This project was completed as part of the Google Data Analytics Capstone course offered through Coursera. I Huynh Kim Long, followed the steps of the data analysis process—ask, prepare, process, analyze, share, and act—to solve Case Study 2: ***How Can a Wellness Technology Company Play It Smart?*** using common tools such as Microsoft Excel, Google Sheets, and BigQuery

1. Ask Phase

a. Background

As a junior data analyst at Bellabeat marketing analytics team, I am tasked with analyzing smart device usage data to uncover insights into consumer behavior. My findings will be used to guide marketing strategies for Bellabeat's products and will be presented to the executive team for strategic decision-making.

Bellabeat, founded by Urška Sršen and Sando Mur, is a high-tech company specializing in health-focused smart products for women. Leveraging Sršen's artistic background, Bellabeat combines design with technology to empower women through data on activity, sleep, stress, and reproductive health. Since its founding in 2013, the company has expanded globally, with products available through online retailers and its own e-commerce site. Bellabeat heavily invests in digital marketing, including Google Search, social media, and video ads. Sršen has requested an analysis of smart device usage data to uncover growth opportunities and guide marketing strategies.

b. Products

Bellabeat offers a range of products designed to enhance wellness through data such as:

- The Bellabeat app provides users with insights on activity, sleep, stress, menstrual cycles, and mindfulness, syncing with their smart devices.
- The Leaf tracker can be worn as a bracelet, necklace, or clip to monitor activity, sleep, and stress.
- The Time watch combines classic design with smart technology to track wellness metrics, while the Spring water bottle ensures proper hydration by tracking daily water intake.
- The subscription-based membership for personalized guidance on various aspects of health and wellness.

c. Business Task

Analyze smart device usage trends to understand how consumers interact with non-Bellabeat smart devices. Apply these insights to develop recommendations for marketing strategies specifically tailored for Bellabeat's products.

2. Prepare Phase

Sršen suggests using a Kaggle dataset (**FitBit Fitness Tracker Data** (CC0: Public Domain, dataset made available through **Mobius**) that contains minute-level fitness tracker data from 30 Fitbit users, covering physical activity, heart rate, and sleep monitoring. She notes that while this dataset is valuable for exploring daily habits, it may have limitations and recommends considering additional data to address these as I continue my analysis

a. **Description data sources**

This dataset (file name: mturkfitbit_export_4.12.16-5.12.16), including 18 csv files, collected from a survey conducted via Amazon Mechanical Turk between April 12, 2016 – May 12, 2016, includes minute-level data from 30 Fitbit users.

I used Microsoft Excel and Fitbit Data Dictionary downloaded from the Kaggle page. I found the dataset includes data from 33 users, tracking physical activity, sleep time, weight information, and step counts. It can be parsed by session ID or timestamp and reflects variations in tracking behaviors and preferences due to different Fitbit device types used. However, with a sample size of just 33 participants, there are several considerations:

- A sample of 33 participants may not be large enough to accurately represent the broader population of Fitbit users. This can lead to sampling bias and limit the generalizability of findings.
- If the participants are self-selected (e.g., they chose to share their data), they might be more enthusiastic or frequent users of Fitbit devices, which can skew results.
- Small sample sizes can result in higher variability and less reliable estimates of patterns or trends. Individual differences among the 33 users may have a significant impact on the overall findings.
- With fewer samples, the statistical power of analyses might be reduced, making it harder to detect significant effects or relationships.
- The lack of demographic details, particularly the absence of a female gender indicator given Bellabeat's focus on women, and the outdated nature of the data further complicate the interpretation and relevance of the findings.

To align with Bellabeat's business goals, the specific business task, and the credibility of the data, I focused on analyzing criteria such as active users, the duration of smart device usage, and preferred features like heart monitoring, sleep tracking, and step count. I selected the following files for my analysis:

- File 'dailyActivity_merged' contains daily logs for 33 users, detailing three activity types—light, fairly active, and very active—along with the distance covered and time spent on each. Distances are based on step data provided in kilometers, and sedentary time is recorded separately. The dataset also includes total steps taken and calories burned.
- File 'hourlySteps_merged' provides hourly step counts for 33 users, categorized in a 24-hour format. However, there's a variance between the total steps recorded here and in the 'dailyActivity_merged' dataset, likely due to differences in device usage. Consequently, this case study will use this dataset only for analyzing steps by time of day.
- File 'sleepDay_merged' includes data for 24 users, detailing total minutes spent asleep and minutes spent in bed but not asleep

For this project, I use Microsoft Excel and BigQuery_unbilling version for data cleaning. I begin by checking all datasets for common issues such as: irrelevant data, data blank spaces, duplicates, and other inconsistencies. The following steps were taken within each dataset:

- Added underscores between words in column names

- Added column ‘Date’ through date function Changing ‘DateTime’
- Sorted and filtered the dataset by ID to determine the number of unique users (by Filter and Unique records only on Data tab).
- Used Excel’s ‘Remove Duplicates’ tool to check and eliminate duplicate entries.
- Formatted date data into the MM/DD/YY format.
- Applied number formatting to all numerical data, ensuring either no decimals or up to two decimal places.
- Sorted the data by date to identify the first and last recorded dates, revealing a 31-day activity period.
- Split the DateTime column into separate Date, Hour and Time of day columns for easier analysis by using the ‘Text to Columns’ tool.
- Standardized time data by formatting it into the 00:00:00 format.
- Verified ID entries and other columns using the LEN function to ensure data consistency and uniform length.
- Verified the length of text using the LEN function to ensure it was 10 characters long.

After completing the cleaning process, only 3 rows of duplicate information were identified in the 'sleepDay_merged'. These duplicates were removed before proceeding with the analysis in the next Phase..

3. Process Phase

In this analysis, I will focus on using BigQuery SQL and Google Sheets to create data visualizations for the stakeholders.

I opened the BigQuery Console, selected 'Create Project,' and entered the project name: ‘`my-project-gda-cs02-bellabeat`’, then created a new dataset for Bellabeat, named it 'bellabeat_data,' and imported three (03) file csv as I mentioned above into the Bellabeat dataset. I started my work by following:

a. Verify that all IDs have the same length

```
SELECT Id
FROM `my-project-gda-cs02-bellabeat.bellabeat_data.dailyActivity_merged_1`
WHERE
  LENGTH(CAST(Id as STRING)) > 10 OR LENGTH(CAST(Id as String)) < 10
SELECT Id
 FROM `my-project-gda-cs02-bellabeat.bellabeat_data.hourlySteps_merged`
WHERE
  LENGTH(CAST(Id as STRING)) > 10 OR LENGTH(CAST(Id as String)) < 10
```

```
Query results
There is no data to display.
```

b. Clean the data

- **Finding Duplicates**

```
SELECT Id, Date,
  COUNT(*) as num_of_id
FROM `my-project-gda-cs02-bellabeat.bellabeat_data.dailyActivity_merged_1`
GROUP BY
  Id, Date
HAVING
  num_of_id > 1
```

Query results
There is no data to display.

```
SELECT Id,Date,
  COUNT(*) as num_of_id
FROM `my-project-gda-cs02-bellabeat.bellabeat_data.hourlySteps_merged`
GROUP BY
  Id, Date
HAVING
  num_of_id > 24
--24 hours in a day
```

Query results
There is no data to display.

```
SELECT Id,Date,
  COUNT(*) as num_of_id
FROM `my-project-gda-cs02-bellabeat.bellabeat_data.sleepDay_merged`
GROUP BY
  Id, Date
HAVING
  num_of_id > 1
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON

Row	Id	Date	num_of_id
1	8378563200	2016-04-25	2
2	4388161847	2016-05-05	2
3	4702921684	2016-05-07	2

After identifying 3 duplicate rows in the 'sleepDay_merged' dataset, I created a new table named sleepDay_merged_new and removed the duplicates. Similarly, duplicate rows in the 'sleepDay_merged' table were also removed. The new 'sleepDay_merged_new' table now contains only distinct values.

```
CREATE or REPLACE TABLE
`my-project-gda-cs02-bellabeat.bellabeat_data.sleepDay_merged_new`
AS SELECT *
FROM
(
  SELECT *,
  ROW_NUMBER()
  OVER (PARTITION BY Id, Date)
  row_number
  FROM `my-project-gda-cs02-bellabeat.bellabeat_data.sleepDay_merged`
)
WHERE row_number = 1
```

Query results

This statement created a new table named sleepDay_merged_new.

Afterward, I verified the new table to ensure that all duplicates had been successfully removed.

```
SELECT Id,Date,
  COUNT(*) as num_of_id
FROM `my-project-gda-cs02-bellabeat.bellabeat_data.sleepDay_merged_new`
GROUP BY
  Id, Date
HAVING
  num_of_id > 1
```

Query results

There is no data to display.

- **Remove irrelevant data**

During the checking and cleaning process, I discovered some zero (0) values in the columns of the 'dailyActivity_merged_1' dataset, particularly in the Total_Steps column. Since I used the BigQuery unbilled version, which does not allow Data Manipulation Language (DML) queries like INSERT, UPDATE, DELETE, or MERGE, I addressed this by creating and replace `dailyActivity_merged_1` dataset:

```
CREATE OR REPLACE TABLE
`my-project-gda-cs02-bellabeat.bellabeat_data.dailyActivity_merged_1` AS
SELECT *
FROM `my-project-gda-cs02-bellabeat.bellabeat_data.dailyActivity_merged_1`
WHERE NOT
(
  Total_Steps = 0
);
```

Then recheck again:

```
SELECT
  Id,
  Count(*) as num_of_zero_steps
```

```
FROM `my-project-gda-cs02-bellabeat.bellabeat_data.dailyActivity_merged_1`
WHERE
  Total_Steps = 0
GROUP BY Id
ORDER BY num_of_zero_steps
```

Query results

There is no data to display.

- **Find null data**

```
SELECT *
 FROM `my-project-gda-cs02-bellabeat.bellabeat_data.dailyActivity_merged_1`
 WHERE Id IS NULL

SELECT *
FROM `my-project-gda-cs02-bellabeat.bellabeat_data.sleepDay_merged_new`
WHERE Id IS NULL

SELECT *
FROM `my-project-gda-cs02-bellabeat.bellabeat_data.hourlySteps_merged`
WHERE Id IS NULL
```

Query results

There is no data to display.

4. Analyze Phase

Before conducting the analysis, I identify global trends in wearable smart devices and key indicators related to smart device user activity to guide my approach.

The global wearable technology market [1] was valued at USD 61.30 billion in 2022 and is projected to grow at a CAGR of 14.6% from 2023 to 2030. This growth is driven by the increasing adoption of smart wearables for health monitoring, with companies like Fitbit, Samsung, Noise, and Fossil introducing advanced devices. For example, Xiaomi's Watch S1 Series, launched in March 2022, features 117 fitness modes, tracks multiple health metrics, and integrates with Amazon Alexa. The rise in consumer spending on personal care, combined with the demand from athletes for advanced features, is expected to further boost market growth. However, the market faces challenges from counterfeit products and low-cost fraudulent devices, which may impede overall industry expansion.

The U.S. wearable technology market, valued at USD 19.92 billion in 2023, is projected to grow at a CAGR of 12.8% from 2024 to 2030, accounting for 27.7% of the global market. This growth is driven by increasing health consciousness and a rising demand for entertainment. Leading companies like Apple, Alphabet, and Fitbit are advancing the market with devices that monitor vital signs such as heart rate, oxygen levels, and blood pressure, leading to wider adoption and further market expansion.

---

1 https://www.grandviewresearch.com/industry-analysis/wearable-technology-market#:~:text=Report%20Overview,consumers%20is%20driving%20industry%20growth.

Based on the study revealing wearable device trends among U.S. adults[2], here is some additional interesting information:

- **Usage Rates and Patterns:**
    - Approximately 49% of U.S. adults using wearable devices report daily use. This rate is lower for those with cardiovascular disease (CVD) at 38%, compared to 48% among individuals at risk for CVD.
    - About 12% of all U.S. adults using wearable devices have not used them in the past month.
    - Among users with access to wearable devices, 25% of those with CVD and 13% of at-risk individuals reported not using their devices in the preceding month.
    - In a national study, about 72 million U.S. adults (one-third) use wearable devices, but usage is significantly lower among those with CVD (18% reporting use).
- **Willingness to Share Data:**
    - 82% of U.S. adults using wearable devices are willing to share their health data with their physician, which equates to about 57.5 million people.
    - Willingness to share data is slightly higher among individuals with CVD (83%) and those at risk for CVD (81%).
    - There are no significant differences in willingness to share data across key demographic, clinical, or socioeconomic subgroups.
- **Barriers and Challenges:**
    - Factors such as being 65 or older, having only a high school education, and low household income are associated with lower wearable device use.
    - Key barriers to broader device use include cost and technological accessibility, particularly for older adults and those from low-income households.
    - Coverage by health insurance providers could improve device adoption, but evidence supporting their role in disease management is needed.
    - Lower use among individuals with less education and older adults indicates a need for user interface improvements and targeted training.
- **Potential for Improvement:**
    - One in four individuals with CVD who own a wearable device did not use it in the past month, suggesting ownership alone is insufficient for health benefits.
    - Social incentive-based gamification strategies and compliance-promoting initiatives need empirical evaluation to enhance device use and integration into health interventions.
    - Despite the potential of wearables to enhance cardiovascular health management, they are underused among those with CVD compared to the general population.

[2] https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2805753

The national public health agency in the United States, the Centers for Disease Control and Prevention (CDC) [3] have recommendations for adults:

- Physical activity is one of the most important things you can do for your health.
- Adults need at least 150 minutes of moderate-intensity physical activity a week, such as 30 minutes a day, 5 days a week.
- Adults also need 2 days of muscle-strengthening activity each week

As reported by healthcare media publishing company, MedicineNet[4], pedometers classify activity as follows:

- Sedentary: Less than 5,000 steps daily
- Low active: About 5,000 to 7,499 steps daily
- Somewhat active: About 7,500 to 9,999 steps daily
- Active: More than 10,000 steps daily
- Highly active: More than 12,500 steps daily

Based on these figures above, I categorize users based on their daily number of steps as follows:

- Sedentary - Less than 5000 steps a day.
- Lightly active - Between 5000 and 7499 steps a day.
- Fairly active - Between 7500 and 9999 steps a day.
- Very active - More than 10000 steps a day. Classification has been made per the

Findings from the National Institutes of Health, part of the U.S. Department of Health and Human Services [5] Experts recommend that adults sleep between 7 and 9 hours a night. Adults who sleep less than 7 hours a night may have more health issues than those who sleep 7 or more hours a night. Sleeping more than 9 hours a night is not necessarily harmful and may be helpful for young adults, people who are recovering from sleep deprivation, and people who are sick. For these reasons, I recommend a target sleep duration of 8 hours for adults in my analysis.

I started my analysis by BigQuery as below:

- **Categorize the type of users (User Level)**

```
WITH
  daily_average AS (
  SELECT
    Id,
```

[3] https://www.cdc.gov/physical-activity-basics/guidelines/adults.html?CDC_AAref_Val=https://www.cdc.gov/physical activity/basics/adults/index.htm

[4] https://www.medicinenet.com/how_many_steps_a_day_is_considered_active/article.htm

[5] https://www.nhlbi.nih.gov/health/sleep/how-much-sleep#:~:text=Language%20switcher&text=Experts%20recommend%20that%20adults%20sleep,or%20more%20hours%20a%20night.

```
    AVG(Total_Steps) AS totalsteps_mean,
  FROM
    `my-project-gda-cs02-bellabeat.bellabeat_data.dailyActivity_merged_1`
  GROUP BY
    Id
  ORDER BY
    totalsteps_mean
),
--Categorize each user based on their usage level
 users AS (
SELECT
  Id,
  AVG(totalsteps_mean) as avg_total_steps,
  CASE
  WHEN AVG(totalsteps_mean) < 5000 THEN 'Sedentary'
  WHEN AVG(totalsteps_mean) BETWEEN 5001 AND 7500 THEN 'Lightly Active'
  WHEN AVG(totalsteps_mean) BETWEEN 7501 AND 10000 THEN 'Fairly Active'
  WHEN AVG(totalsteps_mean) > 10000 THEN 'Very Active'
  END AS user_level
FROM daily_average
GROUP BY
  Id
ORDER BY avg_total_steps
),
 user_level_counts AS (
    SELECT user_level, COUNT(*) AS total
    FROM users
    GROUP BY user_level
  ),
  total_user_level_counts AS (
    SELECT SUM(total) AS total_user_level
    FROM user_level_counts
  ),
  user_level_percentages AS (
    SELECT user_level, CAST(total AS FLOAT64) /
total_user_level_counts.total_user_level AS total_percent
    FROM user_level_counts, total_user_level_counts
    WHERE 1 = 1
  )
SELECT user_level,
total_percent,
FROM user_level_percentages
```

Query results

JOB INFORMATION | RESULTS | CHART | JS

Row	user_level	total_percent
1	Lightly Active	0.272727272727...
2	Sedentary	0.212121212121...
3	Very Active	0.212121212121...
4	Fairly Active	0.303030303030...

I exported the query results to a CSV file to create a chart in the next phase.

- **Determine the number of steps and sleep duration for the user in week**

```
WITH
-- Join two (02) tables
  daily_activity_sleep AS
  (
    SELECT
      dailyActivity_merged_1.Total_Steps,
      sleepDay_merged.Total_Minutes_Asleep,
      dailyActivity_merged_1.Id AS id,
      dailyActivity_merged_1.Date AS date
    FROM `my-project-gda-cs02-bellabeat.bellabeat_data.dailyActivity_merged_1`
AS dailyActivity_merged_1
    INNER JOIN
      `my-project-gda-cs02-bellabeat.bellabeat_data.sleepDay_merged` AS
sleepDay_merged
    ON
      dailyActivity_merged_1.Id = sleepDay_merged.Id AND
      dailyActivity_merged_1.Date = sleepDay_merged.Date
  )
-- Calculate the weekly averages for total steps and total asleep per week
SELECT
  day_of_week,
  ROUND(AVG(Total_Steps), 2) AS ave_totalsteps_perday,
  ROUND(AVG(Total_Minutes_Asleep), 2) AS ave_minutesasleep_perday
FROM
  (
    SELECT *,
      CASE
        WHEN (EXTRACT(DAYOFWEEK FROM date) = 1) THEN 'Mon'
        WHEN (EXTRACT(DAYOFWEEK FROM date) = 2) THEN 'Tue'
        WHEN (EXTRACT(DAYOFWEEK FROM date) = 3) THEN 'Wed'
        WHEN (EXTRACT(DAYOFWEEK FROM date) = 4) THEN 'Thu'
        WHEN (EXTRACT(DAYOFWEEK FROM date) = 5) THEN 'Fri'
        WHEN (EXTRACT(DAYOFWEEK FROM date) = 6) THEN 'Sat'
        WHEN (EXTRACT(DAYOFWEEK FROM date) = 7) THEN 'Sun'
      END AS day_of_week
    FROM daily_activity_sleep
  )
```

```
GROUP BY day_of_week;
```

Query results

JOB INFORMATION RESULTS CHART JSON EXECUTION

Row	day_of_week	ave_totalsteps_perda	ave_minutesasleep_p
1	Wed	9182.69	404.54
2	Thu	8022.86	434.68
3	Fri	8205.35	402.37
4	Sat	7901.4	405.42
5	Sun	9948.69	420.81
6	Mon	7297.85	452.75
7	Tue	9339.85	418.83

I exported the query results to a CSV file to create a chart in the next phase.

- **Calculate usage Rates**

To understand device usage, I categorized users based on their device usage over a 31-day period:

- High Use: Users who use their device 21-31 days.
- Moderate Use: Users who use their device 11-20 days.
- Low Use: Users who use their device 01-10 days.

```
WITH
  -- Join two tables with primary keys
  daily_activity_and_sleep AS (
    SELECT
      daily_activity.Id AS Id,
      COUNT(*) AS num_of_use
    FROM
      `my-project-gda-cs02-bellabeat.bellabeat_data.dailyActivity_merged_1` AS
daily_activity
    INNER JOIN
      `my-project-gda-cs02-bellabeat.bellabeat_data.sleepDay_merged_new` AS
daily_sleep
    ON
      daily_activity.Id = daily_sleep.Id
      AND daily_activity.Date = daily_sleep.Date
    GROUP BY
      daily_activity.Id
  ),
  -- Filtering user data based on daily sleep and activity patterns
  usages AS (
    SELECT
      Id,
      SUM(num_of_use) AS day_used,
      CASE
        WHEN SUM(num_of_use) BETWEEN 1 AND 10 THEN 'low use'
```

```
        WHEN SUM(num_of_use) BETWEEN 11 AND 20 THEN 'moderate use'
        WHEN SUM(num_of_use) BETWEEN 21 AND 31 THEN 'high use'
      END AS usage
    FROM daily_activity_and_sleep
    GROUP BY Id
  ),
  -- Counting the frequency of device usage
  usage_summary AS (
    SELECT
      usage,
      COUNT(*) AS total
    FROM usages
    GROUP BY usage
  ),
  -- Calculating the average frequency and total duration of device usage
  usage_percentage AS (
    SELECT
      usage,
      total,
      total_usage,
      CAST(total AS FLOAT64) / total_usage AS total_percentage
    -- Selecting it FROM usage summary, and finding the total usage
    FROM (
      SELECT
        usage,
        total,
        SUM(total) OVER () AS total_usage
      FROM usage_summary
    )
  )
SELECT
  usage,
  total_percentage,
  CONCAT(CAST(ROUND(total_percentage * 100, 1) AS STRING), '% (', CAST(total
AS STRING), ')') AS labels
FROM usage_percentage;
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS

Row	usage	total_percentage	labels
1	high use	0.5	50% (12)
2	low use	0.375	37.5% (9)
3	moderate use	0.125	12.5% (3)

I exported the query results to a CSV file to create a chart in the next phase.

- **Determine how long users wear their smart device each day**

```
WITH
  -- Join two tables with primary keys
  daily_activity_and_sleep AS (
    SELECT
      daily_activity.Id AS Id,
      COUNT(*) AS num_of_use
    FROM
      `my-project-gda-cs02-bellabeat.bellabeat_data.dailyActivity_merged_1` AS
daily_activity
    INNER JOIN
      `my-project-gda-cs02-bellabeat.bellabeat_data.sleepDay_merged_new` AS
daily_sleep
    ON
      daily_activity.Id = daily_sleep.Id
      AND daily_activity.Date = daily_sleep.Date
    GROUP BY
      daily_activity.Id
  ),
  -- Filtering user data based on daily sleep and activity patterns
  usages AS (
    SELECT
      Id,
      SUM(num_of_use) AS day_used,
      CASE
        WHEN SUM(num_of_use) BETWEEN 1 AND 10 THEN 'low use'
        WHEN SUM(num_of_use) BETWEEN 11 AND 20 THEN 'moderate use'
        WHEN SUM(num_of_use) BETWEEN 21 AND 31 THEN 'high use'
      END AS usage
    FROM daily_activity_and_sleep
    GROUP BY Id
  ),
  -- Counting the frequency of device usage
  usage_summary AS (
    SELECT
      usage,
      COUNT(*) AS total
    FROM usages
    GROUP BY usage
  ),
  -- Calculating the average frequency and total duration of device usage
  usage_percentage AS (
    SELECT
      usage,
      total,
      total_usage,
      CAST(total AS FLOAT64) / total_usage AS total_percentage
    FROM (
```

```
      SELECT
        usage,
        total,
        SUM(total) OVER () AS total_usage
      FROM usage_summary
    )
  ),
  -- Generating a new subquery for daily device usage
  daily_activity_used AS (
    SELECT
      daily_activity_new.*,
      usages.usage
    FROM
      `my-project-gda-cs02-bellabeat.bellabeat_data.dailyActivity_merged_1` AS
daily_activity_new
    LEFT JOIN
      usages ON daily_activity_new.Id = usages.Id
  ),
  -- Calculating and categorizing minutes worn by users
  minutes_worn AS (
    SELECT *,
      CASE
        WHEN minutes_worn_percentage = 100 THEN 'All day'
        WHEN minutes_worn_percentage >= 50 AND minutes_worn_percentage < 100
THEN 'More than half day'
        WHEN minutes_worn_percentage > 0 AND minutes_worn_percentage < 50 THEN
'Less than half day'
      END AS worn
    FROM (
      SELECT
        *,
        (Very_Active_Minutes + Fairly_Active_Minutes + Lightly_Active_Minutes
+ Sedentary_Minutes) AS total_worn_minutes,
        (Very_Active_Minutes + Fairly_Active_Minutes + Lightly_Active_Minutes
+ Sedentary_Minutes) / 1440 * 100 AS minutes_worn_percentage
      FROM daily_activity_used
    )
  ),
  -- Summarize the distribution of user-worn time into categories
  worn_summary AS (
    SELECT
      worn,
      COUNT(*) AS total
    FROM minutes_worn
    GROUP BY worn
  ),
  -- Calculating percentages for each worn time category
  worn_percentage AS (
    SELECT
```

```
    worn,
    total,
    SUM(total) OVER() AS total_worn
  FROM worn_summary
),
-- High use category
minutes_worn_highuse AS (
  SELECT
    worn,
    total / totals AS total_percentage,
    CONCAT(ROUND(total / totals * 100, 2), '%') AS labels
  FROM (
    SELECT
      worn,
      COUNT(*) AS total,
      SUM(COUNT(*)) OVER () AS totals
    FROM minutes_worn
    WHERE usage = 'high use'
    GROUP BY worn
  )
),
-- Moderate use category
minutes_worn_moderateuse AS (
  SELECT
    worn,
    total / totals AS total_percentage,
    CONCAT(ROUND(total / totals * 100, 2), '%') AS labels
  FROM (
    SELECT
      worn,
      COUNT(*) AS total,
      SUM(COUNT(*)) OVER() AS totals
    FROM minutes_worn
    WHERE usage = 'moderate use'
    GROUP BY worn
  )
),
-- Low use category
minutes_worn_lowuse AS (
  SELECT
    worn,
    total / totals AS total_percentage,
    CONCAT(ROUND(total / totals * 100, 2), '%') AS labels
  FROM (
    SELECT
      worn,
      COUNT(*) AS total,
      SUM(COUNT(*)) OVER() AS totals
    FROM minutes_worn
```

```
        WHERE usage = 'low use'
        GROUP BY worn
      )
    )
-- Final selection for low usage category
SELECT  *
FROM minutes_worn_lowuse;
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS

Row	worn	total_percentage	label
1	More than half day	0.210045662100...	21%
2	Less than half day	0.018264840182...	1.83%
3	All day	0.771689497716...	77.17%

I exported the query results to a CSV file to create a chart in the next phase.

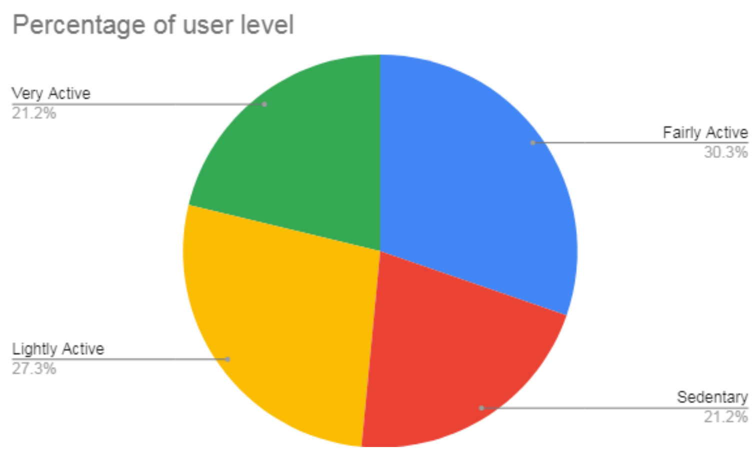
5. **Share Phase**

The summary of my findings are listed below

- **Sample bias:**

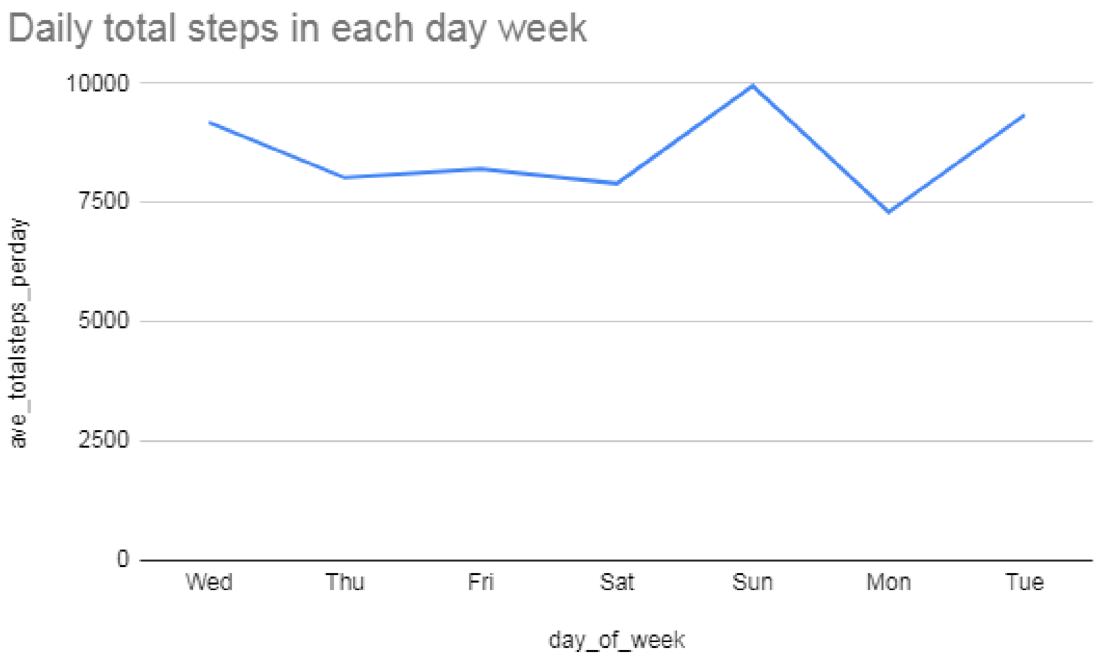
A sample size of 33 may not represent the broader Fitbit user population, risking sampling bias and limited generalizability. Self-selected participants could skew results if they are more enthusiastic users. Small sample sizes lead to higher variability and less reliable patterns, reducing statistical power and making it difficult to detect significant effects. Additionally, missing demographic details and outdated data complicate interpretation.

- **The type of users**

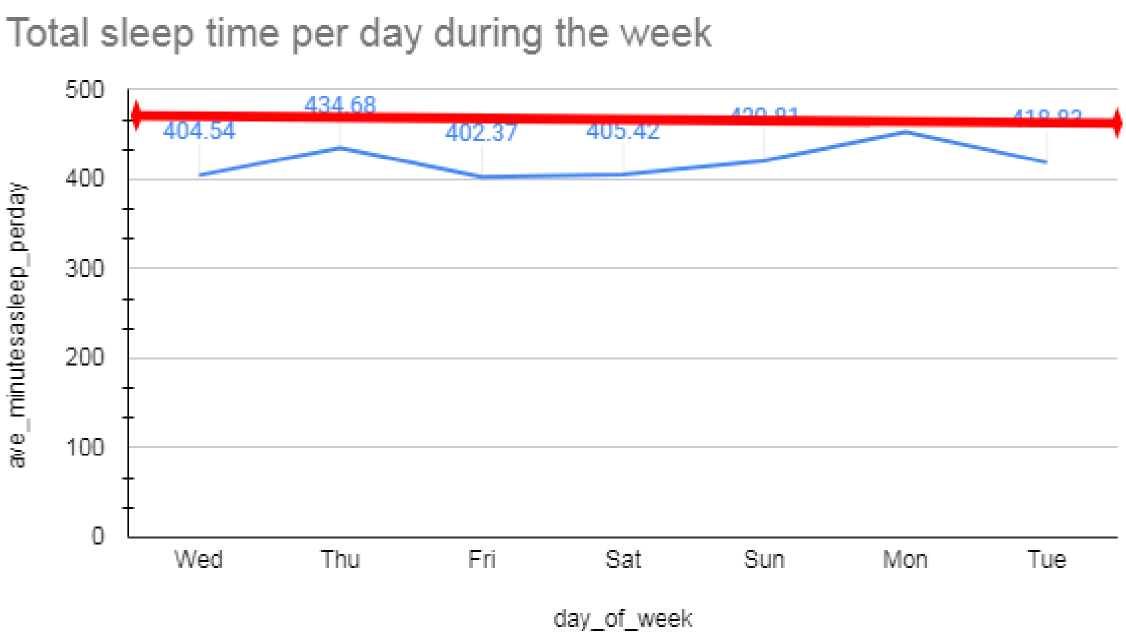


The active users are categorized into three types: very active, fairly active, and lightly active, and together they account for nearly 80% of users. However, the proportion of lightly active and sedentary users remains high, at nearly 50%.

- **The number of steps and sleep duration for the user in week**

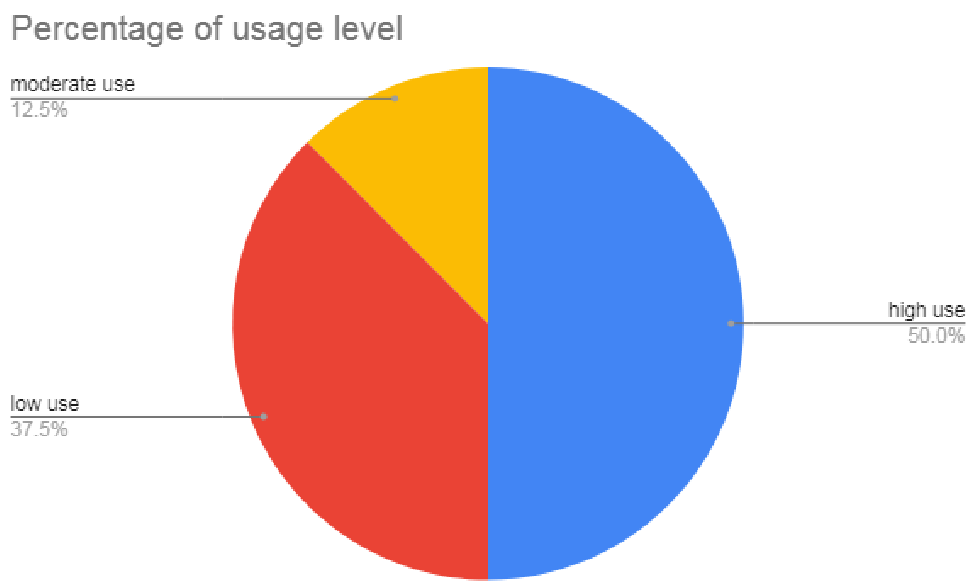


Based on the graphs above, we can conclude users meet the recommended daily step count of 7,500 steps on all days except Mondays and highest steps on Sundays



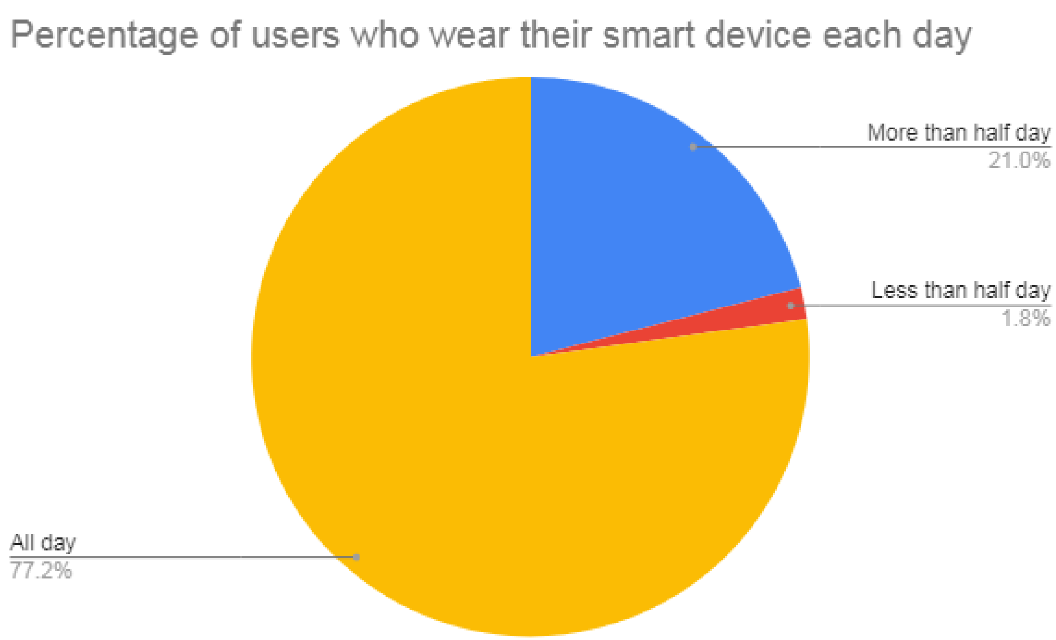
Based on the graphs above, users do not consistently achieve the recommended 8 hours (#480 minutes) of sleep per night.

- **Usage Rates**



The results in the graph show that 50% of users frequently use their device between 21 and 31 days. Additionally, nearly 38% of users balance their smart gadget usage between 11 and 20 days, while 12% of users rarely use their Intelligent device, only 1 to 10 days.

- **How long users wear their smart device each day**



6. **Act phase**

Recommendations for Marketing Strategy

- Create a survey to gather more detailed data for future analysis.
- The proportion of lightly active and sedentary users remains high, at nearly 50%. To address this, we should create a gamification marketing campaign targeted at this segment, encouraging them to use a smart device more frequently to enhance their health.
- Improve sleep tracking devices or apps to monitor sleep patterns, set users goals and give users personalized feedback based on their sleep data, helping them identify patterns and make adjustments to improve their sleep duration.
- Implement gamification strategies by rewarding users for meeting their sleep goals. For example, users could earn points or badges for consistently getting 8 hours of sleep.
- Build a subscription-based membership program that includes content such as infographics, videos, and articles highlighting the crucial benefits of good sleep. The program should also offer tips on optimizing the sleep environment, including reducing screen time before bed, maintaining a consistent sleep schedule, and creating a relaxing bedtime routine.
- Offer programs or challenges that focus on building healthy sleep habits. Provide practical tips and support through online communities or workshops.
- With 50% of users categorized as high-use, we can focus on enhancing the product's design, improving any difficult-to-understand interfaces, addressing battery issues, and refining physical features.
- Collaborate with healthcare providers to develop new health-related applications. These smartphone apps should offer personalized recommendations to support users' fitness goals and promote an active lifestyle.