BUDT704: Data Processing and Analysis in Python

Project Title: Personalized Fitness Insights for Gym Members

Team Members:

- Abhishek Bhosale
- Kaushik Kannan
- Henry Kangten
- Yancy Longin
- Pranav Nair

1. Introduction

Physical inactivity has become a pressing global health issue, with nearly one-third (31%) of adults worldwide failing to meet recommended physical activity levels. This equates to approximately 1.8 billion people in 2022 alone. Our project addresses this concern by leveraging data science to create tailored fitness programs that drive engagement and improve health outcomes. The goal is to reduce physical inactivity to 13%.

Our mission is to enhance gym member engagement and retention through data-driven fitness insights. Our interactive system will assess current member health metrics, create customized workout plans, and refine recommendations through continuous data processing.

2. Business Objectives

1. **Enhance Engagement**: Utilize data-driven fitness recommendations to promote consistent gym attendance.
2. **Drive Retention**: Use predictive models to foster long-term member commitment.
3. **Personalization**: Offer workout recommendations based on age, gender, experience level, and body composition.

3. Dataset Description

Source: Gym Members Exercise Dataset

Key Attributes:

- **Age**: Age of the member.
- **Gender**: Male or Female.
- **BMI**: Body Mass Index.
- **Fat Percentage**: Percentage of body fat.
- **Workout Frequency**: Number of workouts per week.
- **Workout Type**: Type of workout (Yoga, Cardio, Strength, etc.).
- **Calorie Burn**: Calories burned per session.
- **Heart Rate Metrics**: Resting heart rate, maximum heart rate, and heart rate reserve.
- **Hydration**: Amount of water consumed during a session.

4. Data Cleaning and Preprocessing

1. **Data Cleaning**:

   - Renamed columns to follow consistent naming conventions.
   - Identified and removed missing, null, or duplicate values.
   - Corrected data types for attributes like age, workout type, and BMI.

2. **Feature Engineering**:

   - Created new features like **Age Group** (Youth, Adult, Senior) and **BMI Category** (Underweight, Normal, Overweight, Obese).
   - Added calculated fields such as **Hydration per Session** and **Workout Intensity** (based on heart rate data).

5. Exploratory Data Analysis (EDA)

5.1 Age and Fitness Journey

- **Finding**: Fitness commitment is consistent across all age groups, with the 26-35 age group recording the highest engagement (3.37 workout days per week).
- **Visualization**: Bar chart showing workout frequency across age groups.
- **Impact on Decisions**: This insight informs the creation of targeted engagement strategies, such as personalized workout challenges or rewards for this age group to sustain high participation levels.
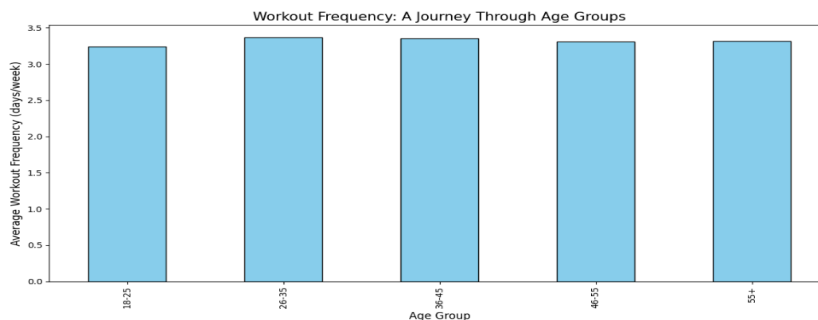

5.2 Gender Workout Dynamics

- **Finding**: Males have higher variability in calorie burns, while females display more consistent burns. Workout preferences differ slightly, with males favoring Yoga and females preferring Cardio.

- **Visualization**: Stacked bar chart of workout type preferences by gender and a boxplot of calorie burns.
- **Impact on Decisions**: Understanding gender preferences helps in designing inclusive workout schedules and marketing fitness programs that appeal to both genders.


Workout Type Preferences: Battle of the Genders
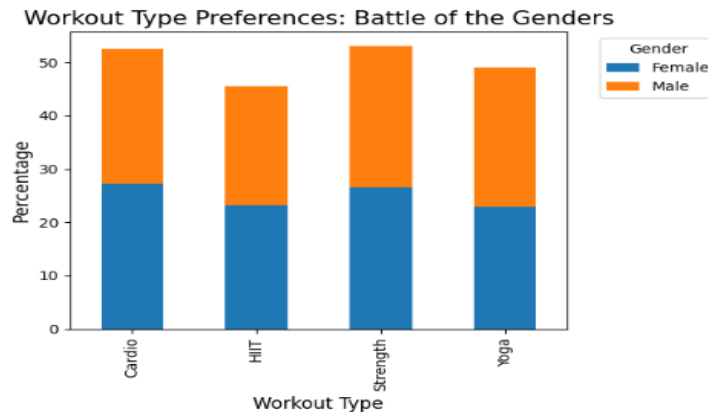
5.3 Body Composition Insights

- **Finding**: Body composition, particularly BMI and fat percentage, significantly impacts calorie burn. Members with higher BMI categories display distinct fat percentage clusters.
- **Visualization**: Scatter plot of BMI vs. fat percentage, pie chart of calorie distribution across BMI groups.
- **Impact on Decisions**: This insight supports the development of customized fitness plans targeting specific BMI categories, promoting body recomposition and healthier outcomes for members.


Body Composition: BMI vs Fat Percentage

Calorie Burn Across BMI Categories
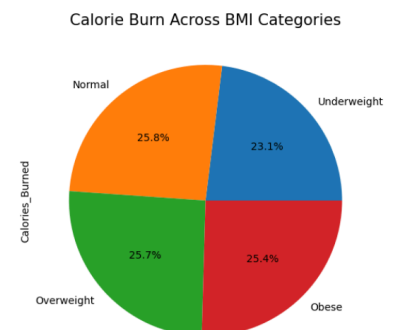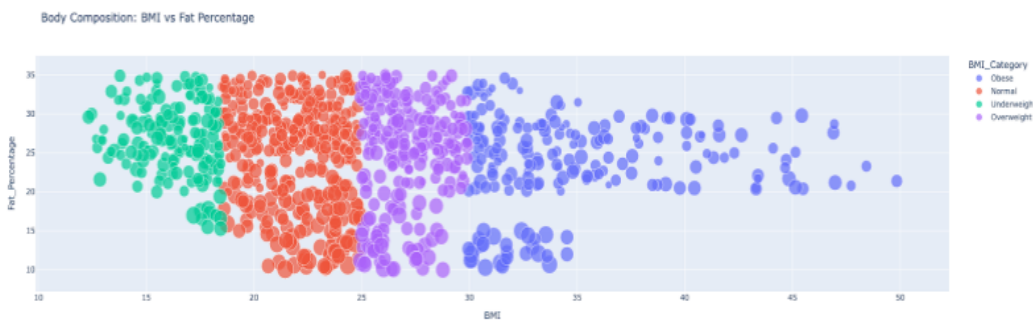
5.4 Fitness Progression Insights

- **Finding**: Experience correlates positively with workout frequency, session duration, and calorie burn. Advanced members prefer Strength and HIIT workouts, while beginners favor Yoga and Cardio.
- **Visualization**: Boxplot showing workout frequency and calorie burn by experience level.
- **Impact on Decisions**: The insight highlights the need for progressive fitness programs that evolve with member experience, promoting long-term engagement and reducing dropout rates.
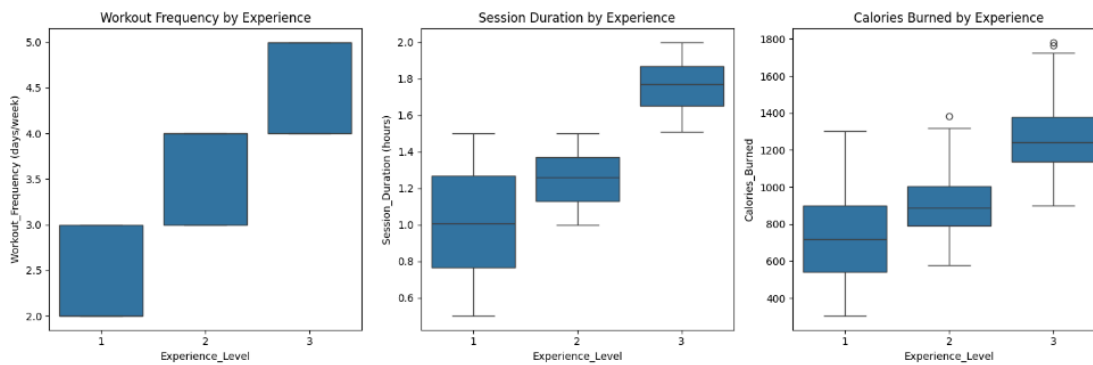

5.5 Heart Health Insights

- **Finding**: Higher calorie burns are associated with greater workout intensity, as reflected by higher maximum heart rates. Variability in heart rate reserve across workout types demonstrates the differing cardiovascular demands of HIIT vs. Yoga.
- **Visualization**: Scatter plot of resting vs. maximum heart rate, color-coded by calorie burn, and a boxplot of heart rate reserve by workout type.
- **Impact on Decisions**: Insights into heart health inform the design of intensity-based workout plans and the promotion of cardiovascular health initiatives.

Heart Rate Reserve Across Workout Types
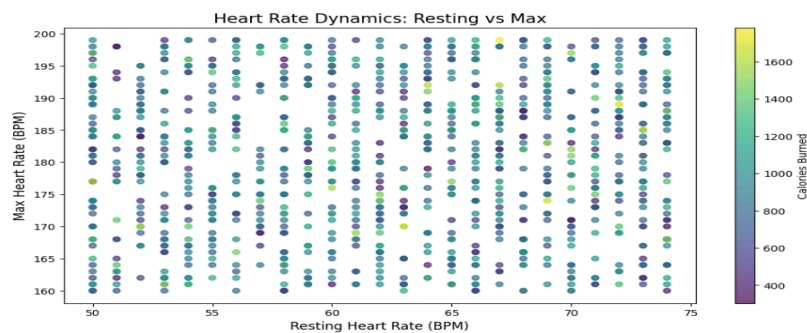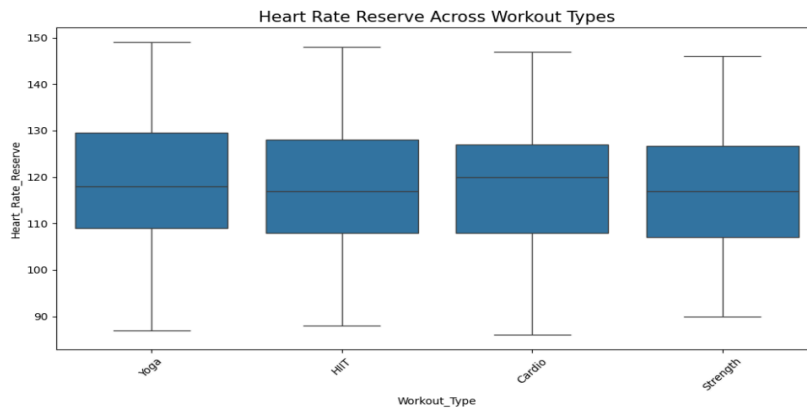
5.6 Hydration and Performance

- **Finding**: Members who consume more water during workouts exhibit longer session durations and higher calorie burns.
- **Visualization**: Scatter plot of water intake vs. calorie burn, color-coded by session duration.
- **Impact on Decisions**: This insight supports gym initiatives promoting hydration through educational campaigns and water station placements to boost member performance and engagement.

Hydration Impact: Water Intake vs Calorie Burn
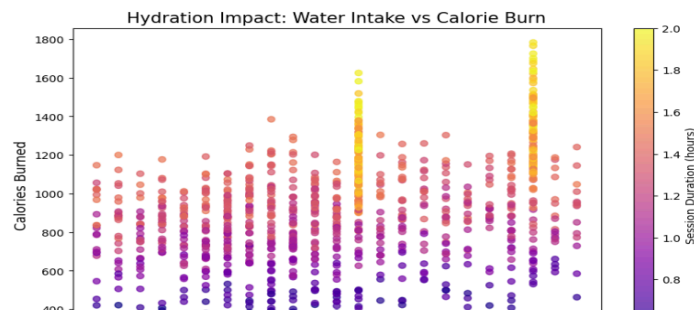
6. Data Modeling

6.1 Clustering Analysis

- **Method**: K-means clustering segmented members into 4 distinct profiles:
  - **Cluster 0**: Engaged Middle-aged (Age 37.38, BMI 25.08, 4.47 workouts/week)
  - **Cluster 1**: Older Moderate (Age 47.90, BMI 20.95, 3.10 workouts/week)
  - **Cluster 2**: Young Beginners (Age 26.10, BMI 23.85, 2.93 workouts/week)
  - **Cluster 3**: Older Weight Focus (Age 43.54, BMI 35.05, 2.88 workouts/week)

6.2 Predictive Models

- **User Input-Output Predictive Model**

  - **Objective**: Provide personalized fitness recommendations based on user inputs like age, gender, and BMI.
  - **Impact**: Enhances user experience by offering tailored workout recommendations and predictive accuracy.
  - **Features**: Real-time feedback and dynamic fitness plan suggestions.
- **Decision Tree Classifier**

  - **Objective**: Predict member experience level.
  - **Performance**: 78% accuracy, 1.00 F1-score for experienced members.
- **Decision Tree Regressor**

  - **Objective**: Predict calories burned based on session duration.
  - **Performance**: R-squared = 0.79, Mean Squared Error = 17,202.

7. Results and Discussion

- **Member Profiles**: Identified four distinct profiles for targeted fitness programming.
- **Predictive Modeling**: Models achieved strong performance.
- **User Input-Output Model**: Offers real-time, personalized fitness recommendations, enhancing user experience and member satisfaction.

Sample Input & Output

```
 PERSONALIZED FITNESS RECOMMENDATION SYSTEM
1. Get Workout Plan Recommendation
2. Health Risk Assessment
3. Exit
Enter your choice (1/2/3): 1
Enter your age: 27
Enter your BMI: 26
Enter your gender (Male/Female): Male
```

```
 PERSONALIZED WORKOUT RECOMMENDATIONS
Recommended Workout Type: Strength
Recommended Workout Frequency: 4 days/week
Recommended Water Intake: 2.6 liters/day
Estimated Calories Burned per Day: 863 calories
```

8. Recommendations

1. **Tailored Fitness Programs**: Offer personalized workout plans based on cluster membership.
2. **Hydration Campaigns**: Encourage water intake during workouts to boost performance.
3. **Engagement by Age**: Design age-specific programs targeting the 26-35 age group.
4. **Gender-Specific Workouts**: Create gender-specific workout plans to match preferences.

9. Limitations and Future Work

- **Data Gaps**: Absence of detailed daily activity logs and member feedback.
- **Model Improvements**: Real-time tracking and live feedback enhancements.

10. References

1. Guide to Data Science Project Reports - The Clever Programmer
2. Data Science Projects with Source Code - Interview Query
3. Data Science Portfolio Guide - KDnuggets