

## Abandono de clientes

**Dificuldade:** ★★☆☆☆

**Introdução:** o abandono de clientes é um grande problema para todos os business. Identificar quando um cliente está mais propenso a abandonar a empresa pode se tornar um diferencial competitivo.

**Problema:** prever se um cliente da Telco vai cancelar ou não o seu plano. Trata-se de um problema de classificação. Crie uma pipeline de dados para que novas observações sejam processadas automaticamente.

Dataset: [LINK](#)

### Material Recomendado

- [Guia completo EDA](#)
- [Feature Selection](#)
- [Feature Engineering](#)

### Material Obrigatório

- Regressão Logística
  - [Teoria - explicação acadêmica](#)
  - [Teoria simplificada e aplicação](#)
- Decision Trees e Random Forest
  - [Teoria](#)
  - [Aplicação](#)
- F1 Score, Precision e Recall
  - [Fórmulas](#)
  - [Explicação detalhada](#)

### **Dicas:**

- O dataset não é muito grande, então tente pensar e responder algumas questões através dos dados.
- Use funções do tipo "One Hot Encoder" para as variáveis categóricas.
- Compare resultados com e sem padronização de variáveis.

### **Resultado:**

- Esse problema é importante por trazer discussões como F1 Score e métricas importantes para problemas de classificação, onde ao aplicar somente a função `.score()`, perdem-se informações importantes.
- Uma performance considerada boa seria um F1 Score em torno de 0.6
- Verifique os coeficientes do modelo de regressão logística para ver o que mais influencia no seu problema.
- Benchmark de solução: [LINK](#)