**Bellabeat Smart Device Data Analysis**

**Introduction**

Bellabeat is a high-tech company founded in 2013 by Urška Sršen and Sando Mur, focused on creating beautifully designed, health-focused smart products for women. With a mission to empower women through technology, Bellabeat provides innovative products that track activity, sleep, stress, and reproductive health, helping users better understand their wellness habits. Their product line includes the **Leaf wellness tracker**, the **Time smart watch**, and the **Spring smart water bottle**, all of which connect to the Bellabeat app. Bellabeat also offers a subscription-based membership that provides personalized guidance on nutrition, activity, and mindfulness. Over the years, Bellabeat has expanded globally and positioned itself as a leader in the women's wellness tech space. The company leverages both traditional and digital marketing channels, with a strong emphasis on data-driven growth strategies.

**1. Ask**

**Business Task**

**Objective**:  
As a junior data analyst at Bellabeat, I was tasked with analyzing **smart device usage data** to provide insights into how consumers use non-Bellabeat smart devices. This analysis aims to help Bellabeat’s executive team better understand trends in smart device usage and apply those insights to one of Bellabeat’s products, influencing future marketing strategies. The analysis will focus on identifying trends in consumer habits from the Fitbit Tracker Dataset and presenting actionable recommendations.

**Key Questions**:

1. What are some trends in smart device usage?
2. How could these trends apply to Bellabeat customers?
3. How could these trends influence Bellabeat’s marketing strategy?

**Stakeholders**:

* **Urška Sršen**: Bellabeat’s co-founder and Chief Creative Officer.
* **Sando Mur**: Co-founder and key member of Bellabeat’s executive team.
* **Bellabeat Marketing Analytics Team**: Responsible for analyzing data and guiding marketing strategy.

**2. Prepare**

**Data Sources**

**Primary Data Source**:

* **Fitbit Fitness Tracker Data**: A publicly available dataset from Kaggle. It contains personal fitness tracker data from 30 Fitbit users who consented to the submission of their daily activity, steps, heart rate, and sleep monitoring data. The dataset includes minute-level output for physical activity, heart rate, and sleep patterns, which can be analyzed to explore user habits. This data is available under the **CC0: Public Domain** license.

**Key Data Files**:

1. **dailyActivity\_merged.csv** – Contains data on steps, distance, and calories burned.
2. **heartrate\_seconds\_merged.csv** – Provides heart rate data.
3. **sleepDay\_merged.csv** – Tracks sleep patterns, including total sleep minutes and time in bed.
4. **hourlySteps**\_**merged** – Contains hourly data on steps.

**Data Format**:  
The dataset is structured in a **long format**, where each row represents an individual user’s activity at a specific time point.

**Data Integrity**:

* The data was verified for completeness and credibility. Some potential biases may exist since the data comes from a limited sample of 30 users, which may not represent the larger population. Another issue is that the data is too old, it may no longer represent current trends in **smart device usage** because technology has evolved significantly since the data was collected.
* However, it offers useful insights into user behavior trends.

**3. Process**

**3.1 Tools Used**:

* **Excel**: Used for basic data filtering, sorting, and preliminary analysis.
* **SQL Server**: Used for longer data file **heartrate\_seconds\_merged.csv**, containing more than 24 lakhs rows, which coluld not be handled in excel worksheet because of its limitations upto almost 10 lakh rows.

**3.2 Uploading the selected CSV file**

Imported following CSV file into different excel worksheets.

**dailyActivity\_merged.csv**

**sleepDay\_merged.csv**

**hourlySteps\_merged.csv**

The following file is too large (more than 24 lakhs rows) to handle in a excel worksheet, so imported it into a **sql server** table.

**heartrate\_seconds\_merged.csv**

**3.3 Data Cleaning and Manipulation**

**Data Cleaning Steps**:

1. **Duplicate Removal in Excel** : Checked for duplicate rows and removed using Remove Duplicates feature in Data Tools.

There were no duplicate rows in *dailyActivity\_merged.csv*and *hourlySteps\_merged.csv*. From *sleepDay\_merged.csv* 3 duplicate rows were removed.

1. **Duplicate Removal in Sql Server** : Scanned the *heartrate\_seconds\_merged* table in **sql server** for any duplicate entry using the following query, and found no duplicate entry :

SELECT Count(\*) Id, Time

FROM heartrate\_seconds\_merged

GROUP BY Id, Time

HAVING Count(\*)>1

1. **Missing Values**: Checked for missing or incomplete data points, especially in critical columns such as Id, Activity Date, TotalSteps, Calories, HeartRate, and TotalMinutesAsleep.
2. **Data Transformation**:
   * There were inconsistencies with columns having dates across all the csv files imported. For data in excel sheets used **Text to Columns** feature bring dates to a consistent format.
   * To eliminate inconsistencies in *hourlySteps\_merged* data in excel first splited the ActivityHour column as date and hours in two separate columns with **Text to Columns** feature. Then again used the same feature to remove inconsistencies present with date values.
   * The *heartrate\_seconds\_merged* data in **sql server** also had the same issue with date values. To handle it, first splitted the column named as Time (with datetime value) into two separate columns using following query:

UPDATE heartrate\_seconds\_merged

SET

ActivityDate = CAST(SUBSTRING(Time, 1, CHARINDEX(' ', Time) - 1) AS DATE),

ActivityTime = CAST(SUBSTRING(Time, CHARINDEX(' ', Time) + 1, LEN(Time)) AS TIME);

* + All other columns also checked for proper data types, such as converting nmbers stored as text to numeric data where necessary.
  + Changed formats of numeric columns as per need.

1. **Summarize and Import table:**

As mentioned earlier, the *heartrate\_seconds\_merged* has been uploaded in sql server table, due to very large size of data. Summarized the table on Id+date imported to excel in dailyActivity\_merged workbook as *HeartRate\_datewise*

SELECT ID, ActivityDate, AVG(Value) as AverageHeartRate INTO AverageHeartRate\_Date

FROM heartrate\_seconds\_merged

GROUP BY ID, ActivityDate

1. **Merging Tables :**

DailyActivity, SleepDay and HeartRate\_datewise table were merged (Joined) using power query’s merge query feature using left join on Id and Date. This was done to find the pattern how the participants are using different feature of the fitness tracker device.

Then loaded back the merged table into a worksheet named **dailyActivity\_sleepDay\_heartRate**

**4. Analyze**

**Sorting of tables:** The first thing we did is, sorted most of the tables in excel and sql server on the combination of Id and Date columns, and a few tables on Id, Date and Hour.

**Unique Participants Count:** We started our analysis with counting number of unique participants, using Unique and Count function. We found there are 33 participants.

**Date-wise Count of Users/Participants:** Created a pivot table from *dailyActivity\_sleepDay\_heartRate* to calculate datewise count of users for different features i.e. Total Steps, Total Distance, Calorie burned, Sleep Time, Heart Rate. This table illustrates how users are tracking different activity and paramenters with different features of the device.

**Users/Participant-wise Average of Daily Activity Parameters:** Created pivot table to calculate average of Total Steps, Total Distance, Calorie Burned. This has been done to know if there is any co-relation between Total Steps and Calories Burned, Total Distance and Calories Burned.

**Users/Participants per day using different features:**  Used pivot table created earlier for Date-wise Count of Users/Participants for this analysis. Calculated average number of users for each matrics being used, to know average percentage of users/day is using a certain feature.

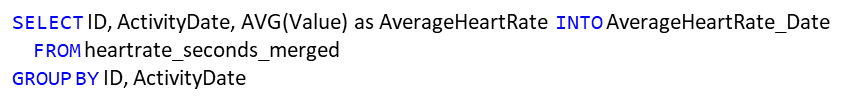
**User/Participant Grouping**: Created pivot tables to know how frequently the users are using different features of Fitness Tracker to track different matrics. We focused on 3 matrics, Total Steps, Sleep Time and Heart Rate

* **Power Users**: Participants who tracked metrics almost daily (90% or more of the days).
* **Frequent Users**: Participants who tracked metrics 60-90% of the days.
* **Occasional Users**: Participants who tracked metrics less than 60% of the days.
* **Inactive Users**: Participants who never tracked certain metrics.

**Hourly average steps:** Calculated hourly average of steps using power query’s Group By feature. This was done to know users’ activeness (behaviour) during various time span of the day.

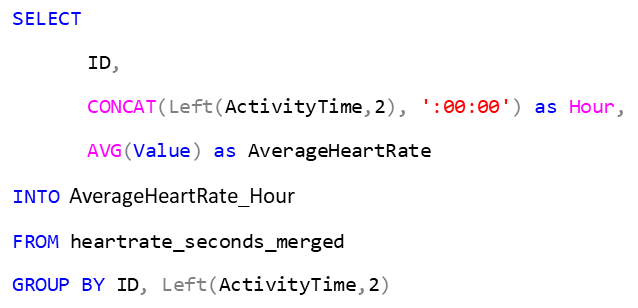
**User-wise hourly steps:** Created pivot table to calculate user-wise hourly average of steps. This can be used to know behaviour (activeness) of a particular user or a bunch of users during various time span.

**Date-wise average heart rate:** Calculated date-wise heart rate average for different users in sql server. The following query used to do this:



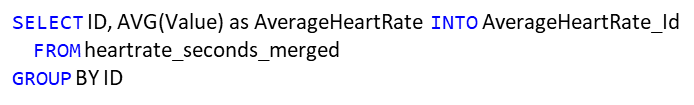
Then the result was imported to a excel worksheet for further analysis.

**Hourly average heart rate:** Calculated hourly average of heart rate for different users with the following query:



Result then imported to the excel worksheet.

**Average heart rate:** With following query calculated user wise average of heart rate”



**Date wise Sleep Minutes:** This was already there we just need to sort on the ID and Date combination. Also created a pivot table to help us visualize it.

**Average Sleep Minutes:** Calculated user-wise average sleep minutes using power query. This was done to comparare sleeping behaviour of different users.

**5. Share/Visualization**

**Trends Identified**:

1. **Daily Activity Tracking Patterns**: A large percentage of participants track daily steps, total distance, and calories burned. The activity tracking chart shows that the users tracking Total Steps metrics are also tracking Total Distance and Calories burned. The line chart clearly shows this pattern as lines for all these three metrics are overlapping each other, which is re-confirmed by the bar chart as well. This indicates that all these three metrics are linked. So, if a user is tracking Total Steps, then Total Distance and Calories burned are also tracked by default.
2. **Hourly Activity Patterns:** Hourly Steps charts show that people are more active between 8 AM in the morning to 8 PM in the evening, whereas less active before and after this period. But both hourly charts show there is a sharp dip in steps between 2PM and 4PM.
3. **Co-relation between metrics:** Steps vs Calories and Distance vs Calories charts show a clear relation between the metrics. The more steps people add the more calorie is burned, similarly the more distance they cover the more calorie is bured.
4. **Feature Usage Pattern:** The analysis shows that more than 90% of users are using Total Steps, Total Distance and Calorie feature where as only 40% are tracking sleep and 33% tracking heart rate. The users grouping based on the usage of different features also indicates that most of the users belong to the power user group for Total Steps metrics means they are using this metrics almost daily, where as this number is smaller for Sleep and Heart Rate metrics.
5. **Sleep Patterns**: Not many of the participants tracked sleep data, The Date vs Sleep Minutes line chart for multiple users shows break in line for many users, that means many of users tracking Sleep Metrics are not using this feature regularly. Also, there is a lot of variations in sleep minutes.
6. **Heart Rate Tracking**: Only a few participants are using this feature only 14 out of 33. Date wise chart show that few of them are consistently tracking heart rate whereas few are occationally. This chart also indicates that there is only a small variation in the daily average of this metrics across the logged period. But the hourly average chart shows a lot of variation in heart rate, this may be because of intensity of activenss during the day. But the third chart comparing average heart rate between the users indicates a big variation in the metrics.

**Visualizations and Key Findings**

1. **Line and Bar Chart of Activity Tracing Count**:

Clearly indicates that Total Steps, Total Steps and Calories burned metrics are inter connected by default.

1. **Line and Bar Chart of Steps Over Time**:

Show a clear-peaks during or arround working hours of the day, indicating user activity pattern.

1. **Line Charts for Sleep Tracking**:

Highlights major variation in sleep duration for some participants, similarly the comparison of average sleep between the users also indicates this variation.

1. **Line Charts for Heart Rate**:

Marjority of the participants showing a deviation from the ideal heart rate, which provides an opportunity for bellabeat to highlight this feature to make people aware for heart health.

**Key Insights**:

* **Most Popular Metrics**: Steps, distance, and calories are the most tracked metrics, showing high user engagement in physical activity monitoring.
* **Underutilized Features**: Sleep and Heart rate tracking are used less frequently, which presents an opportunity for Bellabeat to encourage more users to engage with heart health and sleep tracking features.

**6. High-Level Recommendations for Bellabeat Marketing Strategy**

Based on the data analysis, I recommend the following high-level strategies for Bellabeat’s marketing team:

**1. Emphasize the Popular Metrics (Steps, Distance, Calories)**

* **Leverage Popularity**: Since Total Steps, Total Distance, and Calories Burned are the most tracked metrics, Bellabeat should focus marketing on these features, emphasizing how their device tracks these metrics effectively.
* **Highlight Health Impact**: Promote these metrics as vital indicators of daily physical activity and overall fitness. Showcase success stories or case studies on how consistent activity monitoring has improved users' health.

**2. Promote Sleep Tracking Features**

* **Educational Campaigns**: Since sleep tracking is underutilized, Bellabeat can launch educational campaigns explaining the importance of quality sleep in overall well-being. Use content like blog posts, videos, and social media posts to demonstrate how sleep data can help users improve their sleep habits and mental health.
* **Personalized Insights**: Offer personalized insights or notifications that remind users to track their sleep regularly.

**3. Increase Focus on Heart Rate Tracking**

* **Heart Health Awareness Campaign**: Many users are not actively tracking heart rate. Create campaigns around heart health awareness, particularly highlighting the connection between daily activity and heart rate patterns. Bellabeat can highlight how irregular heart rate patterns can signal potential health risks, promoting proactive health monitoring.
* **Push Notifications & Reminders**: Encourage users to track their heart rate daily by providing reminders, especially during periods of physical activity. Educating users about the importance of resting heart rate and active heart rate can be part of this strategy.

**4. Personalized Health Dashboard**

* Promote Bellabeat’s ability to provide a personalized health dashboard where users can see trends across multiple dimensions of their health (steps, calories, heart rate, sleep) and how each of these affects their overall well-being.

**5. Target Specific User Groups with Personalized Campaigns**

* **Power Users**: Since there are many "power users" who track steps almost daily, Bellabeat can focus on engaging this group further with advanced fitness tracking features, such as goal-setting, insights into activity patterns, and suggestions for improving fitness levels.
* **Non-Power Users (Sleep & Heart Rate)**: For users who are less active in tracking sleep and heart rate, target them with campaigns that highlight the benefits of these features. Offer personalized insights based on minimal engagement (e.g., “You haven’t tracked sleep in a while, here’s why it matters!”).

1. **Product Features and Innovation Focus**

* **Smart Alerts**: Bellabeat could introduce smart alerts or personalized recommendations based on users’ real-time activity data. For instance, if someone has been inactive for too long or hasn’t tracked their heart rate in a few days, a gentle reminder could encourage better engagement

**9. Address the 2-4 PM Activity Dip**

* **Boost Midday Activity**: The sharp dip in activity between 2 PM and 4 PM suggests a common inactive period. Bellabeat can launch app features that encourage users to stay active during this time. For example, reminders or notifications encouraging a short walk, stretching exercises, or hydration during this period.

**Analysis by: Anil Jain**