

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;
```

The screenshot shows the MySQL Workbench interface. In the top-left pane, the schema 'strava_fitness' is selected. The main pane displays a query editor with the following SQL code:

```
14 GROUP BY day_name, activity_date, day_of_week
15 ORDER BY day_name, activity_date, day_of_week
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
```

Below the query editor is a results grid showing the output of the query. The columns are 'day_name', 'activity_date', and 'day_of_week'. The data includes rows for '2020-01-01' through '2020-01-07'. The bottom pane shows the 'Actions' history with 37 entries, mostly showing '1 row(s) returned'.

4. SQL Analysis & Queries

Query 1: Calculate Average Daily Steps

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

The screenshot shows the MySQL Workbench interface. In the top-left pane, the schema 'strava_fitness' is selected. The central pane displays the SQL editor with the following code:

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

The results pane shows a single row of data:

day_name	avg_steps
SUNDAY	6507.4722

The bottom pane shows the 'Actions' tab of the history panel, listing the following queries:

#	Time	Action	Message	Duration / Batch
96	20:24:01	SELECT AVG(Calories) AS avg_daily_calories FROM daily_activity LIMIT 0, 1000	1 rows(s) returned	0.000 sec / 0.000 sec
97	20:24:02	SELECT DAYNAME(ActivityDate) AS day_name, AVG(TotalSteps) AS avg_st	1 rows(s) returned	0.000 sec / 0.000 sec
98	20:24:02	ALTER TABLE daily_activity MODIFY ActivityDate DATE	Error Code: 1292. Incorrect date value: '4/12/2016' for column 'ActivityDate' at row 1 / 0.047 sec	
99	20:25:43	USE strava_fitness	0 rows(s) affected	0.016 sec
100	20:25:49	SELECT * FROM daily_activity LIMIT 10	10 rows(s) returned	0.000 sec / 0.000 sec
101	20:25:49	SELECT AVG(TotalSteps) AS avg_daily_steps FROM daily_activity LIMIT 0, 1000	1 rows(s) returned	0.000 sec / 0.000 sec
102	20:25:49	SELECT AVG(TotalSteps) AS avg_daily_steps FROM daily_activity LIMIT 0, 1000	1 rows(s) returned	0.000 sec / 0.000 sec
103	20:25:49	SELECT AVG(Calories) AS avg_daily_calories FROM daily_activity LIMIT 0, 1000	1 rows(s) returned	0.000 sec / 0.000 sec
104	20:25:49	SELECT AVG(Calories) AS avg_daily_calories FROM daily_activity LIMIT 0, 1000	1 rows(s) returned	0.016 sec / 0.000 sec
105	20:25:49	ALTER TABLE daily_activity MODIFY ActivityDate DA	1252. Incorrect date value: '4/12/2016' for column 'ActivityDate' at row 1 / 0.062 sec	

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;
```

The screenshot shows the MySQL Workbench interface. In the top-left pane, the 'Schemas' tree shows the 'stress_fitness' schema selected. The central pane displays the results of two queries:

```
S1 * SELECT AVG(TotalSteps) AS avg_daily_steps  
      FROM daily_activity;  
S2  
S3
```

The result grid for S1 shows a single row: day_name | avg_steps | 2015-04-12 | 6527.4722. Below the results, the 'Output' tab shows the history of actions taken:

#	Time	Action	Message	Duration / Batch
96	20:24:01	SELECT AVG(Calories) AS avg_daily_calories FROM daily_activity LIMIT 0, 1000	1 rows(s) returned	0.000 sec / 0.000 sec
97	20:24:02	SELECT DAYNAME(ActivityDate) AS day_name, AVG(TotalSteps) AS avg_steps FROM daily_activity	1 rows(s) returned	0.000 sec / 0.000 sec
98	20:24:02	ALTER TABLE daily_activity MODIFY ActivityDate DATE	Error Code: 1292 Incorrect date value: '4/12/2016' for column 'ActivityDate' at row 1 / 0.047 sec	
99	20:25:43	USE stress_fitness	0 rows(s) affected	0.016 sec
100	20:25:43	SELECT * FROM daily_activity LIMIT 10	10 rows(s) returned	0.000 sec / 0.000 sec
101	20:25:43	SELECT AVG(TotalSteps) AS avg_daily_steps FROM daily_activity LIMIT 0, 1000	1 rows(s) returned	0.000 sec / 0.000 sec
102	20:25:43	SELECT AVG(TotalSteps) AS avg_daily_steps FROM daily_activity LIMIT 0, 1000	1 rows(s) returned	0.000 sec / 0.000 sec
103	20:25:43	SELECT AVG(Calories) AS avg_daily_calories FROM daily_activity LIMIT 0, 1000	1 rows(s) returned	0.000 sec / 0.000 sec
104	20:25:43	SELECT DAYNAME(ActivityDate) AS day_name, AVG(Calories) AS avg_daily_calories FROM daily_activity	1 rows(s) returned	0.016 sec / 0.000 sec
105	20:25:43	ALTER TABLE daily_activity MODIFY ActivityDate DATE	292 Incorrect date value: '4/12/2016' for column 'ActivityDate' at row 1 / 0.062 sec	

Result:

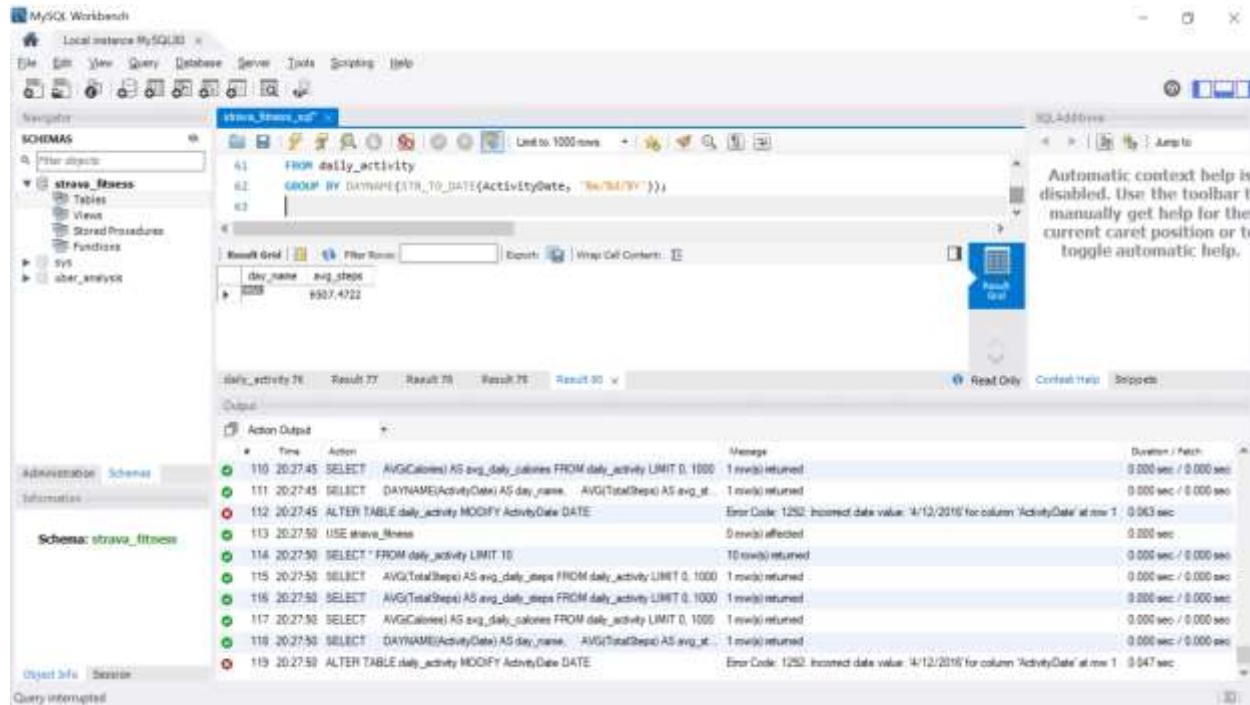
Average daily calories burned ≈ **1,983 calories**

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;
```



The screenshot shows the MySQL Workbench interface with the following details:

- File Bar:** File, Edit, View, Query, Database, Server, Tools, Scripting, Help.
- Navigator:** Schemas (strava_fitness), Tables, Views, Stored Procedures, Functions, sys, aber_analytics.
- Query Editor:** Title: 'new_query_107', Content:

```
61 FROM daily_activity
62 GROUP BY DAYNAME(STR_TO_DATE(ActivityDate, '%m/%d/%Y'))
```
- Results Grid:** Shows the output of the query:

day_name	avg_steps
Sunday	9507.4722
- Output Tab:** Shows the history of actions taken:

Action	Time	Action	Message	Duration / Fetch
110	20:27:45	SELECT AVG(Calories) AS avg_daily_calories FROM daily_activity LIMIT 0, 1000	1 rows(s) returned	0.000 sec / 0.000 sec
111	20:27:45	SELECT DAYNAME(ActivityDate) AS day_name, AVG(TotalSteps) AS avg_st	1 rows(s) returned	0.000 sec / 0.000 sec
112	20:27:45	ALTER TABLE daily_activity MODIFY ActivityDate DATE	Error Code: 1292: Incorrect date value: '4/12/2016' for column 'ActivityDate' at row 1	0.063 sec
113	20:27:50	USE strava_fitness	0 rows(s) affected	0.000 sec
114	20:27:50	SELECT * FROM daily_activity LIMIT 10	10 rows(s) returned	0.000 sec / 0.000 sec
115	20:27:50	SELECT AVG(TotalSteps) AS avg_daily_steps FROM daily_activity LIMIT 0, 1000	1 rows(s) returned	0.000 sec / 0.000 sec
116	20:27:50	SELECT AVG(TotalSteps) AS avg_daily_steps FROM daily_activity LIMIT 0, 1000	1 rows(s) returned	0.000 sec / 0.000 sec
117	20:27:50	SELECT AVG(Calories) AS avg_daily_calories FROM daily_activity LIMIT 0, 1000	1 rows(s) returned	0.000 sec / 0.000 sec
118	20:27:50	SELECT DAYNAME(ActivityDate) AS day_name, AVG(TotalSteps) AS avg_st	1 rows(s) returned	0.000 sec / 0.000 sec
119	20:27:50	ALTER TABLE daily_activity MODIFY ActivityDate DATE	Error Code: 1292: Incorrect date value: '4/12/2016' for column 'ActivityDate' at row 1	0.047 sec

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