**Thaís Amador**
Business Intelligence & Data Analytics

# GYM MEMBER ANALYTICS:

## Fitness Data Analysis & Personalized Training Insights

+52 963 114 5925
thaisamador1@gmail.com

# Introduction

## Context

The fitness industry has become increasingly data-driven, with gyms leveraging analytics to understand member behavior, improve engagement, and design personalized training programs.

This project focuses on analyzing historical gym member data to uncover behavioral patterns, segment members based on fitness and engagement metrics, and create a decision tree model to automatically classify new members.

## Dataset Description

The dataset used comes from Kaggle – Gym Members Exercise Dataset, containing individual level information on gym members, including:

- Demographics: age, gender, experience level.
- Biometrics: weight, height, BMI, fat percentage.
- Workout details: type of workout, workout frequency, session duration, average and maximum heart rate, calories burned.
- Hydration habits: daily water intake.

## Project Objective

The main goals of this analysis are to:

- Clean and preprocess biometric and workout tracking data.
- Perform an exploratory data analysis (EDA) to identify key health and activity trends.
- Apply K-Means clustering to segment members based on performance and engagement.
- Build a decision tree model to classify new members into the identified clusters.
- Develop an interactive Power BI dashboard for visualization and decision-making.

## Tools and Technical Approach

- The Gym Members Exercise dataset was processed and analyzed in R using a combination of data manipulation, statistical, and visualization packages. Core tools included dplyr and tidyr for data cleaning and transformation, ggplot2 for custom visualizations, factoextra and cluster for clustering analysis, rpart for decision tree modeling, and corrplot for correlation analysis.
- The final visual reporting and interactive KPIs were developed in Power BI.

## Methodology

- Data Cleaning & Transformation: Converted categorical fields (Gender, Workout_Type) to factors, validated biometric and workout variable ranges.
- Exploratory Data Analysis: Generated histograms, boxplots, and correlation matrices to identify trends, detect outliers, and analyze relationships among key metrics.
- Clustering: Applied K-Means segmentation using demographic, biometric, and workout features.
- Predictive Modeling: Built a decision tree model to classify new members into clusters based on Age, BMI, Calories_Burned, and Workout_Frequency.
- Visualization: Designed an interactive Power BI dashboard displaying KPIs, cluster distributions, and behavioral insights filtered by demographics and workout type.

# R Analysis and Results

## Data Exploration

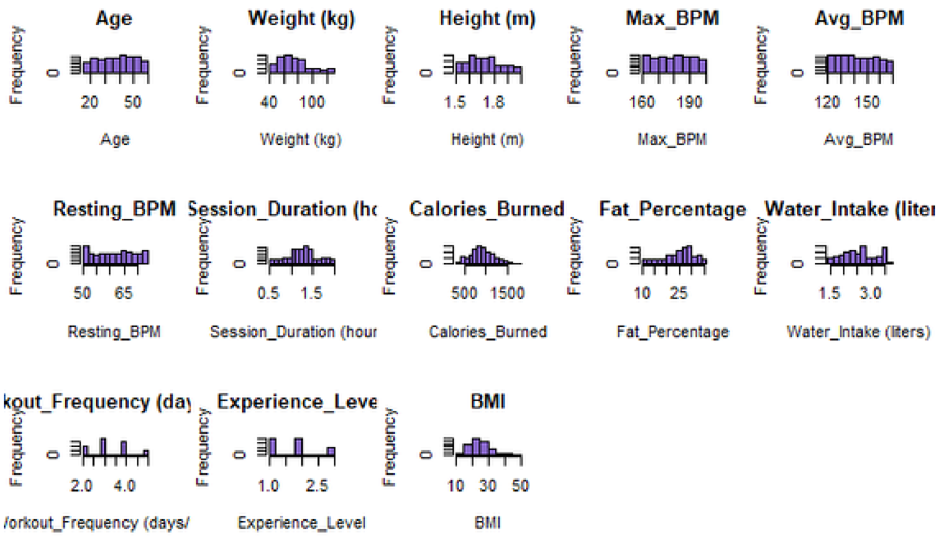
The dataset was loaded and examined in R to understand its structure, completeness, and quality. A summary function provided an overview of column data types, missing value percentages, and key statistics.

Key Findings:

- The dataset contained 973 samples and 15 columns.
- Categorical variables such as Gender and Workout_Type were stored as character type and converted to factors for proper analysis.
- Age ranges from 18 to 59 years, with most members between 25 and 45.
- Weight spans 45.9 to 120.3 kg, and Height ranges from 1.54 to 1.93 m; values are plausible for a gym population.
- BMI varies from 12.3 to 49.8, capturing underweight to obese profiles.

## Exploratory Data Analysis (EDA)

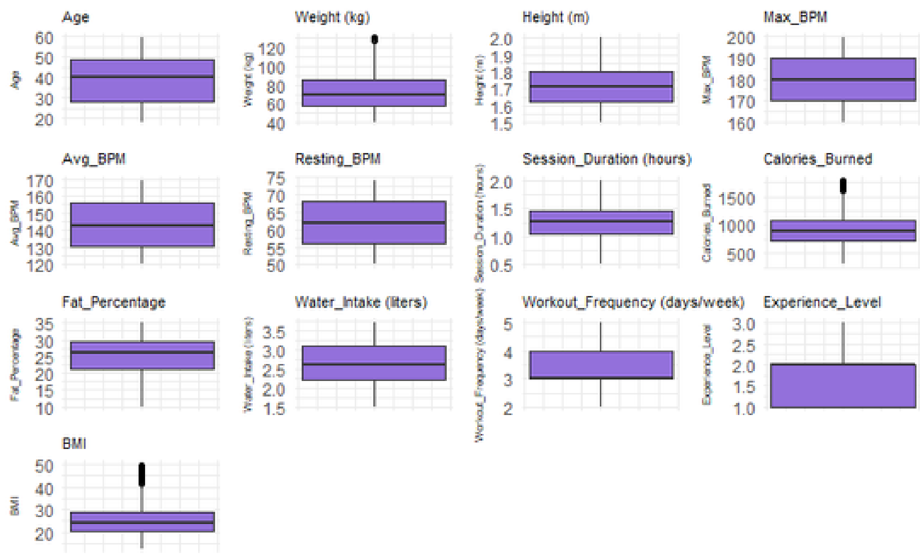
### Distribution of Numerical Features




- The dataset shows a young to middle-aged gym population (18–59 years, majority 25–45).
- Weight & Height: Predominantly 55–80 kg and 1.65–1.80 m, producing BMI values mostly in the 20–30 range.
- BMI: Includes healthy, overweight, and a smaller number of obese members (>30).
- Heart Rates:
- Max_BPM: 170–190 bpm.
- Avg_BPM: 140–160 bpm.
- Resting_BPM: 55–70 bpm.
- Workout Sessions: Usually 1–1.5 h, with Calories_Burned centered around 600–1,200 kcal per session (max 1,783).
- Fat Percentage: 22–30% common range, with isolated cases <15% and >35%.
- Water Intake: 2–3 liters/day is most common.
- Workout Frequency: 3–5 days/week dominates, with a higher share in Experience Level 2–3.

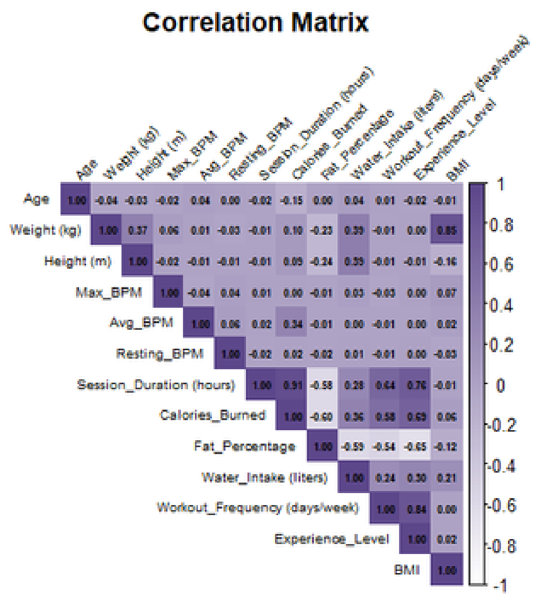
### Outlier Analysis

- Weight: Upper outliers >120 kg (possible obesity cases).
- BMI: Outliers >35–40 and <18.5 (obese/underweight extremes).
- Calories_Burned: Isolated values above 1,500 kcal.
- Most other metrics, including height, BPMs, and water intake, showed no extreme anomalies.




### Correlation Matrix

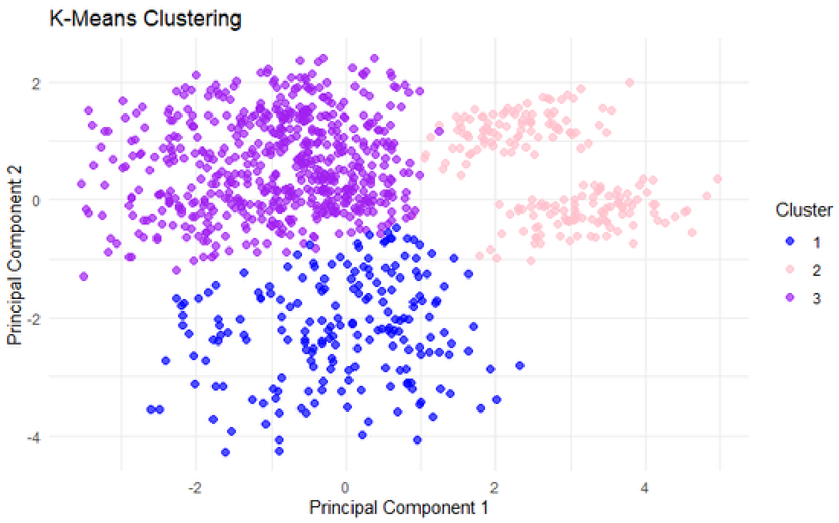
- Session_Duration ↔ Calories_Burned: +0.91, longer sessions burn more calories.
- Workout_Frequency ↔ Experience_Level: +0.84, more experienced members train more often.
- Workout_Frequency ↔ Calories_Burned: +0.76, higher training frequency increases total burn.
- Calories_Burned ↔ Fat_Percentage: –0.60, higher burn associates with lower fat %.
- Water_Intake ↔ Fat_Percentage: –0.59, better hydration correlates with lower fat %.
- Workout_Frequency ↔ Fat_Percentage: –0.65 , frequent training reduces fat %.



### Clustering Analysis (K-Means)

A K-Means clustering model was applied to segment members based on demographic, biometric, and training-related variables:

- Features used: Age, Weight, Height, BMI, Calories_Burned, Session_Duration, Workout_Frequency, Fat_Percentage, Avg_BPM.
- Cluster Selection: The Elbow Method indicated k = 3 as the optimal number of clusters.
- Visualization: Principal Component Analysis reduced the data to two dimensions for plotting.




### Cluster Profiles

### Cluster 1 – Advanced Performers (Blue)

- Age: 38 years (avg)
- Weight: 107.3 kg
- Height: 1.79 m
- Session Duration: 1.75 h
- Calories Burned: 1,263 kcal/session
- Highest training frequency and experience level.
- Profile: High-performance, advanced athletes training 5–6 days/week.

### Cluster 2 – Consistent Core (Pink)

- Age: 38 years
- Weight: 73.5 kg
- Height: 1.73 m
- Session Duration: 1.14 h
- Calories Burned: 861 kcal/session
- Trains regularly (4 days/week) at moderate intensity.
- Profile: Balanced, consistent gym-goers.

### Cluster 3 – New Starters (Purple)

- Age: 39 years
- Weight: 62.5 kg
- Height: 1.70 m
- Session Duration: 1.13 h
- Calories Burned: 799 kcal/session
- Lower frequency (3 days/week) and beginner-level experience.
- Profile: Inactive or entry-level members needing guidance.

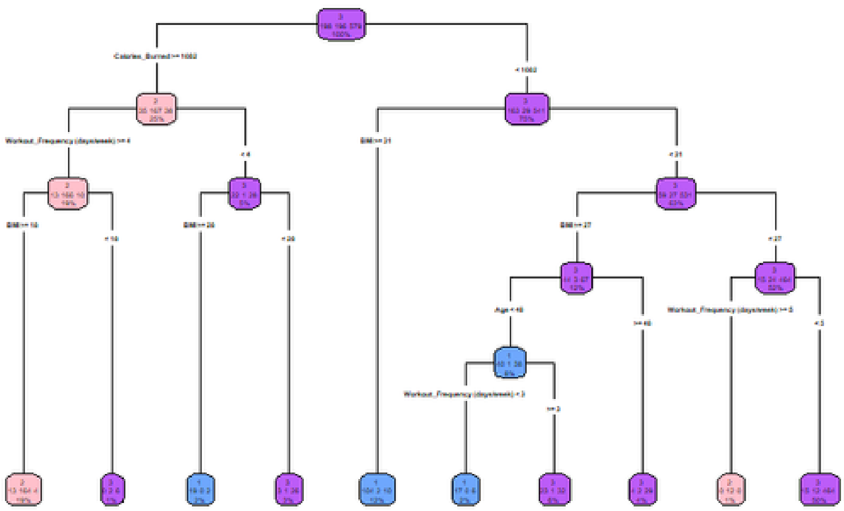
### Decision Tree Analysis

A CART (Classification and Regression Tree) model was built using the K-Means cluster labels as the target variable to create a rule-based approach for segment assignment.

- Target: Cluster (1, 2, or 3)
- Predictors: Calories_Burned, BMI, Workout_Frequency (days/week), Age.
- Goal: Provide an interpretable set of rules to classify new members into performance segments.

### Key Decision Rules

1. Calories_Burned ≥ 1032
    - Workout_Frequency ≥ 4 days/week → Cluster 2 (Consistent Core) if BMI ≤ 26, else Cluster 1 (Advanced Performers).
    - Workout_Frequency < 4 days/week → Cluster 3 (New Starters).
2. Calories_Burned < 1032
    - BMI > 31 → Cluster 1 (Advanced Performers).
    - BMI ≤ 31
        - BMI ≤ 27
            - Age < 48 and Workout_Frequency ≤ 3 → Cluster 3 (New Starters).
            - Else → Cluster 2 (Consistent Core).
        - BMI > 27 and Workout_Frequency < 5 → Cluster 3 (New Starters), else Cluster 2 (Consistent Core).



- Calories burned is the primary splitter, indicating its strong predictive power in determining performance level.
- BMI acts as a secondary driver, differentiating between higher- and lower-performance profiles.
- Workout frequency refines the classification, separating consistent performers from occasional gym-goers.
- The tree's simplicity allows easy operationalization: by collecting just 3–4 variables, gym staff can assign new members to tailored workout programs.

# Dashboard Overview

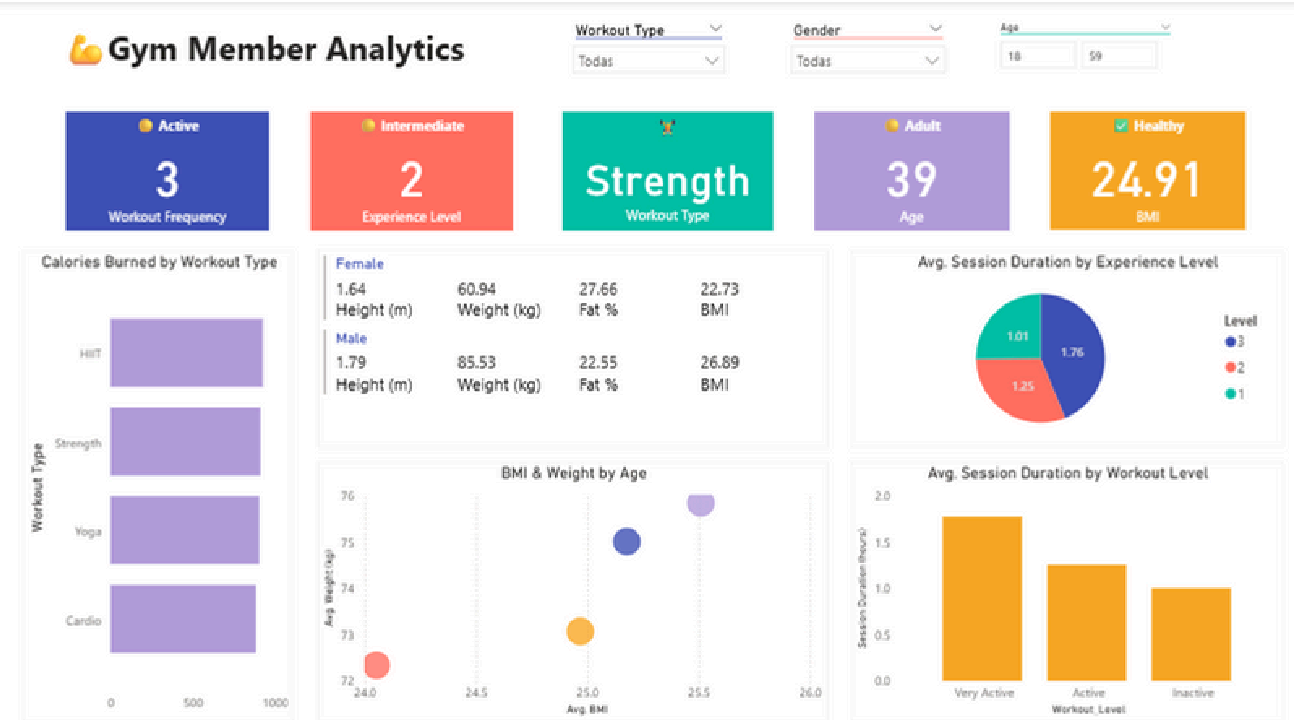
An interactive dashboard was developed in Power BI to consolidate key performance metrics, explore demographic and biometric patterns, and visualize workout behavior across different member segments.

Dashboard: *The Gym Member Analytics Dashboard file is available in the project folder.*

## Main KPIs

- **Workout Frequency:** Average of 3 days/week (Active).
- **Experience Level:** Predominantly Intermediate (Level 2).
- **Preferred Workout Type:** Strength training.
- **Average** Age: 39 years (Adult category).
- **Average** BMI: 24.91 (Healthy category).

## Gym Member Analytics Dashboard




- Calories Burned by Workout Type: HIIT sessions lead in average calories burned, followed closely by Strength training, indicating these as the most intensive workouts.
- Gender Comparison:
  - Female: Average height 1.64 m, weight 60.94 kg, fat % 27.66, BMI 22.73.
  - Male: Average height 1.79 m, weight 85.53 kg, fat % 22.55, BMI 26.89.

These values highlight gender-based differences in composition and workout impact.

- BMI & Weight by Age: Reveals that members in older age brackets (36–45) tend to have higher average weight and BMI compared to younger groups.
- Avg. Session Duration by Experience Level:
  - Level 3 (Advanced): 1.76 hours/session.
  - Level 2 (Intermediate): 1.25 hours/session.
  - Level 1 (Beginner): 1.01 hours/session.
- Avg. Session Duration by Workout Level: Very active members spend nearly twice as long per session compared to inactive ones.

# Conclusions & Recommendations

## Conclusions

- Members classified as Advanced not only train more frequently but also have significantly longer sessions and higher calorie expenditure, reinforcing the link between experience and workout intensity.
- HIIT and Strength training are the most effective workout types in terms of calorie burn, appealing to both male and female members.
- Gender-based differences in biometric indicators are evident, with males generally presenting higher weight, height, and BMI, while females have higher average body fat percentage despite lower body weight.
- Older members (36–45) tend to have higher BMI and weight compared to younger age groups, highlighting the importance of age-sensitive training programs.
- Members in the Very Active category show almost double the session duration of inactive members, indicating that engagement levels are a key driver of workout quality and outcomes.

## Recommendations

- Tailored Training Programs: Use clustering and decision tree insights to assign workout plans by segment, ensuring beginners have gradual progression and advanced members receive performance-focused training.
- Promote High-Impact Workout Types: Encourage broader adoption of HIIT and Strength sessions for members aiming to maximize calorie burn and fitness gains.
- Gender-Specific Strategies: Develop nutrition and training plans addressing the differing composition profiles of male and female members.
- Age-Sensitive Interventions: For members over 35, include targeted programs focusing on weight management, joint health, and functional strength.
- Engagement Incentives: Introduce challenges or rewards to encourage inactive or low-frequency members to increase training days and session length.
- Ongoing Monitoring: Leverage the Power BI dashboard for real-time tracking of performance metrics, ensuring programs remain aligned with member progress and retention goals.