

Case Study: How Can a Wellness Technology Company Play It Smart?

Viviana Bhandari

2022-10-31

Introduction

This is the case study: How Can a Wellness Technology Company Play It Smart? from the Google Data Analytics Certificate course from Cloudera.

Business Task

Analyze smart device fitness data to help unlock new growth opportunities for the company Bellabeat.

Goal

Select one Bellabeat product and apply the insights found. The questions I will focus on my analysis are:

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?

Data Sources Used

- Google spreadsheet: to change the date format to datetime
- Google Cloud Sandbox: to run some queries and check data
- RStudio: Documentation with Visuals for final presentation (R Markdown)
- Kaggle data set FitBit Fitness Tracker Data (<https://www.kaggle.com/datasets/arashnic/fitbit>) (CC0: Public Domain, dataset made available through Mobius (<https://www.kaggle.com/arashnic>))

Key Stakeholders

- Urška Sršen: Bellabeat's cofounder and Chief Creative Officer
- Sando Mur: Mathematician and Bellabeat's cofounder; key member of the Bellabeat executive team

Prepare

1. Install & Load packages

```
install.packages("readr")
```

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

```
library(readr)

install.packages("dplyr")
```

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

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

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

```
install.packages('tidyverse')
```

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

```
library(tidyverse)
```

```
## — Attaching packages
## —————
## tidyverse 1.3.2 —
```

```
## ✓ ggplot2 3.3.6      ✓ purrr 0.3.4
## ✓ tibble 3.1.8       ✓ stringr 1.4.1
## ✓ tidyr 1.2.1        ✓ forcats 0.5.2
## — Conflicts ————— tidyverse_conflicts() —
## * dplyr::filter() masks stats::filter()
## * dplyr::lag()     masks stats::lag()
```

2. Create Data frames

```
dailySteps <- read_csv("Fitabase Data 4.12.16-5.12.16/dailySteps_merged.csv", show_col_types = FALSE)
dailyIntensities <- read_csv("Fitabase Data 4.12.16-5.12.16/dailyIntensities_merged.csv", show_col_types = FALSE)
dailyCalories <- read_csv("Fitabase Data 4.12.16-5.12.16/dailyCalories_merged.csv", show_col_types = FALSE)
```

Observations

1. The data collected was between April 12th, 2016 and May 9th, 2016

- Data is outdated
- Short period of time

```
min(dailySteps$ActivityDay)
```

```
## [1] "4/12/2016"
```

```
max(dailySteps$ActivityDay)
```

```
## [1] "5/9/2016"
```

```
min(dailyIntensities$ActivityDay)
```

```
## [1] "4/12/2016"
```

```
max(dailyIntensities$ActivityDay)
```

```
## [1] "5/9/2016"
```

```
min(dailyCalories$ActivityDay)
```

```
## [1] "4/12/2016"
```

```
max(dailyCalories$ActivityDay)
```

```
## [1] "5/9/2016"
```

2. This data may be a mixed between women and men since there is no column representing genre. For the case study exercise purpose, I will assume it's only related to women given that that is the audience for Bellabeat.

3. There are no null values in any file

```
is.null(dailySteps)
```

```
## [1] FALSE
```

```
is.null(dailyIntensities)
```

```
## [1] FALSE
```

```
is.null(dailyCalories)
```

```
## [1] FALSE
```

4. There are only 33 participants, and this may not be representative of the population

```
n_distinct(dailyCalories$Id)
```

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

```
n_distinct(dailySteps$Id)
```

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

```
n_distinct(dailyIntensities$Id)
```

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

Exploring a few key tables

- dailySteps data.

```
head(dailySteps)
```

```
## # A tibble: 6 × 3
##       Id ActivityDay StepTotal
##   <dbl> <chr>         <dbl>
## 1 1503960366 4/12/2016      13162
## 2 1503960366 4/13/2016      10735
## 3 1503960366 4/14/2016      10460
## 4 1503960366 4/15/2016       9762
## 5 1503960366 4/16/2016     12669
## 6 1503960366 4/17/2016       9705
```

- Identify all the columns in the dailySteps data.

```
colnames(dailySteps)
```

```
## [1] "Id" "ActivityDay" "StepTotal"
```

- dailyIntensities data.

```
head(dailyIntensities)
```

```
## # A tibble: 6 × 10
##       Id Activ...1 Seden...2 Light...3 Fairl...4 VeryA...5 Seden...6 Light...7 Moder...8 VeryA...9
##   <dbl> <chr>      <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 1.50e9 4/12/2...    728     328     13     25       0     6.06    0.550    1.88
## 2 1.50e9 4/13/2...    776     217     19     21       0     4.71    0.690    1.57
## 3 1.50e9 4/14/2...   1218     181     11     30       0     3.91    0.400    2.44
## 4 1.50e9 4/15/2...    726     209     34     29       0     2.83    1.26     2.14
## 5 1.50e9 4/16/2...    773     221     10     36       0     5.04    0.410    2.71
## 6 1.50e9 4/17/2...    539     164     20     38       0     2.51    0.780    3.19
## # ... with abbreviated variable names 1ActivityDay, 2SedentaryMinutes,
## # 3LightlyActiveMinutes, 4FairlyActiveMinutes, 5VeryActiveMinutes,
## # 6SedentaryActiveDistance, 7LightActiveDistance, 8ModeratelyActiveDistance,
## # 9VeryActiveDistance
```

- Identify all the columns in the dailyIntensities data.

```
colnames(dailyIntensities)
```

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

- dailyCalories data.

```
head(dailyCalories)
```

```
## # A tibble: 6 × 3
##       Id ActivityDay Calories
##   <dbl> <chr>      <dbl>
## 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
```

- Identify all the columns in the dailyCalories data.

```
colnames(dailyCalories)
```

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

Visuals & Analysis

Merging datasets & taking a sample of the data

```
df_list <- list(dailySteps, dailyIntensities, dailyCalories)
```

```
set.seed(500)
```

```
df_list %>% reduce(full_join, by='Id')
```

```
## # A tibble: 833,248 × 14
##       Id Activity...1 StepT...2 Activ...3 Seden...4 Light...5 Fairl...6 VeryA...7 Seden...8
##       <dbl> <chr>         <dbl> <chr>         <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 1503960366 4/12/2016      13162 4/12/2...      728    328    13    25    0
## 2 1503960366 4/12/2016      13162 4/12/2...      728    328    13    25    0
## 3 1503960366 4/12/2016      13162 4/12/2...      728    328    13    25    0
## 4 1503960366 4/12/2016      13162 4/12/2...      728    328    13    25    0
## 5 1503960366 4/12/2016      13162 4/12/2...      728    328    13    25    0
## 6 1503960366 4/12/2016      13162 4/12/2...      728    328    13    25    0
## 7 1503960366 4/12/2016      13162 4/12/2...      728    328    13    25    0
## 8 1503960366 4/12/2016      13162 4/12/2...      728    328    13    25    0
## 9 1503960366 4/12/2016      13162 4/12/2...      728    328    13    25    0
## 10 1503960366 4/12/2016      13162 4/12/2...      728    328    13    25    0
## # ... with 833,238 more rows, 5 more variables: LightActiveDistance <dbl>,
## # ModeratelyActiveDistance <dbl>, VeryActiveDistance <dbl>,
## # ActivityDay <chr>, Calories <dbl>, and abbreviated variable names
## # 1ActivityDay.x, 2StepTotal, 3ActivityDay.y, 4SedentaryMinutes,
## # 5LightlyActiveMinutes, 6FairlyActiveMinutes, 7VeryActiveMinutes,
## # 8SedentaryActiveDistance
```

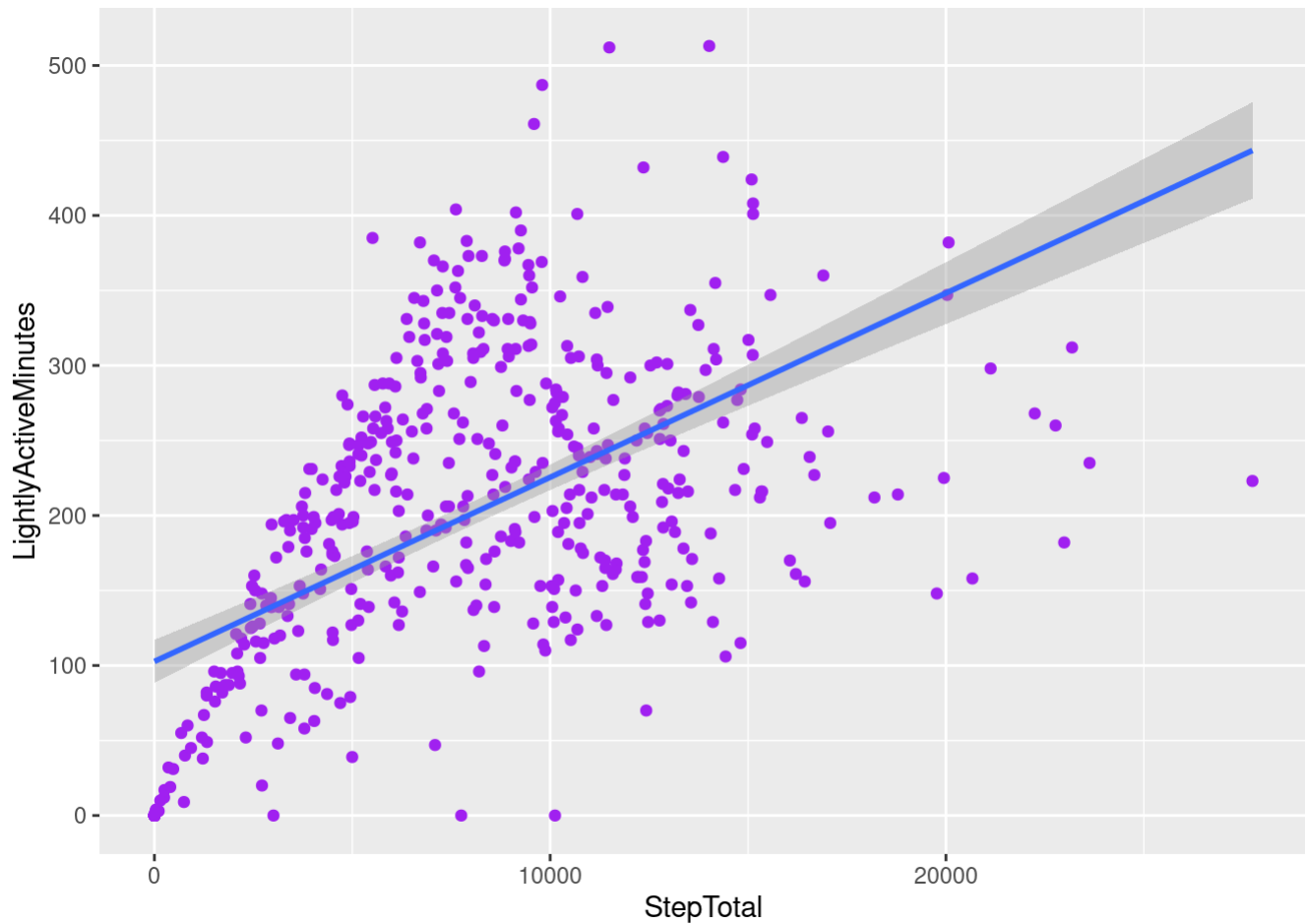
```
sample_data <- data.frame(df_list)
```

```
df_combinedStepsIntensitiesCalories<- sample_n(sample_data, 500)
```

According to WebMd article: (<https://www.webmd.com/a-to-z-guides/news/20221005/turns-out-there-is-something-special-about-ten-thousand-steps-a-day#:~:text=A%20study%20in%20JAMA%20Neurology,miles%2C%20depending%20on%20your%20stride.>) “Turns Out There Is Something Special About 10,000 Steps a Day”, walking about 10,000 steps a day can decrease cardiovascular disease (heart disease, stroke and heart failure), 13 types of cancer, and dementia.

1. There is positive correlation between total # of steps and Light active minutes.

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

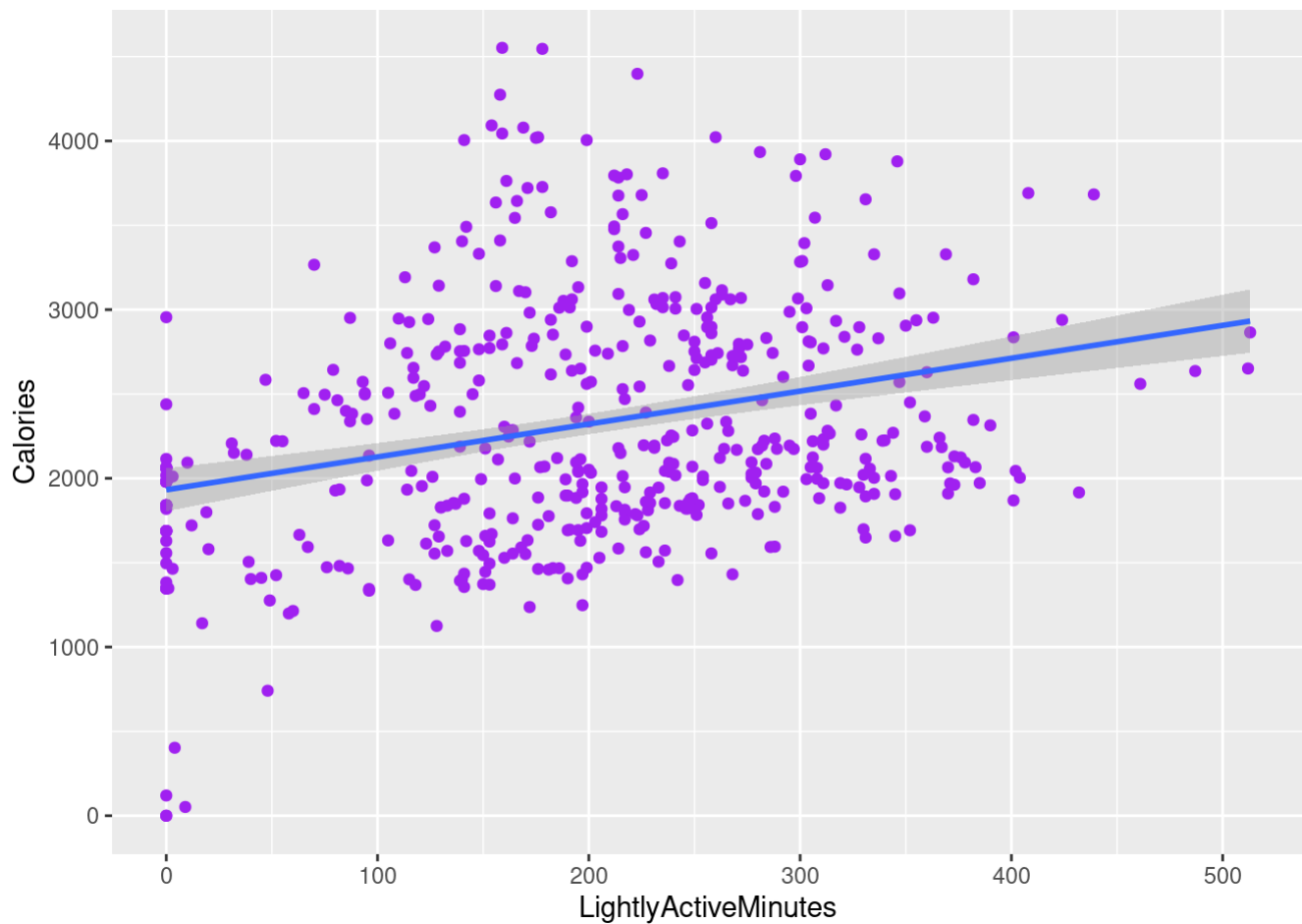


According to the Medical News Today article: “Factors that influence calorie burning” (<https://www.medicalnewstoday.com/articles/what-exercise-burns-the-most-calories#factors>), the more intense an activity is, the more calories it burns.

Intensity & Calories Correlation

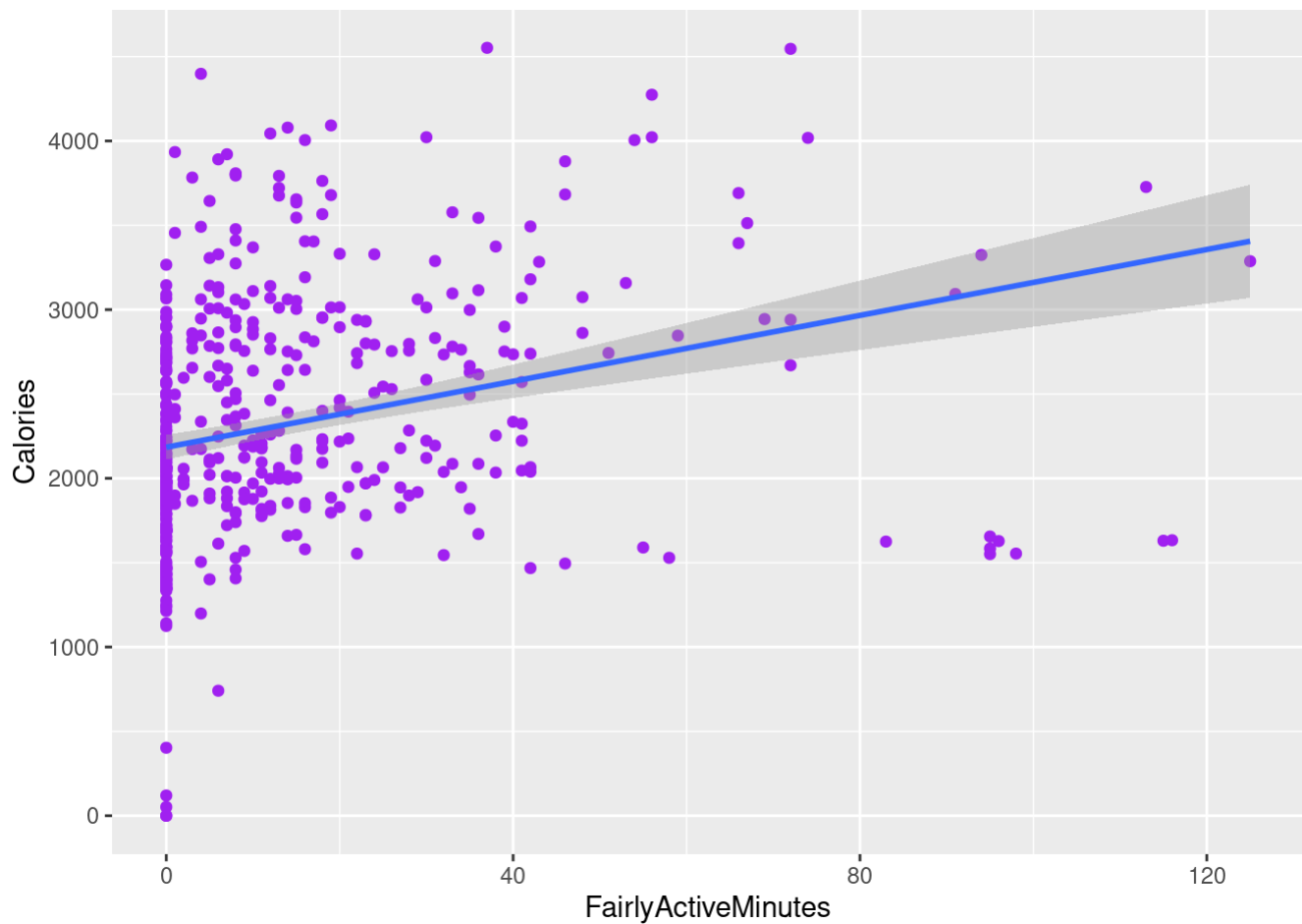
1. There is a positive correlation between light intensity activities and burning calories. However, most users are spending more than 1hr to burn high amounts of calories.

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



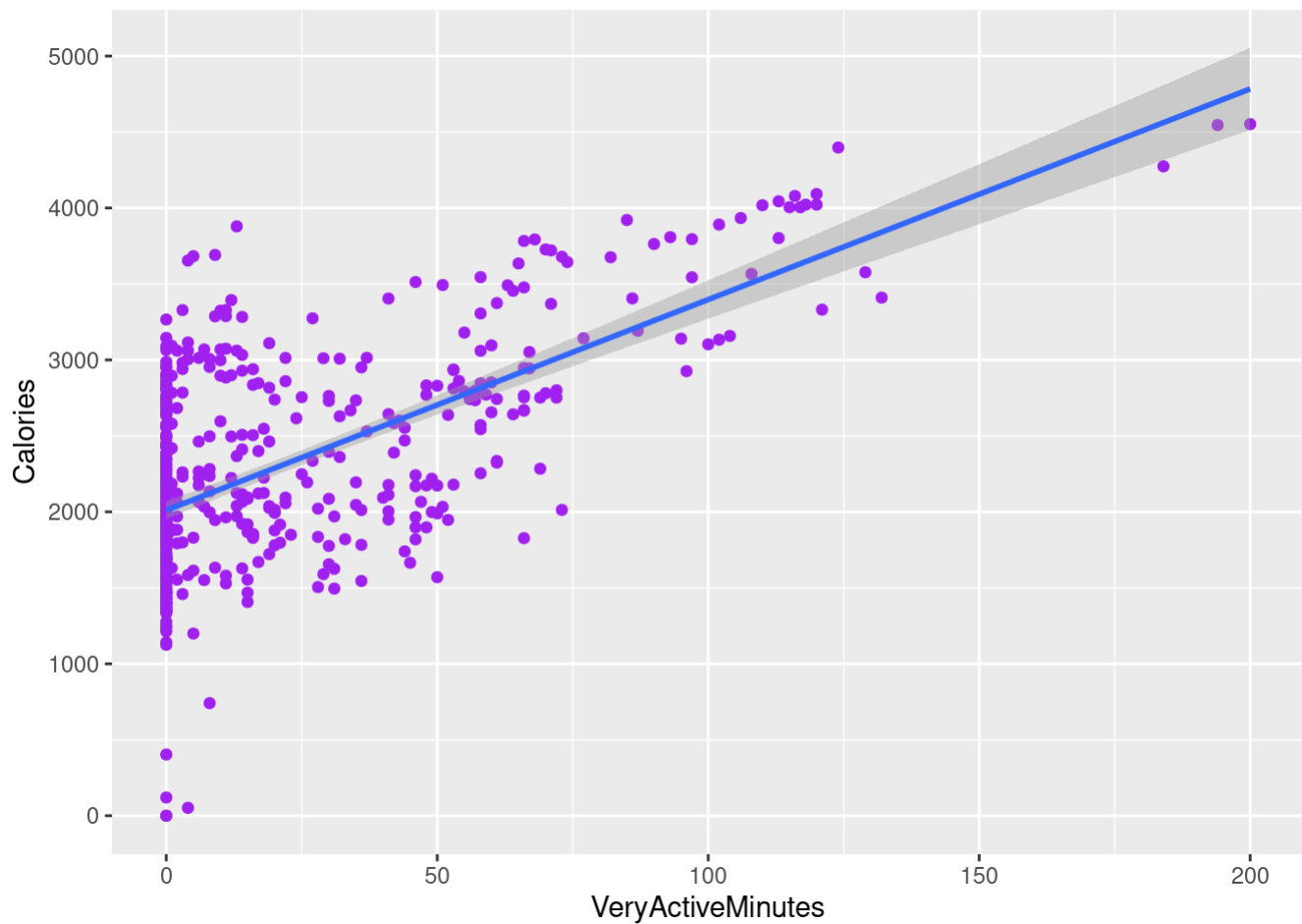
2. The correlation between fairly active intensity and calories is positive and users are spending around 40 minutes to burn high amount of calories.

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

3. The correlation between very active intensity and calories is the most positive and users can spend less than 30 min to burn a high amount of calories.

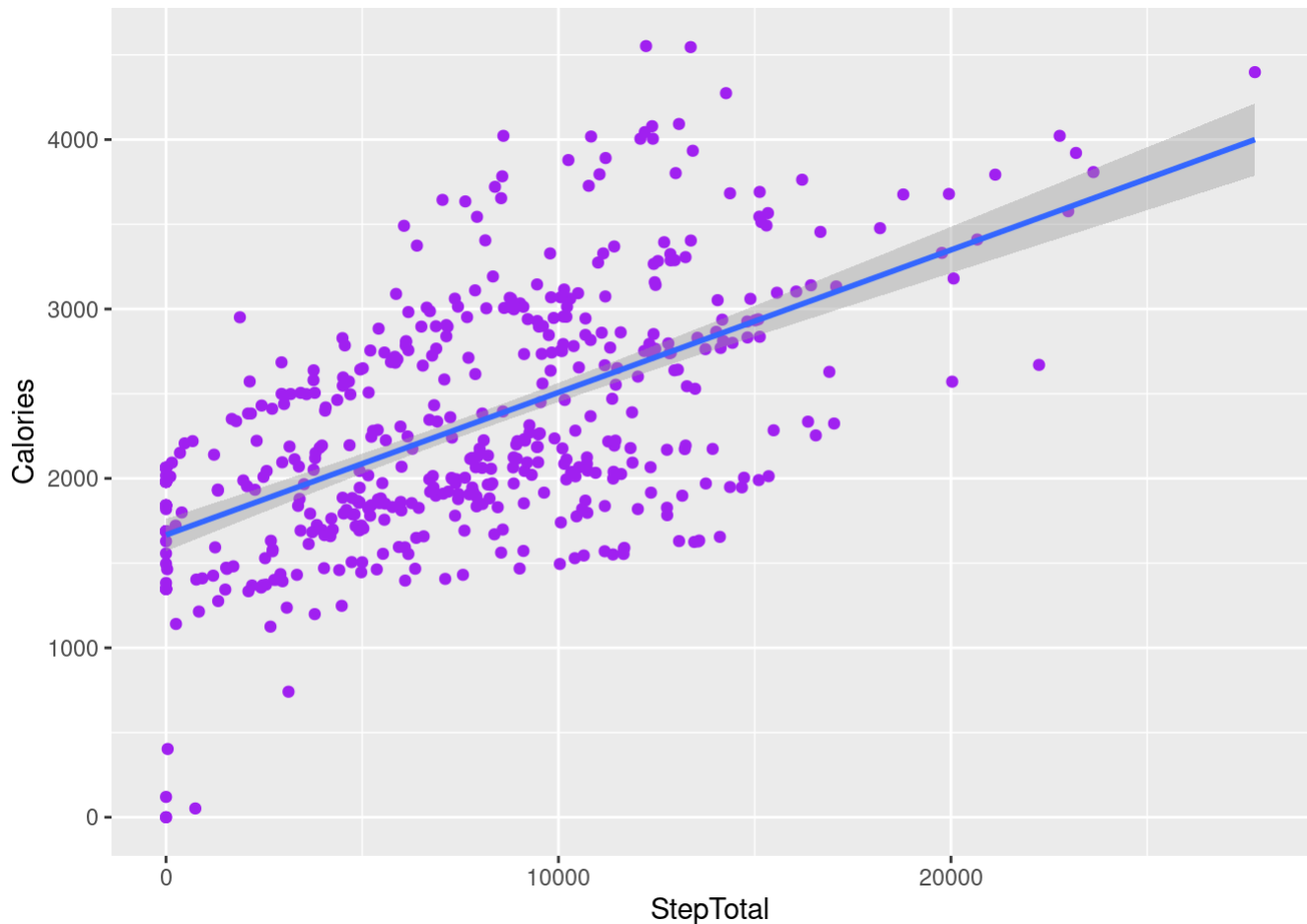
```
## `geom_smooth()` using formula 'y ~ x'
```



Steps & Calories correlation

1. The correlation between the total of steps and calories burned is positive. Less than 10000 steps can help burning some calories

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



Recommendations

Based on the data analyzed, the marketing strategy can be around 2 groups:

- a. Weight lost: the audience in this group are looking to lose weight and may be more likely to add or increment their activity intensity, therefore, Bellabeat can provide a weekly exercise routine with activities that are rated between lightly, fairly and very intense and show the amount of calories that could be burned with each exercise.
- b. Heart health: the audience in this category want to keep track of their heart condition. The marketing strategy can focus on steps. Bellabeat can introduce a routine where users can select the number of steps they want to start with per day, and the app will increment the steps daily until they reach the recommended 10000 steps per day.