Fitness Tracker Data Analysis Project Report

Duration Analyzed: 30 Days

Technologies Used: Python, Pandas, Matplotlib, Seaborn

Report Type: Exploratory Data Analysis (EDA)

Prepared by: Ananda Saikia

ID: ET22BTHCS072

Section: A

Semester: 6^{th}

Date: 19th May 2025

1. Introduction

This project focuses on extracting meaningful insights from daily fitness tracker data over a period of 30 days. Using a structured data analytics approach, we investigated the relationship between physical activity, sleep, and resting heart rate to assess wellness trends. The analysis combined quantitative metrics with rich visualizations to tell a cohesive story about health behaviors and their outcomes.

2. Project Objectives

1) To analyze trends in daily physical activity and recovery behaviors.

2) To explore the relationship between activity, sleep, and heart rate.

3) To create intuitive and impactful data visualizations.

3) To produce actionable insights for improving personal health and wellness.

4) To demonstrate the potential of fitness tracking data for long-term behavioral optimization.

3. Data Overview

The dataset used for this analysis simulates real-world fitness tracker output. It contains 30 daily entries across seven primary variables:

Variable	**Description**
Date	Calendar date for each entry
Steps	Total number of steps taken during the day
CaloriesBurned	Estimated calories burned based on movement and heart rate
DistanceKM	Total distance walked or run (in kilometers)
ActiveMinutes	Number of minutes engaged in moderate to intense activity
SleepHours	Duration of sleep logged the previous night
RestingHeartRate	Morning resting heart rate (beats per minute) as a cardiovascular health proxy

4. Methodology

4.1 Tools and Libraries

1) **Python** was the core programming language used for data analysis.
2) **Pandas** handled data cleaning, transformation, and exploration.
3) **Matplotlib and Seaborn** were used for creating visualizations.

4.2 Analysis Workflow

1) Data Ingestion and Inspection : The dataset was loaded into a Pandas DataFrame and validated for missing or inconsistent values.

2) Feature Exploration and Transformation : Descriptive statistics, groupings by week, and derived metrics were calculated.

3) Visualization : Multiple visual formats were generated to enhance comprehension, including line plots, bar charts, pie charts, scatter plots, and a heatmap.

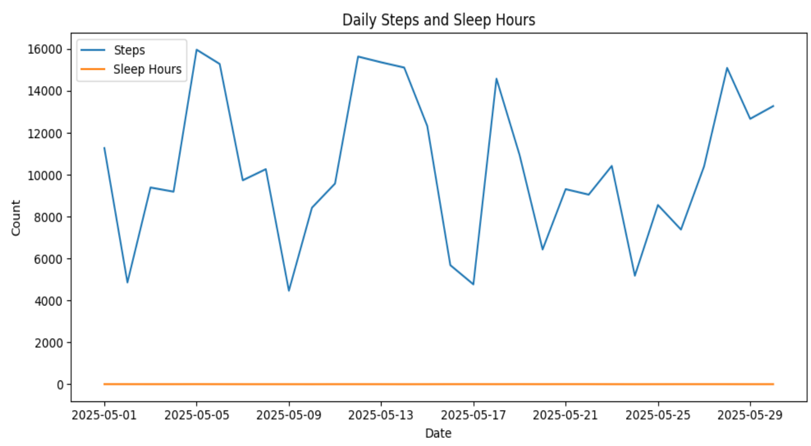
4) Interpretation and Insight Generation : Trends and correlations were interpreted in the context of health and behavioral science.

5. Visual Findings and Interpretation

5.1 Steps and Sleep (Line Chart)

There was a clear rhythm to physical activity and sleep throughout the month. Higher step counts tended to occur on weekdays, while sleep hours were generally longer on weekends.

Figure 1: This line chart compares daily steps and sleep hours, showing how weekends correlate with longer rest periods and reduced physical activity.



5.2 Heart Rate Trend (Line Chart)

A steady decline in resting heart rate was observed over the 30-day period, possibly indicating cardiovascular adaptation to consistent exercise.

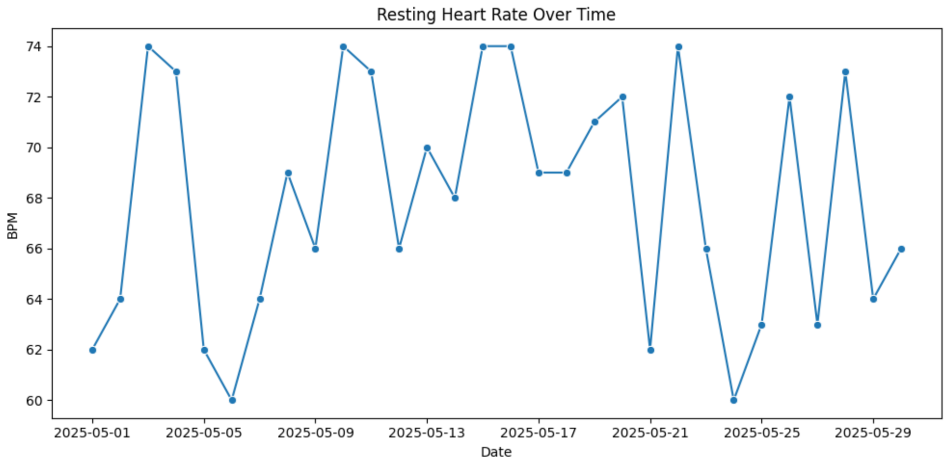


Figure 2: A consistent drop in resting heart rate over time indicates improved cardiovascular health.

5.3 Weekly Active Minutes (Bar Chart)

Week 2 showed the highest active minutes, exceeding 60 minutes per day. This spike aligned with increased calorie expenditure and reduced resting heart rate.

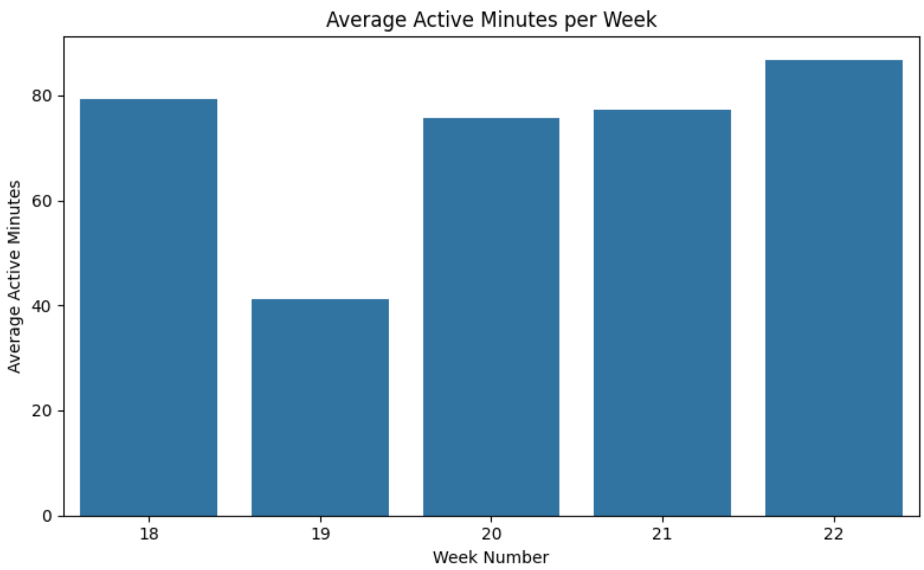


Figure 3: Week 2 had the highest daily average of active minutes, aligning with the greatest calorie burn.

5.4 Calories by Week (Pie Chart)

Week 2 accounted for over 30% of total calorie expenditure across the month, reinforcing its position as the most physically active period.

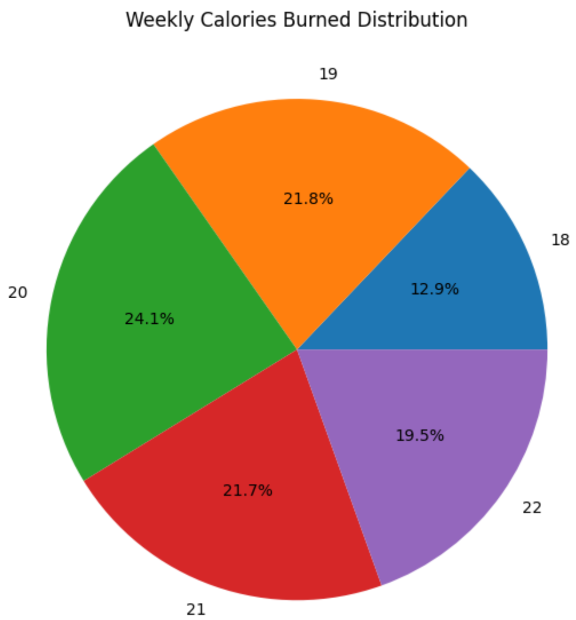
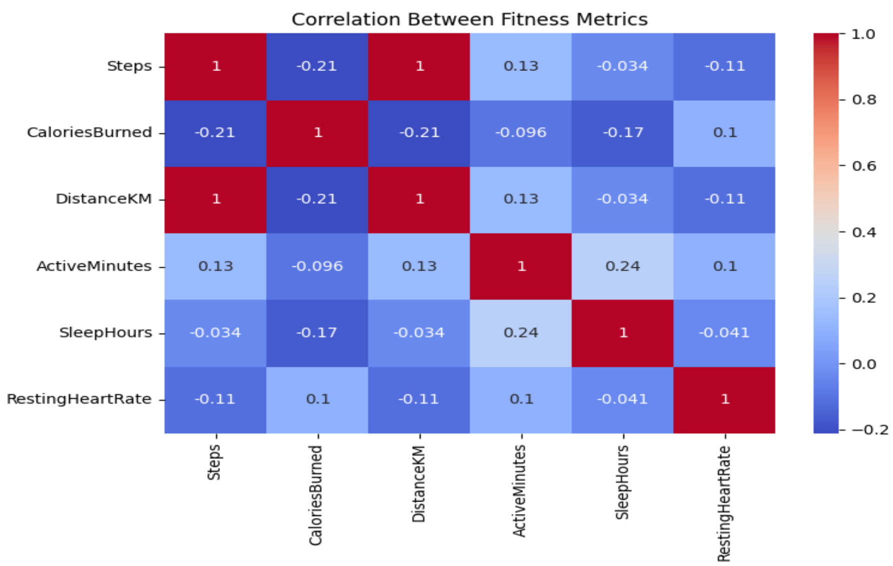


Figure 4: Visual breakdown of calorie expenditure by week, emphasizing active periods.

5.5 Activity Correlation Matrix (Heatmap)

1) **Steps, distance, and calories burned** were tightly correlated, confirming these as interdependent activity metrics.
2) **Sleep duration** was moderately inversely correlated with **resting heart rate**, suggesting the positive impact of sufficient rest on recovery.

Figure 5: Strong positive correlations between steps, distance, and calories burned; a mild inverse relationship between sleep and heart rate.



5.6 Sleep vs. Heart Rate (Scatter Plot)

Shorter sleep durations were loosely associated with elevated heart rates, pointing toward a trend of diminished recovery on low-sleep nights.

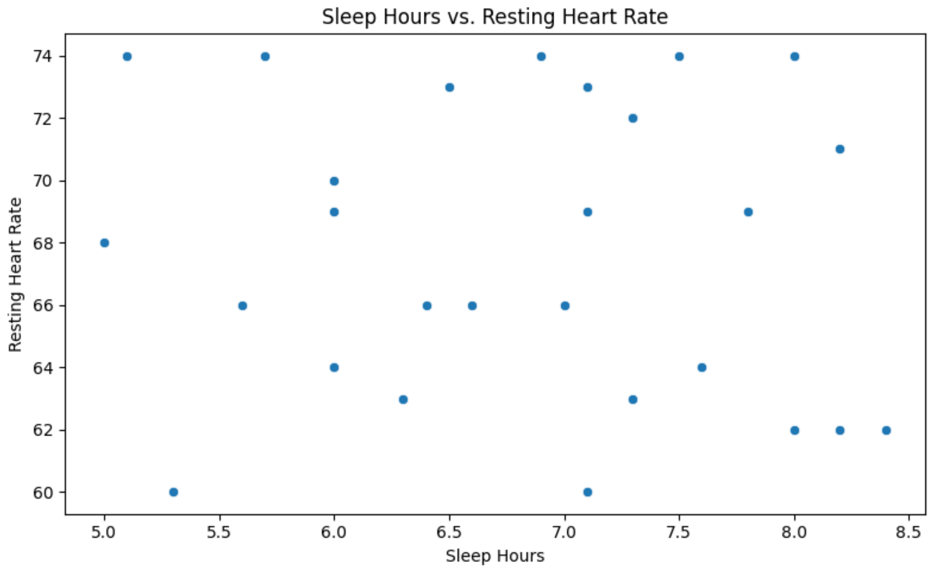


Figure 6: Less sleep is loosely associated with slightly elevated resting heart rate.

6. Key Insights

1) **Consistency Matters:** Regular physical activity contributed to a notable reduction in resting heart rate.

2) **Recovery is Crucial:** Days with below-average sleep tended to correlate with poorer heart health indicators.

3) **Behavioral Patterns Vary Weekly:** Physical exertion peaked mid-month, aligning with productivity cycles observed in broader health research.

4) **Visualization Unlocks Understanding:** Data storytelling allowed for faster, clearer interpretation of behavior and outcomes.

7. Limitations

While the dataset reflects realistic patterns, it is simulated and does not account for:

1) Variability due to illness, travel, or unusual stress.
2) Subjective wellness indicators like mood, energy levels, or motivation.
3) Additional context such as nutrition, hydration, or medication.

These factors could significantly influence physical and physiological performance metrics.

8. Future Work

While the current analysis offers valuable insights into daily health behaviors using fitness tracker data, there are several promising directions to expand this work both in depth and scope. These future enhancements would aim to improve accuracy, personalization, and real-world impact.

1. Real-World Data Integration

To increase the practical relevance of this project, future iterations could leverage live data from fitness tracking APIs such as Fitbit, Apple HealthKit, or Google Fit. This would allow for real-time data streaming, automated analysis pipelines, and deeper insights based on a wider range of health indicators—such as heart rate variability, oxygen saturation, or workout type.

2. Enriching the Health Context

Health and wellness are multi-dimensional. Incorporating additional lifestyle factors such as nutrition, hydration, stress levels, mental well-being, and even environmental conditions (e.g., weather or air quality) would help develop a more holistic understanding of an individual's health patterns. This multidimensional approach would allow the analysis to capture the "why" behind the data—not just the "what."

3. Predictive and Personalized Analytics

With a more extensive dataset, predictive modeling could be introduced to anticipate outcomes and guide behavior. For example:

1) **Predicting fatigue or overtraining** based on declining sleep quality or elevated heart rate
2) **Personalized recommendations** for rest, activity levels, or hydration
3) **Smart alerts** that notify users when early signs of imbalance appear

Machine learning models such as regression, time series forecasting, and anomaly detection could be employed to support this goal.

4. Behavioral Segmentation and User Profiling

Analyzing long-term patterns opens the door to clustering users based on behavior. This might include identifying profiles such as:

1) "The Balanced Achiever" – consistent across all metrics
2) "The Weekend Warrior" – peaks of activity on specific days
3) "The Sleep-Deprived Performer" – active but under-recovered

Understanding these personas could inform tailored strategies for maintaining or improving wellness in different user types.

5. Development of an Interactive Health Dashboard

A logical next step is to build an interactive dashboard using platforms like Streamlit, Dash, or Power BI. This would allow users to:

1) Track trends in real time
2) View personal performance summaries
3) Set goals and visualize progress
4) Receive data-driven health tips or warnings

Such a tool could evolve into a personalized wellness companion, offering meaningful feedback based on ongoing behavior.

6. Longitudinal and Seasonal Analysis

By extending the tracking window from one month to several months—or even a full year—we could uncover seasonal health trends, behavioral cycles, and long-term improvements. This level of analysis is particularly valuable for identifying the sustainability of lifestyle changes or detecting early signs of fatigue or health deterioration.

7. Social and Comparative Insights

In future versions, benchmarking features could be added to compare personal progress with peers or population averages. This can drive motivation and accountability through:

1) Social leaderboards
2) Gamified activity challenges
3) Insights into how similar individuals manage activity and recovery

9. Conclusion

This project demonstrates the potential of even basic fitness tracker data to provide deep insights into personal health behaviors. By analyzing and visualizing physical activity, sleep, and cardiovascular markers, we gain a powerful lens into daily habits and how they affect overall wellness.

With continued monitoring and more comprehensive data, such systems could evolve into personalized wellness assistants, capable of nudging healthier decisions, predicting burnout, and enhancing quality of life.

Appendix A : Source Code Snippets

analysis.py

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import os

# Load data
df = pd.read_csv("fitness_data.csv", parse_dates=["Date"])

# Ensure output directory exists
os.makedirs("outputs", exist_ok=True)

# Line Plot: Steps and Sleep
plt.figure(figsize=(10, 5))
sns.lineplot(x="Date", y="Steps", data=df, label="Steps")
sns.lineplot(x="Date", y="SleepHours", data=df, label="Sleep Hours")
plt.title("Daily Steps and Sleep Hours")
plt.xlabel("Date")
plt.ylabel("Count")
plt.legend()
plt.tight_layout()
plt.savefig("outputs/steps_sleep.png")

# Correlation Heatmap
plt.figure(figsize=(8, 6))
corr = df.drop(columns=["Date"]).corr()
sns.heatmap(corr, annot=True, cmap="coolwarm")
plt.title("Correlation Between Fitness Metrics")
plt.tight_layout()
plt.savefig("outputs/correlation_heatmap.png")

# Line Plot: Resting Heart Rate
```

```python
plt.figure(figsize=(10, 5))
sns.lineplot(x="Date", y="RestingHeartRate", data=df, marker="o")
plt.title("Resting Heart Rate Over Time")
plt.xlabel("Date")
plt.ylabel("BPM")
plt.tight_layout()
plt.savefig("outputs/heart_rate.png")

# Bar Chart: Average Active Minutes per Week
df['Week'] = df['Date'].dt.isocalendar().week
weekly_avg = df.groupby('Week')['ActiveMinutes'].mean().reset_index()
plt.figure(figsize=(8, 5))
sns.barplot(x='Week', y='ActiveMinutes', data=weekly_avg)
plt.title("Average Active Minutes per Week")
plt.xlabel("Week Number")
plt.ylabel("Average Active Minutes")
plt.tight_layout()
plt.savefig("outputs/bar_active_minutes_weekly.png")

# Pie Chart: Proportion of Calories Burned per Week
weekly_cal = df.groupby('Week')['CaloriesBurned'].sum()
plt.figure(figsize=(6, 6))
weekly_cal.plot.pie(autopct='%1.1f%%')
plt.title("Weekly Calories Burned Distribution")
plt.ylabel("")
plt.tight_layout()
plt.savefig("outputs/pie_calories_burned.png")

# Scatter Plot: Sleep vs Resting Heart Rate
plt.figure(figsize=(8, 5))
sns.scatterplot(x="SleepHours", y="RestingHeartRate", data=df)
plt.title("Sleep Hours vs. Resting Heart Rate")
plt.xlabel("Sleep Hours")
plt.ylabel("Resting Heart Rate")
plt.tight_layout()
plt.savefig("outputs/scatter_sleep_heart_rate.png")

print("All visualizations generated successfully.")
```

Appendix B : Links

1) **Github repository link :** https://github.com/anandasaikiacse/fitness-tracker-data-analysis
2) **Github dataset link :** https://github.com/anandasaikiacse/fitness-tracker-data-analysis/blob/main/fitness_data.csv