# Detailed Project Report: Walking Analysis

# Unveiling Patterns in Personal Mobility Data

## Introduction

**Project Overview**: An extensive examination of Apple Health walking data is carried out in this study. The main objective is to investigate the complex interrelationships and connections among three essential variables: energy expenditure, step length, and gait asymmetry. The purpose of this study is to gain an important understanding of individual mobility patterns.

## Motivation

### Purpose and Personal Interest

The fundamental inspiration for this project came from a strong interest in personal health and fitness statistics. The purpose of using data from my regular walks was to discover insights that could lead to improved health and personal well-being.
The research also answers a bigger question about how current technologies, such as Apple Health, might be utilized to examine and enhance one's lifestyle.

## Data Source

### Data Collection and Origin

The datasets for this study were obtained straight from my Apple Health app, guaranteeing that the data used was accurate, personal, and relevant.
The data was collected meticulously over several months, and it included several parameters linked to walking activities such as step length, walking asymmetry, and active energy burned. This extensive collection period was designed to provide a complete picture of my walking patterns over time.

## Exploratory Data Analysis (EDA)

**Data Overview**: The investigation delves into the intricacies of each indicator, using datasets that include walking step length, walking asymmetry percentage, and active energy burned.

### Descriptive Statistics:

- Initial exploration involved summarizing key statistical measures for each dataset.
- Issues of missing or inconsistent data, especially in the Walking Step Length dataset, were identified and addressed accordingly.

### Visualization of EDA

**Histograms**: The histograms for each statistic provide a visual picture of its distribution.

- The Walking Step Length and Walking Asymmetry Percentage histograms revealed the range and frequency of these measurements. ( Appendix1, Appendix2)

- The Active Energy Burned histogram highlighted the variability in daily energy expenditure. (Appendix3)

**Time-Series Analysis**: Time-series plots were created for each metric to visualize their trends over time. These plots were instrumental in identifying patterns and seasonal variations in the data. (Appendix4, Appendix5, Appendix6)

# Correlation Analysis

## Multiple Linear Regression Model:

- The model quantified the relationships among the variables.
- The research found that step length and walking asymmetry have limited explanatory power for energy expenditure (R-squared = 0.052).
- The coefficients indicated that there was a modest inverse link between step length and energy burned, as well as a positive relationship between walking asymmetry percentage and energy burned.

## Machine Learning Models

### Clustering with K-Means

- aims to identify patterns in data using three basic criteria.
- The Elbow Method revealed an ideal cluster count, revealing natural segmentations in the walking data. (Appendix7)

### Classification using Random Forest

- A classification model was created to predict energy expenditure categories using step duration and asymmetry.
- The model attained an accuracy of about 52.94%.
- The confusion matrix and feature importance plots offered more information about the model's prediction skills and the impact of each feature.

# Advanced Visualizations

- **Confusion Matrix Visualization**: This provided an intuitive sense of the model's prediction accuracy, focusing on the true and false classifications. (Appendix8)
- **Feature Importance Plot**: Determined the relative importance of step length and walking asymmetry in predicting energy expenditure. (Appendix9)

# Findings

## In-Depth Analysis of Walking Data

### 1. Understanding of Step Length and Walking Asymmetry:

- The exploratory research and visualizations found interesting patterns in step length and walking asymmetry. We noticed a wide variety of step lengths and asymmetry percentages, showing that walking habits vary from day to day.

- Histogram displays of these metrics provided a clear image of their distribution, revealing a large range in step lengths and a concentration of walking asymmetry percentages around lower values.

### 2. Correlation Between Step Length and Walking Asymmetry:

- Multiple Linear Regression study revealed important insights into the association between step length and walking asymmetry. Contrary to our initial expectations, we discovered a minor inverse association between these two parameters. Walking asymmetry decreased as step length rose, albeit this association was not significant. This research reveals a complex relationship between the length of one's steps and the symmetry of their gait.

### 3. Energy Expenditure Insights:

- The time-series study of active energy burned revealed oscillations over time, which reflected changes in daily activity levels. The presentation of this data aided in understanding how energy expenditure changed with different walking patterns.
- However, the Multiple Linear Regression model demonstrated that the variance in energy expenditure could not be robustly predicted just based on step length and walking asymmetry. The model's low R-squared value suggested that these parameters alone explained only a small part of the variation in energy burned.

### 4. Machine Learning Model Insights:

- The Random Forest Classification algorithm, which was designed to categorize days with high and low energy expenditure based on step length and walking asymmetry, demonstrated reasonable accuracy. This finding corroborated the idea that the relationship between these walking parameters and energy expenditure is complicated and possibly impacted by other unaccounted-for variables.
- The feature importance analysis from the classification model revealed which metric between step length and walking asymmetry had a greater influence on predicting energy expenditure categories. Both features contributed to the model, although their predictive potential was limited.

### Summary of Key Learnings

- The analysis underscored the complexity of human walking patterns and their relationship with energy expenditure. While there are correlations between step length, walking asymmetry, and energy burned, these relationships are not straightforward or strong enough to make definitive predictions.
- It was evident that additional factors, possibly including the intensity of the walking activity, individual physiological differences, or even environmental conditions, play a significant role in determining energy expenditure during walking.

## Limitations and Future Work

### Challenges

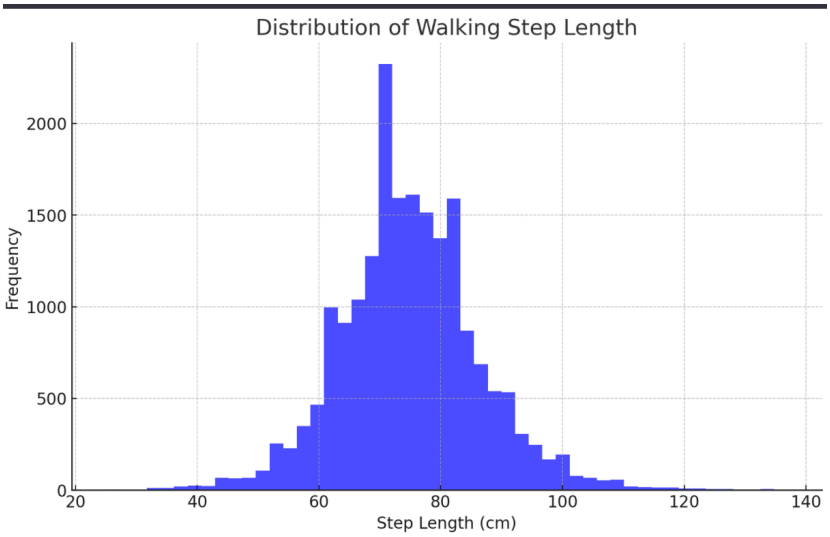
- **Data Limitations**: The analysis was restricted to self-collected data, which may limit its generalizability. Future editions could benefit from a more diversified data collection, which includes a variety of environmental conditions and demographics.
- **Model Complexity**: The models utilized gave basic insights, but they also highlighted the need for more advanced analytical techniques to completely capture the deep linkages within the data.

## Future Directions

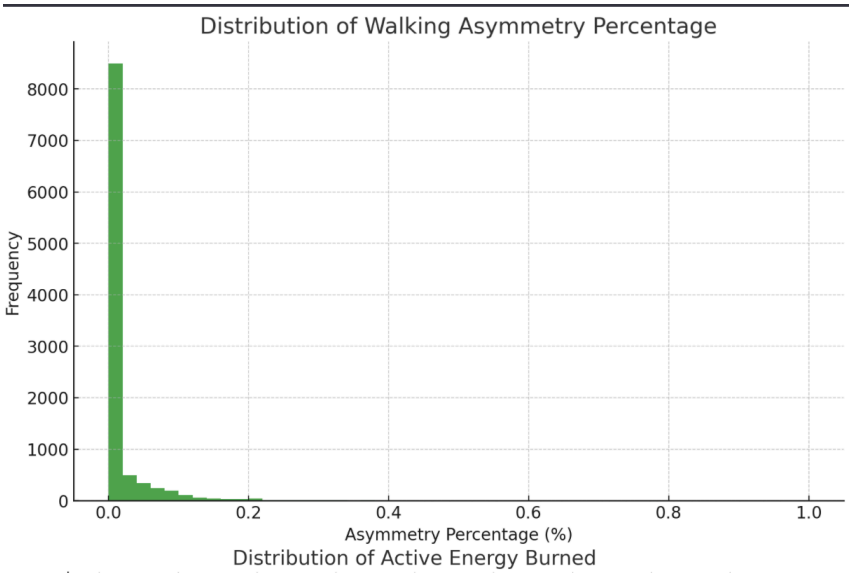
- **Data Enrichment:** Plans are underway to integrate more data kinds, such as heart rate and environmental parameters. This would allow for a more complex study.
- **Advanced Analytical Techniques**: There is a growing interest in more complex statistical models and machine learning algorithms, which have the potential to disclose deeper insights and more accurate forecasts..

# Appendices

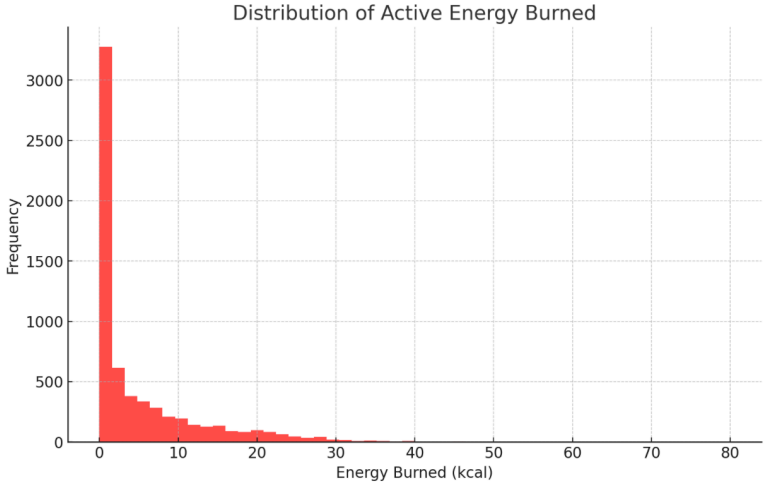
Appendix1:



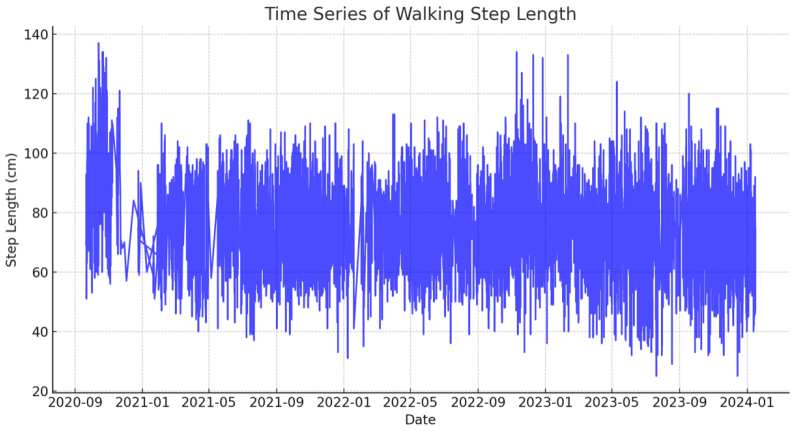

Appnedix2:



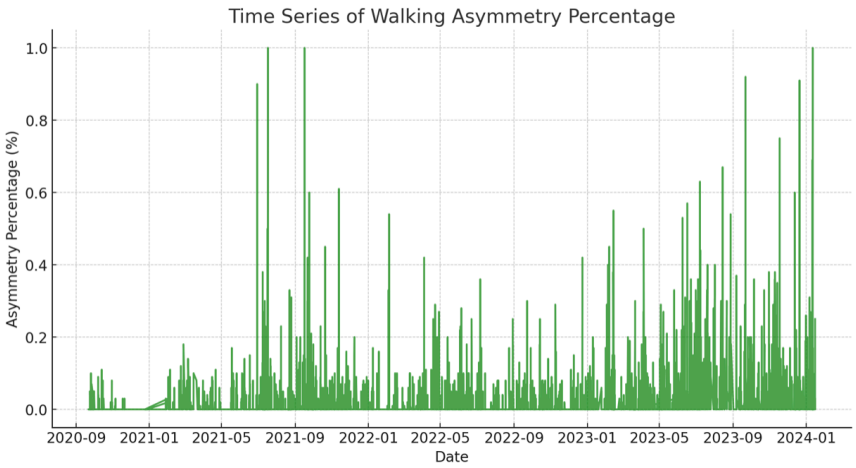

Appendix3:



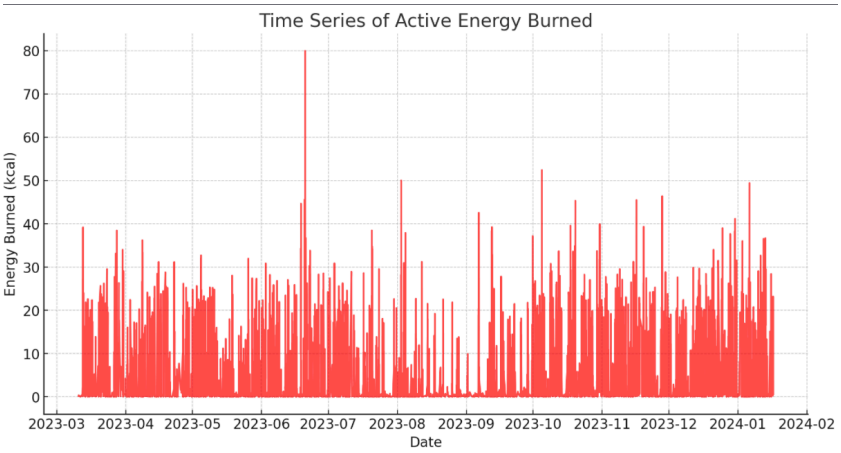
Appendix4:



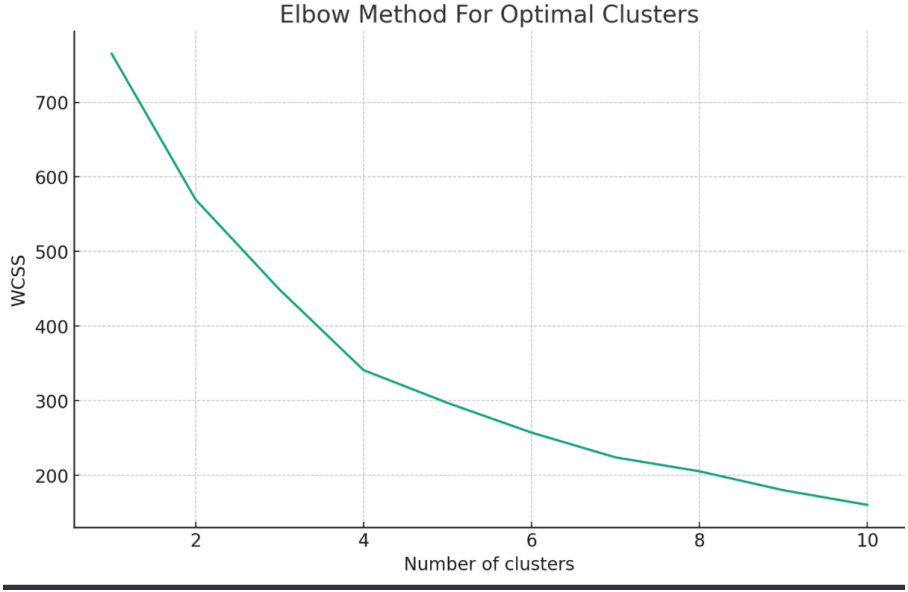

Appendix5:



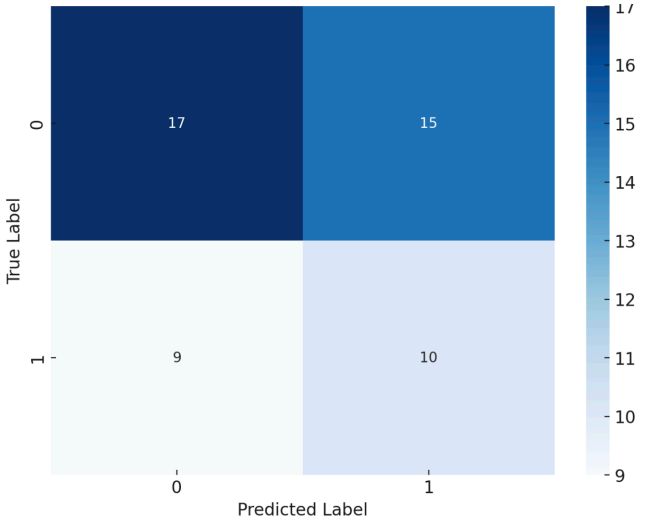

Appendix6:




Appendix7:



Appendix8:




Appendix9: