



# Bootcamp Data Analytics 2025

Challenge Frequência e Medidas



# Desafio 1: O perfil do Catálogo da Netflix

**Dificuldade: Média.**

Esse desafio tem como objetivo responder algumas perguntas através da análise de dados, aprofundando o conhecimento em frequência e medidas.

Os dados que deverão ser utilizado nesse script foram baixados do kaggle, e podem ser acessados através do link:

<https://www.kaggle.com/code/shivamb/netflix-shows-and-movies-exploratory-analysis>

Este conjunto de dados consiste na lista de filmes e suas respectivas e suas informações.

Variáveis:

- **show\_id** - id único do filme/série.
- **title** - título do filme ou série
- **director** - diretor do filme ou série
- **cast** - elenco do filme ou série
- **country** - país do filme ou série
- **date\_added** - data que foi adicionado no Netflix
- **reading score** - ano de lançamento original do filme
- **rating** - classificação da televisão
- **duration** - duração total do filme ou série.
- **listed\_in** - categoria ou gênero do filme ou série.
- **description** - descrição do filme ou série.
- **type** - tipo de filme ou série

Os dados são arquivos em .csv.



## Tarefas:

### 1. Exploração inicial:

- Quantas linhas e colunas tem o dataset?
- Quais são os tipos das variáveis e se há valores ausentes?

## 2. **Análises de frequência:**

- Qual a proporção de filmes vs. séries no catálogo?
- Qual o gênero mais frequente?

## 3. **Análises estatísticas:**

- Qual a média, mediana e moda do tempo de duração dos filmes?
- Qual o filme mais curto e mais longo?

## 4. **Visualização de dados:**

- Criar um gráfico de barras para mostrar a quantidade de títulos por gênero.
- Criar um histograma para analisar a distribuição da duração dos filmes.



## **Atividade extra:**

- Quais são os 5 países que possuem mais produções no catálogo?