Bellabeat_Case_Study

Sushant

2022-06-26

About Bellabeat

Bellabeat is a high-tech manufacturer of health-focused products for women. Bellabeat is a successful small company, but they have the potential to become a larger player in the global smart device market.The Bellabeat app provides users with health data related to their activity, sleep, stress,menstrual cycle, and mindfulness habits. This data can help users better understand their current habits and make healthy decisions. The Bellabeat app connects to their line of smart wellness products. By 2016, Bellabeat had opened offices around the world and launched multiple products. Bellabeat products became available through a growing number of online retailers in addition to their own e-commerce channel on their website. The company has invested in traditional advertising media, such as radio, out-of-home billboards, print, and television, but focuses on digital marketing extensively. Bellabeat invests year-round in Google Search, maintaining active Facebook and Instagram pages, and consistently engages consumers on Twitter. Additionally, Bellabeat runs video ads on Youtube and display ads on the Google Display Network to support campaigns around key marketing dates

Our Products

- Leaf: Bellabeat's classic wellness tracker can be worn as a bracelet, necklace, or clip. The Leaf tracker connects to the Bellabeat app to track activity, sleep, and stress.
- Time: This wellness watch combines the timeless look of a classic timepiece with smart technology to track user activity, sleep, and stress. The Time watch connects to the Bellabeat app to provide you with insights into your daily wellness.
- Spring: This is a water bottle that tracks daily water intake using smart technology to ensure that you areappropriately hydrated throughout the day. The Spring bottle connects to the Bellabeat app to track your hydration levels.

Case Scenario (The Ask Phase)

Ms. Sršen (CEO of Company) asks you to analyze smart device usage data in order to gain insight into how consumers use non-Bellabeat smart devices. She then wants you to select one Bellabeat product to apply these insights to in your presentation. These questions will guide your analysis:

- 1. What are some trends in smart device usage?
- 2. How could these trends apply to Bellabeat customers?
- 3. How could these trends help influence Bellabeat's marketing strategy?

My approach of solving problem :

I have used my basic approach that includes 7 phases : Ask - Prepare - Process - Analyse - Share - Act .

Preparing for data

The data for this analysis will come from FitBit Fitness Tracker Data on Kaggle. These 18 datasets were generated by respondents to a distributed survey via Amazon Mechanical Turk between 03.12.2016–05.12.2016.

Thirty eligible Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring. Individual reports can be parsed by export session ID (column A) or timestamp (column B). Variation between output represents use of different types of Fitbit trackers and individual tracking behaviors / preferences.

Limitation In data :

The data has incudes only 30 participants which has reduced the sample size and most of the data is of weekdays so there may be inaccuracies in our analysis.

Processing of given data

Installing and loading of packages:

```
install.packages("janitor")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.2'
## (as 'lib' is unspecified)
```

```
install.packages("tidyverse")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.2'
## (as 'lib' is unspecified)
```

```
install.packages("dplyr")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.2'
## (as 'lib' is unspecified)
```

```
install.packages("ggplot2")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.2'
## (as 'lib' is unspecified)
```

```
install.packages("sqldf")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.2'
## (as 'lib' is unspecified)
```

```
library(tidyverse)
```

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

## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.7     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1

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

```
library(janitor)
```

```
##
## Attaching package: 'janitor'

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

```r
library(dplyr)
library(ggplot2)
library(sqldf)
```

```
## Loading required package: gsubfn

## Loading required package: proto

## Loading required package: RSQLite
```

Import all datasets

I have already joined and cleaned the datasets in Bigquery and saved these datasets in internal system and now lets just rerieve them here

```r
daily_sleep<- read.csv("sleepDay_merged.csv")
daily_activity1<- read.csv("dailyActivity_merged.csv")
daily_calories2<- read.csv("dailyCalories_merged.csv")
daily_calories_burnt<- read.csv("dailyActivity_merged.csv")
weight <- read.csv("Weight.csv")
daily_distance<- read.csv("daily_distance.csv")
```

Look for summary of our datasets

```r
head(daily_activity1)
```

```
##           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
```

```
head(daily_distance)
```

```
##           Id ActivityDate VeryActiveDistance ModeratelyActiveDistance
## 1 1624580081   2016-05-01              36019                     4.19
## 2 1644430081   2016-04-14              11037                     2.56
## 3 1644430081   2016-04-19              11256                     2.53
## 4 1644430081   2016-04-28               9405                     2.32
## 5 1644430081   2016-04-30              18213                     3.14
## 6 1644430081   2016-05-03              12850                     4.09
##   SedentaryActiveDistance LightActiveDistance
## 1                    0.02                1.91
## 2                    0.00                5.10
## 3                    0.00                5.30
## 4                    0.00                4.31
## 5                    0.00                9.46
## 6                    0.00                4.54
```

```
head(daily_calories2)
```

```
##           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
```

```
head(weight)
```

```
##           Id TotalSteps TotalDistance   BMI WeightKg
## 1 4319703577      10780          7.23 27.45     72.4
## 2 4319703577      10780          7.23 27.38     72.3
## 3 5577150313      12087          9.08 28.00     90.7
## 4 5577150313      14269         10.66 28.00     90.7
## 5 5577150313      12231          9.14 28.00     90.7
## 6 5577150313      10830          8.09 28.00     90.7
```

Analysing the Datasets

I have Manipulated the daily_sleep dataset to get the clean and easy to undertand data .

```
daily_sleep_clean <- clean_names(daily_sleep)

daily_sleep_clean <- daily_sleep %>%
  mutate(Total_bed_time = TotalTimeInBed/60 , total_hours_asleep = TotalMinutesAsleep/60)

glimpse(daily_sleep_clean)
```

```
## Rows: 413
## Columns: 7
## $ 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~
## $ Total_bed_time     <dbl> 5.766667, 6.783333, 7.366667, 6.116667, 11.866667, ~
```

```
## $ total_hours_asleep <dbl> 5.450000, 6.400000, 6.866667, 5.666667, 11.666667, ~
```

The daily average distance of users tracked by smart watch

```
daily_activity1<- read.csv("dailyActivity_merged.csv")
head(daily_activity1)
```

```
##           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
```

```
 avg_total_distance<- daily_activity1 %>%
   group_by(Id) %>%

   summarise(avg_distance= mean(TotalDistance))

 head(avg_total_distance)
```

```
## # A tibble: 6 x 2
##           Id avg_distance
##        <dbl>        <dbl>
## 1 1503960366        7.81
## 2 1624580081        3.91
## 3 1644430081        5.30
## 4 1844505072        1.71
## 5 1927972279        0.635
## 6 2022484408        8.08
```

calories burnt by users on daily basis

```
daily_calories2<- read.csv("dailyCalories_merged.csv")


 daily_calories_burnt<- read.csv("dailyActivity_merged.csv")


daily_cal_burnt<- daily_calories_burnt %>%
  select(TotalDistance , Calories) %>%
  rename(total_distance = TotalDistance)

head(daily_cal_burnt)
```

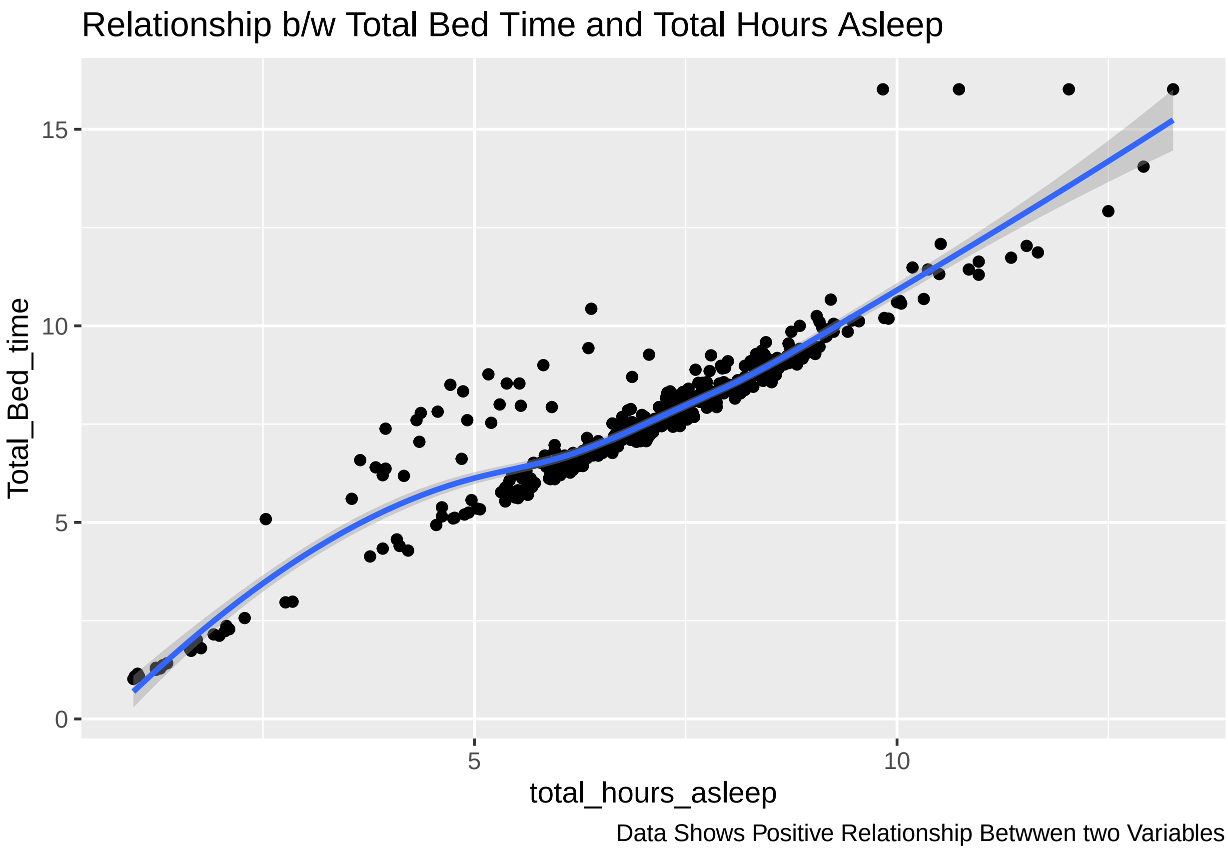
```
##   total_distance Calories
## 1           8.50     1985
## 2           6.97     1797
## 3           6.74     1776
## 4           6.28     1745
## 5           8.16     1863
## 6           6.48     1728
```

Sharing the data

In this phase , I have used ggplot2 function to create visualiations for depicting the trends and relationships.

```
ggplot(data=daily_sleep_clean)+ geom_point(mapping= aes(x=total_hours_asleep , y= Total_bed_time)) +
  geom_smooth(mapping= aes(x=total_hours_asleep , y= Total_bed_time))+
  labs(title="Relationship b/w Total Bed Time and Total Hours Asleep" ,
  caption=paste0("Data Shows Positive Relationship Betwwen two Variables"),
  x="total_hours_asleep" , y= "Total_Bed_time")
```

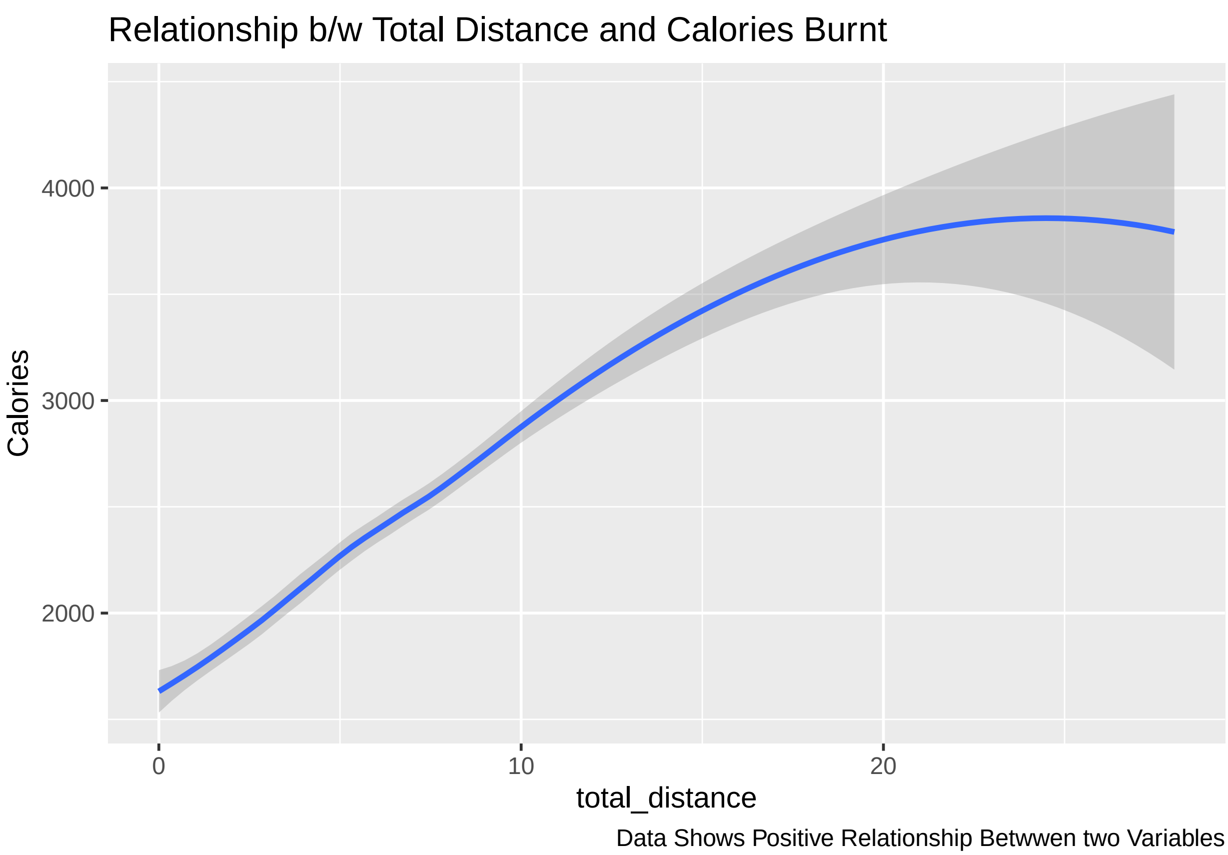
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



The depicts a positive relationship between total minutes asleep and total time in bed. For the most part, the time participants spent asleep and the time they spent in bed was very similar.

```
ggplot(data=daily_cal_burnt)+geom_smooth(mapping = aes(x=total_distance , y= Calories)) +
  labs(title="Relationship b/w Total Distance and Calories Burnt" ,
       caption=paste0("Data Shows Positive Relationship Betwwen two Variables"))
```

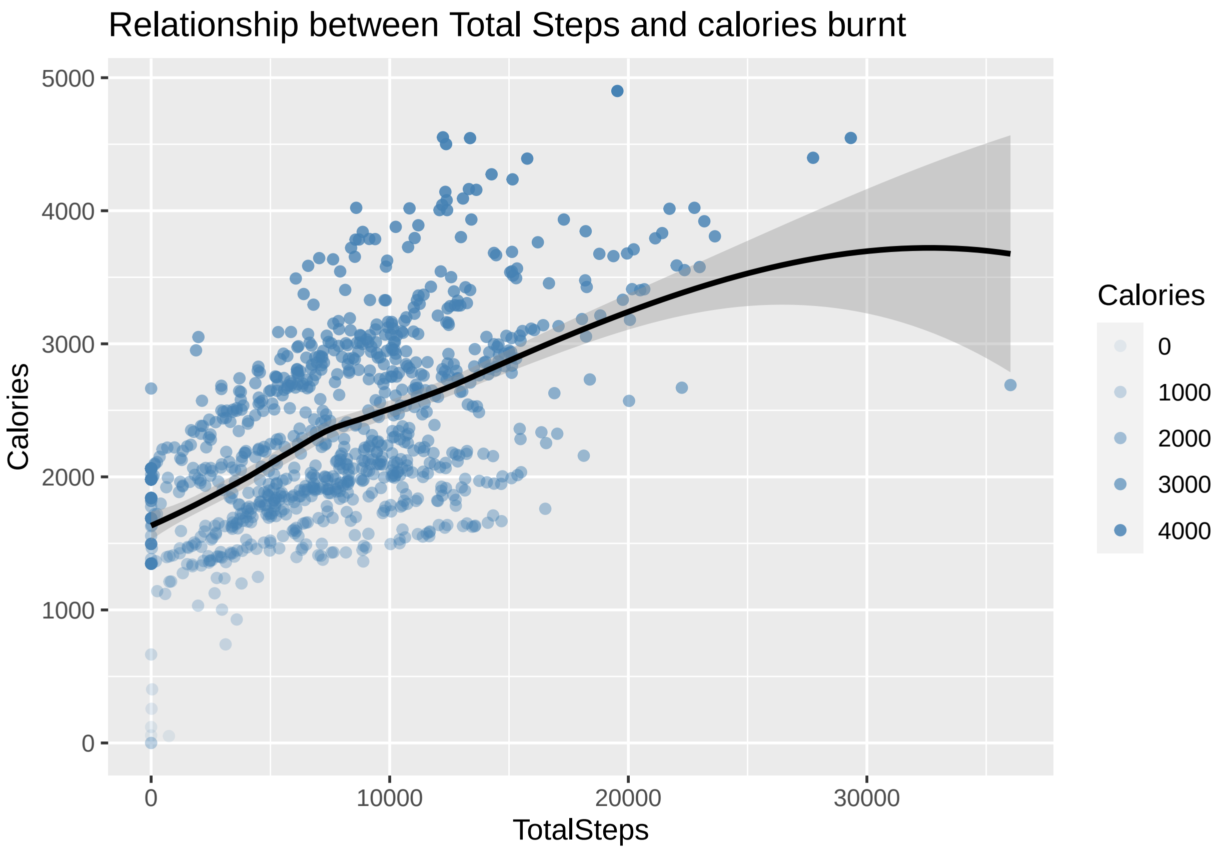
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



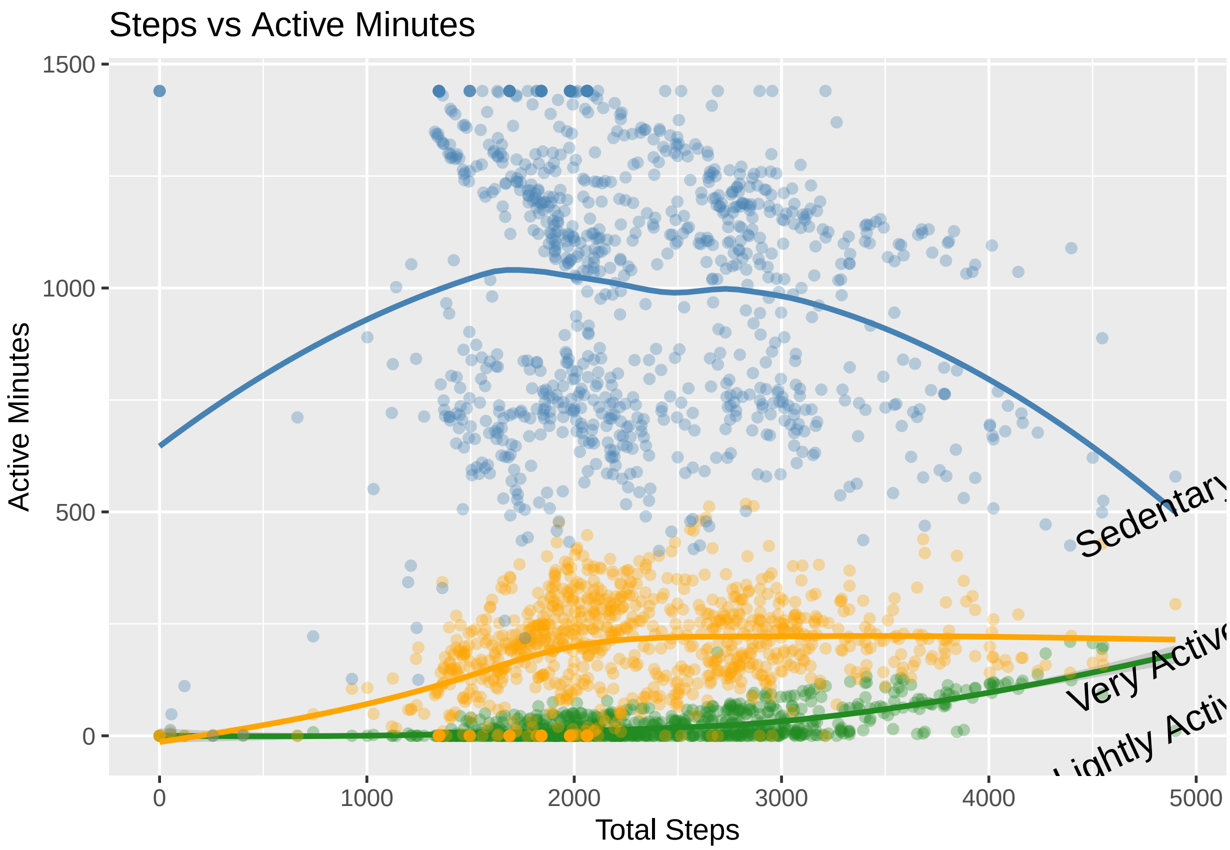
As we can see it shows a positive trend between total distance and total daily calories burned. As the participants moved a greater distance, the number of calories they burned also increased.

```
ggplot(data=daily_calories_burnt)+geom_point(mapping = aes(x=TotalSteps , y=Calories  ,alpha = Calories
  geom_smooth(mapping = aes(x=TotalSteps , y=Calories ), color = "black") +
  labs(title = "Relationship between Total Steps and calories burnt" )
```

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

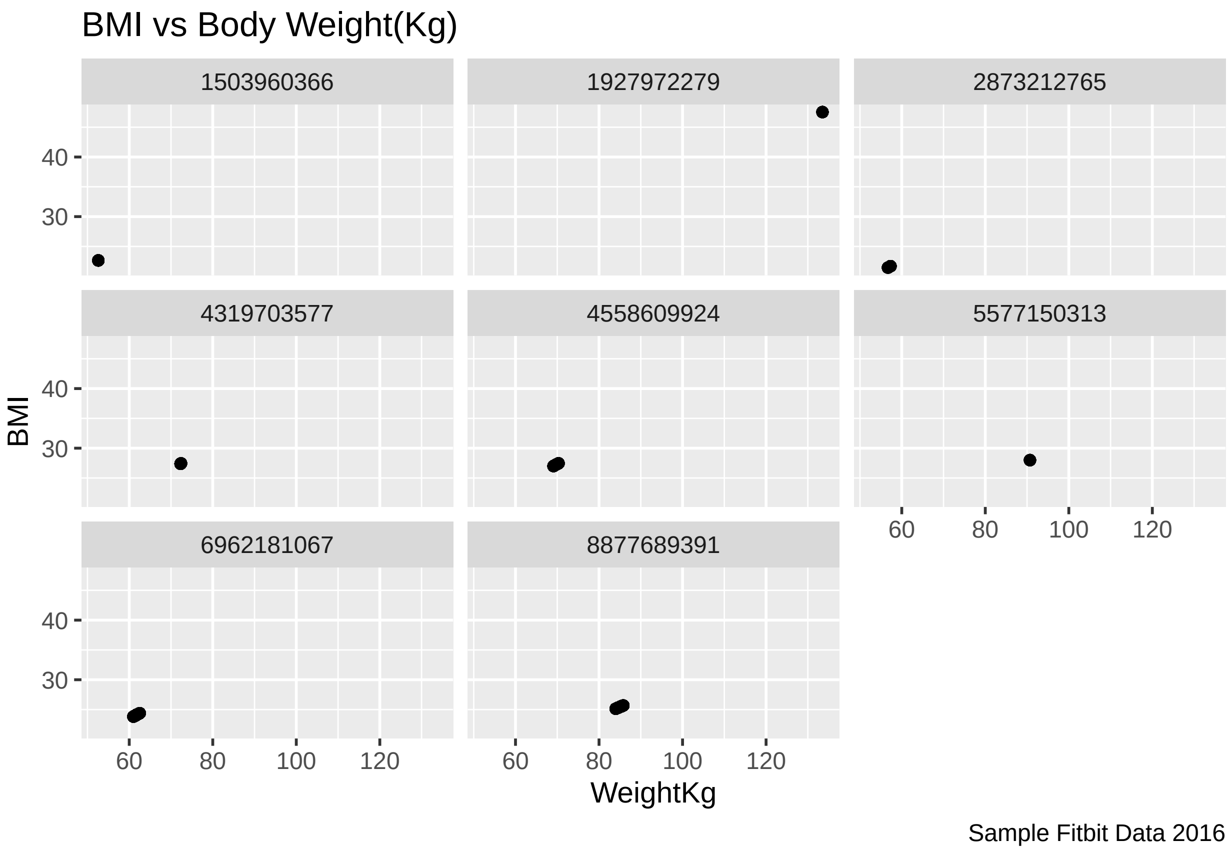



```
ggplot(data=daily_calories_burnt)+ geom_point(mapping=aes(x=Calories, y=VeryActiveMinutes), color = "fo
 geom_smooth(method = loess,formula =y ~ x,mapping=aes(x=Calories, y=VeryActiveMinutes, color=VeryActiv
 annotate("text", x=4800, y=160, label="Very Active", color="black", size=5, angle =25) +

 geom_point(mapping=aes(x=Calories, y=LightlyActiveMinutes), color = "orange", alpha = 1/3) +

 geom_smooth(method = loess,formula =y ~ x,mapping=aes(x=Calories, y=LightlyActiveMinutes, color=Lightl

 annotate("text", x=4800, y=0, label="Lightly Active", color="black", size=5 , angle=25)+

 geom_point(mapping=aes(x=Calories, y=SedentaryMinutes), color = "steelblue", alpha = 1/3) +

 geom_smooth(method = loess,formula =y ~ x,mapping=aes(x=Calories, y=SedentaryMinutes, color=Sedentary

 annotate("text", x=4800, y=500, label="Sedentary", color="black", size=5 , angle=25)+
 # #annotate("text", x=4800, y=160, label="Very Active", color="black", size=5, angle =25)+
 # annotate("text", x=3500, y=0, label="Lightly Active", color="black", size=5 , angle=25)+
 # annotate("text", x=4800, y=500, label="Sedentary", color="black", size=5 , angle=25)+

 labs(x = "Total Steps", y = "Active Minutes", title="Steps vs Active Minutes")
```



Comparing the three active levels to the total steps, we see most data is concentrated on users who take about 5000 to 15000 steps a day. These users spent an average between 8 to 13 hours in sedentary, 5 hours in lightly active, and 1 to 2 hour for fairly and very active.

```
ggplot(data = weight) + geom_point(mapping = aes(x=WeightKg , y= BMI)) +

 facet_wrap(~Id ) +
labs(title = "BMI vs Body Weight(Kg)" , caption  = "Sample Fitbit Data 2016" )
```



These graphs shows the Body mass index of different users according to there Body weight . The data requires expert level recognisation as from graph it only shows the more the body weight more is the BMI level but as we know data we have has limitation and this may or may not be accurate.

Now what we can do to influnce Bellabeat marketing strategy with our findings?

This is our Act Phase :

* Transform the Mobile App :-

Bellabeat app doesn't have a feature to share Data a in social-media platforms like fitbit and also they doesn't have 30 days trial subscription which is very importantly to induce in the app to come up with the competition. It needs to send daily reminders and notifications to users for example there daily goals and achivements.

* Features to introduce in smart watch :-

Our smart watch requires features like heart beat reader , The accuracy of our device needs to be updated . It should have an feature to conect with any mobile device .

* What to provide to Bellabeat User:-

- 1. Offer 30-day free trial subscription.
- 2. Offer reduced subscription fee when a member refers a friend.
- 3. Offer discounts for Bellabeat smart device products with membership.

- 4. Partner with health & fitness companies and offer discounts for members.

For Some information on how customers usually use there fitness watch click herelink.