Fitbit Analyze

Chung-Yu Su

2023-04-18

Introduction

Case Study 2: How Can a Wellness Technology Company Play It Smart? Welcome to the Bellabeat data analysis case study! In this case study, you will perform many real-world tasks of a junior data analyst. You will imagine you are working for Bellabeat, a high-tech manufacturer of health-focused products for women, and meet different characters and team members. In order to answer the key business questions, you will follow the steps of the data analysis process: ask, prepare, process, analyze, share, and act. Along the way, the Case Study Roadmap tables — including guiding questions and key tasks — will help you stay on the right path. By the end of this lesson, you will have a portfolio-ready case study. Download the packet and reference the details of this case study anytime. Then, when you begin your job hunt, your case study will be a tangible way to demonstrate your knowledge and skills to potential employers. ## Scenario You are a junior data analyst working on the marketing analyst team at Bellabeat, a high-tech manufacturer of health-focused products for women. Bellabeat is a successful small company, but they have the potential to become a larger player in the global smart device market. Urška Sršen, cofounder and Chief Creative Officer of Bellabeat, believes that analyzing smart device fitness data could help unlock new growth opportunities for the company. You have been asked to focus on one of Bellabeat's products and analyze smart device data to gain insight into how consumers are using their smart devices. The insights you discover will then help guide marketing strategy for the company. You will present your analysis to the Bellabeat executive team along with your high-level recommendations for Bellabeat's marketing strategy.

Characters and products

● Characters

○ Urška Sršen: Bellabeat's cofounder and Chief Creative Office

○ Sando Mur: Mathematician and Bellabeat's cofounder; key member of the Bellabeat executive team

○ Bellabeat marketing analytics team: A team of data analysts responsible for collecting, analyzing, andreporting data that helps guide Bellabeat's marketing strategy. You joined this team six months ago and have been busy learning about Bellabeat''s mission and business goals — as well as how you, as a junior data analyst, can help Bellabeat achieve them.

● Products

○ Bellabeat app: The Bellabeat app provides users with health data related to their activity, sleep, stress,menstrual cycle, and mindfulness habits. This data can help users better understand their current habits andmake healthy decisions. The Bellabeat app connects to their line of smart wellness products. ○ Leaf: Bellabeat's classic wellness tracker can be worn as a bracelet, necklace, or clip. The Leaf tracker connectsto the Bellabeat app to track activity, sleep, and stress. ○ Time: This wellness watch combines the timeless look of a classic timepiece with smart technology to track user activity, sleep, and stress. The Time watch connects to the Bellabeat app to provide you with insights into your daily wellness. ○ Spring: This is a water bottle that tracks daily water intake using smart technology to ensure that you areappropriately hydrated throughout the day. The Spring bottle connects to the Bellabeat app to track your hydration levels. ○ Bellabeat membership: Bellabeat also offers a subscription-based membership program for users. Membership gives users 24/7 access to fully personalized guidance on nutrition, activity, sleep, health and beauty, and mindfulness based on their lifestyle and goals.

About the company

Urska Srsen and Sando Mur founded Bellabeat, a high-tech company that manufactures health-focused smart products. Srsen 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. By 2016, Bellabeat had opened offices around the world and launched multiple products. Bellabeat products became available through a growing number of online retailers in addition to their own e-commerce channel on their website. The company has invested in traditional advertising media, such as radio, out-of-home billboards, print, and television, but focuses on digital marketing extensively. Bellabeat invests year-round in Google Search, maintaining active Facebook and Instagram pages, and consistently engages consumers on Twitter. Additionally, Bellabeat runs video ads on Youtube and display ads on the Google Display Network to support campaigns around key marketing dates.

Sršen knows that an analysis of Bellabeat's available consumer data would reveal more opportunities for growth. She has asked the marketing analytics team to focus on a Bellabeat product and analyze smart device usage data in order to gain insight into how people are already using their smart devices. Then, using this information, she would like high-level recommendations for how these trends can inform Bellabeat marketing strategy.

啟用 Packages

library(tidyverse)

```
## —— Attaching core tidyverse packages ————————————————————————————————
———————————— tidyverse 2.0.0 ——
## ✔ dplyr     1.1.1     ✔ readr     2.1.4
## ✔ forcats   1.0.0     ✔ stringr   1.5.0
## ✔ ggplot2   3.4.2     ✔ tibble    3.2.1
## ✔ lubridate 1.9.2     ✔ tidyr     1.3.0
## ✔ purrr     1.0.1
## —— Conflicts ————————————————————————————————————————————————————————
———————————————————————— tidyverse_conflicts() ——
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
 errors
```

library(ggplot2)
library(GGally)

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

匯入資料集

Active <- read_csv("/Users/chungyu/Downloads/Fitabase Data 4.12.16-5.12.16/dailyActivity_merged.csv")

```
## Rows: 940 Columns: 15
## —— Column specification ——————————————————————————————————————————————
————————————————————————————————————————————————————————
## Delimiter: ","
## chr  (1): ActivityDate
## dbl (14): Id, TotalSteps, TotalDistance, TrackerDistance, LoggedActivitiesDi...
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
Intensities <- read_csv("/Users/chungyu/Downloads/Fitabase Data 4.12.16-5.12.16/dailyIntensities_merged.csv")
```

Rows: 940 Columns: 10
—— Column specification ——
Delimiter: ","
chr (1): ActivityDay
dbl (9): Id, SedentaryMinutes, LightlyActiveMinutes, FairlyActiveMinutes, Ve...
##
ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.

```
Sleep <- read_csv("/Users/chungyu/Downloads/Fitabase Data 4.12.16-5.12.16/sleepDay_merged.csv")
```

Rows: 413 Columns: 5
—— Column specification ——
Delimiter: ","
chr (1): SleepDay
dbl (4): Id, TotalSleepRecords, TotalMinutesAsleep, TotalTimeInBed
##
ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.

```
Weight <- read_csv("/Users/chungyu/Downloads/Fitabase Data 4.12.16-5.12.16/weightLogInfo_merged.csv")
```

Rows: 67 Columns: 8
—— Column specification ——
Delimiter: ","
chr (1): Date
dbl (6): Id, WeightKg, WeightPounds, Fat, BMI, LogId
lgl (1): IsManualReport
##
ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.

瀏覽原始檔案

```
head(Active)
```

A tibble: 6 × 15
Id ActivityDate TotalSteps TotalDistance TrackerDistance
<dbl> <chr>   <dbl>   <dbl>   <dbl>

```
## 1 1503960366 4/12/2016        13162      8.5        8.5
## 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
## # ℹ 10 more variables: LoggedActivitiesDistance <dbl>,
## #   VeryActiveDistance <dbl>, ModeratelyActiveDistance <dbl>,
## #   LightActiveDistance <dbl>, SedentaryActiveDistance <dbl>,
## #   VeryActiveMinutes <dbl>, FairlyActiveMinutes <dbl>,
## #   LightlyActiveMinutes <dbl>, SedentaryMinutes <dbl>, Calories <dbl>
```

head(Intensities)

```
## # A tibble: 6 × 10
##         Id ActivityDay SedentaryMinutes LightlyActiveMinutes FairlyActiveMinutes
##      <dbl> <chr>                  <dbl>                <dbl>               <dbl>
## 1  1.50e9 4/12/2016                728                  328                  13
## 2  1.50e9 4/13/2016                776                  217                  19
## 3  1.50e9 4/14/2016               1218                  181                  11
## 4  1.50e9 4/15/2016                726                  209                  34
## 5  1.50e9 4/16/2016                773                  221                  10
## 6  1.50e9 4/17/2016                539                  164                  20
## # ℹ 5 more variables: VeryActiveMinutes <dbl>, SedentaryActiveDistance <dbl>,
## #   LightActiveDistance <dbl>, ModeratelyActiveDistance <dbl>,
## #   VeryActiveDistance <dbl>
```

head(Sleep)

```
## # A tibble: 6 × 5
##         Id SleepDay       TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
##      <dbl> <chr>                      <dbl>              <dbl>          <dbl>
## 1 1503960366 4/12/2016 12:0⋯              1                327            346
## 2 1503960366 4/13/2016 12:0⋯              2                384            407
## 3 1503960366 4/15/2016 12:0⋯              1                412            442
## 4 1503960366 4/16/2016 12:0⋯              2                340            367
## 5 1503960366 4/17/2016 12:0⋯              1                700            712
## 6 1503960366 4/19/2016 12:0⋯              1                304            320
```

head(Weight)

```
## # A tibble: 6 × 8
##         Id Date      WeightKg WeightPounds   Fat   BMI IsManualReport   LogId
##      <dbl> <chr>        <dbl>        <dbl> <dbl> <dbl> <lgl>            <dbl>
## 1 1503960366 5/2/2016 ⋯      52.6         116.    22  22.6 TRUE          1.46e12
## 2 1503960366 5/3/2016 ⋯      52.6         116.    NA  22.6 TRUE          1.46e12
```

## 3 1927972279 4/13/2016···	134.	294.	NA	47.5 FALSE	1.46e12
## 4 2873212765 4/21/2016···	56.7	125.	NA	21.5 TRUE	1.46e12
## 5 2873212765 5/12/2016···	57.3	126.	NA	21.7 TRUE	1.46e12
## 6 4319703577 4/17/2016···	72.4	160.	25	27.5 TRUE	1.46e12

資料清理

清理缺失值

```
Active <- Active %>% distinct() %>% drop_na()
Intensities <- Intensities %>% distinct() %>% drop_na()
Sleep <- Sleep %>% distinct() %>% drop_na()
```

增加活動時間分鐘與小時的個別總和

```
Active <- Active %>%
  mutate(TotalActiveMinute = SedentaryMinutes + LightlyActiveMinutes +      FairlyActive
Minutes + VeryActiveMinutes) %>%
  mutate(TotalActiveHour = TotalActiveMinute/60)
head(Active)
```

```
## # A tibble: 6 × 17
##          Id ActivityDate TotalSteps TotalDistance TrackerDistance
##       <dbl> <chr>            <dbl>        <dbl>          <dbl>
## 1 1503960366 4/12/2016        13162          8.5            8.5
## 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
## # ℹ 12 more variables: LoggedActivitiesDistance <dbl>,
## #   VeryActiveDistance <dbl>, ModeratelyActiveDistance <dbl>,
## #   LightActiveDistance <dbl>, SedentaryActiveDistance <dbl>,
## #   VeryActiveMinutes <dbl>, FairlyActiveMinutes <dbl>,
## #   LightlyActiveMinutes <dbl>, SedentaryMinutes <dbl>, Calories <dbl>,
## #   TotalActiveMinute <dbl>, TotalActiveHour <dbl>
```

分割日期與時間

```
Sleep <- Sleep %>%
  separate(SleepDay, c("Date" , "Time") ," ") %>%
  mutate( TotalHourInBed = TotalTimeInBed/60)
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 410 rows [1, 2, 3, 4, 5, 6,
## 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].
```

```
head(Sleep)
```

```
## # A tibble: 6 × 7
##          Id Date       Time   TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
##       <dbl> <chr>      <chr>              <dbl>              <dbl>          <dbl>
## 1 1503960366 4/12/2016 12:0…                 1                327            346
## 2 1503960366 4/13/2016 12:0…                 2                384            407
## 3 1503960366 4/15/2016 12:0…                 1                412            442
## 4 1503960366 4/16/2016 12:0…                 2                340            367
## 5 1503960366 4/17/2016 12:0…                 1                700            712
## 6 1503960366 4/19/2016 12:0…                 1                304            320
## # ℹ 1 more variable: TotalHourInBed <dbl>
```

```r
Weight <- Weight %>%
  separate(Date,c("Date","Time","AMPM") , " ")
head(Weight)
```

```
## # A tibble: 6 × 10
##          Id Date   Time  AMPM  WeightKg WeightPounds   Fat   BMI IsManualReport
##       <dbl> <chr>  <chr> <chr>    <dbl>        <dbl> <dbl> <dbl> <lgl>
## 1 1503960366 5/2/2… 11:5… PM       52.6         116.    22  22.6 TRUE
## 2 1503960366 5/3/2… 11:5… PM       52.6         116.    NA  22.6 TRUE
## 3 1927972279 4/13/… 1:08… AM      134.          294.    NA  47.5 FALSE
## 4 2873212765 4/21/… 11:5… PM       56.7         125.    NA  21.5 TRUE
## 5 2873212765 5/12/… 11:5… PM       57.3         126.    NA  21.7 TRUE
## 6 4319703577 4/17/… 11:5… PM       72.4         160.    25  27.5 TRUE
## # ℹ 1 more variable: LogId <dbl>
```

合併每日活動以及睡眠資料

```r
combined_data <- left_join(Sleep , Active ,by = c('Id'  = 'Id' , 'Date' = "ActivityDate"))
combined_data <- combined_data %>%
  mutate(TotalHour = TotalHourInBed + TotalActiveHour) %>%
  mutate(TotalHourAsleep = TotalMinutesAsleep / 60)
head(combined_data)
```

```
## # A tibble: 6 × 24
##          Id Date       Time   TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
##       <dbl> <chr>      <chr>              <dbl>              <dbl>          <dbl>
## 1 1503960366 4/12/2016 12:0…                 1                327            346
## 2 1503960366 4/13/2016 12:0…                 2                384            407
## 3 1503960366 4/15/2016 12:0…                 1                412            442
## 4 1503960366 4/16/2016 12:0…                 2                340            367
## 5 1503960366 4/17/2016 12:0…                 1                700            712
## 6 1503960366 4/19/2016 12:0…                 1                304            320
## # ℹ 18 more variables: TotalHourInBed <dbl>, TotalSteps <dbl>,
## #   TotalDistance <dbl>, TrackerDistance <dbl>, LoggedActivitiesDistance <dbl>,
## #   VeryActiveDistance <dbl>, ModeratelyActiveDistance <dbl>,
```

```
## #   LightActiveDistance <dbl>, SedentaryActiveDistance <dbl>,
## #   VeryActiveMinutes <dbl>, FairlyActiveMinutes <dbl>,
## #   LightlyActiveMinutes <dbl>, SedentaryMinutes <dbl>, Calories <dbl>,
## #   TotalActiveMinute <dbl>, TotalActiveHour <dbl>, TotalHour <dbl>, …
```

消除活動時間與睡眠時間相加超過 24 小時的資料

```
combined_data %>% filter(TotalHour <= 24)
```

```
## # A tibble: 255 × 24
##            Id Date    Time  TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
##         <dbl> <chr>   <chr>             <dbl>              <dbl>          <dbl>
##  1 1503960366 4/12/20… 12:0…                1                327            346
##  2 1503960366 4/13/20… 12:0…                2                384            407
##  3 1503960366 4/15/20… 12:0…                1                412            442
##  4 1503960366 4/16/20… 12:0…                2                340            367
##  5 1503960366 4/19/20… 12:0…                1                304            320
##  6 1503960366 4/20/20… 12:0…                1                360            377
##  7 1503960366 4/21/20… 12:0…                1                325            364
##  8 1503960366 4/23/20… 12:0…                1                361            384
##  9 1503960366 4/24/20… 12:0…                1                430            449
## 10 1503960366 4/25/20… 12:0…                1                277            323
## # ℹ 245 more rows
## # ℹ 18 more variables: TotalHourInBed <dbl>, TotalSteps <dbl>,
## #   TotalDistance <dbl>, TrackerDistance <dbl>, LoggedActivitiesDistance <dbl>,
## #   VeryActiveDistance <dbl>, ModeratelyActiveDistance <dbl>,
## #   LightActiveDistance <dbl>, SedentaryActiveDistance <dbl>,
## #   VeryActiveMinutes <dbl>, FairlyActiveMinutes <dbl>,
## #   LightlyActiveMinutes <dbl>, SedentaryMinutes <dbl>, Calories <dbl>, …
```

```
head(combined_data)
```

```
## # A tibble: 6 × 24
##            Id Date    Time  TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
##         <dbl> <chr>   <chr>             <dbl>              <dbl>          <dbl>
## 1 1503960366 4/12/2016 12:0…                1                327            346
## 2 1503960366 4/13/2016 12:0…                2                384            407
## 3 1503960366 4/15/2016 12:0…                1                412            442
## 4 1503960366 4/16/2016 12:0…                2                340            367
## 5 1503960366 4/17/2016 12:0…                1                700            712
## 6 1503960366 4/19/2016 12:0…                1                304            320
## # ℹ 18 more variables: TotalHourInBed <dbl>, TotalSteps <dbl>,
## #   TotalDistance <dbl>, TrackerDistance <dbl>, LoggedActivitiesDistance <dbl>,
## #   VeryActiveDistance <dbl>, ModeratelyActiveDistance <dbl>,
## #   LightActiveDistance <dbl>, SedentaryActiveDistance <dbl>,
## #   VeryActiveMinutes <dbl>, FairlyActiveMinutes <dbl>,
```

```
## #   LightlyActiveMinutes <dbl>, SedentaryMinutes <dbl>, Calories <dbl>,
## #   TotalActiveMinute <dbl>, TotalActiveHour <dbl>, TotalHour <dbl>, …
```

睡眠情況分類

```
combined_data$SleepCon = case_when(combined_data$TotalHourAsleep >= 7 ~ "Enough", T
RUE ~ "Not Enough")
head(combined_data)

## # A tibble: 6 × 25
##          Id Date      Time  TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
##       <dbl> <chr>     <chr>             <dbl>              <dbl>          <dbl>
## 1 1503960366 4/12/2016 12:0⋯               1                327            346
## 2 1503960366 4/13/2016 12:0⋯               2                384            407
## 3 1503960366 4/15/2016 12:0⋯               1                412            442
## 4 1503960366 4/16/2016 12:0⋯               2                340            367
## 5 1503960366 4/17/2016 12:0⋯               1                700            712
## 6 1503960366 4/19/2016 12:0⋯               1                304            320
## # ℹ 19 more variables: TotalHourInBed <dbl>, TotalSteps <dbl>,
## #   TotalDistance <dbl>, TrackerDistance <dbl>, LoggedActivitiesDistance <dbl>,
## #   VeryActiveDistance <dbl>, ModeratelyActiveDistance <dbl>,
## #   LightActiveDistance <dbl>, SedentaryActiveDistance <dbl>,
## #   VeryActiveMinutes <dbl>, FairlyActiveMinutes <dbl>,
## #   LightlyActiveMinutes <dbl>, SedentaryMinutes <dbl>, Calories <dbl>,
## #   TotalActiveMinute <dbl>, TotalActiveHour <dbl>, TotalHour <dbl>, …
```

距離誤差

```
combined_data$Distance_Error <- abs(combined_data$TotalDistance - combined_data$Tracker
Distance)
head(combined_data)

## # A tibble: 6 × 26
##          Id Date      Time  TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
##       <dbl> <chr>     <chr>             <dbl>              <dbl>          <dbl>
## 1 1503960366 4/12/2016 12:0⋯               1                327            346
## 2 1503960366 4/13/2016 12:0⋯               2                384            407
## 3 1503960366 4/15/2016 12:0⋯               1                412            442
## 4 1503960366 4/16/2016 12:0⋯               2                340            367
## 5 1503960366 4/17/2016 12:0⋯               1                700            712
## 6 1503960366 4/19/2016 12:0⋯               1                304            320
## # ℹ 20 more variables: TotalHourInBed <dbl>, TotalSteps <dbl>,
## #   TotalDistance <dbl>, TrackerDistance <dbl>, LoggedActivitiesDistance <dbl>,
## #   VeryActiveDistance <dbl>, ModeratelyActiveDistance <dbl>,
## #   LightActiveDistance <dbl>, SedentaryActiveDistance <dbl>,
## #   VeryActiveMinutes <dbl>, FairlyActiveMinutes <dbl>,
```

```
## #   LightlyActiveMinutes <dbl>, SedentaryMinutes <dbl>, Calories <dbl>,
## #   TotalActiveMinute <dbl>, TotalActiveHour <dbl>, TotalHour <dbl>, …
```

sum(combined_data$Distance_Error)*#計算出誤差總和*

```
## [1] 2.06
```

#變更日期格式
combined_data$Date <- as.Date(combined_data$Date,format = "%m/%d/%y")

combined_data$weekday <- weekdays(combined_data$Date)
head(combined_data)

```
## # A tibble: 6 × 27
##         Id Date       Time  TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
##      <dbl> <date>     <chr>             <dbl>              <dbl>          <dbl>
## 1  1.50e9 2020-04-12 12:0…                 1                327            346
## 2  1.50e9 2020-04-13 12:0…                 2                384            407
## 3  1.50e9 2020-04-15 12:0…                 1                412            442
## 4  1.50e9 2020-04-16 12:0…                 2                340            367
## 5  1.50e9 2020-04-17 12:0…                 1                700            712
## 6  1.50e9 2020-04-19 12:0…                 1                304            320
## # ℹ 21 more variables: TotalHourInBed <dbl>, TotalSteps <dbl>,
## #   TotalDistance <dbl>, TrackerDistance <dbl>, LoggedActivitiesDistance <dbl>,
## #   VeryActiveDistance <dbl>, ModeratelyActiveDistance <dbl>,
## #   LightActiveDistance <dbl>, SedentaryActiveDistance <dbl>,
## #   VeryActiveMinutes <dbl>, FairlyActiveMinutes <dbl>,
## #   LightlyActiveMinutes <dbl>, SedentaryMinutes <dbl>, Calories <dbl>,
## #   TotalActiveMinute <dbl>, TotalActiveHour <dbl>, TotalHour <dbl>, …
```

實際距離與裝置距離相差不多

分析資料

使用者睡眠情況

```
sleep_enough <- sum(combined_data$SleepCon == "Enough")
not_sleep_enough <- sum(combined_data$SleepCon == "Not Enough")
total_sleepcon <- sleep_enough + not_sleep_enough
sep = sleep_enough / total_sleepcon
nsep = not_sleep_enough / total_sleepcon
sleep_percent <- c(sep,nsep)
spdf <- data.frame(category = c("sleep enough","Not sleep enough" ),sleep_percent)

ggplot(data = spdf , aes(x = "",y = sleep_percent,fill = category)) +
  geom_bar(stat = "identity")+
  scale_fill_manual(values = c( "#009E73", "#F0E442")) +
  coord_polar(theta = "y") +
  theme_void() +
  labs(fill="") +
  geom_text(aes(label=category), position=position_stack(vjust=0.5))
```
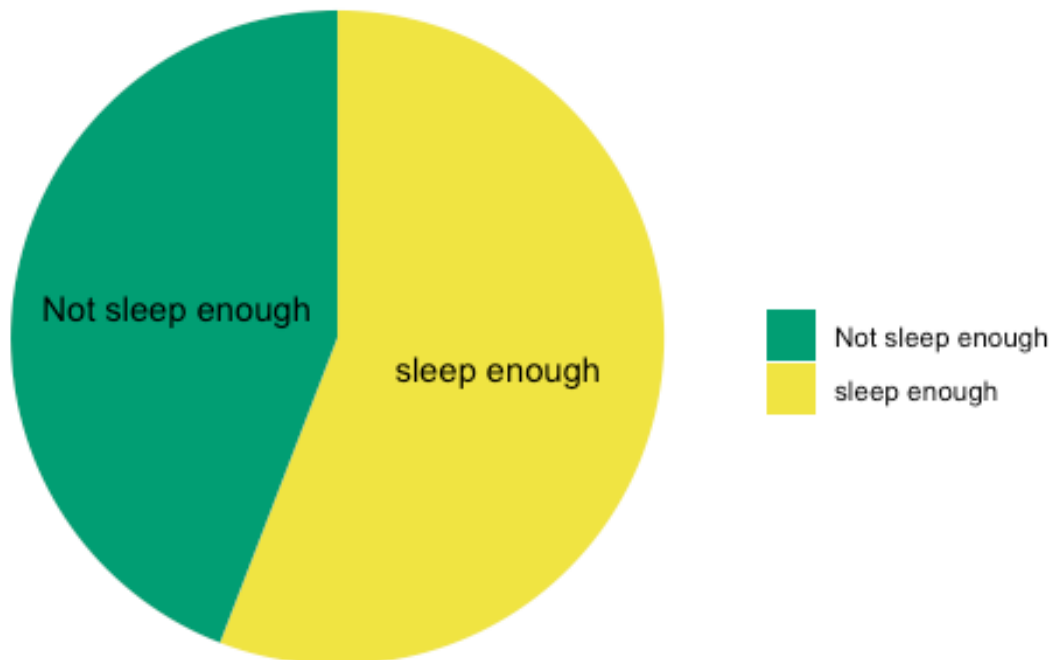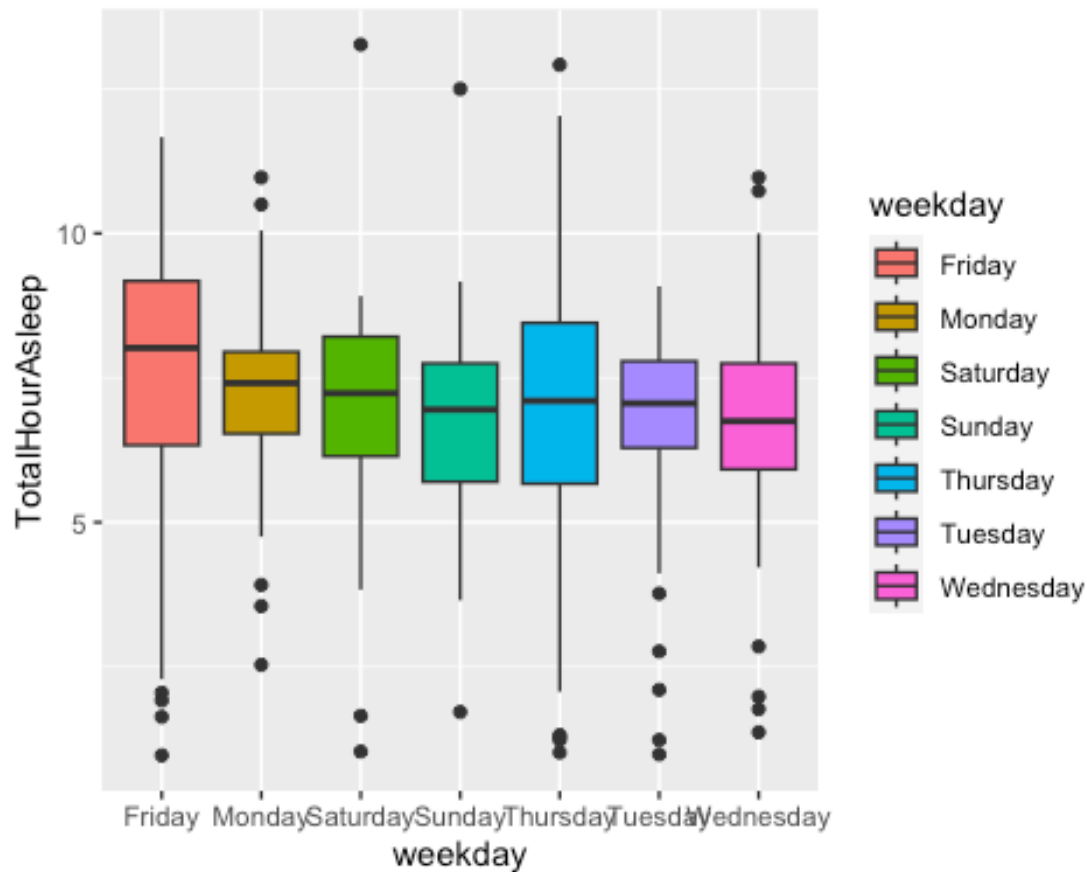
上圖可以得知有睡超過七小時的使用者佔多數，但人數相差不遠，也就是說使用者睡眠不足的情況頻繁發生。

使用者一週中睡眠時間多寡

```
ggplot(data = combined_data) +
  geom_boxplot(mapping = aes(x = weekday , y = TotalHourAsleep , fill = weekday))
```


由上圖可以觀察使用者一週中睡眠時間平均相差不遠，但在星期四與星期五睡眠時間範圍較廣。

使用者睡眠時間與消耗熱量之間的關係

```
ggplot(data = combined_data) +
  geom_point(mapping = aes(x = TotalHourAsleep , y = Calories,color = SleepCon)) +
  geom_smooth(aes(x = TotalHourAsleep , y = Calories),color = 'purple',alpha = 0.2)

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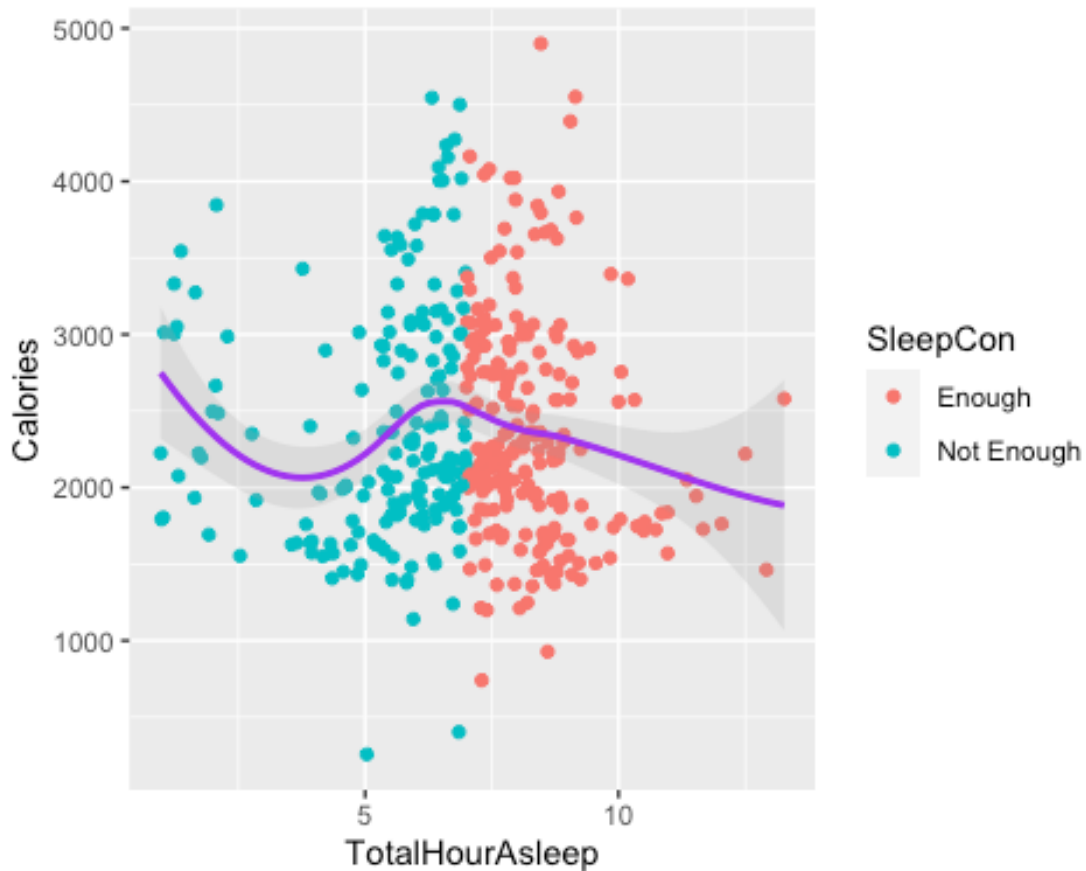

上圖可以得知使用者的睡眠時間大多集中在 5~10 個小時之間，而睡眠時間的多寡
與消耗的熱量並沒有太大的關係。

活動時間

活動型態佔活動百分比

```
Very_p <- sum(combined_data$VeryActiveMinutes)/sum(combined_data$TotalActiveMinute)
Fairly_p <- sum(combined_data$FairlyActiveMinutes)/sum(combined_data$TotalActiveMinute)
Lightly_p <- sum(combined_data$LightlyActiveMinutes)/sum(combined_data$TotalActiveMinute)
Sedentary_p <- sum(combined_data$SedentaryMinutes)/sum(combined_data$TotalActiveMinute)
Active_percent <- data.frame(category = c("Very","Fairly","Lightly","Sedentary") , value = c(Very_p,Fairly_p,Lightly_p,Sedentary_p))


ggplot(data = Active_percent , aes(x = "",y = value ,fill = category)) +
  geom_bar(stat = "identity")+
  scale_fill_manual(values = c( "#48D1CC", "#F0E68C","#6495ED","#DDA0DD")) +
  coord_polar(theta = "y") +
  theme_void() +
  labs(fill="") +
  geom_text(aes(label=category), position=position_stack(vjust=0.5),check_overlap = "TRUE") +
  labs(title = "Active Percent")
```

Active Percent
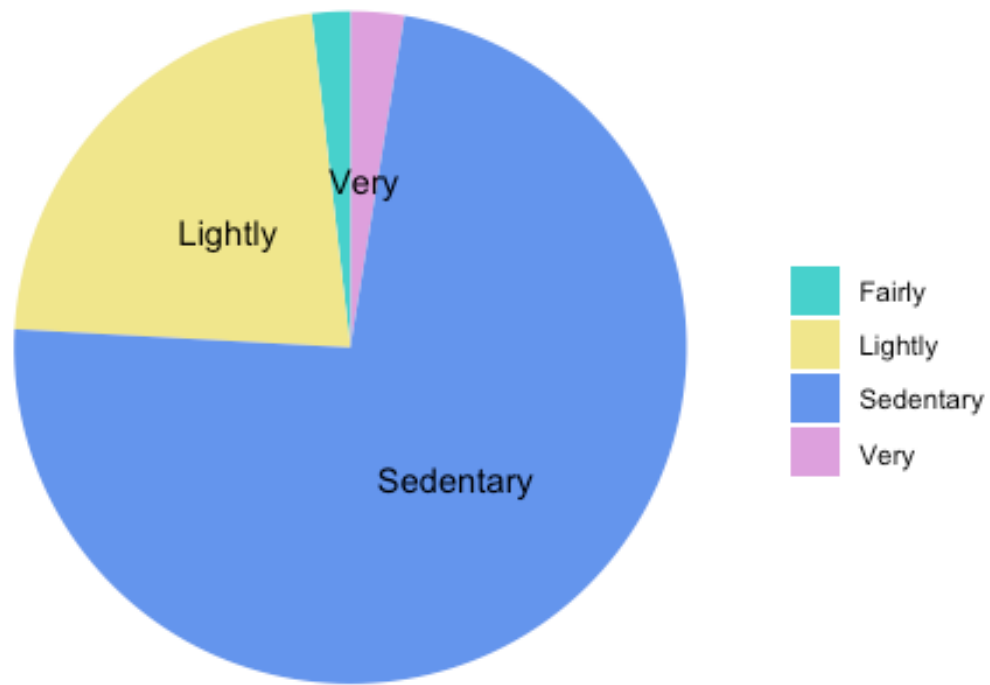
可以看到使用者大部分時間都處於久坐狀態或是輕度活動的狀態，較少進行中高強度的運動。

久坐時間與睡眠狀況

```
ggplot(data = combined_data, mapping = aes(x = weekday , y = SedentaryMinutes)) +
  geom_col(aes(fill = SleepCon),position = "stack") +
  scale_fill_manual(values = c("#E69F00", "#56B4E9", "#009E73", "#F0E442", "#0072B2")) +
  theme_minimal()
```
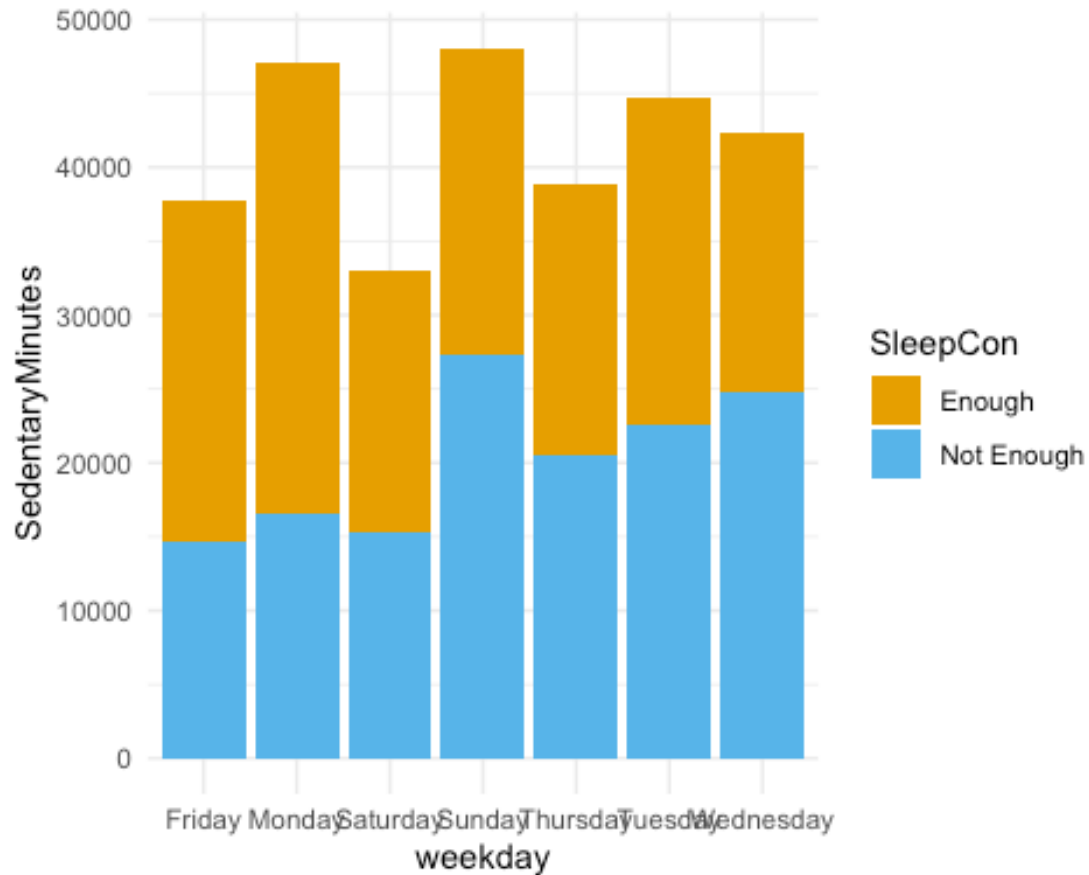

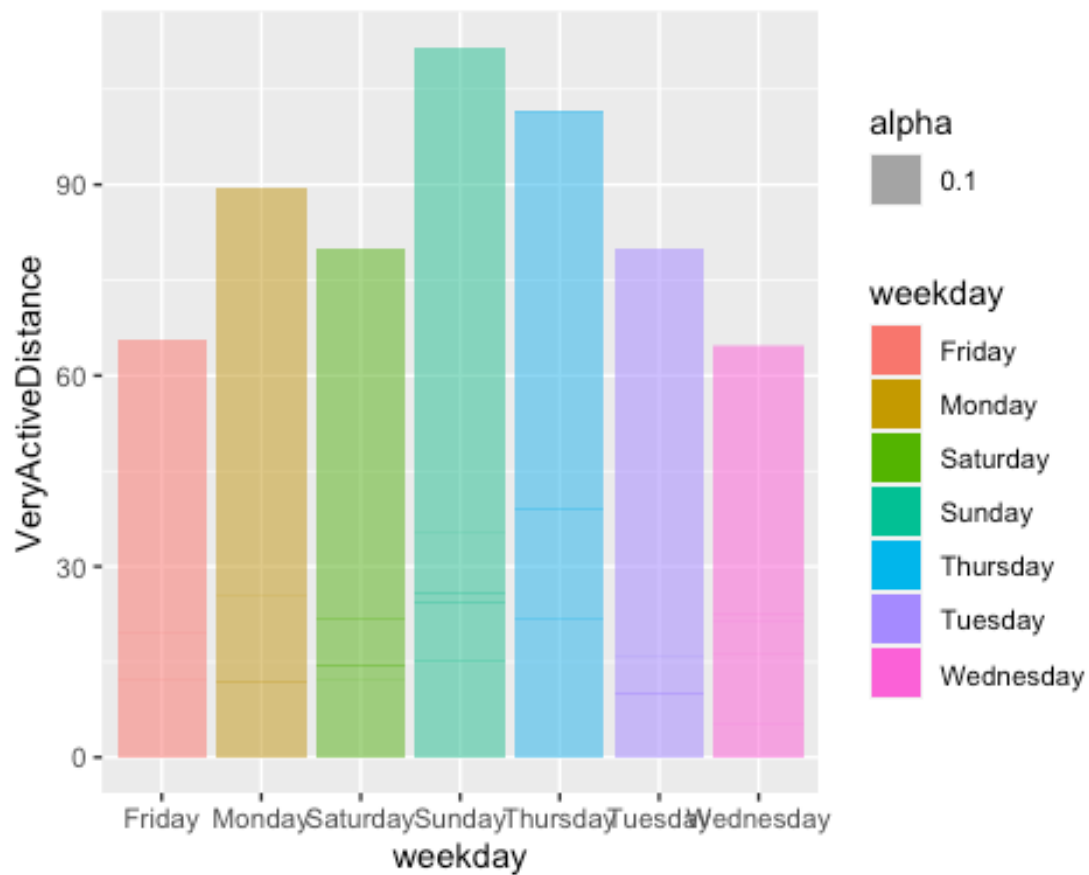
由上圖可以看到，久坐時間與睡眠時間沒有太大的相關性，以星期天為例，雖然久坐的時間很長，但睡眠不足的使用者還是佔多數。

```
ggplot(data = combined_data) +
  geom_col(mapping = aes(x =weekday,y = VeryActiveMinutes,fill = weekday,alpha = 0.1))
```
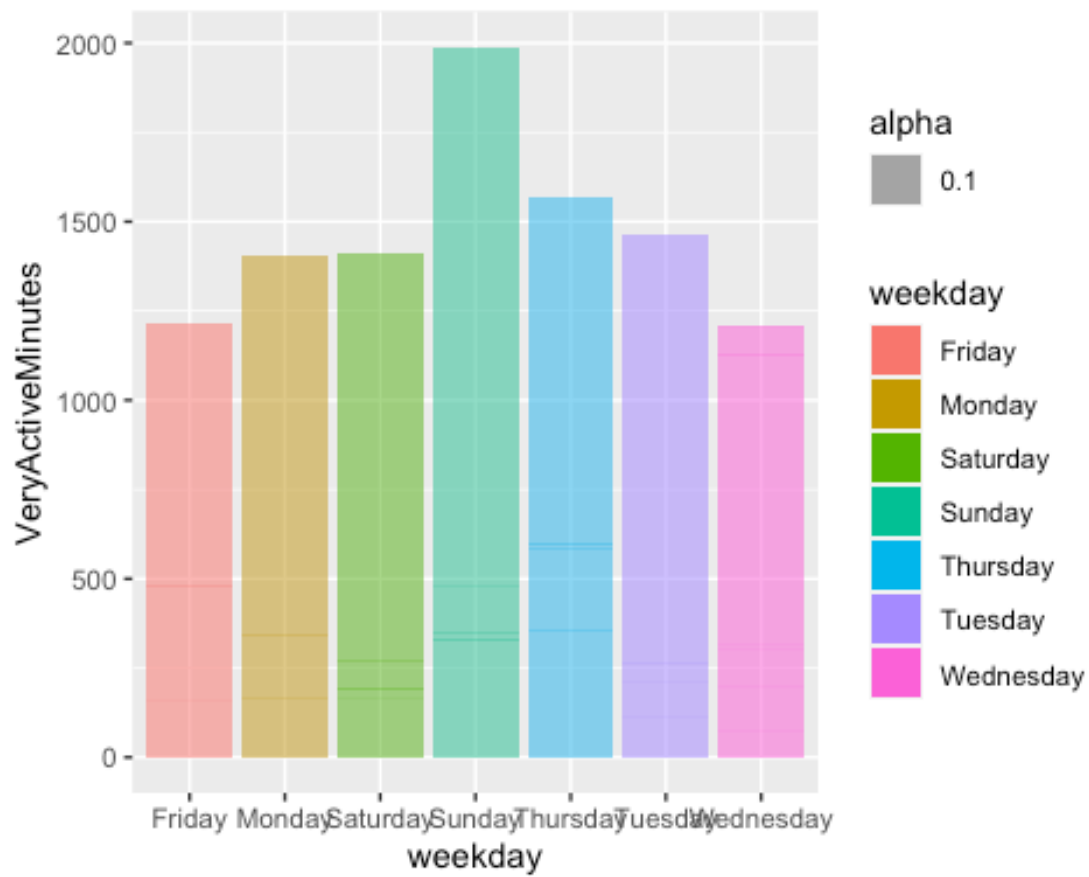
```
ggplot(data = combined_data) +
  geom_col(mapping = aes(x =weekday,y = VeryActiveDistance,fill = weekday,alpha = .1))
```
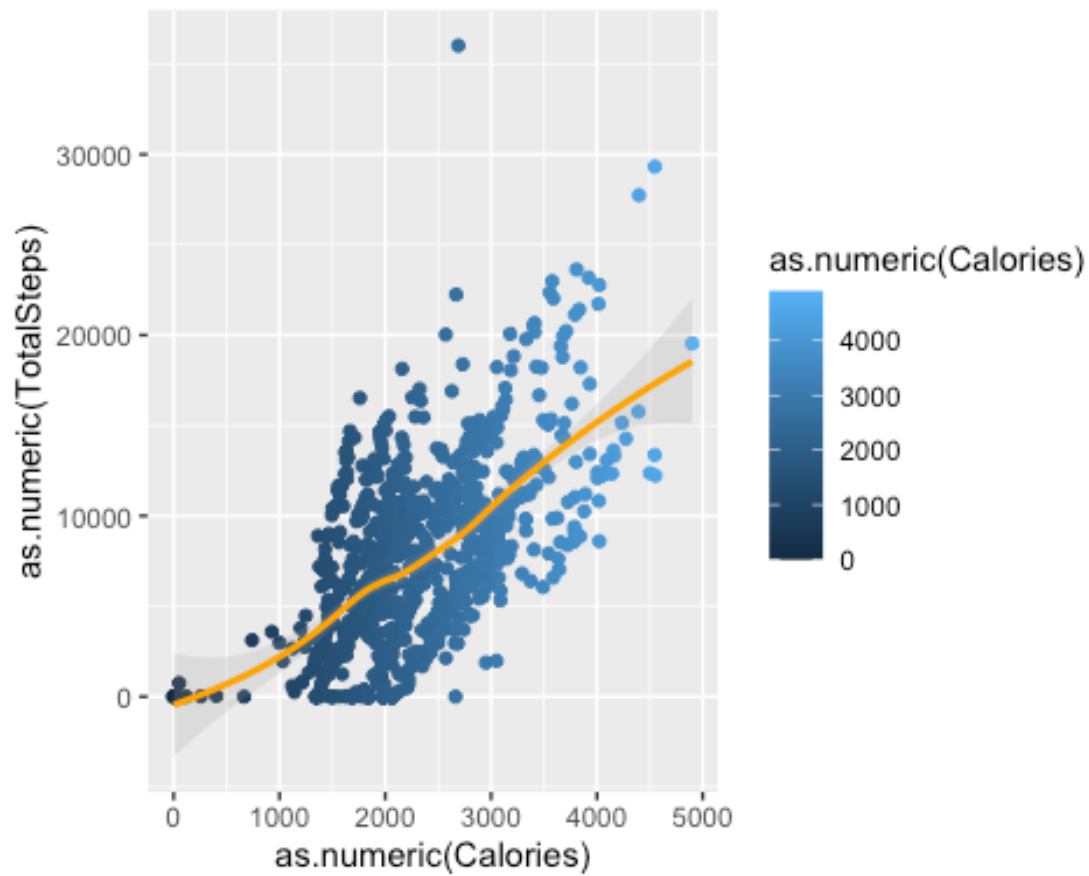

由以上兩張圖顯示，星期四與星期日使用者最常進行高強度運動。

## 消耗熱量分析

移動步數與消耗熱量的關係

```
ggplot(data = Active , aes(x = as.numeric(Calories) , y =as.numeric(TotalSteps))) +
        geom_point(mapping = aes(color =as.numeric(Calories))) +
        geom_smooth(color = 'orange',alpha = 0.2)
```

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

```
ggplot(data = Active , aes(x = as.numeric(Calories) , y =as.numeric(TotalDistance))) +
          geom_point(mapping = aes(color =as.numeric(Calories))) +
          geom_smooth(color = 'pink',alpha = 0.2)
```

`geom_smooth()` using method = 'loess' and formula = 'y ~ x'
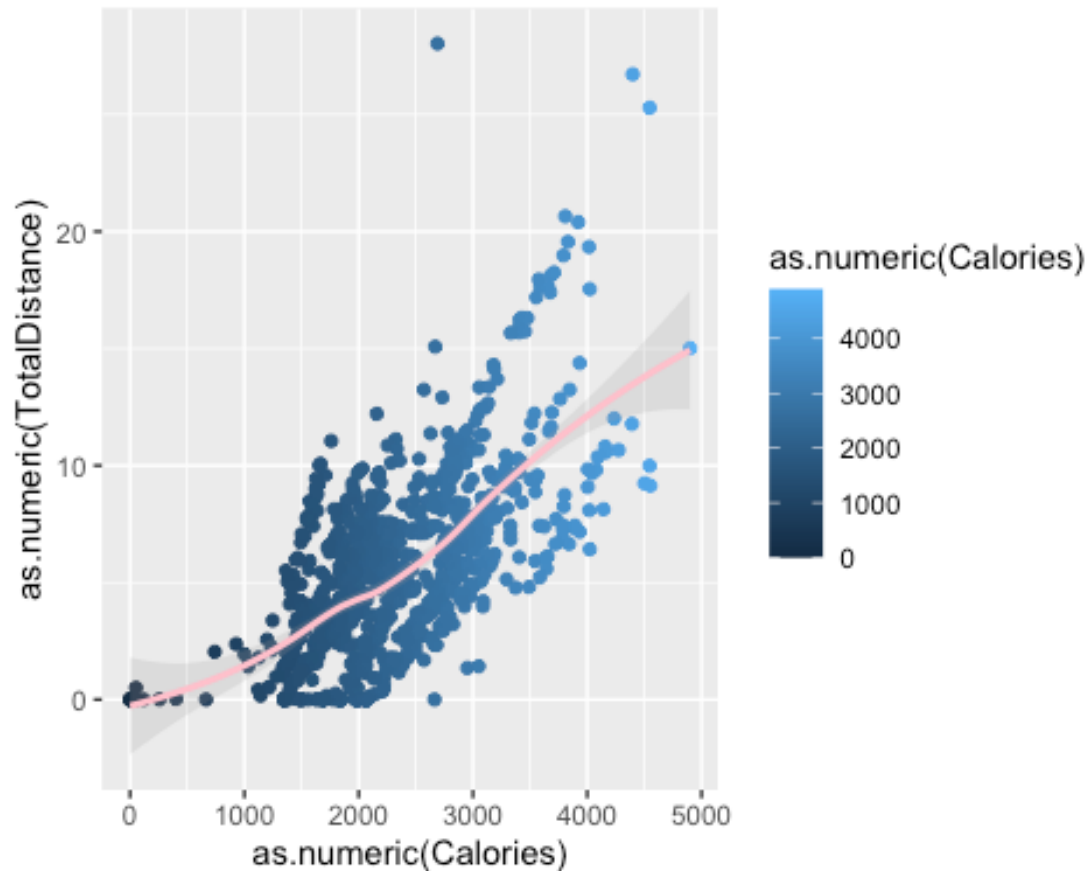
經由觀察趨勢線，可以發現移動步數、距離都與消耗的熱量成正向關係。

相關係數表

```
Picked_Active_1 <- Active %>%
  select(TotalSteps,Calories,TotalDistance)
Cor_Picked_Active_1 <- cor(Picked_Active_1)
head(Cor_Picked_Active_1)

##                TotalSteps  Calories TotalDistance
## TotalSteps      1.0000000 0.5915681     0.9853688
## Calories        0.5915681 1.0000000     0.6449619
## TotalDistance   0.9853688 0.6449619     1.0000000
```

library(corrplot)

corrplot 0.92 loaded

corrplot(Cor_Picked_Active_1,
      tl.col = "black",
      method = "color",
      col = colorRampPalette(c("#ffa500", "#FFFFFF", "#BC8F8F"))(100),
      addCoef.col = "black")
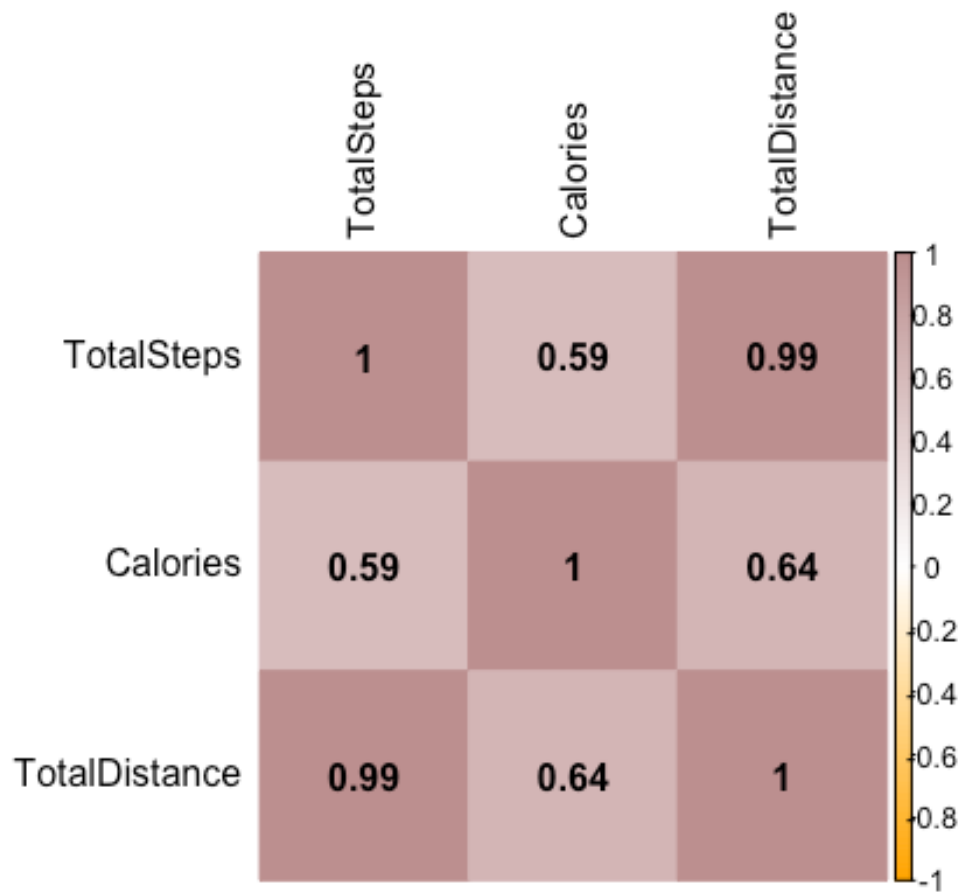
結論

根據分析結果得出了幾個結論：

1. 使用者的睡眠時間普遍不足
2. 睡眠時間與消耗的熱量以及久坐的時間較無相關性
3. 在活動時間中，使用者大多處於久坐或是輕度活動狀態
4. 使用者通常在星期四與星期日進行高強度運動
5. 移動的步數、距離都與消耗的熱量為正相關，其中「移動距離與消耗熱量的相關係數」較「移動步數與消耗熱量的相關係數」高，而移動的距離與步數具有高度相關性。

依照分析的結果，提出幾項可以改善的建議：

1. 針對使用者的睡眠時間進行優化，設定獎勵機制或是睡眠提醒促使使用者增加睡眠時間。
2. 使用者較少進行中高強度的運動，導致這結果的主要原因除了使用者不喜歡運動之外，另一個原因可能是因為裝置的關係，一般而言，在進行球類等高強度運動時，大多運動員會將佩戴型裝置拔下，因此公司可以研發別種型態的裝置類型。
3. 與健身房、舞蹈教室等運動課程合作推出優惠方案，鼓勵進行中高強度運動，並在 App 放置廣告以及推廣運動的優點。