

**TRABALHO DE BIG DATA EM PYTHON**

**ANÁLISE DE DADOS DE ACIDENTES RODOVIÁRIOS DA POLÍCIA RODOVIÁRIA FEDERAL (PRF) NO BRASIL**

**SALVADOR 2024.2**



**TRABALHO DE BIG DATA EM PYTHON**

**NATANAEL HENRIQUE ENCARNAÇÃO DAS MANDIAS (202303622121) LAÍS MEDEIROS COSTA GONÇALVES (202308705301)**

**LETÍCIA MEDEIROS COSTA GONÇALVES (202308705296) ROBERTA SUED NASCIMENTO GOMES DE SANTANA (202308425986)**

**PAULO HENRIQUE LEAL DOS SANTOS (202303706049)**

Atividade avaliativa da disciplina ARA0168 Tópicos de Big Data em Python, ministrada pelo docente Roney Malaguti.

**SALVADOR 2024.2**

**SUMÁRIO**

**1. INTRODUÇÃO........................................................................**

**2. OBJETIVO DO PROJETO................................................**

**3. ESTRUTURA DOS DADOS............................................**

**4. PROCESSAMENTO DOS DADOS.......................................................**

**5. ANÁLISES EXPLORATÓRIAS SIMPLES................................................**

**5.1 Distribuição Temporal dos Acidentes..........................................**

**5.2 Distribuição dos Acidentes por Dia da Semana..........................**

**5.3 Distribuição dos Acidentes por Horário do Dia...........................**

**5.4 Acidentes por Unidades Federativa (UF).....................................**

**5.5 Análise por Tipo de Acidente.......................................**

**5.6 Condições Meteorológicas durante Acidentes...........................**

**5.7 Classificação da Severidade dos Acidentes...............................**

**5.8 Severidade por Tipo de pista...............................**

**5.9 Impacto do Traçado da Via...............................**

**5.10 Número de Veículos e Severidade dos Acidentes......................**

**6. ANÁLISES AVANÇADAS COM APRENDIZADO DE MÁQUINAS.........**

**6.1 Classificação da Gavidade do Acidente (Árvore de Decisão)....**

**6.2 Clusterização dos Acidentes (K-Means)...................................**

**6.3 Regressão para Estimar o Número de Feridos............................**

**7. CONCLUSÕES........................**

**8. RELATÓRIO...........................................**

# Análise de Dados de Acidentes Rodoviários da Polícia Rodoviária Federal (PRF) no Brasil

1. **Introdução**

Este trabalho visa apresentar uma análise detalhada de acidentes rodoviários ocorridos em estradas federais do Brasil, com base nos dados disponibilizados pela Polícia Rodoviária Federal (PRF) de 2007 a 2020. Utilizando ferramentas de análise de dados e aprendizado de máquina, buscamos identificar padrões de ocorrência e severidade de acidentes, bem como os fatores ambientais e humanos que

influenciam esses incidentes. A análise tem o propósito de fornecer insights para políticas de segurança pública, auxiliando no planejamento de recursos e na prevenção de acidentes.

# Objetivo do Projeto

O principal objetivo deste projeto é realizar uma análise exploratória e preditiva sobre os acidentes rodoviários no Brasil. Especificamente, buscamos identificar padrões que possam apoiar o desenvolvimento de estratégias de segurança e de resposta a emergências, respondendo a perguntas como:

* + Quais condições meteorológicas aumentam a probabilidade de ocorrência de acidentes?
  + Quais fatores contribuem para a severidade dos acidentes?
  + Existe alguma correlação entre a localização geográfica e a frequência dos acidentes?

Esses insights podem apoiar a PRF e outras agências na formulação de políticas e campanhas de conscientização, assim como na otimização de recursos operacionais.

# Estrutura dos Dados

Cada conjunto de dados representa um ano de registro de acidentes e compartilha uma estrutura padronizada, com os seguintes campos:

* + **id**: Identificador único do acidente
  + **data\_inversa**: Data do acidente
  + **dia\_semana**: Dia da semana do acidente
  + **horario**: Horário do acidente
  + **uf**: Unidade Federativa onde ocorreu o acidente
  + **br**: Número da rodovia federal
  + **km**: Localização exata em quilômetros
  + **municipio**: Município de ocorrência
  + **causa\_acidente**: Causa provável do acidente
  + **tipo\_acidente**: Classificação do tipo de acidente
  + **classificacao\_acidente**: Nível de severidade
  + **fase\_dia**: Fase do dia (ex. manhã, tarde, noite)
  + **sentido\_via**: Sentido da via
  + **condicao\_metereologica**: Condições climáticas
  + **tipo\_pista**: Tipo de pista (ex. simples, dupla)
  + **tracado\_via**: Características da via (ex. reta, curva)
  + **uso\_solo**: Tipo de área ao redor (urbano, rural)
  + **pessoas, mortos, feridos\_leves, feridos\_graves, ilesos, ignorados, feridos, veiculos**: Dados sobre vítimas e veículos envolvidos.

Para os anos de 2007 a 2015, todos os arquivos incluíram o campo **ano**, enquanto de 2016 a 2020 esse campo foi derivado do campo **data\_inversa**.

# Processamento dos Dados

Os dados foram carregados e padronizados a partir de arquivos CSV. O processo de tratamento dos dados incluiu:

* + **Preenchimento de Campos Ausentes**: O campo **ano**, ausente em alguns arquivos mais recentes, foi derivado do campo **data\_inversa** para assegurar a continuidade temporal.
  + **Padronização de Valores**: Para campos com discrepâncias ou

inconsistências, aplicamos técnicas de limpeza e padronização de valores, assegurando consistência e confiabilidade na análise.

# Análises Exploratórias Simples

Inicialmente, realizamos dez análises exploratórias para obter uma compreensão inicial dos dados e identificar padrões relevantes:

# Distribuição Temporal dos Acidentes

Elaboramos gráficos de linha que mostram a quantidade de acidentes ao longo dos anos, permitindo observar tendências gerais e identificar períodos de aumento ou diminuição na incidência.

# Distribuição dos Acidentes por Dia da Semana

Identificamos os dias da semana com maior ocorrência de acidentes, fornecendo

informações sobre períodos críticos para planejamento de operações e campanhas de conscientização.

# Distribuição dos Acidentes por Horário do Dia

Analisamos os acidentes ao longo do dia para identificar horários de maior risco e relacionar com o fluxo de tráfego.

# Acidentes por Unidade Federativa (UF)

Mapeamos as UFs com maior número de acidentes, oferecendo uma visão regional dos dados e ajudando a identificar áreas que requerem mais atenção.

# Análise por Tipo de Acidente

Classificamos os acidentes de acordo com o tipo (ex. colisão, capotamento) para compreender as categorias mais frequentes e auxiliar na identificação de riscos específicos.

# Condições Meteorológicas durante Acidentes

Analisamos a relação entre condições climáticas e acidentes, determinando se fatores como chuva e neblina aumentam o risco de acidentes.

# 5.7 Classificação da Severidade dos Acidentes

Investigamos a distribuição dos acidentes em termos de severidade, analisando as proporções de acidentes leves, graves e fatais.

# 5.8 Severidade por Tipo de Pista

Avaliamos a relação entre o tipo de pista e a gravidade dos acidentes, especialmente entre pistas simples e duplas.

# Impacto do Traçado da Via

Analisamos a influência do traçado da via (reta, curva) na ocorrência e gravidade dos acidentes.

# Número de Veículos e Severidade dos Acidentes

Exploramos a correlação entre o número de veículos envolvidos e a severidade dos acidentes.

# Análises Avançadas com Aprendizado de Máquina

Para aprofundar a análise, aplicamos técnicas de aprendizado de máquina para identificar padrões ocultos e prever aspectos específicos dos acidentes.

# Classificação da Gravidade do Acidente (Árvore de Decisão)

Usamos o algoritmo de árvore de decisão para prever a gravidade dos acidentes com base em variáveis como tipo de acidente, condição meteorológica e tipo de pista. Esse modelo destacou fatores críticos para a severidade dos acidentes, sendo especialmente útil para identificar condições de risco elevado.

**Resultados**: A condição meteorológica e o tipo de pista mostraram ser os fatores mais significativos na previsão da gravidade.

# Clusterização dos Acidentes (K-Means)

Empregamos o algoritmo K-Means para identificar grupos de acidentes com características semelhantes, baseados em variáveis como fase do dia, condição meteorológica e tipo de pista. Esse método ajudou a identificar cenários específicos de alto risco.

**Resultados**: Identificamos clusters que representam diferentes condições de risco, como acidentes em condições chuvosas e à noite, evidenciando cenários que exigem intervenções específicas.

# Regressão para Estimar o Número de Feridos

Aplicamos regressão linear para prever o número de feridos em acidentes, utilizando variáveis como tipo de acidente, condição meteorológica e fase do dia. Esse modelo fornece uma ferramenta para antecipar os recursos necessários em cenários de emergência.

**Resultados**: O tipo de acidente e as condições meteorológicas apresentaram forte correlação com o número de feridos, auxiliando na alocação de equipes de resposta.

# Conclusões

As análises exploratórias e preditivas realizadas revelaram uma série de padrões importantes nos dados de acidentes:

* + Existe uma variação significativa na frequência dos acidentes ao longo da semana e do dia, com certos períodos sendo mais críticos.
  + Condições meteorológicas adversas e pistas simples estão associadas a uma maior severidade dos acidentes.
  + Modelos de aprendizado de máquina destacaram fatores cruciais, como condição climática e tipo de pista, para a severidade dos acidentes, fornecendo uma base sólida para intervenções.

A utilização de aprendizado de máquina possibilitou uma análise preditiva aprofundada, permitindo criar modelos que antecipam a gravidade e os impactos dos acidentes. Esses resultados são valiosos para a PRF e outras entidades no desenvolvimento de campanhas de prevenção e no planejamento de respostas a emergências.

1. **RELATÓRIO**

Para o desenvolvimento do trabalho, relatamos então que:

O aluno Natanael Henrique Encarnação das Mandias, contribuiu totalmente na realização do trabalho.

O aluno Paulo Henrique Leal dos Santos, contribuiu totalmente na realização do trabalho.

A aluna Laís Medeiros Costa Gonçalves, contribuiu totalmente na realização do trabalho.

A aluna Letícia Medeiros Costa Gonçalves, contribuiu totalmente na realização do trabalho.

A aluna Roberta Sued Nascimento Gomes de Santana, contribuiu totalmente na realização do trabalho.

Link do Google Colab:

<https://colab.research.google.com/drive/1RZJABNGluHsaSLEioid-7XIUK4kbO_YP#scrollTo=WfSjjelpXo0S>

GitHub dos alunos:

Paulo Henrique Leal dos Santos: <https://github.com/lealhenriq/analisededadostopicodebigdataempython>

Laís Medeiros Costa Gonçalves: <https://github.com/LaisMedeiros02/Trabalho-de-Big-Data-em-Python>

Letícia Medeiros Costa Gonçalves: <https://github.com/Leticiagoncalves01/Trabalho-de-Big-Data-em-Python>

Roberta Sued Nascimento Gomes de Santana:

Natanael Henrique Encarnação das Mandias: <https://github.com/Natanaelhenriqu/Trabalho-de-Big-Data-em-Python.git>