

Bellabeat Case Study

Anatole Wilson

2023-06-10

About this Scenario

As a data analyst for the marketing team for Bellabeat, a high-tech manufacturer of health-focused products for women, I am looking at ways that this successful small marketing company can become a larger player in the global smart device market. I will be looking at the smart device data from Bellabeat's products, and selecting one product from the data, I will recommend some steps the company can take to grow the support for the device and help grow the company.

Questions for this Analysis

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?

Business Task

Identify potential opportunities and make some recommendations that will increase Bellabeat usage.

Preparing the data for RStudio

Loading Packages

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.2      v readr      2.1.4
## v forcats    1.0.0      v stringr    1.5.0
## v ggplot2    3.4.2      v tibble     3.2.1
## v lubridate  1.9.2      v tidyr      1.3.0
## v purrr      1.0.1
## -- 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 datasets

For this project, I will use the [FitBit Tracker Data][id].

```
activity <- read.csv("C:/Fitabase Data 4.12.16-5.12.16/dailyActivity_merged.csv")
calories <- read.csv("C:/Fitabase Data 4.12.16-5.12.16/hourlyCalories_merged.csv")
intensities <- read.csv("C:/Fitabase Data 4.12.16-5.12.16/hourlyIntensities_merged.csv")
sleep <- read.csv("C:/Fitabase Data 4.12.16-5.12.16/sleepDay_merged.csv")
weight <- read.csv("C:/Fitabase Data 4.12.16-5.12.16/weightLogInfo_merged.csv")
```

A quick check using the head() function to make sure that everything was imported correctly.

```
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
```

```
head(intensities)
```

```
##           Id           ActivityHour TotalIntensity AverageIntensity
## 1 1503960366 4/12/2016 12:00:00 AM          20          0.333333
## 2 1503960366 4/12/2016 1:00:00 AM           8          0.133333
## 3 1503960366 4/12/2016 2:00:00 AM           7          0.116667
## 4 1503960366 4/12/2016 3:00:00 AM           0          0.000000
## 5 1503960366 4/12/2016 4:00:00 AM           0          0.000000
## 6 1503960366 4/12/2016 5:00:00 AM           0          0.000000
```

There are some problems with the ActivityHour data. The data needs to be converted to a date/time format and then split into separate date and time categories.

```
intensities$ActivityHour=as.POSIXct(intensities$ActivityHour, format="%m/%d/%Y %I:%M:%S %p", tz=Sys.timezone())
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.timezone())
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 and Summarizing Data

First, we'll see how many people were involved in each of the studies.

```
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
```

While we can make some very limited assumptions about the number of people in the first four studies, 8 people is not enough to form any conclusions about the weight category.

Here we'll take a look at some of the statistics of the data.

```
# 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
```

```
# explore num of active minutes per category
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
```

```
# 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
```

```
# 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
```

```
# 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
```

Looking at the data, we can see:

- Most participants are only lightly active.
- The average total steps per day is 7638, which is almost twice as much as the average person walks, but less than the 10,000 steps a day (or 150 minutes) of walking that is recommended by the CDC.
- The average sleep time is seven hours, but few FitBit wearers recorded their sleep.
- Sleepers show that they spend a half hour more in bed than they sleep.

*** Merging Data

To visualize the data sets, I'm going to merge the activity and sleep data sets using an inner join.

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

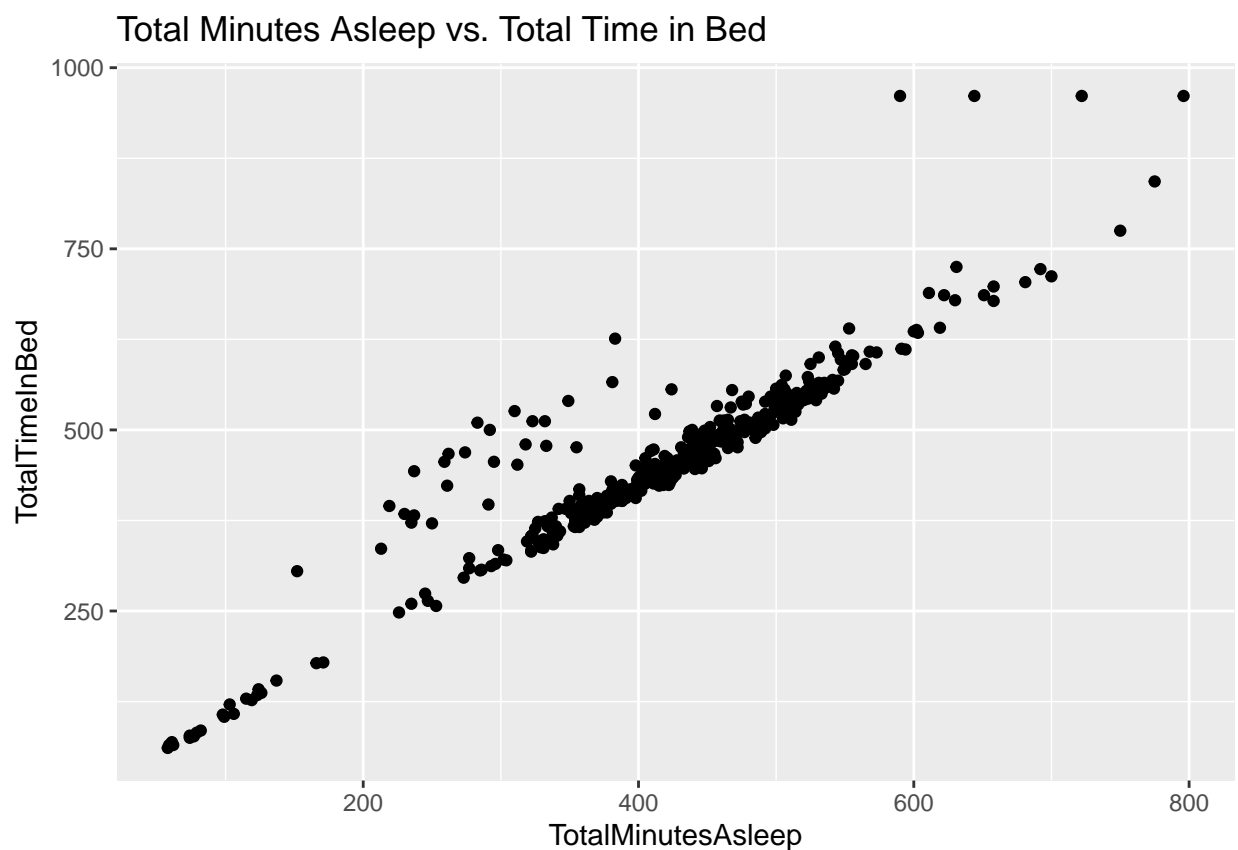
```
##      Id      date  SleepDay TotalSleepRecords TotalMinutesAsleep
## 1 1503960366 04/12/16 2016-04-12             1                327
## 2 1503960366 04/13/16 2016-04-13             2                384
## 3 1503960366 04/15/16 2016-04-15             1                412
## 4 1503960366 04/16/16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

The first visualization shows the difference between the total time in bed and the total minutes asleep. You can see there can be a notable difference between the two. This could be an area where we could use a notification to help people spend less time in bed.

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

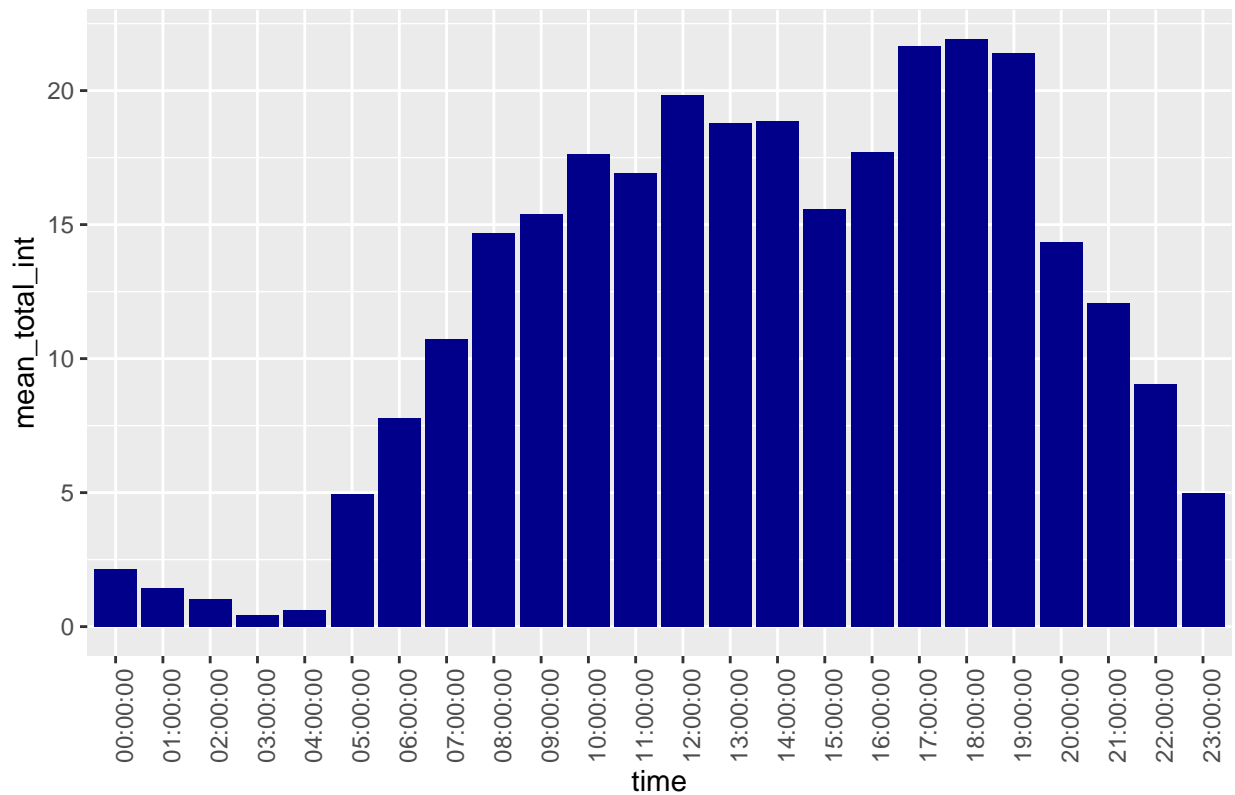


```
int_new <- intensities %>%
  group_by(time) %>%
  drop_na() %>%
  summarise(mean_total_int = mean(TotalIntensity))

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

```
## Warning in geom_histogram(stat = "identity", fill = "darkblue"): Ignoring
## unknown parameters: 'binwidth', 'bins', and 'pad'
```

Average Total Intensity vs. Time



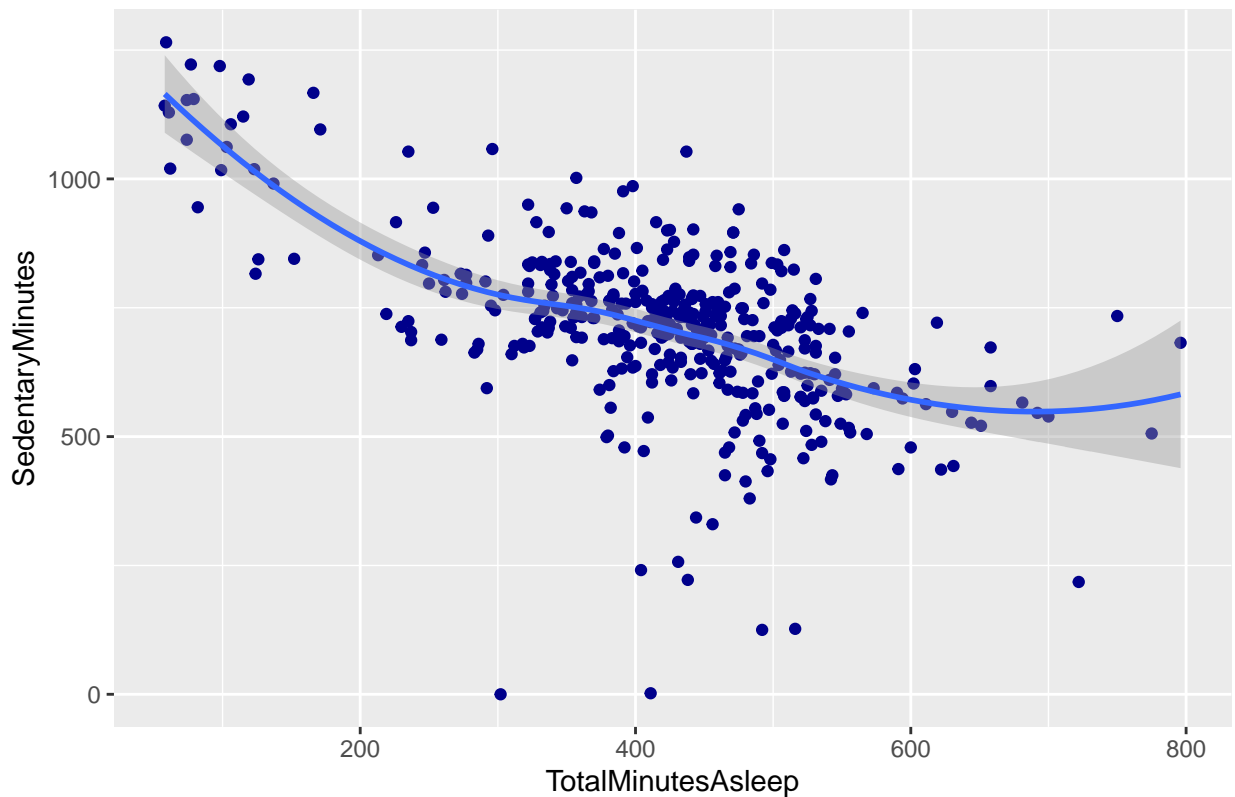
- People are most active between Noon and two and five PM and seven PM.

Lets look at the relationship between Total Minutes Asleep and 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



- There's a negative relationship between Sedentary Minutes and Sleep Time

Summarizing recommendations

Collecting data like this on FitBit and Bellabeat users has allowed Bellabeat to empower women with knowledge about their own health and habits. Since its founding, Bellabeat has grown rapidly and quickly positioned itself as a tech-driven wellness company for women.

After analyzing the Fitbit data, as small a sample as it was, I think I have some ideas that will help influence Bellabeat marketing strategy.

Target Audience

Women who work full-time jobs (based on the sedentary and active times) and spend a lot of time at computers or focused on their work or in meetings (according to the sedentary time data).

These women do light activity to stay healthy, but they probably need to do more, and could probably use encouragement and motivation to do more.

Key Message for the Online Campaign

Bellabeat is more than just another smart device. With the features it has, including new features that track menstruation functions, Bellabeat can be a partner who helps women develop healthy habits and motivate them all through the day with recommendations and advice.

Ideas to Improve the Bellabeat App

- Average total steps per day is 7638, which is below what the 8000 to 10,000 the CDC recommends. There should be a setting that encourages women to take 8 to 10 thousand steps. the encouragement could be an alarm, along with articles on the site.
- the Bellabeat app could have reminders that tell people either at noon or at five o'clock that it's time to go out and get exercise.
- It would also be good if the app could measure other forms of exercise, as not everyone just walks or runs, but could include other activities, or at least a way to time them as well as walking or running.
- If app users want improve their sleep, they need to reduce sedentary time and and could probably use a reminder not to spend too much time in bed waiting for sleep, and one to tell them a better time for sleeping. This could be followed up with articles on the site that offer pre-bedtime meditation and articles that tell people how to fall asleep more efficiently.
- Using information about menstrual cycles, the app could help users plan their cycles better (although this of course is not an exact science) and deal with all the aspects of this that, as a man, I don't really know enough to tell you about.

[id]<https://www.kaggle.com/datasets/arashnic/fitbit>