Case Study 2: How Can a Wellness Technology Company Play It Smart?

Introduction

Bellabeat, a high-tech manufacturer of health and fitness products for women have tasked me, a junior data analyst, with analyzing fitness device data to generate insights for how the company should move forward with their marketing strategy. The stakeholders include Urška Sršen the cofounder and CCO, Sando Mur the mathematician and cofounder, the rest of the executives, and the Bellabeat marketing analytics team. The insights generated by the data analysis will be applied to one of Bellabeat's products. Specifically I chose the Leaf, a wellness tracker that can be worn as a bracelet, necklace, or clip due to its versatile and compact nature. As a result, I expect it to be quite popular among the line of products Bellabeat has to offer.

Ask

The Business Task: What are some trends found in the smart device data set, how can they be applied to Bellabeat customers, and how should Bellabeat adapt their marketing strategy for the Leaf product based on these observations?

Prepare

The data set used for this case study is open source and provided by user Mobius on Kaggle. There are 18 csv files on FitBit users who consented to the submission of their data including physical activity, heart rate, steps, and sleep from 4/12/2016 to 5/12/2016. The main limitations of the provided data set is the small sample size and potentially outdated data as it is 7 years old. Despite these limitations, I will operate under the assumption this work is preliminary research to help guide Bellabeat in a general direction to narrow down market strategy options.

Process

For this dataset, I will be using Excel and BigQuery to clean the data as they are the most appropriate tools for small to mid-size sets. Below is documentation of the changes I made using the respective programs. A full list of queries used on the dataset is documented in the bellabeatquery.sql file on [my GitHub page](). Due to the limited nation of the data, there will be some additional queries I showcase had the dataset been sufficient or simply for generating fun insights.

Excel:

- Remove Duplicates was used to remove identically duplicate data values.
- Clear formatting and the find and replace features were also used to make the data format more consistent while removing any spaces between values.

BigQuery

- ALTER TABLE, DROP COLUMN was used to remove "Fat" column as only two rows had values out of the 50, indicating the data will not generate any meaningful insights.
- ALTER TABLE and UPDATE were used to create a new column that adds up the 3 types of active minutes as ActivityHours.
- Count distinct query was used at the end to check the number of users whose data was submitted for the corresponding datasets; dailyActivity_merged.csv had 33, sleepDay_merged.csv had 24, and weightLogInfo_merged.csv had 8.
- Based on the results of count distinct querying, the weightLogInfo_merged.csv data set will not be used due to insufficient data.

- Instead, we will be focusing on dailyActivity_merged and sleepDay_merged as the two CSVs aggregate data from the other files. The two files were uploaded to BigQuery as tables named “dailyactivity” and “sleepday” respectively.

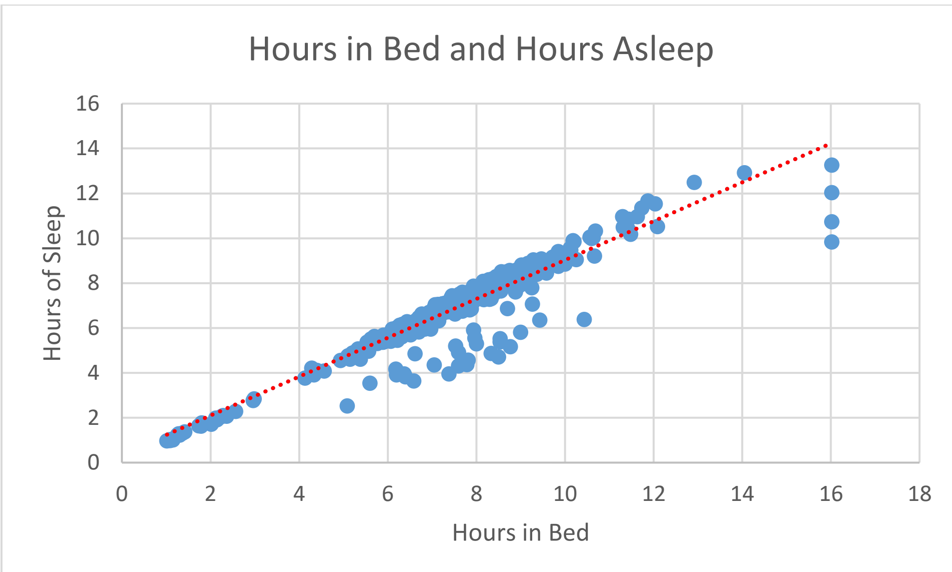
Analyze and Share

Using SELECT and AVG, I created a table that gave the averages for the activity and sleep data along with converting minutes to hours. The calculations show users on average take 7637.91 steps, move 5.5 miles, are active for 3.79 hours, sedentary for 16.52 hours, burn 2303.61 calories, sleep for 7 hours, and spend 7.64 hours in bed a day. According to Thom Rieck of the Mayo Clinic, the average American takes 3,000 to 4,000 steps a day, which equates to 1.5-2 miles. Additionally, the Mayo Clinic states the U.S. Department of Health and Human Services recommends people to be active for 2.5 hours a week. To compare users with those metrics, I used case statement queries to categorize the average number of steps and active hours, then grouped them by user ID. It seems the FitBit users in the dataset well-surpass the average American and U.S. Department of Health and Human Service’s recommendation.

Too look at the users’ sleep activity I did a similar case statement to create a table. For entries where users got less than 5 hours of sleep, I labeled them as sleep deprived, between 5 and 7 as insufficient sleep, 7 to 9 as sufficient sleep, and more than 9 hours as overslept. In this query, I included a count of each of these categories with the result showing most results in the sufficient and insufficient sleep buckets. It was quite rare for users to oversleep or be sleep deprived.

Next we will check for correlations between various variables using the CORR query in SQL. To also check for correlation between sleep and activity levels, I used a JOIN query to create a new table. I found the new table I created only had 31 entries so the data does not seem sufficient. Using left and right join clauses to verify, I can confirm many users tracking their activity do not log sleep data and vice versa. As a result, analysis will focus on variables within their own tables.

Hours Asleep and Total Hours in Bed

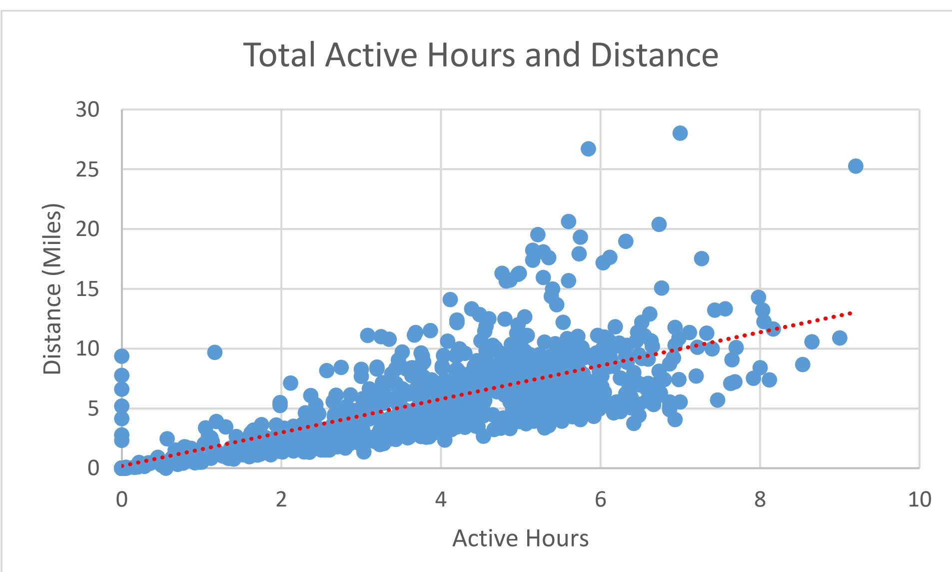


I found total hours in bed and hours asleep to be positively correlated with the value being 0.9304. Another trend I noted was hours in bed were always slightly longer than hours asleep. This makes sense to me as people tend to spend some time winding down before falling asleep.

Total Steps Taken and Calories Burnt

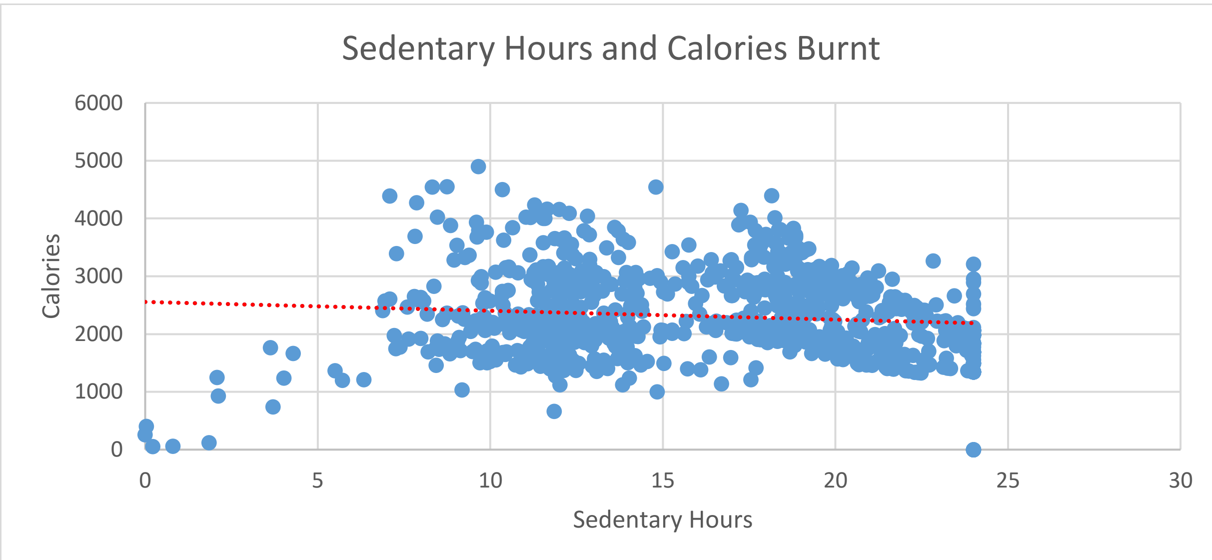
For total steps taken and calories, the r value was 0.5916, indicating a moderately positive correlation. I was surprised to see the r value not as high, as there are other means to burning calories and the intensity of the action matters more.

Total Active Hours and Total Distance



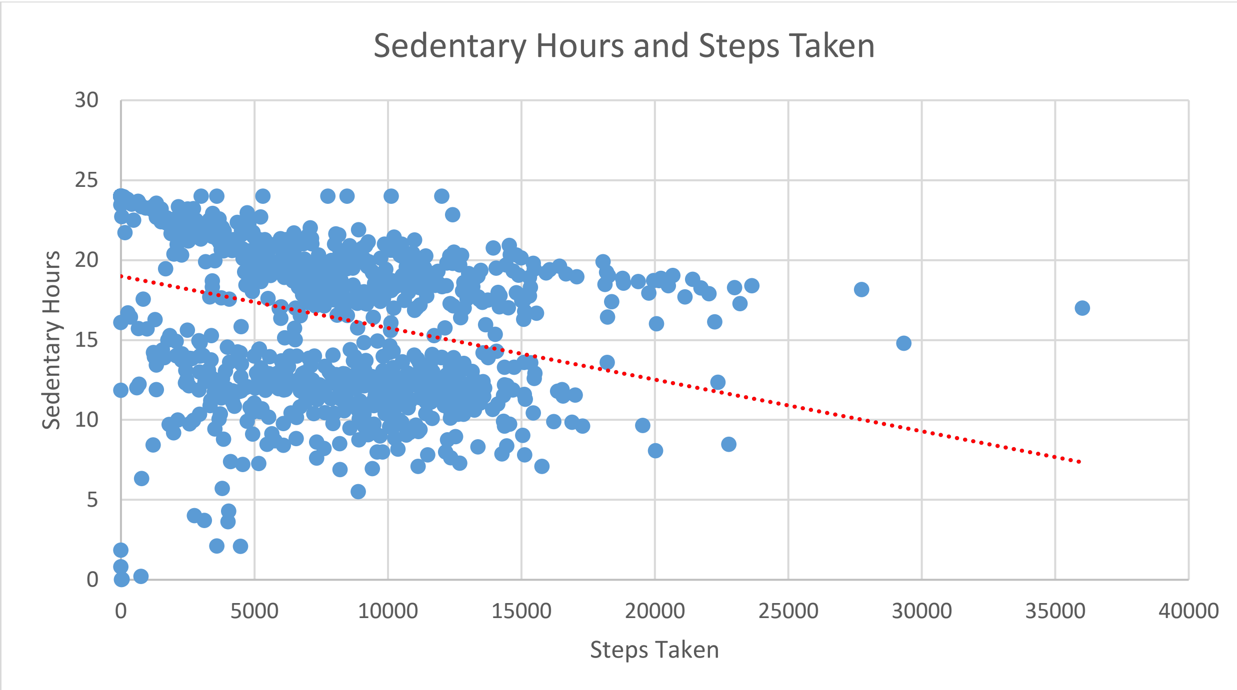
The r value for total active hours and distance is 0.7226, indicating as users are more active, the distance they travel is higher. This result is not that surprising as some cardio activities involve moving distances while other physical activities do not require it.

Sedentary Hours and Calories Burnt



As for the correlation between sedentary hours and calories, it is -0.107, meaning slightly negative. This is interesting to me as the r value is so close to 0 that there is almost no correlation at all. Potentially, this shows regardless of how long the users are sedentary for, they'll burn a wide range of calories depending on other factors such as active hours, distance traveled, or steps taken instead.

Sedentary Hours and Steps Taken



As for sedentary hours and steps, there is a slightly negative correlation of -0.3275. While not as strong as I had anticipated, I believe it still makes sense. While users could spend long amounts of time sedentary, they can still get large quantities of steps in a small amount of time through running, jogging, or power walking.

Act

Due to the limited data source provided to me, here are the following marketing and development recommendations for Bellabeat's marketing team and executives moving forward. Additional points emphasize the use of surveys and questionnaires to help guide the marketing and data analytics by providing more data.

1. For the sleep tracking function of its products, Bellabeat could consider making these features more known to its user base. This can be done through promoting it in advertisements or emphasizing its presence through the manual.
2. The wide range of activity data demonstrates Bellabeat should play into this by enabling and highlighting features allow users to input their activity or fitness goal for the device to track. These notifications and alerts serve as reminders and motivate individuals to stick to their goals whether its health maintenance or weight loss.
3. To motivate users to reach their sleep or activity goals, one milestone reward could be a free trial subscription to Bellabeat's membership program. This allows users to sample bonus features and maximize the Leaf's use.
4. Whether to pursue promoting the sleep tracking function or any other features of the Leaf, Bellabeat should survey its user base to assess which features are popular and unpopular among them. This also allows Bellabeat's developers to know whether certain features need tuning. For example perhaps the sleep feature isn't being used because the Leaf is uncomfortable to wear during sleep or its feature does not fufill its consumers' needs of monitoring sleep apnea.
5. To be more confident in moving forward with a marketing strategy, Bellabeat should expand the sample size of its data by either using historical user data, gathering more user data, or finding more FitBit data as a proxy.

Works Cited

Riek, Thom. (n.d.). *10,000 steps a day: Too low? Too high?* Retrieved May 30, 2023 from: https://www.mayoclinic.org/healthy-lifestyle/fitness/in-depth/10000-steps/art-20317391