



Universidade Federal do Rio de Janeiro

Escola Politécnica

MBA em Big Data, Business Intelligence e Business Analytics  
(MB3B)

**BENEFÍCIOS DA UTILIZAÇÃO DE ANÁLISES DE DADOS,  
INDICADORES DE PERFORMANCE E MINERAÇÃO DE DADOS NO  
MERCADO DE SEGUROS**

Autor:

---

Luciano Bastos Grangeiro Filho, B. Sc.

Orientador:

---

Manoel Villas Boas Junior, M. Sc.

Coorientador:

---

José Otávio Santi Moreira, M. Sc.

Examinador:

---

Norberto Ribeiro Bellas, M. Sc.

Examinador:

---

Vinicius Drumond Gonzaga, M. Sc.

**Rio de Janeiro  
Agosto de 2023**

## Declaração de Autoria e de Direitos

Eu, **Luciano Bastos Grangeiro Filho** CPF 663.830.803-34, autor da monografia ***BENEFÍCIOS DA UTILIZAÇÃO DE ANÁLISES DE DADOS, INDICADORES DE PERFORMANCE E MINERAÇÃO DE DADOS NO MERCADO DE SEGUROS***, subscrevo para os devidos fins, as seguintes informações:

1. O autor declara que o trabalho apresentado na defesa da monografia do curso de Pós-Graduação, Especialização MBA em Big Data, Business Intelligence e Business Analytics da UFRJ é de sua autoria, sendo original em forma e conteúdo.
2. Excetuam-se do item 1 eventuais transcrições de texto, figuras, tabelas, conceitos e ideias, que identifiquem claramente a fonte original, explicitando as autorizações obtidas dos respectivos proprietários, quando necessárias.
3. O autor permite que a UFRJ, por um prazo indeterminado, efetue em qualquer mídia de divulgação, a publicação do trabalho acadêmico em sua totalidade, ou em parte. Essa autorização não envolve ônus de qualquer natureza à UFRJ, ou aos seus representantes.
4. O autor declara, ainda, ter a capacidade jurídica para a prática do presente ato, assim como ter conhecimento do teor da presente Declaração, estando ciente das sanções e punições legais, no que tange a cópia parcial, ou total, de obra intelectual, o que se configura como violação do direito autoral previsto no Código Penal Brasileiro no art.184 e art.299, bem como na Lei 9.610.
5. O autor é o único responsável pelo conteúdo apresentado nos trabalhos acadêmicos publicados, não cabendo à UFRJ, aos seus representantes, ou ao(s) orientador(es), qualquer responsabilização/ indenização nesse sentido.
6. Por ser verdade, firmo a presente declaração.

Rio de Janeiro, 05 de agosto de 2023.

---

Luciano Bastos Grangeiro Filho

## **UNIVERSIDADE FEDERAL DO RIO DE JANEIRO**

Av. Athos da Silveira, 149 - Centro de Tecnologia, Bloco H, sala - 212,  
Cidade Universitária Rio de Janeiro – RJ - CEP 21949-900.

Este exemplar é de propriedade Escola Politécnica da Universidade Federal do Rio de Janeiro, que poderá incluí-lo em base de dados, armazenar em computador, microfilmear ou adotar qualquer forma de arquivamento.

Permitida a menção, reprodução parcial ou integral e a transmissão entre bibliotecas deste trabalho, sem modificação de seu texto, em qualquer meio que esteja ou venha a ser fixado, para pesquisa acadêmica, comentários e citações, desde que sem finalidade comercial e que seja feita a referência bibliográfica completa.

Os conceitos expressos neste trabalho são de responsabilidade do(s) autor(es).

## **DEDICATÓRIA**

Este trabalho é dedicado às mulheres da minha vida. Minha mãe, que sempre acreditou em mim e me incentivou. E minha esposa que tem sido o alicerce da nossa família e me suporta em todas as minhas decisões, assim como embarca em todas as aventuras que aparecem em nossas vidas. Esse suporte e coragem me impulsionam a ser uma pessoa melhor, e espero que consiga fazer por você parte do que faz por mim e por nossa família.

## **AGRADECIMENTO**

Assim como às dedico esse trabalho, também agradeço minha esposa e a minha mãe por todos os sacrifícios, carinho e amor. Agradeço aos meus filhos Benicio e Pedro por todos os dias me ensinarem o que é o amor em sua forma mais pura, sem esperar nada em troca. Agradeço a Deus por todos os desafios e oportunidades que ele coloca em minha jornada diariamente. E, por fim, agradeço aos meus orientadores José Otávio Santi Moreira e Manoel Villas Boas Junior pela compreensão durante o processo de conclusão deste trabalho, e por me guiarem com orientações e sugestões valiosas.

## RESUMO

Este trabalho tem como propósito apresentar soluções para a geração de inteligência de mercado e de melhoria contínua para suportar uma seguradora de veículos a prosperar por meio da análise de dados. As análises focaram na utilização de ferramentas e sistemas de inteligência de negócios para determinar a probabilidade da ocorrência de sinistros de veículos em contratos já existentes e em adições à carteira de clientes gerando dessa forma não só melhorias operacionais, mas também financeiras.

A inteligência de mercado tem uma relação direta com os resultados e prosperidade das seguradoras, pois podem permitir que as empresas foquem em um planejamento estratégico adequado e elaborem indicadores de desempenho para acompanhar a eficiência operacional, nesse caso nas concessões de seguros. Para a elaboração deste trabalho foi utilizado um conjunto de dados *open-source* com mais de 58 mil apólices de seguro extraído da plataforma *Kaggle*.

Esse tema foi escolhido para confirmar a adaptabilidade da inteligência de negócios e a confiança nas análises de dados, além dos resultados destas que foram de extrema importância para que a conclusão alcançada fosse significativa e relevante.

As análises propostas têm como base diversos dados relacionados à segurança dos veículos explorados, além de dados dos segurados e apólices. As ferramentas e sistemas utilizados para essas análises servirão para retroalimentar as informações e indicadores quando da existência de novos dados, auxiliando na avaliação sobre o grau de assertividade do modelo, e por consequência gerando dados para avaliação dos tomadores de decisão.

Através das ferramentas, sistemas e análises propostas neste trabalho a expectativa é que a seguradora aumente sua acuracidade na concessão de seguros em mais de 30%, o que por consequência irá auxiliar na confirmação de uma *performance* operacional e financeira mais consistente.

Palavras-Chave: Inteligência de mercado, Análise, Dados, Estratégico.

## **ABSTRACT**

This paper's purpose is to present answers and solutions focusing on how continuous improvement and business intelligence can support a vehicle insurance company to thrive through data analysis. The analyses performed focused on utilizing business intelligence tools and systems to determine the chance of potential insurance claims of current contacts or new additions generating not only operational, but also financial improvements.

The business intelligence explored has a direct relation with the results and prosperity of companies, as they allow the focus on an adequate strategic planning and the creation of performance indicators to follow-up on operational efficiency, in this case on insurance concession. The elaboration of this paper was based on an open-source data set with over fifty-eight thousand insurance contracts that was extracted from the Kaggle platform.

The subject-matter of this paper was chosen focusing on confirming the adaptability of business intelligence practices and the reliance in the data analysis and its outcome, which can become of extreme importance to the achieved conclusions, and for its relevance in the insurance market – confirming that the practices explored are not dependent of processes and areas.

The analysis suggested have a significant amount of data as basis. The tools and systems utilized in the analysis will feed the information and indicators in the existence of new data, supporting the assessment about the assertiveness of the model and, by consequence generating new data for the decision makers to evaluate.

Through the tools, systems and analysis here proposed it is expected that the insurance company increase its accuracy in more than 30% when awarding insurance policies, securing an improved operational and financial results as consequence.

Keywords: Business intelligence, Analysis, Data, Strategy.

## SIGLAS

<b>ABNT</b>	Associação Brasileira de Normas Técnicas
<b>BI</b>	<i>Business Intelligence</i>
<b>BPM</b>	<i>Business Process Management</i>
<b>CNseg</b>	Confederação Nacional das Empresas de Seguros Gerais, Previdência Privada e Vida
<b>DSS</b>	<i>Decision Support System</i>
<b>DW</b>	<i>Data Warehouse</i>
<b>ERP</b>	<i>Enterprise Resource Planning</i>
<b>ETL</b>	<i>Extract, Transform &amp; Load</i>
<b>KPI</b>	<i>Key Performance Indicator</i>
<b>MBA</b>	<i>Master in Business Administration</i>
<b>MB3B</b>	<i>MBA em Big Data, Business Intelligence e Business Analytics</i>
<b>SUSEP</b>	Superintendência de Seguros Privados
<b>UFRJ</b>	Universidade Federal do Rio de Janeiro

## LISTA DE FIGURAS

<b>Figura 2.1</b>	Partes envolvidas um seguro	7
<b>Figura 2.2</b>	Participações de mercado brasileiro de seguros	8
<b>Figura 2.3</b>	Componentes da inteligência de negócios	9
<b>Figura 2.4</b>	Ciclo de vida <i>Data Warehouse</i>	11
<b>Figura 2.5</b>	Evolução da análise de dados	12
<b>Figura 2.6</b>	Tipo de análise de dados	12
<b>Figura 2.7</b>	ETL <i>Power Query</i>	14
<b>Figura 3.1</b>	Demonstrativo de 5, das 44 colunas, da base “ <i>Car Insurance Claim Prediction</i> ”	18
<b>Figura 3.2</b>	Tradução de dados das células em editor de <i>query</i>	21
<b>Figura 3.3</b>	Tradução de dados booleanos em editor de <i>query</i>	21
<b>Figura 3.4</b>	Conclusão da fase <i>Load</i>	22
<b>Figura 3.5</b>	Relações entre estudo adicional e bases de dados	25
<b>Figura 3.6</b>	Ferramenta de análises <i>Power BI</i>	27
<b>Figura 3.7</b>	Demonstração de gráfico do <i>Power BI</i>	28
<b>Figura 3.8</b>	Demonstração da linguagem M	29
<b>Figura 4.1</b>	Outliers apólices de seguros acima da duração estabelecida	33
<b>Figura 4.2</b>	Análise e exploração dos dados gerais, base para demais análises	35
<b>Figura 4.3</b>	Dados de natureza pessoal analisados	36
<b>Figura 4.4</b>	Conexão entre idade e aumento nos sinistros	37
<b>Figura 4.5</b>	Medias de sinistros por densidade	37
<b>Figura 4.6</b>	Aderência das análises de sinistralidade por cidade e idade	38
<b>Figura 4.7</b>	Aderência dos dados dos veículos à média geral	39
<b>Figura 4.8</b>	Modelo com a idade de 5 anos e seu impacto nos dados dos veículos	40
<b>Figura 4.9</b>	Modelo com os itens de segurança avaliados	41
<b>Figura 4.10</b>	Situações neutras	41
<b>Figura 4.11</b>	Situações esperadas	42
<b>Figura 4.12</b>	Exemplo de indicadores resultantes das análises	43
<b>Figura 5.1</b>	Modelo da Linguagem M executada em nova carteira de clientes	45
<b>Figura 5.2</b>	Resultados alcançados com o aprendizado de máquina	46

## LISTA DE TABELAS

<b>Tabela 3.1</b>	Cabeçalhos e validação dados	19
<b>Tabela 3.2</b>	Demonstração de perguntas a serem avaliadas	26
<b>Tabela 4.1</b>	Análise quantitativa dos dados booleanos	31

# Sumário

<b>Capítulo 1: Introdução.....</b>	<b>01</b>
1.1 – Tema .....	02
1.2 - Justificativa .....	02
1.3 – Objetivos.....	03
1.4 - Delimitação .....	03
1.5 – Metodologia.....	04
1.6 – Descrição .....	04
<b>Capítulo 2: Embasamento Teórico .....</b>	<b>06</b>
2.1 – Ciência Atuarial em seguros .....	07
2.2 – <i>Business Intelligence</i> (BI).....	08
2.3 – <i>Data Warehouse e Data Flow</i> .....	10
2.4 – Análises de dados .....	11
2.4.1 – Análises Descritivas .....	13
2.4.2 – Análises Preditivas .....	13
2.4.3 – Análises Prescritivas.....	13
2.5 – Ferramentas para análises de dados .....	13
2.5.1 – <i>Power BI</i> .....	13
2.6 – Indicadores de <i>Performance</i> .....	15
2.6.1 – Indicadores Estratégicos .....	15
2.6.2 – Indicadores Táticos.....	15
2.7 – Data Mining.....	16
<b>Capítulo 3: Propostas Tecnológicas .....</b>	<b>17</b>
3.1 – Avaliação da data set .....	17
3.1.1 – <i>Extract, Transform and Load</i> .....	19
3.1.2 – Análise da base de dados disponível .....	22
3.1.3 – Aplicação das técnicas para análises de dados .....	25
3.2 – Utilizando o <i>Power BI</i> como ferramenta de inteligência. ....	26
3.3 – Ajustando o modelo de aprendizado da máquina.....	28
<b>Capítulo 4: Resultados Obtidos.....</b>	<b>30</b>
4.1 – Dados desconsiderados das análises.....	30
4.2 – Análises de booleanos duplicados ou repetidos .....	31
4.3 – Análises de outliers.....	32
4.4 – Análises exploratórias dos dados.....	33
4.4.1 – Análises dos dados gerais da base de dados .....	34
4.4.2 – Análises dos dados de natureza pessoal .....	35
4.4.3 – Análises dos dados do veículo.....	38
4.4.4 – Análises dos dados de segurança.....	40
4.5 – Indicadores de <i>performance</i> .....	42
<b>Capítulo 5: Conclusão e Trabalhos Futuros .....</b>	<b>44</b>
5.1 – Conclusão .....	44

5.2 – Trabalhos Futuros .....	46
<b>Referências Bibliográficas .....</b>	<b>47</b>

# CAPÍTULO 1

## Introdução

O projeto aqui proposto tem como objetivo explorar os benefícios existentes nas análises de dados em suas diversas vertentes e como essas análises alimentam os processos e decisões das empresas, de forma a gerar maior confiança nas informações para a tomada de decisões.

O objeto de análise em questão será uma empresa fictícia de seguros, Seguradora Fictícia S.A. A decisão por utilizar uma empresa fictícia se suportou na severidade e confidencialidade dos dados existentes em empresas do ramo, que envolvem desde questões estratégicas a questões legais relacionadas à privacidade de dados.

Embora baseado em uma empresa fictícia, o projeto buscou ser aderente às possibilidades do dia a dia de empresas do ramo e, por isso, utilizou-se de uma base de dados *open-source*<sup>1</sup> de seguros contendo informações reais. A base de dados chamada “*Car Insurance Claim Prediction*” está disponível na plataforma *Kaggle*<sup>2</sup> através do endereço: [kaggle.com/datasets/ifteshanajnin/carinsuranceclaimprediction-classification?select=train.csv](https://kaggle.com/datasets/ifteshanajnin/carinsuranceclaimprediction-classification?select=train.csv) e havia sido atualizada 6 meses antes da elaboração desse trabalho.

A plataforma *Kaggle* se destaca entre as comunidades para assuntos relacionados a Ciência de Dados. Na plataforma são disponibilizados *data sets*<sup>3</sup>, dicas, fóruns e vários outros recursos relacionados à Ciência de Dados. As informações da base de dados “*Car Insurance Claim Prediction*” estavam disponíveis em inglês e se referiam a outra localidade, que não é esclarecida pelo autor do data set. De acordo com sua descrição, a base de dados contém informações sobre os segurados e seguros, como idades do carro e proprietário, densidade populacional da cidade e outras, focando na disponibilização de recursos para identificar o risco associado a sinistros nos meses subsequentes, mas sem entrar em informações financeiras estratégicas.

As informações disponibilizadas na base de dados foram adaptadas ao contexto nacional para maior compreensão e percepção, frente aos resultados alcançados.

---

<sup>1</sup> Open-source: Se refere a código aberto, ou seja, que pode ser visualizada por qualquer pessoa.

<sup>2</sup> Kaggle: Se refere à plataforma de datasets open-source disponível em <https://www.kaggle.com/>.

<sup>3</sup> Data set: Coleções de dados ou base de dados.

## 1.1 – Tema

O tema central deste projeto foca na utilização de várias das ferramentas de *Business Intelligence*<sup>4</sup> e *Business Analytics*<sup>5</sup>, que são temas centrais explorados durante este curso de MBA (*Master in Business Administration*).

A proposta é que através das análises de dados realizadas em ferramentas de inteligências de negócios, seja alcançado um patamar onde a tomada de decisões seja beneficiada pelas informações já processadas, aumentando a acuracidade e sucesso em novos contratos.

Adicionalmente, esses resultados continuarão a ser explorados através do conceito de *Data mining*<sup>6</sup>, onde a estrutura das análises previamente executadas será automaticamente aplicada a novas entradas na base de dados visando um resultado semelhante em termos de análises.

## 1.2 – Justificativa

A abordagem da inteligência de negócios visa transformar dados em informações, que se transformam em decisões e por fim ações (SHARDA; DELEN; TURBAN, 2019). Desta forma a inteligência de negócios deve, e precisa, ser explorada em diversos ambientes para demonstrar sua capacidade.

A área de seguros vem sendo tema de exploração contínua da ciência atuarial, que visa garantir a solvência financeira e econômica das seguradoras (SILVA, 2022). No entanto, a beleza da inteligência de negócios está no fato de que esta pode ser aliada a qualquer ciência, inclusive a atuarial.

Desta forma o tema aqui proposto visa demonstrar como é possível alcançar resultados expressivos em diversas áreas de negócios ao aplicar com afinco as diversas possibilidades que existem nas análises de dados e ferramentas de inteligência de negócios inclusive aumentando os desafios quando se exploram áreas não previamente conhecidas ou amplamente discutidas, como a área de seguros.

---

<sup>4</sup> Business Intelligence: Se refere à inteligência de negócios que envolve desde coleta a compartilhamento de informações.

<sup>5</sup> Business Analytics: Se refere à análise de negócios e a geração de informações para tomada de decisão.

<sup>6</sup> Data Mining: Se refere à mineração de dados estratégicos.

### 1.3 – Objetivos

O objetivo geral deste projeto é demonstrar como os tomadores de decisão de empresas no mercado de seguros de veículos podem se beneficiar com a aplicação de análises de dados internos e indicadores de *performance* de forma contínua, através do benefício do incremento da mineração de dados.

Para alcançamos tal resultado, foram explorados diversos componentes da inteligência de negócios, desde *Data Warehouses*<sup>7</sup> até *Dashboards*<sup>8</sup>. Dentro destes componentes, o objetivo do projeto foi de explorar as diferentes etapas existentes de forma a demonstrar de forma realista a aplicação de um projeto desta magnitude.

Além de demonstrar fidedignamente as etapas de um projeto de inteligência de negócios, buscamos alinhar esse conceito com o de análises de dados, principalmente as análises descritivas, que focam em explorar o que já ocorreu, e preditivas, que focam em identificar o que vai acontecer com base em dados existentes.

Para ambos os casos buscamos resultados que fossem demonstrados de forma clara e objetiva, através de indicadores de *performance*, para que diante de uma necessidade de decisão houvesse o máximo de clareza o possível.

Assim como as etapas até aqui foram de extrema importância, o objetivo do projeto sempre foi o da melhoria contínua e geração de informação relevante. Para isso, nada mais adequado do que o foco que foi mantido em como tornar as etapas anteriores práticas e com o mínimo de interação possível. Esse objetivo foi suportado pela busca constante de ferramentas que permitissem uma mineração de dados eficiente, suportando situações futuras.

### 1.4 – Delimitação

Esse projeto focou única e exclusivamente em análises do mercado de seguros de veículos, que é o segundo maior no Brasil. As demais áreas de atuação deste mercado não foram consideradas nas análises aqui detalhadas.

Adicionalmente, vale reforçar que os dados aqui inseridos são dados de natureza *open-source* e a empresa referente fictícia, simulando da forma mais próxima o possível à realidade

---

<sup>7</sup> Data Warehouse: Se refere a um sistema de armazenamento de grandes volumes de dados de várias fontes diferentes.

<sup>8</sup> Dashboard: Se refere a um painel visual que apresenta um conjunto de informações de forma centralizada.

do mercado local. Ainda assim, algumas limitações como modelos de veículos e área de atuação existem e foram observadas, como veículos não encontrados com facilidade no mercado brasileiro e mercado simulado restrito ao estado do Rio de Janeiro. De toda forma, entendemos não haver impacto para o projeto diante da aderência dos dados às análises aplicadas.

## 1.5 – Metodologia

A metodologia utilizada no projeto baseou-se em pesquisa bibliográfica e aplicação de conceitos e teorias identificadas durante a duração do curso, como as dos diferentes tipos de análise de dados como dos diferentes tipos de indicadores de *performance*.

As teorias foram colocadas em prática em todas as etapas do projeto onde foi simulada a existência de um *Data Warehouse*, que deu suporte a um processo de extração, transformação e carga de dados chegando até às análises através de simulações e perguntas chave. Foi também utilizada uma ferramenta de inteligência de negócios de amplo acesso, *Power BI*, que suportou a simulação e criação dos indicadores resultantes das análises realizadas previamente, assim como manteve o registro de toda as atividades executadas durante o projeto para viabilizar a aplicação do aprendizado de máquinas semiautomático.

## 1.6 – Descrição

Este projeto está estruturado em 5 capítulos, sendo este o capítulo introdutório.

No capítulo 2 será explorado o embasamento teórico, onde a revisão bibliográfica demonstrara os diferentes entendimentos e conceitos buscados. Desde entendimentos das origens da atividade de seguros, passando por seus dados e cálculos, atividade de monitoramento, conceitos de inteligência de negócios, *Data Warehouse*, análise de dados dos três principais tipos existentes, diferentes ferramentas de inteligência de negócios e suas atuações, concluindo com os conceitos de indicadores de *performance* e aprendizado de máquina.

No capítulo 3 as propostas para explorar os dados de forma a alcançar o objetivo serão apresentadas. Aqui foram avaliados os dados existentes, de diversas formas, e utilizada a ferramenta de inteligência de negócios escolhida, incluindo os ajustes necessários para o aprendizado e execução semiautomática.

No capítulo 4 apresentamos os resultados obtidos, que foram executados por etapas que são comuns aos processos e análise de dados. Avaliações sobre a relevância dos dados, existência de dados booleanos duplicados ou repetidos, e existência de *outliers*<sup>9</sup> foram realizadas antes da exploração dos dados que resultaram no modelo final.

Por fim, no capítulo 5 foram apresentadas as conclusões obtidas com esse projeto.

---

<sup>9</sup> Outlier: Considerado como um ponto fora da curva, ou do padrão.

# CAPÍTULO 2

## Embasamento Teórico

No Brasil, o mercado de seguros é fiscalizado por uma Autarquia Federal vinculada ao Ministério da Economia - Superintendência de Seguros Privados (SUSEP). Esta autarquia tem como missão estimular o desenvolvimento do mercado de seguros garantindo a livre concorrência, estabilidade do mercado e respeito ao consumidor (SUSEP, 2022).

Duas das principais finalidades da SUSEP é zelar pela liquidez e solvência das sociedades que integram o mercado - Seguradoras; e pela defesa dos interesses dos consumidores de mercados supervisionados (SUSEP, 2022). De acordo com a Confederação Nacional das Empresas de Seguros Gerais, Previdência Privada e Vida, Saúde Suplementar e Capitalização (CNseg), o setor de seguros movimentou R\$ 306,4 bilhões em 2021, um aumento de 11,9% comparado ao período anterior (CNseg, 2022).

O 9º relatório de análise e acompanhamento dos mercados supervisionados da SUSEP, emitido em 2021, indica que o mercado de seguro de automóveis é o segundo maior mercado de seguros no Brasil com 28% de *market share*<sup>10</sup>, o que proporcionalmente representaria um mercado de aproximadamente R\$ 86 bilhões. Em contrapartida, o mercado de seguros automotivos possui individualmente a maior representatividade na taxa de sinistros, chegando a 63,1% dos sinistros registrados em 2021 (CNseg, 2022).

Para evitar prejuízos, as seguradoras atuando no Brasil são autorizadas a estabelecer seus prêmios, ou taxas que serão cobradas aos consumidores, de acordo com suas próprias regras. Nessa etapa a Ciência Atuarial entra em atuação considerando métodos matemáticos, estatísticos, financeiros e demais. Esses dados envolvem desde o perfil do motorista até a chance de uma sinistralidade (Seguro Auto, 2022).

A aplicação da análise de dados a problemas empresariais é uma habilidade-chave em *Business Intelligence* (SHARDA; DELEN; TURBAN, 2019). Neste conceito a análise de dados e o suporte computadorizado suportará as seguradoras em suas decisões, que necessitam cada vez mais de agilidade, rapidez, estratégia e tática para extrair o máximo de informações dos dados e conhecimentos existentes no processo (SHARDA; DELEN; TURBAN, 2019).

---

<sup>10</sup> Market share: Participação no mercado analisado.

## 2.1 – Ciência Atuarial em seguros

De forma direta, o propósito da atividade atuarial é maximizar os lucros das operações de seguros (SILVA, 2022).

A história e criação da ciência atuarial tem conexão direta com o início das atividades de seguros. Esta última teve seu início há mais de dois mil anos através de uma prática voltada para o compartilhamento, ou divisão, dos riscos. Comerciantes passaram a dividir perdas que ocorriam dentro de um grupo, evitando que um indivíduo arcasse sozinho com o prejuízo. Essa prática se expandiu entre comerciantes de alimentos, animais e transportadores, que dividiam a carga para garantir o sucesso da jornada (SILVA, 2022).

Hoje existem diversos tipos de seguros em atuação no Brasil. Normalmente, o mercado de seguros envolve pelo menos três partes: a seguradora; o segurado; e o objeto do seguro. De forma simples, o propósito do seguro é transferir o risco do objeto segurado à seguradora, mantendo assim o segurado e sua atividade fim garantidos, ou protegidos, contra os riscos acordados entre as partes, conforme demonstrado na Figura 2.1.

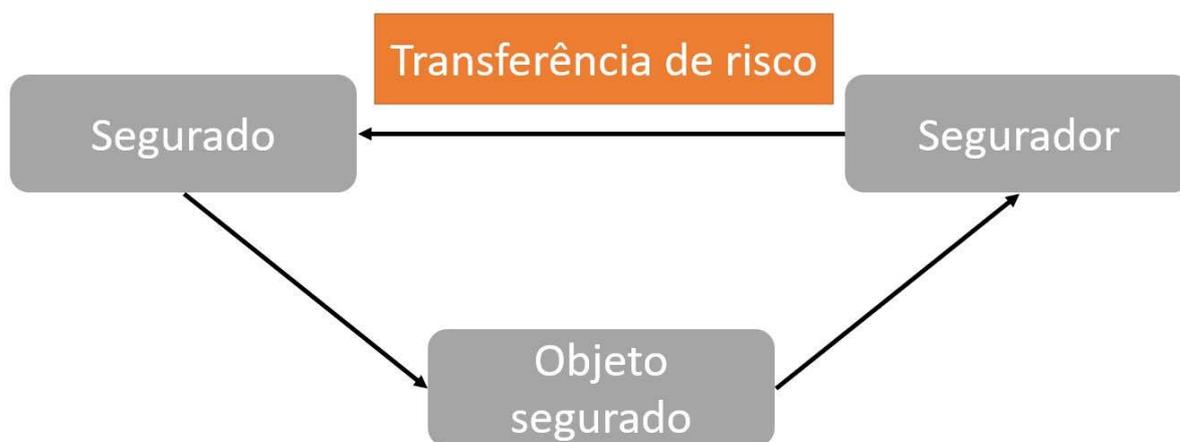


Figura 2.1 – Partes envolvidas um seguro.

Fonte: Elaborada pelo autor., 2023.

A contratação de um seguro envolve vários elementos, sendo os principais: o contrato; o beneficiário (ou segurado); o prêmio, que consiste no valor pago pelo segurado; a franquia, que consiste no valor pago pelo segurado em caso de sinistro para acessar sua cobertura; o sinistro, que consiste na materialização do risco e na cobertura deste pela seguradora; e a indenização, que consiste no valor recebido pelo segurado pelo prejuízo decorrente do sinistro.

Esses elementos são válidos para todos os tipos de seguro, incluindo o seguro de automóveis que é o foco deste projeto.

Conforme demonstrado na Figura 2.2, o seguro de automóveis representa 28% do total do mercado de seguros brasileiro, sendo a segunda modalidade mais representativa. A interseção entre a ciência atuarial e de seguros se dá para garantir a solvência financeira e econômica das seguradoras (SILVA, 2022). Essa garantia pode ser extremamente beneficiada pelo uso da inteligência de negócios através da análise de dados e mineração de dados, associando ainda mais o resultado operacional das seguradoras à eficiência dos modelos utilizados.

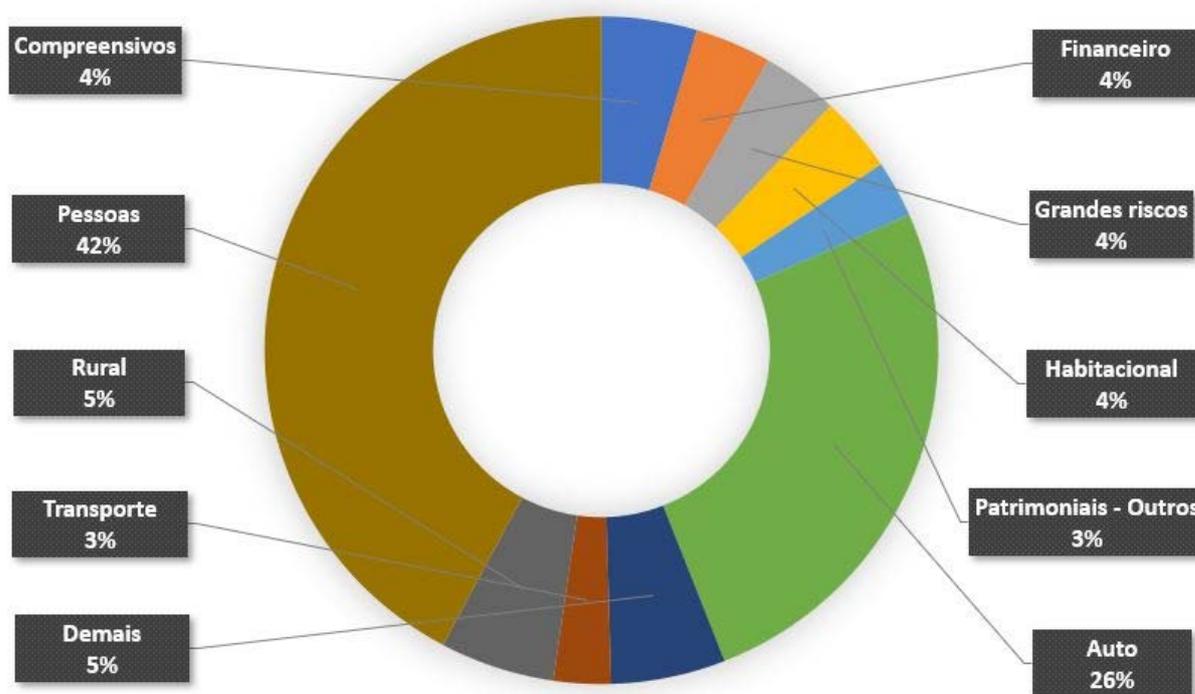


Figura 2.2 – Participações de mercado brasileiro de seguros.

Fonte: Adaptado de SUSEP, 2021, P.16.

## 2.2 – *Business Intelligence* (BI)

Gestores de negócios precisam das informações certas, na hora certa e no lugar certo. Esse antigo mantra foi assumido nas abordagens modernas de inteligência de negócios (*Business Intelligence*).

Inteligência de negócios é um termo guarda-chuva que combina arquitetura, ferramentas, bases de dados, ferramentas analíticas, aplicativos e metodologias utilizadas para extrair informações precisas dos dados disponíveis para as empresas e tomadores de decisão. O principal objetivo da inteligência de negócios é possibilitar o acesso interativo a esses dados, permitindo a manipulação e análises apropriadas, que envolvem desempenhos históricos e atuais, transformando estes dados em informações, que se transformam em decisões e por fim ações (SHARDA; DELEN; TURBAN, 2019).

Um sistema de inteligência de negócios contém quatro componentes principais: um *Data Warehouse* (DW) que armazena dados atuais e históricos de toda a empresa; ferramentas de BI que são utilizadas para manipulação, mineração e análise dos dados existentes no DW; ferramentas de *Business Process Management*<sup>11</sup> (BPM) que busca otimização dos resultados através de análises automatizadas; e uma interface com o usuário que visa facilitar o acesso aos dados já extraídos, analisados, processados e transformados em informação (SHARDA; DELEN; TURBAN, 2019). A dinâmica desses componentes é demonstrada na Figura 2.3.

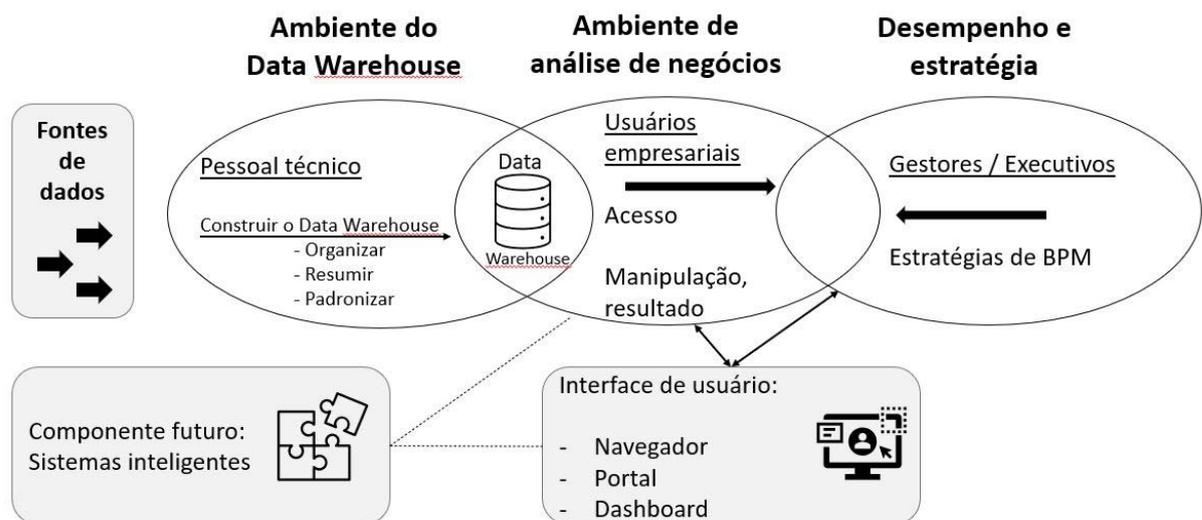


Figura 2.3 – Componentes da inteligência de negócios.

Fonte: *The data Warehousing Institute*, 2003, P. 32.

Pela Figura 2.3 fica claro que não há como pensar em inteligência de negócios sem pensar na fonte de dados, que evoluiu para uma verdadeira engenharia em *Data Warehousing* cada vez mais adaptada às necessidades de cada negócio.

<sup>11</sup> Business Process Management: Conceito que une tecnologia e gestão no gerenciamento de processos.

### 2.3 – *Data Warehouse e Data Flow*

Considerando sua importância para a inteligência de negócios, no que consiste o *Data Warehousing*? Basicamente em construir armazéns de dados onde a história da empresa, seus clientes, fornecedores e operações se mantêm disponíveis e acessíveis para consultas e análises (MACHADO, 2013).

Ainda segundo MACHADO, um projeto de *Data Warehouse* deve buscar os seguintes resultados: 1. Informação disponível para gestão; 2. Visão de curvas de comportamento; 3. Agilidade de ferramentas para apoio à decisão; 4. Maior abrangência de visão de indicadores; 5. Recursos mais abrangentes para análises de negócios; e 6. Necessidades e expectativas atendidas por tecnologia da informação.

Através dos pontos demonstrados acima, dentre outras questões, confirma-se nossa afirmação de que a fonte de dados é essencial para inteligência de negócios. Não só isso, mas atualmente a inteligência de negócios impõe influência significativa em como os *Data Warehouse* são projetados.

A Figura 2.4 demonstra o ciclo de vida, incluindo os comportamentos, de um *Data Warehouse*. Aqui fica visível como ao longo do tempo há uma mudança entre um foco nos recursos utilizados para uma priorização na utilização pelo usuário. *Data Flows*<sup>12</sup>, ou fluxo de dados, são parte integrante desse amadurecimento dos dados através das ferramentas de ETL (*Extract*<sup>13</sup>, *Transform*<sup>14</sup>, *Load*<sup>15</sup>). Uma vez definidos os focos das extrações, transformações e carregamento dos dados as corporações estarão prontas para executar as análises destes em seus sistemas, transformando dados em informações gerenciáveis.

---

<sup>12</sup> Data Flow: Fluxo de dados nos processos de sistemas operacionais.

<sup>13</sup> Extract: Extração dos dados.

<sup>14</sup> Transform: Transformação dos dados

<sup>15</sup> Load: Carga dos dados para análise.

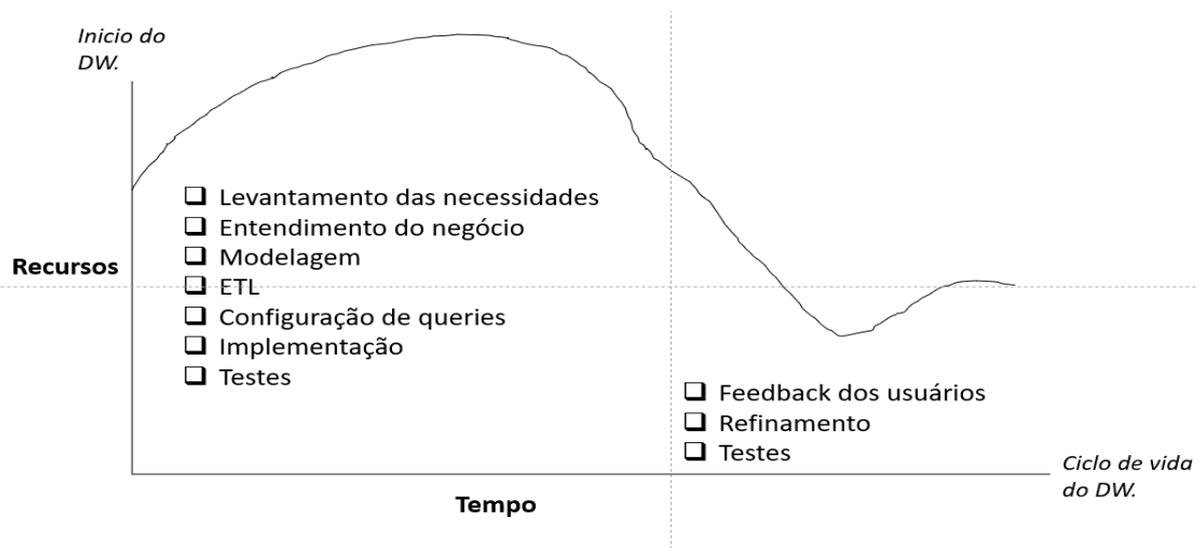


Figura 2.4 – Ciclo de vida *Data Warehouse*.

Fonte: Adaptado de MACHADO, 2013.

## 2.4 – Análises de dados

A análise de dados pode ser vista como o processo de desenvolvimento de decisões ou recomendações práticas para ações baseadas em vislumbres gerados por dados históricos (SHARDA; DELEN; TURBAN, 2019).

Conforme demonstrado na Figura 2.5, as análises de dados em empresas são o resultado de uma evolução natural que se iniciou com os sistemas de apoio à decisão (*DSS – Decision Support System*), que tiveram suas primeiras versões em atividade durante a década de 1970, evoluindo para o conceito de *Big Data*<sup>16</sup> durante a década de 2010 onde, além dados internos, dados externos passaram a ser extraídos de redes sociais para avaliações de diversas naturezas.

<sup>16</sup> Big data: Análises de dados em conjuntos muito grandes.

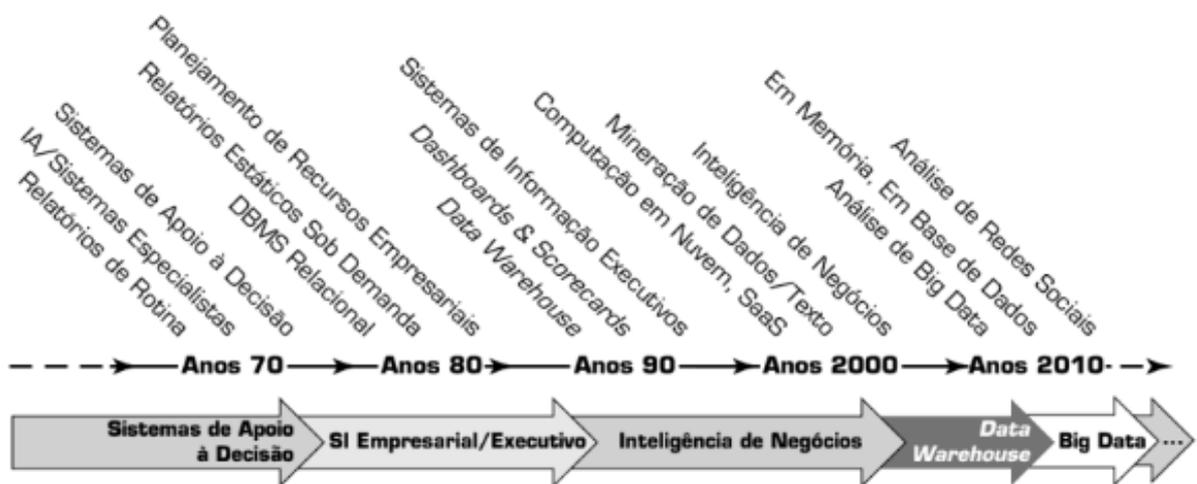


Figura 2.5 – Evolução da análise de dados.

Fonte: SHARDA; DELEN; TURBAN, 2019.

Conforme a Figura 2.6 – “Tipos de análise de dados”, existem três conceitos concretos em análises de negócios: Descritivas; Preditivas e Prescritivas.



Figura 2.6 – Tipo de análise de dados.

Fonte: SHARDA; DELEN; TURBAN, 2019.

### **2.4.1 – Análises Descritivas**

Análises Descritivas tem como foco conhecer o que está acontecendo na organização e entender tendências e causas (SHARDA; DELEN; TURBAN, 2019). Este tipo de análise foca na extração e avaliação dos relatórios internos, disponibilizando visualizações e consultas para a gestão.

### **2.4.2 – Análises Preditivas**

Análises Preditivas se baseiam em técnicas que podem ser consideradas como mineração de dados na tentativa de determinar a probabilidade de certos eventos, de interesse da organização, acontecerem (SHARDA; DELEN; TURBAN, 2019). Este tipo de análise foca em identificar e minerar dados internos, de qualquer formato, para munir a gestão com informações sobre diferentes probabilidades, para facilitar a tomada de decisão.

### **2.4.3 – Análises Prescritivas**

Análises Prescritivas reconhecem o momento da organização, avaliam o que pode vir a acontecer e recomendam uma decisão na busca pelo melhor desempenho (SHARDA; DELEN; TURBAN, 2019). Este tipo de análise tem uma natureza administrativa e foca na decisão que deve ser tomada pela gestão.

## **2.5 – Ferramentas para análises de dados**

Ferramentas de análises de dados podem auxiliar organizações em diversas formas. Ao compreender e utilizar essas ferramentas há um ganho nas decisões de alocação de recursos e precificação de produtos, por exemplo. As ferramentas também podem ser utilizadas para identificar tendências e compreender seus consumidores. Essa gama de informações pode ser disponibilizada através de relatórios e *dashboards*. O Microsoft *Power BI* foi considerado uma das melhores ferramentas de análises de dados em 2023 (FORBES, 2023).

### 2.5.1 – Power BI

A ferramenta *Power BI* é uma coleção de serviços de softwares, aplicativos e conectores que juntos transformam dados não relacionados em informações coerentes, interativas e visuais (MICROSOFT, 2023).

Um dos serviços fornecidos pelo *Power BI* para seus usuários é o de conectar *Data Warehouses* com a ferramenta, facilitando a interação e análise sem nenhum dano ou compromisso aos dados originais. Essa conexão é feita pelo aplicativo *Power Query* que conecta, extrai e carrega os dados para uma visualização dentro da ferramenta (MICROSOFT, 2023). O funcionamento do *Power Query* é demonstrado na Figura 2.7 abaixo.

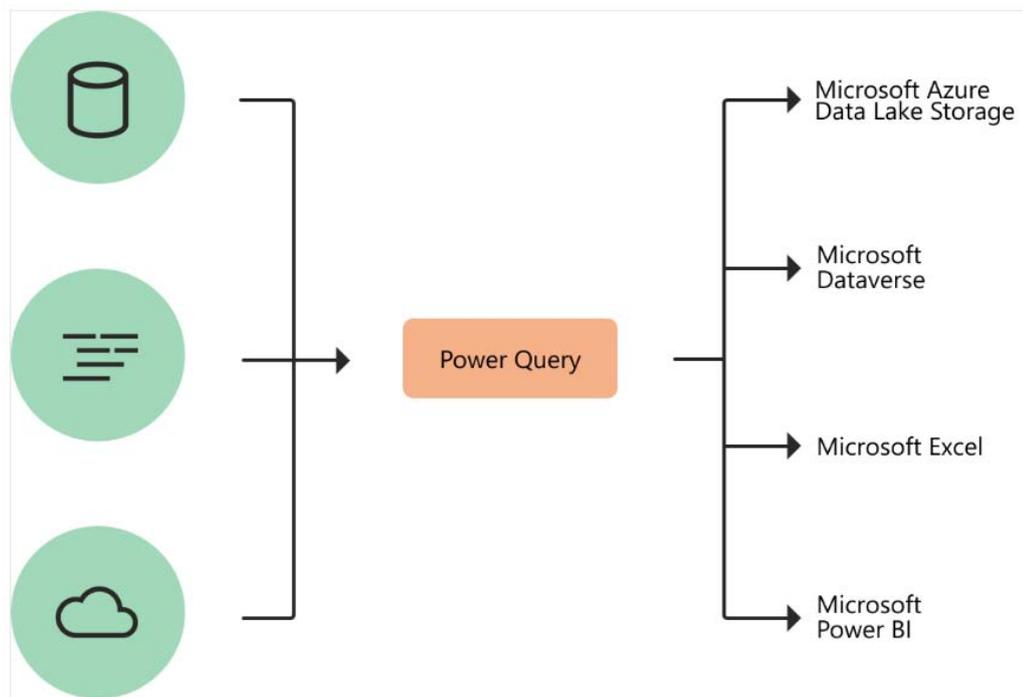


Figura 2.7 – ETL *Power Query*.

Fonte: MICROSOFT, 2023.

O *Power Query* conta ainda com um editor que permite a execução de transformações das mais simples, como a exclusão de uma coluna, às análises avançadas e complexas através de junções e fusões de dados de bases distintas. Em casos em que as transformações exijam configurações específicas, não suportadas pelo editor, é possível utilizar o mecanismo de script do *Power Query* que tem como base a linguagem M (MICROSOFT, 2023). A linguagem M se

refere ao *script*<sup>17</sup> da Microsoft, que possui suas particularidades, mas também semelhanças com outros *scripts* de análises de dados.

As análises realizadas no *Power BI*, com auxílio da *Power Query*, são registradas na ferramenta de forma que atualizações em bases de dados resultem na continuidade das análises. Esses registros na ferramenta podem ser usados de diversas formas, desde mineração de dados, utilizado neste trabalho, até aprendizado de máquinas.

## **2.6 – Indicadores de *Performance***

Indicadores de *performance*, como sugere o nome, focam em demonstrar através de indicadores como está o desempenho de diversas áreas e atividades. Esses indicadores variam e podem monitorar desde acompanhamentos operacionais, mapeamentos de melhorias, até dados financeiros. Existe uma parcela de indicadores que são considerados como chaves pelas organizações, que também conhecidos como KPIs – *Key Performance Indicator*<sup>18</sup>. Esses indicadores chaves se referem a um conjunto de medidas que focam nos aspectos organizacionais que são mais críticos para o sucesso imediato e futuro das organizações (PARMENTER, 2007).

### **2.6.1 – Indicadores Estratégicos**

Estão ligados à missão e visão da organização. São indicadores que estão associados ao planejamento e retratam o futuro esperado para a empresa. Tem como base cenários internos, como modelo de negócios e capital intelectual, e externos, como políticas do setor e situação econômica (SITEWARE, 2022).

### **2.6.2 – Indicadores Táticos**

Indicadores táticos estão ligados a um planejamento de curto a médio prazo, que busca suportar o planejamento da empresa através da execução de ações gerenciais dentro de um

---

<sup>17</sup> Script: Linguagem de programação de dados.

<sup>18</sup> Key Performance Indicator: Indicadores estratégicos chaves para os resultados dos processos corporativos.

período de 1 a 3 anos, sempre buscando o atingimento dos objetivos através da concretização das ações (SITEWARE, 2022).

## **2.7 – Data Mining**

Enquanto o aprendizado de máquinas tem o foco em executar suas ações sem orientações a mineração de dados (*Data Mining*), que pode ser considerada uma irmã do aprendizado de máquinas, foca no desenvolvimento do conhecimento e descobertas através da exploração de bases de dados(RAJAGOPAL; KUMAR; MAGAR; JYOTHI, 2022).

A mineração de dados requer técnicas de aprendizado de máquina, enquanto o contrário não se demonstra verdadeiro. É possível utilizar aprendizado de máquina em projetos que não necessitam de mineração de dados, mas é praticamente impossível a mineração de dados sem considerar pelo menos preceitos básicos de aprendizado máquina. Como exemplo é possível dizer que é necessário aprendizado de máquina para ensinar um robô a dirigir um carro, enquanto a mineração de dados demonstrará quais carros se demonstraram mais seguros. (RAJAGOPAL; KUMAR; MAGAR; JYOTHI, 2022).

# CAPÍTULO 3

## Propostas Tecnológicas

Este projeto se propôs a demonstrar como a análise de dados, aliada a ferramentas de inteligência de negócios, pode auxiliar uma empresa de seguros de automóveis operando no Estado do Rio de Janeiro a ter sucesso em um mercado que ao mesmo tempo que é promissor trás desafios como uma taxa média de 63% de sinistros, percentual extremamente desafiador para a rentabilidade do negócio.

A empresa de seguros que aqui fora explorada, a Seguradora Fictícia S.A., é uma empresa fictícia, criada com o propósito único de viabilizar a execução desse projeto. A empresa fictícia contou com uma base de dados de natureza *open-source*, ou seja, dados reais de código aberto que possibilitaram uma aderência ao propósito e resultado das avaliações realizadas.

A proposta aqui mencionada focou em explorar os desafios que envolveram as três fontes de dados que são base de qualquer operação de seguros: seguradora, segurado e objeto (ou bem). Essa exploração foi realizada através de perguntas e análises que tinham como intuito de explorar a base de dados da Seguradora Fictícia S.A., determinando tendencias que poderiam resultar em geração de informações à gestão. Essas tendências nasceram de análises dos dados realizadas através da ferramenta de inteligência de negócios *Power BI*, e foram exploradas até que se obteve um refinamento dos testes, que passaram então a ser usados como indicadores gerando *dashboards* interativos. Estes *dashboards* tinham como principal objetivo uma atualização contínua através de mineração dos dados adicionados a cada aquisição de contratos por novos clientes.

### 3.1 – Avaliação do *data set*

As análises aqui propostas foram realizadas em uma base de dados *open-source*, que tem a estrutura demonstrada na Figura 3.1. Essa base de dados será a fonte de informações da Seguradora Fictícia S.A., onde exploraremos todas as etapas destes dados de desde o início do ETL (*Extract, Transform and Load*) até o resultado atingido com a aplicação das técnicas de

análises de dados, que sempre foram o foco deste projeto por visar alimentar a gestão de informações relevantes.

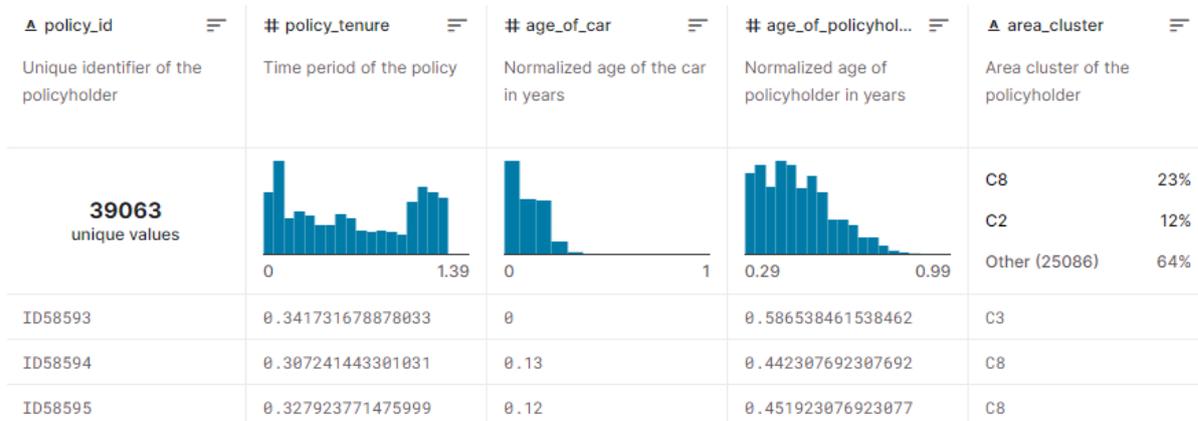


Figura 3.1 – Demonstrativo de 5, das 44 colunas, da base “*Car Insurance Claim Prediction*”.

Fonte: *Kaggle*, 2023.

Conforme mencionado por MACHADO, 2013, o nascimento de um *Data Warehouse* passa pelo levantamento das necessidades, entendimento do negócio e um processo de ETL. Fato é que sem um *Data Warehouse* não há análise contínua e, por isso, este é de extrema relevância também para este projeto.

Sabemos que a necessidade da Seguradora Fictícia S.A. é de progredir em um mercado em que o nível de sinistros atinge uma média de 63% das apólices, colocando assim em risco as margens de lucro em um mercado potencialmente rentável.

Sabemos também que a Seguradora Fictícia S.A. está focada inicialmente em uma localidade, que é o Estado do Rio de Janeiro, o que por si só traz diferentes desafios por reduzir a gama de clientes, que apresentam grande variação de potencial de acordo com cidade, bairro e outras características pessoais. Assim, considerando ser um negócio que possui um risco adicional por sua atuação limitada, a Seguradora Fictícia S.A. necessita da máxima precisão em sua concessão de seguros. Entendemos que neste mercado as empresas de seguros são livres para taxar seus clientes de acordo com suas premissas. No entanto isso aumenta o desafio, pois elevar os valores das apólices pode colocar uma seguradora fora do mercado, fazendo assim que a avaliação de potenciais clientes seja vital para tentar controlar e prever o volume de sinistros.

Com os pontos acima em mente iniciamos a avaliação da nossa base de dados através da extração, transformação e carga (*Extract, Transform and Load – ETL*), gerando uma base inicial para todo o projeto.

### 3.1.1 – *Extract, Transform and Load*

Como mencionado, nossa base de dados é de natureza *open-source* e, portanto, a extração (*Extract*) dos dados fora simulada pelo *download* da base para exploração. Vale aqui mencionar que esta não é a única forma da fase de extração de dados. Em organizações que dispõem de sistemas de planejamento de recursos empresariais, os famosos ERPs, é possível ter uma extração mais frequente e até contínua, que pode evoluir para um modelo sem necessidade de interferência humana.

Na execução da transformação (*Transform*) dos dados, nós iniciamos uma configuração de consultas e modelagens de processamentos de dados, para que quando estes fossem carregados no sistema de inteligência de negócios se mantivessem em formato adequado ao propósito do projeto.

Efetivamente, podemos dizer que nossa primeira transformação foi converter uma base *.csv* (*Comma-Separated Values*, ou valores separados por vírgulas) em uma tabela com 58.592 linhas, sem contar cabeçalhos, e 44 colunas, o que totalizou 2.578.048 células contendo dados para posterior exploração. A seguir, como demonstrado na Tabela 3.1, criamos uma base de referência para promover a transformação dos dados de cabeçalho originais (em inglês) para português e verificamos a existência de dados em branco, garantindo que não haveria nenhum impacto desta natureza em nossa avaliação. Por se tratar de dados de cabeçalho, e, portanto, de referência essas mudanças foram aplicadas manualmente em um editor de *query*<sup>19</sup>.

Tabela 3.1 – Cabeçalhos e validação dados.

<b>Cabeçalhos Original</b>	<b>Cabeçalhos Português</b>	<b>Dados em branco</b>
<i>policy_id</i>	APOLICE_IDENT	0
<i>policy_tenure</i>	APOLICE_DURACAO	0
<i>age_of_car</i>	VEICULO_IDADE	0
<i>age_of_policyholder</i>	IDADE_SEGURADO	0
<i>area_cluster</i>	CIDADE_CLUSTER	0
<i>population_density</i>	CIDADE_DENSIDADE	0
<i>Make</i>	VEICULO_FABRICANTE	0
<i>Segment</i>	VEICULO_SEGMENTO	0
<i>Model</i>	VEICULO_MODELO	0
<i>fuel_type</i>	COMBUSTIVEL	0

<sup>19</sup> Query: Termo utilizado para definir consultas realizadas em dados.

<i>max_torque</i>	TORQUE MAX	0
<i>max_power</i>	POTENCIA MAX	0
<i>engine_type</i>	MOTOR TIPO	0
<i>Airbags</i>	AIRBAGS	0
<i>is_esc</i>	ESTABILIDADE CONT	0
<i>is_adjustable_steering</i>	VOLANTE CONT	0
<i>is_tpms</i>	PNEU CONT(PRESSAO)	0
<i>is_parking_sensors</i>	ESTACIONAMENTO SENS	0
<i>is_parking_camera</i>	ESTACIONAMENTO CAM	0
<i>rear_brakes_type</i>	FREIO TIPO	0
<i>Displacement</i>	MOTOR LOCAL	0
<i>Cylinder</i>	MOTOR CILINDROS	0
<i>transmission_type</i>	CAMBIO TIPO	0
<i>gear_box</i>	CAMBIO MARCHAS	0
<i>steering_type</i>	VOLANTE TIPO	0
<i>turning_radius</i>	DIMENSAO RAO	0
<i>Length</i>	DIMENSAO COMPRIMENTO	0
<i>Width</i>	DIMENSAO LARGURA	0
<i>Height</i>	DIMENSAO ALTURA	0
<i>gross_weight</i>	DIMENSAO PESO	0
<i>is_front_fog_lights</i>	NEBLINA DIANTEIRO	0
<i>is_rear_window_wiper</i>	JANELATRASEIRA LIMPADOR	0
<i>is_rear_window_washer</i>	JANELATRASEIRA LAVADOR	0
<i>is_rear_window_defogger</i>	JANELATRASEIRA AQUECEDOR	0
<i>is_brake_assist</i>	FREIO SISTAUXILIO	0
<i>is_power_door_locks</i>	TRAVAS AUTOMATICO	0
<i>is_central_locking</i>	TRAVA CENTRAL	0
<i>is_power_steering</i>	VOLANTE AUTOMATICO	0
<i>is_driver_seat_height_adjustable</i>	ASSENTO AJUSTAVEL	0
<i>is_day_night_rear_view_mirror</i>	RETROVISOR AJUSTE(NOTURNO)	0
<i>is_ecw</i>	MOTOR ALERTA	0
<i>is_speed_alert</i>	VELOCIDADE ALERTA	0
<i>ncap_rating</i>	AVALIACAO SEGURANCA	0
<i>is_claim</i>	SINISTRO	0

Fonte: Elaborada pelo autor, 2023.

Concluídas as fases relacionadas aos cabeçalhos, também traduzimos dados identificados nas células do nosso *data set* relacionadas aos tipos de volante, tipos de freio, tipos de câmbio e tipos de combustível, assim como todos os dados booleanos (*Yes* = Sim, *No* = Não) existentes nas 17 colunas que se referiam a características dos veículos segurados como: estabilidade controlada e sistema de auxílio de frenagem, dentre outros. Essas alterações também foram feitas em um editor de *query*, conforme demonstrado nas Figuras 3.2 e 3.3.

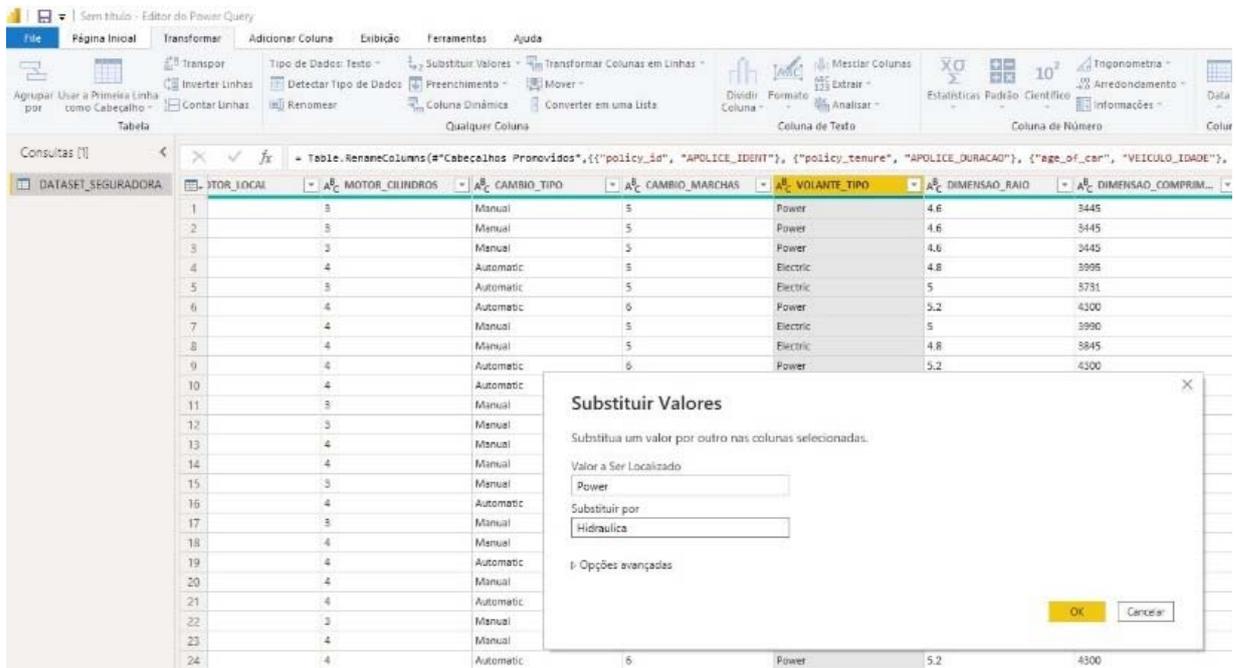


Figura 3.2 – Tradução de dados das células em editor de *query*.

Fonte: Elaborada pelo autor, 2023.

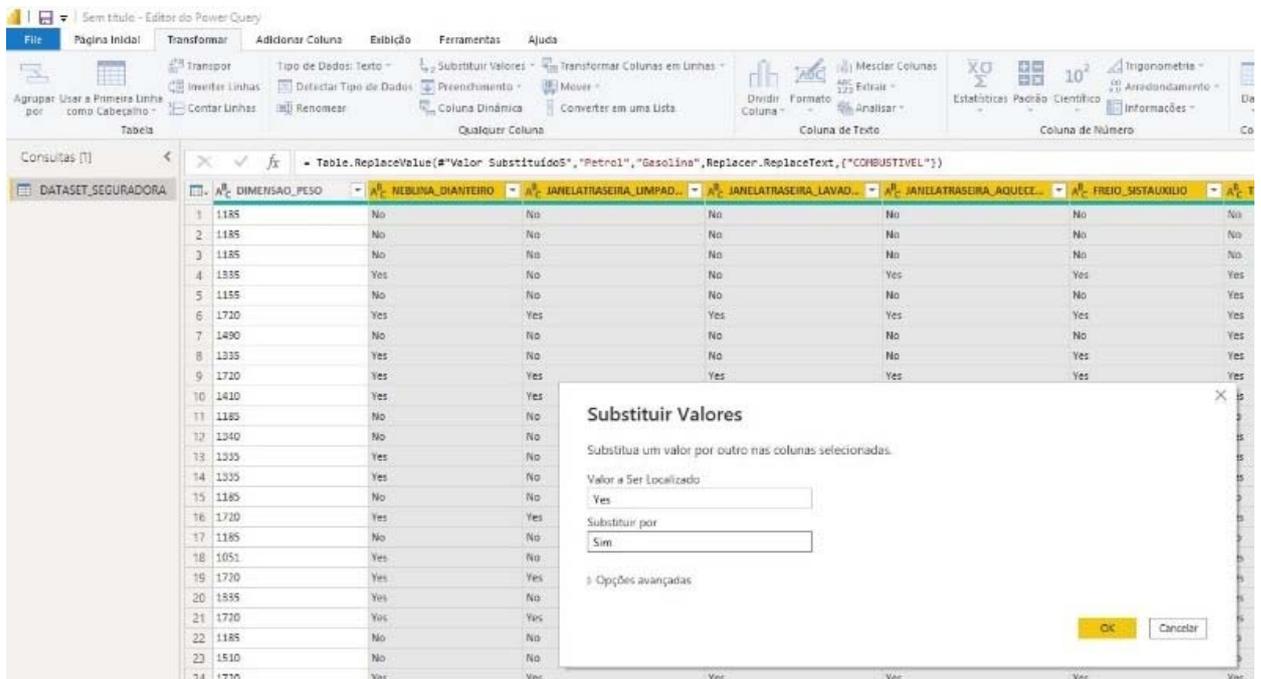


Figura 3.3 – Tradução de dados booleanos em editor de *query*.

Fonte: Elaborada pelo autor, 2023.

As alterações acima demonstradas resultaram na conclusão da fase de transformação e no início da fase de carga (*Load*). A fase de carga tem como principal objetivo levar os dados já tratados para o sistema de inteligência de negócios, sem que haja interferências e alterações

nos dados que poderiam posteriormente resultar em informações errôneas. Após a carga os dados ficam então disponíveis no sistema de inteligência de negócios conforme demonstrado na Figura 3.4, onde é possível ter a confirmação da carga sem erros e uma ideia de como a os dados ficam disponíveis para posteriores análises.

**Adicionar dados ao seu relatório**

Depois de carregados, seu



Importar dados do Excel



Importar dados do SQL Server

Colar dados em uma tabela em branco

Experimentar um conjunto de dados de exemplo

**Carregar**

✓ DATASET\_SEGURADORA  
58.592 linhas carregadas.

Obter dados de outra fonte →

APOLICE_IDENT	APOLICE_DURACAO	VEICULO_IDADE	IDADE_SEGURADO	CIDADE_CLUSTER	CIDADE_DENSIDADE
ID00039	1.12413817796631	0.12	0.403846153846154	C8	8794
ID00464	1.18649452780634	0.12	0.519230769230769	C8	8794
ID00561	0.849526879904547	0.12	0.365384615384615	C8	8794
ID00843	0.132301418307524	0.12	0.384615384615385	C8	8794
ID00929	0.375046619810046	0.12	0.490384615384615	C8	8794
ID01223	1.11815771047196	0.12	0.528846153846154	C8	8794
ID01523	1.11471327396674	0.12	0.5	C8	8794
ID01684	1.04834764508133	0.12	0.673076923076923	C8	8794
ID01848	0.377778578838847	0.12	0.605769230769231	C8	8794
ID02215	0.620690010095491	0.12	0.538461538461538	C8	8794
ID02331	0.740437841351261	0.12	0.576923076923077	C8	8794
ID02455	1.10549521881323	0.12	0.653846153846154	C8	8794
ID02892	0.500579903078288	0.12	0.423076923076923	C8	8794
ID03299	0.123936804569009	0.12	0.509615384615385	C8	8794
ID03310	1.06015485006702	0.12	0.346153846153846	C8	8794

Figura 3.4 – Conclusão da fase *Load*.

Fonte: Elaborada pelo autor, 2023.

Concluída então a fase de extração, transformação e carga (*ETL*), o foco se volta para a fase e análise dos dados disponíveis. Estas análises são feitas tanto na estrutura da base, como através de elaborações de perguntas chave e utilização das técnicas de análises de dados mencionadas no capítulo anterior.

### 3.1.2 – Análise da base de dados disponível

Uma vez em posse da base de dados devidamente traduzida, e cientes da inexistência de dados nulos ou em branco, inicia-se uma avaliação do conteúdo das demais células existentes na base para posterior aplicação de técnicas que possibilitarão o atingimento do resultado esperado com o projeto. Ao avaliar a estrutura da base, que foi carregada no sistema de inteligência de negócios, os seguintes pontos foram identificados:

- As células com dados relacionados a períodos (duração da apólice, idade do veículo e idade do segurado) foram criadas em formato percentual, mas sem informações de referências. Essa natureza demanda uma avaliação extra por parte do projeto, para que esse percentual seja aplicado sobre uma informação aderente com a realidade do mercado.
- As células com dados relacionados a cidade e bairros foram inseridas na base de dados *open-source* apenas como referência. O que significa que não havia detalhes sobre a localidade a ser explorada, mas que não invalida o projeto por possibilitar uma relação entre os locais ali relacionados e uma base de localidade própria.
- As células com dados relacionados aos veículos (fabricante, segmento e modelo) também possuíam uma natureza de referência. No entanto, neste caso havia outros dados de suporte na base *open-source* como número de cilindros, referência do motor, torque e potência.

Considerando a necessidade de se incluir os dados faltantes para possibilitar o avanço do projeto, um estudo adicional foi realizado na tentativa de identificar e incluir as informações faltantes. Esse estudo resultou nas seguintes definições:

- A duração a ser considerada para uma apólice de seguros é de 12 meses (1 ano). Esta informação foi obtida em consultas aos dados da Superintendência de Seguros Privados (SUSEP), que esclarece que esta é a duração média de um seguro de veículos é de 12 meses. Apesar de não haver limitação para este período no mercado, a Seguradora Fictícia S.A. não possuía interesse em contratos maiores pela exposição ao risco.
- A idade máxima do veículo segurado foi definida em 15 anos. Essa informação também foi explorada em consulta à mesma superintendência, que não define limite para a idade do veículo segurado. No entanto, consultas aos maiores concorrentes no mercado de seguros demonstraram a existência de um limite de 15 anos para a idade do veículo.

- A idade do cliente segurado foi definida em 77 anos. Aqui foi considerada a idade média do brasileiro, que na data da realização deste projeto era de 77 anos, mas que deve ser mantida atualizada sempre que houver revisões.
- A localidade explorada foi o Estado do Rio de Janeiro que, como mencionado previamente, é a área de atuação da Seguradora Fictícia S.A. As cidades foram elencadas de acordo com sua representatividade por densidade.
- Os dados de segmentos dos veículos foram extraídos da base internacional com tal propósito.
- Os dados relacionados aos fabricantes e modelos dos veículos foram pesquisados e identificados, utilizando como referência as informações existentes na base de dados (número de cilindros, referência do motor, torque e potência).

Os dados resultantes desse estudo adicional foram inseridos em tabelas em Excel, que foram posteriormente inseridas no sistema de inteligência de negócios para serem adicionados à base através da relação entre as referências das colunas já existentes e os dados adicionais, conforme demonstrado na Figura 3.5.

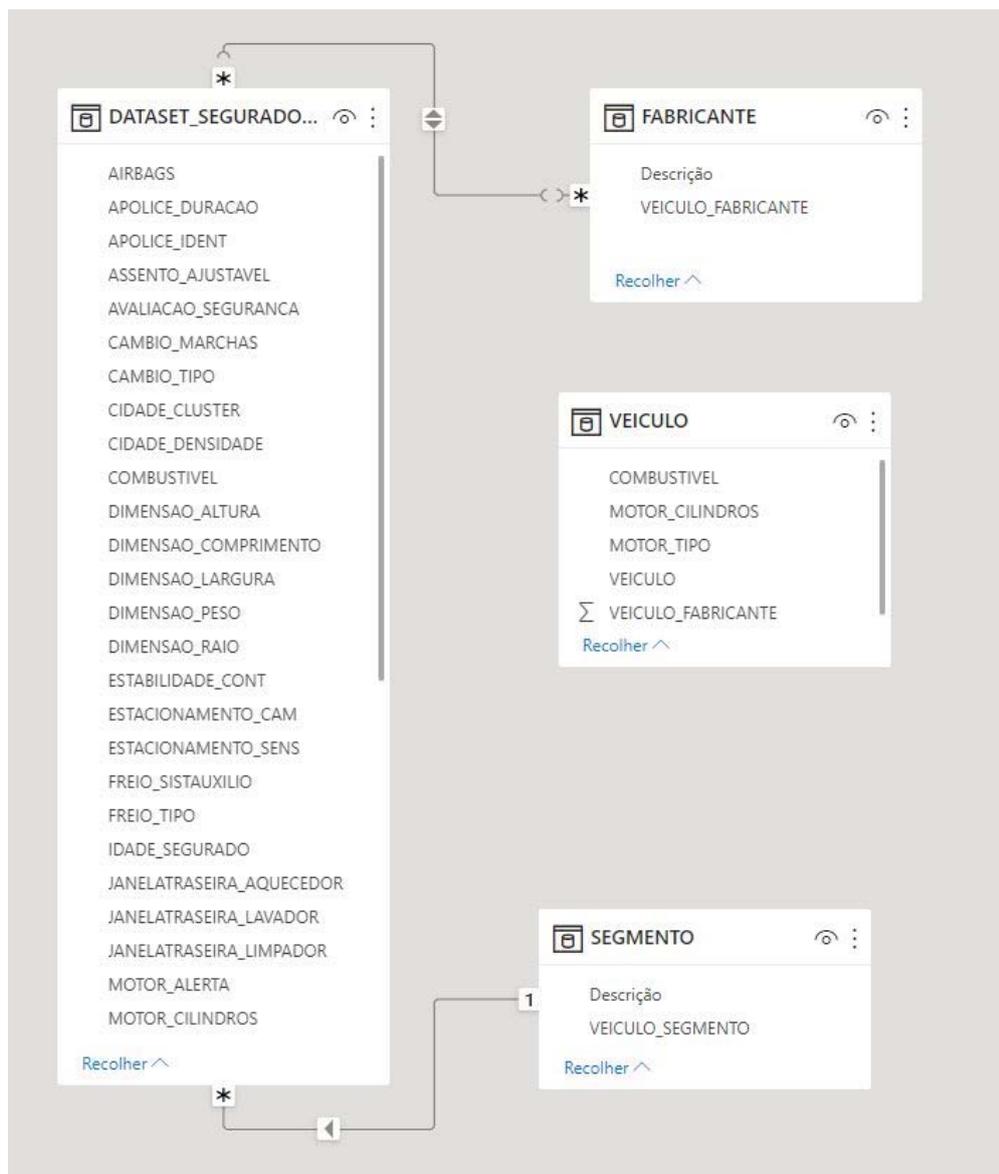


Figura 3.5 – Relações entre estudo adicional e bases de dados.

Fonte: Elaborada pelo autor, 2023.

Com as informações refinadas e disponíveis para análises atingimos um momento mais maduro do ciclo de vida da nossa base de dados, onde então passamos a considerar os testes a serem executados para avançar com o projeto.

### 3.1.3 – Aplicação das técnicas para análises de dados

Com a base de dados pronta para o início das análises exploramos as diferentes técnicas para o atingimento do resultado esperado: segurança na tomada de decisão na concessão de seguros.

Toda e qualquer técnica de análise de dados tem um início em comum, um problema. Em nosso projeto temos clareza de qual é o propósito, como mencionado no parágrafo anterior, mas isso apenas não é o suficiente. Todo resultado deve ser considerado como um incentivo para definição de como chegar àquele patamar.

Em nosso projeto utilizamos um modelo de elaboração de perguntas como técnica de execução das análises. Essas perguntas estavam sempre relacionadas com o resultado esperado e tinham naturezas descritivas, que focavam no que já havia acontecido, e preditivas, que focavam no que ainda iria acontecer.

Com as perguntas elaboradas o resultado do projeto passou a depender da exploração dos dados para responder essas perguntas. Podemos dizer que o segundo passo da aplicação das técnicas de análises é o de minerar os dados no sistema de inteligência de negócios para obter as respostas aos questionamentos, que são demonstrados na Tabela 3.2.

Tabela 3.2 – Demonstração de questões a serem avaliadas.

<b>Tipo</b>	<b>Perguntas</b>
Descritivo	Qual a relação entre a avaliação de segurança do veículo e sinistros?
Preditivo	Carros com maior número de itens de segurança possuem menor número de sinistros?
Descritivo	A densidade populacional influencia nos sinistros?
Descritivo	As idades do carro e motorista tem influência no número de sinistros?
Descritivo	A potência do veículo tem influência no número de sinistros?

Fonte: Elaborada pelo autor, 2023.

Um outro passo, ou técnica, utilizado se refere ao formado em como essas respostas serão demonstradas. Aqui a técnica aplicada foi o de transformar respostas em informações e gráficos, que possibilitarão a geração de informação para a gestão através de *dashboards* e indicadores sobre, facilitando a tomada de decisão.

### **3.2 – Utilizando o Power BI como ferramenta de inteligência**

Como mencionado no embasamento teórico, a ferramenta *Power BI* é extremamente utilizada e considerada um dos melhores sistemas de inteligência de negócios em uso. Isso, em parte, se dá pela conjunção de softwares que compõem o *Power BI*, como por exemplo o *Power Query*, que foi o editor utilizado para transformar os dados inseridos em nossa *data set*, demonstrado na seção anterior.

Avançando no projeto, a característica de inteligência de negócios do *Power BI* passou a ser mais explorada nas análises e combinações de dados. As respostas às perguntas elaboradas na seção 3.1.3 – “Técnicas para análises de dados” foram extraídas da base de dados através de comparações de campos e colunas, utilizando ferramentas de filtros e visualizações. Todos esses artifícios suportados pela ferramenta, conforme demonstrado na Figura 3.6.

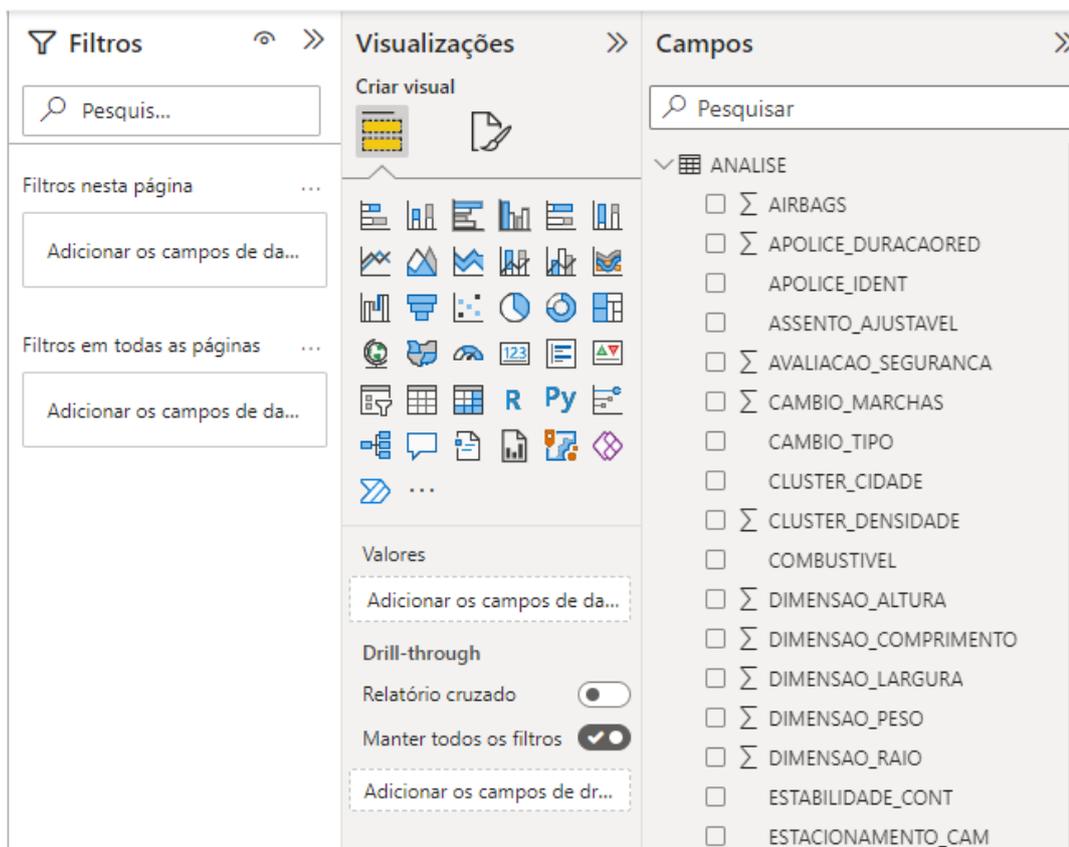


Figura 3.6 – Ferramenta de análises *Power BI*.

Fonte: Microsoft *Power BI*, 2023, 2023.

As conjunções de softwares, que resultam nas ferramentas disponíveis no *Power BI*, possuem extrema relevância não só para identificar as respostas às técnicas implementadas, mas para uma adequada demonstração desses resultados. Todas as “respostas” foram demonstradas através de gráficos, como o da Figura 3.7, que posteriormente são transformados em um *dashboard* retroalimentado que disponibiliza a informação através de indicadores de forma prática e amigável, facilitando a utilização pelos tomadores de decisão.

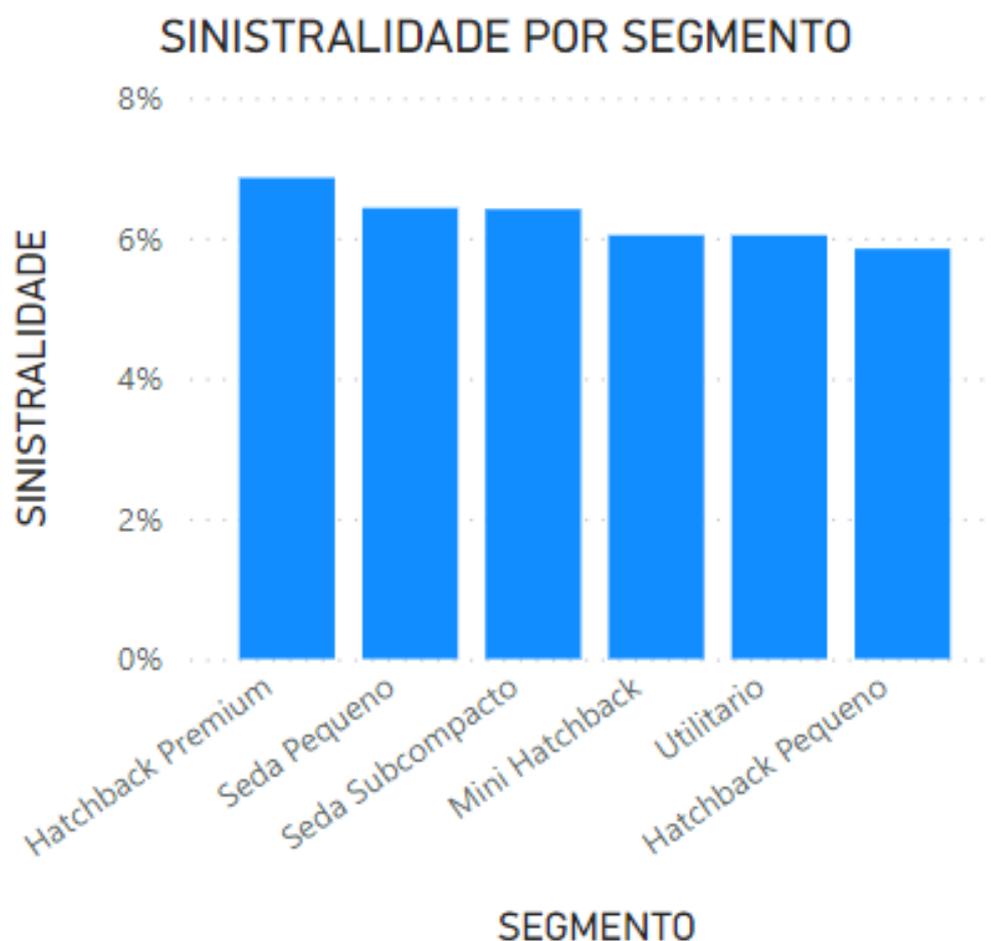


Figura 3.7 – Demonstração de gráfico do *Power BI*.

Fonte: Elaborada pelo autor, 2023.

### 3.3 – Ajustando o modelo para a mineração dos dados

O modelo de *data mining* utilizado neste projeto é voltado para a inclusão de dados no sistema de inteligência de negócios, como uma parametrização com variáveis definidas. No modelo explorado neste trabalho, o esperado é que os dados sejam sempre minerados e posteriormente analisado, de *data sets* atualizados, com mais conteúdo do que as bases iniciais.

O ponto chave do modelo de *data mining* proposto é garantir que todas as técnicas inseridas no projeto e os indicadores definidos e demonstrados através de *dashboards* estejam atualizados. Dessa forma, precisamos ter certeza de que todas as etapas que foram demonstradas até aqui continuarão a ser executadas nos dados diante de eventuais necessidades.

Para ter sucesso nesse desafio contamos mais uma vez com o software complementar do *Power BI* (editor do *Power Query*) e sua capacidade de transformar e adicionar células,

enquanto tudo é inserido automaticamente em um  $\log^{20}$  na linguagem M. Na Figura 3.8 é possível ver uma demonstração dessa linguagem.

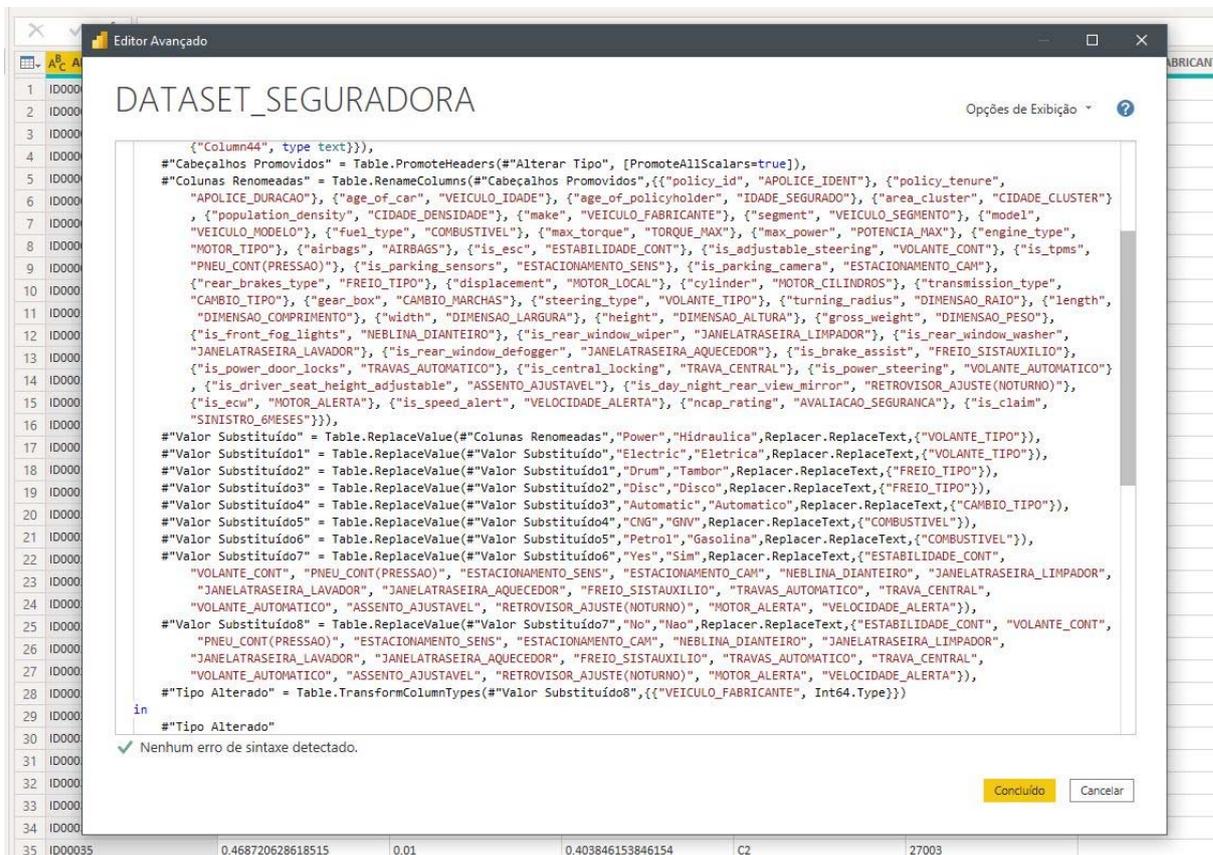


Figura 3.8 – Demonstração da linguagem M.

Fonte: Elaborada pelo autor, 2023.

A linguagem demonstrada é coerente com as etapas executadas no editor e nos permite ter certeza que ao aplicar as etapas necessárias para transformar, carregar e analisar os dados, salvando-os em uma sintaxe que poderia ser reaplicada, teremos uma base fidedigna e que resultará em resultados validos sempre que utilizada.

<sup>20</sup> Log: Registro mantido pelos sistemas detalhando todas as etapas das análises executadas.

# CAPÍTULO 4

## Resultados Obtidos

Conforme discorrido sobre nas propostas tecnológicas, o foco deste projeto é demonstrar os benefícios existentes em aliar as análises de dados com ferramentas de inteligência de negócios no contexto de uma empresa de seguros.

A expectativa é que os resultados aqui apresentados suportem as propostas mencionadas. Para isto, os resultados serão apresentados na seguinte sequência: Dados desconsiderados das análises; Análises de booleanos duplicados ou repetidos; Análises de outliers; e Análises exploratórias dos dados.

### 4.1 – Dados desconsiderados das análises

As avaliações e análises executadas na base de dados da Seguradora Fictícia S.A. confirmaram a existência de uma gama significativa de dados a serem explorados com 2.578.048 células.

Considerando a expressividade desses números e análises superficiais conduzidas junto com as transformações da base de dados, foram identificadas 8 colunas, ou 468.736 células, que não apresentam benefícios para uma exploração futura. São essas colunas:

- TORQUE\_MAX: Torque é uma medida de força. A força de um veículo não está relacionada diretamente com sua velocidade e nem influência em equipamentos de segurança.
- MOTOR\_TIPO, MOTOR\_LOCAL, MOTOR\_CILINDROS: As coordenadas relacionadas ao motor auxiliaram na identificação dos modelos dos veículos, mas individualmente não apresentavam nenhum dado a ser explorado.
- CAMBIO\_MARCHAS: A quantidade de marchas não implicava em informações que pudessem auxiliar as análises.
- DIMENSAO\_COMPRIMENTO, DIMENSAO\_LARGURA, DIMENSAO\_ALTURA: As dimensões de forma isolada foram substituídas por uma concatenação destas informações.

## 4.2 – Análises de booleanos duplicados ou repetidos

A base de dados da Seguradora Fictícia S.A. conta com 17 colunas de dados booleanos, que representam um total de 996.062 células. Essas colunas e dados foram traduzidos para melhor representar o resultado deste projeto, mas além da tradução uma análise mais minuciosa para garantir a relevância de todos esses dados se fazia necessária.

Com isso em mente, listamos todas as colunas de dados booleanos organizadas pela contagem de suas informações “Sim” ou “Não”. O resultado é demonstrado na Tabela 4.1 abaixo:

Tabela 4.1 – Análise quantitativa dos dados booleanos.

Dados	Sim	Não	Total
ESTABILIDADE_CONT	18.401	40.191	58.592
VOLANTE_CONT	35.526	23.066	58.592
PNEU_CONT(PRESSAO)	14.018	44.574	58.592
ESTACIONAMENTO_SENS	56.219	2.373	58.592
ESTACIONAMENTO_CAM	22.888	35.704	58.592
NEBLINA_DIANTEIRO	33.928	24.664	58.592
JANELATRASEIRA_LIMPADOR	16.958	41.634	58.592
JANELATRASEIRA_LAVADOR	16.958	41.634	58.592
JANELATRASEIRA_AQUECEDOR	20.515	38.077	58.592
FREIO_SISTAUXILIO	32.177	26.415	58.592
TRAVAS_AUTOMATICO	42.435	16.157	58.592
TRAVA_CENTRAL	42.435	16.157	58.592
VOLANTE_AUTOMATICO	57.383	1.209	58.592
ASSENTO_AJUSTAVEL	34.291	24.301	58.592
RETROVISOR_AJUSTE(NOTURNO)	22.283	36.309	58.592
MOTOR_ALERTA	42.435	16.157	58.592
VELOCIDADE_ALERTA	58.229	363	58.592

Fonte: Elaborada pelo autor, 2023.

Uma análise quantitativa e explorativa dos dados booleanos demonstrou a existência de cinco colunas com informações idênticas em quantidade. Avaliando as informações mais profundamente identificamos que a coluna “MOTOR\_ALERTA” se tratava de um falso

negativo, pois apesar de ser idêntica em quantidade, os dados das células eram distintos entre si. Já os dados das colunas “JANELATRASEIRA\_LIMPADOR” e “JANELATRASEIRA\_LAVADOR” eram idênticos, assim como os das colunas “TRAVAS\_AUTOMATICO” e “TRAVAS\_CENTRAL”.

Desta forma, e visando uma maior acuracidade e eficiência nas análises as colunas “JANELATRASEIRA\_LIMPADOR” e “TRAVAS\_AUTOMATICO” não foram utilizadas como recursos durante a continuidade do projeto.

### **4.3 – Análises de outliers**

A tradução de um outlier pode ser considerada como “um ponto fora da curva”. Essa tradução é um conceito que representa com precisão o que buscamos ao analisar a existência de outliers na base de dados.

Iniciamos essa análise retornando à base, e verificando como ela se apresentava após os resultados das análises anteriores. Após concluídas tais etapas identificamos que havíamos definido 35 colunas como foco do projeto, sendo destas: 15 colunas de dados booleanos; 12 colunas de dados dos veículos; 2 colunas com dados das cidades, 3 colunas com dados dos seguros; 1 coluna com duração da apólice; 1 coluna com dados de idade do veículo; e 1 coluna com dados de idade dos segurados.

Enquanto 32 colunas não possuíam dados limitantes, as colunas envolvendo duração de seguros, idade do veículo e idade dos segurados haviam sido limitadas durante o desenvolvimento da metodologia a ser aplicada. Com base nas definições previamente estabelecidas, foram criadas colunas condicionais na base de dados definindo o que seriam apólices, ou dados, válidos para a continuidade das análises.

Em relação à duração das apólices, foi definido que todas as apólices com duração até 12 meses eram válidas, enquanto as demais foram consideradas inválidas por não serem de interesse da Seguradora Fictícia S.A. Conforme demonstrado na Figura 4.1, a análise de outliers das apólices resultou em 24,69% de apólices inválidas (em vermelho), ou acima do período de interesse, que por consequência foram excluídas da base de dados através de filtros nas análises realizadas.

## ANALISE OUTLIERS - APOLICES

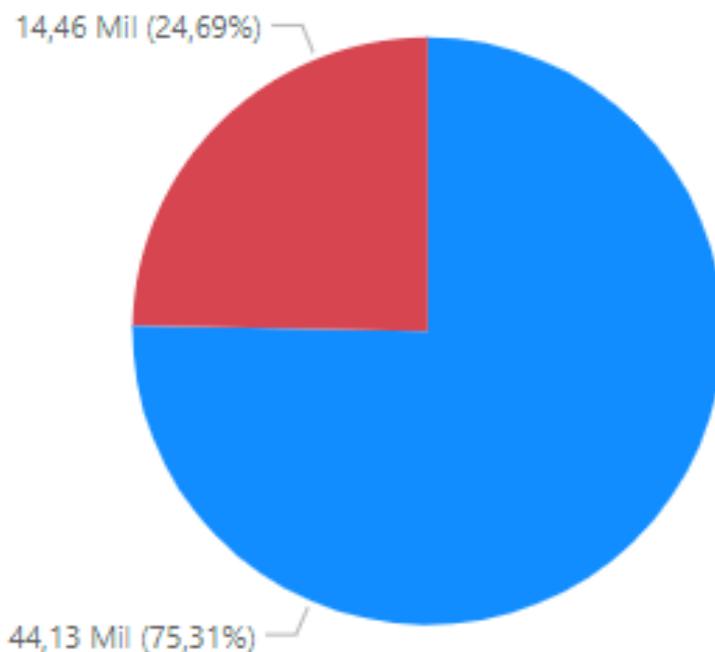


Figura 4.1 – Outliers apólices de seguros acima da duração estabelecida.

Fonte: Elaborada pelo autor, 2023.

As análises de outliers na idade dos veículos tomou como base a limitação estabelecida de veículos com no máximo 15 anos, sendo os veículos acima de 15 anos considerados inválidos. Já as análises de idade dos segurados tomaram como base a limitação estabelecida de 77 anos, sendo os segurados acima dessa idade considerado inválidos. Para as análises de outliers em idade dos segurados foi adicionado a limitação legal de 18 anos para dirigir. Portanto, todos os segurados abaixo de 18 anos e acima de 77 anos foram considerados inválidos. Para ambas as análises não foram identificados dados inválidos, mesmo assim o processo de exclusão de dados inválidos se manteve o mesmo para fins de aprendizado da máquina.

### 4.4 – Análises exploratórias dos dados

Concluídas as análises prévias, ou preparatórias, as análises exploratórias foram iniciadas. Considerando todas as exclusões, por descon sideração de colunas ou filtros de linhas

de outliers, a base de dados da Seguradora Fictícia S.A. havia saído de 44 colunas com 58.592 linhas para 35 colunas com 44.128 linhas, ou de 2.578.048 células para 1.544.480 células com dados a serem explorados.

A partir das propostas tecnológicas descritas no capítulo 03, apresenta-se a sequência de resultados obtidos para a solução do problema estudado.

#### **4.4.1 – Análises dos dados gerais da base de dados**

Conforme mencionado anteriormente, a base de dados da Seguradora Fictícia S.A. fora devidamente tratada, chegando a um total de 44.128 dados válidos de apólices para serem exploradas. Iniciamos as análises gerais verificando quantos sinistros existiam na base e identificamos um total de 2.543 sinistros, com uma média geral de 5,77% de ocorrência. Conhecer essa média era de extrema relevância para a continuidade das análises.

Após avaliar apólices e sinistros começamos a avaliar a base por segmento dos veículos e fabricantes. Verificamos a existência de uma concentração de apólices no segmento *hatchback*, e suas variações, enquanto o segmento de utilitários era o menos explorado pela seguradora. Adicionalmente identificamos também uma concentração relevante de veículos da fabricante Suzuki em nossa base.

A concentração de veículos da Suzuki se tornou relevante, pois ao avaliar a variação de avaliações das notas de segurança atribuídas aos veículos por empresa especializada, a Suzuki apareceu com a média mais baixa, incluindo três veículos com nota zero, enquanto Honda e Tata tinham as melhores avaliações.

O último passo da análise geral foi listar todas as cidades onde a seguradora opera e suas densidades, que serviram de base para a continuidade das análises. As avaliações gerais podem ser observadas na Figura 4.2.

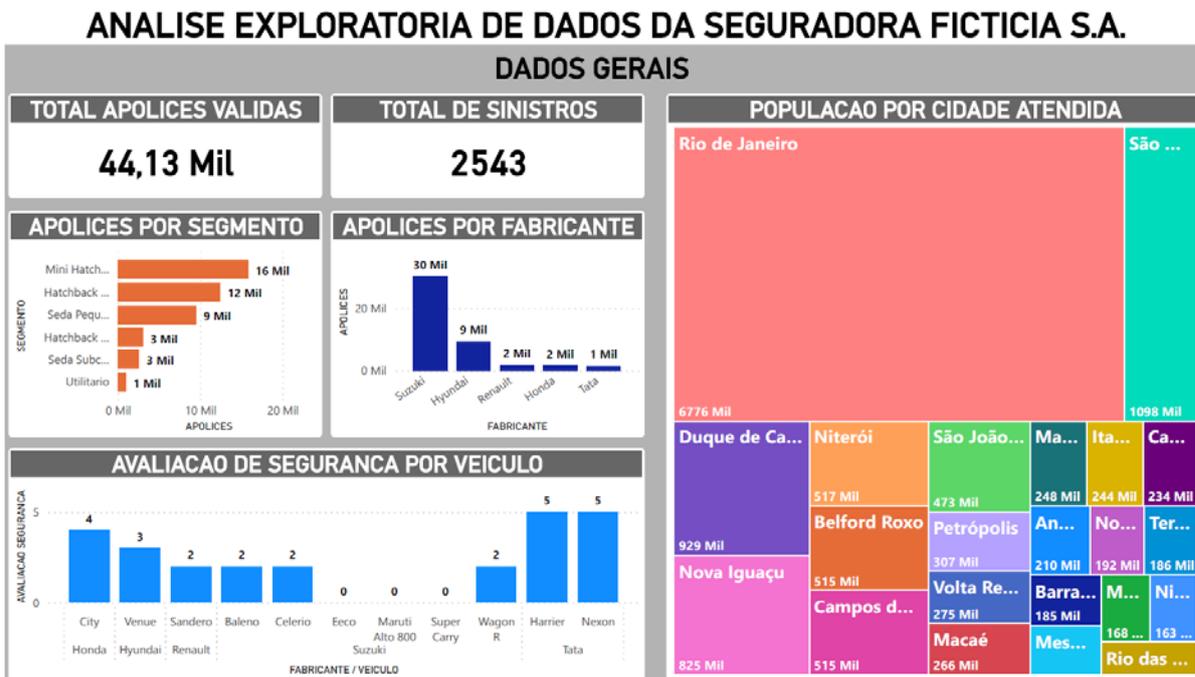


Figura 4.2: Análise e exploração dos dados gerais, base para demais análises.

Fonte: Elaborada pelo autor, 2023.

Na continuidade das análises os dados foram segregados por natureza entre pessoais, veículo e segurança. As análises levaram em consideração dados extraídos das análises resultaram em indicadores para facilitar a tomada de decisão da gestão.

#### 4.4.2 – Análises dos dados de natureza pessoal

As análises de dados de natureza pessoal focaram principalmente nos potenciais impactos existentes na quantidade de sinistros devido à idade dos segurados e cidade onde residiam, conforme demonstrado na Figura 4.3.

# ANALISE EXPLORATORIA DE DADOS DA SEGURADORA FICTICIA S.A.

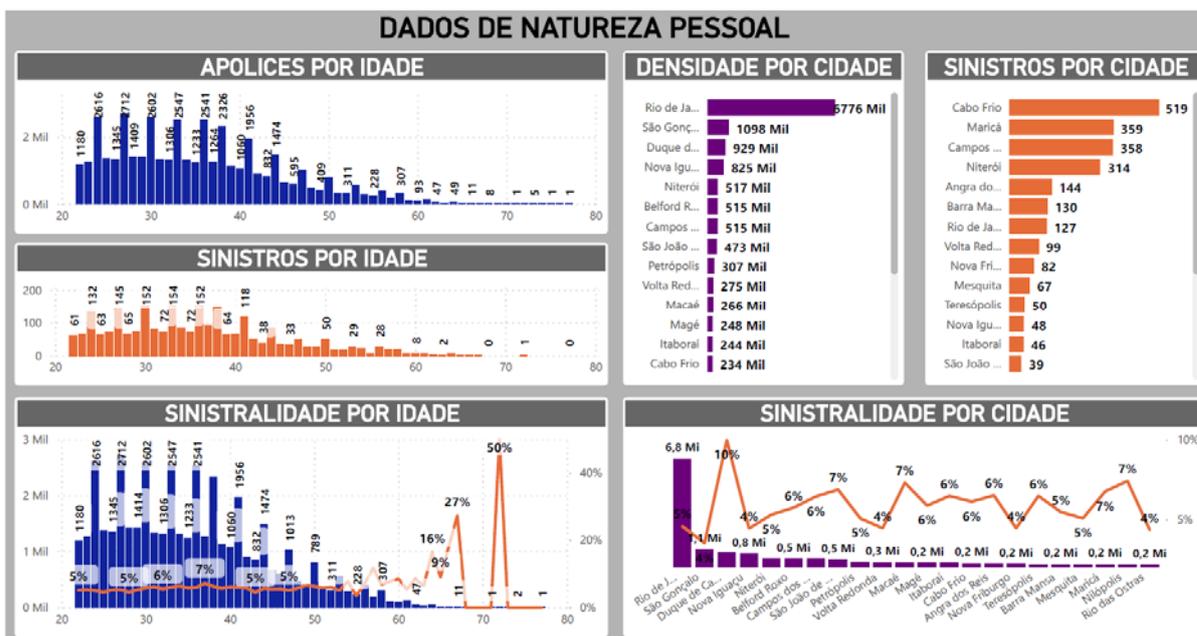


Figura 4.3: Dados de natureza pessoal analisados.

Fonte: Elaborada pelo autor, 2023.

Foi identificada uma concentração de apólices e sinistros entre segurados de 22 a 50 anos, e uma análise da sinistralidade por idade resultou em aumentos significativos nas idades de 37, 54, 57, 60, 62, 64, 65, 66, 67 e 72 anos. Desconsiderando as sinistralidades descontinuadas, que podem ser consideradas como outliers, chegamos a uma conclusão de que nas idades entre 60 e 67 anos há uma conexão com os sinistros, conforme apresentado na Figura 4.4.



Fonte: Elaborada pelo autor, 2023.

Ao analisar o gráfico da Figura 4.6 é possível perceber que as cidades de São Gonçalo, Nova Iguaçu, Volta Redonda, Nova Friburgo e Rio das Ostras apresentam uma média de sinistros de 4%, quase 2% abaixo da média geral. Enquanto as cidades de Duque de Caxias, São João do Meriti, Macaé, Maricá e Nilópolis apresentam média superior a 7%, representando uma exposição maior ao risco, em especial Duque de Caxias que obteve média de 10% de sinistro.

Para avaliar a aderência das conclusões alcançadas, uma análise combinada foi realizada na base, demonstrada na Figura 4.6. Ao combinar as análises de idade e densidade percebeu-se uma aderência ao modelo e este passou a ser um dos indicadores que serão explorados para novos seguros.

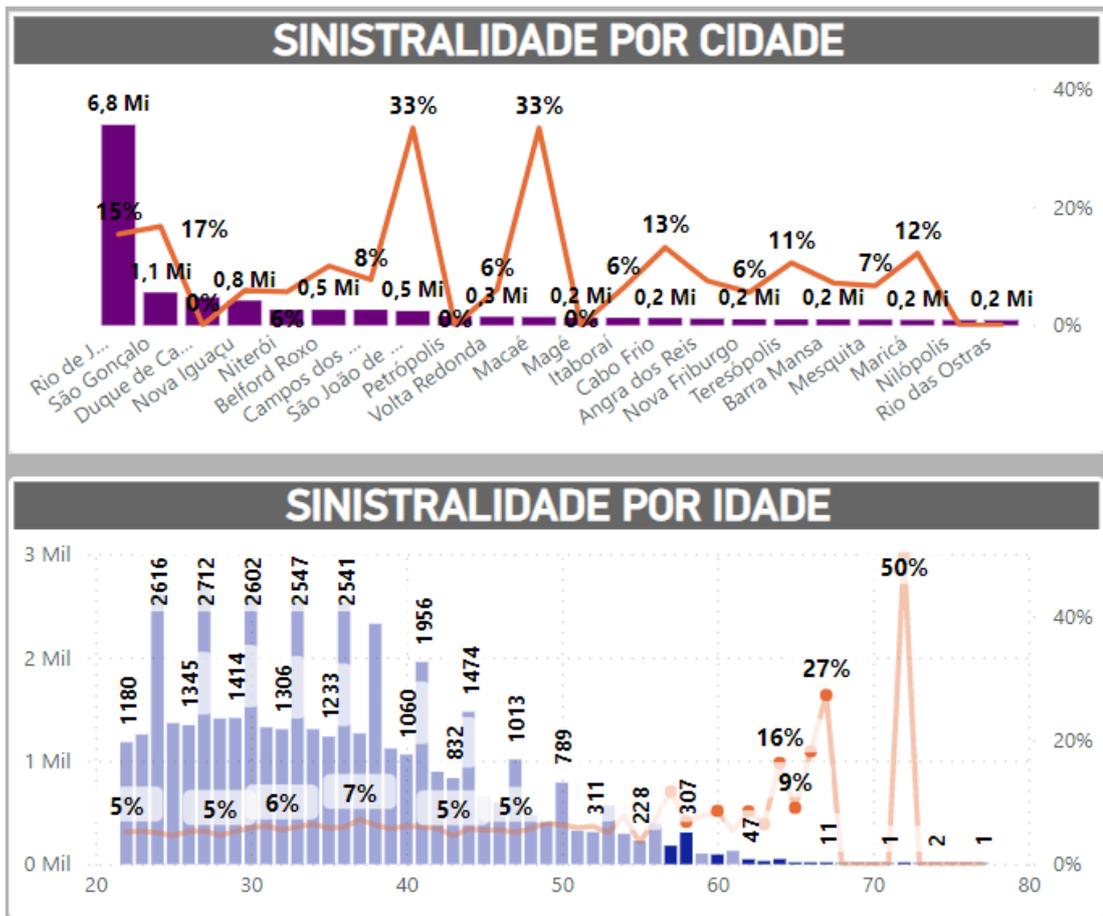


Figura 4.6: Aderência das análises de sinistralidade por cidade e idade.

Fonte: Elaborada pelo autor, 2023.

#### 4.4.3 – Análises dos dados do veículo

Em análise aos dados dos veículos, demonstrados na Figura 4.7, observou-se uma aderência das características com a média geral com uma exceção, veículos com idade igual a cinco anos.

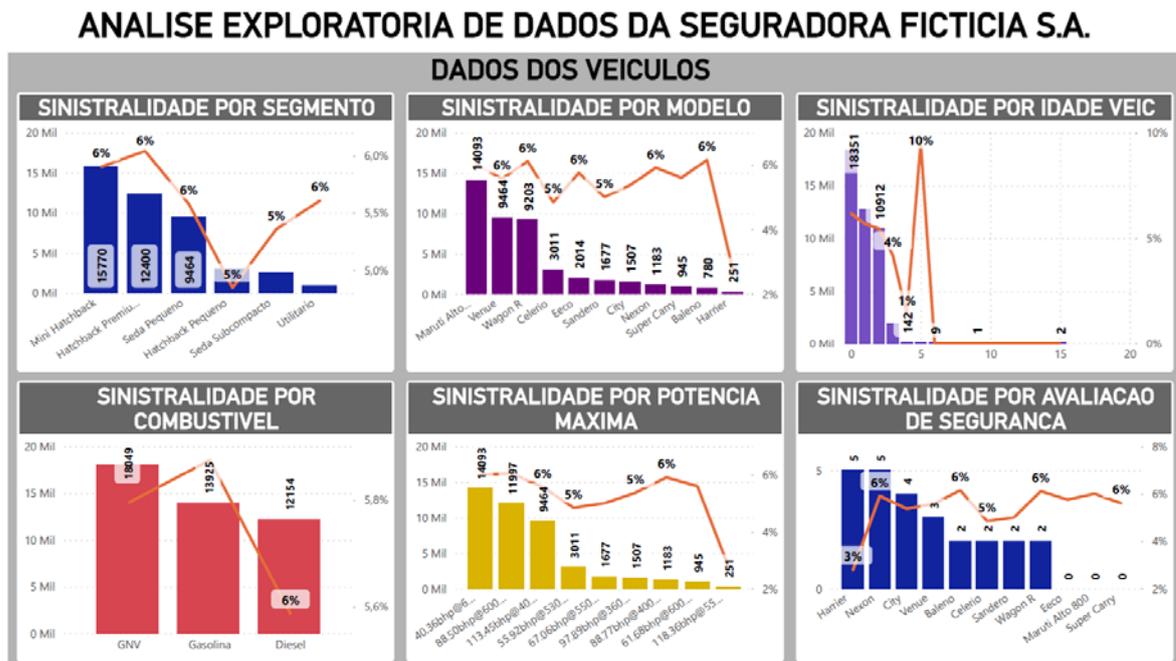


Figura 4.7: Aderência dos dados dos veículos à média geral.

Fonte: Elaborada pelo autor, 2023.

Uma análise mais a fundo nos veículos de cinco anos mostrou que essas ocorrências refletiam nos demais modelos e estavam relacionadas principalmente aos veículos “Wagon R” e “Venue” que apresentavam o fim de sua garantia neste mesmo período, conforme demonstrado na Figura 4.8. Desta forma, a união entre esses modelos específicos e a idade de cinco anos foi inserida nos indicadores.

## ANALISE EXPLORATORIA DE DADOS DA SEGURADORA FICTICIA S.A.

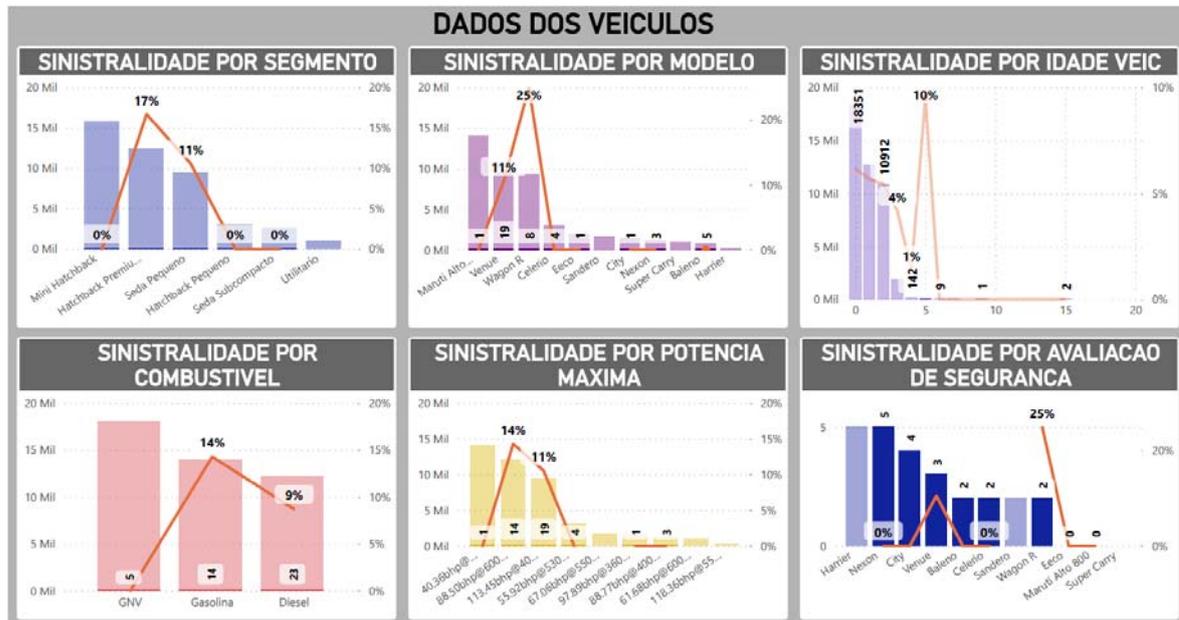


Figura 4.8: Modelo com a idade de 5 anos e seu impacto nos dados dos veículos.

Fonte: Elaborada pelo autor, 2023.

### 4.4.4 – Análises dos dados de segurança

As análises a dados de segurança tiveram como foco 13 itens no total que foram analisados quanto à sua existência na ocorrência de sinistros, enquanto a referência de sinistros se manteve em comparação à média geral, seguindo os mesmos critérios das análises anteriores.

Desta forma, deve se ter em mente que enquanto as diferenças não parecem ser tão representativas se considerando os percentuais demonstrados, que resultaram em diferenças decimais. No entanto, ao compararmos as diferenças decimais entre si, com base na média geral de 5,77% verificamos que os valores se tornam expressivos.

Na Figura 4.9 é possível observar todos os itens de segurança que foram avaliados.

## ANALISE EXPLORATORIA DE DADOS DA SEGURADORA FICTICIA S.A.

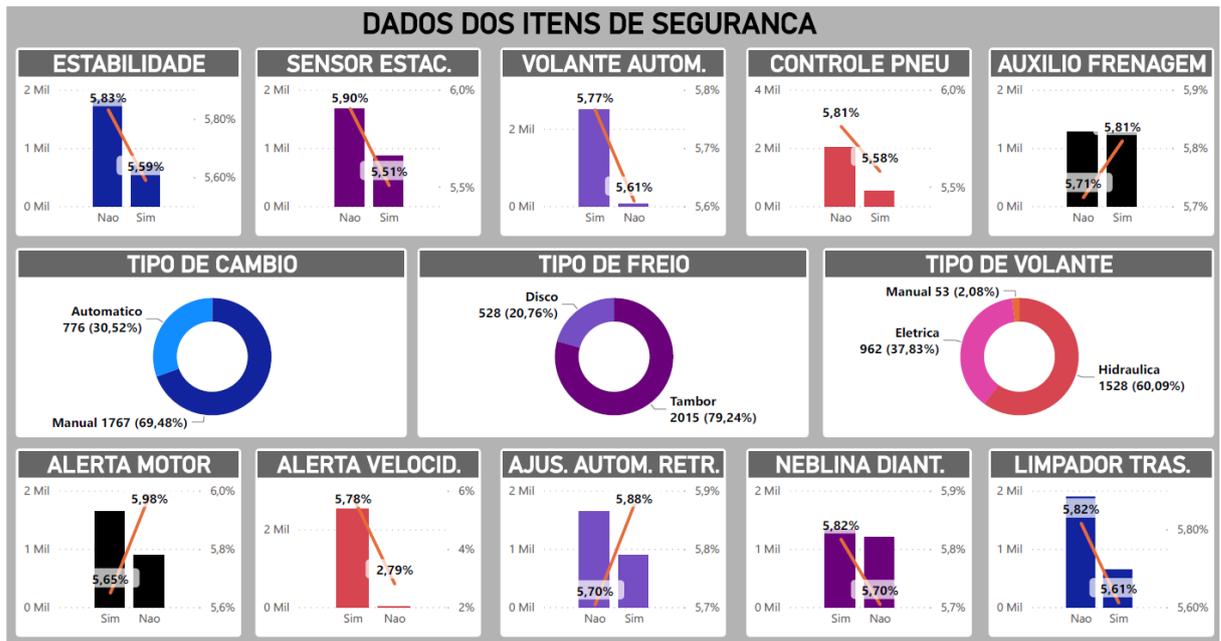


Figura 4.9: Modelo com os itens de segurança avaliados.

Fonte: Elaborada pelo autor, 2023.

Nas análises de dados de segurança três resultados foram observados: situações neutras em que não existe diferença material diante da existência do item; situações esperadas em que os carros que não possuíam os itens estiveram envolvidos em mais sinistros; e situações inesperadas em que os carros que possuíam os itens estiveram envolvidos em mais sinistros.

Como exemplo das situações neutras, a Figura 4.10 apresenta o item de sistema de auxílio de frenagem, que diante da análise realizada não apresentou diferença material quando comparados os veículos que possuíam e não possuíam o item, tendo uma média semelhante de sinistros.

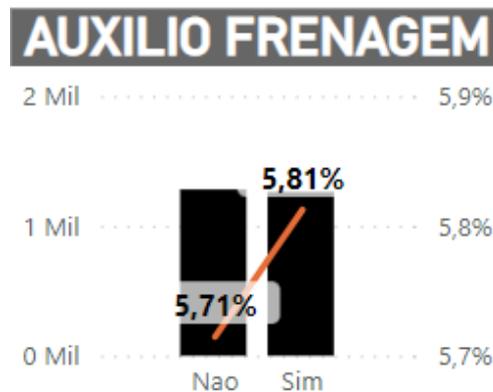


Figura 4.10: Situações neutras.

Fonte: Elaborada pelo autor, 2023.

As situações esperadas ocorreram para quatro itens de segurança e apresentaram diferenças significativas quando comparadas às quantidades de sinistros. Para sensores de estacionamento, veículos que não possuíam o item apresentaram 48% a mais de sinistros. Para controles de estabilidade, veículos que não possuíam o item apresentaram 62% a mais de sinistros. Para limpadores traseiros, veículos que não possuíam o item apresentaram 66% a mais de sinistros. E para controle de pressão dos pneus, apresentado na Figura 4.11, veículos que não possuíam apresentaram 73% a mais de sinistros.

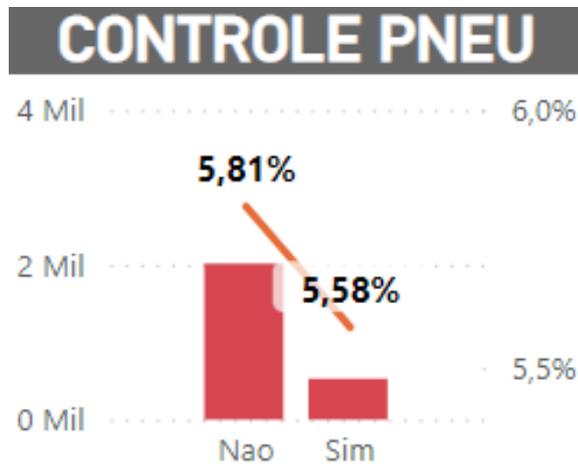


Figura 4.11: Situações esperadas.

Fonte: Elaborada pelo autor, 2023.

Já as situações inesperadas ocorreram para quatro itens de segurança, a saber: volante automático, alerta de motor e alerta de velocidade. Adicionalmente, foi identificada uma incidência maior de sinistros em veículos com câmbio manual, com freios tambor e direções hidráulicas.

Com exceção das situações neutras e inesperadas, que não apresentam itens necessariamente utilizáveis para o modelo, os resultados alcançados nas análises de itens de segurança foram inseridos nos indicadores.

#### 4.5 – Indicadores de *performance*

Concluídas as análises, os resultados foram inseridos em um modelo de indicador de *performance* visando demonstrar a existência de apólices consideradas como “alto risco”, evitando assim um aumento de situações de sinistros para a Seguradora Fictícia S.A. A Figura 4.12 é uma demonstração desses indicadores.

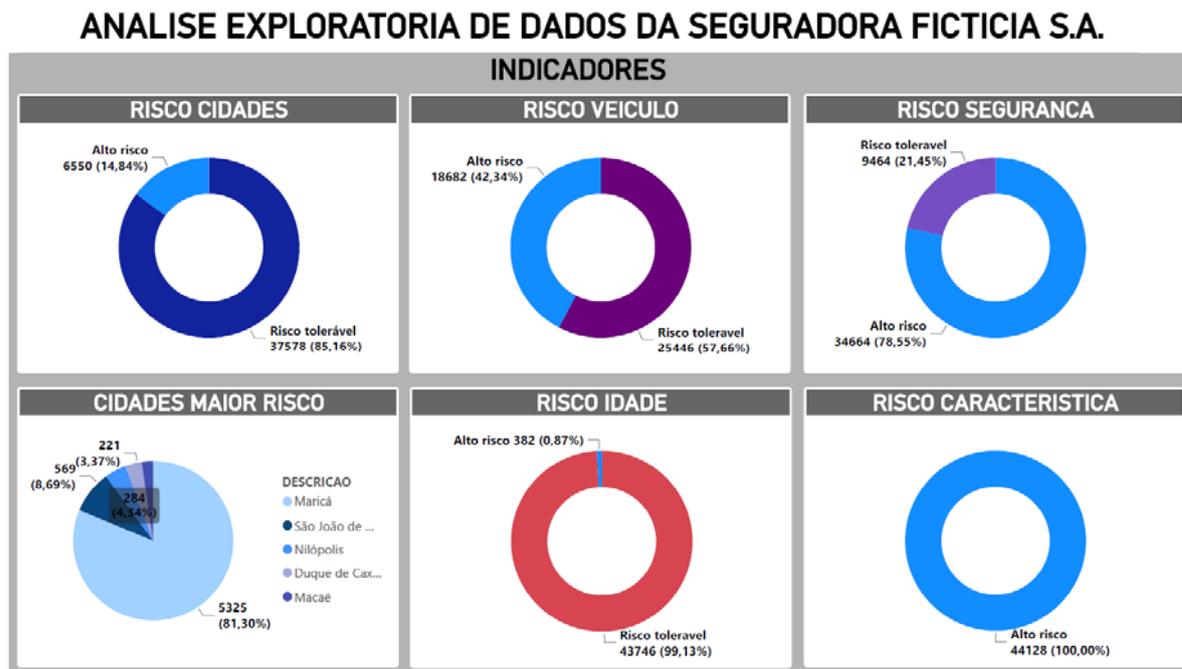


Figura 4.12: Exemplo de indicadores resultantes das análises.

Fonte: Elaborada pelo autor, 2023.

# CAPÍTULO 5

## Conclusão e Trabalhos Futuros

### 5.1 – Conclusão

O projeto aqui apresentado teve como propósito demonstrar a diversidade existente para a aplicação de análises de dados e como essa prática, aliada a outras ferramentas, pode incrementar os processos das companhias, não só na área de seguros de veículos, mas em todas as organizações.

As diversas ferramentas hoje disponíveis suportam as pessoas que se desafiarem a ampliar seus conhecimentos e expandir os benefícios às companhias através da eficiência e prática profissional.

Sempre foi, desde o primeiro capítulo, o foco desse projeto de trazer um resultado claro e aplicável a processos diversos. Iniciamos o projeto com uma visão sobre o tema, objetivo, delimitações e metodologia a ser aplicada.

No capítulo dois foram explorados os conceitos e teorias em uma revisão bibliográfica que apresentou conceitos sobre o mercado explorado, além de todos os recursos que foram utilizados ao longo do projeto.

No capítulo três a metodologia a ser aplicada foi detalhada, com apresentações dos conceitos e práticas que foram utilizadas ao longo do projeto, sempre suportadas na bibliografia que foi apresentada.

No capítulo quatro foram apresentados os resultados, explorando cada etapa do trabalho desenvolvido, incluindo a elaboração dos indicadores para suporte futuro às decisões, que visavam desde o início incrementar a acuracidade das decisões operacionais em pelo menos 30%, que no caso deste projeto seria semelhante a munir os tomadores de decisão com informações que pudessem reduzir em pelo menos 30% os riscos relacionados à concessão de seguros para clientes propensos a sinistros.

Através do editor de Linguagem M demonstrado na Figura 5.1, é possível observar como o modelo foi reinserido em uma outra base de clientes onde a etapa mineração de dados foi concluída, gerando informações sobre clientes não existentes durante as etapas anteriores do projeto.

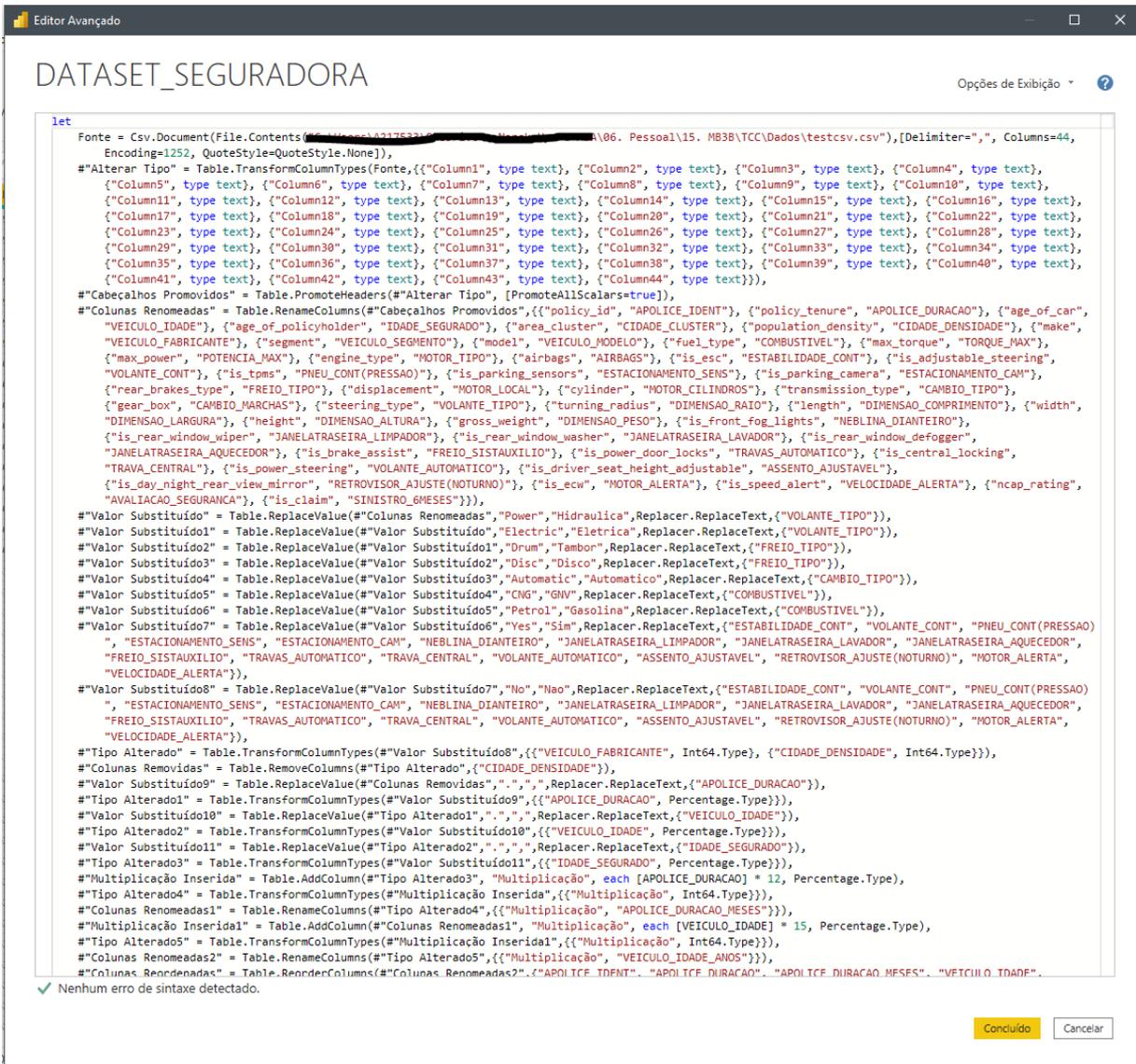


Figura 5.1: Modelo da Linguagem M executada em nova carteira de clientes.

Fonte: Elaborada pelo autor, 2023.

Após reprocessar o modelo em uma nova carteira de clientes, obtivemos resultados válidos para um potencial carteira com 29.338 que já estava livre de *outliers* e demais tratamentos previamente aplicados. Essa carteira, apresentada na Figura 5.2, resultou em indicadores que apresentavam para os riscos considerados como altos (resultados dos testes previamente realizados) resultados relevantes e que muniram a gestão com informações.

## ANALISE EXPLORATORIA DE DADOS DA SEGURADORA FICTICIA S.A.

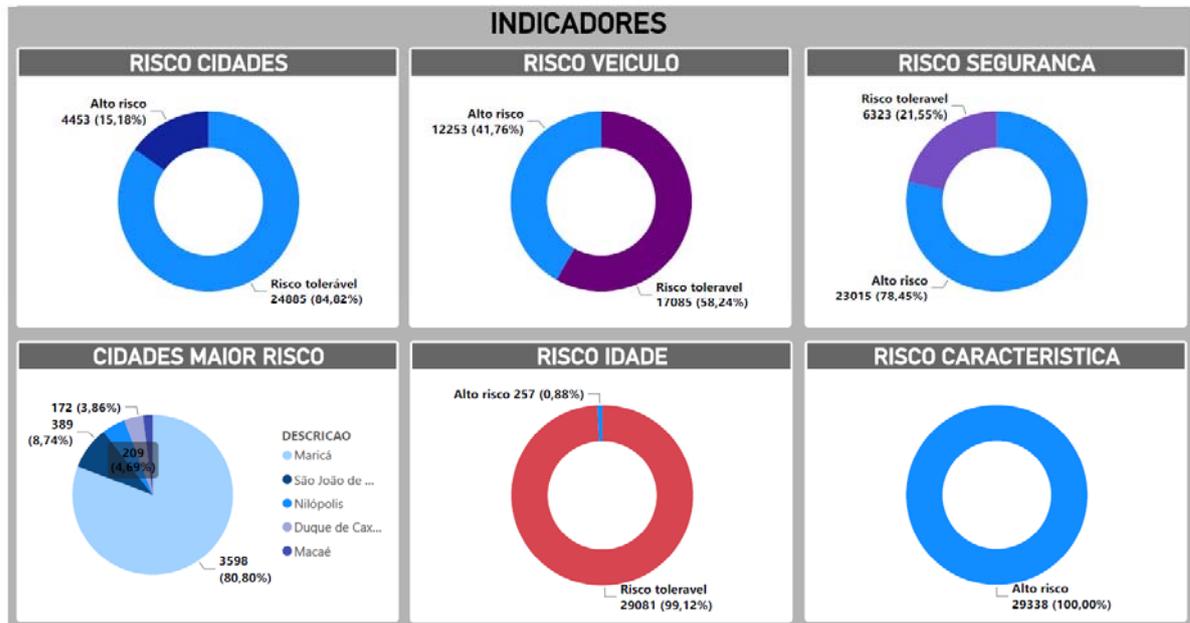


Figura 5.2: Resultados alcançados com o aprendizado de máquina.

Fonte: Elaborada pelo autor, 2023.

Considerando o exposto acima, a conclusão é que o projeto teve sucesso em munir a gestão com informações que são capazes de direcionar as decisões operacionais de forma adequada, reduzindo em pelo menos 30% a concessão de seguros a clientes com alto risco de sinistros.

### 5.2 – Trabalhos Futuros

A sequência desse projeto sugere a exploração de outros desafios corporativos seja em ramos de atuação das empresas ou processos a serem analisados. De qualquer forma o fator principal será a continuidade em extrair o máximo o possível de benefícios das análises de dados combinadas a outras possibilidades e ferramentas.

# Referências Bibliográficas

FATURAMENTO do setor de seguros. CNseg: **Confederação Nacional das Empresas de Seguros Gerais, Previdência Privada e Vida, Saúde Suplementar e Capitalização**. CNseg, 2022. Disponível em: <https://cnseg.org.br/noticias/cnseg-setor-de-seguros-alcanca-r-306-4-bilhoes-alta-de-11-9.html>. Acesso em: 31/03/2023.

INDICADORES estratégicos. **Gestão Estratégica – Tudo sobre a gestão de indicadores estratégicos, táticos e operacionais nas empresas**, 2022. Disponível em: <https://www.siteware.com.br/gestao-estrategica/indicadores-estrategicos/>. Acesso em: 02/05/2023.

INSTITUCIONAL SUSEP - **Superintendência de Seguros Privados, Ministério da Fazenda, SUSEP**, 2022. Disponível em: <https://www.gov.br/susep/pt-br/aceso-a-informacao/institucional/sobre-a-susep>. Acesso em: 31/03/2023.

MACHADO, Felipe Nery Rodrigues. **Tecnologia e projeto de Data Warehouse**. 6ª Edição. Saraiva, 2013.

NETSUITE **diagnostic analytics**. ORACLE, 2021. Disponível em: <https://www.netsuite.com/portal/resource/articles/data-warehouse/diagnostic-analytics.shtml>. Acesso em: 15/04/2023.

PARMENTER, David. **Key Performance Indicators – Developing, Implementing and Using winning KPIs**. 1ª Edição. John Wiley & Sons, 2007.

POWER BI **documentation**. Microsoft Learn. **Power BI**, 2023. Disponível em: <https://learn.microsoft.com/en-us/power-bi/>. Acesso em: 18/04/2023.

POWER QUERY **documentation**. Microsoft Learn. **Power Query**, 2023. Disponível em: <https://learn.microsoft.com/en-us/power-query/>. Acesso em: 18/04/2023.

SAIBA a importância de ter um seguro. **Seguro Auto.org, 2022**. Disponível em: <https://modesttipittolseguros.com.br/blog/183/saiba-a-importancia-de-ter-um-seguro-e-por-que-ele-e-um-servico-de-protecao-indispensavel>. Acesso em: 31/03/2023.

SILVA, Sandra. **Seguros, previdência e capitalização**. Senac – São Paulo, 2022.

SHARDA, Ramesh; DELEN, Dursun; TURBAN, Efraim. **Business Intelligence e Análise de Dados para gestão do negócio**. 4ª Edição. Bookman, 2019.

RAJAGOPAL, Balaji Ramkumar; KUMAR, Dr. B. Muthu; MAGAR, Shyamsundar Pralhad; JYOTHI, M. Lakshaga. **Basic Fundamentals of Machine Learning**. 1ª Edição. AGPH Books, 2022.

THE BEST DATA ANALYTICS TOOLS OF 2023. **Forbes advisor**, 2023. Disponível em: <https://www.forbes.com/advisor/business/software/best-data-analytics-tools/>. Acesso em 18/04/2023.