

## Segmentação de clientes

**Dificuldade:** ★★☆☆☆

**Introdução:** Neste projeto, você irá trabalhar com modelos de clusterização a fim de segmentar os clientes com base nos valores anuais gastos de cada cliente. Você trabalhará com técnicas de exploração de dados, pré-processamento, transformação de features e clustering.

**Problema:** Aplicar técnicas de Unsupervised Learning para descrever melhor a variação nos diferentes tipos de clientes com os quais um distribuidor atacadista interage. Colunas a serem utilizadas: FRESH, MILK, GROCERY, FROZEN, DETERGENTS\_PAPER e DELICATESSEN.

**Dataset:** [LINK](#)

### Material Recomendado

- Principal Component Analysis
  - [\[VÍDEO\]](#)
  - [PCA em Python](#)
- [Dados distorcidos](#)
- [Como selecionar o número correto de clusters?](#)

### Material Obrigatório

- [K-Means Clustering em Python](#)

### **Dicas:**

- Comece analisando a distribuição e a correlação entre cada uma das features. Os dados seguem uma distribuição normal?
- Tente identificar outliers. Teste a clusterização com e sem esses pontos.
- Tente escolher qual o melhor número de componentes para se trabalhar. Pense em variância explicada, por exemplo..

### **Resultado:**

- O número mínimo de componentes do PCA corresponde a 2.
- O número de clusters ótimo é 2.
- Benchmark de solução: [LINK](#)