

Table of Contents

1. Summary	2
2. Ask	2
2.1 Business task	2
2.2 Stakeholders	2
3. Prepare	2
3.1 Data Used	2
3.2 Accessibility and Privacy of data	2
3.3 Data Summary	2
3.4 Data Organization	2
3.5 Data Limitations and Integrity	3
4. Process	3
4.1 Datasets Selected	3
4.2 Use Excel to Clean data	3
4.3 Upload Datasets to SQL	3
5. Analyze and Share	4
5.1 Margining datasets	4
5.2 User verification	5
5.3 Logged users	5
5.4 Users types	6
5.5 Logged user's summary	7
5.6 compare calories, total steps per weekday	9
5.7 compare total steps, total sleep per weekday	9
7. Act	10
7.1 Conclusion & Recommendations	10

1. Summary

Bellabeat is a high-tech manufacturer of health-focused products for women. Bellabeat employees believe that analyzing smart device fitness data could help unlock new growth opportunities for the company and gain insight into how consumers are using their smart devices.

The data was collected on activity, sleep, stress, and reproductive health has allowed Bellabeat to empower women with knowledge about their own health and habits and help them to make health decisions.

2. Ask

2.1 Business task

Analyze smart device Fitbit fitness data to gain insights from the people usage their smart devices, and know the trends of their smart devices and apply it to Bellabeat customers to make growth of Bellabeat marketing strategy.

2.2 Stakeholders

Urška Sršen: Bellabeat's cofounder and Chief Creative Officer.

Sando Mur: Mathematician and Bellabeat's cofounder.

Bellabeat marketing analytics team: A team of data analysts.

3. Prepare

3.1 Data Used

The data source is **FitBit Fitness Tracker data** stored in Kaggle.

3.2 Accessibility and Privacy of data

The metadata published by MÖBIUS, and under public domain, waiving all rights to the work and allowing for the dataset to be copied, modified, distributed and performed without asking for permission

3.3 Data Summary

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. Additionally, "Variation between output represents use of different types of Fitbit trackers and individual tracking behaviors preferences."

3.4 Data Organization

The dataset of **FitBit Fitness Tracker data** includes 18 datasets tracks 33 users during 31 days. Downloaded from Kaggle in .CSV file formats includes long and wide formats.

3.5 Data Limitations and Integrity

The dataset was collected in 2016 which make the dataset outdated, and the dataset was during 03.12.2016 to 05.12.2016 which is during 31 days which make the period of the time very short. It has a limitation of size 33 users. Also, there is a lack of demographic information, which create sample bias.

4. Process

4.1 Datasets Selected

Datasets Selected for the analysis:

- Daily_Activity_Merged
- Daily_Calories_Merged
- Daily_Intensities_Merged
- Daily_Steps_Merged
- Sleep_Day_Merged

4.2 Use Excel to Clean data

Each dataset cleaned using Excel by these steps:

1. Sort the data by Id and filtered to get how many users
2. Checked for duplicated data
3. Split date and time using text to column and use only date for this analysis
4. Convert date to MM/DD/YY format
5. Format numeric data with decimal to decimal up to 2
6. Sort the data by date to get first and last date to find the datasets period of time
7. Checked length for Id column to make sure its correct

The cleaning result is that here was 3 duplicated columns removed and Split date and time in Sleep_Day_Merged table

4.3 Upload Datasets to SQL

Upload the needed FitBit Fitness Tracker tables for this analysis into BigQuery:

Cleaned tables uploaded are:

- Daily_Activity_Merged
- Daily_Calories_Merged
- Daily_Intensities_Merged
- Daily_Steps_Merged
- Sleep_Day_Merged

5. Analyze and Share

5.1 Margining datasets

Join all the selected datasets that are daily and share same column to use if for the analysis

```
SELECT
A.Id,
A.Calories,
* EXCEPT(Id,
Calories,
ActivityDay,
SleepDay,
SedentaryMinutes,
LightlyActiveMinutes,
FairlyActiveMinutes,
VeryActiveMinutes,
SedentaryActiveDistance,
LightActiveDistance,
ModeratelyActiveDistance,
VeryActiveDistance),
I.SedentaryMinutes,
I.LightlyActiveMinutes,
I.FairlyActiveMinutes,
I.VeryActiveMinutes,
I.SedentaryActiveDistance,
I.LightActiveDistance,
I.ModeratelyActiveDistance,
I.VeryActiveDistance
FROM
`unified-surface-365117.Fitbit_data.dailyActivity_merged` A
LEFT JOIN
`unified-surface-365117.Fitbit_data.dailyCalories_merged` C
ON
A.Id = C.Id
AND A.ActivityDate=C.ActivityDay
AND A.Calories = C.Calories
LEFT JOIN
`unified-surface-365117.Fitbit_data.dailyIntensities_merged` I
ON
A.Id = I.Id
AND A.ActivityDate=I.ActivityDay
AND A.FairlyActiveMinutes = I.FairlyActiveMinutes

AND A.LightActiveDistance = I.LightActiveDistance
AND A.LightlyActiveMinutes = I.LightlyActiveMinutes
AND A.ModeratelyActiveDistance = I.ModeratelyActiveDistance
AND A.SedentaryActiveDistance = I.SedentaryActiveDistance
AND A.SedentaryMinutes = I.SedentaryMinutes
AND A.VeryActiveDistance = I.VeryActiveDistance
AND A.VeryActiveMinutes = I.VeryActiveMinutes
```

```

LEFT JOIN
`unified-surface-365117.Fitbit_data.dailySteps_merged` S
ON
A.Id = S.Id
AND A.ActivityDate=S.ActivityDay
LEFT JOIN
`unified-surface-365117.Fitbit_data.sleepDay_merged` Sl
ON
A.Id = Sl.Id
AND A.ActivityDate=Sl.SleepDay;

```

5.2 User verification

Check number of users by count the ids to get number of participants in this analysis

```

SELECT count(distinct Id) AS number_of_users FROM `unified-surface-365117.Fitabase_Data.dailyActivity_merged`

```

Number of the participants in these datasets are 33 participants

5.3 Logged users

Count number of times users use their tracker per month by Ids.

```

SELECT Id, COUNT( Id) AS number_of_logging
FROM `unified-surface-365117.Fitbit_data.Merged_fitbit_data`
GROUP BY Id
ORDER BY number_of_logging DESC

```

- 21 users logged in 31 times
- 3 users logged in 30 times
- 2 users logged in 29 times
- 1 user logged in 28 times
- 2 users logged in 26 times
- 1 user logged in 20 times
- 1 user logged in 19 times
- 1 user logged in 18 times
- 1 user logged in 4 times

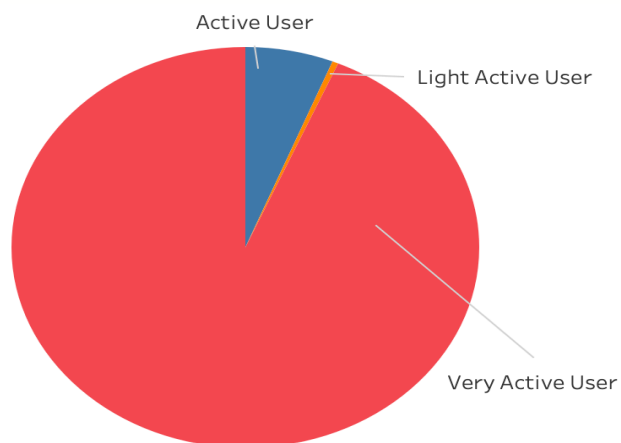
As result 64% of users logged in the whole time, 82% of the users logged in the whole or missed 1 to 2 days.

5.4 Users types

Specify users' types by number of times uses the tracker, because there are no demographics variables in these datasets. Assign users activity depending on the logging during 31 days:

- Very Active user logging 31 to 21 times
- Active User logging 20 to 10 times
- Light Active User logging 9 to 0 times

```
SELECT COUNT(Id) AS number_of_logged,  
CASE  
WHEN COUNT(Id) BETWEEN 21 AND 31 THEN 'Very Active User'  
WHEN COUNT(Id) BETWEEN 10 AND 20 THEN 'Active User'  
WHEN COUNT(Id) BETWEEN 0 AND 9 THEN 'Lightly Active User'  
END Usage_Type  
FROM `unified-surface-365117.Fitbit_data.Merged_fitbit_data`  
GROUP BY Id  
ORDER BY number_of_logged DESC
```

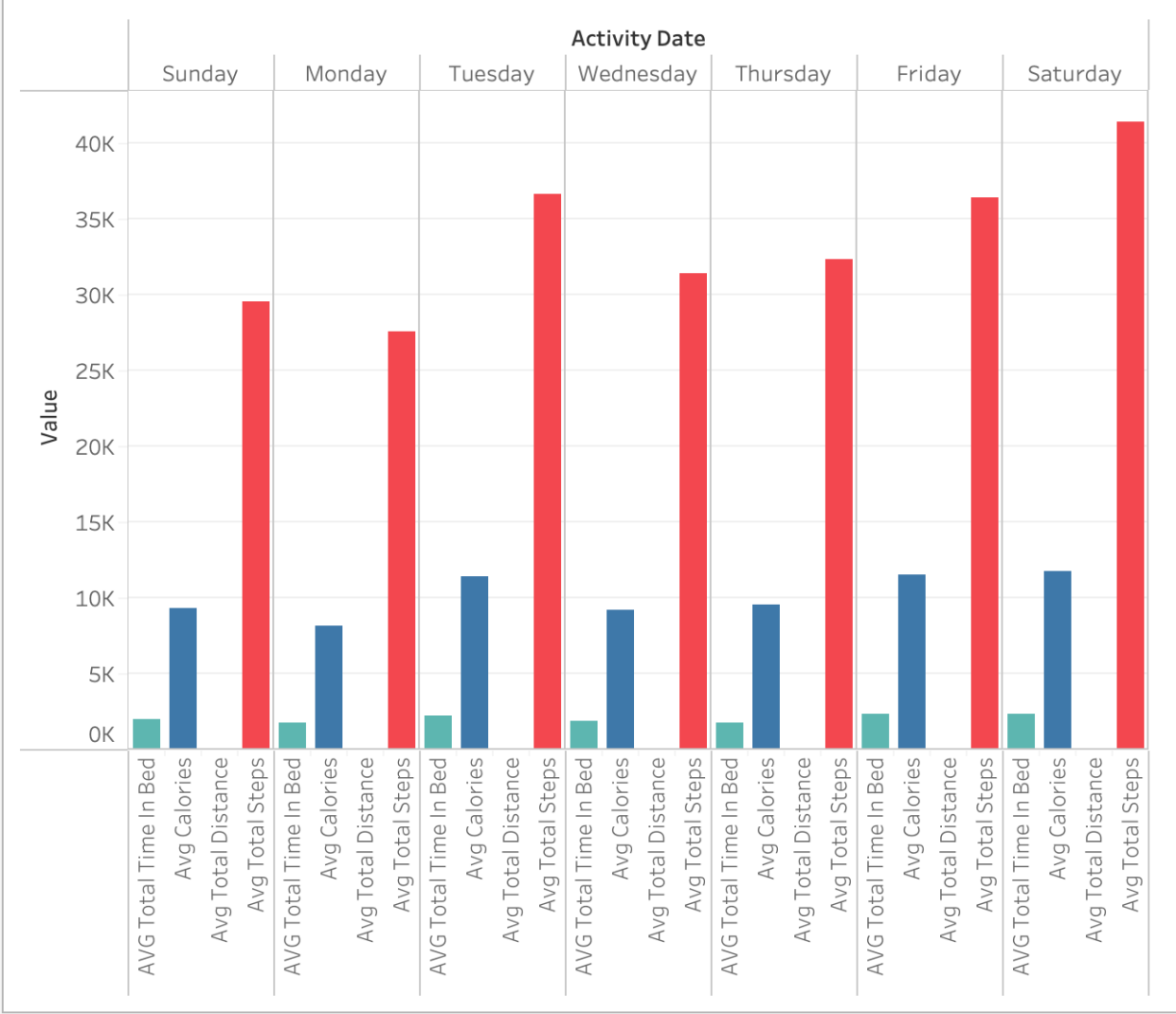


As shown from the pie chart most of the users are very active users, who use the tracker from 31 to 21 times in a month.

5.5 Logged user’s summary

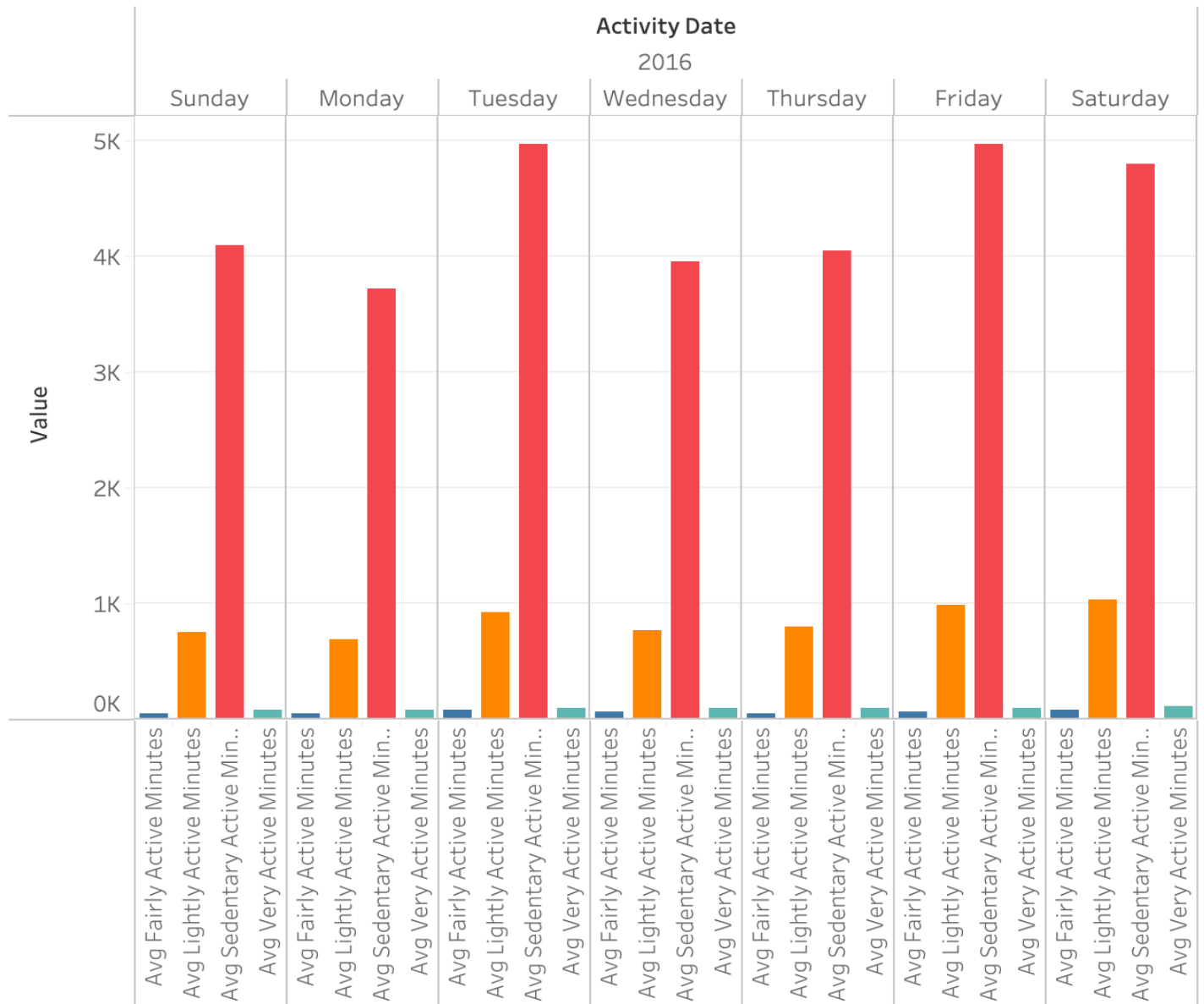
Take a brief look at Average for total steps, total distance, total time in bed, calories per weekday.

```
SELECT ActivityDate,
AVG(TotalSteps) AS Avg_Total_Steps,
AVG(TotalDistance) AS Avg_Total_Distance,
AVG(TotalTimeInBed) AS AVG_Total_Time_In_Bed,
AVG(calories) AS Avg_Calories,
FROM `unified-surface-365117.Fitbit_data.Merged_fitbit_data`
GROUP BY ActivityDate
```



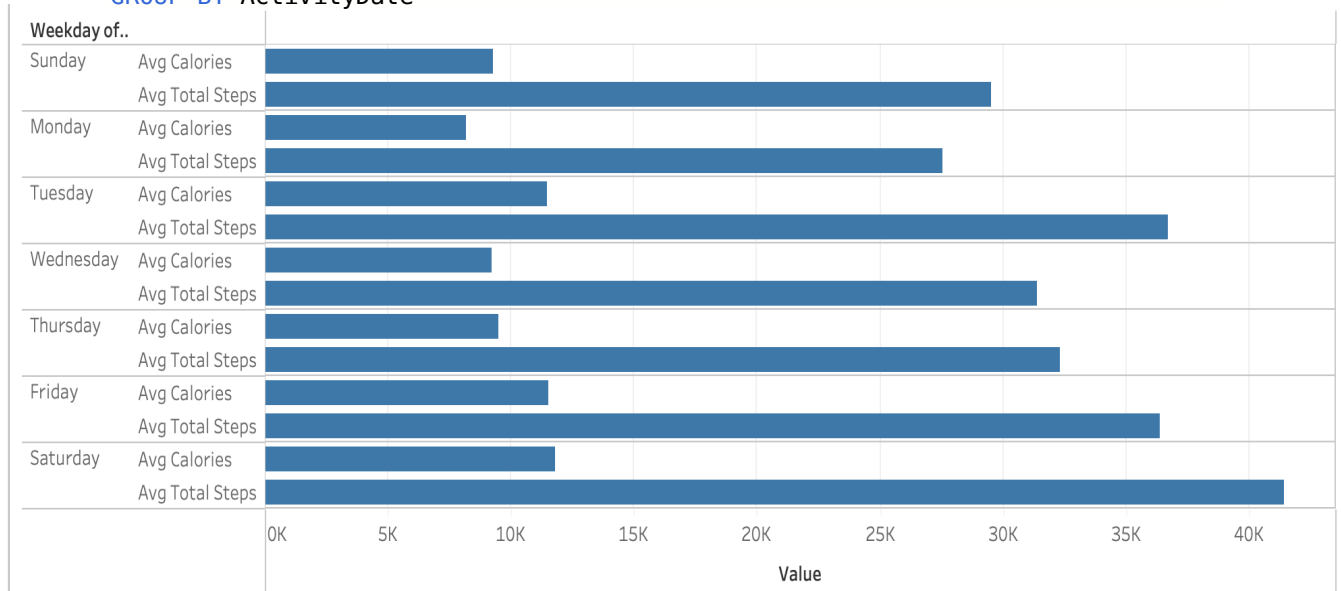
Then, look at all the active minutes types per weekday, count Average of very active minutes, fairly active minutes, lightly active minutes, sedentary active minutes

```
SELECT ActivityDate,
AVG(VeryActiveMinutes) AS Avg_Very_Active_Minutes,
AVG(FairlyActiveMinutes) AS Avg_Fairly_Active_Minutes,
AVG(SedentaryMinutes) AS Avg_Sedentary_Active_Minutes,
AVG(LightlyActiveMinutes) AS Avg_Lightly_Active_Minutes
FROM `unified-surface-365117.Fitbit_data.Merged_fitbit_data`
GROUP BY ActivityDate
```



5.6 compare calories, total steps per weekday

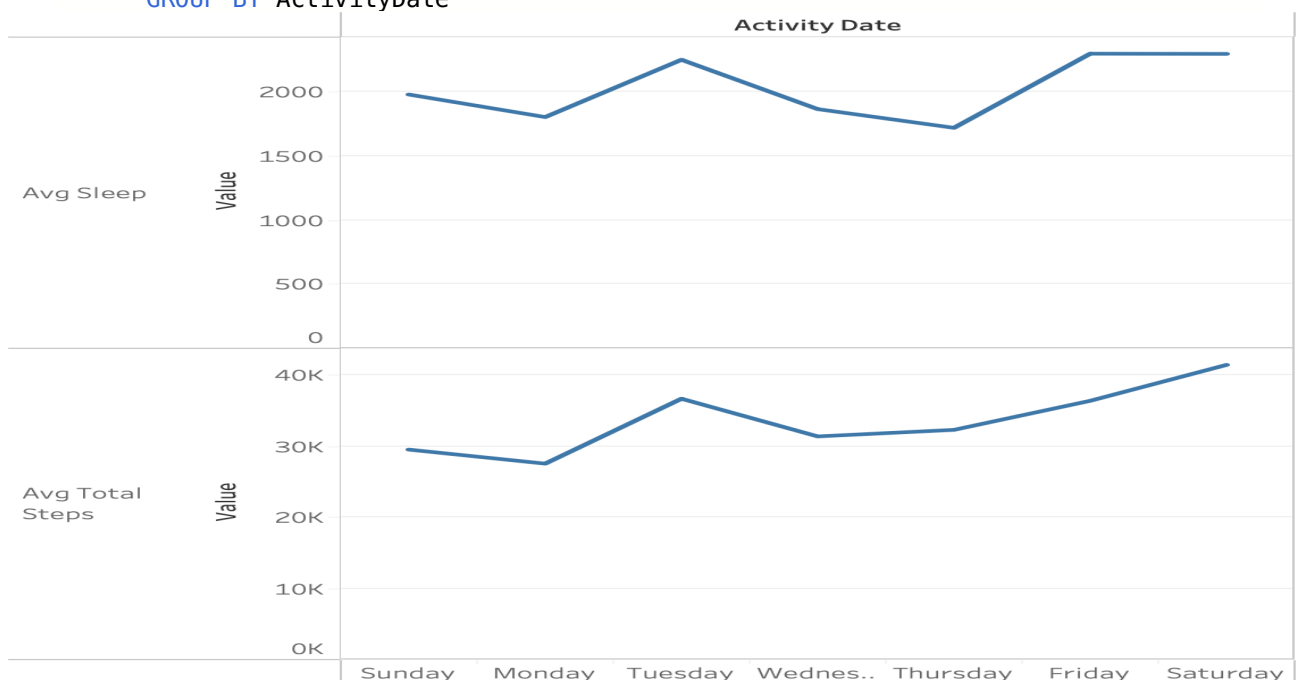
```
SELECT ActivityDate, AVG(TotalSteps) AS Avg_Total_Steps, AVG(Calories) AS Avg_Calories
FROM `unified-surface-365117.Fitbit_data.Merged_fitbit_data`
GROUP BY ActivityDate
```



More steps you take, more calories to burn, each user can check their status for the steps they took each day of week and the calories they burn.

5.7 compare total steps, total sleep per weekday

```
SELECT ActivityDate, AVG(TotalSteps) AS Avg_Total_Steps, AVG(TotalTimeInBed) AS Avg_Sleep
FROM `unified-surface-365117.Fitbit_data.Merged_fitbit_data`
GROUP BY ActivityDate
```



6. Act

6.1 Conclusion & Recommendations

Bellabeat is to empower women to control their health. Based on the result more people use the tracker, more they can control their sleep and calories. I recommend to have demographics about the users in the dataset, and to have recent dataset, because the dataset is old and for a small period of time to prevent the dataset from bias and to have more accurate conclusion.

After analyzing the dataset, I found some insights that may help the Bellabeat to improve.

1. Provide notification for steps:
As a result, most of the people who use the tracker are sedentary active, by providing them a notification every day as a reminder and every 5000 steps. also, provide them an encourage message, so they will continue to reach 10000 steps or more per day.
2. Provide graph analysis about Calories:
How many calories the spent each day, and the goal they want to reach every month, when they reach the goal give them notification.
3. Make a sleep notification:
Provide notification about sleeping time a part of a routine, and a relaxation mode to help them relax and sleep.