

Aula 18 - Implementação com Scikit-learn de Algoritmos de Clusterização: K-means.

Profa. Gabrielly Queiroz

Colab

<https://colab.research.google.com/drive/1d7LLKMUUu00oI60VKLiWLqn3qvmo1Zxp?usp=sharing>

Atividade

Utilize o algoritmo K-Means para realizar o agrupamento de clientes com base nas colunas de idade, renda anual e pontuação de consumo. Leia os dados do seguinte arquivo CSV:

https://raw.githubusercontent.com/tirthajyoti/Machine-Learning-with-Python/master/Datasets/Mall_Customers.csv

Selecione apenas as colunas numéricas necessárias para o agrupamento: Age, Annual Income (k\$) e Spending Score (1-100).

Aplique o K-Means com o melhor valor de k, determinado com base no método do cotovelo e no Silhouette Score.

Adicione ao DataFrame a coluna com o número do cluster correspondente a cada cliente.

Visualize os agrupamentos utilizando um gráfico de dispersão colorido por cluster.

Por fim, exiba a inércia do modelo e utilize groupby() para mostrar as médias das variáveis por grupo e descrever os perfis de clientes encontrados.