Bellabeat product analysis

Dolapo Oyewumi

2022-04-07

```
install.packages("tidyverse")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)
```

```
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   1.0.8
## 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()
```

comments: for this project, i will be using FitBit Fitness Tracker Data uploading the data

```
daily_activity <- read.csv("dailyActivity_merged.csv")
daily_calories <- read.csv("dailyCalories_merged.csv")
daily_intensities <- read.csv("dailyIntensities_merged.csv")
daily_steps <- read.csv("dailySteps_merged.csv")
daily_sleep <- read.csv("sleepDay_merged.csv")
weight <- read.csv("weightLogInfo_merged.csv")
```

comments: identify all columns in the data

```
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
```

```
colnames(daily_activity)
```

```
##  [1] "Id"                       "ActivityDate"
##  [3] "TotalSteps"               "TotalDistance"
##  [5] "TrackerDistance"          "LoggedActivitiesDistance"
##  [7] "VeryActiveDistance"       "ModeratelyActiveDistance"
##  [9] "LightActiveDistance"      "SedentaryActiveDistance"
## [11] "VeryActiveMinutes"        "FairlyActiveMinutes"
## [13] "LightlyActiveMinutes"     "SedentaryMinutes"
## [15] "Calories"
```

```
colnames(daily_calories)
```

```
## [1] "Id"          "ActivityDay" "Calories"
```

```
colnames(daily_intensities)
```

```
##  [1] "Id"                       "ActivityDay"
##  [3] "SedentaryMinutes"         "LightlyActiveMinutes"
##  [5] "FairlyActiveMinutes"      "VeryActiveMinutes"
##  [7] "SedentaryActiveDistance"  "LightActiveDistance"
##  [9] "ModeratelyActiveDistance" "VeryActiveDistance"
```

```
colnames(daily_steps)
```

```
## [1] "Id"          "ActivityDay" "StepTotal"
```

```
colnames(daily_sleep)
```

```
## [1] "Id"                 "SleepDay"           "TotalSleepRecords"
## [4] "TotalMinutesAsleep" "TotalTimeInBed"
```

```
colnames(weight)
```

```
## [1] "Id"             "Date"           "WeightKg"       "WeightPounds"
## [5] "Fat"            "BMI"            "IsManualReport" "LogId"
```

comments:checking the numbers of users in the data

```
colnames(weight)
```

```
## [1] "Id"             "Date"           "WeightKg"       "WeightPounds"
## [5] "Fat"            "BMI"            "IsManualReport" "LogId"
```

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

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

```
n_distinct(daily_calories$Id)
```

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

```
n_distinct(daily_intensities$Id)
```

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

```
n_distinct(daily_steps$Id)
```

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

```
n_distinct(daily_sleep$Id)
```

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

```
n_distinct(weight$Id)
```

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

comments: Observations: majority of the users don't log in their weight. We only have 8 people logged in so far and the number is too small to make a reasonable conclusion. Users should be encouraged to log in their weight. Also, not all consumers remember to log in their sleeptime

Analyzing

```
daily_activity %>%
  select(TotalSteps,
         TotalDistance,
         SedentaryMinutes) %>%
  summary()
```

```
##    TotalSteps    TotalDistance    SedentaryMinutes
##  Min.   :    0   Min.   : 0.000   Min.   :   0.0
##  1st Qu.: 3790   1st Qu.: 2.620   1st Qu.: 729.8
##  Median : 7406   Median : 5.245   Median :1057.5
##  Mean   : 7638   Mean   : 5.490   Mean   : 991.2
##  3rd Qu.:10727   3rd Qu.: 7.713   3rd Qu.:1229.5
##  Max.   :36019   Max.   :28.030   Max.   :1440.0
```

analyse the daily_activity by categories

```
daily_activity %>%
  select(VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes) %>%
  summary()
```

```
##  VeryActiveMinutes FairlyActiveMinutes LightlyActiveMinutes
##  Min.   :  0.00    Min.   :  0.00      Min.   :  0.0
##  1st Qu.:  0.00    1st Qu.:  0.00      1st Qu.:127.0
##  Median :  4.00    Median :  6.00      Median :199.0
##  Mean   : 21.16    Mean   : 13.56      Mean   :192.8
##  3rd Qu.: 32.00    3rd Qu.: 19.00      3rd Qu.:264.0
##  Max.   :210.00    Max.   :143.00      Max.   :518.0
```

comments:majority of the users are lightly active. the total step per day needs to be increased. (7638 total steps per day) which is bad for the health. According to mayo clinic, Active is more than 10,000 steps per day. Users should be encouraged to be more active daily to improve good health.

```
daily_calories %>%
  select(Calories) %>%
  summary()
```

```
##     Calories
##  Min.   :   0
##  1st Qu.:1828
##  Median :2134
##  Mean   :2304
##  3rd Qu.:2793
##  Max.   :4900
```

```
daily_sleep %>%
  select(TotalSleepRecords, TotalMinutesAsleep, TotalTimeInBed) %>%
  summary()
```

```
##  TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
##  Min.   :1.000     Min.   : 58.0      Min.   : 61.0
##  1st Qu.:1.000     1st Qu.:361.0      1st Qu.:403.0
##  Median :1.000     Median :433.0      Median :463.0
##  Mean   :1.119     Mean   :419.5      Mean   :458.6
##  3rd Qu.:1.000     3rd Qu.:490.0      3rd Qu.:526.0
##  Max.   :3.000     Max.   :796.0      Max.   :961.0
```

comments: The average time sleep time is quite inadequate. The average sleep time for an adults is 7- 9 hours. Users should be encouraged to sleep more

```
weight %>%
  select(WeightKg, BMI) %>%
  summary()
```

```
##     WeightKg           BMI
##  Min.   : 52.60   Min.   :21.45
##  1st Qu.: 61.40   1st Qu.:23.96
##  Median : 62.50   Median :24.39
##  Mean   : 72.04   Mean   :25.19
##  3rd Qu.: 85.05   3rd Qu.:25.56
##  Max.   :133.50   Max.   :47.54
```

comments: users average weight is 72.04kg. Users should be encouraged to watch their weight

merging data

```
step_calories <- merge(daily_calories, daily_steps, by = "Id")
head(step_calories)
```

```
##           Id ActivityDay.x Calories ActivityDay.y StepTotal
## 1 1503960366     4/12/2016     1985     4/12/2016     13162
## 2 1503960366     4/12/2016     1985     4/13/2016     10735
## 3 1503960366     4/12/2016     1985     4/14/2016     10460
## 4 1503960366     4/12/2016     1985     4/15/2016      9762
## 5 1503960366     4/12/2016     1985     4/16/2016     12669
```

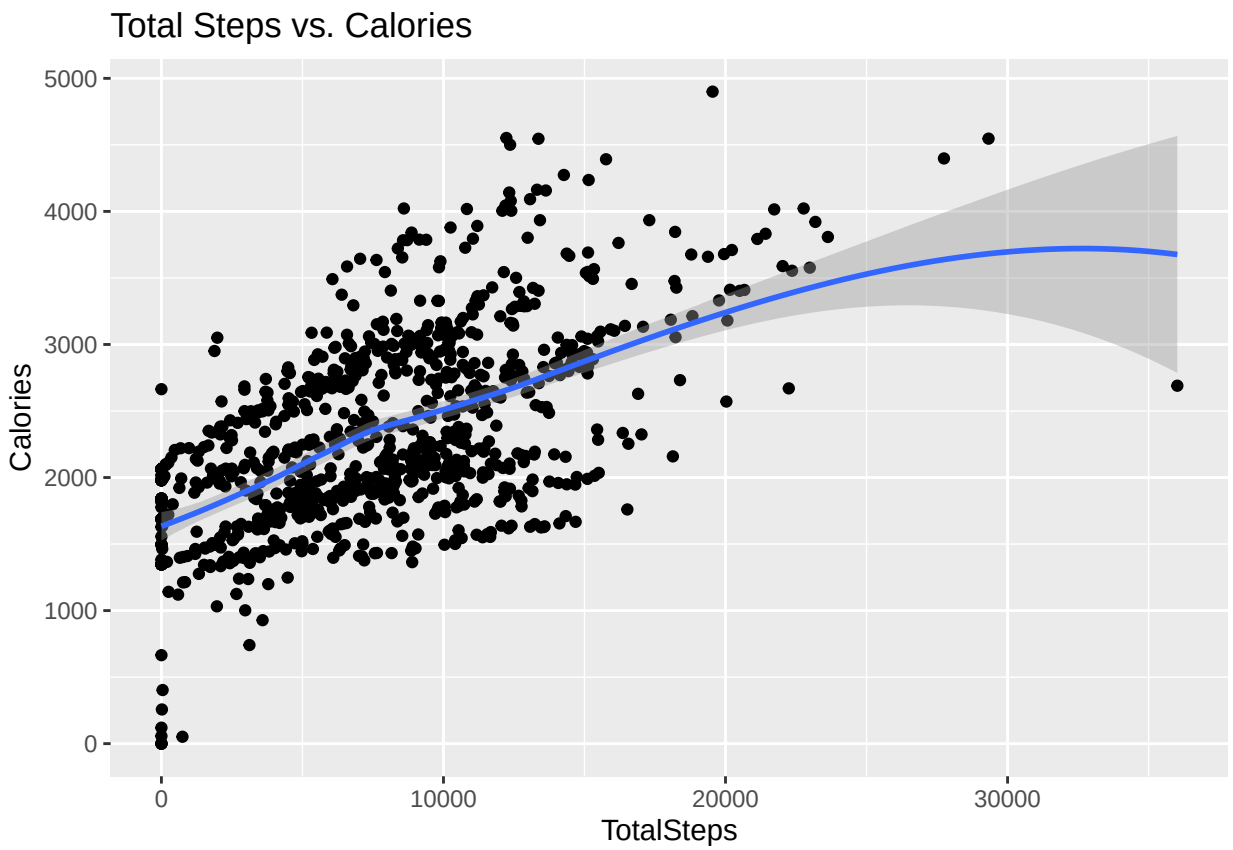
```
## 6 1503960366      4/12/2016      1985      4/17/2016       9705
```

#visualization

comments: want to check the correlation between Calories and Total Steps

```
ggplot(data=daily_activity, aes(x=TotalSteps, y=Calories)) +
geom_point() + geom_smooth() + labs(title="Total Steps vs. Calories")
```

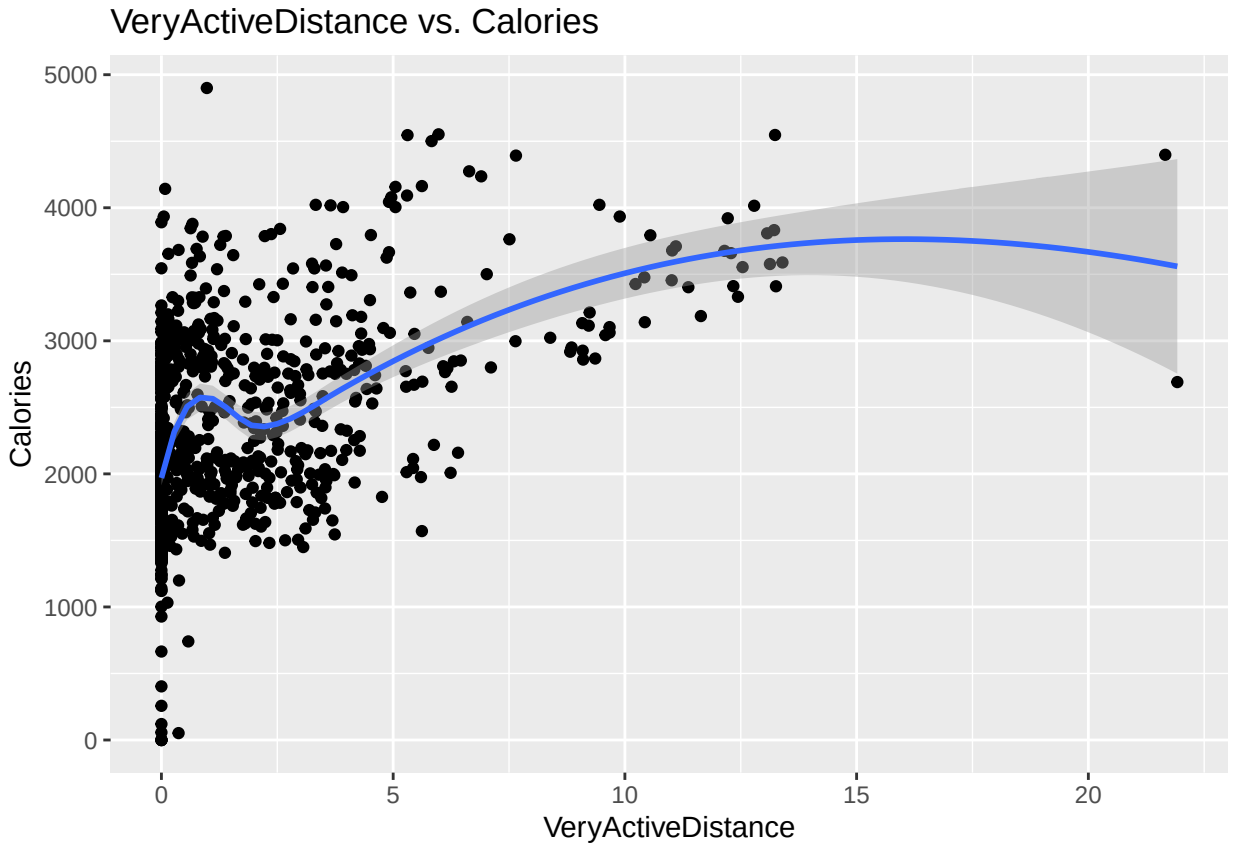
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



comments: there's a positive correlation between total steps and calories. The more active the users, the more calories burnt

```
ggplot(data=daily_activity, aes(x=VeryActiveDistance, y=Calories)) +
geom_point() + geom_smooth() + labs(title="VeryActiveDistance vs. Calories")
```

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



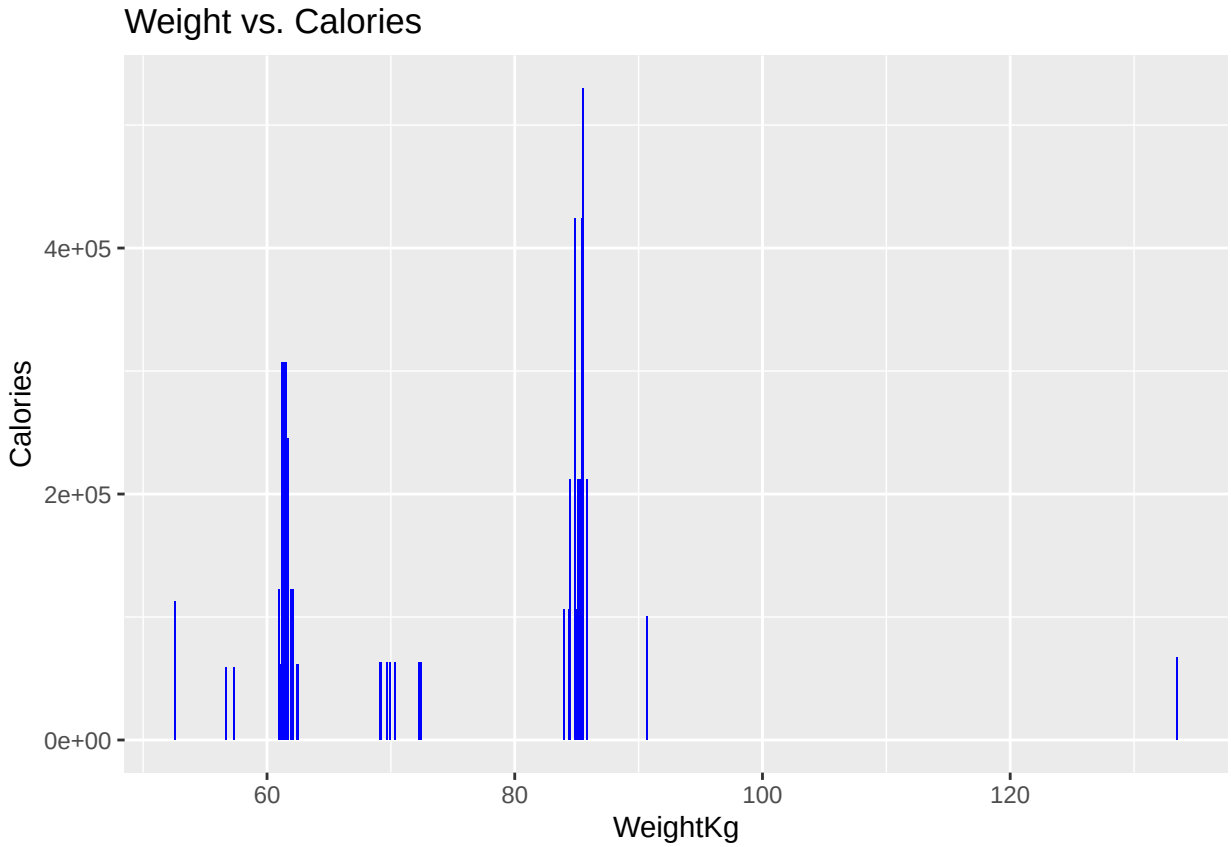
comments: this shows that very active distance burns more calories than Total steps. users who are looking forward to burn calories should be encouraged to be more active

```
weight_calories <- merge(daily_calories, weight, by = "Id")
head(weight_calories)
```

```
##           Id ActivityDay Calories                 Date WeightKg WeightPounds
## 1 1503960366   4/16/2016     1863 5/2/2016 11:59:59 PM     52.6     115.9631
## 2 1503960366   4/16/2016     1863 5/3/2016 11:59:59 PM     52.6     115.9631
## 3 1503960366   4/18/2016     1921 5/2/2016 11:59:59 PM     52.6     115.9631
## 4 1503960366   4/18/2016     1921 5/3/2016 11:59:59 PM     52.6     115.9631
## 5 1503960366   4/15/2016     1745 5/2/2016 11:59:59 PM     52.6     115.9631
## 6 1503960366   4/15/2016     1745 5/3/2016 11:59:59 PM     52.6     115.9631
##   Fat   BMI IsManualReport        LogId
## 1  22 22.65           True 1.462234e+12
## 2  NA 22.65           True 1.462320e+12
## 3  22 22.65           True 1.462234e+12
## 4  NA 22.65           True 1.462320e+12
## 5  22 22.65           True 1.462234e+12
## 6  NA 22.65           True 1.462320e+12
```

```
ggplot(data=weight_calories, aes(x=WeightKg, y=Calories)) +
geom_histogram(stat = "identity", fill='blue') +
labs(title="Weight vs. Calories")
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



comments: There's a correlation with weight and calories. Users who are concerned about their weight,should watch their calories intake and are encouraged to be more active.

RECOMMEDATION

- Total steps per day needs to be increased. The average Total steps is 7638as against the minimum of 10,000 steps recommended by mayo clinic. Users are encouraged to be more active to improve good health.
- Users should be encouraged to increase the hours for sleeping. The average sleep time is inadequate. The ideal sleep time is 7- 9 hours daily
- There's a correlation between calories consumed and weight. Users who intends to reduce in weight should reduce calories intake. The company can suggest foods low in calories to them.
- Analyses showed that the more the steps or active a user is, the higher the calories burnt. Users who looked forward to burning calories should be more active in their daily lives.