Bellabeat Case Study

Marshall Tolleson

2025-08-13

# Bellabeat Marketing Guidance

### A case study for the Grow with Google Data Analytics certificate

## Introduction

The executive and marketing teams of Bellabeat, a fitness device company, would like to analyze usage for fitness trackers in general to guide their future marketing efforts. They have a suite of fitness products available and want to know how publicly available data might help guide how they market some of their devices or subscription. They have provided a dataset sourced from Kaggle user Mobius which details user data from 33 distinct users including heart rate, activity intensity, and sleep data.

This report will identify which features were most used on the fitness trackers,and any discernible habits from the individuals included in the dataset. After analysis, inferences will be made about which Bellabeat products these results would be most compatible with, and how these insights might guide the marketing team in their future strategic campaigns.

## Data Preparation

The dataset, titled “FitBit Fitness Tracker Data” available at <https://www.kaggle.com/datasets/arashnic/fitbit> was generated by voluntary responses to an Amazon Mechanical Turk survey in 2016. Participants consented to submit their personal tracker data, and access to this data is under a CC0:Public Domain license. The dataset includes 18 tables which have been summarized in 2 tables, one for daily activity and one for sleep only. Analysis of the summary tables will give insight into which features common to fitness tracking devices are being used by the survey participants and their general fitness levels. The sex of the participants is not listed in the data, which may hamper interpretation and application for Bellabeat since their products are designed for women.

The data are mostly clean, but there are obvious typographical errors (such as input of “50” calories in a day from a user who averaged 1500) that would skew summary statistics, so these have been removed. Some “dirty data” values have been retained to allow for “usage” analysis, i.e. a participant may have logged calories only for one day but not their activity. Data cleaning was carried out in Microsoft Excel.

## Analysis

### Feature Usage

It is easy to determine how many participants logged any activity including steps, a workout, or calories, by checking for distinct values in the ID column. We can do the same for the sleep summary table to see how many users used their device to log sleep.

distinct\_count\_activity <- length(unique((dailyActivity\_merged$Id)))  
print(distinct\_count\_activity)

## [1] 33

distinct\_count\_sleep <- length(unique((sleepDay\_merged$Id)))  
print(distinct\_count\_sleep)

## [1] 24

Every survey participant used their fitness device to log some activity, however nearly one third of survey participants did not use their device to track their sleep.

Even though each participant used their devices to log activities, a preliminary glance at the data reveal that not every participant used every feature. To determine how many participants used each feature, we can set up a table that summarizes the average value for each participant in each column:

mean\_table <- dailyActivity\_merged %>%  
 group\_by(Id) %>%  
 summarise(across(2:14, ~ mean(.x, na.rm=F))) %>%  
 select(1:5,13,14)  
print(mean\_table, n=33)

## # A tibble: 33 × 7  
## Id TotalSteps TotalDistance TrackerDistance LoggedActivitiesDistance  
## <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1503960366 12117. 7.81 7.81 0   
## 2 1624580081 5744. 3.91 3.91 0   
## 3 1644430081 7283. 5.30 5.30 0   
## 4 1844505072 2580. 1.71 1.71 0   
## 5 1927972279 916. 0.635 0.635 0   
## 6 2022484408 11371. 8.08 8.08 0   
## 7 2026352035 5567. 3.45 3.45 0   
## 8 2320127002 4717. 3.19 3.19 0   
## 9 2347167796 10077. 6.73 6.73 0   
## 10 2873212765 7556. 5.10 5.10 0   
## 11 3372868164 6862. 4.71 4.71 0   
## 12 3977333714 11338. 7.76 7.76 0   
## 13 4020332650 2267. 1.63 1.63 0   
## 14 4057192912 3838 2.86 2.86 0   
## 15 4319703577 7511. 5.05 5.05 0   
## 16 4388161847 10814. 8.39 8.39 0   
## 17 4445114986 4797. 3.25 3.25 0   
## 18 4558609924 7685. 5.08 5.08 0   
## 19 4702921684 8572. 6.96 6.96 0   
## 20 5553957443 8613. 5.64 5.64 0   
## 21 5577150313 8304. 6.21 6.21 0   
## 22 6117666160 7047. 5.34 5.34 0   
## 23 6290855005 5650. 4.27 4.27 0   
## 24 6775888955 2520. 1.81 1.81 0.0754  
## 25 6962181067 9795. 6.59 6.52 0.324   
## 26 7007744171 11323. 8.02 7.58 2.12   
## 27 7086361926 9372. 6.39 6.39 0   
## 28 8053475328 14763. 11.5 11.5 0   
## 29 8253242879 6482. 4.67 4.67 0   
## 30 8378563200 8718. 6.91 6.91 1.12   
## 31 8583815059 7199. 5.62 5.62 0   
## 32 8792009665 1920. 1.23 1.23 0   
## 33 8877689391 16040. 13.2 13.2 0   
## # ℹ 2 more variables: SedentaryMinutes <dbl>, Calories <dbl>

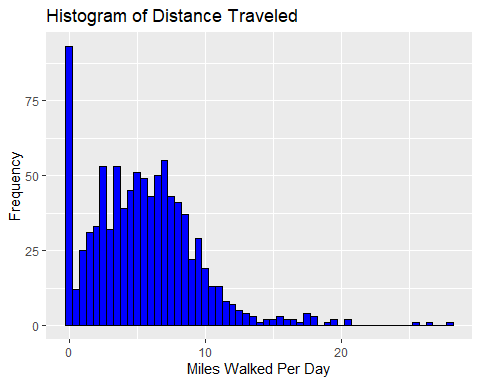
This table shows us several things, intended and otherwise. First, only 4 of the 33 participants are using their devices to log activity distances. Second, there may be some discrepancies in the data since several users are averaging greater than 15 sedentary hours a day and are not logging their activity distance, but are logging greater than 10 miles traveled each day with a corresponding large number of steps. Either these individuals are training for distance race without using their fitness device as a training aid (not logging their workout distance) or the fitness tracker is counting distance traveled in a vehicle as walked miles. By calculating their miles per hour by dividing the “Very Active Distance” value by the “Very Active Minutes” value for each day, the data suggest these individuals were able to maintain a 6 min/mile pace or faster for multiple hours each day, which would make them world class marathoners. It is more likely that the fitness device was not functioning properly.

### User Profile

Bellabeat may also be interested in the habits of the average fitness device user according to these data. A summary table for the average value of each feature is below:

selected\_features <- dailyActivity\_merged %>%  
 select(3:5, 14, 15)  
  
selected\_means <- selected\_features %>%  
 summarise(across(1:5, ~ mean(.x, na.rm = F)))  
print(selected\_means)

## TotalSteps TotalDistance TrackerDistance SedentaryMinutes Calories  
## 1 7669.691 5.512585 5.498173 995.3793 2312.632

The average fitness device user walks 7670 steps for a total of 5.5 miles, is sedentary for 15 hours including sleep time, and consumes 2300 calories in a day, which would describe the lifestyle of a moderately active working adult. If it would be valuable to the marketing team to know the distribution of these activities throughout the day, a closer look at the hourly or by the minute datasets would be useful. Plotting the total distance on a histogram will give a better idea of the distribution of frequencies of distance walked per day and demonstrates that a large number of fitness device users walk very little, despite the average being 5.5 miles. 

This trend is present in each of the selected commonly used features: there is a group of likely typographical errors as outliers, there is a roughly right-skewed normal curve around the mean, and a large group of users near zero. This suggests that there are potentially three different audiences available for the Bellabeat marketing team to address: high activity, moderate activity, and low activity users.

## Discussion and Recommendations

These data align most accurately with the “Leaf” and “Time” wearable devices. These devices track user activity similar to that listed in the above dataset and analysis. According to that analysis, the following approaches are recommended for future marketing efforts:

For the Low Activity User

* Since two thirds of users don’t track sleep, don’t waste time marketing the sleep functions of the devices.
* Almost no users were logging workout activities, so most campaigns targeting the average consumer should not focus on tracking workouts or similar metrics.
* The common theme of not tracking sleep or workouts suggests that most users are looking for a device that seamlessly blends with their current habits, so heavier wearable tech like the “Leaf” may not have percieved value. The “Time” fitness watch is more likely to be what they are looking for.

For the Moderate Activity User

* These users are much like the Low Activity User, except that they are more likely to be monitoring their results and may be more receptive to useful features in the app. Features like goal tracking or sharing activities may even lead them to start logging their activities and would drive interaction with the app.

For the High Activity User

* These users would find value in activity logging and with more in depth analysis like what would be available with the membership subscription. Promoting your products at races or events that attract highly active people would be an effective way to be seen as a quality fitness tracking device.
* This group would also be more likely to monitor every aspect of their wellness, including sleep, so the sleep tracking functionality may be useful in a campaign targeting this group.
* Highly active users will also have higher expectations for device performance, so if Bellabeat devices are known to malfunction less, and not do things like count car rides as active distance travelled, that could be an important part of this kind of ad campaign. “Log your runs, not your car rides.”