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

ANKITA SINHA

SUMMARY

The following case study analyses how users of different smart devices vary in usage trends, which in turn can be studied to inform Bellabeat's marketing strategy. The analysis is based on a publicly available Fitbit dataset collected in 2016, which having some limitations (small sample size, outdated, and lacking key demographic details), provides insights into some users' physical activity patterns. Exploratory data analysis revealed some strong correlations between TotalSteps and Calories, as well as between lifestyle activity minutes and heart rate zones of individuals. Additionally, the users' other patterns were noticed like the percentage of them who record their activity manually versus the ones who depend on the tracking device, and statistical tests confirmed these differences which are laid out as important proofs for further marketing strategy improvement.

INTRODUCTION

Urška Sršen and Sando Mur founded Bellabeat, a high-tech company that manufactures health-focused smart products. Sršen used her background as an artist to develop beautifully designed technology that informs and inspires women around the world. Collecting data on activity, sleep, stress, and reproductive health has allowed Bellabeat to empower women with knowledge about their own health and habits. Since it was founded in 2013, Bellabeat has grown rapidly and quickly positioned itself as a tech-driven wellness company for women.

Scenario

To act as a junior data analyst working on the marketing analyst team at Bellabeat, a high-tech manufacturer of health-focused products for women. Bellabeat is a successful small company, but they have the potential to become a larger player in the global smart device market. Urška Sršen, cofounder and Chief Creative Officer of Bellabeat, believes that analyzing smart device fitness data could help unlock new growth opportunities for the company. The task is to focus on one of Bellabeat's products and analyze smart device data to gain insight into how consumers are using their smart devices. The insights gained will then help guide marketing strategy for the company. The analysis has to be presented to the Bellabeat executive team along with high-level recommendations for Bellabeat's marketing strategy.

DATA ANALYSIS STEP-BY-STEP GUIDE

1. Ask Phase

Sršen asked to analyze smart device usage data in order to gain insight into how consumers use non-Bellabeat smart devices. These questions will guide the 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?

Business Task:

Analyse the trends in smart device usage, identifying marketing strategies based on a sample of users' habits and usage of smart health devices. Based on the patterns found, to identify the intersection of trends in most smart health device users and Bellabeat customers in particular so that these insights drive marketing strategies.

Some questions answered:

● What is the problem you are trying to solve?
Firstly to analyse the trends in smart device usage, and then use the gained insights to solve the main business task- identifying marketing strategies based on a sample of users' habits and usage of smart health devices. What motivates them to buy, keep using, frequency of usage, health problems to be addressed, what exactly they want to track, etc. Based on these patterns our goal is to identify the intersection of trends in most smart health device users and Bellabeat customers in particular so that the marketing team can further take actions into monetising those patterns and needs. Fulfilling the gap between need and availability of changes and modification in existing strategy.

● How can your insights drive business decisions?
The insights I derive can point to the shortfalls in the company, customer retention rate and the outcome of existing marketing strategy- which would also account for the growth percentage, revenue change in profit year on year, which products are doing better than the others, faults in existing products (if we do analyse all the products instead of one), create new enquiries, see if better deals might improve sales. My insights can show the important areas of health people are targeting more and hence how can the product cater to optimise those needs in a better way.

Stakeholders

- Urška Sršen: Bellabeat's cofounder and Chief Creative Officer
- Sando Mur: Mathematician and Bellabeat's cofounder; key member of the Bellabeat executive team
- Bellabeat marketing analytics team: A team of data analysts responsible for collecting, analyzing, and reporting data that helps guide Bellabeat's marketing strategy.
- Product team: Responsible for development of the software and security of the tracking device and associated data.
- Investors: For the Bellabeat company, who are vouching for the marketing upgrades and results.

2. Prepare Phase

Source and Storage of the Data:

FitBit Fitness Tracker Data (CC0: Public Domain, dataset made available through Mobius): This Kaggle data set contains personal fitness tracker from thirty (34) 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.
This data set has some limitations, mainly the small sample size, limited data only for 2 months and which is not aggregated in the same format, format issues, incomplete sleep and weight data, non-inclusive and non-definitive group of people.

This dataset generated by respondents to a distributed survey via Amazon Mechanical Turk between 03.12.2016-05.12.2016 (2 month worth of data separated into 2 datasets for each month). Variation between output represents use of different types of Fitbit trackers and individual tracking behaviors / preferences. I have only considered some of the data files (in .csv) which seemed relevant to answer the business goals and improve marketing.

The data is downloaded from Kaggle and stored into local device, from where it is loaded and imported to Google Sheets and BigQuery for analysis steps.

Potential Biases and Limitations

When analyzing the given Fitbit Fitness Tracker data, it's important to acknowledge its potential limitations:

- *Self-Selection Bias:* Since Fitbit users are generally more health-conscious than the average population, activity levels in the dataset may be higher than what's typical.
- *Limited Demographics*: The dataset lacks detailed demographic information, such as age and gender, making it difficult to segment users effectively. This also means we can't confirm whether the data aligns with Bellabeat's target female audience.
- *Small Sample Size & Timeframe*: With data collected from only thirty users over two months, it may not be representative of broader populations or long-term trends.
- *Data Accuracy:* The reliability of Fitbit data can vary due to device limitations and inconsistent usage patterns among users.

While these factors introduce some constraints, the dataset still provides valuable insights into user activity trends. However, any conclusions should be interpreted with caution, keeping these biases in mind.

Data Overview

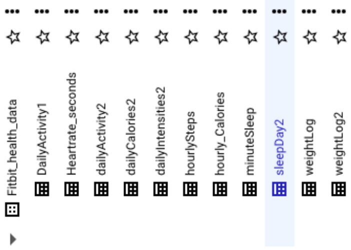
We will use the following primary datasets for this analysis:

- dailyActivity_merged.csv: Provides daily step count, distance, activity levels, and calories burned.
- dailyCalories_merged.csv: Daily summary of calories burned.
- Heartrate_seconds_merged.csv: Heartrate of user broken down and recorded every 5-10 seconds.
- sleepDay_merged.csv: Daily sleep records including total sleep time and time in bed.

And some other csv files listed as examples below and later:

Dataset Description

Dataset Name	Description
dailyActivity_merged.csv	Tracks user activity levels, distance, and calories burned.
dailyCalories_merged.csv	Summarizes calories burned per day.
dailyIntensities_merged.csv	Breaks down activity intensity by minutes and distance.
hourlySteps_merged.csv	Provides hourly step count metrics.
hourlyCalories_merged.csv	Calories burned per hour.
sleepDay_merged.csv	Sleep data tracked per day as TotalTimeAsleep and TimeInBed



(Sample Database tables from Big Query)

Data integrity and credibility

Data is in **Wide Format** where each row represents a single entity (e.g., user, date).

- Different variables (e.g., Id, heart_rate, sleeptime) are stored in separate columns.

The dataset, collected in 2016, is outdated and may not accurately reflect current health behaviors. The small sample size (34 participants) is not representative of the general female population, and key factors like age, diet, and health conditions are missing. Also, the data have been collected for just 2 months. Due to these limitations, the data lacks reliability and validity for making strong business recommendations. The analysis should be supplemented with more extensive and up-to-date data.

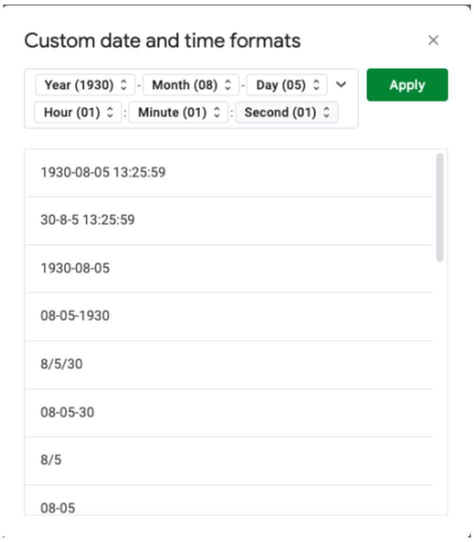
ROCCC System Evaluation:

- Reliability: Low—small, non-representative sample.
- Originality: Low—third-party data source.
- Comprehensiveness: Low—lacks key demographic details.
- Currency: Low—data is seven years old.
- Citation: Low—unverified third-party source.

<u>CLEANING STEPS:</u>

Import to Big Query or Google Sheets first,
Irrelevant data- ignored-
Considered some of these for preparing-

dailyActivity.csv imported into Big Query
heartrate_seconds.csv to Big Query- error in date time format- fixed it in Google Sheets-



YYYY-MM-DD HH:MM:SS FORMAT
Then imported the modified csv into Big Query

Same for hourlyCalories & hourlySteps
Weight_log- Date changed format and Kg, BMI rounded to 2 decimal places,

Same way for second folder- 12/4/2016 to 12/5/2016: Considered only these files- dailyActivity, dailyCalories, dailyIntensities, sleepDay, weightLog
And formatted date and weight, BMI up to 2 decimal places

In BigQuery Editor, removed duplicate rows, null values and outliers from dailyActivity1.csv:
For Outliers removed the top 1% extreme outliers-

```
CREATE TABLE proud-skein-437112-c4.Fitbit_health_data.final_daily_activity1 AS
WITH percentiles AS (
    SELECT
        APPROX_QUANTILES(TotalSteps, 100)[OFFSET(99)] AS steps_99th,
        APPROX_QUANTILES(Calories, 100)[OFFSET(99)] AS calories_99th
    FROM proud-skein-437112-c4.Fitbit_health_data.DailyActivity1
)
SELECT d.*
FROM proud-skein-437112-c4.Fitbit_health_data.DailyActivity1 d
JOIN percentiles p
ON d.TotalSteps <= p.steps_99th OR d.Calories <= p.calories_99th;
```

3. Process Phase

The dataset is downloaded, imported into Google Sheets for easy comparison and cleaning- to make variable formats correct and check for errors while loading csv. The rest of the cleaning and correctly modifying the dataset is done in BigQuery as follows.

I am choosing SQL and Sheets as the tools for this Data Analysis process, since the data is not too huge or not too small. I will focus my analysis here (and explore R in another version of this case study) due to the easy accessibility, ease of creating data visualization to share my results with stakeholders.

Some of the Data Preparation Steps (also done later with analyse processes):

- Calculated Total Daily Calories for the first month data from the hourly Calories.csv for each consumer, and stored results in a new table daily_calories1:

```
CREATE TABLE proud-skein-437112-c4.Fitbit_health_data.daily_calories1 AS
SELECT
    Id,
    DATE(ActivityHour) AS activity_date,
    SUM(calories) AS total_daily_calories
FROM proud-skein-437112-c4.Fitbit_health_data.hourly_Calories
GROUP BY Id, activity_date;
```

- Join the 2 tables for 2 months to form one table that has all Total Calorie data for each user per day, as merged_daily_calories:

```
CREATE TABLE proud-skein-437112-c4.Fitbit_health_data.merged_daily_calories AS
SELECT
    COALESCE(a.Id, b.Id) AS id,
    COALESCE(a.activity_date, b.ActivityDay) AS activityDate,
    COALESCE(a.total_daily_calories, 0) + COALESCE(b.Calories, 0) AS total_calories
FROM proud-skein-437112-c4.Fitbit_health_data.daily_calories1 AS a
FULL OUTER JOIN proud-skein-437112-c4.Fitbit_health_data.dailyCalories2 AS b
ON a.Id = b.Id AND a.activity_date = b.ActivityDay;
```

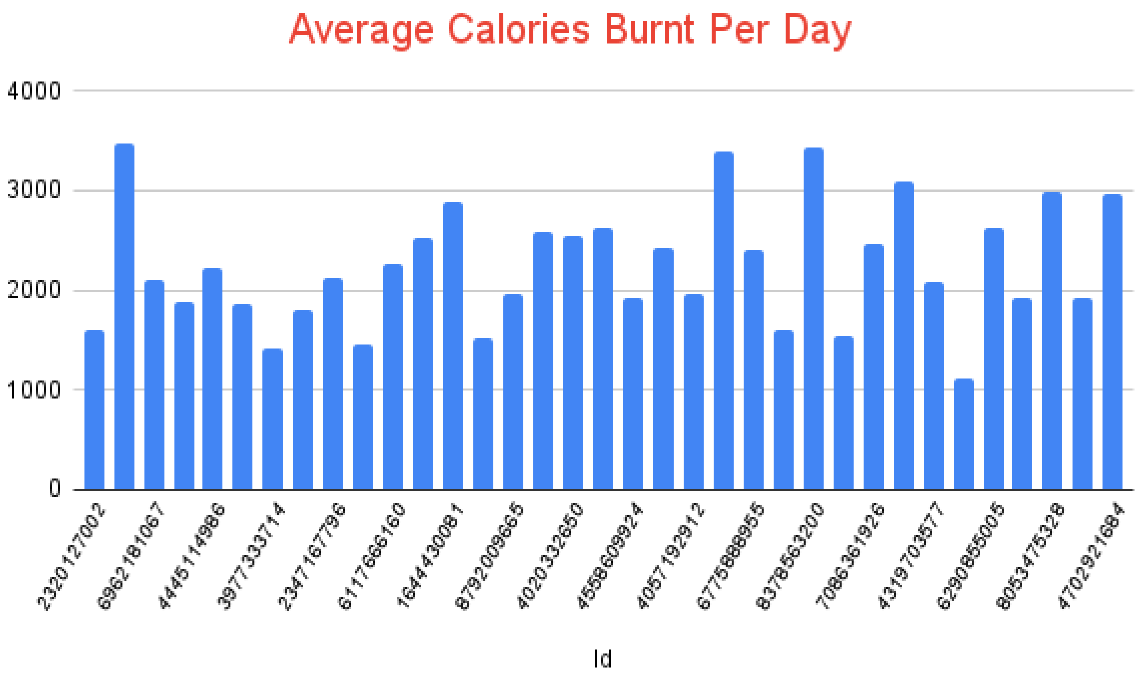
- Average Daily Calories per user rounded to 2 decimal places:

```
CREATE TABLE proud-skein-437112-c4.Fitbit_health_data.avg_daily_calories AS
SELECT
    Id,
    ROUND(AVG(total_calories), 2) AS avg_calories
FROM proud-skein-437112-c4.Fitbit_health_data.merged_daily_calories
GROUP BY Id;
```

4. Analyze Phase

The datasets used here are properly organized, formatted, joined into newer tables and correctly considered for metrics evaluation and subsequent source of insight generation. I have explored some of the relevant trends and relationships using my analysis methods and t answer our original business tasks. Every trend is cross-checked with our business goals and the marketing strategy needs are kept in mind while trying to draw conclusions.

Comparison of Average Calories per day for different users:



HEART RATE ZONES

Answers the question of categorizing the users depending on how much time they spend in each Heart Rate Zone, i.e., to understand if they are active throughout the day and what is their activity type.

To calculate the **time spent in different heart rate zones** per user **per day**, we can use SQL to categorize heart rate values into different zones and sum the total duration spent in each zone.

Defining the Heart Rate Zones

We categorize heart rate values into **four zones**:

- **Resting (<100 BPM)**
- **Light Activity (100-130 BPM)**
- **Moderate Activity (130-160 BPM)**
- **Intense Activity (>160 BPM)**

```sql
CREATE TABLE proud-skein-437112-c4.Fitbit_health_data.heart_rate_zones AS
SELECT
    id,
    DATE(Time) AS activityDate,

    -- Calculate total duration in each zone
    SUM(CASE WHEN Value < 100 THEN 10 ELSE 0 END) AS resting_time_seconds,
    SUM(CASE WHEN Value BETWEEN 100 AND 130 THEN 10 ELSE 0 END) AS
light_activity_time_seconds,
    SUM(CASE WHEN Value BETWEEN 130 AND 160 THEN 10 ELSE 0 END) AS
moderate_activity_time_seconds,
    SUM(CASE WHEN Value > 160 THEN 10 ELSE 0 END) AS intense_activity_time_seconds,

    -- Total active time (excluding resting)
    SUM(CASE WHEN Value >= 100 THEN 10 ELSE 0 END) AS total_active_time_seconds

FROM proud-skein-437112-c4.Fitbit_health_data.Heartrate_seconds
GROUP BY id, activityDate;
```

Sample Table Output:

Row	id	activityDate	resting_time_sec	light_activity_tim	moderate_activit	intense_activity_	total_active_time
1	2022484408	2016-04-02	42800	150	0	0	150
2	2022484408	2016-04-10	44570	5360	0	0	5360
3	2022484408	2016-04-11	42960	4200	2320	0	6420
4	2022484408	2016-04-06	45250	7540	1810	0	9210
5	2022484408	2016-04-12	11160	4100	90	0	4150
6	2022484408	2016-04-03	48980	3050	0	0	3050
7	2022484408	2016-04-05	44980	7780	1330	0	8970
8	2022484408	2016-04-08	48930	4970	1180	0	5950
9	2022484408	2016-04-04	43520	1600	1820	0	3370
10	2026352035	2016-04-02	4390	0	0	0	0
11	2347167796	2016-03-29	87240	2320	110	0	2420
12	2347167796	2016-04-03	82140	3550	0	0	3550
13	2347167796	2016-03-31	86130	1600	0	0	1600
14	2347167796	2016-04-02	83170	8390	600	0	8930
15	2347167796	2016-04-10	79640	6890	980	0	7810
16	2347167796	2016-04-09	61720	9220	60	0	9260
17	2347167796	2016-04-08	88210	2920	300	0	3210
18	2347167796	2016-04-01	85130	500	0	0	500
19	2347167796	2016-04-11	81360	3450	60	0	3500

FREQUENT EXERCISERS

I analyzed how many users frequently exercise by identifying those who spend significant time in moderate or intense heart rate zones using SQL:

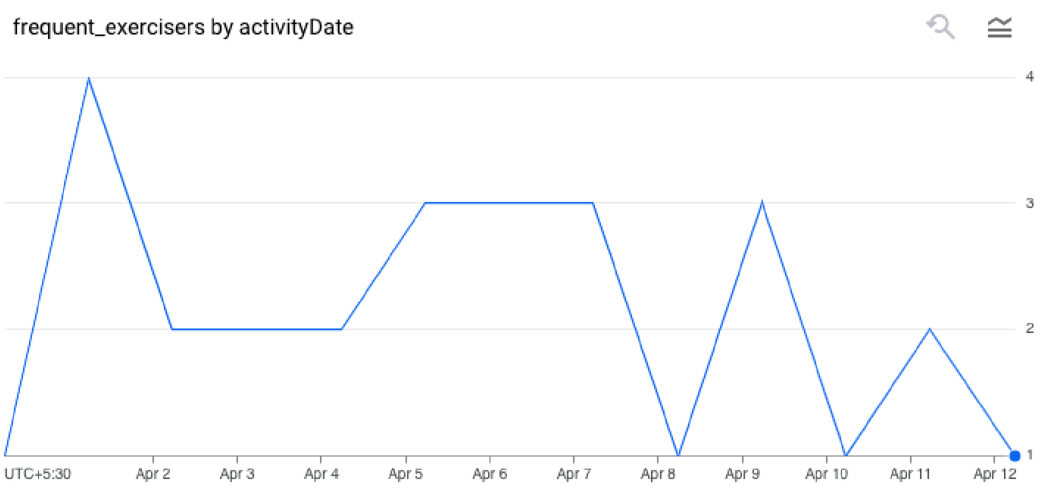
Defining "Frequent Exercisers"

A user can be considered a **frequent exerciser** if they spend: At least **30 minutes (1800 seconds) per day** in **moderate or intense activity**.

SQL Query to Count Frequent Exercisers

```
SELECT
    activityDate,
    COUNT(DISTINCT id) AS frequent_exercisers
FROM proud-skein-437112-c4.Fitbit_health_data.heart_rate_zones
WHERE (moderate_activity_time_seconds + intense_activity_time_seconds) >= 1800
GROUP BY activityDate
ORDER BY activityDate;
```

- Filtered users who spent at least 1800 seconds (30 min) in moderate or intense activity.
- Counted unique users per day who meet this threshold.
- Grouped by activityDate to track daily trends in frequent exercisers.



activityDate	frequent_exercisers
2016-03-31	1
2016-04-01	4
2016-04-02	2
2016-04-03	2
2016-04-04	2
2016-04-05	3
2016-04-06	3
2016-04-07	3
2016-04-08	1
2016-04-09	3

FURTHER ANALYSIS

1. Find the Most Active Users:

```
SELECT
    Id,
    COUNT(activityDate) AS active_days
FROM proud-skein-437112-c4.Fitbit_health_data.heart_rate_zones
WHERE (moderate_activity_time_seconds + intense_activity_time_seconds) >= 1800
GROUP BY Id
ORDER BY active_days DESC
LIMIT 10;
```

This shows the **top 10 most active users** based on how many days they exercised.

Output:

Id	active_days
6962181067	9
8877689391	8
2022484408	6
5577150313	2
4558609924	1
7007744171	1
6775888955	1

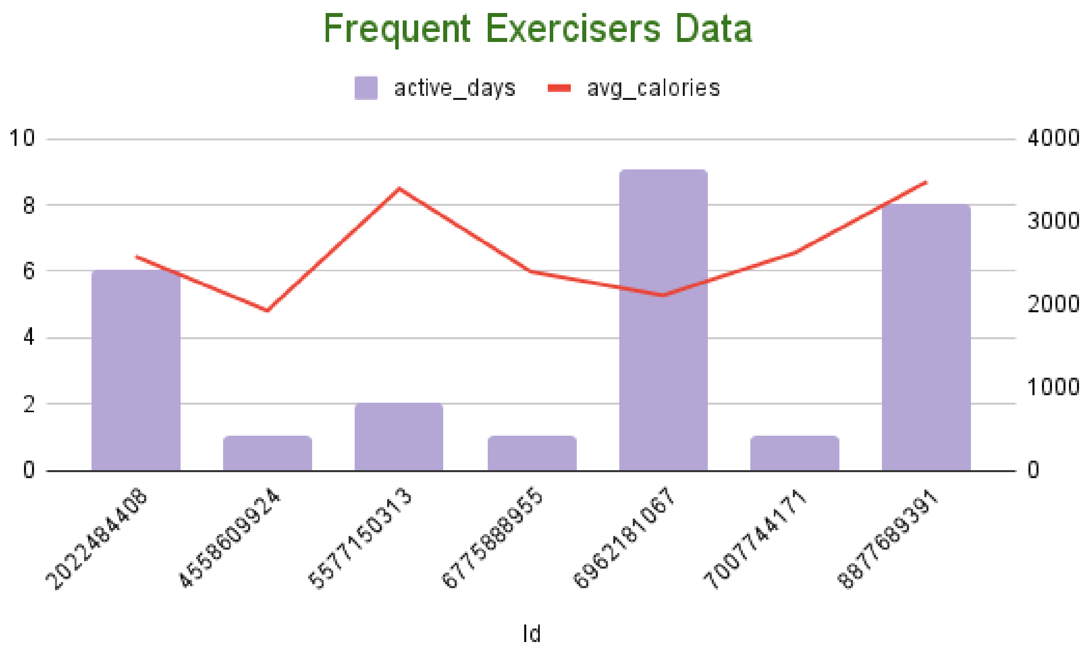
2. Compare with Average Calories Burned Per Day

Joined with the **daily calorie table** to see if **higher exercise time = more calories burned**.

```
SELECT
    fe.Id,
    fe.active_days,
    adc.avg_calories
FROM proud-skein-437112-c4.Fitbit_health_data.frequent_exercisers fe
INNER JOIN proud-skein-437112-c4.Fitbit_health_data.avg_daily_calories adc
ON fe.id = adc.id
ORDER BY fe.id;
```

Output:

Id	active_days	avg_calories
2022484408	6	2580.66
4558609924	1	1927.26
5577150313	2	3397.79
6775888955	1	2395.02
6962181067	9	2111.27
7007744171	1	2625.25
8877689391	8	3483.16



Inspect Correlation: Does more frequent exercise = higher calorie burn?
Cannot conclude from this small dataset as there are 3 people whose average calories are the greatest but their active days are least in the group.

DAILY ACTIVITY COMPARISONS

1. Average daily steps per user:

```
CREATE TABLE proud-skein-437112-c4.Fitbit_health_data.avg_steps_per_user AS
SELECT
    Id,
    ROUND(AVG(TotalSteps), 2) AS avg_steps
FROM proud-skein-437112-c4.Fitbit_health_data.final_daily_activity1
GROUP BY Id;
```

Count users whose average > 10000 steps

```
SELECT COUNT(*) AS users_above_10k
FROM proud-skein-437112-c4.Fitbit_health_data.avg_steps_per_user
WHERE avg_steps > 10000;
```

Output:
users_above_10k
6

2. **How many users log their activity without tracker (manually):**

```
SELECT COUNT(DISTINCT Id) AS users_logging_manually
FROM proud-skein-437112-c4.Fitbit_health_data.final_daily_activity1
WHERE LoggedActivitiesDistance > 0.0;
```

Output:
users_logging_manually
6

Why?- Increase advertisement and understand concerns from them, is it because they don't find the tracking data accurate? Or is it because they are unable to use the sync and tracking features properly. Address these in the next marketing campaign accordingly.

3. **Count of distinct user Id's in the entire dataset for comparison:**

```
SELECT COUNT(DISTINCT Id) AS total_distinct_users
FROM proud-skein-437112-c4.Fitbit_health_data.final_daily_activity1;
```

Output:
total_distinct_users
35

4. **Targeting Moderately Active Users for Engagement**

- **Who?** Users with FairlyActiveMinutes and LightlyActiveMinutes higher than VeryActiveMinutes.
- **Why?** They are interested in staying fit but might need motivation.
- **Potential Products:**
  - Fitness challenge programs, step-based rewards, gamified fitness apps.
  - Wearables with step reminders and coaching features.
  - Budget-friendly fitness equipment (resistance bands, home workout gear).
- **Marketing Strategy:**
  - Offer gamified fitness challenges with rewards (e.g., discounts for reaching 100K monthly steps).
  - Promote step-tracking competitions among users.

```
SELECT DISTINCT Id
FROM proud-skein-437112-c4.Fitbit_health_data.final_daily_activity1
WHERE FairlyActiveMinutes > VeryActiveMinutes
AND LightlyActiveMinutes > VeryActiveMinutes;
```

Output:
Id
1644430081
2022484408
….

5. Weekday vs Weekend activity (to track seasonal changes if any)

Questions we are looking to answer:

- Do users take more steps on weekends or weekdays?
- Are they more active (VeryActiveMinutes) on weekends?
- Are people more sedentary on weekdays due to work schedules?
- Do users burn more calories on weekends or weekdays?

I. Create a New Table with Weekday vs. Weekend Categorization

```
CREATE TABLE proud-skein-437112-c4.Fitbit_health_data.categorized_activity AS
SELECT
    Id,
    ActivityDate,
    CASE
        WHEN EXTRACT(DAYOFWEEK FROM ActivityDate) IN (1, 7) THEN 'Weekend'  -- 1 = Sunday, 7
= Saturday
        ELSE 'Weekday'
    END AS DayType,
    TotalSteps,
    VeryActiveMinutes,
    FairlyActiveMinutes,
    LightlyActiveMinutes,
    SedentaryMinutes,
    Calories
FROM proud-skein-437112-c4.Fitbit_health_data.final_daily_activity1;
```

II. Store Aggregated Weekend vs. Weekday Activity in a New Table

```
CREATE TABLE proud-skein-437112-c4.Fitbit_health_data.weekday_vs_weekend_analysis AS
SELECT
    DayType,
    ROUND(AVG(TotalSteps), 2) AS avg_steps,
    ROUND(AVG(VeryActiveMinutes), 2) AS avg_very_active,
    ROUND(AVG(FairlyActiveMinutes), 2) AS avg_fairly_active,
    ROUND(AVG(LightlyActiveMinutes), 2) AS avg_lightly_active,
    ROUND(AVG(SedentaryMinutes), 2) AS avg_sedentary,
    ROUND(AVG(Calories), 2) AS avg_calories
FROM proud-skein-437112-c4.Fitbit_health_data.categorized_activity
GROUP BY DayType;
```

Results:

Row	DayType	avg_steps	avg_very_active	avg_fairly_active	avg_lightly_active	avg_sedentary	avg_calories
1	Weekend	6434.34	16.44	10.69	176.84	995.34	2207.95
2	Weekday	6528.61	16.35	14.08	166.75	995.06	2173.21

Conclusion: Not much difference in the average steps and other metrics

SLEEP DAY

My goal here is to analyse the Sleep Quality, hence assess the time users spend in bed versus the time they are sleeping. This gives us the Sleep Efficiency calculated as below:

We define **Sleep Efficiency (%)** as:

Sleep Efficiency = (TotalMinutesAsleep / TotalTimeInBed) × 100

We can segment users into different categories:

- **High Sleep Efficiency (≥85%)** – Good sleep habits
- **Moderate Sleep Efficiency (70-85%)** – May need improvement
- **Low Sleep Efficiency (<70%)** – Poor sleep quality

```
SELECT
    Id,
    AVG(TotalTimeInBed) AS avg_time_in_bed,
    AVG(TotalMinutesAsleep) AS avg_minutes_asleep,
    AVG(TotalSleepRecords) AS avg_sleep_records,
    ROUND((AVG(TotalMinutesAsleep) / NULLIF(AVG(TotalTimeInBed), 0)) * 100, 2) AS
avg_sleep_efficiency,
    CASE
        WHEN ROUND((AVG(TotalMinutesAsleep) / NULLIF(AVG(TotalTimeInBed), 0)) * 100, 2) >= 85
             AND AVG(TotalSleepRecords) < 2 THEN 'High Efficiency - Uninterrupted'
        WHEN ROUND((AVG(TotalMinutesAsleep) / NULLIF(AVG(TotalTimeInBed), 0)) * 100, 2)
BETWEEN 70 AND 84.99
             AND AVG(TotalSleepRecords) BETWEEN 2 AND 4 THEN 'Moderate Efficiency - Some
Fragmentation'
        ELSE 'Low Efficiency - Fragmented Sleep'
    END AS SleepCategory
FROM proud-skein-437112-c4.Fitbit_health_data.sleepDay2
GROUP BY Id;
```

Sample Output:

Row	Id	avg_time_in_bed	avg_minutes_asleep	avg_sleep_records	avg_sleep_efficiency	SleepCategory
1	1503960366	383.2000000000...	360.2799999999...	1.080000000000...	94.02	High Efficiency - Uninterrupted
2	1644430081	346.0	294.0	1.0	84.97	Low Efficiency - Fragmented Sl...
3	1844505072	961.0	652.0	1.0	67.85	Low Efficiency - Fragmented Sl...
4	1927972279	437.8	417.0	1.6	95.25	High Efficiency - Uninterrupted
5	2026352035	537.6428571428...	506.1785714285...	1.0	94.15	High Efficiency - Uninterrupted
6	2320127002	69.0	61.0	1.0	88.41	High Efficiency - Uninterrupted
7	2347167796	491.3333333333...	446.8	1.0	90.94	High Efficiency - Uninterrupted
8	3977333714	461.1428571428...	293.6428571428...	1.142857142857...	63.68	Low Efficiency - Fragmented Sl...
9	4020332650	379.75	349.375	1.0	92.0	High Efficiency - Uninterrupted
10	4319703577	501.9615384615...	476.6538461538...	1.038461538461...	94.96	High Efficiency - Uninterrupted
11	4388161847	426.2083333333...	403.125	1.291666666666...	94.58	High Efficiency - Uninterrupted

Related business strategy mentioned later.

5. Share Phase

We answered some of the business questions and most of the business task, which was to analyse the limited dataset to understand scopes of improvement in marketing strategy, which would help Bellabeat grow its customer base. We found some correlations between user's activity levels and their calories, which helped us categorise them into different lifestyles- which would help us for targeted ads and recommendations. We can add features in the app that would label their activities and hence send cookies that will determine their next recommendations.

- My findings give enough new data that we can use further to make newer products and improve strategy for our existing products.
- The data and insights found here have to be compared against the user data from Bellabeat's own tracking system, which would help us understand the differences and fill in gaps better.
- The data visualizations are given above and effective dashboards can be created on other platforms like Tableau with the given output csv files and tables.
- The communication with stakeholders become easy with this one document and a presentation about which metrics we considered and what our next actionable steps might look like.

6. Act Phase

Business Insights & Marketing Strategy Suggestions

I. According to the Sleep Quality Data obtained above, here are some high-level recommendations.

Targeted Product Marketing:

- **Users with Low Efficiency (<70%)**: Promote **sleep aids, melatonin supplements, smart pillows, or relaxation apps**.
- **Moderate Efficiency (70-85%)**: Suggest **better bedding, noise machines, or blue-light blocking glasses**.
- **High Efficiency (≥85%)**: Encourage maintenance with **subscription-based wellness programs**.

Personalized Sleep Reports:

- Offer **customized sleep reports** via email or app notifications.
- Include **tips & recommended products** based on their sleep efficiency trends.

Loyalty & Referral Programs:

- Reward users for **consistent sleep tracking** and engagement with related wellness content.
- Provide discounts on **smart sleep devices** for frequent app users.

II. Targeting moderately active users established before in previous section.

III. Creating Alerts for people with Cardiovascular Risk:

- We can add features like small vibrations to the watch if the heart rate goes beyond a certain limit and there is no activity recorded or no steps.
- For people with these alerts on, we can device another program for better weight management.

IV. Address Sedentary Behavior:

- Wellness Nudges: Bellabeat app can integrate “active break” reminders or tips to reduce sitting time—features that not only encourage movement but also support overall wellness.

V. Use these insights for new Product Launches:

- Leverage Usage Trends: Use insights from the Fitbit data as a starting point to explore how Bellabeat’s own products can fill gaps in the market.
- Supplement with New Data: Recognize the limitations of the current dataset (small sample size, outdated, and lacking demographic details). Bellabeat should consider investing in new, more comprehensive studies to confirm these trends and further refine product features and marketing strategies.

Data Limitations & Next Steps

- Sample Size: The dataset’s small sample and age (2016) mean the trends may not fully represent current or broader consumer behavior. To target the female audience is Bellabeat’s goal, so we need to obtain proper demographic data and then make concrete decisions.
- Additional Research: It is advisable to integrate additional, up-to-date data sources—ideally with broader demographic coverage—to validate and enrich these insights before large-scale strategy shifts. Wider range of dates and differences between male and female patterns should also be observed.