**Capstone Project – Bellabeat Case Study**

**Company Details**

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. 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 analyse smart device usage data in order to gain insight into how people are already using their smart devices. Then, using this information, she would like high-level recommendations for how these trends can inform Bellabeat marketing strategy.

So, we have to follow the complete data cycle to get the desired results:

1. **Ask**

Ask the right questions around and understand the problem statement. Here the company wants to understand the customers behaviour pattern to smart devices and then apply those to Bellabeat marketing strategy and acquire more customers to the subscription plan.

1. **Prepare**

Use the dataset provided to us and prepare it to be utilised. This Kaggle data set contains personal fitness tracker from thirty Fitbit users. Thirty eligible Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring. It includes information about daily activity, steps, and heart rate that can be used to explore users’ habits. We have to use this, sort it, filter it, check it for bias and get it ready for the next phase.

1. **Process**

Now comes the nitty gritty part, cleaning the dataset, removing outliers, missing values, null values, errors in the dataset and get the unbiased, clean and complete dataset to analyse. We use R language to this.

First, we analyse which data we want, so out of all the data we use:

* dailyActivity\_merged.csv
* dailyCalories\_merged.csv
* dailyIntensities\_merged.csv
* minuteMETsNarrow\_merged.csv
* sleepDay\_merged.csv
* dailySteps\_merged.csv
* heartrate\_seconds\_merged.csv.

We check out the glimpse of each data to check the data types, any errors in it. We find some irregularities in it: -

* Date column in minuteMET and sleepDay is char data type which needs to be changed to numeric data type to perform the calculations.
* The ID and date field is common for all and we can connect the sheets based on these columns when needed.
* The dailycalories, dailyintensities, dailysteps sheet info is already there on the dailyactivity sheet so we exclude these sheets from the dataset.

Now we check for null values using the summary function and find out weight data has null values and we remove them.

We were told that the data is for 30 users so we cross check this information.

We verify that and find out that the data in heartrate is not even 50% accurate so we exclude it from the results.

Now we check for duplicates and we find 3 duplicates in sleepDay data so we remove them.

We change the date column from char to datetime data type for all tables.

We change the minute METs to daily METs.

1. **Analyze**

We build different relationships to analyse the data we just processed above.

* Relationship between Total Steps and Calories burnt
* Relationship between Total Distance and Calories burnt
* Relationship between Total Steps and METs
* Relationship between Total Distance and METs
* Users who sleep less than 7 hours
* Users who sleep more than 7 hours
* Relationship between calories burnt in 5-week interval

**PROBLEMS AND ASSUMPTIONS WITH THE DATASET**

(1) There is no information about the age, gender, height and weight of the user. So, assuming that the users age was in the similar range.

(2) Heartrate\_seconds table had less than 50% of data. Henceforth couldn’t use that assumed it won’t affect the analysis.

(3) As per the data dictionary the food\_logs and floor\_logs data were supposed to be present which were missing. Assuming that won’t affect the analysis.

(4) The minuteMETsNarrow\_merged data set has data for 33 and not 30 users. So assumed that a little high number wont effect.

(5) The data in the wide\_merged data sets were similar to the narrow\_merged data sets. Hence didn’t used them for the analysis.

**Summary Report**

Number of steps are directly proportional to the number of calories burnt.

Distance is directly proportional to the number of calories burnt.

Users are more energetic and active during the first week and then it becomes stagnant.

There are users who are sleeping less than 7 hours (not recommended)

The number of METs (Metabolic Rate) increase as the number of Total steps increase. Metabolic rate is the rate of energy expended per unit of time. Hence, we can say that the number of calories burn increase as the METs increases.

**Recommendations**

As per the analysis I would suggest the usage of leaf which can be worn in multiple ways and can get more precise data about the user other than any other products which have a particular time or functionality.

Secondly, Product should capture the age, gender, weight and height as soon as the user wears it in first place.

Thirdly, a notification should be generated with suggestions when user is having less hours or more hours of sleep.

Fourth, Users who are staying consistent with their activity should be praised with “congratulation” notifications and also a summary on how they are staying active and doing good for their health. It will motivate them more.

Fifth, Users who are not active or have stopped burning calories should be notified with health benefits and advantages of workouts.

Sixth, Users who are regularly active and using Bellabeat products should get some discount on their newly launched products.

Food intake, activity date and time, sleep hours must be generated automatically by the product and not manually by the user that way we can avoid human errors.

Product should have regular reminders when user have stopped working or eating unhealthy food or sleeping for less hours as it may affect their health.

**NOTE:- All the working code has been shown in the R file attached separately with this.**