# Aula 05: Conversões entre Array e Séries

## NumPy e o Conceito de Array

A biblioteca **NumPy** é a base de muitas outras bibliotecas de análise de dados no Python. Ela é otimizada para trabalhar com grandes conjuntos de dados de forma rápida e eficiente.

O principal objeto do NumPy é o **array**, um tipo de dado que se parece com uma lista, mas com algumas diferenças cruciais:

- **Velocidade:** Arrays são muito mais rápidos que listas para operações matemáticas.
- **Eficiência de Memória:** Ocupam menos espaço na memória, ótimo para grandes volumes de dados.
- **Funcionalidades Matemáticas:** O NumPy oferece uma gama de funções matemáticas e estatísticas pré-construídas, otimizadas para arrays.

Podemos criar um array a partir de uma lista em Python usando a função np.array():

```
import numpy as np
import pandas as pd

lista_numeros = [10, 20, 30, 40, 50]
meu_array = np.array(lista_numeros)

print(meu_array)
# Saída: [10 20 30 40 50]
print(type(meu_array))
# Saída: <class 'numpy.ndarray'>

# Comparando com uma série do Pandas
minha_serie = pd.Series(lista_numeros)
print(minha_serie)
print(type(minha_serie))
```

## Média e Mediana

- **Média:** É a soma de todos os valores dividida pela quantidade de valores. É muito sensível a valores extremos.
- **Mediana:** É o valor central de um conjunto de dados ordenado. Não é afetada por valores extremos.

A principal diferença para a análise de dados é que a **mediana** é uma medida de tendência central mais robusta, especialmente quando os dados possuem valores atípicos (outliers) que podem distorcer a média.

## Atividade Guiada: Medidas de Tendência Central

Vamos usar o NumPy para calcular e analisar a média e a mediana de um conjunto de dados:

```
# Usando dados do nosso exemplo de produtos
# Extraindo a coluna 'preco' do arquivo vendas_produtos.csv com Numpy

precos_array = np.genfromtxt('../Aula03/vendas_produtos.csv', delimiter=',', skip_header=1,
dtype=None, encoding='utf-8', usecols=3)
print(precos_array)
print(type(precos_array))
```

1. **Calcule a média** usando a função `np.mean()`:
   ```
   media = np.mean(precos_array)
   print(f"Média dos preços: R$ {media:.2f}")
   ```

2. **Obtenha a mediana** usando a função `np.median()`.
   ```
   mediana = np.median(precos_array)
   print(f"Mediana dos preços: R$ {mediana:.2f}")
   ```

3. **Calcule a distância entre a média e a mediana**:
   ```
   distancia = (media - mediana) / mediana
   print(f"Distância entre a média e a mediana: {distancia * 100:.2f}%")
   ```

Agora precisamos interpretar o comportamento desses elementos estatísticos:

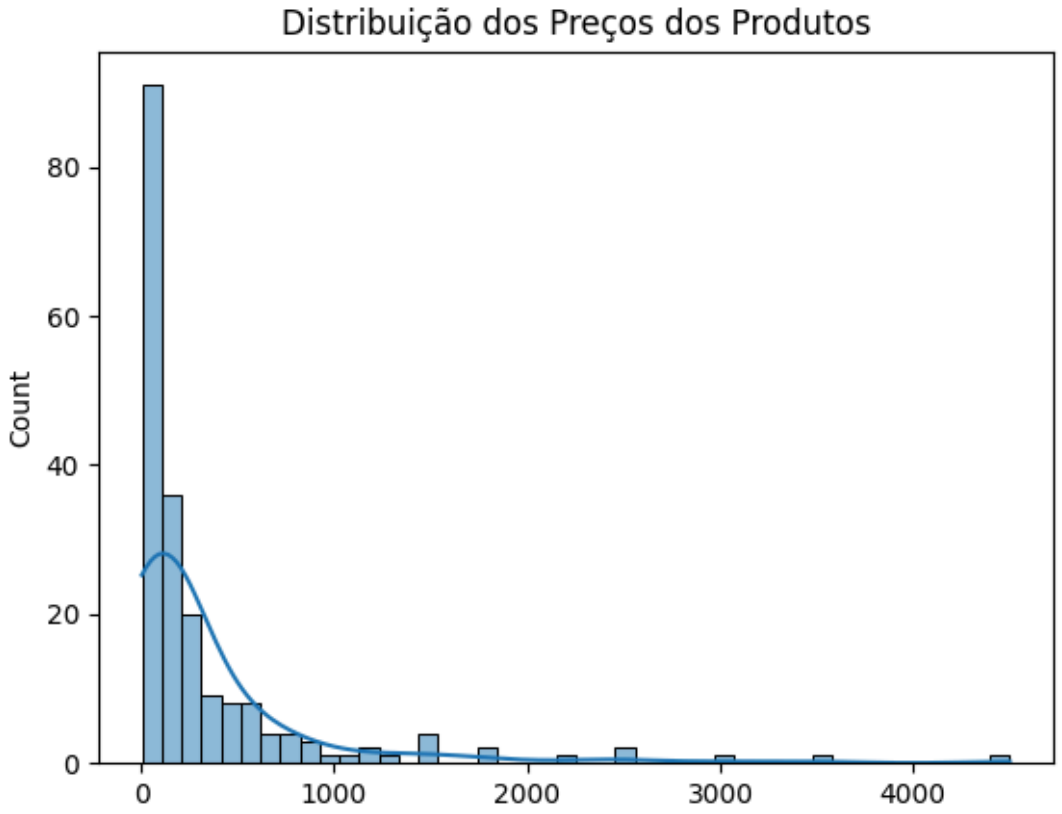
```
if abs(distancia) <= 0.10:
    print("A média tende a ser uma medida de tendência central confiável.")
elif abs(distancia) < 0.25:
    print("A média pode estar sofrendo uma influência moderada de valores extremos.")
else:
    print("A média tende a não ser uma medida de tendência central confiável.")

# Verificando a direção da influência
if media > mediana:
    print("A influência é dos valores mais altos da distribuição.")
elif media < mediana:
    print("A influência é dos valores mais baixos da distribuição.")
```

Por fim, podemos facilitar a interpretação construindo um histograma:

```
# Visualizando a distribuição dos preços
import matplotlib.pyplot as plt  # Importando a biblioteca Matplotlib
import seaborn as sns  # Importando a biblioteca Seaborn
```

```
sns.histplot(precos_array, kde=True) # kde=True adiciona a curva de densidade, traduzindo seria "Kernel Density Estimate"
plt.title('Distribuição dos Preços dos Produtos')
plt.show()
```




## Atividade Prática

1. Utilizem uma coluna quantitativa dos dados que vocês escolheram para trabalhar nas aulas anteriores.
2. Importem esses dados para um array NumPy.
3. Calculem a média e a mediana.
4. Calculem a distância entre elas e interpretem o resultado de acordo com as regras que discutimos.
5. Debatam sobre o que os resultados indicam a respeito da distribuição dos dados.

***Desafio!***

Calculem os quartis (Q1, Q2 e Q3) dos dados e verifiquem se a média está dentro ou fora da faixa interquartil. Discutam o que isso pode significar para a confiabilidade da média como medida de resumo.