

Projeto de Análise de Dados 2025

Relatório Final

[15 DE ABRIL]

UF0814 - Programação em Linguagem SQL Avançada

Formadora: Ana Grade

Formanda: Ivandra Santos



Análise do Dataset Titanic

1. Resumo do Dataset

891 linhas e 12 colunas

687 valores ausentes de 'Cabin', substituídos por 'Unknown'

177 valores ausentes de 'Age', substituídos pela mediana da coluna.

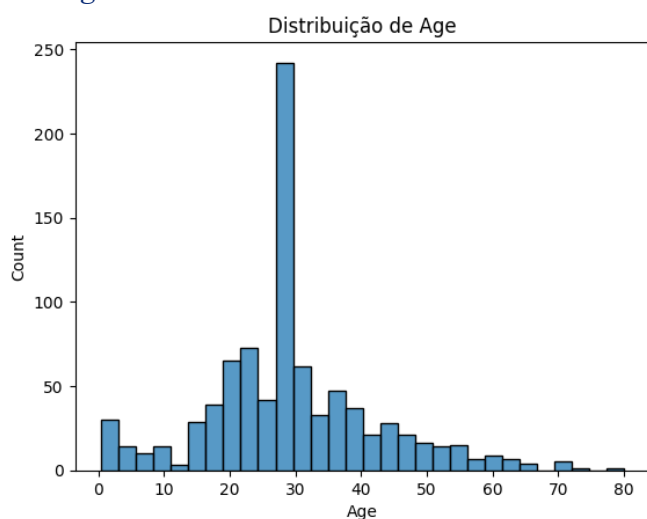
2 valores ausente de 'Embarked', removidos registros aquando ausência de valor

Tabela de Estatísticas Descritivas

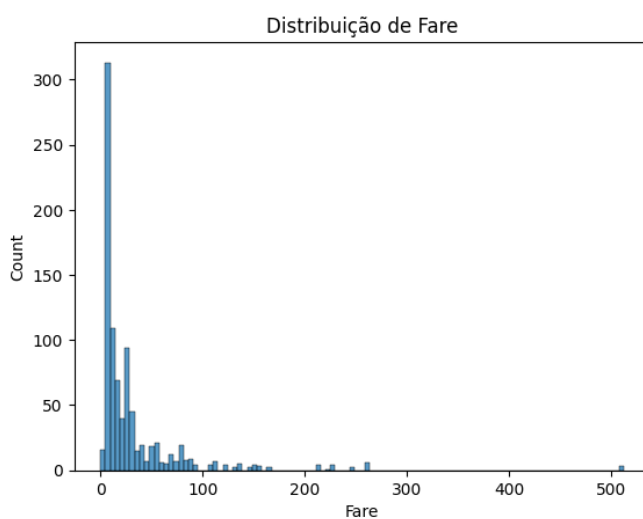
	PassengId	Survived	Pclass	Age	SibSp	Parch	Fare
Contagem	891	891	891	714	891	891	891
Média	446	0,39	2,31	29,70	0,52	0,38	32,20
Desv Padrão	257,35	0,49	0,84	14,53	1,10	0,81	49,70
Mínima	1	0	1	0,42	0	0	0
Máxima	891	1	3	80	8	6	512,33
Quartil 25%	223,50	0	2	20,13	0	0	7,91
Quartil 50%	446	0	3	28	0	0	14,45
Quartil 75%	668,50	1	3	38	1	0	31

2. Gráficos e Visualizações

Histogramas

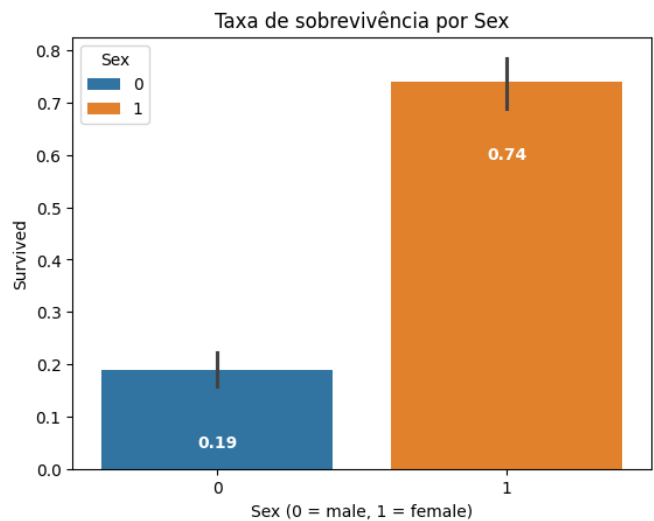


Observação da concentração das faixas etárias

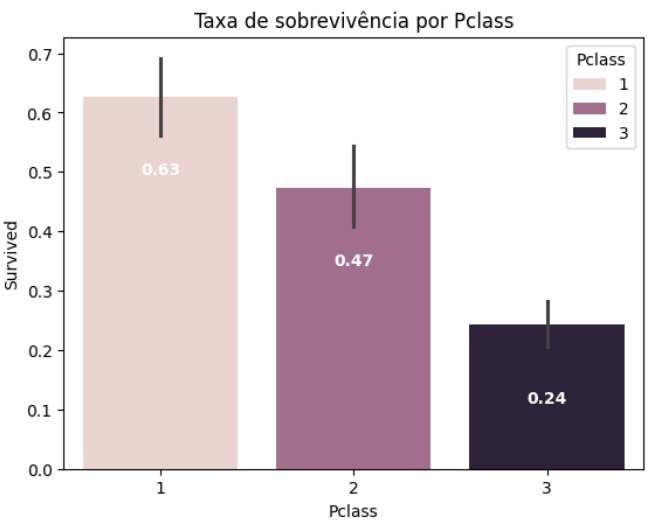


Observação de assimetrias e de valores extremos

Gráficos de Barras para Sobrevivência

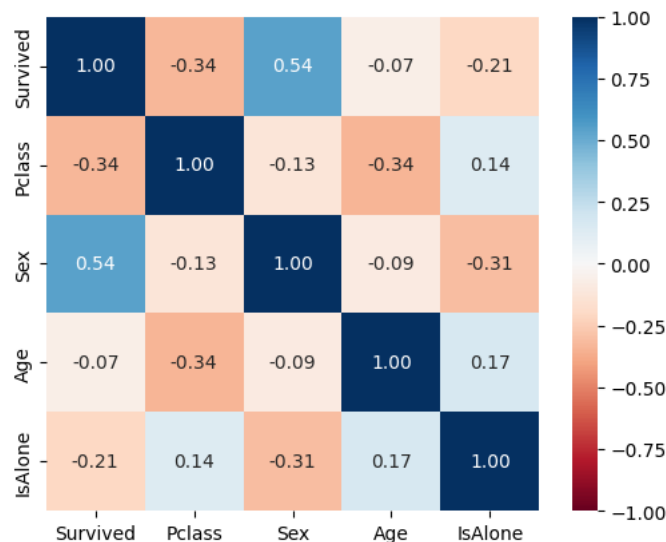


Female (Sex=1) - maior taxa de sobrevivência;
Male (Sex=0) - menor taxa de sobrevivência.



Taxa de sobrevivência mais alta para Passageiros de 1ª classe (Pclass=1), comparando às classes inferiores.

Heatmap de Correlações



As variáveis 'Sex' e 'Pclass' apresentaram correlações notáveis com 'Survived'

3. Resultados dos Modelos

Tabela com Métricas de Desempenho

	Acurácia	F1-Score
Árvore de Decisão	0,792135	0,741259
KNN	0,691011	0,580153

Matriz de Confusão

	Árvore de Decisão	Previsto	
		Negativo	Positivo
	Real		
	Negativo	88	21
	Positivo	16	53

	KNN	Previsto	
		Negativo	Positivo
	Real		
	Negativo	85	24
	Positivo	31	38

4. Conclusões

A análise revelou que passageiros do sexo feminino (Sex=1) têm uma probabilidade significativamente maior de sobrevivência. Os passageiros de 1ª classe (Pclass=1) foram os que apresentaram as maiores taxas de sobrevivência neste grupo. Passageiros acompanhados (IsAlone=0) têm maior taxa de sobrevivência em relação aos que viajavam sozinhos (IsAlone=1).

A Árvore de Decisão apresentou melhores métricas (acurácia e F1-Score) quando comparada ao modelo KNN, o que indica uma maior capacidade preditiva na classificação dos casos de sobrevivência.

5. Recomendações e Observações Finais

Os resultados desta análise exploratória indicam que, desde que os dados comportamentais e as características dos indivíduos esteja bem documentada é possível aplicar este sistema preditivo em outros contextos históricos. Com base no conhecimento dos grupos e das taxas de sobrevivência, é possível definir estratégias para melhorar e aumentar a segurança em realidades idênticas.

Em suma, é sempre importante ter em consideração o contexto dos dados e realizar uma análise cuidada para que a adaptação em realidades idênticas seja a mais adequada possível.