# STAT3612 Group Project: Data Science Challenge in Predicting Credit Card Delinquency

Release date: March 13, 2017

**Problem description:**

This project is to build a statistical or machine learning model for predicting the 60-day delinquency in a credit card portfolio. The training dataset consists of 6655 cards owned by 2000 customers together with the following attributes:

- `Age`: card duration since the origination
- `FICO`: credit score of the customer
- `Line`: credit line of limit of card
- `Balance`: outstanding balance of card
- `Payment`: minimum required payment of card

These varaibles are observed when the cards are all "current" (i.e. heathy without delinquency). The target response of interest is their follow-up performance in the next 12 months as measured by `FlagD60`, an indicator for 60 days past due (i.e. payment delinquency).

The model performance will be assessed based on a test data that consists of 1000 additional customers, in total 3306 cards. We use the single-valued ranking criteria:

$$F = \frac{2 \cdot P \cdot R}{P + R}, \quad \text{with} \quad P = \frac{\mathsf{TP}}{\mathsf{TP} + \mathsf{FP}}, \quad R = \frac{\mathsf{TP}}{\mathsf{TP} + \mathsf{FN}}$$

based on TP (true positive), FP (false positive) and FN (false negative) rates.

**Schedule:**

- Team formation: March 13 - 19 (At least 3 members, at most 4 members per team)
- Data release: March 20 (morning)
- Competition period: March 20 – April 18 (One and only one submission every 3 days)
- Oral presentation: April 24 & 27 (In-class presentation with peer assessment)
- Written report due: May 5 (Final report with detailed methodology and results)

**Remarks:**

- There is no restriction for the methods or software packages to be used.
- All the members in the same team share the same score (30% of final grade).