STRAVA FITNESS DATA ANALYSIS

SQL Internship Project Report

---

1. Project Overview

Project Title: Strava Fitness Data Analysis Using SQL
Project Type: SQL Data Analysis
Contribution: Individual
Database Tool: MySQL Workbench
Dataset: Daily Activity Fitness Data

Project Description

This project focuses on analyzing daily fitness activity data using SQL. The objective is to extract meaningful insights related to physical activity patterns, including step counts and calorie expenditure. The analysis helps in understanding user behavior and highlights the importance of data cleaning for time-based analysis.

---

2. Dataset Description

The dataset contains daily fitness activity records with the following key attributes:

- **Id** – Unique user identifier
- **ActivityDate** – Date of activity (stored as text)
- **TotalSteps** – Total number of steps taken
- **TotalDistance** – Distance covered
- **Calories** – Calories burned
- **LoggedActivitiesDistance** – Distance logged manually

Note: The `ActivityDate` column is stored in text format (`MM/DD/YYYY`), which affects direct date-based analysis.

---

3. Database Setup :

Database Name

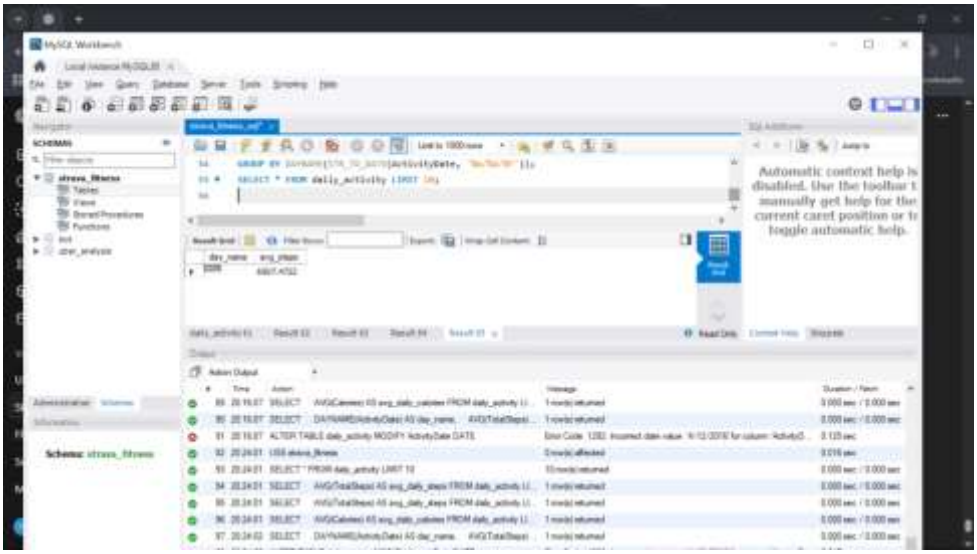
`strava_fitness`

Table Name

`daily_activity`

Sample Data Preview

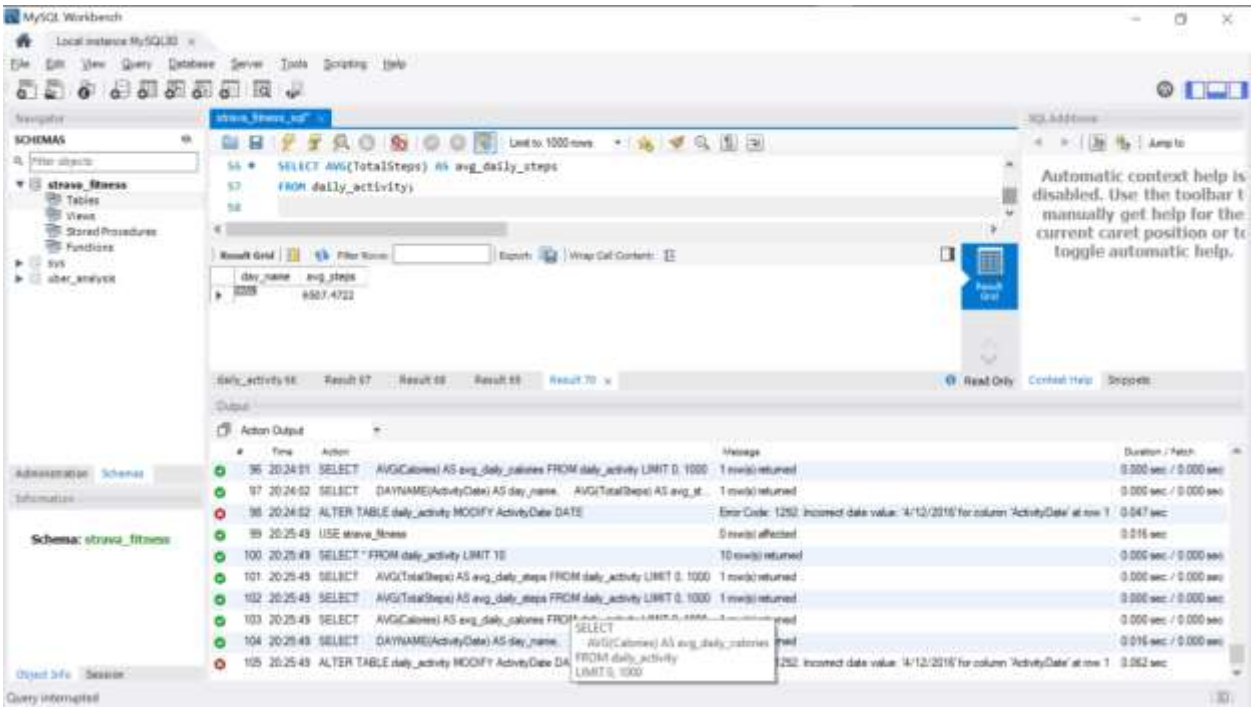
```
SELECT * FROM daily_activity LIMIT 10;
```



4. SQL Analysis & Queries

Query 1: Calculate Average Daily Steps

```
SELECT AVG(TotalSteps) AS avg_daily_steps
FROM daily_activity;
```



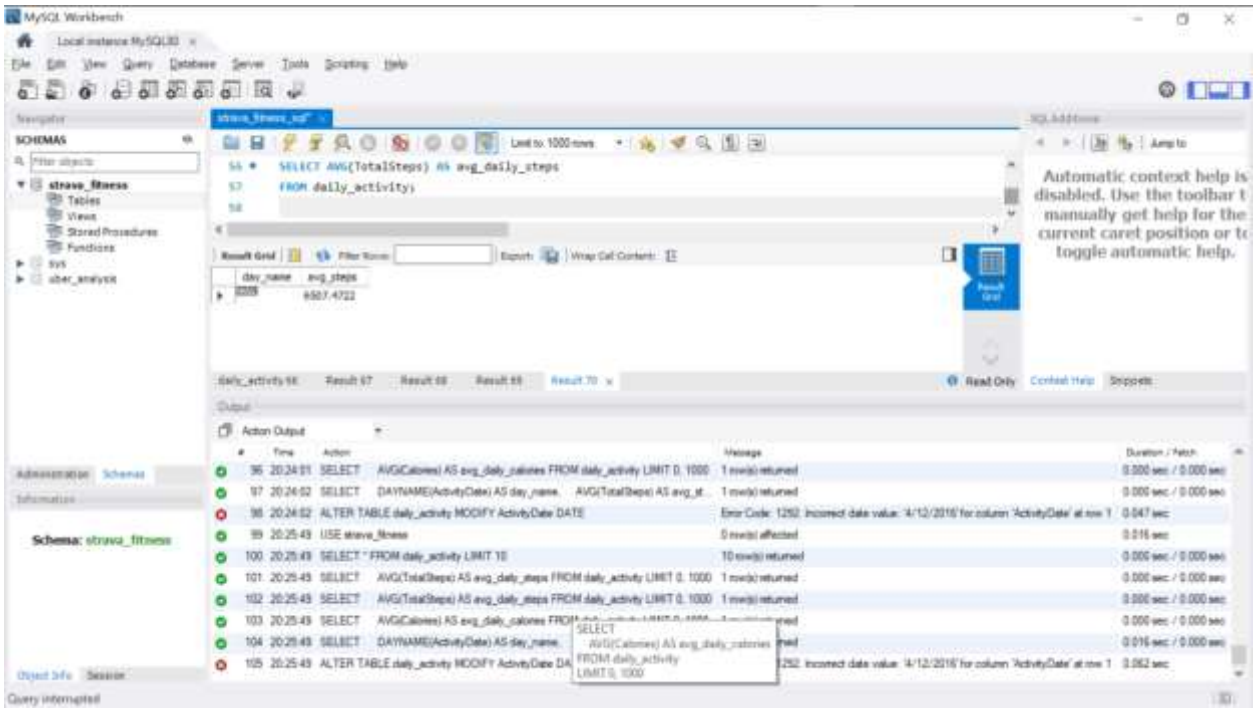
Result:
Average daily steps ≈ **6,507 steps**

Insight:
The average number of steps recorded per user per day is approximately **6,507 steps**, indicating a

moderately active lifestyle. This suggests that users engage in regular movement but may not consistently reach the recommended 10,000 daily steps.

Query 2: Calculate Average Daily Calories Burned

```
SELECT AVG(Calories) AS avg_daily_calories
FROM daily_activity;
```

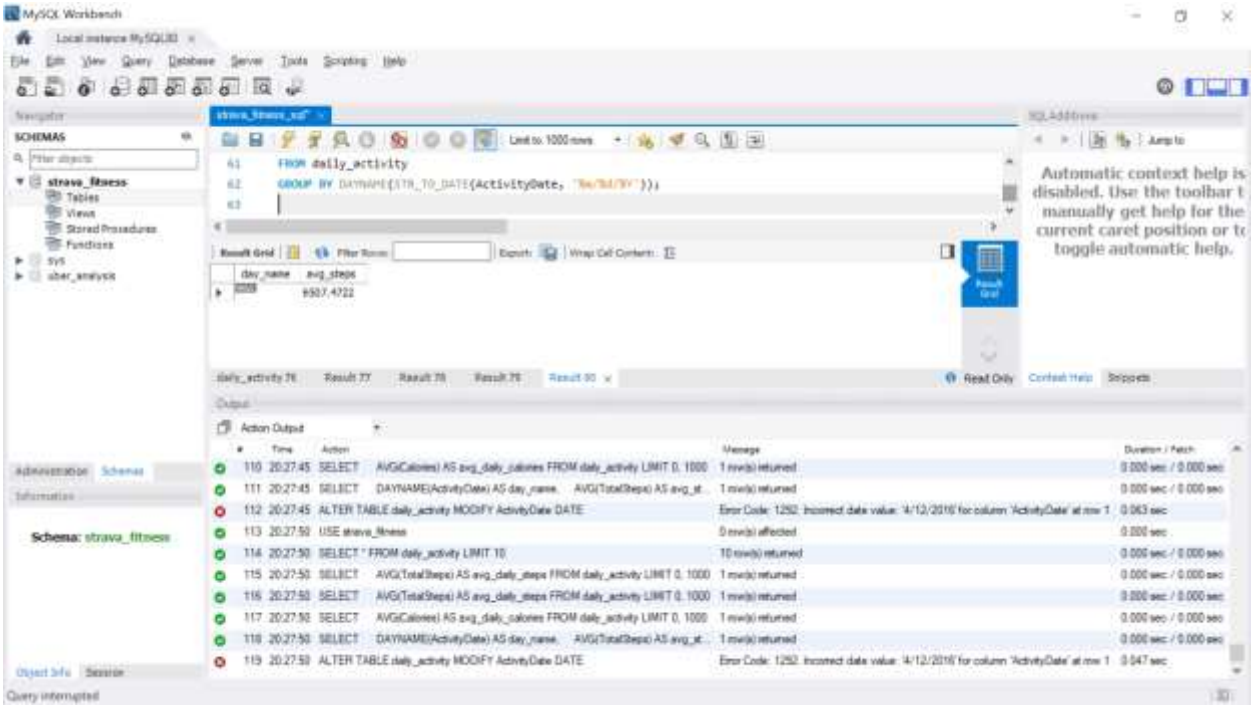


Result:
Average daily calories burned ≈ **1,983 calories**

Insight:
Users burn around **1,983 calories per day** on average. This reflects a realistic level of daily physical activity and energy expenditure typical of non-athletic individuals.

Query 3: Average Steps by Day of the Week

```
SELECT
DAYNAME(STR_TO_DATE(ActivityDate, '%m/%d/%Y')) AS day_name,
AVG(TotalSteps) AS avg_steps
FROM daily_activity
GROUP BY day_name;
```



Insight:
An attempt to analyze average steps by day of the week resulted in **NULL values** due to the `ActivityDate` column being stored as text rather than a proper DATE datatype.

---

5. Data Quality & Challenges

- The `ActivityDate` column is stored as **TEXT**, not as a DATE datatype.
- Direct date functions such as `DAYNAME()` fail without conversion.
- Date conversion using `STR_TO_DATE()` partially resolves the issue but highlights the importance of standardized data formats.

Learning Outcome:
This demonstrates the importance of **data cleaning and preprocessing** before performing advanced analytics.

---

6. Key Insights Summary

- Users take an average of **6,507 steps per day**, indicating moderate activity levels.
- Average daily calorie burn is around **1,983 calories**.
- Date-related analysis was impacted due to inconsistent date formats.
- Proper data type handling is critical for accurate SQL analysis.

---

7. Conclusion

This SQL-based analysis successfully extracted meaningful insights from fitness activity data. The project demonstrates core SQL skills including data querying, aggregation, and handling real-world data issues. Despite minor data quality challenges, valuable insights were obtained, making this project suitable for internship-level evaluation.

---

8. Tools & Technologies Used

- **Database:** MySQL
- **Interface:** MySQL Workbench
- **Language:** SQL

---

9. Appendix

SQL File

- `strava_fitness.sql`

Screenshots Included

- Database & table structure
- Sample data preview
- Average steps query result
- Average calories query result