

# Introdução ao desafio

Este desafio foi desenvolvido para a conclusão da trilha Data Science. Caso tenha alguma dúvida sobre o desafio, pode entrar em contato com a equipe ASCAN;

## Recomendações para o desafio:

- Toda informação que possa ser representada em gráficos deve ser representada, escolher o gráfico adequado para o dado em questão faz parte do desafio;
- Entenda do que o dado se trata, pesquise, teste, pense!;
- Seja curioso, passe por cada coluna olhando elas isoladamente e também procurando se há alguma relação com as demais colunas.
- Utilize boas práticas de programação em Python e em notebooks, segue um [link](#) para mais orientações;
- Todo gráfico deve ter um título e está em um tamanho adequado, lembre-se das boas políticas de visualização.
- Faça métodos (functions) para reaproveitar código;
- Documente o seu código, lembre-se que outros pessoas vão avaliá-lo;
- **Perguntas guias** ou **Perguntas chaves** devem ser respondidas diretamente por você durante a elaboração do notebook, sempre com dados reforçando seus argumentos.
- E o mais importante, divirta-se!

## Dados

Para este desafio vamos utilizar um dos dados mais utilizados na comunidade de Ciência de Dados, o tão famoso TITANIC! Esta versão foi modificada para este desafio com o intuito de facilitar e direcionar melhor os desafiantes. Recomendo que antes de iniciar você se familiarize com o contexto do dataset, o que basicamente foi um navio famoso no início do século 20 que naufragou.

Vou deixar dois links sobre o tema abaixo:

- [Youtube](#)
- [Wikipedia](#)

Metadata:

Variável	Descrição	Valores Chaves
sobreviveu	Se sobreviveu ou não ao acidente.	Não sobreviveu = False, Sobreviveu = Yes
classe_bilhete	Classe do bilhete comprado.	1=Primeira Classe, 2=Segunda Classe, 3=Terceira Classe
genero	Gênero do passageiro.	M=Masculino, F=Feminino
idade	Idade em anos e meses.	Min=0.42 ~ Max=80.0
irmaos_conjuges_abordo	Número de irmãos/cônjuges a bordo do Titanic.	Min=0 ~ Max=6
pais_filhos_abordo	Número de pais/filhos a bordo do Titanic.	Min=0 ~ Max=8
numero_bilhete	Número do Bilhete.	Combinação de texto e número.

Variável	Descrição	Valores Chaves
preco_bilhete	Preço pago no Bilhete.	Min=0.00 ~ Max=512.32
cabine_numero	Número da Cabine do Passageiro.	Combinação de texto e número.
local_embarque	Cidade que o passageiro embarcou.	Valores válidos=['Southampton', 'Cherbourg', 'Queenstown']

### Referência do dado

**ATENÇÃO: Você pode usar outras versões dos pacotes como também outras ferramentas de visualização de dados no lugar do seaborn , como exemplo o plotly !!!**

```
In [ ]: !pip install pandas==1.1.4 seaborn==0.9.0 -q
```

```
In [ ]: import pandas as pd
import seaborn as sns

df = pd.read_csv("titanic.csv").set_index("id_passageiro")
df.head()
```

## EDA

Aqui vamos iniciar a parte de exploratório, as categorias do DataFrame já foram previamente identificadas, mas nada impede que você faça transformações para estas. Fique livre para identificar novos pontos e trazer para a sua defesa no futuro.

Novamente: **Perguntas guias** ou **Perguntas chaves** devem ser respondidas diretamente por você durante a elaboração do notebook, sempre com dados reforçando seus argumentos.

```
In [ ]: numerica = [
    "idade",
    "preco_bilhete",
    "irmaos_conjuges_abordo",
    "pais_filhos_abordo",
    "preco_bilhete",
]

categorica = [
    "classe_bilhete",
    "genero",
    "numero_bilhete",
    "cabine_numero",
    "local_embarque",
    "sobreviveu",
]
```

## Visualize e entenda todas as colunas numéricas.

Perguntas guias para as categorias numéricas:

- Existem dados faltantes no seu conjunto?
- Quais as suas distribuições?

- Temos outliers neste subconjunto? (utilize a definição numérica de outliers apresentado na trilha para fazer essa verificação)
- Existe alguma correlação entre essas variáveis numéricas? (não precisa olhar para outras categorias no momento)

In [ ]: `## Seu código`

In [ ]:

In [ ]:

## Visualize e entenda todas as colunas categóricas.

### Perguntas guias para as categorias categóricas:

- Existem dados faltantes no seu conjunto?
- Conseguimos extrair alguma informação nas colunas `numero_bilhete` e `cabine_numero` ? ([Dica](#))
- Existe alguma casualidade ou correlação entre as categorias?
- Existe alguma relação entre a `classe_bilhete` e o `cabine_numero` ?

In [ ]: `## Seu código`

In [ ]:

In [ ]:

### Perguntas chave para as categorias categóricas:

- Há alguma influência de pertencer a algum grupo de alguma cabine na sua sobrevivência?
- Há algum padrão no `numero_bilhete` ?
- Há algum padrão na `cabine_numero` ?

In [ ]: `### Seu código`

In [ ]:

In [ ]:

## Perguntas do Desafio:

Qual o fator que foi mais importante e menos importante para determinar a sobrevivência para esta população?

In [ ]: `### Seu código`

Sua resposta: .....

**Você pode dizer que houve alguma influência do poder aquisitivo na sobrevivência do acidente?**

In [ ]: `### Seu código`

Sua resposta: .....

**O protocolo marítimo da época do Titanic tinha foco em dar prioridade a crianças e mulheres, analisando os dados podemos dizer se este protocolo foi seguido?**

In [ ]: `### Seu código`

Sua resposta: .....

**Ter uma família no barco influenciou negativamente ou positivamente para a sua sobrevivência?**

In [ ]: `### Seu código`

Sua resposta: .....

**Qual o perfil de quem não sobreviveu? E qual é o perfil de quem sobreviveu?**

- Entenda Perfil como: Gênero, Idade e Classe Social. Caso você identifique alguma nova característica pode adicioná-la ao perfil.

In [ ]: `### Seu código`

Sua resposta: .....

Se você sobreviveu até aqui, meus parabéns! =D