

# Processo seletivo - Cientista de Dados

## Orientações

1. O candidato será avaliado quanto:
  - a) Entendimento do problema
  - b) Clareza do Código
  - c) Técnica adequada
  - d) Capacidade do modelo predizer adequadamente os clientes com propensão a evasão
  - e) Apresentação das decisões tomadas para alcance do resultado e explicação da avaliação do modelo
2. O candidato deverá disponibilizar à banca, via e-mail, o notebook com o código fonte utilizado para realização do desafio.
  1. [Thiago.f.soares@caixa.gov.br](mailto:Thiago.f.soares@caixa.gov.br)
  2. [pedro.miranda@caixa.gov.br](mailto:pedro.miranda@caixa.gov.br)

## Questões

1. Realizar a análise exploratória dos dados do arquivo "base\_churn.csv" e apresentar modelo(s) que predizem a propensão de evasão de clientes da instituição financeira a partir da base de dados disponibilizada.

No conjunto de dados do arquivo "base\_churn.csv" temos as seguintes variáveis:

- CustomerId: identificação do cliente;
- Surname: sobrenome do clientes;
- CreditScore: pontuação de credito, 0 alto risco de inadimplência e 1000 clientes com baixo risco de inadimplência;
- Geography: país que o serviço é oferecido;
- Gender: sexo do cliente;
- Age: idade do cleinte;
- Tenure: um indicativo de estabilidade no emprego, em que 0 significa pouca estabilidade e 10 muita estabilidade.
- Balance: saldo da conta corrente;
- NumOfProducts: número de produtos bancários adquiridos;

- HasCrCard: se tem cartão de credito ou não, (Sim = 1 e Não = 0);
- IsActiveMember: se é um cliente com conta ativa, (Ativo = 1) ;
- EstimatedSalary: salário estimado;
- Exited: cliente deixou de ser cliente do banco ou não (Churn = 1).