

BellaBeat_notebook

Glen

2022-04-30

BellaBeat Project For Google Data Analytics Professional Certification

Hello! My name is Glen Rodrigues, I am from India and this is my capstone project for Google Data Analytics Certification. So in this i will be analysing BellaBeat kaggle dataset [Click_here_for_dataset](#)

About the company

Urška Sršen and Sando Mur founded Bellabeat, a high-tech company that manufactures health-focused smart products. Sršen used her background as an artist to develop beautifully designed technology that informs and inspires women around the world. Collecting data on activity, sleep, stress, and reproductive health has allowed Bellabeat to empower women with knowledge about their own health and habits. Since it was founded in 2013, Bellabeat has grown rapidly and quickly positioned itself as a tech-driven wellness company for women.

Goals and Purpose

This Kaggle data set contains personal fitness tracker from thirty fitbit users. Thirty eligible Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring. It includes information about daily activity, steps, and heart rate that can be used to explore users' habits. Sleep is essential to protect the mental and physical health of an individual, in addition to improving the quality of life. However, the function of sleep has not always been clear and, until recently, the main function was believed to be the avoidance of sleep deprivation. There are many more attributes like heart rate, daily Calories burned so we can analyze this data and explore what customers have potential increased risk of development mentioned diseases and give them appropriate recommendations.

Lets load the packages

```
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()

sleepDay_merged <- read_csv("sleepDay_merged.csv")

## New names:
## * `` -> ...6
## * `` -> ...7
## * `` -> ...8
```

```
## * `` -> ...9

## Rows: 413 Columns: 9
## -- Column specification -----
## Delimiter: ","
## chr (1): SleepDay
## dbl (4): Id, TotalSleepRecords, TotalMinutesAsleep, TotalTimeInBed
## lgl (4): ...6, ...7, ...8, ...9
##
## 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.
DailyActivity <- 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.
heart<-read_csv("heartrate_seconds_merged.csv")

## Rows: 1048575 Columns: 3
## -- Column specification -----
## Delimiter: ","
## chr (1): Time
## dbl (2): Id, Value
##
## 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.
head(sleepDay_merged)

## # A tibble: 6 x 9
##       Id SleepDay TotalSleepRecor~ TotalMinutesAsl~ TotalTimeInBed ...6 ...7
##       <dbl> <chr>         <dbl>         <dbl>         <dbl> <lgl> <lgl>
## 1  1.50e9 4/17/20~           1           700           712 NA   NA
## 2  1.50e9 5/8/201~           1           594           611 NA   NA
## 3  1.50e9 4/24/20~           1           430           449 NA   NA
## 4  1.50e9 4/15/20~           1           412           442 NA   NA
## 5  1.50e9 4/30/20~           1           404           425 NA   NA
## 6  1.50e9 4/13/20~           2           384           407 NA   NA
## # ... with 2 more variables: ...8 <lgl>, ...9 <lgl>
colnames(sleepDay_merged)

## [1] "Id"           "SleepDay"      "TotalSleepRecords"
## [4] "TotalMinutesAsleep" "TotalTimeInBed" "...6"
## [7] "...7"         "...8"          "...9"
glimpse(sleepDay_merged)

## Rows: 413
## Columns: 9
## $ Id           <dbl> 1503960366, 1503960366, 1503960366, 1503960366, 150~
```

```
## $ SleepDay          <chr> "4/17/2016 12:00:00 AM", "5/8/2016 12:00:00 AM", "4~
## $ TotalSleepRecords <dbl> 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, ~
## $ TotalMinutesAsleep <dbl> 700, 594, 430, 412, 404, 384, 383, 369, 366, 361, 3~
## $ TotalTimeInBed     <dbl> 712, 611, 449, 442, 425, 407, 403, 396, 393, 384, 3~
## $ ...6               <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ ...7               <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ ...8               <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ ...9               <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
```

Sleep

It's considered normal to fall asleep in 10 to 20 minutes. If you are falling asleep faster or longer than this, your body may be trying to tell you something. If your issue is falling asleep too quickly, it may mean that you are not sleeping enough. In this case, it would be best to fit sleeping into your schedule, preferably in an eight-hour span each night. On the other hand, if it takes you longer than an hour to fall asleep, it's a sign that your body is sleeping too much or that you are struggling with insomnia. The latter can occur if a situation in your life is preventing you from relaxing.

In data we have 2 columns (TotalMinutesAsleep and Time_Took_to_sleep) we will use the visualization to demonstrate the sleep cycle of customer 1 in a Month. Research has shown that adult woman should sleep 7hr a day (normal_sleep). National Sleep Foundation guidelines¹ advise that healthy adults woman need between 7 and 9 hours of sleep per night.

sleep_cycle of customer_1

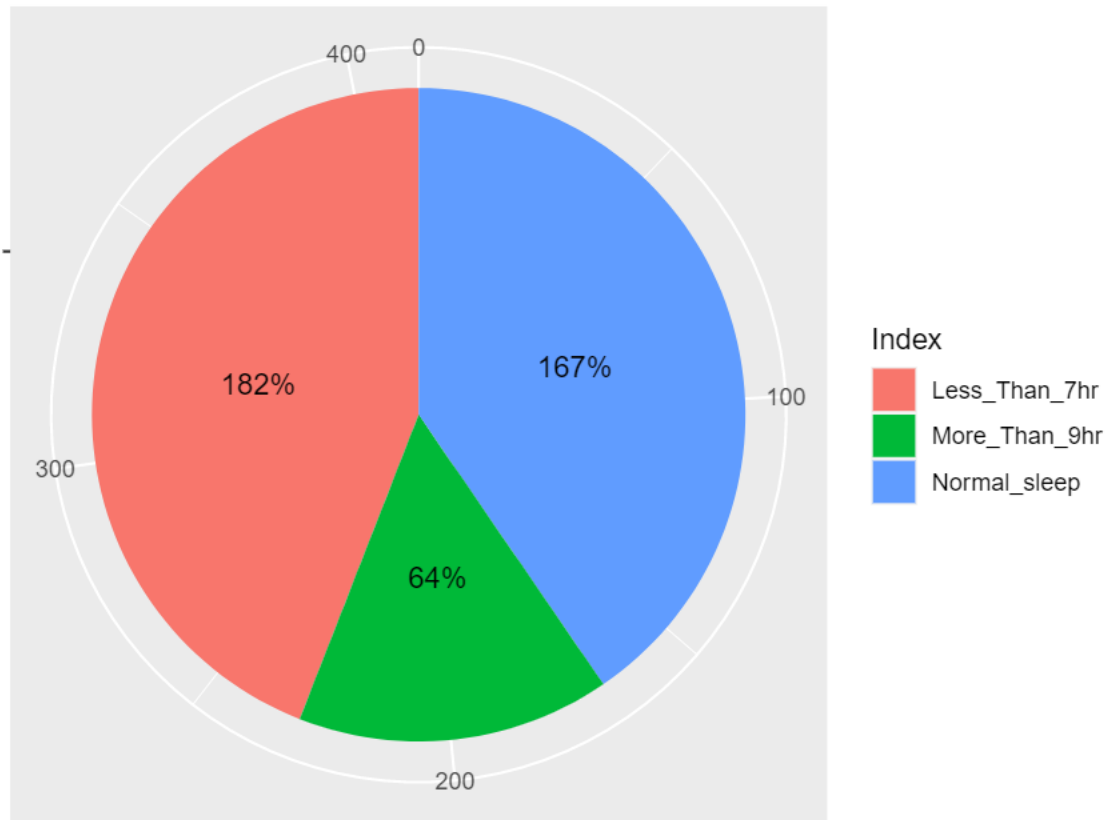
```
yy <- sleepDay_merged %>%
  filter(Id==1503960366)%>%
  drop_na()
```

420min=7hr 520min=9hr

```
gg <- sleepDay_merged %>%
  summarize(Less_Than_7hr = sum(TotalMinutesAsleep < 420), More_Than_8hr = sum(TotalMinutesAsleep > 520))
```

```
df <- data.frame("group" = c('Less_Than_7hr', 'Normal_sleep', 'More_Than_9hr'),
  "value" = c(182, 167, 64))
```

```
ggplot(df, aes(x="", y=value, fill=group)) +
  geom_bar(stat="identity", width=1) +
  coord_polar("y", start=0) +
  geom_text(aes(label = paste0(value, "%")), position = position_stack(vjust=0.5)) +
  labs(x = NULL, y = NULL, fill = "Index")
```



As we can see this customer_1 has more number of Less_Than_7hr sleep and moderate number of normal_sleep so this can be used for recommending customer to sleep more to improve health.

Relation between steps and Calories

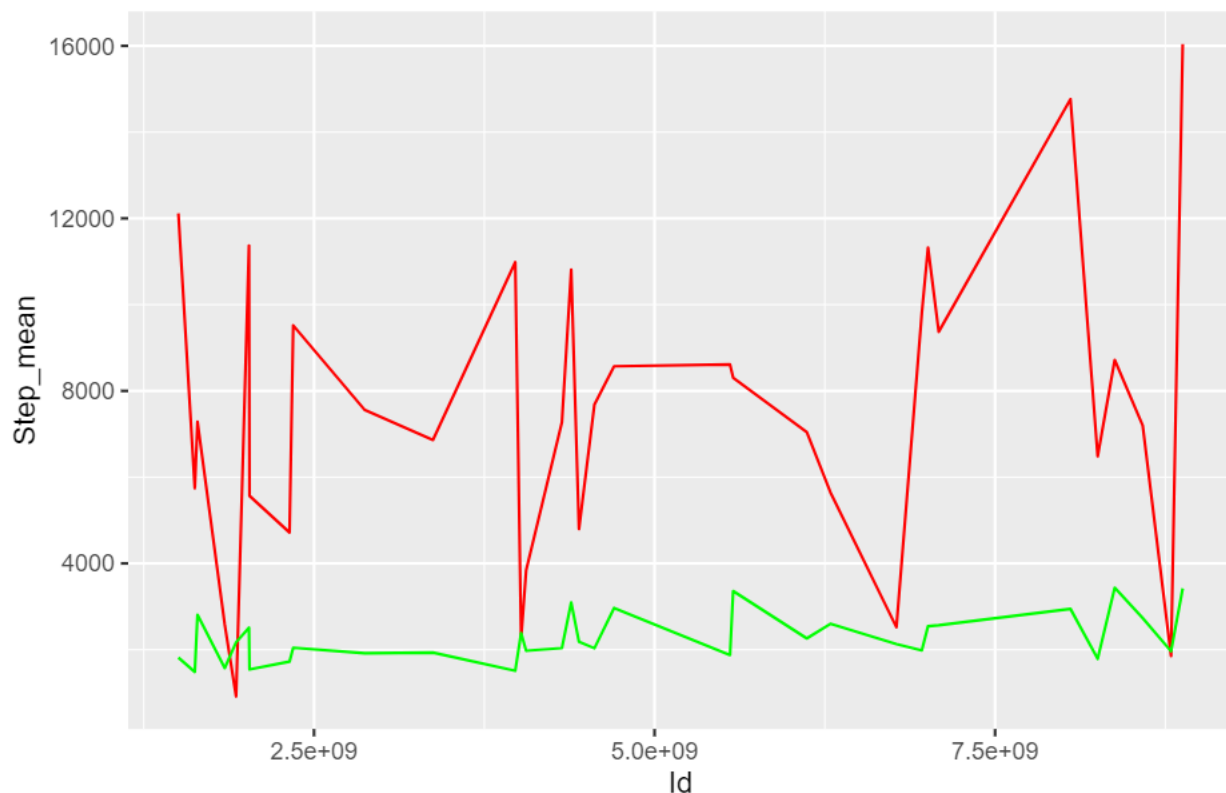
Research(opens in new tab) from the London School of Economics found that people are more likely to have a lower weight if they regularly engage in high-impact walking than other vigorous activities like going to the gym.

```
Daily_Activity <- DailyActivity %>%
  select(Id,TotalSteps,Calories)
```

```
Act <- Daily_Activity %>%
  group_by(Id) %>%
  summarize(Step_mean=mean(TotalSteps),Calories_Mean=mean(Calories))
```

```
ggplot(Act,aes(Id))+
  geom_line(aes(y=Step_mean),color="red")+
  geom_line(aes(y=Calories_Mean),color="Green")+
  ggtitle("Relation between steps and Calories")
```

Relation between steps and Calories



As we can see that more the steps walked more the calories will burn .Red color line indicates the steps and green line indicated the calories burn .

Heart_rate

So we have heart_rate data of 7 customers collected through BellaBand device.The data is already cleaned and small,we will just have a look on the heart rate of each customers.So data set includes field like(ID,date,Values).We will first find the mean heart_rate of every customers then we will have a look on the Heart Rate (beats per min).

```
Heart_rate <- heart%>%
  select(Id,Value)
Heart_rate
```

```
## # A tibble: 1,048,575 x 2
##       Id Value
##   <dbl> <dbl>
## 1 2022484408    97
## 2 2022484408   102
## 3 2022484408   105
## 4 2022484408   103
## 5 2022484408   101
## 6 2022484408    95
## 7 2022484408    91
## 8 2022484408    93
## 9 2022484408    94
## 10 2022484408    93
## # ... with 1,048,565 more rows
```

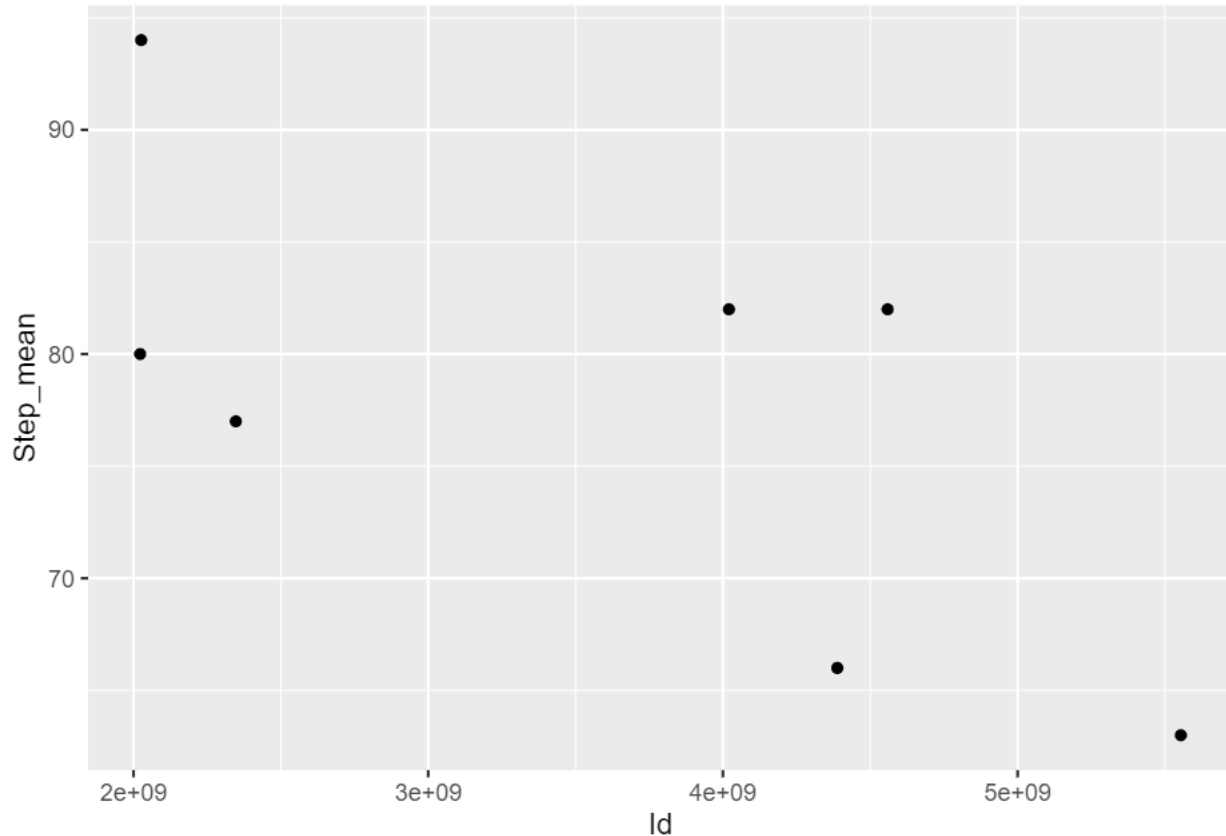
```
heart_Act <- Heart_rate%>%
  group_by(Id) %>%
  summarize(Step_mean=round(mean(Value),digits=0))
heart_Act
```

```
## # A tibble: 7 x 2
##       Id Step_mean
##   <dbl>   <dbl>
## 1 2022484408      80
## 2 2026352035      94
## 3 2347167796      77
## 4 4020332650      82
## 5 4388161847      66
## 6 4558609924      82
## 7 5553957443      63
```

Your heart rate is a measure of how fast your heart beats and is also an important indicator of good health. A good heart rate differs from individual to individual, and it depends upon your age and the kind of physical work you do. 72 beats per minute is a good heart rate. According to the American Heart Association, your risk of dying from a heart attack is lower if your heart rate is below 80 beats per minute. A normal resting heart rate ranges from 60 to 100 beats per minute. However, it is healthier to have a heart rate that is at the lower end of the range.

Now we will check the heart_rate of every customer by using scatter plot graph.

```
show<-ggplot(heart_Act)+
  geom_point(aes(x=Id,y=Step_mean))
show
```



As we can see that some customers heart_rate was abnormal (its was above 80)which is not good, normally heart_rate below 80 is consider as good.

Conclusion

1.Using the give data of sleep duration we found out the sleep cycle of customer1(ID=1503960366)and we found the lack of sleep he being getting so this may cause depression and dementia .We can notify the customer to have a proper sleep to maintain his health.

2.We compare the steps and burn of calories with the help of graph and with the help of Research from the London School of Economics we found that we more the steps great number of calories will burn .

3.Heart rate data provide us the information of customer heart beat per minute.We check whether the heart rate is below 80 or above 80 .we can use analysis insights to pay customers' attention on high heart rate and visit doctor for a regular check up. Analysis was performed only fot two customers, but in future this kind of analysis (or even better updated one) can be made for every customer.