MY BELLABEAT CASE STUDY USING EXCEL AND TABLEAU

TABLE OF CONTENTS

Contents

INTRODUCTION

Bellabeat is a high-tech driven manufacturer of health-focused smart products for women, aiming to empower them with knowledge about their health and habits. It is a successful small company, and has the potential to become a larger player in the global smart device market. Their app and devices collect data on activity, sleep, stress, menstrual cycle, and mindfulness habits.

To unlock new growth opportunities in the global smart device market, Bellabeat aims to analyze smart device fitness data. They are keen to utilize consumer data for strategic insights and want to focus on analyzing the usage patterns of one of their products to optimize their marketing strategies: Bellabeat app.

BUSINESS TASK

Examining user data from non-Bellabeat smart devices that track activity metrics like steps and calories burned, as well as sleeping patterns. The objective is to derive meaningful insights into user behavior and how they engage with their smart devices to monitor their activities. These insights, in the form of trends identified from the data, will be utilized to inform and shape upcoming marketing strategies.

The key stakeholders are:

- Urska Srsen – Co founder and Chief creative officer
- Sando Mur – Co founder and mathematician
- Bellabeat marketing analytics team

DATA SOURCES

A public data set named **Fitbit fitness tracker data** containing the personal fitness data of 30 Fitbit users, made available by Mobius and stored on Kaggle website.

LICENSING AND PRIVACY OF DATA

The data has been confirmed to be open source and so can be used without asking for permission under copyright laws, the platform where the data is hosted, is recognized for its security measures and operates under appropriate licensing.

DATA ORGANIZATION

The 18 CSV documents are available with each document representing different quantitative data tracked by Fitbit.

Table Name	Type	Description
dailyActivity_merged	Excel CSV	Daily Activity tracking (Steps, Distance, Intensities, Calories) for 33 users over 31 days
dailyCalories_merged	Excel CSV	Daily Calorie data for 33 users over 31 days
dailyIntensities_merged	Excel CSV	Daily Intensity (Sedentary, Lightly Active, Fairly Active, Very Active) for 33 users over 31 days
dailySteps_merged	Excel CSV	Daily Steps data for 33 users over 31 days
heartrate_seconds_merged	Excel CSV	Heart Rate logs for 7 users
hourlyCalories_merged	Excel CSV	Hourly Calorie data for 33 users over 31 days
hourlyIntensities_merged	Excel CSV	Hourly Intensity data for 33 users over 31 days
hourlySteps_merged	Excel CSV	Hourly Steps data for 33 users over 31 days
minuteCaloriesNarrow_merged	Excel CSV	Calorie data captured every minute for 33 users over 31 days (Every minute in a single row)
minuteCaloriesWide_merged	Excel CSV	Calorie data captured every minute for 33

		users (Every minute in a single column) over 31 days
minuteIntensitiesNarrow_merged	Excel CSV	Intensity data by minute for 33 users (Every minute in a single row) over 31 days
minuteIntensitiesWide_merged	Excel CSV	Intensity data by minute for 33 users (Every minute in a single column) over 31 days
minuteMETsNarrow_merged	Excel CSV	METs (Energy Ratio) data for 33 users over 31 days
minuteSleep_merged	Excel CSV	Sleep logs by minute for 24 users over 31 days
minuteStepsNarrow_merged	Excel CSV	Steps tracked every minute for 33 users (Every minute in a single row) over 31 days
minuteStepsWide_merged	Excel CSV	Steps tracked every minute for 33 users (Every minute in a single column) over 31 days
sleepDay_merged	Excel CSV	Daily sleep logs (count, minutes, time in bed) for 24 users over 31 days
weightLogInfo_merged	Excel CSV	Daily weight tracking (Kg, Pounds) with BMI calculation for 8 users over 30 days

OBJECTIVES

Considering the datasets at my disposal, I intend to analyze user behavior primarily within daily time frames to identify overarching trends. I will concentrate on examining the daily activity and sleep datasets, sorting and filtering them using an Excel spreadsheet.

CREDIBILITY OF THE DATASET

The data originates from a reliable source, but it's worth noting that the sample size is quite small, comprising only 30 users over 31 days. This small sample size raises concerns about potential sampling bias. Additionally, an important caveat is the absence of any demographic information in the dataset.

TOOLS USED

I made use of Excel as my primary tool for data cleaning and manipulation due to its relative ease of use and accessibility and Tableau for visualization.

IMPORTING DATASETS INTO EXCEL

Importing and optimizing FitBit Fitness Tracker data sets involves the removal of redundant datasets and aggregation of essential datasets required for querying and analyzing specific information. Considering my emphasis on identifying high-level trends in the data for analysis, I opted to import and exclusively utilize the "Dailyactivities_merged", "SleepDay_merged" and "Hourlysteps_merged" datasets.

I created a CHANNGELOG documenting my cleaning and formatting process within Excel CLICK HERE

DATA CLEANING AND MANIPULATION

In order to better visualize my data and prepare it for analysis, I made sure to check for errors and inconsistencies in the data frames that are the most relevant to solving the business task (Dailyactivities_merged, Sleepday_merged and Hourlysteps_merged). I performed this process in an Excel worksheet

Dailyactivities_merged

- **Date Formatting**: Formatting column (Activitydate**)** to Date.
- **Number Formatting**: Formatted columns (ID, Totalstpeps, Totaldistance, Trackerdistance, Veryactivedistance, Moderatelyactivedistance, Lightlyactivedistance, Sedentaryactivedistance, Loggedactivedistance, Fairlyactiveminutes, Lightlyactiveminutes, Sedentaryminutes) to Number formats and also rounded up their figures to a uniform 2 decimal place.

TotalDistance	TrackerDistance
8.5	8.5
6.96999979	6.96999979
6.739999771	6.739999771
6.28000021	6.28000021
8.159999847	8.159999847
6.480000019	6.480000019
8.590000153	8.590000153
9.880000114	9.880000114
6.679999828	6.679999828
6.340000153	6.340000153
8.130000114	8.130000114
9.039999962	9.039999962
6.409999847	6.409999847
9.800000191	9.800000191
8.789999962	8.789999962
12.21000004	12.21000004
8.529999733	8.529999733
7.150000095	7.150000095
9.25	9.25
6.809999943	6.809999943
9.710000038	9.710000038

TotalDistance	TrackerDistance
14.12	14.12
10.29	9.48
8.50	8.50
7.57	7.57
8.41	8.41
7.49	7.49
7.77	7.77
6.74	6.74
7.78	7.78
6.83	6.83
5.98	5.98
6.12	6.12
6.08	6.08
5.20	5.20
6.05	6.05
5.88	5.88
4.43	4.43
3.39	3.39
2.74	2.74
2.20	2.20
1.64	1.64
0.47	0.47
0.00	0.00

- Duplicates: The dataset contains no duplicates
- Null values: Used the function COUNTBLANK() to search for null values but found none.
- Added a new column called “Daysoftheweek” using the function TEXT (B2,” DDDD”) to extract days of the week from each individual date for further analysis.

Id	ActivityDate
1503960366	4/12/2016
1503960366	4/13/2016
1503960366	4/14/2016
1503960366	4/15/2016
1503960366	4/16/2016
1503960366	4/17/2016
1503960366	4/18/2016
1503960366	4/19/2016
1503960366	4/20/2016
1503960366	4/21/2016
1503960366	4/22/2016
1503960366	4/23/2016
1503960366	4/24/2016
1503960366	4/25/2016
1503960366	4/26/2016
1503960366	4/27/2016
1503960366	4/28/2016
1503960366	4/29/2016
1503960366	4/30/2016
1503960366	5/1/2016
1503960366	5/2/2016

Id	ActivityDate	day of the week
8053475328	4/12/2016	Tuesday
7007744171	4/12/2016	Tuesday
1503960366	4/12/2016	Tuesday
5553957443	4/12/2016	Tuesday
7086361926	4/12/2016	Tuesday
2320127002	4/12/2016	Tuesday
1644430081	4/12/2016	Tuesday
6962181067	4/12/2016	Tuesday
4388161847	4/12/2016	Tuesday
2347167796	4/12/2016	Tuesday
3977333714	4/12/2016	Tuesday
4020332650	4/12/2016	Tuesday
5577150313	4/12/2016	Tuesday
4319703577	4/12/2016	Tuesday
8378563200	4/12/2016	Tuesday
4702921684	4/12/2016	Tuesday
1844505072	4/12/2016	Tuesday
4558609924	4/12/2016	Tuesday
2026352035	4/12/2016	Tuesday
4445114986	4/12/2016	Tuesday
8792009665	4/12/2016	Tuesday
1927972279	4/12/2016	Tuesday
6117666160	4/12/2016	Tuesday

Sleepday_merged

- Date Formatting: Formatting column (Sleepday) to Date
- Number formatting: Formatted columns (id, Totalstepsrecords, Totalminutesasleep, Totaltimeinbed) to Number style formats.
- Duplicates: Found and removed 3 duplicates, leaving 410 unique values.

Hourlysteps_merged

- Date Formatting: Formatted column (Activityhour) to data style
- Numeric Formatting: Formatting columns (id, Stepstotal) to a number style format
- Duplicates: The dataset contains no duplicates
- Added column named “HourOfDay” extracting the hour format (h:mm AM/PM) from ActivityHour column to be used in later analysis.

Id	ActivityHour	HourOfDay
1503960366	4/12/2016 0:00	12:00:00 AM
1503960366	4/12/2016 1:00	1:00:00 AM
1503960366	4/12/2016 2:00	2:00:00 AM
1503960366	4/12/2016 3:00	3:00:00 AM
1503960366	4/12/2016 4:00	4:00:00 AM
1503960366	4/12/2016 5:00	5:00:00 AM
1503960366	4/12/2016 6:00	6:00:00 AM
1503960366	4/12/2016 7:00	7:00:00 AM
1503960366	4/12/2016 8:00	8:00:00 AM

DATA ANALYSIS

- ## USER ID COUNT

When using pivot tables to go through the intended datasets to be analyzed, I discovered that although 30 user ids were the supposed ideal number of users documented, the actual user count found was slightly varied in my respective data sets; Dailyactivity_merged (33 users), Hourlysteps_merged (33 users) and Sleepday_merged (24 users) respectively.

This discrepancy likely arises from several factors. In the case of the dailyactivities and hourlysteps datasets, the higher-than-expected number of users (33) could be attributed to certain users utilizing multiple smart devices. On the other hand, the lower-than-expected number of users (24) in the Sleepday dataset may result from some users choosing not to share their activity tracking information publicly. This variance in user counts can be attributed to these specific scenarios.

Row Labels	COUNT ID
1503960366	33
1624580081	
1644430081	
1844505072	
1927972279	
2022484408	
2026352035	
2320127002	
2347167796	
2873212765	

Row Labels	COUNT ID
1503960366	33
1624580081	
1644430081	
1844505072	
1927972279	

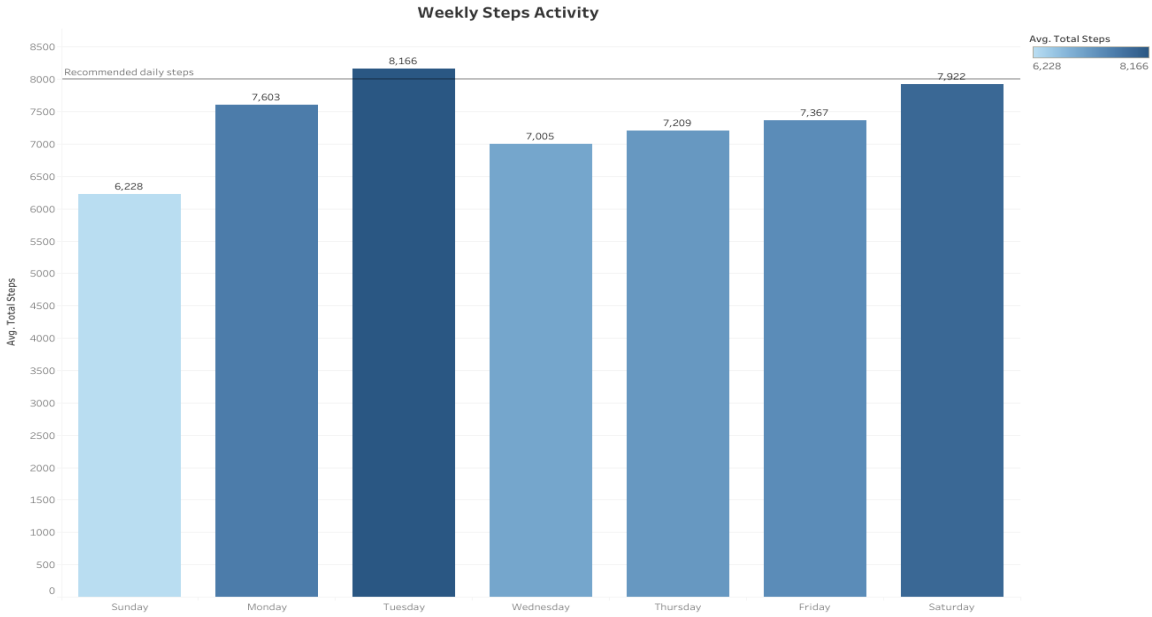
Row Labels	COUNT ID
1503960366	24
1644430081	
1844505072	
1927972279	
2026352035	
2320127002	

DATA VISUALIZATION IN TABLEAU

Subsequently, I imported the refined dataset into Tableau, leveraging its capabilities to craft informative and interactive visualizations and dashboards. This approach allows for a more profound comprehension of the dataset under examination

- ## WEEKLY STEPS ACTIVITY

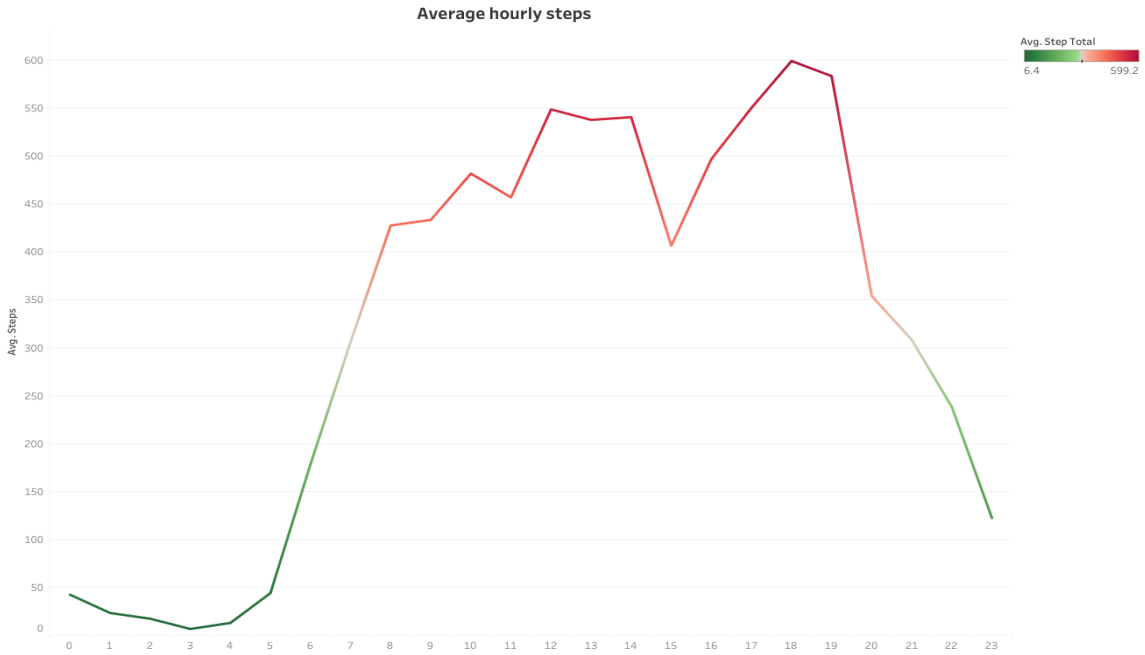
Utilizing the "dailyactivities_merged" dataset, I will construct a plot in tableau that visualizes the distribution of total steps per weekday. This analysis aims to provide insights into user behavior patterns regarding their smart device usage, specifically identifying the most and least active days of the week.



Based on the above plot, we can observe that, on average, Tuesday emerges as the most active day, while Sunday appears to be the least active. This discrepancy is likely influenced by the fact that Sunday is typically associated with being a day of rest and so many users are likely to be at home, engaging in less physical activity. Saturday, however, ranks closely behind Tuesday, possibly due to the weekend effect where people often engage in health-focused activities like running and exercising, leading to a notable increase in their overall activity levels. On average, weekdays exhibit higher user engagement in various activities, as reflected by their average total daily steps.

- ## AVERAGE HOURLY STEPS

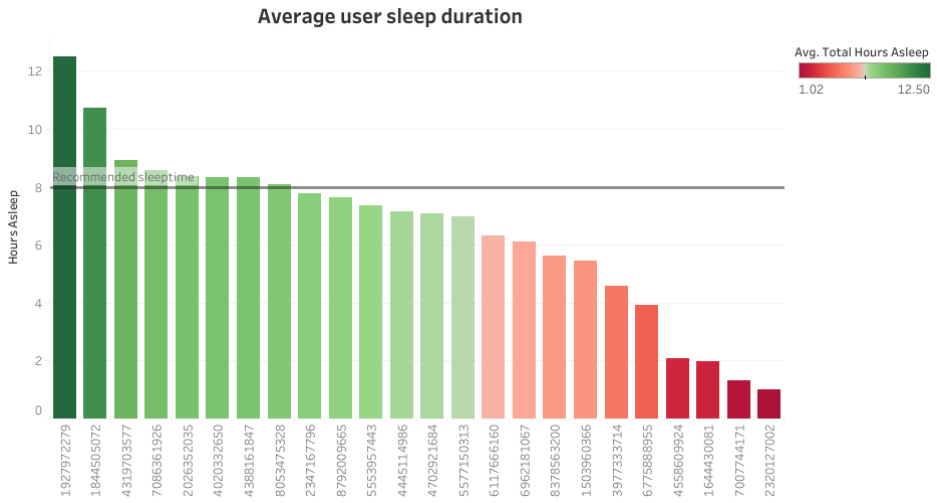
Following my data analysis, I identified Tuesdays as the most active day of the week. Now, I will delve deeper to determine, on average, the most active hours of the day for the Fitbit users.



The visualization provides a detailed breakdown of daily activity levels, focusing on the average steps taken by Fitbit users. It's evident that users are typically more active during daytime hours, starting from 8 a.m. and continuing until around 7 p.m. The activity levels tend to decrease after 8 p.m. On average, we observe peak activity during two distinct periods: from 12 p.m. to 2 p.m. and from 5 p.m. to 7 p.m. Interestingly, the least active time during the 24-hour period is at 3 a.m., which aligns with the common sleeping hours for most people.

- ## USER SLEEP QUALITY

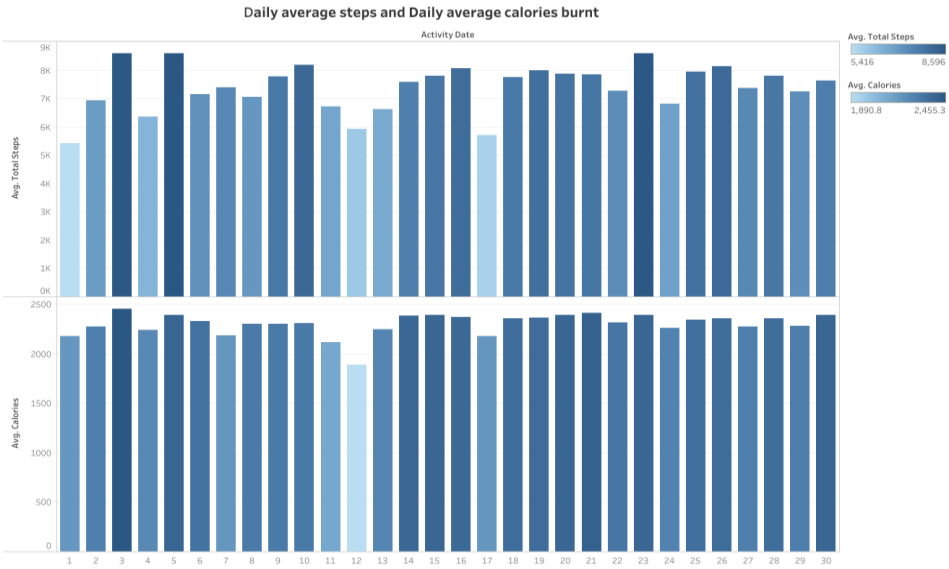
The duration of user sleep is a crucial parameter tracked by smart devices and holds a lot of significance health wise. Leveraging the available data, I constructed a visualization that provided deeper insights into user sleep patterns.

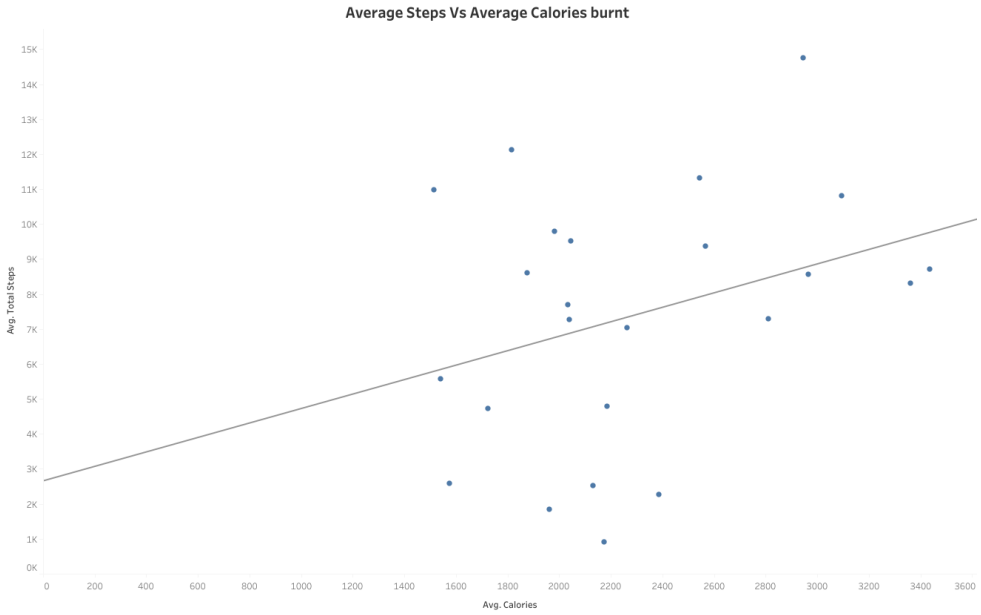


The visualization above indicates a notable trend among Fitbit users, with the majority failing to attain sufficient sleep, as their average sleep durations consistently fall below the recommended daily 8-hour threshold.

- CORRELATIONS

I will attempt to determine if there is a relationship between steps taken by users and the number of calories burnt.





From the first visualization above, it's not immediately apparent how the daily average total steps and daily calories relate when considered individually. Consequently, opting for a scatter plot to underscore their connection reveals a positive correlation between these variables. This correlation arises from the fact that increased walking (total steps) is a form of physical activity, leading to a higher expenditure of calories. It's important to emphasize, however, that correlation does not imply causation, signifying that one variable is not the sole cause of changes in the other, and vice versa.

To view more interactive charts and dashboards on Tableau CLICK HERE.

CONCLUSION

Based on my analysis, I've observed that users tend to be more active on weekdays compared to weekends, with Tuesdays being the peak day for activity. Daily activity shows peaks between 12 PM to 2 PM and 5 PM to 7 PM, while nighttime exhibits lower activity levels, likely indicating rest. It's worth noting that most Fitbit users do not achieve the recommended 8 hours of sleep, with a substantial number falling short of this target. Furthermore, I found a positive relationship between total steps and total calories, as calories burned is directly linked to the number of steps taken. However, it's essential to emphasize that correlation doesn't imply causation, meaning that total steps alone do not cause increased calorie burn and vice versa.

RECOMMENDATIONS

- My recommendation for Bellabeat is to implement improved engagement strategies to maintain user interest in their activity levels. One effective approach could involve introducing a daily steps streak system, which users can share on social media. This would not only motivate individuals to stay consistent in their physical activity but also encourage their friends and family to participate, fostering a sense of community and collective commitment to personal health.

- Bellabeat should consider providing users with access to their historical activity data for informational purposes. This will empower users to take charge of their individual fitness journeys, allowing them to monitor their progress and track their improvements over time. This access to past activity data can serve as a valuable tool for users in achieving their health and fitness goals.

- The Bellabeat app could enhance user experience by offering features that provide notifications to guide users on optimal bedtime preparations (wind down), ensuring they consistently achieve the recommended sleep duration. This proactive approach can contribute to better sleep habits and overall health for app users.

- A beneficial strategy for Bellabeat would be to periodically deliver health tips to users, emphasizing the benefits of adhering to the daily recommended steps and sleep patterns. This serves as a gentle reminder to encourage users to make mindful decisions that contribute to their overall well-being.