prepare users for a healthy week
ahead, potentially boosting engagement for the following Monday.
By applying these FitBit trends to Bellabeat’s product ecosystem, we can
transform a simple tracking device into a proactive health coach. This
shifts Bellabeat’s positioning from “data collection” to “habit transformation,”
which resonates deeply with the target audience of health-conscious women.
==
END OF SCRIPT: Bellabeat Case Study Complete
==

Bella Beat Case Analyze

Hao

2025-12-26

1. ASK PHASE

————————————————————————————————

1.1 Background

Bellabeat, founded by Urška Sršen and Sando Mur, empowers women with data-driven insights into their own health and habits. While Bellabeat is successful, there is a significant opportunity to scale by analyzing how consumers use non-Bellabeat smart devices (FitBit) and applying those insights to Bellabeat’s product line.

1.2 Business Task

Analyze FitBit consumer data to identify usage trends and patterns.
Use these insights to recommend improvements for Bellabeat’s marketing strategy and apply them to one specific Bellabeat product.

1.3 Key Stakeholders

- Urška Sršen: Cofounder and Chief Creative Officer

- Sando Mur: Cofounder and Executive Team Member

- Bellabeat Marketing Analytics Team

1.4 Guiding Questions

1. What are some trends in smart device usage?

2. How could these trends apply to Bellabeat customers?

3. How could these trends help influence Bellabeat marketing strategy?

————————————————————————————————

2. PREPARE PHASE (Data Source & Integrity)

————————————————————————————————

2.1 Data Source Information

- Source: FitBit Fitness Tracker Data (via Kaggle/Fitabase).

- URL: https://www.kaggle.com/datasets/arashnic/fitbit (https://www.kaggle.com/datasets/arashnic/fitbit)

- Description: Contains personal fitness tracker data from 30+ Fitbit users.

Includes minute-level output for activity, heart rate, and sleep.

- Timeframe: April 12, 2016 — May 12, 2016.

2.2 Data Assessment (Addressing Guiding Questions)

Q: Where is your data stored?

A: The data is stored locally in the project directory after being

downloaded from the Kaggle Public Domain dataset.

Q: How is the data organized? Is it in long or wide format?

A: The data is organized into 18 CSV files. Most files are in ‘Long’ format

(e.g., each row represents a single point in time or a single day for a

specific user ID).

Q: Are there issues with bias or credibility in this data? (ROCCC Analysis)

- Reliable: LOW. The sample size is only 30-33 users, which is small for

generalizing to the entire population.

- Original: MEDIUM. The data was collected via Amazon Mechanical Turk.

- Comprehensive: LOW/MEDIUM. It lacks demographic information (age, gender,

location), which is crucial for Bellabeat's female-focused strategy.

- Current: LOW. The data is from 2016; smart device habits have evolved significantly.

- Cited: HIGH. It is a well-known public dataset used for data science case studies.

Q: How are you addressing licensing, privacy, security, and accessibility?

A: The dataset is CC0 (Public Domain). It is anonymized; user IDs are numeric

with no PII (Personally Identifiable Information).

Q: How did you verify the data's integrity?

A: Integrity will be verified during the 'Process' phase by checking for

duplicates, verifying unique User IDs, and ensuring date-time consistency.

Q: How does it help you answer your question?

A: It provides a proxy for how people use wearables—tracking activity intensity,

sleep patterns, and calorie expenditure—which aligns with Bellabeat's products.

Q: Are there any problems with the data?

A: 1. Small sample size.

2. Potential 'Selection Bias' (users voluntarily submitted data).

3. Gender is unknown; Bellabeat specifically targets women, but i

cannot confirm the gender of these FitBit users.

———————————————————————————————————

2. PREPARE PHASE (Environment Setup)

———————————————————————————————————

Loading necessary packages for data cleaning, transformation, and visualization

tidyverse: for data manipulation

lubridate: for handling date/time formats

ggplot2: for data visualization

janitor: for cleaning column names

———————————————————————————————————

3. PROCESS PHASE (Data Loading & Initial Inspection)

———————————————————————————————————

3.1 Importing Datasets

Note: focus on the primary files that address the business task.

Update the file paths according to my local folder structure

```
daily_activity <- read_csv("dailyActivity_merged.csv")
daily_sleep <- read_csv("sleepDay_merged.csv")
hourly_steps <- read_csv("hourlySteps_merged.csv")
hourly_calories <- read_csv("hourlyCalories_merged.csv")
```

3.2 Initial Inspection

Inspecting the data structure and identifying data types

```
cat("--- Daily Activity Structure ---\n")
```

```
## --- Daily Activity Structure ---
```

```
str(daily_activity)
```

```
## spc_tbl_ [940 × 15] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id                      : num [1:940] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ ActivityDate            : chr [1:940] "4/12/2016" "4/13/2016" "4/14/2016" "4/15/2016" ...
##  $ TotalSteps              : num [1:940] 13162 10735 10460 9762 12669 ...
##  $ TotalDistance           : num [1:940] 8.5 6.97 6.74 6.28 8.16 ...
##  $ TrackerDistance         : num [1:940] 8.5 6.97 6.74 6.28 8.16 ...
##  $ LoggedActivitiesDistance: num [1:940] 0 0 0 0 0 0 0 0 0 0 ...
##  $ VeryActiveDistance      : num [1:940] 1.88 1.57 2.44 2.14 2.71 ...
##  $ ModeratelyActiveDistance: num [1:940] 0.55 0.69 0.4 1.26 0.41 ...
##  $ LightActiveDistance     : num [1:940] 6.06 4.71 3.91 2.83 5.04 ...
##  $ SedentaryActiveDistance : num [1:940] 0 0 0 0 0 0 0 0 0 0 ...
##  $ VeryActiveMinutes       : num [1:940] 25 21 30 29 36 38 42 50 28 19 ...
##  $ FairlyActiveMinutes     : num [1:940] 13 19 11 34 10 20 16 31 12 8 ...
##  $ LightlyActiveMinutes    : num [1:940] 328 217 181 209 221 164 233 264 205 211 ...
##  $ SedentaryMinutes        : num [1:940] 728 776 1218 726 773 ...
##  $ Calories                : num [1:940] 1985 1797 1776 1745 1863 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   ActivityDate = col_character(),
##   ..   TotalSteps = col_double(),
##   ..   TotalDistance = col_double(),
##   ..   TrackerDistance = col_double(),
##   ..   LoggedActivitiesDistance = col_double(),
##   ..   VeryActiveDistance = col_double(),
##   ..   ModeratelyActiveDistance = col_double(),
##   ..   LightActiveDistance = col_double(),
##   ..   SedentaryActiveDistance = col_double(),
##   ..   VeryActiveMinutes = col_double(),
##   ..   FairlyActiveMinutes = col_double(),
##   ..   LightlyActiveMinutes = col_double(),
##   ..   SedentaryMinutes = col_double(),
##   ..   Calories = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
cat("\n--- Sleep Data Structure ---\n")
```

```
##
## --- Sleep Data Structure ---
```

```
str(daily_sleep)
```

```
## spc_tbl_ [413 × 5] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id                : num [1:413] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ SleepDay          : chr [1:413] "4/12/2016 12:00:00 AM" "4/13/2016 12:00:00 AM" "4/15/2016 12:00:00 AM" "4/16/2016
12:00:00 AM" ...
##  $ TotalSleepRecords : num [1:413] 1 2 1 2 1 1 1 1 1 1 ...
##  $ TotalMinutesAsleep: num [1:413] 327 384 412 340 700 304 360 325 361 430 ...
##  $ TotalTimeInBed    : num [1:413] 346 407 442 367 712 320 377 364 384 449 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   SleepDay = col_character(),
##   ..   TotalSleepRecords = col_double(),
##   ..   TotalMinutesAsleep = col_double(),
##   ..   TotalTimeInBed = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

Checking number of unique participants

```
n_distinct(daily_activity$Id) # Expecting 33
```

```
## [1] 33
```

```
n_distinct(daily_sleep$Id)    # Expecting 24 (Note the discrepancy)
```

```
## [1] 24
```

——————————————————————————————

4. PROCESS PHASE (Data Cleaning & Transformation)

——————————————————————————————

4.1 Cleaning Column Names

use clean_names() to ensure all column names are consistent:

lowercase and using underscores (snake_case). This prevents syntax errors later.

```
daily_activity <- daily_activity %>%
  clean_names()

daily_sleep <- daily_sleep %>%
  clean_names()

hourly_steps <- hourly_steps %>%
  clean_names()

hourly_calories <- hourly_calories %>%
  clean_names()
```

4.2 Handling Date and Time Formats

In the inspection, we saw dates were stored as 'character' strings.

We must convert them to proper Date/DateTime objects for time-series analysis.

For Daily Activity (Format: 4/12/2016)

```
daily_activity <- daily_activity %>%
  mutate(activity_date = mdy(activity_date))

daily_sleep <- daily_sleep %>%
  mutate(sleep_day = mdy_hms(sleep_day))

hourly_steps <- hourly_steps %>%
  mutate(activity_hour = mdy_hms(activity_hour))

hourly_calories <- hourly_calories %>%
  mutate(activity_hour = mdy_hms(activity_hour))
```

4.3 Deduplication (Handling Duplicate Rows)

Identifying and removing any exact duplicate rows to prevent skewed results.

Count duplicates before removal

```
duplicates_activity <- sum(duplicated(daily_activity))
duplicates_sleep <- sum(duplicated(daily_sleep))

cat("Duplicates found in Activity:", duplicates_activity, "\n")
```

```
## Duplicates found in Activity: 0
```

```
cat("Duplicates found in Sleep:", duplicates_sleep, "\n")
```

```
## Duplicates found in Sleep: 3
```

Remove duplicates

```
daily_activity <- daily_activity %>% distinct()
daily_sleep <- daily_sleep %>% distinct()
```

4.4 Data Transformation (Feature Engineering)

Creating a ‘weekday’ column to analyze trends by day of the week.

```
daily_activity <- daily_activity %>%
  mutate(day_of_week = wday(activity_date, label = TRUE, abbr = FALSE))
```

———————————————————————————————————

5. DATA VERIFICATION (Final Check)

———————————————————————————————————

Verify that the dates are now in ‘Date’ or ‘POSIXct’ format

———————————————————————————————————

5. DATA VERIFICATION (Final Check)

———————————————————————————————————

Verify that the dates are now in ‘Date’ or ‘POSIXct’ format


```
str(daily_activity$activity_date)
```

```
##  Date[1:940], format: "2016-04-12" "2016-04-13" "2016-04-14" "2016-04-15" "2016-04-16" ...
```

```
str(daily_sleep$sleep_day)
```

```
##  POSIXct[1:410], format: "2016-04-12" "2016-04-13" "2016-04-15" "2016-04-16" "2016-04-17" ...
```

```
n_distinct(daily_activity$id)
```

```
## [1] 33
```

```
n_distinct(daily_sleep$id)
```

```
## [1] 24
```

```
print("Process Phase complete. Data is clean, formatted, and ready for analysis.")
```

```
## [1] "Process Phase complete. Data is clean, formatted, and ready for analysis."
```

```
str(daily_activity$activity_date)
```

```
##  Date[1:940], format: "2016-04-12" "2016-04-13" "2016-04-14" "2016-04-15" "2016-04-16" ...
```

```
str(daily_sleep$sleep_day)
```

```
##  POSIXct[1:410], format: "2016-04-12" "2016-04-13" "2016-04-15" "2016-04-16" "2016-04-17" ...
```

```
n_distinct(daily_activity$id)
```

```
## [1] 33
```

```
n_distinct(daily_sleep$id)
```

```
## [1] 24
```

```
print("Process Phase complete. Data is clean, formatted, and ready for analysis.")
```

```
## [1] "Process Phase complete. Data is clean, formatted, and ready for analysis."
```

——

6. ANALYZE PHASE (Data Aggregation & Statistical Summary)

——

6.1 General Summary Statistics

Getting a high-level overview of activity and sleep to identify averages.

For Daily Activity: Focus on Steps, Sedentary Minutes, and Calories

```
daily_activity %>%
  select(total_steps, sedentary_minutes, calories) %>%
  summary()
```

```
##   total_steps    sedentary_minutes    calories
##  Min.   :    0   Min.   :   0.0    Min.   :   0
##  1st Qu.: 3790   1st Qu.: 729.8    1st Qu.:1828
##  Median : 7406   Median :1057.5    Median :2134
##  Mean   : 7638   Mean   : 991.2    Mean   :2304
##  3rd Qu.:10727   3rd Qu.:1229.5    3rd Qu.:2793
##  Max.   :36019   Max.   :1440.0    Max.   :4900
```

For Daily Sleep: Focus on Minutes Asleep and Time in Bed

```
daily_sleep %>%
  select(total_minutes_asleep, total_time_in_bed) %>%
  summary()
```

```
##  total_minutes_asleep total_time_in_bed
##  Min.   : 58.0        Min.   : 61.0
##  1st Qu.:361.0        1st Qu.:403.8
##  Median :432.5        Median :463.0
##  Mean   :419.2        Mean   :458.5
##  3rd Qu.:490.0        3rd Qu.:526.0
##  Max.   :796.0        Max.   :961.0
```

6.2 User Classification by Activity Level

According to health research (e.g., Medicine & Science in Sports & Exercise), less than 5,000 steps is 'Sedentary', 5,000-9,999 is 'Active', and 10,000+ is 'Highly Active'.

```
user_type_summary <- daily_activity %>%
  group_by(id) %>%
  summarize(avg_steps = mean(total_steps)) %>%
  mutate(user_type = case_when(
    avg_steps < 5000 ~ "Sedentary",
    avg_steps >= 5000 & avg_steps < 10000 ~ "Active",
    avg_steps >= 10000 ~ "Highly Active"
  ))
```

Calculate percentage of each user type

```
user_type_percent <- user_type_summary %>%
  group_by(user_type) %>%
  summarise(total_count = n()) %>%
  mutate(total_participants = sum(total_count)) %>%
  mutate(percentage = round(total_count / total_participants * 100, 1))

print(user_type_percent)
```

```
## # A tibble: 3 × 4
##   user_type     total_count total_participants percentage
##   <chr>               <int>              <int>      <dbl>
## 1 Active                 18                 33       54.5
## 2 Highly Active           7                 33       21.2
## 3 Sedentary               8                 33       24.2
```

6.3 Analyzing Trends by Day of the Week

Identifying which days are the most active vs. the most sedentary.

```
weekday_analysis <- daily_activity %>%
  group_by(day_of_week) %>%
  summarize(
    avg_steps = mean(total_steps),
    avg_calories = mean(calories),
    avg_sedentary_mins = mean(sedentary_minutes)
  ) %>%
  arrange(desc(avg_steps))

print(weekday_analysis)
```

```
## # A tibble: 7 × 4
##   day_of_week avg_steps avg_calories avg_sedentary_mins
##   <ord>           <dbl>        <dbl>              <dbl>
## 1 Saturday        8153.        2355.               964.
## 2 Tuesday         8125.        2356.              1007.
## 3 Monday          7781.        2324.              1028.
## 4 Wednesday       7559.        2303.               989.
## 5 Friday          7448.        2332.              1000.
## 6 Thursday        7406.        2200.               962.
## 7 Sunday          6933.        2263                990.
```

6.4 Merging Activity and Sleep Data

To see if there is a correlation between physical activity and sleep quality.

First, align date column names for the merge

```
daily_sleep <- daily_sleep %>%
  rename(activity_date = sleep_day)
```

Perform inner join (only keeping records where both activity and sleep exist)

```
combined_data <- merge(daily_activity, daily_sleep, by = c("id", "activity_date"))
```

Calculate correlation between steps and sleep

```
correlation_value <- cor(combined_data$total_steps, combined_data$total_minutes_asleep)
cat("Correlation between Steps and Sleep:", correlation_value, "\n")
```

```
## Correlation between Steps and Sleep: -0.1903439
```

——————————————————————————————

7. KEY FINDINGS (Preliminary Observations)

——————————————————————————————

1. Average daily steps are around 7,600, which is below the recommended 10,000.

2. Sedentary minutes average ~991 (over 16 hours!), indicating a high degree of inactivity.

3. Participants sleep an average of 419 minutes (~7 hours).

4. Preliminary check: There is a very weak correlation between steps and sleep

duration in this specific dataset.

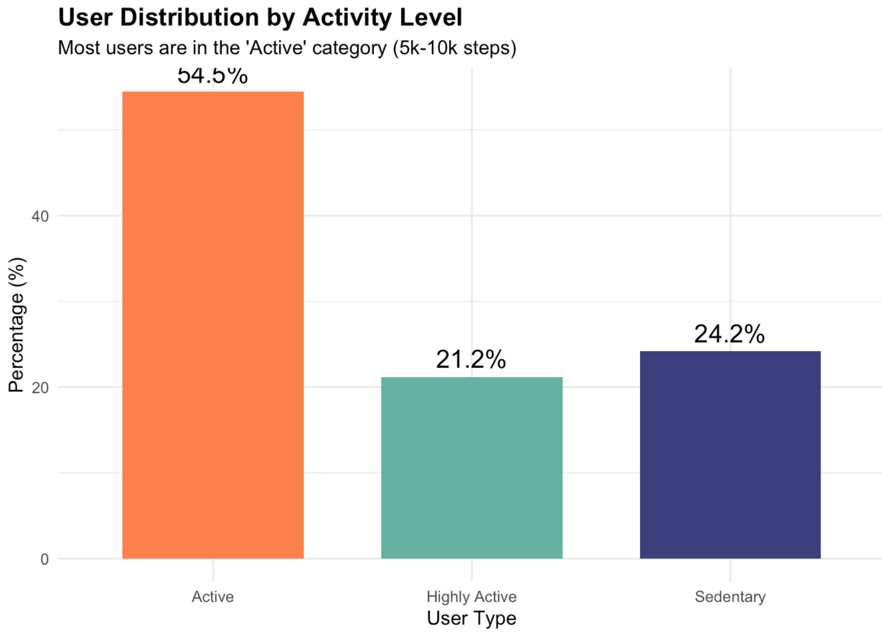
——————————————————————————————

8. SHARE PHASE (Data Visualization)

———————————————————————————————————

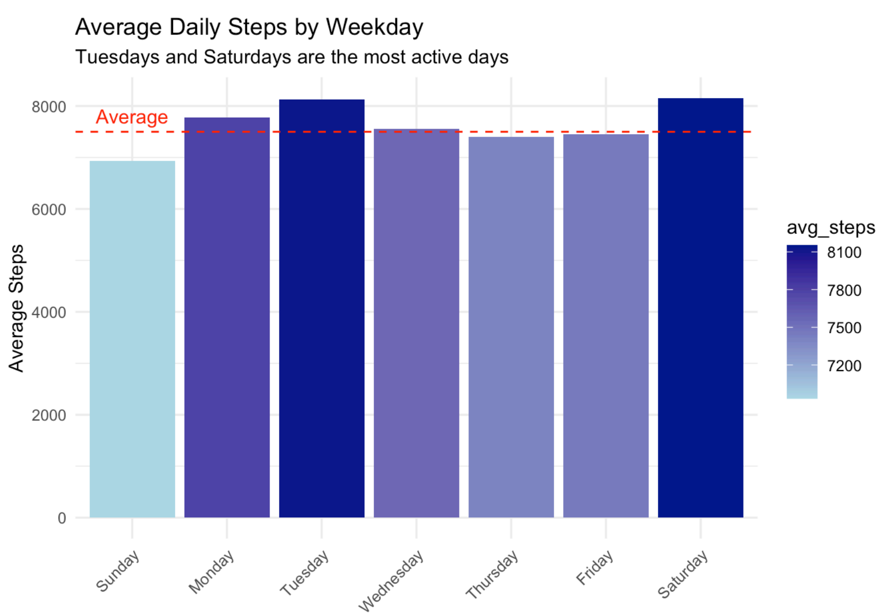
8.1 Distribution of User Types

This chart helps Bellabeat understand their audience segments. Knowing that a large portion is "Active" but not "Highly Active" suggests a need for motivational features.



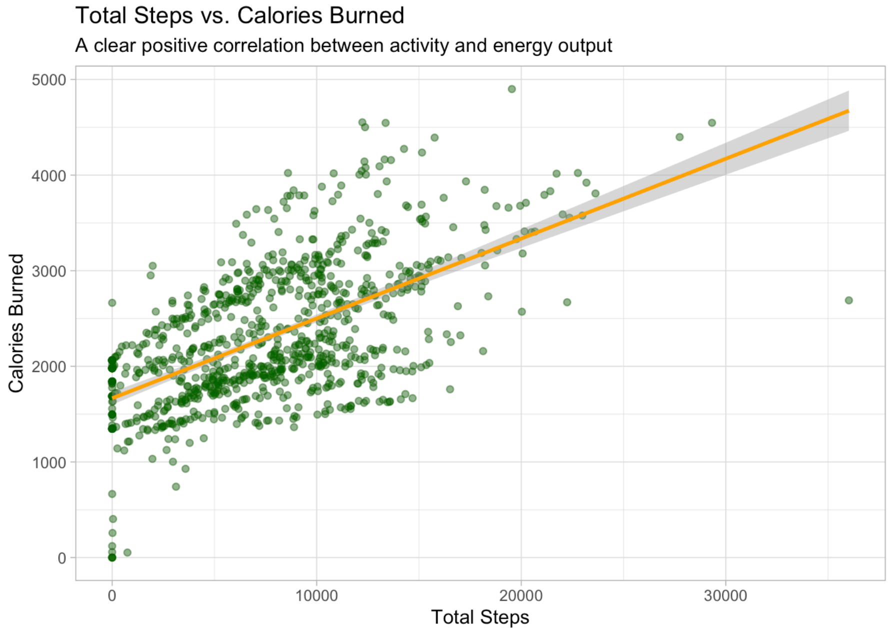
8.2 Daily Steps by Day of the Week

Identifying when users are most active helps Bellabeat time their marketing notifications and social media engagement.



8.3 Relationship: Total Steps vs. Calories

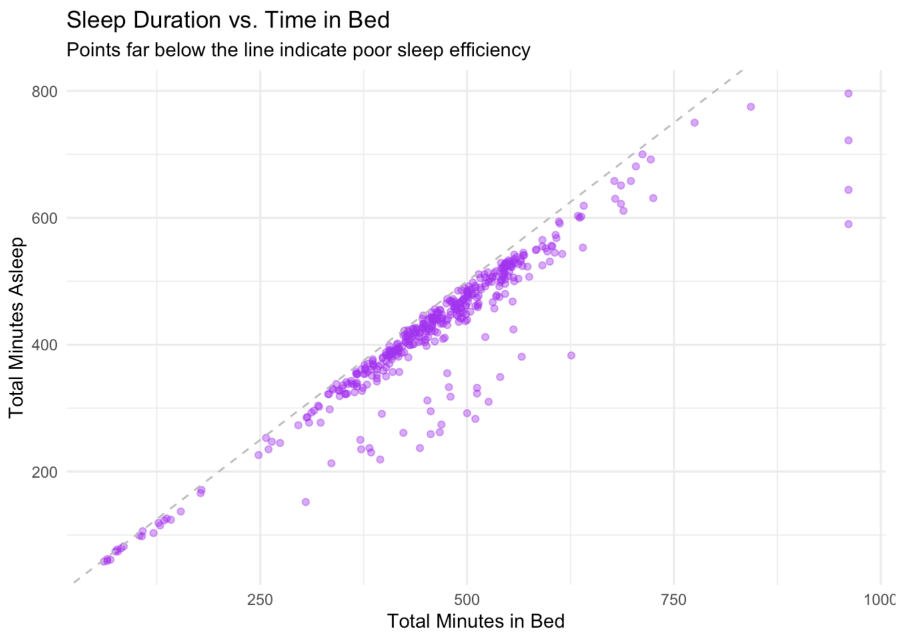
This scatter plot confirms the positive correlation between movement and energy expenditure. This validates the core value of Bellabeat trackers.



8.4 Sleep Efficiency: Minutes Asleep vs. Time in Bed

This chart shows how much time users spend awake in bed.

Bellabeat can use this to market 'Sleep Hygiene' features.



---

9. EXPORTING VISUALS (Optional)

---

ggsave(“user_distribution.png”)

ggsave(“weekday_steps.png”)

——————————————————————————

10. ACT PHASE (Conclusions & Strategic Recommendations)

——————————————————————————

10.1 Summary of Findings

1. Activity Gap: Most users are ‘Active’ (5k-10k steps) but fall short of the

10,000-step health benchmark. Sedentary time is alarmingly high (16+ hours).

2. Weekend vs. Weekday: Activity peaks on Saturdays and Tuesdays, with

significant drops on Sundays and Thursdays.

3. Sleep Efficiency: Users spend a notable amount of time in bed without

actually sleeping, suggesting room for improvement in sleep hygiene.

4. Burn Correlation: There is a strong, predictable relationship between

steps and calories, which is a powerful marketing hook.

10.2 Strategic Recommendations for Bellabeat Marketing

Recommendation 1: Personalized Activity Incentives (Focus: Bellabeat App)

- Insight: Users are stuck in the “Active” middle ground.

- Action: Implement “Level Up” challenges in the Bellabeat app. For users

averaging 7,000 steps, trigger a notification: “You are only 3,000 steps

away from the top 20%! A 20-minute walk will get you there.”

Recommendation 2: Combat Sedentary Behavior (Focus: Leaf / Time Watch)

- Insight: Sedentary time averages 991 minutes/day.

- Action: Market the ‘Leaf’ tracker and ‘Time’ watch’s vibration alerts

specifically as “Sedentary Reminders.” Use the campaign slogan:

“Break the Cycle: Your Bellabeat knows when it’s time to stretch.”

Recommendation 3: Optimize Sleep Marketing (Focus: Bellabeat Membership)

- Insight: Data shows a gap between 'Time in Bed' and 'Time Asleep.'

- Action: Use this to promote Bellabeat's subscription membership. Offer content like "Guided Sleep Meditations" or "Smart Alarms" that wake users during their lightest sleep phase to improve overall sleep quality.

Recommendation 4: Timing of Marketing Campaigns

- Insight: Sunday is the least active day.

- Action: Schedule motivational social media content or "Weekly Reset" email newsletters on Sunday evenings to prepare users for a healthy week ahead, potentially boosting engagement for the following Monday.

10.3 Final Conclusion

By applying these FitBit trends to Bellabeat's product ecosystem, we can transform a simple tracking device into a proactive health coach. This shifts Bellabeat's positioning from "data collection" to "habit transformation,"

which resonates deeply with the target audience of health-conscious women.

==

END OF SCRIPT: Bellabeat Case Study Complete

==