SQL Cleaning & Insights Report

1. Executive Summary

You cleaned datasets related to activity, heart rate, calories, sleep, and minute-level metrics. The cleaning work consistently addressed **duplicates, null values, inconsistent formats, unit conversions, and outliers**. After cleaning, exploratory queries were used to extract insights into **activity patterns, sleep quality, calorie expenditure, and correlations across datasets**.

---

2. Dataset-by-Dataset Review

dailyActivity

Cleaning Actions:

- Removed duplicate records by user-date.
- Standardized activity date formats.
- Converted numeric fields to proper datatypes.
- Handled null values and trimmed text fields.

Insights:

- Average daily steps are below 10,000 for many users.
- Weekday activity is higher than weekend activity.
- Higher VeryActiveMinutes drive higher calorie burn.

---

dailyCalories

Cleaning Actions:

- Ensured one calorie record per user-date.
- Validated calorie fields as numeric, removed erroneous values.
- Aligned dates with other daily datasets.

Insights:

- Calories correlate with active minutes and steps.
- Extreme calorie days align with long exercises.

---

dailyIntensities

Cleaning Actions:

- Consolidated intensity metrics and validated logical totals.

- Removed duplicates.

Insights:

- High sedentary minutes are common.
- Shifts to fairly/very active minutes boost calorie burn.

---

dailySteps

Cleaning Actions:

- Deduplicated user-date records.
- Validated step counts, flagged unrealistic outliers.

Insights:

- Wide variation in step counts across users.
- Weekday-weekend step differences observed.

---

heartrateSeconds

Cleaning Actions:

- Standardized timestamp formats to DATETIME.
- Removed duplicate entries.
- Validated heart rate ranges, flagged outliers.

Insights:

- Identified resting heart rate patterns.
- Heart rate spikes align with high-activity periods.

---

hourlyCalories

Cleaning Actions:

- One hourly entry per user-hour; validated sums with daily totals.
- Handled missing hours.

Insights:

- Calorie burn peaks late afternoon/evening.

---

hourlyIntensities

Cleaning Actions:

- Validated hourly bins of activity intensities.
- Standardized columns for joining.

Insights:

- Activity intensity increases during commute hours and evenings.

---

hourlySteps

Cleaning Actions:

- Reconciled hourly totals with minute-level steps.
- Removed duplicates.

Insights:

- Steps peak morning, lunch, and evening.

---

minuteCaloriesNarrow

Cleaning Actions:

- Deduplicated timestamps, normalized to DATETIME.
- Removed implausible values.

Insights:

- Detected bursts of high-calorie burn matching exercise intervals.

---

minuteIntensitiesNarrow

Cleaning Actions:

- Standardized minute-level intensity values.
- Ensured consistency with daily/hourly datasets.

Insights:

- Activity is usually clustered into 20–60 minute sessions.

---

minuteMETsNarrow

Cleaning Actions:

- Validated METs values, filtered invalid data.
- Aligned units across datasets.

Insights:

- METs accurately reflect exercise intensity across users.

---

minuteSleep

Cleaning Actions:

- Normalized timestamps and removed duplicates.
- Handled missing values and flagged fragmented sleep logs.

Insights:

- Many users experience fragmented sleep or lower sleep efficiency.

---

minuteStepsNarrow

Cleaning Actions:

- Validated non-negative minute steps.
- Aggregated for comparison with daily/hourly totals.

Insights:

- Revealed micro-activity patterns like short walks and bursts.

---

sleepDay

Cleaning Actions:

- Standardized date column and removed duplicates.
- Validated minutes asleep vs. time in bed.

Insights:

- Average sleep < 7 hours.
- Correlation between low sleep and reduced next-day activity.

The SQL-based cleaning and analysis of 14 fitness datasets successfully transformed raw, inconsistent data into reliable and structured tables. By removing duplicates, standardizing formats, and validating values, the datasets were prepared for accurate analysis and visualization. From these cleaned datasets, meaningful insights were drawn regarding user activity levels, calorie expenditure, sleep efficiency, and heart rate patterns.

The findings highlight key behavioral trends such as lower-than-recommended activity and sleep levels, strong correlations between activity intensity and calorie burn, and consistent engagement among users who log multiple metrics. These insights provide a valuable foundation for building interactive dashboards in Power BI or Tableau and can guide strategic decisions for personalized wellness recommendations and user engagement strategies.

Overall, the project demonstrates the importance of thorough data cleaning and structured exploratory analysis in deriving actionable insights that can support both **business growth** and **user well-being**.