

# Uso de Business Intelligence para mensuração de acidentes rodoviários através da base de dados da Polícia Rodoviária Federal

Lucca Carnauba<sup>1</sup>, Mayson Medeiros<sup>1</sup>, Edison Camilo<sup>1</sup>, Jailton Cardoso<sup>1</sup>, Flávio Medeiros<sup>1</sup>

<sup>1</sup>Coordenação de Informática – Instituto Federal de Alagoa (IFAL)  
Código Postal – CEP: 57020-600 – Maceió – AL – Brazil

{lcpr1,mfms1,edison,jailton,flavio.medeiros}@ifal.edu.br

**Abstract.** *This study analyzes accidents on Brazilian federal highways with data from the DPRF (2019-2023) with the aim of answering four research questions. The methodology used was exploratory and quantitative research using Business Intelligence techniques. The results indicate that weather conditions have little influence on the occurrence of traffic accidents. However, a significant increase in the number of accidents was observed during weekends and after 5 pm.*

**Resumo.** *Este estudo analisa acidentes nas rodovias federais brasileiras com dados do DPRF (2019-2023) com o objetivo de responder a quatro questões de pesquisa. A metodologia utilizada foi a pesquisa exploratória e quantitativa utilizando técnicas de Business Intelligence. Os resultados indicam que as condições climáticas têm pouca influência na ocorrência de acidentes de trânsito. No entanto, observou-se um aumento significativo no número de acidentes durante os finais de semana e após às 17 horas.*

## 1 Introdução

O Brasil possui cerca de 1.720.700 km de estradas e rodovias. A malha rodoviária federal, excluindo vias planejadas, soma 75.257 km, dos quais 65.735 km (87,34%) são pavimentados e 9.522 km (12,65%) não são pavimentados. Desses, 11.429 km são administrados por concessão federal [Ministério da Infraestrutura 2019].

O Brasil enfrenta grandes desafios de mobilidade com sua vasta malha rodoviária, sendo os acidentes de trânsito a segunda principal causa de morte desde os anos 1980, afetando principalmente jovens adultos e tendo grande impacto econômico [Rocha 2009]. A frota de veículos cresceu de 80 milhões em 2013 para 120 milhões em 2023 [Confederação Nacional do Transporte 2023], com 392.929 mortes registradas entre 2010 e 2019, especialmente entre jovens de 15 a 29 anos [Ipea 2023]. Com o avanço tecnológico da última década, é importante aprofundar o conhecimento sobre o tema e buscar alternativas para mitigar esses impactos.

*Business Intelligence* (BI) converte dados brutos em informações significativas, facilitando decisões baseadas em dados. BI consiste em uma série de arquiteturas e técnicas com bancos de dados, *data warehousing* e mineração de dados, que transformam dados brutos em informações úteis para fornecer suporte às decisões

[Detoni Junior et al. 2019]. O processo de *data warehousing* trata de quatro fases: extração de dados, transformação do dado bruto, carga dos dados no repositório e consulta aos dados [Pinto 2020].

Neste estudo desenvolvemos a ferramenta BI-SafeRoad, disponível no link (<https://github.com/carnaubarosario/Datatran-Consolidado>), utilizando técnicas de BI e data warehousing, que visa obter conhecimentos sobre os acidentes de trânsito utilizando a base de dados aberta mantida pela Polícia Rodoviária Federal [PRF 2024], que registra acidentes nas rodovias federais brasileiras desde 2007. O objetivo é identificar os principais fatores de risco, características dos envolvidos, tendências e padrões dos acidentes, e suas implicações para a saúde e mobilidade.

A relevância do estudo, e também da criação da ferramenta BI-SafeRoad, se destaca pela escassez de literatura acadêmica que utiliza a base de dados aberta Datatran [PRF 2024] para analisar acidentes nas rodovias federais, contribuindo para o avanço do conhecimento científico podendo gerar intervenções nas estradas que podem reduzir os índices de acidentes e mortalidade nas rodovias. A análise detalhada desses fatores é essencial para tomadas de decisões, como por exemplo, instalação de sonorizadores e radares, ampliar a sinalização das estradas, entre outras.

## 2 Metodologia

O presente artigo é uma pesquisa exploratória que visa investigar os acidentes de trânsito nas rodovias federais brasileiras. A pesquisa tem objetivo de responder a quatro questões de pesquisa, identificar as principais variáveis importantes e estabelecer as bases para estudos futuros.

Neste estudo, adotamos uma abordagem quantitativa baseado nos dados coletados na base de dados aberta da PRF [PRF 2024] e analisado os dados através de uma ferramenta de *dashboard* criada no *Power BI*. Para o desenvolvimento do trabalho selecionamos os cinco últimos anos de dados abertos disponíveis pela PRF.

Para realizar a metodologia de pesquisa foram realizadas buscas por artigos na base de dados Google Scholar e na plataforma de periódico da CAPES, usando as palavras-chave "acidentes em rodovias", "*business intelligence*", "Datatran PRF" e seus sinônimos "acidente em rodovia federal", "BI", "inteligência de negócios", "base de dados da Polícia Rodoviária Federal", "dados abertos acidentes PRF". A partir dessas palavras-chave, foi gerada uma *string* de busca utilizada nas bases mencionadas.

É importante destacar que o termo "BI" ou "Inteligência de Negócios" foi incluído para verificar trabalhos que utilizam essa tecnologia, considerando o período de 2019 a 2023 para a seleção das publicações. A Tabela 1 mostra as *strings* de busca utilizadas para a busca. A Tabela 2 apresenta as questões de pesquisa.

**Tabela 1. String de Busca para busca de estudos**

<b>Strings de Busca</b>
("Acidentes em Rodovias" OR "Acidente em Rodovia Federal" OR "Datatran PRF" OR "Base de dados da Polícia Rodoviária Federal" OR "Dados Abertos Acidentes PRF") AND ("Business Intelligence" OR "BI" OR "Inteligência de Negócios")

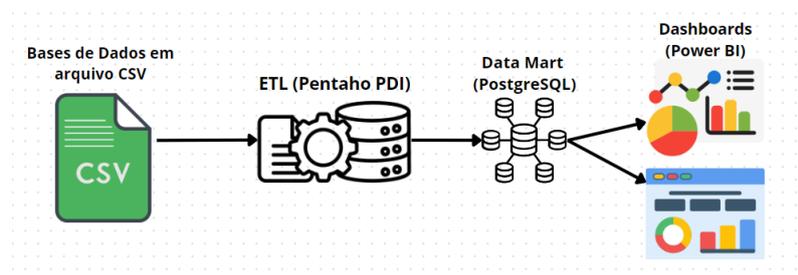
**Tabela 2. Questões de Pesquisa**

<b>QP</b>	<b>Descrição da Questão de Pesquisa</b>
QP1	Quais os principais tipos e causas de acidentes em rodovias federais?
QP2	Quais os estados e rodovias possuem maior registro de acidentes e vítimas fatais?
QP3	Qual o comportamento temporal dos acidentes?
QP4	Quais as influências das condições meteorológicas no número de acidentes e vítimas fatais?

**Fonte: Autoria própria (2024).**

### 3 BI-SafeRoad

A ferramenta criada BI-SafeRoad utiliza a arquitetura de BI (Figura 1), baseada na estrutura da arquitetura do [Kimball and Ross 2011]. A Figura 1 mostra a arquitetura da ferramenta, que parte de uma base de dados aberta de um arquivo csv e seu metadados, disponibilizado pela PRF [PRF 2024], passando pelos processo de ETL (*Extract, Transformation, Load*) para tratamento e carga dos dados.

**Figura 1. Arquitetura do BI-SafeRoad**

Segundo [Kimball and Ross 2011], o *data mart* utiliza a técnica de modelagem dimensional esquema estrela que divide as tabelas entre fato e dimensões. No modelo dimensional criado para a carga de dados no *data mart* (Figura 2) utiliza a tabela fato acidentes tendo como principais métricas: número de mortos, número de envolvidos e número de veículos. As dimensões criadas no modelo dimensional foram: causa

acidente, tipo acidente, localidade, hora, tipo pista, condição meteorológica e tempo.

O modelo dimensional criado para a carga de dados no *data mart* (Figura 2) utiliza a tabela fato acidentes tendo como principais métricas: número de mortos, número de envolvidos e número de veículos. Já as dimensões criadas no modelo dimensional foram: causa acidente, tipo acidente, localidade, hora, tipo pista, condição meteorológica e tempo.

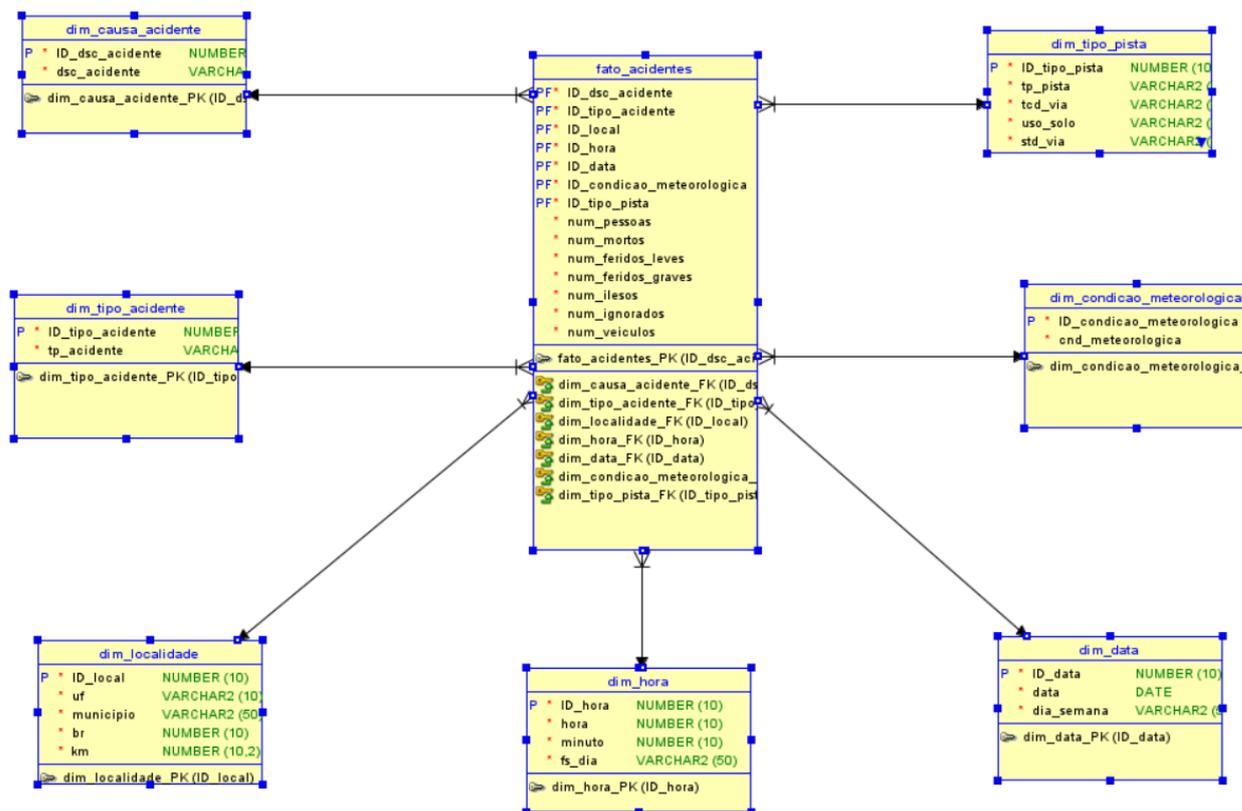


Figura 2. Modelo Dimensional do BI-SafeRoad

Utilizamos a ferramenta *Pentaho Data Integration* PDI para realizar o processo de ETL dos dados abertos da PRF, conforme Figura 3, a fim de carregar as tabelas dimensões e a tabela fato. Pentaho PDI, também conhecido como Kettle é uma ferramenta de integração de dados de código aberto que permite a criação de *pipelines* de ETL de dados de diversas fontes, assim facilitando futuras análises e o relatório de informações. [Detoni Junior et al. 2019]

Apesar da base de dados estar bem estruturada originalmente, no processo de ETL foram encontrados alguns dados errôneos, como valores com hífen em colunas numéricas. Para corrigir esse erro, foram utilizados *steps* adequados do Pentaho PDI para tratar esse tipo de erro.

A Figura 3 apresenta um processo de ETL real do projeto, que é a carga da tabela fato acidentes. O primeiro *step* é a leitura da base de dados consolidada, onde vai ser carregado todos os campos de dados da base. Após a leitura da base, acontecem uma série de *lookups*, que servem para associar os dados de entrada às dimensões através das chaves primárias das dimensões. O primeiro *lookup* é da dimensão causa do acidente, que

busca a chave primária da dimensão e seus campos. Após esse primeiro *lookup*, vem os *lookups* para a dimensão condição meteorológica, seguido por dimensão data, dimensão hora, dimensão localidade, dimensão tipo de acidente e tipo de pista. Por fim, a tabela fato é carregada, consolidando as informações para análise.



Figura 3. Processo ETL no *Pentaho PDI*

O Sistema de Gerenciamento de Banco de Dados (SGBD) escolhido para o *data mart* foi o PostgreSQL, que possui licença de código aberto e é conhecido por ser uma opção mais robusta de acesso gratuito [PostgreSQL 2024]. Já a ferramenta para a criação do *dashboard* foi escolhido o *Power BI*. O *Power BI Desktop* é uma ferramenta gratuita da *Microsoft* que permite a conexão, transformação e visualização de dados. Ele é amplamente utilizado para criar relatórios e *dashboards* interativos, que podem ser compartilhados com outras pessoas através dos serviços do *Power BI* [Microsoft 2024].

O *dashboard* desenvolvido permite uma visualização clara e precisa dos KPIs total de acidentes, taxa de mortalidade e média de acidentes por estado (Figura 4), proporcionando conhecimentos valiosos para a tomada de decisões.

Além disso, a interatividade do *Power BI* permite que os usuários filtrem (por período de data, ano, estado e rodovia) e explorem os dados de maneira personalizada, ajustando as visualizações de acordo com suas necessidades específicas.

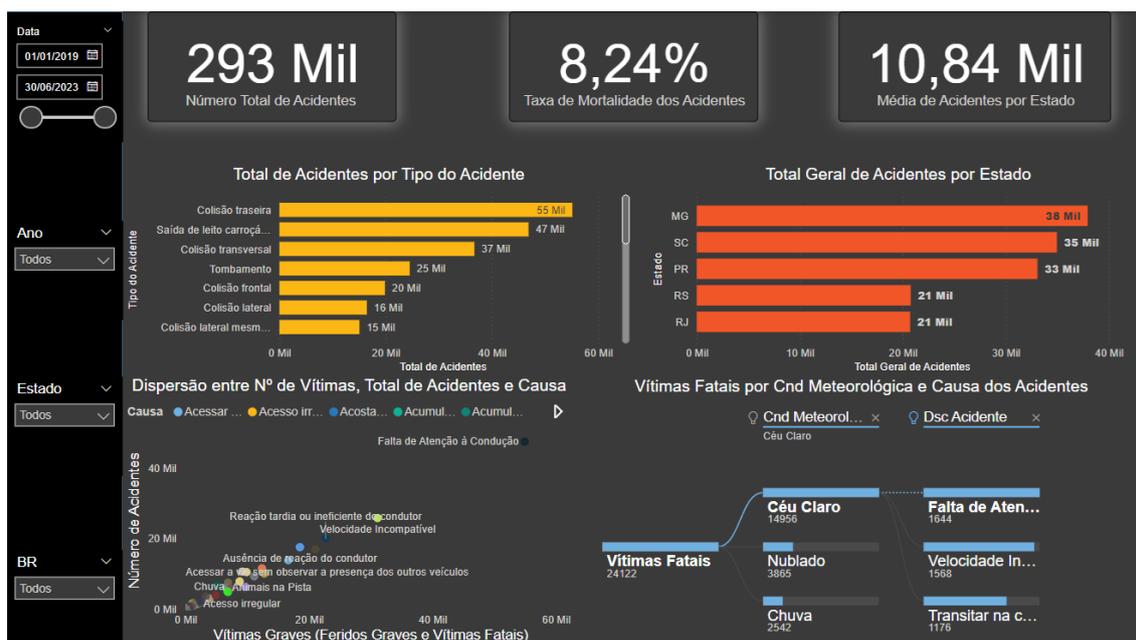


Figura 4. *Dashboard BI-SafeRoad*

A Figura 5 mostra o comportamento temporal dos acidentes, incluindo pico de horário, número de acidentes por mês, dias com mais acidentes e quantidade de acidentes anual. Além disso, a figura traz indicadores como a taxa de crescimento anual de acidentes, o mês com mais acidentes e o dia com maior média de acidentes.

## 4 Resultados

A análise de *dashboard* revelou que a colisão traseira foi o tipo de acidente mais frequente, com cerca de 55 mil registros, e a principal causa foi a falta de atenção do condutor, responsável por 47.433 acidentes. Minas Gerais liderou em número de ocorrências. Aproximadamente 15 mil vítimas fatais ocorreram em condições climáticas favoráveis, com as principais causas sendo falta de atenção do condutor e velocidade incompatível. No total, 8% dos acidentes resultaram em vítimas fatais, com uma média de aproximadamente 11 mil acidentes por estado, totalizando cerca de 293 mil acidentes no período.

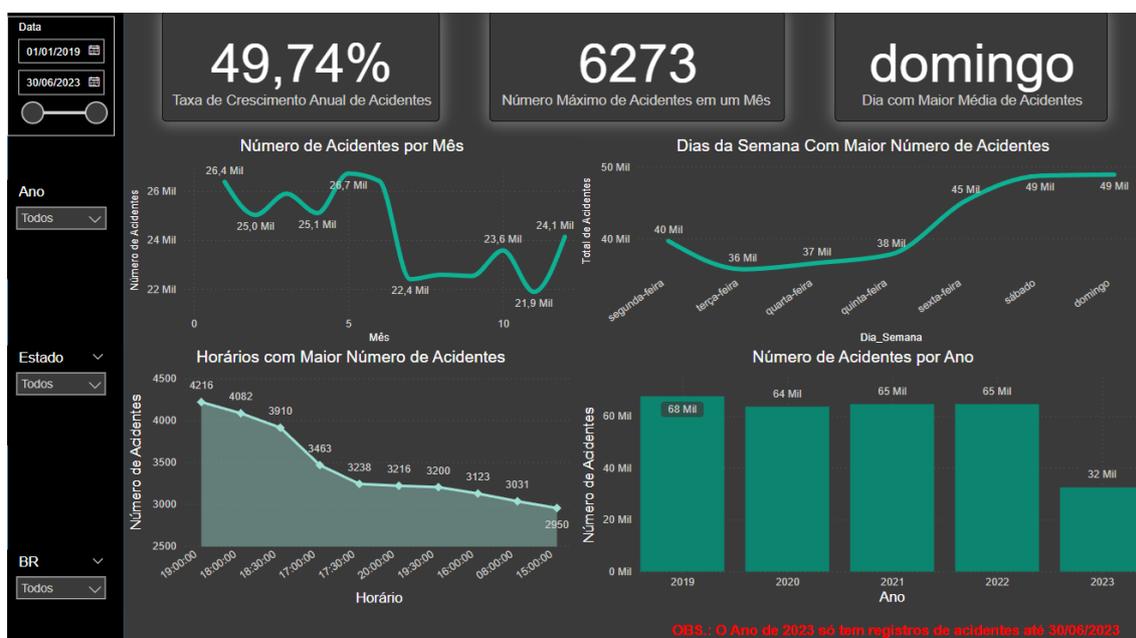


Figura 5. *Dashboard* BI-SafeRoad - Série temporal dos acidentes

Já a análise temporal dos acidentes identificou que o número máximo de acidentes em um mês foi 6.273 e que os domingos tiveram a maior média de acidentes. Os gráficos mostraram variação mensal de acidentes, dias da semana com mais ocorrências, horários de pico de acidentes após as 17 horas, e a evolução anual dos acidentes, destacando 2019 como o ano com maior incidência. Essas visualizações fornecem uma compreensão detalhada dos padrões de acidentes nas rodovias federais.

## 5 Conclusão

A ferramenta BI-SafeRoad permitiu responder de maneira eficaz às questões propostas nesta pesquisa. Através do uso do *Power BI*, desenvolvemos um *dashboard* in-

terativo que oferece uma visão abrangente e detalhada dos acidentes em rodovias federais apoiando a tomada de decisões e o desenvolvimento de estratégias de prevenção de acidentes. O BI-SafeRoad também permitiu correlacionar dados meteorológicos com os registros de acidentes, evidenciando como diferentes condições climáticas afetam a frequência e a gravidade dos acidentes.

Para trabalhos futuros recomendamos a inclusão de um novo KPI que considere o número de acidentes por quilômetro de rodovia, pois, as rodovias mais extensas tendem a apresentar um maior número absoluto de acidentes, e a análise por quilômetro pode fornecer uma compreensão mais precisa e coerente. Como o trabalho se limita ao período de 2019 a 2023, sugerimos a inclusão de outros anos na coleta de dados. Por fim, poderiam ser utilizadas outras ferramentas de BI, como ferramentas de mineração de dados.

Outro fator importante a ser considerado é o fluxo de veículos nas rodovias durante diferentes períodos. A intensidade do tráfego pode influenciar significativamente o número de acidentes, e a integração dessa variável nos futuros *dashboards* permitirá uma análise mais robusta e detalhada.

## Referências

- Detoni Junior, W., Oening, A. P., and Marcilio, D. C. (2019). Business intelligence: Uma revisão de literatura. [https://www.semanaacademica.com.br/system/files/artigos/business\\_intelligence\\_uma\\_revisao\\_de\\_literatura.pdf](https://www.semanaacademica.com.br/system/files/artigos/business_intelligence_uma_revisao_de_literatura.pdf). Acesso em: 28 ago. 2024.
- Ipea (2023). Estudo aponta aumento de 13,5% em mortes no trânsito. <https://www.ipea.gov.br/portal/categorias/45-todas-as-noticias/noticias/13899-estudo-aponta-aumento-de-13-5-em-mortes-no-transito?> Acesso em: 11 fev. 2024.
- Kimball, R. and Ross, M. (2011). *The data warehouse toolkit: the complete guide to dimensional modeling*. John Wiley & Sons.
- Microsoft (2024). Power bi desktop. Acesso em: 31 ago. 2024.
- Pinto, M. V. (2020). *Processo de Data Warehousing*. Marcus Vinicius Pinto.
- PostgreSQL (2024). PostgreSQL: License. <https://www.postgresql.org/about/licence/>. Acesso em: 29 ago. 2024.
- PRF (2024). Dados abertos da prf. <https://www.gov.br/prf/pt-br/aceso-a-informacao/dados-abertos/dados-abertos-da-prf>. Acesso em: 11 jul. 2024.
- Rocha, G. d. S. (2009). *Caracterização dos acidentes de trânsito e vítimas no município de Rio Branco-Acre*. PhD thesis, Universidade de São Paulo.