Notebook Bellabeat Case Study

Klaus Keisers

2022-03-31

Bellabeat Case Study

Case Study 2: How Can a Wellness Technology Company Play It Smart?

Author: Klaus Keisers

Intro

In this case study, I am a junior data analyst who is working for the marketing analytics team at Bellabeat, a high-tech company that designs health tracking products for women. This hypothetical scenario is provided by Google's Data Analytics Certificate Program through Coursera, and I will be outlining the standard data analysis pathway throughout this project (ask, prepare, process, analyze, share, and act).

Contents:

1. Ask
2. Prepare
3. Process
4. Analyze
5. Share
6. Act

1. Ask:

Business Task:

Analyzing data from smart devices outside of your company to gain information that helps the company to unlock new growth opportunities.

Stakeholders:

- Urska Srsen: Chief Creative Officer and Cofounder
- Sando Mur: Key Member of the Bellabeat executive team
- Bellabeat marketing team: Team of data analyst

2. Prepare:

The data is public data from FitBit Fitness Tracker Data. It's a dataset from thirty fitbit users that includes minute-level output for physical activity, heart rate, and sleep monitoring. It's a good database segmented in several tables with different aspects of the data of the device with lots of details about the user behaviour.

I am going to focus on daily patterns like Activity, Calories, Intensities and Steps. So I am using just the tables who are representing this kind of data.

```r
##installing and running the needed packages for this caase study
install.packages("tidyverse")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)
```

```r
library(tidyverse)
```

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

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.8
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(ggplot2)
library(readr)
```

```r
##importing the needed dataset to see daily patterns
daily_activity <- read_csv("/cloud/project/Bellabeat_Case_Study/dailyActivity_merged.csv")
```

```
## Rows: 940 Columns: 15
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (1): ActivityDate
## dbl (14): Id, TotalSteps, TotalDistance, TrackerDistance, LoggedActivitiesDi...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
daily_calories <- read_csv("/cloud/project/Bellabeat_Case_Study/dailyCalories_merged.csv")
```

```
## Rows: 940 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): ActivityDay
## dbl (2): Id, Calories
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
daily_intensities <- read_csv("/cloud/project/Bellabeat_Case_Study/dailyIntensities_merged.csv")
```

```
## Rows: 940 Columns: 10
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): ActivityDay
## dbl (9): Id, SedentaryMinutes, LightlyActiveMinutes, FairlyActiveMinutes, Ve...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
daily_steps <- read_csv("/cloud/project/Bellabeat_Case_Study/dailySteps_merged.csv")
```

```
## Rows: 940 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): ActivityDay
## dbl (2): Id, StepTotal
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
sleep <- read_csv("/cloud/project/Bellabeat_Case_Study/sleepDay_merged.csv")
```

```
## Rows: 413 Columns: 5
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): SleepDay
## dbl (4): Id, TotalSleepRecords, TotalMinutesAsleep, TotalTimeInBed
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
weight_log <- read_csv("/cloud/project/Bellabeat_Case_Study/weightLogInfo_merged.csv")
```

```
## Rows: 67 Columns: 8
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Date
## dbl (6): Id, WeightKg, WeightPounds, Fat, BMI, LogId
## lgl (1): IsManualReport
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
##preview of the datasets

head(daily_activity)
```

```
## # A tibble: 6 x 15
##         Id ActivityDate TotalSteps TotalDistance TrackerDistance LoggedActivitie~
##      <dbl> <chr>             <dbl>         <dbl>           <dbl>            <dbl>
## 1  1.50e9 4/12/2016         13162          8.5             8.5                0
## 2  1.50e9 4/13/2016         10735          6.97            6.97               0
## 3  1.50e9 4/14/2016         10460          6.74            6.74               0
## 4  1.50e9 4/15/2016          9762          6.28            6.28               0
## 5  1.50e9 4/16/2016         12669          8.16            8.16               0
## 6  1.50e9 4/17/2016          9705          6.48            6.48               0
## # ... with 9 more variables: VeryActiveDistance <dbl>,
## #   ModeratelyActiveDistance <dbl>, LightActiveDistance <dbl>,
## #   SedentaryActiveDistance <dbl>, VeryActiveMinutes <dbl>,
## #   FairlyActiveMinutes <dbl>, LightlyActiveMinutes <dbl>,
## #   SedentaryMinutes <dbl>, Calories <dbl>
```

```
head(daily_calories)
```

```
## # A tibble: 6 x 3
##           Id ActivityDay Calories
##        <dbl> <chr>          <dbl>
## 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)
```

```
## # A tibble: 6 x 10
##           Id ActivityDay SedentaryMinutes LightlyActiveMinutes FairlyActiveMinu~
##        <dbl> <chr>                  <dbl>                <dbl>             <dbl>
## 1 1503960366 4/12/2016                728                  328                13
## 2 1503960366 4/13/2016                776                  217                19
## 3 1503960366 4/14/2016               1218                  181                11
## 4 1503960366 4/15/2016                726                  209                34
## 5 1503960366 4/16/2016                773                  221                10
## 6 1503960366 4/17/2016                539                  164                20
## # ... with 5 more variables: VeryActiveMinutes <dbl>,
## #   SedentaryActiveDistance <dbl>, LightActiveDistance <dbl>,
## #   ModeratelyActiveDistance <dbl>, VeryActiveDistance <dbl>
```

```
head(daily_steps)
```

```
## # A tibble: 6 x 3
##           Id ActivityDay StepTotal
##        <dbl> <chr>           <dbl>
## 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(weight_log)
```

```
## # A tibble: 6 x 8
##           Id Date       WeightKg WeightPounds   Fat   BMI IsManualReport   LogId
##        <dbl> <chr>         <dbl>        <dbl> <dbl> <dbl> <lgl>            <dbl>
## 1 1503960366 5/2/2016 ~     52.6         116.    22  22.6 TRUE           1.46e12
## 2 1503960366 5/3/2016 ~     52.6         116.    NA  22.6 TRUE           1.46e12
## 3 1927972279 4/13/2016~    134.          294.    NA  47.5 FALSE          1.46e12
## 4 2873212765 4/21/2016~     56.7         125.    NA  21.5 TRUE           1.46e12
## 5 2873212765 5/12/2016~     57.3         126.    NA  21.7 TRUE           1.46e12
## 6 4319703577 4/17/2016~     72.4         160.    25  27.5 TRUE           1.46e12
```

```
head(sleep)
```

```
## # A tibble: 6 x 5
##           Id SleepDay           TotalSleepRecor~ TotalMinutesAsl~ TotalTimeInBed
##        <dbl> <chr>                         <dbl>            <dbl>          <dbl>
## 1 1503960366 4/12/2016 12:00:0~                1              327            346
## 2 1503960366 4/13/2016 12:00:0~                2              384            407
## 3 1503960366 4/15/2016 12:00:0~                1              412            442
## 4 1503960366 4/16/2016 12:00:0~                2              340            367
## 5 1503960366 4/17/2016 12:00:0~                1              700            712
## 6 1503960366 4/19/2016 12:00:0~                1              304            320
```

Credibility

I will be using the ROCCC framework to demonstrate the credibility.

Reliable - The dataset contains secondary data collected via a distributed survey by Amazon Mechanical Turk

Original - It is a public data set. So it is not original.

Comprehensive - There ere only 30 probands. So its unlikely that it covers a wide range of different variables.

Current - The data was collected between 03.12.2016 - 05.12.2016.

Cited - I didnt find any information whether these work has been cited alot.

3.Process:

Checking the Data for Errors and Cleaning the Data

-Duplicated data

-Irrelevant data

-Inconsistensies in the number of rows

-Inconsistensies in the number of participants

-NULL values

-Missing values

Duplicated data:

```
##removing all duplicated data
daily_activity<-daily_activity[!duplicated(daily_activity),]
daily_calories<-daily_calories[!duplicated(daily_calories),]
daily_intensities<-daily_intensities[!duplicated(daily_intensities),]
daily_steps<-daily_steps[!duplicated(daily_steps),]
sleep<-sleep[!duplicated(sleep),]
weight_log<-weight_log[!duplicated(weight_log),]

##Checking how many rows were removed
sum(duplicated(daily_activity))
```

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

```
sum(duplicated(daily_calories))
```

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

```
sum(duplicated(daily_intensities))
```

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

```
sum(duplicated(daily_steps))
```

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

```
sum(duplicated(sleep))
```

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

```
sum(duplicated(weight_log))
```

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

Irrelevant Data

```
## removing daytime from the data just leaving the date
updated_sleep_table <- sleep %>%
    separate(SleepDay, c("Date", "Time"), " ")
```

```
## 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, ...].
```

```
updated_weight_log <- weight_log %>%
    separate(Date, c("Date", "Time"), " ")
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 67 rows [1, 2, 3, 4,
## 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].
```

```
new_sleep_table <- subset(updated_sleep_table, select = -c(Time))
new_weight_log <- subset(updated_weight_log, select = -c(Time))
```

Inconsistensies in the number of rows

```
nrow(daily_activity)
```

```
## [1] 940
```

```
nrow(daily_calories)
```

```
## [1] 940
```

```
nrow(daily_intensities)
```

```
## [1] 940
```

```
nrow(daily_steps)
```

```
## [1] 940
```

```
nrow(new_sleep_table)
```

```
## [1] 410
```

```
nrow(new_weight_log)
```

```
## [1] 67
```

It seems like that 940 rows is the ideal or "normal" Number of Rows. At the sleep and weight table there are missing some rows. These 2 Tables are not completely accurate.

Inconsistensies in the number of participants

```
daily_activity$Id %>% n_distinct()
```

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

```
daily_calories$Id %>% n_distinct()
```

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

```
daily_intensities$Id %>% n_distinct()
```

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

```
daily_steps$Id %>% n_distinct()
```

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

```
new_sleep_table$Id %>% n_distinct()
```

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

```
new_weight_log$Id %>% n_distinct()
```

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

The number of probands differs in the different tables. The are supposed to be 30 participants in the data. Now it seems that there usually 33.Just the wheight_log and the sleeping_table are missing some participants.

Null Values

```
is.null(daily_activity)
```

```
## [1] FALSE
```

```
is.null(daily_calories)
```

```
## [1] FALSE
```

```
is.null(daily_steps)
```

```
## [1] FALSE
```

```
is.null(daily_intensities)
```

```
## [1] FALSE
```

```
is.null(new_sleep_table)
```

```
## [1] FALSE
```

```
is.null(new_weight_log)
```

```
## [1] FALSE
```

Missing values

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

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

```
sum(is.na(daily_calories))
```

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

```
sum(is.na(daily_intensities))
```

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

```
sum(is.na(daily_steps))
```

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

```
sum(is.na(new_sleep_table))
```

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

```
sum(is.na(new_weight_log))
```

```
## [1] 65
```

There are missing values in the weight log in the "Fat" Column. Because of so many missing values it is invalid so i am going to remove the Fat Column from the table.

```
##Removing the Fat Column and changing the name of the sleeping_log
weight_log <- subset(new_weight_log, Select = -c(Fat))
sleep_log <- new_sleep_table
```

4. Analyze:

Summarizing data

```
sapply(list(daily_activity, daily_calories, daily_intensities, daily_steps, sleep_log, weight_log),summa
```

```
## [[1]]
##        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
##
## [[2]]
##        Id             ActivityDay          Calories
##  Min.   :1.504e+09   Length:940         Min.   :   0
##  1st Qu.:2.320e+09   Class :character   1st Qu.:1828
##  Median :4.445e+09   Mode  :character   Median :2134
##  Mean   :4.855e+09                      Mean   :2304
##  3rd Qu.:6.962e+09                      3rd Qu.:2793
##  Max.   :8.878e+09                      Max.   :4900
##
## [[3]]
##        Id              ActivityDay        SedentaryMinutes LightlyActiveMinutes
##  Min.   :1.504e+09   Length:940         Min.   :   0.0   Min.   :  0.0
##  1st Qu.:2.320e+09   Class :character   1st Qu.: 729.8   1st Qu.:127.0
##  Median :4.445e+09   Mode  :character   Median :1057.5   Median :199.0
##  Mean   :4.855e+09                      Mean   : 991.2   Mean   :192.8
##  3rd Qu.:6.962e+09                      3rd Qu.:1229.5   3rd Qu.:264.0
##  Max.   :8.878e+09                      Max.   :1440.0   Max.   :518.0
##  FairlyActiveMinutes VeryActiveMinutes SedentaryActiveDistance
##  Min.   :  0.00      Min.   :  0.00    Min.   :0.000000
##  1st Qu.:  0.00      1st Qu.:  0.00    1st Qu.:0.000000
##  Median :  6.00      Median :  4.00    Median :0.000000
##  Mean   : 13.56      Mean   : 21.16    Mean   :0.001606
##  3rd Qu.: 19.00      3rd Qu.: 32.00    3rd Qu.:0.000000
##  Max.   :143.00      Max.   :210.00    Max.   :0.110000
##  LightActiveDistance ModeratelyActiveDistance VeryActiveDistance
##  Min.   : 0.000      Min.   :0.0000           Min.   : 0.000
##  1st Qu.: 1.945      1st Qu.:0.0000           1st Qu.: 0.000
##  Median : 3.365      Median :0.2400           Median : 0.210
##  Mean   : 3.341      Mean   :0.5675           Mean   : 1.503
##  3rd Qu.: 4.782      3rd Qu.:0.8000           3rd Qu.: 2.053
##  Max.   :10.710      Max.   :6.4800           Max.   :21.920
##
## [[4]]
##        Id              ActivityDay          StepTotal
##  Min.   :1.504e+09   Length:940         Min.   :    0
##  1st Qu.:2.320e+09   Class :character   1st Qu.: 3790
##  Median :4.445e+09   Mode  :character   Median : 7406
##  Mean   :4.855e+09                      Mean   : 7638
##  3rd Qu.:6.962e+09                      3rd Qu.:10727
##  Max.   :8.878e+09                      Max.   :36019
##
## [[5]]
##        Id                 Date           TotalSleepRecords TotalMinutesAsleep
##  Min.   :1.504e+09   Length:410         Min.   :1.00      Min.   : 58.0
##  1st Qu.:3.977e+09   Class :character   1st Qu.:1.00      1st Qu.:361.0
##  Median :4.703e+09   Mode  :character   Median :1.00      Median :432.5
##  Mean   :4.995e+09                      Mean   :1.12      Mean   :419.2
##  3rd Qu.:6.962e+09                      3rd Qu.:1.00      3rd Qu.:490.0
##  Max.   :8.792e+09                      Max.   :3.00      Max.   :796.0
##  TotalTimeInBed
##  Min.   : 61.0
##  1st Qu.:403.8
##  Median :463.0
##  Mean   :458.5
##  3rd Qu.:526.0
##  Max.   :961.0
##
## [[6]]
##        Id                 Date              WeightKg       WeightPounds
##  Min.   :1.504e+09   Length:67          Min.   : 52.60   Min.   :116.0
##  1st Qu.:6.962e+09   Class :character   1st Qu.: 61.40   1st Qu.:135.4
##  Median :6.962e+09   Mode  :character   Median : 62.50   Median :137.8  
##  Mean   :7.009e+09                      Mean   : 72.04   Mean   :158.8  
##  3rd Qu.:8.878e+09                      3rd Qu.: 85.05   3rd Qu.:187.5  
##  Max.   :8.878e+09                      Max.   :133.50   Max.   :294.3  
##                                                                         
##       Fat             BMI        IsManualReport      LogId            
##  Min.   :22.00   Min.   :21.45   Mode :logical   Min.   :1.460e+12  
##  1st Qu.:22.75   1st Qu.:23.96   FALSE:26        1st Qu.:1.461e+12  
##  Median :23.50   Median :24.39   TRUE :41        Median :1.462e+12  
##  Mean   :23.50   Mean   :25.19                   Mean   :1.462e+12  
##  3rd Qu.:24.25   3rd Qu.:25.56                   3rd Qu.:1.462e+12  
##  Max.   :25.00   Max.   :47.54                   Max.   :1.463e+12  
##  NA's   :65
```

Most important Activity Data:

1. Calories:

-Avg: 2304 Calories

-Max: 4900 Calories

2. Intensities:

-Avg of Very Active Minutes: 21.16

-Avg of Fairly Active Minutes: 13.56

-Avg of Lightly Active Minutes: 199.0

-Avg of Sedentary Active Minutes: 991.2

3. Steps:

-Avg: 7638 Steps

4. Sleep

-Total Minutes Asleep Avg: 419.2

-Total Time In Bed Avg: 458.5

5. Weight

-Average of Weight in kg: 72.04

-Average of BMI: 25.19

5. Share

Steps vs. Weight

Steps vs. Calories

Intensities vs. Calories

Steps vs. Sleep

Time taken to fall Asleep

Sleep vs. Weight

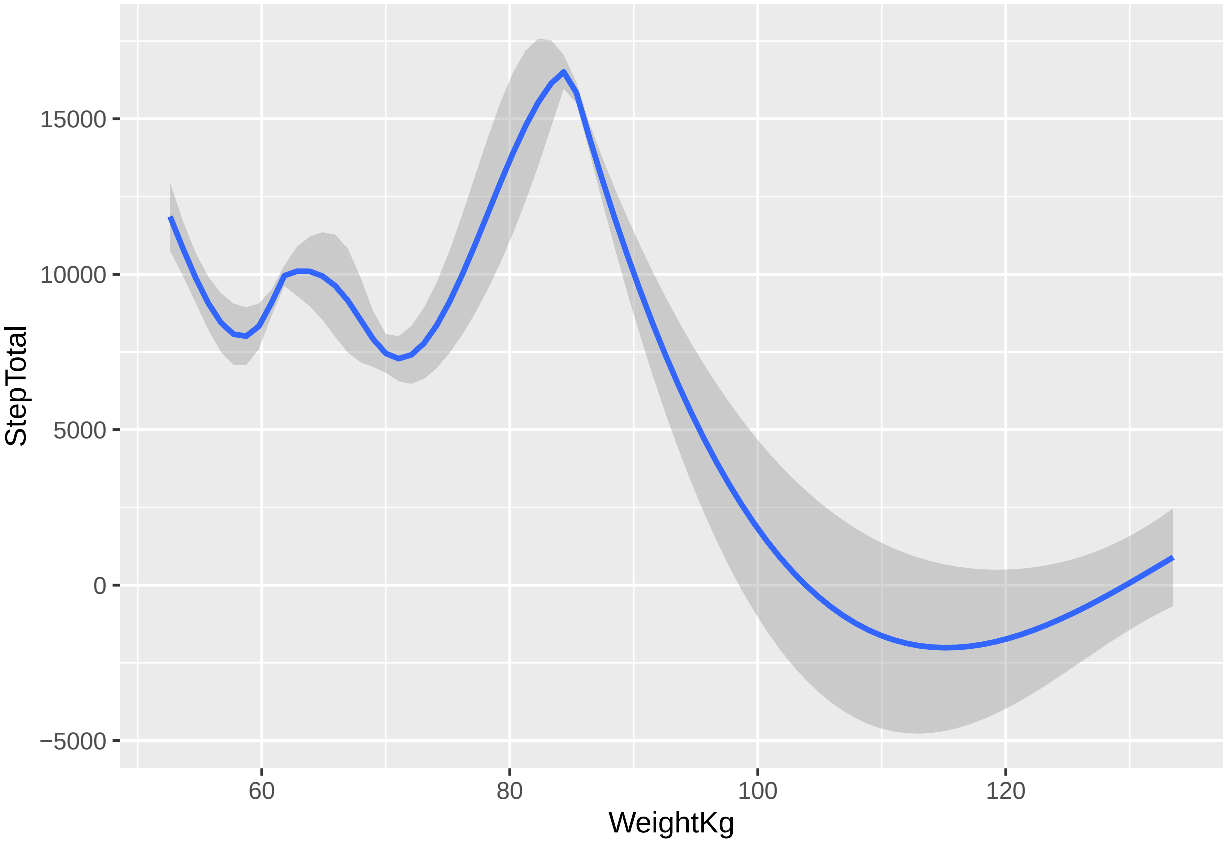
I want to see if there some Correlation between these topics. For example I want to see if it makes a difference how many steps a person took and how many minutes he slept.

```
## library dplr for JOIN Function, so that I can combine tables
library(dplyr)
steps_vs_weight <- daily_steps %>% inner_join(weight_log,by="Id")
print(steps_vs_weight)
```

```
## # A tibble: 2,076 x 10
##            Id ActivityDay StepTotal Date     WeightKg WeightPounds   Fat   BMI
##         <dbl> <chr>           <dbl> <chr>       <dbl>        <dbl> <dbl> <dbl>
##  1 1503960366 4/12/2016       13162 5/2/2016     52.6         116.    22  22.6
##  2 1503960366 4/12/2016       13162 5/3/2016     52.6         116.    NA  22.6
##  3 1503960366 4/13/2016       10735 5/2/2016     52.6         116.    22  22.6
##  4 1503960366 4/13/2016       10735 5/3/2016     52.6         116.    NA  22.6
##  5 1503960366 4/14/2016       10460 5/2/2016     52.6         116.    22  22.6
##  6 1503960366 4/14/2016       10460 5/3/2016     52.6         116.    NA  22.6
##  7 1503960366 4/15/2016        9762 5/2/2016     52.6         116.    22  22.6
##  8 1503960366 4/15/2016        9762 5/3/2016     52.6         116.    NA  22.6
##  9 1503960366 4/16/2016       12669 5/2/2016     52.6         116.    22  22.6
## 10 1503960366 4/16/2016       12669 5/3/2016     52.6         116.    NA  22.6
## # ... with 2,066 more rows, and 2 more variables: IsManualReport <lgl>,
## #   LogId <dbl>
```

```
steps_vs_weight_graphic <- ggplot(data=steps_vs_weight) +
  geom_smooth(mapping = aes(x = WeightKg, y = StepTotal))
print(steps_vs_weight_graphic)
```

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

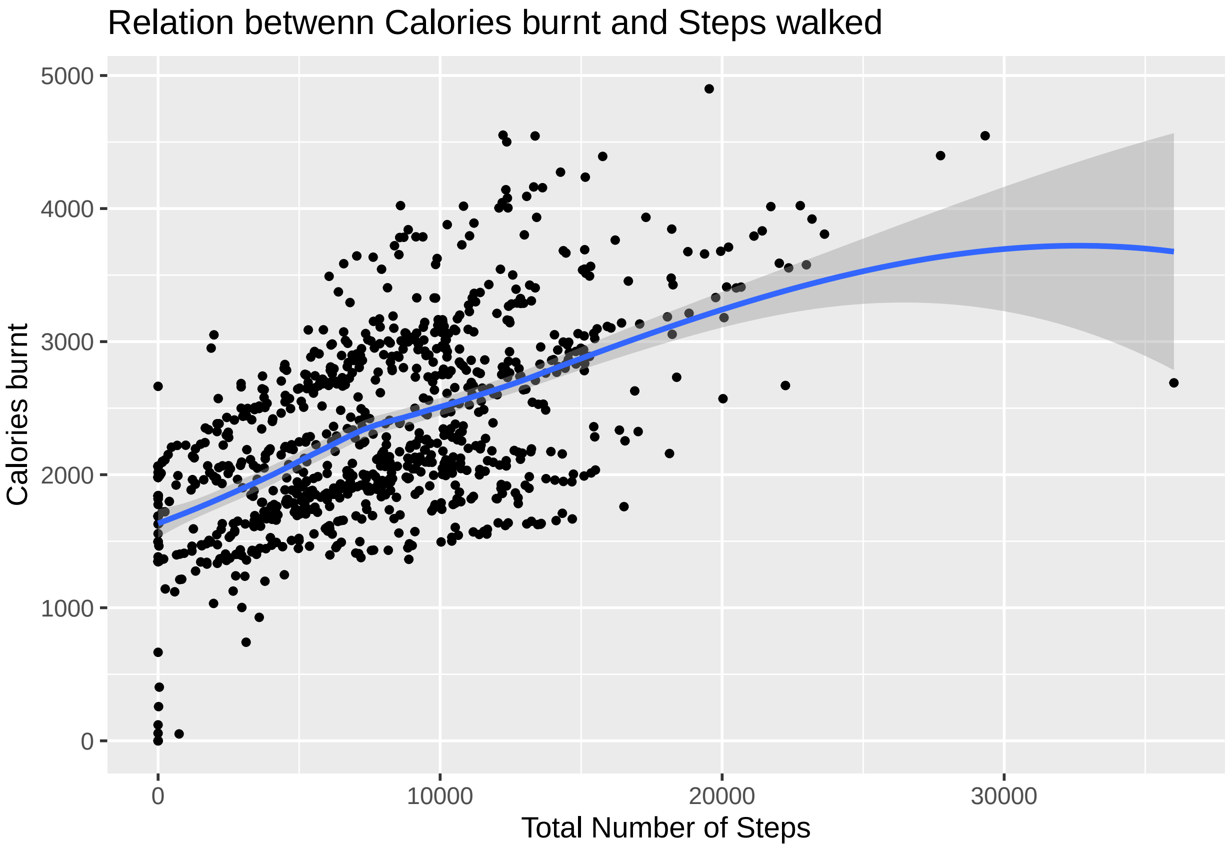


There is no obvious Correlation between Weight of the Person and the Step Total he took during the testing time.

Steps vs. Calories

```
library(ggplot2)

ggplot(data=daily_activity) +
  geom_jitter(width= .5, size=1, mapping = aes(x = TotalSteps, y = Calories))+
  geom_smooth(mapping = aes(x = TotalSteps, y = Calories))+
  labs(x="Total Number of Steps", y="Calories burnt",title = "Relation betwenn Calories burnt and Steps
```

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



There is a correlation between the Number of Steps someone walked and how much Calories he burnt. This can be used to show the customer that tracking your steps can help you to burn more Calories.

Intensities vs Calories:

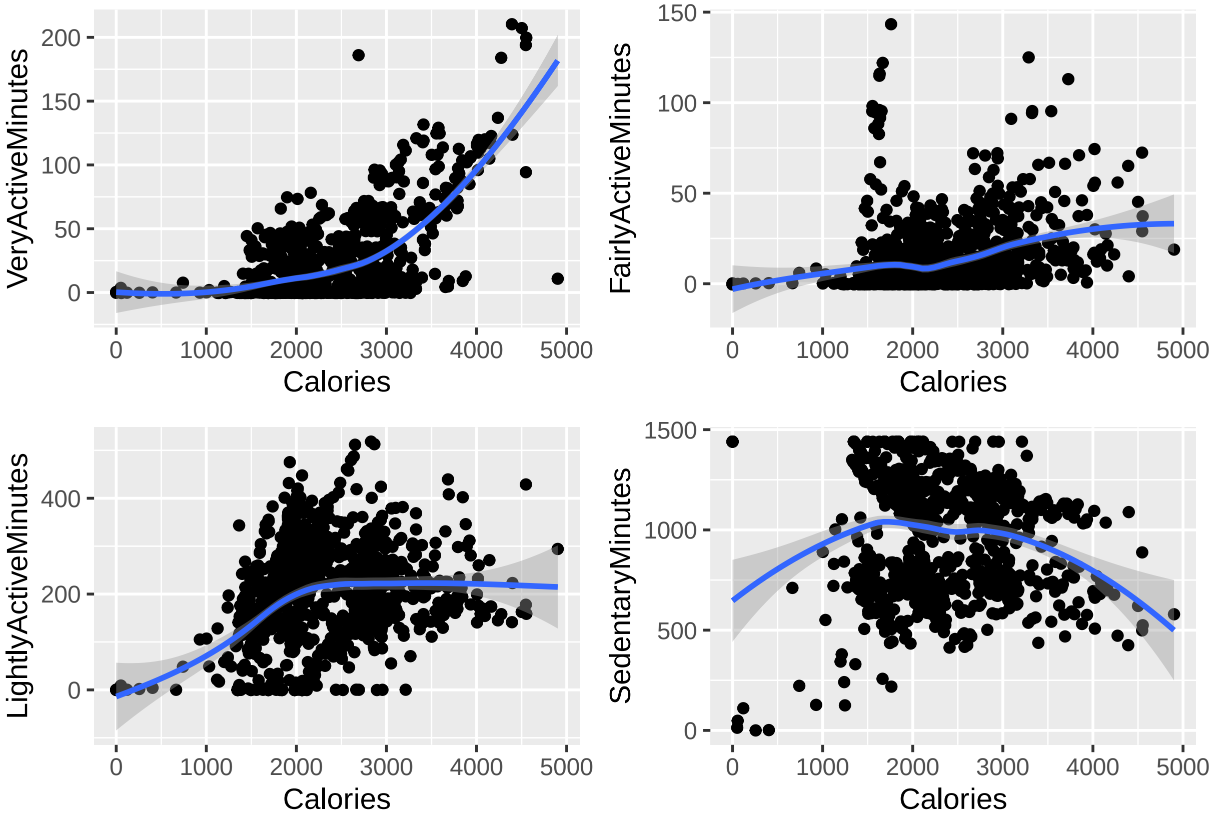
```
## install new package to show different graph side by side
install.packages("patchwork")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)
```

```
library(patchwork)

VeryActiveMinutes_graph <- ggplot(data=daily_activity) +
  geom_jitter(mapping = aes(x = Calories, y = VeryActiveMinutes))+
  geom_smooth(mapping = aes(x = Calories, y = VeryActiveMinutes))
```

```
FairlyActiveMinutes_graph <- ggplot(data=daily_activity) +
  geom_jitter(mapping = aes(x = Calories, y = FairlyActiveMinutes))+
  geom_smooth(mapping = aes(x = Calories, y = FairlyActiveMinutes))

LightlyActiveMinutes_graph <- ggplot(data=daily_activity) +
  geom_jitter(mapping = aes(x = Calories, y = LightlyActiveMinutes))+
  geom_smooth(mapping = aes(x = Calories, y = LightlyActiveMinutes))

SedentaryMinutes_graph <- ggplot(data=daily_activity) +
  geom_jitter(mapping = aes(x = Calories, y = SedentaryMinutes))+
  geom_smooth(mapping = aes(x = Calories, y = SedentaryMinutes))

VeryActiveMinutes_graph + FairlyActiveMinutes_graph + LightlyActiveMinutes_graph + SedentaryMinutes_grap
```

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

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



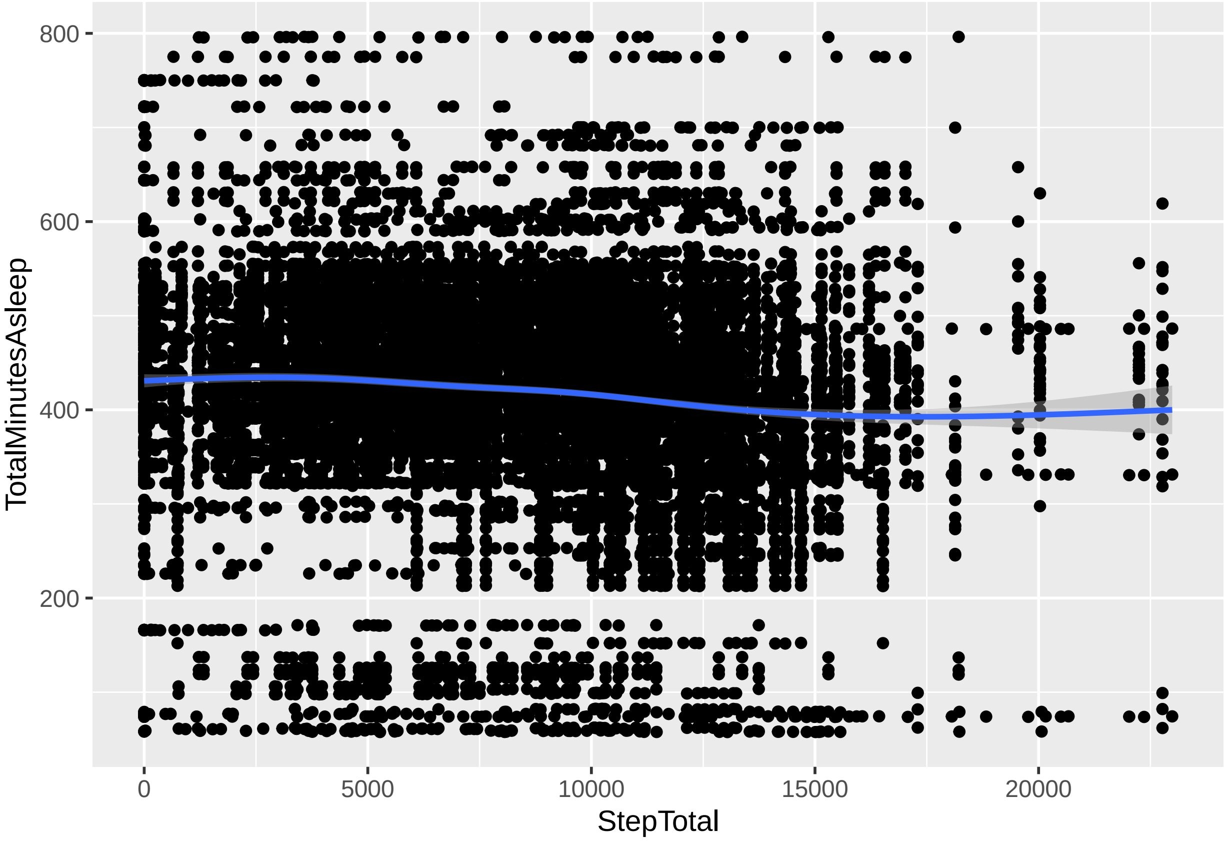
The Data outcome is as expected. If you have more active Minutes it is more likely that you burnt more Calories.

Steps vs Sleep

```
steps_vs_sleep <- daily_steps %>% inner_join(sleep_log,by="Id")
steps_vs_sleep_graph <- ggplot(data=steps_vs_sleep) +
  geom_jitter(mapping = aes(x = StepTotal, y = TotalMinutesAsleep))+
  geom_smooth(mapping = aes(x = StepTotal, y = TotalMinutesAsleep))
```

```
print(steps_vs_sleep_graph)
```

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

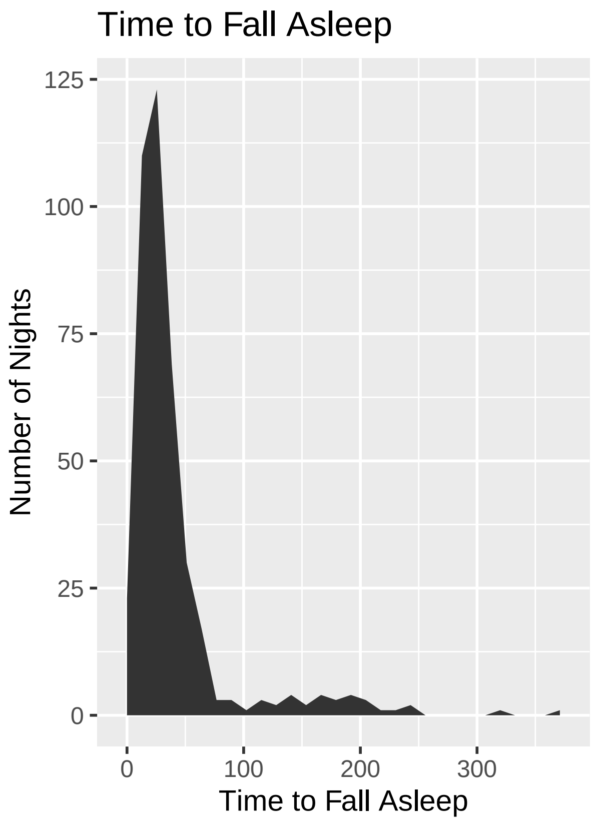


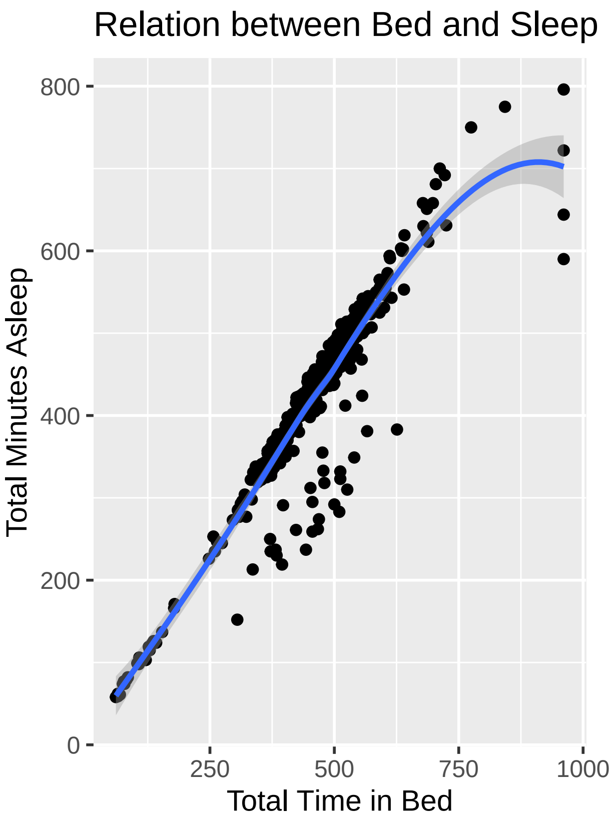
There seems to be no effect from the Steps someone took and the Minutes of Sleep he got.

Time taken to fall Asleep

```
updated_sleep_log <- sleep_log %>% mutate(time_taken_to_sleep = TotalTimeInBed-TotalMinutesAsleep)
Time_Falling_Asleep_Graph <- ggplot(data=updated_sleep_log, mapping = aes(x= time_taken_to_sleep))+
  geom_area(stat = "bin", bins= 30)+
  labs(x="Time to Fall Asleep", y= "Number of Nights", title = "Time to Fall Asleep")

inbed_vs_asleep_graph <- ggplot(data = updated_sleep_log, mapping = aes(x = TotalTimeInBed, y = TotalMi
  geom_point(mapping = aes(x = TotalTimeInBed, y = TotalMinutesAsleep))+
  geom_smooth(mapping = aes(x = TotalTimeInBed, y = TotalMinutesAsleep))+
  labs(x="Total Time in Bed", y= "Total Minutes Asleep", title = "Relation between Bed and Sleep")

Time_Falling_Asleep_Graph + inbed_vs_asleep_graph
```

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





Usually the Probands dont need more than ca. 75 Minutes to fall asleep.

If you stay longer in Bed you will also sleep longer. Besides you stay longer than 700 Minutes. Then the Amount of Sleep you will get will decline.

6. Act:

Summary of the relevant gained information:

- Not very suprising is that if someone takes more steps or has more active Minutes he burns more Calories.
- The Correlation between Very Active Minutes and Calories burnt seems to be very strong. If you have 50 Minutes of very Active Minutes its likely that you burn more than 3000 Calories
- The Steps someone took during the day doesnt have any Effect on his Sleep

Growth Opportunities:

- Help Customers to integrate a short amount of Time, where they are going to be high active, in there daylife. If they get more high active minutes into their daily life, they will burn more Calories.
- Encourage Customers to walk more during the day. More Steps -> More Calories burnt.