

# Bellabeat Case Study

Michael McCarthy

This is a capstone analysis project for the Google Data Analytics Course on Coursera. The following document details the analysis of Fitbit data using Excel to make business recommendations for a new and growing fitness technology company called Bellabeat. The Excel sheet and accompanying dashboard are available as well.

## Contents

1. Summary .....	2
2. Ask.....	2
2.1 Business Task .....	2
2.2 Initial Questions .....	2
3. Prepare.....	3
<b>3.1 Dataset Information</b> .....	3
3.2 Dataset Organization .....	3
3.3 Accessibility and Privacy of Data.....	4
3.4 Dataset Creation .....	4
3.5 Dataset Reliability and Credibility .....	4
4. Process .....	5
4.1 Data Cleaning and Formatting .....	5
4.2 Duplicates.....	5
4.3 Nulls .....	5
4.4 Rename Columns .....	5
4.5 Join Tables using Power Query .....	5
5 Analyze and Share .....	6
5.1 Activity Level - Steps .....	6
5.2 Activity Timing – Steps by Hour .....	7
5.3 Sleep and Activity Connections.....	7
5.3.1 Sleep and Steps .....	7
5.3.2 Sleep and Sedentary Minutes .....	8
5.3.3 Activity Intensity and Sleep.....	8
5.4 Daily Usage.....	9
6. Act .....	10

# 1. Summary

Bellabeat is a high-tech manufacturer of health-focused products for women. They are small but have potential to become a major player in the global smart device market. This analysis is designed to discover trends that will unlock growth opportunities for the company.

Bellabeat offers wearable smart devices as a watch, water bottle, and bracelet, necklace, or clip that track activity, sleep, stress, and water intake. All devices connect to their Bellabeat app which provides users with the data collected, menstrual cycle, and recommendations based on user data.

This analysis will be done in Excel and focus on how trends and connections in the activity and wellness data can help advertise to consumer needs and possibly retool their app product to better meet consumer demands.

## 2. Ask

### 2.1 Business Task

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

Stakeholders

- Urška Sršen - *Co founder and Chief Creative Officer. Looking for growth opportunities through analysis of fitness data*
- Sando Mur - *Mathematician and Co-founder. Executive team.*
- Bellabeat Marketing Analytics Team

### 2.2 Initial Questions

1. Most popular day for logged activities?
2. How do general sleep patterns compare to medically recommended patterns and amounts?
3. What time of day are users exercising or most active? How can Bellabeat help possible issues.
4. How often are features used? Logging activities, tracking sleep, water?
5. Length and intensity of activities?
6. How many users track sleep?
7. Sedentary data? Are most users working at a desk, manual labor, etc.?
8. Would sleep suggestions / reminders be useful? Lots of time in bed but not sleeping? Bad sleep habits?
9. Activity amounts and intensity compared to recommended amounts/type of exercise?
10. Could dietary goals be included? Meal recommendations? Is there any data to show a use or need for this service?
11. Stand walk reminders if not already included. Long periods of sedentary?
12. Are there any links between activity and sleep?
13. Is all data present? Enough to analyze for all categories and devices?

# 3. Prepare

## 3.1 Dataset Information

The data source used is Fitbit Fitness Tracker data. This dataset is stored in Kaggle and made available through Mobius.

## 3.2 Dataset Organization

The data is split up into 18 CSV documents. Each document either holds a different type of data collected or it holds data collected using different time frames. For example, there is a Daily Calories, Hourly Calories, and Minute Calories.

Files being used are:

- DailyActivity\_merged
- hourlySteps\_merged
- sleepDay\_merged

Data File Name	Description
dailyActivity_merged	CSV - Daily Activity over 31 days of 33 users. Tracking daily: Steps, Distance, Intensities, Calories
dailyCalories_merged	CSV - Daily Calories over 31 days of 33 users
dailyIntensities_merged	CSV - Daily Intensity over 31 days of 33 users. Measured in Minutes and Distance, dividing groups in 4 categories: Sedentary, Lightly Active, Fairly Active, Very Active
dailySteps_merged	CSV - Daily Steps over 31 days of 33 users
heartrate_seconds_merged	CSV - Exact day and time heartrate logs for just 7 users
hourlyCalories_merged	CSV - Hourly Calories burned over 31 days of 33 users
hourlyIntensities_merged	CSV - Hourly total and average intensity over 31 days of 33 users
hourlySteps_merged	CSV - Hourly Steps over 31 days of 33 users
minuteCaloriesNarrow_merged	CSV - Calories burned every minute over 31 days of 33 users (Every minute in single row)
minuteCaloriesWide_merged	CSV - Calories burned every minute over 31 days of 33 users (Every minute in single column)
minuteIntensitiesNarrow_merged	CSV - Intensity counted by minute over 31 days of 33 users (Every minute in single row)
minuteIntensitiesWide_merged	CSV - Intensity counted by minute over 31 days of 33 users (Every minute in single column)
minuteMETsNarrow_merged	CSV - Ratio of the energy you are using in a physical activity compared to the energy you would use at rest. Counted in minutes
minuteSleep_merged	CSV - Log Sleep by Minute for 24 users over 31 days. Value column not specified
minuteStepsNarrow_merged	CSV - Steps tracked every minute over 31 days of 33 users (Every minute in single row)
minuteStepsWide_merged	CSV - Steps tracked every minute over 31 days of 33 users (Every minute in single row)
sleepDay_merged	CSV - Daily sleep logs, tracked by: Total count of sleeps a day, Total minutes, Total Time in Bed
weightLogInfo_merged	CSV - Weight track by day in Kg and Pounds over 30 days. Calculation of BMI.5 users report weight manually 3 users not. In total there are 8 users

### 3.3 Accessibility and Privacy of Data

This data is open source as confirmed by the metadata. The work has been designated to the public domain by the owner through waiving all rights to the work under copyright law, including all relating neighboring rights, to the extent allowed by law. The data and work can be copied, modified, distributed, and used for commercial purposes without asking permission.

### 3.4 Dataset Creation

These datasets were generated by respondents to a survey distributed by Amazon Mechanical Turk between 03/12/2016 - 05/12/2016. Thirty eligible Fitbit users consented to the submission of personal tracker data. Variation in user data represents difference in devices or usage preferences.

### 3.5 Dataset Reliability and Credibility

The dataset is limited by its small size (30 users) and could be subject to sampling bias for the same reason. There is no certainty that this dataset is representative of the whole population. At the time of this analysis, the data (2016) is not current (2022) and habits or trends could have changed. The data's collection period is also very short which does not allow for a complete analysis due to differences in behaviors throughout the year.

# 4. Process

## 4.1 Data Cleaning and Formatting

Dates and timestamps are inconsistent across datasets. All date and timestamp records will be separate date and timestamp into separate columns.

- sleepDay\_merged, remove timestamp from SleepDay
- hourlySteps\_merged, separate timestamp and date. Date and Time will have new columns.
- Check – all dates are in format of m/d/y

## 4.2 Duplicates

Check data for duplicates and remove any that exist.

- 3 duplicates found in sleepDay and removed

## 4.3 Nulls

Check data for Nulls and resolve any that exist.

- No Nulls exist.

## 4.4 Rename Columns

- DailyActivity – added space between words in column names.
- DailyActivity – ActivityDay -> Activity Date
- SleepDay – added space between words in column names.
- SleepDay – SleepDay -> Sleep Date
- HourlySteps – Added space for all column names

## 4.5 Join Tables using Power Query

DailyActivity table and SleepDay table were joined using Power Query using an Inner Join. This was done to connect sleep data to activity data and find connections between activity levels and amount of sleep.

# 5 Analyze and Share

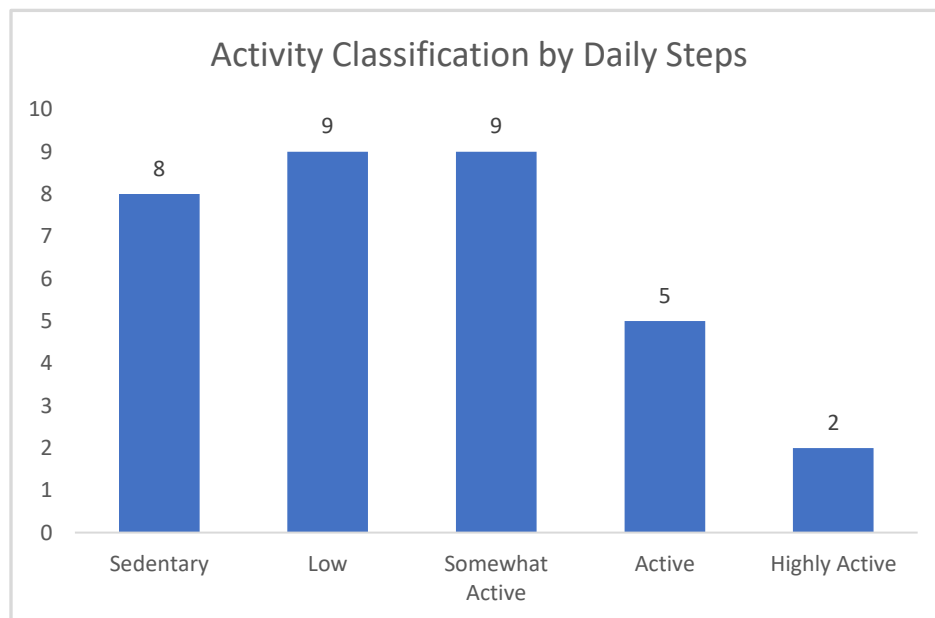
## 5.1 Activity Level - Steps

As an overview analysis of general activity levels, we will look at user daily steps to estimate their lifestyle activity level. This can be classified using information from <https://www.10000steps.org.au/articles/healthy-lifestyles/counting-steps/>.

- Sedentary - Less than 5,000 steps / day
- Low – 5,000 – 7,499 steps/day
- Somewhat Active – 7,500 - 9,999 steps / day
- Active - 10,000 - 12,499 steps / day
- Highly Active - 12,500+ steps / day

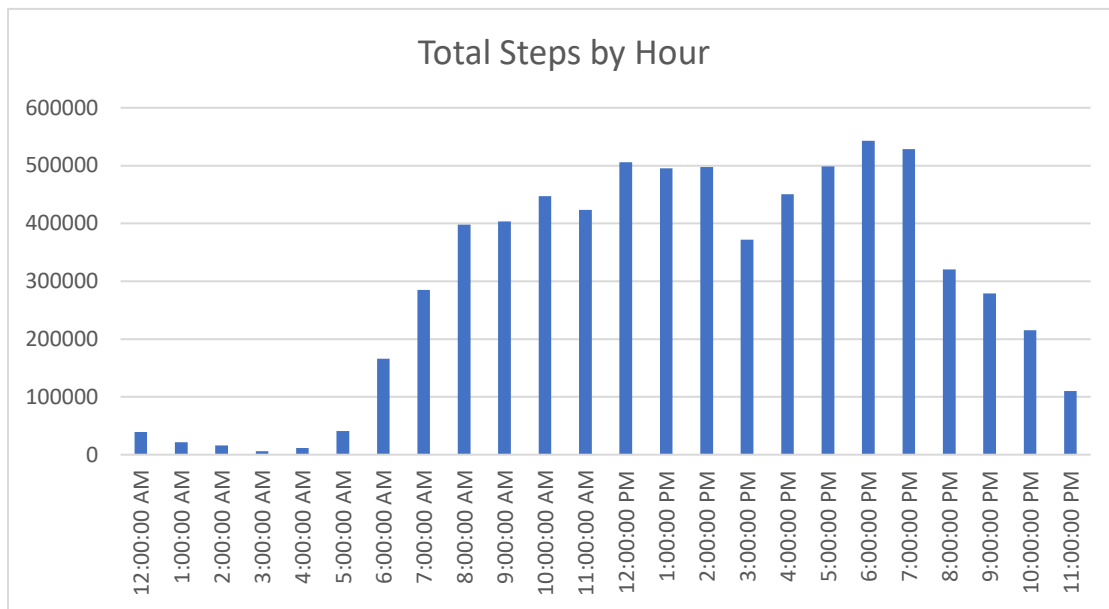
Most people are either Sedentary (25%) or fall into the range of being moderately active (53%). Device reminders to get up and welk, or even schedule walks using the device, would be beneficial to increase basic activity levels.

Could provide data amongst all users in certain age brackets on their percentile score on steps/day. Example: 27 years old and 9,000 steps/day could fall at the 67<sup>th</sup> percentile. Not bad, but not great and could improve. This would incorporate a level of competitiveness if desired by the user.



## 5.2 Activity Timing – Steps by Hour

Using data of the number of steps taken per hour throughout the day, a pivot table was created to allow aggregation of steps by hour. Below is the chart created.



There is a large dip in steps between 2pm – 4pm. Possible reasons:

- Afternoon crash: reminders to drink water (could be tied to water bottle product), go outside, get up and move.
  - User could decide what type of reminders they want.
- Busiest time at work (9-5 schedule): reminders to breathe, reset, meditate, mindfulness moment.

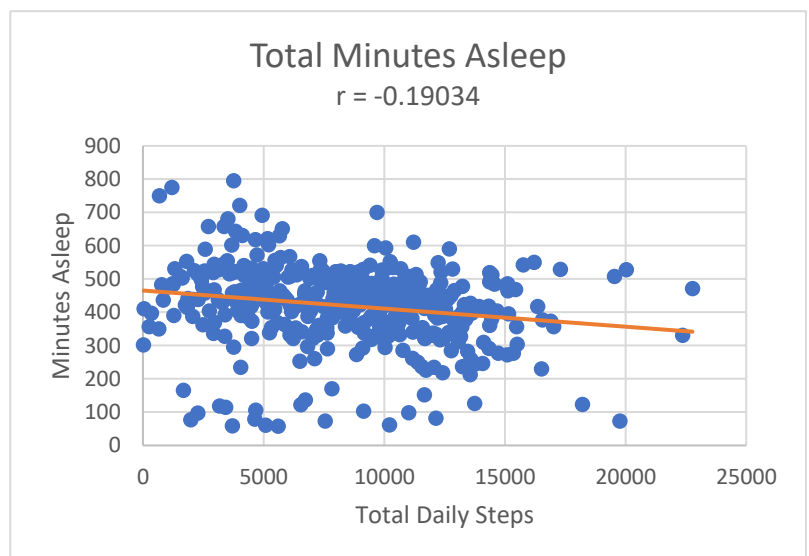
## 5.3 Sleep and Activity Connections

Using Power Query, I combined the DailyActivity and SleepDay tables with an Inner Join.

### 5.3.1 Sleep and Steps

The relationship between Steps and Sleep is shown graphically below.

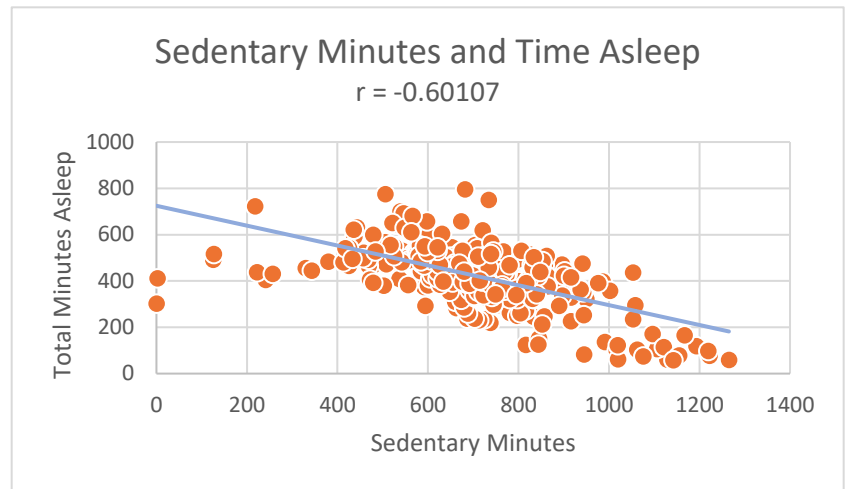
This graph with the included correlation coefficient shows that there is not enough evidence to show a connection between Time Asleep and Daily Steps taken.



### 5.3.2 Sleep and Sedentary Minutes

There is a connection, albeit moderate, between the amount of sleep users are getting and the amount of the day they are sedentary. Less sleep correlates with more sedentary time.

This unhealthy trend could be detected by the user's device and account. Reminders could be sent to the user on how to fix these issues and reasons why it is important to move and sleep more.

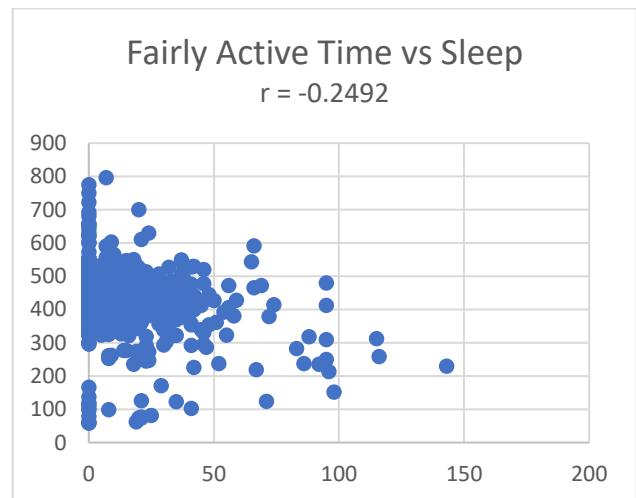
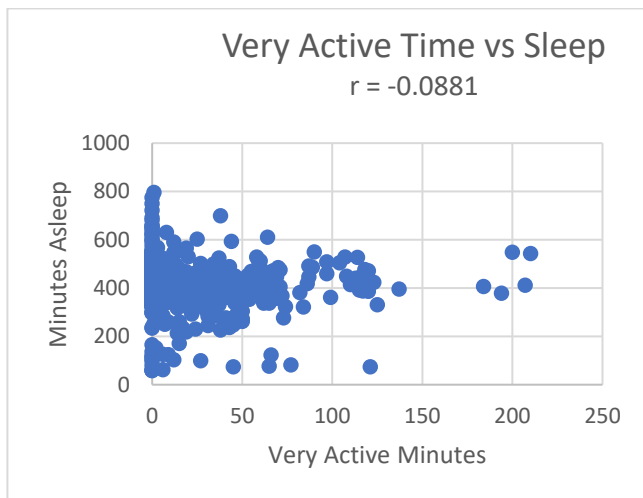


A similar analysis could be done for users on a rolling average basis. This would show users how trends in their sleep effects trends in their activity which in turn effects their overall health. The app could offer sleep reminders along with resources to either help fall asleep or track types of sleep under premium subscriptions similar to other popular sleep apps.

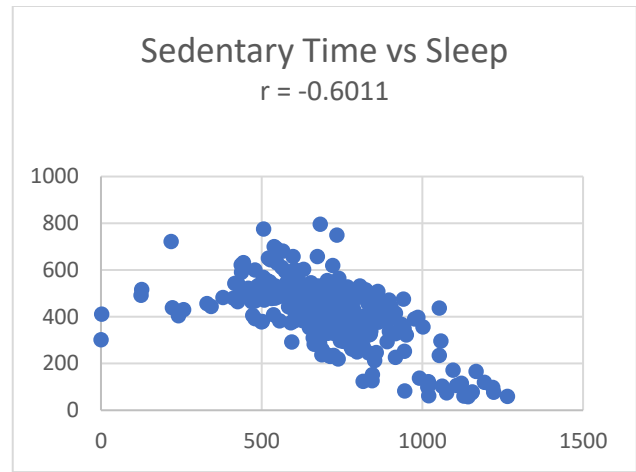
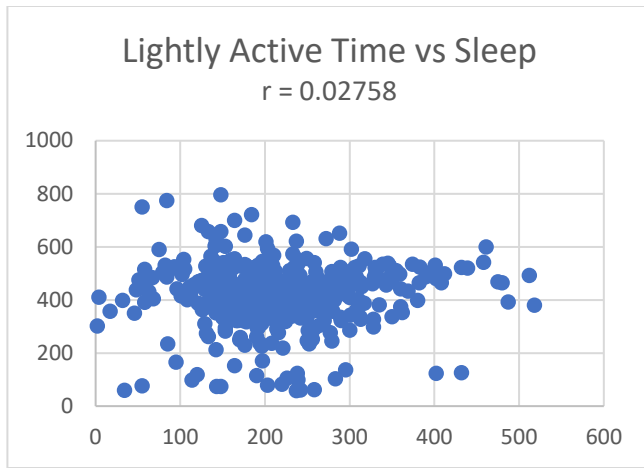
### 5.3.3 Activity Intensity and Sleep

When looking for a connection between sleep and amount of activity at each level, almost no correlation was found. The exception is Sleep and Sedentary Time. This can be seen in the graph and r-value of -0.6011. While not a very strong correlation, it is enough to suggest a relationship.

This finding can be used for app recommendations or reminders when users are either sedentary too often or not sleeping enough.







## 5.4 Daily Usage

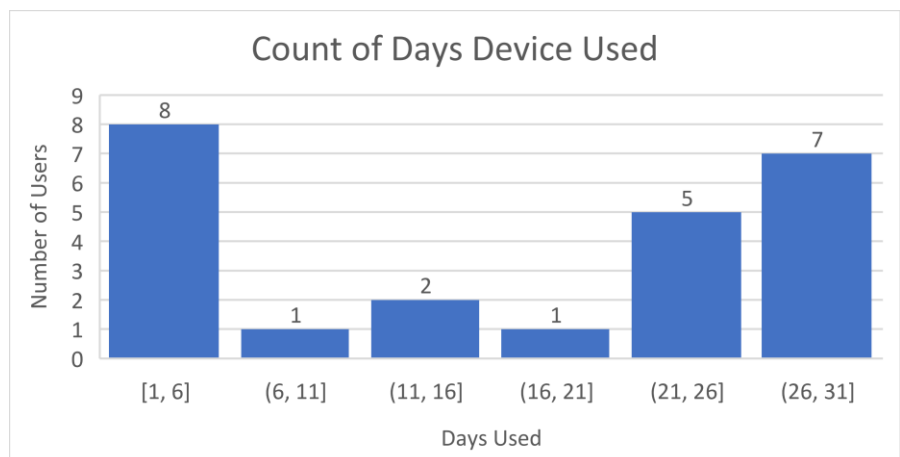
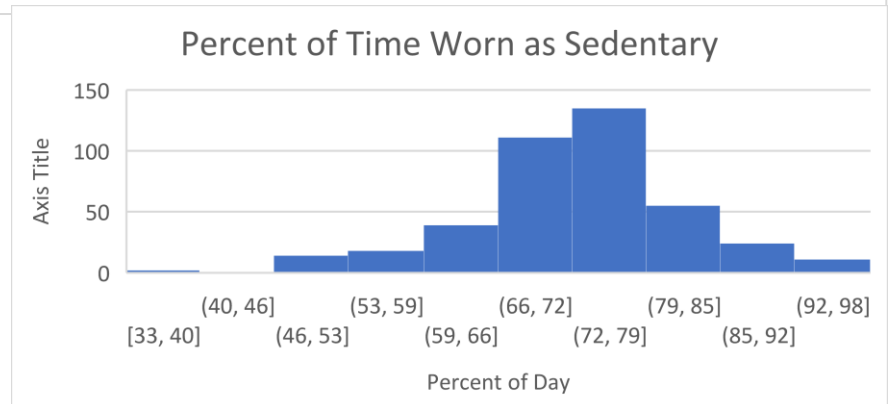
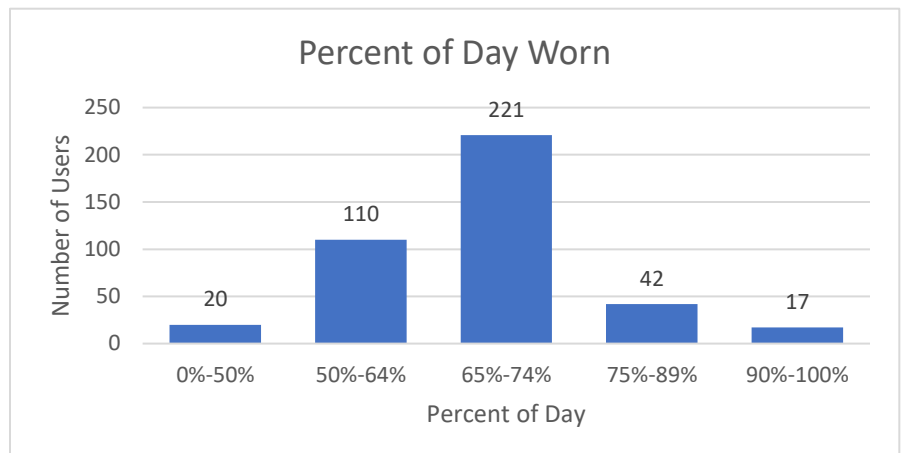
Users tend to wear their devices around 60% - 70% of the day while around 66%-80% of the time worn is spent being sedentary. Within the dataset, users either wore their device for the majority of days

Users should try to be active for a larger portion of their day. Studies show the health benefits of staying active.

[Light Activity Benefits](#)

[Staying Mobile](#)

[General Activity and Exercise](#)



## 6. Act

Bellabeat's mission is to support women to live healthier lives and take control of their health data. The data collected and used is small in scale so further collection and analysis would need to be done for more concrete results.

Users would benefit from these analyses being reproduced on their personalized data as the main target is young and adult women. With larger amounts of data and more specific data being collected on the target audience, new and unique insights could be found for a more specific and targeted marketing strategy.

Below are the recommendations for the Bellabeat app.

Recommendation	Description
1. Daily Notifications and reminders on app and wearable devices to move and walk more frequently	<p>Users were classified into 5 categories based on daily steps. Analysis showed 79% of users do not walk enough to get general health benefits from walking.</p> <p>Based on previous data (when user gets most steps in per day) the app would send recommendations and notifications when users are not on track to meet daily step goal.</p>
2. Sleep monitor with data analysis	<p>Data showed a near strong negative correlation (-0.60107) between time asleep and time spent sedentary. This shows more time asleep means less time sedentary which has health benefits.</p> <p>Users could set sleep schedules and receive reminders or instructions on how to better their sleep (meditation, breathing techniques, device shut off reminders). Some of these sleep features could be wrapped into a premium subscription.</p>
3. Incentives to wear device and stay active	<p>One third of users only wore their device between one and six days. Without data it would be hard to provide recommendations for the user's health. In app incentives, points, could be given for every day the device was worn for 75% of the day or more.</p> <p>Points could also be given for achieving step goals or ranking on an overall or age specific leaderboard for step count. These points could all be redeemed for Bellabeat merchandise or products that would help in brand advertising.</p>
4. Device "wearability"	<p>Users wore the device for an average of 67.5% of the day and only 29% of users wore their device more than 83% of the days recorded.</p> <p>The design of the device should be continuously improved to make sure the battery is long lasting, device can be worn for all occasions, water/damage resistant, and comfort.</p>