Bellabeat_Case_Study

Ian Lim

7/2/2021

Introduction & Goals

This case study is part of the Google Data Analytics Certification Capstone project.For this case study, we are tasked with providing insights regarding how consumers are using their smart devices by analyzing competitor's data (FitBit) to improve Bellabeat's marketing strategy.

Main Questions to Answer & Goals from this Analysis:

- How are customers of FitBit mainly using their fitness trackers?
- What features are most commonly used or in demand?
- How does this compare to what Bellabeat's product has to offer?
- Suggestion to Bellabeats on their product.

Dataset

The dataset for this analysis can be found from https://www.kaggle.com/arashnic/fitbit

Installing Packages

For this analysis, we will be using R packages such as tidyverse, ggplot2, and sqldf.

```
library(tidyverse)

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

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.2     v dplyr   1.0.6
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1

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

library(ggplot2)
library(sqldf)

## Loading required package: gsubfn
## Loading required package: proto

## Loading required package: RSQLite
```

Reading CSV files and Creating Dataframes

```
setwd("C:/Users/Ian Lim/Documents/R/FitBit-Data")
daily_activity <- read.csv("dailyActivity_merged.csv")
daily_calories <- read.csv("dailyCalories_merged.csv")
sleep_day <- read.csv("sleepDay_merged.csv")
daily_intensities <- read.csv("dailyIntensities_merged.csv")
weight_log <- read.csv("weightLogInfo_merged.csv")
```

Exploring each Data Frame

Daily Activity

```
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

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"
```

```
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~
```

Daily Calories

```
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
```

```
colnames(daily_calories)

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

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~
```

Sleep Day

```
head(sleep_day)

##           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

colnames(sleep_day)

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

glimpse(sleep_day)

## 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~
```

Daily Intensities

```
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
```

```
colnames(daily_intensities)
```

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

```
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~
```

Weight Log

```
head(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(weight_log)
```

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

```
glimpse(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,~
```

Quick Analysis

All 5 data frames have a common ‘ID’ field which can be used to merge/join the data frames. We can see that daily_activity, daily_calories, and daily_intensities have the same number of rows/observations. By closer observations, we can also see that the details in daily_calories and daily_intensities seems to already exist in daily_activity. To confirm this, we need to check if the values match for the respective IDs.

Let’s use SQL syntax to check. To do this, we need to create a temporary data frame as it would not work if the number of columns in two data frames are different.

Checking for daily_calories

daily_calories has 3 columns, so we need to create a temporary data frame with 3 columns.

```
daily_activity2 <- daily_activity %>%
  select(Id, ActivityDate, Calories)

head(daily_activity2)

##           Id ActivityDate 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
```

Seems good, now let’s see the intersect of daily_activity2 with daily_calories and check the number of rows of the intersect.

```
sql_check1 <- sqldf('SELECT * FROM daily_activity2 INTERSECT SELECT * FROM
daily_calories')

head(sql_check1)

##           Id ActivityDate 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

nrow(sql_check1)

## [1] 940
```

Checking for daily_intensities

daily_intensities has 10 columns, so we need to create a temporary data frame with 10 columns.

```
daily_activity3 <- daily_activity %>%
  select(Id, ActivityDate, SedentaryMinutes, LightlyActiveMinutes,
FairlyActiveMinutes, VeryActiveMinutes, SedentaryActiveDistance,
LightActiveDistance, ModeratelyActiveDistance, VeryActiveDistance)

head(daily_activity3)

##           Id ActivityDate 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
```

Seems good, now let’s repeat the same intersect check as per above.

```
sql_check2 <- sqldf('SELECT * FROM daily_activity3 INTERSECT SELECT * FROM
daily_intensities')

head(sql_check2)

##           Id ActivityDate 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
```

```
nrow(sql_check2)
```

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

As we can see, both checks result in the number of rows still being 940. In this case, we can say that the values are the same for all 3 data frames. This also means that we can exclude both the daily_calories and daily_intensities table from our analysis as the information they contain can be found in daily_activity.

Further Analysis

Continuing from above, we are now left with 3 data frames:

- daily_activity
- sleep_day
- weight_log

Let’s find out how many distinct IDs are there in all 3.

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

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

```
n_distinct(sleep_day$Id)
```

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

```
n_distinct(weight_log$Id)
```

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

Let’s look at the summary for all 3 as well.

Daily Activity

```
daily_activity %>%
  select(TotalSteps,
         TotalDistance,
         SedentaryMinutes,
         VeryActiveMinutes) %>%
  summary()
```

```
##    TotalSteps    TotalDistance    SedentaryMinutes VeryActiveMinutes
##  Min.   :    0   Min.   : 0.000   Min.   :   0.0   Min.   :  0.00  
##  1st Qu.: 3790   1st Qu.: 2.620   1st Qu.: 729.8   1st Qu.:  0.00  
##  Median : 7406   Median : 5.245   Median :1057.5   Median :  4.00  
##  Mean   : 7638   Mean   : 5.490   Mean   : 991.2   Mean   : 21.16  
##  3rd Qu.:10727   3rd Qu.: 7.713   3rd Qu.:1229.5   3rd Qu.: 32.00  
##  Max.   :36019   Max.   :28.030   Max.   :1440.0   Max.   :210.00  
```

Sleep Day

```
sleep_day %>%
  select(TotalSleepRecords,
  TotalMinutesAsleep,
  TotalTimeInBed) %>%
  summary()
```

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

Weight Log

```
weight_log %>%
  select(WeightKg,
  BMI) %>%
  summary()
```

```
##     WeightKg           BMI       
##  Min.   : 52.60   Min.   :21.45  
##  1st Qu.: 61.40   1st Qu.:23.96  
##  Median : 62.50   Median :24.39  
##  Mean   : 72.04   Mean   :25.19  
##  3rd Qu.: 85.05   3rd Qu.:25.56  
##  Max.   :133.50   Max.   :47.54  
```

Merging the data frame together

Let’s first combine sleep_day and daily_activity. I choose to include all values as indicated by using “all=TRUE”. To check for this, we use n_distinct to make sure the value will be 33.

```
combined_data <- merge(sleep_day, daily_activity, by='Id', all=TRUE)
n_distinct(combined_data$Id)
```

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

Then, we add weight_log to the newly combined data frame, repeating the steps above.

```
combined_data_all <- merge(combined_data, weight_log, by='Id', all=TRUE)
n_distinct(combined_data_all$Id)
```

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

```
head(combined_data_all)
```

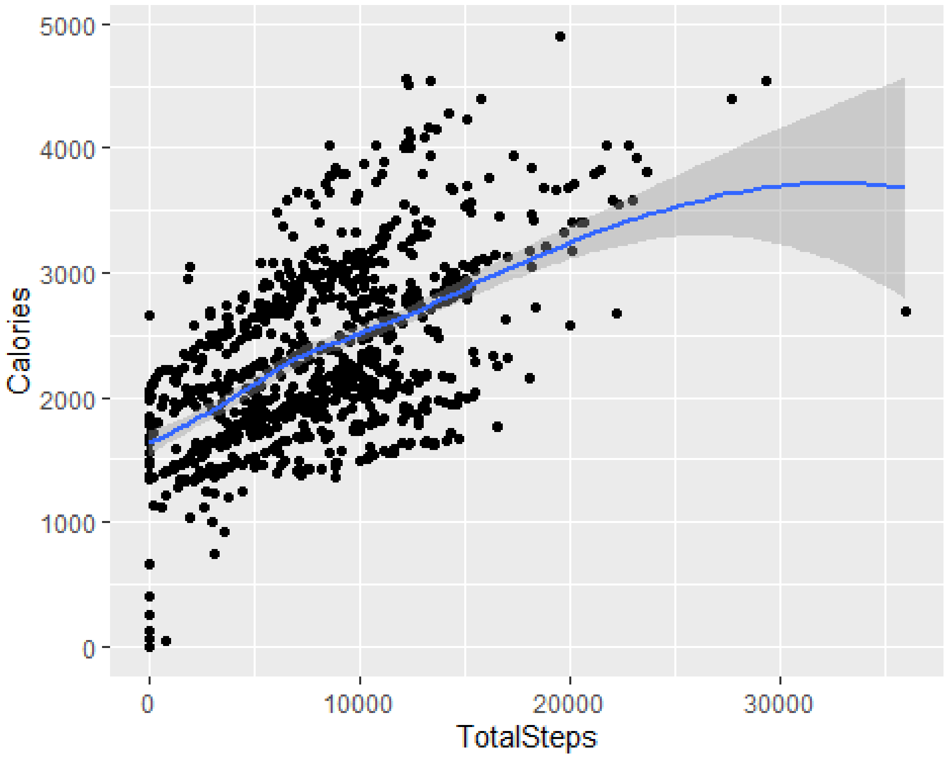
```
##           Id             SleepDay TotalSleepRecords TotalMinutesAsleep
## 1 1503960366 5/6/2016 12:00:00 AM                 1                334
## 2 1503960366 5/6/2016 12:00:00 AM                 1                334
## 3 1503960366 5/7/2016 12:00:00 AM                 1                331
## 4 1503960366 5/7/2016 12:00:00 AM                 1                331
## 5 1503960366 5/6/2016 12:00:00 AM                 1                334
## 6 1503960366 5/6/2016 12:00:00 AM                 1                334
##   TotalTimeInBed ActivityDate TotalSteps TotalDistance TrackerDistance
## 1            367    4/14/2016      10460          6.74            6.74
## 2            367    4/14/2016      10460          6.74            6.74
## 3            349     5/1/2016      10602          6.81            6.81
## 4            349     5/1/2016      10602          6.81            6.81
## 5            367    4/22/2016      12764          8.13            8.13
## 6            367    4/22/2016      12764          8.13            8.13
##   LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 1                        0               2.44                     0.40
## 2                        0               2.44                     0.40
## 3                        0               2.29                     1.60
## 4                        0               2.29                     1.60
## 5                        0               4.76                     1.12
## 6                        0               4.76                     1.12
##   LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
## 1                3.91                       0                30
## 2                3.91                       0                30
## 3                2.92                       0                33
## 4                2.92                       0                33
## 5                2.24                       0                66
## 6                2.24                       0                66
##   FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories
## 1                  11                  181             1218     1776
## 2                  11                  181             1218     1776
## 3                  35                  246              730     1820
## 4                  35                  246              730     1820
## 5                  27                  130             1217     1827
## 6                  27                  130             1217     1827
##                   Date WeightKg WeightPounds Fat   BMI IsManualReport
## 1 5/2/2016 11:59:59 PM     52.6     115.9631  22 22.65           True
## 2 5/3/2016 11:59:59 PM     52.6     115.9631  NA 22.65           True
## 3 5/2/2016 11:59:59 PM     52.6     115.9631  22 22.65           True
## 4 5/3/2016 11:59:59 PM     52.6     115.9631  NA 22.65           True
## 5 5/2/2016 11:59:59 PM     52.6     115.9631  22 22.65           True
## 6 5/3/2016 11:59:59 PM     52.6     115.9631  NA 22.65           True
##          LogId
## 1 1.462234e+12
## 2 1.462320e+12
## 3 1.462234e+12
## 4 1.462320e+12
## 5 1.462234e+12
## 6 1.462320e+12
```

Looks good, let’s begin plotting some graphs.

Plotting some Graphs

Relationship between Total Steps and Calories

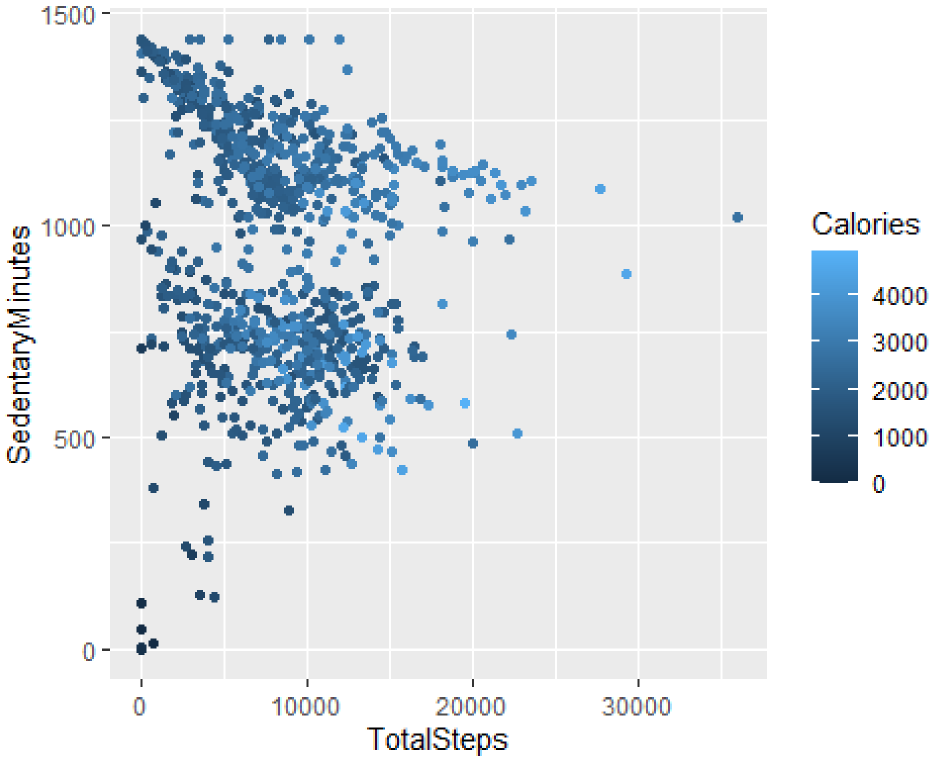
```
ggplot(data = daily_activity, aes(x = TotalSteps, y = Calories))+
geom_point() + geom_smooth()

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



We can see a rather clear indication from the graph above that people who took more steps tends to burn more calories. Nothing out of the expectation here.

Relationship between Total Steps and Sedentary Minutes

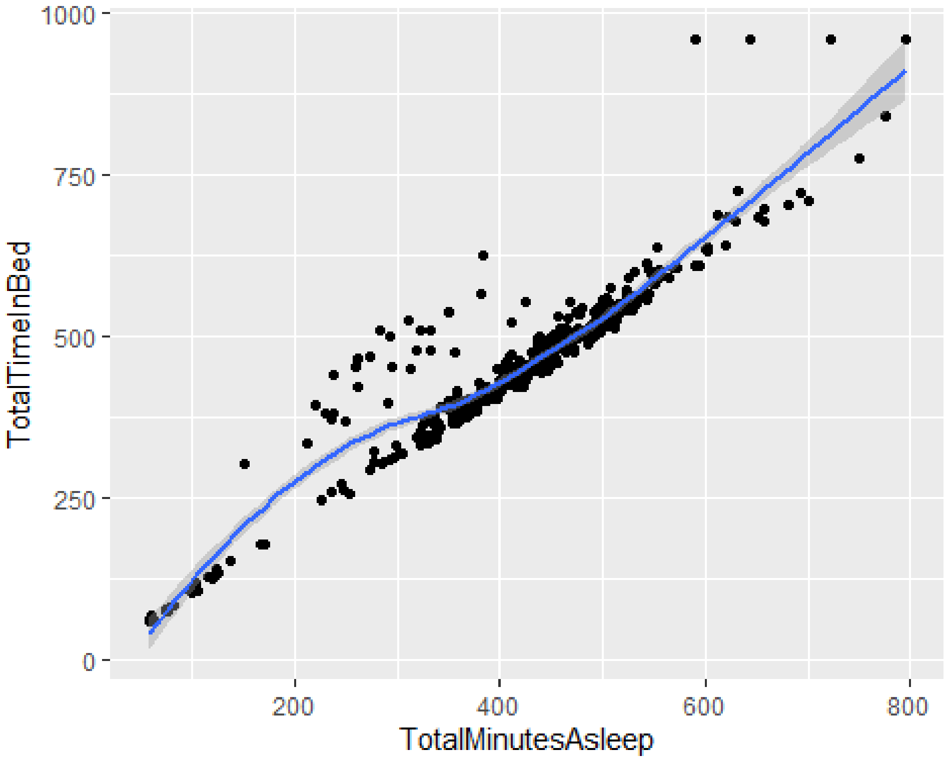
```
ggplot(data = daily_activity, aes(x = TotalSteps, y = SedentaryMinutes, color
= Calories)) + geom_point()
```



There is a negative relationship between the total number of steps and sedentary minutes which makes sense. We can also see that as the number of steps increase the total number of calories burned also trend upwards.

Relationship between Minutes Asleep and Time in Bed

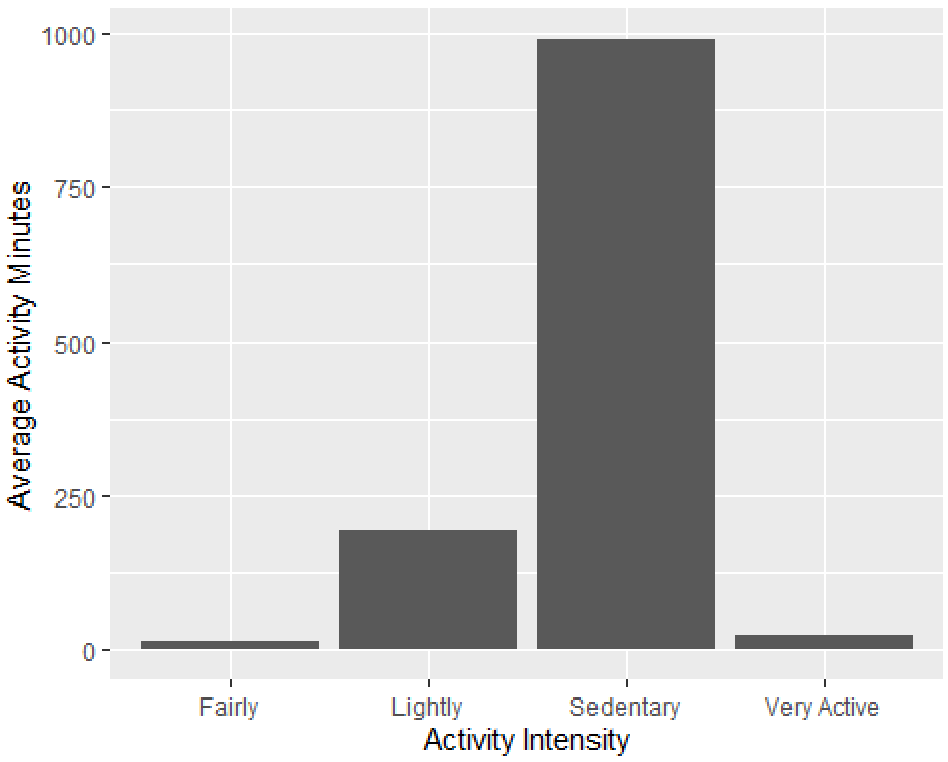
```
ggplot(data = sleep_day, aes(x = TotalMinutesAsleep, y = TotalTimeInBed))+
geom_point() + geom_smooth()

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



It is almost a linear relationship between the total minutes asleep and the total time in bed. There are a few exception but I would assume that some people do other activities in bed other than sleeping such as using their mobile phone and watching movies.

Average Minutes of Activity by Intensity

```
mean_activity_min <- c(991,192,13,21)
activity_intensity <- c("Sedentary","Lightly","Fairly", "Very Active")
minute_intensity <- data.frame(mean_activity_min, activity_intensity)

ggplot(data = minute_intensity, aes(x = activity_intensity, y =
mean_activity_min)) + labs(x = 'Activity Intensity', y = 'Average Activity
Minutes') + geom_col()
```



From the graph above, we can see that the intensity that records the highest average minutes for FitBit users are Sedentary minutes which should not be very surprising as it is very reasonable to expect the majority of the public to be somehwere in the middle of both extremes.

Key Takeaways

1. We need to collect more data to get a more accurate representation of the smart devices market.It is even better if we can obtain data from more than just one competitor to get a better estimation of the market.

2. We notice that FitBit users consider certain matrix more important as indicated by the different number of distinct users logging their information. In the analysis, we find FitBit users take the time to input calories, step taken, and intensity of activity but fail or refuse to track their sleep and weight, possibly because it is optional. We can try to simplify the method of inputing the information for these matrix or improve the design UI/UX which could possibly increase the users' tendency to log these information.

3. We also notice that FitBit does not track water intake which Bellabeat offer. This could be a marketing focus/strategy for Bellabeat as we offer a more comprehensive tracking for our overall fitness.

4. There is no indication from the data that FitBit offers any recommendation or advice to their users to improve their current fitness based on what they have inputted. Bellabeat could try to provide their users with weekly summary and updates on

their users' fitness level and provide some suggestions on how they can further improve their health. Allowing users to set a fitness goal and include a progression which the users can track should be considered. This could be an area that Bellabeat can explore and possibly utilize to gain a bigger market share.

TL;DR Recommendations for Bellabeat

1. More data is needed for a more accurate analysis.
2. Consider improving the design UI/UX or simplify the data input process to encourage users to input the information.
3. Bellabeat tracks more information than FitBit making it a more complete fitness tracker.
4. Consider providing users with weekly/monthly updates on their fitness and activity level along with some suggestions to further improve their current condition. Goals setting and tracking is a possible option.