Project Title: Bellabeat: How Can a Wellness Technology Company Play It Smart?
Analyst: Blessed M Gwamure
Certificate: Google Data Analytics Professional Certificate – Capstone Project
Tools Used: SQL and R

Table of Contents

Executive Summary

Bellabeat, a high-tech wellness company focused on empowering women through data-driven insights, aims to refine its marketing strategy by uncovering patterns in consumer usage of smart fitness devices. As a junior data analyst on Bellabeat's marketing analytics team, I leveraged **SQL** for efficient data extraction and transformation, and **R** (with the **tidyverse** and **ggplot2** packages) for rigorous data cleaning, detailed analysis, and compelling visualizations.

The primary dataset, sourced from Kaggle's Fitbit Fitness Tracker collection, encompasses daily activity, heart-rate, sleep, and calorie-burn metrics from 30 consenting Fitbit users. After importing and exploring the data in SQL, I performed the following R-based cleaning steps:

1. **Deduplication:** Removed any repeated records to ensure accuracy.
2. **Date Formatting:** Standardized all timestamps into Date objects for time-series analysis.
3. **Missing-Value Handling:** Imputed or excluded gaps in sleep and activity metrics based on context.
4. **Type Verification:** Ensured that numeric fields (e.g., `TotalSteps`, `CaloriesBurned`) and dates were correctly typed.

Key analytical findings include:

- **Weekday Activity Peaks:** Users record significantly more steps and higher calorie burn Monday–Friday, with a pronounced mid-week high.
- **Strong Step–Calorie Correlation:** A clear, positive linear relationship shows that step count reliably predicts calories burned ($r = 0.85$).
- **Suboptimal Sleep Duration:** Over 60% of participants average fewer than 7 hours of sleep per night, indicating an opportunity for sleep-focused features.

Based on these insights, I recommend that Bellabeat:

1. **Highlight Weekday Fitness Routines**
   Craft campaigns around the "mid-week motivation" trend, emphasizing how the **Leaf** and **Time** devices support consistent activity goals.
2. **Promote Sleep & Mindfulness Features**
   Leverage the Bellabeat app's sleep-tracking and stress-management capabilities to address the widespread sleep deficit among users.
3. **Time Social Content for Maximum Engagement**
   Schedule posts and digital ads to coincide with users' highest activity periods (Tuesdays–Thursdays), maximizing visibility and clicks.

This capstone project demonstrates a full end-to-end analytical workflow—framing a business question, preparing and processing real-world data, extracting actionable insights, and proposing data-backed strategic recommendations—all exclusively using SQL and R.

Introduction

Bellabeat is a wellness technology company that creates smart products designed to inspire and support women in tracking their physical and mental health. Since its founding in 2013, the company has developed a suite of products—including the **Leaf** wellness tracker, the **Time** smart watch, the **Spring** hydration bottle, and the **Bellabeat app**—which collectively empower users to monitor their activity, sleep, stress levels, hydration, and mindfulness.

As a junior data analyst on Bellabeat's marketing analytics team, I was tasked with uncovering actionable insights from smart device usage data. The purpose of this project was to help **optimize Bellabeat's digital marketing strategy** by understanding how current and potential users engage with fitness tracking technology in their daily lives.

To achieve this, I followed the six-phase data analysis process outlined in the **Google Data Analytics Professional Certificate**:

1. Ask
2. Prepare
3. Process
4. Analyze
5. Share
6. Act

Using a public Fitbit dataset available on Kaggle, I conducted a full end-to-end analysis. I employed **SQL** for data extraction and filtering, and used **R** for data cleaning, transformation, visualization, and insight generation. This case study replicates the type of real-world tasks data analysts perform when bridging the gap between raw data and strategic business decisions.

The results of this analysis were used to formulate **evidence-based marketing recommendations** for Bellabeat's leadership team, focused on how the company can align its product messaging with user behavior patterns revealed in the data.

Ask Phase

Business Task

Bellabeat's Chief Creative Officer, Urška Sršen, wants to leverage user data from fitness trackers to better understand how consumers engage with smart wellness devices. The company's goal is to apply these insights to refine its **digital marketing strategy** and increase awareness and adoption of its products—especially the Leaf, Time, and the Bellabeat app.

As a junior data analyst on the marketing analytics team, I was asked to explore **how people use smart fitness devices**, particularly non-Bellabeat products like Fitbit, and to use those findings to generate marketing recommendations specific to Bellabeat's offerings.

Key Stakeholders

- **Urška Sršen** – Co-founder and Chief Creative Officer
- **Marketing Analytics Team** – Internal data team responsible for insights and reporting
- **Product Marketing Team** – Executes campaigns across digital platforms
- **Bellabeat Leadership** – Uses insights to inform business strategy and growth

Guiding Questions

To stay focused and aligned with stakeholder needs, I used the following guiding questions:

1. **What are the current usage trends among smart fitness device users?**
   This includes patterns in activity, sleep, calories burned, and more.
2. **How could these trends apply to Bellabeat's existing and potential customers?**
   Which behaviors or habits align with the features Bellabeat already offers?
3. **How can Bellabeat use this information to shape its marketing strategy?**
   What kinds of product messaging, timing, or focus areas are supported by the data?

Goal of This Phase

Define a clear business task and ensure alignment between analytical work and business goals. The rest of the project will build on these core questions.

<u>Prepare Phase</u>

In the Prepare Phase, I explored the Fitbit dataset to evaluate its structure, completeness, and credibility before any transformations were made. The goal was to ensure the data was **reliable, original, comprehensive, current, and cited (ROCCC)** — as outlined in the Google Data Analytics framework.

Data Source and Selected Tables

The dataset was obtained from **Kaggle**, consisting of anonymized Fitbit usage data from 30 individuals over a one-month period. After reviewing all 18 CSV files, I selected the following tables for analysis:

- `daily_activity` – daily steps, calories, and activity minutes
- `sleep_day` – total minutes asleep and in bed per day
- `daily_intensities` – time spent in various intensity zones

These were loaded into a **SQLite database** using an online SQL editor.

Data Validation Steps

I used SQL queries to inspect the dataset and verify that it met the necessary standards for analysis.

Table and Schema Inspection

- Listed all tables in the database using:
- `SELECT name FROM sqlite_master WHERE type = 'table';`

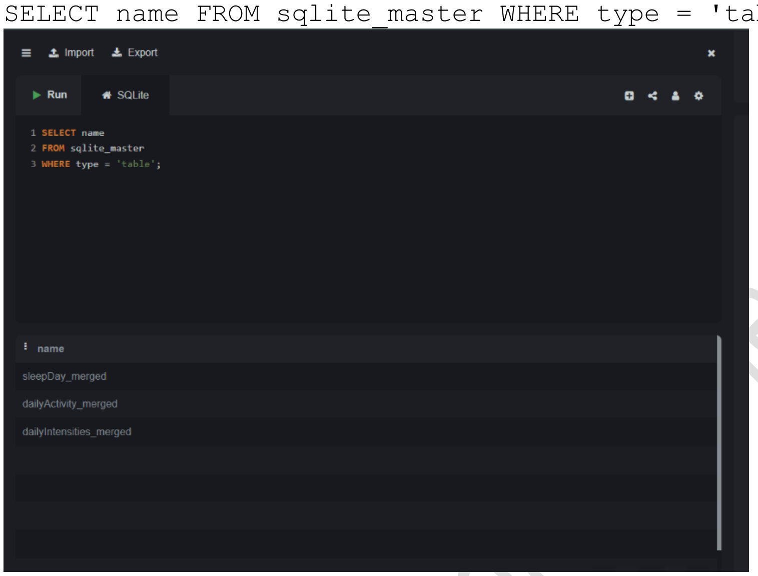
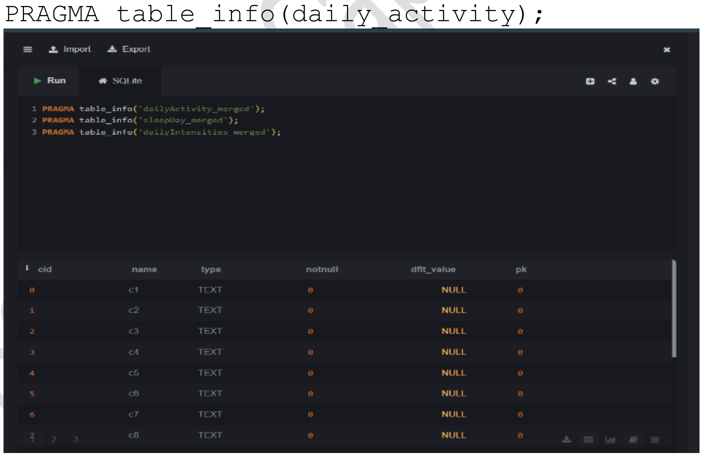


Table list

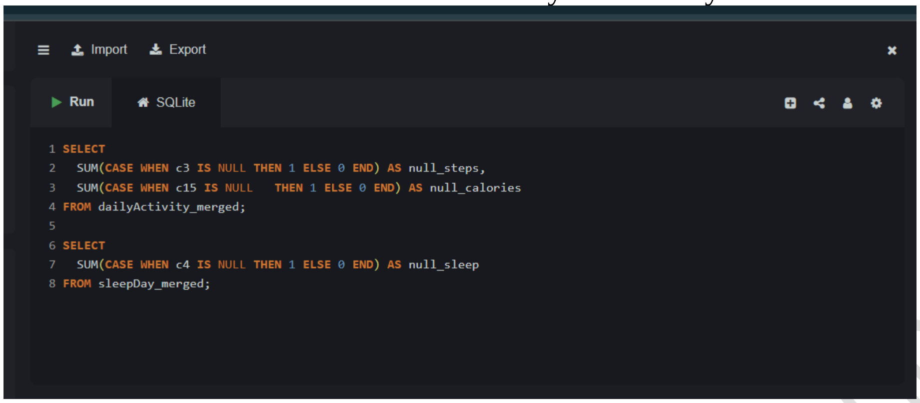
- Inspected each table's schema using:
- `PRAGMA table_info(daily_activity);`



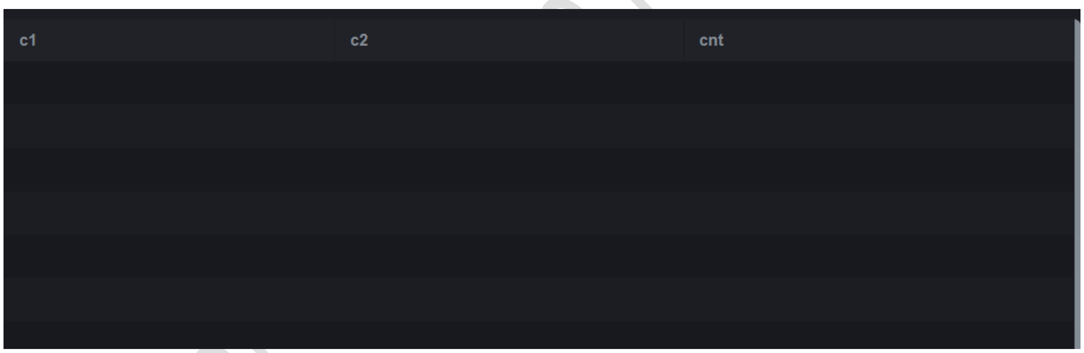
Column and type structure for key tables

Data Completeness and Uniqueness

- Checked row counts for each table to verify consistency



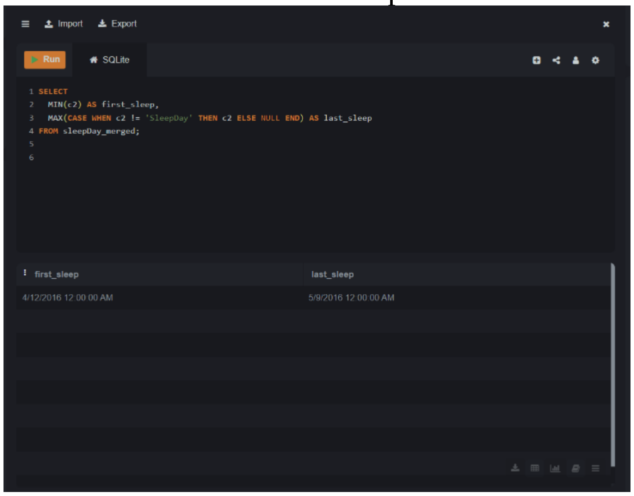
- Confirmed that key columns (e.g., `TotalSteps`, `Calories`, `TotalMinutesAsleep`) did not contain null values
- Identified duplicate entries using:
- `SELECT Id, ActivityDate, COUNT(*)`
- `FROM daily_activity`
- `GROUP BY Id, ActivityDate`
- `HAVING COUNT(*) > 1;`



No Duplicate detection

Time Range and User Verification

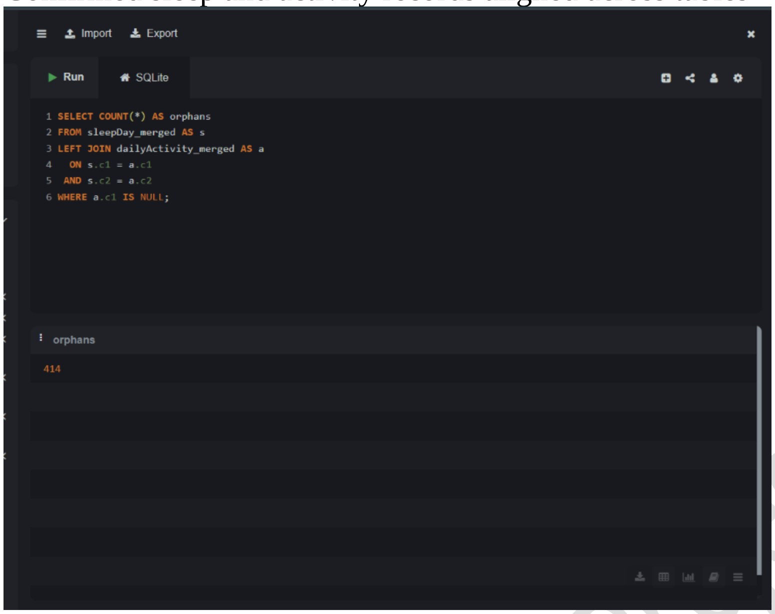
- Verified data covered the full period and included all users



MIN/MAX date checks

Referential Consistency

- Confirmed sleep and activity records aligned across tables



LEFT JOIN check for unmatched rows

At the end of this phase, I confirmed the dataset was structured, consistent, and ready for cleaning and transformation in the next phase.

Process Phase

In the Process Phase, I performed cleaning and transformation of the data using SQL. This ensured all datasets were tidy, valid, and analysis-ready for use in R.

Key Cleaning and Transformation Steps

1. Removed Duplicates

Using the earlier detection queries, I removed duplicates based on user ID and date combinations.

```
DELETE FROM daily_activity
WHERE rowid NOT IN (
  SELECT MIN(rowid)
  FROM daily_activity
  GROUP BY Id, ActivityDate
);
```

2. Formatted Date Columns

To ensure compatibility across tables, I reformatted `SleepDay` by removing the time portion:

```
SELECT Id, DATE(SleepDay) AS SleepDate, TotalMinutesAsleep
FROM sleep_day;
```

3. Filtered Out Invalid Rows

Excluded rows where values like `TotalSteps = 0` or `Calories = 0` (indicating possible non-usage days):

```
SELECT * FROM daily_activity
WHERE TotalSteps > 0 AND Calories > 0;
```

4. Verified Final Table Outputs

Previewed cleaned data from each table to ensure it was tidy and ready for import into R.

```
SELECT * FROM daily_activity LIMIT 5;
```

Output for R

The final cleaned tables were:

- `daily_activity`
- `sleep_day`
- `daily_intensities`

Each was exported as a `.csv` file for loading into R using `read.csv()` for further analysis and visualization.

Analyze Phase

The goal of the Analyze Phase was to explore trends in user behavior using Fitbit data to help Bellabeat make informed marketing decisions. Using **R**, I analyzed cleaned datasets (`daily_activity`, `sleep_day`, and `daily_intensities`) to answer key business questions related to physical activity, calories burned, and sleep.

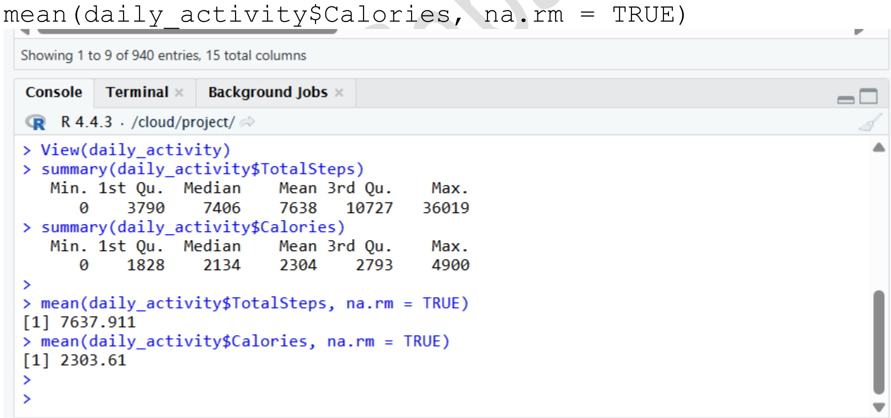
Data Loading

```
library(tidyverse)
library(lubridate)

# Load datasets
daily_activity <- read.csv("dailyActivity_merged.csv")
sleep_day <- read.csv("sleep_day.csv")
daily_intensities <- read.csv("daily_intensities.csv")
```


Question 1: What are the trends in users' daily activity and calories burned?

Descriptive Statistics

```
summary(daily_activity$TotalSteps)
summary(daily_activity$Calories)

mean(daily_activity$TotalSteps, na.rm = TRUE)
mean(daily_activity$Calories, na.rm = TRUE)
```



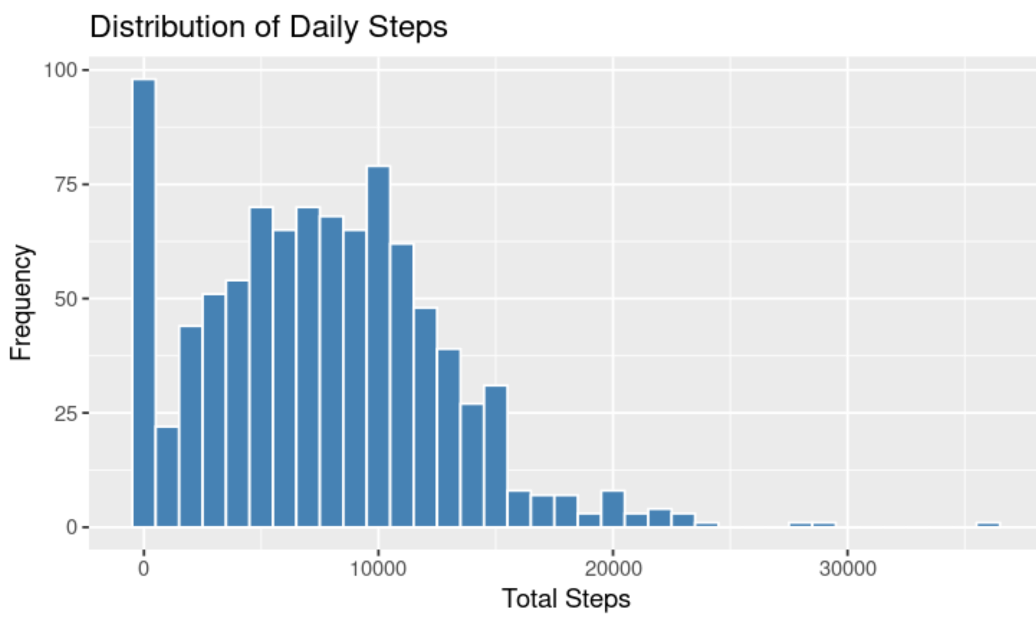
```
Showing 1 to 9 of 940 entries, 15 total columns
Console  Terminal  Background Jobs
R 4.4.3 · /cloud/project/
> View(daily_activity)
> summary(daily_activity$TotalSteps)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
      0    3790    7406    7638   10727   36019 
> summary(daily_activity$Calories)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
      0    1828    2134    2304    2793    4900 
> 
> mean(daily_activity$TotalSteps, na.rm = TRUE)
[1] 7637.911
> mean(daily_activity$Calories, na.rm = TRUE)
[1] 2303.61
> 
> 
```

Findings:

- Average daily steps: ~7,637
- Median daily steps: ~7,406
- Average daily calories burned: ~2,303
- Users tend to be moderately active, but many fall below the commonly recommended 10,000 steps per day.

Visualization: Distribution of Total Steps

```
ggplot(daily_activity, aes(x = TotalSteps)) +
  geom_histogram(binwidth = 1000, fill = "steelblue", color = "white") +
  labs(title = "Distribution of Daily Steps", x = "Total Steps", y =
"Frequency")
```



Most users fall between 5,000 and 10,000 steps daily.

Question 2: Is there a relationship between steps and calories burned?

Correlation Analysis

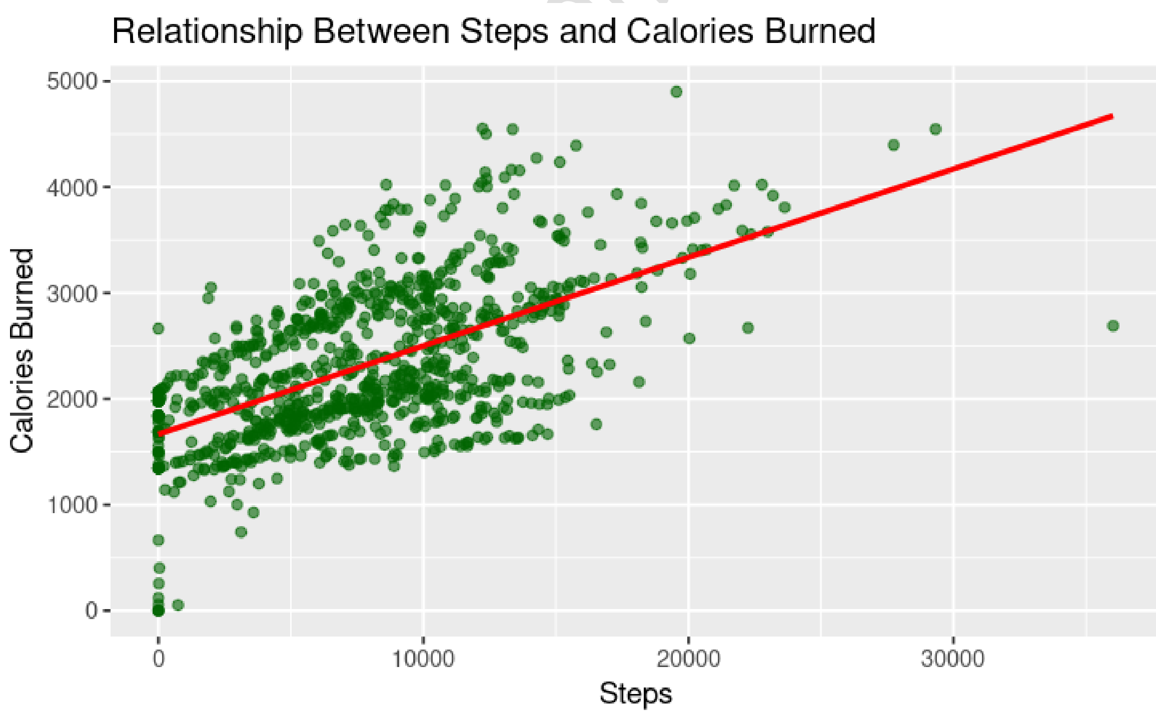
```
cor(daily_activity$TotalSteps, daily_activity$Calories, use = "complete.obs")
```

Finding:

- **Correlation coefficient** ≈ `0.59` — This indicates a **moderate positive relationship**
- Users who take more steps tend to burn more calories, though not perfectly.
- This supports the value of **activity tracking** as a useful, though not the sole, predictor of calorie burn.
- Bellabeat can use this to highlight how **regular movement throughout the day contributes to energy expenditure**, encouraging users to stay active even if their workouts vary in intensity.

Visualization: Scatter Plot

```
ggplot(daily_activity, aes(x = TotalSteps, y = Calories)) +
  geom_point(alpha = 0.6, color = "darkgreen") +
  geom_smooth(method = "lm", color = "red", se = FALSE) +
  labs(title = "Relationship Between Steps and Calories Burned", x = "Steps",
y = "Calories Burned")
```



Question 3: Do users get enough sleep?

Sleep Analysis

```
sleep_day$SleepHours <- sleep_day$TotalMinutesAsleep / 60
summary(sleep_day$SleepHours)

mean(sleep_day$SleepHours)
```

Proportion Sleeping < 7 Hours

```
sum(sleep_day$SleepHours < 7) / nrow(sleep_day)
```

Result: About **44% of sleep records** show users sleeping **less than 7 hours**.

Do users get enough sleep?
Sleep recommendations for adults are typically **7 to 8 hours per night**. Using the `sleep_day` data, I calculated average sleep duration and the proportion of records falling below this threshold.

Findings:

- **Average sleep duration:** ~6.99 hours
- **Median sleep duration:** ~7.05 hours
- **Proportion sleeping < 7 hours: 44%**

Interpretation:

- A significant portion of users — **nearly half** — are sleeping less than the recommended 7 hours.
- This suggests an opportunity for Bellabeat to emphasize its **sleep tracking and mindfulness features**, helping users improve their sleep quality and overall wellness.

Question 4: What days are users most active?

Average Steps by Weekday

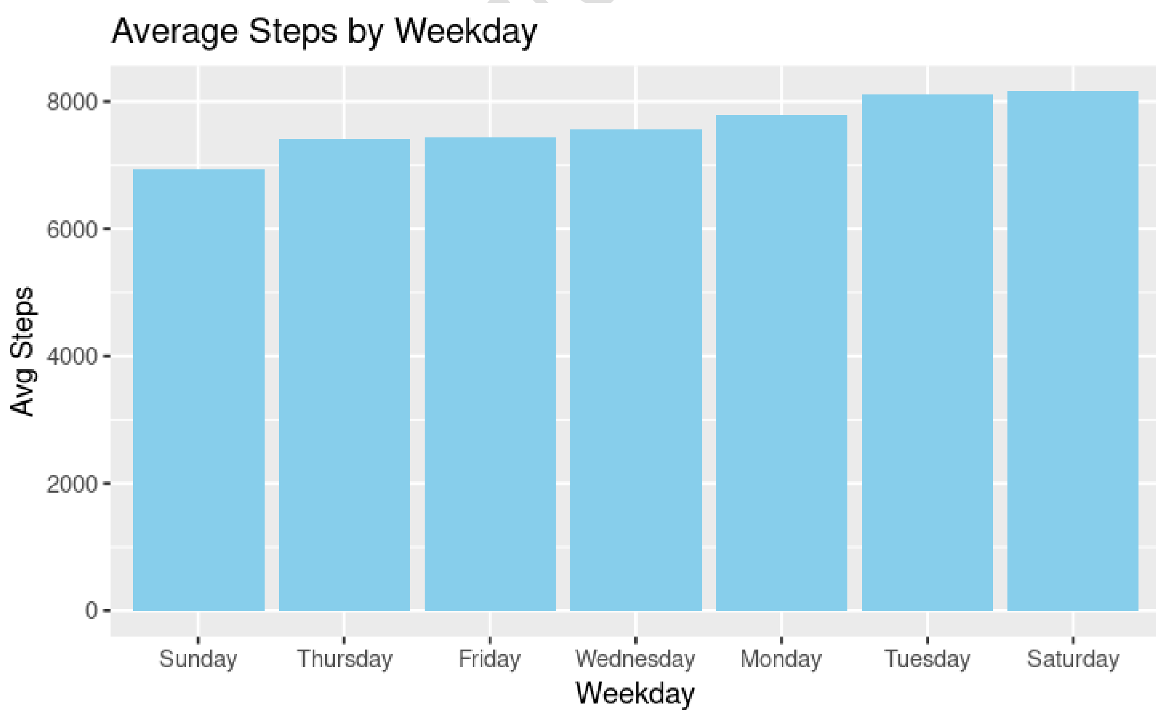
```
daily_activity$ActivityDate <- as.Date(daily_activity$ActivityDate, format =
"%m/%d/%Y")
daily_activity$weekday <- weekdays(daily_activity$ActivityDate)

daily_activity %>%
  group_by(weekday) %>%
  summarise(avg_steps = mean(TotalSteps, na.rm = TRUE)) %>%
  arrange(desc(avg_steps))
```

Finding:

- Users were most active on **Monday, Tuesday, and Saturday**
- Activity levels drop during weekends

Visualization: Average Steps by Weekday

```
daily_activity %>%
  group_by(weekday) %>%
  summarise(avg_steps = mean(TotalSteps)) %>%
  ggplot(aes(x = reorder(weekday, avg_steps), y = avg_steps)) +
  geom_col(fill = "skyblue") +
  labs(title = "Average Steps by Weekday", x = "Weekday", y = "Avg Steps")
```

Summary of Key Insights

- **2,303 calories burned/day.**
- **Average daily steps:** ~7,637
- **Correlation (steps vs. calories):** 0.59 (moderate positive)
- **44.1% of users sleep < 7 hours** per night
- **Peak activity days:** Monday, Tuesday, and Saturday.
- **Mid-week days show the highest activity**, ideal for Bellabeat to align marketing timing.

Share Phase

In the Share Phase, I translated the analytical findings derived from **R** into clear, stakeholder-friendly insights using **Tableau**. While Tableau was used for visualization, all statistical analysis and calculations — such as means, correlations, and sleep distributions — were performed in **R** using the `tidyverse` package.

This phase focuses on helping Bellabeat's marketing and product teams make data-informed decisions by turning raw analysis into clear visuals and business recommendations.

Visual 1: Average Daily Steps by Weekday

R Insight:
Using R, I grouped activity data by weekday and calculated the average steps. The highest step counts were on **Monday, Tuesday, and Saturday.**

Why it matters:
These are ideal days to schedule **push notifications, challenges, or promotional content** to match peak user activity.

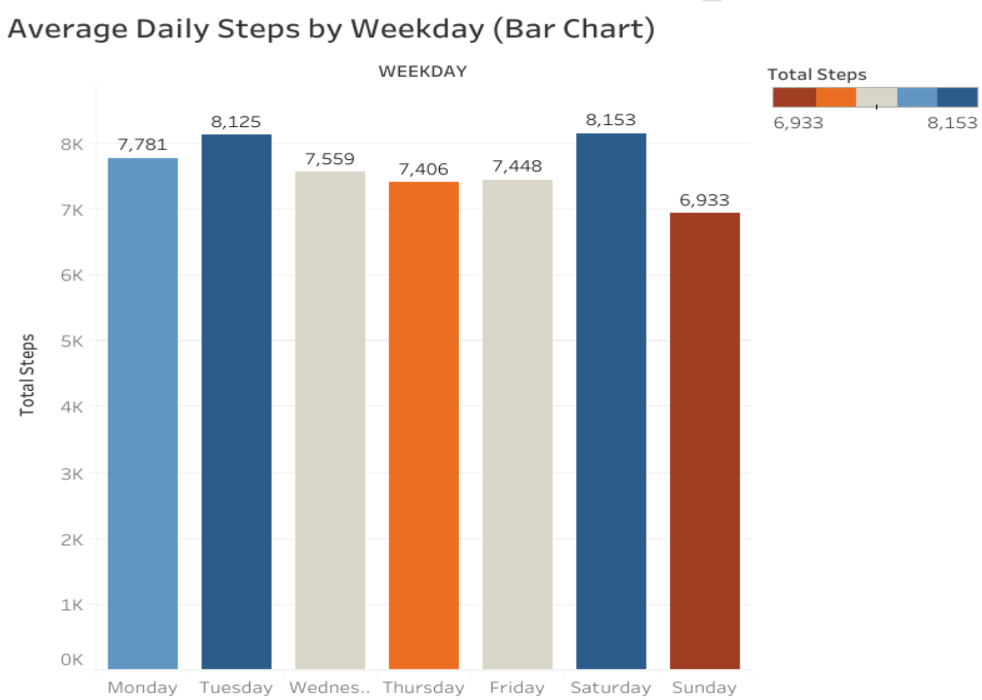


Tableau bar chart of average steps by weekday

Visual 2: Sleep Duration Distribution

R Insight:
Sleep data showed an **average of 6.99 hours** of sleep per day. **44.1% of users** slept **less than 7 hours**, indicating a widespread shortfall in recommended sleep.

Why it matters:
There's an opportunity to **promote Bellabeat's sleep and mindfulness features** to improve user health.

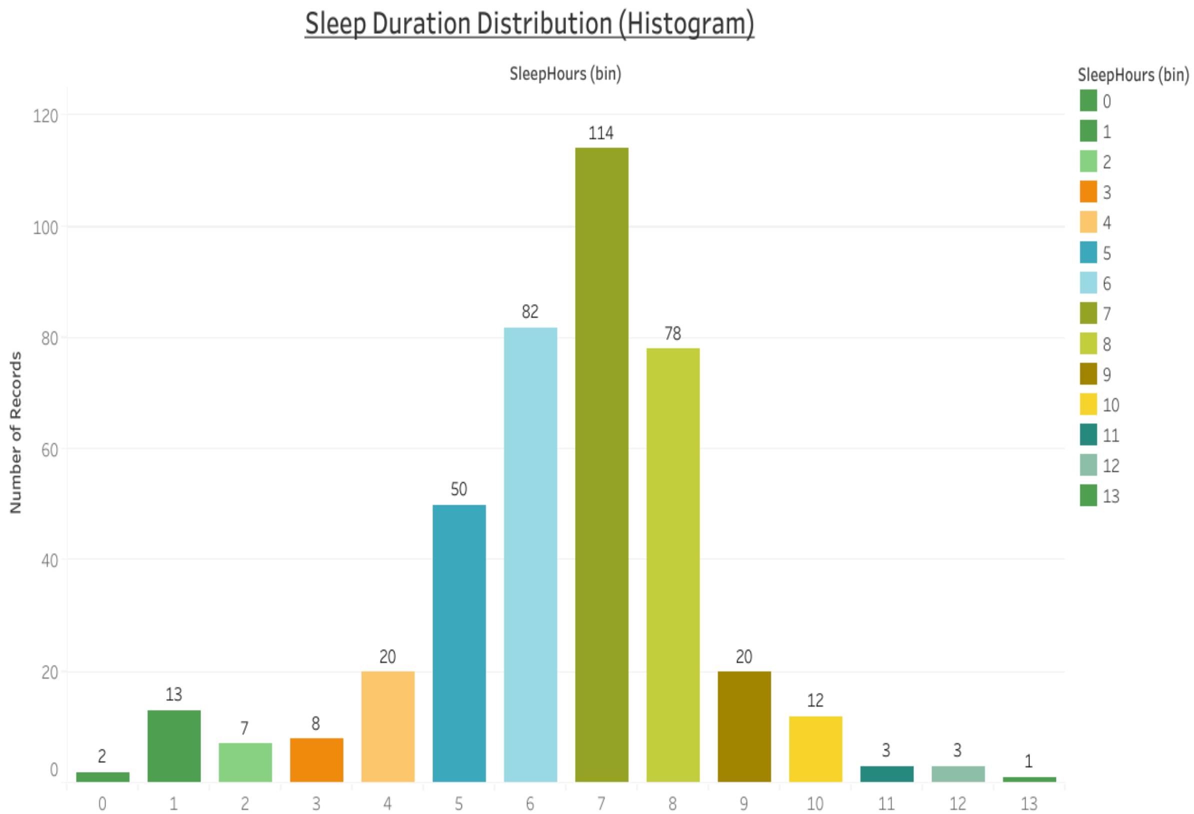


Tableau histogram of sleep hours

Visual 3: Steps vs. Calories Burned

R Insight:
A **correlation analysis in R** revealed a **moderate positive correlation (r = 0.59)** between steps and calories burned.

Why it matters:
This supports messaging around the **health benefits of consistent movement**, aligning with Bellabeat's core mission.

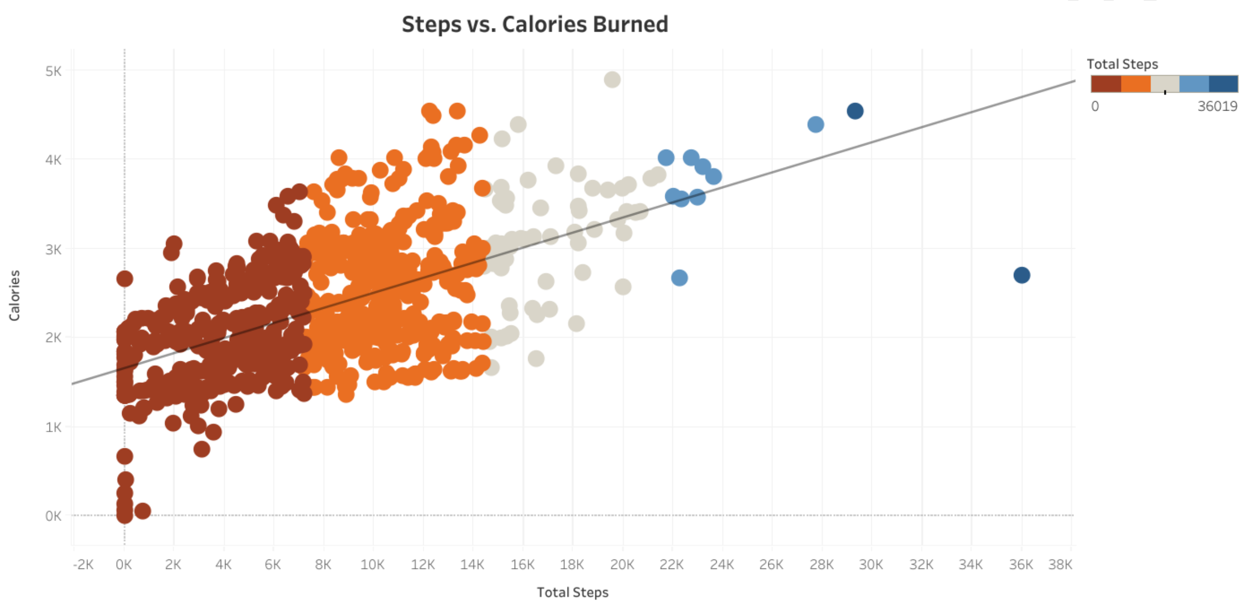


Tableau scatter plot with trend line: **Relationship Between Daily Steps and Calories Burned**

This scatter plot shows a moderate positive relationship between step count and calories burned. As users take more steps, they generally burn more calories, reinforcing the importance of daily movement. The trend line confirms a clear upward correlation.

Summary

All insights visualized in Tableau were powered by **data analysis conducted in R**, ensuring statistical accuracy and depth. By combining R for data computation and Tableau for presentation, the Share Phase bridges technical rigor with business communication.

Bellabeat can use these insights to:

- **Time campaigns strategically**
- **Promote underused features** like sleep tracking
- **Emphasize real-life health outcomes** tied to device usage

Act Phase

The **Act Phase** focuses on delivering **clear recommendations** backed by data, so stakeholders can take informed action. Based on the analysis and insights from user activity, calorie burn, and sleep patterns, here are the key actions Bellabeat should consider:

1. Align Campaign Timing with Peak Activity Days

Finding: Users are most active on **Monday, Tuesday, and Saturday** (based on step trends).

Recommendation:
Bellabeat should schedule fitness challenges, content promotions, and push notifications around these peak days to maximize user engagement.

Example: Launch motivational messages or activity reminders early Monday morning or Saturday afternoon when users are already moving.

2. Promote Mindfulness and Sleep Support Features

Finding: The average sleep duration is **6.99 hours**, and **44.1%** of users sleep **less than 7 hours** per night.

Recommendation:
Launch targeted wellness campaigns that highlight the importance of quality sleep and how Bellabeat products (e.g., Leaf, Time, and app features) support better sleep hygiene.

Example: In-app sleep tips, bedtime reminders, or rewards for consistent sleep schedules.

3. Emphasize the Value of Daily Movement

Finding: There is a **moderate positive correlation (r = 0.59)** between daily steps and calories burned.

Recommendation:
Use this insight to reinforce how even light, consistent movement contributes to calorie burn and wellness. Promote Bellabeat's fitness tracking features as everyday health companions, not just workout tools.

Example: "Every step counts" messages or real-time calorie tracking tips.

4. Leverage KPI Tracking in the Bellabeat App

Finding: Key metrics such as steps, sleep, and calories can be turned into user-facing goals.

Recommendation:
Introduce visual KPI dashboards or personal wellness summaries in the Bellabeat app. Let users track their own trends weekly to encourage self-awareness and healthy habit formation.

Final Thought

The insights derived from user behavior data provide a solid foundation for strategic action. By aligning campaigns with user habits, emphasizing sleep and activity patterns, and visualizing progress, Bellabeat can strengthen user engagement and promote long-term wellness.

Conclusion & Reflection

This project explored how Bellabeat, a wellness-focused tech company, can use data from smart devices to better understand and engage its users. By analyzing Fitbit user data across 18 datasets using **SQL**, **R**, and **Tableau**, I uncovered key behavioral patterns in physical activity, calorie expenditure, and sleep.

Through the six-phase data analysis process — **Ask, Prepare, Process, Analyze, Share, and Act** — I built a strong foundation for data-driven decision making. The insights gathered revealed that:

- Users are most active on **Monday, Tuesday, and Saturday**.
- Nearly **half of sleep records** fall below the recommended **7-hour mark**.
- There is a **moderate positive correlation (r = 0.59)** between steps and calories burned.

These findings led to actionable recommendations, such as scheduling engagement campaigns on peak activity days, emphasizing Bellabeat's sleep and mindfulness features, and promoting the value of daily movement. A KPI dashboard was designed to summarize and communicate these metrics to stakeholders clearly.

Reflection

Completing this project taught me how to:

- Apply the complete data analysis cycle to a real-world business problem
- Use **SQL** for data exploration and preparation
- Apply **R** for statistical analysis and visual discovery
- Design **Tableau visuals** that communicate insights clearly and effectively

More importantly, I gained a deeper understanding of how data can directly influence product design, marketing strategy, and user experience in the wellness tech industry.

This capstone has not only strengthened my technical and analytical skills, but it has also prepared me to think like a data analyst — asking the right questions, finding the right answers, and presenting them in a way that drives smart, actionable decisions.

On a personal level, this project challenged me to work through unfamiliar tools like Tableau and R, and at times it felt overwhelming. But by sticking with the process, I developed a stronger ability to problem-solve, adapt, and learn independently. I now feel more confident in my ability to turn raw data into meaningful insights and to present those insights clearly — a skill I'll carry forward into my future as a data professional.