HOW CAN A WELLNESS COMPANY PLAY IT SMART? A CASE STUDY ON BELLABEAT

DIPTANGSU RAY

A project report submitted in conformity with the requirements
for the degree of Master's of Science in Information Technology

Department of Mathematical and Physical Sciences
Faculty of Graduate Studies
Concordia University of Edmonton

HOW CAN A WELLNESS COMPANY PLAY IT SMART? A CASE STUDY ON BELLABEAT

DIPTANGSU RAY

Approved:

Rossitza Marinova, Ph.D.

Supervisor | Date

Committee Member Name, Ph.D.

Committee Member | Date

Alison Yacyshyn, Ph.D.

Dean of Graduate Studies | Date

Abstract

Fitness trackers can help improve physical activity, and thereby reduce the risks of diseases such as diabetes, obesity, as well as hypertension. Despite these health related benefits, the way these devices are being used by the end user, play a crucial role for the persistent success of fitness trackers. To understand how these factors influence continuous use of fitness trackers and investigate the particular role of health and wellness improvement, this case study has analysed public domain fitness tracker data. Based on a rigorous iterative thematic data analysis, where the emphasis has been on matching deep diving with data and literature, a thematic framework has been developed and it identifies the prominent determinants. The findings propose new innovative ways through which these products and its ecosystem can be improved and how these devices can be marketed more efficiently. This case study will help researchers and academics to uncover as well as visualise user perceptions concerning fitness trackers and also provide practitioners with workable suggestions for ensuring their continuous use.

Keywords: data, metadata, physical activity, weight, sleep, calories, smart devices, wearable technology, data analysis, data manipulation, visualizations

Contents

List of Tables

List of Figures

1 Introduction

1.1 Background

Bellabeat is a company which deals in high-tech manufacturing of health focused smart devices for women. The company was founded in the year 2013 and three years later in 2016 they opened offices across the globe alongside creating multiple products under their product portfolio [8]. Bellabeat's mission is concerned with empowering women through providing them with data and insights to better discover themselves. Cofounded by Urška Sršen and Sandro Mur, Bellabeat quickly gained prominence. Despite of being a successful company, cofounder, and CEO Urška Sršen believes that BEllabeat has the potential of becoming a larger player in light of the global smart device market. She also believes that analysis of smart device fitness data could help unlock certain new growth opportunities for the company. This case study project will emphasize upon one of Bellabeat's products alongside analysing smart device data in order for gaining insights into how the consumers are in turn using their smart devices. Evidently, these insights will facilitate in guiding the company's marketing strategy.

1.2 Products

Following are the most popular offerings from Bellabeat's product portfolio [8]:

- Bellabeat App: The Bellabeat App offers users health-based data related to their activities, stress, sleep, menstrual cycle, as well as mindfulness habits. In turn this data can help the users better understand their current habits alongside helping them make healthy decisions. The Bellabeat App connects to the company's line of smart wellness products.
- Leaf: Leaf is Bellabeat's classic wellness tracker, and this device can be worn as a bracelet, necklace, or even as a clip. This Leaf tracker in turn connects to the Bellabeat App to track activity, sleep, as well as stress.
- Time: Time is Bellabeat's wellness watch, and this device combines the timeless look of a classic timepiece with that of contemporary smart technology, in order to track the users' activity sleep, and stress. This Time watch also connects to the same Bellabeat App to offer insights into the users' daily wellness.
- Spring: Spring is a smart water bottle which tracks the users' daily water intake with the use of smart technology and thereby ensuring that the users are appropriately hydrated throughout the day. The Spring bottle also connects with the Bellabeat App and helps users track their hydration levels.
- Bellabeat Membership: The company also offers their users a subscription-based membership program. With this membership, the users can have 24/7 access to fully personalised and customised guidance on sleep, nutrition, sleep, health and beauty, as well as mindfulness based on their goals and lifestyles.

This case study project is focused on the analysis of smart device usage data to gain meaningful insights into how the modern consumer use non-Bellabeat smart devices. Furthermore, the project has selected one product from Bellabeat's portfolio and has applied the insights from the data analysis for recommending marketing strategy for the same product.

1.3 Rationale

For the purpose of this project, a multidimensional analytical model is being used, and this is because otherwise it would be indispensable to investigate the determinants in regard to the continuous use of fitness trackers. Firstly, data cleaning has been done on Microsoft Excel and an initial data analysis has been conducted on SQL with 21 leading questions that helped uncover certain useful insights about the data and its quality. Furthermore, based on the findings of the initial analysis, a detailed data analysis has been carried out in R. This second level of analysis is the foundation for the overall findings of the project and acts as the footwork of the recommendations that has been made regarding how Bellabeat can further their marketing strategy. To sum up, by incorporating the specific constructs of data analysis, data storytelling, and deep diving with the data, this project has managed to develop an integrative framework when it comes to accounting for the multidimensional nature of user interaction with non-Bellabeat fitness trackers. To further investigate the particular role of Bellabeat's marketing team's strategy, this project has emphasized upon taking into account all of the user activities namely physical activity, sleep activity, calories burned information, weight log information, and physical activity data tracking preferences and have carried a detailed analysis to understand how these data features are significant in light of the metadata, if there is an corelation and what is the nature of such corelations, before deep diving with the data to performing a curated analytics on how the refined and filtered metadata can serve the purpose of this case study i.e. delivering insights and recommendations for Bellabeat's marketing team. Also, this project has carried out the analysis based on the metadata for 33 users, which differentiates this case study project from others that has been carried out on the same FitBit Fitness Tracker Data.

1.4 Objectives and Outcomes

1.4.1 Objectives

The objectives for this case study project are as below:

1. To discover what are some to the trends in smart device usage.
2. To understand how these trends could apply to Bellabeat's consumers.
3. To understand how these trends could help influence Bellabeat's marketing strategy.

1.4.2 Outcomes

The outcomes for this case study project are as below:

1. A clear and concise summary of the business task.
2. A detailed description of all data sources used.
3. Clear documentation of the data cleaning and data manipulation processes.
4. A summary of the data analysis carried out.
5. Supporting visualisations alongside key findings.
6. Top high-level content-based recommendations based on the data analysis and its findings.

2 Literature review

2.1 Wearable fitness tracker technology

If taken into consideration the fact that how the smart wearable devices market has gained prominence and traction over the course of the last decade, it can be seen that new fitness trackers and different kinds of smartwatches are being released to the global consumer markets every year [16]. These devices in turn comes equipped with different sensors, algorithms, as well as accompanying mobile applications. Moreover, with the recent advances in mobile sensor technology, privately collected physical activity-based data can actually be used as an addition to the existing methods for health data collection. Also, data collected from these devices have certain potential applications in healthcare, wellness, and personal hygiene. With an increasing number of diverse brands, there is need for an overview of how the users i.e., the consumers are using these devices and their features when deciding how to improve their health, wellness, and fitness goals.

According to the World Health Organization, 150 minutes of moderate intensity physical activity is recommended each week for adults and 60 minutes for children as well as adolescents [11]. But 25% of adults and over 80% of adolescents fail to achieve the recommended physical activity targets. Findings from the Tromsø Study suggests that only 30.4% of women and 22% of men manages to reach the recommended target [10]. Low physical activity is contemporarily the fourth leading risk factor for morality globally. Despite of the fact that there is limited evidence that using wearable fitness trackers will help improve health, these devices are still very popular, and hence new fitness devices keep appearing on the global consumer market on a regular basis. According to reports, in the year 2016, vendors shipped an estimated 102 million devices globally, as compared to 82 million in 2015. Furthermore, 57% of these devices were sold by top five brands: Fitbit, Xiaomi, Apple, Garmin, and Samsung. In the first quarter of 2017, 18 percent increment in the number of devices sold was marked, as compared to the first quarter of 2016. Thus, with a large number of available devices and brands, it is very difficult for users to navigate through an ever-growing list of brands and devices with various capabilities, price, and quality. Contrarily, it is also difficult for new market entrants like Bellabeat to compete with the already existing prominent players alongside creating a niche consumer base.

2.2 The company under consideration and theoretical background

Bellabeat being a new entrant as compared to the already existing prominent players, must find ways such that they could expand their global presence and in this regard the goal is to analyse public datasets concerned with non-Bellabet device-based fitness data to unlock trends and insight into how consumers are using these devices, and this could help Bellabeat's marketing team better understand the consumer's usage habits. This way the company can better improve their product strategy as well as marketing strategy.

Apart from the fact that fitness trackers being mobile health technology for private users, there also is no common understanding in existing research on how fitness tracker is defined. Hence one of the very first concerns is to clarify what actually constitutes a fitness tracker. It must be understood that these fitness trackers come in various shapes, such as wristbands, clips or textiles to track, analyse, communicate, and to monitor the minutiae of the users' everyday lives. Despite of these differences in appearances and functionality, all devices are worn directly on the body to unobtrusively collect the users' physiological data [15]. Secondly, in contrast to the health applications, health websites, and smartphones, these devices tend to collect the users' physiological data which usually cannot be accessed in other ways [9]. Besides conventional functionalities like number of steps on the floor, the biometrical sensors fitted on these devices are also capable of collecting PHI such as physical activity with an oximeter, or electrodermal sensors on these devices measure the users' stress levels. The devices can also be programmed with fitness goals and incorporate GPS in order to track runs and thereby inform the users on their sleep duration and quality via accelerometer [12]. Therefore, the users receive personalised, immediate, as well as goal-oriented feedback based on the functional range of their devices.

Thirdly, the aspect of functional range in light of these fitness trackers indeed has seen an expansion in recent past. By continually gathering PHI, fitness trackers allow individuals to monitor an array of medical risk factors and thereby provide the users with direct access to their PHI data. The analysis of the users' PHI is done on connected devices such as mobile devices, tablets, or PCs. This is because these applications use advanced data analytics as well as benchmarking in order for generating insights into the different aspects of the users' health status. To sum up, these fitness trackers are designed for private users and are worn by them on the body as small digital devices with biometrical sensors to continuously generate PHI without the need for health professionals.

3 Methodology

3.1 Business Task

Bellabeat, a wellness and technology company with a mission to empower women to reach their full potential, requires help with marketing the products in their portfolio. Bellabeat's product portfolio includes devices: Leaf, Ivy, and Time. These devices can track health data namely hydration, heart rate, menstrual cycles, sleep, and activities. In this scenario, Bellabeat's marketing team requests recommendations based on competitor data. For the purpose of this case study, Bellabeat's competitor Fitbit's user data will be analysed in order to reveal user trends in the wellness smart device market. The findings from this analysis will offer insights into areas of growth opportunity for Bellabeat going forward.

3.2 Data Sources

The data source is called 'FitBit Fitness Tracker Data' and can be found on popular data science and coding platform Kaggle [[14]]. The datasets were originally sourced from a survey which in turn was performed on Amazon Mechanical Turk workers for a research study, that collected Fitbit tracking data. According to the original study, 30 participants were surveyed in total. However, data associated with 33 users can actually be found in the dataset. Furthermore, no demographic information about the users is provided, such as age, sex, height, etc. Also, the exact Fitbit device models used by the participants is also not specified. As a result, it is being considered that the variation across datasets is potentially due to the varying device models worn by the participants, and for varying user preferences. The data for the purpose of this case study project is from the timeline April 12, 2016 – May 12, 2016. This data includes 33 users over 4 datasets which tracks data: physical activity, sleep time, weight information, and steps count.

1. 'Daily Activity Merged' dataset includes the daily activity logs for the 33 users. This dataset compiles 3 activity types, the distance covered for undertaking these activities, and time spent performing them; in minutes. As such, the 3 different activity types are: light, fairly active, and very active. The distance columns are not defined but are based on the step data which is provided in kilometers. The minutes spent without activity has been categorised as sedentary time. Furthermore, this dataset also includes the total steps taken, and the total calories burned.

2. "Hourly Steps Merged' dataset includes the IDs for all the 33 users, and thereby expands the daily steps taken into hourly increments which is categorised in 24 hour format. As mentioned earlier, there exists a certain variance between that of total steps calculated in this dataset as compared to the daily logs within the 'Daily Activity Merged' dataset above, is likely because of the device usage. As a result of this variance, this case study project will be using the step

information in this dataset only for the purpose of analysis concerned with steps per time of day.

3. 'Sleep Day Merged', dataset details 24 user IDs, their total minutes spent asleep, and total minutes spent on bed but not asleep. It can be found mentioned on Fitbit's website that their smartwatch tracks heart rate, alongside movement patterns in order for determining if the user is awake or asleep. The website also states that the 'Awake' category includes when the users are somewhere in a sleep cycle but are restless and thus wakes up briefly.

4. 'Weight Log Info Merged' dataset, records details of only 8 user's IDs, their weight in kg and lbs, BMI, and if the data was logged manually or automatically. This dataset also includes a column 'Fat'; but it can be found only twice, i.e., data is bucketed in only two cells.

3.3 Initial Data Cleaning

For the purpose of this project and ease of use, the initial data cleaning process involves tools: Microsoft Excel and SQL. The cleaning process starts with checking all the datasets for same discrepancies viz., blank spaces/empty buckets, data duplication, and other inconsistencies in the data. Below is the change log for the data cleaning on Microsoft Excel:

1. Shared changes across every table:
   Removing blank spaces through conditional formatting Verifying the User Id column entries are uniform, i.e., 10 characters in length through LEN function Adding underscores between words in column names Adding column 'Day' through date function Changing 'DateTime' columns into two separate ones: 'Date' and 'Time' through INT function

2. Activity:
   Changing the column name '`activitydate`' to 'Date' Changing column name 'steps' Removing 'TrackerDistance'. 'Logged_Distance, as well as 'Sedentary Active Distance' columns

3. Sleep:
   Changing column name 'sleepday' to 'Date' Substracting 'TimeAsleem' from column 'Total Time In Bed' and creating new column 'Time Awake' from the results Removing column 'Total Sleep Records'

4. Weight:
   Changing column name 'Is Manual Report' to 'Report Type' Changing column 'Report Type' responses from True/False to Manual/Automatic respectively Removing column 'Fat' Removing column 'LogId' `Pyt_hon`

3.4 Data Manipulation and Analysis using SQL

This section lists all of the queries that has been used for this project. Furthermore, each individual query is in turn separated by number and their respective leading question.

1. How many unique user IDs are in each of the tables?
   Full join gives results from all of the tables which are listed independently of the other tables's content.

2. How many of the users overlap in each of the tables?
   Counting the number of distinct IDs found in all of the tables listed. Inner join results in only the matching Ids found in all of the listed tables.

3. Which specific user IDs are in or are lacking from each of the tables?
   Verifying the IDs are consistent across tables and shows which IDs are shared or are absent from all the tables. Full join always generates results which includes the null values from all of the tables that have been joined.

4. Which user IDs overlap sets?
   Changing the join method to inner join as it results in only the user IDs which are found in all of the tables listed, excluding the null values. This in turn results in the user IDs which overlap all of the tables. These users in turn will be used later on in the form of the 'overlap' group, in order to observe if there are any trends in light of the users who use all sets, versus the ones that does not. Analysis: Only 6 users overlap in all of the datasets used.

5. How much activity are the users performing on an average?
   Average Activity data and the grouping is by user Id. The table is saved as dataset: 'activity avgs by id'. Analysis: 21 among the 33 users in turn tracked data for the entire month. Analysis: 20 users recorded at least 7,000 steps, 7 users recorded over 10,000 steps, and 14 users recorded below 7,000 steps. Analysis: 20 users are getting at least 20 minutes of a combination of very active level of activity and moderate level of activity. Many exceeded 20 minutes and 6 users recorded over an hour of this level of activity on an average.

6. How much sleep do users get on an average?
   Compiling the sleep data into averages by user IDS. The table saved as dataset: 'sleep totals by id'. Analysis: Only 3 users recorded their sleep for the entire month. Analysis: 15 of the 24 users did actually complete at least half of the month at 15 daily logs or more. Analysis: 12 users got at least 7 hours of sleep, and the other 12 users got less than 7 hours of sleep. Analysis: Most of the users have time disrupted from sleep, and 19 among the 24 users had more than 15 minutes of awake time during their sleep cycle.

7. Combining activity with sleep averages.
   Combining the sleep data with activity data and grouping by Id into a single table.

8. On which days does the most activity and least activity take place on?
The average user activity and sleep data is thus grouped by day in order to observe the day-to-day as well as the weekly trends. Assigning a numerical value to the days such that they can in turn be ordered correctly (Otherwise SQL orders them automatically in an alphabetical order) The table is saved as dataset: 'avg activity day' Analysis: On an average the users are recording over 7,000 steps, with Sunday being the exception. The users are also recording over 8,000 steps on Tuesdays and Saturdays. Analysis: The users are found to be meeting the weekly recommended 150 minutes – 300 minutes of activity and this activity is a combination of vigorous activity level and moderate activity level, with 243.44 minutes on an average. Analysis: The users are found to be travelling over 5 kilometers on an average daily. Analysis: The number of calories burned is found to be consistent around 2300 kilocalorie daily with Sundays and Thursdays being the only exceptions. Analysis: Saturdays are found to be the most active days with 244 minutes of combined activity (very active, fairly active, and lightly active levels), and the least activity is recorded on Sundays with 208 minutes Analysis: Monday is found to be the most sedentary day with 1027.9 minutes, while Thursday is the least with 961.9 minutes.

9. On which days do the users record the most and the least sleep?
Adding column for minutes asleep and converting to hours and counting how many of each day is included in its grouped row. The table is saved as dataset: 'avg sleep by day' Analysis: Users are found to be getting the most sleep along with the recommended at least 7 hours on Sundays, Wednesdays, and Saturdays. For the rest of the week, the users got 6.7 – 6.9 hours of sleep. The more sleep users get, the greater the amount of time they spend in bed awake throughout the week with Sundays recording an average of 50 minutes of restless sleep.

10. Combining activity and sleep data.
Selecting the sleep and activity level columns from that of the activity and sleep datasets. Joining the datasets to have both activity and sleep data into one single set.

11. Observing if there are any activity trends over time?
Averaging all the results by date and thereby sorting them into one row in regard to each specific date. The table is saved as 'avg act dates' Analysis: The users gradually stopped logging activity data over the month with the largest decline found to be occurring from May 8th – May 12th; 27 users to 21 users. Analysis: Users were found to have at least 7,000 steps on 27 of the 31 days and less then 7,000 on only 4 days.

12. Are there any sleep trends over time?
Analysis: Users did not use the sleep tracker consistently as seen by the variance in logs from day-to-day. Analysis: 16 of the 31 days the users met the recommended 7 hours of sleep. Analysis: It is also found that when averaged,

users do get 7 hours of sleep over month long timeline.

13. Combining activity and sleep data
    Combining the average activity and sleep data into one single table and thereby organising by date

14. What are the average weights and how often have they been logged?
    Finding the number of times users logged weight data and average weight data. The table is saved as dataset: 'weight avgs' Analysis: The average weight is found to be 171.54 pounds Analysis: The average BMI is 27.98 Analysis: Only 8 of the users tracked weight and among those, only 2 of the users checked weight significant number of times (24 and 30 logs respectively).

15. When are the users tracking their weights? (The following two queries are used to group the data before joining into one table)
    Finding the first weight data logged. The table is saved as: 'weight start date'. Finding the last weight data logged. The table is saved as: 'weight end date'. What are the changes in weight over time? Calculating the percentage change from start weight to end weight, and combining start weight and end weight data into one single table. The table is saved as: 'Weight percent change'. Analysis: 3 of the 8 users only checked their weight 2 – 5 times. However the logs were found to be spaced out in regard to the difference from their start date and end date being on average 19 days. Analysis: Among the weight users, the 2 users that has the most logs, recorded the most percentage change in weight. The users with 2 logs recorded little change between their 2 logs when compared.

16. How does the number of steps vary by day?
    Finding the average steps by day Ordering results by day of week. Case assigning numeric value to Days, such that the order is based on the value and not on the alphabetical order. Analysis: It is found that Saturday has the highest sleep count and the least is on Sunday.

17. How does the steps vary by time?
    Finding the number of steps per hour. Analysis: It is found that the users tend to gradually increase their number of steps as the morning progresses (3am – 10am). Analysis: It is found that step count wavers up and down in the afternoon. There is a peak in the evening before a rapid decline into the night.

18. Are there any differences in the activity performed by the users that tracked activity, sleep, and weight compared to those who did not track weight?
    Compiling user Ids that overlap across activity, sleep, and weight datasets. The table is saved as dataset: 'overlap ids'. Compiling average activity data and number of activity logs for the six overlapped user IDs. This way the data is pulled only for the six user Ids found in both the sleep and weight datasets. This way the data is pulled only for the six user Ids found in both the sleep

and weight datasets.

19. How do the overlap users' activity and their sleep data vary by day? (The following two queries are being used in order for grouping the data before joining into one single table)
Averaging the activity steps of the six user IDs which overlap across activity, sleep, and weight datasets. Designating to pull the data from the six user Ids in the overlap dataset. Averaging sleep activity data for the six users with data in activity, sleep, and weight. The table is saved as dataset: 'overlap sleep avg'. Combining datasets into one with both the sleep data and activity data for the six user IDs which overlapped datasets.

20. What percentage does each activity make up of an average day for all the users? Calculating the percentage each of the activity types make up on an average day. Calculating the total average minutes of combined activity types. Averaging activity minutes from datasets grouped by weekdays into one daily average. Dataset saved as 'daily avg percents'. It is furthermore imported into Microsoft Excel: The number format is then changed from decimals to percentages. Analysis: Most time is found to be spent on sedentary (59.1%) (16.5 hours) and the least time spent found on fairly active level of activity ((0.8 percent). Analysis: Activity light, fair, and very combined makes up 12.8 percent (3.79 hours) of daily time. Analysis: Sleep makes up 25% (6.99 hours).

21. What percentage does each activity make up of an average day for the overlap users IDs?
Calculating the percentage of each activity type out of the total daily activity in minutes. Adding all the activity minutes to determine the average daily total. Combining the sleep data with activity level data. The table is saved as dataset: 'overlap weekly act percents'.

3.5 Data Manipulation and Analysis using R

3.5.1 Preparation and Setup

The following libraries are imported:

tidyverse: The tidyverse is an opinionated collection of R packages designed for data science. All packages share an underlying design philosophy, grammar, and data structures [1].

lubridate: Date-time data can be frustrating to work with in R. R commands for date-times are generally unintuitive and change depending on the type of date-time object being used. Moreover, the methods used with date-times must be robust to time zones, leap days, daylight savings times, and other time related quirks, and R lacks these capabilities in certain situations. Lubridate makes it easier to do the things R does with date-times and possible to do things R does not [2].

Hms: The hms package provides a simple class for storing durations or time-of-day values and displaying them in the hh:mm:ss format. This class is intended to simplify data exchange with databases, spreadsheets, and other data sources. It stores values as numeric vector that contains the number of seconds since midnight. It furthermore supports construction from explicit hour, minute, or second values alongside supporting the coercion to and from various data types, including POSIXt. Moreover, hms can also be used as a column in a data frame and is based on the difftime class. The values can exceed the 24-hour boundary or be negative. Also, by default, fractional seconds up to a microsecond are displayed, regardless of the value of the 'digits.secs' option [3].

Scales: One of the most difficult parts of any graphics package is scaling, converting from data values to perceptual properties. The inverse of scaling, making guides (legends and axes) that can be used to read graph, is often even harder! The scales packages provide the internal scaling infrastructure used by ggplot2 and gives the user tools to override the default breaks, labels, transformations, and palettes [4].

Janitor: The janitor functions expedite the initial data exploration and cleaning that comes with any new dataset [5].

Skimr: Skimr is designed to provide necessary statistics about variables in data frames, tibbles, data tables and vectors. It is opinionated in its defaults, but easy to modify. In base R, the most similar functions are summary() for vectors and data frames and fivenum() for numeric vectors [6].

Highcharter: Highcharter is a R wrapper of HighCharts javascript library and its module. It helps in creating various charts with the same style like scatter, bubble, time series, heatmaps, treemaps, bar charts, etc. It supports various R objects, alongside supporting Highstocks charts, and Choropleths. It has a pipeline style which is efficient and effective for R programming and offers a large variety of themes with good aesthetics [7].

The datasets 'dailyActivity merged.csv' and 'sleepDay merged.csv' are imported. As seen in Figure 1, and as seen in Figure 2,

```
Rows: 940 Columns: 15
── Column specification ────────────────────────────────────────────────────────
──────
Delimiter: ","
chr  (1): ActivityDate
dbl (14): Id, TotalSteps, TotalDistance, TrackerDistance, LoggedActivitiesDistance, VeryActiveDistance,
M...

ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Figure 1: dailyActivity merged.csv

Displaying the head of the datasets as seen in Figure 3, and as seen in Figure 4,

```
Rows: 413 Columns: 5
── Column specification ────────────────────────────────────────────────────────
Delimiter: ","
chr (1): SleepDay
dbl (4): Id, TotalSleepRecords, TotalMinutesAsleep, TotalTimeInBed

ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Figure 2: sleepDay merged.csv

Id <dbl>	ActivityDate <chr>	TotalSteps <dbl>	TotalDistance <dbl>	TrackerDistance <dbl>
1503960366	4/12/2016	13162	8.50	8.50
1503960366	4/13/2016	10735	6.97	6.97
1503960366	4/14/2016	10460	6.74	6.74
1503960366	4/15/2016	9762	6.28	6.28
1503960366	4/16/2016	12669	8.16	8.16
1503960366	4/17/2016	9705	6.48	6.48

6 rows | 1-5 of 15 columns

Figure 3: dailyActivity merged.csv head

Id <dbl>	SleepDay <chr>	TotalSleepRecords <dbl>	TotalMinutesAsleep <dbl>	TotalTimeInBed <dbl>
1503960366	4/12/2016 12:00:00 AM	1	327	346
1503960366	4/13/2016 12:00:00 AM	2	384	407
1503960366	4/15/2016 12:00:00 AM	1	412	442
1503960366	4/16/2016 12:00:00 AM	2	340	367
1503960366	4/17/2016 12:00:00 AM	1	700	712
1503960366	4/19/2016 12:00:00 AM	1	304	320

6 rows

Figure 4: sleepDay merged.csv head

A glimpse of the data as seen in Figure 5, and as seen in Figure 6,

```
Rows: 940
Columns: 15
$ Id                       <dbl> 1503960366, 1503960366, 1503960366, 1503960366, 1503960366, 150396036
6, 15…
$ ActivityDate             <chr> "4/12/2016", "4/13/2016", "4/14/2016", "4/15/2016", "4/16/2016", "4/1
7/201…
$ TotalSteps               <dbl> 13162, 10735, 10460, 9762, 12669, 9705, 13019, 15506, 10544, 9819, 127
64, …
$ TotalDistance            <dbl> 8.50, 6.97, 6.74, 6.28, 8.16, 6.48, 8.59, 9.88, 6.68, 6.34, 8.13, 9.0
4, 6.…
$ TrackerDistance          <dbl> 8.50, 6.97, 6.74, 6.28, 8.16, 6.48, 8.59, 9.88, 6.68, 6.34, 8.13, 9.0
4, 6.…
$ LoggedActivitiesDistance <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0,…
$ VeryActiveDistance       <dbl> 1.88, 1.57, 2.44, 2.14, 2.71, 3.19, 3.25, 3.53, 1.96, 1.34, 4.76, 2.8
1, 2.…
$ ModeratelyActiveDistance <dbl> 0.55, 0.69, 0.40, 1.26, 0.41, 0.78, 0.64, 1.32, 0.48, 0.35, 1.12, 0.8
7, 0.…
$ LightActiveDistance      <dbl> 6.06, 4.71, 3.91, 2.83, 5.04, 2.51, 4.71, 5.03, 4.24, 4.65, 2.24, 5.3
6, 3.…
$ SedentaryActiveDistance  <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.0
0, 0.…
$ VeryActiveMinutes        <dbl> 25, 21, 30, 29, 36, 38, 42, 50, 28, 19, 66, 41, 39, 73, 31, 78, 48, 1
6, 52…
$ FairlyActiveMinutes      <dbl> 13, 19, 11, 34, 10, 20, 16, 31, 12, 8, 27, 21, 5, 14, 23, 11, 28, 12,
34, …
$ LightlyActiveMinutes     <dbl> 328, 217, 181, 209, 221, 164, 233, 264, 205, 211, 130, 262, 238, 216,
279,…
$ SedentaryMinutes         <dbl> 728, 776, 1218, 726, 773, 539, 1149, 775, 818, 838, 1217, 732, 709, 81
4, 8…
$ Calories                 <dbl> 1985, 1797, 1776, 1745, 1863, 1728, 1921, 2035, 1786, 1775, 1827, 194
9, 17…
```

Figure 5: dailyActivity merged.csv

The unique user Ids found in the two datasets are as seen in Figure 7,

As mentioned earlier, this dataset has some limitations, i.e., no demographic information about the users is not provided. Some of the limitations are as below:

- Gender
- Age
- Height
- Weight
- Device Type
- Activity Type
- Activity intensity
- Sleep quality

```
Rows: 413
Columns: 5
$ Id                 <dbl> 1503960366, 1503960366, 1503960366, 1503960366, 1503960366, 1503960366, 1503
9603…
$ SleepDay           <chr> "4/12/2016 12:00:00 AM", "4/13/2016 12:00:00 AM", "4/15/2016 12:00:00 AM",
"4/16…
$ TotalSleepRecords  <dbl> 1, 2, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1,…
$ TotalMinutesAsleep <dbl> 327, 384, 412, 340, 700, 304, 360, 325, 361, 430, 277, 245, 366, 341, 404, 3
69, …
$ TotalTimeInBed     <dbl> 346, 407, 442, 367, 712, 320, 377, 364, 384, 449, 323, 274, 393, 354, 425, 3
96, …
```

Figure 6: sleepDay merged.csv head

Data	Number of Respondents
daily_activity	33
sleep_day	24

Figure 7: Displaying the two data sets with number of respondents.

3.5.2 Creating new variables in daily activity v2

Assuming that the respondents are in the age group of 18 – 64 years, the World Health Organization (WHO) recommends that at least 150 to 300 minutes of moderate-intensity aerobic physical activity or at least 75 minutes – 150 minutes of vigorous-intensity aerobic physical activity in a 24-hour day. Now, for the purpose of this project, it is being defined that vigorous-intensity activity as the 'very active minutes' column. Furthermore, to describe moderate-intensity activity, a new column is being created 'moderate active minutes' which is being defined as 'fairly active minutes'. Also, an 'activity level' column is being created which is based on the 'moderate active minutes' and 'very active minutes' columns. As seen in Figure 8,

New Variable	Definition
moderate_active_minutes	fairly_active_minutes + lightly_active_minutes
activity_level	based on moderate_active_minutes and very_active_minutes

Figure 8: Displaying the new variables created in daily activity v2 dataset and their respective definitions.

3.5.3 Creating new variables in sleep day v2

A 'sleep duration' column is created based on 'total minutes asleep'. A numerical vector is now being converted in minutes to a period object measured in hours and minutes. Now, assuming that the participants are in the age group of 18 years – 64 years, the Centers for Disease Control and Prevention (CDC) recommends 7 hours or more sleep per day. As such, based on the 'sleep duration', a new column 'recommended sleep' is being created. Furthermore getting 7 hours or more sleep is being defined as 'sufficient' and getting less than that as 'short'. Another new column 'sleep efficiency' is created based on a widely published operational definition, the ratio of total sleep time to time in bed (multiplied by 100 to yield a percentage). As seen in Figure 9,

New Variable	Definition
sleep_duration	based on total_minutes_asleep
recommended_sleep	7 or more hours of daily sleep is "sufficient", getting less is "short"
sleep_efficiency	total_minutes_asleep / total_time_in_bed

Figure 9: Displaying the new variables created in sleep day v2 dataset and their respective definitions.

3.5.4 Combining data

As seen in Figure 10, the data head for the combined dataset is being displayed.

id <fctr>	date <date>	day_of_week <ord>	activity_level <fctr>	total_steps <dbl>	total_distance <dbl>
1503960366	2016-04-12	Tuesday	Active	13162	8.50
1503960366	2016-04-13	Wednesday	Normal	10735	6.97
1503960366	2016-04-15	Friday	Normal	9762	6.28
1503960366	2016-04-16	Saturday	Normal	12669	8.16
1503960366	2016-04-17	Sunday	Normal	9705	6.48
1503960366	2016-04-19	Tuesday	Normal	15506	9.88

6 rows | 1-6 of 22 columns

Figure 10: Combined data head.

As such, the number of unique user Ids in the datasets are as shown in Figure 11:

Data	Number of Respondents
daily_activity_v2	33
sleep_day_v2	24
combined_data	24

Figure 11: The number of unique user Ids in the datasets.

As seen in Figure 12, it can be observed that majority of the respondents tracked over 20 observations during the sample time period.

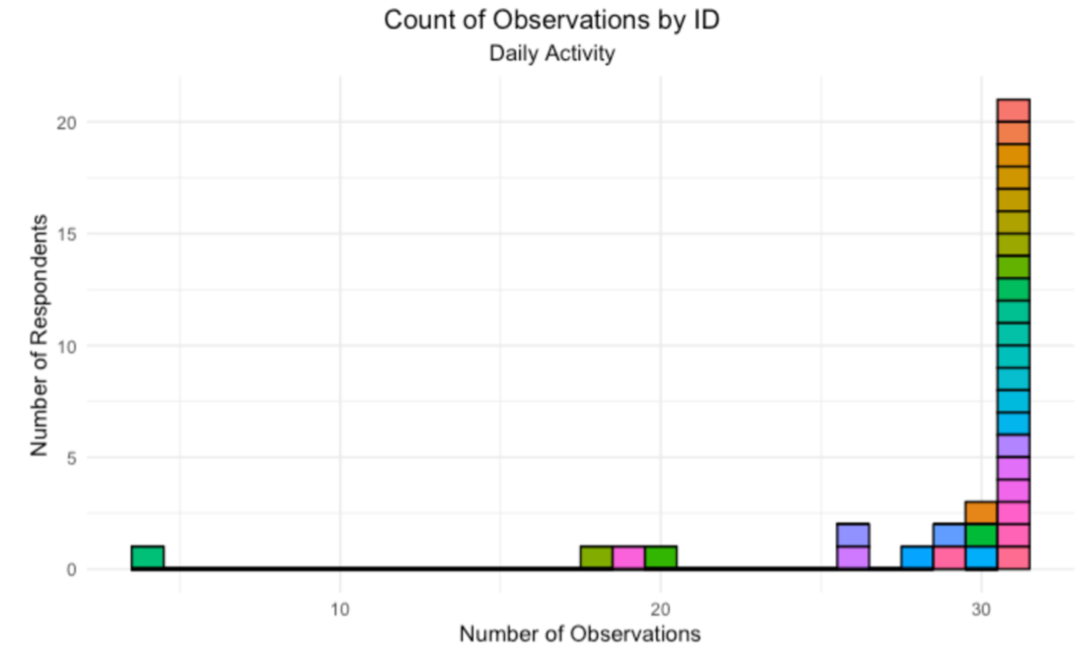

Figure 12: Count of Observations by ID.

Now, checking how many steps did the respondents take in a day, As seen in Figure 14, As seen in Figure 13, it is found that the respondents took an average of 7638 steps in a day.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	3790	7406	7638	10727	36019

Figure 13: Average Steps Took by Respondents in a Day.

4 Findings

4.1 Results from SQL

4.1.1 Activity Findings

Figure 14–16 show the observation for the activity findings.

As seen in Figure 14, activity is mostly light level. However, the users are still meeting the recommended weekly minimum of 150 minutes of activity. A trend can be observed here that the users are becoming less active towards the end of the week across all the activity levels.

Figure 15 shows the average calories burned per minute of different activity levels.

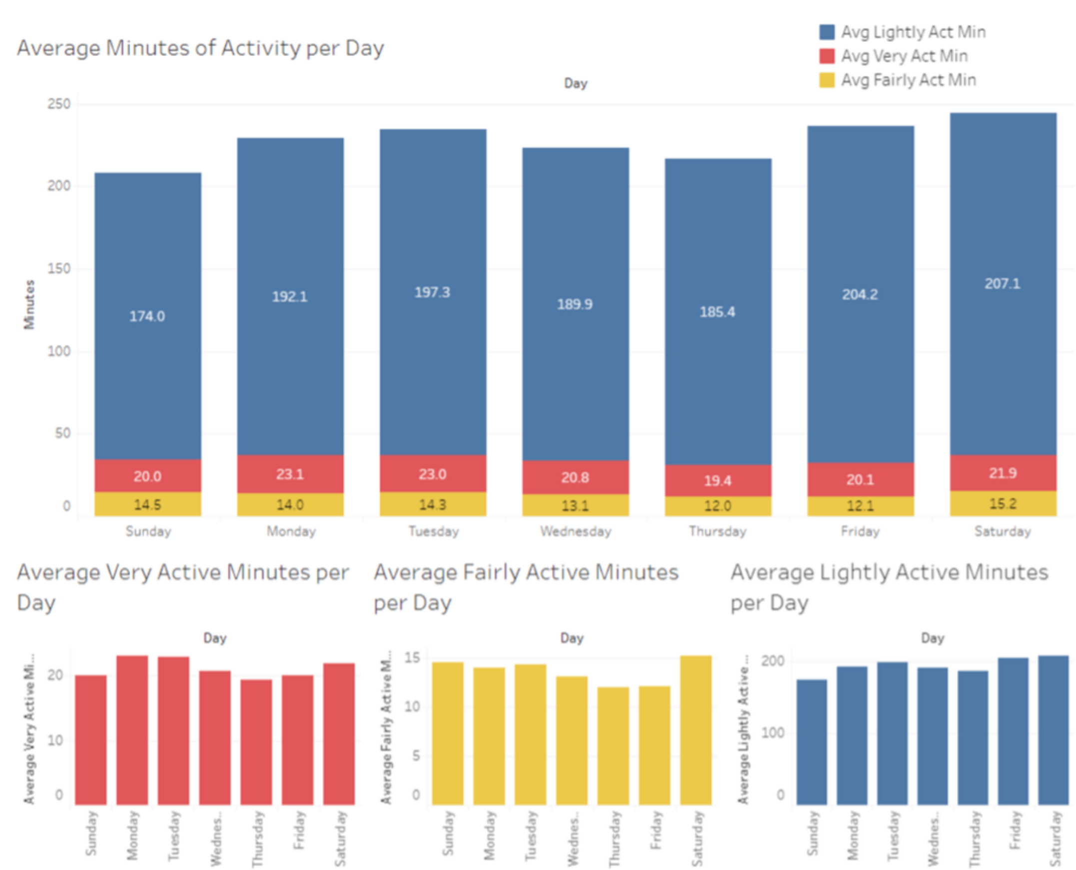

Figure 14: Average minutes of activity per day

The results indicate there is a certain positive correlation between the more vigorous the activity level the more calories burned.

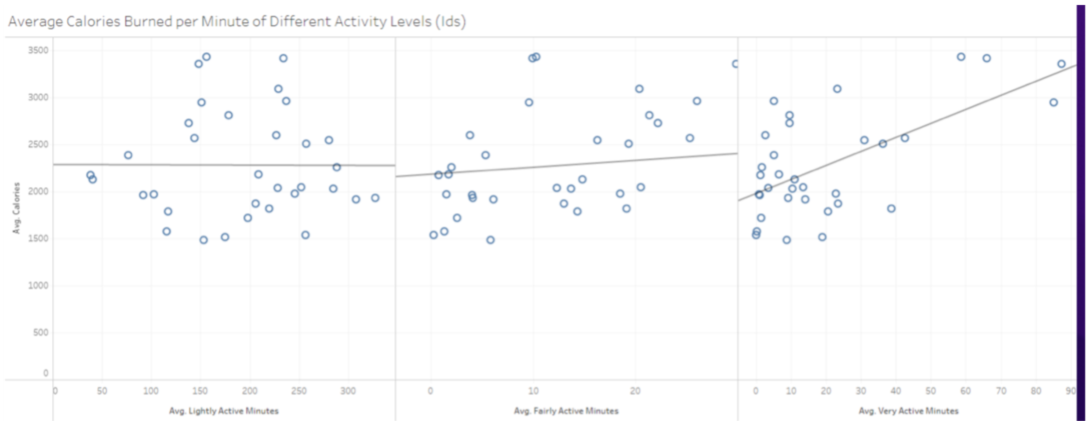

Figure 15: Average calories burned per minute of different activity levels (Ids).

Figure 16 presents the number of users who logged activity data over time. It can be seen that the users gradually stopped tracking activity data over time. Among the 33 users, only 22 completed the month.

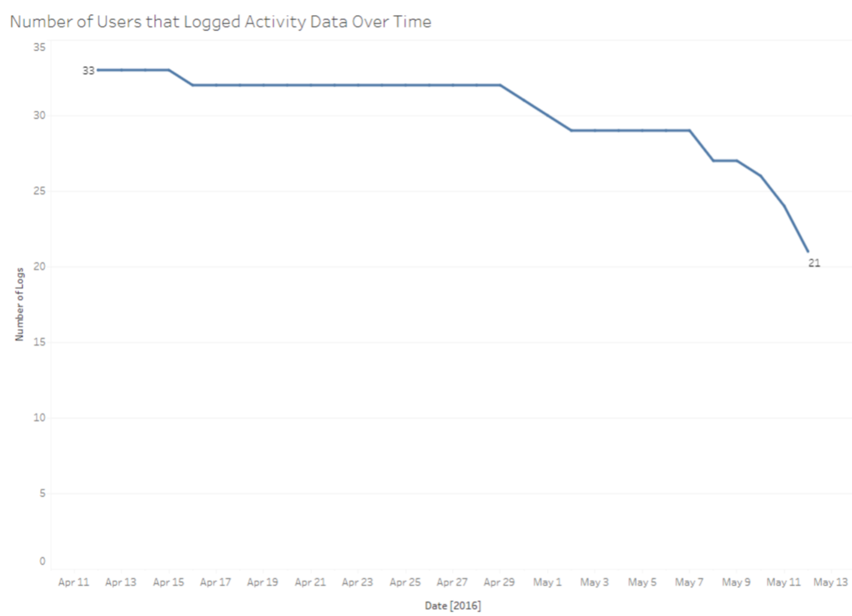

Figure 16: Number of users that logged activity data over time.

4.1.2 Steps Findings

Figure 17–20 show the observation for the activity findings.

As seen in Figure 17, it can be observed that the users are actually fulfilling close to 7,000 steps. Furthermore it can also be seen that there is a certain noticeable decline when it comes to the number of users who are tracking their steps; as well as the ones who did not track their steps, recorded significantly less steps.

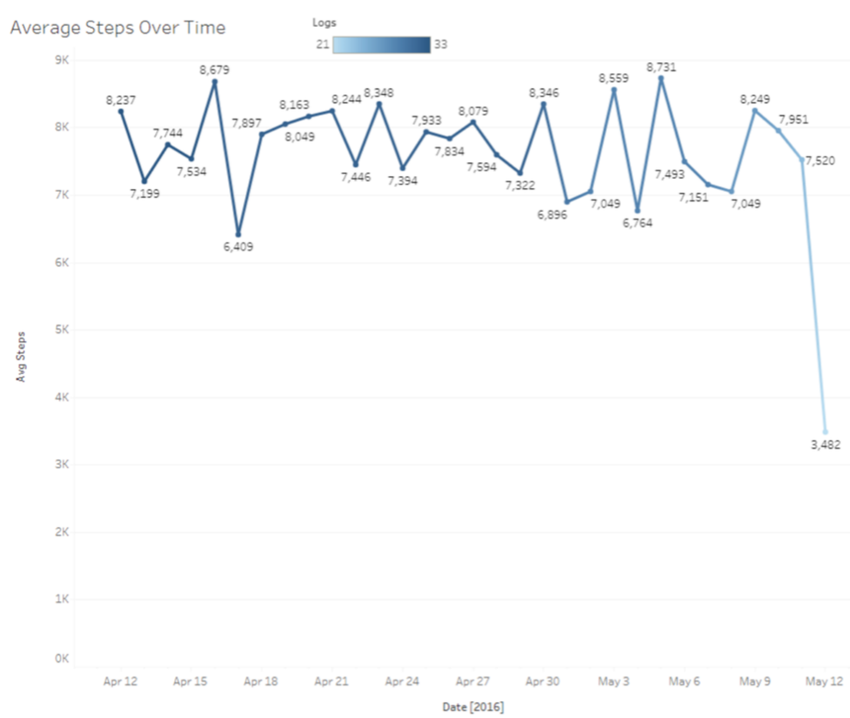

Figure 17: Average Steps Over Time.

From Figure 18, it can be seen that the users indeed are typically taking more steps on Tuesdays and Saturdays with a plateau in mid week.

From Figure 19, it is observed that on an average the peak number of steps usually occurs in the evening and then rapidly declines.

In Figure 20, it can be observed that there exists a certain positive correlation between taking more steps and burning more calories.

4.1.3 Sleep Findings

Figure 21–26 show the observation for the sleep findings.

As seen in Figure 21, it can be observed that there exists a positive corelation between burning more calories and taking more steps.

As seen in Figure 22, it can be observed that the two outliers are now removed in order to have a clearer view of the data. As a result, it can be seen that the longer the user sleeps, the more time awake they are actually experiencing throughout their sleep cycles.

As seen in Figure 23, it can be observed that the users tend to get similar amounts of sleep all across the week. However, it can also be seen that they vary from getting

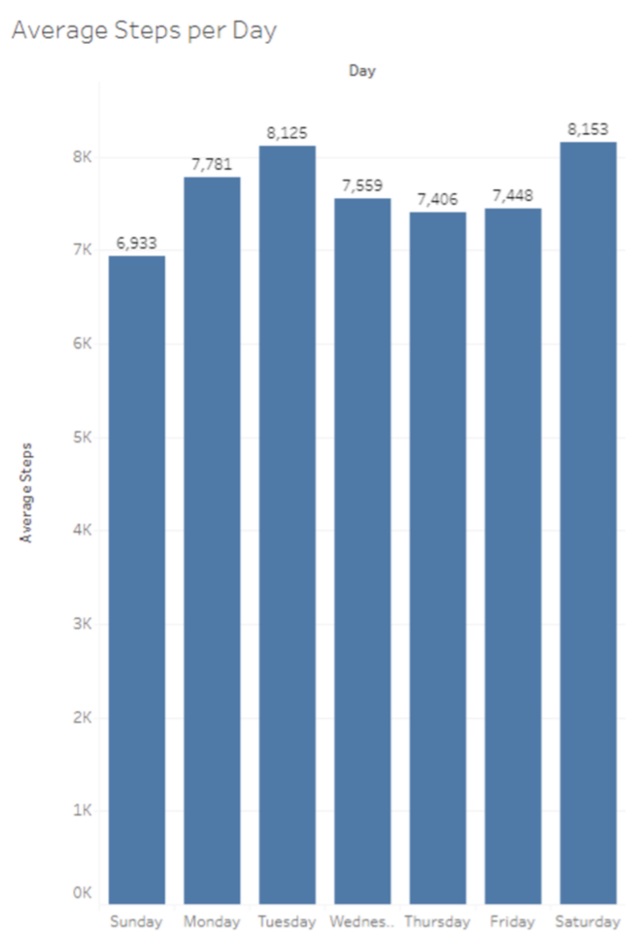

Figure 18: Average Steps per Day.

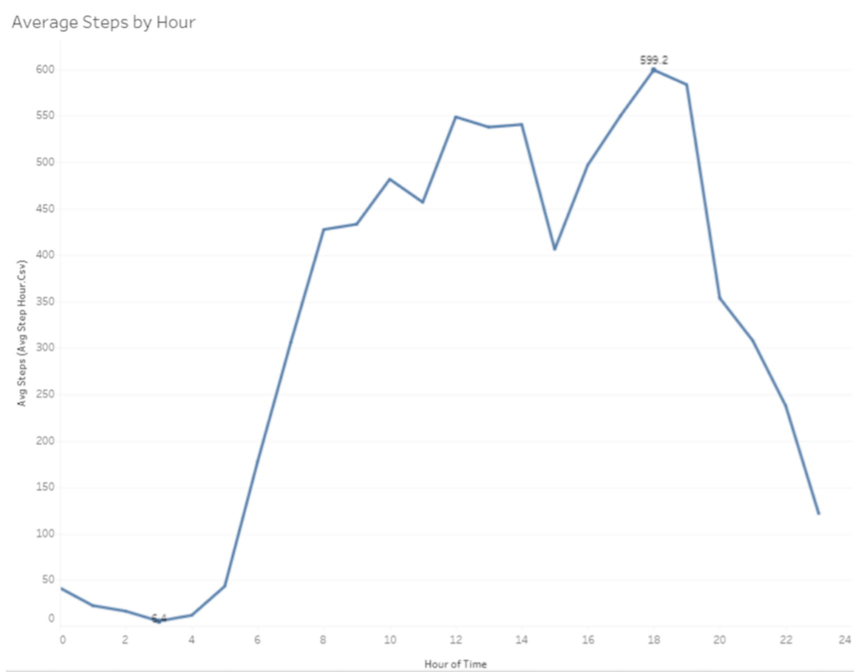

Figure 19: Average Steps by Hour.

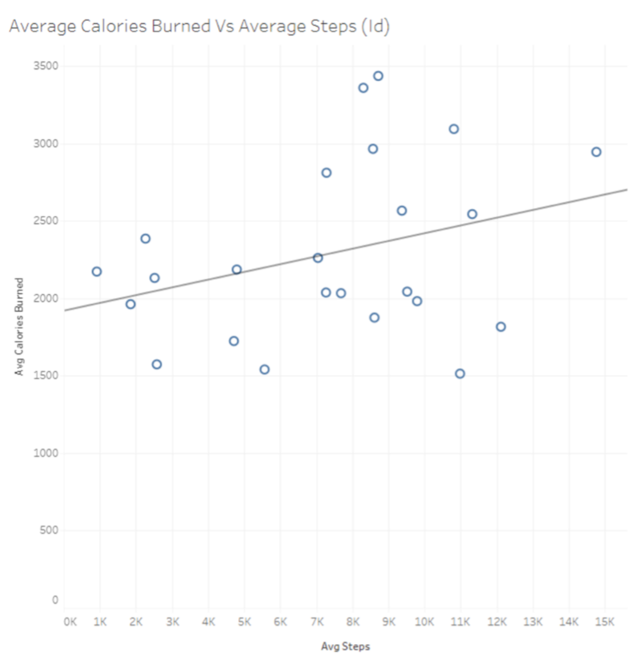

Figure 20: Average calories burned Vs. Average Steps (Id)

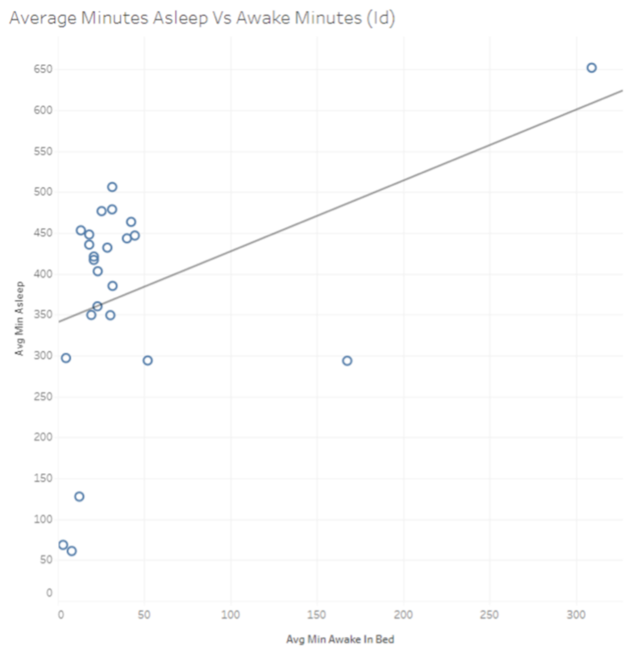

Figure 21: Average Minutes Asleep Vs. Awake Minutes (Id).

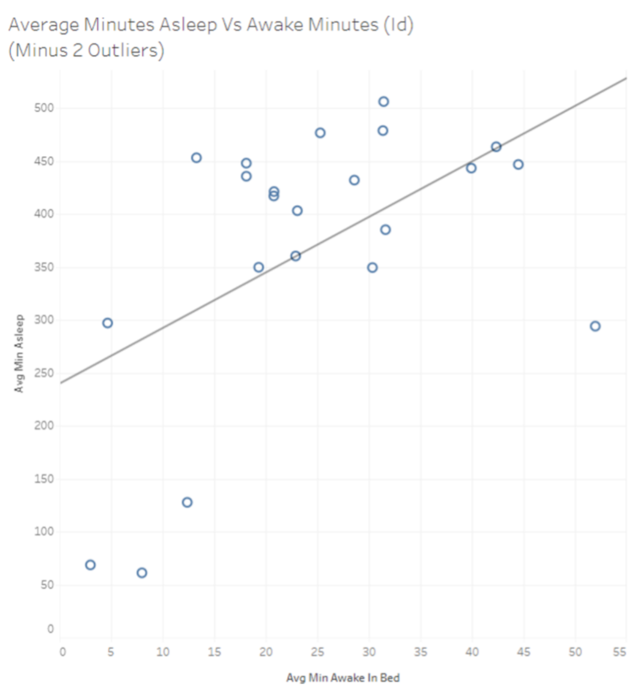

Figure 22: Average Minutes Asleep Vs. Awake Minutes (Id); (Minus 2 Outliers).

the recommended amount to less than throughout the week. Also, users get the most sleep on Sundays, but there is a decline in the beginning as well as at the end of the week.

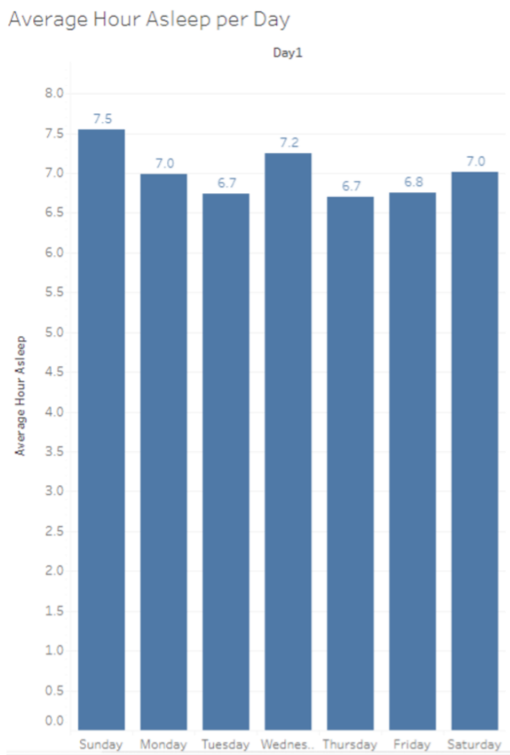

Figure 23: Average Hour Asleep per Day

As seen in Figure 24, it can be observed that users are spending more time in bed awake on the weekends as compared to the weekdays.

As seen in Figure 25, it is found that sleep varies over time but averages around 7 hours a night.

As seen in Figure 26, it can be seen that not all the users tracked sleep data every night. The users are found to be tracking their sleep inconsistently.

4.1.4 Daily Time Spent Findings

Figure 27–32 show the observation for the activity findings.

As seen in Figure 27, it can be found that the data for All Users and Overlap Users have been visualized.

As seen in Figure 28, it can be seen that there does not appear to be a significant difference in daily activity between the users who track weight and those who does not. It can also be seen that the greatest variance is that the overlap users tend to spend less sedentary time (−2%) and more time asleep (+1.5%). Furthermore, the

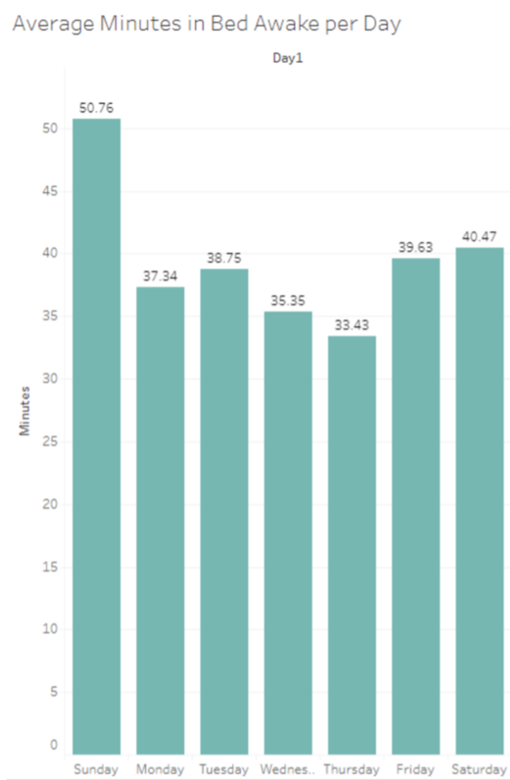

Figure 24: Average Minutes in Bed Awake per Day.

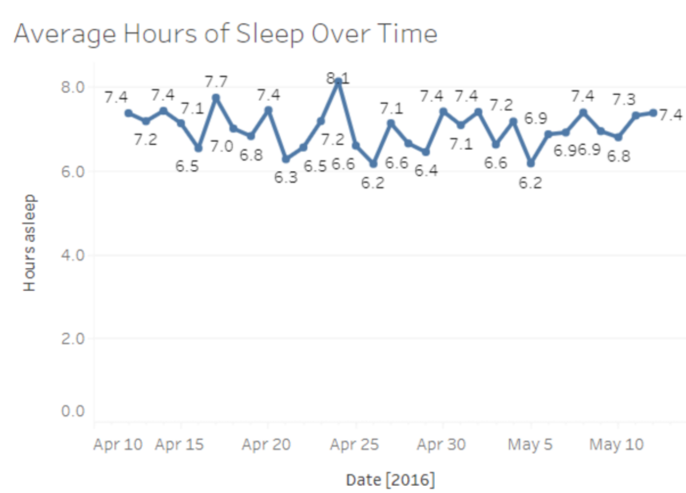

Figure 25: Average Hours of Sleep Over Time.

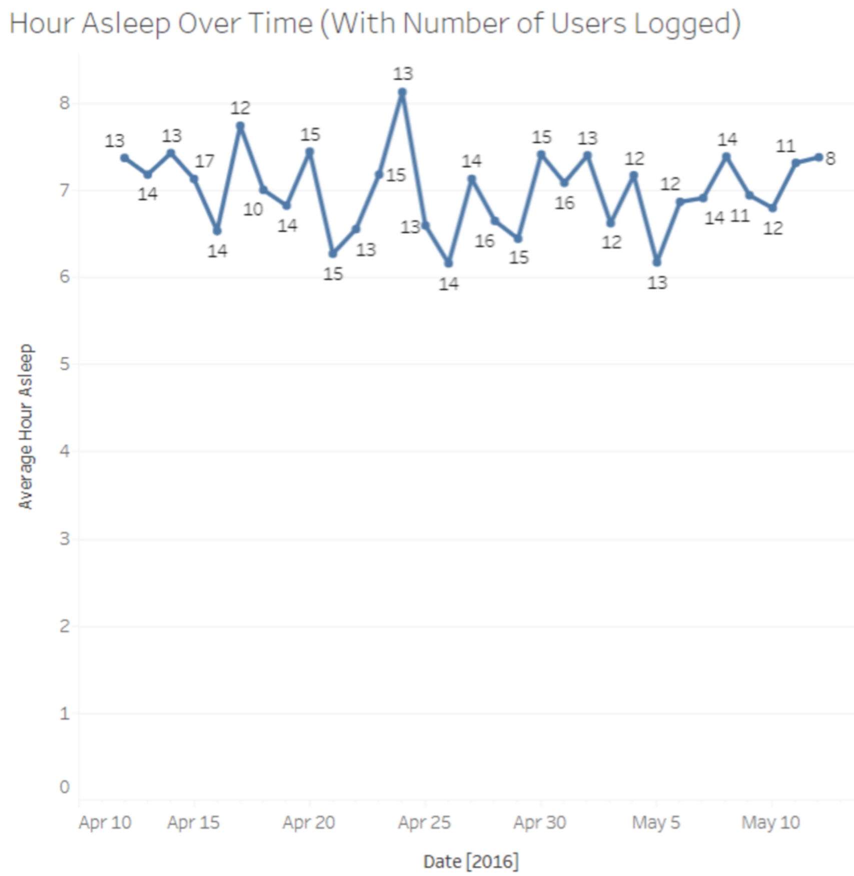

Figure 26: Hour Asleep Over Time (With Number of Users Logged).

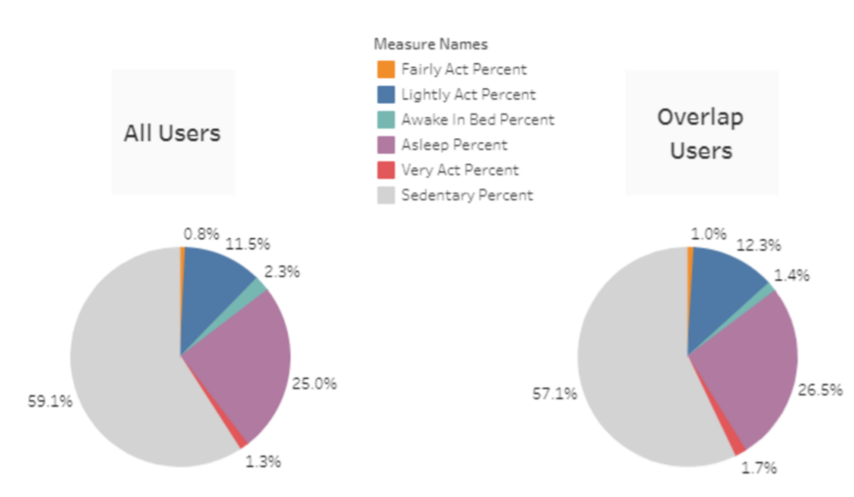

Figure 27: Visualizing data for All Users and Overlap Users, in the form of Pie Charts.

weight users might also be getting more rested sleep, as they get more sleep and less time awake in bed compared to the rest of the users.

Users	very_act_percen	fairly_act_percent	lightly_act_percent	sedentary_percent	asleep_percent	awake_in_bed_percent
all	1.30%	0.80%	11.50%	59.10%	25.00%	2.30%
overlap	1.70%	1.00%	12.30%	57.10%	26.50%	1.40%
Change	0.40%	0.20%	0.80%	-2.00%	1.50%	-0.90%

Figure 28: Visualizing data for All Users and Overlap Users, in the form of heat map with the deeper the color the more the intensity of change and vice versa.

As seen in Figure 29, 4 of the users with longs ranging between 2 – 5, performed infrequent logs and this might suggest that users prefer periodic weight check-ins over daily monitoring.

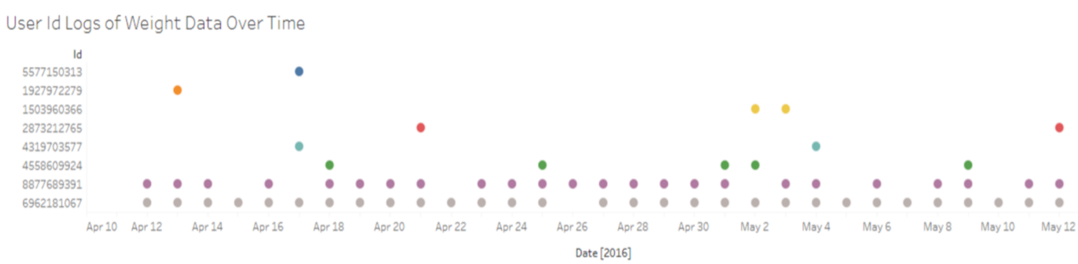

Figure 29: User Id Logs of Weight Data Over Time.

As seen in Figure 30, it can be observed that the users with more weight logs had a greater percentage change in body weight than those who only logged a few times. As such, this could indicate that the users who track weight are more motivated to lose weight and thereby wants to lose more.

As seen in Figure 31, it can be seen that among the 8 users who did track their weight data, only 2 did so a significant number of times.

As seen in Figure 32, it can be observed that overall, the users are tracking their activity data the most frequently as well as consistently. Furthermore, it can be seen that sleep monitoring is popular but much less consistent. Finally, weight tracking is almost insignificant in comparison to all the other types of activity monitoring. This suggests that users are more interested in physical activity tracking.

4.2 Results from R

4.2.1 How many steps did the respondents take in a day?

Figure 33–34 show the observation for how many steps the respondents took in a day.

As seen in Figure 33, respondents are found to be taking the most steps on Tuesdays and Saturdays, and the least number of steps were found to be taken on Sundays.

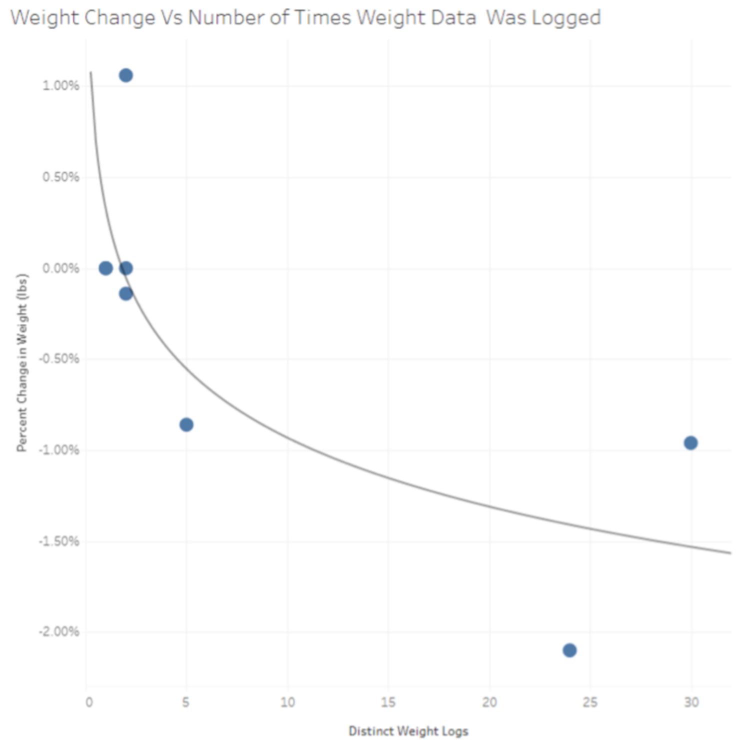

Figure 30: Weight Change Vs. Number of Times Weight Data Was Logged.

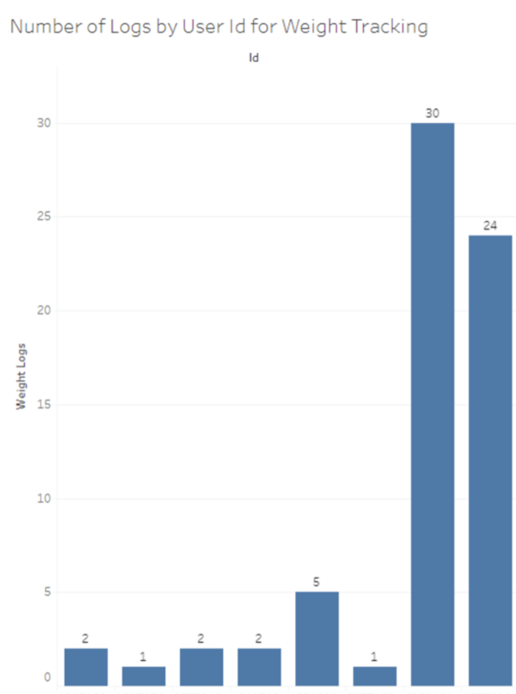

Figure 31: Number of Logs by User Id for Weight Tracking.

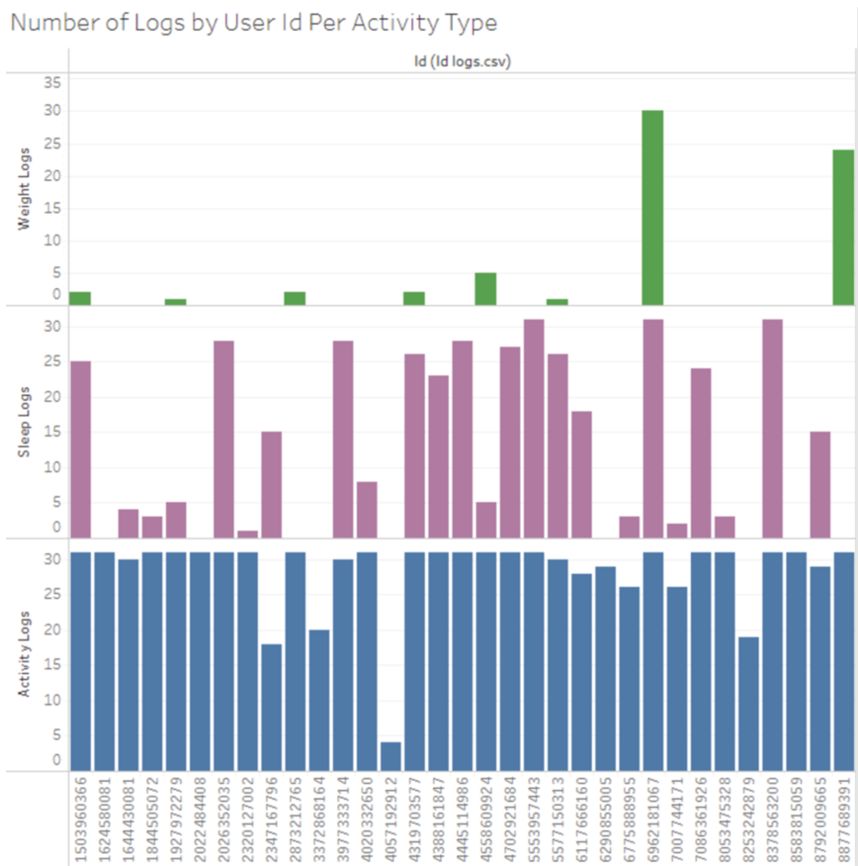

Figure 32: Number of Logs by User Id Per Activity Type.

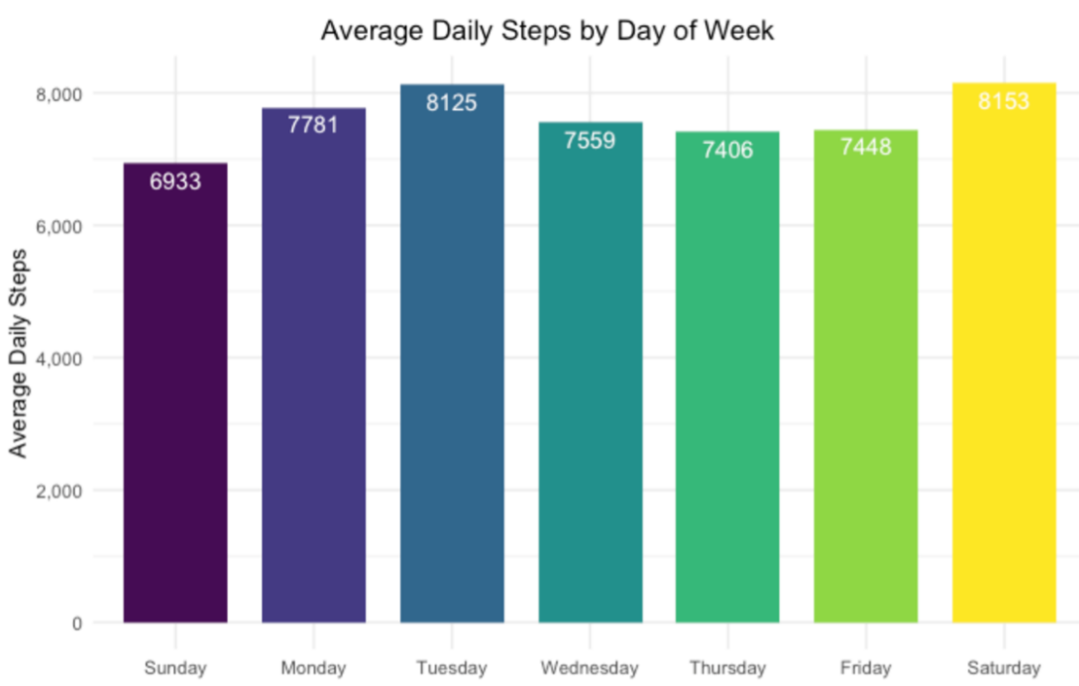

Figure 33: Average Daily Steps by Day of Week.

As seen in Figure 34, the respondents are found to be taking at least four times more steps on normal days and active days as compared to sedentary days.

4.2.2 What were the activity levels of the users?

Figure 35–37 show the observation for the activity levels of the users.

As seen in Figure 35, it can be observed that close to half of all observations were tracked as normal activity. Also, 27% of observations are at a sedentary level and 22.2% are at an active level.

As seen in Figure 36, summarizing by activity level, which days had the most observations.

As seen in Figure 37, it can be observed that sedentary activity levels were mostly recorded on Thursdays, normal activity levels were mostly recorded on Tuesdays and Wednesdays, and active activity levels were recorded mostly on Tuesdays and Saturdays.

4.2.3 How many calories did respondents burn in a day?

Figure 38–40 show the observation for the activity findings.

As seen in Figure 38, it can be observed that the respondents burned an average 2304 calories in a day.

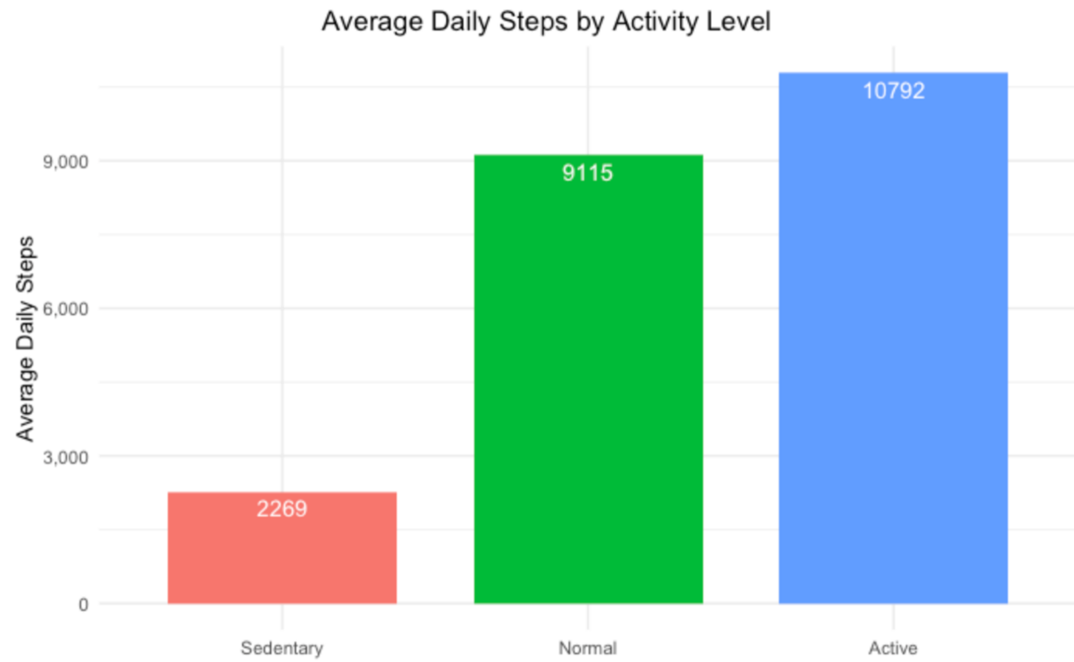

Figure 34: Average Daily Steps by Activity Level.

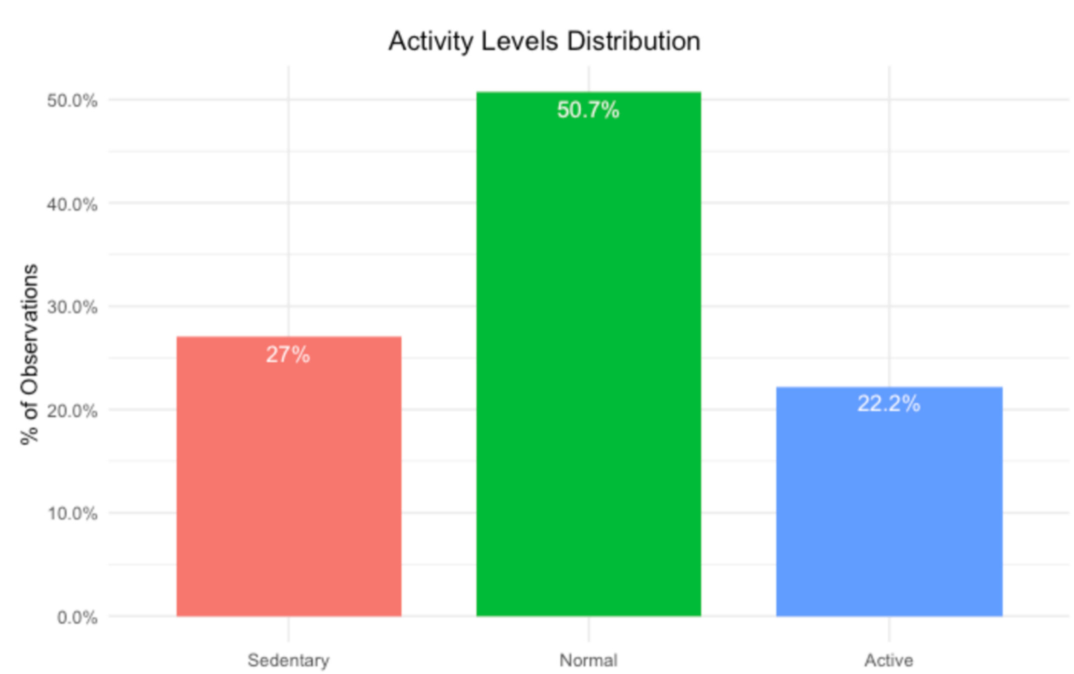

Figure 35: Activity Levels Distribution.

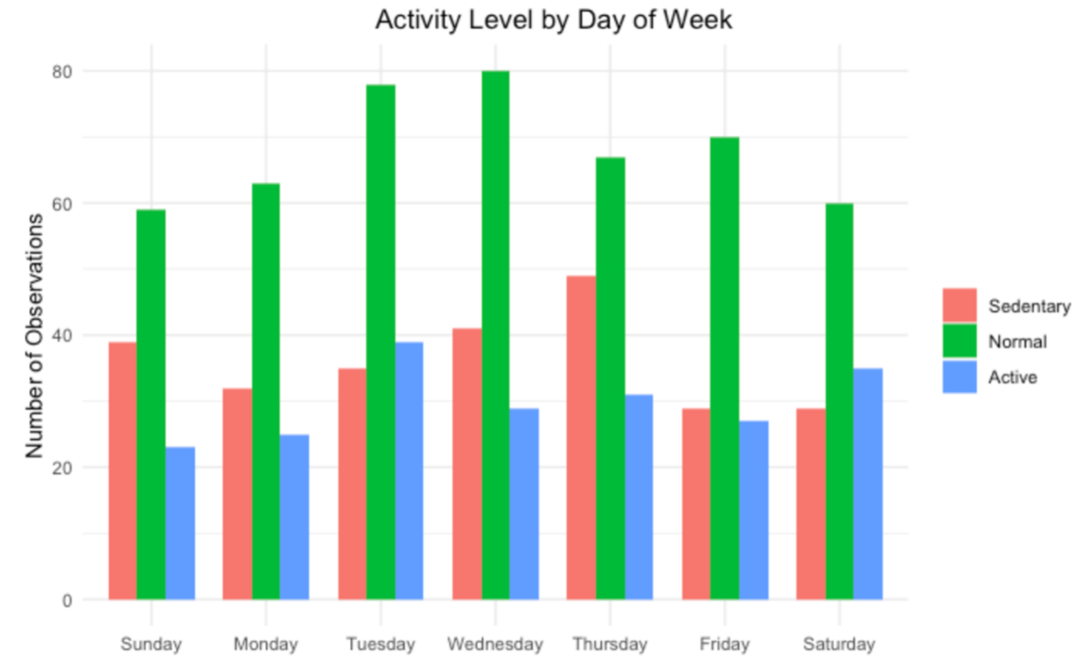

Figure 36: Activity Level by Day of Week.

Activity Level	Days with Most Observations
Sedentary	Thursday
Normal	Tuesday, Wednesday
Active	Tuesday, Saturday

Figure 37: Three different Activity Levels paired with Days with Most Observations.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	1828	2134	2304	2793	4900

Figure 38: The respondents burned an average 2304 calories in a day.

As seen in Figure 39, the respondents burned about 100 calories less that the average 2304 on Thursdays.

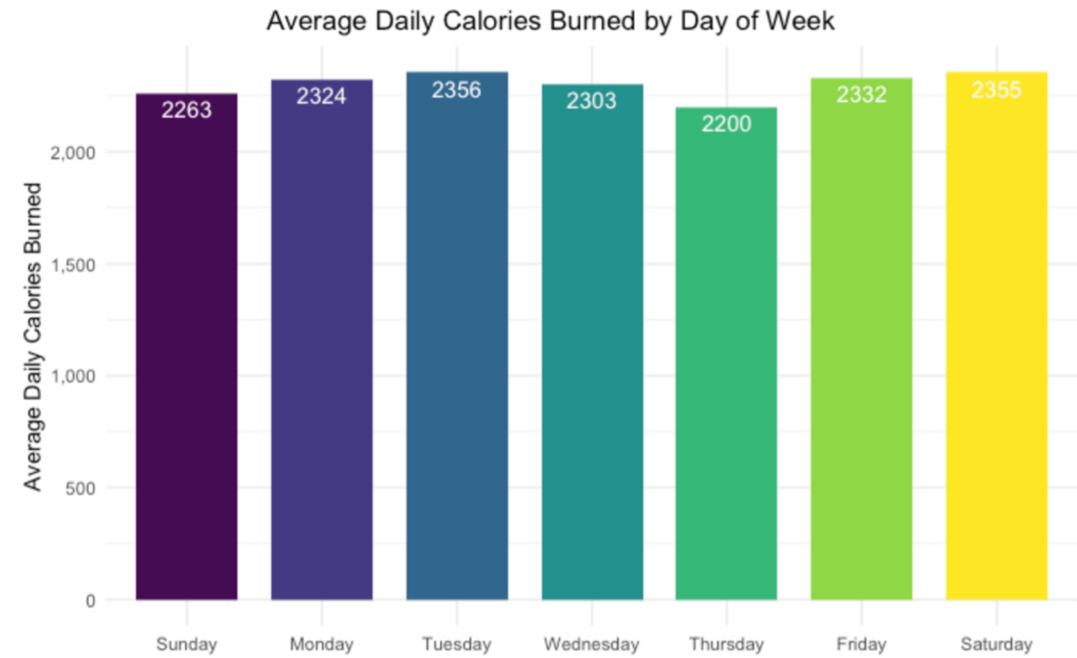

Figure 39: Average Daily Calories Burned by Day of Week.

As seen in Figure 14, the respondents burned at least 500 more calories on normal days and active days as compared to sedentary days.

4.2.4 Is there a relationship between steps taken and calories burned?

Upon testing for association between total steps taken and calories, Pearson's product moment-moment correlation coefficient is found to be 0.59 and this means there is a positive linear trend and that the sampled individuals are scattered around the line of best fit. Upon plotting the regression line between total steps taken and calories:

As seen in Figure 41, a moderate correlation can be seen between taking more steps and burning more calories.

4.2.5 Sleep Day

From Figure 42, the sleep day findings count is being displayed.

4.2.6 How much sleep did respondents get in a day?

Figure 43–46 show the observation for the sleep day findings.

As seen in Figure 43, the respondents tracked an average of close to 7 hours of sleep in a day.

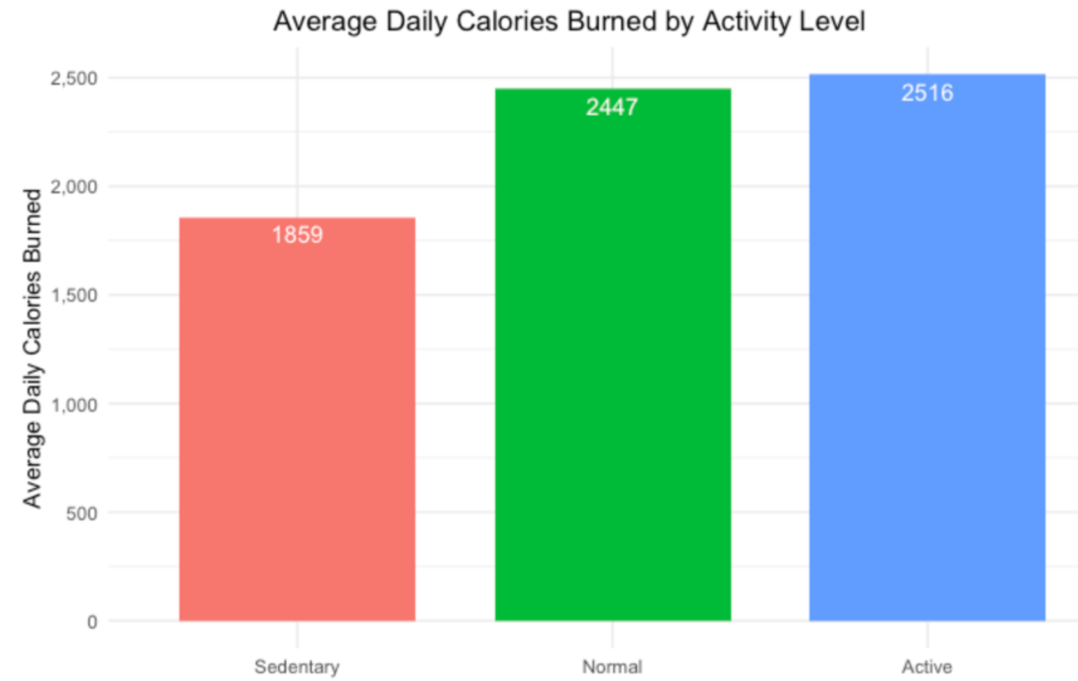

Figure 40: Average Daily Calories Burned by Activity Level.

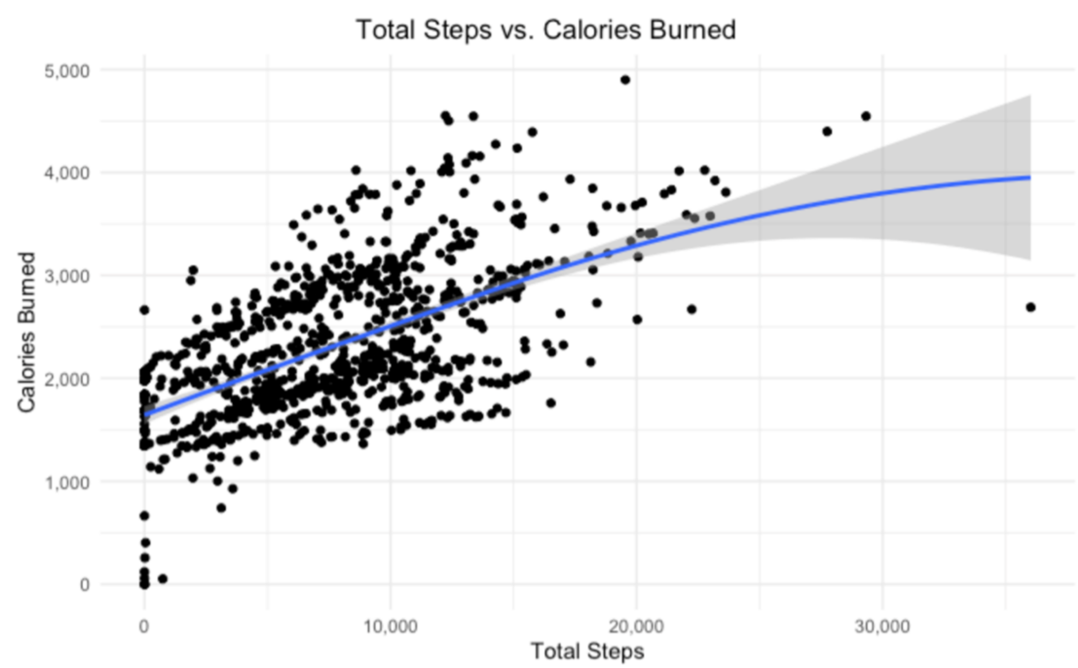

Figure 41: Total Steps Vs. Calories Burned

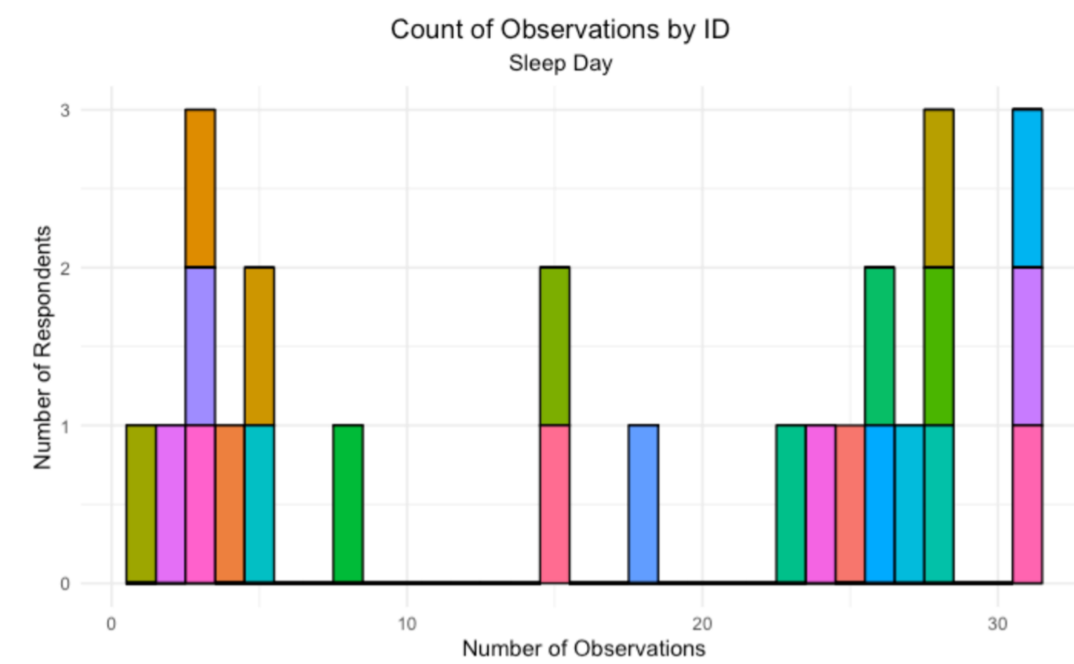

Figure 42: Count of Observations by ID.

```
                    Min.                   1st Qu.                    Median
Mean
                 "58M 0S"              "6H 1M 0S"           "7H 12M 30S" "6H 59M 10.3902439024
387S"
                 3rd Qu.                      Max.
              "8H 10M 0S"            "13H 16M 0S"
```

Figure 43: Average hours of sleep in a day.

As seen in Figure 44, it can be observed that the respondents got the most sleep on Sundays. Also, the respondents got sufficient sleep on Mondays, Wednesdays, and Saturdays. However, they did not sleep more than 7 hours on Tuesdays, Thursdays, and Fridays.

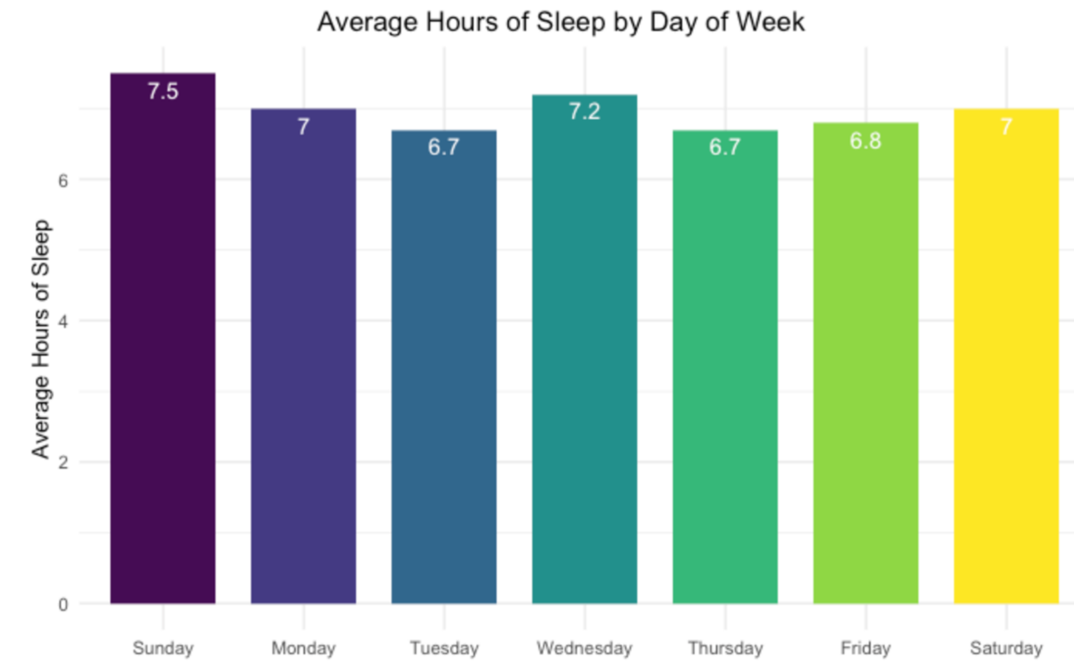

Figure 44: Average Hours of Sleep by Day of Week.

As seen in Figure 45, it can be observed which days of the week, the users got recommended amount of sleep and on which days they did not.

As seen in Figure 46, it can be seen that the respondents who tracked sufficient sleep, got 3 hours more sleep a day then respondents who did not get enough sleep.

4.2.7 Did the respondents get enough sleep?

As seen in Figure 47, it is found that the respondents tracked getting enough sleep for 56 percent of observations.

4.2.8 How efficiently were respondents sleeping?

Figure 48–50 show the observation for how efficiently the respondents were sleeping.

As seen in Figure 48, the respondents tracked an average of 92

As seen in Figure 49, It can be observed that the respondents tracked 91 percent to 92 percent sleep efficiency throughout the week.

As seen in Figure 50, it can be seen that the respondents who got enough sleep, tracked 5 percent more sleep efficiency.

Day of Week	Recommended Sleep
Sunday	Sufficient
Monday	Sufficient
Tuesday	Short
Wednesday	Sufficient
Thursday	Short
Friday	Short
Saturday	Sufficient

Figure 45: Day of Week and Recommended Sleep record.

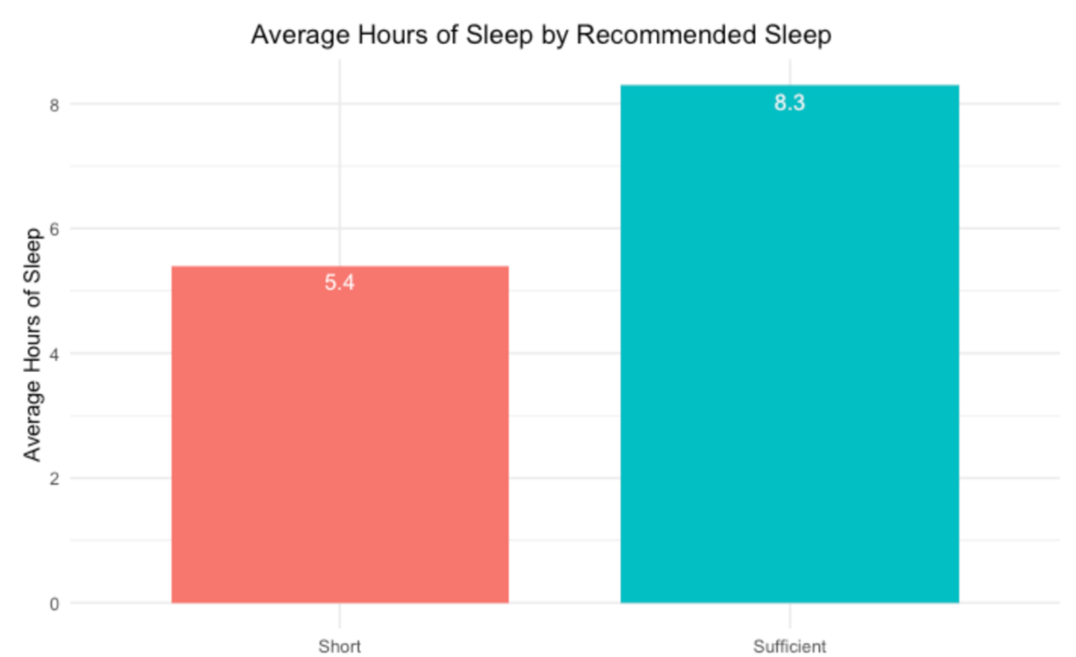

Figure 46: Average Hours of Sleep by Recommended Sleep.

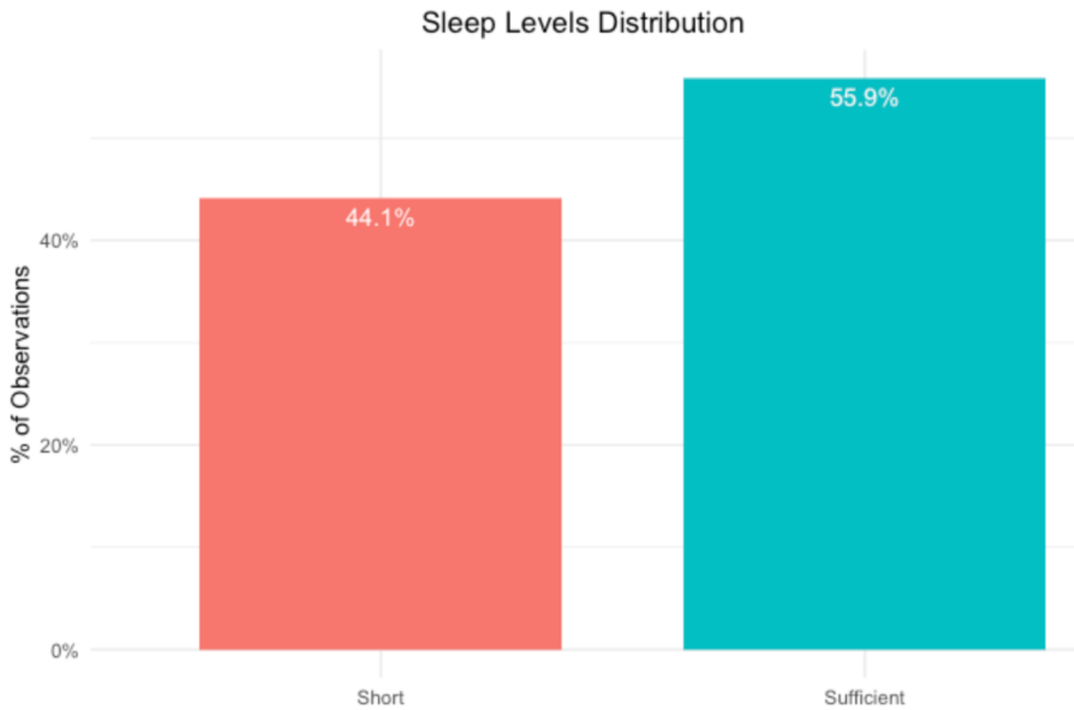

Figure 47: Observation:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.4984	0.9118	0.9426	0.9165	0.9606	1.0000

Figure 48: Average sleep efficiency of the respondents.

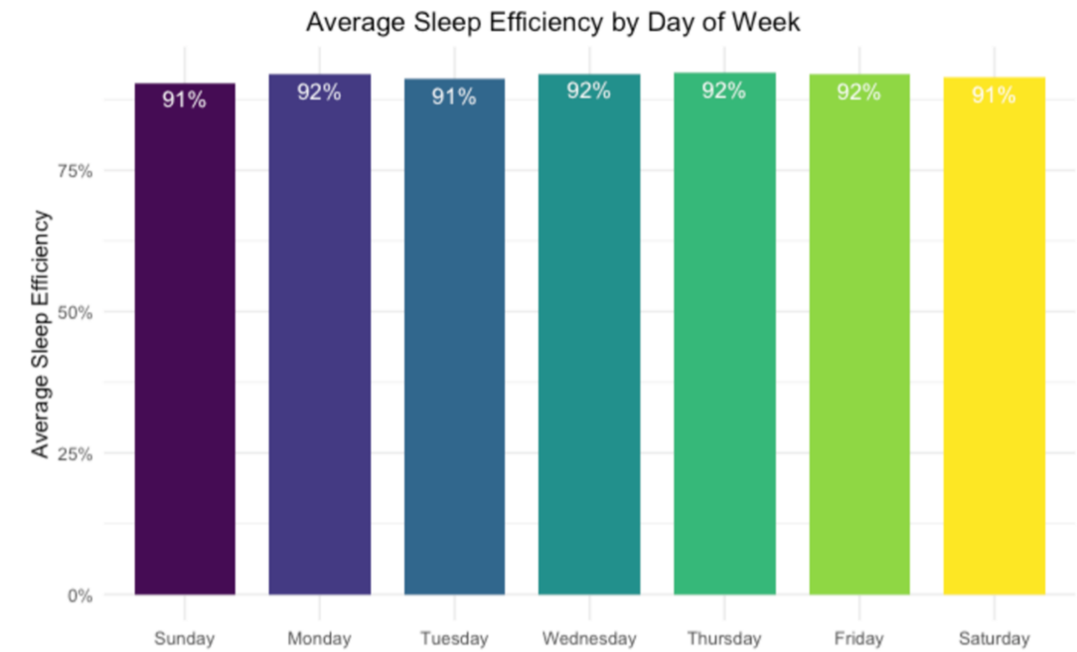

Figure 49: Average Sleep Efficiency by Day of Week.

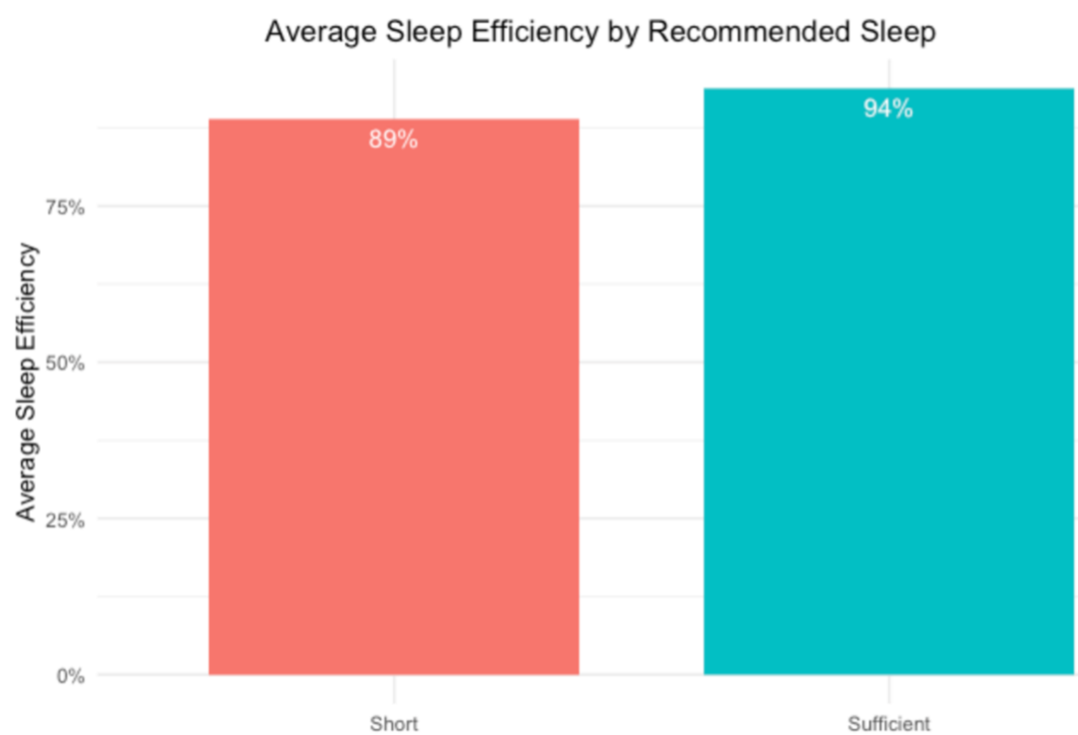

Figure 50: Average Sleep Efficiency by Recommended Sleep.

4.2.9 Is there a relationship between total time in bed and total time asleep?

In this case, the correlation coefficient is found to be 0.93 and this means that there is indeed a positive linear trend and that the sampled individuals are scattered around the line of best fit. Upon plotting a regression line between time in bed and total time asleep:

As seen in Figure 51, it can be observed that there is a strong correlation between total time in bed and total time asleep.

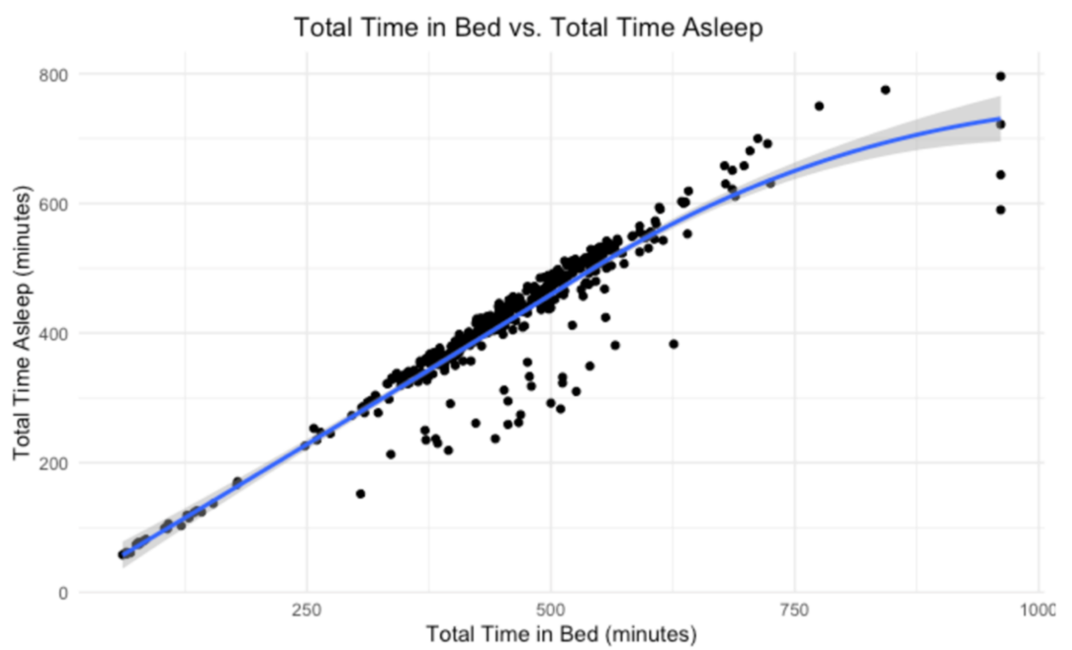

Figure 51: Totsal Time in Bed Vs. Total Time Asleep.

4.2.10 Combined data

4.2.11 How is activity level related to sleep time?

As seen in Figure 52, it can be seen that the respondents got close to 30 minutes more sleep on sedentary days as compared to normal days.

4.2.12 How is activity level related to sleep efficiency?

As seen in Figure 53, it can be seen that the respondents tracked 91 percent to 93% sleep efficiency.

4.2.13 How is getting enough sleep related to total steps taken in a day?

As seen in Figure 54, it can be seen that the respondents took close to 1500 more steps on days when they did not get enough sleep.

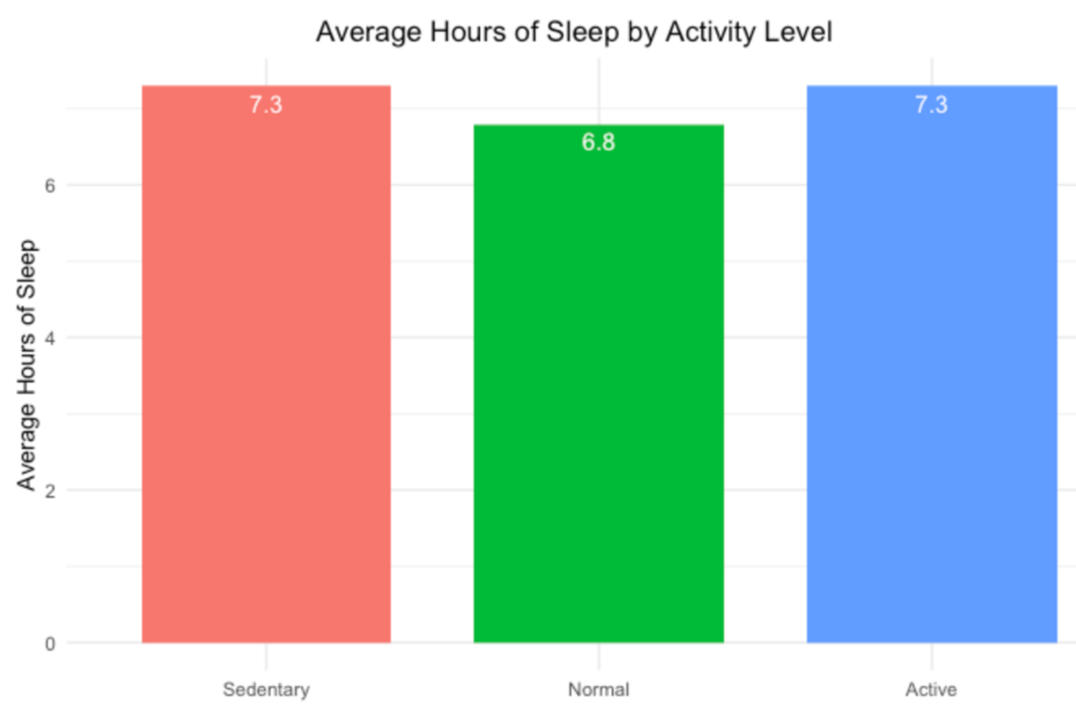

Figure 52: Average Hours of Sleep by Activity Level.

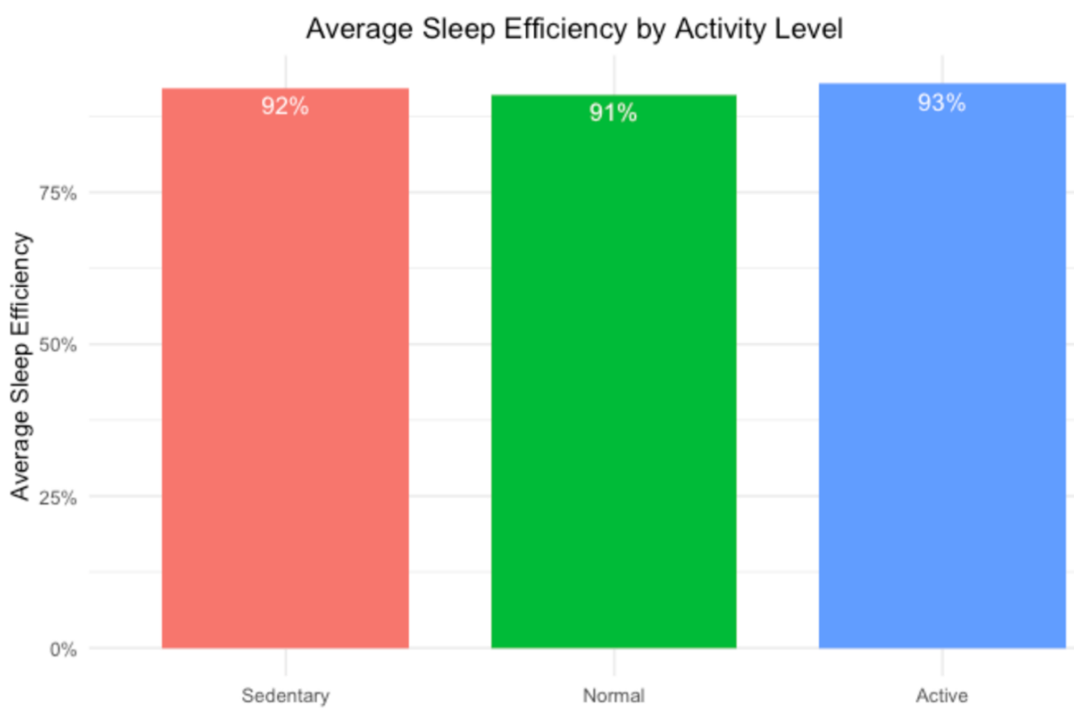

Figure 53: Average Sleep Efficiency by Activity Level.

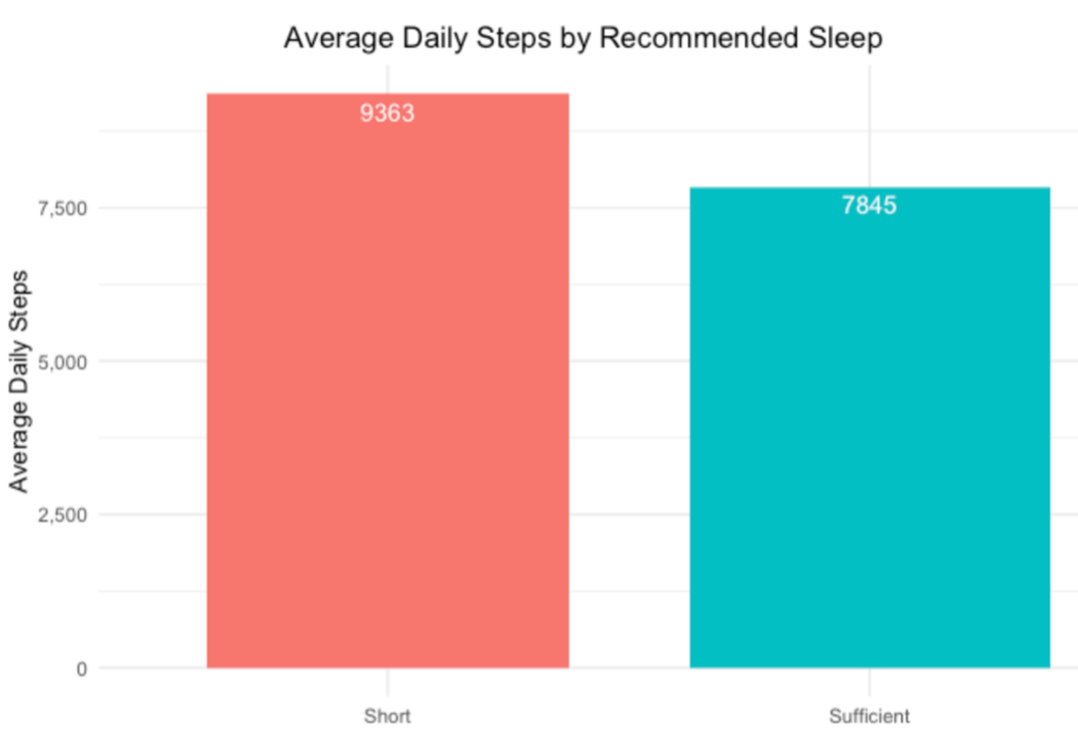

Figure 54: Average Daily Steps by Recommended Sleep.

4.2.14 How is getting enough sleep related to calories burned?

As seen in Figure 55, it can be observed that the respondents burned close to 44 more calories on days when they did not get enough sleep.

4.2.15 Is there a relationship between total steps taken in a day and total minutes asleep?

In this case, the correlation coefficient is found to be -0.19 and this means that there is a weak negative correlation between the total steps and total minutes asleep.

4.2.16 Is there a relationship between calories burned in a day and total minutes asleep?

In this case the correlation coefficient is found to be -0.03 and this means there is no correlation between calories burned and total minutes asleep.

4.2.17 Main Findings

Figure 56–61 show the observation for the main findings.

As seen in Figure 56, the respondents took at least four times more steps on normal days than sedentary days.

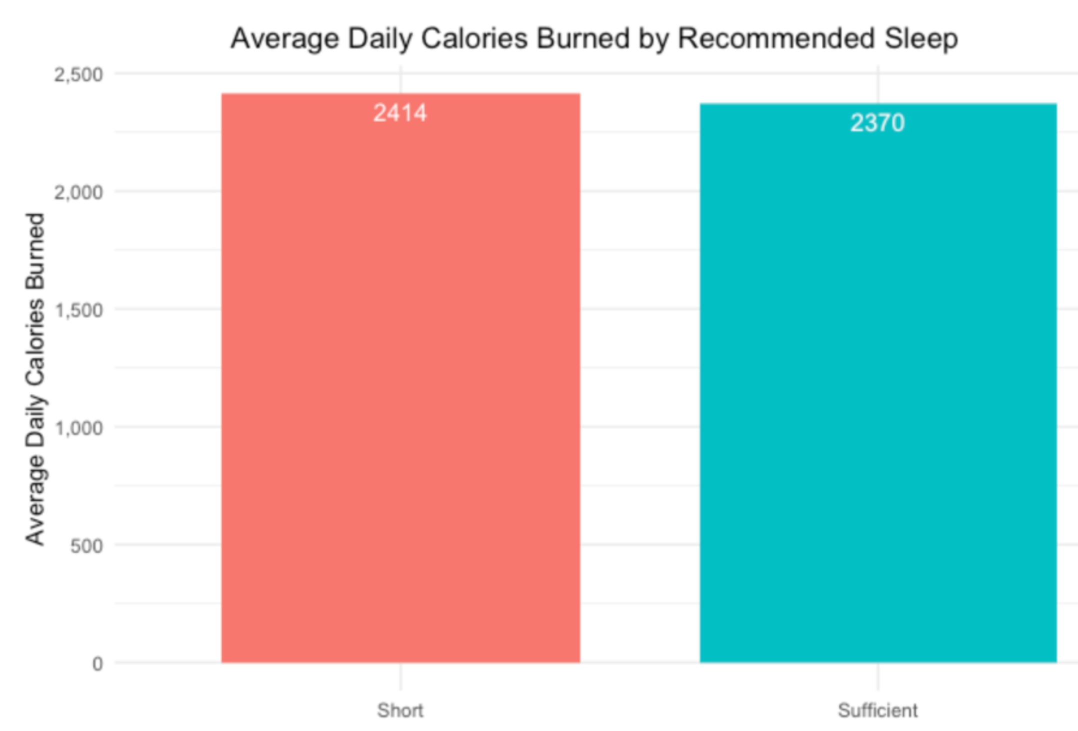

Figure 55: Average Daily Calories Burned by Recommended Sleep.

As seen in Figure 57, the respondents burned at least 500 more calories on normal days then sedentary days.

As seen in Figure 58, the respondents got about 30 minutes more sleep on sedentary days than normal days.

As seen in Figure 59, the respondents took the most steps on Tuesdays and Saturdays, and the least steps on Sundays.

As seen in Figure 60, respondents burned about 100 calories less than average of 2304 on Thursdays.

As seen in Figure 61, the respondents got the most sleep on Sundays. Respondents got sufficient sleep on Mondays, Wednesdays, and Saturdays but did not sleep more than 7 hours on Tuesdays, Thursdays, and Fridays.

4.3 Summary of Findings

The summary of our findings are listed below:

- Respondents took more steps, burned more calories, and got more sleep on active days. Thus, it can be understood that the fitness tracker may be motivating users to develop healthy habits and increase wellness by providing metrics on their activities and sleep schedule to make better decisions.

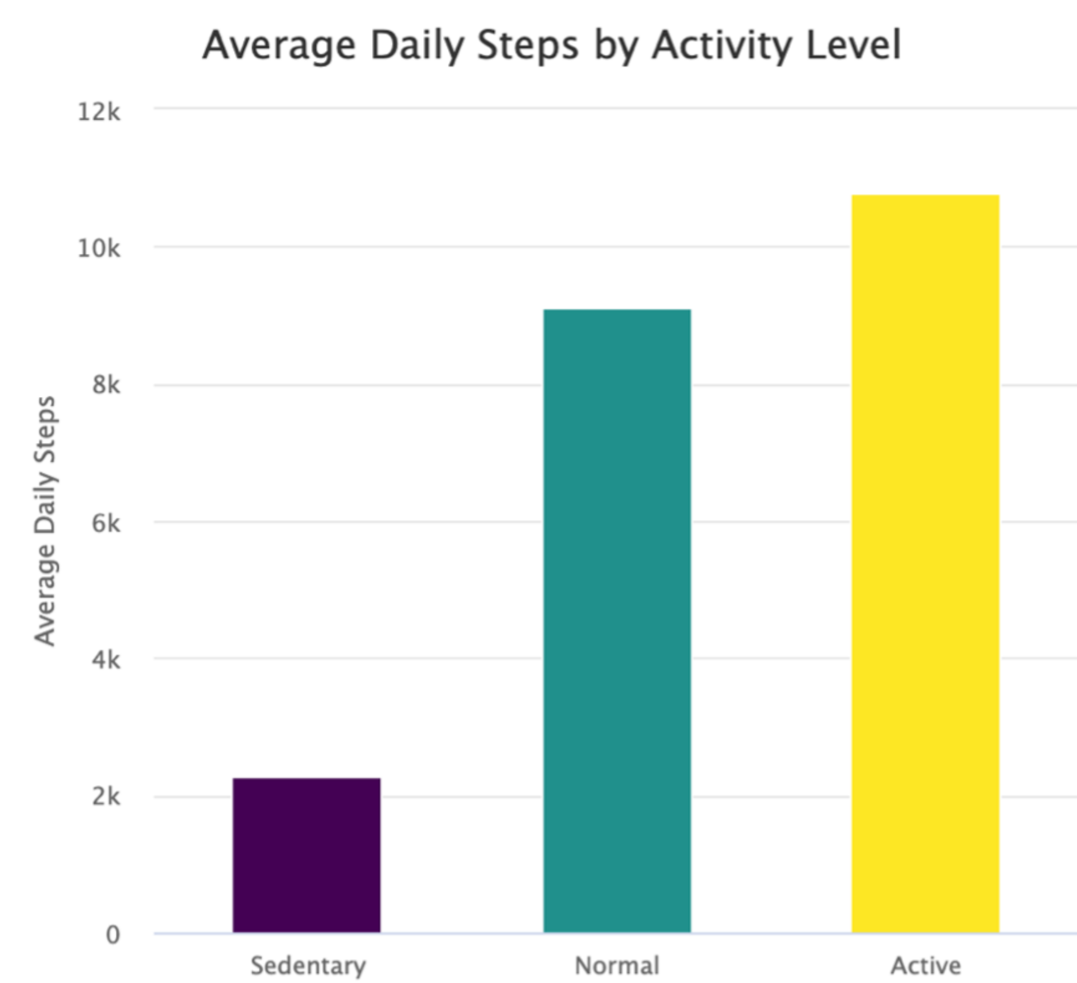

Figure 56: Average Daily Steps by Activity Level.

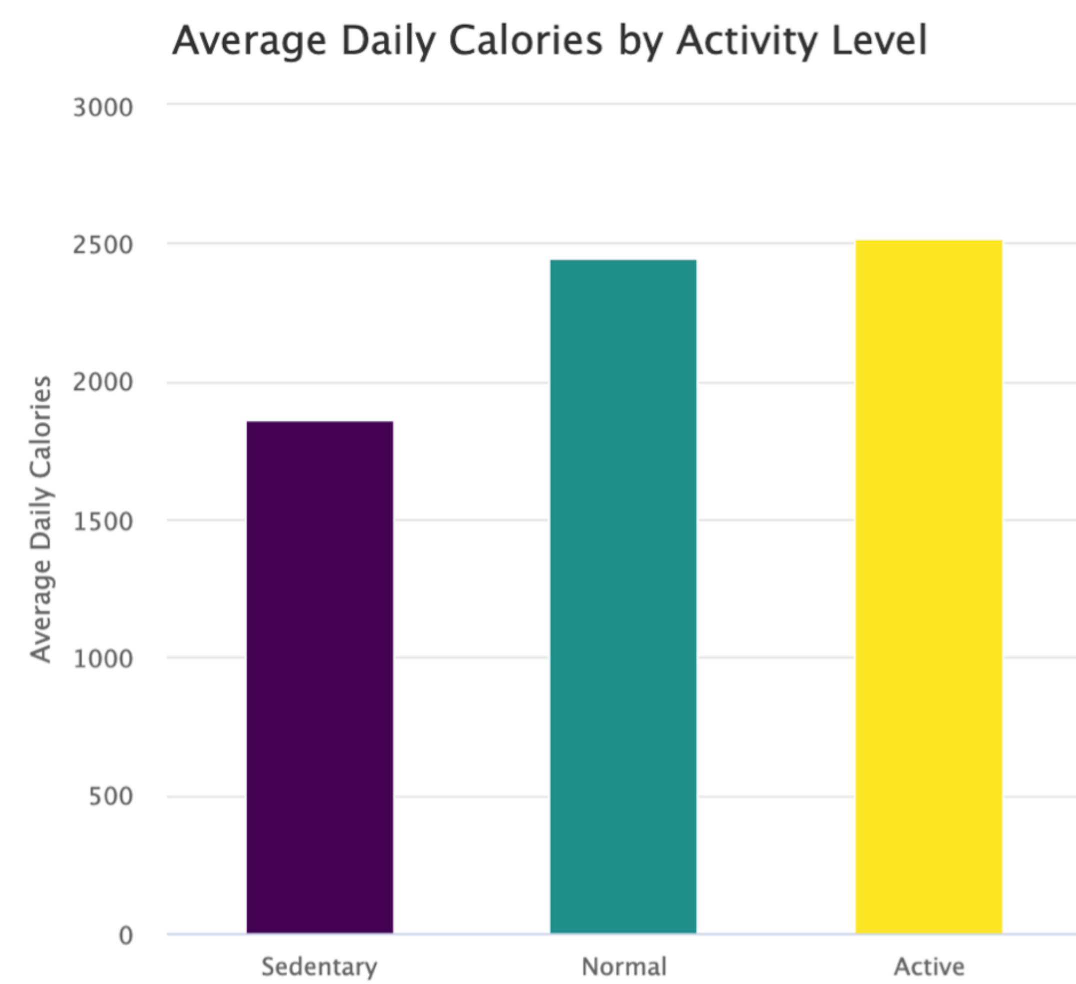

Figure 57: Average Daily Calories by Activity Level.

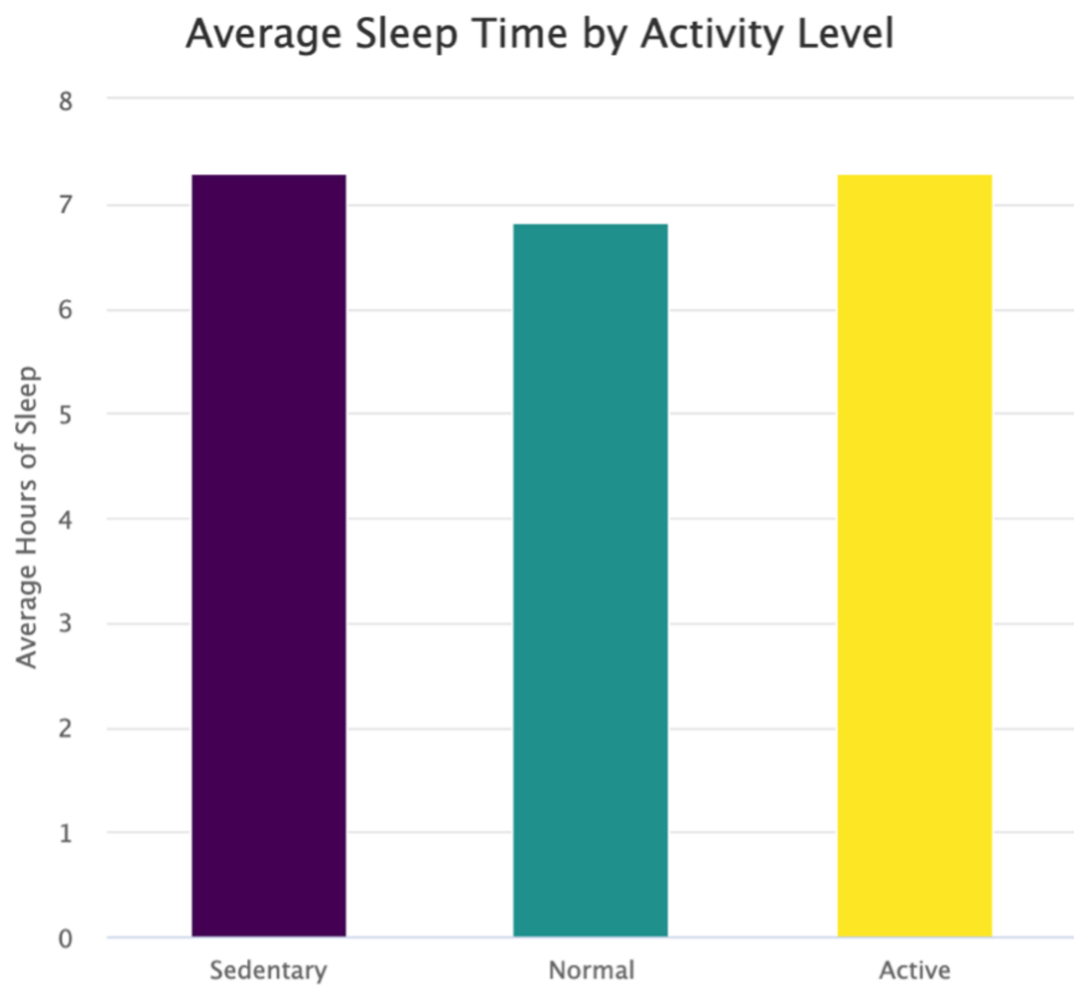

Figure 58: Average Sleep Time by Activity Level.

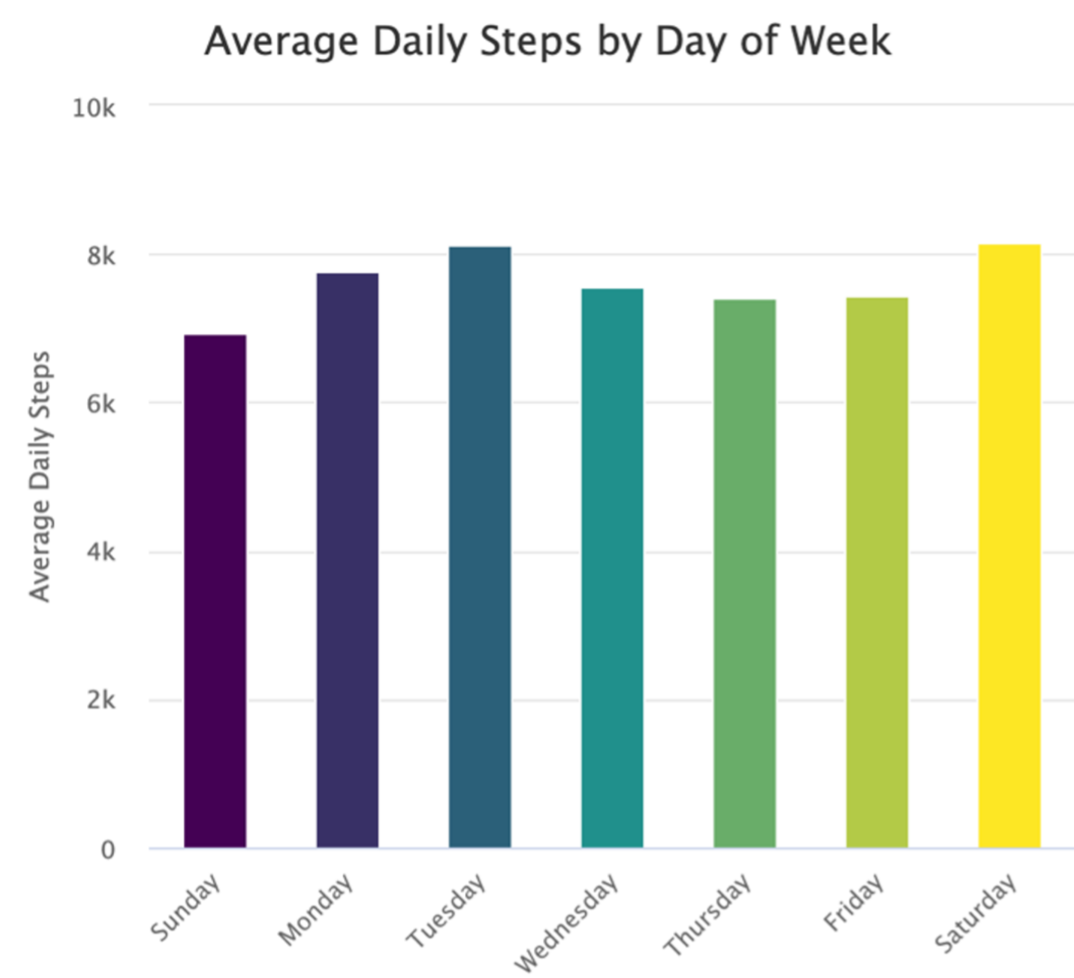

Figure 59: Average Daily Steps by Day of Week.

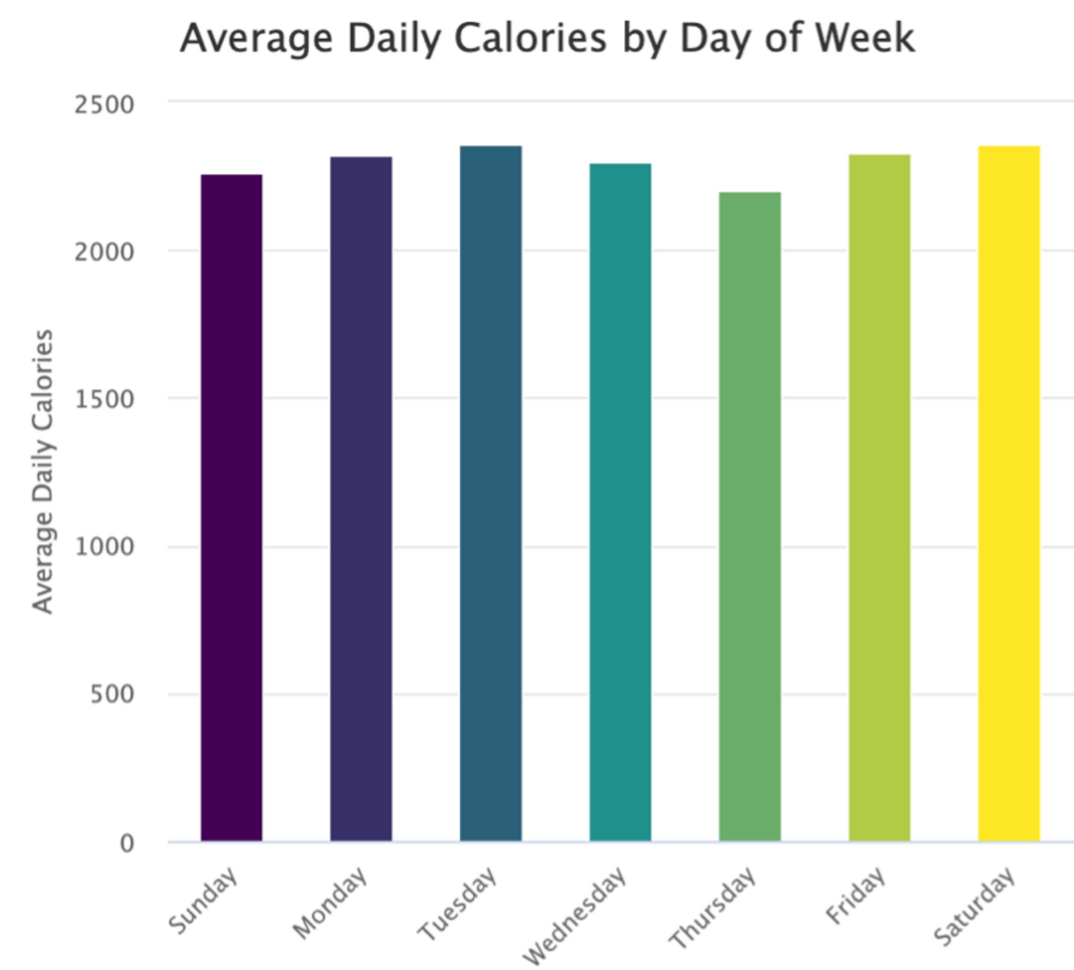

Figure 60: Average Daily Calories by Day of Week.

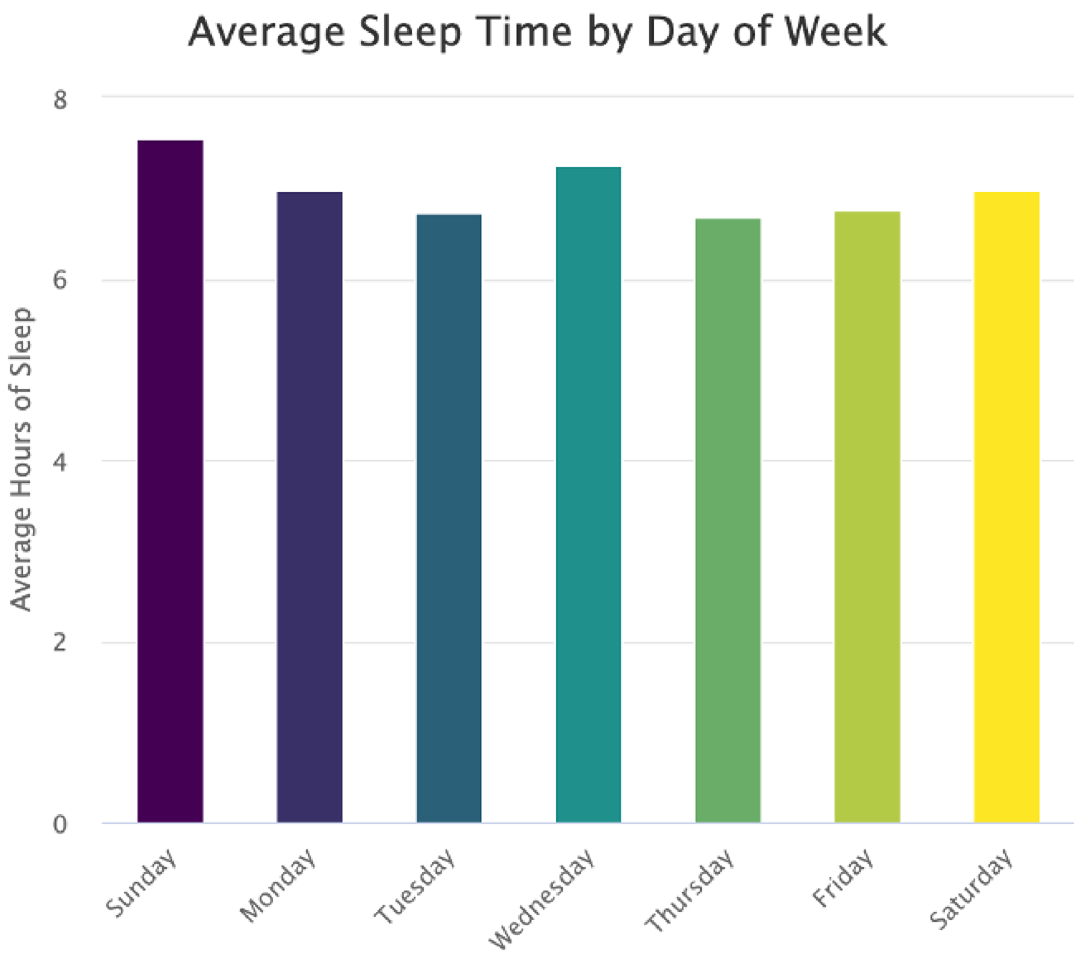

Figure 61: Average Sleep Time by Day of Week.

- It is also found that there is a variation in the number of steps taken, calories burned, and sleep time by the day of week. Although respondents tracked the least steps and most sleep on Sundays, there are differences in activity levels and hours of sleep throughout the week that may be further investigated with additional data.

5 Recommendations

It must be noted that this project has taken the insights and the findings from the data analysis and has applied to the Bellabeat App by highlighting its features, and the improvements in notifications and visualizations.

In particular, the Bellabeat App can benefit the users as below:

- Establish regular activity and sleep habits.
- Measure progress towards achieving wellness goals.
- Improve decision making about health.

Therefore, it can be stated that the Bellabeat App can offer guidance on exercise, sleep, and stress. As an incentive, there may be awards and recognition for achieving wellness goals and sustaining healthy habits.

Based on the summary of the analysis and the findings from the metadata, the following recommendations can be made:

1. To market the features that Bellabeat offers which Fitbit users are already using:
    - Expand on the options to track physical activity with specific exercises tracking and coaching memberships for new workout plans.
    - Users are meeting the recommended time of moderate-vigorous active minutes so they are likely interested to see how they are performing day to day and can set new goals with the Bellabeat App.
    - Maintain active minutes or challenge themselves with the coaching membership.
    - CDC also recommends resistance training which could be taught and monitored with the coaching features.
    - Most users spend the majority of their active minutes performing light level exercise which users could better expand on with the ability to assign specific exercises such as yoga or walking to their daily activity.
2. To market improved wellness by targeting areas of health tracking that the Fitbit data suggests users could use improvement in:

- Meet the recommended 7 hours of sleep with the sleep tracker and monitor their sleep cycles for better sleep health.
- Use the stress meter as a metric to evaluate how their sleep and activity are possibly impacting their mental health.
- Get more than 7,000 steps (to decrease mortality rates) with the option to set up count goals.
- Increase step and activity frequency with the inactivity alarm feature to improve midday decreased activity.
- The coaching membership can be used to round out the midweek plateau of activity displayed in the Fitbit data by being held accountable and/or given work out schedules.

3. To avoid marketing that focuses on weight loss, instead promote the positive and well-rounded health feature: the Wellness Score. This sets Bellabeat apart from its competitors.
    - The data suggests users do not have an interest in weight tracking as it was the least used feature.
    - The wellness score calculates overall health of the body and mind, not just a number on the scale which users may find refreshing and uplifting. It can also help them to stay motivated to track consistently by seeing a snapshot of their performance day to day.
    - A focus on meditation, hydration, and other aspects that make up a healthy life imprints the brand's positive values on consumers.
    - Marketing this feature will expand the ways current fitness tracking consumers can monitor their health but also widen the demographic to less fit consumers as a way to enter the health world in a non-intimidating way.

5.1 Summary

Bellabeat should continue to market their positive and well-rounded health monitoring services. Emphasize how their products, the Ivy and Leaf, can do all of the physical fitness tracking that Fitbit does that consumer like, but with the addition of the wellness score tool that can help them monitor other key aspects (hydration, stress, menstrual cycle) that make up a healthy life in an upbeat and stylish way.

5.2 How these trends influence marketing strategy

Following are the ways, the trends can be influencing Bellabeat's marketing strategy:

- Clearly communicate to the audience through every channel what is Bellabeat wearable technology and how to use the devices to fully reap its benefits. Taylor the communication for different user groups.
- For the user to get a complete picture of their health status through data, being able to record sleep is important. Understanding what are the reasons that prevents users to wear the device during sleep should be one of the priorities and it could be done through focus groups or customer surveys.
- Offer guided goal setting programs where users can set fitness and wellness goals. Help make their data accessible for visualising the progress towards reaching it. Communicate what are the different types of metrics they can keep track of and what are they measuring with those metrics.

6 Conclusion

Over the course of the past few years, a large increase in the available brands of wearable devices can be noted, and more devices are being released with additional sensors. However, for activity tracking, some sensors are more relevant than others are. This case study has focused on analysing public domain fitness tracker data in order to unlock insights into consumer trends and usage patterns of these devices and has presented findings which are most relevant in light of collecting PA data for the purpose of research. However, deciding which wearable to use will depend on several additional factors. The wearable landscape is constantly changing with new devices being released and as new vendors enter or leaves the market or are acquired by larger vendors. What currently are considered relevant devices and brands will therefore change over time, and each research project should be carefully consider which brand and device to use.

7 Limitations and future scope

To begin with, it must be highlighted that the public domain dataset i.e., the metadata used for this case study has its limitations. There is no demographic information on the users, there is no clarity of the Fitbit devices the users wore and if at all they were same devices, there is a variance in the metadata as a result. Also, the way the data was originally captured is not qualitative enough to have analysed and therefore required serious cleaning and manipulation prior to the actual analysis. For the purpose of future work, the metadata can be combined with other relevant available data to uncover deeper insights into trends, and usage patterns of the users which definitely will lead to better recommendation for Bellabeat's product portfolio as well as their marketing team's strategy.

References

[1] Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686.

[2] Garrett Grolemund, Hadley Wickham (2011). Dates and Times Made Easy with lubridate. Journal of Statistical Software, 40(3), 1-25, https://www.jstatsoft.org/v40/i03/.

[3] Kirill Müller (2021). hms: Pretty Time of Day. R package version 1.1.1, https://CRAN.R-project.org/package=hms.

[4] Hadley Wickham and Dana Seidel (2020). scales: Scale Functions for Visualization. R package version 1.1.1, https://CRAN.R-project.org/package=scales.

[5] Sam Firke (2021). janitor: Simple Tools for Examining and Cleaning Dirty Data. R package version 2.1.0, https://CRAN.R-project.org/package=janitor.

[6] Elin Waring, Michael Quinn, Amelia McNamara, Eduardo Arino de la Rubia, Hao Zhu and Shannon Ellis (2021). skimr: Compact and Flexible Summaries of Data. R package version 2.1.3, https://CRAN.R-project.org/package=skimr.

[7] Joshua Kunst (2022). highcharter: A Wrapper for the 'Highcharts' Library. R package version 0.9.4, https://CRAN.R-project.org/package=highcharter.

[8] Bellabeat. "Ivy Features" Bellabeat, https://bellabeat.com/ivy/features/.

[9] Brinton, Julia E, et al. "Establishing Linkages between Distributed Survey Responses and Consumer Wearable Device Datasets: A Pilot Protocol." JMIR Research Protocols, vol. 6, no. 4, 2017, https://doi.org/10.2196/resprot.6513.

[10] CDC. "Assessing Your Weight" CDC, September 17, 2020, https://www.cdc.gov/healthyweight/assessing/index.html.

[11] CDC. "How much physical activity do adults need?" CDC, October 7, 2020, https://www.cdc.gov/physicalactivity/basics/adults/index.htm.

[12] CDC. "How Much Sleep Do I Need?" CDC, March 2, 2017, https://www.cdc.gov/sleep/aboutsleep/howmuchsleep.html.

[13] Fitbit. "What should I know about Fitbit sleep stages?" Fitbit, 2022, https://help.fitbit.com/articles/enUS/Helparticle/2163.htm.

[14] Möbius. "Fitbit Fitness Tracker Data." Kaggle, 16 Dec. 2020, https://www.kaggle.com/arashnic/fitbit.

[15] Asimakopoulos, Stavros, Grigorios Asimakopoulos, and Frank Spillers. "Motivation and user engagement in fitness tracking: Heuristics for mobile healthcare wearables." In Informatics, vol. 4, no. 1, p. 5. MDPI, 2017, https://www.mdpi.com/2227-9709/4/1/5.

[16] Michaelis, Jessica R., Michael A. Rupp, James Kozachuk, Baotran Ho, Daniela Zapata-Ocampo, Daniel S. McConnell, and Janan A. Smither. "Describing the user experience of wearable fitness technology through online product reviews." In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Vol. 60, no. 1, page(s): 1073-1077, 2016, https://doi.org/10.1177/1541931213601248.