Health Data Analysis Report

1. Introduction & Objectives

This project leverages personal health data exported from Apple HealthKit to investigate walking/running habits, specifically focusing on:

Step Count

Walking/Running Distance

Step Length

Active Energy Burned

Through daily and monthly aggregations, as well as correlation and regression analysis, the project aims to answer the following questions:

How do daily step counts and distances vary over time?

What relationships exist between step length, distance, and total energy burned?

Are there noticeable trends or improvements in monthly totals for steps, distance, and active energy?

What gaps or anomalies in the data might impact interpretation?

2. Data Overview

2.1 Sources & Timeframe

Apple Health Export (XML) converted into separate CSV files for step count, distance, step length, and active energy.

Data spans 2020–2025, though coverage may vary (some months or days have missing records).

2.2 Tools

Jupyter Notebook, primary environment for data cleaning, exploratory data analysis (EDA), and documentation. Pandas, data wrangling, aggregation, and feature engineering. NumPy: numerical computations. Matplotlib & Seaborn: Visualization of trends, correlations, and distributions.

2.2 Key Metrics

Daily Step Count (Steps)

Daily Walking/Running Distance (km)

Daily Average Step Length (cm)

Daily Active Energy Burned (kcal)

Derived Metrics:

Steps per km = total steps / total distance

kcal per step = total kcal / total steps

kcal per km = total kcal / total distance

3. Daily Trends

3.1 Daily Step Count Trends

Summary:

Step counts generally range from 800 to 1,300 per day, with some higher peaks.

Notable gaps around mid-2022 reflect periods when the device wasn't with me.

Interpretation: Sustained daily steps suggest relatively consistent walking habits. Data gaps must be accounted for when assessing long-term averages.

3.2 Daily Walking/Running Distance

Key Observations:

Most values cluster between ~0.3 km and 1.0 km, with spikes above 1.5 km.

Gaps around mid-2022 mirror those in step count, again indicating missing or untracked days.

Implication: Regular daily walking/running suggests moderate exercise routines. Spikes may correlate with deliberate workouts or events.

3.3 Daily Average Step Length

Range: Typically 40–90 cm, with some outliers above 120 cm.

Shifts: The data split (mid-2022) highlights missing records or device inactivity.

Possible Causes of Variation: Changes in footwear, terrain, or personal fitness can influence step length.

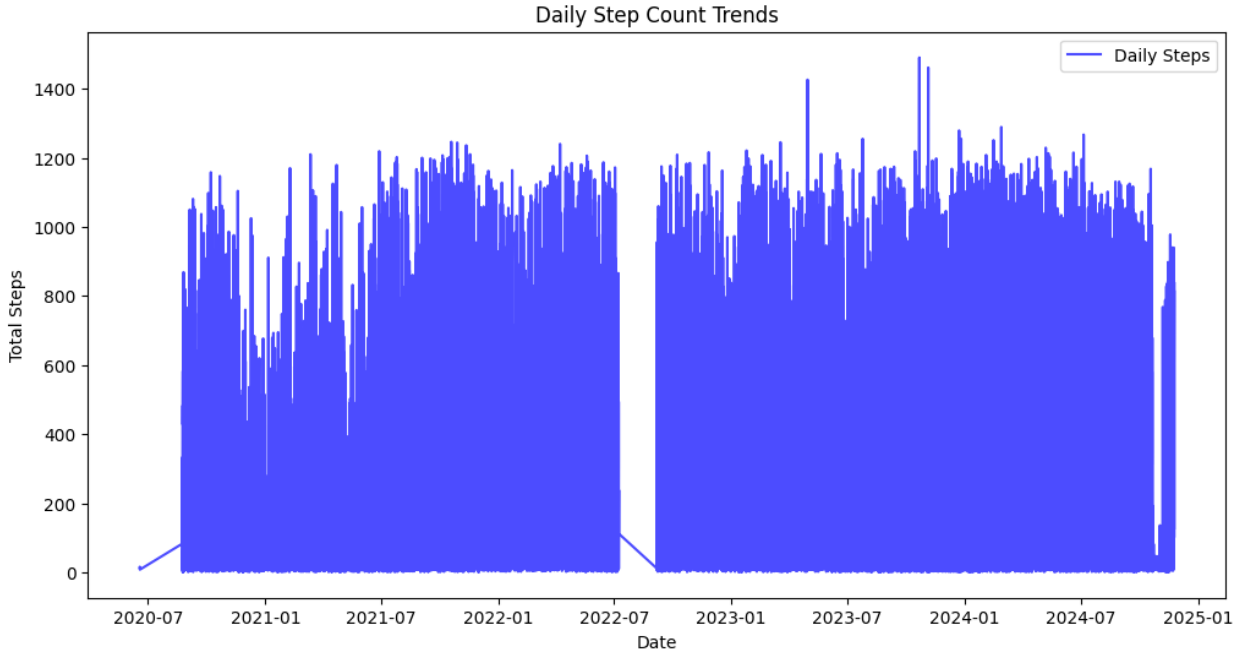
3.4 Daily Active Energy Burned

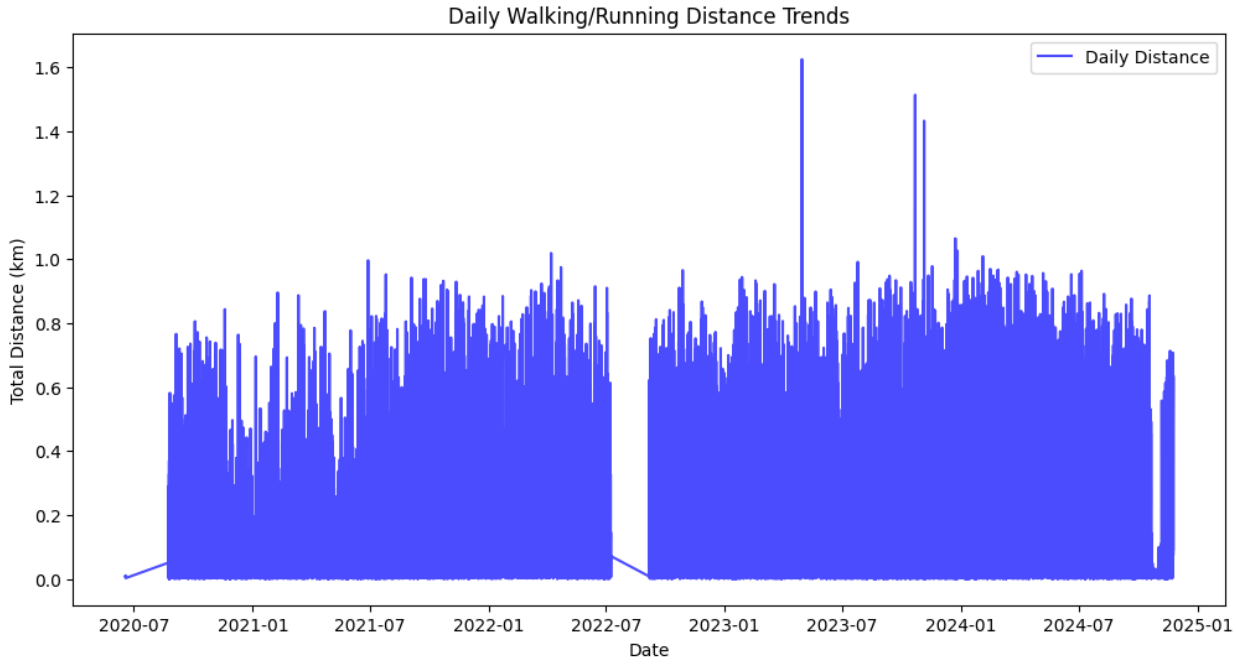
Observations:

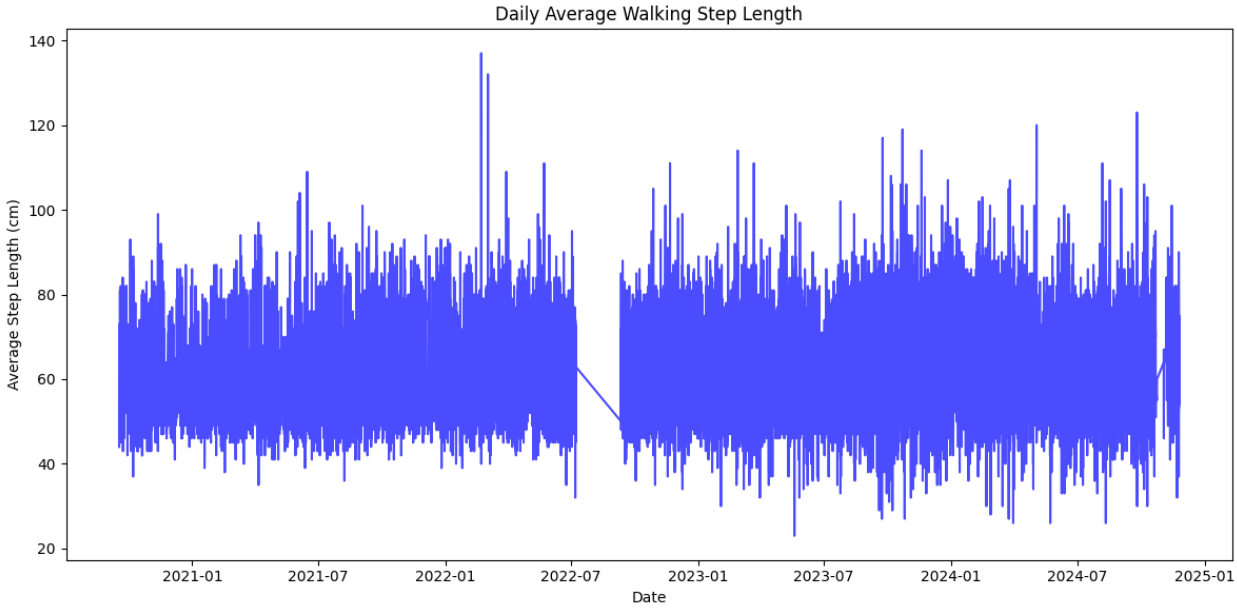
Values often hover around 20–60 kcal per day, with sporadic peaks above 80 kcal.

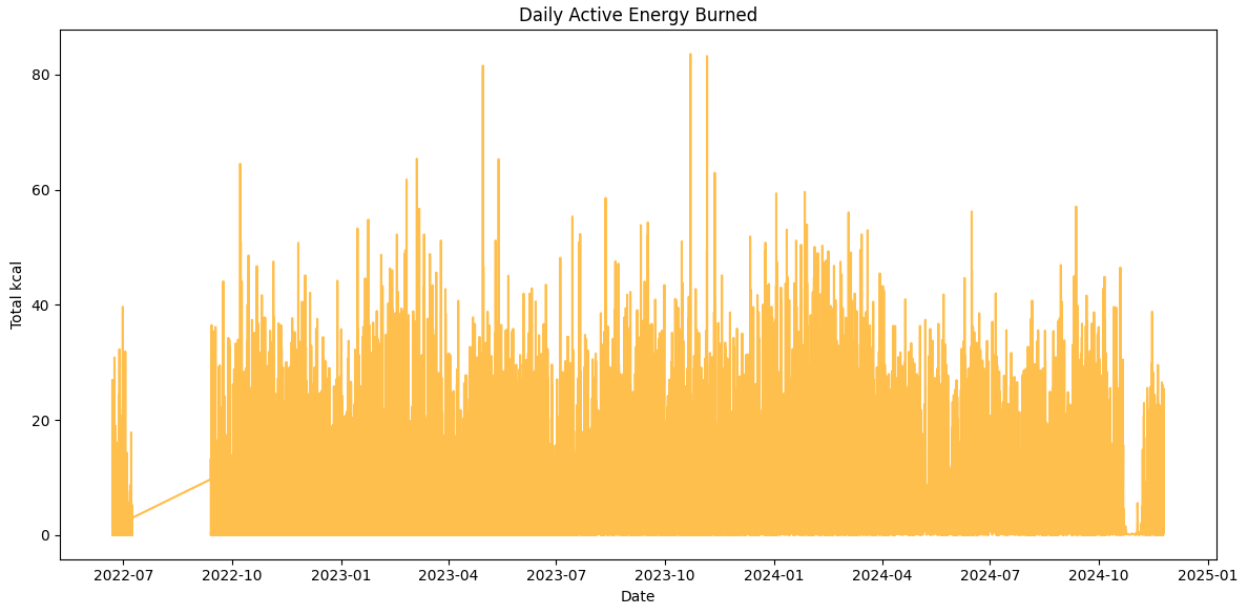
A noticeable break/inconsistency around mid-2022 aligns with the other metrics' data gap.

Interpretation: Higher spikes likely indicate particularly active days or additional exercise.




Daily Walking/Running Distance Trends
Daily Distance
Total Distance (km)
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
2020-07
2021-01
2021-07
2022-01
2022-07
2023-01
2023-07
2024-01
2024-07
2025-01
Date


Daily Average Walking Step Length
Average Step Length (cm)
20
40
60
80
100
120
140
2021-01
2021-07
2022-01
2022-07
2023-01
2023-07
2024-01
2024-07
2025-01
Date



4. Monthly Analyses

4.1 Monthly Total Steps

Trend:

A steady climb from below 100k steps in early months to nearly 500k steps in later months.

Occasional dips could be due to travel, health issues, or incomplete device usage.

Takeaway: Rising step totals suggest a general increase in walking frequency or duration over time.

4.2 Monthly Total Walking/Running Distance

Observations:

Early months show <50 km, while peak months exceed 300 km, indicating a substantial uptick in activity.

Minor plateaus or declines hint at lifestyle or seasonal disruptions (e.g., busy work, weather changes).

Practical Insight: Monitoring distance is a tangible measure of overall effort. Sustained growth in monthly distance underscores improved or extended walking routines.

4.3 Monthly Active Energy Burned

Key Points:

Sharp growth from a few thousand kcal per month to over 12,000 kcal at peak.

Suggests more frequent or intensive exercise regimens, especially in mid-to-late 2024.

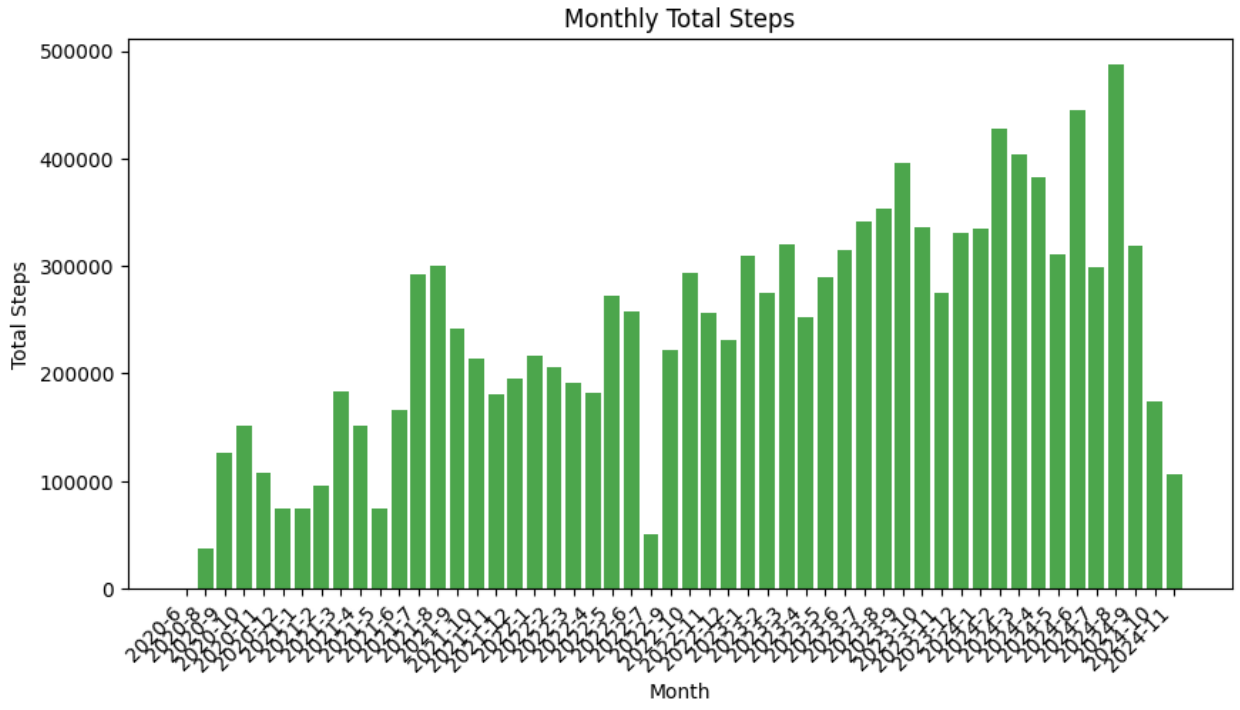
Context: Active energy is closely tied to both step count and intensity, reinforcing the idea of consistent workouts or longer daily walks.

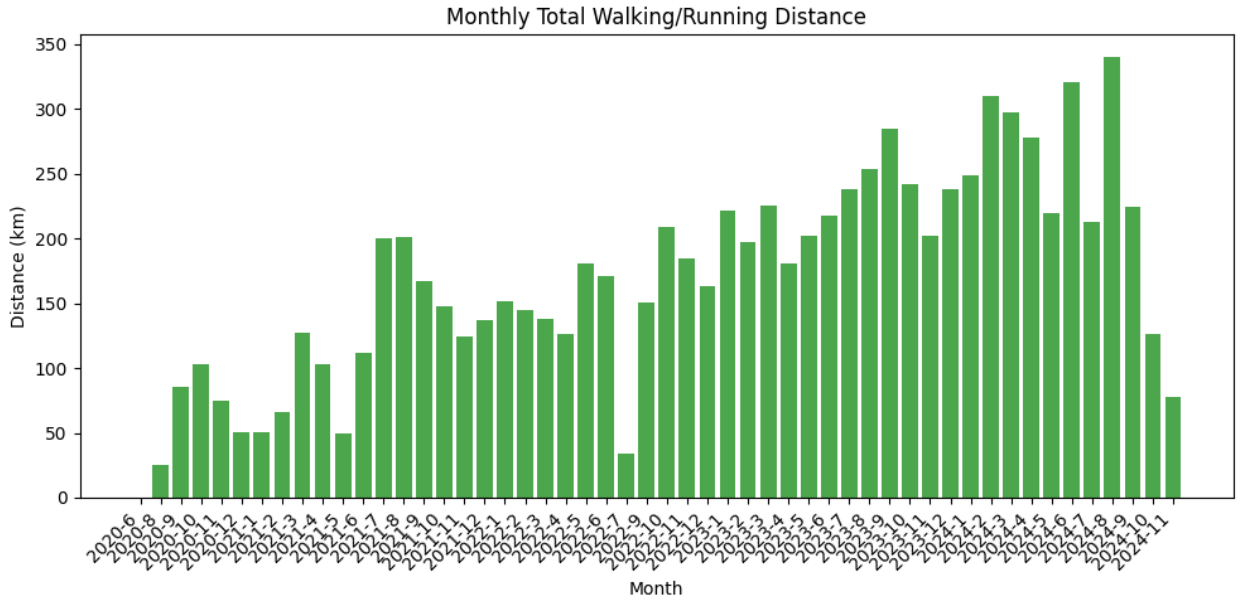
4.4 Monthly Average Walking Step Length

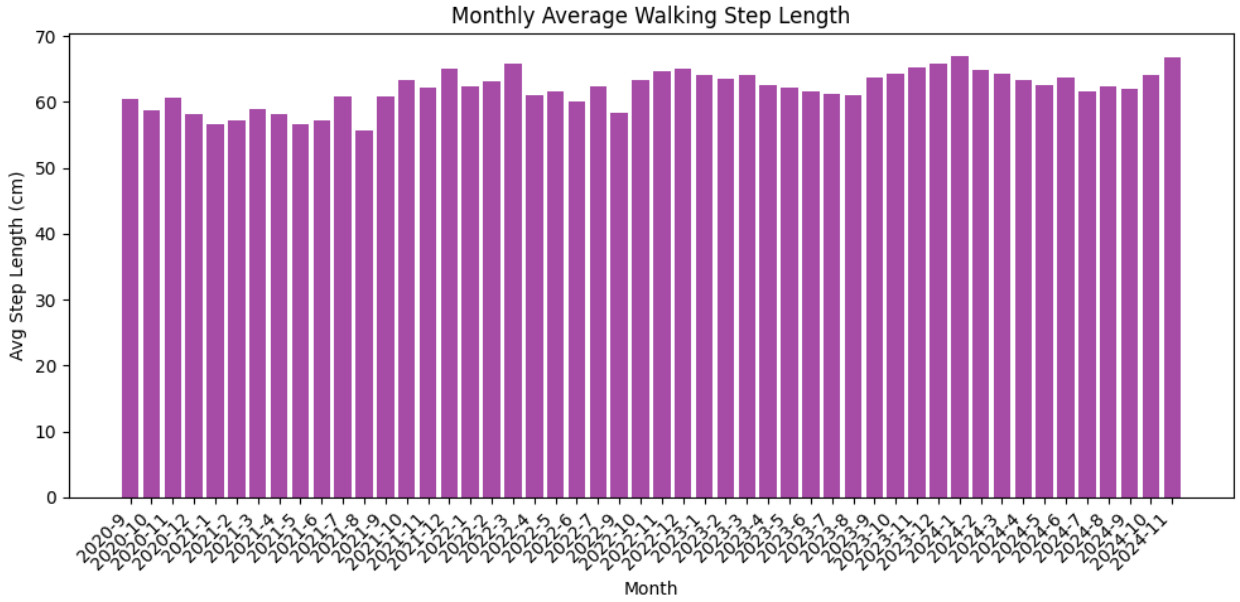
Range: Generally 55–65 cm, peaking around 70 cm.

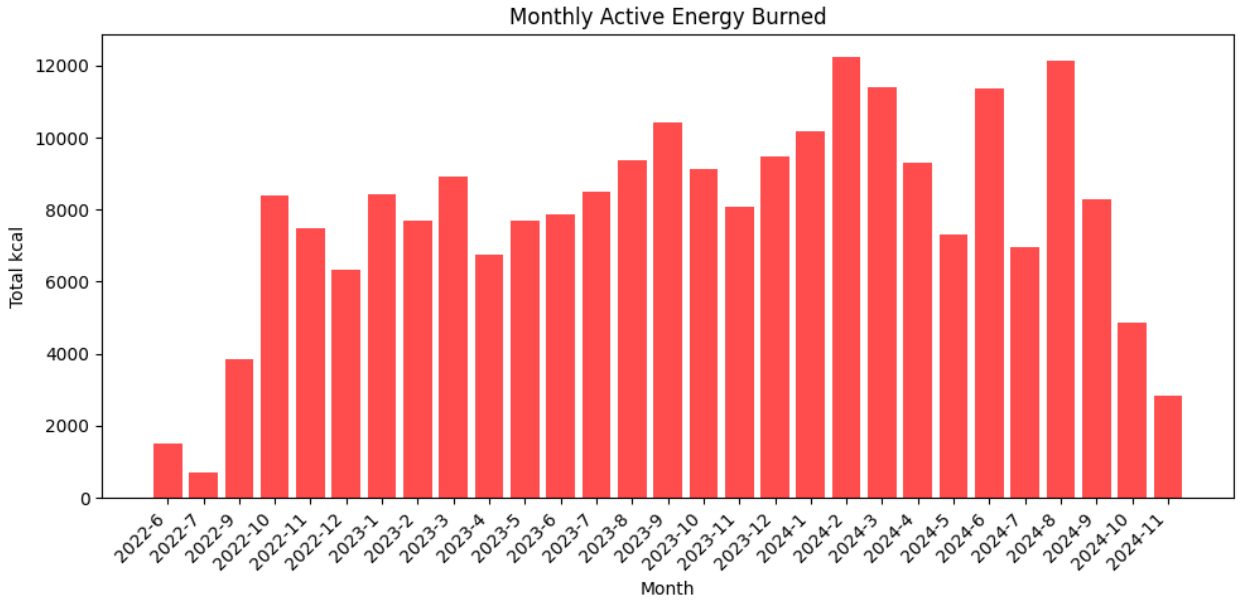
Observation: Step length has modest month-to-month fluctuations, with a gradual trend toward slightly longer strides.

Link to Efficiency: Longer step lengths often reduce total steps needed per km, potentially reflecting better walking mechanics.




Monthly Total Walking/Running Distance
Distance (km)
0
50
100
150
200
250
300
350
Month


Monthly Average Walking Step Length
Avg Step Length (cm)
0
10
20
30
40
50
60
70
Month



5. Correlation & Regression Analysis

5.1 Correlation Matrix

Highlights:

Total Steps and Total Distance are nearly perfectly correlated (~1.00).

Avg Step Length shows a negative correlation with Steps per km, consistent with longer strides reducing steps needed.

kcal per step strongly associates with Total kcal, indicating that on days of higher overall energy burn, each step is also more “costly.”

5.2 Regression: Step Length vs. Steps per km

Finding: Clear negative slope suggests that as step length increases, steps per km decrease.

Confidence Interval: The shaded region shows moderate confidence in the linear fit, though real-world walking can vary due to pace, terrain, or form.

5.3 Daily Step Length vs. Steps per km

What It Shows

X-axis: Average Step Length (cm) per day.

Y-axis: Steps per km (i.e., how many steps are taken to walk/run 1 km).

Key Observations

Negative Trend: The cluster of points generally slopes downward, when average step length is higher, the total steps needed per kilometer is lower.

Range of Values: Step length ranges from roughly 35 cm (very short strides) to over 80 cm (quite long strides). Steps per km vary from about 1,200 to nearly 1,800 steps per km.

Possible Interpretations

Longer Strides, Fewer Steps: This aligns with the intuitive idea that if I take bigger steps, I cover more ground with each step, thus needing fewer steps to walk the same distance.

5.4 Monthly Activity Trends

What It Shows

X-axis: Sequential months of data (labeled as 0, 10, 20, etc., though they likely correspond to actual calendar months/years).

Y-axis: Aggregated values (steps, distance in km, active energy in kcal) for each month.

Lines:

Blue Solid Line: Total Steps (Monthly).

Orange Dashed Line: Total Distance (Monthly).

Green Dotted Line: Total Energy Burned (Monthly).

Key Observations

Significant Growth in Steps: The blue line starts near zero and climbs above 400,000 steps per month, indicating a substantial increase in monthly walking activity over time.

Minimal Distance & Energy Lines: The orange and green lines appear small in comparison—likely because steps are in the hundreds of thousands, whereas distance (in km) and total energy (in kcal) are on a different numerical scale.

Spikes and Dips: Some months have steep peaks or drops in step count. These could correlate with seasonal changes, holidays.

Possible Interpretations

Consistency & Trend: Overall, there's a positive trajectory in monthly steps. Even if distance or energy lines look small on the same scale, they likely also show an increasing trend if plotted separately.

5.5 Pair Plot of Key Activity Metrics

What It Shows

Diagonal Plots: Distributions (histograms or density plots) for each individual metric (e.g., total_steps, total_distance_km, total_kcal, avg_step_length_cm, steps_per_km, kcal_per_step, kcal_per_km).

Off-Diagonal Scatter Plots: Pairwise relationships between each pair of metrics.

Key Observations

Diagonals:

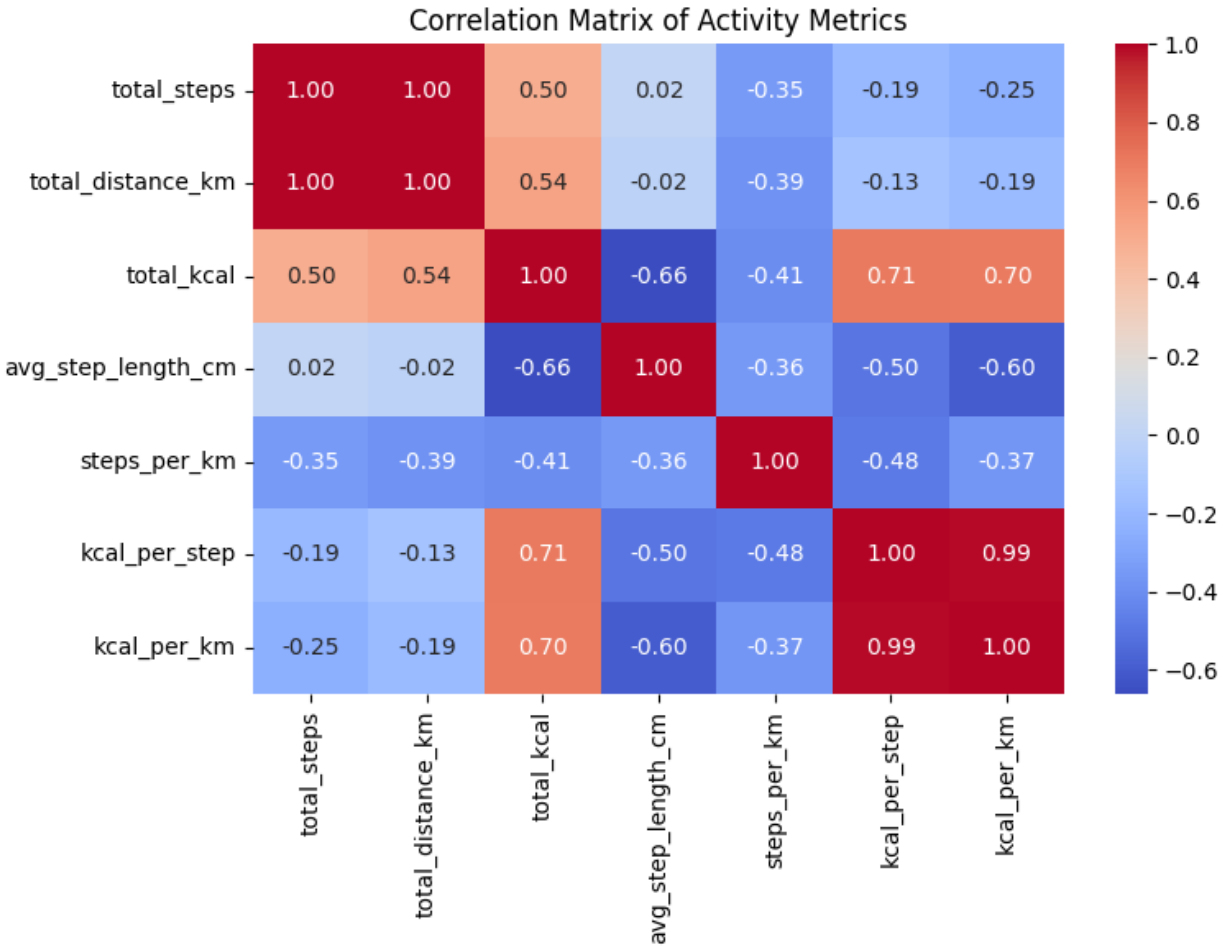
Reveal how each metric is distributed (e.g., is it skewed to the right, mostly clustered around one range, etc.).

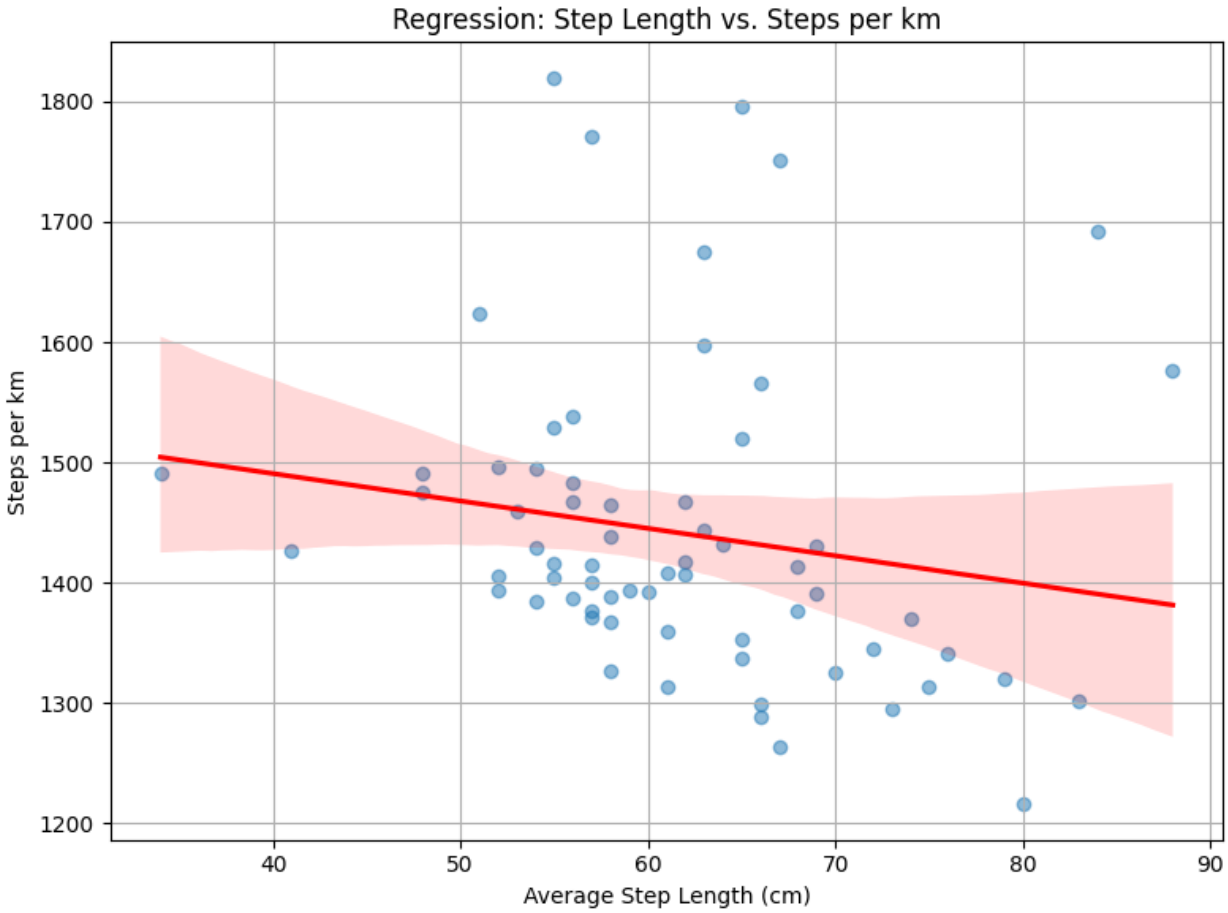
Scatter Points:

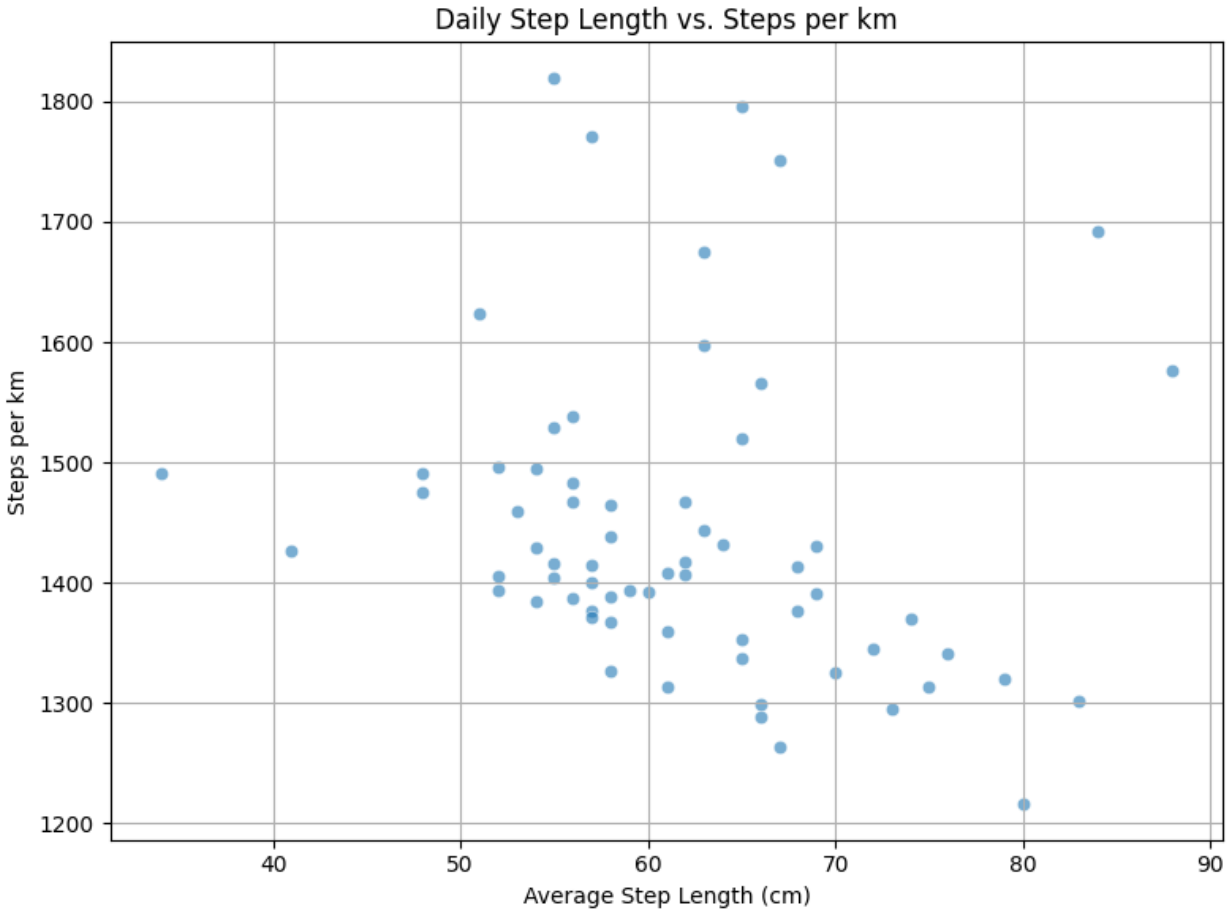
total_steps vs. total_distance_km might show a near-linear relationship if I consistently walk the same stride length.

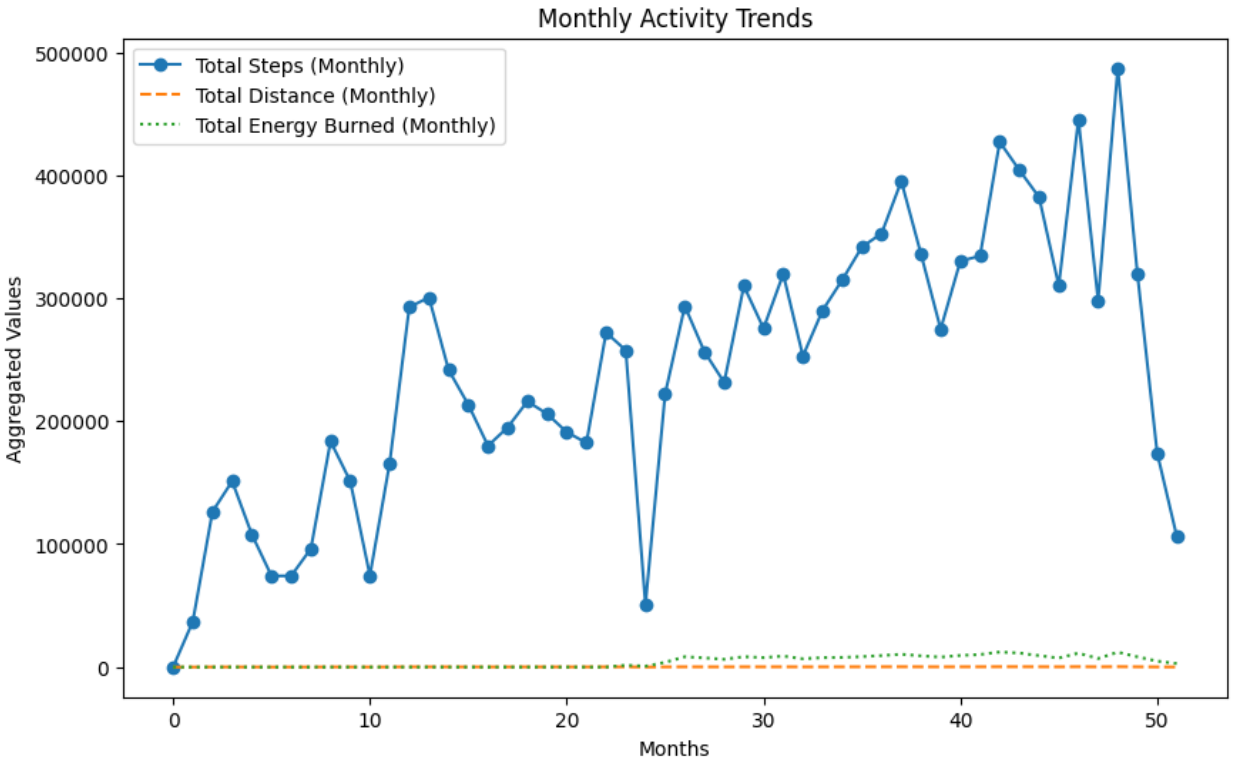
avg_step_length_cm vs. steps_per_km typically shows a negative slope.

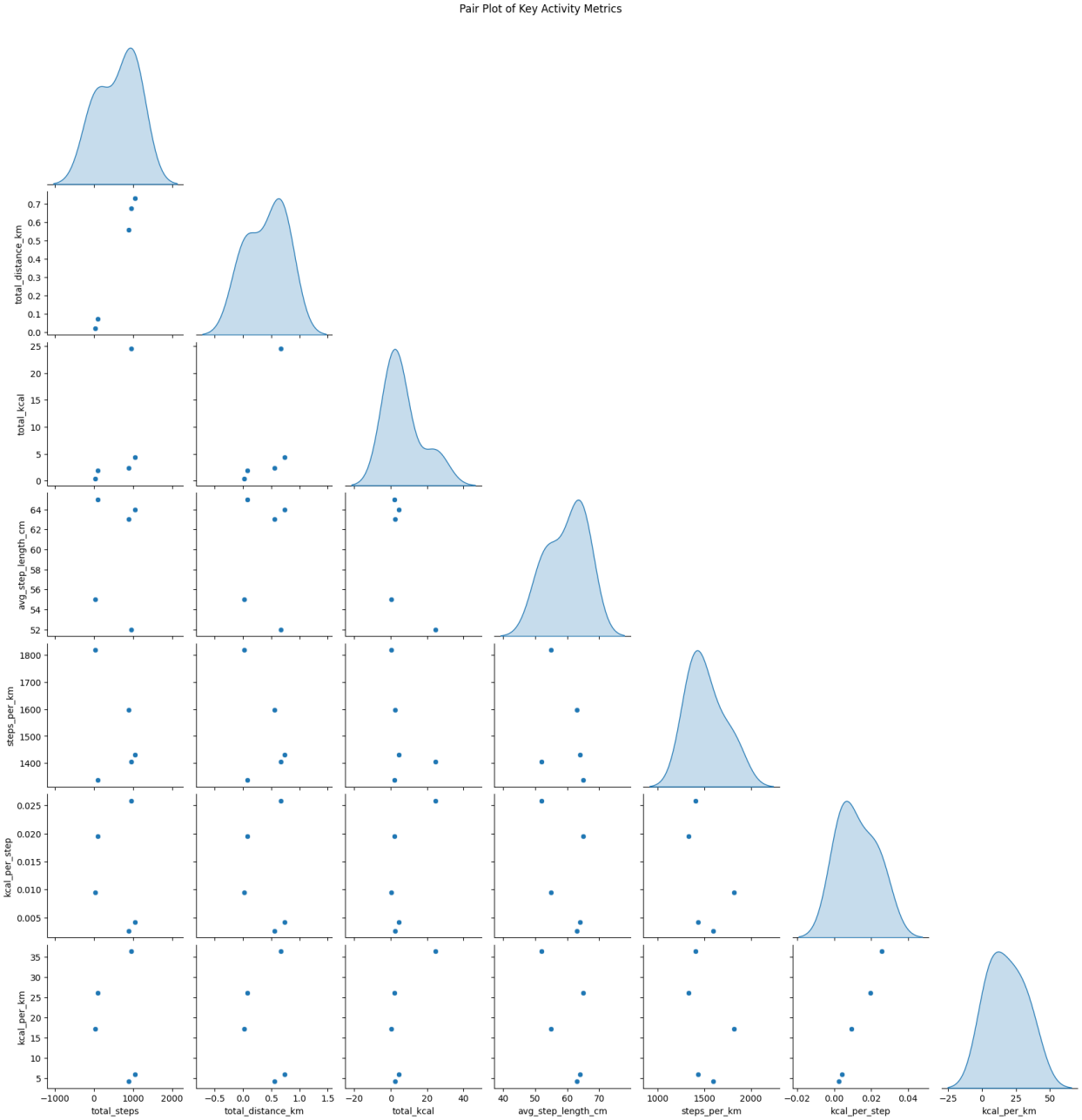
total_kcal vs. kcal_per_step might show strong positive correlation, days with higher total kcal also have higher kcal expended per step.


Correlation Matrix of Activity Metrics
total_steps
total_distance_km
total_kcal
avg_step_length_cm
steps_per_km
kcal_per_step
kcal_per_km
1.00 1.00 0.50 0.02 -0.35 -0.19 -0.25
1.00 1.00 0.54 -0.02 -0.39 -0.13 -0.19
0.50 0.54 1.00 -0.66 -0.41 0.71 0.70
0.02 -0.02 -0.66 1.00 -0.36 -0.50 -0.60
-0.35 -0.39 -0.41 -0.36 1.00 -0.48 -0.37
-0.19 -0.13 0.71 -0.50 -0.48 1.00 0.99
-0.25 -0.19 0.70 -0.60 -0.37 0.99 1.00
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6


Regression: Step Length vs. Steps per km
Steps per km
Average Step Length (cm)
1800
1700
1600
1500
1400
1300
1200
40
50
60
70
80
90


Daily Step Length vs. Steps per km
Steps per km
1800
1700
1600
1500
1400
1300
1200
40
50
60
70
80
90
Average Step Length (cm)


Monthly Activity Trends
Total Steps (Monthly)
Total Distance (Monthly)
Total Energy Burned (Monthly)
Aggregated Values
Months
500000
400000
300000
200000
100000
0
0
10
20
30
40
50


Pair Plot of Key Activity Metrics
total_steps
total_distance_km
total_kcal
avg_step_length_cm
steps_per_km
kcal_per_step
kcal_per_km

6. Key Findings

Increasing Activity Over Time: Monthly metrics (steps, distance, and energy) generally show a positive outline.

Impact of Missing Data: Mid-2022 and other isolated gaps highlight that “zero” readings often reflect untracked days.

Stride Efficiency: The inverse relationship between average step length and steps per km confirms that longer strides can reduce total steps needed over a given distance.

Energy Variability: Periodic spikes in daily or monthly active energy suggest more intense exercise sessions.

7. Limitations

Missing Points: All data depends on use of the iPhone. That’s why some data is missing at ceratin points, because device was not with me.

8. Next Steps

Seasonal & Weekday Analysis

Break down monthly or daily data by season (Winter, Spring, Summer, Autumn) or weekday/weekend to uncover deeper patterns.

Integrate More Variables

Combine heart rate, sleep patterns, or caloric intake data to build a more holistic profile of health and fitness.

9. Conclusion

This analysis demonstrates a clear improvement in walking activity, from step count and distance to energy burn, over the 4 year period. The relationship between longer step length and fewer steps per km supports the importance of efficiency. While device gaps complicate some findings, the general upward trend in monthly metrics provides a positive look on exercise habits and suggests room for continued progress with consistent tracking.