Fitbit Fitness Tracker Data Analysis

#### This is my Capstone Project for the Google Data Analytics Professional Certificate, which I have recently completed.

To begin, let’s explore the context. Bellabeat, a high-tech manufacturer of health-focused products for women. Since it was founded in 2013, Bellabeat has grown rapidly and quickly positioned itself as a tech-driven wellness company for women. Bellabeat is a successful small company, but they have the potential to become a larger player in the global smart device market. Urška Sršen, co-founder and Chief Creative Officer of Bellabeat, believes that analyzing smart device fitness data could help unlock new growth opportunities for the company.

Let’s go through each step of the analysis process.

##### **Step 1: Ask**

#### **Business Task Summary:**

The main task is to analyze trends in smart device usage, particularly focusing on how consumers are interacting with non-Bellabeat devices, and apply these insights to one Bellabeat product. The goal is to gain insights that will help Bellabeat better understand its customers, improve its products, and refine its marketing strategies.

**These questions will guide our 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?

**The key stakeholders include:**

Urška Sršen — Chief Creative Officer and Bellabeat’s Co-founder.

Sando Mur — Mathematician and Bellabeat’s Co-founder.

Bellabeat’s marketing analytics team — a team of data analysts.

##### **Step 2: Prepare**

These datasets were made available through Mobius in Kaggle and licensed under the CCO Public Domain. According to the dataset information available on Zenodo.org, these datasets were generated by respondents to a distributed survey via Amazon Mechanical Turk between 03.12.2016–05.12.2016**.**

The datasets downloaded were in a zip file from which I extracted 18 CSV files representing 18 data sets a brief period from April to May 2016.

A publicly available dataset containing personal fitness tracker data from 30 users. It includes daily on physical activity, heart rate, sleep monitoring, and steps. That the data is stored in a long format with each ID having data in multiple rows.

Dataset owner: Furberg, R., Brinton, J., Keating, M., & Ortiz, A. (2016). Crowd-sourced Fitbit datasets 03.12.2016-05.12.2016 [Data set]. [Zenodo](http://doi.org/10.5281/zenodo.53894)

**After reviewing the data and based on it, I understood the following:**

**Self-reporting Bias**: Users may inaccurately log activities, which could affect the accuracy of the data. For instance, over-reporting steps or under-reporting inactive time.

**Selection Bias**: The dataset may not represent a wide demographic, as users of Fitbit products may primarily belong to certain income groups or geographic locations, which could limit generalizability.

**Data Size**: The dataset contains information from only **30 users**, which is a very small sample. This means it doesn’t represent larger fitness trends or the global population. Because the data is limited to these 30 users.

**Short Date Range**: The data covers only **April to May 2016**, which is a short period. This means it might miss seasonal trends or long-term changes in user behavior, limiting the depth of the analysis and making it harder to generalize findings over a longer timeframe.

##### **Step 3: Process**

**What tools are you choosing and why?**

For data processing I worked with Excel.

Cleaning was done using Excel and SQL. The following steps were taken for each dataset.

FirstI checked for blanks.

I sorted and filtered the data.

Then I checked for duplicate values and in minutesleep\_merged 543 duplicates. were removed and 187978 remained while in sleepday\_merged 3 duplicates were removed and 410 unique values remained.

Checking for missing or NA and zero values

Filter Out Outliers (total steps > 0)

split date-time columns into two separate columns: date, time.I formatted the designated date and time columns to MM/DD/YYYY and the time columns to hh:mm:ss and split the merged date/time column using INT().

I also added a column with the days of the week.Finally, rows with data not in the correct format were deleted.

Renamed the columns and simplified them for better understanding.

**Step 4: Analyze**

I used the SQL import wizard to upload my Excel files for analysis. After reviewing the four tables (dailyactivity, dailycalories, dailyintensities, and dailysteps), I wanted to check if they matched by user IDs and activity dates. To do this, I ran an INNER JOIN query to compare them.

**select activity.calories, calories.calories**

**from `steel-spark-430803-r3.Bellabeat\_case.daily\_activity` activity**

**INNER JOIN `steel-spark-430803-r3.Bellabeat\_case.daily\_calories` calories**

**ON activity.Id = calories.Id**

**AND activity.date = calories.`activity day`;**

**SELECT activity.Id, activity.totalsteps, steps.`step total`,**

**FROM `steel-spark-430803-r3.Bellabeat\_case.daily\_activity` activity**

**INNER JOIN `steel-spark-430803-r3.Bellabeat\_case.daily\_steps` steps**

**ON activity.id = steps.id**

**AND activity.date = steps.`activity day`;**

**SELECT activity.`sedentary minutes`,**

**intensities.`sedentary minutes`**

**FROM `steel-spark-430803-r3.Bellabeat\_case.daily\_activity` activity**

**INNER JOIN `steel-spark-430803-r3.Bellabeat\_case.daily\_intensities` intensities**

**ON activity.id = intensities.id**

**AND activity.date = intensities.`activity day`**

After checking, it became clear that the data in dailyactivity, dailycalories, dailyintensities, and dailysteps are the same as dailyactivity. Therefore, we will work with the data in the table.

To find out how often each user logged in or wore their Fitbit watches, I ran a query to count the number of entries for each user ID.

**select Id,**

**COUNT(Id) as total\_id**

**from `steel-spark-430803-r3.Bellabeat\_case.daily\_activity`**

**GROUP BY (Id)**

**ORDER BY(total\_id) ASC**

After running the query, we observed that users logged activities anywhere from 4 to 31 times. Based on this, we can classify users into three categories: light activity, moderate activity, and active users.

**SELECT id,**

**COUNT(id) AS total\_uses,**

**CASE**

**WHEN COUNT(Id) BETWEEN 23 AND 31 THEN 'active\_user'**

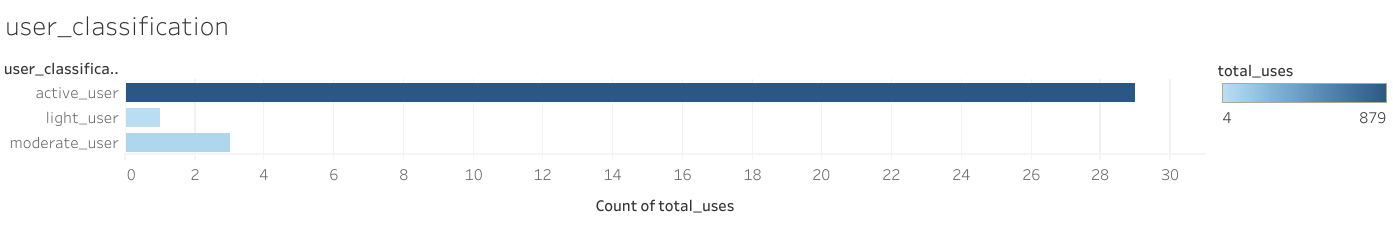
**WHEN COUNT(Id) BETWEEN 14 and 22 THEN 'moderate\_user'**

**WHEN COUNT(Id) BETWEEN 4 and 13 THEN 'light\_user'**

**END AS user\_classification**

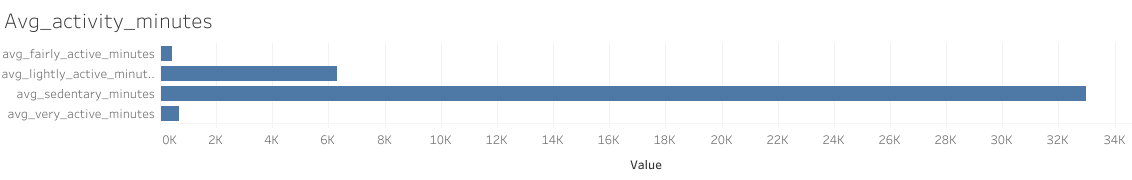
**FROM `steel-spark-430803-r3.Bellabeat\_case.daily\_activity`**

**GROUP BY id**

The majority of users fall into the Active Users category, indicating a high level of engagement with the Fitbit devices. However, the percentage of Moderate and Light Users suggests an opportunity to encourage more consistent usage among these groups. Strategies could be implemented to increase user engagement, such as personalized activity goals or reminders.

Then I wanted to take a look at the highest average number of minutes spent actively.

Throughout the day one Fitbit user would spend 80% of the time being sedentary and 20% of the time doing various types of physical activity. This indicates that Fitbit is not only used as sportswear accessory but also casually during the day.

****

**SELECT `day of week` as day\_of\_week,**

**ROUND(Avg(`very active minutes`),2) AS avg\_very\_active\_minutes,**

**ROUND(Avg(`fairly active minutes`),2) AS avg\_fairly\_active\_minutes,**

**ROUND(Avg(`lightly active minutes`),2) AS avg\_lightly\_active\_minutes,**

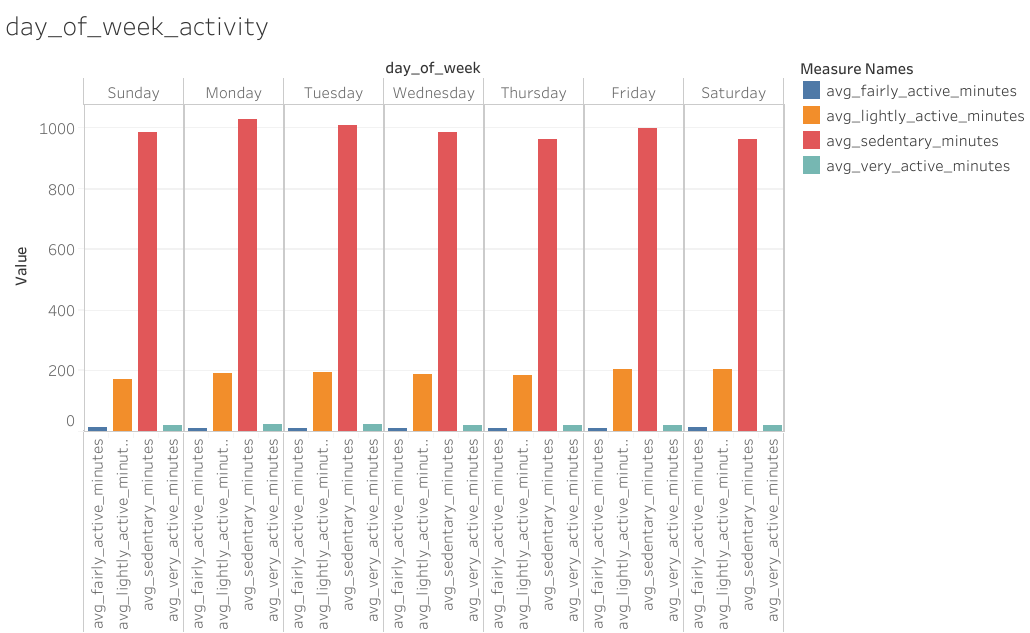
**ROUND(Avg(`sedentary minutes`),2) AS avg\_sedentary\_minutes**

**FROM `steel-spark-430803-r3.Bellabeat\_case.daily\_activity`**

**GROUP BY day\_of\_week**

**ORDER BY 2,3,4,5 DESC**

Next, I wanted to find out which day is less active.

****

Throughout the day one Fitbit user would spend 80% of the time being sedentary and 20% of the time doing various types of physical activity. This indicates that Fitbit is not only used as sportswear accessory but also casually during the day.

Based on the research and recommendations that the steps indicated in the article

<https://pubmed.ncbi.nlm.nih.gov/14715035/>

It is clear from the article that <5000 steps per day can be used as a “sedentary lifestyle index”;

5000-7499 steps per day are typical for daily activity excluding sports/exercise and may be considered “low activity”;

7500-9999 likely involves some volitional action (and/or increased demands on professional activity) and may be considered “fairly active”; and (iv).

>or=10,000 steps/day indicates the point that should be used to classify people as “active.”

People who take >12,500 steps/day are likely to be classified as “highly active.”

**SELECT Id, ROUND(AVG(` total steps`),2) AS avg\_totalsteps,**

**CASE**

**WHEN ROUND(AVG(` total steps`),2) < 5000 THEN 'sedentary\_lifestyle'**

**WHEN ROUND(AVG(` total steps`),2) BETWEEN 5000 and 7499 THEN 'low\_active'**

**WHEN ROUND(AVG(` total steps`),2) BETWEEN 7500 and 9999 THEN 'somewhat\_active'**

**WHEN ROUND(AVG(` total steps`),2) BETWEEN 10000 and 12499 THEN 'active'**

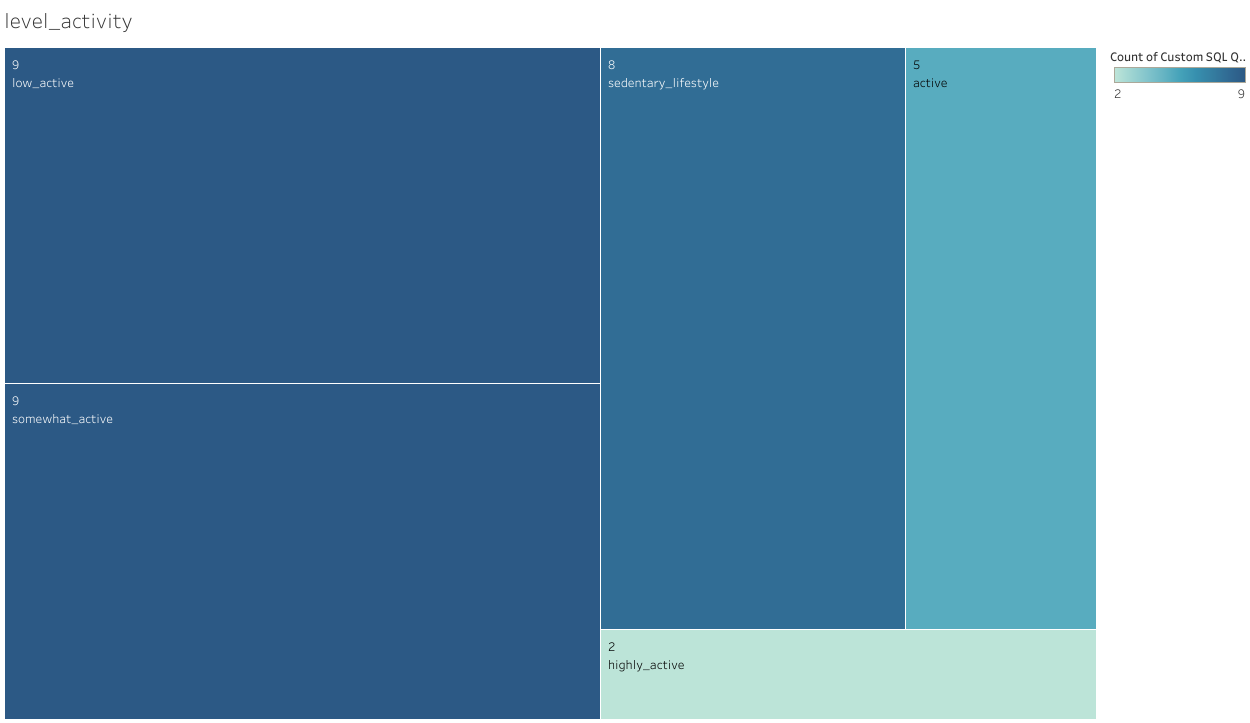
**WHEN ROUND(AVG(` total steps`),2) > 12499 THEN 'highly\_active'**

**END AS total\_steps\_indices**

**FROM `steel-spark-430803-r3.Bellabeat\_case.daily\_activity`**

**Group by Id**

**ORDER BY avg\_totalsteps**

****

The above query showed that 8 users fell under the sedentary lifestyle category:

9 users were classified as low active

9 users as somewhat active

5 as active and

2 as highly active

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.

[**https://www.who.int/news-room/fact-sheets/detail/physical-activity**](https://www.who.int/news-room/fact-sheets/detail/physical-activity)

I checked and made an analysis based on the given data and analysis.

**SELECT Id,**

**avg( `very active minutes`) + avg(`fairly active minutes`) + avg(`lightly active minutes`) AS total\_avg\_active\_minutes,**

**CASE**

**WHEN avg( `very active minutes`) + avg(`fairly active minutes`) + avg(`lightly active minutes`) >= 300 THEN 'exceeds\_WHO\_recommendation'**

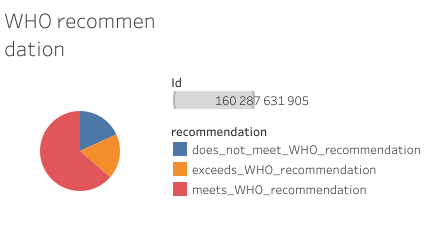
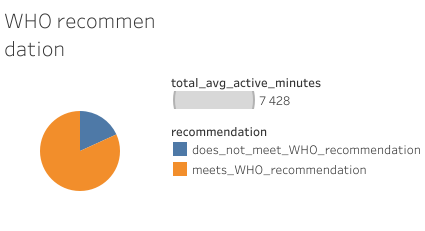
**WHEN avg(`very active minutes`) + avg(`fairly active minutes`) + avg(`lightly active minutes`) BETWEEN 150 AND 300 THEN 'meets\_WHO\_recommendation'**

**WHEN avg( `very active minutes`) + avg(`fairly active minutes`) + avg(`lightly active minutes`) < 150 THEN 'does\_not\_meet\_WHO\_recommendation'**

**END AS recommendation**

**FROM `steel-spark-430803-r3.Bellabeat\_case.daily\_activity`**

**GROUP BY Id;**



**SELECT Id,**

**avg( `very active minutes`) + avg(`fairly active minutes`) + avg(`lightly active minutes`) AS total\_avg\_active\_minutes,**

**CASE**

**WHEN avg( `very active minutes`) + avg(`fairly active minutes`) + avg(`lightly active minutes`) >= 150 THEN 'meets\_WHO\_recommendation'**

**WHEN avg( `very active minutes`) + avg(`fairly active minutes`) + avg(`lightly active minutes`) < 150 THEN 'does\_not\_meet\_WHO\_recommendation'**

**END AS recommendation**

**FROM `steel-spark-430803-r3.Bellabeat\_case.daily\_activity`**

**GROUP BY Id;**

Participant Activity Analysis:

Participants Meeting Recommendations: 21

Participants Exceeding Recommendations: 6

Participants Not Meeting Recommendations: 6

**SELECT `day of week`,**

**ROUND(AVG(` total steps`),2) AS avg\_step,**

**ROUND(AVG(`total distance`),2) AS avg\_dist,**

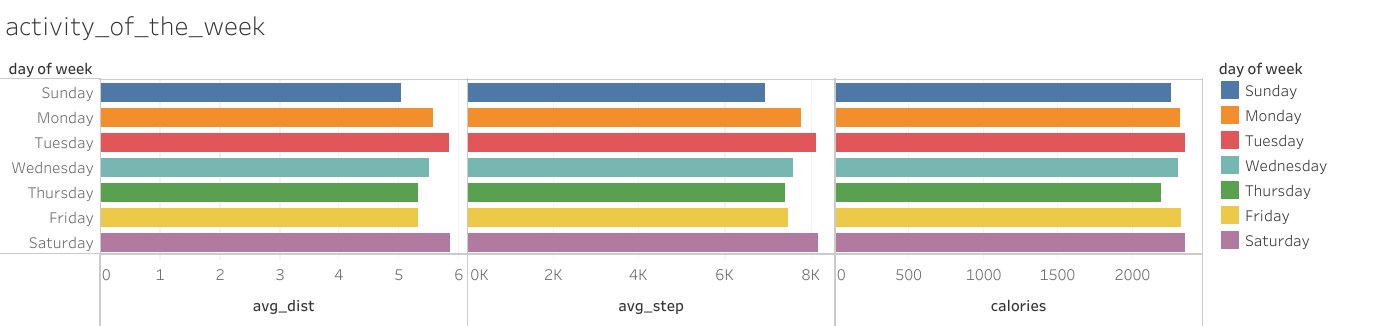
**ROUND(AVG(calories),2) AS calories,**

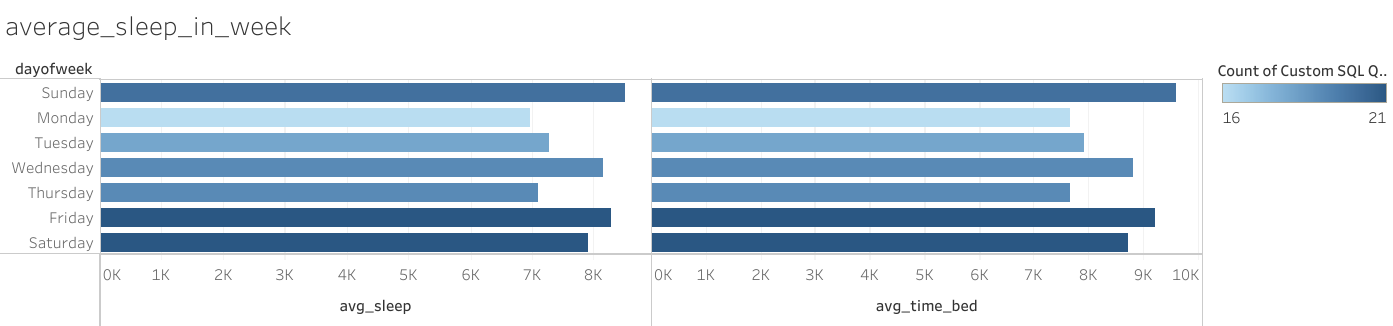
**FROM `steel-spark-430803-r3.Bellabeat\_case.daily\_activity`**

**GROUP BY `day of week`**

**ORDER BY 2 DESC**

As shown in the graph, the most active days were Tuesday and Saturday.





Based on an article from the Sleep Foundation by Dr. Elizabeth Rausch-Phung and colleagues, “Most healthy people fall asleep within 15 to 20 minutes of lying down.” To explore the connection between sleep and physical activity, I examined the total steps recorded in the daily activity table by using a JOIN statement to query the tables.

**SELECT**

**sd.Id,**

**ROUND(AVG(da.`total steps`), 2) AS avg\_steps,**

**ROUND(AVG(da.calories), 2) AS avg\_calories,**

**ROUND(AVG(sd.TotalMinutesAsleep), 2) AS avg\_minutes\_asleep,**

**ROUND(AVG(sd.TotalTimeInBed), 2) AS avg\_time\_in\_bed,**

**ROUND(AVG(sd.TotalTimeInBed - sd.TotalMinutesAsleep), 2) AS avg\_time\_awake**

**FROM**

**`steel-spark-430803-r3.Bellabeat\_case.sleep\_day` sd**

**INNER JOIN**

**`steel-spark-430803-r3.Bellabeat\_case.daily\_activity` da**

**ON**

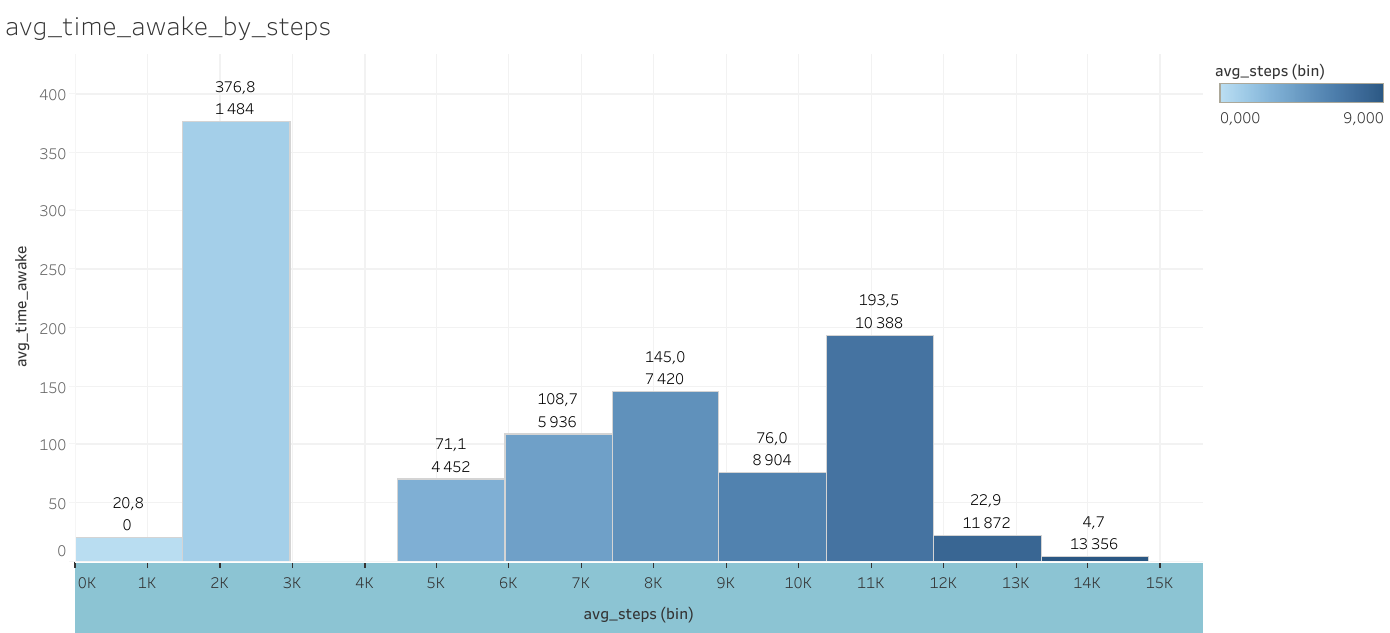
**sd.Id = da.Id**

**GROUP BY**

**sd.Id**

**ORDER BY**

**avg\_minutes\_asleep DESC;**

****

On doing this, I found that the highest average time awake was 376 minutes and the lowest was 4 minute but the total steps for the former was an average of 1484 steps indicating a sedentary lifestyle and the latter was 13 356 indicating an active lifestyle according to the categorization of Catrine Tudor-Locke et al in their How many steps/day are enough?

<https://pubmed.ncbi.nlm.nih.gov/14715035/?source=post_page-----c18835475563-------------------------------->

## **Share**

These are some of the insights I deduced from my analysis.

Throughout the day one Fitbit user would spend 80% of the time being sedentary and 20% of the time doing various types of physical activity. This indicates that Fitbit is not only used as sportswear accessory but also casually during the day.

The graph clearly shows that on Monday, Tuesday and Friday the amount of average sedentary activity increased.

The majority of users fall into the Active Users category, indicating a high level of engagement with the Fitbit devices. However, the percentage of Moderate and Light Users suggests an opportunity to encourage more consistent usage among these groups. Strategies could be implemented to increase user engagement, such as personalized activity goals or reminders.

The most active days were Tuesday and Saturday.

**Act**

Here are my top three recommendations.

Bellabeat’s marketing strategy should focus on educating users about the importance of staying active for their overall health and well-being. The key features of the watch, such as tracking daily activity goals, progress monitoring, and reminders to stay active, should be at the forefront of this message. Promoting the “get up and move” feature can encourage users to consistently complete their activity rings each day, making them feel more accomplished.

Additionally, since Bellabeat is primarily targeted at women, incorporating period tracking could add significant value. It could help women feel more confident in staying active even during menstruation, offering helpful reminders on how to keep moving at that time of the month. This would make it easier for users to achieve their goals every day, regardless of their menstrual cycle.

In general, people who take more steps tend to sleep better (spend less time awake at night). However, there are exceptions where people with a high step count actually spend more time awake. Overall, there is a correlation, but it's not always straightforward.