



# Bellabeat Case Study

Data Driven Insights for Wellness & Marketing Strategy

## Krishnanunni H Pillai

Tech Enthusiast | Data Analyst | B.Tech Student

[View Project](#)

[Download PDF](#)

[Contact Me](#)

### Business Task

Analyze smart device usage data to identify user behavior patterns that can help Bellabeat optimize its marketing strategy and improve user engagement. The goal is to uncover trends in physical activity, sleep, and overall wellness behaviors using wearable device data. These findings will guide the development of personalized wellness content, targeted marketing campaigns, and product improvements to increase user satisfaction and market competitiveness. Analyze smart device usage data to identify user behavior patterns that can help Bellabeat optimize its marketing strategy and improve user engagement.

### Company Background

- Bellabeat is a wellness-focused technology company founded in 2013 by Urska Srsen and Sando Mur.
- The company designs smart devices and apps tailored to women's health and lifestyle tracking.
- Bellabeat offers a subscription-based program that provides personalized wellness guidance based on user data.
- The company operates globally and relies on digital marketing, including social media and targeted ads, to reach and engage its audience.

### Dataset Overview

## Fitbit Fitness Tracker Data (Public Domain - Kaggle)

- Data collected from 30 Fitbit users who volunteered to share their minute-level fitness data.
- Includes multiple CSV files covering daily activity, steps, calories burned, sleep duration, weight logs, BMI, and body fat percentage.
- Covers a wide range of health behaviors: physical activity, rest, hydration, and body composition.
- Selected for its relevance to Bellabeat user base and the wellness insights it can generate.
- Limitations: Small sample size, not Bellabeat-specific, mostly anonymized data.
- Tracks: Steps, Calories, Sleep, Weight, Fat %, BMI
- Public domain dataset from Kaggle

## Data Cleaning

Using **R** with the **tidyverse**, **lubridate**, and **janitor** packages,



I imported and reviewed multiple CSV files to check the structure and quality of the data.

I standardized column names with **janitor::clean\_names()** to keep uniformity across the datasets.

I also converted the date and time fields using lubridate functions for consistent formatting.

I removed duplicate records to keep the data unique, and I handled missing values and **NA** entries with suitable cleaning methods. Next, I filtered and joined key datasets, including daily activity, sleep, and weight logs, to support a combined analysis.

To improve data usability, I extracted the latest **BMI** value for each user and rounded timestamps to make the records fit together better. Finally, I checked the data structure and fixed any inconsistencies across datasets to establish a clean and reliable basis for analysis.

## Key Insights

### 1. Active Hours

#### Findings:

Analysis of the daily activity data revealed that most users tend to engage in physical activity during the late afternoon and early evening hours, particularly between 4 PM and 7 PM. This time window consistently shows the highest step counts and active minutes.

#### Analysis process:

importing the needed dataset for analysis

```
steps <- read.csv("minuteStepsNarrow_merged.csv")
View(steps)
```

#### Cleaning & Organizing the data

```
steps$ActivityMinute <- as.POSIXct(steps$ActivityMinute, format='
View(steps)

#Extract Hour from Time column
steps$Hour <- hour(steps$ActivityMinute)
View(steps)

#Finding total steps by hour
# Total steps by hour
hourly_steps <- steps %>%
  group_by(Hour) %>%
  summarise(AverageSteps = mean(Steps))
View(steps)

#Plotting the values in bar graph
ggplot(hourly_steps, aes(x=Hour,y=AverageSteps))+
  geom_line()+
  geom_point()+
  labs(title = "Average Steps by Hour of Day", x="Hour", y="Averag

#Time Period wise Analysis
#Morning-AfternoonEvening-Night, kindaaa

steps <- steps %>%
```

```
mutate(TimeOfDay = case_when(  
  Hour >= 5 & Hour < 12 ~ "Morning",  
  Hour >= 12 & Hour < 16 ~ "Afternoon",  
  Hour >= 16 & Hour < 19 ~ "Evening",  
  TRUE ~ "Night"  
))
```

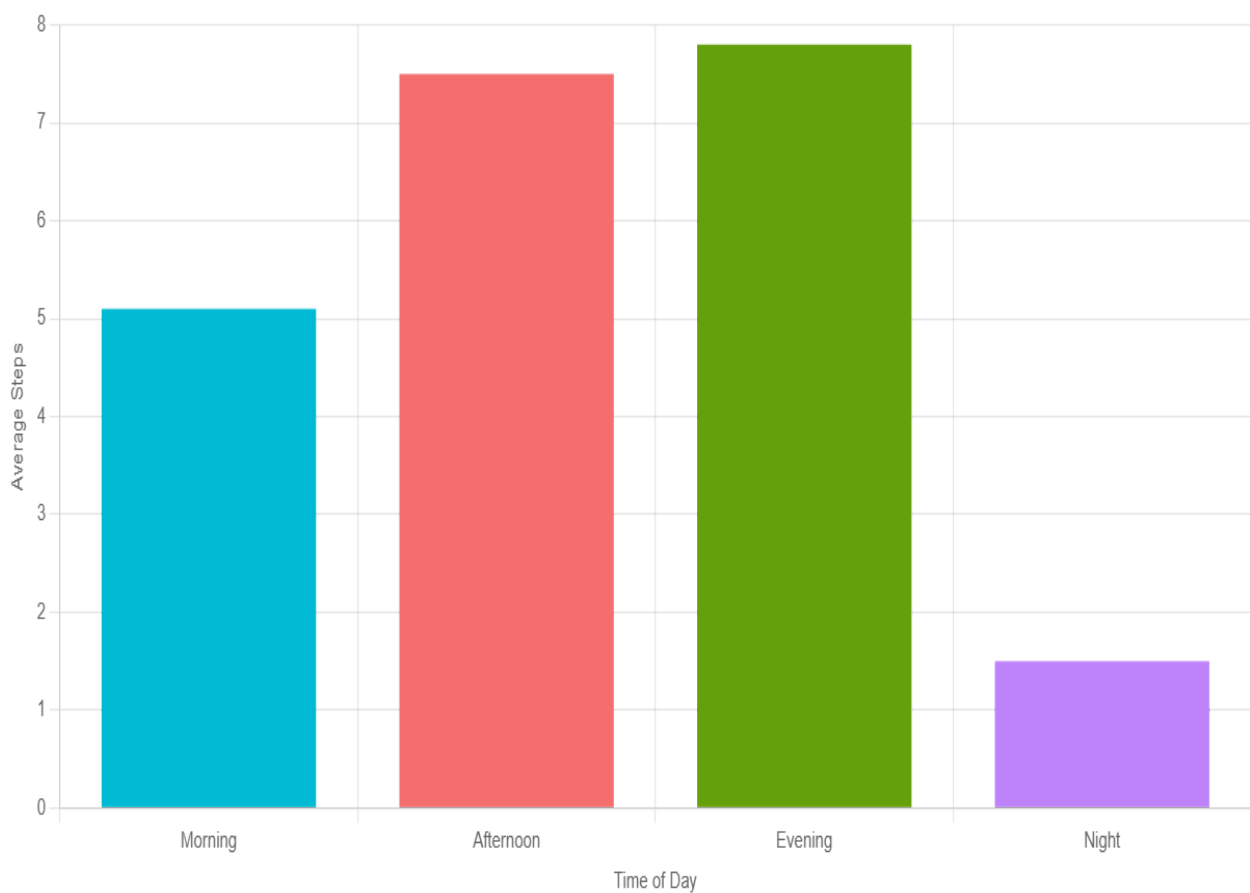
```
#group by time period  
period_steps <- steps %>%  
  group_by(TimeOfDay) %>%  
  summarise(AverageSteps=mean(Steps))
```

## Creating the graph for better understanding

```
#Visualizing  
ggplot(period_steps, aes(x=TimeOfDay, y=AverageSteps, fill=TimeOfDay)) +  
  geom_bar(stat = "identity") +  
  labs(title="Average Steps by Time of Day", x="Time of Day", y="Average Steps")
```

### Result :

This bar chart shows that users are most active in the afternoon and evening, with fewer steps recorded during morning and night hours.



**Action:**

Bellabeat should schedule motivational push notifications, reminders, or app based challenges during this time to match users' natural behavior patterns. This could improve engagement and support healthy habits when users are most open to them.

## 2. Average Daily Steps

**Findings:**

Users take an average of 6,565 steps per day. This is below the commonly recommended target of 10,000 steps per day for good health. The distribution of step counts also shows that many users regularly fall short of this goal. This may indicate a need for prompts or goal-setting features.

**Analysis process :**

### Loading the DataSet

```
hourly <- read.csv("/hourlySteps_merged.csv")
View(hourly)
```

### Cleaning & Organizing the data

```
#Activity hour to date-time
hourly$ActivityHour <- as.POSIXct(hourly$ActivityHour,format="%m/%c

#Extract date(without time)
hourly$Date <- as.Date(hourly$ActivityHour)
View(hourly)

#Aggregate to daily steps per user
daily_steps <- hourly %>%
  group_by(Id,Date) %>%
  summarise(DailyStepTotal = sum(StepTotal), .groups = "drop")
View(daily_steps)

#Calculating average steps users take on a day

average_daily_steps <- mean(daily_steps$DailyStepTotal)
print(average_daily_steps)

#Average steps per user
user_avg_steps <- daily_steps %>%
  group_by(Id) %>%
  summarise(AverageDailySteps = mean(DailyStepTotal))
View(user_avg_steps)
```

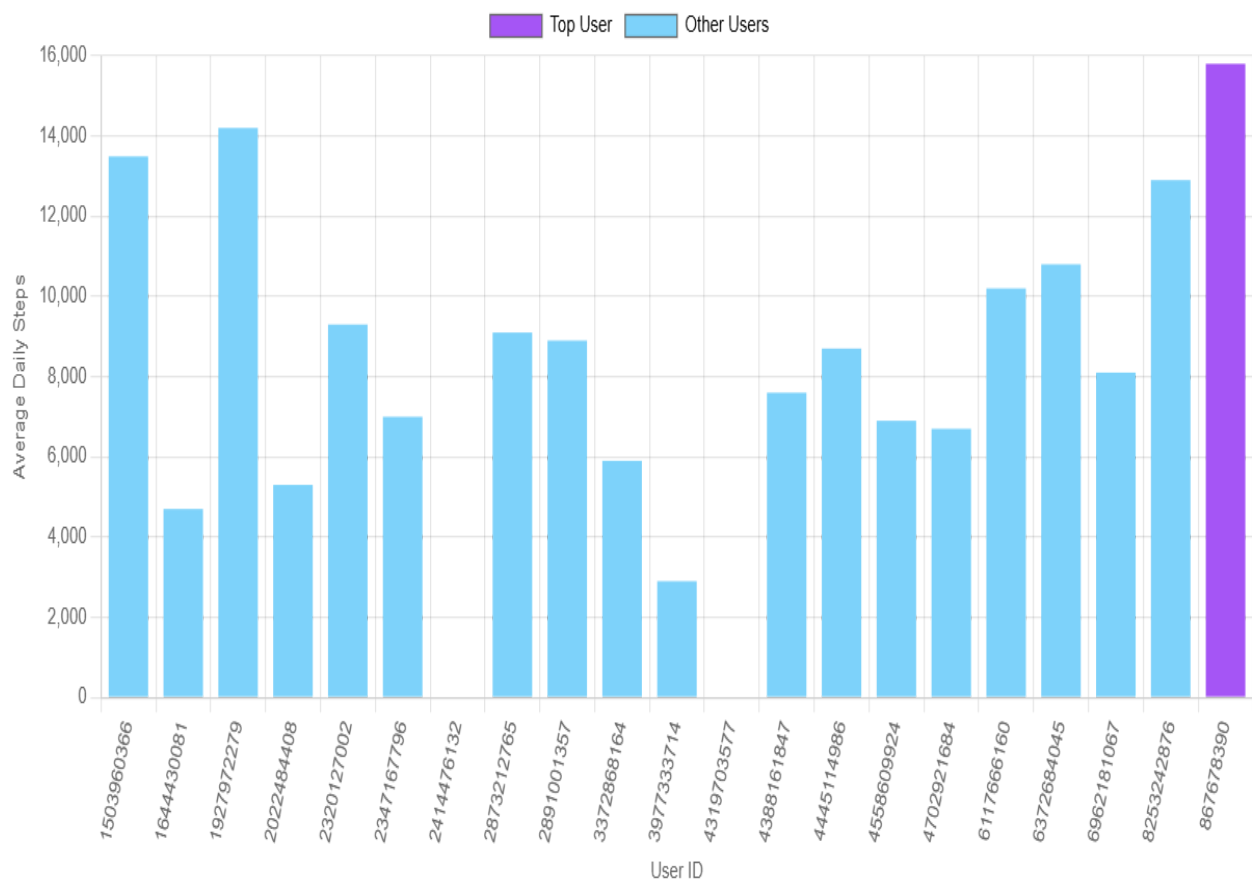
### Visualizing the Analysis

```
#visualizing
```

```
ggplot(user_avg_steps, aes(x = factor(Id), y = AverageDailySteps, f  
  geom_col() +  
  scale_fill_manual(values = c("Top User" = "purple", "Other Users"  
  labs(title = "Average Daily Steps by User", x = "User ID", y = "Av  
  theme_minimal() +  
  theme(axis.text.x=element_text(angle = 45, hjust = 1))
```

### Result:

This visual highlights daily steps per user, with one user significantly outperforming others in step count.



### Action:

Encourage users to set and gradually increase their daily step goals using gamified elements in the app.

Features like virtual challenges, achievement badges, and community leaderboards can motivate users to stay active and track progress toward healthier habits.

## 3. Sleep & Activity Correlation

## Findings:

Users who sleep between six and eight hours a night have the highest average step counts when compared to those who sleep less than six or more than eight hours. This suggests that the optimal amount of sleep and increased physical activity are closely related.

## Analysis process:

### Loading the dataset

```
sleep <- read.csv("/minuteSleep_merged.csv")
View(sleep)

steps <- read.csv("/hourlySteps_merged.csv")
View(steps)
```

### Cleaning & Organizing the data

```
# Convert date column to Date type
sleep$date <- as.Date(sleep$date, format = "%m/%d/%Y %I:%M:%S %p")

daily_sleep <- sleep %>%
  group_by(Id, Date) %>%
  summarise(TotalMinutesAsleep = sum(value), .groups = "drop")

steps$ActivityHour <- as.POSIXct(steps$ActivityHour, format = "%m/%d/%Y %I:%M:%S %p")
steps$date <- as.Date(steps$ActivityHour)

daily_steps <- steps %>%
  group_by(Id, Date) %>%
  summarise(TotalSteps = sum(StepTotal), .groups = "drop")

merged_data <- inner_join(daily_steps, daily_sleep, by = c("Id", "Date"))
View(merged_data)

print(paste("Rows after merge:", nrow(merged_data)))

clean_data <- merged_data %>%
  mutate(SleepCategory = case_when(
    TotalMinutesAsleep < 360 ~ "Less than 6 hrs",
    TotalMinutesAsleep < 480 ~ "6-8 hrs",
    TRUE ~ "More than 8 hrs"
  ))
avg_steps_by_sleep <- clean_data %>%
  group_by(SleepCategory) %>%
  summarise(AverageSteps = mean(TotalSteps), .groups = "drop")
```

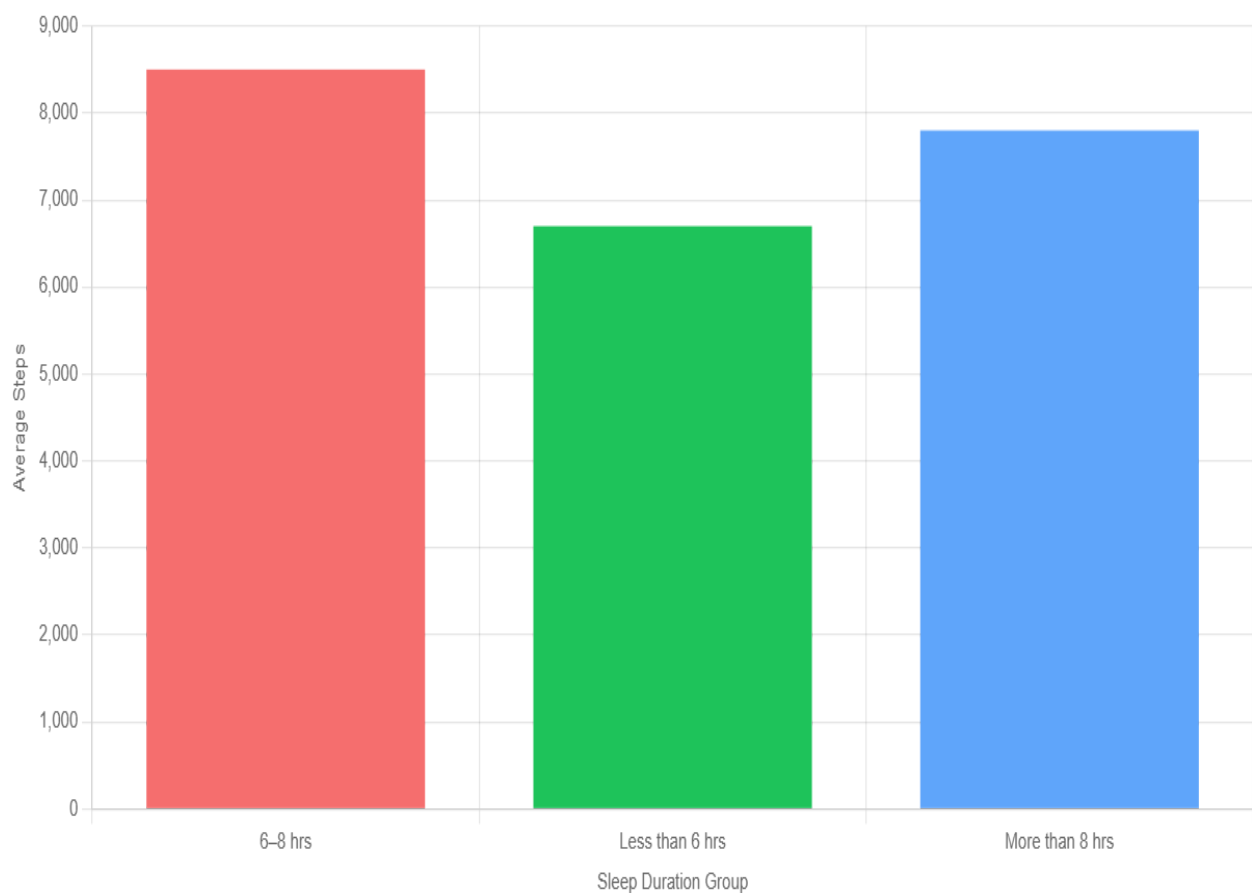


### Visualizing the analysis

```
ggplot(avg_steps_by_sleep, aes(x = SleepCategory, y = AverageSteps,  
  geom_col(width = 0.6) +  
  labs(title = "Average Daily Steps by Sleep Duration",  
    x = "Sleep Duration Group",  
    y = "Average Steps") +  
  theme_minimal() +  
  theme(legend.position = "none"))
```

### Result:

Users who sleep between 6 to 8 hours tend to have higher average daily steps compared to those with shorter or longer sleep durations.



### Action:

Implementing personalized sleep tracking and insights into the Bellabeat app. Promote healthy sleep habits through wellness challenges focused on getting 6–8 hours of sleep with educational content and gentle prompting designed to enhance sleep and daily activity.

## 4. BMI Analysis



## Findings:

According to their weight logs, the average BMI is 25.73, which means the average user falls just into the overweight category according to WHO. Mean weight is 73.44 kg, mean percent body fat is 16%. Although these thresholds may differ between age groups and sexes, the BMI insight indicates an overall trend that a subset of users may be good candidates for additional support in weight management.

## Analysis process:

### Importing data

```
data <- read.csv("/WeightLogInfo_merged.csv")
View(data)
```

### Cleaning & Organizing the data

```
# Apply clean names and fix date
data <- data %>%
  clean_names() %>%
  mutate(date = mdy_hms(date)) # Convert to proper datetime format

# View first rows and structure
View(data)
head(data)
glimpse(data)
summary(data)

# Unique users
n_distinct(data$id)

# Summary stats
weight_summary <- data %>%
  summarise(
    avg_wght_kg = mean(weight_kg, na.rm = TRUE),
    avg_bmi = mean(bmi, na.rm = TRUE),
    avg_fat = mean(fat, na.rm = TRUE)
  )
View(weight_summary)
print(weight_summary)

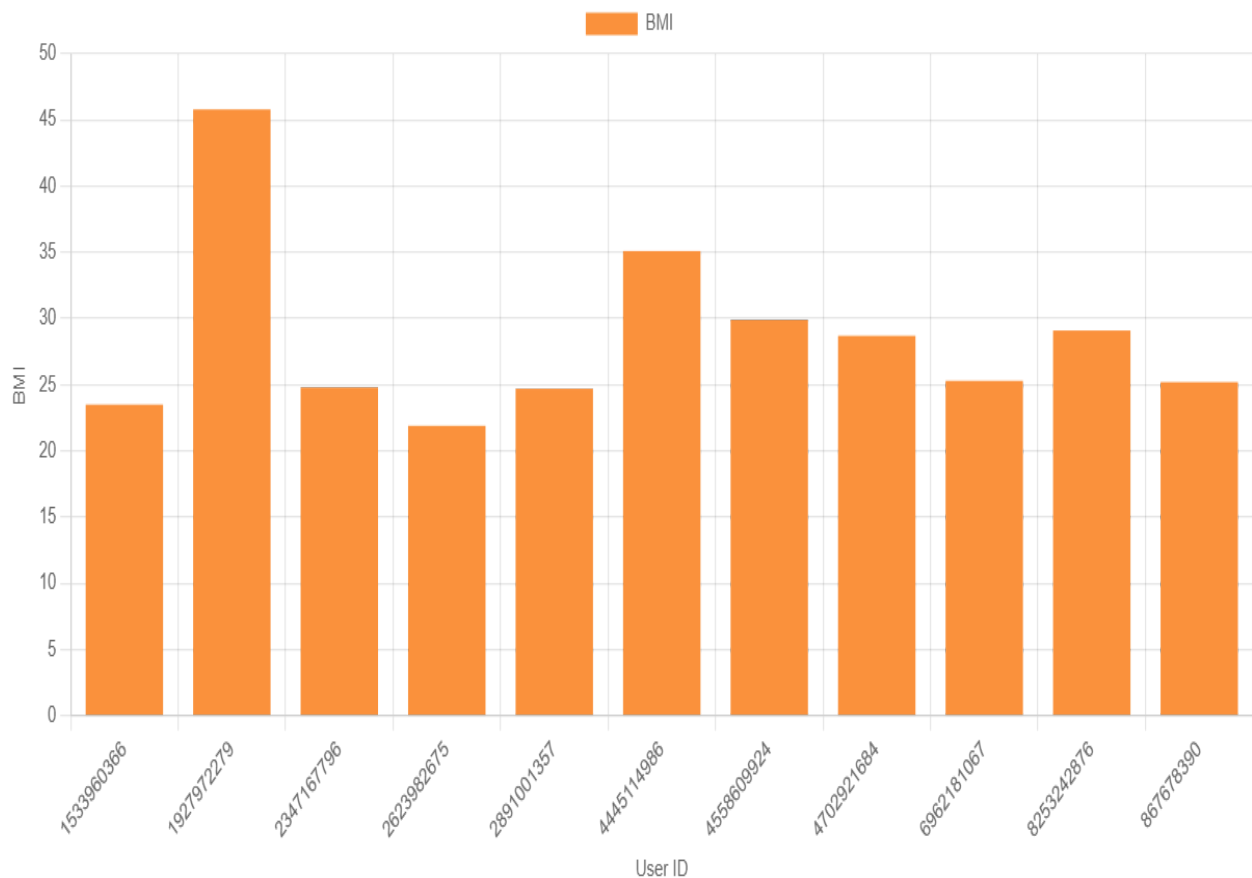
# Filter for latest BMI per user (and drop NAs)
bmi_per_user <- data %>%
  filter(!is.na(bmi)) %>%
  group_by(id) %>%
  arrange(desc(date)) %>%
  slice(1) %>%
  ungroup()

# Plot BMI per user
ggplot(bmi_per_user, aes(x = factor(id), y = bmi)) +
  geom_bar(stat = "identity", fill = "coral", width = 0.8) +
  labs(title = "BMI per User", x = "User ID", y = "BMI") +
  theme(axis.text.x = element_text(angle = 45, size = 6.7))
```

## Visualizing the data

```
#Per user analysis
data$bmi_category <- cut(data$bmi,
                          breaks = c(0, 18.5, 24.9, 29.9, Inf),
                          labels = c("Underweight", "Normal", "Overv
```

**Results:** This chart presents the BMI values of users. Most fall within the healthy range, though a few outliers show elevated BMI levels.



### Action:

Bellabeat is to implement components for BMI tracking and mindfulness. This involves setting personalised goals, education about having a healthy BMI, and regular prompts to weigh in. For Bellabeat the aim of being able to provide users with a better understanding and control over their BMI, the result is a more nurturing and catered health and wellness dynamic.

## Recommendations

- **Peak Activity Times:** The majority of users get moving between 4 PM and 7 PM when you could be sure to prompt app usage or a challenge.
- **Sleep Activity Connection:** Those who get 6 and up to 8 hours of sleep daily achieve higher average daily step counts and make the point that rest helps you keep moving.
- **Low Step Count:** There are a few days around 6565 steps on average. It is less than the generally accepted 10,000 steps/day this points toward potential inspirational features or coaching.
- **The sleep halves:** According to Zee and colleagues, the man or woman who sleeps under 6 hours or over 8 hours are equally less active than the normal sleeping individual, showing a kind of U-shape curve relationship between activity expenditure and sleeping duration.
- **BMI beyond healthy range:** Taking a closer look at users' BMI, the average BMI is 25.73, slightly beyond the healthy range (18.5-24.9), with which many of the users are categorized in the overweight range, indicating a potential need for personalized weight management support.

## Final Conclusion

Analysis of Fitbit data exposes important health trends, which Bellabeat can reference in an ongoing effort to improve user engagement and marketing. Notable behavioral insights (i.e., added activity during the evening, beneficial effect of 6–8 h of sleep on activity levels, and the slightly higher average BMI outcome) provide a good basis for designing personalized wellness inspirations.

By scheduling product features according to user behavior such as boosting activity during peak-hours, ensuring enough sleep, managing optimal BMI, etc., Bellabeat's products can offer more compelling, engaging and relevant wellness experiences. Adopting these data-based approaches not only encourages healthier lifestyles, but also enhances consumer retention, app usage, and brand promotion in the tight health-tech market.

This case study demonstrates how thoughtful data analysis can translate into strategic decisions that benefit both the user and the business, positioning Bellabeat as a smart, responsive, and wellness-centric company in the digital health market.

## Portfolio Summary

For this project, I examined the activity, sleep, and weight trends of 30 Fitbit users. After cleaning, transforming, and visualizing the data in R, I found users to be most active during evening hours from 4 to 7 PM. On average, users took around 6,565 steps a day. Additionally, those who slept between 6 to 8 hours were more likely to be active.

Analysis on BMI indicated that the average user is classified as overweight with a BMI of

25.73 which indicates that tailored intervention may be beneficial. These insights can help shape targeted marketing campaigns and wellness initiatives for Bellabeat.

© 2025 Krishnanunni H Pillai | Data Analytics Portfolio