

Google Case Study 2: Bellabeat Smart Device Usage

Introduction

This case study details a data analysis project undertaken for the Google Data Analytics Professional Certificate course, specifically the Capstone Project, focusing on BellaBeat's smart devices usage.

The project centers on Bellabeat, a company specializing in health-focused products for women. Despite its current success as a small firm, Bellabeat aims to expand its presence in the global smart device market. The task for the marketing analyst team is to analyze data from smart devices to understand consumer usage patterns. These insights will inform strategic marketing decisions aimed at propelling the company's growth. The findings and key recommendations will be presented to Bellabeat's executive leadership.

Products

The Bellabeat app offers health data related to activity, sleep, stress, menstrual cycle, and mindfulness habits, helping users understand their habits and make healthier choices. It connects to Bellabeat's line of smart wellness products:

1. Leaf: A versatile wellness tracker worn as a bracelet, necklace, or clip, it monitors activity, sleep, and stress, linking with the Bellabeat app.

2. Time: A wellness watch blending classic design with smart technology to track activity, sleep, and stress, also connecting to the Bellabeat app.
3. Spring: A smart water bottle tracking daily hydration levels, ensuring adequate hydration throughout the day through the Bellabeat app.
4. Bellabeat Membership: A subscription service offering personalized guidance on nutrition, activity, sleep, health, and mindfulness tailored to users' lifestyles and goals.

This case study specifically focuses on the TIME product, examining how users utilize this smart device for monitoring their activity and sleep patterns.

Tools used in Analysis

- Data Cleaning: Microsoft Excel
- Data Analysis: Microsoft SQL Server
- Visualization: Tableau Public

Data Analysis Process

To effectively analyze Bellabeat data and address key business questions while making informed recommendations, the data analysis process will adhere to the following key steps:

- 1) Ask relevant questions
- 2) Prepare the data for analysis
- 3) Process the data to ensure accuracy and relevance
- 4) Analyze the data to derive meaningful insights
- 5) Share findings through clear and actionable reports or presentations
- 6) Act upon the insights to drive strategic decisions and improvements.
- 7) These steps ensure a structured approach to deriving insights from data for informed business decisions at Bellabeat.

Ask Phase:

To tackle the business objective of analyzing how users utilize the Time smart device, we'll focus on several key findings to guide our analysis:

1. Proportion of Calories per Distance by Day: Examining how calorie expenditure relates to the distance covered each day.
2. Average Steps per Hour: Analyzing the average number of steps taken per hour throughout the day to understand activity patterns.

3. Total Steps & Total Distance by ID: Aggregating total steps and distance covered by individual users (ID) to assess overall activity levels.

4. Maximum Sleep Duration (in hours) by Day: Identifying the longest sleep duration recorded each day to understand users' sleep habits.

5. Correlation of Steps with Active Minutes: Investigating how closely steps and active minutes (of physical activity) correlate, indicating engagement levels.

6. Correlation of Sleep with Active Minutes by Day: Examining the relationship between sleep duration and active minutes each day to understand the impact of activity on sleep patterns.

7. Connection of total minutes in bed and total minutes asleep: Aggregating the relationship between the minutes spent in bed and the effective minutes spent asleep.

8. Total Steps vs Calories Burned: Visualizing the correlation between the number of steps walked and the amount of calories burned

These findings will provide insights into how users interact with the Time smart device, informing strategic decisions and recommendations for Bellabeat.

Prepare Phase:

The data for this analysis is sourced from a public dataset available on Kaggle, which includes 18 files containing detailed information on daily activity, sleep, weight, calories, and intensities. This dataset is structured in a mix of long and wide formats, with each file focusing on different aspects of smart device usage. It's important to note that this data is not from Bellabeat itself, potentially introducing biases or credibility issues. Additionally, the dataset is not current, covering only April and May of 2016, which may limit the reflection of current trends in smart device usage.

Moreover, the dataset lacks information on Bellabeat's Spring product and the Bellabeat app, restricting the analysis primarily to insights related to the Time smart device. These considerations are crucial for interpreting the findings and making recommendations based on the available data.

Cleaning in Excel:

Based on the observations and updates to clean the data effectively:

Removal of Unwanted Tables: Tables like 'daily_steps', 'daily_calories', and 'daily_intensities' were part of the 'daily_activity' file, and were removed since they were validated as redundant using the MATCH function.

Rename File: Proper naming conventions were applied to maintain data integrity across the dataset.

Check Data Types: Relevant columns such as TotalSteps, TotalDistance, VeryActiveMinutes, LittleActiveMinutes, FairlyActiveMinutes, and SedentaryMinutes were converted to numeric values where necessary to ensure consistency and accuracy in analysis.

Check for Duplicates: Duplicate values were identified across two files and removed using the 'Remove Duplicates' function to avoid skewing analysis results.

Check for Blanks: The dataset was thoroughly checked for incomplete or blank values in all columns to ensure completeness and reliability of the data.

Validate Column Values: All column values in their respective files were validated using filters to ensure data quality and accuracy.

Sorting the Table: Tables were sorted in ascending order based on the 'Date' column to facilitate chronological analysis and insights.

These steps ensure that the dataset is cleaned and prepared for accurate analysis, minimizing potential errors and ensuring robust findings to guide strategic decisions.

Process Phase:

During this phase, a thorough examination of the key findings related to how users utilize Bellabeat smart devices was conducted using Microsoft SQL Server.

1. Proportion of Calories per Distance by Day:

SQL Query: [Bellabeat Case Study Complete
Code](<https://github.com/JJRockzzzz/Google-Data-Analytics-Capstone-Project/blob/main/Bellabeat%20Case%20Study%20SQL.sql>)

2. Average Steps per Hour

SQL Query: [Bellabeat Case Study Complete
Code](<https://github.com/JJRockzzzz/Google-Data-Analytics-Capstone-Project/blob/main/Bellabeat%20Case%20Study%20SQL.sql>)

3. Total Steps & Total Distance by ID

SQL Query: [Bellabeat Case Study Complete
Code](<https://github.com/JJRockzzzz/Google-Data-Analytics-Capstone-Project/blob/main/Bellabeat%20Case%20Study%20SQL.sql>)

4. Maximum Sleep Duration (in hours) by Day

SQL Query: [Bellabeat Case Study Complete
Code](<https://github.com/JJRockzzzz/Google-Data-Analytics-Capstone-Project/blob/main/Bellabeat%20Case%20Study%20SQL.sql>)

5. Correlation of Steps with Sleeping Minutes

SQL Query: [Bellabeat Case Study Complete

Code](<https://github.com/JJRockzzzz/Google-Data-Analytics-Capstone-Project/blob/main/Bellabeat%20Case%20Study%20SQL.sql>)

Result:

```
> cor(combined_activity_and_sleep$TotalMinutesAsleep,  
+      combined_activity_and_sleep$TotalTimeInBed, method = "pearson")  
[1] 0.9304224
```


6. Correlation of Sleep with Sleep and Sedentary Minutes by Day

SQL Query: [Bellabeat Case Study Complete

Code](<https://github.com/JJRockzzzz/Google-Data-Analytics-Capstone-Project/blob/main/Bellabeat%20Case%20Study%20SQL.sql>)

Result:

```
> cor(combined_activity_and_sleep$TotalMinutesAsleep,  
+      combined_activity_and_sleep$SedentaryMinutes, method = "pearson")  
[1] -0.6010731
```

![[image]][sleep active minutes per day](https://github.com/JJRockzzzz/Google-Data-Analytics-Capstone-Project/blob/main/corr_sleep_sedentary.png)

Analyze Phase:

After processing the valuable insights, we will now explore explanatory visualizations that illustrate users' healthy daily activities and correlate key parameters using Tableau

Public(https://public.tableau.com/views/BellaBeatSmartDeviceUsageDashboard/Dashboard2?:language=en-US&:display_count=n&:origin=viz_share_link). The engaging visualizations are provided below:

Visualization 1: Proportion of Calories per Distance by Day

Visualization 2: Average Steps per Hour

Visualization 3: Total Steps & Total Distance by ID

Visualization 4: Maximum Sleep Duration (in hours) by Day

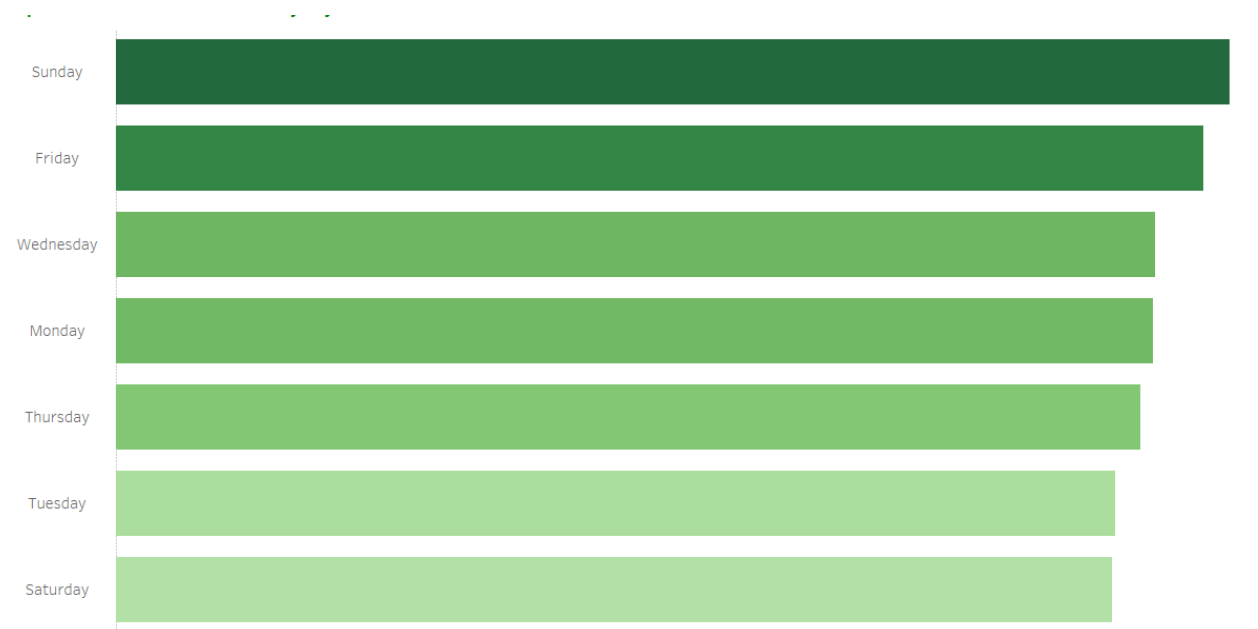
Visualization 5: Correlation of Steps with Active Minutes

Visualization 6: Correlation of Sleep with Active Minutes by Day

These visualizations help in understanding the usage patterns of the Bellabeat Time smart device and provide insights into users' daily activities and health metrics.

1. Proportion of Calories per Distance by Day:

The analysis of the proportion of calories per distance by day reveals interesting insights into users activity level. Specifically, it is observed that Sundays exhibit a relatively high proportion of calories burned per distance compared to other days of the week.

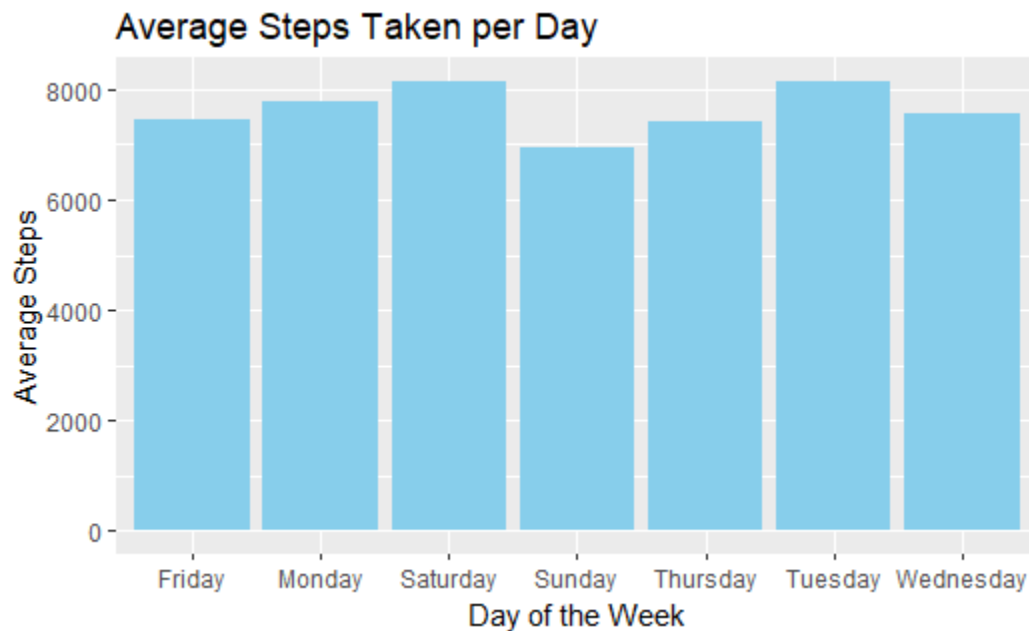


Visualization 1

![[image]](https://github.com/JJRockzzzz/Google-Data-Analytics-Capstone-Project/blob/main/calories_distance.png)

2. Average Steps per Day

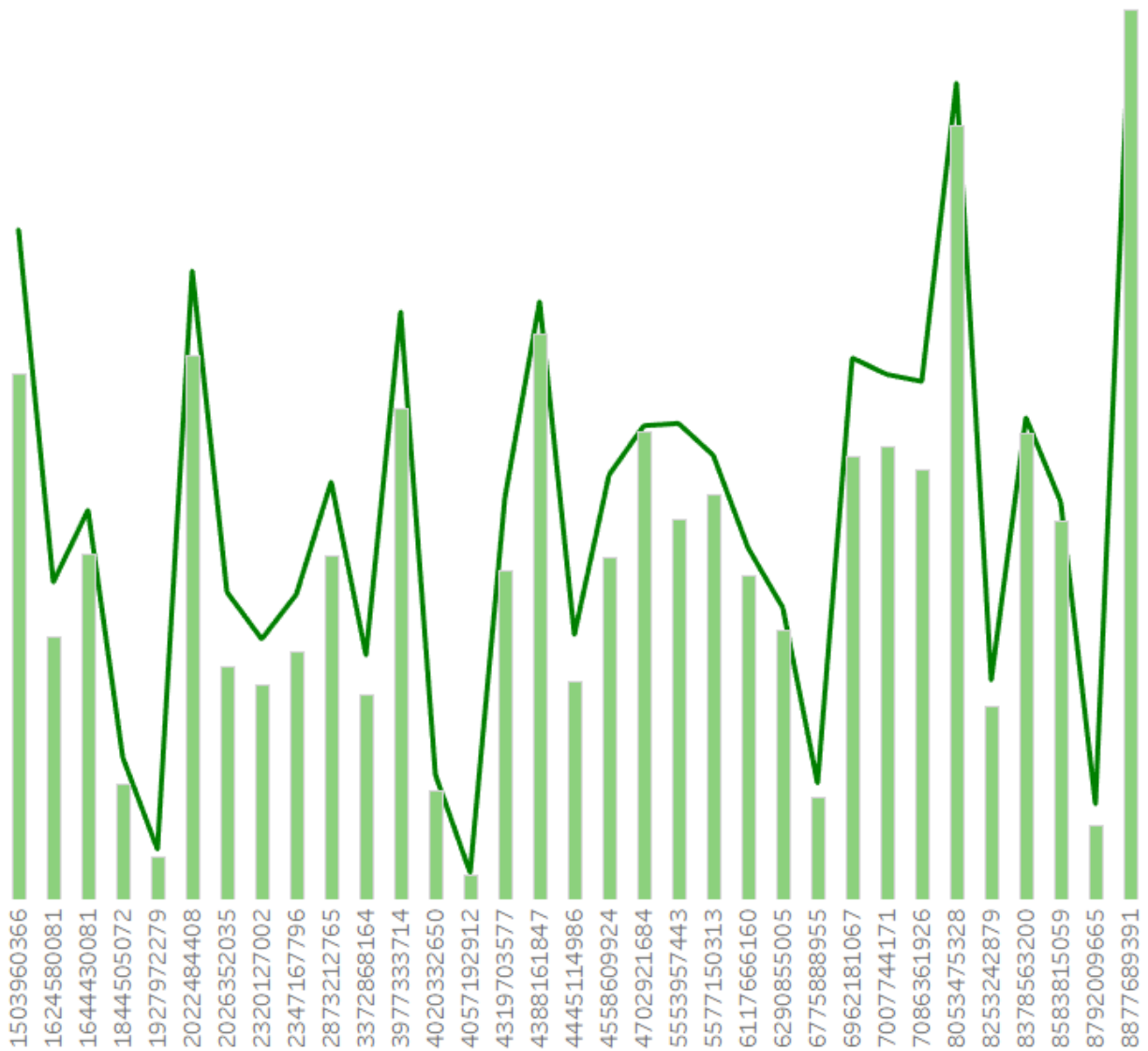
The visualization showcasing the average steps taken per day reveals interesting patterns in users activity throughout the day. It is evident that the highest average number of steps is recorded during which day. This finding suggests that users are more active and tend to engage in higher levels of physical activity during this day.



Visualization 2

3. Total Steps & Total Distance by ID

This key finding highlights the relationship between total steps and total distance covered by individual users. The analysis reveals that user ID '8877689391' recorded the highest number of steps, totaling 497,241, and covered a distance of 409.4 units. This indicates that this particular user has been highly active, taking a significant number of steps and covering a considerable distance.

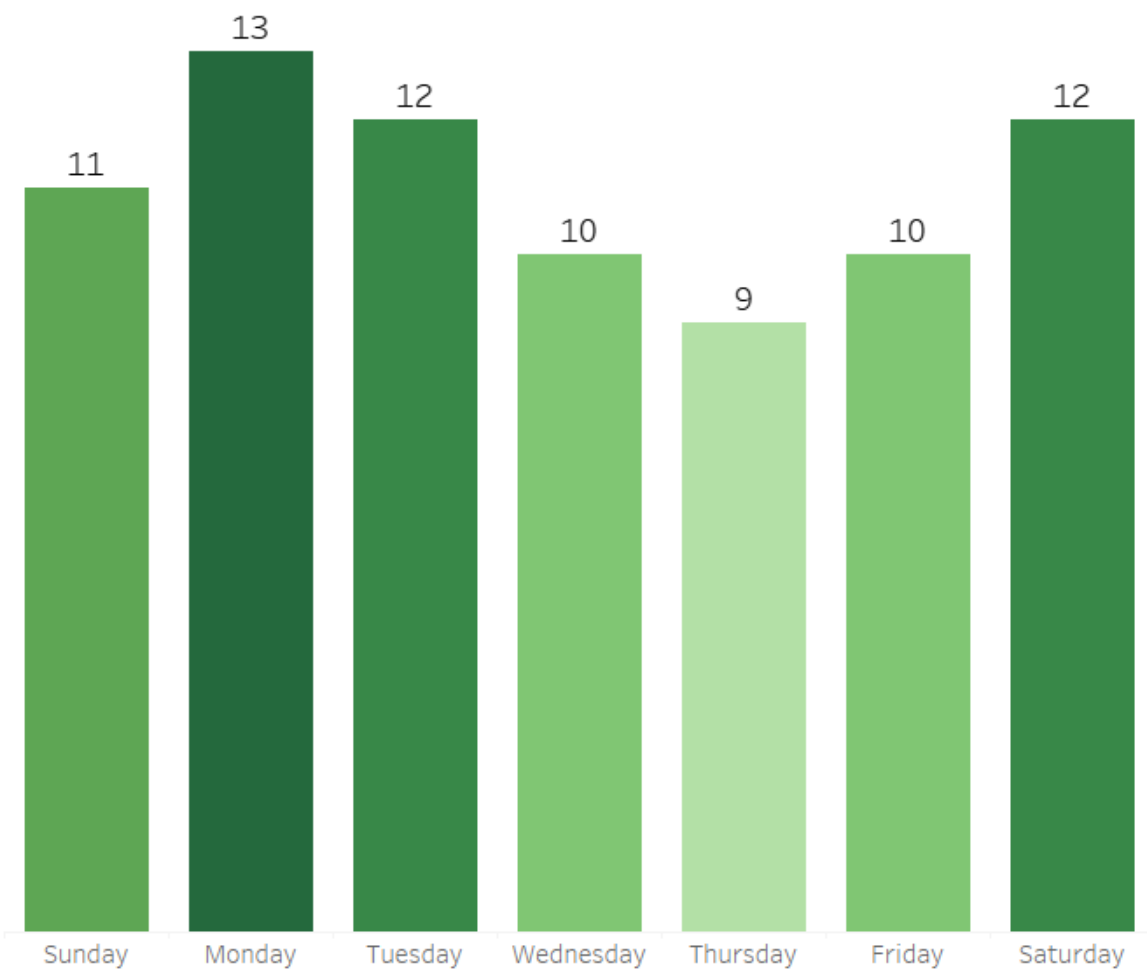


Visualization 3

![[image]](https://github.com/JJRockzzzz/Google-Data-Analytics-Capstone-Project/blob/main/distance_id.png)

4. Maximum Sleep Duration (in hours) by Day

The sleep duration was categorized in days, and it is identified that maximum sleep duration is found to be 13 'hours' on 'Monday'. This finding suggests that users tend to have longer sleep periods on Mondays compared to other days of the week.



Visualization 4

5. Correlation of Steps with Active Minutes (Very Active and Sedentary Minutes)

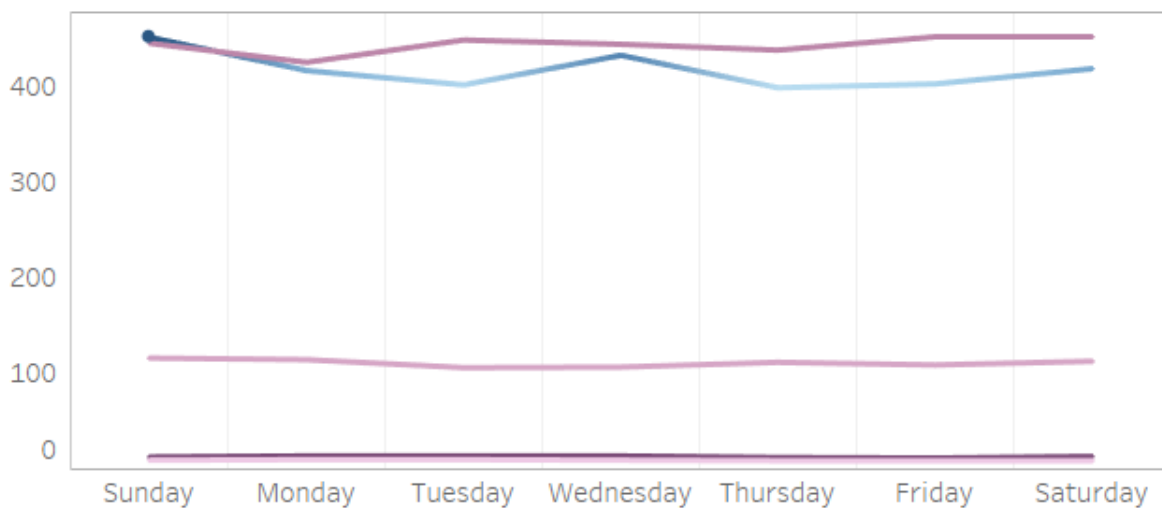
The analysis reveals a clear correlation between total steps, very active minutes, and sedentary minutes. As the total steps increase, the very active minutes also tend to increase, while the sedentary minutes decrease. This suggests that individuals who take more steps engage in more intense physical activity and spend less time in a sedentary state. These findings align with the recommendations of the Centers for Disease Control and Prevention (CDC), which suggests engaging in at least 30 minutes of moderate physical activity every day for optimal health. However, the data indicates that the majority of users in the dataset spend a considerable amount of time in a sedentary state, with relatively low levels of very active minutes.

```
> cor(daily_activity$TotalSteps, daily_activity$Calories, method = "pearson")  
[1] 0.5915681
```

Visualization 5

6. Correlation of Sleep with Each Active Minutes by day

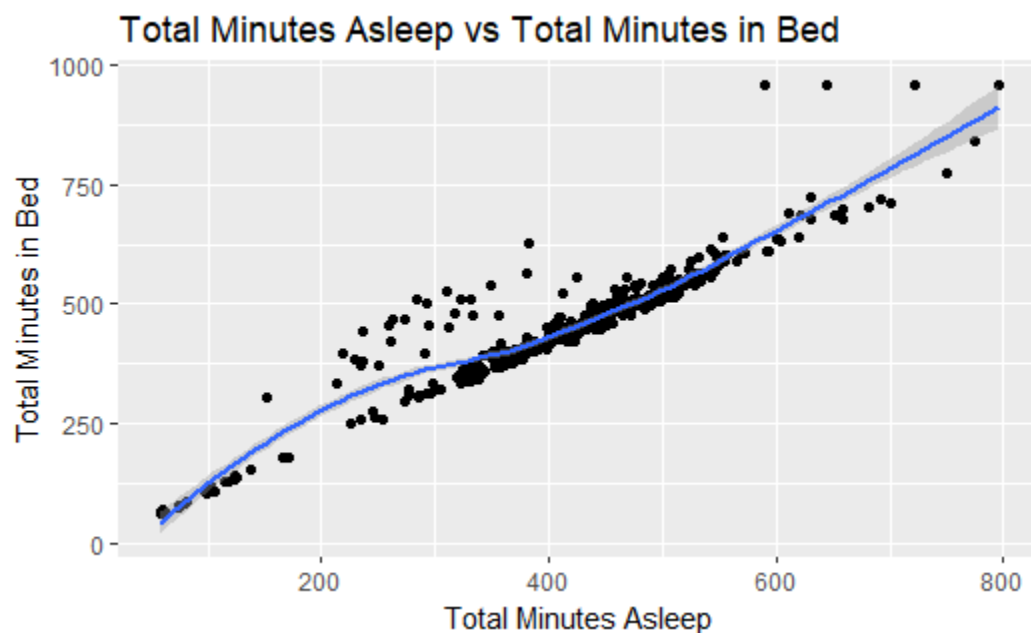
The analysis of sleep and activity minutes over the week reveals that, on average, users spent approximately 6.98 hours sleeping, 13.31 hours in a sedentary state, 3.33 hours in light activity, 0.27 hours in fairly active moments, and 0.40 hours in very active periods. This highlights that users tend to spend a significant amount of time in sedentary behavior during the weekdays. Additionally, the data indicates that users allocate maximum time for sleep and have minimal engagement in activities.



Visualization 6

7. Connection of total minutes in bed and total minutes asleep

The analysis of sleep versus minutes in bed over the week reveals that, on average, users spent approximately 6.98 hours actually sleeping while recording around 7.50 hours in bed. This indicates that users spend about 0.52 hours awake while in bed, either before falling asleep or after waking up. This correlation highlights that while users allocate a substantial amount of time to rest, a notable portion of this time is not spent in actual sleep. This data suggests opportunities for users to improve their sleep efficiency by minimizing the time spent awake in bed, thereby potentially enhancing overall sleep quality and well-being.



Visualization 7

8. Total Steps vs Calories Burned

The analysis of total steps versus calories burned over the week reveals that, on average, users took approximately 7,500 steps daily, resulting in an average calorie burn of around 2,300 calories per day. This correlation highlights that users who engage in more physical activity, as measured by their step count, tend to burn more calories. This data underscores the importance of maintaining an active lifestyle to achieve higher calorie expenditure, which is crucial for overall health and weight management. By increasing their daily step count, users can effectively enhance their calorie burn and contribute to their fitness goals.



Visualization 8

Act Phase

Key Takeaways

Based on the findings of this analysis, several key takeaways can be derived:

1. Sunday Activity: Users tend to prioritize their health and engage in physical activities on Sundays. This suggests that people might deliberately choose one

day of the weekend for active pursuits, while the other day is spent in more sedentary activities.

2. Evening Activity Peak: There is a peak in average steps per hour between 5 and 6 PM, indicating that individuals are most active in the evening.

3. Sedentary Behavior: Despite the positive correlation between steps and active minutes, most users in the dataset still spend a significant amount of time in a sedentary state, underscoring the need to promote more physical activity and reduce sedentary behavior.

4. Sleep vs. Activity: The analysis shows that users are meeting the minimum sleep standard of 7 hours per night. However, even with sufficient sleep, their active time remains relatively low, with a substantial portion of the day spent sedentary.

Recommendations

Based on the key findings of Bellabeat smart devices usage, here are some recommendations for the business:

1. Personalized Activity Reminders: Develop a feature within Bellabeat smart devices that provides personalized activity reminders to users. These reminders, based on individual activity patterns, can encourage more active behaviors throughout the day.

2. Sleep Optimization Guidance: Offer sleep optimization guidance through Bellabeat smart devices. This could include recommendations on sleep duration and quality, along with tips for improving sleep habits. The goal is to help users establish healthy sleep routines and maximize sleep benefits.

3. Active Time Tracking: Enhance the activity tracking capabilities of Bellabeat smart devices by providing users with detailed insights into their active time distribution throughout the day. This can help users understand their activity patterns and motivate them to increase active minutes by incorporating activities like cycling, yoga, and walking.

4. Activity Challenges and Rewards: Implement activity challenges and rewards within the Bellabeat app. This can create a sense of competition and motivation among users to achieve their activity goals. Rewards can include virtual badges, discounts on Bellabeat products, or access to exclusive content.

By implementing these recommendations, Bellabeat can enhance the user experience, increase engagement with smart devices, and further establish itself as a leading provider of health and wellness solutions.

Visualization

```
> daily_activity %>%
+   select(VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes,
+           SedentaryMinutes) %>%
+   summary()
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
> daily_steps %>% select(StepTotal) %>% summary()
StepTotal
Min.      : 0
1st Qu.: 3790
Median : 7406
Mean    : 7638
3rd Qu.:10727
Max.    :36019
> cleaned_daily_sleep %>% select(TotalMinutesAsleep) %>% summary()
TotalMinutesAsleep
Min.      : 58.0
1st Qu.:361.0
Median :432.5
Mean    :419.2
3rd Qu.:490.0
Max.    :796.0
```

