

Google Data Analytics Professional Certificate study case report

Julián David Candela

2023-08-31

This report was created as an assignment for the Google Data Analytics Professional Certificate.

Hypothetical case: The Bellabeat company believes that analyzing physical activity data from smart devices could unlock new business opportunities for the company. I've been asked analyze the FitBit Fitness Tracker Data set from Kaggle to learn how users are using their smart devices and present high-level recommendations for Bellabeat's marketing strategy.

Installing and loading packages and libraries

```
install.packages('tidyverse')  
library(tidyverse)
```

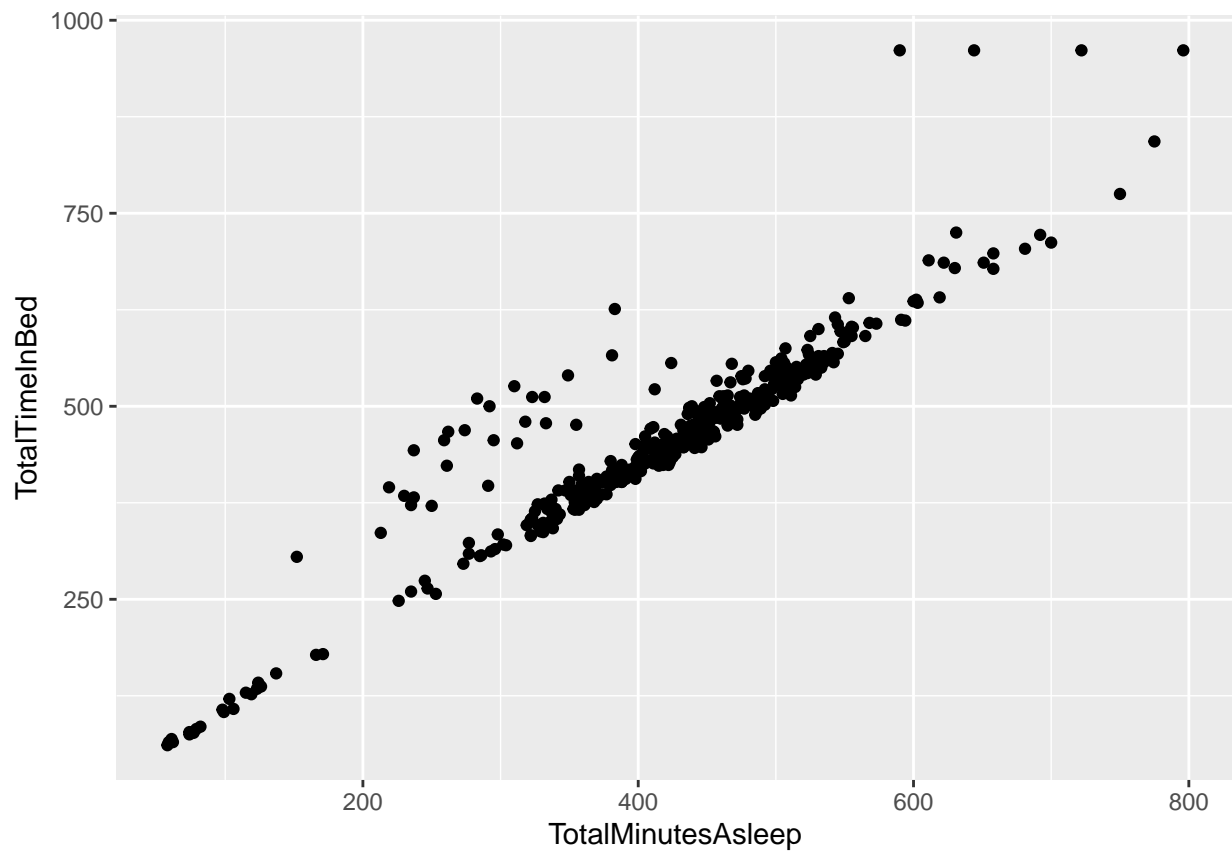
Loading the data

```
daily_activity <- read.csv("dailyActivity_merged.csv")  
sleep_day <- read.csv("sleepDay_merged.csv")  
weight <- read.csv("weightLogInfo_merged.csv")  
calories <- read.csv("dailyCalories_merged.csv")
```

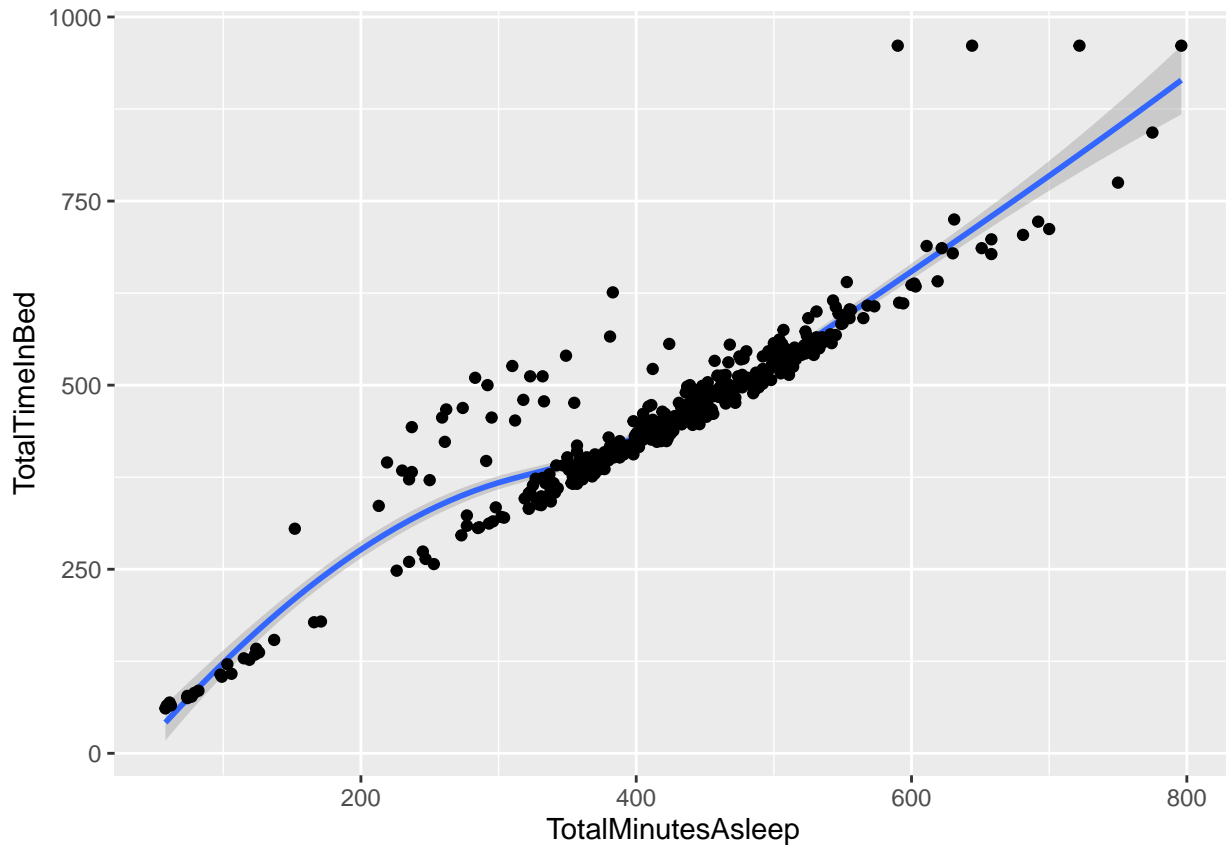
Plotting

First, we plot the variable TotalMinutesAsleep vs the variable TotalTimeInBed in order to study the behavior of the data. We also add a trend line.

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



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



Correlation

Since the plotting of the variables TotalMinutesAsleep vs TotalTimeInBed suggests they are correlated, we verify this by calculating the correlation coefficient.

```
cor(x=sleep_day$TotalMinutesAsleep, y=sleep_day$TotalTimeInBed)
```

```
## [1] 0.9304575
```

Averages of the variables

Now, we calculate the average of the variables TotalTimeInBed and TotalMinutesAsleep as an estimator of the amount of time a user is in bed and sleeps. We also calculate the minimum and maximum of these variables.

```
sleep_day %>% drop_na() %>% summarize(average_time_bed=mean(TotalTimeInBed), min_time_bed=min(TotalT
```

```
##   average_time_bed min_time_bed max_time_bed
## 1           458.6392           61          961
```

```
sleep_day %>% drop_na() %>% summarize(average_time_asleep=mean(TotalMinutesAsleep), min_time_asleep=
```

```
##   average_time_asleep min_time_asleep max_time_asleep
## 1           419.4673           58          796
```

Analysis by user

now, we want to make some analysis by user in order to have a better understanding of the behavior of the data.

```
sleep_day %>% group_by(Id) %>% drop_na() %>% summarize(average_time_bed=mean(TotalTimeInBed),min_time_b
```

```
## # A tibble: 24 x 4
##       Id average_time_bed min_time_bed max_time_bed
##   <dbl>         <dbl>         <int>         <int>
## 1 1503960366         383.           264           712
## 2 1644430081         346           127           961
## 3 1844505072         961           961           961
## 4 1927972279         438.           178           775
## 5 2026352035         538.           380           607
## 6 2320127002          69            69            69
## 7 2347167796         491.           386           602
## 8 3977333714         461.           305           626
## 9 4020332650         380.            77           541
## 10 4319703577         502.            65           722
## # i 14 more rows
```

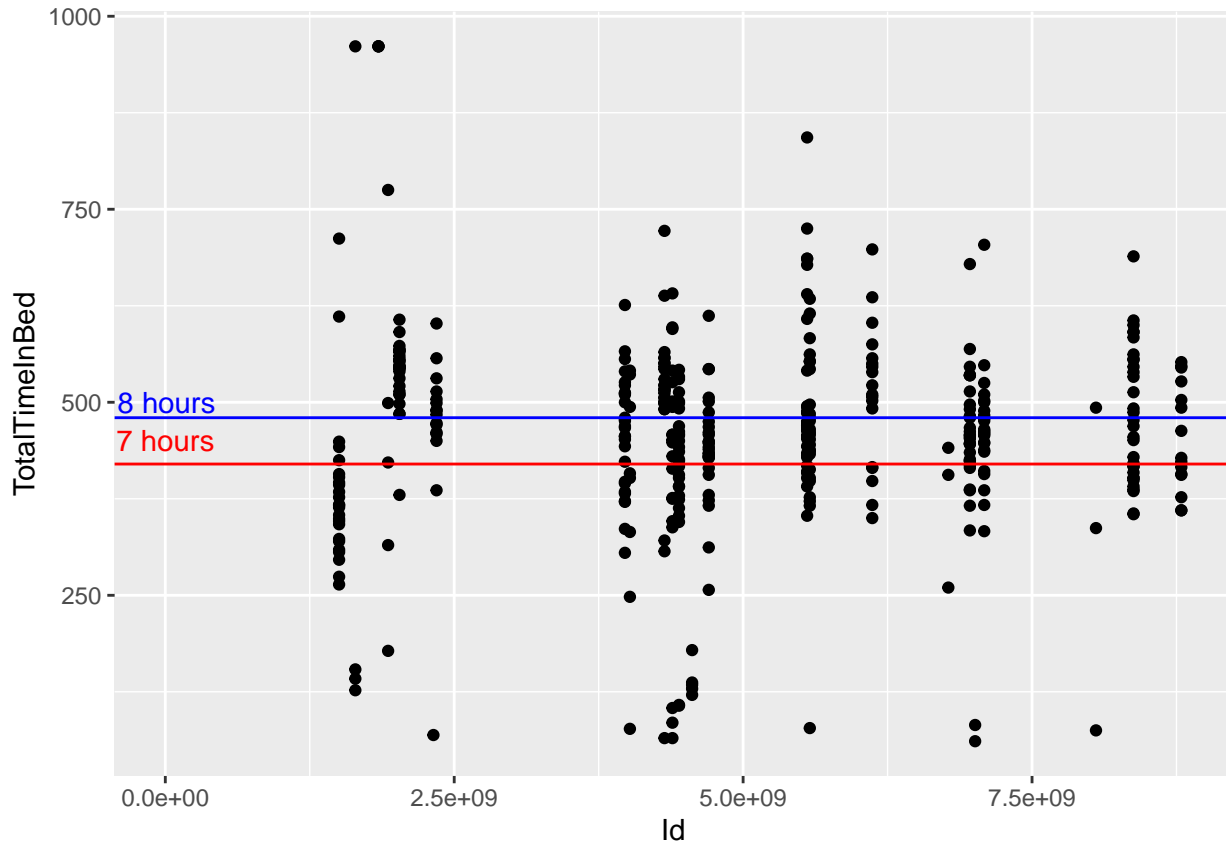
```
sleep_day %>% group_by(Id) %>% drop_na() %>% summarize(mean_time_asleep=mean(TotalMinutesAsleep), min
```

```
## # A tibble: 24 x 4
##       Id mean_time_asleep min_time_bed max_time_bed
##   <dbl>         <dbl>         <int>         <int>
## 1 1503960366         360.           245           700
## 2 1644430081         294           119           796
## 3 1844505072         652           590           722
## 4 1927972279         417           166           750
## 5 2026352035         506.           357           573
## 6 2320127002          61            61            61
## 7 2347167796         447.           374           556
## 8 3977333714         294.           152           424
## 9 4020332650         349.            77           501
## 10 4319703577         477.            59           692
## # i 14 more rows
```

Comparing TotalMinutesAsleep vs the recommended amount of sleep

we want to compare how often users sleep an amount of hours below the recommended amount of sleep

```
ggplot(data=sleep_day, aes(x=Id, y=TotalTimeInBed)) + geom_point() + geom_hline(yintercept = 420, color=
```



Summary of results:

- The correlation coefficient $r=0.93$ guarantee that the variables TotalTimeInBed and TotalMinutesAsleep are highly correlated.
- The average TotalTimeInBed is 458.63 min, that is approximately 7.8 hours.
- The average Totaltimeasleep is 419.46 min, that is approximately 7 hours.

Observations:

- Despite that, in average, the amount of total time for a user to be asleep is reasonable good, making the analysis by user we found out that actually some of the users don't sleep enough.

Recommendations:

- The data suggests that almost all the users sleep approximately the same amount of time that they are in bed; however, the data also shows that in many cases this last amount of time is not enough for a user to rest well . We recommend to add an alert to the device that suggest the user to go to bed early.