

Instruções: Entregar via e-mail em Google Colab ou similares. Não se esqueça de abrir uma caixa de texto no começo do arquivo e inserir um link do seu trabalho

Trabalho 8: Data de Entrega 07/06/2025 até às 8h00min

1. Uma abordagem muito utilizada para a escolha do valor de k , em métodos de agrupamento como o KMeans, é o método do cotovelo. Pesquise e explique como esse método funciona.
2. Vamos considerar **PIB per capita**, **expectativa de vida** e **população** de países das Américas e Europa no ano de 2007. Faça a leitura e seleção dos dados através do código a seguir.

```
# Dataset com países, PIB, expectativa de vida e população
url = 'https://raw.githubusercontent.com/plotly/datasets/master/gapminderDataFiveYear.csv'
df = pd.read_csv(url)

# Filtra apenas dados da Europa e América no ano mais recente (2007)
# Corrected filtering using .isin()
df_euro = df[(df['continent'].isin(['Europe', 'Americas'])) & (df['year'] == 2007)].copy()

# Seleção de variáveis
features = ['gdpPerCap', 'lifeExp', 'pop']
countries = df_euro['country'].values
X = df_euro[features].values
```

Realize os seguintes procedimentos:

- Normalize os dados selecionados;
- Aplique o método de **agrupamento hierárquico** com ligação do tipo **ward**, e exiba o dendrograma com os nomes dos países;
- Utilize o **método do cotovelo** para escolher um valor adequado de k para o algoritmo KMeans;
- Aplique o KMeans com o número de grupos que considerar adequado;
- Liste e Plote os agrupamentos obtidos, utilizando o **PIB per capita** e a **expectativa de vida** como eixos, e identifique visualmente os países no gráfico.

Após realizar os passos acima, responda de forma discursiva:

- (a) Qual a importância da normalização neste contexto?
- (b) Que tipo de estrutura o dendrograma revelou? Quantos grupos são sugeridos visualmente?
- (c) O que representa a “inércia” no método do cotovelo? Qual valor de k foi escolhido e por quê?
- (d) Ao observar o gráfico com os clusters do KMeans, que padrões podem ser percebidos entre os países? Os agrupamentos parecem coerentes?
- (e) Escolha dois clusters distintos e interprete suas características principais com base nas variáveis analisadas.
- (f) Cite uma vantagem e uma limitação do uso do KMeans neste tipo de análise.