

# Projeto 02 – Análise do Titanic

---

## Nota

Este projeto faz uma análise dos dados de parte dos tripulantes e passageiros do Titanic. Os dados analisados estão descritos no site <https://www.kaggle.com/c/titanic/data>.

## Perguntas realizadas

- Qual o percentual de sobreviventes de cada classe por sexo?
- Qual o percentual de sobreviventes por idade?
- A existência de parentes influenciou nas chances de sobrevivência?

## Descrição do que foi feito na investigação

Para a investigação foi escolhida a ferramenta Jupyter Notebook, por conseguir descrever melhor a análise, gerando código documentado e exibindo gráficos. Junto com a ferramenta escolhida, foram utilizadas as bibliotecas pandas, numpy, matplotlib e seaborn.

A investigação foi dividida nas seguintes etapas:

1. Leitura dos dados em CVS;
2. Limpeza dos dados;
3. Análise inicial do conjunto de dados;
4. Salvando os dados (após limpeza);
5. Perguntas (juntamente com as respostas em forma de gráfico);
6. Conclusão.

## Descrição da limpeza

Inicialmente, foram descartadas algumas colunas consideradas irrelevantes para a análise dos dados. Após esta etapa, as colunas foram renomeadas de forma padronizada. Na situação de valores faltantes, em alguns casos foram descartadas as linhas e em outros se assumiu a média

da coluna para preencher os valores faltantes. Ao final, foram removidas as duplicações dos registros.

## Resultado Final

A seguir são descritos breves resumos das análises de cada questionamento levantado.

### Qual o percentual de sobreviventes de cada classe por sexo?

Observando o gráfico gerado, podemos ver que o percentual de mulheres sobreviventes foi bem maior do que o dos homens. Também, independentemente do sexo, a classe do ticket influenciou nas chances de sobrevivência. De modo que os melhores tickets (primeira classe) davam maiores chances do que os piores tickets (terceira classe).

### Qual o percentual de sobreviventes de cada classe por idade?

De acordo com o gráfico de pizza, pode-se observar que as crianças e os adolescentes tiveram prioridade na evacuação, dada as maiores chances de sobrevivência.

### A existência de parentes influenciou nas chances de sobrevivência?

Quando observamos o gráfico de barras, podemos verificar que as pessoas que tinham maiores chances de sobrevivência eram as que possuíam exatamente 1 irmão ou cônjuge. Também, aqueles que possuíam de 1 a 3 parentes, que fossem pais ou filhos, tinham mais de 50% de chance de sobreviver.

Outro fato, que podemos observar no gráfico apresentado, é que se o passageiro possuísse 4 ou mais parentes à bordo que fossem do tipo pais/filhos ou irmãos/conjuge, suas chances de sobrevivência foram abaixo de 20%. Isto leva a crer que, possivelmente, este passageiro tinham muitos parentes para cuidar em um momento de desespero, diminuindo suas chances de sobrevivência.

## Limitações

Em uma análise inicial, percebeu-se a existência de dados faltantes nos valores de algumas propriedades dos passageiros. Após o descarte de algumas propriedades que foram consideradas irrelevantes, foi verificada a existência das seguintes colunas com valores faltantes: Age e Embarked.

Para cada uma destas propriedades, foram tomadas as seguintes decisões:

- **Age:** Foram encontrados 177 registros com valores faltantes de uma amostra total de 891, representando aproximadamente 20% dos dados. Como tal amostragem é bastante represen-

tativa e para não perder as demais informações, foi decidido substituir os valores faltantes pela média das idades.

- **Embarked:** Como foram encontrados apenas 2 registros com os valores faltantes, decidiu-se efetuar a exclusão dos mesmos, pois estes não representavam uma amostragem significativa.