

Bellabeat Case Study

Author: Jessica E S. Bauer

Tools:

- Excel
- SQL with (SSMS)
- Tableau

Skills:

- Data Cleaning
- Data Analysis
- Data Visualization
- Research
- Presentation

Scenario:

Founded in 2013 by Urška Sršen, Bellabeat has grown rapidly and quickly positioned itself as a tech-driven wellness company for women. The company offers 5 different, beautifully designed products that inform and inspire healthy habits in women around the world. While Bellabeat has invested in traditional marketing strategies, their strongest focus is on digital marketing through the major players of Google Search Engine, Facebook, Instagram, Twitter, and YouTube. Bellabeat is a successful small company that believes business analysis of customer data will gain insights into new opportunities for growth. The recommendations from these insights will help guide the marketing strategy for the company.

The Phases

Ask:

Task: Analyze non-Bellabeat (FitBit) data to gain insights into customer trends to help guide the marketing strategy for Bellabeat’s future growth.

Key Stakeholders: Urška Sršen and Sando Mur (executive team); Bellabeat Marketing Analytics Team

Objective: Are people meeting recommended health and fitness goals?

Prepare:

Data Source: Third Party. These datasets were generated by respondents to a distributed survey via Amazon Mechanical Turk between April 2016 to May 2016.

Dataset: FitBit Fitness Tracker Data (CC0: Public Domain, dataset made available by Mobius <https://www.kaggle.com/datasets/arashnic/fitbit>). This Kaggle dataset contains a personal fitness tracker from 30 eligible FitBit users who 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’ behaviors and habits.

Credibility: (Using ROCCC Approach)

Reliability: This data is not reliable. There is no information about margin of error and the small sample size (30 participants) limits the amount of analysis that can be performed and challenges the confidence interval.

Original: The dataset was generated by respondents to a distributed survey via Amazon Mechanical Turk between April 12 and May 12, 2016. Thirty eligible Fitbit users consented to the submission of personal tracker data. Since data was collected by a third party, unsure of reliability or origin.

Comprehensive: This data is not comprehensive. No information is provided about the participants such as gender, age, nationality, health state, etc. This could indicate that the data was not collected at random. If there is bias in the data, it can skew the results and not provide a fair analysis. Data was collected at a granular level by minutes for physical activity, heart rate, weight, and sleep monitoring.

Current: This data is not current, having been collected between the dates of April 12 and May 12, 2016. It may not represent current trends in fitness tracker usage.

Cited: Amazon Mechanical Murk created the dataset, but we have no information whether they are a credible third-party source.

Process:

1. After downloading the CSV files from Kaggle, I opened them in Excel to get a feel for the data and do a little exploring. Deciding to do my analysis on weekly averages, I isolated out the 3 tables I felt would provide me with the best data for my objective:
 - a. dailyActivity_merged
 - b. sleepDay_merged
 - c. weightLogInfo_merged
2. Adjusted the format of the date columns to a format that SSMS would accept for importing files.
3. Imported the 3 tables into SSMS for EDA and cleaning. Tables were checked for common columns (id and date) and consistency in format of similar columns. The lengths of the ids were all consistent at either 11 or 12 digits.
4. Checking for distinct users, I found
 - a. dailyActivity had 33 unique users
 - b. sleepDaily had 24 unique users
 - c. weightLogInfo had 8 unique users
5. Odd that there were 33 unique users when there were supposedly only 30 participants. Supports questionable reliability of the data. There could be several reasons why the unique users are not consistent across all three tables, one being certain features may not have been used by all participants either by choice or due to the type of tracker they wore. Also, some data may have needed to be entered manually vs being tracked by the device
6. Checked tables for distinct rows to remove duplicates and nulls. Only dailyActivity has extra rows (3) which were removed.
7. Columns were renamed and integer datatypes were rounded to the 2nd decimal point for consistency across all tables.
8. Previewing each table showed all changes took and everything looked ready for further exploration.
9. Exported each table into an excel file from SSMS and uploaded them initially into Tableau.

Analyze:

Steps and Active Minutes being main components of tracking fitness levels, I did a few whisker plots of total steps, each of the 4 active minutes categories (sedentary, lightly, fairly, and very), and sleep.

Steps and Active Minutes all had one outlier on May 12, which may mean that data was not collected for the whole day and stopped partway through, thus giving a much lower value on that day. Sleep had no outliers.



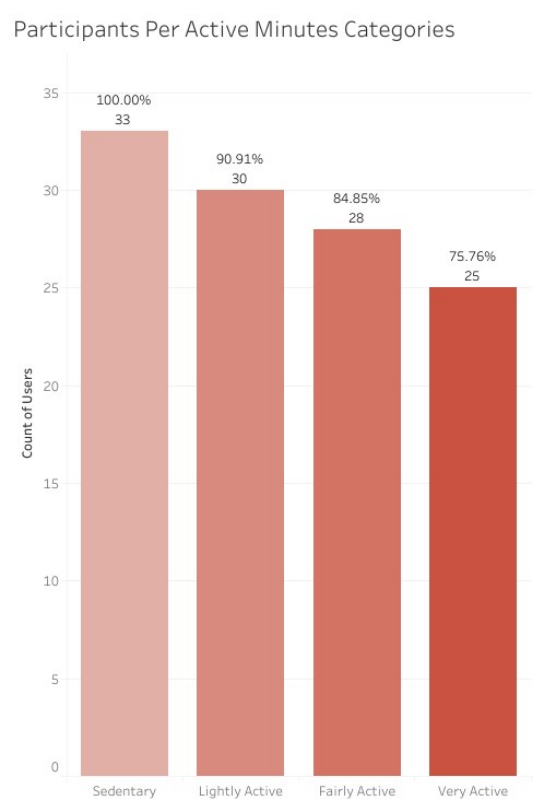
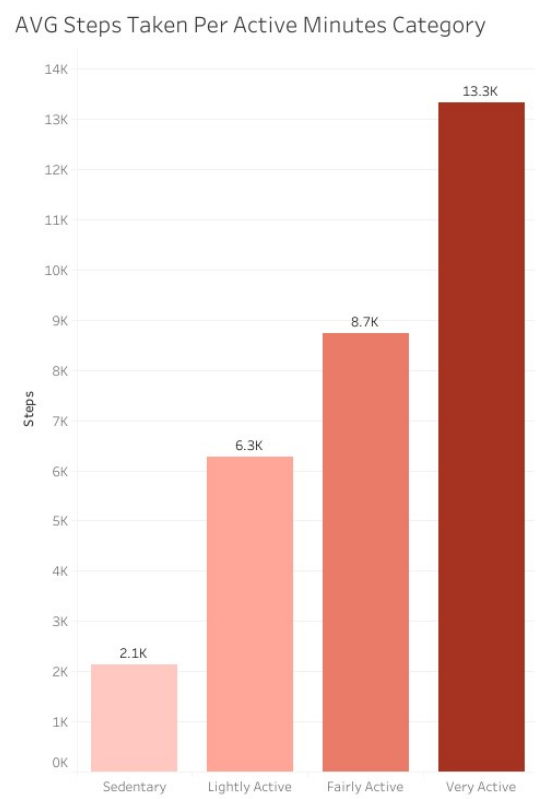
10,000 steps a day was a marketing ploy created in Japan in 1964 because the Manpo-Kei translated to 10,000-step-meter and was a catchy name for the Yamasa new pedometer. It was not based on any scientific evidence. Research studies found that between 7000 and 8000 steps a day is plenty for increasing longevity and decreasing health risks.^{1 2}

Our participants averaged 7.4K steps per day, falling within the 7000-8000 steps per day range. Only on Sunday did they fall just short of the 7000-step minimum.



According to most pedometers:³
Sedentary = > 5000 Steps
Lightly Active = 5000 to 7999 Steps
Fairly Active = 8000 to 9999 Steps
Very Active = <= 10000 Steps

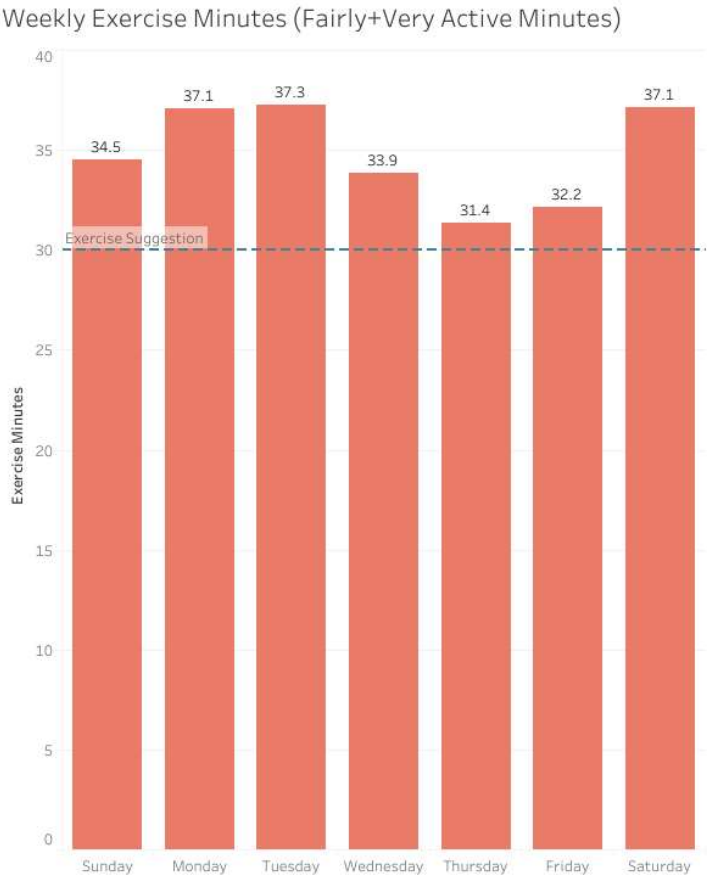
Most participants appeared to be quite active with over 75% of them hitting the very active level of steps. Sedentary, lightly, and fairly active categories fell below the median of steps taken for each category.



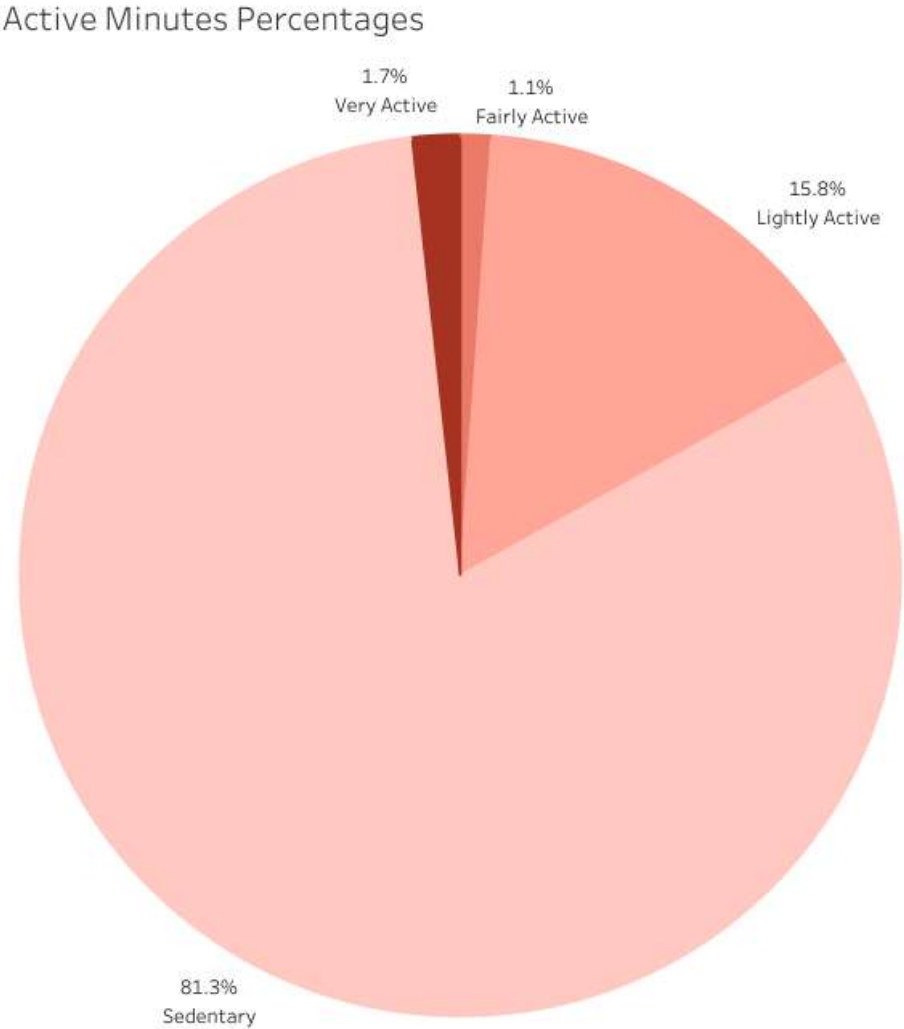
¹ <https://www.foodnetwork.com/healthyeats/fitness-and-wellness/how-many-steps-should-you-take-every-day>
² <https://www.insider.com/walking-10000-steps-for-health-not-science-based-in-marketing-2021-7>
³ https://www.medicinenet.com/how_many_steps_a_day_is_considered_active/article.htm

It is recommended that adults get at least 150 active minutes (fairly to very active) a week or at least 30 minutes 5 days a week.⁴

This group of participants was above average with over 30 minutes of exercise all 7 days of the week, even on Sunday, which has the lowest step count.



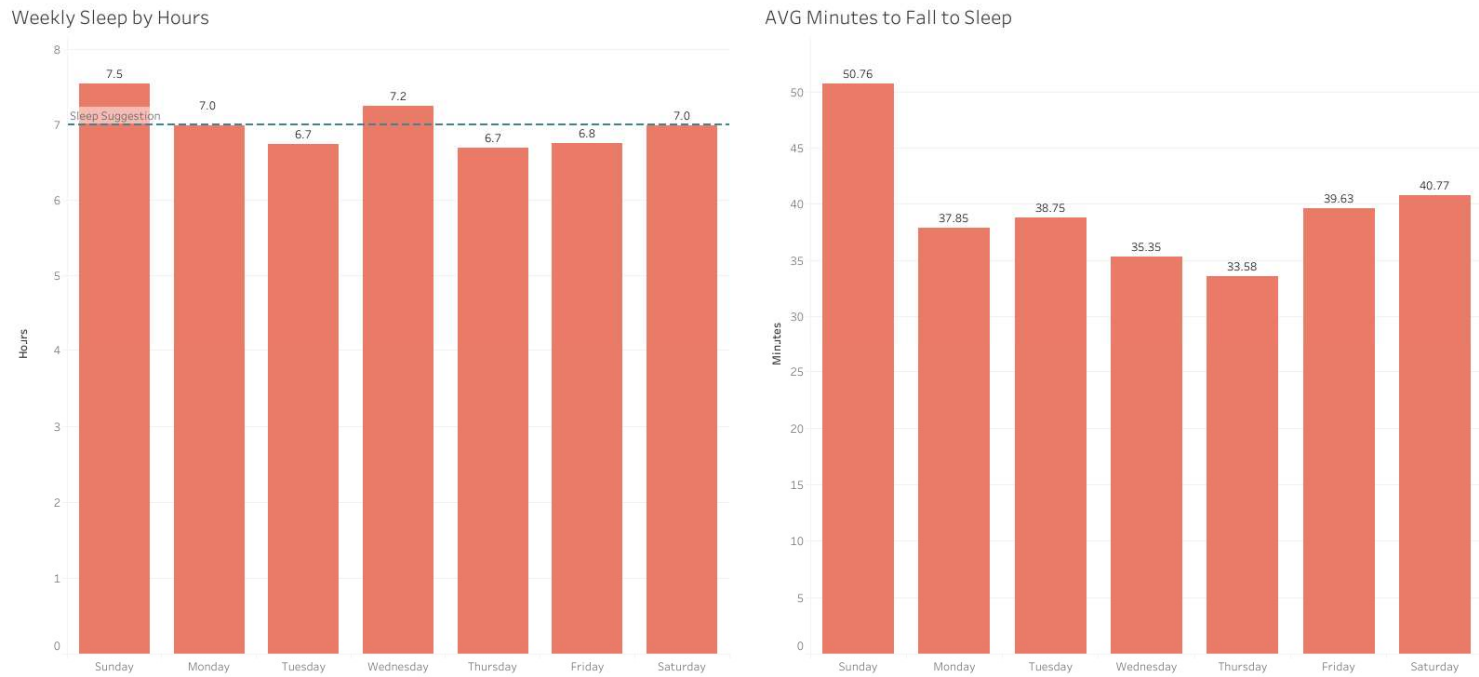
However, looking at the percentage of active minutes by category, over 81% fell under Sedentary, with only 3% within the Fairly and Very Active minutes categories.



⁴ <https://www.cdc.gov/physicalactivity/basics/adults/index.htm#>

Our participants slept around 7 hours on average per night. The CDC recommends for adults ages 18-60, 7 or more hours per night.⁵

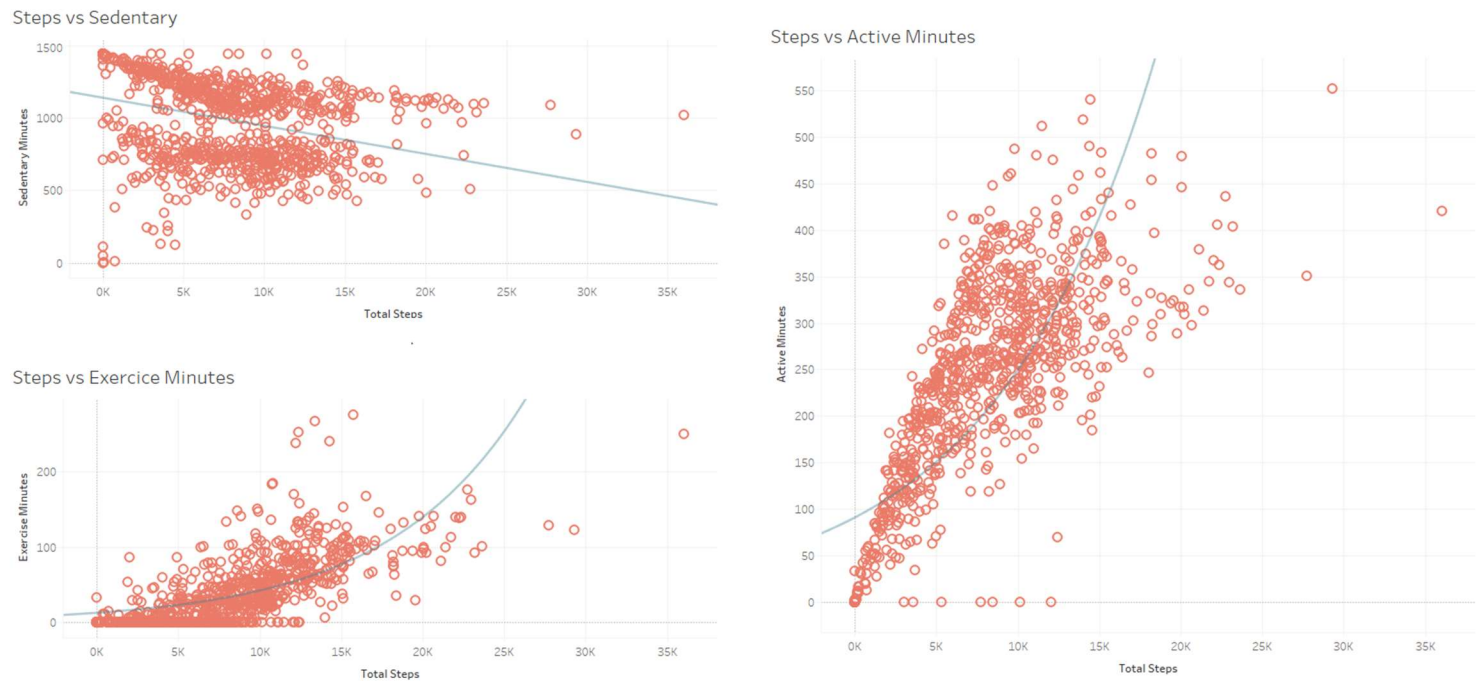
Sunday had the highest average of sleep hours, most likely due to the weekend and having time off. Average hours then dipped down at the beginning of the week only to spike on Wednesday, between the two lowest days Tuesday and Thursday. The hours then began to climb back up for the weekend.



TRENDS TO NOTE:

More steps taken equates to higher levels of activity.⁶

- Active Minutes = Lightly+Fairly+Very Active Minutes
- Exercise Minutes = Fairly+Very Active Minutes

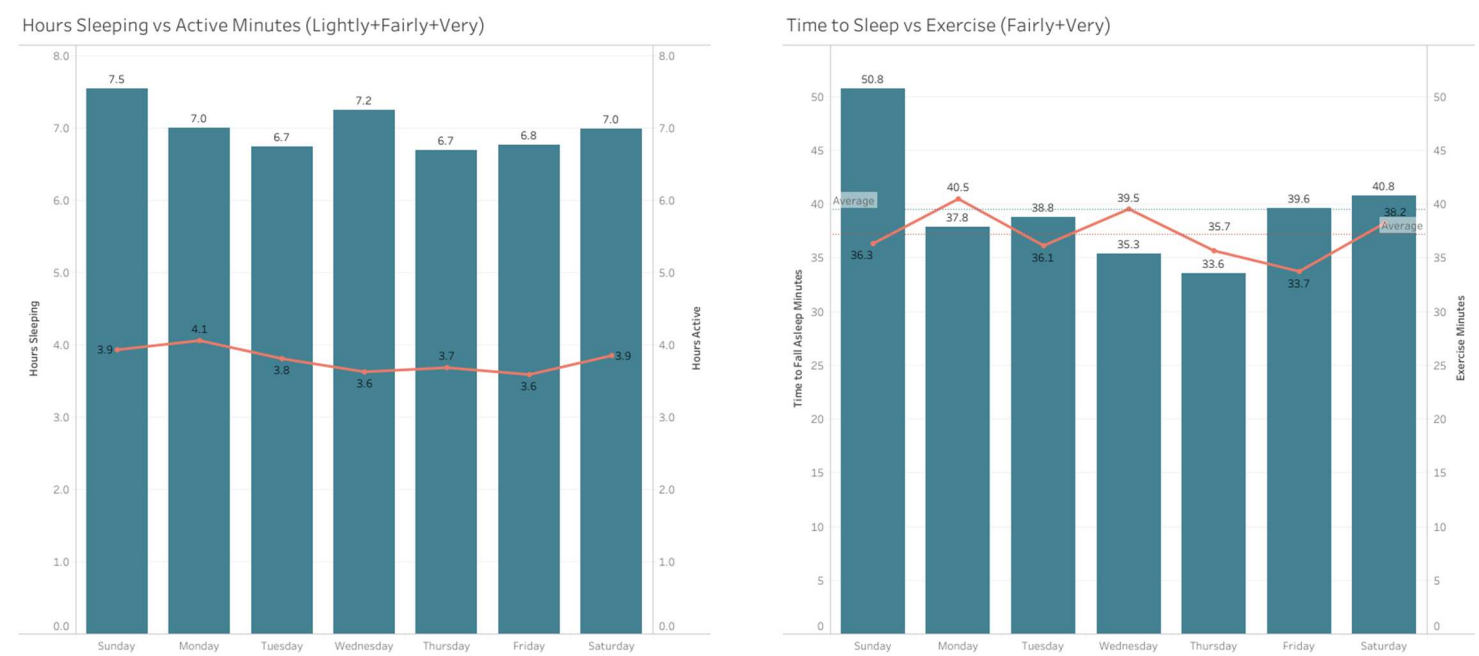


⁵ https://www.cdc.gov/sleep/about_sleep/how_much_sleep.html
⁶ <https://extension.sdstate.edu/light-moderate-and-vigorous-activity>

No surprise, the more steps taken or minutes active, the more calories burned. And the reverse can be seen comparing sedentary minutes vs calories in that the longer someone doesn't move, the less calories they burn.



Increased activity does not appear to directly relate to time spent sleeping nor does exercise appear to help people fall asleep faster. 3 of the 4 days where the average exercise minutes were below average, people did take longer to fall asleep, except Thursday, which had the second lowest time for exercise, but also the lowest time it took to fall asleep.



Share:

See Links Below

Act:

Conclusions:

Our participants hit all the recommendations for steps, activity and even sleep. They were within or above a scientific recommended range for steps for all but Sunday. Their step average per category fell below the median for each, so while active they aren't pushing the limits.

Exercise Minutes (Fairly+Very) averaged over 30 minutes every day of the week. However, those same Exercise Minutes accounted for less than 3% of total overall minutes, indicating that our participants likely performed a daily exercise and then were relatively sedentary the rest of the time. Over 81% of minutes accounted for little to no activity.

While the participants averaged 7 hours of sleep a night, 3 of the nights were below average, only 2 hit average and the remaining 2 days were above, with Sunday having the most hours at 7.5. The downward slope of exercise minutes on Wednesday with the spike in sleep time that day suggests that people were tired by the middle of the week and took it easy.

I did not end up using any data from the weights table based on the extremely small number of participants (8) that logged at least 1 weight, and only one of those eight logged each day, making the data very unreliable to base any fitness analysis on. However, this does indicate that people are not the best at manually logging stats and that an automated system would work better.

Recommendations:

Bellabeat Marketing Team should develop ways to encourage users to move more throughout the day. Exercising is great and people are hitting their recommended active minutes a week, but over 80% of the time they aren't moving enough to even be considered Lightly Active (ie household chores and daily tasks). Push notifications with suggestions and tips on how to move more such as taking the stairs instead of the elevator or taking a walk around the neighborhood after dinner could be beneficial in encouraging people to move more often. These suggestions should be sent throughout the day to remind people to get up every hour for a few minutes and move around.

Perhaps gamification can be applied to certain activities, allowing users to challenge and compete against friends and family and allow them to earn points or badges as positive reinforcement for adopting healthier moving habits.

Due to sleep just making the minimum recommended hours, I would suggest Bellabeat offering tips to creating healthy bedtime routines and suggestions to crafting a restive sleep environment to encourage less laying in bed and more sleeping.

Due to the extremely limited data available for weight, it may be best to have that be automatically acquired by the device rather than dependent on manually logging the data. Fitbit offers their own scale, Aria that connects to their app as well as the Withings Scale, who they partner with to share data. This is an area that could be explored so people could monitor their weight over time without having to remember to enter it. Stepping on a scale only takes a few seconds.

Changes start with a single step.