



A base de dados Titanic

## Sobre a base Titanic

O Titanic foi um navio que naufragou e ficou marcado por ter sido uma tragédia que vitimou muitas pessoas. Os dados sobre o naufrágio foram registrados e alguns deles foram agrupados dando origem a uma base de dados para ser utilizada em diversas aplicações, dentre elas, a mineração de dados.

Na base constam informações sobre sobreviventes, mortos e outras informações importantes.

## Questionamentos interessantes para responder através de análises

- Passageiros de quais classes tiveram maior probabilidade de sobreviver?
- Qual foi a probabilidade de sobrevivência por sexo?
- Quais faixas etárias tiveram mais probabilidade sobreviver?

## Atributos

Atributos presentes na base:

- Survived - sobrevivente: 0 = Não, 1 = Sim
- Pclass - classe: 1 = 1ª, 2 = 2ª, 3 = 3ª
- Sex - sexo : male, female
- Age - idade: idade em valor numérico
- SibSp: Quantidade de irmãos e cônjuges a bordo
- Parch: Quantidade de pais e crianças a bordo
- Ticket - bilhete: Número do bilhete
- Fare - tarifa: Valor pago pelo bilhete
- Cabin - cabine: Número da cabine
- Embarked - porto embarque: C = Cherbourg, Q = Queenstown, S = Southampton

## Avaliação de atributos

Dependendo dos objetivos da análise, quais perguntas deverão ser respondidas, alguns dados podem não ser interessantes.

Deve ser realizada uma avaliação prévia das bases antes de começar a mineração e retirar os atributos que não irão contribuir com os objetivos estabelecidos.

## Avaliação de atributos

Para responder:

**“Passageiros de quais classes tiveram maior probabilidade de sobreviver?”**

Os atributos que podem não ser interessantes:

- Sibsp - Quantidade de irmãos e cônjuges a bordo;
- Parch - Quantidade de pais e crianças a bordo;
- Ticket - Bilhete: Número do bilhete;
- Cabin - Cabine: Número da cabine;
- Embarked - embarque: Local de embarque.

## Outras características

A base conta com valores NaN, que são atributos que não foram devidamente preenchidos. Essas situações devem ser tratadas antes dos dados começarem a ser minerados, pois eles podem afetar os resultados que serão obtidos ao final das análises.



Obrigada!

Ana Laurentino