**Case Study: How Can a Wellness Technology Company Play It Smart?**

**Introduction**

1. This is a data analysis case study on Bellabeat, a high-tech manufacturer of health-focused products for women. Bellabeat is a successful small company, but they have the potential to become a larger player in the global **smart device** market. Urška Sršen and Sando Mur, cofounder and Chief Creative of Bellabeat, believes that analyzing smart device fitness data could help unlock new growth oppounities for the company. The company has 5 focus products: bellabeat app, leaf, time, spring and bellabeat membership.

Here we are going to focus on one of Bellabeat’s products and analyze smart device data to gain insight into how consumers are using their smart devices. The insights discover from the analysis will then help guide marketing strategy for the company. This case study will focus on the analysis process: ask, prepare, process, analysis, share and act.

Ask

**BUSINESS TASK: Analyze Fitbit data to gain insight and help guide marketing strategy for Bellabeat to grow as a global player.**

Primary stakeholders: Urška Sršen and Sando Mur, executive team members.

Secondary stakeholders: Bellabeat marketing analytics team.

**Prepare**

**FitBit Fitness Tracker Data** (CC0: Public Domain, dataset made available through Mobius):

https://www.kaggle.com/arashnic/fitbit

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 user’s habits.

The dataset has 18 CSV. The data also follow a ROCCC approach:

* Reliability: The data is from 30 FitBit users who consented to the submission of personal tracker data and generated by from a distributed survey via Amazon Mechanical Turk.
* Original: The data is from 30 FitBit users who consented to the submission of personal tracker data via Amazon Mechanical Turk.
* Comprehensive: Data minute-level output for physical activity, heart rate, and sleep monitoring. While the data tracks many factors in the user activity and sleep, but the sample size is small and most data is recorded during certain days of the week.
* Current: Data is from March 2016 to May 2016. Data is not current so the users habit may be different now.
* Cited: Unknown.

The dataset has limitations:

* Only 30 user data is available. The central limit theorem general rule of n≥30 applies and we can use the t test for statstic reference. However, a larger sample size is preferred for the analysis.
* Upon further investigation with n\_distinct() to check for unique user Id, the set has 33 user data from daily activity, 24 from sleep and only 8 from weight. There are 3 extra users and some users did not record their data for tracking daily activity and sleep.
* For the 8-user data for weight, 5 users manually entered their weight and 3 recorded via a connected wifi device (eg: wifi scale).
* Most data is recorded from Tuesday to Thursday, which may not be comprehensive enough to form an accurate analysis.

**Process**

* Choosing the appropriate tools to use for data cleaning process. ( R programming)
* Check the data for errors.
* Transform the data so we can work with it effectively.
* Document the cleaning and manipulation process.

Examine the data, check for NA, and remove duplicates for three main tables: daily\_activity, sleep\_day and weight:

* dim(sleep\_day)
* sum(is.na(sleep\_day))
* sum(duplicated(sleep\_day))
* sleep\_day <- sleep\_day[!duplicated(sleep\_day), ]