Case Study: BellaBeat

Kahlyll Wilson

February 20 , 2023

NOTE: code and files can be found in Github Repository

Overview

The following case study was completed apart of the Google Data Analytics Professionals course. It is based on the fictional Company BellaBeat a high-tech manufacturer of health-focused products for women.

In this scenario I was tasked with helping Bellabeat become a potential big player in the global smart device market. The Cofounder and Chief Creativte Officer,Urška Sršen, believes that analyzing smart device fitness data could help unlock new growth opportunites for the company. Along with Sršen another key stake holder is Sando Mur a mathmetichian and Bellabeats cofounder.

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 I have been tasked with completing the following requirements for the key stakeholders:

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 to in your presentation

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

1. What are some trends in smart device usuage?
2. How could the trends apply to the bellabeat customer?
3. How can both my team members and stakeholders use these insishts to make data driven decisoins

Analysis

Data gathering and cleaning phase

The analysis was completed using the FitBit Fitness Tracker Data (FFTD). The FFTD dataset is open to the public and can be found on the Kaggle website.

For the analysis I first cleaned and organized the data using excel when I pulled the raw data file from FFTD I noticed that 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. Using this feature I was able to bring in the data from the four spread sheets and compile it into one data set that I could analyze.

During this stage I noticed something interesting from the data. When it came to overall logging the active distance with the fitness tracker it was very rare or noe at all. There were also inconsistent metrics being recorded when users are either Very active or Moderately active. However, when it it came to light active distance tracking, the results were overwhelmingly positive. With showing results that at least everyday each user was engagin in light activity.

Analysis phase

How anlaysis was conducted

For the analysis portion of this case study it was completed using SQL and 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?

Once I had my results I took them out of the data base and stored them in a seperate folder. Then Used a combination of R and Tableu 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 data there was an outlier throwing off my day and, in each plot, it was a single point on May 12th. For each metric there was a significant drop in the data recorded. There are no significant holidays on this day unless a vast majority of people celebrate National Limerick Day and 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 averge total distance traveled and the total distance tracked are the same because that means that the fitness tracker is working properly and tracking all the miles the users walk.

After reviewing the data for all the data for the averages for across all users. I wanted to see what the users were averaging per 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 treadline to see how it would progress. The Plot shows a downward trend showing that as time goes on users burn less calories

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 treadline to see how it would progress. The Plot shows a downward trend showing that as time goes on users burn less calories

```
  library(ggplot2)
  library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

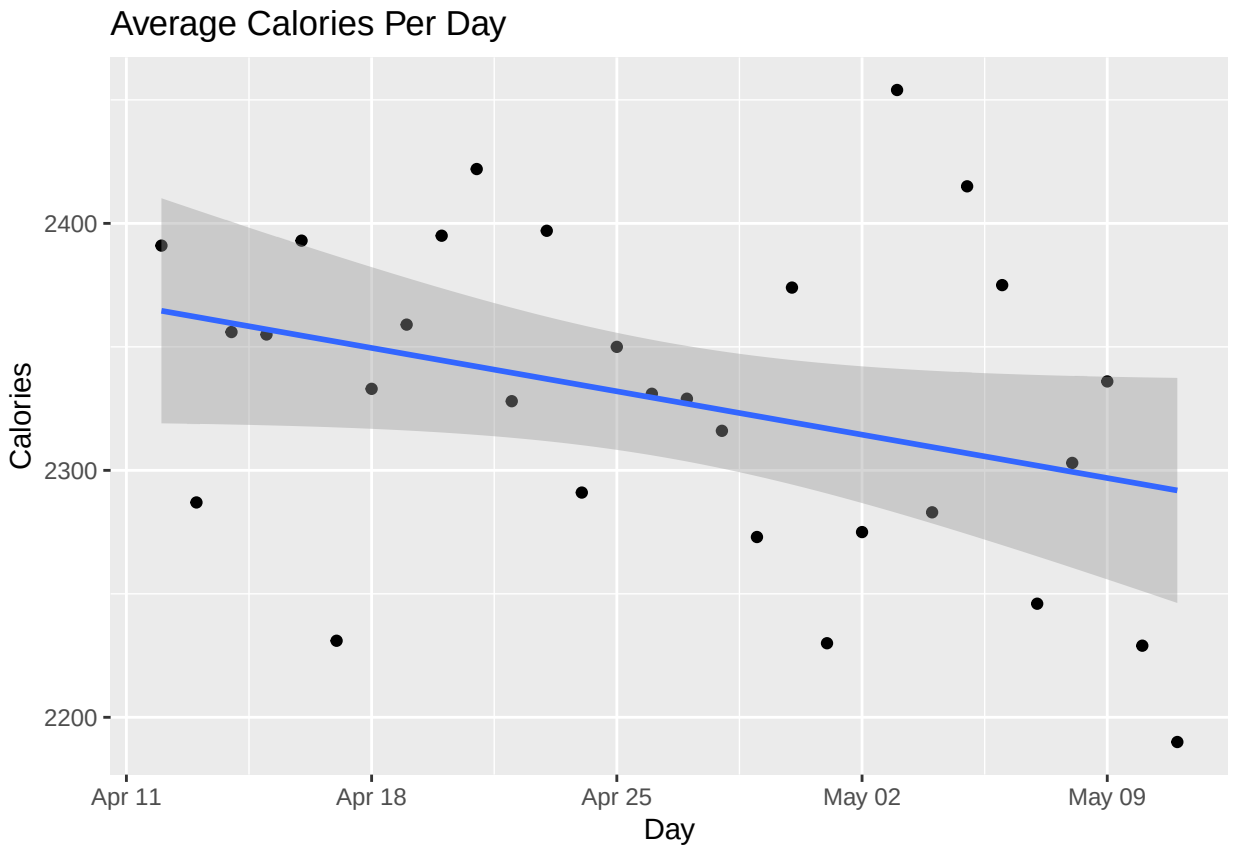
```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
# 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'
```


Average Calories Per Day
Calories
2400
2300
2200
Apr 11
Apr 18
Apr 25
May 02
May 09
Day

When viewing the results for the steps. The data showed as time went on users were taking less steps per day.

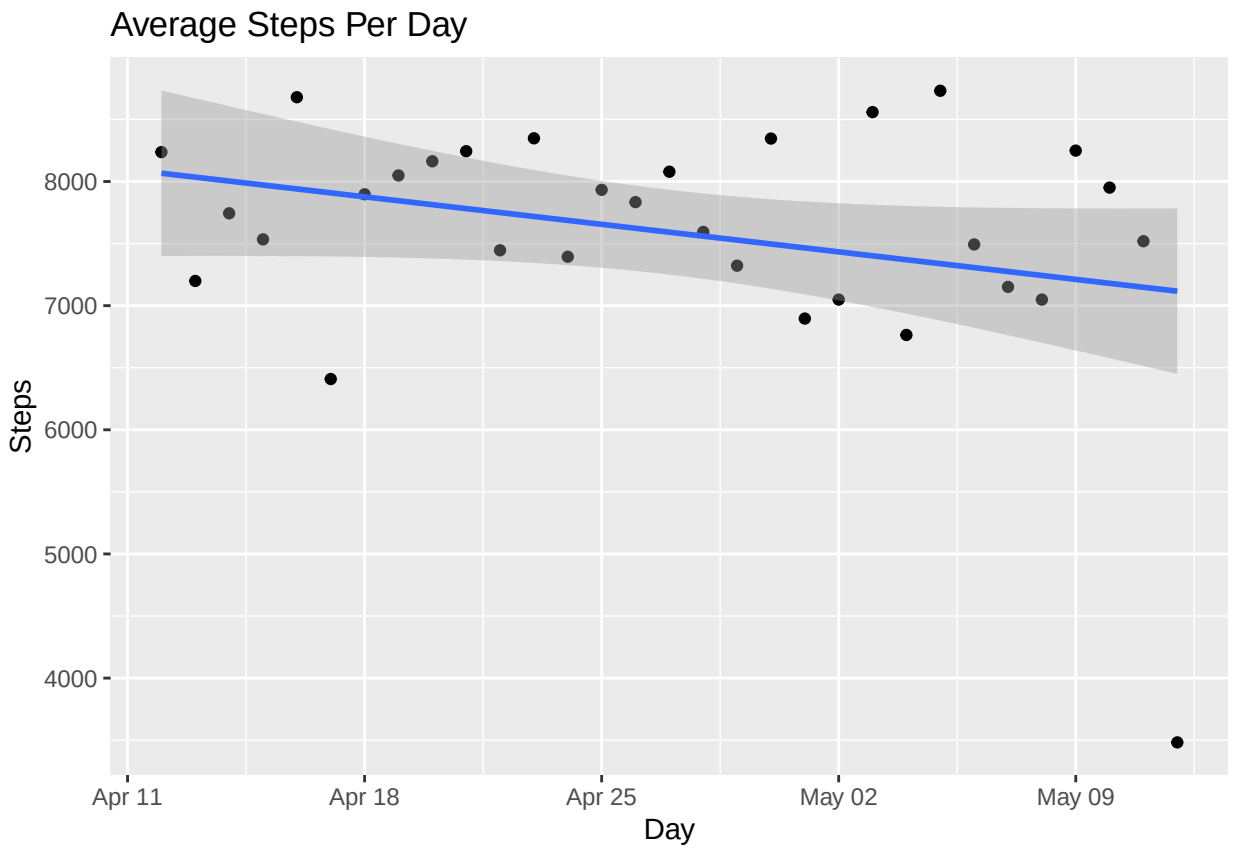
```
library(ggplot2)
library(dplyr)

# 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 the distance users traveled less as time went on.

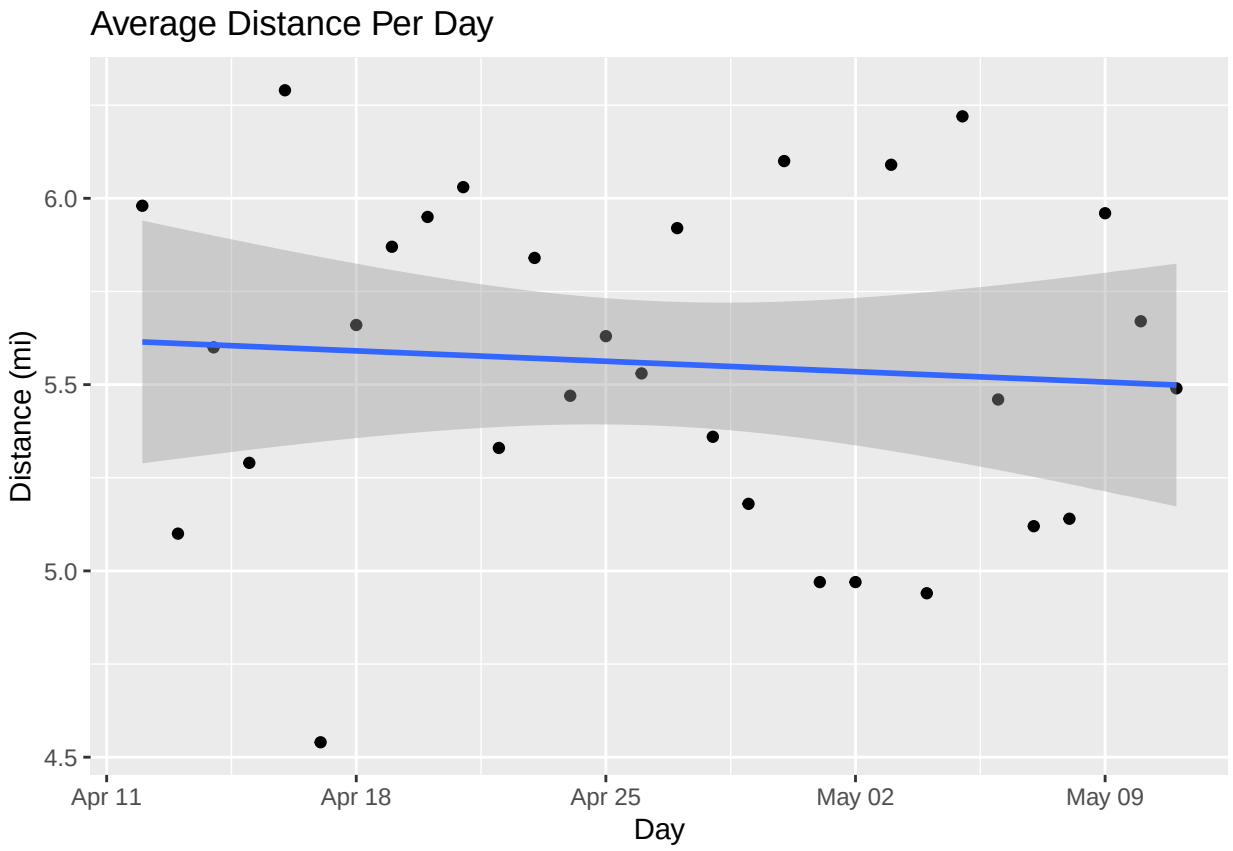
```
library(ggplot2)
library(dplyr)

# 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 the distance being tracked was consitant with the distance traveled. As people traveled less there were less distance being tracked.

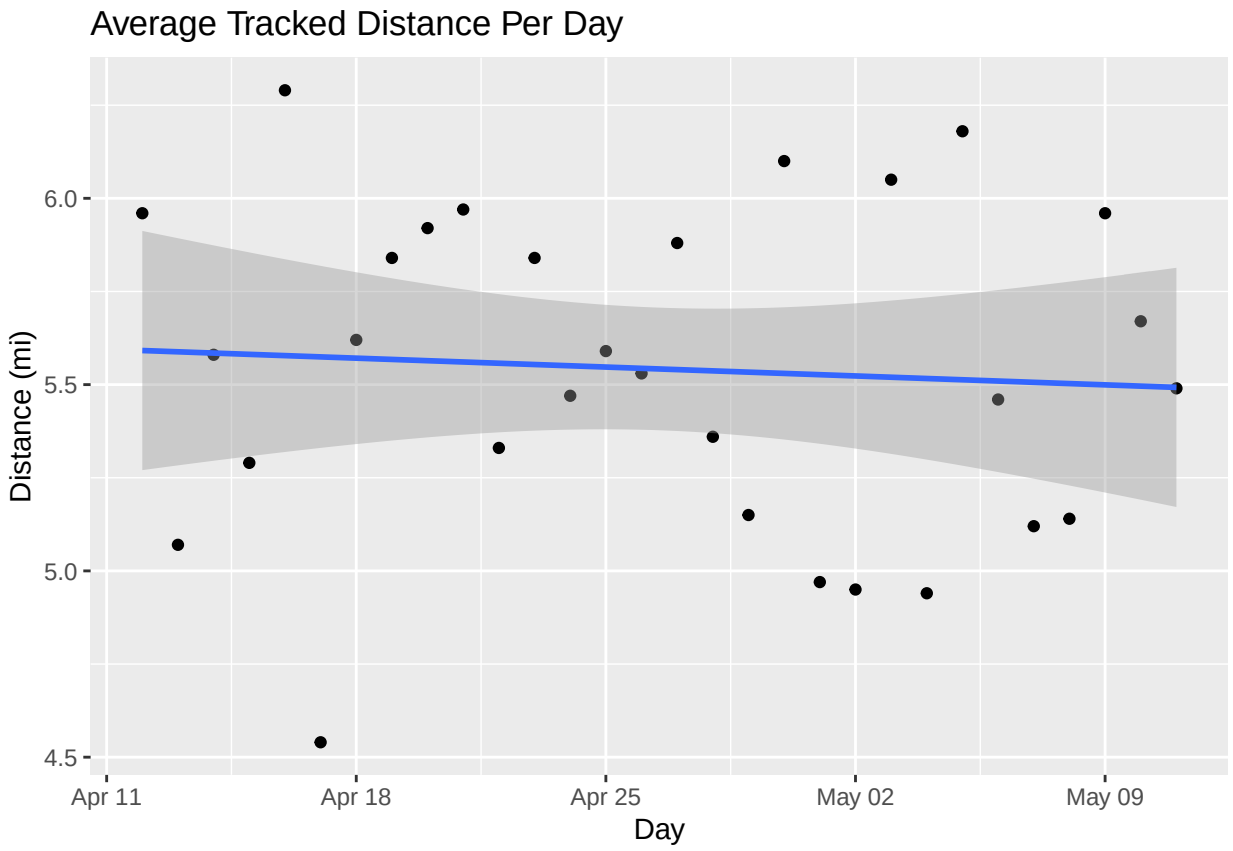
```
library(ggplot2)
library(dplyr)

# loads tracked distance file
library(readxl)
tracked_distance <- read_excel("/Users/kahlyllwilson/Desktop/excel sheets/Average Distance Tracked.xls
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 ='Distan
```

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



The full combined graph can be viewed below or on Tableu

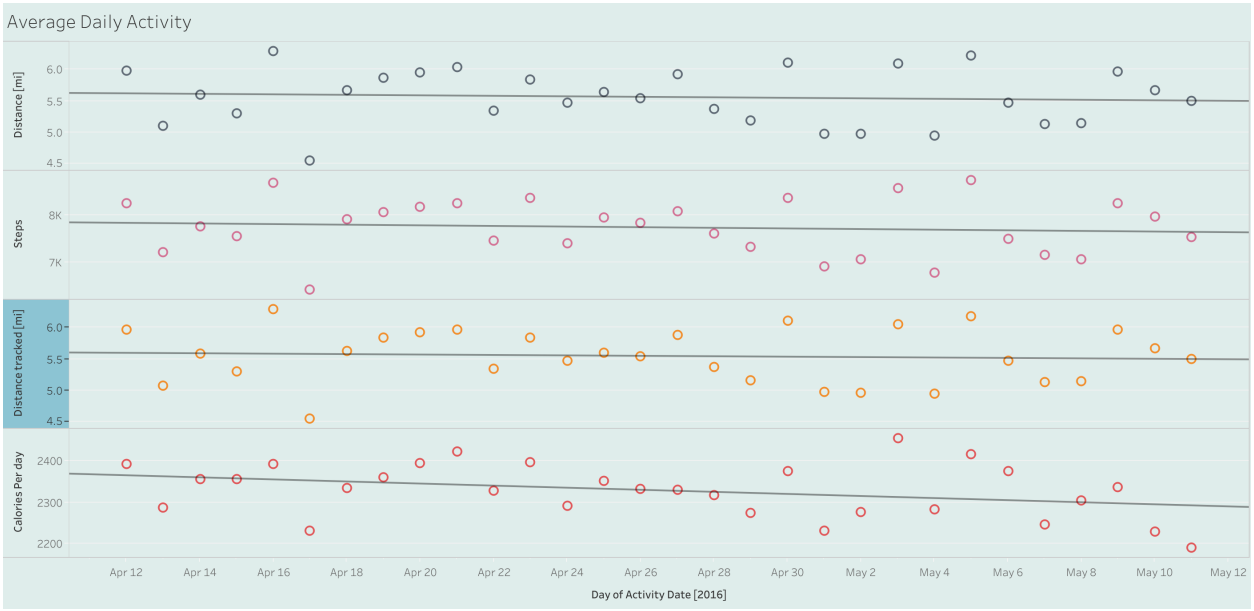


Figure 1: Tableu Table: Activity Over Time

Conclusions