

PROJETO TITANIC

Nesta apresentação vamos explorar o projeto Titanic e como desenvolver um modelo de machine learning para prever sobreviventes. Prepare-se para se aprofundar em análise de dados e modelagem.

Gabriel Resende Meireles



Entendendo os Dados

01 Oferta de Dados

Exploração do conjunto de dados e definição de quais variáveis são relevantes para o modelo.

02 Criando Novas Variáveis

Criação de novas variáveis com base nas informações existentes para tornar os modelos mais precisos.

03 Checagem da Qualidade

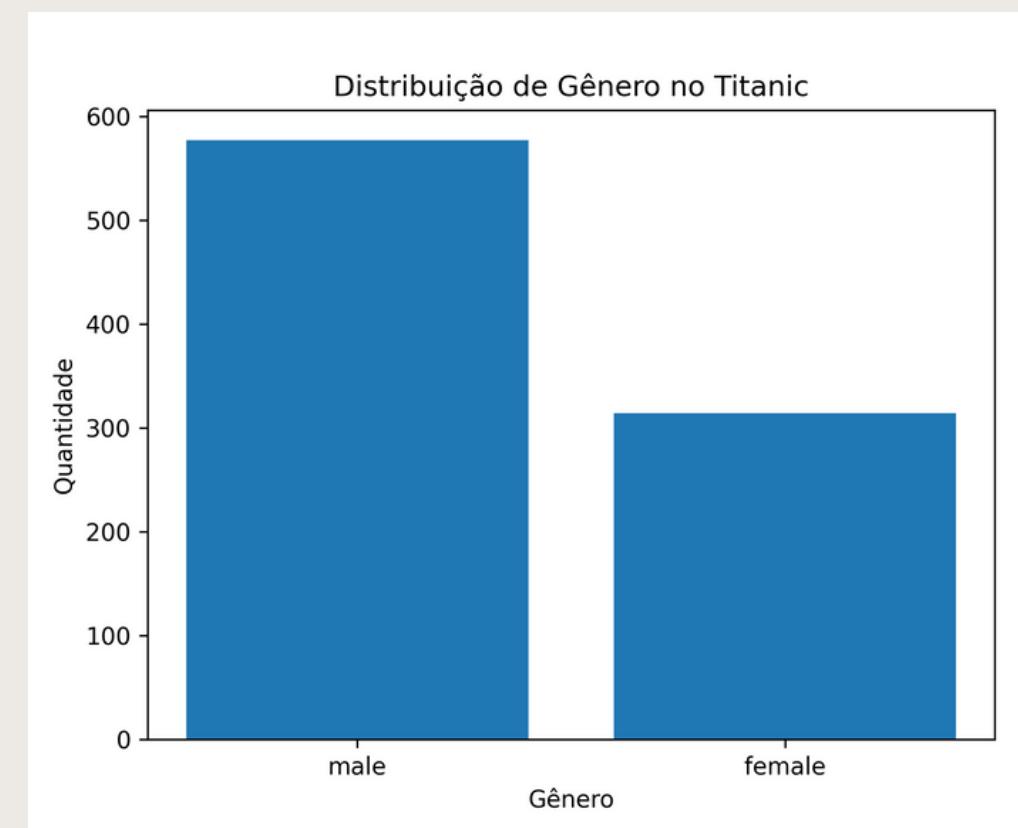
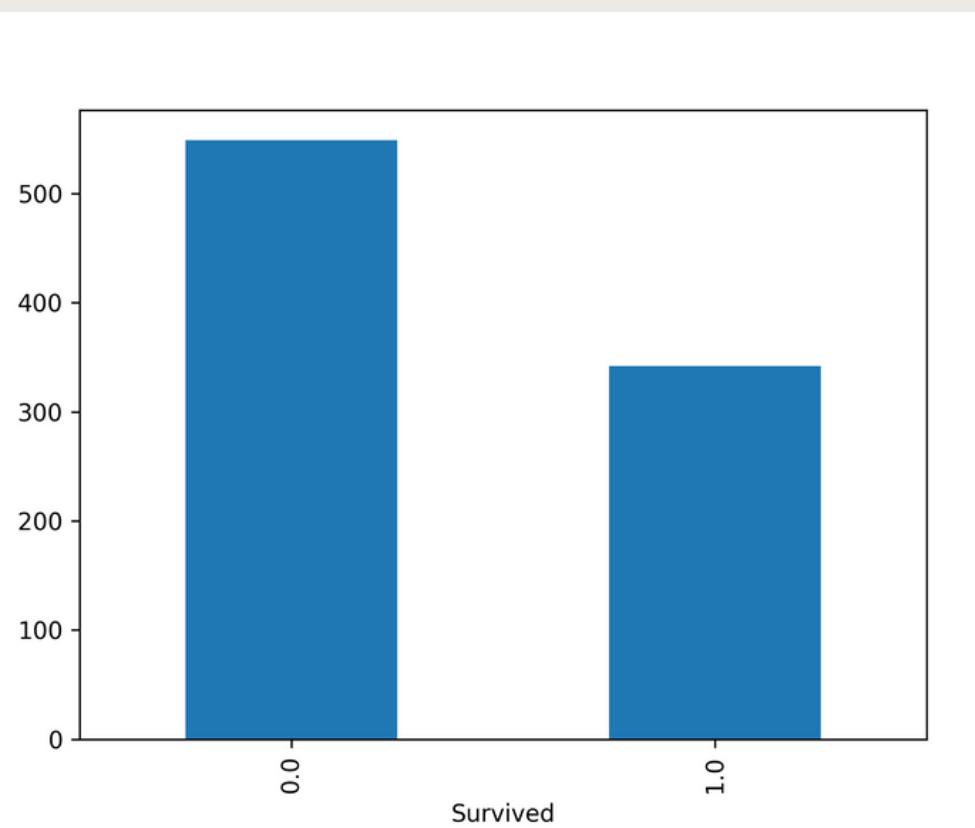
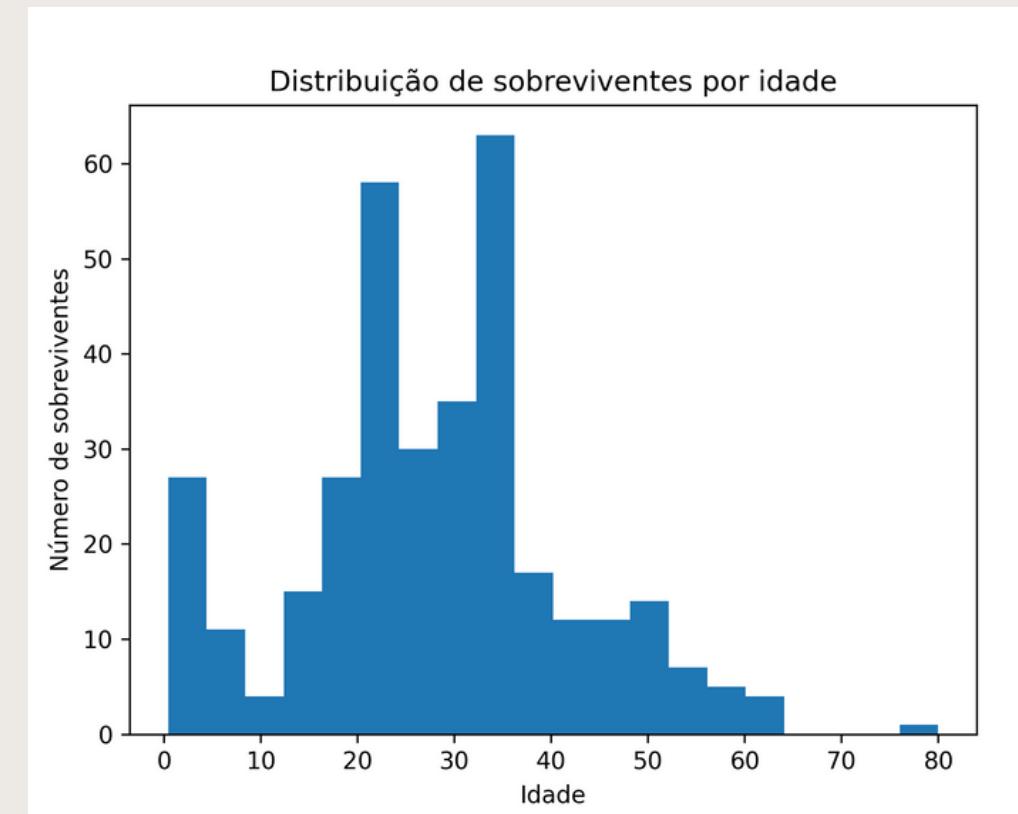
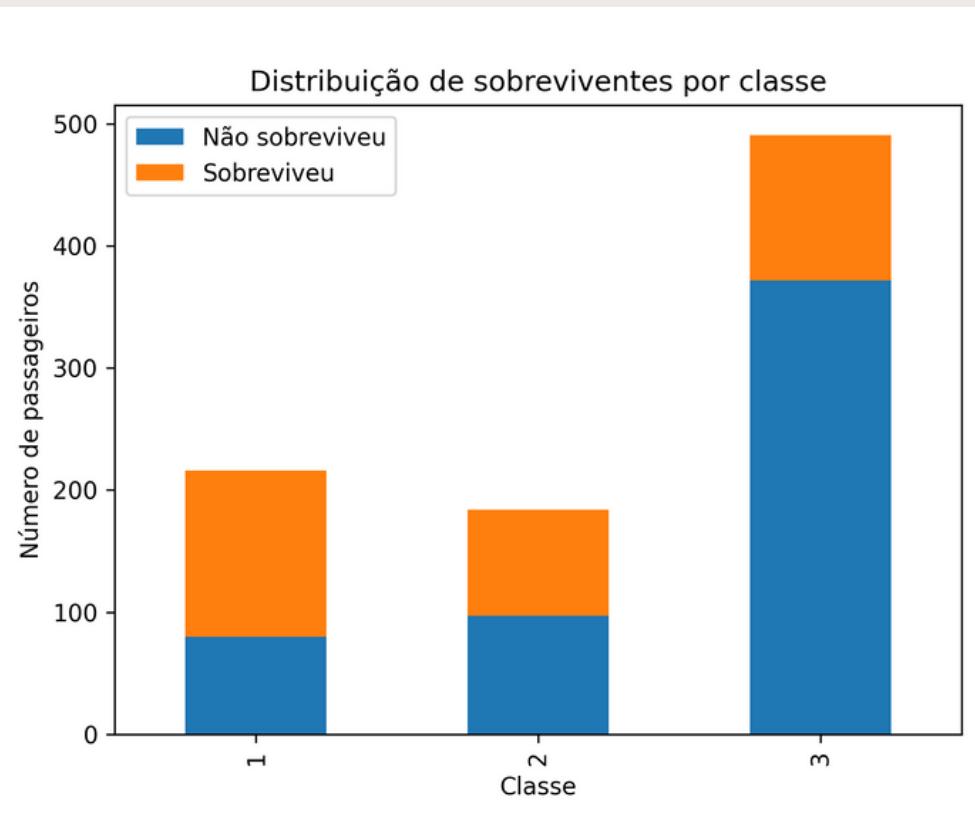
Correções nos dados, como remoção de valores faltantes e eliminação de outliers.

04 Divisão do Conjunto de Dados

Divisão do conjunto de dados em treino e teste para avaliar a eficácia do modelo.

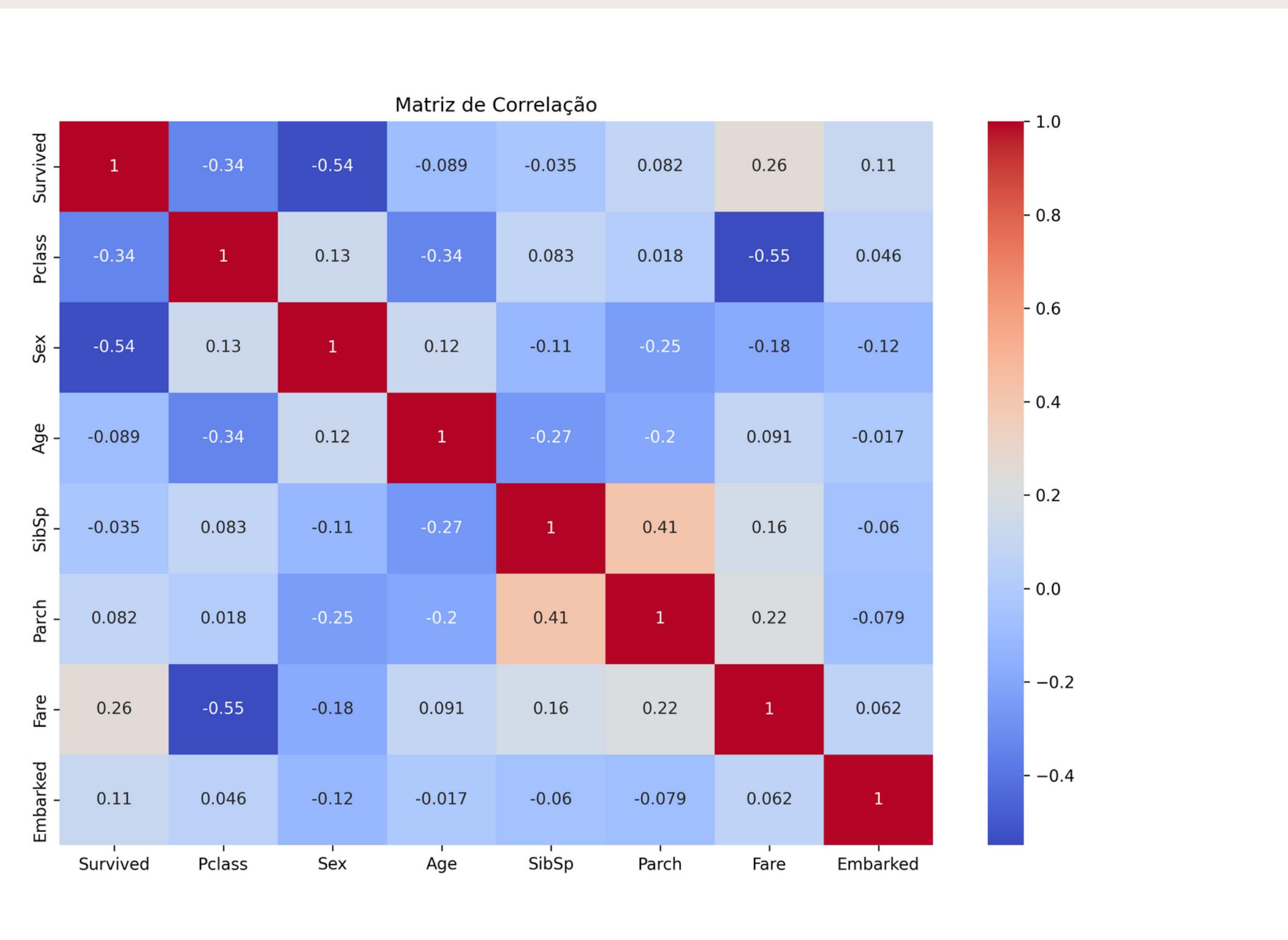
Análise Exploratória dos dados

Nesses gráficos temos exemplos de alguns dados presentes no set, colunas como "Idade", "Gênero", "Classe" e "Sobreviventes" são informações cruciais que serão utilizadas pelo modelo para predição almejada.



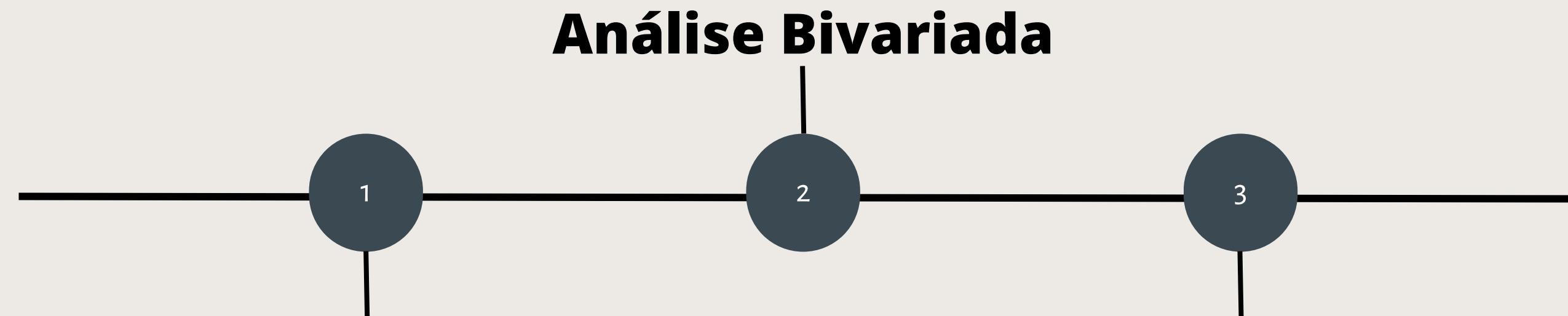
Análise Exploratória dos dados

Na matriz de correlação podemos observar a relação entre as colunas, assim percebemos que as colunas "Survived" e "Sex" possuem uma correlação grande assim como as colunas "Pclass" e "Fare"



Análise Exploratória dos dados

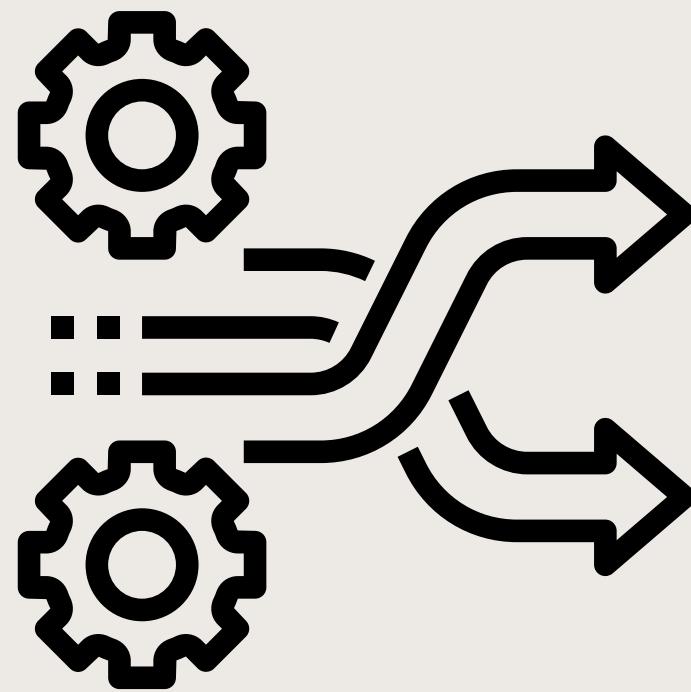
Analisa como duas variáveis se relacionam uma com a outra, em busca de possíveis correlações.



Explora uma variável por vez e visualiza suas características.

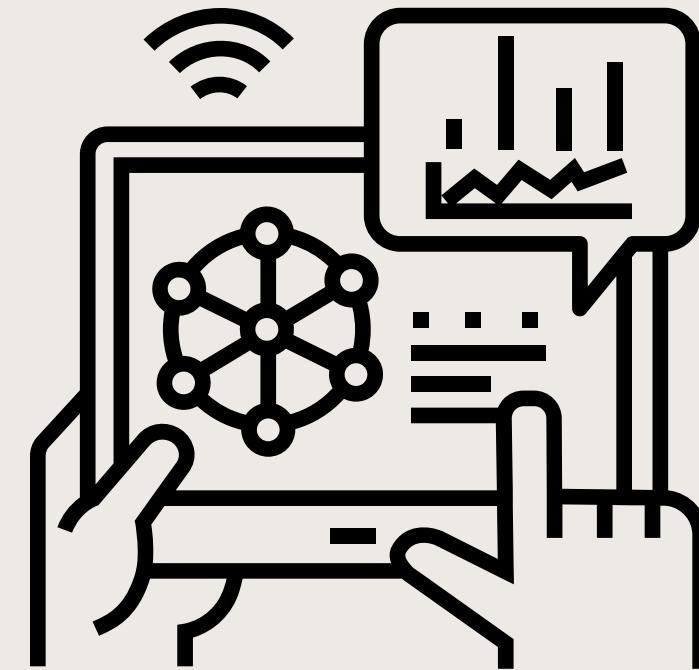
Examina mais de duas variáveis, buscando possíveis relações significativas.

Preparando os Dados para Modelagem



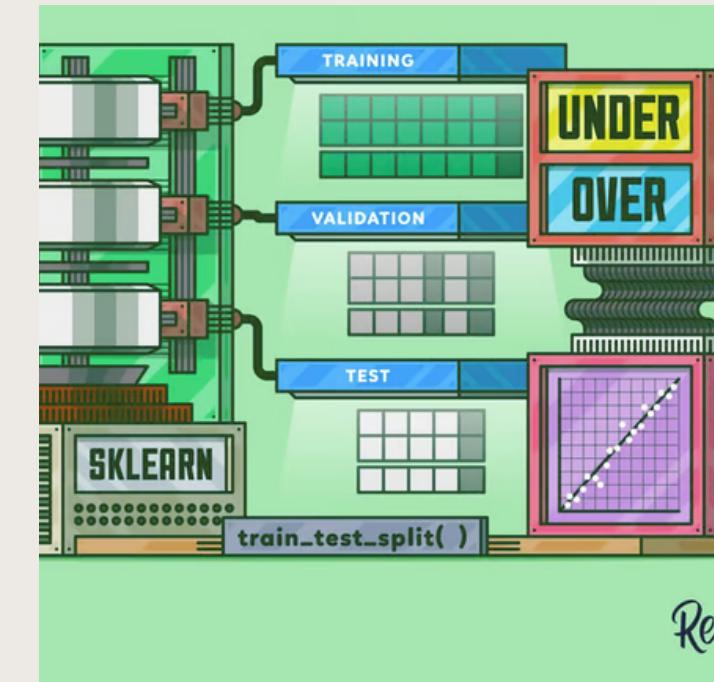
Padronização e Normalização

Aplicamos esses processos para melhorar
a precisão do modelo



Engenharia de Características

Geração de novas características que
possam ajudar o modelo na obtenção de
resultados.



Divisão do conjunto de dados

Geração dos conjuntos de treino e teste
para a validação do modelo.

Construção e Avaliação do Modelo

- •
- •
- •
- •

Escolha do Algoritmo

Seleção do algoritmo mais apropriado para o nosso conjunto de dados.

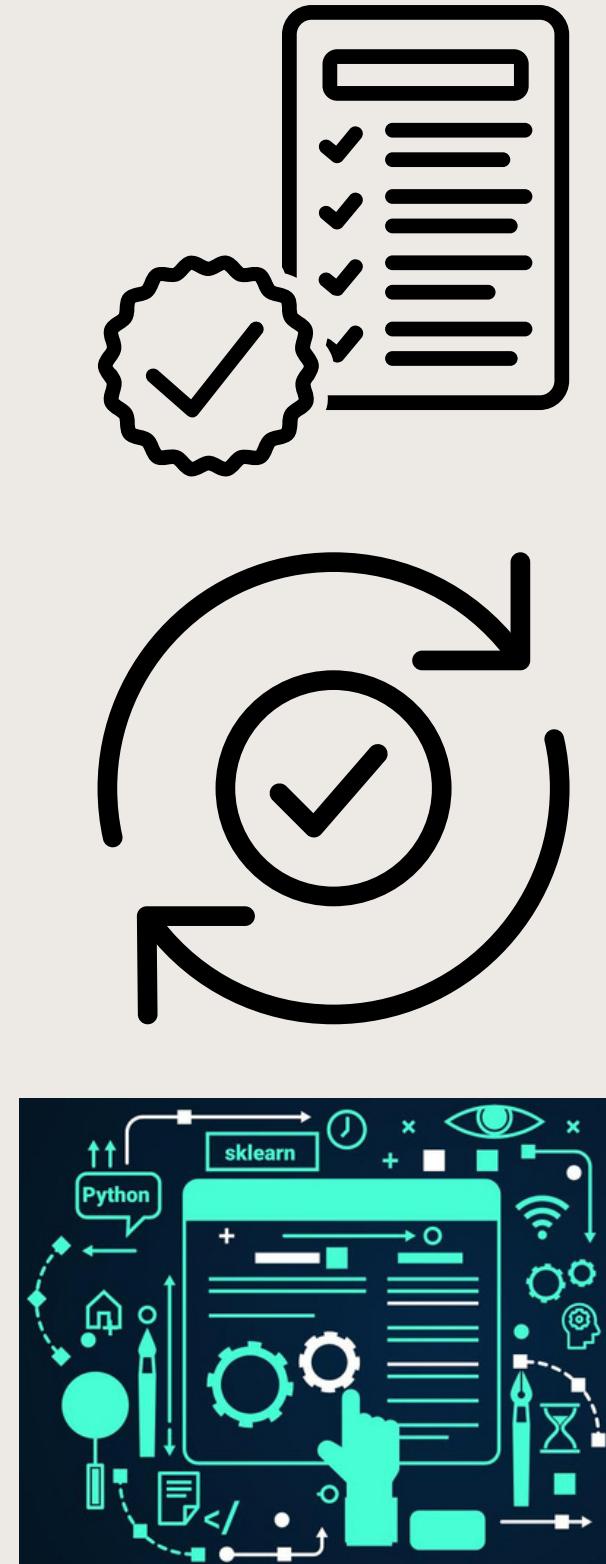
Construção do Modelo

Construção do modelo de machine learning para prever sobreviventes no Titanic.

Avaliação do Modelo

Avaliação da precisão do modelo com base nas métricas disponíveis para avaliação de modelos de classificação.

Melhorando o Modelo



Regularização

Para evitar o overfitting, utilizamos técnicas de regularização.

Cross-validation

Utilizamos esse método para aumentar a confiabilidade do nosso modelo.

Tuning de Hiperparâmetros

Utilizamos esse processo para encontrar a melhor configuração de hiperparâmetros para otimizar a performance do modelo

Resultados

Random Forest

Fold 1: 0.81818181818182
Fold 2: 0.8461538461538461
Fold 3: 0.795774647887324
Fold 4: 0.8309859154929577
Fold 5: 0.823943661971831
Acurácia média: 0.82

Decision tree

Fold 1: 0.7972027972027972
Fold 2: 0.8041958041958042
Fold 3: 0.8169014084507042
Fold 4: 0.795774647887324
Fold 5: 0.7887323943661971
Acurácia média: 0.80

K-Nearest Neighbors

Fold 1: 0.8391608391608392
Fold 2: 0.8391608391608392
Fold 3: 0.823943661971831
Fold 4: 0.8028169014084507
Fold 5: 0.8028169014084507
Acurácia média: 0.82