CS614 – Prof. Berardi
Module 1

This module will use data collected from a sample of users of a popular app for meditation and wellness. It consists of all app activity between 2018 and 2020 for 1,000 users, grouped by sessions, which are defined as an instance of app usage. The data includes the following variables:

ID: user ID
start_date: date on which the user's account was created
date: date of the app session
duration_in_sec: duration of the app session
session_type: one of eight categories into which each session is classified (body, breathe, masterclass, meditation, mood_check_in, music, sleep, soundscape
program_title: title of the session program
selected_mood: one of 13 moods that participants can select when completing a 'mood_check_in' session. 0 otherwise
model_estimated_age: Estimated age of participant according to proprietary app algorithm
model_estimated_gender: Estimated gender of participant according to proprietary app algorithm

This data was collected immediately during the rollout of the *Mood Tracking* feature of the app, which prompts participants to periodically log in and record their mood by selecting an emoji+description that best represents how they are feeling. Therefore, the analysis for this module will focus on this feature. Work with your partner to complete the following tasks:

A. Defined Research Questions

a) Calculate summary statistics for the use of the *Mood Tracking* feature by answering the following questions. i.) What proportion of total sessions were associated with this feature? ii.) How many users used this feature at least once?
b) For those users who have used the *Mood Tracking* feature at least once, create a histogram showing the number of mood check-ins for each user. If you found it necessary to make any changes to your figure to make it readable, please detail in your writeup. Also include the mean, median, and standard deviation of your values.
c) Create a separate bar graph for the reported mood for male and female users (ignoring the 'no data' and 'uncertain' gender estimates). Compare and contrast these graphs.
d) Examine the relationship between reported mood and age by calculating the average age of the user for each reported mood. What two moods are associated with older participants? Two moods associated with younger participants? Do the ages for the four moods you've identified differ by male vs. female gender?

B. Student-Determined Research Question

A second research question related to *Mood Tracking* should be developed and executed by each group. Address the research question by creating data a visualization, quantification and/or statistical test as

needed. In your write-up, please clearly describe the question you have decided to pursue and the methodology you used for this task. This is your chance to demonstrate your creativity re: data analysis!

Here are some example research questions to get you thinking.

- What type of sessions immediately proceeds or follows different reported moods?
- Do certain moods precede or follow other moods?
- When in their enrollment do people use the *Mood Tracking*? Immediately after signing up? During periods of high/low app usage?
- Related to the previous question, does *Mood Tracking* appear to occur in bunches or as isolated incidents.
- What is the probability of using the *Mood Tracking* feature and/or reporting different types of moods at different times of day?

For these or other questions you come up with, you may find it useful to group moods by valence (e.g. positive vs negative)….or not.

C. Wishlist

If you could waive a magic wand and have one more variable for this data set, what would it be? Why?

Answer Format

Please put all answers in R Markdown (html or pdf), with one submission per group. (You are grouped as teams in Canvas.) I recognize that it is not possible to produce a polished writeup in the time allotted for this assignment, but do your best to organize your work clearly and make it easy to follow. If you are not asked for a comparison or explanation for a given section, just give the answer and/or figure that was requested (along with he code required to produce it). Your answer for Section C does not need to be long – 1 to 3 sentences should suffice.