Case Study Outline Table

Title	How Can a Wellness Technology Company Play It Smart?
Industry focus	Health-focused smart products.
Problem statement	Utilizing data analytics concepts to analyze smart device data to gain insight into how consumers are using their smart devices.
Business use case (What are you solving?)	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's marketing strategy?
Goals/metrics	Find the usage patterns for smart devices in order to gain insight into how consumers use non-Bellabeat smart devices and identify how the information can be used to improve Bellabeat products.
Deliverables	Insights to guide the marketing strategy for the company. A report and presentation outlining the insights.
Are datasets available?	Yes.
Dataset list	The dataset can be downloaded from the link given in the section below.
The website to scrape the data needed	https://www.kaggle.com/datasets/arashnic/fitbit?resource=download

Data Analytics Phases

Ask

Key Characters

○ Urška Sršen: Bellabeat's co-founder and Chief Creative Officer

○ Sando Mur: Mathematician and Bellabeat's cofounder; a key member of the Bellabeat executive team

○ Bellabeat marketing analytics team: A team of data analysts responsible for collecting, analyzing, and reporting data that helps guide Bellabeat's marketing strategy.

- What is the problem you are trying to solve?
- How can your insights drive business decisions?

The objective is to analyze consumer behavior patterns in using smart devices. The expected outcomes should help:

- Identify key trends in how consumers interact with smart devices.
- Uncover potential areas for improvement within the business strategy based on these consumer insights.

- Pinpoint specific opportunities for the marketing department to leverage consumer behavior for greater engagement and positive outcomes.

Goals

These will guide my 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's marketing strategy?

Prepare

Data Acquisition and Characteristics:

Source: The data was obtained from Kaggle, a reputable online data repository known for its diverse datasets and frequent updates. The most recent update for this specific dataset occurred on March 2nd, 2024. However, the data was gathered in 2016, making it difficult to analyze recent trends.

Format: The data is stored in CSV format, a widely used and accessible format for data analysis.

Structure: The data is organized in a long format, suitable for time series analysis or exploring individual user behavior.

Scope: While relevant to the research topic, the data is limited to Fitbit users and does not encompass users of other smart devices. This limitation is considered during analysis and interpretation. Also, the data provided does not specify the demographics of the people in question, making it difficult to tailor responses/analyses relevant to them. For instance, gender, age range, and occupation are among a few factors that could affect a person's lifestyle. The data period given for analysis was too restricted and could not allow an opportunity for the analysis to be more detailed, analyzing and reviewing user trends for a period longer than 2 days.

Data Provenance:

The dataset was contributed by Mobius, with Arash Nicoomanesh, a data scientist and researcher, listed as the author.

The cited sources link to researchers affiliated with a non-profit research institute, suggesting a credible and potentially unbiased data collection process.

Process

Data was uploaded into R and used R functions to review and clean it. I used R functions to create visuals that clearly represent the cleaned data. The daily activities dataset shows the users' information in a single day. It outlines the information that FitBit tracks in a day.

Cleaning

The process begins with installing and loading packages in RStudio, followed by installing and loading the lubridate package in the tidyverse.

```
install.packages("tidyverse")
install.packages("here")
install.packages("janitor")
```

```r
install.packages("lubridate")
install.packages("skimr")

library(tidyverse)
library(here)
library(janitor)
library(lubridate)
library(skimr)
```

The next step involves

Importing data

The data is from Mobius on Kaggle FitBit Fitness Tracker Data.

the `**read_csv()**` function is used to import data from a **.csv** in the project folder called *"dailyActivity_merged.csv"* and saved as a data frame called `*dailyactivity_df*`:

```r
> read.csv("dailyActivity_merged.csv")
read.csv("dailyActivity_merged.csv")
read.csv("hourlyCalories_merged.csv")
read.csv("hourlyIntensities_merged.csv")
read.csv("hourlySteps_merged.csv")
read.csv("minuteSleep_merged.csv")
read.csv("weight_merged.csv")

> dailyactivity_df <- read.csv("dailyActivity_merged.csv")
hourlyCalories_df <- read.csv("hourlyCalories_merged.csv")
hourlyIntensities_df <- read.csv("hourlyIntensities_merged.csv")
hourlySteps_df <- read.csv("hourlySteps_merged.csv")
minuteSleep_df <- read.csv("minuteSleep_merged.csv")
minuteSleep_df <- read.csv("weight_merged.csv")
```

Inspect and clean data

The `**glimpse()**` function is used to preview the data to get a better understanding of the data by getting a summary view of the dataset.

```r
> View('hourlyCalories_df')
View('hourlyIntensities_df')
View('hourlySteps_df')
View('minuteSleep_df')
View('weight_df')

> glimpse(dailyactivity_df)
glimpse(hourlyCalories_df)
glimpse(hourlyIntensities_df)
```

glimpse(hourlySteps_df)
glimpse(minuteSleep_df)
glimpse('weight_df')

Some of the datasets contained the same columns during the review process. To make it easy to access the information, the **'join()'** function was used to unite the different columns into one dataset. The **'inner_join()'** function returns the entries common in the tables involved.
> inner_join(hourlyCalories_merged, hourlyIntensities_merged, hourlySteps_merged, by = c ("Id", "ActivityHour"))

Checking the number of distinct participants
The **n_distinct()** function is used to count the exact number.
n_distinct('daily_activity$Id')
n_distinct('hourly_activity$Id')
n_distinct('daily_sleep$Id')
n_distinct('weight$Idf')

Checking duplicates
sum(duplicated(daily_activity))
sum(duplicated(daily_sleep))
sum(duplicated(hourly_activity))

Analysis of Dataset Participant Numbers and Implications
Following the initial data exploration, a closer examination of the number of unique participants represented within each dataset was conducted. This analysis is crucial to assess the reliability and generalizability of any subsequent findings. Here is a breakdown of the findings and their implications:

"weight" Dataset:

This dataset contained data from only 8 unique participants. This sample size is considered too small to draw statistically significant conclusions or conduct meaningful analyses. This limitation is taken into consideration as the data is reviewed.

"daily_activity" and "hourly_activity" Datasets:

Both the "daily_activity" and "hourly_activity" datasets offered a more promising sample size, with 33 unique participants each. While not ideal for large-scale population studies, this sample size allows for preliminary explorations and potential trend identification. However, it's important to acknowledge the limitations of a relatively small sample and interpret any results with caution.

"daily_sleep" Dataset:

The "daily_sleep" dataset presented the smallest participant pool, with only 24 unique participants. Similar to the "daily_activity" and "hourly_activity" datasets, this sample size facilitates exploratory analysis but necessitates careful interpretation due to potential limitations in generalizability.

Decision on Further Analysis:

Despite the relatively small sample sizes in the remaining datasets ("daily_activity," "hourly_activity," and "daily_sleep"), analysis was conducted, acknowledging the limitations. Here are some potential approaches:
Focus on Descriptive Statistics: Descriptive statistics can be utilized to summarize the data, providing insights into central tendencies and variability within the sample.
Exploratory Data Analysis: Employing exploratory data analysis techniques to identify potential trends and relationships within the data. However, any conclusions drawn will be specific to this limited sample and may not be representative of the broader population.
Highlighting Limitations: If we decide to present any findings based on these datasets, it's crucial to clearly outline the limitations due to the small sample size. This transparency will ensure stakeholders understand the potential for bias and the need for further research with a larger participant pool.

Analysis/Share
For the analysis, I created visualizations using Google Sheets after exporting select cleaned-up and reviewed data. I utilized average durations of the reviewed data to create the visuals.

Below is a clear presentation of the averages used to draw conclusions regarding the data in question. The table chart represents the average weight(kg), BMI, sleep(hours), steps, and distance.

Metric	Value
Weight (kg)	72.04
BMI	25.19
Sleep (hours)	6.9
Steps	7638
Distance (km)	5.49

The analysis delves into the characteristics of the typical user within this dataset. The data shows the general weight range for most users and what that could mean for their health.

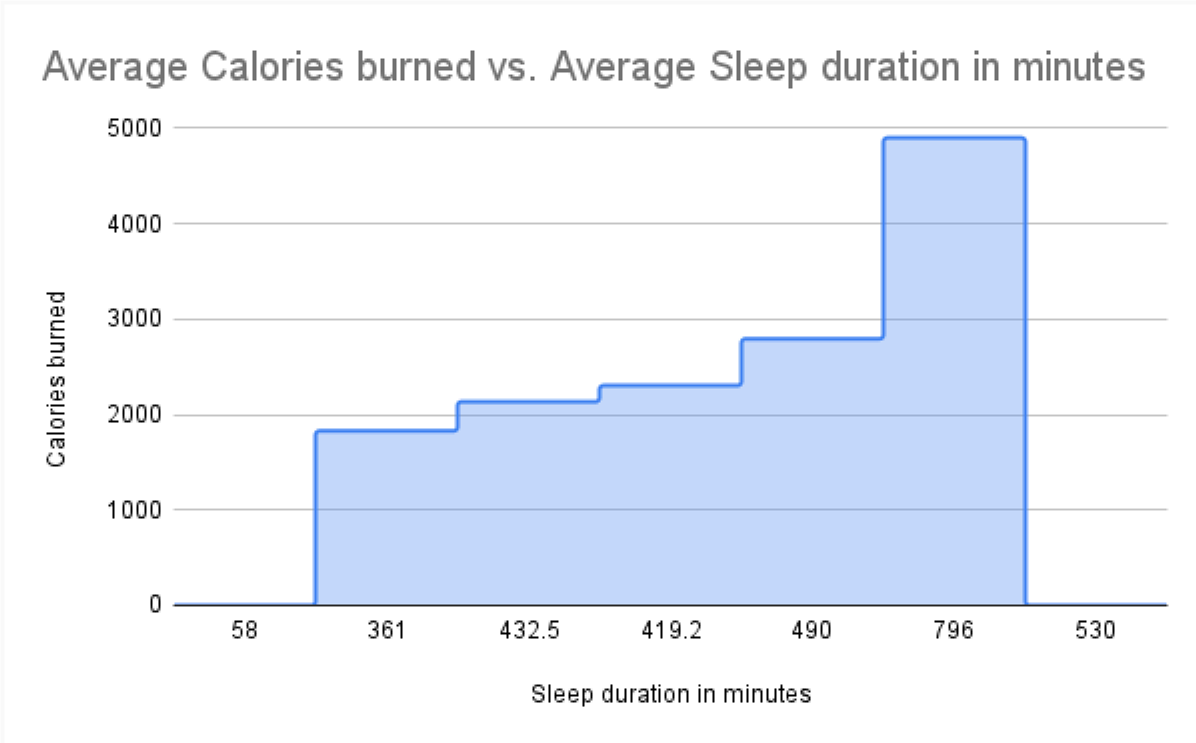
Weight(Body Composition): The average user carries a weight of approximately 72.04 kilograms. To delve deeper, we can calculate the Body Mass Index (BMI) based on this weight, which comes out to be 25.19. This value falls within the "overweight" category according to the World Health Organization (WHO) classification.

Activity Level: When it comes to physical activity, the average user primarily engages in light activities. While the specific breakdown of activity types is not explicitly mentioned, this suggests a tendency towards lower-intensity movements throughout the day.

Sleep Patterns: On average, users dedicate roughly 6.9 hours to sleep each night. This falls slightly below the National Sleep Foundation's recommended range of 7-8 hours of sleep for adults.

Daily Movement: The average user takes approximately 7638 steps per day. This translates roughly to 5.49 kilometers of daily travel distance. While these figures provide some insight into activity levels, it is important to consider factors like stride length and terrain variations, which can influence the actual distance covered.

A correlation considered is the relationship between the total sleep in minutes versus the calories burned. The aim was to identify if a user slept more hours, it assisted with calorie burning.



Sleep and Calorie Expenditure

The analysis reveals a notable trend—individuals who slept between 6 and 7 hours exhibited calorie burning, with only a few outliers deviating from this pattern. This initial observation sparks curiosity: Could there be a connection between sleep duration in this specific range and increased calorie expenditure?

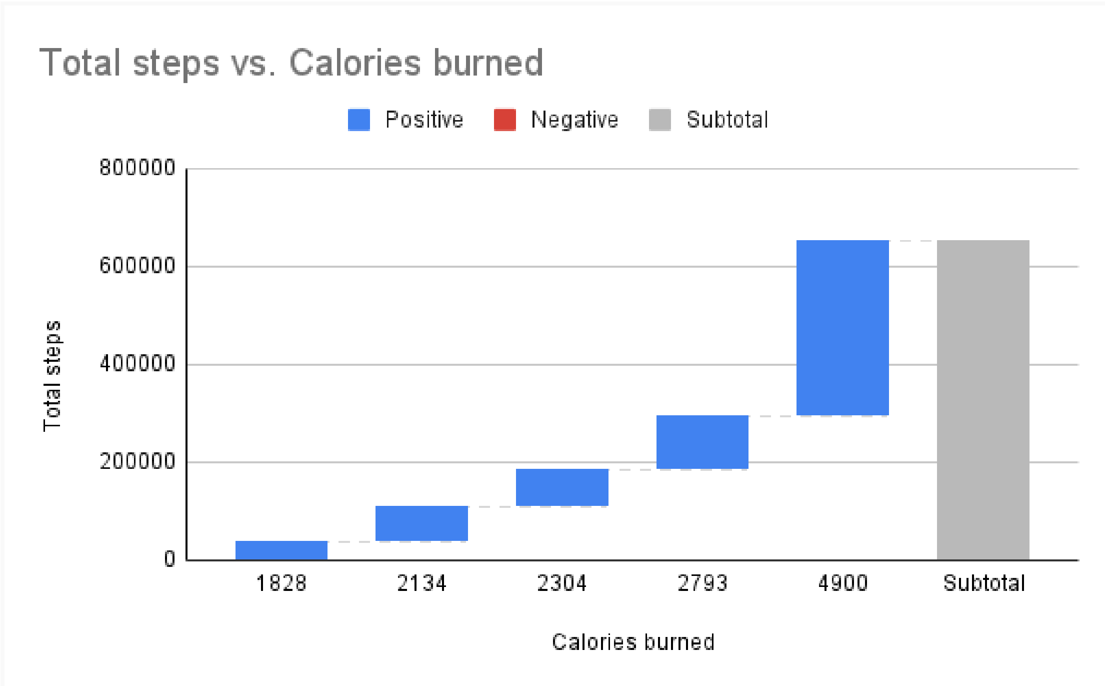
Weight Loss Considerations: Aligning Sleep and Calorie Burning

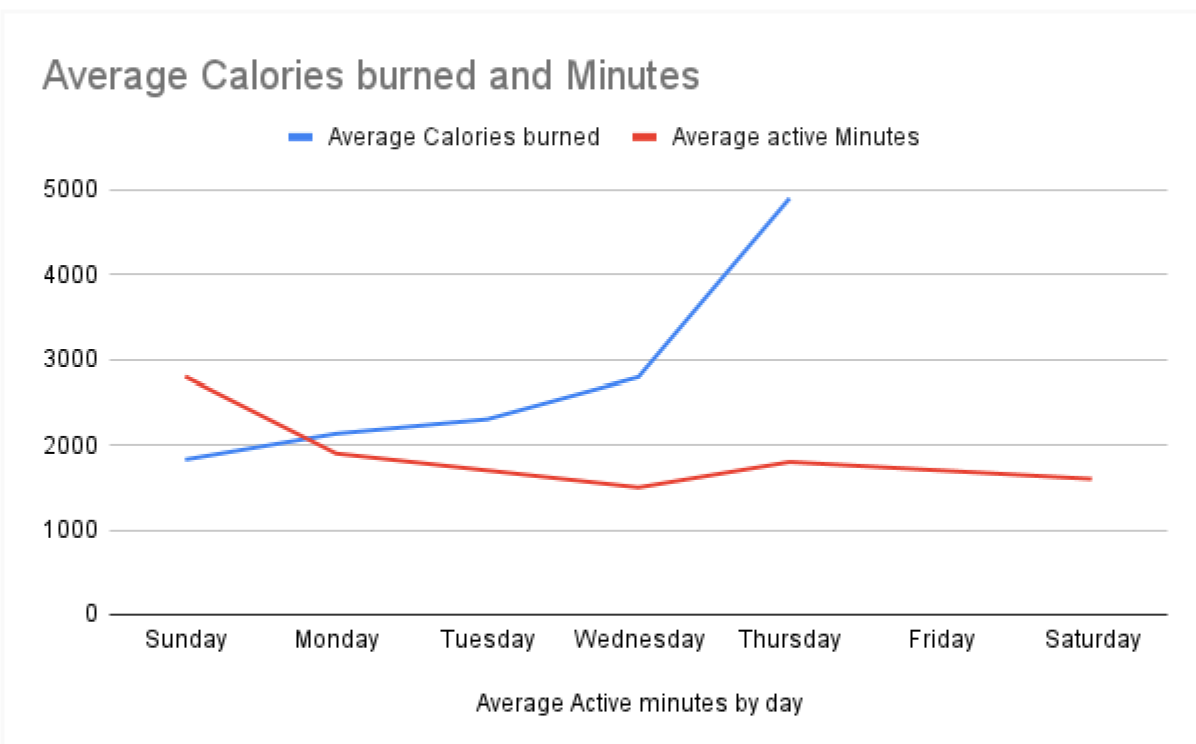
Intriguingly, shifting focus solely to weight loss and considering only calorie expenditure, another interesting alignment emerges. The data suggests a potential correlation between the recommended sleep range of 5.2 to 9.4 hours and higher calorie burning. This overlap with the 6-7-hour sleep range observed earlier warrants further investigation.

Could Adequate Sleep Be a Secret Weapon for Weight Loss?

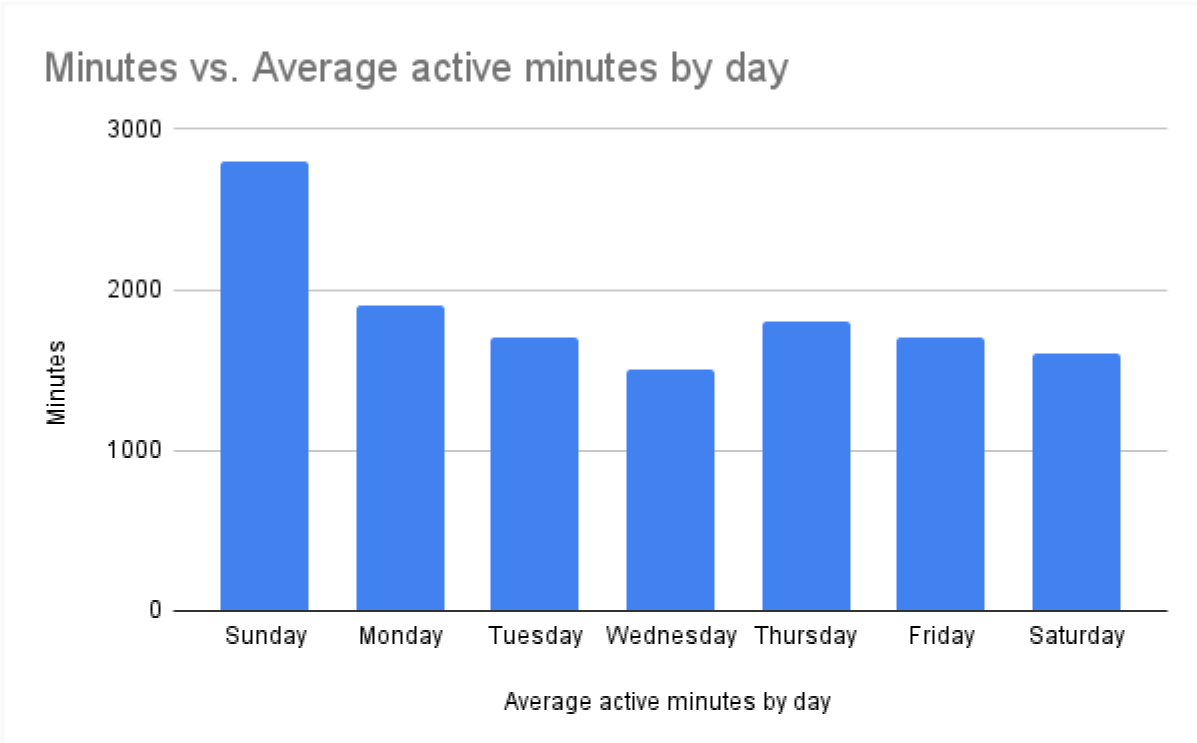
While the data cannot definitively establish causality, it does raise a fascinating question: could adequate sleep, particularly within the recommended range of 5.2 to 9.4 hours, play a role in weight management by influencing calorie expenditure? This potential link deserves further exploration through more comprehensive studies that account for additional factors like individual metabolisms, activity levels, and dietary habits.

Below confirms the notion that the more users move, the more calories they burn. The correlation between movement and calorie burning is likely because being more active burns more calories. Exercise helps the body use energy from calories stored in muscles and fat. The more intense the activity, the more calories users burn.





Below are the average active minutes in a day. People typically have more free time on weekends, leading to increased activity levels. This could include exercise, running errands, socializing, or engaging in hobbies that involve physical movement. Hence, the increased activity on Sunday.



Act

The marketing team could partner with fitness instructors or influencers to create workout routines or challenges that can be integrated with the company's products.

The more intense the activity, the more calories users burn. The marketing team can tailor specific activities that assist in calorie burning the more a person moves while incentivizing them with badges or redeemable points for a small reward from some of the products for every movement completed in a year or a given period.