How Can a Wellness Technology Company Play It Smart?

Angela Iao

Last updated Aug 2022

Scenario

I am 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.

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.

About Bellabeat

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, maintains 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 marketing strategy.

Ask

Business Task

The **key business task** is to analyze smart device usage data to gain insights into how customers use smart devices. I will apply those insights and make data-driven recommendations for the Bellabeat marketing analytics team.

Stakeholders

- **Urška Sršen**: Bellabeat's co-founder and Chief Creative Officer
- **Sandro Mur**: Mathematician and Bellabeat co-founder; 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.

Prepare

Data Source

Data source: https://www.kaggle.com/arashnic/fitbit (https://www.kaggle.com/arashnic/fitbit) (CC0: Public Domain, dataset made available through Mobius)

The dataset, FitBit Fitness Tracker Data, consists of 18 CVS files that are organized in a long format. The dataset contains personal fitness tracker data from 30 Fitbit users, generated from a distributed survey via Amazon Mechanical Turk between March 12, 2016, to May 12, 2016. 30 participants consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring. It includes information about daily activity, steps, and sleep data that can be used to explore users' habits.

Data Integrity

Please take into consideration that this dataset has various limitations and may not provide reliable insights. Highly recommend finding more accurate and reliable data sources.

Limitations

Reliability | *Low* | The dataset is incomplete due to the small sample size (n = 30). The data source does not mention whether the sample is selected at random, potentially sampling bias.

Original | *Low* | The dataset is generated from a survey via Amazon Mechanical Turk, so it may not be accurate.

Comprehensive | *Low* | The data has sufficient data, however, the parameters and attributes are undefined and unclear.

Current | *Low* | The survey was done 5 years ago in 2016, not representative of present times. The dataset is outdated.

Process

Dataset Parameters

There are 18 CSV files in this dataset:

Weight

1. **weightloginfo** → Id, Date, WeightKg, WeightPounds, Fat, BMI, IsManualReport, LogID

METs (Metabolic Equivalent of Tasks

2. **minuteMETsNarrow** → id, ActivityMinute, METs

Steps

3. **minuteStepsWide** → id, ActivityHour, StepsNN

4. **minuteStepsNarrow** → id, ActivityMinute, Steps
5. **hourlySteps** → id, ActivityHour, StepTotal
6. **dailySteps** → id, ActivityDay, StepTotal

Sleep

7. **sleepDay** → id, SleepDay, TotalSleepRecords, TotalMinutesAsleep, TotalTimeInBed
8. **minuteSleep** → id, date, value, logid

Intensities

9. **minuteIntensitiesWide** → id, ActivityHour, IntensityNN
10. **minuteIntensitiesNarrow** → id, ActivityHour, Intensity
11. **hourlyIntensities8** → id, ActivityHour, TotalIntensity, AverageIntensity
12. **dailyIntensities** → id, ActivityDay, Sedentary Minutes, LightlyActiveMinutes, FairlyActiveMinutes, VeryActiveMinutes, SedentaryActiveDistance, LightActiveDistance, ModeratelyActiveDistance, VeryActiveDistance

Calories

13. **minuteCaloriesWide** → Id, ActivityHour, CaloriesNN
14. **minuteCaloriesNarrow** → Id, ActivityHour, Calories
15. **hourlyCalories** → id, ActivityHour, Calories
16. **dailyCalories** → id, ActivityDay, Calories

Heart Rate

17. **heartrate_seconds** → id, time, value

Activity

18. **dailyActivity** → id, ActivityDate, TotalSteps, TotalDistance, TrackerDistance, LoggedActiviesDistance, VeryActiveDistance, ModeratelyActiveDistance, LightActiveDistance, SedentaryActiveDistance, VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes, SedentaryMinutes, Calories

Code

For this report, I will explore how activeness impacts users' sleep using R. To analyse that theme, I will analyze data from the following data files:

- dailyActivity
- dailySteps
- sleepDay

Importing Packages

```
install.packages(c("tidyverse", "dplyr", "tidyr", "janitor", "readr", "skimr", "ggplot2"), repos = "http://cran.us.r-project.org")
```

```
## Installing packages into 'C:/Users/Angela/OneDrive/Documents/R/win-library/4.1'
## (as 'lib' is unspecified)
```

```
## package 'tidyverse' successfully unpacked and MD5 sums checked
## package 'dplyr' successfully unpacked and MD5 sums checked
## package 'tidyr' successfully unpacked and MD5 sums checked
## package 'janitor' successfully unpacked and MD5 sums checked
## package 'readr' successfully unpacked and MD5 sums checked
## package 'skimr' successfully unpacked and MD5 sums checked
## package 'ggplot2' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Angela\AppData\Local\Temp\RtmpqsNyoA\downloaded_packages
```

Loading Packages

```
library(lubridate)
```

```
## Warning: package 'lubridate' was built under R version 4.1.3
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.1.3
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
```

```
## Warning: package 'tidyr' was built under R version 4.1.3
```

```
library(readr)
```

```
## Warning: package 'readr' was built under R version 4.1.3
```

```
library(janitor)
```

```
## Warning: package 'janitor' was built under R version 4.1.3
```

```
##
## Attaching package: 'janitor'
```

```
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```
library(skimr)
```

```
## Warning: package 'skimr' was built under R version 4.1.3
```

Importing Data

```
## Rows: 940 Columns: 15
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (1): ActivityDate
## dbl (14): Id, TotalSteps, TotalDistance, TrackerDistance, LoggedActivitiesDi...
##
## 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.
## Rows: 940 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): ActivityDay
## dbl (2): Id, StepTotal
##
## 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.
```

Taking a First Look at Our Data

```
#generals to our data
head(dailyActivity)
```

```
## # A tibble: 6 x 15
##       Id Activ~1 Total~2 Total~3 Track~4 Logge~5 VeryA~6 Moder~7 Light~8 Seden~9
##    <dbl> <chr>     <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1 1.50e9 4/12/2~   13162    8.5     8.5        0    1.88   0.550    6.06       0
## 2 1.50e9 4/13/2~   10735    6.97    6.97       0    1.57   0.690    4.71       0
## 3 1.50e9 4/14/2~   10460    6.74    6.74       0    2.44   0.400    3.91       0
## 4 1.50e9 4/15/2~    9762    6.28    6.28       0    2.14   1.26     2.83       0
## 5 1.50e9 4/16/2~   12669    8.16    8.16       0    2.71   0.410    5.04       0
## 6 1.50e9 4/17/2~    9705    6.48    6.48       0    3.19   0.780    2.51       0
## # ... with 5 more variables: VeryActiveMinutes <dbl>,
## #   FairlyActiveMinutes <dbl>, LightlyActiveMinutes <dbl>,
## #   SedentaryMinutes <dbl>, Calories <dbl>, and abbreviated variable names
## #   1: ActivityDate, 2: TotalSteps, 3: TotalDistance, 4: TrackerDistance,
## #   5: LoggedActivitiesDistance, 6: VeryActiveDistance,
## #   7: ModeratelyActiveDistance, 8: LightActiveDistance,
## #   9: SedentaryActiveDistance
## # i Use `colnames()` to see all variable names
```

```
skim_without_charts(dailyActivity) #no missing values
```

Data summary

Name	dailyActivity
Number of rows	940
Number of columns	15
_____	
Column type frequency:	
character	1
numeric	14
_____	
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
ActivityDate	0	1	8	9	0	31	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p
Id	0	1	4.855407e+09	2.424805e+09	1503960366	2.320127e+09	4.445115e+
TotalSteps	0	1	7.637910e+03	5.087150e+03	0	3.789750e+03	7.405500e+
TotalDistance	0	1	5.490000e+00	3.920000e+00	0	2.620000e+00	5.240000e+
TrackerDistance	0	1	5.480000e+00	3.910000e+00	0	2.620000e+00	5.240000e+
LoggedActivitiesDistance	0	1	1.100000e-01	6.200000e-01	0	0.000000e+00	0.000000e+
VeryActiveDistance	0	1	1.500000e+00	2.660000e+00	0	0.000000e+00	2.100000e-

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p
ModeratelyActiveDistance	0	1	5.700000e-01	8.800000e-01	0	0.000000e+00	2.400000e-
LightActiveDistance	0	1	3.340000e+00	2.040000e+00	0	1.950000e+00	3.360000e+
SedentaryActiveDistance	0	1	0.000000e+00	1.000000e-02	0	0.000000e+00	0.000000e+
VeryActiveMinutes	0	1	2.116000e+01	3.284000e+01	0	0.000000e+00	4.000000e+
FairlyActiveMinutes	0	1	1.356000e+01	1.999000e+01	0	0.000000e+00	6.000000e+
LightlyActiveMinutes	0	1	1.928100e+02	1.091700e+02	0	1.270000e+02	1.990000e+
SedentaryMinutes	0	1	9.912100e+02	3.012700e+02	0	7.297500e+02	1.057500e+
Calories	0	1	2.303610e+03	7.181700e+02	0	1.828500e+03	2.134000e+

```
str(dailyActivity)
```

```
## spec_tbl_df [940 x 15] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id                       : num [1:940] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ ActivityDate             : chr [1:940] "4/12/2016" "4/13/2016" "4/14/2016" "4/15/2016" ...
##  $ TotalSteps               : num [1:940] 13162 10735 10460 9762 12669 ...
##  $ TotalDistance            : num [1:940] 8.5 6.97 6.74 6.28 8.16 ...
##  $ TrackerDistance          : num [1:940] 8.5 6.97 6.74 6.28 8.16 ...
##  $ LoggedActivitiesDistance: num [1:940] 0 0 0 0 0 0 0 0 0 0 ...
##  $ VeryActiveDistance       : num [1:940] 1.88 1.57 2.44 2.14 2.71 ...
##  $ ModeratelyActiveDistance: num [1:940] 0.55 0.69 0.4 1.26 0.41 ...
##  $ LightActiveDistance      : num [1:940] 6.06 4.71 3.91 2.83 5.04 ...
##  $ SedentaryActiveDistance : num [1:940] 0 0 0 0 0 0 0 0 0 0 ...
##  $ VeryActiveMinutes        : num [1:940] 25 21 30 29 36 38 42 50 28 19 ...
##  $ FairlyActiveMinutes      : num [1:940] 13 19 11 34 10 20 16 31 12 8 ...
##  $ LightlyActiveMinutes     : num [1:940] 328 217 181 209 221 164 233 264 205 211 ...
##  $ SedentaryMinutes         : num [1:940] 728 776 1218 726 773 ...
##  $ Calories                 : num [1:940] 1985 1797 1776 1745 1863 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   ActivityDate = col_character(),
##   ..   TotalSteps = col_double(),
##   ..   TotalDistance = col_double(),
##   ..   TrackerDistance = col_double(),
##   ..   LoggedActivitiesDistance = col_double(),
##   ..   VeryActiveDistance = col_double(),
##   ..   ModeratelyActiveDistance = col_double(),
##   ..   LightActiveDistance = col_double(),
##   ..   SedentaryActiveDistance = col_double(),
##   ..   VeryActiveMinutes = col_double(),
##   ..   FairlyActiveMinutes = col_double(),
##   ..   LightlyActiveMinutes = col_double(),
##   ..   SedentaryMinutes = col_double(),
##   ..   Calories = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
head(dailySteps)
```

```
## # A tibble: 6 x 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
```

```
skim_without_charts(dailySteps) #no missing values
```

Data summary

Name	dailySteps
Number of rows	940
Number of columns	3
_____	
Column type frequency:	
character	1
numeric	2
_____	
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
ActivityDay	0	1	8	9	0	31	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	
Id	0	1	4.855407e+09	2.424805e+09	1503960366	2.320127e+09	4445114986.0	696218
StepTotal	0	1	7.637910e+03	5.087150e+03	0	3.789750e+03	7405.5	1

```
str(dailySteps)
```

```
## spec_tbl_df [940 x 3] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id         : num [1:940] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ ActivityDay: chr [1:940] "4/12/2016" "4/13/2016" "4/14/2016" "4/15/2016" ...
##  $ StepTotal  : num [1:940] 13162 10735 10460 9762 12669 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   ActivityDay = col_character(),
##   ..   StepTotal = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
head(sleepDay)
```

```
##           Id              SleepDay TotalSleepRecords TotalMinutesAsleep
## 1 1503960366 4/12/2016 12:00:00 AM                 1                327
## 2 1503960366 4/13/2016 12:00:00 AM                 2                384
## 3 1503960366 4/15/2016 12:00:00 AM                 1                412
## 4 1503960366 4/16/2016 12:00:00 AM                 2                340
## 5 1503960366 4/17/2016 12:00:00 AM                 1                700
## 6 1503960366 4/19/2016 12:00:00 AM                 1                304
##   TotalTimeInBed
## 1            346
## 2            407
## 3            442
## 4            367
## 5            712
## 6            320
```

```
skim_without_charts(sleepDay) #no missing values
```

Data summary

Name	sleepDay
Number of rows	413
Number of columns	5
_____	
Column type frequency:	
character	1
numeric	4
_____	
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
SleepDay	0	1	20	21	0	31	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	
Id	0	1	5.000979e+09	2.06036e+09	1503960366	3977333714	4702921684	696218
TotalSleepRecords	0	1	1.120000e+00	3.50000e-01	1	1	1	
TotalMinutesAsleep	0	1	4.194700e+02	1.18340e+02	58	361	433	
TotalTimeInBed	0	1	4.586400e+02	1.27100e+02	61	403	463	

```
str(sleepDay)
```

```
## 'data.frame':    413 obs. of  5 variables:
##  $ Id               : num  1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ SleepDay         : chr  "4/12/2016 12:00:00 AM" "4/13/2016 12:00:00 AM" "4/15/2016 12:00:00 AM" "4/16/2
016 12:00:00 AM" ...
##  $ TotalSleepRecords : int  1 2 1 2 1 1 1 1 1 1 ...
##  $ TotalMinutesAsleep: int  327 384 412 340 700 304 360 325 361 430 ...
##  $ TotalTimeInBed    : int  346 407 442 367 712 320 377 364 384 449 ...
```

Cleaning the Dataset

```
#Cleaning- formatting dates from string to date and correcting data types
#Cleaning- adding another column 'Date' for activity and sleep so that I can merge later
activity <- dailyActivity %>%
  mutate(ActivityDate = mdy(ActivityDate)) %>%
  mutate(TotalSteps = as.integer(TotalSteps)) %>%
  mutate(Date = ActivityDate)

sleep <- sleepDay %>%
  mutate(SleepDay = mdy_hms(SleepDay)) %>%
  mutate(TotalMinutesAsleep = as.numeric(TotalMinutesAsleep)) %>%
  mutate(TotalTimeInBed = as.numeric(TotalTimeInBed)) %>%
  mutate(Date = SleepDay)

steps <- dailySteps %>%
  mutate(ActivityDay = mdy(ActivityDay)) %>%
  mutate(Date = ActivityDay)
```

Exploring Data

```
activity %>%
  select(TotalSteps, VeryActiveMinutes, FairlyActiveMinutes,
         LightlyActiveMinutes, SedentaryMinutes) %>%
  summary(activity)
```

```
##    TotalSteps    VeryActiveMinutes FairlyActiveMinutes LightlyActiveMinutes
##  Min.   :    0   Min.   :  0.00    Min.   :  0.00      Min.   :  0.0
##  1st Qu.: 3790   1st Qu.:  0.00    1st Qu.:  0.00      1st Qu.:127.0
##  Median : 7406   Median :  4.00    Median :  6.00      Median :199.0
##  Mean   : 7638   Mean   : 21.16    Mean   : 13.56      Mean   :192.8
##  3rd Qu.:10727   3rd Qu.: 32.00    3rd Qu.: 19.00      3rd Qu.:264.0
##  Max.   :36019   Max.   :210.00    Max.   :143.00      Max.   :518.0
##  SedentaryMinutes
##  Min.   :   0.0
##  1st Qu.: 729.8
##  Median :1057.5
##  Mean   : 991.2
##  3rd Qu.:1229.5
##  Max.   :1440.0
```

```
sleep %>%
  select(Date, TotalSleepRecords, TotalMinutesAsleep, TotalTimeInBed) %>%
  summary()
```

```
##       Date                       TotalSleepRecords TotalMinutesAsleep
##  Min.   :2016-04-12 00:00:00   Min.   :1.000     Min.   : 58.0
##  1st Qu.:2016-04-19 00:00:00   1st Qu.:1.000     1st Qu.:361.0
##  Median :2016-04-27 00:00:00   Median :1.000     Median :433.0
##  Mean   :2016-04-26 12:40:05   Mean   :1.119     Mean   :419.5
##  3rd Qu.:2016-05-04 00:00:00   3rd Qu.:1.000     3rd Qu.:490.0
##  Max.   :2016-05-12 00:00:00   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
```

```
steps %>%
  select(StepTotal) %>%
  summary()
```

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

```
#Exploring- distinct users in each data file
users_activity <- unique(dailyActivity$Id)
length(users_activity) #33
```

```
## [1] 33
```

```
users_steps <- unique(dailySteps$Id)
length(users_steps) #33
```

```
## [1] 33
```

```
users_sleep <- unique(sleepDay$Id)
length(users_sleep) #24
```

```
## [1] 24
```

```
#Exploring- counting number of entries
nrow(sleep) #413
```

```
## [1] 413
```

```
nrow(activity) #940
```

```
## [1] 940
```

```
#Exploring- how many times each user logged in
sleepDay %>% count(Id)
```

```
##            Id  n
## 1  1503960366 25
## 2  1644430081  4
## 3  1844505072  3
## 4  1927972279  5
## 5  2026352035 28
## 6  2320127002  1
## 7  2347167796 15
## 8  3977333714 28
## 9  4020332650  8
## 10 4319703577 26
## 11 4388161847 24
## 12 4445114986 28
## 13 4558609924  5
## 14 4702921684 28
## 15 5553957443 31
## 16 5577150313 26
## 17 6117666160 18
## 18 6775888955  3
## 19 6962181067 31
## 20 7007744171  2
## 21 7086361926 24
## 22 8053475328  3
## 23 8378563200 32
## 24 8792009665 15
```

Analyze

Looking more deeply into these variables:

- TotalSteps (steps users take per day)
- SedentaryMinutes (minutes users stay sedentary per day (i.e. sitting and lying down))
- TotalMinutesSleep (minutes users are sleeping per day)
- TotalTimeInBed (minutes users are in bed)
- Activity Level (very active, fairly active, lightly active, sedentary)
    - StepsActivityLevel (categorize users' activity levels according to recommended steps per day 1 (https://www.medicinenet.com/how_many_steps_a_day_is_considered_active/article.htm))
    - ActivityLevelMins (categorize users by activity levels according to users daily activity minutes)
- SleepSufficiency (categorize whether users had enough sleep (at least 420 minutes or 7 hours 2 (https://www.sleepfoundation.org/sleep-hygiene/what-is-healthy-sleep)) or not enough sleep (less than 420 minutes or 7 hours)
- AwakeInBed (minutes that users are in bed but not asleep)

Code

Merging Data

```
#Merging activity and sleep in a df by 'Id' and 'Date', excluding columns I don't need
sleep_activity <- merge(sleep, activity, by = c('Id', 'Date')) %>%
  select(-c(TotalDistance, TrackerDistance, LoggedActivitiesDistance,
            VeryActiveDistance, ModeratelyActiveDistance, LightActiveDistance, SedentaryActiveDistance))
View(sleep_activity)

#filtering out the data that captured sedentary minutes of 1440 mins (24 hrs)
sleep_activity_cleaned <- sleep_activity %>%
  filter(SedentaryMinutes != 1440)

#filtered out the data that was tracked when user was not using it
sleep_activity_cleaned <- sleep_activity %>%
  filter(SedentaryMinutes != 1440)
summary(sleep_activity_cleaned)
```

```
##        Id                 Date
##  Min.   :1.504e+09   Min.   :2016-04-12 00:00:00
##  1st Qu.:3.977e+09   1st Qu.:2016-04-19 00:00:00
##  Median :4.703e+09   Median :2016-04-27 00:00:00
##  Mean   :5.001e+09   Mean   :2016-04-26 12:40:05
##  3rd Qu.:6.962e+09   3rd Qu.:2016-05-04 00:00:00
##  Max.   :8.792e+09   Max.   :2016-05-12 00:00:00
##     SleepDay                   TotalSleepRecords TotalMinutesAsleep
##  Min.   :2016-04-12 00:00:00   Min.   :1.000     Min.   : 58.0
##  1st Qu.:2016-04-19 00:00:00   1st Qu.:1.000     1st Qu.:361.0
##  Median :2016-04-27 00:00:00   Median :1.000     Median :433.0
##  Mean   :2016-04-26 12:40:05   Mean   :1.119     Mean   :419.5
##  3rd Qu.:2016-05-04 00:00:00   3rd Qu.:1.000     3rd Qu.:490.0
##  Max.   :2016-05-12 00:00:00   Max.   :3.000     Max.   :796.0
##  TotalTimeInBed   ActivityDate          TotalSteps    VeryActiveMinutes
##  Min.   : 61.0   Min.   :2016-04-12   Min.   :   17   Min.   :  0.00
##  1st Qu.:403.0   1st Qu.:2016-04-19   1st Qu.: 5206   1st Qu.:  0.00
##  Median :463.0   Median :2016-04-27   Median : 8925   Median :  9.00
##  Mean   :458.6   Mean   :2016-04-26   Mean   : 8541   Mean   : 25.19
##  3rd Qu.:526.0   3rd Qu.:2016-05-04   3rd Qu.:11393   3rd Qu.: 38.00
##  Max.   :961.0   Max.   :2016-05-12   Max.   :22770   Max.   :210.00
##  FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes    Calories

##  Min.   :  0.00      Min.   :  2.0        Min.   :   0.0   Min.   : 257
##  1st Qu.:  0.00      1st Qu.:158.0        1st Qu.: 631.0   1st Qu.:1850
##  Median : 11.00      Median :208.0        Median : 717.0   Median :2220
##  Mean   : 18.04      Mean   :216.9        Mean   : 712.2   Mean   :2398
##  3rd Qu.: 27.00      3rd Qu.:263.0        3rd Qu.: 783.0   3rd Qu.:2926
##  Max.   :143.00      Max.   :518.0        Max.   :1265.0   Max.   :4900
```

The mean for sedentary minutes and total steps are below average for the recommended level for healthy adults. For sedentary minutes, the average time users have been sedentary is 717 minutes per day, almost 12 hours which is considered a high average. The mean for total steps is 8541 which is considered low for the recommended at least 10,000 steps per day. 3 (https://ijbnpa.biomedcentral.com/articles/10.1186/1479-5868-8-79#:~:text=In%20summary%2C%20at%20least%20in,populace%20%5B3%2C%2023%5D.)

Analyzing Data

On average, how many users are active? (at least 10k steps a day)

```
avg_steps <- sleep_activity_cleaned %>%
  group_by(Id) %>%
  summarise(AverageSteps = mean(TotalSteps, na.rm = TRUE))

active_users_steps <- filter(avg_steps, AverageSteps >= 10000)
count(active_users_steps) #5
```

```
## # A tibble: 1 x 1
##       n
##   <int>
## 1     5
```

```
active_users_steps_percent <- nrow(active_users_steps)/nrow(avg_steps)*100
print(active_users_steps_percent ) #20.83%
```

```
## [1] 20.83333
```

5 out of 24 or around 21% of smart device users are considered active or walk at least 10,000 steps a day. That is a relatively low number, considering that smart device trackers are meant to measure users' activity and health to motivate users towards their fitness goals.

How many users are sedentary? (less than 5k steps a day)

```
sedentary_users_steps <- filter(avg_steps, AverageSteps < 5000)
count(sedentary_users_steps) #5
```

```
## # A tibble: 1 x 1
##       n
##   <int>
## 1     5
```

```
sedentary_users_steps_percent <- nrow(sedentary_users_steps)/nrow(avg_steps)*100
print(sedentary_users_steps_percent) #20.83%
```

```
## [1] 20.83333
```

5 out of 33 or around 21% of users are considered sedentary or inactive, meaning they walk less than 5,000 steps per day. That is a relatively low number which is good because we want to encourage more exercise and activeness.

On average, how many users are healthy sleepers? (at least 420 minutes a night)

```
avg_sleep_minutes <- sleep_activity_cleaned %>%
  group_by(Id) %>%
  summarise(AverageSleepMinutes = mean(TotalMinutesAsleep, na.rm = TRUE))
summary(avg_sleep_minutes$AverageSleepMinutes) #377.6 or 6.29 hrs
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    61.0   336.3   419.1   377.6   449.3   652.0
```

On average, users sleep for 377.6 minutes or 6.29 hours. That average falls below the recommended level of 420 minutes or 7 hours 2 (https://www.sleepfoundation.org/sleep-hygiene/what-is-healthy-sleep).

```
#create new columns in sleep_activity_cleaned df
sleep_activity_cleaned$TotalLoggedMinutes = sleep_activity_cleaned$TotalTimeInBed +
  sleep_activity_cleaned$VeryActiveMinutes +
  sleep_activity_cleaned$FairlyActiveMinutes +
  sleep_activity_cleaned$LightlyActiveMinutes +
  sleep_activity_cleaned$SedentaryMinutes

sleep_activity_cleaned$TotalActiveMinutes = sleep_activity_cleaned$VeryActiveMinutes +
  sleep_activity_cleaned$FairlyActiveMinutes +
  sleep_activity_cleaned$LightlyActiveMinutes


#creating new column to categorize the activity level of users based on steps and determine whether they had s
ufficient sleep
sleep_steps <- sleep_activity_cleaned %>%
  mutate(StepsActivityLevel = case_when(
    TotalSteps < 5000 ~ "Inactive",
    TotalSteps >= 5000 & TotalSteps < 8750 ~ "Lightly Active",
    TotalSteps >= 8750 & TotalSteps < 12500 ~ "Fairly Active",
    TotalSteps >= 12500 ~ "Very Active"
  ), SleepSufficiency = case_when(
    TotalMinutesAsleep < 420 ~ "Not Enough",
    TotalMinutesAsleep >= 420 ~ "Enough"
  ))


#Analyzing- correlation b/w SedentaryMinutes and TotalMinutesAsleep
cor(sleep_steps$SedentaryMinutes, sleep_steps$TotalMinutesAsleep) #-0.599394
```

```
## [1] -0.599394
```

The correlation between SedentaryMinutes and TotalMinutesAsleep is -0.599394 or -0.6, which is considered to be a moderately strong negative correlation. [go to Figure 5]

```
#Analyzing- counting how many observations in EnoughSleep
sleep_steps %>% count(SleepSufficiency) #Enough- 231 observations; Not Enough- 182 observations
```

```
##   SleepSufficiency   n
## 1           Enough 231
## 2       Not Enough 182
```

There are 231 observations (55.9%) that were categorized as ‘Enough’ sleep, meaning users had at least 420 minutes of sleep that day. There are 182 observations (44.1%) that were categorized as ‘Not Enough’ sleep, meaning users had less than 420 minutes of sleep that day. This result is relatively consistent with our previous analysis of healthy sleep users and sleep minutes. [go to Figure 3]

```
#Analyzing- creating new column to categorize the activity level (based on activity minutes)
mean(sleep_activity_cleaned$SedentaryMinutes) # 712.1695
```

```
## [1] 712.1695
```

```
mean(sleep_activity_cleaned$LightlyActiveMinutes) # 216.8547
```

```
## [1] 216.8547
```

```
mean(sleep_activity_cleaned$FairlyActiveMinutes) # 18.03874
```

```
## [1] 18.03874
```

```
mean(sleep_activity_cleaned$VeryActiveMinutes) # 25.18886
```

```
## [1] 25.18886
```

```
sleep_activity_mins <- sleep_steps %>%
  group_by(Id) %>%
  mutate(AwakeInBed = TotalTimeInBed - TotalMinutesAsleep,
         ActivityLevelMins = case_when(
           SedentaryMinutes > 712.1965 &
             LightlyActiveMinutes < 216.8547 &
             FairlyActiveMinutes < 18.03874 &
             VeryActiveMinutes < 25.18886 ~ "Inactive",
           SedentaryMinutes < 712.1965 &
             LightlyActiveMinutes > 216.8547 &
             FairlyActiveMinutes < 18.03874 &
             VeryActiveMinutes < 25.18886 ~ "Lightly Active",
           SedentaryMinutes < 712.1965 &
             LightlyActiveMinutes < 216.8547 &
             FairlyActiveMinutes > 18.03874 &
             VeryActiveMinutes < 25.18886 ~ "Fairly Active",
           SedentaryMinutes < 712.1965 &
             LightlyActiveMinutes < 216.8547 &
             FairlyActiveMinutes < 18.03874 &
             VeryActiveMinutes > 25.18886 ~ "Very Active"),
         SleepSufficiency = case_when(
           TotalMinutesAsleep < 420 ~ "Not Enough",
           TotalMinutesAsleep >= 420 ~ "Enough")) %>%
  drop_na()


#Analyzing- correlation b/w AwakeInBed and TotalMinutesAsleep
cor(sleep_activity_mins$AwakeInBed, sleep_activity_mins$TotalMinutesAsleep) #-0.1385993
```

```
## [1] -0.1385993
```

```
#On average, how long does it take users to sleep?
mean(sleep_activity_mins$AwakeInBed) #38.83 mins
```

```
## [1] 38.82993
```

Note: The number of rows in this data has dropped in comparison to sleep_steps data because some observations had ambiguous results and did not fit the case criteria.

The correlation between AwakeInBed and TotalMinutesAsleep is -0.1385993 or -0.14, which is considered a weak negative correlation. It is not correlated enough to infer a strong relationship between those two variables. [go to Figure 6]

On average, it takes users 38.83 minutes to fall asleep. That is higher than the healthy recommended level of 10 to 20 minutes 4 (https://www.healthline.com/health/healthy-sleep/how-long-does-it-take-to-fall-asleep#:~:text=It%20should%20take%20between%2010,fall%20asleep%20much%20more%20quickly). Users are having a

harder time falling asleep. This could be due to other factors such as screen time before bed, stress, user's health or medical records, etc.

Share

Using data visualization to show trends and relationships amongst variables.

Aims to answer:

- Does regular physical activity lead to better sleep?
- How does being active vs being sedentary impact my sleeping schedule? Is there a relationship between sleep and activeness?
- Do users spend a considerable amount of time being sedentary?
- How does this translate in terms of improving Bellabeat?

Code

Loading Package

```
#Visualization- look at relationship b/w total steps and calories
library(ggplot2)
```

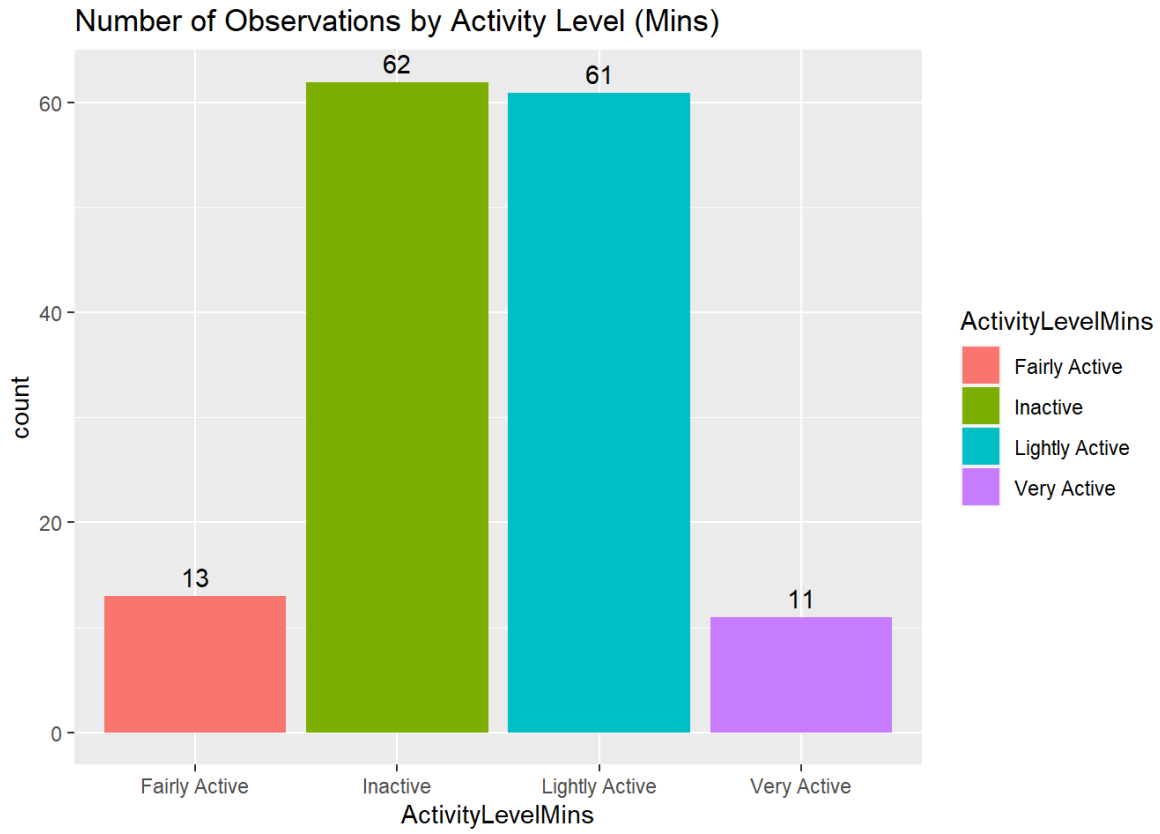
```
## Warning: package 'ggplot2' was built under R version 4.1.3
```

Visualizing Data

Figure 1: Number of Observations by Activity Level (Mins.)

Who is our target audience (the highest number of users) based on activity level (Minutes)?

```
ggplot(data = sleep_activity_mins, aes(ActivityLevelMins)) +
  geom_bar(aes(fill = ActivityLevelMins)) +
  geom_text(stat = 'count', aes(label = ..count..), vjust =  -0.5) +
  labs(title = 'Number of Observations by Activity Level (Mins)')
```

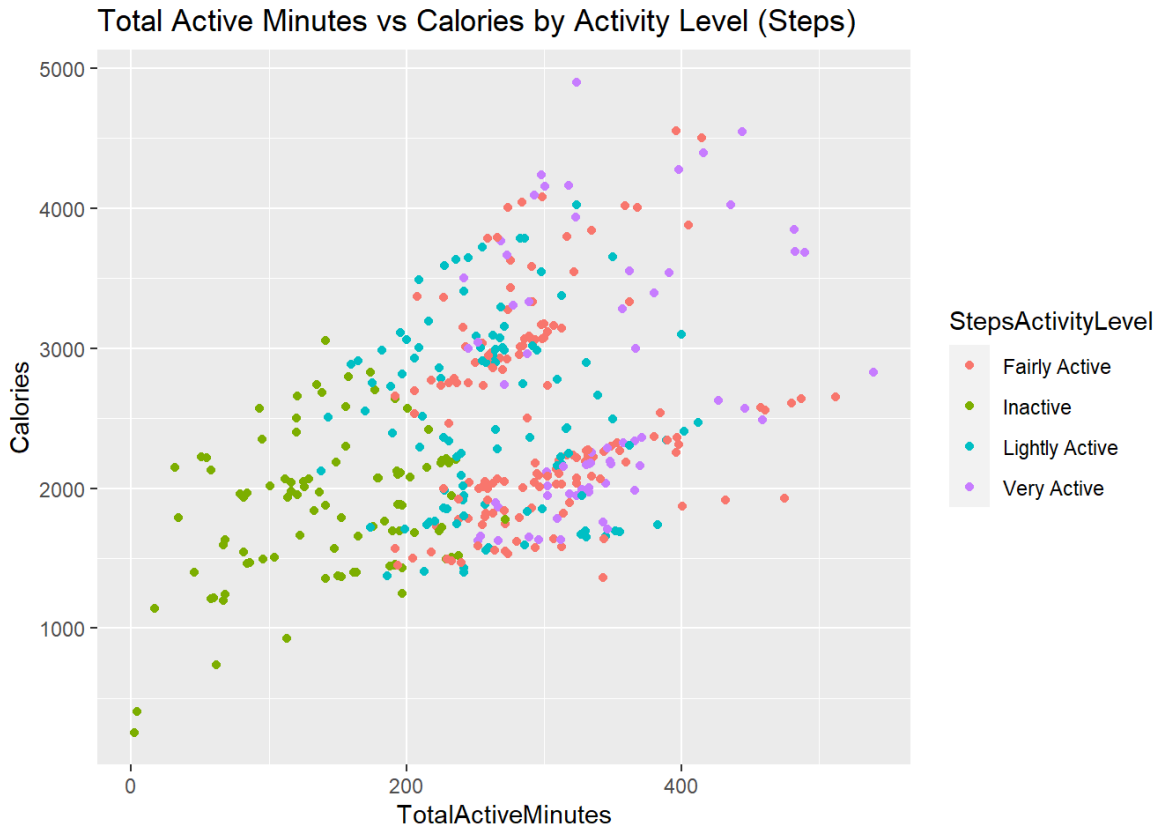


Most observations are from users that spend the most time being inactive or sedentary (i.e. sitting, lying down) and lightly active (i.e. standing, doing household chores, walking).

These are users that wear and use Bellabeat products the most. If the marketing team wants to bring in more users, they can target these niche audiences.

Figure 2: Total Active Minutes vs. Calories by Activity Level (Steps)

```
#Total Active Minutes vs Calories by Activity Level (Steps)
ggplot(sleep_steps, mapping = aes(
  x = TotalActiveMinutes, y = Calories, fill = StepsActivityLevel, color = StepsActivityLevel)) +
  geom_point() +
  labs(title = "Total Active Minutes vs Calories by Activity Level (Steps)")
```



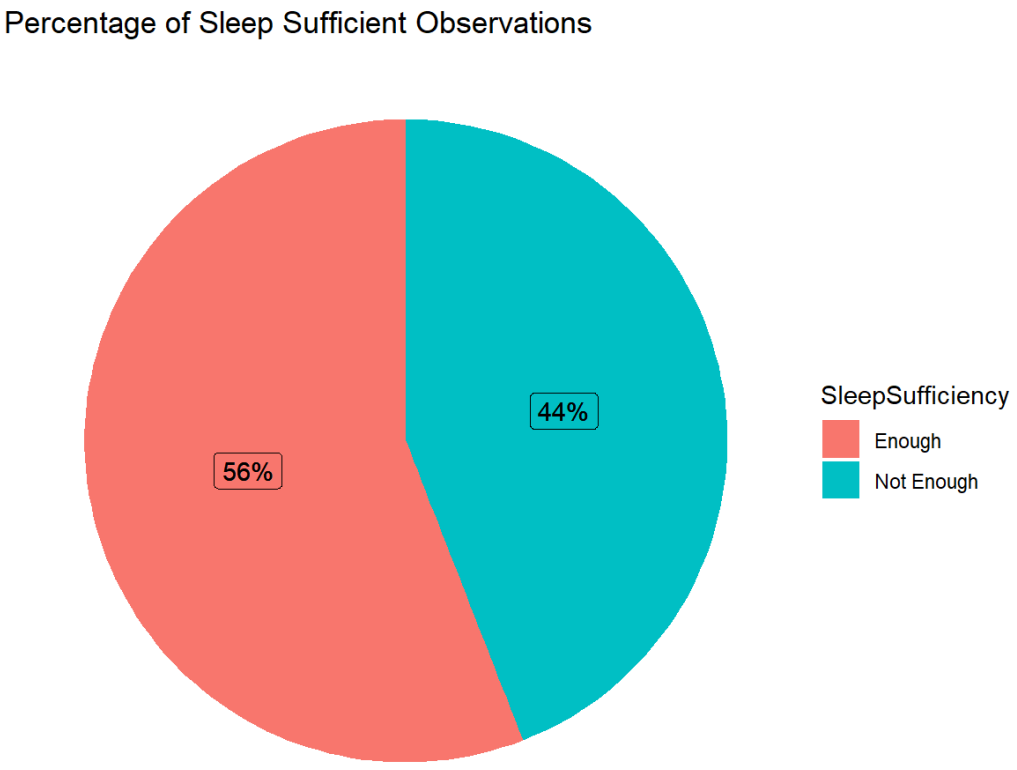
This graph looks at the total active minutes (Sedentary Minutes + LightlyActiveMinutes + FairlyActiveMinutes + VeryActiveMinutes) and the calories burned at each activity level.

There is a positive relationship between TotalActiveMinutes and Calories. The more time users are active, the more calories users burn. The activity level data is consistent with this analysis. Users who are very active spend the most time being active and burn the most calories per day; users who are inactive spend most of their time being sedentary and burn the least calories per day.

Figure 3: Percentage of Sleep Sufficient Observations

Are users getting enough sleep?

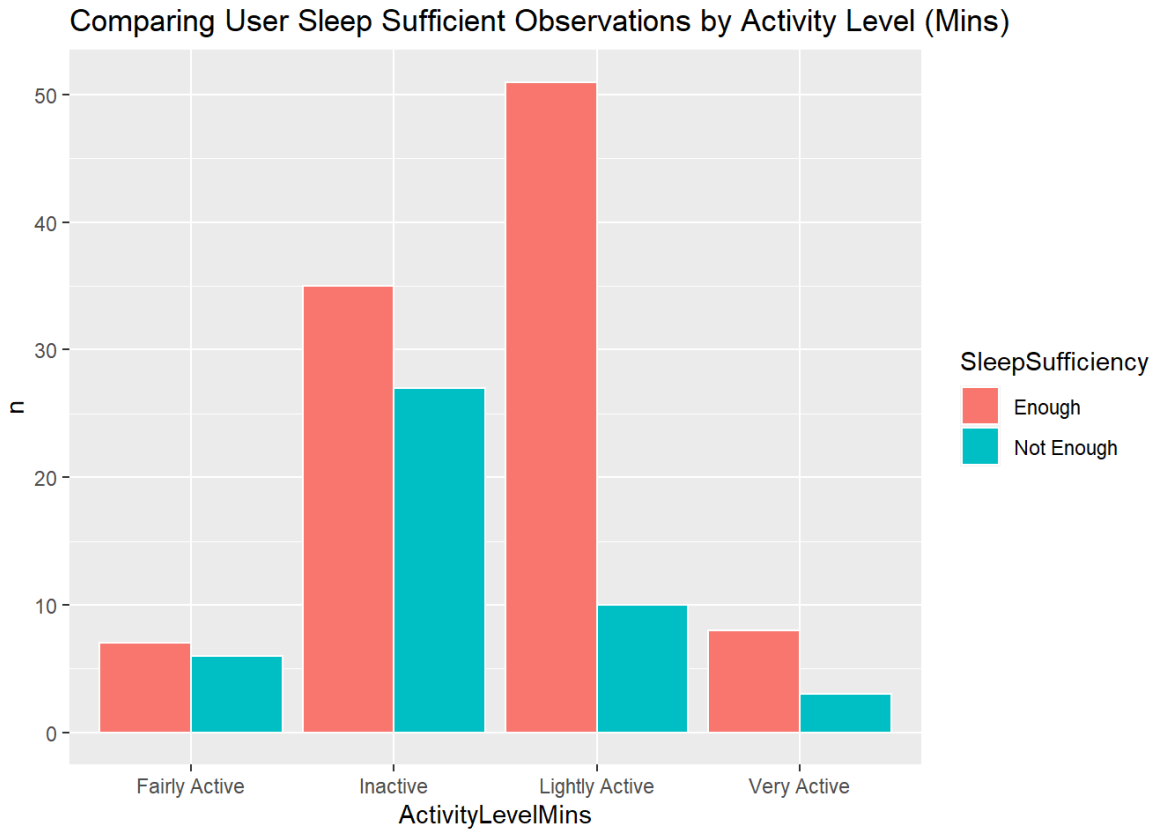
```
#creating a df to get an aggregate of observations (n) and the percentage for sleep sufficiency
pie_chart <- sleep_steps %>%
  group_by(SleepSufficiency) %>%
  count() %>%
  ungroup() %>%
  mutate(percent = n / sum(n)) %>%
  arrange(percent) %>%
  mutate(labels = scales::percent(percent))

#making a pie chart with annotations
ggplot(pie_chart, aes(x = "", y = percent, fill = SleepSufficiency)) +
  geom_col() +
  geom_label(aes(label = labels), position = position_stack(vjust = 0.5), show.legend = FALSE) +
  geom_text(aes(label = labels), position = position_stack(vjust = 0.5), show.legend = FALSE) +
  coord_polar(theta = "y") +
  theme_void() +
  labs(title = "Percentage of Sleep Sufficient Observations")
```



This visualization is consistent with our previous analysis. More than half of the users (56%) get enough sleep, meaning at least 420 minutes or 7 hours a day.

Figure 4: Comparing User Sleep Sufficient Observations by Activity Level (Minutes)

```
sleep_activity_mins %>%
  group_by(SleepSufficiency) %>%
  count(ActivityLevelMins) %>%
  ggplot(sleep_activity2, mapping = aes(x = ActivityLevelMins, y = n)) +
  geom_bar(aes(fill = SleepSufficiency),
           stat = "identity", color = "white",
           position = position_dodge(0.9)) +
  labs(title = "Comparing User Sleep Sufficient Observations by Activity Level (Mins)")
```



There is a clear trend; As user activity level decreases, users are less likely to get sufficient sleep (at least 420 minutes or 7 hours a night).

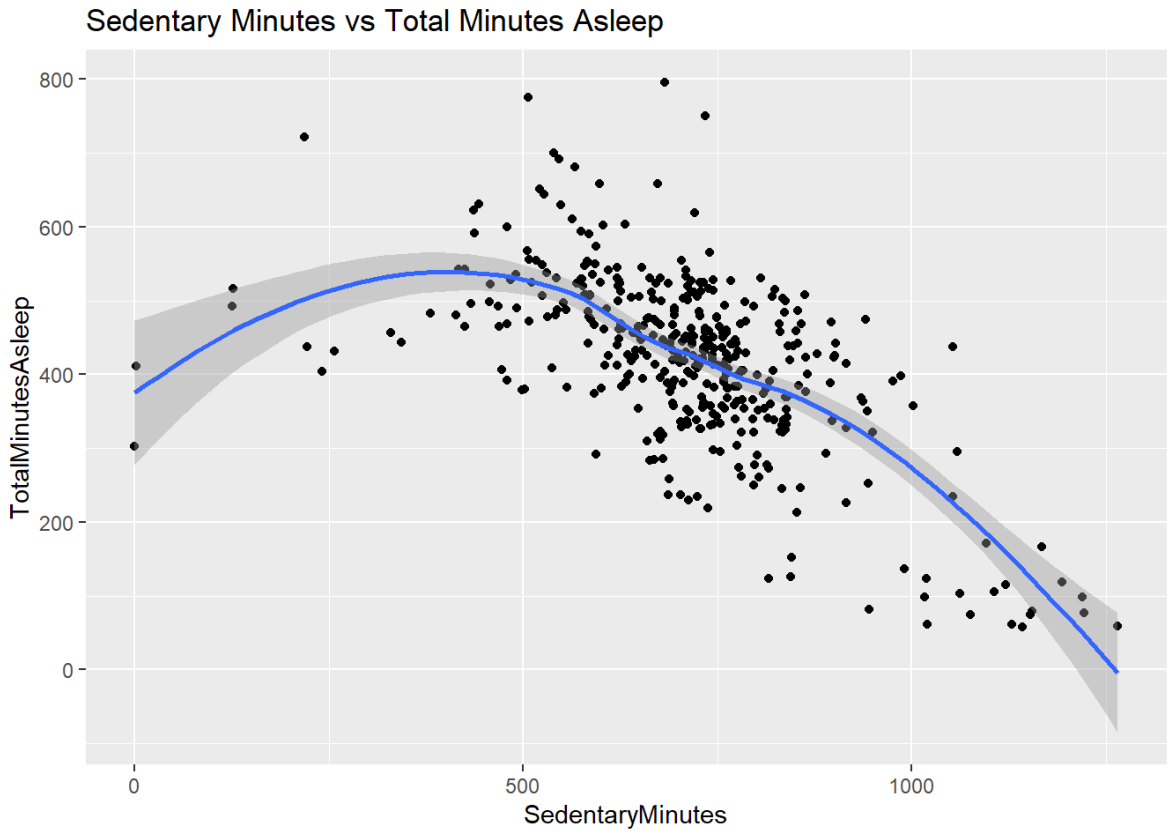
Poor sleep may contribute to physical inactivity 5 (https://journals.sagepub.com/doi/abs/10.1177/1559827614544437). The group that are least likely to get enough sleep is the sedentary group. While many sedentary users have enough sleep, there is still a high number of 'Not Enough' sleep observations compared to other activity levels.

Figure 5: Sedentary Minutes vs. Total Minutes Asleep

How does being sedentary impact sleep?

```
ggplot(sleep_activity_cleaned, mapping = aes(x = SedentaryMinutes, y = TotalMinutesAsleep)) + geom_point() + g
eom_smooth() + labs(title = "Sedentary Minutes vs Total Minutes Asleep")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



It is evident that there is a negative relationship between Sedentaryminutes and TotalMinutesAsleep. The more users stay sedentary, the fewer users sleep.

This can be concerning for the user's health because Bellabeat users spend a lot of time sedent (on average, 571.6 minutes or 9.52 hours a day). Inactiveness can be strongly associated with insomnia and restlessness. Increasing exercise or physical activity can improve sleep duration and alertness in the daytime 6 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6352043/).

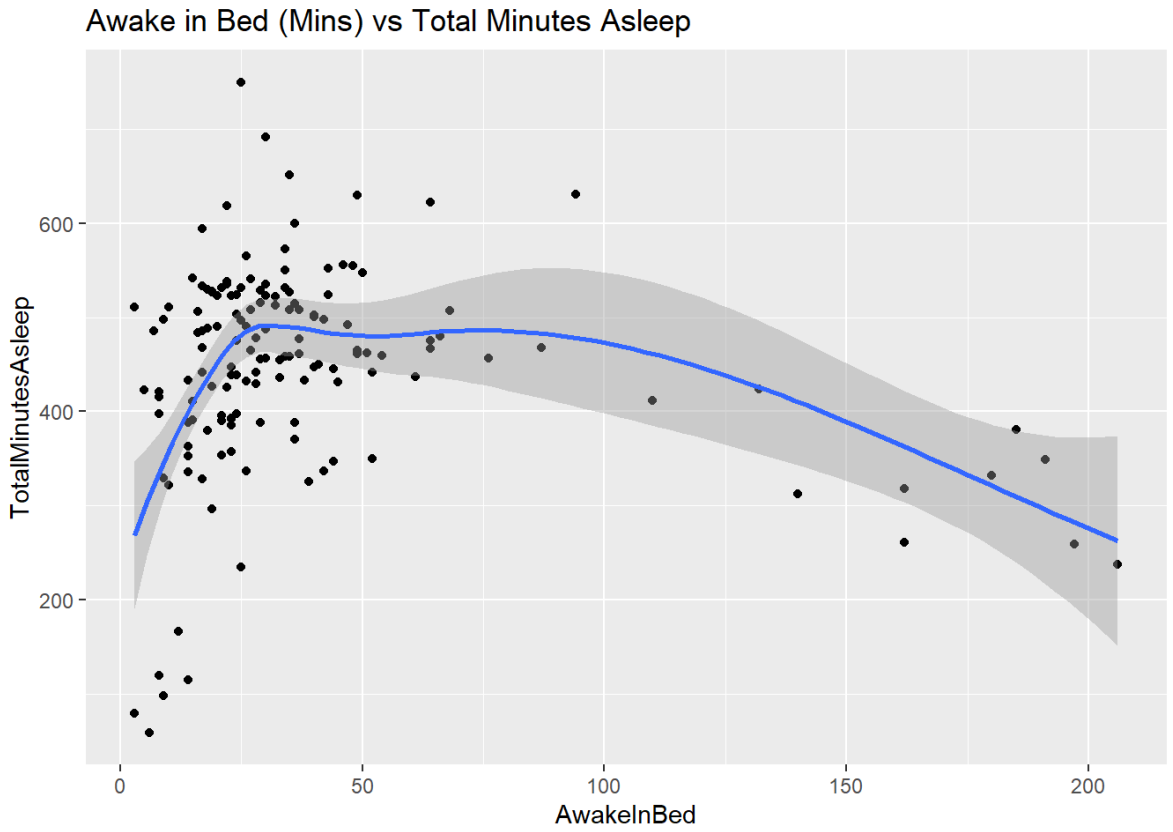
In this study, researchers examined the association between sedentary behaviours and sleep time and found that physical activity increased with sleep duration in younger respondents (aged 20-39) 7 (https://www.sciencedirect.com/science/article/abs/pii/S0091743514002035%20%E2%86%92%20physical%20activity%20and%20sle

Figure 6: AwakeInBed vs. TotalMinutesAsleep

How does being awake in bed relate to sleep?

```
ggplot(data = sleep_activity_mins, mapping = aes(x = AwakeInBed, y = TotalMinutesAsleep)) +
  geom_point() +
  geom_smooth() +
  labs(title = "Awake in Bed (Mins) vs Total Minutes Asleep")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



This graph shows a decreasing concavity, showing a negative relationship between AwakeInBed and TotalMinutesAsleep. The longer users stay awake in bed, the fewer users sleep. The correlation between these two variables shows a weak relationship.

Act

Limitations

If the dataset included demographic data, we can analyze the differences in sleep and activity by age. Age can affect sleep and activity, so having that extra data can provide more additional and insightful analysis.

Key Findings

- On average, users stay sedentary for 717 minutes per day.
- On average, users take 8541 steps per day which are lower than recommended (10k steps/day).
- 21% of users are active and take at least 10,000 steps per day.
- 21% of users are inactive and take less than 5,000 steps per day.
- On average, users sleep for 377.6 minutes per night.
- 50% of users get at least 420 minutes of sleep per night.
- There is a strong negative correlation between SedentaryMinutes and TotalMinutesAsleep.
- On average, it takes users 38.83 minutes to fall asleep.
- Most of the users are between inactive and lightly active.
- The more time users stay active, the more calories they burn.
- More than half of the users (56%) get sufficient sleep (at least 420 minutes).
- As user activity level decreases, users are less likely to get sufficient sleep.
- The longer users stay awake in bed, the fewer users sleep.

Recommendations

Upgrades

==> **Improve machine learning.**

- Continue to improve on collecting accurate data to encourage users to wear their Bellabeat tracker more and take control of their own fitness and health journey.

==> **Improve wearability**

- Ensure the Bellabeat tracker is comfortable while staying trendy to wear for users to encourage logging participation.

New Features

==> **Allow users to set up a wake-up and sleep buzz alert.**

- Sending a reminder for bedtime will encourage more sleep by reducing screen time. According to a poll, “95% of people said they regularly use some type of electronics within an hour of bedtime.” 8 (https://www.sleepfoundation.org/bedroom-environment/technology-in-the-bedroom) We can get lost in our electronics at night before bed and feel not tired because we’re stimulated.

==> **Meditation, breathing, and other mindful exercises.**

- Meditation and breathing exercises can improve the quality of your sleep. It can help users who are sleeping less or have trouble sleeping relax and bring calmness 9 (https://www.sleepfoundation.org/insomnia/treatment/meditation). This new feature in the Time Watch will incorporate guided breathing exercises that users can pick from throughout the day.

User Growth

==> **Connect to users’ medical records.**

- Bellabeat can personalize each user’s health journey by connecting to their medical records. That way, doctors or providers can better track your health and self-manage their health. Bellabeat can expand in the healthcare industry by working with healthcare insurance companies to use Bellabeat for patients that need regular fitness tracking 10 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6746089/).

==> **Rewards point system.**

- Bellabeat can create a rewards point system for unlocking achievements, upholding streaks, or even completing a fun fitness challenge. As users gather points, they can trade them in for rewards such as free workout programs, deals with active equipment, and much more.

Next Up…

There is additional data in the dataset to analyze to provide more extensive insights such as users’ logs on weight, intensities, heart rate, and METs.

Sources

[1] How Many Steps a Day Is Considered Active? (https://www.medicinenet.com/how_many_steps_a_day_is_considered_active/article.htm)

[2] Healthy Sleep: What Is It and Are You Getting It? (https://www.sleepfoundation.org/sleep-hygiene/what-is-healthy-sleep)

[3] How many steps/day are enough? for adults - International Journal of Behavioral Nutrition and Physical Activity (https://ijbnpa.biomedcentral.com/articles/10.1186/1479-5868-8-79#:~:text=In%20summary%2C%20at%20least%20in,populace%20%5B3%2C%2023%5D)

[4] How Long Does It Take to Fall Asleep? Average Time and Tips (https://www.healthline.com/health/healthy-sleep/how-long-does-it-take-to-fall-asleep#:~:text=It%20should%20take%20between%2010,fall%20asleep%20much%20more%20quickly)

[5] Bidirectional Relationship Between Exercise and Sleep: Implications for Exercise Adherence and Sleep Improvement - Christopher E. Kline, 2014 (https://journals.sagepub.com/doi/abs/10.1177/1559827614544437)

[6] Are Sedentary Behaviors Associated with Sleep Duration? A Cross-Sectional Case from Croatia (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6352043/)

[7] Associations between physical activity, sedentary time, sleep duration and daytime sleepiness in US adults (https://www.sciencedirect.com/science/article/abs/pii/S0091743514002035)

[8] Technology in the Bedroom (https://www.sleepfoundation.org/bedroom-environment/technology-in-the-bedroom)

[9] Can Meditation Treat Insomnia (https://www.sleepfoundation.org/insomnia/treatment/meditation)

[10] Wearable Health Technology and Electronic Health Record Integration: Scoping Review and Future Directions (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6746089/)