**Introduction**

Bellabeat is a high tech manufacturer of fitness and health focused products for women. They offer a variety of technologies, including jewelry-like wellness trackers, a smart-watch, and smart water bottles. These products track user data relating to their activity, sleep, stress, menstrual cycle, and mindfulness habits, which is readily available on their free mobile app.

Currently, Bellabeat primarily focuses on digital marketing, including Facebook and X(Twitter), and runs ads on Youtube and the Google Display Network. While this has proven successful in the past, Bellabeat would like to explore the data from user of other smart device companies to determine whether they should modify their marketing strategy. Perhaps smart watches are not popular amoungst the older generation, in which case, Bellabeat should shift its focus towards more traditional forms of marketing where the target audience is that older generation……………………….

This study focuses on analyzing those trends, and applying what we learn to inform future Bellabeat marketing strategy for its smart watch.

**Ask**

Business Task – Analyze trends in non-Bellabeat smart watch usage to inform marketing strategy for Bellabeat’s smart watch

Key Stakeholders

Urška Sršen - Bellabeat’s cofounder and Chief Creative Officer

Sando Mur - Mathematician and Bellabeat’s cofounder

**Prepare**

**Where is your data stored?**

This data was made publicly available on Kaggle

**How is the data organized?**

The data is stored in csv format, and split across many spreadsheets. Each spreadsheet tracks its own individual statistic, such as daily calories burned, daily steps, weight, etc..

**Is it in long or wide format?**

The data is stored in long format

**Are there issues with bias or credibility in this data?**

The majority of the data recorded comes from the device itself, therefore with respect to that data, there is little bias or credibility. However, in the calories burned datasets, there are entries where the data reads 0 calories burned. This data is not credible, as it is likely the user wasn’t wearing their device during that day and should be omitted during cleaning. Additionally, the weight log data includes self reported data. While this shouldn’t be a big concern, as there is no incentive for users to intentionally misreport their weight, it is something that should be taken into consideration.

**Does your data ROCCC?**

Reliable –

Original – the data is original, as it is first party data collected directly from FitBit users with their consent

Comprehensive

Current – the data is from 7 years ago and not current. However, for the educational purposes of this case study, this is okay

Cited -

**How are you addressing licensing, privacy, security, and accessibility?**

**How did you verify the data’s integrity?**

**How does it help you answer your question?**

**Are there any problems with the data?**

dailyActivity\_merged- tracks total steps and distance, and seperates that into very, moderately, light and active distances as well as the minutes for each of those categories as well, including sedentary minutes. It also tracks calories burned in total. Each day for the whole month

dailyCalories\_merged- tracks calories burned for each day for the whole month

dailyIntensities\_merged- tracks minutes and distances across four subcategories: very, moderately, lightly, and sedentary activity levels each day for the whole month

dailySteps\_merged- tracks daily steps for the whole month

heartrate\_seconds\_merged – tracks heartrate multiple times every minute of every day for the whole month

hourlyCalories\_merged – tracks calories burned every hour of every day for the whole month

hourlyIntensities\_merged- tracks total intensity for every hour and average intensity per minute every day for the whole month

hourlySteps\_merged- tracks total steps for every hour of every day for the whole month

minuteCaloriesNarrow\_merged- tracks caloires burned every minute of every hour of every day for the whole month

minuteCaloriesWide\_merged- minuteCaloriesNarrow but in wider format for every hour

minuteIntensitiesNarrow\_merged- measurs intensity for every minute across the whole month

minuteIntensitiesWide\_merged- minuteIntensitiesNarrow\_merged but in wider format for every hour

minuteMETsNarrow\_merged- tracks the ratio of the energy used vs at rest every minute

minuteSleep\_merged- no idea

minuteStepsNarrow\_merged- steps for every minute

minuteStepsWide\_merged- steps for every minute, wide format

sleepDay\_merged- measure the number of sleeps, total minutes slept, and total time in bed for each day

weightLogInfo\_merged- tracks weight of X users at various times in pounds and kg, as well as their BMI, and whether they manually reported their weight

add the number of users

**Prepare**

For this study, we will only use these datasets

dailyActivity\_merged

sleepDay\_merged

libraries used

'readxl', 'Rcpp'

install.packages("magrittr") # package installations are only needed the first time you use it

install.packages("dplyr") # alternative installation of the %>%

library(magrittr) # needs to be run every time you start R and want to use %>%

library(dplyr) # alternatively, this also loads %>%

Import daily activity

library(readxl)

dailyActivity\_merged <- read\_excel("Development/Fitabase Data 4.12.16-5.12.16/dailyActivity\_merged.xlsx")

View(dailyActivity\_merged)

Import sleep day

sleepDay\_merged <- read\_excel("Development/Fitabase Data 4.12.16-5.12.16/sleepDay\_merged.xlsx")

View(sleepDay\_merged)

Convert sedentary distances to 0 – you cant have any positive sedentary distance