Bellabeat customers trends

Tanjim Islam

2023-10-01

##installing packages

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

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)
```

```r
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ----------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
install.packages("skimr")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)
```

```r
install.packages("here")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)
```

```r
install.packages("janitor")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)
```

```r
install.packages("dplyr")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)
```

```r
install.packages("ggplot2")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)
```

```r
library(skimr)
library(here)
```

```
## here() starts at /cloud/project
library(janitor)

##
## Attaching package: 'janitor'
##
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
library(dplyr)
library(ggplot2)
```

Transfering Data to R

```
daily_activity <- read.csv("dailyActivity_merged.csv")
heartrate_sec <- read.csv("heartrate_seconds_merged.csv")
minute_steps <- read.csv("minuteStepsNarrow_merged.csv")
sleep_day <- read.csv("sleepDay_merged.csv")
weight_log <- read.csv("weightLogInfo_merged.csv")
```

Data cleaning & Formatting

```
glimpse(daily_activity)
```

```
## Rows: 940
## Columns: 15
## $ Id                       <dbl> 1503960366, 1503960366, 1503960366, 150396036~
## $ ActivityDate             <chr> "4/12/2016", "4/13/2016", "4/14/2016", "4/15/~
## $ TotalSteps               <int> 13162, 10735, 10460, 9762, 12669, 9705, 13019~
## $ TotalDistance            <dbl> 8.50, 6.97, 6.74, 6.28, 8.16, 6.48, 8.59, 9.8~
## $ TrackerDistance          <dbl> 8.50, 6.97, 6.74, 6.28, 8.16, 6.48, 8.59, 9.8~
## $ LoggedActivitiesDistance <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ VeryActiveDistance       <dbl> 1.88, 1.57, 2.44, 2.14, 2.71, 3.19, 3.25, 3.5~
## $ ModeratelyActiveDistance <dbl> 0.55, 0.69, 0.40, 1.26, 0.41, 0.78, 0.64, 1.3~
## $ LightActiveDistance      <dbl> 6.06, 4.71, 3.91, 2.83, 5.04, 2.51, 4.71, 5.0~
## $ SedentaryActiveDistance  <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ VeryActiveMinutes        <int> 25, 21, 30, 29, 36, 38, 42, 50, 28, 19, 66, 4~
## $ FairlyActiveMinutes      <int> 13, 19, 11, 34, 10, 20, 16, 31, 12, 8, 27, 21~
## $ LightlyActiveMinutes     <int> 328, 217, 181, 209, 221, 164, 233, 264, 205, ~
## $ SedentaryMinutes         <int> 728, 776, 1218, 726, 773, 539, 1149, 775, 818~
## $ Calories                 <int> 1985, 1797, 1776, 1745, 1863, 1728, 1921, 203~
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"
glimpse(sleep_day)
```

```
## Rows: 413
```

```
## Columns: 5
## $ Id                <dbl> 1503960366, 1503960366, 1503960366, 1503960366, 150~
## $ SleepDay          <chr> "4/12/2016 12:00:00 AM", "4/13/2016 12:00:00 AM", "~
## $ TotalSleepRecords <int> 1, 2, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ TotalMinutesAsleep <int> 327, 384, 412, 340, 700, 304, 360, 325, 361, 430, 2~
## $ TotalTimeInBed     <int> 346, 407, 442, 367, 712, 320, 377, 364, 384, 449, 3~
```

```r
daily_activity <- daily_activity %>%
  mutate(ActivityDate = as.Date(ActivityDate, format = "%m/%d/%Y"))
```

##Distinct Participants

```r
daily_activity %>%
  summarise(Participants = n_distinct(daily_activity$Id))
```

```
##   Participants
## 1           33
```

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

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

```r
n_distinct(minute_steps$Id)
```

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

```r
n_distinct(heartrate_sec$Id)
```

```
## [1] 14
```

```r
n_distinct(weight_log$Id)
```

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

```r
n_distinct(weight_log$Id)
```

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

##Analize the data

```r
daily_activity %>%
  select(Id,
         TotalSteps,
         TotalDistance,
         Calories) %>%
  summary()
```

```
##       Id               TotalSteps     TotalDistance       Calories
##  Min.   :1.504e+09   Min.   :    0   Min.   : 0.000   Min.   :   0
##  1st Qu.:2.320e+09   1st Qu.: 3790   1st Qu.: 2.620   1st Qu.:1828
##  Median :4.445e+09   Median : 7406   Median : 5.245   Median :2134
##  Mean   :4.855e+09   Mean   : 7638   Mean   : 5.490   Mean   :2304
##  3rd Qu.:6.962e+09   3rd Qu.:10727   3rd Qu.: 7.713   3rd Qu.:2793
##  Max.   :8.878e+09   Max.   :36019   Max.   :28.030   Max.   :4900
```

```r
heartrate_sec %>%
  select(Id,
         Time,
         Value) %>%
  summary()
```

```
##       Id                Time                Value
##  Min.   :2.022e+09   Length:2483658    Min.   : 36.00
##  1st Qu.:4.388e+09   Class :character   1st Qu.: 63.00
##  Median :5.554e+09   Mode  :character   Median : 73.00
##  Mean   :5.514e+09                      Mean   : 77.33
##  3rd Qu.:6.962e+09                      3rd Qu.: 88.00
##  Max.   :8.878e+09                      Max.   :203.00
```

```r
weight_log %>%
  select(Id,
         WeightKg,
         Fat,
         BMI)%>%
  summary()
```

```
##       Id                WeightKg          Fat              BMI
##  Min.   :1.504e+09   Min.   : 52.60   Min.   :22.00   Min.   :21.45
##  1st Qu.:6.962e+09   1st Qu.: 61.40   1st Qu.:22.75   1st Qu.:23.96
##  Median :6.962e+09   Median : 62.50   Median :23.50   Median :24.39
##  Mean   :7.009e+09   Mean   : 72.04   Mean   :23.50   Mean   :25.19
##  3rd Qu.:8.878e+09   3rd Qu.: 85.05   3rd Qu.:24.25   3rd Qu.:25.56
##  Max.   :8.878e+09   Max.   :133.50   Max.   :25.00   Max.   :47.54
##                                       NA's   :65
```
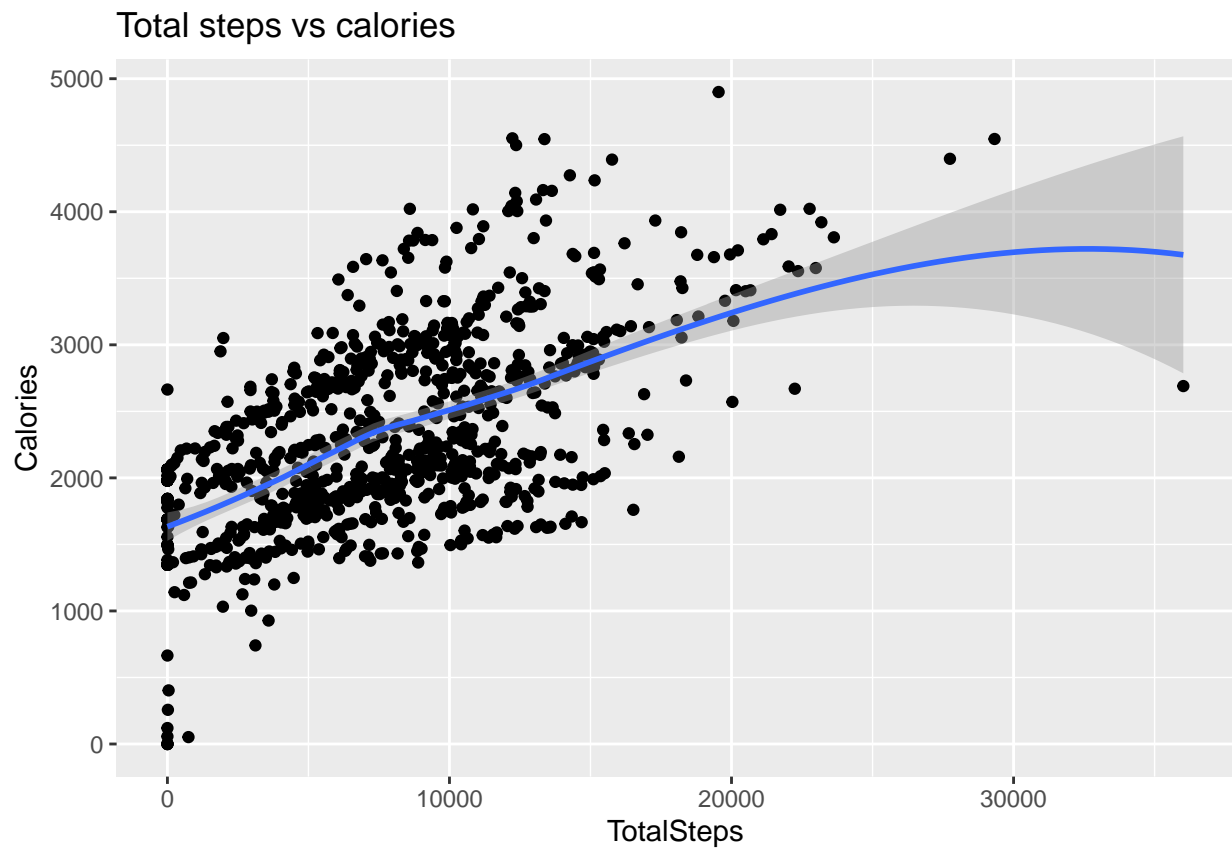
##Combining the data

```r
combined_data <- merge(sleep_day, daily_activity, by="Id")
```

##Data Visualization

```r
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'
```

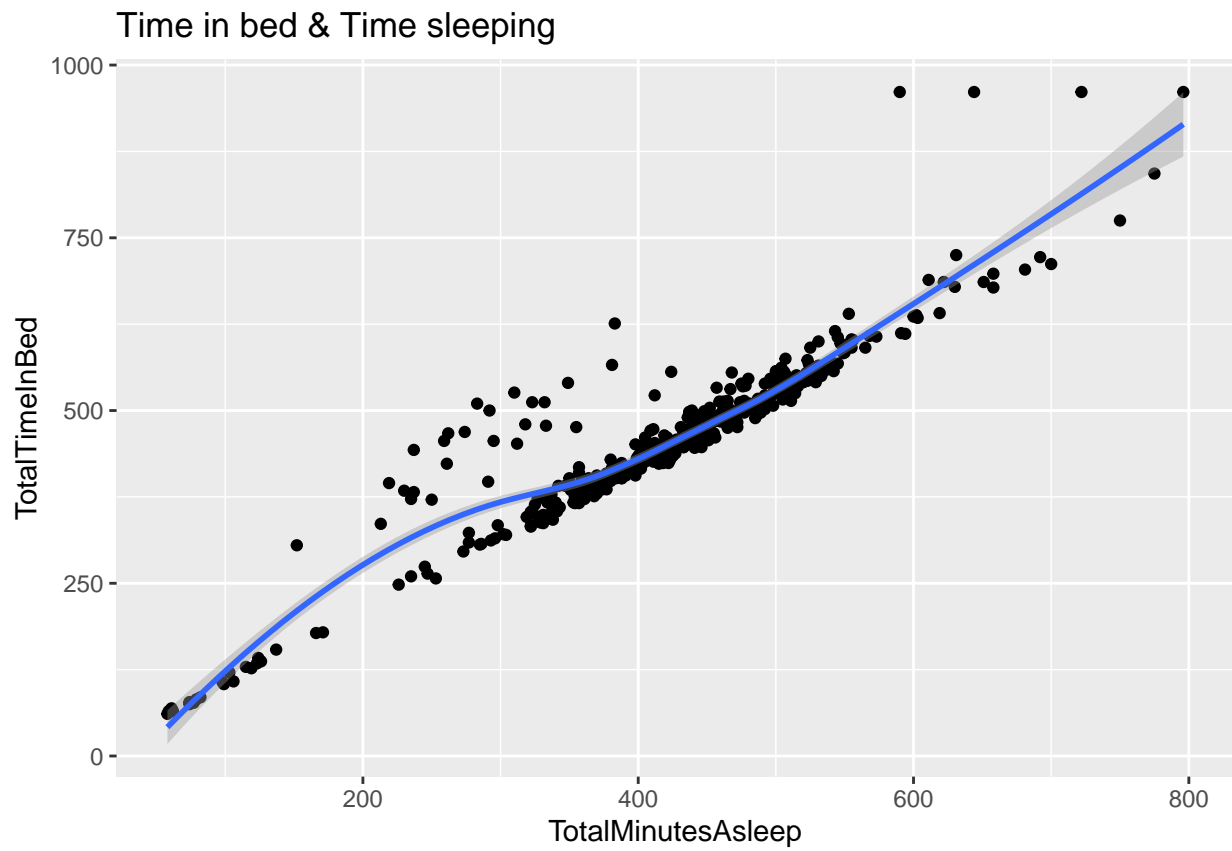## Total steps vs calories


```
ggplot(data = combined_data, aes(x = TotalMinutesAsleep, y = TotalSteps)) +
  geom_col() +
  labs(x = "Total Minutes Asleep", y = "Total Steps", title = "Relationship Between Total Minutes Asleep
```

Relationship Between Total Minutes Asleep and Total Steps
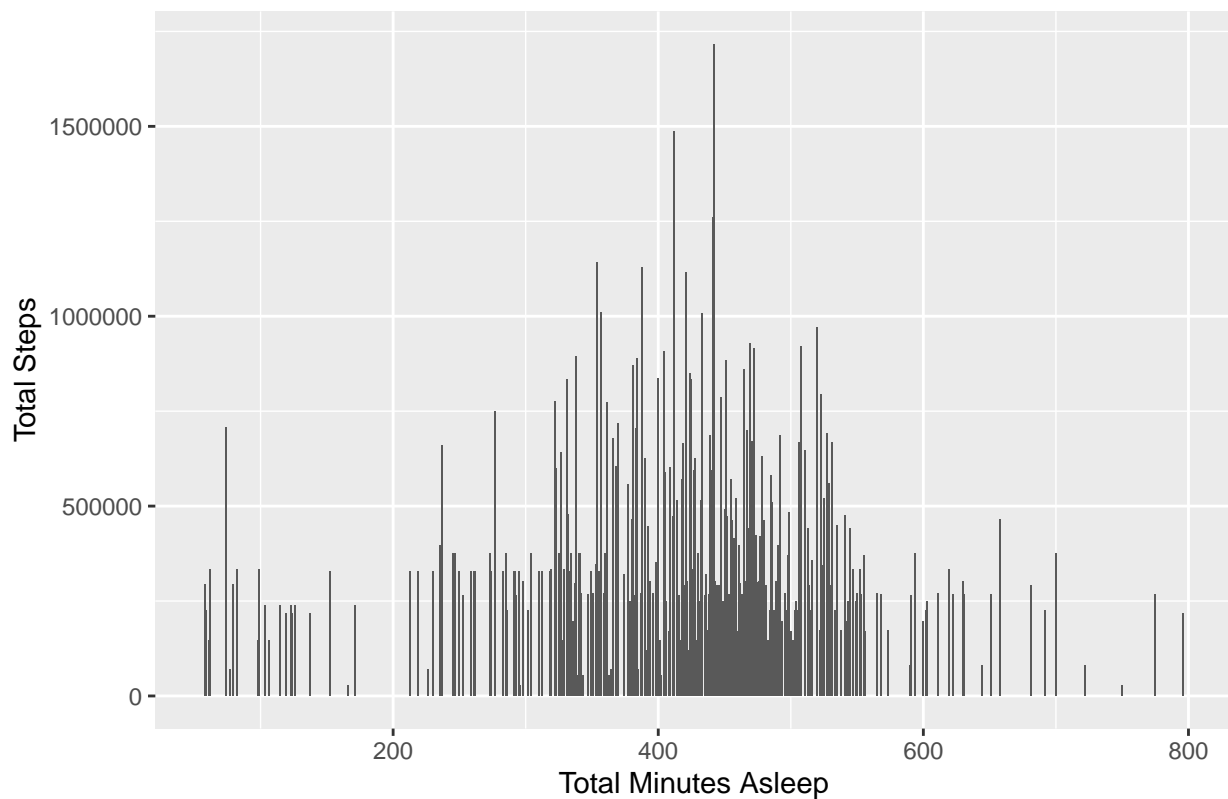

```
ggplot(data = sleep_day, aes(x = TotalMinutesAsleep, y = TotalTimeInBed)) +
  geom_point() + geom_smooth() + labs(title = "Time in bed & Time sleeping")
```

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

Time in bed & Time sleeping

knit