Business Analytics: Homework 3

## Chengjun Liu, 7116099552

* Understand situation and need: We try to distinguish influence on delinquent credit accounts with different attributes paying debt by 13 different actions and find the suitable action for each delinquent credit accounts.
* Data: The data are collected from internal resources and credit bureau records to expand slice choices. We only select accounts which are delinquent and taken 1 of actions.
* General approach: We try to roll and slice customer records by months and select suitable records. We divide records according to 13 actions. For each action, we build individual statistical models to judge whether delinquent accounts with different attributes pay debt under one specific action. We input attributes of delinquent accounts into different action models to decide which action to take.
* Procedures
  + Select influential attributes
    - We select and focus on several main attributes of delinquent accounts. Here we suggest purchase/payment frequency, purchase category, single/total/average payment amount, cash advance amount/frequency, account state (delinquent or not, how long), gender, age, region. If possible, we can add more variables.
    - The variables are before the observation month and in a period of usually 6 months.
  + Define good records and find suitable records
    - We hope that the debt will be paid right after the actions, so we only observe the month after actions. Because we focus on one-month delinquent account, the accounts in bucket 1 in the observation month and with actions are included, while others are excluded.
    - Roll the records by month and filter through included records. If next month bucket goes 2, it’s bad (still delinquent). If bucket goes 0, it’s good (paid). If bucket remains 1, it’s indeterminate. Collect those records.
  + Build 13 action models and find best action
    - We try to predict whether delinquent accounts will or not pay debt under specific actions and compare probability of paying debt for each action. The statistical model to do this binary classification is logistic regression model.
    - We divide records according to 13 actions and build different logistic regression models for each action, using attributes as variable and paying or not as predicted variable. We divide test and training datasets and do training and testing to tune 13 models.
    - To predict reaction of delinquent accounts under different actions and select the one that works best, we compare the probability generates by 13 models because under logistic regression model the predicted variable is calibrated to probability so that it can be compared with those from other models. We select the best action by finding the one with largest probability of being good.