## **Step 1 – Per-User Aggregates**

After cleaning and merging the raw daily activity and sleep logs, the dataset was aggregated at the **user level**. This allows comparisons across participants and highlights differences in overall lifestyle patterns, rather than just focusing on day-to-day fluctuations.

### **Query Executed**

SELECT

Id,

ROUND(AVG(total\_steps),0) AS avg\_steps,

ROUND(AVG(total\_minutes\_asleep)/60,2) AS avg\_sleep\_hours,

ROUND(AVG(calories),0) AS avg\_calories,

SUM(non\_wear\_day) AS non\_wear\_days

FROM `coursera-473006.Bellabeat.leaf\_dataset`

GROUP BY Id

ORDER BY avg\_steps DESC;

### **What the Query Does**

* AVG(total\_steps): Each user’s average daily steps.
* AVG(total\_minutes\_asleep)/60: Converts minutes to hours and computes average daily sleep.
* AVG(calories): Each user’s average daily calories burned.
* SUM(non\_wear\_day): Total number of non-wear days per user.
* Results are ordered by steps from most active to least active.

### **Results (based on exported file)**

* **Users:** 33 unique users.
* **Sleep data coverage:** Only 24 users have recorded sleep data → ~9 users show NULL for avg\_sleep\_hours.
* **Steps:** Wide variation, from fewer than 3,000 steps/day on average to over 16,000.
* **Calories:** Strongly scales with steps, confirming consistency in activity tracking.
* **Non-wear days:** Some users wore the device nearly every day, while others had multiple non-wear days — indicating differences in adherence.

### **Key Observations**

* **High-activity outliers:** A small group of users averaged 15,000–16,000+ steps/day, but many of them did not record sleep data.
* **Balanced patterns:** Users averaging ~9,000 steps/day tended to sleep around 7+ hours per night, suggesting a healthier balance between activity and rest.
* **Sedentary profiles:** Lower-step users often had fewer calories burned and more sedentary minutes, aligning with less active lifestyles.
* **Device engagement:** Non-wear days highlight reliability issues — if the device is not worn consistently, data gaps emerge.

### **Why This Matters for Bellabeat**

* These user-level profiles reveal **segments** of behavior (e.g., highly active but inconsistent sleepers, moderately active with balanced sleep, sedentary non-wearers).
* Such segmentation can guide Bellabeat’s marketing:  
  + Encourage **active women** to use Leaf for better recovery and sleep tracking.
  + Position Leaf as a motivator for **sedentary users** to adopt healthier routines.
  + Highlight the importance of **consistent device use** to get meaningful insights, addressing adherence challenges.

## **📝 Step 2: Organize and Format the Data**

### **1. Review Data Types (Schema Check)**

Query executed:

SELECT

column\_name,

data\_type

FROM `coursera-473006.Bellabeat.INFORMATION\_SCHEMA.COLUMNS`

WHERE table\_name = 'leaf\_dataset';

🔹 **Purpose**: Ensured each variable had the correct type (e.g., activity\_date = DATE, total\_steps = INT64, sleep\_hours = FLOAT).  
 🔹 **Result**: Data types were consistent and ready for analysis. No further type casting was needed.

### **2. Count Rows and Check Missing Values**

Query executed:

SELECT

COUNT(\*) AS total\_rows,

COUNTIF(total\_minutes\_asleep IS NULL) AS missing\_sleep,

COUNTIF(total\_steps IS NULL) AS missing\_steps

FROM `coursera-473006.Bellabeat.leaf\_dataset`;

🔹 **Purpose**: Assessed completeness of the dataset.  
 🔹 **Result**:

* **Total rows**: 940
* **Missing sleep records**: 530
* **Missing steps**: 0

📌 **Interpretation**: The dataset is complete for activity data, but about half of the rows have missing sleep data. This is expected, as not all users track sleep.

### **3. Convert Sleep Minutes to Hours**

Query executed:

SELECT

Id,

activity\_date,

total\_steps,

ROUND(total\_minutes\_asleep/60, 2) AS sleep\_hours,

calories

FROM `coursera-473006.Bellabeat.leaf\_dataset`;

🔹 **Purpose**: Converted total\_minutes\_asleep into hours (sleep\_hours) for easier interpretation and downstream analysis.  
 🔹 **Result**: Created a user-friendly metric that aligns with common sleep recommendations (hours/day).

### **4. Create a Final Ready Dataset**

Query executed:

CREATE OR REPLACE TABLE `coursera-473006.Bellabeat.leaf\_dataset\_ready` AS

SELECT

Id,

activity\_date,

total\_steps,

sedentary\_minutes,

calories,

ROUND(total\_minutes\_asleep/60, 2) AS sleep\_hours,

total\_time\_in\_bed,

sleep\_anomaly,

non\_wear\_day

FROM `coursera-473006.Bellabeat.leaf\_dataset`;

🔹 **Purpose**: Compiled all relevant fields into a single clean dataset, formatted for analysis.  
 🔹 **Result**: leaf\_dataset\_ready created, containing 940 rows with standardized columns (sleep\_hours added, anomalies/flags preserved).

✅ **Summary of Step 2**

* Schema validated: column types correct.
* Data completeness checked: 940 rows, sleep missing in ~56% of cases.
* Sleep converted to hours for interpretability.
* Final cleaned dataset (leaf\_dataset\_ready) prepared for analysis.

# **Step 3 – Perform Calculations**

### **Goal**

Transform daily records into **user-level insights** and uncover **relationships** between activity, calories, and sleep.

### **✅ 1. Daily Averages Across the Whole Dataset**

**Query:**

SELECT

ROUND(AVG(total\_steps),0) AS avg\_steps,

ROUND(AVG(calories),0) AS avg\_calories,

ROUND(AVG(sleep\_hours),2) AS avg\_sleep\_hours,

ROUND(AVG(total\_time\_in\_bed)/60,2) AS avg\_time\_in\_bed

FROM `coursera-473006.Bellabeat.leaf\_dataset\_ready`;

**Results:**

* **7,638 steps/day**
* **2,304 calories/day**
* **6.99 hours of sleep/day**
* **7.64 hours in bed/day**

**Interpretation:** On average, users get close to 7 hours of sleep, but only slightly above 7,600 daily steps — below the often-cited 10,000-step guideline.

### **✅ 2. Per-User Aggregates**

**Query:**

SELECT

Id,

ROUND(AVG(total\_steps),0) AS avg\_steps,

ROUND(AVG(sleep\_hours),2) AS avg\_sleep\_hours,

ROUND(AVG(calories),0) AS avg\_calories,

SUM(non\_wear\_day) AS non\_wear\_days

FROM `coursera-473006.Bellabeat.leaf\_dataset\_ready`

GROUP BY Id

ORDER BY avg\_steps DESC;

**Results Snapshot:**

* Some users average **16,000+ steps/day**, but several have **missing sleep data (NULL)**.
* Moderate users (~9,000 steps/day) report **7+ hours of sleep**, indicating balanced patterns.
* Calories scale with steps (more activity → higher energy expenditure).

**Interpretation:** Users can be segmented into **high-activity/short-sleepers vs. moderate/steady sleepers** — a useful starting point for marketing personas.

### **✅ 3. Correlation Checks**

**Query:**

SELECT

CORR(total\_steps, sleep\_hours) AS corr\_steps\_sleep,

CORR(total\_steps, calories) AS corr\_steps\_calories

FROM `coursera-473006.Bellabeat.leaf\_dataset\_ready`;

**Results:**

* Steps vs. Sleep: **-0.19** → weak negative correlation (more steps = slightly less sleep).
* Steps vs. Calories: **+0.59** → moderate positive correlation (steps strongly drive calories burned).

**Interpretation:** The device accurately captures calorie burn via steps, but higher activity may slightly reduce sleep duration.

### **✅ 4. Segment Users by Sleep Duration**

**Query:**

SELECT

CASE

WHEN ROUND(AVG(sleep\_hours),2) < 6 THEN 'Short Sleepers (<6h)'

WHEN ROUND(AVG(sleep\_hours),2) BETWEEN 6 AND 8 THEN 'Normal Sleepers (6–8h)'

ELSE 'Long Sleepers (>8h)'

END AS sleep\_group,

ROUND(AVG(total\_steps),0) AS avg\_steps,

ROUND(AVG(calories),0) AS avg\_calories,

COUNT(DISTINCT Id) AS users

FROM `coursera-473006.Bellabeat.leaf\_dataset\_ready`

GROUP BY sleep\_group;

**Results:**

* **Normal Sleepers (6–8h):** ~7,693 steps, 2,411 calories, largest group.
* **Short Sleepers (<6h):** ~7,693 steps, 2,261 calories.
* **Long Sleepers (>8h):** ~7,172 steps, 2,134 calories.

**Interpretation:** Normal sleepers are the most balanced group. Long sleepers are less active, while short sleepers burn fewer calories despite similar steps.

### **✅ 5. Anomaly Checks**

**Query:**

SELECT

COUNTIF(sleep\_anomaly = 1) AS extreme\_sleep\_days,

COUNTIF(non\_wear\_day = 1) AS non\_wear\_days

FROM `coursera-473006.Bellabeat.leaf\_dataset\_ready`;

**Results:**

* **Extreme sleep days:** 9
* **Non-wear days:** 79

**Interpretation:** The dataset contains a manageable number of anomalies, but non-wear days could impact consistency — highlighting the need for user engagement strategies.

### **📌 Step 3 – Summary**

* **Average user**: 7,600 steps/day, 7 hours of sleep, 2,300 calories.
* **Correlations**: Steps ↔ Calories (strong), Steps ↔ Sleep (slightly negative).
* **Segments**: Most users are normal sleepers (6–8h), showing balanced patterns.
* **Data quality**: Some anomalies exist, but not enough to undermine analysis.

**🔹 Step 4 – Identify Trends and Relationships**

**Goal:** Understand patterns in the dataset: how activity, calories, sleep, and sedentary minutes are connected. This reveals which behaviors Bellabeat should emphasize in marketing or product design.

### **1. Correlations Between Key Metrics**

**Query Run:**

SELECT

CORR(total\_steps, sleep\_hours) AS corr\_steps\_sleep,

CORR(total\_steps, calories) AS corr\_steps\_calories,

CORR(sedentary\_minutes, sleep\_hours) AS corr\_sedentary\_sleep

FROM `coursera-473006.Bellabeat.leaf\_dataset\_ready`;

**Results:**

* **Steps ↔ Calories:** **+0.59** → strong positive correlation (more steps → more calories burned).
* **Steps ↔ Sleep:** **-0.19** → weak negative correlation (active days don’t mean better sleep; sometimes slightly less).
* **Sedentary ↔ Sleep:** **-0.60** → moderate negative correlation (longer sedentary periods linked with less sleep).

**Interpretation:** Calories are strongly driven by activity, making fitness a clear focus. Sedentary time is an important negative health factor Bellabeat can target (reminders to move, posture nudges, etc.). Sleep has weaker, less reliable ties to steps.

### **2. User Segmentation by Sleep Duration**

**Query Run:**

SELECT

CASE

WHEN ROUND(AVG(sleep\_hours),2) < 6 THEN 'Short Sleepers (<6h)'

WHEN ROUND(AVG(sleep\_hours),2) BETWEEN 6 AND 8 THEN 'Normal Sleepers (6–8h)'

ELSE 'Long Sleepers (>8h)'

END AS sleep\_group,

ROUND(AVG(total\_steps),0) AS avg\_steps,

ROUND(AVG(calories),0) AS avg\_calories,

COUNT(DISTINCT Id) AS users

FROM `coursera-473006.Bellabeat.leaf\_dataset\_ready`

GROUP BY sleep\_group;

**Results:**

* **Normal Sleepers (6–8h):** ~7,693 steps, 2,411 calories (largest group, balanced).
* **Short Sleepers (<6h):** ~7,693 steps, 2,261 calories (similar activity but less recovery).
* **Long Sleepers (>8h):** ~7,172 steps, 2,134 calories (less active overall).

**Interpretation:** Normal sleepers are the most balanced and healthiest. Short sleepers still walk but burn fewer calories, while long sleepers are more sedentary. This segmentation can support personalized marketing:

* Push **recovery and sleep insights** to short sleepers.
* Push **activity goals** to long sleepers.
* Promote **balanced lifestyle tracking** to normal sleepers.

### **3. Sedentary Behavior Averages**

**Query Run:**

SELECT

ROUND(AVG(sedentary\_minutes),0) AS avg\_sedentary\_minutes,

ROUND(AVG(total\_steps),0) AS avg\_steps,

ROUND(AVG(sleep\_hours),2) AS avg\_sleep\_hours

FROM `coursera-473006.Bellabeat.leaf\_dataset\_ready`;

**Results:**

* **Avg sedentary time:** ~991 minutes/day (~16.5 hours).
* **Avg steps:** ~7,638/day.
* **Avg sleep:** ~7 hours.

**Interpretation:** Despite moderate activity, users are sedentary for most of the day — a key opportunity for Bellabeat to promote movement reminders and micro-activity features.

### **4. Extreme Cases & Anomalies**

**Query Run:**

SELECT

COUNTIF(sleep\_anomaly = 1) AS extreme\_sleep\_days,

COUNTIF(non\_wear\_day = 1) AS non\_wear\_days,

MIN(total\_steps) AS min\_steps,

MAX(total\_steps) AS max\_steps,

MIN(calories) AS min\_calories,

MAX(calories) AS max\_calories

FROM `coursera-473006.Bellabeat.leaf\_dataset\_ready`;

**Results:**

* **Extreme sleep days:** 9.
* **Non-wear days:** 79.
* **Steps range:** 0 → 36,019.
* **Calories range:** 0 → 4,900.

**Interpretation:** There are clear **data anomalies and non-wear gaps**, reflecting real-world tracker behavior. For Bellabeat, this reinforces the need to design **engagement features** that keep users consistently wearing the device.

✅ **Step 4 Summary**:

* Activity ↔ Calories is the strongest, most consistent relationship.
* Sedentary behavior is very high and negatively tied to sleep → opportunity for interventions.
* Sleep has weaker patterns and varies widely between users.
* User segmentation (short, normal, long sleepers) reveals lifestyle clusters Bellabeat could target with **different product/marketing strategies**.

# **📊 Step 5 – Investigating Trends**

Goal: Move beyond averages and correlations to uncover **deeper behavioral patterns** among users.

### **🔹 Step 5.1 – Segment Users by Sleep Group**

**Query:**

-- Segment users by sleep group

SELECT

CASE

WHEN avg\_sleep\_hours < 6 THEN 'Short Sleepers (<6h)'

WHEN avg\_sleep\_hours BETWEEN 6 AND 8 THEN 'Normal Sleepers (6–8h)'

ELSE 'Long Sleepers (>8h)'

END AS sleep\_group,

ROUND(AVG(avg\_steps),0) AS avg\_steps,

ROUND(AVG(avg\_calories),0) AS avg\_calories,

COUNT(DISTINCT Id) AS users

FROM `coursera-473006.Bellabeat.user\_summary`

GROUP BY sleep\_group

ORDER BY avg\_steps DESC;

**Results:**

* Normal Sleepers (6–8h): ~7,693 steps, 2,411 calories, **14 users**.
* Short Sleepers (<6h): ~7,693 steps, 2,261 calories, **8 users**.
* Long Sleepers (>8h): ~7,172 steps, 2,134 calories, **11 users**.

**Interpretation:**

* Normal sleepers = most balanced (highest activity + calorie burn).
* Short sleepers = same steps but lower calorie burn (less intensity).
* Long sleepers = less active overall.

### **🔹 Step 5.2 – Active vs. Less Active Users**

**Query:**

-- Compare more active vs less active users

SELECT

CASE

WHEN avg\_steps >= (SELECT AVG(avg\_steps) FROM `coursera-473006.Bellabeat.user\_summary`)

THEN 'More Active Users'

ELSE 'Less Active Users'

END AS activity\_group,

ROUND(AVG(avg\_sleep\_hours),2) AS avg\_sleep,

ROUND(AVG(avg\_calories),0) AS avg\_calories,

COUNT(\*) AS users

FROM `coursera-473006.Bellabeat.user\_summary`

GROUP BY activity\_group;

**Results:**

* More Active Users: ~5.97h sleep, 2,502 calories, **16 users**.
* Less Active Users: ~6.68h sleep, 2,076 calories, **17 users**.

**Interpretation:**

* More active users burn **significantly more calories**, but sleep **less**.
* Less active users sleep longer but burn fewer calories.
* There’s a trade-off: activity intensity ↔ sleep duration.

### **🔹 Step 5.3 – Engagement Levels (Non-Wear Days)**

**Query:**

-- Categorize users by engagement (non-wear days as proxy)

SELECT

CASE

WHEN total\_non\_wear\_days = 0 THEN 'Highly Engaged'

WHEN total\_non\_wear\_days BETWEEN 1 AND 10 THEN 'Moderately Engaged'

ELSE 'Low Engagement'

END AS engagement\_group,

COUNT(\*) AS users,

ROUND(AVG(avg\_steps),0) AS avg\_steps,

ROUND(AVG(avg\_sleep\_hours),2) AS avg\_sleep

FROM `coursera-473006.Bellabeat.user\_summary`

GROUP BY engagement\_group

ORDER BY users DESC;

**Results:**

* Highly Engaged: 16 users, ~8,751 steps, 5.7h sleep.
* Moderately Engaged: 15 users, ~6,996 steps, 6.9h sleep.
* Low Engagement: 2 users, ~1,592 steps, 6.4h sleep.

**Interpretation:**

* The more engaged users (wear device consistently) → **far more active**.
* Highly engaged = higher activity but slightly less sleep.
* Low engagement = very low activity (device not worn often).
* Engagement is a **proxy for success** → those who track, move more.

## **✅ Step 5 Takeaways**

1. **Normal sleepers (6–8h)** show the healthiest balance (best group to target for wellness marketing).
2. **More active users** burn more calories but sacrifice sleep — device can highlight recovery.