**Bellabeat – How can wellness technology company play it smart?**

**Introduction**

Urška Sršen and Sando Mur founded Bellabeat, a high-tech company that manufactures health-focused smart products. Sršen used her background as an artist to develop beautifully designed technology that informs and inspires women around the world. Collecting data on activity, sleep, stress, and reproductive health has allowed Bellabeat to empower women with knowledge about their own health and habits. Since it was founded in 2013, Bellabeat has grown rapidly and quickly positioned itself as a tech-driven wellness company for women. By 2016, Bellabeat had opened offices around the world and launched multiple products. Bellabeat products became available through a growing number of online retailers in addition to their own e-commerce channel on their website. The company has invested in traditional advertising media, such as radio, out-of-home billboards, print, and television, but focuses on digital marketing extensively. Bellabeat invests year-round in Google Search, maintaining active Facebook and Instagram pages, and consistently engages consumers on Twitter. Additionally, Bellabeat runs video ads on YouTube and display ads on the Google Display Network to support campaigns around key marketing dates.

**Key Stakeholders**

Urška Sršen: Bellabeat’s cofounder and Chief Creative Officer Sando Mur: Mathematician, Bellabeat’s cofounder and key member of the Bellabeat executive team Bellabeat marketing analytics team: A team of data analysts guiding Bellabeat’s marketing strategy.

Using the Case Study Roadmap as a guide, this analysis will follow the steps of the data analysis process: **Ask, Prepare, Process, Analyze, Share, and Act.**

# ****Step 1: Ask****

### **Guiding questions**

* What is the problem you are trying to solve?  
  Sršen knows that an analysis of Bellabeat’s available consumer data would reveal more opportunities for growth. She has asked the marketing analytics team to focus on a Bellabeat product and Analyze smart device usage data in order to gain insight ]into how people are already using their smart devices.
* How can your insights drive business decisions?b  
  Using this information, would produce high-level recommendations on how these trends can inform Bellabeat marketing strategy.
* Business Task: Analyze FitBit fitness tracker data to gain insights into how consumers are using the FitBit app and discover trends for Bellabeat marketing strategy.

### **Business questions:**

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 marketing strategy?**

**Deliverables:**

1. **A clear summary of the business task.**
2. **A description of all data sources used.**
3. **Documentation of any cleaning and manipulating of data.**
4. **A summary of our analysis.**
5. **Support visualization and key findings.**
6. **Top recommendations based on our analysis.**

**Step 2: Prepare**

I prepared the data for analysis using the “Case Study Roadmap” as a guide:

**Guiding Questions:**

* Where is your data stored?  
  Data is publicly available on Kaggle: [FitBit Fitness Tracker Data](https://www.kaggle.com/arashnic/fitbit).
* How is the data organized?  
  Data is stored in 18 csv files.
* Are there issues with bias or credibility in this data? Does your data ROCCC?  
  This Kaggle data set contains personal fitness tracker from 30 FitBit users. They were generated by respondents to a distributed survey via ***Amazon Mechanical Turk*** *between 12.04.2016-12.05.2016.*
* How are you addressing licensing, privacy, security, and accessibility?  
  30 eligible Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring. A good data source is ROCCC which stands for Reliable, Original, Comprehensive, Current, and Cited:

# Reliable — LOW — **Not reliable as it only has 30 respondents** Original — LOW — **Third party provider (**Amazon Mechanical Turk**)** Comprehensive — MED — **Parameter’s match most of Bellabeat product´s parameters** Current — LOW — **Data is 5 years old and may not be relevant** Cited — LOW — **Data collected from third party, hence unknown** “Overall, this dataset is considered “bad quality data” and it **is not recommended** to produce business recommendations based on this data”

* **How did you verify the data’s integrity?**  
  As data is collected in a survey, we are unable to ascertain its integrity or accuracy.
* **How does it help you answer your question?**  
  This data explores smart device users’ daily habits. It includes information about daily activity, steps, sleep habits and heart rate, that can be used to explore users’ habits and find some trends.
* **Are there any problems with the data?**  
  Data was collected **5 years ago** in 2016. Users’ daily activity, fitness and sleeping habits, diet and food consumption may have changed since then. Data may not be timely or relevant. Sample size of 30 FitBit users is not representative of the entire fitness population.

**Step 3: Process**

Process the data by cleaning and ensuring that it is correct, relevant, complete and free of error and outlier.

There are many different CSV files in this dataset, but I decided to concentrate in two CSVs: “dailyActivity\_merged.csv” and “sleepDay\_merged.csv”

### **Guiding questions:**

* **What tools are you choosing and why?**  
  I am using R for data cleaning, transformation and visualization. R provides an accessible language to organize, modify, clean data frames and create insightful data visualizations.
* **Have you ensured your data’s integrity?**  
  Data was collected in a survey, so I am unable to ascertain its integrity or accuracy.
* **What steps have you taken to ensure that your data is clean?**  
  I removed duplicates, NA´s, cleaned, formatted and converted date formats.
* **How can you verify that your data is clean and ready to Analyze?**  
  I checked all cleaning operations above and assured accuracy.
* **Have you documented your cleaning process so you can review and share those results?**  
  yes, I used R Markdown and Code Chunks.

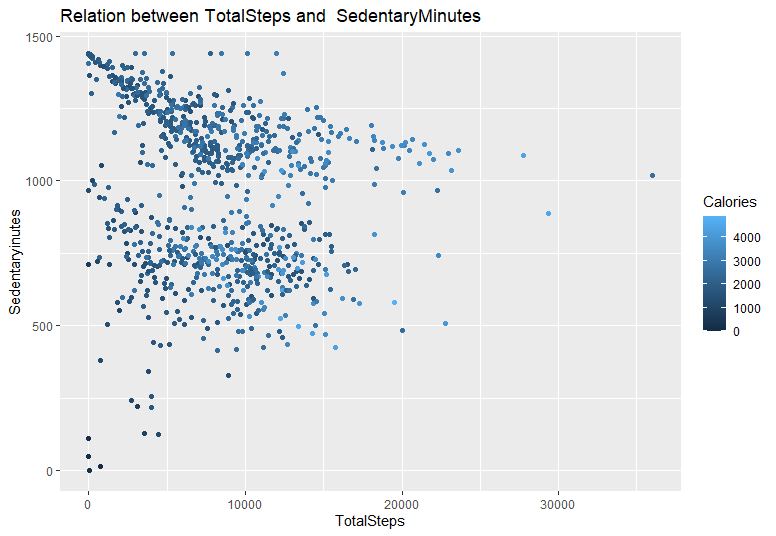
# ****Step 4: Analyze****

Now that the data is stored appropriately and has been prepared for analysis, we can start putting it to work.

## **Plotting a few explorations on user activities:**

What’s the relationship between steps taken in a day and sedentary minutes? How could this help inform the customer segments that we can market to?

* **we are observing a negative relationship between total steps taken and sedentary minutes.**
* **We can also note that sedentary time is not necessarily related to calories burned. Nm,hkjlbl v**
* **We also see that calories generally trend positively with total steps taking.**



### **What’s the relationship between minutes asleep and time in bed? You might expect it to be almost completely linear - are there any unexpected trends?**

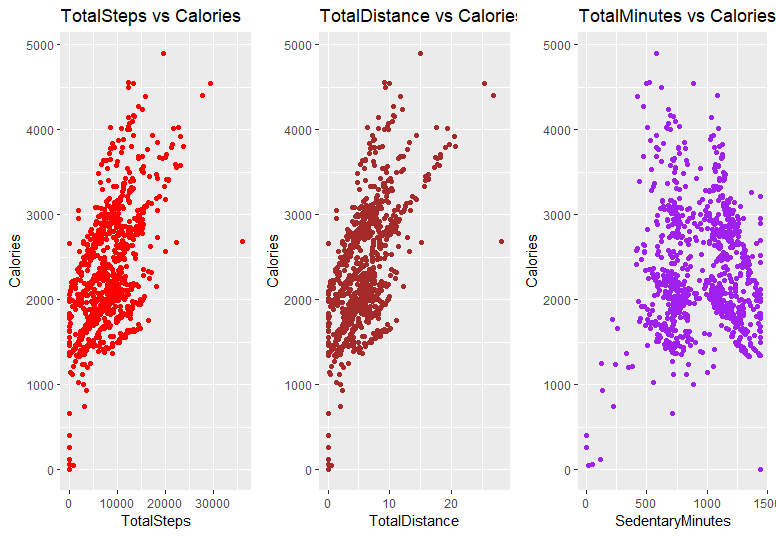
* **Observing the next graph, We find some outliers. Some of these data points spent a lot of time in bed, but didn’t actually sleep, and then a small batch that slept a whole bunch and spent more time in bed.**

### **What could these trends tell you about how to help market this product?**

* **We could definitely market consumers to use their watch to better monitor their time in bed against their sleep time.** ## Or areas where you might want to explore further?
* **I wonder which days of week users often spend more time logging? How does this relates to the sedentary minutes??**

### **Interpreting statistical findings:**

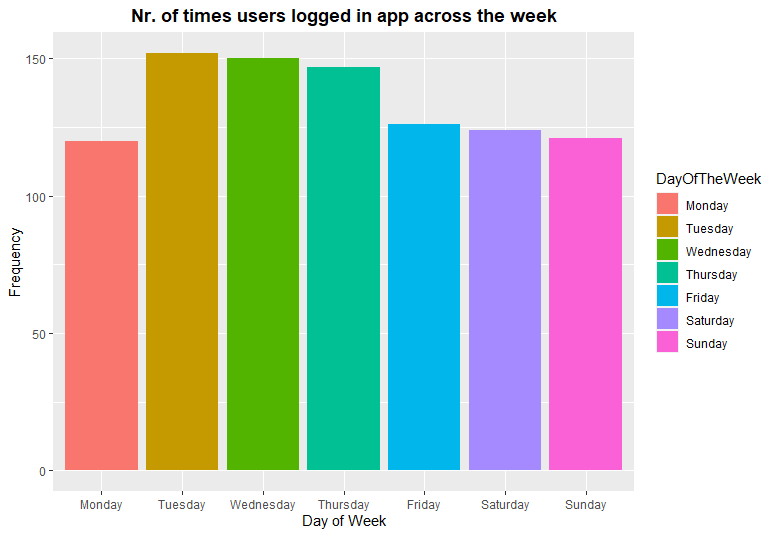
* On average, users logged 7,652 steps or 5.4km which is not adequate. As recommended by OMS, an adult female has to aim at least 10,000 steps or 8km per day to benefit from general health, weight loss and fitness improvement.
* Sedentary users are the majority logging on average 990.4 minutes or 20 hours making up 81% of total average minutes.
* Noting that average calories burned is 2,308 calories equivalent to 0.3 Kg. Could not interpret into detail as calories burned depend on several factors such as the age, weight, daily tasks, exercise, hormones and daily calorie intake.



* By quickly analyzing these 3 scenarios, we observe “TotalDistance” is more closely related to “Calories”.

# ****Step 5: Share****

## **Plot Nr. of times users logged in app across the week**



### **In this bar chart, we are looking at the frequency of FitBit app usage in terms of days of the week.**

* **We discovered that users prefer or remember (giving them the doubt of benefit that they forgotten) to track their activity on the app during midweek from Tuesday to Friday.**
* **Noting that the frequency dropped on Friday and continue on weekends and Monday.**

## **Plot Calories burned for every step taken:**



### **Calories burned for every step taken**

The scatter plot is showing:

* We have a positive correlation, meaning that Calories are burned for every step taken
* We observed that intensity of calories burned increase when users are at the range of > 0 to 15,000 steps with calories burn rate cooling down from 15,000 steps onwards.

Noted a few outliers: \* Zero steps with zero to minimal calories burned. \* 1 observation of > 35,000 steps with < 3,000 calories burned. \* Deduced that outliers could be due to natural variation of data, change in user’s usage or errors in data collection (i.e. miscalculations, data contamination or human error).

## **Plot Calories burned for every Hour logged:**



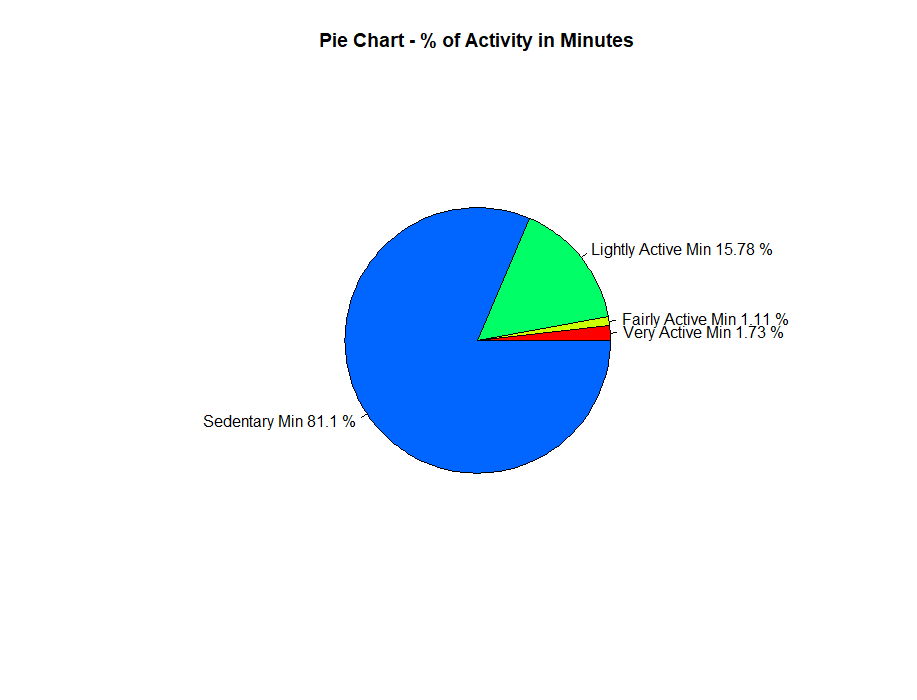
### **Calories burned for every hour logged**

The scatter plot is showing:

* A weak positive correlation whereby the increase of hours logged does not translate to more calories being burned. That is largely due to the average sedentary hours (blue line) plotted at the 16 to 17 hours range.

Again, we can see a few outliers:  
 **The same zero value outliers.** **An unusual red dot at the 24 hours with zero calorie burned which may be due to the same reasons as above.**

## **Plot % of Activity in Minutes**



### **Percentage of Activity in Minutes**

As seen from the pie chart:

* **Sedentary minutes takes the biggest slice at 81.10%.**
* **This indicates that users are using the FitBit app to log daily activities such as daily commute, inactive movements (moving from one spot to another) or running errands.**
* **The App is rarely used to track fitness (ie. running) according to the minor percentage of Fairly Active Activity (1.11%) and Very Active Activity (1.73%).**
* **This is highly discouraging as FitBit app was developed to encourage fitness.**

**Step 6: Act**

In the final step, we will be delivering our insights and providing recommendations based on our analysis. Here, we revisit our business questions and share with you our high-level business recommendations.

### **1. What are some trends in smart device usage?**

**Majority of users (81.10%) are using the FitBit app to track sedentary activities and not using it for tracking their health habits. Users prefer to track their activities during weekdays as compared to weekends - perhaps because they spend more time outside on weekdays and stay in on weekends. Data also tells us that most users log in their calories, steps taken, etc, and fewer log their sleep data.**

### **2. How could these trends apply to Bellabeat customers?**

**Both companies develop products focused on providing women with their health, habit and fitness data and encouraging them to understand their current habits and make healthy decisions. These common trends surrounding health and fitness can very well be applied to Bellabeat customers. Bellabeat could easily market these types of costumers by telling them smart-devices could help them start their journey by measuring how much they’re moving and how these moments of activity would benefit them to live longer!**

### **3. How could these trends help influence Bellabeat marketing strategy?**

**It is well documented that moderate-to-vigorous physical activity is protective against chronic disease. Conversely, emerging evidence indicates the deleterious effects of prolonged sitting, so in a need to change both behaviours, self-monitoring of behavior is one of the most robust behaviour-change techniques available. Bellabeat marketing team can encourage users by educating and equipping them with knowledge about fitness benefits, suggest different types of exercise (ie. simple 10 minutes exercise on weekday and a more intense exercise on weekends) and calories intake and burnt rate information on the Bellabeat app. On weekends, Bellabeat app can also prompt notification to encourage users to exercise. By marketing these devices to consumers, Bellabeat provides a unique opportunity for individuals to change their behavior, become more physically active and increase their life expectancy.**