# SQL ANALYSIS

### Dailyactivity Merged

Insight: Users with the highest total steps often have the highest calorie burn, indicating strong correlation between activity and energy expenditure.

### Sleepday Merged

Insight: Average sleep duration across users is around 6.8 hours, with some users frequently recording less than 5 hours, indicating poor sleep quality.

### Heartrate Seconds Merged

Insight: Most users have a resting heart rate range of 60-75 bpm. Peak heart rate activity aligns with active minutes.

### Weightloginfo Merged

Insight: There is a moderate positive correlation between weight and BMI. Many users have consistent weight logs indicating regular tracking.

### Dailycalories Merged

Insight: Top users burn between 2,000-3,500 calories daily. Days with higher steps directly show higher calorie values.

### Dailysteps Merged

Insight: Step counts vary widely. A few users regularly hit 10,000+ steps/day, showing strong consistency.

### Dailyintensities Merged

Insight: Users average 20-45 minutes of active time daily. Sedentary minutes dominate time distribution.

### Hourlycalories Merged

Insight: Calorie burn spikes during morning and evening hours, aligning with typical workout periods.

### Hourlysteps Merged

Insight: Steps are clustered between 6-9 AM and 5-8 PM, indicating preferred activity hours.

### Hourlyintensities Merged

Insight: Intensity is highest during post-work hours. Early mornings show medium intensity.

### Minutecaloriesnarrow Merged

Insight: Minute-level calorie burn remains low, but high variance is seen during workout sessions.

### Minutecalorieswide Merged

Insight: Same trends as narrow - highlights granular fluctuations during high movement times.

### Minuteintensitiesnarrow Merged

Insight: Active minutes spike in intervals - often in 30-minute blocks, likely workouts.

### Minuteintensitieswide Merged

Insight: Supports intensity patterns seen in other files - highly granular data.

### Minutestepsnarrow Merged

Insight: Minute-by-minute step analysis reveals short bursts of intense activity.

### Minutestepswide Merged

Insight: Same structure as narrow - shows how user movement peaks in intervals.

### Minutesleep Merged

Insight: Users mostly stay in light sleep. Deep sleep periods are limited, with frequent interruptions.

1 .DATA CLEANING

1. Remove Duplicates:

- Used ROW_NUMBER() window function to retain only unique rows per user/date

- Applied DISTINCT inserts or DELETE commands depending on table structure

2. Fix Data Types:

- Converted string dates using STR_TO_DATE()

- Ensured numeric columns like Steps, Calories, BPM are of correct INT/FLOAT types

3. Handle Null Values:

- Removed rows with critical NULLs (e.g., missing ID or Date)

- Filtered out blank strings using TRIM() and LEN()

4. Remove Outliers:

- Checked for negative or extremely high values in:

- Steps (e.g., > 100,000)

- Calories (e.g., > 20,000)

- Heart Rate (e.g., BPM > 220 or < 30)

5. Standardize Columns:

- Unified column formats for dates

- Renamed columns where necessary using ALTER TABLE ... CHANGE

- Converted inconsistent strings (upper/lowercase, units, etc.)

2.BASIC ANALYSIS (SQL)

1. Aggregated Steps per User:

   SELECT Id, SUM(TotalSteps) FROM dailyActivity_merged GROUP BY Id;

2. Average Calories per Day:

   SELECT Date, ROUND(AVG(Calories), 2) FROM dailyActivity_merged GROUP BY Date;

3. Active Days per User:

   SELECT Id, COUNT(DISTINCT Date) FROM dailyActivity_merged GROUP BY Id;

- I Cleaned table creation scripts for all 18 datasets
- Data load commands using LOAD DATA LOCAL INFILE
- Follow-up cleaning & transformation queries per table