



# Segmentação de Clientes em E-commerce com K-means

Lucas Fortolan Sampaio, Engenharia da Computação, Fundação Hermínio Ometto (FHO)

# O Desafio do E-commerce

O e-commerce gera um vasto volume de dados, mas muitas empresas ainda lutam para convertê-los em insights estratégicos para retenção e fidelização. A análise tradicional não acompanha a complexidade das informações, ocultando padrões relevantes.

## Crescimento Online

Vendas online quase dobraram desde a pré-pandemia, atingindo 17,8% do faturamento total do varejo em fev/2025.

## Dados Comportamentais

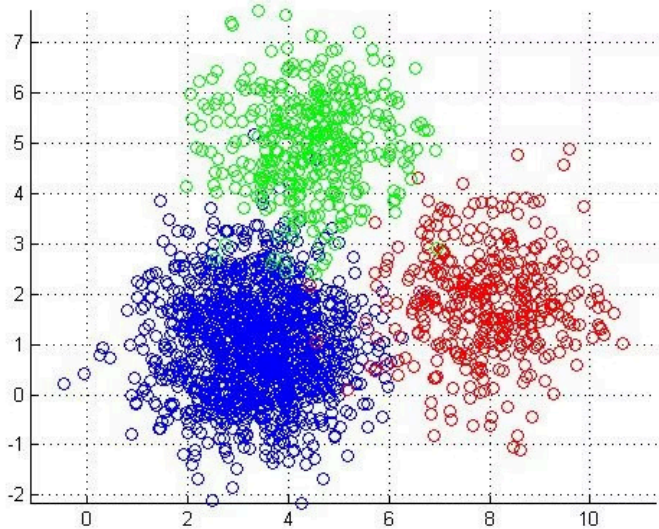
Grandes volumes de dados comportamentais disponíveis, mas subutilizados.

## Marketing Não Personalizado

A análise tradicional não acompanha a complexidade dos dados, resultando em ações de marketing genéricas.

# Solução Proposta: K-means

Segmentar clientes com base no histórico de compras.



## IA Não Supervisionada

Aplicação do K-means para identificar grupos de clientes com características semelhantes.



## Segmentação de Clientes

Identificar grupos e engajamento para ações de marketing personalizadas.



## Otimização de ROI

Aumento do ticket médio e melhoria do relacionamento com o cliente, otimizando o retorno sobre investimento.

# Materiais e Ferramentas

O estudo utiliza uma base de dados do Kaggle de uma empresa do Reino Unido, com 541.909 registros transacionais de 2010-2011. ID da Venda, ID do Cliente, Quantidade de Itens vendidas, Valor do produto Individual e Data da Venda.

## Bibliotecas Python

- **Pandas:** Leitura, manipulação e limpeza de dados.
- **NumPy:** Operações numéricas de baixo nível.
- **Scikit-learn:** StandardScaler, KMeans, Silhouette Score.
- **Matplotlib & Seaborn:** Visualizações gráficas

# Pré-processamento de Dados

O pré-processamento envolve limpeza, normalização e transformação dos dados para adequação às técnicas de análise.

01

## Limpeza e Normalização

Tratamento de valores ausentes, duplicidades e padronização de formatos.

03

## Padronização com StandardScaler

Uniformização das escalas das variáveis RFM para evitar distorções.

02

## Técnica RFM

Cálculo de Recência, Frequência e Valor Monetário para cada cliente.

04

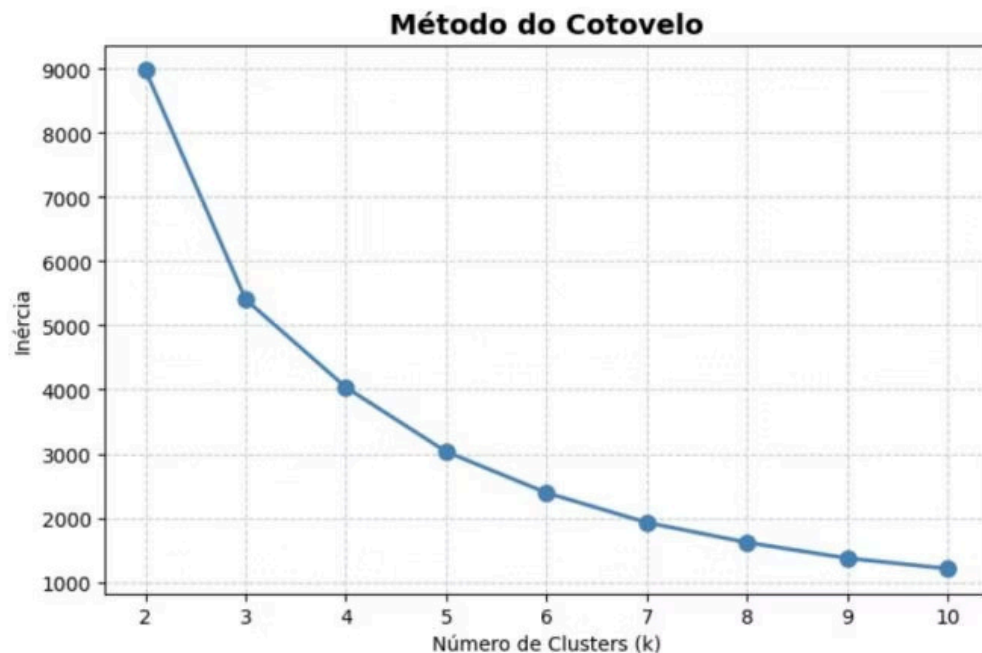
## Modelagem

Aplicação do K-means para criação de clusters de clientes.

# Aplicação do K-means e Avaliação

## Método do Cotovelo

Determinação do número ótimo de clusters ( $k=4$ ), identificando o ponto de equilíbrio entre grupos e variância.



## Índice de Silhouette

Média de 0,61, indicando boa coesão interna e separação entre os clusters, validando a consistência dos resultados.

### Coeficiente de Silhouette

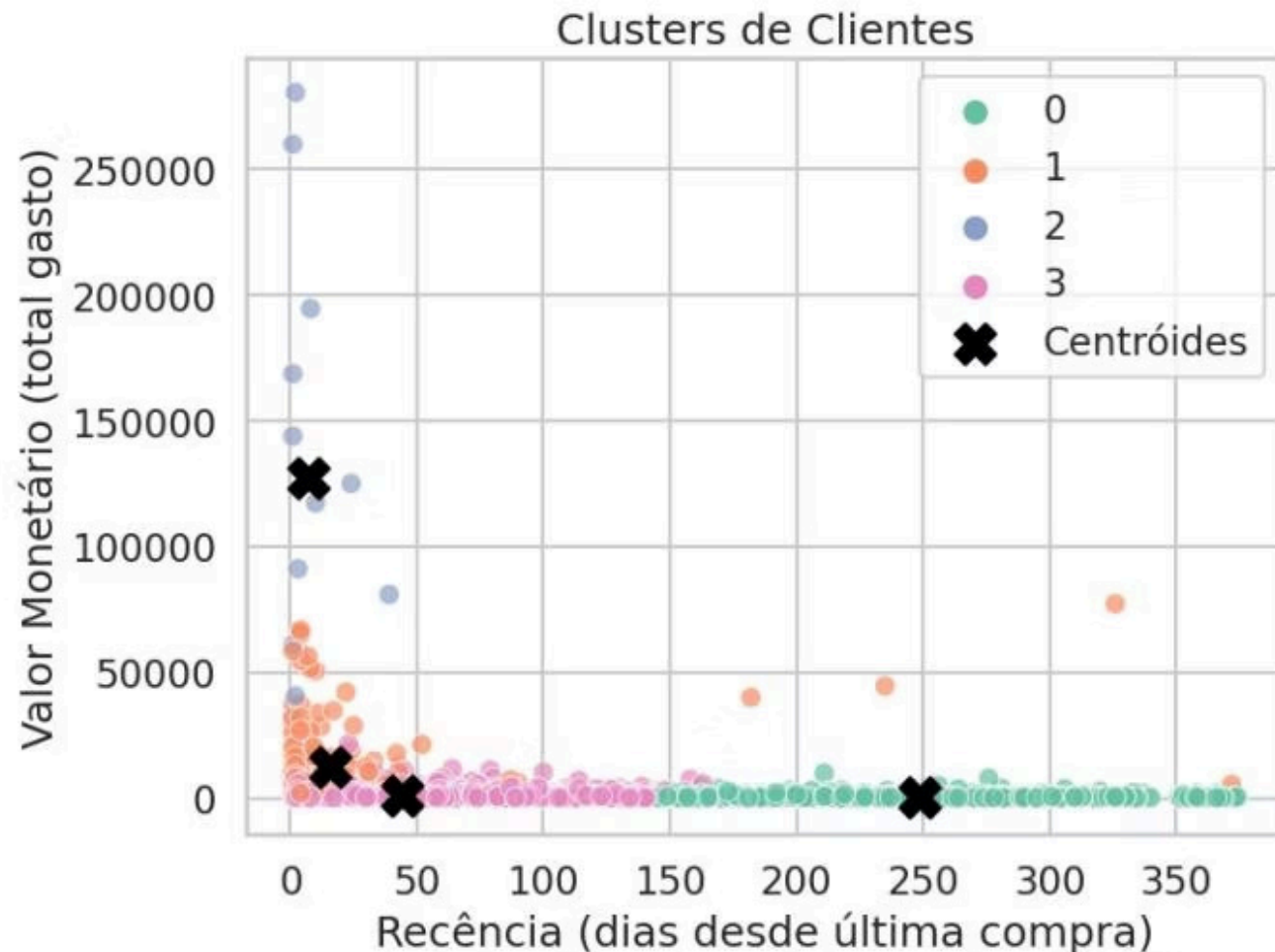
$$s = \frac{b - a}{\max(a, b)}$$

# Clusters de Clientes

A segmentação dos clientes após a análise RFM com K-Means revelou quatro clusters distintos.

Cada cor representa um cluster, e os símbolos "X" indicam os centróides (posição média de cada grupo).

RFM (Recência, Frequência e Valor Monetário)



# Análise de Componentes Principais (PCA)

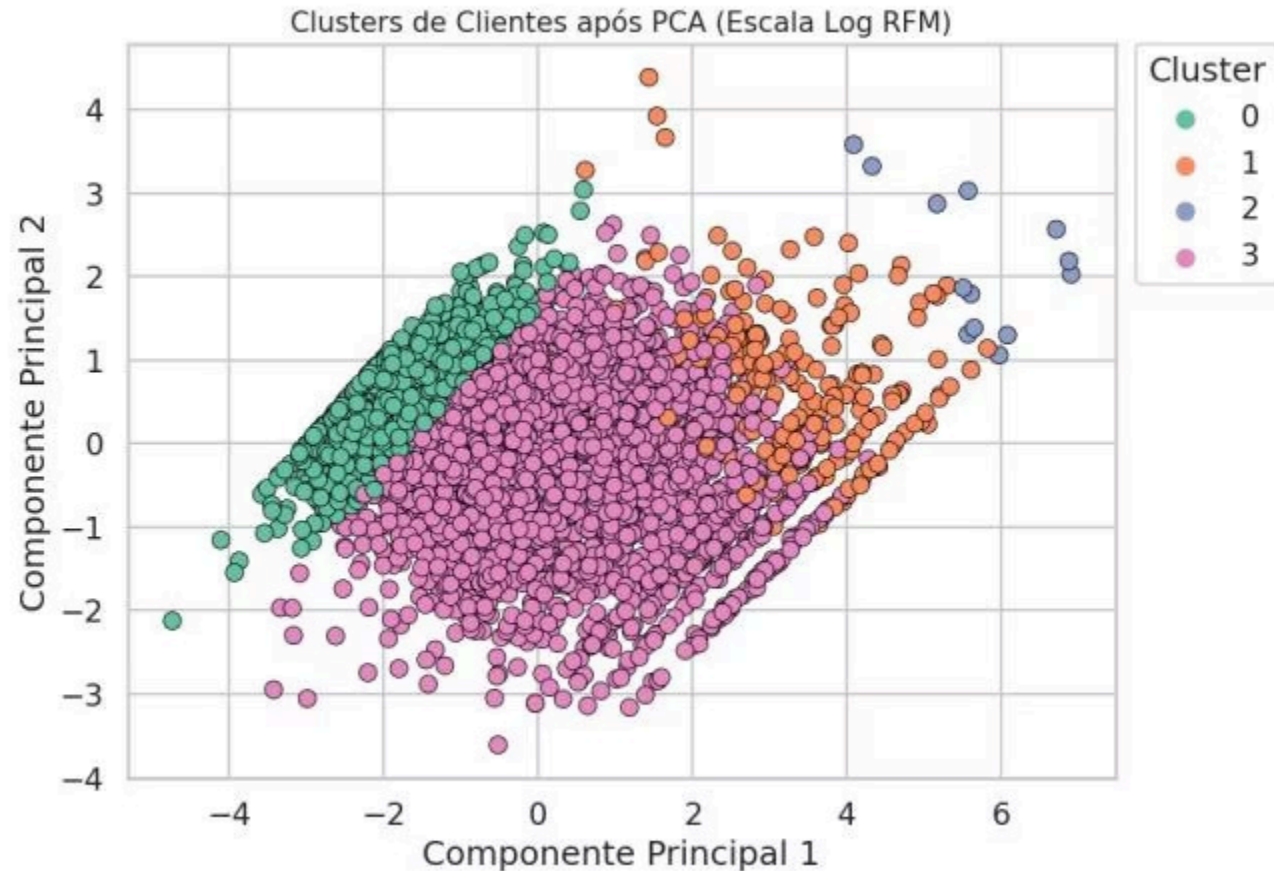
A PCA foi aplicada para visualizar a estrutura espacial dos grupos e validar a separação entre os clusters.

## Eixo Horizontal (PC1)

Correlacionado com a magnitude do consumo (Frequência e Valor Monetário).

## Separação dos Grupos

Distinção clara entre os agrupamentos, com o Cluster 3 como zona de transição.





# Perfis dos Clusters

A análise descritiva dos clusters revela padrões de comportamento distintos entre os clientes.

Cluster	% Clientes	Recência	Freq. Média (ano)	Monetário	Valor Total (£)	% Receita
0	24.48	248.56	1.60	476.33	505863.04	5.69
1	4.86	15.67	22.00	12435.09	2623803.29	29.52
2	0.30	7.38	82.50	127187.96	1653443.47	18.60
3	70.36	43.92	3.70	1344.72	4104099.09	46.18

**Cluster 3:** Maior parte da base (70,36%), gasto moderado, 46,18% da receita.

**Cluster 2 (Outlier):** Grupo seleteo (0,30%), maior valor monetário médio, 18,60% da receita.

**Cluster 1:** Clientes fiéis, alta frequência e ticket médio elevado, 29,52% do faturamento.

**Cluster 0:** 24,48% dos clientes, maior recência média, indícios de desengajamento.

Conferência do Código

# Estratégias de Marketing Direcionadas

Os resultados fornecem insumos estratégicos para ações de marketing personalizadas.

## Cluster 3

Programas de fidelização.

## Cluster 0

Estratégias de reativação e retenção.

## Cluster 1

Ações de cross-selling e recomendações para aumentar o ticket médio.

## Clusters 1 e 2

Prioritários para alocação de recursos de marketing e relacionamento.

**Cluster 1:** Clientes fiéis, alta frequência e ticket médio elevado, 29,52% do faturamento.

**Cluster 3:** Maior parte da base (70,36%), gasto moderado, 46,18% da receita.

**Cluster 0:** 24,48% dos clientes, maior recência média, indícios de desengajamento.

**Cluster 2 (Outlier):** Grupo seletivo (0,30%), maior valor monetário médio, 18,60% da receita.