Analysis Report: Bellabeat Case Study

Date Written: 8/25/23 Link to Original Code [Here]

1. Introduction

1.1 Summary

This project is a capstone case study for the Google Data Analytics Certificate. For this case study we seek to improve the Bellabeat app using insight drawn from Fitbit data, a competitor product with similar customer base.

Bellabeat is a high tech company that creates health related products with the purpose of improving women's health. Created in 2013, Bellabeat grew into a tech driven wellness company with a global consumer base. Their products and services emphasize the use of smart devices to monitor the user's activities, sleep, and stress.

The Bellabeat app provides users with health data related to their activity, sleep, stress, menstrual cycle, and mindfulness habits. This data can help users better understand their current habits and make healthy decisions. The Bellabeat app connects to their line of smart wellness products.

Our analysis found several new insights on the Fitbit user base which could be applied to Bellabeat's customer base. Using users' lifestyle and exercise habits as a basis, the Bellabeat app could give users recommendations tailored to the time of day. This will increase current user's reliance on the Bellabeat app, encouraging users to continue using the app.

1.2 Questions

1. What is the user base average activity level over the course of 2 months?
2. What hours are users most active on average?
3. Which weekdays are the users on average most and least active over the course of the 2 months?
4. On average, how long do users spend in bed?
5. Is there a relationship between how long users sleep and their daily workout intensity?

1.3 Outline

Section 2 will discuss the background of the dataset, dataset origin, limitations and cleaning process applied. Section 3 contains our analysis of the question mentioned in Section 1.2. Each subsection of Section 3 discusses the methodology, analysis and conclusion of a single question. Section 4 contains a summary of our new insights, recommendations and new questions to pursue.

2. Data

2.1 Dataset Used

The dataset, FitBit Fitness Tracker Data, is provided by Mobius on the Kaggle platform. The data was obtained via survey by Amazon Mechanical Turk, a third party data source. The data consist of daily, hourly and minute information on the user's daily activity, calories burned, sleep, and etc. For this case study, we will focus on the daily and hourly data because they will provide more insight than the minute data. The dataset can be found here[1]. Information regarding the dataset's columns can be found here[2], provided by Fitbit.

2.2 Limitations

The survey's sample size consists of only 33 users, so the data is not an accurate representation of the customer base. In addition, Fitbit product is marketed to people of all genders as opposed to Bellabeat's focus on women. Although they have the same focus on promoting healthier lifestyle, the data contains a mixture of males and females. This may cause discrepancy between expectation and reality. This data was collected from the time period 3/12/2016 to 5/12/2016 so the information is outdated.

2.3 Licensing & Privacy

This dataset is licensed under CC0: Public Domain thus freely available online for public use. This dataset may be copied, modified and distributed for any purpose without needing attribution or permission. All participants in the survey are anonymous.

2.4 Summary of Cleaning Process

These steps were all done in R.

1. Merge hourly_intensity, hourly_calories and hourly_steps data to create an hourly_activity table.
2. Change column name for clarity.
3. Change inaccurate columns into proper data type
4. Check for missing values and found none
5. Check for duplicate rows and remove them

---

[1] Fitbit Fitness Tracker Data
[2] Fitbit Dictionary defining features and their implications

6. Merge daily_sleep data into daily activity data.

3. Analysis

For our analysis section, the entire process was performed in Rstudio.

3.1 Average User's Activity Level

3.1.1 Method & Analysis

To gauge user activity during the day, we looked at each user's average daily sedimentary minutes, lightly active minutes, fairly active minutes, and very active minutes. For sleeping habits, we examined the user's average sleep duration. Afterwards we checked total hours recorded to check data integrity.

For activities, we consider 1 minute of walking as equivalent to 1 minute of light activity and sedentary activity as just being stationary to put the situation into perspective. The metric we used to convert non-light activity minute to light active minute is:

1 sedimentary minute = 0 lightly active minute
1 fairly active minute = 2 lightly active minute
1 very active minute = 3 lightly active minute

After converting the activity minutes, we bin users into groups based on the average lightly active minutes worth of exercise a user does in a day. The following is how we categorize a user as sedentary, lightly active, fairly active, and very active:

Sedimentary users: 0 - 60 min of lightly active minute
Lightly Active users: 61 - 180 min of lightly active minute
Fairly Active users: 181 - 300 min of lightly active minute
Very Active users: 301+ min of lightly active minute

Analysis Report: Bellabeat Case Study

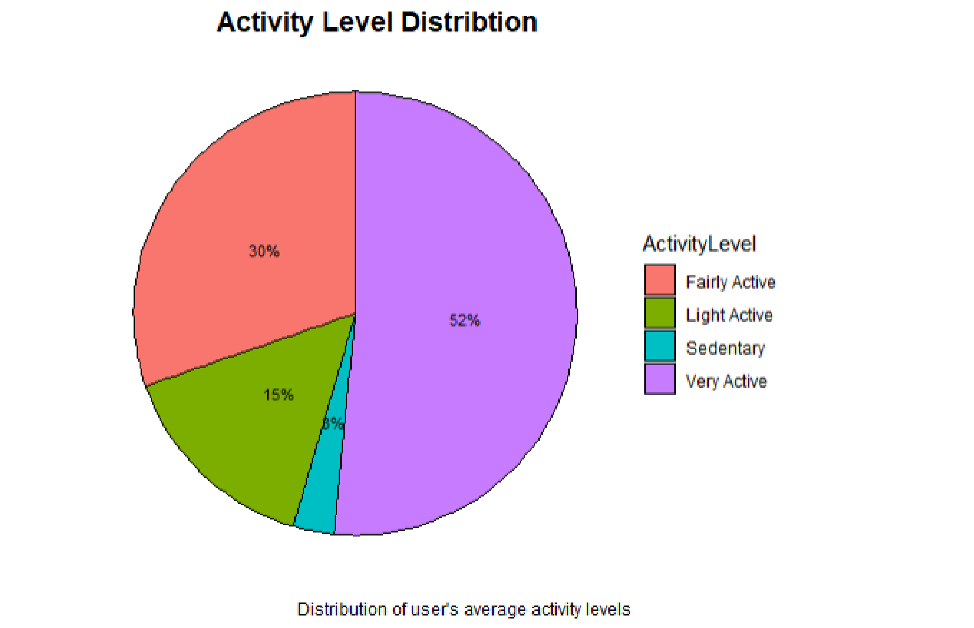

Figure 1

In addition, while investigating the user's overall activity level, we found conflicting information within Fitbit data. We compared rows with missing sleep data to rows with sleep data and found the rows missing sleep data to have much higher sedentary minutes than rows with sleep data. We can conclude rows without sleep data include their sleep time as a part of sedentary minutes while the ones with sleep data kept their sleep time separate from sedentary minutes. This explains why some sedentary minutes entries are abnormally high for rows missing sleep data.

Figure 2

Analysis Report: Bellabeat Case Study

The inaccuracy of the sleeping data in Figure 2 suggests a couple potential problems.

1. Rows less than 1440 minutes implies either the device is not recording properly to or the user's device was not active at some point during the day. Distinction is difficult without further user input.

2. Rows larger than 1440 minutes should not be possible because this implies there are more than 24 hours being recorded in one day. Potential causes are: Mishandling of information, a system error that causes devices to double count minute data or data from an adjacent day is being carried over.

We found users on average spend 458.5 minutes or roughly 7.7 hours asleep when taking the average of all time asleep entries.

3.1.2 Conclusion

Of the 33 users data, 52% are very active, 30% are fairly active, 15% are lightly active and the remaining are 3% sedentary. In addition, users on average spend 458.5 minutes or 7.7 hours sleeping.

Determining the reason for the inconsistency is outside the scope of this analyst report, but this outlines an issue to avoid with Bellabeat products. For the Bellabeat app, having a feature that accurately differentiate the two would be key in providing more accurate lifestyle recommendation.

3.2 Highest Hour of Activity

3.2.1 Method & Analysis

We are not provided a given activity level per minute so our previous methodology would not apply here. In addition, total steps within an hour is inaccurate because it doesn't record activities like swimming accurately. Thus our best metric of comparison is hourly intensity. In Figure 3, we took the average hourly intensity of every entry for each hour.

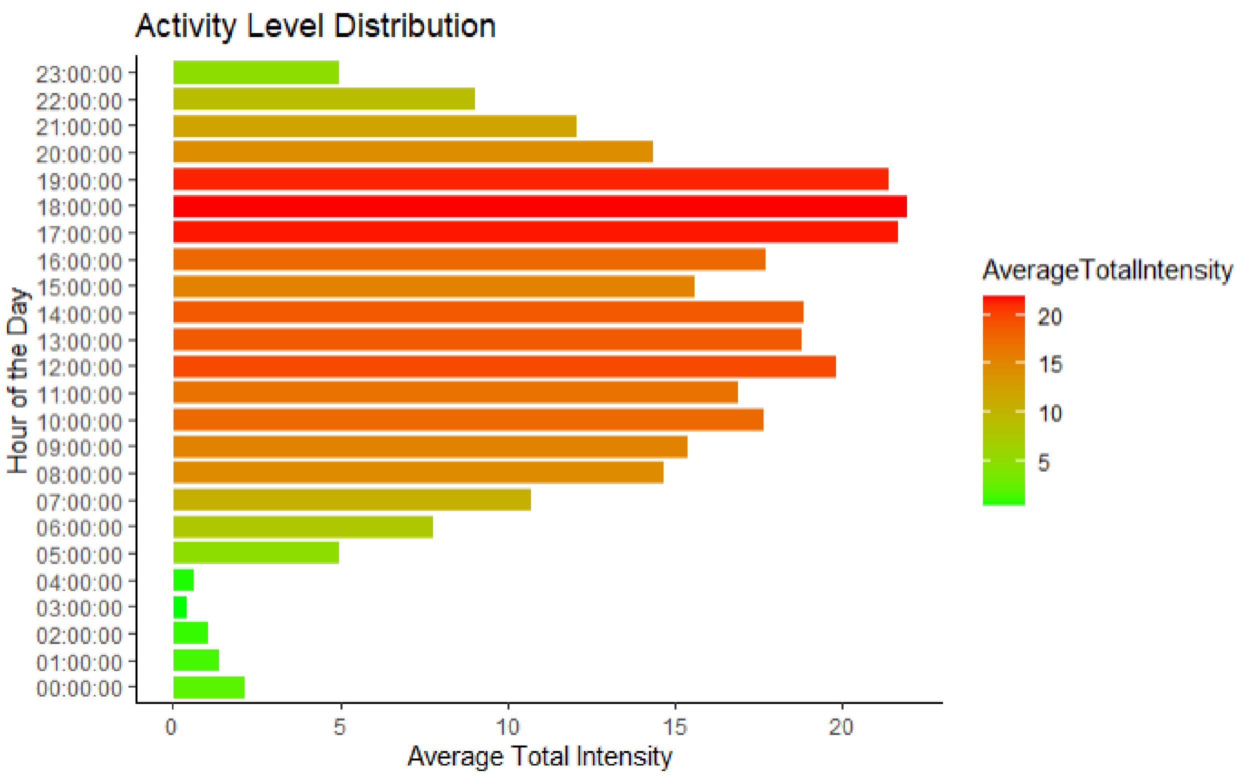

Figure 3

Based on the histogram, users are most active between 12pm - 2pm and 5pm - 7pm. The 12pm - 2pm and 5pm - 7pm time frames draw parallels to lunch break hours and after work hours in a normal 9-5 work day.

3.2.2 Conclusion

Users are most active between 12pm - 2pm and 5pm - 7pm. The 12pm - 2pm time frame most likely corresponds to work lunch break since people commute to find lunch. The 5pm - 7pm probably corresponds to after work hours when users commute back home or have free time to exercise.

3.3 Weekday Activities Levels

3.3.1 Method & Analysis

Following the same logic as section 3.2, we will be using hourly intensity as our primary metric. We took the average intensity of every entry and sorted them by weekday. The results are shown in Table 1 below.

Weekday <chr>	AverageTotalIntensity <dbl>
Friday	12.09309
Monday	12.11220
Saturday	12.90086
Sunday	10.98377
Thursday	11.92690
Tuesday	12.44278
Wednesday	11.75867

Table 1

Table 1 indicates users experience their most intense workout during Saturday while Sunday has the least intensity. Saturday and Sunday are usually non-work days meaning users have time to indulge in their personal hobby. Since Sunday has no work, users may be more inclined to go out to party, clubs, etc on Saturday night. On the other hand, users might be taking it easy on Sunday because they have work the following morning.

3.3.2 Conclusion

Users experience their most intense workout during Saturday while Sunday has the least intensity. We believe most users are off work during the weekend, presenting users more time to engage in different activities. Saturday is likely the day people designate to enjoy outdoor activities. Sunday is the designated resting period where people stay home and relax.

3.4 Time Spent Awake In Bed

3.4.1 Method & Analysis

To determine the period of time a user is laying in bed awake, we take the difference between time spent in bed and time asleep then average the results. We found users on average lie in bed awake for 39.3 minutes. Upon a brief inspection of the raw data set, we found records suggesting users occasionally either have trouble falling asleep, suffer from restless nights, trouble getting out of bed or even all 3.

3.4.2 Conclusion

Analysis Report: Bellabeat Case Study

Fitbit Users on average lay in bed awake for approximately 39.3 minutes everyday. We suggest including a function that to remind users on their phone to sleep or notify users they have been staying in bed for too long.

3.5 Relationship Between Amount of Sleep and Daily Workout Intensity

3.5.1 Method & Analysis

We plotted the length of time users were asleep against their activity that day and applied a linear regression on the data. In addition, we signify the number of different instances a person was recorded sleeping with rgb coloring.

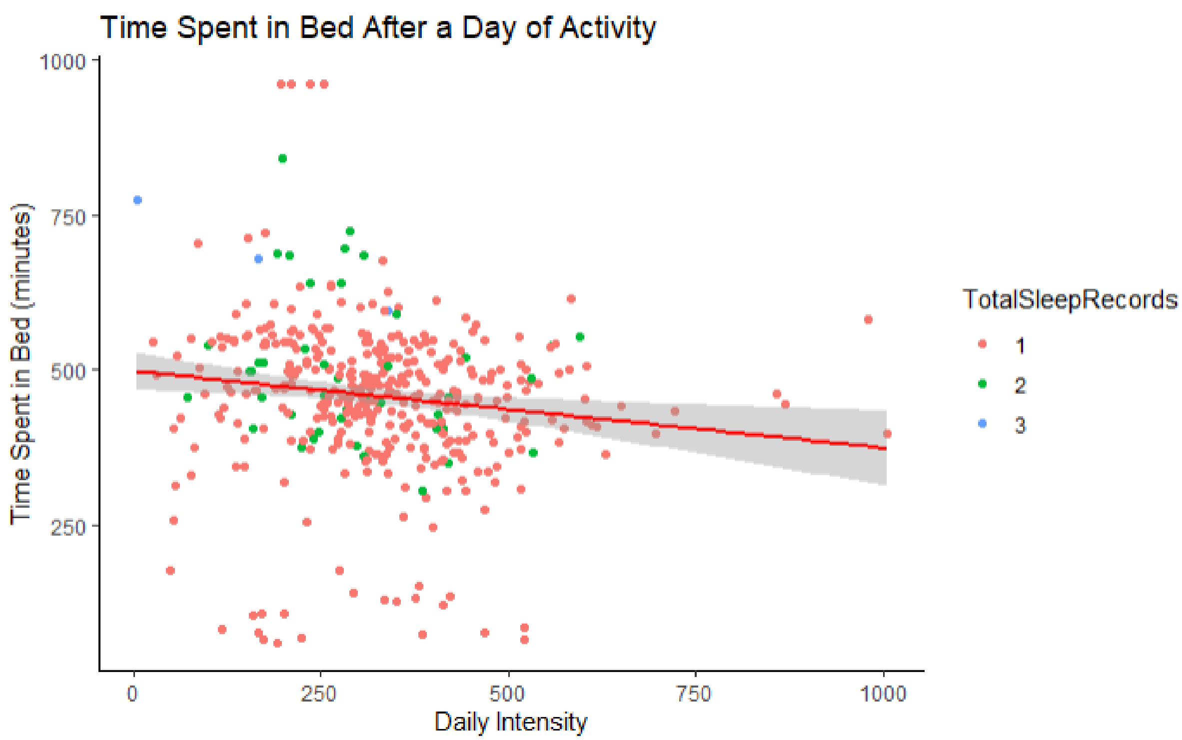

Figure 4

From Figure 4, the near flat slope indicates there is little correlation between how intense a user's day was and the amount of hours they spend in bed based on available data. Upon further inspection, records show users who sleep three times a day have notably spent more time in bed for an amount of daily activity than users who slept only once or twice. However, there is little to no difference in records of whether the user has slept once or twice. However, we will note there aren't enough users who slept three times in a day to make any conclusions.

Something to consider when trying to interpret the data. Depending on when a person sleeps, their period of sleep could span over two days rather than one. The impact of one day's worth of activity will sometimes affect the next day's sleep record. Similarly, the previous day's workout activity could be affecting the current day's sleep record. The situation becomes more complicated when we consider periods of sleep recorded in irregular hours. Given the complexity of this situation, for this analysis, we are working under the assumption that the daily time spent in bed begins upon leaving the bed and ends the following morning.

3.5.2 Conclusion

There is little connection between user workout intensity and amount of sleep a user receives per day. However we will admit our analysis for this section doesn't properly encapsulate the problem. Further data will be needed to properly assess this question.

4. Conclusions

In conclusion, the insights derived from Fitbit will be valuable assets in the continued improvement of the Bellabeat app. Although Bellabeat's focus audience is different from Fitbit's focus audience, we believe our insight will improve the services our company provides. The following is a list of our results along with recommendations for the next step to take.

1. Of the customer base, 52% are very active, 30% are fairly active, 15% are lightly active and the remaining are 3% sedentary.
2. There is conflicting information in how stationery minutes are recorded. Having a system that could differentiate inactivity and sleeping would improve Bellabeat' system health recommendations
3. Users on average spend 40 mins in bed awake and sleep for 7.7 hours per day. This implies people either have difficulty sleeping or rising up. Adding a setting to notify users on their phone to sleep or leave better after a certain period of time would be a good addition to the app's functionality.
4. Users are most active between 12 pm - 2 pm and 5 pm - 7 pm. An app notification as a reminder to exercise would be most effective during this time period. Giving recommendations to nearby food establishments would encourage exercise.
5. Users are most active during Saturday and least active on Sunday. We believe this is because people use Saturday to meet friends while Sunday is used to stay home and relax. We suggest having a feature that recommends interesting nearby places to encourage activity on Saturday. On Sunday we could recommend healthy at home activities or healthy recipes to cook.
6. There is little to no relationship between the user's daily workout intensity and hours slept.

For future works, we suggest using data taken from Bellabeat's user base rather Fitbit or a different competitor product that focus on women health. Information on whether a user is pregnant would provide us another angle to view our customer base needs. Additional information on what Bellabeat product a customer uses and how often they use the product would provide key info on what product needs improvement.

This concludes my report.