

Curso de Data Science



Aula 07 - Análise de dados

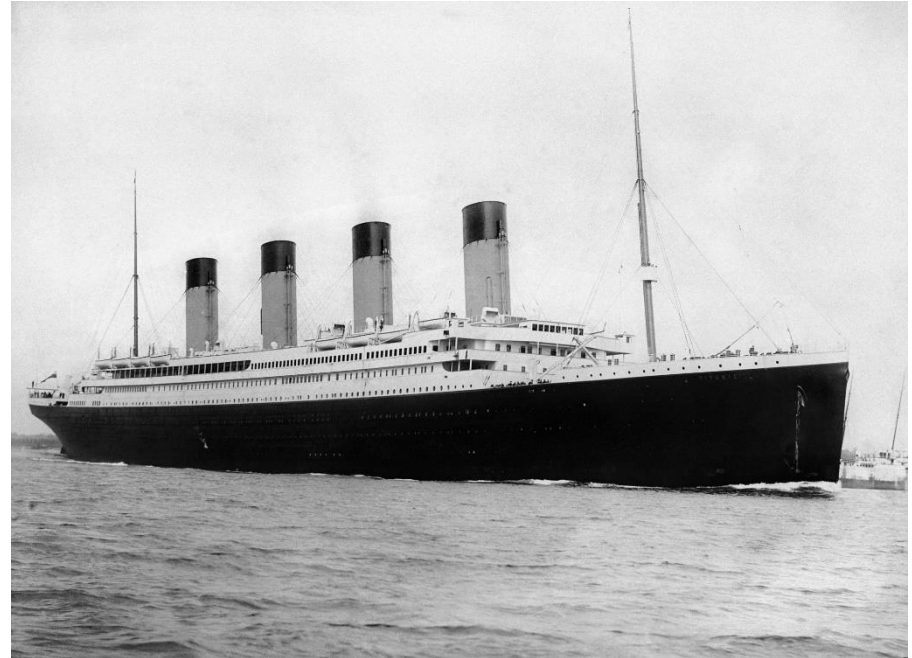
O que você irá aprender nesta aula?

- Entendimento do problema
- Obtenção de dados

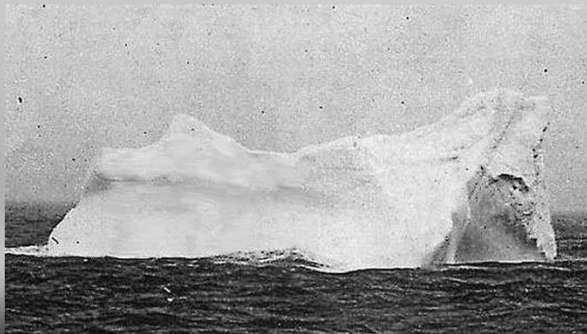
Sobre o problema

Titanic

- Navio Britânico
- Afundou nos Oceano Atlântico Norte em 1912
- Colidiu com um Iceberg
- De aproximadamente 2224 passageiros e tripulantes, mais de 1500 morreram
- Até hoje, é o único registro de navio naufragado após colisão com Iceberg



Sobre o problema



Sobre o problema

Queremos uma base de dados com informações dos passageiros, para que possamos analisar quais fatores podem contribuir com o fato do passageiro sobreviver ou não ao naufrágio.

Obtenção dos dados

- Usaremos a base de dados do Titanic disponível na plataforma online Kaggle.
- Ela oferece bancos de dados para aplicações de Ciência de dados e Aprendizado de máquina, além de organizar competições nessas áreas

The logo for Kaggle, featuring the word "kaggle" in a lowercase, blue, sans-serif font.

Obtenção dos dados

Arquivos dentro do dataset:

- `train.csv`: Deve ser usado para treinar o modelo de Machine Learning. Neste conjunto de dados, é informado se o passageiro sobreviveu ou não
- `test.csv`: Deve ser usado para testar o modelo treinado. A performance do modelo é testada a partir da predição feita nesse conjunto de teste, portanto não é informado se o passageiro sobreviveu ou não
- `gender_submission.csv`: modelo (gabarito) de como as previsões devem ser enviadas para a plataforma do Kaggle, em caso de participação na competição

Análise exploratória dos dados1

PassengerId: Número de identificação
do passageiro

Survived: Informa se o passageiro
sobreviveu ao desastre

0 = Não

1 = Sim

Pclass: Classe do bilhete

1 = 1ª Classe

2 = 2ª Classe

3 = 3ª Classe

Name: Nome do passageiro

Sex: Sexo do passageiro

Age: Idade do passageiro

SibSp: Quantidade de cônjuges e
irmãos a bordo

Parch: Quantidade de pais e filhos
a bordo

Ticket: Número da passagem

Fare: Preço da Passagem

Cabin: Número da cabine do passageiro

Embarked: Porto no qual o passageiro
embarcou

C = Cherbourg

Q = Queenstown

S = Southampton

Então, nesta aula vimos:

- Entendimento do problema
- Obtenção de dados

Muito obrigado!