



UNIVERSIDADE NOVE DE JULHO - UNINOVE  
PLANEJAMENTO EM INFORMATICA

EXCLUÍDOS OS DADOS SOBRE OS AUTORES EM ATENDIMENTO A  
LGPD - LEI GERAL DE PROTEÇÃO DE DADOS

ANÁLISE E PREVISÃO DA GRAVIDADE DE ACIDENTES  
FERROVIÁRIOS

São Paulo  
2024

**EXCLUÍDOS OS DADOS SOBRE OS AUTORES EM ATENDIMENTO A  
LGPD - LEI GERAL DE PROTEÇÃO DE DADOS**

**ANÁLISE E PREVISÃO DA GRAVIDADE DE ACIDENTES  
FERROVIÁRIOS**

Projeto apresentado a Universidade Nove de Julho - UNINOVE, como parte dos requisitos obrigatórios para obtenção do título de TECNÓLOGO EM TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS.

Prof. Orientador: Edson Melo de Souza, Dr.

**São Paulo  
2024**

# RESUMO

---

**Contexto:** A segurança ferroviária é uma preocupação crescente, então entender e prever a gravidade dos acidentes é fundamental. Este estudo busca analisar dados históricos de acidentes ferroviários usando técnicas de aprendizado de máquina. Queremos encontrar padrões que possam ajudar a evitar acidentes graves no futuro. Criamos um modelo preditivo que considera várias características dos acidentes para prever sua gravidade, oferecendo informações valiosas para melhorar a segurança. **Objetivo:** O objetivo principal deste projeto é desenvolver um modelo de aprendizado de máquina que consiga prever a gravidade dos acidentes ferroviários com base em várias variáveis dos acidentes. Uma previsão precisa pode ajudar muito na implementação de medidas preventivas e na melhoria contínua da segurança ferroviária. **Método:** Para alcançar isso, fizemos uma análise exploratória detalhada dos dados e construímos um modelo de Random Forest para classificação. A metodologia seguiu várias etapas, são elas: Preparação dos Dados: Primeiro, tratamos os dados faltantes e codificamos variáveis categóricas para garantir a integridade e consistência dos dados; Divisão dos Dados: Dividimos os dados em conjuntos de treino e teste para validar o modelo; Treinamento do Modelo: Usamos o RandomForestClassifier para treinar o modelo com os dados de treino; Avaliação do Modelo: Avaliamos o desempenho do modelo com base em métricas como acurácia, precisão, recall e F1-score; Análise de Importância das Características: Identificamos as características que mais influenciam a previsão da gravidade dos acidentes. **Resultados:** Os resultados mostraram uma acurácia de 84% para o modelo. As características mais importantes para prever a gravidade dos acidentes foram a informação sobre o condutor do trem (Equipagem), o número de óbitos no acidente e a natureza do acidente (como descarrilamento, colisão, etc.). Essas características são essenciais para prever a gravidade e podem orientar ações preventivas específicas para reduzir acidentes graves. **Conclusão:** A principal conclusão deste estudo é que há uma tendência crescente no número de acidentes ferroviários ao longo dos anos. Se essa tendência continuar, podemos esperar um aumento contínuo no número de acidentes, reforçando a necessidade de intervenções preventivas. O modelo se mostrou eficaz na previsão da gravidade dos acidentes, e os insights sugerem que melhorias focadas, como o treinamento adequado dos condutores e a manutenção rigorosa da infraestrutura ferroviária, são essenciais para aumentar a segurança.

**Palavras-chave:** Segurança, ferroviária, acidentes, gravidade, modelo, preditivo, Análise, exploratória, Random Forest, Acurácia, fatalidades, descarrilamento, colisão.

# ABSTRACT

---

**Contextualization:** Railway safety is a growing concern, so understanding and predicting the severity of accidents is crucial. This study aims to analyze historical data on railway accidents using machine learning techniques. We seek to find patterns that can help prevent serious accidents in the future. We have developed a predictive model that considers various accident characteristics to forecast their severity, providing valuable insights to enhance safety. **Objective:** The main objective of this project is to develop a machine learning model that can predict the severity of railway accidents based on various accident variables. Accurate prediction can greatly assist in implementing preventive measures and continuously improving railway safety. **Method:** To achieve this, we conducted a detailed exploratory analysis of the data and built a Random Forest model for classification. The methodology followed several steps: Data Preparation: Firstly, we handled missing data and encoded categorical variables to ensure data integrity and consistency; Data Splitting: We divided the data into training and testing sets to validate the model; Model Training: We used the RandomForestClassifier to train the model with the training data; Model Evaluation: We assessed the model's performance based on metrics such as accuracy, precision, recall, and F1-score; Feature Importance Analysis: We identified the features that most influence the prediction of accident severity. **Results:** The results showed an accuracy of 84% for the model. The most important features for predicting accident severity were information about the train crew, the number of fatalities in the accident, and the nature of the accident (such as derailment, collision, etc.). These features are crucial for predicting severity and can guide specific preventive actions to reduce serious accidents. **Conclusion:** The main conclusion of this study is that there is a growing trend in the number of railway accidents over the years. If this trend continues, we can expect a continuous increase in accidents, reinforcing the need for preventive interventions. The model proved effective in predicting accident severity, and insights suggest that focused improvements, such as proper training of drivers and rigorous maintenance of railway infrastructure, are essential for enhancing safety.

**Keywords:** Safety, railway, accidents, severity, model, predictive, exploratory, analysis, Random Forest, Accuracy, fatalities, derailment, collision.

# SUMÁRIO

---

|   |           |
|---|-----------|
| <b>Lista de Ilustrações</b>                                     | <b>7</b>  |
| <b>1 Introdução</b>   | <b>8</b>  |
| 1.1 Contexto . . . . .  | 8         |
| 1.2 Objetivo . . . . .  | 8         |
| 1.3 Justificativa . . . . .                                     | 8         |
| 1.4 Metodologia . . . . .                                       | 8         |
| 1.5 Estrutura do Trabalho . . . . .                             | 9         |
| 1.6 Contribuição do Estudo . . . . .                            | 9         |
| <b>2 Fundamentação Teórica</b>                                  | <b>10</b> |
| 2.1 Random Forest . . . . .                                     | 10        |
| 2.2 Biblioteca Pandas . . . . .                                 | 10        |
| 2.3 Python . . . . .  | 10        |
| 2.4 Python com aprendizado de máquina . . . . .                 | 10        |
| 2.5 Biblioteca Seaborn com o matplotlib . . . . .               | 11        |
| 2.6 Métodos de aprendizado de máquina . . . . .                 | 11        |
| 2.7 Índice de acidentes por trem em Curitiba . . . . .          | 11        |
| 2.8 Principais causas de acidentes com trem . . . . .           | 11        |
| 2.9 Mapa Ferroviario . . . . .                                  | 11        |
| 2.10 Google colab . . . . .                                     | 12        |
| <b>3 Metodologia</b>  | <b>13</b> |
| 3.1 Visão Geral . . . . .                                       | 13        |
| 3.2 Coleta de Dados . . . . .                                   | 13        |
| 3.2.1 Lista das colunas usadas para a coleta de dados . . . . . | 13        |
| 3.2.2 Exemplo de saída de dados do Dataframe . . . . .          | 14        |
| 3.3 Limpeza de dados . . . . .                                  | 15        |
| 3.3.1 Exemplo de ajuste feito nos dados . . . . .               | 15        |
| 3.4 Análise Exploratória de Dados (EDA) . . . . .               | 15        |
| 3.4.1 Distribuição Temporal . . . . .                           | 15        |
| 3.4.2 Gravidade dos Acidentes . . . . .                         | 17        |
| 3.4.3 Causas dos Acidentes . . . . .                            | 18        |
| 3.5 Modelagem . . . . .   | 18        |
| 3.5.1 Regressão Linear . . . . .                                | 18        |
| 3.5.2 Modelo de Classificação . . . . .                         | 19        |
| 3.6 Visualização dos Resultados . . . . .                       | 20        |

|          |  |           |
|----------|--|-----------|
| 3.6.1    | Boxplots . . . . .   | 20        |
| <b>4</b> | <b>Análise dos Resultados</b>                                | <b>22</b> |
| 4.1      | Análise Descritiva . . . . .                                 | 22        |
| 4.1.1    | Distribuição dos Acidentes ao Longo dos Anos . . . . .       | 22        |
| 4.1.2    | Frequência dos acidentes por Município . . . . .             | 22        |
| 4.2      | Boxplots . . . . .   | 22        |
| 4.2.1    | Número de Acidentes por Ano . . . . .                        | 22        |
| 4.2.2    | Número de Acidentes por Município em 2019 . . . . .          | 22        |
| 4.3      | Regressão Linear . . . . .                                   | 22        |
| 4.4      | Gráfico de barras horizontal . . . . .                       | 23        |
| 4.4.1    | Número de Acidentes por Tipo de Transporte do Trem . . . . . | 23        |
| <b>5</b> | <b>Conclusões</b>  | <b>24</b> |
|          | <b>Referências Bibliográficas</b>                            | <b>25</b> |
|          | <b>Anexos</b>  | <b>26</b> |
| A        | : Título . . . . .   | 26        |

## LISTA DE ILUSTRAÇÕES

---

|     |   |    |
|-----|---|----|
| 3.1 | Lista das colunas inclusas no Dataframe . . . . .   | 14 |
| 3.2 | Exemplo de saída de dados do Dataframe. . . . .   | 14 |
| 3.3 | Exemplo de ajuste dos dados separando de dia, mês e ano na coluna de data<br>dos acidentes. . . . . | 15 |
| 3.4 | Gráfico de quantidade de acidentes registrados por ano em 2004. . . . .                             | 16 |
| 3.5 | Gráfico comparando os acidentes graves e não graves. . . . .  | 17 |
| 3.6 | Gráfico com as causas diretas entre os acidentes . . . . .  | 18 |
| 3.7 | Gráfico da regressão linear . . . . .   | 19 |
| 3.8 | Boxplot da distribuição dos acidentes anuais por município em 2004 . . . . .                        | 20 |
| 3.9 | Boxplot da distribuição dos acidentes anuais por município em 2019 . . . . .                        | 21 |

# 1 INTRODUÇÃO

---

## 1.1 CONTEXTO

Os acidentes ferroviários representam um problema significativo de segurança e logística no Brasil. Apesar das melhorias tecnológicas e regulatórias ao longo dos anos, o número de acidentes ainda é preocupante e impacta tanto a operação das linhas ferroviárias quanto a segurança das comunidades próximas. O estudo desses acidentes pode fornecer insights importantes para a prevenção de futuros incidentes.

## 1.2 OBJETIVO

O principal objetivo deste projeto acadêmico é analisar os dados históricos de acidentes ferroviários no Brasil para identificar padrões, tendências e fatores contributivos. Com isso, esperamos fornecer insights para a melhoria da segurança ferroviária e a redução do número de acidentes.

## 1.3 JUSTIFICATIVA

A escolha do tema se justifica pela relevância do transporte ferroviário no Brasil, tanto para o transporte de cargas quanto de passageiros. Acidentes ferroviários não só resultam em perdas humanas e materiais, mas também afetam a eficiência do sistema logístico do país. A análise detalhada dos acidentes pode ajudar a identificar áreas críticas e desenvolver estratégias de prevenção mais eficazes.

## 1.4 METODOLOGIA

Para alcançar os objetivos propostos, seguimos uma metodologia estruturada em várias etapas:

- **Coleta de Dados:** Utilizamos uma base de dados que contém informações detalhadas sobre cada acidente ferroviário ocorrido no Brasil.
- **Limpeza de Dados:** Tratamos os dados ausentes, corrigimos tipos de dados incorretos e padronizamos valores categóricos.
- **Análise Exploratória:** Exploramos a distribuição dos acidentes ao longo do tempo, localização geográfica e gravidade.
- **Modelagem:** Aplicamos técnicas de aprendizado de máquina, como regressão linear e Random Forest, para prever tendências e classificar a gravidade dos acidentes.



- **Visualização:** Criamos visualizações para comunicar os resultados de forma clara e eficaz.

## 1.5 ESTRUTURA DO TRABALHO

Este trabalho está estruturado em cinco capítulos:

- **Capítulo 1: Introdução** - Apresenta o contexto, objetivo, justificativa e a metodologia do projeto.
- **Capítulo 2: Fundamentação Teórica** - Revisa a literatura relevante sobre segurança ferroviária, análise de dados e aprendizado de máquina.
- **Capítulo 3: Metodologia** - Detalha os métodos e técnicas utilizados na coleta, limpeza e análise dos dados.
- **Capítulo 4: Análise dos Resultados** - Apresenta os resultados das análises e modelos.
- **Capítulo 5: Conclusão** - Discute as conclusões do estudo e propõe recomendações para a melhoria da segurança ferroviária.

## 1.6 CONTRIBUIÇÃO DO ESTUDO

Esperamos que este estudo contribua para um melhor entendimento dos acidentes ferroviários no Brasil e forneça informações úteis para a formulação de políticas de segurança e a implementação de melhorias na infraestrutura ferroviária. Através da análise dos dados, buscamos identificar padrões e tendências que possam ser usados para prevenir futuros acidentes e reduzir seus impactos.

## 2 FUNDAMENTAÇÃO TEÓRICA

---

### Resumo do capítulo

*Neste capítulo, a gente vai falar sobre os principais conceitos e estudos que serviram de base para a nossa análise e previsão de acidentes ferroviários. A ideia é entender melhor a segurança ferroviária, a análise de dados e o aprendizado de máquina.*

### 2.1 RANDOM FOREST

Segundo um artigo da Escola Britânica de Artes criativas e Tecnologia ([EBAC, 2024](#)) Random Forest é um algoritmo de machine learning que usa vários subconjuntos de dados de treinamento para construir uma série de árvores de decisão. A nossa ideia principal em escolher o Random Forest partiu do ponto de procurarmos uma ferramenta que nos ajudasse com a previsão de acidentes com base nos dados já obtidos.

### 2.2 BIBLIOTECA PANDAS

A biblioteca Pandas é uma ferramenta para uso da linguagem Python, de código aberto e gratuita. Esse software é utilizado para a análise e limpeza de dados, bem como construção de gráficos e manipulação de tabelas.

O Python é muito utilizado para lidar com grandes estruturas de dados, por conta de pacotes como a biblioteca Pandas. Eles auxiliam o programador a trabalhar com áreas como machine learning, cibersegurança, data mining, data science, programação web, entre outros ([COUTINHO, Thiago, 2021](#)).

### 2.3 PYTHON

O Python é uma linguagem de programação amplamente usada em aplicações da Web, desenvolvimento de software, ciência de dados e machine learning (ML). Os desenvolvedores usam o Python porque é eficiente e fácil de aprender e pode ser executada em muitas plataformas diferentes. O software Python pode ser baixado gratuitamente, integra-se bem a todos os tipos de sistema e agiliza o desenvolvimento ([AWS, 2024](#)).

### 2.4 PYTHON COM APRENDIZADO DE MÁQUINA

Segundo ([CORBO, Antony, 2023](#)) A linguagem Python traz uma quantidade excepcional de força e versatilidade para os ambientes de aprendizado de máquina. A sintaxe simples do idioma simplifica a validação de dados e simplifica os processos de coleta de dados, processamento, refinamento de dados, limpeza, organização e análise.

## 2.5 BIBLIOTECA SEABORN COM O MATPLOTLIB

Aplicamos a biblioteca Seaborn na elaboração dos gráficos e segundo [SCHÖNHOFEN, Ricardo D \(2017\)](#) para a criação de gráficos visualmente legíveis, a biblioteca Seaborn é responsável em melhorar o visual dos gráficos gerados nas análises.

## 2.6 MÉTODOS DE APRENDIZADO DE MÁQUINA

Segundo a [IBM \(2024\)](#) aprendizado de máquina (Machine Learning), é definido pelo uso de conjuntos de dados rotulados para treinar algoritmos para classificar dados ou prever resultados com precisão. À medida que os dados de entrada são alimentados no modelo, o modelo ajusta seus pesos até que tenha sido ajustado adequadamente.

## 2.7 ÍNDICE DE ACIDENTES POR TREM EM CURITIBA

De acordo uma matéria da [CBN Curitiba \(2023\)](#), um relatório divulgado pela Agência Nacional de Transportes Terrestres (ANTT) aponta que a cidade de Curitiba registrou o maior índice de acidentes em linhas férreas em todo o país. Através dessa matéria, em nossa análise, tivemos como comprovar a veracidade dos dados, e pudemos elaborar uma análise gráfica de que Curitiba realmente é a Cidade que lidera os casos de acidentes ferroviários no País.

## 2.8 PRINCIPAIS CAUSAS DE ACIDENTES COM TREM

Uma de nossas análises aponta que a maioria dos acidentes ferroviários ocorrem pela natureza de descarrilhamento. De acordo com o artigo da [SUMMIT MOBILIDADE \(2023\)](#), podemos saber porque a causa da maioria dos acidentes estarem relacionados ao descarrilhamento. Segundo eles, as principais causas de acidentes de trem no Brasil, que geram descarrilamentos e atropelamentos, estão ligadas a estrutura precária da malha ferroviária e os problemas no material rodante.

## 2.9 MAPA FERROVIARIO

O mapa ferroviário tem um papel importante na pesquisa e análise de dados. Pesquisadores e analistas utilizam esses mapas para estudar padrões de tráfego, avaliar a eficácia de medidas de segurança e desenvolver modelos preditivos para prevenir acidentes ([ANTF, 2024](#)).

## 2.10 GOOGLE COLAB

O Google Colaboratory ou Google Colab, como é popularmente conhecido, é uma plataforma gratuita oferecida pelo Google. Ela fornece um ambiente de notebook Jupyter hospedado na nuvem que permite escrever e executar código Python. Além disso, dá para visualizar os resultados diretamente no navegador, sem a necessidade de configuração ou instalação local de software. (BORATINO, Carolina, 2024).

## 3 METODOLOGIA

---

### Resumo do capítulo

*Neste capítulo, vamos detalhar a metodologia e os passos que seguimos para analisar os dados dos acidentes ferroviários e chegar aos resultados finais. Essa abordagem inclui a coleta de dados, limpeza, análise exploratória, modelagem e visualização dos dados.*

### 3.1 VISÃO GERAL

No nosso projeto, começamos coletando dados sobre acidentes ferroviários, que continham informações detalhadas como data, local, gravidade e causas dos acidentes. Em seguida, limpamos os dados para tratar valores ausentes, corrigir tipos de dados e padronizar valores categóricos. Fizemos uma análise exploratória para entender a distribuição temporal, geográfica e a gravidade dos acidentes, gerando gráficos para facilitar a visualização.

Depois, aplicamos técnicas de aprendizado de máquina, como regressão linear para prever a tendência de acidentes ao longo dos anos e Random Forest para classificar a gravidade dos acidentes. Avaliamos a performance dos modelos com métricas de precisão, recall e f1-score.

Finalizamos com a criação de várias visualizações, incluindo boxplots e gráficos de barras, para comunicar os resultados de forma clara.

### 3.2 COLETA DE DADOS

A primeira etapa do nosso projeto foi a coleta dos dados sobre acidentes ferroviários. Utilizamos uma base de dados que contém informações detalhadas sobre cada acidente, como data, local, gravidade, causas, entre outros detalhes.

#### 3.2.1 Lista das colunas usadas para a coleta de dados

A Figura 3.1 mostra todas as colunas usadas no Dataframe.

```
▼ Listando as colunas do arquivo de dados

# Listando as colunas do arquivo de dados
df.columns

Index(['Gravidade', 'Concessionaria', 'Data_Ocorrencia', 'Hora_Ocorrencia',
      'UF', 'Municipio', 'Linha', 'Quilômetro_Inicial', 'Estação_Anterior',
      'Estação_Posterior', 'Perímetro_Urbano', 'Causa_direta', 'Natureza',
      'Interrupção', 'Prefixo', 'Serviço_Transporte', 'N_feridos', 'N_obitos',
      'Mercadoria', 'Equipagem', 'Outra_Ferrovia', 'PN'],
      dtype='object')
```

**Figura 3.1** – *Lista das colunas inclusas no Dataframe*

Fonte: Autor

### 3.2.2 Exemplo de saída de dados do Dataframe

A Figura 3.2 mostra a saída de dados do Dataframe.

**Figura 3.2** – *Exemplo de saída de dados do Dataframe.*

```
# 3 Amostras aleatórias da tabela
df.sample(3)
```

| Gravidade | Concessionaria | Data_Ocorrencia | Hora_Ocorrencia | UF | Municipio | Linha                |
|-----------|----------------|-----------------|-----------------|----|-----------|----------------------|
| Sim       | RMO            | 24/11/2009      | 06:24           | SP | Sorocaba  | Mairinque - Bauru    |
| Sim       | RMS            | 25/02/2013      | 23:30           | PR | Pinhais   | Paranaguá - Uvaranas |
| Sim       | FTL            | 04/04/2018      | 12:36           | MA | Codo      | Tronco São Luis      |

Fonte: Autor

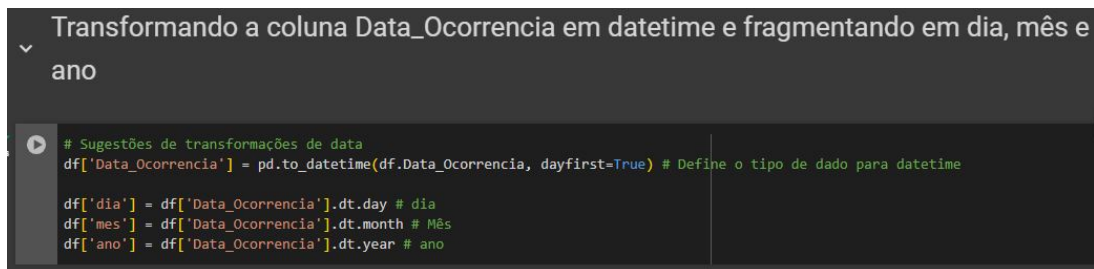
### 3.3 LIMPEZA DE DADOS

Na etapa de limpeza de dados, tratamos valores ausentes nas colunas críticas, corrigimos tipos de dados incorretos e padronizamos valores categóricos. Removemos duplicatas e garantimos a consistência dos dados para preparar a base para análises precisas.

#### 3.3.1 Exemplo de ajuste feito nos dados

A Figura 3.3 mostra o a separação do dia, mês e ano dos valores do tipo date na coluna do Dataframe.

**Figura 3.3** – *Exemplo de ajuste dos dados separando de dia, mês e ano na coluna de data dos acidentes.*



```
Transformando a coluna Data_Ocorrencia em datetime e fragmentando em dia, mês e ano

# Sugestões de transformações de data
df['Data_Ocorrencia'] = pd.to_datetime(df.Data_Ocorrencia, dayfirst=True) # Define o tipo de dado para datetime

df['dia'] = df['Data_Ocorrencia'].dt.day # dia
df['mes'] = df['Data_Ocorrencia'].dt.month # Mês
df['ano'] = df['Data_Ocorrencia'].dt.year # ano
```

Fonte: Autor

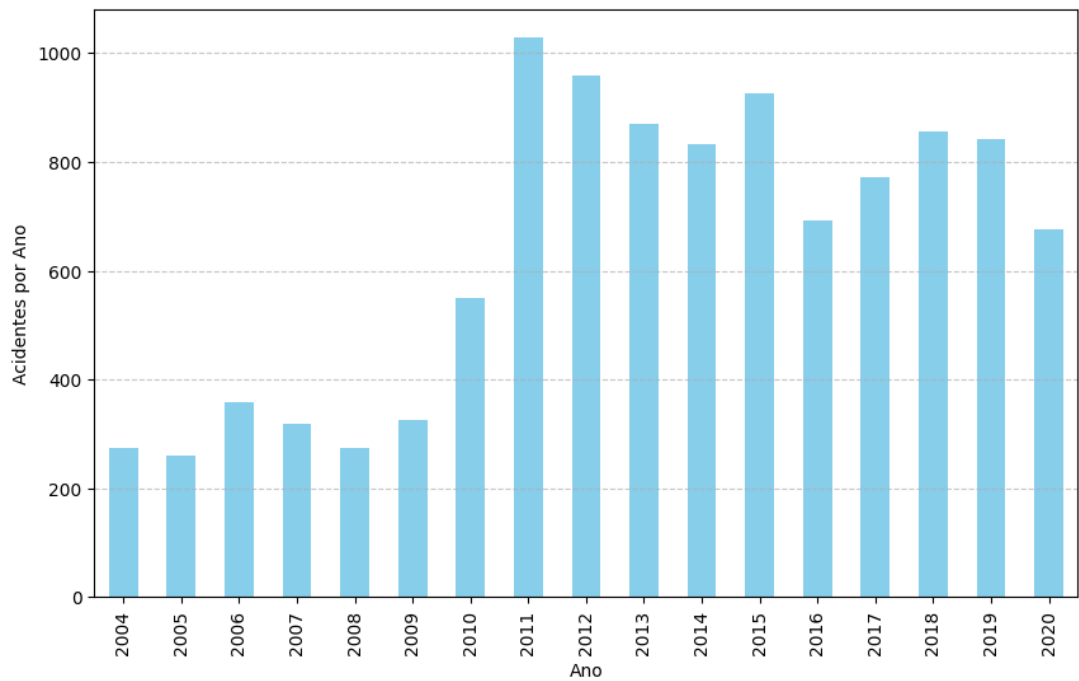
### 3.4 ANÁLISE EXPLORATÓRIA DE DADOS (EDA)

Com os dados limpos, fizemos uma Análise Exploratória de Dados (EDA) para entender melhor a distribuição e as características dos acidentes ferroviários.

#### 3.4.1 Distribuição Temporal

A Figura 3.4 mostra o gráfico de quantidade de acidentes registrados por ano.

**Figura 3.4** – *Gráfico de quantidade de acidentes registrados por ano em 2004.*



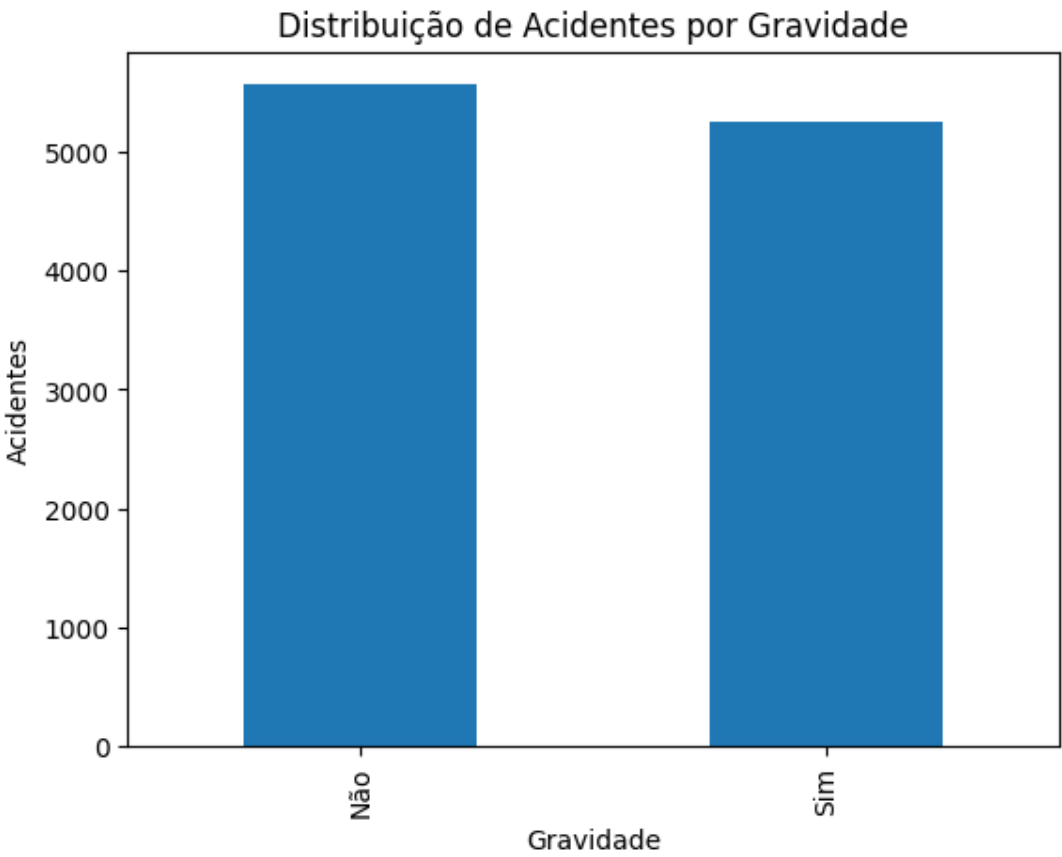
Fonte: Autor



3.4.2 Gravidade dos Acidentes

A Figura 3.5 mostra o gráfico comparando os acidentes graves dos não graves.

**Figura 3.5** – *Gráfico comparando os acidentes graves e não graves.*

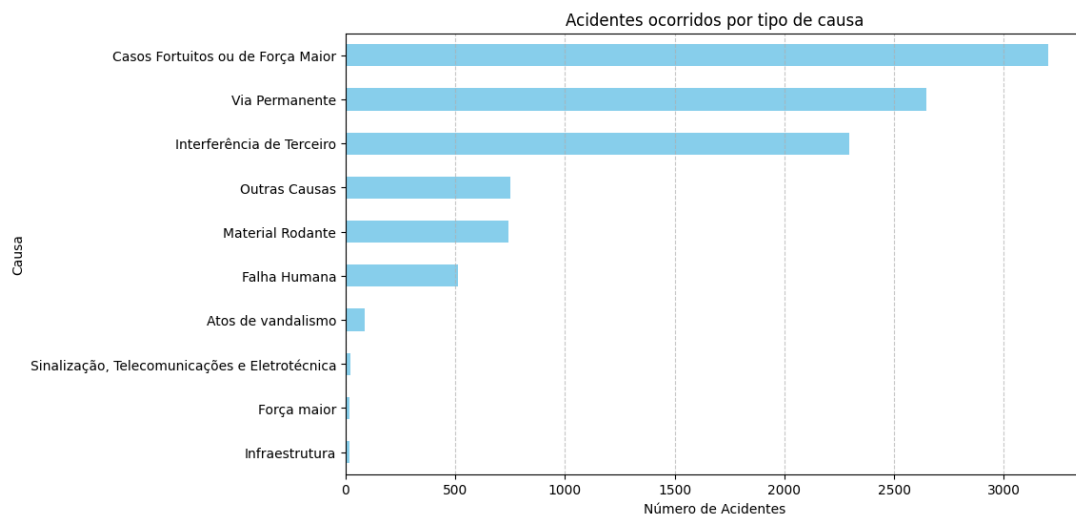


Fonte: Autor

### 3.4.3 Causas dos Acidentes

A Figura 3.6 mostra o gráfico com as causas diretas mais comuns entre os acidentes.

**Figura 3.6** – *Gráfico com as causas diretas entre os acidentes*



Fonte: Autor

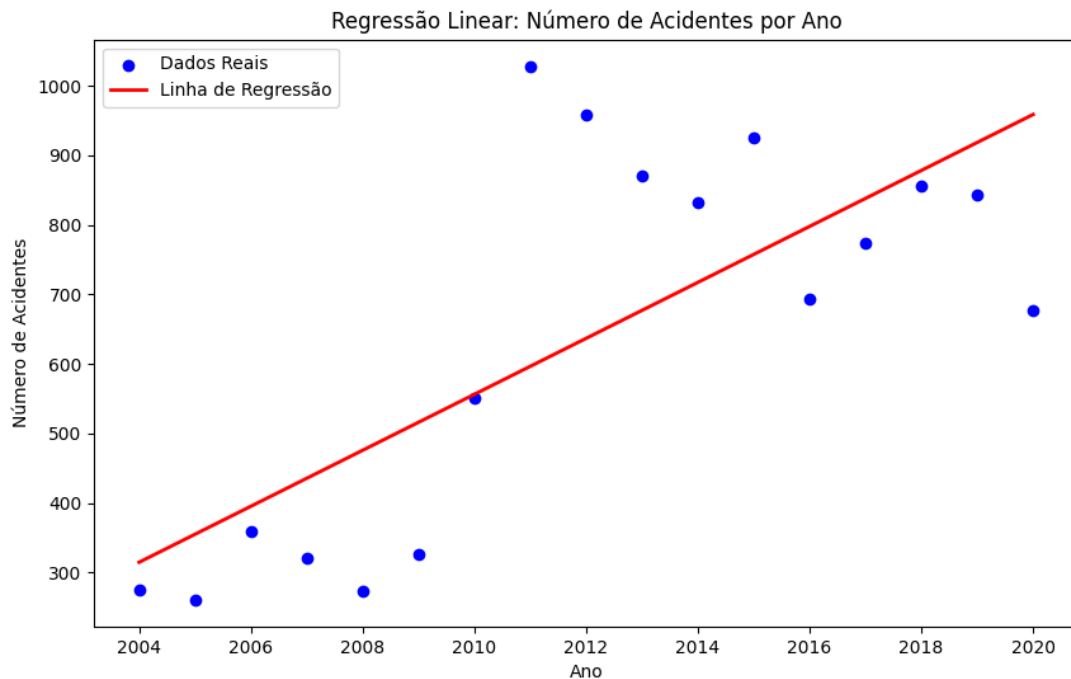
## 3.5 MODELAGEM

A etapa de modelagem incluiu a aplicação de técnicas de aprendizado de máquina para prever e classificar acidentes.

### 3.5.1 Regressão Linear

Utilizamos a regressão linear para prever o número de acidentes ao longo dos anos. Os resultados mostraram uma tendência de aumento dos acidentes.

A Figura 3.7 mostra o gráfico da regressão linear para prever e classificar acidentes.

**Figura 3.7** – *Gráfico da regressão linear*

Fonte: Autor

### 3.5.2 Modelo de Classificação

Usamos um modelo de Random Forest para classificar a gravidade dos acidentes. A performance do modelo foi avaliada usando métricas de precisão, recall e f1-score.

Trecho do código do Modelo de Classificação:

```

1 from sklearn.ensemble import RandomForestClassifier
2 from sklearn.metrics import classification_report
3
4 # Prepara o dos dados
5 X = df[['Equipagem', 'N_obitos', 'Natureza', 'Mercadoria', 'N_feridos',
6        'Servi o_Transporte', 'Per metro_Urbano', 'UF', 'Concessionaria', 'PN', 'Outra_Ferrovia']]
7
8 # Treinamento do modelo
9 modelo = RandomForestClassifier()
10 modelo.fit(X, y)
11
12 # Avalia o do modelo
13 y_pred = modelo.predict(X_test)
14 print(classification_report(y_test, y_pred))

```

**Code Listing 3.1:** *Código do Modelo de Classificação*

### 3.6 VISUALIZAÇÃO DOS RESULTADOS

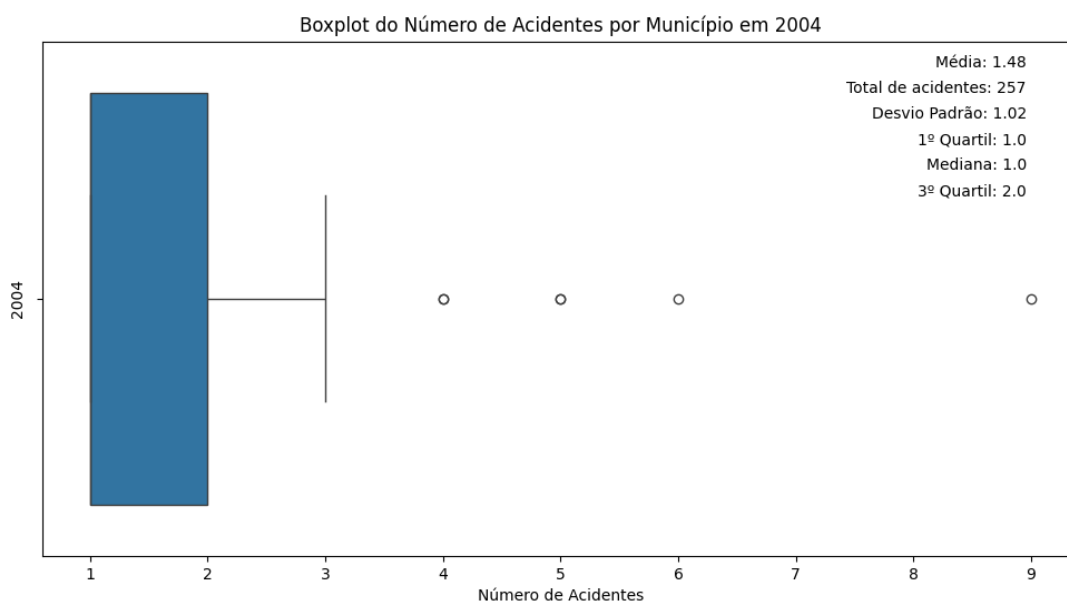
Para comunicar nossos achados de forma clara, criamos várias visualizações.

#### 3.6.1 Boxplots

Utilizamos boxplots para analisar a distribuição dos acidentes por ano e por município.

A Figura 3.8 mostra o gráfico boxplot da distribuição dos acidentes anuais por município em 2004.

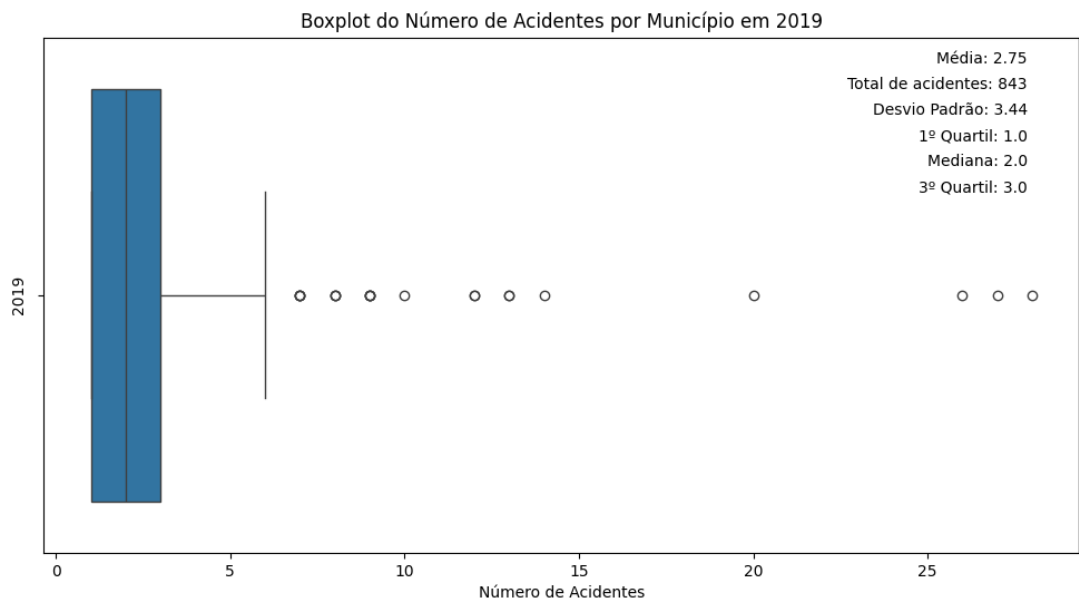
**Figura 3.8** – *Boxplot da distribuição dos acidentes anuais por município em 2004*



Fonte: Autor

A Figura 3.9 mostra o gráfico boxplot da distribuição dos acidentes anuais por município em 2019.

**Figura 3.9** – *Boxplot da distribuição dos acidentes anuais por município em 2019*



Fonte: Autor

## 4 ANÁLISE DOS RESULTADOS

---

### Resumo do capítulo

*Neste capítulo, vamos analisar os principais e mais relevantes resultados obtidos durante o projeto, focando nas descobertas e insights gerados a partir dos dados de acidentes ferroviários.*

### 4.1 ANÁLISE DESCRITIVA

#### 4.1.1 Distribuição dos Acidentes ao Longo dos Anos

Observamos um aumento constante no número de acidentes ferroviários ao longo dos anos. Este crescimento pode ser atribuído a diversos fatores, como aumento do tráfego ferroviário, deterioração da infraestrutura ou mudanças nos padrões de reporte de acidentes.

#### 4.1.2 Frequência dos acidentes por Município

Identificamos que alguns municípios registram um número significativamente maior de acidentes. Este padrão sugere que pode haver problemas específicos nessas localidades, como infraestrutura inadequada ou maior volume de tráfego ferroviário.

### 4.2 BOXPLOTS

#### 4.2.1 Número de Acidentes por Ano

A análise dos boxplots mostrou uma variação significativa no número de acidentes por ano. Alguns anos apresentaram picos de acidentes, indicando possíveis eventos atípicos ou períodos de maior vulnerabilidade.

#### 4.2.2 Número de Acidentes por Município em 2019

Para 2019, o boxplot revelou que a maioria dos municípios teve um número baixo de acidentes, mas alguns poucos registraram números extremamente altos. A mediana, média e os quartis forneceram uma visão detalhada da distribuição desses acidentes.

### 4.3 REGRESSÃO LINEAR

A regressão linear realizada para prever o número de acidentes com base nos anos revelou uma tendência crescente, com um coeficiente de inclinação positivo. Isso indica

que, se nenhuma medida significativa for tomada, o número de acidentes pode continuar aumentando nos próximos anos.

Coeficiente (Inclinação): 40.24

Intercepção: -80326.57

Esses valores sugerem que o número de acidentes aumenta, em média, cerca de 40 acidentes por ano. A intercepção negativa, embora menos intuitiva, é comum em modelos lineares e reflete a extrapolação do modelo para além do período analisado.

#### 4.4 GRÁFICO DE BARRAS HORIZONTAL

##### 4.4.1 Número de Acidentes por Tipo de Transporte do Trem

Este gráfico revelou que alguns tipos de transporte têm uma maior incidência de acidentes. Ao focar nos seis tipos de transporte mais comuns, conseguimos uma visualização mais clara e manejável, que facilita a identificação de padrões específicos.

## 5 CONCLUSÕES

---

A análise dos resultados confirma nossa hipótese inicial de que os acidentes ferroviários estão aumentando ao longo dos anos. A distribuição geográfica e por tipo de transporte aponta para áreas e categorias específicas que precisam de mais atenção. Os modelos preditivos e de classificação que desenvolvemos são ferramentas valiosas para antecipar e impossibilitar riscos. Com base nessas descobertas, recomendamos implementar medidas preventivas direcionadas, melhorar a infraestrutura ferroviária e monitorar continuamente para reduzir a ocorrência de acidentes no futuro.

Isso é essencial para garantir a segurança e eficiência das operações ferroviárias e proteger a vida de passageiros e trabalhadores do setor.



## REFERÊNCIAS BIBLIOGRÁFICAS

---

ANTF. *Mapa Ferroviário*. 2024. Disponível em: <<https://www.antf.org.br/mapa-ferroviario/>>. Acesso em: 28 abril 2024. Citado na pág. 11.

AWS. *O que é Python?* 2024. Disponível em: <<https://aws.amazon.com/pt/what-is/python/>>. Acesso em: 04 maio 2024. Citado na pág. 10.

BORATINO, Carolina. *Google Colab: o que é e como usar?* 2024. Disponível em: <<https://www.hostgator.com.br/blog/google-colab/>>. Acesso em: 04 maio 2024. Citado na pág. 12.

CBN Curitiba. *Curitiba lidera índice de acidentes por trem em todo o país, segundo ANTT*. 2023. Disponível em: <<https://cbncuritiba.com.br/materias/curitiba-lidera-indice-de-acidentes-por-trem-em-todo-o-pais-segundo-antt/>>. Acesso em: 28 abril 2024. Citado na pág. 11.

CORBO, Antony. *Is Python Good for Machine Learning?* 2023. Disponível em: <<https://builtin.com/machine-learning/python-machine-learning>>. Acesso em: 27 abril 2024. Citado na pág. 10.

COUTINHO, Thiago. *Descubra o que é e o porquê de tantos programadores utilizarem a Biblioteca Pandas*. 2021. Disponível em: <<https://voitto.com.br/blog/artigo/biblioteca-pandas>>. Acesso em: 27 abril 2024. Citado na pág. 10.

EBAC. *O que é Random Forest*. 2024. Disponível em: <<https://ebaonline.com.br/blog/random-forest-seo>>. Acesso em: 27 abril 2024. Citado na pág. 10.

IBM. *Métodos de aprendizado de máquina*. 2024. Disponível em: <<https://www.ibm.com/br-pt/topics/machine-learning>>. Acesso em: 27 abril 2024. Citado na pág. 11.

SCHÖNHOFEN, Ricardo D. *Biblioteca Seaborn com o matplotlib*. 2017. Disponível em: <<https://www.vooo.pro/insights/biblioteca-seaborn-com-o-matplotlib/>>. Acesso em: 27 abril 2024. Citado na pág. 11.

SUMMIT MOBILIDADE. *Principais causas de acidentes com trem*. 2023. Disponível em: <<https://summitmobilidade.estadao.com.br/guia-do-transporte-urbano/como-evitar-acidentes-de-trem/#:~:text=Conforme%20dados%20da%20ANTT%2C%20problemas,per%C3%ADodo%20de%202006%20a%202013.>> Acesso em: 28 abril 2024. Citado na pág. 11.

## ANEXOS

---

### A : TÍTULO

Segundo a ABNT (Associação Brasileira de Normas Técnicas), os anexos são documentos criados por terceiros, e usados pelo autor.

### Publicação

1. MIGUEL, D. A. A.; **Acidentes Ferroviários No Brasil: Análise Comparativa Com A União Europeia**. Anais do 34º Congresso de Pesquisa e Ensino em Transportes, ANPET, 2020.
2. JUNIOR, Luiz S.. **Entendendo a biblioteca NumPy**. Medium, [S.I.], 2018.Disponível em:<https://medium.com/ensina-ai/entendendo-a-biblioteca-numpy-4858fde63355>. Acesso em: 7 de abril 2024
3. MACHADO, Alysson. **Biblioteca Matplotlib do Python**. Medium, [S.I.], 2020.Disponível em:<https://medium.com/@alyssonmachado388/biblioteca-matplotlib-do-python-5e32104439d0>. Acesso em: 7 de abril 2024
4. SCHÖNHOFEN, Ricardo D.. **Biblioteca Seaborn com o matplotlib**. Vooo, [S.I.],2017.Disponível em:<https://www.vooo.pro/insights/biblioteca-seaborn-com-o-matplotlib/>. Acesso em: 7 de abril 2024
5. COUTINHO, Thiago. **O que é a biblioteca Pandas?**. Voitto,[S.I.], 2021.Disponível em:<https://www.voitto.com.br/blog/artigo/biblioteca-pandas>. Acesso em: 7 de abril 2024