

Merit America Google Data Analytics Certificate Capstone Project

Case Study #2: Bellabeat Analysis

Author: Joseph Solorio Arellano

Date: 06/24/2024

Email: solorioarellanojoseph@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 provided the task of conducting a case study, with the scenario of myself as 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 Not only is this dataset outdated, the sample size is too small to conduct a proper analysis.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 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 to make the data more visibly pleasing

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 "aesthetic-kite-427400-k2"

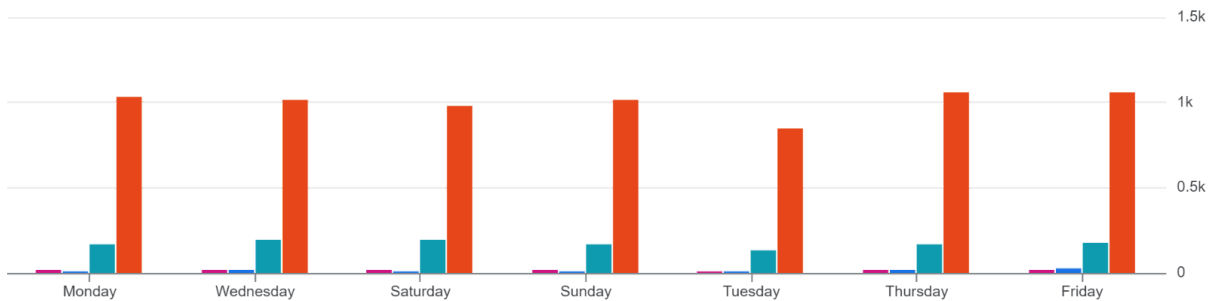
I wanted to start off by finding out if there was a certain day of the week that users were most active or least active.

```

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;

```

Avg_Very_Active_Minutes, Avg_Fairly_Active_Minutes, Avg_Lightly_Active_Minutes, Sedentary_Minutes by ActivityDay



Reviewing the results from the graph above, it shows that no matter which day of the week, the different types of minutes show similar results for each day.

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(TrackerDistance) AS Min_Total_Distance,
  MAX(TrackerDistance) AS Max_Total_Distance,
  AVG(TrackerDistance) 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,

```

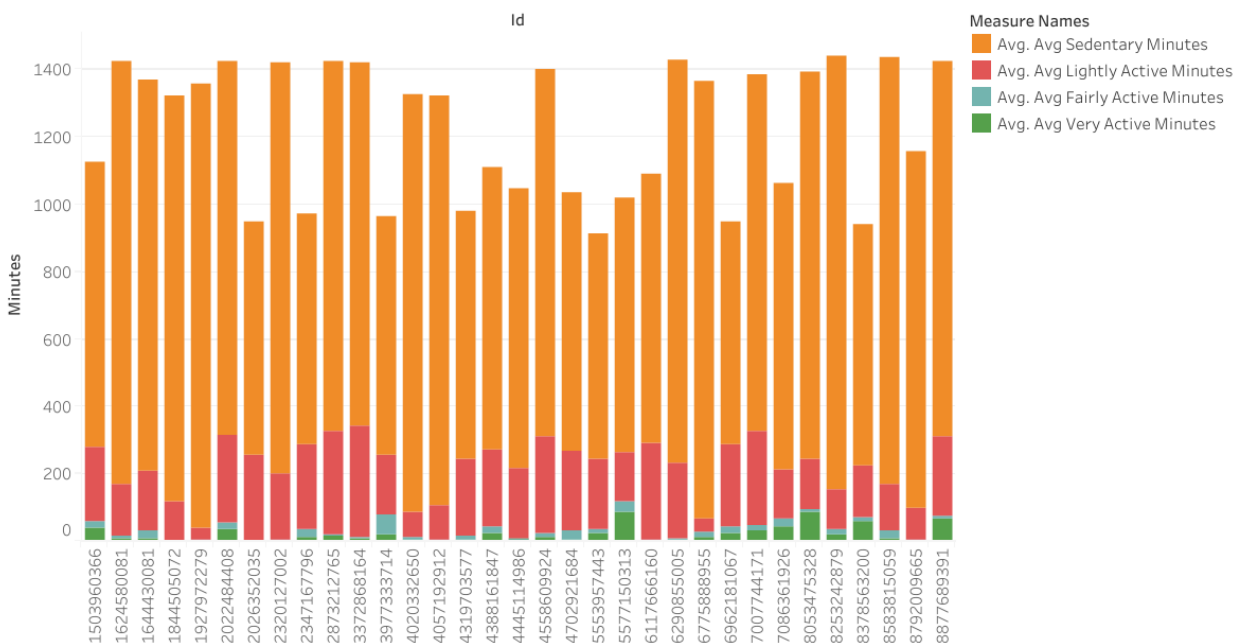
```

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



By looking at the chart above, you can see that a lot of the activity levels show more sedentary minutes than other categories.

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.

```

SELECT
    id,
    AVG(TotalSteps) AS Avg_Total_Steps,
    AVG(TotalTrackerDistance) AS Avg_Total_Distance,
    AVG(Calories) AS Avg_Calories_Burned

```

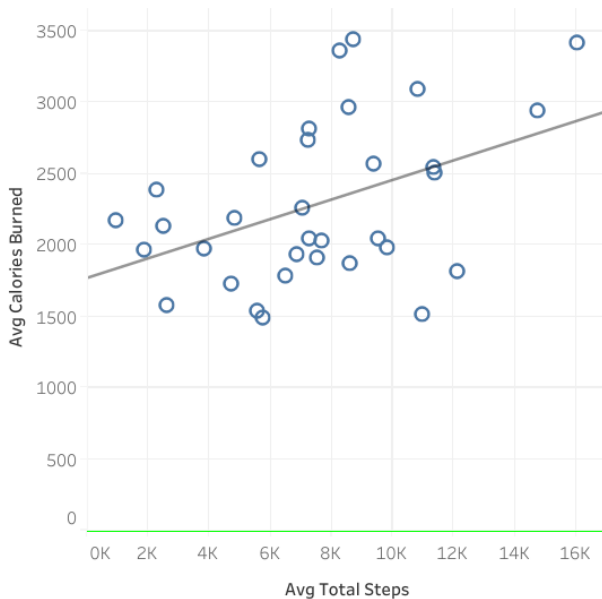
```

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;

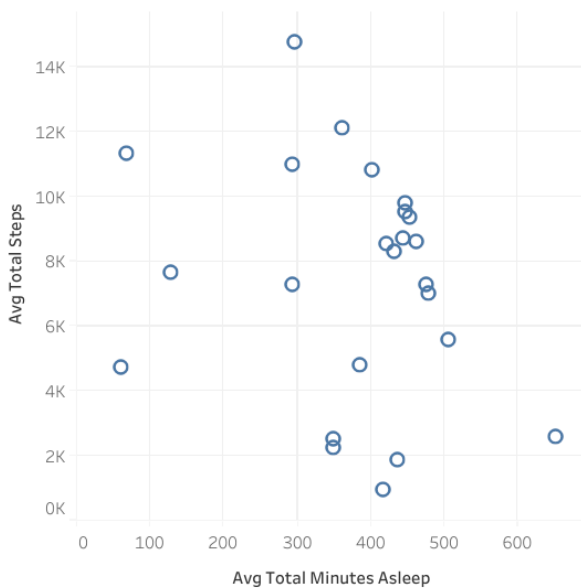
```

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

AvgTotalSteps_AvgCaloriesBurned



AvgTotalMinutesAsleep_AvgTotalSteps



“AvgTotalSteps_AvgCaloriesBurned” chart confirms the theory that the more steps you take, the more calories you burn.

“AvgTotalMinutesAsleep_AvgTotalSteps” chart reveals 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.

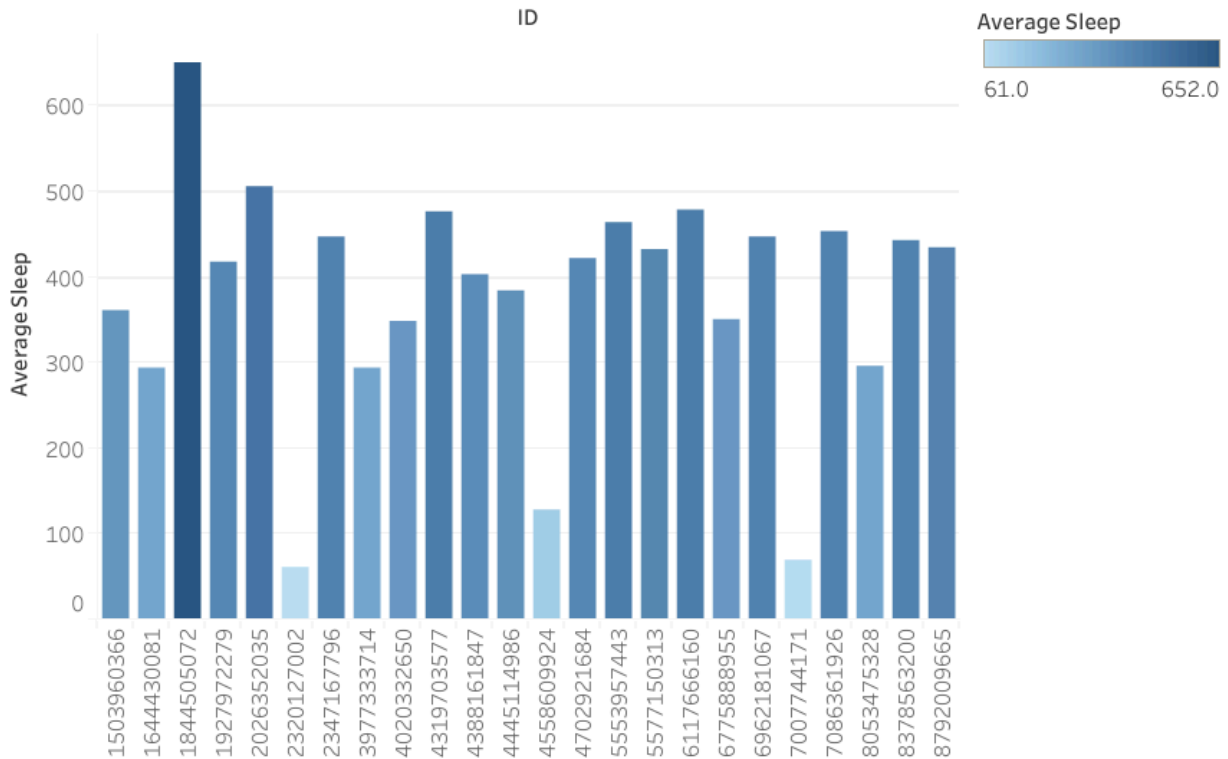
```
/*average minutes asleep by ID*/
```

```
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 either a particular program where after a certain amount of steps that they achieve a day or week, the customer can choose the foundation that is partnered with Bellabeat so it can be a symbiotic relationship with both organizations. Another way to incentivize users to log more steps is having a reward program where the consumer can win gift cards, discounts to various training facilities, diet programs, or cash back to purchase items from an authorized reseller.
2. Based on some of the graphs displayed above, Bellabeat could tailor their products towards individual user-needs in order to encourage more product use. Not everyone uses the same program or tracking information for themselves. Some consumers may use the step tracker more

than the sleep tracker because they are more comfortable using the step tracker. Bellabeat could diversify their products to use only certain programs like the step tracker or sleep tracker individually.

3. With $\frac{1}{3}$ of the users not getting the recommended amount of sleep, on average, Bellabeat could institute more emphasis with the sleep tracking program. For example, sleep reminders to turn off devices and set an alarm for their normal routine can be a great way to incorporate the sleep tracking program into the everyday life of a consumer.