



Desafio para área de Data Science

Parabéns pela aprovação na primeira etapa do processo seletivo para a área de Data Science da DHAUZ!

Seguindo com o nosso processo, nesta etapa você deverá realizar um desafio técnico, detalhado abaixo. O objetivo é avaliar a sua capacidade de exploração de dados, criatividade na elaboração de hipóteses e metodologia na resolução de problemas. Use o seu conhecimento para investigar e extrair insights.

Você pode utilizar a linguagem, ferramentas e frameworks que se sentir mais confortável para elaborar a solução. Recomendamos o desenvolvimento do racional e exploração em uma ferramenta como por exemplo o Jupyter Notebook para facilitar a inclusão de comentários e gráficos explicativos.

Sinta-se livre para utilizar qualquer outra funcionalidade não listada acima que demonstre suas habilidades!

O projeto deve ser feito em um repositório no Github e o seu link enviado no final do desafio.

1. Evasão de clientes - Telecom

Evasão ou *churn*, em seu sentido mais amplo, é uma medida do número de indivíduos ou itens que saem de um grupo coletivo durante um período específico.

Você foi contratado pela DHAUZ como cientista de dados para analisar uma base de dados de clientes de uma empresa de Telecomunicações e sua tarefa é investigar os dados em busca de *insights* que possam ajudar a empresa a evitar evasões e também construir um modelo preditivo que possa antecipar essas evasões, de modo que a empresa tenha tempo hábil para agir com ações de retenção.

Informações sobre o dataset:

- As informações são anonimizadas por questões de privacidade
- Base de dados no link: https://dhauz-challenges.s3.amazonaws.com/churn_challenge.zip
- Um dicionário é fornecido juntamente aos dados

Comece respondendo as seguintes questões:

- Elabore hipóteses e visualizações envolvendo a variável *churn* e, pelo menos, outras três variáveis presentes no dataset;
- Desenvolva um modelo preditivo de classificação para identificar possíveis evasões e utilize métricas adequadas para argumentar a efetividade do modelo;
- (BÔNUS) Qual o perfil de cliente que possui maior probabilidade de evasão?