

Google Data Analytics Certificate Capstone Project

Case Study: Bellabeat Analysis

Author: Rachel Allison

Date: 05/24/2024

Email: rachelahammo@gmail.com

LinkedIn: [linkedin.com/in/rachel--h](https://www.linkedin.com/in/rachel--h)

Tools used: Excel, SQL, Tableau

Introduction

For my Capstone project for the Google Data Analytics Certificate I analyzed data from Bellabeat, a high-tech manufacturer of health-focused products for women.

Ask

These questions guided my analysis:

1. What are some trends in smart device usage?
2. How could these trends apply to Bellabeat customers?
3. How could these trends help influence Bellabeat marketing strategy?

Prepare

I used the public data [FitBit Fitness Tracker Data](#) from Kaggle. The data contains personal fitness tracker data from thirty fitbit users. There are 18 different datasets containing data collected between 03/12/2016-05/12/2016.

There are limitations with the datasets:

- The sample size consists of 30 unique users over a 30 day period
- The data is outdated as it comes from 2016
- The data does not contain any demographic information

Because of these limitations in the datasets I chose to do an overview analysis using two of the datasets: "dailyActivity_merged" and "sleepDay_merged".

Process

I first renamed the datasets:

dailyActivity_merged = daily_activity

sleepDay_merged = sleep_day

I then loaded the datasets into Excel to clean the data.

daily_activity

Formatted date to MM/DD/YYYY

Added **ActivityDay** column to show the day of the week in addition to date

sleep_day

Removed 3 duplicate rows

Formatted date to MM/DD/YYYY

Analyze & Share

I loaded the cleaned datasets into BigQuery to analyze using SQL.

I first wanted to find the number of days each unique ID used their device.

```
/*Finding number of uses for each ID number*/  
SELECT  
  Id, COUNT (Id) AS Total_ID  
FROM  
  `treeproject-415820.bellabeat_casestudy.daily_activity`  
GROUP BY  
  Id  
ORDER BY  
  Total_ID DESC;
```

I found that 21 users used their devices for 31 days.

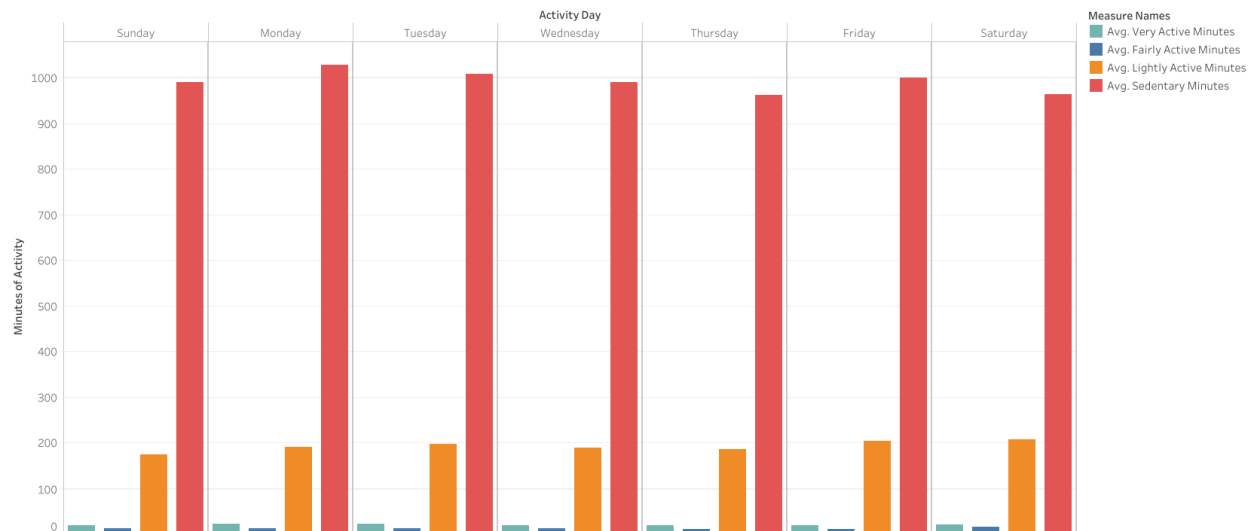
# of users	# of days device used
21	31
3	30
2	29
1	28
2	26
1	20
1	19
1	18
1	4

I then wanted to see if users' activity level differs by day of the week.

This showed me that activity levels do not show a significant difference by day of the week. I also noticed that the majority of activity is sedentary or light activity.

```
/*Finding average activity level by day of the week to see if/what days users are the most and 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
  `treeproject-415820.bellabeat_casestudy.daily_activity`
GROUP BY
  ActivityDay;
```

Activity levels by days of the week

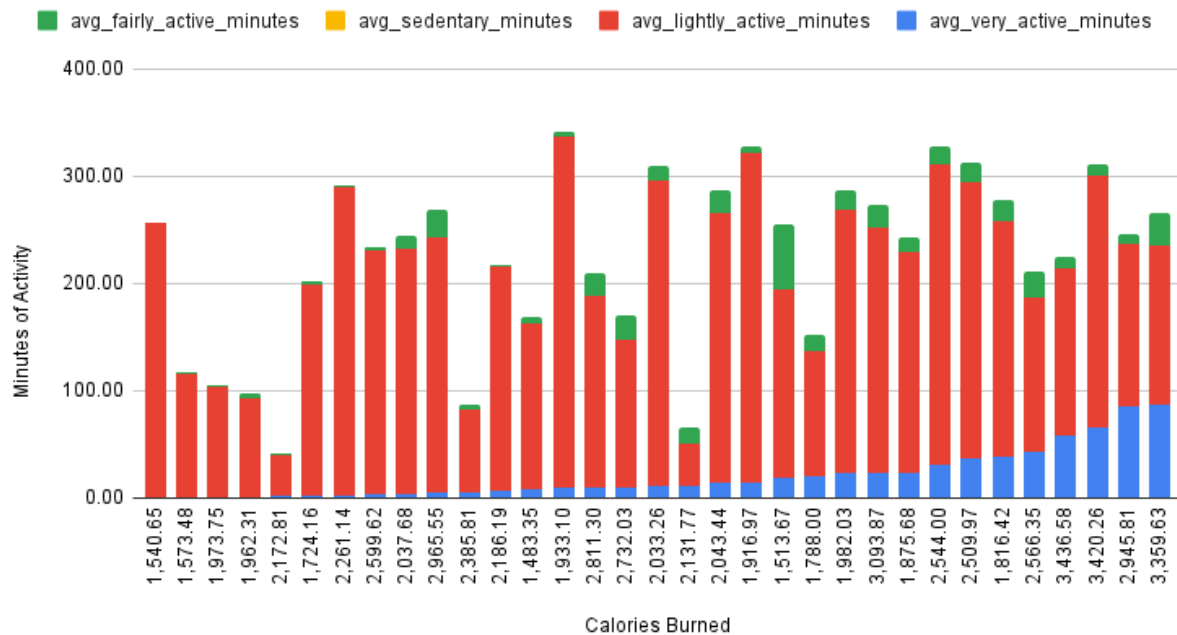


I looked at activity level and compared it to calories burned.

This confirmed that the majority of the calories burned are from light activity. This also shows that users who had more very active minutes tended to burn more calories.

```
/*Finding average of TotalSteps, VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes,
SedentaryActiveMinutes, and Calories*/
SELECT
  Id,
  AVG(TotalSteps) AS avg_total_steps,
  AVG(VeryActiveMinutes) AS avg_very_active_minutes,
  AVG(FairlyActiveMinutes) AS avg_fairly_active_minutes,
  AVG(LightlyActiveMinutes) AS avg_lightly_active_minutes,
  AVG(SedentaryActiveDistance) AS avg_sedentary_minutes,
  AVG(Calories) AS avg_calories,
FROM
  `treeproject-415820.bellabeat_casestudy.daily_activity`
GROUP BY
  Id;
```

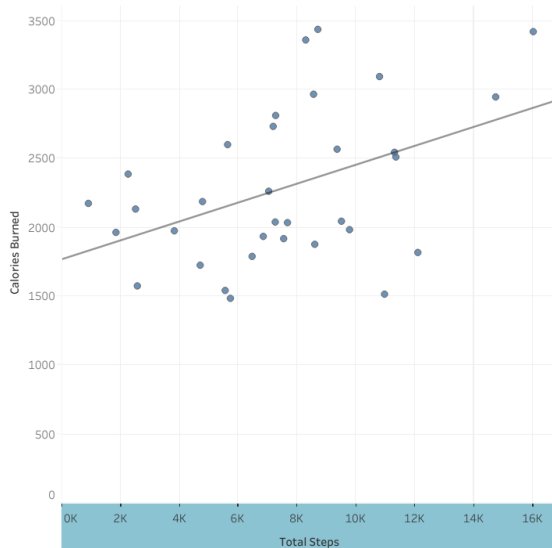
Calories burned from Activity



I wanted to see if higher step count would equate to higher calories burned. This showed me that there is a trend between step count and calories burned.

```
/*Comparing average TotalSteps and Calories*/
SELECT
  Id,
  ROUND (AVG(TotalSteps), 2) AS avg_total_steps,
  ROUND (AVG(Calories), 2) AS avg_calories,
FROM
  `treeproject-415820.bellabeat_casestudy.daily_activity`
GROUP BY
  Id;
```

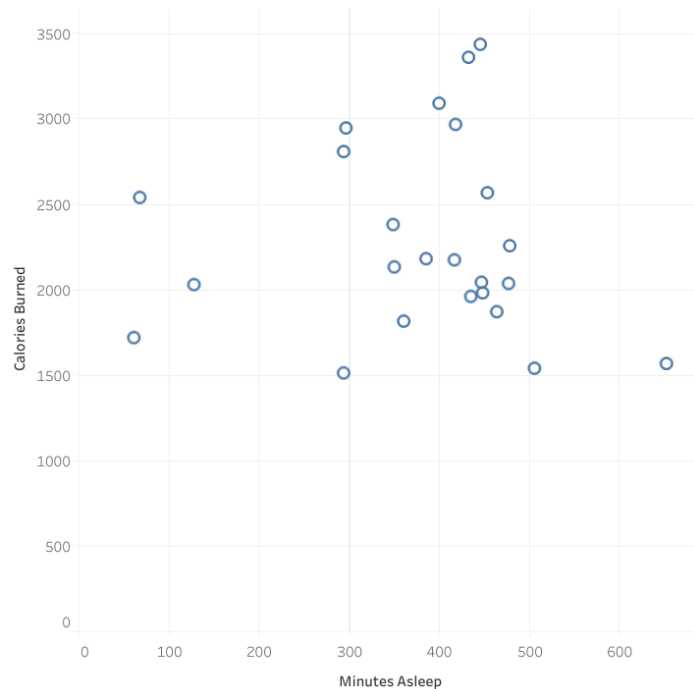
Calories burned vs Total Steps



Finally I wanted to compare step count, calories, and sleep time.
This did not show a significant trend in minutes slept to calories burned.

```
/*Comparing average TotalSteps, Calories, and TotalMinutesAsleep*/  
SELECT  
  Activity.Id,  
  AVG(TotalSteps) AS avg_total_steps,  
  AVG(Calories) AS avg_calories,  
  AVG(TotalMinutesAsleep) AS avg_total_minutes_asleep,  
FROM `treeproject-415820.bellabeat_casestudy.daily_activity` AS Activity  
JOIN  
  `treeproject-415820.bellabeat_casestudy.sleep_day` AS Sleep  
ON  
  Activity.Id = Sleep.Id  
GROUP BY  
  Activity.Id;
```

Calories burned vs Total Minutes Asleep



[View Slideshow presentation](#)

[View Visuals on Tableau](#)

Act

My findings conclude that:

1. Users seem more likely to do very light activity than very active activity
2. The more steps taken the more calories burned
3. Majority of users day is sedentary

From these findings these are the next steps the stakeholders could take:

- In-App competition with friends and family.
- Shareable rewards or badges for hitting certain step count.
- Reminders throughout the day to get up and move around
- Create in-app workouts that vary in time and intensity
- Walking/Running playlists to link to spotify or apple music
- Partner with popular athletic wear company

I would also advise conducting a new study on Bellabeat customers specifically as to get a better representation of current customers that includes demographics.