**Exploratory Data Analysis (EDA) Report Summary**

**Title: Body Composition and Workout Routines Analysis for Gym Members to Optimize Physical Fitness**

**1. Introduction**

* **Objective:**
  + **Assess the gym members’ body composition and workout routines.**
  + **Explore the relationships between key factors such as workout frequency, session duration, calories burned, and body fat percentage.**
  + **Provide actionable insights to develop personalized workout routines for optimizing members' fitness levels.**

**2. Data Overview**

* **Dataset 1: *gym\_members\_exercise\_tracking***
  + **Records: 973**
  + **Features: 15 (e.g., Age, Gender, BMI, Fat Percentage, Workout Frequency, Calories Burned)**
  + **Source: [Gym Members Exercise Dataset on Kaggle](https://www.kaggle.com/datasets/valakhorasani/gym-members-exercise-dataset)**
* **Dataset 2: *com\_corp\_mta***
  + **Records: 278**
  + **Features: 8 (e.g., Time, BMI, Fat Percentage, Weight, Muscle Mass, Bone Mass)**
  + **Source: [Own registers via Garmin Index S2 scale](https://connect.garmin.com/modern/weight)**

**3. Data Cleaning and Feature Engineering**

* **Missing Values: No missing values detected.**
* **Outlier Detection:**
  + **Removed records based on specific criteria:**
    - **Unusual Fat Percentage: <3% or >50%**
    - **Extreme Weights: <30 kg or >200 kg**
    - **Unusual Heights: <1.2 m or >2.5 m**
* **Feature Engineering:**
  + **Created new columns for BMI Status and Fat Status based on standard health guidelines.**
  + **Estimated Muscle Mass Percentage and Basal Metabolic Rate (BMR) using calculated values.**

**Sources:**

* **National Health and Nutrition Examination Survey (NHANES)**
  + **\*\*Website\*\*: [NHANES Data](<https://www.cdc.gov/nchs/nhanes/index.htm>)**
  + **\*\*Description\*\*: NHANES provides detailed health and nutritional data, including body composition metrics for a diverse sample of the U.S. population. This dataset includes body fat percentage, lean body mass, and other anthropometric measurements.**
* **Revised Harris-Benedict Equations (1990)**
  + **\*\*Reference\*\*: Mifflin, M. D., St Jeor, S. T., Hill, L. A., Scott, B. J., Daugherty, S. A., & Koh, Y. O. (1990). \*A new predictive equation for resting energy expenditure in healthy individuals\*. The American Journal of Clinical Nutrition.**
  + **\*\*Description\*\*: This study provides updated coefficients for the Harris-Benedict equation, improving its accuracy. It is often referenced as the revised Harris-Benedict equation and is commonly used in modern BMR calculations.**

**4. Analysis and Key Metrics**

**Body Mass Index (BMI)**

* **Categories:**
  + **Underweight: BMI < 18.5**
  + **Normal Weight: 18.5–24.9**
  + **Overweight: 25–29.9**
  + **Obesity: BMI ≥ 30**

**Source: \*\*World Health Organization (WHO)\*\*: WHO offers standardized guidelines for BMI categories, which are used globally for assessing underweight, normal weight, overweight, and obesity in both men and women.**

**Body Fat Percentage:**

* **Healthy ranges differ by gender and age:**
  + **Typical Muscle Mass Percentage for Men by Age:**
    - **Ages 20-39\*\*: 33-39%**
    - **Ages 40-59\*\*: 30-36%**
    - **Ages 60-79\*\*: 27-34%**
  + **Typical Muscle Mass Percentage for Women by Age:**
    - **Ages 20-39\*\*: 24-30%**
    - **Ages 40-59\*\*: 22-28%**
    - **Ages 60-79\*\*: 19-25%**

**Source:\*\*American Council on Exercise (ACE)\*\*: ACE provides ranges for body fat percentage according to age and fitness levels for both men and women, widely referenced in health and fitness.**

* **Calculated using a formula based on gender, age, and experience level.**
* **Factors in lean body mass and adjusts for age.**

**5. Exploratory Analysis Results**

* **Descriptive Statistics:**
  + **Summarized key features (e.g., mean, median BMI and Fat Percentage).**
* **Correlation Analysis:**
  + **Examined relationships using a heatmap for a comprehensive overview.**
  + **Key insights:**
    - **Higher BMI values correlated with lower workout frequency.**
    - **Positive correlation between session duration and calories burned.**

**Visualizations:**

* **Histograms for Age, BMI, and Fat Percentage distributions.**
* **Scatter plots for relationships like session duration vs. calories burned.**
* **Bar charts showing the frequency of different workout types.**

**6. Key Findings**

* **BMI Analysis: Higher BMI often linked to lower workout frequency.**
* **Fat Percentage: High fat percentage associated with lower workout frequency and intensity.**
* **Session Duration: Longer sessions generally resulted in higher calories burned.**

**7. Limitations**

* **Sample Size: The relatively small sample may limit generalizability.**
* **Missing Health Metrics: Important data like blood pressure and muscle mass are absent.**
* **Potential Bias: Data may be skewed towards experienced gym members who track their workouts diligently.**

**8. Recommendations**

* **For High Fat Percentage Members:**
  + **Increase workout frequency to 4-5 days/week.**
  + **Consider higher intensity sessions if calories burned are consistently low.**
* **Data Collection: Include additional health metrics (e.g., waist measurements, muscle mass) for a more holistic analysis.**
* **Personalized Plans: Use insights to tailor workout plans based on body composition analysis.**

**9. Conclusion**

* **Summary: The EDA provided valuable insights into the relationship between workout routines and body composition, highlighting areas for potential improvement.**
* **Next Steps:**
  + **Collect more comprehensive health data for better analysis.**
  + **Tailor workout recommendations to optimize fitness outcomes for gym members.**

**Appendix**

* **Code Snippets: Python code used for data cleaning, feature engineering, and visualizations.**
* **Additional Visualizations: Supplementary charts supporting the analysis.**

**All inside the eda\_mta.jpynb**

**Tools Used**

* **Data Analysis: Pandas, NumPy**
* **Visualizations: Matplotlib, Seaborn**
* **Documentation: Jupyter Notebook**

**Note: This summarized report aims to provide a clear and actionable overview of the EDA findings, focusing on optimizing body composition and workout routines based on the data available.**