# Analysis

I started by working with the daily data tables, beginning with the daily activity table. I selected the calories and steps columns, grouped the data by day, and then calculated statistics such as the average, maximum value, minimum value, and total sum. To do this, I used the following query:

```
SELECT
  FORMAT_TIMESTAMP("%A", Activity.ActivityDate)
AS day,
  SUM(TotalSteps) AS sum_steps,
  MIN(TotalSteps) AS min_steps,
  MAX(TotalSteps) AS max_steps,
  AVG(TotalSteps) AS avg_steps,
  SUM(Calories) AS sum_calories,
  MIN(Calories) AS min_calories,
  MAX(Calories) AS max_calories,
  AVG(Calories) AS avg_calories
FROM
  `bellabeat-
456023.fitbit_fitness_tracker.dailyActivity` AS
Activity
GROUP BY
  FORMAT_TIMESTAMP("%A", Activity.ActivityDate)
ORDER BY
  sum_calories DESC
```

The result of this query produced the following table. The main finding is that Saturdays are the most active days, both in terms of steps taken and calories burned.

| day | Saturday | Tuesday | Friday | Wednesday | Sunday | Monday | Thursday |
|---|---|---|---|---|---|---|---|
| sum_steps | 1520997 | 1449779 | 1346789 | 1397353 | 1254616 | 1310275 | 1309953 |
| min_steps | 14 | 8 | 42 | 8 | 16 | 8 | 7 |
| max_steps | 29326 | 23186 | 23014 | 24136 | 36019 | 20779 | 21129 |
| avg_steps | 8.843 | 7.966 | 7.696 | 8.077 | 7.468 | 8.241 | 8.037 |
| sum_calories | 415357 | 406847 | 406258 | 406212 | 383773 | 369234 | 367727 |
| min_calories | 787 | 50 | 403 | 52 | 489 | 1223 | 257 |
| max_calories | 4547 | 4286 | 4196 | 4079 | 4552 | 4234 | 4236 |
| avg_calories | 2.415 | 2.235 | 23.215 | 2.348 | 2.284 | 2.322 | 2.256 |

Next, I decided to perform the same analysis, but this time distinguishing between weekdays and weekends. The query used for this is shown below.

```
SELECT
  CASE
    WHEN FORMAT_TIMESTAMP("%A",
Activity.ActivityDate) = "Saturday"
      OR FORMAT_TIMESTAMP("%A",
Activity.ActivityDate) = "Sunday" THEN "Weekend"
    ELSE "Weekday"
  END AS dayType,
  SUM(TotalSteps) AS sum_steps,
  MIN(TotalSteps) AS min_steps,
  MAX(TotalSteps) AS max_steps,
  AVG(TotalSteps) AS avg_steps,
  SUM(Calories) AS sum_calories,
  MIN(Calories) AS min_calories,
  MAX(Calories) AS max_calories,
  AVG(Calories) AS avg_calories
FROM
  `bellabeat-
456023.fitbit_fitness_tracker.dailyActivity` AS
Activity
GROUP BY
  CASE
    WHEN FORMAT_TIMESTAMP("%A",
Activity.ActivityDate) = "Saturday"
      OR FORMAT_TIMESTAMP("%A",
Activity.ActivityDate) = "Sunday" THEN "Weekend"
    ELSE "Weekday"
  END
ORDER BY
  avg_calories DESC
```

The results of the query are presented in the following table, where it can be seen that, on average, more steps are taken and more calories are burned on weekends compared to weekdays.

| dayType | Weekend | Weekday |
|---|---|---|
| sum_steps | 2775613 | 6814149 |
| min_steps | 14 | 7 |
| max_steps | 36019 | 24136 |
| avg_steps | 8.164 | 7.998 |
| sum_calories | 799130 | 1956278 |
| min_calories | 489 | 50 |
| max_calories | 4552 | 4286 |
| avg_calories | 2.350 | 2.296 |

Finally, for the daily activity table, the times and distances for the different activity levels were aggregated. The goal was to analyze the users' level of sedentary behavior. The query used is shown below.

```
SELECT
  SUM(TotalDistance) AS totalDistance,
  SUM(TrackerDistance) AS trackerDistance,
  SUM(LoggedActivitiesDistance) AS
loggedActivitiesDistance,
  SUM(VeryActiveDistance) AS veryActiveDistance,
  SUM(ModeratelyActiveDistance) AS
moderatelyActiveDistance,
  SUM(LightActiveDistance) AS lightActiveDistance,
  SUM(SedentaryActiveDistance) AS
sedentaryActiveDistance,
  SUM(VeryActiveMinutes) AS veryActiveMinutes,
  SUM(FairlyActiveMinutes) AS fairlyActiveMinutes,
  SUM(LightlyActiveMinutes) AS lightlyActiveMinutes,
  SUM(SedentaryMinutes) AS sedentaryMinutes

FROM
  `bellabeat-
456023.fitbit_fitness_tracker.dailyActivity` AS
Activity
```

The results of the query are shown below, highlighting that the greatest distance is covered during light activity and that users spend most of their time in a sedentary state.

| | |
|---|---|
| totalDistance | 6.855,63 |
| trackerDistance | 6.854,31 |
| loggedActivitiesDistance | 37,56 |
| veryActiveDistance | 1.835,34 |
| moderatelyActiveDistance | 721,83 |
| lightActiveDistance | 4.286,61 |
| sedentaryActiveDistance | 2,16 |
| veryActiveMinutes | 25.442,00 |
| fairlyActiveMinutes | 17.372,00 |
| lightlyActiveMinutes | 248.471,00 |
| sedentaryMinutes | 1.124.191,00 |

For the sleep table, the data was grouped by each sleep record in order to analyze users' sleep quality and the amount of time they spend either sleeping or lying in bed. The query used is shown below:

```
SELECT
  TotalSleepRecords,
  COUNT(TotalSleepRecords) AS count_SleepRecords,
  SUM(TotalMinutesAsleep) AS minutesAsleepByRecord,
  SUM(TotalTimeInBed) AS timeInBedByRecord,
  AVG(TotalMinutesAsleep) AS avgSleepByRecord,
  AVG(TotalTimeInBed) AS avgInBedByRecord
FROM
  `bellabeat-456023.fitbit_fitness_tracker.sleepDay`
GROUP BY
  TotalSleepRecords
ORDER BY
  avgSleepByRecord DESC
```

The results are shown in the following table, where it is notable that very few users have three sleep sessions in a day, while the majority have only one sleep session. Users with three sleep sessions sleep more than 8 hours, those with two records sleep around 8 hours, and those with only one sleep session sleep less than 8 hours.

| TotalSleepRecords | 3 | 2 | 1 |
|---|---|---|---|
| count_SleepRecords | 3 | 43 | 367 |
| minutesAsleepByRecord | 1932,00 | 19485,00 | 151823,00 |
| timeInBedByRecord | 2049,00 | 21504,00 | 165865,00 |
| avgSleepByRecord | 644,00 | 453,14 | 413,69 |
| avgInBedByRecord | 683,00 | 500,09 | 451,95 |

For the hourly data analysis, all tables were merged and analyzed together. The day was divided into four time intervals: morning, afternoon, evening, and night. The average of each metric was then calculated for each time period. The query used is shown below.

```
DECLARE
  MORNING_START,
  MORNING_END,
  AFTERNOON_END,
  EVENING_END INT64;
SET
  MORNING_START = 6;
SET
  MORNING_END = 12;
SET
  AFTERNOON_END = 18;
SET
  EVENING_END = 21;
SELECT
  CASE
    WHEN EXTRACT(HOUR FROM calories.ActivityHour) >
MORNING_START AND
      EXTRACT(HOUR FROM calories.ActivityHour) <=
MORNING_END THEN "Morning"
    WHEN EXTRACT(HOUR FROM calories.ActivityHour) >
MORNING_END AND
      EXTRACT(HOUR FROM calories.ActivityHour) <=
AFTERNOON_END THEN "Afternoon"
    WHEN EXTRACT(HOUR FROM calories.ActivityHour) >
AFTERNOON_END AND
      EXTRACT(HOUR FROM calories.ActivityHour) <=
EVENING_END THEN "Evening"
    ELSE "Night"
  END AS time_of_day,
  AVG(calories.Calories) AS avg_calories,
  AVG(steps.StepTotal) AS avg_steps,
  AVG(intensities.TotalIntensity) AS avg_intensity
FROM
  `bellabeat-
456023.fitbit_fitness_tracker.hourlyCalories` AS
calories
  INNER JOIN
  `bellabeat-
456023.fitbit_fitness_tracker.hourlyIntensities` AS
intensities
  ON calories.Id = intensities.Id AND
calories.ActivityHour = intensities.ActivityHour
  INNER JOIN
  `bellabeat-456023.fitbit_fitness_tracker.hourlySteps`
AS steps
  ON calories.Id = steps.Id AND calories.ActivityHour =
steps.ActivityHour
GROUP BY
  time_of_day
```

The results of the query are shown below. It is evident that the afternoon is the most active time of day, while the night is the least active in terms of steps, calories burned, and intensity.

| time_of_day | avg_calories | avg_steps | avg_intensity |
|---|---|---|---|
| Night | 74,66 | 63,98 | 3,23 |
| Morning | 105,37 | 432,57 | 15,39 |
| Evening | 105,46 | 405,41 | 15,36 |
| Afternoon | 113,26 | 486,83 | 17,91 |

The minute-level sleep table was used to analyze when during the day users are asleep. To do this, the data was grouped by each hour of the day, and using the "value" column, it was determined whether the user was awake, restless, or asleep. Then, the number of minutes spent in each of these states was counted for every hour.

```
SELECT
  EXTRACT(HOUR FROM date) AS hour,
  COUNT(CASE WHEN value = 1 THEN 1 END) AS
minutes_asleep,
  COUNT(CASE WHEN value = 2 THEN 1 END) AS
minutes_restless,
```

```
  COUNT(CASE WHEN value = 3 THEN 1 END) AS
minutes_awake
FROM
  `bellabeat-456023.fitbit_fitness_tracker.minuteSleep`
GROUP BY
  hour
ORDER BY
  hour
```

The results of the query are shown below, where it is observed that users are generally more awake around midday. However, the data seems to contain some inconsistencies: at all hours there are more records of people sleeping than awake. This could be due to an issue in data collection or a malfunction of the Fitbit device. It may also stem from the interpretation of the "value" column, which was obtained from external sources rather than the official dataset documentation.

| hour | minutes_asleep | minutes_restless | minutes_awake |
|---|---|---|---|
| 0 | 31990 | 1830 | 357 |
| 1 | 36017 | 2255 | 242 |
| 2 | 39923 | 2458 | 366 |
| 3 | 43214 | 2382 | 224 |
| 4 | 41694 | 2567 | 256 |
| 5 | 38435 | 2501 | 173 |
| 6 | 28587 | 2703 | 255 |
| 7 | 18542 | 2025 | 212 |
| 8 | 11208 | 1391 | 118 |
| 9 | 7190 | 564 | 68 |
| 10 | 4212 | 354 | 90 |
| 11 | 1906 | 211 | 79 |
| 12 | 938 | 126 | 41 |
| 13 | 1035 | 112 | 33 |
| 14 | 994 | 111 | 9 |
| 15 | 1137 | 145 | 47 |
| 16 | 1067 | 141 | 29 |
| 17 | 728 | 94 | 18 |
| 18 | 508 | 71 | 75 |
| 19 | 677 | 81 | 63 |
| 20 | 1485 | 354 | 166 |
| 21 | 5721 | 842 | 336 |
| 22 | 13998 | 1707 | 439 |
| 23 | 24800 | 2008 | 345 |

Finally, the weight tables were analyzed, using measurements in kilograms, as they are the most common in my region. First, basic statistics related to BMI (Body Mass Index) and users' weight were obtained. The query used is shown below.

```
SELECT
  MAX(WeightKg) AS max_weight,
  MIN(WeightKg) AS min_weight,
  AVG(WeightKg) AS avg_weight,
  MAX(BMI) AS max_BMI,
  MIN(BMI) AS min_BMI,
  AVG(BMI) AS avg_BMI
FROM
  `bellabeat-
456023.fitbit_fitness_tracker.weightLogInfo`
```

The results of the query are shown below, where users appear to belong to different weight categories based on their BMI.

| | |
|---|---|
| max_weight | 133,50 |
| min_weight | 52,60 |
| avg_weight | 72,50 |
| max_BMI | 47,54 |
| min_BMI | 21,45 |
| avg_BMI | 25,37 |

Finally, based on the BMI (Body Mass Index) data, users were grouped into four categories: Underweight (less than 18.5), Normal weight (18.5 – 24.9), Overweight (25.0 – 29.9), and Obese (30.0 or higher). This classification was used to analyze the types of users present in the Fitbit dataset. The query used is shown below.

```
SELECT
  COUNT(CASE WHEN BMI < 18.5 THEN 1 END) AS
Underweight,
  COUNT(CASE WHEN BMI >= 18.5 AND BMI < 25.0 THEN 1
END) AS NormalWeight,
  COUNT(CASE WHEN BMI >= 25.0 AND BMI < 30.0 THEN 1
END) AS Overweight,
  COUNT(CASE WHEN BMI >= 30.0 THEN 1 END) AS Obese
FROM
  `bellabeat-
456023.fitbit_fitness_tracker.weightLogInfo`
```

The results are shown in the following table, with a notable number of users falling into the overweight category.

| | |
|---|---|
| Underweight | 0 |
| NormalWeight | 52 |
| Overweight | 45 |
| Obese | 3 |