

Relatório Final - Projeto Titanic

Dataset Utilizado

O dataset escolhido foi o Titanic, disponível no Kaggle ([link](#)). Este conjunto de dados foi selecionado pois permite facilmente a aplicação de técnicas de regressão linear (predição numérica) e regressão logística (classificação binária).

O dataset contém:

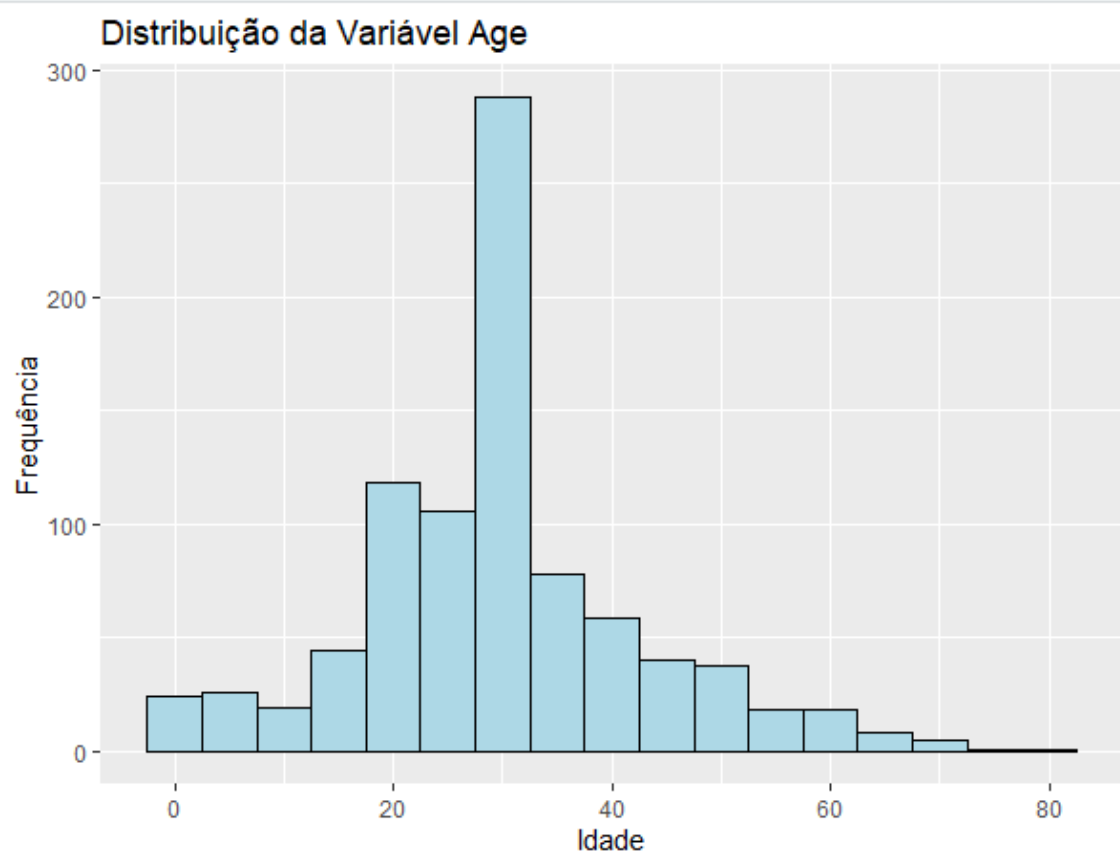
- 891 registros (passageiros)
- 12 variáveis, incluindo idade (Age), tarifa (Fare), classe (Pclass), sexo (Sex), entre outras.

Análise Exploratória

Inicialmente, foi feita uma análise exploratória dos dados, incluindo a distribuição da variável Age e o gráfico de dispersão entre Age e Fare:

Distribuição da variável Age:

Interpretação: A maioria dos passageiros possuía idade entre 20 e 30 anos, com pico significativo nessa faixa.



Regressão Logística (Classificação)

Para a regressão logística, utilizamos a variável Age para prever a sobrevivência dos passageiros (Survived).

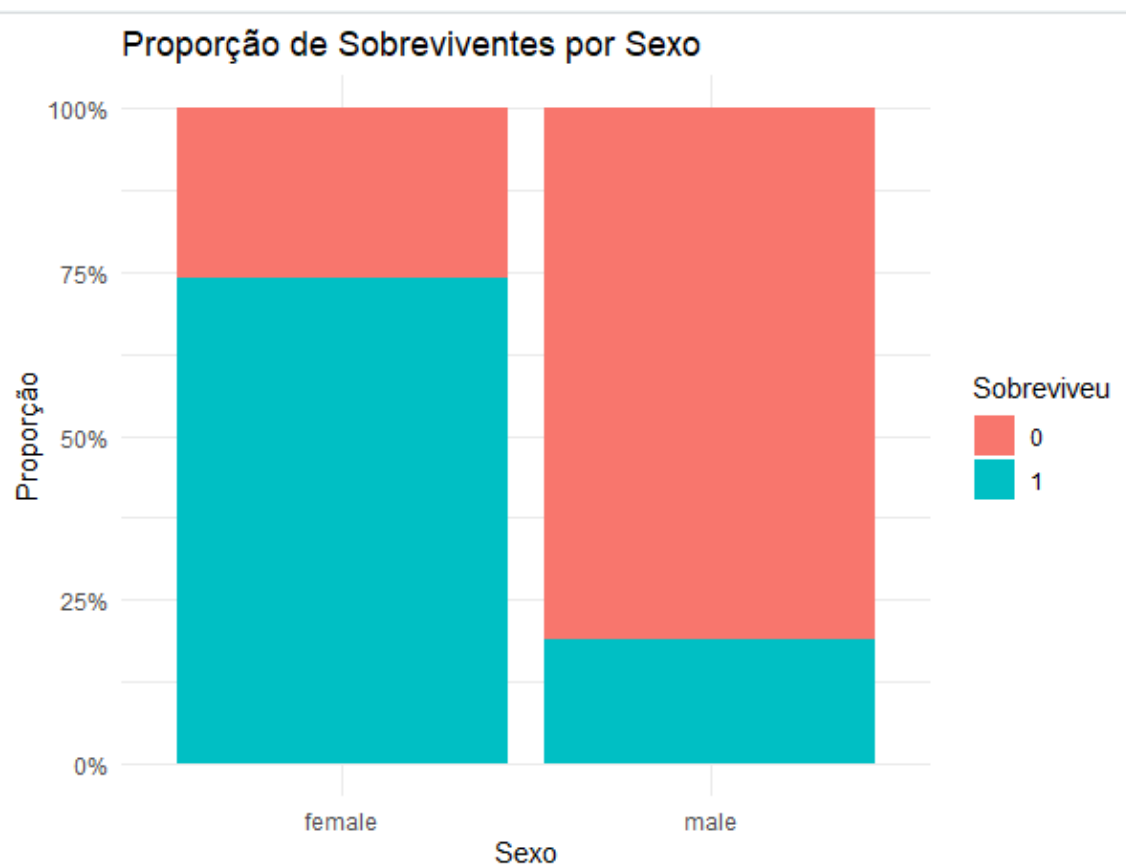
Resultado da Classificação:

Para um passageiro com 40 anos, a probabilidade de sobrevivência foi estimada em:

"probabilidade": 0.90,

"classe": 1

Interpretação: O modelo prevê que a probabilidade desse passageiro sobreviver é aproximadamente 90%. Como essa probabilidade não é inferior a 50%, a classificação resultante foi "sobrevivente" (classe 1).



Podemos ver que por diversos fatores a quantidade de mulheres que sobreviveram foi muito superior



E que dependendo da classe da pessoa esse valor pode variar, pois os que tinham mais dinheiro saíram antes