

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

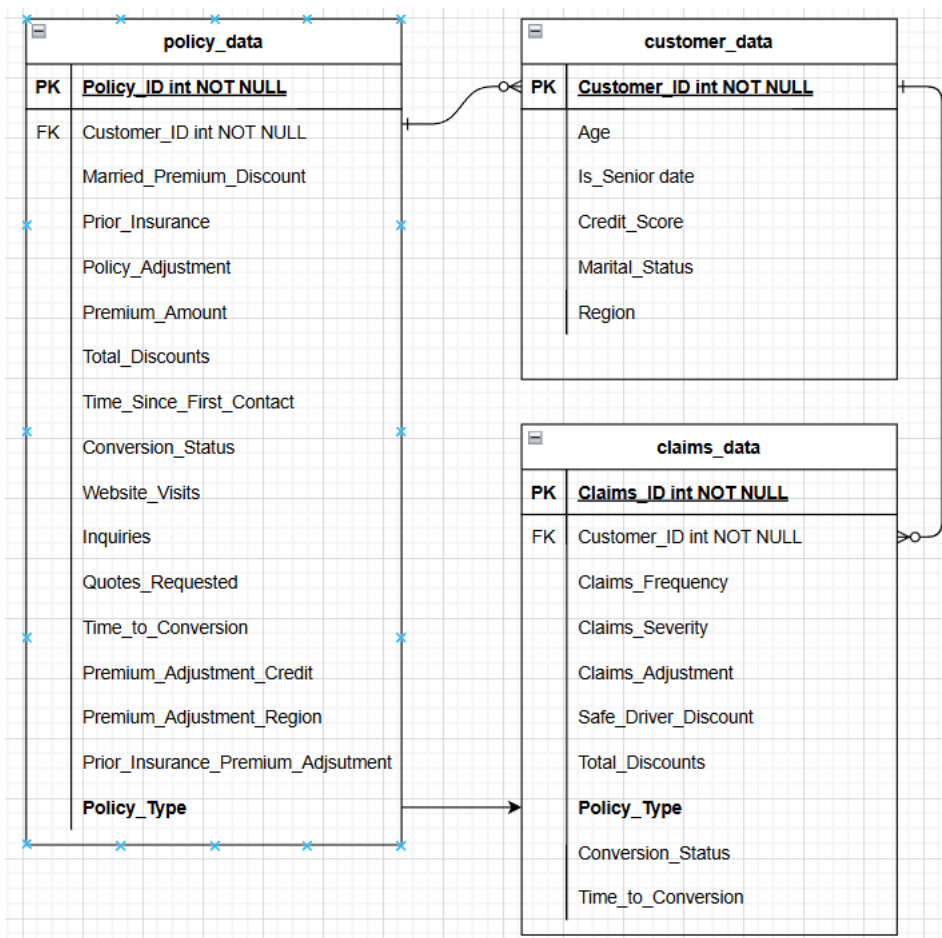
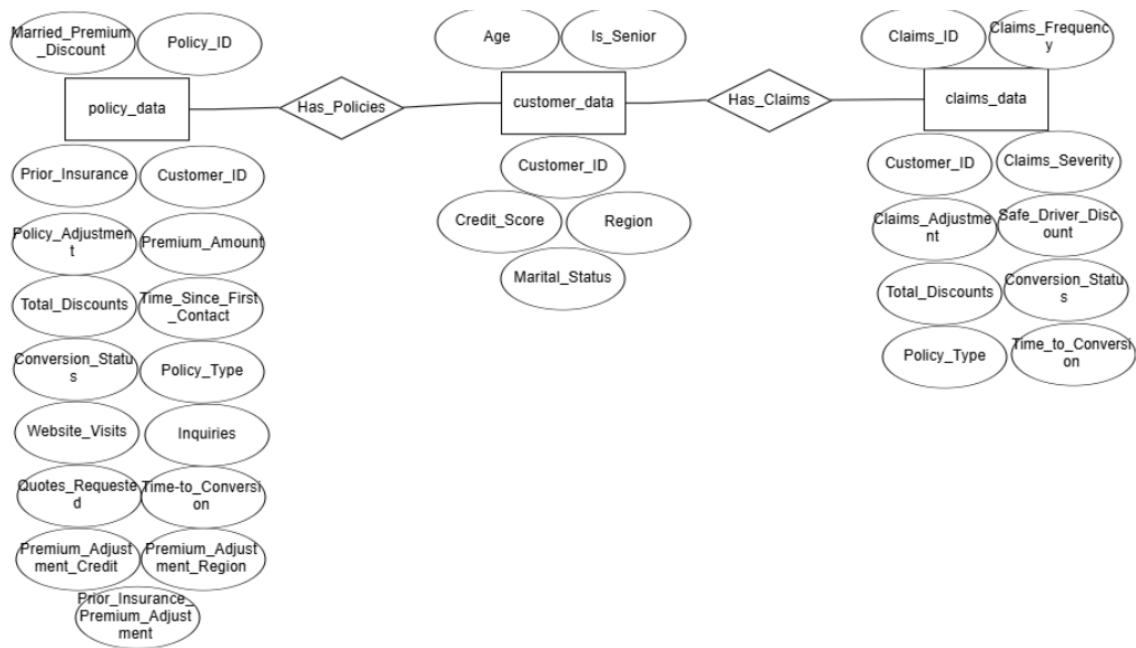
3. claims_data - A tabela claims_data contém todas as informações importantes sobre sinistros que podem ser usadas para análise de dados e tendências.

- **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]  
)  
)
```

Dashboard em Power BI para Análise de Dados de Seguros

Introdução

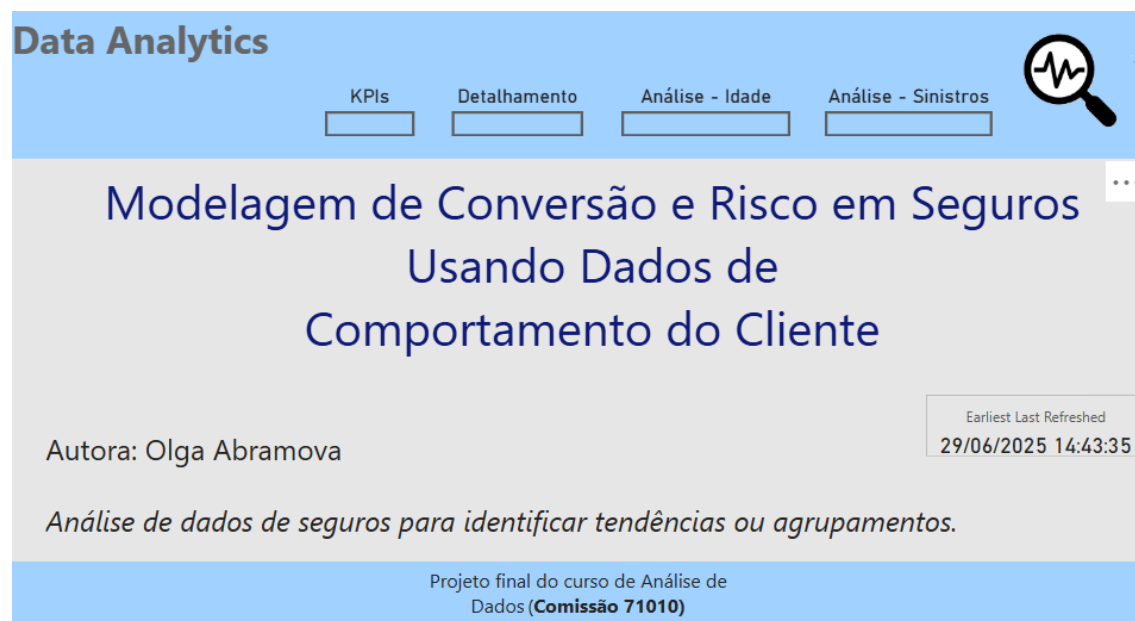
O dashboard do projeto foi desenvolvido utilizando o Power BI com o objetivo de analisar as informações de seguros utilizadas neste estudo. O painel interativo é composto por cinco abas principais, cada uma com funcionalidades específicas que contribuem para uma compreensão abrangente dos dados.

Estrutura do Dashboard

1. Aba "Capa"

A aba "Capa" representa a página de abertura do projeto e contém:

- O tema do projeto,
- O nome da autora,
- A data da última atualização do dashboard,
- E botões de navegação entre as páginas, que facilitam o deslocamento entre as demais abas do relatório.

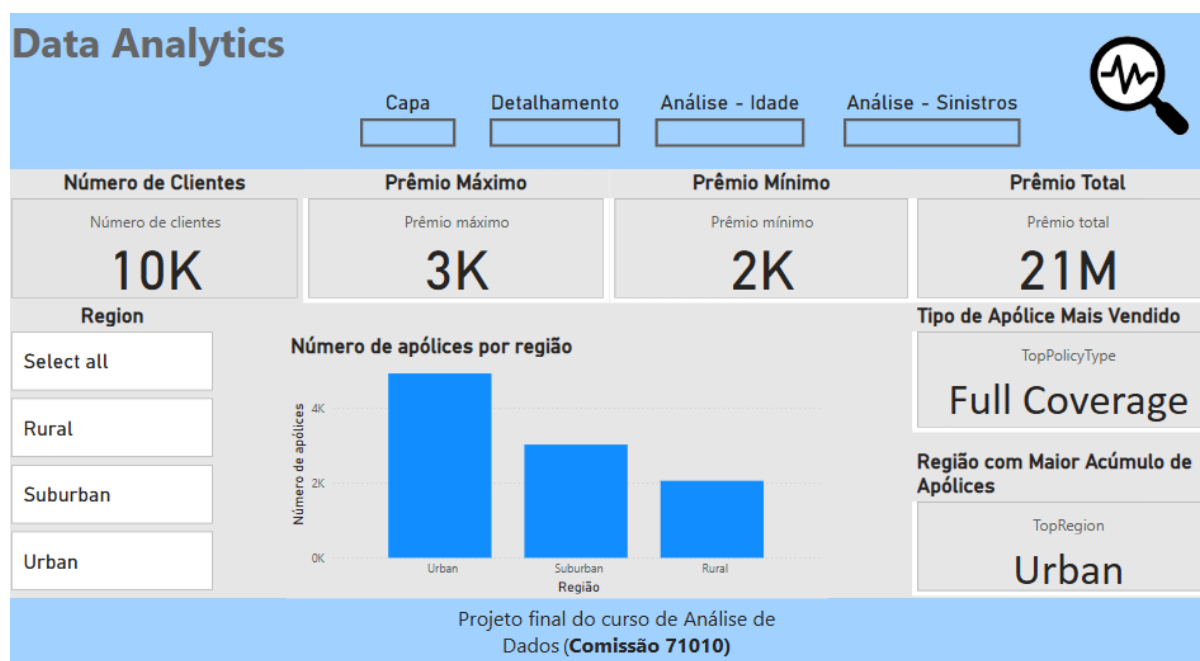


2. Aba "KPIs"

A aba "KPIs" apresenta os principais indicadores de desempenho (Key Performance Indicators), que fornecem uma visão geral do portfólio da seguradora. Entre as informações destacadas estão:

- Os valores máximos e mínimos dos prêmios,
- O número total de segurados,
- O tipo de apólice mais vendida,
- E a região com maior número de segurados.

Esses indicadores possibilitam uma análise de alto nível sobre o desempenho da carteira da companhia de seguros.



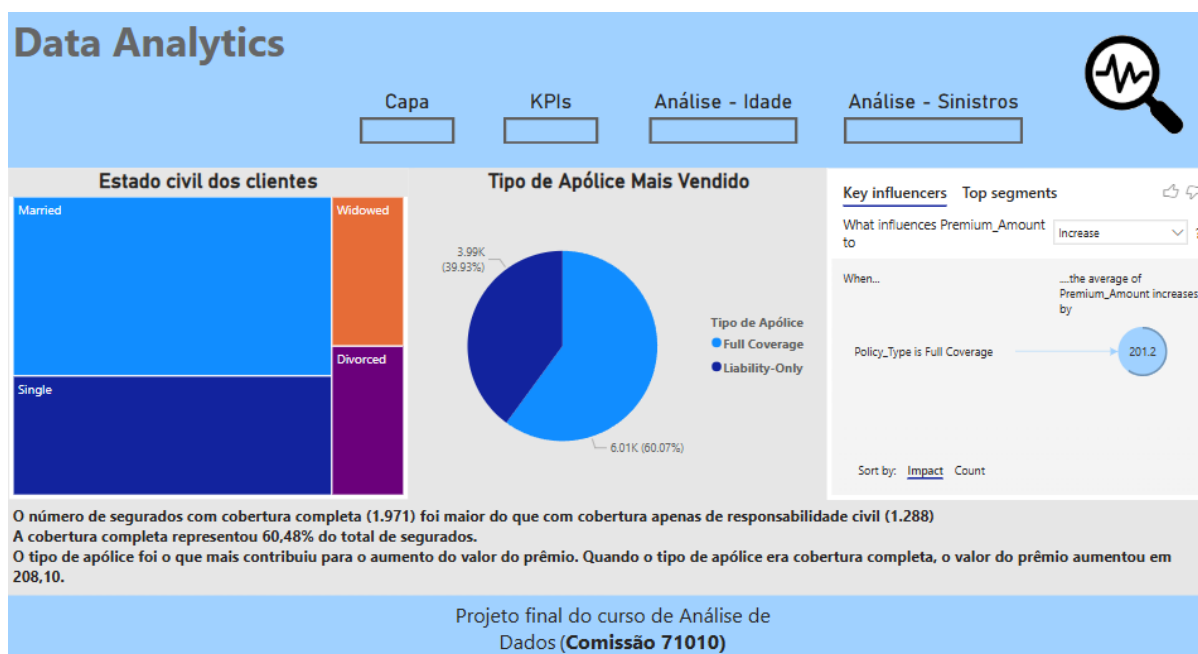
3. Aba "Detalhamento"

A aba "Detalhamento" contém:

- Um gráfico que mostra a proporção de segurados de acordo com o estado civil, permitindo compreender preferências comportamentais dos segurados.
 - Por exemplo, observa-se que à medida que uma pessoa se casa, aumenta a necessidade de adquirir um seguro como forma de proteção contra possíveis perdas financeiras.



- Também é evidenciado que a apólice mais vendida é a "Cobertura Total", o que indica a preferência dos segurados por uma cobertura mais abrangente.
- Além disso, foi incluído um gráfico interativo de "Influenciadores Principais" (Key Influencers), que mostra os fatores que mais impactam no aumento ou redução do valor do prêmio, podendo ser classificados por impacto ou quantidade de ocorrências.

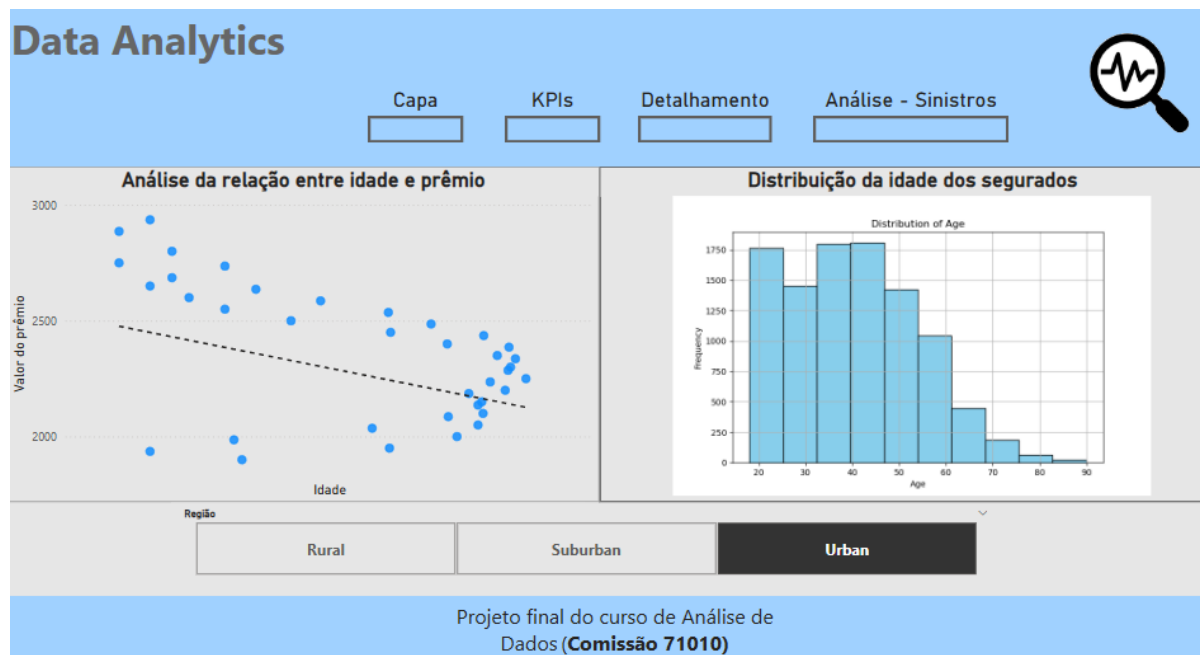


4. Aba "Análise - Idade"

A aba "Análise - Idade" traz uma análise detalhada da variável Idade:

- Mostra a relação entre idade e valor do prêmio cobrado,
- Apresenta a distribuição da idade entre todos os segurados,
- E permite filtrar os dados por região, facilitando a identificação de padrões por localização geográfica.





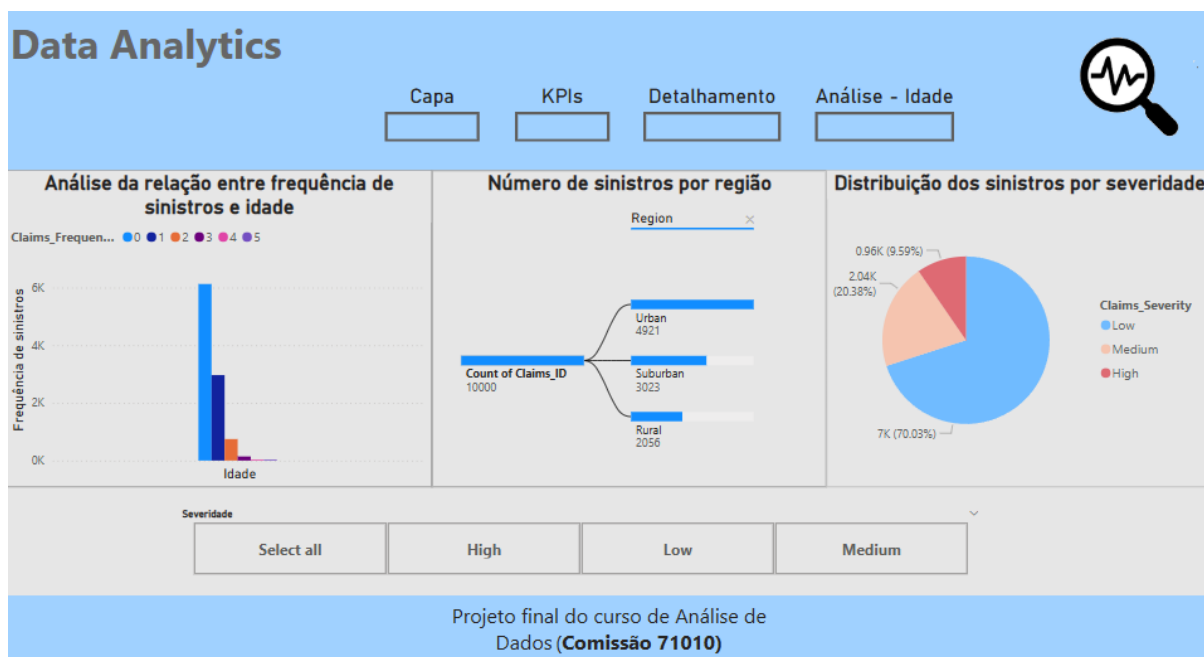
5. Aba "Análise - Sinistros"

A aba "Análise - Sinistros" oferece uma visão abrangente sobre:

- A relação entre idade e frequência de sinistros,
- O número de sinistros por região,
- E a distribuição dos sinistros por grau de severidade.

Para refinar a análise, foi adicionado um filtro interativo (slicer) que permite segmentar os dados conforme a gravidade dos sinistros.





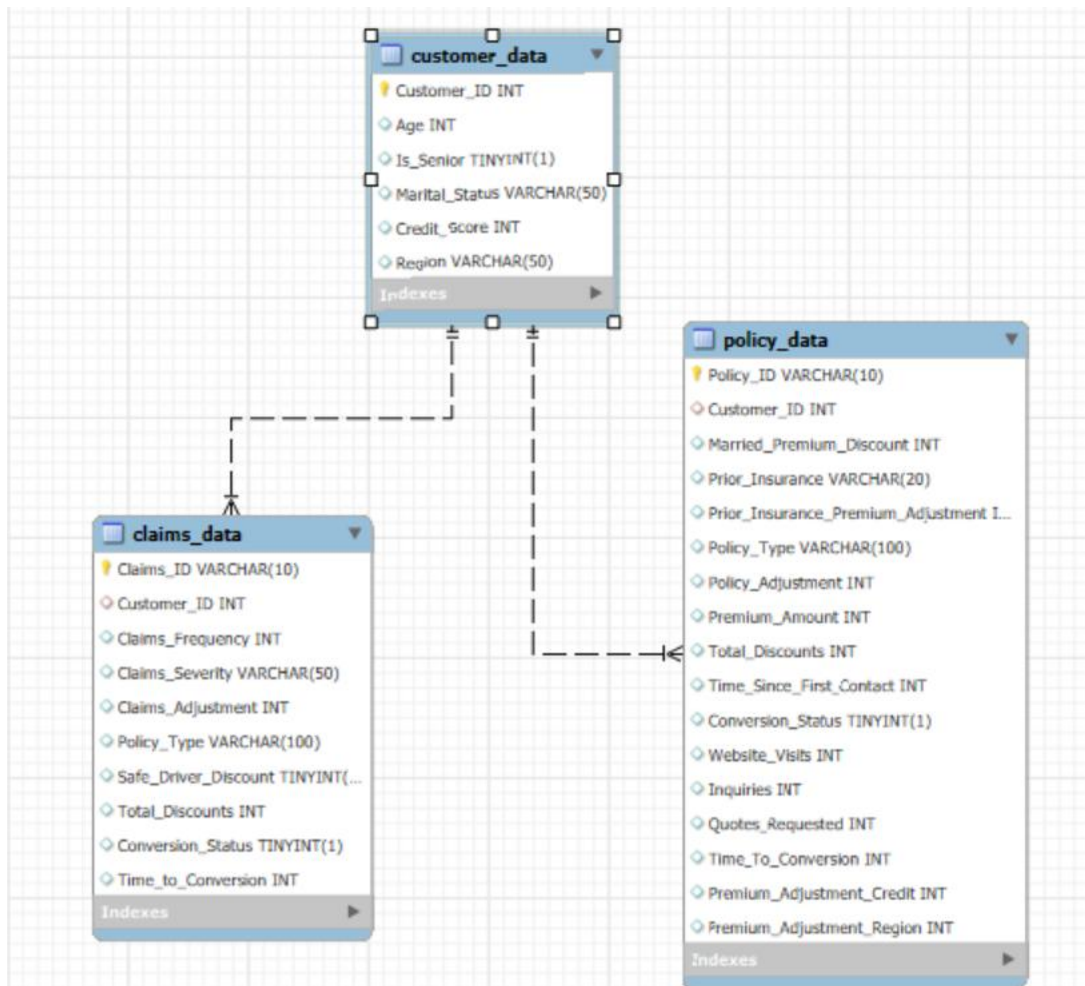
O dashboard criado no Power BI proporciona uma visão completa e interativa dos dados de seguros utilizados no projeto. Com suas cinco abas temáticas, permite a exploração de diferentes dimensões do comportamento dos segurados, tendências de sinistros e critérios de precificação de prêmios, facilitando a tomada de decisão estratégica por parte das seguradoras.

Diagrama de Relacionamento — Modelagem de Dados no Power BI

Durante o processo de análise, o diagrama entidade-relacionamento (ER) no Power BI passou por ajustes significativos em relação à estrutura inicial apresentada no início do projeto. À medida que a análise evoluiu, tornou-se necessário aprimorar o modelo de dados para garantir maior flexibilidade, escalabilidade e precisão nas visualizações e cálculos desenvolvidos.

Diagrama **Entidade-Relacionamento (ER)** utilizando linguagem SQL:





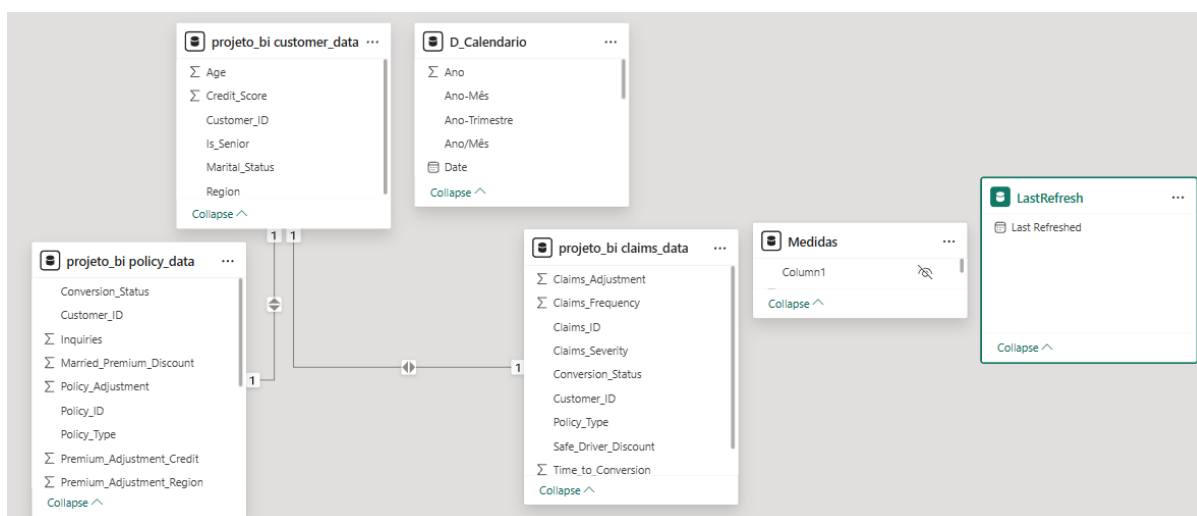
Como parte dessa evolução, foram criadas novas tabelas auxiliares que desempenharam um papel fundamental no apoio à análise e construção do dashboard:

- **D_Calendario:** Tabela de calendário utilizada para permitir análises temporais, como filtragem por data, agrupamentos por ano, mês e dia, além de permitir a correta vinculação de fatos com dimensões de tempo.
- **Medidas:** Tabela dedicada exclusivamente à centralização de medidas DAX (KPIs), proporcionando uma organização mais eficiente do modelo e facilitando a manutenção e reutilização dos cálculos em diferentes visualizações.
- **LastRefresh:** Tabela criada para registrar a data e hora da última atualização do relatório, permitindo que o usuário final visualize com clareza a atualidade das informações apresentadas no dashboard.

Com essas alterações, o modelo final de dados se tornou mais robusto e alinhado às melhores práticas de modelagem dimensional no Power BI.



A seguir, apresentamos o diagrama **entidade-relacionamento (ER)** final do projeto:



Resumo Executivo

O dashboard final para a análise de dados de seguros foi desenvolvido com sucesso e encontra-se organizado em cinco abas principais: **Capa, KPIs, Detalhamento, Análise - Idade e Análise - Sinistros**. Cada uma dessas seções apresenta informações cruciais que auxiliam na avaliação dos dados utilizados, com o objetivo de identificar quais atributos dos segurados influenciam diretamente no valor do prêmio e como o número de sinistros está distribuído entre diferentes regiões e níveis de gravidade.

Este relatório teve como objetivo principal identificar os fatores que são indicadores relevantes no cálculo dos prêmios de seguro, em relação ao risco potencial que esses fatores representam para as seguradoras.

Análise Realizada

Durante a análise, observamos que variáveis como população da região, faixa etária, incidência e gravidade dos sinistros possuem grande influência na determinação do valor do prêmio. Além disso, comportamentos humanos, como direção segura e estado civil, também impactam significativamente o valor final, sendo considerados para a aplicação de descontos adicionais ou acréscimos (loadings) por parte das seguradoras.

Através de diversos cálculos e visualizações, conseguimos identificar relações importantes que confirmam a hipótese inicial:

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



No entanto, ao longo da análise, o foco foi redirecionado para a relação entre idade e o valor do prêmio, a qual se mostrou ser um fator ainda mais determinante do que a pontuação de crédito do segurado, inicialmente considerada como foco principal.

A análise visual demonstra que, conforme a idade aumenta, o valor do prêmio tende a diminuir. Essa descoberta fornece às seguradoras uma base sólida para ajustar suas estratégias de marketing e personalizar campanhas para diferentes faixas etárias, de forma mais eficaz.

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.

Conclui-se que fatores como idade, região, incidência de sinistros e comportamentos individuais exercem influência direta na precificação de prêmios de seguros. O relatório evidencia que a idade do segurado tem um papel mais relevante na determinação do valor do prêmio do que outros fatores inicialmente considerados, como a pontuação de crédito.

Esses insights permitem que as seguradoras adotem abordagens mais estratégicas na segmentação de clientes e na precificação personalizada, visando mitigar riscos e otimizar os resultados operacionais.

