NeuroFlow Data Associate Take-Home Project

Goals of this Project

We are asking you to complete this take-home project because we are impressed with you, but would like to see how your skills will translate to our specific context. Please feel free to do any research or preparation you wish, but do not plan to spend more than a few hours on this project. We're not seeking the perfect solution, just trying to get a deeper understanding of your technical skills and how you approach new problems.

Part 1

Our Problem

For this project, we'd like you to consider data from what we call "Subjective Metrics", which are mood, stress, rumination, and sleep tracking. In our iOS and Android apps, we ask users to rate these four elements on a scale of 1-5, with 1 being negative and 5 being positive.

Every morning at 8am, we send all of our patients a notification to rate their mood and their sleep on a circular scale. The full five options are Awful, Bad, Okay, Good, Great! Awful is 1 and Great is 5.

Users can also rate their rumination ("Are you having trouble letting go of something that happened today?") with options "Not at all", "Slightly", "Moderately", "Very", "Extremely" and they can rate their stress ("Are you stressed about something that might happen in the near future?") on the same scale.

The clinical purpose of these activities is to help patients track their sleep, mood, rumination, and stress over time to see if therapy is making a difference. The provider can see this data too, building the basis of a conversation they can have together.

We currently have the problem of not being able to well visualize progress for these four metrics to mental health providers and their patients.

Your Solution

The attached zip contains four zip files in CSV format, one for each of these four metrics we track. The sample data is from one internal team user.

For each line in each file, the first value represents a timestamp in seconds representing the time the measurement was made, and the second value represents the value of the response.

Given the information you have and any light research you'd like to do, what insights can you draw. What are 2-3 additional pieces of information that would be important to monitor? Feel free to use open source tools, such as Python or R to analyze the data set. We are not looking for a production-ready code, but we will assess both your approach to visualization and your technical abilities.

When complete, please send us a Github link or other access to a repo to review what you've done, with any necessary instructions on how to run your code locally. Let us know if you have any questions as you work through this assignment.

Part 2

Our Problem

We'd like to see how you write SQL for a given question. Often our business counterparts will ask us for a quick query to answer a question. In this case, the question is: how many users completed an exercise in their first month?

As context, our platform is a tool that clinical providers use to assign different kinds of exercises to their patients. Typical exercises may be to write a journal entry, complete a meditation session or fill out a clinically validated questionnaire. The purpose of these exercises are to help their patient stay engaged and ultimately feel better faster, so the earlier this feature becomes sticky for the patients, the longer they'll stay engaged.

We want to identify patterns in patient behavior with respect to exercises, compare this over time, and find the driving factors for their exercise completion rates.

Your Solution

Assume you have two tables in our company's database:

- 'users' table, with columns 'users', 'created_at'
- 'exercises' table, with columns 'questionnaire_id', 'user_id', 'exercise_completion_date'

Write a single SQL query that breaks up the users based on the month that they signed up (their cohort month), and determines the percentage of users that have a completed exercise in their first month for each cohort over the months (e.g., the 2018 January cohort has x% of users completing an exercise in their first month, 2018 February cohort has x% of users completing an exercise in their first month, etc.).

When complete, please send us a Github link, a sql file, or something similar to review what you've done.

Again, let us know if you have any questions as you work through this assignment.