

# Merit America Google Data Analytics Certificate Capstone Project

## Case Study: 2

### Bellabeat Analysis

Author: Alexandra Wilkison

Date: 08/07/2023

Email: a.wilkison1026@gmail.com

[GitHub](#)

[LinkedIn](#)

## Ask - Introduction

Welcome to my Capstone project for the Google Data Analytics Certificate through the Data Analytics program with Merit America.

I was tasked with conducting a case study, imagining myself to be a marketing analyst for the company, **Bellabeat**, a high-tech manufacturer of health-focused products for women.

## About the company

Bellabeat is a high-tech company that manufactures health-focused, smart products. Collecting data on activity, sleep, stress, and reproductive health has allowed Bellabeat to empower women with knowledge about their own health and habits, and grow rapidly as a company.

## Key Stakeholders

- Urška Sršen: Cofounder and Chief Creative Officer
- Sando Mur: Cofounder and Mathematician

## Questions

- What are some trends in smart device usage?
- How could these trends apply to Bellabeat customers?
- How could these trends help influence Bellabeat's marketing strategy?

## Deliverables

- 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 the analysis
- Supporting visualizations and key findings
- Top high-level recommendations based on analysis

## PREPARE – Upload and Inspect Data

### Data Source

#### [FitBit Fitness Tracker Data](#)

- 18 datasets were generated by respondents to a distributed survey via Amazon Mechanical Turk between 03.12.2016-05.12.2016.
- 30 eligible Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring.

### Limitations

The data collected consists of 30 users over a period of 30 days and was conducted in 2016. Not only is this dataset outdated, the sample size is too small to conduct a proper analysis. The data also does not include any information on the demographics for the participants, meaning the data might not represent Bellabeat's target audience- women.

This case study will function as a high-level overview, showcasing my analysis skills and providing broad insights.

## PROCESS – Clean Data

### Clean Data In Excel

I chose to work with the following datasets in one Excel workbook:

- dailyActivity\_merged
- sleepDay\_merged
- hourlySteps\_merged
- hourlyCalories\_merged

These datasets were chosen out of a total of 18 datasets available because the information in the majority of the datasets can be found in the three previously listed.

I took the following steps to **clean and organize** each dataset:

I loaded the three datasets into the same Excel workbook and renamed the different sheets as follows:

- dailyActivity\_merged = daily\_activity
- sleepDay\_merged = sleep\_day
- hourlySteps\_merged = hourly\_steps
- hourlyCalories\_merged = hourly\_calories

I used conditional formatting in the daily\_activity sheet to visually differentiate between the active distance and active minutes columns.

I filtered the ID column in each sheet to verify the number of unique ID numbers were in each dataset with the following results:

- daily\_Activity: 33
- sleep\_day: 24
- hourly\_steps: 33
- hourly\_calories: 33

Formatted date data in MM/DD/YYYY date format

Removed the LoggedActivitiesDistance column from the daily\_activity sheet because there wasn't enough data input to result in significant analysis.

Added columns in the sleep\_day sheet called "TotalTimeAsleep" and "TotalTimeInBed" to show the times as hours

Created a new column in daily\_activity sheet named "Activity Day" to show what day of the week each entry was logged, based off of the dates in the column "Activity Date."

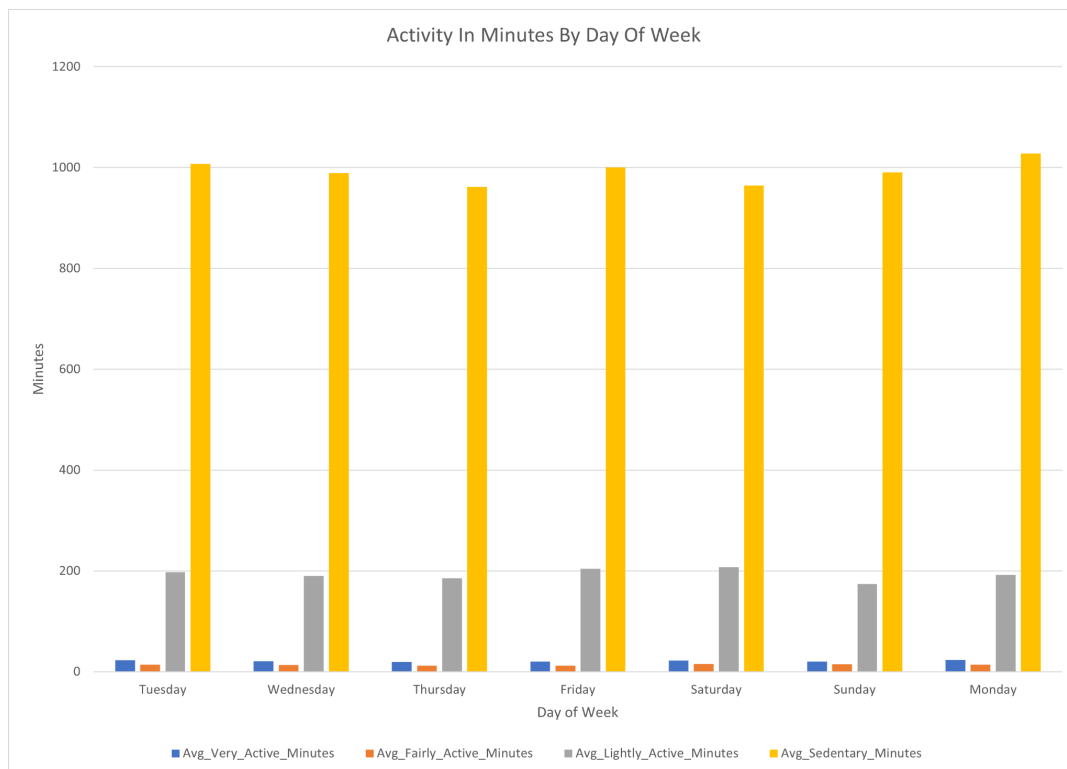
## ANALYZE & SHARE

I chose to work in BigQuery for my analysis. I uploaded the datasets to BigQuery under the project name "data-analytics-capstone-394814.Bellabeat\_Case\_Study"

I wanted to start off by finding out if there was a certain day of the week that users were most active or least active.

```
/*Activity levels in minutes listed by days of the week to see if users were more active on a certain day of the week*/
```

```
SELECT ActivityDay,  
  ROUND (avg(VeryActiveMinutes), 2) AS Avg_Very_Active_Minutes,  
  ROUND (avg(FairlyActiveMinutes), 2) AS Avg_Fairly_Active_Minutes,  
  ROUND (avg(LightlyActiveMinutes), 2) AS Avg_Lightly_Active_Minutes,  
  ROUND (avg(SedentaryMinutes), 2) AS Avg_Sedentary_Minutes,  
FROM  
  `data-analytics-capstone-394814.Bellabeat_Case_Study.daily_activity`  
GROUP BY  
  ActivityDay;
```



The results showed that there were no major differences in activity levels from one day to the next.

From here, I decided to examine the minimum, maximum, and average of total steps, total distance, calories, and activity levels grouped by ID.

/\*the minimum, maximum, and average of total steps, total distance, calories, and activity levels grouped by ID\*/

SELECT

id,  
MIN(TotalSteps) AS Min\_Total\_Steps,  
MAX(TotalSteps) AS Max\_Total\_Steps,  
AVG(TotalSteps) AS Avg\_Total\_Steps,  
MIN(TotalTrackerDistance) AS Min\_Total\_Distance,  
MAX(TotalTrackerDistance) AS Max\_Total\_Distance,  
AVG(TotalTrackerDistance) AS Avg\_Total\_Distance,  
MIN(Calories) AS Min\_Total\_Calories,  
MAX(Calories) AS Max\_Total\_Calories,  
AVG(Calories) AS Avg\_Total\_Calories,  
MIN(VeryActiveMinutes) AS Min\_Very\_Active\_Minutes,  
MAX(VeryActiveMinutes) AS Max\_Very\_Active\_Minutes,  
AVG(VeryActiveMinutes) AS Avg\_Very\_Active\_Minutes,  
MIN(FairlyActiveMinutes) AS Min\_Fairly\_Active\_Minutes,  
MAX(FairlyActiveMinutes) AS Max\_Fairly\_Active\_Minutes,  
AVG(FairlyActiveMinutes) AS Avg\_Fairly\_Active\_Minutes,  
MIN(LightlyActiveMinutes) AS Min\_Lightly\_Active\_Minutes,  
MAX(LightlyActiveMinutes) AS Max\_Lightly\_Active\_Minutes,  
AVG(LightlyActiveMinutes) AS Avg\_Lightly\_Active\_Minutes,  
MIN(SedentaryMinutes) AS Min\_Sedentary\_Minutes,  
MAX(SedentaryMinutes) AS Max\_Sedentary\_Minutes,  
AVG(SedentaryMinutes) AS Avg\_Sedentary\_Minutes

FROM

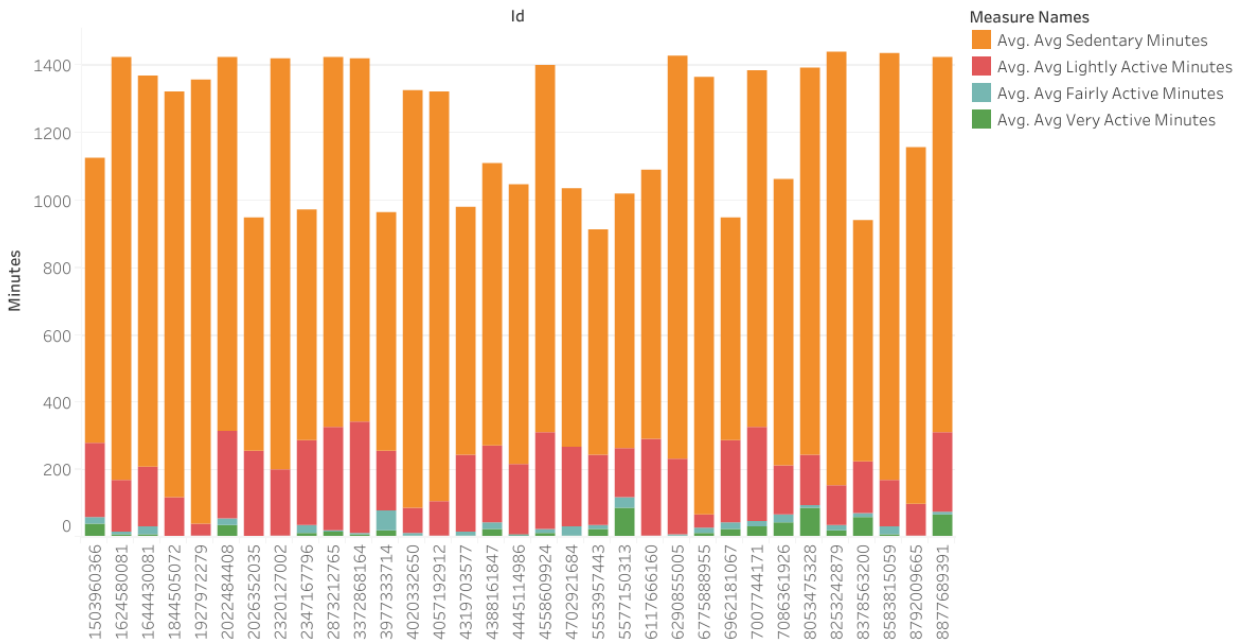
`data-analytics-capstone-394814.Bellabeat\_Case\_Study.daily\_activity`

GROUP BY

Id;

After looking through the results of this query, I chose to focus on the Activity Levels in Minutes by ID number.

### Activity Levels



This showed that the majority of the logged activity was considered “Sedentary.”

I wanted to see if users were reaching the recommended 10,000 steps per day as well as examine possible correlations between steps and calories burned, as well as steps and sleep.

(Information on the 10,000 step recommendation and benefits can be found with JAMA - Journal of American Medical Association [here](#))

(Information for the recommended 7-9 hours of sleep can be found with the National Institute of Health [here](#))

```
SELECT
  id,
  AVG(TotalSteps) AS Avg_Total_Steps,
  AVG(TotalTrackerDistance) AS Avg_Total_Distance,
  AVG(Calories) AS Avg_Calories_Burned
FROM
  `data-analytics-capstone-394814.Bellabeat_Case_Study.daily_activity`
GROUP BY
  id;
/*total steps vs total minutes asleep by id*/
SELECT
  Activity.Id,
  AVG(Activity.TotalSteps) AS Avg_Total_Steps,
  AVG(Sleep.TotalMinutesAsleep) AS Avg_Total_Minutes_Asleep,
FROM
  `data-analytics-capstone-394814.Bellabeat_Case_Study.daily_activity` AS Activity
```

JOIN

```
`data-analytics-capstone-394814.Bellabeat_Case_Study.sleep_day` AS Sleep
```

ON

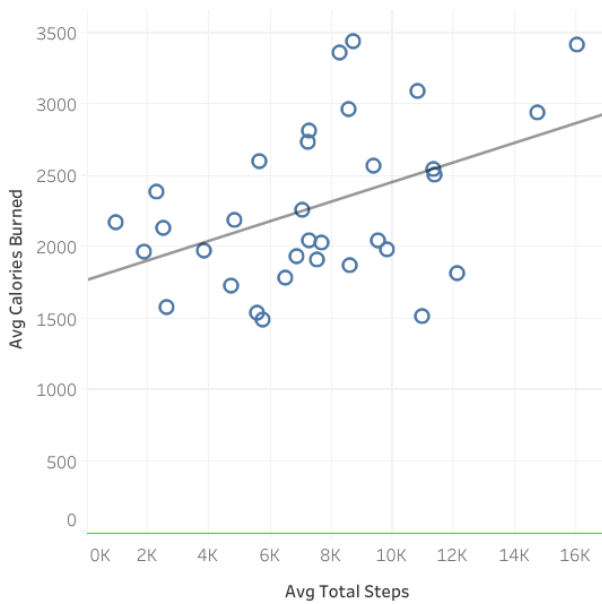
```
Activity.Id = Sleep.Id
```

GROUP BY

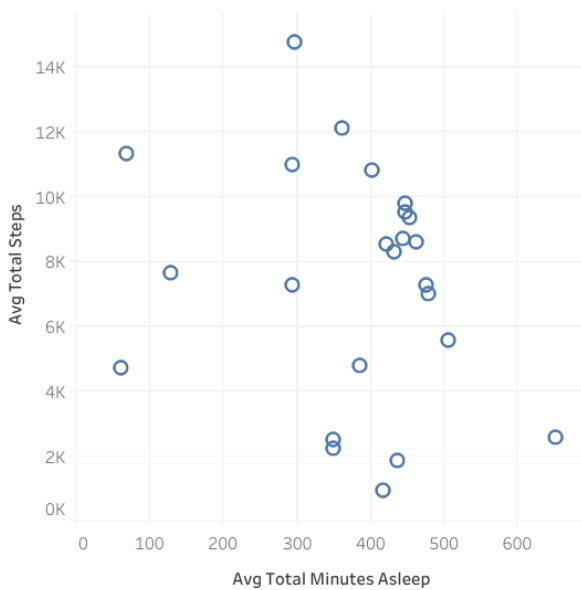
```
Activity.Id;
```

With the results from those queries, I was able to generate the following two charts:

AvgTotalSteps\_AvgCaloriesBurned



AvgTotalMinutesAsleep\_AvgTotalSteps



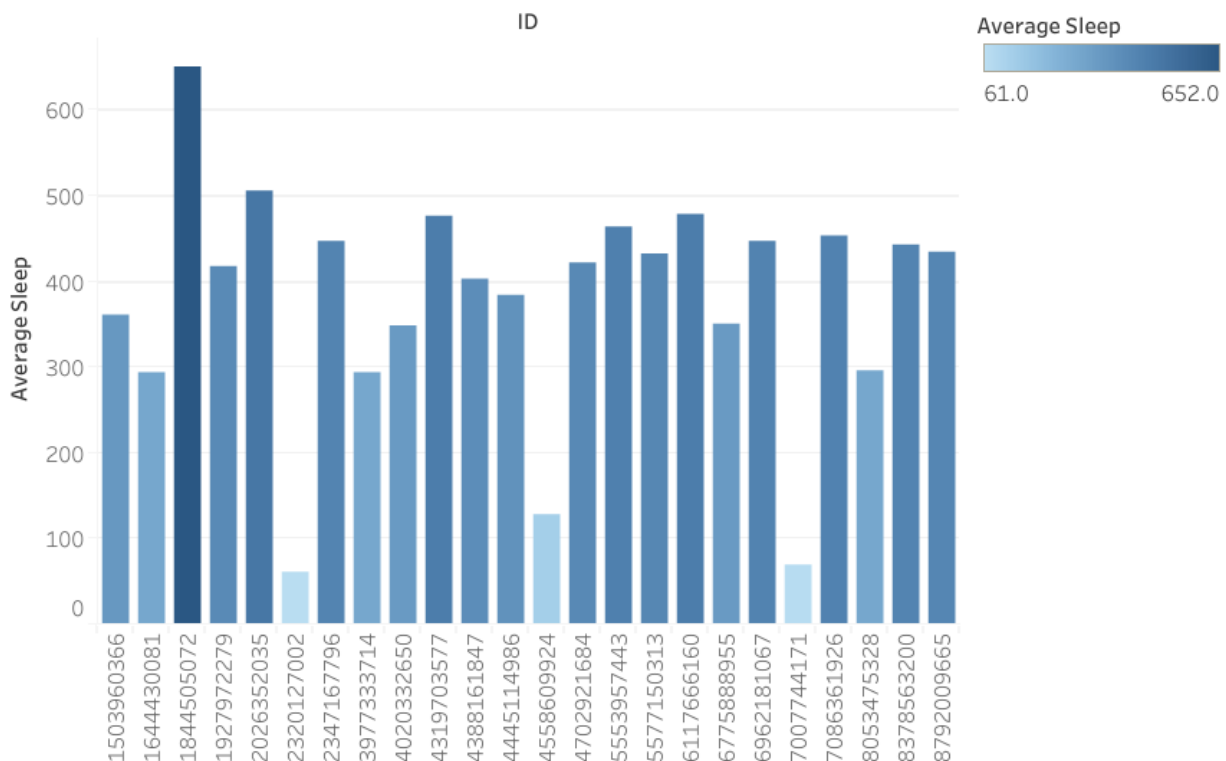
“AvgTotalSteps\_AvgCaloriesBurned” confirms that the more steps you take, the more calories you burn.

“AvgTotalMinutesAsleep\_AvgTotalSteps” revealed that only 5 of the users logged an average of over 10,000 steps and only 2 users logged 8 or more hours of sleep on average. (8 hours = 480 minutes).

Lastly, I wanted to examine the average sleep logged by each user more closely.

```
/*average minutes asleep by ID*/  
SELECT ID, AVG(TotalMinutesAsleep) AS average_sleep  
FROM `data-analytics-capstone-394814.Bellabeat_Case_Study.sleep_day`  
GROUP BY ID;
```

AvgSleepByUserID



The results showed me that 1/3 of the users (11) were logging average sleep times below the recommended amount.

## ACT

### Conclusion:

Through my analysis I was able to uncover the following:

- There were no major differences in activity levels from one day to the next.
- The majority of the logged activity was considered “Sedentary.”
- The more steps you take, the more calories you burn.
- Only 5 of the users logged an average of over 10,000 steps and only 2 users logged 8 or more hours of sleep on average.
- 1/3 of the users (11) were logging average sleep times below the recommended amount.

### High-Level Recommendations:

1. I would recommend that Bellabeat incentivize their users to log more steps. This could be through give-back programs where Bellabeat donates a certain dollar amount to charitable organizations for every 10,000 steps the users log. You could also allow the users to bank their votes for every 10,000 steps so they can then vote for which organization they want their step contribution to go to. Another incentive could be banking step points to redeem for Bellabeat merch or other health and wellness items or for coupons.
2. Bellabeat could tailor their products towards individual user-needs in order to encourage more product use. One user might be more interested in logging their steps, while another would find more benefit in sleep reminders. When users get their fitbit, they could set up their interests like "step goals, daily mantras, sleep reminders, etc. If the user experience is tailored to what the individual user is interested in, you're likely to see growth.
3. Only 24 of the 33 users logged their sleep during the study. This could be due to the discomfort some people feel wearing watches or other accessories while sleeping. Bellabeat could target efforts towards design to make their products more comfortable.
4. With  $\frac{1}{3}$  of the users not getting the recommended amount of sleep, on average, Bellabeat could incorporate sleep reminders for their users to encourage them to use the feature.