Bellabeat Case Study

mahmoud hassan

March 2016

### About a company

Bellabeat, 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. 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

### Business task

Identify potential opportunities for growth and recommendations for the Bellabeat marketing strategy improvement based on trends in smart device usage.

### Questions for the 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 marketing strategy

### What Are We Talking About?

How Do **Annual Members** And **Casual Riders** Use Cyclistic bikes Differently?

* I used Cyclistic’s historical trip data to analyze and identify trends.to visit data source click here [link](https://divvy-tripdata.s3.amazonaws.com/index.html)

##### first install.packages()

library("tidyr")  
library("tidyverse")

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ ggplot2 3.4.0 ✔ dplyr 1.0.10  
## ✔ tibble 3.1.8 ✔ stringr 1.5.0   
## ✔ readr 2.1.3 ✔ forcats 0.5.2   
## ✔ purrr 0.3.5   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library("lubridate")

## Loading required package: timechange  
##   
## Attaching package: 'lubridate'  
##   
## The following objects are masked from 'package:base':  
##   
## date, intersect, setdiff, union

library("janitor")

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

library("dplyr")  
library("scales")

##   
## Attaching package: 'scales'  
##   
## The following object is masked from 'package:purrr':  
##   
## discard  
##   
## The following object is masked from 'package:readr':  
##   
## col\_factor

library("ggplot2")

#### Importing datasets

For this project, I will use FitBit Fitness Tracker [Data](https://www.kaggle.com/datasets/arashnic/fitbit).

#### importing and cleaning data

* Import csv files without blank data which does not read as (NA)

setwd("D:/data analysis/capstone project/case study 2/archive (1)/Fitabase Data 4.12.16-5.12.16")  
activity <-   
 read.csv("dailyActivity\_merged.csv",stringsAsFactors = F)  
calories <-   
 read.csv("hourlyCalories\_merged.csv",stringsAsFactors =F )  
intensities <-  
 read.csv("hourlyIntensities\_merged.csv",stringsAsFactors =F )  
sleep <-  
 read.csv("sleepDay\_merged.csv",stringsAsFactors = F)  
weight <-   
 read.csv("weightLogInfo\_merged.csv",stringsAsFactors = F)

I already checked the data in Google Sheets. I just need to make sure that everything were imported correctly by using View() and head() functions.

head(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

#### convert data as character to date and time

##### intensities

convert data as character to date and time and add new columns to time and date

intensities$ActivityHour<-as.POSIXct(intensities$ActivityHour,  
 format="%m/%d/%Y %H:%M:%S")  
intensities$time<-format(intensities$ActivityHour,format = "%H:%M:%S")  
intensities$date<-format(intensities$ActivityHour,"%m:%d:%Y ")

##### calories

convert data as character to date and time and add new columns to time and date

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")

##### weight

weight$Date<-as.POSIXct(weight$Date,format="%m/%d/%Y %I:%M:%S %p", tz=Sys.timezone())  
weight$Date<-format(weight$Date,format = "%m/%d/%Y")

##### activity

activity$ActivityDate<-as.POSIXlt(activity$ActivityDate,format = "%m/%d/%Y")  
activity$date <- format(activity$ActivityDate, format = "%m/%d/%y")

##### sleep

sleep$SleepDay<-as.POSIXlt(sleep$SleepDay,format="%m/%d/%Y %H:%S")  
sleep$date <- format(sleep$SleepDay, format = "%m/%d/%y")

### Exploring and summarizing data

n\_distinct(activity$Id)

## [1] 33

n\_distinct(calories$Id)

## [1] 33

n\_distinct(intensities$Id)

## [1] 33

n\_distinct(sleep$Id)

## [1] 24

n\_distinct(weight$Id)

## [1] 8

This information tells us about number participants in each data sets.

There is 33 participants in the activity, calories and intensities data sets, 24 in the sleep and only 8 in the weight data set. 8 participants is not significant to make any recommendations and conclusions based on this data.

##### summary statistics of the activity table

activity %>% select(TotalSteps,TotalDistance,SedentaryMinutes,Calories,VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes) %>%   
 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   
## 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

* Average sedentary time is 991 minutes or 16 hours. Definately needs to be reduced!
* The majority of the participants are lightly active.

##### summary statistics of the Calories table

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

##### summary statistics of the sleep table

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

* On the average, participants sleep 1 time for 433 Minutes or 7 hours.

##### summary statistics of the weight table

weight %>%  
 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 data

Before beginning to visualize the data, I need to merge two data sets. I’m going to merge (inner join) activity and sleep on columns Id and date (that I previously created after converting data to date time format).

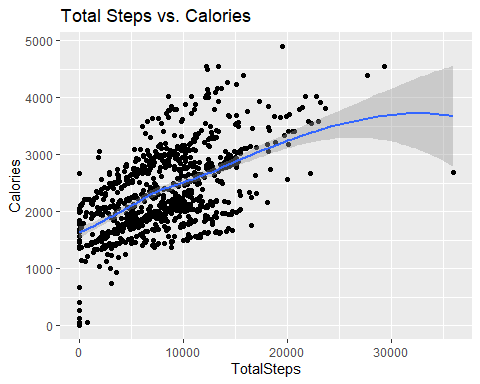
merged\_data<-merge(activity,sleep,by=c("Id","date"))  
  
head(merged\_data)

## Id date ActivityDate TotalSteps TotalDistance TrackerDistance  
## 1 1503960366 04/12/16 2016-04-12 13162 8.50 8.50  
## 2 1503960366 04/13/16 2016-04-13 10735 6.97 6.97  
## 3 1503960366 04/15/16 2016-04-15 9762 6.28 6.28  
## 4 1503960366 04/16/16 2016-04-16 12669 8.16 8.16  
## 5 1503960366 04/17/16 2016-04-17 9705 6.48 6.48  
## 6 1503960366 04/19/16 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 SleepDay  
## 1 13 328 728 1985 2016-04-12  
## 2 19 217 776 1797 2016-04-13  
## 3 34 209 726 1745 2016-04-15  
## 4 10 221 773 1863 2016-04-16  
## 5 20 164 539 1728 2016-04-17  
## 6 31 264 775 2035 2016-04-19  
## TotalSleepRecords TotalMinutesAsleep TotalTimeInBed  
## 1 1 327 346  
## 2 2 384 407  
## 3 1 412 442  
## 4 2 340 367  
## 5 1 700 712  
## 6 1 304 320

#### Visualization

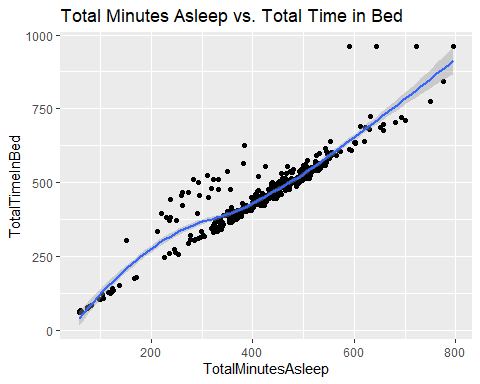
ggplot(data=activity, aes(x=TotalSteps, y=Calories)) +   
 geom\_point() + geom\_smooth() + labs(title="Total Steps vs. Calories")

## `geom\_smooth()` using method = 'loess' and formula = 'y ~ x'

 - I see positive correlation here between Total Steps and Calories, which is obvious - the more active we are, the more calories we burn.

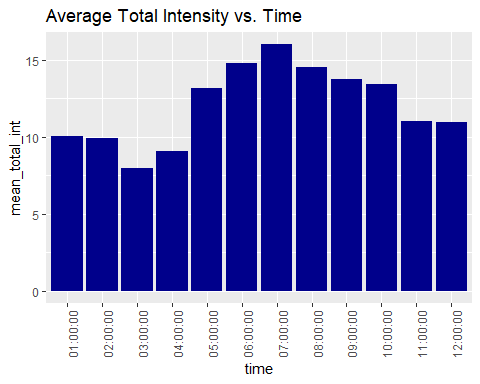
ggplot(data=sleep, aes(x=TotalMinutesAsleep, y=TotalTimeInBed)) +   
 geom\_point()+ geom\_smooth()+labs(title="Total Minutes Asleep vs. Total Time in Bed")

## `geom\_smooth()` using method = 'loess' and formula = 'y ~ x'

 The relationship between Total Minutes Asleep and Total Time in Bed looks linear. So if the Bellabeat users want to improve their sleep, we should consider using notification to go to sleep.

int\_new <- 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 = "identity", fill='darkblue') +  
 theme(axis.text.x = element\_text(angle = 90)) +  
 labs(title="Average Total Intensity vs. Time")

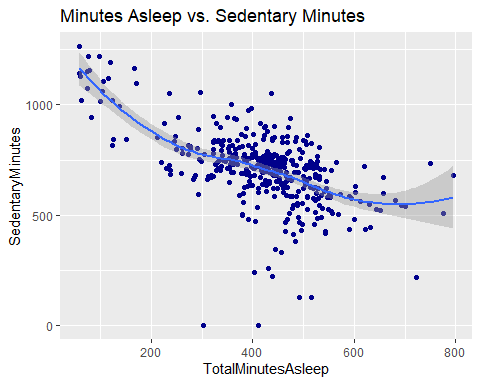
## Warning in geom\_histogram(stat = "identity", fill = "darkblue"): Ignoring  
## unknown parameters: `binwidth`, `bins`, and `pad`

 \* After visualizing Total Intensity hourly, I found out that people are more active between 5 am and 10pm.

* Most activity happens between 5 pm and 7 pm - I suppose, that people go to a gym or for a walk after finishing work. We can **use this time in the Bellabeat app to remind and motivate users to go for a run or walk**.

ggplot(data=merged\_data, aes(x=TotalMinutesAsleep, y=SedentaryMinutes)) +   
geom\_point(color='darkblue') + geom\_smooth() +  
 labs(title="Minutes Asleep vs. Sedentary Minutes")

## `geom\_smooth()` using method = 'loess' and formula = 'y ~ x'



* Here we can clearly see the negative relationship between Sedentary Minutes and Sleep time.
* As an idea: **if Bellabeat users want to improve their sleep, Bellabeat app can recommend reducing sedentary time**.
* Keep in mind that we need to support this insights with more data, because correlation between some data doesn’t mean causation.

#### recommendations for the business

After analyzing FitBit Fitness Tracker Data, I found some insights that would help influence Bellabeat marketing strategy.

* Women who work full-time jobs (according to the hourly intensity data) and spend a lot of time at the computer/in a meeting/ focused on work they are doing (according to the sedentary time data).
* These women do some light activity to stay healthy (according to the activity type analysis). Even though they need to improve their everyday activity to have health benefits. They might need some knowledge about developing healthy habits or motivation to keep going.

##### The key message for the Bellabeat online campaign

The Bellabeat app is not just another fitness activity app. It’s a guide (a friend) who empowers women to balance full personal and professional life and healthy habits and routines by educating and motivating them through daily app recommendations.

##### Ideas for the Bellabeat app

* As there is no gender information about the participants, I assumed that all genders were presented and balanced in this data set.
* If users want to lose weight, it’s probably a good idea to control daily calorie consumption. Bellabeat can suggest some ideas for low-calorie lunch and dinner.
* If users want to improve their sleep, Bellabeat should consider using app notifications to go to bed.
* Most activity happens between 5 pm and 7 pm - I suppose, that people go to a gym or for a walk after finishing work. Bellabeat can use this time to remind and motivate users to go for a run or walk.
* As an idea: if users want to improve their sleep, the Bellabeat app can recommend reducing sedentary time.