1.2

Quality

Reliability:
The reliability of the dataset can be partially assessed based on internal consistency and plausibility. The data appears plausible (e.g., weight, height, heart rate values) and aligns with expected ranges for gym-goers. However, no source documentation or data collection methods are provided, so we cannot verify the dataset's origin, how data was gathered, or the accuracy. Potential issues include uniformity of measurements. For example, are all metrics in consistent units?

Judgment: Medium reliability. Verification would require more information about data sources and methods.

Detail

Level of Detail:

The dataset captures a wide range of variables, from demographic data (age, gender) to fitness specifics (heart rate, workout type, session duration). It includes potentially valuable metrics like calories burned, water intake, and BMI.

Helpfulness:

The level of granularity is suitable for tracking fitness progress or studying correlations between lifestyle choices and health. However, some details could enhance the dataset, such as workout intensity ratings, dietary habits, or long-term tracking for trend analysis.

Documentation

Clarity:

The column names are clear, and units are specified (e.g., kg, liters). However, detailed metadata explaining calculation methods (e.g., how was BMI or calorie burn calculated?) is absent. There is no source documentation provided within the dataset file or as an accompanying document.

Ease of Finding:

Without documentation, assumptions must be made about methodology, which limits usability.

Interrelation

Connections to Other Data:

It could benefit from linking with datasets like:

- Dietary Intake: To examine correlations between nutrition and performance.
- Medical Records: For analyzing the impact of workouts on chronic conditions.
- Geographic Data: To study exercise trends in different regions.

Ease of Connection:

Linking would require compatible identifiers (e.g., timestamps, user IDs). Since this dataset lacks such identifiers, integrating it with other data might involve significant preprocessing.

Use

Potential Uses:

- Fitness Trends Analysis: Identifying workout patterns and their effects on health metrics.
- Personalized Fitness Recommendations: Suggesting optimal routines based on demographic and physiological data.
- Predictive Analytics: Predicting health outcomes based on workout behaviors.

Missing Information:

- Longitudinal data for trend analysis over time.
- Data about injuries, motivation levels, or goals.
- External factors influencing workouts (e.g., weather, schedules).
- Anonymized member IDs for tracking individual progress.

Discoverability

Ease of Finding:

- If this dataset came from an open data repository, its ease of discoverability depends on the platform's search functionality and domain coverage.
- Alternatives may exist in public health datasets, gym software APIs, or fitness device data, though accessibility might vary.

Experience:

- The dataset seems suitable for analyzing exercise patterns, but the lack of comprehensive metadata makes validation challenging.

1.3

Interest in the Dataset

The first thing that made me interested in this dataset in the first place has to do with my interest in fitness and health.

This dataset is intriguing because it provides a comprehensive snapshot of gym members' physical activity and health metrics. It bridges personal fitness tracking with potential insights into broader health trends. Understanding how different variables—such as age, gender, and workout types—affect outcomes like calorie burn, BMI, and resting heart rate offers valuable information for personal trainers, gym managers, and fitness enthusiasts.

Additionally, the dataset's diversity allows the exploration of correlations and predictive analytics, making it versatile for fitness-related applications or research.

Why It's Interesting

1. Practical Applications:
    - This data mirrors real-world gym settings and can inform decisions about workout plans, fitness app features, and gym management strategies.
2. Health and Wellness Trends:
    - Exploring how exercise impacts physical metrics like BMI or resting heart rate aligns with growing interest in preventive healthcare and fitness.
3. Custom Recommendations:
    - The dataset supports building models to personalize fitness regimens, contributing to the trend of individualized health and wellness.

Questions a Database Application Could Help Answer

1. Personalized Fitness Insights:
    - What types of workouts lead to the highest calorie burn for individuals of different genders and age groups?
    - How does BMI change with workout frequency or type over time?
2. Health Monitoring:
    - Are certain workout types correlated with lower resting BPM or higher fat loss?
    - What is the average resting BPM for advanced-level members compared to beginners?
3. Gym Management Decisions:
    - Which workouts are most popular based on frequency and duration?
    - How does water intake vary across different workout types or session durations?
4. Trend Analysis:

- Are younger members more likely to engage in HIIT, while older members prefer yoga or strength training?
- How does workout frequency change with age or BMI?

5. Predictive Modeling:
    - Can we predict calories burned based on demographic and workout data?
    - What factors most strongly influence fat percentage reduction over time?

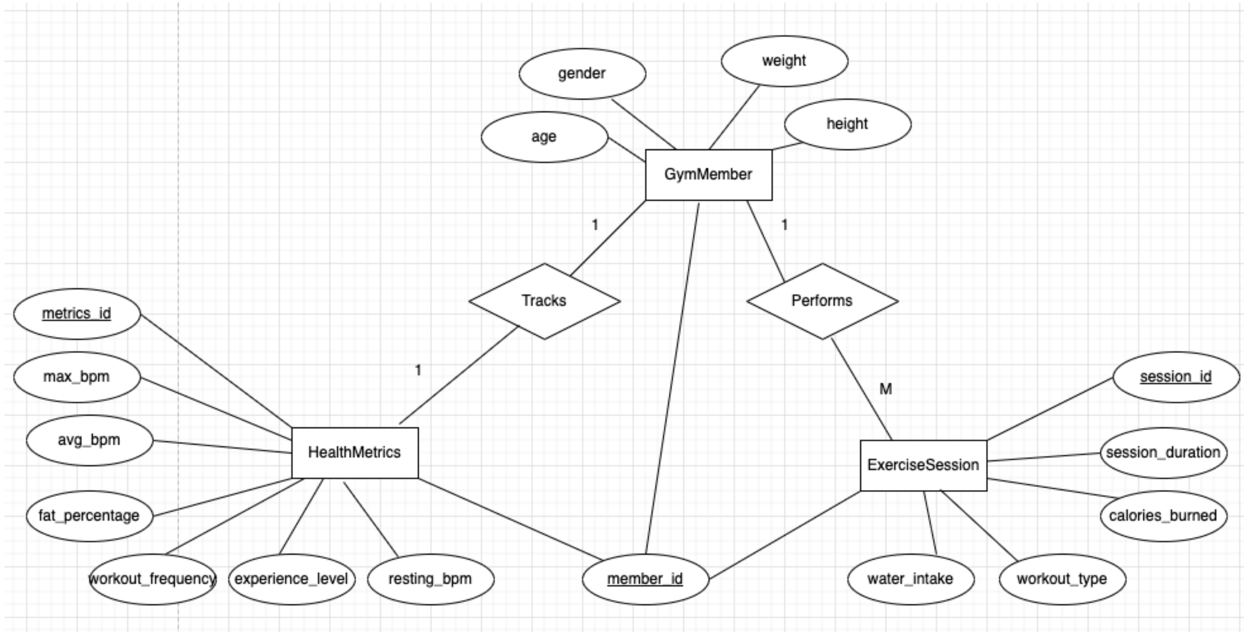
Missing Questions the Dataset Could Answer with Enhancements

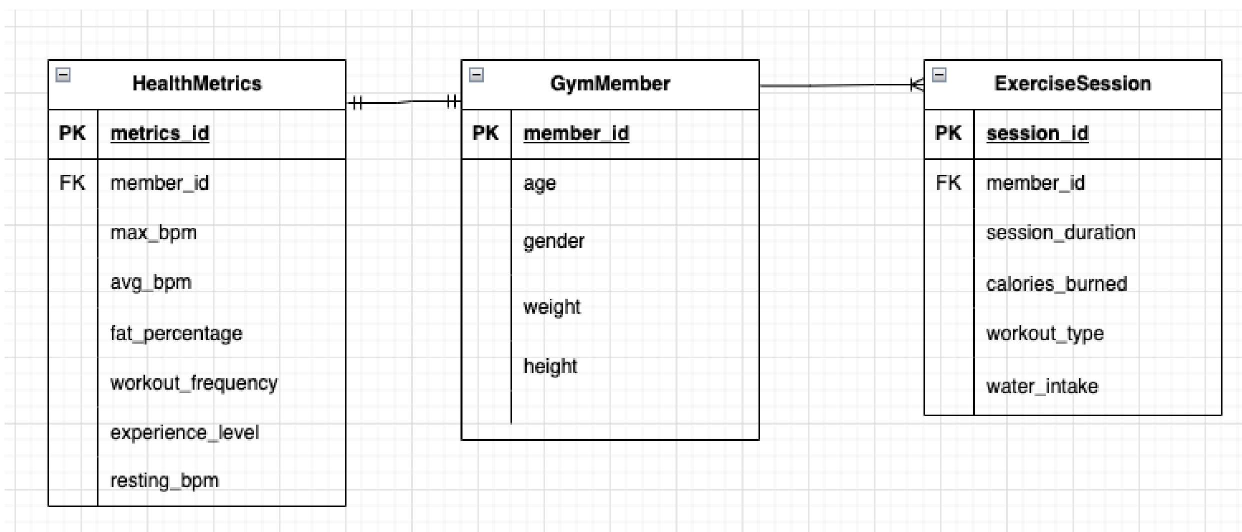
1. Behavioral Insights:
    - What motivates members to increase their workout frequency?
    - Are members more likely to exercise consistently if their workout types vary?
2. Longitudinal Trends:
    - How do metrics like BMI, resting BPM, and calories burned evolve over months or years?
3. External Influences:
    - How do seasonal or weather changes impact workout frequency?
    - Does gym attendance increase during specific times of the day or week?
4. Outcome-Based Questions:
    - What combination of exercise and water intake leads to the highest fat percentage reduction?
    - Do members with higher experience levels have better resting BPM or BMI outcomes?

Value of Database Integration

A robust database application could enable efficient querying, trend visualization, and predictive modeling. For example, integrating this dataset with other health or fitness data (like dietary habits or wearable device stats) could unlock more profound insights, making it a valuable tool for health and fitness innovation.

2.1, 2.2





2.3

Database Tables and Fields

1. GymMember
   - Fields:
     - member_id (Primary Key)
     - age

        - gender
        - weight
        - height
        - bmi
2. ExerciseSession
    - Fields:
        - session_id (Primary Key)
        - member_id (Foreign Key)
        - session_duration
        - calories_burned
        - workout_type
        - water_intake
3. HealthMetrics
    - Fields:
        - metrics_id (Primary Key)
        - member_id (Foreign Key)
        - max_bpm
        - avg_bpm
        - resting_bpm
        - fat_percentage
        - workout_frequency
        - experience_level

Normalization Analysis

1NF (First Normal Form)

Requirements:

- Each column contains atomic values (no multi-valued attributes).
- Each record is uniquely identifiable.

Evaluation:

- Each table satisfies 1NF:
    - All columns contain atomic values.
    - Primary keys (member_id, session_id, metrics_id) uniquely identify records.

2NF (Second Normal Form)

Requirements:

- Satisfies 1NF.

- No partial dependency (i.e., non-primary key attributes depend on the entire primary key).

Evaluation:

- Each table satisfies 2NF:
    - In ExerciseSession and HealthMetrics, all non-primary key attributes depend on the entire primary key.
    - GymMember does not have a composite key, so no partial dependency exists.

3NF (Third Normal Form)

Requirements:

- Satisfies 2NF.
- No transitive dependency (i.e., non-primary key attributes must not depend on other non-primary key attributes).

Evaluation:

- GymMember: bmi might depend on weight and height, suggesting a transitive dependency. To fix this, I remove bmi and calculate it dynamically when needed.
- ExerciseSession: No transitive dependencies.
- HealthMetrics: No transitive dependencies.

Adjustment:

- Remove bmi from GymMember.

Revised Database Tables

1. GymMember
    - Fields:
        - member_id (Primary Key)
        - age
        - gender
        - weight
        - height
2. ExerciseSession
    - Fields:
        - session_id (Primary Key)
        - member_id (Foreign Key)
        - session_duration
        - calories_burned
        - workout_type
        - water_intake

3. HealthMetrics
    - Fields:
        - metrics_id (Primary Key)
        - member_id (Foreign Key)
        - max_bpm
        - avg_bpm
        - resting_bpm
        - fat_percentage
        - workout_frequency
        - experience_level

BCNF (Boyce-Codd Normal Form)

Requirements:

- Every determinant must be a candidate key.

Evaluation:

- All tables are in BCNF since:
    - Primary keys determine all other attributes.
    - No non-trivial functional dependencies violate BCNF.

4NF (Fourth Normal Form)

Requirements:

- Satisfies BCNF.
- No multi-valued dependencies.

Evaluation:

- The current structure avoids multi-valued dependencies since there are no repeated groups or sets of related data.

Justification for Stopping at 3NF/BCNF

1. No Multi-Valued Dependencies: The data structure is simple and doesn't require further normalization to 4NF.
2. Practical Trade-Offs: Over-normalizing (e.g., splitting workout types into a separate table) would complicate queries without significant benefit for this dataset.
3. Data Retrieval Efficiency: Denormalization (if necessary) can optimize performance for common queries.

3.1

```sql
-- Step 1: Create and configure the database
CREATE DATABASE IF NOT EXISTS gym_management;
```

```sql
-- Step 3: Create a new user with appropriate privileges
CREATE USER IF NOT EXISTS 'gym_user'@'127.0.0.1' IDENTIFIED BY 'gymowner123';
GRANT ALL PRIVILEGES ON gym_management.* TO 'gym_user'@'127.0.0.1';
FLUSH PRIVILEGES;
```

```sql
-- Step 2: Create the raw_data table for loading CSV data
CREATE TABLE IF NOT EXISTS raw_data (
    Age INT,
    Gender VARCHAR(10),
    `Weight (kg)` FLOAT,
    `Height (m)` FLOAT,
    Max_BPM INT,
    Avg_BPM FLOAT,
    Resting_BPM INT,
    `Session_Duration (hours)` FLOAT,
    Calories_Burned FLOAT,
    Workout_Type VARCHAR(20),
    Fat_Percentage FLOAT,
    `Water_Intake (liters)` FLOAT,
    `Workout_Frequency (days/week)` INT,
    Experience_Level VARCHAR(20),
    BMI FLOAT
);
```

```sql
-- Step 2: Create the GymMember table
CREATE TABLE IF NOT EXISTS GymMember (
    member_id INT AUTO_INCREMENT PRIMARY KEY,
    age INT NOT NULL,
    gender ENUM('Male', 'Female') NOT NULL,
    weight FLOAT NOT NULL,
    height FLOAT NOT NULL,
    bmi FLOAT NOT NULL
);

-- Step 3: Create the ExerciseSession table
CREATE TABLE IF NOT EXISTS ExerciseSession (
    session_id INT AUTO_INCREMENT PRIMARY KEY,
    member_id INT NOT NULL,
    session_duration FLOAT NOT NULL,
    calories_burned FLOAT NOT NULL,
    workout_type ENUM('Yoga', 'HIIT', 'Cardio', 'Strength') NOT NULL,
    water_intake FLOAT NOT NULL,
    FOREIGN KEY (member_id) REFERENCES GymMember(member_id) ON DELETE CASCADE
);

-- Step 4: Create the HealthMetrics table
CREATE TABLE IF NOT EXISTS HealthMetrics (
    metrics_id INT AUTO_INCREMENT PRIMARY KEY,
    member_id INT NOT NULL,
    max_bpm INT NOT NULL,
    avg_bpm FLOAT NOT NULL,
    resting_bpm INT NOT NULL,
    fat_percentage FLOAT NOT NULL,
    workout_frequency INT NOT NULL,
    experience_level ENUM('Beginner', 'Intermediate', 'Advanced') NOT NULL,
    FOREIGN KEY (member_id) REFERENCES GymMember(member_id) ON DELETE CASCADE
);
```

3.2

How the Data Was Added

1. Step 1: Sample Data Preparation
    - Data was taken from a subset of the provided dataset.
    - Each record was manually formatted into SQL INSERT statements.
2. Step 2: Execution

    - The INSERT INTO commands were executed in the MySQL environment sequentially, starting with GymMember, followed by ExerciseSession and HealthMetrics.
3. Step 3: Validation
    - After inserting data into each table, a SELECT * query was run to confirm the data was correctly inserted.

3.3

What Works Well

1. Normalization and Structure:
    - The database is well-normalized to at least 3NF, ensuring minimal data redundancy and clear relationships between entities (GymMember, ExerciseSession, and HealthMetrics).
    - Attributes are grouped logically, with demographic information in GymMember, session-specific details in ExerciseSession, and health-specific metrics in HealthMetrics.
2. Data Integrity:
    - Using FOREIGN KEY constraints ensures that relationships between gym members and their associated sessions or health metrics are maintained. For example, deleting a GymMember automatically cascades to related entries in other tables.
    - Primary keys (member_id, session_id, and metrics_id) uniquely identify records, avoiding duplication.
3. Flexibility for Analysis:
    - The structure allows for querying specific insights (e.g., workout trends, calorie burn comparisons, or health improvements) without over-complicating relationships.

What Could Be Improved

1. Derived Fields (e.g., BMI):
    - Including BMI as an attribute in the original dataset introduces a redundancy since it can be derived from Weight and Height. While it's been removed from the database, users might expect to store this value for quick access.
    - Improvement: Dynamically calculate BMI in queries rather than storing it.
2. Lack of Member IDs in Dataset:
    - The dataset lacks unique member identifiers (member_id), requiring sequential numbering to assign member_id values. This might cause problems if future datasets include overlapping members or inconsistent ordering.

    - Improvement: Ensure all future datasets include a unique identifier for each member to simplify merging and updates.
3. Workout_Type Limitation:
    - The current ENUM for Workout_Type limits the type of workouts to predefined values (e.g., Yoga, HIIT). Adding new workout types would require schema modifications.
    - Improvement: Replace ENUM with a separate WorkoutType table to allow dynamic updates and flexibility.

Final Reflection

The database effectively reflects the dataset's structure, relationships, and attributes, supporting robust querying and analysis. However, addressing derived attributes, unique identifiers, and flexible workout types would enhance scalability and ease of use.

3.4

What types of workouts lead to the highest calorie burn for individuals of different genders and age groups?

SELECT gm.gender, gm.age, es.workout_type, AVG(es.calories_burned) AS avg_calories_burned FROM GymMember gm JOIN ExerciseSession es ON gm.member_id = es.member_id GROUP BY gm.gender, gm.age, es.workout_type ORDER BY avg_calories_burned DESC;

Are certain workout types correlated with lower resting BPM or higher fat loss?

SELECT es.workout_type, AVG(hm.resting_bpm) AS avg_resting_bpm, AVG(hm.fat_percentage) AS avg_fat_percentage FROM ExerciseSession es JOIN HealthMetrics hm ON es.member_id = hm.member_id GROUP BY es.workout_type ORDER BY avg_fat_percentage ASC, avg_resting_bpm ASC;

Which workouts are most popular based on frequency and duration?

```
SELECT
   es.workout_type,
   COUNT(es.session_id) AS session_count,
   AVG(es.session_duration) AS avg_duration
FROM
   ExerciseSession es
GROUP BY
   es.workout_type
ORDER BY
```

```
    session_count DESC, avg_duration DESC;
```

Are younger members more likely to engage in high-intensity workouts like HIIT?

```
SELECT CASE WHEN gm.age < 30 THEN 'Young' WHEN gm.age BETWEEN 30 AND 50
THEN 'Middle-aged' ELSE 'Older' END AS age_group, es.workout_type, COUNT(es.session_id)
AS session_count FROM GymMember gm JOIN ExerciseSession es ON gm.member_id =
es.member_id WHERE es.workout_type = 'HIIT' GROUP BY age_group, es.workout_type
ORDER BY session_count DESC;
```

Can we predict calories burned based on demographic and workout data?

```
SELECT
    gm.gender,
    gm.age,
    es.workout_type,
    es.session_duration,
    es.calories_burned
FROM
    GymMember gm
JOIN
    ExerciseSession es ON gm.member_id = es.member_id;
```

Q1: What motivates members to increase their workout frequency?

- The dataset lacks motivational or behavioral data, such as survey responses or feedback.

Q2: How do seasonal or weather changes impact workout frequency?

- The dataset does not include timestamps or external weather data.

Q3: What combination of exercise and water intake leads to the highest fat percentage reduction?

- The dataset does not provide temporal trends to track fat percentage changes over time.