

Case Study: BellaBeat

Kahlyll Wilson

February 23 , 2023

NOTE: code and files can be found in the full [Github Repository](#)

Overview

I completed the following case study as a part of the Google Data Analytics Professional certificate. It is based on the fictional company BellaBeat, a high-tech manufacturer of health-focused products for women.

In this scenario, I was tasked as a data analyst to help BellaBeat become a potential big player in the global smart device market. The co-founder and Chief Creative Officer (CEO), Urška Sršen, believed that analyzing smart device fitness data could help unlock new growth opportunities for the company. Along with Sršen, another key stakeholder was Sando Mur, a mathematician and Bellabeat's co-founder.

The company offers a variety of products such as:

- Bellabeat App
- Leaf: a classic wellness tracker
- Time: a wellness watch combines timeless look of a classic timepiece with smart technology to track user activity
- Spring: This is a water bottle that tracks daily water intake using smart technology
- Bellabeat membership: a subscription-based membership program for users.

Buisness Task

In this scenario, Sršen and Mur have tasked me with completing the following requirements:

1. Analyze smart device usage data in order to gain insight into how consumers use non-Bellabeat smart devices
2. Select one Bellabeat product to apply these insights

Through completing the tasks this report will explain to key stakeholders:

1. What are some trends in smart device usage?
2. How could the trends apply to the Bellabeat customer?
3. How can both my team members and stakeholders use these insights to make data driven decisions?

Analysis

For my analysis, I focused on the fitness tracker to complete this report. I completed the analysis using the [FitBit Fitness Tracker Data \(FFTD\)](#). The FFTD dataset is open to the public and can be found on the [Kaggle website](#).

Data gathering and cleaning phase

Before analyzing the data, I first cleaned and organized the data using excel when I pulled the raw data file from FFTD. I noticed there were multiple spread sheets that had overlapping data. To make sure I was not looking at repeated data, I combined all of them into one file using the XLOOKUP function. [With the one excel sheet I imported it to a relational databases](#).

During this stage of the analysis, I noticed something interesting about the data. I observed that users rarely or never logged their data with the fitness tracker, and there were inconsistent metrics recorded when users were very active or moderately active. However, there was ample data recorded for the light activity metric. At this stage, it became clear that users were engaging in some form of activity every day.

Analysis phase

How analysis was conducted

For the analysis portion of this case study, I completed it using [SQL and the BigQuery database](#).

I broke the large data down to answer the query's below:

- What was the Average activity per day?
- What was each users Average Activity?
- What was the Average of all the users?

I stored these results into a [seperate folder](#), then used a combination of R and Tableau to visualize the results.

Results

The following metrics were used to complete analysis:

- Total Calories Burned
- Total Steps taken
- Total Distance Traveled
- Total Distance Tracked

Through analysis of these metrics, I found something interesting in the data. I noticed that in each of the datasets there was an outlier throwing off the trendline of my scatterplot. The error came from a single point on May 12th. For each metric, there was a significant drop in the data recorded on that day. There are no significant reasons or holidays on this day that would make the drop in activity seem reasonable across all users. Unless a vast majority of people celebrate National Limerick Day & World Migratory Bird Day, and I am just unaware. I decided it was safe to remove this data point.

From the data I found that the average user:

- Burned 2,304 Calories
- Took 7,638 Steps
- Traveled 5 miles
- Tracked 5 miles

It is a good sign that the average total distance traveled and the total distance tracked are the same because that means that the fitness tracker works properly and tracking all the miles the users are active.

After reviewing the averages for each metric overall, I analyzed the results by day. The first metric I analyzed was calories per day.

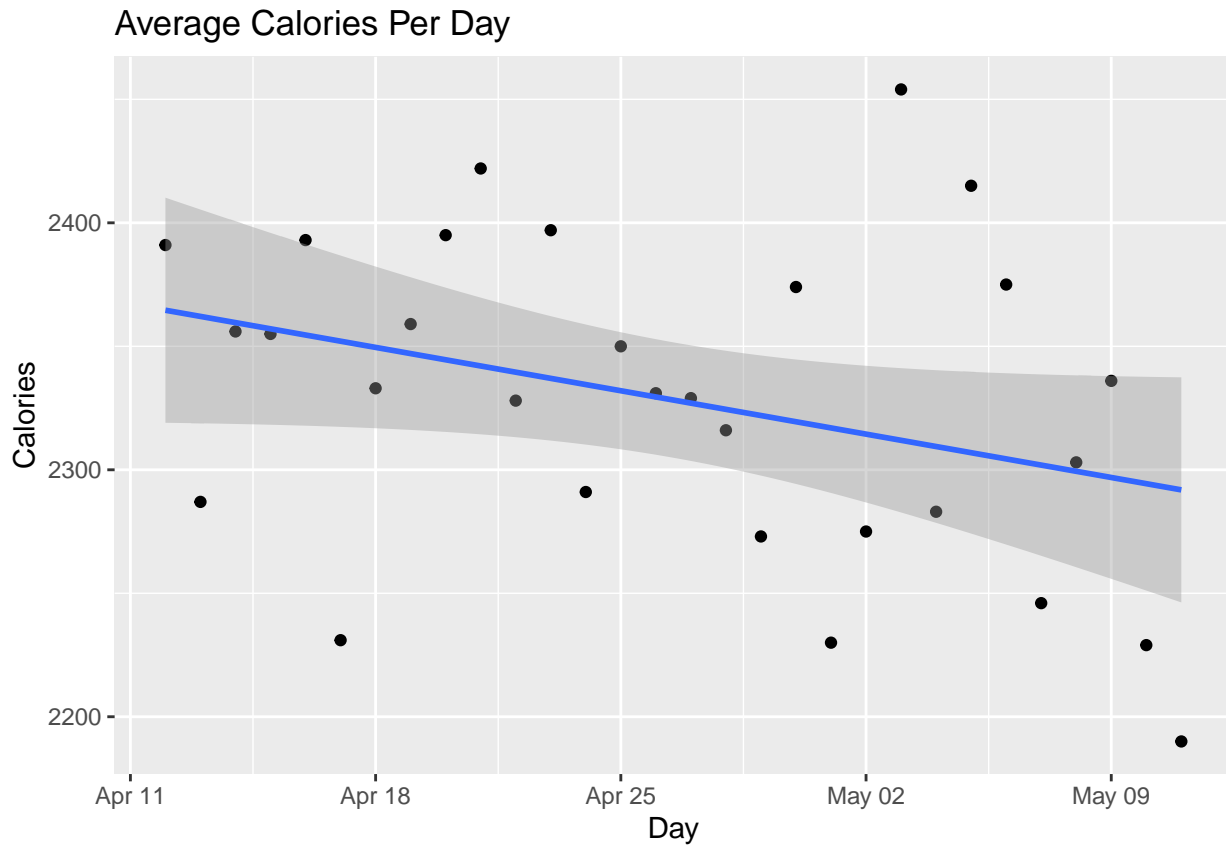
After isolating just the calories in SQL, I used the code below to plot a scatterplot and included a trend line to see what direction the data was moving. The plot shows a downward trend revealing that as time goes on, users burn fewer calories.

```
# loads calories file
library(readxl)
calories <- read_excel("/Users/kahlyllwilson/Desktop/excel sheets/Average Calories per day.xls")
calo = data.frame(calories)

#Filters out the outlier
filt_calo = filter(calo, Calories_Per_day > 1900)

# Makes scatter plot for the Average Calories per day
ggplot(data= filt_calo, mapping = aes(x=ActivityDate, y=Calories_Per_day)) +geom_point() +
geom_smooth(method = lm) + ggtitle('Average Calories Per Day') +
labs(x = 'Day', y='Calories')
```

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



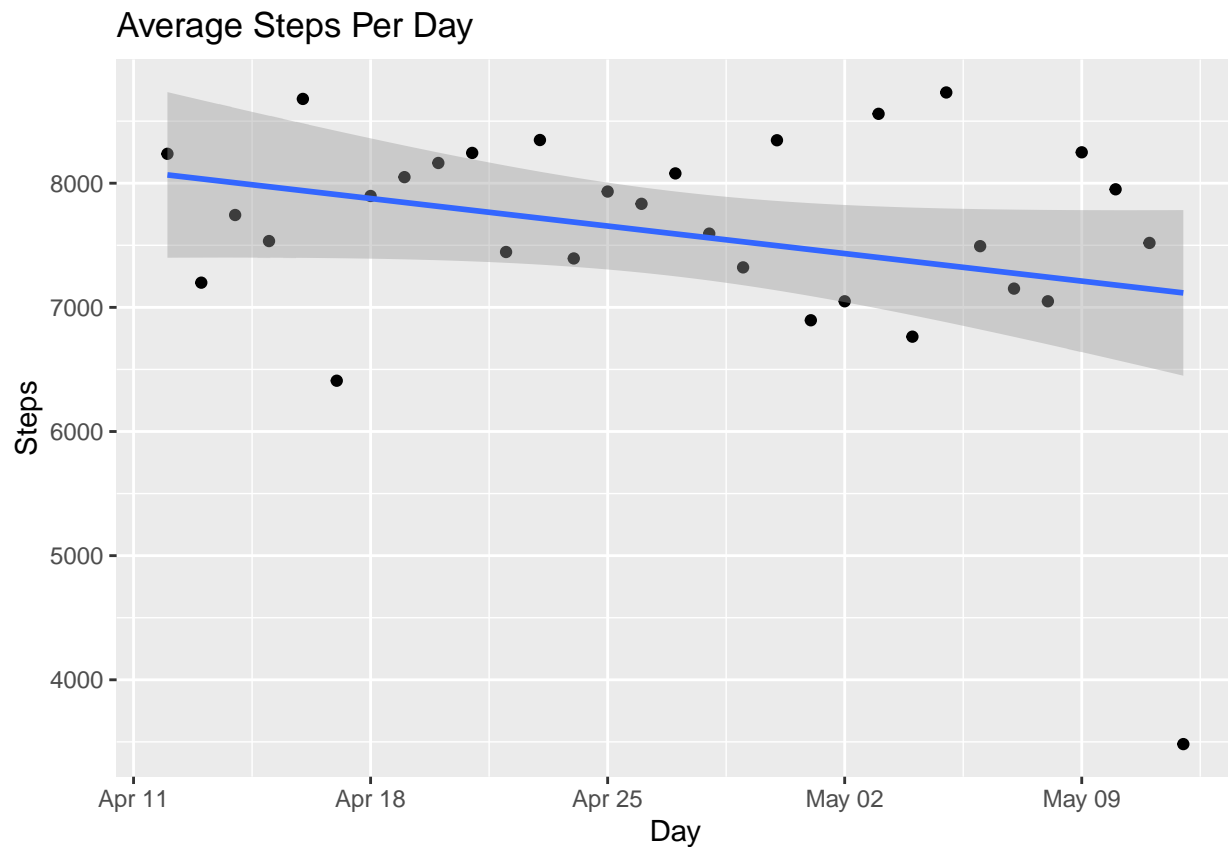
When viewing the results for the steps, I found that as time went on, users took fewer steps per day.

```
# loads steps file
library(readxl)
steps <- read_excel("/Users/kahlyllwilson/Desktop/excel sheets/Average steps per day.xls")
ste = data.frame(steps)

#Filters out the outlier
filt_ste = filter(ste, Average_Total_steps > 2.5)

# Makes scatter plot for the Average Tracked Distance per day
ggplot(data=filt_ste, mapping = aes(x=ActivityDate, y=Average_Total_steps)) +geom_point() +
  geom_smooth(method = lm) + ggtitle('Average Steps Per Day') +
  labs(x = 'Day', y ='Steps')
```

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



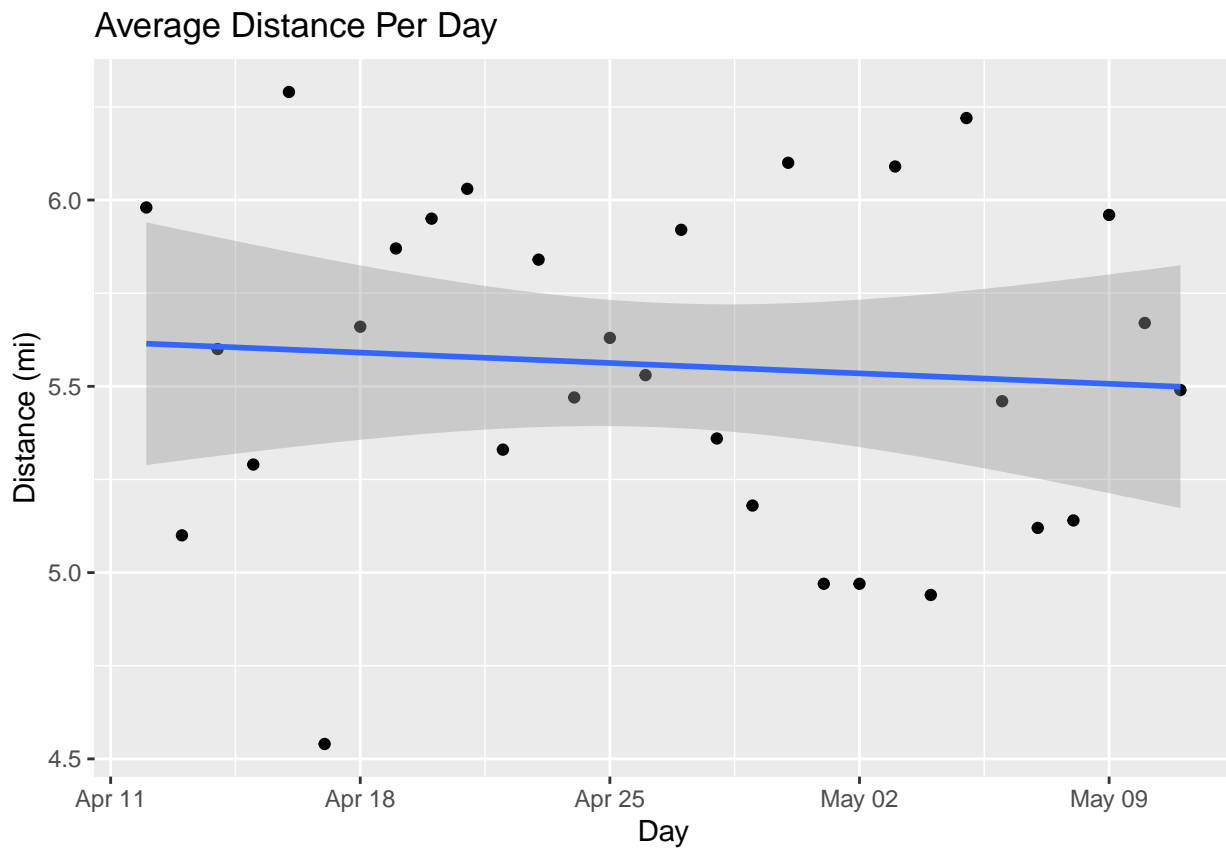
When viewing the results for the distance traveled. The data showed as time went on, the distance users traveled was mostly the same.

```
# loads distance file
library(readxl)
distance <- read_excel("/Users/kahlyllwilson/Desktop/excel sheets/Average Distance Traveled.xls")
dis = data.frame(distance)

#Filters out the outlier
filt_dis= filter(dis, Distance_traveled_Per_Day > 2.5)

# Makes scatter plot for the Average Tracked Distance per day
ggplot(data=filt_dis, mapping = aes(x=ActivityDate, y=Distance_traveled_Per_Day)) +geom_point() +
  geom_smooth(method = lm) + ggtitle('Average Distance Per Day') +
  labs(x = 'Day', y = 'Distance (mi)')
```

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



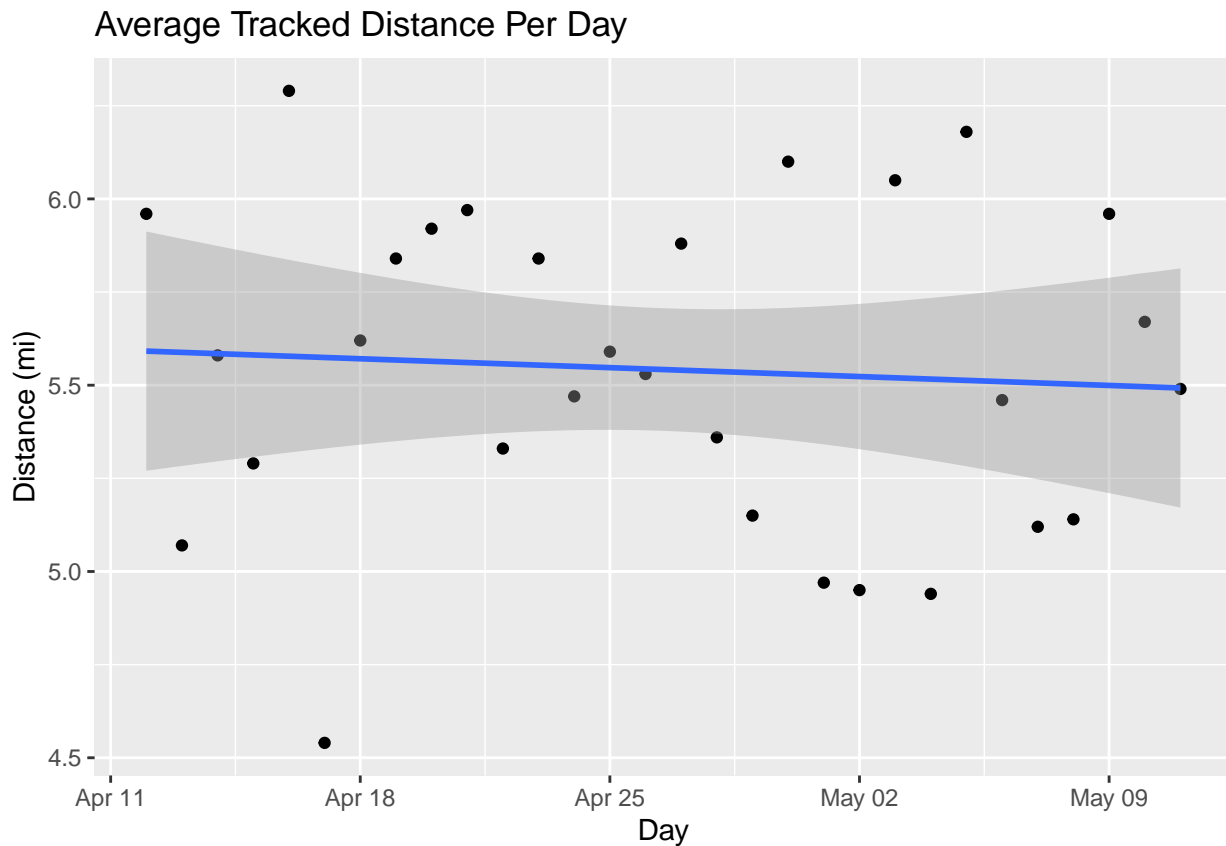
When viewing the results for the distance traveled. The data showed as time went on; the distance being tracked was constant with the distance traveled. This is a positive result, as it shows the tracking on the fitness tracker is accurate.

```
# loads tracked distance file
library(readxl)
tracked_distance <- read_excel("/Users/kahlyllwilson/Desktop/excel sheets/Average Distance Tracked.xlsx")
trac = data.frame(tracked_distance)

#Filters out the outlier
filt_trac = filter(trac, Distance_tracked_Per_Day > 2.5)

# Makes scatter plot for the Average Tracked Distance per day
ggplot(data=filt_trac, mapping = aes(x=ActivityDate, y=Distance_tracked_Per_Day)) +geom_point() +
  geom_smooth(method = lm) + ggtitle('Average Tracked Distance Per Day') +
  labs(x = 'Day', y ='Distance (mi)')
```

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



The full combined graph can be viewed below or on [Tableau](#)

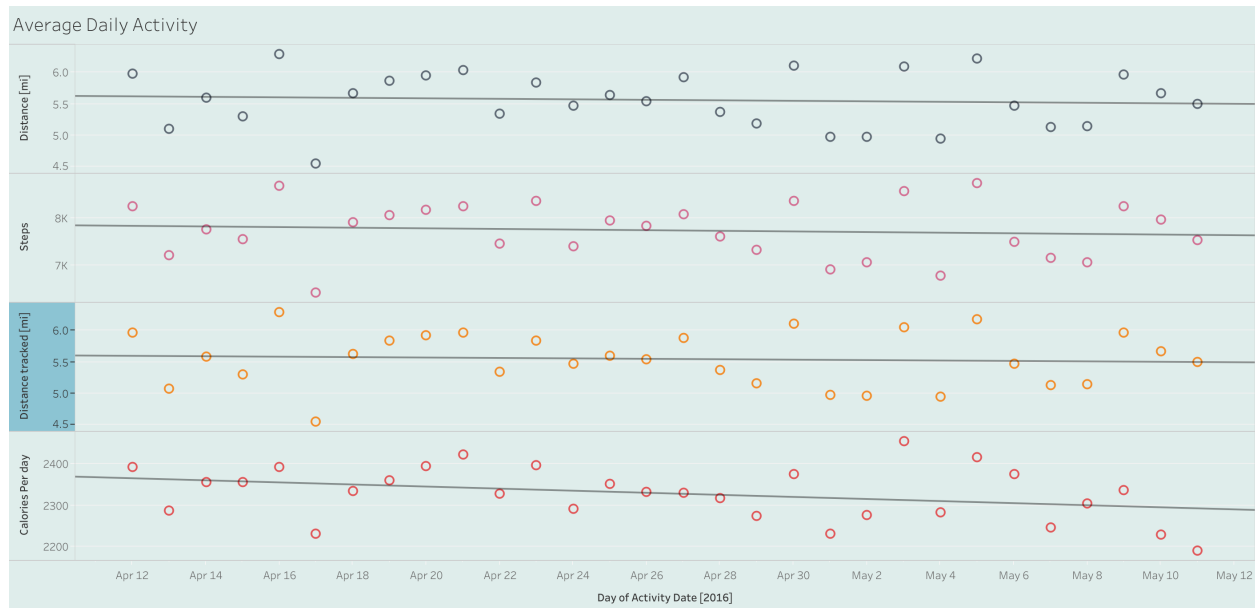


Figure 1: Tableau Table: Activity Over Time

Conclusions

The analysis of the fitness tracker was successful as it provided valuable insights into how users interact with the product. Based on my sample size, I found that users engage more in light activities such as walking and light bike rides rather than moderate or very active activities such as running and rock climbing.

I also noticed that users are not logging their activity, which could be due to forgetfulness or not feeling the need to do so. To address this, I suggest adding a feature that automatically logs the data for the user while still giving them the option to manually log their activity.

Furthermore, I observed that as time passed, the calories burned decreased. This might be a problem as our customers use the product to lose weight or maintain their healthy lifestyle. To address this issue, I propose integrating a weekly fitness goal feature into our app. Users can create their own goals or choose from a few generic ones like walking 10,000 steps per week or running 5 miles per week. The fitness tracker can track their progress daily, and every Sunday the user can receive a report of their activity with suggestions for improvement.

In addition to weekly goals, I suggest having daily advice blogs that take no longer than two minutes to read. These blogs can provide guidance on promoting healthy living and help users stay on track towards their fitness goals.