Bellabeat Case Study-R

***About the company*:** Bellabeat, a high-tech company that manufactures health-focused smart products. Urška Sršen and Sando Mur founded Bellabeat. 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 established an international market with multiple products.

***Business Task*:** Analyze smart device usage data in order to gain insight into how consumers use non-Bellabeat smart devices.

***Questions*:**

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

***Data source*:** The stakeholder suggested to use the FitBit Fitness Tracker Data (CC0: Public Domain, datasets made available through Mobius)

```
# Installing Packages

if (!requireNamespace("tidyverse", quietly = TRUE)) {
  install.packages("tidyverse")
}
install.packages("janitor")
```

Preparation

```
## Installing package into 'C:/Users/91988/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
```

```
## package 'janitor' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\91988\AppData\Local\Temp\RtmpikCIKu\downloaded_packages
```

```
install.packages("readr")
```

```
## Installing package into 'C:/Users/91988/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)

## package 'readr' successfully unpacked and MD5 sums checked

## Warning: cannot remove prior installation of package 'readr'

## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying
## C:\Users\91988\AppData\Local\R\win-library\4.3\00LOCK\readr\libs\x64\readr.dll
## to C:\Users\91988\AppData\Local\R\win-library\4.3\readr\libs\x64\readr.dll:
## Permission denied

## Warning: restored 'readr'

##
## The downloaded binary packages are in
##  C:\Users\91988\AppData\Local\Temp\RtmpikCIKu\downloaded_packages
```

```
install.packages("lubridate")
```

```
## Installing package into 'C:/Users/91988/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)

## package 'lubridate' successfully unpacked and MD5 sums checked

## Warning: cannot remove prior installation of package 'lubridate'

## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying
## C:\Users\91988\AppData\Local\R\win-library\4.3\00LOCK\lubridate\libs\x64\lubridate.dll
## to
## C:\Users\91988\AppData\Local\R\win-library\4.3\lubridate\libs\x64\lubridate.dll:
## Permission denied

## Warning: restored 'lubridate'

##
## The downloaded binary packages are in
##  C:\Users\91988\AppData\Local\Temp\RtmpikCIKu\downloaded_packages
```

```
# Loading 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   1.5.1
## v ggplot2   3.5.0     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- 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
```

```r
library("dplyr")
library("readr")
library("janitor")
```

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

```r
library("ggplot2")
library("lubridate")
```

```r
# Importing datasets and assigning new names

daily_activity <- read.csv("D:/PROJECTS/Bellabeat/mturkfitbit_export_4.12.16-5.12.16/Fitabase Data 4.12
daily_calories <- read.csv("D:/PROJECTS/Bellabeat/mturkfitbit_export_4.12.16-5.12.16/Fitabase Data 4.12
daily_intensities <- read.csv("D:/PROJECTS/Bellabeat/mturkfitbit_export_4.12.16-5.12.16/Fitabase Data 4
daily_steps <- read.csv("D:/PROJECTS/Bellabeat/mturkfitbit_export_4.12.16-5.12.16/Fitabase Data 4.12.16
daily_sleep <- read.csv("D:/PROJECTS/Bellabeat/mturkfitbit_export_4.12.16-5.12.16/Fitabase Data 4.12.16
weight_log_info <- read.csv("D:/PROJECTS/Bellabeat/mturkfitbit_export_4.12.16-5.12.16/Fitabase Data 4.1
```

```r
#A quick preview of datasets

head(daily_activity)
```

Examine the datasets

```
##           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
```

```
head(daily_calories)
```

```
##           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
```

```
head(daily_intensities)
```

```
##           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
```

```
head(daily_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
```

```r
head(daily_steps)
```

```
##           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
```

```r
head(weight_log_info)
```

```
##           Id                  Date WeightKg WeightPounds Fat   BMI
## 1 1503960366  5/2/2016 11:59:59 PM     52.6     115.9631  22 22.65
## 2 1503960366  5/3/2016 11:59:59 PM     52.6     115.9631  NA 22.65
## 3 1927972279  4/13/2016 1:08:52 AM    133.5     294.3171  NA 47.54
## 4 2873212765 4/21/2016 11:59:59 PM     56.7     125.0021  NA 21.45
## 5 2873212765 5/12/2016 11:59:59 PM     57.3     126.3249  NA 21.69
## 6 4319703577 4/17/2016 11:59:59 PM     72.4     159.6147  25 27.45
##   IsManualReport        LogId
## 1           True 1.462234e+12
## 2           True 1.462320e+12
## 3          False 1.460510e+12
## 4           True 1.461283e+12
## 5           True 1.463098e+12
## 6           True 1.460938e+12
```

```r
# View structure of the imported dataframes

str(daily_activity)
```

```
## '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 ...
```

```
str(daily_calories)
```

```
## '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 ...
```

```
str(daily_intensities)
```

```
## '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 ...
```

```
str(daily_sleep)
```

```
## '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 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 ...
```

```
str(daily_steps)
```

```
## '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 ...
```

```
str(weight_log_info)
```

```
## 'data.frame':    67 obs. of  8 variables:
##  $ Id            : num  1.50e+09 1.50e+09 1.93e+09 2.87e+09 2.87e+09 ...
##  $ Date          : chr  "5/2/2016 11:59:59 PM" "5/3/2016 11:59:59 PM" "4/13/2016 1:08:52 AM" "4/21/2
##  $ WeightKg      : num  52.6 52.6 133.5 56.7 57.3 ...
##  $ WeightPounds  : num  116 116 294 125 126 ...
##  $ Fat           : int  22 NA NA NA NA 25 NA NA NA NA ...
##  $ BMI           : num  22.6 22.6 47.5 21.5 21.7 ...
##  $ IsManualReport: chr  "True" "True" "False" "True" ...
##  $ LogId         : num  1.46e+12 1.46e+12 1.46e+12 1.46e+12 1.46e+12 ...
```

```
#Checking "unique user Ids" in the data frame

n_distinct(daily_activity$Id)
```

Process

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

```
n_distinct(daily_calories$Id)
```

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

```
n_distinct(daily_intensities$Id)
```

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

```
n_distinct(daily_steps$Id)
```

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

```
n_distinct(daily_sleep$Id)
```

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

```
n_distinct(weight_log_info$Id)
```

```
## [1] 8
```

This information further proves that the daily_activity dataset contains the data from daily_intensities, daily_calories and daily_steps.There are 33 users in all 4 of those datasets.

I'll be using daily_activity, daily_sleep and weight_log_info datasets here, since the daily_intensities, daily_calories and daily_steps were already included in the daily_activity dataset.

```
#Identifying duplicates

sum(duplicated(daily_activity))
```

Identifying duplicates, missing values and remove duplicates

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

```r
sum(duplicated(daily_sleep))
```

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

```r
sum(duplicated(weight_log_info))
```

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

The daily_sleep data frame contains 3 duplicate entries, will be eliminated those by applying the distinct() and drop_na() function to ensure that only unique and non null records are retained for analysis.

```r
# Remove duplicates and null values

daily_activity <- daily_activity %>%
  distinct() %>%
  drop_na()
daily_sleep <- daily_sleep %>%
  distinct() %>%
  drop_na()
weight_log_info <- weight_log_info %>%
  distinct() %>%
  drop_na()
```

```r
# Ensure duplicates were removed

sum(duplicated(daily_sleep))
```

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

```r
# Ensure NULL values were removed

sum(is.na(daily_activity))
```

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

```r
sum(is.na(daily_sleep))
```

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

```r
sum(is.na(weight_log_info))
```

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

```r
# Cleaning and standardizing column name

daily_activity <- daily_activity %>%
  clean_names() %>%
  rename_with(tolower)
```

```r
daily_sleep <- daily_sleep %>%
  clean_names() %>%
  rename_with(tolower)
weight_log_info <- weight_log_info %>%
  clean_names() %>%
  rename_with(tolower)
```

```r
# Ensure that column names to lowercase across the data frame

colnames(daily_activity)
```

```
##  [1] "id"                          "activity_date"
##  [3] "total_steps"                 "total_distance"
##  [5] "tracker_distance"            "logged_activities_distance"
##  [7] "very_active_distance"        "moderately_active_distance"
##  [9] "light_active_distance"       "sedentary_active_distance"
## [11] "very_active_minutes"         "fairly_active_minutes"
## [13] "lightly_active_minutes"      "sedentary_minutes"
## [15] "calories"
```

```r
colnames(daily_sleep)
```

```
## [1] "id"                   "sleep_day"            "total_sleep_records"
## [4] "total_minutes_asleep" "total_time_in_bed"
```

```r
colnames(weight_log_info)
```

```
## [1] "id"               "date"             "weight_kg"        "weight_pounds"
## [5] "fat"              "bmi"              "is_manual_report" "log_id"
```

```r
# Daily_sleep

daily_sleep <- daily_sleep %>%
  separate(sleep_day, c('date', 'time')," ")
```

Standardizing date format: format Date of some data frame to match with other data frames.

```
## Warning: Expected 2 pieces. Additional pieces discarded in 410 rows [1, 2, 3, 4, 5, 6,
## 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].
```

```r
# Weight_log

weight_log_info <- weight_log_info %>%
  separate(date, c('date', 'time'), " ")
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 2 rows [1, 2].
```

```r
# Ensure columns has been seperated

head(daily_sleep)
```

```
##           id      date     time total_sleep_records total_minutes_asleep
## 1 1503960366 4/12/2016 12:00:00                   1                  327
## 2 1503960366 4/13/2016 12:00:00                   2                  384
## 3 1503960366 4/15/2016 12:00:00                   1                  412
## 4 1503960366 4/16/2016 12:00:00                   2                  340
## 5 1503960366 4/17/2016 12:00:00                   1                  700
## 6 1503960366 4/19/2016 12:00:00                   1                  304
##   total_time_in_bed
## 1               346
## 2               407
## 3               442
## 4               367
## 5               712
## 6               320
```

```r
head(weight_log_info)
```

```
##           id      date     time weight_kg weight_pounds fat   bmi
## 1 1503960366  5/2/2016 11:59:59      52.6      115.9631  22 22.65
## 2 4319703577 4/17/2016 11:59:59      72.4      159.6147  25 27.45
##   is_manual_report       log_id
## 1             True 1.462234e+12
## 2             True 1.460938e+12
```

```r
# Rename the activity_date column name of daily_activity data frame

daily_activity <- daily_activity %>%
  rename(date = activity_date)

head(daily_activity)
```

```
##           id      date total_steps total_distance tracker_distance
## 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
##   logged_activities_distance very_active_distance moderately_active_distance
## 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
##   light_active_distance sedentary_active_distance very_active_minutes
## 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
##   fairly_active_minutes lightly_active_minutes sedentary_minutes 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
```

```
# Change data type

daily_activity$date <- mdy(daily_activity$date)
daily_sleep$date <- mdy(daily_sleep$date)
weight_log_info$date <- mdy(weight_log_info$date)
```

```
# Ensure data type has been changed

head(daily_activity)
```

```
##           id       date total_steps total_distance tracker_distance
## 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
##   logged_activities_distance very_active_distance moderately_active_distance
## 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
##   light_active_distance sedentary_active_distance very_active_minutes
## 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
##   fairly_active_minutes lightly_active_minutes sedentary_minutes 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
```

```r
head(daily_sleep)
```

```
##           id       date     time total_sleep_records total_minutes_asleep
## 1 1503960366 2016-04-12 12:00:00                    1                  327
## 2 1503960366 2016-04-13 12:00:00                    2                  384
## 3 1503960366 2016-04-15 12:00:00                    1                  412
## 4 1503960366 2016-04-16 12:00:00                    2                  340
## 5 1503960366 2016-04-17 12:00:00                    1                  700
## 6 1503960366 2016-04-19 12:00:00                    1                  304
##   total_time_in_bed
## 1               346
## 2               407
## 3               442
## 4               367
## 5               712
## 6               320
```

```r
head(weight_log_info)
```

```
##           id       date     time weight_kg weight_pounds fat   bmi
## 1 1503960366 2016-05-02 11:59:59      52.6      115.9631  22 22.65
## 2 4319703577 2016-04-17 11:59:59      72.4      159.6147  25 27.45
##   is_manual_report       log_id
## 1             True 1.462234e+12
## 2             True 1.460938e+12
```

```r
# View the structure of datasets

str(daily_activity)
```

```
## 'data.frame':    940 obs. of  15 variables:
##  $ id                       : num  1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ date                     : Date, format: "2016-04-12" "2016-04-13" ...
##  $ total_steps              : int  13162 10735 10460 9762 12669 9705 13019 15506 10544 9819 ...
##  $ total_distance           : num  8.5 6.97 6.74 6.28 8.16 ...
##  $ tracker_distance         : num  8.5 6.97 6.74 6.28 8.16 ...
##  $ logged_activities_distance: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ very_active_distance     : num  1.88 1.57 2.44 2.14 2.71 ...
##  $ moderately_active_distance: num  0.55 0.69 0.4 1.26 0.41 ...
##  $ light_active_distance    : num  6.06 4.71 3.91 2.83 5.04 ...
##  $ sedentary_active_distance : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ very_active_minutes      : int  25 21 30 29 36 38 42 50 28 19 ...
##  $ fairly_active_minutes    : int  13 19 11 34 10 20 16 31 12 8 ...
##  $ lightly_active_minutes   : int  328 217 181 209 221 164 233 264 205 211 ...
##  $ sedentary_minutes        : int  728 776 1218 726 773 539 1149 775 818 838 ...
##  $ calories                 : int  1985 1797 1776 1745 1863 1728 1921 2035 1786 1775 ...
```

```r
str(daily_sleep)
```

```
## 'data.frame':    410 obs. of  6 variables:
##  $ id                  : num  1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
```

```
##  $ date                : Date, format: "2016-04-12" "2016-04-13" ...
##  $ time                : chr  "12:00:00" "12:00:00" "12:00:00" "12:00:00" ...
##  $ total_sleep_records : int  1 2 1 2 1 1 1 1 1 1 ...
##  $ total_minutes_asleep: int  327 384 412 340 700 304 360 325 361 430 ...
##  $ total_time_in_bed   : int  346 407 442 367 712 320 377 364 384 449 ...
```

```
str(weight_log_info)
```

```
## 'data.frame':    2 obs. of  9 variables:
##  $ id              : num  1.50e+09 4.32e+09
##  $ date            : Date, format: "2016-05-02" "2016-04-17"
##  $ time            : chr  "11:59:59" "11:59:59"
##  $ weight_kg       : num  52.6 72.4
##  $ weight_pounds   : num  116 160
##  $ fat             : int  22 25
##  $ bmi             : num  22.6 27.5
##  $ is_manual_report: chr  "True" "True"
##  $ log_id          : num  1.46e+12 1.46e+12
```

```
# Summary of total_steps and calories

daily_activity %>%
  select(total_steps,calories) %>%
  summary()
```

Analysis

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

```
# Summary of total time in bed and total minutes asleep

daily_sleep %>%
  select(total_time_in_bed,total_minutes_asleep) %>%
  summary()
```

```
##  total_time_in_bed total_minutes_asleep
##  Min.   : 61.0     Min.   : 58.0
##  1st Qu.:403.8     1st Qu.:361.0
##  Median :463.0     Median :432.5
##  Mean   :458.5     Mean   :419.2
##  3rd Qu.:526.0     3rd Qu.:490.0
##  Max.   :961.0     Max.   :796.0
```

```r
# Average minutes of activity

activity <- daily_activity %>%
  summarise(very_active_minutes = round(mean(very_active_minutes )),
  fairly_active_minutes = round(mean(fairly_active_minutes)),
  lightly_active_minutes = round(mean(lightly_active_minutes)),
  sedentary_minutes = round(mean(sedentary_minutes)))
```

```r
activity_table <- pivot_longer(activity,
                               cols = c("very_active_minutes", "fairly_active_minutes", "lightly_active
                               names_to = "activity",
                               values_to = "minutes")
```

```r
# View the table

head(activity_table)
```

```
## # A tibble: 4 x 2
##   activity               minutes
##   <chr>                    <dbl>
## 1 very_active_minutes         21
## 2 fairly_active_minutes       14
## 3 lightly_active_minutes     193
## 4 sedentary_minutes          991
```

```r
# Summary of daily_sleep

summary(daily_sleep)
```

```
##        id                 date                time
##  Min.   :1.504e+09   Min.   :2016-04-12   Length:410
##  1st Qu.:3.977e+09   1st Qu.:2016-04-19   Class :character
##  Median :4.703e+09   Median :2016-04-27   Mode  :character
##  Mean   :4.995e+09   Mean   :2016-04-26
##  3rd Qu.:6.962e+09   3rd Qu.:2016-05-04
##  Max.   :8.792e+09   Max.   :2016-05-12
##  total_sleep_records total_minutes_asleep total_time_in_bed
##  Min.   :1.00        Min.   : 58.0        Min.   : 61.0
##  1st Qu.:1.00        1st Qu.:361.0        1st Qu.:403.8
##  Median :1.00        Median :432.5        Median :463.0
##  Mean   :1.12        Mean   :419.2        Mean   :458.5
##  3rd Qu.:1.00        3rd Qu.:490.0        3rd Qu.:526.0
##  Max.   :3.00        Max.   :796.0        Max.   :961.0
```

```r
# Summary of weight_log_info

summary(weight_log_info)
```

```
##        id                 date                time             weight_kg
##  Min.   :1.504e+09   Min.   :2016-04-17   Length:2           Min.   :52.60
##  1st Qu.:2.208e+09   1st Qu.:2016-04-20   Class :character   1st Qu.:57.55
```

```
##  Median :2.912e+09   Median :2016-04-24   Mode  :character   Median :62.50
##  Mean   :2.912e+09   Mean   :2016-04-24                      Mean   :62.50
##  3rd Qu.:3.616e+09   3rd Qu.:2016-04-28                      3rd Qu.:67.45
##  Max.   :4.320e+09   Max.   :2016-05-02                      Max.   :72.40
##  weight_pounds        fat              bmi        is_manual_report
##  Min.   :116.0   Min.   :22.00   Min.   :22.65   Length:2
##  1st Qu.:126.9   1st Qu.:22.75   1st Qu.:23.85   Class :character
##  Median :137.8   Median :23.50   Median :25.05   Mode  :character
##  Mean   :137.8   Mean   :23.50   Mean   :25.05
##  3rd Qu.:148.7   3rd Qu.:24.25   3rd Qu.:26.25
##  Max.   :159.6   Max.   :25.00   Max.   :27.45
##      log_id
##  Min.   :1.461e+12
##  1st Qu.:1.461e+12
##  Median :1.462e+12
##  Mean   :1.462e+12
##  3rd Qu.:1.462e+12
##  Max.   :1.462e+12
```

```r
# The number of steps taken correlate with the total calories burned

ggplot(data = daily_activity, aes(x = total_steps, y = calories)) +
  geom_point() +
  geom_smooth() +
  labs(title = "The Relationship between Total Steps and Calories Burned",
       x = "Total Steps",
       y = "Calories Burned")
```

Share

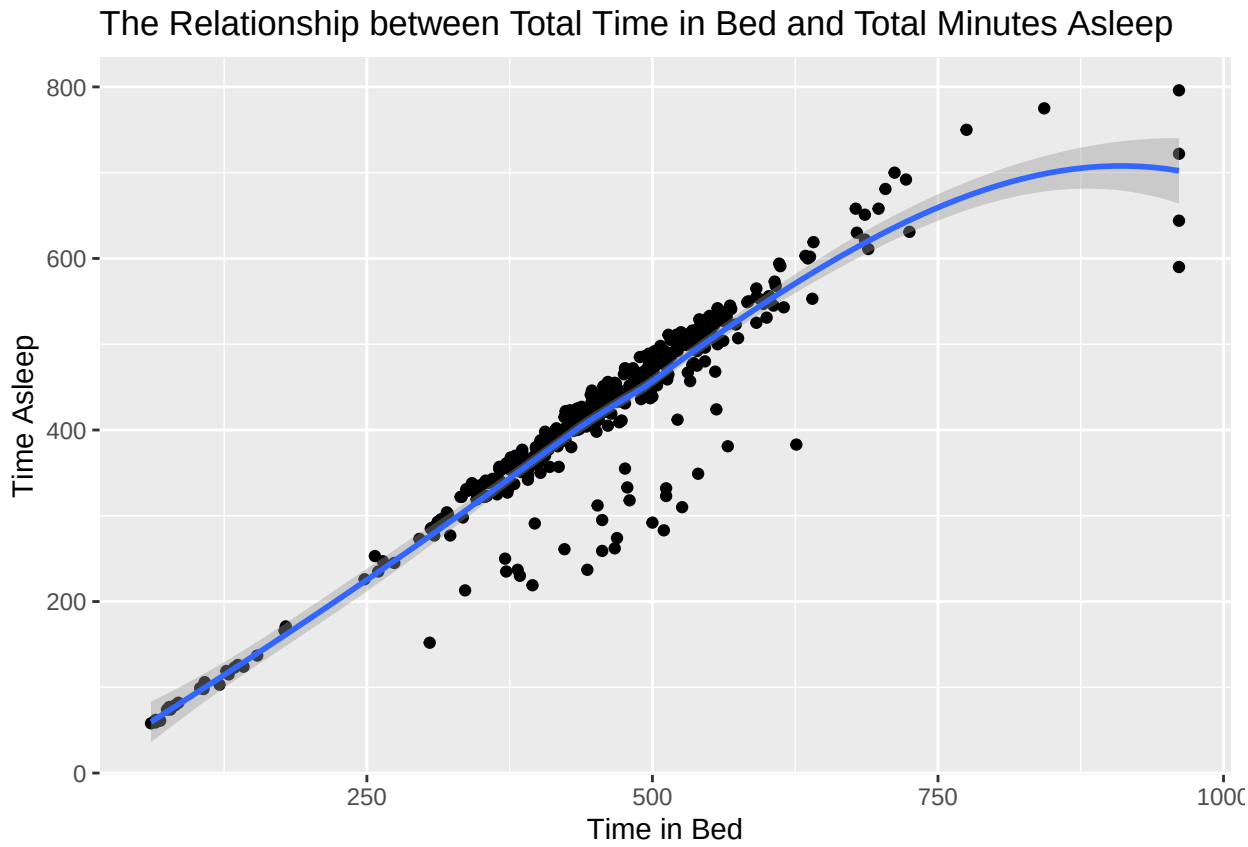
```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

The visualization of the relationship between Total Steps and Calories Burned leads to clear insight that there is a positive correlation between the two. This means that taking more steps corresponds with burning more calories.

```
# Minutes spent asleep correlate with the total time spent in bed

ggplot(data = daily_sleep, aes(x = total_time_in_bed, y = total_minutes_asleep) ) +
  geom_point() +
  geom_smooth() +
  labs(title = "The Relationship between Total Time in Bed and Total Minutes Asleep", x= "Time in Bed",
```

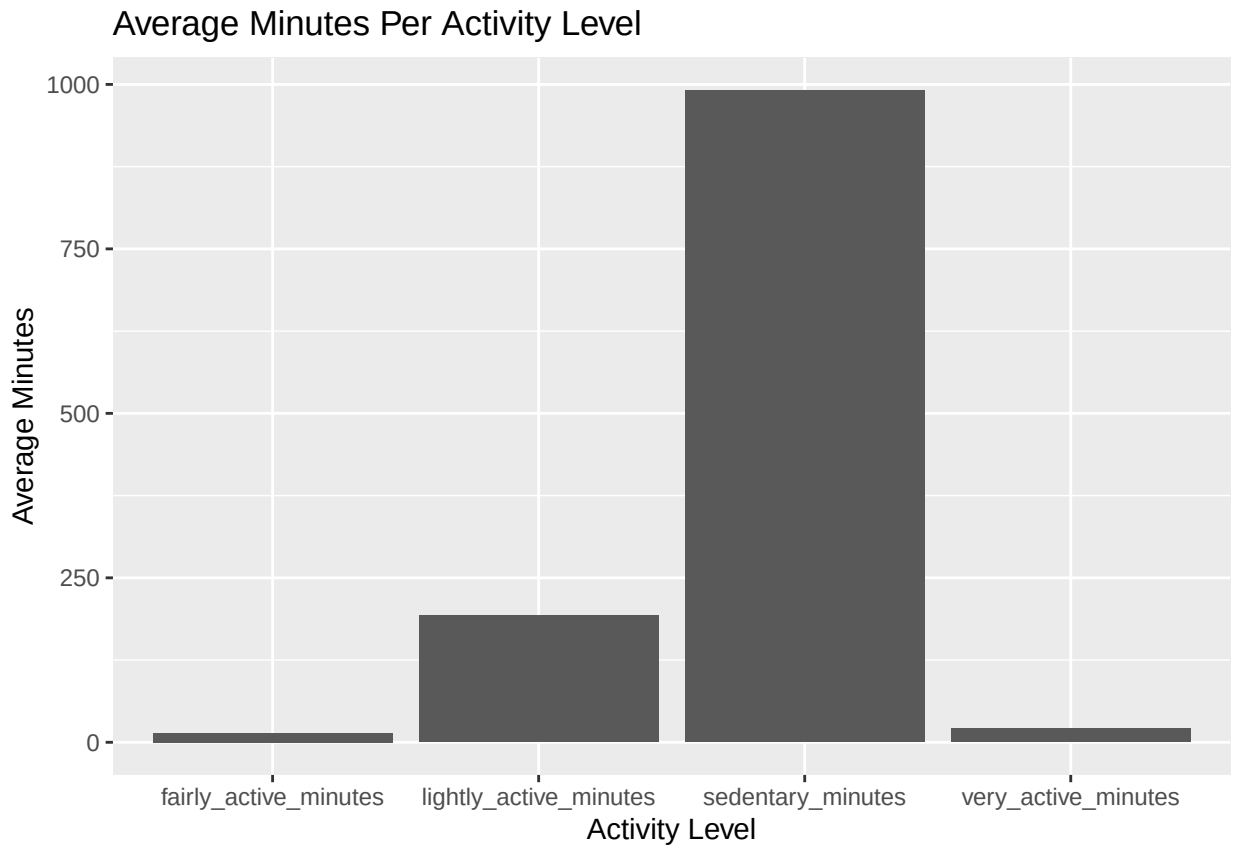
```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



The visualization of the relationship between Total Minutes ans Total Time Spend in the Bed leads to clear insight that there is a significant positve correlation between the number of minutes one spends asleep and amount of time spent in bed. This suggests that having more sleep is often correlated with spending more time in bed.

```
# Average minutes per activity level

ggplot(data = activity_table) +
  geom_col(mapping = aes(x = activity, y= minutes)) +
  labs(title = "Average Minutes Per Activity Level", x ="Activity Level", y="Average Minutes")
```



The visualization of Average Minutes Per Activity Level leads to clear insight that the majority of users are sedentary with only a small percentage being Fairly or Very Active.

Act To improve the results of this analysis I would suggest to collect data of our own smart devices and acting on those results. I have found some interesting insights and trends, concerning our business task, which can help to improve parts of our App and the App to user interaction.

***Trends in the Use of FitBit Smart Devices*:**

- Average Total Steps is 7638 per day.
- Average Calories Burn is 2304 per day.
- The average sedentary time is 991minutes (16.5 hours). means most of the users are sedentary.
- On average, the participants take around 7 hours/day sleep.
- Most participants spend the majority of their movement in the lightly active category.

***Company Recommendations*:**

1. Motivation to take more steps: I've discovered a strong correlation between the number of steps taken (7638 on average per day) and the number of calories burned (2304 on average per day). The average daily step count is 7638, which is somewhat below than the recommended amount for health benefits, according to CDC data. We are aware that 8,000 to 10,000 steps/day needed to maintain a healthy lifestyle. Therefore, it is recommended that the app sends them reminders including inspirational quotes to encourage users to complete the 8000 steps each day.

2. Increased attention to alarm-linked sleep cycles: I have found that the average person sleeps for roughly seven hours. Furthermore, I observed that not everyone tracks their sleep using the gadgets. In order to ensure that he gets at least eight hours of sleep every day. It is recommended that the user has the option to set a desirable time to wake up. And the application does the calculations and notifies the user when bedtime is almost approaching. It's additionally, so that he sleeps at least 8 hours daily

3. Notification of activity: Very high sedentary time (nearly 16 hours) a day of inactivity can cause a number of health problems. Is it because the majority of users engage in sedentary behaviour that they are not recording their actions or they are not moving at all. It is recommended that they improve the device's daily activity notifications, which will prompt users to log their actions. Provide a goal function to encourage them to move.

***Further Recommendations*:** Points and Rewards System: Since we knowing that notifications don't work for everyone, an in-app points and reward system could be created. Users could earn points and rewards for completing training, completing daily steps, getting eight hours of sleep, and engaging in active hours, among other accomplishments.