



# Case Study: How Can a Wellness Technology Company Play It Smart?

This project was completed as a part of the Google Data Analytics Capstone Course offered through Coursera.

## Business Task:

Bellabeat, a high-tech manufacturer of health-focused products for women collects data on activity, sleep, stress, and reproductive health to empower women with knowledge about their own health and habits. As a Junior Data Analyst working on the marketing Analyst Team at Bellabeat I was asked to focus on one of Bellabeat's products and analyze smart device data to gain insight into how consumers are using them

The products offered by the company are:

1. Bellabeat app: The Bellabeat app provides users with health data related to their activity, sleep, stress, menstrual cycle, and mindfulness habits. This data can help users better understand their current habits and make healthy decisions. The Bellabeat app connects to their line of smart wellness products.
2. Leaf: Bellabeat's classic wellness tracker can be worn as a bracelet, necklace, or clip. The Leaf tracker connects to the Bellabeat app to track activity, sleep, and stress.
3. Time: This wellness watch combines the timeless look of a classic timepiece with smart technology to track user activity, sleep, and stress. The Time watch connects to the Bellabeat app to provide you with insights into your daily wellness.
4. Spring: This is a water bottle that tracks daily water intake using smart technology to ensure that you are appropriately hydrated throughout the day. The Spring bottle connects to the Bellabeat app to track your hydration levels.
5. Bellabeat membership: Bellabeat also offers a subscription-based membership program for users. Membership gives users 24/7 access to fully personalized guidance on nutrition, activity, sleep, health and beauty, and mindfulness based on their lifestyle and goals.

Although the company has invested in several advertising Medias, an analysis of Bellabeat's available consumer data would reveal more opportunities for growth.

### Questions for 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?

### Preparing Data:

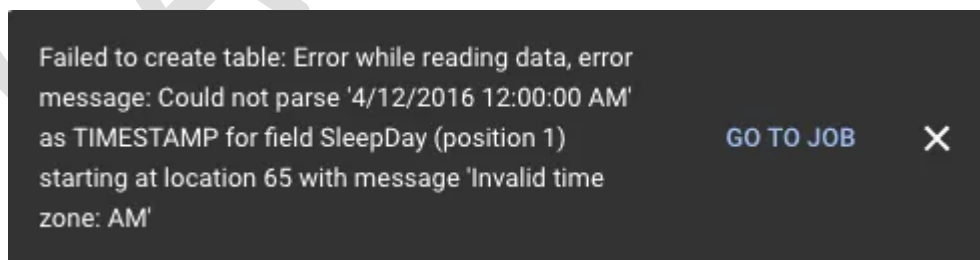
FitBit Fitness Tracker Data is a data set that contains personal fitness tracker from thirty fitbit users. Thirty eligible Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring. It includes information about daily activity, steps, and heart rate that can be used to explore users' habits. The data set had 18 CSV files and my focus was on weightLogInfo\_merged, sleepDay\_merged, dailyActivity\_merged, hourlyCalories\_merged, hourlySteps\_merged. The reasons for choosing these files was understand how active hours/exercise correlated to sleep as well as the member's BMI. The data is very small hence cannot be enough to represent a year's worth of trend of activity.

### Data Processing:

I used BigQuery as my tool to process data. The plan was to process large files of data and then use tableau to create trends for data visualization.

I created a bellabeat data set. Within the data set I uploaded five CSV files and renamed them to make them more consistent and clear. For example: the weightLogInfo\_merged was changed to weight only.

While importing the weightLogInfo\_merged file and sleepDay\_merged I encountered the following error:



*Figure 1 Error for Invalid Time Zone*

To remove this error I opened the excel files and formatted the date columns to Date-Time format. I saved the edited files as new files so the original data was not tempered. I then uploaded the files again on BigQuery.

## Data Analysis:

Now that I had all my datasets imported, I used the scheme to understand the data better and to ensure there was no issue with naming.

## Daily Activity:

<input type="checkbox"/>	Field name	Type	Mode	Key	Collation	Default Value	Policy Tags	Description
<input type="checkbox"/>	<a href="#">Id</a>	INTEGER	NULLABLE					
<input type="checkbox"/>	<a href="#">ActivityDate</a>	DATE	NULLABLE					
<input type="checkbox"/>	<a href="#">TotalSteps</a>	INTEGER	NULLABLE					
<input type="checkbox"/>	<a href="#">TotalDistance</a>	FLOAT	NULLABLE					
<input type="checkbox"/>	<a href="#">TrackerDistance</a>	FLOAT	NULLABLE					
<input type="checkbox"/>	<a href="#">LoggedActivitiesDistance</a>	FLOAT	NULLABLE					
<input type="checkbox"/>	<a href="#">VeryActiveDistance</a>	FLOAT	NULLABLE					
<input type="checkbox"/>	<a href="#">ModeratelyActiveDistance</a>	FLOAT	NULLABLE					
<input type="checkbox"/>	<a href="#">LightActiveDistance</a>	FLOAT	NULLABLE					
<input type="checkbox"/>	<a href="#">SedentaryActiveDistance</a>	FLOAT	NULLABLE					
<input type="checkbox"/>	<a href="#">VeryActiveMinutes</a>	INTEGER	NULLABLE					
<input type="checkbox"/>	<a href="#">FairlyActiveMinutes</a>	INTEGER	NULLABLE					
<input type="checkbox"/>	<a href="#">LightlyActiveMinutes</a>	INTEGER	NULLABLE					
<input type="checkbox"/>	<a href="#">SedentaryMinutes</a>	INTEGER	NULLABLE					
<input type="checkbox"/>	<a href="#">Calories</a>	INTEGER	NULLABLE					

Figure 2 Daily activity dataset

## Calories:

<input type="checkbox"/>	Field name	Type	Mode	Key	Collation	Default Value	Policy Tags	Description
<input type="checkbox"/>	<a href="#">Id</a>	INTEGER	NULLABLE					
<input type="checkbox"/>	<a href="#">ActivityDay</a>	DATE	NULLABLE					
<input type="checkbox"/>	<a href="#">Calories</a>	INTEGER	NULLABLE					

[EDIT SCHEMA](#) [VIEW ROW ACCESS POLICIES](#)

Figure 3 Calories dataset

## Weight:

weight

QUERY

SHARE

COPY

SNAPSHOT

DELETE

EXPORT

REFRESH

SCHEMA

DETAILS

PREVIEW

LINEAGE

Filter

Enter property name or value

<input type="checkbox"/>	Field name	Type	Mode	Key	Collation	Default Value	Policy Tags	Description
<input type="checkbox"/>	<a href="#">Id</a>	INTEGER	NULLABLE					
<input type="checkbox"/>	<a href="#">Date</a>	STRING	NULLABLE					
<input type="checkbox"/>	<a href="#">WeightKg</a>	FLOAT	NULLABLE					
<input type="checkbox"/>	<a href="#">WeightPounds</a>	FLOAT	NULLABLE					
<input type="checkbox"/>	<a href="#">Fat</a>	INTEGER	NULLABLE					
<input type="checkbox"/>	<a href="#">BMI</a>	FLOAT	NULLABLE					
<input type="checkbox"/>	<a href="#">IsManualReport</a>	BOOLEAN	NULLABLE					
<input type="checkbox"/>	<a href="#">LogId</a>	FLOAT	NULLABLE					

EDIT SCHEMA

VIEW ROW ACCESS POLICIES

Figure 4 Weight dataset

## Hourly Steps:

hourly\_steps

QUERY

SHARE

COPY

SCHEMA

DETAILS

PREVIEW

LINEAGE

Filter

Enter property name or value

<div></div>	Field name	Type	Mode	Collation	Default Value
<div></div>	<a href="#">Id</a>	INTEGER	NULLABLE		
<div></div>	<a href="#">ActivityHour</a>	TIMESTAMP	NULLABLE		
<div></div>	<a href="#">StepTotal</a>	INTEGER	NULLABLE		
<div></div>	<a href="#">Date</a>	DATE	NULLABLE		
<div></div>	<a href="#">Hour</a>	TIME	NULLABLE		

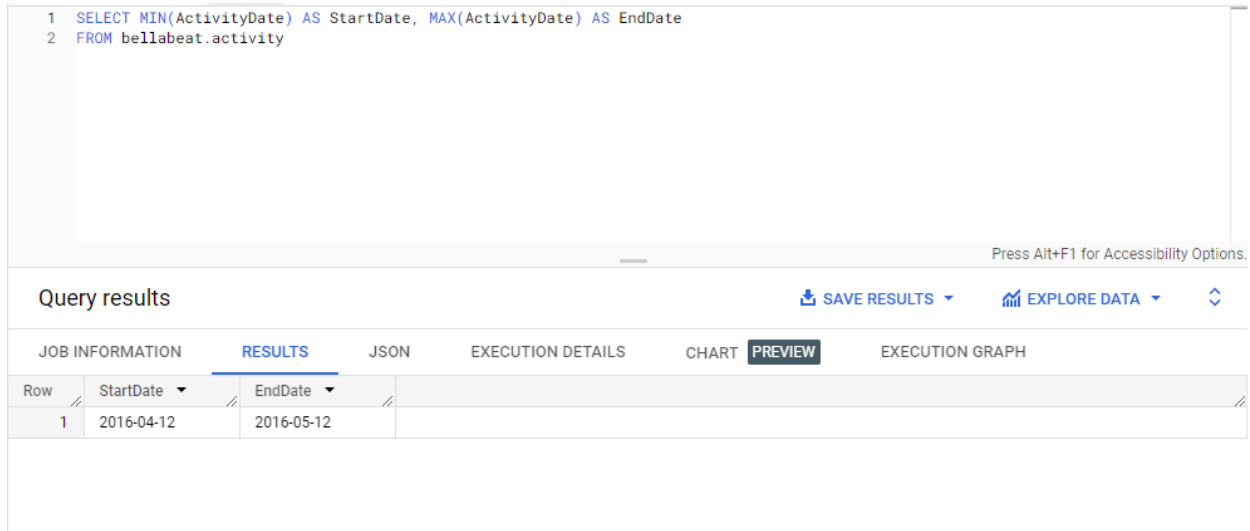
EDIT SCHEMA

VIEW ROW ACCESS POLICIES

Figure 5 Hourly steps dataset

All the columns were in PascalCase, generally snake\_case is preferred but I continued to work with the original formatting.

Using the following query for activity I found out that the data provided was only of a month.



```
1 SELECT MIN(ActivityDate) AS StartDate, MAX(ActivityDate) AS EndDate
2 FROM bellabeat.activity
```

Query results

JOB INFORMATION RESULTS JSON EXECUTION DETAILS CHART PREVIEW EXECUTION GRAPH

Row	StartDate	EndDate
1	2016-04-12	2016-05-12

Figure 6 Min and Max Date

The next thing I noticed was that the data as not consistent. Some data sets did not have consistent values. The weight dataset has 67 entries from only 8 members while the sleep dataset has 413 entries, with varying amounts of from 24 total members. These values were obtained using the following queries:

```
1 -- Members who recored their weight --
2 SELECT COUNT(DISTINCT Id) AS Members
3 FROM bellabeat.weight;
4
5 -- Members who recorded their calories --
6 SELECT COUNT(DISTINCT Id) AS Members
7 FROM bellabeat.calories;
8
9 --Number of weight enteries by each memebr --
10 SELECT DISTINCT(COUNT(Id)) AS EntryCount, Id
11 FROM bellabeat.weight
12 GROUP BY Id
13
```

Figure 7 Evaluating data

Row	EntryCount	Id
1	5	4558609924
2	2	4319703577
3	2	1503960366
4	24	8877689391
5	2	2873212765
6	1	5577150313
7	1	1927972279
8	30	6962181067

Figure 8 Evaluation results

I had to ensure there was no duplicates in the datasets. So I used the following queries to look for any:

```

1 SELECT Id, ActivityDate, COUNT(*)
2 FROM bellabeat.activity
3 GROUP BY Id, ActivityDate
4 HAVING COUNT(*) > 1;
5
6 SELECT Id, ActivityDay, COUNT(*)
7 FROM bellabeat.calories
8 GROUP BY Id, ActivityDay
9 HAVING COUNT(*) > 1;
10
11 SELECT Id, Date, COUNT(*)
12 FROM bellabeat.weight
13 GROUP BY Id, Date
14 HAVING COUNT(*) > 1;

```

Figure 9 Identifying duplicates

```

1 SELECT Id, ActivityHour, StepTotal, COUNT(*)
2 FROM bellabeat.hourly_steps
3 GROUP BY Id, ActivityHour, StepTotal
4 HAVING COUNT(*) > 1;
5
6 SELECT Id, SleepDay, COUNT(*) AS Count
7 FROM bellabeat.sleep
8 GROUP BY Id, SleepDay
9 HAVING COUNT(*) > 1;

```

Figure 10 Identifying duplicates

Row	Id	SleepDay	Count
1	4388161847	05-05-16	2
2	4702921684	05-07-16	2
3	8378563200	4/25/2016 12:00:00 AM	2

Figure 11 Duplicates found in sleep dataset

The dataset Sleep had 3 duplicates, so I used the following query for the sleep dataset. This enabled me to pull all the entries except the duplicates which was saved as a table in BigQuery called sleep\_clean. The number of entries dropped to 410 which ensured there were no duplicates left

```

1 WITH sleep_clean
2 AS (
3   SELECT Id, SleepDay, TotalSleepRecords, TotalMinutesAsleep, TotalTimeInBed,
4     ROW_NUMBER() OVER (PARTITION BY Id, SleepDay
5       ORDER BY Id) AS Duplicate
6   FROM bellabeat.sleep
7 )
8
9 SELECT *
10 FROM sleep_clean
11 WHERE Duplicate = 1

```

Press Alt+F1 for Accessibility Options

Figure 12 Query for excluding duplicates

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS	CHART	PREVIEW	EXECUTION GRAPH
Row	Id	SleepDay	TotalSleepRecords	TotalMinutesAsleep	TotalTimeInBed	Duplicate	
1	6117666160	05-05-16	1	392	415	1	
2	8378563200	4/29/2016 12:00:00 AM	1	527	562	1	
3	1503960366	05-03-16	1	273	296	1	
4	1927972279	04-12-16	3	750	775	1	

Results per page: 50 1 - 50 of 410

Figure 13 Results without duplicates

Once all the duplicates were removed, my data was consistent for further transformation.

For the sleep\_clean dataset I did wanted to know the time the members took to fall asleep. I also had to ensure that the results did have values of 0.

```

1 SELECT
2   *,
3   (ROUND(TotalMinutesAsleep/60,1)) AS HoursAsleep,
4   (TotalTimeInBed - TotalMinutesAsleep) AS MinutesAwake
5 FROM bellabeat.sleep_clean
6 WHERE
7   (TotalTimeInBed - TotalMinutesAsleep) > 0;

```

Figure 14 Formatting sleep dataset

Row	ipDay	TotalSleepRecords	TotalMinutesAsleep	TotalTimeInBed	Duplicate	HoursAsleep	MinutesAwake
1	3/2016 12:00:00 AM	1	486	493	1	8.1	7
2	17-16	1	74	75	1	1.2	1
3	3/2016 12:00:00 AM	1	331	337	1	5.5	6
4	12-16	1	796	961	1	13.3	165
5	3/2016 12:00:00 AM	1	124	142	1	2.1	18
6	3/2016 12:00:00 AM	1	119	127	1	2.0	8
7	18-16	1	137	154	1	2.3	17
8	3/2016 12:00:00 AM	1	171	179	1	2.9	8
9	1/2016 12:00:00 AM	1	126	137	1	2.1	11

Figure 15 Results after formatting

For activity I created columns for sedentary hours and total active hours as my focus was not on the specific types but the general information. I also ensured that the data did not have values of 0.



```

1 SELECT
2   Id, ActivityDate, TotalSteps, TotalDistance,
3   (ROUND(SedentaryMinutes/60,1)) AS SedentaryHours,
4   (ROUND((VeryActiveMinutes + FairlyActiveMinutes + LightlyActiveMinutes)/60,1)) AS TotalActiveHours, Calories
5 FROM
6   bellabeat.activity
7 WHERE
8   (VeryActiveMinutes + FairlyActiveMinutes + LightlyActiveMinutes) > 0 AND Calories > 0;

```

Figure 16 Formatting activity dataset

Row	Id	ActivityDate	TotalSteps	TotalDistance	SedentaryHours	TotalActiveHours	Calories
1	1624580081	2016-05-01	36019	28.03000068664...	17.0	7.0	2690
2	1644430081	2016-04-14	11037	8.020000457763...	18.8	5.3	3226
3	1644430081	2016-04-19	11256	8.180000305175...	18.3	5.7	3300
4	1644430081	2016-04-28	9405	6.840000152587...	19.3	4.7	3108
5	1644430081	2016-04-30	18213	13.23999977111...	13.6	8.0	3846
6	1644430081	2016-05-03	12850	9.340000152587...	18.6	5.4	3324
7	2022484408	2016-04-20	15112	10.67000007629...	17.6	6.5	2897
8	2022484408	2016-05-09	13379	9.390000343322...	17.7	6.3	2709
9	2347167796	2016-04-14	10129	6.699999809265...	11.8	4.3	2010

Results per page: 50 1 - 50 of 857

Figure 17 Results after formatting

Once all the data was formatted I merged the sleep and activity data into one dataset.

```

1 SELECT
2   activity.Id,
3   activity.ActivityDate,
4   activity.TotalSteps,
5   activity.SedentaryHours,
6   activity.TotalActiveHours,
7   sleep.SleepDay,
8   sleep.HoursAsleep, |
9   sleep.MinutesAwake
10  FROM bellabeat.activity AS activity
11  INNER JOIN bellabeat.sleep AS sleep
12  ON activity.ActivityDate = sleep.SleepDay AND activity.Id = sleep.Id
13  ORDER BY activity.TotalActiveHours DESC;

```

Figure 18 Merging sleep and activity datasets

For the weight data set I removed all the columns that were not required and created a Weight Status column. I also ensured that there were not null cells.

```

1 SELECT
2   * EXCEPT(Fat,WeightPounds,LogId,IsManualReport),
3   CASE
4     WHEN BMI < 18.5 THEN "Underweight"
5     WHEN BMI < 25.0 THEN "Healthy"
6     WHEN BMI < 30.0 THEN "Overweight"
7     ELSE "Obese" END AS WeightStatus
8 FROM bellabeat.weight
9 WHERE BMI > 0;

```

Figure 19 Query for weight status

Row	Id	Date	WeightKg	BMI	WeightStatus
1	4558609924	05-02-16	69.19999695	27.04000092	Overweight
2	4558609924	4/18/2016 11:59:59 PM	69.69999695	27.25	Overweight
3	4558609924	05-09-16	69.09999847	27.0	Overweight
4	4558609924	05-01-16	69.90000153	27.31999969	Overweight
5	4558609924	4/25/2016 11:59:59 PM	70.30000305	27.45999908	Overweight
6	4319703577	4/17/2016 11:59:59 PM	72.40000153	27.45000076	Overweight
7	4319703577	05-04-16	72.30000305	27.37999916	Overweight
8	1503960366	05-02-16	52.59999847	22.64999962	Healthy
9	1503960366	05-03-16	52.59999847	22.64999962	Healthy
10	8877689391	05-12-16	84.0	25.13999939	Overweight

Figure 20 Results after formatting

Lastly I created columns with the total and average steps and calories grouped by the hour. This will later enable me to have the data organized so that I could analyze hourly trends.

```

1 -- Creating temporary tables --
2 CREATE TEMP TABLE avg_hourly_calories
3 AS
4   SELECT ActivityHour, SUM(Calories) AS TotalCalories, ROUND(AVG(Calories),0) AS AvgCalories
5   FROM bellabeat.hourly_calories
6   WHERE Calories > 0
7   GROUP BY ActivityHour
8   ORDER BY ActivityHour;
9
10 CREATE TEMP TABLE avg_hourly_steps
11 AS
12   SELECT ActivityHour, SUM(StepTotal) AS TotalSteps, ROUND(AVG(StepTotal),0) AS AvgSteps
13   FROM bellabeat.hourly_steps
14   WHERE StepTotal > 0
15   GROUP BY ActivityHour
16   ORDER BY ActivityHour;
17
18 -- Joining the temp tables --
19
20 SELECT steps.ActivityHour, steps.TotalSteps, steps.AvgSteps, calories.TotalCalories, calories.AvgCalories
21 FROM avg_hourly_steps AS steps
22 FULL OUTER JOIN avg_hourly_calories AS calories
23 ON steps.ActivityHour = calories.ActivityHour
24 ORDER BY steps.ActivityHour;

```

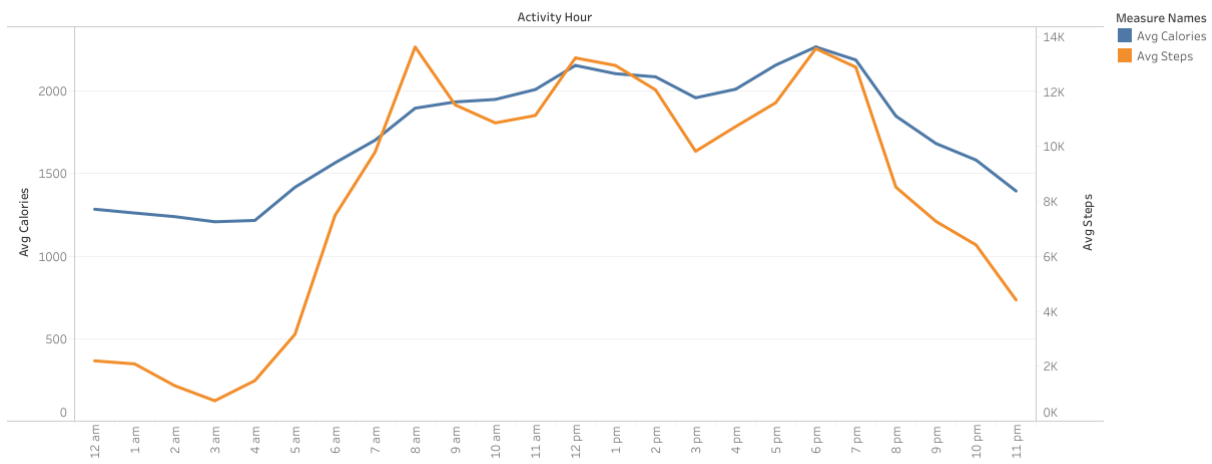
Figure 21 Temporary table

My data was completely transformed. I exported all the finalized/cleaned data files and exported them to my computer. I then used Tableau for data visualization and observe the trends.

## Data Visualization:

In Tableau Public I uploaded my datasets and created multiple visualizations. I began looking at hourly calories and steps to determine what time of day users are most and least active. The most calories and steps taken were during 5–7 PM. There was unsurprisingly less activity between 9 PM and 5 AM; however, during the day there was a decrease in both steps and calories at 3 PM as well.

Average Calories and Steps per hour



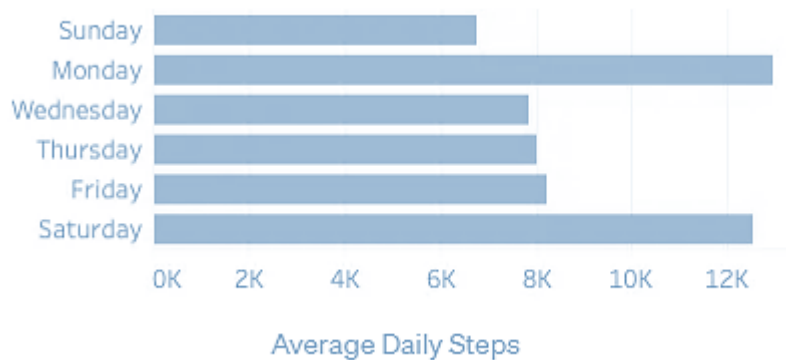
The trends of Avg Calories and Avg Steps for Activity Hour Hour. Colour shows details about Avg Calories and Avg Steps. The view is filtered on Activity Hour Hour, which excludes Null.

Figure 22

Walking 10,000 steps daily, on average, was correlated with a healthy BMI. Additionally, more sedentary hours in a day was correlated with an Overweight or Obese BMI. Healthy and Overweight BMIs correlated to 5 active hours, on average, which suggests the intensity of the activity during the active hours is important.

BMI Classification	Avg. Steps	Avg. Active Hours	Avg. Sedentary Hours
Healthy	10,095	5	11
Overweight	7,830	5	18
Obese	356	1	16

Figure 23



Monday and Saturday had the highest average daily steps with both above 12,000 steps. Sunday had the lowest daily steps with 6,746 steps on average.

Figure 24

## Conclusion:

### **High Level Recommendations for Bellabeat's Marketing Strategy**

As mentioned above, the ultimate business task was to help guide marketing strategy for Bellabeat. Below are my recommendations based on the findings discovered through analyzing Fitbit data:

1. Users are most active in the evening following the end of work. Bellabeat could host live events, in person and virtual, targeted for those hours to encourage the highest amount of participation. This could be linked to the Bellabeat membership to increase subscriptions.
2. Though encouraging 10,000 steps daily has been widely used it is relevant and correlated to a healthier BMI. The Bellabeat app should have a notification and markers to encourage users to walk 10,000 steps daily. A point system could be created to help encourage users which could be used to attend the live events or get discounts on other products.
3. Sunday was found to be a day with little activity and steps as well as less sleep too. Bellabeat should make a point to increase reminders for users to make Sunday a self-care day. Mindfulness and sleep should be emphasized on Sunday to prepare for the week ahead.
4. The Bellabeat app must encourage more sleep the night of more active days to ensure recovery. The app could be set to send informative articles on recovery and the importance of sleep after particularly active days.