

## Case Study Overview

Bellabeat is a tech focused health and wellness company that utilizes smart devices and wearable tech to track a number of wellness factors. In this analysis I will be exploring usage trends for similar smart devices to inform Bellabeat's marketing strategy.

### Step 1: Determine Goals and Stakeholders

#### **Business Task**

Deliver marketing strategy recommendations based on analysis of smart device usage data.

#### **Key Stakeholders**

Urška Sršen - Cofounder and Chief Creative Officer

Sando Mur - Mathematician and Cofounder

Bellabeat marketing analytics team

### Step 2: Locate and Prepare Data

**Dataset:** FitBit Fitness Tracker Data ([link](#)) 2016 - Public domain data set downloaded from Kaggle. From the content description:

"This dataset 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. Individual reports can be parsed by export session ID (column A) or timestamp (column B). Variation between output represents use of different types of Fitbit trackers and individual tracking behaviors / preferences."

#### **Assessment of Credibility**

The dataset has a number of issues that will affect its overall usefulness and will inform the analysis and recommendations to follow.

- The dataset utilizes a very small sample size of 30 participants who consented to the collection of their personal tracker data over a short period of time. Any trends present in the data may not be reliable indicators.
- The dataset is fairly old (2016). User habits may have changed significantly making it difficult to know whether this usage data is consistent with more contemporary trends.
- There is very little identifying information about the users whose data was collected for this study. While this is a good thing for user privacy, it also makes it difficult to determine whether any external factors such as age, sex, or preexisting health conditions might have contributed to any trends in the data.

Due to all of these potential problems with the dataset, it would be best to use the findings in this report as a guide for further analysis using more comprehensive and current data.

### Step 3: Processing and Cleaning Data

My first step of processing the data was to download the dataset and uploading it into BigQuery

I focused on uploading and analyzing the daily activity summary as it aggregates steps, calories, and activity level into a single dataset, as well as the heart rate data and the sleep data. With these metrics we should be able to see if there are noticeable trends between a variety of factors.

When attempting to upload the datasets for heart rate and sleep, it became clear that the Date and Time columns of both datasets were incorrectly formatted for analysis. I brought them into Google Sheets to reformat the column before reuploading them into BigQuery.

With all of the datasets formatted properly, I began to scan through the schema and previews for the data to see what might be useful and what I might need to check for accuracy.

- First, I knew I should check the datasets for any 0 values where they might skew any calculations. By using the following query I only found one row where 'TotalDistance' was 0 and removed it.

```
Untitled 3 [RUN] [SAVE] [SHARE] [SCHEDULE] [MORE]
1 SELECT *
2 FROM `skilled-nation-381416.Fitbit_user_data.daily_activity`
3 WHERE TotalDistance = 0
```

- Second I checked the datasets for duplicated data. I used the following query and found 3 instances for duplicate entries in the 'Sleep' table then removed the duplicates by replacing the table with only distinct rows.

```
Untitled 3 [RUN] [SAVE] [SHARE] [SCHEDULE] [MORE]
1 SELECT Id, SleepDay, TotalMinutesAsleep, COUNT(*)
2 FROM Fitbit_user_data.sleep
3 GROUP BY Id, SleepDay, TotalMinutesAsleep
4 HAVING COUNT(*) > 1
```

```
Untitled 3 [RUN] [SAVE] [SHARE] [SCHEDULE] [MORE]
1 CREATE OR REPLACE TABLE `Fitbit_user_data.sleep` AS [
2 SELECT DISTINCT *
3 FROM `Fitbit_user_data.sleep`]
```

- Finally, I wanted to count the number of distinct users in each dataset and confirm which users were present in all three datasets for further analysis. By using the following query, I found that 12 user IDs exist in all datasets.

```

Untitled 3  [RUN] [SAVE] [SHARE] [SCHEDULE] [MORE]
1 SELECT DISTINCT hr.Id
2 FROM `Fitbit_user_data.average_daily_heartrates` AS hr
3 INNER JOIN `Fitbit_user_data.daily_activity` AS da
4 ON hr.Id = da.Id
5 INNER JOIN `Fitbit_user_data.sleep` AS sl
6 ON hr.Id = sl.Id

```

With the data prepared and processed I began to consider what analysis would be most useful and might provide the client with actionable steps toward a marketing strategy.

#### Step 4: Analysis

I first composed the following query to create a new table specifically aggregating the data from the 12 users who existed in all of the data sets.

```

1 CREATE TABLE Fitbit_user_data.all_metrics AS
2 SELECT cr.Id,
3 CASE WHEN COUNT(sl.Id) >= 10 THEN AVG(TotalMinutesAsleep) END AS avg_sleep,
4 AVG(Value) AS avg_heartrate,
5 ROUND(AVG(TotalSteps)) AS steps,
6 ROUND(AVG(VeryActiveMinutes)) AS very_active,
7 ROUND(AVG(FairlyActiveMinutes)) AS fairly_active,
8 ROUND(AVG(LightlyActiveMinutes)) AS lightly_active,
9 ROUND(AVG(SedentaryMinutes)) AS sedentary
10 FROM `Fitbit_user_data.crossref` AS cr
11 INNER JOIN `Fitbit_user_data.daily_activity` AS da
12 ON cr.Id = da.Id
13 INNER JOIN `Fitbit_user_data.heartrate` AS hr
14 ON cr.Id = hr.Id
15 INNER JOIN `Fitbit_user_data.sleep` AS sl
16 ON cr.Id = sl.Id
17 GROUP BY cr.Id

```

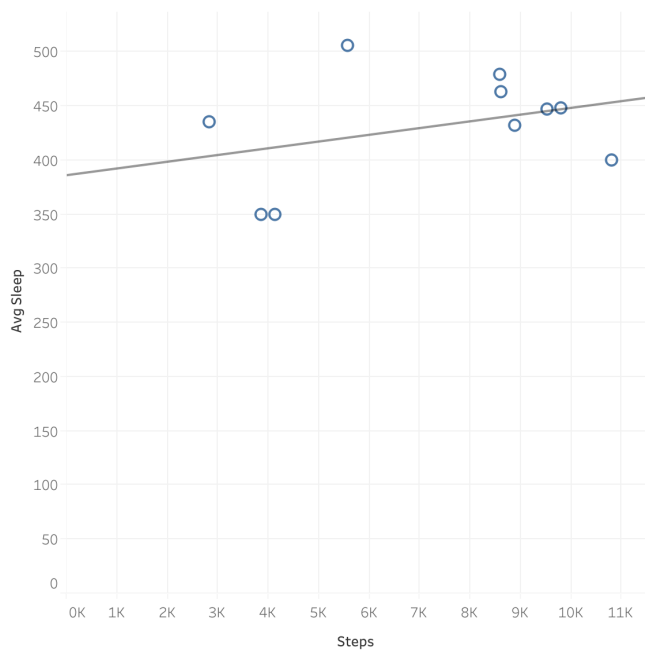
I was curious to know more information about the following questions:

1. Is there a relationship between activity level and the time spent asleep? A positive correlation might indicate that increased activity helps ensure more restful sleep.
2. Is there a trend in the general activity levels, heart rate, or sleep patterns of the participants? If there are consistent clusters within all of the metrics it may indicate larger trends in the user base.

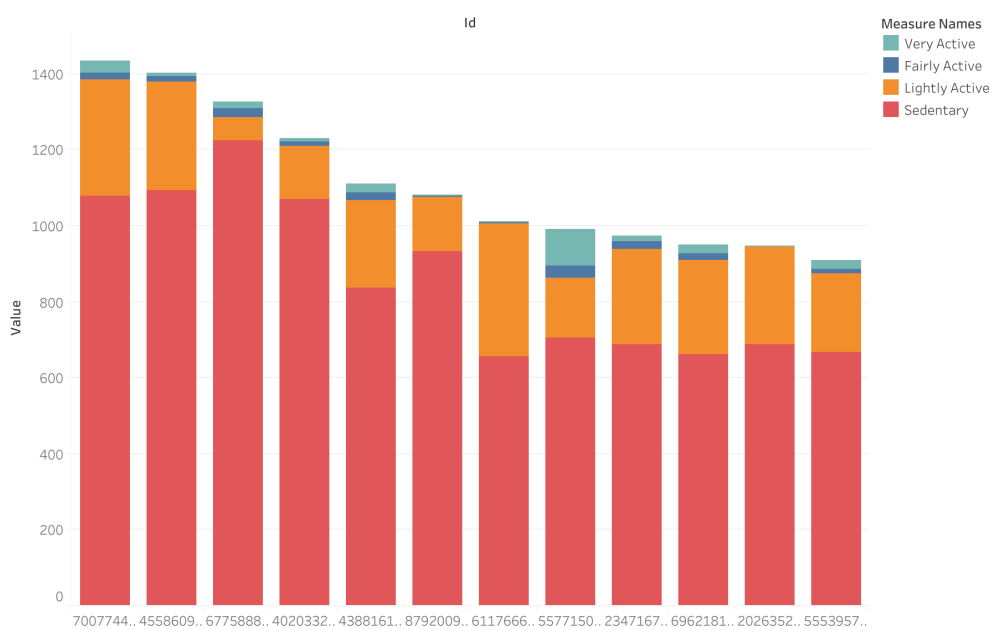
Using only the 12 participants for whom there was data for all of the metrics, I was not able to determine any significant and clear trends. But there were some potential indicators that might guide future analysis after expanding the analysis to include more participants.

- There seems to be a slight indication that increased activity metrics like step count and active minutes have a positive correlation with time spent asleep.
- In general, the 12 users are sedentary for around 77% of their logged activity time, lightly active for 20%, and fairly or very active for 1-3%.

Sleep and Step Count



Activity Breakdown



I decided I needed to expand the sample to further investigate any relationships. Since the heart rate metric was not seeming to indicate any strong correlation or use case, I removed it from the dataset to find 24 users for whom there was data for both sleep and activity level.

Total Metrics Over Time

RUN

SAVE QUERY

SHARE

SCHEDULE

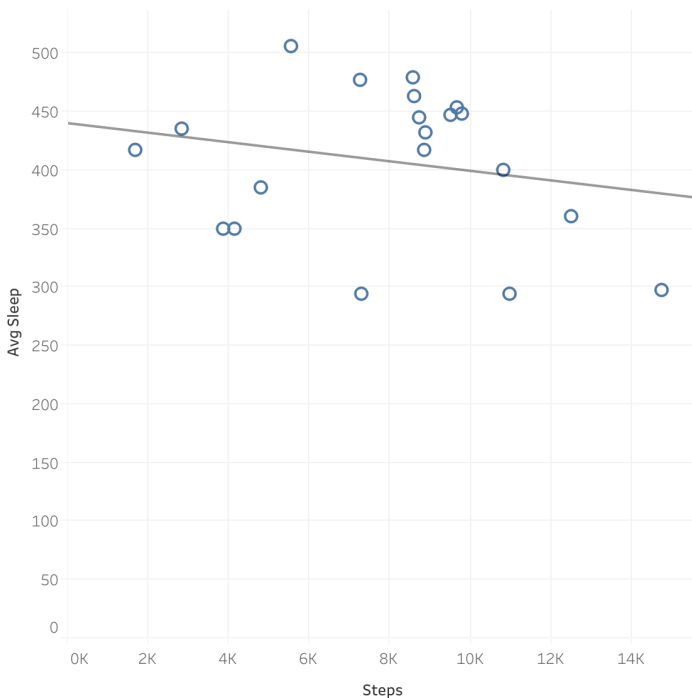
MORE

```
1 SELECT DISTINCT ActivityDate,
2 SUM(TotalSteps) AS Steps,
3 SUM(SedentaryMinutes) AS Sedentary,
4 SUM(LightlyActiveMinutes) AS LightlyActive,
5 SUM(FairlyActiveMinutes) AS FairlyActive,
6 SUM(VeryActiveMinutes) AS VeryActive,
7 FROM `skilled-nation-381416.Fitbit_user_data.daily_activity`
8 GROUP BY ActivityDate
9 ORDER BY ActivityDate
```

After manipulating the data I discovered that my previous hypothesis regarding activity level and sleep quality was no longer supported by the data.

- The most significant cluster lands between 8,000 and 10,000 steps on average and 400 to 500 minutes asleep on average. But more or less steps do not show a clear increase or decrease in time spent asleep.

Sleep and Step Count (Expanded Sample)



In addition, the distribution of activity time was only further weighted towards sedentary time when analyzing the larger sample size

- On average, users were sedentary for 79% of their logged activity time, lightly active for 18% of the time, and fairly or very active for 1-3%

Activity Breakdown (Expanded Sample)



Overall, it was difficult to see clear, marketable trends in the user data upon initial investigation. Again, however, the analysis could perhaps be more useful in guiding future research and investigation. The data might indicate that the users who submitted their data were generally more sedentary but I wondered if their investment in a smart fitness device might cause an increase in activity over time.

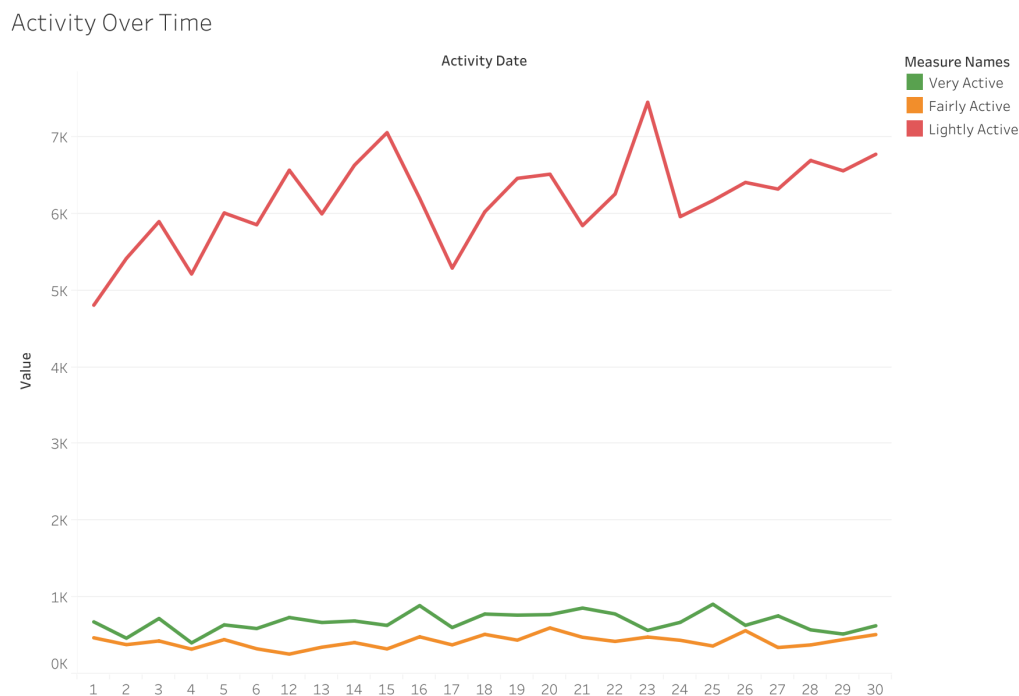
Using the following query, I created a count of the total logged activity minutes and steps for each activity day. Even with a fairly short period of time, I was hoping that there would be some small indication that total steps or activity level would increase and sedentary minutes would decrease slightly.

The first thing I noticed was that there were a number of days where the total of all recorded minutes were significantly higher or lower than average which would indicate that fewer or more users logged their activity those days. This could be due to a number of factors including the devices being forgotten, out of battery, or other circumstances. To get a useful trend line, I needed to maintain a somewhat consistent total of minutes logged. I amended the query to the following to create a new table with a total minute count that I could use to filter out the very low and very high minute days.

```
🔍 Total Metrics Over Time RUN SAVE QUERY SHARE SCHEDULE MORE
```

```
1 CREATE TABLE Fitbit_user_data.TotalsOverTime AS
2 SELECT DISTINCT ActivityDate,
3 SUM(TotalSteps) AS Steps,
4 SUM(SedentaryMinutes) AS Sedentary,
5 SUM(LightlyActiveMinutes) AS LightlyActive,
6 SUM(FairlyActiveMinutes) AS FairlyActive,
7 SUM(VeryActiveMinutes) AS VeryActive,
8 SUM(SedentaryMinutes+LightlyActiveMinutes+FairlyActiveMinutes+VeryActiveMinutes) AS TotalMinutes
9 FROM `skilled-nation-381416.Fitbit_user_data.daily_activity`
10 GROUP BY ActivityDate
11 ORDER BY ActivityDate
```

After removing the outlier dates, the data still does not show an undeniably significant change over time but there is a more clear indication that there is perhaps some trend of increased lightly active minutes.



## Conclusions

The data is unfortunately very limited in its capability to describe significant trends in smart device usage, however this initial analysis may serve as a guide for future research which can guide confident business decisions.

There were some interesting insights gained from this analysis despite its limited nature.

1. The users who submitted their data seem to have a similar breakdown of activity levels. Primarily sedentary with some light activity and a small amount of intense activity.
2. The users seemed to show a slight trend of increasing their amount of light physical activity the longer they were using their device
3. Some users may have shown early indications of an increase in quality of sleep by becoming more active

Based on these findings, it would be beneficial to collect new data in order to focus on these potential trends and address some of the shortcomings of the existing data.

- A larger sample size would allow for more comprehensive market trend data, and a longer collection period would give Bellabeat the ability to focus on how usage changes over time.
- Collecting new data will ensure that the marketing strategy is addressing contemporary trends in the market and the ways that modern consumers approach smart fitness devices.
- Collecting some voluntary data on demographics would also help contextualize the data. The existing dataset has very little context and a better understanding of who your customers are in addition to how they use your devices would allow for a more cohesive brand strategy.

I recommend a 3 month period of data collection and surveys focused on new users and how their usage changes over time with a target of at least 100 participants. The resulting data should verify some of the trends found here and may reveal new insights allowing Bellabeat to discover marketing strategies to bring in new users and retain them.