Capstone.R

r2272095

2024-01-16

```
#Introduction

###This is a capstone project for google data analytics course,

###Adeeba Amreen

###Date 16th January 2024

#Objective

###Introducing marketing strategies for bellabeat fitbit devices for the customers
###who use the smart devices.


###I'm going to analyse the data based on correlation activity level, sleep and calories burnt.

#Description of the data used

#I would like to express my gratitude to Möbius for supplying the pertinent dataset needed to study the
#License: CC0: Public Domain
#Source: https://zenodo.org/record/53894#.X9oeh3Uzaao
#Privacy: These datasets were generated by respondents to a distributed survey via Amazon Mechanical Tu


library('tidyverse')
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     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 error
```

```
library('janitor')
```

```
##
## Attaching package: 'janitor'
##
## The following objects are masked from 'package:stats':
```

```
##
##     chisq.test, fisher.test
```

```
library('skimr')
library('here')
```

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

```
library('dplyr')
library(lubridate)
library(ggplot2)

##Uploading the required datasets

library(readr)
dailyActivity_merged <- read_csv("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.
```

```
sleepDay_merged <- read_csv("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.
```

```
##Summary of the dataset

head(dailyActivity_merged)
```

```
## # A tibble: 6 x 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>
```

```r
head(sleepDay_merged)
```

```
## # A tibble: 6 x 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
```

```r
str(dailyActivity_merged)
```

```
## spc_tbl_ [940 x 15] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id                       : num [1:940] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ ActivityDate             : chr [1:940] "4/12/2016" "4/13/2016" "4/14/2016" "4/15/2016" ...
##  $ TotalSteps               : num [1:940] 13162 10735 10460 9762 12669 ...
##  $ TotalDistance            : num [1:940] 8.5 6.97 6.74 6.28 8.16 ...
##  $ TrackerDistance          : num [1:940] 8.5 6.97 6.74 6.28 8.16 ...
##  $ LoggedActivitiesDistance: num [1:940] 0 0 0 0 0 0 0 0 0 0 ...
##  $ VeryActiveDistance       : num [1:940] 1.88 1.57 2.44 2.14 2.71 ...
##  $ ModeratelyActiveDistance: num [1:940] 0.55 0.69 0.4 1.26 0.41 ...
##  $ LightActiveDistance      : num [1:940] 6.06 4.71 3.91 2.83 5.04 ...
##  $ SedentaryActiveDistance : num [1:940] 0 0 0 0 0 0 0 0 0 0 ...
##  $ VeryActiveMinutes        : num [1:940] 25 21 30 29 36 38 42 50 28 19 ...
##  $ FairlyActiveMinutes      : num [1:940] 13 19 11 34 10 20 16 31 12 8 ...
##  $ LightlyActiveMinutes     : num [1:940] 328 217 181 209 221 164 233 264 205 211 ...
##  $ SedentaryMinutes         : num [1:940] 728 776 1218 726 773 ...
##  $ Calories                 : num [1:940] 1985 1797 1776 1745 1863 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   ActivityDate = col_character(),
##   ..   TotalSteps = col_double(),
##   ..   TotalDistance = col_double(),
##   ..   TrackerDistance = col_double(),
##   ..   LoggedActivitiesDistance = col_double(),
##   ..   VeryActiveDistance = col_double(),
##   ..   ModeratelyActiveDistance = col_double(),
##   ..   LightActiveDistance = col_double(),
##   ..   SedentaryActiveDistance = col_double(),
##   ..   VeryActiveMinutes = col_double(),
##   ..   FairlyActiveMinutes = col_double(),
##   ..   LightlyActiveMinutes = col_double(),
##   ..   SedentaryMinutes = col_double(),
##   ..   Calories = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```r
str(sleepDay_merged)
```

```
## spc_tbl_ [413 x 5] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id                : num [1:413] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ SleepDay          : chr [1:413] "4/12/2016 12:00:00 AM" "4/13/2016 12:00:00 AM" "4/15/2016 12:00:0
```

```
##  $ TotalSleepRecords : num [1:413] 1 2 1 2 1 1 1 1 1 1 ...
##  $ TotalMinutesAsleep: num [1:413] 327 384 412 340 700 304 360 325 361 430 ...
##  $ TotalTimeInBed    : num [1:413] 346 407 442 367 712 320 377 364 384 449 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   SleepDay = col_character(),
##   ..   TotalSleepRecords = col_double(),
##   ..   TotalMinutesAsleep = col_double(),
##   ..   TotalTimeInBed = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
#Cleaning dates, converting the activity date and Sleep day


dailyActivity_merged$ActivityDate <- as.Date(dailyActivity_merged$ActivityDate, "%m/%d/%Y")
sleepDay_merged$SleepDay <- parse_date_time(sleepDay_merged$SleepDay, orders = 'mdy HMS')
sleepDay_merged$SleepDay <- as.Date(sleepDay_merged$SleepDay, "%m/%d/%y %h:%m:%s")

#Merging the datasets of activity and sleep timing using left join and replacing NA with zero

daily_activity_sleep <- merge(x= dailyActivity_merged, y= sleepDay_merged,
                              by.x = c("Id", "ActivityDate"), by.y = c("Id", "SleepDay"), all.x = TRUE)
daily_activity_sleep [is.na(daily_activity_sleep)] <- 0


head(daily_activity_sleep)
```

```
##           Id ActivityDate TotalSteps TotalDistance TrackerDistance
## 1 1503960366   2016-04-12      13162          8.50            8.50
## 2 1503960366   2016-04-13      10735          6.97            6.97
## 3 1503960366   2016-04-14      10460          6.74            6.74
## 4 1503960366   2016-04-15       9762          6.28            6.28
## 5 1503960366   2016-04-16      12669          8.16            8.16
## 6 1503960366   2016-04-17       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
##   TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
## 1                 1                327            346
## 2                 2                384            407
## 3                 0                  0              0
## 4                 1                412            442
## 5                 2                340            367
## 6                 1                700            712
```

```
#Summary of merged data set
summary(daily_activity_sleep)
```

```
##        Id              ActivityDate          TotalSteps    TotalDistance
##  Min.   :1.504e+09   Min.   :2016-04-12   Min.   :    0   Min.   : 0.000
##  1st Qu.:2.320e+09   1st Qu.:2016-04-19   1st Qu.: 3795   1st Qu.: 2.620
##  Median :4.445e+09   Median :2016-04-26   Median : 7439   Median : 5.260
##  Mean   :4.858e+09   Mean   :2016-04-26   Mean   : 7652   Mean   : 5.503
##  3rd Qu.:6.962e+09   3rd Qu.:2016-05-04   3rd Qu.:10734   3rd Qu.: 7.720
##  Max.   :8.878e+09   Max.   :2016-05-12   Max.   :36019   Max.   :28.030
##  TrackerDistance  LoggedActivitiesDistance VeryActiveDistance
##  Min.   : 0.000   Min.   :0.000           Min.   : 0.000
##  1st Qu.: 2.620   1st Qu.:0.000           1st Qu.: 0.000
##  Median : 5.260   Median :0.000           Median : 0.220
##  Mean   : 5.489   Mean   :0.110           Mean   : 1.504
##  3rd Qu.: 7.715   3rd Qu.:0.000           3rd Qu.: 2.065
##  Max.   :28.030   Max.   :4.942           Max.   :21.920
##  ModeratelyActiveDistance LightActiveDistance SedentaryActiveDistance
##  Min.   :0.0000           Min.   : 0.000      Min.   :0.000000
##  1st Qu.:0.0000           1st Qu.: 1.950      1st Qu.:0.000000
##  Median :0.2400           Median : 3.380      Median :0.000000
##  Mean   :0.5709           Mean   : 3.349      Mean   :0.001601
##  3rd Qu.:0.8050           3rd Qu.: 4.790      3rd Qu.:0.000000
##  Max.   :6.4800           Max.   :10.710      Max.   :0.110000
##  VeryActiveMinutes FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes
##  Min.   :  0.00    Min.   :  0.00      Min.   :  0          Min.   :   0.0
##  1st Qu.:  0.00    1st Qu.:  0.00      1st Qu.:127          1st Qu.: 729.0
##  Median :  4.00    Median :  7.00      Median :199          Median :1057.0
##  Mean   : 21.24    Mean   : 13.63      Mean   :193          Mean   : 990.4
##  3rd Qu.: 32.00    3rd Qu.: 19.00      3rd Qu.:264          3rd Qu.:1229.0
##  Max.   :210.00    Max.   :143.00      Max.   :518          Max.   :1440.0
##     Calories    TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
##  Min.   :   0   Min.   :0.0000    Min.   :  0.0      Min.   :  0.0
##  1st Qu.:1830   1st Qu.:0.0000    1st Qu.:  0.0      1st Qu.:  0.0
##  Median :2140   Median :0.0000    Median :  0.0      Median :  0.0
##  Mean   :2308   Mean   :0.4899    Mean   :183.7      Mean   :200.9
##  3rd Qu.:2796   3rd Qu.:1.0000    3rd Qu.:417.5      3rd Qu.:450.5
##  Max.   :4900   Max.   :3.0000    Max.   :796.0      Max.   :961.0
```

```
#Data Analysis


#Merge Data Frames

#We are utilizing a left join to combine the data while combining two data frames because of a discrepa
```

```r
daily_activity_sleep <- merge(x= dailyActivity_merged, y= sleepDay_merged,
                              by.x = c("Id", "ActivityDate"), by.y = c("Id", "SleepDay"), all.x = TRUE)
daily_activity_sleep [is.na(daily_activity_sleep)] <- 0
head(daily_activity_sleep)
```

```
##           Id ActivityDate TotalSteps TotalDistance TrackerDistance
## 1 1503960366   2016-04-12      13162          8.50            8.50
## 2 1503960366   2016-04-13      10735          6.97            6.97
## 3 1503960366   2016-04-14      10460          6.74            6.74
## 4 1503960366   2016-04-15       9762          6.28            6.28
## 5 1503960366   2016-04-16      12669          8.16            8.16
## 6 1503960366   2016-04-17       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
##   TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
## 1                 1                327            346
## 2                 2                384            407
## 3                 0                  0              0
## 4                 1                412            442
## 5                 2                340            367
## 6                 1                700            712
```

```r
# Create Categories
#Based on sleep, distance and calories burnt

daily_activity_sleep <- daily_activity_sleep %>%
  mutate(sleep_categories = case_when(
    TotalMinutesAsleep >360 & TotalMinutesAsleep <= 480 ~ "6h-8h",
    TotalMinutesAsleep > 480 ~ "> 8h",
    TRUE ~ "< 6h"
  )) %>%
  mutate(calorie_categories = case_when(
    Calories > 1500 & Calories <= 2500 ~ "1.5k-2.5k",
    Calories > 2500 ~ "> 2.5k",
    TRUE ~ "< 1.5k"
```

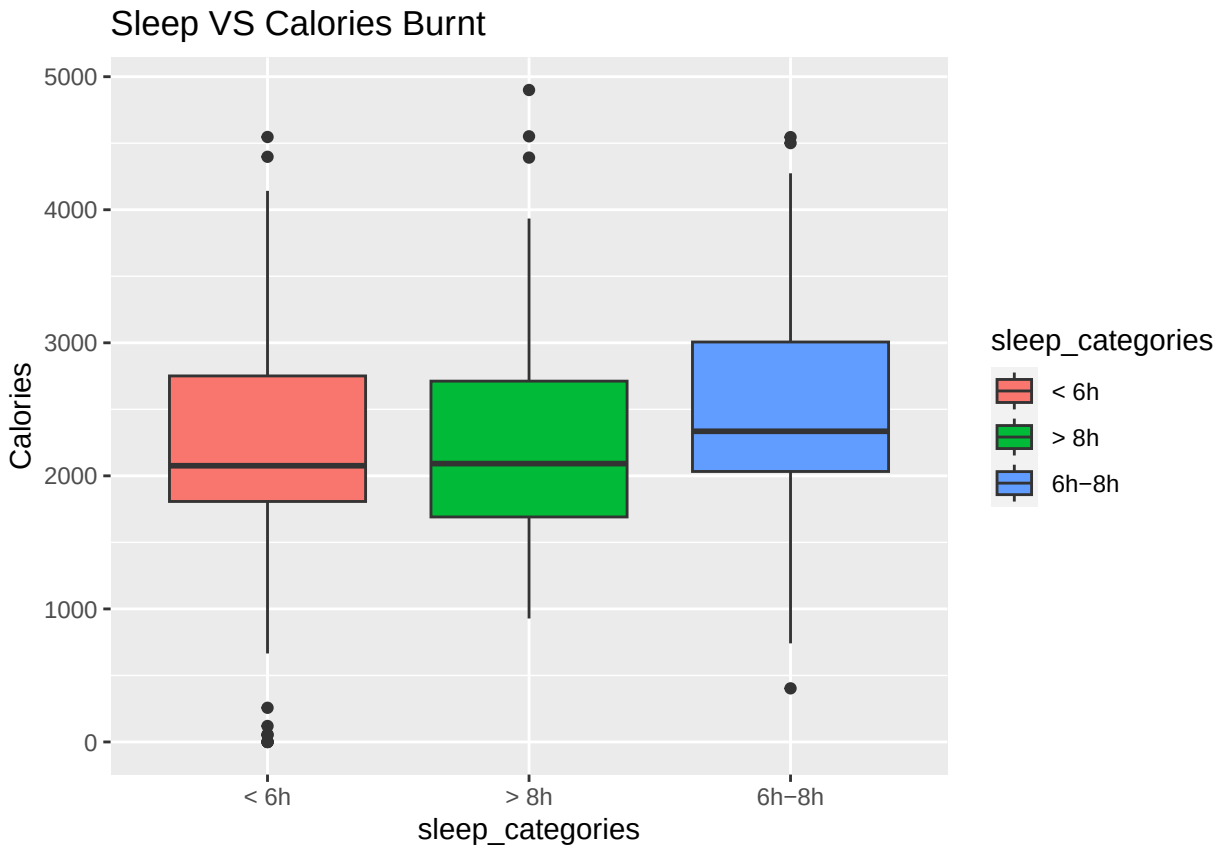
```
  ))


head(daily_activity_sleep)
```

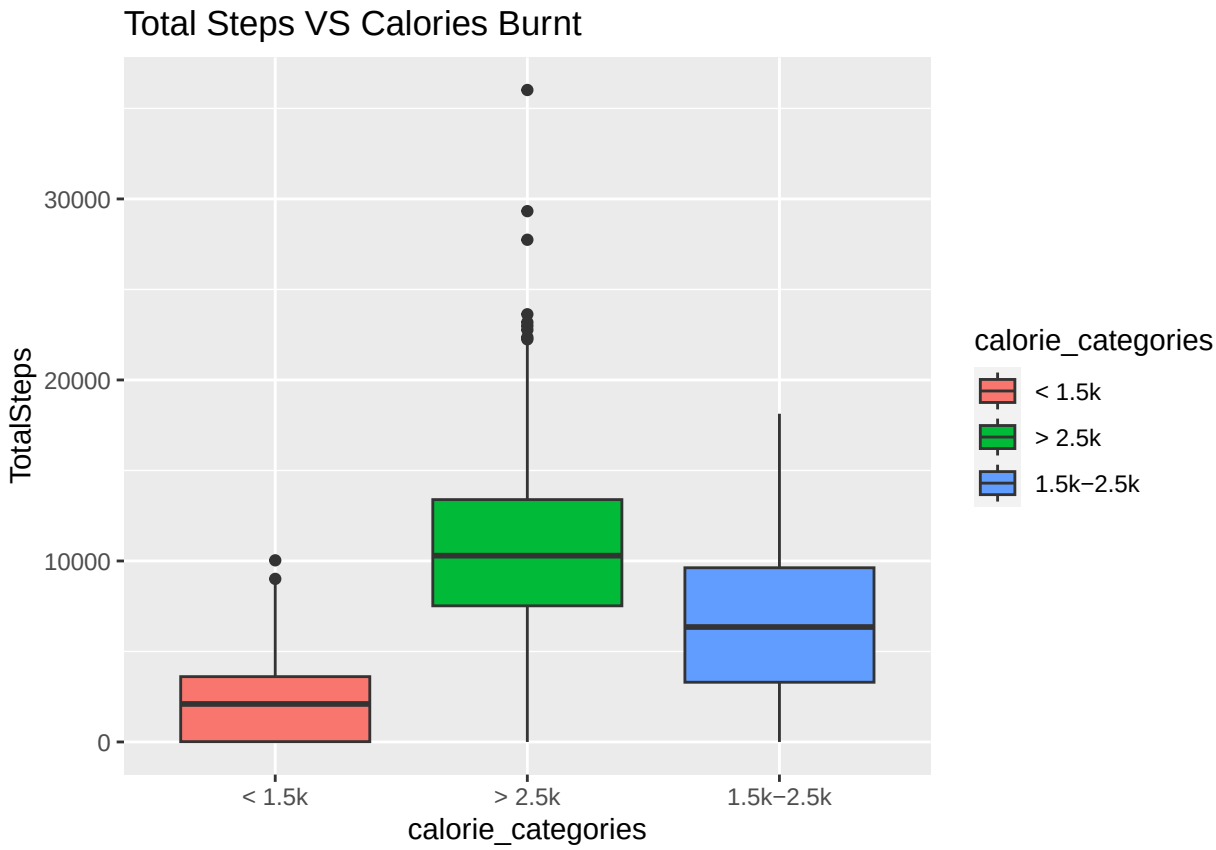
```
##           Id ActivityDate TotalSteps TotalDistance TrackerDistance
## 1 1503960366   2016-04-12      13162          8.50            8.50
## 2 1503960366   2016-04-13      10735          6.97            6.97
## 3 1503960366   2016-04-14      10460          6.74            6.74
## 4 1503960366   2016-04-15       9762          6.28            6.28
## 5 1503960366   2016-04-16      12669          8.16            8.16
## 6 1503960366   2016-04-17       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
##   TotalSleepRecords TotalMinutesAsleep TotalTimeInBed sleep_categories
## 1                 1                327            346             < 6h
## 2                 2                384            407            6h-8h
## 3                 0                  0              0             < 6h
## 4                 1                412            442            6h-8h
## 5                 2                340            367             < 6h
## 6                 1                700            712             > 8h
##   calorie_categories
## 1          1.5k-2.5k
## 2          1.5k-2.5k
## 3          1.5k-2.5k
## 4          1.5k-2.5k
## 5          1.5k-2.5k
## 6          1.5k-2.5k
```

```
#Creating visualization

ggplot(data= daily_activity_sleep) +
  geom_boxplot(mapping= aes(x=sleep_categories, y= Calories, fill= sleep_categories))+ggtitle("Sleep VS
```




```
ggplot(data= daily_activity_sleep) +
  geom_boxplot(mapping= aes(x=calorie_categories, y= TotalSteps, fill= calorie_categories))+ggtitle("To
```




```
#Summary of Data Analysis

### 1. There is a correlation between number of sleep hours and the calories burnt.
### 2. There is a correlation between number of steps and the calories burnt.
### 3. It is evident from the box plot that people who sleep 6-8 hours burn relatively more calories tha
### 4. It is also evident that number of step taken effects the calories burnt, people who have average


#Business Recommendations

### There is a strong correlation between proper number of hours of sleep and calories, so it is importa
### Number of steps are significant in calories reduction, so it is essential for people to track the nu
### The marketing strategies, can be made to inform customers about how efficiently the bellabeat produc
```