Bellabeat Case Study

Table of Contents

Scenario
For this case study, I am working as a Junior Data Analyst on the Marketing Analytics Team at Bellabeat.

Bellabeat is a high-tech manufacturer of health-focused products for women. Since its foundation, Bellabeat has been growing rapidly and has the potential of becoming a major player in the global smart device market. The current marketing strategy focuses on digital marketing, but they also invest on a smaller scale in traditional media.

The Chief Creative Officer believes that analyzing consumer data would reveal more opportunities for growth. The Marketing Analytics Team has been asked to focus on one of the products and analyze smart device data to gain insight into how consumers are using their smart devices. These insights would later be used to inform marketing strategies. The product we are focusing on is the Ivy Health Tracker.

Ask Phase
Questions for analysis:

a. What are some trends in smart device usage?
b. How do these trends apply to Bellabeat customers?
c. How could these trends help influence Bellabeat's marketing strategy?

The management and stakeholders are:

- **Urška Sršen:** Co-founder and Chief Creative Officer
- **Sando Mur:** Co-founder and mathematician – key member of executive team
- **Bellabeat marketing analytics team:** analysts responsible for collecting, analyzing, and reporting data that helps guide the marketing strategy.

* The stakeholders expect to receive a presentation of the analysis with high-level recommendations for the marketing strategy, based on insights gained by analyzing how consumers use non-Bellabeat smart devices.

Business Task
Understand how consumers are using their smart devices based on usage data to discover opportunities for growth and inform and guide marketing strategies.

Prepare Phase

Description of the data

The CCO suggested the use of the following open-source dataset: **Fitbit Fitness Tracker Data.**

This dataset can be found in Kaggle and is available through Mobius. It consists of 33 distinct participants (Fitbit Users) and was generated by the submission of personal tracker data through a Survey via Amazon Mechanical Turk. The data encompasses daily data (steps, distance, intensities, calories burned, and sleep time) and hourly data (calories burned, heart rate, steps), from April 11, 2016, to May 12, 2016.

The entire dataset contains 18 CSV documents, of which the following 8 were used in this analysis:

- Daily Activity
- Daily Calories
- Daily Intensities
- Daily Steps
- Daily Sleep
- Hourly Calories
- Hourly Intensities
- Hourly Steps

Data considerations

The sample size is small and might not be representative of the population. Additionally, it is not specified how the participants were chosen, which could imply sampling bias. Since it lacks any demographic information, we cannot assume that it accurately reflects Bellabeat customers' behavior. Furthermore, the data is outdated (from 2016), and encompasses only one month.

Process Phase

Tools

The Data was analyzed using SQL in BigQuery.

Data cleansing

The process of cleaning the data included checking for duplicates, verifying number of users and dates, changing data types, separating columns, and renaming columns. Below are some examples of this process:

- Changing data type and format of columns containing date and time.

```
-- separating ActivityHour in 2 columns (date and time)
SELECT
  Id
  ,DATE(PARSE_TIMESTAMP('%m/%d/%Y %I:%M:%S %p', ActivityHour)) AS ActivityDate
  ,TIME(PARSE_TIMESTAMP('%m/%d/%Y %I:%M:%S %p', ActivityHour)) AS ActivityTime
  ,calories
FROM
  `fitbit.calories_hour`
```

- Verifying number of users in each table.

```
-- Verify amount of participants in each (daily) table
SELECT
  'calories' AS table_name, COUNT(DISTINCT Id) AS total_ids
FROM
  `belladata.calories_day`
UNION ALL
SELECT
  'intensities' AS table_name, COUNT(DISTINCT Id) AS total_ids
FROM
  `belladata.intensities_day`
UNION ALL
SELECT
  'sleep' AS table_name, COUNT(DISTINCT Id) AS total_ids
FROM
  `belladata.sleep_day`
UNION ALL
SELECT
  'steps' AS table_name, COUNT(DISTINCT Id) AS total_ids
FROM
  `belladata.steps_day`
```

- Verifying dates.

```sql
-- Verify first and last dates
SELECT
  MIN(ActivityDate) AS min_date,
  MAX(ActivityDate) AS max_date
FROM
  `belladata.calories_hour`
```

Analyze Phase

We started by examining how many users were active each day.

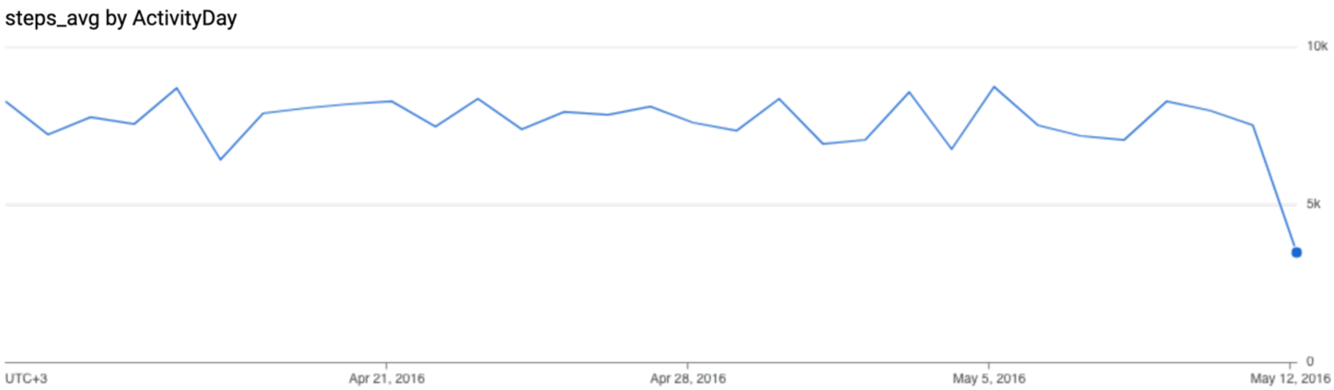
```
-- How many users used the device each day
WITH active_users AS (
  SELECT
    ActivityDate
    ,COUNT(DISTINCT Id) AS active_users
  FROM
    `belladata.activity_day`
  GROUP BY
    ActivityDate
  ORDER BY
    ActivityDate
  ),
total_count AS (
  SELECT
    COUNT(DISTINCT Id) AS total_users
  FROM
    `belladata.activity_day`
  )
SELECT
  users.ActivityDate
  ,users.active_users
  ,ROUND((users.active_users / total.total_users) * 100, 2) AS users_pcnt
FROM
  active_users users
CROSS JOIN
  total_count total
ORDER BY
  users.ActivityDate
```

This shows that, initially, 100% of the participants used the device, but this percentage gradually decreased throughout the month. By the end of the month, only 64% of the participants were using their devices. This indicated that not all the users utilize their devices consistently: after three weeks, more than 10% did not use their device consistently, and after one month, this percentage increased to nearly 35%.

Steps Function

Looking into the trends and averages in the steps data, it is evident that the average number of steps remains relatively consistent throughout the month, with a notable decrease just in the last day.

```sql
-- average steps per day (date)
SELECT
  ActivityDay,
  ROUND(AVG(StepTotal), 2) AS steps_avg
FROM
  `belladata.steps_day`
GROUP BY
  ActivityDay
ORDER BY
  ActivityDay
```



Analyzing the steps per day of the week, we observe that Saturday and Tuesday register the highest number of steps, while Sunday records the fewer number of steps.

```sql
-- average steps per weekday
WITH steps_average AS (
  SELECT
  EXTRACT(DAYOFWEEK FROM ActivityDay) AS day_num
  ,FORMAT_DATE('%A', ActivityDay) AS day_of_week
  ,ROUND(AVG(StepTotal), 2) AS step_avg
FROM
  `belladata.steps_day`
GROUP BY
  day_of_week
  ,day_num
ORDER BY
  step_avg DESC
  )
SELECT
--  day_num
  steps_average.day_of_week
  ,steps_average.step_avg
  ,RANK() OVER(ORDER BY step_avg DESC) AS step_rank
FROM
  steps_average
```

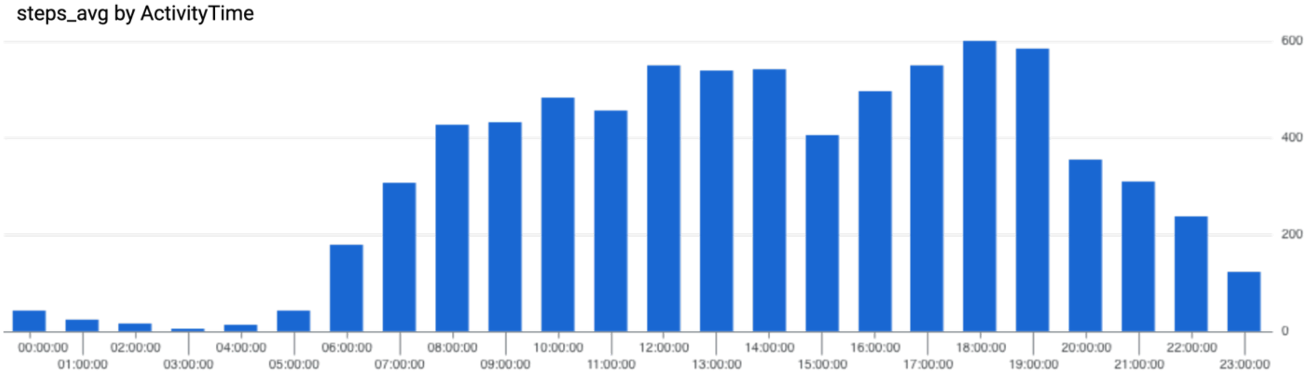
```
ORDER BY
  day_num
```

Row	day_of_week	step_avg	step_rank
1	Sunday	6933.23	7
2	Monday	7780.87	3
3	Tuesday	8125.01	2
4	Wednesday	7559.37	4
5	Thursday	7405.84	6
6	Friday	7448.23	5
7	Saturday	8152.98	1

Looking closer into the hourly steeps, we observe that, on average, users take the highest number of steps between 12:00 to 15:00 and 17:00 to 20:00.

```
-- average steps per hour of the day
SELECT
  ActivityTime
  ,ROUND(AVG(StepTotal), 2) AS steps_avg
FROM
  `belladata.steps_hour`
GROUP BY
  ActivityTime
ORDER BY
  ActivityTime
```

steps_avg by ActivityTime



This could be interpreted as most users working regular hours (from 8:00 to 17:00), allowing for more steps during lunch breaks or after work hours. The lowest number of steps occurs between 23:00 and 6:00, indicating that most users might be sleeping during this period. Sunday appears to be the preferred rest day, while Saturday is the day with additional steps. However, all weekdays show an above-average number of steps (based on https://www.medicalnewstoday.com/articles/average-steps-per-day).

To compare the number of steps with the rest of the data, we proceeded to analyze the calories data.

Calories burned Function

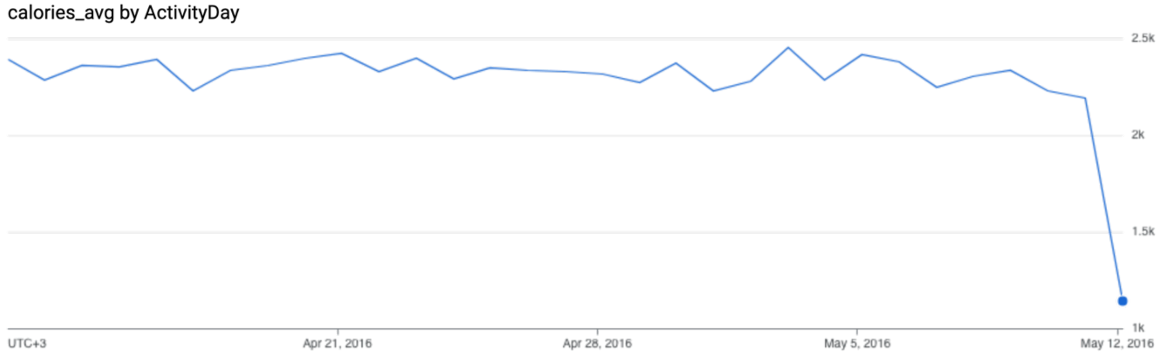
Using 2000 calories as the threshold for considering users active or inactive, it appears that 60% (20 out of 33) of the participants can be classified as active, based on the average calories recorded for each individual throughout the entire month.

```
-- average calories burned per user
SELECT
  DISTINCT Id
  ,ROUND(AVG(calories),2) AS calories_avg
FROM
  `belladata.calories_day`
GROUP BY
  Id
ORDER BY
  calories_avg DESC
```

Similar to what was observed with the steps count, there isn't much variation in the calories burned throughout the month, except for the last day, which shows a significant decrease.

```
-- average calories burned throughout the month
SELECT
  ActivityDay,
  ROUND(AVG(calories),2) AS calories_avg
FROM
  `belladata.calories_day`
GROUP BY
  ActivityDay
ORDER BY
  ActivityDay
```

calories_avg by ActivityDay



Similar to the steps count, Saturday and Tuesday are the days with the higher number of calories burned, whereas Sunday and Thursday reflect the lowest numbers.

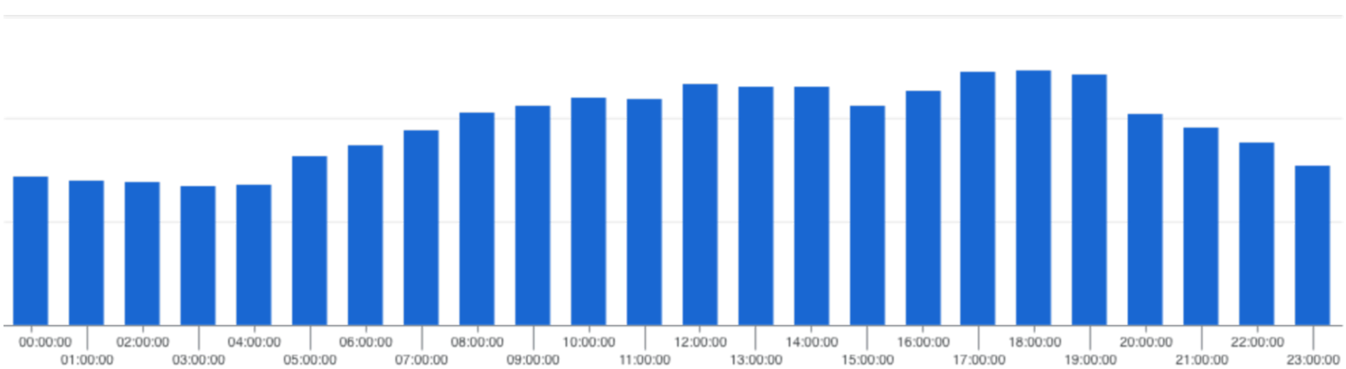
```
-- average calories burned and steps per day (weekday)
WITH calories_avg AS (
  SELECT
    EXTRACT(DAYOFWEEK FROM ActivityDay) AS day_num
    ,FORMAT_DATE('%A', ActivityDay) AS day_of_week
    ,ROUND(AVG(calories),2) AS calories_avg
  FROM
    `belladata.calories_day`
  GROUP BY
    day_of_week
    ,day_num
    ),
steps_average AS (
  SELECT
    FORMAT_DATE('%A', ActivityDay) AS day_of_week
    ,ROUND(AVG(StepTotal), 2) AS step_avg
  FROM
    `belladata.steps_day`
  GROUP BY
    day_of_week
  )
SELECT
  calories_avg.day_of_week
  ,calories_avg.calories_avg
  ,steps_average.step_avg
  ,RANK() OVER(ORDER BY calories_avg.calories_avg DESC) AS calories_rank
  ,RANK() OVER(ORDER BY step_avg DESC) AS step_rank
FROM
  calories_avg
JOIN
  steps_average
ON calories_avg.day_of_week = steps_average.day_of_week
ORDER BY
  day_num ASC
```

Row	day_of_week	calories_avg	step_avg	calories_rank	step_rank
1	Sunday	2263.0	6933.23	6	7
2	Monday	2324.21	7780.87	4	3
3	Tuesday	2356.01	8125.01	1	2
4	Wednesday	2302.62	7559.37	5	4
5	Thursday	2199.57	7405.84	7	6
6	Friday	2331.79	7448.23	3	5
7	Saturday	2354.97	8152.98	2	1

Similar observations arise when analyzing the calories burned per hour. Although the changes throughout the day are more subtle, the periods with the highest number of calories burned are between 12:00 to 15:00 and 17:00 to 20:00.

```
-- average calories burned per hour
SELECT
  ActivityTime
  ,ROUND(AVG(calories), 2) AS steps_avg
FROM
  `belladata.calories_hour`
GROUP BY
  ActivityTime
ORDER BY
  ActivityTime
```

calories_avg by ActivityTime



On average, the participants burned more than 2000 calories each weekday. This is also true for all the dates throughout the entire month except for the last one.

Intensities Function

According to the intensities data, it appears that the majority of the calories were burned during sedentary and light activity moments.

```
SELECT
  DISTINCT(Id) AS Id
  ,ActivityDay
  ,FORMAT_DATE('%A', ActivityDay) AS day_of_week
  ,EXTRACT(DAYOFWEEK FROM ActivityDay) AS day_num
  ,AVG(sedentary_total) AS sedentary_avg
  ,AVG(light_total) AS light_avg
  ,AVG(active_total) AS active_avg
FROM
  (
  SELECT
```

```
    Id
    ,ActivityDay
    ,SUM(SedentaryMinutes) AS sedentary_total
    ,SUM(LightlyActiveMinutes) AS light_total
    ,SUM(VeryActiveMinutes) AS active_total
  FROM
    `belladata.intensities_day`
  GROUP BY
    ActivityDay
    ,Id
  ORDER BY
    Id,
    ActivityDay
  )
GROUP BY
  Id
  ,day_num
  ,ActivityDay
ORDER BY
  activityDay
```

The average time of light activity exceeds three hours for almost every day of the week (except one with 2.9h), while the active minutes range only between 10 and 15 for each day. This indicated that, when viewed by the day of the week, the average user is moderately active, showing very little to no (highly) active time.

```
SELECT
  *
  ,ROUND((light + active) / 60, 1) AS non_sedentary_hours
  ,RANK() OVER(ORDER BY light + active DESC) AS active_rank
FROM
  (
  SELECT
    day_of_week
    ,ROUND(AVG(sedentary_avg), 2) AS sedentary
    ,ROUND(AVG(light_avg), 2) AS light
    ,ROUND(AVG(active_avg), 2) AS active
  FROM
    `belladata.intensities_user`
  GROUP BY
    day_of_week
    ,day_num
  ORDER BY
    day_num
  )
```

Row	day_of_week	sendentaty	light	active	non_sedentary_hours	active_rank
1	Saturday	964.28	207.15	15.2	3.7	1
2	Tuesday	1007.36	197.34	14.34	3.5	3
3	Friday	1000.31	204.2	12.11	3.6	2
4	Wednesday	989.48	189.85	13.1	3.4	5
5	Sunday	990.26	173.98	14.53	3.1	7
6	Thursday	961.99	185.42	11.96	3.3	6
7	Monday	1027.94	192.06	14.0	3.4	4

If we consider the sum of the lightly active time and very active time as the workout time, we can conclude that, on average, every participant has workouts lasting at least 40 minutes and up to 5.5 hours.

```
-- intensities per user
SELECT
  *
  ,ROUND((light + active) / 60, 1) AS non_sedentary_hours
FROM
  (
    SELECT
    DISTINCT Id
    ,ROUND(AVG(sedentary_avg), 2) AS sendentaty
    ,ROUND(AVG(light_avg), 2) AS light
    ,ROUND(AVG(active_avg), 2) AS active
  FROM
    `belladata.intensities_user`
  GROUP BY
    Id
  )
ORDER BY
  5 DESC
```

Based on this, we can conclude that the participants are moderately active individuals who make steps and burn more calories than the average, although they might not frequently engage in very active workouts or activities. Saturday and Tuesday are among the most active days, consistent with the steps and calories burned. Similarly, Sunday and Thursday appear to be the least active days.

Sleep Function

Out of the 33 participants, only 24 used the Sleep Function, and among them, only 12.5% (3 out of 24) used it every day.

```
-- sleep recordins per users througout the month
SELECT
  id,
  COUNT(DISTINCT ActivityDate) AS use_cnt
FROM
  `belladata.sleep_day`
GROUP BY
  id
ORDER BY
  2 DESC
```

The preferred days of the week for using the sleep function are mid-week, specifically Wednesday, Tuesday and Thursday.

```
-- uses and users per week - totals and percentages
WITH weekdays AS (
  SELECT
    FORMAT_DATE('%A', ActivityDate) AS day_of_week,
    COUNT(Id) AS times_used,
    COUNT(DISTINCT Id) AS users
  FROM
    `belladata.sleep_day`
  GROUP BY
    day_of_week
  ),
total_cnt AS (
  SELECT
    COUNT(DISTINCT Id) AS total_users,
    COUNT(Id) AS total_uses
  FROM
    `belladata.sleep_day`
  )
SELECT
  day.day_of_week,
  day.times_used,
  day.users,
  ROUND(day.times_used / total.total_uses * 100, 2) AS total_used_pcnt,
  ROUND(day.users / total.total_users * 100, 2) AS total_users_pcnt
FROM
  weekdays day
CROSS JOIN
  total_cnt total
```

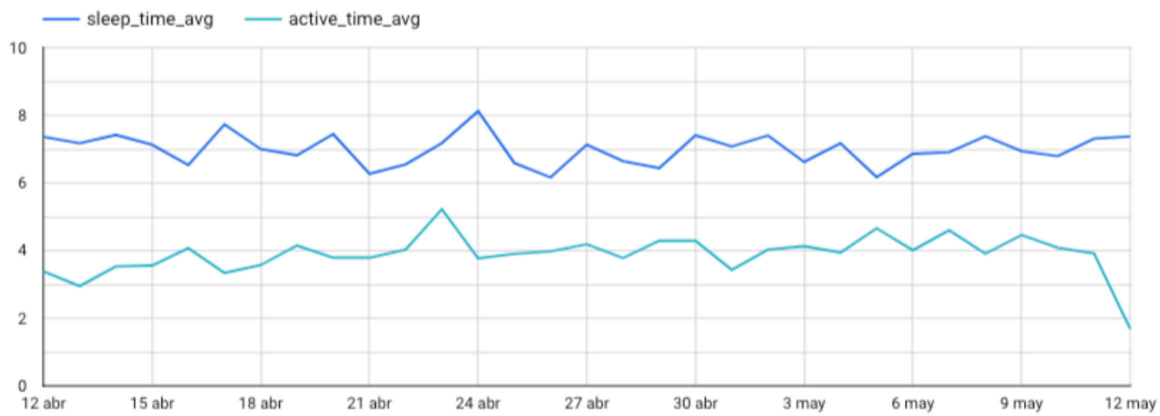
Row	day_of_week	times_used	users	total_used_pcnt	total_users_pcnt
1	Wednesday	66	19	15.98	79.17
2	Tuesday	65	18	15.74	75.0
3	Thursday	65	19	15.74	79.17
4	Saturday	58	21	14.04	87.5
5	Friday	57	21	13.8	87.5
6	Sunday	55	20	13.32	83.33
7	Monday	47	16	11.38	66.67

Assuming that having more than one sleep recording on the same date means that the user took a nap, it appears that napping is not common. Even though half of the participants using the sleep function (12 out of 24) had naps at least once, naps were recorded only 46 times out of 413 instances (11%).

```
SELECT
  COUNT(DISTINCT Id) AS users_napped
  ,COUNT(TotalSleepRecords) AS times_napped
FROM
  `belladata.sleep_day`
WHERE
  TotalSleepRecords > 1
```

The average sleep time from all the users through this month was 7 hours. There seems to be a relationship between the number of active hours and the hours slept, when observed throughout the month. However, this relationship cannot be seen when analyzed by user.

```
SELECT
  DISTINCT ActivityDay
  ,ROUND(AVG(non_sedentary_time), 2) AS active_time_avg
  ,ROUND(AVG(time_asleep), 2) AS sleep_time_avg
FROM
  `belladata.intensities_sleep`
GROUP BY
  1
```



The sleep function is not used every day, if used at all. Although the analysis suggests that the majority of the participants using the sleep function do not take naps, supporting the assumption that most of them are engaged in work or regular activities within an 8:00 to 17:00 schedule, this cannot be conclusively accepted due to lack of sufficient data. Moreover, many participants have insufficient sleep recording and some present minimal hours of sleep. More data is needed to assess if those could be considered outliers.

Share Phase

Refer to the PowerPoint Presentation.

Act Phase

Based on the previous analysis, the recommendations for initial steps are as follow:

- Develop two distinct marketing campaigns:
  - Target individuals with sedentary lifestyles (potential customers), highlighting the advantages of this kind of wearable devices. The analysis indicates that the users of these devices tend to burn more calories and take more steps than average, potentially attracting users seeking for these changes in their lifestyles.
  - Focus on individuals with active lifestyles (current customers), emphasizing the features that are currently underutilized (such as sleep tracking and weight monitoring) and their potential benefits. Under the assumption that current and potential users are unaware of the advantages of these features, it is recommended to run small campaigns for social media to spotlight those benefits.
- Implement in-app notifications for the current users, encouraging them to explore the different features. These notifications could include motivational messages, progress tracking within the general trend and personalized progress tracking over time.
- Define social media goals and Key Performance Indicators to measure the retention and engagement of customers with the brand. The KPIs might include indicators such as customer retention rate and usage of the different features.
- Define customer profiles. One specific suggestion is to create advertisements targeting women interested in this type of wearable devices but hesitant to use them in specific occasions due the perceived lack of “fashionable” element. This effort might aim to engage both women considering BellaBeat as their primary wearable device and those seeking an alternative device for specific occasions.

Next steps:

- Expand the analysis by utilizing a larger sample size and extending the timeframe. Considering the declining popularity of Fitbit devices in recent years, it is advised to incorporate data from other wearable devices, like Apple and Garmin, if available. This approach will provide a more comprehensive understanding of the current market landscape and user behavior.

- Compare this data with internal data, to understand which Bellabeat features should be prioritized within the larger market. It's important to note that while studies indicate that data collected via platforms like Mturk – used for this study – tend to have more demographic diversity than other online data collection methods, the lack of available demographics prevents confirmation for this study. Bellabeat serves a specific niche, and this sample might not adequately represent this population. (Reference: https://www.researchprotocols.org/2017/4/e66#ref10)

- If corroborated by further analyses, consider developing a campaign targeting individuals seeking to improve sleeping habits, by showing the correlation between calories burned and activity levels, and the potential benefits on sleep quantity, as a key focus for this campaign.