**Trabalho é individual para o último dia de aula**

## **Instruções:**

1) Baixar os datasets em: <https://www.kaggle.com/c/home-credit-default-risk/data>

2) Explorar conforme instruções:

* application\_train.csv: Esse csv possui 122 colunas: explorar 40 colunas sendo as primeiras 30 obrigatórias e 10 adicionais a escolher.
* 30 colunas obrigatórias:

SK\_ID\_CURR  
TARGET  
NAME\_CONTRACT\_TYPE  
CODE\_GENDER  
FLAG\_OWN\_CAR  
FLAG\_OWN\_REALTY  
CNT\_CHILDREN  
AMT\_INCOME\_TOTAL  
AMT\_CREDIT  
AMT\_ANNUITY  
AMT\_GOODS\_PRICE  
NAME\_TYPE\_SUITE  
NAME\_INCOME\_TYPE  
NAME\_EDUCATION\_TYPE  
NAME\_FAMILY\_STATUS  
NAME\_HOUSING\_TYPE  
REGION\_POPULATION\_RELATIVE  
DAYS\_BIRTH  
DAYS\_EMPLOYED  
DAYS\_REGISTRATION  
DAYS\_ID\_PUBLISH  
OWN\_CAR\_AGE  
FLAG\_MOBIL  
FLAG\_EMP\_PHONE  
FLAG\_WORK\_PHONE  
FLAG\_CONT\_MOBILE  
FLAG\_PHONE  
FLAG\_EMAIL  
OCCUPATION\_TYPE  
CNT\_FAM\_MEMBERS

* alunos deverão se organizar para que todas as colunas sejam exploradas
* escolher dois arquivos, para exploração e join/merge/concat com o 'application\_train.csv'

HomeCredit\_columns\_description.csv  
POS\_CASH\_balance.csv  
credit\_card\_balance.csv  
installments\_payments.csv  
application\_train.csv  
bureau.csv  
previous\_application.csv  
bureau\_balance.csv  
sample\_submission.csv

3) objetivos:

* entendimento do problema de negócio
* carregamento do dataset
* identificação e tratamento de NaN
* exploração univariada, bivariada
* visualizações
* insights de negócio
* formatação do dataset para algoritmos de machine learning
* criação de features: agregação de dados, junção de dois ou mais dataframes

4) entregável: jupyter notebook. Não será necessário apresentação em powerpoint, nem relatório.

5) entrega limite: 8/agosto/2018

6) trabalho individual. Haverá aplicação sistema anti-plágio nas submissões. (<http://www.fgv.br/mailing/codigo-etica-conduta/24/>) suspeitas de plágio estarão sujeitos a arguição oral de 5 mins por aluno no ultimo dia de aula (11/agosto/2018),

7) avaliação:

* jupyter notebook e arguição (se houver)

Data Description

application\_{train|test}.csv

* + This is the main table, broken into two files for Train (with TARGET) and Test (without TARGET).
  + Static data for all applications. One row represents one loan in our data sample.
* bureau.csv
  + All client's previous credits provided by other financial institutions that were reported to Credit Bureau (for clients who have a loan in our sample).
  + For every loan in our sample, there are as many rows as number of credits the client had in Credit Bureau before the application date.
* bureau\_balance.csv
  + Monthly balances of previous credits in Credit Bureau.
  + This table has one row for each month of history of every previous credit reported to Credit Bureau – i.e the table has (#loans in sample \* # of relative previous credits \* # of months where we have some history observable for the previous credits) rows.
* POS\_CASH\_balance.csv
  + Monthly balance snapshots of previous POS (point of sales) and cash loans that the applicant had with Home Credit.
  + This table has one row for each month of history of every previous credit in Home Credit (consumer credit and cash loans) related to loans in our sample – i.e. the table has (#loans in sample \* # of relative previous credits \* # of months in which we have some history observable for the previous credits) rows.
* credit\_card\_balance.csv
  + Monthly balance snapshots of previous credit cards that the applicant has with Home Credit.
  + This table has one row for each month of history of every previous credit in Home Credit (consumer credit and cash loans) related to loans in our sample – i.e. the table has (#loans in sample \* # of relative previous credit cards \* # of months where we have some history observable for the previous credit card) rows.
* previous\_application.csv
  + All previous applications for Home Credit loans of clients who have loans in our sample.
  + There is one row for each previous application related to loans in our data sample.
* installments\_payments.csv
  + Repayment history for the previously disbursed credits in Home Credit related to the loans in our sample.
  + There is a) one row for every payment that was made plus b) one row each for missed payment.
  + One row is equivalent to one payment of one installment OR one installment corresponding to one payment of one previous Home Credit credit related to loans in our sample.
* HomeCredit\_columns\_description.csv
  + This file contains descriptions for the columns in the various data files.

