

## Desafio Extra - Análise Exploratória (EDA) - Titanic

Link Github: [https://github.com/dqz22/Desafio\\_Extra\\_Senai.git](https://github.com/dqz22/Desafio_Extra_Senai.git)

Aluno: Diego Alves Queiroz

1.0 - Pergunta a ser respondida ao final da análise: Quais fatores influenciaram na sobrevivência ?

2.0 - Carregar o arquivo e entender melhor sobre os dados na qual iria trabalhar: Quantidade de linhas, colunas, se haviam valores nulos (NaN), quais dados seriam relevantes para análise, etc.

2.1 - Verifiquei que não haviam linhas duplicadas, porém haviam 177 valores nulos na coluna 'Age', 687 na coluna 'Cabin' e 2 na coluna 'Embarked'.

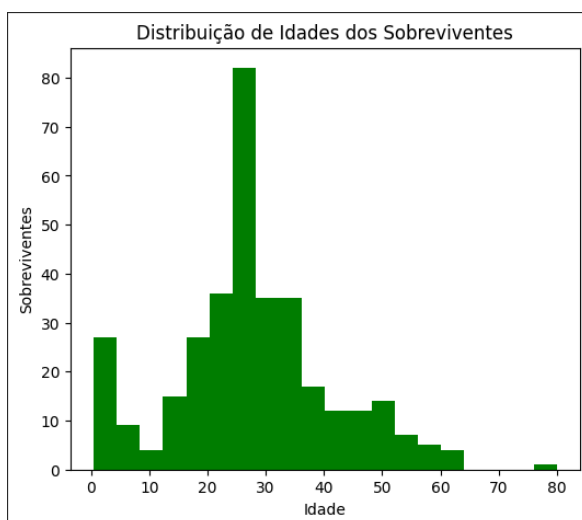
Os valores nulos na coluna 'Age' foram substituídos pela mediana dos valores. Nas demais colunas atribuímos o valor '0' para os valores nulos.

2.2 - Analisando os dados que o arquivo possuía decidi que iria focar a análise nos seguintes dados:

- *Survived* (Sobreviventes)
- *Age* (Idade)
- *Sex* (Genero)
- *Pclass* (Classe Social)
- *Fare* (Valor pago na passagem)

3.0 - Iniciando a análise geral dos dados identificou que aproximadamente 38% dos passageiros sobreviveram

3.1 - Sobreviventes x Idade: Pude observar que a faixa etária com o maior número de sobreviventes estava entre 20 a 35 anos. Essa faixa etária também era a que possuía o maior número de passageiros no início da viagem. Percentualmente, a faixa etária com a maior taxa de sobrevivência foi de 0 até 10 anos, com aproximadamente 59%.

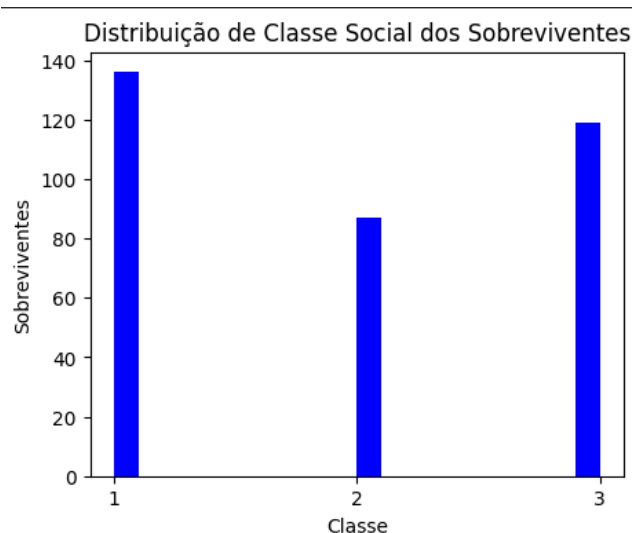


	% Não Sobreviveu	% Sobreviveu
Faixa_Etaria		
0-10	40.625000	59.375000
11-20	61.739130	38.260870
21-30	66.584767	33.415233
31-40	55.483871	44.516129
41-50	61.627907	38.372093
51-60	59.523810	40.476190
61-70	76.470588	23.529412
71-80	80.000000	20.000000

3.2 - Sobreviventes x Classe Social: A classe social com o maior número de sobreviventes foi a 1ª Classe, com aproximadamente 63%, totalizando 136 passageiros. Na 2ª Classe, 47% dos passageiros sobreviveram. Já na 3ª Classe, o número de sobreviventes é expressivo, 119 passageiros, porém esse valor é equivalente a 24% de todos os passageiros da 3ª Classe.

*Survived = 0 (Não sobreviveu)*

*Survived = 1 (Sobreviveu)*



Survived	0	1
Pclass		
1	37.037037	62.962963
2	52.717391	47.282609
3	75.763747	24.236253

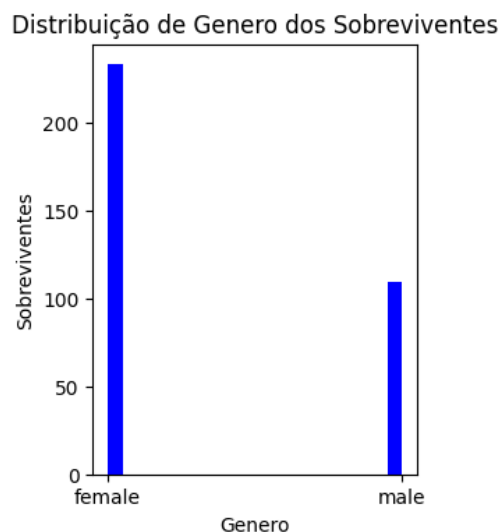
3.3 - Sobreviventes x Gênero: As mulheres foram a maioria entre os sobreviventes, 74%, 233 passageiras ao total. Os sobreviventes homens foram 109 passageiros, equivalente a 19%.

*Survived = 0 (Não sobreviveu)*

*Survived = 1 (Sobreviveu)*

*Female = Mulher*

*Male = Homem*

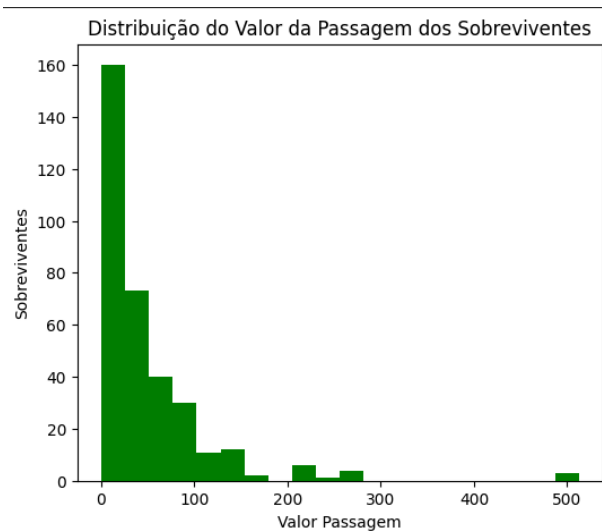


Survived	0	1
Sex		
female	25.796178	74.203822
male	81.109185	18.890815

3.4 - Sobreviventes x Valor pago na passagem : O maior número de sobreviventes está dentro da faixa de 0 a 50 Libras Esterlinas, 232 sobreviventes, mas este valor representa apenas 32% do total desta faixa de valor. O número de passageiros que pagaram mais de 50 Libras caiu drasticamente, porém todos tiveram uma taxa de sobrevivência acima de 64%.

*Survived = 0 (Não sobreviveu)*

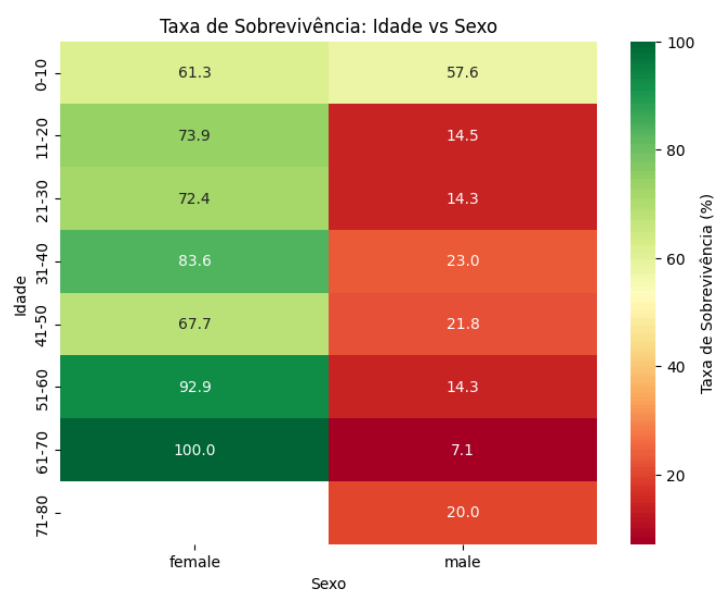
*Survived = 1 (Sobreviveu)*

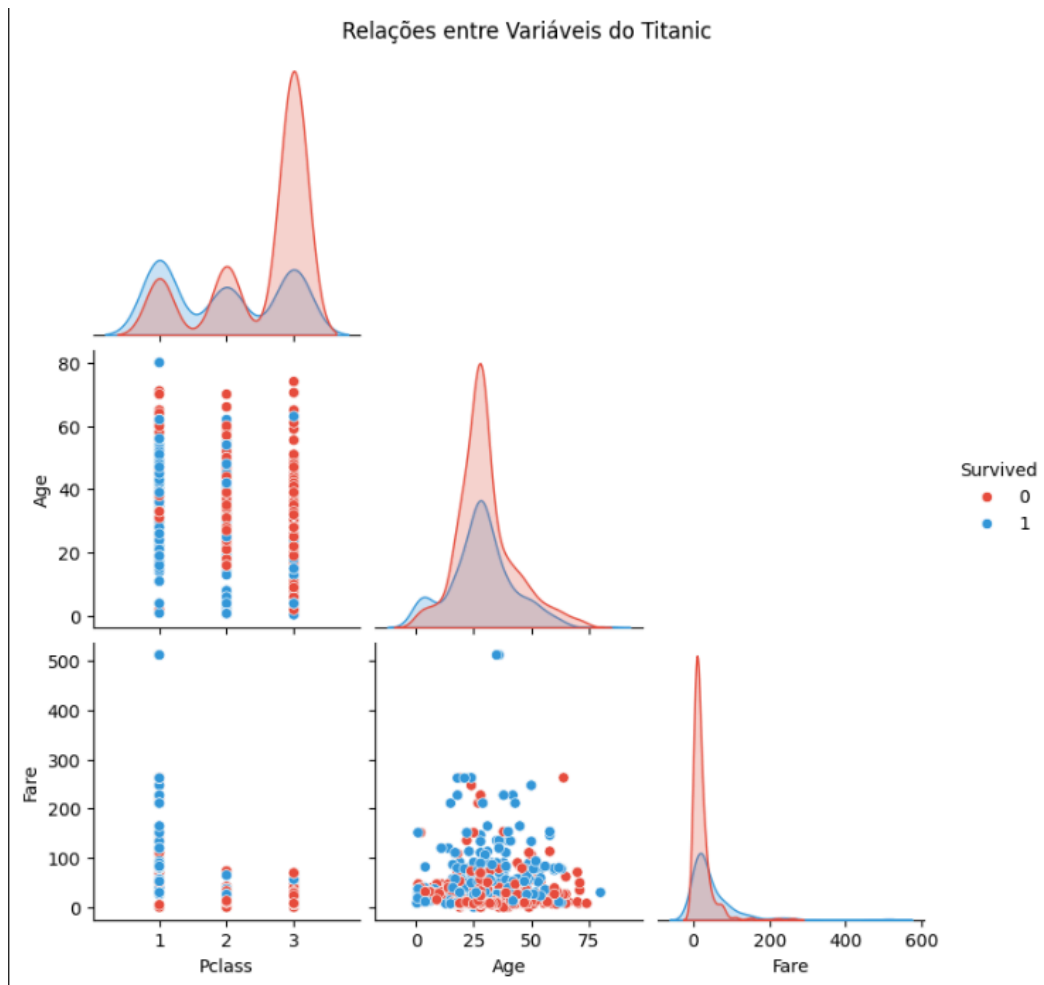


Survived	0	1
Faixa_Precos		
0-50	67.597765	32.402235
51-100	34.579439	65.420561
101-150	20.833333	79.166667
151-200	33.333333	66.666667
201-300	35.294118	64.705882
300+	0.000000	100.000000

#### 4.0 - Cruzando indicadores

4.1 - Sobreviventes x Idade x Gênero: Podemos visualizar que a taxa de sobreviventes entre as mulheres entre 0-10 anos foi maior que 61%, para as mulheres com 11 até 20 anos o valor sobe para mais 72%. As passageiras com 51 até 60 anos a taxa é maior que 92%. No caso dos homens, a faixa etária que possui a maior taxa de sobrevivência é a 0 a 10 anos, com 57%, nas demais faixas a taxa de sobrevivência é igual ou menos a 23%.





## 5.0 - Conclusões e Hipóteses

Os fatores que mais influenciaram na sobrevivência foram o género e a classe social, onde pudemos observar que as mulheres tiveram uma taxa de sobrevivência muito maior que as dos homens, principalmente as mulheres da 1ª e 2ª Classe. A hipótese que surge é que as pessoas da 1ª e 2ª classe possuíam melhores acessos aos botes salva-vidas, e que provavelmente as mulheres tiveram prioridade durante a evacuação do navio.

Embora que após a análise realizada, os dados nos mostre que o principal fator que influenciou na sobrevivência foi a classe social, existem outros fatores não quantificáveis que possivelmente também podem ter influenciado neste resultado, por exemplo a atitude dos homens que podem ter focado seus esforços em ajudar as mulheres. Outro ponto que também influencia nos dados são os valores nulos na coluna 'Age' (Idade) o que pode ter afetado os valores encontrados, ou até mesmo a subnotificação de passageiros, principalmente na 3ª Classe, afetando diretamente os indicadores analisados.