my project

2022-07-21

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

INTRODUCTION Bellabeat, a high-tech manufacturer of health-focused products for women is a successful small company which has the potential to become a larger player in the

global smart device market. Urška Sršen, cofounder and Chief Creative Officer of Bellabeat, believes that analyzing smart device fitness data could help unlock new growth opportunities for the company. Bellabeat has 5 focus products: Bellabeat app, leaf, time, spring and Bellabeat membership. This analysis is meant to focus on one of Bellabeat´s products and analyze smart device data to gain insight into how consumers are using their smart devices. Insights from this analysis would guide into providing marketing strategies for Bellabeat´s products. The following steps will be used in the analysis: Ask, Prepare, Process, Analyse, Share, Act

ASK:

Primary stakeholders: Urška Sršen and Sando Mur, executive team members.

Secondary stakeholders: Bellabeat marketing team.

● What is the problem you are trying to solve? FOCUS: To have an insight into women´s daily habit via analysis of data from other smart devices (FitBit Fitness Tracker Data).

● How can your insights drive business decisions? FOCUS: My insight on this project should bring a result that will make Bellabeat improve on their products especially the product that would make women particularly more interested in buying Bellabeat smart devices. One can also use this insight to discover better marketing strategies for Bellabeat products.

PREPARE

• Where is your data stored? The data is publicly available on Kaggle FitBit Fitness Tracker Data.

• How is the data organized? Is it in long or wide format? The fitness tracker data provided includes 18 csv format files arranged in long formats.

Are there issues with bias or credibility in this data? The data set was uploaded by Möbius on Kaggle set and was reported to have been generated by thirty respondents in a distributed survey via Amazon Mechanical Turk between 03.12.2016-05.12.2016. It doesn't state how the selection of participants or distribution were made thereby making it impossible to question the credibility or biasness of the data collation. It is important that data is Reliable, Original, Comprehensive, Current, and Cited (ROCCC).

• Reliable=LOW — Not reliable as 30 participants appears too low for such survey. • Original = MEDIUM — The data was collected via Third party provider (Amazon Mechanical Turk) • Comprehensive =MEDIUM — Parameters match most of Bellabeat product´s parameters • Current =LOW — The data is six years old and may not be relevant today. • Cited =LOW — Data collected from third party • OVERALL: It is difficult to give a good rating to the data collection based on the ROCC model and therefore cannot recommend the use of the dataset for business recommendations.

• How are you addressing licensing, privacy, security, and accessibility? It is written that 30 eligible Fitbit users (participants) consented to the survey. The data is publicly available on Kaggle.

• How did you verify the data's integrity? I could not verify the data´s integrity as it was a survey made by a third-party company. • How does it help you answer your question? The data from this survey should be able to provide information on how much people use in order to determine products and marketing strategies for Bellabeat.

• Are there any problems with the data? Yes, one major problem with the data is that it is not current and may not satisfy current recommendations. I would not be able to totally ignore the fact that the data could be original and reliable despite being provided by a third-party company. Newer technologies might have been improved while participants habits might have changed

PROCESS

load required libraries

```
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   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.1     v forcats 0.5.1
```

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

```
library(readr)
library(knitr)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

loading required data

```
daily_activity <- 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.
```

```
sleeping_time <- 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.
```

```
#daily_calories <- read_csv("dailyCalories_merged.csv")
minute_sleep <- read_csv("minuteSleep_merged.csv")
```

```
## Rows: 188521 Columns: 4
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): date
## dbl (3): Id, value, logId
```

```
##
## 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.
```

Exploring the data checking data dimension

```
dim(daily_activity)
```

```
## [1] 940  15
```

```
dim(sleeping_time)
```

```
## [1] 413   5
```

```
#dim(daily_calories)
dim(minute_sleep)
```

```
## [1] 188521      4
```

summarize the data

```
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
```

```
#summary(daily_calories)
summary(sleeping_time)
```

```
##        Id              SleepDay         TotalSleepRecords TotalMinutesAsleep
##  Min.   :1.504e+09   Length:413         Min.   :1.000     Min.   : 58.0
##  1st Qu.:3.977e+09   Class :character   1st Qu.:1.000     1st Qu.:361.0
##  Median :4.703e+09   Mode  :character   Median :1.000     Median :433.0
##  Mean   :5.001e+09                      Mean   :1.119     Mean   :419.5
##  3rd Qu.:6.962e+09                      3rd Qu.:1.000     3rd Qu.:490.0
##  Max.   :8.792e+09                      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
```

```
summary(minute_sleep)
```

```
##        Id                date               value           logId
##  Min.   :1.504e+09   Length:188521      Min.   :1.000   Min.   :1.137e+10
##  1st Qu.:3.977e+09   Class :character   1st Qu.:1.000   1st Qu.:1.144e+10
##  Median :4.703e+09   Mode  :character   Median :1.000   Median :1.150e+10
##  Mean   :4.997e+09                      Mean   :1.096   Mean   :1.150e+10
##  3rd Qu.:6.962e+09                      3rd Qu.:1.000   3rd Qu.:1.155e+10
##  Max.   :8.792e+09                      Max.   :3.000   Max.   :1.162e+10
```

Listing the variables in each data

```
str(daily_activity)
```

```
## spec_tbl_df [940 x 15] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id                      : num [1:940] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ ActivityDate            : chr [1:940] "4/12/2016" "4/13/2016" "4/14/2016" "4/15/2016" ...
##  $ TotalSteps              : num [1:940] 13162 10735 10460 9762 12669 ...
##  $ TotalDistance           : num [1:940] 8.5 6.97 6.74 6.28 8.16 ...
##  $ TrackerDistance         : num [1:940] 8.5 6.97 6.74 6.28 8.16 ...
##  $ LoggedActivitiesDistance: num [1:940] 0 0 0 0 0 0 0 0 0 0 ...
##  $ VeryActiveDistance      : num [1:940] 1.88 1.57 2.44 2.14 2.71 ...
##  $ ModeratelyActiveDistance: num [1:940] 0.55 0.69 0.4 1.26 0.41 ...
##  $ LightActiveDistance     : num [1:940] 6.06 4.71 3.91 2.83 5.04 ...
##  $ SedentaryActiveDistance : num [1:940] 0 0 0 0 0 0 0 0 0 0 ...
##  $ VeryActiveMinutes       : num [1:940] 25 21 30 29 36 38 42 50 28 19 ...
##  $ FairlyActiveMinutes     : num [1:940] 13 19 11 34 10 20 16 31 12 8 ...
##  $ LightlyActiveMinutes    : num [1:940] 328 217 181 209 221 164 233 264 205 211 ...
##  $ SedentaryMinutes        : num [1:940] 728 776 1218 726 773 ...
##  $ Calories                : num [1:940] 1985 1797 1776 1745 1863 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   ActivityDate = col_character(),
##   ..   TotalSteps = col_double(),
##   ..   TotalDistance = col_double(),
##   ..   TrackerDistance = col_double(),
##   ..   LoggedActivitiesDistance = col_double(),
##   ..   VeryActiveDistance = col_double(),
##   ..   ModeratelyActiveDistance = col_double(),
##   ..   LightActiveDistance = col_double(),
##   ..   SedentaryActiveDistance = col_double(),
##   ..   VeryActiveMinutes = col_double(),
##   ..   FairlyActiveMinutes = col_double(),
##   ..   LightlyActiveMinutes = col_double(),
##   ..   SedentaryMinutes = col_double(),
##   ..   Calories = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
#str(daily_calories)
str(sleeping_time)
```

```
## spec_tbl_df [413 x 5] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id                : num [1:413] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ SleepDay          : chr [1:413] "4/12/2016 12:00:00 AM" "4/13/2016 12:00:00 AM" "4/15/2016 12:00:00 AM" "4/
16/2016 12:00:00 AM" ...
##  $ TotalSleepRecords : num [1:413] 1 2 1 2 1 1 1 1 1 1 ...
##  $ TotalMinutesAsleep: num [1:413] 327 384 412 340 700 304 360 325 361 430 ...
##  $ TotalTimeInBed    : num [1:413] 346 407 442 367 712 320 377 364 384 449 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   SleepDay = col_character(),
##   ..   TotalSleepRecords = col_double(),
##   ..   TotalMinutesAsleep = col_double(),
##   ..   TotalTimeInBed = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(minute_sleep)
```

```
## spec_tbl_df [188,521 x 4] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id   : num [1:188521] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ date : chr [1:188521] "4/12/2016 2:47:30 AM" "4/12/2016 2:48:30 AM" "4/12/2016 2:49:30 AM" "4/12/2016 2:50:
30 AM" ...
##  $ value: num [1:188521] 3 2 1 1 1 1 1 2 2 2 ...
##  $ logId: num [1:188521] 1.14e+10 1.14e+10 1.14e+10 1.14e+10 1.14e+10 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   date = col_character(),
##   ..   value = col_double(),
##   ..   logId = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

ANALYSE

Checking the number of unique id

```
n_distinct(daily_activity$Id)
```

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

```
#n_distinct(daily_calories$Id)
n_distinct(sleeping_time$Id)
```

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

checking for missing values

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

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

```
sum(is.na(sleeping_time))
```

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

```
sum(is.na(minute_sleep))
```

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

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

It appears that there are 940 missing variables in the activity table.Dates appears as “NA” Solution:(This was fix by reuploading the dataframe dailyActivity_merged.csv) again.

Renaming the column “ActivityDate” in daily_activity dataframe to “Date”

```
daily_activity <- daily_activity %>%
  rename(Date = ActivityDate)
```

From the summary, it seems the date column is being recognised wrongly as a character format. changing the character format to “date format”

```
 daily_activity[['Date']] <-as.POSIXct(daily_activity[['Date']],format = "%m/%d/%Y")
```

checking for duplicate records

```
sum(duplicated(daily_activity))
```

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

```
#sum(duplicated(daily_calories))
sum(duplicated(sleeping_time))
```

```
## [1] 3
```

```
sum(duplicated(minute_sleep))
```

```
## [1] 543
```

Having 52 duplicates in the daily_activity data sets. removing duplicates.

```
daily_activity <- daily_activity %>%
 distinct()
```

```
daily_activity <- daily_activity %>%
  distinct()
```

Removing duplicates from sleeping_time

```
sleeping_time <- sleeping_time %>%
  distinct()
```

Confirm duplicate removal

```
sum(duplicated(daily_activity))
```

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

```
sum(duplicated(sleeping_time))
```

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

Changing the sleepDay datatype (character datatype) in the sleeping_time datasets to date format and rename as “date”.

```
sleeping_time <- sleeping_time %>%
  rename(Date = SleepDay)
```

```
 sleeping_time[['Date']] <-as.POSIXct(sleeping_time[['Date']],format = "%m/%d/%Y")
```

Warning: Problem while computing `Date = as_date(Date, format = "%m/%d/%Y %I:%M:%S %p", tz = Sys.timezone())`. ℹ `tz` argument is ignored by `as_date()`

Recheck the sleeping_time dataset

```
head(sleeping_time)
```

```
## # A tibble: 6 x 5
##          Id Date                TotalSleepRecords TotalMinutesAsl~ TotalTimeInBed
##       <dbl> <dttm>                          <dbl>            <dbl>          <dbl>
## 1   1.50e9 2016-04-12 00:00:00                 1              327            346
## 2   1.50e9 2016-04-13 00:00:00                 2              384            407
## 3   1.50e9 2016-04-15 00:00:00                 1              412            442
## 4   1.50e9 2016-04-16 00:00:00                 2              340            367
## 5   1.50e9 2016-04-17 00:00:00                 1              700            712
## 6   1.50e9 2016-04-19 00:00:00                 1              304            320
```

```
minute_sleep[['date']] <- as.POSIXct(minute_sleep[['date']],
                                   format = "%m/%d/%Y")
```

Recheck the minute_sleep dataset

```
head(minute_sleep)
```

```
## # A tibble: 6 x 4
##           Id date                value       logId
##        <dbl> <dttm>              <dbl>       <dbl>
## 1 1503960366 2016-04-12 00:00:00     3 11380564589
## 2 1503960366 2016-04-12 00:00:00     2 11380564589
## 3 1503960366 2016-04-12 00:00:00     1 11380564589
## 4 1503960366 2016-04-12 00:00:00     1 11380564589
## 5 1503960366 2016-04-12 00:00:00     1 11380564589
## 6 1503960366 2016-04-12 00:00:00     1 11380564589
```

commment: I renamed column ActivityDate in daily_activity to “Date”

```
daily_activity <- daily_activity %>%
  rename(Date = Date)
```

Recheck the daily_activity dataset

```
head(daily_activity)
```

```
## # A tibble: 6 x 15
##           Id Date                TotalSteps TotalDistance TrackerDistance
##        <dbl> <dttm>                   <dbl>         <dbl>           <dbl>
## 1 1503960366 2016-04-12 00:00:00      13162          8.5             8.5
## 2 1503960366 2016-04-13 00:00:00      10735          6.97            6.97
## 3 1503960366 2016-04-14 00:00:00      10460          6.74            6.74
## 4 1503960366 2016-04-15 00:00:00       9762          6.28            6.28
## 5 1503960366 2016-04-16 00:00:00      12669          8.16            8.16
## 6 1503960366 2016-04-17 00:00:00       9705          6.48            6.48
## # ... with 10 more variables: LoggedActivitiesDistance <dbl>,
## #   VeryActiveDistance <dbl>, ModeratelyActiveDistance <dbl>,
## #   LightActiveDistance <dbl>, SedentaryActiveDistance <dbl>,
## #   VeryActiveMinutes <dbl>, FairlyActiveMinutes <dbl>,
## #   LightlyActiveMinutes <dbl>, SedentaryMinutes <dbl>, Calories <dbl>
```

I lost my dates

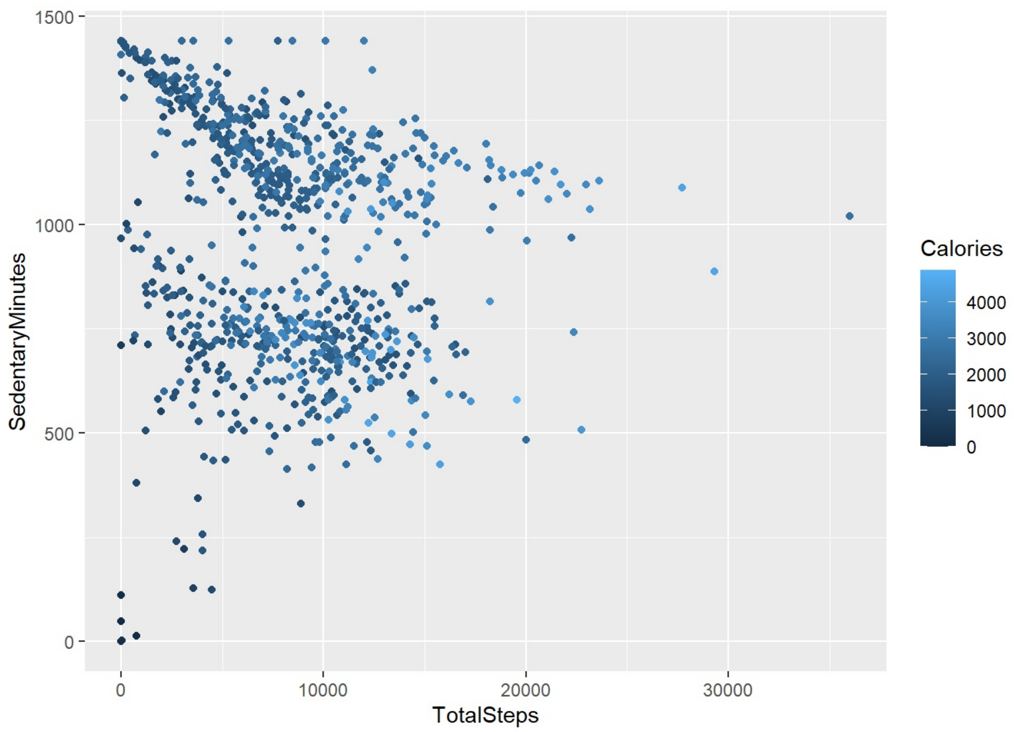
Cleaning minute_sleep data sets: Change datatype of date and bring out as date..

```
minute_sleep <- minute_sleep %>%
  rename(Date = date)
```

VISUALISATION

Activity: Analysing the daily_activity data set using ggplot.Is there any relationships between the calories burnt per day and the sedentary or daily steps

```
ggplot(data=daily_activity, aes(x=TotalSteps, y=SedentaryMinutes, color = Calories)) + geom_point()
```

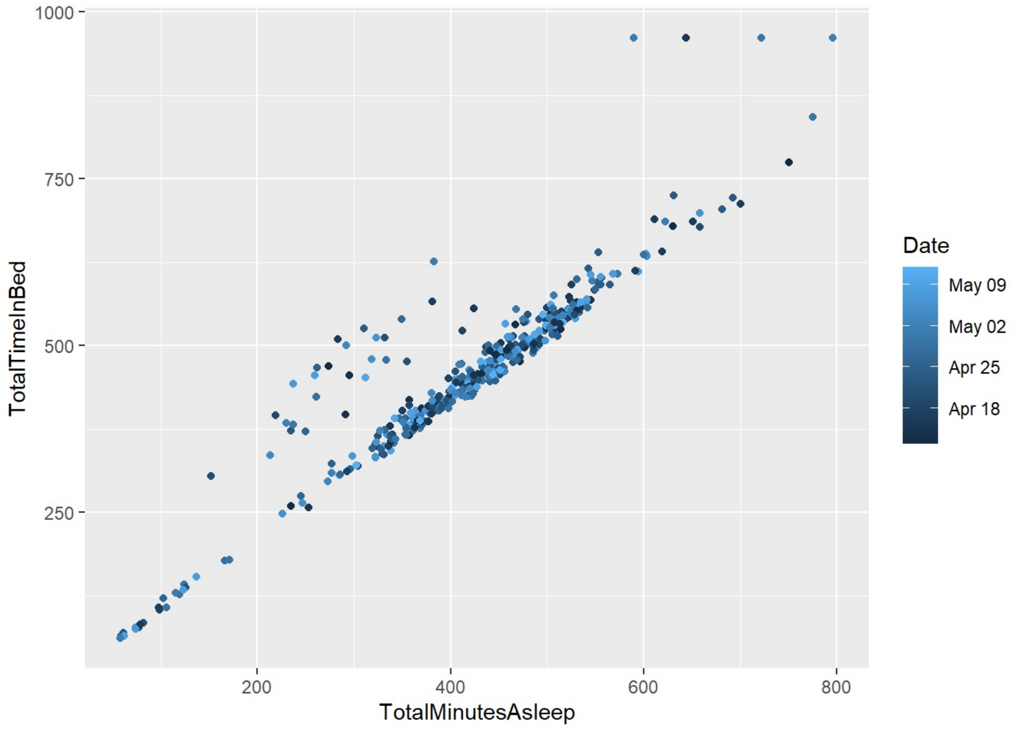


From the above ggplot: The calories expended has not really reflected a relationship with the sedentary minutes.

The sedentary minutes tends to reflet a negative relationship with the steps taking per day.

visualising the sleeping time data set using ggplot

```
ggplot(data=sleeping_time, aes(x=TotalMinutesAsleep, y=TotalTimeInBed)) + geom_point(aes(color=Date))
```



In general, it is easy to say from the plot that there is a relationship between time spent in bed and the total amount of minutes in bed. Quite a large number of participants had total sleeping minutes between 300- 600 minutes(5-10 hours) with the highest sleeping minutes around 800 minutes. Very few participants spent more time in bed with corrensponding more total sleeping minutes. Sleep appears not to be a major problem of these sets of women as the recommended sleeping hours 8 hours a day according to WHO.

Which Bellabeat product can we market using the above plot:

We could market the Bella smartwatch to consumers to better manage their sleeping habits. These habits could even give an overview of days of the week the consumers sleep best as well as give information about a potential sleeping problem among the consumers.

To check how actively participants use Fitbits device on weekly basis:

Add a column for day of the week in daily_activity dataframe:

```
daily_activity <- daily_activity %>% mutate(Weekday = weekdays(as.Date(Date, "%m/%d/%Y")))
```

```
## Warning in as.POSIXlt.POSIXct(x, tz = tz): unknown timezone '%m/%d/%Y'
```

```
head(daily_activity)
```

```
## # A tibble: 6 x 16
##           Id Date                TotalSteps TotalDistance TrackerDistance
##        <dbl> <dttm>                   <dbl>         <dbl>           <dbl>
## 1 1503960366 2016-04-12 00:00:00      13162          8.5             8.5
## 2 1503960366 2016-04-13 00:00:00      10735          6.97            6.97
## 3 1503960366 2016-04-14 00:00:00      10460          6.74            6.74
## 4 1503960366 2016-04-15 00:00:00       9762          6.28            6.28
## 5 1503960366 2016-04-16 00:00:00      12669          8.16            8.16
## 6 1503960366 2016-04-17 00:00:00       9705          6.48            6.48
## # ... with 11 more variables: LoggedActivitiesDistance <dbl>,
## #   VeryActiveDistance <dbl>, ModeratelyActiveDistance <dbl>,
## #   LightActiveDistance <dbl>, SedentaryActiveDistance <dbl>,
## #   VeryActiveMinutes <dbl>, FairlyActiveMinutes <dbl>,
## #   LightlyActiveMinutes <dbl>, SedentaryMinutes <dbl>, Calories <dbl>,
## #   Weekday <chr>
```

Add a column for day of the week in sleeping_time dataframe:

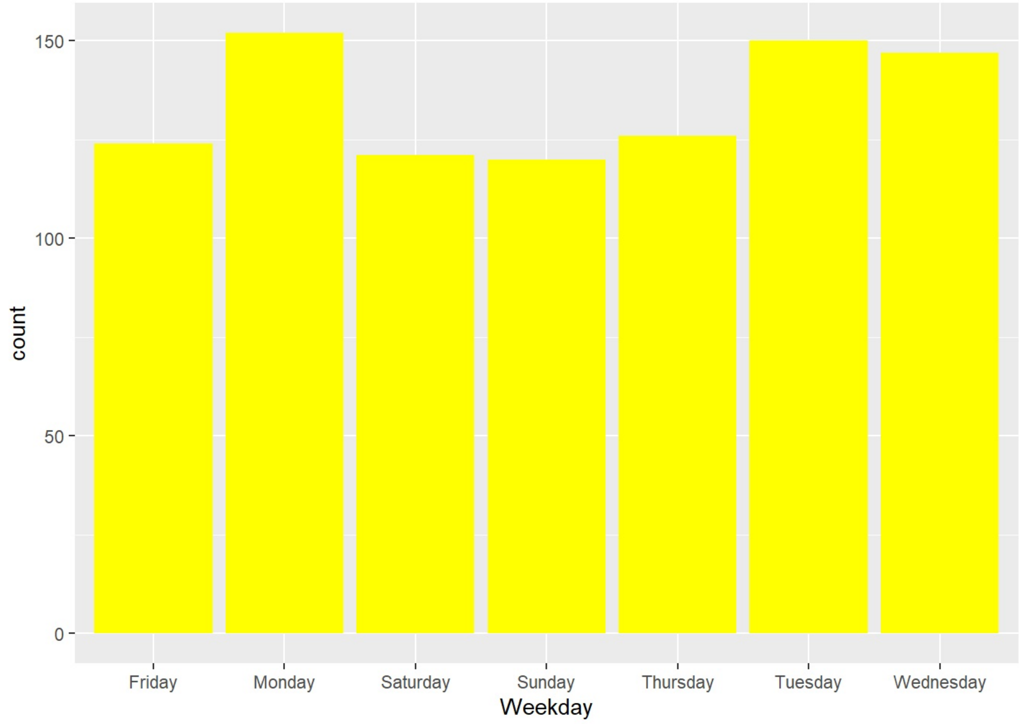
```
sleeping_time<- sleeping_time %>%
mutate(Weekday = weekdays(as.Date(Date, "%m/%d/%Y")))
```

```
## Warning in as.POSIXlt.POSIXct(x, tz = tz): unknown timezone '%m/%d/%Y'
```

Creating a barchart visualization to view how participants use record their data

daily_activity:

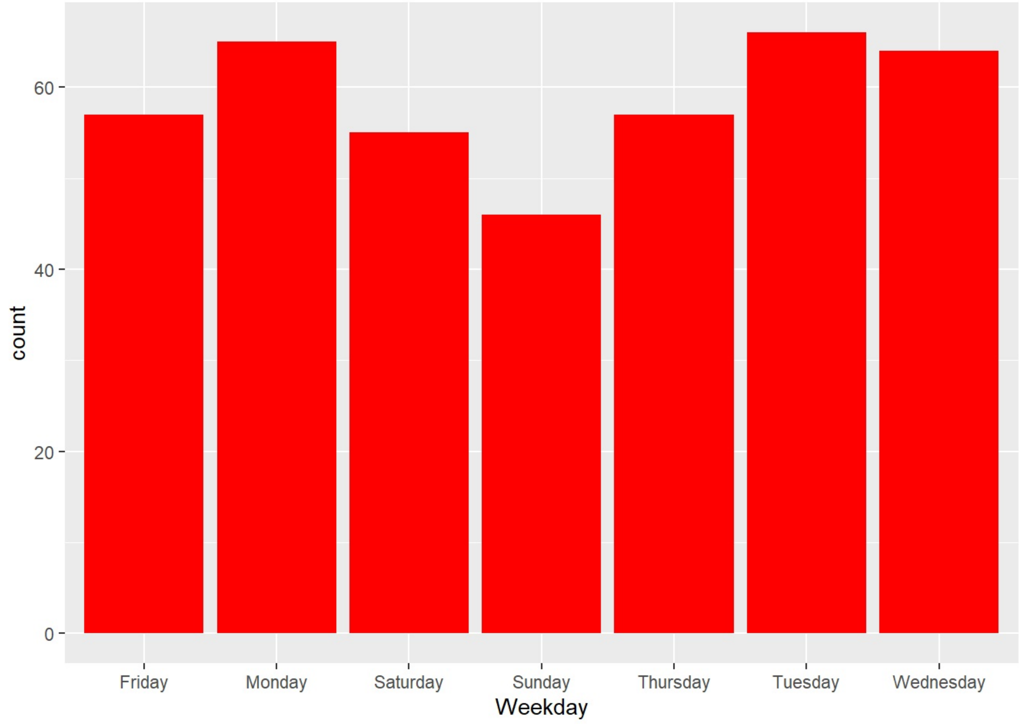
```
ggplot(data=daily_activity, aes(x=Weekday))+geom_bar(fill="yellow")
```



This barchart shows that participants record their daily activities mostly on Tuesdays and Wednesday and Thurdays in a unifrom manner. Although records from other days apprears to be lower but also relatively uniform.

Barchart sleeping_time:

```
ggplot(data=sleeping_time, aes(x=Weekday))+geom_bar(fill="red")
```



Similar to daily_activity records, sleeping_time records were more on Tuesdays,Wednesday and Thurdays while other days had relatively lower records but non-uniform. It could be good to figure out a reason for these differences and may be improve on what might have been causes hinderances in the record of these activities.

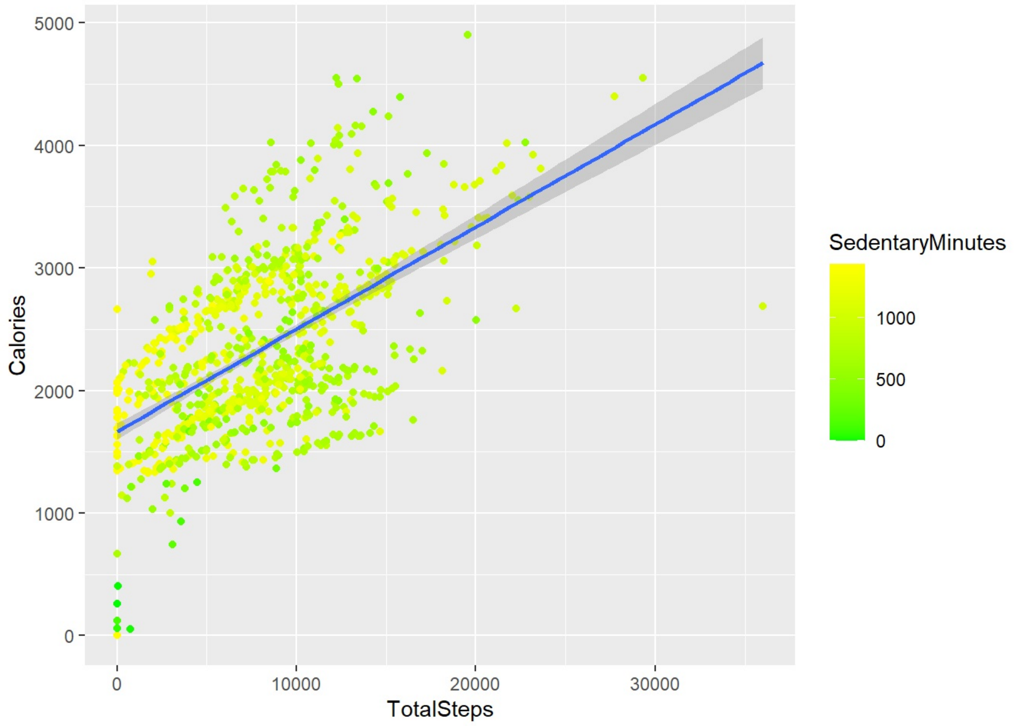
let´s try to merge these tables together to get an insight into common features in the datasets

```
activitysleep <- merge(daily_activity, sleeping_time, by = c("Id","Date"))
```

```
combined_data <- merge(activitysleep, minute_sleep, by = c("Id","Date"))
```

```
ggplot(data=daily_activity, aes(x=TotalSteps, y = Calories, color=SedentaryMinutes))+
  geom_point()+
  stat_smooth(method=lm)+
  scale_color_gradient(low="green", high="yellow")
```

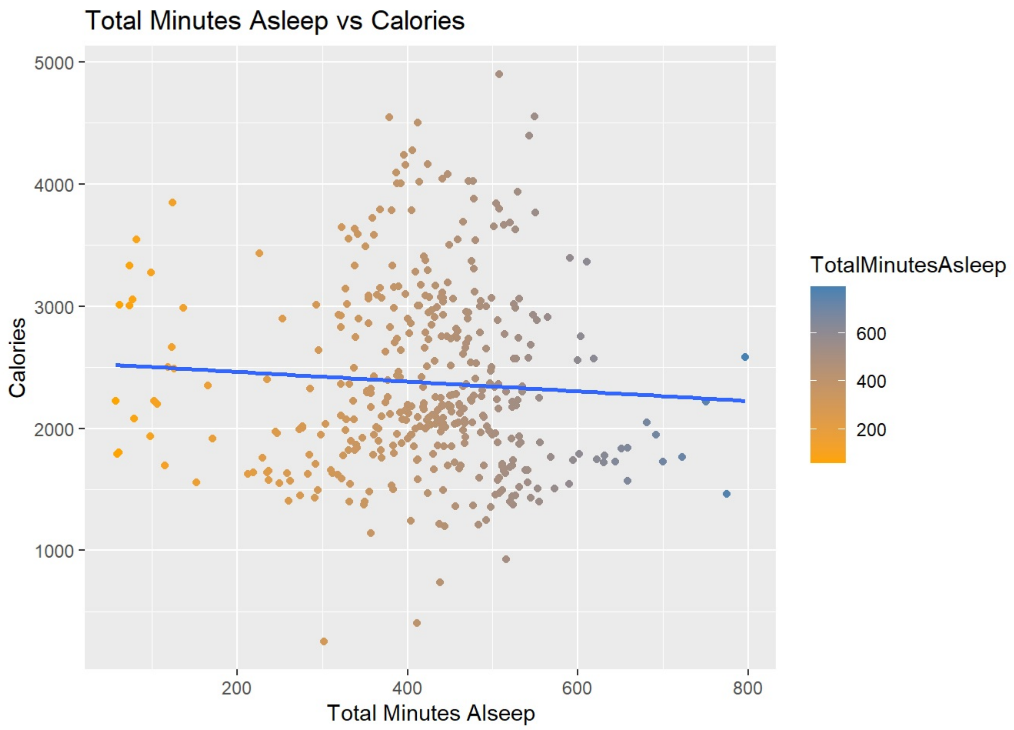
```
## `geom_smooth()` using formula 'y ~ x'
```



The above plot shows a positive relationship between calories burnt and steps taking with most steps less than 15,000 step and most calories burnt between 1500 to 2500 calories. One can not really figure out how participants with high sedentary activity still manage to burn so much same calories as those taking longer steps.

```
ggplot(data=combined_data, aes(x=TotalMinutesAsleep, y = Calories, color=TotalMinutesAsleep))+
  geom_point()+
  labs(title="Total Minutes Asleep vs Calories")+
  xlab("Total Minutes Alseep")+
  stat_smooth(method=lm)+
  scale_color_gradient(low="orange", high="steelblue")
```

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




```
awake_in_bed <- mutate(sleeping_time, Awake_time = TotalTimeInBed - TotalMinutesAsleep)
```

```
awake_in_bed <- awake_in_bed %>%
  filter(Awake_time >= 55) %>%
  group_by(Id) %>%
  arrange(Awake_time)
```

About 13 women spend more than 55 minutes in bed before they fall asleep.

ACT

Sleep: About 54% of participants who gave their sleeping records spend 55 minutes or more in bed before falling asleep

Activity: Sedentary minutes take up 81% of participants daily activity meaning that much of this reading were taking while they were idle.
The calories expended has not really reflected a relationship with the sedentary minutes. However, average mean calories per day is 2304 calories which is relatively high enough when compared to medically recommended 2000 calories per day.

• Users spend on average 12 hours a day in sedentary minutes and 4 hours lightly active.

Recommendations for Bellabeat marketing team: It would be good to have more recent data for this analysis. Since majority of the participants are not meeting up with the recommended time spent on daily activities, the marketing crew could suggest promo for participants/Bellabeat customers who meet up with the medically recommended daily body activity requirement. Bellabeat product could also be designed to set up reminders for the users in case they tend to forget to record or use their smart devices.