Bellabeat Case Study

Coursera Google Data Analytics Capstone

1. Summary

Bellabeat is a high-tech company that manufactures health focused smart products. Their products are the Bellabeat app, leaf a wellness tracker, Time a wellness watch, and Spring a hydration tracking water bottle. Using their products customers can track data on activities, sleep, stress, and reproductive health to empower women with knowledge about their health and habits.

The focus of this case study is to analyze fitness tracking data to help unlock new growth opportunities for Bellabeat. This report will focus on Bellabeat app, with some insights into how current products, potential new products, and new app features can help the Bellabeat app meet and surpass company goals.

2. Ask Phase

2.a Business Task

Identify trends in how consumers use non-Bellabeat smart devices to apply insights to Bellabeats's Marketing strategy.

Stakeholders:

- Urška Sršen - Bellabeat cofounder and Chief Creative Officer
- Sando Mur - Bellabeat cofounder and key member of Bellabeat executive team
- Bellabeat Marketing Analytics team

3. Prepare Phase

3.a Dataset used

The data source used for this case Study is FitBit Fitness Tracker Data. This dataset is stored on Kaggle and was made available through Mobius. (https://www.kaggle.com/datasets/arashnic/fitbit)

3.b Accessibility and privacy of data

The dataset is open source. The owner has dedicated the work to the public domain by waiving all of his or her rights to the work worldwide under copyright law, including all related and neighboring rights, to the extent allowed by law. You can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission.

3.c Information about the dataset

The dataset was generated by respondents to a distributed survey via Amazon Mechanical Turk between 03.12.2016-05.12.2016. Thirty eligible Fitbit users consented to the submission of personal tracker data,

including minute-level output for physical activity, heart rate, and sleep monitoring. Variation between output represents use of different types of Fitbit trackers and individual tracking behaviors / preferences.

3.d Data Organization

There are 18 CSV documents available to use in the dataset. Each CSV has different or merged quantitative data tracked by Fitbit. Every user has a unique ID and is tracked by day and time.

3.e Data Credibility and Integrity

Out of the 18 CSV documents available four was chosen for this study. Daily_activity was chosen because it is the merged data of several individual CSV's and it had 33 participants. hourly_steps was chosen with 33 participants. sleep_log and weight_log datasets was also chosen even though the sleep log had 24 participants and the weight log only had 8 participants, which is below the sample size threshold of 30.

There was no data given on gender so it was assumed that the participants were female, since Bellabeat is a female health and wellness company. no other information was given about the participants, no other biases could be accounted for.

4. Process Phase

4.a Tools used

For this phase of the case study, I chose to use Excel and Rstudio to process the data.

4.b Cleaning and Formatting

I began by downloading the CSVs to my computer from Kaggle then previewing the datasets to become familiar with what I had been provided. I chose my 4 documents, because they offered the best overview of the participants, without having to much redundant data or to much of a micro view that maybe useful for customers but not for our purpose. I changed the file formats to excel workbooks so that I could format and standardize the file naming conventions, then using excel I made changes to the column names to make them easier to understand, such as making all of the date information have Date as the titles instead of "date of log" or "date of activity".

Next, I uploaded the new workbooks to RStudio for the rest of the cleaning and formatting. In RStudio I installed the tidyverse package. Using the n_distinct() function for Id columns of each of the datasets to determine the number of customers in each dataset. As previously discussed, this gave a insight to the integrity of the data, as well as an insight into the usage of the app. I noticed that not all of the date formats were not the same across the 4 data sets, and the hourly_steps dataset had date and time data in the same column. To conquer splitting the date and time data I first had to add a date and time column using this code:

```
hourly_steps_1_0 %>%

mutate(ActivityHour = dmy_hm(ActivityHour)) %>%

separate(ActivityHour, into = c('date', 'time'), sep=' ', remove = FALSE)
```

To standardize the date format I applied the next code to all 4 datasets. Code:

```
hourly_steps_1_0$Date=as.POSIXct(hourly_steps_1_0$Date, format="%m/%d/%Y %I:%M:%S %p", tz=Sys.timezone())

hourly_steps_1_0$Date <- format(hourly_steps_1_0$Date, format = "%m/%d/%y")
```

Using the sum() and duplicated() functions I found that only the sleep_log had 3 duplicates, then used this code to remove those duplicates:

```
sleep_log_1_0 <- sleep_log_1_0 %>%

distinct() %>%

drop_na()
```

5. Analyze Phase and Share Phase

5.a Tool used

Rstudio was used to generate plots and analyze data with the ggplot() function in the tidyverse package.

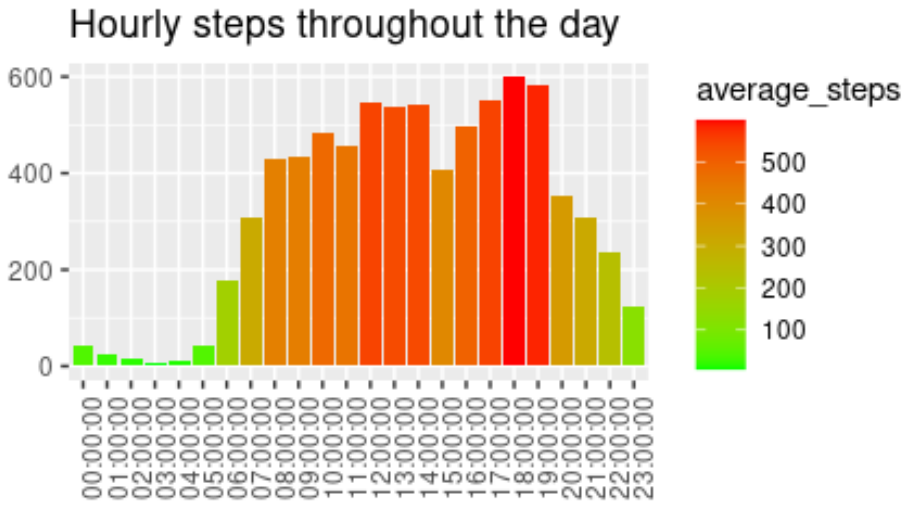
5.b Getting to know the customer

How healthy are our participants? Body mass index (BMI) is an estimate of body composition based on weight and height. BMI is a general tool that for the average person, it is not complete without taking activity level and body fat percentage into account, but it is a great place to start. In general, the scale is: below 18.5 is underweight, 18.5-24.9 is normal, 25-29.9 is overweight, and 30 or above is obese. From the available data our average user is overweight at a BMI of 25.19, with the lowest BMI of 21.45 and highest BMI of 47.54. Another general tool for health is number steps per day.

Our participants averaged 7638 steps per day with an average distance of 5.5 miles traveled. In a 2011 study submitted to the International Journal of Behavioral Nutrition and Physical Activity, it was found that adults take between 4,000 and 18,000 steps per day. It also set forth a scale for activity: sedentary is less than 5,000, low activity is 5,000-7,499, somewhat active is 7,500-9,999, active is 10,000-12,499, and highly active was over 12,500 steps. This puts our average user barely in the somewhat active category. The last barometer for the general health of our customers is their daily calories and activity.

Level of Activity	Average Minutes per Day
Sedentary	991.2
Lightly Active	192.8
Faily Active	13.56
Very Active	21.16

The American Heart Association recommends that adults should get roughly 21 minutes of moderate-intensity aerobic activity (150 min/week) or 10 min of vigorous aerobic activity per day (75 min/week). Our participants on average meet these requirements, while burning on average 2304 calories a day. Finally, to get a full picture, lets examine our customers steps throughout the day.



From the graph it is seen that the activity period is from 7:00 A.M. to 7:00P.M., with the bulk of the steps coming from around lunch time, and after work hours.

In summary, the average customer works a sedentary job, sets out time to work out, meets activity standards, but still falls into a somewhat active and overweight category. In the table above it shows an average of 16.5 hours of sedentary time, could this be a limiting factor the health success of our customers? The next section will explore the date more deeply to find correlations in app participation, app engagement and test what limiting factors such as sedentary time might be hindering customer and company growth goals.

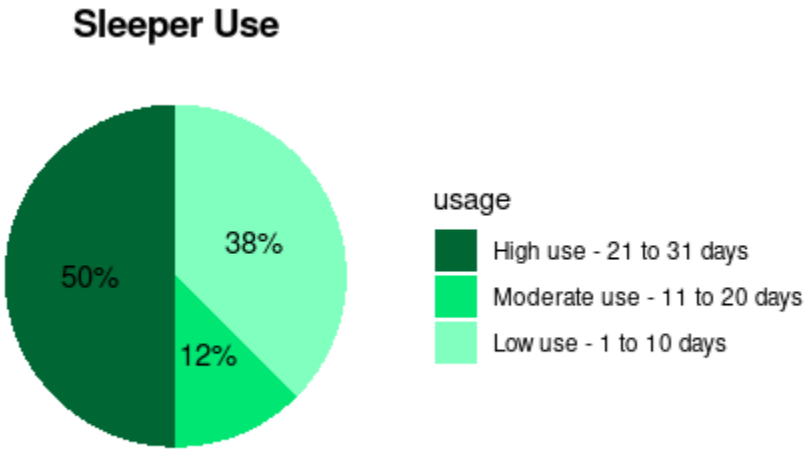
5.c In-depth Analysis

First, app participation can be considered through this table:

Data set	Participants
Daily Activity	33
Sleep_Log	24
Weight_Log	8

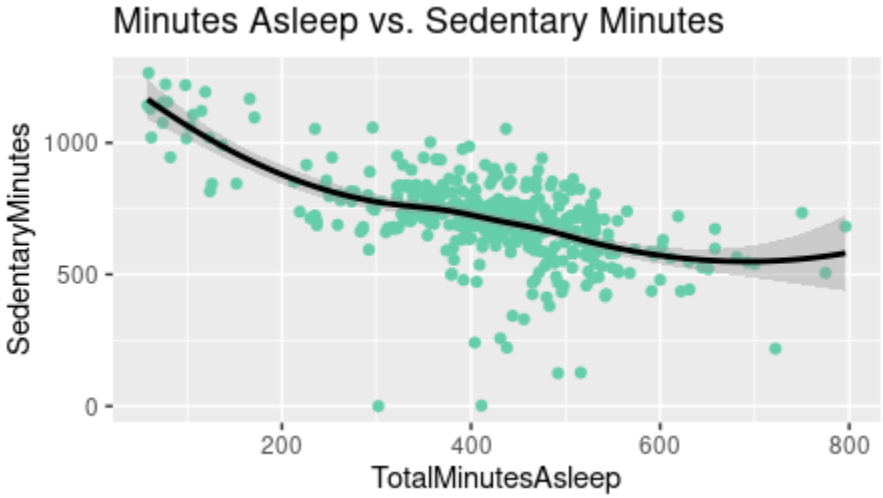
This table was generated as discussed in the data credibility and integrity sections. From this table and the data collected we can see that the total number of participants was 33 that means that, 72% of the users engaged with tracking sleep and only 24% used the product to track weight, BMI and body fat percentage. To put this in perspective of Bellabeat products 33 people would use the Leaf or Time products, of those 33, 72% sleep wearing those products, and 24% would utilize the weight tracking features of the

Bellabeat app. There was no data provided for hydration so the Spring lacks representation. Of the 8 users who used the weight tracking feature only one user logged more than 2 times and they only logged 5 times. The table below shows the usage of the 72% who sleep with the Leaf or Time products.



This would show a 50% retention or drop off over the course of 30 days in users.

As we seen above our customers have a lot of sedentary time and tracking sleep seems to be important to a majority of users.



This graph shows that the more time spent sedentary the less time users sleep, and everyone wants more sleep.

6. Conclusion

Bellabeat's goal is to empower women with knowledge about their own health and habits. Using that knowledge women can take back control of their health, and dial in how macro or micro they want that control. The focus I selected was the Bellabeat app and how it interfaces with each product.

I would first recommend a deep look into this business task using Bellabeats data as well as user polls. This data set was missing key demographic information such as age, and it was also missing key data to analyze all of Bellabeats products such as the Spring. The limited data also did not have data to analyze app features like stress and menstrual cycle tracking. This data set also did not have enough participants to be completely credible, but I believe that the insights that can be gained from the lack of data are just as important as the ones one can gain from data. Following I will discuss the insights found along the way and the recommendation those insights produced. The break down will be user insights, weight tracking engagement, Sleeper use fall off and lack of data insights.

It was found that while our 33 participants met the number of intense or moderately intense exercise, they are still on average only somewhat active and overweight. Their hourly steps data even supports that they most likely have a sedentary job. Setting a base line step goal for users in the Bellabeat app of 10,000 steps per day, along with push notifications every half hour to get up, stretch and get a few steps in will get users into the active range of steps, and help reduce their sedentary time. The added benefit of this will also help them increase their sleep minutes.

From the data it was shown that the users have a poor engagement and commitment to the weight log feature. To get a better picture of this problem it would be best to poll users and get more data on what would make this feature more beneficial to their health. Along with this poll here are some recommendations to poll and consider:

- Education on how to use body metrics and what they mean, and the benefits of low body fat percentage such as losing visceral fat (fat around the organs) provided by Mindfulness habits
- Having daily healthy recipes and food tips in the Mindfulness habits
- Self set fitness, body metric, and nutrient goals. Along with calorie intake and expenditure tracking. This could allow over time for tailored, slow and control dieting through the app. This would need a new app feature to track all body metrics and calculate a controlled calorie drop as well as keeping nutrients up.
- Fitness point system and competition with friends.

- A new product a scale that connects to the app so that the user doesn't need to manually input the metrics in app.

It was also found that while the sleep tracking function was used by the majority of participants only 50% stayed engaged. It would be good to poll the users as to how they feel about wearing the Time or Leaf at night. This could be due to charging patterns, or the comfort level of the products to sleep in. multiple batteries or a specific sleep focused product may be the answer.

Finally, it was discovered that the data lacked insights into the Spring, menstrual health data, and stress data. If the Bellabeat products has this data it would be beneficial to preform another case study with Bellabeat product data and new data from the recommended polls. Some last recommendations in regards to the lack of data:

- Perspiration tracking with Spring hydration tracking for the right amount of water intake (don't forget salt is important too!)
- Cornering the market in menstrual health tracking could be aided by adding temperature tracking to the Leaf and Time. By adding a fertility awareness method (FAM) to the Bellabeat app along with body temperature tracking, would allow for users to predict their cycle better and notice changes in health quicker.
- Subscription scale based on how much is tracked (if you only want to track steps and tell the time why pay for water tracking)

Bellabeat offers a tool to measure, track, and take control of their health. A Bellabeat user could wake up after a good rest, weigh themselves on a Bellabeat scale, that measures and updates their body metrics to their app. By the time they are ready to start their day Bellabeat has used their health goals, body metrics and yesterday's successes to set new goals, suggest a new recipe for breakfast or article on intermittent fasting. At work they get a notification to stretch and take a sip from the Spring since yesterday they had a harder workout, they may need more water than normal. Then right before their work out they notice a friend just earned a few more points in a fitness challenge and they decide to pick a more difficult challenge today. Finally, they get a warning that PMS symptoms may be on the way and to watch out for cravings, or that now is the perfect time to start a new family. Bellabeat is more than a tool, it is a friend watching out for your health and giving you as much control as you want over your health.