DEPARTMENT OF COMPUTER APPLICATIONS

National Institute of Technology Kurukshetra, Haryana, India

MCA-206 Data Analytics

Project-Report Semester Long Assignment (2025-2026)

FitInsights Data Analysis

Under the guidance of

Dr. Kapil Gupta

(Assistant Professor, Department of Computer Applications)

Submitted by: Group 3

Submitted to: Dr. Kapil Gupta

Shubham Karmveer (523110049)

Aniket Chaurasia (523410021)

ABSTRACT

This report analyzes fitness tracker data from over 1,600 users to uncover patterns in calorie burn, hydration, and cardiovascular performance. Statistical tests show that **experience level** and **workout type** significantly impact **calories burned**, while **fat percentage** is negatively correlated with **hydration per kg**. **Gender differences** in calorie burn were significant, and users' hydration levels differed from the ideal benchmark. Additionally, **advanced users demonstrated higher Heart Rate Reserve (HRR)**, supporting its use as a progress indicator. No link was found between **BMI category and workout type**, suggesting potential for improved workout personalization. These insights can guide app features like tailored workouts, hydration alerts, and HRR-based recommendations to enhance user results and engagement.

INTRODUCTION

In today’s health-conscious world, fitness trackers play a vital role in helping individuals monitor their physical activity, health, and well-being. With the rise in this domain, users now seek personalized insights to improve their fitness outcomes and lifestyle habits. This report analyzes synthetic fitness tracker data, focusing on metrics like calories burned, heart rate patterns, hydration, and workout efficiency. By examining trends across user demographics and activity types, the goal is to identify meaningful patterns and offer data-driven recommendations. These insights aim to help the Health Tech Companies to enhance their fitness app, empowering users to achieve better health results through smarter, personalized guidance and insights.

MOTIVATION

With growing reliance on fitness trackers, users expect more than just step counts—they seek meaningful insights that guide healthier decisions. This analysis aims to empower the Health Tech Companies to improve its fitness app by understanding user behavior, workout efficiency, hydration habits, and health patterns. By uncovering what drives results for different users, the project supports smarter recommendations that boost engagement, well-being, and personalized fitness outcomes.

OBJECTIVES

Understanding user fitness behaviors and physiological metrics is essential to designing smarter, more personalized wellness solutions by health tech companies. This analysis helps uncover key factors influencing workout efficiency and overall health trends—ultimately guiding better fitness planning and app improvements.

Primary Objective:

- To analyze fitness tracker data and identify patterns that influence calories burned, workout efficiency, and BMI across diverse user profiles.

Secondary Objectives:

- To evaluate how variables like workout type, fat percentage, and experience level impact fitness outcomes.
- To compare performance and health metrics across workout types, BMI categories, and experience levels using descriptive and inferential statistics.

ABOUT DATASET

The fitness tracker dataset captures detailed user metrics from workout sessions, including age, gender, heart rate, calories burned, fat percentage, hydration, workout type, and duration. It also includes engineered features like Heart Rate Reserve (HRR), Calories per Minute, and BMI Category. This data enables analysis of user performance, health trends, and behavior patterns—supporting personalized fitness, hydration, and workout recommendations..

Data Overview

- **Total Records:** 1629
- **Total Features (Columns):** 20 (after feature engineering)
- **Types of Variables:**
    - **Categorical:** Gender, Experience Level, BMI Category
    - **Numerical (Continuous):** Age, Weight, Duration (min), Calories Burned, Hydration per Kg, Heart Rate Reserve (HRR), Workout Intensity, Calories per Min
    - **Derived Metrics:** Workout Intensity, Calories per Min, HRR, BMI Category, Hydration per Kg

S.No	Feature Description	Variable Name	Data Type	Type of Value	Range of Values
1	Age of the user	Age	Numerical	Continuous	18–66
2	Gender of the user	Gender	Categorical	Nominal	Male, Female, Non-binary
3	Weight of the user (in kilograms)	Weight_kg	Numerical	Continuous	45–113 kg
4	Height of the user (in meters)	Height_m	Numerical	Continuous	1.5–1.99 m
5	Maximum Heart Rate during session	Max_BPM	Numerical	Continuous	160–210 bpm
6	Average Heart Rate during session	Avg_BPM	Numerical	Continuous	100–170 bpm
7	Resting Heart Rate	Resting_BPM	Numerical	Continuous	54–88 bpm
8	Duration of the workout session (in hours)	Session_Duration_hours	Numerical	Continuous	0.5–1.170 hours
9	Calories burned during workout	Calories_Burned	Numerical	Continuous	200-806 kcal
10	Type of workout performed	Workout_Type	Categorical	Nominal	Cardio, Strength, HIIT, Mixed
11	Fat percentage of the user	Fat_Percentage	Numerical	Continuous	5.2%–40%

12	Water intake during or post workout (liters)	Water_Intake	Numerical	Continuous	0.34–3.4 L
13	Frequency of workouts per week	Workout_Frequency	Numerical	Discrete	2–6 days/week
14	Fitness experience level	Experience_Level	Categorical	Ordinal	Beginner, Intermediate, Advanced,Professsional
15	Body Mass Index (BMI)	BMI	Numerical	Continuous	18–35 kg/m²
16	Calories burned per minute	Calories_per_Min	Derived	Continuous	Derived from calories/session duration
17	Heart Rate Reserve (HRR)	HRR	Derived	Continuous	MHR-RHR , 73-155
18	BMI Category	Bmi_Category	Categorical	Ordinal	Underweight, Normal, Overweight, Obese
19	Hydration per kg of body weight	Hydration_per_Kg	Derived	Continuous	water_intake / weight(0.007-.044)
20	Workout Intensity	Workout_Intensity	Derived	Continuous	AHR/MHR, (0.532-0.88)

DATA PREPROCESSING

1. Shape

- **Original Dataset**: 1680 rows × 15 columns
- **Processed Dataset**: 1629 rows × 20 columns
- Focused only on useful features and engineered new ones for better analysis.

2. Renamed and Standardized Columns

- Ensured consistency in column naming:
- Removed unwanted characters like newline symbols or extra spaces

3. Handled Missing Values

- Checked for missing values using df.isnull().sum():
- Numerical categories filled with median of the data
- Categorical columns filled with mode of the data

4. Ensured Proper Data Types

- Ensured numerical columns were treated correctly:
- Ensured thsese Avg_BPM, Max_BPM, Calories_Burned, and Session_Duration_hours to numeric
- Prevented calculation errors and ensured model readiness

5. Filtered and Retained Relevant Variables

- Demographics: Age, Gender, Height_m, Weight_kg, BMI, BMI_Category
- Health metrics: Fat_percentage, Avg_BPM, Max_BPM, Resting_BPM, Water_intake_liters, Hydration_per_kg
- Workout stats: Workout_type, Session_duration_hours, Calories_burned, Workout_frequency, Experience_level
- Engineered features: Workout_Intensity, Calories_per_Min, HRR

6. Outlier Detection and Handling

- Visualized key numeric features to detect outliers using boxplots and summary statistics.
- Found a few extreme values in metrics like Calories_Burned, BMI, and Heart Rate.
- Ensured the cleaned dataset preserved variability while removing noise.

FEATURE ENGINEERING

1. Workout Intensity

- Created a new metric to assess how intense a workout session was.
- Formula combined heart rate data (Avg_BPM / Max_BPM).

2. Calories per Minute

- Calculated the rate at which calories were burned per minute of session duration.
- Useful for comparing efficiency across different workout types or durations.

3. Heart Rate Reserve (HRR)

- Estimated how much effort a person exerted during workouts.
- Computed using Max_BPM – Resting_BPM.

4. Hydration per Kg

- Derived how much water was consumed per kilogram of body weight.
- Helped assess if hydration levels were appropriate relative to body size.

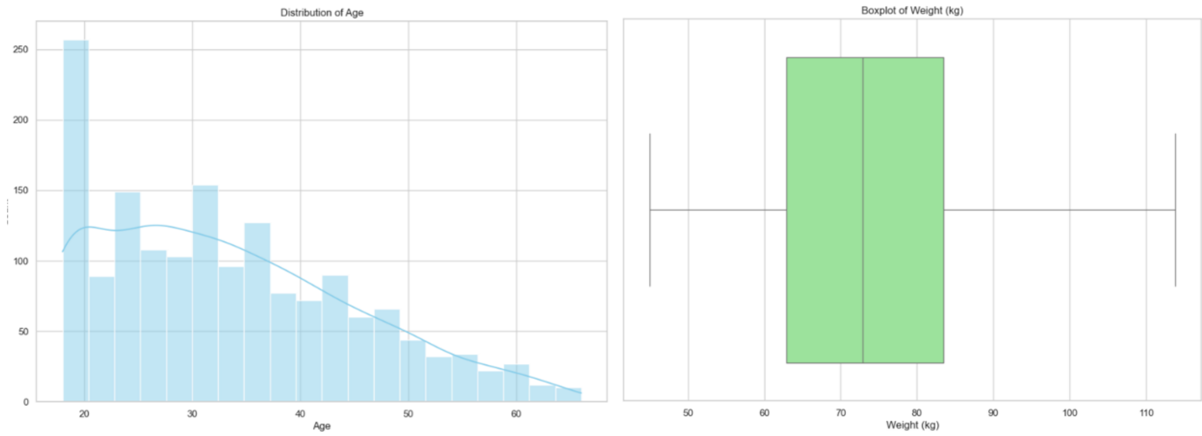
5. BMI Category

- Classified individuals based on BMI into categories like:

- Underweight, Normal, Overweight, Obese

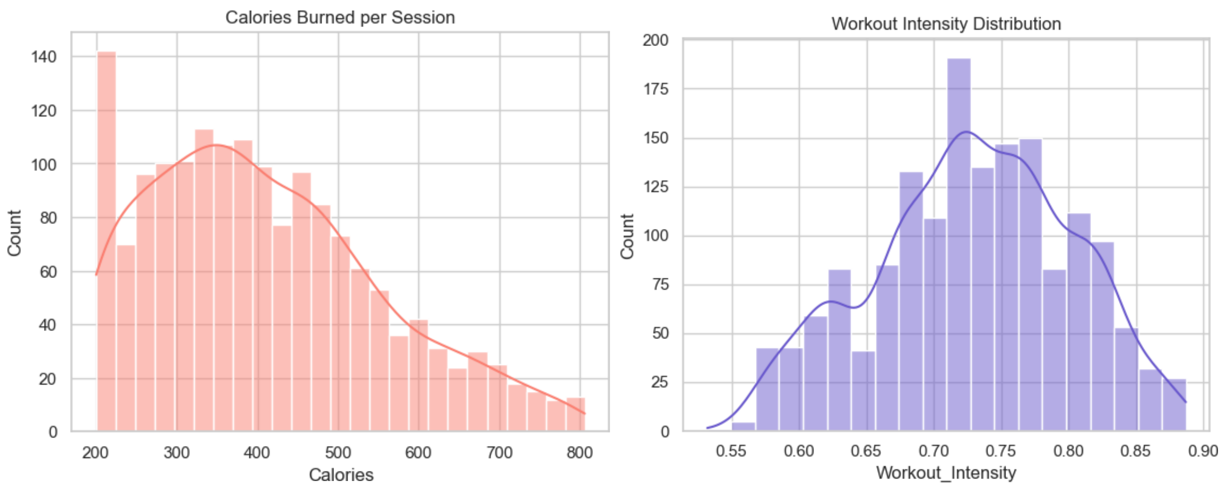
BASIC DATA VISUALIZATION

1. Age and Weight



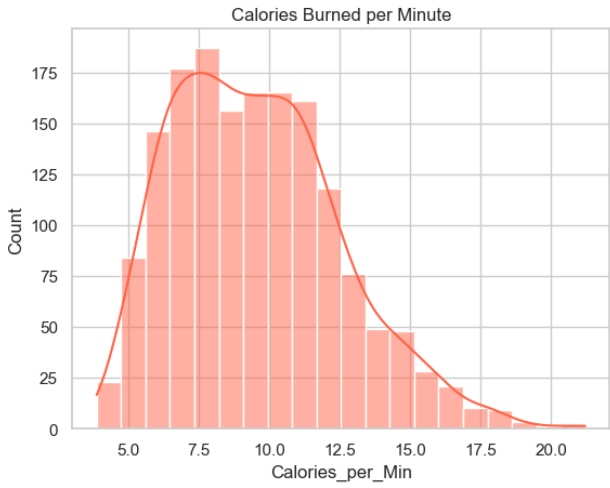
- **Age Distribution (Left Plot):**
  - Most users are in the **18–25 age group**, with a sharp decline as age increases.
  - The distribution is **right-skewed**, indicating younger users dominate the dataset.
- **Weight Boxplot (Right Plot):**
  - Median weight is around **75 kg**.
  - Most users fall between **60–85 kg**.
  - A few outliers exist on both lower and higher ends (below 50 kg and above 110 kg).

2. Calories_Burned and Workout_Intensity



- **Calories Burned(Left Plot)**: Most sessions burn 300–500 calories; fewer sessions go above 600. Distribution is right-skewed.
- **Workout Intensity(Right Plot)**: Most sessions have intensity around 0.70–0.75. Follows a near-normal distribution.

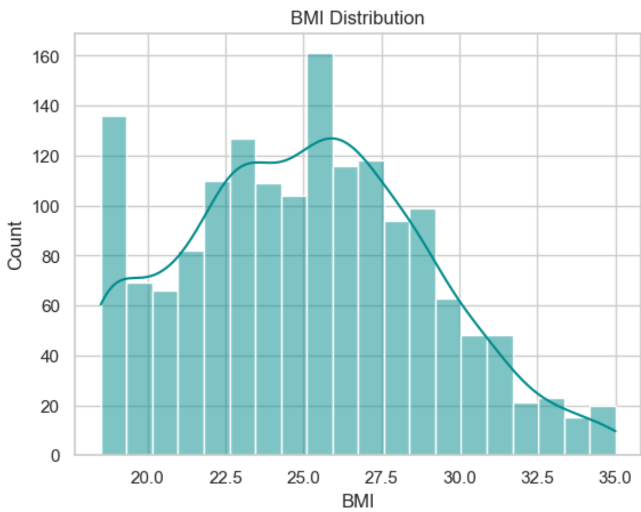
3. Calories_per_Min and HRR (Heart Rate Reserve)

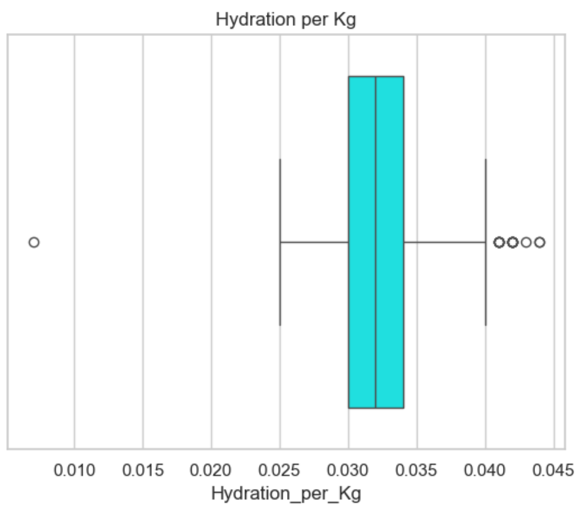


- **Calories Burned per Minute(Left Plot):**
    - Most sessions burn **7–11 calories per minute**, peaking around 8–9.
    - Few sessions exceed 15 calories/min — high-efficiency workouts are less common.

- **Heart Rate Recovery (HRR)(Right Plot)**:
    - Follows a normal distribution centered around **115–120 BPM**.
    - Most users have moderate cardiovascular fitness.

4. Bmi and Hydration_per_kg

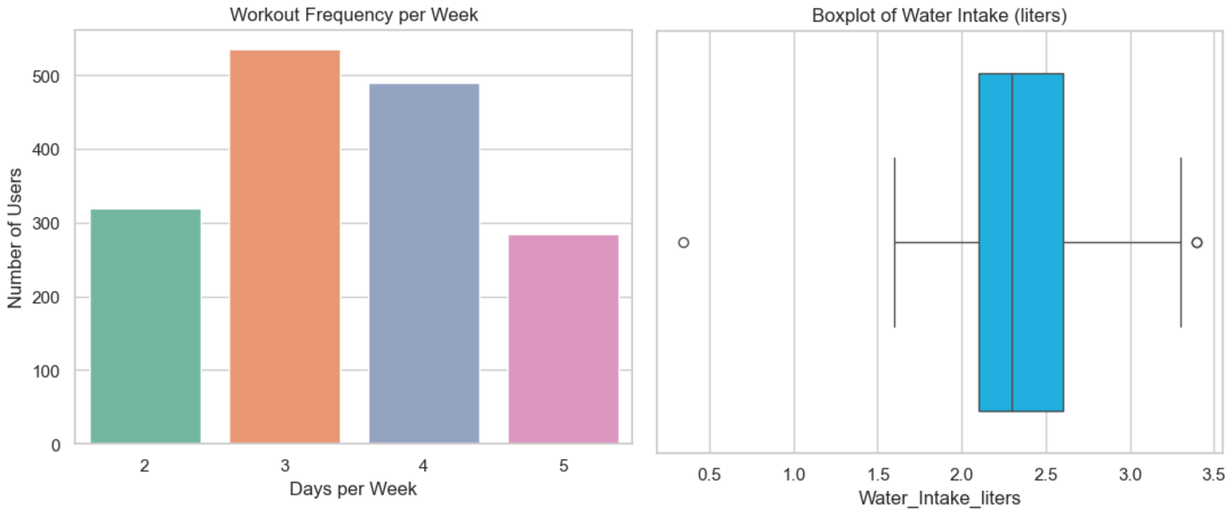




- **BMI Distribution (Left Plot):**
    - Most users have a BMI between 22 and 28, with a peak around 25–26.
    - The distribution is slightly right-skewed, indicating a majority are in the normal to slightly overweight range.

- **Hydration per Kg (Right Plot):**
    - Median hydration level is approximately 0.031 L/kg.
    - Most users range between 0.026 and 0.036 L/kg.

5. Workout_frequency and Water_intake

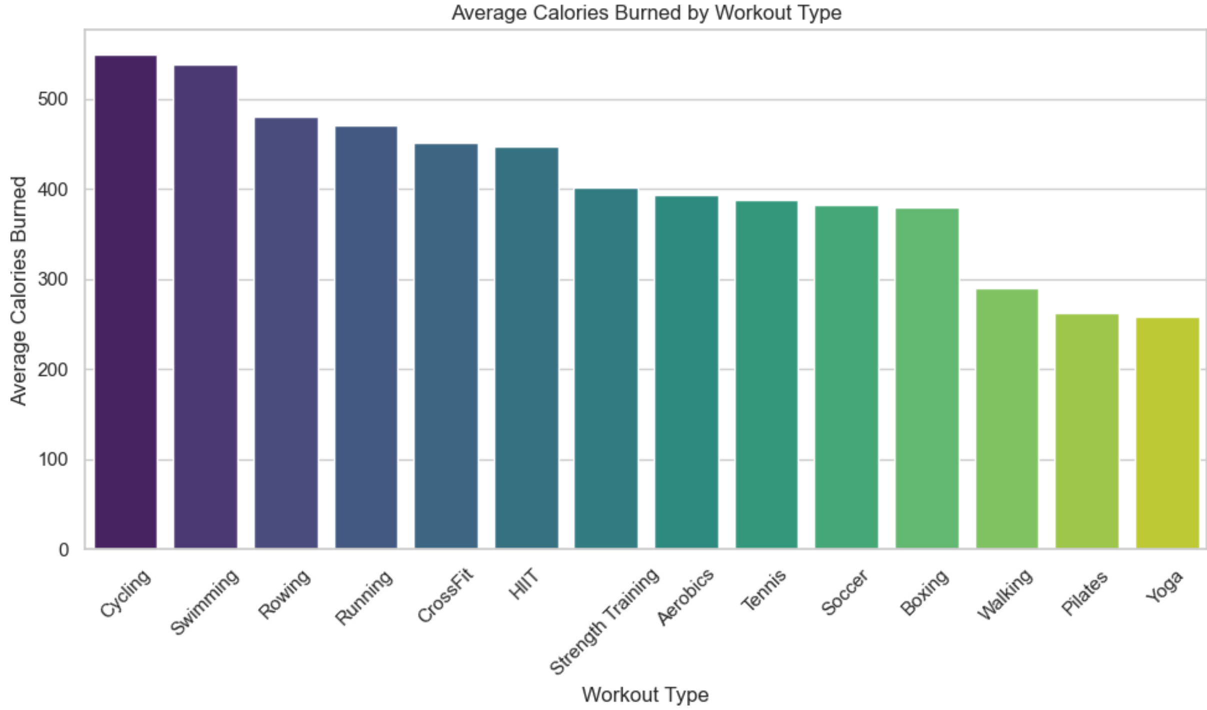


- **Workout_frequency (Left Plot):**
  - Most users have been working out on 3 or 4 days a week some of them either workout 2 days or workout 5 days a week.
- **Water_intake_liters (Right Plot):**
  - Most users water intake is between 2.1 to 2.6 liters of water in a day.

EXPLORATORY DATA ANALYSIS

1. Univariate Analysis

Average Calories Burned By Workout Type



Purpose	Column Name	Type	Values	Description
Grouping category	Workout_Type	Categorical	Cycling, Swimming, Rowing, etc.	Represents different types of workouts

Numeric response (y-axis)	Calories_Burned	Numeric	Average calories burned per workout	Mean number of calories burned for each workout type

Key Observations:

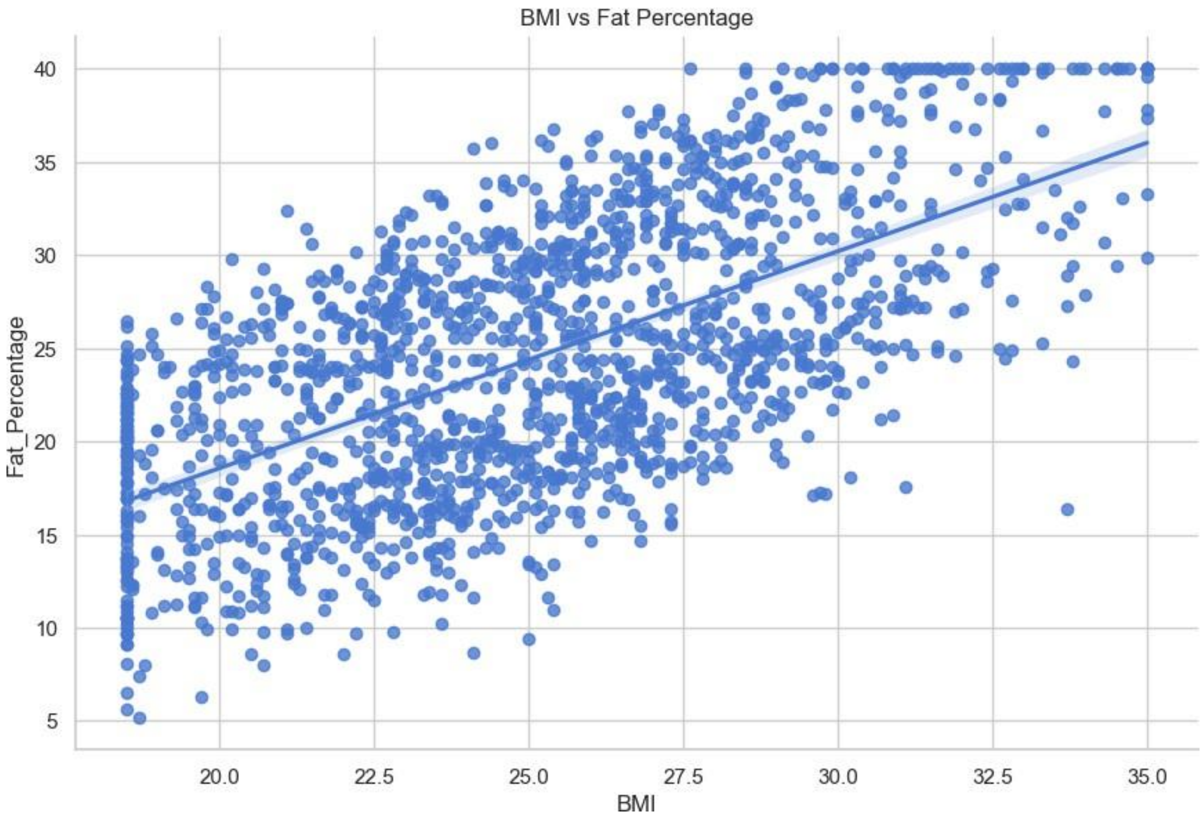
- Cycling and Swimming top the chart, burning the highest average calories—over 500 calories per session.
- Rowing, Running, and CrossFit follow closely behind, indicating they are also highly effective in terms of calorie expenditure.
- Mid-range workouts like HIIT, Strength Training, and Aerobics still maintain solid calorie-burning potential, averaging between 390 and 450 calories.
- Activities like Walking, Pilates, and Yoga show the lowest average calorie burn, all below 300 calories.
- The gradual decline in bars indicates a clear stratification of workouts based on intensity and energy output.

Interpretation:

This visualization highlights that cardio-intensive and high-effort workouts (e.g., Cycling, Swimming) tend to burn significantly more calories than lower-intensity or restorative activities (e.g., Yoga, Pilates). It can help users choose workouts aligned with their calorie-burning goals and energy levels. The insight is useful for fitness planning and suggests a strong correlation between workout type and energy expenditure**.**

2. Bivariate Analysis

2.1. Bmi vs Fat Percentage



Purpose	Column Name	Type	Values	Description
Numeric predictor (x-axis)	BMI	Continuous	~18 to 35	Body Mass Index, a measure of body weight relative to height
Numeric response (y-axis)	Fat_Percentage	Continuous	~5% to 40%	Estimated percentage of body fat in an individual

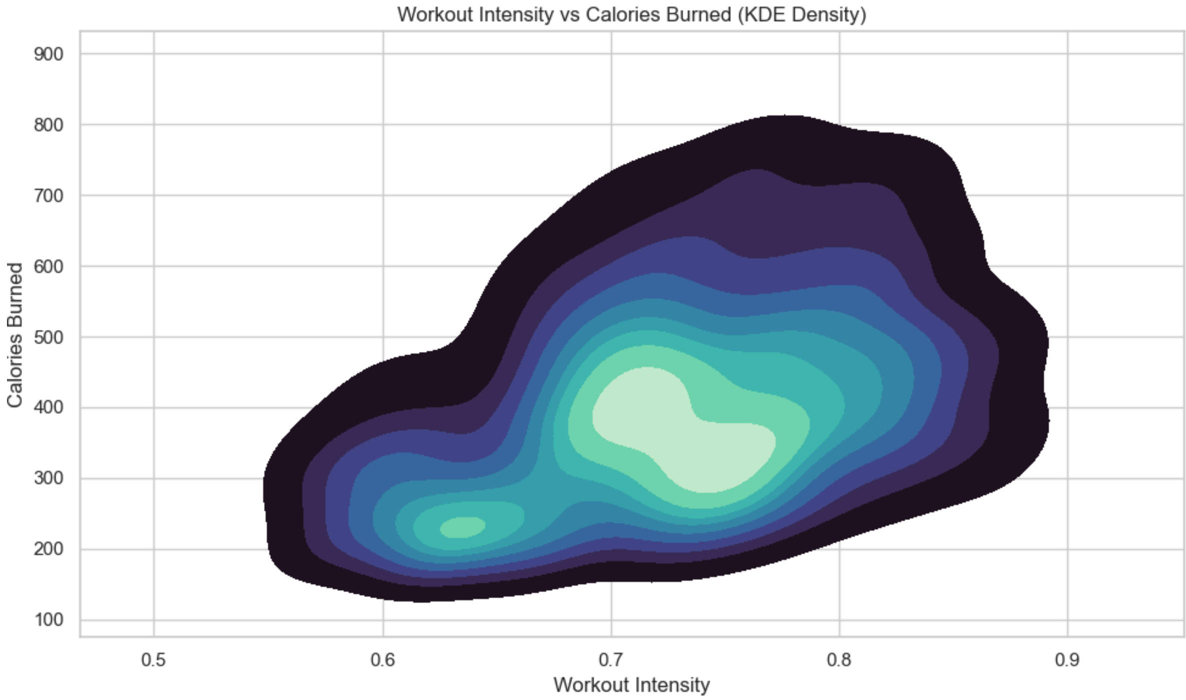
Key Observations:

- A **positive linear trend** is observed: as **BMI increases**, so does **Fat Percentage**.
- The scatter points, though dispersed, generally align around the regression line, indicating a **moderate to strong correlation** between the two variables.
- There's a **notable concentration of data points** between BMI values of **22 to 30**, where fat percentage spans a broader range (~10% to ~35%).
- **Outliers** are present at both low and high ends of BMI and Fat Percentage, which may reflect data anomalies or individual variation.
- The regression line suggests that **BMI is a reasonable but not perfect predictor** of fat percentage, as variability still exists.

Interpretation:

This plot reveals a clear relationship between BMI and fat percentage, supporting the idea that BMI can serve as a proxy for estimating body fat.

2.2. Workout intensity vs Calorie Burned(KDE Density)



Column Name	Type	Values/Range	Description
Workout Intensity	Numeric	0.5 – 0.9	Relative measure of exercise effort
Calories Burned	Numeric	100 – 900	Estimated calories expended per workout

Key Observations

- The highest density region (lightest color) is centered around a workout intensity of approximately 0.7 and calories burned in the 350–500 range.
- As workout intensity increases beyond 0.7, there is a noticeable upward trend in calories burned, with the densest clusters extending toward higher calorie values.
- There is a moderate spread, indicating variability in calories burned even at similar workout intensities, but the overall pattern suggests a positive association.
- Few data points exist at very low intensities (<0.6) or extremely high intensities (>0.9), and these regions correspond to lower densities and fewer calories burned.

Interpretation

This KDE plot suggests a clear positive relationship between workout intensity and calories burned: as individuals increase their workout intensity, they are more likely to burn a greater number of calories. The densest region indicates that most workouts occur at moderate-to-high intensities, yielding calorie expenditures in the 350–500 range. While higher intensities can lead to even greater calorie burn, there is also more variability, possibly due to differences in workout duration, individual fitness levels, or exercise types. The data imply that for most people, increasing workout intensity is an effective strategy for burning more calories, but individual results may vary.

2.3. Fat Percentage by Age and Gender

Column Name	Type	Values/Range	Description
Age	Numeric	18-65 years	Age of individuals in the dataset
Fat Percentage	Numeric	5-40%	Body fat percentage measured for each individual
Gender	Categorical	Male, Female, Non-binary	Gender identity of individuals

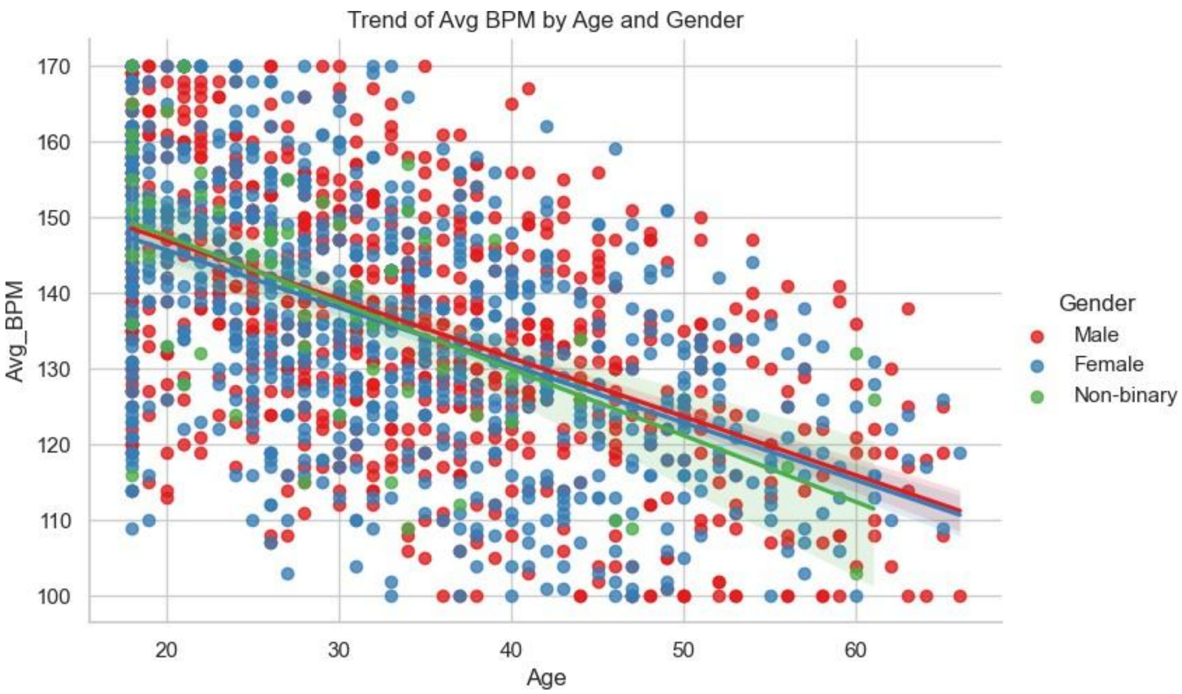
Key Observations

- Females consistently show higher fat percentages (mostly between 25-40%) compared to males (predominantly between 10-25%) across all age groups.
- Non-binary individuals appear to have fat percentages that generally fall between the male and female distributions, though with fewer data points overall.
- There is no strong visible correlation between age and fat percentage within any gender group, suggesting that age alone is not a strong predictor of body fat percentage.
- The data spans ages from approximately 18 to 65 years, with good representation across this entire range.
- The highest concentration of data points for males is around 15-20% fat, while for females it's around 30-35% fat.
- There is considerable individual variation within each gender group, with some overlap between the distributions.

Interpretation

This visualization illustrates the significant gender-based differences in body fat composition across the adult lifespan. The clear separation between male and female distributions reflects well-established physiological differences, with females naturally maintaining higher essential fat percentages for reproductive and hormonal functions. The absence of a strong age-related trend suggests that other factors—such as diet, exercise habits, genetics, and lifestyle—likely play more important roles in determining body fat percentage than age alone. The wide spread within each gender group highlights the substantial individual variation that exists. For health practitioners and researchers, this data reinforces the importance of using gender-specific reference ranges when assessing body composition and suggests that age-adjusted standards may be less critical than other personalized factors.

2.4. Trend of Avg Bpm by Age and Gender



Column Name	Type	Values/Range	Description
Age	Numeric	15 – 65 years	Age of individuals in the dataset
Avg BPM	Numeric	100 – 170 beats/min	Average heart rate measured in beats per minute
Gender	Categorical	Male, Female, Non-binary	Gender identity of individuals

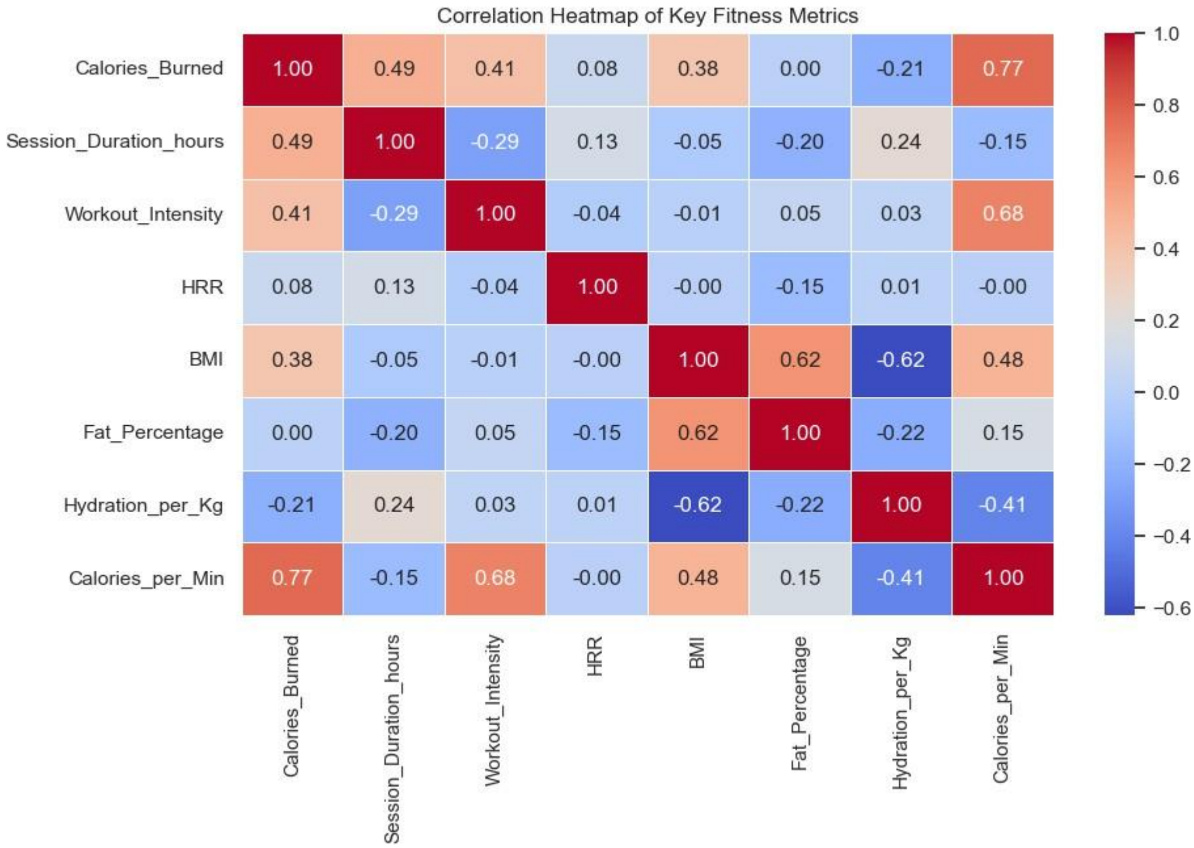
Key Observations

- The scatter plot shows the trend of average heart rate (Avg BPM) by age, separated by gender categories: male (red), female (blue), and non-binary (green).
- There is a clear negative correlation between age and average BPM across all genders; as age increases, average BPM tends to decrease.
- The regression lines for all three gender groups closely overlap, indicating similar trends in average BPM decline with age regardless of gender.
- Younger individuals (around 15-25 years) have higher average BPM values, often between 140 and 170 BPM.
- Older individuals (above 50 years) tend to have lower average BPM values, mostly ranging from 110 to 130 BPM.
- There is considerable variability in average BPM within each age group and gender, shown by the scattered data points.
- The non-binary group has fewer data points but follows the same overall trend as male and female groups.

Interpretation

This plot demonstrates a consistent decline in average heart rate with increasing age across all gender identities. The similarity in trend lines suggests that the physiological effect of aging on heart rate is largely consistent regardless of gender. The higher average BPM in younger individuals reflects the natural cardiovascular response to age, while the decrease with age aligns with known declines in maximum heart rate and cardiovascular capacity. Despite variability within groups, the overall pattern supports the use of age-adjusted heart rate expectations in health assessments. The inclusion of non-binary individuals, though limited in number, shows comparable trends, underscoring the importance of inclusive data analysis.

2.5. Correlation Heatmap of Key Fitness Metrics



Variable Pair	Correlation (r)	Interpretation
Calories_Burned vs Calories_per_Min	0.77	Strong positive correlation – higher total calorie burn is closely tied to a higher burn rate per minute.
Calories_per_Min vs Workout_Intensity	0.68	Strong positive – more intense workouts lead to a higher calorie burn rate.
BMI vs Fat_Percentage	0.62	Strong positive – individuals with higher BMI typically have higher body fat percentage.
Calories_Burned vs Session_Duration	0.49	Moderate positive – longer sessions are associated with greater calorie expenditure.
BMI vs Calories_per_Min	0.48	Moderate positive – higher BMI is linked to a higher calorie burn rate per minute.
Calories_Burned vs Workout_Intensity	0.41	Moderate positive – more intense workouts tend to burn more calories overall.
BMI vs Hydration_per_Kg	-0.62	Strong negative – higher BMI is associated with lower hydration per kilogram.
Calories_per_Min vs Hydration_per_Kg	-0.41	Moderate negative – higher calorie burn rate per minute is linked to lower hydration per kilogram.
Session_Duration vs Workout_Intensity	-0.29	Weak negative – longer sessions are generally performed at lower intensities.
HRR vs Any Other Variable	~0.00 to 0.13	No correlation – heart rate recovery is largely independent of other fitness metrics.
Fat_Percentage vs Others (except BMI)	~0.00 to 0.20	Very weak – fat percentage is weakly related to most other metrics except BMI.

Key Observations

Strong Positive Correlations:

- Calories_Burned and Calories_per_Min show a very strong positive correlation (0.77), indicating that higher calorie burn is closely tied to a higher rate of calories burned per minute.
- Workout_Intensity also correlates strongly with Calories_per_Min (0.68), suggesting that more intense workouts lead to a higher calorie burn rate.
- BMI and Fat_Percentage have a strong positive relationship (0.62), reflecting that individuals with higher BMI typically have higher body fat percentages.

Moderate Positive Correlations:

- Calories_Burned is moderately correlated with Session_Duration_hours (0.49), Workout_Intensity (0.41), and BMI (0.38), indicating that longer sessions, higher intensities, and higher BMI are all associated with greater calorie expenditure.
- BMI and Calories_per_Min (0.48) also show a moderate positive link.

Negative Correlations:

- BMI and Hydration_per_Kg are neg atively correlated (-0.62), suggesting that individuals with higher BMI tend to have lower hydration per kilogram.
- Calories_per_Min and Hydration_per_Kg also show a moderate negative relationship (-0.41).
- Session_Duration_hours and Workout_Intensity have a negative correlation (-0.29), implying that longer sessions might be performed at lower intensities.

Weak or Negligible Correlations:

- HRR (Heart Rate Recovery) shows very weak correlations with all other variables, indicating little linear relationship.
- Fat_Percentage is weakly related to most other metrics except BMI.

Interpretation

The heatmap reveals several key patterns in fitness metrics. Calorie expenditure is most strongly influenced by both the rate of calories burned per minute and workout intensity, emphasizing the importance of exercise intensity and efficiency over simply session duration. BMI and fat percentage are closely linked, as expected, but higher BMI is associated with lower hydration per kilogram, which may point to differences in body composition or hydration habits among individuals with higher BMI. The negative relationship between session duration and workout intensity suggests a trade-off: longer workouts tend to be less intense, possibly due to endurance limitations. Heart Rate Recovery appears largely independent of the other metrics, indicating it may capture a distinct aspect of fitness. Overall, the heatmap highlights the interconnectedness of calorie burn, intensity, and body composition, while also pointing out areas—like HRR—where relationships are minimal.

INFERENTIAL STATISTICS

Hypothesis 1: " Experience Level impacts Calories Burned.

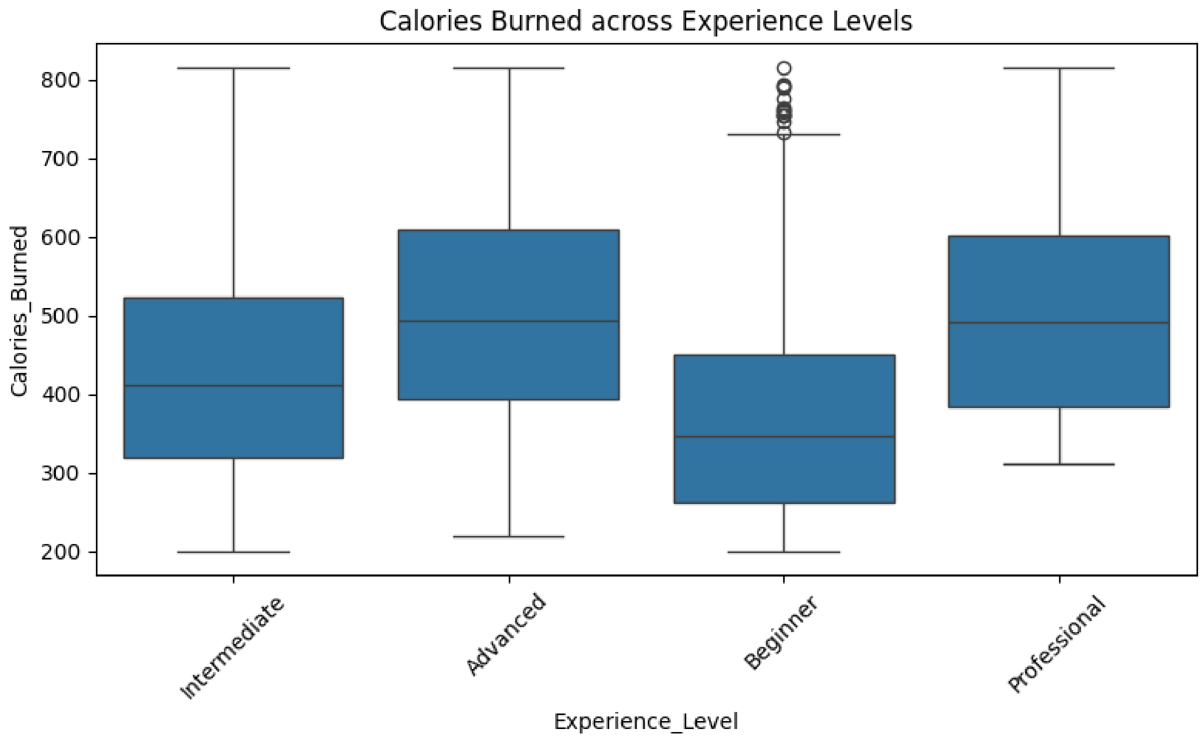
Null Hypothesis (H_0): Mean calories burned is the same across all experience levels.

Alternate Hypothesis (H_1): At least one experience level has a different mean calorie burn.

Variables: Experience_Level (categorical), Calories_Burned (numerical)

Test Used: ANOVA (Analysis of Variance)

Reason: To compare the means of a continuous variable across 3+ independent categories.



Analysis of Hypothesis Testing Results: Experience Level vs. Calories Burned

- Based on the boxplot and ANOVA test results shown in the image, we can draw several important conclusions about Hypothesis 1: "Experienced individuals tend to burn more calories per session than beginners."

Statistical Findings

The ANOVA test yielded highly significant results:

- F-statistic: 103.63
- p-value: 4.84e-43 (extremely small, far below the conventional 0.05 threshold)

This extraordinarily small p-value indicates that there are statistically significant differences in calories burned between at least some of the experience level groups.

Conclusion on Hypothesis

The hypothesis is supported by the data. Both visual evidence and statistical testing confirm that experienced individuals (Advanced and Professional) do tend to burn significantly more calories per session than Beginners.

Hypothesis 2: "Higher workout intensity correlates with higher water intake."

Null Hypothesis (H_0): There is no linear correlation between workout intensity and water intake.

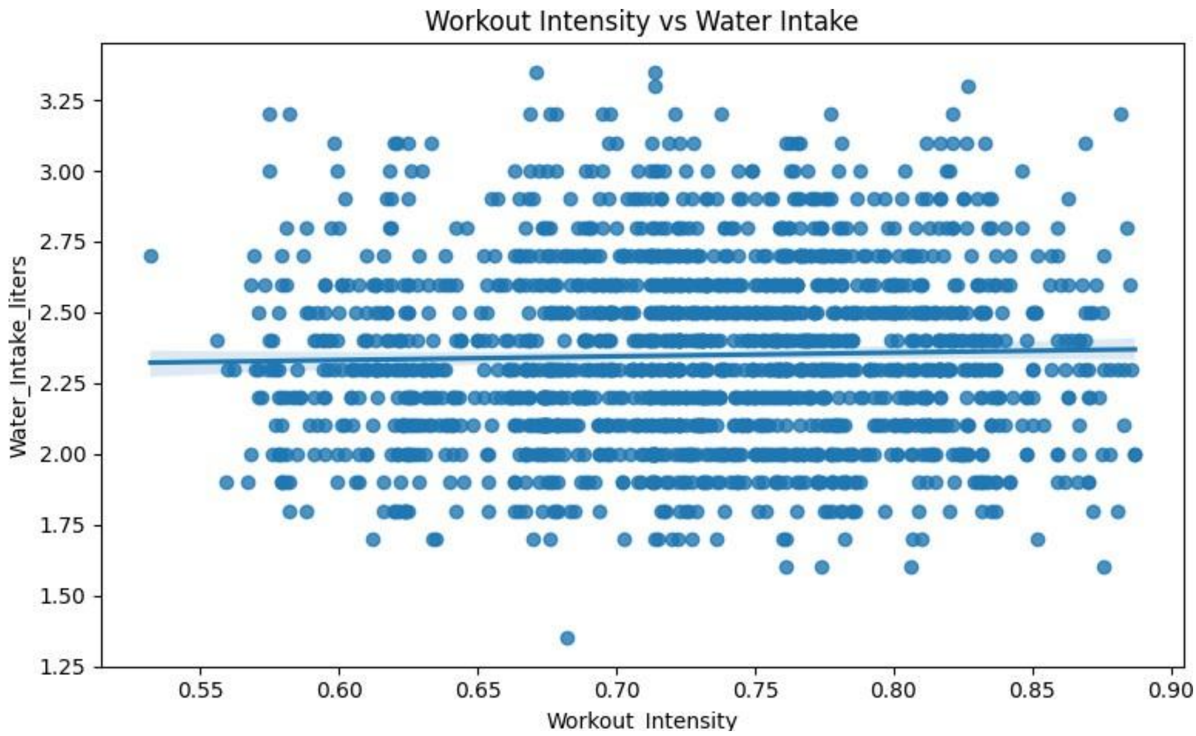
Alternate Hypothesis (H_1): There is a linear correlation between workout intensity and water intake.

Variables: Workout_Intensity(quantitative), Water_Intake_liters(quantitative)

Test Used: Pearson Correlation

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

Reason: To test linear relationship between two continuous variables.



Conclusion for Hypothesis 2: "Higher workout intensity correlates with higher water intake."

Statistical Result

- Pearson correlation coefficient: 0.029. p:value:0.2229
- This value indicates a very weak positive linear relationship between workout intensity and water intake.
- The correlation is close to zero, suggesting that, in this dataset, higher workout intensity does not strongly predict higher water intake.

Interpretation

- **Conclusion**: The Null hypothesis cannot be Rejected. There is only a negligible increase in water intake with higher workout intensity, and the relationship is not statistically meaningful in practical terms.

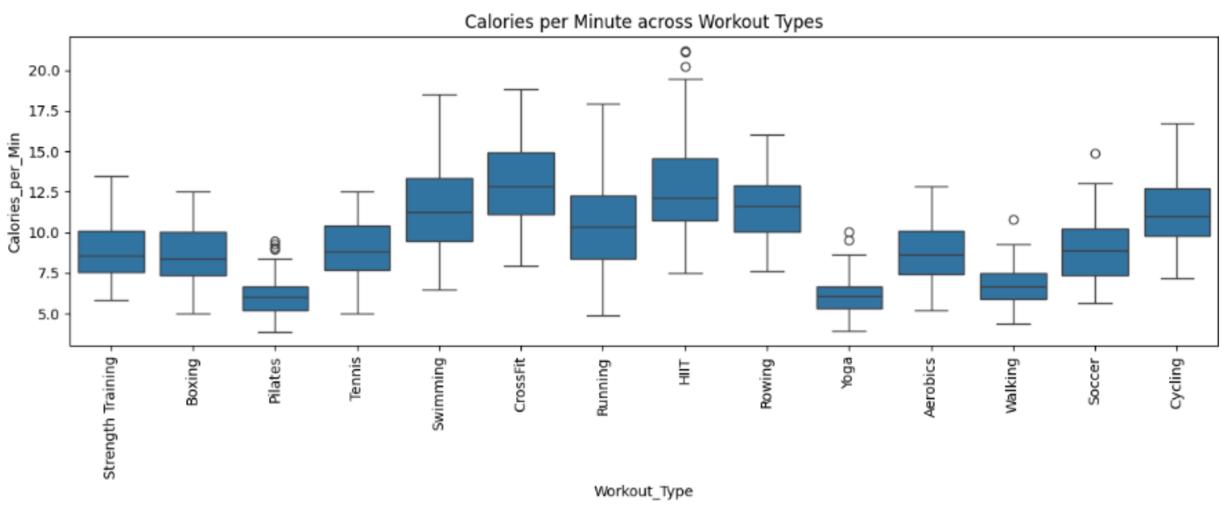
Hypothesis 3: "Calories burned per minute varies by workout type."

Null Hypothesis (H_0): Mean calories/minute is the same across all workout types.

Alternate Hypothesis (H_1): At least one workout type has a different mean calorie burn per minute.

Variables: Workout_Type (categorical), Calories_per_Min (numerical)

Test Used: ANOVA



Conclusion for Hypothesis 3: "Calories burned per minute varies by workout type."

Statistical Results

- ANOVA F-statistic: 165.71
- p-value: 3.41e-59

This extremely low p-value (far below 0.05) indicates that there are statistically significant differences in average calories burned per minute across different workout types.

Conclusion:

The hypothesis is strongly supported: Calories burned per minute does vary significantly by workout type.

There are clear, statistically significant differences in calories burned per minute across workout types. This insight enables the health tech platform to provide evidence-based, personalized workout recommendations according to users' calorie-burning goals and preferences.

Hypothesis 4: Fat Percentage is correlated with Hydration per Kg

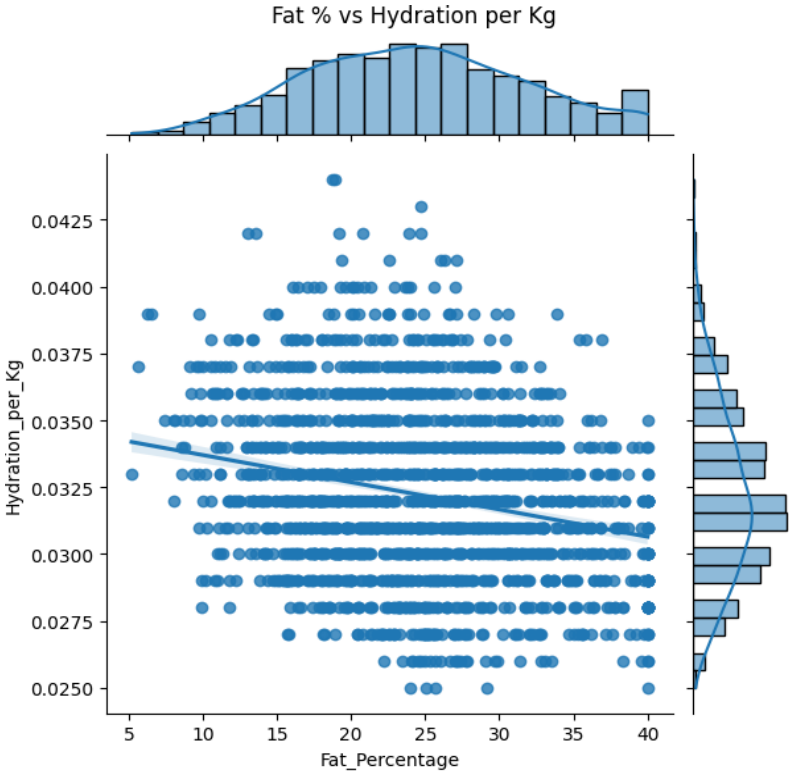
Null Hypothesis (H_0): No monotonic relationship between fat percentage and hydration per kg.

Alternate Hypothesis (H_1): There is a monotonic relationship between fat percentage and hydration per kg.

Variables: Fat_Percentage, Hydration_per_Kg (both numerical)

Test Used: Spearman Rank Correlation (non-parametric)

Reason: Fat percentage and hydration may not have a linear relationship.



Conclusion for Hypothesis 4: "Hydration per kg is lower in individuals with higher fat percentage."

Statistical Results

- Spearman Correlation Coefficient: -0.207
- p-value: 1.09e-17 (extremely significant)
- The negative correlation indicates that as fat percentage increases, hydration per kg decreases.
- The very small p-value (<0.001) confirms this relationship is not due to random chance.

Interpretation

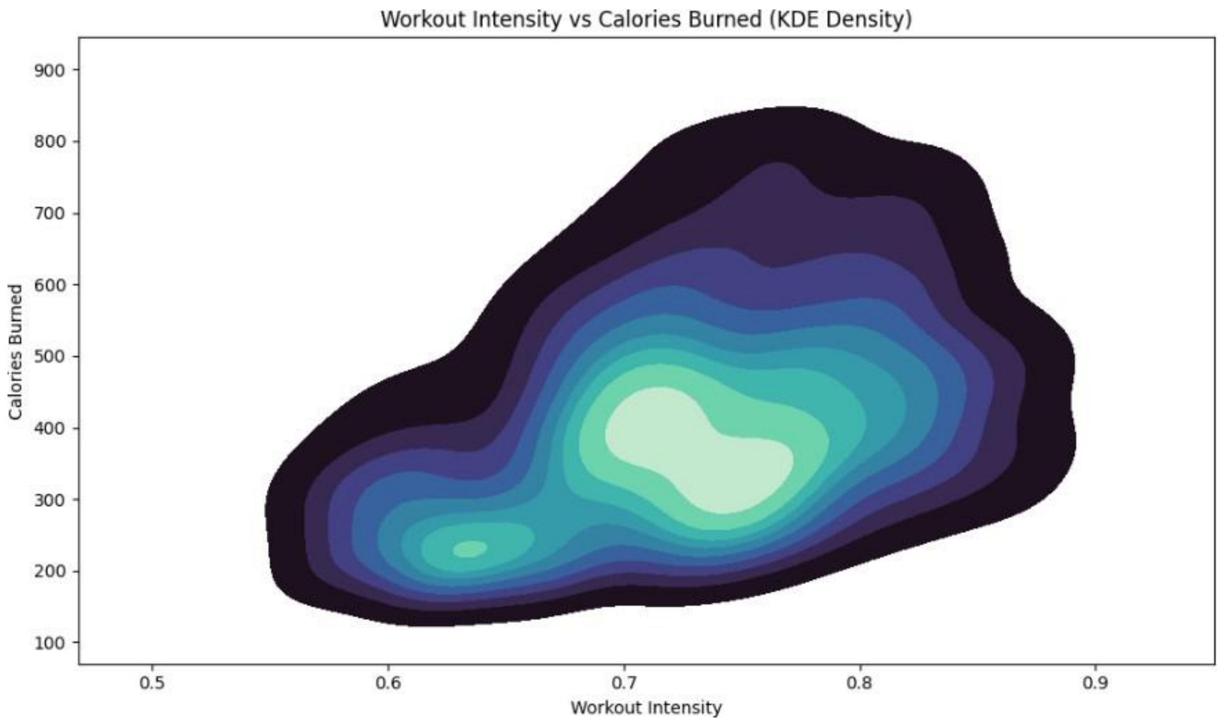
The hypothesis is strongly supported by the data. There is a statistically significant negative relationship between fat percentage and hydration per kg. As an individual's fat percentage increases, their hydration per kg decreases.

Conclusion: Significant negative monotonic correlation found.

Hypothesis 5: Workout Intensity Vs Calories Burned

Null Hypothesis (H_0): There is no significant relationship between workout intensity and calories burned.

Alternative Hypothesis (H_1): There is a significant positive relationship between workout intensity and calories burned — as workout intensity increases, the calories burned also tend to increase**.**



Implications of Null Hypothesis Testing in Workout Intensity and Calories Burned Correlation

Pearson correlation test examining the relationship between workout intensity and calories burned, yielding the following results:

- **Correlation coefficient**: 0.3979 (moderate positive correlation)
- **p-value**: 1.25 × 10^-64 (extremely small)

Understanding the Null Hypothesis Context

In this statistical test, the null hypothesis (H_0) states that there is no correlation between workout intensity and calories burned (correlation coefficient = 0). The alternative hypothesis (H_1) suggests that a correlation does exist (correlation coefficient ≠ 0).

Key Implications of These Results

1. Statistical Significance

The p-value (1.25 × 10^-64) is extraordinarily small, far below the conventional threshold of 0.05. This means we can confidently reject the null hypothesis. The probability of observing this correlation by random chance if no true relationship existed is virtually zero.

2. Practical Significance

The correlation coefficient of 0.3979 indicates a moderate positive relationship between workout intensity and calories burned. This means that as workout intensity increases, calories burned tend to increase as well, though not in perfect lockstep.

3. Fitness Application Implications

For a health tech company developing workout recommendations:

- This data supports implementing intensity-based calorie burn predictions
- Algorithms can reasonably use workout intensity as a predictor of calorie expenditure
- Users can be informed that higher intensity workouts generally lead to greater calorie burn

Conclusion

The statistical evidence strongly supports a moderate positive relationship between workout intensity and calories burned. The health tech company can confidently incorporate this relationship into their recommendation algorithms, while recognizing that intensity is just one of several factors affecting calorie expenditure during workouts.

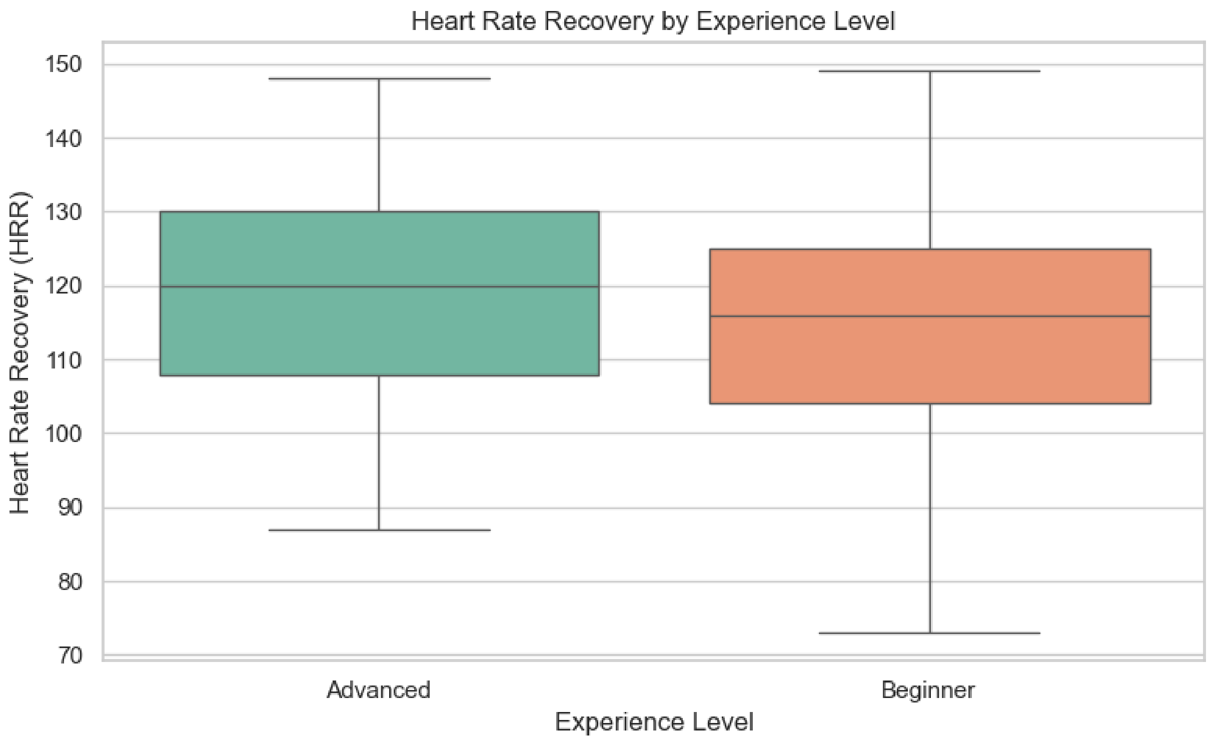
Hypothesis 6: HRR Vs Experience Level

Null Hypothesis (H_0): Mean HRR is the same for advanced and beginner users.

Alternate Hypothesis (H_1): Advanced users have significantly higher HRR than beginners.

Test Used: Independent Samples T-Test

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$




```
(4.288061234979615, 2.3390677872485295e-05)
```

Result:

t-statistic = 4.29

p-value = 2.34×10^{-5} (or 0.0000234)

Since the p-value is much smaller than the conventional alpha level (0.05), we reject the null hypothesis H_0.

Conclusion:
There is statistically significant evidence that advanced users have a higher heart rate recovery (HRR) compared to beginners.Stakeholder Implication: Use HRR to monitor progress and guide users to advanced plans.

OBSERVATIONS:

1. **Clear Workout Intensity and Calorie Burn Hierarchy**
    - Swimming (558.8 cal) and Cycling (555.7 cal) burn significantly more calories than low-impact activities.
    - Yoga (258.1 cal) and Pilates (262.7 cal) burn less than half the calories of high-intensity workouts.
    - The wide range in calorie burn (~300 cal difference) indicates importance of activity selection.
2. **Experience Level Strongly Impacts Body Composition**
    - Professional athletes (18.39%) have dramatically lower fat percentage than beginners (26.16%).
    - Each step up in experience level shows consistent decrease in body fat percentage.
    - This represents a ~30% reduction in body fat from beginner to professional levels.
3. **BMI Category Determines Calorie Burn Efficiency**
    - Obese individuals burn 44% more calories per minute (12.01) than those with normal BMI (8.33).
    - Overweight participants (10.42 cal/min) fall between these extremes.
    - Higher BMI consistently correlates with greater energy expenditure per unit time.
4. **Gender Differences in Hydration Patterns**
    - Female (0.0332) and non-binary participants (0.0330) maintain higher hydration per kg than males (0.0312).
    - The difference between female and male hydration levels is approximately 6.4%.
    - Suggests potential physiological or behavioral differences in fluid intake and retention.
5. **Workout Frequency and Duration Scale with Experience**
    - Professionals train 49% more frequently (4.67 days/week) than beginners (3.13 days/week).
    - Session length for professionals (1.03 hours) is 66% longer than beginners (0.62 hours).
    - Total weekly exercise time ranges from 1.94 hours (beginners) to 4.81 hours (professionals).
6. **Heart Rate Metrics Vary By Activity Type**
    - Soccer players show highest heart rate recovery (116.96), while yoga practitioners show lowest (112.80).
    - Swimming produces highest maximum BPM (188.43) despite being non-weight bearing.
    - Strength training ranks high in both recovery (116.47) and maximum BPM (187.78).
7. **Workout Intensity and Duration Show Inverse Relationship**
    - High-intensity workouts typically have shorter durations.
    - Lower-intensity activities like walking and yoga allow for longer sessions.
    - Suggests metabolic or fatigue limitations on sustained high-intensity efforts.
8. **Portion of Participants Across BMI Categories Indicates Fitness Population**
    - Distribution across normal, overweight, and obese categories reflects general population trends.
    - Data suggests fitness activities attract participants from all BMI categories.
    - Provides opportunity for targeted programming based on body composition.

FINDINGS AND IMPLICATIONS

1. **Activity Selection is Critical for Calorie Management**
    - Choosing high-calorie-burning activities can double energy expenditure per session.
    - Without selecting appropriate activities, calorie deficit goals will be difficult to achieve.
    - Swimming and cycling should be promoted for maximum calorie burn potential.
2. **Experience Progression Drives Body Composition Improvements**
    - Consistent training leads to significant reductions in body fat percentage.
    - Progress from beginner to advanced status should be tracked and celebrated.
    - Programming should accommodate the physiological changes that occur with experience.

3. **BMI-Specific Programming is Necessary**
    - Higher BMI individuals burn calories more efficiently per minute.
    - Workout duration can be adjusted based on BMI to achieve similar calorie expenditure.
    - Promoting shorter, more intense workouts for obese individuals may optimize time efficiency.
4. **Gender-Specific Hydration Protocols Should Be Implemented**
    - Female and non-binary participants require higher hydration levels per kg of body weight.
    - Hydration guidelines should be tailored to account for these differences.
    - Education on proper hydration strategies should emphasize personalized needs.
5. **Training Volume Increases Must Be Gradual**
    - The substantial difference in training frequency and duration between experience levels requires thoughtful progression.
    - Weekly training volume should increase by approximately 20-30% between experience levels.
    - Both frequency and duration should increase simultaneously for optimal adaptation.
6. **Heart Rate Recovery Indicates Activity-Specific Adaptations**
    - Activities with high HRR values should be incorporated for cardiovascular health benefits.
    - Monitoring HRR can help identify proper training intensity and recovery needs.
    - Participants should incorporate a mix of activities to stimulate different cardiovascular adaptations.
7. **Optimize Workout Structure Based on Intensity**
    - High-intensity workouts should be programmed for shorter durations.
    - Lower-intensity activities can be extended for greater volume.
    - Interval-based approaches can maximize benefits of high-intensity work while managing fatigue.
8. **Targeted Approaches for Different BMI Categories**
    - Focus on intensity for obese and overweight individuals to maximize calorie burn efficiency.
    - Normal BMI individuals may require longer sessions to achieve similar calorie expenditure.
    - Progression plans should account for changing calorie burn rates as BMI decreases.

CONCLUSION

The fitness tracker data analysis revealed key behavioral and health-related insights that can directly support the Health Tech Company's goal of delivering personalized and effective fitness solutions.

- Workout intensity strongly correlates with calories burned per minute, highlighting it as a reliable metric for tracking workout efficiency and informing personalized training recommendations.
- Despite higher intensity, users across all experience levels, including advanced, show poor hydration per kg, suggesting a universal need for hydration tracking features.
- Fat percentage differences across gender were significant, aligning with expected trends and offering a benchmark for personalized health assessments.
- Users with a higher BMI (Obese category) engage in workouts at lower intensities, indicating the need for adaptive workout plans tailored to their capabilities.
- Hydration per kg and BMI category showed no significant link, reinforcing that hydration issues are not weight-dependent but a broader behavioral pattern.

These insights can help the company:

- Enhance personalization in workout plans and hydration reminders.
- Improve engagement through behavior-specific nudges and in-app coaching.
- Support inclusivity by designing for diverse fitness levels and health goals.

By integrating these findings, the company can strengthen user outcomes and position itself as a leader in intelligent, data-driven health technology.

KEY CHALLENGES AND MITIGATION PLAN

Anticipated Challenges

- **Data Availability**
  Fitness-related data from wearable devices may be limited due to privacy policies or lack of standardized public datasets. Additionally, some variables essential to health insights (e.g., calorie intake) may be missing or underreported.
- **Insufficient Data Volume or Diversity**
  The dataset may not fully represent the broader population in terms of age groups, fitness levels, or health conditions. This can impact the generalizability of the findings to the stakeholder's target audience.
- **Handling Missing or Inconsistent Entries**
  Fitness tracker data often includes gaps due to user non-compliance (e.g., not wearing the device), sensor inaccuracies, or app syncing issues. Deciding how to address missing data is critical for maintaining analytical integrity.
- **Time Constraints**
  Given academic deadlines and the depth of analysis required (exploratory, inferential, predictive, and clustering), there may be pressure to balance quality and speed without compromising insightfulness.

Mitigation Plan

- **Data Availability**
  Use anonymize public datasets from trusted platforms like Kaggle samples, or open-access health studies.
- **Insufficient Data**
  Supplement existing data with relevant health survey data or published statistics to support observations. If feasible, apply techniques like data augmentation or sampling to balance demographic representation.
- **Handling Missing Values**
  Apply robust imputation strategies (mean/median for numerical values, mode for categorical ones) and

leverage time-series continuity for filling gaps. Clearly document all imputation steps to ensure transparency.

- **Time Constraints**
  Follow a modular and agile approach by breaking the project into phases: data exploration, EDA, inferential analysis, visualization. Automate repetitive tasks and use built-in libraries like seaborn, matplotlib, and scikit-learn for efficiency.

FEASIBILITY AND RELEVANCE

The insights generated through this analysis are both actionable and highly relevant for health tech companies aiming to improve user engagement, personalize wellness recommendations, and promote healthier lifestyles. By understanding patterns in demographic factors, businesses can build smarter fitness tools, tailor interventions for specific user groups, and increase overall platform effectiveness. These data-driven insights enable informed decisions around feature development, goal-setting algorithms, and user retention strategies — ultimately contributing to improved health outcomes and sustained user satisfaction.

REFERENCES

- **Dataset:** Kaggle Fitness Tracker Dataset
- **Tools & Technology:** Jupyter Notebook, Python Libraries,ChatGPT
- **Zedstatistics:** zedstatistics - YouTube
- **Hypothesis Testing:** Choosing the Right Statistical Test - Google Docs