Bellabeat Data Analysis

Salah Eddine EL KHIRANI

2024-08-08

## Introduction

This R Markdown document contains the analysis for the Bellabeat case study. The goal is to analyze smart device usage data to gain insights into how consumers use non-Bellabeat smart devices, and then select one Bellabeat product to apply these insights to in our final presentation.

## Business Task

Our main business tasks are:

1. Identify trends in smart device usage
2. Apply these trends to Bellabeat customers
3. Use these insights to influence Bellabeat’s marketing strategy

## Data Preparation

First, we’ll load the necessary libraries and import our data:

library(tidyr)  
library(dplyr)  
library(lubridate)  
library(janitor)  
  
# Specify the directory containing the CSV files  
data\_dir <- "C:/Users/33775/Documents/Data Analyst/Other/bellabeat data/archive/mturkfitbit\_export\_4.12.16-5.12.16/Fitabase Data 4.12.16-5.12.16/"  
  
# Get a list of all CSV files in the directory  
csv\_files <- list.files(data\_dir, pattern = "\*.csv", full.names = TRUE)  
  
# Loop through each file, read the CSV, and assign it to a variable  
for (csv\_file in csv\_files) {  
 # Extract the file name without the extension  
 file\_name <- tools::file\_path\_sans\_ext(basename(csv\_file))  
   
 # Create a variable name based on the file name  
 var\_name <- make.names(file\_name)  
   
 # Read the CSV file  
 data <- read.csv(csv\_file)  
   
 # Assign the data frame to a variable with the name derived from the file name  
 assign(var\_name, data)  
}  
  
# Verify that the variables have been created  
ls()

## [1] "csv\_file" "csv\_files"   
## [3] "dailyActivity\_merged" "dailyCalories\_merged"   
## [5] "dailyIntensities\_merged" "dailySteps\_merged"   
## [7] "data" "data\_dir"   
## [9] "file\_name" "heartrate\_seconds\_merged"   
## [11] "hourlyCalories\_merged" "hourlyIntensities\_merged"   
## [13] "hourlySteps\_merged" "minuteCaloriesNarrow\_merged"   
## [15] "minuteCaloriesWide\_merged" "minuteIntensitiesNarrow\_merged"  
## [17] "minuteIntensitiesWide\_merged" "minuteMETsNarrow\_merged"   
## [19] "minuteSleep\_merged" "minuteStepsNarrow\_merged"   
## [21] "minuteStepsWide\_merged" "sleepDay\_merged"   
## [23] "var\_name" "weightLogInfo\_merged"

## Limitations of Data

1. Where is your data stored?
   * The data is stored on Kaggle and can be accessed through the following link: [FitBit Fitness Tracker Data](https://www.kaggle.com/datasets/arashnic/fitbit).
2. How is the data organized? Is it in long or wide format?
   * The data is organized in long format, where each row represents a single observation and includes multiple entries for each user across different times or dates.
3. Are there issues with bias or credibility in this data? Does your data ROCCC?
   * The data may have some limitations in terms of sample size and representativeness, as it only includes data from thirty users who consented to share their information. This could introduce selection bias. Additionally, the credibility of the data is dependent on the accuracy of the Fitbit devices and the honesty of the users. To ensure the data ROCCC (Reliable, Original, Comprehensive, Current, Consistent), it’s important to cross-verify with other data sources and understand the context in which the data was collected.
4. How are you addressing licensing, privacy, security, and accessibility?
   * The data is under the CC0: Public Domain license, which means it can be freely used for any purpose. Privacy and security are maintained as the dataset does not contain personally identifiable information (PII). Accessibility is ensured as the data is available publicly on Kaggle.
5. How did you verify the data’s integrity?
   * Data integrity can be verified by checking for consistency in data entries, such as ensuring there are no missing or duplicate records and that the data values are within expected ranges. Cross-referencing with similar datasets or performing sanity checks can also help verify integrity.
6. How does it help you answer your question?
   * The data provides insights into users’ daily habits related to physical activity, heart rate, and sleep patterns. These insights are crucial for understanding trends and behaviors in smart device usage, which can inform Bellabeat’s marketing strategies and product improvements.
7. Are there any problems with the data?
   * Potential problems with the data include:
     + Small sample size, which may not be representative of the broader population.
     + Incomplete data or missing values that could affect the analysis.
     + Possible inaccuracies in self-reported or device-recorded data.
     + Lack of demographic diversity, which may limit the generalizability of the findings. Process

## Data Cleaning and Processing

Number of distinct users in each dataset

n\_distinct(dailyActivity\_merged$Id)

## [1] 33

n\_distinct(hourlyIntensities\_merged$Id)

## [1] 33

n\_distinct(sleepDay\_merged$Id)

## [1] 24

n\_distinct(weightLogInfo\_merged$Id)

## [1] 8

The analysis reveals the number of participants within each dataset. There are 33 participants in the activity and intensities datasets, 24 in the sleep dataset, and only 8 in the weight dataset. The small sample size of 8 participants in the weight data is insufficient to draw any meaningful conclusions or make reliable recommendations based on this data.

Now, let’s clean and process our data:

#remove duplicates and NA  
dailyActivity\_merged <- dailyActivity\_merged %>%  
 distinct() %>%  
 drop\_na()  
  
  
hourlyIntensities\_merged <- hourlyIntensities\_merged %>%  
 distinct() %>%  
 drop\_na()  
  
sleepDay\_merged <- sleepDay\_merged %>%  
 distinct() %>%  
 drop\_na()  
  
weightLogInfo\_merged <- weightLogInfo\_merged %>%  
 distinct() %>%  
 drop\_na(Id, BMI, WeightKg)  
  
# Display summary and view data  
summary(dailyActivity\_merged)

## Id ActivityDate TotalSteps TotalDistance   
## Min. :1.504e+09 Length:940 Min. : 0 Min. : 0.000   
## 1st Qu.:2.320e+09 Class :character 1st Qu.: 3790 1st Qu.: 2.620   
## Median :4.445e+09 Mode :character Median : 7406 Median : 5.245   
## Mean :4.855e+09 Mean : 7638 Mean : 5.490   
## 3rd Qu.:6.962e+09 3rd Qu.:10727 3rd Qu.: 7.713   
## Max. :8.878e+09 Max. :36019 Max. :28.030   
## TrackerDistance LoggedActivitiesDistance VeryActiveDistance  
## Min. : 0.000 Min. :0.0000 Min. : 0.000   
## 1st Qu.: 2.620 1st Qu.:0.0000 1st Qu.: 0.000   
## Median : 5.245 Median :0.0000 Median : 0.210   
## Mean : 5.475 Mean :0.1082 Mean : 1.503   
## 3rd Qu.: 7.710 3rd Qu.:0.0000 3rd Qu.: 2.053   
## Max. :28.030 Max. :4.9421 Max. :21.920   
## ModeratelyActiveDistance LightActiveDistance SedentaryActiveDistance  
## Min. :0.0000 Min. : 0.000 Min. :0.000000   
## 1st Qu.:0.0000 1st Qu.: 1.945 1st Qu.:0.000000   
## Median :0.2400 Median : 3.365 Median :0.000000   
## Mean :0.5675 Mean : 3.341 Mean :0.001606   
## 3rd Qu.:0.8000 3rd Qu.: 4.782 3rd Qu.:0.000000   
## Max. :6.4800 Max. :10.710 Max. :0.110000   
## VeryActiveMinutes FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes  
## Min. : 0.00 Min. : 0.00 Min. : 0.0 Min. : 0.0   
## 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.:127.0 1st Qu.: 729.8   
## Median : 4.00 Median : 6.00 Median :199.0 Median :1057.5   
## Mean : 21.16 Mean : 13.56 Mean :192.8 Mean : 991.2   
## 3rd Qu.: 32.00 3rd Qu.: 19.00 3rd Qu.:264.0 3rd Qu.:1229.5   
## Max. :210.00 Max. :143.00 Max. :518.0 Max. :1440.0   
## Calories   
## Min. : 0   
## 1st Qu.:1828   
## Median :2134   
## Mean :2304   
## 3rd Qu.:2793   
## Max. :4900

View(dailyActivity\_merged)  
  
# Filter and mutate the data in place  
dailyActivity\_merged <- dailyActivity\_merged %>%   
 filter(TotalSteps != 0 & Calories != 0) %>%  
 mutate(ActivityDate = as.Date(ActivityDate, "%m/%d/%Y"))  
  
  
# Separate the Time column into Date and Time columns  
hourlyIntensities\_merged <- hourlyIntensities\_merged %>%  
 separate(ActivityHour, into = c("Date", "Time"), sep = " ")

## Warning: Expected 2 pieces. Additional pieces discarded in 22099 rows [1, 2, 3, 4, 5, 6,  
## 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].

sleepDay\_merged <- sleepDay\_merged %>%  
 separate(SleepDay, into = c("Date", "Time"), sep = " ")

## Warning: Expected 2 pieces. Additional pieces discarded in 410 rows [1, 2, 3, 4, 5, 6,  
## 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].

# Convert Date and Time columns to appropriate formats using lubridate  
hourlyIntensities\_merged <- hourlyIntensities\_merged %>%  
 mutate(Date = mdy(Date),  
 Time = hms(Time))  
  
sleepDay\_merged <- sleepDay\_merged %>%  
 mutate(Date = mdy(Date),  
 Time = hms(Time))  
  
#Clean and rename columns  
  
dailyActivity\_merged = clean\_names(dailyActivity\_merged)  
hourlyCalories\_merged = clean\_names(hourlyCalories\_merged)  
hourlyIntensities\_merged = clean\_names(hourlyIntensities\_merged)  
sleepDay\_merged = clean\_names(sleepDay\_merged)  
weightLogInfo\_merged = clean\_names(weightLogInfo\_merged)  
dailyActivity\_merged <-dailyActivity\_merged %>%  
 rename(date = activity\_date)

Merge data To prepare for data visualization, the first step is to merge the activity and sleep datasets. This process involves performing an inner join on the Id and Date columns, which I created after converting the original data to the correct date-time format. Merging these datasets ensures that activity and sleep information are aligned by participant and date, providing a cohesive foundation for subsequent analysis.

# Merge daily activity and sleep data  
daily\_activity\_sleep\_merged <- merge(dailyActivity\_merged, sleepDay\_merged, by=c ("id", "date"))  
glimpse(daily\_activity\_sleep\_merged)

## Rows: 410  
## Columns: 19  
## $ id <dbl> 1503960366, 1503960366, 1503960366, 1503960…  
## $ date <date> 2016-04-12, 2016-04-13, 2016-04-15, 2016-0…  
## $ total\_steps <int> 13162, 10735, 9762, 12669, 9705, 15506, 105…  
## $ total\_distance <dbl> 8.50, 6.97, 6.28, 8.16, 6.48, 9.88, 6.68, 6…  
## $ tracker\_distance <dbl> 8.50, 6.97, 6.28, 8.16, 6.48, 9.88, 6.68, 6…  
## $ logged\_activities\_distance <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0…  
## $ very\_active\_distance <dbl> 1.88, 1.57, 2.14, 2.71, 3.19, 3.53, 1.96, 1…  
## $ moderately\_active\_distance <dbl> 0.55, 0.69, 1.26, 0.41, 0.78, 1.32, 0.48, 0…  
## $ light\_active\_distance <dbl> 6.06, 4.71, 2.83, 5.04, 2.51, 5.03, 4.24, 4…  
## $ sedentary\_active\_distance <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0…  
## $ very\_active\_minutes <int> 25, 21, 29, 36, 38, 50, 28, 19, 41, 39, 73,…  
## $ fairly\_active\_minutes <int> 13, 19, 34, 10, 20, 31, 12, 8, 21, 5, 14, 2…  
## $ lightly\_active\_minutes <int> 328, 217, 209, 221, 164, 264, 205, 211, 262…  
## $ sedentary\_minutes <int> 728, 776, 726, 773, 539, 775, 818, 838, 732…  
## $ calories <int> 1985, 1797, 1745, 1863, 1728, 2035, 1786, 1…  
## $ time <Period> 12H 0M 0S, 12H 0M 0S, 12H 0M 0S, 12H 0M …  
## $ total\_sleep\_records <int> 1, 2, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1…  
## $ total\_minutes\_asleep <int> 327, 384, 412, 340, 700, 304, 360, 325, 361…  
## $ total\_time\_in\_bed <int> 346, 407, 442, 367, 712, 320, 377, 364, 384…

Let’s have a look at summary statistics of the data sets

# Summary statistics for daily activity  
summary(dailyActivity\_merged)

## id date total\_steps total\_distance   
## Min. :1.504e+09 Min. :2016-04-12 Min. : 4 Min. : 0.00   
## 1st Qu.:2.320e+09 1st Qu.:2016-04-18 1st Qu.: 4923 1st Qu.: 3.37   
## Median :4.445e+09 Median :2016-04-26 Median : 8053 Median : 5.59   
## Mean :4.858e+09 Mean :2016-04-26 Mean : 8319 Mean : 5.98   
## 3rd Qu.:6.962e+09 3rd Qu.:2016-05-03 3rd Qu.:11092 3rd Qu.: 7.90   
## Max. :8.878e+09 Max. :2016-05-12 Max. :36019 Max. :28.03   
## tracker\_distance logged\_activities\_distance very\_active\_distance  
## Min. : 0.000 Min. :0.0000 Min. : 0.000   
## 1st Qu.: 3.370 1st Qu.:0.0000 1st Qu.: 0.000   
## Median : 5.590 Median :0.0000 Median : 0.410   
## Mean : 5.964 Mean :0.1178 Mean : 1.637   
## 3rd Qu.: 7.880 3rd Qu.:0.0000 3rd Qu.: 2.275   
## Max. :28.030 Max. :4.9421 Max. :21.920   
## moderately\_active\_distance light\_active\_distance sedentary\_active\_distance  
## Min. :0.0000 Min. : 0.000 Min. :0.00000   
## 1st Qu.:0.0000 1st Qu.: 2.345 1st Qu.:0.00000   
## Median :0.3100 Median : 3.580 Median :0.00000   
## Mean :0.6182 Mean : 3.639 Mean :0.00175   
## 3rd Qu.:0.8650 3rd Qu.: 4.895 3rd Qu.:0.00000   
## Max. :6.4800 Max. :10.710 Max. :0.11000   
## very\_active\_minutes fairly\_active\_minutes lightly\_active\_minutes  
## Min. : 0.00 Min. : 0.00 Min. : 0.0   
## 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.:146.5   
## Median : 7.00 Median : 8.00 Median :208.0   
## Mean : 23.02 Mean : 14.78 Mean :210.0   
## 3rd Qu.: 35.00 3rd Qu.: 21.00 3rd Qu.:272.0   
## Max. :210.00 Max. :143.00 Max. :518.0   
## sedentary\_minutes calories   
## Min. : 0.0 Min. : 52   
## 1st Qu.: 721.5 1st Qu.:1856   
## Median :1021.0 Median :2220   
## Mean : 955.8 Mean :2361   
## 3rd Qu.:1189.0 3rd Qu.:2832   
## Max. :1440.0 Max. :4900

Based on the summary statistics of the dailyActivity\_merged dataset, several key insights can be drawn:

1. Activity Levels Vary Widely:
   * The Total Steps recorded by participants range from as few as 4 steps to a maximum of 36,019 steps per day. The mean number of steps is 8,319, indicating that while some participants are very active, a significant portion may have relatively low daily activity.
   * Very Active Minutes also show considerable variation, with a range from 0 to 210 minutes per day. The mean is 23 minutes, suggesting that while some participants engage in high-intensity activities, many may not be frequently involved in such exercises.
2. Distance Covered:
   * Total Distance and Tracker Distance metrics show a wide range, with some participants covering nearly 28 miles in a day. The average distance covered per day is about 6 miles. This aligns with the step count, reinforcing the notion that participant activity levels vary significantly.
   * Lightly Active Distance is the most substantial contributor to overall distance, with a mean of 3.64 miles, suggesting that most participants engage in light activities regularly.
3. Sedentary Behavior:
   * The Sedentary Minutes mean is 955.8 minutes (nearly 16 hours) per day, indicating that participants spend a large portion of their day in sedentary activities. This insight could be crucial for health recommendations, highlighting the need for strategies to reduce sedentary time.
4. Calories Burned:
   * The number of Calories burned varies from 52 to 4,900 calories per day, with an average of 2,361 calories. This wide range reflects different levels of physical activity and possibly varying body compositions and metabolic rates among participants.
5. Moderate and Light Activities:
   * The Moderately Active Distance and Fairly Active Minutes have lower means, indicating that moderate activity is less common among participants compared to light activity. Most participants seem to engage more in light physical activities rather than moderate or intense activities. – Recommendations:

* Encourage Increased Physical Activity: Participants with low step counts and very active minutes might benefit from personalized recommendations to increase their daily physical activity.
* Target Sedentary Time: Given the high average sedentary time, strategies to break up prolonged periods of inactivity could be beneficial.
* Personalized Interventions: With significant variability in activity levels, interventions could be tailored based on individual behavior patterns to effectively promote healthier lifestyles.

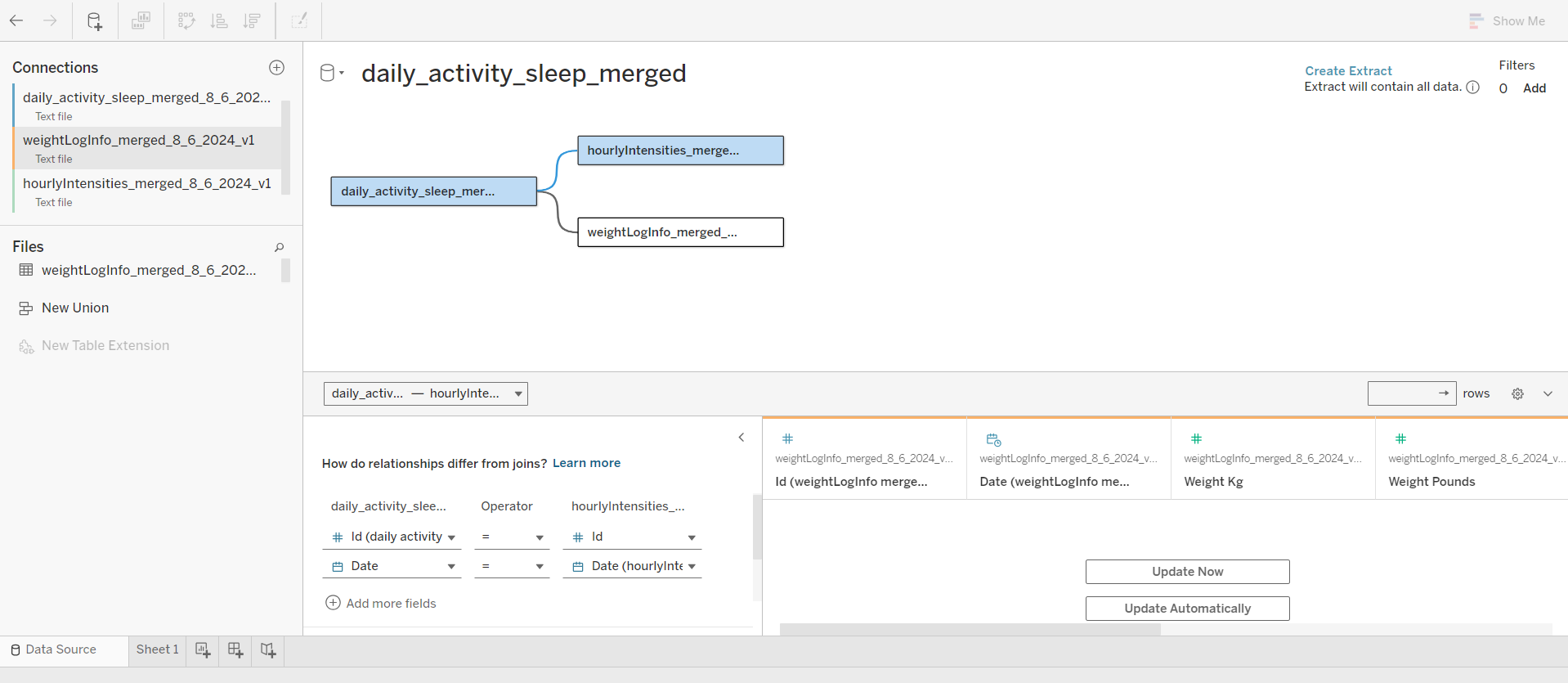
## Data Visualization

The analysis phase will be conducted in Tableau. To facilitate this, I’ll export the required data into CSV files and save them in the designated output directory. This ensures that the data is readily available for visualization and further analysis in Tableau.

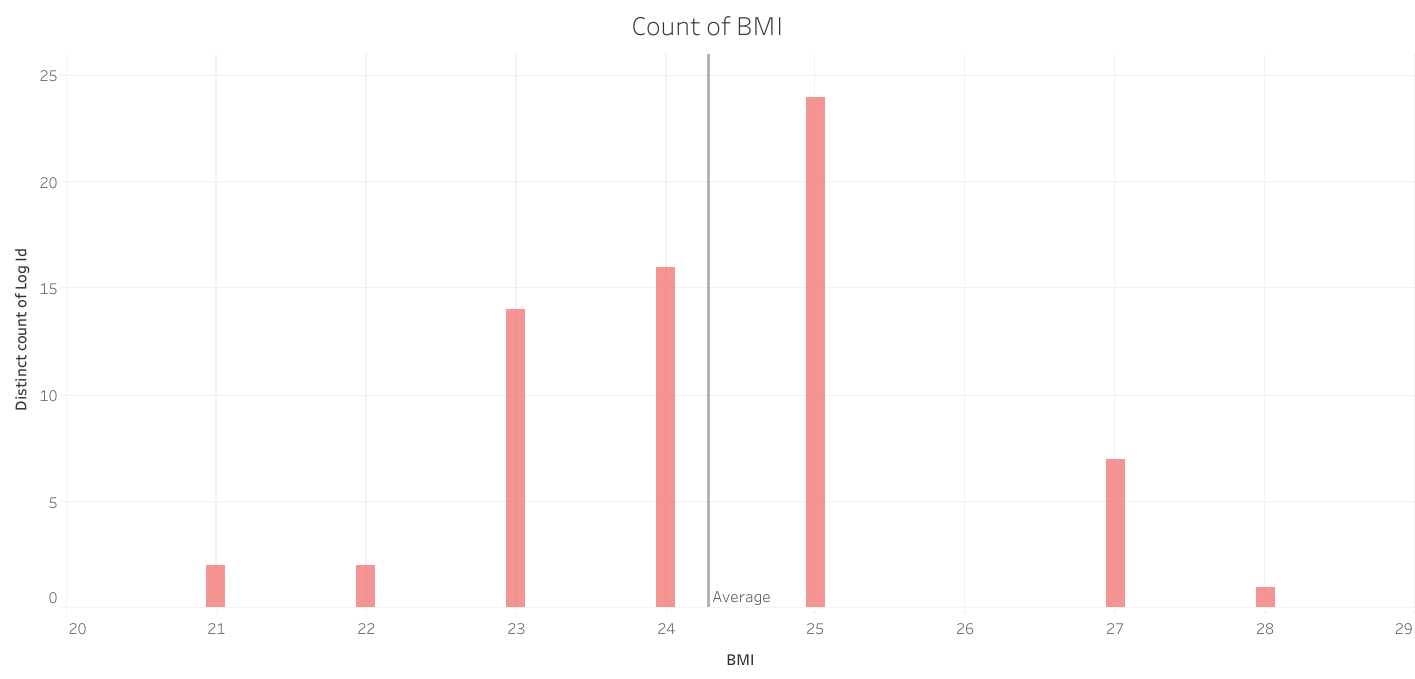
# Specify the Output directory   
Output\_dir <- "C:/Users/33775/Documents/Data Analyst/Other/OutPut Bellabeat/"  
  
# Define the output file path  
output\_file\_path <- file.path(Output\_dir, paste0("daily\_activity\_sleep\_merged\_8\_6\_2024\_v1", ".csv"))  
  
# Write the data frame to a CSV file  
write.csv(daily\_activity\_sleep\_merged, file = output\_file\_path, row.names = FALSE)  
  
# Define the output file path  
output\_file\_path <- file.path(Output\_dir, paste0("weightLogInfo\_merged\_8\_6\_2024\_v1", ".csv"))  
  
# Write the data frame to a CSV file  
write.csv(weightLogInfo\_merged, file = output\_file\_path, row.names = FALSE)  
  
# Define the output file path  
output\_file\_path <- file.path(Output\_dir, paste0("hourlyIntensities\_merged\_8\_6\_2024\_v1", ".csv"))  
  
# Write the data frame to a CSV file  
write.csv(hourlyIntensities\_merged, file = output\_file\_path, row.names = FALSE)

The initial step I completed in Tableau was establishing connections between the three tables, as illustrated in the visualization below.

knitr::include\_graphics("C:/Users/33775/Documents/Data Analyst/Other/OutPut Bellabeat/tableau connexion.png")

 Now, let’s create some visualizations to better understand our data:

knitr::include\_graphics("C:/Users/33775/Downloads/BMI.png")



The bar chart visualizes the distribution of BMI (Body Mass Index) values among participants, with the count of distinct Log IDs represented on the vertical axis.

##Insights:

Concentration Around Normal BMI: The majority of participants have BMI values clustered around 23 to 25, indicating that most of the population falls within a healthy weight range.

Fewer Outliers: There are fewer participants with BMI values below 23 and above 27. This suggests that there are relatively few underweight or overweight participants in the dataset.

Average BMI Marker: The chart includes an average BMI marker, showing that the average BMI of the participants is around 25. This is on the higher end of the healthy range, bordering on overweight.

Small Sample Size: The distribution is not very broad, which might indicate that the dataset has a limited number of participants. This could affect the generalizability of the insights drawn from this data.

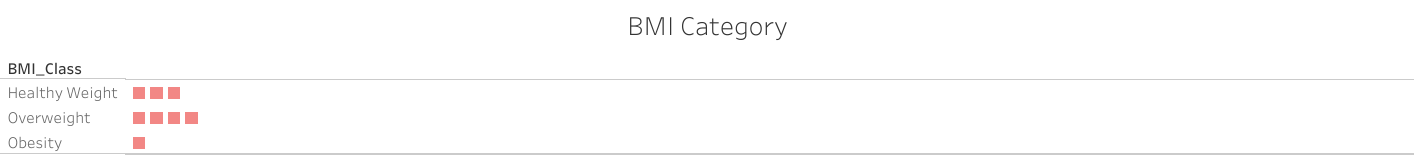
Overall, the data suggests that most participants are within the normal weight range, with a small portion potentially at risk of being overweight.

Then, i created a calculated measure to categorize BMI into four distinct classes, following the guidelines provided by the Australian Government’s Department of Health (source: health.gov.au):

* Underweight: BMI less than 18.5
* Healthy Weight: BMI between 18.5 and 24.9
* Overweight but Not Obese: BMI between 25 and 29.9
* Obese Class I: BMI between 30 and 34.9
* This categorization allows for a more detailed analysis of participants’ BMI distributions according to standardized health guidelines.

Mesure Formula : IF [BMI1]<18.5 THEN “Underweight” ELSEIF [BMI1]>=18.5 AND [BMI1]<25 THEN “Healthy Weight” ELSEIF [BMI1]>=25 AND [BMI1]<30 THEN “Overweight” ELSEIF [BMI1]>=30 THEN “Obesity” END

knitr::include\_graphics("C:/Users/33775/Downloads/BMI CLASS.png")



Based on our sample, it’s evident that there are no individuals in the underweight category. As observed from the distribution above, the data reveals that 3 participants fall within the healthy weight range, 4 are categorized as overweight, and 1 is classified as obese.

knitr::include\_graphics("C:/Users/33775/Downloads/Distribution Activities.png")

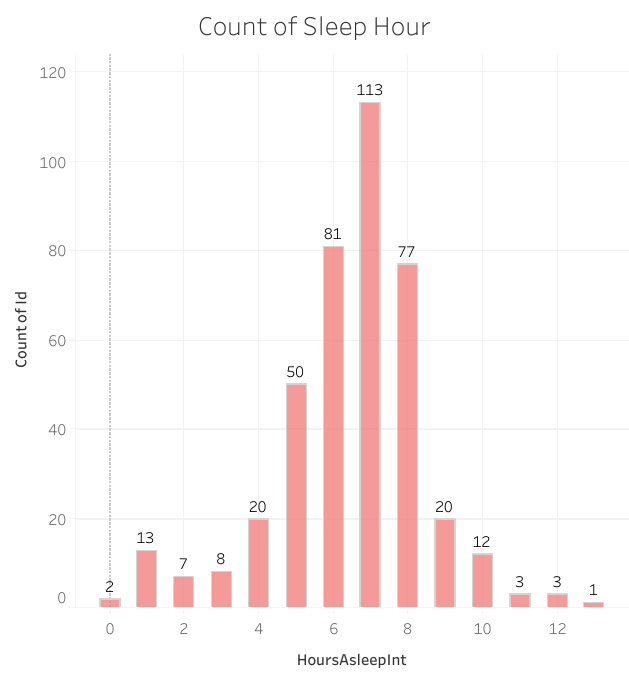


Based on the “Distribution Activities” chart shown in the image, a few key insights can be drawn:

The majority of the distribution is concentrated in the “Sedentary Minutes” and “Total Time In Bed” categories, indicating that most people spend a significant portion of their time being sedentary or in bed. The “Fairly Active Minutes” and “Lightly Active Minutes” categories show more moderate levels of activity, suggesting that a substantial portion of the population engages in some level of physical activity. The “Very Active Minutes” category has the lowest distribution, implying that a relatively small percentage of the population spends a significant amount of time in vigorous physical activity.

In summary, the chart suggests that the general population tends to be more sedentary or spend a lot of time in bed, while moderate and high levels of physical activity are less common.

knitr::include\_graphics("C:/Users/33775/Downloads/Count of Sleep Hour.png")



Based on the “Count of Sleep Hour” chart, the key insights are:

1- The distribution of sleep hours shows a clear peak at 8 hours, with 113 counts recorded for this sleep duration. 2- There is a secondary peak at 7 hours of sleep, with 81 counts. 3- The chart exhibits a normal distribution pattern, with the highest frequency around the 8-hour sleep duration and tapering off on both the shorter and longer sleep hour ends. 4- There is a range of sleep durations from 0 hours up to 12 hours, indicating a wide variation in individual sleep patterns. 5- The counts for sleep durations below 6 hours and above 10 hours are relatively low, suggesting these are less common sleep patterns compared to the peaks around 7-8 hours. 6- The data points to a clear preference or tendency for most individuals to sleep around 7-8 hours per night, which is generally considered the recommended sleep duration for adults.

In summary, the chart shows a normal distribution pattern in the count of sleep hours, with the majority of individuals sleeping around 7-8 hours per night, and fewer people exhibiting shorter or longer sleep durations.

knitr::include\_graphics("C:/Users/33775/Downloads/Calories & Hours asleep Vs Total Steps.png")



The image shows a scatter plot of “Calories & Hours asleep” versus “Total Steps” for a dataset.

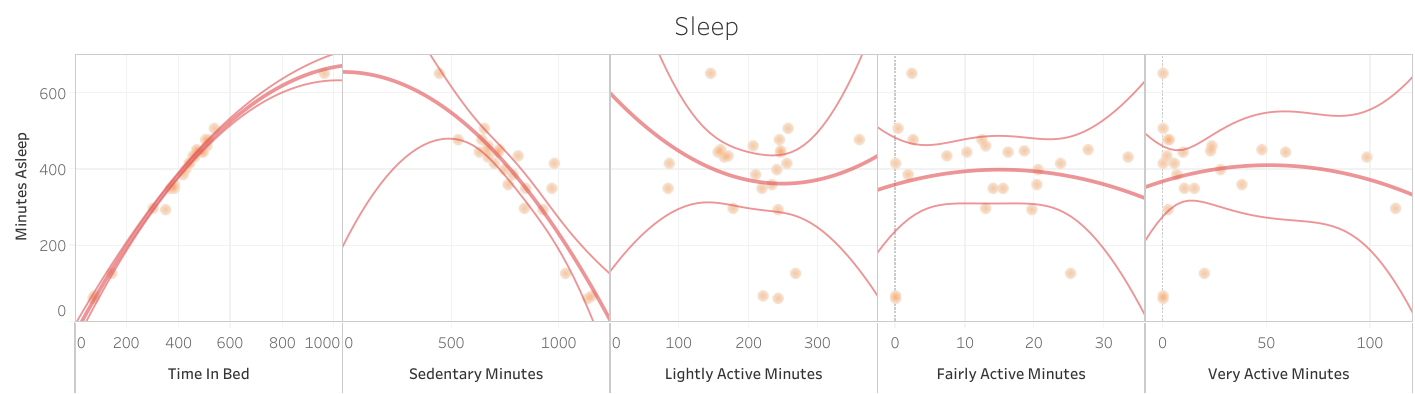
1- Positive correlation between hours of sleep and total steps: The scatter plot actually shows a positive correlation between hours of sleep and total steps. As the hours of sleep increase, the total number of steps taken generally increases as

2- Relationship between calories and total steps: The scatter plot also shows a positive correlation between calories burned and total steps. As the total number of steps increases, the number of calories burned also tends to increase.

3- Variation in the data: There is a significant amount of variation in the data points, indicating that the relationship between sleep, calories, and steps is not perfectly linear. There are many outliers and data points that do not strictly follow the overall trend.

4- Potential insights for Bellabeat: This data could provide insights for Bellabeat on how their customers use fitness trackers and how sleep, activity, and calorie burn are related. Bellabeat could use these insights to inform their product development, marketing strategies, and recommendations to customers.

knitr::include\_graphics("C:/Users/33775/Downloads/Sleep.png")



Positive Correlation (Time in Bed):

1- Vertical Positive Trend Line: The first chart (Time in Bed vs. Minutes Asleep) shows a strong positive correlation, evidenced by the upward sloping trend line. This indicates that the more time users spend in bed, the longer they sleep. This is a clear and expected relationship, confirming that time in bed is a significant factor in determining sleep duration. Negative Correlation (Sedentary Minutes):

2- Vertical Negative Trend Line: The second chart (Sedentary Minutes vs. Minutes Asleep) shows a negative correlation. The trend line slopes downward, suggesting that higher sedentary minutes may be associated with less sleep, though the relationship is not as strong as with time in bed. This could indicate that spending excessive time being inactive might negatively impact sleep quality. No Significant Correlation (Lightly Active, Fairly Active, and Very Active Minutes):

3- Horizontal Trend Lines: The last three charts (Lightly Active, Fairly Active, and Very Active Minutes vs. Minutes Asleep) show horizontal trend lines, indicating little to no correlation between these activity levels and sleep duration. This suggests that moderate to high levels of activity during the day do not have a direct or significant impact on the amount of sleep.

– Recommendations for Bellabeat:

Encourage Healthy Bedtime Routines: Since time in bed has a strong positive correlation with sleep duration, Bellabeat could emphasize the importance of a consistent bedtime routine in its app. Features such as bedtime reminders and sleep goals could help users improve their sleep duration.

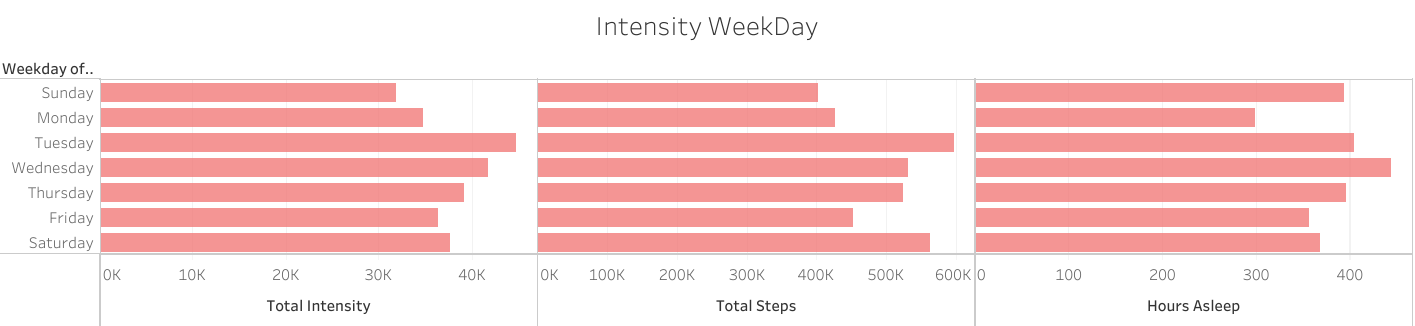
Monitor and Manage Sedentary Behavior: The negative correlation between sedentary minutes and sleep duration suggests that reducing sedentary behavior might improve sleep. Bellabeat could introduce features that encourage users to take breaks from inactivity, such as gentle reminders to move after prolonged periods of sitting.

Balance Activity Levels: Although lightly, fairly, and very active minutes show little correlation with sleep duration, it’s still important to maintain a balanced level of daily activity for overall health. Bellabeat could continue to promote balanced activity levels while focusing on other factors, like stress management and mindfulness, that might more directly affect sleep.

Personalized Sleep Insights: Bellabeat could provide users with personalized insights and suggestions based on their specific activity and sleep patterns. For instance, if a user has high sedentary minutes but low sleep duration, the app could recommend ways to integrate more active breaks during the day.

Comprehensive Health Monitoring: Since sleep is influenced by various factors beyond just physical activity, Bellabeat might consider offering more comprehensive health monitoring, including stress levels, diet, and mental well-being, to provide users with a holistic approach to improving their sleep and overall health.

knitr::include\_graphics("C:/Users/33775/Downloads/Intensity WeekDay.png")



1- Total Intensity:

Highest on Wednesdays: The “Total Intensity” is at its peak on Wednesday, indicating that users tend to be most active in terms of intensity in the middle of the week. This could be due to mid-week exercise routines or other physically demanding activities typically scheduled for this day. Lowest on Fridays and Sundays: Fridays and Sundays show lower levels of total intensity, suggesting that users may be winding down towards the weekend or taking rest days.

2- Total Steps:

Consistent Step Counts: The “Total Steps” metric remains relatively consistent across most weekdays, with slight peaks on Tuesday and Wednesday. This indicates that users maintain a fairly steady level of daily walking or movement throughout the week, with only slight variations. Notable Drop on Friday: There is a noticeable decrease in total steps on Friday, possibly due to the beginning of the weekend, when users may be less active.

3- Hours Asleep:

Higher Sleep Duration Midweek: Sleep duration is higher on Wednesday and Thursday, suggesting that users may be catching up on rest midweek, perhaps after the initial rush of the start of the week. Lower Sleep Duration on Friday: Friday shows a slight drop in sleep duration, which could be due to social activities or changes in routine as the weekend begins.

* Recommendations for Bellabeat:

Promote Midweek Activity: Bellabeat could encourage users to maintain or increase their physical activity intensity on less active days like Friday and Sunday by offering challenges or personalized workout suggestions.

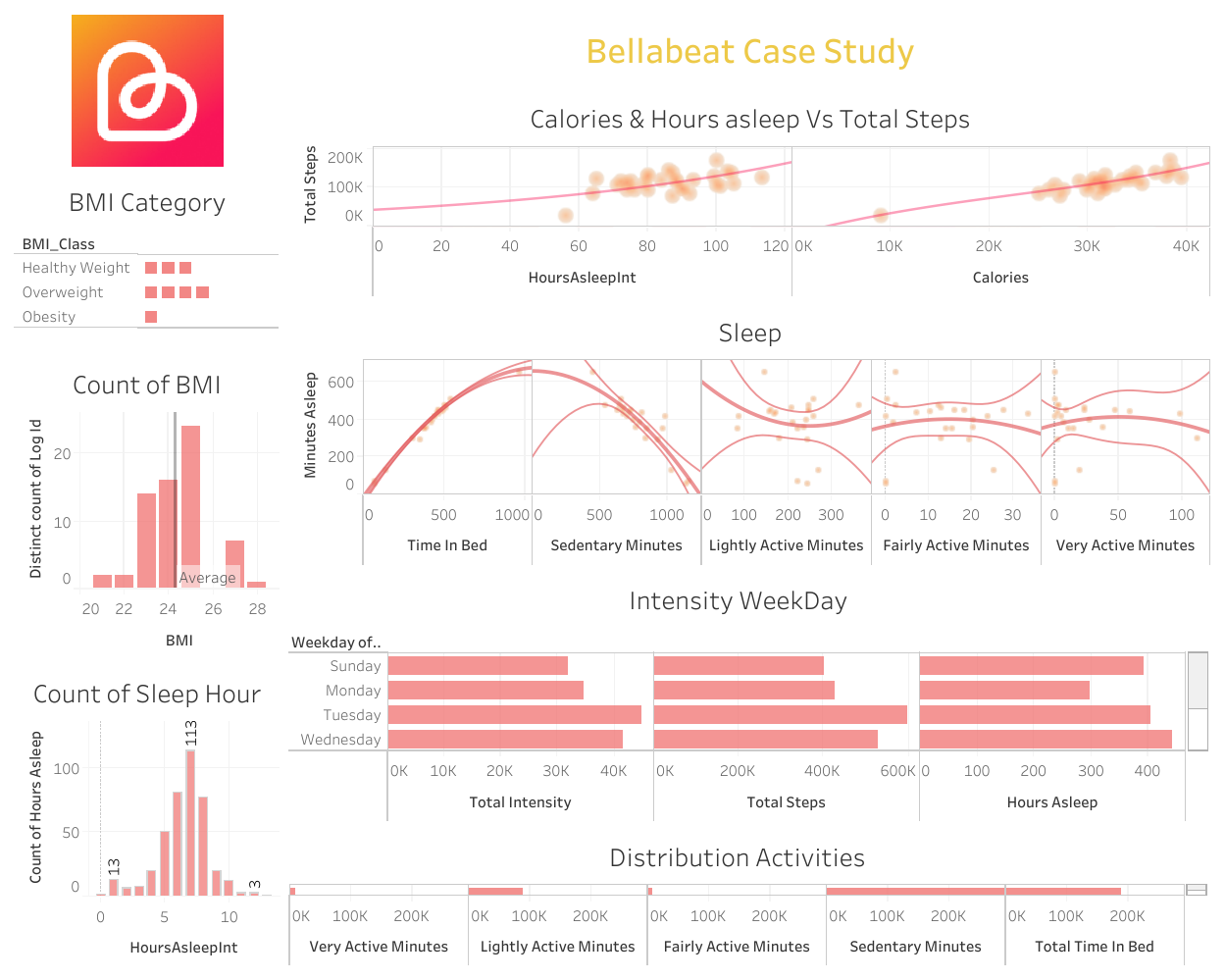
Weekend Sleep Optimization: Since there is a drop in sleep duration on Fridays, Bellabeat might recommend strategies for maintaining consistent sleep patterns throughout the week, including during the weekend. This could involve reminders to maintain a consistent bedtime or tips for winding down after a busy week.

Consistent Activity: Given the consistency in step counts across the week, Bellabeat could emphasize the importance of regular daily movement and help users identify opportunities to increase steps on less active days, particularly Friday.

Rest Day Encouragement: The lower intensity on Sundays could be an indication of rest days. Bellabeat might offer content on the importance of rest and recovery, helping users to understand how to balance intense activities with adequate rest for optimal health.

## Here is the final outcome of our work: a dashboard meticulously created in Tableau.

knitr::include\_graphics("C:/Users/33775/Downloads/Bellabeat Dashboard.png")



This comprehensive dashboard for the Bellabeat Case Study Click here to view the Tableau Dashboard provides a wealth of insights into user activity patterns, sleep habits, and health metrics. The visualizations effectively showcase the relationships between various aspects of users’ daily routines, including steps taken, calories burned, sleep duration, and activity intensity across different days of the week. The BMI distribution chart offers a quick overview of the user base’s health status, while the sleep analysis graphs reveal intricate patterns in sleep quality and duration relative to different activity levels. The weekday intensity chart highlights how user behaviors vary throughout the week, potentially informing targeted marketing strategies. This dashboard serves as a powerful tool for Bellabeat to understand their customers’ habits and tailor their products and services accordingly, ultimately supporting their goal of becoming a larger player in the global smart device market for women’s health.

Based on the analysis provided in the Bellabeat case study, here are the key findings and recommendations: ## Key Findings 1. Activity Patterns: - There’s wide variation in activity levels among users, with daily steps ranging from 4 to 36,019 (mean: 8,319). - Most users engage in light activity, with less time spent on moderate or intense activities. - Sedentary time is high, averaging nearly 16 hours per day.

1. Sleep Patterns:
   * Most users sleep between 7-8 hours per night, with a peak at 8 hours.
   * There’s a positive correlation between time in bed and sleep duration.
   * A negative correlation exists between sedentary minutes and sleep duration.
2. BMI Distribution:
   * The majority of users fall within the healthy to overweight BMI range.
   * No users were classified as underweight in the sample.
3. Weekly Activity Trends:
   * Wednesdays show peak activity intensity.
   * Fridays and Sundays have the lowest activity intensity.
   * Step counts remain relatively consistent throughout the week, with a slight drop on Fridays.
4. Calorie Burn and Steps:
   * There’s a positive correlation between total steps and calories burned.
5. Sleep and Activity Relationship:
   * Light, fair, and very active minutes show little correlation with sleep duration.

## Recommendations

1. Personalized Activity Goals:
   * Implement personalized step and activity goals based on individual user patterns to encourage increased physical activity, especially for those with low step counts.
2. Sedentary Time Alerts:
   * Introduce features that remind users to move after prolonged periods of inactivity, helping to reduce the high average sedentary time.
3. Sleep Optimization:
   * Develop a sleep tracking feature that provides insights on optimal bedtime routines and sleep duration based on individual patterns.
   * Offer tips for maintaining consistent sleep patterns, especially on weekends.
4. Midweek Motivation:
   * Capitalize on the midweek activity peak by introducing special challenges or rewards for maintaining high activity levels on Wednesdays.
5. Weekend Activity Boost:
   * Create weekend-specific challenges or suggestions to combat the lower activity levels observed on Fridays and Sundays.
6. Holistic Health Approach:
   * Integrate features that connect activity levels, sleep patterns, and BMI to provide users with a comprehensive view of their health.
7. Smart Notifications:
   * Implement an intelligent notification system that encourages activity based on real-time data and weekly trends.
8. Educational Content:
   * Provide in-app educational content about the importance of balancing activity levels, reducing sedentary time, and maintaining consistent sleep patterns.
9. Social Features:
   * Introduce community challenges or friend competitions to boost motivation, particularly on less active days.
10. Stress Management:
    * Consider adding stress tracking and management features, as stress can impact both activity levels and sleep quality.

By implementing these recommendations, Bellabeat can create a more engaging, personalized experience for users, potentially leading to improved health outcomes and increased user retention.

## Next Steps

To further improve this analysis, we could:

1. Collect more data over a longer period to identify long-term trends.
2. Gather demographic information to segment users and tailor marketing strategies.
3. Conduct a survey to understand user motivations and preferences.