Project Proposal: Fitness Data Analysis using SQL

Project Title:
Fitness Data Analysis and Insights

Project Objective:
The objective of this project is to analyze the workout data of users in a fitness tracker dataset. Through SQL queries, the goal is to extract insights regarding various aspects of user fitness and health. This includes understanding which workouts burn the most calories, identifying the fittest age group, evaluating the impact of workouts on mood, and categorizing users by their activity levels, among other insights. The project aims to demonstrate SQL skills by employing a variety of SQL techniques such as aggregation, grouping, correlation analysis, and custom health scoring.

SQL Skills Used:

1. **Basic Aggregation:**
   Using SQL aggregate functions like `AVG()`, `SUM()`, and `COUNT()` to compute averages, totals, and counts across different dimensions (e.g., workout types, users, etc.).

2. **Group By and Ordering:**
   Grouping data by specific fields such as "Workout Type", "User ID", or "Age" and ordering the results to highlight top performers or trends.

3. **Date and Time Functions:**
   Using `strftime()` to extract specific time-related data, such as the month or year, from date fields to analyze patterns over time (e.g., calories burned per month).

4. **Conditional Aggregation:**
   Implementing conditional logic with `CASE WHEN` to segment data (e.g., BMI categories, activity levels) and evaluate results based on specific conditions.

5. **Correlation Analysis:**
   Computing Pearson correlation coefficients using SQL to evaluate relationships between variables like calories burned and heart rate, or steps taken and calories burned. This showcases the ability to calculate and interpret relationships in data.

6. **Custom Scoring System:**
   Creating a custom health score that combines multiple fitness metrics like calories burned, steps taken, water intake, sleep hours, and body fat percentage to evaluate the overall health of users.

7. **Subqueries and Complex Joins:**
   Although not used directly in the provided queries, this would be an area to explore in more advanced analysis, such as joining different datasets or calculating additional metrics based on subqueries.

Queries Used for Analysis:

1. **Which Workout Burns the Most Calories:**
   This query calculates the average calories burned for each workout type and orders the results to find the workout type that burns the most calories on average.

2. **What Age Group Burns the Most Calories by Gender and Workout Type:**
   This query analyzes which age groups, broken down by gender and workout type, burn the most calories, helping to identify trends across different demographic groups.

3. **Fittest Age Group:**
   This query calculates a fitness score for each age group by combining multiple factors like calories burned, steps taken, water intake, sleep, VO2 max, and body fat percentage. The query ranks the age groups based on their fitness scores to identify which age group is the fittest.

4. **Calories Burned Over Time:**
   This query calculates the average calories burned per month to observe how calories burned trend over time. It can help identify peak workout months.

5. **Workout Frequency by Month:**
   This query calculates how many workouts each user did in each month, helping to understand user consistency and workout patterns.

6. **Seasonality of Certain Workouts:**
   This query tracks which workout types are most common in each month, giving insight into seasonal workout preferences or trends.

7. **Most Active Users:**
   This query identifies the top users who have taken the most steps in total, providing a list of the most physically active users in the dataset.

8. **High Burners:**
   This query identifies the top users who have burned the most calories in total, showing which users are the most dedicated to burning calories.

9. **Most Common Workout by User:**
   This query identifies the most common workout type for each user, showing what workout they do the most often.

10. **Correlation Between Calories Burned and Heart Rate:**
    This query calculates the Pearson correlation between calories burned and heart rate, helping to understand how heart rate affects calories burned during workouts.

11. **Correlation Between Steps Taken and Calories Burned:**
    This query calculates the Pearson correlation between the number of steps taken and calories burned, analyzing how step count influences caloric expenditure.

12. **Segment Users by BMI:**
    This query calculates BMI (Body Mass Index) for each user and categorizes them into weight categories such as underweight, normal, overweight, and obese.

13. **Segment Users by Activity Level:**
    This query classifies users into activity levels (highly active, moderately active, inactive) based on their total calories burned and steps taken.

14. **Mood Change by Workout Type:**
    This query tracks how different workout types affect users' mood before and after workouts, providing insight into the emotional impact of different types of exercise.

15. **Health Score Based on Multiple Metrics:**
    This query calculates a custom health score for each user based on various health metrics like calories burned, steps taken, sleep hours, VO2 max, and body fat percentage. The score combines these factors to rank users by their overall health.

Conclusion:

This project showcases the ability to perform a comprehensive analysis of fitness and health data using SQL. The analysis provides valuable insights into workout habits, user activity levels, correlations between different fitness metrics, and how age and gender influence workout performance. The use of advanced SQL functions such as aggregation, grouping, date functions, correlation analysis, and custom scoring provides a detailed understanding of the dataset and demonstrates proficiency in SQL query writing.

This project can be extended further by integrating additional datasets, performing more advanced statistical analysis, or even building a dashboard for real-time visualization of these insights.