# Data Challenge

The goal of this data challenge is to see your skills in programming with data in action on a data set similar in nature to real data you might end up analyzing if you joined our team. We recognize that your time is valuable and have crafted this challenge to be something that can be completed within an hour. We aren't imposing a strict time limit, but we think people who are a great fit for this role will be able to complete the challenge in about an hour.

Use the programming language of your choosing -- please share the code you used to solve as well as a short write-up for the questions.

## The Data

We have a feature where members can chat asynchronously with their coach. For this challenge, we have a small sample data set containing information on when members and coaches are sending messages.

The messages between a given member and their coach happen within a specific "chat channel". We have a chat messages table that contains a record of every message sent, and links the message to the user that sent it via the user_id. The channel_id shows what channel the message was part of. So if a member sends a message to a coach, and then the coach responds, those messages would be part of the same chat channel. The chat_users table has information about the users -- it identifies whether a user was a member or a coach. You can assume that every chat channel has only two users in it, one member and one coach (but feel free to validate that yourself!).

**Schemas:**

chat messages
https://drive.google.com/drive/folders/0AN16eM0n24gnUk9PVA

| Column name | Data type | Description |
|---|---|---|
| user_id | string | Unique identifier for each user, whether member or coach |
| channel_id | string | Unique identifier for each "chat channel", which stores the conversations between a particular member-coach pair |
| message_time | timestamp | The time a message was sent in UTC time. |
| message_id | int | Unique identifier for each message |

Chat_users
https://drive.google.com/file/d/1rFh28r72tcHTwcuN25qbDubb0weDnBNu/view?usp=sharing

| Column name | Data type | Description |
|---|---|---|
| user_id | string | Unique identifier for each user, whether member or coach |
| chat_role | string | The type of user -- member or a coach |

## The Questions

Part 1: What is the average and median time between a 'member' sending a message and a coach responding to that message?

Part 2: If choosing between average and median response time, which do you think best captures how quickly messages are being responded to? What are some limitations with both approaches? What are some other metrics that one might consider for response times?

Part 3: Are there patterns to when members and coaches send messages by day of week or time of day? Create some plots and/or tables to show how message volume for each role varies by day of week and time of day.

### What's next?

Please enter your submission in the submission link provided from your recruiter. As the next step of our interview process, we'll chat about what you found in this data challenge to get more insight into how you think about solving problems.