2023

bellabeat

BELLABEAT CASE STUDY

Marketing Analysis for new growth opportunities

Melissa V Aristizabal

BUSINESS TASK

ABOUT THE COMPANY

Bellabeat is a high-tech company that manufactures health-focused smart products. Collecting data on activity, sleep, stress, and reproductive health has allowed Bellabeat to empower women with knowledge about their own health and habits. Bellabeat has grown rapidly and quickly positioned itself as a tech-driven wellness company for women.

Key Stakeholders

• Urška Sršen Cofounder and Chief Creative Officer.
• Sando Mur Cofounder and key member of the 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.
• Customers Everyone who purchases their product or use Bellabeat's services.

Business task

To focus on one of Bellabeat's products and analyze smart device data to gain insights into how customers are using their smart devices. To identify trends in the Smart device usage, how this trends can be applied to Bellabeat customers and provide insights on a marketing strategy for growth opportunities based on these trends.

PREPARE

Dataset storage

Data Organization & verification

Licensing, privacy, security & accessibility

Information from the data

PREPARE

Dataset storage

The data source we are going to use is a FitBit fitness tracker data from a public domain stored in Kaggle. The data comes from Thirty eligible Fitbit users that gave the consent to the submission of their personal tracker data, including: minute-level output for physical activity, heart rate, and sleep monitoring.

The data found is about daily activity, steps, and heart rate that can be useful as it is used to explore users' habits.

Data Organization & verification

The available data set consists on 18 CSV documents. The csv documents are divided in different categories that represents quantitative data tracked by Fitbit. The dataset files are in long format as I found the categories are in the columns and data information is in the rows.

What all files have in common is User ID, date, time and general data about distance, and counts. This dataset doesn't have any demographic information and based on this, there's a possibility of sampling bias. The dataset is not current as it is from 2016 and is only 2 months of time length.

Licensing, privacy, security & accessibility

From the score calculated by Kaggle this dataset it has and accurate completeness, its update frequency is from 2 years. It is open to the public domain by waiving all of his or her rights to the work worldwide under copyright law, including all related and neighboring rights, to the extent allowed by law. You can distribute, copy and modify even for commercial purposes.

Information from the data

We can confirm the data matches the products. So this can be helpful to identify trends and customers device usage. There's an analysis that can be made in order to create the inisghts the stakeholders are looking for. The source is known from the data.

The data's integrity can be confirmed by checking the data comes from a third party source the quality of its content is not questionable and we can confirm the source of it.

PROCESS

Installing packages and opening libraries

Dataset preview

Format & Clean

How many users we have

Duplicates

Remove duplicates and N/A

Rename & format columns

Consistency of date and time

Merging Datasets

In this process phase I will use R_studio platform to analyze the data, clean it and create the visualizations.In this case study I will be using R_Studio as R can manage this dataset in an more practical way as it also generates clear and useful visualizations for my analysis.

Installing packages and opening libraries
First I will upload the following packages to do the analysis and create the visualizations:

```
tidyverse
here
skimr
janitor
lubridate
ggpubr
ggrepel
```

Dataset preview
We will preview our selected data frames and check the summary of each column.

```
hourly_steps <- read_csv("hourlySteps_merged.csv")
View(hourly_steps)
daily_activity <- read_csv("Daily_Activity_Merged.csv")
View(daily_activity)
daily_sleep <- read_csv("sleepDay_merged.csv")
View(daily_sleep)
```

Format & Clean
After uploading our dataset is important to clean it first in order to find any erros or inconsistencies before we analyze it.

How many users we have
We need to check first if the dataset have duplicates, null rows and verify how many unique users are in each dataframe.

```
head(daily_activity)
str(daily_activity)

head(daily_sleep)
str(daily_sleep)

head(hourly_steps)
str(hourly_steps)
n_unique(daily_activity$Id)
n_unique(daily_sleep$Id)
n_unique(hourly_steps$Id)

-33
-24
-33
```

Duplicates
Lets verify how many duplicate users are in the data frames.

```
sum(duplicated(daily_activity))
sum(duplicated(daily_sleep))
sum(duplicated(hourly_steps))
0
3
0
```

Remove duplicates and N/A
We found some duplicates in the dataframe of daily sleep with drop_na we can eliminate those duplicates.

```
daily_activity <- daily_activity %>%
  distinct() %>%
  drop_na()
daily_sleep <- daily_sleep %>%
  distinct() %>%
  drop_na()
hourly_steps <- hourly_steps %>%
  distinct() %>%
  drop_na()
```

We will verify that the duplicates have been removed

```
sum(duplicated(daily_sleep))
0
```

Rename & format columns
We have some columns that doesn’t have the right format titles. Since we are going to merge some columns later on we need to modify them.

```
clean_names(daily_activity)
daily_activity <- rename_with(daily_activity, tolower)

clean_names(daily_sleep)
daily_sleep <- rename_with(daily_sleep, tolower)

clean_names(hourly_steps)
hourly_steps <- rename_with(hourly_steps, tolower)
```

Consistency of date and time
Next, we will change format of date-time format for daily activity and daily sleep, to do this the function will be as_date and we will remove time.

```
daily_activity <- daily_activity %>%
  rename(date = activitydate) %>%
  mutate(date = as_date(date, format = "%m/%d/%Y"))

daily_sleep <- daily_sleep %>%
  rename(date = sleepday) %>%
  mutate(date = as_date(date,format ="%m/%d/%Y %I:%M:%S %p", tz=Sys.timezone()))
```

With _head function we can check the new dataframes
Last but not least we will change from string to date.

```
hourly_steps<- hourly_steps %>%
  rename(date_time = activityhour) %>%
  mutate(date_time = as.POSIXct(date_time,format ="%m/%d/%Y %I:%M:%S %p" , tz=Sys.timezone()))
head(hourly_steps)
```

Merging Datasets

To verify some correlations of the data we have I am going to merge the data frames daily_sleep & daily_activity.

```
daily_activity_sleep <- merge(daily_activity, daily_sleep, by=c ("id", "date"))
glimpse(daily_activity_sleep)
```

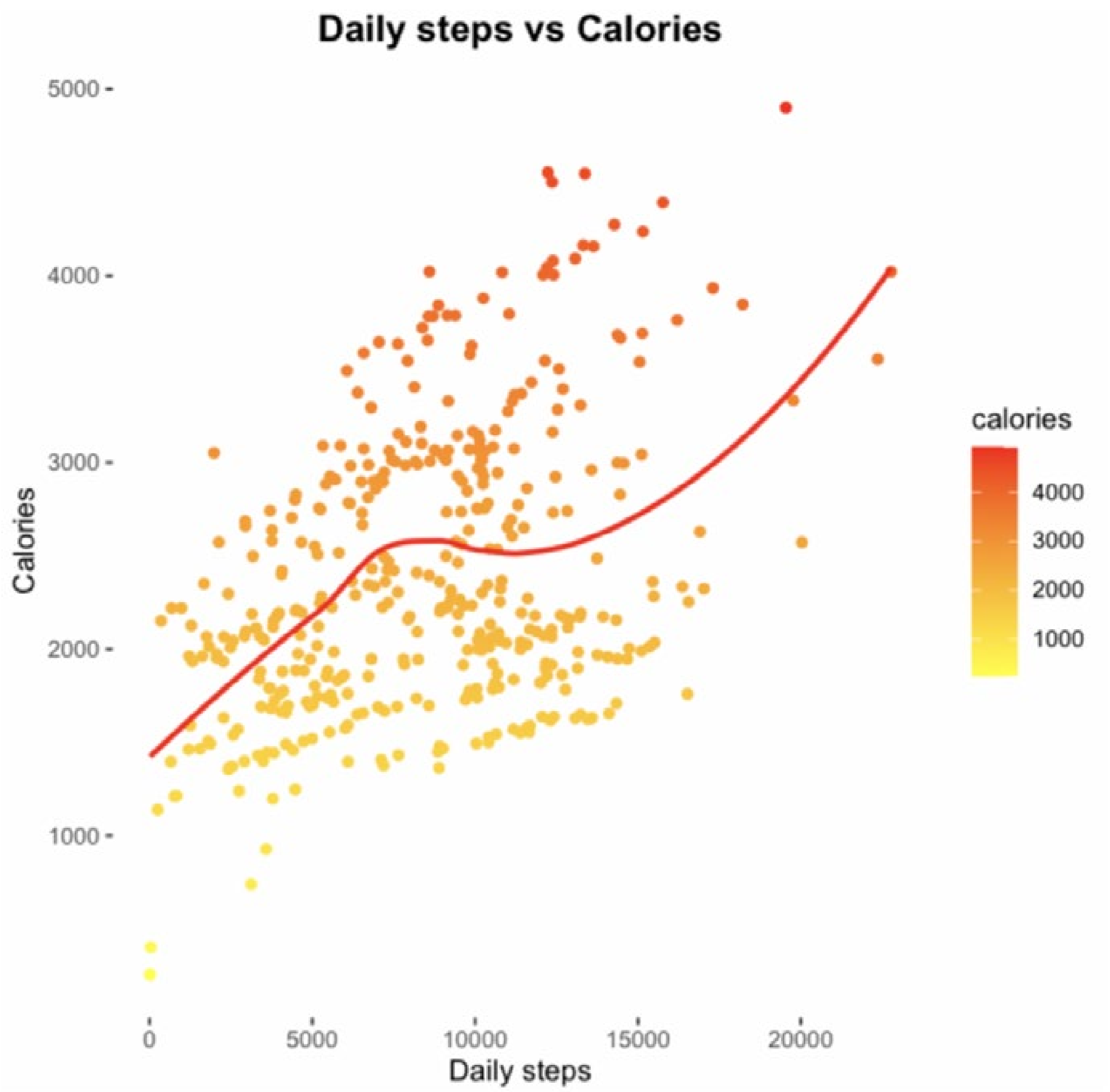
ANALYZE

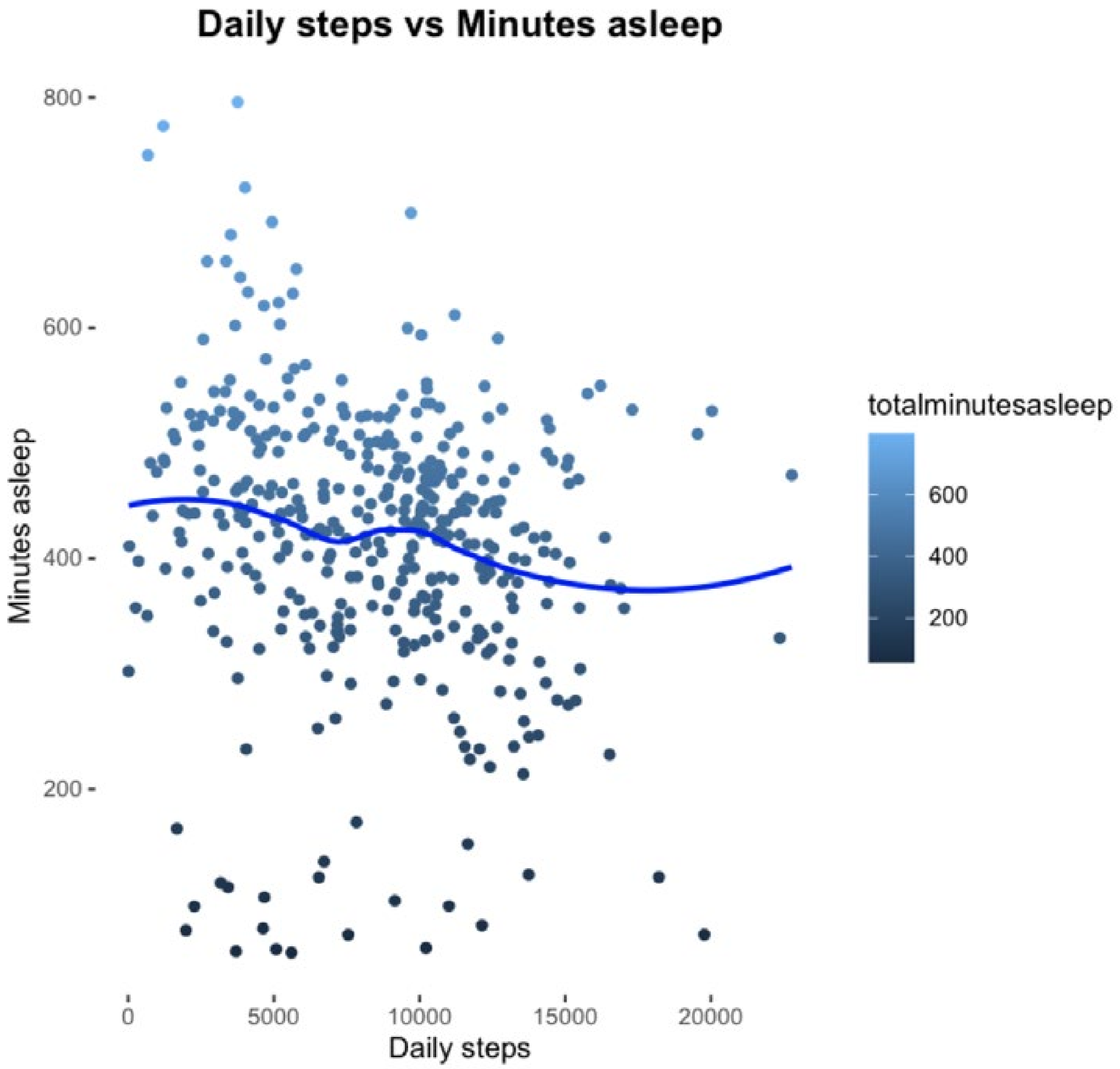
Let's find some trends in between data frames to determine if this information is going to be useful to resolve our business task and if there is some approach insights we can get for the BellaBeat's marketing Strategy.

SHARE

Correlations

We are going to verify some correlations in between Calories and Daily Steps:





We can clearly see a correlation between Daily Steps and Calories, the more steps made the more calories burned. We can focus on customers that doesn’t achieve the goal on the daily average steps and encourage them to do the correct daily_steps.

Let’s find the correlation between Minutes Asleep and Daily Steps
Based on our plots we can see users sleep a day.

We can see a positive correlation between total time in bed and minutes asleep, this can help us focus on users that spend more time in bed and encourage them to sleep the average hours to have better rest.

Define type of users

We want to determine the type of users with the data we have. We can classify the users by activity considering the daily amount of steps. We can categorize users as follows:

-Sedentary
-Lightly active
-Fairly active
-Very active

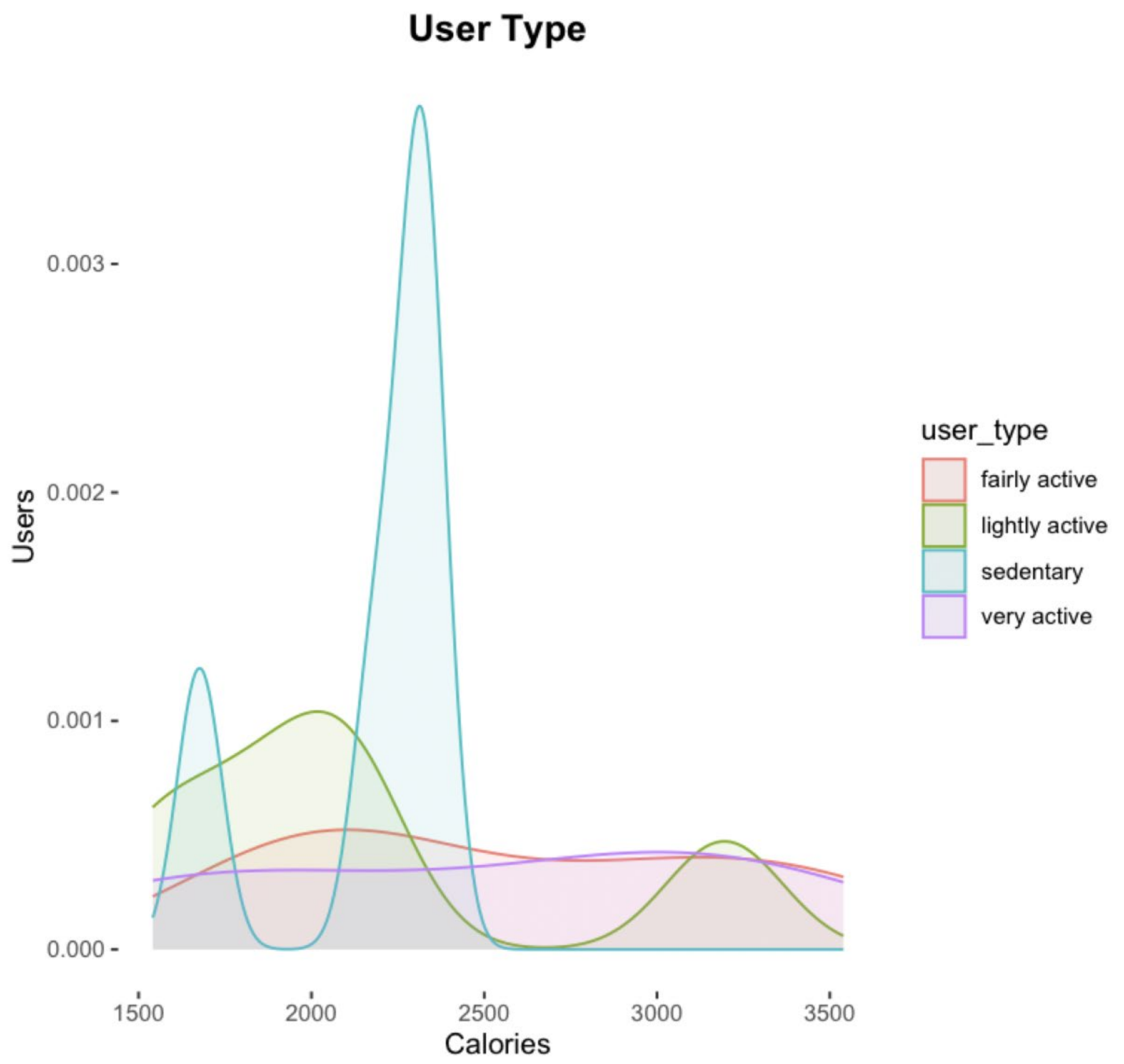
First we will calculate the daily steps average by user compare to the daily average steps in general.

```
daily_average <- daily_activity_sleep %>%
  group_by(id) %>%
  summarise (mean_daily_steps = mean(totalsteps), mean_daily_calories = mean(calories), mean_daily_sleep =
mean(totalminutesasleep))
```

Now we can to plot the average user type to analize trends or which category is the most commmon in between them.

```
user_type <- daily_average %>%
  mutate(user_type = case_when(
    mean_daily_steps < 5000 ~ "sedentary",
    mean_daily_steps >= 5000 & mean_daily_steps < 7499 ~ "lightly active",
    mean_daily_steps >= 7500 & mean_daily_steps < 9999 ~ "fairly active",
    mean_daily_steps >= 10000 ~ "very active"
```

we can see that users are fairly distributed by their activity considering the daily amount of calories burned. We can determine that even if most of the users are fairly distributed on these 4 categories the sedentary ones can have some notifications to encourage them and use the device more.

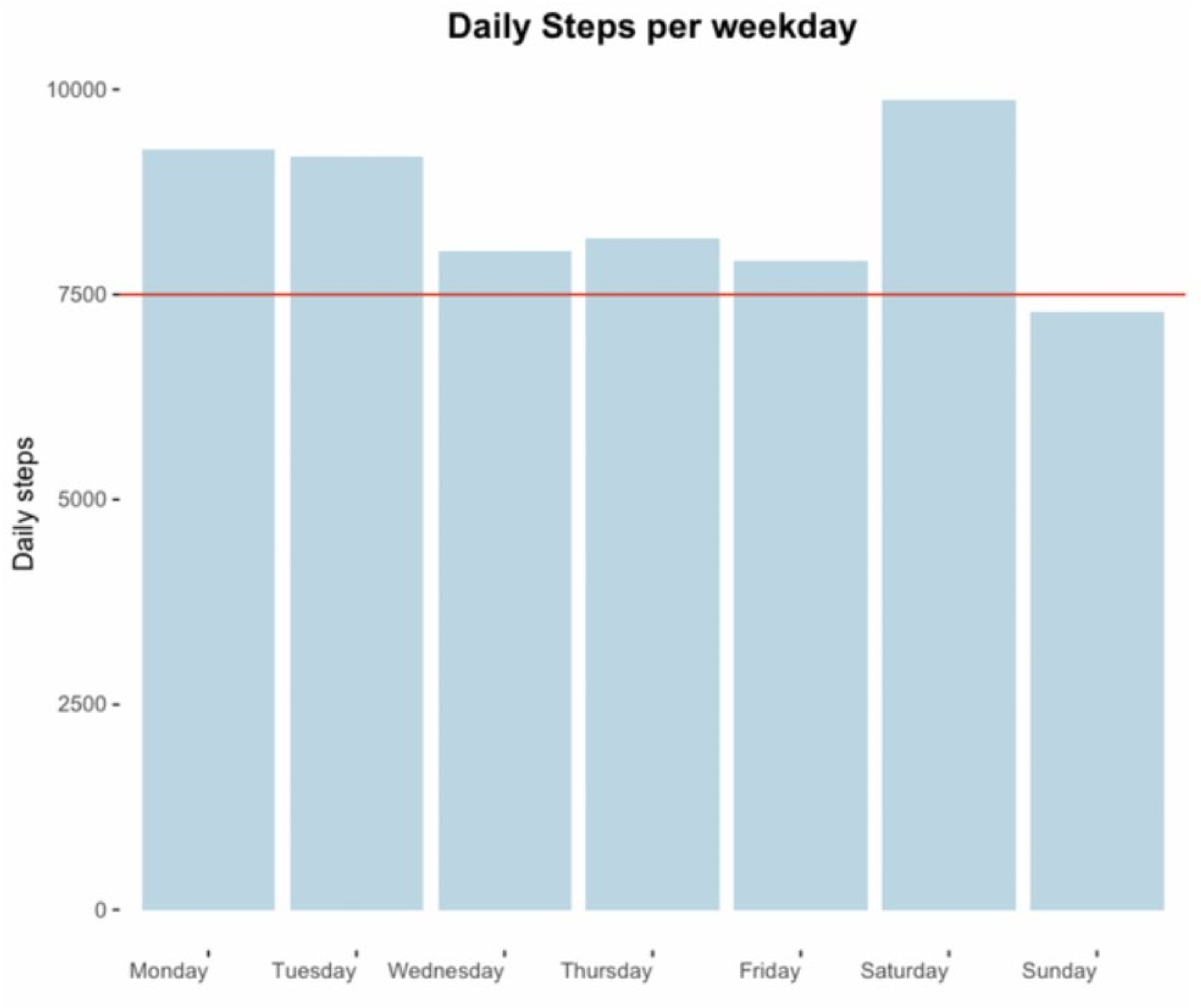


Steps & Weekdays

```
weekday_steps_sleep <- daily_activity_sleep %>%
  mutate(weekday = weekdays(date))

weekday_steps_sleep$weekday
<-ordered(weekday_steps_sleep$weekday, levels=c("Monday",
"Tuesday", "Wednesday", "Thursday",
"Friday", "Saturday", "Sunday"))

 weekday_steps_sleep <-weekday_steps_sleep%>%
  group_by(weekday) %>%
  summarize (daily_steps = mean(totalsteps), daily_sleep =
mean(totalminutesasleep))
```



In the graphs above we can determine the following:
Users walk daily the recommended amount of steps of 7500 besides Sunday’s.

ACT PHASE

Based on the analysis made we can conclude that the analysis can be wider if we can collect more data since the data we have can be biased since we don't have demographic data from the users. The period of time is too short and doesn't let use analyze deeper for trends. For example it can be useful to develop marketing strategies depending on young or adult women.

What we can work with based on this data analyzed is that:

1.Notifications on daily goals for customers that walk below the average 7,500 steps daily and persuade customers on Sunday's to do more outdoor activities focused more on the sedentary customers. It can be showing outdoor activities around their area (with demographic data) and type of activities they can do.

2.We can conclude from the analysis on the plots that there is a good correlation between Daily steps and Calories so we can provide more informative content about the healthy benefits on reaching the goal by having more than 7,500 steps per day.

3.More content as newsletter on novelties providing customer information about the benefits on having good night sleep 1 hr more taking into account the average on most of the customers that have 6 hrs of sleep. Persuade them on 5 minutes meditations to help them sleep as well as breathing exercises.

4.Create challenges were the customers can earn rewards.For example a monthly challenge were all users of BellaBeat's community can join to motivate them achieve the daily steps and outdoor activities on Sunday's. By doing this they can earn discounts on BellaBeat's products or unlock levels to share in between the community.

5.We can personalized content on customers based on the user type. Divide the content in the different 4 categories so we can persuade customers to use more the app or to take advantage of the benefits on using our products.

6.A daily notification content on the BellaBeat features that make our company unique and why is a revolutionary product in the health and wellness industry. The features to make relevant can be: wearable on every occasion product. Easy chargeable. Elegant. Water-resistant. Stress-free. Customizable. Long-lasting batteries