Bellabeat casestudy

Yanchen Wu

9/9/2021

Introduction

Background: Bellabeat is a high-tech company that manufactures health-focused smart products. The company is founded by Urška Sršen and Sando Mur.Since it was founded in 2013, Bellabeat has grown rapidly and quickly positioned itself as a tech-driven wellness company for women. Sršen knows that an analysis of Bellabeat's available consumer data would reveal more opportunities for growth. She asks you to analyze smart device usage data in order to gain insight into how consumers use non-Bellabeat smart devices. She then wants you to select one Bellabeat product to apply these insights to in your presentation.

Questions for analysis: 1.What are some trends in smart device usage? 2.How could these trends apply to Bellabeat customers? 3.How could these trends help influence Bellabeat marketing strategy?

Business task: Analyze trend of smart device usage using datasets and apply the trend to Bellabeat customers. Provide high-level recommendations to Bellabeat marketing strategy based on the analysis.

Source of data: FitBit Fitness Tracker Data (CC0: Public Domain, dataset made available through Mobius): A Kaggle data set contains personal fitness tracker from thirty Fitbit users. Thirty eligible 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.

Potential issues with the data: The dataset contains only 30 participants of the Fitbit users, which is not a big population to start with. Moreover, the gender of the participants are unspecified, which could cause inaccurate predictions. Since this is the dataset given, I would still use this dataset and try to give as much insights as I could

Product select

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. The data provided records 30 Fitbit user's daily activity, Leaf is the product that serves similar purpose as Fitbit. Therefore, I believe Leaf is the best product to reflect the analysis outcome.

Cleaning data using SQL and EXCEL

In this procedure, I used EXCEL to check errors for data since the filter and sorting function allow me to scan the data quickly for errors. Then I use SQL to clean the data since some of the datasets are too large to

work on EXCEL. The data has been cleaned beforehand and there is no formatting errors. However, there is a major problem in the data where 77 rows of data in "dailyactivity_merged" dataset that has total_steps of 0 and sedentary time as 1440 minutes (equals to 24 hours). This indicates the participants did not wear Fitbit at all during these days. If we include these data in the statistics, it will greatly underestimate the average activity of the participants and reflect inaccurate research outcome.

(There are also around 53% of the data within "dailyactivity_merged" has the issue where Fitbit did not track the entire day of participant's activity. It is likely the participants did not wear the handband all 24 hours within the day. I still decided to keep these data since they reflected the participant's activity during the day and deleting all the problematic data will result insufficient data for analyze. Therefore, the participant's daily activity will be mildly underestimated)

The same issue happens in "hourlyCalories_merged", "hourlyIntensities_merged", and "hourlySteps_merged" dataset. I managed to delete all the problematic data in the hourly datasets that correspond to the "dailyactivity_merged". Changelog and SQL code are attached at the end for reference.

Data Analysis using R

I chose R as the data analysis tool since R is really powerful for making plots. #### Package setup

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.1     v dplyr   1.0.6
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
daily <- read.csv('cleaned/dailyActivity_merged.csv', header =TRUE)
hourly <- read.csv('cleaned/Hourly_cleaned.csv', header =TRUE)
```

Loading files

```
hourly$Hour <- factor(hourly$Hour, levels = c("0:00","1:00","2:00","3:00","4:00","5:00","6:00","7:00","8
daily$Dayoftheweek <- factor(daily$Dayoftheweek, levels = c("Mon","Tue","Wed","Thu","Fri","Sat","Sun"))
```

Sorting data

```
summary(daily$TotalSteps)
```
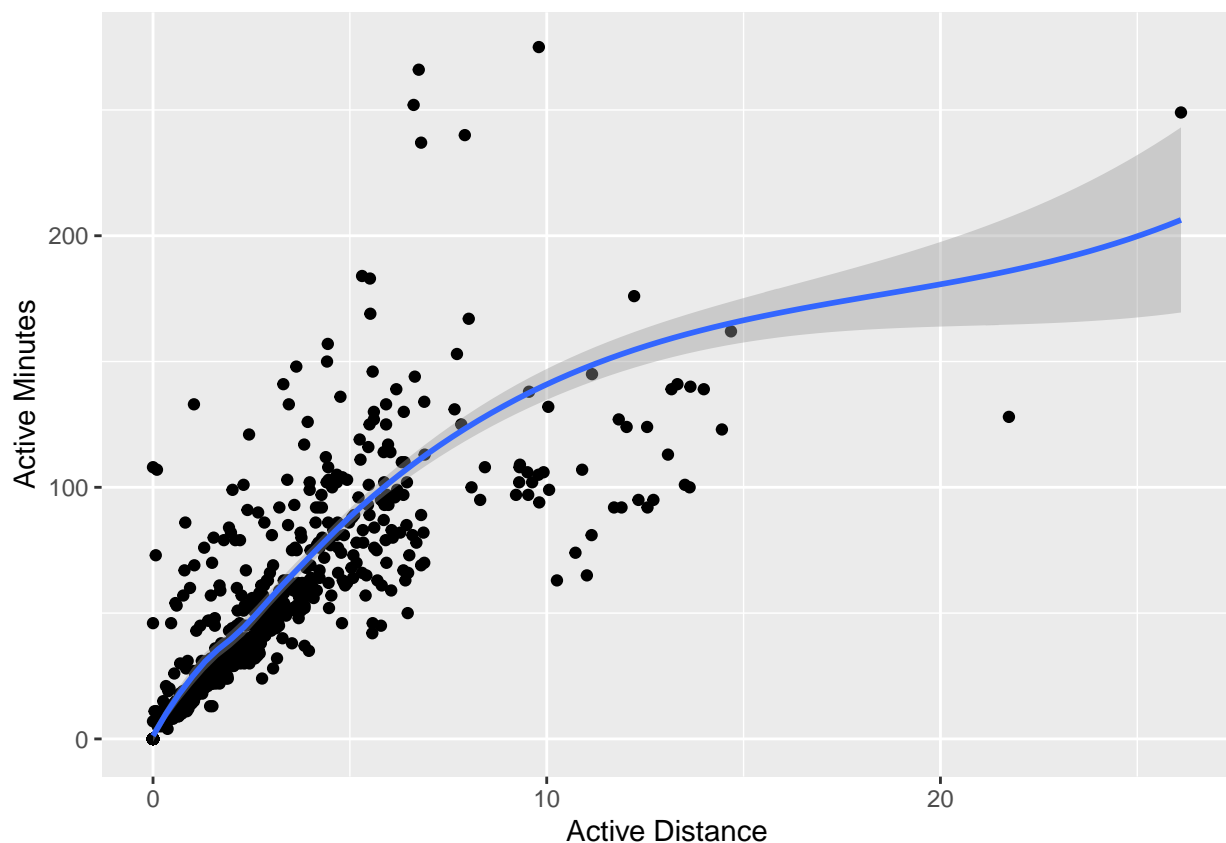
Analysis between Active distance and Active minutes

```
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      4    4923    8053    8319   11092   36019
```

What's worth mentioning is that the average totalsteps of the 30 participants everyday are around 8000 steps. Based on the research done by US National Library of Medicine National Institutes of Health, the average U.S. adults walks around 6500 steps per day. The fitbit participants take considerably more steps than the average U.S. adults every day.

```
ggplot(data = daily) +
  geom_point(mapping = aes(x = VeryActiveDistance + ModeratelyActiveDistance, y = VeryActiveMinutes + Fa
  geom_smooth(mapping = aes(x = VeryActiveDistance + ModeratelyActiveDistance, y = VeryActiveMinutes + F
  xlab('Active Distance') + ylab('Active Minutes')
```

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


We can see from the chart that there is a clear positive correlation between the active minutes and active distance. The graph indicates that the major exercise activity for the 30 participants are walking and running.

```
ggplot(data = hourly) +
  geom_histogram(mapping = aes(x = Hour, y = TotalIntensity), stat = "summary") +
  theme(axis.text.x = element_text(angle = 90)) +
  xlab('Hour during the day') + ylab('Total Intensity')
```

Analysis of active time slot

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```

```
## No summary function supplied, defaulting to 'mean_se()'
```
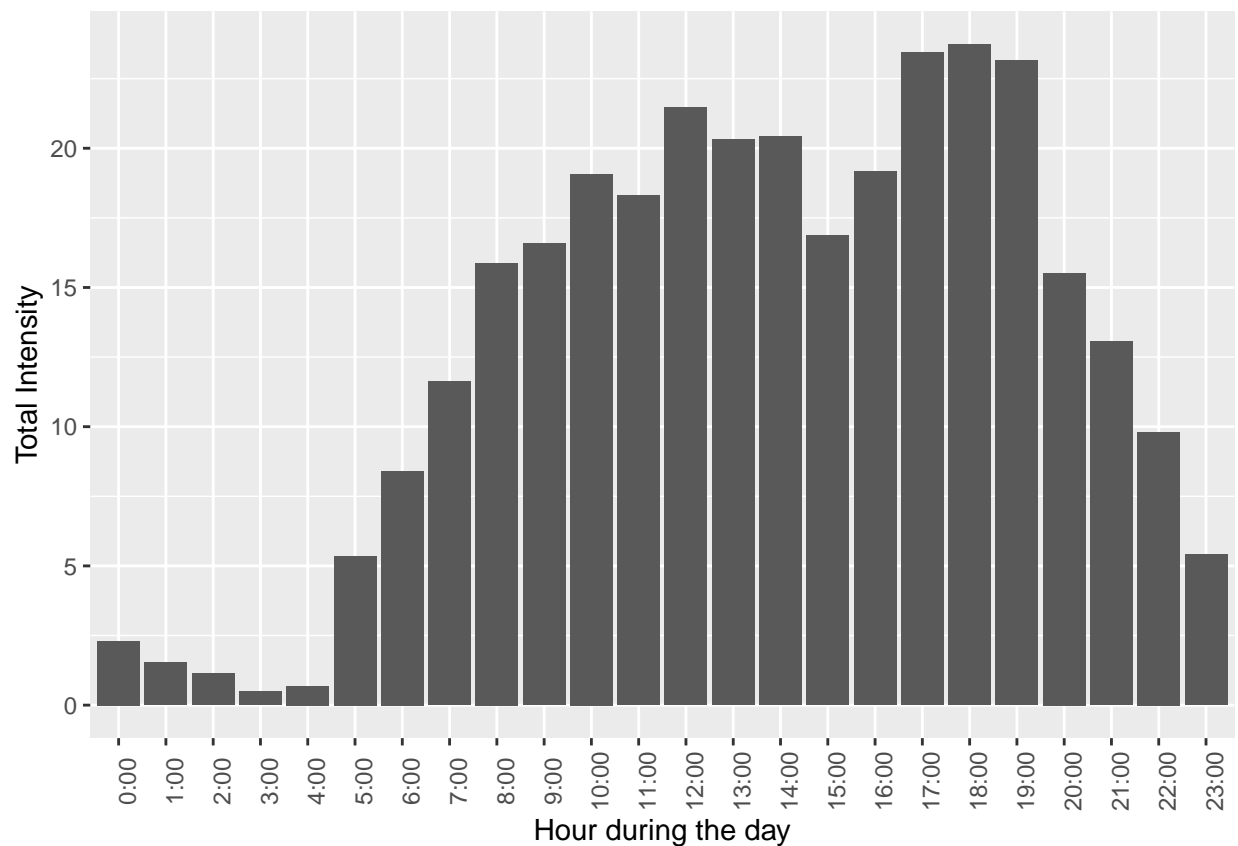

The data shows that the total intensity concentrates mostly during 5PM to 7PM and 12PM to 2PM. Given people usually get off from work during these time slots, participants could use these times for exercise.

Conclusion of the analysis and recommendations for Bellabeat marketing strategy

The dataset involved with the research recorded very detailed data of the participants' daily activity. From this dataset, we can conclude that:

1. The Fitbit users on average takes 23% more steps than average American adults every day.
2. The main exercise preferred by the Fitbit participants are walking and running
3. The participants has the greatest amount of exercise at 5PM to 7PM and 12PM to 2PM

Based on the data analysis, I recommend that Bellabeat sponsor sport events that targets female population. For example, a marathon that raises awareness to breast cancer. Bellabeat can set up booth to display their product, and also provide service like try wearing the Leaf and track their activity. The advantages of sponsoring such event are: 1. it helps Bellabeat to filter out their potential customers. Bellabeat position itself as a tech-driven wellness company for women. Especially for the women who has awareness to keep track of their health and do exercise. The attendants of those events are much more likely to be interested in Bellabeat's product. 2. Sports-related events provide a perfect place for the attends to try Bellabeat product on-the-spot. People are more likely to grow attachment and accept a product after they have tried the product. Therefore increase the successful rate of selling the product. 3. Sponsoring sports event can potentially raise Bellabeat's reputation. Sponsoring marathon for breast cancer shows that Bellabeat cares about women wellness. It could also reinforce the company's marketing position in people's mind. More and more people would know Bellabeat is a tech-driven wellness company for woman when the name is mentioned.

Some extra insights   Based on the daily activity dataset, more than 8% of the time the participants did not wear Fitbit band at all during the day. Adding a function that allows the user to set up a reminder when they are not wearing Leaf could increase the customer stickiness. (Reflects how often customers uses the product) It would be nice if I could have access to Bellabeat's dataset about similar topics. Then I would be able to produce more accurate insights and better recommandations.

Thank you for reading!

SQL Query wrote for analysis:

https://console.cloud.google.com/bigquery?sq=574119381719:71ff4d23cd6b48429c5f878911547ddd

Reference

Tudor-Locke et al. (2011, July 28). How many steps/day are enough? For adults. The international journal of behavioral nutrition and physical activity. Retrieved September 14, 2021, from link.