Title: “Bellabeat Case Study with R”

Author: Oluwatobi Owamokele

Date: ‘2022-05-31’

The Company - Bellabeat

Bellabeat is a high-tech manufacturer of health-focused smart products for women. The stakeholders and founders are Urška Sršen and Sando Mur founded Bellabeat. Sršen used her background as an artist to develop beautifully designed technology that informs and inspires women around the world. Collecting data on activity, sleep, stress, and reproductive health has allowed Bellabeat to empower women with knowledge about their own health and habits. Since it was founded in 2013, Bellabeat has grown rapidly and quickly positioned itself as a tech-driven wellness company for women.

Questions for 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?

Business Task

Identifying opportunities for growth based on market trends for Bellabeat’s fitness tracker and recommendations to improve marketing strategy in order to thrive better in the market.

About dataset

Data was got from public dataset- https://www.kaggle.com/datasets/arashnic/fitbit (CC0: Public Domain, dataset made available through Mobius): This Kaggle data set contains personal fitness tracker from fitbit users. These Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and

sleep monitoring. It includes information about daily activity, steps, sleep patterns, intensities of activities, weight ,heart rate that can be used to explore users' habits.

Loading packages for analysis

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

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

```
install.packages("lubridate")
```

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

```
install.packages("tidyr")
```

```
## 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)
```

Loading packages

```
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.
3.1 ──

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

## ── Conflicts ────────────────────────────────────────── tidyverse_conflict
s() ──
## ✕ dplyr::filter() masks stats::filter()
## ✕ dplyr::lag()    masks stats::lag()
```

```
library(lubridate)
```

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

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

library(tidyr)
library(dplyr)
library(ggplot2)
```

Import datasets

```
activity <- read.csv("dailyActivity_merged.csv")
calories <- read.csv("dailyCalories_merged.csv")
intensities <- read.csv("dailyIntensities_merged.csv")
sleep <- read.csv("sleepDay_merged.csv")
steps <- read.csv("dailySteps_merged.csv")
weight_log_info <- ("weightLogInfo_merged.csv")

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
```

Data Cleaning

I already noticed the error in date format when I initially loaded the datasets and this error is present in all. Before continuing this analysis, the datatype must be changed correctly to date and datetime.

Formatting datasets

```
activity$ActivityDate <- as.Date(activity$ActivityDate, "%m/%d/%Y")
activity$date <- format(activity$ActivityDate, format = "%m/%d/%y")
calories$ActivityDay <- as.Date(calories$ActivityDay , "%m/%d/%Y")
intensities$ActivityDay <- as.Date(intensities$ActivityDay, "%m/%d/%Y")
sleep$SleepDay <- as.POSIXct(sleep$SleepDay, format= "%m/%d/%Y %I:%M:%S %p",
tz= Sys.timezone())
sleep$date <- format(sleep$SleepDay, format = "%m/%d/%y")
steps$ActivityDay <- as.Date(steps$ActivityDay, "%m/%d/%Y")
```

Checking for duplicates

```
sum(duplicated(activity))

## [1] 0

sum(duplicated(calories))

## [1] 0

sum(duplicated(intensities))

## [1] 0

sum(duplicated(steps))

## [1] 0

sum(duplicated(sleep))

## [1] 3
```

Remove duplicates

```
sleep <- sleep %>%
  distinct() %>%
  drop_na()
```

Confirm duplicates have been removed

```
sum(duplicated(sleep))

## [1] 0
```

Summarizing datasets

```
# activity
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
```

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

```
##     Calories
##  Min.   :   0
##  1st Qu.:1828
##  Median :2134
##  Mean   :2304
##  3rd Qu.:2793
##  Max.   :4900
```

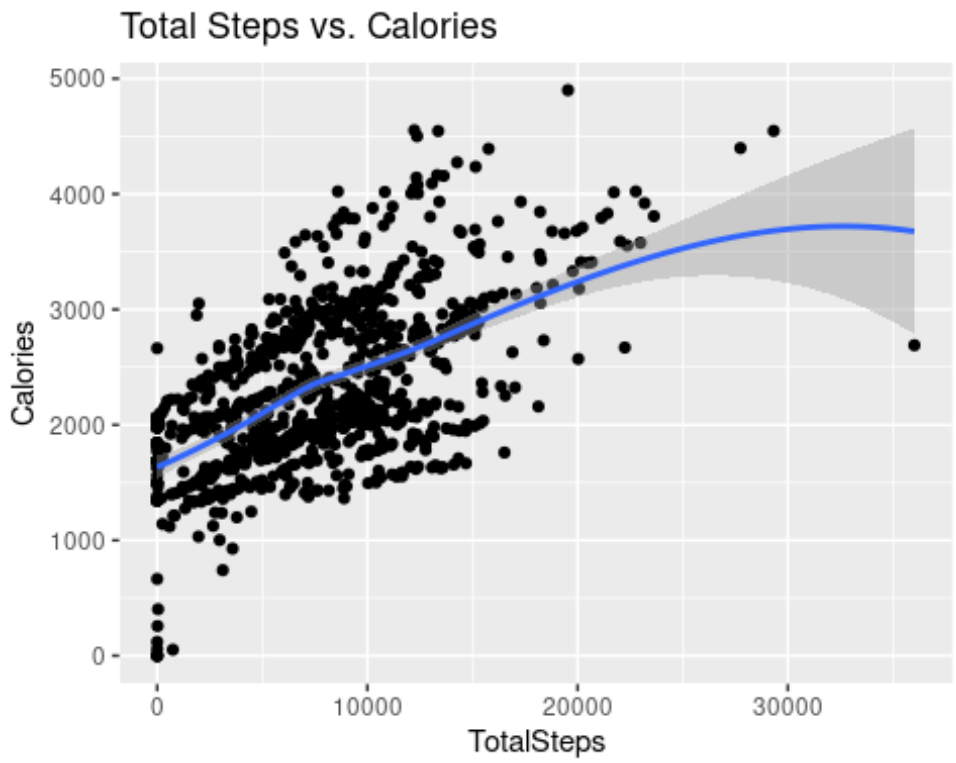
```
#sleep
  sleep %>%
  select(TotalMinutesAsleep, TotalTimeInBed) %>%
  summary()
```

```
##  TotalMinutesAsleep TotalTimeInBed
##  Min.   : 58.0      Min.   : 61.0
##  1st Qu.:361.0      1st Qu.:403.8
##  Median :432.5      Median :463.0
##  Mean   :419.2      Mean   :458.5
##  3rd Qu.:490.0      3rd Qu.:526.0
##  Max.   :796.0      Max.   :961.0
```

Visualization

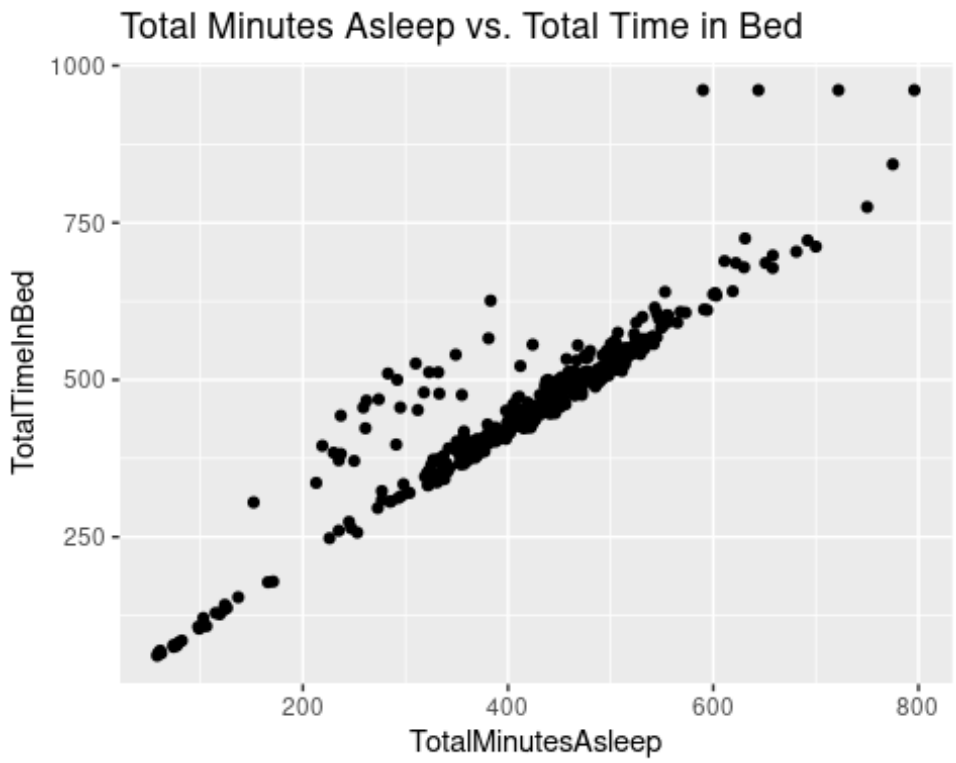
```
ggplot(data=activity, aes(x=TotalSteps, y=Calories)) +
  geom_point() + geom_smooth() + labs (title="Total Steps vs. Calories")
```

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



There is visibly a positive correlation between Total steps and Calories which shows us that more steps taken every day will lead to more calories being burned

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



This graphs shows a positive linear correlation. Hence, the more time they spend in bed, the likelihood they fall asleep

Grouping users by steps taken

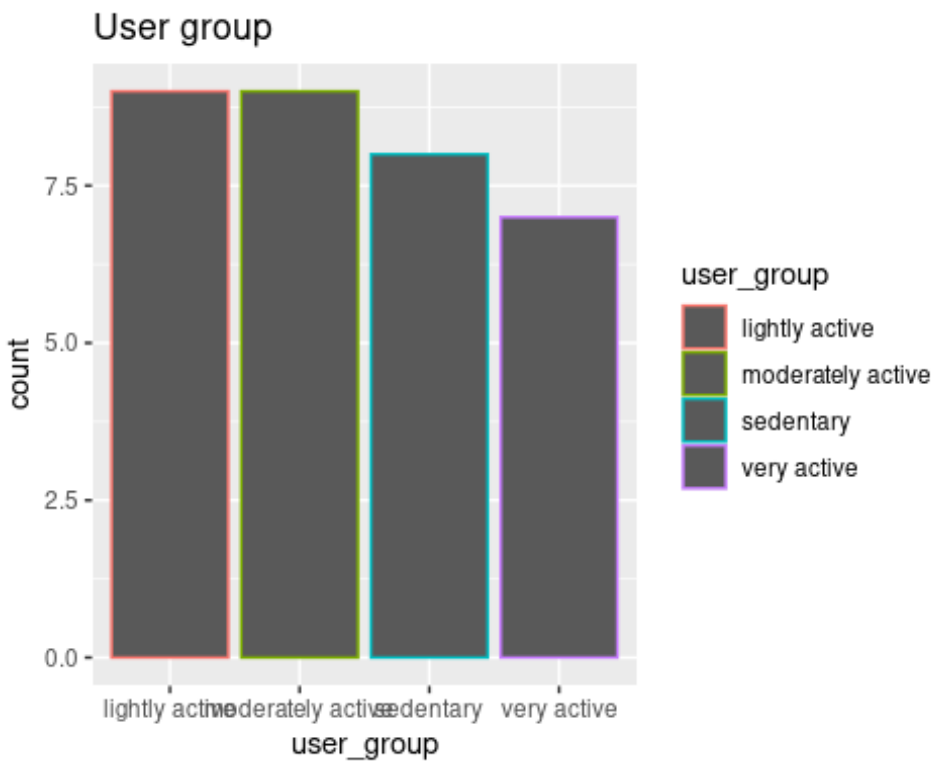
```
daily_average_steps <- activity %>%
  group_by(Id) %>%
  summarise (mean_steps = mean(TotalSteps))
head(daily_average_steps)
```

```
## # A tibble: 6 × 2
##           Id mean_steps
##        <dbl>      <dbl>
## 1 1503960366     12117.
## 2 1624580081      5744.
## 3 1644430081      7283.
## 4 1844505072      2580.
## 5 1927972279       916.
## 6 2022484408     11371.
```

```
user_group <- daily_average_steps %>%
  mutate(user_group = case_when(
    mean_steps < 5000 ~ "sedentary",
    mean_steps >= 5000 & mean_steps <= 7499 ~ "lightly active",
    mean_steps >= 7500 & mean_steps <= 9999 ~ "moderately active",
    mean_steps >= 10000 ~ "very active"
  ))
head(user_group)
```

```
## # A tibble: 6 × 3
##           Id mean_steps user_group
##        <dbl>      <dbl> <chr>
## 1 1503960366     12117. very active
## 2 1624580081      5744. lightly active
## 3 1644430081      7283. lightly active
## 4 1844505072      2580. sedentary
## 5 1927972279       916. sedentary
## 6 2022484408     11371. very active
```

```
ggplot(data=user_group)+
  geom_bar(mapping = aes(user_group, color=user_group))+
  labs(title="User group")
```



Conclusion and Recommendation

1. Majority of users are lightyly active and moderately active.
2. Sedentary minutes was approx 991 minutes which is too high for someone aiming for fitness.
3. Many of the users on the average sleep less than six hours which isn't the best but should attain at least seven hours of sleep or more.
4. There should be a notification to alert users to get moving and monitor their sleep patterns.
5. There should be a monitor to help users keep track of their weight and maintain healthy BMI.

6. Notifications should be put in place to alert user to workout for at least twenty(20) minutes each day and to take water to stay hydrated.
7. In summary,the fitness tracker application needs a bit of revamping by programmers which will help achieve the above goals and help users achieve a more healthy lifestyle and thus stand out in the competitive market.

This is my first project using R. Comments and recommendations are highly appreciated.