

# Projeto final do curso de Análise de Dados (Comissão 71010)

---

*Modelagem de Conversão e Risco em Seguros  
Usando Dados de Comportamento do Cliente*

---

Autora: Olga Abramova

## **Introdução:**

Este projeto se propõe a analisar os fatores que influenciam a conversão de clientes e o risco em um pipeline de vendas de seguros. Utilizando um conjunto de dados gerado sinteticamente, buscamos explorar a relação entre o comportamento do cliente, seu histórico de seguro, pontuação de crédito e a taxa de conversão. A análise será fundamentada em variáveis demográficas, comportamentais e relacionadas às apólices, com foco na validação de uma hipótese que propõe que clientes com seguro prévio e pontuação de crédito mais alta são mais propensos a converter e necessitam de menos ajustes no prêmio.

O conjunto de dados utilizado possui 27 colunas e 10.000 linhas, abrangendo informações mensuráveis como idade, frequência de sinistros, descontos, histórico de seguro e resultados de conversão. Este estudo se mostra relevante para seguradoras que buscam otimizar suas estratégias de marketing, ajustar prêmios de forma justa e identificar clientes com alto risco ou alta probabilidade de conversão.

Consideraremos a seguinte hipótese:

### **Hipótese:**

*Clientes com histórico prévio de seguro e pontuações de crédito mais altas têm maior probabilidade de conversão e demandam menos ajustes no prêmio.*

### **Descrição da Metodologia SMART dos Dados:**

**S – Specific / Específico:** O conjunto de dados analisa os fatores que influenciam a conversão de clientes e o risco em um pipeline de vendas de seguros. Ele inclui atributos demográficos, comportamentais e relacionados às apólices.

**M – Measurable / Mensurável:** O conjunto é composto por 27 colunas e 10.000 linhas, abrangendo dados mensuráveis como idade, frequência de sinistros, descontos, histórico prévio de seguro e resultados de conversão.

**A - Achievable / Alcançável:** os objetivos são realistas e alcançáveis. Os dados foram gerados sinteticamente, eliminando preocupações com privacidade enquanto preservam a validade analítica. Eles contêm informações suficientes para elaborar uma análise da hipótese proposta.

**R - Relevant / Relevante:** Este projeto é relevante para seguradoras que buscam otimizar estratégias de marketing, ajustar prêmios de forma justa e identificar clientes de alto risco ou com alta probabilidade de conversão.

**T - Time-bound / Com prazo definido:** Embora as informações sobre o período de tempo não estejam explicitamente presentes, as colunas “Time\_Since\_First\_Contract” / “Tempo\_Desde\_o\_Primeiro\_Contato” e “Time\_to\_Conversion” / “Tempo\_para\_Conversão” contêm dados suficientes para permitir uma análise temporal da jornada dos clientes.