

Capstone Project Proposal

Project Title: *Fitbit Tracker Data Analysis*

Proposed By: *Miguel Paredes*

GitHub Repository: https://github.com/M0pare02/Fitbit_Data_Analysis

Project Overview

Using data from multiple users, this study aims to show how factors such as sleep duration, physical activity, and workout intensity affect overall health and wellness. The expected outcome is to identify optimal habits that improve fitness and develop strategies for efficient training.

Data Sources

Dataset name: Fitbase Data

Source: Kaggle.com

Relevant fields: heart rate, daily intensity, calories, steps, sleep, weight

Link

<https://www.kaggle.com/datasets/sawandikirby/bellabeat-case-study-fitbit-data/data>

Research Objectives

- **Primary Question(s):** Understanding the relationship between user habits and fitness.
- **Secondary/Exploratory Questions:** How similar habits affect different individuals.
 - Findings may provide information for personalized fitness coaching and health habits recommendations.

Data Preparation Approach

Every table has its own independent User ID , together with date, we can obtain a natural key

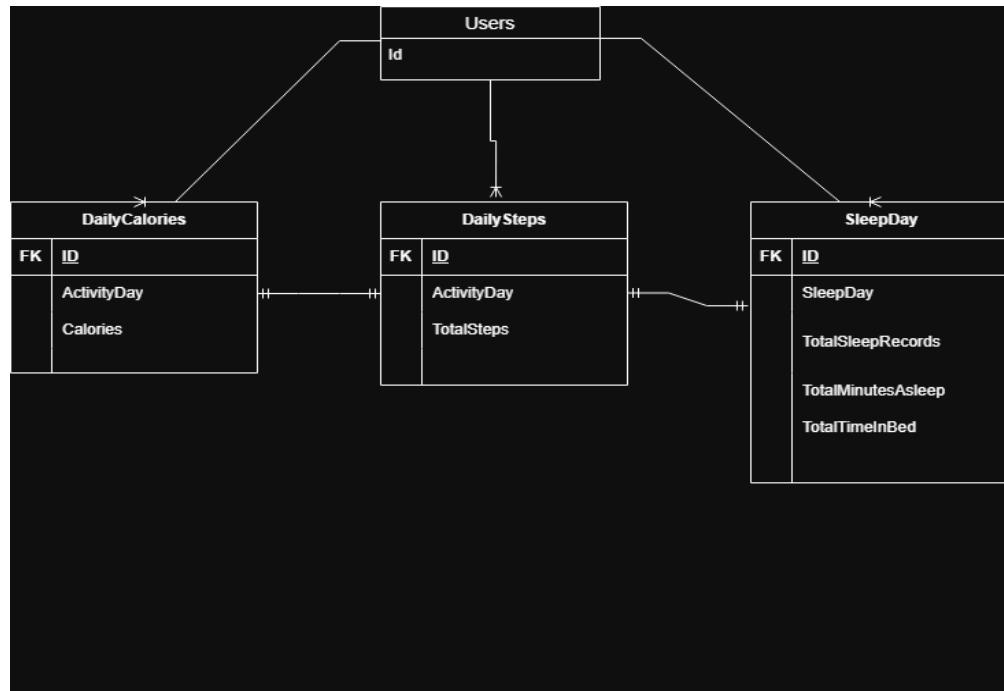
Rows with rare missing values will be dropped or filled in depending on context.

May create new variables such as sleep efficiency (minutes asleep/time in bed), average resting heart rate (average of heart rate logs while resting), weighted time in heart rate zones (zone1, zone2, zone3, zone4).

Each ID is a user with one-to-many relationships to all the other tables.

Each table uses ID and Date to create a natural key that connects them for each user's daily logs.

I don't know which tables I will end up using. This is an example of what the ERD will probably look like.



Current Status

- Collected fitbit datasets and inspected the values. Did initial exploratory checks, row counts, missing values checks, and duplicate records.

Deliverables

- Finish cleaning datasets, standardize date/time, handle missing values, remove duplicates, and outliers.
- Create new aggregation and derived indices, and merge datasets across multiple tables.
- Build SQL tables, produce visualizations for the different correlations.
- Provide proper documentation in the README.

Stretch Goals

- Integrate another dataset with relevant information.
- Create interactive visualizations.

Project Timeline

Phase 1:

- Acquire Datasets
- Clean the data
- Set up ERD

Phase 2:

- Implement python functions for aggregation and merging
- Generate correlation visualizations to see how data relate to each other
- Identify outliers

Phase 3:

- Create derived indices(sleep efficiency, active to rest ratio, etc)
- Run correlation analyses to test hypotheses
- Document

Phase 4:

- Polish everything
- Write README
- Prepare for presentation
- Attempt to do stretch goals

Additional Considerations

Data is only for female users, must keep that in mind, haven't been able to find good mixed/male data that are similar.