

# Roteiro de Atividades - Ciências de Dados

## 1 Orientações Gerais

- Utilize este roteiro de atividades para consolidar os conceitos apresentados na primeira aula sobre descrição e visualização de dados com Python.
- Crie um novo arquivo no Google Colab e execute cada exercício em uma ou mais células. Teste e observe os resultados antes de prosseguir para o próximo.
- Aproveite para experimentar variações dos códigos apresentados para garantir que tenha entendido a atividade.
- Caso encontre dificuldades, revise o material da aula correspondente ao tópico em questão.
- Mantenha o notebook organizado com comentários explicativos sobre o que cada código faz e suas principais descobertas.

## 2 Preparação

Antes de começar, importe os pacotes necessários e carregue os dados que serão utilizados:

```
import math
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from seaborn import load_dataset

# Carregando iris
iris = sns.load_dataset('iris')
tips = sns.load_dataset('tips')
```

## 3 Parte 1: Operações Básicas e Estruturas de Dados

### 3.1 Exercício 1.1 - Operações Aritméticas

1. Calcule as seguintes operações:
  - $15 + 8$
  - $25 * 3$
  - $100 / 7$  (divisão comum)
  - $100 // 7$  (divisão inteira)
  - $100 \% 7$  (resto da divisão)
  - $2^{**}8$  (potenciação)
2. Use as funções matemáticas para calcular:
  - Logaritmo natural de 50
  - Logaritmo base 10 de 1000
  - Raiz quadrada de 64
  - Seno de  $\pi/4$  (use `math.pi`)

### 3.2 Exercício 1.2 - Variáveis e Atribuições

1. Crie uma variável `altura` com valor 1.75

2. Crie uma variável `peso` com valor 70
3. Calcule o IMC ( $\text{peso} / \text{altura}^2$ ) e armazene em uma variável `imc`
4. Verifique o resultado

### 3.3 Exercício 1.3 - Listas e Arrays

1. Crie uma lista com as idades: [23, 34, 45, 28, 31, 29]
2. Converta essa lista em um array NumPy
3. Multiplique todos os valores por 2 usando a lista e depois usando o array
4. Compare os resultados obtidos

### 3.4 Exercício 1.4 - Sequências

1. Crie uma sequência de números de 1 a 20 usando `range()`
2. Crie um array com 7 valores igualmente espaçados entre 0 e 100 usando `np.linspace()`
3. Repita o número 5 dez vezes usando `np.repeat()`

## 4 Parte 2: Explorando os Dados - Dataset Iris

---

### 4.1 Exercício 2.1 - Primeiras Explorações

1. Visualize as primeiras 10 linhas do dataset iris
2. Verifique as dimensões do dataset (shape)
3. Examine os tipos de dados de cada coluna
4. Verifique se há valores ausentes no dataset. Se houver, exclua estes valores.

### 4.2 Exercício 2.2 - Seleção e Filtragem

1. Selecione apenas a coluna 'sepal length (cm)'
2. Selecione as colunas 'petal length (cm)' e 'petal width (cm)' simultaneamente
3. Selecione as linhas de 20 a 30
4. Filtre apenas as observações da espécie 'setosa'
5. Filtre as observações onde 'petal length (cm)' > 4.0

## 5 Parte 3: Estatística Descritiva - Dataset Tips

---

### 5.1 Exercício 3.1 - Variáveis Qualitativas

1. Calcule a frequência absoluta da variável 'day' (dia da semana)
2. Calcule a frequência relativa da variável 'time' (almoço/jantar)
3. Crie um gráfico de barras para a variável 'smoker'

### 5.2 Exercício 3.2 - Variáveis Quantitativas

1. Gere um resumo descritivo completo da variável 'total\_bill'
2. Calcule separadamente a média, mediana e desvio padrão de 'tip'
3. Crie um histograma da variável 'total\_bill' com 8 intervalos
4. Crie um boxplot da variável 'tip'

### 5.3 Exercício 3.3 - Quartis e Medidas de Posição

1. Calcule os quartis (25%, 50%, 75%) da variável 'total\_bill'
2. Identifique qual é o valor do percentil 90 da variável 'tip'

### 5.4 Exercício 3.4 - Padronização (Z-score)

1. Calcule o Z-score da variável 'total\_bill'
2. Adicione esta nova variável ao dataset como 'zscore\_bill'

3. Verifique que a média do Z-score é aproximadamente 0 e o desvio padrão é 1
4. Crie dois histogramas lado a lado: um da variável original e outro do Z-score

## 6 Parte 4: Medidas de Associação

---

### 6.1 Exercício 4.1 - Associação entre Variáveis Qualitativas (Dataset Tips)

1. Crie uma tabela de contingência entre 'day' e 'time'
2. Calcule as frequências relativas por linha (marginal por linha)
3. Calcule as frequências relativas por coluna (marginal por coluna)
4. Crie um gráfico de barras agrupadas mostrando a relação entre 'day' e 'smoker'

### 6.2 Exercício 4.2 - Associação entre Variáveis Quantitativas (Dataset Iris)

1. Crie um gráfico de dispersão entre 'sepal length (cm)' e 'petal length (cm)'
2. Calcule a covariância entre essas duas variáveis
3. Calcule a correlação entre essas duas variáveis

### 6.3 Exercício 4.3 - Associação entre Variável Quantitativa e Qualitativa

Com dataset Tips:

1. Calcule a média de 'tip' por categoria de 'time' (almoço/jantar)
2. Crie um boxplot de 'total\_bill' por 'day'
3. Faça um gráfico de pontos (pointplot) mostrando a média de 'tip' por 'smoker' com barras de erro **Com dataset Iris:**
4. Calcule um resumo descritivo de 'petal width (cm)' por espécie
5. Crie boxplots de 'sepal width (cm)' por espécie

### 6.4 Exercício 4.4 - Análises Multivariadas

1. **Dataset Tips:** Crie um gráfico de dispersão de 'total\_bill' vs 'tip', colorindo os pontos por 'smoker'
2. **Dataset Iris:** Faça um pairplot das variáveis quantitativas, colorindo por espécie

## 7 Desafios Extras

---

### 7.1 Desafio 1

Usando o dataset Tips, investigue se existe diferença no comportamento de gorjeta entre fumantes e não fumantes em diferentes dias da semana. Use tabelas de contingência e gráficos apropriados.

### 7.2 Desafio 2

No dataset Iris, identifique qual espécie possui a maior variabilidade nas medidas. Compare os coeficientes de variação (desvio padrão / média) de cada medida para cada espécie.

### 7.3 Desafio 3

Crie uma nova variável no dataset Tips chamada 'tip\_rate' que represente a porcentagem de gorjeta em relação à conta total. Analise essa nova variável descritivamente e investigue sua relação com outras variáveis do dataset.