



UNIVERSIDADE FEDERAL DA PARAÍBA  
CENTRO DE CIÊNCIAS SOCIAIS APLICADAS  
COORDENAÇÃO DE CIÊNCIA DE DADOS PARA NEGÓCIOS

# **Dos Dados ao Alerta: Utilizando Machine Learning para Identificar e Prevenir Riscos em Trechos Críticos das Rodovias Nordestinas**

DISCENTE: ELAINE REGINA REIS SOUSA  
ORIENTADOR: PROFESSOR DR. JORGE HENRIQUE NORÕES VIANA



# Resumo do Projeto

O projeto visa o desenvolvimento de uma aplicação inteligente para identificação e previsão de trechos críticos nas rodovias federais do Nordeste, considerando fatores históricos, estruturais e climáticos.

O resultado esperado é uma interface interativa georreferenciada, que possibilite:

- Identificar os trechos com maior risco de acidentes graves;
- Simular cenários de risco em rodovias planejadas ou em construção;
- Subsidiar decisões preventivas no planejamento viário e políticas públicas de segurança.

# Justificativa – A Urgência do Problema no Nordeste

- **Impacto Social e Econômico:**

- No Brasil, estima-se cerca de **40 mil mortes por ano** devido a acidentes de trânsito na última década (IPEA, 2025).
- Os custos relacionados a acidentes em rodovias brasileiras somam aproximadamente **R\$40 bilhões ao ano** (Carvalho, 2020).

- **O Cenário Crítico do Nordeste:**

- **Infraestrutura Deficitária:** Uma pesquisa da CNT (2024) revela que **70,3% das rodovias** da região Nordeste são classificadas entre regular, ruim ou péssima.
- **Risco Elevado:** Em 2016, a chance de um acidente ser letal era, em média, **78% maior** para vítimas de ocorrências nessa região (Barroso et al., 2019).
- **Motocicletas:** Em 2014, o Nordeste concentrou **44% de todos os óbitos** decorrentes de acidentes com motocicletas nas rodovias federais brasileiras (Carvalho and Rabetti, 2015).

Diante desse cenário, a metodologia preditiva baseada em dados se torna relevante, inovando ao integrar dados históricos de acidentes e condições climáticas para classificar trechos críticos e simular o risco em vias planejada

# Objetivos

- **Objetivo Geral**

- Desenvolver um modelo analítico baseado em dados históricos e meteorológicos que permita identificar trechos críticos nas rodovias federais do Nordeste e prever o risco de acidentes em trechos novos, utilizando dados da PRF e do INMET.

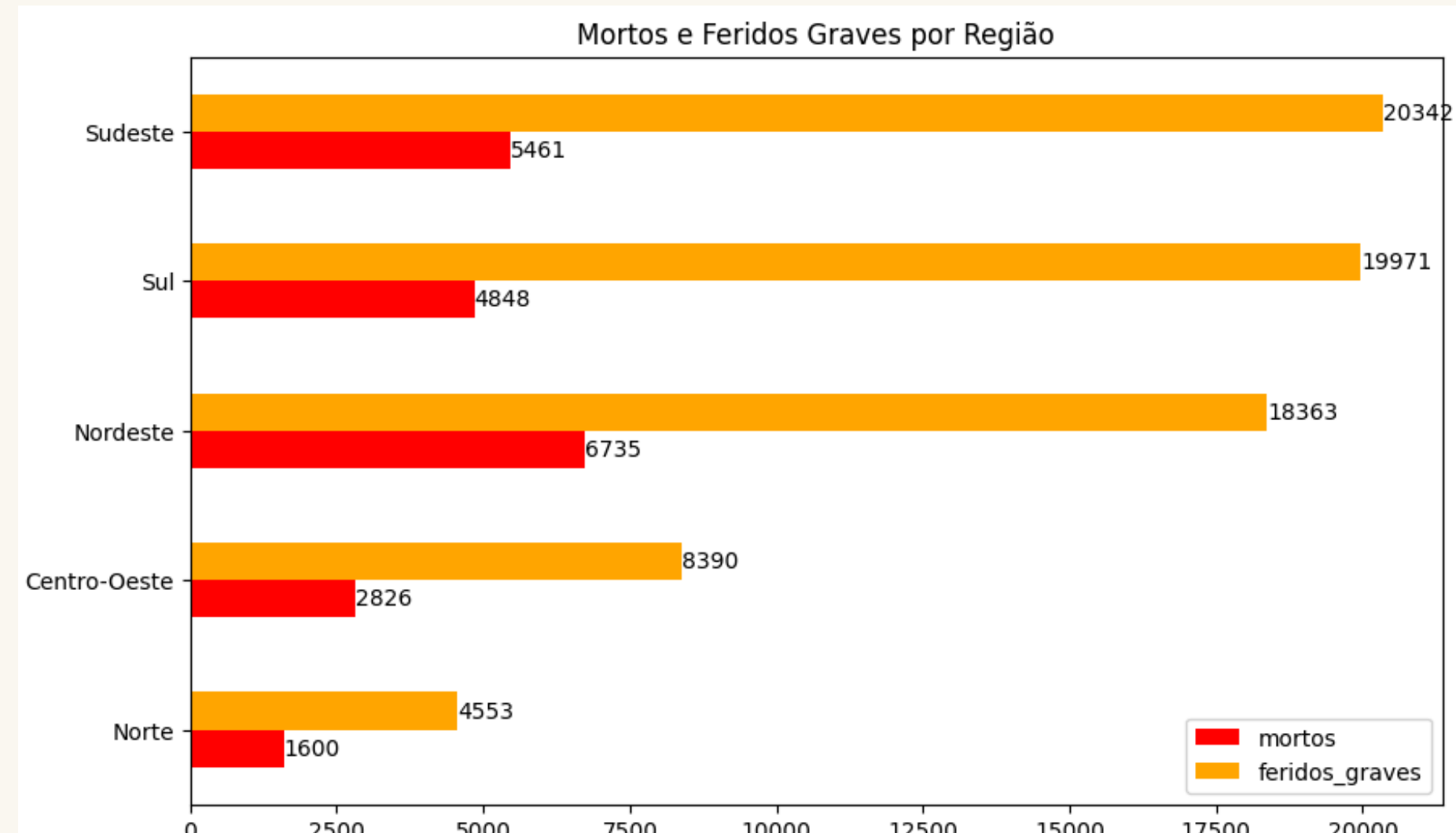
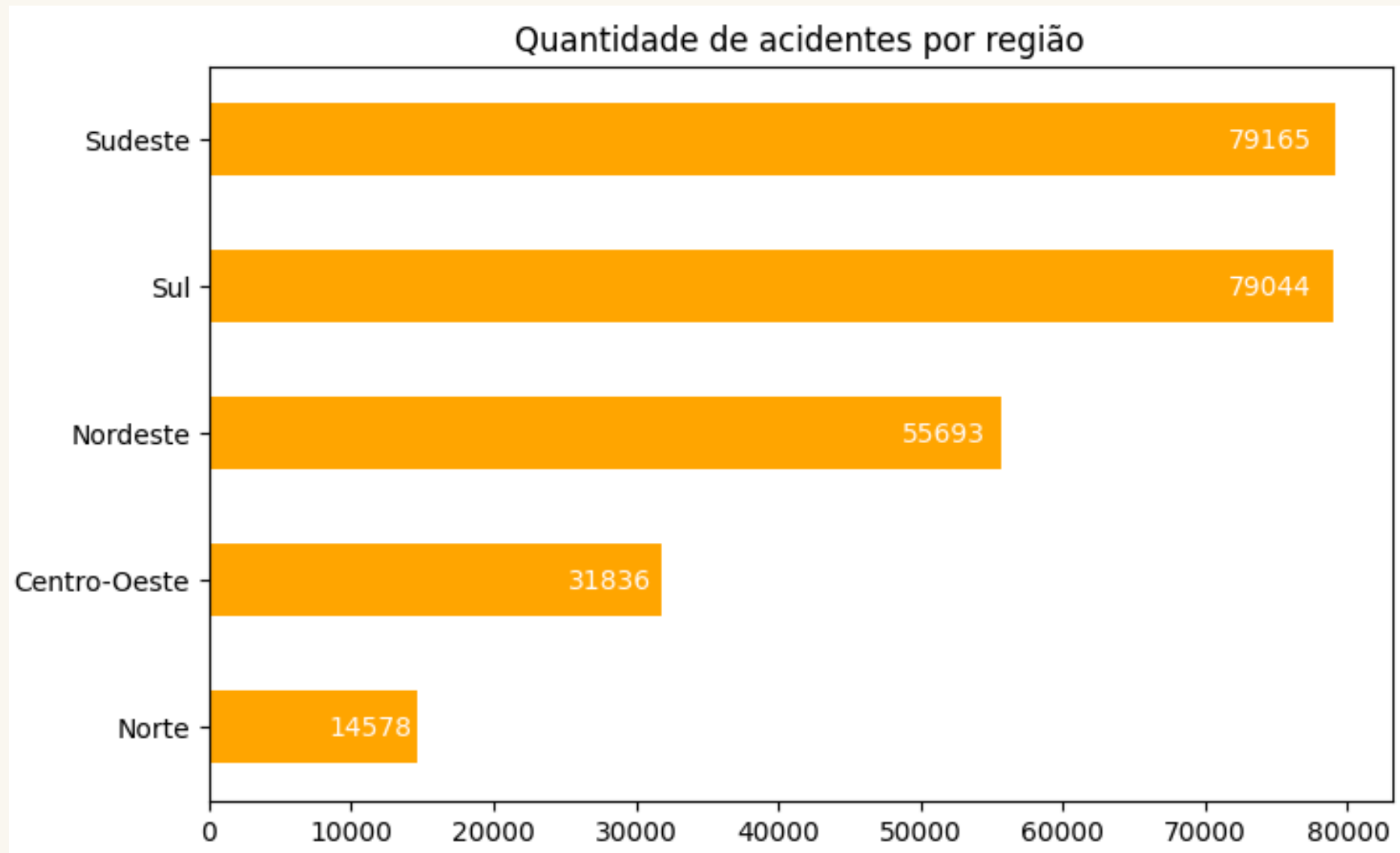
- **Objetivos Específicos**

1. Coletar, integrar e tratar dados abertos da Polícia Rodoviária Federal (PRF) e dados meteorológicos do INMET.
2. Analisar padrões de ocorrência e gravidade dos acidentes com base em variáveis como tipo de pista, clima, horário, localização e sazonalidade.
3. Modelar a probabilidade de risco de acidentes em trechos existentes, utilizando técnicas de aprendizado supervisionado.
4. Criar uma aplicação interativa que visualize os trechos classificados por risco e permita simulações para trechos novos com entrada manual de parâmetros.

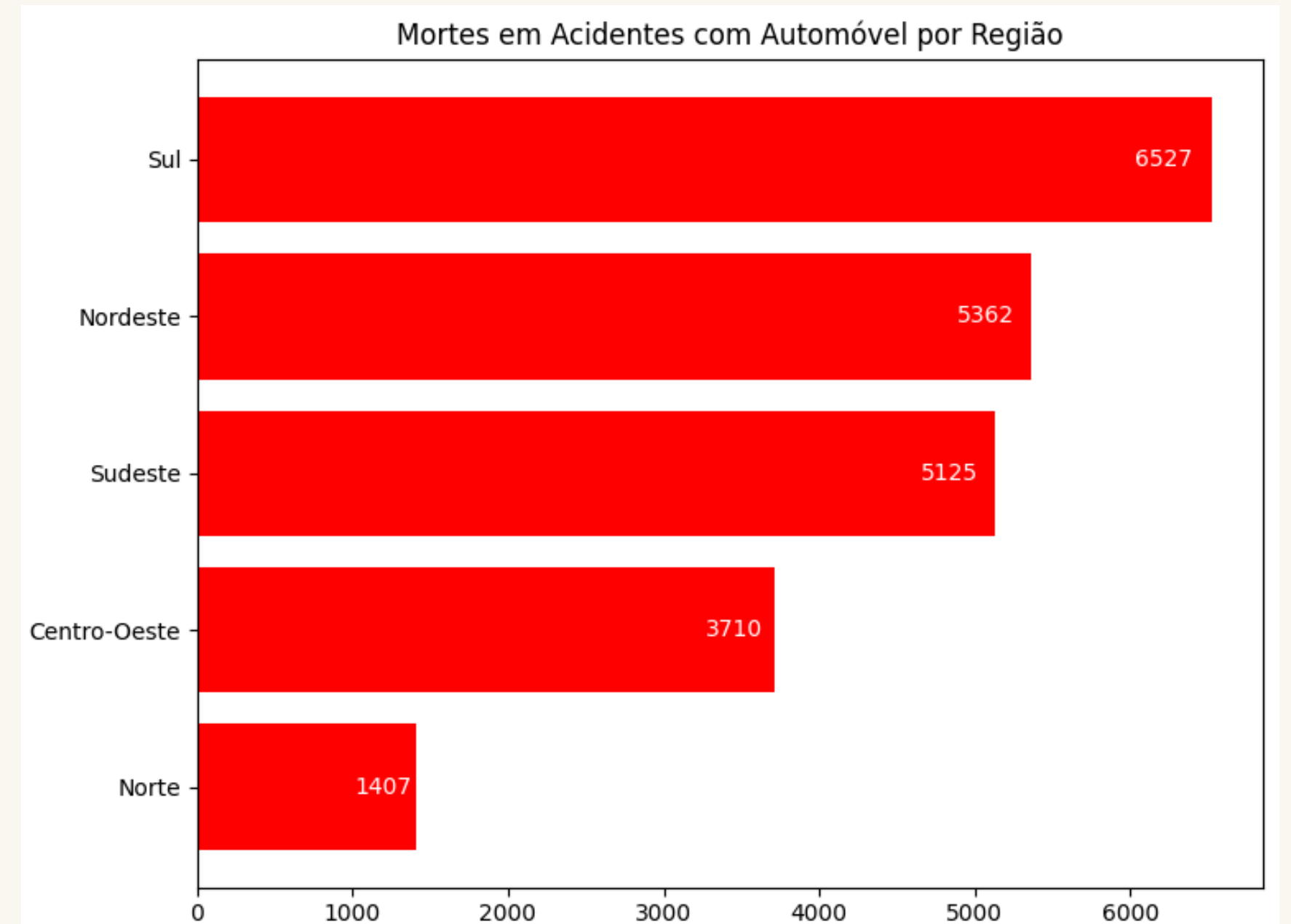
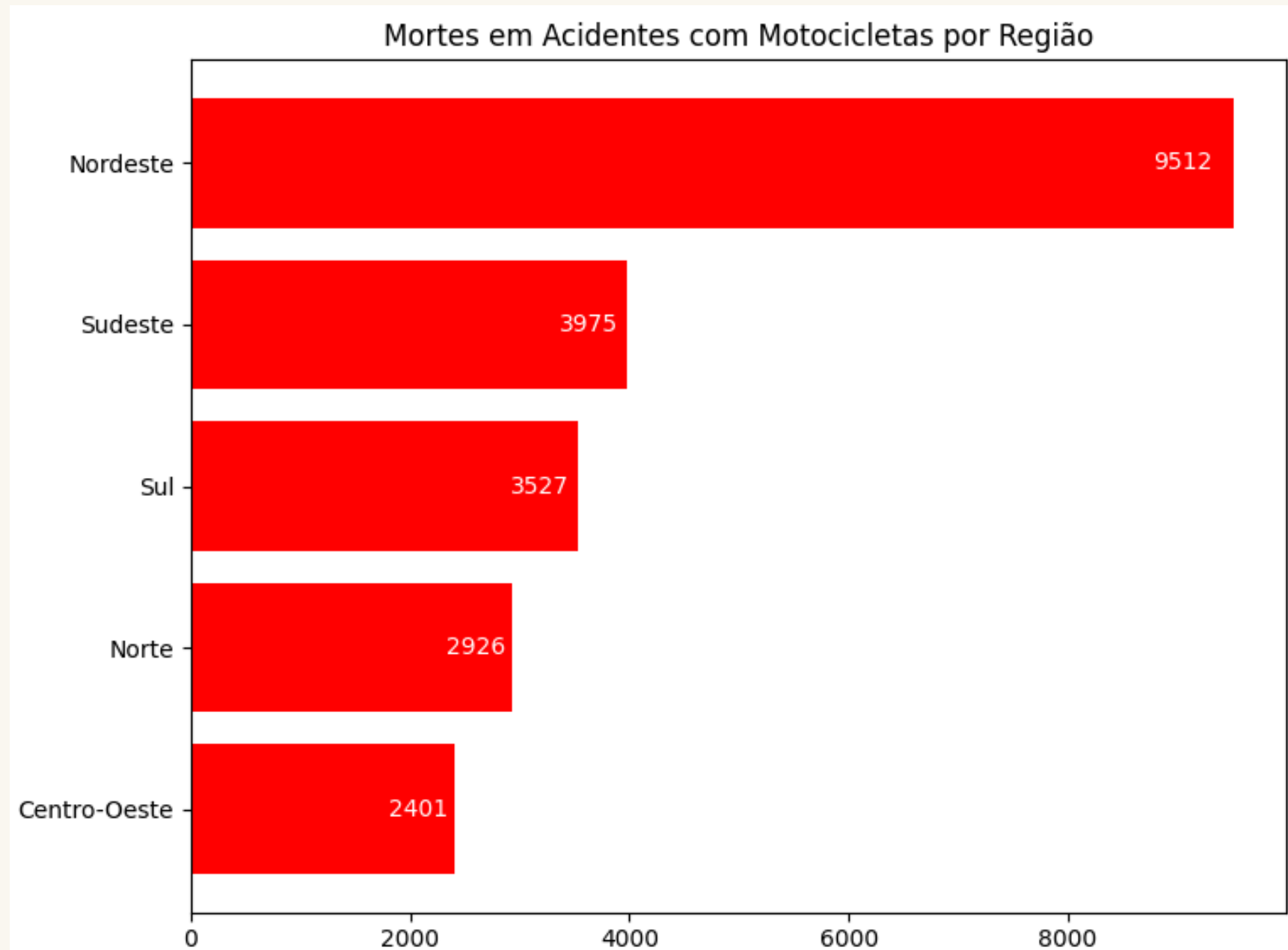
# Dados

- Fontes de Dados
  - **Polícia Rodoviária Federal (PRF):** Dados de acidentes agrupados por pessoa, abrangendo o período de 2017 a 2025, com 37 atributos.
  - **Instituto Nacional de Meteorologia (INMET):** Fornece dados meteorológicos, especificamente a variável precipitação total (mm).
- Variáveis Relevantes
  - **Gravidade (Variável-Alvo):** Classificada como 1 (Morte ou Ferimentos Graves) e 0 (Caso contrário). Essa escolha prioriza a identificação de ocorrências severas.
  - **Atributos Explicativos:** UF, BR, KM, dia da semana, fase do dia, sentido da via, condição meteorológica, tipo de pista, traçado da via, tipo de veículo, precipitação total diária no dia do acidente.

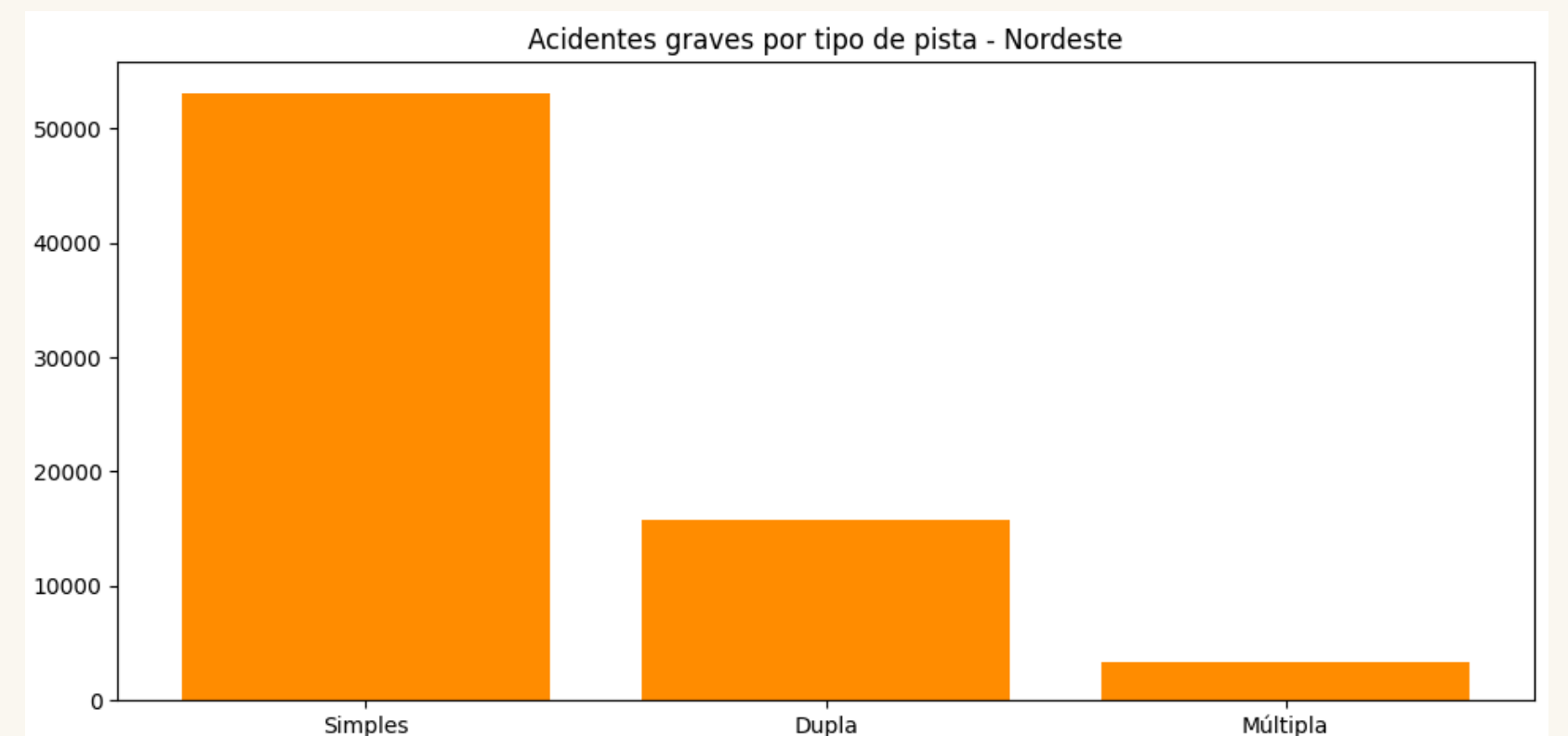
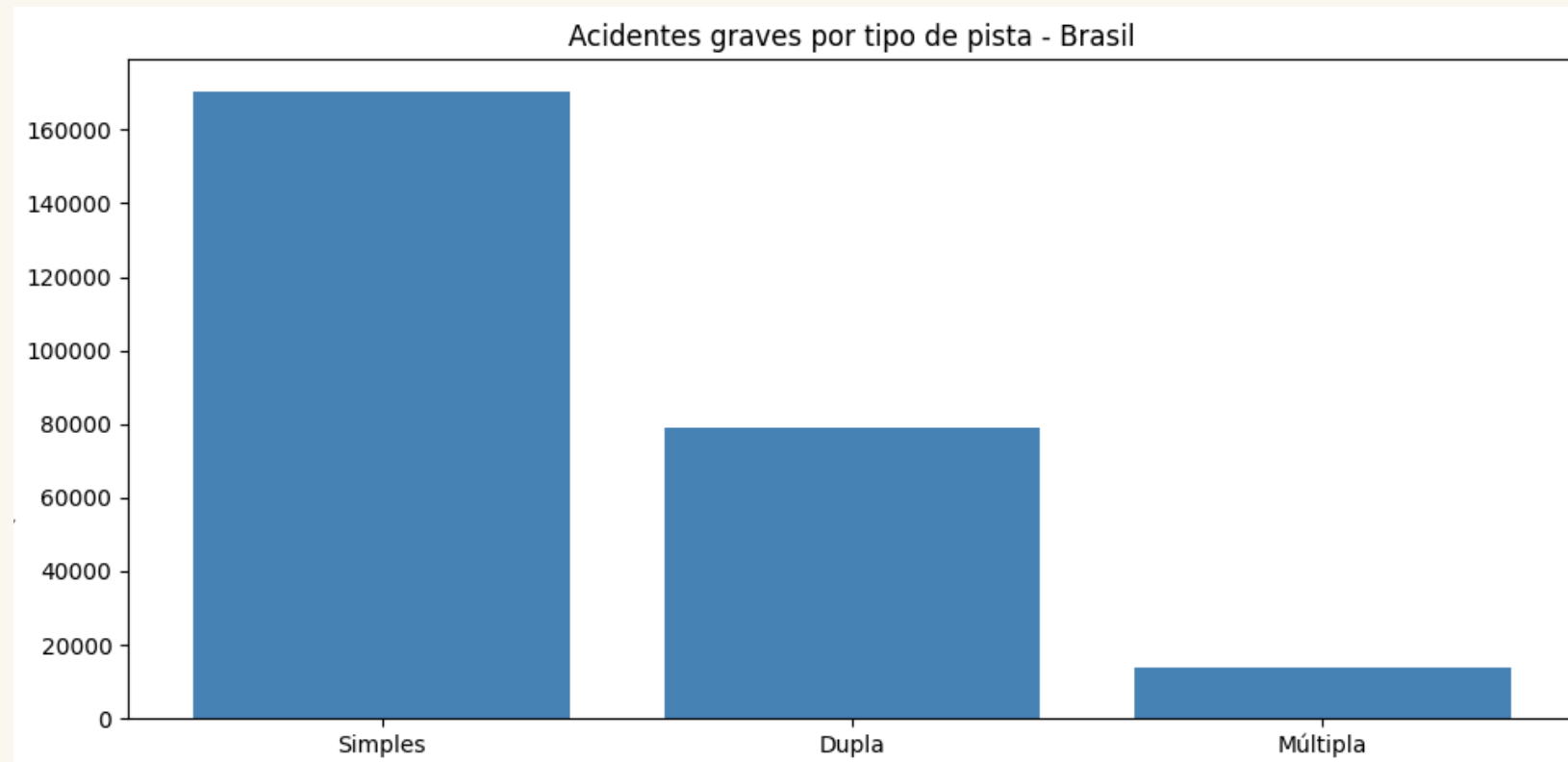
# Exploração dos dados



# Exploração dos dados

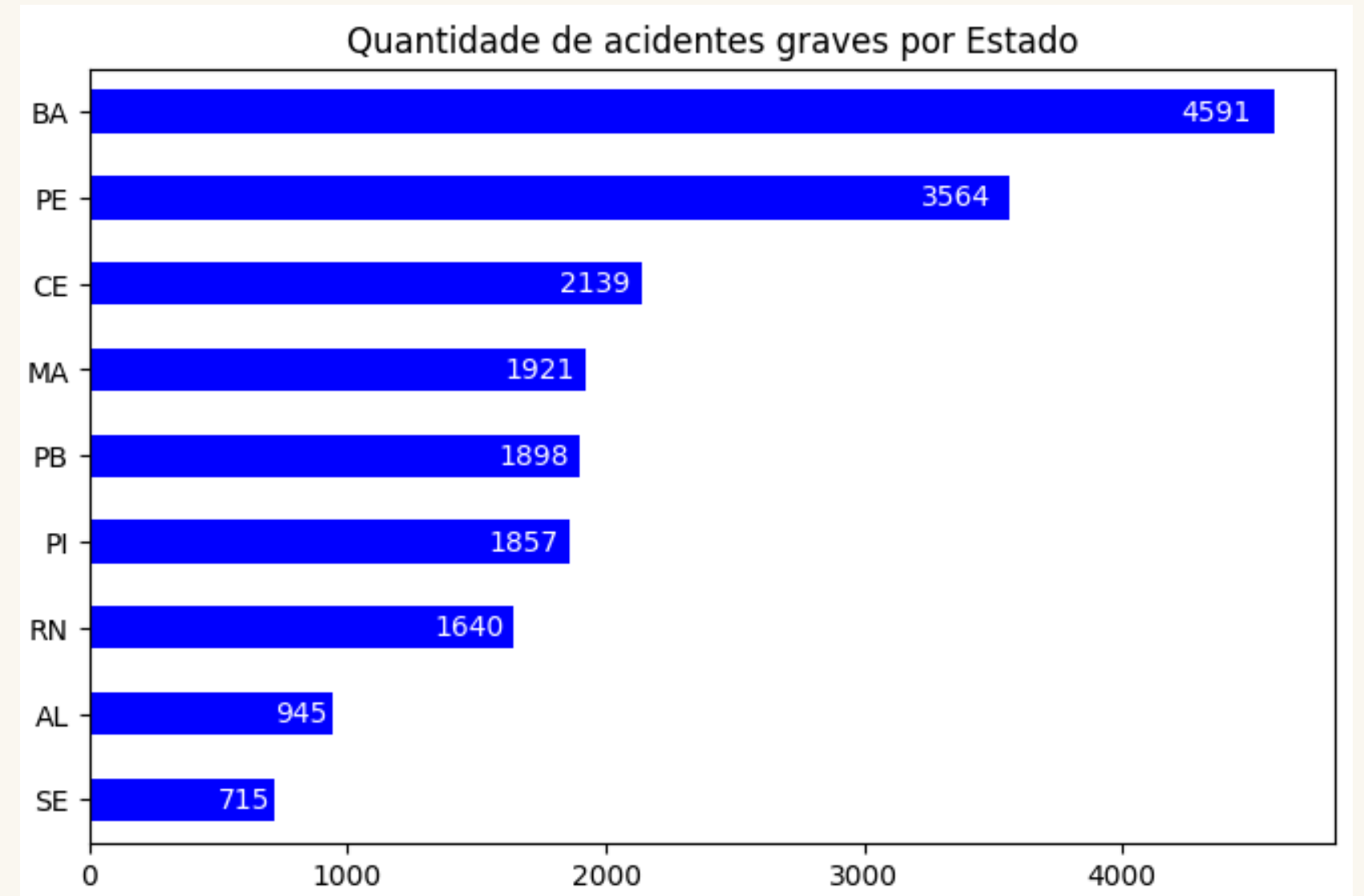
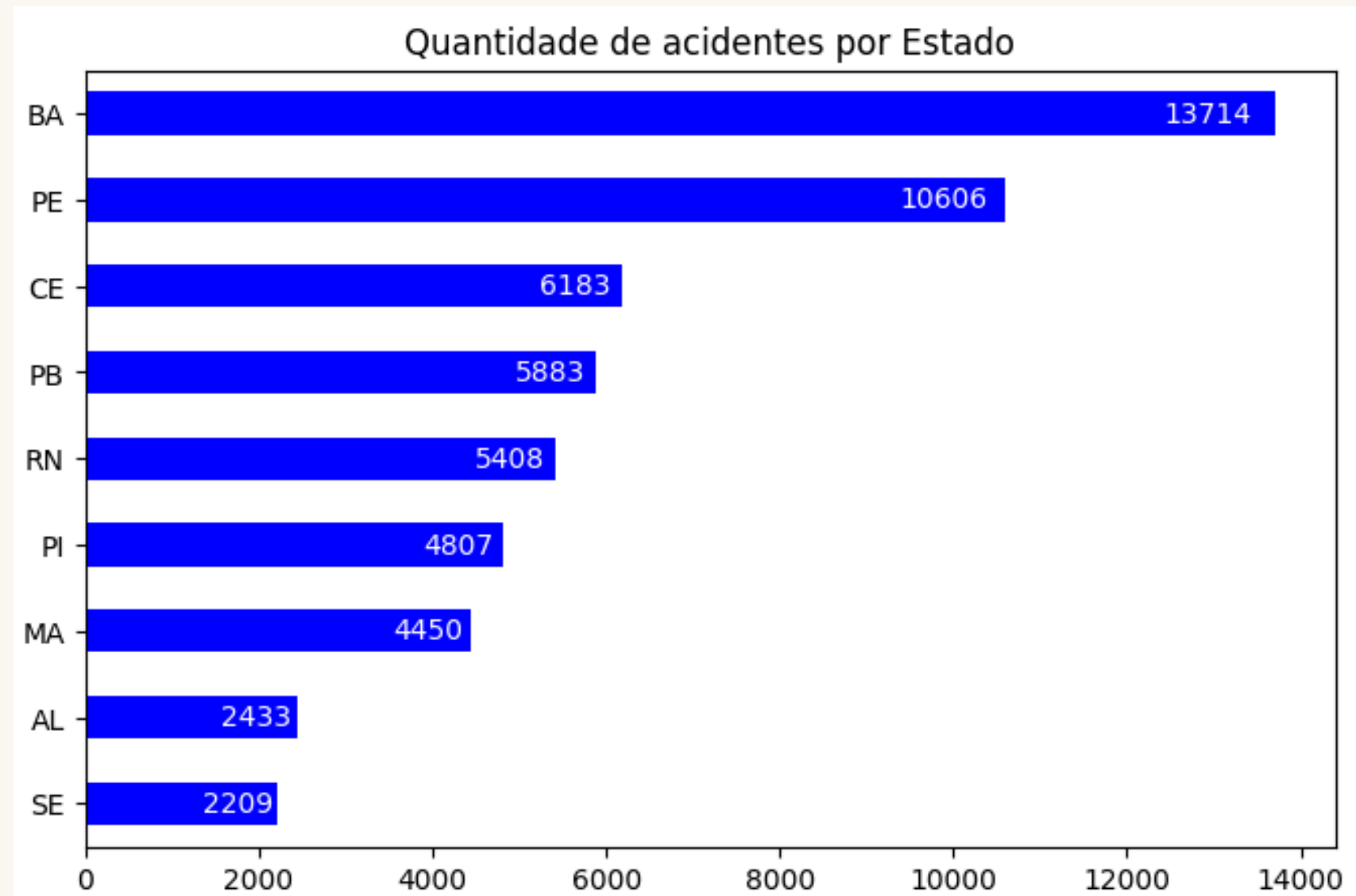


# Exploração dos dados

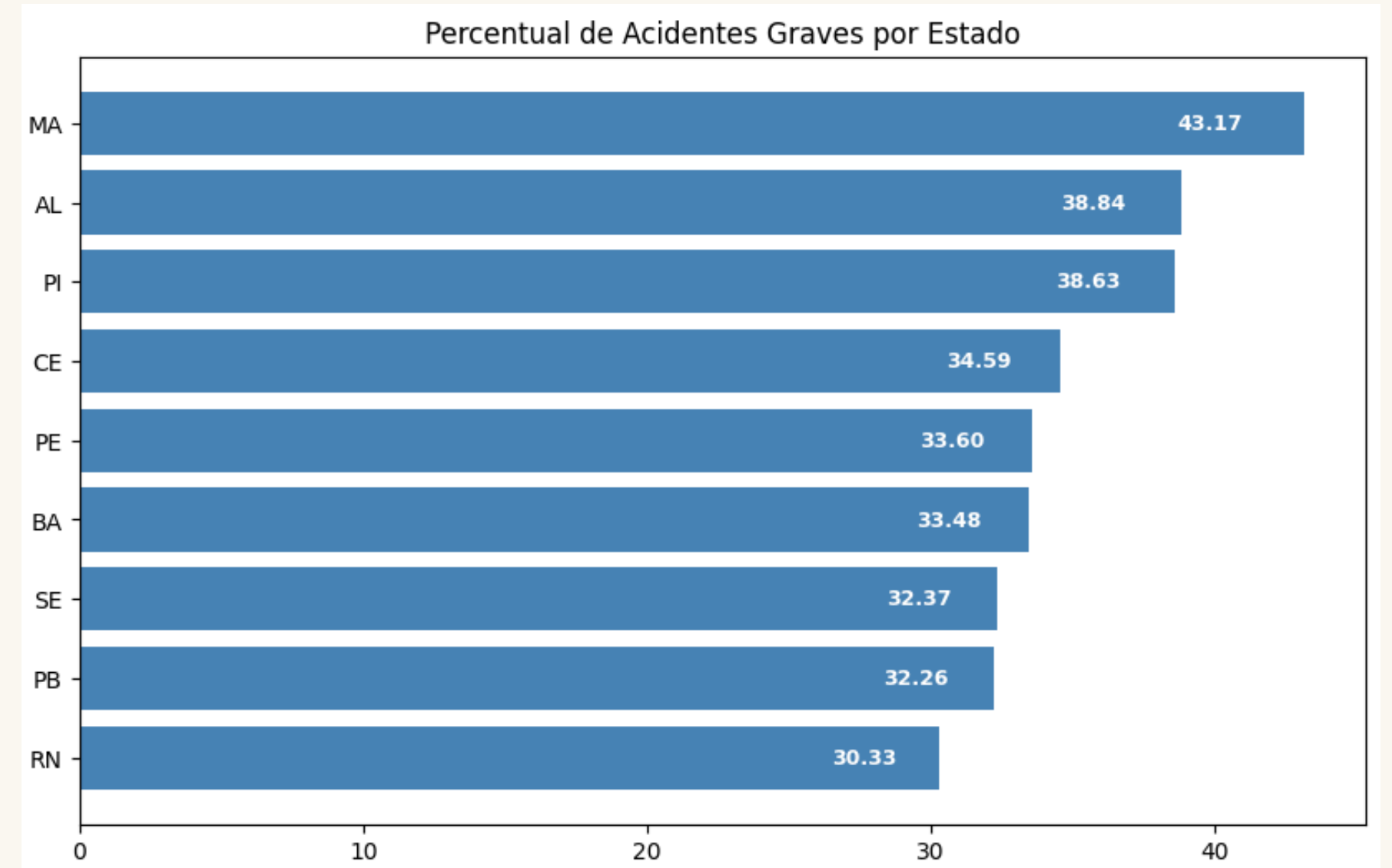
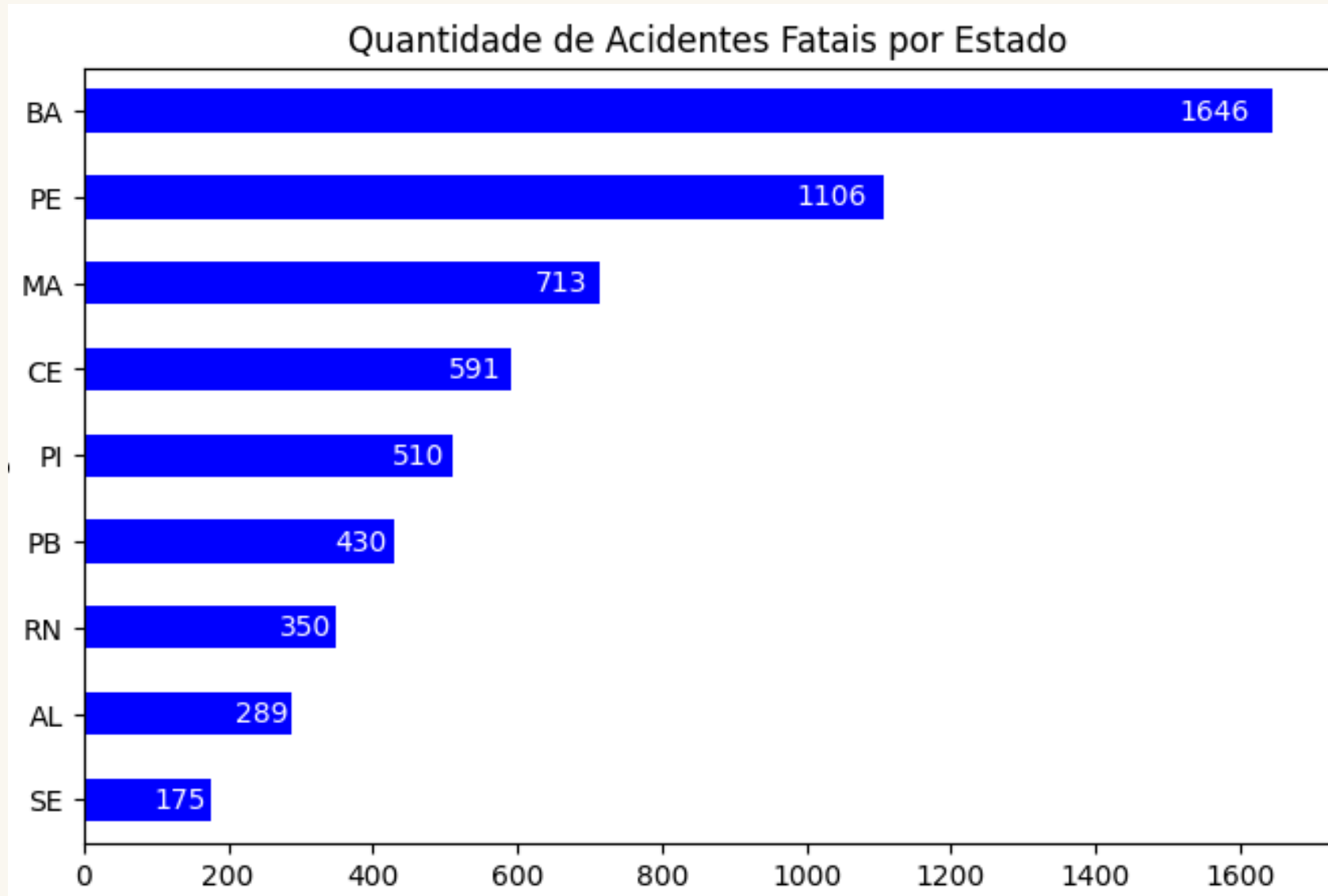




# Exploração dos dados



# Exploração dos dados



# Metodologia

- **O Processo de Aprendizado de Máquina** (Lantz, 2013)
  1. **Coleta de Dados:** Obtenção das bases da **PRF** e **INMET**.
  2. **Preparação dos Dados:** Integração, limpeza, tratamento de valores ausentes e transformação de variáveis em formato adequado.
  3. **Treinamento, Otimização e Validação:** Aplicação dos algoritmos e ajuste de hiperparâmetros.
  4. **Melhoria do Modelo:** Avaliação comparativa e seleção do modelo mais robusto.
- **Algoritmos Supervisionados Seleccionados**
  - **Regressão Logística:** Utilizada para prever respostas qualitativas, modelando a probabilidade de uma observação pertencer a uma categoria específica (acidente grave ou não). É valorizada pela sua simplicidade e interpretabilidade.
  - **Gradient Boosting Machines (GBM):** Técnica robusta de aprendizado em conjunto (ensemble learning) que constrói sequencialmente múltiplas árvores de decisão simples, visando corrigir os erros dos modelos anteriores para atingir alta performance e acurácia.
- **Avaliação de Modelos**
  - Para lidar com o desequilíbrio natural dos dados (acidentes graves são minoria), será utilizada a técnica de Validação Cruzada **Estratificada (Stratified k-Fold Cross-Validation)**. A estratificação garante que cada subconjunto de teste preserve a mesma proporção de acidentes graves e não graves, resultando em estimativas de desempenho mais estáveis e confiáveis.

# Comunicação e Aplicação Prática dos Resultados

Os resultados serão disponibilizados em uma aplicação interativa.

- **Plataforma Interativa**

- Um relatório interativo desenvolvido, a princípio, através da plataforma Looker Studio, permitirá a exploração dos dados e a visualização dos trechos críticos.

- **Principais Funcionalidades da Aplicação:**

1. **Relatórios Dinâmicos:** Dashboards interativos que apresentam indicadores-chave, como número e gravidade de acidentes, tipos de veículos envolvidos e condições meteorológicas.
2. **Mapeamento Georreferenciado:** Permite a visualização direta dos trechos críticos em mapas interativos, com a possibilidade de filtragem por período, tipo de via e condição climática.
3. **Simulações Preditivas:** Aplicação dos modelos de Machine Learning para estimar a probabilidade de ocorrência de acidentes em trechos específicos, incluindo vias em fase de projeto ou construção.

Essa aplicação visa apoiar a tomada de decisões estratégicas por órgãos públicos, contribuindo diretamente para a redução de acidentes fatais e o fortalecimento do planejamento viário preventivo no Nordeste.

◦

# Contribuições Esperadas

Apoiar a tomada de decisões estratégicas por órgãos públicos, contribuindo para a redução de acidentes fatais e o fortalecimento do planejamento viário preventivo.

## Próximos Passos

- Melhoria dos hiperparâmetros
- Testar outros modelos
- Avaliar a necessidade de criação de uma base balanceada

# Referências Bibliográficas

Barroso, G. T., Bertho, A. C. S., and Veiga, A. d. C. (2019). A letalidade dos acidentes de trânsito nas rodovias federais brasileiras em 2016. *Revista Brasileira de Estudos de População*, 36:e0074. Disponível em: <https://doi.org/10.20947/S0102-3098a0074>.

Carvalho, C. and Rabetti, M. (2015). Acidentes de trânsito nas rodovias federais brasileiras: caracterização, tendências e custos para a sociedade. Instituto de Pesquisa Econômica Aplicada—IPEA.

Carvalho, C. H. R. d. (2020). Custos dos acidentes de trânsito no brasil: estimativa simplificada com base na atualização das pesquisas do IPEA sobre custos de acidentes nos aglomerados urbanos e rodovias.

CNT (2024). Pesquisa CNT de rodovias 2024. Acesso em: 3 ago. 2025.

Instituto de Pesquisa Econômica Aplicada (IPEA) (2025). Taxa de mortes no trânsito está associada ao desenvolvimento econômico. Acessado em: 3 ago. 2025.

Obrigada!