

CHAPTER THREE: METHODOLOGY

3.0 Introduction

This chapter outlines the methodology employed to analyze the data on the top 50 exercises for the human body. The research utilized statistical and computational techniques to explore patterns, build predictive models, and uncover insights from the dataset. This section explains the research design, target population, sampling techniques, data collection methods, data analysis procedures, ethical considerations, and limitations of the study.

3.1 Research Design

This study adopts a **descriptive and analytical research design**. The descriptive aspect involves summarizing key attributes, such as calorie burn, difficulty level, and equipment needs. The analytical approach includes regression modeling, clustering, and data visualization to derive deeper insights. This design is appropriate as it allows the identification of patterns, relationships, and clusters within the exercise data without experimental manipulation.

3.2 Target Population

The target population includes exercises that cater to diverse fitness levels and target multiple muscle groups. The dataset comprises 50 exercises selected for their popularity, effectiveness, and variety. Each exercise is characterized by parameters like sets, reps, calorie burn, difficulty level, and required equipment. The population is relevant as it represents a wide spectrum of fitness routines applicable to individuals with varying goals and capabilities.

3.3 Sample Size and Sampling Techniques

The sample consists of all 50 exercises listed in the dataset, representing a complete enumeration rather than a subset. Since the dataset was pre-curated to represent diversity in exercises, no additional sampling techniques were necessary. The decision to use the entire dataset ensures comprehensive analysis and minimizes the risk of omitting critical patterns.

3.4 Data Collection Methods

The dataset was obtained in CSV format and contains eight variables:

- a) Name of the exercise
- b) Sets
- c) Reps
- d) Calories burned in 30 minutes
- e) Benefit
- f) Target muscle group
- g) Equipment needed
- h) Difficulty level

Data was pre-collected and formatted for analysis, eliminating the need for field data collection. Consistency checks and transformations (e.g., converting variables to numeric or factor types) were performed to ensure analytical readiness.

3.5 Data Analysis

Statistical Techniques

- a) **Descriptive Analysis:** Summary statistics and visualizations (e.g., bar plots, histograms) were used to describe the distribution of variables like calorie burn and difficulty level.
- b) **Regression Analysis:**
 - A linear regression model was developed to predict calorie burn based on sets, reps, and difficulty level.
 - A logistic regression model was applied to classify exercises into beginner and advanced/intermediate categories using equipment type and reps.
- c) **Clustering:** K-means clustering grouped exercises based on sets and reps, revealing patterns in exercise intensity.
- d) **Aggregate Analysis:** Averages for sets and reps were computed for different target muscle groups.

Tools

Data analysis was performed using R, with visualization packages like ggplot2. Regression models and clustering were executed using standard statistical libraries.

3.6 Ethical Considerations

Since the dataset contains no personal or sensitive information, ethical concerns are minimal. However, care was taken to ensure accuracy, transparency, and reproducibility of analysis. Proper citations and acknowledgments were made for the data source.

3.7 Limitations of the Study

- a) **Limited Dataset Size:** With only 50 exercises, the findings may not generalize to all fitness routines.
 - i. **Mitigation:** Focus on diverse exercises within the sample to maximize representativeness.
- b) **Pre-Curated Data:** The dataset may lack variability in some variables, such as equipment needs.
 - i. **Mitigation:** Utilize clustering to reveal hidden patterns within available data.
- c) **No Experimental Validation:** Results rely on pre-existing data without real-world testing.
 - i. **Mitigation:** Suggest practical applications for future experimental validation.

This methodology provides a comprehensive framework to analyze the dataset, ensuring that insights are both data-driven and actionable.