Google Data Analytics: Bellabeat Case Study

Melody Cheng

2025-07-28

Hello! This case study is part of the **Google Data Analytics Professional Certificate** on Coursera. Over the course of the program, I developed skills in data cleaning, analysis, and visualization using tools such as R, SQL, and Tableau. In this case study, I will walk through my analytical approach using the six-step process: *Ask, Prepare, Process, Analyze, Share, and Act.*

Objective

The goal of this case study is to help **Bellabeat**, a wellness technology company, **gain uncover trends and insights** into consumer behavior through smart device usage data. By understanding how consumers interact with wellness technology, Bellabeat aims to better position its products and tailor its campaigns to boost user engagement and growth. The analysis focuses on identifying behavioral patterns from existing smart device data and translating those findings into **actionable recommendations**.

Urška Sršen, Bellabeat's co-founder and Chief Creative Officer, has tasked the marketing analytics team with leveraging public data to explore user habits around activity, sleep, and wellness. The ultimate aim is to apply these insights to one of Bellabeat's own products—such as the Leaf, Time, or Spring—and propose strategies to **increase brand impact** and **customer adoption**.

About Bellabeat

Bellabeat is a wellness technology company that creates smart, health-focused products for women. Their product lineup includes the **Leaf** wellness tracker, the **Time** smartwatch, and the **Spring** smart water bottle—all of which sync with the **Bellabeat app** to track activity, sleep, hydration, and mindfulness. Bellabeat also offers a membership program with personalized health guidance. The brand combines elegant design with technology to help women make informed lifestyle choices.

Although Bellabeat has experienced fast growth and international expansion, the company is now seeking deeper insights from data to stay competitive in the growing market for smart wellness devices. By shifting its focus toward data-driven marketing, Bellabeat hopes to increase its user base and strengthen its presence in the global wellness tech space.

Step 1: Ask

These are the questions that will guide my analysis

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

Business tasks

- Identify user behavior and wellness trends in smart devices using FitBit Fitness Tracker Data
- Apply these insights to Bellabeat's products
- Recommend a data-driven marketing strategy to help Bellabeat grow

Step 2: Prepare

Data Source

- **Dataset**: FitBit Fitness Tracker Data

  FitBit Fitness Tracker Data
- **Publisher**: Mobius on Kaggle
- **License**: CC0 Public Domain

  CC0 Public Domain
- **Users**: 30 Fitbit users who voluntarily shared personal tracker data
- Format: Data is in long format, stored in multiple CSV files that spans from March 12, 2016 to May 12, 2016.

Step 3: Process

Tools Used

I'm using **RStudio** for this project because it provides a powerful and user-friendly environment for data cleaning, transformation, and visualization, making it ideal for analyzing large datasets in this case study.

Load Packages

You need to install these packages first before loading if you havn't install it before.

```
library(tidyverse)
library(ggplot2)
library(dplyr)
library(tidyr)
library(lubridate)
library(readr)
library(janitor)
```

Import Datasets and Standardize Column Names

I'm using datasets the contain user informations from April 12, 2016 to May 12, 2016 and standardizing column names to all lowercases.

```
activity <- read.csv("dailyActivity_merged.csv") %>% clean_names()
calories <- read.csv("dailyCalories_merged.csv") %>% clean_names()
steps <- read.csv("dailySteps_merged.csv") %>% clean_names()
sleep <- read.csv("sleepDay_merged.csv") %>% clean_names()
intensities <- read.csv("dailyIntensities_merged.csv") %>% clean_names()
```

View Datasets

```
glimpse(activity)
glimpse(calories)
glimpse(steps)
glimpse(sleep)
glimpse(intensities)
```

—> After viewing these datasets, I noticed that the Day/Date column in each datasets are stored as **Character** column, which it have not yet converted to **Date** type.

Clean Datasets

Fix formatting from text to date

```
activity$activity_date <- as.Date(activity$activity_date, format = "%m/%d/%Y")
calories$activity_day <- as.Date(calories$activity_day, format = "%m/%d/%Y")
steps$activity_day <- as.Date(steps$activity_day, format = "%m/%d/%Y")
sleep$sleep_day_dt <- strptime(sleep$sleep_day, format = "%m/%d/%Y %I:%M:%S %p")
sleep$sleep_day <- as.Date(sleep$sleep_day_dt)
intensities$activity_day <- as.Date(intensities$activity_day, format = "%m/%d/%Y")
```

Remove duplicate rows keeps only unique rows

```
activity <- activity %>% distinct()
calories <- calories %>% distinct()
steps <- steps %>% distinct()
sleep <- sleep %>% distinct()
intensities <- intensities %>% distinct()
```

Check if there is missing values

```
check_missing <- function(df) {
  colSums(is.na(df)) }

check_missing(activity)
```

```
##                          id              activity_date
##                           0                          0
##                 total_steps             total_distance
##                           0                          0
##            tracker_distance logged_activities_distance
##                           0                          0
##        very_active_distance moderately_active_distance
##                           0                          0
##       light_active_distance   sedentary_active_distance
##                           0                          0
##         very_active_minutes       fairly_active_minutes
##                           0                          0
##      lightly_active_minutes           sedentary_minutes
##                           0                          0
##                    calories
##                           0
```

```
check_missing(calories)
```

```
##           id activity_day     calories
##            0            0            0
```

```
check_missing(steps)
```

```
##           id activity_day   step_total
##            0            0            0
```

```
check_missing(sleep)
```

```
##                   id            sleep_day  total_sleep_records
##                    0                    0                    0
## total_minutes_asleep    total_time_in_bed         sleep_day_dt
##                    0                    0                    0
```

```r
check_missing(intensities)
```

```
##                          id               activity_day
##                           0                          0
##           sedentary_minutes     lightly_active_minutes
##                           0                          0
##       fairly_active_minutes        very_active_minutes
##                           0                          0
##    sedentary_active_distance      light_active_distance
##                           0                          0
## moderately_active_distance        very_active_distance
##                           0                          0
```

—> Based on the result, there's **no missing values** for all datasets.

Step 4: Analyze

Explore datasets

Check how many unique participants are these in each datasets

```r
n_distinct(activity$id)
```

```
## [1] 33
```

```r
n_distinct(calories$id)
```

```
## [1] 33
```

```r
n_distinct(steps$id)
```

```
## [1] 33
```

```r
n_distinct(sleep$id)
```

```
## [1] 24
```

```r
n_distinct(intensities$id)
```

```
## [1] 33
```

—> There's 33 unique users in every datasets except the sleep dataset has 24 unique users

Quick summary statistics on specific columns

```r
activity %>%
    select(total_steps,
         total_distance,
         sedentary_minutes) %>%
    summary()
```

```
##   total_steps    total_distance   sedentary_minutes
##  Min.   :    0   Min.   : 0.000   Min.   :   0.0
##  1st Qu.: 3790   1st Qu.: 2.620   1st Qu.: 729.8
##  Median : 7406   Median : 5.245   Median :1057.5
##  Mean   : 7638   Mean   : 5.490   Mean   : 991.2
##  3rd Qu.:10727   3rd Qu.: 7.713   3rd Qu.:1229.5
##  Max.   :36019   Max.   :28.030   Max.   :1440.0
```

```r
sleep %>%
    select(total_sleep_records,
```

```
                 total_minutes_asleep,
                 total_time_in_bed) %>%
    summary()
```

```
##  total_sleep_records total_minutes_asleep total_time_in_bed
##  Min.   :1.00        Min.   : 58.0        Min.   : 61.0    
##  1st Qu.:1.00        1st Qu.:361.0        1st Qu.:403.8    
##  Median :1.00        Median :432.5        Median :463.0    
##  Mean   :1.12        Mean   :419.2        Mean   :458.5    
##  3rd Qu.:1.00        3rd Qu.:490.0        3rd Qu.:526.0    
##  Max.   :3.00        Max.   :796.0        Max.   :961.0    
```

—> From this summary, we can see that the average sedentary is 991.2 minutes, which is about 16.5 hours. The average steps per day is 7638, and the average minutes asleep is 419.2 (about 7 hours). Therefore, we should keep in mind that we should **increase users' daily steps and sleep time**.

Adding new column in sleep dataset

```
sleep <- sleep %>%
    mutate(difference = total_time_in_bed - total_minutes_asleep)
```

Rename columns for merging datasets

```
steps <- steps %>%
  rename(
    date = activity_day
  )

intensities <- intensities %>%
  rename (
    date = activity_day
  )

sleep <- sleep %>%
  rename(
    date = sleep_day
  )
```

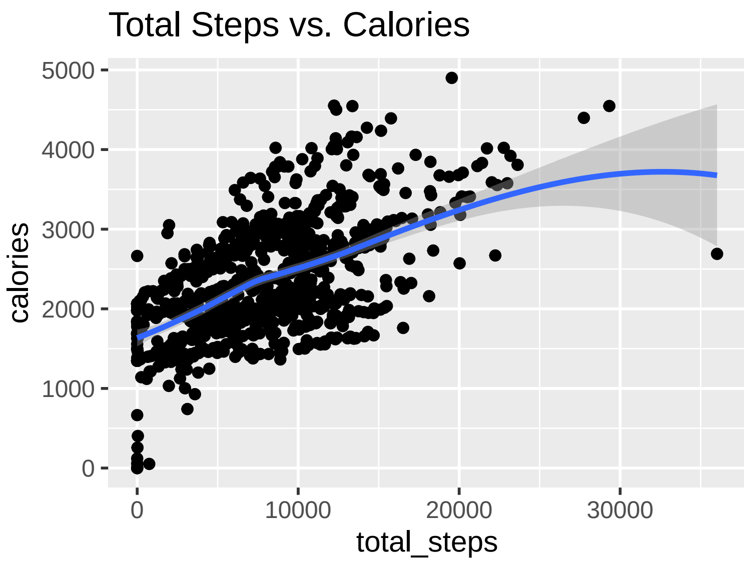
Merge datasets

```
sleep_steps <- merge(sleep, steps, by=c("id", "date"))
intensities_sleep <- merge(intensities, sleep, by=c("id", "date"))
```

Analyze datasets

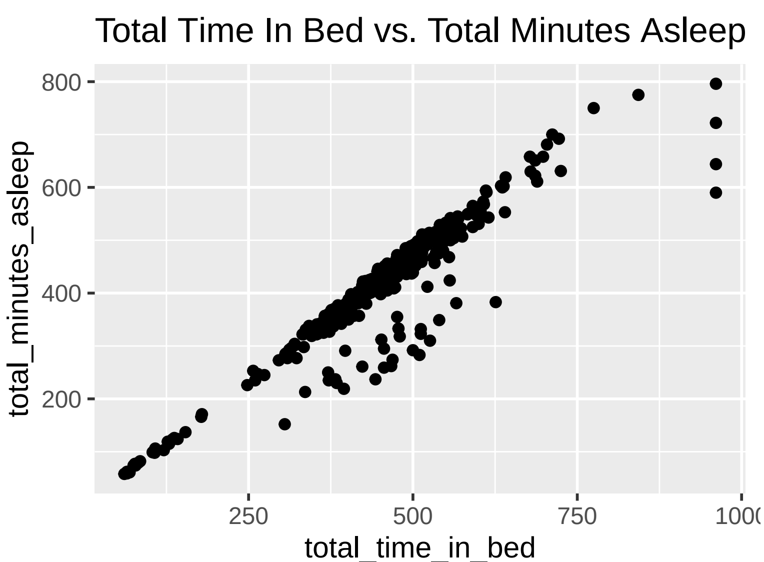
Total Steps vs. Calories

```
ggplot(data=activity, aes(x=total_steps, y=calories)) +
  geom_point() + geom_smooth() + labs(title="Total Steps vs. Calories")
```



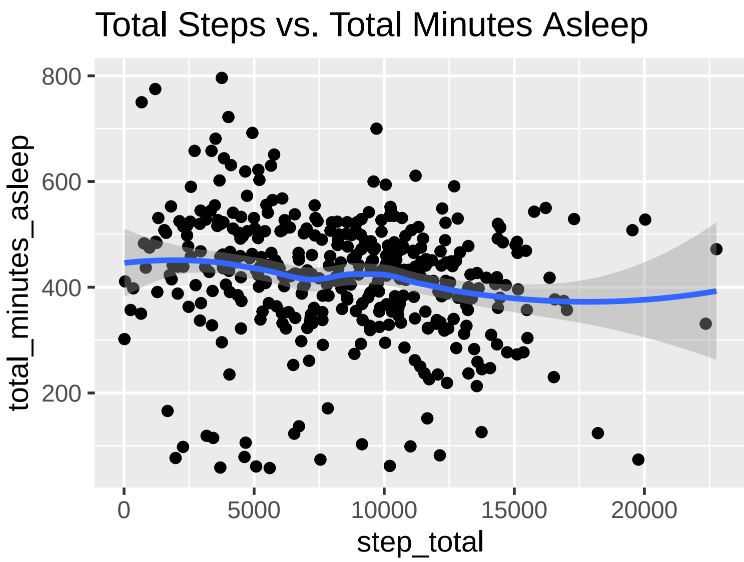
Total Time In Bed vs. Total Minutes Asleep

```
ggplot(data=sleep, aes(x=total_time_in_bed, y=total_minutes_asleep)) +
  geom_point() + labs(title="Total Time In Bed vs. Total Minutes Asleep")
```



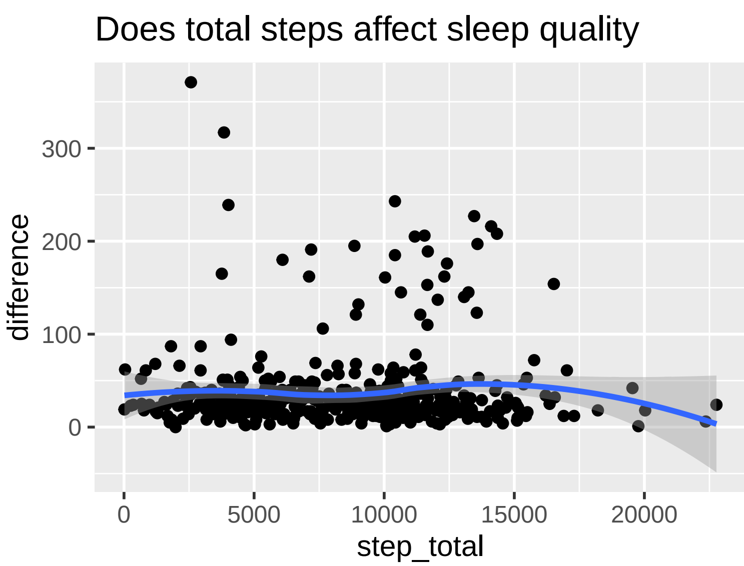
Total Steps vs. Total Minutes Asleep

```
ggplot(data=sleep_steps, aes(x=step_total, y=total_minutes_asleep)) +
  geom_point() + geom_smooth() + labs(title="Total Steps vs. Total Minutes Asleep")
```



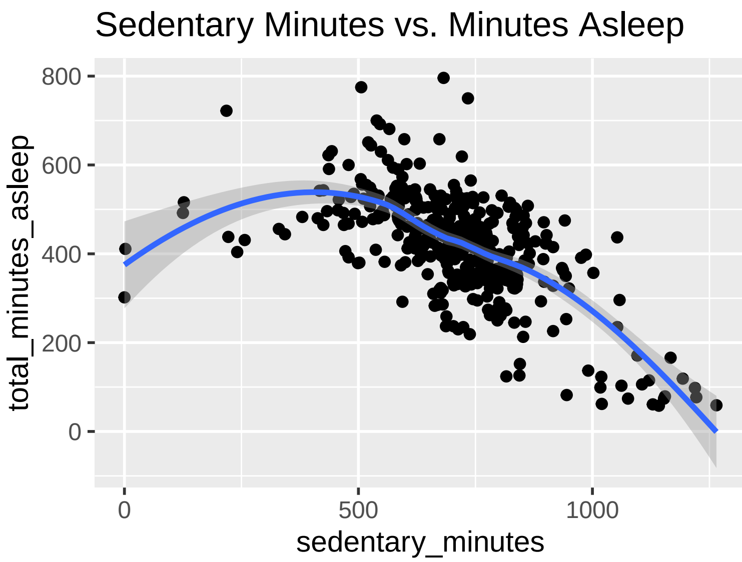
Does total steps affect sleep quality

```
ggplot(data=sleep_steps, aes(x=step_total, y=difference)) +
  geom_point() + geom_smooth() + labs(title="Does total steps affect sleep quality")
```



Sedentary Minutes vs. Minutes Asleep

```
ggplot(data=intensities_sleep, aes(x=sedentary_minutes, y=total_minutes_asleep)) +
  geom_point() + geom_smooth() + labs(title="Sedentary Minutes vs. Minutes Asleep")
```



Step 5: Share

Total Steps vs. Calories —

We can see that there's a **positive correlation** between total steps and calories burned. Which means the more steps you take, the more calories will be burned.

Total Time In Bed vs. Total Minutes Asleep —

There's a **strong positive linear relationship** between time in bed and minutes asleep, suggesting that the more time someone spends in bed, the more sleep they generally get. However, there's still many points below the linear line, indicating people often spend more time in bed than actually sleeping.

Total Steps vs. Total Minutes Asleep —

The trend line is relatively flat, implying that there's **no significant linear relationship** between step count and sleep duration. This suggests that how many steps a person takes doesn't have a clear impact on how long they sleep.

Total Steps vs. Sleep Quality —

Y-axis: difference = total time in bed - total minutes asleep

This plot investigates if being more active improves sleep efficiency. From this graph we can see the **relationship is weak** and the overall data is scattered, indicating that there's **no strong direct correlation**.

Sedentary Minutes vs. Minutes Asleep —

There's a slightly inverse relationship between sedentary minutes and minutes asleep. Sleep duration is **decreasing while the sedentary minutes is between 500 and 1000**. This suggests that in order to **increase sleep duration**, users must **decrease sedentary time**.

Step 6: Act

Recommendation 1: Encourage Active Breaks to Reduce Sedentary Minutes

Many users in the dataset spend a large portion of their day being sedentary, which can negatively affect overall wellness. Bellabeat can implement **in-app push notifications** reminding users to **take short, active breaks** throughout the day—such as standing up, stretching, or walking for 2–5 minutes every hour. These can be personalized based on users' typical activity patterns and may improve daily energy levels while aligning with Bellabeat's wellness mission.

Recommendation 2: Promote Healthy Sleep Habits by Encouraging Earlier Bedtimes

Analysis of sleep data shows that some users fall short of the recommended 8 hours of sleep. Bellabeat can create a feature in the app that **reminds users to start preparing for bed** at a consistent time—especially if late sleep patterns are detected. This could include mindfulness audio, hydration tracking, and dim-light mode to reduce screen stimulation. Emphasizing sleep quality supports users' health and deepens engagement with the Bellabeat ecosystem.

Recommendation 3: Motivate Users to Increase Daily Step Counts

Users with higher step counts consistently burn more calories and show more balanced activity profiles. Bellabeat can introduce a **daily step challenge** that motivates users to gradually increase their step goal based on their historical averages. Friendly competition, badges, and milestone rewards can create a sense of progress and make Bellabeat's products not just tracking tools but motivational partners.

> Thank you for taking the time to review this case study. Working on the Bellabeat project has been an insightful experience that allowed me to apply data analytics skills to a real-world business challenge. Through this analysis, I was able to explore user behavior, uncover actionable trends, and develop marketing recommendations that align with Bellabeat's mission of empowering women's wellness. I look forward to applying these skills to future data-driven projects and continuing to grow as an analyst.