

1. Objetivo do projeto

Analisar o perfil de risco de segurados da categoria tarifária "Automóveis comuns (passeio nacional e importado)" utilizando dados históricos de sinistros. O projeto busca identificar características de risco por meio de variáveis como tipo de veículo, perfil do cliente e região, com o intuito de otimizar a seleção de riscos, minimizar perdas financeiras e melhorar a subscrição de seguros. Além disso, garantirá a confiabilidade das informações por meio de etapas de limpeza, tratamento e análise exploratória dos dados, resultando em insights acionáveis para a tomada de decisões.

2. A fonte de dados utilizada e suas características

Os dados utilizados neste projeto são provenientes da base de dados AutoSeg, disponibilizada pela SUSEP. Para a análise, foram utilizados dados históricos de sinistros e prêmios de seguros de automóveis entre os anos de 2015 e 2019. A principal fonte de dados foi a tabela **arq_casco_comp**, que contém informações detalhadas sobre sinistros, prêmios, e características dos veículos segurados, permitindo uma análise abrangente e fundamentada do perfil de risco dos segurados.

3. Qual é a categoria tarifária que você irá analisar?

"Automóveis comuns (passeio nacional e importado)" - Essa escolha é justificada pelo tamanho da base de dados, que é ampla e permite uma análise mais robusta dos sinistros nessa categoria.

4. Dentro dessa categoria tarifária, quais riscos você não aceitará de acordo com a Política de Subscrição da sua empresa?

- **Veículos com mais de 10 anos:**
Veículos mais antigos tendem a ter maior probabilidade de falhas mecânicas e custos de manutenção elevados. Além disso, a desvalorização pode afetar a compensação em caso de sinistro, aumentando o risco para a seguradora.
- **Pessoas com menos de 25 anos:**
Motoristas mais jovens, especialmente aqueles com menos de 25 anos, estatisticamente apresentam maior índice de acidentes e sinistros. Essa faixa etária é considerada de maior risco devido à inexperiência ao volante e comportamentos de direção mais arriscados.
- **Carros que não sejam de passeio ou para passeio:**
Excluindo veículos comerciais e outros tipos de automóveis, a análise se concentra em carros de passeio, que possuem perfis de uso e riscos diferentes. Essa abordagem ajuda a garantir que a análise seja mais precisa e direcionada, evitando variáveis que poderiam distorcer os resultados.

5. Qual é a experiência histórica que você irá utilizar? Por que você escolheu esses períodos?

Utilizei os dados de 2015 a 2019, excluindo 2020 devido à pandemia de COVID-19, que poderia distorcer os números de sinistros devido a lockdowns e outras restrições. Essa escolha permite uma análise mais precisa das tendências históricas sem a interferência de anomalias externas.

6. Como você avaliou a razoabilidade dos dados?

A razoabilidade dos dados foi avaliada verificando a quantidade de sinistros por ano, região e tipo de veículo. Valores muito fora do padrão foram ajustados para evitar distorções, e conferi se as informações de prêmios e sinistros estavam consistentes e faziam sentido.

7. Quais transformações você realizou para melhorar a confiabilidade da informação?

- Limpeza de dados: Exclusão de registros duplicados e tratamento de valores ausentes, especialmente nas colunas de prêmios e sinistros.
- Tratamento de outliers: Valores extremos, como prêmios ou sinistros muito altos ou baixos, foram analisados e, quando identificados como erros ou situações atípicas não representativas, foram removidos ou ajustados. Isso evitou que esses outliers distorcessem as médias e outros indicadores chave durante a análise.
- Criação de novas variáveis: Foram derivadas variáveis de risco, como "frequência de sinistros por segurado", para melhorar a análise.
- Unificação de dados: Os dados dos semestres foram unificados em anos, facilitando a visualização e a análise de tendências.

8. Qual é o tamanho da sua amostra?

Após filtrar de acordo com os riscos que não estou disposto a aceitar, a amostra ficou com 71.076 registros. Verifiquei que talvez haja um erro após realizar os filtros, então o tamanho da amostra pode sofrer alterações futuras.

9. Você acredita que seus dados têm credibilidade, tanto em relação ao tamanho da amostra quanto em relação a sua qualidade?

Sim, os dados têm credibilidade. A amostra é representativa, cobrindo 2015 a 2019, e foi tratada para minimizar inconsistências. Além disso, as fontes, como a SUSEP, são confiáveis, o que reforça a qualidade das informações, apesar de suspeitar que a amostra possa ser alterada, ainda sim ela me possibilitou realizar análises significantes.

10. Existe alguma limitação dessa informação que você deve documentar e comunicar?

As limitações incluem:

- Dificuldade na categorização dos veículos por marca e modelo, apesar de conseguir vincular as informações de COD_TARIF com a coluna CODIGO da table auto_cat, me sinto inseguro com a qualidade dos dados após o MERGE.
- Dificuldade em aplicar filtros para a categoria tarifaria escolhida e os riscos que não serão aceitos.

<https://github.com/victorvsaraujo/AutoSeg-Risk-Evaluation>