A companhia de seguros Proteja Seu Amanhã quer resolver algumas tarefas com a ajuda de aprendizado de máquina e você precisa avaliar a possibilidade de fazê-lo.

* Tarefa 1: Encontrar clientes semelhantes a um determinado cliente. Isso vai ajudar os agentes da empresa com tarefas de marketing.
* Tarefa 2: Predizer se um novo cliente provavelmente receberá um pagamento de seguro. Um modelo de predição pode ser melhor do que um modelo dummy?
* Tarefa 3: Predizer o número de pagamentos de seguro que um novo cliente provavelmente receberá usando um modelo de regressão linear.
* Tarefa 4: Proteger os dados pessoais dos clientes sem prejudicar o modelo da tarefa anterior.

É necessário desenvolver um algoritmo de transformação de dados que tornaria difícil recuperar informações pessoais se os dados caíssem nas mãos erradas. Isso é chamado de *mascaramento de dados* ou *ofuscação de dados*. Mas os dados devem ser protegidos de forma que a qualidade dos modelos de aprendizado de máquina não piore. Você não precisa escolher o melhor modelo, só prove que o algoritmo funciona corretamente.

Instruções do Projeto

1. Carregue os dados.
2. Verifique se os dados estão livres de problemas — não há dados ausentes, valores extremos e assim por diante.
3. Trabalhe em cada tarefa e responda às perguntas feitas no modelo de projeto.
4. Tire conclusões com base em sua experiência trabalhando no projeto.

Há algum pré-código no modelo do projeto, sinta-se à vontade para usá-lo, alguns deles precisam ser concluídos primeiro. Além disso, há dois apêndices no modelo de projeto com informações úteis.

**Descrição de Dados**

O conjunto de dados é armazenado no arquivo /datasets/insurance\_us.csv. [Você pode baixar o conjunto de dados aqui.](https://practicum-content.s3.us-west-1.amazonaws.com/datasets/insurance_us.csv)

* Características: sexo, idade, salário e número de familiares do segurado.
* Alvo: número de pagamentos de seguro recebidos por um segurado nos últimos cinco anos.