

Case Study: Bellabeat Analysis

Author: Lillian Pineda

Date: 2023-08-07

Email: lpineda2191@gmail.com

Kaggle: <https://www.kaggle.com/datasets/lillianpineda/bellabeat-capstone-project>

LinkedIn: <https://www.linkedin.com/in/lillianpineda2191/>

Introduction:

In this particular research project, I examined data from Bellabeat, a San Francisco-based company specializing in smart devices. My focus was on addressing essential business inquiries and generating practical and valuable observations from the data.

1. Ask:

How can we better market our smart devices based on user trends?

What are some trends in smart device usage?

How could these trends apply to Bellabeat customers?

2. Prepare:

This data set does have its limitations. It was collected from 2016-03-12 to 2016-05-12 and had a participant base of 30. I only looked at two of the tables from this dataset, dailyActivity_merged and sleepDay_Merged.

3. Process:

Daily activities_merged data:

- Fixed structural errors such as typos, capitalization, incorrect spelling
- Eliminated duplicates
- Changed decimals to whole numbers
- Eliminated tracked distance column
- Renamed total distance column to total tracked distance
- Eliminated logged activities as most of the values were and wouldn't have a significant impact on data
- Added days of the week column (titled ActivityDay)

HourlySteps and HourlyCaloriesMerged:

- Split ActivityDate column into 2 columns; ActivityDate and ActivityTime
- Added AM/PM column

SleepDayMerged:

- Added column TotalMinutesAsleep in Hours

Made All 4 Google sheets in CSV files

4. Analyze:

I uploaded the datasets to BigQuery under the project name “first-sandbox-project5723.Bellabeat.daily_activity”

My objective was to determine the frequency of tracker usage for each individual user.

```
SELECT
  Id, COUNT (Id) AS Total_Id
FROM
  `first-sandbox-project5723.Bellabeat.daily_activity`
GROUP BY
  Id
ORDER BY
  Total_Id DESC;
```

The participants wore their trackers for varying durations, ranging from 4 to 31 days, with a significant number of individuals leaning towards longer periods.

Following that, I aimed to categorize users into subgroups based on their frequency of tracker usage:

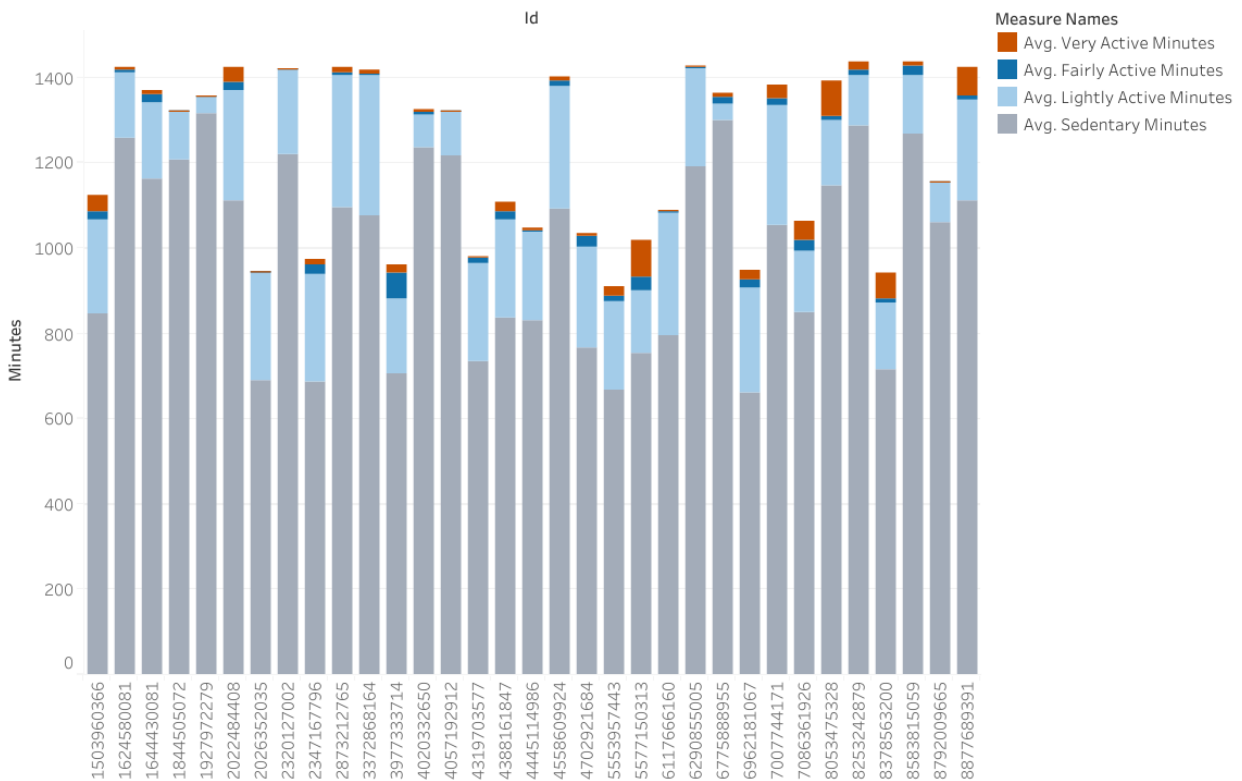
- Light Users: 0-14 days
- Moderate Users: 15-24 days
- Active Users: 25-31 days

```
(SELECT
  Id,
  COUNT(Id) AS Total_Logged_Uses,
CASE
  WHEN COUNT (Id) BETWEEN 25 AND 31 THEN 'Active User'
  WHEN COUNT (Id) BETWEEN 15 AND 24 THEN 'Moderate User'
  WHEN COUNT (Id) BETWEEN 0 AND 14 THEN 'Light User'
END as Usage_Type
FROM
  `first-sandbox-project5723.Bellabeat.daily_activity`
GROUP BY
  Id
ORDER BY
  Total_Logged_Uses)
```

Subsequently, my intention was to focus solely on calculating the average minutes for each category of users based on their respective IDs.

```
SELECT
  Id,
  avg(VeryActiveMinutes) AS Avg_Very_Active_Minutes,
  avg(FairlyActiveMinutes) AS Avg_Fairly_Active_Minutes,
  avg(LightlyActiveMinutes) AS Avg_Lightly_Active_Minutes,
  avg(SedentaryMinutes) AS Avg_Sedentary_Minutes,
FROM
  `first-sandbox-project5723.Bellabeat.daily_activity`
GROUP BY
  Id;
```

Average Activity Levels

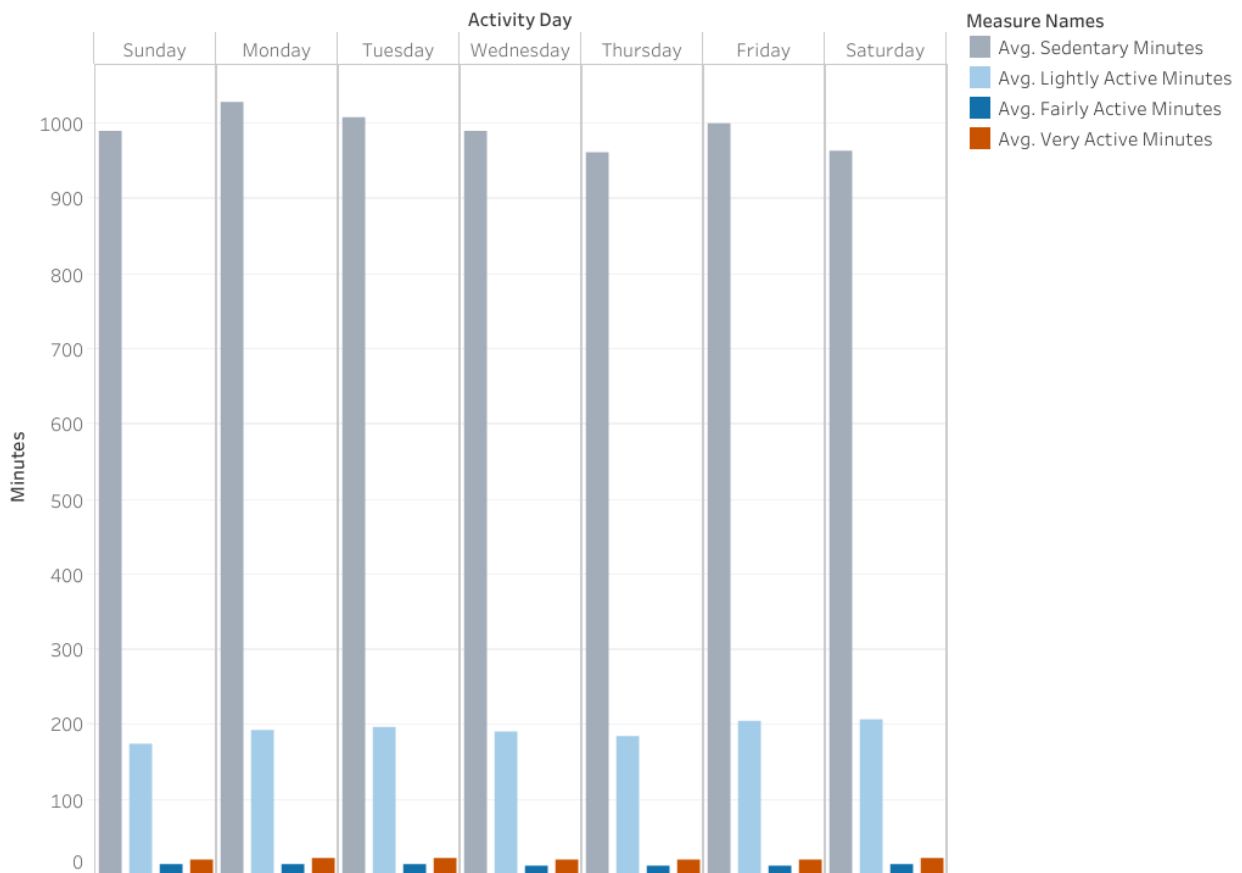


The analysis revealed that a significant number of users predominantly remained at the sedentary activity level for the majority of the time.

To investigate further, I sought to ascertain whether there was a pattern of increased activity on specific days of the week among the users.

```
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  
  `first-sandbox-project5723.Bellabeat.daily_activity`  
GROUP BY  
  ActivityDay;
```

Daily Average Activity Levels



The examination of activity levels on different days did not indicate any significant variations.

In order to gain more insights, I decided to conduct a more thorough investigation into the correlation between the number of steps taken and the calories burned by the users.

```
SELECT
  Id,
  AVG(TotalSteps) AS Avg_Total_Steps,
  AVG(TotalTrackedDistance) AS Avg_Total_Distance,
  AVG(Calories) AS Avg_Calories_Burned
FROM
  `first-sandbox-project5723.Bellabeat.daily_activity`
GROUP BY
  Id;
```

Steps vs. Calories

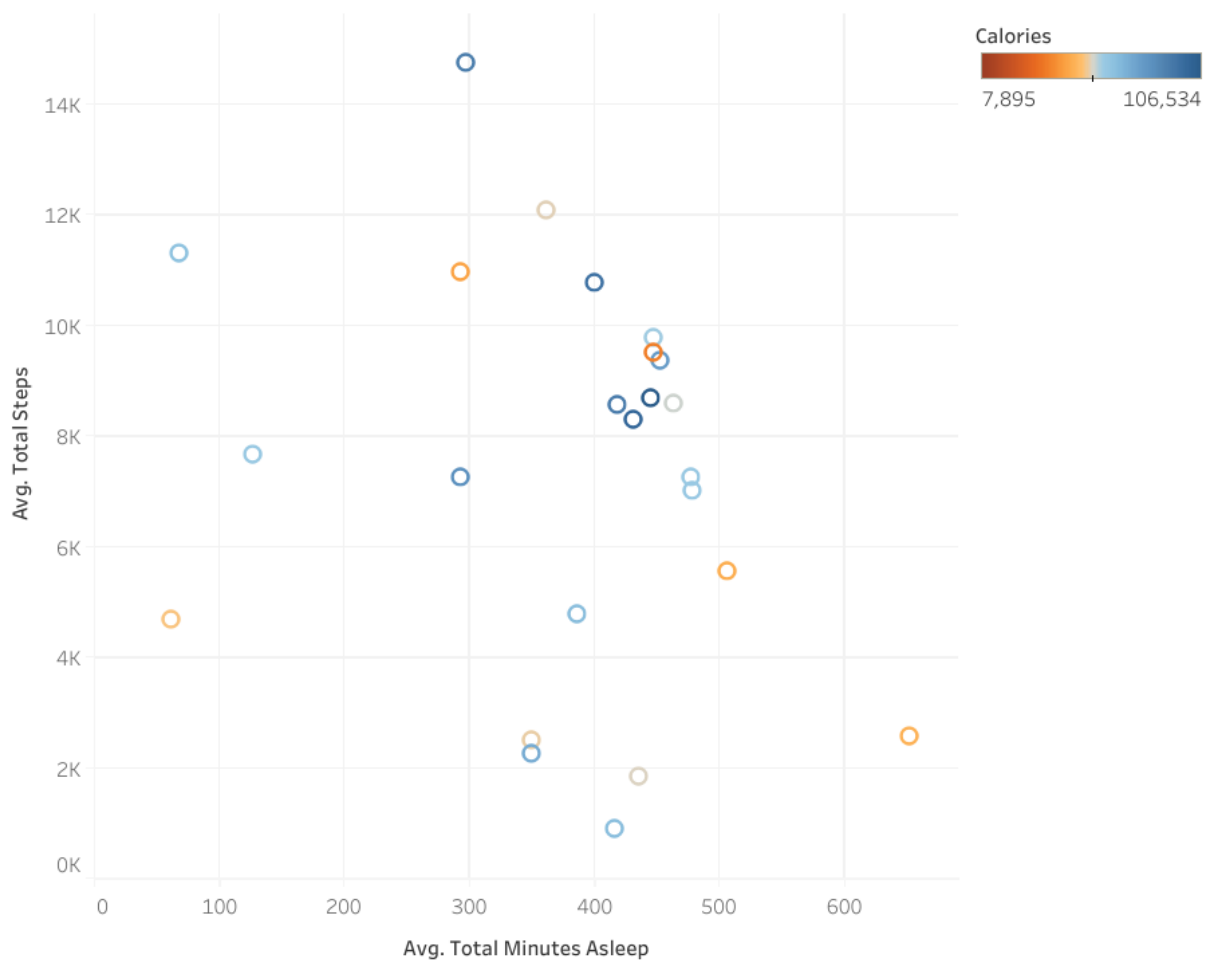


The analysis revealed a positive correlation between the total number of steps taken by the users and the calories burned.

Next, I aimed to examine how this relationship compared to the users' sleep patterns. To do this, I combined the activity and sleep tables to draw meaningful insights.

```
SELECT
    Activity.Id,
    AVG(Activity.TotalSteps) AS Avg_Total_Steps,
    AVG(Activity.Calories) AS Avg_Calories,
    AVG(Sleep.TotalMinutesAsleep) AS Avg_Total_Minutes_Asleep,
FROM
    `first-sandbox-project5723.Bellabeat.daily_activity` AS Activity
JOIN
    `first-sandbox-project5723.Bellabeat.daily_sleep` AS Sleep
ON
    Activity.Id = Sleep.Id
GROUP BY
    Activity.Id;
```

Time Asleep vs. Total Steps



The graphical representation indicates that out of all the users, only five individuals managed to exceed the recommended 10,000 daily steps. Interestingly, just two of these users were able to achieve an average of 5.5 hours or more of sleep per day.

5. Share:

Here is the link to my Tableau visualizations for this project. I invite you to take a look at the dashboard and explore the insights

[Lillian Pineda - Tableau Dashboard](#)

6. Act:

Data Trends:

- The majority of users predominantly remained in the sedentary activity category, with relatively limited time spent in the light, moderate, and active categories.
- Users, on average, slept for less than 8 hours per night.
- Participant consistency in tracking sleep was low, with only 24 individuals recording their sleep data.

Recommendations:

- Bellabeat could implement incentives to encourage users to strive for the recommended 10,000 daily steps. This might encompass earning badges or accumulating points that can be redeemed for merchandise or other rewards.
- Introducing a feature that sends reminders to users when they have been in the sedentary category for an extended period could motivate them to become more active and break prolonged periods of inactivity.