

**Oficina 7**

CD-intermediario > ofc7.py > ...

```
1 import numpy as np
2 import pandas as pd
3 import matplotlib.pyplot as plt
4 from sklearn.cluster import KMeans
5 from sklearn.preprocessing import StandardScaler
6
7 """
8 Você foi contratado como analista de dados para uma loja on-line que deseja entender melhor os padrões de
9 compra de seus clientes.
10 A empresa possui dados sobre várias características dos clientes, como o valor médio gasto, a frequência
11 de compras e o número de categorias de produtos comprados. No entanto, esses dados não possuem rótulos que
12 definem quais grupos de clientes existem.
13 Seu objetivo é segmentar os clientes em grupos semelhantes, usando um algoritmo de clustering.
14 """
15
16 # 1. Primeiro, gere o Dataset:
17 # Utilize o Python para criar um dataset sintético com pelo menos 300 registros de clientes;
18 np.random.seed(0)
19 n_samples = 300
20
21 """
22 O dataset deve conter pelo menos 3 features que representam diferentes aspectos do comportamento do cliente,
23 como:
24 'valor_medio_gasto': representa o valor médio gasto por compra.
25 'frequencia_compras': representa o número de compras feitas em um mês.
26 'categorias_produtos': representa o número de categorias de produtos comprados.
27 """
28
29 data = {
30     'valor_medio_gasto': np.random.randn(n_samples) * 20 + 200,
31     'frequencia_compras': np.random.randn(n_samples) * 5 + 15,
32     'categorias_produtos': np.random.randn(n_samples) * 2 + 8
33 }
34
35 df = pd.DataFrame(data)
36
37 # ajustar valores para não ter valores negativos
38 df['valor_medio_gasto'] = df['valor_medio_gasto'].clip(lower=0)
39 df['frequencia_compras'] = df['frequencia_compras'].clip(lower=0)
40 df['categorias_produtos'] = df['categorias_produtos'].clip(lower=0)
41
42 # normalizar os dados (importante para K-means)
43 scaler = StandardScaler()
44 X_scaled = scaler.fit_transform(df)
45
46 # aplicar K-means (vamos escolher k=4 clusters)
47 k = 4
48 kmeans = KMeans(n_clusters=k, random_state=42)
49 kmeans.fit(X_scaled)
50
51 # adicionar a coluna do cluster no dataframe original
52 df['cluster'] = kmeans.labels_
53
54 # visualizar os clusters com um gráfico 3D
55 from mpl_toolkits.mplot3d import Axes3D
56
57 fig = plt.figure(figsize=(10,7))
58 ax = fig.add_subplot(111, projection='3d')
59
60 colors = ['red', 'blue', 'green', 'orange']
61 for cluster in range(k):
62     cluster_data = df[df['cluster'] == cluster]
63     ax.scatter(
64         cluster_data['valor_medio_gasto'],
65         cluster_data['frequencia_compras'],
66         cluster_data['categorias_produtos'],
67         s=50,
68         c=colors[cluster],
69         label=f'Cluster {cluster}'
70     )
71
72 ax.set_xlabel('Valor Médio Gasto')
73 ax.set_ylabel('Frequência de Compras')
74 ax.set_zlabel('Categorias de Produtos')
75 ax.set_title('Segmentação de Clientes via K-means')
76 ax.legend()
77 plt.show()
```

### Justificativa da implementação:

O K-means é um algoritmo simples e eficiente para segmentar clientes em grupos baseados em características numéricas contínuas, como valor gasto, frequência e diversidade de produtos. Ele agrupa os dados em clusters com centróides, facilitando a interpretação dos grupos.

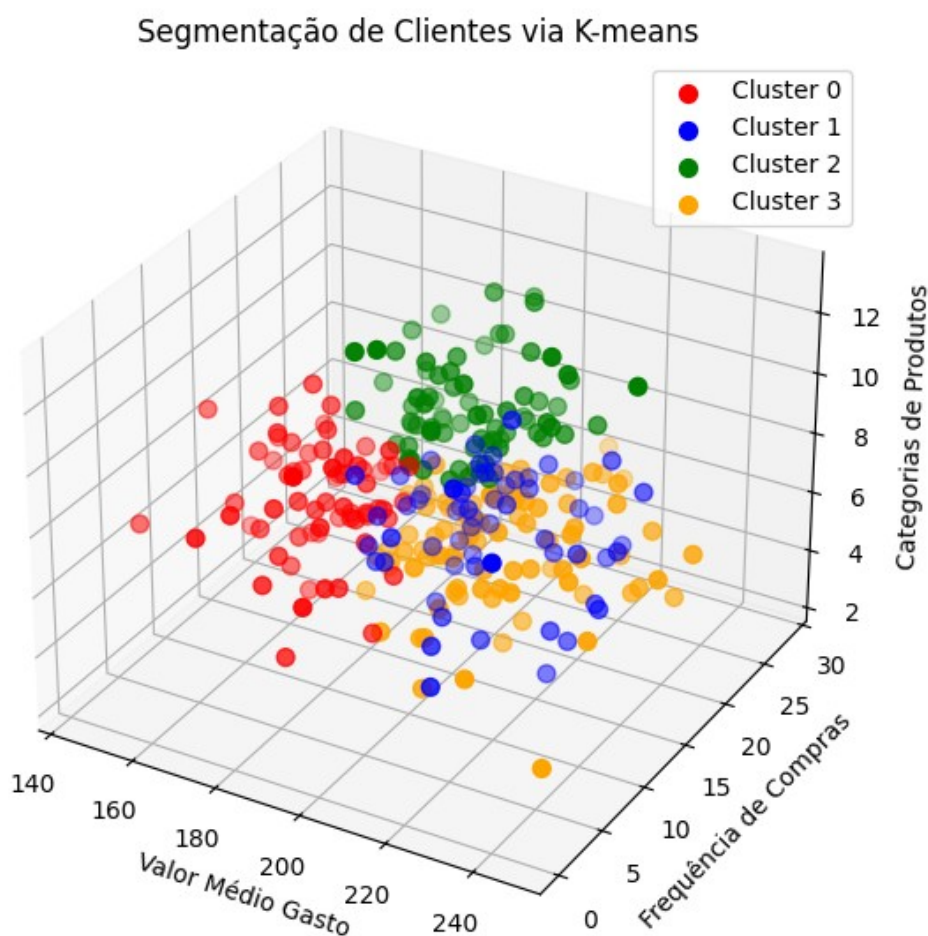
### Vantagens:

- Fácil de implementar e interpretar.
- Rápido em datasets com tamanho moderado, como esse com 300 clientes
- Produz clusters esféricos que, neste contexto, ajudam a entender perfis típicos de consumo

### Limitações:

- É necessário definir previamente o número de clusters. Aqui, escolhemos 4 clusters, mas a escolha pode precisar de análise mais detalhada
- Sensível a outliers, que podem distorcer a posição dos centróides.
- Assume que clusters são convexos e aproximadamente do mesmo tamanho.

### Output:



Cada cluster representa um grupo de clientes com comportamento de compra similar, por exemplo:

- Clientes que gastam pouco, compram pouco e em poucas categorias.
- Clientes que gastam muito, compram frequentemente e em várias categorias.
- Clientes intermediários, etc.

Com essa segmentação, a loja pode criar estratégias personalizadas, como promoções direcionadas, campanhas específicas para cada perfil, ou até melhorar o sortimento de produtos para grupos mais diversificados.