**Passo 1. Baixar os dados**

Model Fitness forneceu a você dados CSV contendo dados sobre rotatividade em um determinado mês e informações sobre o mês anterior. O conjunto de dados inclui os seguintes campos:

* 'Churn' — a rotatividade do mês em questão
* Campos de dados atuais:
  + Dados do mês anterior
    - 'gender'
    - 'Near\_Location' — se o cliente morar ou trabalhar na vizinhança onde a academia está localizada
    - 'Partner' — se o usuário for um funcionário de uma companhia parceira (a academia tem empresas parceiras cujos funcionários conseguem descontos; nesses casos, a academia armazena informações sobre clientes de são funcionários)
    - Promo\_friends — se o cliente originalmente se inscreveu através de uma oferta "traga um amigo" eles normalmente usam o código de promoção do amigo quando pagam pela primeira filiação)
    - 'Phone' — se o usuário fornece o seu número de telefone
    - 'age' (idade)
    - 'Lifetime' — o tempo (em meses) desde a primeira vez que o cliente veio à academia
* Dados do log de frequência e compras e dados sobre status de filiação atual
  + 'Contract\_period' — 1 mês, 3 meses, 6 meses, ou um ano
  + 'Month\_to\_end\_contract' — os meses remanescentes até que o contrato expira
  + 'Group\_visits' — se o cliente participa de sessões em grupo
  + 'Avg\_class\_frequency\_total' — frequência média de idas por semana por toda a vida do cliente
  + 'Avg\_class\_frequency\_current\_month' — frequência média de visitas por semana durante o mês corrente
  + 'Avg\_additional\_charges\_total' — a quantidade total de dinheiro gasto em outros serviços da academia: café, artigos esportivos, cosméticos, massagem, etc.

Caminho do arquivo: /datasets/gym\_churn\_us.csv. [Baixe o conjunto de dados](https://practicum-content.s3.us-west-1.amazonaws.com/datasets/gym_churn_us.csv).

**Passo 2. Realize análise exploratória dos dados (AED)**

* Olhe para o conjunto de dados: ele contém alguma característica ausente? Estude a média de valores e desvio padrão (use o método describe()).
* Observe a média dos valores médios das características em dois grupos: para aqueles que ficaram (use o método groupby()).
* Faça histogramas de barra e distribuições de características para aqueles que saíram (rotatividade) e aqueles que ficaram.
* Construa a matriz de correlação e a exiba.

**Passo 3. Construa um modelo para predizer a rotatividade de clientes**

Construa um modelo de classificação binária para clientes onde a variável objetivo é a saída de usuários do próximo mês.

* Divida os dados de treinamento e validação em dois conjuntos usando a função train\_test\_split().
* Treine o modelo no conjunto com dois métodos:
  + regressão logística
  + floresta aleatória
* Avalie acurácia, precisão e sensibilidade para ambos os modelos usando dados de validação. Use-os para comparar os modelos. Qual modelo rendeu melhores resultados?

Lembre-se de indicar o parâmetro random\_state quando dividir os dados e definir o algoritmo.

**Passo 4. Crie agrupamentos de clientes**

Defina ao lado colunas com dados sobre rotatividade e identifique agrupamentos do objeto (cliente):

* Padronize os dados.
* Use a função linkage() para construir a matriz das distâncias baseada na matriz de características padronizada e construa um dendrograma. Perceba: renderizar o dendrograma pode demorar um tempo! Use o gráfico resultante para estimar o número de agrupamentos que você pode destacar.
* Treine o modelo de agrupamento com o algoritmo K-means e preveja agrupamentos de clientes. (Deixe que o número de agrupamentos seja n=5,para que seja fácil comparar seus resultados com os de outros estudantes. No entanto, na vida real, ninguém vai te dar essas dicas, para que você tenha que decidir baseados nos gráficos das etapas anteriores.)
* Olhe para os valores médios das características para agrupamentos. Nada chama a sua atenção?
* Faça distribuições de características para os agrupamentos. Você notou alguma coisa?
* Calcule a taxa de rotatividade para cada agrupamento (use o método groupby()). Eles diferem em termos de taxa de rotatividade? Quais agrupamentos são propensos a sair, e quais são leais?

**Passo 5. Chegue a conclusões e faça recomendações básicas sobre trabalhar com clientes**

Tirar conclusões e formular recomendações sobre a estratégia de interação e retenção de clientes.

Você não precisa entrar em detalhes. Três ou quatro princípios essenciais e exemplos da sua implementação na forma de passos de marketing específicos vai funcionar.