Predicting the Amount of Calories Burned During a Workout

Kerry Yu

2024-12-03

Contents

Introduction **1**
What is our goal? 1
Why? 2

Data Description **2**

Introduction

What is our goal?

The goal of this project is to build a model that will predict the amount of calories a person will burn during a workout. The big question of day is **"How many calories will a person burn during a specific workout session?"**. The calories a person burns represents the total energy expenditure during physical

activity. It is influenced by physiological attributes (weight, age, sex) and external factors in workout sessions (type, intensity, duration).

Why?

At the start of this school year, I was really interested in bettering my health and my fitness, so I started going to the gym in an attempt to lose some weight. In this pursuit, I realized that being able to understand the factors that influence the amount of calories burned provides helpful information on improving workout routines. By being able to estimate the amount calories a person burns during a workout, they can change up their routines to align with their own fitness goals, such as weight loss, endurance building, and strength improvement.

Data Description

The dataset for this project was retrieved from Kaggle, titled **"Gym Members Exercise Tracking Dataset"** by user Vala Khorasani.

First, we'll have to load our packages and read in our data to be able to work with it.

```
fitness <- read.csv("/Users/kerryyu/Documents/131/Final Project/data/gym_members_exercise_tracking.csv")
set.seed(076)
fitness <- clean_names(fitness)

head(fitness)
```

```
##   age gender weight_kg height_m max_bpm avg_bpm resting_bpm
## 1  56   Male      88.3     1.71     180     157          60
## 2  46 Female      74.9     1.53     179     151          66
## 3  32 Female      68.1     1.66     167     122          54
## 4  25   Male      53.2     1.70     190     164          56
## 5  38   Male      46.1     1.79     188     158          68
## 6  56 Female      58.0     1.68     168     156          74
##   session_duration_hours calories_burned workout_type fat_percentage
## 1                   1.69            1313         Yoga           12.6
## 2                   1.30             883         HIIT           33.9
## 3                   1.11             677       Cardio           33.4
## 4                   0.59             532     Strength           28.8
## 5                   0.64             556     Strength           29.2
## 6                   1.59            1116         HIIT           15.5
##   water_intake_liters workout_frequency_days_week experience_level   bmi
## 1                 3.5                           4                3 30.20
## 2                 2.1                           4                2 32.00
## 3                 2.3                           4                2 24.71
## 4                 2.1                           3                1 18.41
## 5                 2.8                           3                1 14.39
## 6                 2.7                           5                3 20.55
```

Let's play around with our data to see what we are working with.

```
dim(fitness)
```

```
## [1] 973  15
```

So as we can see, the dataset has a total of 973 observations with 15 variables. Normally with any dataset there is a high chance for mising values to exist.

```
vis_miss(fitness)
```