CaseStudy2

Chuan Liu

2025-04-25

ASK

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?

PREPARE

1.Access data: This is public data from the website: https://www.kaggle.com/datasets/arashnic/fitbit, given by the class.

2.The data consists of 18 csv files: the files for recorded the activities time, calories consumption, sleep time and status and etc. from 2016/4 to 2016/5, but I am choosing to focus on these 6:

dailyActivity_merged.csv dailySteps_merged.csv sleepDay_merged.csv weightLogInfo_merged.csv

3.The data is first-party data collected by Fitbit, so there is a low chance of bias, but due to it being the company's own data the credibility is very high.

4.The data is open source and covered by licence : https://creativecommons.org/publicdomain/zero/1.0/

PROCESS

Library packages:

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

Load files:

```
daily_activity <- read.csv("/cloud/project/dailyActivity_merged.csv")
daily_steps <- read.csv("/cloud/project/dailySteps_merged.csv")
sleep_day <- read.csv("/cloud/project/sleepDay_merged.csv")
weight_info <- read.csv("/cloud/project/weightLogInfo_merged.csv")
```

Let's check the column name and cell type for each table:

```
head(daily_activity)
```

```
##           Id ActivityDate TotalSteps TotalDistance TrackerDistance
## 1 1503960366    4/12/2016      13162          8.50            8.50
```

```
## 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
##   LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 1                        0              1.88                     0.55
## 2                        0              1.57                     0.69
## 3                        0              2.44                     0.40
## 4                        0              2.14                     1.26
## 5                        0              2.71                     0.41
## 6                        0              3.19                     0.78
##   LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
## 1                6.06                       0                25
## 2                4.71                       0                21
## 3                3.91                       0                30
## 4                2.83                       0                29
## 5                5.04                       0                36
## 6                2.51                       0                38
##   FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories
## 1                  13                  328              728     1985
## 2                  19                  217              776     1797
## 3                  11                  181             1218     1776
## 4                  34                  209              726     1745
## 5                  10                  221              773     1863
## 6                  20                  164              539     1728
```

```
head(daily_steps)
```

```
##           Id ActivityDay StepTotal
## 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
```

```
head(sleep_day)
```

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

```
head(weight_info)
```

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

Cleaning the date for "sleep_day" and "weight_info" files:

```
sleep_day$ActivityDate <- parse_date_time(sleep_day$SleepDay, orders = "mdy HMS")
sleep_day$ActivityDate <- as.Date(sleep_day$ActivityDate)
weight_info$Date <- parse_date_time(weight_info$Date, orders = "mdy HMS")
weight_info$Date <- as.Date(weight_info$Date)
head(sleep_day)
head(weight_info)
```

Join all tables in to one big table:

```
daily_steps <- daily_steps %>% rename(ActivityDate = ActivityDay)
merge_1 <- merge(daily_activity, daily_steps, by = c("Id","ActivityDate"))
merge_1$ActivityDate <- as.Date(merge_1$ActivityDate, "%m/%d/%Y")
daily_total <- merge(merge_1, sleep_day, by = c("Id", "ActivityDate"))
```

Summary:

```
summary(daily_total)
```

```
##        Id             ActivityDate          TotalSteps    TotalDistance
##  Min.   :1.504e+09   Min.   :2016-04-12   Min.   :   17   Min.   : 0.010
##  1st Qu.:3.977e+09   1st Qu.:2016-04-19   1st Qu.: 5206   1st Qu.: 3.600
##  Median :4.703e+09   Median :2016-04-27   Median : 8925   Median : 6.290
##  Mean   :5.001e+09   Mean   :2016-04-26   Mean   : 8541   Mean   : 6.039
##  3rd Qu.:6.962e+09   3rd Qu.:2016-05-04   3rd Qu.:11393   3rd Qu.: 8.030
##  Max.   :8.792e+09   Max.   :2016-05-12   Max.   :22770   Max.   :17.540
##  TrackerDistance  LoggedActivitiesDistance VeryActiveDistance
##  Min.   : 0.010   Min.   :0.0000           Min.   : 0.00
##  1st Qu.: 3.600   1st Qu.:0.0000           1st Qu.: 0.00
##  Median : 6.290   Median :0.0000           Median : 0.57
##  Mean   : 6.034   Mean   :0.1131           Mean   : 1.45
##  3rd Qu.: 8.020   3rd Qu.:0.0000           3rd Qu.: 2.37
##  Max.   :17.540   Max.   :4.0817           Max.   :12.54
##  ModeratelyActiveDistance LightActiveDistance SedentaryActiveDistance
##  Min.   :0.0000           Min.   :0.010       Min.   :0.0000000
##  1st Qu.:0.0000           1st Qu.:2.540       1st Qu.:0.0000000
##  Median :0.4200           Median :3.680       Median :0.0000000
##  Mean   :0.7502           Mean   :3.807       Mean   :0.0009201
##  3rd Qu.:1.0400           3rd Qu.:4.930       3rd Qu.:0.0000000
##  Max.   :6.4800           Max.   :9.480       Max.   :0.1100000
##  VeryActiveMinutes FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes
```

```
##  Min.   :  0.00    Min.   :  0.00      Min.   :  2.0        Min.   :   0.0
##  1st Qu.:  0.00    1st Qu.:  0.00      1st Qu.:158.0        1st Qu.: 631.0
##  Median :  9.00    Median : 11.00      Median :208.0        Median : 717.0
##  Mean   : 25.19    Mean   : 18.04      Mean   :216.9        Mean   : 712.2
##  3rd Qu.: 38.00    3rd Qu.: 27.00      3rd Qu.:263.0        3rd Qu.: 783.0
##  Max.   :210.00    Max.   :143.00      Max.   :518.0        Max.   :1265.0
##     Calories      StepTotal       SleepDay         TotalSleepRecords
##  Min.   : 257   Min.   :   17   Length:413         Min.   :1.000
##  1st Qu.:1850   1st Qu.: 5206   Class :character   1st Qu.:1.000
##  Median :2220   Median : 8925   Mode  :character   Median :1.000
##  Mean   :2398   Mean   : 8541                      Mean   :1.119
##  3rd Qu.:2926   3rd Qu.:11393                      3rd Qu.:1.000
##  Max.   :4900   Max.   :22770                      Max.   :3.000
##  TotalMinutesAsleep TotalTimeInBed
##  Min.   : 58.0      Min.   : 61.0
##  1st Qu.:361.0      1st Qu.:403.0
##  Median :433.0      Median :463.0
##  Mean   :419.5      Mean   :458.6
##  3rd Qu.:490.0      3rd Qu.:526.0
##  Max.   :796.0      Max.   :961.0
```

Check the sleeping categories percent for each individual user type:

```
sleep_and_user_type <- daily_total %>%
group_by(Id) %>%
reframe(
user_type = factor(case_when(
    SedentaryMinutes > mean(SedentaryMinutes) &
    LightlyActiveMinutes < mean(LightlyActiveMinutes) &
    FairlyActiveMinutes < mean(FairlyActiveMinutes) &
    VeryActiveMinutes < mean(VeryActiveMinutes) ~ "Sedentary",
    SedentaryMinutes < mean(SedentaryMinutes) &
    LightlyActiveMinutes > mean(LightlyActiveMinutes) &
    FairlyActiveMinutes < mean(FairlyActiveMinutes) &
    VeryActiveMinutes < mean(VeryActiveMinutes) ~ "Lightly Active",
    SedentaryMinutes < mean(SedentaryMinutes) &
    LightlyActiveMinutes < mean(LightlyActiveMinutes) &
    FairlyActiveMinutes > mean(FairlyActiveMinutes) &
    VeryActiveMinutes < mean(VeryActiveMinutes) ~ "Fairly Active",
    SedentaryMinutes < mean(SedentaryMinutes) &
    LightlyActiveMinutes < mean(LightlyActiveMinutes) &
    FairlyActiveMinutes < mean(FairlyActiveMinutes) &
    VeryActiveMinutes > mean(VeryActiveMinutes) ~ "Very Active",
),levels=c("Sedentary", "Lightly Active", "Fairly Active", "Very Active")),
sleep_type = factor(case_when(
    mean(TotalMinutesAsleep) < 360 ~ "Bad Sleep",
    mean(TotalMinutesAsleep) > 360 & mean(TotalMinutesAsleep) <= 480 ~ "Normal Sleep",
    mean(TotalMinutesAsleep) > 480 ~ "Over Sleep",
),levels=c("Bad Sleep", "Normal Sleep", "Over Sleep")),
total_sleep = sum(TotalMinutesAsleep)) %>%
drop_na() %>%
group_by(user_type) %>%
summarise(
  bad_sleepers = sum(sleep_type == "Bad Sleep"),
  normal_sleepers = sum(sleep_type == "Normal Sleep"),
```

```
  over_sleepers = sum(sleep_type == "Over Sleep"),
  total=n(),
  total_sleep = sum(total_sleep)) %>%
group_by(user_type) %>%
summarise(
    bad_sleepers = bad_sleepers / total,
    normal_sleepers = normal_sleepers / total,
    over_sleepers = over_sleepers / total,
    total_sleep = total_sleep)
```

Plot total steps within each week day:

```
daily_steps$ActivityDate <- as.Date(daily_steps$ActivityDate, format = "%m/%d/%Y")
step_summary <- daily_steps %>%
  group_by(ActivityDate) %>%
  summarise(total_steps = sum(StepTotal, na.rm = TRUE))
step_summary <- step_summary %>%
  mutate(
    day_label = format(ActivityDate, "%d"),
    month_label = format(ActivityDate, "%b"),
    week_day = format(ActivityDate, "%a")
  )
```

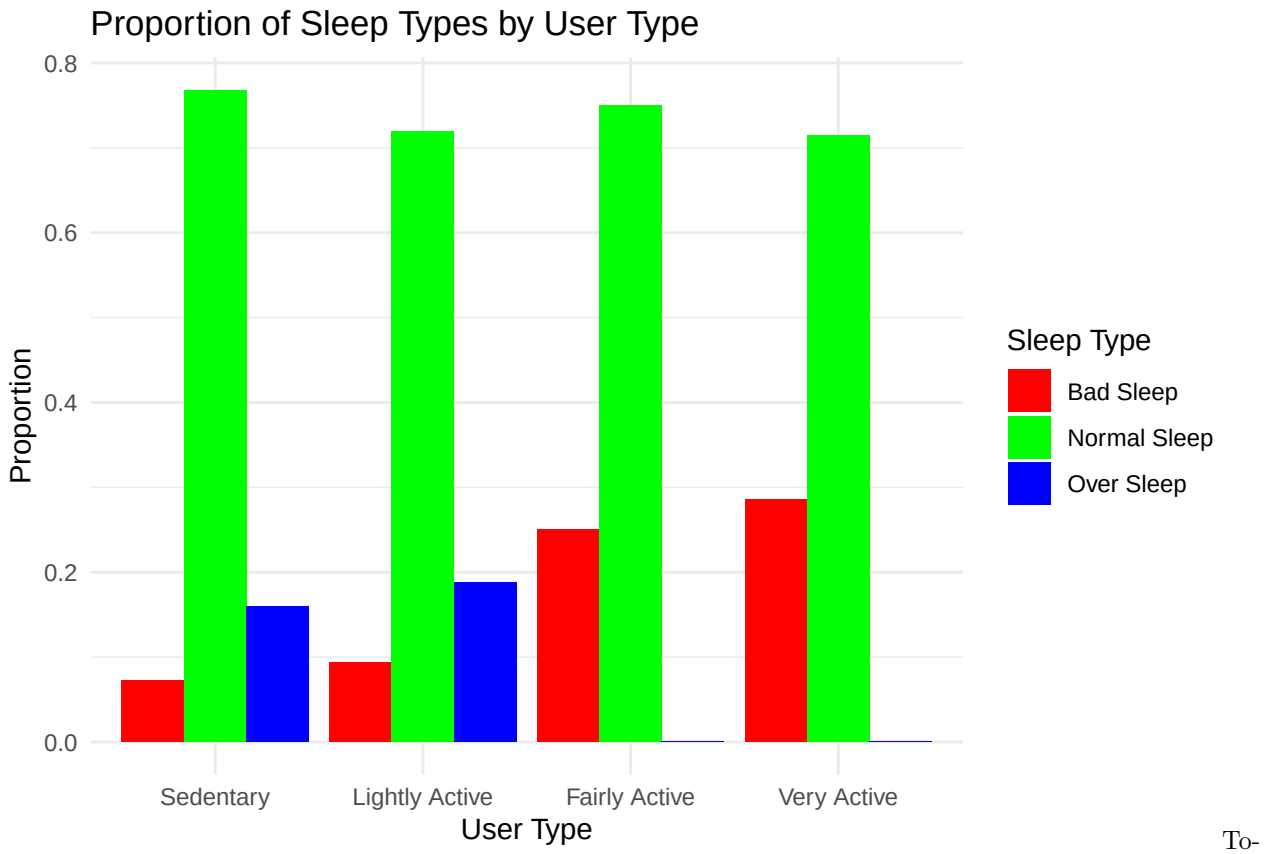
Show the relationship between sleep time and calories:

```
daily_activity$ActivityDate <- as.Date(daily_activity$ActivityDate, format = "%m/%d/%Y")
sleep_day$ActivityDate <- as.Date(sleep_day$SleepDay, format = "%m/%d/%Y")
sleep_calories <- daily_activity %>%
  select(Id, ActivityDate, Calories) %>%
  inner_join(
    sleep_day %>% select(Id, ActivityDate, TotalMinutesAsleep),
    by = c("Id", "ActivityDate")
  )
```

ANALYZE

Sleep Type VS Activity Status

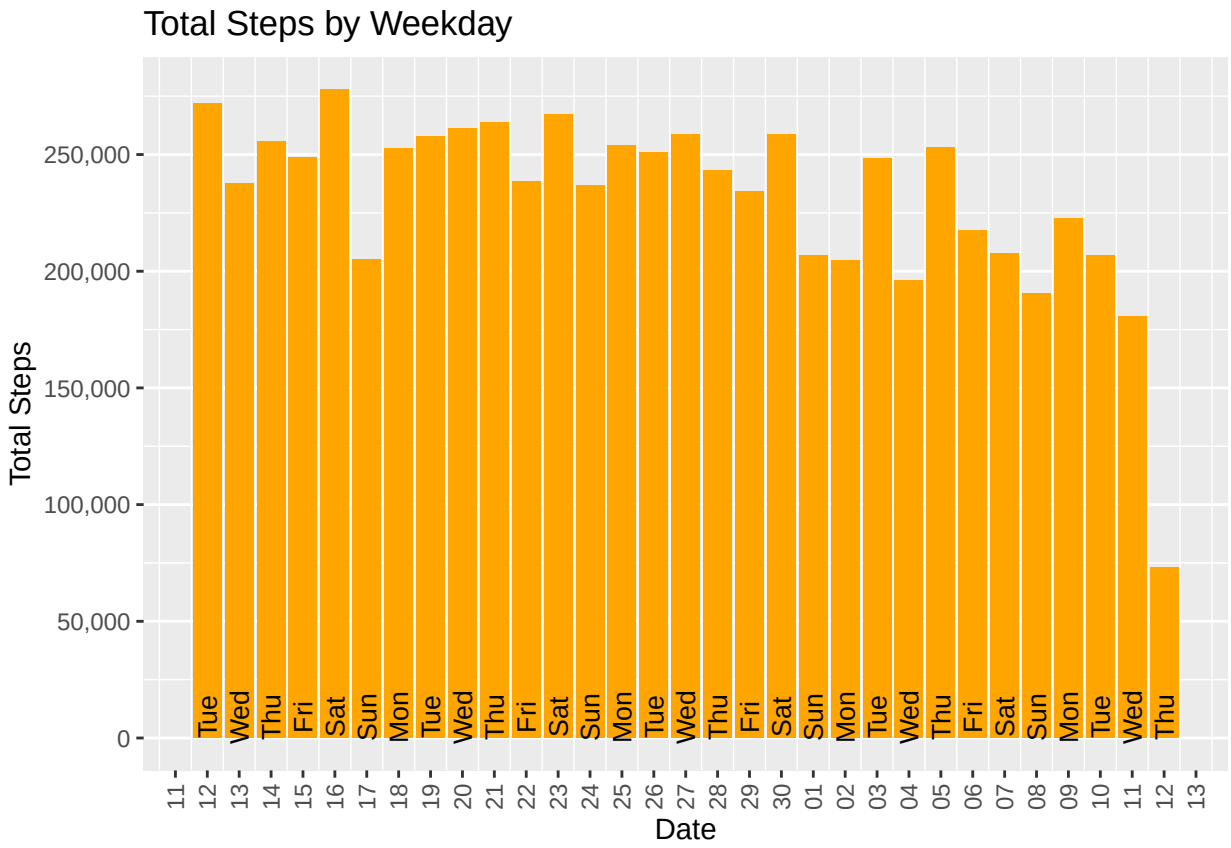
```
sleep_long <- sleep_and_user_type %>%
  pivot_longer(cols = c(bad_sleepers, normal_sleepers, over_sleepers),
               names_to = "sleep_type",
               values_to = "proportion")
sleep_long$sleep_type <- factor(sleep_long$sleep_type,
                                levels = c("bad_sleepers", "normal_sleepers", "over_sleepers"),
                                labels = c("Bad Sleep", "Normal Sleep", "Over Sleep"))
ggplot(sleep_long, aes(x = user_type, y = proportion, fill = sleep_type)) +
  geom_bar(stat = "identity", position = "dodge") +
  scale_fill_manual(values = c("Bad Sleep" = "red", "Normal Sleep" = "green", "Over Sleep" = "blue")) +
  labs(title = "Proportion of Sleep Types by User Type",
       x = "User Type", y = "Proportion", fill = "Sleep Type") +
  theme_minimal()
```



To-

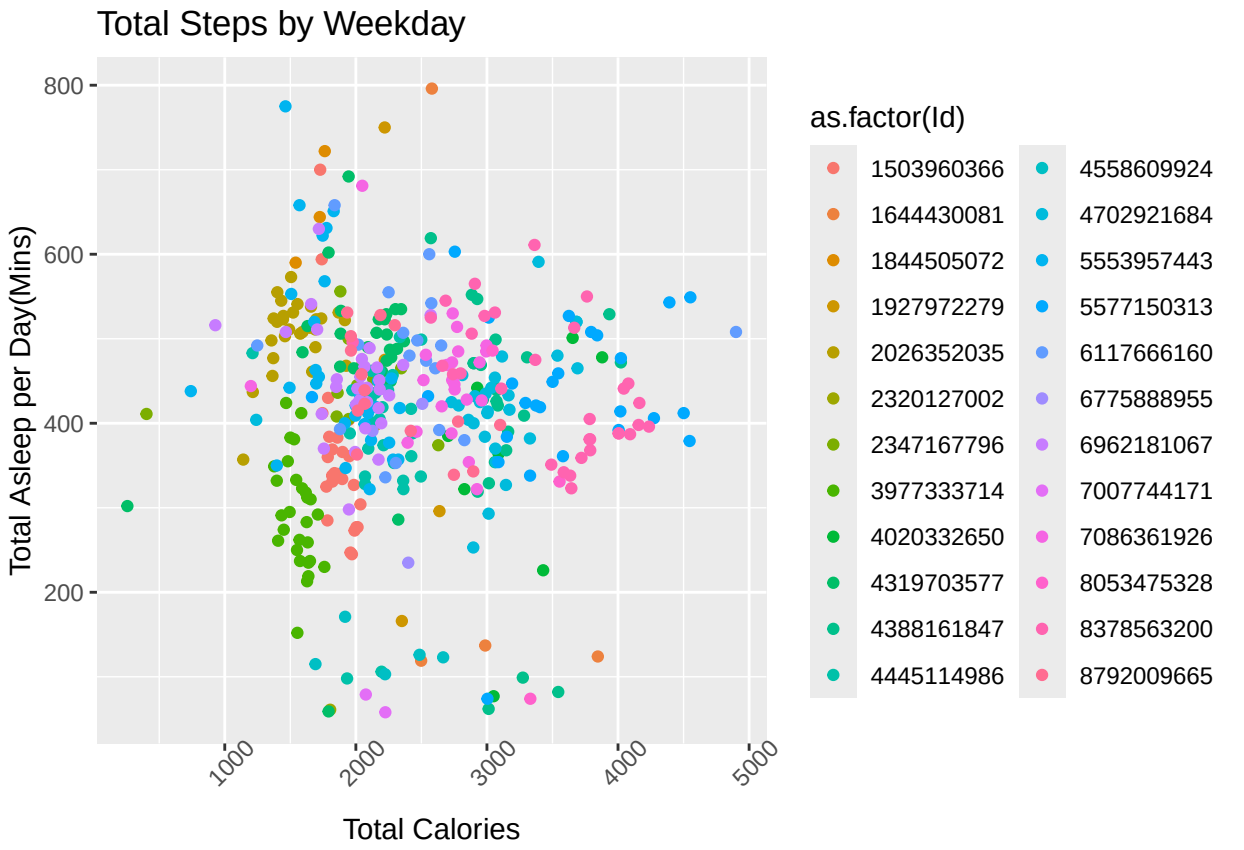
tal steps by weekday:

```
ggplot(data = step_summary, aes (x = ActivityDate, y = total_steps)) +
   geom_bar(stat = "identity", fill = "orange") +
   geom_text(aes(label = week_day), vjust = 0.5, hjust = 0.5, color = "black", size = 3.5,
             angle = 90, y = 9000) +
     theme(
     axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1),
     panel.grid.major.x = element_blank()
   ) +
      labs(x = "Date", y = "Total Steps", title = "Total Steps by Weekday") +
   scale_x_date(
     date_breaks = "1 day",
     date_labels = "%d"
   ) +
    scale_y_continuous(labels = scales::comma, breaks = pretty_breaks(n = 5))
```



Sleep time with calories plot:

```
ggplot(data = sleep_calories, aes(x = Calories, y = TotalMinutesAsleep, color = as.factor(Id),
                                  group = Id))+
  geom_point()+
  theme(axis.text.x = element_text(angle = 45))+
  labs(x = "Total Calories", y = "Total Asleep per Day(Mins)", title = "Total Steps by Weekday")
```



SHARE

1. What are some trends in smart device usage?

Both insufficient and excessive exercise can affect the quality of our sleep.

2. How could these trends apply to Bellabeat customers?

While supervising users who do not exercise enough, it is also necessary to remind those who exercise excessively.

3. How could these trends help influence Bellabeat marketing strategy?

We can not only target on insufficient exercise user, but also some Sports expert. To help both of them adjust their physical conditions and sleep quality.

ACT

1. Increase awareness of the effects of excessive exercise and target marketing to people who exercise regularly.
2. In our products, improve the definition of excessive movement and monitoring mechanism.