

Fitbit Daily Activity

Jose Juan Lopez

2024-01-25

Setting up my environment

Load libraries.

```
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   1.5.0
## v ggplot2    3.4.4      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.1
## v purrr      1.0.1
```

```
## -- 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(here)
```

```
## here() starts at /Users/jjlof/Documents/Data Analyst/Bellabeat/R Data
```

```
library(skimr)
```

```
library(janitor)
```

```
##
```

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

```
##
```

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

```
##
```

```
##      chisq.test, fisher.test
```

```
library(readr)
```

```
library(scales)
```

```
##
```

```
## Attaching package: 'scales'
```

```
##
```

```
## The following object is masked from 'package:purrr':
```

```
##
```

```
##      discard
```

```
##
```

```
## The following object is masked from 'package:readr':
```

```
##
```

```
##      col_factor
library(gridExtra)

##
## Attaching package: 'gridExtra'
##
## The following object is masked from 'package:dplyr':
##
##      combine
library(lubridate)
library(grid)
library(gridExtra)
library(magrittr)

##
## Attaching package: 'magrittr'
##
## The following object is masked from 'package:purrr':
##
##      set_names
##
## The following object is masked from 'package:tidyr':
##
##      extract
library(patchwork)
```

Read documents

In this analysis I choose to analyze:

- DailyActivityMerged_2020_05_12.csv
- DailyCaloriesMerged_2020_05_12.csv
- DailyIntensitiesMerged_2020_05_12.csv
- DailyStepsMerged_2020_05_12.csv
- SleepDayMerged_2020_05_12.csv

I consider these databases to help to understand the behavior of people's devices and identify the new way which needs **marketing strategy** of Bellabeat.

Showing daily activity data

```
dActivity <- read.csv('~/Documents/Data Analyst/Bellabeat/Q2-Fitbit-Data/DailyActivityMerged_2020_05_12.csv')
head(dActivity)
```

```
##      Id ActivityDate TotalSteps TotalDistance TrackerDistance
## 1 1503960366 4/12/2016    13162         8.50           8.50
## 2 1503960366 4/13/2016    10735         6.97           6.97
## 3 1503960366 4/14/2016    10460         6.74           6.74
## 4 1503960366 4/15/2016     9762         6.28           6.28
## 5 1503960366 4/16/2016    12669         8.16           8.16
## 6 1503960366 4/17/2016     9705         6.48           6.48
##      LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 1                        0              1.88              0.55
## 2                        0              1.57              0.69
```

```
## 3          0          2.44          0.40
## 4          0          2.14          1.26
## 5          0          2.71          0.41
## 6          0          3.19          0.78
##   LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
## 1          6.06          0          25
## 2          4.71          0          21
## 3          3.91          0          30
## 4          2.83          0          29
## 5          5.04          0          36
## 6          2.51          0          38
##   FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories
## 1          13          328          728    1985
## 2          19          217          776    1797
## 3          11          181         1218    1776
## 4          34          209          726    1745
## 5          10          221          773    1863
## 6          20          164          539    1728
```

```
str(dActivity)
```

```
## 'data.frame':   940 obs. of  15 variables:
## $ Id          : num  1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
## $ ActivityDate : chr   "4/12/2016" "4/13/2016" "4/14/2016" "4/15/2016" ...
## $ TotalSteps   : int   13162 10735 10460 9762 12669 9705 13019 15506 10544 9819 ...
## $ TotalDistance : num   8.5 6.97 6.74 6.28 8.16 ...
## $ TrackerDistance : num   8.5 6.97 6.74 6.28 8.16 ...
## $ LoggedActivitiesDistance: num   0 0 0 0 0 0 0 0 0 0 ...
## $ VeryActiveDistance : num   1.88 1.57 2.44 2.14 2.71 ...
## $ ModeratelyActiveDistance: num   0.55 0.69 0.4 1.26 0.41 ...
## $ LightActiveDistance : num   6.06 4.71 3.91 2.83 5.04 ...
## $ SedentaryActiveDistance : num   0 0 0 0 0 0 0 0 0 0 ...
## $ VeryActiveMinutes : int   25 21 30 29 36 38 42 50 28 19 ...
## $ FairlyActiveMinutes : int   13 19 11 34 10 20 16 31 12 8 ...
## $ LightlyActiveMinutes : int   328 217 181 209 221 164 233 264 205 211 ...
## $ SedentaryMinutes : int   728 776 1218 726 773 539 1149 775 818 838 ...
## $ Calories       : int   1985 1797 1776 1745 1863 1728 1921 2035 1786 1775 ...
```

```
colnames(dActivity)
```

```
## [1] "Id"          "ActivityDate"
## [3] "TotalSteps"  "TotalDistance"
## [5] "TrackerDistance" "LoggedActivitiesDistance"
## [7] "VeryActiveDistance" "ModeratelyActiveDistance"
## [9] "LightActiveDistance" "SedentaryActiveDistance"
## [11] "VeryActiveMinutes" "FairlyActiveMinutes"
## [13] "LightlyActiveMinutes" "SedentaryMinutes"
## [15] "Calories"
```

Showing daily calories data

```
dCalories <- read.csv('~/.Documents/Data Analyst/Bellabeat/Q2-Fitbit-Data/DailyCaloriesMerged_2020_05_12
head(dCalories)
```

```
##           Id ActivityDay Calories
## 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
```

```
str(dCalories)
```

```
## 'data.frame': 940 obs. of 3 variables:
## $ Id : num 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
## $ ActivityDay: chr "4/12/2016" "4/13/2016" "4/14/2016" "4/15/2016" ...
## $ Calories : int 1985 1797 1776 1745 1863 1728 1921 2035 1786 1775 ...
```

```
colnames(dCalories)
```

```
## [1] "Id" "ActivityDay" "Calories"
```

Showing daily intensities data

```
dIntensities <- read.csv('~/.Documents/Data Analyst/Bellabeat/Q2-Fitbit-Data/DailyIntensitiesMerged_2020')
head(dIntensities)
```

```
##      Id ActivityDay SedentaryMinutes LightlyActiveMinutes
## 1 1503960366 4/12/2016           728                328
## 2 1503960366 4/13/2016           776                217
## 3 1503960366 4/14/2016          1218                181
## 4 1503960366 4/15/2016           726                209
## 5 1503960366 4/16/2016           773                221
## 6 1503960366 4/17/2016           539                164
##      FairlyActiveMinutes VeryActiveMinutes SedentaryActiveDistance
## 1              13              25                0
## 2              19              21                0
## 3              11              30                0
## 4              34              29                0
## 5              10              36                0
## 6              20              38                0
##      LightActiveDistance ModeratelyActiveDistance VeryActiveDistance
## 1              6.06              0.55              1.88
## 2              4.71              0.69              1.57
## 3              3.91              0.40              2.44
## 4              2.83              1.26              2.14
## 5              5.04              0.41              2.71
## 6              2.51              0.78              3.19
```

```
str(dIntensities)
```

```
## 'data.frame': 940 obs. of 10 variables:
## $ Id : num 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
## $ ActivityDay : chr "4/12/2016" "4/13/2016" "4/14/2016" "4/15/2016" ...
## $ SedentaryMinutes : int 728 776 1218 726 773 539 1149 775 818 838 ...
## $ LightlyActiveMinutes : int 328 217 181 209 221 164 233 264 205 211 ...
## $ FairlyActiveMinutes : int 13 19 11 34 10 20 16 31 12 8 ...
## $ VeryActiveMinutes : int 25 21 30 29 36 38 42 50 28 19 ...
## $ SedentaryActiveDistance : num 0 0 0 0 0 0 0 0 0 0 ...
## $ LightActiveDistance : num 6.06 4.71 3.91 2.83 5.04 ...
## $ ModeratelyActiveDistance: num 0.55 0.69 0.4 1.26 0.41 ...
## $ VeryActiveDistance : num 1.88 1.57 2.44 2.14 2.71 ...
```

```
colnames(dIntensities)
```

```
## [1] "Id" "ActivityDay"
## [3] "SedentaryMinutes" "LightlyActiveMinutes"
## [5] "FairlyActiveMinutes" "VeryActiveMinutes"
## [7] "SedentaryActiveDistance" "LightActiveDistance"
## [9] "ModeratelyActiveDistance" "VeryActiveDistance"
```

Showing daily steps data

```
dSteps <- read.csv('~/.Documents/Data Analyst/Bellabeat/Q2-Fitbit-Data/DailyStepsMerged_2020_05_12.csv')
head(dSteps)
```

```
##           Id ActivityDay StepTotal
## 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
```

```
str(dSteps)
```

```
## 'data.frame':    940 obs. of  3 variables:
## $ Id           : num  1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
## $ ActivityDay: chr   "4/12/2016" "4/13/2016" "4/14/2016" "4/15/2016" ...
## $ StepTotal   : int   13162 10735 10460 9762 12669 9705 13019 15506 10544 9819 ...
```

```
colnames(dSteps)
```

```
## [1] "Id" "ActivityDay" "StepTotal"
```

Showing sleep day data

```
sDay <- read.csv('~/.Documents/Data Analyst/Bellabeat/Q2-Fitbit-Data/SleepDayMerged_2020_05_12.csv')
head(sDay)
```

```
##           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
```

```
str(sDay)
```

```
## '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"
## $ TotalSleepRecords : int  1 2 1 2 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 ...
```

```
colnames(sDay)
```

```
## [1] "Id"          "SleepDay"      "TotalSleepRecords"
## [4] "TotalMinutesAsleep" "TotalTimeInBed"
```

Empty Values

After to showing daily users activity, got to search duplicate data for disposal

Searching duplicate data

```
sum(duplicated(dActivity))
```

```
## [1] 0
```

```
sum(duplicated(dCalories))
```

```
## [1] 0
```

```
sum(duplicated(dIntensities))
```

```
## [1] 0
```

```
sum(duplicated(dSteps))
```

```
## [1] 0
```

```
sum(duplicated(sDay))
```

```
## [1] 3
```

Cleaning duplicate data

```
sd <- sDay
nrow(dActivity)-sum(complete.cases(dActivity))
```

```
## [1] 0
```

```
nrow(dCalories)-sum(complete.cases(dCalories))
```

```
## [1] 0
```

```
nrow(dIntensities)-sum(complete.cases(dIntensities))
```

```
## [1] 0
```

```
nrow(dSteps)-sum(complete.cases(dSteps))
```

```
## [1] 0
```

```
nrow(sd)-sum(complete.cases(sd))
```

```
## [1] 0
```

Format and timetable

Subsequently, activity dates, days and sleep day format will be changed in each of these databases.

Showing activity date modified at daily activity data

```
dActivity$ActivityDate <- as.Date (dActivity$ActivityDate, format = "%m/%d/%Y")
head(dActivity)
```

```
##           Id ActivityDate TotalSteps TotalDistance TrackerDistance
## 1 1503960366 2016-04-12      13162           8.50           8.50
## 2 1503960366 2016-04-13      10735           6.97           6.97
## 3 1503960366 2016-04-14      10460           6.74           6.74
## 4 1503960366 2016-04-15       9762           6.28           6.28
## 5 1503960366 2016-04-16      12669           8.16           8.16
## 6 1503960366 2016-04-17       9705           6.48           6.48
##   LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 1                        0                1.88                   0.55
## 2                        0                1.57                   0.69
## 3                        0                2.44                   0.40
## 4                        0                2.14                   1.26
## 5                        0                2.71                   0.41
## 6                        0                3.19                   0.78
##   LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
## 1                  6.06                      0                 25
## 2                  4.71                      0                 21
## 3                  3.91                      0                 30
## 4                  2.83                      0                 29
## 5                  5.04                      0                 36
## 6                  2.51                      0                 38
##   FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories
## 1                   13                   328                728    1985
## 2                   19                   217                776    1797
## 3                   11                   181               1218    1776
## 4                   34                   209                726    1745
## 5                   10                   221                773    1863
## 6                   20                   164                539    1728
```

```
str(dActivity)
```

```
## 'data.frame':   940 obs. of  15 variables:
## $ Id           : num  1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
## $ ActivityDate : Date, format: "2016-04-12" "2016-04-13" ...
## $ TotalSteps   : int   13162 10735 10460 9762 12669 9705 13019 15506 10544 9819 ...
## $ TotalDistance : num   8.5 6.97 6.74 6.28 8.16 ...
## $ TrackerDistance : num   8.5 6.97 6.74 6.28 8.16 ...
## $ LoggedActivitiesDistance: num  0 0 0 0 0 0 0 0 0 0 ...
## $ VeryActiveDistance : num  1.88 1.57 2.44 2.14 2.71 ...
## $ ModeratelyActiveDistance: num  0.55 0.69 0.4 1.26 0.41 ...
## $ LightActiveDistance : num  6.06 4.71 3.91 2.83 5.04 ...
## $ SedentaryActiveDistance : num  0 0 0 0 0 0 0 0 0 0 ...
## $ VeryActiveMinutes : int   25 21 30 29 36 38 42 50 28 19 ...
## $ FairlyActiveMinutes : int   13 19 11 34 10 20 16 31 12 8 ...
## $ LightlyActiveMinutes : int   328 217 181 209 221 164 233 264 205 211 ...
## $ SedentaryMinutes : int   728 776 1218 726 773 539 1149 775 818 838 ...
```

```
## $ Calories : int 1985 1797 1776 1745 1863 1728 1921 2035 1786 1775 ...
```

Showing activity date modified at daily calories data

```
dCalories$ActivityDay <- as.Date (dCalories$ActivityDay, format = "%m/%d/%Y")
head(dCalories)
```

```
##           Id ActivityDay Calories
## 1 1503960366 2016-04-12    1985
## 2 1503960366 2016-04-13    1797
## 3 1503960366 2016-04-14    1776
## 4 1503960366 2016-04-15    1745
## 5 1503960366 2016-04-16    1863
## 6 1503960366 2016-04-17    1728
```

```
str(dCalories)
```

```
## 'data.frame': 940 obs. of 3 variables:
## $ Id : num 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
## $ ActivityDay: Date, format: "2016-04-12" "2016-04-13" ...
## $ Calories : int 1985 1797 1776 1745 1863 1728 1921 2035 1786 1775 ...
```

Showing activity date modified at daily intensities data

```
dIntensities$ActivityDay <- as.Date (dIntensities$ActivityDay, format = "%m/%d/%Y")
head(dIntensities)
```

```
##           Id ActivityDay SedentaryMinutes LightlyActiveMinutes
## 1 1503960366 2016-04-12           728           328
## 2 1503960366 2016-04-13           776           217
## 3 1503960366 2016-04-14          1218           181
## 4 1503960366 2016-04-15           726           209
## 5 1503960366 2016-04-16           773           221
## 6 1503960366 2016-04-17           539           164
##   FairlyActiveMinutes VeryActiveMinutes SedentaryActiveDistance
## 1              13              25              0
## 2              19              21              0
## 3              11              30              0
## 4              34              29              0
## 5              10              36              0
## 6              20              38              0
##   LightActiveDistance ModeratelyActiveDistance VeryActiveDistance
## 1              6.06              0.55              1.88
## 2              4.71              0.69              1.57
## 3              3.91              0.40              2.44
## 4              2.83              1.26              2.14
## 5              5.04              0.41              2.71
## 6              2.51              0.78              3.19
```

```
str(dIntensities)
```

```
## 'data.frame': 940 obs. of 10 variables:
## $ Id : num 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
## $ ActivityDay : Date, format: "2016-04-12" "2016-04-13" ...
## $ SedentaryMinutes : int 728 776 1218 726 773 539 1149 775 818 838 ...
## $ LightlyActiveMinutes : int 328 217 181 209 221 164 233 264 205 211 ...
```



```
## $ FairlyActiveMinutes : int 13 19 11 34 10 20 16 31 12 8 ...
## $ VeryActiveMinutes   : int 25 21 30 29 36 38 42 50 28 19 ...
## $ SedentaryActiveDistance : num 0 0 0 0 0 0 0 0 0 0 ...
## $ LightActiveDistance  : num 6.06 4.71 3.91 2.83 5.04 ...
## $ ModeratelyActiveDistance: num 0.55 0.69 0.4 1.26 0.41 ...
## $ VeryActiveDistance   : num 1.88 1.57 2.44 2.14 2.71 ...
```

Showing activity date modified at daily steps data

```
dSteps$ActivityDay <- as.Date (dSteps$ActivityDay, format = "%m/%d/%Y")
head(dSteps)
```

```
##           Id ActivityDay StepTotal
## 1 1503960366 2016-04-12    13162
## 2 1503960366 2016-04-13    10735
## 3 1503960366 2016-04-14    10460
## 4 1503960366 2016-04-15     9762
## 5 1503960366 2016-04-16    12669
## 6 1503960366 2016-04-17     9705
```

```
str(dSteps)
```

```
## 'data.frame':   940 obs. of  3 variables:
## $ Id           : num  1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
## $ ActivityDay: Date, format: "2016-04-12" "2016-04-13" ...
## $ StepTotal  : int  13162 10735 10460 9762 12669 9705 13019 15506 10544 9819 ...
```

Showing activity date modified at sleep day data

```
sDay$SleepDay <- sDay$SleepDay %>%
  as.POSIXct(format = "%m/%d/%Y %I:%M:%S %p") %>%
  as.Date(format = "%m/%d/%Y %I:%M:%S %p")
head(sDay)
```

```
##           Id SleepDay TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
## 1 1503960366 2016-04-12                1                327            346
## 2 1503960366 2016-04-13                2                384            407
## 3 1503960366 2016-04-15                1                412            442
## 4 1503960366 2016-04-16                2                340            367
## 5 1503960366 2016-04-17                1                700            712
## 6 1503960366 2016-04-19                1                304            320
```

```
str(sDay)
```

```
## '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      : Date, format: "2016-04-12" "2016-04-13" ...
## $ TotalSleepRecords : int  1 2 1 2 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 ...
```

Analysis

Once the data has been cleaned, the investigation continues to find the data that will help to create visualizations for decision-making.

First, this begins with a participants count to see if they are participating in each report, otherwise, any reports that don't have all parties will be ruled out because they hampered the analysis results

```
n_distinct(dActivity$Id)
```

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

```
n_distinct(dCalories$Id)
```

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

```
n_distinct(dIntensities$Id)
```

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

```
n_distinct(dSteps$Id)
```

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

```
n_distinct(sDay$Id)
```

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

In this case, the document *SleepDayMerged_2020_05_12* is incomplete and will not be used in the analysis.

Data frames

Then, I started to get the data frames of each daily user activity.

Showing daily activity data frame

```
dActivity %>%
```

```
select(TotalSteps, TotalDistance, VeryActiveDistance, ModeratelyActiveDistance, LightActiveDistance, Se
```

```
##      TotalSteps      TotalDistance      VeryActiveDistance ModeratelyActiveDistance
##  Min.       :    0      Min.       : 0.000      Min.       : 0.000      Min.       :0.0000
## 1st Qu.: 3790      1st Qu.: 2.620      1st Qu.: 0.000      1st Qu.:0.0000
## Median : 7406      Median : 5.245      Median : 0.210      Median :0.2400
## Mean   : 7638      Mean   : 5.490      Mean   : 1.503      Mean   :0.5675
## 3rd Qu.:10727      3rd Qu.: 7.713      3rd Qu.: 2.053      3rd Qu.:0.8000
## Max.    :36019      Max.    :28.030      Max.    :21.920      Max.    :6.4800
## LightActiveDistance SedentaryActiveDistance SedentaryMinutes VeryActiveMinutes
##  Min.       : 0.000      Min.       :0.000000      Min.       : 0.0      Min.       : 0.00
## 1st Qu.: 1.945      1st Qu.:0.000000      1st Qu.: 729.8      1st Qu.: 0.00
## Median : 3.365      Median :0.000000      Median :1057.5      Median : 4.00
## Mean   : 3.341      Mean   :0.001606      Mean   : 991.2      Mean   : 21.16
## 3rd Qu.: 4.782      3rd Qu.:0.000000      3rd Qu.:1229.5      3rd Qu.: 32.00
## Max.    :10.710      Max.    :0.110000      Max.    :1440.0      Max.    :210.00
## FairlyActiveMinutes LightlyActiveMinutes
##  Min.       : 0.00      Min.       : 0.0
## 1st Qu.: 0.00      1st Qu.:127.0
## Median : 6.00      Median :199.0
## Mean   : 13.56      Mean   :192.8
## 3rd Qu.: 19.00      3rd Qu.:264.0
## Max.    :143.00      Max.    :518.0
```

Showing daily calories data frame

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

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

Showing daily intensities data frame

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

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

Showing daily steps data frame

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

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

Construction data

As soon as the results of the data frame, are made a comparison with daily calories and daily steps about daily activity

```
datacs <- merge(dCalories,dSteps)  
head(datacs)
```

```
##      Id ActivityDay Calories StepTotal  
## 1 1503960366 2016-04-12      1985      13162  
## 2 1503960366 2016-04-13      1797      10735  
## 3 1503960366 2016-04-14      1776      10460  
## 4 1503960366 2016-04-15      1745       9762
```

```
## 5 1503960366 2016-04-16 1863 12669
## 6 1503960366 2016-04-17 1728 9705
```

```
str(dataacs)
```

```
## 'data.frame': 940 obs. of 4 variables:
## $ Id : num 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
## $ ActivityDay: Date, format: "2016-04-12" "2016-04-13" ...
## $ Calories : int 1985 1797 1776 1745 1863 1728 1921 2035 1786 1775 ...
## $ StepTotal : int 13162 10735 10460 9762 12669 9705 13019 15506 10544 9819 ...
```

After that, is chosen the calories and step total data to understand average than the average number of calories burned through the footsteps of the users

```
dataacs %>%
  select(Calories, StepTotal) %>%
  summary()
```

```
##      Calories      StepTotal
## Min.   : 0      Min.   : 0
## 1st Qu.:1828    1st Qu.: 3790
## Median :2134    Median : 7406
## Mean   :2304    Mean   : 7638
## 3rd Qu.:2793    3rd Qu.:10727
## Max.   :4900    Max.   :36019
```

The calorie data gained from such users is instrumental in investigating and identifying the days of the week on which people have an intensive until sedentary lifestyle

```
ad <- merge(dActivity, dSteps)
head(dataacs)
```

```
##      Id ActivityDay Calories StepTotal
## 1 1503960366 2016-04-12 1985 13162
## 2 1503960366 2016-04-13 1797 10735
## 3 1503960366 2016-04-14 1776 10460
## 4 1503960366 2016-04-15 1745 9762
## 5 1503960366 2016-04-16 1863 12669
## 6 1503960366 2016-04-17 1728 9705
```

```
str(dataacs)
```

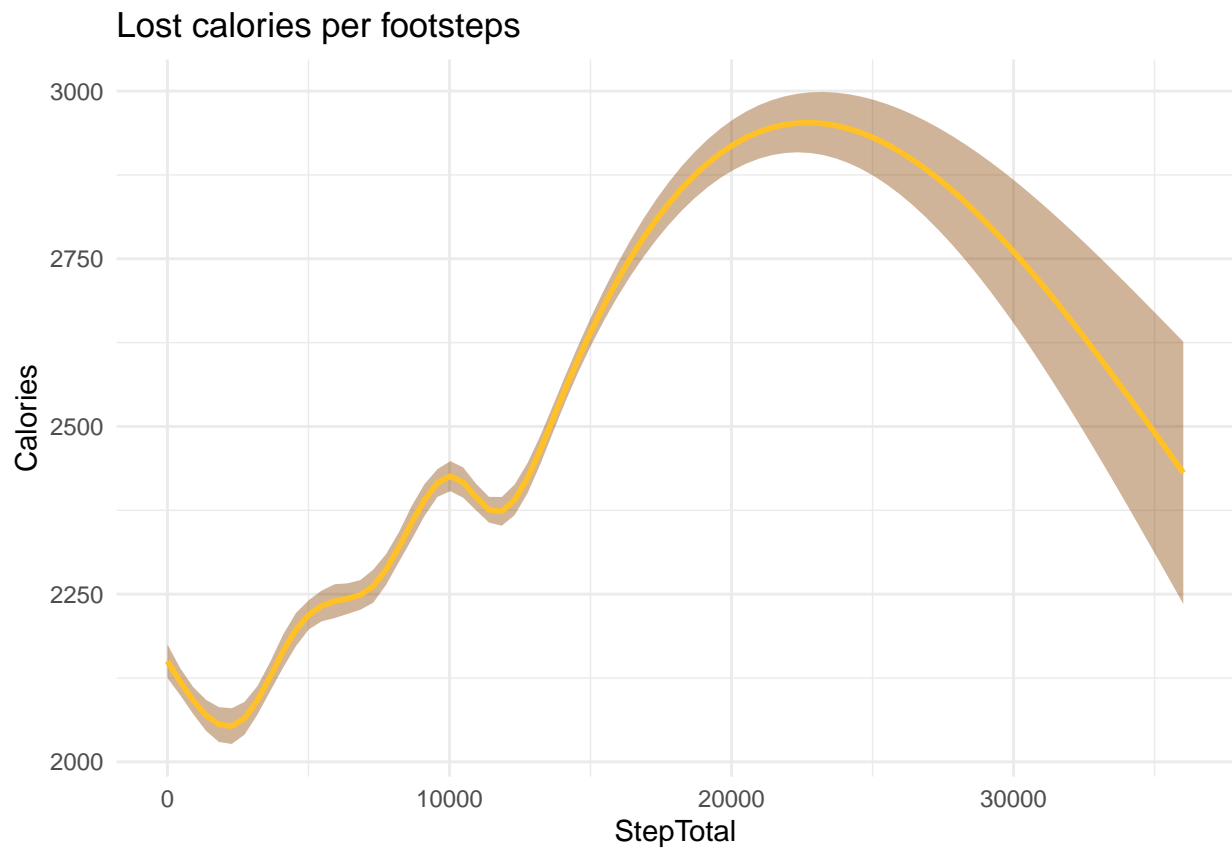
```
## 'data.frame': 940 obs. of 4 variables:
## $ Id : num 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
## $ ActivityDay: Date, format: "2016-04-12" "2016-04-13" ...
## $ Calories : int 1985 1797 1776 1745 1863 1728 1921 2035 1786 1775 ...
## $ StepTotal : int 13162 10735 10460 9762 12669 9705 13019 15506 10544 9819 ...
```

Showing charts

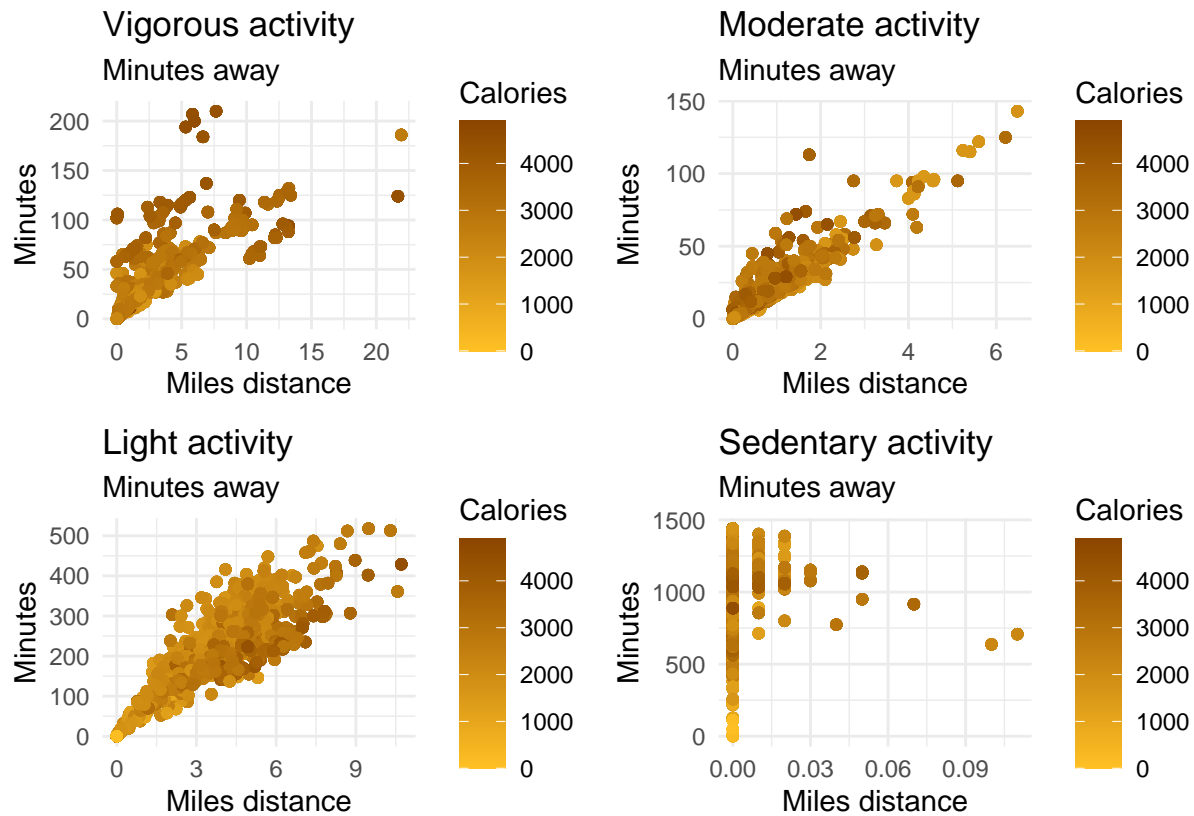
Previous results, will be created charts with the purpose to shows the performance of caloric losses under the users' walking rhythm and the period time which passes due type of activity that performs in their day to day lives

```
ggplot(data=ad) + geom_smooth(mapping=aes(x=StepTotal,y=Calories),color = "goldenrod1", fill="darkorange")
```

```
## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```



```
c1 <- ggplot(data=ad, aes(x=VeryActiveDistance, y=VeryActiveMinutes)) + geom_point(aes(color=Calories),
c2 <- ggplot(data=ad, aes(x=ModeratelyActiveDistance, y=FairlyActiveMinutes)) + geom_point(aes(color=Ca.
c3 <- ggplot(data=ad, aes(x=LightActiveDistance, y=LightlyActiveMinutes)) + geom_point(aes(color=Calori
c4 <- ggplot(data=ad, aes(x=SedentaryActiveDistance, y=SedentaryMinutes)) + geom_point(aes(color=Calori
c1 + c2 + c3 + c4 & theme_minimal()
```



Average

The average lets us know the user's total steps per day.

```
dSteps %>%
  group_by(Id) %>%
  summarise(StepTotalAverage = mean(StepTotal))
```

```
## # A tibble: 33 x 2
##       Id StepTotalAverage
##   <dbl>         <dbl>
## 1 1503960366      12117.
## 2 1624580081       5744.
## 3 1644430081       7283.
## 4 1844505072       2580.
## 5 1927972279        916.
## 6 2022484408      11371.
## 7 2026352035       5567.
## 8 2320127002       4717.
## 9 2347167796       9520.
## 10 2873212765       7556.
## # i 23 more rows
```

Conclusion

In conclusion, for this analysis, it can be seen that despite having limited information shared by the Bellabeat team, the trends in the data align. This is because it demonstrates that depending on the effort exerted and the amount of rest each user gets, their calorie fluctuates in their daily lives.

Now, the recommendations that would be given to the Bellabeat team to improve the marketing of their products are as follows: In addition to tracking the physical activity of their users, they also need to gather data on their age, weight, and height. This will enable them to devise a strategy for rewards that motivates users to continue using their products.

References

I would like to thank the following websites that helped me with create my case:

- Kaggle [Click here](#)
- Coursera [Click here](#)
- Stack Overflow [Click here](#)