

TEMPLATE PARA ENTREGA DO PROJETO DA DISCIPLINA
Ciência de Dados e Inteligência Artificial
Fase 1

Nome do estudante: Dionatan Henrique Santos de Castro

Síntese do conjunto de dados como apresentado na sua origem:

Apresente uma breve descrição, acerca do conjunto de dados escolhido, indicando as informações contidas nos dados selecionados. É importante, ao apresentar sua síntese, que você inclua elementos que permitam compreender o significado e a aplicação dos dados escolhidos. (Mínimo de 2 parágrafos)

O conjunto de dados selecionado refere-se aos **acidentes de trânsito registrados no município de Porto Alegre**, contendo informações geográficas, temporais e sobre a gravidade dos eventos. Entre os diversos atributos disponíveis, foram escolhidas dez variáveis de interesse: o **tipo de acidente (tipo_acid)**, a **região da cidade (regiao)**, o **dia da semana (dia_sem)**, além dos indicadores de consequências, como **feridos**, **feridos graves (feridos_gr)** e **mortes**. Também foram incluídas as coordenadas geográficas **longitude** e **latitude**, que permitem a espacialização dos acidentes, a **data** de ocorrência e a variável **UPS (Unidade Padrão de Severidade)**, que representa a gravidade do evento e será utilizada como atributo alvo na análise.

Essas informações permitem não apenas compreender o **contexto de cada acidente**, mas também realizar análises estatísticas e geográficas que auxiliem na identificação de padrões relevantes. Por exemplo, é possível verificar **quais tipos de acidentes geram maior severidade**, se existem **regiões ou dias da semana mais críticos**, ou ainda analisar a distribuição espacial para subsidiar políticas públicas de segurança viária. Assim, o conjunto de dados escolhido possibilita uma visão ampla sobre os fatores associados à ocorrência e à gravidade dos acidentes, fornecendo subsídios importantes para a **prevenção e redução de vítimas no trânsito**.

Quantidade de linhas e colunas do conjunto de dados:

Descreva apenas a quantidade de linhas e de colunas selecionadas.

O conjunto de dados contém 6.690 linhas e 10 colunas.

Escolha pelo menos 10 colunas totalmente preenchidas (as mais importantes) e, para cada coluna (inclusive para a coluna alvo):

- Qual o nome e o que representa?
- Qual o tipo de dados? Nominal/Ordinal/Numérico/Data e/ou hora?
- Quais são os valores considerados válidos?
- Quantos valores distintos aparecem na coluna?
- Qual o menor e o maior valor, e qual a moda?
- Os valores da coluna são numéricos? Qual a média e qual o desvio-padrão? Qual a mediana?

Você pode apresentar essas informações da maneira como entender mais conveniente, como por exemplo em uma tabela, parágrafos, etc. É importante que sejam inseridas as imagens e as evidências do passo a passo realizado na atividade.

1. data

O que representa: Data do acidente

Tipo de dados: Data

Valores válidos: Formato DD/MM/AAAA (ex.: 01/01/2010 a 31/12/2025)

Valores distintos: 2034

Menor/Maior/Moda: Menor: 01/01/2025; Maior: 27/07/2025; Moda: 28/05/2025

2. tipo_acid

O que representa: Tipo de acidente

Tipo de dados: Nominal

Valores válidos: Categorias (ex.: "Colisão", "Atropelamento")

Valores distintos: 10

Menor/Maior/Moda: Moda: COLISÃO

3. regioao

O que representa: Área da cidade

Tipo de dados: Nominal

Valores válidos: Região (ex.: Centro)

Valores distintos: 4

Menor/Maior/Moda: Moda: Leste

4. dia_sem

O que representa: Dia da semana

Tipo de dados: Nominal

Valores válidos: Segunda a Domingo

Valores distintos: 7

Menor/Maior/Moda: Moda: Quarta-feira

5. feridos

O que representa: Feridos de acidente

Tipo de dados: Numérico

Valores válidos: ≥ 0

Valores distintos: 11

Menor/Maior/Moda: Menor: 0; Maior: 7; Moda: 0

Média/Desvio/Mediana: Média: 0,41; Desvio-padrão: 1.53; Mediana: 0

6. feridos_gr

O que representa: Feridos graves

Tipo de dados: Numérico

Valores válidos: ≥ 0

Valores distintos: Menor: 0; Maior: 5; Moda: 0

Menor/Maior/Moda: Média: 0,08; Desvio-padrão: 3.80; Mediana: 0

7. mortes

O que representa: Vítimas fatais

Tipo de dados: Numérico

Valores válidos: ≥ 0

Valores distintos: 4

Menor/Maior/Moda: Menor: 0; Maior: 1; Moda: 0

Média/Desvio/Mediana: Média: 0, Desvio-padrão: 17.28; Mediana: 0

8. latitude

O que representa: Coordenada geográfica

Tipo de dados: Numérico

Valores válidos: > -30.2426

Valores distintos: 44.235

Menor/Maior/Moda: Menor: -30.2426; Maior: -29.9604; Moda: -30.0238

Média/Desvio/Mediana: Média: -30.0476; Desvio-padrão: -0.00144129; Mediana: -30.0384

9. longitude

O que representa: Coordenada geográfica

Tipo de dados: Numérico

Valores válidos: > -51.2597

Valores distintos: 42.740

Menor/Maior/Moda: Menor: -51.2597; Maior: -51.0449; Moda: -51.219

Média/Desvio/Mediana: Média: -51.183; Desvio-padrão: -0.000744379; Mediana: -51.1902

10. Ups

O que representa: Unidade padrão de severidade

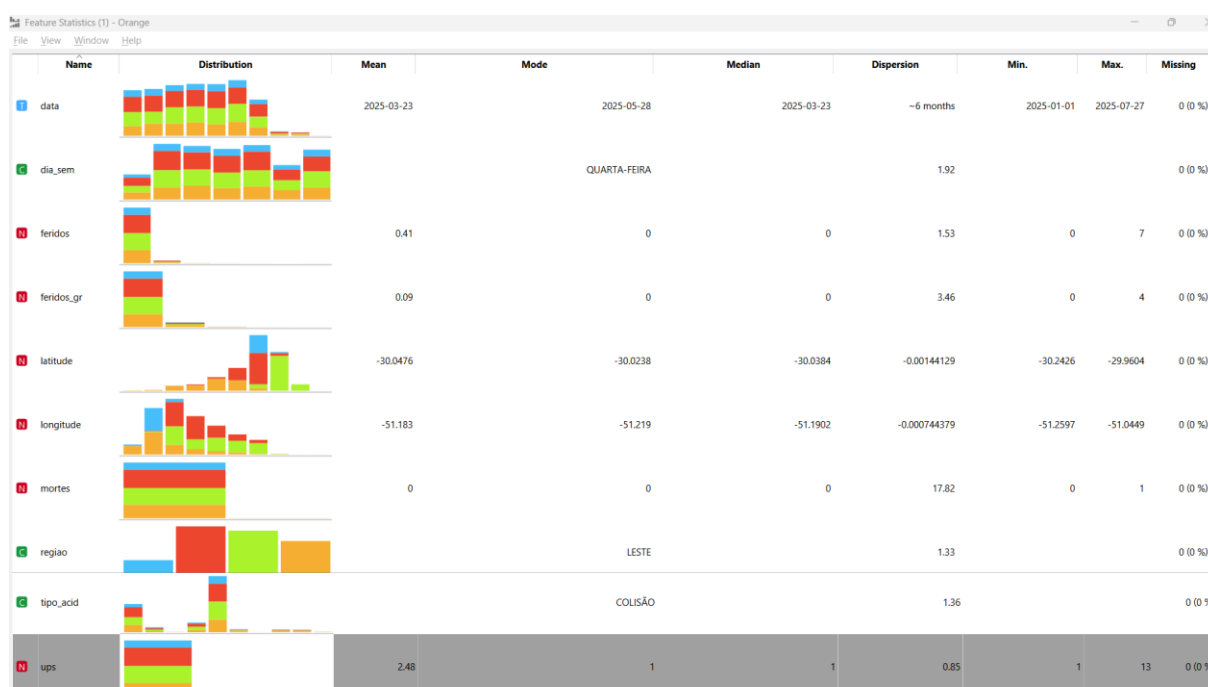
Tipo de dados: Numérico

Valores válidos: > 0

Valores distintos: 3

Menor/Maior/Moda: Menor: 1; Maior: 13; Moda: 1

Média/Desvio/Mediana: Média: 2,48; Desvio-padrão: 0,85; Mediana: 1



Descrição da coluna alvo e potenciais aplicações:

Informe qual a coluna alvo escolhida e descreva possíveis aplicações usando aprendizado de máquina que possam utilizá-la.

A coluna alvo escolhida é a **UPS (Unidade Padrão de Severidade)**, que representa a gravidade dos acidentes por meio de um peso atribuído conforme os danos causados.

Do ponto de vista acadêmico, esse atributo pode ser utilizado em modelos de aprendizado de máquina para:

- **Classificação ou regressão da severidade** dos acidentes, a partir de variáveis como horário, local, tipo de veículo e condições do evento;
- **Análise preditiva de risco**, permitindo identificar padrões associados a acidentes mais graves;
- **Subsídio a políticas públicas**, apoiando a priorização de intervenções viárias e campanhas educativas;
- **Gestão da mobilidade urbana**, com foco na redução de acidentes fatais e graves.

Assim, a utilização da UPS como variável alvo possibilita transformar dados históricos em conhecimento aplicado à prevenção e mitigação de acidentes de trânsito.

Link para acesso aos dados selecionados:

Inclua aqui o link para o conjunto de dados escolhido. No caso de ter sido feito um recorte/amostragem do conjunto original, documentar quais foram os passos feitos para produzir o recorte.

ATENÇÃO: O conjunto de dados do link deve estar PRONTO para ser carregado no Orange.

Link para o conjunto de dados escolhido:

<https://dadosabertos.poa.br/dataset/acidentes-de-transito-acidentes>

Para chegar ao recorte do conjunto de dados, foram consideradas apenas as datas disponíveis do ano de 2025 (01/01/2025 - 27/07/2025), este filtro foi feito através do widget Select Rows. Além disso, foram descartadas as linhas vazias de **todas as colunas** através do widget Preprocess.