

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 - A tabela customer_data contém informações básicas sobre o cliente que podem ajudar a agrupá-lo em uma categoria específica para auxiliar na análise de dados.

- **PK:** Customer_ID – INT NOT NULL
- Age - INT
- Is_Senior - BOOL
- Marital_Status – VARCHAR(50)
- Credit_Score - INT
- Region – VARCHAR(50)

2. policy_data - A tabela policy_data contém todas as informações vitais sobre a apólice que o cliente possui e pode ajudar a analisar aspectos numéricos das apólices de seguro.

- **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

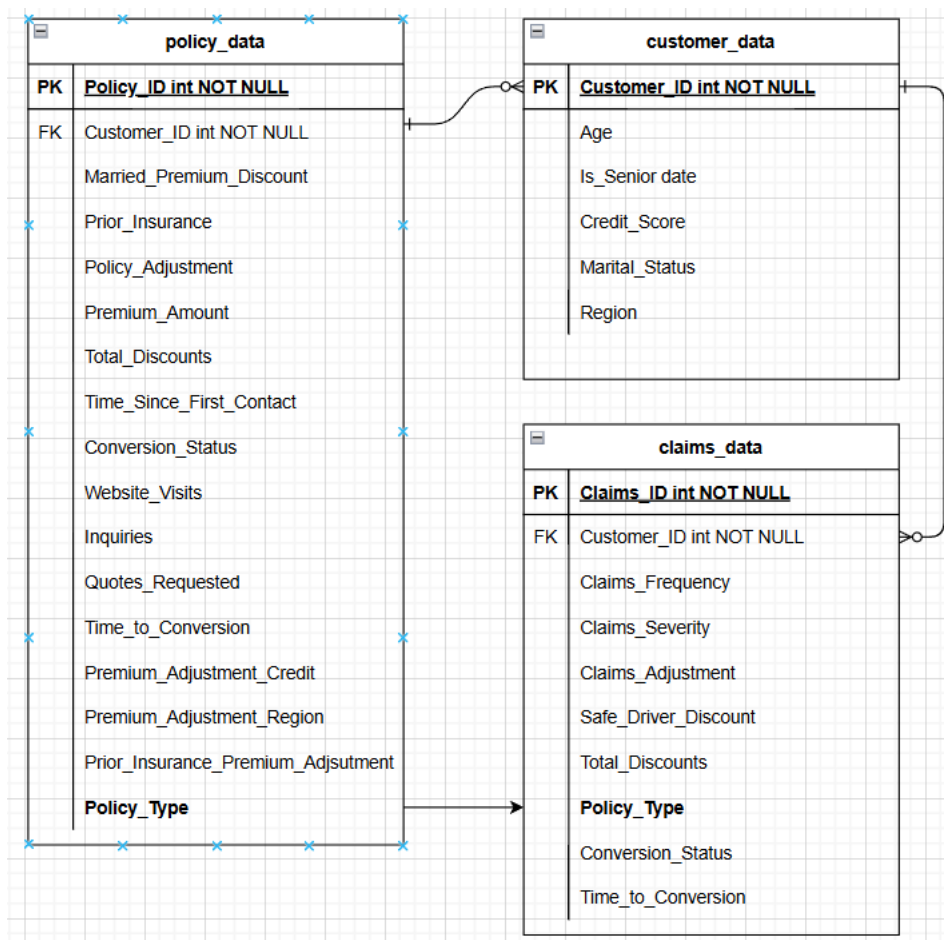
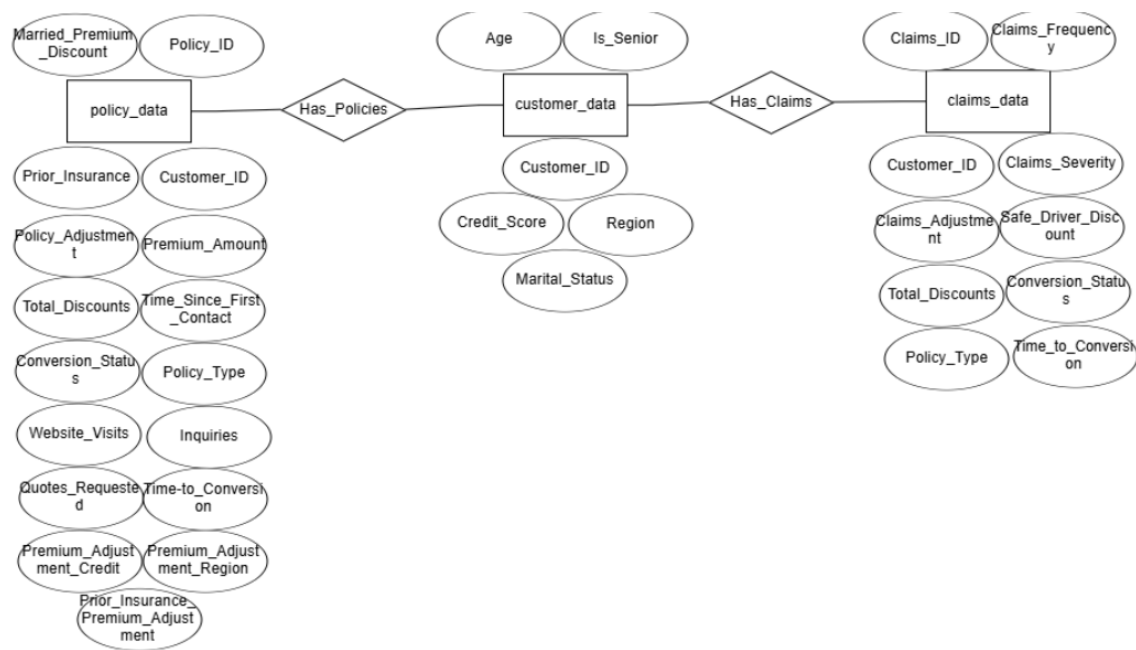
3. claims_data - A tabela claims_data contém todas as informações importantes sobre sinistros que podem ser usadas para análise de dados e tendências.

- **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.

Relatório de Análise de Dados em Power BI***Introdução***

O presente relatório tem como objetivo descrever a estrutura e os principais insights obtidos a partir da análise de um conjunto de dados do setor de seguros, utilizando a ferramenta Power BI.

Estrutura dos Dados

O conjunto de dados foi carregado no Power BI e está organizado em três tabelas distintas:

- `customer_data` (dados de clientes)
- `policy_data` (dados de apólices)
- `claims_data` (dados de sinistros)

Relações relevantes entre essas tabelas foram devidamente configuradas na aba Model View, garantindo a integridade e a coerência das análises.

Medidas Criadas

Foram desenvolvidas medidas com base nos dados numéricos, com o intuito de facilitar a visualização das principais informações e tendências presentes no conjunto de dados. Essas medidas foram essenciais para o desenvolvimento de indicadores-chave de desempenho (KPIs) e análises estratégicas.



Considerações sobre a Aba D_Calendario

Foi criada uma aba denominada D_Calendario, com o objetivo de introduzir uma série temporal de dados ao relatório. No entanto, como o conjunto de dados atual não contém informações baseadas em datas, essa aba não será utilizada nesta versão do relatório. Ressalta-se, contudo, que a estrutura foi preparada para possibilitar a incorporação futura de análises temporais e projeções, caso dados cronológicos venham a ser disponibilizados.

Painel de Análise (Dashboard)

Foi elaborado um painel interativo no Power BI composto por quatro abas:

- **Capa:** Apresenta o tema e o escopo do relatório.
- **Visão Geral:** Exibe os principais KPIs e valores calculados com base nas medidas desenvolvidas.
- **Detalhamento:** Traz visualizações mais específicas, permitindo uma análise aprofundada.
- **Análise:** Mostra as principais tendências identificadas no conjunto de dados.

Medidas Relevantes

A aba de Medidas contém seis cálculos numéricos considerados cruciais para a tomada de decisão em ambiente corporativo. Abaixo, segue a descrição de cada medida:

- **Número de Clientes:** Representa o total de clientes ativos no portfólio da empresa. Esse indicador é útil na definição de metas para o crescimento do negócio no próximo ano.
 - **Formula:** Número de clientes =
COUNTROWS(customer_data)
- **Prêmio Máximo (Maximum Premium):** Indica o valor mais alto cobrado por uma apólice no portfólio.
 - **Formula:** Prêmio máximo =
MAX(policy_data[Premium_Amount])
- **Prêmio Mínimo (Minimum Premium):** Reflete o menor valor de prêmio praticado. Ambas as medidas permitem comparar a política de preços da empresa com a dos concorrentes, servindo como base para decisões estratégicas.
 - **Formula:** Prêmio mínimo =
MIN(policy_data[Premium_Amount])



- **Prêmio Total** (Total Premium): Soma dos prêmios de todas as apólices. Essa medida possibilita avaliar o crescimento do portfólio frente às metas estabelecidas, além de servir de referência para projeções de atividades futuras.
 - **Formula:** Prêmio total =

$$\text{SUMX}(\text{policy_data}, (\text{policy_data}[\text{Premium_Amount}] - \text{policy_data}[\text{Married_Premium_Discount}] + \text{policy_data}[\text{Policy_Adjustment}] + \text{policy_data}[\text{Premium_Adjustment_Credit}] + \text{policy_data}[\text{Premium_Adjustment_Region}] - \text{policy_data}[\text{Total_Discounts}]))$$
- **Tipo de Apólice Mais Vendido** (Top Policy Type): Identifica o tipo de apólice com maior volume de vendas. Essa informação é valiosa para o planejamento de ações comerciais e definição de foco estratégico.
 - **Formula:** TopPolicyType =
 -- tipo de apólice mais adquirida

```

CALCULATE (
    SELECTEDVALUE(policy_data[Policy_Type]),
    TOPN(1,
        SUMMARIZE(
            'policy_data',
            'policy_data'[Policy_Type],
            "PolicyCount", COUNT(policy_data[Customer_ID])
        ),
        [PolicyCount]
    )
)

```
- **Região com Maior Acúmulo de Apólices** (Top Region): Determina a região com maior concentração de apólices. Tal dado auxilia na alocação eficiente de recursos, com o objetivo de manter o nível de serviço diante do crescimento da carteira.
 - **Formula:** TopRegion =
 --região com o maior número de clientes

```

CALCULATE (
    SELECTEDVALUE(customer_data[Region]),
    TOPN(1,
        SUMMARIZE(
            'customer_data',
            'customer_data'[Region],
            "PolicyCount", COUNT('customer_data'[Customer_ID])
        )
    )
)

```



```
),  
[PolicyCount]  
)  
)
```

Conclusão

A análise realizada por meio do Power BI permitiu a extração de informações estratégicas relevantes, contribuindo para o monitoramento do desempenho atual da carteira de seguros e apoiando o planejamento de ações futuras. O painel desenvolvido oferece uma visão clara e objetiva, promovendo uma base sólida para a tomada de decisão gerencial.

