Bellabeat Case Study

Rudhra

2024-07-31

Case Study : How Can a Wellness Technology Company Play It Smart?

Shareholder asked the marketing analytics team to focus on a Bellabeat product and analyze smart device usage data to gain insight into how people are already using their smart devices. Then, using this information, she would like high-level recommendations for how these trends can inform Bellabeat marketing strategy.

Business task: The business task involves analyzing consumer usage data of smart devices, specifically non-Bellabeat devices, to gain insights into user behavior. The objective is to inform and develop a targeted marketing strategy for Bellabeat products, leveraging the identified trends and insights.

Questions to guide the analysis

What are some trends in smart device usage?

How could these trends apply to Bellabeat customers?

How could these trends help influence Bellabeat marketing strategy?

Case Study Roadmap – Ask

Guiding questions

What is the problem you are trying to solve?

Bellabeat is a **small company** but striving to become a larger player in the global smart device market. Looking for **Growth opportunity** and **marketing strategy** to become a larger player in the global market.

How can your insights drive business decisions?
My insights should help the Bellabeat company to add the user smart device usage insights in to the Bellabeat product features which will benefit the woman's health and develop the marketing strategy accordingly.

Key tasks

Identify the business task : To analyze smart device usage data to gain insight into how consumers use non-Bellabeat smart devices and to select one Bellabeat product to apply these insights to in my presentation.

Consider key stakeholders: Bellabeat executive team
Urška Sršen: Bellabeat's cofounder and Chief Creative Officer
Sando Mur: Mathematician and Bellabeat's cofounder; key member of the Bellabeat executive team

Deliverable :

A clear statement of the business task: To analyze smart device usage data to gain insight into how consumers use non-Bellabeat smart devices and to select one Bellabeat product to apply these insights in to my presentation of analysis to the Bellabeat executive team along with my high-level recommendations for Bellabeat's marketing strategy to unlock new growth opportunities for the company.

Case Study Roadmap - Prepare

Where is your data stored?
https://www.kaggle.com/datasets/arashnic/fitbit (CC0: Public Domain, dataset made available through Mobius)

How is the data organized? Is it in long or wide format?
It's in wide format

How are you addressing licensing, privacy, security, and accessibility?

Bellabeat should ensure user consent for data collection, implement strong encryption for storing sensitive information and anonymize data for analysis and maintain the privacy of user information in smart device usage data.

Installed and loaded the tidyverse packages

```
install.packages('tidyverse')

## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.4'
## (as 'lib' is unspecified)

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse
2.0.0 ──
## ✓ dplyr      1.1.4     ✓ readr      2.1.5
## ✓ forcats    1.0.0     ✓ stringr    1.5.1
## ✓ ggplot2    3.5.1     ✓ tibble     3.2.1
## ✓ lubridate 1.9.3     ✓ tidyr      1.3.1
## ✓ 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
```

Loaded CSV files and Created a dataframe named 'daily_activity'

```
daily_activity <- read.csv("dailyActivity_merged.csv")
```

Created another dataframe for the sleep data.

```
sleep_day <- read.csv("sleepDay_merged.csv")
```

Explored a few key tables

Looked at the daily_activity data.

```
head(daily_activity)

##            Id ActivityDate TotalSteps TotalDistance TrackerDistance
## 1 1503960366    3/25/2016      11004          7.11            7.11
## 2 1503960366    3/26/2016      17609         11.55           11.55
## 3 1503960366    3/27/2016      12736          8.53            8.53
## 4 1503960366    3/28/2016      13231          8.93            8.93
## 5 1503960366    3/29/2016      12041          7.85            7.85
## 6 1503960366    3/30/2016      10970          7.16            7.16
##    LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 1                         0               2.57                     0.46
## 2                         0               6.92                     0.73
```

```
## 3                           0              4.66                    0.16
## 4                           0              3.19                    0.79
## 5                           0              2.16                    1.09
## 6                           0              2.36                    0.51
##   LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
## 1                4.07                       0                33
## 2                3.91                       0                89
## 3                3.71                       0                56
## 4                4.95                       0                39
## 5                4.61                       0                28
## 6                4.29                       0                30
##   FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories
## 1                  12                  205              804     1819
## 2                  17                  274              588     2154
## 3                   5                  268              605     1944
## 4                  20                  224             1080     1932
## 5                  28                  243              763     1886
## 6                  13                  223             1174     1820
```

Identifed all the columns in the daily_activity data.

```
colnames(daily_activity)

##  [1] "Id"                  "ActivityDate"
##  [3] "TotalSteps"          "TotalDistance"
##  [5] "TrackerDistance"     "LoggedActivitiesDistance"
##  [7] "VeryActiveDistance"  "ModeratelyActiveDistance"
##  [9] "LightActiveDistance" "SedentaryActiveDistance"
## [11] "VeryActiveMinutes"   "FairlyActiveMinutes"
## [13] "LightlyActiveMinutes" "SedentaryMinutes"
## [15] "Calories"
```

Looked at the sleep_day data.

```
head(sleep_day)

##            Id               SleepDay TotalSleepRecords TotalMinutesAsleep
## 1 1503960366 4/12/2016 12:00:00 AM                 1                327
## 2 1503960366 4/13/2016 12:00:00 AM                 2                384
## 3 1503960366 4/15/2016 12:00:00 AM                 1                412
## 4 1503960366 4/16/2016 12:00:00 AM                 2                340
## 5 1503960366 4/17/2016 12:00:00 AM                 1                700
## 6 1503960366 4/19/2016 12:00:00 AM                 1                304
##   TotalTimeInBed
## 1            346
## 2            407
## 3            442
## 4            367
```

```
## 5               712
## 6               320
```

Identified all the columns in the daily_activity data.

```
colnames(sleep_day)
```

```
## [1] "Id"                 "SleepDay"         "TotalSleepRecords"
## [4] "TotalMinutesAsleep" "TotalTimeInBed"
```

Noted that both datasets have the 'Id' field which can be used to merge the datasets.

Understanding some summary statistics

How many unique participants are there in each dataframe?

```
n_distinct(daily_activity$Id)
```

```
## [1] 35
```

```
n_distinct(sleep_day$Id)
```

```
## [1] 24
```

There are more participants in the daily activity dataset than the sleep dataset.

No. of observations in each dataframe

```
nrow(daily_activity)
```

```
## [1] 457
```

```
nrow(sleep_day)
```

```
## [1] 413
```

quick summary statistics about each data frame

summary of daily activity dataframe

```
daily_activity %>%
  select(TotalSteps,
         TotalDistance,
         SedentaryMinutes) %>%
  summary()
```

```
##     TotalSteps      TotalDistance      SedentaryMinutes
##  Min.   :    0   Min.   : 0.000    Min.   :  32.0
##  1st Qu.: 1988   1st Qu.: 1.410    1st Qu.: 728.0
##  Median : 5986   Median : 4.090    Median :1057.0
##  Mean   : 6547   Mean   : 4.664    Mean   : 995.3
##  3rd Qu.:10198   3rd Qu.: 7.160    3rd Qu.:1285.0
##  Max.   :28497   Max.   :27.530    Max.   :1440.0
```

For the sleep data frame

```
sleep_day %>%
  select(TotalSleepRecords,
         TotalMinutesAsleep,
         TotalTimeInBed) %>%
  summary()

##  TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
##  Min.   :1.000     Min.   : 58.0      Min.   : 61.0
##  1st Qu.:1.000     1st Qu.:361.0      1st Qu.:403.0
##  Median :1.000     Median :433.0      Median :463.0
##  Mean   :1.119     Mean   :419.5      Mean   :458.6
##  3rd Qu.:1.000     3rd Qu.:490.0      3rd Qu.:526.0
##  Max.   :3.000     Max.   :796.0      Max.   :961.0
```

Plotted a few explorations

```
  ggplot(data=sleep_day, aes(x=TotalMinutesAsleep, y=TotalTimeInBed))+
    geom_point(color="purple") +
  labs(title="Total Minutes Asleep vs. Total Time In Bed", x="Total Minutes
Asleep", y="Total Time In Bed") +
  theme_minimal()
```

Total Minutes Asleep vs. Total Time In Bed
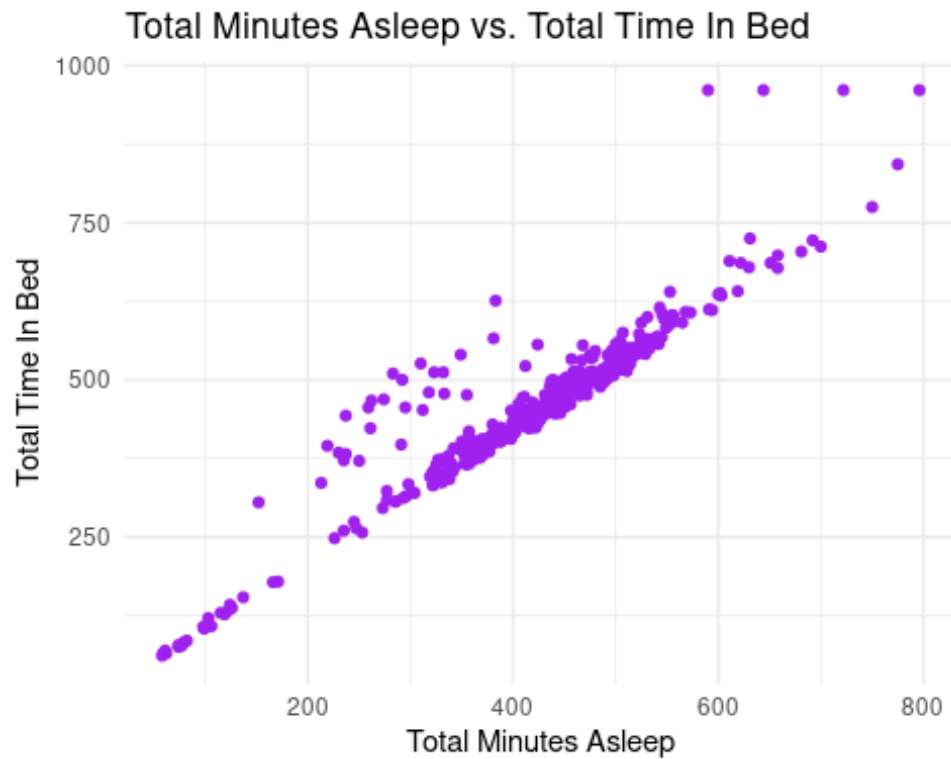
- Positive correlation between TotalMinutesAsleep and TotalTimeInBed, meaning that as the total time in bed increases, the total minutes asleep also increases.
- Ideally, if someone spends more time in bed, they should get proportionally more sleep. A steep slope suggests high sleep efficiency, where most of the time in bed is spent sleeping.

Total Steps per Day

Steps vs. Date

```
ggplot(daily_activity, aes(x = ActivityDate, y = TotalSteps)) +
  geom_line(color="violet") +
  labs(title = "Total Steps per Day", x = "Date", y = "Total Steps")
```
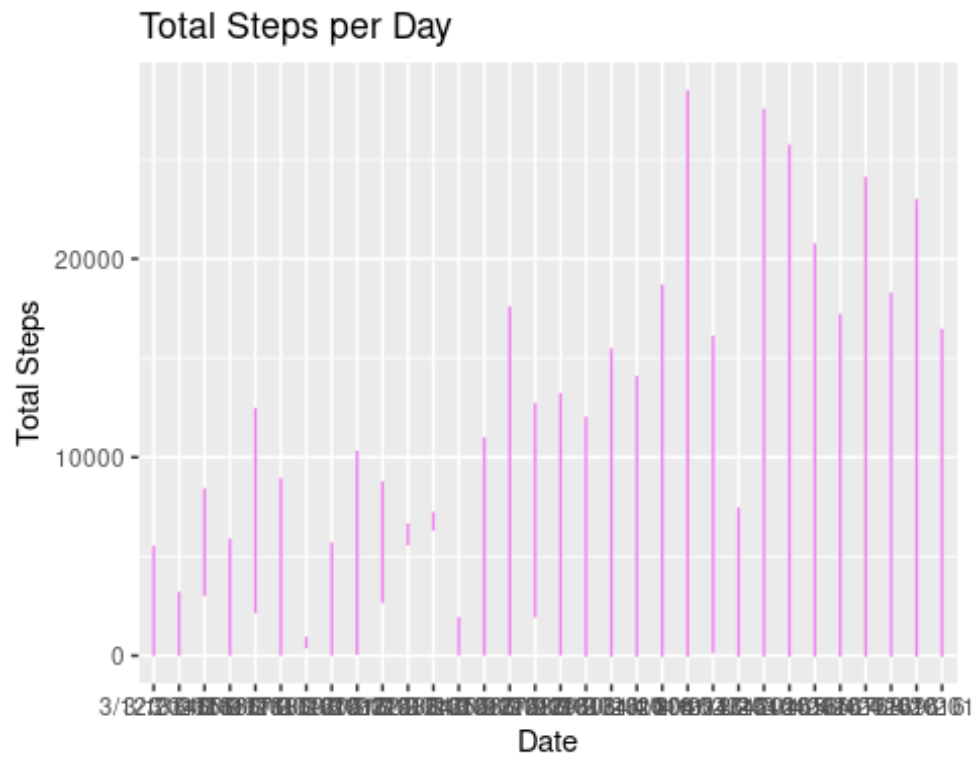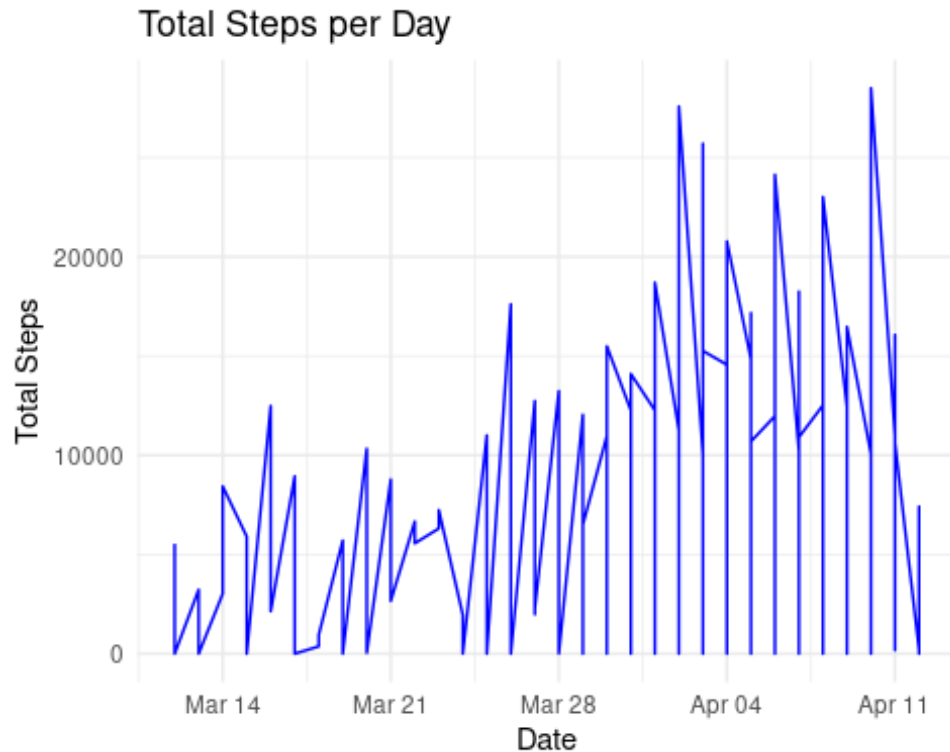
Total Steps per Day

Convert ActivityDate to Date format

```
daily_activity$ActivityDate <- as.Date(daily_activity$ActivityDate,
format="%m/%d/%Y")
```

Plot Total Steps per Day

```
ggplot(daily_activity, aes(x=ActivityDate, y=TotalSteps)) +
  geom_line(color="blue") +
  labs(title="Total Steps per Day", x="Date", y="Total Steps") +
  theme_minimal()
```

Total Steps per Day

Insights:

- **Trends Over Time**: The line plot of total steps per day reveals fluctuations in activity levels. Peaks and troughs may indicate more active days versus less active days, possibly corresponding to weekends or specific events.
- **Consistency**: Users show varying levels of consistency in their daily steps. Some periods may have more consistent activity, while others show significant variation.

Recommendations:

- **Encourage Daily Goals**: To increase consistency, recommend users set daily step goals and use reminders or notifications to stay on track.

- **Analyze Patterns**: Investigate further into the days with lower activity to understand potential barriers (e.g., weather, work commitments) and offer solutions to maintain activity levels.

Calories vs. Total Steps

```
ggplot(daily_activity, aes(x=TotalSteps, y=Calories)) +
  geom_point(color="darkgreen") +
  geom_smooth(method="lm", color="red") +
  labs(title="Calories Burned vs. Total Steps", x="Total Steps",
y="Calories") +
  theme_minimal()

## `geom_smooth()` using formula = 'y ~ x'
```


Insights:

Positive Correlation: There is a positive correlation between the number of steps taken and the calories burned, as shown by the upward trend in the scatter plot and the linear regression line

Variation: While the overall trend is positive, there is some variation in calories burned for a given number of steps, possibly due to differences in individual metabolism, intensity of activity, or additional exercises not captured by steps alone.

Recommendations:

- **Personalized Goals**: Tailor calorie-burning goals to individual users based on their activity patterns and metabolic rates.
- **Encourage More Activity**: Promote activities that can help burn more calories even with fewer steps, such as high-intensity interval training (HIIT) or strength training.

Steps to merge sleep data with Activity Data in R

Merge Activity and Sleep Data

```
combined_data <- merge(sleep_day, daily_activity, by="Id")
```

Convert SleepDay to Date format
```
sleep_day$SleepDay <- as.Date(sleep_day$SleepDay, format="%m/%d/%Y")
```

Inspect Merged Data:
```
n_distinct(combined_data$Id)

## [1] 24

head(combined_data)

##            Id              SleepDay TotalSleepRecords TotalMinutesAsleep
## 1 1503960366 4/12/2016 12:00:00 AM                 1                327
## 2 1503960366 4/12/2016 12:00:00 AM                 1                327
## 3 1503960366 4/12/2016 12:00:00 AM                 1                327
## 4 1503960366 4/12/2016 12:00:00 AM                 1                327
## 5 1503960366 4/12/2016 12:00:00 AM                 1                327
## 6 1503960366 4/12/2016 12:00:00 AM                 1                327
##   TotalTimeInBed ActivityDate TotalSteps TotalDistance TrackerDistance
## 1            346   2016-04-09      12432          8.10            8.10
## 2            346   2016-04-12        224          0.14            0.14
## 3            346   2016-04-10      10057          6.98            6.98
## 4            346   2016-03-26      17609         11.55           11.55
## 5            346   2016-04-08      12521          7.94            7.94
## 6            346   2016-03-27      12736          8.53            8.53
##   LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 1                        0               2.59                     0.59
## 2                        0               0.00                     0.00
## 3                        0               4.00                     0.49
## 4                        0               6.92                     0.73
## 5                        0               3.31                     0.90
## 6                        0               4.66                     0.16
##   LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
## 1                4.92                       0                32
## 2                0.13                       0                 0
```

```
## 3               2.48                    0                44
## 4               3.91                    0                89
## 5               3.74                    0                46
## 6               3.71                    0                56
##   FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories
## 1                  15                  248              738     1883
## 2                   0                    9               32       50
## 3                  13                  168              737     1755
## 4                  17                  274              588     2154
## 5                  22                  212             1160     1895
## 6                   5                  268              605     1944
```

summary(combined_data)

```
##        Id              SleepDay          TotalSleepRecords
TotalMinutesAsleep
##  Min.   :1.504e+09  Length:5388        Min.   :1.000    Min.   : 58.0
##  1st Qu.:3.977e+09  Class :character   1st Qu.:1.000    1st Qu.:359.0
##  Median :4.445e+09  Mode  :character   Median :1.000    Median :428.0
##  Mean   :4.846e+09                     Mean   :1.113    Mean   :415.7
##  3rd Qu.:6.776e+09                     3rd Qu.:1.000    3rd Qu.:485.0
##  Max.   :8.792e+09                     Max.   :3.000    Max.   :796.0
##  TotalTimeInBed   ActivityDate        TotalSteps     TotalDistance
##  Min.   : 61.0  Min.   :2016-03-12  Min.   :    0  Min.   : 0.000
##  1st Qu.:399.0  1st Qu.:2016-04-02  1st Qu.: 3358  1st Qu.: 2.390
##  Median :459.0  Median :2016-04-05  Median : 7583  Median : 5.550
##  Mean   :453.9  Mean   :2016-04-04  Mean   : 7568  Mean   : 5.295
##  3rd Qu.:522.0  3rd Qu.:2016-04-09  3rd Qu.:11080  3rd Qu.: 7.710
##  Max.   :961.0  Max.   :2016-04-12  Max.   :25701  Max.   :20.140
##  TrackerDistance  LoggedActivitiesDistance VeryActiveDistance
##  Min.   : 0.000  Min.   :0.0000           Min.   : 0.000
##  1st Qu.: 2.390  1st Qu.:0.0000           1st Qu.: 0.000
##  Median : 5.550  Median :0.0000           Median : 0.290
##  Mean   : 5.271  Mean   :0.1834           Mean   : 1.206
##  3rd Qu.: 7.710  3rd Qu.:0.0000           3rd Qu.: 2.140
##  Max.   :20.140  Max.   :5.4569           Max.   :16.820
##  ModeratelyActiveDistance LightActiveDistance SedentaryActiveDistance
##  Min.   :0.0000           Min.   : 0.000      Min.   :0.000000
##  1st Qu.:0.0000           1st Qu.: 1.960      1st Qu.:0.000000
##  Median :0.3600           Median : 3.500      Median :0.000000
##  Mean   :0.6588           Mean   : 3.353      Mean   :0.001218
##  3rd Qu.:0.9200           3rd Qu.: 4.690      3rd Qu.:0.000000
##  Max.   :6.4000           Max.   :12.510      Max.   :0.100000
##  VeryActiveMinutes FairlyActiveMinutes LightlyActiveMinutes
SedentaryMinutes
##  Min.   : 0.00   Min.   : 0.0    Min.   : 0.0    Min.   : 32.0
##  1st Qu.: 0.00   1st Qu.: 0.0    1st Qu.:139.0   1st Qu.: 654.0
##  Median : 5.00   Median : 10.0   Median :205.0   Median : 738.0
##  Mean   : 20.83  Mean   : 14.9   Mean   :193.2   Mean   : 792.6
##  3rd Qu.: 33.00  3rd Qu.: 24.0   3rd Qu.:257.0   3rd Qu.: 869.0
```

```
##  Max.   :202.00     Max.   :141.0         Max.   :586.0         Max.   :1440.0
##    Calories
##  Min.   :   0
##  1st Qu.:1819
##  Median :2133
##  Mean   :2212
##  3rd Qu.:2624
##  Max.   :4430
```

Plot Total Sleep Time vs. SleepDay

```
ggplot(combined_data, aes(x = ActivityDate, y = TotalMinutesAsleep)) +
  geom_point(color = "purple") +
  labs(title = "Total Sleep Time per Day", x = "AcitivityDate", y = "Total
Minutes Asleep")+
  theme_minimal()
```


Total Sleep Time per Day
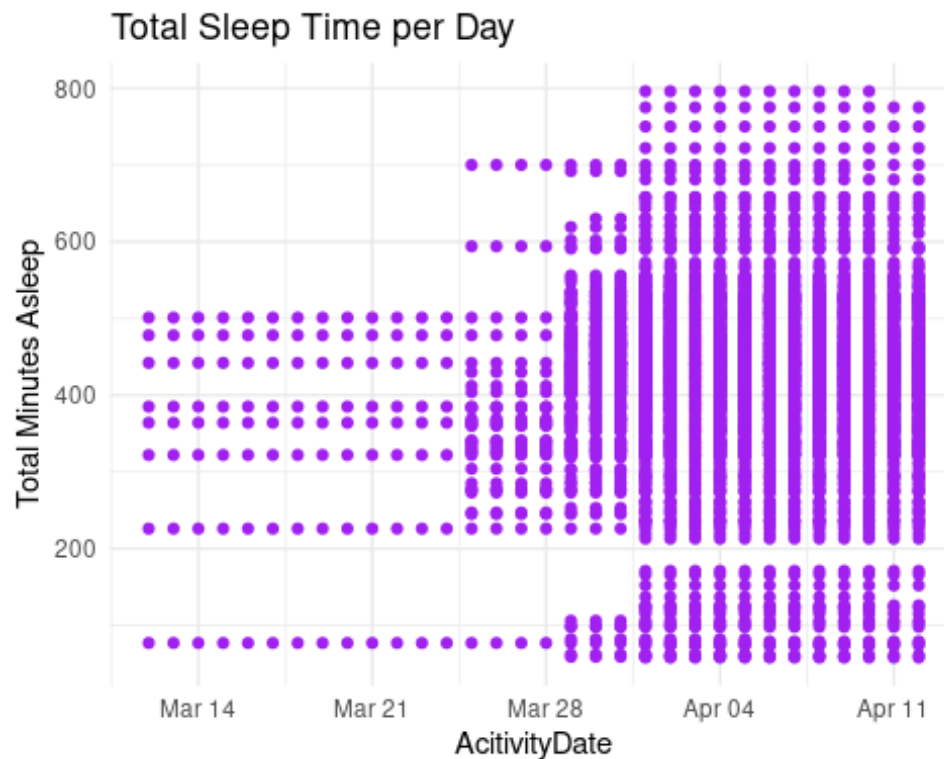
Recommendations

Comprehensive Health Monitoring:

- **Integration of Activity and Sleep Data:** Ensure seamless integration of activity and sleep data within the Bellabeat app to provide holistic health insights.

- **Personalized Recommendations:** Utilize machine learning to offer personalized recommendations based on individual user data, promoting better sleep and activity habits.

- **User Engagement:**
  Interactive Dashboards: Create interactive dashboards that allow users to visualize their sleep and activity data over time, helping them identify patterns and make improvements.

- **Gamification:** Incorporate gamification elements like badges, challenges, and leaderboards to motivate users to improve their sleep and activity levels.

Plot very active minutes vs. sedentary minutes

Very Active Minutes
```
ggplot(daily_activity, aes(x = VeryActiveMinutes, y = SedentaryMinutes))+
geom_point(color = "violet")+
labs(title = "Correlation of very active and Sedentary", x =
"VeryActiveMinutes",
     y = "SedentaryMinutes")
```

Correlation of very active and Sedentary
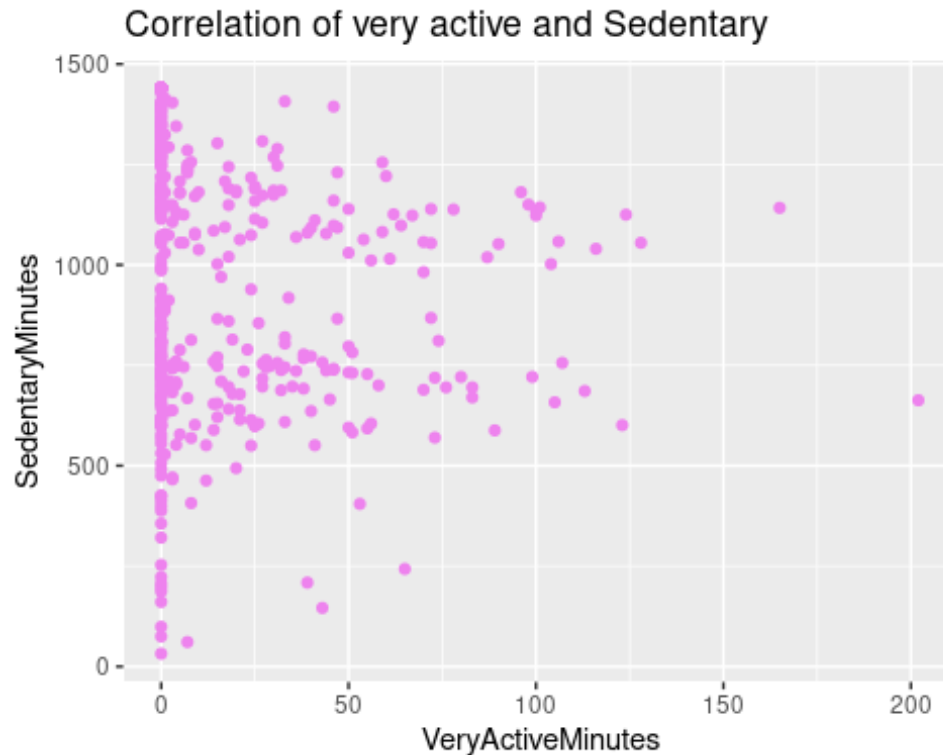
Insights

Activity Levels :

The graph shows how often users engage in very active minutes. Most users have a relatively low number of very active minutes per day, with fewer instances of very active minutes

Recommendations:

- **Increase High-Intensity Activities:** Encourage users to incorporate more high-intensity activities into their routines, such as running, cycling, or aerobic exercises.
- **Set Achievable Milestones:** Set incremental goals for increasing very active minutes to avoid overwhelming users and to gradually improve their fitness levels.
- **Group Challenges:** Introduce group challenges or social features to motivate users to increase their very active minutes through friendly competition.
- **Regular Feedback:** Provide regular feedback and progress reports to users to keep them motivated and informed about their activity patterns.
- **Gamification:** Introduce gamification elements such as badges, rewards, and leaderboards to make physical activity more engaging and fun.
- **Personalized Insights:** Use machine learning algorithms to provide personalized Insights and recommendations based on individual user data, helping them achieve their fitness goals more effectively.

- **Holistic Approach:** Promote a holistic approach to wellness by integrating other health metrics such as sleep, nutrition, and mental well-being into the platform.
- **Community Engagement:** Foster a sense of community among users by enabling them to share their achievements, participate in group activities, and support each other's fitness journeys.

High-Level Content Recommendations for Marketing strategy

1. Targeted Campaigns for Fitness Enthusiasts:

- **Highlight Advanced Fitness Tracking Features:** Emphasize Bellabeat's advanced fitness tracking capabilities that help users achieve their fitness goals, including step count, active minutes, and distance covered.
- **Utilize Success Stories and Testimonials:** Share inspiring success stories and testimonials from users who have improved their fitness using Bellabeat devices.

2. Sleep Quality Campaigns:

- **Emphasize the Importance of Sleep Quality:** Promote the benefits of good sleep and how Bellabeat devices can monitor and enhance sleep patterns.
- **Create Educational Content:** Develop blogs and videos focusing on sleep health, tips for better sleep, and the role of Bellabeat devices in achieving quality sleep.

3. Health Analytics Promotion:

- **Showcase Detailed Health Analytics:** Highlight the detailed heart rate and stress monitoring capabilities of Bellabeat devices, showcasing how these features can provide valuable health insights.
- **Use Data-Driven Marketing:** Incorporate data-driven insights and visualizations in marketing materials to clearly demonstrate the benefits of using Bellabeat devices.

4. Goal Achievement Stories:

- **Share User Stories:** Highlight user stories and testimonials about achieving health and fitness goals with the help of Bellabeat devices.
- **Organize Community Events:** Create challenges and events that encourage users to set and share their progress towards their health and fitness goals.

5. Integration Benefits:

- **Promote Seamless Integration:** Emphasize the integration of Bellabeat devices with popular health and fitness apps, highlighting the convenience of having all wellness data in one place.

- **Showcase Added Value:** Demonstrate the added value of integration, such as more comprehensive health tracking and easier data management.

6. Smart Reminders Feature:

- **Emphasize Convenience and Benefits:** Promote the smart reminders feature that helps users stay on top of their health by reminding them to move, hydrate, and manage stress.
- **Create Educational Content:** Develop content explaining the importance of these reminders and how they can contribute to maintaining a healthy routine.

These recommendations are designed to leverage the insights gained from the analysis of smart device usage data and align with Bellabeat's goals of becoming a larger player in the global smart device market.