Bellabeat’s Report

# Introduction

Urška Sršen and Sando Mur founded Bellabeat, a high-tech company that manufactures smart products focused on health. Urška leveraged her experience as an artist to develop elegantly designed technology that informs and inspires women worldwide. Collecting data on activity, sleep, stress, and reproductive health has enabled Bellabeat to empower women with knowledge about their own health and habits. Since its founding in 2013, Bellabeat has grown rapidly and quickly established itself as a wellness company focused on technology for women.

In 2016, Bellabeat opened offices around the world and launched several products. Bellabeat products became available through a growing number of online retailers, in addition to its own e-commerce channel on its website. The company invested in traditional advertising media, such as radio, billboards, print media, and television, but focuses largely on digital marketing. Bellabeat invests year-round in Google Search, maintaining active pages on Facebook and Instagram, and consistently engaging consumers on Twitter. Additionally, Bellabeat runs video ads on YouTube and display ads on the Google Display Network to support campaigns on important marketing dates.

Urška knows that an analysis of Bellabeat's available consumer data would reveal more growth opportunities. She asked the marketing analytics team to focus on one Bellabeat product and analyze the usage data of smart devices to gain insights into how people are already using their smart devices. Through this information, she would like to obtain excellent recommendations on how these trends can guide Bellabeat's marketing strategy.

Urška requests that you analyze the usage data of smart devices to gain insights into how consumers use smart devices that are not from Bellabeat. Then, she wants you to select a Bellabeat product to apply these insights in your presentation.

# Business Task

The problem we are addressing involves exploring how the use of smart health devices can uncover new insights, trends, and features in the current market landscape. Our goal is to harness these findings to enhance Bellabeat's offerings.

To achieve this, we need to enhance both our app and smart devices to deliver superior service and products to our clients. By doing so, we aim to introduce innovative materials that will enable Bellabeat to capture a larger market share.

The main stakeholders for this analysis are:

* Urška Sršen: Co-founder and CEO of Bellabeat;
* Sando Mur: Co-founder of Bellabeat; key member of Bellabeat's executive team;
* Bellabeat marketing analytics team.

## Step by step:

Objective: Improve customer catch and retention of the Bellabeat’s products.

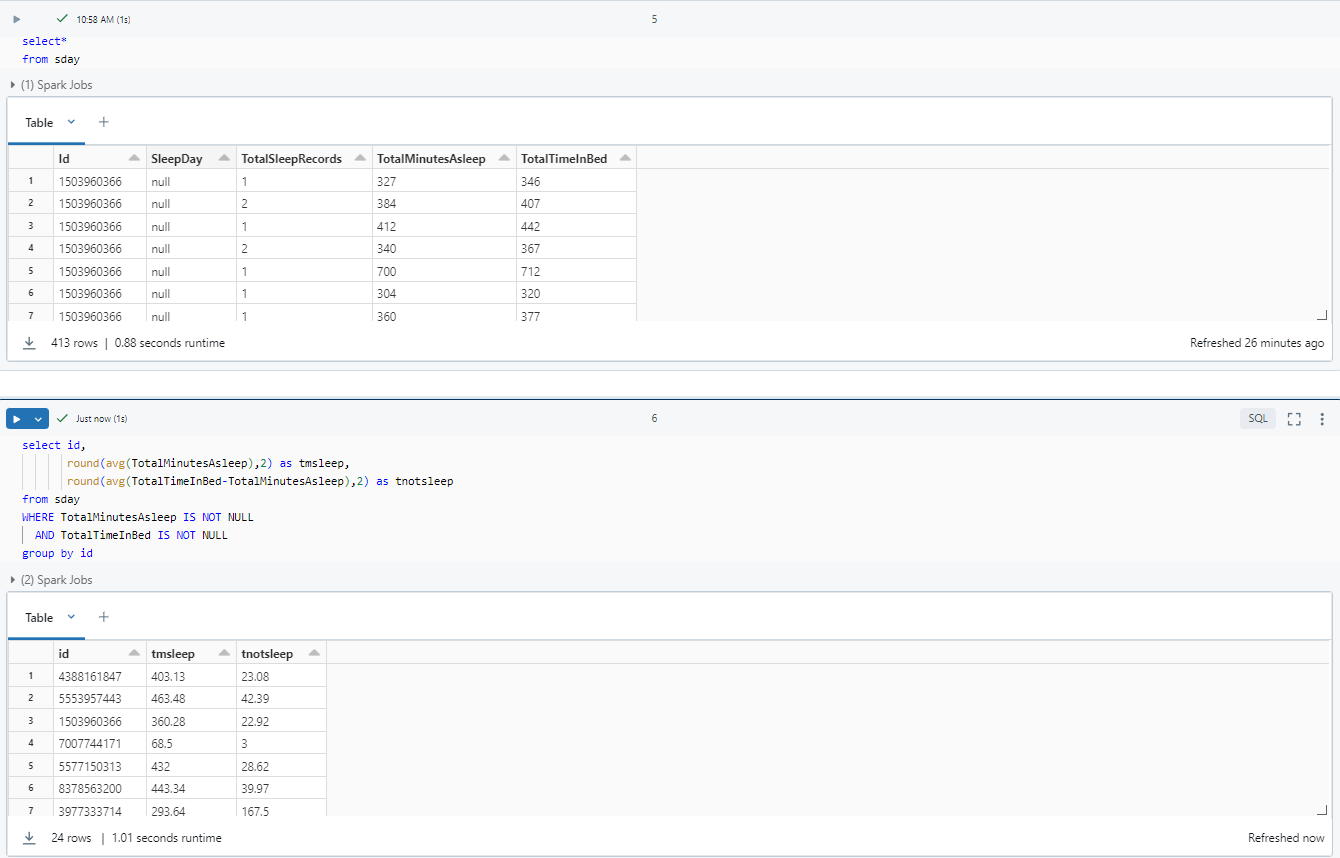
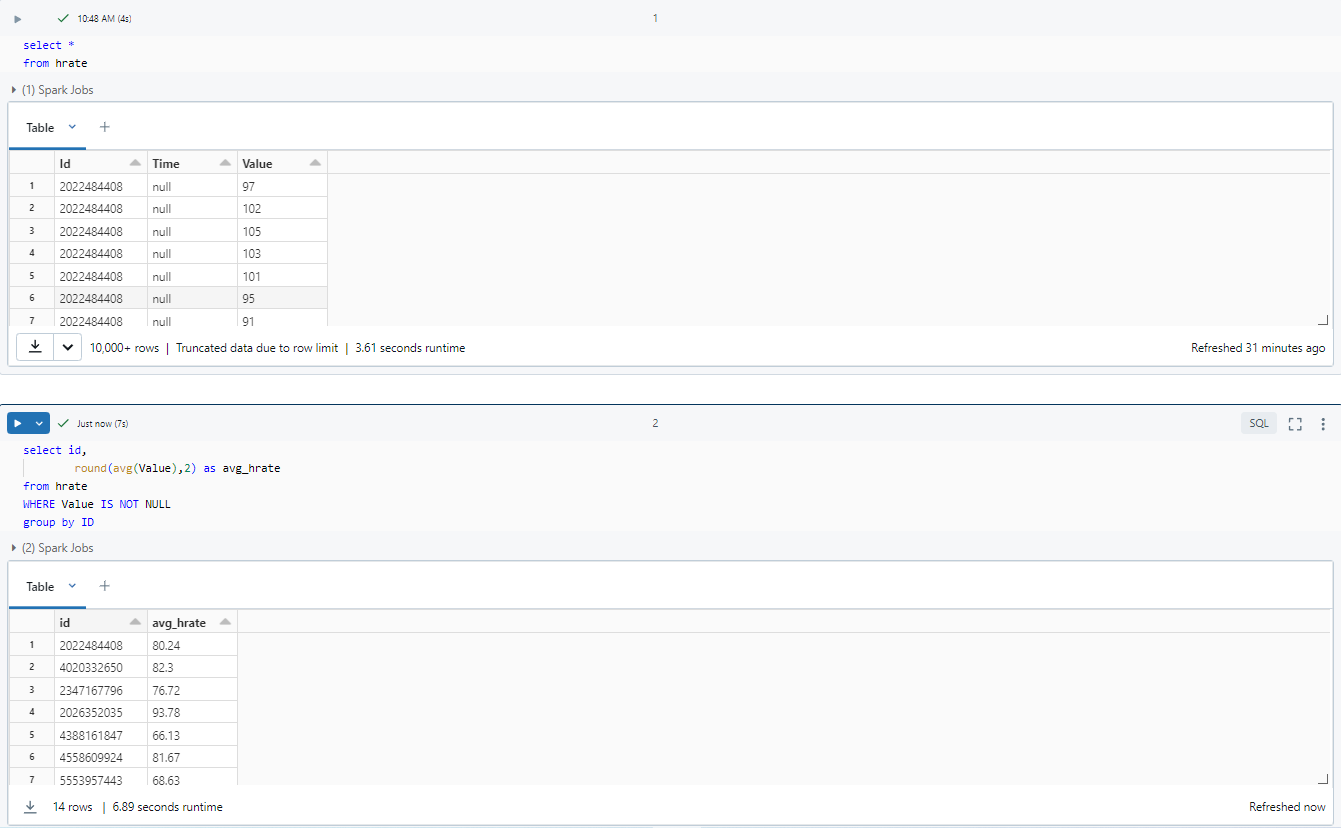
Task Description: perform an analysis over the existing market of intelligent devices and create insights to improve performance and product delivery to our customers and so grow Bellabeat’s market steak in specific for their Leaf product, a bracelet that keeps track of activity, sleep and stress.

## Key Steps:

1. Data Collection:
   * <https://www.kaggle.com/arashnic/fitbit> a public data collection published on Kaggle. This dataset provides a record (in a long format) of personal fitness tracking for thirty-three Fitbit users;
   * Regarding this data, it’s important to state that there’s only 33 users for this dataset, who could probably create low liability over this data. Besides the number, some users are not keeping track of their daily activities;
   * Note about this dataset: Thirty eligible Fitbit users consented to the submission of personal tracker data, including minute-by-minute results of physical activity, heart rate, and sleep monitoring. The data covers information on daily activities, steps, and heart rate that can be used to explore users' habits;
   * The data frame analysed in the case study cover the time period starting from April 12th, 2016, and ending on May 12th, 2016.
2. Data Cleaning and Preparation:
   * Remove duplicates, correct errors, and fill in missing data. Using Excel, we’ve looked over possible mistakes such as duplicates, for example, records with the same hour and date.
   * Structure the data in a way that facilitates efficient analysis, using SQL in Databricks It was created a set of four new tables about means of activity, heart rate, sleep and weight measurement, with the following code:

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

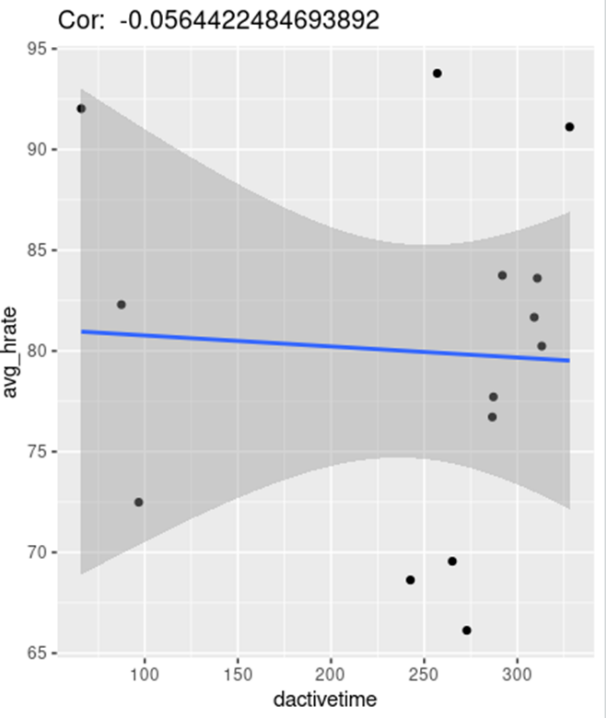
* + After cleaning it was possible to obtain two different sets of tables:

A diagram of a dataset

Description automatically generated with medium confidenceA diagram of data processing

Description automatically generated

* + The first set is a group of tables chosen to make an analysis by record, and the second dataset it was a group of tables adjusted into averages per ID of Dataset Chosen.
  + The overall objective it was to see by one hand isolate events and in the other hand to stablish a relationship between several indicators such as average heart rate, sleep, weight, activity and its ID (the user).
  + To create the data visualizations and metrics for analysis, R was the platform used. Below are the steps taken to generate all metrics. Some joins were necessary to establish relationships between tables, as not all tables contain records for the same IDs. For example, some users did not log their weight or heart rate, necessitating adjustments based on the available data:

> install.packages("tidyverse") # Installing tidyverse

> library(tidyverse) **# Uploading**

>c1hrate\_actime <- inner\_join(dactivetime, hrate, by = "id") **# Creating a table with heart rate and activity**

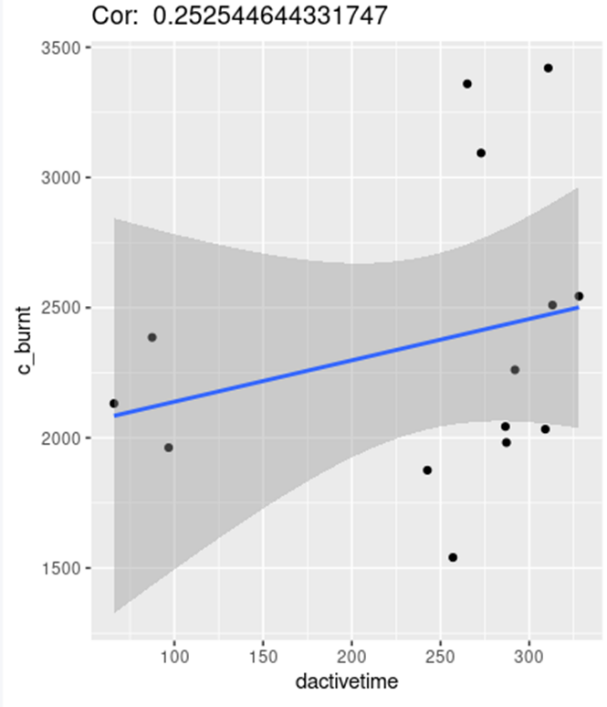
> str(c1hrate\_actime) spc\_tbl\_ [14 × 4] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame) **# Basic stats**

> ggplot(data=c1hrate\_actime, aes(x=dactivetime, y=avg\_hrate))

+ geom\_point()

+ geom\_smooth(method='lm', formula='y~x')

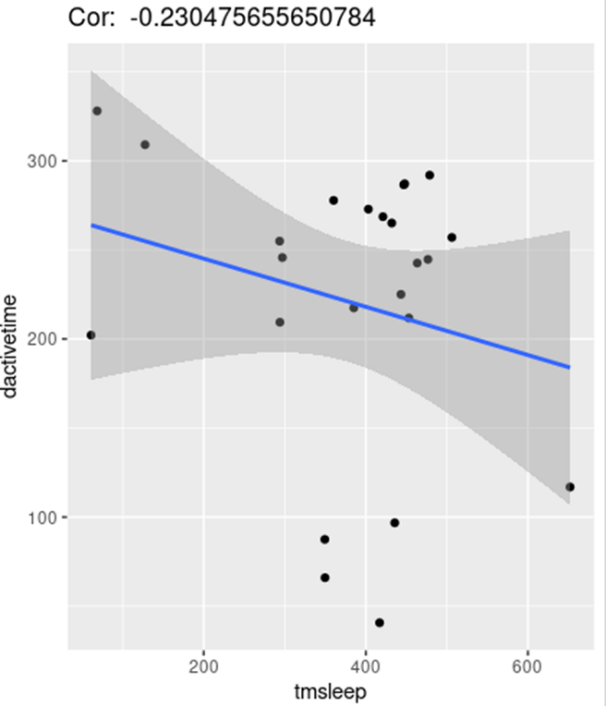
+ ggtitle((paste("Cor: ",cor(c1hrate\_actime$dactivetime,c1hrate\_actime$avg\_hrate)))) **# Correlation between hrate and activity**

>ggplot(data=c1hrate\_actime, aes(x=dactivetime, y=c\_burnt))

+ geom\_point()

+geom\_smooth (method='lm', formula='y~x')

+ggtitle((paste("Cor: ",cor (c1hrate\_actime$ dactivetime, c1hrate\_actime$ c\_burnt)))) **# Correlation between calories burned and atcivity**



c2sleep\_actime <- inner\_join (dactivetime, sleep, by = "id") **# Creating a table with sleep and activity**

> str(c2sleep\_actime)spc\_tbl\_ [24 × 5] (S3: spec\_tbl\_df/ tbl\_df/tbl/data.frame) **# basic stats**

> ggplot(data=c2sleep\_actime, aes (x=tmsleep, y=dactivetime))

+ geom\_point()

+ geom\_smooth(method='lm', formula='y~x')

+ggtitle((paste("Cor: ",cor(c2sleep\_actime$tmsleep, c2sleep\_actime$dactivetime)))) **# correlation between sleep and activity**

A graph with a line and dots

Description automatically generated> c3weight\_actime <- inner\_join(wkg, dactivetime, by = "id") **# Creating a table with weight and activity**

> str(c3weight\_actime)

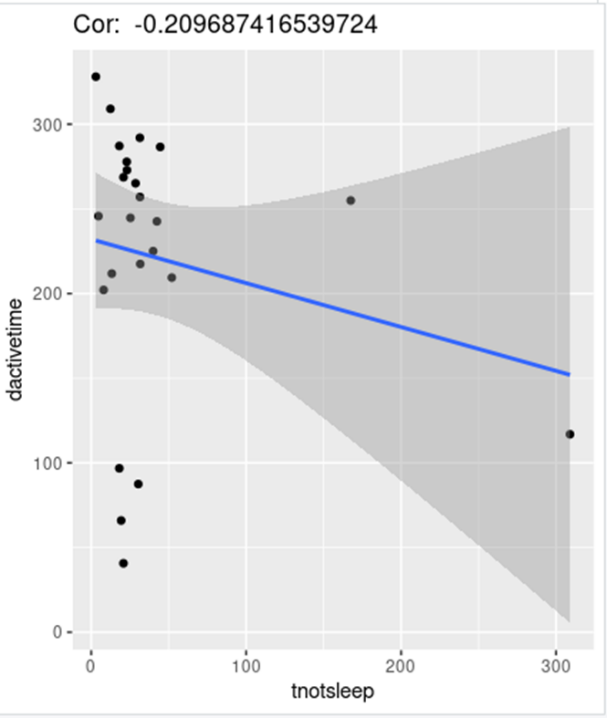
spc\_tbl\_ [8 × 4] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame) **# Basic stats**

> ggplot(data= c3weight\_actime, aes (x=wKG, y=dactivetime)) +

+geom\_point() +

+geom\_smooth(method='lm', formula='y~x')+

+ggtitle((paste("Cor:",cor(c3weight\_actime$ wKG, c3weight\_actime$dactivetime)))) **# Correlation between weight and activity**



> ggplot(data=c2sleep\_actime, aes( x=tnotsleep, y=dactivetime)) +

+ geom\_point() +

+geom\_smooth(method='lm', formula='y~x')+

+ggtitle((paste("Cor: ",cor( c2sleep\_actime$tnotsleep, c2sleep\_actime$dactivetime)))) **# Correlation between time not sleeping but in bed and activity**

Then it was time log and prepare the chosen dataset:

> str(dailyActivity\_merged)

spc\_tbl\_ [940 × 15] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame) **# Basic stats**

> str(heartrate\_seconds\_merged)

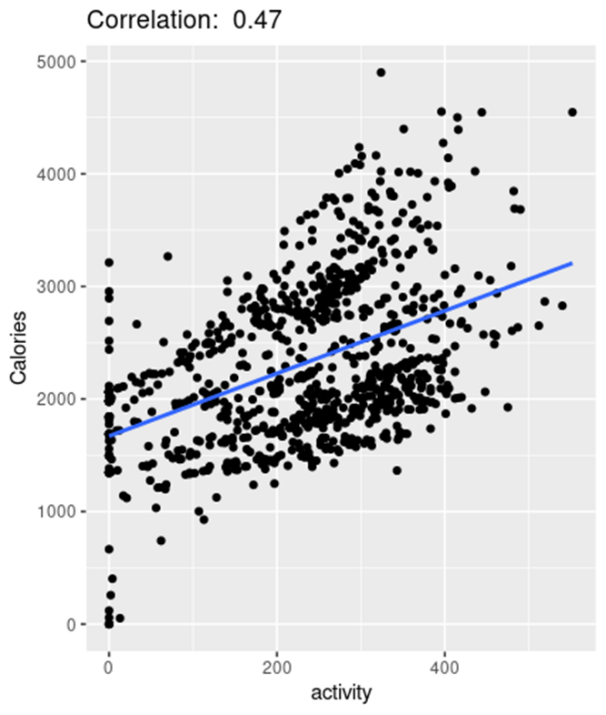
spc\_tbl\_ [2,483,658 × 3] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame) **# Basic stats**

> str(sleepDay\_merged)

spc\_tbl\_ [413 × 5] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame) **# Basic stats**

> str(weightLogInfo\_merged)

spc\_tbl\_ [67 × 8] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame) **# Basic stats**

> dailyActivity\_merged$activity<-dailyActivity\_merged$VeryActiveMinutes+dailyActivity\_merged$FairlyActiveMinutes+dailyActivity\_merged$LightlyActiveMinutes **#Creating a column with the sum of all the activity**

> ggplot(data = dailyActivity\_merged, aes(x = activity, y = Calories)) +

+ geom\_point() +

+ geom\_smooth(method = "lm", formula = y ~ x, se = FALSE)+ ggtitle(paste("Correlation: ", round(cor(dailyActivity\_merged$activity,dailyActivity\_merged$Calories),2))) **#** **Correlation between calories and activity**

A graph with a line and a blue line

Description automatically generated

> ggplot(data = dailyActivity\_merged, aes(x = TotalSteps, y = Calories)) +

+ geom\_point() +

+ geom\_smooth(method = "lm", formula = y ~ x, se = FALSE)+ ggtitle(paste("Correlation: ", round(cor(dailyActivity\_merged$TotalSteps,dailyActivity\_merged$Calories),2))) **#** **Correlation between calories and total steps**

A graph with a line and a blue line

Description automatically generated

> ggplot(data = dailyActivity\_merged, aes(x = TotalDistance, y = Calories)) +

+ geom\_point() +

+ geom\_smooth(method = "lm", formula = y ~ x, se = FALSE)+ ggtitle(paste("Correlation: ", round(cor(dailyActivity\_merged$TotalDistance,dailyActivity\_merged$Calories),2))) **#** **Correlation between calories and Total distance**

1. **Data Analysis**:

A diagram of data processing

Description automatically generated

This datased gave origin to three main tables in R:

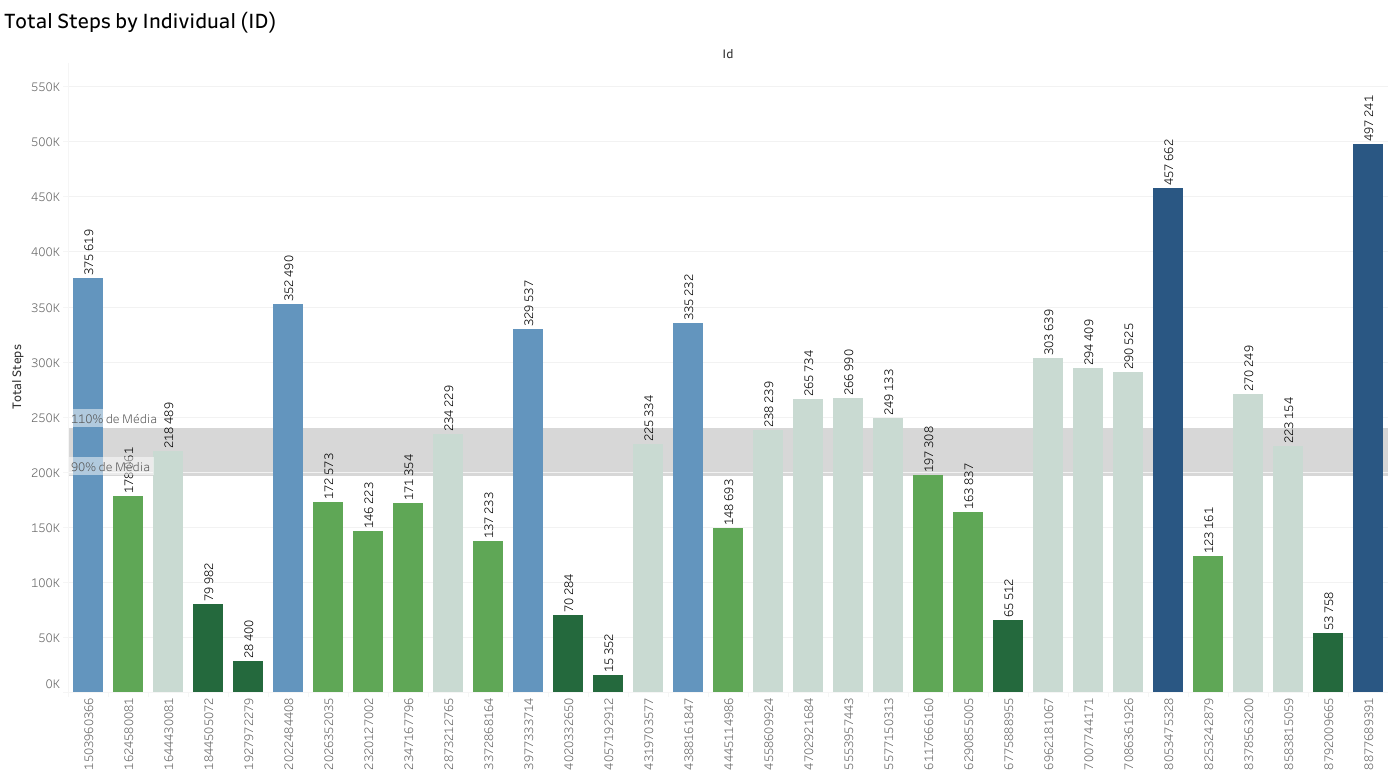
* Active\_time to heart\_rate – C1
  + A table with 14 record that permitted to create a correlation between the heart rate and the activity of the user, we had a result of -0,06, so a user who practice more sport has a lower heart rate, however the correlation it’s to weak so be relevant.
  + The same table helped making another correlation between calories burnt and activity, we had the result of 0,25, which means the number of calories burnt grows when the activity is incremented.
* Active\_time to sleep – C2
  + A table with 24 records that permitted to create a correlation between the hours of sleep and the activity of the user, we had a result of -0,23, so a user who practice more sport has a lower sleep time, the same happens with the time in bed before or without sleeping, with a correlation of -0,20
* Active\_time to weight – C3
  + A table with 8 records that permitted to create a correlation between the weight and the activity of the user, we had a result of -0,86, so a user who practice more sport has a lower weight.

A diagram of a dataset

Description automatically generated with medium confidence

This datased gave origin to three main tables in R:

* Due to the problem of low level of records we just explore the main table dailyActivity\_merged:
  + By making a correlation of 0,47 between calories and activity, It was possible to determine that each time the activity increases so increases the calories burnt;
  + The same happens with the number of steps, with a correlation of 0,59, each time we increment the number of steps so increments de calories burnt;
  + Thus, we stand with a correlation of 0,64 betweene total distance and calories, activity does increase the consumption of calories.



1. **Recommendations**:

First is important to state that we have a small sample of data, and more data must be gathered so we can improve our research and so have more meaningful results.

After processing and analysing results from this sample, we can state the following:

1. Due the number of users (ID’s) of the sample we can state that not all customers are using the complete pack of features of their devices, for example, out of 33 only 14 measure heart rate, only 22 measure the sleep and only 8 give inputs of weight, this is the first question we must ask: Why aren’t more users using the full capabilities of the device? is it difficult to input data or measure? is the device user-friendly?
2. The next question stands with the performance between device and app: Is it easy to establish a connection? Does it take a lot of battery to make it?
3. Which devices could we give to the user so they could improve the experience and at the same time record more data? For example, develop devices like weight scales.
4. The relation between user, app and device must be as easy as possible, it’s supposed to make the user comfortable and at the same time give him a slight will to use it frequently;
5. About the data, is possible to state that activity increases life quality. From reducing heart rate or weight to increase the calories burn being active can make us healthier and stronger;
6. Having this conclusions can help us develop a new way to interact with the customer:
   1. The leaf bracelet must have signals like vibrating alarms warning the user that is getting out of battery, that the user has not been active for a long period of time, or that accomplished any objective for the day or month….;
   2. Then this signals must bring warning messages to the app, smooth inputs to the user so it not looks like an order but a suggestion;
   3. Bringing inspirational messages to the user, can make him having the motivation to be more active;
   4. About the plans, we can input some metrics like averages to inform the user that they are performing well, for example, informing them that they accomplished the average number of steps for the day, or that they burned more calories that 80% of the users.

Being in constant interaction and support of our customers can make them feel supported and encouraged to go further. So, it’s important to show how he can improve and reach their goals