

ANALYSIS REPORT

BELLABEAT



Introduction

[Bellabeat](#), a high-tech manufacturer of health-focused products for women, targets a global health-conscious community. It offers a subscription-based membership program for users giving them access to personalized guidance on having a healthy lifestyle. Inspiring women to be the best version of themselves by encouraging them to be healthier, happier, and more mindful.

Co-founder of Bellabeat, Mrs. Urška Sršen states that Bellabeat empowers women around the world by helping them reach their fullest potential through the use of revolutionary technology that easily tracks their overall health, and aids in achieving a mind and body connection throughout different stages of life.

Objective

Bellabeat, with the potential to become a large player in the global smart device market with the help of analyzing smart device fitness data, may achieve so. In this case study, we will focus on one of Bellabeat's products and analyze smart device data to gain insight into how customers are using their smart devices.

Stakeholders:

- Urška Sršen: Bellabeat's co-founder and Chief Creative Officer
 - Sando Mur: Mathematician and Bellabeat's co-founder; a key member of the Bellabeat executive team
-

A good way to start this case study would be by asking the right questions, such as,

- What is the business task?

To analyze data from smart devices and gain insight into how is a smart device being used and how can it help Bellabeat to grow.

- What are the future possibilities of growth?

Due to Covid-19, everybody is starting to get more and more health-conscious. Hence, the market will only grow in the future.

We should start by researching products with similar functionality and explore user experience while using the product to identify usage trends.

Preparing the data

- We will use [Fitbit Fitness Tracker Data](#) for this analysis, this dataset is made available by Mobius.
- The dataset is CC0, i.e, no copyright was reserved by the provider
- Thirty Fitbit users consented to submit their personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring.

Does the data ROCCC?

Reliability: Low - As the dataset was collected from thirty different users(gender unknown).

Originality: Low - As it is third-party data the originality cannot be confirmed.

Comprehensive: Normal - the dataset contains data on daily activity, calories intake, steps, and Intensities.

Current: Normal - the data is 5 years old but is still relevant.

Cited: High - All the sources and the data collectors were well documented.

We have a total of 18 datasets but we will only focus on the merged datasets of activities done in a day, i.e,

dailyActivity_merged.csv

dailyCalaroies_merged.csv

dailyIntensities_merged.csv

dailySteps_merged.csv

sleepDay_merged.csv

weightlogInfo_merged.csv

Processing the data

In this case study, we will use R Programming language to conduct the analysis and present meaningful visualizations. We will first look at the integrity of data and then do the necessary processing steps.

We will use *dailyActivity_merged.csv*, *sleepDay_merged.csv*, and *dailySteps_merged.csv* datasets in this analysis. When we look into the data we see that there are many inconsistencies such as date format, irrelevant columns, missing data, and more.

When we explore the datasets, we come across the fact that 'Id' is common in all the datasets - this can be used to merge datasets if need be.

Also if we look into the *daily_activity* dataset you will see data of *daily_steps* is already present in *daily_activity\$TotalSteps*. This is the reason why other datasets were not loaded. Hence, we will not use *dailySteps_merged.csv* further in the analysis.

Analyzing the data

We used *n_distinct()* and *nrow()* to get the value of the number of unique users and the number of observations in a dataset respectively. As *Id* is the only column that can be used to identify different users, we will use it to calculate the number of unique users.

The analysis shows that in *daily_activity* there are 33 different users and 940 observations while in *sleep_day* there are 24 different users and 413 observations.

We used *summarize()* to look more into the statistics of the dataset.

The summary of *daily_activity* shows that the average number of steps taken in a day by a user is 7406, which is more than the minimum recommended steps per day but less than recommended steps per day for health by CDC. Looking the *SedentaryMinutes* we get to know that an average person is sitting idle and not doing any physical activity for about 16.52 hours(991.2 minutes), which is very high. On average CDC recommends a person to do vigorous activity for at least 75 minutes a week, from the dataset we can see that on average 21.16 minutes is the time spent doing a vigorous activity which amounts to 148.12 minutes. So we can say on average a Fitbit user is doing more activity than recommended and availing more health benefits. Furthermore, on average a Fitbit user is burning 2304 calories per day which depends is good or not on your goal. To lose weight on average 3500

calories need to be burned according to CDC. On average a person burns 1800 calories a day, in that sense a Fitbit user is doing better.

The summary of *sleep_day* shows that on average a Fitbit user is sleeping for 419.5 minutes(6.99 hours) which is just touches the recommended sleep of 7 hours in a day. If you look at the time spent on bed, the average time comes out to be 458.6 which means that a person is spending 39.1 minutes laying on the bed. According to Health Central, people should not spend more than 1 hour in bed awake.

We used Inner Join to combine *daily_activity* and *sleep_day* to gain more insight into the relationship between activity done during a day and sleep cycle. Looking at *combined_data* we get to know that on average more the active minutes spent by a user more is their minutes asleep. That is the more you burn calories better get your sleep cycle.

Visualization

We used *ggplot()* a function in the *ggplot2* library in R to visualize data and look into patterns and trends that the dataset contains.

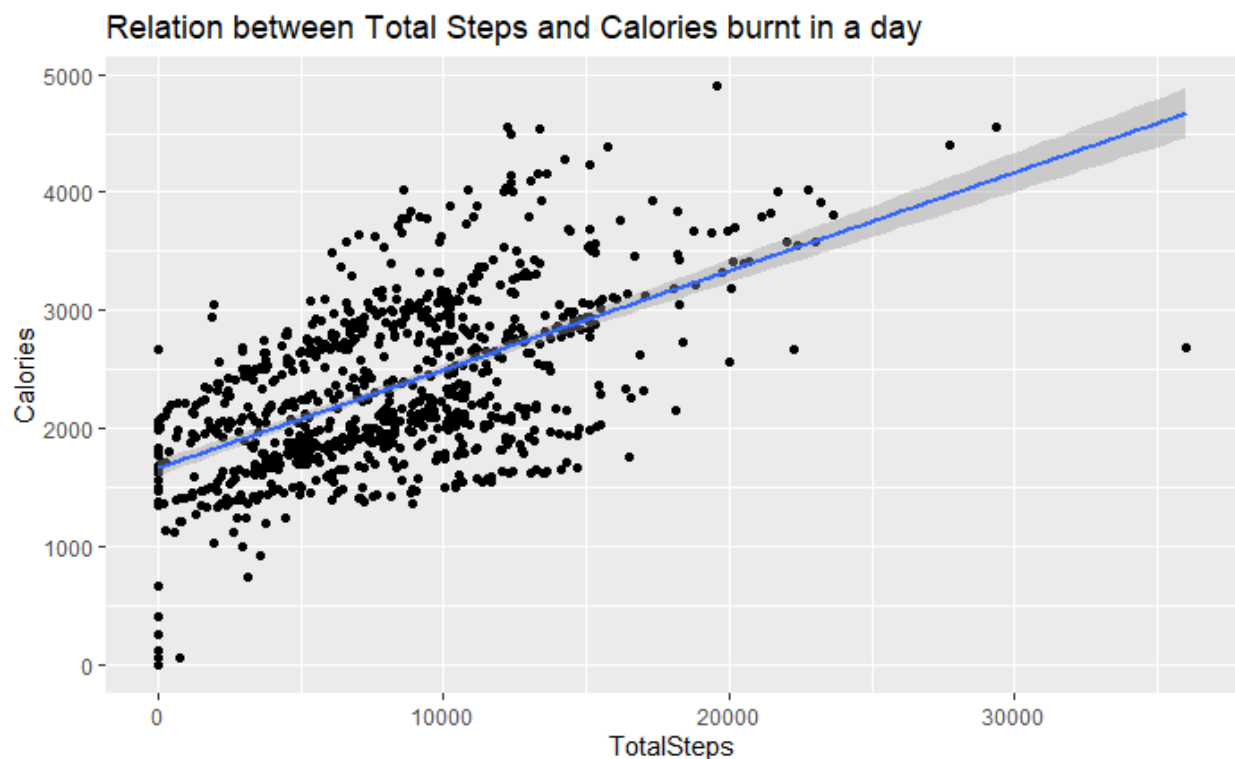


Fig 1

As we can see in Figure 1 there is a positive relation between Total steps and Calories burnt, which tells us that more steps taken by a fitbit user, more calories were burnt by them.

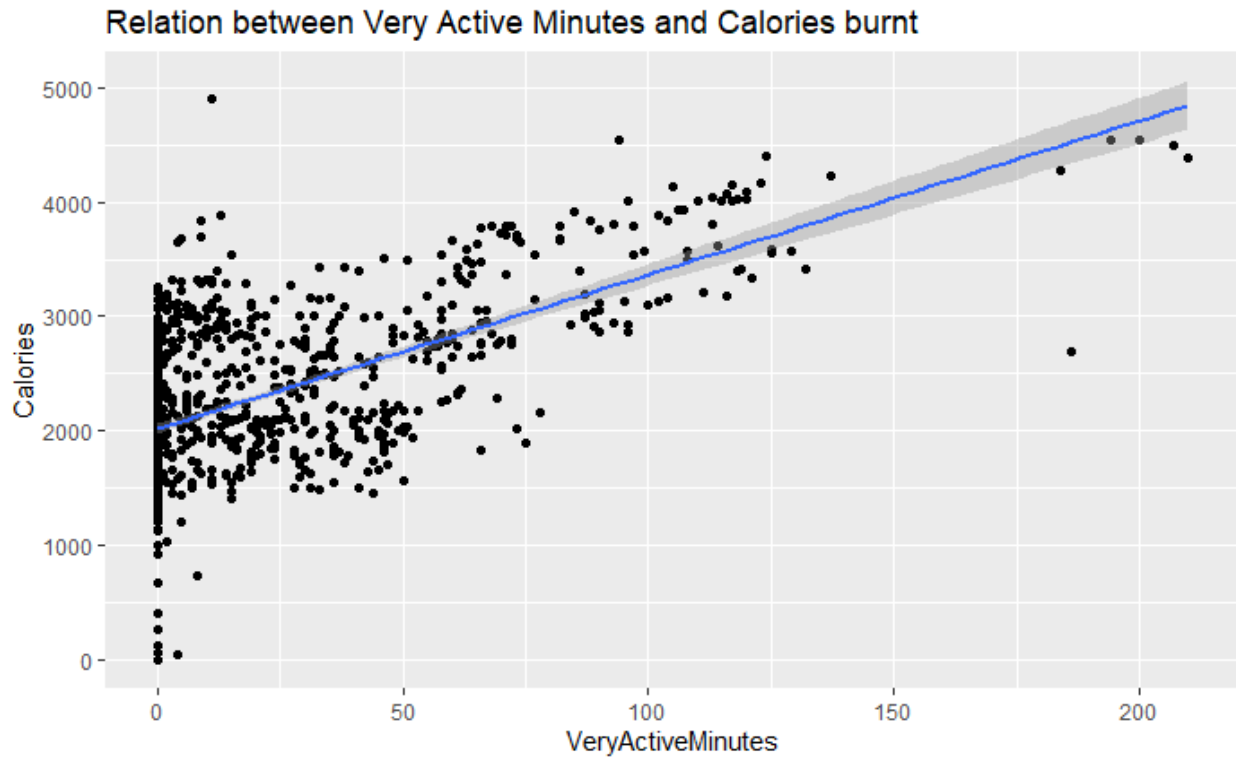


Fig 2

As seen in Figure 2, the relation between Active minutes and Calories burnt is positive meaning the more a person spends doing the activity more calories will be burnt by that person.

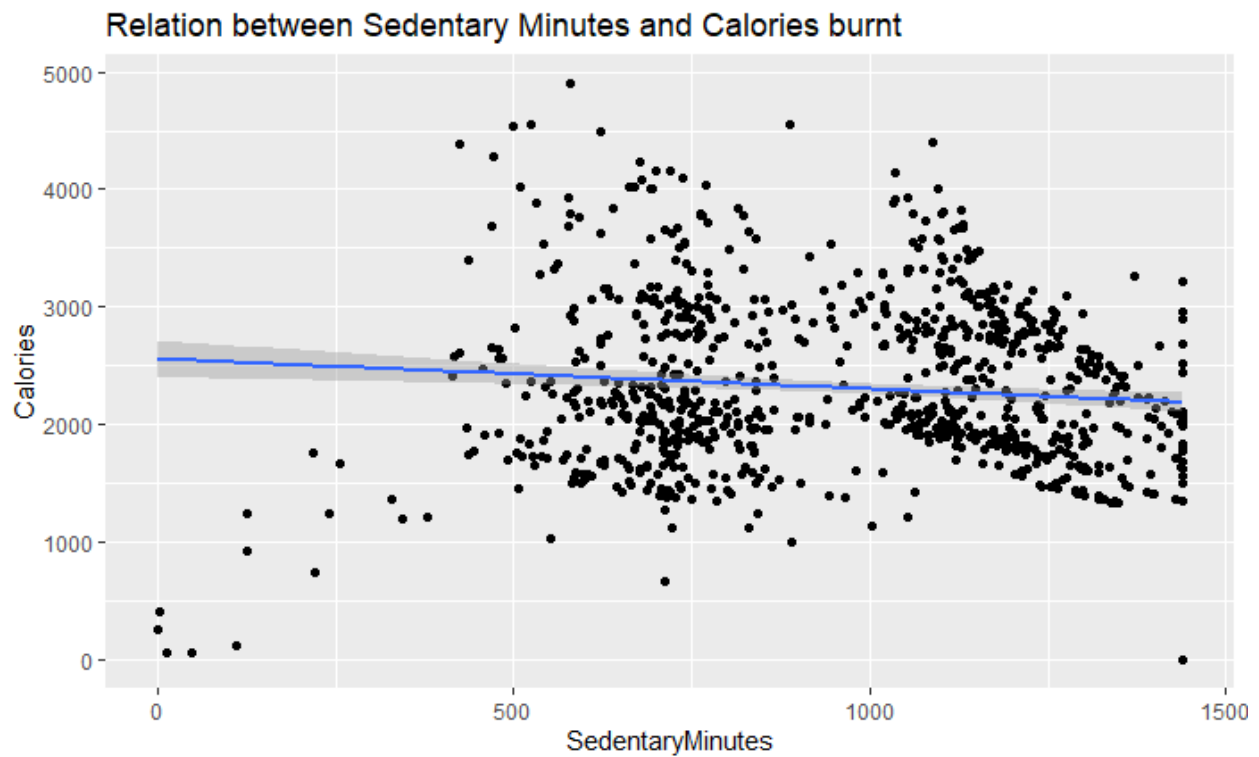


Fig 3

As seen in Figure 3, the relation between Sedentary minutes and Calories burnt comes to be almost linear inclined toward negative side. This means that if a user is idle the calories burnt will not be affected much but as the idle time increase the calories burnt will be less than that of a person who is not idle for that amount of time.

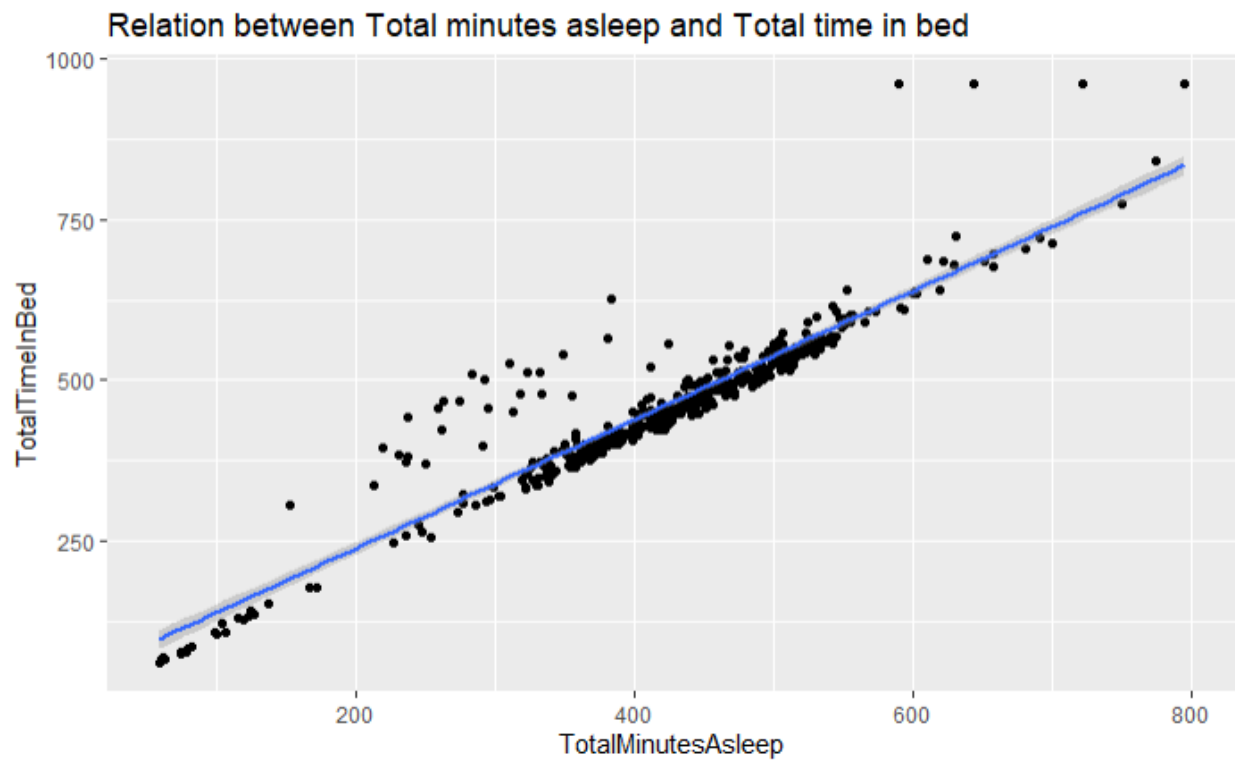


Fig 4

Figure 4 tells us the relation between Total time asleep and Total time in bed. As most of the points are near the line, it tells us that most of the time spent in the bed is similar to time asleep.

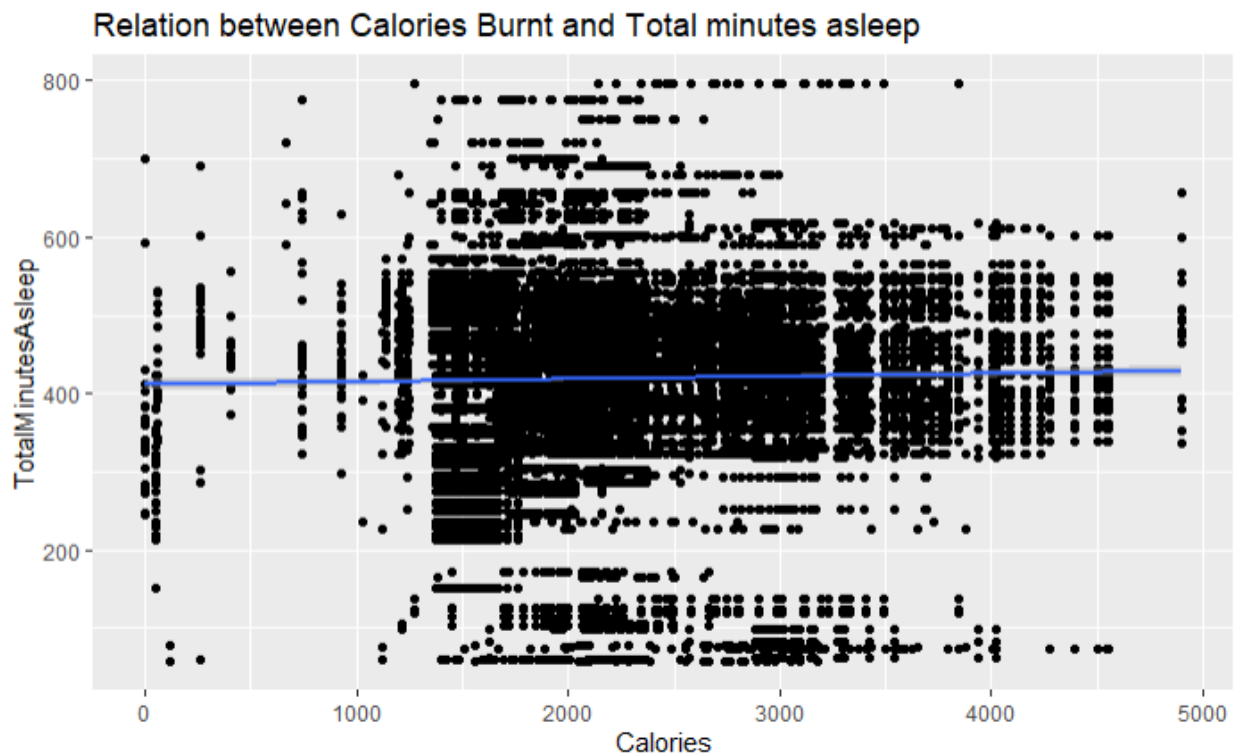


Fig 5

As we see in Figure 5, the relation between calories burnt and minutes asleep is almost linear slightly inclined towards positive relation. This means that the more calories burnt in a day make a person sleep for a longer duration, this makes a healthy cycle.

Recommendations and act

Bellabeat grew rapidly because of its focus to empowering women by collecting data on activity, sleep, stress, and reproductive health. They provided women knowledge about their own health and habit but the world is also advancing more rapidly. To catch up and to grow their company more, Bellabeat should change their model of only providing information to start recommending their user-health tips using the data they gather.

In this analysis, there is a clear trend that non-active members are living a negative lifestyle. It is also established in the analysis that:

- More time spent doing nothing had a negative relation with calories burnt
 - Whereas more active minutes had a positive relation with calories burnt
 - Active users have a positive relationship with sleep.
-

Recommendations

Applying all the insight gained from the analysis, we will now make some data-driven decisions

- Bellabeat should start focusing on collecting data and encourage their users to achieve fitness goals
 - Bellabeat should start notifying its user if sedentary minutes exceed a certain amount.
 - Bellabeat should set daily/weekly goals and encourage its user to complete them.
 - Bellabeat can further enhance their sleep tracking function to promote the sleep/non-active relation. Use this as an incentive to purchase Bellabeat products: create a better sleeping habit by being more active in everyday life.
-