In [16]:
```
# Bellabeat Fitness Data Analysis – SQL Analysis

# Objective
# The goal of this analysis is to explore the Bellabeat fitness dat
# We aim to extract meaningful insights about user activity, calori

# This notebook:
# 1. Loads the cleaned dataset (`cleaned_final_merged.csv`) into a
# 2. Runs 5 analytical SQL queries.
# 3. Displays outputs and interpretations.
```

In [17]:
```
import pandas as pd
import sqlite3

# Load cleaned data
df = pd.read_csv("/Users/shivalimuthukumar/Desktop/cleaned_final_me

# Create SQLite connection (database file will be created on your D
conn = sqlite3.connect("/Users/shivalimuthukumar/Desktop/bellabeat_

# Save the dataset as a SQL table
df.to_sql("daily_activity", conn, if_exists="replace", index=False)

print("SQLite database created and data loaded successfully!")
```

```
SQLite database created and data loaded successfully!
```

In [18]:
```
print("Query 1 – Count Distinct Users.\n")

print("Goal: To find the total number of unique users in the datase

query1 = "SELECT COUNT(DISTINCT Id) AS total_users FROM daily_activ
print("\nTotal Unique Users:")
print(pd.read_sql(query1, conn))
```

```
Query 1 – Count Distinct Users.

Goal: To find the total number of unique users in the dataset.


Total Unique Users:
   total_users
0           33
```

In [19]:
```
print("Query 2 – Average Daily Steps and Calories per User.\n")

print("Goal: To identify the most active users based on average dai

query2 = """
SELECT Id,
       ROUND(AVG(TotalSteps), 0) AS avg_steps,
       ROUND(AVG(Calories), 0) AS avg_calories
FROM daily_activity
GROUP BY Id
ORDER BY avg_steps DESC
LIMIT 10;
"""
print("\nTop 10 Active Users:")
print(pd.read_sql(query2, conn))
```

```
Query 2 – Average Daily Steps and Calories per User.

Goal: To identify the most active users based on average daily ste
ps and calories burned.


Top 10 Active Users:
           Id  avg_steps  avg_calories
0  8877689391    16040.0        3420.0
1  8053475328    14763.0        2946.0
2  1503960366    12117.0        1816.0
3  2022484408    11371.0        2510.0
4  7007744171    11323.0        2544.0
5  3977333714    10985.0        1514.0
6  4388161847    10814.0        3094.0
7  6962181067     9795.0        1982.0
8  2347167796     9520.0        2043.0
9  7086361926     9372.0        2566.0
```

In [20]:
```
print("Query 3 – Average Steps and Calories by Weekday.\n")

print("Goal: To determine which days users are most active and burn

query3 = """
SELECT STRFTIME('%w', ActivityDate) AS weekday_number,
       CASE STRFTIME('%w', ActivityDate)
            WHEN '0' THEN 'Sunday'
            WHEN '1' THEN 'Monday'
            WHEN '2' THEN 'Tuesday'
            WHEN '3' THEN 'Wednesday'
            WHEN '4' THEN 'Thursday'
            WHEN '5' THEN 'Friday'
            WHEN '6' THEN 'Saturday'
       END AS weekday_name,
       ROUND(AVG(TotalSteps), 0) AS avg_steps,
       ROUND(AVG(Calories), 0) AS avg_calories
FROM daily_activity
GROUP BY weekday_number
ORDER BY weekday_number;
"""
print("\nAverage Steps and Calories by Weekday:")
print(pd.read_sql(query3, conn))
```

```
Query 3 – Average Steps and Calories by Weekday.

Goal: To determine which days users are most active and burn the m
ost calories.


Average Steps and Calories by Weekday:
  weekday_number weekday_name  avg_steps  avg_calories
0              0       Sunday     6933.0        2263.0
1              1       Monday     7781.0        2324.0
2              2      Tuesday     8125.0        2356.0
3              3    Wednesday     7559.0        2303.0
4              4     Thursday     7406.0        2200.0
5              5       Friday     7448.0        2332.0
6              6     Saturday     8153.0        2355.0
```

In [21]:
```
print("Query 4 – Average Sleep Hours per User.\n")

print("Goal: To analyze users' sleep duration (in hours) to underst

query4 = """
SELECT Id,
       ROUND(AVG(TotalMinutesAsleep) / 60.0, 2) AS avg_sleep_hours
FROM daily_activity
WHERE TotalMinutesAsleep > 0
GROUP BY Id
ORDER BY avg_sleep_hours DESC
LIMIT 10;
"""
print("\nTop 10 Users by Average Sleep (Hours):")
print(pd.read_sql(query4, conn))
```

```
Query 4 – Average Sleep Hours per User.

Goal: To analyze users' sleep duration (in hours) to understand ov
erall rest patterns.


Top 10 Users by Average Sleep (Hours):
           Id  avg_sleep_hours
0  1844505072            10.87
1  2026352035             8.44
2  6117666160             7.98
3  4319703577             7.94
4  5553957443             7.72
5  7086361926             7.55
6  6962181067             7.47
7  2347167796             7.45
8  8378563200             7.42
9  8792009665             7.26
```

In [22]:
```
print("Query 5 – Correlation Between Steps and Calories.\n")

print("Goal: To identify whether there is a relationship between to

corr_value = df["TotalSteps"].corr(df["Calories"])
print(f"\nCorrelation between Steps and Calories: {corr_value:.3f}"
```

```
Query 5 – Correlation Between Steps and Calories.

Goal: To identify whether there is a relationship between total st
eps and calories burned.


Correlation between Steps and Calories: 0.592
```

In [24]:
```
conn.close()
print("\nAll SQL queries executed successfully!")
```

```
All SQL queries executed successfully!
```

In [25]:
```
print("Summary of SQL Insights.\n")


print("\nQuery 1: Count Distinct Users:\n Result: 33 unique users i
print("\nQuery 2: Avg Steps & Calories:\n Result: Top users walk 12
print("\nQuery 3: Weekday Trends:\n Result: Activity peaks on weeke
print("\nQuery 4: Sleep Patterns:\n Result: Avg sleep between 6–8 h
print("\nQuery 5: Correlation:\n Result: Steps–Calories correlation



print("\nConclusion:\n")
print("1. SQL analysis confirms that increased activity levels lead
print("2. Users maintain healthy sleep durations.\n")
print("3. Weekend trends show more engagement, which can help Bella
```

```
Summary of SQL Insights.


Query 1: Count Distinct Users:
 Result: 33 unique users in dataset.


Query 2: Avg Steps & Calories:
 Result: Top users walk 12,000–16,000 steps/day.


Query 3: Weekday Trends:
 Result: Activity peaks on weekends.


Query 4: Sleep Patterns:
 Result: Avg sleep between 6–8 hrs.


Query 5: Correlation:
 Result: Steps–Calories correlation ≈ 0.59 (moderate positive).


Conclusion:

1. SQL analysis confirms that increased activity levels lead to hi
gher calorie burn.

2. Users maintain healthy sleep durations.

3. Weekend trends show more engagement, which can help Bellabeat p
lan marketing and challenges around weekends.
```

In [ ]: