DATA-DRIVEN STRATEGIES FOR WELLNESS TECHNOLOGY GROWTH

Bellabeat data analysis case study

Carolina Abelho

May 2025

Contents

INTRODUCTION

This report details a comprehensive data analysis project for Bellabeat, a company dedicated to women's health. The core goal was to analyze smart device usage data and uncover key consumer behavior patterns. Organized around the Ask, Prepare, Process, Analyze, Share, and Act phases, this document outlines the entire analytical process and presents data-driven insights and recommendations to enhance Bellabeat's marketing and product strategies.

ASK

DEFINING THE BUSINESS TASK

Identify key consumer behavior patterns and trends from Bellabeat smart device usage data to develop data-driven marketing strategies that drive growth and boost customer engagement.

PREPARE

DATA SOURCES USED

PRIMARY DATA SOURCE

The dataset for this analysis originates from Fitbit Fitness Tracker Data, publicly available through Mobius on Kaggle under a CC0 Public Domain license. This data was collected from 30 eligible Fitbit users who voluntarily submitted their personal tracker data via Amazon Mechanical Turk (MTurk) over a 62-day period, spanning March 12, 2016, to May 12, 2016.

SELECTED DATA SOURCES AND ORGANIZATION

The data was provided in 29 CSV files, organized into two folders corresponding to the following date ranges:

1. March 12, 2016 – April 11, 2016
2. April 12, 2016 – May 12, 2016

After careful consideration of data completeness, consistency, temporal coverage, and the ability to aggregate, the following files were chosen as the primary sources for analysis, with aggregation to daily and hourly levels where applicable:

hourlySteps: Provides a reliable and consistent baseline for hourly step counts.

hourlyCalories: Offers data for calorie expenditure, allowing for accurate aggregation to daily levels.

minuteIntensities: Provides detailed intensity information at the minute level, crucial for accurate aggregation of activity levels at different temporal resolutions.

minuteMETs: This minute-level file is the *only* source for METs data, making it essential for analyzing activity intensity in terms of metabolic equivalents.

minuteSleep: Offers the most detailed sleep records, enabling flexible aggregation to hourly and daily sleep metrics.

heartrate_seconds: Provides heart rate data at second-level, allowing for the calculation of average heart rate at minute, hourly, and daily values.

This selection prioritizes granularity where available to maximize the potential for accurate aggregation and detailed insights into user behavior across different time scales.

The analysis will focus on the records of the following health metrics present in the provided datasets:

1. **Steps:** Represents the total number of steps taken by the user, available as total daily and hourly aggregates

2. **Calories:** Indicates the estimated total energy expenditure, most likely in kilocalories (kcal), available as total calories burned daily or hourly.
3. **Minutes in each Intensity Level:** Represents the total duration (in minutes) spent in each of the following activity intensity levels: sedentary, lightly active, fairly active, and very active.

   In the *minutesIntensities* file, intensity levels are encoded numerically: 0 (sedentary), 1 (lightly active), 2 (fairly active), and 3 (very active) - this pairing was done after comparison with dailyData file. Hourly aggregates were derived by summing the occurrences of each level within each hour.

   For daily or hourly summaries, the values represent the total minutes within that period for each intensity level. The sum of minutes across all four intensity levels for any given day cannot exceed 1440 minutes (24 hours).

4. **METs** (Metabolic Equivalent of Task):

   A first evaluation on Minute-level METs files shows that values range from 10 to 189. It is hypothesized that these values may need to be scaled by a factor of ten for accurate interpretation. Furthermore, while a MET value of zero is unexpected, aggregation to hourly or daily levels will utilize the average MET value to provide a representative measure of intensity over time.

   "*METs are a measure of the energy cost of physical activities relative to resting metabolism. One MET is defined as the oxygen consumption rate while at rest (approximately 3.5 ml O2/kg/min).*"

   Reference: National Library of Medicine, https://pubmed.ncbi.nlm.nih.gov/2204507/

   The following activity intensity categories are proposed to correspond with specific ranges of METs:

   Low-intensity: ≤ 1.5 METs
   Light-intensity: 1.6 to 2.9 METs
   Moderate-intensity: 3.0 to 5.9 METs
   High-intensity: ≥ 6.0 METs

   Reference: https://pmc.ncbi.nlm.nih.gov/articles/PMC11110193/

5. **Heart Rate Data:** Represents the user's heart rate in Beats Per Minute (BPM), recorded at the second level and aggregated to minute, hourly, and daily averages for analysis.

6. **Sleep Data:** Minute-level records indicating the user's sleep state. The 'value' column ranges from 1 to 3, likely representing different sleep levels or stages.

   The following relationships are used when aggregating sleep data in minutes to hour or daily level: TotalMinutesAsleep corresponds to the sum of minutes at sleep level 1;

TotalTimeInBed represents the sum of minutes across all sleep levels (1+2+3); and TotalSleepRecords equals the count of unique logId values.

NOTE: To ensure accurate interpretation, the numerical encoding used for sleep stages and activity intensity levels was cross-referenced with the descriptive text categories found in the SleepDay and dailyData files, respectively.

PROCESS

DOCUMENTATION OF CLEANING AND MANIPULATION OF DATA

This section details the data cleaning and manipulation applied to the Bellabeat fitness tracking data. Utilizing **Google Sheets** for initial steps and **RStudio** (with lubridate, dplyr, and tidyr) for complex transformations, these procedures ensured data quality and consistency for the subsequent analysis.
The goal was to create a clean and reliable dataset that accurately reflects user activity, calorie expenditure, intensity levels, metabolic equivalents, sleep patterns, and heart rate.

A. KEY ISSUES AND CONSIDERATIONS:

The data cleaning process encompassed several steps:

- **Standardization of temporal data:**

Recognized inconsistencies in date and time formats across different files, in columns named ‘ActivityHour’, ‘ActivityMinute’, and ‘date’ (for sleep and heart rate data). All temporal data was standardized.

Also, combined DateTime columns were split into separate 'Date' and 'Time' columns ensuring uniformity and facilitating accurate temporal comparisons and aggregations.

- **Removal of duplicate records:**

To address potential data redundancy, especially arising from the overlapping data on April 12, 2016, between the two time period files, duplicate entries were identified and removed. This process was performed based on unique User ID and timestamp combinations. This step ensured that each record represented a unique observation for a given user at a specific time.

- **Standardization of time representation for aggregation:**

For accurate aggregation of minute-level data to hourly intervals, the 'Time' column, initially provided in HH:MM:SS format, was transformed to represent the start of each hour (HH:00:00). This ensured that all minute-level records falling within the same hour were correctly grouped for subsequent aggregation.

- **Aggregation of granular data for consistency:**

To create consistent hourly datasets for key metrics, minute-level data was aggregated. This involved:

Intensities: Counting the occurrences of each intensity level (0-sedentary, 1-lightly active, 2-fairly active, 3-very active) within each hour to determine the total minutes spent at each level. This was then pivoted into separate columns for each intensity.

METs: Counting the occurrences of minute-level records falling into predefined MET intensity categories (Low METs: ≤ 1.5, Light METs: 1.6 - 2.9, Moderate METs: 3.0 - 5.9, High METs: ≥ 6.0) within each hour, which were subsequently pivoted into separate columns representing the total minutes in each MET level.

Sleep: Counting the occurrences of each sleep level (1, 2, 3) within each hour to determine the total minutes spent at each level, followed by pivoting these counts into separate sleep level columns.

Heart Rate: Averaging the second-level heart rate values for all records within each hour to generate an hourly average Beats Per Minute (BPM) value for each user.

- **Handling missing values resulting from pivoting:**

Following the pivoting of intensity, METs and sleep level data, null values were replaced with 0. This signifies that for a particular hour, if no minute-level records existed for a specific intensity or sleep level, the duration spent at that level was zero minutes.

B. GENERAL STEPS TAKEN DURING THE CLEANING PROCESS:

- Import of data files corresponding to the two time periods.
- Combination of data from both time periods into a single dataset.
- Splitting of the DateTime column into separate 'Date' (DD/MM/YYYY format) and 'Time' (HH:MM:SS in 24-hour format) columns.
- Removal of duplicate records based on 'Id', 'Date', and 'Time'.
- Transformation of the 'Time' column to represent the start of the hour for aggregation purposes.
- Aggregation of granular (minute/second level) data to an hourly level for Intensities, METs, Sleep, and Heart Rate.
- Categorization of METs values into predefined intensity levels.
- Pivoting of Intensities, METs, and Sleep data so that each value category became a separate column representing total minutes per hour.
- Export of the final processed data for each metric to a CSV file.

C. DETAILED DATA PROCESSING STEPS BY METRIC:

DATA PROCESSING TASKS FOR ‘STEPS’ AND ‘CALORIES’ METRICS, USING SHEETS

- **Locale setting:**

  Ensure the locale setting is configured to “United States” to facilitate consistent interpretation of date and time formats during subsequent conversions.

- **Data import:**

  Import both data files, corresponding to the two time periods, into separate sheets;

  Convert the data in each sheet into a Table object to facilitate structured referencing of the data.

- **Combining data:**

  Create a new sheet to house the combined dataset.

  Use a formula ={Table2[#ALL];Table1[#ALL]} to vertically stack the data from both tables. The order of the tables is crucial for the subsequent duplicate removal process. Records in the second dataset are prioritized due to its more comprehensive data per user compared to the first.

- **Date/Time transformation and splitting:**

  Convert the date/time column to the DateTime data type, by explicitly specifying the format is recognized as dd/mm/yyyy hh:mm:ss (24-hour format).

  To enable splitting, convert the DateTime column to Text format.

  Create one column to the right of the original date/time column

  Split the text-based date/time column into the Date and Hour columns using the "space" character as the delimiter. Label the columns Date and Time.

  Change data type of Date to date (dd/mm/yyyy) and Time to time (hh:mm:ss 24-hour format)

- **Duplicate removal:**

  Select the Id, Date, and Hour columns.

  Select the duplicate removal option: Data → Data clean-up → Remove duplicates to remove duplicate records, particularly those from April 12, 2016, which are present in both datasets from the two time periods

- **Export to CSV:** Export the resulting cleaned dataset into a CSV file: File → Download → Comma-separated values (.csv).

DATA PROCESSING TASKS FOR ‘INTENSITIES’, ‘METs’, ‘SLEEP’, AND ‘HEART RATE’ METRICS, USING RSTUDIO

The processing was performed using R, with the following libraries:

'lubridate': For handling date and time data.

'dplyr': For data manipulation and transformation.

'tidyr': For data restructuring (pivoting).

TASKS APPLICABLE TO ALL METRICS

- **Data import:**

  Import both datasets, corresponding to the two time periods

- **Combining data:**

  The two data files for each metric were combined into a single data frame.

- **Date/Time transformation and splitting:**

  The date/time column was converted into separate 'Date' and 'Time' columns. The 'parse_date_time()' function from the 'lubridate' package was used.

  The 'format()' function was used to extract the date (in "DD/MM/YYYY" format) and time (in "HH:MM:SS" format).

- **Check for missing values**

  The code checks for missing values (NAs) in the 'Date' and 'Time' columns for each metric. The 'is.na()' function was used in conjunction with 'summarise()' from the 'dplyr' package to count the number of missing values in each column.

- **Duplicate removal:**

  Duplicate records were removed based on the combination of 'Id', 'Date', and 'Time' columns.

- **Extract hour from time:**

  The 'Time' column for each metric was processed to extract the hour component using 'floor_date()' and 'format()'.

METRIC-SPECIFIC STEPS:

➔ **INTENSITIES**

The data was aggregated by Id, ActivityDate, ActivityHour, and Intensity using the group_by() and summarize() functions from the dplyr package. For each group, the number of minutes was calculated using the n() function within summarize().

The data was transformed using the pivot_wider() function from the tidyr package. This step is crucial for analyzing time spent in different intensity levels. The Intensity column's values were used to create new columns. The values in the new columns represent the number of minutes spent at each activity level for each hour.

➔ **METs Data**

The 'METs' values were divided by 10 to correct the data.

The 'METs' values were categorized into four categories: "LowMETs", "LightMETs", "ModerateMETs", and "HighMETs".

The data was pivoted, similar to the intensity data. The 'METs_category' column's values were used to create new columns, representing minutes spent in each METs category per hour. The columns were ordered.

➔ **SLEEP**

The data was aggregated by 'Id', 'Date', 'Hour', and 'value'.

For each group, the number of minutes and the number of distinct 'logId' values were calculated.

The data was pivoted.

The row with '11/03/2016' was removed from the Date column, to eliminate out of range value.

➔ **HEART RATE**

The data was aggregated to hourly level, calculating the average heart rate ('Value') for each hour, user, and date.

FINAL STEP:

- **Export to CSV:**

  The final processed data for each metric was exported to a CSV file using the write.csv() function. The row.names = FALSE argument was used to prevent row numbers from being written to the file.

D. REASONS FOR DATA RESTRUCTURING AND AGGREGATION

The transformation process primarily involved restructuring the granular (minute/second level) raw data into a more manageable hourly format for all metrics, with the exception of "calories" and "steps", which were already at an hourly level. This aggregation to an hourly resolution facilitates the analysis of temporal patterns within the 24-hour cycle and enables direct comparisons between different types of activity and their corresponding intensity levels. The dataset underwent consistent formatting, with a specific focus on standardizing date records. Additionally, missing values were identified to improve data cleanliness and facilitate subsequent analytical procedures. This hourly organization provides a structured foundation for potential future aggregation to daily and weekday levels, which will be valuable for identifying trends across broader temporal scales.

E. VERIFICATION OF DATA CLEANLINESS AND READINESS FOR ANALYSIS

To ensure the dataset was adequately clean and prepared for meaningful analysis across the all previous metrics, the data was carefully verified applying the following steps:

- **Temporal consistency:**

  For activity-related metrics (Steps, Calories, Intensities, METs), the data was verified to primarily span the intended period of March 12, 2016, to May 12, 2016, covering the full 24-hour cycle. A minor initial inconsistency in the sleep data's date range, including March 11, 2016, was identified and will be addressed by excluding the out-of-range date. The heart rate data spans from March 29, 2016, to May 12, 2016, indicating missing data from March 12, 2016, to March 28, 2016.

- **Unique user identification:**

  A total of 35 unique user identifiers were consistently identified across the activity-related datasets (Steps, Calories, Intensities, and METs), which is a discrepancy from the initially stated 30 eligible users. This suggests potential inconsistencies in user participation or data submission.

  For the sleep and heart rate datasets, 25 and 15 unique users were identified, respectively.

  Analysis of data completeness revealed significant variability across metrics. Specifically, as noted in the metric-specific overviews, four users across the activity metrics have less than 35 tracked days, and a substantial portion of sleep and heart rate data users have similarly limited records. The smaller number of users in the sleep and heart rate data also necessitates careful consideration regarding the generalizability of findings for those specific metrics.

- **Range checks and plausibility:**

The observed ranges for key continuous variables were assessed for plausibility. Hourly step counts (0 to 10,565) and calorie expenditure (42.16 to 948.49 kcal) exhibited ranges that are conceivable within the context of varying activity levels. Average heart rate values were generally within an acceptable physiological range, although some low and high values (below or above the typical 60-100 bpm) were observed, potentially indicative of highly fit athletes but requiring careful interpretation. Similarly, average nightly sleep duration ranged from 6.5 to 9.5 hours for most users, but notably low averages (around 2 hours) in some cases warrant further investigation.

- **Assessment of overall data completeness:**

  The completeness of records varied significantly across users for all metrics, with a notable subset having substantially fewer than the expected 62 tracked days. For heart rate data, tracked days vary considerably: one user with 44 days, four users with 42 days, and the remaining users with fewer than 40 days, including two with only 5 and 3 tracked days. The implications of this variability will be considered in the analytical phase.

- **Logical integrity checks:**

  For hourly Intensities, METs, and Sleep data, minutes consistently summed to 60 per hour, confirming valid aggregation.

The data cleaning and manipulation procedures outlined before were critical for preparing data for analysis. By addressing inconsistencies, removing redundancies, transforming data to a consistent hourly level (where applicable), and verifying the cleanliness of the resulting datasets, a more reliable and structured foundation has been established for analysing user behavior and trends. The exported CSV files are now ready for further exploration in the subsequent phases of this analysis.

ANALYSE

SUMMARY OF THE ANALYSIS

The analysis of the Bellabeat user data reveals several key patterns in activity, sleep, and heart rates.

Daily activity levels show a clear trend, with most users exhibiting peak activity and calorie expenditure during the daytime hours, typically between late morning and early evening. Step counts, active minutes, and MET levels all increase during this period, while sedentary behavior decreases. Heart rate follows an expected pattern, with lower rates during sleep and elevated rates during waking hours.

A strong positive correlation exists between physical activity and calorie expenditure. Both the duration of activity (active minutes) and the intensity of activity (time spent in higher MET levels) are positively associated with the number of calories burned. While step count also shows a positive correlation, it is less strong than that of active time and intensity.

Weekly activity patterns indicate a tendency for slightly reduced activity across most metrics on Sundays.

Sleep patterns on weekends suggest potentially longer overall sleep duration on Saturdays and a marginal increase in sleep levels 2 and 3 on both Saturday and Sunday nights. The distribution of sleep stages across the night shows light sleep being more prominent in the initial hours, a consistent presence of sleep level 2, and an increase in sleep Level 3 later in the sleep period. The limited and inconsistent sleep data prevent definitive conclusions regarding individual relationships between activity and sleep.

Individual user data reveals notable variations in activity, calorie expenditure, and heart rate that do not consistently align with overall trends.

Analysis of heart rate reveals a strong physiological response of heart rate to changes in physical activity levels. Lower heart rates are characteristic of periods of rest and minimal movement, while even light activity elicits a noticeable increase. As the intensity of activity escalates through fairly active to very active states, the average heart rate continues to rise significantly. This clear relationship highlights the dynamic nature of heart rate as an indicator of the body's physiological response to varying levels of physical exertion throughout the day. Just like sleep records, the scarce heart rate records limits conclusions about relationships between activity and heart rate.

In conclusion, the data indicates a general pattern of daytime activity driving calorie expenditure, with subtle weekly variations. However, significant individual differences and limitations in the consistency and availability of sleep and heart rate data suggest the need for further investigation with more comprehensive information to fully understand user behaviors.

SHARE

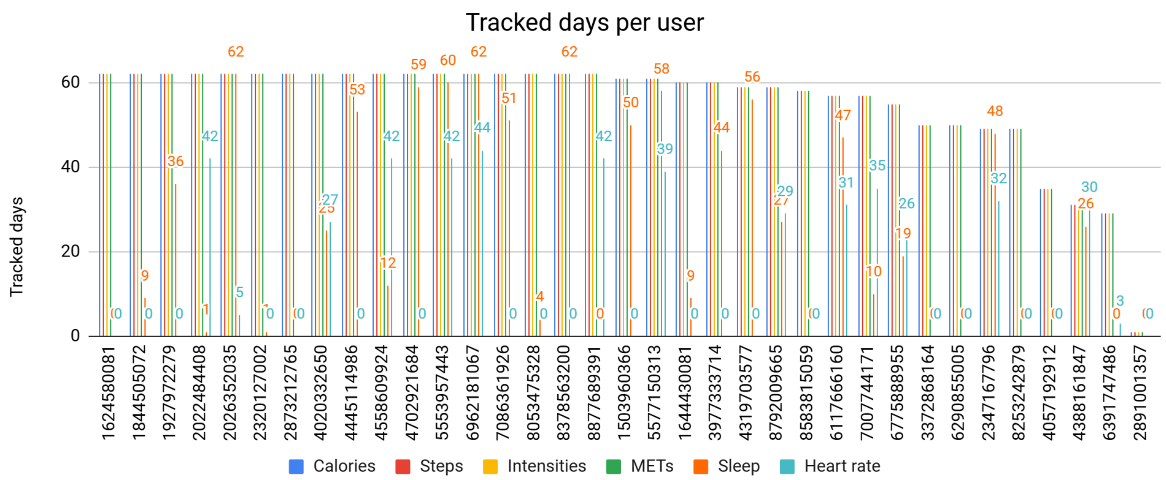
SUPPORTING VISUALIZATIONS AND KEY FINDINGS

A. DATA OVERVIEW: SCOPE & COVERAGE OF ANALYSIS

The analysis utilized smart device usage data collected over a **62-day period**, from March 12th to May 12th, 2016.

There were **35 unique users** that provided activity-related data (steps, calories, active minutes, METs). Sleep data was available for only 25 of these users and heart rate data was available for only 15 of these users.

The extent of data tracking varied considerably among users across all metrics. A substantial portion of users had fewer than the expected 62 tracked days. This ranged from users with just over a month of data to some with severely limited records, including single-day entries.



This chart illustrates the number of tracked days per user for each metric.

The key metrics analyzed included:

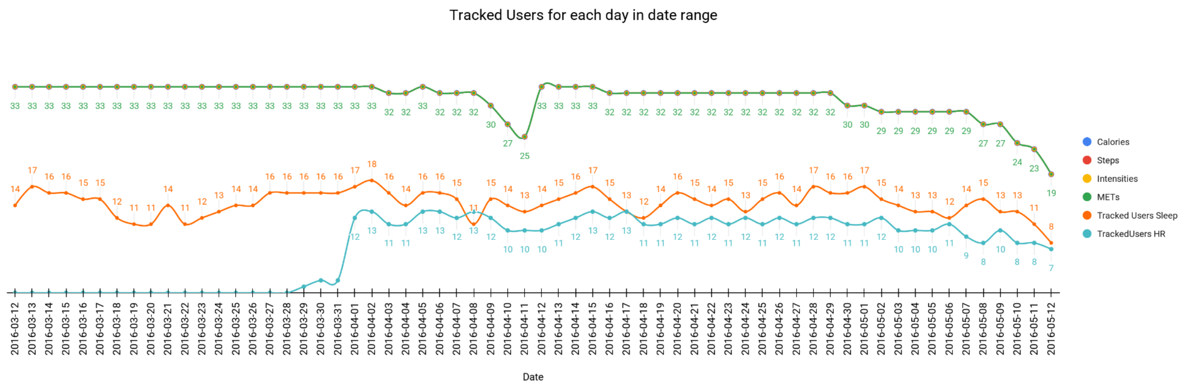
- **Activity**: Daily steps, active minutes (light, fairly, very active), and METs
- **Energy**: Daily calorie expenditure.
- **Sleep**: Time asleep and distribution of sleep stages, from level 1 to 3.
- **Heart Rate**: Hourly heart rate measurements.

B. DAILY USER COVERAGE: CONSISTENCY ACROSS THE DATA RANGE

The number of users tracking activity daily varied significantly, from a maximum of 33 users to a low of 19 users by the end of the period, indicating a drop-off in sustained engagement.

Sleep data shows intermittent daily logging, often below the 25 unique users identified.

Heart rate data is sparser, with daily tracking significantly below 15 unique users, and records starting only from March 29th, indicating a partial tracking period.



This chart illustrates the number of unique users providing activity data each day over the 62-day analysis period.

Implication on daily trends: The fluctuating daily user counts mean that daily average trends for activity, sleep, and heart rate are based on a changing subset of users, impacting the generalizability of day-to-day comparisons.

C. INSIGHTS OVERVIEW

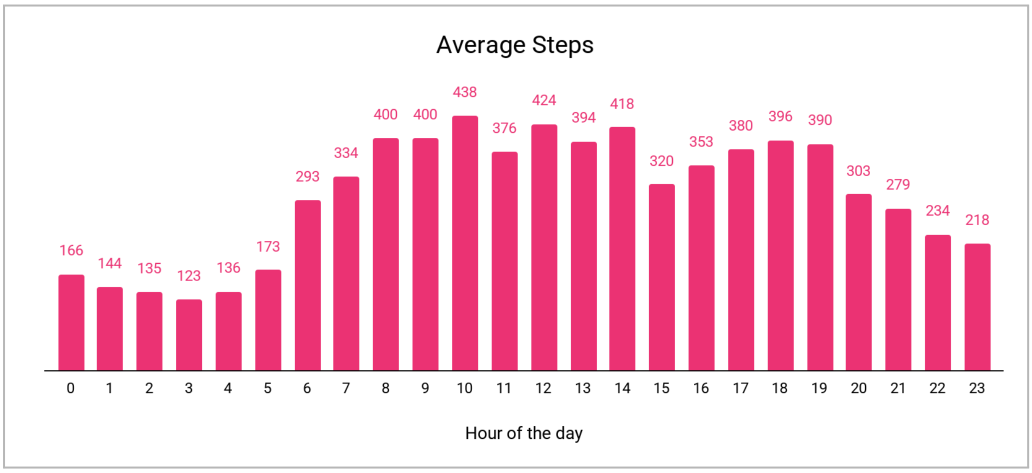
Analysis of Bellabeat smart device usage data reveals consistent daily and weekly activity patterns, a strong correlation between activity and calorie expenditure, and initial insights into sleep behaviors.

However, user behaviors often deviate from overall averages, highlighting unique individual needs, and the depth of analysis for sleep and heart rate was limited by inconsistent data availability.

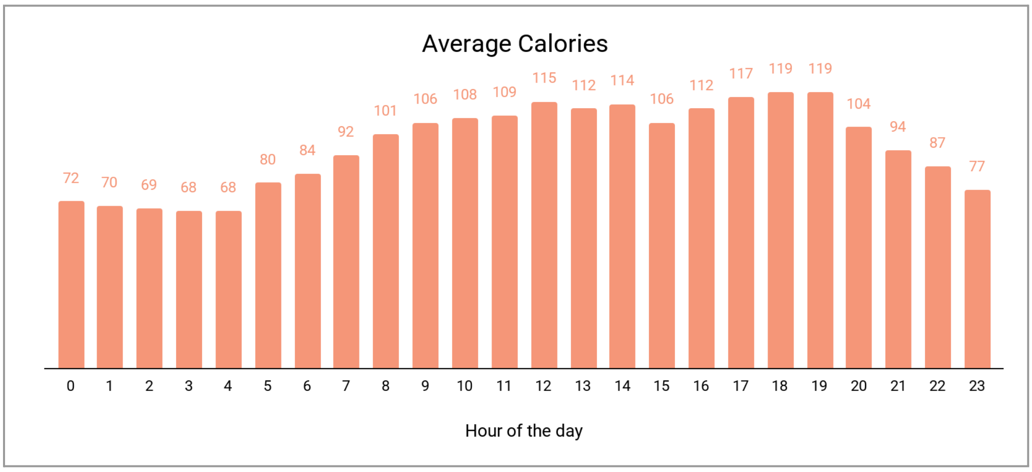
D. UNDERSTANDING DAILY RHYTHMS

Step counts and generally peak during the late morning, around lunchtime, **calorie expenditure and** in the early evening (8-10 AM, 12-2 PM, and 6-7 PM).

Calorie expenditure mirrors step count trends, with the highest burn observed around midday and early evening, indicating a direct relationship between movement and energy use.

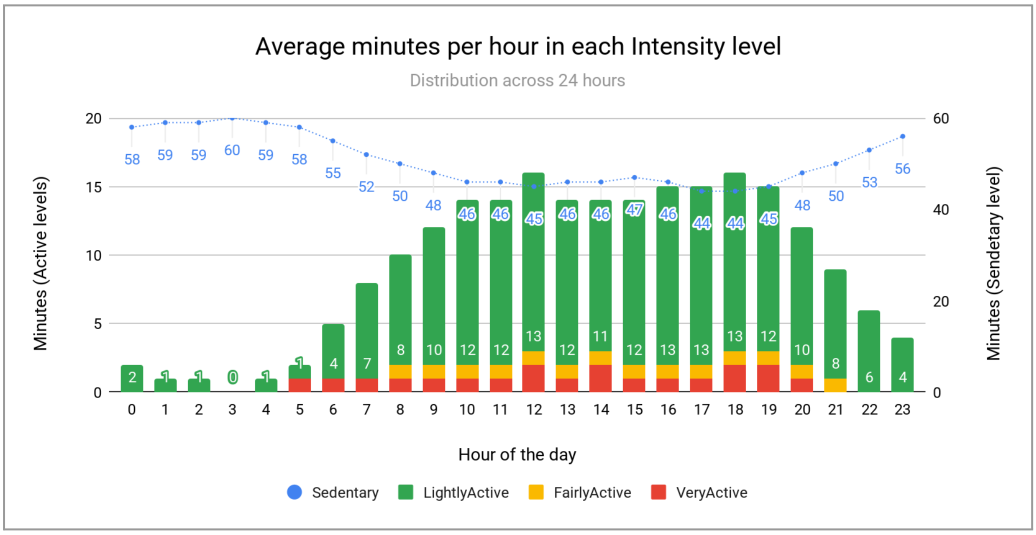


Average daily step count distribution across 24 hours

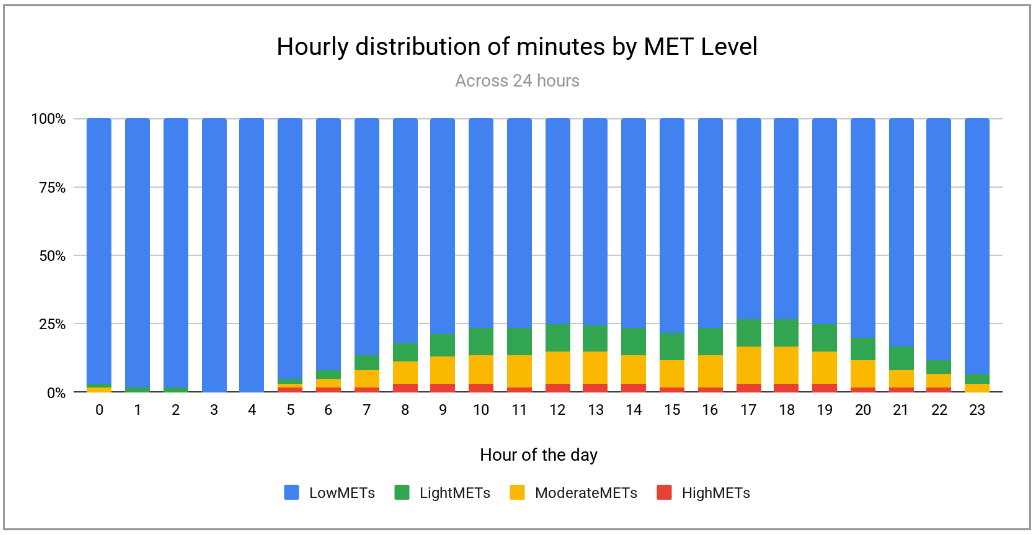


Average daily calorie expenditure distribution across 24 hours

Examination of **activity levels** indicates an inverse relationship between sedentary behavior and active engagement throughout the day.

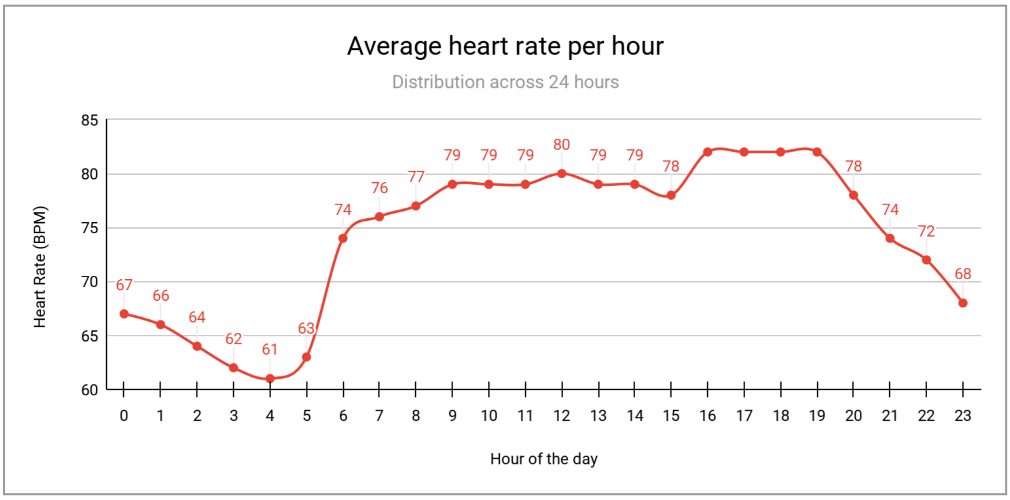


The amount of time spent in sedentary activities decreases as the day progresses, while the total minutes spent in active levels (LightlyActive, FairlyActive, and VeryActive) increase. Notably, the peaks in total active minutes align with the peaks in calorie expenditure.

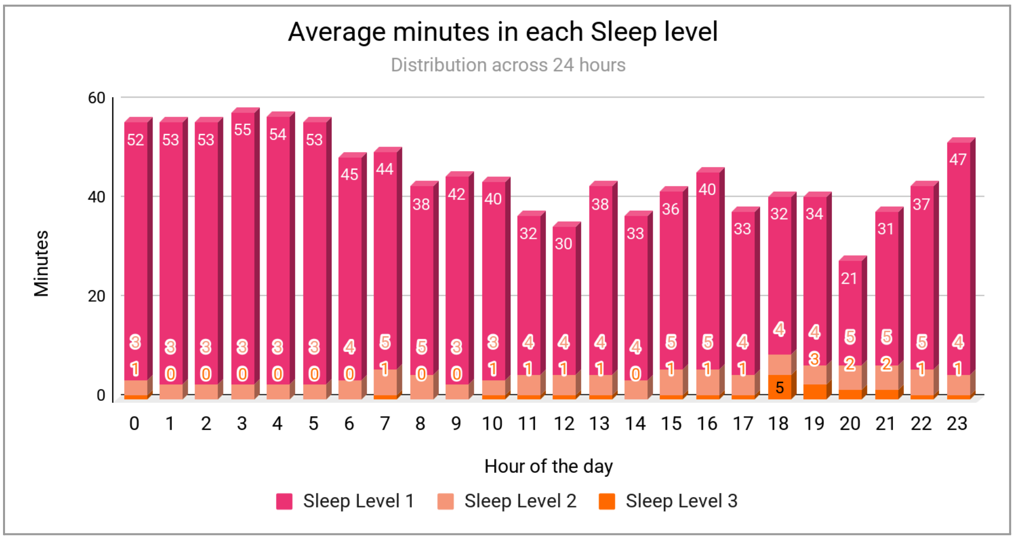


The 100% stacked bar chart for METs further supports this, illustrating a decreasing proportion of sedentary time (Low METs ≤ 1.5) and an increasing proportion of time spent in higher MET intensity levels as the day unfolds.

Regarding heart rates, the lowest averages are recorded during the nighttime hours (0AM and 5AM), ranging from 61 to 67 bpm; a sharp increase occurs at 6AM, rising to an average of 74 bpm. It then remains relatively stable throughout the main part of the day, from 7AM to 8PM, fluctuating between 74 and 82 bpm, before gradually declining in the evening hours leading up to sleep.



Hourly analysis reveals distinct **sleep** stage patterns. Sleep Level 1 dominates late night and early morning (23:00-05:00), peaking around 04:00, then becoming less prominent and more variable. Sleep Level 2 maintains a consistent presence, averaging 3-5 minutes hourly throughout the day. Conversely, Sleep Level 3 is minimal from 00:00-05:00 but begins to increase around 06:00, showing significant fluctuations and peaking substantially around



18:00.

The analysis of hourly data across key metrics indicates a clear pattern of increased physical activity and energy expenditure during the daytime hours, with specific peaks in the morning, around lunchtime, and in the early evening. Sedentary behavior is lowest during these active periods. Heart rate has the lowest values observed during sleep and a sustained elevation during the waking hours.

E. UNDERSTANDING ACTIVITY INTENSITY ACROSS THE WEEK

Across all days, the most active periods are concentrated between approximately 16:00 and 19:00, representing when users accumulate the most active minutes.

Average active minutes per hour across each day of the week

	Hour of the day																							
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Mon	2	1	1	0	1	3	6	10	11	12	14	13	15	13	12	11	13	16	17	16	12	8	5	3
Tue	2	1	1	0	1	4	6	10	10	12	14	14	16	13	12	11	14	16	17	16	13	11	8	4
Wed	2	1	1	0	1	3	6	9	11	12	12	13	15	13	14	13	15	17	17	16	12	9	7	4
Thu	2	1	1	0	0	3	7	9	12	13	14	13	15	13	14	13	14	16	15	15	12	11	7	4
Fri	2	1	1	1	1	3	7	10	11	12	11	13	14	13	13	12	14	15	16	16	13	9	8	5
Sat	3	2	1	1	1	1	2	6	9	13	15	17	17	17	16	16	15	15	15	15	12	10	8	5

Sun	3	2	1	1	0	1	2	5	8	12	15	14	15	15	16	15	14	15	13	14	11	8	5	3

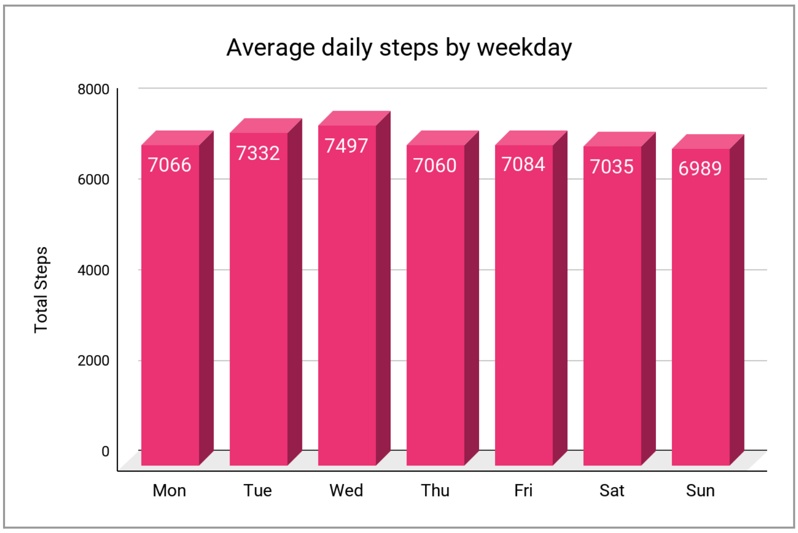
Activity levels are significantly lower during nighttime hours (22:00 - 06:00), indicating periods of rest and sleep.

Weekdays (Monday to Friday) show a more focused midday peak (around noon) and sustained activity between 17:00 and 19:00.

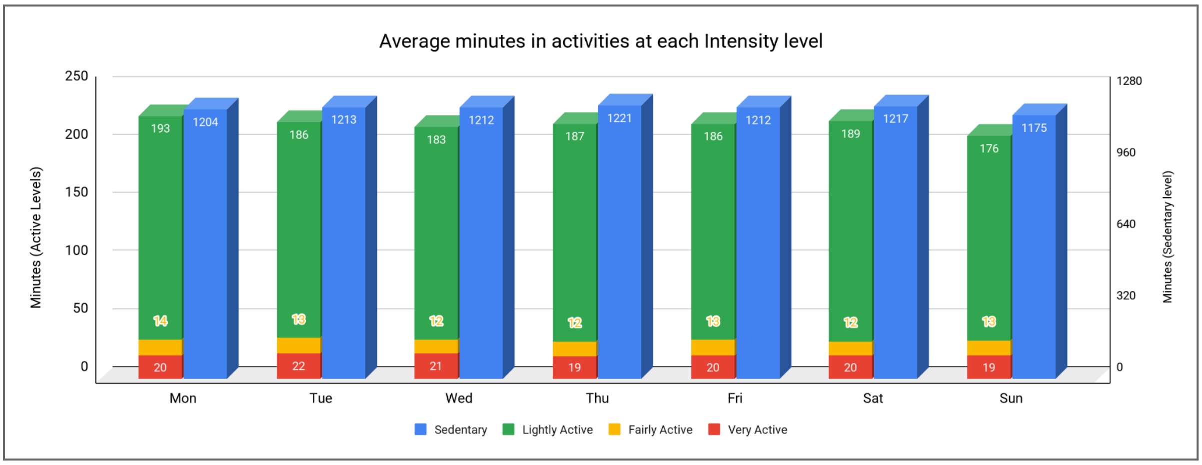
Weekends exhibit a broader, more sustained period of activity throughout the middle of the day (10:00 to 19:00), possibly reflecting more leisure-based activities.

F. WEEKEND SHIFT: ACTIVITY AND SLEEP

Average daily step counts and active minutes tend to be slightly lower on weekends compared to weekdays, with Sundays often showing the lowest average activity.

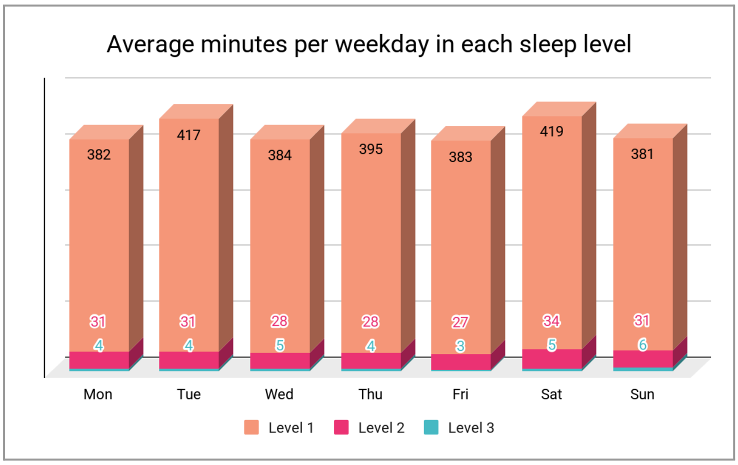


Comparison of average daily steps between weekdays and weekends



Comparison of average daily active minutes between weekdays and weekends

Average sleep duration may be slightly longer on Saturdays, suggesting users might prioritize rest on weekends.



Comparison of average daily sleep duration between weekdays and weekends.

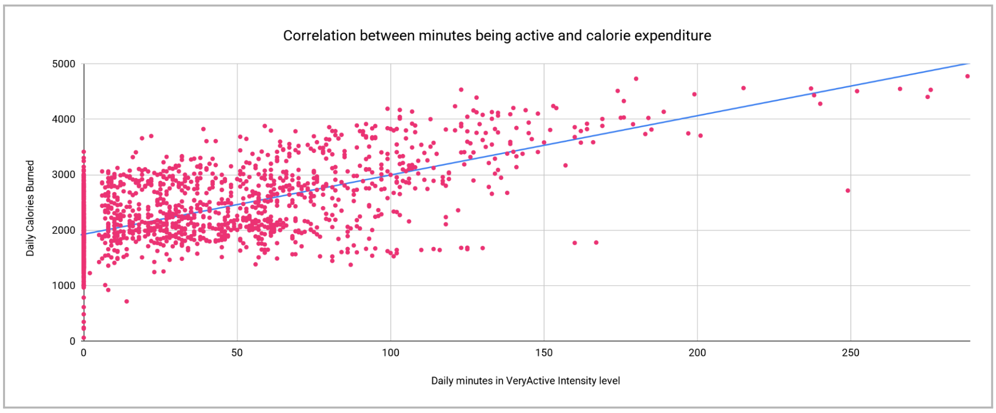
G. ACTIVE ENGAGEMENT FUELS CALORIE EXPENDITURE

Both daily step count and total active minutes show a clear positive relationship with calories burned, indicating that increased activity leads to higher energy expenditure.

CORRELATION	Calories	StepTotal	ActiveMinutes	HighMETsMinutes
Calories	1.00	0.46	**0.67**	**0.69**
StepTotal		1.00	0.56	0.65
ActiveMinutes			1.00	**0.73**
HighMETsMinutes				1.00

Table: Correlation factor between variables

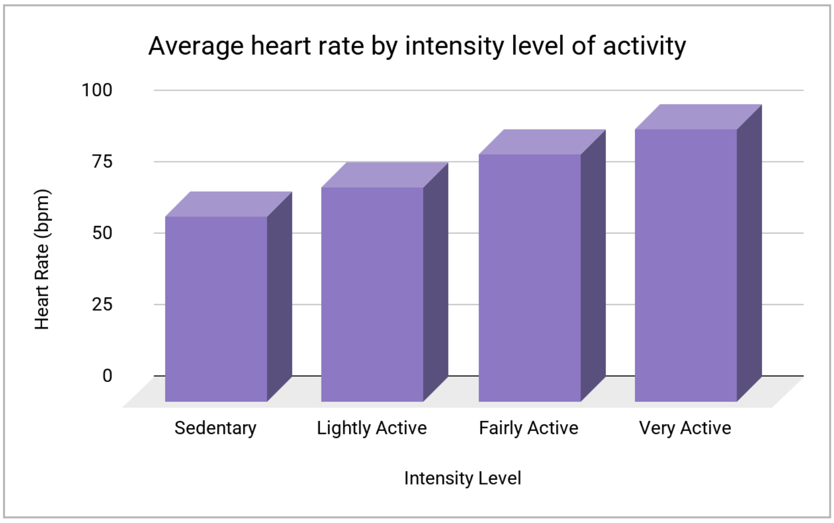
The correlation between total active minutes and calories burned is stronger than that between total steps and calories, suggesting that the duration and intensity of movement are more influential than just the sheer number of steps.



Analysis confirmed that minutes spent in high METs activities have an even stronger link to calorie expenditure, emphasizing the efficiency of higher intensity workouts.

H. HEART RATE VARIABILITY IN RELATION TO ACTIVITY LEVELS

Analysis reveals that the level of physical activity influences heart rate. As users transition from sedentary to increasingly intense activity levels, a corresponding and progressive increase in their average heart rate is observed.

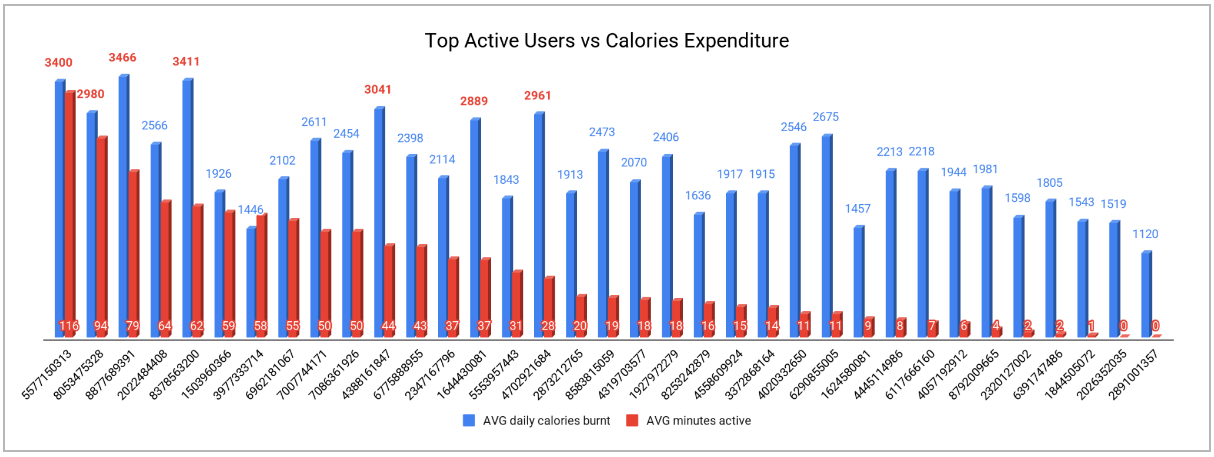


The average heart rate during Very Active periods (95 BPM) is approximately 31 BPM higher than during Sedentary periods (64 BPM). This underscores the considerable impact of physical effort on the cardiovascular system.

I. USER-SPECIFIC TRENDS:

ACTIVITY & CALORIE EXPENDITURE NUANCES

A deeper look into individual user data reveals variations that challenge broad assumptions.

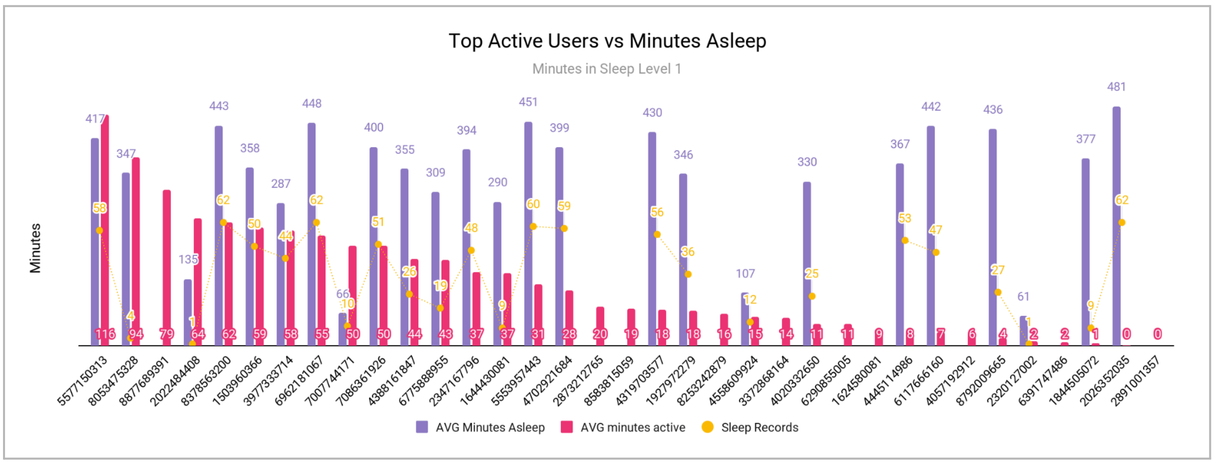


User 4702921684 exhibited a high daily average calorie expenditure (nearly 3000 kcal) with a surprisingly low average of only 28 active minutes per day. In contrast, User 1503960366 averaged 1926 calories per day despite logging a significantly higher 59 active minutes daily.

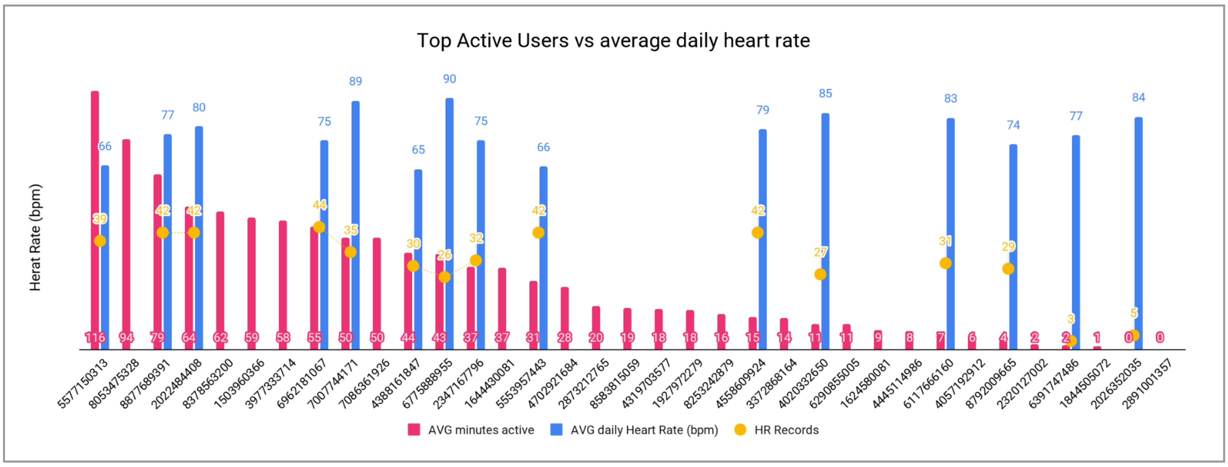
This suggests that factors beyond just logged "active minutes" (like individual metabolic rate, intensity of non-tracked activity, or unique data logging behaviors) play a significant role in individual calorie expenditure.

SLEEP & HEART RATE DATA LIMITATIONS

The expected link between increased activity and longer sleep wasn't consistently supported due to limited and sparse sleep data. Sleep records were significantly scarcer than activity data. For instance, User 8053475328 had 62 days of activity but only 4 days of sleep. This inconsistency prevents a robust comparison of activity and sleep at the individual level.



Heart rate records are even rarer, available for only 15 unique users, making broad inferences challenging. Individual heart rate averages (active User 5577150313 at 56 bpm vs. less active User 611766160 at 83 bpm) do not consistently align with activity minutes, suggesting physiological differences that cannot be fully explored with the limited data.



CONCLUSION

This analysis of Bellabeat smart device usage data has revealed several key insights into user behavior, highlighting both opportunities and challenges for the company. While users demonstrate consistent daily and weekly activity patterns, with a clear link between activity and calorie expenditure, significant individual variability exists. Furthermore, the limited and inconsistent data for sleep and heart rate presents a challenge for developing a truly holistic understanding of user well-being.

ACT

TOP HIGH-LEVEL INSIGHTS BASED ON ANALYSIS

Based on the previous analysis of user behavior patterns, the following data-driven recommendations are proposed to enhance Bellabeat's product and marketing strategies:

ENGAGEMENT TIMING
Send notifications and challenges during the times users are most active. This way, they're more likely to see and act on them.

SHOW HOW ACTIVITY BURNS CALORIES
Clearly explain how more active minutes and intense workouts lead to burning more calories. This helps users focused on fitness and see the real value of their Bellabeat device.

MAKE IT PERSONAL
Create app features that give users tailored insights and goals. Everyone's different, so the app should adapt to their unique activity and body responses.

IMPROVE DATA COLLECTION
Focus on ways to get more consistent and complete sleep and heart rate data. Better data here means we can offer users richer, more personalized health advice and features.

By acting on these recommendations, Bellabeat can enhance user engagement, strengthen its value proposition, and drive sustainable growth in the competitive health and wellness market.