



UNIVERSIDADE
FEDERAL DO CEARÁ
CAMPUS DE CRATEÚS

RELATÓRIO ANÁLISE EXPLORATÓRIA DE DADOS E VISUALIZAÇÕES

DISCIPLINA: CIÊNCIA DOS DADOS
DOCENTE: RENAN GOMES VIEIRA

CARLOS EDUARDO RODRIGUES PITA – 509630

FRANCISCO RIAN RIBEIRO MEDEIRO – 552915

MARLON MELO MOURA – 55324

Crateús/CE
13/11/2025

Este relatório apresenta um resumo geral do que foi realizado na atividade de análise exploratória de dados e visualizações, abordando os principais aspectos, métodos utilizados e resultados obtidos.

ÍNDICE

1. Objetivo
2. Dados Utilizados
3. Resumo do Processamento
4. Principais Achados e Visualizações
5. Limitações
6. Recomendações
7. Principais Desafios e Aprendizados

1. OBJETIVOS

Este trabalho tem como objetivo central desenvolver as habilidades de Análise Exploratória de Dados (EDA) e Visualização de Dados. A partir de um conjunto de dados públicos, foi realizada uma investigação para caracterizar o dataset, processá-lo e, por fim, formular e responder perguntas de interesse sobre os dados. O foco foi extrair informações através de análises estatísticas e, principalmente, de visualizações, buscando identificar padrões que possam ter relevância social ou de segurança pública.

2. DADOS UTILIZADOS

A análise foi realizada sobre o conjunto de dados de Acidentes Automobilísticos de 2025, disponibilizado pela Polícia Rodoviária Federal (PRF).

- Fonte: [Portal de Dados Abertos da PRF \(Gov.br\)](#).
- Conteúdo: O dataset registra as ocorrências de acidentes em rodovias federais brasileiras.
- Dimensões: O conjunto de dados original possui 47.192 instâncias (acidentes) e 30 atributos (colunas).
- Período: Os dados abrangem o período de 01 de janeiro de 2025 a 31 de agosto de 2025.
- Considerações Éticas: Sendo dados públicos sob o Plano de Dados Abertos da PRF, eles seguem a Lei de Acesso à Informação (LAI) e a Política de Dados Abertos do Executivo Federal, não apresentando questões de privacidade que impeçam a análise.

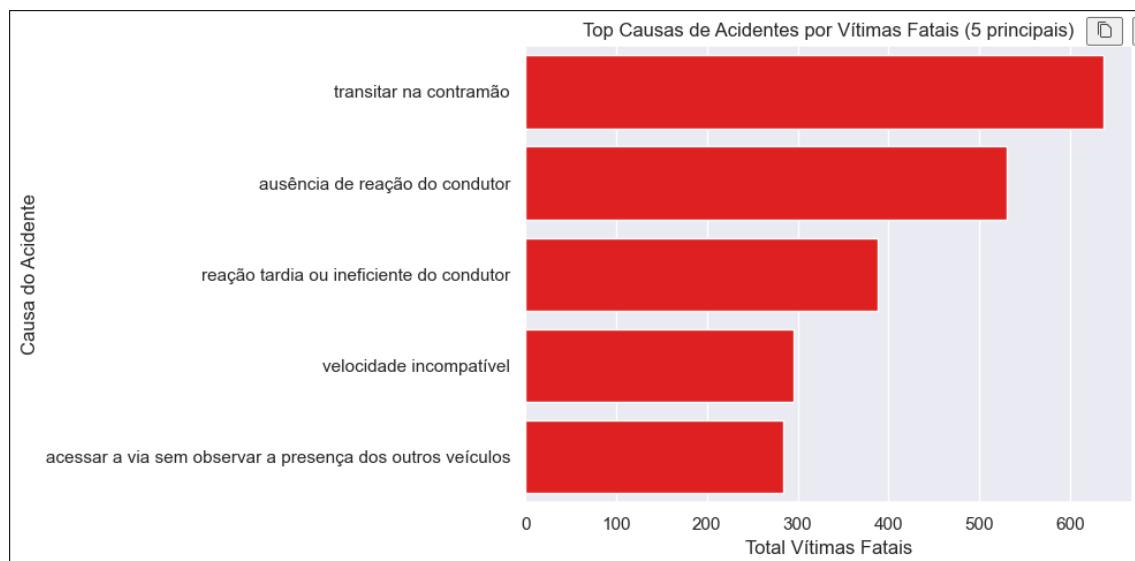
3. RESUMO DO PROCESSAMENTO

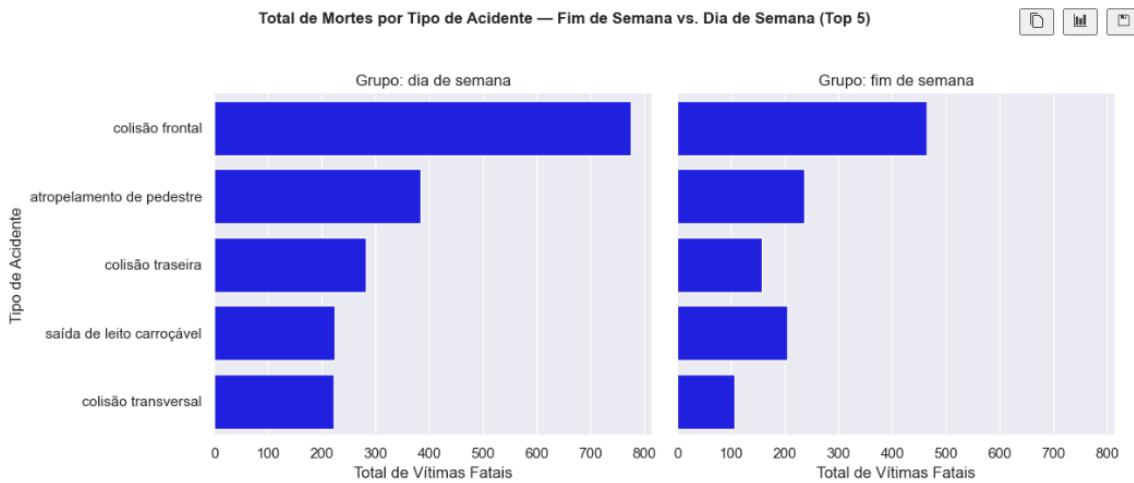
Antes da visualização, foi executada uma etapa robusta de pré-processamento para garantir a qualidade dos dados. As principais ações incluíram o carregamento e configuração com a importação das bibliotecas (pandas, seaborn) e carregamento do arquivo datatran2024.csv, houve também o tratamento de dados faltantes onde verificou-se que colunas como 'classificacao_acidente', 'regional', 'delegacia' e 'uop' possuíam valores nulos, sendo assim aplicada a função 'impute_missing_categorical' para preencher as lacunas com dados faltantes. Ademais na conversão de tipo, a função 'convert_data_types' foi responsável por corrigir os tipos de dados de colunas que eram lidas como texto (objeto) como 'data_inversa' (para o datetime) e colunas numéricas, novas colunas também foram criadas através da função 'Add_new_columns' para auxiliar nas análises e por fim o conteúdo textual de diversas colunas foi padronizado usando 'standardize_text_content' para unificar categorias (por exemplo, "Sem Vítimas vs. sem vítimas).

4. PRINCIPAIS ACHADOS E VISUALIZAÇÕES

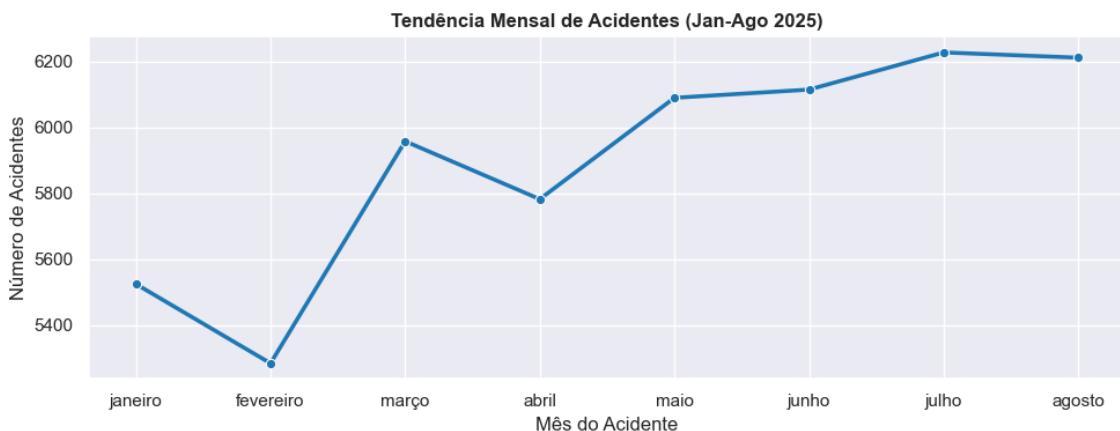
A análise exploratória buscou responder perguntas chave sobre a natureza dos acidentes.

A análise das principais causas de acidentes (Top Causas de Acidentes por Vítimas Fatais (5 principais)) mostra que "transitar na contramão" e "ausência de reação do condutor" são as mais frequentes. Consequentemente, ao cruzar as causas com o tipo de acidente (Total de Mortes por Tipo de Acidente – Fim de Semana vs. Dia da Semana (Top 5)), o gráfico revela que a "Colisão frontal" é, de longe, a causa mais fatal, superando significativamente as demais, o que revela uma relação semântica entre esses gráficos.

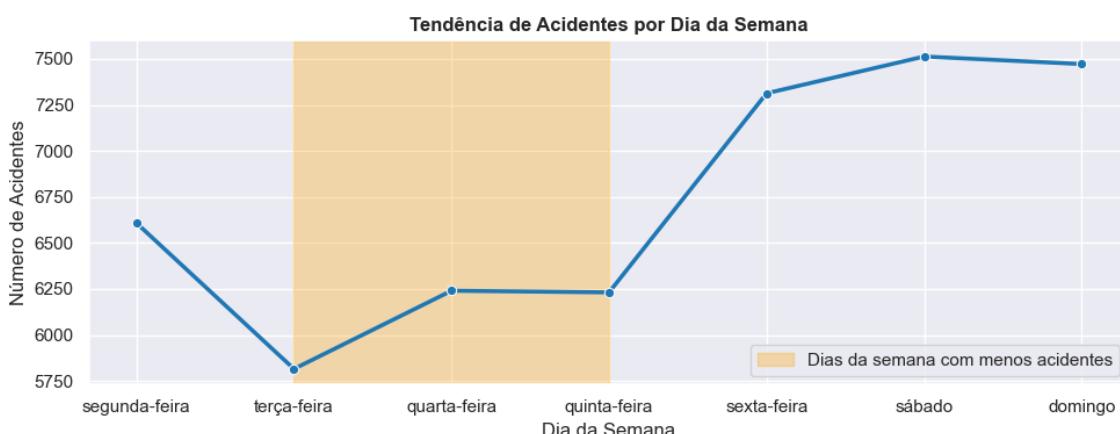




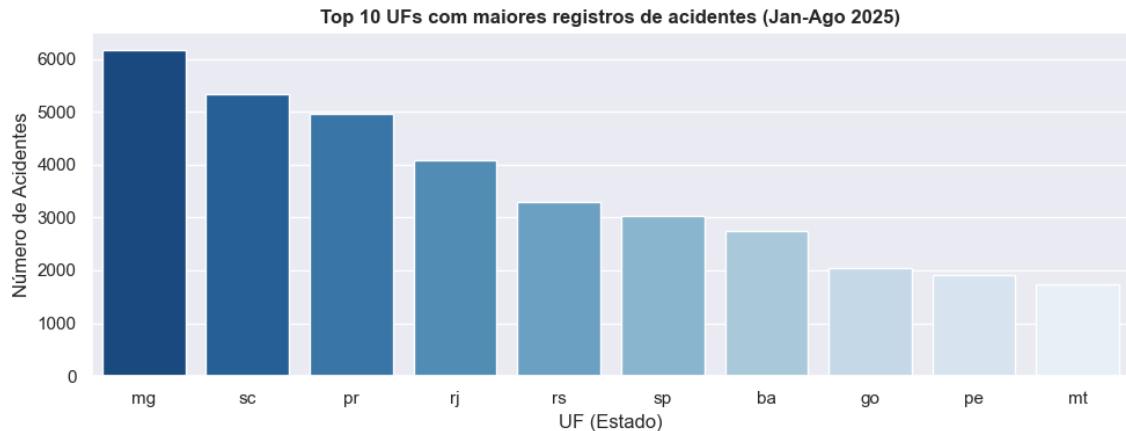
Outro insight que podemos ter é uma análise temporal que revelou dois padrões claros, existindo uma tendência mensal, onde o número de acidentes apresentou uma tendência de crescimento ao longo do ano. O ponto mais baixo ocorreu em fevereiro, seguido por um aumento quase contínuo que atingiu o pico em julho (Tendência Mensal de Acidentes (Jan-Ago 2025)).



Junto a tendência mensal, temos também uma tendêncial semanal, mostrando que ocorrem significativamente mais acidentes nos fins de semana, com os picos de ocorrências concentrados no sábado, domingo e sexta-feira. A terça-feira, em contrapartida, é o dia com menor número de registros.



O gráfico 'Top 10 UFs com maiores registros de acidentes (Jan-Ago 2025)' mostra que a distribuição geográfica dos acidentes não é homogênea. O gráfico abaixo mostra que três estados concentram a maior parte das ocorrências no período: Minas Gerais (MG), Santa Catarina (SC) e Paraná (PR). É perceptível que a região Sul chama bastante atenção por conta de seus 3 estados estarem no top 5 do gráfico.



5. LIMITAÇÕES

Apesar dos insights gerados, a análise possui limitações que devem ser consideradas como o período incompleto, pois o dataset cobre apenas os primeiros 8 meses de 2025 (Jan-Ago). Isso impede a análise de tendências de fim de ano (feriados de Natal e Ano Novo), que historicamente impactam a accidentalidade. Junto a isso também tem a possibilidade de acidentes em divisas de dividem estados, poucos dados que podem contribuir para algum tipo de erro quando feita uma análise mais localizada.

6. RECOMENDAÇÕES

Com base nos achados e dados que foram cruzados, recomendamos o foco em ultrapassagem (Contramão), dado que "transitar na contramão" é a causa mais letal, recomenda-se um aumento na fiscalização em rodovias de pista simples e em trechos com ultrapassagem proibida, que frequentemente resultam em "colisões frontais". Reforço da fiscalização no fim de semana, pois o aumento drástico de acidentes a partir de sexta-feira sugere a necessidade de reforço no patrulhamento e campanhas de direção defensiva especificamente voltadas para o tráfego de fim de semana. E por fim alocação de recursos maior para os estados no topo (MG, SC, PR), os dados indicam que os estados de MG, SC e PR devem ser prioritários na alocação de recursos da PRF para fiscalização e ações preventivas, visando reduzir o número absoluto de ocorrências.

7. PRINCIPAIS DESAFIOS E APRENDIZADOS

Acreditamos que os principais desafios dizem respeito exatamente ao que aprendemos, o trabalho permitiu aplicar na prática o fluxo completo de análise de dados. Foi um exercício com um ganho de conhecimento de limpeza e transformação de dados brutos usando pandas, e de como traduzir dados limpos em *insights* visuais tentando ser claro e direto usando seaborn e matplotlib, permitindo responder perguntas complexas de forma objetiva demonstrando os dados plotados. Junto a isso houve também o aprendizado na visualizações dessas informações através de gráficos que facilitaram e orientaram o entendimento dos dados dos quais a gente estava utilizando.

[Link Repositório Github](#)