

Credit risk game

- Individual assignment.
 - Upload your decision(s) before 11:55PM (EDT) on October 15th, 2020.
 - You are required to provide your R code (upload it on ZoneCours).
 - This round of business simulation is worth 10% of your final grade.
-

Context:

In order to minimize loss and maximize profits from lending, banks require careful assessment of their borrowers. This has led to credit risk computation becoming a commonly used application of statistics and data science. You are working with a large bank to help them optimize their profits from financing individuals who apply for a loan. The bank has provided you with the records and results of lending money to some customers. You will be using this information and R to predict the risk of default when lending money to new customers and decide how many and which individuals to lend money to. Your objective is to maximize the banks' profits obtained from these loans.

Data:

A zip file containing the data is available on ZoneCours. The file *CreditGame_TRAIN.csv* contains records of previous loans provided by the bank. They include features about the customers and information about their default status. The file *CreditGame_Applications.csv* contains information about current loan applications. You must decide who on this list get their loan approved.

How to play the game:

You need to prepare a one column CSV file with the list of IDs of the customers whose loan application you accept from *CreditGame_Applications.csv*. Upload this decision on <https://dsgame.hec.ca/play> (consult ZoneCours for instructions on setting up your access, and joining the game). When you upload a decision, you will get immediate feedback on the profit that the loans have generated for the bank after 24 months. The platform allows for multiple uploads per person, up to 70, which means that you may try many different solutions.

On the upload platform, you will see not only your results but also those of the whole group. While the game is being played, you will see the "interim leaderboard." Before the deadline of the assignment, you must select your final decision as one of your uploads. When the game ends, the "real-life leaderboard" will be unveiled and will prevail for the final ranking. The "real-life leaderboard" plays the role of a test set: it is a holdout sample that is kept until the end to measure the performance of the final answers of everybody.

Method:

Your objective is based on a business outcome: profit. The process of prediction will involve cleaning, analyzing, modeling, and getting results. We do not give further instructions on the methods used; you are on your own for that. There is no single good answer and multiple strategies that can support the business problem reasonably well. We expect each student to come up with their own approach.

Disclaimer:

There are additional nuances for credit risk assessment in a real-life setting. Banks need to abide by the Basel accords and must comply with some rules to assess their credit risk. Credit risk typically implies a need to interpret the results of a model, and some standards in methodology apply. Although this business simulation is very realistic, both in terms of context and data, it does not depict those field-specific particularities.

Evaluation:

Look for the baseline on the upload platform. It corresponds to the profit made when all customer applications are approved. To get a passing grade, you must do better than that!

The evaluation will be based on the results at the end of the game. Each student must select one of their uploads as their final answer, and that answer will prevail. You get:

- 0% - if you are below the baseline on the interim leaderboard; at least 50% if your profit is above.
- 100% - Top 3 students on the “real-life leaderboard.”
- The rest of the marks will be linearly interpolated using the following equation with values from the “real-life leaderboard”:

$$50\{1 + (x - B)/(T - B)\}$$

where x is your profit, B the baseline, and T the profit of the third best student.

You must upload your R code on ZoneCours. It will not be reviewed systematically, only if some precisions are needed. Your grades could be reduced if irregularities are found in the R code.

Variables:

Variable name	Description
ID_TRAIN	Unique borrower ID
TYP_FIN	Type of funding requested (Car, Mortgage, or Credit)
NB_EMPT	Number of borrowers
R_ATD	Total Debt Amortization(TDA) Ratio, i.e., monthly financial commitments over monthly income
PRT_VAL	The requested loan amount over the value of the goods
DUREE	Requested loan duration
AGE_D	Age of the borrower
REV_BT	Gross Income
REV_NET	Net Income
TYP_RES	Residence Type – P: Owner, L: Tenant, A: Others
ST_EMPL	Employment Status – R: Regular, P: Part-Time, T: Self Employed
MNT_EPAR	Savings Value
NB_ER_6MS	Number of transactions refused due to insufficient funds in the last 6 months
NB_ER_12MS	Number of transactions refused due to insufficient funds in the last 12 months
NB_DEC_12MS	Number of overdrafts in the last 12 months
NB_OPER	Total number of transactions in record
NB_SATI	Total number of satisfactory transactions in record (No payment delay)
NB_COUR	Number of current transactions
NB_INTR_1M	Number of inquiries in the last month
NB_INTR_12M	Number of inquiries in the last 12 months
PIR_DEL	Worst current delinquency
NB_DEL_30	Number of 30–59 day delinquency in the last 12 months
NB_DEL_60	Number of 60–89 day delinquency in the last 12 months
NB_DEL_90	Number of 90+ day delinquency in the last 12 months
MNT_PASS	Value of financial Liabilities
MNT_ACT	Value of financial Assets
MNT_AUT_REN	Total authorized amount of revolving credit
MNT_UTIL_REN	Total used amount of revolving credit
MNT_DEMANDE	Loan amount requested
Target_0	Default is considered when payment is 90 days or more late within 24 months 1: Default, 0: Did not default