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

Seguem instruções:

1) Baixar dataset 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 univariadaj, 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)