Bellabeat case study

Using Excel, SQL & Tableau

BY: SARAN SOMASEKARAN

Table of Contents

1. Company Summary

1.a Company Background
Bellabeat is a health tech company that specializes in creating wearable technology for women. Their products focus on providing holistic insights and tracking for women's health and wellness, including activity tracking, stress management, and menstrual cycle tracking. Bellabeat aims to empower women to take control of their health by providing personalized data and insights.

1.b Company Products
The company's wearable products include:

1. **Ivy** – "a health tracker disguised as smart jewellery"
2. **Time** – "an elegant hybrid wellness watch"
3. **Leaf** – available in three styles: chakra, urban, & crystal and can be worn as a necklace, bracelet or clip. This was Bellabeat's classic wellness tracker.
4. **Spring** – a "smart water bottle" designed to track your drinking / hydration habits.

All the company's wearables sync to their Bellabeat app where members can check their metrics. The wearables track activity (steps taken, distanced travelled, calories burned and activity minutes) & sleep and through the app you can also track your menstrual cycle, hydration (if not using Spring) & meditation. Their IVY wearable also tracks heart rate metrics.

Bellabeat also offers a Wellness Coach app with "unlimited access to 400+ education video, written, and audio content from areas of beauty, fitness, mindfulness, women's health and more" through their Bellabeat+ membership.

2. ASK Phase

2.a Business Task
Utilizing the Fitbit Fitness Tracker Data, identify some trends in smart device usage, how these trends can be applied to Bellabeat's customers and how they can help influence Bellabeat's marketing strategy.

2.b Stakeholders
- **Urška Sršen** – Bellabeat's cofounder and Chief Creative Officer
- **Sando Mur** – Mathematician and Bellabeat's cofounder
- **Bellabeat's marketing analytics team** – a team of data analytics

3. PREPARE Phase

3.a Data Used

The data source used for this case study is **FitBit Fitness Tracker Data.** This dataset was downloaded from Kaggle where it was uploaded by Möbius.

3.b Accessibility & Usage of Data

The dataset was published by Möbius to Kaggle.com under the CC0: Public Domain Creative Common License – waiving all rights to the work and allowing for the dataset to be copies, modified, distributed and performed without asking for permission. Möbius cited the dataset from Zendo: Furberg, Robert; Brinton, Julia; Keating, Michael ; Ortiz, Alexa [Source Here]

3.c Data Summary

According to the dataset information submitted on Zenodo.org,"this 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." Additionally, "Variation between output represents use of different types of Fitbit trackers and individual tracking behaviors / preferences."

3.d Data Organization

Eighteen datasets were downloaded from the **FitBit Fitness Tracker Data.** The datasets downloaded in .csv file format and included long and wide formats. The datasets chosen for analysis below included a user count of 33 participants over a 31 day period of time.

3.e Data Limitations & Integrity

The **FitBit Fitness Tracker Data** was collected in 2016 making the datasets outdated for current trend analysis. Additionally, while the data initially states a time range of 03-12-2016 to 05-12-2016, after data verification, the data collected was only during a 31-day period (04-12-2016 to 05-12-2016). Since the data only included instances over a 31-day period, the timeframe for a more insightful analysis is relatively small.

Lastly, the sample size itself could create a sample bias. While a sample size of 30 (*our data later shows a sample size of 33*) will hold up within the CLT theorm, a larger sample size will be more representative of the population and would increase the confidence interval. Additionally, since there was no demographic information collected it will be hard to see if we have a true representation of a national or global population. This lack of demographic information will also limit recommendations on the target audience (including gender, location, age and job status) and where to market to them. Considering that Bellabeat is primarily targeted to women and individuals who menstruate, having demographics would have bolstered any recommendations after analysis.

4. PROCESS Phase

4.a Datasets Selected
For the case study analysis the following datasets were chosen:

- Daily_Activity_Merged
- Daily_Sleep_Merged
- Hourly_Steps_Merged
- Hourly_Intensity_Merged
- Hourly_Calories_Merged
- Heart_Rate_Merged
- Weight_Log_Merged

4.b Using SQL to Clean Data
Each dataset was cleaned using SQL. The following steps were taken within each dataset:

- GROUP BY clause to obtain the count of unique users by ID in each dataset.
- COUNTIF function to check for duplicate data within each dataset.
- CAST function to format date data into MM/DD/YY format.
- CAST function to format numerical data into Number format with up to 2 decimals.
- MIN and MAX functions to find the first and last date of the dataset.
- SUBSTRING function to separate Date and Hour into two columns when needed for later analysis.
- LEN function to check the length of the ID and other columns to ensure data uniformity.

After the cleaning process was finished, only 3 rows of duplicate information were found within the Daily_Sleep_Merged file. These were removed before analysis.

5. ANALYZE & SHARE Phases

https://github.com/Amrish6/BellabeatProject/blob/main/Analysing_Fitabase.sql

5.a SQL Dataset Upload

Uploaded the following clean data sets:

- Daily_Activity_Cleaned
- sleepDay_Cleaned
- Heartrate_seconds_Cleaned
- HourlyCalroie_Cleaned
- HourlyIntensity_Cleaned
- HourlySteps_Cleaned
- weightLogInfo_Cleaned

5.b User Verification

Checked for # of participants by counting number of distinct Ids in each dataset.

I repeated the SQL query above with each dataset (changing the FROM clause each time) and recieved these results:

- Daily_Activity_Cleaned - 33
- sleepDay_Cleaned - 24
- Heartrate_seconds_Cleaned - 7
- HourlyCalroie_Cleaned - 33
- HourlyIntensity_Cleaned - 33
- HourlySteps_Cleaned - 33
- weightLogInfo_Cleaned - 8

Both the Heart Rate and Weight Log datasets do not include enough data to move forward with analysis. These datasets will not be used.

5.c User Insights

5.c.i User Usage of Wearables

First, I wanted to see how many times each of the users wore/used the FitBit tracker:

See Google Sheets for Results

Number of times logged data	Number of Users
31 times	21 users
30 times	3 users
29 times	2 users
28 times	1 users
26 times	2 users
20 times	1 user
19 times	1 user
18 times	1 user
4 times	1 user

64% of users logged data for the entire data time period (04-12-2016 to 05-12-2016). When you add in the users who only missed 1-3 days that percentage jumps up to 82% of users who logged data or wore their FitBit Tracker consistently over the month-long period.

Next, I wanted to breakdown the users by how much they wore their FitBit Fitness Tracker. I created three user types:

- **Active User -** wore their tracker for 25-31 days
- **Moderate User -** wore their tracker for 15-24 days
- **Light User** - wore their tracker for 0 to 14 days

(See Google Sheets for Results)

- **Active User** - 93.5%
- **Moderate User** - 6.1%
- **Light User** - 0.4%

5.c.ii User Data Summary
Next, I wanted to take a closer look at the MIN, MAX, & AVG of total steps, total distance, calories and activity levels by Id.

[(See Google Sheet for Results)]

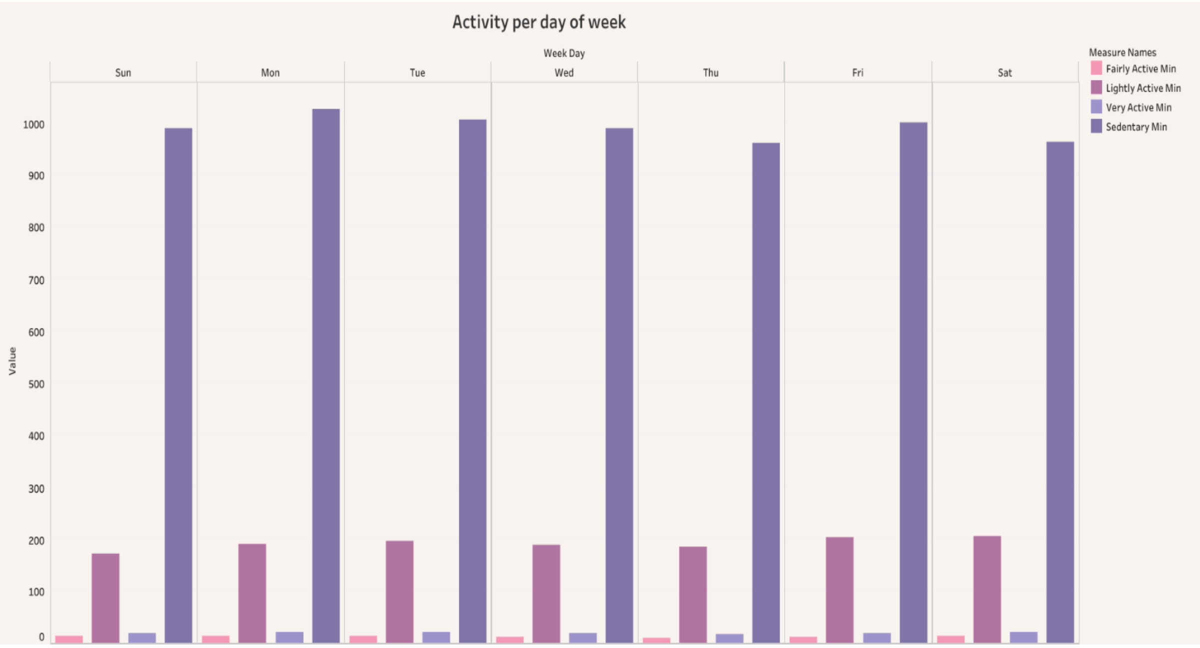
Next I wanted to narrow my results to just the averages of the different types of minutes by Id.

[(See Google Sheet for Results)]

These results showed that the average minutes of the Sedentary activity level was the highest for each distinct Id.

Lastly, I wanted to take a look at average active minutes by week day before moving on to user types.

(See Google Sheets Results)



Again, through this query we see that Sedentary Minutes are the highest type of active minutes. What is noticeable is that there is no real difference in type of active minute total by week day. It seems users are consistent in their active minute output each day. This information could show that Bellabeat could leverage activity goals for users to meet as users might already being trying to meet personal activity goals each day and Bellabeat could enourage higher activity goals to increase daily active minutes that are very active or fairly active.

5.d User Types by Total Steps

A Healthline.com article ("How many steps do I need a day?") written by Sara Lindberg in 2019 cited a 2011 study by Tudor-Locke et. al. titled "How many steps/day are enough? for adults" which found that 10,000 steps/day is a reasonable target for healthy adults. Lindberg (2019) breaks down activity level by steps into three categories based off the 2011 study by Tudor-Locke et. al.:

- **Inactive:** less than 5,000 steps per day
- **Average (somewhat active):** ranges from 7,500 to 9,999 steps per day
- **Very Active:** more than 12,500 steps per day

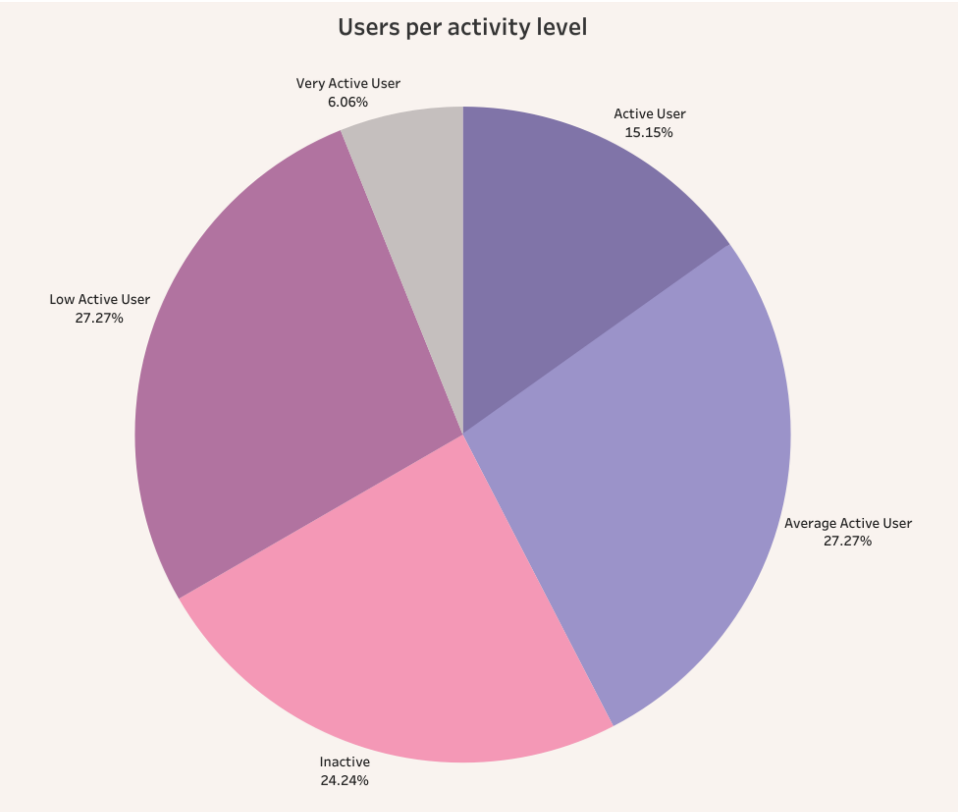
Sources: Healthline Article | 2011 Study

I will be using the above activity categories to create user types for the distinct Ids within the Daily Activity dataset. I'm interested to see how these categories may be broken down.

When creating my SQL query I realized that the above categories from the Healthline article left some gaps. So I created two other categories:

- **Low Active User:** 5,000 to 7,499 steps
- **Active User:** 10,000 to 12,499 steps

(See Google Sheets for Results)



Here are the Results:

- **Inactive User:** 8 users
- **Low Active User:** 9 users
- **Average Active User:** 9 users
- **Active User:** 5 users
- **Very Active User:** 2 users

If we break this down between Non & Low Active users and users who are considered 'Active' we can see that 17 users (52%) are little to non-active and 16 (48%) users are considered active. So almost a 50/50 split on type of users.

This split is pretty close to the results earlier of looking at active minutes. If we took out the summation of the Lightly Active Minutes from the query, we saw 17 users (52%) met CDC recommendations of 150 active minutes a week, 13 users (39%) did not meet CDC guidelines and 3 users (9%) did not have data from that week. After breaking down the sample into user types, I wanted to look at some more activity comparisons.

5.e Calories, Steps & Active Minutes by ID

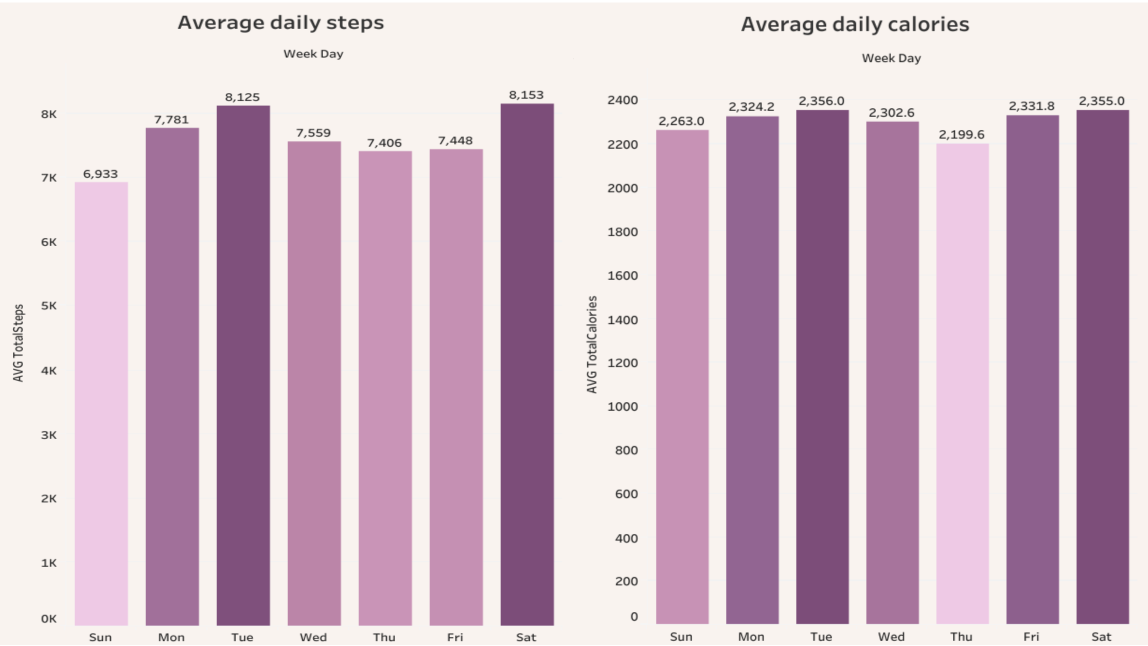
Now that we have more insights into our users' activity levels, I'm interested in seeing what their logged calories in relation to their steps and active minutes can tell us. This may be easier to see as a data visualization, but I'd still like to explore this in SQL as well for the practice.

[(See Google Sheets for Results)]

5.f Total Steps, Total Calories and Total Distance by Week Day

Next, I wanted to take a look at average steps, Calories and Distance by day to see if users were more active on certain days of the week.

([See Google Sheets for Results])



After running the query, there wasn't a whole lot of difference between each day in terms of average steps and Calories. With that said, Saturday had the highest average steps as well as the beginning of each week (Monday and Tuesday). We could potentially infer from this that the users wanted to be more active right after the weekend of rest (Sunday with lowest total steps & Friday not too far behind) & that Saturday allowed for more time for activity & movement. Mid-week (Tuesday and Wednesday) had the highest average calories as well as the Saturday.

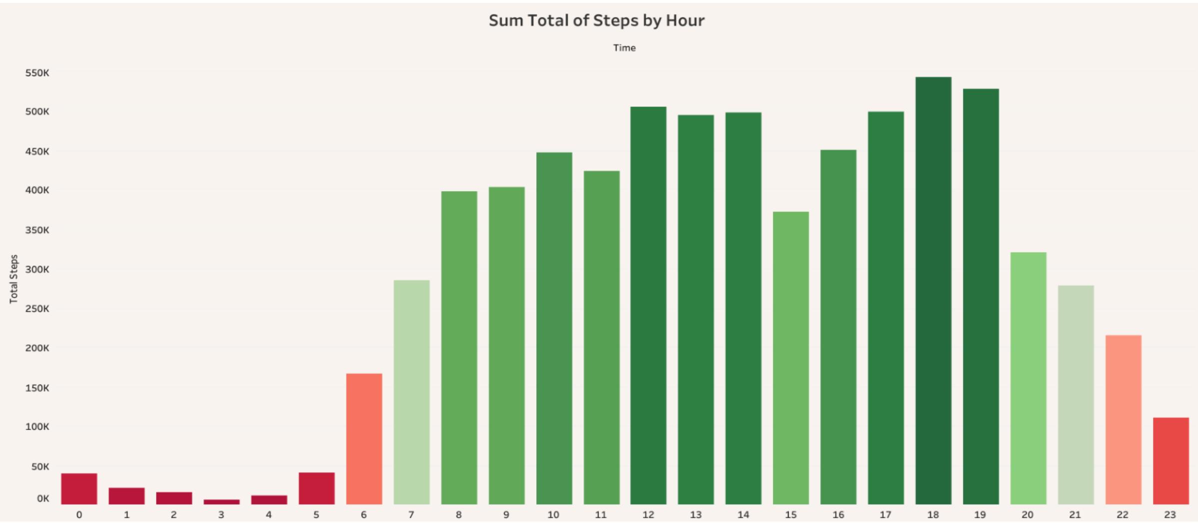
5.g Total Steps by Hour

Next, I wanted to take a look at Total Steps taking by Hour to see what time of day our users were most active.

[(see spreadsheet results)]

The top 5 hours of steps recorded were:

1. 18:00:00 (6pm) – 542,848 steps
2. 19:00:00 (7pm) – 528,552 steps
3. 12:00:00 (12pm) – 505,848 steps
4. 17:00:00 (5pm) – 498,511 steps
5. 14:00:00 (2pm) – 497,813 steps



5.h Deeper Look into Sleep

Then I wanted to explore the Sleep habits of the users and how it compares to activity level.

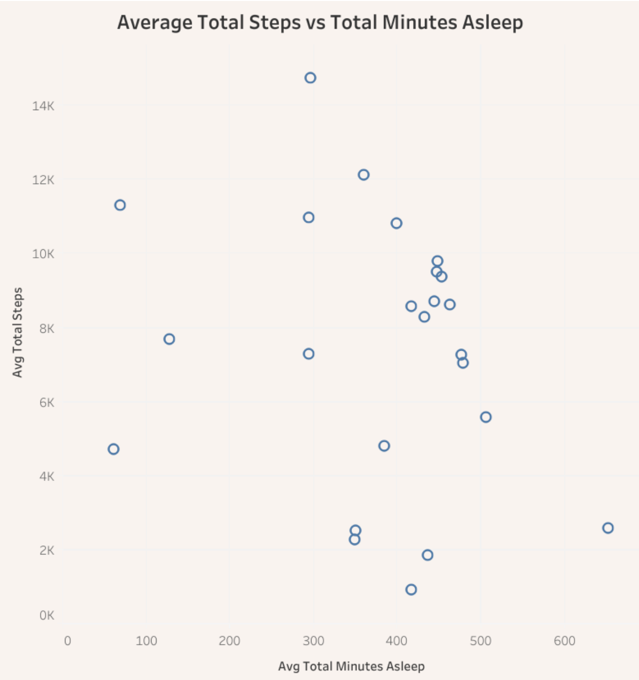
First, I looked at which date had the most minutes of sleep from all the users.

(See Google Sheets Results)

Next I looked at average minutes slept, total steps and calories by user Id.

(See Google Sheets for Results)

Then I visualized average total steps against average total minutes slept to see any type of correlation

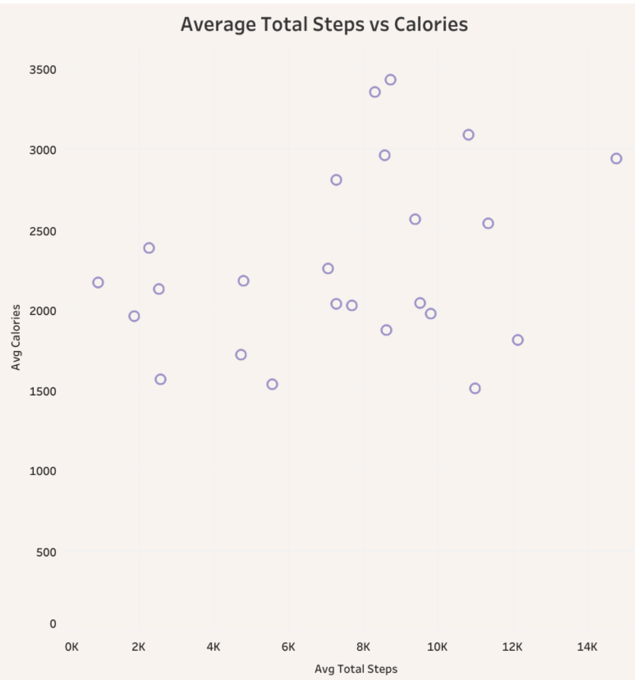


The graphs shows that most users who got at least 5 hrs. asleep had higher step counts. With this said, most users were not averaging the recommended 10,000 steps a day as noted in the Healthline article cited above.

5.i Average Total Steps vs Calories

look at average minutes slept, total steps and calories by user Id.

(See Google Sheets for Results)



There is **high correlation** between average total step and calories

6. ACT Phase

6.a Conclusion & Recommendations

Bellabeat's women-centric, holistic approach paired with smart insights and body positivity has led to the creation of wearable technology for women. These products empower women to utilize data to improve their overall health.

Since Bellabeat focuses strongly on a female audience for their products, I would recommend that the company look into using their own marketing and user data or conduct their own data collection to gain further insights and trends. I'd also recommend using a larger sample size if possible, in order to increase the confidence interval. Since the data utilized in this case study did not include demographic information, I'm unable to give a more detailed recommendation or ensure there was no sampling bias.

With that said, I was able to see some trends in the FitBit Fitness Tracker Data utilized in this case study.

Recommendations

App Notifications

- To ensure users are wearing their FitBits consistently, Bellabeat should consider sending a daily notification reminding them to charge their wearable device at the end of the day.
- More than half of the users are not reaching the recommended 10,000 steps per day. Bellabeat should send notifications throughout the day to remind users to "get up & move" or "complete their step goal for the day" to increase their overall step count.
- The data showed that users typically take the most steps during lunch hours or between 5 pm to 7 pm. Bellabeat should utilize this information to send personalized app notifications to recommend users start their walk or workout if they haven't begun their exercise routine at their usual time of increased steps. Moreover, they can promote workout routines, plans or classes tailored to these times of the day.
- Lastly, users who averaged five hours of sleep or more had a higher average step total. Bellabeat can help users increase their sleep time by sending a notification to "wind down for sleep" at a specific time based on their sleep habits, potentially helping them achieve better health outcomes.

Marketing

- The high level of activity among users, with 93.5% of them being consistent in utilizing their FitBits, suggests that Bellabeat should focus its marketing efforts on customers who are already invested in wellness and using fitness trackers. Highlighting the unique features of Bellabeat's products, especially those designed for women, may entice these customers to switch to Bellabeat's products for more targeted insights.
- Since users tend to be most active during lunchtime and in the late afternoon to early evening, Bellabeat should target customers who have set schedules, such as those with regular jobs or parents with a predictable daily routine. By focusing on these customer groups, Bellabeat can tailor its products and marketing messages to fit their needs and increase the likelihood of adoption and continued use of its products.

[Tableau Viz]