Hi candidate:

Banks play a crucial role in market economies. They decide who can get finance and on what terms and can make or break investment decisions. For markets and society to function, individuals and companies need access to credit.

Credit scoring algorithms, which make a guess at the probability of default, are the method banks use to determine whether or not a loan should be granted. This assessment requires participants to improve on the state of the art in credit scoring, by predicting the probability that somebody will experience financial distress in the next two years.

The goal of this assessment is to build a model that borrowers can use to help make the best financial decisions.

Historical data are provided on over 250,000 borrowers. About the files:

- cs-test.csv – this is the test set you must use to generate the final results for your model
- cs-training.csv – this is the training set you are to use to build your model
- sampleEntry.csv – when you build your model and test it on cs-test.csv data, you'll need to generate results in this exact format which we will use to determine your model's predictive accuracy. **This is just an example file for your model output**. To be precise, your final results should reside on the first 2 columns , with column 1 (titled Id), containing the number of rows in ascending order (will follow the exact same order as in the cs-test.csv), and column 2 (titled Probability) being your probability of default for the customer on that respective row.
- Data Dictionary.xls – data dictionary for all fields in the training and test data set

On completion of this assessment, you will need to email back to the assessment invigilator:

- Code (Python or R only)
- Any relevant additional material to supplement the code, methodology, and analytics (docs, ppt, pdfs, notebooks, etc)
- Predictions on the cs-test.csv in the same format as the sampleEntry.csv file. Ideally in .csv format, but any text format would be fine.
- Feedback on your assessment experience so we can further improve the process

You will not be given any further help from the assessment invigilator on the content in this dataset and on these instructions. Part of the assessment is your ability to interpret what needs to be done given the task at hand.