Bellabeat How Can a Wellness Technology Company Play It Smart?

Ivan Alarcon

2022-03-22

# Statement of business task

Bellabeat is a high-tech company that manufactures health-focused smart products collecting data that informs and inspires women around the world, Bellabeat can benefit from analyze the data of the different serves and products so the company can create new servers or campaigns and fined new opportunities or strategies that can help the company to grow.

Starting points:

1. Identified the trends that the industry has   
2. Identified the most used functionalities  
3. How to apply the new trends in our products that satisfied the costumers and the market demand

# Prepare the data

• FitBit Fitness Tracker Data dataset made available through Mobius in Kaggle  
• The data is under CC0 Public Domain licenses (Which give us the permission to use this data set)  
• This dataset generated by respondents to a distributed survey via Amazon Mechanical Turk between 03.12.2016-05.12.2016  
• Thirty eligible Fitbit users consented to the submission of personal tracker data

ROCCC analysis

• Reliable - Bad, we do not know if the people who did the survey are all women

• Original - Bad, is a third-party information  
• Comprehensive - Medium contain some data type errors   
• Current - Bad, the data is out of date   
• Cited - Good, the sources are well documented

# Process the data

library(tidyverse)  
library(dplyr)  
library(ggplot2)

# Upload the dataset

daily\_activity <- read.csv("C:/Users/augus/OneDrive/Documentos/Data\_Analysis/8\_Capstone\_project/Fitabase\_Data/dailyActivity\_merged.csv")  
daily\_calories <- read.csv("C:/Users/augus/OneDrive/Documentos/Data\_Analysis/8\_Capstone\_project/Fitabase\_Data/dailyCalories\_merged.csv")  
daily\_intensities <- read.csv("C:/Users/augus/OneDrive/Documentos/Data\_Analysis/8\_Capstone\_project/Fitabase\_Data/dailyIntensities\_merged.csv")  
daily\_sleep <- read.csv("C:/Users/augus/OneDrive/Documentos/Data\_Analysis/8\_Capstone\_project/Fitabase\_Data/sleepDay\_merged.csv")  
daily\_weight\_log <- read.csv("C:/Users/augus/OneDrive/Documentos/Data\_Analysis/8\_Capstone\_project/Fitabase\_Data/weightLogInfo\_merged.csv")

When we have all the dataset, we can start to explore the data.

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 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

head(daily\_weight\_log)

## 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

colnames(daily\_activity)

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

colnames(daily\_calories)

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

colnames(daily\_intensities)

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

colnames(daily\_sleep)

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

colnames(daily\_weight\_log)

## [1] "Id" "Date" "WeightKg" "WeightPounds"   
## [5] "Fat" "BMI" "IsManualReport" "LogId"

glimpse(daily\_activity)

## Rows: 940  
## Columns: 15  
## $ Id <dbl> 1503960366, 1503960366, 1503960366, 150396036~  
## $ ActivityDate <chr> "4/12/2016", "4/13/2016", "4/14/2016", "4/15/~  
## $ TotalSteps <int> 13162, 10735, 10460, 9762, 12669, 9705, 13019~  
## $ TotalDistance <dbl> 8.50, 6.97, 6.74, 6.28, 8.16, 6.48, 8.59, 9.8~  
## $ TrackerDistance <dbl> 8.50, 6.97, 6.74, 6.28, 8.16, 6.48, 8.59, 9.8~  
## $ LoggedActivitiesDistance <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~  
## $ VeryActiveDistance <dbl> 1.88, 1.57, 2.44, 2.14, 2.71, 3.19, 3.25, 3.5~  
## $ ModeratelyActiveDistance <dbl> 0.55, 0.69, 0.40, 1.26, 0.41, 0.78, 0.64, 1.3~  
## $ LightActiveDistance <dbl> 6.06, 4.71, 3.91, 2.83, 5.04, 2.51, 4.71, 5.0~  
## $ SedentaryActiveDistance <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~  
## $ VeryActiveMinutes <int> 25, 21, 30, 29, 36, 38, 42, 50, 28, 19, 66, 4~  
## $ FairlyActiveMinutes <int> 13, 19, 11, 34, 10, 20, 16, 31, 12, 8, 27, 21~  
## $ LightlyActiveMinutes <int> 328, 217, 181, 209, 221, 164, 233, 264, 205, ~  
## $ SedentaryMinutes <int> 728, 776, 1218, 726, 773, 539, 1149, 775, 818~  
## $ Calories <int> 1985, 1797, 1776, 1745, 1863, 1728, 1921, 203~

glimpse(daily\_calories)

## Rows: 940  
## Columns: 3  
## $ Id <dbl> 1503960366, 1503960366, 1503960366, 1503960366, 1503960366~  
## $ ActivityDay <chr> "4/12/2016", "4/13/2016", "4/14/2016", "4/15/2016", "4/16/~  
## $ Calories <int> 1985, 1797, 1776, 1745, 1863, 1728, 1921, 2035, 1786, 1775~

glimpse(daily\_intensities)

## Rows: 940  
## Columns: 10  
## $ Id <dbl> 1503960366, 1503960366, 1503960366, 150396036~  
## $ ActivityDay <chr> "4/12/2016", "4/13/2016", "4/14/2016", "4/15/~  
## $ SedentaryMinutes <int> 728, 776, 1218, 726, 773, 539, 1149, 775, 818~  
## $ LightlyActiveMinutes <int> 328, 217, 181, 209, 221, 164, 233, 264, 205, ~  
## $ FairlyActiveMinutes <int> 13, 19, 11, 34, 10, 20, 16, 31, 12, 8, 27, 21~  
## $ VeryActiveMinutes <int> 25, 21, 30, 29, 36, 38, 42, 50, 28, 19, 66, 4~  
## $ SedentaryActiveDistance <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~  
## $ LightActiveDistance <dbl> 6.06, 4.71, 3.91, 2.83, 5.04, 2.51, 4.71, 5.0~  
## $ ModeratelyActiveDistance <dbl> 0.55, 0.69, 0.40, 1.26, 0.41, 0.78, 0.64, 1.3~  
## $ VeryActiveDistance <dbl> 1.88, 1.57, 2.44, 2.14, 2.71, 3.19, 3.25, 3.5~

glimpse(daily\_sleep)

## Rows: 413  
## Columns: 5  
## $ Id <dbl> 1503960366, 1503960366, 1503960366, 1503960366, 150~  
## $ SleepDay <chr> "4/12/2016 12:00:00 AM", "4/13/2016 12:00:00 AM", "~  
## $ TotalSleepRecords <int> 1, 2, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~  
## $ TotalMinutesAsleep <int> 327, 384, 412, 340, 700, 304, 360, 325, 361, 430, 2~  
## $ TotalTimeInBed <int> 346, 407, 442, 367, 712, 320, 377, 364, 384, 449, 3~

glimpse(daily\_weight\_log)

## Rows: 67  
## Columns: 8  
## $ Id <dbl> 1503960366, 1503960366, 1927972279, 2873212765, 2873212~  
## $ Date <chr> "5/2/2016 11:59:59 PM", "5/3/2016 11:59:59 PM", "4/13/2~  
## $ WeightKg <dbl> 52.6, 52.6, 133.5, 56.7, 57.3, 72.4, 72.3, 69.7, 70.3, ~  
## $ WeightPounds <dbl> 115.9631, 115.9631, 294.3171, 125.0021, 126.3249, 159.6~  
## $ Fat <int> 22, NA, NA, NA, NA, 25, NA, NA, NA, NA, NA, NA, NA, NA,~  
## $ BMI <dbl> 22.65, 22.65, 47.54, 21.45, 21.69, 27.45, 27.38, 27.25,~  
## $ IsManualReport <chr> "True", "True", "False", "True", "True", "True", "True"~  
## $ LogId <dbl> 1.462234e+12, 1.462320e+12, 1.460510e+12, 1.461283e+12,~

We cand see that daily\_activity and daily\_intensities are very similar if not equals

glimpse(daily\_activity)

## Rows: 940  
## Columns: 15  
## $ Id <dbl> 1503960366, 1503960366, 1503960366, 150396036~  
## $ ActivityDate <chr> "4/12/2016", "4/13/2016", "4/14/2016", "4/15/~  
## $ TotalSteps <int> 13162, 10735, 10460, 9762, 12669, 9705, 13019~  
## $ TotalDistance <dbl> 8.50, 6.97, 6.74, 6.28, 8.16, 6.48, 8.59, 9.8~  
## $ TrackerDistance <dbl> 8.50, 6.97, 6.74, 6.28, 8.16, 6.48, 8.59, 9.8~  
## $ LoggedActivitiesDistance <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~  
## $ VeryActiveDistance <dbl> 1.88, 1.57, 2.44, 2.14, 2.71, 3.19, 3.25, 3.5~  
## $ ModeratelyActiveDistance <dbl> 0.55, 0.69, 0.40, 1.26, 0.41, 0.78, 0.64, 1.3~  
## $ LightActiveDistance <dbl> 6.06, 4.71, 3.91, 2.83, 5.04, 2.51, 4.71, 5.0~  
## $ SedentaryActiveDistance <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~  
## $ VeryActiveMinutes <int> 25, 21, 30, 29, 36, 38, 42, 50, 28, 19, 66, 4~  
## $ FairlyActiveMinutes <int> 13, 19, 11, 34, 10, 20, 16, 31, 12, 8, 27, 21~  
## $ LightlyActiveMinutes <int> 328, 217, 181, 209, 221, 164, 233, 264, 205, ~  
## $ SedentaryMinutes <int> 728, 776, 1218, 726, 773, 539, 1149, 775, 818~  
## $ Calories <int> 1985, 1797, 1776, 1745, 1863, 1728, 1921, 203~

glimpse(daily\_intensities)

## Rows: 940  
## Columns: 10  
## $ Id <dbl> 1503960366, 1503960366, 1503960366, 150396036~  
## $ ActivityDay <chr> "4/12/2016", "4/13/2016", "4/14/2016", "4/15/~  
## $ SedentaryMinutes <int> 728, 776, 1218, 726, 773, 539, 1149, 775, 818~  
## $ LightlyActiveMinutes <int> 328, 217, 181, 209, 221, 164, 233, 264, 205, ~  
## $ FairlyActiveMinutes <int> 13, 19, 11, 34, 10, 20, 16, 31, 12, 8, 27, 21~  
## $ VeryActiveMinutes <int> 25, 21, 30, 29, 36, 38, 42, 50, 28, 19, 66, 4~  
## $ SedentaryActiveDistance <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~  
## $ LightActiveDistance <dbl> 6.06, 4.71, 3.91, 2.83, 5.04, 2.51, 4.71, 5.0~  
## $ ModeratelyActiveDistance <dbl> 0.55, 0.69, 0.40, 1.26, 0.41, 0.78, 0.64, 1.3~  
## $ VeryActiveDistance <dbl> 1.88, 1.57, 2.44, 2.14, 2.71, 3.19, 3.25, 3.5~

We can do the compassion of daily\_activity, daily\_calories and daily\_intensities to see if they have the same information.

identical(daily\_activity$Id, daily\_calories$Id)

## [1] TRUE

identical(daily\_activity$ActivityDate, daily\_calories$ActivityDay)

## [1] TRUE

identical(daily\_activity$Calories, daily\_calories$Calories)

## [1] TRUE

identical(daily\_activity$Id, daily\_intensities$Id)

## [1] TRUE

identical(daily\_activity$ActivityDate, daily\_intensities$ActivityDay)

## [1] TRUE

identical(daily\_activity$SedentaryMinutes, daily\_intensities$SedentaryMinutes)

## [1] TRUE

identical(daily\_activity$LightlyActiveMinutes, daily\_intensities$LightlyActiveMinutes)

## [1] TRUE

identical(daily\_activity$FairlyActiveMinutes, daily\_intensities$FairlyActiveMinutes)

## [1] TRUE

identical(daily\_activity$VeryActiveMinutes, daily\_intensities$VeryActiveMinutes)

## [1] TRUE

identical(daily\_activity$SedentaryActiveDistance, daily\_intensities$SedentaryActiveDistance)

## [1] TRUE

identical(daily\_activity$LightActiveDistance, daily\_intensities$LightActiveDistance)

## [1] TRUE

identical(daily\_activity$ModeratelyActiveDistance, daily\_intensities$ModeratelyActiveDistance)

## [1] TRUE

identical(daily\_activity$VeryActiveDistance, daily\_intensities$VeryActiveDistance)

## [1] TRUE

As they are equals, we can use only 3 datasets.

# Data analysis

n\_distinct(daily\_activity$Id)

## [1] 33

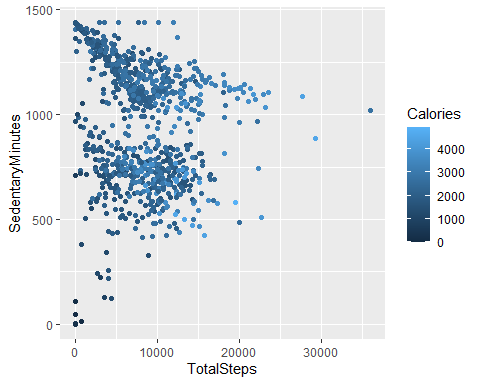
n\_distinct(daily\_sleep$Id)

## [1] 24

We can see that one dataset have more Id’s that the other, we need to have this in consideration if we need to combine the datasets.

## Steps Vs Sedentary minutes

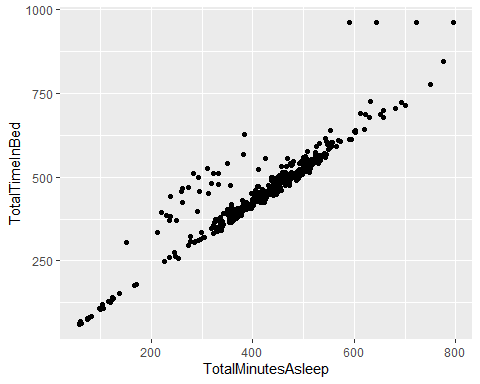
ggplot(data=daily\_activity, aes(x=TotalSteps, y=SedentaryMinutes, color = Calories)) + geom\_point()



As we can see those how have more steps burn more calories than those how have fewer even if have many sedentary minutes.

## Relation between time in bed and time sleeping

ggplot(data=daily\_sleep, aes(x=TotalMinutesAsleep, y=TotalTimeInBed)) + geom\_point()



We can see that the relationship between time sleeping and time in bed is lineal.

Combining daily\_sleep and daily\_activity

combined\_sleep\_activity <- merge(daily\_sleep, daily\_activity, by="Id", all = TRUE)  
n\_distinct(combined\_sleep\_activity$Id)

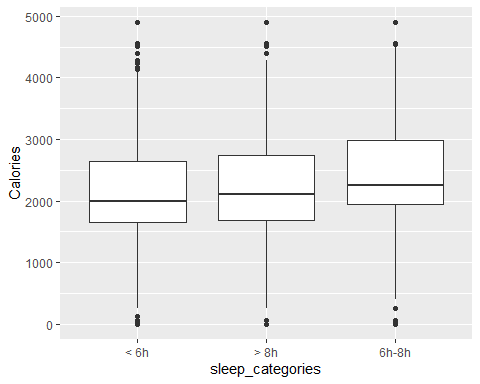
## [1] 33

Now we want to see if there is a relationship between a person with less calories and those who sleep well. we divide the time like this: less than 6h, 6h-8h and more than 8h.

combined\_sleep\_activity\_2 <- combined\_sleep\_activity %>%  
 mutate(sleep\_categories = case\_when(TotalMinutesAsleep >360 & TotalMinutesAsleep <= 480 ~ "6h-8h",  
 TotalMinutesAsleep > 480 ~ "> 8h",TRUE ~ "< 6h"))

**Good sleep and Loss of calories.**

ggplot(data= combined\_sleep\_activity\_2) + geom\_boxplot(mapping= aes(x=sleep\_categories, y= Calories))



Combining daily\_weight\_log and daily\_activity

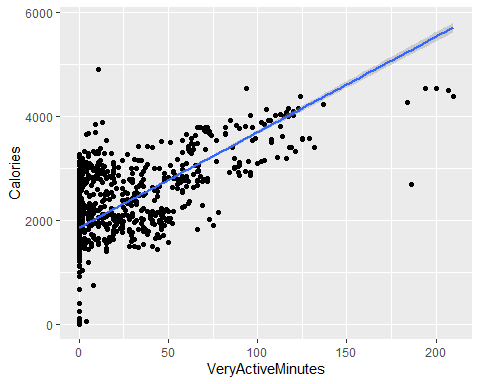
combined\_weight\_activity <- merge(daily\_weight\_log, daily\_activity, by="Id", all = TRUE)  
n\_distinct(combined\_weight\_activity$Id)

## [1] 33

## Loss of calories vs Very active minutes.

ggplot(data=combined\_weight\_activity, aes(x=VeryActiveMinutes, y=Calories)) + geom\_point() + stat\_smooth(method=lm)

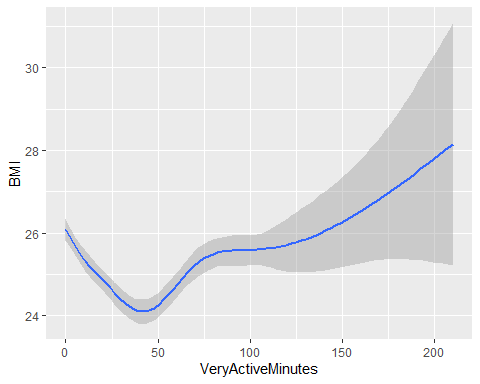
## `geom\_smooth()` using formula 'y ~ x'



**BMI vs Very active minutes**

ggplot(data=combined\_weight\_activity, aes(x=VeryActiveMinutes, y=BMI)) + geom\_smooth()

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



Combine daily\_sleep and daily\_weight\_log

combined\_sleep\_weight <- merge(daily\_sleep, daily\_weight\_log, by="Id", all = TRUE)  
n\_distinct(combined\_sleep\_weight$Id)

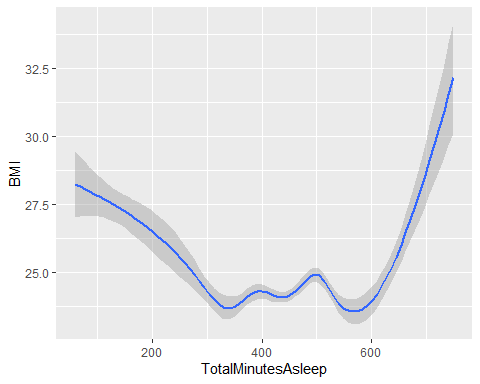
## [1] 26

## Relationship BMI vs Minutes asleep.

ggplot(data=combined\_sleep\_weight, aes(x=TotalMinutesAsleep, y=BMI)) + geom\_smooth()

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

## Warning: Removed 321 rows containing non-finite values (stat\_smooth).

 **Recommendations**

Based on the analysis that we perform we find some recommendations for the businesses.

1. The relationship between the activity and the calorie burn is evident, we can include feature when you have been too much time inactive, so in that way you could stretch your legs and walk for a while.
2. We can see a relation between sleep and the loss of calories, a feature where the user can see the quality of sleep and help them to improve their development.
3. Is also evident that people that walk more burn more calories, so we could include a feature that base of the information give the client a minimal steps per day, considering the sleep the BMI and other indicators.