



UNIVERSIDADE FERNANDO PESSOA
FACULDADE DE CIÊNCIAS E TECNOLOGIA

Projeto Análise de Dados

Análise Integrada de Dados de Acidentes e
Condições Climáticas nos EUA (2021)

Aluno

Mário Pinto 2022121295

Professor

João Viana

Ano letivo 2024/2025

Índice

1	Introdução	2
2	Objetivo da Análise	2
3	Importação e Integração dos Dados	2
4	Tratamento de Dados Ausentes e Outliers	3
5	Análise Exploratória (EDA)	3
6	Outros Resultados Relevantes	5
7	Discussão dos Resultados	5
8	Conclusão	6

1 Introdução

A ocorrência de acidentes rodoviários representa uma das principais causas de lesões e mortes em diversos países, incluindo os Estados Unidos. Entre os múltiplos fatores que influenciam esses incidentes, as condições climáticas desempenham um papel determinante, afetando tanto o comportamento dos condutores quanto a segurança das infraestruturas viárias.

Este projeto tem como objetivo analisar a relação entre variáveis meteorológicas e a severidade dos acidentes rodoviários ocorridos nos EUA durante o ano de 2021. Através da integração de dois conjuntos de dados — *US Accidents* e *Weather Events* — procura-se identificar padrões e correlações significativas que possam contribuir para uma melhor compreensão do impacto das condições atmosféricas nos acidentes.

A análise contempla diferentes dimensões climáticas, como temperatura, precipitação, visibilidade e eventos extremos (chuva intensa, nevoeiro, gelo, entre outros), cruzando esses dados com a gravidade dos acidentes registados. Esta abordagem permite uma avaliação abrangente e fundamentada, com o potencial de apoiar o desenvolvimento de estratégias preditivas e preventivas para a gestão da segurança rodoviária em condições adversas.

2 Objetivo da Análise

Avaliar se há padrões significativos entre eventos climáticos (chuva, neve, visibilidade reduzida, etc.) e a severidade dos acidentes. Os principais objetivos são:

- Identificar correlações entre clima e severidade;
- Detectar padrões em variáveis como temperatura, humidade e visibilidade;
- Avaliar o impacto de diferentes tipos de eventos meteorológicos.

3 Importação e Integração dos Dados

Fontes de Dados

- US Accidents (Kaggle): `US_Accidents_March23.csv`
- Weather Events (Kaggle): `WeatherEvents_Jan2016-Dec2022.csv`

Processamento

1. Filtragem para o ano de 2021;
2. Conversão de datas e extração da hora arredondada;
3. Merge com `merge_asof()` por ZIP e hora;
4. Criação de `df_merged`.

4 Tratamento de Dados Ausentes e Outliers

- Colunas com mais de 50% de valores nulos foram removidas;
- Variáveis numéricas foram imputadas com a mediana, categóricas com o modo;
- Outliers em `Distance(mi)` removidos com base no percentil 99;
- A variável `Severity` foi convertida para inteiro e filtrada para valores 1–4.

5 Análise Exploratória (EDA)

Distribuições das Variáveis Climáticas

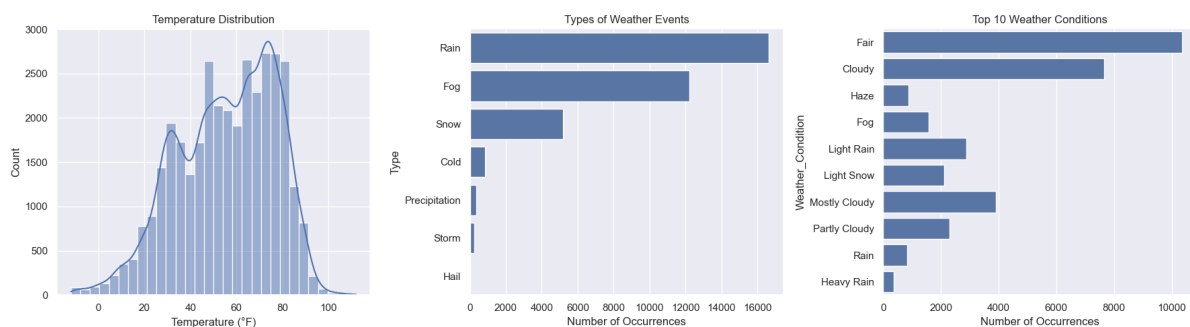


Figura 1: Distribuições das Variáveis Climáticas em 2021

A distribuição da temperatura mostra uma concentração entre os 40°F e 70°F, intervalo onde ocorreram mais acidentes. Nos eventos climáticos, destacam-se *Rain*, *Fog* e *Snow* como os mais frequentes. Em relação às condições meteorológicas reportadas, *Fair* e *Cloudy* são as mais comuns, seguidas de *Light Rain* e *Mostly Cloudy*, indicando que muitos acidentes ocorrem mesmo em tempo aparentemente estável.

Frequência de Acidentes por Tipo de Evento

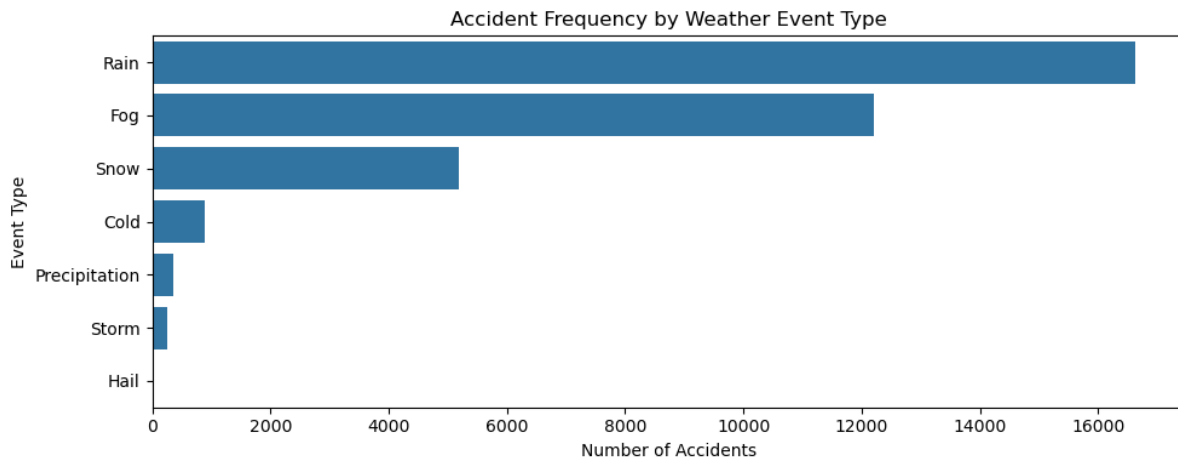


Figura 2: Frequência de Acidentes por Tipo de Evento Climático

Tipos de eventos como *Rain* (chuva), *Snow* (neve) e *Fog* (nevoeiro) destacam-se como os mais frequentemente associados a acidentes, indicando elevada exposição combinada com risco inerente à condição atmosférica.

Severidade Média por Tipo de Evento

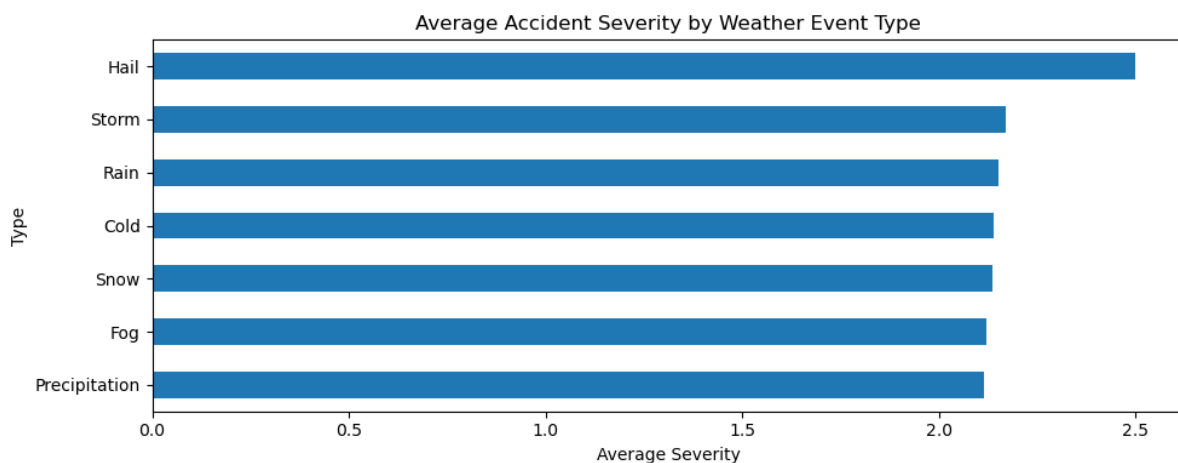


Figura 3: Severidade Média dos Acidentes por Tipo de Evento Climático

Eventos como *Fog*, *Ice* e *Freezing Rain* apresentam severidade média mais elevada, sugerindo maior gravidade associada a estas condições em comparação com eventos mais comuns como *Rain* e *Snow*.

6 Outros Resultados Relevantes

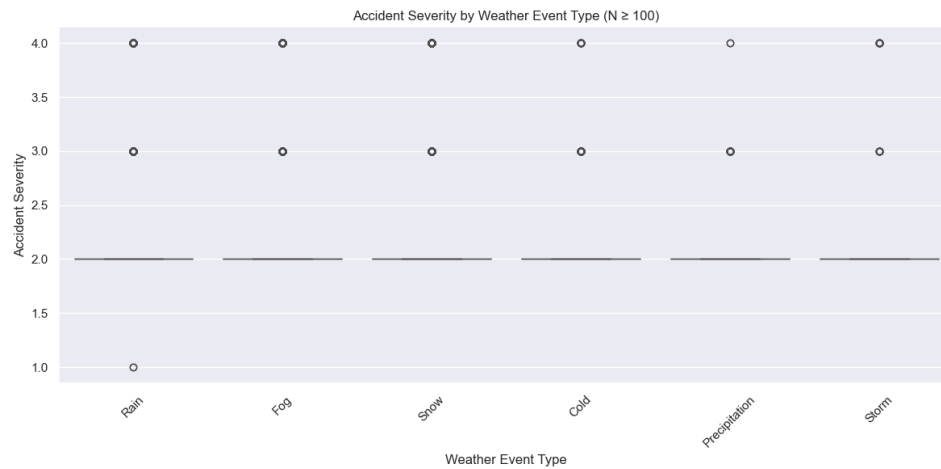


Figura 4: Distribuição da Severidade por Tipo de Evento Climático ($N \geq 100$)

A figura apresenta a distribuição da severidade dos acidentes em diferentes tipos de eventos climáticos com pelo menos 100 ocorrências. Observa-se que a maioria dos acidentes ocorreu com severidade 2, independentemente do tipo de evento. No entanto, eventos como *Snow*, *Fog* e *Storm* evidenciam maior dispersão, indicando ocorrências mais graves (níveis 3 e 4) com alguma frequência. Já condições como *Rain* e *Precipitation* apresentam uma distribuição mais concentrada, predominantemente na severidade moderada. Estes resultados sugerem que certos eventos extremos têm potencial de elevar a gravidade dos acidentes, mesmo sendo menos frequentes.

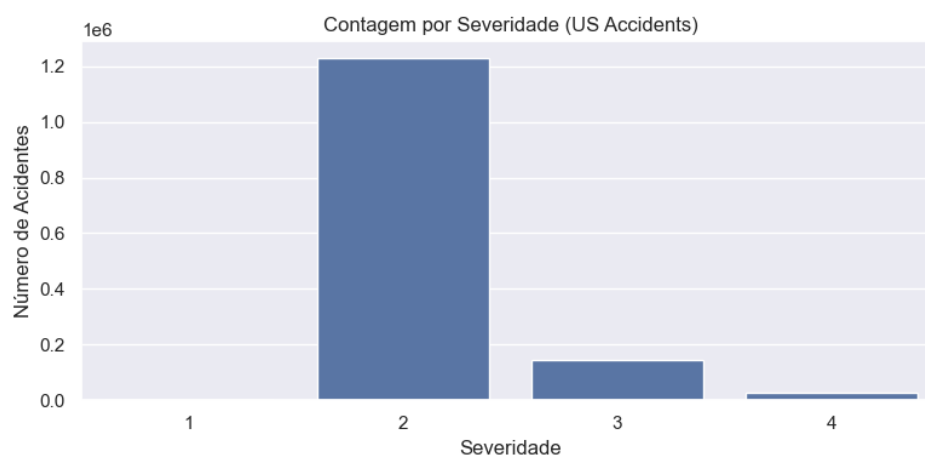


Figura 5: Contagem Total por Severidade dos Acidentes (1–4)

7 Discussão dos Resultados

- A maioria dos acidentes ocorreu durante eventos meteorológicos adversos, com destaque para chuva (*Rain*), neve (*Snow*) e nevoeiro (*Fog*). Estes tipos de eventos

apresentaram elevada frequência e, no caso do nevoeiro, severidade média significativamente superior;

- A visibilidade revelou-se uma das variáveis mais relevantes, apresentando uma correlação negativa moderada com a severidade dos acidentes ($r \approx -0,22$). Acidentes com baixa visibilidade tendem a ser mais graves, conforme evidenciado nos boxplots;
- Condições térmicas extremas, tanto em regiões de frio intenso quanto de calor excessivo, estão associadas a uma maior variabilidade na severidade dos acidentes, ainda que com correlação fraca;
- A severidade 1 representa apenas cerca de 1% dos casos reportados, enquanto as categorias 2 e 3 concentram a maioria dos acidentes, indicando uma predominância de acidentes de gravidade moderada;
- Eventos como *Ice* e *Freezing Rain*, embora menos frequentes, mostraram-se particularmente críticos em termos de severidade, justificando maior atenção preventiva por parte das autoridades de tráfego.

8 Conclusão

A análise realizada demonstra que a integração de dados meteorológicos com registos de acidentes rodoviários permite identificar relações significativas entre condições atmosféricas e a severidade dos acidentes. Evidenciou-se que eventos como chuva, neve e nevoeiro estão fortemente associados a uma maior frequência e gravidade de acidentes. Destaca-se a visibilidade como a variável com maior correlação negativa com a severidade, o que indica que acidentes sob baixa visibilidade tendem a ser mais graves. Outros fatores como a temperatura e a precipitação também apresentaram impacto relevante, especialmente sob condições extremas.

Esta abordagem integrada fornece uma base sólida para o desenvolvimento de sistemas de alerta e modelos preditivos baseados em condições meteorológicas em tempo real. A análise pode ser facilmente adaptada para contextos regionais, ajudando autoridades e operadores a mitigar riscos em ambientes rodoviários.