Bellabeat Case Study

Rahmanudeen

About company

Bellabeat, a high-tech manufacturer of health-focused products 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

Questions for the analysis

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

Business task

Identify potential opportunities for growth and recommendations for the Bellabeat marketing strategy improvement based on trends in smart device usage.

Installing and Loading Packages

```r
# Run these once in your R console (not every time you knit)
# install.packages("tidyverse")
# install.packages("ggplot2")
# install.packages("lubridate")
# install.packages("tidyr")
# install.packages("dplyr")

library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.2     v tibble    3.2.1
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.4
## -- Conflicts ----------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(lubridate)
library(dplyr)
library(ggplot2)
library(tidyr)
```

Importing the Data

```
activity <- read.csv("/Users/rahu/RStudio/dailyActivity_merged.csv")
calories <- read.csv("/Users/rahu/RStudio/hourlyCalories_merged.csv")
intensities <- read.csv("/Users/rahu/RStudio/hourlyIntensities_merged.csv")
sleep <- read.csv("/Users/rahu/RStudio/sleepDay_merged.csv")
weight <- read.csv("/Users/rahu/RStudio/weightLogInfo_merged.csv")
```

Data Cleaning and Date Formatting

```
# Intensities
intensities$ActivityHour <- as.POSIXct(intensities$ActivityHour, format="%m/%d/%Y %I:%M:%S %p", tz=Sys.
intensities$time <- format(intensities$ActivityHour, format = "%H:%M:%S")
intensities$date <- format(intensities$ActivityHour, format = "%m/%d/%y")

# Calories
calories$ActivityHour <- as.POSIXct(calories$ActivityHour, format="%m/%d/%Y %I:%M:%S %p", tz=Sys.timezo
calories$time <- format(calories$ActivityHour, format = "%H:%M:%S")
calories$date <- format(calories$ActivityHour, format = "%m/%d/%y")

# Activity
activity$ActivityDate <- as.POSIXct(activity$ActivityDate, format="%m/%d/%Y", tz=Sys.timezone())
activity$date <- format(activity$ActivityDate, format = "%m/%d/%y")

# Sleep
sleep$SleepDay <- as.POSIXct(sleep$SleepDay, format="%m/%d/%Y %I:%M:%S %p", tz=Sys.timezone())
sleep$date <- format(sleep$SleepDay, format = "%m/%d/%y")
```

Exploring the Data

Unique Users per Dataset

```
n_distinct(activity$Id)
```

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

```
n_distinct(calories$Id)
```

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

```
n_distinct(intensities$Id)
```

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

```
n_distinct(sleep$Id)
```

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

```
n_distinct(weight$Id)
```

```
## [1] 8
```

Summary: Activity

```
activity %>%
  select(TotalSteps, TotalDistance, SedentaryMinutes, Calories) %>%
  summary()
```

```
##    TotalSteps     TotalDistance    SedentaryMinutes    Calories
##  Min.   :    0   Min.   : 0.000   Min.   :   0.0   Min.   :   0
##  1st Qu.: 3790   1st Qu.: 2.620   1st Qu.: 729.8   1st Qu.:1828
##  Median : 7406   Median : 5.245   Median :1057.5   Median :2134
##  Mean   : 7638   Mean   : 5.490   Mean   : 991.2   Mean   :2304
##  3rd Qu.:10727   3rd Qu.: 7.713   3rd Qu.:1229.5   3rd Qu.:2793
##  Max.   :36019   Max.   :28.030   Max.   :1440.0   Max.   :4900
```

Summary: Active Minutes

```
activity %>%
  select(VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes) %>%
  summary()
```

```
##  VeryActiveMinutes FairlyActiveMinutes LightlyActiveMinutes
##  Min.   :  0.00    Min.   :  0.00      Min.   :  0.0
##  1st Qu.:  0.00    1st Qu.:  0.00      1st Qu.:127.0
##  Median :  4.00    Median :  6.00      Median :199.0
##  Mean   : 21.16    Mean   : 13.56      Mean   :192.8
##  3rd Qu.: 32.00    3rd Qu.: 19.00      3rd Qu.:264.0
##  Max.   :210.00    Max.   :143.00      Max.   :518.0
```

Summary: Calories

```
calories %>%
  select(Calories) %>%
  summary()
```

```
##     Calories
##  Min.   : 42.00
##  1st Qu.: 63.00
##  Median : 83.00
##  Mean   : 97.39
##  3rd Qu.:108.00
##  Max.   :948.00
```

Summary: Sleep

```
sleep %>%
  select(TotalSleepRecords, TotalMinutesAsleep, TotalTimeInBed) %>%
  summary()
```

```
##  TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
##  Min.   :1.000     Min.   : 58.0      Min.   : 61.0
##  1st Qu.:1.000     1st Qu.:361.0      1st Qu.:403.0
##  Median :1.000     Median :433.0      Median :463.0
##  Mean   :1.119     Mean   :419.5      Mean   :458.6
##  3rd Qu.:1.000     3rd Qu.:490.0      3rd Qu.:526.0
##  Max.   :3.000     Max.   :796.0      Max.   :961.0
```

Summary: Weight

```
weight %>%
  select(WeightKg, BMI) %>%
  summary()
```

```
##     WeightKg          BMI
##  Min.   : 52.60   Min.   :21.45
##  1st Qu.: 61.40   1st Qu.:23.96
##  Median : 62.50   Median :24.39
##  Mean   : 72.04   Mean   :25.19
##  3rd Qu.: 85.05   3rd Qu.:25.56
##  Max.   :133.50   Max.   :47.54
```

Merging Sleep and Activity Data

```
merged_data <- merge(sleep, activity, by = c("Id", "date"))
head(merged_data)
```

```
##           Id     date   SleepDay TotalSleepRecords TotalMinutesAsleep
## 1 1503960366 04/12/16 2016-04-12                 1                327
## 2 1503960366 04/13/16 2016-04-13                 2                384
## 3 1503960366 04/15/16 2016-04-15                 1                412
## 4 1503960366 04/16/16 2016-04-16                 2                340
## 5 1503960366 04/17/16 2016-04-17                 1                700
```

```
## 6 1503960366 04/19/16 2016-04-19                          1          304
##   TotalTimeInBed ActivityDate TotalSteps TotalDistance TrackerDistance
## 1           346   2016-04-12      13162          8.50            8.50
## 2           407   2016-04-13      10735          6.97            6.97
## 3           442   2016-04-15       9762          6.28            6.28
## 4           367   2016-04-16      12669          8.16            8.16
## 5           712   2016-04-17       9705          6.48            6.48
## 6           320   2016-04-19      15506          9.88            9.88
##   LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 1                        0               1.88                     0.55
## 2                        0               1.57                     0.69
## 3                        0               2.14                     1.26
## 4                        0               2.71                     0.41
## 5                        0               3.19                     0.78
## 6                        0               3.53                     1.32
##   LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
## 1                6.06                       0                25
## 2                4.71                       0                21
## 3                2.83                       0                29
## 4                5.04                       0                36
## 5                2.51                       0                38
## 6                5.03                       0                50
##   FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories
## 1                  13                  328              728     1985
## 2                  19                  217              776     1797
## 3                  34                  209              726     1745
## 4                  10                  221              773     1863
## 5                  20                  164              539     1728
## 6                  31                  264              775     2035
```

Visualizations

Total Steps vs. Calories

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

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```
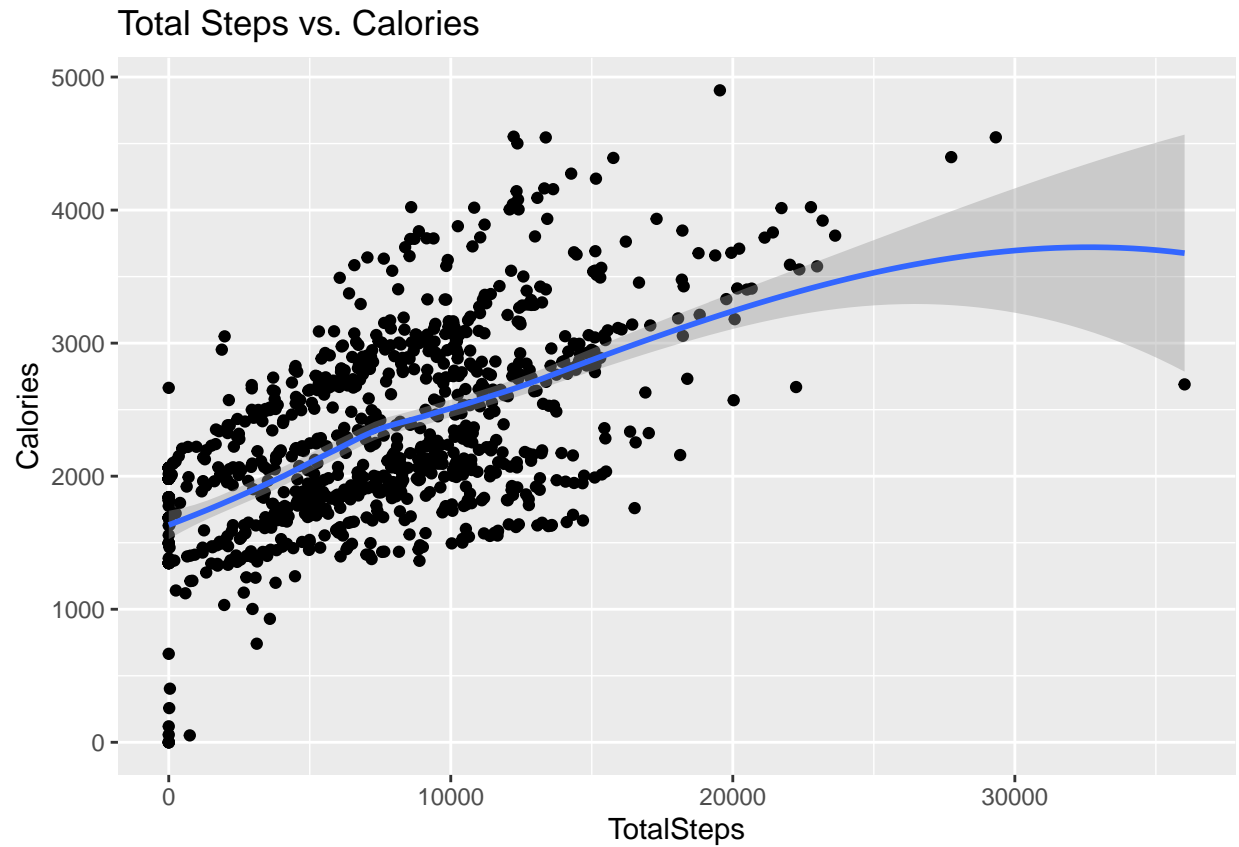
I see positive correlation here between Total Steps and Calories, which is obvious - the more active we are, the more calories we burn.

Minutes Asleep vs. Time in Bed

```
ggplot(data = sleep, aes(x = TotalMinutesAsleep, y = TotalTimeInBed)) +
  geom_point() +
  labs(title = "Total Minutes Asleep vs. Total Time in Bed")
```

Total Minutes Asleep vs. Total Time in Bed
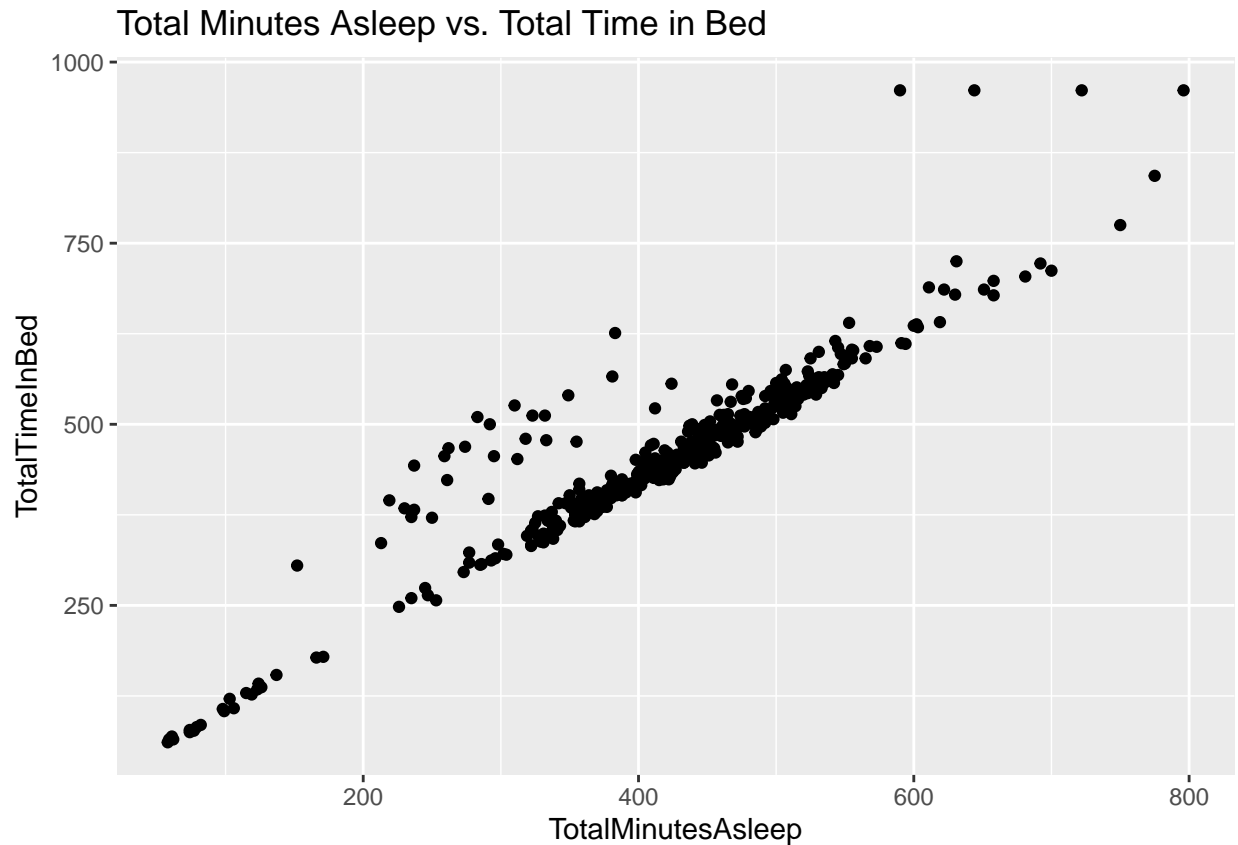
The relationship between Total Minutes Asleep and Total Time in Bed looks linear. So if the Bellabeat users want to improve their sleep, we should consider using notification to go to sleep.

Let's look at intensities data over time (hourly).

Average Intensity by Time of Day

```r
int_new <- intensities %>%
  group_by(time) %>%
  drop_na() %>%
  summarise(mean_total_int = mean(TotalIntensity))

ggplot(data = int_new, aes(x = time, y = mean_total_int)) +
  geom_col(fill = "darkblue") +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(title = "Average Total Intensity vs. Time", x = "Time", y = "Mean Total Intensity")
```

Average Total Intensity vs. Time
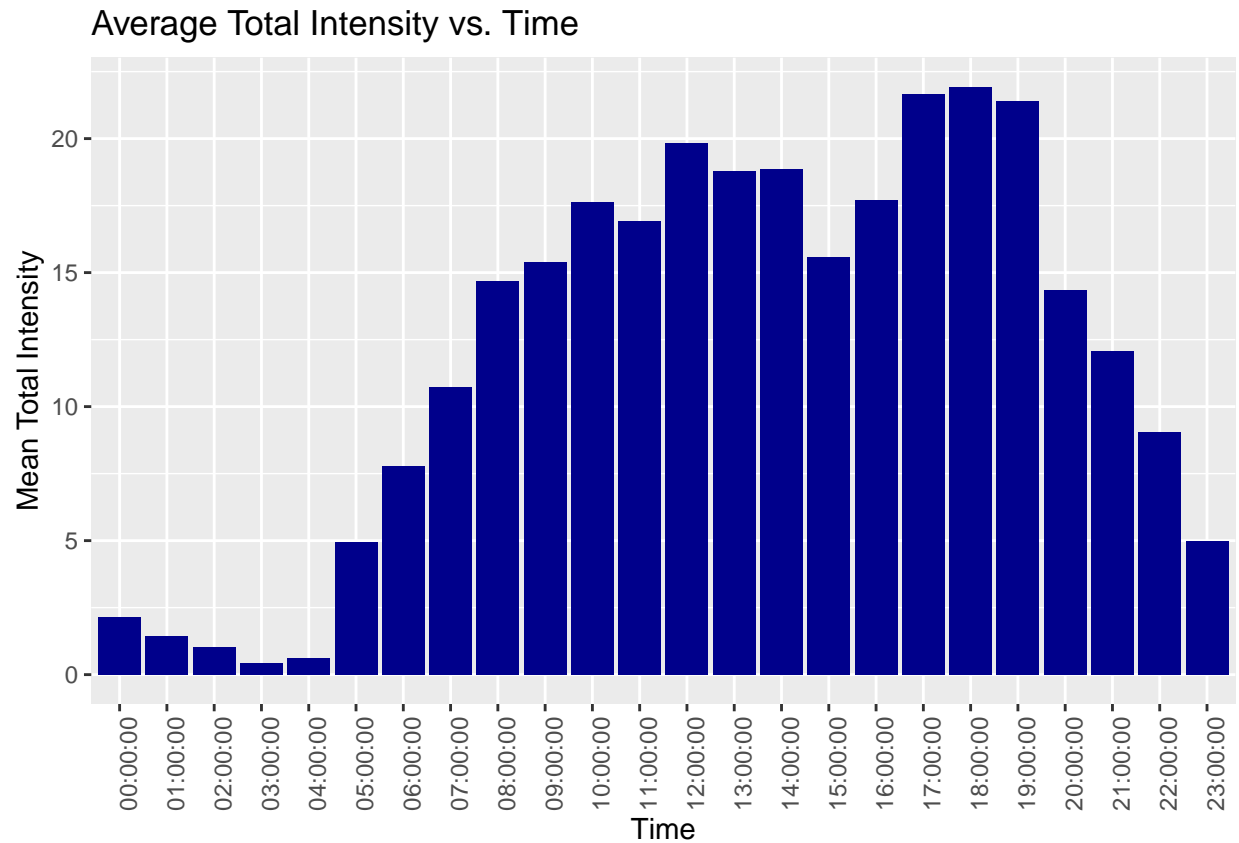
After visualizing Total Intensity hourly, I found out that people are more active between 5 am and 10pm.

Most activity happens between 5 pm and 7 pm - I suppose, that people go to a gym or for a walk after finishing work. We can use this time in the Bellabeat app to remind and motivate users to go for a run or walk.

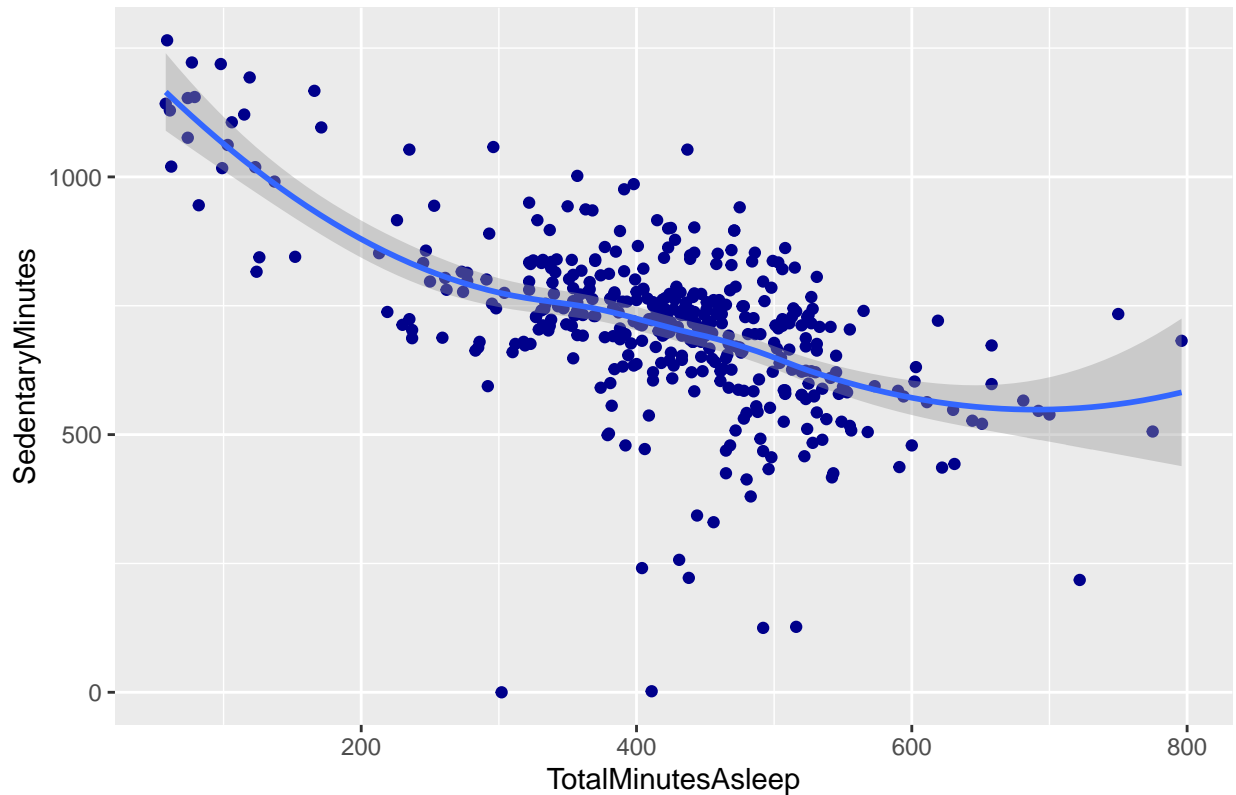Let's look at the relationship between Total Minutes Asleep and Sedentry Minutes.

Sleep vs. Sedentary Minutes

```
ggplot(data = merged_data, aes(x = TotalMinutesAsleep, y = SedentaryMinutes)) +
  geom_point(color = 'darkblue') +
  geom_smooth() +
  labs(title = "Minutes Asleep vs. Sedentary Minutes")
```

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

Minutes Asleep vs. Sedentary Minutes

Here we can clearly see the negative relationship between Sedentary Minutes and Sleep time.

As an idea: if Bellabeat users want to improve their sleep, Bellabeat app can recommend reducing sedentary time.

Keep in mind that we need to support this insights with more data, because correlation between some data doesn't mean causation.

Summarizing recommendations for the business

As we already know, 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.

After analyzing FitBit Fitness Tracker Data, I found some insights that would help influence Bellabeat marketing strategy.

Target audience

Women who work full-time jobs (according to the hourly intensity data) and spend a lot of time at the computer/in a meeting/ focused on work they are doing (according to the sedentary time data).

These women do some light activity to stay healthy (according to the activity type analysis). Even though they need to improve their everyday activity to have health benefits. They might need some knowledge about developing healthy habits or motivation to keep going.

As there is no gender information about the participants, I assumed that all genders were presented and balanced in this data set. The key message for the Bellabeat online campaign

The Bellabeat app is not just another fitness activity app. It's a guide (a friend) who empowers women to balance full personal and professional life and healthy habits and routines by educating and motivating them through daily app recommendations.

Ideas for the Bellabeat app

Average total steps per day are 7638 which a little bit less for having health benefits for according to the CDC research. They found that taking 8,000 steps per day was associated with a 51% lower risk for all-cause mortality (or death from all causes). Taking 12,000 steps per day was associated with a 65% lower risk compared with taking 4,000 steps. Bellabeat can encourage people to take at least 8 000 explaining the benefits for their health.

If users want to lose weight, it's probably a good idea to control daily calorie consumption. Bellabeat can suggest some ideas for low-calorie lunch and dinner.

If users want to improve their sleep, Bellabeat should consider using app notifications to go to bed.

Most activity happens between 5 pm and 7 pm - I suppose, that people go to a gym or for a walk after finishing work. Bellabeat can use this time to remind and motivate users to go for a run or walk.

As an idea: if users want to improve their sleep, the Bellabeat app can recommend reducing sedentary time.

Thank you for your interest to my Bellabeat Case Study!

This is my first project using R. I would appreciate any comments and recommendations for improvement!