

## **Internship Report Summary – Fitness Data Analysis (STRAVA)**

### **SQL- PDF**

#### **Objective:**

To analyze, clean, and visualize fitness data from wearable devices using SQL

#### **SQL Data Cleaning Summary:**

Cleaned 7 datasets using MySQL:

daily\_activity, daily\_calories, daily\_steps

daily\_intensities, hourly\_steps, hourly\_calories, hourly\_intensities

Removed over 10,000+ duplicates using advanced SQL queries (ROW\_NUMBER, GROUP BY).

Ensured data consistency by fixing date formats, null values, and redundant rows.

#### **Data Insights Extracted from SQL**

As an observation, Peak Intensity Hours found at Morning (7–9 AM) and Evening (6–8 PM).

Lowest Intensity found at early morning and late night.

User activity spikes during typical exercise hours—suggests routines are aligned with personal fitness goals or work-life balance.

Top Active Users shows users with most consistent and highest activity

Day of Week patterns

Most Active Days found are Saturday and Sunday , Least Active days are Wednesday and Friday

Insight found as the most Users prefer weekend workouts. This supports aligning fitness challenges or group runs with weekends.

Large number of inactive hours detected across all users

While comparing distance Vs calories, users burning high calories are strongly aligned with high tracker distance and step count.

The tracker data found consistent and valid

#### **Project Title:**

SQL-Based Fitness Tracker Data Cleaning and Analysis for STRAVA Reporting

## **Tools Used:**

MySQL 8.0

Steps, Intensity, Sleep, Calories, etc.

## **Data Cleaning Overview**

I cleaned 7 major datasets using SQL operations like DELETE, GROUP BY, HAVING, and ROW\_NUMBER():

Dataset Key Cleaning Tasks

### **Daily\_Activity**

Removed duplicate dates per user, checked for nulls, verified columns with excessive 0s

### **Daily\_Calories**

Removed entries with 0 calories, fixed date formats

### **Daily\_Intensities**

Cleaned duplicates, checked for nulls

### **Daily\_Steps**

Removed duplicates, verified date structure

### **Hourly\_Calories**

Deleted 24-duplicate rows per user per hour

### **Hourly\_Steps**

Used ROW\_NUMBER() to remove hourly duplicates

### **Hourly\_Intensities**

Cleaned with row\_id, grouped and filtered duplicates, extracted patterns by hour and day.