

# 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.

## **Usuário Final e Nível de Aplicação da Análise**

Este projeto tem como **usuários finais** os **gestores e analistas das áreas de vendas, marketing, subscrição e pricing** no setor de seguros. São profissionais responsáveis pela análise de desempenho comercial, definição de estratégias de conversão de leads, modelagem de risco e desenvolvimento de políticas de precificação.

A análise proposta se aplica principalmente aos níveis **tático** e **estratégico** da organização:

- **Nível Tático:**  
A análise permite que equipes de vendas e marketing identifiquem padrões de comportamento e perfil de clientes com maior probabilidade de conversão. Isso possibilita ajustes em abordagens comerciais, priorização de leads e campanhas segmentadas com base em dados históricos e demográficos.
- **Nível Estratégico:**  
Os insights gerados também contribuem para a tomada de decisões de longo prazo, como a definição de políticas de subscrição, desenvolvimento de novos produtos, revisão de critérios de aceitação e precificação de seguros. A validação de hipóteses relacionadas à influência do histórico de seguro e pontuação de crédito no comportamento do cliente pode impactar diretamente o direcionamento estratégico da empresa.

Portanto, a análise ultrapassa o escopo operacional e se posiciona como uma ferramenta de apoio à tomada de decisão em níveis mais elevados da organização.



## **Diagrama de Entidade-Relacionamento (DER)**

### ***Entidades e Atributos:***

#### **1. customer\_data**

- **PK:** Customer\_ID – INT NOT NULL
- Age - INT
- Is\_Senior - BOOL
- Marital\_Status – VARCHAR(50)
- Credit\_Score - INT
- Region – VARCHAR(50)

#### **2. policy\_data**

- **PK:** Policy\_ID – VARCHAR(10) NOT NULL
- **FK:** Customer\_ID → customer\_data(Customer\_ID) – INT NOT NULL
- Married\_Premium\_Discount - INT
- Prior\_Insurance – VARCHAR(20)
- Prior\_Insurance\_Premium\_Adjustment - INT
- Policy\_Type – VARCHAR(100)
- Policy\_Adjustment - INT
- Premium\_Amount - INT
- Total\_Discounts - INT
- Time\_Since\_First\_Contact - INT
- Conversion\_Status - BOOL
- Website\_Visits - INT
- Inquiries - INT
- Quotes\_Requested - INT
- Time\_to\_Conversion - INT
- Premium\_Adjustment\_Credit - INT
- Premium\_Adjustment\_Region - INT



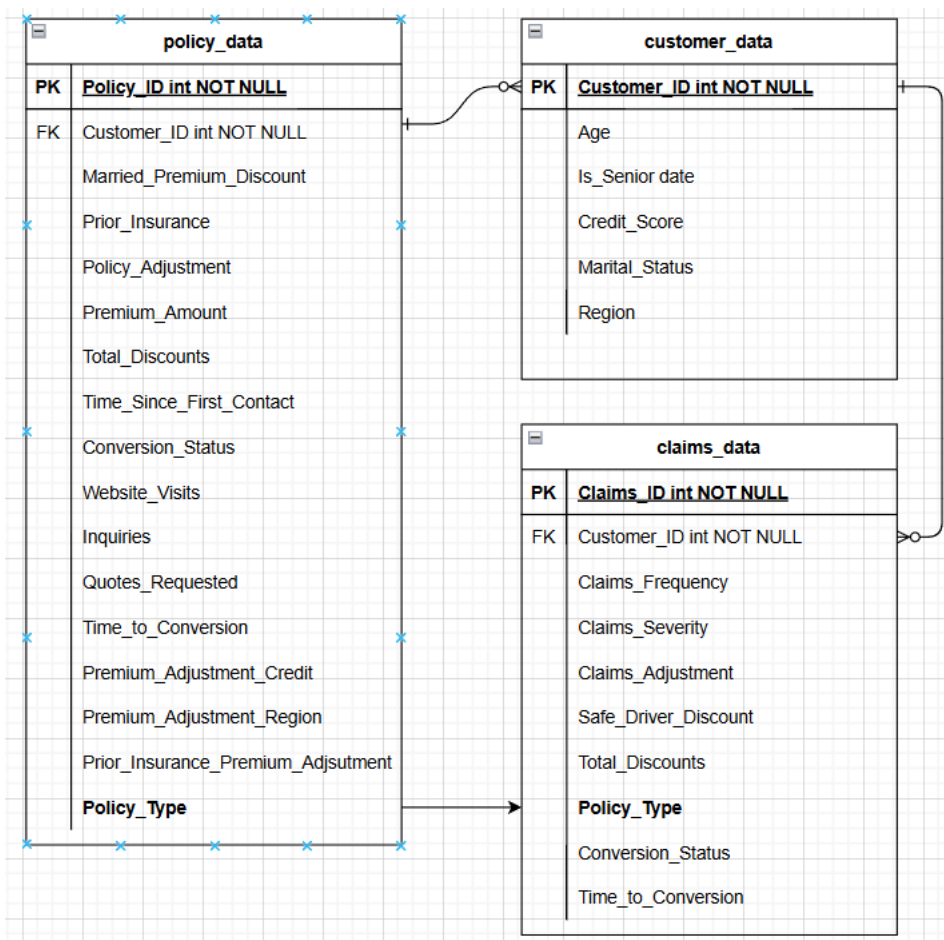
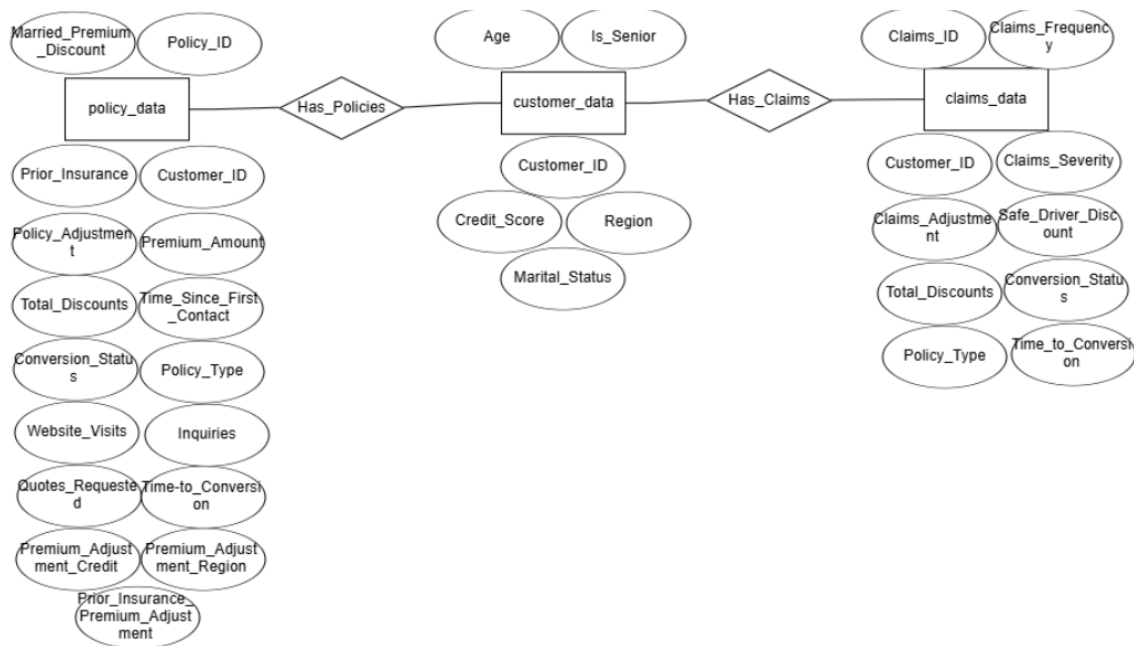
### 3. claims\_data

- **PK:** Claims\_ID – VARCHAR(10) NOT NULL
- **FK:** Customer\_ID → customer\_data(Customer\_ID) – INT NOT NULL
- Claims\_Frequency - INT
- Claims\_Severity – VARCHAR(50)
- Claims\_Adjustment - INT
- Policy\_Type – VARCHAR(100)
- Safe\_Driver\_Discount - BOOL
- Total\_Discounts - INT
- Conversion\_Status - BOOL
- Time\_to\_Conversion - INT

#### **Relacionamentos:**

- **customer\_data** (1) ↔ (N) **policy\_data**  
Cada cliente pode ter várias apólices.
- **customer\_data** (1) ↔ (N) **claims\_data**  
Cada cliente pode ter múltiplos registros de sinistros (claims).
- **Policy\_Type** aparece em ambas **policy\_data** e **claims\_data**, podendo ser usado para análises cruzadas, mas não é uma entidade separada neste contexto (a não ser que haja uma tabela referencial para tipos de apólice).



**DER – Descrição Visual:**

**Observações:**

- Há chave estrangeira em `policy_data` e `claims_data` ligando ao mesmo `customer_id`, centralizando o cliente como entidade principal.
- Alguns campos aparecem em mais de uma tabela (como `conversion_status`, `total_discounts`, `policy_type`), o que permite análises comparativas e consistência de métricas ao longo do pipeline.
- Pode-se estender esse modelo com uma tabela de referência para **`policy_type`** caso se deseje normalização adicional.

