

# Projeto de Análise de Sinistros Automotivos: Estratégia de Dados para o Setor Privado

## 1. Dissertação sobre o Problema

### 1.1 Descrição Detalhada do Problema

Para empresas do setor privado, como locadoras de veículos e seguradoras, os sinistros automotivos representam um desafio operacional e financeiro complexo e multifacetado. A problemática vai muito além do custo imediato de um reparo; ela permeia toda a cadeia de valor do negócio, impactando a rentabilidade, a eficiência operacional e a satisfação do cliente.

**Para Locadoras de Veículos:** O problema central é a gestão de uma frota geograficamente dispersa e constantemente exposta a riscos. Cada sinistro resulta em uma cascata de perdas:

- **Custos Diretos:** Despesas com reparos, franquias de seguro e, em casos de perda total, a substituição do ativo. Um evento de grande escala, como as enchentes no Rio Grande do Sul, pode resultar na perda de milhares de veículos, com prejuízos que ultrapassam centenas de milhões de reais, especialmente porque muitas apólices de frotas não cobrem desastres naturais.
- **Custos Indiretos (Perda de Receita):** Um veículo parado para reparo é um ativo que não está gerando receita. A indisponibilidade da frota afeta diretamente a capacidade de atender à demanda, resultando em perda de oportunidades de negócio.
- **Custos Operacionais:** A gestão de sinistros consome tempo e recursos significativos, envolvendo processos administrativos, logística de reboque e reparo, e comunicação com clientes e seguradoras.

- **Impacto na Experiência do Cliente:** Um sinistro é um ponto de atrito significativo na jornada do cliente. A forma como a locadora gerencia a situação — desde a assistência imediata até a substituição do veículo — pode definir a percepção do cliente sobre a marca e influenciar sua fidelidade futura.

**Para Seguradoras e Associações de Proteção Veicular:** A problemática reside na precificação precisa do risco e na gestão eficiente dos custos de sinistros.

- **Assimetria de Informação e Precificação de Risco:** Tradicionalmente, a precificação de seguros baseia-se em dados demográficos estáticos (idade, sexo, local de residência), que são proxies imperfeitos para o risco real. Essa abordagem leva a subsídios cruzados, onde motoristas seguros pagam por aqueles de alto risco, e expõe a seguradora a uma seleção adversa.
- **Custos Operacionais de Sinistros:** O processo de regulação de um sinistro é caro e complexo. Envolve a verificação da apólice, investigação da ocorrência, avaliação de danos, combate a fraudes e, finalmente, o pagamento da indenização. Ineficiências nesse processo aumentam os custos operacionais e o tempo de resolução, impactando negativamente a satisfação do cliente.
- **Fraudes:** Sinistros fraudulentos representam um custo bilionário para o setor, exigindo investimentos robustos em detecção e prevenção.

## 1.2 Importância e Relevância do Problema

A relevância deste problema é tanto econômica quanto estratégica. Para as locadoras, a otimização da gestão de sinistros é uma alavanca direta para a melhoria da margem de lucro e para a construção de uma marca resiliente e confiável. Para as seguradoras, a capacidade de entender e precificar o risco de forma granular é a essência do seu modelo de negócio e um fator crítico de competitividade.

Em um mercado cada vez mais competitivo, com a ascensão de novos modelos de mobilidade, a experiência do cliente tornou-se um campo de batalha decisivo. Uma gestão de sinistros que é percebida como lenta, burocrática ou injusta pode levar à perda de clientes para concorrentes mais ágeis. Por outro lado, uma abordagem proativa e baseada em dados pode transformar um evento negativo (o sinistro) em uma oportunidade de demonstrar eficiência e cuidado, fortalecendo a lealdade do cliente.

### 1.3 Como a Análise de Dados Pode Ajudar a Solucionar o Problema

A análise de dados oferece um caminho para transformar a gestão de sinistros de um processo reativo e custoso para uma função estratégica e proativa. Ao analisar dados públicos de sinistros em conjunto com dados internos (telemática, histórico do cliente), as empresas podem:

- **Identificar "Hotspots" de Risco:** Utilizando análise geoespacial, é possível mapear as áreas geográficas (cidades, rodovias, bairros) com maior concentração de acidentes.<sup>1</sup> Para uma locadora, isso permite otimizar rotas, alertar clientes sobre trechos perigosos e informar decisões sobre onde alocar veículos com mais recursos de segurança. Para uma seguradora, permite refinar a precificação de risco com base na geografia e direcionar campanhas de prevenção.
- **Prever a Severidade e o Custo dos Sinistros:** Modelos de machine learning podem prever a probabilidade de um acidente ser grave com base em fatores como tipo de colisão, condições climáticas e características da via.<sup>4</sup> Isso permite que as seguradoras priorizem sinistros de alto custo potencial desde o primeiro aviso, otimizando a alocação de ajustadores e reduzindo o tempo de resolução.
- **Desenvolver Produtos Inovadores (UBI):** A análise de dados é a espinha dorsal do Seguro Baseado no Uso (*Usage-Based Insurance* - UBI). Ao combinar dados de telemática (como o motorista dirige) com dados contextuais (onde e quando ele dirige), as seguradoras podem oferecer apólices personalizadas que recompensam o comportamento seguro com descontos, atraindo clientes de baixo risco e incentivando uma condução mais segura.
- **Melhorar a Experiência do Cliente:** Ao antecipar riscos e agilizar processos, a análise de dados melhora a jornada do cliente. Uma locadora pode, por exemplo, usar a telemática para detectar uma colisão automaticamente e iniciar o processo de assistência sem que o cliente precise ligar. Uma seguradora pode usar a análise para acelerar a aprovação de sinistros de baixo risco, aumentando a satisfação e a retenção.

Em suma, a análise de dados permite que locadoras e seguradoras passem de uma visão agregada e reativa para uma compreensão granular e preditiva do risco, gerando mais valor para o negócio, reduzindo perdas e melhorando a experiência do cliente.

## 2. Levantamento das Fontes de Dados

Para a realização deste projeto, foram utilizadas fontes de dados públicas e não confidenciais, garantindo a transparência e a reprodutibilidade da análise.

Nome da Fonte	Descrição dos Dados	Tipo de Dados	Método de Acesso
<b>Polícia Rodoviária Federal (PRF) - Dados Abertos</b>	Dados detalhados de todos os Boletins de Acidente de Trânsito (BAT) registrados em rodovias federais desde 2007. Inclui informações cruciais como localização (latitude, longitude), causa, tipo, classificação de gravidade, condições da via e do tempo, veículos e pessoas envolvidas. <sup>7</sup>	Estruturados	Download direto de arquivos anuais no formato .csv, compactados em .zip. <sup>7</sup>
<b>Secretaria Nacional de Trânsito (SENATRAN)</b>	Fornecer estatísticas consolidadas sobre a frota de veículos (RENAVAM), condutores habilitados (RENACH) e, de forma crescente, estatísticas de sinistros através do RENAEST. Esses dados são úteis para contextualizar o tamanho da frota em diferentes municípios e estados.	Estruturados	Download de anuários em .pdf e acesso a painéis de dados. O RENAEST está em fase de implantação para coleta via aplicativo.

<b>Instituto Brasileiro de Geografia e Estatística (IBGE)</b>	Dados sobre a frota de veículos por município e estimativas populacionais. Essencial para normalizar os dados de acidentes e calcular taxas de risco (ex: acidentes por 10.000 veículos).	Estruturados	Consulta interativa na plataforma Cidades@IBGE e download de tabelas. Algumas APIs e pacotes (como sidrar para R) também permitem o acesso programático. <sup>11</sup>
<b>Órgãos Estaduais e Municipais (Exemplos)</b>	Diversos órgãos estaduais (como SEJUSP-MG) e municipais (como CET-SP ou BHTrans <sup>14</sup> ) também disponibilizam dados de acidentes. Embora não utilizados diretamente neste núcleo de análise, eles representam fontes valiosas para análises mais localizadas.	Estruturados	Varia por órgão, mas geralmente inclui painéis interativos e, em alguns casos, download de arquivos .csv.

A principal fonte para este projeto são os dados da PRF, devido à sua abrangência nacional, granularidade e disponibilidade consistente em formato aberto e legível por máquina.<sup>15</sup>

### 3. Análise Exploratória de Dados (EDA)

A Análise Exploratória de Dados (EDA) é o processo de investigação inicial dos dados para descobrir padrões, identificar anomalias, testar hipóteses e verificar suposições com a ajuda de estatísticas resumidas e representações gráficas.

### 3.1 Metodologia de Processamento de Dados

Para realizar a análise, foi desenvolvido um script em Python que automatiza todo o fluxo de trabalho, desde a coleta até a preparação dos dados para visualização. As principais etapas do script são:

1. **Coleta e Extração:** O script inicia baixando os dados de acidentes da PRF (para o ano de 2023, neste caso) a partir de um link do Google Drive. Os dados, que estão em um arquivo .zip, são extraídos automaticamente.
2. **Limpeza e Pré-processamento:** Os dados brutos são carregados em um DataFrame do Pandas. É realizada uma limpeza essencial, que inclui a conversão de colunas de data para o formato datetime, a correção de formatos numéricos (latitude, longitude) e a remoção de registros com informações críticas ausentes.
3. **Engenharia de Features:** Novas colunas (features) são criadas para enriquecer a análise, como ano, mes e hora do acidente. Uma coluna alvo, `nivel_gravidade`, é criada para classificar os acidentes em "Grave ou Fatal" e "Leve ou Sem Vítimas", simplificando o problema para a modelagem.
4. **Modelagem Preditiva:** Um modelo de Machine Learning (`RandomForestClassifier`) é treinado utilizando features como dia da semana, fase do dia, condição meteorológica e tipo de acidente para prever a gravidade.
5. **Geração de Previsões:** O modelo treinado é utilizado para gerar previsões de gravidade e a probabilidade de um acidente ser grave para *todos os registros* do dataset. Essas informações são adicionadas como novas colunas na tabela principal.
6. **Análise de Importância:** O script também extrai a importância de cada variável do modelo, identificando quais fatores mais influenciam na previsão da gravidade de um acidente. O resultado é salvo como um gráfico de barras e um arquivo .csv.
7. **Exportação:** Ao final, o script gera uma única planilha no formato Excel (.xlsx), contendo os dados limpos, as features criadas e as previsões do modelo. Este arquivo é o produto final, pronto para ser importado em ferramentas de Business Intelligence como o Looker Studio para a criação dos dashboards.

Este processo garante que a análise seja reprodutível, eficiente e que os dados estejam devidamente estruturados para a etapa de visualização e extração de insights.

## 4. Relatório de Insights

Com base na análise dos dados processados, desenvolvemos uma análise mais profunda para extrair insights acionáveis para locadoras e seguradoras.

#### 4.1 Resumo dos Principais Achados e Relevância para o Negócio

A análise revelou dois conjuntos principais de insights:

1. **O Risco é Geograficamente Concentrado:** Os acidentes não ocorrem de forma aleatória. Existem "hotspots" claros, ou seja, trechos de rodovias e áreas urbanas que concentram um número desproporcional de sinistros. A identificação desses hotspots é fundamental para a gestão de risco. Uma locadora pode usar essa informação para desviar frotas de áreas de alto risco ou para equipar veículos que circulam nessas áreas com tecnologia de segurança adicional. Uma seguradora pode usar os hotspots para refinar seus modelos de precificação de risco geográfico.
2. **A Gravidade do Acidente é Preditível:** Fatores como o tipo de acidente (ex: colisão frontal), condições climáticas (ex: chuva), fase do dia (ex: noite) e tipo de via (ex: pista simples) são fortes preditores da gravidade de um sinistro. Entender o peso de cada um desses fatores permite a criação de estratégias de mitigação direcionadas.

#### 4.2 Fatores Determinantes para a Gravidade dos Sinistros

Para quantificar o impacto de cada fator, o modelo de *Random Forest* foi analisado para extrair a importância das variáveis. O gráfico abaixo ilustra os 15 fatores mais relevantes que o modelo utiliza para prever se um acidente será grave ou fatal.

---

Análise dos Resultados:

O gráfico de importância das variáveis revela insights cruciais:

- **Tipo de Acidente é Crítico:** O fator de maior peso é o tipo\_acidente\_Colisão frontal. Isso confirma que colisões frontais são, de longe, o tipo de sinistro com maior potencial de resultar em ferimentos graves ou fatais. Em segundo lugar,

Atropelamento de Pedestre também figura entre os mais importantes.

- **Condições de Visibilidade e Horário:** A fase\_dia\_Plena Noite aparece como o segundo fator mais importante, destacando os riscos inerentes à condução noturna, seja pela menor visibilidade ou pela fadiga do condutor.
- **Comportamento e Ambiente:** Fatores como causa\_acidente\_Velocidade Incompatível e tracado\_via\_Curva também estão no topo da lista, mostrando que a combinação de comportamento de risco com características desafiadoras da via aumenta significativamente a probabilidade de um desfecho grave.

### Sugestões de Ações com Base nos Insights:

- **Para Locadoras:**
  - **Gestão de Risco de Frota:** Utilizar o mapa de hotspots para definir "zonas de alerta". Veículos que entram nessas zonas podem ter seus parâmetros de condução monitorados mais de perto via telemática.
  - **Alertas ao Cliente:** Integrar os insights ao aplicativo da locadora, enviando notificações proativas aos clientes, como: "Você está entrando em uma área com histórico de acidentes em dias de chuva. Redobre a atenção."
- **Para Seguradoras:**
  - **Subscrição Inteligente:** Incorporar os fatores de risco (geográficos e comportamentais) nos modelos de precificação. Um cliente que dirige frequentemente à noite em pistas simples e em regiões chuvosas representa um risco maior.
  - **Desenvolvimento de UBI:** Usar os fatores de maior importância como base para um modelo de *Pay-How-You-Drive* (PHYD). Pontuar os motoristas com base em quão frequentemente eles se expõem a condições de alto risco.

## 5. Visualização de Dados com os Resultados

Para que os insights sejam consumidos de forma eficaz por gestores e equipes operacionais, foi desenvolvido um dashboard no Looker Studio, dividido em três painéis estratégicos, cada um com um público-alvo específico.

### 5.1 Dashboard - Visão Geral



Este painel oferece uma visão macro da sinistralidade, permitindo uma compreensão rápida das tendências e dos principais pontos de atenção. É a porta de entrada para a análise, ideal para executivos e gestores que precisam de um resumo do cenário geral.

---

### **Insights e Ações:**

#### **Distribuição Geográfica:**

O *Mapa de Calor de Acidentes* evidencia uma concentração significativa de ocorrências nas regiões Sudeste, Sul e Nordeste. Esse padrão sugere a necessidade de priorizar rotas com maior volume de sinistros em estratégias de prevenção, fiscalização e cobertura securitária.

#### **Fatores Humanos como Principal Causa:**

O gráfico de *Principais Causas de Acidente* revela que fatores relacionados à conduta do motorista, como “Reação Tardia” e “Ausência de Reação”, representam a maior parte dos sinistros. Isso indica um forte potencial de impacto por meio de programas de capacitação, uso de tecnologias embarcadas e campanhas educativas.

#### **Gravidade por Tipo de Acidente:**

A análise de gravidade mostra que certos tipos de ocorrência, como *colisão com objeto* e *colisão traseira*, apresentam maior proporção de casos fatais. Esse dado é crucial para reforçar medidas de segurança veicular e melhorar sinalização e infraestrutura em locais com maior incidência desses acidentes.

#### **Prioridades Regionais:**

O gráfico de *Incidência por Estado* aponta que Minas Gerais, Santa Catarina e Paraná lideram em volume de sinistros. Essa informação possibilita o direcionamento de recursos e ações específicas por região, otimizando os investimentos em prevenção e assistência.

## **5.2 Dashboard - Produto para Gestor de Frotas**

Este painel é desenhado especificamente para as necessidades de um gestor de frotas de uma locadora de veículos. O foco está em otimizar a operação, reduzir custos e garantir a segurança dos clientes e dos ativos.

---

## Insights e Ações:

### Causas Ligadas ao Tipo de Acidente:

O gráfico *Principais Tipos de Acidente por Causa* revela que a combinação de **falhas humanas** (como reação tardia, velocidade incompatível e ausência de atenção) resulta em acidentes graves como **colisão frontal, capotamento e colisão traseira**. Isso sugere que **intervenções comportamentais e tecnológicas**, como alertas de fadiga ou velocidade, podem reduzir significativamente a severidade dos sinistros.

### Fatores de Risco Críticos:

A análise de *Fatores de Risco* aponta variáveis com alto peso na ocorrência de sinistros, como:

- **Colisão frontal**
- **Céu claro (condição meteorológica)** – sugerindo que acidentes ocorrem mesmo em boas condições, reforçando o papel do fator humano
- **Pista simples e trechos em linha reta**, onde pode haver excesso de confiança  
Esses fatores orientam decisões sobre **investimentos em tecnologias embarcadas, alertas em tempo real** e treinamentos específicos para situações de risco previsível.

### Zonas de Alta Incidência:

O novo *Mapa de Calor* reflete áreas críticas de sinistros, com grande concentração nas regiões Sul, Sudeste e Centro-Oeste. A visualização reforça a necessidade de monitoramento e **ação geográfica direcionada**, seja com campanhas locais ou ajustes de rota.

## 5.3 Dashboard - Produto para Seguradoras

Este painel é uma ferramenta estratégica para subscritores, analistas de risco e gestores de sinistros de seguradoras. Ele traduz os dados de acidentes em insights financeiros e de risco, permitindo uma precificação mais justa e uma gestão de sinistros mais eficiente.

---

## Insights e Ações:

### Cálculo Dinâmico de Risco:

A *Calculadora de Nota de Risco* considera variáveis como **estado, cidade, tipo de pista, condições meteorológicas e dia da semana**, permitindo estimar uma **nota de risco média (ex: 1,86 – Risco Médio)**. Essa funcionalidade é útil para:

- Precificação de apólices personalizadas
- Ajustes de rota preventiva
- Monitoramento preditivo de segurança operacional

### Concentração de Acidentes por Trecho:

O *Mapa de Calor por Rodovia* destaca com clareza os trechos mais críticos da malha rodoviária, especialmente nas regiões Sudeste e Sul. Esse recurso facilita:

- Planejamento logístico com rotas mais seguras
- Identificação de pontos que demandam infraestrutura ou sinalização reforçada

### Fatores de Risco Mais Relevantes:

O painel de *Fatores de Risco* mantém como principais variáveis:

- Colisão frontal
- Pistas simples
- Céu claro (quando o excesso de confiança pode ser um risco)  
Esse conjunto reforça a necessidade de **ações educativas e uso de tecnologia embarcada mesmo em condições aparentemente seguras.**

## **Tipos de Acidente Mais Comuns:**

*Queda de ocupante, tombamento e colisão traseira* continuam figurando entre os tipos mais recorrentes. Isso indica:

- A importância de políticas de **capacitação de condutores**
- Reforço no uso de **cintos de segurança, estabilizadores de carga** e monitoramento de direção defensiva.

## **Entregáveis do Projeto**

Este documento e os artefatos gerados cumprem os entregáveis definidos para o projeto:

1. **Documento de Dissertação:** A Seção 1 deste relatório detalha a problemática, sua relevância e o papel da análise de dados.
2. **Relatório de Fontes de Dados:** A Seção 2 documenta as fontes de dados públicos, seus tipos e métodos de coleta.
3. **Metodologia de Análise de Dados:** A Seção 3.1 descreve o processo automatizado via script Python para a EDA e preparação dos dados.
4. **Relatório de Insights:** A Seção 4 apresenta os principais achados da análise, incluindo a análise de importância das variáveis, sua relevância para o negócio e sugestões de ações.
5. **Visualização de Dados:** A Seção 5 descreve a estrutura e os insights dos dashboards interativos criados no Looker Studio para os diferentes públicos-alvo.

Para a entrega final, o script Python, os relatórios de importância das variáveis, este documento em formato Markdown e o link para o dashboard interativo seriam organizados em um repositório no GitHub, garantindo a documentação completa, a reprodutibilidade da análise e a facilidade de acesso para todas as partes interessadas.

## **Referências citadas**

1. Spatiotemporal Analysis of Traffic Accidents Hotspots Based on Geospatial Techniques, acessado em julho 12, 2025,

- <https://www.mdpi.com/2220-9964/11/4/260>
2. Geração de mapas de hotspots em redes de ruas para predição de crimes - Universidade Federal do Ceará, acessado em julho 12, 2025, [https://repositorio.ufc.br/bitstream/riufc/51515/3/2020\\_dis\\_fcfunesjunior.pdf](https://repositorio.ufc.br/bitstream/riufc/51515/3/2020_dis_fcfunesjunior.pdf)
  3. GIS-Based Spatial Analysis of Accident Hotspots: A Nigerian Case Study - MDPI, acessado em julho 12, 2025, <https://www.mdpi.com/2412-3811/7/8/103>
  4. sugatagh/Road-Traffic-Accident-Severity-Classification - GitHub, acessado em julho 12, 2025, <https://github.com/sugatagh/Road-Traffic-Accident-Severity-Classification>
  5. SrikantShubam/Road-Accident-Class-severity-classification: Python machine learning project that classifies road accidents' severity using various algorithms. It's a valuable resource for individuals interested in machine learning and predictive modeling. - GitHub, acessado em julho 12, 2025, <https://github.com/SrikantShubam/Road-Accident-Class-severity-classification>
  6. Accident Severity Prediction in Big Data Using Auto-Machine Learning - Scientia Iranica, acessado em julho 12, 2025, [https://scientiairanica.sharif.edu/article\\_23141.html](https://scientiairanica.sharif.edu/article_23141.html)
  7. Dados Abertos da PRF — Polícia Rodoviária Federal - Portal Gov.br, acessado em julho 12, 2025, <https://www.gov.br/prf/pt-br/aceso-a-informacao/dados-abertos/dados-abertos-da-prf>
  8. Pesquisas com base nos dados abertos da PRF - Vias Seguras, acessado em julho 12, 2025, [http://vias-seguras.com/dados\\_da\\_prf](http://vias-seguras.com/dados_da_prf)
  9. PLANO DE GOVERNO 2025 – 2028 - O TEMPO, acessado em julho 12, 2025, [https://eleicoes2024candidatosapi.otempo.com.br/api/proposals/download/SP/2024SP250001885611\\_01.pdf](https://eleicoes2024candidatosapi.otempo.com.br/api/proposals/download/SP/2024SP250001885611_01.pdf)
  10. Secretaria de Estado de Justiça e Segurança Pública - SEJUSP | MG.GOV.BR, acessado em julho 12, 2025, [https://www.mg.gov.br/instituicao\\_unidade/secretaria-de-estado-de-justica-e-seguranca-publica-sejusp](https://www.mg.gov.br/instituicao_unidade/secretaria-de-estado-de-justica-e-seguranca-publica-sejusp)
  11. sidrar: An Interface to IBGE's SIDRA API, acessado em julho 12, 2025, <https://cran.r-project.org/web/packages/sidrar/sidrar.pdf>
  12. Dados do IBGE com o Pacote SIDRAR - YouTube, acessado em julho 12, 2025, <https://www.youtube.com/watch?v=EvBP4vfOkYs>
  13. open-geodata/br\_ibge\_sidra: Consumir dados da API do IBGE: SIDRA - GitHub, acessado em julho 12, 2025, [https://github.com/open-geodata/br\\_ibge\\_sidra](https://github.com/open-geodata/br_ibge_sidra)
  14. Conjunto de dados - Portal de Dados Abertos, acessado em julho 12, 2025, [https://dados.pbh.gov.br/dataset/?\\_groups\\_limit=0&organization=bhtrans&tags=Acidentes](https://dados.pbh.gov.br/dataset/?_groups_limit=0&organization=bhtrans&tags=Acidentes)
  15. Dados Abertos — Polícia Rodoviária Federal - Portal Gov.br, acessado em julho 12, 2025, <https://www.gov.br/prf/pt-br/aceso-a-informacao/dados-abertos>
  16. DASHCARE: Um Dashboard para Análise dos Acidentes nas Rodovias Federais - Repositório de Universidade Federal Rural do Semi-Árido, acessado em julho 12, 2025, <https://repositorio.ufersa.edu.br/bitstreams/e951841b-ac4a-4180-a82a-173ea1401>

[be6/download](#)

17. Acidentes de Trânsito - Base dos Dados, acessado em julho 12, 2025,  
<https://basedosdados.org/dataset/5b33d4d8-954b-4cb2-9c65-ba009dd16f9a>
18. OBSERVATÓRIO disponibiliza painel com dados da PRF, acessado em julho 12, 2025,  
<https://www.onsv.org.br/comunicacao/materias/observatorio-disponibiliza-painel-com-dados-da-prf>