# Ask:

### **Business Task Statement**

**Objective**: Analyze smart device usage data to identify trends in how consumers interact with health-focused smart devices, particularly non-Bellabeat devices, to gain insights that will inform marketing strategies for Bellabeat.

### **Specific Questions to Address**

1. **Trends in Smart Device Usage**: Identify patterns in usage, including frequency of use, types of activities tracked, and features that are most popular among users of non-Bellabeat smart devices.
2. **Relevance to Bellabeat Customers**: Examine how the identified trends in smart device usage align with the behaviors and preferences of Bellabeat's target demographic—health-conscious women—and how these trends could be applicable to enhance user engagement with Bellabeat products.
3. **Influence on Bellabeat Marketing Strategy**: Determine actionable insights from the analysis that can guide Bellabeat's marketing strategies, focusing on product positioning, targeted messaging, and promotional efforts that resonate with consumer habits and preferences uncovered in the analysis.

### **Key Stakeholders**

* **Urška Sršen (Cofounder and Chief Creative Officer)**: Interested in how insights align with Bellabeat’s mission and can empower women with health information.
* **Sando Mur (Cofounder)**: Focused on analytical validation to support marketing strategies for growth.
* **Bellabeat Marketing Analytics Team**: Needs data-driven recommendations to enhance marketing efforts across various channels.

### **Driving Business Decisions**

The insights gathered from this analysis will inform how Bellabeat can:

* Better align its product offerings with user preferences and behaviors.
* Develop targeted marketing campaigns that resonate with potential customers.
* Differentiate Bellabeat’s offerings in a competitive market by addressing unique consumer needs identified in the smart device usage data.

### **Summary**

The business task is to leverage smart device usage data to uncover insights on user behaviors and preferences. These insights will not only inform marketing strategies for a selected Bellabeat product but also enhance customer engagement and drive product adoption in alignment with Bellabeat's mission to empower health-conscious women.

# Prepare:

**Fitbit Data Analysis Preparation Summary**

### **Data Source**

The dataset used is the **Fitbit Fitness Tracker Data** (publicly available through Kaggle under the CC0 license), which includes personal activity data from 30 Fitbit users. The data contains insights into daily habits like activity levels, steps, calories, heart rate, and sleep, which will help explore user behaviors.

### **Data Structure**

The dataset includes various files that track:

* **Daily Activity**: Steps, distance, calories.
* **Sleep Monitoring**: Sleep duration and patterns.
* **Heart Rate**: Minute-by-minute heart rate recordings.

Data is organized in a mix of **long and wide formats**:

* **Long Format** (e.g., minute-level tracking for heart rate).
* **Wide Format** (e.g., daily summaries per user).

### **Data Integrity and Bias Considerations**

* **Data Quality**: Missing values and duplicates have been identified, and appropriate cleaning actions will be taken.
* **Bias**: The dataset represents a small subset of users and may not fully capture broader demographics or fitness levels. Additional data or insights may be needed for a more comprehensive understanding.

### **Privacy, Licensing, and Compliance**

This dataset is under the **CC0 license**, allowing free use with acknowledgment. User consent for data submission was obtained, ensuring ethical handling of personal information.

# 

# Process:

**Fitbit Data Processing Documentation**

**Objective:**

To clean and transform the Fitbit dataset for effective analysis, ensuring that it is organized and free of errors. This step enables the dataset to provide meaningful insights into smart device usage patterns.

**Tools Chosen:**

R Programming Language:

* Chosen for its versatility in data cleaning, manipulation, and analysis.
* R’s robust libraries (e.g., dplyr, tidyr) simplify tasks like data wrangling, transformation, and visualization.

**Data Integrity and Cleaning Steps**

1. Data Loading:

* All CSV files were imported into R using read.csv() and organized in a dedicated folder following consistent naming conventions for easy reference (e.g., daily\_activity.csv, sleep\_data.csv).

1. Initial Checks for Errors:

* Duplicate Removal: Identified and removed duplicate rows to ensure each record is unique, particularly in the sleepDay and dailyActivity files.
* Missing Values: Used summary() to check for missing data. Found and handled missing values by replacing or removing, depending on the variable context.
* Data Types: Converted data types to match analysis needs (e.g., ActivityDate was converted to a Date format).

1. Transformations for Consistency:

* Date Standardization: All date fields were standardized to YYYY-MM-DD format to ensure uniformity across files.
* Format Alignment: For files in a mix of long and wide formats (like minute-level files), reshaped data to a consistent format, making sure timestamps align across entries.

1. Variable Renaming:

* Renamed columns for clarity (e.g., TotalSteps to total\_steps) and standardized across files.

1. Data Documentation:

Documented each cleaning step with comments in the R script to facilitate future reviews and reproducibility.

# Summary of the Data Analysis Process

#### 

#### **1. Data Organization and Format**

* **Load Data:** The first step was to load the dataset (likely a CSV or data file) into an R environment for processing.
* **Check Column Names:** It was important to inspect the column names (colnames(merged\_data)) to ensure the data had been loaded properly and that the column names were consistent with the analysis requirements.
* **Address Missing Data:** Missing values in certain columns like sleepday, totalsleeprecords, etc., were identified, and any missing data was handled appropriately (e.g., using na.rm = TRUE to exclude missing values during aggregations).

#### **2. Data Aggregation and Summary**

* **Calculate Total Steps and Average Steps per Month:** Using group\_by(month\_year) and summarise(), the data was aggregated by month\_year to calculate the total steps and the average steps per month. This provided insight into overall trends over time.
* **Summarize Key Metrics:** Summary statistics such as the total steps, total calories burned, active minutes, etc., were calculated for each month, enabling an overview of the user's activity patterns.

#### **3. Visualizations**

* **Scatter Plots:**
  + **Steps vs Calories Burned:** A scatter plot was created to examine the relationship between total steps (totalsteps) and calories burned (calories.x). Each point represented an observation, showing how the number of steps correlates with calories burned.
  + **Active Minutes vs Calories Burned:** Another scatter plot was created for the relationship between active minutes (veryactiveminutes.x) and calories burned (calories.x).
* **Trendlines:**
  + A **linear regression trendline** was added to both scatter plots using geom\_smooth(method = "lm", se = FALSE, color = "red"). This trendline helped visualize the underlying relationship between the variables.

#### **4. Trends and Insights**

* **Steps and Calories Burned:** The scatter plot between steps and calories burned, coupled with the trendline, helped identify if there's a linear relationship between the two. This could reveal how well activity levels (steps) correlate with calories burned.
* **Active Minutes and Calories Burned:** Similarly, the plot for active minutes versus calories burned helped explore if increased physical activity (active minutes) results in higher calorie expenditure.
* **Aggregated Trends Over Time:** The monthly aggregation allowed for identifying patterns and trends in the data, such as fluctuations in physical activity across different months or the consistency of calories burned over time.

#### **5. Key Findings**

* **Relationships between Variables:** The trendlines confirmed or rejected the hypothesis that more steps and more active minutes correlate with higher calorie burn.
* **Activity Trends Over Time:** The monthly aggregation of steps and calories burned provided insights into whether the user’s activity levels varied by month or were relatively stable.

#### **6. Actionable Insights for Business**

* **User Health Insights:** Understanding the relationship between steps, active minutes, and calories burned could inform personalized health recommendations, helping users track their fitness goals.
* **Improved App Design:** Insights from these relationships could be used to refine a fitness app or dashboard, making it more intuitive for users to monitor their physical activity and calorie expenditure.

### 

### **Futher Investigations**

### **1. Analysis Summary:**

I conducted various analyses to explore the relationships between key activity metrics, such as steps taken, calories burned, active minutes, sedentary time, and sleep duration. Below are the main analyses performed:

* **Scatter Plots with Trendlines:** I visualized correlations between pairs of variables (e.g., steps vs. calories, active minutes vs. calories, sedentary minutes vs. calories) to identify trends and assess the strength of relationships. Trendlines were added to help interpret patterns.
* **Correlation Heatmap:** A heatmap was created to provide an overview of correlations among multiple numeric variables. This helped me identify subtle relationships not immediately visible in scatter plots, such as weak correlations between sedentary time and calories burned.
* **Monthly Reward Progress Visualization:** To encourage user engagement, I calculated monthly goal achievement (days meeting step goals) and visualized it to assess the consistency of goal attainment. This provides a basis for introducing reward-based incentives.

### **2. Next Steps and Recommendations:**

#### **2.1 Personalized Insights and Reminders:**

Based on observed patterns, I recommend implementing personalized insights and reminders that encourage users to maintain activity levels. For example:

* **Weekly Summary Reports:** Highlight each user's weekly performance against step and active minutes goals, with suggestions to boost performance where shortfalls are identified.
* **Smart Reminders for Inactivity:** If sedentary time exceeds a threshold, prompt the user to take short breaks or engage in light activity.

#### **2.2 Gamification and Rewards System:**

Use monthly goal achievements as a framework for introducing gamification. Recommendations include:

* **Progress Badges:** Reward users with virtual badges for consistently meeting daily step goals.
* **Monthly Goal Challenges:** Encourage users to reach monthly milestones (e.g., 20+ goal-met days in a month) and unlock rewards, enhancing motivation and engagement.

#### **2.3 Targeted Health Recommendations:**

Using correlations observed between sedentary time, sleep, and active minutes, provide users with actionable health advice. For instance:

* **Sedentary Time Reduction Tips:** Suggest low-impact activities for users with high sedentary minutes.
* **Improving Sleep Quality:** For users with lower active minutes and high sedentary time, suggest adjustments to improve sleep, such as setting sleep routines or reducing screen time before bed.

#### **2.4 Enhanced Data Visualizations for Progress Tracking:**

To better monitor trends over time:

* **Detailed Timeline Visuals:** Display users' daily progress with timelines that break down each activity metric weekly or bi-weekly, helping users track consistency.
* **More Granular Date Visuals:** Provide finer details on the x-axis of visualizations (e.g., every two weeks) to help users see shorter-term progress within monthly trends.

# Summary of the Analysis and Recommendations

#### 

**Objective**: The goal was to analyze user activity data to identify patterns and provide insights that can drive user engagement and promote healthier habits.

### **Key Insights:**

1. **Activity and Calories Burned**: Active minutes positively correlate with calories burned. Users who engage in more physical activity tend to burn more calories.
2. **Sedentary Time Impact**: High sedentary time negatively correlates with calorie burn, suggesting the importance of reducing inactivity.
3. **Goal Achievement**: Monitoring monthly goal progress shows patterns of user engagement, with certain users consistently meeting their step goals.

### **Recommended Visualizations:**

1. **Scatter Plots**: To show relationships between activity levels (e.g., active minutes, sedentary time) and calories burned.
2. **Correlation Heatmap**: To visually represent correlations between multiple wellness metrics like steps, sedentary time, and calories burned.
3. **Monthly Goal Progress Line Chart**: To track users' goal achievements over time, motivating them with visual progress and rewards.

### **Next Steps:**

1. **Finalize Visualizations**: Clean, well-labeled, and insightful charts that highlight key findings.
2. **Presentation**: Use PowerPoint or Google Slides to present these visualizations to stakeholders, ensuring accessibility by using high-contrast colors and clear labels.
3. **User Engagement Strategy**: Leverage these insights to drive a reward system in the app, encouraging users to achieve their health goals with personalized feedback and incentives.

# Final Action Plan

#### 

#### **1. Business Task Summary**

Analyze user activity data to provide insights for improving engagement and promoting healthy habits through a rewards system. Focus on how activity levels (steps, calories, active/sedentary minutes) impact goal achievements.

#### **2. Data Sources**

* **Merged Data**: User activity metrics (steps, calories, active minutes, sedentary minutes, sleep).
* **Additional Data**: Target goals (if available), timestamps for activity days.

#### **3. Data Cleaning & Manipulation**

* Converted dates to monthly format.
* Handled missing values.
* Aggregated data by month for goal tracking.

#### **4. Analysis Summary**

* **Goal Met Days**: Tracked user progress toward meeting goals.
* **Correlations**: Identified key relationships between activity metrics and calories burned.
* **Visualizations**: Used scatter plots and heatmaps to visualize activity-calories correlations and monthly goal progress.

#### **5. Key Findings**

* **Active Minutes**: Strong correlation with calories burned.
* **Sedentary Time**: Negative correlation with calories burned.
* **Goal Progress**: Monthly goal achievements can be tracked to encourage rewards.

#### **6. Recommendations**

* **Rewards System**: Create personalized rewards based on goal achievements and activity levels.
* **User Notifications**: Encourage more active minutes and less sedentary time via reminders.
* **Personalized Experience**: Use data insights to offer customized challenges and notifications.

### **Next Steps:**

* Implement a rewards system based on goal progress.
* Enhance user engagement with targeted notifications.
* Consider integrating additional data (e.g., diet) for better personalization.
* Conduct A/B testing to determine the most effective incentive strategies.

### **Deliverables:**

* **Portfolio**: Summary of findings with visualizations.
* **Case Study**: Documentation of the analysis process.
* **Presentation**: A concise presentation of insights and recommendations.