Capstone: Bellabeat Case Study

Upasana Sharma

3/16/2022

Introduction:

This is a case study of a company 'Bellabeat' which is a high-tech manufacturer of health focused products for women. The stakeholders of the company want to analyze smart device fitness data could help unlock new growth opportunities for the company.

Product - Time (Bellabeat Smart Watch):

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.

Business Task:

- To recommend the marketing strategy/high-level recommendation for the one of the Bellabeat smart device (smart watch in this case) based on the analysis of Fitbit users' data.
- Analyse trends of another smart device data to gain insights on non-bellabeat users.

Stakeholders:

- Urska Srsen: Bellabeat's cofounder and Chief Creative officer.
- Sando Mur: Bellabeat's cofounder
- Marketing Analytics Team

Ask:

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

Database used:

Kaggle public dataset- Fitbit Fitness Tracker Data (This database has 18 datasets, I have used 3 of them daily_activity, sleep_day and weight_log datasets for my analysis.)

Data Limitations and biases:

- The dataset has only 33 fitbit users and the gender ratio is unknown.
- This data is an year old at time time of this analysis, therefore the trends observed might vary from the current ones.
- No data available for a particular smart device trends.

Preparing the data:

Used R language and Rstudio for this analysis.

Loading all necessary libraries

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(dplyr)
library(ggplot2)
```

Reading CSV files into variables

```
daily_activity <- read.csv("C:/Data/dailyActivity_merged.csv")
sleep_day <- read.csv("C:/Data/sleepDay_merged.csv")
weightlog_info <- read.csv("C:/Data/weightLogInfo_merged.csv")
```

Exploring Tables

Table - dailyActivity_merged

```
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
```

```
str(daily_activity)
```

```
## 'data.frame':    940 obs. of  15 variables:
##  $ Id                       : num  1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ ActivityDate             : chr  "4/12/2016" "4/13/2016" "4/14/2016" "4/15/2016" ...
##  $ TotalSteps               : int  13162 10735 10460 9762 12669 9705 13019 15506 10544 9819 ...
##  $ TotalDistance            : num  8.5 6.97 6.74 6.28 8.16 ...
##  $ TrackerDistance          : num  8.5 6.97 6.74 6.28 8.16 ...
##  $ LoggedActivitiesDistance: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ VeryActiveDistance       : num  1.88 1.57 2.44 2.14 2.71 ...
##  $ ModeratelyActiveDistance: num  0.55 0.69 0.4 1.26 0.41 ...
##  $ LightActiveDistance      : num  6.06 4.71 3.91 2.83 5.04 ...
##  $ SedentaryActiveDistance : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ VeryActiveMinutes        : int  25 21 30 29 36 38 42 50 28 19 ...
##  $ FairlyActiveMinutes      : int  13 19 11 34 10 20 16 31 12 8 ...
##  $ LightlyActiveMinutes     : int  328 217 181 209 221 164 233 264 205 211 ...
##  $ SedentaryMinutes         : int  728 776 1218 726 773 539 1149 775 818 838 ...
##  $ Calories                 : int  1985 1797 1776 1745 1863 1728 1921 2035 1786 1775 ...
```

Table - sleepDay

```
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
```

```
str(sleep_day)
```

```
## 'data.frame':    413 obs. of  5 variables:
##  $ Id                : num  1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ SleepDay          : chr  "4/12/2016 12:00:00 AM" "4/13/2016 12:00:00 AM" "4/15/2016 12:00:00
AM" "4/16/2016 12:00:00 AM" ...
##  $ TotalSleepRecords : int  1 2 1 2 1 1 1 1 1 1 ...
##  $ TotalMinutesAsleep: int  327 384 412 340 700 304 360 325 361 430 ...
##  $ TotalTimeInBed    : int  346 407 442 367 712 320 377 364 384 449 ...
```

Table - weightInfo

```
head(weightlog_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
```

```
str(weightlog_info)
```

```
## 'data.frame':    67 obs. of  8 variables:
##  $ Id            : num  1.50e+09 1.50e+09 1.93e+09 2.87e+09 2.87e+09 ...
##  $ Date          : chr  "5/2/2016 11:59:59 PM" "5/3/2016 11:59:59 PM" "4/13/2016 1:08:52 AM"
"4/21/2016 11:59:59 PM" ...
##  $ WeightKg      : num  52.6 52.6 133.5 56.7 57.3 ...
##  $ WeightPounds  : num  116 116 294 125 126 ...
##  $ Fat           : int  22 NA NA NA NA 25 NA NA NA NA ...
##  $ BMI           : num  22.6 22.6 47.5 21.5 21.7 ...
##  $ IsManualReport: chr  "True" "True" "False" "True" ...
##  $ LogId         : num  1.46e+12 1.46e+12 1.46e+12 1.46e+12 1.46e+12 ...
```

Understanding statistics summary of all three tables

```
n_distinct(daily_activity$Id)
```

```
## [1] 33
```

```
n_distinct(sleep_day$Id)
```

```
## [1] 24
```

```
n_distinct(weightlog_info$Id)
```

```
## [1] 8
```

Processing the data -

Changing the date format - daily_activity

```
daily_activity$ActivityDate = mdy(daily_activity$ActivityDate)
```

Adding a new column ‘day_name’ for the corresponding dates

```
daily_act <- daily_activity %>%
  mutate(day_name = wday(daily_activity$ActivityDate, label = TRUE))
```

Creating a new daily activity table summarized by days of the week

```
days_daily_activity <- daily_act %>%
  group_by(day_name) %>%
  summarise(total_steps = sum(TotalSteps), total_distance = sum(TotalDistance), total_calories = s
um(Calories), Very_active_dist = sum(VeryActiveDistance),
            mod_active_dist = sum(ModeratelyActiveDistance), light_active_dist = sum(LightActiveDi
stance), sed_active_dist = sum(SedentaryActiveDistance),
            very_active_min = sum(VeryActiveMinutes), fairly_active_min = sum(FairlyActiveMinute
s), lightly_active_mins = sum(LightlyActiveMinutes),
            sed_min = sum(SedentaryMinutes)) %>%
  select(day_name, total_steps, total_distance, total_calories, Very_active_dist, mod_active_dist,
light_active_dist, sed_active_dist, very_active_min,
         fairly_active_min, lightly_active_mins, sed_min)
```

View the created table

```
print(days_daily_activity)
```

```
## # A tibble: 7 x 12
##   day_name total_steps total_distance total_calories Very_active_dist
##   <ord>          <int>          <dbl>          <int>            <dbl>
## 1 Sun           838921           608.         273823             180.
## 2 Mon           933704           666.         278905             184.
## 3 Tue          1235001           886.         358114             245.
## 4 Wed          1133906           823.         345393             245.
## 5 Thu          1088658           781.         323337             204.
## 6 Fri           938477           669.         293805             165.
## 7 Sat          1010969           726.         292016             188.
## # ... with 7 more variables: mod_active_dist <dbl>, light_active_dist <dbl>,
## #   sed_active_dist <dbl>, very_active_min <int>, fairly_active_min <int>,
## #   lightly_active_mins <int>, sed_min <int>
```

Changing the format of date - sleep_Day

```
sleep_day$SleepDay = mdy_hms(sleep_day$SleepDay)
```

Adding a new column ‘day_name’ for the corresponding dates

```
sleep_d <- sleep_day %>%
  mutate(day_name = wday(SleepDay, label = TRUE))
```

Creating a new sleep_day table summarized by days of the week

```
days_sleep_day <- sleep_d %>%
  group_by(day_name) %>%
  summarise(tot_sleep_rec = sum(TotalSleepRecords), tot_min_asleep = sum(TotalMinutesAsleep), tot_
time_bed = sum(TotalTimeInBed)) %>%
  select(day_name, tot_sleep_rec, tot_min_asleep, tot_time_bed)
```

View the created table

```
print(days_sleep_day)
```

```
## # A tibble: 7 x 4
##   day_name tot_sleep_rec tot_min_asleep tot_time_bed
##   <ord>            <int>          <int>        <int>
## 1 Sun                 65          24901        27693
## 2 Mon                 52          19685        21440
## 3 Tue                 72          26295        28814
## 4 Wed                 76          28689        31022
## 5 Thu                 67          26154        28327
## 6 Fri                 61          23109        25368
## 7 Sat                 69          24407        26754
```

Combining days_daily_activity and days_sleep_day on 'day_name'

```
combined_activity_sleep <- left_join(days_daily_activity, days_sleep_day, 'day_name')
```

View the combined table

```
print(combined_activity_sleep)
```

```
## # A tibble: 7 x 15
##   day_name total_steps total_distance total_calories Very_active_dist
##   <ord>          <int>          <dbl>          <int>             <dbl>
## 1 Sun           838921           608.         273823              180.
## 2 Mon           933704           666.         278905              184.
## 3 Tue          1235001           886.         358114              245.
## 4 Wed          1133906           823.         345393              245.
## 5 Thu          1088658           781.         323337              204.
## 6 Fri           938477           669.         293805              165.
## 7 Sat          1010969           726.         292016              188.
## # ... with 10 more variables: mod_active_dist <dbl>, light_active_dist <dbl>,
## #   sed_active_dist <dbl>, very_active_min <int>, fairly_active_min <int>,
## #   lightly_active_mins <int>, sed_min <int>, tot_sleep_rec <int>,
## #   tot_min_asleep <int>, tot_time_bed <int>
```

Changing the format of date - weightlog_info

```
weightlog_info$Date = mdy_hms(weightlog_info$Date)
```

Adding a new column 'day_name' for the corresponding dates

```
weightlog_info <- weightlog_info %>%
  mutate(day_name = wday(Date, label = TRUE))
```

Analysing the Data -

Creating a summarized daily_activity table by Id

```
id_daily_activity <- daily_activity %>%
  group_by(Id) %>%
  summarise(no_days = max(ActivityDate) - min(ActivityDate), total_steps = sum(TotalSteps), total_
distance = sum(TotalDistance), tracker_dis = sum(TrackerDistance),
            log_activities_dist = sum(LoggedActivitiesDistance), Very_active_dist = sum(VeryActive
Distance), mod_active_dist = sum(ModeratelyActiveDistance),
            light_active_dist = sum(LightActiveDistance), sed_active_dist = sum(SedentaryActiveDis
tance), very_active_min = sum(VeryActiveMinutes),
            fairly_active_min = sum(FairlyActiveMinutes), lightly_active_mins = sum(LightlyActiveM
inutes), sed_min = sum(SedentaryMinutes),
            total_calories = sum(Calories)) %>%
  select(Id, no_days, total_steps, total_distance, tracker_dis, log_activities_dist, Very_active_d
ist, mod_active_dist, light_active_dist, sed_active_dist,
         very_active_min, fairly_active_min, lightly_active_mins, sed_min, total_calories)
```

Creating a summarized sleep_day table by Id

```
id_sleep_day <- sleep_day %>%
  group_by(Id) %>%
  summarise(no_days = (max(SleepDay) - min(SleepDay)) / 86400,  total_sleep_rec = sum(TotalSleepRe
cords), total_min_asleep = sum(TotalMinutesAsleep),
            total_time_bed = sum(TotalTimeInBed)) %>%
  select(Id, no_days, total_sleep_rec, total_min_asleep, total_time_bed)
```

Creating a summarized weightlog_info table by Id

```
id_weightlog_info <- weightlog_info %>%
  group_by(Id) %>%
  summarise(no_days = (max(Date) - min(Date)) / 86400, max_weight = max(WeightKg), min_weight = mi
n(WeightKg), max_fat = max(Fat), min_fat = min(Fat),
            max_bmi = max(BMI), min_bmi = min(BMI)) %>%
  select(Id, no_days, max_weight, min_weight, max_fat, min_fat, max_bmi, min_bmi)
```

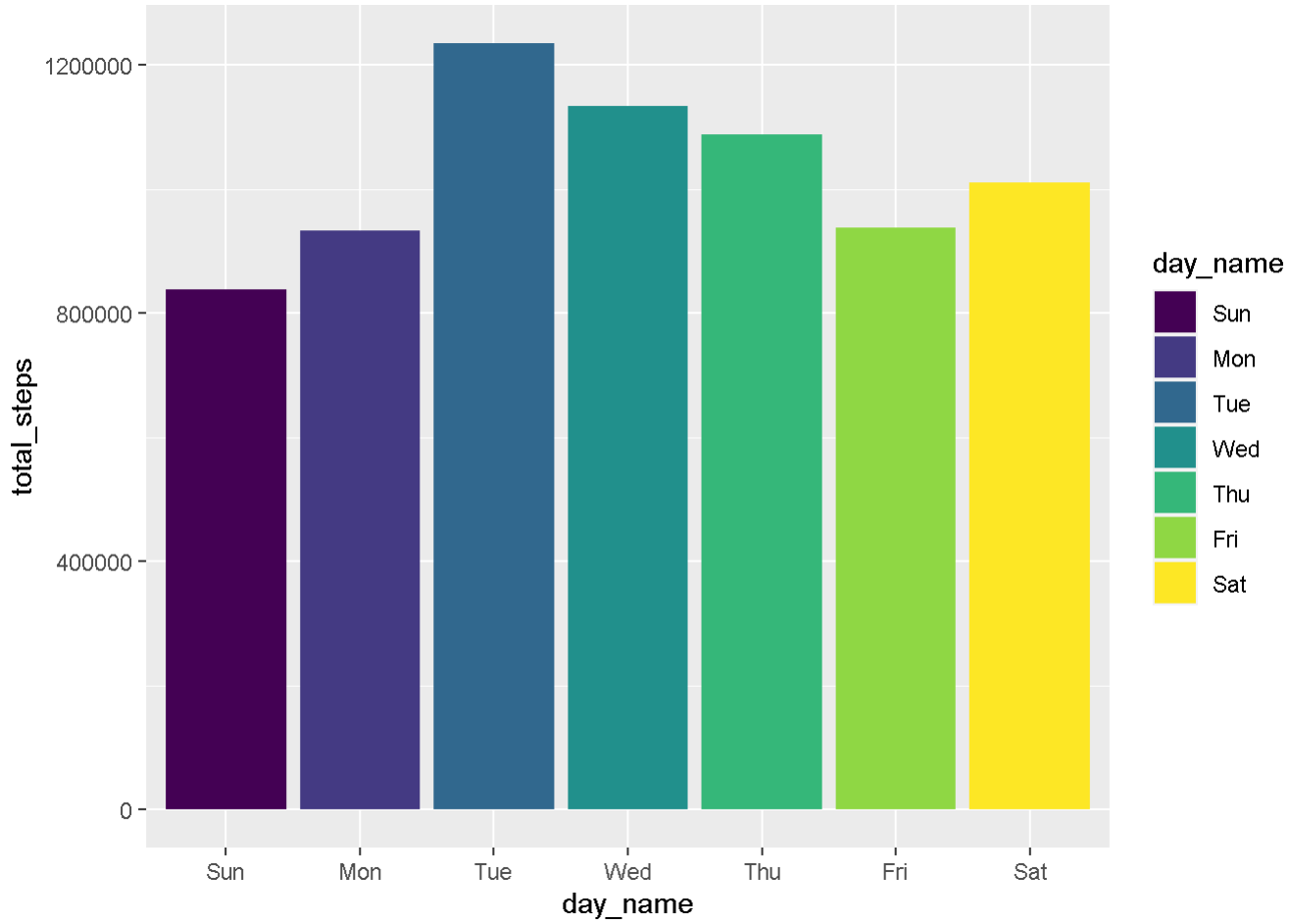
View tables created by Id

```
View(id_daily_activity)
View(id_sleep_day)
View(id_weightlog_info)
```

Visualizing the Data -

total_steps Vs. day_of_week

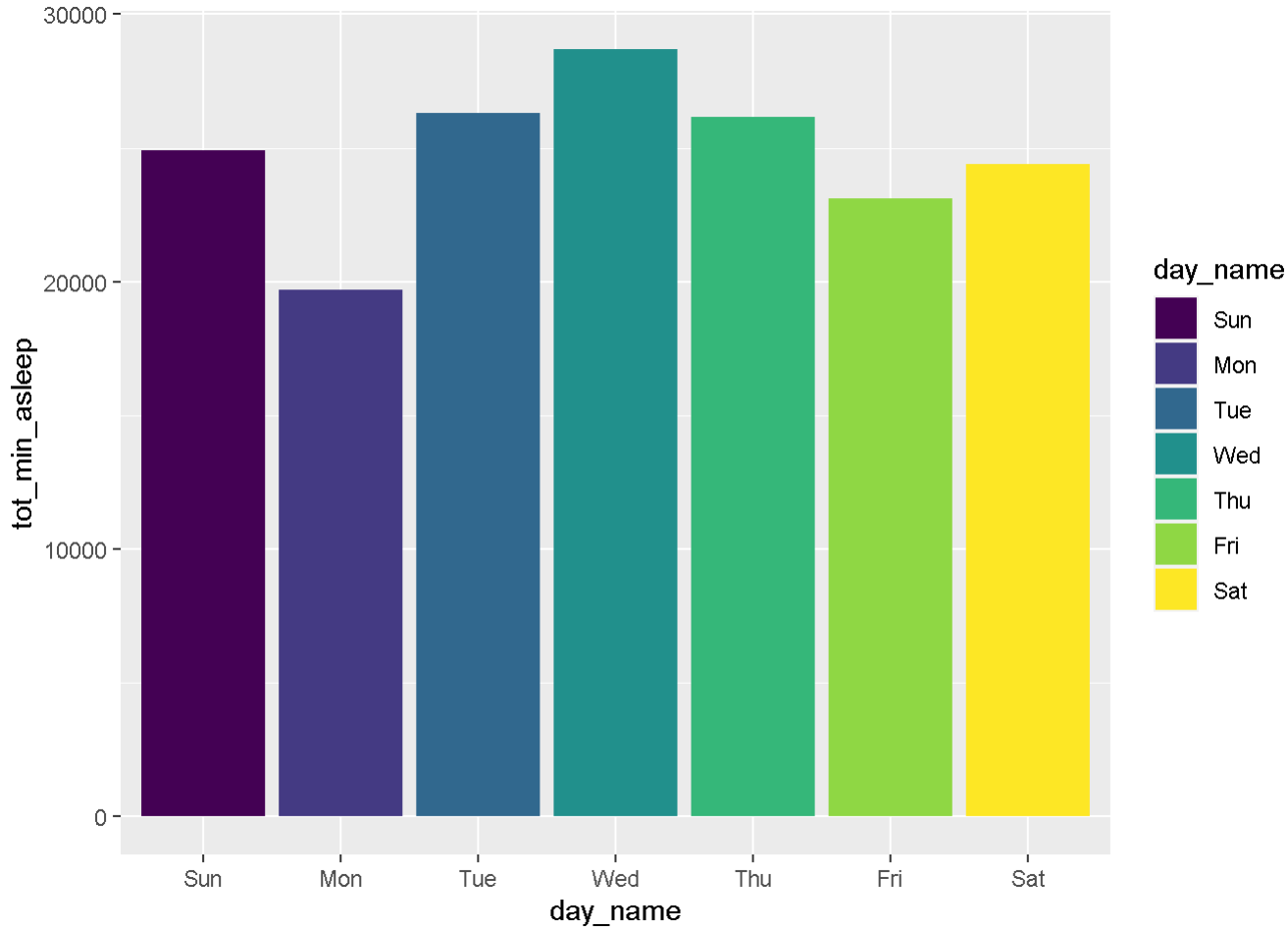
```
ggplot(data = combined_activity_sleep, aes(x = day_name, y = total_steps, fill = day_name)) + geom
_bar(stat = "identity")
```



Observation: Calories has a positive correlation with no. of steps

day_of_week Vs. mins_asleep

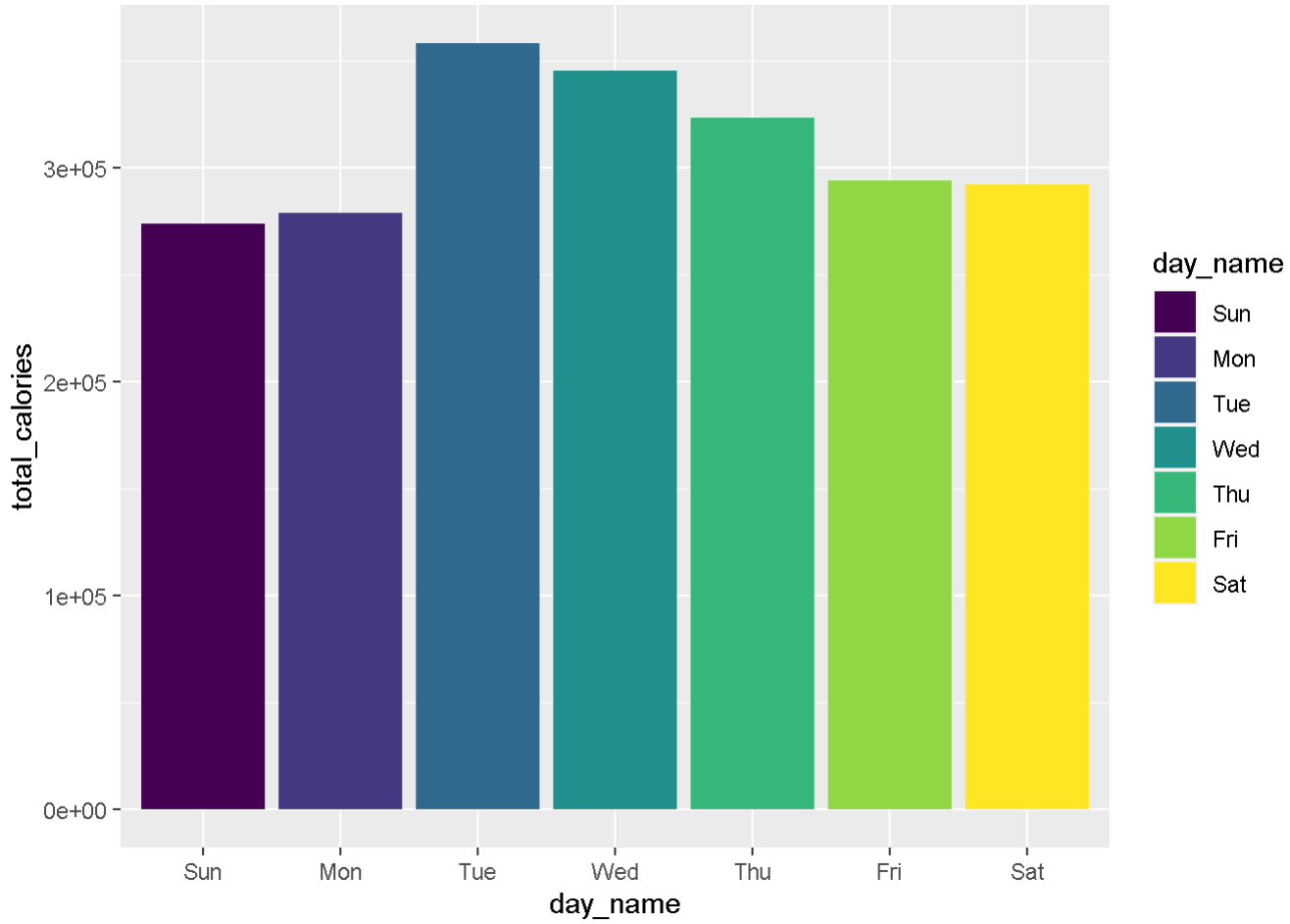
```
ggplot(data = combined_activity_sleep, aes(x = day_name, y = tot_min_asleep , fill = day_name)) +
geom_bar(stat = "identity")
```



Observation: Total mins asleep also has a positive correlation with no. of steps

total_calories Vs. day_of_week

```
ggplot(data = combined_activity_sleep, aes(x = day_name, y = total_calories, fill = day_name)) + g
eom_bar(stat = "identity")
```

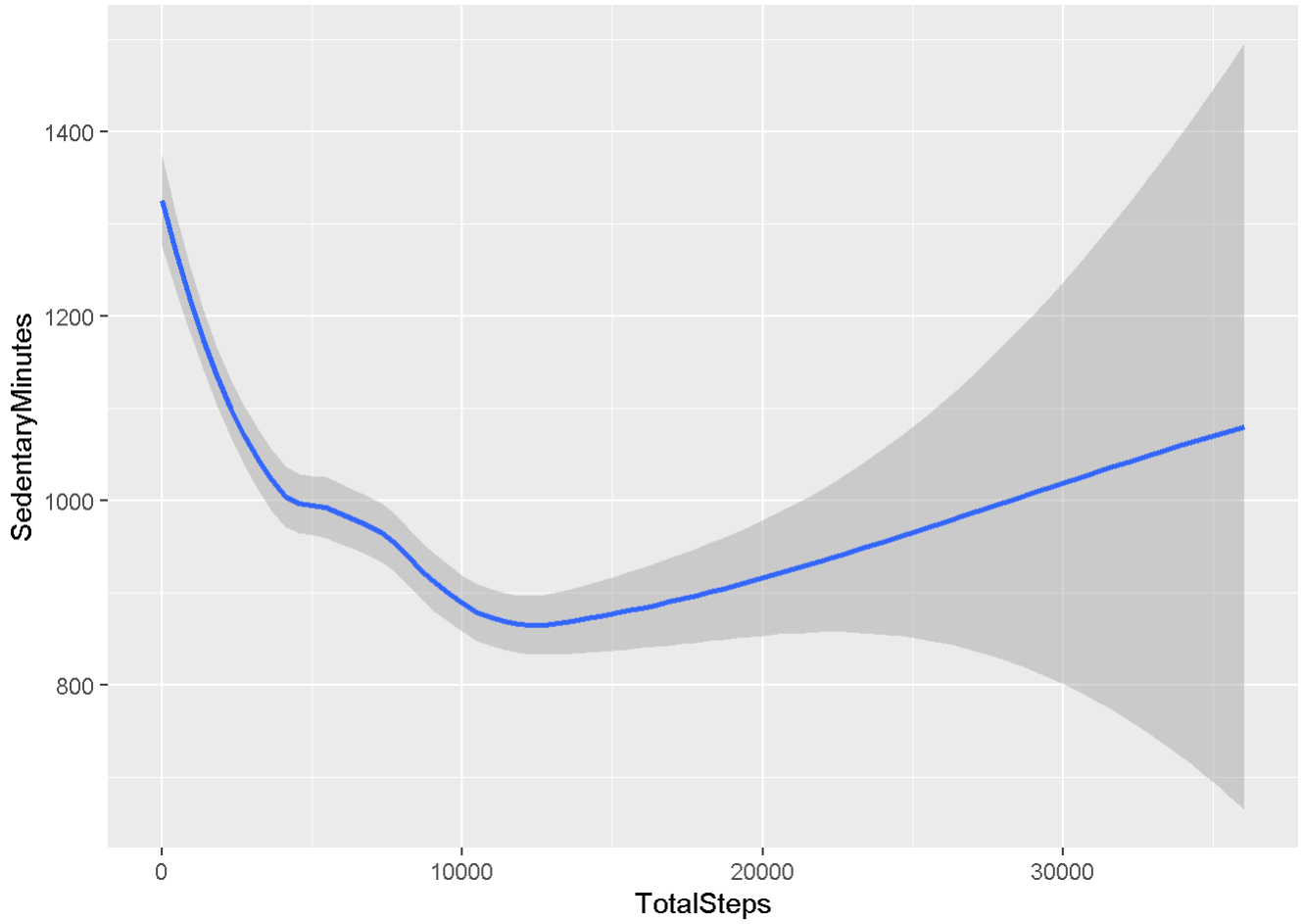


Observation: Calories also has a positive correlation with no. of steps

total_steps Vs. sedentary_mins

```
ggplot(data = daily_activity) + geom_smooth(mapping = aes(x = TotalSteps, y = SedentaryMinutes))
```

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

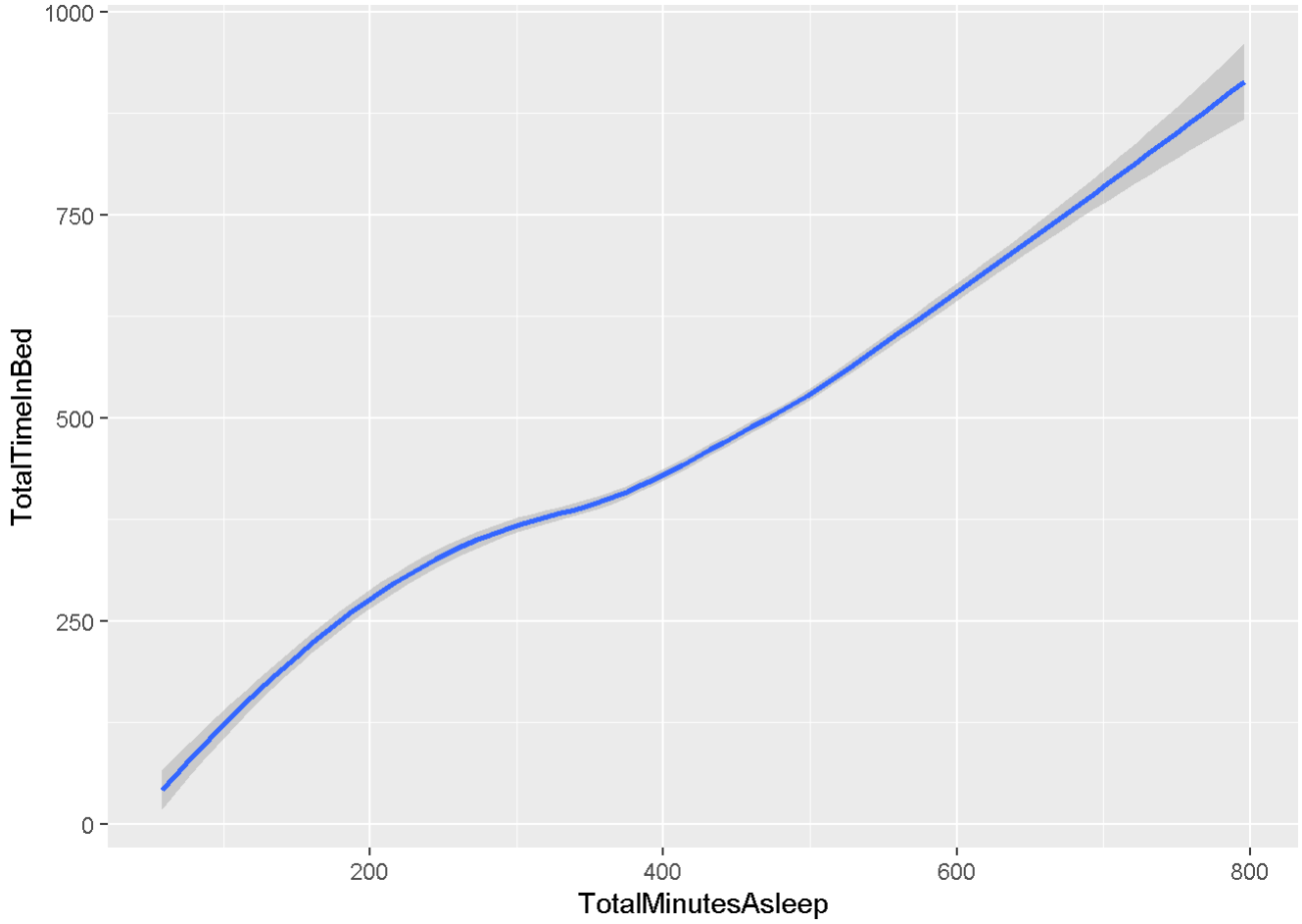


geom_smooth()` using method = 'loess' and formula 'y ~ x' Observation: Sedentary minutes reduces as total steps increases, the least sedentary minutes are b/w 10,000 to 20,000 steps & spikes up after 20,000 steps

total_mins_asleep Vs. total_time_in_bed

```
ggplot(data = sleep_day) + geom_smooth(mapping = aes(x = TotalMinutesAsleep, y = TotalTimeInBed))
```

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




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

Observation: Total time in bed increases with total mins asleep

Conclusions:

- Users that used the smart devices to track steps were found to have their highest record on Tuesday followed by Wednesday.
- Time asleep increases with the increase in steps.
- The tendency to use smart device to track weight information is very low in the users of such devices.
- The brand specialises in Women's products and aims at improving their health by providing the facility to track reproductive health.

Recommendations:

- Bellabeat should form a day-wise weekly social media strategy for customers and should encourage the customers to use the health bands for step tracking for other days too.
- Bellabeat can motivate it's users for tracking more of their sleep patterns on the app by sending reminders each night to improve them.
- Bellabeat should encourage the customers to use their app to track their food habits and behaviours, give them advice on healthy eating and excercising to control their calory intake and burn. This will motivate customers to use their devices even more.

- Bellabeat should interact to customers via posting health awareness content on their social media, and explain how food, sleep, excercise and body weight can influence on hormones and all over their health. This can create a new customer base for the company.