

|  |
| --- |
| Case Study 2: How Can a Wellness Technology Company Play It Smart? A  Bellabeat data analysis case study |
|  |
| November 3  Bellabeat  Authored by: Mohammed Tayfour |



# Table of Contents

# A clear summary of the business task

# A description of all data sources used

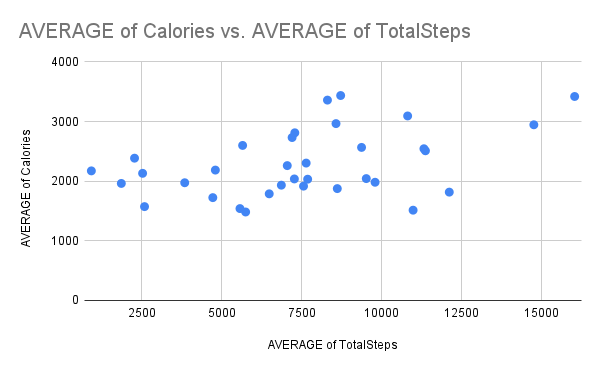
# Documentation of any cleaning or manipulation of data

# A summary of analysis

# Supporting visualizations and key findings

# Top high-level content recommendations based on your analysis

|  |
| --- |
| Summary of the business task I have been tasked to focus on one of Bellabeat’s products and analyze smart device data to gain insights into how consumers are using their smart devices.  The idea is to apply those insights in one of Bellabeat’s products.   1. Data sources used   Sršen recommended that I use public data that explores smart device users’ daily habits, specifically the FitBit Fitness Tracker dataset which was available on the site, Kaggle.com. These datasets were generated by respondents to a distributed survey via Amazon Mechanical Turk between 03.12.2016-05.12.2016. 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  I was also encouraged to overcome any limitations that may be faced with the chosen dataset.  The fitbit dataset was stored in an archive ZIP folder on my desktop Documentation of any cleaning or manipulation of dataImported dataset into Google sheets.Imported activity dataset and sleep datasetsMade sure that there were no nulls/blanks and no duplicatesThe ID number is duplicated, however this is due to the entries being daily for one month each2 pivot charts where created for the activity dataset. The first one is a scatter plot mapping the correlation between the average calories and the average distance tracked. As observed, the higher the distance an individual took, the higher the number of calories that were consumed on average. This is likely due to a non-sedentary lifestyle. |
| 1. Another pivot chart was created for the sleep dataset |



1. A summary on Analysis

On SQL:

* Changed data type to upload dataset
* Performed outer and inner join between activities and sleep\_info tables:

SELECT

        a.Id,

        a.TrackerDistance,

        a.Calories,

        b.TotalMinutesAsleep,

        b.TotalTimeInBed

FROM Fitabase\_data.activity\_info AS A

INNER JOIN

Fitabase\_data.sleep\_info AS B

ON a.Id = B.Id

2)

SELECT

        a.Id,

        a.TotalSteps,

        a.Calories,

        b.TotalMinutesAsleep,

        b.TotalTimeInBed

FROM Fitabase\_data.activity\_info AS A

FULL JOIN

Fitabase\_data.sleep\_info AS B

    ON a.Id = B.Id

On RStudio:

<https://24afb41bbd72474a9865135c1e39f215.app.rstudio.cloud/?view=rmarkdown>

# Supporting visualizations and key findings

# 

# 