

Executive Summary

The objective of this analysis was to uncover non-intuitive patterns in how women use smart health devices, using Fitbit data as a proxy, to inform Bellabeat's product development—particularly the Leaf tracker and Bellabeat app—and support its marketing strategy. Multiple datasets related to physical activity, sleep behavior, heart rate, and calorie expenditure were analyzed to identify meaningful trends in real-world user behavior.

The analysis revealed several key insights:

- Higher levels of daily physical activity were associated with lower average body weight, although this relationship was not strictly linear.
- Users with lower activity levels consistently recorded longer sleep durations than more active users, suggesting that sleep quantity alone may not reflect overall wellness.
- Average heart rate increased steadily with higher activity intensity levels, validating the physiological consistency of the activity classifications.
- Energy expenditure rose significantly with activity intensity, with noticeable differences between weekdays and weekends.

Based on these findings, Bellabeat can take several data-driven actions to improve both product experience and user engagement:

- Develop personalized in-app insights that emphasize activity quality and intensity rather than step count alone.
- Design sleep-related features that focus on sleep efficiency and recovery, not just total sleep duration.
- Leverage heart rate and calorie expenditure data to deliver adaptive feedback and wellness recommendations.
- Launch targeted marketing campaigns aligned with real user behavior patterns, differentiating weekday and weekend activity habits.

By aligning product features and marketing initiatives with observed user behavior, Bellabeat can strengthen user engagement and reinforce its positioning as a data-informed wellness brand.

Background & Business Task

Background (Business Context)

Bellabeat is a wellness technology company founded in 2013 that develops smart devices and health applications designed specifically for women. Its product ecosystem includes wearables such as Leaf, Time, and Spring, along with the Bellabeat app, which tracks physical activity, sleep, heart rate, and daily habits.

Unlike many competitors in the wearable technology market, Bellabeat adopts a holistic, women-centered approach to health, combining physiological data with lifestyle insights, motivation, and self-care. The company's mission is to empower women to better understand their bodies and make informed decisions through data-driven insights.

Operating in a highly competitive market dominated by major players such as Fitbit, Apple, and Garmin, Bellabeat aims to differentiate itself not only through technology, but also through how it interprets and communicates health data tailored to women's needs. In this context, data analysis plays a critical role in strengthening its value proposition and supporting business growth.

Business Task

The primary objective of this analysis is to identify usage patterns and behavioral trends in smart fitness devices using publicly available Fitbit data, in order to:

- Understand how users engage with fitness and health tracking devices in their daily lives.
- Identify key trends in physical activity, sleep behavior, and physiological metrics.
- Generate actionable insights to support Bellabeat's marketing strategy, with a particular focus on improving engagement and relevance for women.

The insights derived from this analysis will enable Bellabeat to refine its marketing messaging, product recommendations, and user communication, ensuring closer alignment with real user behaviors and reinforcing its positioning as a women-focused wellness brand.

Key Stakeholders

- **Urška Sršen (Co-founder & Chief Creative Officer)**: Oversees product vision and ensures alignment with women's wellness needs.
- **Sando Mur (Co-founder & Mathematician)**: Leads data-driven strategic decision-making and analytical direction.
- **Bellabeat Marketing Team**: Primary consumer of the analysis results, responsible for translating insights into growth strategies.
- **Product & Development Team**: Uses findings to enhance app features, notifications, and overall user experience.

Data Sources & Data Cleaning (Methodology)

This section describes the data sources used, their main limitations, and the methodological process applied for data cleaning and preparation prior to exploratory analysis and visualization.

Data Sources

For this analysis, the FitBit Fitness Tracker Data dataset was used. This dataset is publicly available on the Kaggle platform and was collected by Mobius. It contains anonymized data generated by Fitbit devices, including metrics related to daily physical activity, sleep, body weight, and heart rate.

The dataset covers a period between April and May 2016 and consists of records from 30 users who voluntarily agreed to share their data. The information is provided across multiple CSV files, each corresponding to a specific type of metric.

Although the data does not come directly from Bellabeat users, it is considered an appropriate proxy for this analysis, as Bellabeat operates within the same health and wellness tracking market segment. This allows for the exploration of behavioral patterns that can support the development of data-driven strategic recommendations.

Data Reliability & Limitations

While the dataset is useful for identifying general behavioral patterns, several limitations must be considered when interpreting the results:

- The sample size is small (30 users), which limits the generalizability of the findings.
- The data comes from users who voluntarily shared their information, introducing a potential self-selection bias.
- The dataset lacks detailed demographic information such as age, gender, or socioeconomic context. Given that Bellabeat primarily targets women, it is not possible to confirm whether the observed patterns accurately represent its core audience.
- The data was collected in 2016, and therefore may not fully reflect current user habits or technology usage trends.
- It is assumed that users wore their Fitbit devices correctly and consistently, although individual adherence and measurement accuracy cannot be independently verified.

These limitations were taken into account throughout the analysis to ensure a responsible and realistic interpretation of the results.

Data Cleaning Process

The data cleaning and preparation process was conducted using the R programming language, with the objective of ensuring data quality, consistency, and reliability prior to exploratory analysis and visualization.

Datasets Used

The original dataset consisted of 7 CSV files, with the following characteristics:

- daily_activity_merged.csv: 940 observations and 15 variables, including daily activity metrics such as steps, calories burned, distance, and active minutes.
- daily_calories_merged.csv: 940 observations and 3 variables containing daily calorie data.
- daily_steps_merged.csv: 940 observations and 3 variables recording total daily steps.
- daily_intensities_merged.csv: 940 observations and 10 variables detailing daily activity intensity levels.
- sleepDay_merged.csv: 413 observations and 5 variables related to sleep duration and quality.
- weightLogInfo_merged.csv: 67 observations and 8 variables related to body weight and composition.
- heartrate_seconds_merged.csv: 2,483,658 observations and 3 variables capturing heart rate measurements at the second level.

For the core analysis, 6 of these datasets were actively used, excluding daily_intensities_merged.csv. Activity intensity information was instead derived from the aggregated variables already available in daily_activity_merged.csv. This decision simplified the data model without compromising analytical relevance.

Data Cleaning and Preparation Steps

Prior to analysis, the following steps were performed:

- Identification and removal of duplicate records, missing values, and incomplete entries
- Exclusion of users with insufficient data coverage
- Standardization of date and time formats across all datasets
- Transformation of temporal variables to identify day of the week and day type (weekday vs. weekend)
- Integration of multiple datasets using user identifiers and shared dates
- Validation of internal consistency across key metrics such as steps, calories, active minutes, and sleep duration
- Creation of derived variables to support downstream analysis and insight generation

This cleaning process ensured that the final dataset was coherent, comparable, and suitable for addressing the defined business questions.

Tools Used

The analysis and data preparation were conducted using tools designed to support reproducible analysis and effective data visualization:

- **R** as the primary programming language
- **RStudio** as the development environment
- **tidyverse** for data manipulation and transformation
- **dplyr** for filtering, grouping, and summarization
- **lubridate** for date handling and standardization
- **ggplot2** for exploratory and final data visualizations

This toolset enabled a structured, efficient, and reproducible analytical workflow.

Methodological Assumptions

The following methodological assumptions were established during the analysis:

- Users were assumed to wear their Fitbit devices correctly and consistently throughout the data collection period.
- Available daily records were considered representative of users' general habits, despite potential individual variability or atypical days.
- The results are interpreted as exploratory and indicative, rather than definitive or broadly generalizable.
- The integration of multiple data sources was assumed to provide a coherent representation of user behavior, even though some metrics were not uniformly available for all users.

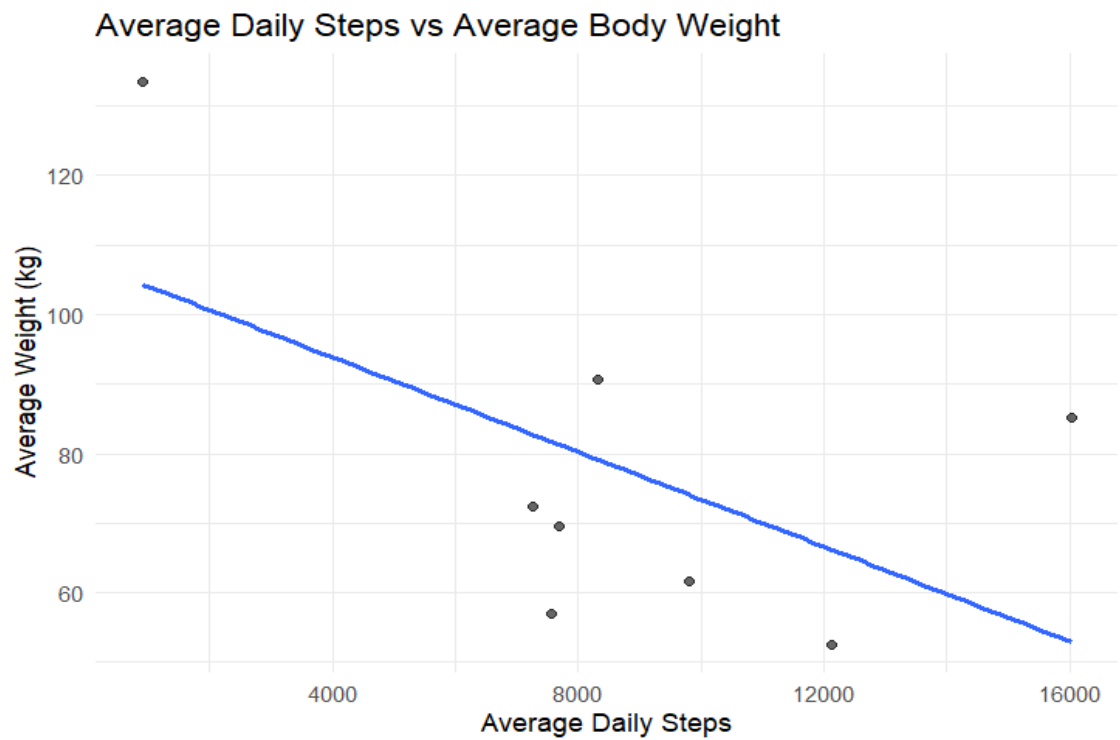
Conclusion

This section provides a clear, transparent, and professional description of the data sources, limitations, and technical decisions underpinning the analysis, establishing a solid foundation for the insights and recommendations presented in subsequent sections.

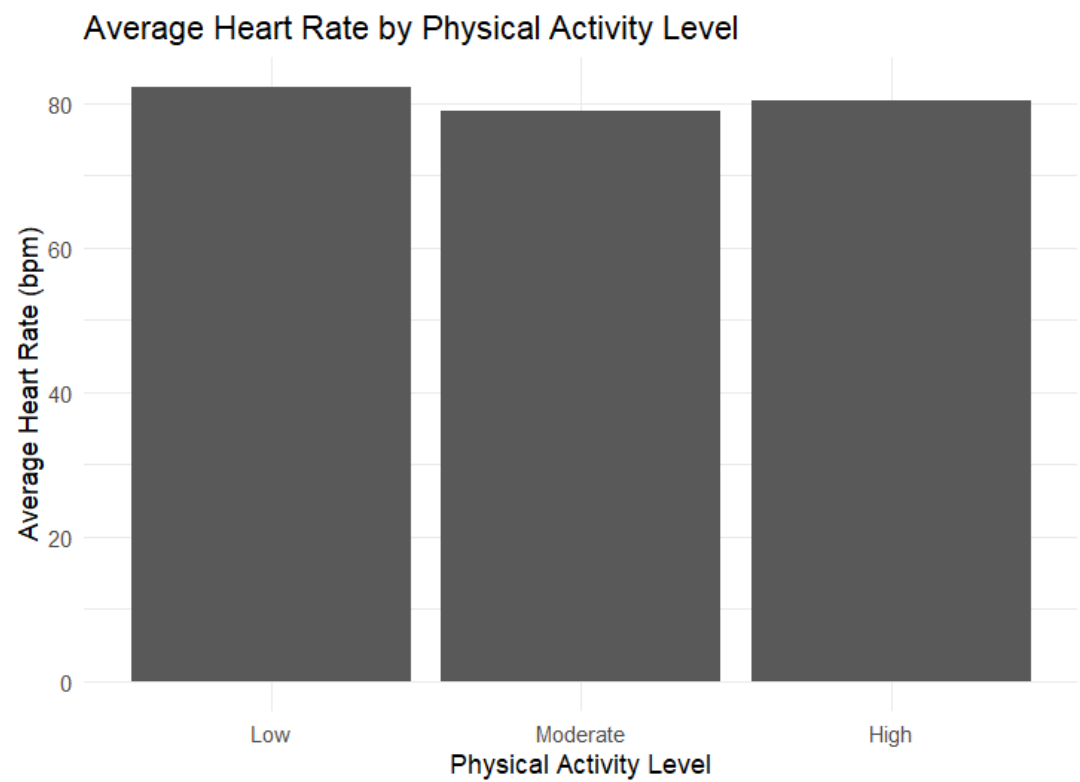
Insight 1: Physical Activity, Body Weight, and Heart Rate Patterns

Key Finding

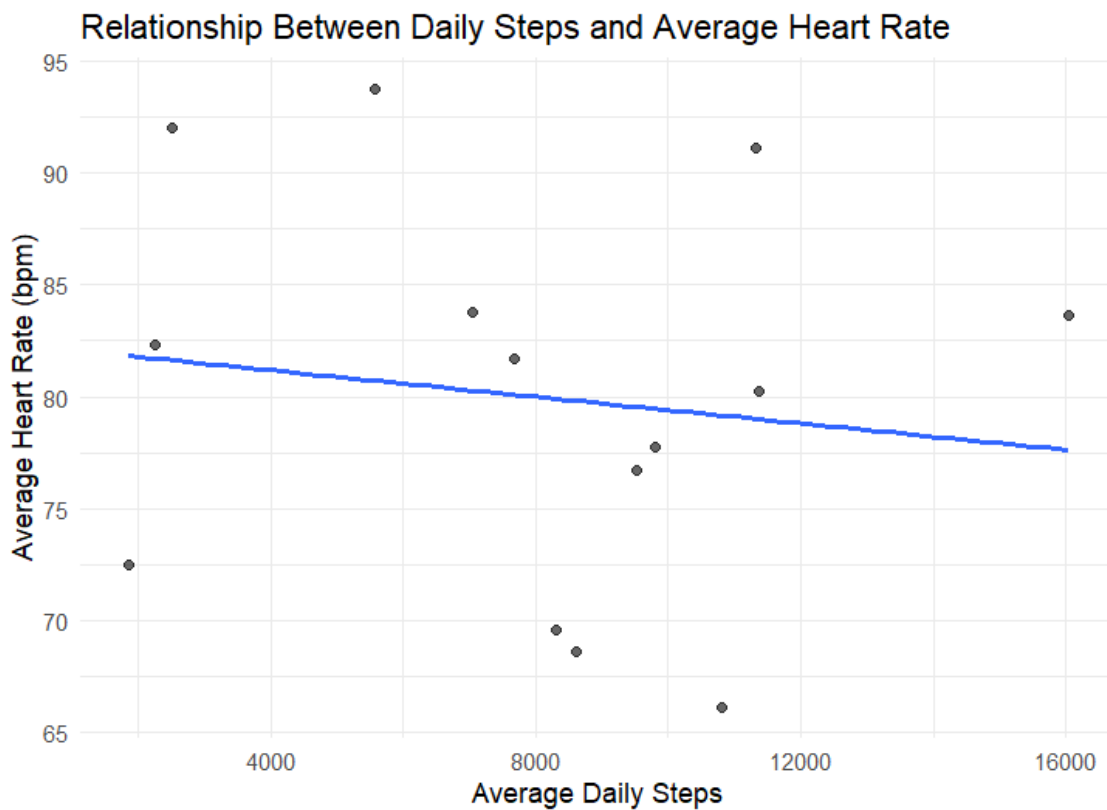
The analysis shows a relationship between users' daily step counts and body weight, where users with higher average daily steps tend to present lower body weight values.



Additionally, the physiological data shows slight variation in average heart rate across activity levels, with users in the 'Moderate' and 'High' activity categories exhibiting marginally lower average heart rates compared to the 'Low' activity group.



Together, these results suggest that the step-based activity classifications are reflected not only in body-related metrics but also in physiological responses, supporting the internal consistency of the activity data.



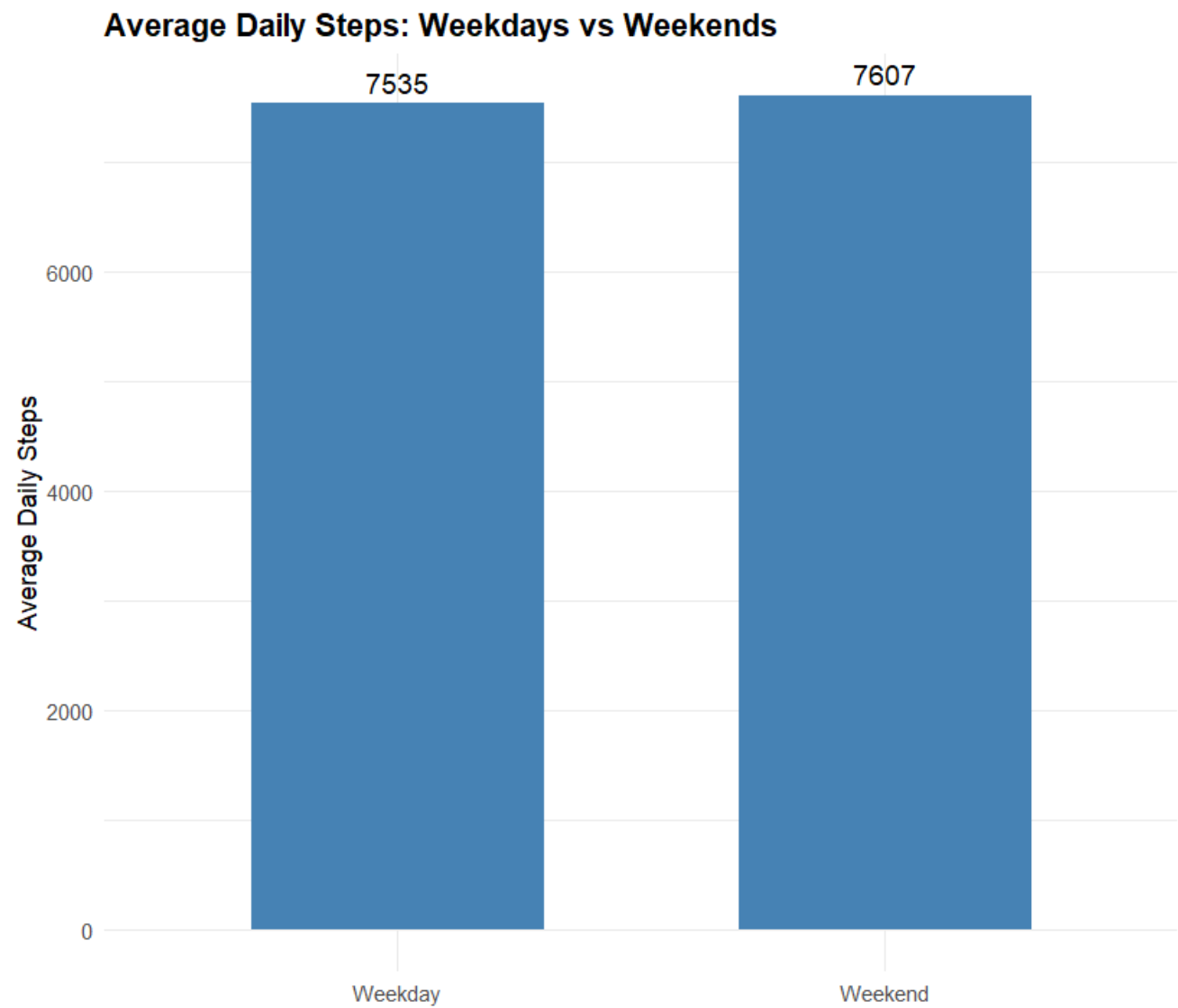
Why this matters for Bellabeat

This insight highlights the value of step-based activity tracking as a meaningful indicator of both physical movement and physiological engagement. It supports the use of daily steps as a core metric for user feedback, progress tracking, and personalized wellness guidance.

Insight 2: Activity Patterns and Energy Expenditure Across the Week

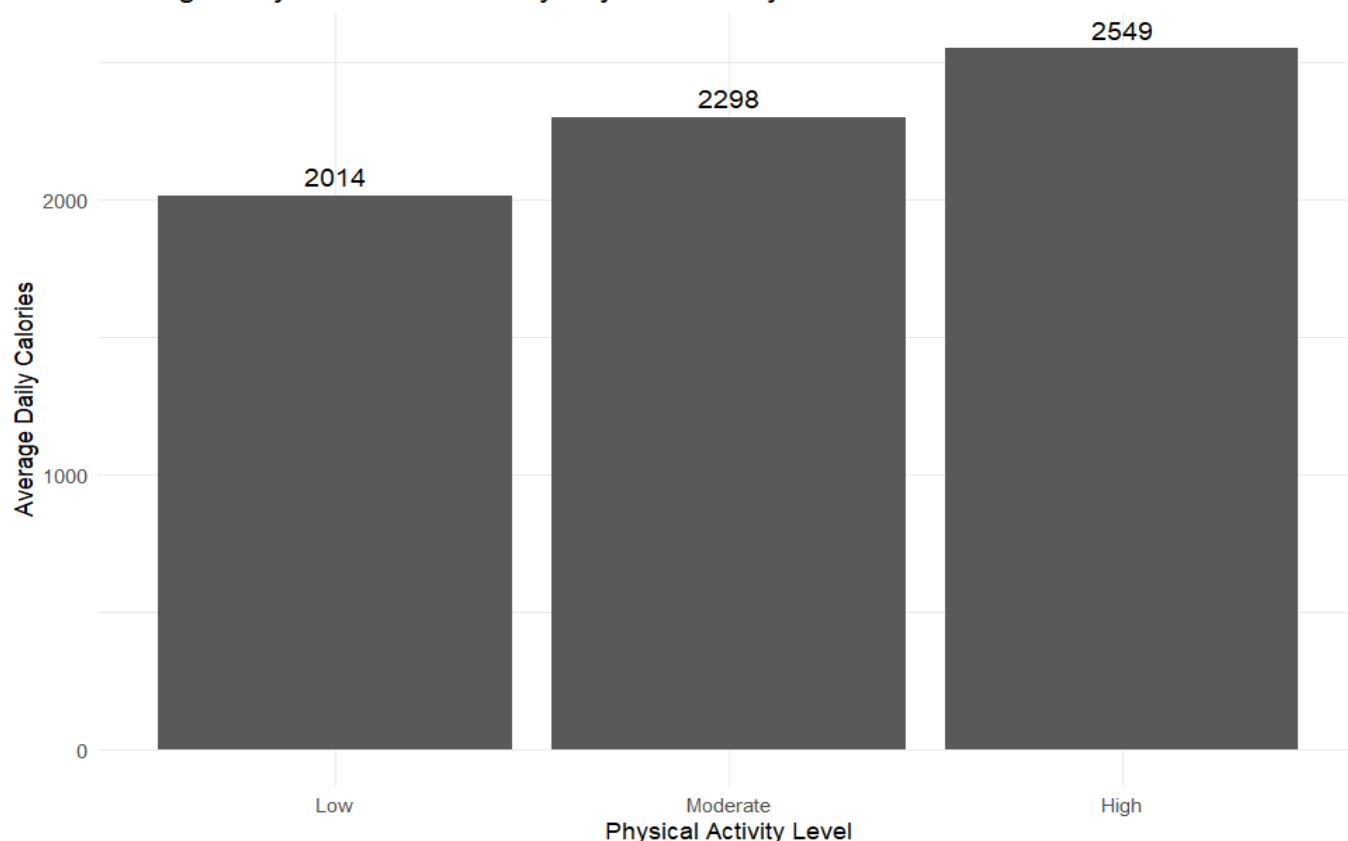
Key Finding

Physical activity patterns remain highly consistent across weekdays and weekends, with only a marginal difference of approximately 1% in average daily steps. Users average slightly more steps on weekends than on weekdays, reinforcing the overall stability of movement behavior throughout the week.



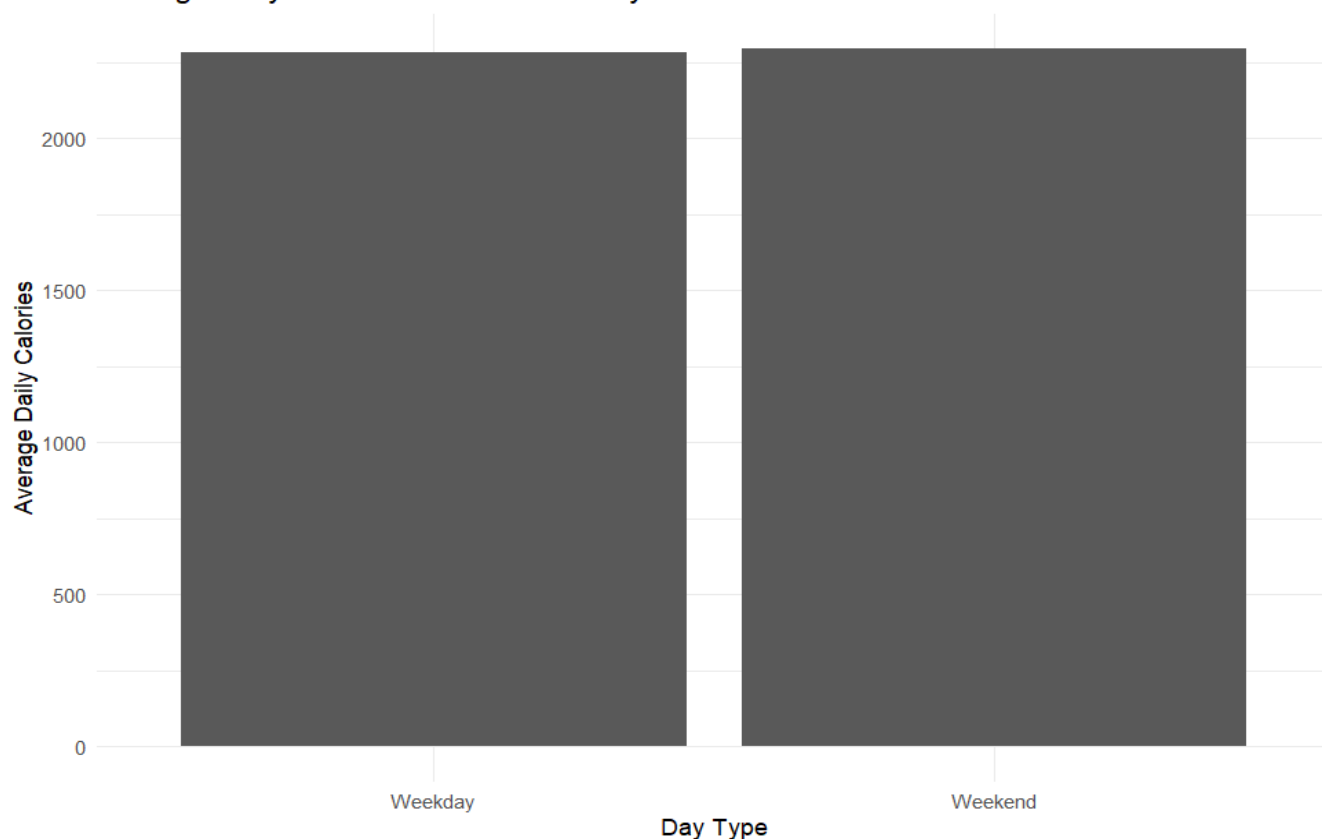
In contrast, energy expenditure increases substantially with higher activity intensity. Users classified in higher activity levels burn approximately 26% more calories than those in the “Low” activity group.

Average Daily Calories Burned by Physical Activity Level



Additionally, calorie burn remains comparable between weekdays and weekends when activity intensity is similar, indicating that energy expenditure is primarily driven by activity level rather than by the type of day.

Average Daily Calories Burned: Weekdays vs Weekends



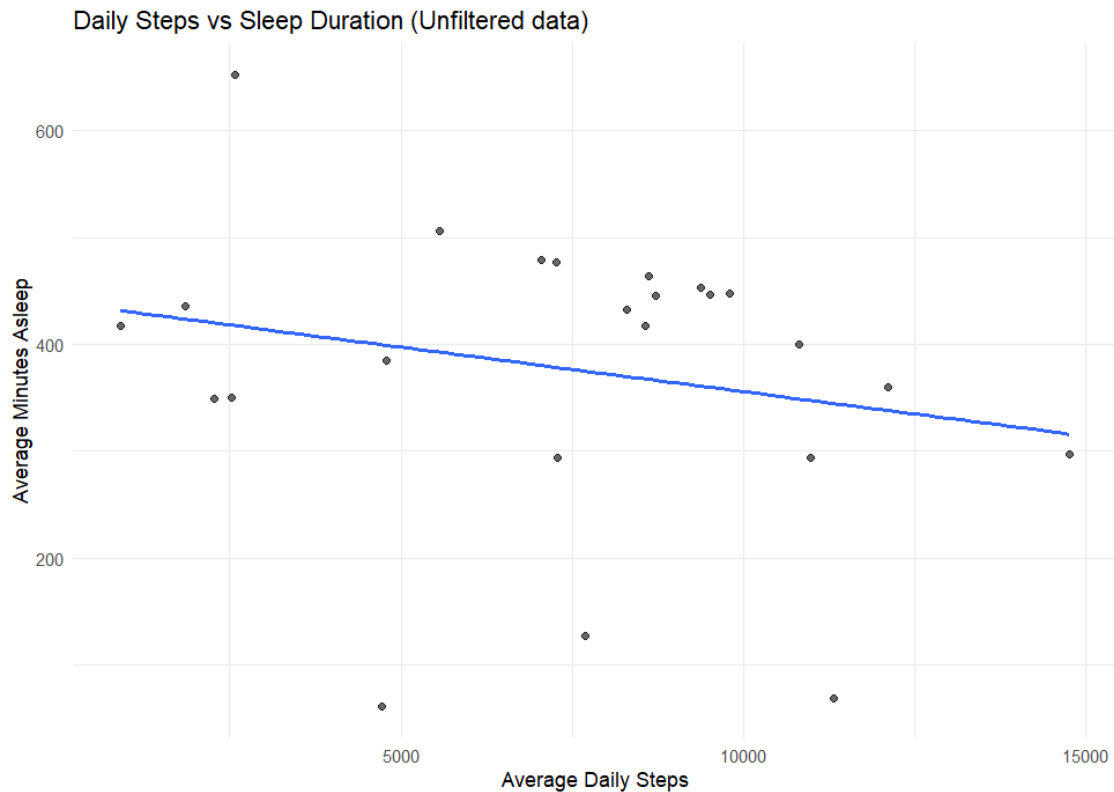
Why this matters for Bellabeat

This finding emphasizes the importance of encouraging higher activity levels regardless of the day of the week. Bellabeat can leverage this insight to promote consistent movement habits and highlight that meaningful energy expenditure can be achieved both on weekdays and weekends.

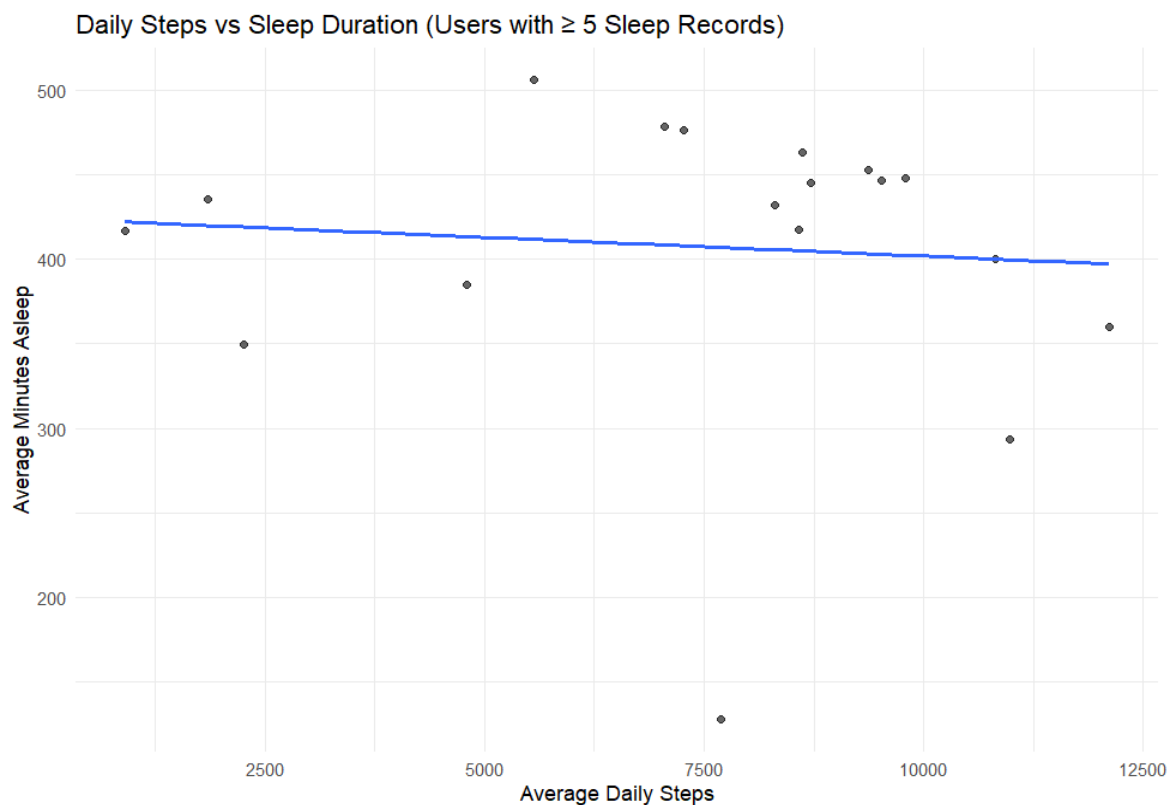
Insight 3: Physical Activity and Sleep Duration & Onset Patterns

Key Finding

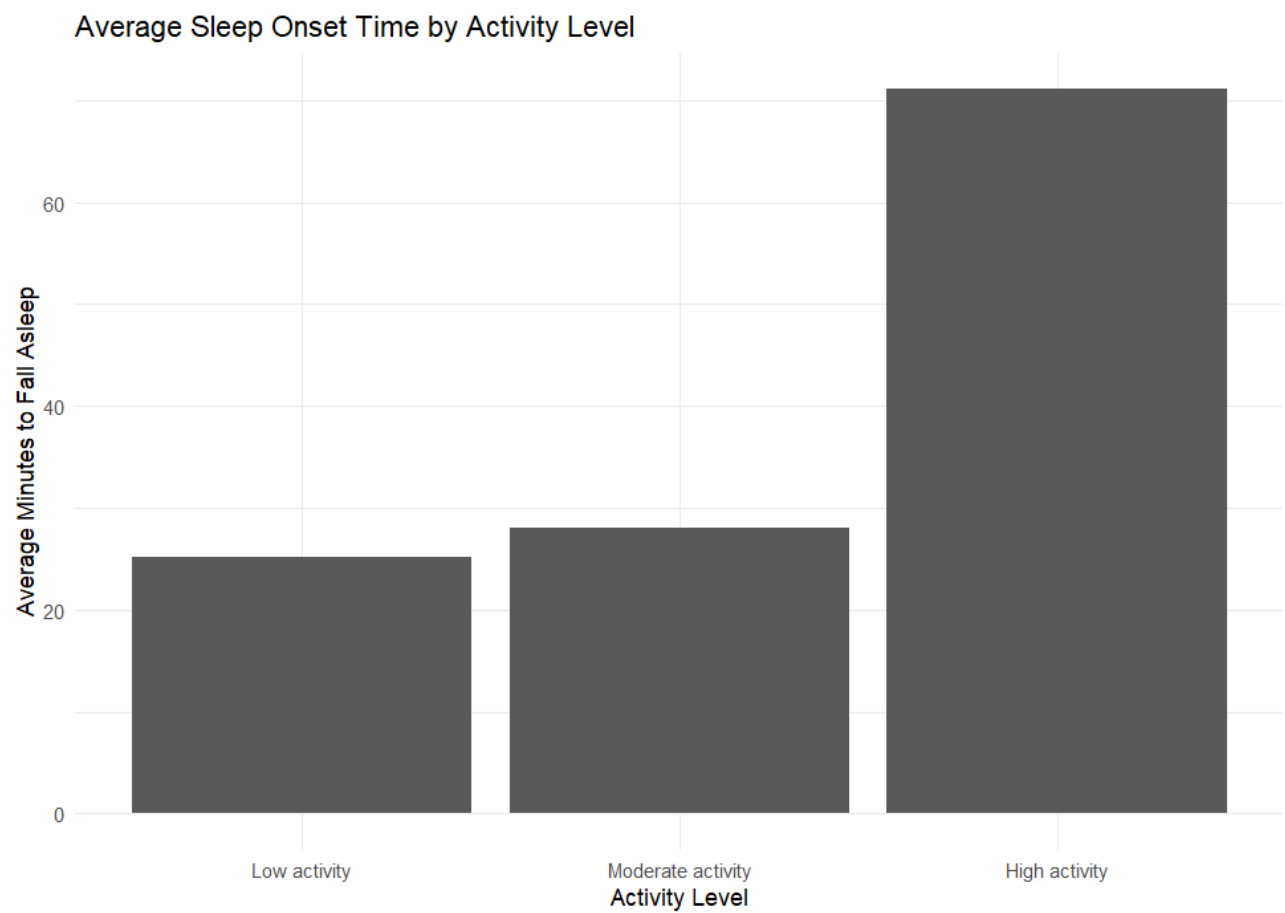
Daily step counts show weak and inconsistent associations with both total sleep duration and the time it takes users to fall asleep. While a slight negative relationship with sleep duration is observed in some scenarios ($r = -0.22$)



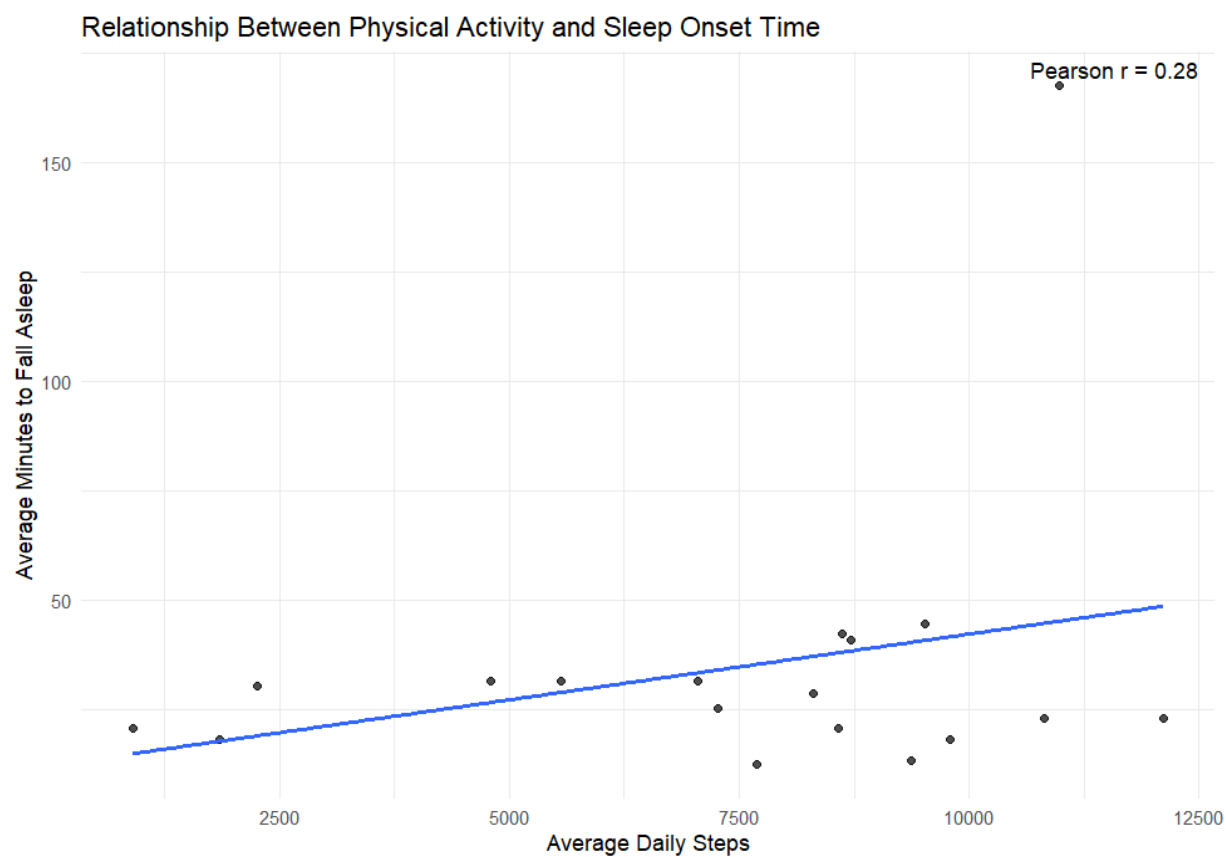
this association diminishes substantially when data quality improves ($r = -0.08$), suggesting limited explanatory power.



Similarly, the relationship between physical activity and sleep onset time is weak. Notably, the “High Activity” group showed longer average sleep onset times, a pattern influenced by variability and outliers.



Overall, these findings indicate that daily physical activity alone is not a strong predictor of sleep outcomes within this dataset.



Why this matters for Bellabeat

This insight suggests that sleep outcomes are influenced by multiple factors beyond daily physical activity alone. Relying solely on step counts to provide sleep coaching could lead to inaccurate user expectations. For Bellabeat, this highlights the importance of delivering holistic sleep insights that integrate activity, routines, and recovery behaviors, rather than relying exclusively on step-based metrics to explain sleep patterns.

Business Recommendations

1) Reframe Marketing from Quantity to Quality (Sleep Score)

Based on: Insight 3 (Sleep vs. Steps)

- **Strategy**: Debunk the assumption that “more steps automatically lead to more sleep hours.” Bellabeat should prioritize Sleep Efficiency and Sleep Quality (Sleep Score) over total sleep duration.
- **Action**: Implement educational in-app notifications. For example, if a user has a highly active day but short sleep duration, the app could display: **“You had an active day. Let’s focus on achieving deep, restorative sleep tonight—even if the duration is shorter.”**

This reframes expectations, reduces user frustration, and reinforces the value of Bellabeat’s sensors and analytics.

Expected Business Impact:

- Increased trust through evidence-based wellness messaging
- Higher adoption of advanced sleep metrics (Sleep Score)
- Reduced churn driven by unrealistic sleep expectations

2) Gamification to Improve Data Integrity (Data Strategy)

Based on: Data limitations (need for ≥ 5 consistent records to detect meaningful trends)

- **Strategy**: Encourage consistent data collection through Logging Streaks.
- **Action**: Reward users who log both activity and sleep for 7 consecutive days with badges or in-app points.

Expected Business Impact:

- Improves user retention through engagement mechanics
- Reduces noise in recommendation models
- Enables more accurate and personalized insights in the long term

3) Smart User Segmentation and “Wind-Down” Alerts

Based on: Insight 3 (Longer sleep latency among highly active users)

- **Strategy**: Identify and address the “Activity Paradox” (wired but tired).
- **Action**: For the High Activity segment, automatically trigger Wind-Down Alerts 60 minutes before bedtime, suggesting guided meditation, breathing exercises, or screen-disconnection reminders.

Expected Business Impact:

- Helps transition the nervous system from alertness to rest
- Directly targets the observed increase in sleep latency
- Strengthens Bellabeat’s role as an active wellness coach

4) Daily Routine Optimization Through Exercise Timing

Based on: Insight 2 (Weekly consistency) and Insight 3 (Sleep latency)

- **Strategy**: Monitor exercise timing to better align activity with circadian rhythms.
- **Action**: Allow users to log the time of their workouts. If data shows that late-evening intense exercise is associated with longer sleep latency, the app should recommend shifting workouts to earlier hours.

Expected Business Impact:

- Positions Bellabeat as a health coach that understands women’s biological rhythms
- Enhances personalization of wellness guidance
- Improves perceived sophistication of the product

5) Incentives for Sustained Weekend Engagement (“Weekend Boost”)

Based on: Insight 2 (High weekday–weekend consistency with marginal weekend increase)

- **Strategy**: Reinforce consistent movement habits across the entire week, while sustaining engagement during lower-structure days.
- **Action**: Launch low-effort “Light Movement” challenges on weekends for sedentary or low-activity users, encouraging gentle activity without compromising rest.

Expected Business Impact:

- Maintains engagement across all seven days of the week
- Supports consistent energy expenditure without promoting overexertion
- Reinforces healthy, sustainable activity habits

Final Note

These recommendations translate analytical insights into concrete product, marketing, and data strategy actions, demonstrating how Bellabeat can leverage behavioral data to enhance user trust, engagement, and personalization—without overstating causal relationships.

Limitations & Next Steps

Limitations

This analysis is subject to several limitations that should be considered when interpreting the findings:

- **Sample Bias & Size Attrition**: The dataset represents a relatively small starting sample (n=30). Furthermore, significant data gaps were identified across specific metrics; for instance, when filtering for users with consistent weight or sleep logs, the sample size was further reduced. This sample attrition weakens the statistical power of the analysis and limits the generalizability of the results to Bellabeat's broader user base.
- **Data Obsolescence**: The data was collected in 2016. Since then, wearable technology accuracy, user adoption rates, and health awareness have evolved significantly. Consequently, these insights may not fully reflect current digital health behaviors.
- **Lack of Demographic Context**: The metadata lacks critical information such as confirmed gender, age, and lifestyle habits. This is a significant constraint given Bellabeat's specific focus on women's health, as it prevents meaningful segmentation (e.g., analyzing patterns based on life stages or menstrual cycles).
- **Correlation vs. Causality**: This is an observational study. While associations between activity levels and sleep patterns were identified, the lack of controlled variables means we cannot infer causality.
- **Data Integrity**: Certain variables, specifically weight logs and minute-by-minute sleep indicators, contained significant amounts of missing or sparse data, reducing the statistical confidence in conclusions related to these specific metrics.

Next Steps

To strengthen future analyses and drive more actionable business growth, the following steps are recommended:

- **First-Party Data Integration**: Transition from public proxy datasets to Bellabeat's internal user data. This would provide high-fidelity, real-time information that is directly aligned with the brand's target audience.
- **Enhanced User Segmentation**: Incorporate demographic and lifestyle variables—such as age brackets, stress scores, and menstrual cycle phases—to enable hyper-personalized marketing and product features.
- **Longitudinal Habit Analysis**: Conduct a long-term study (6–12 months) to identify seasonal trends, habit formation, and the long-term impact of Bellabeat's reminders on user retention.
- **Mixed-Methods Approach**: Combine these quantitative findings with qualitative UX Research (surveys or interviews). Understanding the "why" behind user behavior will help refine the app's interface and content strategy to better meet women's emotional and physical needs.

Conclusion

This analysis provided a data-driven exploration of smart device usage to understand how daily activity and sleep patterns influence user well-being. By examining the relationship between physical activity and rest, the project moved beyond raw metrics to identify actionable opportunities for Bellabeat.

Key Strategic Takeaways

- **Quality over Volume**: The findings indicate that raw activity metrics, such as daily step counts, are not sole predictors of sleep health. Instead, holistic indicators—including sleep quality scores and habit consistency—offer more meaningful insights into overall user wellness.
- **Actionable Intelligence**: From a business perspective, Bellabeat has a significant opportunity to shift its value proposition from promoting volume-based activity goals to delivering personalized, quality-focused health insights. This approach aligns more naturally with women's real-life experiences and long-term health needs.
- **Data as a Roadmap**: While limitations of the current dataset—such as sample attrition and data obsolescence—were identified, these constraints provide a strategic roadmap for improving data maturity and strengthening future analytical confidence.

Final Outlook

The insights derived from this analysis offer a foundational framework for refining Bellabeat's positioning and enhancing user engagement. By bridging the gap between passive tracking and active wellness management, Bellabeat can empower its community with information that reflects individual biological rhythms, guiding future analytical efforts toward a more comprehensive and representative ecosystem.