Bellabeat Case Study

Author: Samantha Houston

Date: 10-1-2024

LinkedIn

Ask

Bellabeat stakeholders have requested the analysis of smart device usage data in order to gain insight into how consumers use non-Bellabeat smart devices. This will help the company identify new growth opportunities in their target audience.

About Company

Urška Sršen and Sando Mur founded Bellabeat, a high-tech company that manufactures health-focused smart products. Sršen used her background as an artist to develop beautifully designed technology that informs and inspires women around the world. Collecting data on activity, sleep, stress, and reproductive health has allowed Bellabeat to empower women with knowledge about their own health and habits. Since it was founded in 2013, Bellabeat has grown rapidly and quickly positioned itself as a tech-driven wellness company for women.

Stakeholders

- Urška Sršen
- Sando Mur
- Bellabeat marketing analytics team

Questions

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

Prepare

The data source of this project was the Fitbit Fitness Tracker dataset that is linked below.

Fitbit Fitness Tracker

During preparation of this dataset the files had to be downloaded and removed from the zip file in which they were stored. The dataset included files for data dating from 3.12.16 to 4.11.16 and 4.12.16 to 5.12.16. For the purposes of this analysis data was used from 3.12.16 to 4.12.16.

Process

For analysis the following dataset were used

- dailyActivity_merged
- hourlySteps_merged
- hourlyCalories_merged

The data cleaning process was conducted with Google Sheets.

1. **Sorting and filtering:** Since some of the sheets did include dates outside of the range I was analyzing I filtered the dates to only show ones relevant to my analysis.
2. **Removing duplicates:** To ensure that my data did not include any duplicates I ran a data cleanup.
3. **Time and Date Formatting:** In order for my sheets to be uploaded to BigQuery later for SQL I had to format all date and time columns in my data that had either a date or time that was not relevant to the analysis. The time was removed from the hourlySteps sheet.
4. **Rounding:** In the dailyActivity sheet the "TotalDistance" and "TrackerDistance" were rounded.
5. **Removing and Adding Columns:** In the dailyActivity sheet the column "LoggedActivity" was removed and a column to reflect "ActivityDay" was added.

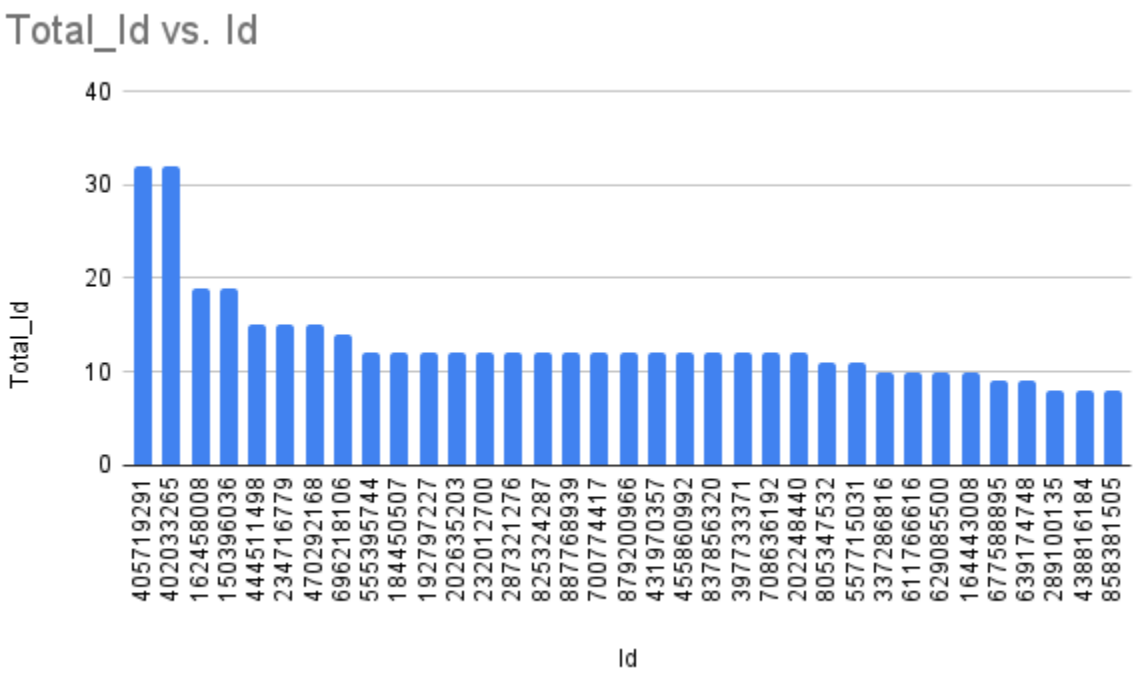
Analyze

During my analysis I uploaded my data sheets to BigQuery from Google Drive. I made different visualizations with my results using Google Sheets and Tableau Public.

I started my analysis off by running a query Total_Id

```
SELECT Id, COUNT(Id) AS Total_Id
```

```
FROM `speedy-method-429216-n4.Wellness_Data.dailyactivity`

GROUP BY Id

ORDER BY Total_Id DESC;
```

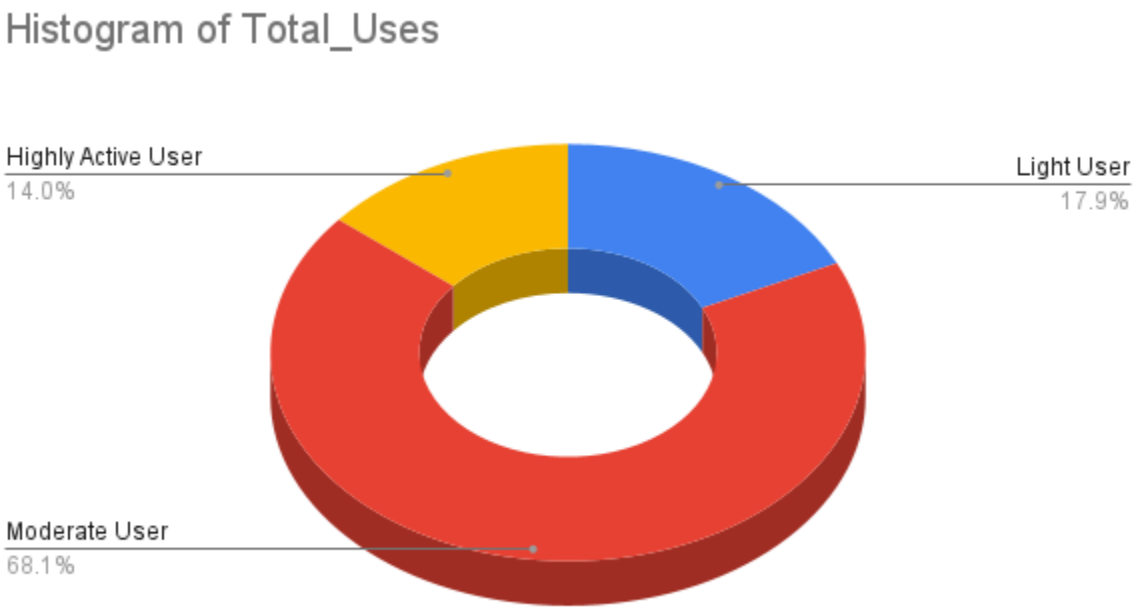


This query indicates that the majority of the users Id fall within the same range when totaled through the amount of times that specific Id shows in the dataset.

Then I wanted to label the users by the level of activity.

```
SELECT Id, COUNT(Id) AS TotalActivity,

CASE

 WHEN COUNT(Id) BETWEEN 0 AND 10 THEN 'Light User'

 WHEN COUNT(Id) BETWEEN 11 AND 22 THEN 'Moderate User'

 WHEN  COUNT(Id) BETWEEN 23 AND 32 THEN 'Highly Active User'
```

```
END as usage_type

FROM `speedy-method-429216-n4.Wellness_Data.dailyactivity`

GROUP BY Id

ORDER BY TotalActivity;
```



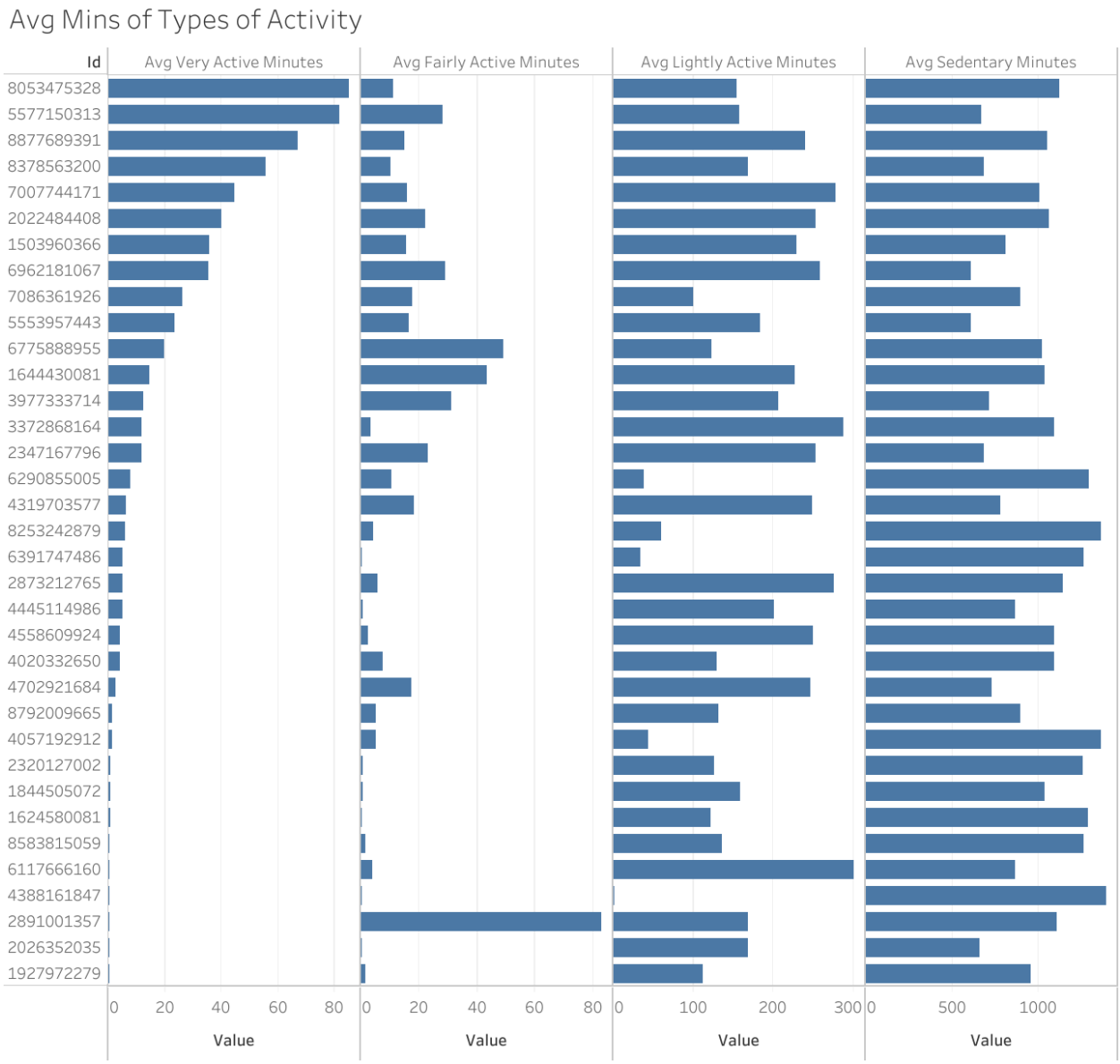
These results show that based on the queries ran so far and the different types of users that majority of the users are considered ‘Moderate’.

In my next query I wanted to identify the types of average active minutes.

```
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,
```

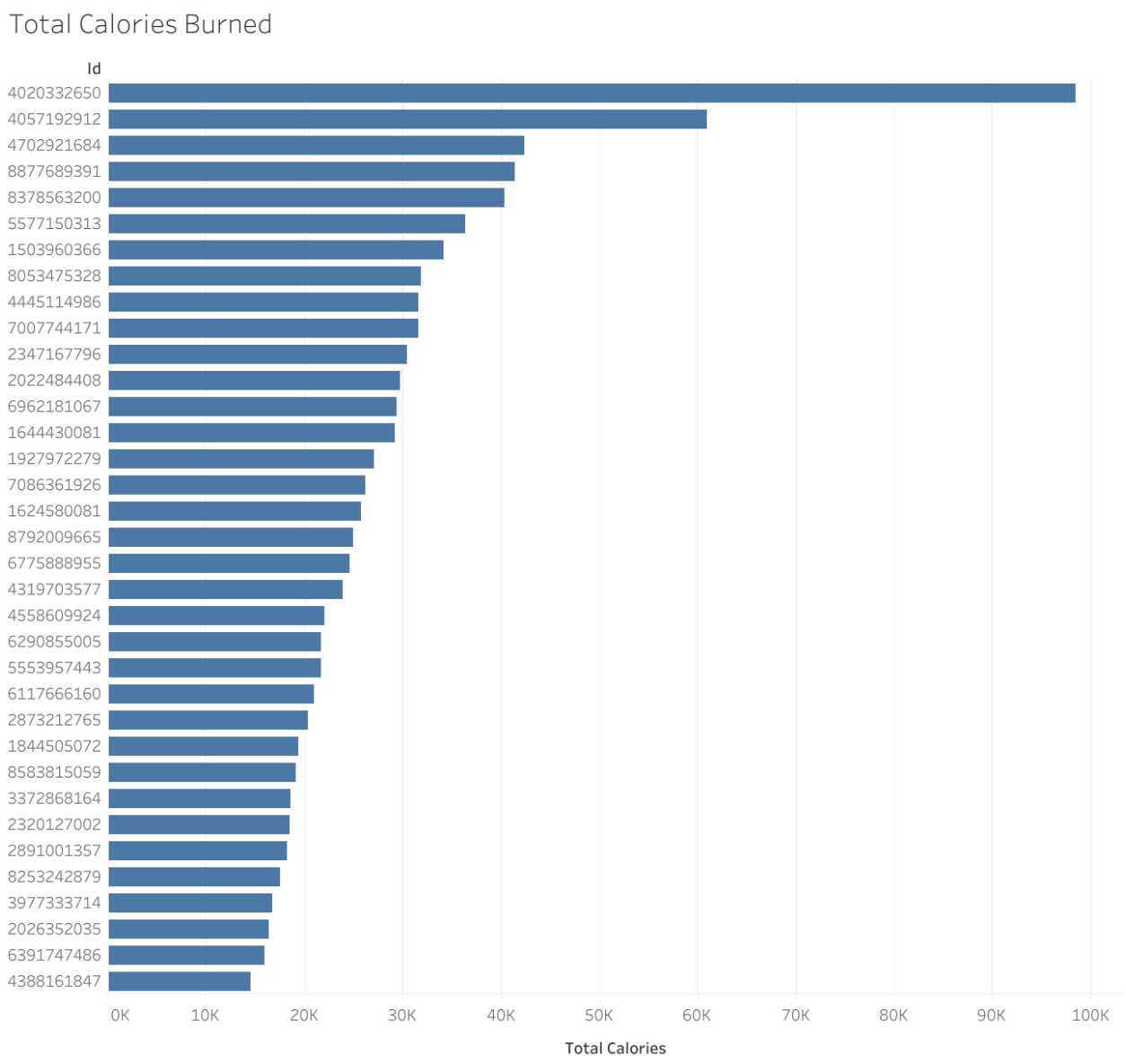
```
FROM `speedy-method-429216-n4.Wellness_Data.dailyactivity`

GROUP BY Id;
```

I then uploaded the report to Google Drive, then uploaded it to Tableau Public to create a visualization.



Now that I have the types of activities label I wanted to know the total calories each user had burned and order it by Id.

```
SELECT Id, SUM(Calories) AS TotalCalories

FROM `speedy-method-429216-n4.Wellness_Data.dailyactivity`

GROUP BY Id;
```

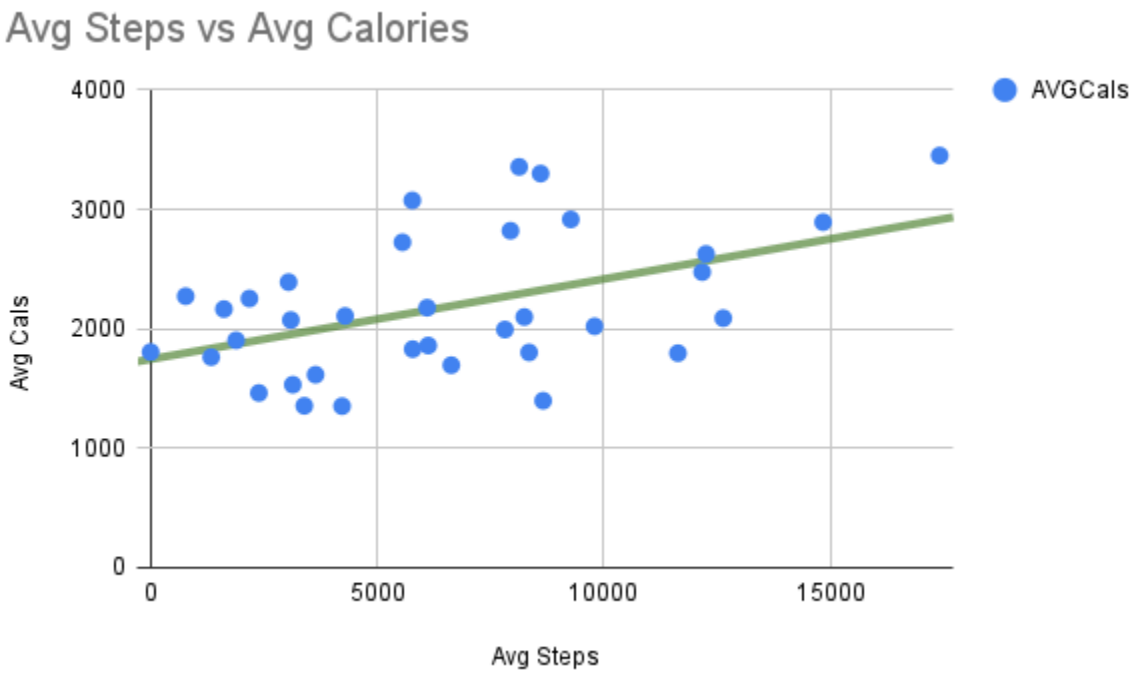


Through this visualization we can see that while there are some users that have burned a substantial amount of calories in comparison to

their counterparts, the majority of the users have similar amounts of total calories burned.

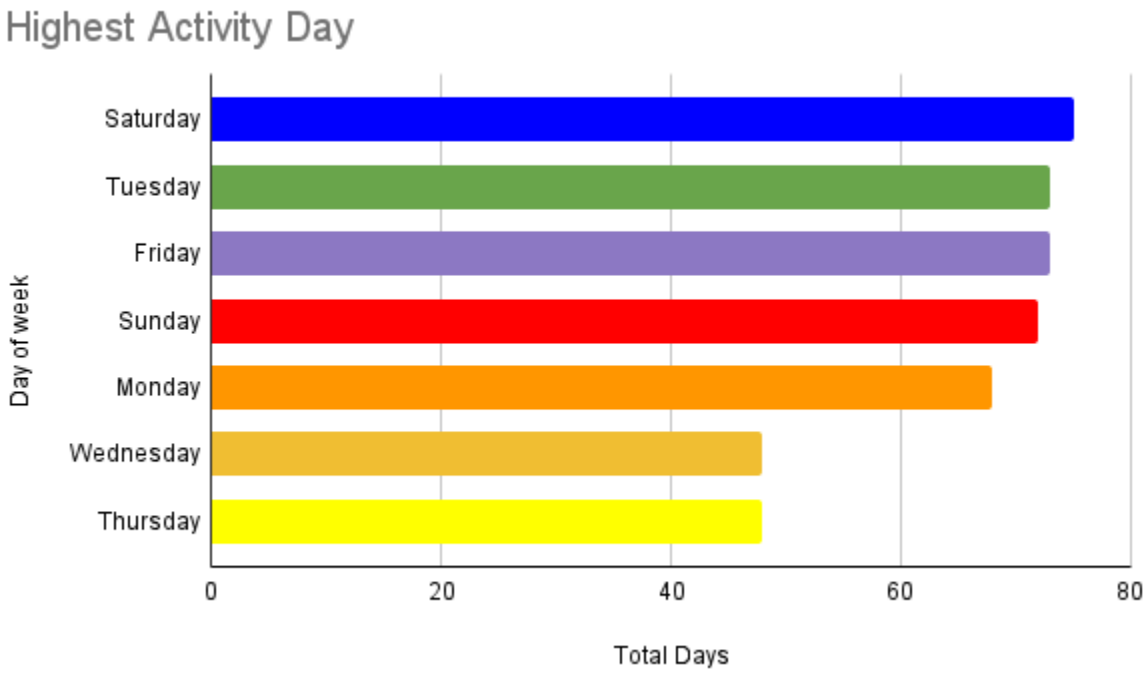
I then took the average of the total calories and total steps of each user

```
SELECT Id, AVG(Calories) AS AVGCals, AVG(TotalSteps) AS AVGSteps

FROM `speedy-method-429216-n4.Wellness_Data.dailyactivity`

GROUP BY Id;
```



For my final analysis of the data I wanted to know which day of the week had the highest activity.

```
SELECT ActivityDay, COUNT(ActivityDay) AS Total_Day

FROM `speedy-method-429216-n4.Wellness_Data.dailyactivity`

GROUP BY ActivityDay

ORDER BY Total_Day DESC;
```



The findings of this query show that the most activity happens on Saturdays for users.

Share

Findings

- The highest days of activity for users were Saturday
- The higher amount of steps users took burned more calories
- Majority of users are moderate users at 68.1%

Act

1. **Incentives for usage:** The more frequently users are active and data is logged offer coupons, rewards, small gifts.
2. **Highlight goals and health:** Offer users more information about setting activity goals and the impact higher activity can have on their health.

3. **Reminders:** Give users activity reminders such as steps, activity levels, and calories burned to reach daily targets.
4. **Increase activity on weekdays:** Encourage users to find time during the week to stay active and increase activity with family walks, walking the neighborhood, group fitness activities.