Personalized Fitness Segmentation with Actionable Insights

Ms. Anju Tiwari[1], Mr. Bhushan Gaonkar[2]
[1]Student,[2]Professor, Institute of Computer Science, Mumbai Educational Trust – MET ICS, Mumbai, India

***Abstract*— This research paper aims to personalize workout plans, reducing injury risk and improving adherence, trying to find clusters based on cardio exercises.**

Keywords— data driven,fitness, cardio, machine learning, KNN

I.INTRODUCTION

A. Fitness Statistics

Fitness is a multifaceted concept that encompasses various types of workouts, each contributing uniquely to caloric expenditure and overall health. Aerobic exercises, such as running, cycling, and swimming, are particularly effective in burning calories due to their ability to elevate heart rates and increase oxygen consumption. Research indicates that the intensity of these workouts plays a crucial role in determining caloric burn; higher intensity aerobic activities lead to greater energy expenditure during and after exercise (Kravitz, 2023). In contrast, resistance training, which focuses on building muscle strength through weight lifting and bodyweight exercises, also significantly impacts caloric expenditure. Not only does resistance training enhance muscle mass, but it also increases resting metabolic rate, allowing individuals to burn more calories even at rest (Loftin et al., 1988).

Moreover, combining aerobic and resistance training can maximize caloric burn and improve overall fitness. This integrated approach allows individuals to engage multiple muscle groups and elevate heart rates, thereby enhancing the effectiveness of their workouts (Toner et al., 1990). Understanding the relationship between different types of workouts and caloric expenditure is essential for individuals seeking to manage their weight and improve their health. The number of calories burned during exercise is influenced by various factors, including the type of exercise, its intensity, and individual fitness levels. For instance, lower body exercises typically result in higher caloric expenditure compared to upper body exercises due to the larger muscle mass involved (Kravitz, 2023).

To accurately measure caloric expenditure, researchers often utilize indirect calorimetry, which assesses oxygen consumption and carbon dioxide production during physical activity. This method provides valuable insights into how different workouts affect energy expenditure and can guide individuals in selecting the most effective exercise regimens (Lusk, 1928). Ultimately, choosing the right workout is crucial for long-term adherence to a fitness routine, which is essential for achieving sustained health benefits and effective weight management. By understanding the dynamics of workout types, intensity, and caloric burn, individuals can make informed decisions that align with their fitness goals and preferences.

In addition to the type and intensity of exercise, the duration and frequency of workouts are critical factors influencing overall calorie burn and fitness outcomes. Studies have shown that consistent engagement in physical activity, even at moderate intensity, can lead to significant improvements in cardiorespiratory fitness, metabolic health, and weight control (Physical Activity Guidelines Advisory Committee, 2018). For example, a meta-analysis evaluating aerobic interventions demonstrated that regular moderate-to-vigorous aerobic exercise sustained over several weeks leads to enhanced cardiovascular function and increased caloric expenditure during daily activities (Swain & Franklin, 2006). Furthermore, the American College of Sports Medicine recommends that adults accumulate at least 150 minutes of moderate aerobic exercise per week for general health benefits, underscoring the importance of exercise frequency and duration in fitness regimens (Garber et al., 2011). These findings reinforce that sustained physical activity, rather than isolated intense sessions alone, is key to promoting long-term energy balance and overall well-being.

Moreover, recent research highlights the role of high-intensity interval training (HIIT) as an efficient method for maximizing calorie burn within shorter workout durations. HIIT alternates brief bouts of intense exercise with recovery periods, eliciting significant post-exercise oxygen consumption (EPOC) that contributes to elevated total daily energy expenditure (Buchheit & Laursen, 2013). Studies comparing HIIT with steady-state aerobic training indicate that despite its shorter duration, HIIT may produce comparable or even superior improvements in metabolic rate and fat oxidation (Gibala et al., 2012). This makes HIIT an attractive option for individuals with limited time who aim to improve fitness and reduce body fat. However, the intensity of HIIT requires careful consideration of individual fitness levels to avoid overtraining or injury. Overall, understanding various exercise modalities, from continuous aerobic training to interval-based workouts, allows fitness enthusiasts and researchers alike to tailor programs that optimize calorie burning and health outcomes.

Machine learning has revolutionized many industries, including the fitness domain, by enabling personalized and adaptive workout experiences through fitness applications. By analyzing large volumes of user data such as activity patterns, heart rate, sleep quality, and calorie intake, machine learning algorithms can tailor fitness programs to individual needs, preferences, and progress levels. This personalization helps users optimize workout effectiveness, maintain motivation, and reduce injury risks by recommending the right intensity and types of exercises. Additionally, fitness apps powered by machine learning can predict users' performance trends and potential setbacks, offering timely interventions or adjustments to goals. Furthermore, these advanced technologies facilitate real-time feedback and coaching by recognizing patterns in movement and suggesting improvements. Overall, integrating machine learning in fitness applications enhances user engagement and promotes healthier lifestyles through smart, data-driven fitness solutions.

II. METHODOLOGY

This paper aims to get optimized fitness program with a data driven approach on behavior clustering

A. Data Collection

Dataset form imported from Kaggle, regarding the fitness with 10000 rows representing 10000 people of varied age group, weight, height and workout type they perform.

Feature	**Data Type**	**Unique Values**	**Range/Description**
User ID	Integer	10,000	Unique identifier for each individual
Age	Integer	42	Range: 18 to 59
Gender	Categorical	3	Categories: Male, Female, Other
Height	Continuous	50	Varies in height
Weight	Continuous	70	Varies in weight
Workout Type	Categorical	6	Categories: e.g., Cardio, Strength, Yoga, etc.
Workout Duration	Continuous	110	Range: 10 to 119 minutes
Calories Burned	Continuous	-	Range: 100 to 999 kcal
Heart Rate	Continuous	100	Range: 80 to 179 bpm
Steps Taken	Continuous	7,767	Range: 1,000 to 19,998 steps
Distance	Continuous	1,449	Range: 0.5 to 15 km
Workout Intensity	Categorical	3	Categories: Low, Moderate, High
Sleep Hours	Continuous	61	Range: 4 to 10 hours
Water Intake	Continuous	1	Constant: 1.9 liters (or possibly erroneous entry)
Daily Calories Intake	Continuous	2,445	Range: 1,500 to 3,999 kcal
Resting Heart Rate	Continuous	40	Range: 50 to 89 bpm

VO2 Max	Continuous	1	Constant value: 38.4 (no variability)
Body Fat	Continuous	1	Constant value: 28.5 (no variability)
Mood Before Workout	categorical	4	Categories: e.g., Happy, Neutral, Sad, Angry
Mood After Workout	categorical	3	Categories: e.g., Happy, Neutral, Tired

B. *Data Pre-processing*

To enable better analysis of age-related trends, the continuous 'Age' variable was transformed into discrete intervals of 5 years (e.g., 15–19, 20–24, ..., up to 90–94). A new categorical column named Age_Group was created to represent these bins, thus facilitating grouped insights across different stages of life.

Additionally, some data quality issues were observed:

- Water Intake had a constant value (1.9 L), suggesting a potential error or placeholder.
- Both VO_2 Max and Body Fat metrics were constant across all entries (38.4 and 28.5, respectively), indicating that these features do not contribute to variance in clustering and may be excluded or treated as constants during modelling.

The preprocessing also involved:

- Handling missing values (e.g., for Mood variables).
- Standardizing numeric features for clustering.
- Encoding categorical variables for machine learning compatibility.

C. *Exploratory Data Analysis*

Histograms were plotted for all selected numerical variables to visualize their distribution patterns. Each variable was represented using 20 bins, allowing for an understanding of the spread, central tendency, and potential skewness or outliers
A correlation matrix heatmap was generated to examine the linear relationships between selected numerical variables. The heatmap used a diverging 'coolwarm' color scheme, with annotated correlation coefficients, to highlight the strength and direction of associations between variables.

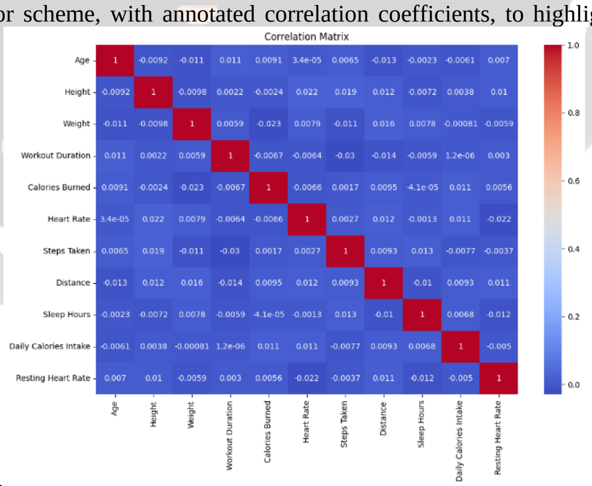


Figure 1: Correlation matrix heatmap

The correlation matrix heatmap reveals that most numerical variables in the dataset exhibit very weak correlations with each other, with correlation coefficients ranging between -0.03 and 0.03. Notable observations include:

- Age has negligible correlation with all other variables, indicating it does not have a linear relationship with health or activity metrics in this dataset.
- Workout Duration and Calories Burned show a very weak negative correlation (-0.0067), suggesting a minimal linear association between exercise time and calories burned.

- Heart Rate and Resting Heart Rate are uncorrelated (-0.022), implying that peak workout heart rate does not reflect resting cardiovascular condition.
- Steps Taken and Distance have a slight positive correlation (0.0093), which aligns with expectations since more steps generally indicate more distance, though the relationship is extremely weak.
- Daily Calories Intake, Sleep Hours, and Workout Intensity (if included in other analyses) also show little to no correlation with primary workout metrics.

Overall, the heatmap suggests that the dataset lacks strong linear dependencies, highlighting the potential need for non-linear modelling techniques to uncover deeper relationships.

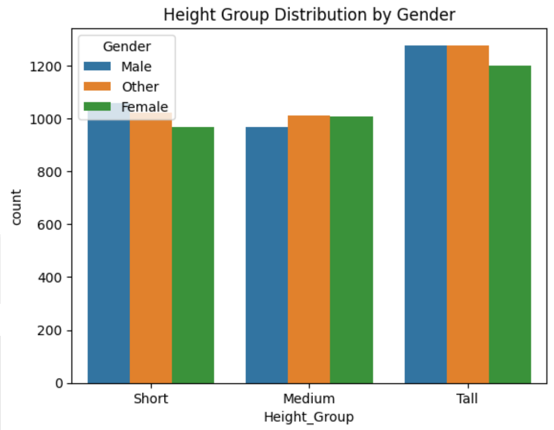


Figure 2: Barplot of height distribution w.r.t gender

Figure 2 illustrates the distribution of individuals across height groups (Short, Medium, Tall) segmented by gender (Male, Female, Other). The 'Tall' category has the highest count across all gender groups, with 'Male' and 'Other' being nearly identical and higher than 'Female'. The 'Medium' category shows relatively balanced representation across genders, while the 'Short' category has slightly higher counts in 'Male' and 'Other' compared to 'Female'.

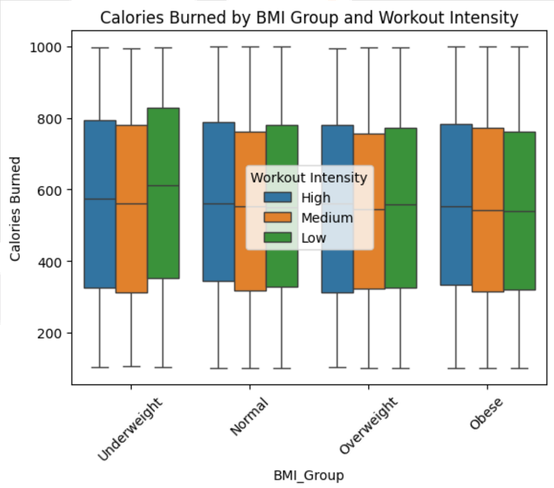


Figure 3: Boxplot of categorized weight vs calories burnt w.r.t worktout intensity

Figure 3 shows the distribution of calories burned across different BMI groups (Underweight, Normal, Overweight, Obese) and workout intensity levels (Low, Medium, High). Across all BMI groups, higher workout intensity is generally associated with higher median calories burned. The spread of calories burned remains consistent across BMI categories, indicating that workout intensity, rather than BMI, is the dominant factor influencing energy expenditure.

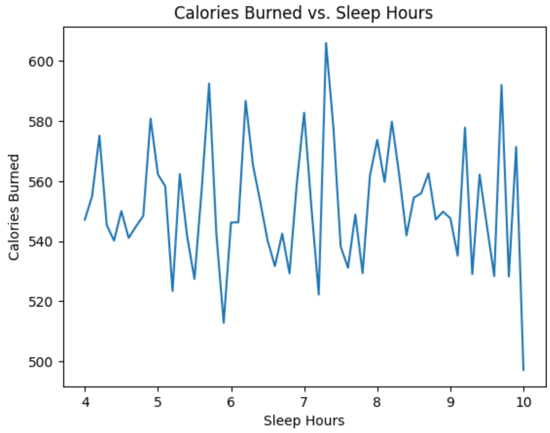


Figure 4 : Line plot of Calories burnt vs Sleep hours

Figure 4 illustrates the relationship between sleep hours and average calories burned. The plot reveals a relatively stable trend, with only minor variations in calories burned across different sleep durations. According to the grouped statistics, individuals sleeping 6 hours or less burned an average of 550.17 calories, those sleeping between 6 and 8 hours burned 553.48 calories, and those with 8 to 10 hours of sleep burned 552.68 calories. This suggests that sleep duration has a negligible impact on calories burned.

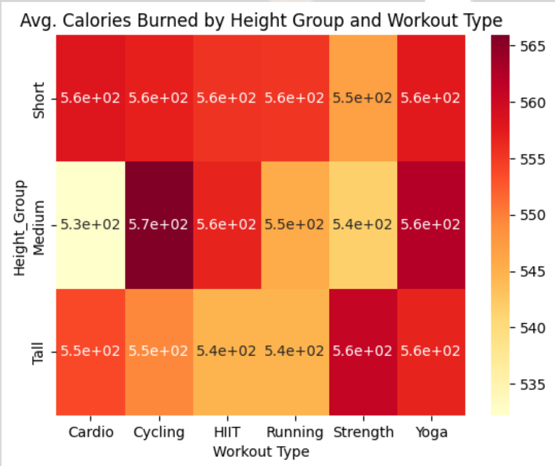


Figure 5. Correlation matrix heatmap of Calories Burned by Height Group and Workout Type

Figure 5 displays a heatmap of the average calories burned across different height groups and workout types. Short individuals exhibit consistently high average calorie expenditure across all workout types, with values around 560 kcal. Medium-height individuals show the highest calorie burn during cycling (~570 kcal), while tall individuals burn more calories during strength and yoga sessions. Overall, workout type has a stronger influence on calorie expenditure than height group.

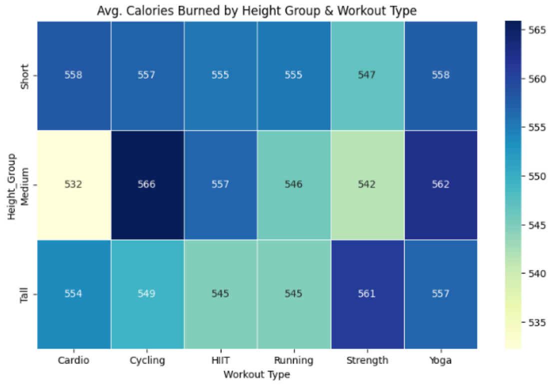


Figure 6 Correlation matrix heatmap if Avg.Calories Burned by Height Group and Workout Type

Figure 6 presents a heatmap depicting the average calories burned across height groups and workout types. Short individuals demonstrate consistently high energy expenditure, especially in cardio, yoga, and cycling (≈558 kcal). Medium-height individuals show a pronounced spike in calorie burn during cycling (566 kcal), similar to a sudden peak on a terrain map, while their lowest output is seen in cardio (532 kcal). Tall individuals exhibit a more balanced distribution, with the highest calories burned during strength training (561 kcal), suggesting a strong correlation between body stature and resistance-based exercises. The pattern resembles a heat engine where different body frames extract energy differently from various workout fuels.

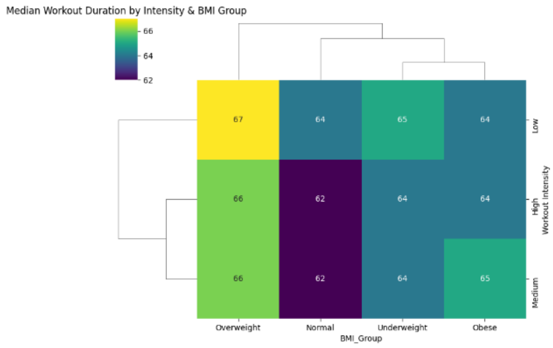


Figure 7. Cluster heatmap of median workout duration by Intensity & BMI Group

This cluster heatmap illustrates the median workout duration segmented by workout intensity and BMI group. A few insights emerge:

- Low-intensity workouts are most prolonged among overweight individuals (67 minutes), suggesting that they may prefer or require extended low-impact sessions.
- High and medium intensities exhibit very similar durations across most BMI categories (~64–66 minutes), indicating a general consistency in session length despite variations in body mass
- The dendrogram reveals that "Normal" and "Underweight" BMI groups cluster closely, reflecting similar workout duration patterns across intensities.
- Interestingly, high-intensity sessions for normal BMI individuals are shortest (62 minutes), perhaps reflecting efficient, high-output routines

The analysis of daily calorie intake by gender and workout type reveals notable patterns in both average consumption and variability. Among female participants, the highest average calorie intake is observed during Cycling (2,772 kcal) and HIIT (2,768 kcal), while the lowest is seen during Running (2,723 kcal). These values suggest that endurance-based workouts may be associated with more controlled dietary intake among females. The variability is highest in Strength training (±739 kcal) and Yoga (±732 kcal), indicating a wide range of dietary habits within these workout categories. For male participants, the highest average intake occurs during Running (2,815 kcal), reflecting the elevated energy demands typically associated with high-intensity endurance activities. Intake across other workout types remains relatively consistent, ranging from 2,731 to 2,752 kcal. However, standard deviation peaks again in Running (±745 kcal), pointing to diverse energy expenditure or nutritional strategies. Individuals identifying as Other show their highest intake during Cycling (2,764 kcal) and Running (2,761 kcal), with slightly lower intake during Strength training and HIIT. Although their average intake values are generally comparable to other groups, variability remains significant, particularly in Cardio (±733 kcal). Overall, Running consistently correlates with higher caloric intake across all genders, likely due to greater exertion levels, while Strength training and Yoga exhibit the most variation in intake, suggesting more individualized or inconsistent nutritional behaviors. These findings underscore the complex interplay between gender, workout type, and dietary practices.

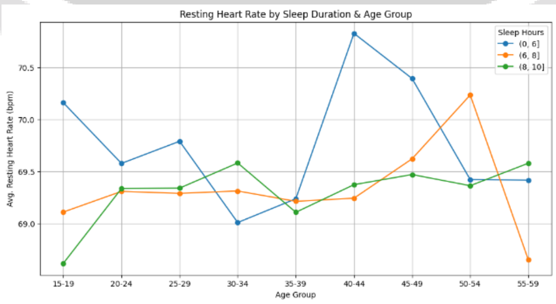


Figure 8. Line graph of Resting Heart Rate by Sleep Duration & Age Group

The Figure 8 reveals several key patterns in the relationship between sleep duration, age, and average resting heart rate:

Elevated Resting Heart Rate with Short Sleep Duration: Across nearly all age groups, individuals sleeping 0–6 hours consistently show higher resting heart rates. This pattern is especially pronounced in the 40–44 and 45–49 age groups, suggesting that insufficient sleep may have a compounding effect on cardiovascular load as individuals age.

Stability in Resting Heart Rate with Adequate Sleep: For the 6–8 and 8–10hour sleep categories, resting heart rates remain relatively stable across age groups. The 8–10hour group demonstrates the most consistent trend, with minor fluctuations, indicating a potential protective effect of longer sleep on cardiac function.

Age-Related Variability: In the short sleep group (0–6 hours), there is greater variability in resting heart rate across age groups, with noticeable peaks at ages 40–44 and 45–49, and a sharp decline at 50–54. This volatility contrasts with the flatter trajectories seen in the longer sleep groups.
Lowest Resting Heart Rates in Long Sleepers (8–10 hours): Particularly in the 15–19 and 55–59 age groups, individuals sleeping 8–10 hours exhibit the lowest resting heart rates. This suggests that both younger and older adults may benefit the most from extended sleep in terms of cardiovascular recovery.
Anomalies in Midlife Sleep Patterns: The 50–54 age group in the 6–8 hour category shows a spike in resting heart rate, which deviates from the preceding and succeeding age groups. This may point to specific physiological or lifestyle stressors affecting this demographic, such as occupational stress or early signs of cardiovascular aging.
Convergence in Middle Age: Between ages 30–39, the differences between sleep groups narrow, particularly between the 6–8 and 8–10 hour categories, suggesting a possible plateau in sleep-related heart rate benefits during early middle age.

D. *Training Machine Learning Model*

K-means clustering was applied to categorize users based on workout and health-related features. Features such as 'Workout Type', 'Workout Duration', 'Workout Intensity', 'Heart Rate', 'Calories Burned', 'Sleep Hours', and 'Daily Calories Intake' were selected for clustering. Categorical features were one-hot encoded, while numerical features were standardized for scaling.

The elbow method was used to determine the optimal number of clusters, resulting in 4 clusters. The K-means algorithm was then applied, and the data was assigned to one of the four clusters. Human-readable labels were assigned to each cluster:
Cluster 0: High-Intensity Over-Trainers
Cluster 1: Balanced Exercisers
Cluster 2: Sedentary Sleep-Deprived
Cluster 3: Endurance Specialists
These cluster labels were added to the dataset for further analysis.
The key statistics per cluster were calculated as follows:

1.Sedentary Sleep-Deprived:
Average Workout Duration: 90.17 minutes
Average Calories Burned: 794.42 kcal
Average Heart Rate: 127.83 bpm
Average Sleep Hours: 6.77 hours
Most Common Mood After Workout: Energized

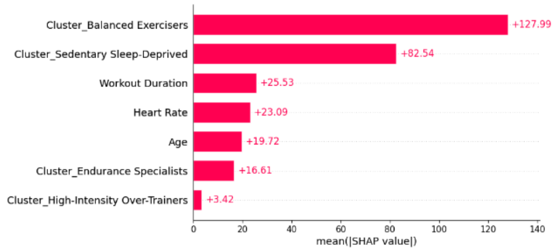
2.High-Intensity Over-Trainers:
Average Workout Duration: 38.34 minutes
Average Calories Burned: 587.36 kcal
Average Heart Rate: 129.40 bpm
Average Sleep Hours: 8.58 hours
Most Common Mood After Workout: Neutral

3. Endurance Specialists:
Average Workout Duration: 38.82 minutes
Average Calories Burned: 524.00 kcal
Average Heart Rate: 131.74 bpm
Average Sleep Hours: 5.36 hours
Most Common Mood After Workout: Neutral

4.Balanced Exercisers:
Average Workout Duration: 90.37 minutes
Average Calories Burned: 299.81 kcal
Average Heart Rate: 129.17 bpm
Average Sleep Hours: 7.19 hours
Most Common Mood After Workout: Fatigued

The clusters were encoded as categorical features using one-hot encoding. These encoded cluster labels were then combined with other features like 'Workout Duration', 'Heart Rate', and 'Age' to form the supervised learning dataset (X_supervised). The target variable for prediction was 'Calories Burned'.An XGBoost regressor model was trained on 80% of the data, with the remaining 20% used for testing. The feature importance analysis revealed the following:
Cluster_Balanced Exercisers: 60.41% importance
Cluster_Sedentary Sleep-Deprived: 36.14% importance
Cluster_Endurance Specialists: 1.11% importance
Cluster_High-Intensity Over-Trainers: 0.61% importance
Workout Duration: 0.60% importance
Age: 0.58% importance
Heart Rate: 0.55% importance

This indicates that the cluster label 'Balanced Exercisers' plays the most significant role in predicting calories burned,



followed by 'Sedentary Sleep-Deprived'.

Figure 9. SHAP bar plot

The SHAP (SHapley Additive exPlanations) bar plot highlights the most influential features contributing to the model's predictions. The x-axis represents the mean absolute SHAP value, which quantifies each feature's overall impact on the output.
Key observations:
Dominant Influence of Clusters: Cluster membership exerts the strongest influence, particularly the "Balanced Exercisers" and "Sedentary Sleep-Deprived" clusters, with mean SHAP values of +127.99 and +82.54 respectively. This suggests that behavioral patterns grouped in these clusters play a crucial role in the model's decision-making, potentially outweighing individual metrics like workout duration or heart rate.

Physical Activity & Recovery Patterns: The high impact of the "Balanced Exercisers" cluster likely reflects the positive predictive power of consistent, moderate physical activity combined with adequate recovery and sleep. Conversely, the substantial influence of the "Sedentary Sleep-Deprived" cluster may indicate high risk or poor health outcomes associated with inactivity and insufficient rest.
Individual Metrics Still Relevant: Workout Duration (+25.53), Heart Rate (+23.09), and Age (+19.72) also contribute meaningfully, although less than the behavioral clusters. These factors likely provide additional granularity within each cluster.
Specialized Training Effects: The "Endurance Specialists" cluster (+16.61) has moderate impact, likely reflecting the physiological adaptations and training consistency characteristic of this group. In contrast, the "High-Intensity Over-Trainers" cluster has minimal influence (+3.42), possibly due to erratic patterns or mixed health signals within that group.
Overall, cluster-based behavioral profiles explain a larger portion of the model's variance than isolated physiological variables, highlighting the importance of holistic lifestyle patterns in predictive modeling of health-related outcomes.

The ANOVA test was performed to determine if there are significant differences in calories burned across the clusters. The resulting p-value was 0.0000, indicating that there are significant differences in calories burned between the clusters ($p < 0.05$).
The Chi-Square test was used to assess whether mood after workout varies by cluster. The p-value obtained was 0.3836, which is greater than 0.05, indicating that mood after workout does not vary significantly across clusters.

Figure 10. LineGraph of Workout Duration Impact by Cluster

The partial dependence plot shows the marginal effect of workout duration on the model's predicted output (e.g., calories burned), with a focus on how this effect varies across different clusters.

E. *Observation*

- Nonlinear Relationship: The relationship between workout duration and predicted outcome is nonlinear, with noticeable fluctuations across the range. This suggests that the model captures complex interactions, possibly modulated by cluster-specific traits such as intensity, rest habits, or metabolic efficiency.

- Higher Sensitivity at Moderate Durations (30–50 minutes): The most pronounced variability and spikes in partial dependence occur between 30 and 50 minutes. This indicates that moderate-duration workouts lead to significant changes in predicted outcomes, possibly due to being a threshold for meaningful exertion across most individuals.

- Stabilization Beyond 60 Minutes: Beyond 60 minutes, the partial dependence curve stabilizes with less fluctuation, suggesting diminishing marginal gains or a plateau in predicted benefit. This may reflect physiological limits or fatigue effects modeled through cluster interaction.

- Inter-Cluster Variability Embedded: The high-frequency variation in the curve, despite being an average across clusters, reflects differing cluster responses to the same workout duration. This implies that some clusters (e.g., sedentary vs. endurance) experience sharply different outcomes from similar durations, reinforcing the importance of personalization in training recommendations.

- Model Sensitivity to Short and Moderate Durations: Below 30 minutes, the predicted outcome increases gradually, implying the model captures low-yield gains from brief workouts, which still contribute positively, especially in inactive populations.

In sum, workout duration exerts a significant and nonlinear influence on predicted outcomes, with the strongest model sensitivity occurring in the 30–50 minute range, shaped by underlying cluster characteristics.

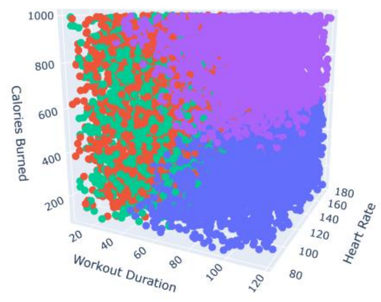


Figure 11. 3D Plot of Cluster w.r.t. Calories Burned, Workout Duration and Heard Rate

Observations on Workout Clusters:

1. Cluster: Balanced Exercisers
Workout Type Distribution: Balanced Exercisers engage in a variety of workout types, with the most frequent being High-Intensity Interval Training (HIIT), Cardio, Running, Strength Training, Cycling, and Yoga. HIIT had the highest occurrence (449), closely followed by Cardio (415) and Running (409). This indicates a balanced approach to different types of exercises.
Age Group Distribution: The cluster primarily consists of individuals in the 45-49 age group (325 occurrences), with other prominent age groups being 50-54 (304 occurrences), and 40-44 (300 occurrences). The youngest age groups (15-19 and 20-24) show lower participation, while there is a complete absence of individuals from the 60-64 and above age groups. This suggests that the cluster is most common among middle-aged adults, with a notable gap in older age groups.
Key Metrics: On average, individuals in this cluster burn 299.81 kcal per session, with a moderate average heart rate of 129.17 bpm. The average workout duration is 90.37 minutes, suggesting that participants in this cluster engage in longer workout sessions. However, there are missing values for the average workout intensity, mood before workout, and mood after workout, which may require further data analysis.

2. Cluster: Endurance Specialists
Workout Type Distribution: The most frequent workout type is HIIT (443 occurrences), followed by Strength Training (430 occurrences), and Running (423 occurrences). Other activities like Yoga, Cycling, and Cardio are less common, which may indicate a strong focus on endurance and strength-based activities in this group. Age Group Distribution: Similar to the Balanced Exercisers, this cluster also predominantly consists of individuals from the 30-34 age group (333 occurrences) and 25-29 (307 occurrences). The presence of middle-aged individuals (50-54 and 45-49) suggests that endurance-focused individuals tend to be active throughout their adulthood, with a slight preference for individuals in their 30s.
Key Metrics: The average calories burned per session in this group is 524.00 kcal, which is significantly higher than the Balanced Exercisers group. The average heart rate is 131.74 bpm, and the average workout duration is much shorter at 38.82 minutes. Sleep hours are relatively lower (5.36 hours), and water intake remains consistent at 1.90 litres. This suggests that Endurance Specialists may focus on shorter, more intense workout sessions but with greater calorie expenditure.

3. Cluster: High-Intensity Over-Trainers
Workout Type Distribution: This cluster's most frequent workout types are Yoga (453 occurrences), followed by Cycling (442 occurrences), and HIIT (419 occurrences). Running and Strength Training appear less frequently, indicating a high preference for low-impact exercises like Yoga and Cycling that allow for more consistent participation. Age Group Distribution: The age distribution here includes individuals from the 20-24 (311 occurrences) and 25-29 (311 occurrences) age groups, highlighting a younger demographic for high-intensity training. There is also a notable presence of individuals in their 50s (300 occurrences), though there is a complete lack of participation from individuals aged 60 and above. Key Metrics: Average calories burned in this group is 587.36 kcal, which is slightly higher than the Endurance Specialists but not as high as the Sedentary Sleep-Deprived group. The average heart rate is 129.40 bpm, and the average workout duration is 38.34 minutes, indicating a preference for higher intensity in a shorter time span. Sleep hours are higher than the previous two clusters, with an average of 8.58 hours, which may contribute to better recovery.

4. Cluster: Sedentary Sleep-Deprived
Workout Type Distribution: Participants in this cluster engage in a wide variety of workouts, with Cardio (429 occurrences), Strength Training (422 occurrences), and HIIT (420 occurrences) being the most frequent types. This suggests a more

balanced approach to exercise, though the frequency of intense exercises like HIIT suggests participants may struggle with consistency.
Age Group Distribution: This cluster is most prevalent among individuals in the 25-29 age group (308 occurrences) and 50-54 (308 occurrences). Other significant age groups include 45-49 (305 occurrences) and 20-24 (305 occurrences). The complete lack of individuals from the 60-64 age group and older suggests that sedentary and sleep-deprived individuals may not extend into the senior demographic. Key Metrics: The average calories burned in this cluster is 794.42 kcal, the highest among all clusters, indicating that individuals in this group engage in longer and more intense workout sessions. Despite this, the average heart rate is slightly lower (127.83 bpm), and the average workout duration is 90.17 minutes, reflecting the more extended sessions. Sleep duration is also notably low at 6.77 hours, which could potentially contribute to feelings of fatigue or inconsistent workout results.

III. CONCLUSION

The clusters reveal distinct workout patterns, with Balanced Exercisers focusing on a mix of workout types and moderate intensity, Endurance Specialists engaging in high-calorie-burning exercises, High-Intensity Over-Trainers prioritizing shorter but intense sessions, and Sedentary Sleep-Deprived individuals showing high calorie expenditure but lower workout consistency and sleep deprivation. These findings suggest that tailored workout recommendations based on age, workout preferences, and sleep patterns could further optimize fitness programs.

IV. REFERENCE

- Buchheit, M., & Laursen, P. B. (2013). High-intensity interval training, solutions to the programming puzzle: Part I: Cardiopulmonary emphasis. *Sports Medicine, 43*(5), 313–338. https://doi.org/10.1007/s40279-013-0029-x
- Garber, C. E., Blissmer, B., Deschenes, M. R., Franklin, B. A., Lamonte, M. J., Lee, I. M., ... & Swain, D. P. (2011). Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: Guidance for prescribing exercise. *Medicine & Science in Sports & Exercise, 43*(7), 1334–1359. https://doi.org/10.1249/MSS.0b013e318213fefb
- Gibala, M. J., Little, J. P., Macdonald, M. J., & Hawley, J. A. (2012). Physiological adaptations to low-volume, high-intensity interval training in health and disease. *The Journal of Physiology, 590*(5), 1077–1084. https://doi.org/10.1113/jphysiol.2011.224725
- Physical Activity Guidelines Advisory Committee. (2018). *2018 Physical Activity Guidelines Advisory Committee scientific report*. U.S. Department of Health and Human Services. https://health.gov/sites/default/files/2019-09/PAG_Advisory_Committee_Report.pdf
- Swain, D. P., & Franklin, B. A. (2006). Comparison of cardioprotective benefits of vigorous versus moderate intensity aerobic exercise. *The American Journal of Cardiology, 97*(1), 141–147. https://doi.org/10.1016/j.amjcard.2005.07.130