Bellabeat

Wellness Technology and Consumers Uses: Bellabeat Case Study

By Mishel Kadosh

Table of Contents

Introduction

Junior data analyst project in the marketing analyst team at 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, co-founder 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 insight into how consumers are using their smart devices. The insights will then help guide the marketing strategy for the company. I will present my analysis to the Bellabeat executive team along with my high-level recommendations for Bellabeat's marketing strategy.

About the company

Urška Sršen and Sando Mur founded Bellabeat, a high-tech company that manufactures health-focused smart products. Sršen used her background as an artist to develop beautifully designed technology that informs and inspires women around the world. Collecting data on activity, sleep, stress, and reproductive health has allowed Bellabeat to empower women with knowledge about their 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.

Products

Bellabeat app: The Bellabeat app provides users with health data related to their activity, sleep, stress, menstrual cycle, and mindfulness habits. This data can help users better understand their current habits and make healthy decisions. The Bellabeat app connects to their line of smart wellness products.

Leaf: Bellabeat's classic wellness tracker can be worn as a bracelet, necklace, or clip. The Leaf tracker connects to the Bellabeat app to track activity, sleep, and stress.

Time: This wellness watch combines the timeless look of a classic timepiece with smart technology to track user activity, sleep, and stress. The Time watch connects to the Bellabeat app to provide you with insights into your daily wellness.

Spring: This is a water bottle that tracks daily water intake using smart technology to ensure that you are appropriately hydrated throughout the day. The Spring bottle connects to the Bellabeat app to track your hydration levels.

Bellabeat membership: Bellabeat also offers a subscription-based membership program for users. Membership gives users 24/7 access to fully personalized guidance on nutrition, activity, sleep, health and beauty, and mindfulness-based lifestyle and goals.

1. Background

Sršen asks to analyze smart device usage data to gain insight into how consumers use non-Bellabeat smart devices. She then wants to select one Bellabeat product to apply these insights to in the presentation. These questions will guide 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?

Stakeholders

Urška Sršen: Bellabeat's co-founder and Chief Creative Officer

Sando Mur: Mathematician and Bellabeat's cofounder; a key member of the Bellabeat executive team

Bellabeat marketing analytics team: A team of data analysts responsible for collecting, analyzing, and reporting data that helps guide Bellabeat's marketing strategy.

The business goal

Analyze Fitbit data, gain insight to identify trends in how consumers use non-Bellabeat smart devices to apply insights into Bellabeat's marketing strategy.

2. The data

Urška Sršen encourages me to use public data that explores smart device users' daily habits. She points me to a specific data set:

About the dataset

The FitBit Fitness Tracker Data dataset contains 18 CSV files with 30 users of FitBit Fitness Tracker Data from Mobius in Kaggle: https://www.kaggle.com/arashnic/fitbit. These datasets were generated by respondents to a distributed survey via Amazon Mechanical Turk between 12/03/2016 – 12/05/2016. The Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring. It includes information about daily activity, steps, and heart rate that can be used to explore users' habits.

Data structure

Each file from the 18 CSV files represents different quantitative data tracked by Fitbit. The data is considered long format since each row is a one-time point per subject, so each subject will have data in multiple rows. Every user has a unique ID in different rows since data is tracked by day and time.

File	File name	File type	File description
1	dailyActivity_merged	Microsoft Excel (CSV)	Daily Activity over 31 days of 33 users. Tracking daily: Steps, Distance, Intensities, Calories
2	dailyCalories_merged	Microsoft Excel (CSV)	Daily Calories over 31 days of 33 users
3	dailyIntensities_merged	Microsoft Excel (CSV)	Daily Intensity over 31 days of 33 users. Measured in Minutes and Distance, dividing groups in 4 categories: Sedentary, Lightly Active, Fairly Active,Very Active
4	dailySteps_merged	Microsoft Excel (CSV)	Daily Steps over 31 days of 33 users
5	heartrate_seconds_merged	Microsoft Excel (CSV)	Exact day and time heartrate logs for just 7 users
6	hourlyCalories_merged	Microsoft Excel (CSV)	Hourly Calories burned over 31 days of 33 users
7	hourlyIntensities_merged	Microsoft Excel (CSV)	Hourly total and average intensity over 31 days of 33 users
8	hourlySteps_merged	Microsoft Excel (CSV)	Hourly Steps over 31 days of 33 users
9	minuteCaloriesNarrow_merged	Microsoft Excel (CSV)	Calories burned every minute over 31 days of 33 users (Every minute in single row)
10	minuteCaloriesWide_merged	Microsoft Excel (CSV)	Calories burned every minute over 31 days of 33 users (Every minute in single column)
11	minuteIntensitiesNarrow_merged	Microsoft Excel (CSV)	Intensity counted by minute over 31 days of 33 users (Every minute in single row)
12	minuteIntensitiesWide_merged	Microsoft Excel (CSV)	Intensity counted by minute over 31 days of 33 users (Every minute in single column)
13	minuteMETsNarrow_merged	Microsoft Excel (CSV)	Ratio of the energy you are using in a physical activity compared to the energy you would use at rest. Counted in minutes
14	minuteSleep_merged	Microsoft Excel (CSV)	Log Sleep by Minute for 24 users over 31 days. Value column not specified
15	minuteStepsNarrow_merged	Microsoft Excel (CSV)	Steps tracked every minute over 31 days of 33 users (Every minute in single row)
16	minuteStepsWide_merged	Microsoft Excel (CSV)	Steps tracked every minute over 31 days of 33 users (Every minute in single column)
17	sleepDay_merged	Microsoft Excel (CSV)	Daily sleep logs, tracked by: Total count of sleeps a day, Total minutes, Total Time in Bed
18	weightLogInfo_merged	Microsoft Excel (CSV)	Weight track by day in Kg and Pounds over 30 days. Calculation of BMI.5 users report weight manually 3 users not.In total there are 8 users

Accessibility and privacy

The metadata of the dataset confirms and verifies that is open-source data (CC0: Public Domain). The data creator has dedicated the work to the public domain by waiving all of his or her rights to the work worldwide under copyright law, including all related and neighboring rights, to the extent allowed by law. It can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission.

Credibility and Integrity

The first limitation is the small sample size (30 users) and the lack of demographic information. Therefore, we could encounter a sampling bias and we are not sure if the sample is representative of the population as a whole. Another problem we encounter is that the dataset is not current, created in 2016 and also the short time of data collection is a limitation, only 2 months long. For these reasons, we will give our case study a project approach.

3. Process and adjust

In the analysis process, I will focus on R programming due to the accessibility, amount of data, and to be able to create data visualization to share my results with stakeholders.

Packages and libraries

I will choose the packages that will help in the analysis and install them.

List of packages:

```r
install.packages("ggpubr")
install.packages("tidyverse")
install.packages("here")
install.packages("skimr")
install.packages("janitor")
install.packages("lubridate")
install.packages("ggrepel")
library(ggpubr)
library(tidyverse)
library(here)
library(skimr)
library(janitor)
library(lubridate)
library(ggrepel)
```

Import and preview datasets

Now, I will upload the datasets that will help me with the business goal. I will focus on specific three dataset files because they contain the most relevant and richest data:

The files are: Daily activity, hourly steps, and sleep day.

```r
daily_activity <- read_csv(file= "/cloud/project/dailyActivity_merged.csv")
hourly_steps <- read_csv(file= "/cloud/project/hourlySteps_merged.csv")
sleep_day <- read_csv(file= "/cloud/project/sleepDay_merged.csv")
```

Due to the small sample data, I won't consider for this analysis the weight file (only 8 Users) and heart rate file (only 7 users).

First, I preview the data frames and view the summary of each column.

```
head(daily_activity)
str(daily_activity)

head(hourly_steps)
str(hourly_steps)

head(sleep_day)
str(sleep_day)
```

Cleaning and formatting

Now that we got to know more about our data structures we will process them to look for any errors and inconsistencies. Before I continue the cleaning, I want to ensure how many unique users are per data frame.

```
> n_unique(daily_activity$Id)
[1] 33
> n_unique(hourly_steps$Id)
[1] 33
> n_unique(sleep_day$Id)
[1] 24
```

Remove duplicates and N/A

I will now look for any duplicates:

```
> sum(duplicated(daily_activity))
[1] 0
> sum(duplicated(hourly_steps))
[1] 0
> sum(duplicated(sleep_day))
[1] 3
```

Then, I will remove the duplicates, and N/A cells and check again the duplicate at sleep_day:

```
> daily_activity <- daily_activity %>%
+   distinct() %>%
+   drop_na()
>
> hourly_steps <- hourly_steps %>%
+   distinct() %>%
+   drop_na()
>
> sleep_day <- sleep_day %>%
+   distinct() %>%
+   drop_na()
> sum(duplicated(sleep_day))
[1] 0
```

Clean and consistent columns

We want to ensure that all column names are using the right syntax and the same format in all datasets since we will merge them later on. We are changing the format of all columns to lower case.

```
clean_names(daily_activity)
daily_activity<- rename_with(daily_activity, tolower)

clean_names(hourly_steps)
hourly_steps <- rename_with(hourly_steps, tolower)

clean_names(sleep_day)
sleep_day <- rename_with(sleep_day, tolower)
```

Now that we have verified that column names and changed to lower case, we will focus on cleaning the date-time format for daily_activity and sleep_day since we will merge both data frames. Since we can disregard the time on the daily_sleep data frame we are using as_date instead of as_datetime.

```
daily_activity <- daily_activity %>%
  rename(date = activitydate) %>%
  mutate(date = as_date(date, format = "%m/%d/%Y"))

sleep_day <- sleep_day %>%
  rename(date = sleepday) %>%
  mutate(date = as_date(date,format ="%m/%d/%Y %I:%M:%S %p" , tz=Sys.timezone()))

head(daily_activity)
head(sleep_day)
```

For our hourly_steps dataset we will convert the date string to date-time:

```
hourly_steps<- hourly_steps %>%
  rename(date_time = activityhour) %>%
  mutate(date_time = as.POSIXct(date_time,format ="%m/%d/%Y %I:%M:%S %p" , tz=Sys

head(hourly_steps)
```

Merging Datasets

I will merge daily_activity and sleep_day to see any correlation between variables by using id and date as their primary keys.

```
activity_and_sleep <- merge(daily_activity, sleep_day, by=c ("id", "date"))
glimpse(activity_and_sleep)
```

4. Analyze and share

I will analyze the trends of the users of FitBit and determine if that can help us with BellaBeat's marketing strategy. First, I will present the Descriptive statistic, then the predictive statistic. At this point, we will use the two main files we worked on before:

1. activity_and_sleep (18 columns, 410 rows), the merged file, contains all records of users' activities.

2. hourly_steps (3 columns, 22099 rows), contains the day and hour of users' activities.

Descriptive analysis

Descriptive analytics looks at data statistically to tell you what happened in the past. Descriptive analytics helps a business understand how it is performing by providing context to help stakeholders interpret information.

Type of users

since we don't have any demographic variables from our sample we want to determine the type of users with the data we have. We can classify the users by activity considering the daily number of steps. I will categorize users into five groups as follows:

1. Very low active - Less than 5000 steps a day.
2. Low active - Between 5000 and 7499 steps a day.
3. Active - Between 7500 and 9999 steps a day.
4. High active - More than 10000 steps a day.
5. Very high active - More than 12500 steps a day.

Classification has been made per the following article at:
https://www.medicinenet.com/how_many_steps_a_day_is_considered_active/article.htm

Table 1

As we can see, the largest group of users is Active (38%), while the smallest group of users is very high active (4%).

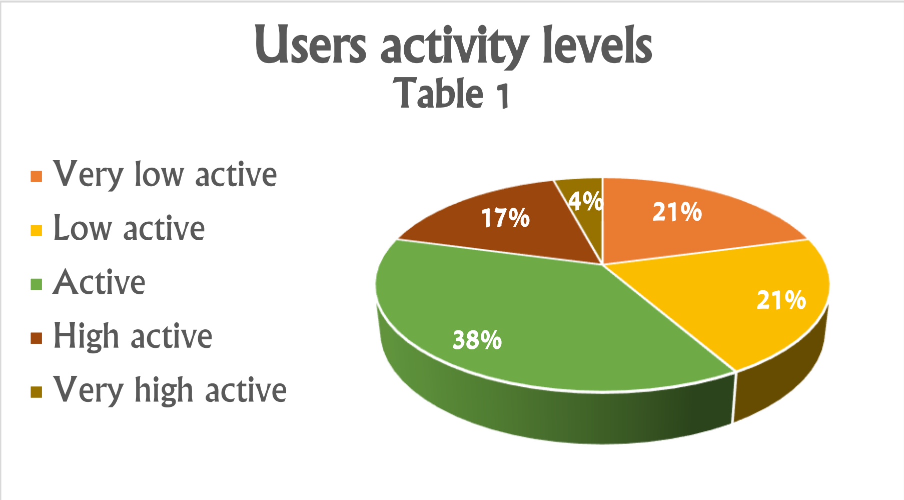


Table 1	
Activity level	**Unique users**
Very low active	5
Low active	5
Active	9
High active	4
Very high active	1
Total	**24**

Table 2

Here we can see some basic calculations about our user's steps that will help us to know better the data.

Table 2	
Function	Value
Average steps	8,514.9
Max steps	22,770
Min steps	17
Median steps	8,913

Table 3

Here we can see more basic calculations but now, about the calories burned by our users. It will help us to know better the user's behavior.

Table 3	
Function	Value
Average calories	2,389.3
Max calories	4,900
Min calories	257
Median calories	2,207

Table 4

Here we can also see basic calculations but now, about the sleep time of our users. It will help us to know better the user's sleep quality.

Table 4	
Function	Value (Hours)
Average sleep time	7
Max sleep time	13
Min sleep time	1
Median sleep time	7

Table 5

Here we can see the total tracked steps of all users, grouped by days. It will help us to better understand week steps activity patterns.

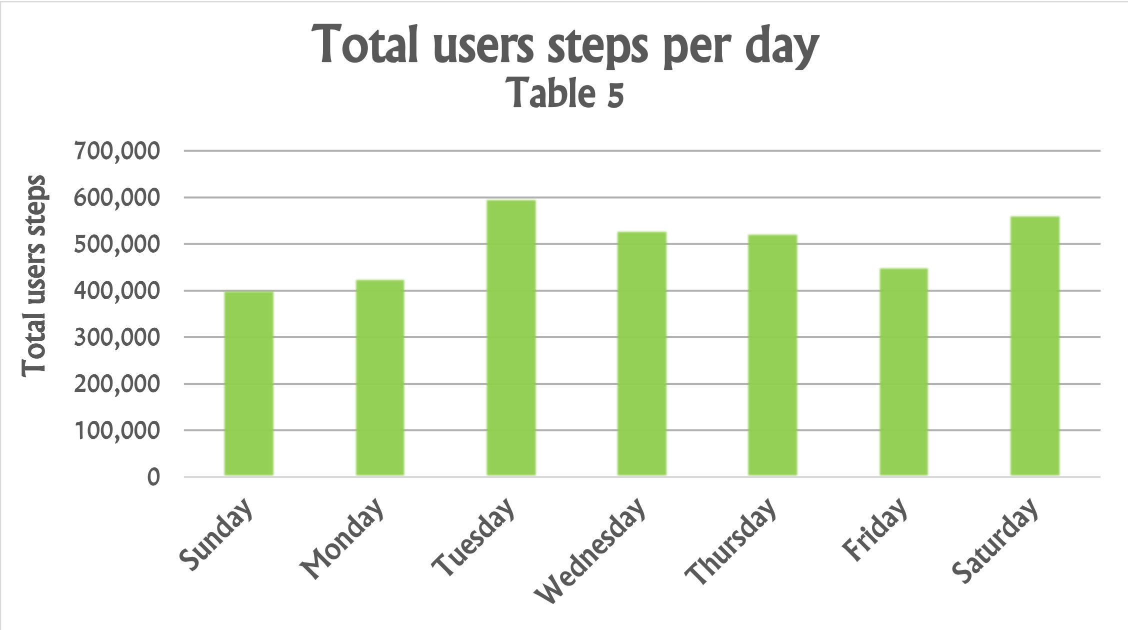


Table 6

We also want to examine the smart devices used by days, to see if there are any differences in groups of users.

Low use - users who use their device between 1 and 10 days.

Mid use - users who use their device between 10 and 20 days.

High use - users who use their device between 21 and 31 days.

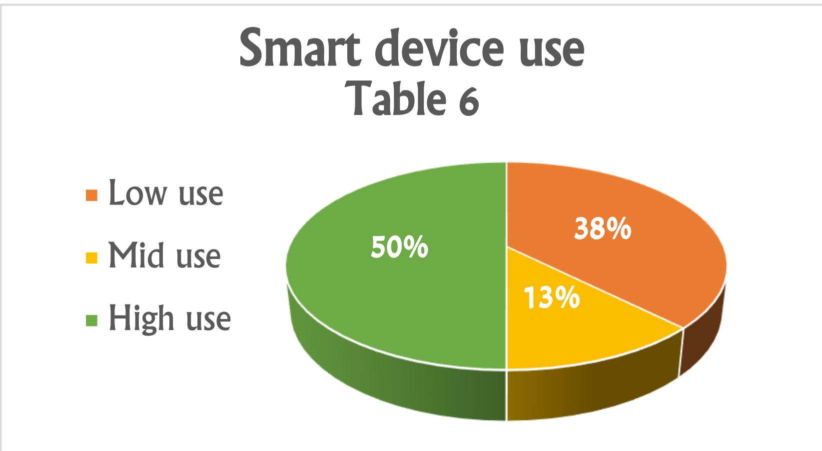


Table 6

Use level	Unique users
Low use	9
Mid use	3
High use	12
Total	**24**

Predictive analysis

Predictive analytics takes historical data and feeds it into a machine learning model that considers key trends and patterns. The model is then applied to current data to predict what will happen next.

Correlations

In statistics, correlation or dependence is any statistical relationship, whether causal or not, between two random variables or bivariate data.

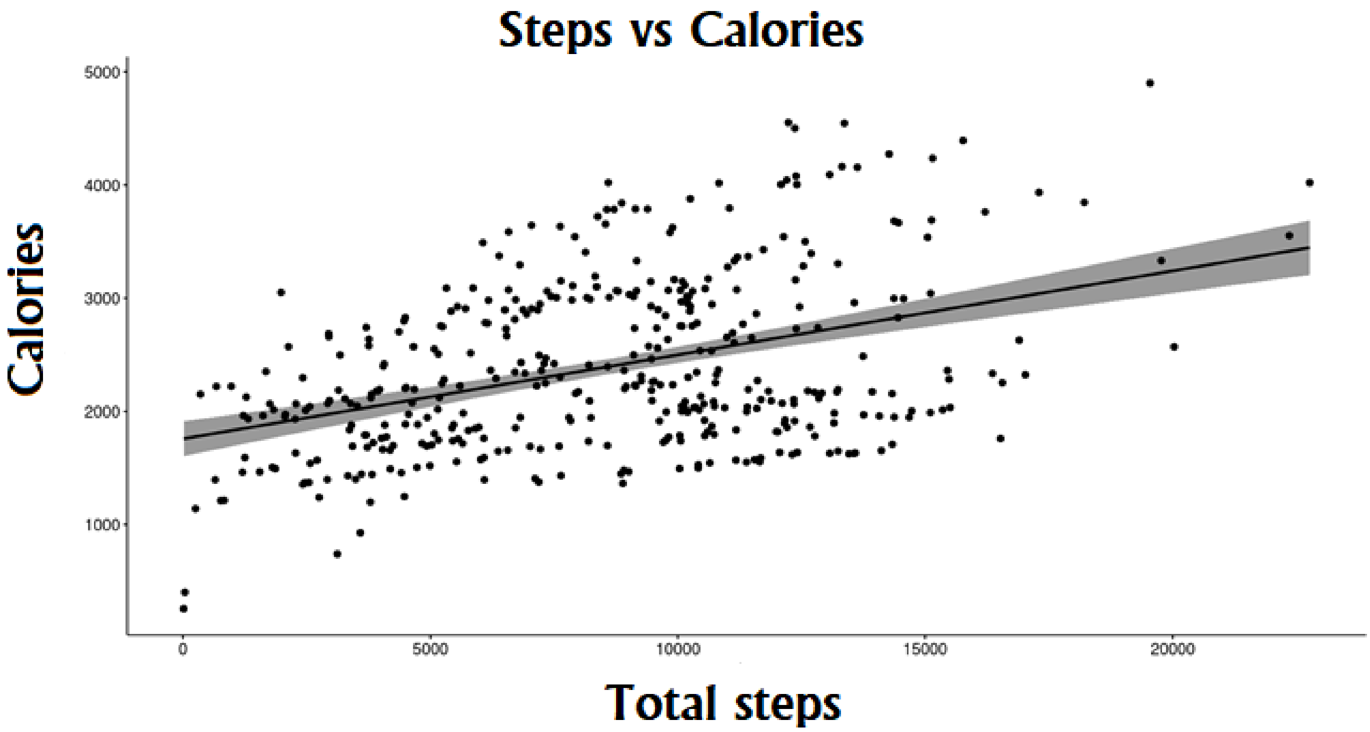
Steps and calories

To examine if there is a relationship between steps and calories, a Pearson test was conducted. As I expect, there is a medium positive correlation between the total steps and the calories, the more steps you take – the more calories you will burn ($r=0.368$, $P<0.001$).

Correlations

		totalsteps	calories
totalsteps	Pearson Correlation	1	.368**
	Sig. (2-tailed)		<.001
	N	414	413
calories	Pearson Correlation	.368**	1
	Sig. (2-tailed)	<.001	
	N	413	414

**. Correlation is significant at the 0.01 level (2-tailed).



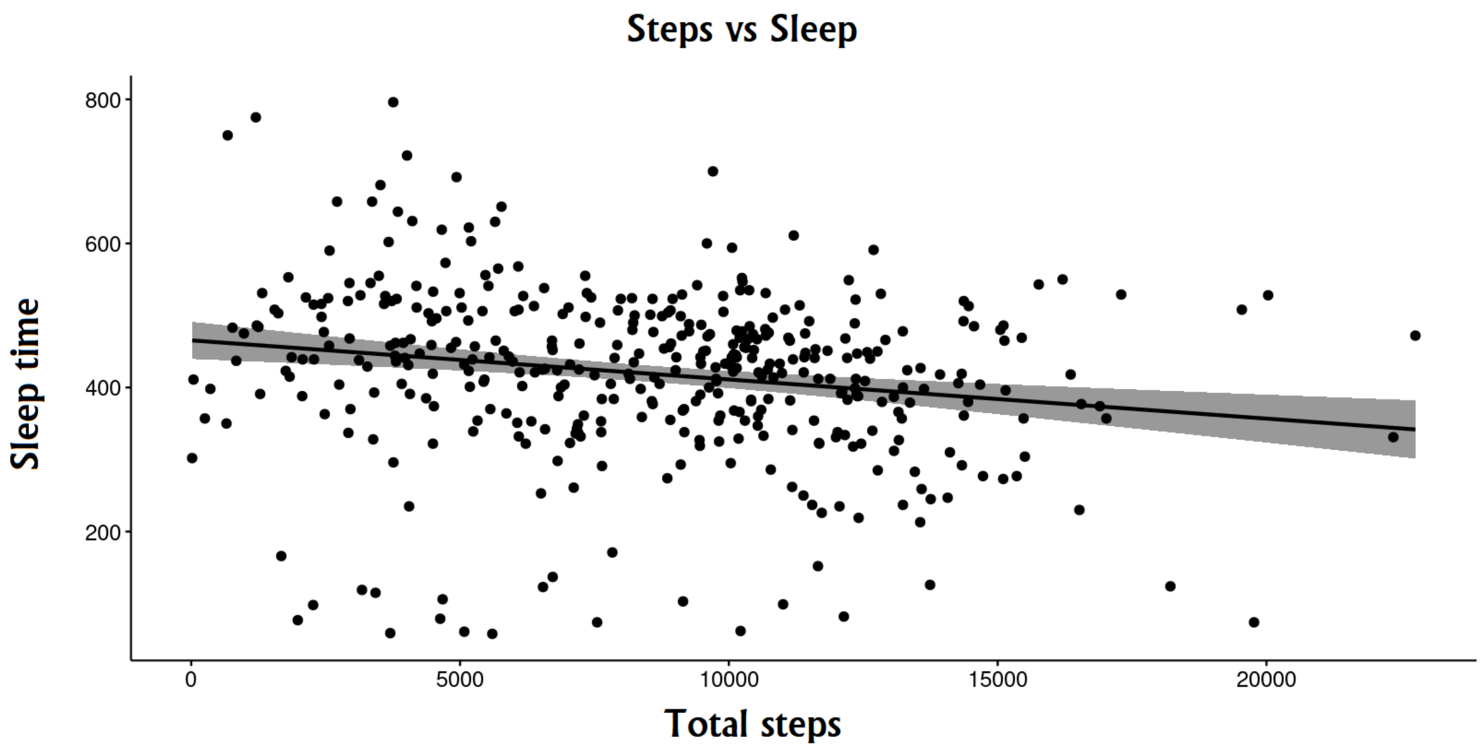
Steps and sleep

To examine if there is a relationship between steps and sleep time, a Pearson test was conducted. Here, there is a weak negative correlation between the total steps and the sleep time, the more steps you take – the less time you will sleep ($r=-0.190$, $P<0.001$).

Correlations

		totalsteps	totalminutesasleep
totalsteps	Pearson Correlation	1	-.190**
	Sig. (2-tailed)		<.001
	N	414	410
totalminutesasleep	Pearson Correlation	-.190**	1
	Sig. (2-tailed)	<.001	
	N	410	410

**. Correlation is significant at the 0.01 level (2-tailed).



5. Conclusions

To answer the business goal and help Bellabeat on the mission, based on our results, I would advise using own tracking data for further analysis. Knowing that our main target is young and adult women I would encourage to continue finding trends to be able to create a marketing strategy focused on them.

Results and applications

After the analysis I have found some trends that may help the online campaign and improve Bellabeat app:

Reward system - I assume that some people don't get motivated by notifications only, so we could create a league with tables and points on our app for a limited period. This engagement would consist in reaching the top places in the league based on the number of steps walked every day. You need to maintain your activity level for some time (maybe a month) to stay in the league. For each league, you would win a certain amount of points that would be redeemable for merchandise or a discount on other Bellabeat products.

Sleep and notification - Based on the analysis I see that the average users sleep less than 8 hours a day (Table 4). We can make a function that the users could track their sleep activity, show them which days are "bad" sleeping days and send them notifications about sleep time. So, the users can improve their sleep. Also, create a page in the app with helpful resources that help customers sleep - ex. breathing advises podcasts with relaxing music, and sleep techniques.

Daily notification - As I classified users into 5 categories I saw that the average of users walks around 8,500 steps (Table 2) daily and most on Tuesday (Table 5). The company needs to encourage customers to reach at least daily recommended steps by CDC - 8,000, send them notifications if they haven't reached the steps yet, and create also posts on the new page in the app, explaining the benefits of reaching that goal. As CDC describes, the more steps you walk the lower the mortality rate.

Limitations and recommendations

As with all reports, there are limitations also in this report. First, the datasets used have a small sample and can be biased since we didn't have any demographic details of users. As we mentioned, we would recommend the company try to collect demographic data from the users to perform more effective analyzes in the future. Second, this case study was conducted as part of a data analyst course. Therefore, the time and resources available to me were relatively limited. Third, the datasets used were created in 2016 so it must be known that the use patterns use of this type of sports equipment in particular and smart devices, in general, have probably changed over the past few years. Thus, I will recommend for future projects and reports to use the most updated data, try to reach more users and expand the sample size, and of course, start tracking demographic data from the users.