Bellbeat-A Case Study

Shashwat Raina

31/03/2022

Introduction

COMPANY NAME: Bellabeat

WHAT IT DOES: Manufactures High-Tech health-focused products for women.

SCENARIO : I'm a junior Data Analyst working on Marketing Analyst Team at Bellabeat. I have been asked to focus on one of Bellabeat's products and analyze smart device data to gain insight into how consumers are using their smart devices. The insights that I discover will help in guiding the marketing strategy for the company.

Business Task

Analyze smart fitness device data to gain insights into how consumers use these devices for their well being. The insights will be used to guide the marketing strategy of the company. And based on the analysis, high level recommendations must be given for Bellabeat's Marketing Strategy.

Stakeholders

The Key Stakeholder are :

- **Urška Sršen**, Bellabeat Co-founder and Chief Creative officer
- **Ando Mur**, Mathematician and Bellabeat's cofounder
- Bellabeat marketing analytics team

Packages Loaded

```
#Importing required libraries
library(tidyverse)
library(lubridate)
library(dplyr)
library(ggplot2)
library(tidyr)
```

Importing Data

```
Activity <- read.csv("~/Downloads/Fitabase Data 4.12.16-5.12.16/dailyActivity_merged.csv")
View(Activity)

Calories <- read.csv("~/Downloads/Fitabase Data 4.12.16-5.12.16/hourlyCalories_merged.csv")
View(Calories)

hourly_Intensities <- read.csv("~/Downloads/Fitabase Data 4.12.16-5.12.16/hourlyIntensities_merged.csv")
View(hourly_Intensities)

sleep <- read.csv("~/Downloads/Fitabase Data 4.12.16-5.12.16/sleepDay_merged.csv")
View(sleep)

weight_LogInfo <- read.csv("~/Downloads/Fitabase Data 4.12.16-5.12.16/weightLogInfo_merged.csv")
View(weight_LogInfo)
```

To have an overview and check whether data was imported correctly, we use head() function

```
head(Activity)
```

Output :

```
##           Id ActivityDate TotalSteps TotalDistance TrackerDistance
## 1 1503960366     04/12/16      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
```

Now, we must fix formatting of data frames to make it suitable for our use.

```
#formatting

#daily intensities data
hourly_Intensities$ActivityHour=as.POSIXct(hourly_Intensities$ActivityHour, format=
"%m/%d/%Y %I:%M:%S %p", tz=Sys.timezone())
hourly_Intensities$date <- format(hourly_Intensities$ActivityHour, format = "%m/%d/
%y")
hourly_Intensities$time <- format(hourly_Intensities$ActivityHour, format = "%H:%M:
%S")

#daily calories data
Calories$ActivityHour=as.POSIXct(Calories$ActivityHour, format="%m/%d/%Y %I:%M:%S %
p", tz=Sys.timezone())
Calories$time <- format(Calories$ActivityHour, format = "%H:%M:%S")
Calories$date <- format(Calories$ActivityHour, format = "%m/%d/%y")

#daily activity data
Activity$ActivityDate=as.POSIXct(Activity$ActivityDate, format="%m/%d/%Y", tz=Sys.t
imezone())
Activity$date <- format(Activity$ActivityDate, format = "%m/%d/%y")

#sleep data
sleep$SleepDay=as.POSIXct(sleep$SleepDay, format="%m/%d/%Y %I:%M:%S %p", tz=Sys.tim
ezone())
sleep$date <- format(sleep$SleepDay, format = "%m/%d/%y")
```

Understanding the data

n_distinct function helps us find the count of unique values of data frames.

```
n_distinct(Activity$Id)
```

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

```
n_distinct(hourly_Intensities$Id)
```

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

```
n_distinct(Calories$Id)
```

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

```
n_distinct(sleep$Id)
```

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

```
n_distinct(weight_LogInfo$Id)
```

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

We get to know the number of participants in each data frames.

- There are *33* participants in Activity, Calories and daily_Intensities data frames, **24** in sleep data frame and only **8** in weight_LogInfo data frame.
- As there are very few participants in weight_LogInfo data frame, no significant recommendation and conclusion can be given.

Let's have a look at the summaries of the above data frames

```
Activity %>%
  select(TotalSteps, TotalDistance, SedentaryMinutes, Calories) %>%
  summary()
```

```
##    TotalSteps    TotalDistance    SedentaryMinutes    Calories
##  Min.   :    0   Min.   : 0.000   Min.   :   0.0   Min.   :   0
##  1st Qu.: 3790   1st Qu.: 2.620   1st Qu.: 729.8   1st Qu.:1828
##  Median : 7406   Median : 5.245   Median :1057.5   Median :2134
##  Mean   : 7638   Mean   : 5.490   Mean   : 991.2   Mean   :2304
##  3rd Qu.:10727   3rd Qu.: 7.713   3rd Qu.:1229.5   3rd Qu.:2793
##  Max.   :36019   Max.   :28.030   Max.   :1440.0   Max.   :4900
```

```
Activity %>%
  select(VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes)%>%
  summary()
```

```
##  VeryActiveMinutes FairlyActiveMinutes LightlyActiveMinutes
##  Min.   :  0.00    Min.   :  0.00      Min.   :  0.0
##  1st Qu.:  0.00    1st Qu.:  0.00      1st Qu.:127.0
##  Median :  4.00    Median :  6.00      Median :199.0
##  Mean   : 21.16    Mean   : 13.56      Mean   :192.8
##  3rd Qu.: 32.00    3rd Qu.: 19.00      3rd Qu.:264.0
##  Max.   :210.00    Max.   :143.00      Max.   :518.0
```

```
hourly_Intensities %>%
  select(TotalIntensity)%>%
  summary()
```

```
##  TotalIntensity
##  Min.   :  0.00
##  1st Qu.:  0.00
##  Median :  3.00
##  Mean   : 12.04
##  3rd Qu.: 16.00
##  Max.   :180.00
```

```
Calories %>%
  select(Calories) %>%
  summary()
```

```
##     Calories
##  Min.   : 42.00
##  1st Qu.: 63.00
##  Median : 83.00
##  Mean   : 97.39
##  3rd Qu.:108.00
##  Max.   :948.00
```

```
sleep %>%
  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_LogInfo %>%
  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
```

We get the following observation from the summaries:

- Average Sedentary time turns out to be 991 minutes which is equal to 16 hours. It should be taken into consideration and must change.
- The Average sleep time for participants is 7 hours at one time.
- Most of the participants (majority) fall in the category of lightly active.
- On Average, participants covered 7638 steps. The number is less than what CDC research found to be more accurate for better health benefits.(They found that taking 8,000 steps per day was associated with a 51% lower risk for all-cause mortality. Taking 12,000 steps per day was associated with a 65% lower risk)

Merging Data

In order to make useful visualization, we must merge data as per requirement. Lets merge Activity and sleep data frames on columns Id and date. (inner joint).

```
#Merging Data

data_merged <- merge(sleep, Activity, by=c('Id', 'date'))
head(data_merged)
```

```
##           Id     date   SleepDay TotalSleepRecords TotalMinutesAsleep
## 1 1503960366 04/12/16 2016-04-12                 1                327
## 2 1503960366 04/13/16 2016-04-13                 2                384
## 3 1503960366 04/15/16 2016-04-15                 1                412
## 4 1503960366 04/16/16 2016-04-16                 2                340
## 5 1503960366 04/17/16 2016-04-17                 1                700
## 6 1503960366 04/19/16 2016-04-19                 1                304
##   TotalTimeInBed ActivityDate TotalSteps TotalDistance TrackerDistance
## 1            346   0016-04-12      13162          8.50            8.50
## 2            407   2016-04-13      10735          6.97            6.97
## 3            442   2016-04-15       9762          6.28            6.28
## 4            367   2016-04-16      12669          8.16            8.16
## 5            712   2016-04-17       9705          6.48            6.48
## 6            320   2016-04-19      15506          9.88            9.88
##   LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 1                        0               1.88                     0.55
## 2                        0               1.57                     0.69
```

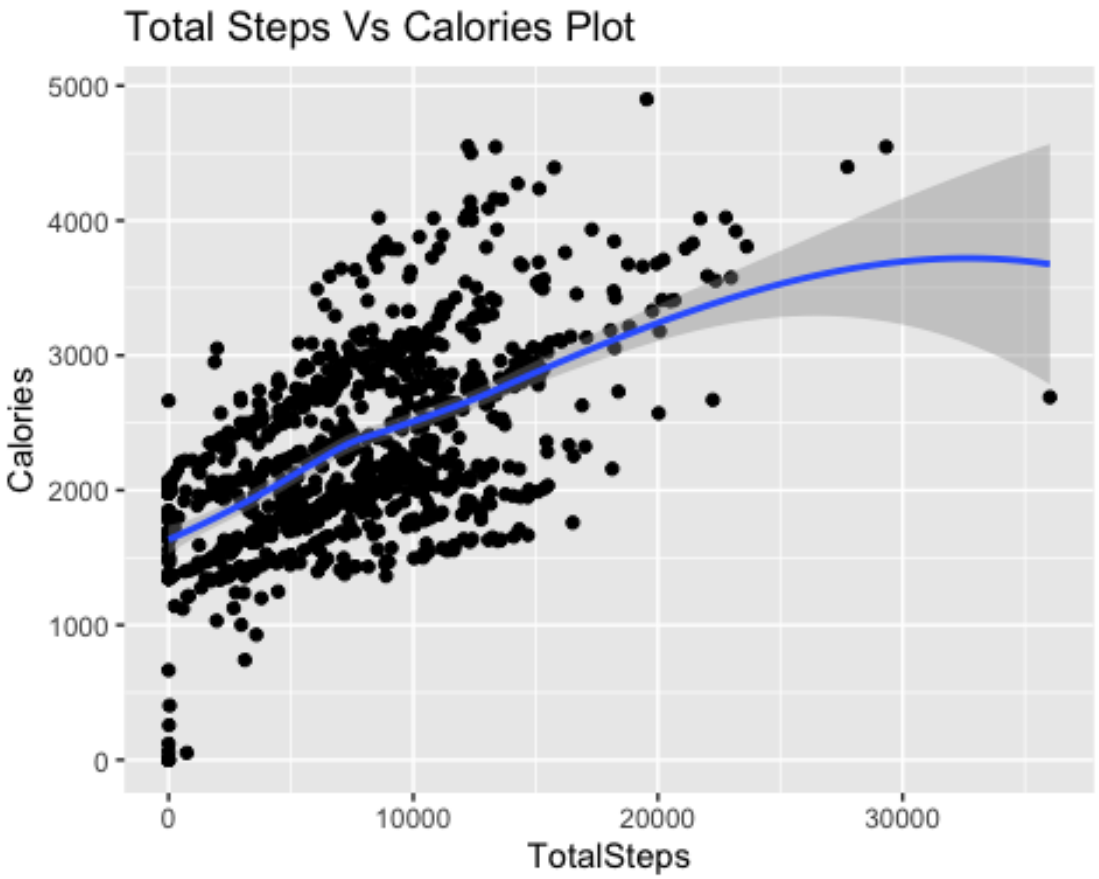
```
## 3                        0               2.14                     1.26
## 4                        0               2.71                     0.41
## 5                        0               3.19                     0.78
## 6                        0               3.53                     1.32
##   LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
## 1                6.06                       0                25
## 2                4.71                       0                21
## 3                2.83                       0                29
## 4                5.04                       0                36
## 5                2.51                       0                38
## 6                5.03                       0                50
##   FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories
## 1                  13                  328              728     1985
## 2                  19                  217              776     1797
## 3                  34                  209              726     1745
## 4                  10                  221              773     1863
## 5                  20                  164              539     1728
## 6                  31                  264              775     2035
```

Visualization : To get the insights from the data

```
#Total Steps vs Calories graph


ggplot(data=Activity, aes(x=TotalSteps, y=Calories)) +
  geom_point() + geom_smooth(method="loess") + labs(title ='Total Steps Vs Calories
Plot')

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

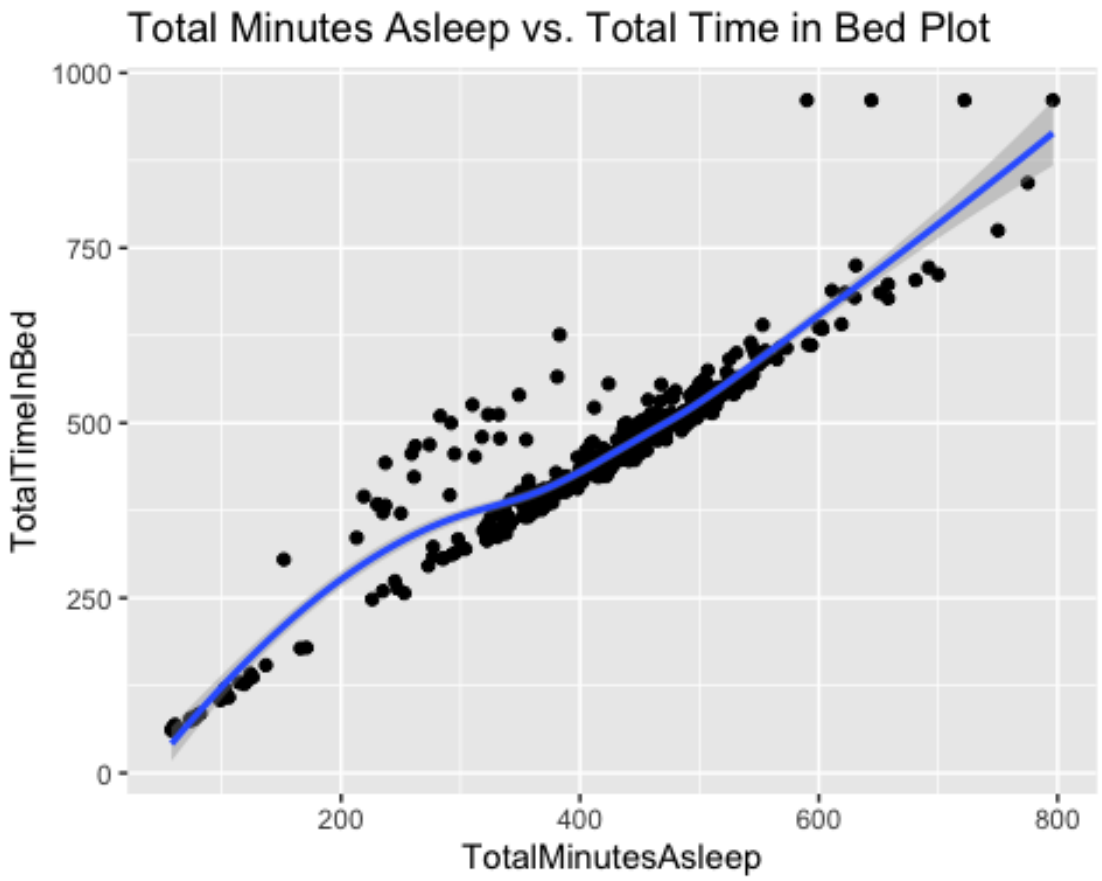


Observation : It shows the basic relation between Total Steps and Calorie burned. Its shows a straight line and have a +ve relation

```
# Total Minutes Asleep vs. Total Time in Bed Graph

ggplot(data=sleep, aes(x=TotalMinutesAsleep, y=TotalTimeInBed)) +
  geom_point() + geom_smooth(method="loess") + labs(title='Total Minutes Asleep vs.
Total Time in Bed Plot')

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



Observations :

- The graph clearly shows the direct relation between Total Minutes Asleep and Total Time in Bed.
- To improve the sleep, Bellabeat App can use the feature of notifications to remind people people go to sleep.

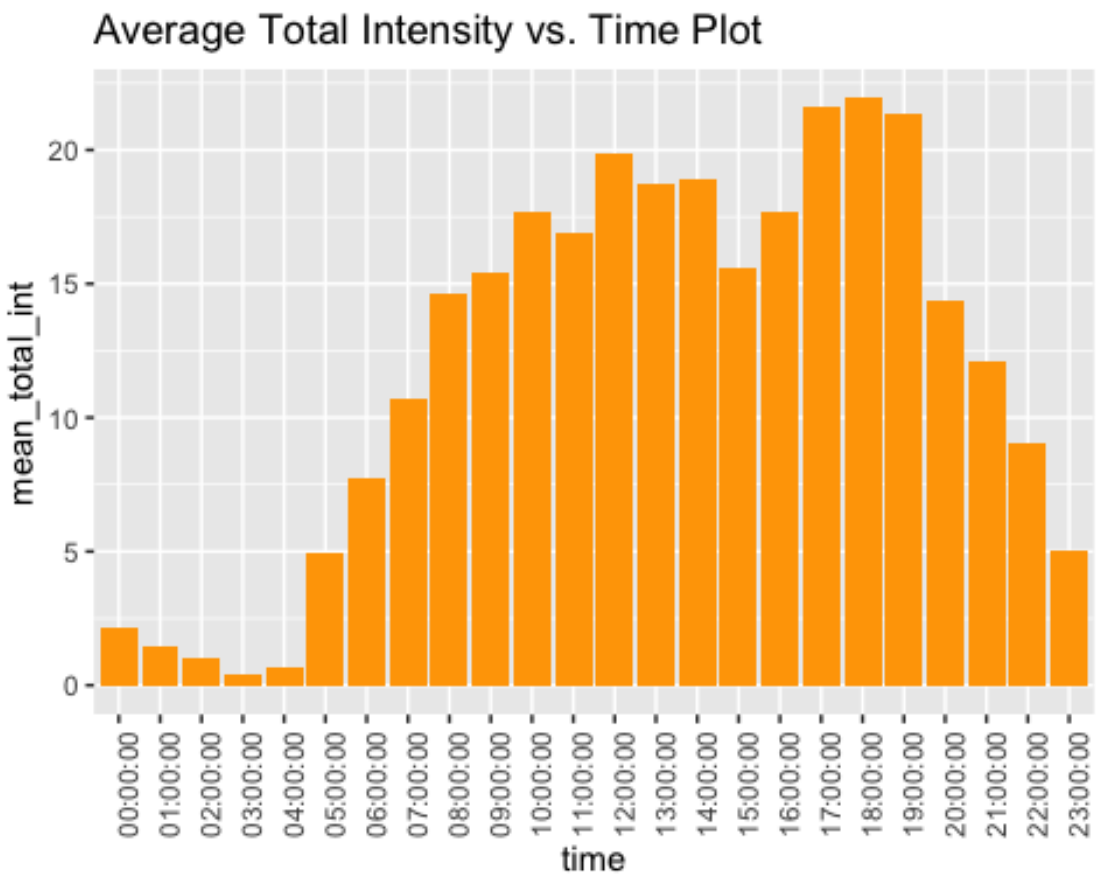
Intensities over time (hourly)

```
#Intensities Data

int_new <- hourly_Intensities %>%
  group_by(time) %>%
  drop_na() %>%
  summarise(mean_total_int = mean(TotalIntensity))

ggplot(data=int_new, aes(x=time, y=mean_total_int)) + geom_histogram(stat = "identi
ty", fill='orange') +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(title="Average Total Intensity vs. Time Plot")

## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



Observations :

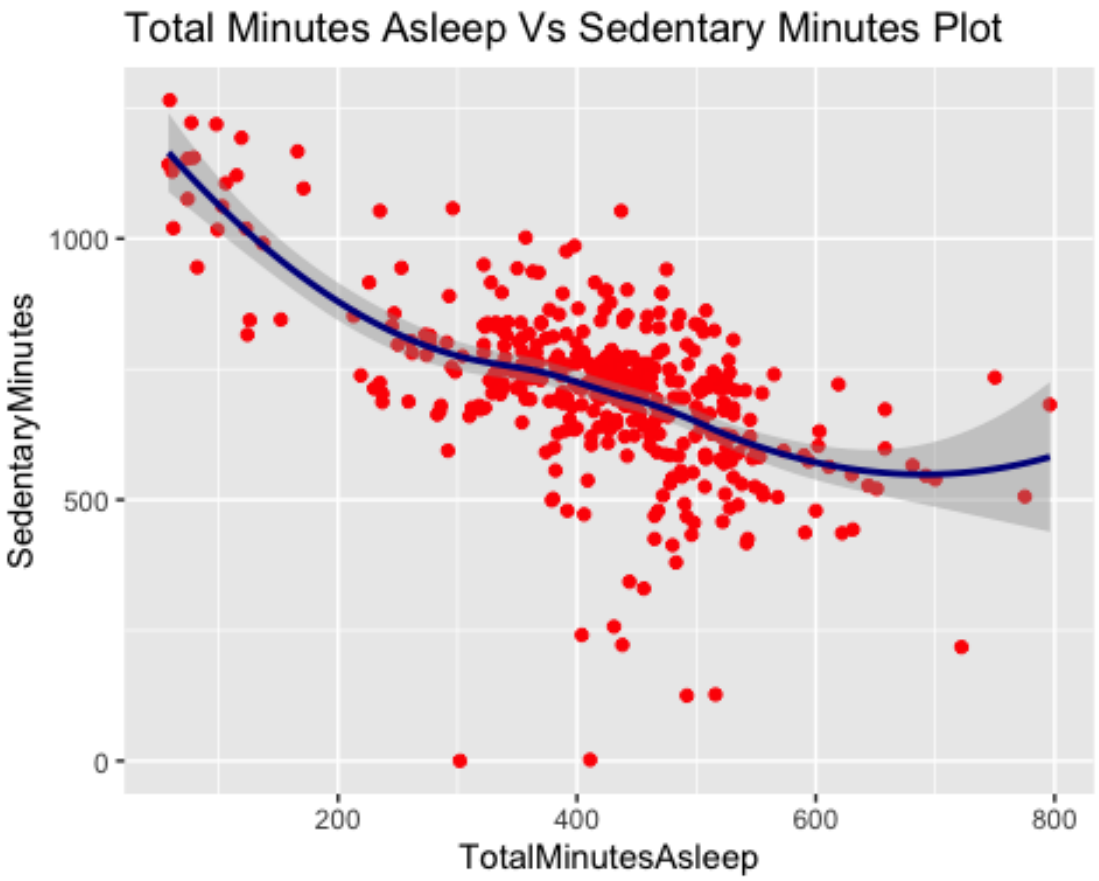
- The graph shows that people are more active between 5 am and 10 pm.
- The busiest time when people are most active is between 5 pm and 7 pm. This is true because people returning from the office. After returning from the office, they go to gym or for a walk.

Let's look at the relationship between Total Minutes Asleep and Sedentry Minutes.

```
#Merged data (sleep and activity)

ggplot(data=data_merged, aes(x=TotalMinutesAsleep, y=SedentaryMinutes)) +
  geom_point(color='red') + geom_smooth(color='darkblue',method="loess") +
  labs(title="Total Minutes Asleep Vs Sedentary Minutes Plot")

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



Observations :

- The graph shows the -ve relationship between Sedentary Minutes and Sleep time.
- Bellabeat can enable user to get reminded about reducing their sedentary time, to improve their sleep.

Conclusion & Recommendations

We analyzed Fitbit Fitness Tracker Data which helped us gain various insights. These insight will help us Bellabeat's marketing strategy. This will also help Bellabeat empower women with knowledge about their own health and habits.

Target Audience

The key target audience are the women with full time jobs and who spend lot of time on computer.

Recommendations

- People who target to loose weight must be encouraged to control and maintain their daily calorie consumption. Bellabeat App can do that by having the record of their user's calorie intake. And based on that, give suggestions to the specific users.
- Bellbeat should use the feature of notifications for People facing sleeping problems. The notification can include poistive messages or basic things to remember before going to sleep. They can also suggest users to reduce sedentary time to improve their sleep
- Focus of Bellabeat's online campaign :

  The Bellabeat app is not just fitness activity app. Its a 'guide' who empowers women to balance between personal and professional life. And top of that, encourage them to inculcate healthy habits and routines Bellabeat can do that by educating them through daily app reminders, guides, e-mails, workshops etc.
- We should also connect with our users through social media.(Facebook, Instagram etc)
- People who are covering steps less than what is recommended by CDC research, must be encouraged by Bellabeat to complete the recommended 8000 steps. Bellabeat can do that by giving them steps challenges and explaining users about its health benefits.

```
--- The End ---
```