# Bellabeat_case_study

Edwin King'ori

2023-11-06

## Contents



Mission Statement

Bellabeat, a high-tech manufacturer of health-focused products for women, believes that analyzing smart device fitness data could help unlock new growth opportunities to become a larger player in the global [smart device] (https://en.wikipedia.org/wiki/Smart_device) market. The objective is 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 discovered will then help guide marketing strategy for the company.

The Analysis Process.

## ASK.

Analyzing smart device usage data in order to gain insight into how customers use non-bellabeat smart devices.

1.1 Stakeholders Involved

Urška Sršen: Bellabeat's cofounder and Chief Creative Officer.

Sando Mur: Mathematician and Bellabeat's cofounder; 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.

1.2 Insight Questions

What are some trends in smart device usage?

How could these trends apply to Bellabeat customers?

How could these trends help influence Bellabeat marketing strategy?

## Prepare.

Key Tasks Involved:

2.1 Identify how it's organized.

The FitBit Fitness Tracker data set contains personal fitness tracker from thirty fitbit users.

Thirty eligible Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring.

2.2 Determining the credibility of data.

Is the Data ROCC? Reliable, Original, Comprehensive, Current, and Cited.

Reliable: Low reliability due to the low number of respondents.

Original: The data was sourced from a third-party, hence, the source of the data cannot be easily verified.

Comprehensive :The data set has 18 csv files with information about daily activity, steps, and heart rate that can be used to explore users' habits. However, It lacks the usre's demographic info.

Time Frame : The data is (8) years old,thus, not relevant. However, we will forego the limitation for the purpose of th capstone study.

Cited: The original dataset contains the required sources information including, publication date, DOI (Document Object Identifier), Authors and file formats.

2.3 Loading Packages

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## 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(lubridate)
library(skimr)
library(dplyr)
library(janitor)
```

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

```
library(ggplot2)
library(selectr)
```

2.4 Importing files

```
activity <- read.csv("dailyActivity_merged.csv")
calories <- read.csv("dailyCalories_merged.csv")
intensities <- read.csv("dailyIntensities_merged.csv")
steps <- read.csv("dailySteps_merged.csv")
sleepday <- read.csv("sleepDay_merged.csv")
```

```r
Weight <- read.csv("weightLogInfo_merged.csv")

head(activity)
```

```
##           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
```

```r
head(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
```

```r
head(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(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
```

```
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
```

```
head(Weight)
```

```
##           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
```

2.5 Data Cleaning

2.5.1 Cleaning Column Names

```r
activity1 <- janitor::clean_names(activity)
calories1 <- janitor::clean_names(calories)
intensities1 <- janitor::clean_names(intensities)
steps1 <- janitor::clean_names(steps)
sleepday1 <- janitor::clean_names(sleepday)
Weight1 <- janitor::clean_names(Weight)

colnames(activity1)
```

```
##  [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(calories1)
```

```
## [1] "id"           "activity_day" "calories"
```

```r
colnames(intensities1)
```

```
##  [1] "id"                         "activity_day"
##  [3] "sedentary_minutes"          "lightly_active_minutes"
##  [5] "fairly_active_minutes"      "very_active_minutes"
##  [7] "sedentary_active_distance"  "light_active_distance"
##  [9] "moderately_active_distance" "very_active_distance"
```

```r
colnames(sleepday1)
```

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

```r
colnames(steps1)
```

```
## [1] "id"           "activity_day" "step_total"
```

2.6 Summarizing data

Checking sample size to decide whether to use for analysis

using the n-distinct() to check the number of unique ID in each data set.

```r
n_distinct(activity$Id)
```

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

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

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

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

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

```r
n_distinct(steps$Id)
```

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

```r
n_distinct(sleepday$Id)
```

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

```r
n_distinct(Weight$Id)
```

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

Using the summary () to get a comprehensive data categorization.

```r
activity1 %>%
  select(very_active_minutes,
         fairly_active_minutes,
         lightly_active_minutes) %>%
  summary()
```

```
##  very_active_minutes fairly_active_minutes lightly_active_minutes
##  Min.   :  0.00      Min.   :  0.00        Min.   :  0.0
##  1st Qu.:  0.00      1st Qu.:  0.00        1st Qu.:127.0
##  Median :  4.00      Median :  6.00        Median :199.0
##  Mean   : 21.16      Mean   : 13.56        Mean   :192.8
##  3rd Qu.: 32.00      3rd Qu.: 19.00        3rd Qu.:264.0
##  Max.   :210.00      Max.   :143.00        Max.   :518.0
```

```r
#calories summary
calories1 %>%
  select(calories) %>%
  summary()
```

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

```r
#Intensities
#summarizing the number of active minutes per category
intensities1 %>%
  select(very_active_minutes,
         fairly_active_minutes,
         lightly_active_minutes) %>%
  summary()
```

```
##  very_active_minutes fairly_active_minutes lightly_active_minutes
##  Min.   :  0.00      Min.   :  0.00        Min.   :  0.0
##  1st Qu.:  0.00      1st Qu.:  0.00        1st Qu.:127.0
##  Median :  4.00      Median :  6.00        Median :199.0
##  Mean   : 21.16      Mean   : 13.56        Mean   :192.8
##  3rd Qu.: 32.00      3rd Qu.: 19.00        3rd Qu.:264.0
##  Max.   :210.00      Max.   :143.00        Max.   :518.0
```

```r
#Sleep
sleepday1 %>%
  select(total_sleep_records,
         total_minutes_asleep,
         total_time_in_bed) %>%
```

```
  summary()
```

```
##  total_sleep_records total_minutes_asleep total_time_in_bed
##  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
```

2.7 Insights From the Summary

From the distinct data above, 33 participants reported their data for activity, Intensities and Steps. Although only 24 participants shared their data for sleep_day, I'll still use the data set for analysis.

Only 8 participants partook in the weight log analysis, which is the smallest data set. Thus, will not use it due to bias concerns

Most participants are lightly active compared to the highly active data.

The average steps is 7400 per day, which is lower than the recommended 8000 steps

The average sedentary time is 991, an equivalent of 16 hours which is lower than the medically required 4-8 hours a day according to this study

Most adults spend an approximated 7 hours of sleep.

# Process and Analyze

3.0 Formatting: Changing the date format, making the tables compatible for a successful merge.

```
# Formatting 'activity1' table
activity1$activity_date <- as.POSIXct(activity1$activity_date, format = "%m/%d/%Y", tz = Sys.timezone()
activity1$date <- format(activity1$activity_date, format = "%m/%d/%y")

# Formatting 'sleepday1' table
sleepday1$sleep_day <- as.POSIXct(sleepday1$sleep_day, format = "%m/%d/%Y %I:%M:%S %p", tz = Sys.timezo
sleepday1$date <- format(sleepday1$sleep_day, format = "%m/%d/%y")

head(activity1)
```

```
##           id activity_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
##       date
## 1 04/12/16
## 2 04/13/16
## 3 04/14/16
## 4 04/15/16
## 5 04/16/16
## 6 04/17/16
```

```
head(sleepday1)
```

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

3.1 Merging and Creating new Variables from merged data

```
merged_data <- merge ( sleepday1, activity1, by = c('id', 'date'))
head(merged_data)
```

```
##           id     date  sleep_day total_sleep_records total_minutes_asleep
## 1 1503960366 04/12/16 2016-04-12                   1                  327
## 2 1503960366 04/13/16 2016-04-13                   2                  384
## 3 1503960366 04/15/16 2016-04-15                   1                  412
## 4 1503960366 04/16/16 2016-04-16                   2                  340
## 5 1503960366 04/17/16 2016-04-17                   1                  700
## 6 1503960366 04/19/16 2016-04-19                   1                  304
##   total_time_in_bed activity_date total_steps total_distance tracker_distance
## 1               346    2016-04-12       13162           8.50             8.50
## 2               407    2016-04-13       10735           6.97             6.97
## 3               442    2016-04-15        9762           6.28             6.28
## 4               367    2016-04-16       12669           8.16             8.16
## 5               712    2016-04-17        9705           6.48             6.48
## 6               320    2016-04-19       15506           9.88             9.88
```

```
##   logged_activities_distance very_active_distance moderately_active_distance
## 1                          0                 1.88                       0.55
## 2                          0                 1.57                       0.69
## 3                          0                 2.14                       1.26
## 4                          0                 2.71                       0.41
## 5                          0                 3.19                       0.78
## 6                          0                 3.53                       1.32
##   light_active_distance sedentary_active_distance very_active_minutes
## 1                  6.06                         0                  25
## 2                  4.71                         0                  21
## 3                  2.83                         0                  29
## 4                  5.04                         0                  36
## 5                  2.51                         0                  38
## 6                  5.03                         0                  50
##   fairly_active_minutes lightly_active_minutes sedentary_minutes calories
## 1                    13                    328               728     1985
## 2                    19                    217               776     1797
## 3                    34                    209               726     1745
## 4                    10                    221               773     1863
## 5                    20                    164               539     1728
## 6                    31                    264               775     2035
```

```r
merged_data2 <- merge(x = steps1, y =  activity1, by = 'id')
head(merged_data2)
```

```
##           id activity_day step_total activity_date total_steps total_distance
## 1 1503960366    4/12/2016      13162    2016-04-12       13162           8.50
## 2 1503960366    4/12/2016      13162    2016-04-13       10735           6.97
## 3 1503960366    4/12/2016      13162    2016-04-14       10460           6.74
## 4 1503960366    4/12/2016      13162    2016-04-15        9762           6.28
## 5 1503960366    4/12/2016      13162    2016-04-16       12669           8.16
## 6 1503960366    4/12/2016      13162    2016-04-17        9705           6.48
##   tracker_distance logged_activities_distance very_active_distance
## 1             8.50                          0                 1.88
## 2             6.97                          0                 1.57
## 3             6.74                          0                 2.44
## 4             6.28                          0                 2.14
## 5             8.16                          0                 2.71
## 6             6.48                          0                 3.19
##   moderately_active_distance light_active_distance sedentary_active_distance
## 1                       0.55                  6.06                         0
## 2                       0.69                  4.71                         0
## 3                       0.40                  3.91                         0
## 4                       1.26                  2.83                         0
## 5                       0.41                  5.04                         0
## 6                       0.78                  2.51                         0
##   very_active_minutes fairly_active_minutes lightly_active_minutes
## 1                  25                    13                    328
## 2                  21                    19                    217
## 3                  30                    11                    181
## 4                  29                    34                    209
## 5                  36                    10                    221
## 6                  38                    20                    164
##   sedentary_minutes calories     date
## 1               728     1985 04/12/16
```

```
## 2               776     1797 04/13/16
## 3              1218     1776 04/14/16
## 4               726     1745 04/15/16
## 5               773     1863 04/16/16
## 6               539     1728 04/17/16
```

```
# Creating two variables: User_type and sleep_quality that will exhibit the distribution of the steps a
merged_df <- merged_data%>%
  mutate(user_type = case_when(
    .$total_steps < 5000 ~ "Sedentary",
    .$total_steps >= 5000 & .$total_steps < 7499 ~ "Low_active",
    .$total_steps >= 7500 & .$total_steps < 9999 ~ "Moderately_active",
    .$total_steps >= 10000 & .$total_steps < 12499 ~ "Active",
    .$total_steps >= 12500  ~ "very_active"
  ),sleep_quality = case_when(
    .$total_minutes_asleep/60 >= 6 & .$total_minutes_asleep/60 <= 8 ~"Sufficient Sleep",
    .$total_minutes_asleep/60 < 6  ~ "InSufficient Sleep",
    .$total_minutes_asleep/60 > 9  ~ "OverSleeping"

  )
  )
head(merged_df)
```

```
##           id     date  sleep_day total_sleep_records total_minutes_asleep
## 1 1503960366 04/12/16 2016-04-12                   1                  327
## 2 1503960366 04/13/16 2016-04-13                   2                  384
## 3 1503960366 04/15/16 2016-04-15                   1                  412
## 4 1503960366 04/16/16 2016-04-16                   2                  340
## 5 1503960366 04/17/16 2016-04-17                   1                  700
## 6 1503960366 04/19/16 2016-04-19                   1                  304
##   total_time_in_bed activity_date total_steps total_distance tracker_distance
## 1               346    2016-04-12       13162           8.50             8.50
## 2               407    2016-04-13       10735           6.97             6.97
## 3               442    2016-04-15        9762           6.28             6.28
## 4               367    2016-04-16       12669           8.16             8.16
## 5               712    2016-04-17        9705           6.48             6.48
## 6               320    2016-04-19       15506           9.88             9.88
##   logged_activities_distance very_active_distance moderately_active_distance
## 1                          0                 1.88                       0.55
## 2                          0                 1.57                       0.69
## 3                          0                 2.14                       1.26
## 4                          0                 2.71                       0.41
## 5                          0                 3.19                       0.78
## 6                          0                 3.53                       1.32
##   light_active_distance sedentary_active_distance very_active_minutes
## 1                  6.06                         0                  25
## 2                  4.71                         0                  21
## 3                  2.83                         0                  29
## 4                  5.04                         0                  36
## 5                  2.51                         0                  38
## 6                  5.03                         0                  50
##   fairly_active_minutes lightly_active_minutes sedentary_minutes calories
## 1                    13                    328               728     1985
## 2                    19                    217               776     1797
## 3                    34                    209               726     1745
```

```
## 4                    10                     221                773      1863
## 5                    20                     164                539      1728
## 6                    31                     264                775      2035
##          user_type      sleep_quality
## 1      very_active InSufficient Sleep
## 2           Active   Sufficient Sleep
## 3 Moderately_active   Sufficient Sleep
## 4      very_active InSufficient Sleep
## 5 Moderately_active       OverSleeping
## 6      very_active InSufficient Sleep
```

# Share

4.1 Visualizations

4.1.1: Examining the relationship between the Steps vs Calories.

```
ggplot(data = merged_data, aes(x = total_steps, y = calories, color = calories))+
geom_point()+
geom_smooth()+
labs(title = 'Relationship between the Steps vs Calories')+
theme(plot.title = element_text(hjust = 0.5),
      panel.border = element_rect(color = 'black',  fill = NA),
      legend.background = element_blank(),
      legend.box.background = element_rect(colour = "black"))
```

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

```
## Warning: The following aesthetics were dropped during statistical transformation: colour
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```


```
ggplot(data = merged_data, aes(x = very_active_minutes, y = calories, color = calories))+
  geom_point()+
  geom_smooth()+
  labs(title = 'Relationship between the Steps vs Calories')+
  theme(plot.title = element_text(hjust = 0.5),
        panel.border = element_rect(color = 'black',  fill = NA),
        legend.background = element_blank(),
        legend.box.background = element_rect(colour = "black"))
```

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

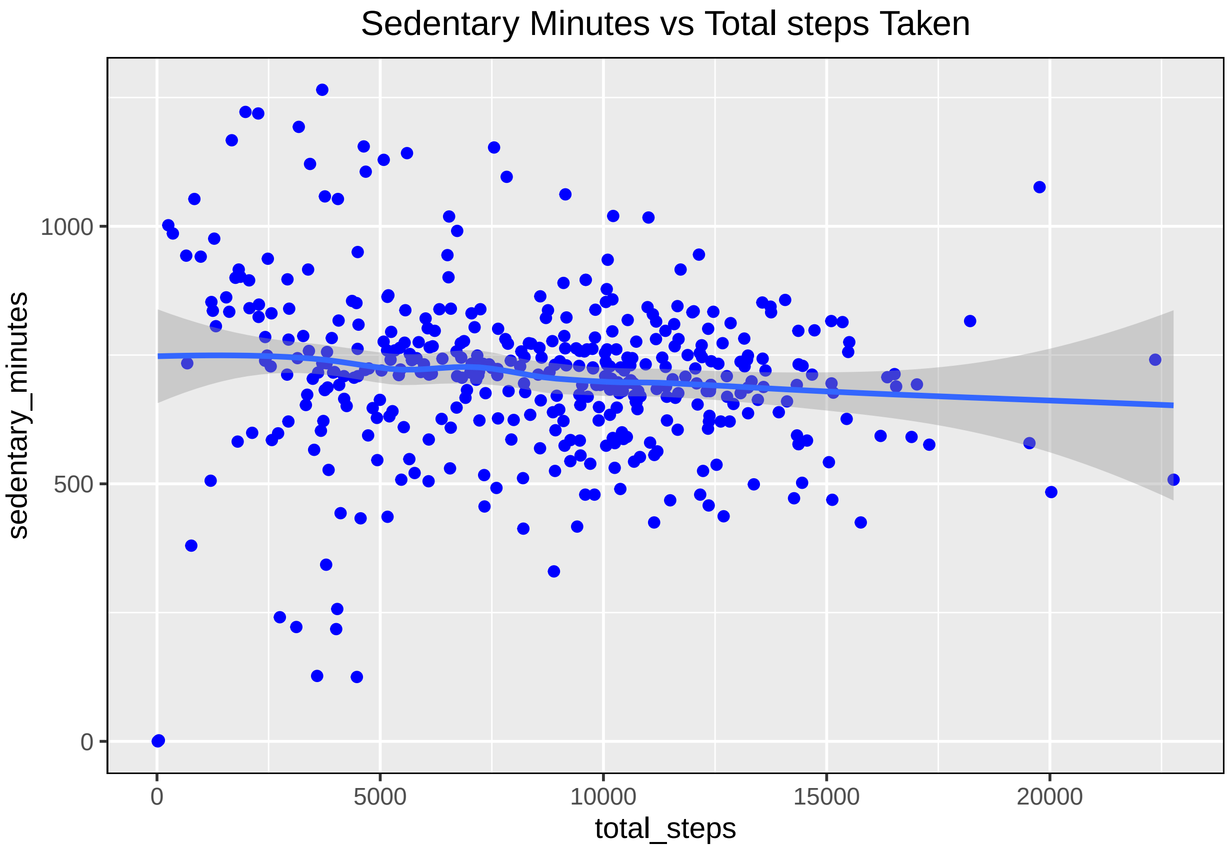
```
## Warning: The following aesthetics were dropped during statistical transformation: colour
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```


Based on the plots above, there is a positive correlation between The Total_steps, Very_active_minutes and Calories. This means that the more active participants are, the more calories they burn.

4.1.2 : Exploring the relationship between steps taken in a day and sedentary minutes

```
ggplot(data = merged_data, aes(x = total_steps, y = sedentary_minutes))+
  geom_point(color = "blue")+
  geom_smooth()+
  labs(title = "Sedentary Minutes vs Total steps Taken")+
  theme(plot.title = element_text(hjust = 0.5),
        panel.border = element_rect(color = 'black',  fill = NA))
```

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




The plot reveals a negative correlation. There is a downward trend in the data points and the smooth line, which suggests a negative relationship between sedentary minutes and the total steps taken. When there is an increase in the number of steps taken, there is an automatic reduction in the amount of time spent in sedentary state.

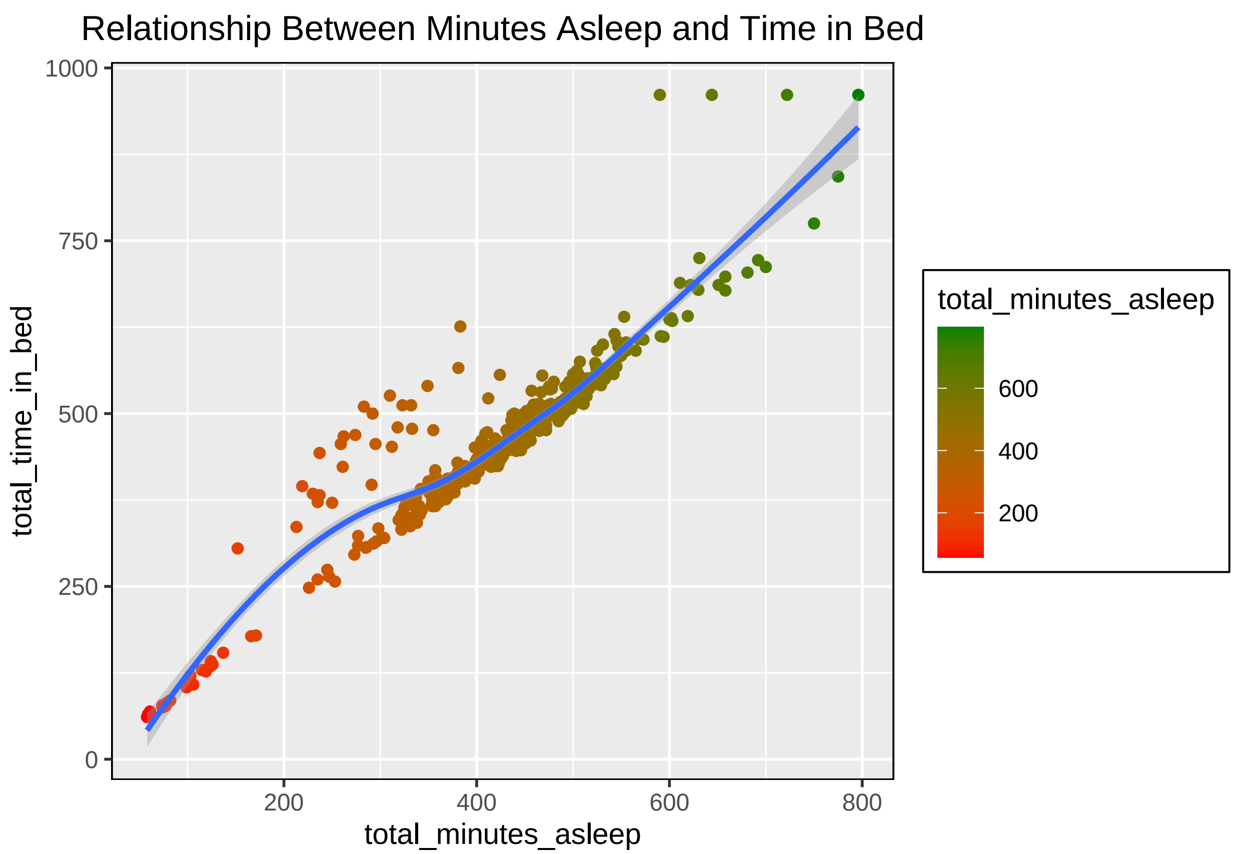
Therefore, the plot implies that participants with a high physical output in the steps taken have a lower amount of time spent in sedentary minutes. Thus, the plot shows the need for the company to sensitize its participants to engage in more physical activity like walking and reduce amount of time spent in sedentary.

4.1.3 : Exploring the Relationship Between Minutes Asleep and Time in Bed

```
ggplot(data = sleepday1, aes(x= total_minutes_asleep, y = total_time_in_bed, color = total_minutes_aslee
  geom_point()+
  geom_smooth()+
  labs(title = "Relationship Between Minutes Asleep and Time in Bed")+
  theme(plot.title = element_text(hjust = 0.5),
        panel.border = element_rect(color = 'black',  fill = NA),
        legend.background = element_blank(),
        legend.box.background = element_rect(colour = "black"))+
  scale_color_gradient(low="red", high ="#008000")
```

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

```
## Warning: The following aesthetics were dropped during statistical transformation: colour
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```




The plot indicates a positive correlation between minutes asleep and time in bed. There is a general upward trend in the data points and the smooth line and suggests a positive correlation between the duration of sleep (minutes asleep) and the total time spent in bed. This means that individuals who tend to sleep for longer durations also spend more time in bed.

Color Variation: The color gradient applied to the data points represents the "Total Minutes Asleep" variable. The color variation from red to green indicates the range of values for total minutes asleep, with red indicating lower values and green indicating higher values. This color gradient provides an additional dimension of information to the plot.

Concentration of Points: If you notice a concentration of data points around a specific range of minutes asleep and time in bed, it suggests that there may be a common pattern or trend within that range. For example, if you observe a cluster of green points with a high time in bed and high minutes asleep, it could indicate that individuals who consistently sleep for longer duration also spend more time in bed.

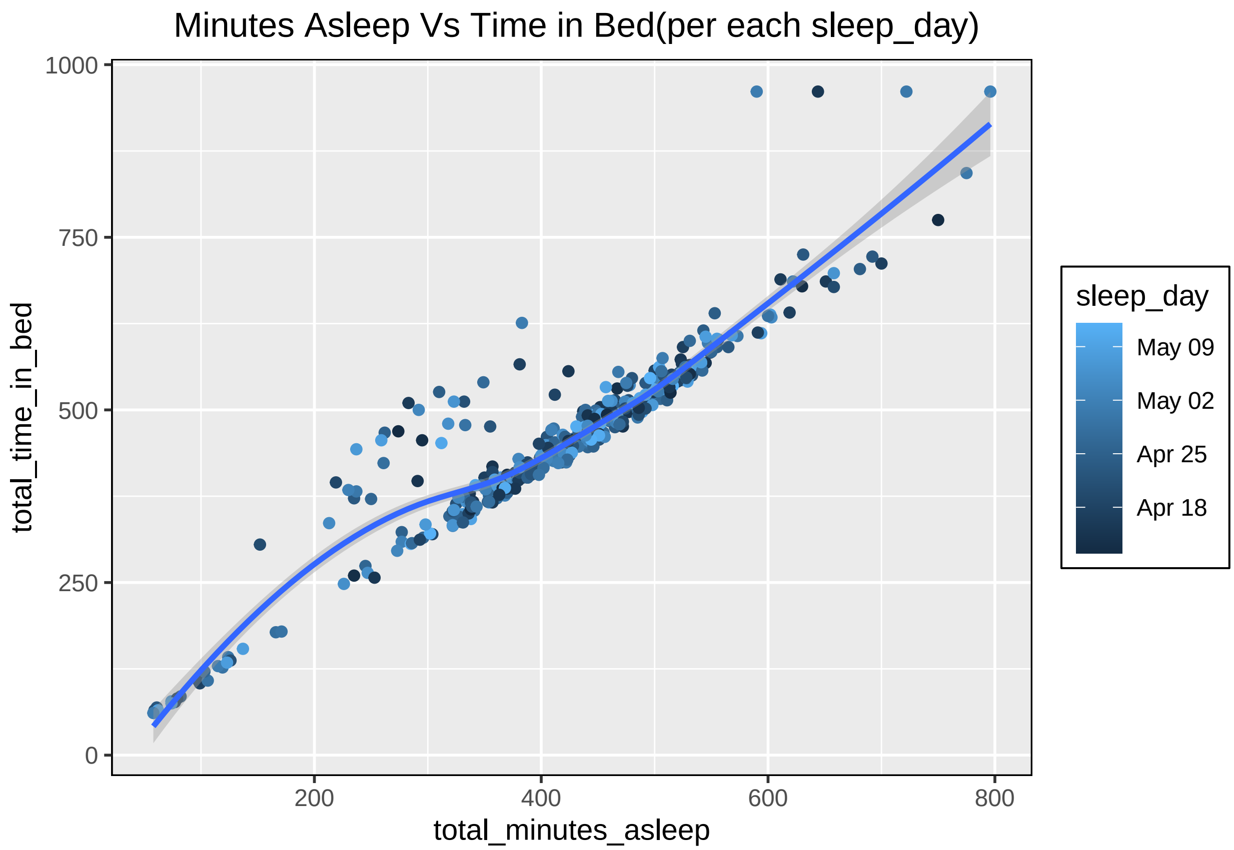
4.1.4 : Exploring the Relationship Between Minutes Asleep and Time in Bed (as per each sleep-day)

```
ggplot(data = sleepday1, aes(x= total_minutes_asleep, y = total_time_in_bed, color = sleep_day))+
  geom_point()+
  geom_smooth()+
  labs(title = " Minutes Asleep Vs Time in Bed(per each sleep_day)")+
  theme(plot.title = element_text(hjust = 0.5),
        panel.border = element_rect(color = 'black',  fill = NA),
        legend.background = element_blank(),
        legend.box.background = element_rect(colour = "black"))
```

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

```
## Warning: The following aesthetics were dropped during statistical transformation: colour
```

```
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```




The general distribution of data points and the smooth line provides insights in the correlation between the minutes asleep and time in bed. if the data points cluster closely around the smooth line, it indicates a strong relationship. Thus, the plot indicates a strong positive relationship between the minutes spent during the participants sleep time and their time in bed.

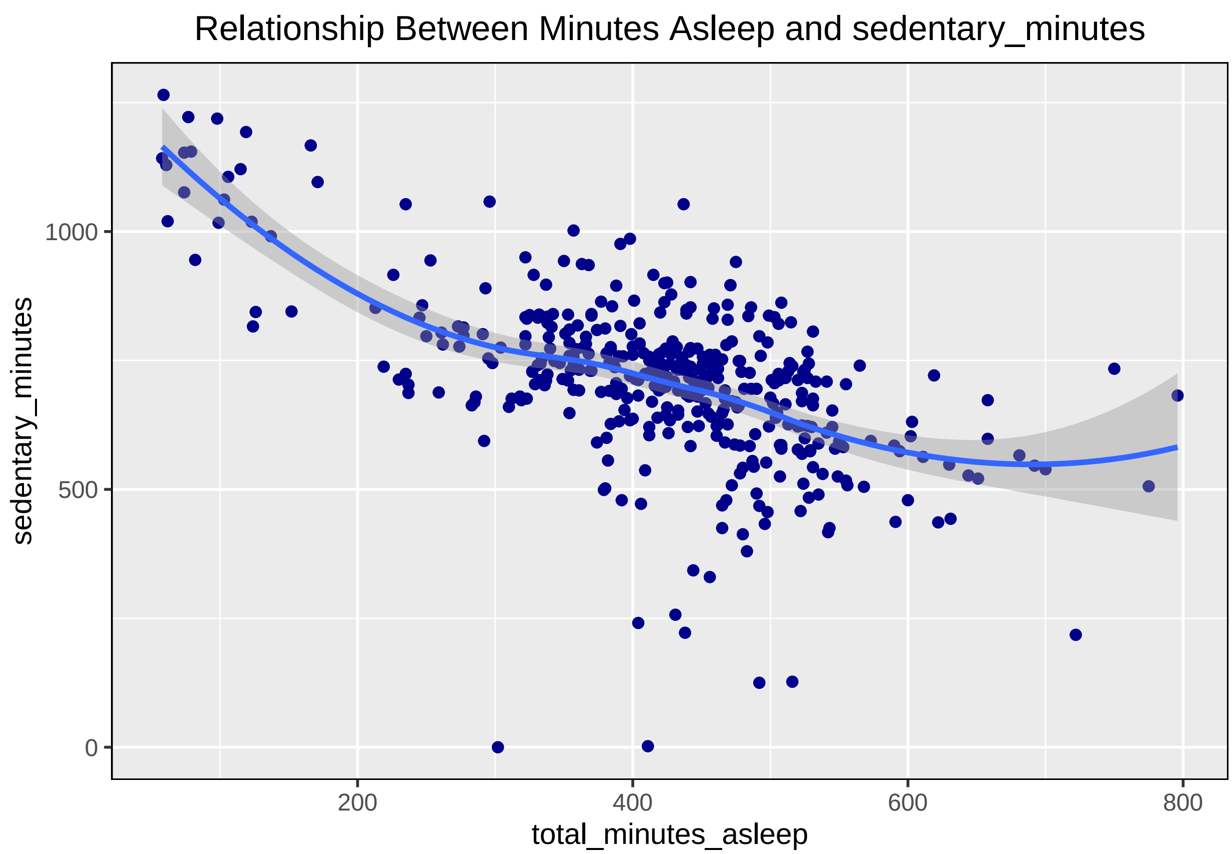
However, the plot also exhibits significant fluctuations and outliers. the fluctuations indicate changes in sleep duration or time in bed for specific sleep days. The Outliers, represented by data points that deviate significantly from the general trend, might suggest unique sleep patterns or abnormal sleep behaviors for certain days.

For this reason, there is need to check the correlation and effects of the fluctuations and outliers. In this case, we can explore the relationship time spent asleep and its impact on sedentary time spent.

4.1.5 : Exploring the Correlation Between Minutes Asleep and sedentary_minutes

```
ggplot(data = merged_data, aes(x= total_minutes_asleep, y = sedentary_minutes))+
  geom_point(color ="darkblue")+
  geom_smooth()+
  labs(title = "Relationship Between Minutes Asleep and sedentary_minutes")+
  theme(plot.title = element_text(hjust = 0.5),
        panel.border = element_rect(color = 'black',  fill = NA))
```

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




The plot confirms a negative relationship between the minutes spent asleep and the amount of sedentary time spent. the fluctuations and outliers in plot 4.1.4 indicate that as the number of minutes spent asleep increases, the number of sedentary minutes tends to decrease. This suggests that individuals who sleep longer may engage in fewer sedentary activities.

There is a rigid/tight cluster of data points around the smoothline compared to the few data points being scattered indicating a convincing negative relationship between the minutes spent asleep and the amount of sedentary time spent.

Recommendation : In this case, for users to improve their sleep, the company can recommend reducing sedentary time

# ACT

Conclusion

The data provided by Fitbit user activity indicates that most participants consistently provided data for the steps, intensity, sleep and calories activities while only few participants tracked weight and heart rate activities.

The data collected has its own limitations, such as being an outdated data set collected in 2016, which does not represent the prevailing consumer behavior. Additionally, the lack of demographic data limits our ability to make conclusive decisions on specific user groups. For this reason, research data provided in the future should contain more recent data and should include demographic information such as device type, location etc, which will provide a comprehensive understanding of consumer behavior.

5.1 Recommendations

For the Bellabeat's success in becoming a global smart device user, the analysis conducted recomends the following business decisions:

The company has the opportunity to enhance its application by integrating a personalized sleep journal feature. This addition would enable users to effectively monitor their sleep patterns and make meaningful adjustments to improve their overall rest quality. By incorporating a sleep journal, users can proactively address sedentary behaviors and prevent excessive sleep, thereby promoting healthier sleep habits. This measure serves as a strategic approach to encourage a balanced and optimized rest routine.

Based on the available data, it is evident that the most engaging activities predominantly take place during the time period of 5pm to 7pm. To further enhance user engagement and promote an active lifestyle, Bellabeat's application can incorporate daily reminders for users to participate in a diverse range of physical exercises, including activities such as running. This proactive approach will not only contribute to an increase in calorie expenditure but also aid in reducing sedentary behavior.

In addition to incorporating exercise reminders and integrating a sleep journal into the application, it is imperative for the company to prioritize the provision of healthy habits and a comprehensive diet program. This diet program will play a crucial role in regulating daily calorie intake by offering tailored low-calorie meal plans to individual users.

Finally, Bellabeat should consider providing a periodic report on a user's performance on a specified time period. Users can benefit from receiving comprehensive weekly performance reports, which serve as valuable tools for motivation and self-improvement. These reports provide users with a holistic overview of their activities and progress over the course of a week. By reviewing these reports, users can gain insights into their achievements, identify areas for improvement, and set goals for the upcoming week. This feature empowers users to stay motivated and actively track their progress, fostering a sense of accountability and continual growth towards their health and wellness goals.

Thank you for reading! This is my first case study and will sincerely appreciate your insights and constructive feedback. While this marks my inaugural case study, I am eager to not only improve my skills in data analysis but also project presentation. I am open to any criticisms and corrections you may provide. Thank you once again!