

# Bellabeat Case Study in R

Sahil Chikode

2025-08-08

## About Company

Urška Srsen and Sando Mur founded Bellabeat, a high-tech company that manufactures health-focused smart products. Srsen 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 [click here](#).

## Question for the analysis

1. What are some trends in smart device usage?
2. How could these trends apply to Bellabeat customers?
3. How could these trends help in influence Bellabeat marketing strategy?

## Loading Packages

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.5.2      v tibble    3.3.0
## v lubridate  1.9.4      v tidyr     1.3.1
## v purrr      1.1.0
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()     masks stats::lag()
```

```
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(dplyr)
library(ggplot2)
library(tidyr)
library(janitor)
```

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

```
library(skimr)
```

## Importing dataset using readr()

```
daily_activity <- read.csv("C:/Users/Admin/OneDrive/Desktop/PROJECTS/Bellabeat/Fitabase Data/dailyActivity_merged.csv")
sleep <- read.csv("C:/Users/Admin/OneDrive/Desktop/PROJECTS/Bellabeat/Fitabase Data/SleepDay_merged.csv")
weight <- read.csv("C:/Users/Admin/OneDrive/Desktop/PROJECTS/Bellabeat/Fitabase Data/weightLogInfo_merged.csv")
calories <- read.csv("C:/Users/Admin/OneDrive/Desktop/PROJECTS/Bellabeat/Fitabase Data/hourlyCalories_merged.csv")
intensities <- read.csv("C:/Users/Admin/OneDrive/Desktop/PROJECTS/Bellabeat/Fitabase Data/hourlyIntensities_merged.csv")
```

```
head(daily_activity)
```

```
##           Id ActivityDate TotalSteps TotalDistance TrackerDistance
## 1 1503960366   3/25/2016     11004           7.11           7.11
## 2 1503960366   3/26/2016     17609          11.55          11.55
## 3 1503960366   3/27/2016     12736           8.53           8.53
## 4 1503960366   3/28/2016     13231           8.93           8.93
## 5 1503960366   3/29/2016     12041           7.85           7.85
## 6 1503960366   3/30/2016     10970           7.16           7.16
##   LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 1                        0                2.57                   0.46
## 2                        0                6.92                   0.73
## 3                        0                4.66                   0.16
## 4                        0                3.19                   0.79
## 5                        0                2.16                   1.09
## 6                        0                2.36                   0.51
##   LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
## 1                4.07                   0                33
## 2                3.91                   0                89
## 3                3.71                   0                56
## 4                4.95                   0                39
## 5                4.61                   0                28
## 6                4.29                   0                30
##   FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories
## 1                 12                205                804     1819
## 2                 17                274                588     2154
## 3                  5                268                605     1944
## 4                 20                224               1080     1932
## 5                 28                243                763     1886
## 6                 13                223               1174     1820
```

```
head(sleep)
```

```
##           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
```

## Fixing formatting of the date

```
# intensities
intensities$ActivityHour=as.POSIXct(intensities$ActivityHour, format="%m/%d/%Y %I:%M:%S %p", tz=Sys.time())
intensities$time <- format(intensities$ActivityHour, format = "%H:%M:%S")
intensities$date <- format(intensities$ActivityHour, format = "%m/%d/%y")

# calories
calories$ActivityHour=as.POSIXct(calories$ActivityHour, format="%m/%d/%Y %I:%M:%S %p", tz=Sys.timezone())
calories$time <- format(calories$ActivityHour, format = "%H:%M:%S")
calories$date <- format(calories$ActivityHour, format = "%m/%d/%y")

# activity
daily_activity$ActivityDate=as.POSIXct(daily_activity$ActivityDate, format="%m/%d/%Y", tz=Sys.timezone())
daily_activity$date <- format(daily_activity$ActivityDate, format = "%m/%d/%y")

# sleep
sleep$SleepDay=as.POSIXct(sleep$SleepDay, format="%m/%d/%Y %I:%M:%S %p", tz=Sys.timezone())
sleep$date <- format(sleep$SleepDay, format = "%m/%d/%y")

# weight
weight$Date=as.POSIXct(weight$Date, format="%m/%d/%Y %I:%M:%S %p", tz=Sys.timezone())
weight$time <- format(weight$Date, format = "%H:%M:%S")
weight$date <- format(weight$Date, format = "%m/%d/%y")
```

The format of data and time is not consistent accross all the dataset. So, I can converted the into single format.

## Summarizing the data.

```
skim_without_charts(daily_activity)
```

Table 1: Data summary

Name	daily_activity
Number of rows	457
Number of columns	16

Column type frequency:

character	1
numeric	14
POSIXct	1

Group variables	None
-----------------	------

#### Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
date	0	1	8	8	0	32	0

#### Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Id	0	1	4.628595e-20293781e+00	3960366347168e+09	57193e+03	91747e+03	77689e+09		
TotalSteps	0	1	6.546560e-50398490e+03	0	1.988000e+50386000e+03	19800e+03	419700e+04		
TotalDistance	0	1	4.660000e-40080000e+00	0	1.410000e+40090000e+00	60000e+00	2753000e+01		
TrackerDistance	0	1	4.610000e-40070000e+00	0	1.280000e+40090000e+00	10000e+00	2753000e+01		
LoggedActivitiesDistance	0	1	1.800000e-8.500000e-01	0	0.000000e+00000000e+00	0000000e+00	6730000e+00		
VeryActiveDistance	0	1	1.180000e-20090000e+00	0	0.000000e+00000000e+00	310000e+00	2092000e+01		
ModeratelyActiveDistance	0	1	4.800000e-8.300000e-01	0	0.000000e+20000000e+00	670000e+00	6.400000e+00		
LightActiveDistance	0	1	2.890000e-20240000e+00	0	8.700000e-2.930000e+00	4060000e+00	1251000e+01		
SedentaryActiveDistance	0	1	0.000000e-00000000e+00	0	0.000000e+00000000e+00	0000000e+00	1000000e+00		
VeryActiveMinutes	0	1	1.662000e-20892000e+01	0	0.000000e+00000000e+00	500000e+00	2012000e+02		
FairlyActiveMinutes	0	1	1.307000e-30621000e+01	0	0.000000e+00000000e+00	600000e+00	6000000e+02		
LightlyActiveMinutes	0	1	1.700700e-10222100e+02	0	6.400000e+0810000e+02	2570000e+02	7020000e+02		
SedentaryMinutes	0	1	9.952800e-30370200e+02	32	7.280000e+0257000e+02	285000e+03	1340000e+03		
Calories	0	1	2.189450e-80354800e+02	0	1.776000e+20362000e+02	2667000e+03	4562000e+03		

#### Variable type: POSIXct

skim_variable	n_missing	complete_rate	min	max	median	n_unique
ActivityDate	0	1	2016-03-12	2016-04-12	2016-04-05	32

`skim_without_charts(sleep)`

Table 5: Data summary

Name	sleep
Number of rows	413
Number of columns	6

Column type frequency:	
character	1
numeric	4
POSIXct	1
<hr/>	
Group variables	None
<hr/>	

#### Variable type: character

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

#### Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Id	0	1	5.000979e+00	0.6036e+01	0.0396036e+01	0.39773337e+01	0.70292168e+01	0.96218106e+01	0.97920096e+01
TotalSleepRecords	0	1	1.120000e+00	0.50000e-01	1	1	1	1	3
TotalMinutesAsleep	0	1	4.194700e+02	2.18340e+02	58	361	433	490	796
TotalTimeInBed	0	1	4.586400e+02	2.27100e+02	61	403	463	526	961

#### Variable type: POSIXct

skim_variable	n_missing	complete_rate	min	max	median	n_unique
SleepDay	0	1	2016-04-12	2016-05-12	2016-04-27	31

```
skim_without_charts(weight)
```

Table 9: Data summary

Name	weight
Number of rows	33
Number of columns	10
<hr/>	
Column type frequency:	
character	3
numeric	6
POSIXct	1
<hr/>	
Group variables	None
<hr/>	

#### Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
IsManualReport	0	1	4	5	0	2	0

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
time	0	1	8	8	0	11	0
date	0	1	8	8	0	14	0

#### Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Id	0	1.00	6.477156e+23	0.8888e+09	0.03960e+09	0.702922e+09	0.962181e+09	0.977689e+09	0.977689e+09
WeightKg	0	1.00	7.344000e+01	5.30000e+01	0.30000e+01	6.170000e+01	6.250000e+01	8.580000e+01	1.296000e+02
WeightPounds	0	1.00	1.619100e+02	3.44000e+01	0.175100e+02	0.260300e+02	0.277900e+02	0.291600e+02	0.257200e+02
Fat	31	0.06	1.600000e+01	1.90000e+00	0.000000e+00	0.300000e+00	0.600000e+00	0.900000e+00	2.200000e+01
BMI	0	1.00	2.573000e+01	0.330000e+01	0.45000e+00	0.110000e+01	0.139000e+01	0.1576000e+01	0.617000e+01
LogId	0	1.00	1.459959e+12	0.88072e+08	0.59382e+08	1.259753e+08	1.259987e+08	1.260160e+08	1.260506e+12

#### Variable type: POSIXct

skim_variable	n_missing	complete_rate	min	max	median	n_unique
Date	0	1	2016-03-30 23:59:59	2016-04-12 23:59:59	2016-04-06 23:59:59	24

```
skim_without_charts(calories)
```

Table 13: Data summary

Name	calories
Number of rows	24084
Number of columns	5
Column type frequency:	
character	2
numeric	2
POSIXct	1
Group variables	None

#### Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
time	0	1	8	8	0	24	0
date	0	1	8	8	0	32	0

#### Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Id	0	1	4.889424e+09	2.156581e+09	0.0396036e+09	0.34716779e+09	0.55860992e+09	0.96218106e+09	0.977689391e+09

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Calories	0	1	9.427000e+01	59.4	42	61	77	104	933

**Variable type: POSIXct**

skim_variable	n_missing	complete_rate	min	max	median	n_unique
ActivityHour	0	1	2016-03-12	2016-04-12 10:00:00	2016-03-27 04:00:00	755

```
skim_without_charts(intensities)
```

Table 17: Data summary

Name	intensities
Number of rows	24084
Number of columns	6
Column type frequency:	
character	2
numeric	3
POSIXct	1
Group variables	None

**Variable type: character**

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
time	0	1	8	8	0	24	0
date	0	1	8	8	0	32	0

**Variable type: numeric**

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Id	0	1	4.889424e+01	2.1566e+01	0.0396036e+01	3.4716779e+01	6.55861e+01	9.62181e+01	8.77689391e+01
TotalIntensity	0	1	1.083000e+01	0.31000e+01	0	0	1.00000e+01	4.00000e+01	180
AverageIntensity	0	1	1.800000e-01	3.400000e-01	0	0	2.00000e-02	2.300000e-01	3

**Variable type: POSIXct**

skim_variable	n_missing	complete_rate	min	max	median	n_unique
ActivityHour	0	1	2016-03-12	2016-04-12 10:00:00	2016-03-27 04:00:00	755

Using library skimr, summarize each dataset without chart.

```
n_distinct(daily_activity$Id)
```

```
## [1] 35
```

```
n_distinct(sleep$Id)
```

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

```
n_distinct(weight$Id)
```

```
## [1] 11
```

```
n_distinct(intensities$Id)
```

```
## [1] 34
```

```
n_distinct(calories$Id)
```

```
## [1] 34
```

This is about number participants in each data sets. There are 35 participants in daily activity data set, 34 participants in intensities and calories data set, 24 in sleep data set and 11 at weight data set.

Now summary statistics of all the data sets.

```
daily_activity %>% select(Id, TotalSteps, TotalDistance, SedentaryMinutes) %>%  
  summary()
```

```
##           Id           TotalSteps   TotalDistance   SedentaryMinutes  
##  Min.      :1.504e+09   Min.      :    0   Min.      : 0.000   Min.      : 32.0  
## 1st Qu.:2.347e+09   1st Qu.: 1988   1st Qu.: 1.410   1st Qu.: 728.0  
## Median :4.057e+09   Median : 5986   Median : 4.090   Median :1057.0  
## Mean   :4.629e+09   Mean   : 6547   Mean   : 4.664   Mean   : 995.3  
## 3rd Qu.:6.392e+09   3rd Qu.:10198   3rd Qu.: 7.160   3rd Qu.:1285.0  
## Max.   :8.878e+09   Max.   :28497   Max.   :27.530   Max.   :1440.0
```

```
avg_indival <- daily_activity %>% group_by(Id) %>% summarise(avg_steps = mean(TotalSteps),  
  avg_distance = mean(TotalDistance),  
  avg_SedentaryActiveDistance = mean(SedentaryActiveDistance))  
print(avg_indival, n = 35) #Prints each row of the output
```

```
## # A tibble: 35 x 4
```

```
##           Id avg_steps avg_distance avg_SedentaryActiveDistance  
##      <dbl>    <dbl>      <dbl>                <dbl>  
## 1 1503960366  11641.        7.61                  0  
## 2 1624580081   4226.        2.75                0.00526  
## 3 1644430081   9275.        6.75                0.00800  
## 4 1844505072   3641.        2.41                  0  
## 5 1927972279   2181.        1.51                  0
```



```
## 6 2022484408 12175.      8.77      0
## 7 2026352035 3393.      2.10      0
## 8 2320127002 3138.      2.12      0
## 9 2347167796 9800.      6.51      0
## 10 2873212765 6637.      4.47      0.00167
## 11 2891001357 774.      0.604      0
## 12 3372868164 6128.      4.22      0.0140
## 13 3977333714 8664.      5.81      0
## 14 4020332650 5777.      4.14      0.00625
## 15 4057192912 1887.      1.39      0.00187
## 16 4319703577 7821.      5.26      0
## 17 4388161847 0      0      0
## 18 4445114986 4293.      2.91      0
## 19 4558609924 5785.      3.82      0
## 20 4702921684 7943.      6.45      0
## 21 5553957443 8355.      5.46      0
## 22 5577150313 8608.      6.45      0
## 23 6117666160 8249.      6.23      0
## 24 6290855005 1618.      1.22      0
## 25 6391747486 1337.      1.07      0
## 26 6775888955 5559      3.99      0
## 27 6962181067 12640.     8.65      0.00714
## 28 7007744171 12260.     8.86      0.00417
## 29 7086361926 6104.      4.09      0.00500
## 30 8053475328 14844.     11.6      0
## 31 8253242879 2390.      1.68      0.00250
## 32 8378563200 8135.      6.45      0
## 33 8583815059 3046.      2.38      0
## 34 8792009665 3095.      1.98      0
## 35 8877689391 17417.     14.1      0.00250
```

```
daily_activity %>%
  select(VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes) %>%
  summary()
```

```
## VeryActiveMinutes FairlyActiveMinutes LightlyActiveMinutes
## Min. : 0.00 Min. : 0.00 Min. : 0.0
## 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 64.0
## Median : 0.00 Median : 1.00 Median :181.0
## Mean : 16.62 Mean : 13.07 Mean :170.1
## 3rd Qu.: 25.00 3rd Qu.: 16.00 3rd Qu.:257.0
## Max. :202.00 Max. :660.00 Max. :720.0
```

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

```
## TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
## Min. :1.000 Min. : 58.0 Min. : 61.0
## 1st Qu.:1.000 1st Qu.:361.0 1st Qu.:403.0
## Median :1.000 Median :433.0 Median :463.0
## Mean :1.119 Mean :419.5 Mean :458.6
## 3rd Qu.:1.000 3rd Qu.:490.0 3rd Qu.:526.0
## Max. :3.000 Max. :796.0 Max. :961.0
```

```
weight %>% select(WeightKg, WeightPounds, BMI) %>%
  summary()
```

```
##      WeightKg      WeightPounds      BMI
##  Min.   : 53.30   Min.   :117.5   Min.   :21.45
## 1st Qu.: 61.70   1st Qu.:136.0   1st Qu.:24.10
##  Median : 62.50   Median :137.8   Median :24.39
##  Mean   : 73.44   Mean   :161.9   Mean   :25.73
## 3rd Qu.: 85.80   3rd Qu.:189.2   3rd Qu.:25.76
##  Max.   :129.60   Max.   :285.7   Max.   :46.17
```

```
weight %>% select(Id, WeightKg) %>% filter(WeightKg > 95)
```

```
##           Id WeightKg
## 1 1927972279    129.6
## 2 4702921684     99.7
```

```
weight %>% select(Id, WeightKg, BMI) %>% filter(BMI > 30)
```

```
##           Id WeightKg  BMI
## 1 1927972279    129.6 46.17
## 2 4445114986     92.4 35.01
```

```
weight %>% select(Id, WeightKg, BMI) %>% filter(BMI >= 25 & BMI < 29.9)
```

```
##           Id WeightKg  BMI
## 1 2891001357     88.4 25.03
## 2 4558609924     69.4 27.14
## 3 4702921684     99.7 26.11
## 4 8253242879     75.6 29.55
## 5 8877689391     85.5 25.61
## 6 8877689391     86.6 25.94
## 7 8877689391     86.0 25.76
## 8 8877689391     86.3 25.83
## 9 8877689391     85.1 25.49
## 10 8877689391     85.0 25.44
## 11 8877689391     85.4 25.56
## 12 8877689391     86.1 25.79
## 13 8877689391     85.8 25.68
```

```
calories %>% select(Calories) %>% summary()
```

```
##      Calories
##  Min.   : 42.00
## 1st Qu.: 61.00
##  Median : 77.00
##  Mean   : 94.27
## 3rd Qu.:104.00
##  Max.   :933.00
```

```
intensities %>% select(TotalIntensity) %>% summary()
```

```
## TotalIntensity
## Min.      : 0.00
## 1st Qu.: 0.00
## Median : 1.00
## Mean   : 10.83
## 3rd Qu.: 14.00
## Max.    :180.00
```

### Insights from summary :

1.Average Steps: 6547, Average Distance: 4.664,  
Average Sedentary Minutes :995.3 min.

- Suggestions : Steps of users should increase, Average steps should be more than 8000. Sedentary min. is 995.3 min. means 16 hours, it should be reduced.

2.The majority of the participants are lightly active.

3.On average, participants sleep 1 time for 7 hours.

- Suggestions : Participants takes good amount of sleep, we should consider using notification to go to sleep.

4.20 users are normal weight, 13 are overweight and 2 are obese.

- Suggestions : I recommend is that according to their BMI we should display there weight category in app. And if category is overweight and obese we should notify to control diet and do exercise.

## Merging columns

```
combined_data <- merge(sleep, daily_activity, by = 'Id', all = TRUE) %>%
  select(-date.x, -date.y, -SedentaryActiveDistance)
head(combined_data)
```

```
##           Id  SleepDay TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
## 1 1503960366 2016-04-12                1                 327           346
## 2 1503960366 2016-04-12                1                 327           346
## 3 1503960366 2016-04-12                1                 327           346
## 4 1503960366 2016-04-12                1                 327           346
## 5 1503960366 2016-04-12                1                 327           346
## 6 1503960366 2016-04-12                1                 327           346
## ActivityDate TotalSteps TotalDistance TrackerDistance
## 1 2016-04-09      12432          8.10          8.10
## 2 2016-04-12        224          0.14          0.14
## 3 2016-04-10     10057          6.98          6.98
## 4 2016-03-26     17609         11.55         11.55
## 5 2016-04-08     12521          7.94          7.94
## 6 2016-03-27     12736          8.53          8.53
```

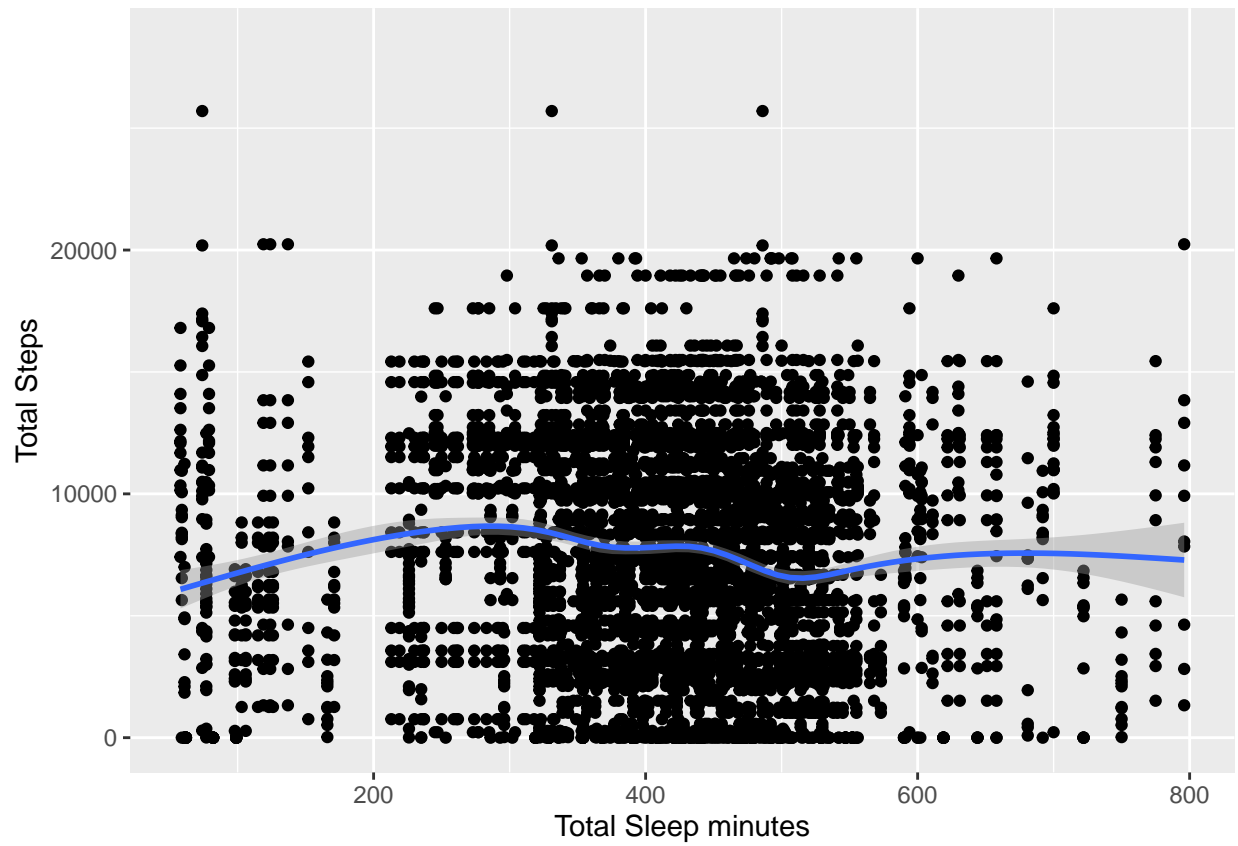
```
##   LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 1                        0                2.59                    0.59
## 2                        0                0.00                    0.00
## 3                        0                4.00                    0.49
## 4                        0                6.92                    0.73
## 5                        0                3.31                    0.90
## 6                        0                4.66                    0.16
##   LightActiveDistance VeryActiveMinutes FairlyActiveMinutes
## 1                4.92                32                15
## 2                0.13                0                 0
## 3                2.48                44                13
## 4                3.91                89                17
## 5                3.74                46                22
## 6                3.71                56                 5
##   LightlyActiveMinutes SedentaryMinutes Calories
## 1                248                738        1883
## 2                 9                 32         50
## 3               168                737        1755
## 4               274                588        2154
## 5               212               1160        1895
## 6               268                605        1944
```

```
n_distinct(combined_data$Id)
```

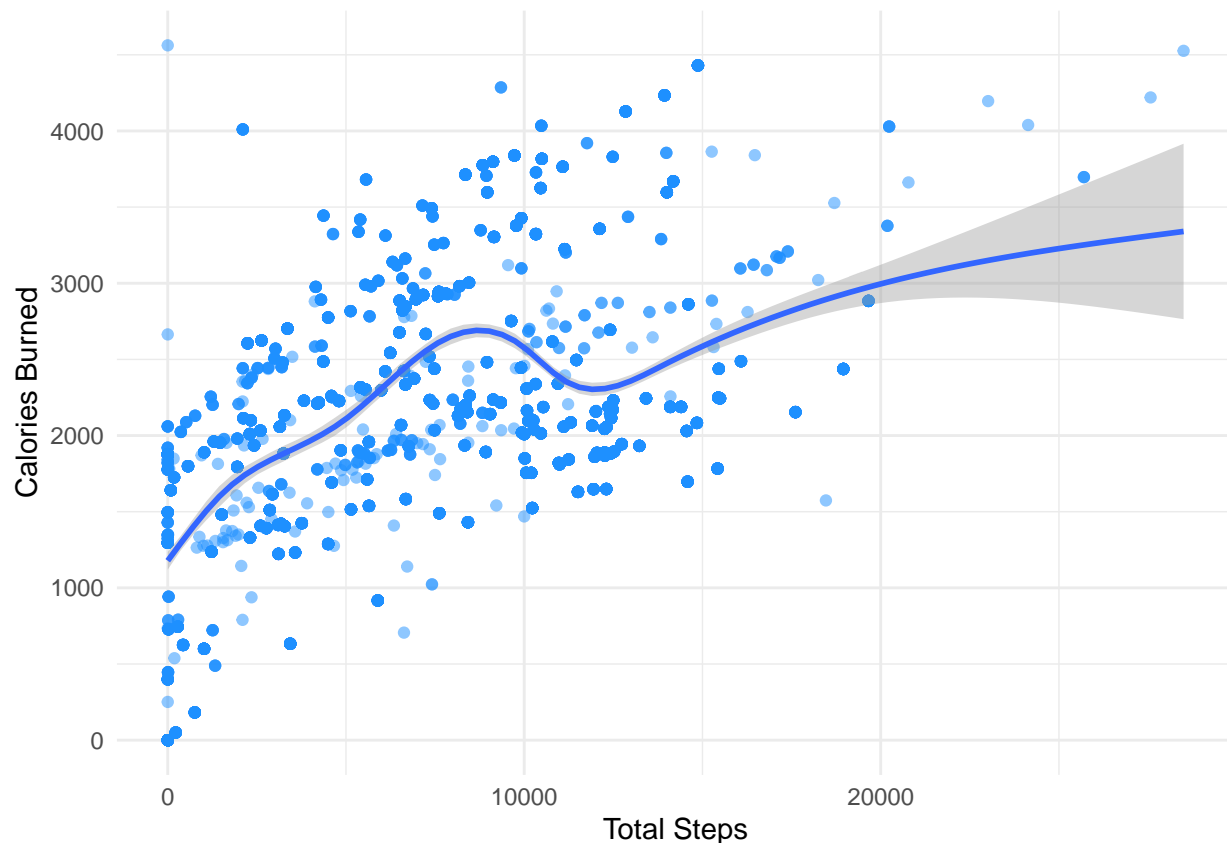
```
## [1] 35
```

Daily activity and sleep data sets are merged for more detailed analysis. I have used outer join to merge data sets, which will include all participants.

## Visualizing Data



- This chart indicates that people who sleep in the 6-7.5 hrs tend to take most steps, on average.
- Most of users sleep between 5-10 hrs and steps between 5000-15000 steps.



- As steps increase, more calories gets burnt.

```
sleep_analysis <- combined_data %>% filter(!is.na(TotalTimeInBed)) %>%
  mutate(SleepEfficiency = TotalMinutesAsleep / TotalTimeInBed * 100) %>%
  arrange(Id, ActivityDate) %>% group_by(Id)
sleep_analysis
```

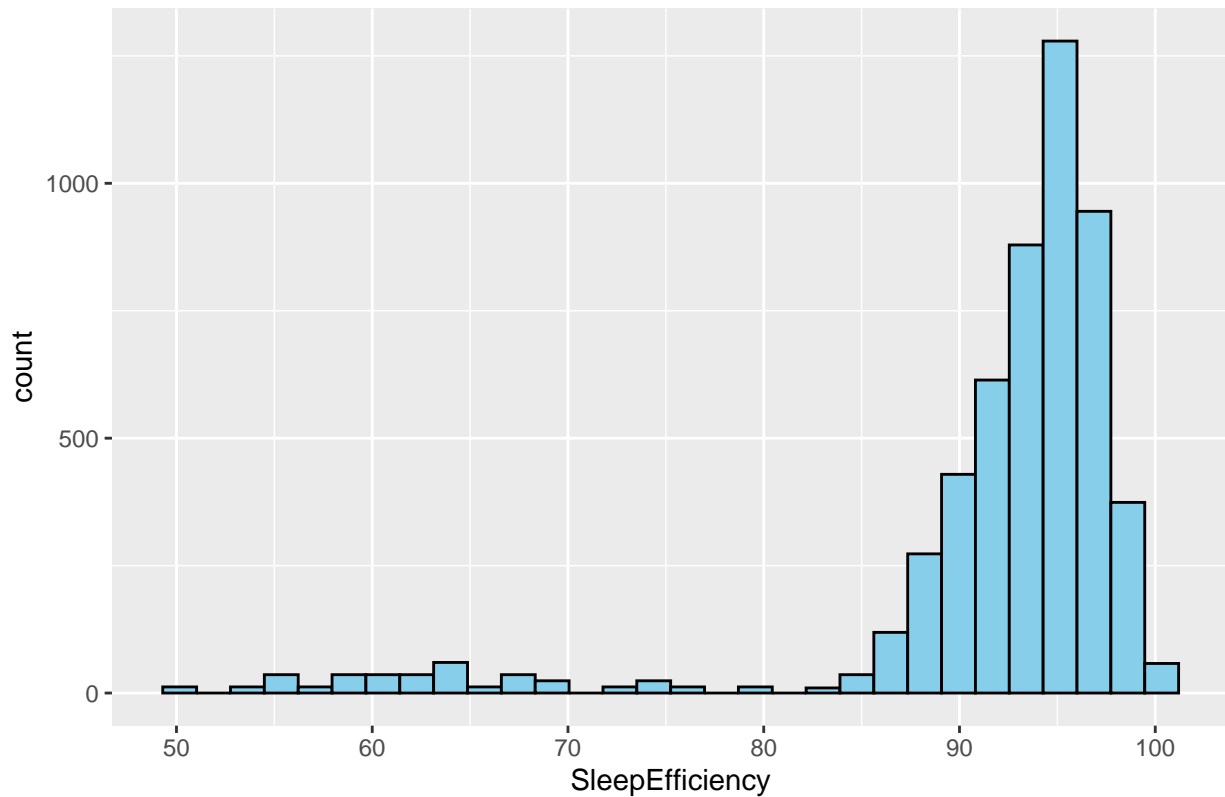
```
## # A tibble: 5,388 x 19
## # Groups:   Id [24]
##       Id SleepDay      TotalSleepRecords TotalMinutesAsleep
##       <dbl> <dtm>              <int>              <int>
## 1 1503960366 2016-04-12 00:00:00             1             327
## 2 1503960366 2016-04-13 00:00:00             2             384
## 3 1503960366 2016-04-15 00:00:00             1             412
## 4 1503960366 2016-04-16 00:00:00             2             340
## 5 1503960366 2016-04-17 00:00:00             1             700
## 6 1503960366 2016-04-19 00:00:00             1             304
## 7 1503960366 2016-04-20 00:00:00             1             360
## 8 1503960366 2016-04-21 00:00:00             1             325
## 9 1503960366 2016-04-23 00:00:00             1             361
## 10 1503960366 2016-04-24 00:00:00             1             430
## # i 5,378 more rows
## # i 15 more variables: TotalTimeInBed <int>, ActivityDate <dtm>,
## #   TotalSteps <int>, TotalDistance <dbl>, TrackerDistance <dbl>,
## #   LoggedActivitiesDistance <dbl>, VeryActiveDistance <dbl>,
```

```
## # ModeratelyActiveDistance <dbl>, LightActiveDistance <dbl>,
## # VeryActiveMinutes <int>, FairlyActiveMinutes <int>,
## # LightlyActiveMinutes <int>, SedentaryMinutes <int>, Calories <int>, ...
```

- Calculating sleep efficiency, arranging through id, date and grouping by Id. `filter(!is.na())` is used keep only rows where `TotalTimeinBed` column has a number.

Understanding sleep quality of the users.

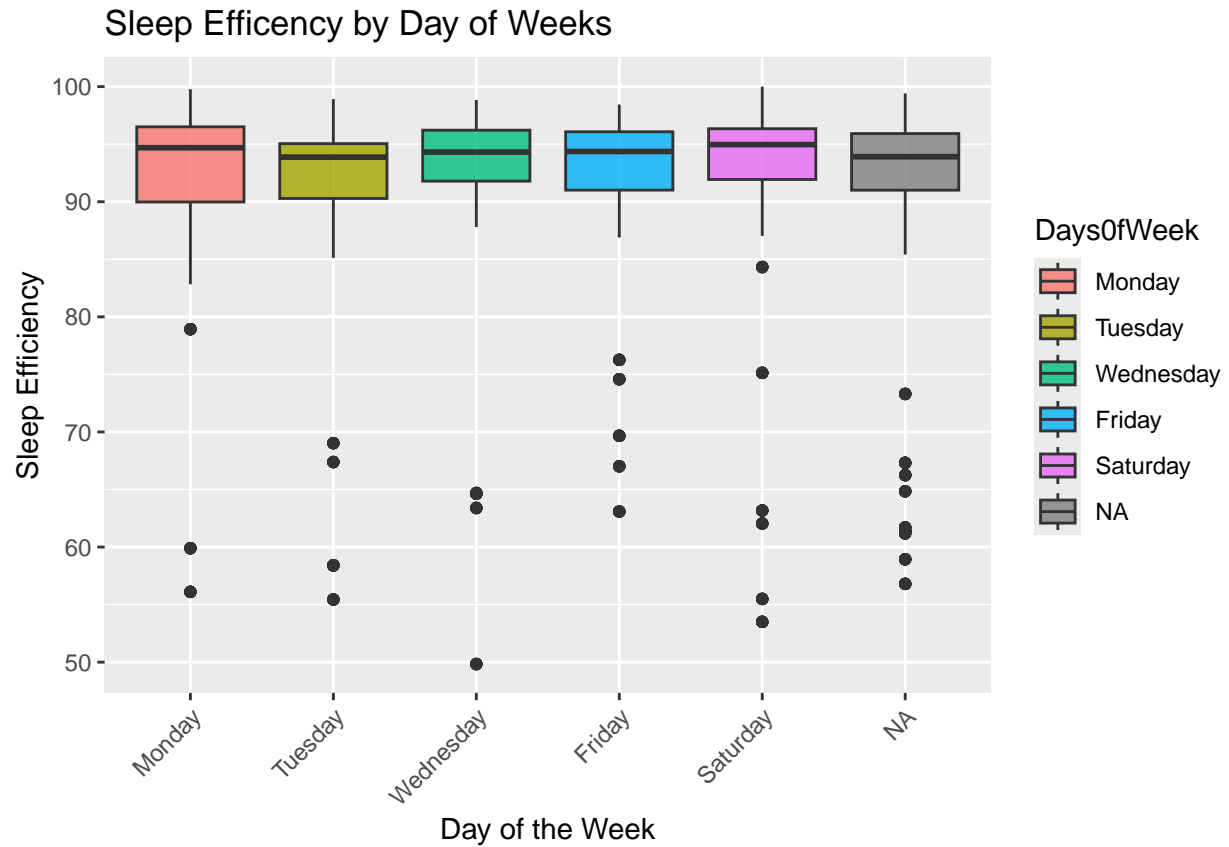
### Sleeping Efficiency (Total Sleep / Total Time in Bed)



- Overall many users sleep quality is good and had efficient sleep.

```
new_sleep_analysis <- sleep_analysis %>% mutate(DaysOfWeek = factor(weekdays(SleepDay), levels = c('Mon', 'Tue', 'Wed', 'Thu', 'Fri', 'Sat', 'Sun')))
```

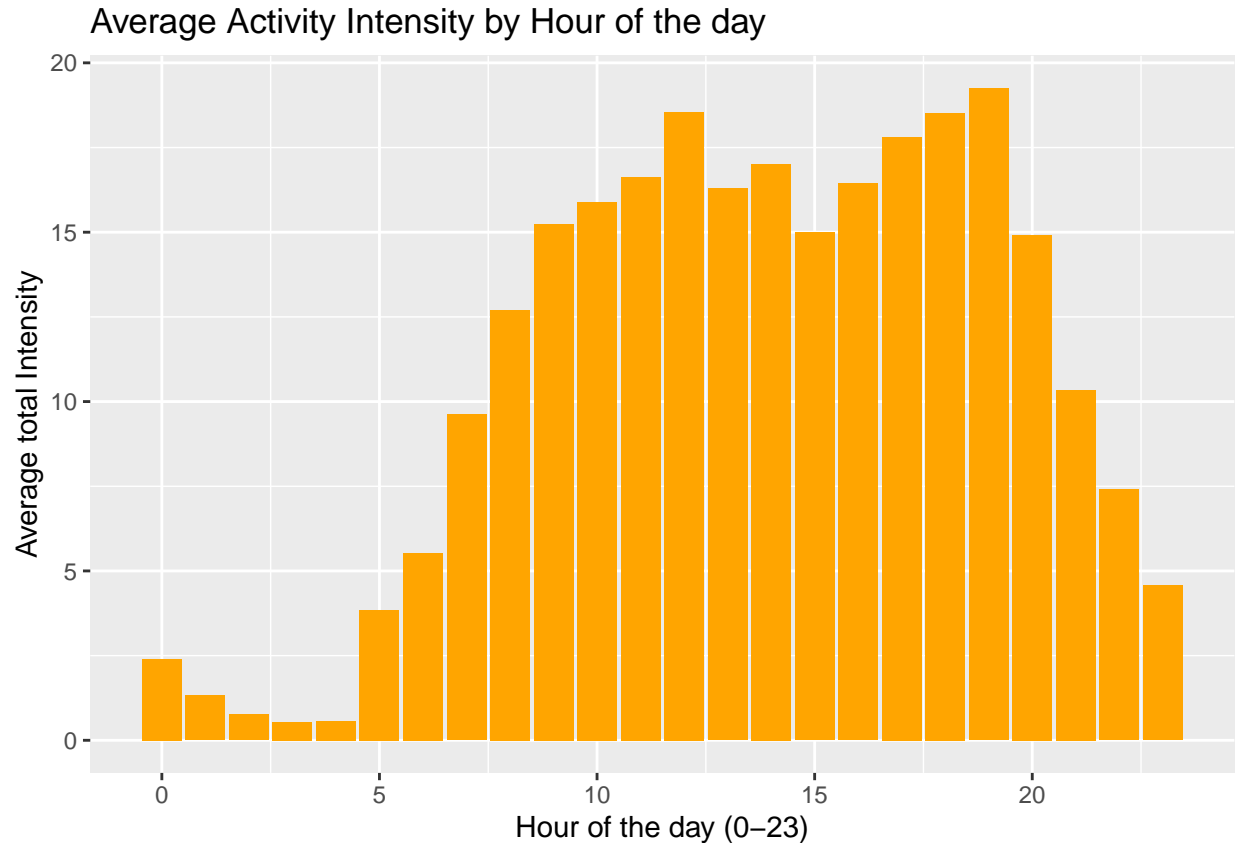
A new column, Day of week was added to analyse the participants sleep efficiency across different days of the week.



```
new_intensities <- intensities %>% mutate(HourofDay = as.numeric(format(ActivityHour, "%H")))
```

A new column, Hour of Day was added to analyse the participants energy intensities across different hours of the day.





- The visual shows that people are active between 5 a.m. to 11 p.m.

## Summary of the Business Case Study.

Analyzing FitBit Fitness Tracker Data, can help BellaBeat to understand the audience and implement market strategy.

### Target audience

Women who work full-time jobs and spend a lot of time at the computer, in a meeting, focused on work.

These women do some light activity to stay healthy. Even though they need to improve their everyday activity to have health benefits. They might need some knowledge about developing healthy habits or motivation to keep going.

- As there is no gender information about the participants, I assumed that all genders were presented and balanced in this data set.

### Recommendation for the Bellabeat.

1. Steps of users should increase, Average steps should be more than 8000, taking 8000 steps per day was associated with a 51% lower risk for all-cause mortality.
2. Participants takes good amount of sleep, Bellabeat should consider using notification to go to sleep.
3. I recommend is that according to user's BMI Bellabeat should display there weight category in app. And if category is overweight and obese Bellabeat should suggest to control diet and do exercise.

4. Sedentary min. is 995.3 min. means 16 hours, it should be reduced. Bellabeat should set some time for sedentary minutes, so after crossing it app will notify users to reduce their sedentary time.