Bellabeat Fitness Tracker Data Analysis

Introduction

Bellabeat is a company headquartered in San Francisco, CA that makes wearable fitness trackers for women. The company's devices track data such as steps, heart rate, respiratory rate, sleep, and cardiac coherence, a measure of the synchronization of heart rate variability and breathing rate. These devices sync with the Bellabeat app, which then performs calculations on the data to give the user further insights.

Business Task

The purpose of this analysis is to find trends and gain insight into how consumers use non-Bellabeat smart fitness devices. Once the analysis is complete, I'm to determine how these trends apply to Bellabeat customers and determine how they could influence Bellabeat marketing strategy.

Data Source

The data used in this analysis is available on Kaggle. The download includes the following files:

- dailyActivity_merged.csv
- dailyCalories_merged.csv
- dailyIntensities_merged.csv
- dailySteps_merged.csv
- heartrate_seconds_merged.csv
- hourlyCalories_merged.csv
- hourlyIntensities_merged.csv
- hourlySteps_merged.csv
- minuteCaloriesNarrow_merged.csv
- minuteCaloriesWide_merged.csv
- minuteIntensitiesNarrow_merged.csv
- minuteIntensitiesWide_merged.csv
- minuteMETsNarrow_merged.csv
- minuteSleep_merged.csv
- minuteStepsNarrow_merged.csv
- minuteStepsWide_merged.csv
- sleepDay_merged.csv
- weightLogInfo_merged.csv

An explanation of the features in these files is located in the Fitabase_Data_Dictionary.pdf file.

These data were collected via a survey conducted on Amazon Mechanical Turk from March 12, 2016 to May 12, 2016. Thirty-three users consented to submit their personal tracker data for this time period.

Data Cleaning and Manipulation

I added each CSV file as a table to a DB file using DB Browser for SQLite so that I could run SQLite queries against it.

In this file, I used the following query to find that sleepDay_merged.csv file had three duplicates, which I deleted manually.

```
SELECT
    Id,
    SleepDay,
    count(DISTINCT SleepDay)
FROM sleepDay_merged
GROUP BY Id, SleepDay
HAVING count(*) > 1
```

The minuteSleep_merged.csv file had 543 duplicates, which I deleted with the following query.

```
DELETE FROM minuteSleep_merged
WHERE ROWID IN
    (SELECT min(ROWID)
    FROM minuteSleep_merged
    GROUP BY Id, date
    HAVING count(*) > 1)
```

Activity durations in the Fitbit data are broken down into four categories: sedentary, lightly active, fairly active, and very active. To simplify analysis, I combined these into two categories in Tableau: high activity and low activity, where high activity is the sum of fairly active and very active minutes and low activity is the sum of sedentary and lightly active minutes.

Summary of Analysis

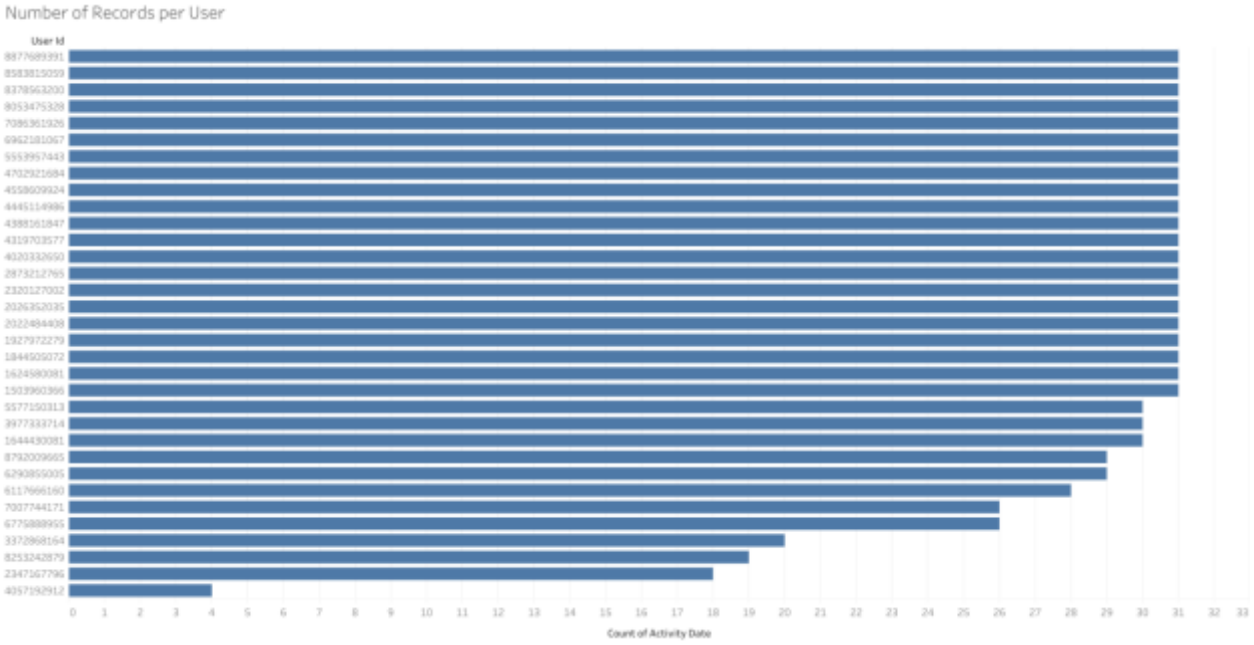
To see how Fitbit users were using their devices, I created several visualizations and performed several calculations. These visualizations showed a positive correlation between intensity levels and step counts, indicating that most users in this set engage in some type of running for their exercise.

Only eight users logged their weight. Of those users, five entered their weight manually and three had their weight entered automatically with a smart scale.

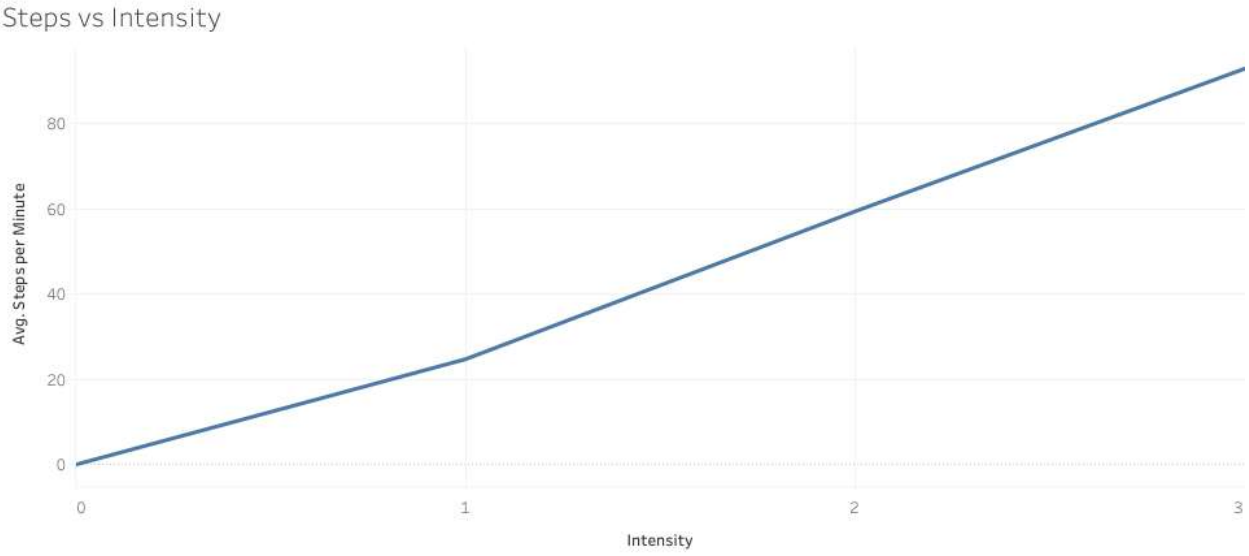
Twenty-four of the 33 total participants logged their sleep, indicating that the other nine did not wear their devices to bed. Of those 24, three wore their devices every night.

Key Findings

Not all users have data for every day during the collection period. These users most likely removed their devices on the days for which there is no data. The following chart shows a per-user breakdown of the number of days for which there are records.



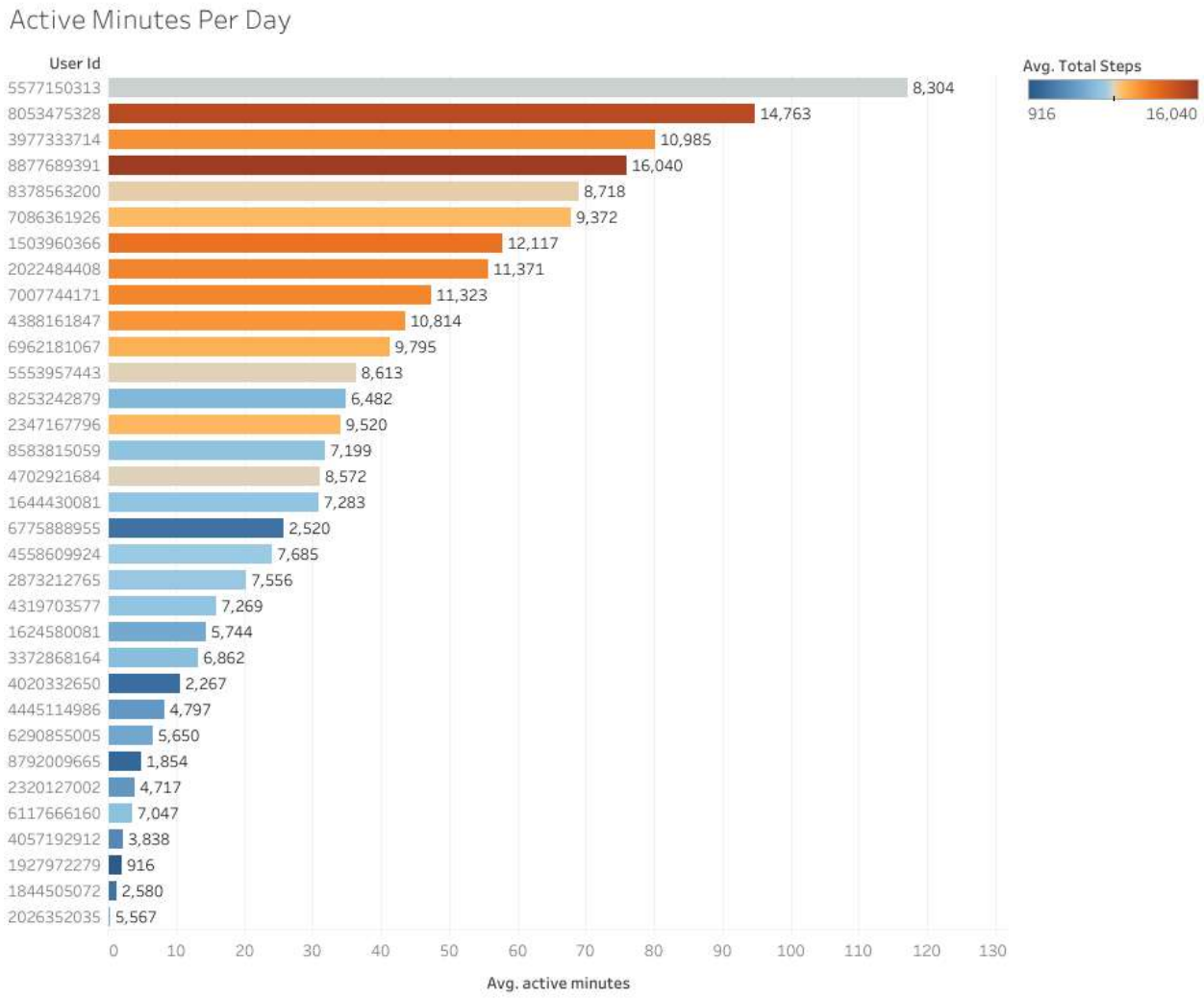
Here we see a positive correlation between the average number of steps taken per minute across all participants and the intensity level. Intensity is a value from zero to three where zero is sedentary, one is light activity, two is moderate activity, and three is very active.
This correlation suggests that users tend to run for their exercise.



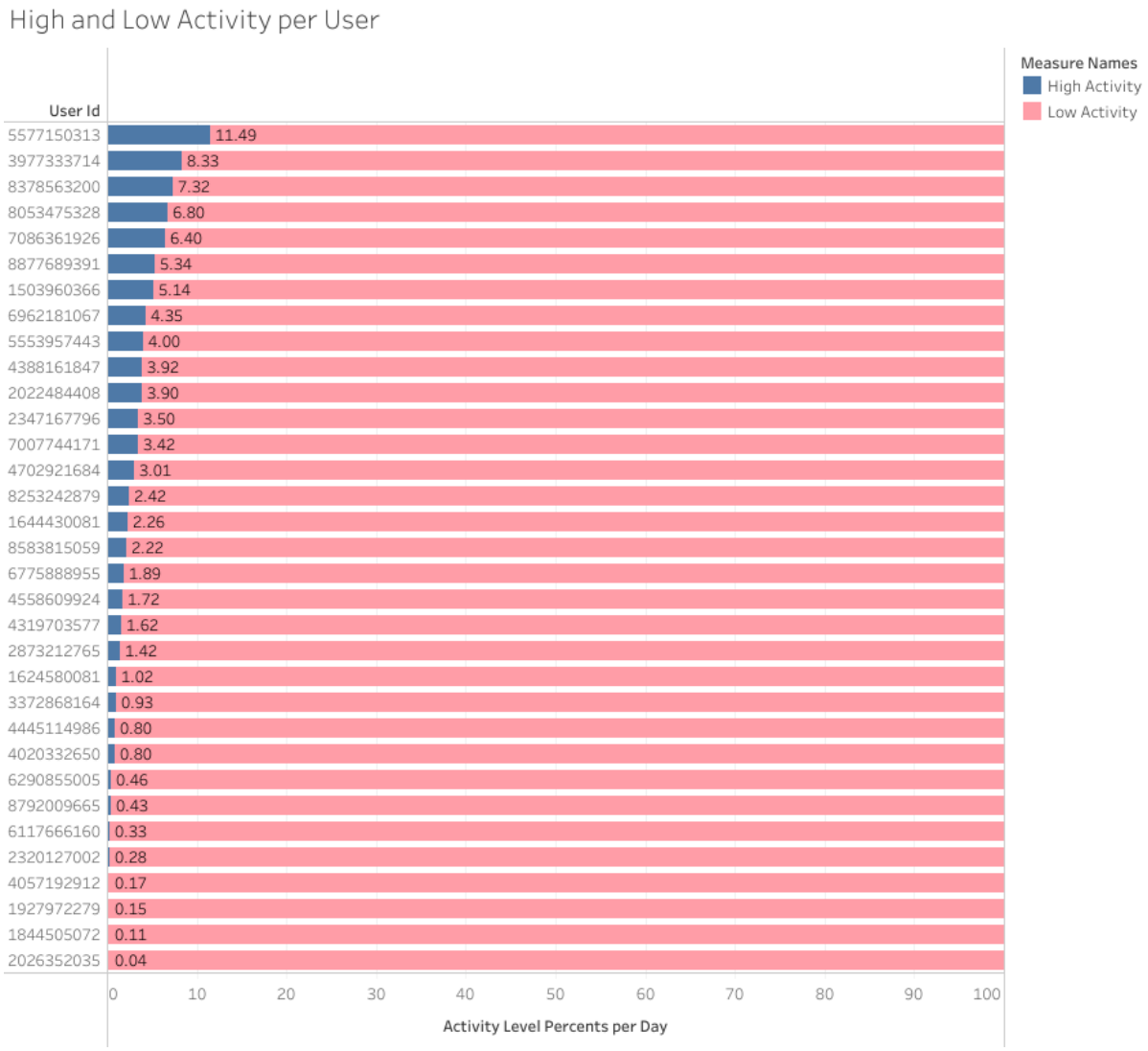
The following chart is a complement to the previous one. It shows the average active minutes per day for each user, with color coding that shows the average total steps per day for each user. From the color coding, we can see that the previous correlation still holds, but is not accurate for every user. In fact, the user with the most active minutes is middle-of-the road when it comes to

average steps per day. Such users were perhaps doing other forms of exercise, such as yoga or weightlifting.

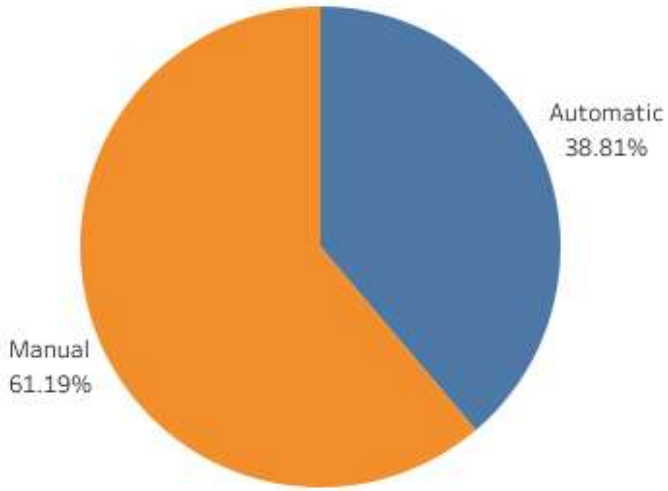
On average, users were active for 35 minutes a day and took 7,638 steps every day.



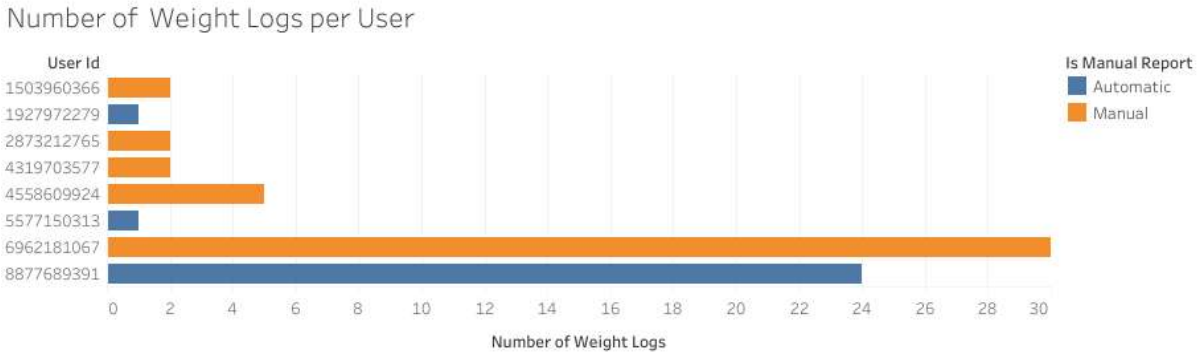
The next chart shows the average percent of the day that each user was engaged in high activity versus low activity. On average, each user spent 2.91% (about 41 minutes) of the day in high activity.



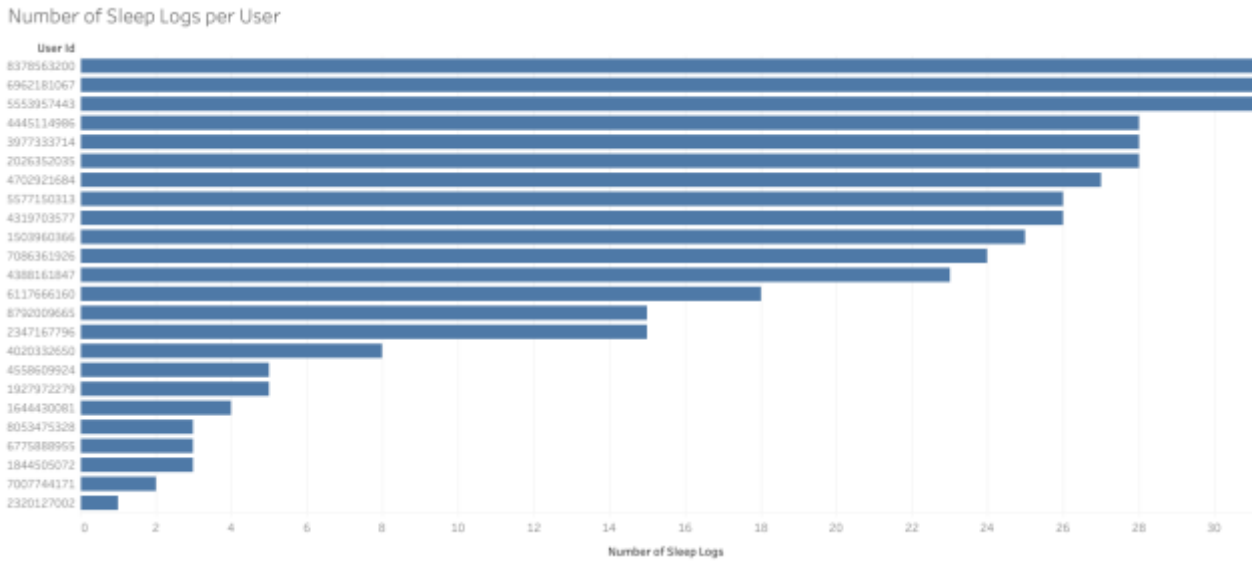
The below pie chart compares the users who entered their weight manually and those who had it entered automatically. Most users from this subset entered their weights manually, but a sizeable minority at approximately 40% used a smart scale.



Interestingly, the users with the ability to automatically log their weight tended to log less than those who entered their weight manually.



The Fitbit devices also track sleep. Out of the 33 participants, nine (27.3%) did not have any sleep records, indicating that they didn't wear their devices to bed. Of the 24 (72.7%) who did have sleep records, only three (12.5%) wore the device every night. On average, those users who wore their device to bed wore it for 17 out of the 31 nights (54.8%).



Recommendations

From this analysis, we see that one of the main problems was that users didn't wear their devices consistently and didn't fully use the capabilities of the device or the app. Bellabeat and Fitbit devices are fairly similar in their operation, so Bellabeat's initial focus should be to create different reminders in the form of push notifications to help users consistently wear the device. The notifications could remind users to do other things as well, such as log weight or water intake, or to exercise.

Since most users ran for exercise, Bellabeat can focus on showing users running in advertisements. To help future customers use their devices more consistently, ads can also show users sleeping with the device and logging different statistics in the app.

Appendix

The SQLite queries used in this analysis are located in the fitbit_sqlite.sqbpro file.