# DORA Manager Take-Home Assignment

To effectively mitigate the risk of charge-off accounts, we need to be able to identify accounts that are most likely to commit fraudulent activities (e.g. bad checks, deposits, making multiple point of sale transactions with no funds available, etc.). The identification of these accounts will be based on account level and transaction level characteristics. According to Verafin and supported by PCU charge-off data, the following are things to consider when attempting to uncover check and ACH-related fraud:

* **Unusual values.** For example, if a customer has historically transferred less than $1,500 but is now attempting a transfer for $5,000, it requires investigation.
* **How well you know the customer.** How old is their account? New accounts are inherently riskier. What is their overall relationship with your institution? If this customer has an established relationship with your institution and carries multiple products there is less risk.
* **Online account activity.** It is important to determine if there was unusual activity within the customer’s online account activity, which may indicate a takeover. For example, has the person recently altered contact information or has the business recently altered payee information via their online account?
* **Quick withdrawals.** This one is closely related with the age of the account. Criminals will often open a new account, sometimes using a stolen identity, fund the account from another institution and then attempt to withdraw the funds before your institution has reviewed the ACH returns.

**Objective:** Build a predictive DL model that identifies the segment of accounts that are most risk of charging off. Provide your source code along with a brief write-up of your findings.

**Dataset:** Consists of 55,676 observations and 12 variables.

* **Status**: Account status (Active, Closed, Dormant, Incomplete, Locked, Restricted, Unfunded)
* **COS**: Whether an account is labeled as charged-off or not (charged-off= 1; not charged-off= 0)
* **Rim\_age**: Tenure as a member (in months)
* **Loan:** Whether an account has a loan product with Provident (has at least one loan= 1; no loan= 0)
* **S\_plus\_c:** Whether an account has only a membership savings account and a checking account or just a membership savings (membership savings and checking account= 1; membership savings only= 0)
* **Fico\_b**: Credit score
* **Ck\_returns:** Number of returned checks since January 2017
* **Max\_days\_neg:** Maximum consecutive number of days an account stays since January 2017
* **Num\_times\_neg:** Number of times an account becomes negative since January 2017
* **Ntriggers:** Number of risk-related alerts received from the credit bureau (e.g. delinquency at another financial institution, bankruptcy, etc.) over the life of the account
* **Avg\_bal6:** Rolling 6-month deposit amount
* **Avg\_bal3:** Rolling 3-month deposit amount

**Assumptions:**

* Since new accounts are inherently riskier, the learning data set has been limited to accounts where the membership has been opened since January 2017.
* Training data should include all charged-off accounts and add randomly sampled non-charged off accounts to make up 20% of the entire observation.
* Training data should include all accounts, charged-off or not.
* Charge-off is defined as COS = 1 and status= Closed.
* Score active accounts that have not charged-off.
* Please document any other assumptions made that are not listed above.