How can a fitness technology company play it smart?

Senura Charith Attanayake

2024-05-22

Business Task

Bellabeat, a high-tech manufacturer of health-focused products for women, wants to unlock new growth opportunities by analyzing smart device data. The key business task is to analyze usage patterns from non-Bellabeat smart devices, derive insights on consumer behavior, and apply these findings to improve Bellabeat's products. The goal is to inform Bellabeat's marketing strategy with data-driven insights and provide high-level recommendations to the executive team on how to optimize their product offerings and market positioning.

Key questions driving the analysis:

1. What are the emerging trends in smart device usage?
2. How can these trends be applied to Bellabeat's customer base?
3. What marketing strategies can be recommended based on these trends to increase Bellabeat's market presence?

The Data

The data used for this analysis comes from a public dataset on smart device usage, specifically the Fitbit Fitness Tracker data. This dataset includes daily activity logs, heart rate, sleep monitoring, and other personal health metrics from 30 Fitbit users who consented to share their data.

Data sources:

Fitbit Fitness Tracker Data: A dataset from Kaggle (https://www.kaggle.com/datasets/arashnic/fitbit) containing detailed information on physical activity, heart rate, and sleep patterns.

Data Structure:

- dailyActivity_merged.csv: Daily step counts, distance, calories burned, and activity levels.
- dailyCalories_merged.csv: Caloric expenditure data on a daily basis.
- dailySteps_merged.csv: Number of steps taken per day.
- heartRate_seconds_merged.csv: Heart rate data at minute-level granularity.
- sleepDay_merged.csv: Data on sleep duration and sleep efficiency.

The dataset allows for detailed exploration of physical activity, sleep patterns, and heart rate, providing rich insights into how users interact with their devices and health data.

Limitations:

It should be noted that this dataset used has significant limitations as this was based on only 30 consenting participants. This is not remotely close to conducting a comprehensive analysis. However, Based on this data we can get a brief idea about certain patterns and trends which may serve as a starting point for further assessment.

For certain dataset only about 8 participants provided comprehensive details, This is not enough to get a thorough idea about user behavior

Data preprocessing

In this preprocessing step, we perform multiple actions to clean and prepare the Fitbit dataset for analysis. Here’s a breakdown of what happens during this stage:

Library Loading:

The tidyverse library is loaded, which includes essential packages for data manipulation, such as ggplot2, dplyr, and readr.

Data Loading:

Various Fitbit CSV datasets are imported using read_csv(). Each dataset contains different information related to users’ daily activities, heart rates, sleep patterns, etc. These datasets are assigned to variables like dailyActivity, dailyCalories, and heartRateSeconds, allowing us to work with them in subsequent analysis steps. Missing Value Detection:

- The code then checks for missing values across all the imported datasets * using colSums(is.na()). This step helps in identifying any gaps in the data, which is critical for maintaining data quality in the analysis. In particular, the weightLogInfo dataset has missing values in the “Fat” column, which are addressed by removing the entire column using select(-Fat). Exploratory Data Summarization:

The skimr and janitor packages are used to generate a clean summary of the datasets. Functions like skim_without_charts() and summary() provide essential information such as the data’s distribution, mean, and other statistical details for each column. This ensures that we have a solid understanding of the structure and content of the datasets before diving into deeper analysis. ### Summary Statistics:

Summary statistics are printed for each dataset, providing a quick snapshot of key metrics such as the number of entries, minimum, maximum, and average values for numeric fields. This helps identify patterns and potential anomalies in the data. Date and Time Handling:

The lubridate package is used to handle date-time formats. For example, the Time column in heartRateSeconds is converted to a POSIXct format to facilitate time-based analysis, enabling accurate plotting of heart rate over time.

Data Transformation:

The mutate() function from dplyr is used to calculate new columns based on existing data. For example, in dailyActivity, the total active minutes are calculated, and activity levels are converted into percentages. This transformation makes it easier to interpret how users spend their time (e.g., sedentary vs. very active). Data Cleaning:

- Cleaning actions include removing unnecessary columns (like the “Fat” column from weightLogInfo) and correcting data types (such as converting date columns to proper date formats). These steps ensure that the data is ready for effective analysis without errors or inconsistencies.*

Analysis

This analysis aims to explore and visualize various aspects of Fitbit data, including daily activities, heart rates, sleep patterns, and weight logs. The goal is to derive meaningful insights that can help users understand their health and activity levels.

We start by importing the Libraries

```
# Load Required Packages
install.packages("tidyverse")
```

```
## Installing package into 'C:/Users/Senura/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
```

```
## package 'tidyverse' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Senura\AppData\Local\Temp\RtmpMt9lwY\downloaded_packages
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.3.3
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✓ dplyr     1.1.4     ✓ readr     2.1.4
## ✓ forcats   1.0.0     ✓ stringr   1.5.1
## ✓ ggplot2   3.4.4     ✓ tibble    3.2.1
## ✓ lubridate 1.9.3     ✓ tidyr     1.3.0
## ✓ purrr     1.0.2
```

```
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✕ dplyr::filter() masks stats::filter()
## ✕ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
install.packages("here")
```

```
## Installing package into 'C:/Users/Senura/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
```

```
## package 'here' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Senura\AppData\Local\Temp\RtmpMt9lwY\downloaded_packages
```

```
library(here)
```

```
## Warning: package 'here' was built under R version 4.3.3
```

```
## here() starts at C:/Users/Senura/Projects/Bellabeat Data Analysis/Data-analysis-project-for-bellabeat
```

```
install.packages("skimr")
```

```
## Installing package into 'C:/Users/Senura/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
```

```
## package 'skimr' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Senura\AppData\Local\Temp\RtmpMt9lwY\downloaded_packages
```

```
library(skimr)
```

```
## Warning: package 'skimr' was built under R version 4.3.3
```

```
install.packages("janitor")
```

```
## Installing package into 'C:/Users/Senura/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
```

```
## package 'janitor' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Senura\AppData\Local\Temp\RtmpMt9lwY\downloaded_packages
```

```
library(janitor)
```

```
## Warning: package 'janitor' was built under R version 4.3.3
```

```
##
## Attaching package: 'janitor'
##
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

Each CSV data set is stored in a variable

Check for missing values in each dataset

```
# Check for missing values
missing_values <- lapply(list(dailyActivity, dailyCalories, dailyIntensities, dailySteps,
                              heartRateSeconds, hourlyCalories, hourlyIntensities,
                              hourlySteps, minuteCaloriesNarrow, minuteCaloriesWide,
                              minuteIntensitiesNarrow, minuteIntensitiesWide,
                              minuteMETsNarrow, minuteSleep, minuteStepsNarrow,
                              minuteStepsWide, sleepDaily, weightLogInfo),
                        function(x) colSums(is.na(x)))
```

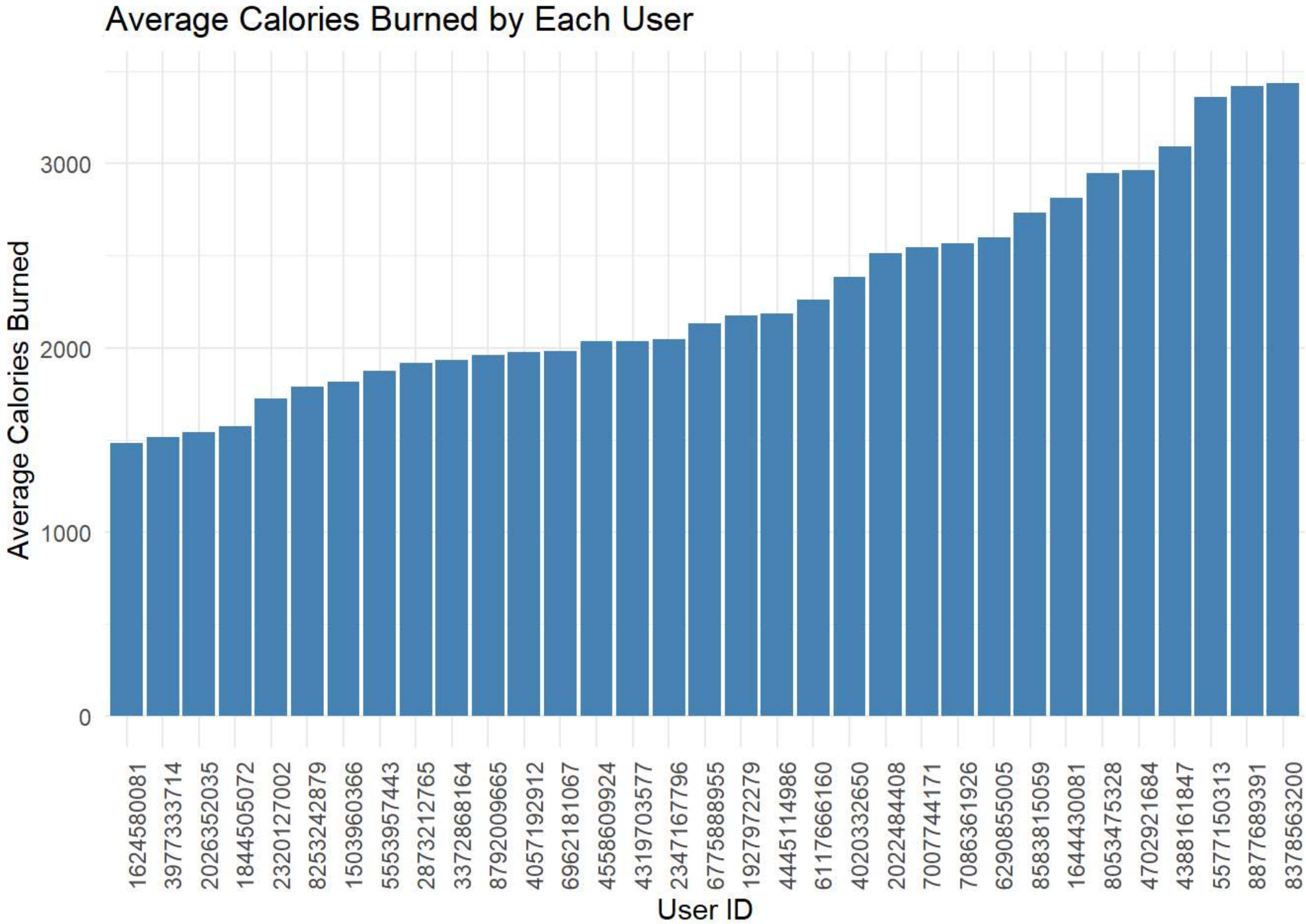
We look at the summary statistics for each dataset

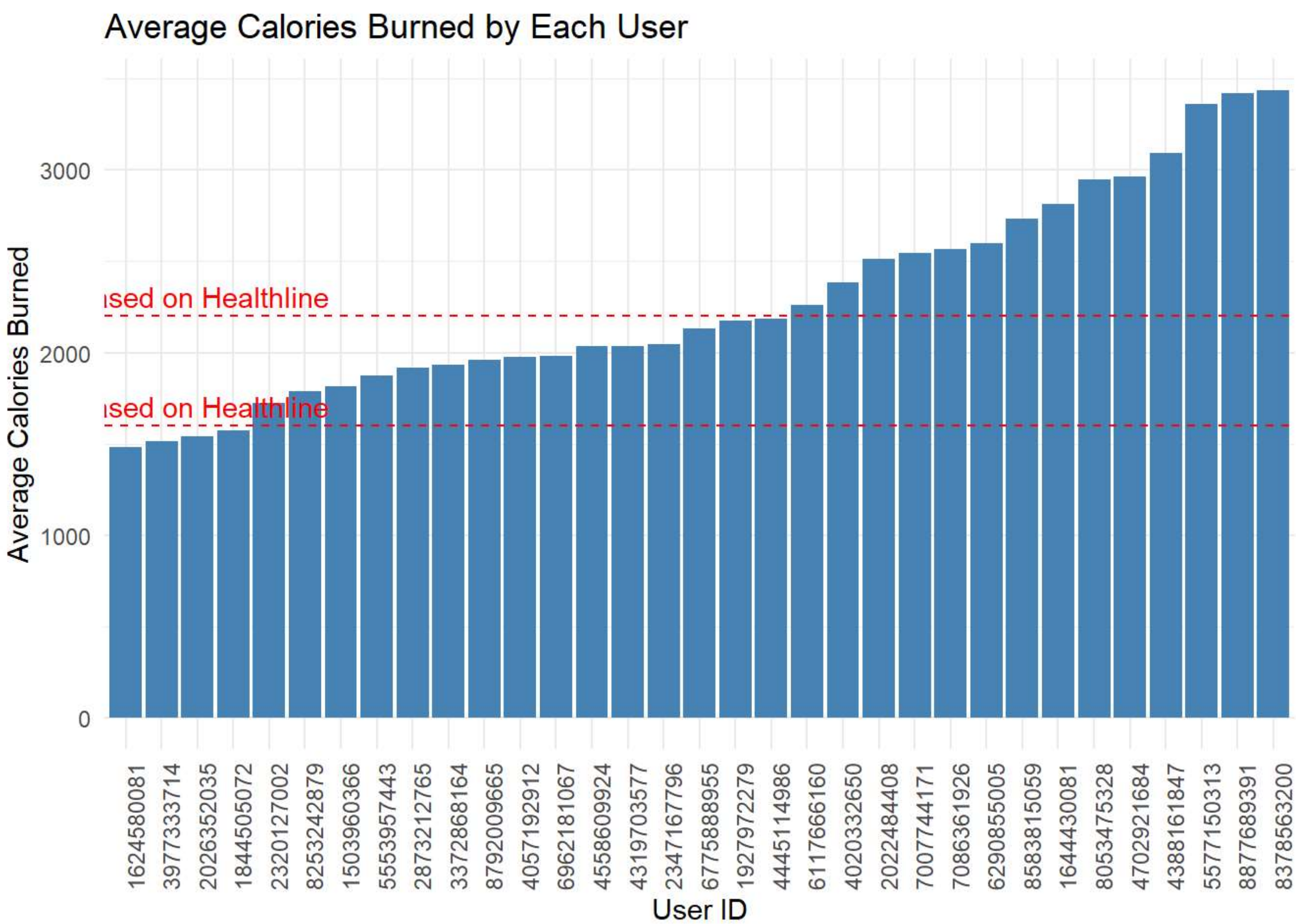
```
# Check for missing values
# Summary statistics for all datasets
lapply(list(dailyActivity, dailyCalories, dailyIntensities, dailySteps,
            heartRateSeconds, hourlyCalories, hourlyIntensities,
            hourlySteps, minuteCaloriesNarrow, minuteCaloriesWide,
            minuteIntensitiesNarrow, minuteIntensitiesWide,
            minuteMETsNarrow, minuteSleep, minuteStepsNarrow,
            minuteStepsWide, sleepDaily, weightLogInfo), summary)
```

Daily Activity Analysis

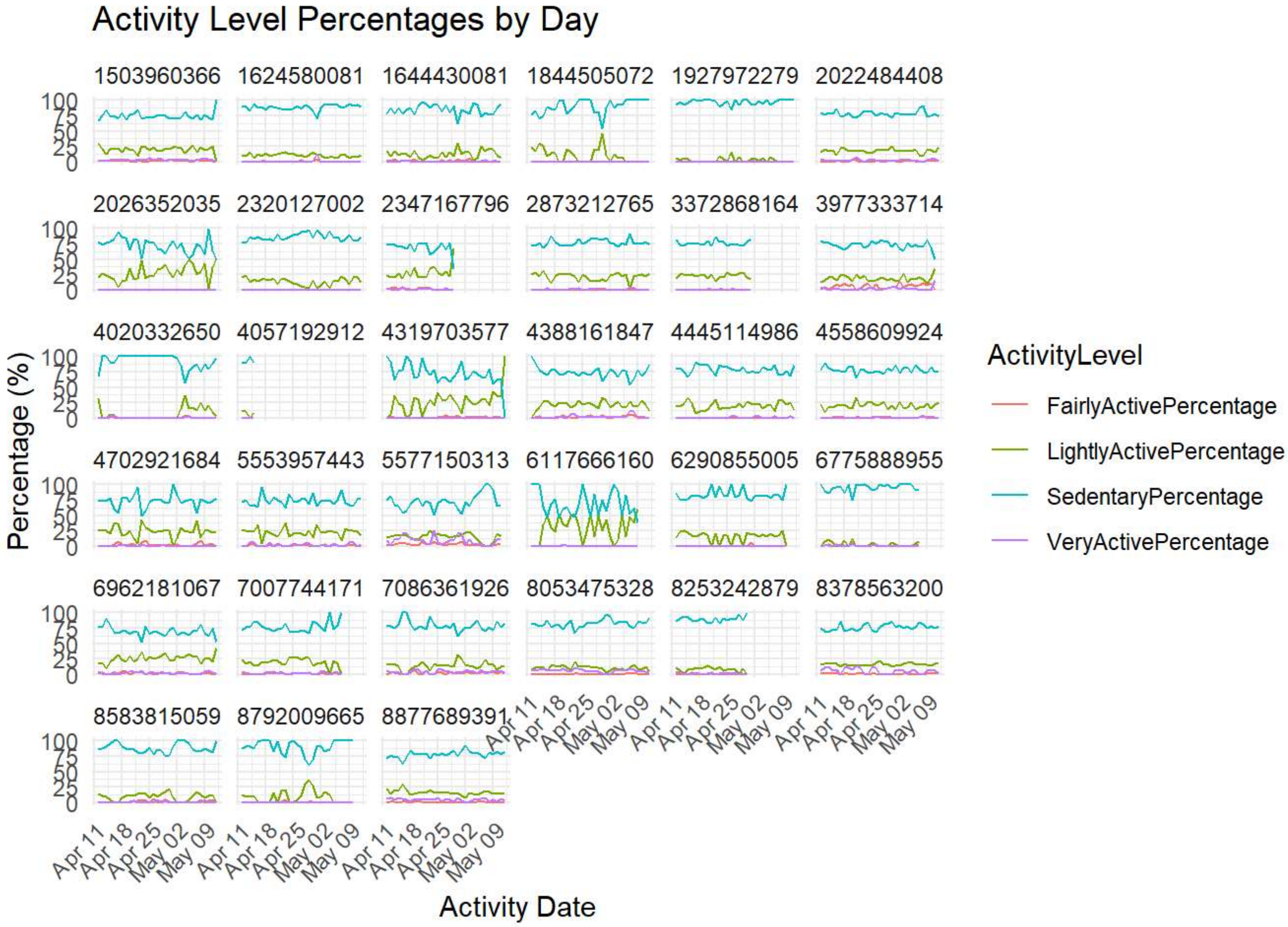
Average Calories Burned

We calculate the average calories burned per user and visualize the results.





Explain the significance of calories burned and its implications for user health.

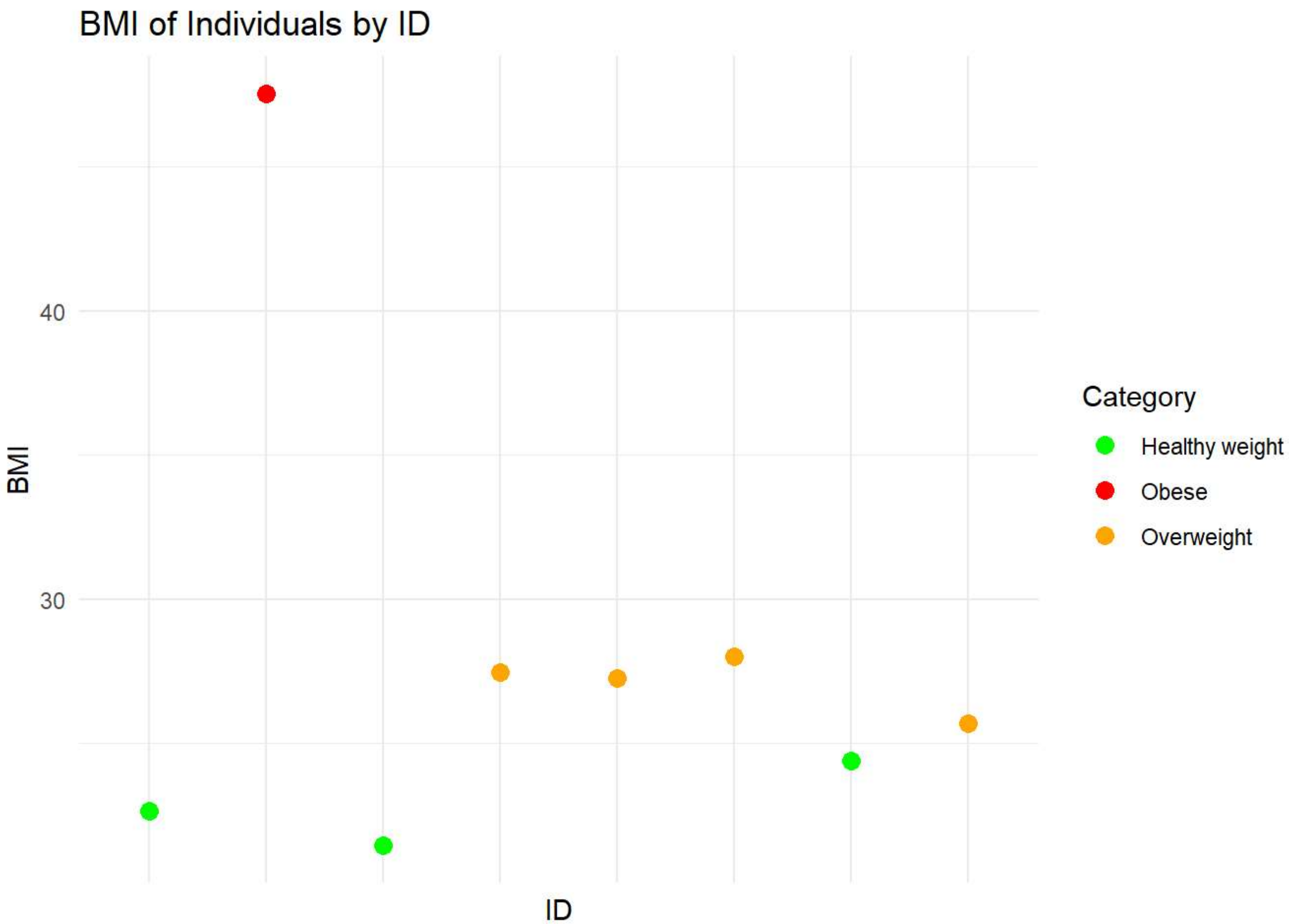


Discuss how different activity levels affect overall health and well-being.

Weight Log Analysis

BMI Calculation

We compute BMI categories based on the earliest records.

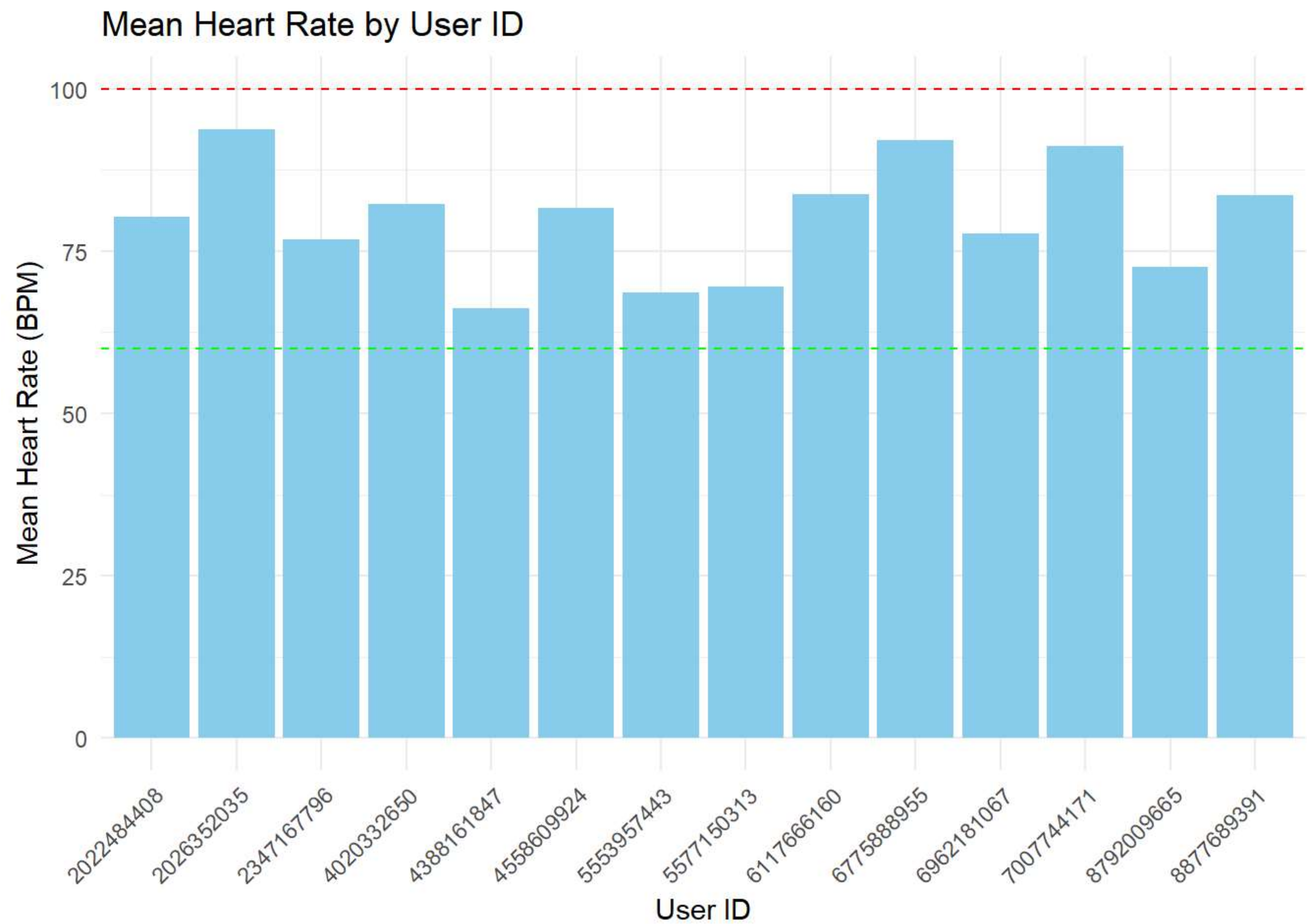


Interpret the BMI categories and their health implications.

Heart Rate Analysis

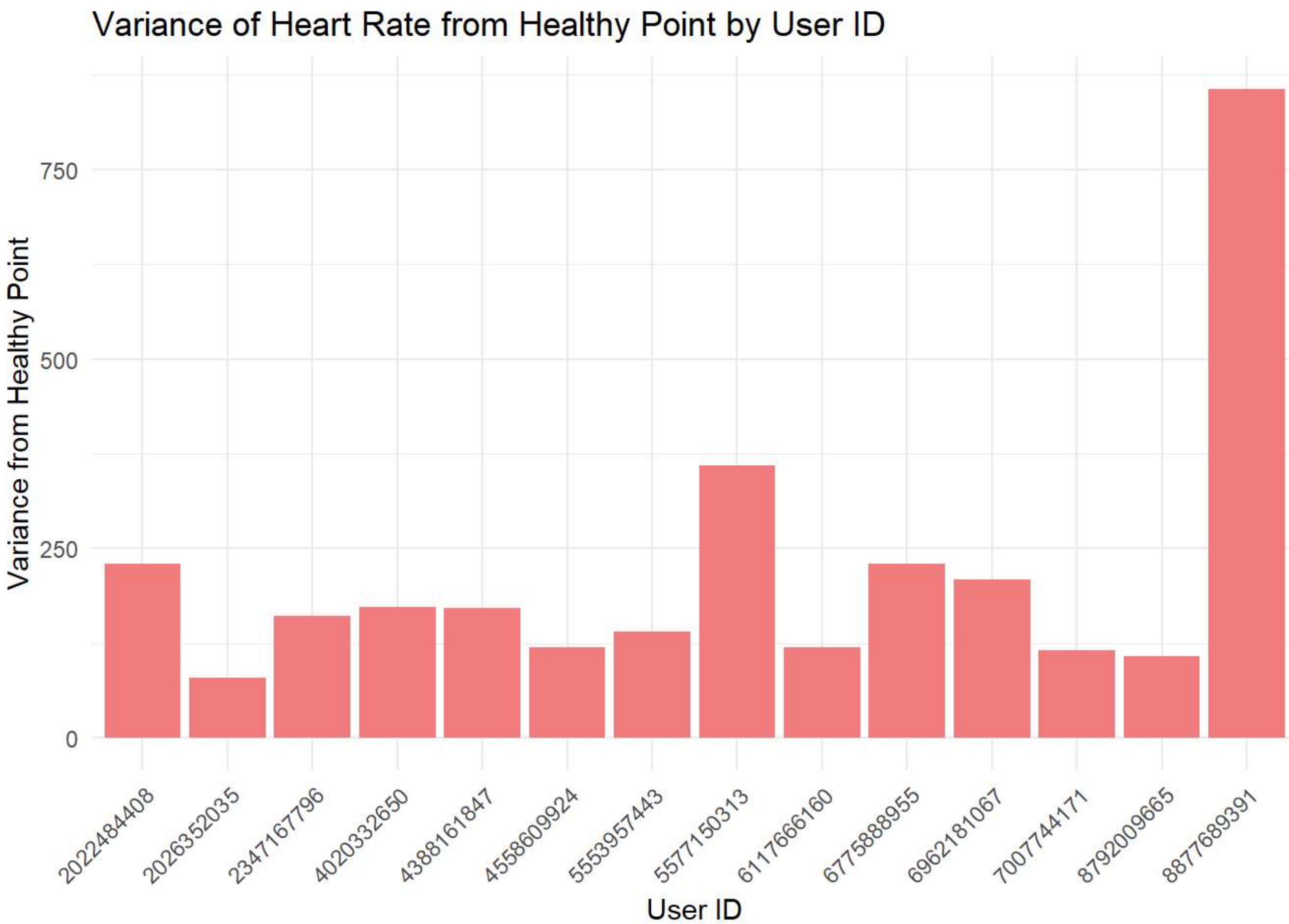
Mean and Variance of Heart Rate

We analyze heart rate data, focusing on the mean and variance.



Discuss the importance of heart rate monitoring in assessing cardiovascular health.

Variance of Heart Rate

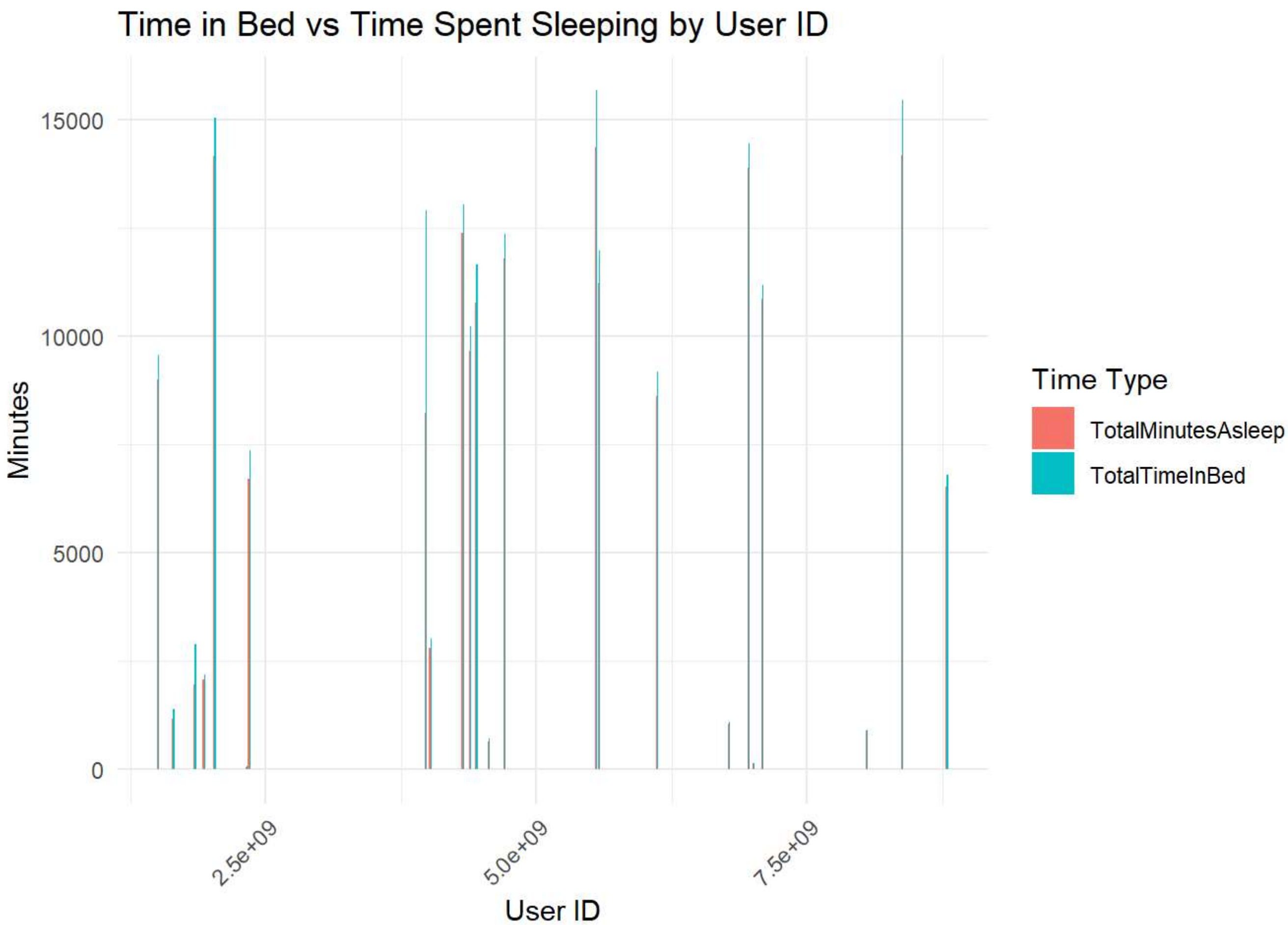


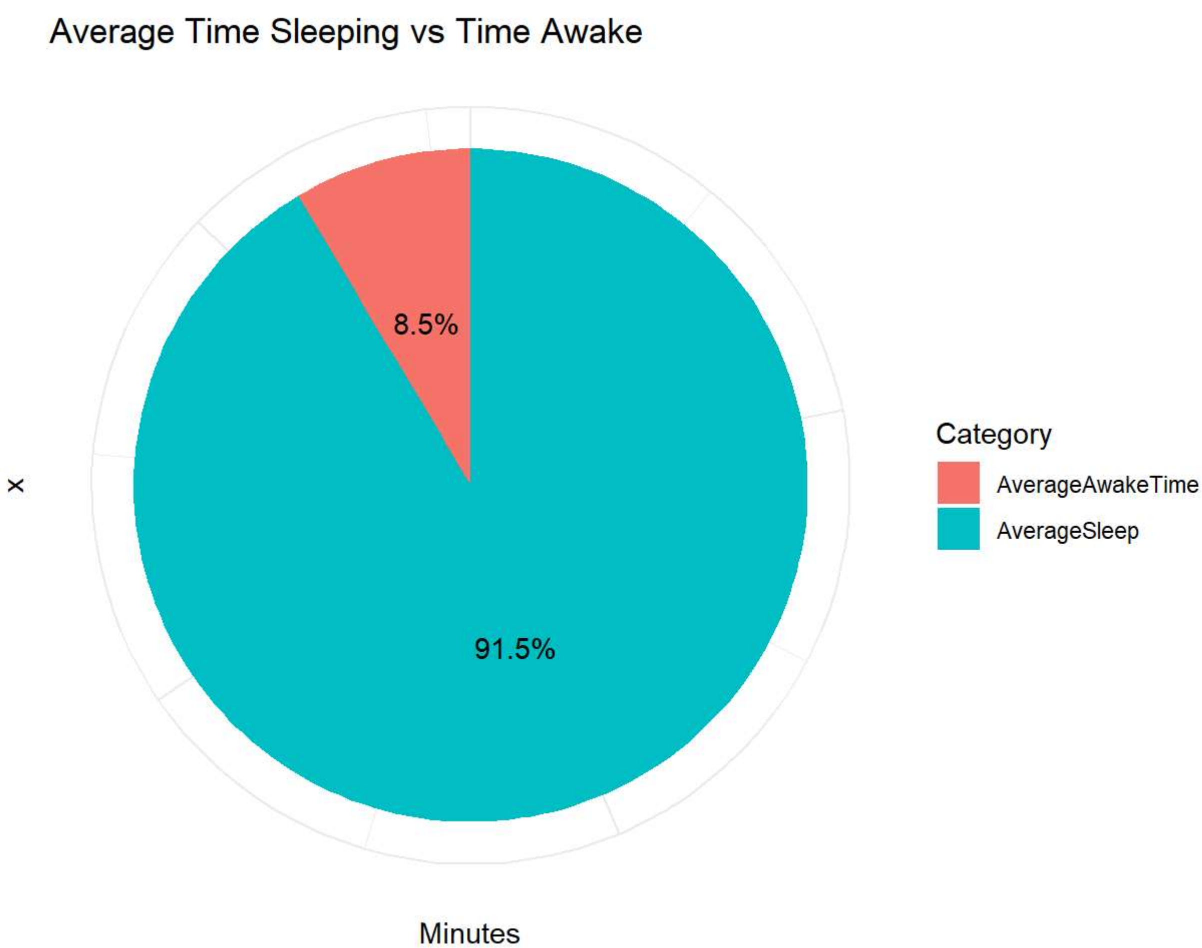
*Elaborate on the significance of heart rate variance for health monitoring.

Sleep Analysis

Sleep Patterns

We evaluate the total time in bed versus the time spent sleeping.





Elaborate on the impact of sleep quality on overall health and productivity.

Summary of Analysis

Conclusion

*The provided dataset is not enough for a complete analysis.

*As for the analyzed user group the majority has a healthy calorie consumption (Calories burned). When comparing with a lower limit of 1600 calories and upper limit of 2200 calories as suggested in this healthline article (https://www.healthline.com/health/fitness-exercise/how-many-calories-do-i-burn-a-day#:~:text=You%20burn%20calories%20daily%20when,your%20body%20and%20activity%20levels.) , The majority of the participating females burn a healthy amount of calories.

*Based on the activity level analysis of each participant, The Sedentary time spent is significantly higher than time spent doing light Activity, Intense activities and time spent being fairly active. The time spent being fairly active is seen to be higher than time spend engaging in Light Activities or being Very Active.

*The BMI graph contains the Earliest BMI's recorded for each recorded user. This serves to give an idea of what sort of a market starts to use wearable fitness technology. 50% of the assessed group are overweight. And only 30% of the users carry a healthy BMI. If we were to assume that the earliest recorded dates for each user indicate their earliest days using the Fitness trackers, This is an indicator that an overweight demographic are more inclined to use fitness technology.

*The mean heartrates for the Users indicate that the assessed user group has healthy heartrates. All of the users have heartrates within the healthy limit (60 - 100).

*The heartrate variance chart shows that 1/14 individuals display a significant variance from an accepted healthy heartrate of 80 BPM.

*The sleep partterns chart indicates that most participants spend minimal awake minutes in bed. And some users display comparatively larger awake minutes in bed. This can be an indicator of participants having trouble falling asleep or participants staying in bed for long before getting off the bed. The analysis found that the average time spent awake in bed was 8.5%. Although this is a small number, When comparing with sleep hours this is significant and features to help minimize this percentage would benefit.

Recommendations based on the Analysis

Future Recommendations

1. **Enhanced Data Collection**: Suggest improvements in data collection for more robust analysis.
2. **User Engagement**: Encourage users to engage with their data regularly.
3. **Features to help achieve less sedentary minutes**: The Bellabeat app could provide reminders to take a walk etc. And help customers achieve better activity.
4. **Features to track calorie intake**: The bellabeat app can accomodate features to track user's calorie intake. The analysis indicated that the majority of the females have a healthy calorie consumption but it also indicated the majority to be overweight. This can be caused by an excess calorie intake and features to track it could be beneficial.
5. **Marketing and features catered to groups hoping to lose weight**: The study indicated that the majority are overweight and indicated that overweight individuals have a special interest in products similar to bellabeat. Therefore, targeted marketing and features beneficial to them would increase sales.
6. **Features to minimize awake minutes in bed**: The analysis indicated that there is still a significant proportion of minutes spent awake in bed among users. This could indicate trouble falling asleep or trouble getting up/ Lack of energy to get up. Bellabeat can implement features in their products to tackle these two scenarios. Such as features to help with stress management or features to track healthy macro and micro nutrient intakes similar to "MyFitnessPal" and thereby providing suggestions to users. This would drive more customers to experience the benefits to real problems they face.

In conclusion, this comprehensive analysis of Fitbit data offers valuable insights into user behavior, ranging from daily activity patterns to heart rate variability and sleep quality. By carefully preprocessing, cleaning, and analyzing the data, we are better equipped to inform Bellabeat's marketing strategy and product development. The insights derived from this study have the *potential to guide future innovations and enhance user engagement with Bellabeat’s health-focused products. Thank you for following along on this data journey, and I look forward to exploring further opportunities for growth and optimization.*

Senura Charith Attanayake