Google Data Analytics Capstone Project: Bellabeat Case Study: How can a Wellness Technology Company play it smart?

\*\*Introduction\*\*

Welcome to the Bellabeat data analysis case study! In this case study, you will perform many real-world tasks of a junior data analyst. You will imagine you are working for Bellabeat, a high-tech manufacturer of health-focused products for women, and meet different characters and team members. In order to answer the key business questions, you will follow the steps of the data analysis process: ask, prepare, process, analyze, share, and act. Along the way, the Case Study Roadmap tables — including guiding questions and key tasks — will help you stay on the right path. By the end of this lesson, you will have a portfolio-ready case study. Download the packet and reference the details of this case study anytime. Then, when you begin your job hunt, your case study will be a tangible way to demonstrate your knowledge and skills to potential employers.

\*\*Scenario\*\*

You are a junior data analyst working on the marketing analyst team at Bellabeat, a high-tech manufacturer of health-focused products for women. Bellabeat is a successful small company, but they have the potential to become a larger player in the global smart device market. Urška Sršen, co-founder and Chief Creative Officer of Bellabeat believes that analyzing smart device fitness data could help unlock new growth opportunities for the company. You have been asked to focus on one of Bellabeat’s products and analyze smart device data to gain insight into how consumers are using their smart devices. The insights you discover will then help guide the marketing strategy for the company. You will present your analysis to the Bellabeat executive team along with your high-level recommendations for Bellabeat’s marketing strategy.

\*\* Characters and Products\*\*

- **Characters** ○ Urška Sršen: Bellabeat’s cofounder and Chief Creative Officer ○ Sando Mur: Mathematician and Bellabeat’s cofounder; a key member of the Bellabeat executive team ○ Bellabeat marketing analytics team: A team of data analysts responsible for collecting, analyzing, and reporting data that helps guide Bellabeat’s marketing strategy. You joined this team six months ago and have been busy learning about Bellabeat’s mission and business goals — as well as how you, as a junior data analyst, can help Bellabeat achieve them.

- **Products** ○ Bellabeat app: 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.

○ Leaf: Bellabeat’s classic wellness tracker can be worn as a bracelet, necklace, or clip. The Leaf Tracker connects to the Bellabeat app to track activity, sleep, and stress.

○ Time: This wellness watch combines the timeless look of a classic timepiece with smart technology to track user activity, sleep, and stress. The Time watch connects to the Bellabeat app to provide you with insights into your daily wellness.

○ Spring: This is a water bottle that tracks daily water intake using smart technology to ensure that you are appropriately hydrated throughout the day. The Spring bottle connects to the Bellabeat app to track your hydration levels.

○ Bellabeat membership: Bellabeat also offers a subscription-based membership program for users. Membership gives users 24/7 access to fully personalized guidance on nutrition, activity, sleep, health and beauty, and mindfulness-based on their lifestyle and goals.

\*\*Background information about the company\*\*

Urška Sršen and Sando Mur founded Bellabeat, a high-tech company that manufactures health-focused smart products. Sršen used her background as an artist to develop beautifully designed technology that informs and inspires women around the world. Collecting data on activity, sleep, stress, and reproductive health has allowed Bellabeat to empower women with knowledge about their own health and habits. Since it was founded in 2013, Bellabeat has grown rapidly and quickly positioned itself as a tech-driven wellness company for women. By 2016, Bellabeat had opened offices around the world and launched multiple products. Bellabeat products became available through a growing number of online retailers in addition to their own e-commerce channel on their website. The company has invested in traditional advertising media, such as radio, out-of-home billboards, print, and television, but focuses on digital marketing extensively. Bellabeat invests year-round in Google Search, maintaining active Facebook and Instagram pages, and consistently engages consumers on Twitter. Additionally, Bellabeat runs video ads on Youtube and display ads on the Google Display Network to support campaigns around key marketing dates. Sršen knows that an analysis of Bellabeat’s available consumer data would reveal more opportunities for growth. She has asked the marketing analytics team to focus on a Bellabeat product and analyze smart device usage data in order to gain insight into how people are already using their smart devices. Then, using this information, she would like high-level recommendations for how these trends can inform Bellabeat's marketing strategy

\*\*Step 1: ASK\*\*

Business Task: Analyze customer's use of an existing competitor to identify potential opportunities for growth and recommendations for the Bellabeat marketing strategy Questions for Analysis:

1. what are some trends in smart usage" . Users are more than likely to not document their total steps if they have forgotten their devices. Another trend I noticed would that there is a positive trend in the total amount of calories being burned greatly depended on the activity state that the participant was engaging in.

2. How could these trends apply to Bellabeat customers? Making sure to emphasis the various hidden gems of what this membership all entails. More specifically, making sure to elaborate more on the automatic tracking function that the watches offer, long-lasting battery life for moments where users may forget to charge their device the night before, including more inspirational stories where members are at the moment of reaching a milestone achievement of steps taken.

3. How could these trends help influence Bellabeat's marketing strategy? Making sure to include stories where members have set a particular goal and have taken the many milestones/perks offered to them through the usage of this device and being an active member of this team has helped them achieve a goal of, for instance, weight loss. Emphasizing the various perks of being a Bellabeat member and how the sleek design of the device, not to forget the longevity of the battery for the device can all help assist with reaching that momentum milestone they have been trying to reach.

Key Stakeholders: ● · Urška Sršen — Bellabeat’s co founder and Chief Creative Officer ● Sando Mur — Mathematician and Bellabeat’s cofounder; a key member of the Bellabeat executive team ● Bellabeat marketing analytics team — A team of data analysts responsible for collecting, analyzing, and reporting data that helps guide Bellabeat’s marketing strategy.

\*\* STEP 2: PREPARE\*\*

The data for this analysis will come from FitBit Fitness Tracker Data on Kaggle. These 18 datasets were generated by respondents to a distributed survey via Amazon Mechanical Turk between 03.12.2016–05.12.2016. Thirty eligible Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring. Individual reports can be parsed by export session ID (column A) or timestamp (column B). Variation between output represents the use of different types of Fitbit trackers and individual tracking behaviors/preferences. Some limitations for this data exist due to the sample size and the absence of key characteristics of the participants, such as gender, age, location, and lifestyle. For this analysis, the datasets for daily activity, daily calories, daily intensities, daily steps, heart rate by seconds, minute METs, daily sleep, and weight log information, will be used. Because of the largeness of the datasets being used, R Studio was used to prepare, process, and complete this analysis of which the many packages and data visualization features available therein can be used to explore the data.

#Setting up the Environment

install.packages("tidyverse")

install.packages(“googlesheets4”)

install.packages("ggplot2")

install.packages("tidyr")

install.packages("skimr")

install.packages("dplyr")

install.packages("viridisLite")

install.packages(“lubridate”)

install.packages(“httr”)

install.packages(“here”)

install.packages(“janitor”)

#Unloading the packages

library(tidyverse)

library(here)

library(janitor)

library(skimr)

library(dplyr)

library(httr)

library(ggplot2)

library(tidyr)

library(virdisLite)

library(lubridate)

Importing Datasets

First, the CSV files were opened in Excel. Secondly, I formatted and changed the time/date from "custom" to both "time" and "short drive". Thirdly, I rounded the columns with numbers in them to two decimal places making sure to maintain consistency within the data. Lastly, I wrapped up importing the files into R Studio due to massive datasets being explored with the data frames being created with more simplified names.

#Importing the Datasets

daily\_activity <- read\_csv("../input/fitbit/Fitabase Data 4.12.16-5.12.16/dailyActivity\_merged.csv")

daily\_calories <- read\_csv("../input/fitbit/Fitabase Data 4.12.16-5.12.16/dailyCalories\_merged.csv")

daily\_intensities <- read\_csv("../input/fitbit/Fitabase Data 4.12.16-5.12.16/dailyIntensities\_merged.csv")

daily\_steps <- read\_csv("../input/fitbit/Fitabase Data 4.12.16-5.12.16/dailySteps\_merged.csv") minute\_METs <- read\_csv("../input/fitbit/Fitabase Data 4.12.16-5.12.16/minuteMETsNarrow\_merged.csv")

heart\_rate\_sec <- read\_csv("../input/fitbit/Fitabase Data 4.12.16-5.12.16/heartrate\_seconds\_merged.csv")sleep

sleep\_day <- read\_csv("../input/fitbit/Fitabase Data 4.12.16-5.12.16/sleepDay\_merged.csv") weight\_log <- read\_csv("../input/fitbit/Fitabase Data 4.12.16-5.12.16/weightLogInfo\_merged.csv")

STEP 3:PROCESS

Viewing the Data Frames: To ensure the data frames were imported correctly, the head() function is used. The colnames() and glimpse() functions were used to explore the data frames and find common features.

#Viewing the Dataframes

#dailyActivity

head(dailyActivity\_merged)

[# ℹ 12 more variables:

# TotalDistance <dbl>,

# TrackerDistance <dbl>,

# LoggedActivitiesDistance <dbl>,

# VeryActiveDistance <dbl>,

# ModeratelyActiveDistance <dbl>,

# LightActiveDistance <dbl>, …

]

colnames(dailyActivity\_merged)

[1] "Id"

[2] "ActivityDate"

[3] "TotalSteps"

[4] "TotalDistance"

[5] "TrackerDistance"

[6] "LoggedActivitiesDistance"

[7] "VeryActiveDistance"

[8] "ModeratelyActiveDistance"

[9] "LightActiveDistance"

[10] "SedentaryActiveDistance"

[11] "VeryActiveMinutes"

[12] "FairlyActiveMinutes"

[13] "LightlyActiveMinutes"

[14] "SedentaryMinutes"

[15] "Calories"

glimpse(dailyActivity\_merged)

$ Id *<dbl>* 15…

$ ActivityDate *<chr>* "4…

$ TotalSteps *<dbl>* 13…

$ TotalDistance *<dbl>* 8.…

$ TrackerDistance *<dbl>* 8.…

$ LoggedActivitiesDistance *<dbl>* 0,…

$ VeryActiveDistance *<dbl>* 1.…

$ ModeratelyActiveDistance *<dbl>* 0.…

$ LightActiveDistance *<dbl>* 6.…

$ SedentaryActiveDistance *<dbl>* 0,…

$ VeryActiveMinutes *<dbl>* 25…

$ FairlyActiveMinutes *<dbl>* 13…

$ LightlyActiveMinutes *<dbl>* 32…

$ SedentaryMinutes *<dbl>* 72…

$ Calories *<dbl>* 19…

#sleepDay

head(sleepDay\_merged)

# A tibble: 6 × 5

Id SleepDay TotalSleepRecords

*<dbl>* *<chr>* *<dbl>*

1 1.50e9 4/12/20… 1

2 1.50e9 4/13/20… 2

3 1.50e9 4/15/20… 1

4 1.50e9 4/16/20… 2

5 1.50e9 4/17/20… 1

6 1.50e9 4/19/20… 1

# ℹ 2 more variables:

# TotalMinutesAsleep <dbl>,

# TotalTimeInBed <dbl>

colnames(sleepDay\_merged)

[1] "Id"

[2] "SleepDay"

[3] "TotalSleepRecords"

[4] "TotalMinutesAsleep"

[5] "TotalTimeInBed"

glimpse(sleepDay\_merged)

Rows: 413

Columns: 5

$ Id *<dbl>* 15039603…

$ SleepDay *<chr>* "4/12/20…

$ TotalSleepRecords *<dbl>* 1, 2, 1,…

$ TotalMinutesAsleep *<dbl>* 327, 384…

$ TotalTimeInBed *<dbl>* 346, 407…

#daily\_calories

head(dailyCalories\_merged)

*<dbl>* *<chr>* *<dbl>*

1 1503960366 4/12/2016 1985

2 1503960366 4/13/2016 1797

3 1503960366 4/14/2016 1776

4 1503960366 4/15/2016 1745

5 1503960366 4/16/2016 1863

6 1503960366 4/17/2016 1728

colnames(dailyCalories\_merged)

[1] "Id" "ActivityDay"

[3] "Calories"

glimpse(dailyCalories\_merged)

Rows: 940

Columns: 3

$ Id *<dbl>* 1503960366, 150…

$ ActivityDay *<chr>* "4/12/2016", "4…

$ Calories *<dbl>* 1985, 1797, 177…

#daily\_intensities

head(dailyIntensities\_merged)

# A tibble: 6 × 10

Id ActivityDay SedentaryMinutes

*<dbl>* *<chr>* *<dbl>*

1 1.50e9 4/12/2016 728

2 1.50e9 4/13/2016 776

3 1.50e9 4/14/2016 1218

4 1.50e9 4/15/2016 726

5 1.50e9 4/16/2016 773

6 1.50e9 4/17/2016 539

# ℹ 7 more variables:

# LightlyActiveMinutes <dbl>,

# FairlyActiveMinutes <dbl>,

# VeryActiveMinutes <dbl>,

# SedentaryActiveDistance <dbl>,

# LightActiveDistance <dbl>,

# ModeratelyActiveDistance <dbl>, …

colnames(dailyIntensities\_merged)

[1] "Id"

[2] "ActivityDay"

[3] "SedentaryMinutes"

[4] "LightlyActiveMinutes"

[5] "FairlyActiveMinutes"

[6] "VeryActiveMinutes"

[7] "SedentaryActiveDistance"

[8] "LightActiveDistance"

[9] "ModeratelyActiveDistance"

[10] "VeryActiveDistance"

glimpse(dailyIntensities\_merged)

Rows: 940

Columns: 10

$ Id *<dbl>* 15…

$ ActivityDay *<chr>* "4…

$ SedentaryMinutes *<dbl>* 72…

$ LightlyActiveMinutes *<dbl>* 32…

$ FairlyActiveMinutes *<dbl>* 13…

$ VeryActiveMinutes *<dbl>* 25…

$ SedentaryActiveDistance *<dbl>* 0,…

$ LightActiveDistance *<dbl>* 6.…

$ ModeratelyActiveDistance *<dbl>* 0.…

$ VeryActiveDistance *<dbl>* 1.…

#daily\_steps

head(daily\_steps)

# A tibble: 6 × 3

Id ActivityDay StepTotal

*<dbl>* *<chr>* *<dbl>*

1 1503960366 4/12/2016 13162

2 1503960366 4/13/2016 10735

3 1503960366 4/14/2016 10460

4 1503960366 4/15/2016 9762

5 1503960366 4/16/2016 12669

6 1503960366 4/17/2016 9705

colnames(daily\_steps)

[1] "Id" "ActivityDay"

[3] "StepTotal"

glimpse(daily\_steps)

Rows: 940

Columns: 3

$ Id *<dbl>* 1503960366, 150…

$ ActivityDay *<chr>* "4/12/2016", "4…

$ StepTotal *<dbl>* 13162, 10735, 1…

#minuteMETsNarrow

head(minute\_METs)

# A tibble: 6 × 3

Id ActivityMinute METs

*<dbl>* *<chr>* *<dbl>*

1 1503960366 4/12/2016 12:00:0… 10

2 1503960366 4/12/2016 12:01:0… 10

3 1503960366 4/12/2016 12:02:0… 10

4 1503960366 4/12/2016 12:03:0… 10

5 1503960366 4/12/2016 12:04:0… 10

6 1503960366 4/12/2016 12:05:0… 12

colnames(minute\_METs)

[1] "Id" "ActivityMinute"

[3] "METs"

glimpse(minute\_METs)

Rows: 1,325,580

Columns: 3

$ Id *<dbl>* 1503960366, …

$ ActivityMinute *<chr>* "4/12/2016 1…

$ METs *<dbl>* 10, 10, 10, …

#heart\_rate\_sec

head(heart\_rate\_sec)

# A tibble: 6 × 3

Id Time Value

*<dbl>* *<chr>* *<dbl>*

1 2022484408 4/12/2016 7:21:00… 97

2 2022484408 4/12/2016 7:21:05… 102

3 2022484408 4/12/2016 7:21:10… 105

4 2022484408 4/12/2016 7:21:20… 103

5 2022484408 4/12/2016 7:21:25… 101

6 2022484408 4/12/2016 7:22:05… 95

colnames(heart\_rate\_sec)

[1] "Id" "Time" "Value"

glimpse(heart\_rate\_sec)

Rows: 2,483,658

Columns: 3

$ Id *<dbl>* 2022484408, 202248440…

$ Time *<chr>* "4/12/2016 7:21:00 AM…

$ Value *<dbl>* 97, 102, 105, 103, 10…

#sleep\_day

head(sleep\_day)

# A tibble: 6 × 5

Id SleepDay TotalSleepRecords

*<dbl>* *<chr>* *<dbl>*

1 1.50e9 4/12/20… 1

2 1.50e9 4/13/20… 2

3 1.50e9 4/15/20… 1

4 1.50e9 4/16/20… 2

5 1.50e9 4/17/20… 1

6 1.50e9 4/19/20… 1

# ℹ 2 more variables:

# TotalMinutesAsleep <dbl>,

# TotalTimeInBed <dbl>

colnames(sleep\_day)

[1] "Id"

[2] "SleepDay"

[3] "TotalSleepRecords"

[4] "TotalMinutesAsleep"

[5] "TotalTimeInBed"

glimpse(sleep\_day)

Rows: 413

Columns: 5

$ Id *<dbl>* 15039603…

$ SleepDay *<chr>* "4/12/20…

$ TotalSleepRecords *<dbl>* 1, 2, 1,…

$ TotalMinutesAsleep *<dbl>* 327, 384…

$ TotalTimeInBed *<dbl>* 346, 407…

#weight\_log

head(weight\_log)

# A tibble: 6 × 8

Id Date WeightKg WeightPounds

*<dbl>* *<chr>* *<dbl>* *<dbl>*

1 1.50e9 5/2/… 52.6 116.

2 1.50e9 5/3/… 52.6 116.

3 1.93e9 4/13… 134. 294.

4 2.87e9 4/21… 56.7 125.

5 2.87e9 5/12… 57.3 126.

6 4.32e9 4/17… 72.4 160.

# ℹ 4 more variables: Fat <dbl>,

# BMI <dbl>, IsManualReport <lgl>,

# LogId <dbl>

colnames(weight\_log)

[1] "Id" "Date"

[3] "WeightKg" "WeightPounds"

[5] "Fat" "BMI"

[7] "IsManualReport" "LogId"

glimpse(weight\_log)

Rows: 67

Columns: 8

$ Id *<dbl>* 1503960366, …

$ Date *<chr>* "5/2/2016 11…

$ WeightKg *<dbl>* 52.6, 52.6, …

$ WeightPounds *<dbl>* 115.9631, 11…

$ Fat *<dbl>* 22, NA, NA, …

$ BMI *<dbl>* 22.65, 22.65…

$ IsManualReport *<lgl>* TRUE, TRUE, …

$ LogId *<dbl>* 1.462234e+12…

Removing data frames:

All the nine data frames contain the “Id” column, so it is possible to merge all of them (with SQL) if needed. The daily\_activity data frame seems to contain data for calories, intensities, and steps. In order to use the daily\_activity frame in place of daily\_calories, daily\_intensities, and daily\_steps, the number of observations must be the same and the observations must match for each ID number. We will load the sqldf package to utilize SQL syntax to determine if the values of daily\_calories, daily\_intensities, and daily\_steps are contained in daily\_activity. However, the number of columns must be the same between the data frames, so a temporary data frame with the important columns is created first.

#Filtering entries with values greater than 0

(Logging\_occurances <- filter(dailyActivity\_merged,LoggedActivitiesDistance > "0"))

#Summarization of Datasets

dailyActivity\_merged %>% select(TotalSteps,TotalDistance,SedentaryMinutes) %>% summary()

sleepDay\_merged %>% select(TotalSleepRecords, TotalMinutesAsleep, TotalTimeInBed) %>% summary()

TotalSleepRecords

Min. :1.000

1st Qu.:1.000

Median :1.000

Mean :1.119

3rd Qu.:1.000

Max. :3.000

TotalMinutesAsleep TotalTimeInBed

Min. : 58.0 Min. : 61.0

1st Qu.:361.0 1st Qu.:403.0

Median :433.0 Median :463.0

Mean :419.5 Mean :458.6

3rd Qu.:490.0 3rd Qu.:526.0

Max. :796.0 Max. :961.0

weightLogInfo\_merged %>% select(WeightPounds, BMI, Id) %>% summary()

WeightPounds BMI

Min. :116.0 Min. :21.45

1st Qu.:135.4 1st Qu.:23.96

Median :137.8 Median :24.39

Mean :158.8 Mean :25.19

3rd Qu.:187.5 3rd Qu.:25.56

Max. :294.3 Max. :47.54

Id

Min. :1.504e+09

1st Qu.:6.962e+09

Median :6.962e+09

Mean :7.009e+09

3rd Qu.:8.878e+09

Max. :8.878e+09

hourlySteps\_merged %>% select(Id, ActivityHour, StepTotal) %>% summary()

Id

Min. :1.504e+09

1st Qu.:2.320e+09

Median :4.445e+09

Mean :4.848e+09

3rd Qu.:6.962e+09

Max. :8.878e+09

ActivityHour

Length:22099

Class :character

Mode :character

StepTotal

Min. : 0.0

1st Qu.: 0.0

Median : 40.0

Mean : 320.2

3rd Qu.: 357.0

Max. :10554.0

dailyCalories\_merged %>% select(Id, ActivityDay, Calories)

# A tibble: 940 × 3

Id ActivityDay Calories

*<dbl>* *<chr>* *<dbl>*

1 1503960366 4/12/2016 1985

2 1503960366 4/13/2016 1797

3 1503960366 4/14/2016 1776

4 1503960366 4/15/2016 1745

5 1503960366 4/16/2016 1863

6 1503960366 4/17/2016 1728

7 1503960366 4/18/2016 1921

8 1503960366 4/19/2016 2035

9 1503960366 4/20/2016 1786

10 1503960366 4/21/2016 1775

# ℹ 930 more rows

# ℹ Use `print(n = ...)` to see more rows

The outputs of the head() function of the temporary data frames created, match the outputs of the head() function for the original data frames. The outputs of the head() function of the SQL data frames match the outputs of the head() function for the temporary data frames. The number of observations for each SQL data frame are equal to 940. Conclusively, the data for the daily\_calories, daily\_intensities, and daily\_steps data frames are contained in daily\_activity. These three data frames will be removed from the analysis for simplicity.

STEP 4: ANALYZE

Summarization of the data: The n\_distinct() and nrow() functions are used to determine the number of unique values and the number of rows in a data frame.

n\_distinct(dailyActivity\_merged $ Id)

n\_distinct(dailyCalories\_merged $ Id)

n\_distinct(dailySteps\_merged $ Id)

n\_distinct(heartrate\_seconds\_merged $ Id)

n\_distinct(minuteSleep\_merged $ Id)

n\_distinct(sleepDay\_merged $ Id)

n\_distinct(weightLogInfo\_merged $ Id)

n\_distinct(hourlySteps\_merged $ Id)

n\_distinct(dailyIntensities\_merged$Id)

> n\_distinct(dailyActivity\_merged $ Id)

[1] 33

> n\_distinct(dailyCalories\_merged $ Id)

[1] 33

> n\_distinct(dailySteps\_merged $ Id)

[1] 33

> n\_distinct(heartrate\_seconds\_merged $ Id)

[1] 14

> n\_distinct(minuteSleep\_merged $ Id)

[1] 24

> n\_distinct(sleepDay\_merged $ Id)

[1] 24

> n\_distinct(weightLogInfo\_merged $ Id)

[1] 8

> n\_distinct(hourlySteps\_merged $ Id)

[1] 33

> n\_distinct(dailyIntensities\_merged$Id)

[1] 33

nrow(daily\_activity) nrow(minute\_METs) nrow(heart\_rate\_sec) nrow(sleep\_day) nrow(weight\_log)

The heart rate and weight log data frames contain a very low number of participants based on the n\_distinct() outputs. Thus, reliable recommendations and conclusions cannot be made solely from these data frames. The summary() function is used to pull key statistics about the data frames.

by\_id <- group\_by(heartrate\_seconds\_merged, Id)

heart\_rate\_summary\_by\_id <- summarise(by\_id, count = n(), mean = mean(Value, na.rm = TRUE), min = min(Value, na.rm = TRUE), max = max(Value, na.rm = TRUE))

print(heart\_rate\_summary\_by\_id)

# A tibble: 14 × 5

Id count mean min max

*<dbl>* *<int>* *<dbl>* *<dbl>* *<dbl>*

1 2.02e9 154104 80.2 38 203

2 2.03e9 2490 93.8 63 125

3 2.35e9 152683 76.7 49 195

4 4.02e9 285461 82.3 46 191

5 4.39e9 249748 66.1 39 180

6 4.56e9 192168 81.7 44 199

7 5.55e9 255174 68.6 47 165

8 5.58e9 248560 69.6 36 174

9 6.12e9 158899 83.7 52 189

10 6.78e9 32771 92.0 55 177

11 6.96e9 266326 77.7 47 184

12 7.01e9 133592 91.1 54 166

13 8.79e9 122841 72.5 43 158

14 8.88e9 228841 83.6 46 180

This summary portends that the average user takes 7638 steps per day, not up to the recommended 10,000 steps for health by the Centre for Disease Control (CDC). On average, users get 21.16 minutes of very active or vigorous activity per day, equating to 148.12 minutes per week. The CDC recommends 75 minutes of vigorous activity per week, so the typical Fitbit user is doing well in this area and achieving additional health benefits. In contrast, participants are averaging 991.2 minutes, or 16.52 hours of sedentary time a day! This is a significant amount of time and can lead to other health issues because the body functions best upright. Scientists have determined that 40 minutes of moderate to vigorous activity a day will balance out the effects of sitting up to 10 hours a day. In addition, this summary shows the average user is burning 2304 calories per day. Researches show the average person in the population burns 1800 calories a day, but burning 3500 is needed to lose a pound of weight. The Fitbit users in this case are burning more than the norm, therefore, they are on track to lose a few pounds per week if they so choose.

#heart\_rate\_summary\_by\_id <- summarise(by\_id, count = n(), mean = mean(Value, na.rm = TRUE), min = min(Value, na.rm = TRUE), max = max(Value, na.rm = TRUE))

# A tibble: 14 × 5

Id count mean min max

*<dbl>* *<int>* *<dbl>* *<dbl>* *<dbl>*

1 2.02e9 154104 80.2 38 203

2 2.03e9 2490 93.8 63 125

3 2.35e9 152683 76.7 49 195

4 4.02e9 285461 82.3 46 191

5 4.39e9 249748 66.1 39 180

6 4.56e9 192168 81.7 44 199

7 5.55e9 255174 68.6 47 165

8 5.58e9 248560 69.6 36 174

9 6.12e9 158899 83.7 52 189

10 6.78e9 32771 92.0 55 177

11 6.96e9 266326 77.7 47 184

12 7.01e9 133592 91.1 54 166

13 8.79e9 122841 72.5 43 158

14 8.88e9 228841 83.6 46 180

Despite the low number of users in the heart rate data frame, the average heart rate of 77 beats per minute (bpm) fits within the “normal” range. The range between 50 to 80 bpm for men, and 53 to 82 bpm for women are considered to be Normal. However, research suggests that it is more important for individuals to determine what is a normal and healthy heartrate for them, and not compare to population levels. This is because resting heart rating between different people can vary by as much as 70 bpm. Changes in resting heartrate over days can be a sign of infection, menstrual cycle effects, body chemical composition, or other acute triggers. Thus, making heartrate a vital health characteristic to monitor.

#minute\_METs minute\_METs %>% select(METs) %>% summary()

# A tibble: 1,325,580 × 3

Id ActivityMinute METs

*<dbl>* *<chr>* *<dbl>*

1 1503960366 4/12/2016 12:00:00 AM 10

2 1503960366 4/12/2016 12:01:00 AM 10

3 1503960366 4/12/2016 12:02:00 AM 10

4 1503960366 4/12/2016 12:03:00 AM 10

5 1503960366 4/12/2016 12:04:00 AM 10

6 1503960366 4/12/2016 12:05:00 AM 12

7 1503960366 4/12/2016 12:06:00 AM 12

8 1503960366 4/12/2016 12:07:00 AM 12

9 1503960366 4/12/2016 12:08:00 AM 12

10 1503960366 4/12/2016 12:09:00 AM 12

# ℹ 1,325,570 more rows

# ℹ Use `print(n = ...)` to see more rows

The summary of minute METs shows the average user has a MET of 14.47. A MET is the division between your working metabolic rate and your resting metabolic rate. One MET is the energy your body consumes when at rest. This means an activity with a MET of four, would require a person to exert four times the energy they do when they are sitting. Therefore, a user averaging 14.47 MET throughout the day is considerably high, which leads to the assumption that the Fitbit is not calculating this data point correctly. Due to this, the minute MET data frame will no longer be used in this analysis.

sleep\_day %>% select(TotalSleepRecords, TotalMinutesAsleep, TotalTimeInBed) %>% summary()

TotalSleepRecords TotalMinutesAsleep

Min. :1.000 Min. : 58.0

1st Qu.:1.000 1st Qu.:361.0

Median :1.000 Median :433.0

Mean :1.119 Mean :419.5

3rd Qu.:1.000 3rd Qu.:490.0

Max. :3.000 Max. :796.0

TotalTimeInBed

Min. : 61.0

1st Qu.:403.0

Median :463.0

Mean :458.6

3rd Qu.:526.0

Max. :961.0

The summary of the sleep data frame displays the average user sleeps once per day for 419.5 minutes, or roughly 7 hours. This falls within the CDC’s recommendations for adults in order to get the proper amount of rest. The average participant is spending 458.6 minutes in bed, or 7.64 hours. This means the typical user is spending 38.6 minutes awake in bed. According to Health Central, people should not spend more than 1 hour in bed awake. This is to prevent a mental link from being formed between being awake and being in bed, which can lead to insomnia.

weight\_log %>% select(WeightPounds, BMI) %>% summary()

WeightPounds BMI

Min. :116.0 Min. :21.45

1st Qu.:135.4 1st Qu.:23.96

Median :137.8 Median :24.39

Mean :158.8 Mean :25.19

3rd Qu.:187.5 3rd Qu.:25.56

Max. :294.3 Max. :47.54

Id

Min. :1.504e+09

1st Qu.:6.962e+09

Median :6.962e+09

Mean :7.009e+09

3rd Qu.:8.878e+09

Max. :8.878e+09

While this data frame has a low number of participants, the average BMI is 25.19. This is considered an overweight BMI. However, BMI can be a screening tool and does not diagnose the body fatness or health of an individual.

STEP 5:SHARE

The ggplot() function of R Studio was used to create data visualizations that depict patterns and trends found in the data frames, which can give us further insights for this project.

no\_steps <- combine\_data\_completed %>%

select(total\_steps, sedentary\_minutes,day\_of\_the\_week, logged\_kms )%>%

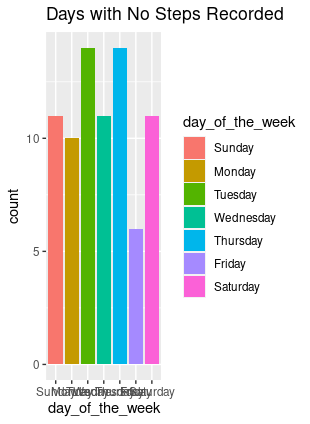
group\_by(day\_of\_the\_week)%>%

filter(total\_steps == 0)

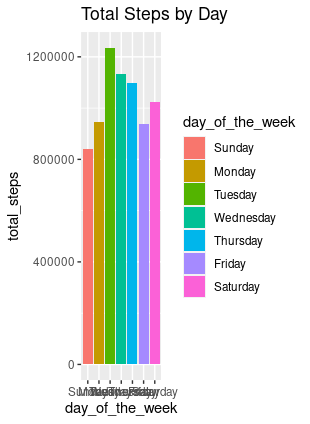
ggplot(data = no\_steps, mapping = aes(x= day\_of\_the\_week, fill = day\_of\_the\_week))+

geom\_bar()+

labs(title = "Days with No Steps Recorded")

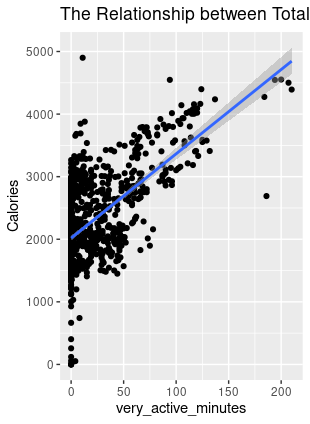


This graph shows the days of the week where no steps were recorded. Friday stands out the most with the number of steps barely making the half way metric. This can be for a number of reasons why participants are not recording their step for this particular day of the week.



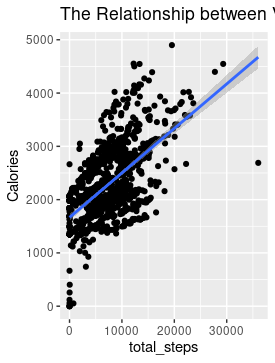
This graph compares the total number of steps taken by day. According to this chat, Tuesday seems to be the most active day where participants are reaching high levels of steps taken. This can surely lead to them achieving the goal of meeting a certain metric for walking per day.

ggplot(data=daily\_activity, aes(x=VeryActiveMinutes, y=Calories)) + geom\_point() + stat\_smooth(method=lm) + labs(title="The Relationship between Very Active Minutes and Total Daily Calories Burned")



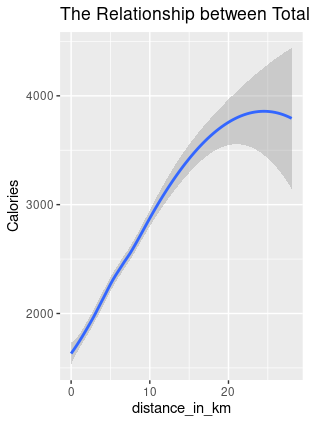
This plot showcases the positive relationship between very active minutes and total daily calories burned. This shows that the more vigorous physical activity the individual did, the more calories being burned.

ggplot(data=daily\_activity, aes(x=TotalSteps, y=Calories))+ geom\_point()+ stat\_smooth(method=lm)+ labs(title="The Relationship between Total Daily Steps and Total Daily Calories Burned")



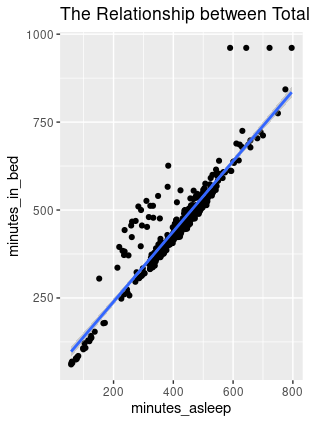
This plot depicts a positive relationship between total daily steps taken and total calories burned. This shows that the more steps taken the more calories being burned.

ggplot(data=daily\_activity, aes(x=TotalDistance, y=Calories))+ geom\_smooth()+ labs(title="The Relationship between Total Distance and Total Daily Calories Burned")



This last plot shows a positive trend between total distance and total daily calories burned. This shows that the more distance being run would likely result in more calories being burned.

ggplot(data=sleep\_day, aes(x=TotalMinutesAsleep, y=TotalTimeInBed))+ geom\_point()+ stat\_smooth(method=lm)+ labs(title="The Relationship between Total Minutes Asleep and Total Time in Bed")



This last plot shows a positive relationship between total minutes asleep and total time in bed. Ironically, the same amount of time spent asleep and in bed was pretty similar.

\*\*STEP 6: ACT\*\*

Bellabeat has been successful since it was founded by empowering women by providing data on their activity, sleep, stress, hydration levels, and reproductive health. Based on analyzing how Fitbit consumers use and respond to features, recommendations can be made to promote further growth for Bellabeat. The Bellabeat app should be completely enhanced and revamped. Rather than simply providing data on users’ health, the app should further encourage users to meet fitness goals and become a social media platform. The CDC recommends working out with a friend in order to feel more motivated, be more adventurous in trying workouts, and to become consistent. The CDC even recommends the use of a social media workout app to connect with friends and reach your goals. The Bellabeat app could become that social media workout app that women turn to, by creating an online community of supportive women ready to prioritize their health. Recommendations for the Bellabeat app: Enable social networking so users can post their favorite workouts, wellness tips, healthy meals, etc. Enable users to add friends and view each other’s activity. Create weekly fitness and wellness challenges to encourage use. Recommend users to get 10,000 steps a day and enable alert notifications to encourage users to meet goals. Recommend users to get at least 7 hours of sleep a night and enable alert notifications to encourage users to meet this. Recommend users get 75 minutes of vigorous activity a week and enable alert notifications to encourage users to meet this. Have health and fitness companies pay for advertising. Encourage users to enter weight and height to track BMI. If users are interested in losing weight, enable notifications to keep users on track to burn necessary calories to meet goals. Enable alert notifications if the user's resting heart rate varies significantly from their normal. Enable notifications to encourage activity if a user has spent an hour in bed awake. Enable notifications to encourage activity if a user has been sedentary for an extended period of time. Recommendations for Bellabeat membership: Partner with health & fitness companies and offer discounts for members. Offer reduced subscription fee when a member refers to a friend. Offer 30-day free trial subscription Offer discounts for Bellabeat smart device products with membership. Recommendations for Bellabeat products: Offer a bundle deal for the Spring and Leaf together. Heavily market Spring as Fitbit does not track hydration levels.

Works Cited

“The Dangers of Sitting: Why Sitting Is the New Smoking.” The Dangers of Sitting: Why Sitting Is the New Smoking — Better Health Channel, 22 Aug. 2020, www.betterhealth.vic.gov.au/health/healthyliving/the-dangers-of-sitting

"3 Reasons to Work out with a Friend.” Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 23 Apr. 2021, www.cdc.gov/diabetes/library/spotlights/workout-buddy.html

“About Adult Bmi.” Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 17 Sept. 2020, www.cdc.gov/healthyweight/assessing/bmi/adult\_bmi/index.html Gornall, Lucy.

“How to Lose Weight: How Many Calories Should i Eat to Lose Weight?” GoodtoKnow, 12 Aug. 2020, www.goodto.com/wellbeing/diets-exercise/what-is-calorie-how-many-lose-wei gt-425557 “CDC

— How Much Sleep Do I Need? — Sleep and Sleep Disorders.” Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 2 Mar. 2017, www.cdc.gov/sleep/about\_sleep/how\_much\_sleep.html Reed, Martin. “Spend Less Time In Bed If You Want More Sleep.” Healthcentral.com, 7 May 2017, www.healthcentral.com/article/spend-less-time-in-bed-if-you-want-more-sle ep Roland, James.

“What Are Mets, and How Are They Calculated?” Healthline, Healthline Media, 21 Oct. 2019, www.healthline.com/health/what-are-mets#calculation Grey, Heather. “Heart Rates Can Vary by 70 Bpm: What That Means for Your Health.” Healthline, Healthline Media, 9 Feb. 2020, www.healthline.com/health-news/what-your-heart-rate-says-about-your-healt h

“How Much Physical Activity Do Adults Need?” Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 7 Oct. 2020, www.cdc.gov/physicalactivity/basics/adults/index.htm Nield, David.

“Scientists Figured out How Much Exercise You Need to ‘Offset’ a Day of Sitting.” ScienceAlert, 26 Nov. 2020, www.sciencealert.com/getting-a-sweat-on-for-30-40-minutes-could-offset-aday-of-sitting-down