**Breakage Model Exercise Brief**

The task is to conduct exploratory analysis and build a model that’ll attempt to predict payment breakage\* using the data provided.

The output should be a presentation explaining the approach and work done, and a well commented code/analysis.

Breakage occurs relatively infrequently and there is little signal within the noise. Therefore, it isn’t easy to predict, so don’t be too concerned if the model performance is not high.

**Expectations**

* There shouldn’t be any need to derive more variables for the model. The dataset is actually incomplete, so this wouldn’t prove valuable in any case.
* If you are coding please use your preferred language of choice, however preferably Python.
* Remember to comment your code so we can understand your analysis.
* It is also an opportunity for you to gain an understanding of the data you may use within the job role.

**About the Datasets**

We have provided a data set called “Breakage” in a pipe delimited CSV:

* The data in each row corresponds to a customer (‘INDIVIDUAL\_ID’) in a given month (‘Transaction\_Month’). The file is therefore said to be on the customer transaction month level.
* The data is filtered to only include records with a payment profile\* of 3 or more.
* The data is mostly comprised of bureau variables (definitions attached). Bureau variables are derived from a customer’s credit file. Each customer has a credit file detailing the status of all their credit accounts. If a value is NULL this means a customer has never had a credit line of that type. If a value is 0 (for example sum\_open\_lines) this means that they have once had an open credit line but they don’t anymore.
* Many of the aggregations are repeated for all credit lines together and for individual credit lines. This is indicated by a code proceeding the variable name. For example, cc\_avg\_arrears\_ratio is the arrears ratio across all of a customer’s credit card. See the DD for all credit line code mappings.

\*Payment Breakage: When a customer misses a payment after having paid in all three of the prior 3 months.

\* Payment Profile: The number of months since the customer last missed a payment