

Case Study 2: How Can a Wellness Technology Company Play It Smart?

A

Bellabeat data analysis case study

DECEMBER 15

Bellabeat

Authored by: Mohammed Tayfour



Logo
Name

Table of Contents

1. A clear summary of the business task
2. A description of all data sources used
3. Documentation of any cleaning or manipulation of data
4. A summary of analysis
5. Supporting visualizations and key findings
6. Top high-level content recommendations based on your analysis

1. 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.

2. 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

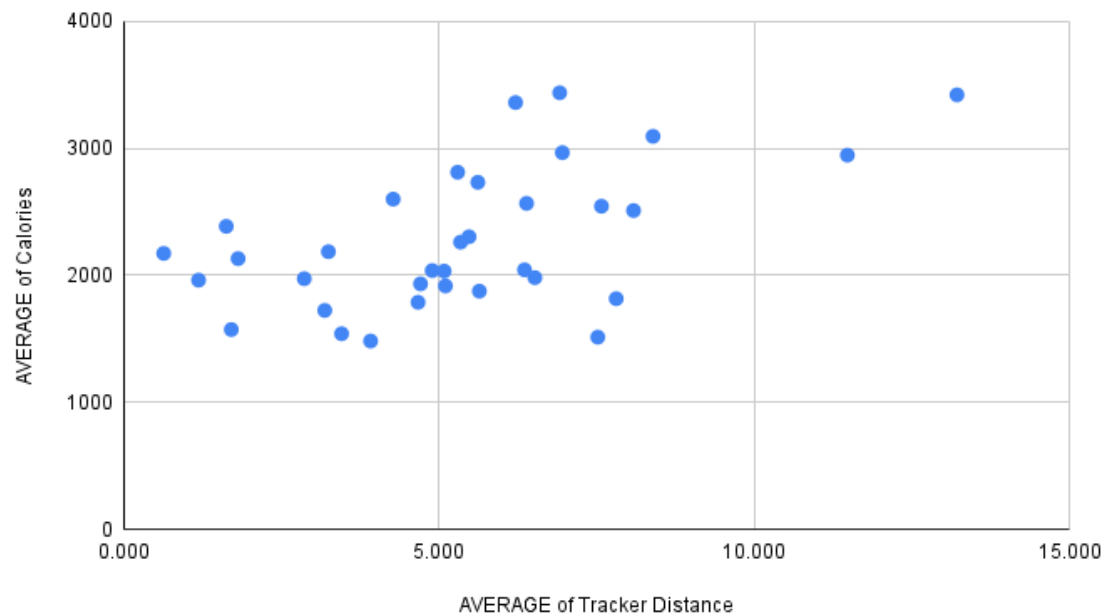
I was also encouraged to overcome any limitations that may be faced with the chosen dataset.

The fitbit dataset was stored in an archive ZIP folder on my desktop

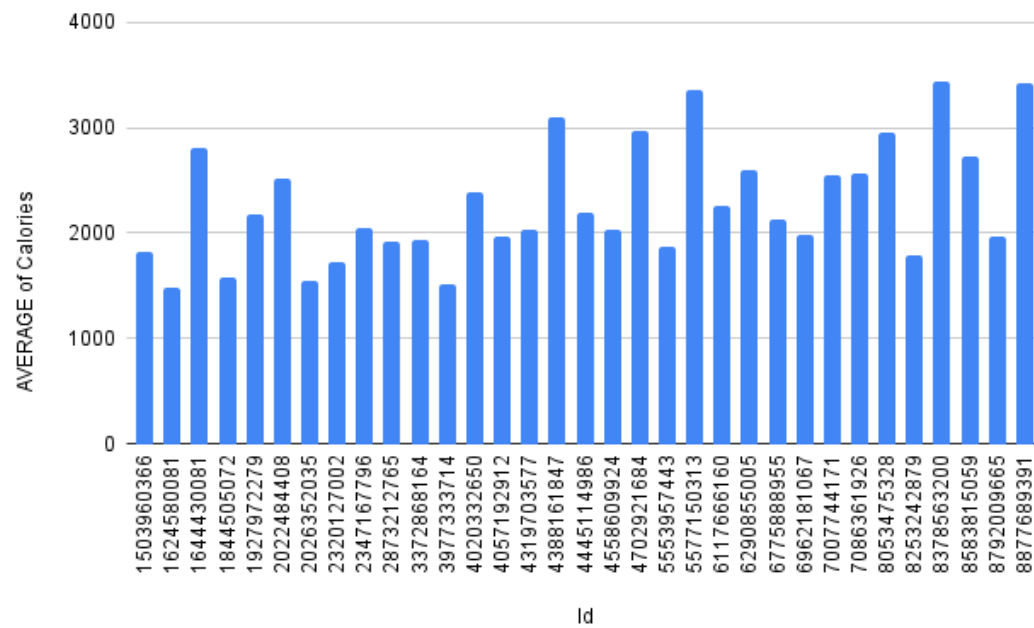
3. Documentation of any cleaning or manipulation of data

1. Imported dataset into Google sheets.
2. Imported activity dataset and sleep datasets
3. Made sure that there were no nulls/blanks and no duplicates
4. The ID number is duplicated, however this is due to the entries being daily for one month each
5. 2 pivot charts were 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.

AVERAGE of Calories vs. AVERAGE of Tracker Distance

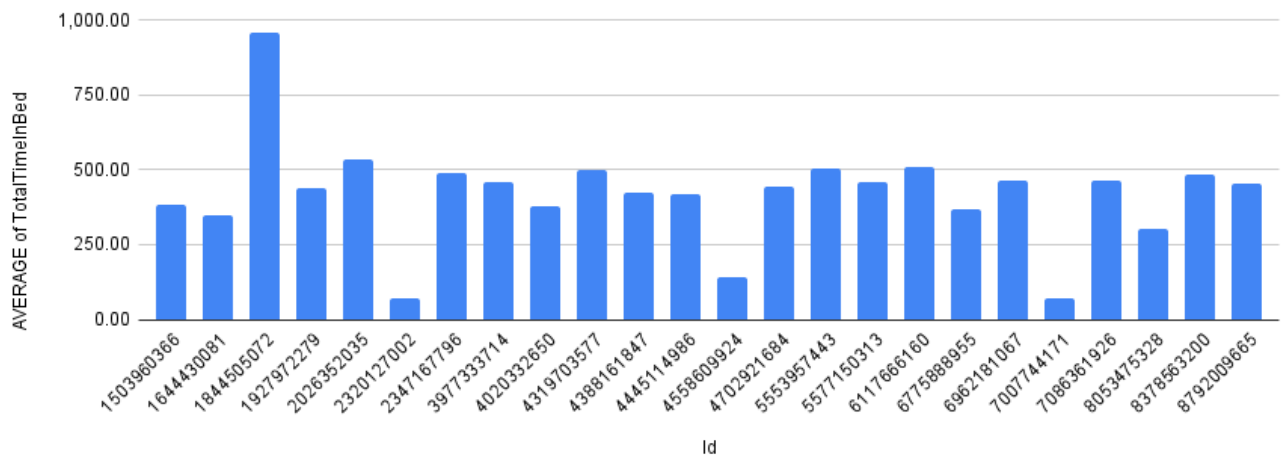


AVERAGE of Calories vs. Id

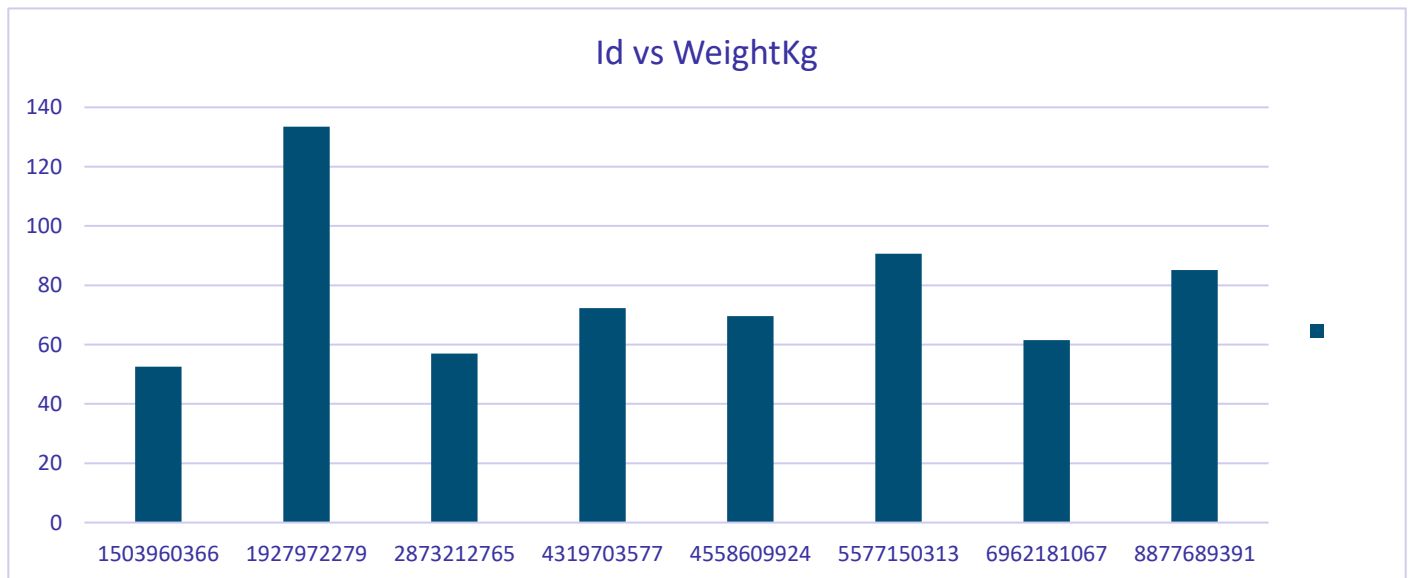


6. Another pivot chart was created for the sleep dataset

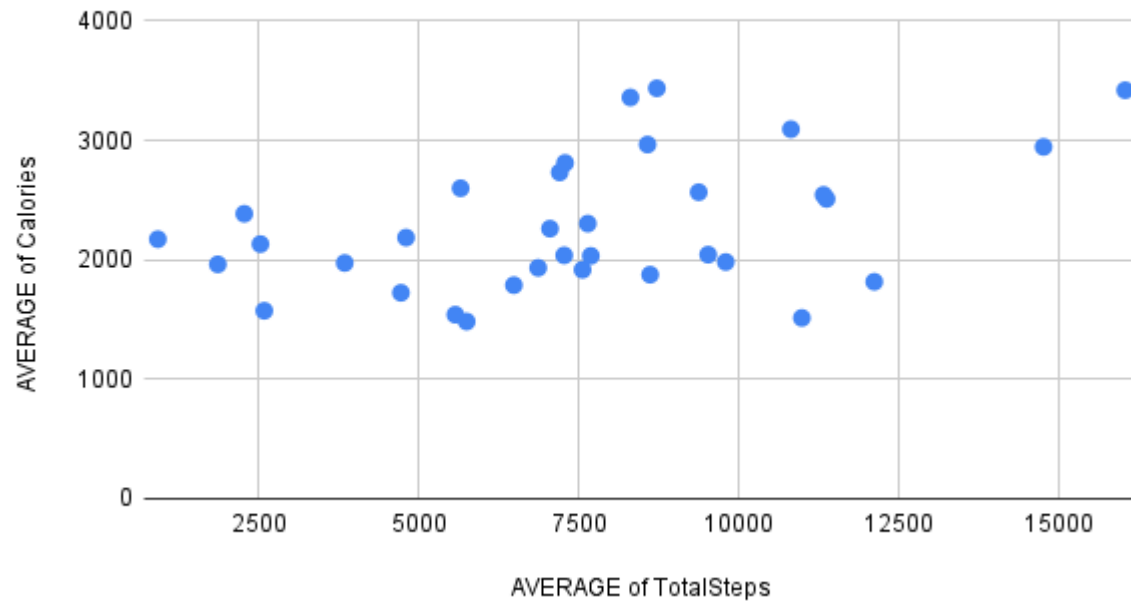
ID vs AVERAGE of TotalTimeInBed



Id vs WeightKg



AVERAGE of Calories vs. AVERAGE of TotalSteps



4. 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)

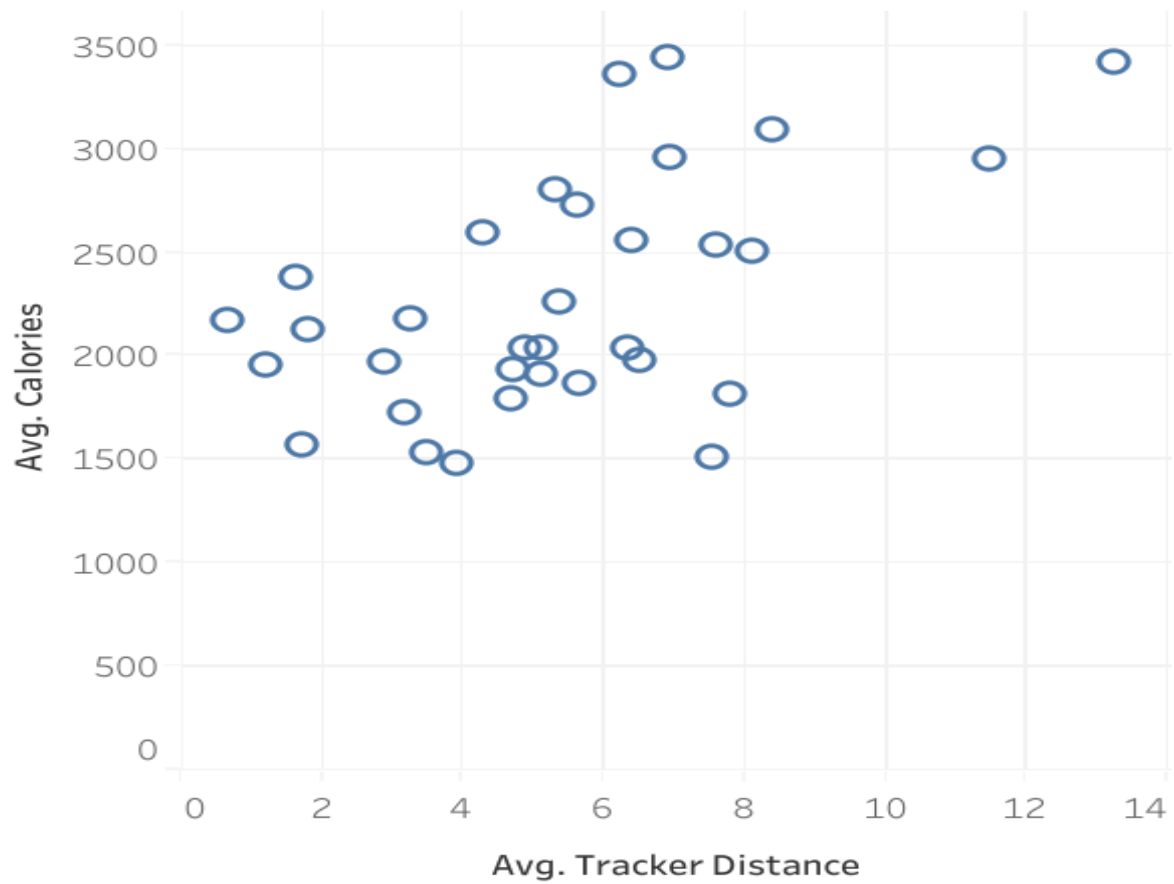
```
SELECT
    a.Id,
    a.TotalSteps,
    a.Calories,
    b.TotalMinutesAsleep,
    b.TotalTimeInBed
FROM Fitabase_data.activity_info AS A
FULL JOIN
Fitabase_data.sleep_info AS B
ON a.Id = B.Id
```

On RStudio:

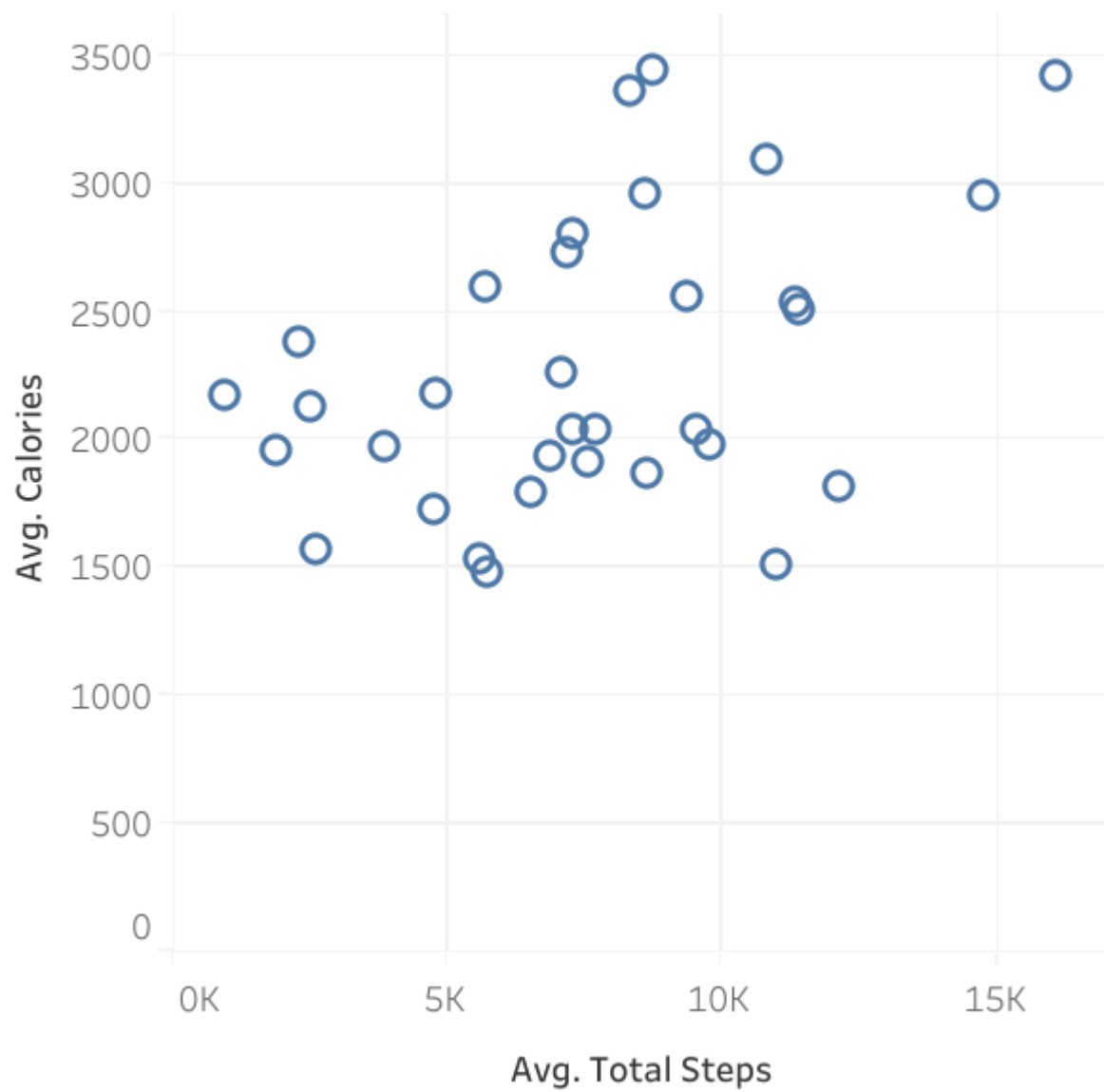
<https://24afb41bbd72474a9865135c1e39f215.app.rstudio.cloud/?view=rmarkdown>

5. Supporting visualizations and key findings

Bellabeat



AVG TOTAL STEPS VS AVG CALORIES



TOTAL MINUTES ASLEEP VS TOTAL TIME IN BED

