



## CURSO | TECNOLOGIA EM CIÊNCIA DE DADOS

### Professor

Carolina Toledo Ferraz

### Componente

Projeto e análise de visualização do conhecimento

### Aula

Aula 2

### Aplicando conhecimento

#### Exercício 1: Análise Univariada com Histogramas e Boxplots

##### Enunciado:

Escolha um **conjunto de dados de sua preferência**. Ele pode ser:

- Um dataset disponível em sites como [Kaggle](https://www.kaggle.com/), UCI Machine Learning Repository, IBGE ou outro repositório aberto.
- Uma planilha em CSV ou Excel disponível na internet.
- Ou até mesmo dados que você mesmo colete (por exemplo: alturas de colegas, notas de uma disciplina, preços de produtos etc).

Com base nesse conjunto de dados, siga os passos:

1. Selecione uma variável numérica do seu conjunto de dados (por exemplo: idade, preço, salário, nota, tempo de execução etc.).
2. Calcule as seguintes medidas descritivas dessa variável:
  - Média, Mediana, Moda, Variância e Desvio Padrão.
3. Plote um histograma dessa variável, escolhendo um número adequado de classes (*bins*).
4. Plote um boxplot da mesma variável, destacando possíveis *outliers*.
5. Interprete os resultados:
  - O histograma mostra uma distribuição simétrica, enviesada à esquerda/direita, ou multimodal?
  - O boxplot indica presença de *outliers*? Eles fazem sentido no contexto dos dados?

**Entregável:** Código em Python contendo as respostas para todas as perguntas, acompanhado da interpretação dos resultados.



### Exercício 2: Visualização de Séries Temporais

#### Instruções:

1. Escolha um conjunto de dados de séries temporais disponível publicamente (exemplos: séries de preços de ações, dados de temperatura diária, consumo de energia, dados de tráfego ou vendas mensais).
  - Dicas: você pode encontrar séries temporais no [Kaggle](#), Google Dataset Search, ou em bibliotecas como `statsmodels` e `seaborn`.
2. Carregue o conjunto de dados em Python.
3. Faça a visualização inicial da série temporal:
  - Use um gráfico de linha simples (`plt.plot`) mostrando a evolução dos dados no tempo.
4. Crie pelo menos duas visualizações adicionais:
  - Decomposição da série temporal (tendência, sazonalidade e resíduo) usando `statsmodels`.
  - Boxplots ou violin plots para observar padrões sazonais (por exemplo: distribuição mensal ou por dia da semana).
  - Rolling mean / média móvel para suavizar a série e identificar tendências.
  - Ou outra visualização de sua escolha que te ajude a interpretar os resultados.
5. Interprete os resultados:
  - Há uma tendência clara (crescente ou decrescente)?
  - Existe algum padrão sazonal?
  - Você encontrou anomalias/outliers?

Entregável: **Código em Python contendo as respostas para todas as perguntas, acompanhado da interpretação dos resultados.**