**titanic**

**Introdução**

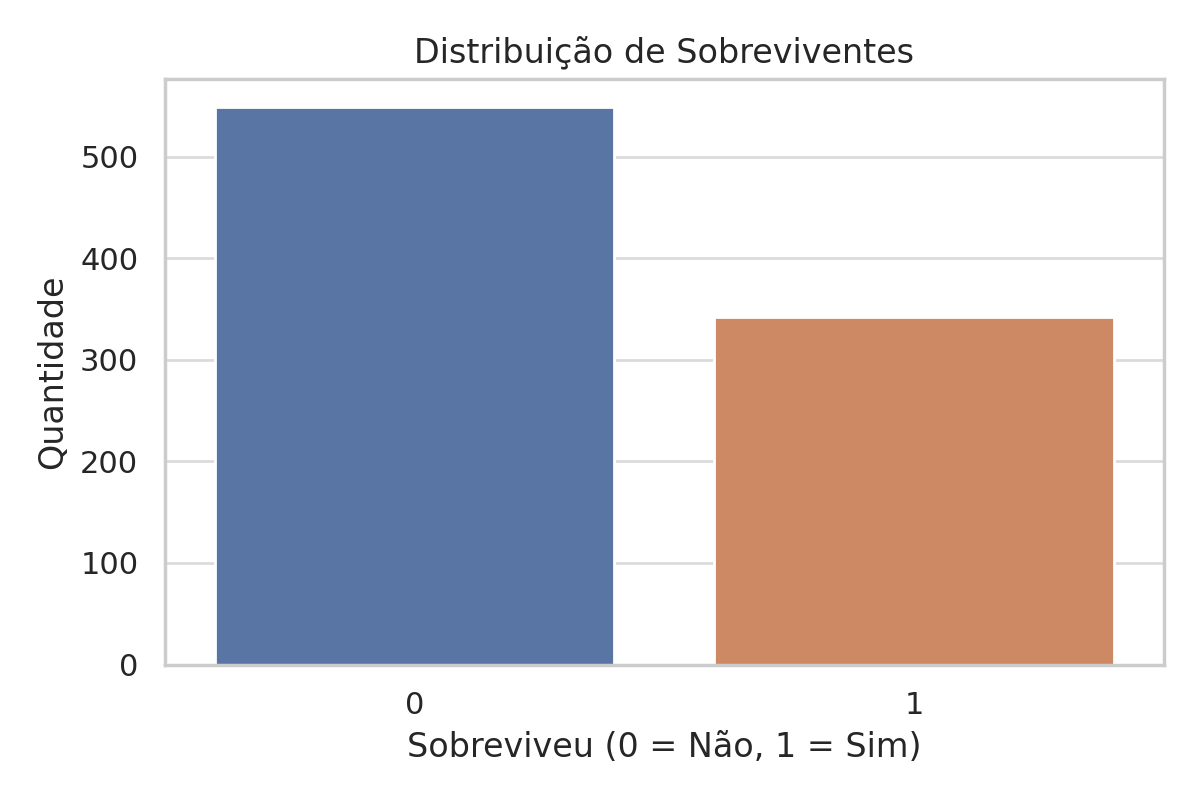
Este relatório tem como objetivo mostrar como resolver o desafio do Titanic na plataforma Kaggle. Este desafio, bem conhecido no campo da ciência de dados, visa prever quais passageiros do Titanic sobreviverão usando informações como idade, gênero, classe social, etc. Por ser um caso clássico e educativo, é amplamente utilizado para praticar análise e processamento de dados e a criação de modelos preditivos.

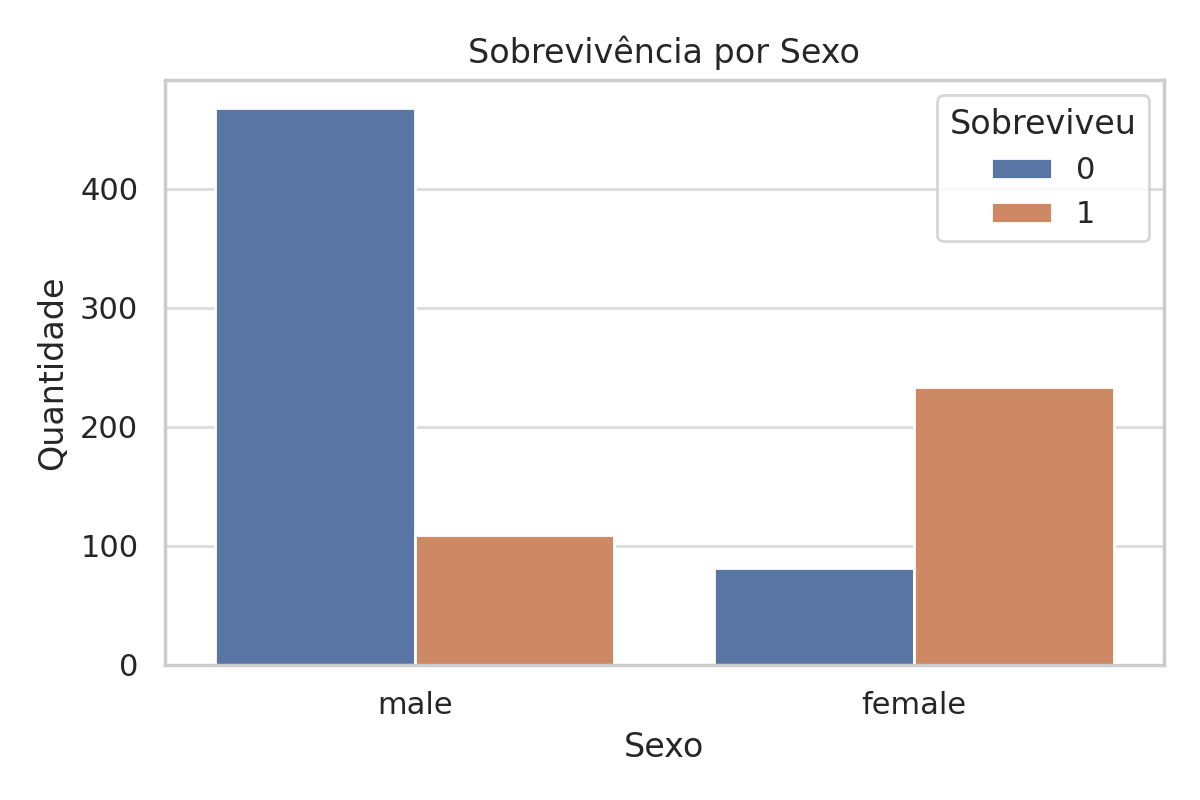
**Metodologia**

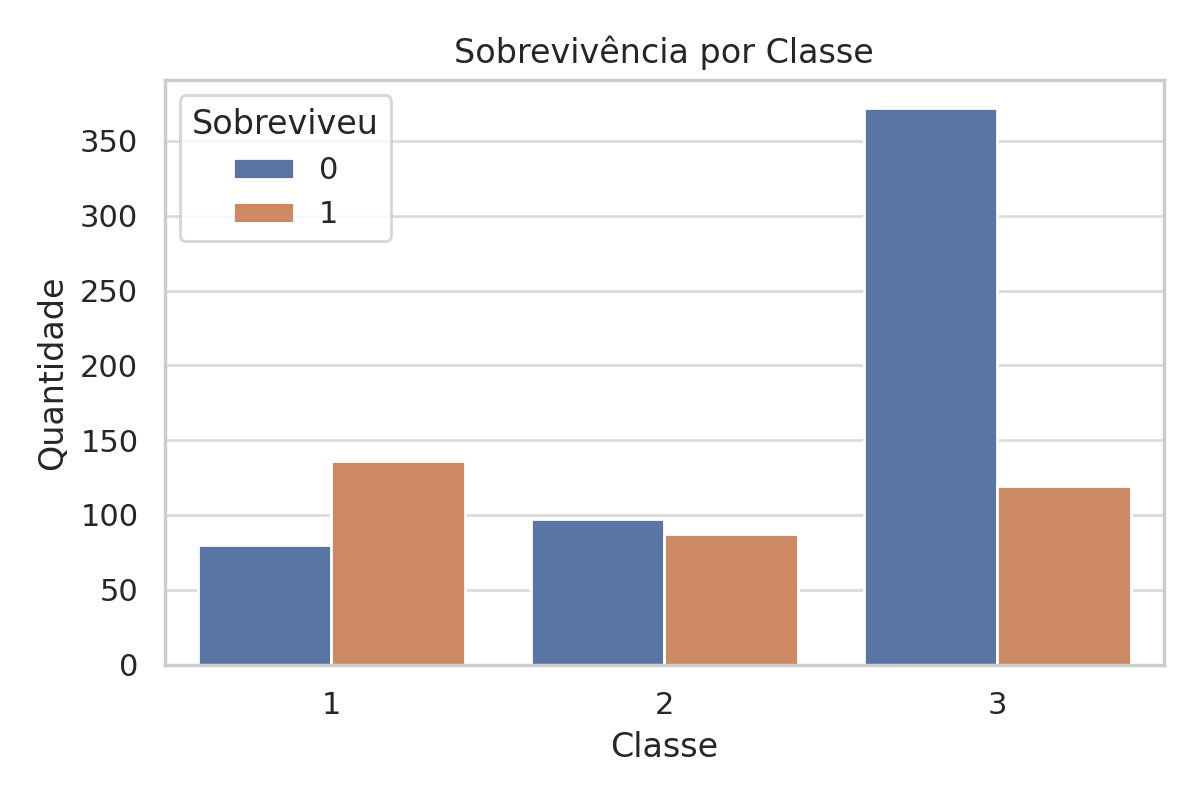
1. A metodologia utilizada incluiu as seguintes etapas:
   * Análise Exploratória de Dados (AED) para compreender as principais variáveis ​​associadas à sobrevivência.
   * Tratamento de dados ausentes: as variáveis ​​numéricas foram imputadas usando a média e as variáveis ​​categóricas foram imputadas usando a moda.
   * As variáveis ​​categóricas foram codificadas usando a Codificação One-Hot.
   * As variáveis ​​numéricas foram normalizadas usando o StandardScaler.
   * Os modelos Random Forest foram treinados usando a validação cruzada de 5 vezes.
   * O desempenho do modelo foi avaliado com base na acurácia.

**Resultados gerados**

Gráficos gerados a partir da Análise Exploratória de Dados, conforme o notebook original:







A acurácia média obtida com validação cruzada de 5 folds foi de 0.8092, com valores individuais: 0.7821, 0.8090, 0.8483, 0.7809, 0.8258.

**Discurssão**

Durante o desafio, encontramos alguns desafios, como valores ausentes em colunas como Idade, Classe e Data de Partida. Além disso, não exploramos todas as variáveis ​​disponíveis, como nomes de passageiros, que poderiam ter sido processadas para extrair títulos (por exemplo, Sr., Sra., etc.).

O desempenho do modelo de floresta aleatória foi satisfatório, mas as áreas que precisam ser aprimoradas incluem:

- Testar outros algoritmos, como aumento de gradiente ou XGBoost;

- Executar ajuste de hiperparâmetros;

- Criar novos recursos a partir de recursos existentes

.**Conclusão**

Durante o processo, foi possível explorar os dados, tratá-los e aplicar um modelo preditivo para tentar prever quem sobreviveu ao naufrágio. O modelo Random Forest teve um bom desempenho, com mais de 80% de acurácia, o que mostra que, mesmo com dados simples, é possível construir soluções eficazes. Além disso, participar da competição no Kaggle ajudou a entender melhor como funciona um projeto aplicado na prática e em um ambiente de desafio real.