Desafio Cientista de Dados

# 

# Introdução

Olá candidato, o objetivo deste desafio é testar os seus conhecimentos sobre a resolução de problemas de análise de dados e aplicação de modelos preditivos. Queremos testar seus conhecimentos dos conceitos estatísticos de modelos preditivos, criatividade na resolução de problemas e aplicação de modelos básicos de machine learning. É importante deixar claro que não existe resposta certa e que o que nos interessa é sua capacidade de descrever e justificar os passos utilizados na resolução do problema.

# Desafio

Seu objetivo é prever o *churn* (abandono de clientes) de um banco de dados fictício de uma instituição financeira. Para isso são fornecidos dois *datasets*: um *dataset* chamado **Abandono\_clientes** composto por 10000 linhas e 13 colunas de informação (*features)*, sendo uma coluna “Exited” composta por dados binários: 1 se o cliente abandonou o banco, 0 se não. O segundo dataset possui 1000 linhas e 12 colunas e não possui a coluna “Exited”. **Seu objetivo é prever essa coluna a partir dos dados enviados e nos enviar para avaliação dos resultados.**

# Entregas

1. Descreva graficamente os dados disponíveis, apresentando as principais estatísticas descritivas. Comente o porquê da escolha dessas estatísticas.
2. Explique como você faria a previsão do **Churn** a partir dos dados. Quais variáveis e/ou suas transformações você utilizou e por quê? Qual tipo de problema estamos resolvendo (regressão, classificação)? Qual modelo melhor se aproxima dos dados e quais seus prós e contras? Qual medida de performance do modelo foi escolhida e por quê?
3. Envie o resultado final do modelo em uma planilha com apenas duas colunas (rowNumber, predictedValues).

# Prazo

Você tem até **7 dias corridos** para a entrega, contados a partir do recebimento deste desafio.

Envie o seu relatório dentro da sua data limite para o e-mail: **selecao.lighthouse@indicium.tech**

Bom trabalho!