

# **Bellabeat Case Study:-**

## **How Can a Wellness Technology Company Play It Smart**

### About the 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.

Bellabeat has a lot of products and services but for this case study we will be focusing on the Bellabeat app, which acts as your personal health hub. By tracking activity, sleep, stress, menstrual cycles, and mindfulness habits, the app empowers you to understand your current practices and make informed decisions to optimize your overall well-being.

This study will pass through the six phases of the data life cycle which are;

- Ask
- Prepare
- Process
- Analyze
- Share
- Act

## **ASK**

### Business task-

Sršen asked me to analyze smart device usage data in order to gain insight into how consumers use non-Bellabeat smart devices. She then wants suggestions and recommendations for their marketing analytics team.

## Key stakeholders-

- Urška Sršen: Bellabeat's co founder and Chief Creative Officer
- Sando Mur: Mathematician and Bellabeat's cofounder; key member of the Bellabeat executive team
- Bellabeat marketing analytics team: A team of data analysts responsible for collecting, analyzing, and reporting data that helps guide Bellabeat's marketing strategy.

## Prepare

The analysis I am conducting utilizes datasets retrieved from Mobius on Kaggle, licensed under the CCO Public Domain. kaggle.com indicates these datasets were collected via a distributed Amazon Mechanical Turk survey between December 3rd, 2016 and December 5th, 2016, with 30 Fitbit users reportedly consenting to share personal tracker data.

Upon initial examination of the 18 extracted CSV files (out of the downloaded 10), I identified several limitations. The data format stores information for each ID across multiple rows (long format). Additionally, data integrity concerns arose. While Kaggle reported 30 consenting Fitbit users, further investigation revealed 35 users over 32 days. Furthermore, the 2016 collection date makes the data outdated for your current needs. The timeframe discrepancy between the stated collection period (December 3rd-5th, 2016) and the observed 31 days also raises questions.

The limited sample size introduces bias, as a larger group would better represent the target population and increase confidence levels. The lack of demographic information (gender, age, location) further hinders generalizability. Since Bellabeat products primarily target women, including demographic details would have significantly enhanced the analysis and subsequent recommendations.

## Process

For this analysis, I will be using the following datasets,

1. dailyactivity\_merged
2. hourlycalories\_merged
3. hourlyintensities\_merged
4. hourlysteps\_merged
5. minutesleep\_merged
6. heartrateseconds\_merged
7. minutesteps\_merged

- First I opened each dataset in Excel to get a feel of the dataset and see what I will be working with.
- Then, I sorted and filtered the data to focus on the relevant information. Next, I tackled data quality.
- I checked for and removed any blank entries, ensuring all data points were populated. Then, I hunted down duplicate values.
- In the 'minutesleep\_merged' dataset, I removed 543 duplicates, leaving 187,978 unique values.
- I also standardized the format of the date and time columns. Dates were converted to a clear month/day/year format (MM/DD/YYYY), while times were formatted to display hours, minutes, and seconds (hh:mm:ss).
- Additionally, I separated the merged date/time column using an integer function (to likely extract individual date and time components). Finally, I had to be ruthless and delete any rows containing data that didn't conform to the established formats. After this thorough cleaning process in Excel, I transitioned the data to SQL for further analysis.

## Analyze

- I started with uploading the files to BigQuery. I uploaded all 7 datasets mentioned above.
- I used Looker Studio to make visualizations using the tables created from SQL queries as it was easy to transport the temporary tables from BigQuery to Looker studios.

## Steps performed in SQL

### 1. To count the number of unique ID in each table

```
“
SELECT COUNT(DISTINCT id) AS unique_id
FROM `harshit-401709.fitbit_data.daily_activity`
“
```

>>Output confirm that there are data for 35 unique users in the daily activity table.

### 2. To know how often each ID was logged in

```
“
SELECT id,
COUNT(id) AS total_id
FROM `harshit-401709.fitbit_data.daily_activity`
GROUP BY id
Order By total_id
“
```

>>The query showed us that activity from users logged in between 8 to 32 times.

### 3. To classify their activities into three classes:

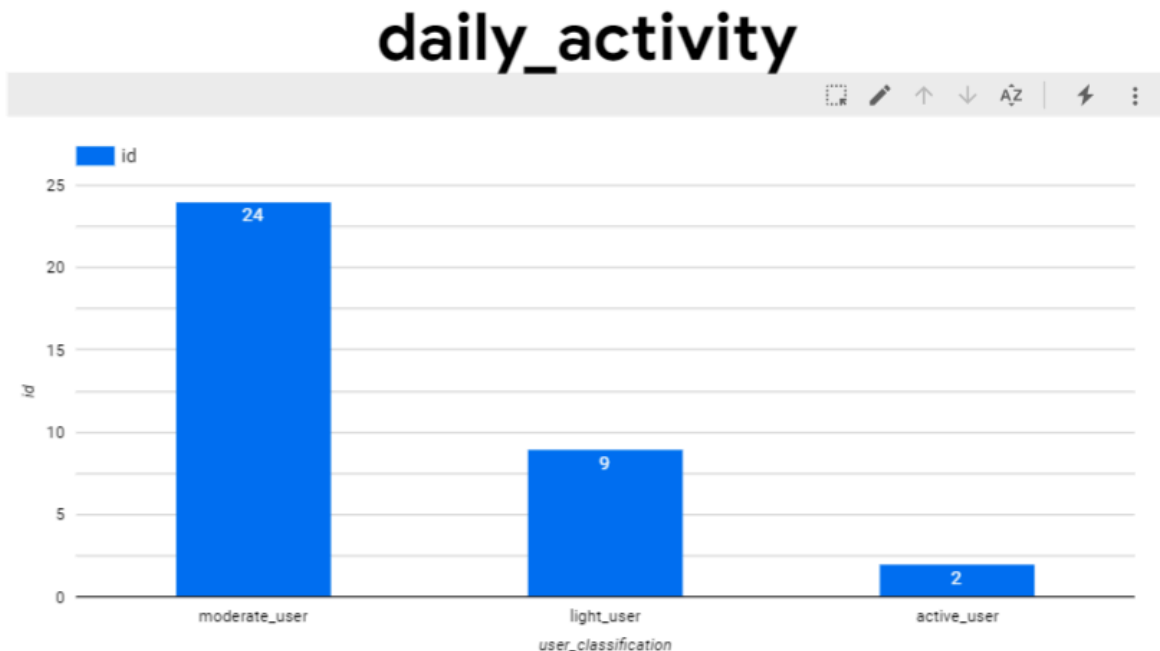
```
“
SELECT id,
COUNT(id) AS total_uses,
CASE
WHEN COUNT(Id) BETWEEN 21 AND 35 THEN 'active_user'
WHEN COUNT(Id) BETWEEN 11 and 20 THEN 'moderate_user'
WHEN COUNT(Id) BETWEEN 0 and 10 THEN 'light_user'
END AS user_classification
FROM `harshit-401709.fitbit_data.daily_activity`
GROUP BY id
“
```

>> The resulting table showed the classification as below:

Active users-2

Moderate users-24

Light users-9



### 4. To find out Avg, Min & Max of total steps, calories, total distance, and other activity levels.

```
“
SELECT Id,
ROUND(MIN(TotalSteps),2) AS min_total_steps,
ROUND(MAX(TotalSteps),2) AS max_total_steps,
ROUND(AVG(TotalSteps),2) AS avg_total_steps,
```

```

ROUND(MIN(TotalDistance),2) AS min_total_dist,
ROUND(MAX(TotalDistance),2) AS max_total_dist,
ROUND(AVG(TotalDistance),2) AS avg_total_dist,
ROUND(MIN(Calories),2) AS min_calories,
ROUND(MAX(Calories),2) AS max_calories,
ROUND(AVG(Calories),2) AS avg_calories,
ROUND(MIN(VeryActiveMinutes),2) AS min_very_active_minutes,
ROUND(MAX(VeryActiveMinutes),2) AS max_very_active_minutes,
ROUND(AVG(VeryActiveMinutes),2) AS avg_very_active_minutes,
ROUND(MIN(FairlyActiveMinutes),2) AS min_fairly_active_minutes,
ROUND(MAX(FairlyActiveMinutes),2) AS max_fairly_active_minutes,
ROUND(AVG(FairlyActiveMinutes),2) AS avg_fairly_active_minutes,
ROUND(MIN(LightlyActiveMinutes),2) AS min_lightly_active_minutes,
ROUND(MAX(LightlyActiveMinutes),2) AS max_lightly_active_minutes,
ROUND(AVG(LightlyActiveMinutes),2) AS avg_lightly_active_minutes,
ROUND(MIN(SedentaryMinutes),2) AS min_sedentary_minutes,
ROUND(MAX(SedentaryMinutes),2) AS max_sedentary_minutes,
ROUND(AVG(SedentaryMinutes),2) AS avg_sedentary_minutes
FROM `harshit-401709.fitbit_data.daily_activity`
GROUP BY id
ORDER BY id

```

“

>>Created table shows the data

## 5. To find out max of activity type

“

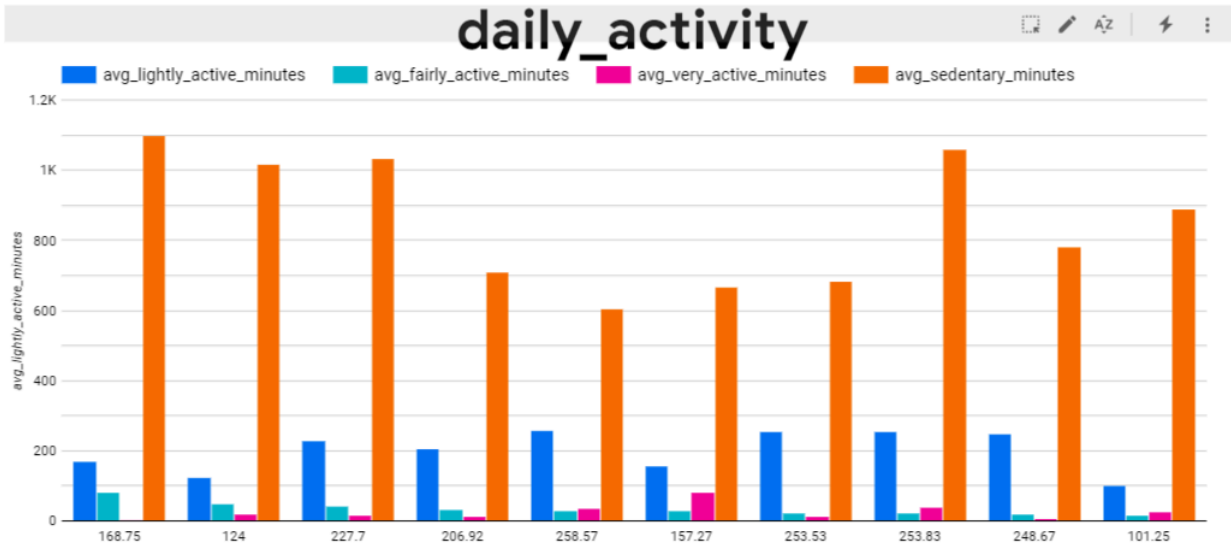
```

SELECT id,
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 `harshit-401709.fitbit_data.daily_activity`
GROUP BY Id
ORDER BY 2,3,4,5 DESC

```

“

>>The result showed that the sedentary minutes column was the highest on average.



### 6. To find out the no. of active and sedentary users.

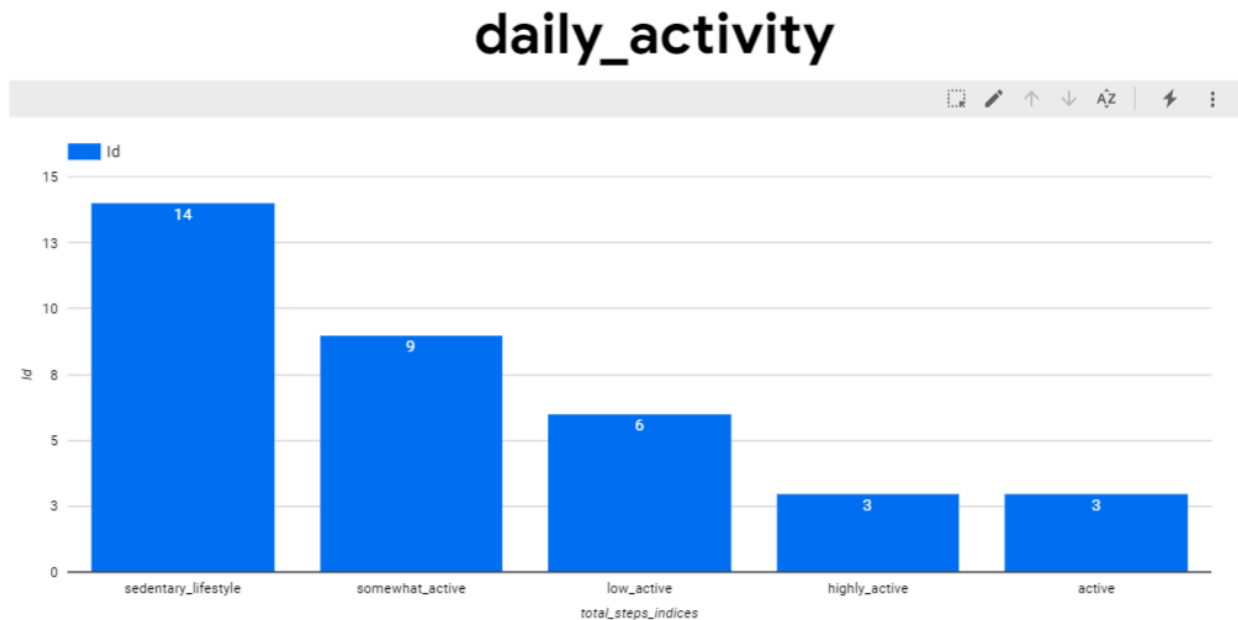
“

```
SELECT Id, ROUND(AVG(TotalSteps),2) AS avg_totalsteps,
CASE
WHEN ROUND(AVG(TotalSteps),2) < 5000 THEN 'sedentary_lifestyle'
WHEN ROUND(AVG(TotalSteps),2) BETWEEN 5000 and 7499 THEN 'low_active'
WHEN ROUND(AVG(TotalSteps),2) BETWEEN 7500 and 9999 THEN 'somewhat_active'
WHEN ROUND(AVG(TotalSteps),2) BETWEEN 10000 and 12499 THEN 'active'
WHEN ROUND(AVG(TotalSteps),2) > 12499 THEN 'highly_active'
END AS total_steps_indices
FROM `harshit-401709.fitbit_data.daily_activity`
Group by Id
ORDER BY avg_totalsteps
```

“

>>14 users have a sedentary lifestyle and only 3 users have a highly active lifestyle.

## 7. To find the lifestyle based on total steps.



-Indices reference used from “How many steps/day are enough?” from Catrine Tudor-Locke and David R Bassett Jr as mentioned below:

-Preliminary pedometer indices for public health proposed the following indices in healthy adults

Sedentary lifestyle index < 5000 steps a day

low-active  $5000 \leq 7499$

somewhat active  $7500 \leq 9999$

Active > 10000

Highly active >12500

>>Based on the above chart we have 14 sedentary lifestyle users, which means people with sedentary lifestyle are using more smart devices. 6 users were classified as low active, 9 users as somewhat active, 3 as active and 3 as highly active.

## 8. To find out how many users are meeting WHO expectations of walking steps.

“

```
SELECT Id,  
avg(veryactiveminutes) + avg(fairlyactiveMinutes) + avg(lightlyactiveMinutes) AS  
total_avg_active_minutes,  
CASE  
WHEN avg(veryactiveminutes) + avg(fairlyactiveminutes) + avg(lightlyactiveminutes) >=  
300 THEN 'exceeds_WHO_recommendation'  
WHEN avg(veryactiveminutes) + avg(fairlyactiveminutes) + avg(lightlyactiveminutes)  
BETWEEN 150 AND 300 THEN 'meets_WHO_recommendation'
```

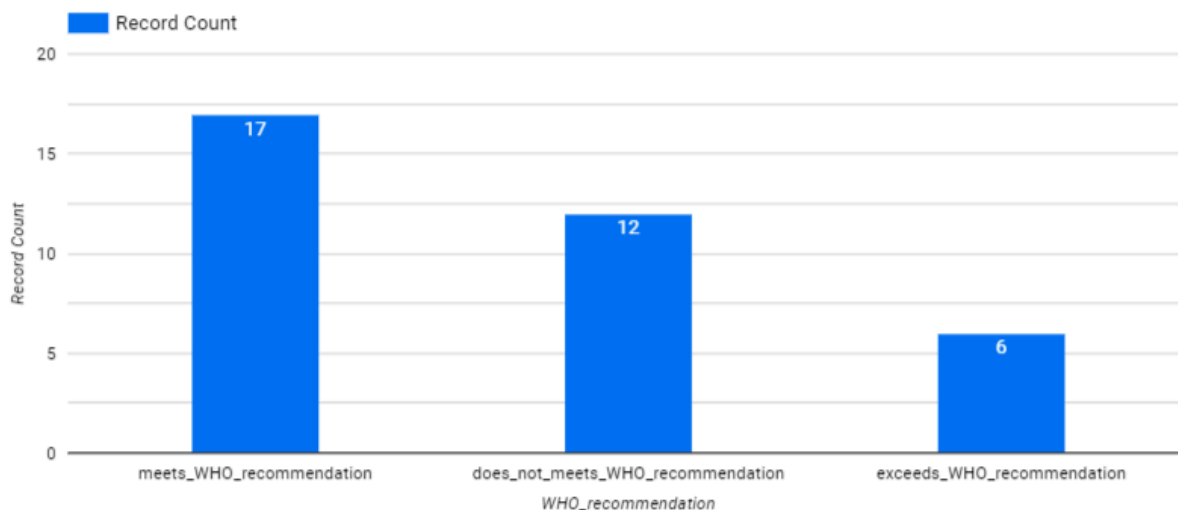
```

WHEN avg(veryactiveminutes) + avg(fairlyactiveminutes) + avg(lightlyactiveminutes)
<150 THEN 'does_not_meets_WHO_recommendation'
END AS WHO_recommendation
FROM `harshit-401709.fitbit_data.daily_activity`
GROUP BY Id

```

“

## daily\_activity



The World Health Organization recommends that adults between the ages of 18–64 years of age should do at least 150–300 minutes of moderate-intensity aerobic physical activity.

This chart suggested that there are 12 people in the dataset who do not meet WHO expectation for walking and 17 people meet with WHO recommendation. Only 6 people were able to exceed recommendations.

### 9. To find out how many users are meeting WHO expectations of walking steps without considering lightly active minutes.

“

```

SELECT Id,
avg(veryactiveminutes) + avg(fairlyactiveMinutes) + avg(lightlyactiveMinutes) AS
total_avg_active_minutes,
CASE
WHEN avg(veryactiveminutes) + avg(fairlyactiveminutes) >= 300 THEN
'exceeds_WHO_recommendation'
WHEN avg(veryactiveminutes) + avg(fairlyactiveminutes) BETWEEN 150 AND 300 THEN
'meets_WHO_recommendation'

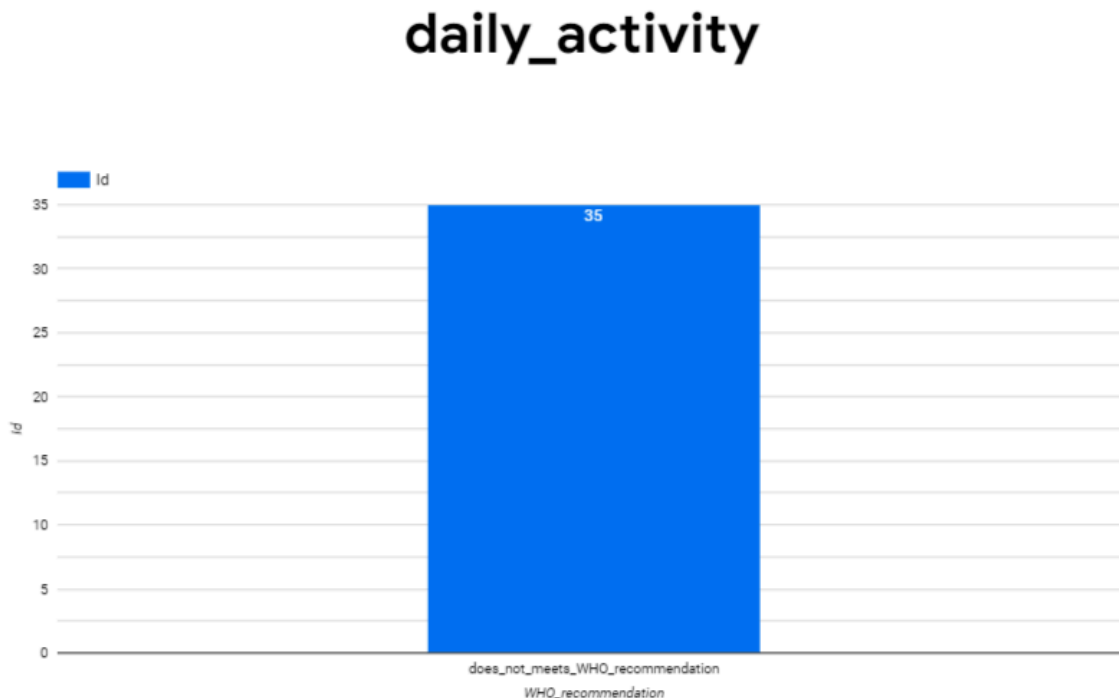
```



```

WHEN avg(veryactiveminutes) + avg(fairlyactiveminutes) <150 THEN
'does_not_meets_WHO_recommendation'
END AS WHO_recommendation
FROM `harshit-401709.fitbit_data.daily_activity`
GROUP BY Id

```



>> the resulting table showed that there were no users meeting WHO expectations when lightly active minutes are not considered.

#### 10. To confirm daily activity calory data is same as hourly\_calorie data.

```

"SELECT dailyactivity.id AS dailyactivity_id,
calories.id AS calories_id,
COUNT(dailyactivity.calories) AS dailyactivity_calories,
COUNT(calories.calories) AS total_hourly_calories
FROM `harshit-401709.fitbit_data.daily_activity` dailyactivity
LEFT JOIN `harshit-401709.fitbit_data.hourly_Calories` calories
ON dailyactivity.id = calories.id
GROUP BY dailyactivity.id, calories.id

```

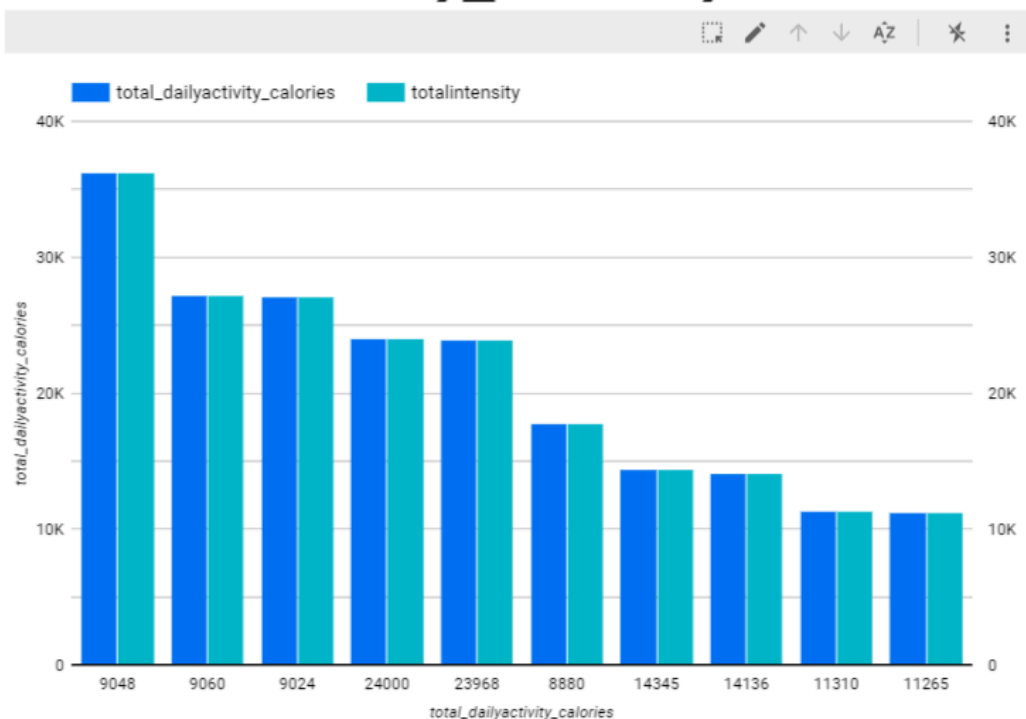
>>This confirms that the data is same in both tables so we will be using daily\_activity table to comparing colories and workout intensities.

## 11. To compare workout intensity with calories burned.

“

```
SELECT dailyactivity.id AS dailyactivity_id,  
intensity.id AS intensity_id,  
COUNT(dailyactivity.calories) AS total_dailyactivity_calories,  
COUNT(TotalIntensity) AS totalintensity  
FROM `harshit-401709.fitbit_data.daily_activity` dailyactivity  
RIGHT JOIN `harshit-401709.fitbit_data.hourly_Intensities` intensity  
ON dailyactivity.id = intensity.id  
GROUP BY dailyactivity.id, intensity.id
```

### daily\_activity



“

>>This chart confirmed that the calories burned were equal to the workout intensity.

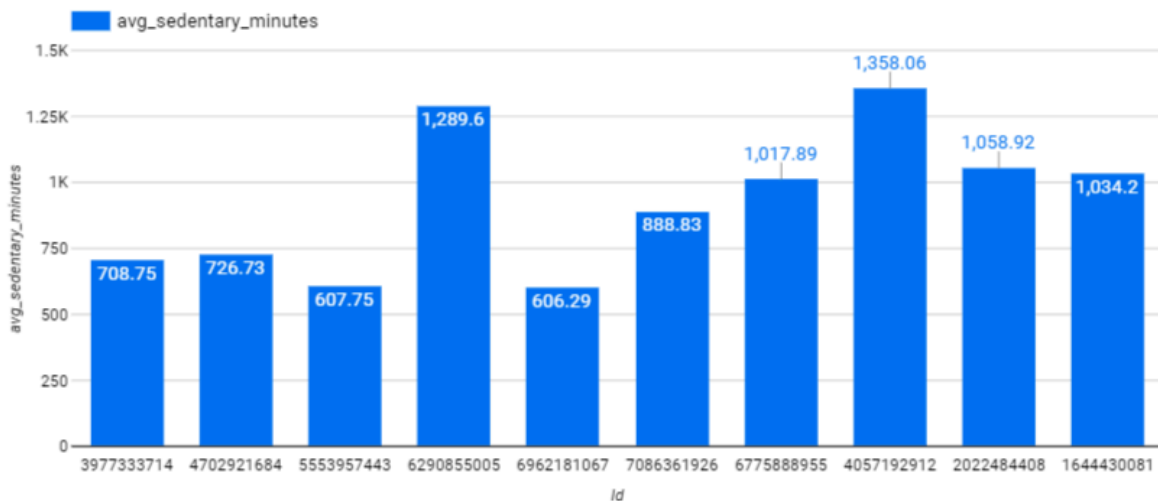
## 12. To find out Average sedentary time per user.

“

```
select Id,  
AVG(SedentaryMinutes) AS avg_sedentary_minutes  
FROM `harshit-401709.fitbit_data.daily_activity`  
Group by Id
```

“

# daily\_activity



>>> The results showed that the average sedentary time comes to 989.34 minutes per user. Which is close to 16 hours per day and needs to be reduced.

## Share

Here are some insights that I got from my analysis,

- Most of the users are moderate users as they contributed 68% of the total users in the dataset whereas 25% users were classified as light users.
- The intensity of exercise equaled calories burned.
- There are 12 people in the dataset who do not meet WHO expectation for walking and 17 people meet with WHO recommendation. Only 6 people were able to exceed recommendations.
- We have 14 sedentary lifestyle users, which means people with sedentary lifestyle are using more smart devices, 6 users were classified as low active, 9 users as somewhat active, 3 as active and 3 as highly active.
- The average sedentary time comes to 989.34 minutes per user. Which is close to 16 hours per day and needs to be reduced.

# Act

Here are my recommendations and Strategies for the Marketing Team (From My Analysis).

## Target Audience:

- **Let's leverage the moderate user majority (68%)!** We can tailor campaigns and messaging to their needs and goals. Motivate them to keep moving forward and celebrate their progress towards a healthier lifestyle.
- **Light users (25%) have potential!** I recommend developing strategies to convert them into more regular users. Educational content or challenges can help them build healthy habits.

## Content and Messaging:

- **Progress, not perfection, is key!** Let's emphasize achieving realistic goals and gradual improvement. We can showcase how Bellabeat can help users become more active, even if they start with small changes.
- **Benefits matter!** Don't just promote exercise; connect it to tangible benefits users care about, such as better sleep, stress management, or improved heart health. The data (e.g., WHO recommendations) supports this focus.
- **Let's tackle sedentary concerns!** I found 14 users with sedentary lifestyles using our devices. We can develop content or challenges specifically targeted at reducing sedentary time. Offering tips and tricks for incorporating more movement throughout the day can be really helpful.

## Product Development:

- **Interesting discovery!** We should investigate why people with sedentary lifestyles (14 users) are using smart devices. This could be an opportunity to tailor app features or challenges to help them become more active.
- **Goal setting and tracking are crucial!** I recommend enhancing the app's functionalities to help users set personalized goals and track their progress towards achieving them. Offering rewards or recognition for reaching milestones can provide extra motivation.

## Additional Strategies:

- **Partnerships can be powerful!** Let's consider partnering with fitness experts or influencers to create engaging content or challenges for Bellabeat users.
- **Community is key!** Fostering a supportive user community within the app or through social media can encourage motivation and knowledge sharing.

**Remember:**

- We should tailor our communication based on user activity level (moderate, light, etc.).
- Data insights are valuable! Let's leverage them to personalize the user experience within the app.
- Continuous monitoring and refinement are essential. We should use user feedback and data analysis to keep improving our marketing efforts.

By implementing these recommendations, I believe we can help the Bellabeat marketing team attract new users, engage existing users more effectively, and ultimately help them achieve their health and wellness goals.