Case Study: Bellabeat Analysis

Author: Yurani Muratalla
Date: November 16, 2024
Email: yuranimuratalla1@gmail.com
Github
LinkedIn

Introduction

In this case study, I analyzed smart device fitness data that could help unlock new growth opportunities for the company Bellabeat, a high-tech manufacturer of health-focused products for women. I prepared and analyzed datasets detailing activities of 30 participants in order to answer key business questions and develop actionable insights.

About the Company

Bellabeat, a high-tech company founded in 2013, manufactures health-focused smart products and devices that empower women with knowledge about their own health and habits. Their technology development and design has helped inform and inspire women around the world, a contribution that has led them to rapid growth and quickly positioned them as a tech-driven wellness company for women. They've launched multiple products geared towards the collection of data on activity, sleep, stress, and reproductive health.

Stakeholders

- **Urška Sršen**–Co Founder and Chief Creative Officer of Bellabeat
- **Sando Mur**– Mathematician and Bellabeat's Cofounder; key member of Bellabeat executive team

Questions to Ask

- What are some trends in smart device usage?
- How could these trends apply to Bellabeat customers?
- How could these trends help influence Bellabeat's Marketing strategies?

PREPARE–Upload and Inspect Data

Data Source: Fitbit Fitness Tracker Data

- (CC0: Public Domain, dataset made available through Mobius): This Kaggle data set contains personal fitness tracker from thirty fitbit users. Thirty eligible Fitbit users consented to the submission of personal tracker data.
- I chose to work with 4 main datasets of the 18: dailyActivity_merged, sleepDay_merged, hourlyCalories_merged, and hourlySteps_merged

Limitations

The data collected consists of 30 users over a period of approximately 2 months (between 03-12-2016 and 05-12-2016) and was conducted in 2016. This dataset is outdated and might not represent the current targeted audience. The sample size is too small to conduct a proper analysis representative of a larger population. The data also does not include information on demographics for the participants, meaning perhaps the data might not entirely represent Bellabeat's target audience of women. Adding another dataset in addition to the current dataset used could address the limitations but for the purpose of this case study the current dataset will function to showcase my analysis skills and provide broad insights.

PROCESS–Clean Data

I used Excel to prepare and clean my data and the 4 datasets I used for my analysis were renamed as follows:

dailyActivity_merged = daily_activity
sleepDay_merged= daily_sleep
hourlyCalories_merged= hourly_calories
hourlySteps_merged= hourly_steps

- Data was separated into 03/12/2016-04/11/2016 and 04/12/2016-05/12/2016 folders so I appended sheets into a single spreadsheet for all datasets used
- Duplicate submissions for the same ID on same dates were deleted from daily_activity dataset
- Removed blank rows from all datasets
- Deleted logged_activities column from daily_activity dataset because there was not enough info to make it relevant to analysis
- Two datasets had conjoined date and time format so I added two additional columns (time and date). I used split texts to extract time and dates into separate columns and have a uniform MM/DD/YY among all data sets

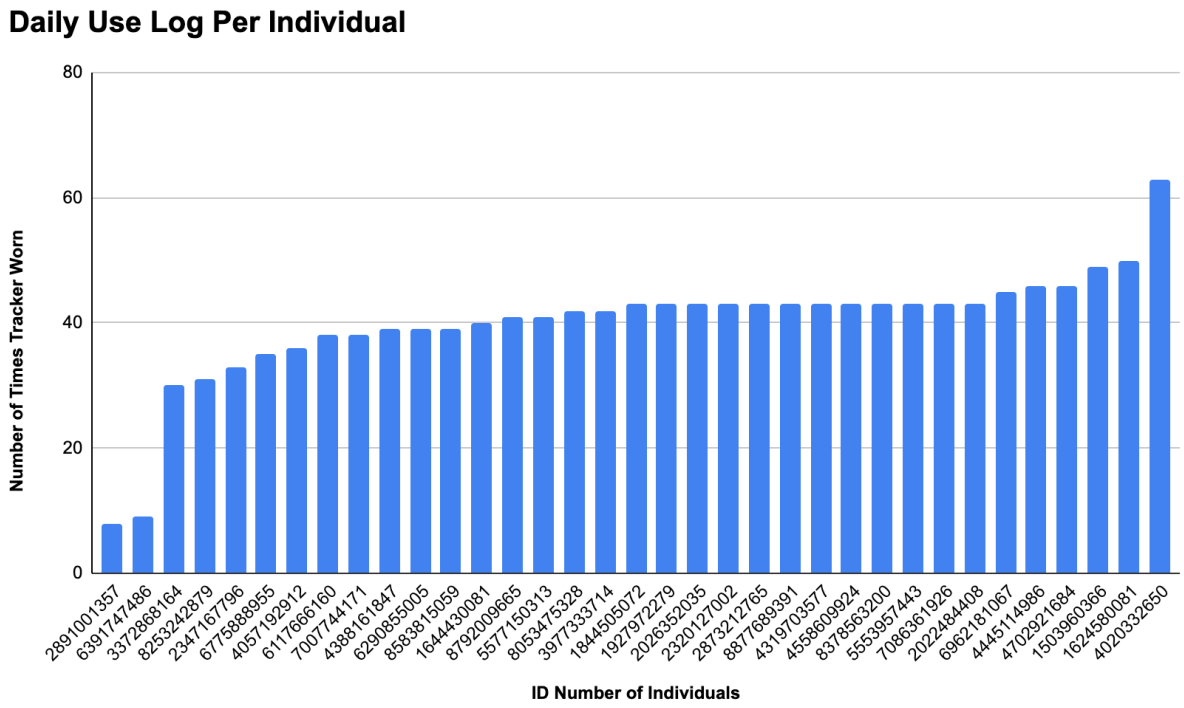
- Column “ActivityDay” was added to daily_activity data set to have a weekday representing each date of activity
- Using filters, I was able to review how many Unique IDs shared their information:
    - Sleep data: 24 individuals
    - Daily activity(including steps,calories, etc.): 34 individuals

ANALYZE AND SHARE

I uploaded my cleaned Excel datasets to BigQuery under the project name “bellabeatcasestudy-442700”

Initially, I wanted to see how often each user wore their tracker and did so by counting the amount of days each user had logged.

```
SELECT
 Id, COUNT (Id) AS Total_days_worn
FROM `bellabeatcasestudy-442700.Bellabeat.daily_activity`
GROUP BY
 Id
 ORDER BY
 Total_days_worn DESC;
```



The results showed that users wore their trackers between 8 and 62 days for the total of the two month period, with a majority wearing or logging activity often or for more than half the day.

Next, I wanted to establish if there was a particular day(s) of the week where users would log more compared to others. I wanted to see if there was a relationship between weekends and weekdays.

```
SELECT
 ActivityDay, COUNT (ActivityDay) AS Days_logged
FROM `bellabeatcasestudy-442700.Bellabeat.daily_activity`
GROUP BY
 ActivityDay
 ORDER BY
 Days_logged DESC;
```

However, results lead to activity being random and there did not seem to be a correlation between more or less activity on weekends versus weekdays.

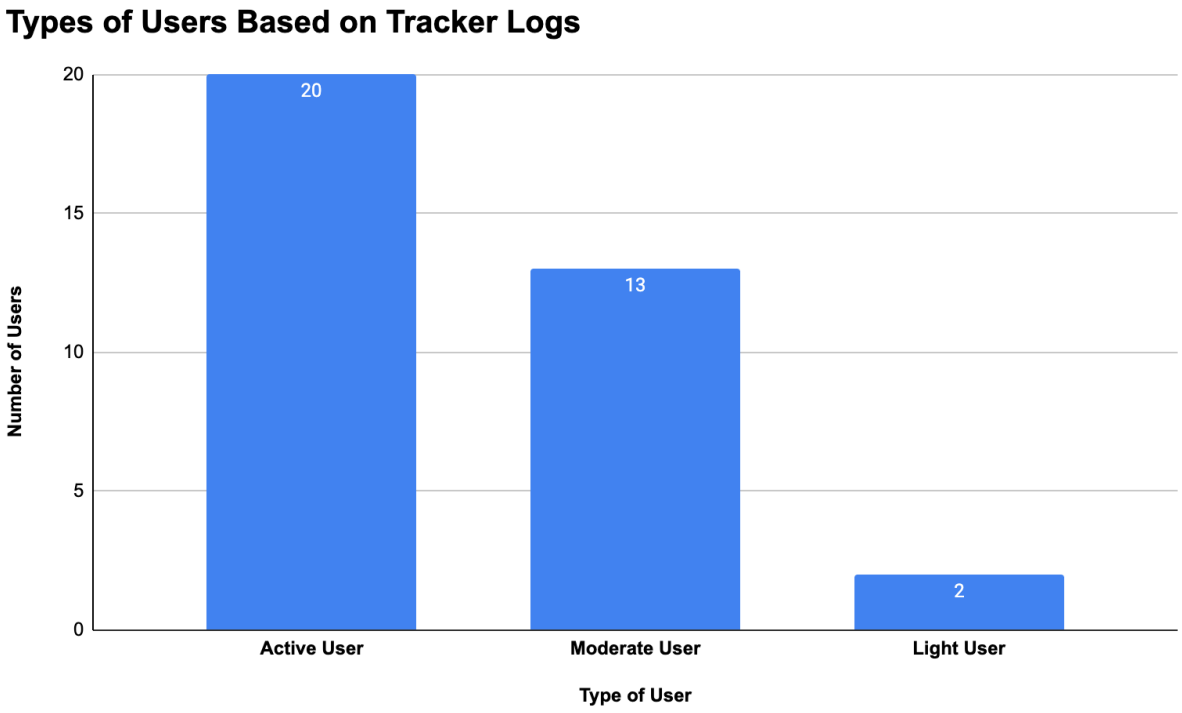
I decided to categorize users into subgroups based on their usage of their trackers and did so by grouping them as follows:

Active Users: 41+ days (Engage frequently)
Moderate Users: 21-41 days (Engage moderately)
Light: 0-20 days (Engage infrequently)

```
SELECT
Id,
 COUNT(Id) AS Total_Logged_Uses,
CASE
 WHEN COUNT (Id) BETWEEN 42 AND 63 THEN 'Active User'
 WHEN COUNT (Id) BETWEEN 21 AND 41 THEN 'Moderate User'
 WHEN COUNT (Id) BETWEEN 0 AND 20 THEN 'Light User'
END as User_Type
FROM `bellabeatcasestudy-442700.Bellabeat.daily_activity`
GROUP BY
```

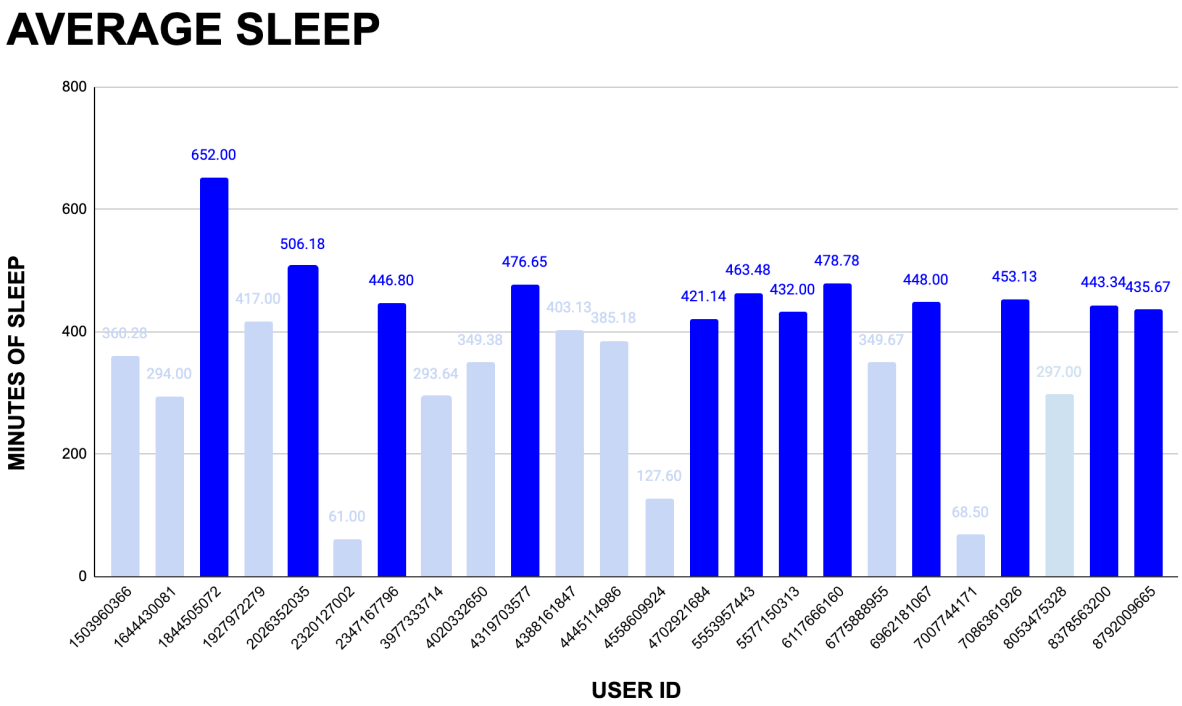
```
 Id
ORDER BY
 Total_Logged_Uses;
```

Results showed that of the 35 total individuals 20 individuals were active users, 13 individuals were moderate users, and 2 individuals were light users.



I wanted to start by examining users' sleeping trends and look at their average sleep time versus average time in bed

```
SELECT
ID,
AVG(TotalMinutesAsleep) AS Avg_sleep
FROM `bellabeatcasestudy-442700.Bellabeat.daily_sleep`
GROUP BY
ID;
```

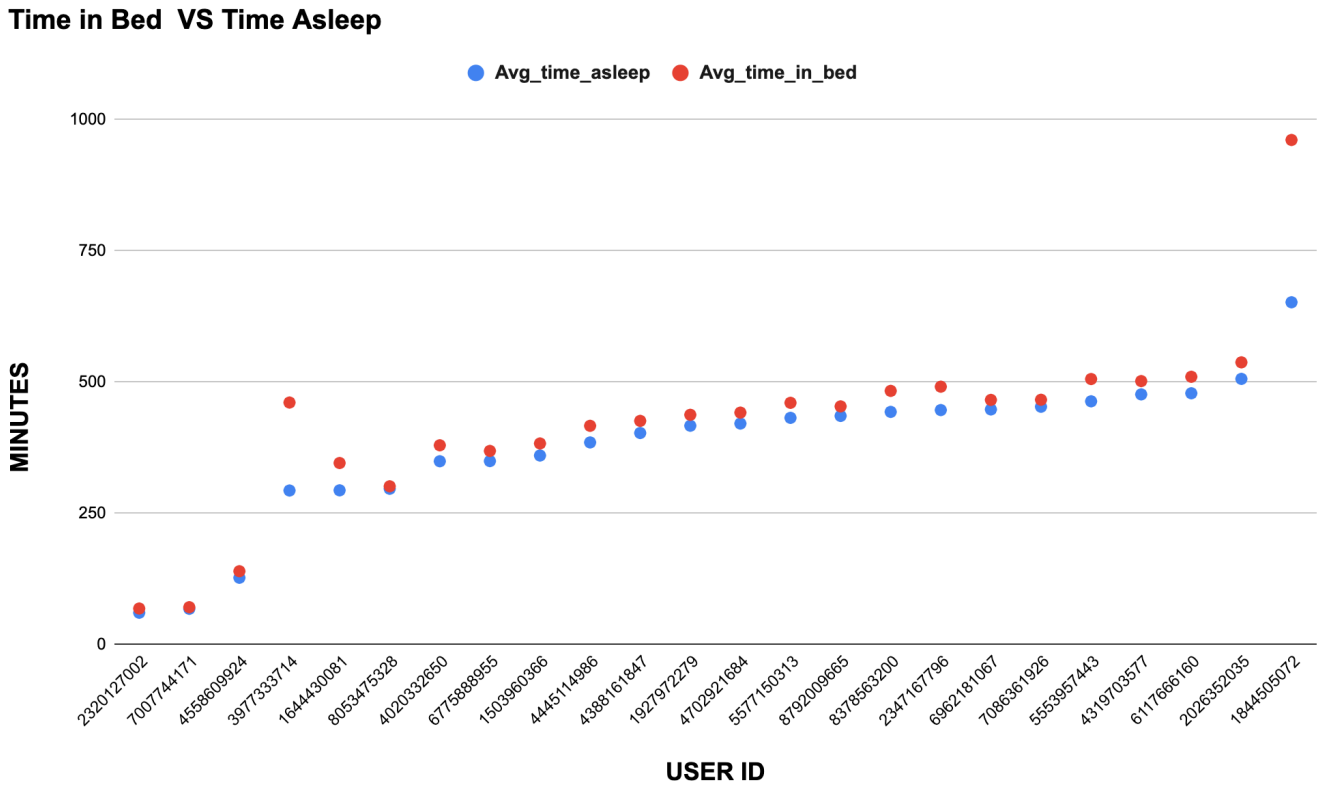


Results showed that not all users (only 24 out of 35) logged their sleep activities and only half of them were sleeping an average of at least 7 hours (420 minutes).

```
SELECT
AVG(TotalMinutesAsleep) AS Avg_time_asleep,
AVG(TotalTimeInBed) AS Avg_time_in_bed,
FROM `bellabeatcasestudy-442700.Bellabeat.daily_sleep`
```

Average time in bed = 458 minutes (7.6 hours)
Average time sleeping = 419 minutes (almost 7 hours)

- Users were spending an average of 39 minutes in bed before falling asleep



I wanted to take a look at the minimum, maximum, and average values for steps, calories, and activities so I ran the following query. After looking at the information, I decided to focus on the average steps and apply conditional formatting to show how many users were walking the recommended daily 10000+ steps (dark green), at least half of the recommended 10000 steps (light green) or less than half (white).

```
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,
```

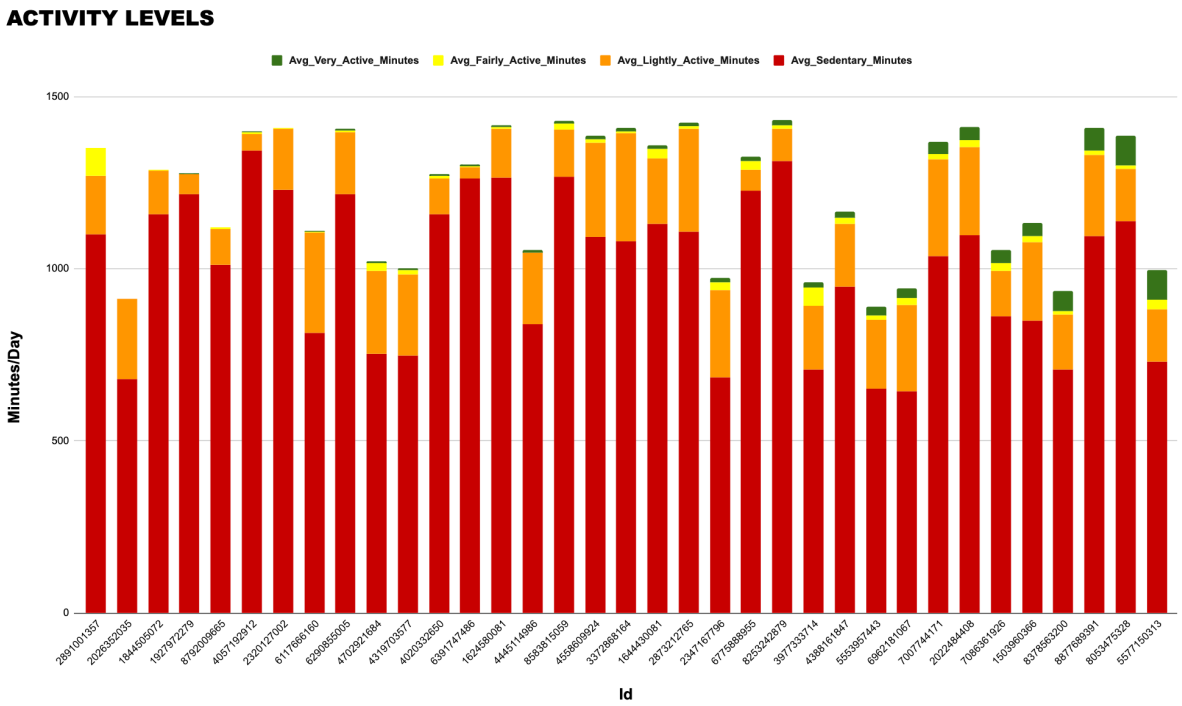
```
 MAX(Calories) AS Max_Total_Calories,
 AVG(Calories) AS Avg_Total_Calories,
 MIN(VeryActiveMinutes) AS Min_Very_Active_Minutes,
 MAX(VeryActiveMinutes) AS Max_Very_Active_Minutes,
 AVG(VeryActiveMinutes) AS Avg_Very_Active_Minutes,
 MIN(FairlyActiveMinutes) AS Min_Fairly_Active_Minutes,
 MAX(FairlyActiveMinutes) 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 `bellabeatcasestudy-442700.Bellabeat.daily_activity`
GROUP BY
 Id;
```

	A	B	C	D	E	F	G	H
	Id	Avg_Total_Steps	Avg_Total_Distance	Avg_Total_Calories	Avg_Very_Active_M	Avg_Fairly_Active_	Avg_Lightly_Active_	Avg_Sedentary_Minutes
	1624580081	5167.2	3.471000002	1433.78	5.66	3.82	141.16	1265.56
	4020332650	4049.761905	2.904920635	2736.063492	4.619047619	6.523809524	104.0793651	1159.015873
	4057192912	2103.972222	1.552777781	1911.333333	1.194444444	4.861111111	49.88888889	1342.416667
	1503960366	12174.79592	7.887755092	1844.632653	38.3877551	18.26530612	227.4489796	849.9591837
	6775888955	3301.228571	2.371999974	2284.257143	13.28571429	23.65714286	61.71428571	1227.028571
	1844505072	2876.023256	1.901627916	1585.325581	0.3023255814	1.139534884	127.4651163	1158.813953
	1927972279	1269.069767	0.8790697649	2195.465116	0.9534883721	1.023255814	58.97674419	1215.697674
	2026352035	4960.139535	3.077674417	1488.976744	0.06976744186	0.1860465116	232.2790698	680.9767442
	2320127002	4276.372093	2.890465086	1670.55814	1.23255814	2.162790698	178.0232558	1228.232558
	2873212765	7299.255814	4.925116265	1855.232558	11.58139535	6.093023256	298.8604651	1108.27907
	2891001357	773.625	0.6037500054	2273.375	0	82.5	168.75	1099.875
	3372868164	6616.933333	4.543333356	1908.833333	10.03333333	3.9	314.2333333	1080.066667
	4388161847	8595.692308	6.67153853	2829.538462	18.41025641	16.17948718	182.3076923	949
	6117666160	7363	5.575000035	2218.552632	1.289473684	2.578947368	291.3947368	814.1315789
	6290855005	4615.846154	3.49051283	2488.333333	4.051282051	5.564102564	179.2051282	1217.794872
	8253242879	4898.064516	3.510645178	1662.193548	14.83870968	10.58064516	94.58064516	1312.516129
	8877689391	16424.32558	13.45790688	3428.883721	66.27906977	11.37209302	236.4651163	1094.255814
	4445114986	4632.369565	3.135217388	2160.630435	6.108695652	1.47826087	206.326087	840.173913
	6391747486	1336.888889	1.074444466	1763.111111	5.111111111	0.6666666667	34.33333333	1261.888889
	7007744171	11619.28947	8.281052621	2570.236842	35.34210526	16.26315789	279.7105263	1038.078947
	8053475328	14784.52381	11.50166659	2932.02381	85.07142857	10.04761905	151.8571429	1139.214286
	2347167796	9647.121212	6.426363642	2033.393939	12.72727273	21.72727273	252.969697	685.6666667
	8792009665	2217	1.419024387	1994.902439	1.12195122	4.43902439	103.5853659	1011.634146
	4319703577	7422.813953	4.993720922	2025.55814	4.418604651	13.97674419	234.3255814	748.1627907
	4558609924	7154.930233	4.729767446	1976.581395	8.697674419	10.65116279	275.1627907	1091.302326
	8378563200	8555.162791	6.784651207	3414.139535	57.76744186	10.27906977	159.6744186	707.2093023
	8583815059	6346.615385	4.951282015	2662.128205	7.820512821	18.02564103	138.0512821	1266.102564
	4702921684	8367.065217	6.789130421	2918.565217	4.413043478	23.17391304	240.5	753.4782609
	6962181067	10679.88889	7.227333355	2015.377778	26.71111111	21.77777778	249.7777778	644.8888889
	5577150313	8385.926829	6.276341433	3343.707317	85.85365854	29.3902439	150.4390244	730.6585366
	5553957443	8540.627907	5.589069798	1855.255814	23.46511628	14.02325581	199.8139535	651.4418605
	3977333714	10321.52381	7.028333279	1480.642857	16.97619048	52.69047619	183.952381	707.8809524
	7086361926	8459.813953	5.747906959	2457.697674	38.02325581	23.20930233	131.9534884	861.1627907
	1644430081	7780.925	5.658750016	2837.575	10.875	26.9	190.775	1129.95
	2022484408	11595.09302	8.276744172	2500.302326	37.34883721	20.23255814	256.4418605	1097.604651
	AVG STEPS	8381.146512						

From the summary above, we see that the average steps per day “Avg_Total_Steps” is 8381 steps and an average of 2050 calories are being burned.
Only 7 users were getting an average of 10k + steps, 15 users were getting at least half of 10k steps, and the rest were not logging even half of the recommended steps.

The following query was used to focus on the Activity Levels per kind of activity, per minute, and per ID.

```
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 `bellabeatcasestudy-442700.Bellabeat.daily_activity`
GROUP BY ID;
```



The visual above represents how most users spend large quantities of time in the sedentary activity level.

```sql
SELECT
 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 `bellabeatcasestudy-442700.Bellabeat.daily_activity`
```

User averages in each category can be were further broken down as follows:

Very Active Avg= 19.69 minutes
Fairly Active Avg= 13.41 minutes
Lightly Active Avg= 185.49 minutes
Sedentary Active Avg= 993.23 minutes

Users were spending an average of 16.55 hours in sedentary activity level.

ACT

My findings and trends:

- Not all users were consistent with logging their sleep with only about 68% tracking it
- Half of users sleep on average less than 7 hours per night
- A majority of users spend an average of 30+ minutes in bed before falling asleep
- On average, not enough users are walking the recommended 10k steps per day, only reaching about 8k on average
- Users spend a majority of their time in a sedentary activity level over the light, moderate, and active categories

High-Level Recommendations

1. Bellabeat could introduce challenges, leaderboards, and other incentives to make tracking more engaging and fun in order to encourage more daily activity and counter the high count of sedentary activity levels. Rewards or discounts on other products or services and features could be provided for meeting activity goals.
2. Bellabeat could integrate sending reminders or prompts throughout the day when daily activity tracked has been inactive for a while to encourage movement or activity.
3. Bellabeat could show its users their trends in their sleep and wake times, helping them spot habits that might be affecting their sleep such as screen-time. The company could also incorporate bedtime reminders in advance to remind users to start unwinding for the night.