ASK

Business Task:

- The business task is to analyze smart device usage data from non-Bellabeat devices in order to identify major user trends. These insights will then be applied to one Bellabeat product to generate high-level marketing recommendations. The ultimate goal is to help Bellabeat strengthen its marketing strategy, increase customer engagement, and expand its position in the global smart wellness device market.

Key Stakeholders:

- **Urška Sršen:** Bellabeat's cofounder and Chief Creative Officer
- **Sando Mur:** Mathematician and Bellabeat's cofounder; key member of the Bellabeat executive team
- **Bellabeat marketing analytics team**

Prepare

- **DailyActivity_merged.csv:**The dataset contains daily activity information for users, including metrics such as total steps, calories burned, distance traveled, sedentary minutes, and time spent in different activity intensities (light, fairly, and very active). The data was collected through personal fitness trackers, which automatically record physical activity and related health metrics.
- **hourlySteps_merged :** This dataset contains the number of steps recorded per user at an hourly level. Each record corresponds to a user ID and a specific timestamp, providing a view of physical activity patterns throughout the day
- **hourlyIntensities_merged :** This dataset captures activity intensity at the hourly level, including total intensity and average intensity scores
- **hourlyCalories_merged :** This dataset records the estimated number of calories burned per user per hour

Process

Data Cleaning and Manipulation Documentation

- Data cleaning involved removing unrealistic records where daily steps were below 1,000 and excluding users with fewer than three days of data to ensure reliability. Ninety duplicate entries were deleted to avoid double-counting, and *ActivityDate* was converted into proper date format. A new variable was created by summing all activity minutes, with daily totals restricted to between 600 and 1,440 minutes. Calorie values above

8,000 kcal per day were treated as outliers and removed. Finally, dataset formatting was adjusted for readability.
- The three datasets (*hourlySteps_merged, hourlyIntensities_merged, and hourlyCalories_merged*) were imported into SQL. Data types were standardized: timestamps to **DateTime**, user IDs to **String**, integer values to **Integer**, and decimal values to **Float**. Using **User ID** as the primary key, the three tables were joined to create a new consolidated dataset, **calories_Intensities_steps_merged**, which combines steps, intensities, and calories for analysis.

Analyze

Summary of Analysis

Discover 1:

- I examined the relationship between different levels of activity (light, fairly active, and very active minutes) and calories burned. The correlation analysis shows that light activity has a weak positive correlation with calories (r = 0.13), fairly active minutes have a moderate correlation (r = 0.35), and very active minutes show the strongest correlation (r = 0.54). These results suggest that higher-intensity activities contribute more significantly to calorie expenditure compared to light or moderate activities.

Discover 2

- The analysis found a strong positive correlation between total steps and calories burned (r = 0.83). However, activity intensity showed an even stronger correlation with calories (r = 0.91). This indicates that although walking more steps contributes to higher energy expenditure, the level of intensity is a more decisive factor. This finding is consistent with Discover 1, where higher-intensity activity minutes were shown to contribute more significantly to calorie burn compared to light or moderate activity.

Discover 3:

- The distribution of activity efficiency (steps per calorie) shows clear variation across users. Most users cluster around the middle range, meaning they require a moderate number of steps to burn one calorie. However, some users fall on the lower end of the scale, indicating higher efficiency—they burn more calories with fewer steps—while others fall on the higher end, meaning they need many more steps for the same energy expenditure. The median efficiency line highlights where the majority of users lie. This
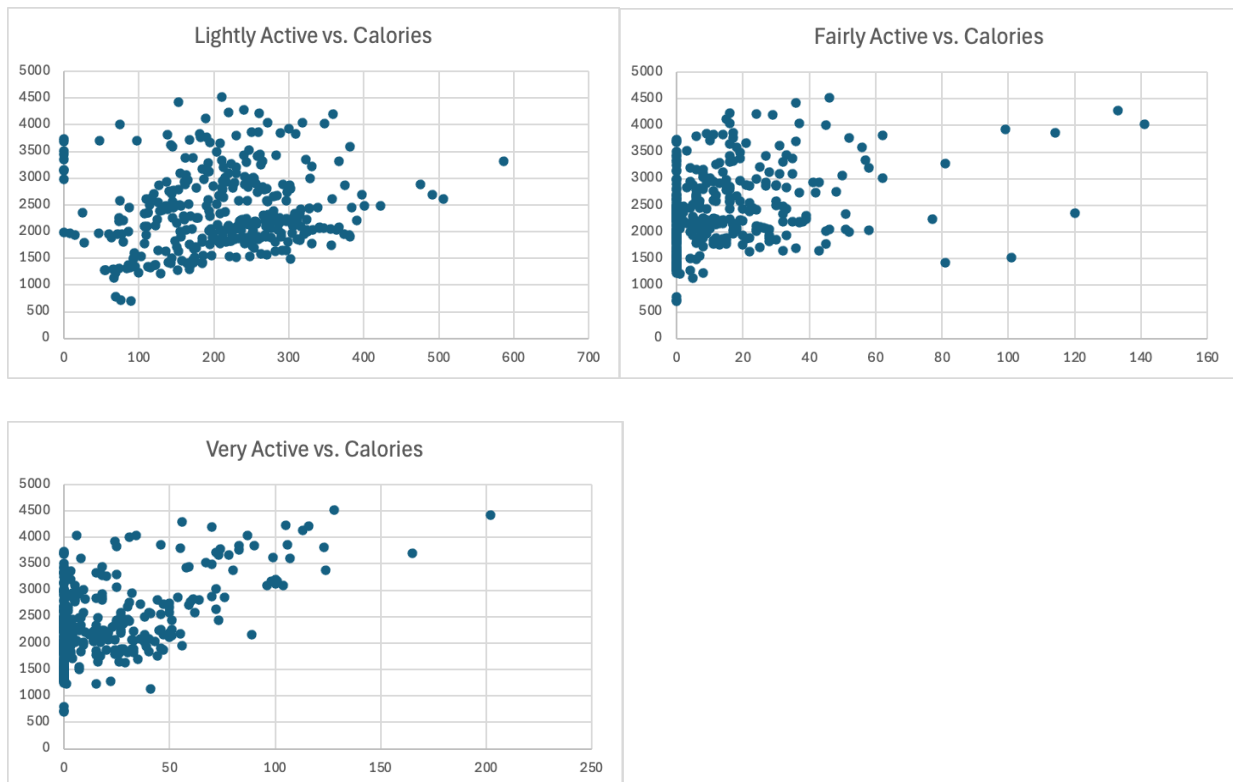
variation suggests that individuals respond differently to physical activity: for some, increasing intensity may significantly improve efficiency, while others may already achieve good calorie burn with relatively fewer steps. The correlation with activity intensity is strong and positive (r ≈ 0.78), indicating that higher-intensity activities make each step more effective at burning calories.
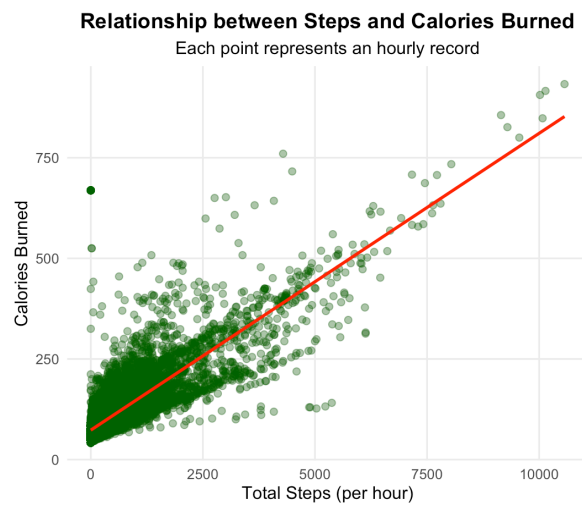
Discover 4:

- The hourly analysis shows that both steps and calorie expenditure follow a similar daily rhythm. Activity peaks in the morning and evening, while midday and late night remain consistently low. This pattern suggests that most users are active before and after work, but remain largely sedentary during the middle of the day. In addition, calorie burn sometimes rises without a proportional increase in steps, indicating that higher-intensity activities also play an important role in energy expenditure.
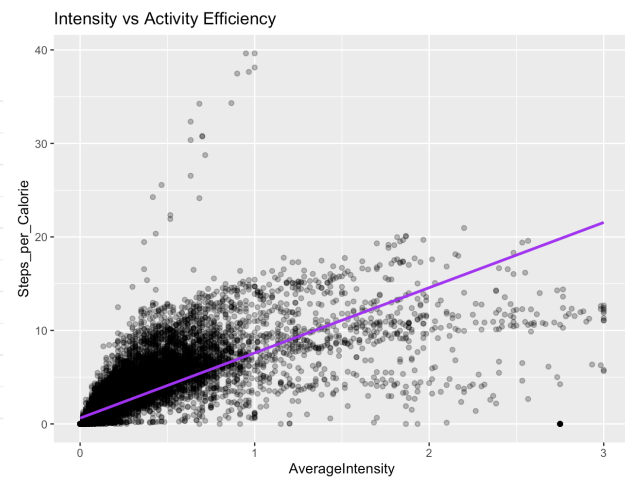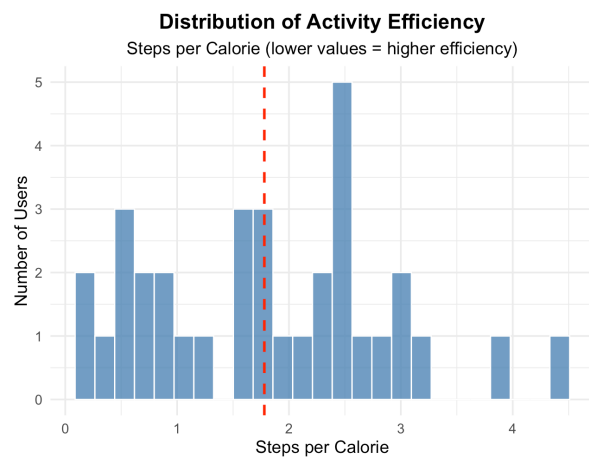
Share

Discover 1:

Discover 2:

Relationship between Steps and Calories Burned
Each point represents an hourly record

Intensity vs Calories
Relationship between Activity Intensity and Calories Burned

Discover 3:

Distribution of Activity Efficiency
Steps per Calorie (lower values = higher efficiency)

Intensity vs Activity Efficiency

Discover 4:

Average Steps by Hour

Average Calories by Hour

Act

Key Discoveries

Discover 1:
Very active minutes have the strongest correlation with calorie expenditure, while light and moderate activity contribute far less.

Discover 2:
Although step counts are strongly correlated with calories burned, activity intensity has an even stronger influence, making "moving harder" more effective than simply "moving more."

Discover 3:
Users show wide variation in activity efficiency (steps per calorie), meaning step goals alone are not effective. Personalized guidance is needed to reflect these differences.

Discover 4:
Activity levels peak in the morning and evening, while midday shows long sedentary periods. Calorie burn sometimes rises without a proportional increase in steps, indicating the role of higher-intensity activity.

Final High-Level Insight

Conclusion:
Calorie burn depends more on **intensity** than on steps, varies by **individual efficiency**, and follows clear **time-of-day patterns**. Bellabeat should move beyond generic step goals and

focus on **personalized, intensity-driven, and time-aware guidance** to help users achieve better health outcomes.

Application for Bellabeat:

- **Product Design:** Add efficiency scores, intensity-based goals, and personalized reminders, especially during midday sedentary periods.

- **Marketing Strategy:** Position Bellabeat as the brand that helps women **"move smarter, not just more,"** emphasizing intensity and timing.

- **User Engagement:** Align prompts with natural activity rhythms and tailor challenges to different efficiency levels.

Next Steps:
 Pilot efficiency and intensity features, develop midday nudges, and launch a campaign highlighting smarter movement.

Additional Data Needed:
 Include heart rate and demographic data to improve accuracy and validate insights across diverse users.