

Bellabeat Case Study — Data Analysis Report

Reza Mahin Mohammadizadeh, Master's in Data Science, University of Padova

2025-12-02

Contents

1	Introduction	1
2	Data	2
2.1	Data Loading	2
2.2	Data Sources	2
2.3	Data Overview and Validation	2
3	Exploratory Data Analysis	3
3.1	Hourly-Level Analysis	3
3.1.1	Distribution of Hourly Step Count	3
3.1.2	Steps vs Calories	4
3.1.3	Steps vs Intensity	5
3.1.4	Average Hourly Activity Pattern	6
3.2	Daily-Level Analysis	6
3.2.1	Daily Step Count Distribution	7
3.2.2	Daily Calories Burned Distribution	8
3.2.3	Daily Steps vs Calories Burned	9
3.2.4	Daily Steps vs Sedentary Minutes	10
3.2.5	Distribution of Daily Activity Minutes	11
3.2.6	Sleep Duration Distribution	12
3.2.7	Sleep Duration vs Daily Steps	13
3.2.8	Average Steps by Weekday	14
4	Insights and Business Recommendations	14
4.1	Key Insights	14
4.2	Business Recommendations	15
5	Conclusion	15

1 Introduction

Bellabeat is a high-tech wellness company that manufactures smart products designed to empower women through health data.

This case study explores **Fitbit user data** to uncover activity, sleep, and heart-rate behavior patterns. By analyzing daily and hourly data, we aim to identify **how users engage with their devices**, discover trends in physical activity and rest, and generate **data-driven recommendations** to improve Bellabeat's marketing and product strategy.

2 Data

2.1 Data Loading

We use the cleaned datasets created during the data-wrangling phase.

The `daily_merged` dataset summarizes daily user metrics such as steps, calories, and sleep, while the `hourly_merged` dataset contains per-hour measures of activity intensity and heart rate.

```
# Move one level up from /Scripts, then into Data/Processed/
daily_merged <- read_csv("../Data/Processed/daily_merged.csv")
hourly_merged <- read_csv("../Data/Processed/hourly_merged.csv")

cat("Daily data rows:", nrow(daily_merged), "\n")

## Daily data rows: 457
cat("Hourly data rows:", nrow(hourly_merged), "\n")

## Hourly data rows: 24084
```

2.2 Data Sources

The data used for this analysis originates from **Fitbit Fitness Tracker Data**, publicly available on Kaggle. It contains anonymized activity, sleep, and weight logs collected from Fitbit users who voluntarily shared their data.

The original dataset includes multiple separate CSV files such as:

- `dailyActivity_merged.csv` – user-level daily step count, distance, and calories.

- `hourlySteps_merged.csv`, `hourlyCalories_merged.csv`, and `hourlyIntensities_merged.csv` – detailed hourly metrics of steps, energy expenditure, and activity intensity.
- `minuteSleep_merged.csv` – minute-level sleep tracking.
- `heartrate_seconds_merged.csv` – second-level heart rate data.
- `weightLogInfo_merged.csv` – manually logged weight entries.

For this project, these raw files were **cleaned, joined, and aggregated** to produce two main analytical datasets:

- `daily_merged`: combines daily activity, calories, and sleep data per user per day.

- `hourly_merged`: merges hourly step count, calories, intensity, and average heart rate.

These processed datasets provide a consistent structure suitable for trend analysis, correlation study, and data visualization.

2.3 Data Overview and Validation

Before conducting any analysis, we first perform a **structural and quality validation** of the processed datasets.

This includes inspecting variable types, counting unique participants, identifying missing values, and verifying date/time coverage to ensure data integrity.

Table 1: Summary of Processed Datasets

Dataset	Rows	Columns	Unique Users	Missing Values
daily_merged	457	25	35	3017
hourly_merged	24084	7	34	21552

Date range (daily): Mar 12, 2016 — Apr 12, 2016

Datetime range (hourly): Mar 12, 2016 — Apr 12, 2016

3 Exploratory Data Analysis

This section explores users' physical activity patterns at both **hourly** and **daily** levels to identify trends in movement, intensity, calories burned, and sleep behavior.

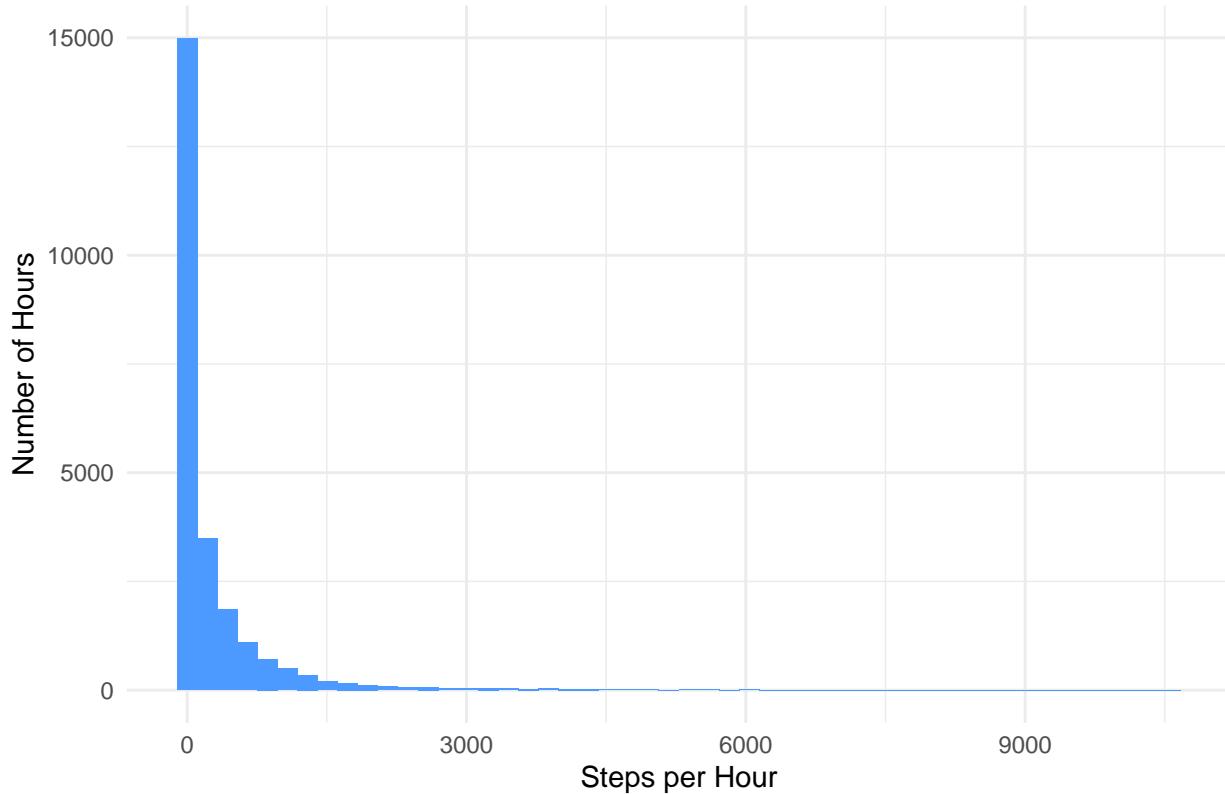
By examining correlations and temporal variations, we gain insights into how users interact with their wearable devices and maintain their activity throughout the day.

3.1 Hourly-Level Analysis

We first explore activity behavior at the **hourly scale** — how movement, intensity, and calories relate across the day.

3.1.1 Distribution of Hourly Step Count

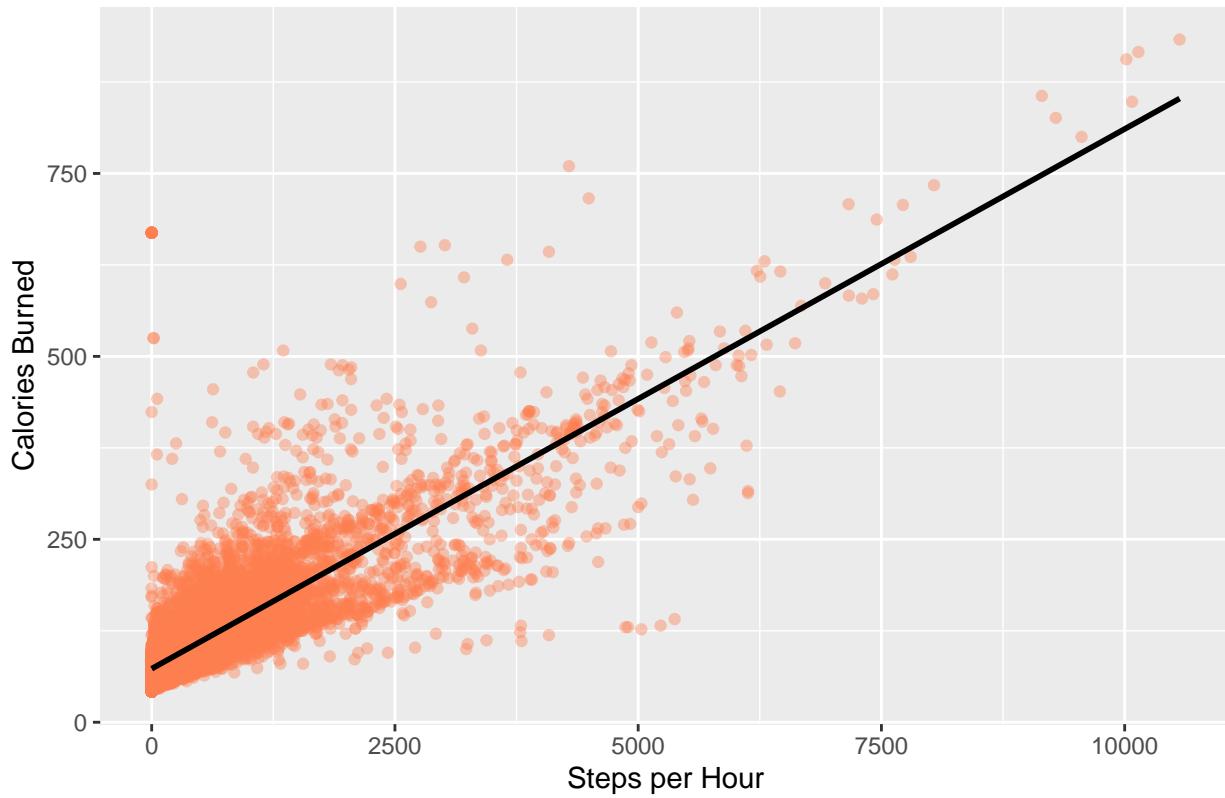
Distribution of Hourly Step Count



Interpretation: Most hourly step counts are close to zero, indicating that users spend many inactive hours during the day. Only a few hours show high activity (2,000+ steps), suggesting bursts of movement such as workouts or commuting periods.

3.1.2 Steps vs Calories

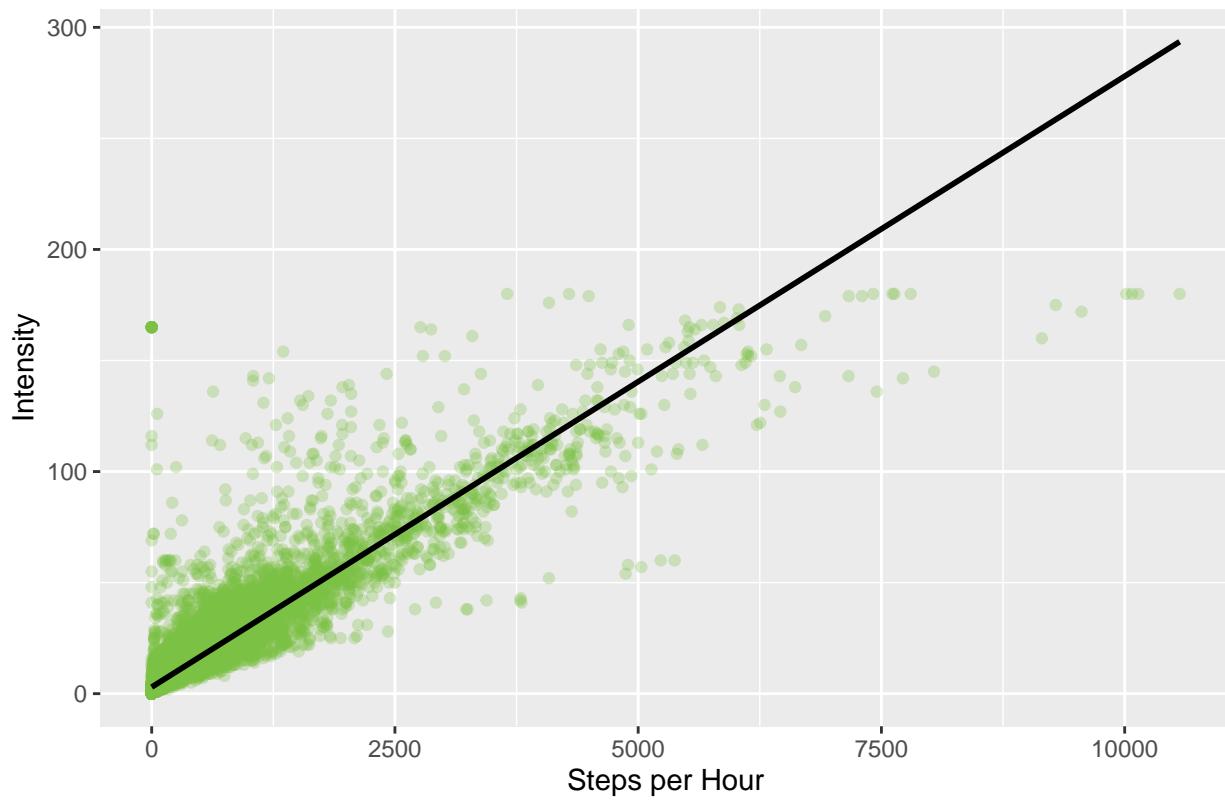
Steps vs Calories Burned (Hourly)



Interpretation: There is a clear positive linear relationship between steps and calories burned per hour. As the number of steps increases, the calories burned also rise, which confirms the logical link between physical activity and energy expenditure.

3.1.3 Steps vs Intensity

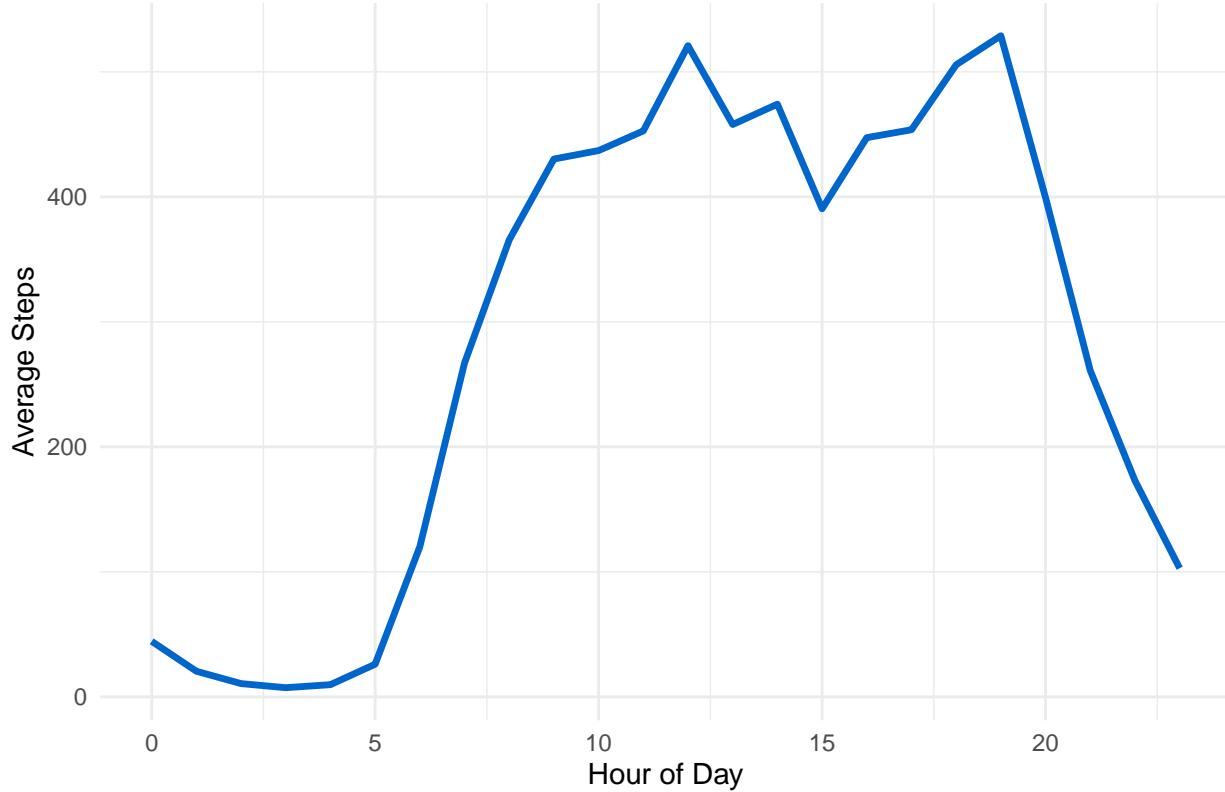
Steps vs Total Intensity (Hourly)



Interpretation: Step counts and total intensity also exhibit a positive correlation, indicating that more movement per hour results in higher recorded activity intensity.

3.1.4 Average Hourly Activity Pattern

Average Steps by Hour of Day



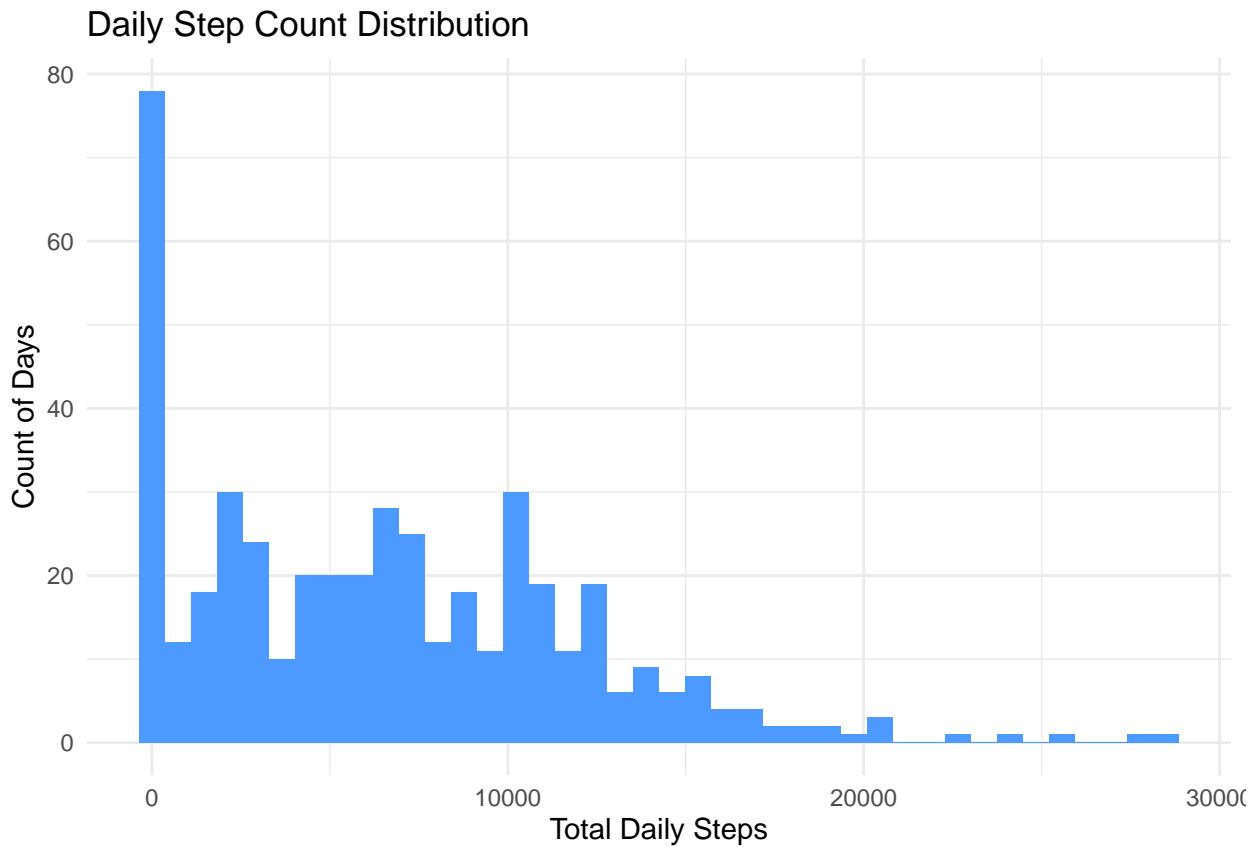
Interpretation: Users show clear daily activity cycles, with step counts rising sharply after 6 AM, peaking in late morning and early evening hours, and declining after 9 PM. This reflects common daily routines of commuting, exercise, and evening movement.

3.2 Daily-Level Analysis

We now shift from hourly activity to **daily trends** to identify long-term lifestyle patterns — examining total steps, calories burned, sedentary time, activity breakdown, and sleep duration.

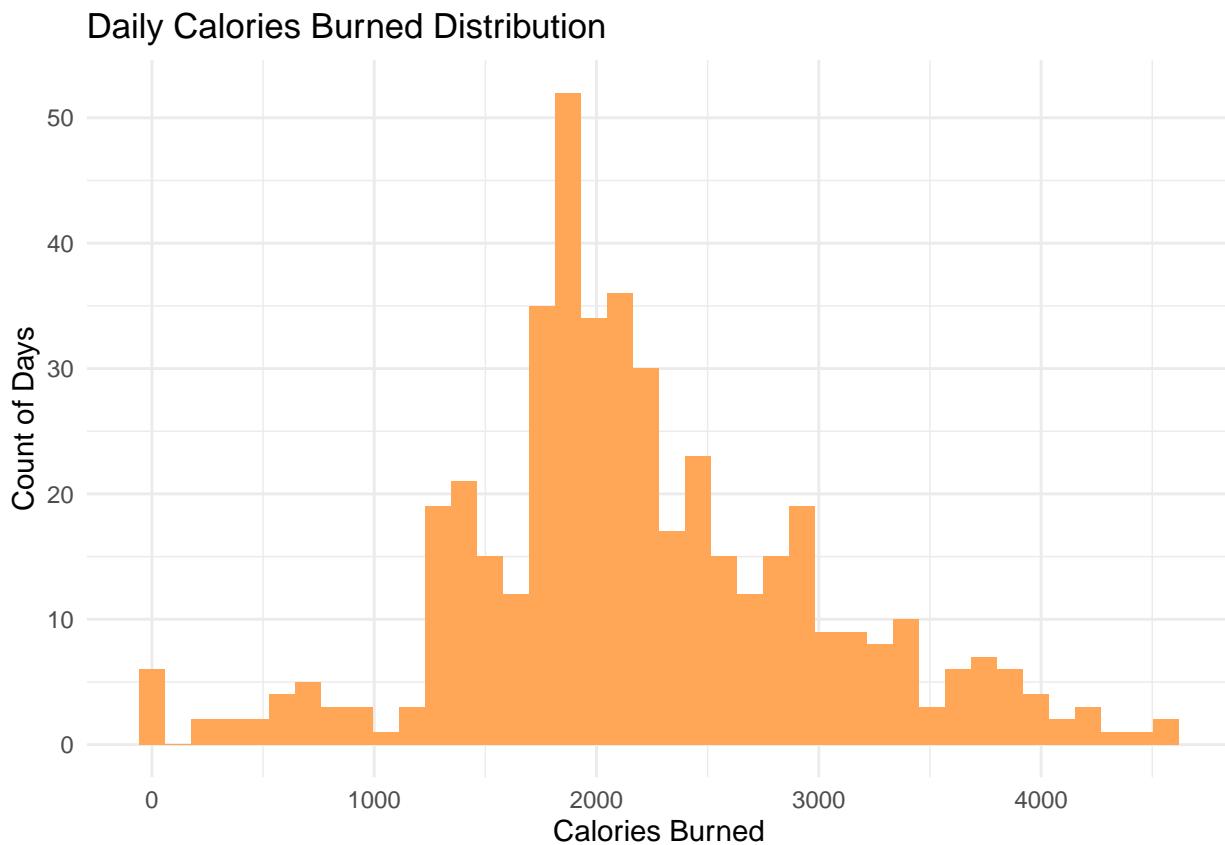
These visualizations help to understand overall behavioral habits and their relationships across different wellness dimensions.

3.2.1 Daily Step Count Distribution



Insight: The step distribution is heavily right-skewed — most users record fewer than 5,000–7,000 steps per day, while only a few achieve over 15,000. This suggests that consistent high activity is uncommon, and many users fall below the typical 10,000-step wellness benchmark.

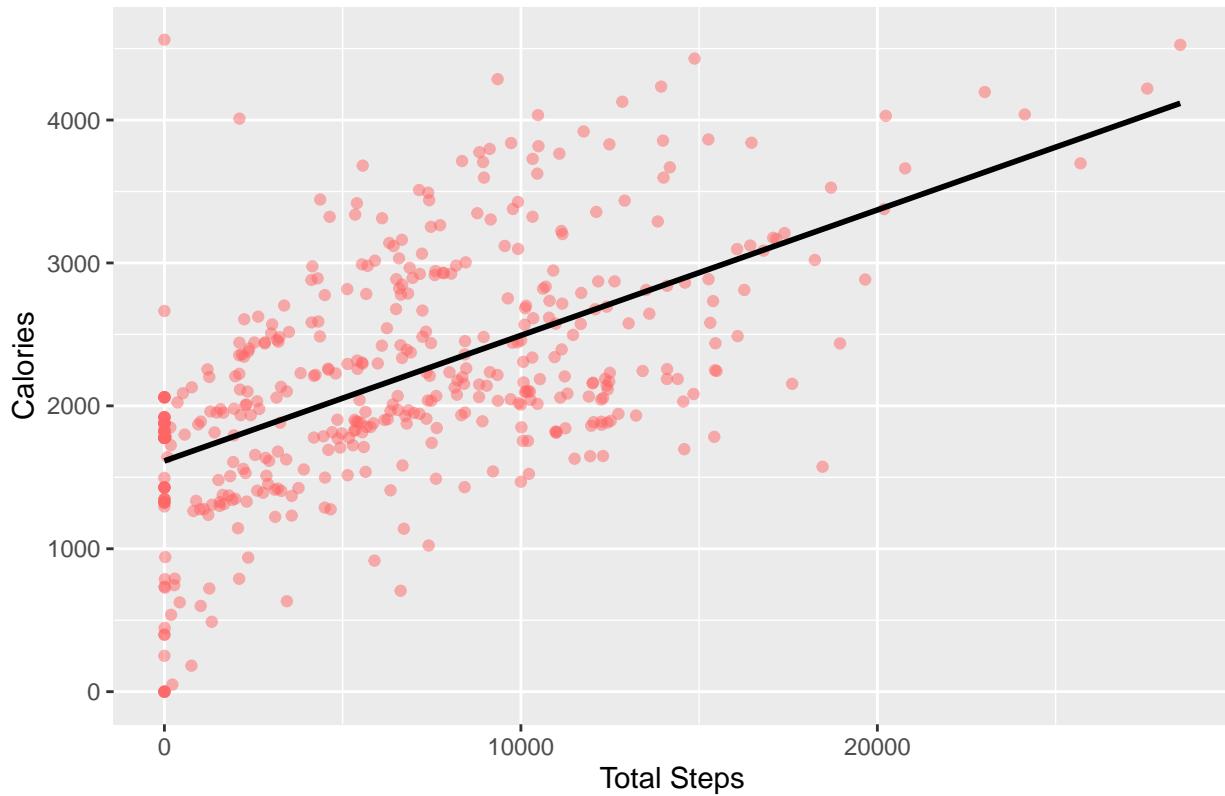
3.2.2 Daily Calories Burned Distribution



Insight: Calories burned follow a normal-like distribution centered around 2,000–2,500 kcal/day, which aligns with moderate lifestyle activity. Outliers on both sides reflect highly sedentary users and more active individuals performing workouts or longer movement periods.

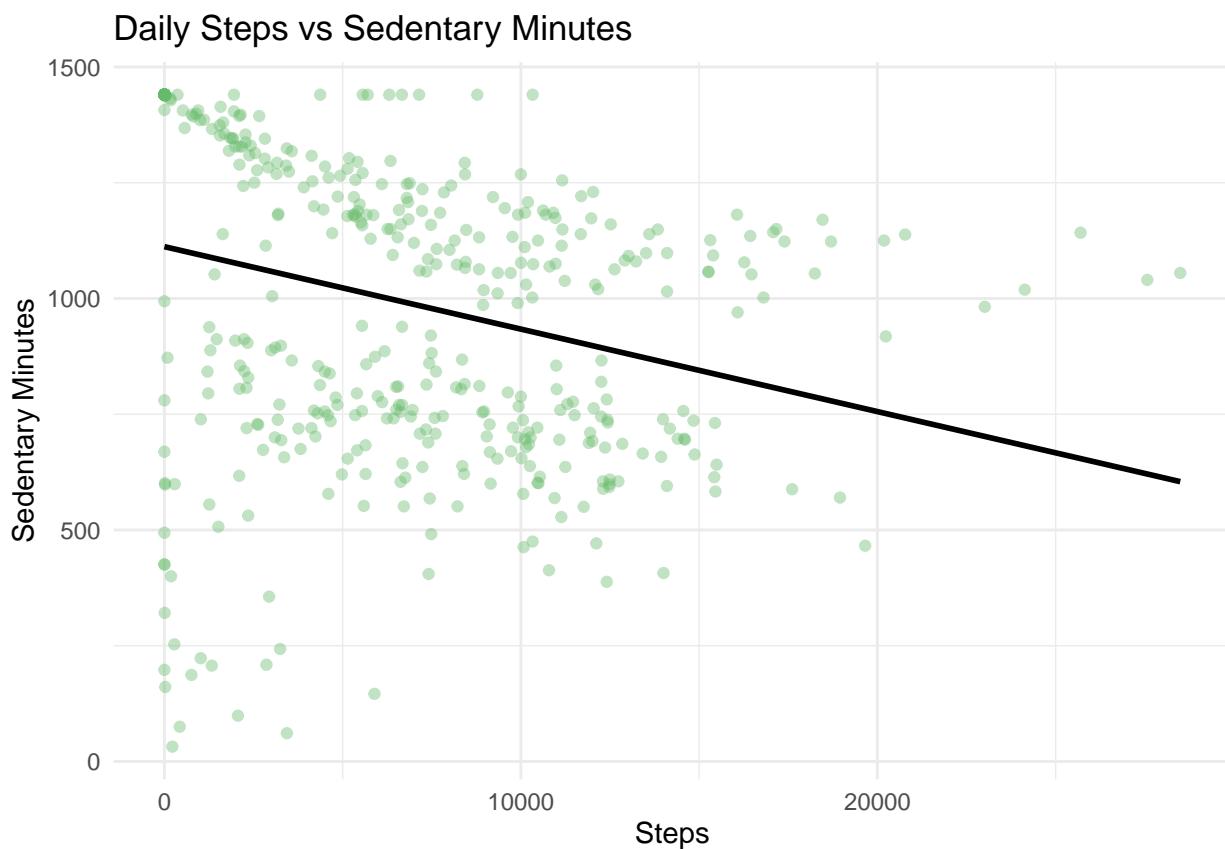
3.2.3 Daily Steps vs Calories Burned

Daily Steps vs Calories Burned



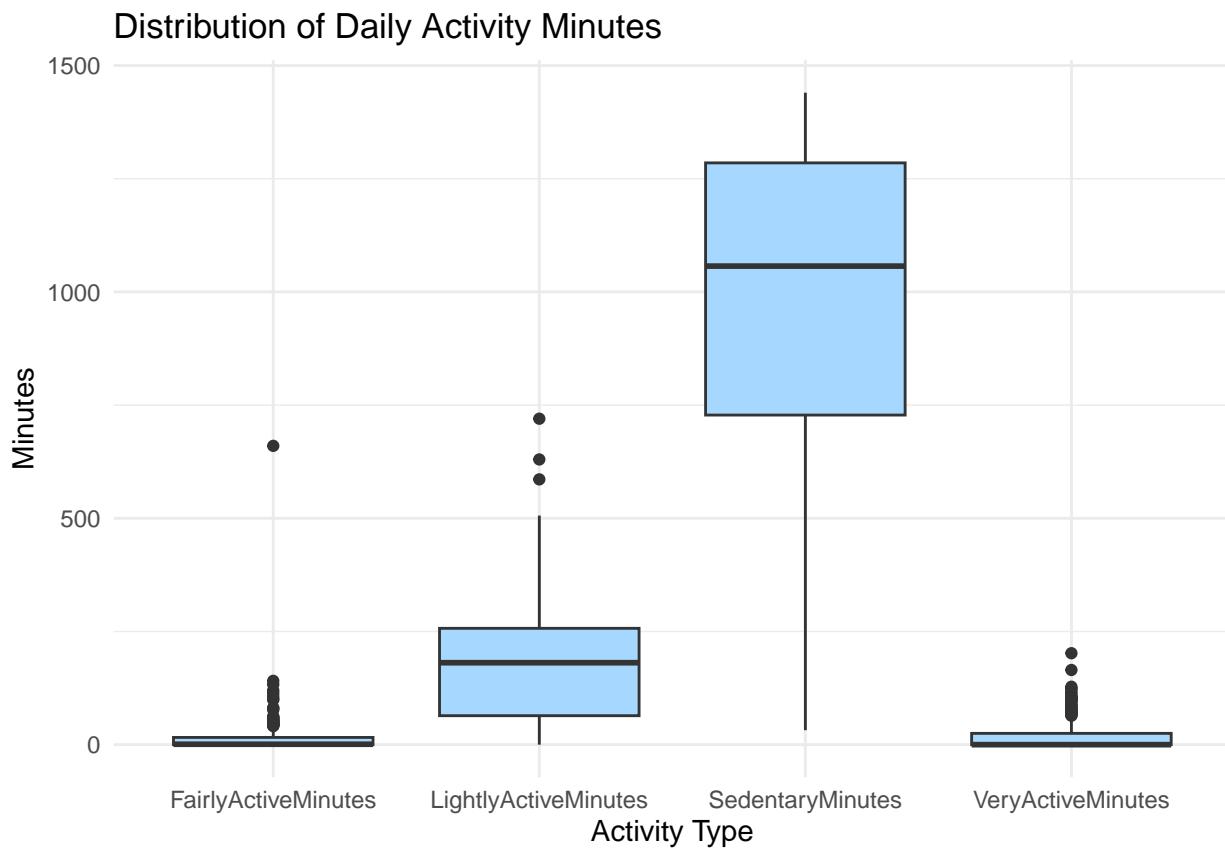
Insight: A strong positive linear trend exists between total daily steps and calories burned. This confirms that step count is a reliable predictor of daily energy expenditure.

3.2.4 Daily Steps vs Sedentary Minutes



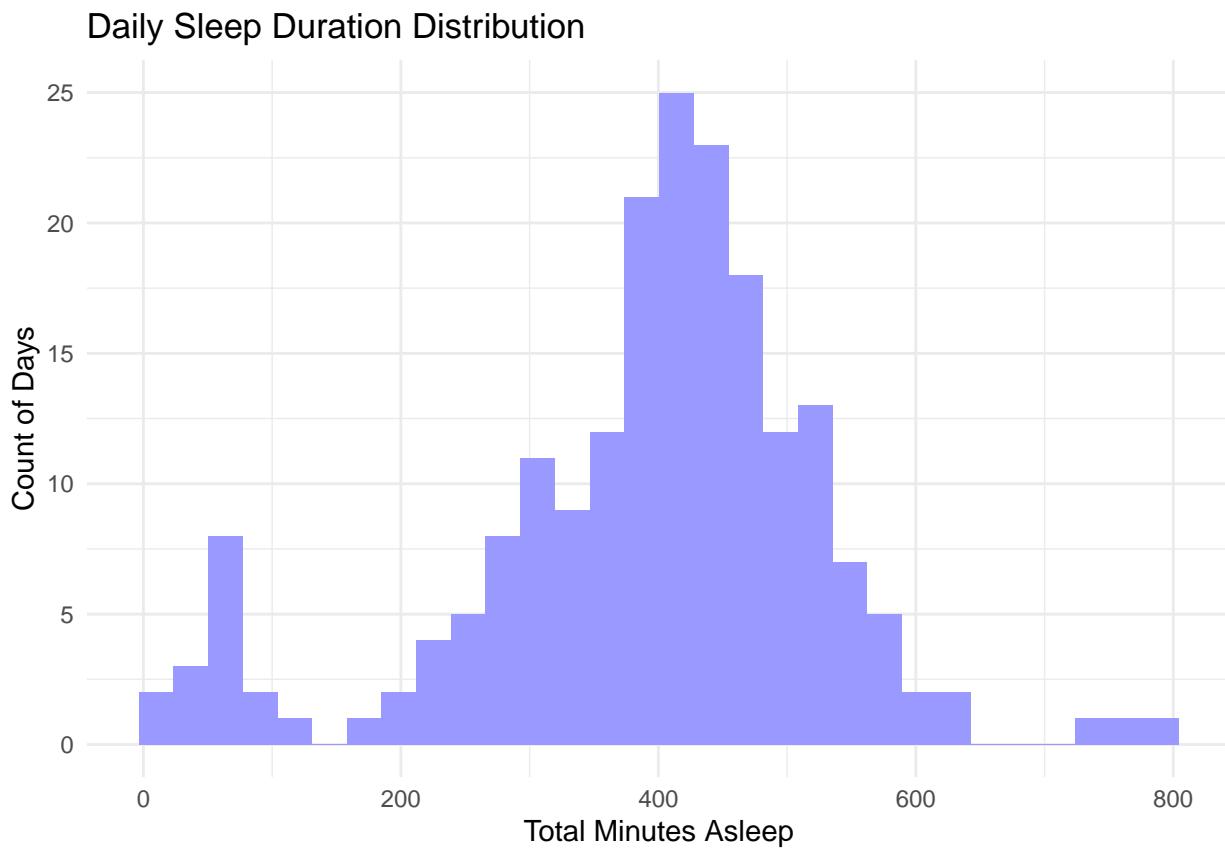
Insight: There is a clear negative relationship between step count and sedentary time — users who walk more tend to spend fewer minutes sitting or inactive. This aligns with expected physical activity patterns in balanced lifestyles.

3.2.5 Distribution of Daily Activity Minutes



Insight: Sedentary time dominates most days, with median values far exceeding any active categories. “Lightly Active” minutes show moderate engagement, whereas “Very Active” and “Fairly Active” minutes are minimal, reflecting limited high-intensity activity.

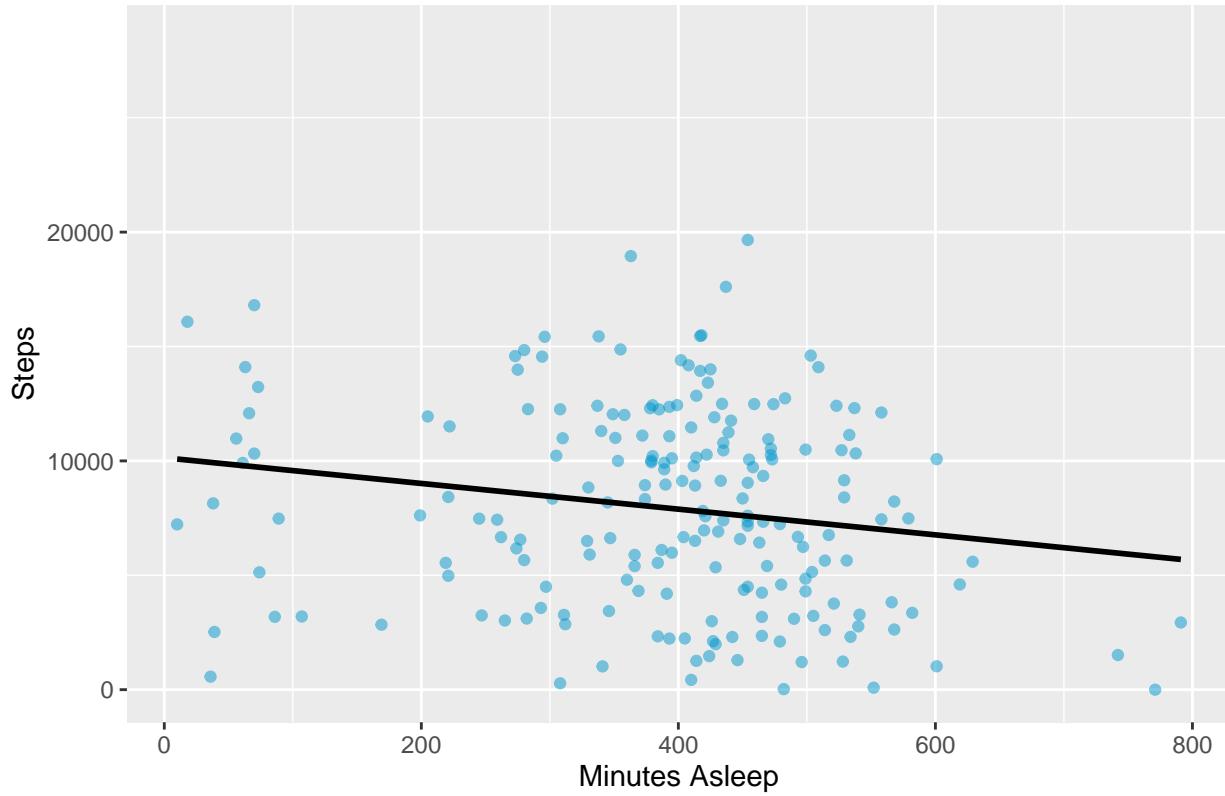
3.2.6 Sleep Duration Distribution



Insight: Most users report 6–8 hours of sleep per night, which fits within recommended sleep duration. However, there is notable variance, with several users showing inconsistent sleep tracking or short sleep durations (<5 hours).

3.2.7 Sleep Duration vs Daily Steps

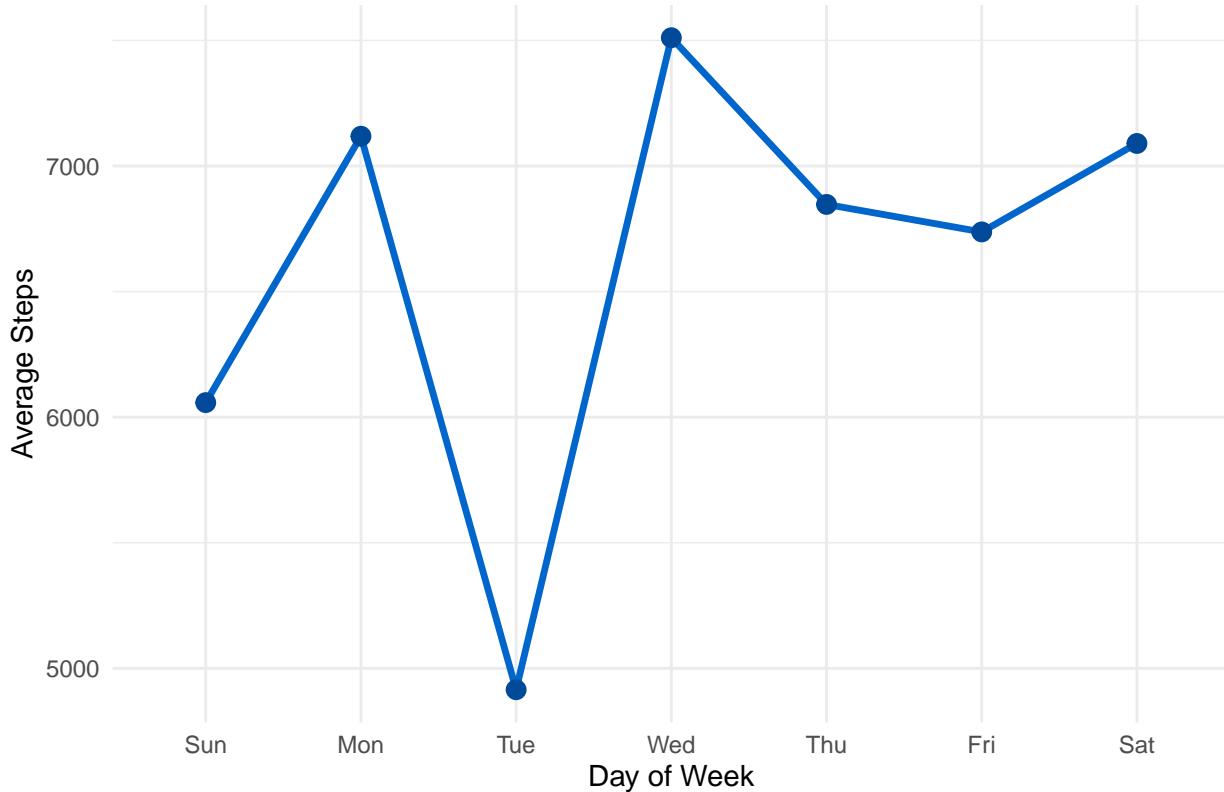
Sleep Duration vs Daily Steps



Insight: The relationship between sleep and activity is weakly negative — users with slightly longer sleep durations tend to take fewer daily steps, possibly due to reduced waking hours or lower energy expenditure.

3.2.8 Average Steps by Weekday

Average Daily Steps by Weekday



Insight: Average steps vary noticeably throughout the week. Users tend to be more active on weekdays, with visible peaks around mid-week (Tuesday–Thursday) and a noticeable drop in activity during weekends. This pattern may indicate that users' physical activity is influenced by structured routines (commuting, work schedules) and declines during leisure days, when they rely less on step-tracking habits.

4 Insights and Business Recommendations

4.1 Key Insights

1. Activity Patterns

- Most users show limited movement during early morning hours (**00:00–06:00**) and peak activity between **07:00–21:00**, with the **highest intensity from 17:00–20:00**.
- A strong positive correlation between **steps, intensity, and calories burned** confirms consistent physical activity tracking across devices.

2. Sedentary Behavior

- Users spend a large portion of their day sedentary, typically **700–1,000 minutes per day** (12–16 hours).
- A clear negative correlation exists between **steps and sedentary minutes**, confirming that active users experience substantially less idle time.

3. Sleep Patterns

- The majority of users record between **350–450 minutes of sleep per night** (6–7.5 hours).

- A weak negative correlation between **sleep duration and daily steps** suggests that higher activity levels may lead to slightly shorter sleep durations.

4. Weekly Trends

- **Wednesday and Monday** emerge as peak activity days, while **Tuesday** shows the lowest activity levels — possibly due to recovery or midweek fatigue.
- The **weekend drop** suggests less structured physical activity and reduced device engagement on leisure days.

5. Heart Rate

- Heart-rate data is limited but demonstrates a **positive correlation with both steps and intensity**, reinforcing its reliability for tracking exertion and calorie estimates.
-

4.2 Business Recommendations

1. Encourage Regular Movement

- Implement **in-app activity reminders** that nudge users to move after long sedentary periods.
- Introduce **micro-goals**, such as 200 steps every 30 minutes or quick “stretch breaks,” to maintain engagement throughout the day.

2. Optimize Sleep Insights

- Offer **personalized sleep recovery tips** — e.g., bedtime reminders or post-activity recovery suggestions based on prior fatigue levels.
- Use correlations between **sleep and activity intensity** to recommend ideal rest durations or recovery windows.

3. Gamify Activity Goals

- Launch **weekday-specific challenges** (e.g., “Move More Tuesdays”) to address low-activity days identified in the data.
- Add **streaks, badges, and milestone rewards** to incentivize consistent daily movement and improve long-term retention.

4. Integrate Heart-Rate Metrics

- Enhance device syncing to include **continuous heart-rate and intensity zone tracking**.
- Provide users with **personalized insights** such as “optimal calorie-burn hours” or “active recovery periods” derived from their heart-rate profiles.

5. Marketing and Growth Strategy

- Position Bellabeat as a **wellness intelligence platform** focused on holistic health — **activity, rest, and recovery**.
 - Target communications toward **young professionals and health-conscious women**, emphasizing balanced lifestyles, habit-building, and self-awareness.
-

5 Conclusion

The analysis of Fitbit-derived data highlights consistent user behavior patterns: **moderate physical activity, prolonged sedentary time, and generally healthy sleep habits**.

By embedding these findings into Bellabeat’s product and marketing strategy, the company can:

- Improve **user engagement** through behavior-based notifications and gamified challenges.
- Support **better health outcomes** by recommending balanced routines across activity and rest.

- Strengthen **brand differentiation** as a data-driven wellness company empowering women to make informed lifestyle decisions.
-