

Bellabeat_CaseStudy

Jung-Cheng Huang

9/7/2021

About the company

Bellabeat, a high-tech company that manufactures health-focused smart products. By collecting data on their users, they want to design more powerful technology that informs and inspires women around the world. Therefore, the company focus on analyzing previous product usage information in order to gain user insights and the marketing strategy. In the end, Bellabeat can become a potential player in the global smart device market.

Phase 1: Ask

1. Identify the business task:

Find the trends or information from the usage of their fitness smart devices and app. Furthermore, give high-level recommendations which can inform Bellabeat marketing strategy in the future.

On the other hand, in one phrase it would be: What's the trend for the user when using our product?

2. Consider key stakeholders:

The main stakeholders here are Urška Sršen, Bellabeat's co-founder and Chief Creative Officer; Sando Mur, Mathematician and Bellabeat's cofounder; And the rest of the Bellabeat marketing analytics team.

Phase 2: Prepare

1. The credibility of the data:

The data set is from FitBit Fitness Tracker Data which 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. It includes information about daily activity, steps, and heart rate that can be used to explore users' habits.

2. Sort and filter the data:

In order to find some useful insights, I need to focus on analyzing the group dataset and personal information. Furthermore, we might see what the relationship between Sleep quality and other variables is. In other words, it is great to approach different variables to understand the result of them. For instance, I am going to analyze each person's data and we can realize the key facts for great sleep quality.

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5      v purrr  0.3.4
## v tibble  3.1.4      v dplyr  1.0.7
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   2.0.1      v forcats 0.5.1
```

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

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

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

```
library(knitr)
```

```
library(dplyr)
```

```
library(ggplot2)
```

Input the datasets

```
Daily_Activity <- read.csv("/Users/dodsonhuang/Desktop/20210904_Bellabeat_Case_Study/Data/Fitabase Data
```

```
Daily_Calories <- read.csv("/Users/dodsonhuang/Desktop/20210904_Bellabeat_Case_Study/Data/Fitabase Data
```

```
Daily_Intensities <- read.csv("/Users/dodsonhuang/Desktop/20210904_Bellabeat_Case_Study/Data/Fitabase D
```

```
Daily_Sleep <- read.csv("/Users/dodsonhuang/Desktop/20210904_Bellabeat_Case_Study/Data/Fitabase Data 4.
```

preview the datasets

```
head(Daily_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
## 1	13	328	728
## 2	19	217	776
## 3	11	181	1218
## 4	34	209	726
## 5	10	221	773
## 6	20	164	539

```
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
```

```
#kable(Daily_Activity[1:10,])
```

```
# Get the summary for each data set
```

```
# Daily Activity
```

```
summary(Daily_Activity)
```

```
##      Id      ActivityDate      TotalSteps      TotalDistance
## Min.   :1.504e+09 Length:940      Min.    : 0      Min.    : 0.000
## 1st Qu.:2.320e+09 Class :character 1st Qu.: 3790   1st Qu.: 2.620
## Median :4.445e+09 Mode  :character Median : 7406   Median : 5.245
## Mean   :4.855e+09      Mean   : 7638   Mean    : 5.490
## 3rd Qu.:6.962e+09      3rd Qu.:10727  3rd Qu.: 7.713
## Max.   :8.878e+09      Max.    :36019  Max.    :28.030
## TrackerDistance LoggedActivitiesDistance VeryActiveDistance
## Min.    : 0.000 Min.    :0.0000 Min.    : 0.000
## 1st Qu.: 2.620 1st Qu.:0.0000 1st Qu.: 0.000
## Median : 5.245 Median :0.0000 Median : 0.210
## Mean   : 5.475 Mean   :0.1082 Mean   : 1.503
## 3rd Qu.: 7.710 3rd Qu.:0.0000 3rd Qu.: 2.053
## Max.   :28.030 Max.   :4.9421 Max.   :21.920
## ModeratelyActiveDistance LightActiveDistance SedentaryActiveDistance
## Min.    :0.0000 Min.    : 0.000 Min.    :0.000000
## 1st Qu.:0.0000 1st Qu.: 1.945 1st Qu.:0.000000
## Median :0.2400 Median : 3.365 Median :0.000000
## Mean   :0.5675 Mean   : 3.341 Mean   :0.001606
## 3rd Qu.:0.8000 3rd Qu.: 4.782 3rd Qu.:0.000000
## Max.   :6.4800 Max.   :10.710 Max.   :0.110000
## VeryActiveMinutes FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes
## Min.    : 0.00 Min.    : 0.00 Min.    : 0.0 Min.    : 0.0
## 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.:127.0 1st Qu.: 729.8
## Median : 4.00 Median : 6.00 Median :199.0 Median :1057.5
## Mean   : 21.16 Mean   :13.56 Mean   :192.8 Mean   : 991.2
## 3rd Qu.: 32.00 3rd Qu.:19.00 3rd Qu.:264.0 3rd Qu.:1229.5
## Max.   :210.00 Max.   :143.00 Max.   :518.0 Max.   :1440.0
##      Calories
## Min.    : 0
## 1st Qu.:1828
## Median :2134
## Mean   :2304
## 3rd Qu.:2793
## Max.   :4900
```

```
# Daily Sleep
Daily_Sleep %>%
  select(TotalSleepRecords, TotalMinutesAsleep, TotalTimeInBed) %>%
  summary()
```

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

Inner join the dataset

```
# Change the format in order to join the two tables(Daily_Activity and Daily_Sleep)
```

```
# Daily_Activity
```

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

```
# Daily_Sleep
```

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

```
# Merge the two tables by Id and date.
```

```
Daily_Activity_with_Sleep <- merge(Daily_Activity, Daily_Sleep, by = c('Id', 'date'))
```

```
# Review the new data set
```

```
Daily_Activity_with_Sleep$Id <- as.factor(Daily_Activity_with_Sleep$Id)
head(Daily_Activity_with_Sleep)
```

```
##           Id      date ActivityDate TotalSteps TotalDistance TrackerDistance
## 1 1503960366 04/12/16 2016-04-12      13162           8.50           8.50
## 2 1503960366 04/13/16 2016-04-13      10735           6.97           6.97
## 3 1503960366 04/15/16 2016-04-15       9762           6.28           6.28
## 4 1503960366 04/16/16 2016-04-16      12669           8.16           8.16
## 5 1503960366 04/17/16 2016-04-17       9705           6.48           6.48
## 6 1503960366 04/19/16 2016-04-19      15506           9.88           9.88
## LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 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
## LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
## 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
##   FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories   SleepDay
## 1          13          328          728    1985 2016-04-12
## 2          19          217          776    1797 2016-04-13
## 3          34          209          726    1745 2016-04-15
## 4          10          221          773    1863 2016-04-16
## 5          20          164          539    1728 2016-04-17
## 6          31          264          775    2035 2016-04-19
##   TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
## 1          1          327          346
## 2          2          384          407
## 3          1          412          442
## 4          2          340          367
## 5          1          700          712
## 6          1          304          320
```

```
summary(Daily_Activity_with_Sleep)
```

```
##           Id           date           ActivityDate
## 8378563200: 32   Length:413   Min.   :2016-04-12 00:00:00
## 5553957443: 31   Class :character 1st Qu.:2016-04-19 00:00:00
## 6962181067: 31   Mode  :character Median :2016-04-27 00:00:00
## 2026352035: 28                                     Mean  :2016-04-26 12:40:05
## 3977333714: 28                                     3rd Qu.:2016-05-04 00:00:00
## 4445114986: 28                                     Max.   :2016-05-12 00:00:00
## (Other)      :235
##   TotalSteps TotalDistance TrackerDistance LoggedActivitiesDistance
## Min.   : 17   Min.   : 0.010   Min.   : 0.010   Min.   :0.0000
## 1st Qu.: 5206 1st Qu.: 3.600   1st Qu.: 3.600   1st Qu.:0.0000
## Median : 8925 Median : 6.290   Median : 6.290   Median :0.0000
## Mean   : 8541 Mean   : 6.039   Mean   : 6.034   Mean   :0.1131
## 3rd Qu.:11393 3rd Qu.: 8.030   3rd Qu.: 8.020   3rd Qu.:0.0000
## Max.   :22770 Max.   :17.540   Max.   :17.540   Max.   :4.0817
##
##   VeryActiveDistance ModeratelyActiveDistance LightActiveDistance
## Min.   : 0.00   Min.   :0.0000   Min.   :0.010
## 1st Qu.: 0.00   1st Qu.:0.0000   1st Qu.:2.540
## Median : 0.57   Median :0.4200   Median :3.680
## Mean   : 1.45   Mean   :0.7502   Mean   :3.807
## 3rd Qu.: 2.37   3rd Qu.:1.0400   3rd Qu.:4.930
## Max.   :12.54   Max.   :6.4800   Max.   :9.480
##
##   SedentaryActiveDistance VeryActiveMinutes FairlyActiveMinutes
## Min.   :0.0000000   Min.   : 0.00   Min.   : 0.00
## 1st Qu.:0.0000000   1st Qu.: 0.00   1st Qu.: 0.00
## Median :0.0000000   Median : 9.00   Median :11.00
## Mean   :0.0009201   Mean   :25.19   Mean   :18.04
## 3rd Qu.:0.0000000   3rd Qu.:38.00   3rd Qu.:27.00
## Max.   :0.1100000   Max.   :210.00   Max.   :143.00
##
##   LightlyActiveMinutes SedentaryMinutes   Calories
```

```
## Min.      : 2.0          Min.      : 0.0      Min.      : 257
## 1st Qu.:158.0          1st Qu.: 631.0    1st Qu.:1850
## Median :208.0          Median : 717.0    Median :2220
## Mean    :216.9          Mean    : 712.2    Mean    :2398
## 3rd Qu.:263.0          3rd Qu.: 783.0    3rd Qu.:2926
## Max.     :518.0          Max.     :1265.0   Max.     :4900
##
##      SleepDay              TotalSleepRecords TotalMinutesAsleep
## Min.      :2016-04-12 00:00:00 Min.      :1.000      Min.      : 58.0
## 1st Qu.:2016-04-19 00:00:00 1st Qu.:1.000      1st Qu.:361.0
## Median :2016-04-27 00:00:00 Median :1.000      Median :433.0
## Mean    :2016-04-26 12:40:05 Mean    :1.119      Mean    :419.5
## 3rd Qu.:2016-05-04 00:00:00 3rd Qu.:1.000      3rd Qu.:490.0
## Max.     :2016-05-12 00:00:00 Max.     :3.000      Max.     :796.0
##
## TotalTimeInBed
## Min.      : 61.0
## 1st Qu.:403.0
## Median :463.0
## Mean    :458.6
## 3rd Qu.:526.0
## Max.     :961.0
##
```

This merged data table combines the two datasets to help us analyze the information, thus we can choose several variables to approach it. For example, we might group the dataset by Id and we could see the interesting insights.

Phase 3 & 4: Process and Analyse

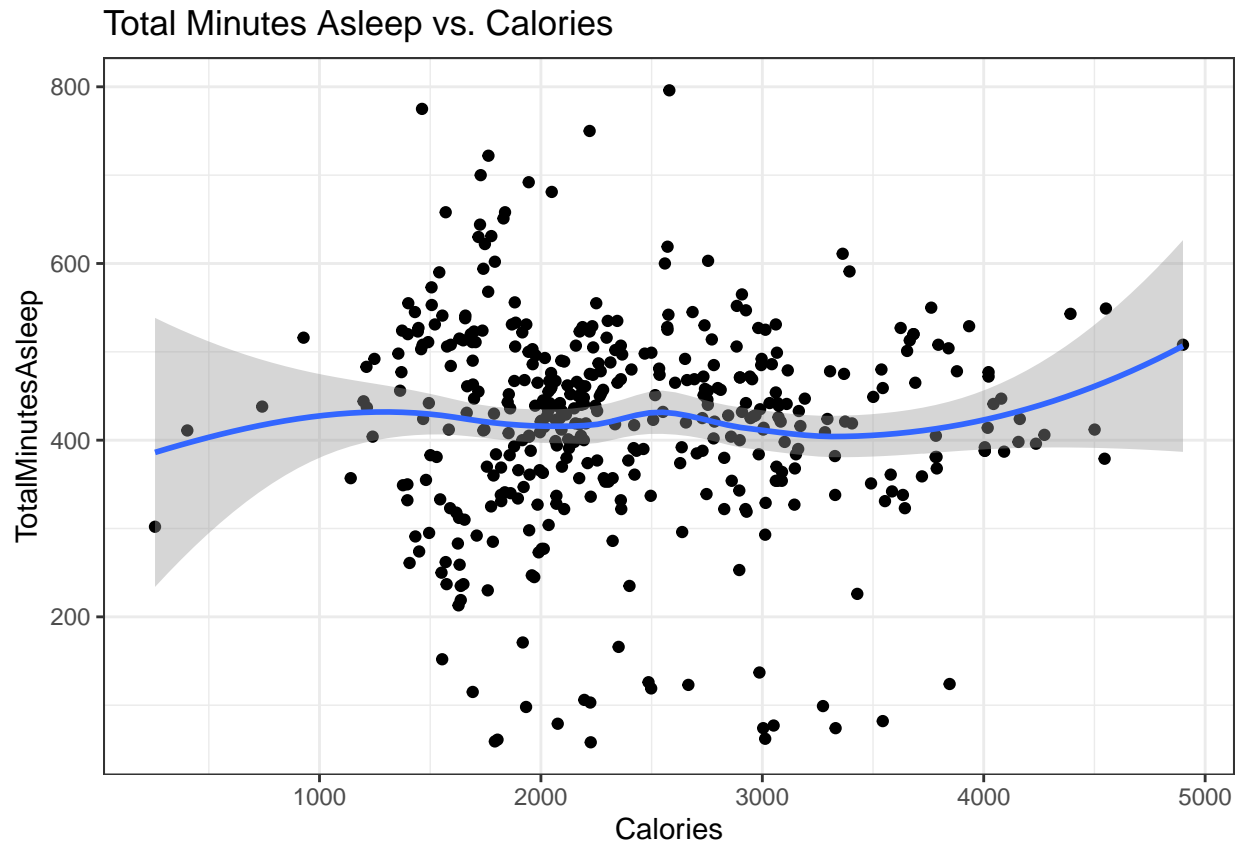
In this process, we will find the relationship for each variable such as Total Minutes Asleep vs. Calories. On the other hand, we might understand why someone has a high sleep time. Therefore, I will use the Plotly function in order to get the dimension plot for this approach.

Let's see some interesting patterns in the following graphs:

Graph: Total Minutes Asleep vs. Calories

```
ggplot(data = Daily_Activity_with_Sleep, aes(x = Calories , y = TotalMinutesAsleep)) +
  geom_point() + geom_smooth() + labs(title = "Total Minutes Asleep vs. Calories") + theme_bw()
```

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



ANALYSIS:

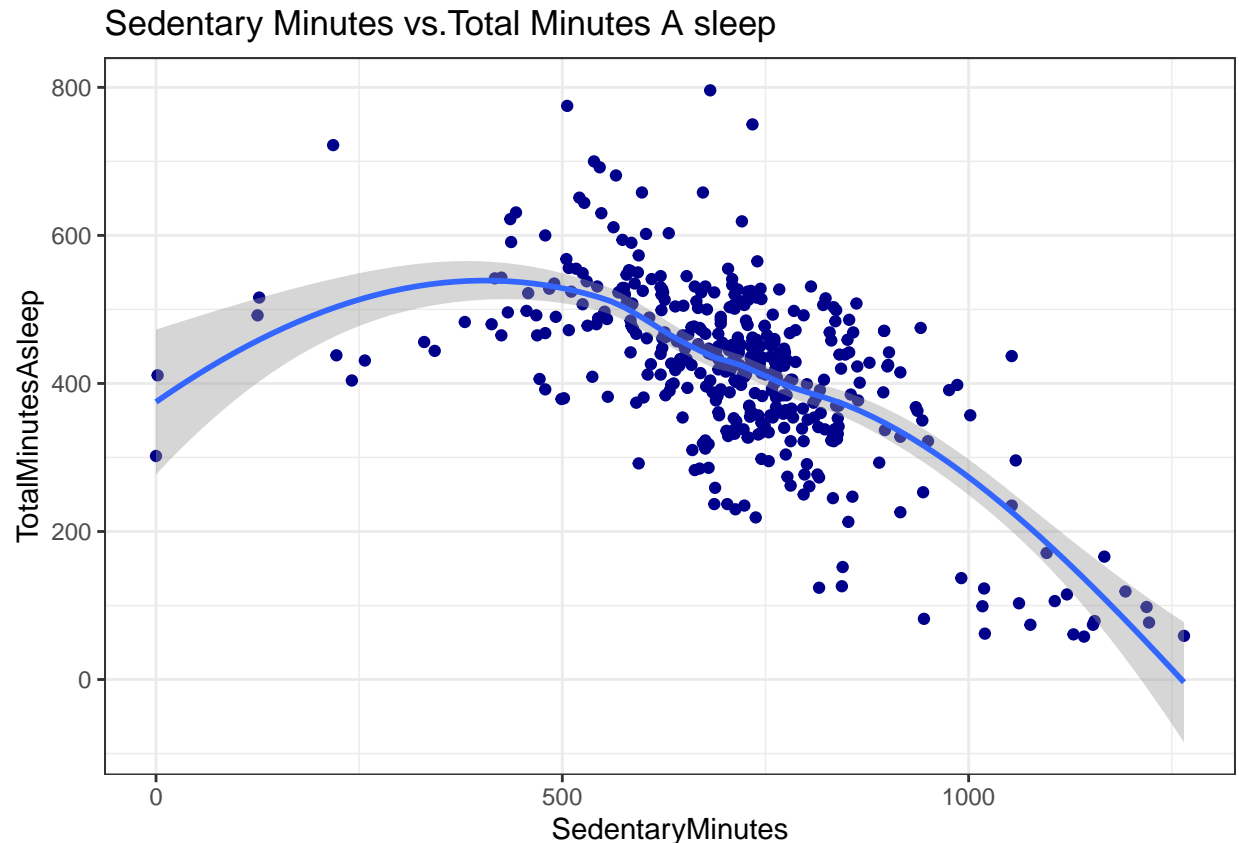
HINT: CDC Recommended Hours of Sleep Per Day for Adult(18–60 years) is 7 or more hours per night.

In the graph, we can see that Total Minutes Asleep and Calories might not have a positive or negative relationship. However, most data are central in average values. Therefore, we can ensure most people who spend 2400 calories could have 420 minutes of sleep per day. On the other side, spending 4000 calories per day would have higher sleep times. To conclude, Spend around 2400 calories can have great sleep quality. But it still depends on people's BMI or other facts.

Graph: Sedentary Minutes vs.Total Minutes A sleep

```
ggplot(data=Daily_Activity_with_Sleep, aes(x=SedentaryMinutes, y=TotalMinutesAsleep)) +
  geom_point(color='darkblue') + geom_smooth() +
  labs(title="Sedentary Minutes vs.Total Minutes A sleep") + theme_bw()
```

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

ANALYSIS:

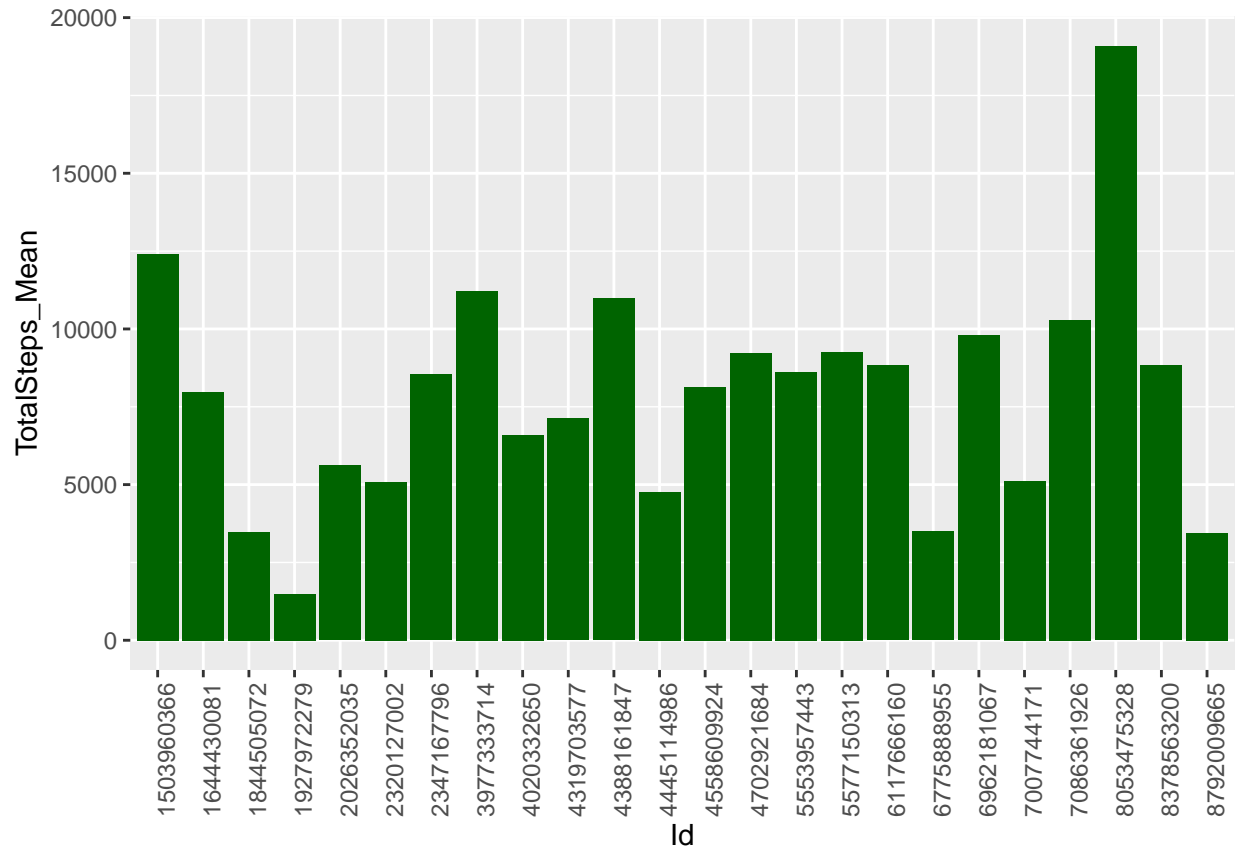
The graph shows that these two variables have a positive relationship which means Sedentary Minutes would impact Total Minutes of sleep. In other words, we can conclude that people need to avoid being sedentary in order to have higher hours of sleep per day. Thus, the app could give a sedentary reminder for people who want to have a higher quality of sleep. Furthermore, providing the suggestion article or new functions in-app will be a great choice for retain the users and increase the number of new customers.

Graph: Individual average of steps record

```
Daily_Activity_with_Sleep_New <- Daily_Activity_with_Sleep %>%
  group_by(Id) %>%
  summarise(TotalSteps_Mean = mean(TotalSteps))%>%
  arrange(desc(TotalSteps_Mean))

ggplot(data = Daily_Activity_with_Sleep_New,
  aes(x = Id,
    y = TotalSteps_Mean,
  )) +
  geom_histogram(stat = "identity", fill = "darkgreen") +
  theme(axis.text.x = element_text(angle = 90))
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



ANALYSIS:

Most people walk more than 5000 steps per day, but the average total steps per day are close to 7700. Therefore, we might use the app coach to remind people who could walk more steps rather than spending too much time on sedentary. In addition, we can interview people who have the lowest steps and highest sties in order to understand the customer deep thoughts.

Graph: Calories, Total Stpes, TotalMinutesAsleep,SedentaryMinutes

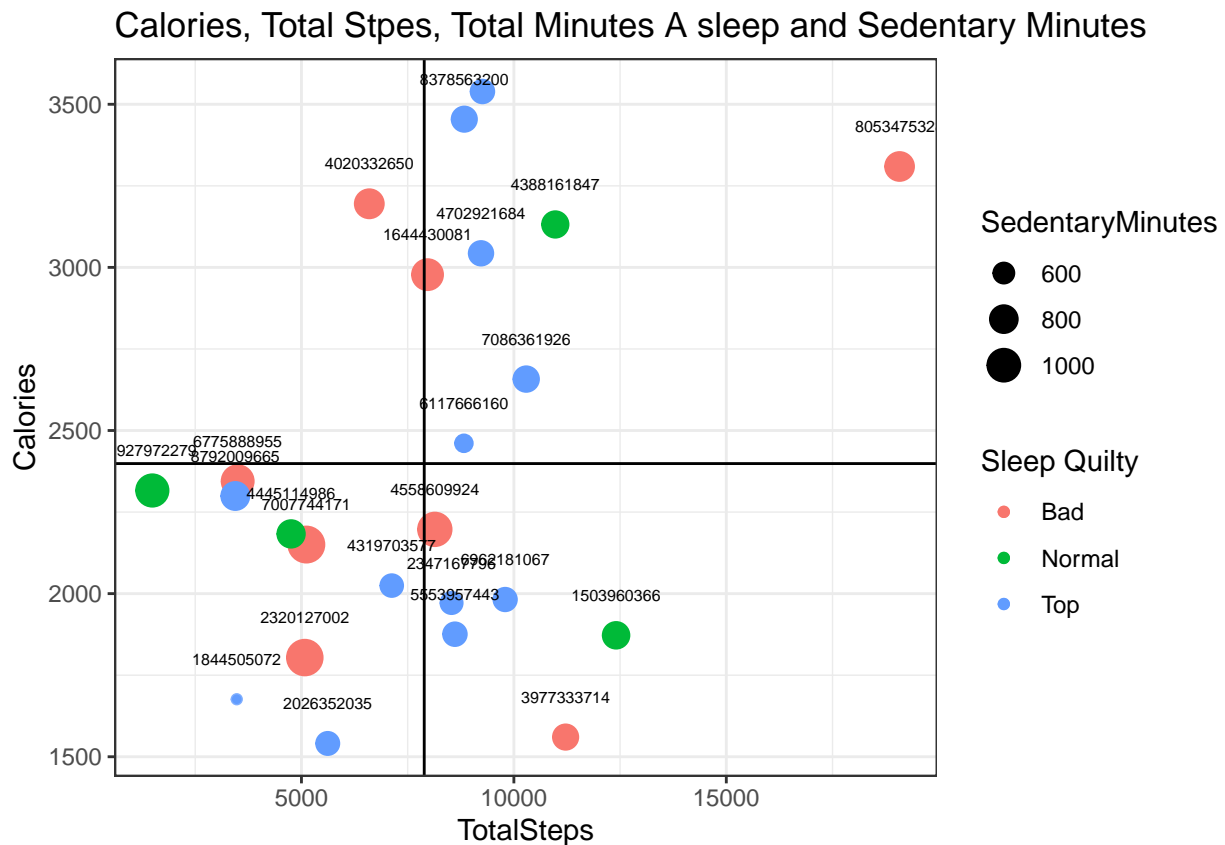
```
# Group by people's Id.
# Calculate the average of each variable.
# Based on the CDC recommendation, I use 6 and 7 hours to determine one's sleep quality and put them in
Daily_Activity_with_Sleep_forGraph <- Daily_Activity_with_Sleep %>%
  group_by(Id) %>%
  summarise( TotalSteps = mean(TotalSteps),
             Calories = mean(Calories),
             TotalMinutesAsleep = mean(TotalMinutesAsleep),
             SedentaryMinutes = mean(SedentaryMinutes)) %>%
  mutate(TotalMinutesAsleep_Group = ifelse(TotalMinutesAsleep > 420, 'Top',
                                           ifelse(TotalMinutesAsleep >= 360
                                                  & TotalMinutesAsleep < 420, 'Normal', 'Bad')) %>%
  arrange(desc(TotalSteps))

head(Daily_Activity_with_Sleep_forGraph)
```

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

##	Id	TotalSteps	Calories	TotalMinutesAsleep	SedentaryMinutes	TotalMinutesAsleep
## 1	8053475328	19079.	3309.	297	837.	Bad
## 2	1503960366	12406.	1872.	360.	759.	Normal
## 3	3977333714	11218	1560.	294.	716.	Bad
## 4	4388161847	10975.	3132.	403.	751.	Normal
## 5	7086361926	10290.	2658.	453.	724.	Top
## 6	6962181067	9795.	1982.	448	662.	Top

```
ggplot(data = Daily_Activity_with_Sleep_forGraph,
  aes( x = TotalSteps,
    y = Calories,
    color = TotalMinutesAsleep_Group)) +
  geom_point(alpha = 0.8) +
  geom_point(aes(size = SedentaryMinutes)) +
  geom_text(aes( label =Id), vjust = -3, size = 2, color ='black') +
  geom_vline( aes( xintercept = mean(TotalSteps))) +
  geom_hline( aes( yintercept = mean(Calories))) +
  labs( title = 'Calories, Total Stpes, Total Minutes A sleep and Sedentary Minutes',
    color = 'Sleep Quilty') +
  theme_bw()
```



ANAYSIS:

As the graph, we can find some interesting patterns from these four variables:

1. If people have a bigger size circle in the graph, they tend to be in the bad sleep quality group. In other

words, people have higher sedentary minutes, the total sleep time would be lower than 6 hours.

2. For the outliers, we can see that the red circle in the top-right corner. It shows this person has higher data for each variable, but she still has bad sleep quality.
3. From the plot, calories could be not a great metric to determine one's sleep quality and we need to research more information for these parts.

Phase 5 : Share

```
library(plotly)

##
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':
##
##   last_plot

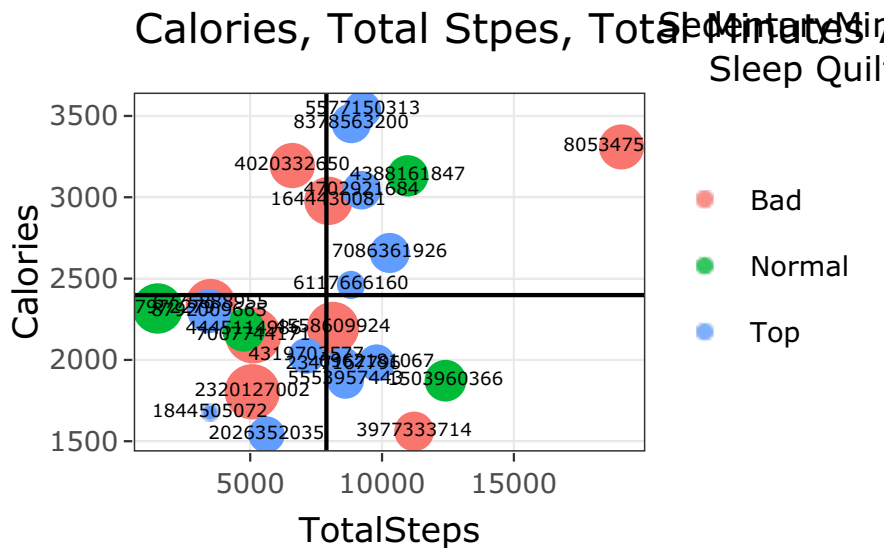
## The following object is masked from 'package:stats':
##
##   filter

## The following object is masked from 'package:graphics':
##
##   layout

DynamicPlot <- ggplot( data = Daily_Activity_with_Sleep_forGraph,
                      aes( x = TotalSteps,
                          y = Calories,
                          color = TotalMinutesAsleep_Group)) +
  geom_point( alpha = 0.8 ) +
  geom_point( aes( size = SedentaryMinutes )) +
  geom_text( aes( label = Id ), vjust = -3, size = 2, color = 'black') +
  geom_vline( aes( xintercept = mean(TotalSteps))) +
  geom_hline( aes( yintercept = mean(Calories))) +

  labs( title = 'Calories, Total Stpes, Total Minutes A sleep and Sedentary Minutes',
        color = 'Sleep Quilty') +
  theme_bw()

ggplotly(DynamicPlot)
```



Key objectives:

Conclusions

We get some useful findings:

As we mention before, we want to test the relationship between each variable and personal information. In order to find the relation and pattern, we combine the four variables into a dynamic graph. Furthermore, we could conclude three final insights and provide some marketing strategies.

1. Most people walk more than 5000 steps per day, but the average total steps per day are close to 7700. Therefore, we might use the app coach to remind people who could walk more steps rather than spending too much time on sedentary.
2. In this dataset, We can ensure most people who spend 2400 calories could have 420 minutes of sleep per day. However, calories could be not a great metric to determine one's sleep quality and we need to research more information for these parts.
3. People need to avoid being sedentary in order to have higher hours of sleep per day. In other words, people have higher sedentary minutes, the total sleep time would be lower than 6 hours.

Ideas for the Bellabeat app

1. Creating the sedentary reminder function for people who want to have higher sleep quality.
2. Adding the new online course or app coach function is a great approach to a new market. To illustrate, most people nowadays prefer to do exercise at home due to pandemics. Therefore, a home gym is a trend in the future and we can find some coach to cooperate for the different kinds of courses for women.
3. Bellabeat membership can provide free courses for our subscribers to retain the user and attracting new customers by providing free-trials sessions.