**Daily Steps Data Analysis Project**

**Motivation**

The primary motivation for this project is to better understand personal physical activity habits and identify areas for improvement in achieving daily fitness goals. Using a dataset collected from iPhone’s Health app and manually logged details, the project explores patterns in daily step counts, active minutes, calories burned, and rest times. This analysis aims to provide actionable insights into the relationship between different variables such as daily step count, goal achievement percentage, and exercise types to improve health outcomes.

**Data Overview**

The dataset combines automatically tracked data from the iPhone’s Health app and manually recorded metrics. It contains the following features:

1. Step Count: Total steps taken daily.

2. Distance (km): Distance covered in kilometers.

3. Calories (kcal): Calories burned through physical activity.

4. Active Time (minutes): Total time spent actively moving.

5. Floor Climbing: Number of floors climbed.

6. Rest Time (minutes): Time spent resting.

7. Goal Achievement (%): Percentage of the daily goal (8000 steps) achieved.

8. Exercise Type: The type of physical activity (e.g., walking, running, none).

9. Weight (kg): Daily weight values.

The dataset spans 90 days and includes 10 columns with no missing data. Data cleaning and feature engineering were conducted to ensure consistency and usability.

**Tools Used**

- Google Colab: Coding and documenting.

- Pandas & NumPy: Data cleaning, manipulation, and structuring.

- Matplotlib & Seaborn: Data visualization.

- SciPy: Statistical analyses.

**Key Findings**

1. Daily Activity Patterns

- Step Count Trends: Weekdays exhibit higher step counts compared to weekends, indicating higher activity levels during workdays.

- Active vs. Rest Time: Days with lower step counts generally show an increase in rest time, suggesting a recovery pattern.

- Daily Step Count Change: Analysis of daily step count changes revealed significant fluctuations, particularly influenced by weekends and exercise type.

**2. Goal Achievement Analysis**

- The daily step goal of 8000 steps was achieved approximately 70-80% of the time.

- Higher goal achievement correlated with increased calories burned and active minutes.

- The relationship between rest time and goal achievement percentage shows varying trends across exercise types (e.g., running, walking).

**3. Exercise Impact**

- Walking is the most common type of physical activity, contributing the most to step counts.

- Running is less frequent but yields higher calorie burn rates and active minutes.

- Visualization of exercise type distributions shows that walking constitutes the majority of physical activities.

**4. Correlation Analysis**

- Positive Correlations:

- Active Time and Calories Burned (r = 0.45).

- Active Time and Step Count (r = 0.38).

- Negative Correlations:

- Rest Time and Active Time (r = -0.17).

- Rest Time and Step Count (r = -0.38).

**5. Weight Analysis**

- Weight showed minimal fluctuations throughout the data collection period, with no significant impact on step counts or calories burned.

- Scatterplots reveal weak relationships between weight and other activity metrics, such as step count and distance covered.

**Hypothesis Testing**

Hypotheses Defined:

- Null Hypothesis (H0): There is no significant relationship between daily active time and calories burned.

- Alternative Hypothesis (H1): Daily active time is significantly and positively correlated with calories burned.

**Methodology**:

Using linear regression analysis, the relationship between active time and calories burned was tested. The results were as follows:

- Slope: 1.807 calories/minute.

- Intercept: 124.984 calories.

- R-squared: 0.203, indicating a moderate positive relationship.

- p-value: 8.30 × 10⁻⁶, which is less than 0.05, allowing rejection of the null hypothesis.

Conclusion from Hypothesis Testing: The null hypothesis was rejected, confirming that daily active time significantly influences calories burned. This supports the claim that increased active time leads to higher calorie expenditure.

**Limitations**

- Timeframe: The dataset spans only three months, limiting the ability to analyze long-term trends or seasonal effects.

- Data Accuracy: Manually recorded data (e.g., exercise type) may contain subjective errors.

- Scope: Advanced machine learning techniques for predictive analysis were not applied, restricting the depth of insights.

**Future Work**

- Seasonal Trends: Extend the dataset to cover at least one year to observe seasonal variations in physical activity.

- Predictive Modeling: Use machine learning models to predict step counts, calories burned, and goal achievement percentages based on historical data.

- Interactive Dashboard: Develop a visualization dashboard for real-time analysis and tracking.

**Conclusion**

This project provided comprehensive insights into daily physical activity patterns, highlighting key correlations between variables such as step count, calories burned, and rest time. The analysis demonstrates the effectiveness of walking as a primary physical activity and underscores the importance of maintaining consistency in daily fitness routines. These findings can serve as a baseline for further studies or the development of fitness optimization tools. The project’s actionable insights provide valuable recommendations for improving physical activity habits and achieving fitness goals more consistently.

**Visualization of Analysis Results**

A diagram of a number of steps

Description automatically generated

A graph with blue lines

Description automatically generated

A graph with green lines

Description automatically generated

A graph of different colored squares

Description automatically generated

A graph of a graph

Description automatically generated with medium confidence

A graph of a graph with blue and green dots

Description automatically generated

A graph of blue bars

Description automatically generated

A graph with purple dots

Description automatically generated

A graph with orange dots

Description automatically generated

A graph with blue and purple dots

Description automatically generated

A graph with red lines and numbers

Description automatically generated

A graph of activity levels

Description automatically generated

A chart with red and blue squares

Description automatically generated

A graph with a red line going up

Description automatically generated