BELLABEAT R MARKDOWN

AGRIPPINE L TOBIAS

2025-10-07

Introduction

Bellabeat is a wellness high-technology company that designs health-focused products for women, including the Bellabeat app, wearable devices, a smart water bottle, and a subscription app. Founded in 2013 by Urška Sršen and Sando Mur, the company has grown quickly and markets its products worldwide.

As a junior data analyst on the marketing team, I was tasked with analyzing smart devices' fitness data and determining how it could help unlock new growth opportunities for Bellabeat. I will focus on one of Bellabeat's products: **Bellabeat's Time wellness.**

The case study follows the six step data analysis process:ASK,PREPARE,PROCESS,ANALYZE,SHARE,ACT

STEP 1: ASK

Business Task Statement

Analyze public smart device data to identify user activity, sleep, and lifestyle patterns. Use insights to recommend marketing strategies for Bellabeat's Time watch.

Guiding Questions

1. What are some trends in smart device usage (How are people using smart devices)?
2. How could these trends apply to Bellabeat customers?
3. How could these trends help influence Bellabeat's marketing strategy?

Key Stakeholders: Urška Sršen (Chief Creative Officer), Sando Mur (Cofounder), Bellabeat Marketing Analytics Team, Executive Team

STEP 2: PREPARE

2.1 Dataset used

The data source used for our case study is FitBit Fitness Tracker Data. This dataset is stored in Kaggle and was made available through Mobius

2.2 Accessibility and privacy of data:

Data used is from Kaggle, a public domain, thus it is open source. The owner has dedicated the work to the public domain by waiving all of his or her rights to work worldwide under copyright law for users to modify, distribute and perform the work, even for commercial purposes, all without asking permission

2.3 Information about our dataset:

These datasets were generated by respondents to a distributed survey via Amazon Mechanical Turk between 03.12.2016-05.12.2016. Thirty eligible Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring. Variation between output represents use of distinct types of Fitbit trackers and individual tracking behaviors / preferences.

2.4 Data Organization and verification:

Data contains 18 CSV documents. Each document represents different quantitative data tracked by Fitbit. The data is considered long since each row is one point per subject. Every user has a unique ID and different rows since data is tracked by day and time.

2.5 Data limitations

1. It is a small sample size (33 distinctive users based on data). After slicing and grouping, the size becomes even smaller and may not properly lead to meaningful results.
2. Demographic information, such as gender, age and health condition, is unavailable. As women is the target audience for Bellabeat, data specifically focused on women will be most ideal.
3. The data dated in 2016; thus it may not reflect current trends. To gain more consumer insights, it will be best to collect more up-to-date data.

STEP 3: PROCESS

R programming language was used for data cleaning, manipulation and analysis

The Fitbit dataset does not specify device type. However, because the Time watch tracks activity, sleep, and stress, I focused on those corresponding data fields in the Fitbit dataset: **DAILY ACTIVITY, SLEEPDAY, HOURLYSTEPS AND HOURLY INTENSITY.**

```
## LOADING PACKAGES
library(tidyverse)

library(lubridate)
library(janitor)

library(broom)
library(data.table)

library(scales)

library(readxl)
```

1. LOAD AND READ FILES

```
library(readr)

activity<-read.csv("C:/Users/User/Desktop/DATA ANALYTICS JOURNEY/BELLABEAT/dailyActivity_merged.csv")
head(activity)

##          Id ActivityDate TotalSteps TotalDistance TrackerDistance
## 1 1503960366    4/12/2016      13162         8.50            8.50
## 2 1503960366    4/13/2016      10735         6.97            6.97
## 3 1503960366    4/14/2016      10460         6.74            6.74
## 4 1503960366    4/15/2016       9762         6.28            6.28
## 5 1503960366    4/16/2016      12669         8.16            8.16
## 6 1503960366    4/17/2016       9705         6.48            6.48
##   LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 1                        0              1.88                     0.55
## 2                        0              1.57                     0.69
## 3                        0              2.44                     0.40
## 4                        0              2.14                     1.26
## 5                        0              2.71                     0.41
## 6                        0              3.19                     0.78
##   LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
## 1                6.06                       0                25
## 2                4.71                       0                21
## 3                3.91                       0                30
## 4                2.83                       0                29
## 5                5.04                       0                36
## 6                2.51                       0                38
##   FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories
## 1                  13                  328              728     1985
## 2                  19                  217              776     1797
## 3                  11                  181             1218     1776
## 4                  34                  209              726     1745
## 5                  10                  221              773     1863
## 6                  20                  164              539     1728

sleepday<-read.csv("C:/Users/User/Desktop/DATA ANALYTICS JOURNEY/BELLABEAT/sleepDay_merged.csv")
head(sleepday)

##          Id              SleepDay TotalSleepRecords TotalMinutesAsleep
## 1 1503960366 4/12/2016 12:00:00 AM                1                327
## 2 1503960366 4/13/2016 12:00:00 AM                2                384
## 3 1503960366 4/15/2016 12:00:00 AM                1                412
## 4 1503960366 4/16/2016 12:00:00 AM                2                340
## 5 1503960366 4/17/2016 12:00:00 AM                1                700
## 6 1503960366 4/19/2016 12:00:00 AM                1                304
```

```
##   TotalTimeInBed
## 1            346
## 2            407
## 3            442
## 4            367
## 5            712
## 6            320
```

```
hourlysteps<-read.csv("C:/Users/User/Desktop/DATA ANALYTICS JOURNEY/BELLABEAT/hourlySteps_merged.csv")
head(hourlysteps)
```

```
##           Id   ActivityHour StepTotal
## 1 1503960366 3/12/2016 0:00         0
## 2 1503960366 3/12/2016 1:00         0
## 3 1503960366 3/12/2016 2:00         0
## 4 1503960366 3/12/2016 3:00         0
## 5 1503960366 3/12/2016 4:00         0
## 6 1503960366 3/12/2016 5:00         0
```

```
intensity<-read.csv("C:/Users/User/Desktop/DATA ANALYTICS JOURNEY/BELLABEAT/hourlyIntensities_merged.csv")
head(intensity)
```

```
##           Id   ActivityHour TotalIntensity AverageIntensity
## 1 1503960366 3/12/2016 0:00             0                 0
## 2 1503960366 3/12/2016 1:00             0                 0
## 3 1503960366 3/12/2016 2:00             0                 0
## 4 1503960366 3/12/2016 3:00             0                 0
## 5 1503960366 3/12/2016 4:00             0                 0
## 6 1503960366 3/12/2016 5:00             0                 0
```

2. **CHANGE DATE FORMATS**

Daily steps,sleepday,hourlysteps,hourlyintensity

```
if(!"ActivityDate" %in% names(activity)) stop("Check activity file: column ActivityDate expected")
activity <- activity %>% mutate(date = mdy(ActivityDate))

if(!"SleepDay" %in% names(sleepday)) stop("Check sleepday file: column SleepDay expected")
sleepday <- sleepday %>% mutate(SleepDay = parse_date_time(SleepDay, orders = c("mdy HMS","mdy","ymd HMS","ymd")),
                                date = as_date(SleepDay),
                                sleep_hours = TotalMinutesAsleep/60,
                                time_in_bed_hours = TotalTimeInBed/60)
if(!"ActivityHour" %in% names(hourlysteps)) stop("Check hourlySteps file: column ActivityHour exp
```

```
ected")
hourlysteps <- hourlysteps %>%
  mutate(ActivityHour = parse_date_time(ActivityHour, orders = c("ymd HMS","mdy HMS","ymd H
M","mdy HM")),
         hour = hour(ActivityHour),
         date = as_date(ActivityHour))
intensity <- intensity %>%
  mutate(ActivityHour = parse_date_time(ActivityHour, orders = c("ymd HMS","mdy HMS","ymd H
M","mdy HM")),
         hour = hour(ActivityHour),
         date = as_date(ActivityHour))
```

3. CREATE DERIVED VARIABLES

Make active minutes, sleep hours, weekday/weekend

```
library(lubridate)

activity <- activity %>%
  mutate(
    VeryActiveMinutes  = coalesce(VeryActiveMinutes, 0),
    FairlyActiveMinutes = coalesce(FairlyActiveMinutes, 0),
    LightlyActiveMinutes = coalesce(LightlyActiveMinutes, 0),
    active_minutes = VeryActiveMinutes + FairlyActiveMinutes + LightlyActiveMinutes,
    total_steps = coalesce(TotalSteps,0),
    calories = coalesce(Calories, 0))

activity <- activity %>%
  mutate(date = mdy(ActivityDate),
         weekday = lubridate:: wday(date,label = TRUE, week_start = 1))



sleepday <- sleepday %>% mutate(sleephours = TotalMinutesAsleep/60)
head(sleepday)

##          Id          SleepDay TotalSleepRecords TotalMinutesAsleep
## 1 1503960366 2016-04-12 12:00:00                1               327
## 2 1503960366 2016-04-13 12:00:00                2               384
## 3 1503960366 2016-04-15 12:00:00                1               412
## 4 1503960366 2016-04-16 12:00:00                2               340
## 5 1503960366 2016-04-17 12:00:00                1               700
## 6 1503960366 2016-04-19 12:00:00                1               304
##   TotalTimeInBed       date sleep_hours time_in_bed_hours sleephours
## 1            346 2016-04-12    5.450000          5.766667   5.450000
## 2            407 2016-04-13    6.400000          6.783333   6.400000
## 3            442 2016-04-15    6.866667          7.366667   6.866667
```

```
## 4          367 2016-04-16    5.666667          6.116667   5.666667
## 5          712 2016-04-17   11.666667         11.866667  11.666667
## 6          320 2016-04-19    5.066667          5.333333   5.066667
```

4. DATA QUALITY CHECKS

A. Standardize column names

```
library(janitor)
activity <- clean_names(activity)
sleepday <- clean_names(sleepday)
hourlysteps <- clean_names(hourlysteps)
intensity <- clean_names(intensity)
colnames(activity)
```

```
##  [1] "id"                       "activity_date"
##  [3] "total_steps"                "total_distance"
##  [5] "tracker_distance"           "logged_activities_distance"
##  [7] "very_active_distance"       "moderately_active_distance"
##  [9] "light_active_distance"      "sedentary_active_distance"
## [11] "very_active_minutes"        "fairly_active_minutes"
## [13] "lightly_active_minutes"     "sedentary_minutes"
## [15] "calories"                   "date"
## [17] "active_minutes"             "total_steps_2"
## [19] "calories_2"                 "weekday"
```

```
colnames(sleepday)
```

```
## [1] "id"                   "sleep_day"            "total_sleep_records"
## [4] "total_minutes_asleep" "total_time_in_bed"    "date"
## [7] "sleep_hours"          "time_in_bed_hours"    "sleephours"
```

```
colnames(hourlysteps)
```

```
## [1] "id"            "activity_hour" "step_total"    "hour"
## [5] "date"
```

```
colnames(intensity)
```

```
## [1] "id"                "activity_hour"     "total_intensity"
## [4] "average_intensity" "hour"              "date"
```

B. REMOVE DUPLICATES

```
sum(duplicated(activity))
```

```
## [1] 0
```

```
sum(duplicated(sleepday))
```

```
## [1] 3
```

```
sum(duplicated(intensity))
```

```
## [1] 0
```

```
sum(duplicated(hourlysteps))
```

```
## [1] 0
```

```
## Removed the 3 duplicates in sleepday dataset ##
sleepday <- sleepday %>% distinct()
sum(duplicated(sleepday))
```

```
## [1] 0
```

No duplicates found in activity, 3 duplicates in sleepday,0 duplicates in Hourly intensity and Hourly Steps

STEP 4: ANALYZE

DAILY ACTIVITY SUMMARY

```
# Finding number of participants in each category
n_distinct(activity$id)
```

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

```
n_distinct(intensity$id)
```

```
## [1] 34
```

```
n_distinct(hourlysteps$id)
```

```
## [1] 34
```

```
n_distinct(sleepday$id)
```

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

To summarize the above data, there are 33 participants in the activity, 34 in hourly intensities and hourly steps datasets and 24 in the sleep dataset. Although the sleepDay dataset contains data for only 24 participants compared to 33–34 in the other datasets, it offers critical insights into user sleep patterns, a key wellness dimension that complements activity and stress metrics tracked by the Bellabeat Time watch. While limited in size, the dataset is sufficiently robust for exploratory analysis, especially when combined with daily activity data.

```
# Summarize daily totals


daily_user_summary <- activity %>%
  group_by(date) %>%
```

```
  summarise(
    daily_steps = sum(total_steps, na.rm = TRUE),
    daily_active_minutes = sum(active_minutes, na.rm = TRUE),
    daily_calories = sum(calories, na.rm = TRUE),
    daily_sedentary = sum(sedentary_minutes, na.rm = TRUE)
  )

summary(daily_user_summary)

##       date             daily_steps     daily_active_minutes daily_calories
##  Min.   :2016-04-12   Min.   : 73129   Min.   :2208         Min.   :23925
##  1st Qu.:2016-04-19   1st Qu.:207128   1st Qu.:6270         1st Qu.:66562
##  Median :2016-04-27   Median :242996   Median :7182         Median :73592
##  Mean   :2016-04-27   Mean   :231601   Mean   :6900         Mean   :69851
##  3rd Qu.:2016-05-04   3rd Qu.:256548   3rd Qu.:7592         3rd Qu.:75475
##  Max.   :2016-05-12   Max.   :277733   Max.   :8535         Max.   :78893
##  daily_sedentary
##  Min.   :13692
##  1st Qu.:29248
##  Median :31415
##  Mean   :30056
##  3rd Qu.:32917
##  Max.   :33959

# Sleep patterns

sleepday %>%
  select(total_sleep_records, total_minutes_asleep, total_time_in_bed) %>%
  summary()

##  total_sleep_records total_minutes_asleep total_time_in_bed
##  Min.   :1.00        Min.   : 58.0        Min.   : 61.0
##  1st Qu.:1.00        1st Qu.:361.0        1st Qu.:403.8
##  Median :1.00        Median :432.5        Median :463.0
##  Mean   :1.12        Mean   :419.2        Mean   :458.5
##  3rd Qu.:1.00        3rd Qu.:490.0        3rd Qu.:526.0
##  Max.   :3.00        Max.   :796.0        Max.   :961.0
```

Observations made from the above summary

Data collected between April 12, 2016, and May 12, 2016 (approximately one month), a relatively short but consistent data collection period. The analysis of the daily user summary shows that, on average, each participant recorded approximately 7,638 steps per day, engaged in about 227 minutes (nearly 4 hours) of active movement, and spent around 991 minutes (approximately 16.5 hours) in sedentary behavior. Users burned an average of 2,304 calories daily. The median step count of 7,406 suggests that most users met or came close to the commonly recommended 7,000–8,000 daily steps associated with moderate physical activity. However, the high average sedentary time indicates that even active users spent a substantial portion of their day inactive, which may reflect work or rest habits. The sleep data

indicates that users averaged approximately 419 minutes (around 7 hours) of sleep per night, spending about 459 minutes (7.6 hours) in bed overall.

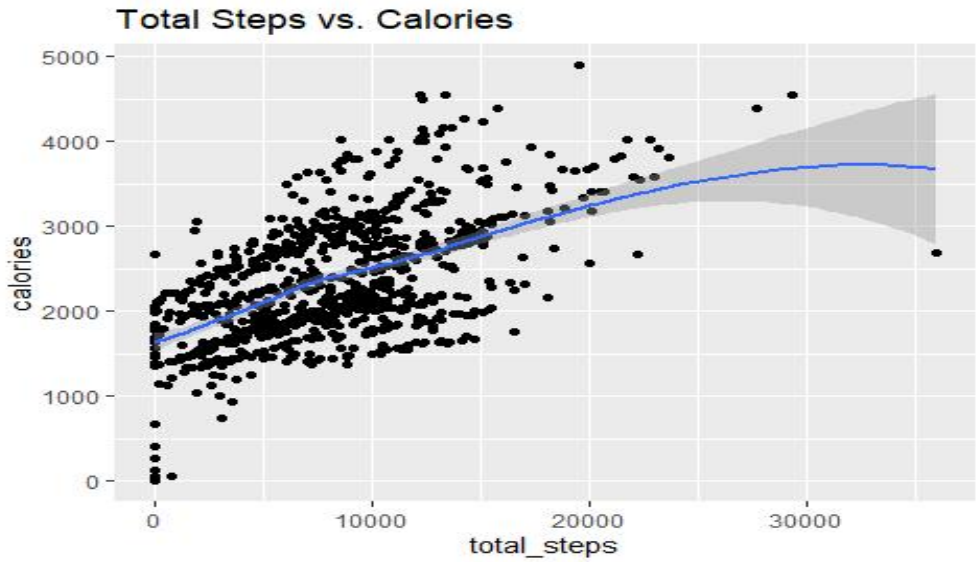
STEP 5: SHARE

```
# Plot daily steps
daily_summary <- activity %>%
  group_by(date) %>%
  summarise(
    total_steps = sum(total_steps, na.rm = TRUE),
    total_active_minutes = sum(active_minutes, na.rm = TRUE),
    total_calories = sum(calories, na.rm = TRUE)
  )
ggplot(daily_summary, aes(x = date, y = total_steps)) +
  geom_line(color = "steelblue") +
  labs(title = "Daily Steps Over Time", x = "Date", y = "Total Steps") +
  theme_minimal()
```

The trend in daily steps over time above shows that total step counts generally remained consistent and relatively high throughout April, averaging between 230,000 and 270,000 total steps per day. However, starting in early May, there is a noticeable downward trend, with step counts gradually declining and a sharp drop in mid-May, possibly indicating reduced user engagement or missing data toward the end of the period

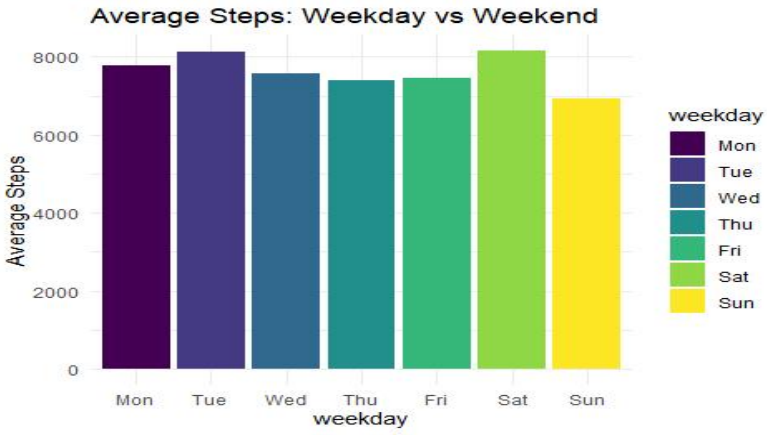
Steps vs. Calories

```
ggplot(data = activity, aes(x = total_steps, y = calories)) + geom_point() + geom_smooth() + labs(title = "Total Steps vs. Calories")

## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



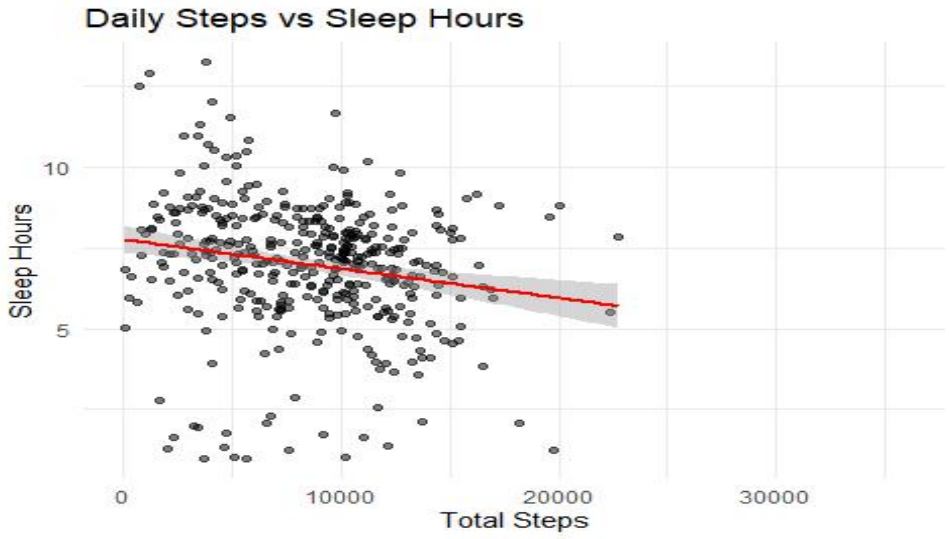
The scatter plot shows a positive correlation between total steps and calories burned, indicating that users who take more steps tend to burn more calories. The relationship appears strongest up to around 20,000 steps.

```
# Activity by Weekday vs Weekend


activity_by_daytype <- activity %>%
  group_by(weekday) %>%
  summarise(
    avg_steps = mean(total_steps, na.rm = TRUE),
    avg_active_minutes = mean(active_minutes, na.rm = TRUE),
    avg_calories = mean(calories, na.rm = TRUE)
  )
ggplot(activity_by_daytype, aes(x = weekday, y = avg_steps, fill = weekday)) +
  geom_col() +
  labs(title = "Average Steps: Weekday vs Weekend", y = "Average Steps") +
  theme_minimal()
```



As shown above, users are generally more active on Saturdays, and Tuesdays and least active on Sundays, thus target campaigns are needed to encourage more activity throughout the week.

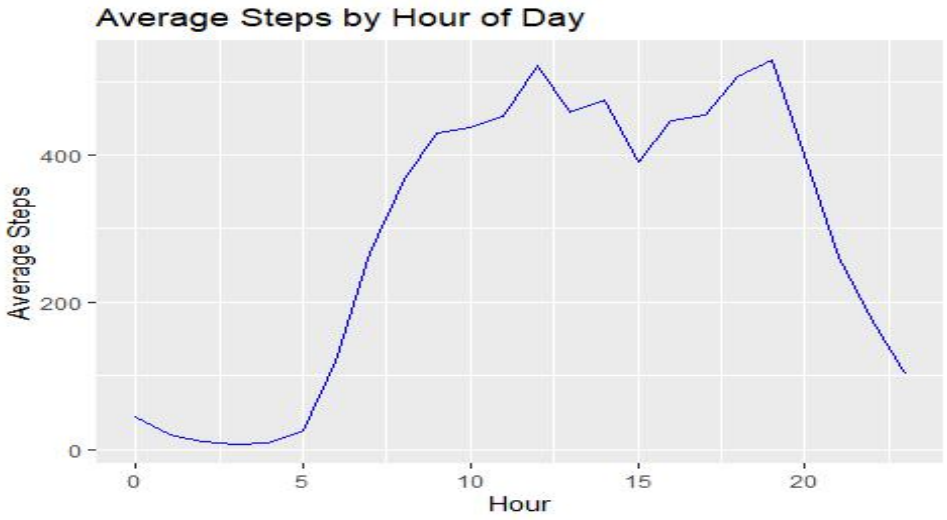
```
# Sleep summary
sleep_summary <- sleepday %>%
  group_by(date) %>%
  summarise(
    avg_sleep_hours = mean(sleep_hours, na.rm = TRUE),
    avg_time_in_bed = mean(time_in_bed_hours, na.rm = TRUE)
  )
sleep_activity <- activity %>%
  left_join(sleepday %>% select(id, date, sleep_hours), by = c("id", "date"))

cor(sleep_activity$total_steps, sleep_activity$sleep_hours, use = "complete.obs")

## [1] -0.1903439

# Activity vs Sleep Correlation

ggplot(sleep_activity, aes(x = total_steps, y = sleep_hours)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "lm", color = "red") +
  labs(title = "Daily Steps vs Sleep Hours", x = "Total Steps", y = "Sleep Hours") +
  theme_minimal()
```



Based on the scatter plot above, there is a negative correlation between sleeping hours and daily steps, as users sleep more hours, total steps decreased. The plot also shows that users can achieve the 10000 steps recommended and also achieve the daily recommended 8 hours of sleep.

```
hourlysteps %>%
  group_by(hour) %>%
```

```
summarise(avg_steps = mean(step_total, na.rm=TRUE)) %>%
ggplot(aes(x=hour, y=avg_steps)) +
geom_line(color="blue") +
labs(title="Average Steps by Hour of Day", x="Hour", y="Average Steps")
```



The line plot of average steps by hour of day above shows clear patterns in user activity:

- Low activity during early morning hours (midnight–6 AM): Users are largely inactive, as expected during sleep hours.
- Morning increase (6–10 AM): Steps rise sharply, likely due to morning routines, commuting, or exercise.
- Midday plateau (10 AM–2 PM): Users maintain moderate activity levels, possibly reflecting daily movement at work or school.
- Afternoon/evening peak (4–8 PM): Highest activity occurs during late afternoon and early evening, likely reflecting post-work exercise, errands, or leisure activities.
- Decline after 8 PM: Steps drop sharply as users prepare for the end of the day.
- Thus,users are most active during morning and late afternoon/evening hours, with minimal activity during early morning and late night.

STEP 6: ACT

After analyzing the FitBit Fitness Tracker data focusing on Bellabeat's Time wellness, I came up with some recommendations for Bellabeat marketing strategy based on trends in smart device usage.

1. As stated previously, the average number of steps per day is 7,638. This is lower than what the CDC recommends (www.nih.gov/news-events/nih-research-matters/how-many-steps-better-health). CDC recommends 10,000 steps per day in mantaining good health.Further, sedentary remains high,thus Bellabeat can promote reminders for movement during sedentary hours and messages on the benefits of exercising.

2. A decline in daily steps from April to May can be explained by a decline in user engagement since users tend to show high enthusiasm early in fitness tracking, followed by a decline as motivation decreases or novelty wears off. Bellabeat could address this decline by integrating personalized reminders, goal adjustments, or gamified challenges to sustain engagement over time. Providing insights and encouragement when activity drops could help users stay consistent in their wellness habits.

3. Analysis of user activity data reveals that users are most active on Saturdays, with moderate activity levels during weekdays. Based on this insight, the following recommendations are proposed to enhance Bellabeat's marketing strategy and product engagement:

a. **Leverage Weekend Engagement** :Launch targeted campaigns such as a "Weekend Wellness Challenge" to encourage users to maximize activity on Saturdays and Sundays. Send app notifications or tips on Friday afternoons or Saturday mornings to motivate weekend activity.

b. **Encourage Weekday Activity** :Share bite-sized wellness content, such as short workouts or mindfulness exercises, to boost weekday engagement.

4. There was a negative correlation between sleep duration and daily steps,however, some users are still able to achieve the recommended 10,000 daily steps while also reaching the 8 hours of sleep guideline. Based on this insight, the following recommendations are proposed for Bellabeat:

a. Emphasize in marketing campaigns that users can maintain both healthy sleep patterns and high daily activity, reinforcing Bellabeat's holistic wellness mission.

b. Share in-app tips and content on integrating movement efficiently into daily routines without sacrificing rest.

c. Provide adaptive step goals based on individual sleep patterns, helping users maintain activity while respecting their rest needs.

d. Send gentle reminders or nudges to encourage a balanced approach to activity and recovery.

5. Users are most active during morning and late afternoon/evening hours, with minimal activity during early morning and late night, thus I reccomend Bellabeat to Optimize Notifications and Reminders

a. Send motivational messages, step reminders, or app nudges during peak activity hours (morning and late afternoon) to reinforce healthy habits.

b. Avoid sending alerts during low-activity periods (midnight–6 AM), as they are unlikely to engage users.

c. Introduce morning or evening step challenges to capitalize on natural activity patterns.