EXPLORING THE RELATIONSHIP BETWEEN EXERCISE TIME AND CALORIES BURNED

A PROJECT REPORT

Submitted by

VISWADHARSHINI TK (2303811724322124)

in partial fulfillment of requirements for the award of the course

AGI1252 - FUNDAMENTALS OF DATA SCIENCE USING R

in

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

K. RAMAKRISHNAN COLLEGE OF TECHNOLOGY

(An Autonomous Institution, affiliated to Anna University Chennai and Approved by AICTE, New Delhi)

SAMAYAPURAM – 621 112

JUNE- 2025

K. RAMAKRISHNAN COLLEGE OF TECHNOLOGY
(AUTONOMOUS)

SAMAYAPURAM – 621 112

BONAFIDE CERTIFICATE

Certified that this project report on **“EXPLORING THE RELATIONSHIP BETWEEN EXERCISE TIME AND CALORIES BURNED”** is the bonafide work of **VISWADHARSHINI TK (230381124322124)** who carried out the project work during the academic year 2024 - 2025 under my supervision.

SIGNATURE

Dr.T. AVUDAIAPPAN, M.E.,Ph.D.,

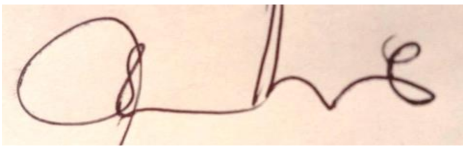
HEAD OF THE DEPARTMENT

PROFESSOR

Department of Artificial Intelligence

K.Ramakrishnan College of Technology (Autonomous)

Samayapuram–621112.



SIGNATURE

Ms.S.Murugavalli., M.E.,(Ph.D).,

SUPERVISOR

ASSISTANT PROFESSOR

Department of Artificial Intelligence

K.Ramakrishnan College of Technology (Autonomous)

Samayapuram–621112.

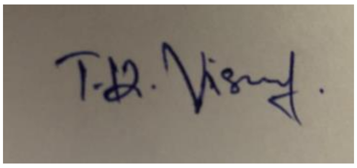
Submitted for the viva-voce examination held on …………**02.06.2025**.

INTERNAL EXAMINER

EXTERNAL EXAMINER

DECLARATION

I declare that the project report on **“EXPLORING THE RELATIONSHIP BETWEEN EXERCISE TIME AND CALORIES BURNED”** is the result of original work done by us and best of our knowledge, similar work has not been submitted to **“ANNA UNIVERSITY CHENNAI”** for the requirement of Degree of **BACHELOR OF TECHNOLOGY**. This project report is submitted on the partial fulfilment of the requirement of the completion of the course **AGI1252 - FUNDAMENTALS OF DATA SCIENCE USING R**



Signature

VISWADHARSHINI TK

Place: Samayapuram

Date: 30.05.2025

ACKNOWLEDGEMENT

It is with great pride that I express our gratitude and in-debt to our institution "**K.Ramakrishnan College of Technology (Autonomous)**", for providing me with the opportunity to do this project.

I glad to credit honourable chairman **Dr. K. RAMAKRISHNAN**, **B.E.,** for having provided for the facilities during the course of my study in college.

I would like to express our sincere thanks to our beloved Executive Director **Dr. S. KUPPUSAMY, MBA, Ph.D.,** for forwarding to my project and offering adequate duration in completing my project.

I would like to thank **Dr. N. VASUDEVAN, M.Tech., Ph.D.,** Principal, who gave opportunity to frame the project the full satisfaction.

I whole heartily thanks to **Dr. T. AVUDAIAPPAN, M.E.,Ph.D.,** Head of the department, **ARTIFICIAL INTELLIGENCE** for providing his encourage pursuing this project.

I express our deep expression and sincere gratitude to my project supervisor **Ms.S.Murugavalli., M.E.,(Ph.D).,** Department of **ARTIFICIAL INTELLIGENCE,** for her incalculable suggestions, creativity, assistance and patience which motivated me to carry out this project.

I render my sincere thanks to Course Coordinator and other staff members for providing valuable information during the course.

I wish to express my special thanks to the officials and Lab Technicians of my departments who rendered their help during the period of the work progress.

INSTITUTE

Vision:

- To serve the society by offering top-notch technical education on par with global standards.

Mission:

- Be a center of excellence for technical education in emerging technologies by exceeding the needs of industry and society.
- Be an institute with world class research facilities.
- Be an institute nurturing talent and enhancing competency of students to transform them as all – round personalities respecting moral and ethical values.

DEPARTMENT

Vision:

- To excel in education, innovation, and research in Artificial Intelligence and Data Science to fulfil industrial demands and societal expectations.

Mission

- To educate future engineers with solid fundamentals, continually improving teaching methods using modern tools.
- To collaborate with industry and offer top-notch facilities in a conducive learning environment.
- To foster skilled engineers and ethical innovation in AI and Data Science for global recognition and impactful research.
- To tackle the societal challenge of producing capable professionals by instilling employability skills and human values.

PROGRAM EDUCATIONAL OBJECTIVES (PEO)

- **PEO1:** Compete on a global scale for a professional career in Artificial Intelligence and Data Science.
- **PEO2:** Provide industry-specific solutions for the society with effective communication and ethics.
- **PEO3** Enhance their professional skills through research and lifelong learning initiatives.

PROGRAM SPECIFIC OUTCOMES (PSOs)

- **PSO1:** Capable of finding the important factors in large datasets, simplify the data, and improve predictive model accuracy.
- **PSO2:** Capable of analyzing and providing a solution to a given real-world problem by designing an effective program.

PROGRAM OUTCOMES (POs)

Engineering students will be able to:

1. **Engineering knowledge:** Apply knowledge of mathematics, natural science, computing, engineering fundamentals, and an engineering specialization to develop solutions to complex engineering problems.
2. **Problem analysis:** Identify, formulate, review research literature and analyze complex engineering problems reaching substantiated conclusions with consideration for sustainable development.
3. **Design/development of solutions:** Design creative solutions for complex engineering problems and design/develop systems/components/processes to meet identified needs with consideration for the public health and safety, whole-life cost, net zero carbon, culture, society and environment as required.
4. **Conduct investigations of complex problems:** Conduct investigations of complex engineering problems using research-based knowledge including design of experiments, modelling, analysis & interpretation of data to provide valid conclusions.
5. **Engineering Tool Usage:** Create, select and apply appropriate techniques, resources and modern engineering & IT tools, including prediction and modelling recognizing their limitations to solve complex engineering problems.
6. **The Engineer and The World:** Analyze and evaluate societal and environmental aspects while solving complex engineering problems for its impact on sustainability with reference to economy, health, safety, legal framework, culture and environment.
7. **Ethics:** Apply ethical principles and commit to professional ethics, human values, diversity and inclusion; adhere to national & international laws.

8. **Individual and Collaborative Team work:** Function effectively as an individual, and as a member or leader in diverse/multi-disciplinary teams.

9. **Communication:** Communicate effectively and inclusively within the engineering community and society at large, such as being able to comprehend and write effective reports and design documentation, make effective presentations considering cultural, language, and learning differences.

10. **Project management and finance:** Apply knowledge and understanding of engineering management principles and economic decision-making and apply these to one's own work, as a member and leader in a team, and to manage projects and in multidisciplinary environments.

11. **Life-long learning:** Recognize the need for, and have the preparation and ability for i) independent and life-long learning ii) adaptability to new and emerging technologies and iii) critical thinking in the broadest context of technological change.

ABSTRACT

The Exploring the Relationship Between Exercise Time and Calories Burned project utilizes data science techniques in R to analyze how exercise duration influences the number of calories burned. By collecting and preprocessing data from fitness datasets or wearable device exports, we construct a feature-rich dataset that captures exercise- related variables such as duration, type of activity, intensity level, and user demographics. Through exploratory data analysis (EDA) and regression modeling, we investigate the strength and nature of the correlation between exercise time and calories burned. Visualization tools and statistical validation techniques are employed to interpret the results effectively. The findings can support personalized fitness recommendations, goal setting, and wellness tracking in health applications, showcasing the practical application of R in real-time health analytics and decision- making. Evaluation is conducted using regression diagnostics, correlation coefficients, and residual analysis. Visualization techniques such as scatter plots, trend lines, and regression curves in R help illustrate the relationship and model accuracy.

ABSTRACT WITH POs AND PSOs MAPPING

CO 5 : BUILD DATA SCIENCE USING R PROGRAMMING FOR SOLVING REAL-TIME PROBLEMS.

ABSTRACT	POs MAPPED	PSOs MAPPED
Apply the knowledge of mathematics, science, engineering fundamentals. Identify, formulate and analyze real-world problems. Design solutions for complex engineering problems. Conduct investigations using data analysis and research methods. Use modern tools like R for engineering applications. Apply contextual knowledge to assess societal, health, safety, legal and cultural issues. Work effectively in multidisciplinary teams. Communicate effectively on complex engineering activities.	**PO1 -3** **PO2 -3** **PO3 -3** **PO4 -3** **PO5 -3** **PO6 -3** **PO7 -3** **PO8 -3** **PO9 -3** **PO10 -3** **PO11-3**	**PSO1 -3** **PSO2 -3**

Note: 1- Low, 2-Medium, 3- High

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

1.1 Objective

The primary goal of this project is to examine the relationship between exercise time and calories burned using R programming. By analyzing real-world fitness data, we aim to identify how duration and other variables impact energy expenditure. The outcome supports personalized fitness planning and enhances the development of data- driven health applications.

1.2 Overview

This project applies statistical analysis and visualization techniques to uncover patterns between exercise time and caloric expenditure. By leveraging datasets from fitness tracking apps or wearable devices, we preprocess and analyze the data using R to find correlations, trends, and predictive insights**.**

1.3 R Programming Used

The R workflow begins with data preprocessing (cleaning, handling missing values, and data normalization). EDA follows, where trends are visualized using plots. Regression modeling (e.g., linear regression) is used to analyze the relationship between variables. The project concludes with evaluation metrics and visualizations to interpret model performance and derive insights.

CHAPTER 2

PROJECT METHODOLOGY

1) Data Collection

Objective: Obtain a dataset containing information about exercise time and calories burned.

- Sources:
    - Use publicly available datasets (e.g., Kaggle, UCI Machine Learning Repository).
    - Simulate synthetic data if real-world data is unavailable.

2)Data Preprocessing

Objective: Clean and prepare the data for analysis.

Tasks:

- Handle missing values using na.omit() or tidyr::fill().
- Convert data types appropriately.
- Filter unrealistic entries (e.g., negative time or calories).

3) Exploratory Data Analysis (EDA)

Objective: Understand the data structure and identify trends or patterns.

Tasks:

- Summary statistics: summary(), mean(), sd().
- Visualizations:
    - Histogram of exercise time: ggplot2::geom_histogram()
    - Scatter plot of exercise time vs. calories burned: ggplot2::geom_point()

4) Statistical Analysis

Objective: Quantify the relationship between exercise time and calories burned.

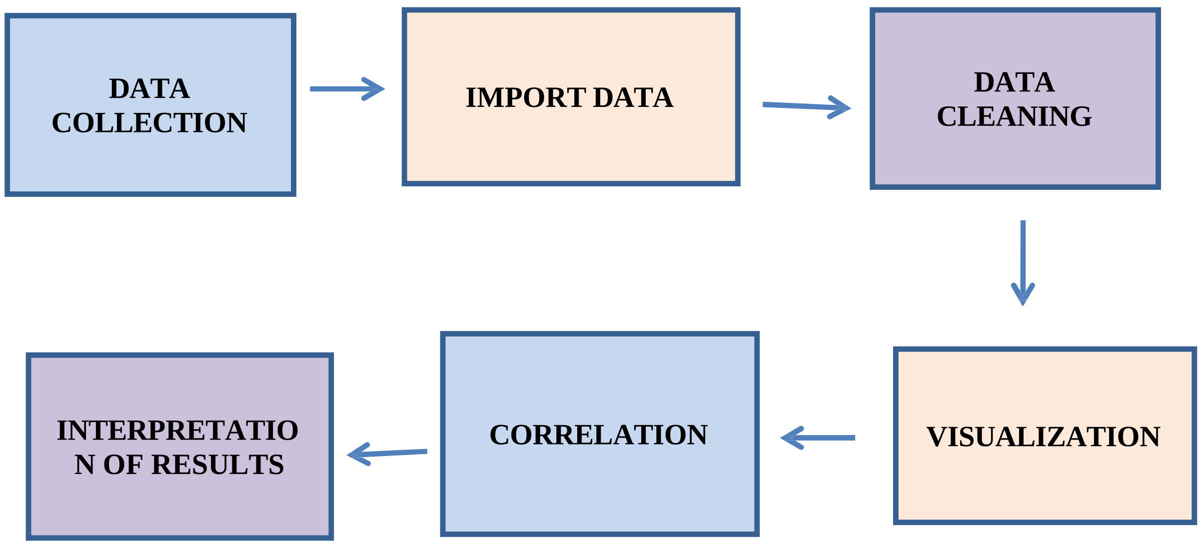
Tasks:

- Compute correlation coefficient.
- Perform linear regression: lm(calories ~ time)
- Check assumptions of regression (normality, homoscedasticity).

5) Model Evaluation and Validation

- Objective: Evaluate how well the model explains the variation in calories burned.
- Tasks:
    - Use R^2 and Adjusted R^2 from regression summary.
    - Residual analysis using diagnostic plots.
    - Cross-validation (e.g., caret::train() with k-fold CV).

2.2 Block Diagram



CHAPTER 3
MODULE DESCRIPTION

3.1 Data Acquisition Module

This module collects exercise and health data from fitness trackers, health APIs, or exported CSV files. Variables include exercise duration, calories burned, activity type, intensity level, and user attributes. The data is imported into R for processing.

3.2 Feature Engineering Module

Raw data is transformed into structured inputs. This involves handling missing values, normalizing continuous variables, and encoding categorical data such as activity types. Features like MET (Metabolic Equivalent of Task) may also be derived.

3.3 Exploratory Data Analysis (EDA) Module

This module uncovers data patterns using visual tools (e.g., scatter plots, histograms) and descriptive statistics. Correlation matrices are computed to understand how exercise time aligns with calorie expenditure across different activity types.

3.4 Genre Clustering Module

Linear regression and multiple regression models are used to quantify the relationship between exercise time and calories burned. Evaluation involves checking R-squared values, residual plots, and statistical significance. Visualization tools like ggplot2 are used to plot regression lines and interpret model results.

CHAPTER 4

CONCLUSION & FUTURE SCOPE

This project successfully demonstrates how R programming and statistical modeling can analyze the link between exercise time and calories burned. Through data preprocessing, visualization, and regression analysis, strong correlations were observed, offering valuable insights for personalized fitness and health tracking. The framework developed can be scaled to include more variables such as heart rate, gender, or age to refine predictions. Future work may involve integrating real-time data streams from fitness APIs and building interactive dashboards for users to track progress dynamically.In conclusion, this project successfully analyzed the relationship between exercise time and calories burned using statistical tools and visualization techniques in R. The insights revealed a clear positive correlation, indicating that as exercise duration increases, so does caloric expenditure—though the rate of increase may vary depending on factors such as activity type, intensity, and individual characteristics.

This analysis not only validates commonly held assumptions about exercise efficiency but also emphasizes the importance of data-driven insights in personal health and fitness. The use of regression models allowed for predictive capabilities, helping estimate calories burned for given durations across different activity categories. These models can be incorporated into health monitoring tools and mobile applications to assist users in setting realistic fitness goals.

APPENDICES

APPENDIX A – SOURCE CODE

```
library(shiny)
library(ggplot2)
library(cluster)
library(caret)
library(plotly)
library(dplyr)

# Simulated dataset (you can replace this with your own CSV)
set.seed(123)
n <- 300
data <- data.frame(
  Time = runif(n, 10, 120),  # Time in minutes
  Calories = runif(n, 10, 120) * 8 + rnorm(n, 0, 40)  # Calories burned with noise
)

# UI
ui <- fluidPage(
  titlePanel("Exploring the Relationship Between Time and Calories Burned"),

  sidebarLayout(
    sidebarPanel(
      sliderInput("timeRange", "Select Time Range (minutes):",
                  min = 10, max = 120, value = c(20, 100)),
      numericInput("new_time", "Enter Time (min) for Prediction:", value = 60),
      actionButton("predict_btn", "Predict Calories")
```

```
    ),

    mainPanel(
      plotlyOutput("scatterPlot"),
      verbatimTextOutput("modelSummary"),
      verbatimTextOutput("predictionOutput")
    )
  )
)

# Server
server <- function(input, output) {

  filtered_data <- reactive({
    data %>% filter(Time >= input$timeRange[1], Time <= input$timeRange[2])
  })

  model_fit <- reactive({
    train(Calories ~ Time, data = filtered_data(), method = "lm")
  })

  output$scatterPlot <- renderPlotly({
    p <- ggplot(filtered_data(), aes(x = Time, y = Calories)) +
      geom_point(color = "blue", alpha = 0.6) +
      geom_smooth(method = "lm", se = TRUE, color = "red") +
      labs(title = "Time vs Calories Burned", x = "Time (minutes)", y = "Calories Burned")
+
      theme_minimal()
```

```
    ggplotly(p)
  })

  output$modelSummary <-
    renderPrint({
    summary(model_fit()$finalModel)
  })

  observeEvent(input$predict_btn,  {
    new_input <- data.frame(Time =
    input$new_time) prediction <-
    predict(model_fit(), new_input)

    output$predictionOutput <- renderPrint({
      cat("Predicted Calories Burned for", input$new_time,
      "minutes:\n") round(prediction, 2)
    })
  })
}

# Run the
application
shinyApp(ui =
ui, server =
server)
```

APPENDIX B – SCREENSHOTS

REFERENCES

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning with Applications in R. Springer.

- Wickham, H., & Grolemund, G. (2016). R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. O'Reilly Media.

- Health datasets from Kaggle, Fitbit, or Apple Health.

- Kassambara, A. (2017). Practical Guide to Cluster Analysis in R. STHDA.