

Google Capstone Project: How Can Bellabeat, A Wellness Technology Company Play It Smart?

Step 1: ASK

Bellabeat is a high-tech manufacturer of health-focused products for women, creating awareness for their health and habits. Bellabeat has grown rapidly in recent years and has recognized itself as a tech-driven wellness company for females.

The co-founder and Chief Creative Officer, Urška Sršen is confident that an analysis of non-Bellebeat consumer data (i.e. Fitbit fitness tracker usage data) and analyzing smart device fitness data could help unlock new growth opportunities for the company which will help in improving their user incredibility and to become a larger player in the global smart device market.

Business Task

Analyse consumers' Fitbit fitness tracker usage data to gain insights and trends for improving Bellabeat's marketing strategy.

Objective

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?

Deliverables

1. A clear summary of the business task
2. A description of all data sources used
3. Documentation of any cleaning or manipulation of data
4. A summary of your analysis

5. Supporting visualizations and key findings

6. Your top high-level content recommendations based on your analysis

Key Stakeholders

- Urška Sršen: Bellabeat's co-founder and Chief Creative Officer
- Sando Mur: Mathematician and Bellabeat's cofounder, a key member of the Bellabeat executive team
- Bellabeat marketing analytics team: A team of data analysts responsible for collecting, analyzing, and reporting data that helps guide Bellabeat's marketing strategy.

Step 2: PREPARE

Information on data source

Public data that explores smart device users' daily habits is used. This Kaggle data set contains a personal fitness tracker from thirty Fitbit users. It includes information about daily activity, steps, and heart rate that can be used to explore users' habits. It is stored in 18 CSV files.

Credibility of data

A good data source is ROOCCC which stands for **R**eliable, **O**riginal, **C**omprehensive, **C**urrent, and **C**ited.

- **Reliable:** Not reliable because this data only contains 30 individuals which is not a representative sample for the overall population of Fitbit users.
- **Original:** Not original as the dataset was generated by a third party i.e., respondents to a distributed survey via Amazon Mechanical Turk.
- **Comprehensive:** Comprehensive as parameters match most of Bellabeat products' parameters but having more data from more individuals would help with the overall comprehensiveness.
- **Current:** Not current as the data was collected 7 years ago and may not be relevant.
- **Cited:** Not cited as it is collected by a third party and the source may or may not be reliable.

Overall, the conclusion drawn is that dataset is of bad quality and not reliable. It is not recommended to produce business recommendations based on this data. However, the general insights could still prove to be useful.

Tool

I am using R for data cleaning, transformation, and visualization.

Loading Packages

```
library(tidyverse)
library(ggplot2)
library(dplyr)
library(readr)
library(tidyr)
library(skimr)
library(scales)
library(janitor)
```

Importing Datasets

```
daily_activity<- read.csv("C:\\Users\\user\\Desktop\\Fitabase Data\\
dailyActivity_Merged.csv")

weight<- read.csv("C:\\Users\\user\\Desktop\\Fitabase Data\\
weightLogInfo_merged.csv")

daily_sleep<- read.csv("C:\\Users\\user\\Desktop\\Fitabase Data\\
sleepDay_merged.csv")
```

Verifying the data

| | Id | ActivityDate | TotalSteps | TotalDistance | TrackerDistance | LoggedActivitiesDistance | |
|---|--------------------|--------------|--------------------------|----------------------|---------------------|--------------------------|----------|
| 1 | 1503960366 | 4/12/2016 | 13162 | 8.50 | 8.50 | | 0 |
| 2 | 1503960366 | 4/13/2016 | 10735 | 6.97 | 6.97 | | 0 |
| 3 | 1503960366 | 4/14/2016 | 10460 | 6.74 | 6.74 | | 0 |
| 4 | 1503960366 | 4/15/2016 | 9762 | 6.28 | 6.28 | | 0 |
| 5 | 1503960366 | 4/16/2016 | 12669 | 8.16 | 8.16 | | 0 |
| 6 | 1503960366 | 4/17/2016 | 9705 | 6.48 | 6.48 | | 0 |
| | VeryActiveDistance | | ModeratelyActiveDistance | | LightActiveDistance | SedentaryActiveDistance | |
| 1 | | 1.88 | | 0.55 | 6.06 | | 0 |
| 2 | | 1.57 | | 0.69 | 4.71 | | 0 |
| 3 | | 2.44 | | 0.40 | 3.91 | | 0 |
| 4 | | 2.14 | | 1.26 | 2.83 | | 0 |
| 5 | | 2.71 | | 0.41 | 5.04 | | 0 |
| 6 | | 3.19 | | 0.78 | 2.51 | | 0 |
| | VeryActiveMinutes | | FairlyActiveMinutes | LightlyActiveMinutes | | SedentaryMinutes | Calories |
| 1 | | 25 | 13 | | 328 | 728 | 1985 |
| 2 | | 21 | 19 | | 217 | 776 | 1797 |
| 3 | | 30 | 11 | | 181 | 1218 | 1776 |
| 4 | | 29 | 34 | | 209 | 726 | 1745 |
| 5 | | 36 | 10 | | 221 | 773 | 1863 |
| 6 | | 38 | 20 | | 164 | 539 | 1728 |

| | Id | Date | weightKg | weightPounds | Fat | BMI | IsManualReport | LogId |
|---|------------|-----------------------|----------|--------------|-----|-------|----------------|--------------|
| 1 | 1503960366 | 5/2/2016 11:59:59 PM | 52.6 | 115.9631 | 22 | 22.65 | True | 1.462234e+12 |
| 2 | 1503960366 | 5/3/2016 11:59:59 PM | 52.6 | 115.9631 | NA | 22.65 | True | 1.462320e+12 |
| 3 | 1927972279 | 4/13/2016 1:08:52 AM | 133.5 | 294.3171 | NA | 47.54 | False | 1.460510e+12 |
| 4 | 2873212765 | 4/21/2016 11:59:59 PM | 56.7 | 125.0021 | NA | 21.45 | True | 1.461283e+12 |
| 5 | 2873212765 | 5/12/2016 11:59:59 PM | 57.3 | 126.3249 | NA | 21.69 | True | 1.463098e+12 |
| 6 | 4319703577 | 4/17/2016 11:59:59 PM | 72.4 | 159.6147 | 25 | 27.45 | True | 1.460938e+12 |

| | Id | SleepDay | TotalSleepRecords | TotalMinutesAsleep | TotalTimeInBed |
|---|------------|-----------------------|-------------------|--------------------|----------------|
| 1 | 1503960366 | 4/12/2016 12:00:00 AM | 1 | 327 | 346 |
| 2 | 1503960366 | 4/13/2016 12:00:00 AM | 2 | 384 | 407 |
| 3 | 1503960366 | 4/15/2016 12:00:00 AM | 1 | 412 | 442 |
| 4 | 1503960366 | 4/16/2016 12:00:00 AM | 2 | 340 | 367 |
| 5 | 1503960366 | 4/17/2016 12:00:00 AM | 1 | 700 | 712 |
| 6 | 1503960366 | 4/19/2016 12:00:00 AM | 1 | 304 | 320 |

```

Rows: 940
Columns: 15
$ Id                <dbl> 1503960366, 1503960366, 1503960366, 1503960366, 15
03960366, 15...
$ ActivityDate      <chr> "4/12/2016", "4/13/2016", "4/14/2016", "4/15/2016"
, "4/16/2016..."
$ TotalSteps        <int> 13162, 10735, 10460, 9762, 12669, 9705, 13019, 155
06, 10544, 9...
$ TotalDistance     <dbl> 8.50, 6.97, 6.74, 6.28, 8.16, 6.48, 8.59, 9.88, 6.
68, 6.34, 8...
$ TrackerDistance   <dbl> 8.50, 6.97, 6.74, 6.28, 8.16, 6.48, 8.59, 9.88, 6.
68, 6.34, 8...
$ LoggedActivitiesDistance <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0...

```

```

$ VeryActiveDistance      <dbl> 1.88, 1.57, 2.44, 2.14, 2.71, 3.19, 3.25, 3.53, 1.
96, 1.34, 4....
$ ModeratelyActiveDistance <dbl> 0.55, 0.69, 0.40, 1.26, 0.41, 0.78, 0.64, 1.32, 0.
48, 0.35, 1....
$ LightActiveDistance     <dbl> 6.06, 4.71, 3.91, 2.83, 5.04, 2.51, 4.71, 5.03, 4.
24, 4.65, 2....
$ SedentaryActiveDistance <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0,...
$ VeryActiveMinutes       <int> 25, 21, 30, 29, 36, 38, 42, 50, 28, 19, 66, 41, 39
, 73, 31, 78...
$ FairlyActiveMinutes     <int> 13, 19, 11, 34, 10, 20, 16, 31, 12, 8, 27, 21, 5,
14, 23, 11, ...
$ LightlyActiveMinutes    <int> 328, 217, 181, 209, 221, 164, 233, 264, 205, 211,
130, 262, 23...
$ SedentaryMinutes        <int> 728, 776, 1218, 726, 773, 539, 1149, 775, 818, 838
, 1217, 732,...
$ Calories                <int> 1985, 1797, 1776, 1745, 1863, 1728, 1921, 2035, 17
86, 1775, 18...

```

```
> glimpse(weight)
```

```

Rows: 67
Columns: 8
$ Id      <dbl> 1503960366, 1503960366, 1927972279, 2873212765, 2873212765,
4319703577, ...
$ Date    <chr> "5/2/2016 11:59:59 PM", "5/3/2016 11:59:59 PM", "4/13/2016 1
:08:52 AM", ...
$ WeightKg <dbl> 52.6, 52.6, 133.5, 56.7, 57.3, 72.4, 72.3, 69.7, 70.3, 69.9,
69.2, 69.1,...
$ WeightPounds <dbl> 115.9631, 115.9631, 294.3171, 125.0021, 126.3249, 159.6147,
159.3942, 15...
$ Fat      <int> 22, NA, NA, NA, NA, 25, NA, NA, NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, ...
$ BMI      <dbl> 22.65, 22.65, 47.54, 21.45, 21.69, 27.45, 27.38, 27.25, 27.4
6, 27.32, 27...
$ IsManualReport <chr> "True", "True", "False", "True", "True", "True", "True", "Tr
ue", "True", "...
$ LogId    <dbl> 1.462234e+12, 1.462320e+12, 1.460510e+12, 1.461283e+12, 1.46
3098e+12, 1....

```

```
> glimpse(daily_sleep)
```

```

Rows: 413
Columns: 5
$ Id      <dbl> 1503960366, 1503960366, 1503960366, 1503960366, 15039603
66, 15039603...
$ SleepDay <chr> "4/12/2016 12:00:00 AM", "4/13/2016 12:00:00 AM", "4/15/
2016 12:00:0...
$ TotalSleepRecords <int> 1, 2, 1, 2, 1, 1, 1, 1, 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, 277, 2
45, 366, 341...
$ TotalTimeInBed     <int> 346, 407, 442, 367, 712, 320, 377, 364, 384, 449, 323, 2
74, 393, 354...

```

There are some records where the step count is zero which means not everyone was consistent in tracking their data each day and some people do not wear it for the whole month.

```

> daily_activity_new<- daily_activity%>%
+   filter(TotalSteps!=0)
> view(daily_activity_new)

```

Removing the zero steps will be a better option to improve the overall analysis.

Separating the date and time

```
> weight_new<- weight%>%
+   separate(Date,into = c("Date","Time"),sep = " ")
Warning message:
Expected 2 pieces. Additional pieces discarded in 67 rows [1, 2, 3, 4, 5,
6, 7, 8, 9, 10, 11,
12, 13, 14, 15, 16, 17, 18, 19, 20, ...].

> daily_sleep_new<- daily_sleep%>%
+   separate(SleepDay,into = c("Date","Time"),sep = " ")
Warning message:
Expected 2 pieces. Additional pieces discarded in 413 rows [1, 2, 3, 4, 5,
6, 7, 8, 9, 10, 11,
12, 13, 14, 15, 16, 17, 18, 19, 20, ...].
```

Counting the distinct ID's in each dataset

```
> n_distinct(daily_activity_new$Id)
[1] 33

> n_distinct(daily_sleep_new$Id)
[1] 24

> n_distinct(weight_new$Id)
[1] 8
```

On analyzing the data we can infer that only 8 people entered their weight, only 24 people entered their sleep data, and also 33 people recorded their daily activity which contradicts the credibility of the data as the data citation says that there are only 30 people in the sample.

Now I will check if there are any duplicate records in the dataset.

```
> nrow(daily_activity_new)
[1] 863
> nrow(daily_sleep_new)
[1] 413
> nrow(weight_new)
[1] 6
> nrow(unique(daily_activity_new))
[1] 863
> nrow(unique(daily_sleep_new))
[1] 410
> nrow(unique(weight_new))
[1] 67
```

To obtain cleaner data duplicate rows should be removed from the sleep data.

```
> daily_sleep_unique <- unique(daily_sleep_new)
```

Now, the data is ready to be analyzed.

Step 4: ANALYSE

I will analyze, the data and identify the trends, relationships, and patterns in this process. It may include using statistical methods.

Pulling general statistics

```
> skim_without_charts(daily_activity_new)
```

```
— Data Summary —————
Name                               values
Number of rows                    863
Number of columns                  15

Column type frequency:
  character                        1
  numeric                         14

Group variables                    None

— Variable type: character —————
  skim_variable n_missing complete_rate min max empty n_unique whitespace
1 ActivityDate      0           1  8  9      0      31          0

— Variable type: numeric —————
  skim_variable      n_missing complete_rate      mean      sd
p0                p25
1 Id                0           1 4.86e+9 2.42e+9 150396
0366 2320127002
2 TotalSteps        0           1 8.32e+3 4.74e+3
4 4923
3 TotalDistance     0           1 5.98e+0 3.72e+0
0 3.37
4 TrackerDistance   0           1 5.96e+0 3.70e+0
0 3.37
5 LoggedActivitiesDistance 0           1 1.18e-1 6.46e-1
0 0
6 VeryActiveDistance 0           1 1.64e+0 2.74e+0
0 0
7 ModeratelyActiveDistance 0           1 6.18e-1 9.05e-1
0 0
8 LightActiveDistance 0           1 3.64e+0 1.86e+0
0 2.34
9 SedentaryActiveDistance 0           1 1.75e-3 7.65e-3
0 0
10 VeryActiveMinutes 0           1 2.30e+1 3.36e+1
0 0
11 FairlyActiveMinutes 0           1 1.48e+1 2.04e+1
0 0
12 LightlyActiveMinutes 0           1 2.10e+2 9.68e+1
0 146.
13 SedentaryMinutes 0           1 9.56e+2 2.80e+2
0 722.
14 calories         0           1 2.36e+3 7.03e+2
52 1856.
      p50      p75      p100
1 4.45e+9 6.96e+9 8.88e+9
2 8.05e+3 1.11e+4 3.60e+4
3 5.59e+0 7.90e+0 2.80e+1
4 5.59e+0 7.88e+0 2.80e+1
```

```

5 0      0      4.94e+0
6 4.10e-1 2.27e+0 2.19e+1
7 3.10e-1 8.65e-1 6.48e+0
8 3.58e+0 4.89e+0 1.07e+1
9 0      0      1.10e-1
10 7      e+0 3.5 e+1 2.1 e+2
11 8      e+0 2.1 e+1 1.43e+2
12 2.08e+2 2.72e+2 5.18e+2
13 1.02e+3 1.19e+3 1.44e+3
14 2.22e+3 2.83e+3 4.9 e+3

```

```
> skim_without_charts(daily_sleep_unique)
```

```

— Data Summary —
Name                               values
Number of rows                    410
Number of columns                  6
Column type frequency:
  character                        2
  numeric                         4
Group variables                    None

```

```
— variable type: character —
```

| | skim_variable | n_missing | complete_rate | min | max | empty | n_unique | whitespace |
|---|---------------|-----------|---------------|-----|-----|-------|----------|------------|
| 1 | Date | 0 | 1 | 8 | 9 | 0 | 31 | 0 |
| 2 | Time | 0 | 1 | 8 | 8 | 0 | 1 | 0 |

```
— variable type: numeric —
```

| | skim_variable | n_missing | complete_rate | mean | sd |
|-----|--------------------|------------|---------------|-------------|---------|
| p0 | | | | | |
| 1 | Id | 0 | 1 | 4994963041. | 2.06e+9 |
| 366 | 3977333714 | 4702921684 | | | |
| 2 | TotalSleepRecords | 0 | 1 | 1.12 | 3.47e-1 |
| 1 | | 1 | | | |
| 3 | TotalMinutesAsleep | 0 | 1 | 419. | 1.19e+2 |
| 58 | 361 | 432. | | | |
| 4 | TotalTimeInBed | 0 | 1 | 458. | 1.27e+2 |
| 61 | 404. | 463 | | | |
| | p75 | p100 | | | |
| 1 | 6962181067 | 8792009665 | | | |
| 2 | 1 | 3 | | | |
| 3 | 490 | 796 | | | |
| 4 | 526 | 961 | | | |

```
> skim_without_charts(weight_new)
```

```

— Data Summary —
Name                               values
Number of rows                    67
Number of columns                  9
Column type frequency:
  character                        3
  numeric                         6
Group variables                    None

```

```
— variable type: character —
```

| | skim_variable | n_missing | complete_rate | min | max | empty | n_unique | whitespace |
|---|---------------|-----------|---------------|-----|-----|-------|----------|------------|
| 1 | Date | 0 | 1 | 8 | 9 | 0 | 31 | 0 |
| 2 | Time | 0 | 1 | 7 | 8 | 0 | 26 | 0 |

| | 3 | IsManualReport | 0 | 1 | 4 | 5 | 0 | 2 | 0 |
|--------------------------|-----------|----------------|---------|-------------|---------|---------|---------|---------|---------|
| — Variable type: numeric | | | | | | | | | |
| skim_variable | n_missing | complete_rate | mean | sd | p0 | p25 | p50 | p75 | p100 |
| 1 Id | 0 | 1 | 7.01e 9 | 1950321944. | 1.50e 9 | 6.96e 9 | 6.96e 9 | 6.96e 9 | 6.96e 9 |
| 2 weightKg | 0 | 1 | 7.20e 1 | 13.9 | 5.26e 1 | 6.14e 1 | 6.25e 1 | 6.25e 1 | 6.25e 1 |
| 3 weightPounds | 0 | 1 | 1.59e 2 | 30.7 | 1.16e 2 | 1.35e 2 | 1.38e 2 | 1.38e 2 | 1.38e 2 |
| 4 Fat | 65 | 0.0299 | 2.35e 1 | 2.12 | 2.2 e 1 | 2.27e 1 | 2.35e 1 | 2.35e 1 | 2.35e 1 |
| 5 BMI | 0 | 1 | 2.52e 1 | 3.07 | 2.15e 1 | 2.40e 1 | 2.44e 1 | 2.44e 1 | 2.44e 1 |
| 6 LogId | 0 | 1 | 1.46e12 | 782994784. | 1.46e12 | 1.46e12 | 1.46e12 | 1.46e12 | 1.46e12 |

I would like to condense each file into only the columns that I want to use for
A more focused analysis.

```
> daily_activity_final <- daily_activity_new %>%
+   select(Id, ActivityDate, TotalSteps, VeryActiveMinutes, FairlyActiveMi
+   nutes, LightlyActiveMinutes, SedentaryMinutes, Calories) %>%
+   rename(Date = ActivityDate)
>
> weight_final <- weight_new %>%
+   select(Id, Date, BMI, WeightPounds, IsManualReport)
>
> daily_sleep_final <- daily_sleep_unique %>%
+   select(Id, Date, TotalMinutesAsleep, TotalTimeInBed)
```

Conclusions Drawn

- The average number of steps logged by users is 8320, which is inadequate. It is recommended to aim a goal of at least 10000 steps per day for weight loss and fitness improvement.
- The average minutes for Very Active is 23.02, for Fairly Active is 14.78, for Lightly Active is 210, and for Sedentary is 955.8.
- On average, users logged 419 minutes of sleep which is approximately 6.98 hours of sleep which is a good sleep time. It is recommended to sleep for 7-8 hours a day for a healthy mind.
- The average BMI is 25.2, but according to WHO recommendations, the healthy BMI range is 18.5-24.9 which means average users have the nutritional status of pre-obesity.

Note: Due to lack of information regarding the data outliers still exist in the data due to which the above values might be slightly skewed.

Step 5: SHARE

I will use visualizations to share my insights and important findings.

```
> VeryActiveMin <- sum(daily_activity_final$VeryActiveMinutes)
> FairlyActiveMin <- sum(daily_activity_final$FairlyActiveMinutes)
> LightlyActiveMin <- sum(daily_activity_final$LightlyActiveMinutes)
> SedentaryMin <- sum(daily_activity_final$SedentaryMinutes)
> TotalMin <- VeryActiveMin + FairlyActiveMin + LightlyActiveMin + Sedenta
ryMin
>
> slices <- c(VeryActiveMin, FairlyActiveMin, LightlyActiveMin, SedentaryMin)
```

```

> lbls <- c("VeryActive","FairlyActive","LightlyActive", "Sedentary")
> pct <- round(slices/sum(slices)*100)
> lbls <- paste(lbls, pct)
> lbls <- paste(lbls, "%", sep="")
> pie(slices, labels = lbls, col = rainbow(length(lbls)), main = "Percentage of Activity in Minutes")

```

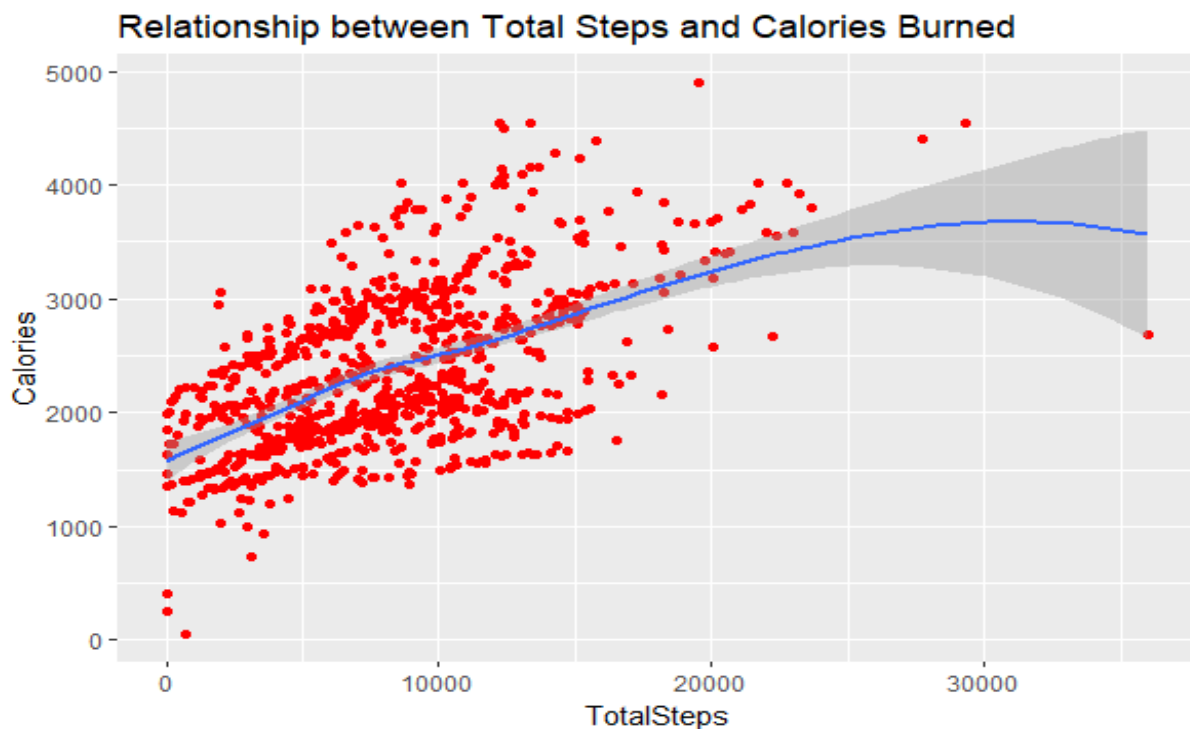


It is visually clear that average users have a sedentary lifestyle as 79% of the activity time is spent sedentary by the average users .

```

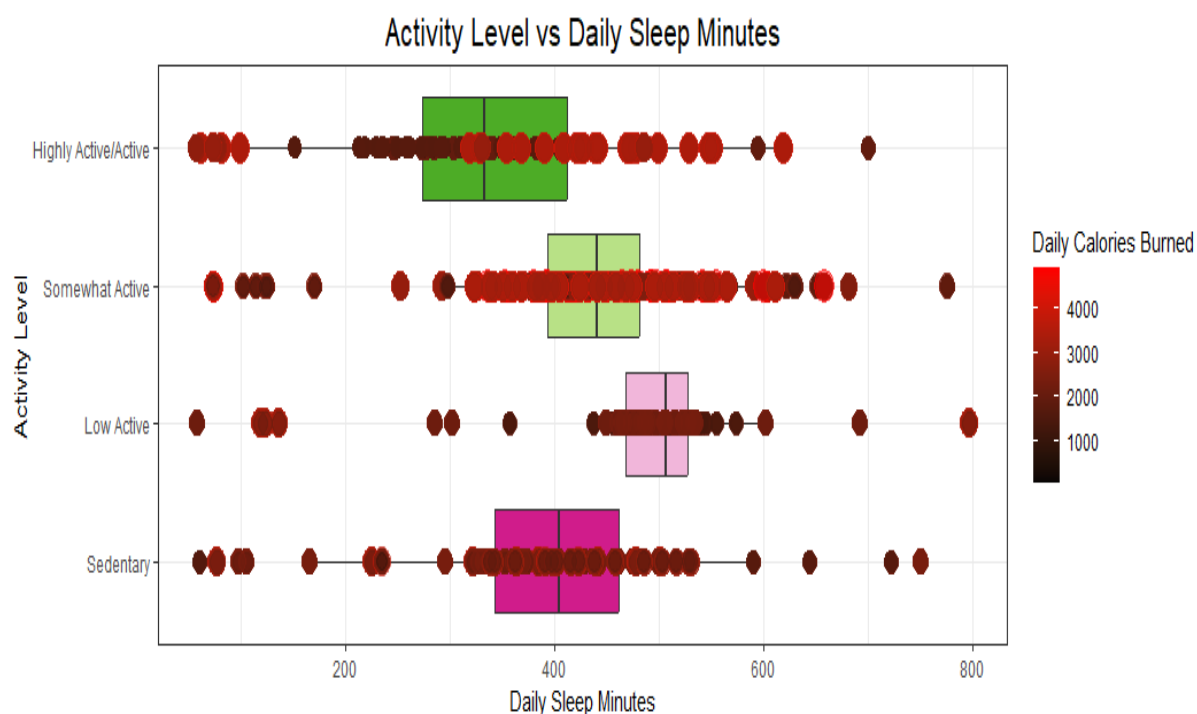
> ggplot(daily_acitivity_final)+ geom_point(mapping=aes(x=TotalSteps,y=Calories),color="Red")+ geom_smooth(mapping = aes(x=TotalSteps,y=Calories))+labs(title = "Relationship between Total Steps and Calories Burned")
`geom_smooth()` using method = 'loess' and formula = 'y ~ x'

```



We discovered that it is a positive correlation, and it is obvious that the more steps an individual takes more the calories are burnt. Without more information regarding the person's age, sex, and height, it would be impossible to say exactly how many calories the person needs to burn to lose weight at a healthy rate. However, they are not burning enough weight as the BMI and weight of the individuals who logged those values did not see an improvement over the month of data collection.

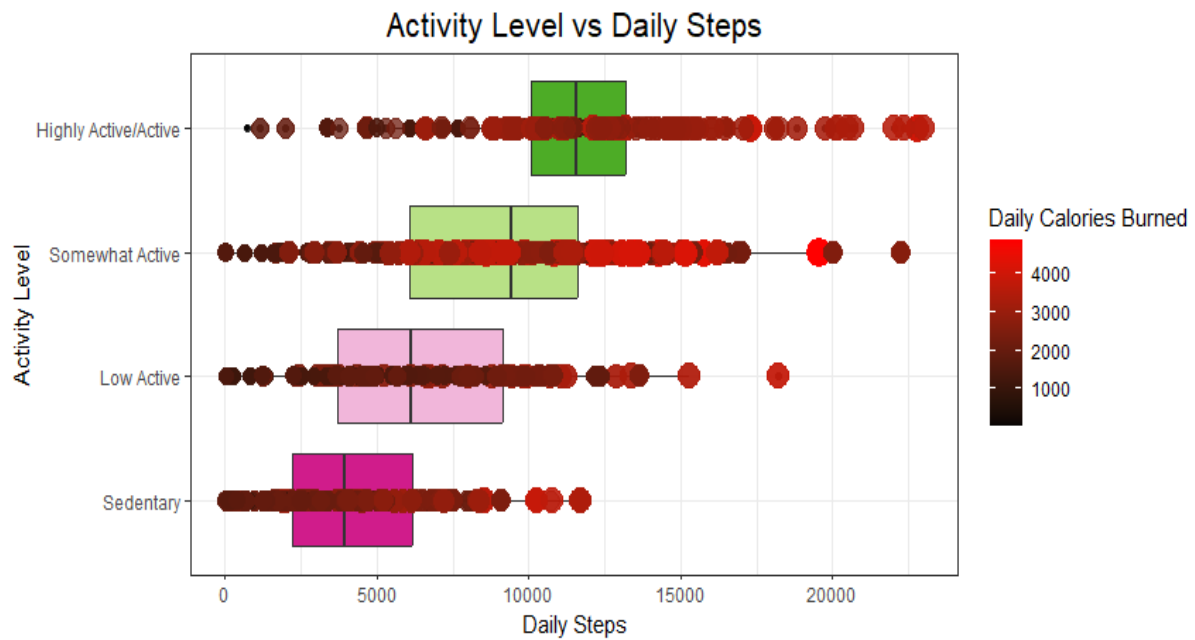
```
> combined_data <- merge(daily_activity_new, daily_sleep_unique, by=
c("Id"))
> combined_data$user_steps <- " "
>
> combined_data_grouped <- combined_data %>%
+   group_by (Id) %>%
+   summarize(average_totalsteps = mean(TotalSteps),
+             average_totalcalories = mean(Calories),
+             average_totaldistance = mean(TotalDistance),
+             average_minutesasleep = mean(TotalMinutesAsleep, na.rm
= TRUE)) %>%
+   mutate(user_steps = case_when(
+     average_totalsteps >= 10000 ~ "Highly Active/Active",
+     average_totalsteps >= 7500 & average_totalsteps < 10000 ~ "Som
ewhat Active",
+     average_totalsteps >= 5000 & average_totalsteps < 7500 ~ "Low
Active",
+     average_totalsteps < 5000 ~ "Sedentary"))
>
> combined_data <- subset(combined_data, select = -user_steps)
>
> combined_data_grouped <- merge(combined_data, combined_data_groupe
d, by= c("Id"))
>
> combined_data_grouped$user_steps <- factor(combined_data_grouped$u
ser_steps, levels = c("Sedentary", "Low Active", "Somewhat Active",
"Highly Active/Active"))
> ggplot(combined_data_grouped, aes(user_steps, TotalMinutesAsleep))
+
+   geom_boxplot(aes(fill= user_steps))+
+   geom_point(alpha = 0.5, aes(size = Calories, color = Calories))+
+   labs(title = "Activity Level vs Daily Sleep Minutes", x = "Activ
ity Level", y = "Daily Sleep Minutes", fill= "Activity Level", color
= "Daily Calories Burned", caption= "Data Source:
Physical activity for campus employees: a university worksite well
ness program")+
+   coord_flip()+
+   scale_fill_brewer(palette="PiYG")+
+   scale_color_gradient(low= "grey2", high= "red")+
+   theme_bw()+
+   theme(plot.title = element_text(hjust = 0.5, size = 16))+
+   theme(plot.caption = element_text(hjust = 1.75))+
+   guides(size = "none", fill = "none")
```



Physical activity for campus employees: a university worksite wellness

I assumed that the more sleep an individual had more active they would be but there is no significant correlation between activity level and daily sleep and surprisingly when people slept the most amount of minutes, they became more sedentary.

```
> ggplot(combined_data_grouped, aes(user_steps, TotalSteps))+
+   geom_boxplot(aes(fill= user_steps))+
+   geom_point(alpha = 0.5, aes(size = Calories, color = Calories))+
+   labs(title = "Activity Level vs Daily Steps", x = "Activity Level", y = "Daily Steps", fill= "Activity Level", size= "", color= "Daily Calories Burned", caption= "Data Source: Physical activity for campus employees: a university worksite wellness program")+
+   coord_flip()+
+   scale_fill_brewer(palette="PiYG")+
+   scale_color_gradient(low= "grey2", high= "red")+
+   theme_bw()+
+   theme(plot.title = element_text(hjust = 0.5, size = 16))+
+   theme(plot.caption = element_text(hjust = 1.75))+
+   guides(size = "none", fill = "none")
```



Data Source: Physical activity for campus employees: a university worksite wellness program

There is a high correlation between daily steps and activity level and also users in more active groups burn more calories per step.

Step 6: ACT

In the final step, we will be delivering our insights and providing recommendations based on our analysis.

Here, we revisit our business questions and share with you our high-level business recommendations.

What are the trends identified?

- The majority of users are using Fitbit to track sedentary activities and not using it for tracking their health habits.
- On average, the average Total Steps per day for the participating individuals was 8053, which is almost 2000 steps below the suggested minimum Total Steps per day.
- There is a high correlation between daily steps and activity level and users in more active groups burn more calories per step.
- There is no significant correlation between activity level and daily sleep and surprisingly when people slept the greatest number of minutes, they became more sedentary.
- The participating individuals did not lose weight, did not improve their BMI or sleep quality, and did not see any overall improvement in their activity levels.

How could these trends help influence Bellabeat's marketing strategy?

- Bellabeat marketing team can encourage users by educating and providing them with knowledge about fitness benefits, suggesting different types of exercise and calorie intake and burnt rate information on the Bellabeat app.
- Bellabeat should introduce a built-in calculator in its products, where the users can enter their details like sex, age, weight, height, and other health information to create precise results.
 - This calculator will notify the user of what their maintenance calories are (and their macros) and how much of a caloric deficit the user needs to be in each day to lose an X amount of lbs each week, based upon their weight goals and time frame.
 - The user would also be notified if they are reaching, have reached, or have passed their daily caloric intake.
- Bellabeat can introduce interesting schemes to motivate users to reach their desired goals. For instance, they can offer free membership to the top 100 most consistent users.
- Bellabeat marketing team can host campaigns and walkathons and introduce how is their product capable of tracking their activities to bring up more awareness towards health among people.