

Bellabeat Fitness Data Analysis Case Study

Introduction

This report summarizes the analysis of fitness-related data from Bellabeat datasets, focusing on activity patterns, caloric burn, and sleep data. The primary objective is to understand the relationship between physical activity and calories burned, along with exploring user behavior regarding sedentary and active segments. The datasets provided consist of various activity logs, including steps, calories burned, and sleep records. The ultimate goal is to segment users based on their activity levels and analyze the efficiency of their activity.

Data Overview

The following datasets were used in this analysis:

1. **daily_activity**: Contains data on daily physical activity, including total steps, distances, and active minutes.
2. **daily_steps**: Logs the total steps taken by users.
3. **daily_calories**: Contains information on the calories burned by users based on their daily activities.
4. **daily_sleep**: Contains records of users' sleep data, such as total sleep time and time spent in bed.

Data Cleaning and Merging

The data underwent a series of cleaning steps to ensure consistency and completeness:

- **Renaming Columns**: Columns were renamed for consistency across datasets.
- **Handling Missing Values**: Missing data, particularly in the `sleep_day` and `total_minutes_asleep` columns, were filled with the mean values to maintain dataset integrity.
- **Merging Data**: The datasets were merged on common identifiers (user ID and date), creating a comprehensive dataset that combines activity, calories, and sleep information.

Exploratory Data Analysis (EDA)

Correlation Analysis

A **correlation matrix** was generated to examine the relationships between different activity and sleep parameters. Key findings include:

- **Total Steps and Calories**: A positive correlation exists between **total steps** and **calories burned**, meaning as users take more steps, they burn more calories.

- **Very Active Minutes and Calories:** Similarly, **very active minutes** show a strong positive correlation with calories burned, indicating that more intense activity leads to a higher caloric burn.

Sedentary Behavior and Caloric Burn

Sedentary minutes were analyzed to see how they influence caloric burn. The following insights were drawn:

- **Descriptive Statistics:** On average, users spend **991.2 minutes** (around 16.5 hours) sedentary each day.
- **Correlation:** A negative correlation between **sedentary minutes** and **calories burned** was found, suggesting that increased sedentary behavior leads to reduced caloric expenditure.

User Segmentation

Users were segmented based on their **sedentary behavior**:

- **Low Sedentary:** Users who spend less than 4 hours a day sedentary.
- **Moderate Sedentary:** Users with 4-8 hours of sedentary behavior.
- **High Sedentary:** Users who are sedentary for more than 8 hours a day.

Interestingly, the **Moderate and High Sedentary groups** burned more calories on average than the **Low Sedentary** group, which could be due to factors like bursts of intense activity or higher energy expenditure.

Activity Classification by Steps

Users were further classified based on their **average daily steps**:

- **Sedentary:** Users with less than 5000 steps.
- **Moderately Active:** Users who take 5000 to 7500 steps.
- **Active:** Users who take 7500 to 10000 steps.
- **Very Active:** Users who take more than 10000 steps.

This classification was visualized in a pie chart, revealing that the majority of users are either **Active** or **Sedentary**.

Activity Efficiency

Efficiency per Step and **Efficiency per Active Minute** were calculated as:

- **Efficiency per Step:** Calories burned per step.
- **Efficiency per Minute:** Calories burned per active minute.

Analysis revealed that users are most **efficient on weekends**, with **Friday** and **Saturday** showing the highest efficiency for both steps and active minutes. This suggests that users engage in more intense activities during these days, potentially due to recreational activities or free time.

Trends in Activity Over Time

The **trend of average steps by day of the week** was analyzed. Users tend to be most active on **Saturday**, with activity levels declining during the workweek. This pattern is consistent with common work-life balance behaviors where users are more active during the weekend and less active during weekdays.

Statistical Testing

ANOVA (Analysis of Variance) was used to test if there are statistically significant differences in caloric burn across different activity categories. The p-value obtained from the test indicates no significant difference, suggesting that caloric burn variations among users are more likely due to individual differences rather than the activity category.

Caloric Goals

A **caloric goal of 2000 calories** was set, and approximately **65.12%** of days met or exceeded this goal. This suggests that a majority of users are engaged in sufficient physical activity to meet basic caloric burn expectations.

Summary of Key Insights

1. **Activity and Caloric Burn:** There is a strong positive correlation between **total steps**, **very active minutes**, and **calories burned**. Users who engage in more intense activities tend to burn more calories.
2. **Sedentary Behavior:** Users with higher sedentary minutes generally burn fewer calories, although some outliers indicate that low sedentary behavior doesn't always correlate with lower caloric burn.
3. **User Segmentation:** Most users fall into the "Active" or "Sedentary" categories. The "Very Active" group is slightly outnumbered by those who are either active or sedentary, indicating a mix of user behaviors.
4. **Weekday Trends:** Users are most active on **Saturdays** and **Fridays**, with reduced activity during weekdays, particularly on **Mondays**.
5. **Efficiency:** The most efficient activities (in terms of calories burned per step and per minute) occur during weekends.

Conclusion

This analysis highlights how activity patterns, including sedentary behavior and physical activity levels, influence daily caloric expenditure. The findings offer insights that can guide product development and marketing strategies for fitness-related devices, encouraging users to maintain consistent activity levels and achieve their fitness goals. Further research could focus on understanding the physiological factors behind these patterns and how users interact with the product over time.