**Instructions: Data Scientist Technical Evaluation (CAFÉ)**

Your task is to predict hospital readmissions for diabetes patients. We are interested in your approach to the problem, how you deal with problems that may arise, your ability to make accurate predictions, the quality of your code, how you assess the quality and predictive performance of your model, and your communication of the results, caveats, next steps, etc. to an audience that is technical but not necessarily expert in machine learning.

You have two hours to complete the challenge. You are free to work in any environment, but we are looking for someone with Python skills and would prefer you work in Python. There are two products you need to produce for this analysis:

1. An html notebook file that contains all of your code and results as well as a description of your process and the results.

2. A csv file that contains your predictions with row identifiers. The `training\_data.csv` file contains binary outcomes and should be used to train your model(s). The `test\_data.csv` file is from the same dataset, but has had the outcome variable removed. The csv file you produce should be titled "predictions.csv" and contain three columns with the names "encounter\_id" (which should match those in `test\_data.csv`), "predicted\_probability", and "predicted\_class".

You have two hours to complete the challenge. Please document your start and end times in your notebook. When you are finished, please email the results to [trevor.smith@healthcatalyst.com](mailto:trevor.smith@healthcatalyst.com), [michael.levy@healthcatalyst.com](mailto:michael.levy@healthcatalyst.com), and [michael.mastanduno@healthcatalyst.com](mailto:michael.mastanduno@healthcatalyst.com). Good luck!