How Can a Wellness Technology Company Play It Smart?

Sutiasmoko

2024-02-13

Scenario

I am a junior data analyst working on the marketing analyst team at Bellabeat, a high-tech manufacturer of health-focused products for women. Bellabeat is a successful small company, but they have the potential to become a larger player in the global smart device market. cofounder and Chief Creative Officer of Bellabeat, believes that analyzing smart device fitness data could help unlock new growth opportunities for the company. I have been asked to focus on one of Bellabeat's products and analyze smart device data to gain insight into how consumers are using their smart devices. The insights I discover will then help guide marketing strategy for the company. I will present my analysis to the Bellabeat executive team along with my high-level recommendations for Bellabeat's marketing strategy.

- Products
- Bellabeat app: The Bellabeat app provides users with health data related to their activity, sleep, stress, menstrual cycle, and mindfulness habits. This data can help users better understand their current habits and make healthy decisions. The Bellabeat app connects to their line of smart wellness products.
- Leaf: Bellabeat's classic wellness tracker can be worn as a bracelet, necklace, or clip. The Leaf tracker connects to the Bellabeat app to track activity, sleep, and stress.
- Time: This wellness watch combines the timeless look of a classic timepiece with smart technology to track user activity, sleep, and stress. The Time watch connects to the Bellabeat app to provide you with insights into your daily wellness.

Ask

Chief Creative Officer asks me to analyze smart device usage data in order to gain insight into how consumers use non-Bellabeat smart devices. She then wants me to select one Bellabeat product to apply these insights to in my presentation. These questions will guide the analysis:

1. What are some trends in smart device usage?
2. How could these trends apply to Bellabeat customers?
3. How could these trends help influence Bellabeat marketing strategy?

I will produce a report with the following deliverables:

1. A clear summary of the business task
2. A description of all data sources used
3. Documentation of any cleaning or manipulation of data
4. A summary of your analysis
5. Supporting visualizations and key findings
6. Your top high-level content recommendations based on your analysis

Produce a Report

1. A clear summary of the business task This analysis focuses on customer behavior regarding non-Bellabeat products, aiming to depict relationships, trends, and distributions of measurements while referring to the minimum and optimal thresholds for steps, calories, active minutes, and sleep as recommended by WHO and various health articles.

2. A description of all data sources used *FitBit Fitness Tracker Data (Link)* (CC0: Public Domain, dataset made available through *Mobius (Link)* This Kaggle data set contains personal fitness tracker from thirty fitbit users. Thirty eligible Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity and sleep monitoring. It includes information about daily activity and steps that can be used to explore users' habits.

3. Documentation of any cleaning or manipulation of data

- Create and update columns

```r
# add columns in to sleepDay
sleepDay <- sleepDay %>%
  mutate (date_hours = mdy_hms(SleepDay),
          date = as.Date(date_hours))

# add columns in to daily_activity
daily_activity <- daily_activity %>%
  mutate (date = mdy(ActivityDate))

# Join data
joined_data <- full_join (daily_activity, sleepDay, by = c("Id" = "Id", "date" = "date"))

# add column DayOfWeek
joined_data <- joined_data %>%
  mutate(DayOfWeek=weekdays(date),
         DayOfWeek = factor(DayOfWeek,
                            levels = c("Monday", "Tuesday", "Wednesday",
                                       "Thursday", "Friday", "Saturday",
                                       "Sunday"))) %>%
  arrange(DayOfWeek)

# add steps category
joined_data <- joined_data %>%
  mutate(Steps_Category = ifelse(TotalSteps < 10000, "< 10000", ">= 10000"))

# create Joined_data_dayofweek
Joined_data_dayofweek <- joined_data %>%
  group_by(Id, DayOfWeek) %>%
  summarise(Count_Id = n_distinct(Id),
            Calories = as.integer(round(mean(Calories))),
            TotalSteps = as.integer(round(mean(TotalSteps))),
            SedentaryMinutes = as.integer(round(mean(SedentaryMinutes))),
            VeryActiveMinutes = as.integer(round(mean(VeryActiveMinutes))),
            TotalMinutesAsleep = as.integer(round(mean(TotalMinutesAsleep))))

# Calculate the total Count_Id for each Steps_Category on each DayOfWeek.
```

```r
total_counts <- Joined_data_dayofweek %>%
  group_by(DayOfWeek) %>%
  summarise(total_count_category = sum(Count_Id))

total_counts_sleep <- Joined_data_dayofweek %>%
  filter(!is.na(TotalMinutesAsleep)) %>%
  group_by(DayOfWeek) %>%
  summarise(total_count_category = sum(Count_Id))

# Steps Calculation & Categorized

Joined_data_dayofweek <- Joined_data_dayofweek %>%
  mutate(Steps_Category = ifelse(TotalSteps < 10000, "Less than 10000", ">= 10000"))

Joined_data_dayofweek_Steps <- Joined_data_dayofweek %>%
  group_by(Steps_Category, DayOfWeek) %>%
  summarise(Count_Id = sum(Count_Id),
            TotalSteps = as.integer(round(mean(TotalSteps)))) %>%
  left_join(total_counts, by = c("DayOfWeek")) %>%
  mutate(percentage = paste0(round(Count_Id / total_count_category * 100), "%")) %>%
  mutate(label_text = paste(Count_Id, ", ", percentage))

# Calories Calculation & Categorized

Joined_data_dayofweek <- Joined_data_dayofweek %>%
  mutate(Calories_Category = ifelse(Calories < 2200, "Less than 2200", ">= 2200"))

Joined_data_dayofweek_Calories <- Joined_data_dayofweek %>%
  group_by(Calories_Category, DayOfWeek) %>%
  summarise(Count_Id = sum(Count_Id),
            Calories = as.integer(round(mean(Calories)))) %>%
  left_join(total_counts, by = c("DayOfWeek")) %>%
  mutate(percentage = paste0(round(Count_Id / total_count_category * 100), "%")) %>%
  mutate(label_text = paste(Count_Id, ", ", percentage))

# Very Active minutes Calculation & Categorized

Joined_data_dayofweek <- Joined_data_dayofweek %>%
  mutate(VeryActiveMinutes_Category = ifelse(VeryActiveMinutes < 22, "Less than 22", ">= 22"))

Joined_data_dayofweek_VeryActiveMinutes <- Joined_data_dayofweek %>%
  group_by(VeryActiveMinutes_Category, DayOfWeek) %>%
  summarise(Count_Id = sum(Count_Id),
            VeryActiveMinutes = as.integer(round(mean(VeryActiveMinutes)))) %>%
  left_join(total_counts, by = c("DayOfWeek")) %>%
  mutate(percentage = paste0(round(Count_Id / total_count_category * 100), "%")) %>%
  mutate(label_text = paste(Count_Id, ", ", percentage))

# Minutes ASleep Calculation & Categorized
Joined_data_dayofweek <- Joined_data_dayofweek %>%
  mutate(TotalMinutesAsleep_Category = ifelse(TotalMinutesAsleep < 420, "Less than 420", ">= 420"))

Joined_data_dayofweek_TotalMinutesAsleep <- Joined_data_dayofweek %>%
```

```r
  filter(!is.na(TotalMinutesAsleep)) %>%
  group_by(TotalMinutesAsleep_Category, DayOfWeek) %>%
  summarise(Count_Id = sum(Count_Id),
            TotalMinutesAsleep = as.integer(round(mean(TotalMinutesAsleep)))) %>%
  left_join(total_counts_sleep, by = c("DayOfWeek")) %>%
  mutate(percentage = paste0(round(Count_Id / total_count_category * 100), "%")) %>%
  mutate(label_text = paste(Count_Id, ", ", percentage))
```

4. A summary of my analysis

- The more steps, the longer the very active minutes and the more calories burned; conversely, the fewer sedentary minutes.

- The calories burned is strongly related to amount of steps and very active minutes.

- From 33 IDs, the majority of step categories are still less than 10,000 steps per day.

- From 33 IDs, the majority of calories burned categories are still less than 2200 calories per day.

- From 33 IDs, the majority of very active minutes categories are still less than 22 minutes per day.

- From 15 IDs, the majority of minutes asleep categories are more than or equal to 420 minutes per day.

5. Supporting visualizations and key findings

- **Relationship to Total Steps**

R Script: Relationship Between Very Active Minutes and Total Steps

```
ggplot(data=Joined_data_dayofweek, aes(x= VeryActiveMinutes, y= TotalSteps)) +
  geom_point() + geom_smooth(method = "lm", se = TRUE) +
  labs(title = "Relationship Between Very Active Minutes and Total Steps",
       x = "Very Active Minutes",
       y = "Total Steps")
```

Relationship Between Very Active Minutes and Total Steps
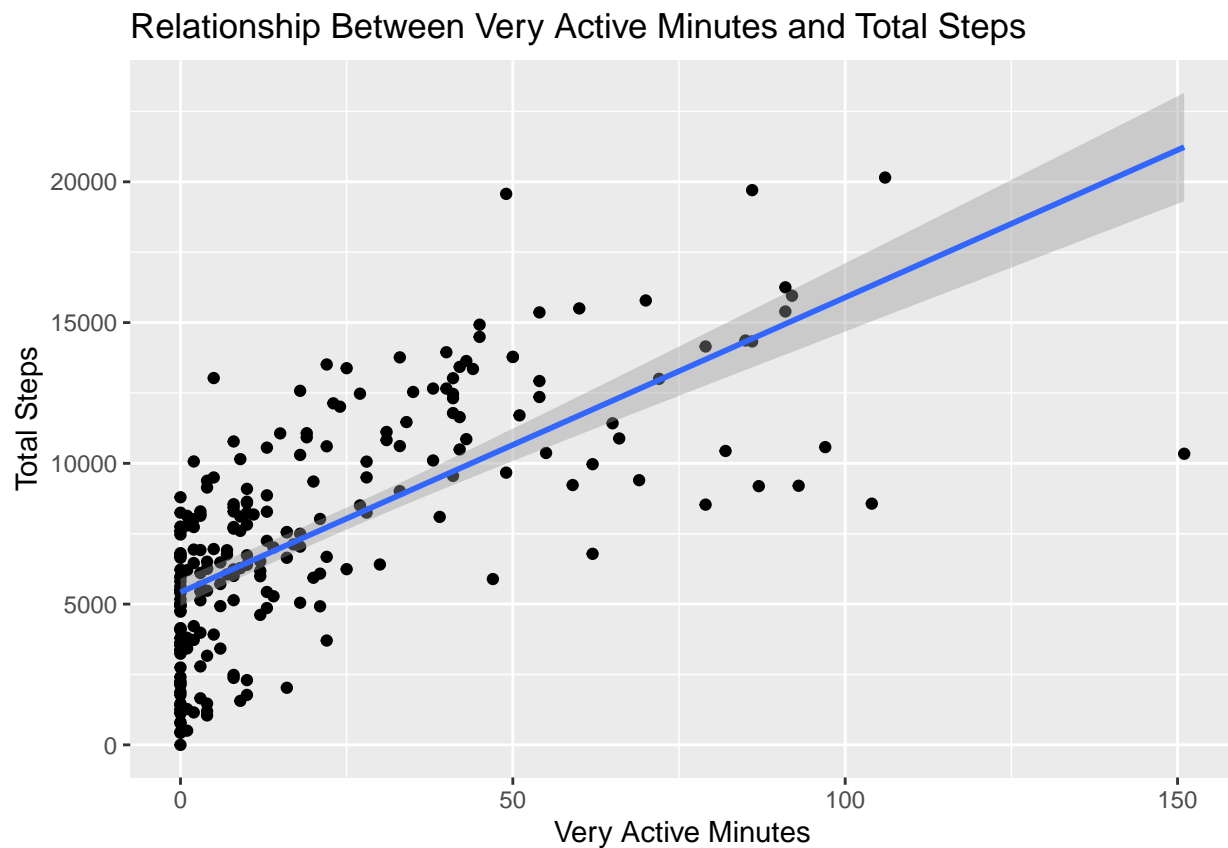
The more steps, the longer the very active minutes.

*R Script: Relationship Between Calories Burned and Total Steps*

```r
ggplot(data=Joined_data_dayofweek, aes(x= Calories, y=TotalSteps)) +
  geom_point() + geom_smooth(method = "lm", se = TRUE) +
  labs(title = "Relationship Between Calories Burned and Total Steps",
       x = "Calories Burned",
       y = "Total Steps")
```

Relationship Between Calories Burned and Total Steps

The more steps, the more calories burned.

R Script: Relationship Between Sedentary Minutes and Total Steps

```
ggplot(data=Joined_data_dayofweek, aes(x= SedentaryMinutes, y= TotalSteps)) +
  geom_point() + geom_smooth(method = "lm", se = TRUE) +
  labs(title = "Relationship Between Sedentary Minutes and Total Steps",
       x = "Sedentary Minutes",
       y = "Total Steps")
```
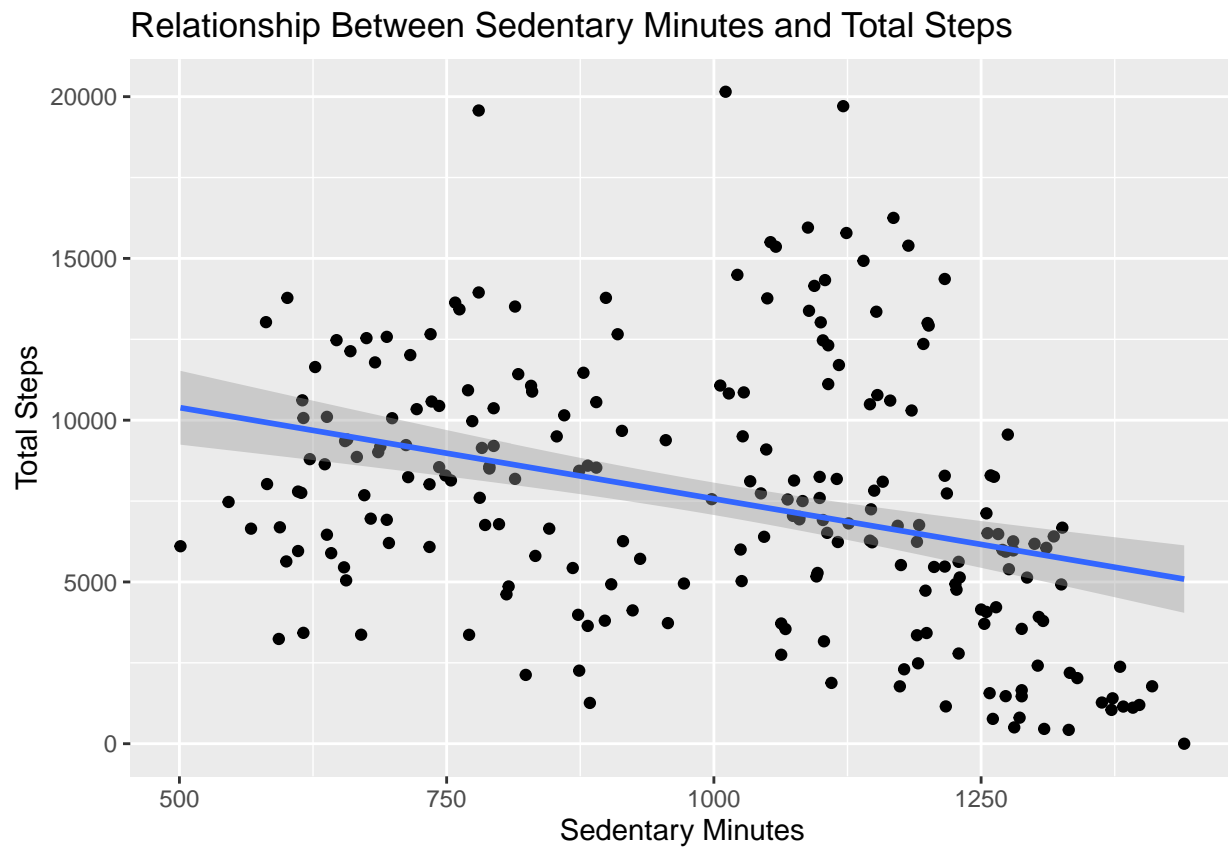

The more steps, the fewer sedentary minutes.

R Script: Calories Burned by VeryActiveMinutes and Steps Category

```
ggplot(Joined_data_dayofweek, aes(x = VeryActiveMinutes_Category , y = Steps_Category, fill = Calories))
  geom_tile() +
  scale_fill_gradient(low = "white", high = "black") +
  labs(title = "Heatmap of Calories Burned by VeryActiveMinutes and Steps Category",
       x = "VeryActiveMinutes Category",
       y = "Steps Category",
       fill = "Calories Burned")+
  theme_minimal()
```
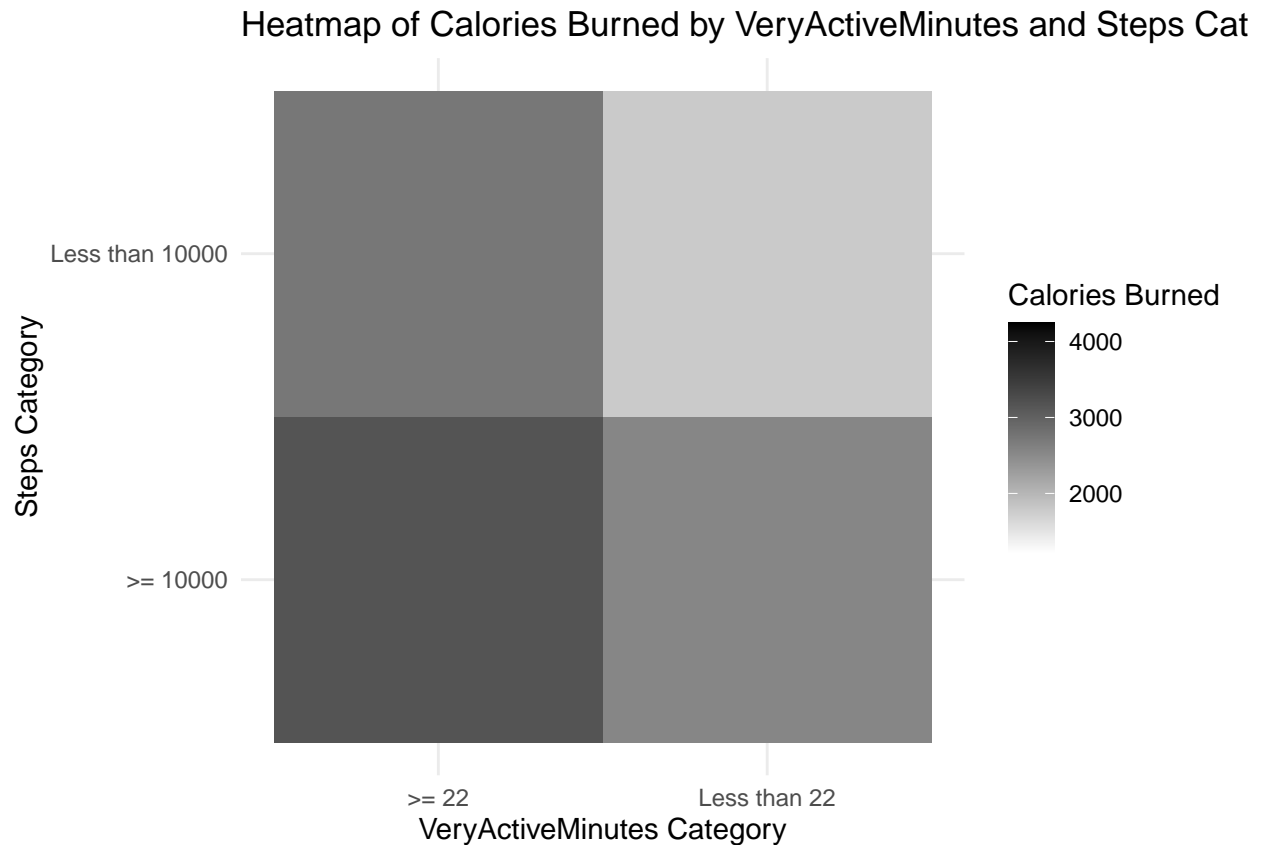
Heatmap of Calories Burned by VeryActiveMinutes and Steps Cat

The calories burned is strongly related to amount of steps and very active minutes.

- **Day of week trend**

Steps Day of week trend

R Script:

```
ggplot(Joined_data_dayofweek_Steps, aes(x = DayOfWeek, y = Count_Id,
                                        group = Steps_Category, fill= Steps_Category)) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = label_text), position=position_stack(vjust = 0.5), colour = "white") +
  labs(title = "Weekly average Steps from 33 IDs",
       x = "Day Of Week",
       y = "Count IDs",
       fill = "Steps Category") +
  theme_minimal()
```
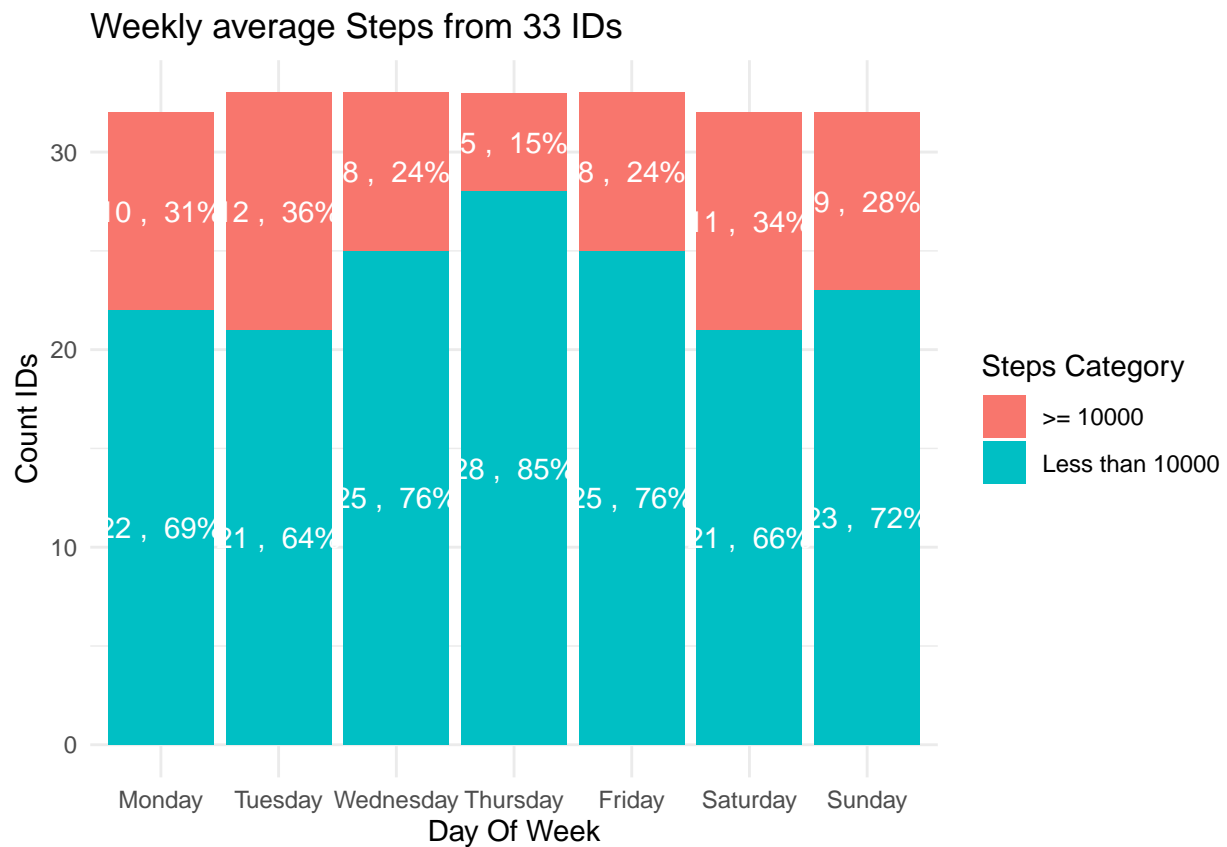

How Many Steps Do I Need a Day? MThe Centers for Disease Control and Prevention (CDC) recommend doing at least 150 minutesTrusted Source of moderate exercise per week, but many people also aim to walk 10,000 steps a day. (Article Link)

From 33 IDs, the majority of step categories are still less than 10,000 steps per day.

Calories Day of week trend

R Script:

```
ggplot(Joined_data_dayofweek_Calories, aes(x = DayOfWeek, y = Count_Id,
                                            group = Calories_Category, fill= Calories_Category)) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = label_text), position=position_stack(vjust = 0.5), colour = "white") +
  labs(title = "Weekly average Calories from 33 IDs",
       x = "Day Of Week",
       y = "Count IDs",
       fill = "Calories Category") +
  theme_minimal()
```

Weekly average Calories from 33 IDs
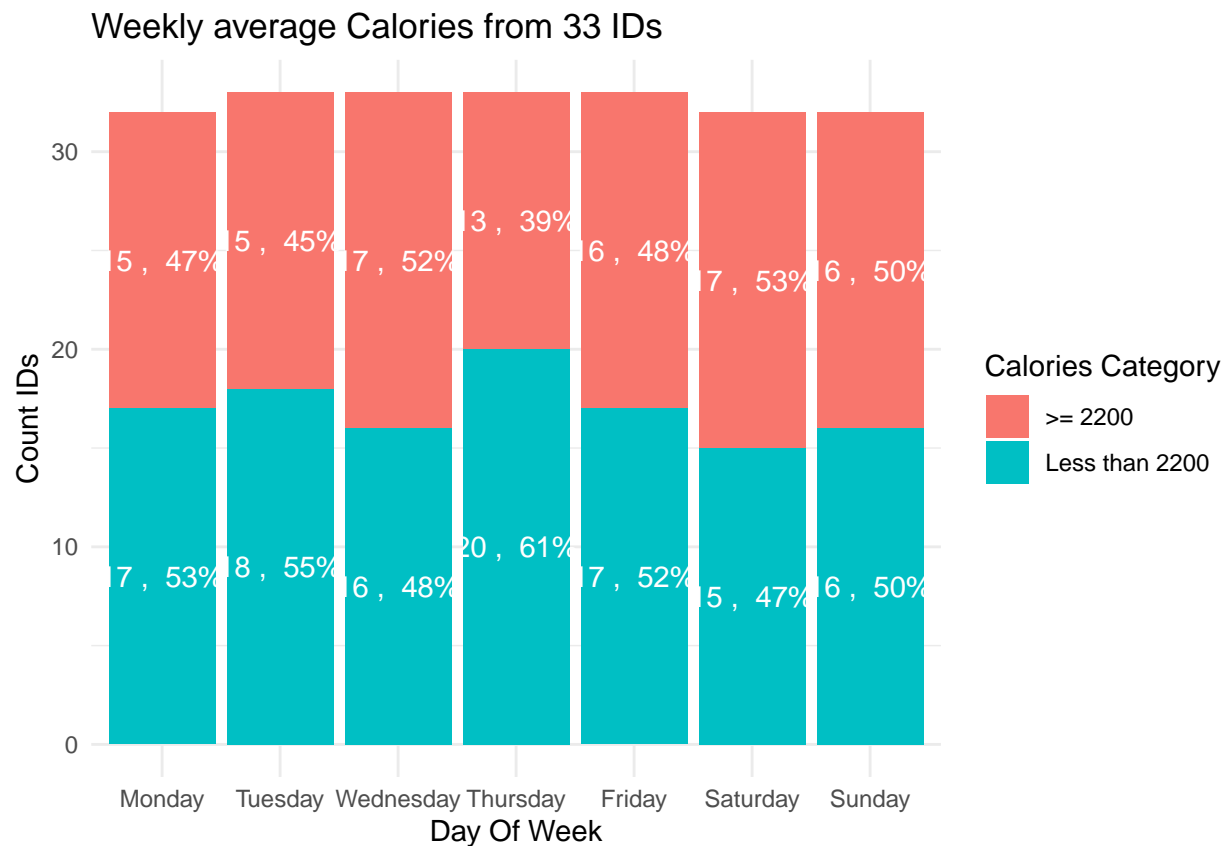
How Many Calories Do I Burn in a Day? While the average male and female need roughly 2,200–3,000 and 1,600–2,200 calories per day, respectively, your needs may differ depending on your height, weight, and activity level. (Article Link)
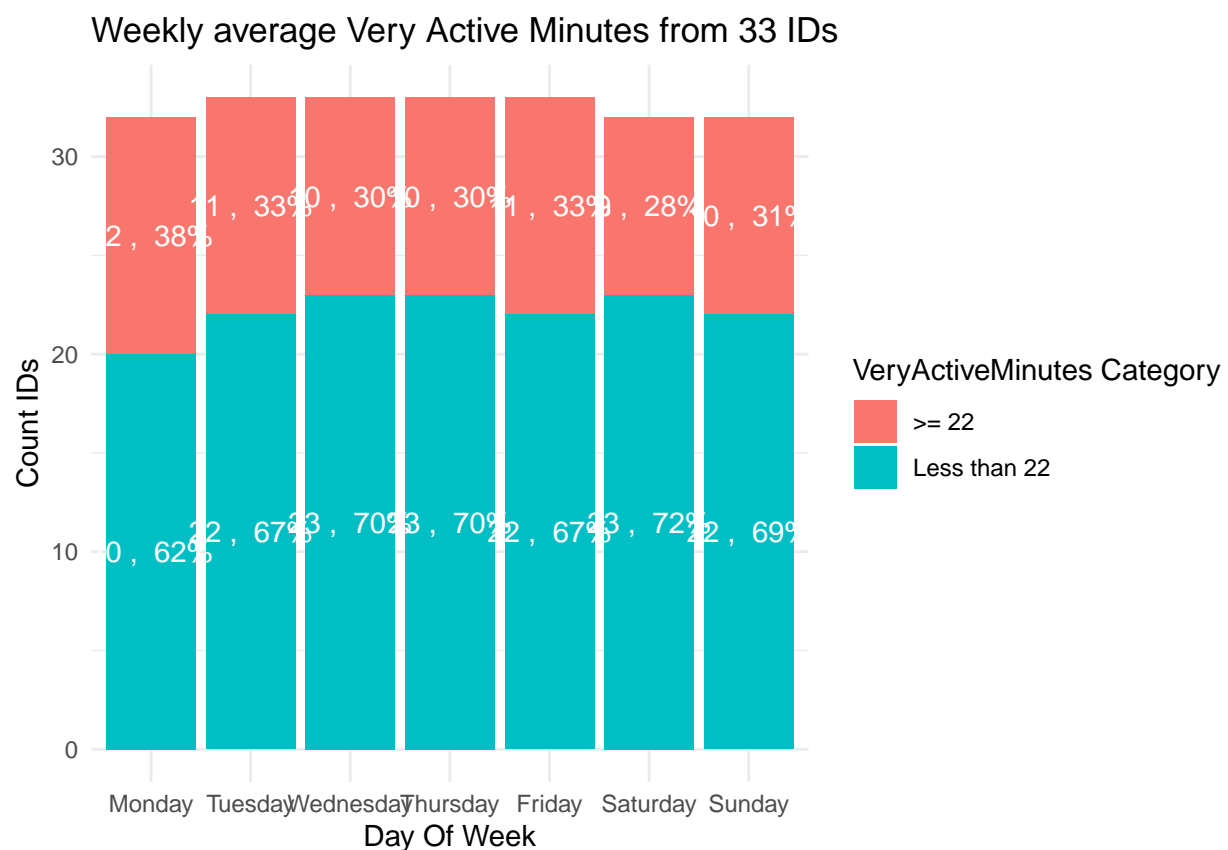
From 33 IDs, the majority of calories burned categories are still less than 2200 calories per day.

Very Active Minutes Day of week trend

R Script:

```r
ggplot(Joined_data_dayofweek_VeryActiveMinutes, aes(x = DayOfWeek, y = Count_Id,
                                                    group = VeryActiveMinutes_Category,
                                                    fill= VeryActiveMinutes_Category)) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = label_text), position=position_stack(vjust = 0.5), colour = "white") +
  labs(title = "Weekly average Very Active Minutes from 33 IDs",
       x = "Day Of Week",
       y = "Count IDs",
       fill = "VeryActiveMinutes Category") +
  theme_minimal()
```


Adults aged 18–64 years should do at least 150–300 minutes of moderate-intensity aerobic physical activity, or at least 75–150 minutes of vigorous-intensity aerobic physical activity; or an equivalent combination of moderate- and vigorous-intensity activity throughout the week. (Article Link) (150minutes/week = 22minutes/day)

From 33 IDs, the majority of very active minutes categories are still less than 22 minutes per day.

Total Minutes Asleep Day of week trend

R Script:

```r
ggplot(Joined_data_dayofweek_TotalMinutesAsleep, aes(x = DayOfWeek, y = Count_Id, group = TotalMinutesAs
  geom_bar(stat = "identity") +
  geom_text(aes(label = label_text), position=position_stack(vjust = 0.5), colour = "white") +
  labs(title = "Weekly Total Minutes Asleep from 15 IDs",
       x = "Day Of Week",
       y = "Count IDs",
       fill = "Minutes Asleep Category") +
  theme_minimal()
```

Weekly Total Minutes Asleep from 15 IDs
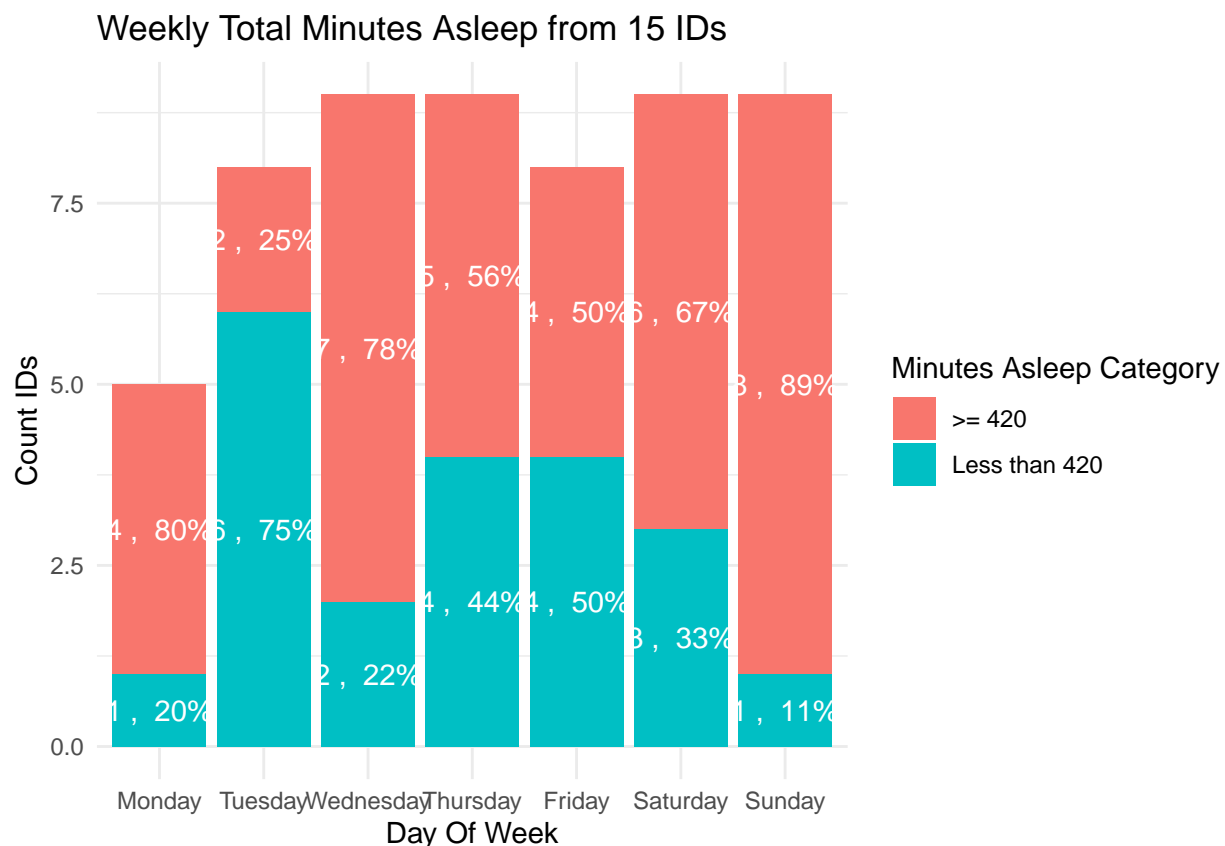
On average, how many hours do you sleep each night? For most healthy adults, guidelines suggest at least seven hours of slumber. (Article Link), (7 hours = 420 minutes)

From 15 IDs, the majority of minutes asleep categories are more than or equal to 420 minutes per day.

6. My top high-level content recommendations based on my analysis "The majority of users generate fewer steps and fewer active minutes than recommended, Bellabeat could offer additional solutions or features to encourage their customers to achieve their daily step and active minutes targets. This could include notifications or prompts to increase activity, or better analysis tools to help customers understand their activity patterns."