

Bellabeat Case Study in R

Miguel Sánchez Hernández

2023-08-11

1. Context

Urška Sršen and Sando Mur founded Bellabeat, a high-tech company that manufactures smart products focused on healthcare. Sršen used her experience as an artist to develop beautifully designed technology that will inform and inspire women around the world. Collecting data on physical activity, sleep, stress and reproductive health has enabled Bellabeat to provide women with insights into their own health and habits. Since its founding in 2013, Bellabeat grew at a breakneck pace and quickly positioned itself as a technology-driven wellness company for women.

It have five products:

- **Bellabeat App:** The Bellabeat app provides users with health data related to their physical activity, sleep, stress, menstrual cycle and mindfulness habits. This data can help users understand their current habits and make healthy choices. The Bellabeat app connects to its line of smart wellness products. ○ **Leaf:** Bellabeat's classic wellness tracking device that can be worn as a bracelet, necklace or clip. The Leaf device connects to the Bellabeat app to track physical activity, sleep and stress.
- **Time:** This wellness watch combines the timeless look of a classic watch with smart technology to track the wearer's physical activity, sleep and stress. The Time watch connects to the Bellabeat app to provide daily wellness information.
- **Spring:** a water bottle that tracks daily water consumption by using smart technology to ensure proper hydration throughout the day. The Spring bottle connects to the Bellabeat app to track hydration levels.
- **Bellabeat Membership:** Bellabeat also offers users a subscription-based membership program. Membership provides users with 24/7 access to fully personalized guidance on nutrition, physical activity, sleep, health and beauty, and mindfulness based on the user's lifestyle and goals.

2. Ask

THE BUSINESS TASK

- Find recommendations for marketing strategies
- Improve the Bellabeat app

- Aport usefull information

3. Prepare

We have 18 CSV files to analize from [Kaggle](#) so we are not going to analize them all.

3.1 Loading packages

```
library(tidyverse)

## — Attaching core tidyverse packages ————— tidyverse
2.0.0 —
## ✓ dplyr      1.1.2      ✓ readr      2.1.4
## ✓ forcats   1.0.0      ✓ stringr    1.5.0
## ✓ ggplot2    3.4.2      ✓ tibble     3.2.1
## ✓ lubridate 1.9.2      ✓ tidyr      1.3.0
## ✓ purrr     1.0.1
## — Conflicts —————
tidyverse_conflicts() —
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all
conflicts to become errors

library(skimr)
library(here)

## here() starts at C:/Users/20mig/Desktop/FITBIT/Fitabase Data 4.12.16-
5.12.16

library(janitor)

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

3.2 Change the directory

```
setwd("C:/Users/20mig/Desktop/FITBIT/Fitabase Data 4.12.16-5.12.16")
```

This is mine but yours can be different

3.3 Load the CSV files

```
dailyActivity<-read_csv("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.

weight<-read_csv("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.

hourlySteps_merged<-read_csv("hourlySteps_merged.csv")

## Rows: 22099 Columns: 3
## — Column specification

```

```

## Delimiter: ","
## chr (1): ActivityHour
## 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.

sleepDay_merged<-read_csv("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.

```

3.4 Preview

```
skim_without_charts(dailyActivity)
```

Data summary

Name dailyActivity
Number of rows 940
Number of columns 15

Column type frequency:

character 1
numeric 14

Group variables None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
ActivityDate	0	1	8	9	0	31	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Id	0	1	4.855407e+09	2.424805e+09	1503966	2.320127e+09	4.445115e+09	6.962181e+09	8.877689e+09
TotalSteps	0	1	7.637910e+03	5.087150e+03	0	3.789750e+03	7.405500e+03	1.072700e+04	3.601900e+04
TotalDistance	0	1	5.490000e+00	3.920000e+00	0	2.620000e+00	5.240000e+00	7.710000e+00	2.803000e+01
TrackerDistance	0	1	5.480000e+00	3.910000e+00	0	2.620000e+00	5.240000e+00	7.710000e+00	2.803000e+01
LoggedActivitiesDistance	0	1	1.100000e-01	6.200000e-01	0	0.000000e+00	0.000000e+00	0.000000e+00	4.940000e+00
VeryActiveDistance	0	1	1.500000e+00	2.660000e+00	0	0.000000e+00	2.100000e-01	2.050000e+00	2.192000e+01
ModeratelyActiveDistance	0	1	5.700000e-01	8.800000e-01	0	0.000000e+00	2.400000e-01	8.000000e-01	6.480000e+00

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
LightActiveDistance	0	1	3.3400 00e+0 0	2.0400 00e+0 0	0	1.9500 00e+0 0	3.3600 00e+0 0	4.7800 00e+0 0	1.0710 00e+0 1
SedentaryActiveDistance	0	1	0.0000 00e+0 0	1.0000 00e- 02	0	0.0000 00e+0 0	0.0000 00e+0 0	0.0000 00e+0 0	1.1000 00e- 01
VeryActiveMinutes	0	1	2.1160 00e+0 1	3.2840 00e+0 1	0	0.0000 00e+0 0	4.0000 00e+0 0	3.2000 00e+0 1	2.1000 00e+0 2
FairlyActiveMinutes	0	1	1.3560 00e+0 1	1.9990 00e+0 1	0	0.0000 00e+0 0	6.0000 00e+0 0	1.9000 00e+0 1	1.4300 00e+0 2
LightlyActiveMinutes	0	1	1.9281 00e+0 2	1.0917 00e+0 2	0	1.2700 00e+0 2	1.9900 00e+0 2	2.6400 00e+0 2	5.1800 00e+0 2
SedentaryMinutes	0	1	9.9121 00e+0 2	3.0127 00e+0 2	0	7.2975 00e+0 2	1.0575 00e+0 3	1.2295 00e+0 3	1.4400 00e+0 3
Calories	0	1	2.3036 10e+0 3	7.1817 00e+0 2	0	1.8285 00e+0 3	2.1340 00e+0 3	2.7932 50e+0 3	4.9000 00e+0 3

```
skim_without_charts(weight)
```

Data summary

Name	weight
Number of rows	67
Number of columns	8

Column type frequency:

character	1
logical	1
numeric	6

Group variables	None
-----------------	------

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
---------------	-----------	---------------	-----	-----	-------	----------	------------

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
Date	0	1	19	21	0	56	0

Variable type: logical

skim_variable	n_missing	complete_rate	mean	count
IsManualReport	0	1	0.61	TRU: 41, FAL: 26

Variable type: numeric

skim_v variable	n_mi ssin g	compl ete_rat e	mean	sd	p0	p25	p50	p75	p100
Id	0	1.00	7.0092 82e+0 9	1.9503 22e+0 9	1.5039 60e+0 9	6.9621 81e+0 9	6.9621 81e+0 9	8.8776 89e+0 9	8.8776 89e+0 9
Weight Kg	0	1.00	7.2040 00e+0 1	1.3920 00e+0 1	5.2600 00e+0 1	6.1400 00e+0 1	6.2500 00e+0 1	8.5050 00e+0 1	1.3350 00e+0 2
Weight Pounds	0	1.00	1.5881 00e+0 2	3.0700 00e+0 1	1.1596 00e+0 2	1.3536 00e+0 2	1.3779 00e+0 2	1.8750 00e+0 2	2.9432 00e+0 2
Fat	65	0.03	2.3500 00e+0 1	2.1200 00e+0 0	2.2000 00e+0 1	2.2750 00e+0 1	2.3500 00e+0 1	2.4250 00e+0 1	2.5000 00e+0 1
BMI	0	1.00	2.5190 00e+0 1	3.0700 00e+0 0	2.1450 00e+0 1	2.3960 00e+0 1	2.4390 00e+0 1	2.5560 00e+0 1	4.7540 00e+0 1
LogId	0	1.00	1.4617 72e+1 2	7.8299 48e+0 8	1.4604 44e+1 2	1.4610 79e+1 2	1.4618 02e+1 2	1.4623 75e+1 2	1.4630 98e+1 2

`skim_without_charts(hourlySteps_merged)`

Data summary

Name	hourlySteps_merged
Number of rows	22099
Number of columns	3

Column type frequency:

character	1
numeric	2

Group variables None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
ActivityHour	0	1	19	21	0	736	0

Variable type: numeric

skim_v ariable	n_mi ssing	comple te_rate	mean	sd	p0	p25	p50	p75	p100
Id	0	1	4.8482 35e+09	2.422 5e+09	15039 60366	23201 27002	44451 14986	69621 81067	88776 89391
StepTo tal	0	1	3.2017 00e+02	6.903 8e+02	0	0	40	357	10554

`skim_without_charts(sleepDay_merged)`

Data summary

Name sleepDay_merged
Number of rows 413
Number of columns 5

Column type frequency:

character 1
numeric 4

Group variables None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
SleepDay	0	1	20	21	0	31	0

Variable type: numeric

skim_vari able	n_mi ssing	comple te_rate	mean	sd	p0	p25	p50	p75	p100
Id	0	1	5.0009 79e+09	2.0603 6e+09	15039 60366	39773 33714	47029 21684	69621 81067	87920 09665
TotalSlee pRecords	0	1	1.1200 00e+00	3.5000 0e-01	1	1	1	1	3
TotalMinu tesAsleep	0	1	4.1947 00e+02	1.1834 0e+02	58	361	433	490	796

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
TotalTimeInBed	0	1	4.586400e+02	1.27100e+02	61	403	463	526	961

Observations:

- 65 missings in the column *Fat* of the weight table.
- We can clearly see the number of rows and columns of each data frame

4. Process

4.1 Clearing variables name

As the data set is a 10 in kaggle we are only going to check the variables name

```
clean_names(dailyActivity)

## # A tibble: 940 × 15
##       id activity_date total_steps total_distance tracker_distance
##       <dbl> <chr>          <dbl>          <dbl>          <dbl>
##  1 1503960366 4/12/2016          13162           8.5           8.5
##  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
##  7 1503960366 4/18/2016          13019           8.59          8.59
##  8 1503960366 4/19/2016          15506           9.88          9.88
##  9 1503960366 4/20/2016          10544           6.68          6.68
## 10 1503960366 4/21/2016           9819           6.34          6.34
## # i 930 more rows
## # i 10 more variables: logged_activities_distance <dbl>,
## #   very_active_distance <dbl>, moderately_active_distance <dbl>,
## #   light_active_distance <dbl>, sedentary_active_distance <dbl>,
## #   very_active_minutes <dbl>, fairly_active_minutes <dbl>,
## #   lightly_active_minutes <dbl>, sedentary_minutes <dbl>, calories <dbl>

clean_names(weight)

## # A tibble: 67 × 8
##       id date weight_kg weight_pounds fat bmi is_manual_report
##       <dbl> <chr>    <dbl>        <dbl> <dbl> <dbl> <lgl>
##  1 1503960366 5/2/...    52.6          116.    22  22.6 TRUE
##  2 1503960366 5/3/...    52.6          116.    NA  22.6 TRUE
```



```

1.46e12
## 3 1927972279 4/13... 134. 294. NA 47.5 FALSE
1.46e12
## 4 2873212765 4/21... 56.7 125. NA 21.5 TRUE
1.46e12
## 5 2873212765 5/12... 57.3 126. NA 21.7 TRUE
1.46e12
## 6 4319703577 4/17... 72.4 160. 25 27.5 TRUE
1.46e12
## 7 4319703577 5/4/... 72.3 159. NA 27.4 TRUE
1.46e12
## 8 4558609924 4/18... 69.7 154. NA 27.2 TRUE
1.46e12
## 9 4558609924 4/25... 70.3 155. NA 27.5 TRUE
1.46e12
## 10 4558609924 5/1/... 69.9 154. NA 27.3 TRUE
1.46e12
## # i 57 more rows

```

```
clean_names(hourlySteps_merged)
```

```

## # A tibble: 22,099 × 3
##       id activity_hour step_total
##       <dbl> <chr>         <dbl>
## 1 1503960366 4/12/2016 12:00:00 AM      373
## 2 1503960366 4/12/2016 1:00:00 AM      160
## 3 1503960366 4/12/2016 2:00:00 AM      151
## 4 1503960366 4/12/2016 3:00:00 AM         0
## 5 1503960366 4/12/2016 4:00:00 AM         0
## 6 1503960366 4/12/2016 5:00:00 AM         0
## 7 1503960366 4/12/2016 6:00:00 AM         0
## 8 1503960366 4/12/2016 7:00:00 AM         0
## 9 1503960366 4/12/2016 8:00:00 AM       250
## 10 1503960366 4/12/2016 9:00:00 AM     1864
## # i 22,089 more rows

```

```
clean_names(sleepDay_merged)
```

```

## # A tibble: 413 × 5
##       id sleep_day total_sleep_records total_minutes_asleep
total_time_in_bed
##       <dbl> <chr>         <dbl>         <dbl>
<dbl>
## 1 1.50e9 4/12/201...      1          327
346
## 2 1.50e9 4/13/201...      2          384
407
## 3 1.50e9 4/15/201...      1          412
442
## 4 1.50e9 4/16/201...      2          340
367

```

```
## 5 1.50e9 4/17/201... 1 700
712
## 6 1.50e9 4/19/201... 1 304
320
## 7 1.50e9 4/20/201... 1 360
377
## 8 1.50e9 4/21/201... 1 325
364
## 9 1.50e9 4/23/201... 1 361
384
## 10 1.50e9 4/24/201... 1 430
449
## # i 403 more rows
```

4.2 Converting date format

```
dailyActivity$ActivityDate<-as.Date(dailyActivity$ActivityDate,format =
"%m/%d/%Y")
hourlySteps_merged<-separate(hourlySteps_merged,ActivityHour,into =
c('Date','Hour'),sep=" ")

## Warning: Expected 2 pieces. Additional pieces discarded in 22099 rows [1,
2, 3, 4, 5, 6,
## 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].

hourlySteps_merged$Date<- as.Date(hourlySteps_merged$Date,format =
'%d','%m','%y')
```

4.3 Merging data frames

```
weight_Activity<-merge(weight,dailyActivity,by=c("Id"))
head(weight_Activity)
```

##	Id	Date	WeightKg	WeightPounds	Fat	BMI
## 1	1503960366	5/2/2016 11:59:59 PM	52.6	115.9631	22	22.65
## 2	1503960366	5/2/2016 11:59:59 PM	52.6	115.9631	22	22.65
## 3	1503960366	5/2/2016 11:59:59 PM	52.6	115.9631	22	22.65
## 4	1503960366	5/2/2016 11:59:59 PM	52.6	115.9631	22	22.65
## 5	1503960366	5/2/2016 11:59:59 PM	52.6	115.9631	22	22.65
## 6	1503960366	5/2/2016 11:59:59 PM	52.6	115.9631	22	22.65

```
## IsManualReport LogId ActivityDate TotalSteps TotalDistance
## 1 TRUE 1.462234e+12 2016-04-16 12669 8.16
## 2 TRUE 1.462234e+12 2016-04-18 13019 8.59
## 3 TRUE 1.462234e+12 2016-04-15 9762 6.28
## 4 TRUE 1.462234e+12 2016-05-08 10060 6.58
## 5 TRUE 1.462234e+12 2016-04-17 9705 6.48
## 6 TRUE 1.462234e+12 2016-04-19 15506 9.88
## TrackerDistance LoggedActivitiesDistance VeryActiveDistance
## 1 8.16 0 2.71
## 2 8.59 0 3.25
## 3 6.28 0 2.14
## 4 6.58 0 3.53
```

```
## 5          6.48          0          3.19
## 6          9.88          0          3.53
##   ModeratelyActiveDistance LightActiveDistance SedentaryActiveDistance
## 1          0.41          5.04          0
## 2          0.64          4.71          0
## 3          1.26          2.83          0
## 4          0.32          2.73          0
## 5          0.78          2.51          0
## 6          1.32          5.03          0
##   VeryActiveMinutes FairlyActiveMinutes LightlyActiveMinutes
SedentaryMinutes
## 1          36          10          221
773
## 2          42          16          233
1149
## 3          29          34          209
726
## 4          44           8          203
574
## 5          38          20          164
539
## 6          50          31          264
775
##   Calories
## 1      1863
## 2      1921
## 3      1745
## 4      1740
## 5      1728
## 6      2035
```

```
sleep_activity<-merge(sleepDay_merged,dailyActivity,by=c("Id"))
head(sleep_activity)
```

```
##           Id           SleepDay TotalSleepRecords TotalMinutesAsleep
## 1 1503960366 4/12/2016 12:00:00 AM                1                327
## 2 1503960366 4/12/2016 12:00:00 AM                1                327
## 3 1503960366 4/12/2016 12:00:00 AM                1                327
## 4 1503960366 4/12/2016 12:00:00 AM                1                327
## 5 1503960366 4/12/2016 12:00:00 AM                1                327
## 6 1503960366 4/12/2016 12:00:00 AM                1                327
##   TotalTimeInBed ActivityDate TotalSteps TotalDistance TrackerDistance
## 1          346   2016-05-07     11992          7.71          7.71
## 2          346   2016-05-06     12159          8.03          8.03
## 3          346   2016-05-01     10602          6.81          6.81
## 4          346   2016-04-30     14673          9.25          9.25
## 5          346   2016-04-12     13162          8.50          8.50
## 6          346   2016-04-13     10735          6.97          6.97
##   LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 1              0          2.46          2.12
```

```
## 2          0          1.97          0.25
## 3          0          2.29          1.60
## 4          0          3.56          1.42
## 5          0          1.88          0.55
## 6          0          1.57          0.69
##   LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
## 1          3.13          0          37
## 2          5.81          0          24
## 3          2.92          0          33
## 4          4.27          0          52
## 5          6.06          0          25
## 6          4.71          0          21
##   FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories
## 1          46          175          833          1821
## 2           6          289          754          1896
## 3          35          246          730          1820
## 4          34          217          712          1947
## 5          13          328          728          1985
## 6          19          217          776          1797
```

5. Analyze

Daily Activity

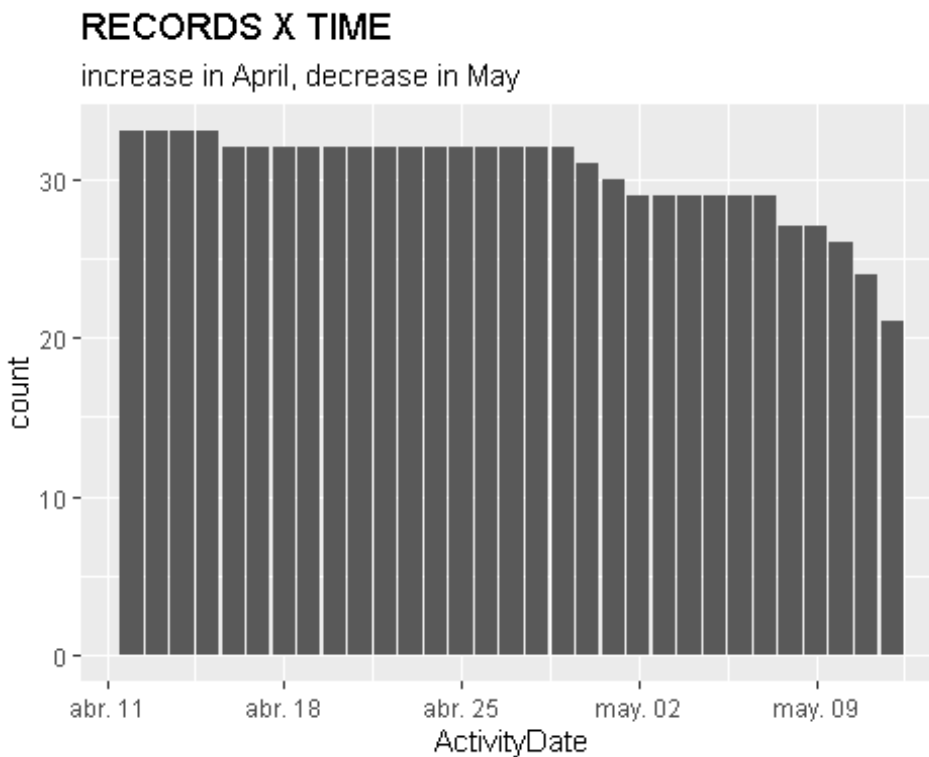
```
dailyActivity%>%
  group_by(Id)%>%
  count(Id)

## # A tibble: 33 × 2
## # Groups:   Id [33]
##       Id     n
##   <dbl> <int>
## 1 1503960366    31
## 2 1624580081    31
## 3 1644430081    30
## 4 1844505072    31
## 5 1927972279    31
## 6 2022484408    31
## 7 2026352035    31
## 8 2320127002    31
## 9 2347167796    18
## 10 2873212765    31
## # i 23 more rows
```

We observe that there are 33 unique Ids in this data and ones have more information than others

```
ggplot(data=dailyActivity)+
  geom_bar(mapping = aes(x=ActivityDate))+
```

```
labs(title="RECORDS X TIME", subtitle = "increase in April, decrease in May")
```

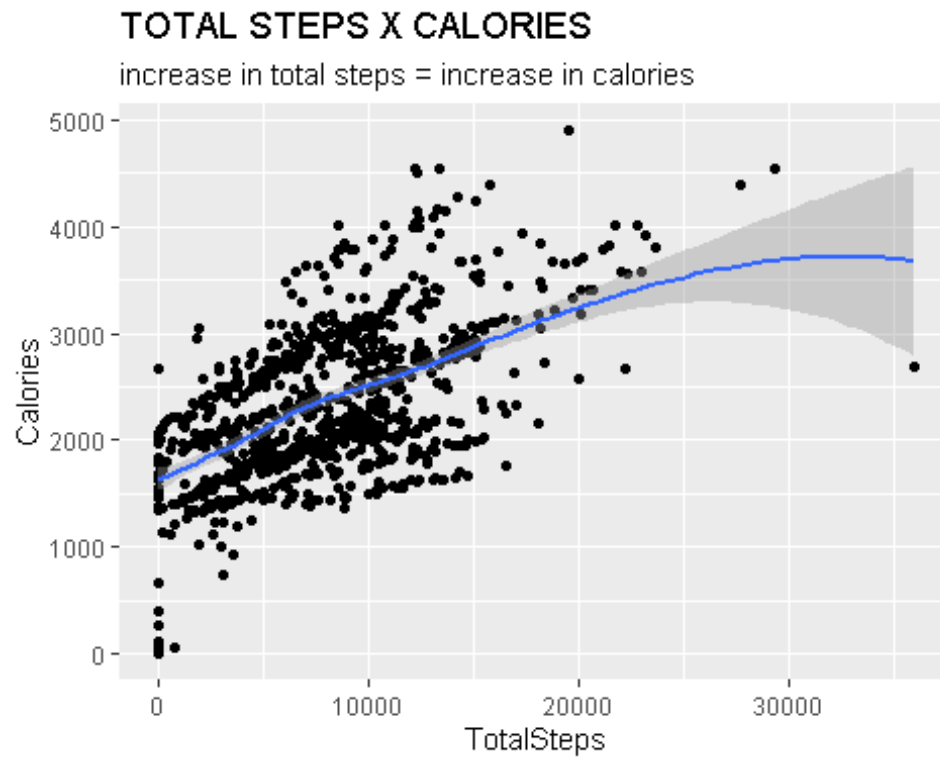


Observations:

- Increase of records in April and decrease of them in May
- Users are more active at the start

```
ggplot(data=dailyActivity)+
  geom_point(mapping = aes(x=TotalSteps,y=Calories),color='black')+
  geom_smooth(mapping = aes(x=TotalSteps,y=Calories),method = 'loess')+
  labs(title="TOTAL STEPS X CALORIES", subtitle = "increase in total steps =
increase in calories")
```

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



Observations:

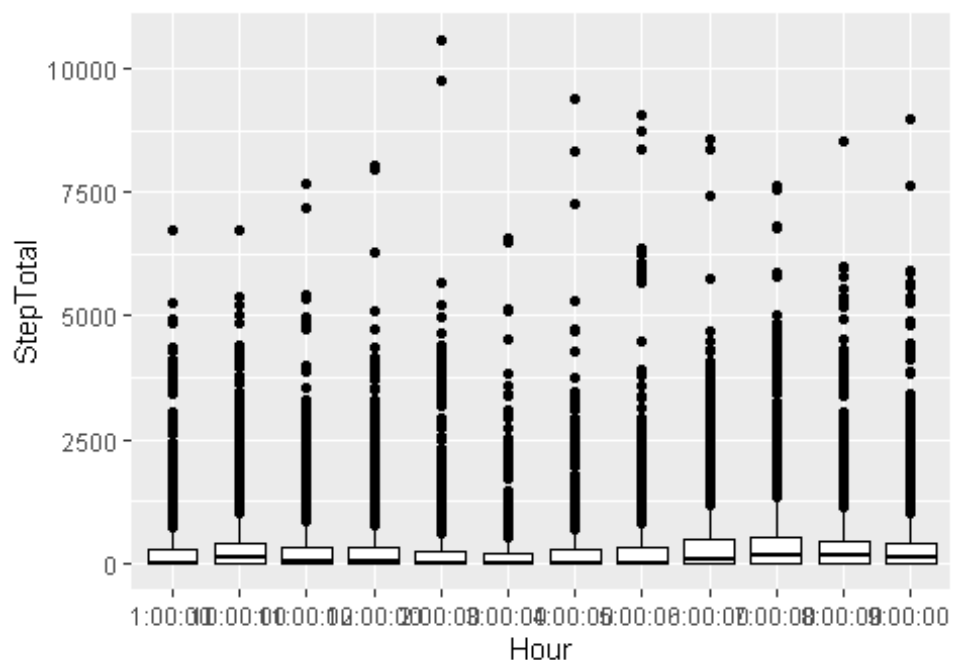
- Correlation between total steps and calories

Hourly Steps

```
ggplot(data=hourlySteps_merged)+
  geom_boxplot(mapping = aes(x=Hour,y=StepTotal),color='black')+
  geom_smooth(mapping = aes(x=Hour,y=StepTotal))+
  labs(title="HOUR X STEPTOTAL",subtitle = "")

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

HOUR X STEPTOTAL



```
hourlySteps_merged%>%
  group_by(Hour)%>%

summarise(media=mean(StepTotal),maximun=max(StepTotal),minimun=min(StepTotal)
)
```

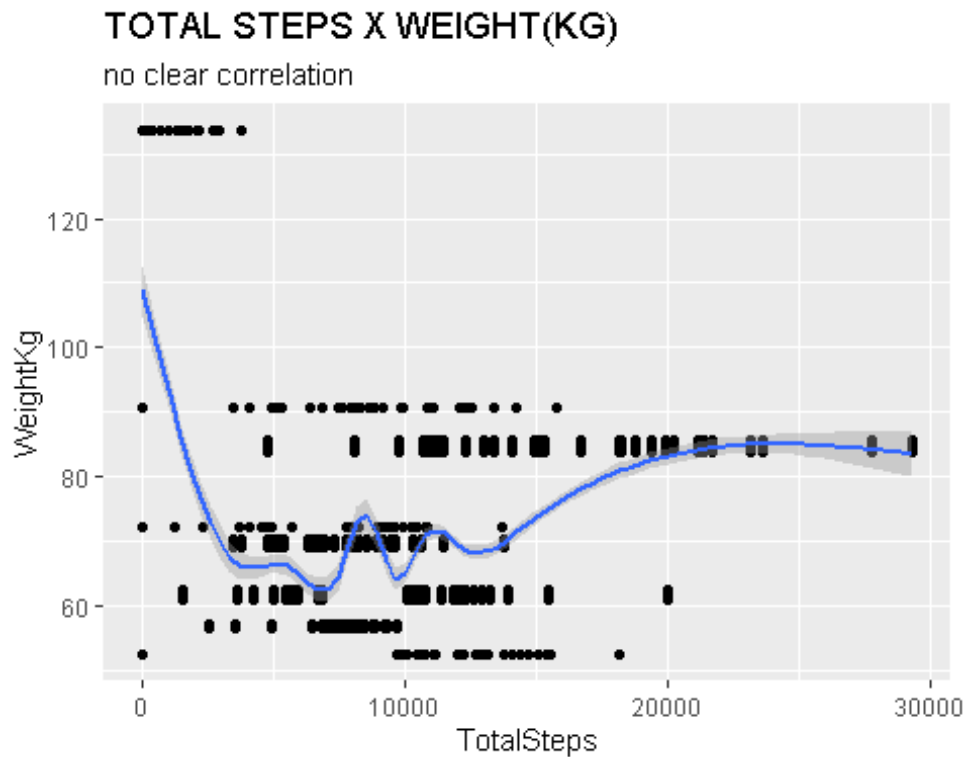
```
## # A tibble: 12 × 4
##   Hour      media maximun minimun
##   <chr>    <dbl>    <dbl>    <dbl>
## 1 10:00:00   361.     6715         0
## 2 11:00:00   292.     7652         0
## 3 12:00:00   294.     8043         0
## 4 1:00:00    279.     6745         0
## 5 2:00:00   277.    10554         0
## 6 3:00:00   204.     6554         0
## 7 4:00:00   251.     9392         0
## 8 5:00:00   293.     9062         0
## 9 6:00:00   386.     8586         0
## 10 7:00:00  443.     7643         0
## 11 8:00:00  391.     8520         0
## 12 9:00:00  372.     8976         0
```

Observations:

- The higher media of steps is at 7:00:00
- The maximun higher is at 2:00:00

Weight and Daily Activity

```
ggplot(data=weight_Activity)+  
  geom_point(mapping = aes(x=TotalSteps,y=WeightKg))+  
  geom_smooth(mapping = aes(x=TotalSteps,y=WeightKg))+  
  labs(title="TOTAL STEPS X WEIGHT(KG)",subtitle="no clear correlation")  
  
## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```



Observations:

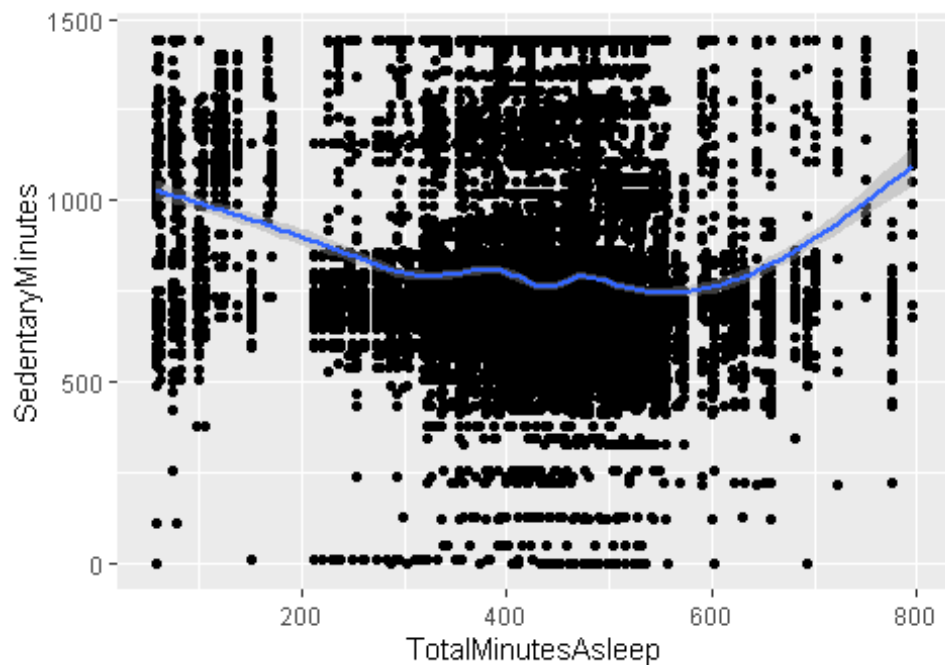
- The people with high weight use to don't walk

Sleep and Daily Activity

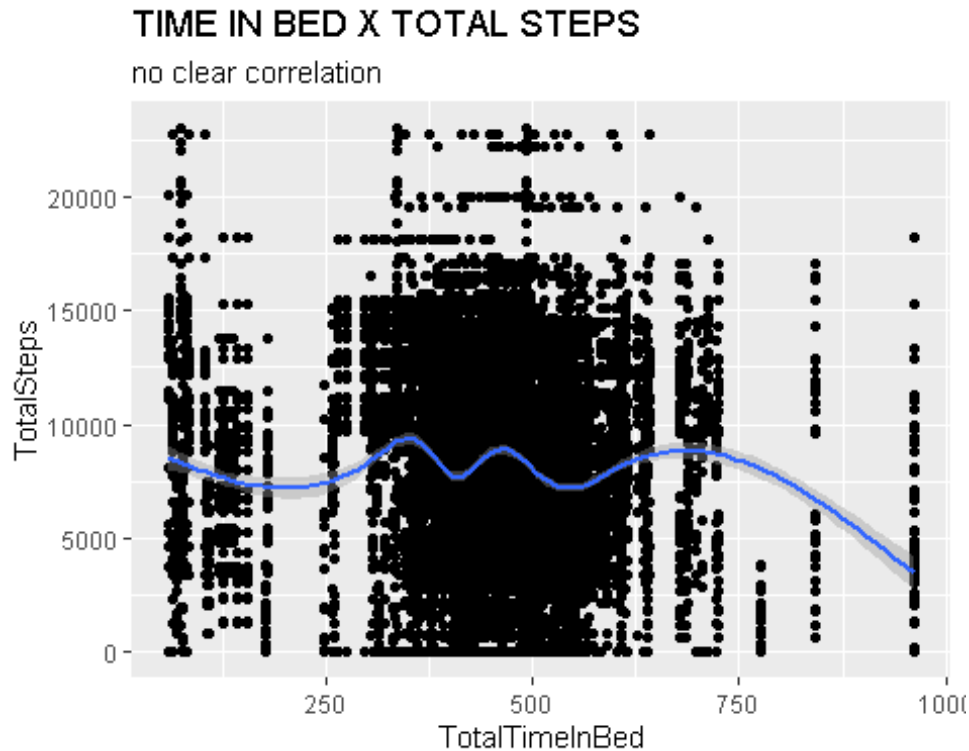
```
ggplot(data=sleep_activity)+  
  geom_point(mapping = aes(x=TotalMinutesAsleep,y=SedentaryMinutes))+  
  geom_smooth(mapping = aes(x=TotalMinutesAsleep,y=SedentaryMinutes))+  
  labs(title = "MINUTES ASLEEP X SEDENTARY TIME",subtitle = "no clear  
correlation")  
  
## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```


MINUTES ASLEEP X SEDENTARY TIME

no clear correlation



```
ggplot(data=sleep_activity)+  
  geom_point(mapping = aes(x=TotalTimeInBed,y=TotalSteps))+  
  geom_smooth(mapping = aes(x=TotalTimeInBed,y=TotalSteps))+  
  labs(title = "TIME IN BED X TOTAL STEPS",subtitle = "no clear correlation")  
## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```



Observations:

- We don't see a clear correlation between minutes asleep and sedentary time
- We don't see a clear correlation between time in bed and total steps

6. Act

Based on the findings, Recommendations are as below:

- We need to update the app once a month to get users active because of the decrease in May
- We could create rewards based on the steps taken and put a bonus at 5:00:00 that lasts 30 minutes so that they start walking at that time and gradually increase the walking time.
- We need to focus on people above 100 kg because they are the ones that least walk.
- We also need to incentive our users to walk at least 10,000 steps per day so they can burn a good amount of calories

Extras

This is my first case study in r and I am very excited if anyone finds any errors please let me know.

[Linkedin](#)

[GitHub](#)