# BELLABEAT

# Bellabeat Case Study

## Using Excel, SQL & Tableau

### **Table of Contents**

* 1. [Company Summary](https://www.kaggle.com/code/yurialejandrofarinas/bellabeat-case-study-excel-sql-tableau#companysummary)
     + [1.1 Company Background](https://www.kaggle.com/code/yurialejandrofarinas/bellabeat-case-study-excel-sql-tableau#companybackground)
     + [1.2 Company Products](https://www.kaggle.com/code/yurialejandrofarinas/bellabeat-case-study-excel-sql-tableau#companyproducts)
  2. [Ask Phase](https://www.kaggle.com/code/yurialejandrofarinas/bellabeat-case-study-excel-sql-tableau#ask)
     + [2.1 Business Task](https://www.kaggle.com/code/yurialejandrofarinas/bellabeat-case-study-excel-sql-tableau#businesstask)
     + [2.2 Stakeholders](https://www.kaggle.com/code/yurialejandrofarinas/bellabeat-case-study-excel-sql-tableau#stakeholders)
* [3. Prepare Phase](https://www.kaggle.com/code/yurialejandrofarinas/bellabeat-case-study-excel-sql-tableau#prepare)
  1. [3.1 Data Used](https://www.kaggle.com/code/yurialejandrofarinas/bellabeat-case-study-excel-sql-tableau#dataused)
  2. [3.2 Accessibility & Usage of Data](https://www.kaggle.com/code/yurialejandrofarinas/bellabeat-case-study-excel-sql-tableau#dataaccessibility)
  3. [3.3 Data Summary](https://www.kaggle.com/code/yurialejandrofarinas/bellabeat-case-study-excel-sql-tableau#datasummary)
  4. [3.4 Data Organization](https://www.kaggle.com/code/yurialejandrofarinas/bellabeat-case-study-excel-sql-tableau#dataorg)
  5. [3.5 Data Limitations & Integrity](https://www.kaggle.com/code/yurialejandrofarinas/bellabeat-case-study-excel-sql-tableau#datalimit)
* [4. Process Phase](https://www.kaggle.com/code/yurialejandrofarinas/bellabeat-case-study-excel-sql-tableau#process)
  1. [4.1 Datasets Selected](https://www.kaggle.com/code/yurialejandrofarinas/bellabeat-case-study-excel-sql-tableau#dataselected)
  2. [4.2 Using Excel to Clean Data](https://www.kaggle.com/code/yurialejandrofarinas/bellabeat-case-study-excel-sql-tableau#using)
  3. [4.3 Next Steps](https://www.kaggle.com/code/yurialejandrofarinas/bellabeat-case-study-excel-sql-tableau#nextsteps)
* [5. Analyze and Share Phases](https://www.kaggle.com/code/yurialejandrofarinas/bellabeat-case-study-excel-sql-tableau#analyze)
  1. [5.1 SQL Dataset Upload](https://www.kaggle.com/code/yurialejandrofarinas/bellabeat-case-study-excel-sql-tableau#sql)
  2. [5.2 User Verification](https://www.kaggle.com/code/yurialejandrofarinas/bellabeat-case-study-excel-sql-tableau#userv)
  3. [5.3 User Insights](https://www.kaggle.com/code/yurialejandrofarinas/bellabeat-case-study-excel-sql-tableau#useri)
     + [5.3.1 User usage of wearable](https://www.kaggle.com/code/yurialejandrofarinas/bellabeat-case-study-excel-sql-tableau#userw)
     + [5.3.2 User Data Summary](https://www.kaggle.com/code/yurialejandrofarinas/bellabeat-case-study-excel-sql-tableau#usersummary)
  4. [5.4 User Types by Activity Levels](https://www.kaggle.com/code/yurialejandrofarinas/bellabeat-case-study-excel-sql-tableau#usertypes1)
     + [5.4.1 User Types by Active Minutes](https://www.kaggle.com/code/yurialejandrofarinas/bellabeat-case-study-excel-sql-tableau#usertypes2)
     + [5.4.2 User Types by Total Steps](https://www.kaggle.com/code/yurialejandrofarinas/bellabeat-case-study-excel-sql-tableau#usertypes3)
  5. [5.5 Calories, Steps & Active Minutes by ID](https://www.kaggle.com/code/yurialejandrofarinas/bellabeat-case-study-excel-sql-tableau#calories)
  6. [5.6 Total Steps by Day](https://www.kaggle.com/code/yurialejandrofarinas/bellabeat-case-study-excel-sql-tableau#totalstepsday)
  7. [5.7 Total Steps by Hour](https://www.kaggle.com/code/yurialejandrofarinas/bellabeat-case-study-excel-sql-tableau#totalstepshour)
  8. [5.8 Deeper Look into Sleep](https://www.kaggle.com/code/yurialejandrofarinas/bellabeat-case-study-excel-sql-tableau#deeper)
* [6. ACT Phase](https://www.kaggle.com/code/yurialejandrofarinas/bellabeat-case-study-excel-sql-tableau#act)
  1. [6.1 Conclusion & Recommendations](https://www.kaggle.com/code/yurialejandrofarinas/bellabeat-case-study-excel-sql-tableau#conclusion)

## 1. Company Summary

#### **1.1 Company Background**

According to Forbes, “Bellabeat is a data-oriented wellness tech company that was founded by Sandro Mur, Urška Sršen, and Lovepreet Singh in 2013” (Robter, 2020). The company is global with offices in London, Hong Kong and Zagreb, but is headquarter in San Francisco. The company focuses on women’s health and wellness with a collection of wearable and non-wearable tech. The four pillars of their brand include: smart insights, women-centric, holistic approach and body positivity. With this in mind, their products focus on including metrics around a woman’s menstrual cycle stating on their website that “Bellabeat helps you get in sync with your natural cycle” and that “through tracking your body’s bio-responses and aligning that data with your hormonal cycle, you’ll always know why you feel how you do.”

#### **1.2 Company Products**

The company’s wearable products include:

1. **Ivy** – “a health tracker disguised as smart jewelry”
2. **Time** – “an elegant hybrid wellness watch”
3. **Leaf** – available in three styles: chakra, urban, & crystal and can be worn as a necklace, bracelet or clip. This was Bellabeat’s classic wellness tracker.
4. **Spring** – a “smart water bottle” designed to track your drinking / hydration habits.

All of the company’s wearables sync to their Bellabeat app where members can check their metrics. The wearables track activity (steps taken, distanced traveled, calories burned and activity minutes) & sleep and through the app you can also track your menstrual cycle, hydration (if not using Spring) & meditation. Their IVY wearable also tracks heart rate metrics.

Bellabeat also offers a Wellness Coach app with “unlimited access to 400+ education video, written, and audio content from areas of beauty, fitness, mindfulness, women’s health and more” through their Bellabeat+ membership.

## 2. ASK Phase

#### **2.1 Business Task**

Utilizing the Fitbit Fitness Tracker Data, identify some trends in smart device usage, how these trends can be applied to Bellabeat’s customers and how they can help influence Bellabeat’s marketing strategy.

#### **2.2 Stakeholders**

1. **Urška Sršen** – Bellabeat’s cofounder and Chief Creative Officer
2. **Sando Mur** – Mathematician and Bellabeat’s cofounder
3. **Bellabeat’s marketing analytics team** – a team of data analytics

## 3. PREPARE Phase

#### **3.1 Data Used**

The data source used for this case study is **[FitBit Fitness Tracker Data.](https://www.kaggle.com/datasets/arashnic/fitbit)** This dataset was downloaded from Kaggle where it was uploaded by Möbius.

#### **3.2 Accessibility & Usage of Data**

The dataset was published by Möbius to Kaggle.com under the CC0: Public Domain Creative Common License – waiving all rights to the work and allowing for the dataset to be copies, modified, distributed and performed without asking for permission. Möbius cited the dataset from Zendo: Furberg, Robert; Brinton, Julia; Keating, Michael ; Ortiz, Alexa [[Source Here]](https://zenodo.org/record/53894#.YMoUpnVKiP9)

#### **3.3 Data Summary**

According to the dataset information submitted on Zenodo.org,"this dataset was generated by respondents to a distributed survey via Amazon Mechanical Turk between 03.12.2016 – 05.12.2016.Thirty eligible Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring." Additionally, "Variation between output represents use of different types of Fitbit trackers and individual tracking behaviors / preferences."

#### **3.4 Data Organization**

Eighteen datasets were downloaded from the **FitBit Fitness Tracker Data.** The datasets downloaded in .csv file format and included long and wide formats. The datasets chosen for analysis below included a user count of 33 participants over a 31 day period of time.

#### **3.5 Data Limitations & Integrity**

The **FitBit Fitness Tracker Data** was collected in 2016 making the datasets outdated for current trend analysis. Additionally, while the data initially states a time range of 03-12-2016 to 05-12-2016, after data verification, the data collected was only during a 31 day period (04-12-2016 to 05-12-2016). Since the data only included instances over a 31 day period, the timeframe for a more insightful analysis is realitively small.

Lastly, the sample size itself could create a sample bias. While a sample size of 30 (our data later shows a sample size of 33) will hold up within the CLT theorm, a larger sample size will be more representative of the population and would increase the confidence interval. Additionally, since there were no demographic information collected it will be hard to see if we have a true representation of a national or global population. This lack of demographic information will also limit recommendations on the target audience (including gender, location, age and job status) and where to market to them. Considering that Bellabeat is primarily targeted to women and individuals who menstruate, having demographics would have bolstered any recommendations after analysis.

## 4. PROCESS Phase

#### **4.1 Datasets Selected**

For the casestudy analysis the following datasets were chosen:

* Daily\_Activity\_Merged
* Daily\_Sleep\_Merged
* Hourly\_Steps\_Merged
* Hourly\_Intensity\_Merged
* Hourly\_Calories\_Merged
* Heart\_Rate\_Merged
* Weight\_Log\_Merged

#### **4.2 Using Excel to Clean Data**

Each dataset was cleaned using Excel. The following steps were taken within each dataset:

* Sorted and filtered data by Id to obtain how many unique users there were within the dataset.
* Checked for duplicate data using the 'duplicate data' tool in Excel
* Formatted date data into MM/DD/YY date format
* Formatted all numerical data into Number format with either no decimils or up to 2 decimials.
* Sorted by date to find the first and last date of the dataset (this is what first indicated only a 31-day period of activity was captured).
* Separated Date and Hour into two columns when needed for later analysis. Utilized the 'Text to Columns' tool to do so.
* Formatted any time data into 00:00:00 format for consistency.
* Checked Id entries and other columns for LEN to make sure the data was correct and uniform in length

After the cleaning process was finished, only 3 rows of duplicate information was found within the Daily\_Sleep\_Merged file. These were removed before analysis.

#### **4.3 Next Steps**

After cleaning the data, I decided to work with SQL to analyze the information versus staying in Excel (although I did go back a few times for visualizations). Why SQL? SQL and Excel are the two biggest skills being requested (& used) within a Data Analysts' job. So I'm hoping to showcase that skillset here.

With that said, Excel and R would have worked for anaylsis as well. Both [MACARENA LACASA](http://https:/www.kaggle.com/code/macarenalacasa/capstone-case-study-bellabeat) and [ANASTASIIA CHEBOTINA](https://www.kaggle.com/code/chebotinaa/bellabeat-case-study-with-r) have good examples of analyzing this data with R.

## 5. ANALYZE & SHARE Phases

#### **5.1 SQL Dataset Upload**

Uploaded **FitBit Fitness Tracker Data** into BigQuery under the project capstone-course-120622

Uploaded the following clean data sets:

* dailyActivity\_merged
* sleepDay\_merged
* heartrate\_seconds\_merged
* hourlyCalroie\_merged
* hourlyIntensity\_merged
* hourlySteps\_merged
* weightLogInfo\_merged

#### **5.2 User Verification**

Checked for # of participants by counting number of distinct Ids in each dataset.

SELECT COUNT (DISTINCT Id) AS Total\_Ids

FROM `capstone-project-120622.fitbit\_tracker\_data.daily\_activity\_merged`

File "/tmp/ipykernel\_19/3808155361.py", line 1

SELECT COUNT (DISTINCT Id) AS Total\_Ids

^

SyntaxError: invalid syntax

I repeated the SQL query above with each dataset (changing the FROM clause each time) and recieved these results:

* Daily\_Activity\_Merged - 33
* Daily\_Sleep\_Merged - 24
* Heart\_Rate\_Merged - 7
* Hourly\_Calorie\_Merged - 33
* Hourly\_Intensity\_Merged - 33
* Hourly\_Steps\_Merged - 33
* Weight\_Log\_Merged - 8

Both the Heart Rate and Weight Log datasets do not include enough data to move forward with analysis. These datasets will not be used.

#### **5.3 User Insights**

##### ***5.3.1 User Usage of Wearbles***

First, I wanted to see how many times each of the users wore/used the FitBit tracker:

In [ ]:

SELECT Id,

COUNT(Id) AS Total\_Id

FROM `capstone-project-120622.fitbit\_tracker\_data.daily\_activity\_merged`

GROUP BY Id

[See Google Sheets for Results](https://docs.google.com/spreadsheets/d/1KdMt0n8w6KQwT1eKY3iTDFH-_99OMCEe8XsAN57T_Mw/edit?usp=sharing)

A black background with a black square

Description automatically generated

64% of users logged data for the entire data time period (04-12-2016 to 05-12-2016). When you add in the users who only missed 1-3 days that percentage jumps up to 82% of users who logged data or wore their FitBit Tracker consistently over the month long period.

Next, I wanted to breakdown the users by how much they wore their FitBit Fitness Tracker. I created three user types:

* **Active User** - wore their tracker for 25-31 days
* **Moderate User** - wore their tracker for 15-24 days
* **Light User** - wore their tracker for 0 to 14 days

In [ ]:

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 Fitbit\_Usage\_Type

FROM `capstone-project-120622.fitbit\_tracker\_data.daily\_activity\_merged`

GROUP BY Id

A blue pie chart with a red triangle

Description automatically generated

SELECT Id,

MIN(TotalSteps) AS Min\_Total\_Steps,

MAX(TotalSteps) AS Max\_Total\_Steps,

AVG(TotalSteps) AS Avg\_Total\_Stpes,

MIN(TotalDistance) AS Min\_Total\_Distance,

MAX(TotalDistance) AS Max\_Total\_Distance,

AVG(TotalDistance) AS Avg\_Total\_Distance,

MIN(Calories) AS Min\_Total\_Calories,

MAX(Calories) AS Max\_Total\_Calories,

AVG(Calories) AS Avg\_Total\_Calories,

MIN(VeryActiveMinutes) AS Min\_Very\_Active\_Minutes,

MAX(VeryActiveMinutes) AS Max\_Very\_Active\_Minutes,

AVG(VeryActiveMinutes) AS Avg\_Very\_Active\_Minutes,

MIN(FairlyActiveMinutes) AS Min\_Fairly\_Active\_Minutes,

MAX(FairlyActiveMinutes) AS Max\_Fairly\_Active\_Minutes,

AVG(FairlyActiveMinutes) AS Avg\_Fairly\_Active\_Minutes,

MIN(LightlyActiveMinutes) AS Min\_Lightly\_Active\_Minutes,

MAX(LightlyActiveMinutes) AS Max\_Lightly\_Active\_Minutes,

AVG(LightlyActiveMinutes) AS Avg\_Lightly\_Active\_Minutes,

MIN(SedentaryMinutes) AS Min\_Sedentary\_Minutes,

MAX(SedentaryMinutes) AS Max\_Sedentary\_Minutes,

AVG(SedentaryMinutes) AS Avg\_Sedentary\_Minutes

From `capstone-project-120622.fitbit\_tracker\_data.daily\_activity\_merged`

Group BY Id

[(*See Google Sheet for Results*)](https://docs.google.com/spreadsheets/d/1HoDXTXxEErJgsgijVUktOIVn3mRK9_uLP2oHxWpajnQ/edit?usp=sharing)

Next I wanted to narrow my results to just the averages of the different types of minutes by Id.

In [ ]:

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 `capstone-project-120622.fitbit\_tracker\_data.daily\_activity\_merged`

GROUP BY Id

[(*See Google Sheet for Results*)](https://docs.google.com/spreadsheets/d/1sMwfhBMBBTXUULmnDyP6aCGFmYu_vEWNZi2tqfPBiV0/edit?usp=sharing)

These results showed that the average minutes of the Sedentary activity level was the highest for each distinct Id.

Lastly, I wanted to take a look at average active minutes by week day before moving on to user types.

Special thanks to Quy5002 for help with figuring out how to create the days of the week in Excel! Go check out Quy's case study [here](https://www.kaggle.com/code/quy5002/google-data-analytics-capstone-project-casestudy-2/notebook).

In [ ]:

SELECT Activity\_Day\_,

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 `capstone-project-120622.fitbit\_tracker\_data.daily\_activity\_merged\_weekday`

GROUP BY Activity\_Day\_

A graph with green and orange bars

Description automatically generated

Again, through this query we see that Sedentary Minutes are the highest type of active minutes. What is noticable is that there is no real difference in type of active minute total by week day. It seems users are consistent in their active minute output each day. This information could show that Bellabeat could leverage activity goals for users to meet as users might already being trying to meet personal activity goals each day and Bellabeat could enourage higher activity goals to increase daily active minutes that are very active or fairly active.

#### **5.4 User Types by Activity Levels**

The CDC recommends 150 minutes of physical activity each week. (<https://www.cdc.gov/physicalactivity/basics/adults/index.htm>)

So I wanted to see the sum of the average minutes of Very Active and Fairly Active to see if each distinct Id was hitting those CDC activity recommendations.

##### ***5.4.1 User Types by Active Minutes***

In [ ]:

SELECT Id,

avg(VeryActiveMinutes) + avg(FairlyActiveMinutes) AS Total\_Avg\_Active\_Minutes

FROM `capstone-project-120622.fitbit\_tracker\_data.daily\_activity\_merged`

GROUP BY Id

[(See Google Sheet for Results)](https://docs.google.com/spreadsheets/d/12-N970f719cAxn5BgufMX6P9zzs5w9odGZyamHgD0Qg/edit?usp=sharing)

The results showed that all users were not hitting an average of 150 minutes of active or fairly active activity level when averaged out.

However if we add in the ‘lightly active’ activity level we see that most of the users now have an average of 150 or more ‘active’ minutes which would correspond to the recommended requirements by the CDC.

In [ ]:

SELECT Id,

avg(VeryActiveMinutes) + avg(FairlyActiveMinutes) + avg(LightlyActiveMinutes) AS Total\_Avg\_Active\_Minutes,

CASE

WHEN avg(VeryActiveMinutes) + avg(FairlyActiveMinutes) + avg(LightlyActiveMinutes) >= 150 THEN 'Meets CDC Recommendation'

WHEN avg(VeryActiveMinutes) + avg(FairlyActiveMinutes) + avg(LightlyActiveMinutes) <150 THEN 'Does Not Meet CDC Recommendation'

END CDC\_Recommendations

FROM `capstone-project-120622.fitbit\_tracker\_data.daily\_activity\_merged`

GROUP BY Id

[(See Google Sheet for Results)](https://docs.google.com/spreadsheets/d/155OhHpucVncd6mU8wzr_onoU7hWpehbA7JR9aLM7oTI/edit?usp=sharing)

* Out of the 33 users, 27 met CDC Activity Length Recommendations and 6 do not on average.

I then broke this down by date – looking at one week of data -- to see if this would also match what I found across the entire date frame of the dataset. I looked at 4-17-2016 to 4-23-2016.

In [ ]:

SELECT Id,

SUM(VeryActiveMinutes) + SUM(FairlyActiveMinutes) + SUM(LightlyActiveMinutes) AS Total\_Avg\_Active\_Minutes,

CASE

WHEN SUM(VeryActiveMinutes) + SUM(FairlyActiveMinutes) + SUM(LightlyActiveMinutes) >= 150 THEN 'Meets CDC Recommendation'

WHEN SUM(VeryActiveMinutes) + SUM(FairlyActiveMinutes) + SUM(LightlyActiveMinutes) <150 THEN 'Does Not Meet CDC Recommendation'

END CDC\_Recommendations

FROM `capstone-project-120622.fitbit\_tracker\_data.daily\_activity\_merged`

WHERE ActivityDate BETWEEN '2016-04-17' AND '2016-04-23'

GROUP BY Id

Here I used SUM versus average since I was only looking at one week of tracked instances versus the whole timeframe of the study.

[(See Google Sheet for Results)](https://docs.google.com/spreadsheets/d/1RH5ENg91sJDziIA5vuY4w8lsMiX-Sb9_K2W9g2DtVDQ/edit?usp=sharing)

The results showed that more users met the CDC Activity Length Recommendations during this first week of data.

* Out of 33 users, 31 met the CDC Activity Length Recommendations, 1 did not and 1 did not have data from this time frame.

I was curious to see what the results would be if we took out the Lightly Active Minutes for the week.

In [ ]:

SELECT Id,

SUM(VeryActiveMinutes + FairlyActiveMinutes) AS Total\_Avg\_Active\_Minutes,

CASE

WHEN SUM(VeryActiveMinutes + FairlyActiveMinutes) >= 150 THEN 'Meets CDC Recommendation'

WHEN SUM(VeryActiveMinutes + FairlyActiveMinutes) <150 THEN 'Does Not Meet CDC Recommendation'

END CDC\_Recommendations

FROM `capstone-project-120622.fitbit\_tracker\_data.daily\_activity\_merged`

WHERE ActivityDate BETWEEN '2016-04-17' AND '2016-04-23'

GROUP BY Id

In this case there was a significant change to the amount of users meeting the CDC Recommendations for amount of active minutes during a week.

* Out of 33 users, 18 met the CDC Recommendations, 14 did not and 1 did not have data from this time frame.

[(See Google Sheet Results)](https://docs.google.com/spreadsheets/d/1gBW6bHzRHzuWzGvv3ady-Ta3e9cJPlR0nn4cr_mlCYQ/edit?usp=sharing)

Before I moved on to looking at steps, calories and sleep tracked, I wanted to explore this information during a week closer to the end of the dataset collection.

So I looked at the dates 2016-05-01 through 2016-05-07 as this was a week before the end of the data collection (2016-05-12, a Thursday) and also the last full week before the end date.

Query with LightlyActiveMinutes being counted:

In [ ]:

SELECT Id,

SUM(VeryActiveMinutes + FairlyActiveMinutes + LightlyActiveMinutes) AS Total\_Avg\_Active\_Minutes,

CASE

WHEN SUM(VeryActiveMinutes + FairlyActiveMinutes + LightlyActiveMinutes) >= 150 THEN 'Meets CDC Recommendation'

WHEN SUM(VeryActiveMinutes + FairlyActiveMinutes + LightlyActiveMinutes) <150 THEN 'Does Not Meet CDC Recommendation'

END CDC\_Recommendations

FROM `capstone-project-120622.fitbit\_tracker\_data.daily\_activity\_merged`

WHERE ActivityDate BETWEEN '2016-05-01' AND '2016-05-07'

GROUP BY Id

* Results show that out of 33 users, 30 met CDC Recommendations and 3 did not have data from this time frame.

[(See Google Sheets for Results)](https://docs.google.com/spreadsheets/d/1lDV0Vd1MTJQss2XhrlzFAEY1dVAaZoL1DO-6KHpkhT8/edit?usp=sharing)

Now to see the results without the LightlyActive Minutes

In [ ]:

SELECT Id,

SUM(VeryActiveMinutes + FairlyActiveMinutes) AS Total\_Sum\_Active\_Minutes,

CASE

WHEN SUM(VeryActiveMinutes + FairlyActiveMinutes) >= 150 THEN 'Meets CDC Recommendation'

WHEN SUM(VeryActiveMinutes + FairlyActiveMinutes) <150 THEN 'Does Not Meet CDC Recommendation'

END CDC\_Recommendations

FROM `capstone-project-120622.fitbit\_tracker\_data.daily\_activity\_merged`

WHERE ActivityDate BETWEEN '2016-05-01' AND '2016-05-07'

GROUP BY Id

* Results show that out of 33 users, 17 met CDC Recommendations, 13 did not & 3 did not have data from this time frame.

[(See Google Sheet for Results)](https://docs.google.com/spreadsheets/d/1JMwJH2b51ecl6mwiW02L5Iv7UAIm-WNeeZilFrXkf00/edit?usp=sharing)

After comparing both results (the results from April & the results from May), I saw no significant changes to length of activity each week. With that said, these weeks are only separated by one week and the overall data collection time frame is pretty small. While the summary of the dataset being used (FitBit Fitness Tracker Data) stated that data was collected from 2016-03-12 through 2016-05-12, there was no data tracked for most users until 2016-04-12. This indicates that there is really only 1 month of data to work with. This isn’t a significant amount of time to see if habits have changed from the beginning of using the wearable to when the data collection ended.

With that said, I’m interested to see how many steps tracked could categorize users even more.

##### ***5.4.2 User Types by Total Steps***

A Healthline.com article (“How many steps do I need a day?”) written by Sara Lindberg in 2019 cited a 2011 study by Tudor-Locke et. al. titled “How many steps/day are enough? for adults” which found that 10,000 steps/day is a reasonable target for healthy adults. Lindberg (2019) breaks down activity level by steps into three categories based off the 2011 study by Tudor-Locke et. al.:

* **Inactive:** less than 5,000 steps per day
* **Average (somewhat active):** ranges from 7,500 to 9,999 steps per day
* **Very Active:** more than 12,500 steps per day

Sources: [Healthline Article](https://www.healthline.com/health/how-many-steps-a-day#How-many-steps-for-weight-loss?) | [2011 Study](https://ijbnpa.biomedcentral.com/articles/10.1186/1479-5868-8-79#Sec3!%5Bimage.png%5D(attachment:164e6137-441b-4400-b1e1-d07f2268749b.png)

I will be using the above activity categories to create user types for the distinct Ids within the Daily Activity dataset. I’m interested to see how these categories may be broken down.

When creating my SQL query I realized that the above categories from the Healthline article left some gaps. So I created two other categories:

* **Low Active User:** 5,000 to 7,499 steps
* **Active User:** 10,000 to 12,499 steps

In [ ]:

SELECT Id,

avg(TotalSteps) AS Avg\_Total\_Steps,

CASE

WHEN avg(TotalSteps) < 5000 THEN 'Inactive'

WHEN avg(TotalSteps) BETWEEN 5000 AND 7499 THEN 'Low Active User'

WHEN avg(TotalSteps) BETWEEN 7500 AND 9999 THEN 'Average Active User'

WHEN avg(TotalSteps) BETWEEN 10000 AND 12499 THEN 'Active User'

WHEN avg(TotalSteps) >= 12500 THEN 'Very Active User'

END User\_Type

FROM `capstone-project-120622.fitbit\_tracker\_data.daily\_activity\_merged`

GROUP BY Id

A pie chart with different colored circles

Description automatically generated

[(See Google Sheets for Results)](https://docs.google.com/spreadsheets/d/1yRj4dD-e4hQnK1eRDQ_jDgeFftUTJmCCSv95WTj0ovw/edit?usp=sharing)

Here are the Results:

* **Inactive User:** 8 users
* **Low Active User:** 9 users
* **Average Active User:** 9 users
* **Active User:** 5 users
* **Very Active User:** 2 users

If we break this down between Non & Low Active users and users who are considered ‘Active’ we can see that 17 users (52%) are little to non-active and 16 (48%) users are considered active. So almost a 50/50 split on type of users.

This split is pretty close to the results earlier of looking at active minutes. If we took out the summation of the Lightly Active Minutes from the query we saw 17 users (52%) met CDC recommendations of 150 active minutes a week, 13 users (39%) did not meet CDC guidelines and 3 users (9%) did not have data from that week. After breaking down the sample into user types, I wanted to look at some more activity comparisons.

#### **5.5 Calories, Steps & Active Minutes by ID**

Now that we have more insights into our users’ activity levels, I’m interested in seeing what their logged calories in relation to their steps and active minutes can tell us. This may be easier to see as a data visualization, but I’d still like to explore this in SQL as well for the practice.

In [ ]:

SELECT Id,

Sum(TotalSteps) AS Sum\_total\_steps,

SUM(Calories) AS Sum\_Calories,

SUM(VeryActiveMinutes + FairlyActiveMinutes) AS Sum\_Active\_Minutes

FROM `capstone-project-120622.fitbit\_tracker\_data.daily\_activity\_merged`

GROUP BY Id

[(See Google Sheets for Results)](https://docs.google.com/spreadsheets/d/1PLFQZN29xgDiXdREScI-pZEjWd-98UiIBTHOvivvpNQ/edit?usp=sharing)

#### **5.6 Total Steps by Day**

Next, I wanted to take a look at average steps by day to see if users were more active on certain days of the week.

Special thanks to Quy5002 for help with figuring out how to create the days of the week in Excel! Go check out Quy's case study [here](https://www.kaggle.com/code/quy5002/google-data-analytics-capstone-project-casestudy-2/notebook).

I was super caught up on trying to do it in SQL that I didn't even think to make the data in Excel. I still want to explore how to do it in SQL as the function I found online wasn't working in BigQuery. Still more for me to learn!

In [ ]:

In Excel: use function TEXT([cell], 'ddd') // thanks again Quy!

SQL:

SELECT Activity\_Day\_,

ROUND (avg(TotalSteps), 2) AS Average\_Total\_Steps,

FROM `capstone-project-120622.fitbit\_tracker\_data.daily\_activity\_merged\_weekday`

GROUP BY Activity\_Day\_

ORDER BY Average\_Total\_Steps DESC

A graph of steps by week

Description automatically generated

After running the query, there wasn't a whole lot of difference between each day in terms of average steps. With that said, Saturday had the highest average steps as well as the beginning of each week (Monday and Tuesday). We could potentially infer from this that the users wanted to be more active right after the weekend of rest (Sunday with lowest total steps & Friday not too far behind) & that Saturday allowed for more time for activity & movement.

#### **5.7 Total Steps by Hour**

Next, I wanted to take a look at Total Steps taking by Hour to see what time of day our users were most active.

In [ ]:

SELECT

ActivityHour,

SUM(StepTotal) AS Total\_Steps\_By\_Hour

FROM `capstone-project-120622.fitbit\_tracker\_data.hourly\_steps\_merged`

GROUP BY ActivityHour

ORDER BY Total\_Steps\_By\_Hour DESC

[(see spreadsheet results)](https://docs.google.com/spreadsheets/d/1MQ2ME3-J409yJcDc0lmjSaV66CvSOgH-dG3sC7dk6uk/edit?usp=sharing)

The top 5 hours of steps recorded were:

1. 18:00:00 (6pm) – 542,848 steps
2. 19:00:00 (7pm) – 528,552 steps
3. 12:00:00 (12pm) – 505,848 steps
4. 17:00:00 (5pm) – 498,511 steps
5. 14:00:00 (2pm) – 497,813 steps

A graph of steps by hour

Description automatically generated

#### **5.8 Deeper Look Into Sleep**

Then I wanted to explore the Sleep habits of the users and how it compares to activity level.

First I looked at which date had the most minutes of sleep from all the users.

In [ ]:

SELECT

SleepDay,

SUM(TotalMinutesAsleep) AS Total\_Minutes\_Asleep

FROM `capstone-project-120622.fitbit\_tracker\_data.daily\_sleep\_merged`

WHERE SleepDay IS NOT NULL

GROUP BY SleepDay

[(See Google Sheets Results)](https://docs.google.com/spreadsheets/d/1_Z2WQiQ0QazrsQqM5ydE2DNGxNx779giiY7fyjG74zs/edit?usp=sharing)

Next I looked at average minutes slept, total steps and calories by user Id.

In [ ]:

SELECT a.Id,

avg(a.TotalSteps) AS AvgTotalSteps,

avg(a.Calories) AS AvgCalories,

avg(s.TotalMinutesAsleep) AS AvgTotalMinutesAsleep,

FROM `capstone-project-120622.fitbit\_tracker\_data.daily\_activity\_merged` AS a

INNER JOIN `capstone-project-120622.fitbit\_tracker\_data.daily\_sleep\_merged` AS s ON a.Id=s.Id

GROUP BY a.Id

[(See Google Sheets for Results)](https://docs.google.com/spreadsheets/d/1ZNuJznSVepDVpw6ZNDGjpGv3BvLRJaEgBGXStO58sHY/edit?usp=sharing)

Then I visualized average total steps against average total minutes slept to see any type of correlationA graph with blue dots

Description automatically generated

The graphs shows that most users who got at least 5 hrs. asleep had higher step counts. With this said, most users were not averaging the recommended 10,000 steps a day as noted in the Healthline article cited above.

## 6. ACT Phase

### **6.1 Conclusion & Recommendations**

Bellabeat's women-centric, holistic approach paired with smart insights and body positivity has led to the creation of wearable technology for women. These products empower women to utilize data to improve their overall health.

Since Bellabeat focuses strongly on a female audience for their products, I would recommend that the company look into using their own marketing and user data or conduct their own data collection to gain further insights and trends. I'd also recommend using a larger sample size if possible in order to increae the confidence interval. Since the data utilized in this case study did not include demographic information, I'm unable to give a more detailed recommendaion or ensure there was no sampling bias.

With that said, I was able to see some trends in the FitBit Fittness Tracker Data utilized in this case study.

#### **App Notifications**

* Through my analysis I found about 7% percent of users were 'low' or 'moderate' when it came to wearing their FitBits. While most of the users wore thier FitBits consistently, **I'd recommend a daily notification at the end of day to 'charge up' the wearable so it's ready for tracking the next day.**
* The analysis also showed that 52% of users are not reaching the recommended 10,000 steps a day. **Adding in notifications throughout the day as reminders to "get up & move" or to "complete your step goal for the day" may help increase total steps for users.**
* My analysis of the data also showed that the most steps taken were during lunch hours or the 5pm - 7pm time frame. This shows that most users have a consistant time they are increasing their steps either through a walk or potentially through a workout. **It may be beneficial to send an app notification recommending users 'start their walk' or 'start their workout' if the app notices that the user hasn't begun to do so at their usual time of increased steps.**
* Lastly, the data showed that users who averaged 5 hours of sleep or more also had a higher average step total. **It may be beneficial to help users increase their sleep time by sending them a notification to 'wind down for sleep' at a certain time based off their sleep habits.**

#### **Marketing Recommendations**

Again, while there were no demographic information to assist with any potential marketing recommendations there were some insights that might be helpful for the marketing team:

* The data showed that 93.5% of users were 'active users' meaning that they utilized their FitBit consistently for 25-31 days throughout the data collection time frame. This high level of activity indicates that this group of users are invested in utilizing their fitness tracker. **I would recommend marketing Bellabeats products to customers who may already own a tracker or are already invested in wellness or learning about their health.** Showcasing this product as woman-focused as its unique features may convience customers who already own a wearable to switch to one of Bellabeat's products for the benefit of more targeted insights.
* The data also showed (as mentioned above) that the total steps tracked were highest during the lunch time frame and the 5pm - 7pm time frame. This indicates that most users have a set routine -- usually fitting in the most activity during lunch or potentially after work. **I'd recommend that Bellabeat market their products to customers most likely living around this particular type of routine -- customers with a set job schedule and parents with a set daily routine.**