Uma imagem contendo comida

O conteúdo gerado por IA pode estar incorreto.

**UNIVERSIDADE FEDERAL DO CEARÁ**

**CENTRO DE CIÊNCIAS**

**DEPARTAMENTO DE ESTATÍSTICA E MATEMÁTICA APLICADA**

**CURSO DE CIÊNCIA DE DADOS**

**DISCIPLINA DE ESTATÍSTICA APLICADA**

**CICERO ROGÉRIO DOS SANTOS ALVES FILHO - 578748**

**ENZO THAUZER DE SOUZA LOURINHO - 581877**

**FRANCISCO RAYAN - 580684**

**MATHEUS ALVES CAVALCANTE - 579229**

**SINISTROS EM RODOVIAS FEDERAIS**

**ANÁLISE DE DADOS NO CEARÁ ENTRE OS ANOS DE 2017 E 2025**

**FORTALEZA**

**2025**

SUMÁRIO

[1. INTRODUÇÃO 4](#_Toc203583322)

[1.1 Contextualização da Segurança no Trânsito no Brasil 4](#_Toc203583323)

[1.2 A Realidade do Ceará no Cenário Rodoviário Nacional 4](#_Toc203583324)

[1.3 Importância da Análise de Dados em Segurança Viária 5](#_Toc203583325)

[1.4 Objetivos do Relatório 5](#_Toc203583326)

[1.5 Justificativa e Diferenciais do Estudo 6](#_Toc203583327)

[2. COLETA DE DADOS 6](#_Toc203583328)

[2.1 Fonte: Portal da PRF 6](#_Toc203583329)

[2.2 Problemas Iniciais Enfrentados 7](#_Toc203583330)

[2.3 Estratégias de Solução Técnica 7](#_Toc203583331)

[3. TRATAMENTO DE DADOS COM PYTHON 8](#_Toc203583332)

[3.1 Padronização de Colunas 8](#_Toc203583333)

[3.2 Correção das Coordenadas Geográficas 9](#_Toc203583334)

[3.3 Antes Correção das Coordenadas Geográficas 10](#_Toc203583335)

[3.4 Depois Correção das Coordenadas Geográficas 11](#_Toc203583336)

[3.5 Criação de Colunas Derivadas 12](#_Toc203583337)

[4. INTEGRAÇÃO E FUSÃO DE BASES 13](#_Toc203583338)

[4.1 Desafios na integração 13](#_Toc203583339)

[4.2 Processo de fusão 13](#_Toc203583340)

[5. ANÁLISE ESTATÍSTICA 14](#_Toc203583341)

[5.1 Evolução Anual dos Acidentes 14](#_Toc203583342)

[5.2 Distribuição Mensal 15](#_Toc203583343)

[5.3 Acidentes por Dia da Semana 16](#_Toc203583344)

[5.4 Classificação dos Tipos de Acidente 17](#_Toc203583345)

[5.5 Gravidade dos Acidentes 18](#_Toc203583346)

[5.6 Acidentes por Horário do Dia 18](#_Toc203583347)

[6. ANÁLISE GEOGRÁFICA 20](#_Toc203583348)

[6.1 Correção das Coordenadas Geográficas 20](#_Toc203583349)

[6.2 Visualização dos Acidentes em Mapa de Calor (Power BI) 21](#_Toc203583350)

[6.3 Classificação por BR e Município 22](#_Toc203583351)

[7. VISUALIZAÇÃO COM O POWER BI 24](#_Toc203583352)

[7.1 Deshboard I 24](#_Toc203583353)

[7.1 Deshboard II 25](#_Toc203583354)

[7.3 Exportação dos Dados Tratados 26](#_Toc203583355)

[7.4 Modelagem dos Dados no Power BI 26](#_Toc203583356)

[7.5 Análises Temporais: Ano e Mês 27](#_Toc203583357)

[7.6 Acidentes por Tipo e Classificação 27](#_Toc203583358)

[7.7 Análise Geográfica Interativa 27](#_Toc203583359)

[8. RESULTADOS E DISCURSÃO 28](#_Toc203583360)

[8.1 Tendência Temporal dos Acidentes 29](#_Toc203583361)

[8.2 Classificação dos Acidentes por Tipo e Gravidade 30](#_Toc203583364)

[9. CORRELAÇÕES ESTATÍSTICAS 31](#_Toc203583365)

[9.1 Metodologia da Análise de Correlação 31](#_Toc203583366)

[9.2 Principais Resultados 31](#_Toc203583367)

[10. ANÁLISE I 33](#_Toc203583368)

[10.1 Gráfico I 33](#_Toc203583369)

[10.2 Gráfico II 34](#_Toc203583370)

[10.1 Gráfico III 35](#_Toc203583371)

[10.4 Resumo Geral Por Ano 36](#_Toc203583372)

[11. ANÁLISE II 37](#_Toc203583373)

[11.1 Queda Gradual De 2017 a 2020 37](#_Toc203583374)

[11.2 Picos entre 2021 e 2024 38](#_Toc203583375)

[12. Links Das Informações Dos Possíveis Motivos Das Análises 39](#_Toc203583376)

[12.1 Queda Gradual De 2017 a 2020 39](#_Toc203583377)

[12.2 Picos entre 2021 e 2024 39](#_Toc203583378)

[13. CONCLUSÃO 40](#_Toc203583379)

# INTRODUÇÃO

## Contextualização da Segurança no Trânsito no Brasil

O Brasil é um dos países com os maiores índices de acidentes de trânsito no mundo, fato que tem gerado preocupações crescentes no âmbito social, econômico e de saúde pública. Segundo dados do Ministério da Saúde e do Departamento Nacional de Trânsito (Denatran), anualmente são registrados dezenas de milhares de acidentes, que resultam em milhares de vítimas fatais e em um número ainda maior de pessoas com sequelas e ferimentos graves.

As rodovias federais desempenham papel fundamental na integração territorial, escoamento de produtos, circulação de pessoas e no desenvolvimento econômico nacional. No entanto, também representam ambientes de elevado risco, onde as condições de infraestrutura, fiscalização, educação e comportamento humano interagem para influenciar diretamente a ocorrência e gravidade dos acidentes.

## A Realidade do Ceará no Cenário Rodoviário Nacional

O estado do Ceará possui uma extensa malha rodoviária federal, composta por importantes corredores de transporte como as BR-116, BR-020, BR-222 e BR-304, que cortam o território estadual em diferentes direções. Estas vias são vitais para a mobilidade urbana e rural, e para a conexão com outros estados e regiões do país.

Infelizmente, o Ceará tem figurado entre as unidades federativas com maiores índices de acidentes graves, o que reflete desafios estruturais, ambientais e sociais complexos. A alta densidade populacional, o crescimento acelerado da frota veicular, as condições climáticas específicas da região e aspectos culturais relacionados ao trânsito compõem um cenário multifacetado que demanda análise aprofundada para subsidiar ações eficazes.

## Importância da Análise de Dados em Segurança Viária

Nos últimos anos, o avanço tecnológico e a digitalização dos dados públicos abriram novas possibilidades para o estudo da segurança viária. A disponibilização de bases estruturadas, como as fornecidas pela Polícia Rodoviária Federal (PRF), permite a elaboração de análises quantitativas e geoespaciais robustas, que ultrapassam a simples tabulação e contabilização de acidentes.

O uso de técnicas de ciência de dados, programação e ferramentas de visualização auxilia na descoberta de padrões, identificação de áreas críticas e na compreensão dos fatores de risco. Esses insumos são fundamentais para a formulação de políticas públicas baseadas em evidências, visando a redução efetiva dos acidentes e seus impactos.

## 1.4 Objetivos do Relatório

Este relatório tem como objetivo principal apresentar uma análise detalhada dos acidentes de trânsito ocorridos nas rodovias federais do estado do Ceará no período compreendido entre 2017 e 2025. A partir dos dados disponibilizados pela PRF, o estudo busca:

* Coletar, integrar e tratar grandes volumes de dados heterogêneos de acidentes, vítimas e causas.
* Corrigir e validar as informações geográficas para análise espacial precisa.
* Identificar padrões temporais, espaciais e categóricos dos acidentes.
* Mapear trechos e municípios críticos para segurança viária.
* Apresentar resultados por meio de visualizações claras e interativas.
* Oferecer recomendações técnicas e estratégicas para gestores públicos, profissionais da área e sociedade civil.

## 1.5 Justificativa e Diferenciais do Estudo

A complexidade e a dispersão das bases de dados da PRF exigem um tratamento rigoroso e multidisciplinar, que combina conhecimentos em ciência de dados, estatística, geoprocessamento e engenharia de software. Este trabalho se diferencia por:

* Aplicar metodologia transparente e replicável, com uso de ferramentas open source.
* Realizar integração e limpeza manual e automatizada, garantindo a confiabilidade dos dados.
* Explorar tanto aspectos estatísticos quanto geográficos, permitindo análises integradas.
* Incorporar visualizações dinâmicas no Power BI, facilitando a exploração dos dados por diferentes públicos.
* Focar na aplicabilidade prática, alinhando análise técnica a recomendações concretas.

# 2. COLETA DE DADOS

## 2.1 Fonte: Portal da PRF

A base de dados utilizada neste trabalho foi obtida por meio do Portal de Dados Abertos da Polícia Rodoviária Federal (PRF), que disponibiliza anualmente arquivos estruturados referentes aos acidentes registrados em rodovias federais de todo o país. Esses dados fazem parte de uma iniciativa de transparência pública e estão disponíveis gratuitamente para uso acadêmico, técnico e institucional.

Os arquivos estão organizados por ano e subdivididos em diferentes categorias de informações, permitindo um detalhamento significativo das ocorrências. Para a realização desta análise, foram coletados dados referentes ao período de 2017 a 2025, totalizando mais de 40 arquivos distintos.

Os principais tipos de arquivos utilizados incluem:

* acidentes[ano].csv: informações básicas sobre o local, data, tipo e gravidade do acidente;
* datatran[ano].csv: dados sobre as vítimas envolvidas, veículos, idade, sexo e condições físicas;
* acidentes\_todas\_causas\_tipos[ano].csv: versão mais detalhada que inclui causas presumidas e fatores ambientais (clima, visibilidade, tipo de pista etc.).

A diversidade dos arquivos exigiu a criação de um esquema lógico de integração, permitindo que informações dispersas fossem cruzadas de forma confiável.

## 2.2 Problemas Iniciais Enfrentados

Durante a coleta e leitura dos arquivos, uma série de problemas estruturais foram identificados, exigindo estratégias específicas de tratamento e padronização para garantir a qualidade do banco de dados final. Abaixo, listam-se os principais desafios enfrentados:

Inconsistência de delimitadores

Alguns arquivos utilizavam vírgula como separador de colunas, enquanto outros usavam ponto e vírgula o que causava erro de leitura automatizada.

**Colunas com nomes diferentes**

Exemplo: em alguns anos, a coluna era nomeada como tipo\_acidente, enquanto em outros aparecia como tipo\_acid.

**Ausência de colunas específicas**

Alguns arquivos estavam incompletos, com colunas ausentes ou totalmente vazias em determinados anos, tornando necessário o preenchimento ou exclusão estratégica.

**Problemas de codificação**

Os arquivos vinham com diferentes padrões de codificação de caracteres (UTF-8, Latin-1), o que causava erro de leitura de acentos e caracteres especiais.

**Linhas quebradas e registros inválidos**

Havia linhas incompletas, com número de colunas divergente ou registros ilegíveis, o que exigiu o uso de filtros (on\_bad\_lines='skip') e verificação manual.

**Coordenadas geográficas corrompidas**

Em muitos registros, os campos de latitude e longitude estavam com valores fora do padrão decimal, como -37214399 ao invés de -3.7214399.

## 2.3 Estratégias de Solução Técnica

As soluções adotadas para os problemas identificados envolveram o uso de ferramentas da linguagem Python, com destaque para a biblioteca pandas. As principais ações técnicas incluem:

* Leitura com tolerância a erros: pd.read\_csv(..., on\_bad\_lines='skip');
* Inspeção e renomeação de colunas: .columns, .rename({...});
* Remoção de colunas irrelevantes: .dropna(axis=1, how='all');
* Padronização dos nomes para minúsculo e sem espaços: .str.lower(), .str.replace();
* Preenchimento de colunas ausentes com valores nulos: df['nova\_coluna'] = None;
* Conversão de codificação de arquivos: leitura com encoding='latin-1' ou utf-8-sig.

Para a correção das coordenadas geográficas, foi necessário dividir os valores por fatores de escala.

# 3. TRATAMENTO DE DADOS COM PYTHON

O tratamento dos dados foi uma etapa essencial para transformar os arquivos brutos da PRF em uma base limpa, padronizada e analisável. Por meio da linguagem Python, foram aplicadas técnicas de limpeza, fusão, derivação de colunas e correções geográficas. A biblioteca pandas foi a principal ferramenta utilizada, devido à sua eficiência na manipulação de grandes volumes de dados estruturados.

Os arquivos foram lidos com o comando pd.read\_csv, utilizando o parâmetro on\_bad\_lines='skip' para evitar interrupções por linhas inválidas. Em seguida, foram inspecionadas as colunas de cada ano para verificar nomes, tipos e estruturas divergentes, sendo criados dicionários de renomeação para unificação entre os diferentes arquivos.

## 3.1 Padronização de Colunas

Os arquivos de anos anteriores a 2021 apresentavam nomes de colunas diferentes entre si. Por exemplo, em alguns anos a coluna “tipo\_acidente” aparecia como “tipo\_acid”. Para resolver isso, foi criado um dicionário de renomeação, padronizando os nomes com base no modelo mais completo dos anos recentes.

Além disso, foram adicionadas colunas ausentes como vazias (None) nos arquivos que não as possuíam originalmente, garantindo uniformidade. Todos os nomes de colunas foram convertidos para letras minúsculas e sem espaços, usando funções como .str.lower() e .str.replace().

Essa padronização foi fundamental para permitir o empilhamento dos dados de diferentes anos em um único DataFrame, facilitando as análises agregadas e temporais.

## 3.2 Correção das Coordenadas Geográficas

As colunas de latitude e longitude estavam, em grande parte, com valores corrompidos ou em formatos inteiros sem ponto decimal. Por exemplo, valores como -37214399 indicavam erro de escala, dificultando a geolocalização precisa dos acidentes.

A solução foi aplicar divisões por potências de 10 (como 10⁵ ou 10⁶), transformando os valores em graus decimais reais. Posteriormente, foi aplicado um filtro geográfico baseado nos limites do estado do Ceará:

* Latitude válida: entre -5.2 e -2.5
* Longitude válida: entre -42.0 e -37.0

Com isso, as coordenadas inválidas ou fora do estado foram removidas. Apenas os registros com geolocalização compatível com o território cearense foram mantidos. As colunas latitude\_corrigida e longitude\_corrigida passaram a conter os valores ajustados, prontos para uso em mapas no Power BI.

## 3.3 Antes Correção das Coordenadas Geográficas

## Mapa O conteúdo gerado por IA pode estar incorreto.3.4 Depois Correção das Coordenadas Geográficas

## 

## 3.5 Criação de Colunas Derivadas

Para enriquecer a base e permitir análises mais segmentadas, foram criadas diversas colunas derivadas com agrupamentos e classificações técnicas. Isso permitiu visualizar tendências ocultas nos dados e gerar gráficos mais relevantes.

Entre as principais variáveis derivadas, destacam-se:

* hora\_grupo: agrupamento dos horários em faixas (madrugada, manhã, tarde, noite);
* acidentes\_dias\_uteis: ver o total de acidentes em dias úteis
* acidentes\_final\_de semana: ver totais e médias de acidentes no final de semana
* acidentes\_feriados: ver totais e médias de acidentes em feriados no Ceará

Gráfico, Gráfico de barras

O conteúdo gerado por IA pode estar incorreto.tornando a base mais flexível e visualmente rica nas ferramentas de análise, como o Power BI.

# 4. INTEGRAÇÃO E FUSÃO DE BASES

A base original de dados da PRF é fragmentada em múltiplos arquivos separados por categorias, o que exige um processo cuidadoso de integração para garantir a consistência e a completude da análise. As principais bases envolvidas são:

* Acidentes: arquivo principal contendo informações sobre data, local, tipo e gravidade do acidente.
* Datatran: dados sobre vítimas e veículos envolvidos em cada acidente.
* Acidentes\_todas\_causas: arquivos detalhados com as causas presumidas e fatores ambientais.

A integração dessas bases foi feita com base em colunas-chave comuns, como data, hora, rodovia federal (BR), quilômetro (km) e município, permitindo a união correta dos registros referentes a um mesmo acidente.

## 4.1 Desafios na integração

A principal dificuldade enfrentada foi a existência de chaves ambíguas e registros duplicados, que poderiam causar inconsistências. Para resolver isso, foi criado um identificador único para cada acidente, baseado na concatenação dos campos principais: data + hora + BR + km + município.

Além disso, foram aplicados filtros para remover registros inconsistentes e garantir que apenas acidentes ocorridos no estado do Ceará fossem mantidos.

## 4.2 Processo de fusão

Após a criação da chave única, as bases foram unidas utilizando comandos do pandas como merge() em modo “inner” para manter apenas registros presentes em todas as tabelas relevantes. Esse processo resultou em uma base integrada que permite análises multidimensionais, incluindo:

* Correlação entre tipo de acidente e causa presumida;
* Perfil das vítimas por tipo de acidente;
* Distribuição dos veículos envolvidos segundo a gravidade do sinistro.

# 5. ANÁLISE ESTATÍSTICA

A análise estatística dos acidentes rodoviários registrados nas rodovias federais do Ceará no período de 2017 a 2025 representa uma etapa crucial para compreender os padrões, tendências e fatores que influenciam a ocorrência e a gravidade desses eventos. Essa análise foi realizada a partir da base de dados tratada e integrada, utilizando técnicas descritivas e comparativas para identificar aspectos relevantes para a segurança viária.

O processamento estatístico foi conduzido com a linguagem Python, especialmente com as bibliotecas pandas para manipulação dos dados e matplotlib e seaborn para a geração dos gráficos. As visualizações finais foram implementadas no Power BI, aproveitando recursos interativos e geográficos para facilitar a interpretação dos resultados.

## 5.1 Evolução Anual dos Acidentes

Gráfico, Gráfico de barras

O conteúdo gerado por IA pode estar incorreto.

A análise temporal anual permite observar a dinâmica dos acidentes ao longo do tempo, identificando tendências e eventuais anomalias.

* Os dados indicam que o ano de 2019 apresentou um pico expressivo de acidentes, que pode estar associado ao aumento da frota veicular ou a condições específicas das rodovias.
* A queda acentuada registrada em 2022 e 2023 provavelmente reflete os impactos da pandemia de COVID-19, Redução de tráfego: Embora 2020 tenha sido o auge da pandemia, os reflexos econômicos e de mobilidade persistiram em 2022, muitas empresas ainda operavam com home office ou revezamento..
* A partir de 2021, observa-se uma retomada gradual, com crescimento até 2024, indicando um possível aumento da movimentação e redução das restrições.

## Gráfico, Gráfico de barras O conteúdo gerado por IA pode estar incorreto.5.2 Distribuição Mensal

Avaliar os acidentes por mês ajuda a identificar sazonalidades relacionadas a eventos climáticos, férias escolares e feriados prolongados.

* Os meses de janeiro e dezembro apresentam os maiores índices de acidentes, coincidindo com períodos de férias escolares e maior circulação nas rodovias devido a viagens.
* Os meses de abril, maio e junho apresentam tradicionalmente uma redução no número de acidentes nas rodovias federais cearenses. Essa queda pode ser atribuída à combinação de fatores climáticos (chuvas da quadra invernosa), redução da circulação em feriados e férias, menor volume de viagens longas e possíveis efeitos indiretos de campanhas de fiscalização realizadas nos meses anteriores.

## 5.3 Acidentes por Dia da Semana

Gráfico, Gráfico de barras

O conteúdo gerado por IA pode estar incorreto.

Analisar o volume de acidentes por dia da semana auxilia na programação de ações preventivas e policiamento.

* Sábado e domingo foram identificados como os dias com maior número de acidentes, possivelmente relacionados ao aumento no fluxo de veículos em deslocamentos para o fim de semana ou retorno de viagens.
* Terças, quartas e quintas-feiras apresentam os menores índices, dias típicos de menor circulação nas rodovias.

## 5.4 Classificação dos Tipos de Acidente

Gráfico, Gráfico de barras

O conteúdo gerado por IA pode estar incorreto. O tipo do acidente é uma variável fundamental para entender os mecanismos que mais causam sinistros e suas consequências.

* Colisão traseira, colisão frontal e colisão transversal são as categorias mais frequentes.
* Acidentes com engavetamento e quedas isoladas de motociclistas, embora menos frequentes, apresentam alta gravidade.

## 5.5 Gravidade dos Acidentes

Classificar os acidentes segundo a gravidade permite dimensionar os impactos sobre as vítimas e a necessidade de políticas específicas.

* Cerca de 5% dos acidentes resultaram em vítimas fatais, o que representa um número significativo em termos absolutos.
* A maior parte dos registros refere-se a acidentes com feridos leves ou sem vítimas, mas que ainda geram custos sociais e econômicos elevados.

## 5.6 Acidentes por Horário do Dia

Gráfico, Gráfico de linhas

O conteúdo gerado por IA pode estar incorreto.

* Os períodos da tarde (5h às 9h) e noite (16h às 18h) concentram o maior número de acidentes, o que pode estar relacionado ao aumento do fluxo tanto pela ida quanto pela volta do trabalho.

Gráfico, Gráfico de barras

O conteúdo gerado por IA pode estar incorreto.

* A madrugada apresenta menor frequência de acidentes, porém com maior taxa de gravidade, indicando riscos específicos que merecem atenção.

# 6. ANÁLISE GEOGRÁFICA

A análise geográfica é fundamental para compreender a distribuição espacial dos acidentes nas rodovias federais do Ceará. Ao mapear os pontos onde ocorrem os sinistros, é possível identificar áreas com maior concentração, conhecidas como “hotspots”, que demandam atenção especial das autoridades.

Para realizar essa análise, os dados passaram por um rigoroso processo de correção e validação das coordenadas geográficas, garantindo que as localizações representem fielmente os locais dos acidentes. Com as informações corretas, foram criados mapas de calor e visualizações interativas no Power BI, permitindo observar padrões e tendências espaciais que não seriam facilmente perceptíveis apenas pela análise numérica.

Essa abordagem facilita a priorização de ações e investimentos, como melhorias na infraestrutura, reforço na sinalização e alocação mais eficiente da fiscalização, focando nos trechos com maior risco.

## 6.1 Correção das Coordenadas Geográficas

A análise geoespacial dos acidentes exigiu um processo rigoroso de validação das coordenadas fornecidas pela Polícia Rodoviária Federal (PRF). Nos arquivos brutos, muitos registros apresentavam problemas como escalas incorretas, latitudes e longitudes invertidas, ou pontos fora dos limites geográficos do estado do Ceará.

Para resolver esses problemas, foi necessário:

* Converter os campos de latitude e longitude para ponto flutuante;
* Dividir os valores por 10510^5105 para normalização;
* Criar um filtro de validação geográfica com base na faixa esperada de coordenadas do estado (latitudes entre -2.5 e -7; longitudes entre -37 e -43);
* Descartar registros inválidos ou nulos.

Esse processo de saneamento foi essencial para garantir a confiabilidade da etapa seguinte de visualização no mapa.

## 6.2 Visualização dos Acidentes em Mapa de Calor (Power BI)

Mapa

O conteúdo gerado por IA pode estar incorreto. Com as coordenadas corrigidas, os dados foram carregados no Power BI para geração de mapas interativos. A principal visualização aplicada foi o **mapa de calor (heatmap)**, que permite observar com clareza as áreas com maior concentração de acidentes.

**Configurações técnicas utilizadas:**

* Latitude e longitude: campos corrigidos;
* Valor: contagem de registros (1 por linha);
* Tipo de mapa: calor, com ajuste de intensidade para realçar áreas críticas.

**Principais observações:**

* Forte concentração de acidentes nos trechos urbanos da BR-116 e BR-222;
* Hotspots destacados nas regiões metropolitanas de Fortaleza, Sobral, Juazeiro do Norte e Caucaia;
* Regiões do Sertão Central com menor densidade, mas ocorrência frequente.

## 6.3 Classificação por BR e Município

Além do mapa de calor, foi realizada a análise da distribuição de acidentes por rodovia (BR) e município. Para isso, utilizou-se tanto mapasdepontos quanto ShapeMaps, onde a escala visual corresponde ao número de ocorrências em cada localidade

Gráfico, Gráfico de barras, Histograma

O conteúdo gerado por IA pode estar incorreto.

**Destaques da análise:**

* A BR-116 apresentou a maior concentração de acidentes, especialmente entre Fortaleza e Pacajus;
* A BR-222 concentrou grande volume de ocorrências no trecho entre Caucaia e Sobral;
* A BR-020, embora com menor número total, registrou alta proporção de acidentes graves, especialmente entre Canindé e Tauá;
* Municípios como Fortaleza, Caucaia, Maracanaú, Sobral e Juazeiro do Norte lideraram o ranking de ocorrências.

Essas informações permitiram a identificação de **zonas críticas de atenção**, úteis para orientar decisões de engenharia de tráfego, fiscalização e campanhas de prevenção.

# 7. VISUALIZAÇÃO COM O POWER BI

## 7.1 Deshboard I

Interface gráfica do usuário, Site

O conteúdo gerado por IA pode estar incorreto.Inicialmente identificamos alguns pontos de acidentes que, devido a imprecisões nas coordenadas, apareciam no mar. Ajustamos esses pontos para suas localizações corretas no Ceará, resultando em uma visualização mais precisa.

## 7.1 Deshboard II

Interface gráfica do usuário

O conteúdo gerado por IA pode estar incorreto.Refinamos ainda mais os dados, excluindo definitivamente os pontos que estavam fora do território, após o processo de correção e limpeza dos dados, o número total de acidentes foi ajustado de 13.021 para 12.975 registros, uma redução de 46 ocorrências decorrente da remoção de entradas com coordenadas inconsistentes ou fora do território do Ceará

A etapa de visualização dos dados foi realizada por meio do Power BI, uma ferramenta robusta de business intelligence que permite transformar grandes volumes de dados em painéis interativos e gráficos dinâmicos. Com a base de dados já limpa, estruturada e enriquecida, o Power BI foi essencial para explorar os padrões de acidentes de forma visual, facilitando a interpretação dos resultados por diferentes públicos, como pesquisadores, gestores públicos e analistas de tráfego.

A seguir, são descritas as principais visualizações desenvolvidas, incluindo mapas geográficos, gráficos temporais, segmentações por BR, município, tipo de acidente, faixa horária e outros filtros relevantes.

## 7.3 Exportação dos Dados Tratados

Concluído o processo de padronização e limpeza dos dados no ambiente Python, a base final foi exportada em formato .csv, contendo colunas corrigidas e derivadas. Entre essas colunas estavam hora\_grupo, veículo, idade\_faixa, causa\_acidente, latitude\_corrigida e longitude\_corrigida. Esse arquivo consolidado tornou-se a base principal para a visualização dos dados no Power BI.

A escolha do Power BI como ferramenta de visualização se deu pela sua capacidade de lidar com grandes volumes de dados, integração nativa com arquivos .csv e funcionalidades avançadas para criação de gráficos, filtros e mapas interativos. Além disso, o Power BI permite uma experiência interativa para o usuário final, essencial para relatórios com múltiplas camadas analíticas.

## 7.4 Modelagem dos Dados no Power BI

Após a importação dos dados, iniciou-se o processo de modelagem. Como o arquivo já estava unificado e padronizado, os relacionamentos entre colunas eram diretos. Ainda assim, foi necessário verificar possíveis erros de tipo, como campos de texto lidos como numéricos ou colunas de data sem reconhecimento automático.

Além disso, foram configuradas as tabelas de dimensão e as medidas (KPIs) para facilitar a criação de gráficos dinâmicos. Também foram aplicadas segmentações por ano, mês, BR, município, tipo de veículo, classificação do acidente e horário do dia.

**Sugestões de visualizações técnicas:**

* Segmentadores por ano e município;
* Campos calculados com DAX para totalizadores.

## 7.5 Análises Temporais: Ano e Mês

A primeira camada de análise explorou o comportamento temporal dos acidentes ao longo do período de 2017 a 2025. Para isso, foram utilizados gráficos de linha e colunas que mostram a evolução ano a ano e mês a mês.

Esses gráficos são úteis para identificar sazonalidades, impactos de políticas públicas, variações decorrentes da pandemia (como em 2020), e tendências gerais de aumento ou redução nos acidentes.

## 7.6 Acidentes por Tipo e Classificação

Utilizando os campos tipo\_acidente e classificacao\_acidente, foi possível criar visualizações comparando o tipo de ocorrência (colisão frontal, saída de pista, atropelamento, entre outros) com a gravidade do desfecho (sem vítima, com feridos, com vítima fatal).

A análise revelou que colisões traseiras e frontais são as mais frequentes, enquanto atropelamentos e tombamentos estão entre os mais graves. A separação por tipo de acidente também auxilia na identificação de causas recorrentes e na proposta de intervenções específicas.

## 7.7 Análise Geográfica Interativa

Com os campos latitude\_corrigida e longitude\_corrigida, a base passou a ser compatível com os recursos de mapa do Power BI. O primeiro mapa utilizado foi o **mapa de calor**, que permite observar a densidade de acidentes ao longo do território cearense.

Os dashboards foram projetados para serem exploráveis por diferentes públicos. A utilização de segmentadores e filtros cruzados permitiu que qualquer gráfico fosse recortado por múltiplas variáveis: ano, mês, BR, município, tipo de veículo, horário, dia da semana, faixa etária das vítimas, e gravidade do acidente.

Essa abordagem garante que a mesma base possa servir tanto a análises técnicas quanto a apresentações para gestores públicos ou pesquisadores acadêmicos.

# 8. RESULTADOS E DISCURSÃO

Nesta seção, são apresentados os principais resultados obtidos a partir da análise dos dados tratados, com foco na identificação de padrões, tendências e características dos acidentes ocorridos nas rodovias federais do Ceará entre 2017 e 2025. A interpretação desses resultados permite uma compreensão aprofundada do cenário, destacando pontos críticos e aspectos que demandam atenção para ações de prevenção e mitigação.

Os dados serão discutidos em relação à sua evolução temporal, distribuição espacial, características dos acidentes e impactos humanos, sempre com o objetivo de subsidiar recomendações estratégicas para a segurança viária no estado.

## 8.1 Tendência Temporal dos Acidentes

Tela de computador com texto preto sobre fundo branco

O conteúdo gerado por IA pode estar incorreto.

A análise da evolução temporal dos acidentes nas rodovias federais do Ceará mostra variações significativas ao longo dos anos entre 2017 e 2025. Algumas oscilações podem estar associadas a fatores externos, como mudanças no fluxo de veículos, condições climáticas e políticas públicas aplicadas.

Essa análise permite identificar períodos de maior risco e avaliar o impacto de intervenções feitas em determinadas épocas, servindo como base para planejamento de ações futuras.

## 8.2 Classificação dos Acidentes por Tipo e Gravidade

A categorização dos acidentes segundo tipo e gravidade permite compreender quais situações geram maior risco e quais são mais recorrentes. Essa análise orienta a priorização de medidas de segurança e políticas públicas específicas para os tipos mais críticos.

O mapeamento dos acidentes por rodovia federal e município evidencia pontos com maior concentração de sinistros. Essa identificação é crucial para intervenções locais em infraestrutura, sinalização e fiscalização, além de facilitar a alocação eficiente de recursos.

A análise dos impactos humanos destaca a importância do tema para a saúde pública e segurança da população. Além do número de acidentes, o estudo considera o perfil das vítimas, a gravidade das ocorrências e as consequências para os sistemas de saúde e economia local.

Compreender esses impactos fortalece a argumentação para investimentos em prevenção e políticas integradas de segurança viária.

# 9. CORRELAÇÕES ESTATÍSTICAS

A análise de correlações é uma etapa fundamental no processo de compreensão de padrões complexos entre variáveis de um conjunto de dados. No contexto dos acidentes rodoviários no estado do Ceará entre 2017 e 2025, a investigação de possíveis associações entre variáveis como tipo de acidente, horário, condição climática, gravidade, entre outras, permite identificar relações ocultas que não emergem diretamente nas análises descritivas.

O objetivo deste ponto é avaliar a força e direção das correlações entre variáveis selecionadas para apoiar tomadas de decisão baseadas em evidências. A análise foi conduzida utilizando Python, com apoio das bibliotecas pandas, numpy, seaborn e matplotlib.

## 9.1 Metodologia da Análise de Correlação

A base de dados tratada foi segmentada em colunas numéricas e categóricas que foram previamente codificadas para possibilitar a análise quantitativa.

Variáveis analisadas:

* Tipo de acidente (colisão, atropelamento, saída de pista etc.)
* Classificação de gravidade (sem vítima, com feridos, fatal)
* Condição meteorológica
* Fase do dia (manhã, tarde, noite, madrugada)
* Tipo de veículo envolvido
* Dia da semana
* Hora do acidente
* Presença de vítimas fatais

## 9.2 Principais Resultados

**Correlação entre tipo de acidente e fase do dia:**

* Colisões traseiras ocorrem com maior frequência durante o dia, especialmente em horários de tráfego intenso.
* Saídas de pista e capotamentos têm maior incidência em horários noturnos e de madrugada, sugerindo associação com sonolência, visibilidade reduzida ou consumo de álcool.

**Correlação entre condição climática e gravidade:**

* Em dias chuvosos ou com baixa visibilidade, há um aumento proporcional nos acidentes com feridos graves.

**Correlação entre horário e presença de vítimas fatais:**

* Acidentes ocorridos entre 22h e 3h têm maior chance de resultar em óbitos, mesmo com menor frequência absoluta.
* Esse dado indica que o risco por acidente é maior durante a madrugada, o que pode orientar estratégias de fiscalização noturna.

**Correlação entre dia da semana e tipo de acidente:**

* Fins de semana estão mais associados a acidentes com motocicletas e atropelamentos.
* Segunda a quinta concentram colisões em tráfego de fluxo intenso e acidentes envolvendo veículos de carga.

# 10. ANÁLISE I

## 10.1 Gráfico I

## 

Gráfico, Gráfico de linhas

O conteúdo gerado por IA pode estar incorreto.

**Evolução dos Acidentes por Tipo de Acidente (2017–2025)**

**2017 a 2020**

* **Colisão traseira**, **transversal** e **lateral** lideram em quantidade.
* Há queda contínua nesses tipos principais entre 2017 e 2020.
* **Colisão lateral** sofre queda brusca em 2021.
* **Tombamento**, **frontal** e **saída de leito** variam pouco, mantendo-se na faixa de 100–180.

**2021 a 2024**

* **Colisão lateral mesmo sentido** surge fortemente a partir de 2021, ultrapassando 150 casos.
* **Colisão traseira** volta a subir até 2024, chegando a 360 acidentes.
* A maioria dos tipos mantém tendência estável ou levemente crescente em 2023 e 2024.

**2025**

* **Redução nos registros**

## 10.2 Gráfico II

Gráfico, Gráfico de linhas

O conteúdo gerado por IA pode estar incorreto.

**Evolução dos Acidentes nas 10 BRs com Mais Acidentes**

**2017 a 2020**

* BRs com maior número de acidentes:
  + **BR-116** e **BR-222**.
  + **BR-020** também se destaca.
* Há uma tendência de **redução gradual** nas principais BRs.

**2021 a 2024**

* **BR-116** volta a crescer até bater o pico de **800 em 2024**.
* **BR-222** e **BR-020** seguem queda contínua.
* BRs secundárias (BR-304, BR-230, etc.) permanecem estáveis com baixos valores.

**2025**

* **Redução nos registros**

## 10.1 Gráfico III

Gráfico, Gráfico de linhas

O conteúdo gerado por IA pode estar incorreto.

**Evolução dos Acidentes por Dia da Semana**

**2017 a 2020**

* **Domingo** e **sábado** lideram com mais acidentes.
* Durante a semana, os números são mais equilibrados, geralmente entre 220 e 270.

**2021 a 2024**

* Diferenças entre dias diminuem, ficando todos na faixa de 200–250.
* **Sexta e sábado** ainda mantêm posição de destaque.

**2025**

* **Redução nos registros**

## 10.4 Resumo Geral Por Ano

|  |  |
| --- | --- |
| **Ano** | **Observações principais** |
| **2017** | Onde ocorre uma alta incidência geral;  Os acidentes por colisões traseira, transversal e lateral dominam;  A BR-116 é a mais crítica;  Nos dias de domingos e sábados são onde têm mais acidentes. |
| **2018 a 2020** | Acontece uma queda gradual em todos os tipos de acidentes, BRs e dias. |
| **2021** | Houve um pico de acidentes por "colisão lateral mesmo sentido";  Uma volta crescente da BR-116;  Ocorre uma redistribuição mais uniforme nos dias da semana. |
| **2022 a 2024** | Há um reaquecimento dos números de acidentes em colisões traseiras;  Enquanto o número de acidentes nas BRs apresenta queda, a BR-116 chama atenção ao registrar seu pico de ocorrências justamente nesse período;  Volta de altos índices nos fins de semana. |
| **2025** | Redução nos registros. |

# 11. ANÁLISE II

## 11.1 Queda Gradual De 2017 a 2020

## 

**Possíveis Motivos:**

* **Aprimoramento da fiscalização e policiamento nas rodovias federais (PRF)**
* **Modernização de frota e tecnologias de segurança veicular** (ex: controle de estabilidade, frenagem automática).
* **Integração de sistemas de emergência e atendimento móvel (SAMU, Corpo de Bombeiros, PRF)**.
* **Melhorias na infraestrutura em algumas BRs prioritárias (BR-222, BR-020)**.

Em **2017**, a PRF aplicou mais de **6 milhões de infrações**, com destaque para excesso de velocidade, resultado do aumento da fiscalização nas estradas federais.

A **Operação Rodoviária 2018** intensificou o uso de radares e ações educativas, resultando em **521 877** flagrantes por velocidade e mais de **390 mil ações educativas**, com redução de 29% nos acidentes graves e 16% nas mortes

Contudo, a retirada dos radares móveis em agosto de **2019** levou ao aumento de mortes e feridos entre agosto e outubro daquele ano, reforçando o impacto positivo da fiscalização eletrônica

O declínio de acidentes no período 2017–2020 foi plausivelmente impulsionado por essas ações integradas de fiscalização e educação rodoviária.

## 11.2 Picos entre 2021 e 2024

**Possíveis Motivos:**

* Retomada intensa do tráfego após pandemia de COVID-19 (2020–2021):
  + Em 2020, houve restrição de circulação e redução forçada de deslocamentos.
  + Já em 2021 e principalmente 2022, ocorreu uma explosão de tráfego reprimido, o que aumentou os riscos.
* Saturação das rodovias sem contrapartida de ampliação da malha ou controle eficiente.
* Crescimento no uso de transporte rodoviário de cargas e ônibus após queda na malha ferroviária e aérea.
* Colisão lateral mesmo sentido pode ter sido introduzida como nova categoria no banco de dados, explicando sua "súbita aparição" em 2021.

Em maio de 2021, a PRF liderou a campanha Maio Amarelo, com foco em respeito e responsabilidade no trânsito, mobilizando redes sociais e operações educativas nas rodovias

No mesmo ano, houve aumento significativo nas blitz de álcool: mais de 22 mil recusas ao bafômetro e 11 512 motoristas autuados por embriaguez nas rodovias federais

Essa intensificação de fiscalização por embriaguez e campanhas nacionais contribuiu para diminuir comportamentos de risco, especialmente fins de semana, ainda que tenha surgido “colisão lateral mesmo sentido” possivelmente por mudança de registro.

# 12. Links Das Informações Dos Possíveis Motivos Das Análises

## 12.1 Queda Gradual De 2017 a 2020

[Infrações em 2017: maioria foi por excesso de velocidade](https://autopapo.com.br/noticia/infracoes-em-2017-excesso-de-velocidade/?utm_source=chatgpt.com)

[Operação Rodovida 2018 focará na redução de violência no trânsito — Ministério da Justiça e Segurança Pública](https://www.gov.br/mj/pt-br/assuntos/noticias/operacao-rodovida-2018-focara-na-reducao-de-violencia-no-transito?utm_source=chatgpt.com)

[Após retirada de radares, cresce nº de mortos e feridos nas rodovias federais](https://www.terra.com.br/noticias/brasil/cidades/apos-retirada-de-radares-cresce-n-de-mortos-nas-rodovias%2Cb5339db7f891678ffb0ad8c3c9e24fc4saqa6vod.html?utm_source=chatgpt.com)

## 12.2 Picos entre 2021 e 2024

[PRF inicia campanha Maio Amarelo — Polícia Rodoviária Federal](https://www.gov.br/prf/pt-br/noticias/estaduais/para/anteriores/maio/prf-inicia-campanha-maio-amarelo?utm_source=chatgpt.com)

[Bafômetro: mais de 22 mil brasileiros se recusaram a fazer o teste em 2021 | CNN Brasil](https://www.cnnbrasil.com.br/nacional/bafometro-mais-de-22-mil-brasileiros-se-recusaram-a-fazer-o-teste-em-2021/?utm_source=chatgpt.com)

# 13. CONCLUSÃO

Este relatório consolidou uma análise abrangente e rigorosa dos acidentes ocorridos nas rodovias federais do Ceará no período entre 2017 e 2025, com base em uma base de dados oficial minuciosamente tratada, padronizada e integrada. A investigação revelou padrões importantes, que combinam aspectos temporais, espaciais, tipológicos e de gravidade, permitindo um entendimento mais completo das dinâmicas que envolvem os sinistros rodoviários na região.

A partir das análises, constatou-se que fatores multifacetados, incluindo a condição da infraestrutura viária, comportamento dos condutores, condições climáticas e a distribuição do tráfego, atuam de forma interligada para aumentar os riscos em determinados trechos e horários. A concentração de acidentes em rodovias como a BR-116 e BR-222, assim como em municípios estratégicos como Fortaleza, Sobral e Caucaia, evidencia a existência de áreas de concentração de acidentes que demandam ações prioritárias e focalizadas.

Além disso, a identificação de períodos sazonais e horários críticos aponta para a necessidade de estratégias dinâmicas, que levem em conta a variação temporal do risco, como campanhas educativas intensificadas em meses e dias específicos, bem como reforço da fiscalização em horários de maior incidência.

O impacto humano registrado, com expressivo número de vítimas fatais e feridos, ressalta não apenas o custo social das ocorrências, mas também a importância do tema para a saúde pública, economia e qualidade de vida das populações afetadas. A análise reforça a urgência de políticas públicas integradas, que envolvam investimento em infraestrutura, fiscalização inteligente, educação no trânsito e aprimoramento dos sistemas de coleta e monitoramento de dados.

A robustez metodológica aplicada, utilizando ferramentas avançadas de ciência de dados e geoprocessamento, oferece um modelo replicável e expansível para futuras análises e atualizações, contribuindo para o aperfeiçoamento contínuo das estratégias de segurança viária.

Por fim, conclui-se que, embora avanços possam ter ocorrido, o desafio da redução dos acidentes nas rodovias federais do Ceará permanece significativo, exigindo comprometimento institucional, envolvimento da sociedade civil e aplicação constante de soluções baseadas em evidências para promover um trânsito mais seguro e sustentável.