Fitbit users marketing data analysis

Mini Project | Mason Phung

Table of contents

01 Introduction

02 Methodology

03 Findings

04 Conclusions

01

Introduction and Objectives

# Bellabeat - A Sweden high-tech company

- Founded by Urška Sršen and Sando Mur in 2013, Bellabeat is known not only as a rapid developing high-tech company that manufactures health-focused smart products, but also a tech-driven wellness company for women.

•By 2016, Bellabeat had launched multiple products along with opening many branches worldwide. Bellabeat products can be found on their own e-commerce channel as well as numerous online retailers.

•Realizing the huge opportunity to grow in the online market, Bellabeat has been focusing on analyzing consumer usage data in order to understand the habit and preferences of their potential customers.

•Bellabeat current products:
- Bellabeat app
- "Leaf" (tracker, can be worn as accessories)
- "Time" (watch)
- "Spring" (water bottle with smart tech)
- Bellabeat membership (subscription based)
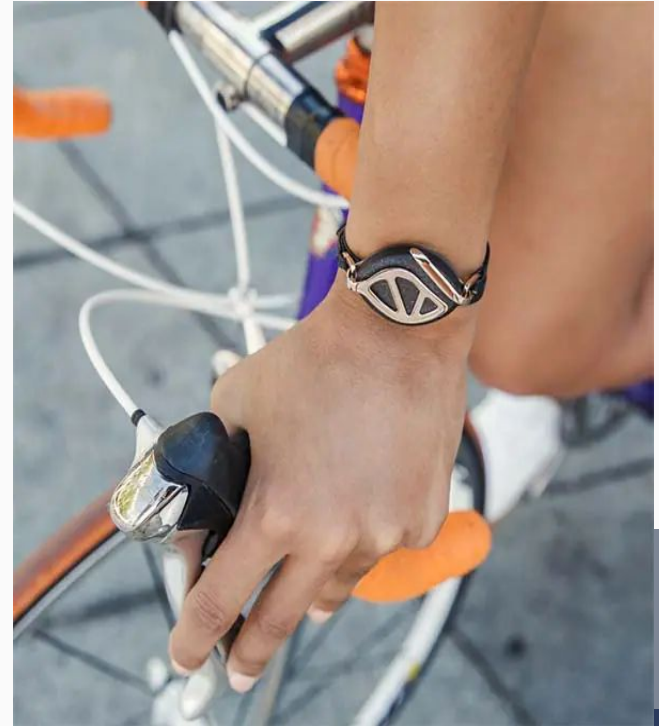
Bellabeat's leaf

(source: bellabeat.com)

Objectives

Our aim

- The analysis of the smart device usage data (Provided by the company)
- Suggestions for future marketing campaigns of Bellabeat's products based on the insights

Key questions

- What are some trends in smart device usage?
- How could these trends apply to Bellabeat campaign?
- How could these trends help influence marketing strategy?

Stakeholders

- Bellabeat marketing team
- Bellabeat executive board

Data

Basic information

- *Fitbit Fitness Tracker data pack* is generated by respondents to a distributed survey via Amazon Mechanical Turk between 2016/03/12 to 2016/05/12. The data is available on Kaggle and Mobius as open-source.

- The data pack includes 18 Comma-separated values files (.csv) record the activities, calories, intensities, foot steps and weight loss of 35 users in 62 days.

- Only **daily activity**, **daily sleep** and **hourly activity** datasets will be used in this project.

Bellabeat's leaf

(source: bellabeat.com)

Data

Limitations

- Small data size with only local population, which may not represent the whole population
- Lack of demographics data
- Data's short date length

-> **Bias possibility**

- Hourly and daily datasets are synced but still have some small errors, NA values.
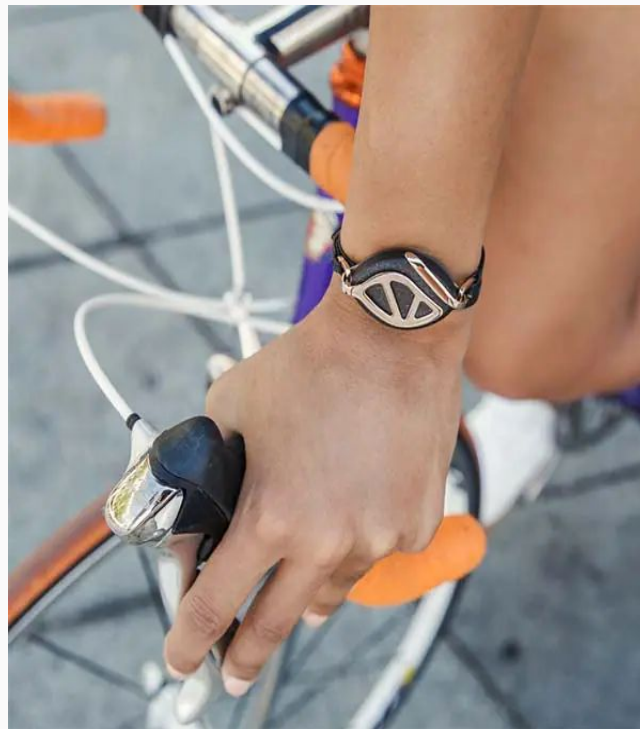
Bellabeat's leaf

(source: bellabeat.com)

02

Methodology

Tools

R

All of the work will be conducted using R programming language.

We mainly use tidyverse

Visualization libraries

To support the project's plotting, multiple Python visualization libraries will be used. Such as:

ggplot2, gridExtra, ggrepel, wesanderson

Data analysis process

01

Data cleaning
tidyverse
- Merge/join datasets
- Date formatting
- Determine missing values, NAs
- Remove unnecessary variables

02

Stats summary
dplyr
Summarize different variables in the datasets to find quick insights

03

Exploratory Descriptive Analysis
dplyr, ggplot2, …
- Use dplyr functions to find activity and, sleep statistics
- Visualize activity and sleep data using ggplot2 and support libraries: gridExtra, ggrepel, wesanderson.

04

Report
- Provide insights gained from the analysis
- Provide comments & suggestions

03

Findings

Exploratory analysis: Daily activity

On daily average

| 82%	18%	7281
Sedentary time	Active time	Walking steps
16.6	3.6	5.2 km
Sedentary hours	Active hours	Walking distance

- People spend most of the time in a day sedentary (82%), and about 3.6 hours a day for different types of activities

- The amount of time recorded from the participants in the data suggests that people wear the tracker most of the time in a day, even during normal activities or workout

Daily average activity time distribution (in minutes)
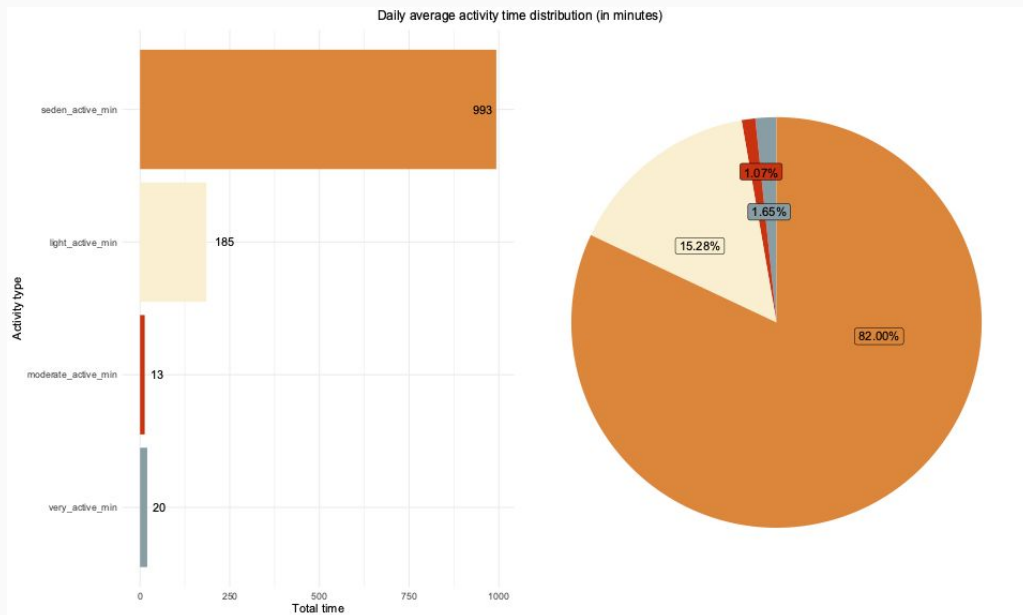
Exploratory analysis: Daily activity

Majority (80% - 28/35) of the participants walk less than 10,000 steps a day

Users daily average steps group distribution

Steps taken
- 0-5,000 steps
- 5,000-10,000 steps
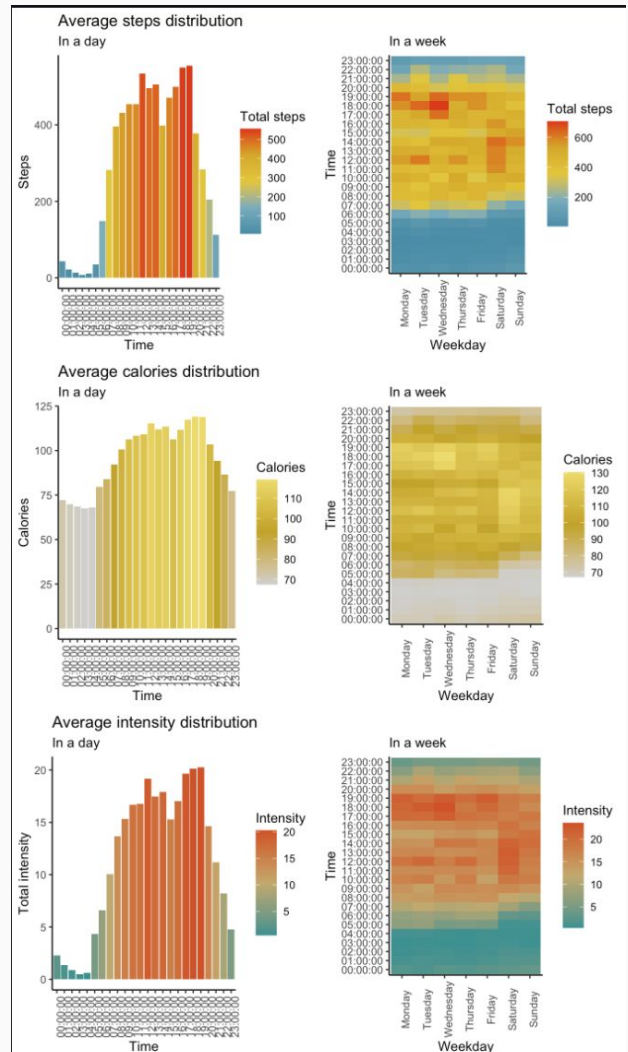- More than 10,000 steps
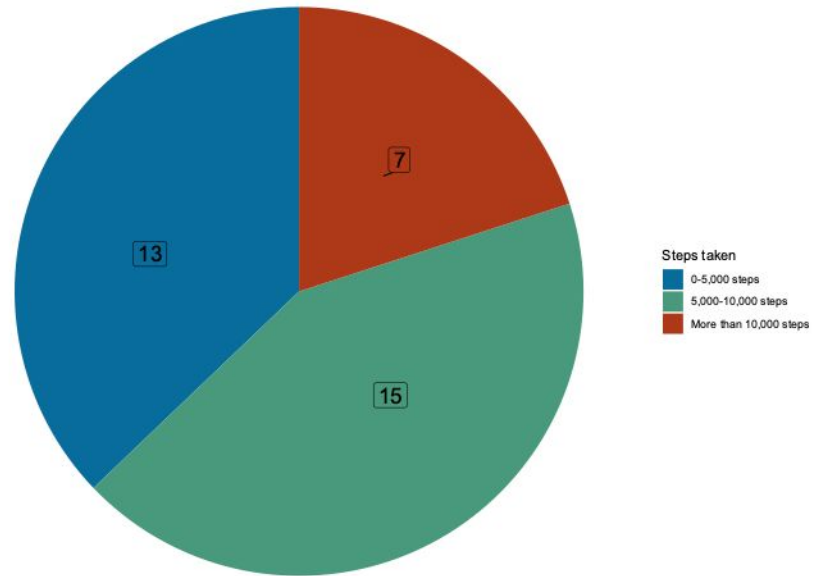
Exploratory analysis: Daily activity

- Most active days: **Saturday**, **Monday**, **Wednesday**

- During the week, people tend to walk less than 15,000 steps

- In the weekend, walking steps boundary opens up to >30,000 and the number of users take more than 15,000 steps/day increases

- **Sunday is the least active day, with the least average steps taken**
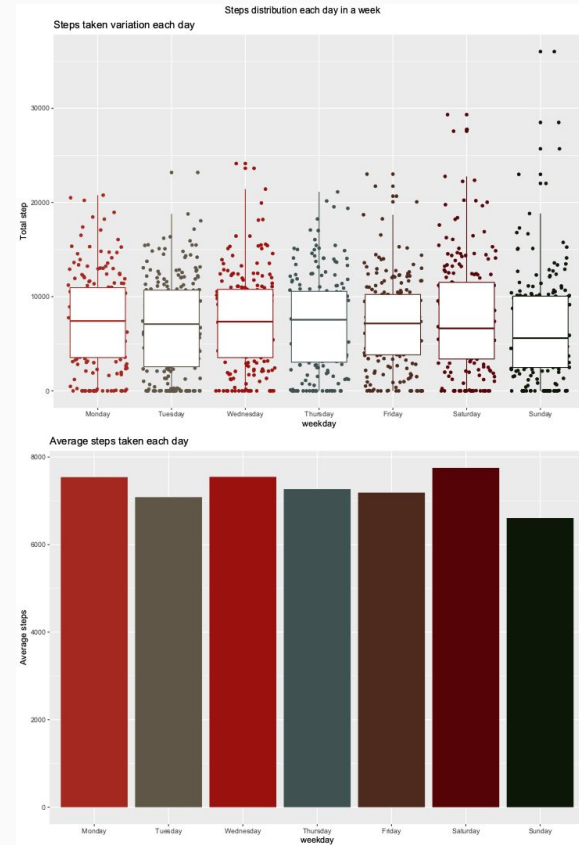
Exploratory analysis: Daily activity

- Common active time: from **7h to 21h**

- **12h-14h** and **17h-19h** are the two most active time with the highest amount of steps taken, calories burnt and total intensity

- These 2 time periods are all meal time (while the latter is the getting off work time, workouts and also people may move more to prepare for their dinner)

- At the weekend, people are usually start their days **later** but **move less in the evening** and still having the **same rest time** at the end of the day.

- **The heat maps suggest the active pattern of normal office workers**

Exploratory analysis: Daily activity

Only

24

users

Possess sleep records out of 35 users

Sunday Wednesday

Longest sleep

7

hours

Average sleep time

Users spend on average

39

mins

In bed without sleeping
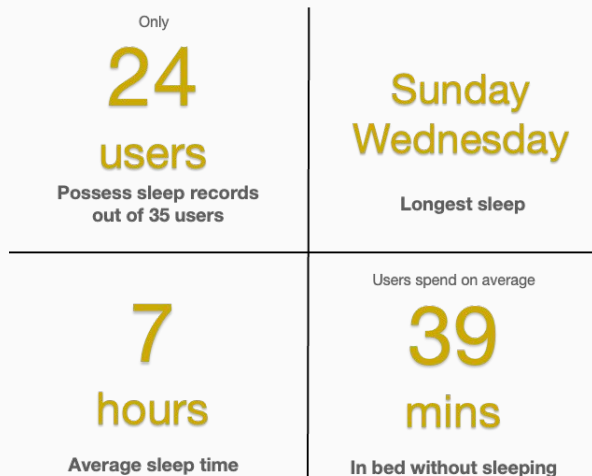
11

users

*don't wear the tracker device during sleep, this can be due to **personal preference** or the device's **lack of convenience** issue (need more information).*

Average sleep time distribution

Daily average amount of time asleep and in bed
In a week

Percentage of asleep time relative to in bed time
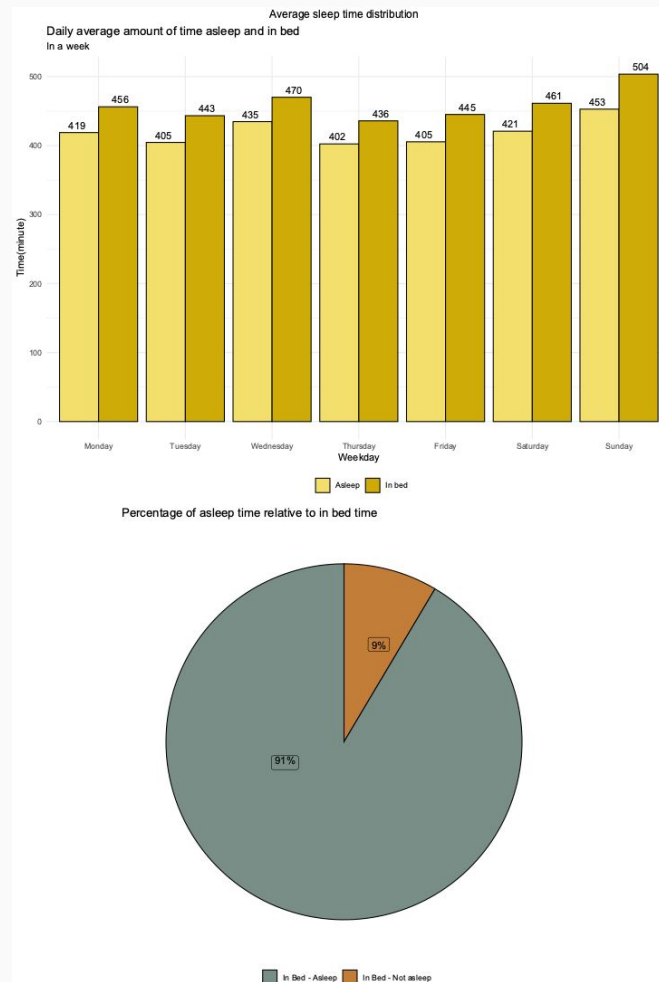
User segmentation

Users are divided into groups based on their daily average steps taken

13
people

Take less than 5,000 steps on daily average

15
people

Take from 5,000 – 10,000 steps a day

7
people

Walk more than 10,000 steps a day
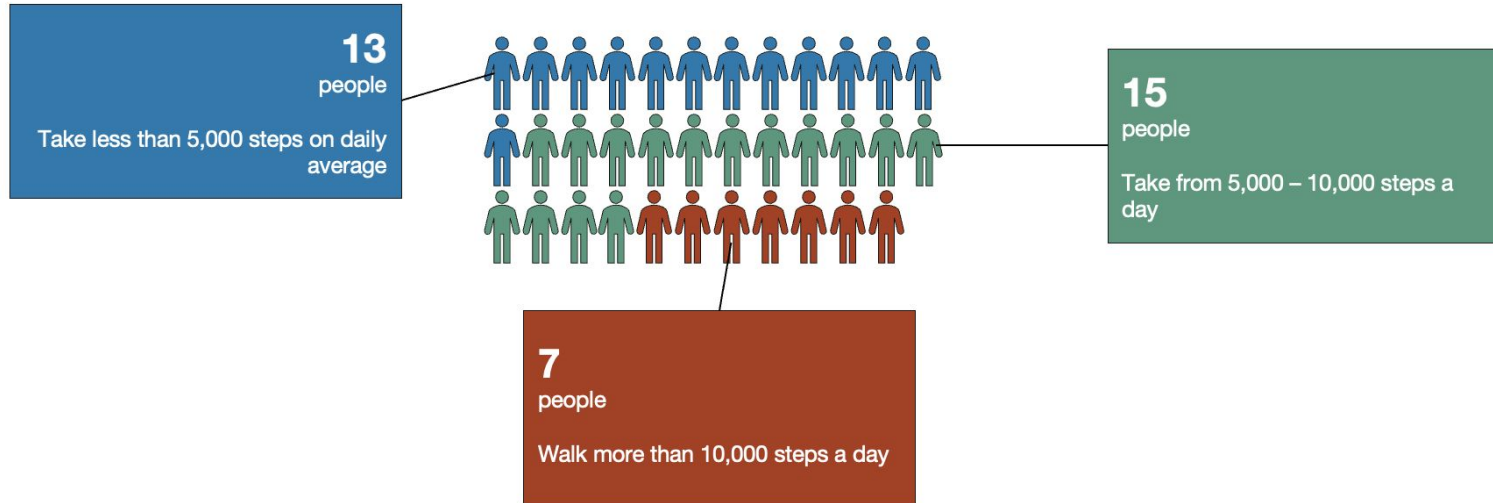
Exploratory research: By group

Group 1: Take less than 5,000 steps on daily average

36
Average tracker using days

2.15 km
Average daily distance travelled

Wednesday
Most active day

- This group of users only spend *2 hours avg/day active*, including light & intensive activities

- Most active time in a day at *17-19pm*

- Heatmap shows evidence of *frequent late night activity*, at sleep hours

- Aside from Wednesday, others day's activities are *quite similar* with *not many differences in travelled steps*

- This group often wears the tracking device, with *4 days avg/week*

- The data suggests a habit of a person who *not very active*, always *walk in short distance* and may *take not much exercises.*

- There is a *disordered pattern* in activity

Total active vs sedentary time

138 mins

11%

1135 mins

Activity type
seden_active_min
active_min

Total step by weekday

Total step

12.5k
10k
7.5k
5k
2.5k
0

| Monday | Tuesday | Wednesday | Thursday | Friday | Saturday | Sunday |
| 3.4k | 3.2k | 4k | 3.5k | 3.6k | 3.6k | 3.3k |

Weekday

Average steps heatmap

Exploratory research: By group

Group 2: Take from 5,000 - 10,000 steps on daily average

40
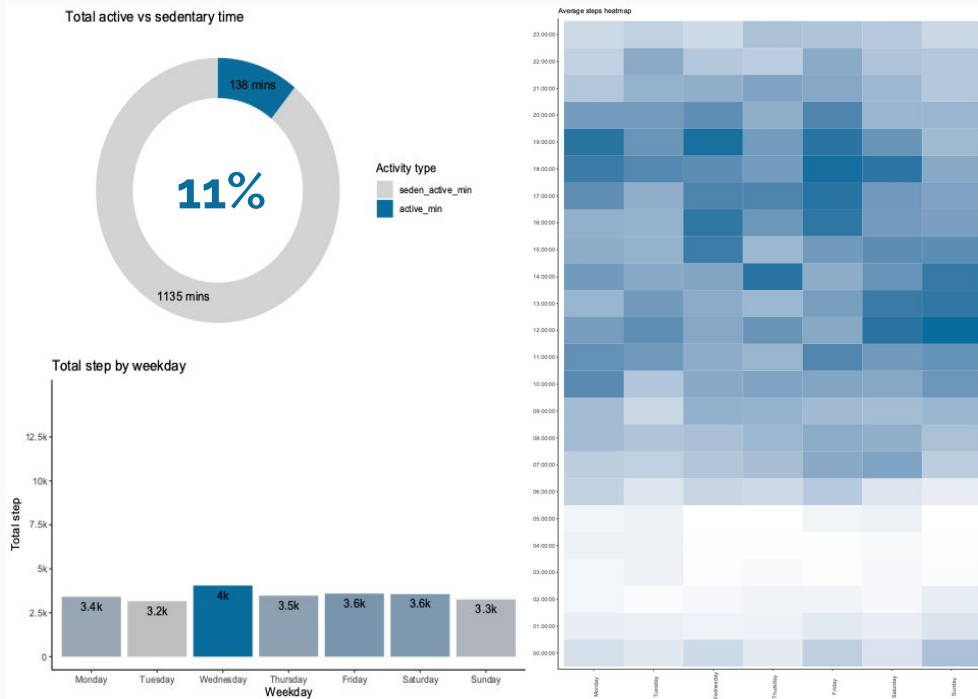
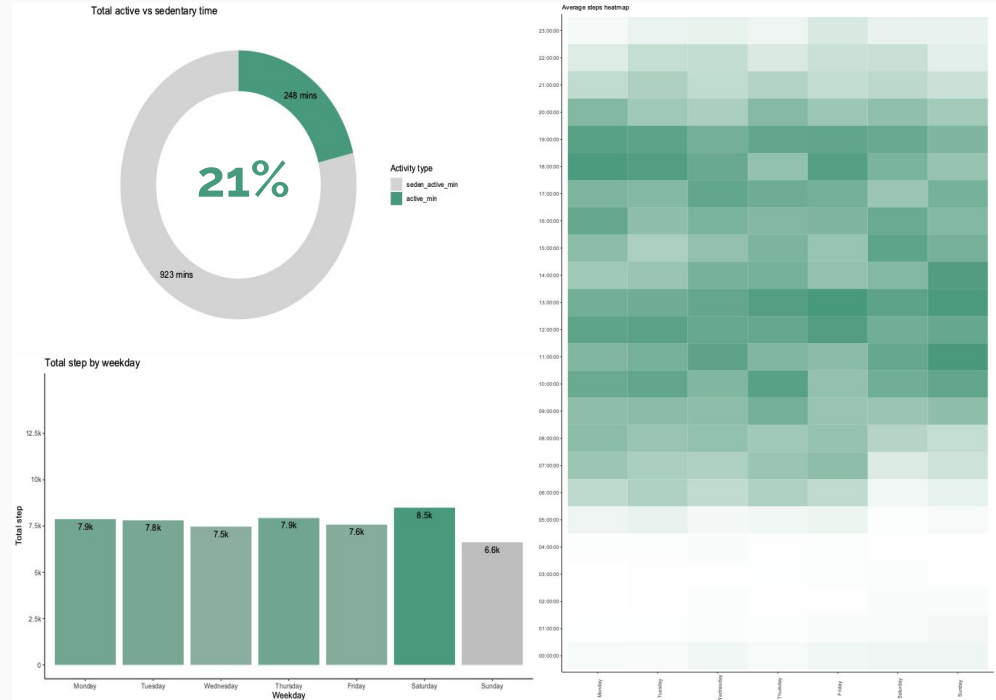Average tracker using days

5.42 km

Average daily distance travelled

Saturday

Most active day

- This group of users only spend *4 hours avg/day active*, including light & intensive activities

- Most active time in a day at *12-13pm and 18-19pm*

- Heatmap shows evidence of *stricker weekly schedule compared to G1*
    - Day usually starts at 6pm weekdays or 7-8pm weekend
    - Less late night activity compared to G1

- Aside from the weekend, the *activity of weekdays are similar*

- This group often wears the tracking device, with *4.4 days avg/week*

- The data suggests that users are *fairly active* and *follow a similar activity habit constantly*

Total active vs sedentary time

21%

248 mins

923 mins

Activity type
- sedan_active_min
- active_min

Average steps heatmap

Total step by weekday

7.9k 7.8k 7.5k 7.9k 7.6k 8.5k 6.6k

Total step

Monday Tuesday Wednesday Thursday Friday Saturday Sunday

Weekday

Exploratory research: By group

Group 3: Take more than 10,000 steps on daily average
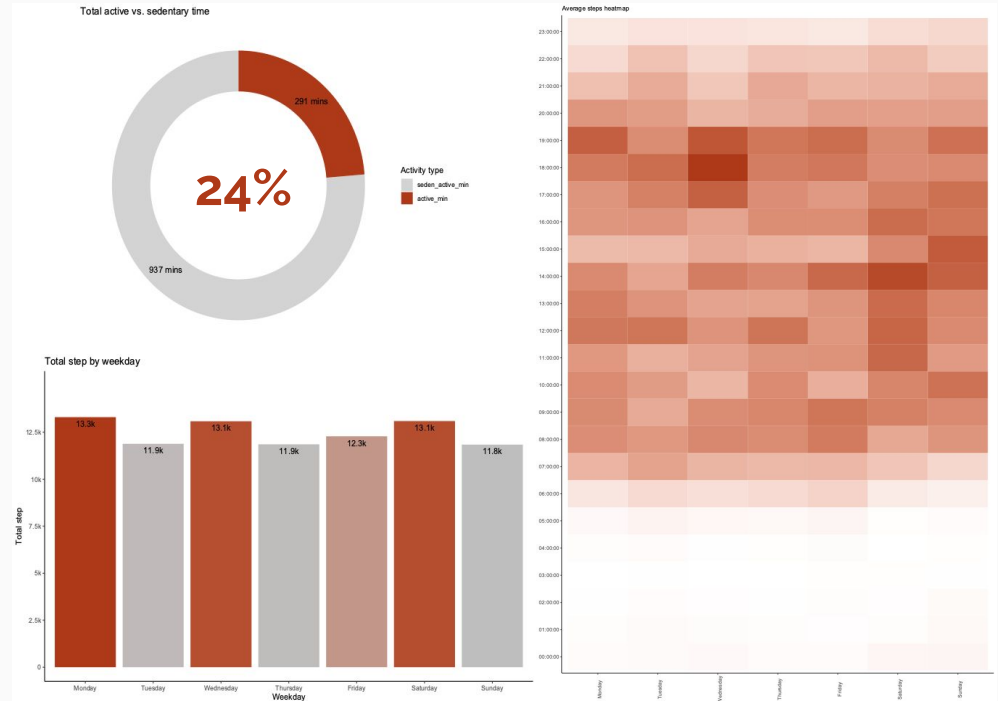
42

Average tracker using days

9.06 km

Average daily distance travelled

Monday

Most active day

- This group of users only spend *4.8 hours avg/day active*, including light & intensive activities

- Most active time in a day at *12-14pm and 18-19pm*

- Heatmap shows evidence of *a very strict weekly schedule*
  - Day usually starts at 6pm weekdays or 7-8pm weekend
  - Less late night activity, almost no activity after midnight

- Mon, Wed, Sat are considerably more active

- This group often wears the tracking device, with *4.6 days avg/week*

- The data suggests that users are *very active* and *follow a similar activity habit constantly*. They with spikes at Mon, Wed and Sat, this suggests that they might having a special exercise/activity on these days

Total active vs. sedentary time

291 mins

24%

937 mins

Activity type

seden_active_min

active_min

Total step by weekday

13.3k 11.9k 13.1k 11.9k 12.3k 13.1k 11.8k

Monday Tuesday Wednesday Thursday Friday Saturday Sunday

Weekday

Total step

Average steps heatmap

04

Conclusions

Comments: Insight summary

- **Tracker**
    - Users wear the tracker most of the time in a day, through different types of activity
    - Active users are more likely to wear the tracker
    - 31% of users doesn't wear the tracker during sleep
    - Users only wear the tracker 63% of the time throughout the survey
- **Users' active habits**
    - Active users are more likely to follow a strict schedule
    - Users spend 82% of the time in a day being sedentary, only be active at certain time periods
    - Users spend 84% of their active time doing light activities
    - Users are more active during meal time
    - Users rest more but travel more during the weekend
- **User's sleeping habits**
    - Users usually have a constant sleep time of ~ 7 hours daily
    - Users spend a considerable amount of time ~ 0.65 hours in bed without sleep

Comments: Suggestions

- Provide **new tracking modes** for different types of activities
- High amount of time not wearing the tracker shows that users may not feel comfortable with the design and consider r**edesign the tracker to fit different situation, environment, temperature..**
- **Create different user profiles**, features for different user groups from less active to very active users.
- Aside from sedentary activities, **create new feature for light activities** as it takes 84% of all the average activities in a day
- Create a **guide** or an **AI** to as an assistant to help reminding, scheduling and reporting
- Provide **health reports**
    - **Day-end report**: people usually spend some time to relax, restless before sleep, so we can provide a report of their health status during the day.
    - **Week-end report** on Sunday when people usually spend their time to rest, be less active and may want to take a look of what had happened throughout the week
- **Encourage users** to be more active by adding:
    - Reward system: game or an experience count system with level-up mechanism
    - Notifications: regular reminds customers to encourage them to be more active

Conclusions

Prediction method: Decision Tree

The test shows an outstanding result of 94% accuracy for decision tree classification. Which should be used for future prediction.

Promising future

Recent testing results show positive outcomes, which is an improvement compared to the past. Most missions were completed successfully.

Geographical aspects

Geography shows an essential impact to the launch sites. Most sites were placed close to transportations for logistics and were far from urban areas.