

# Projeto de Mineração de Dados

**Segmentação de Clientes em Shopping Centers**

Por Victor Renzzo



# Introdução: O Desafio da Segmentação

## Problema

Compreender o comportamento do cliente para otimizar estratégias de marketing em shopping centers.

## Objetivo

Segmentar clientes do "Mall Customers" e descobrir padrões de associação para insights acionáveis.

## Proposta

Utilizar K-means para agrupamento e Apriori para regras de associação, transformando dados em inteligência de negócios.





# Justificativa: Por Que Segmentar?



## Marketing Personalizado

Criar campanhas mais eficazes e direcionadas, aumentando o ROI.



## Otimização de Recursos

Alocar orçamentos de forma inteligente, focando nos segmentos de maior valor.



## Melhora da Experiência

Oferecer produtos e serviços que realmente atendam às necessidades de cada grupo de clientes.



## Aumento de Receita

Impulsionar conversões e o valor médio das compras através de estratégias customizadas.

# Metodologia: A Jornada dos Dados

01

## 1. Carregamento e Descrição

Análise inicial do dataset "Mall Customers" (200 entradas, 5 colunas) para entender a estrutura e identificar valores nulos.

02

## 2. Análise Exploratória (EDA)

Visualização de distribuições (idade, renda, score de gastos) e correlações para insights iniciais.

03

## 3. Pré-processamento

Codificação de gênero, normalização Min-Max e discretização de variáveis para Apriori.

04

## 4. Agrupamento K-means

Determinação do K ótimo (5 clusters) via método do cotovelo e silhueta, seguido pela execução do K-means.

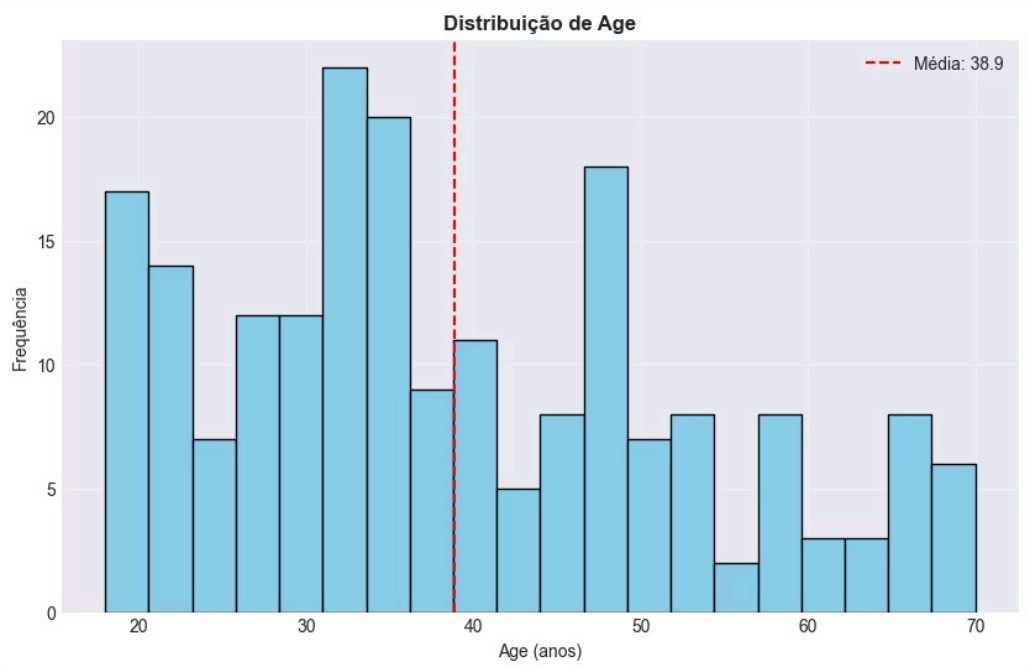
05

## 5. Regras de Associação (Apriori)

Aplicação do algoritmo Apriori para descobrir padrões de compra e comportamento.

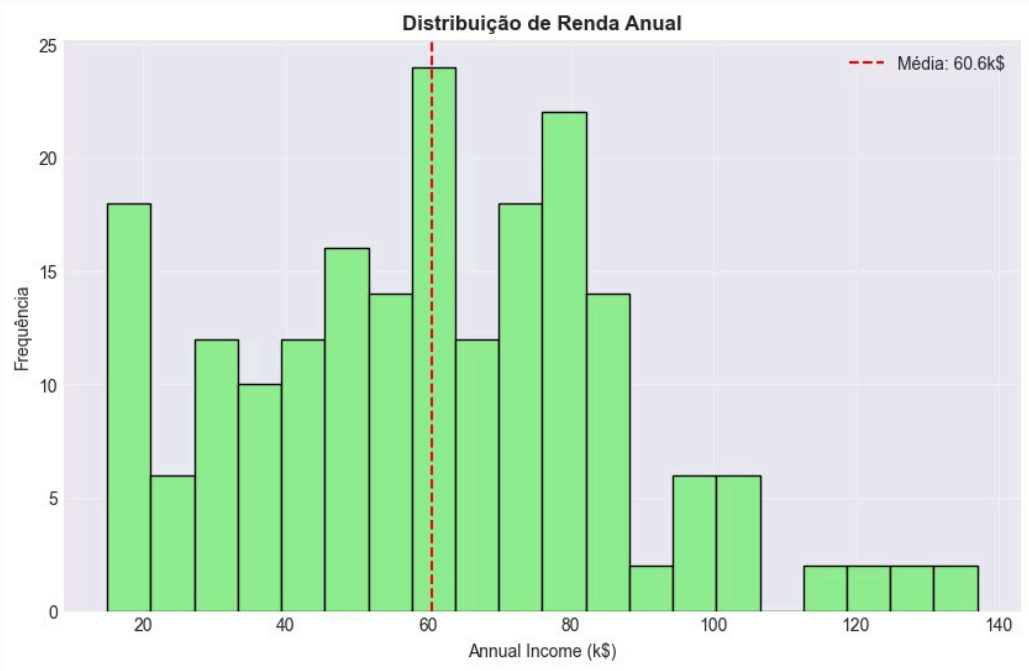
# Análise Exploratória de Dados (EDA)

## Distribuição de Idade



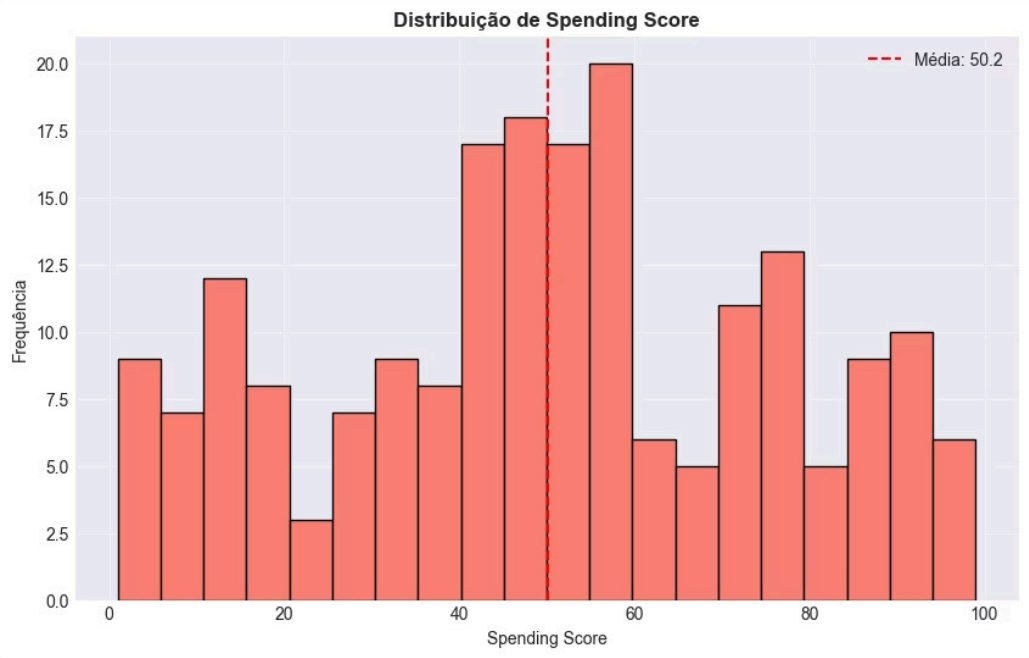
Média de idade: 38.9 anos. A maioria dos clientes está na faixa adulta.

## Distribuição de Renda Anual



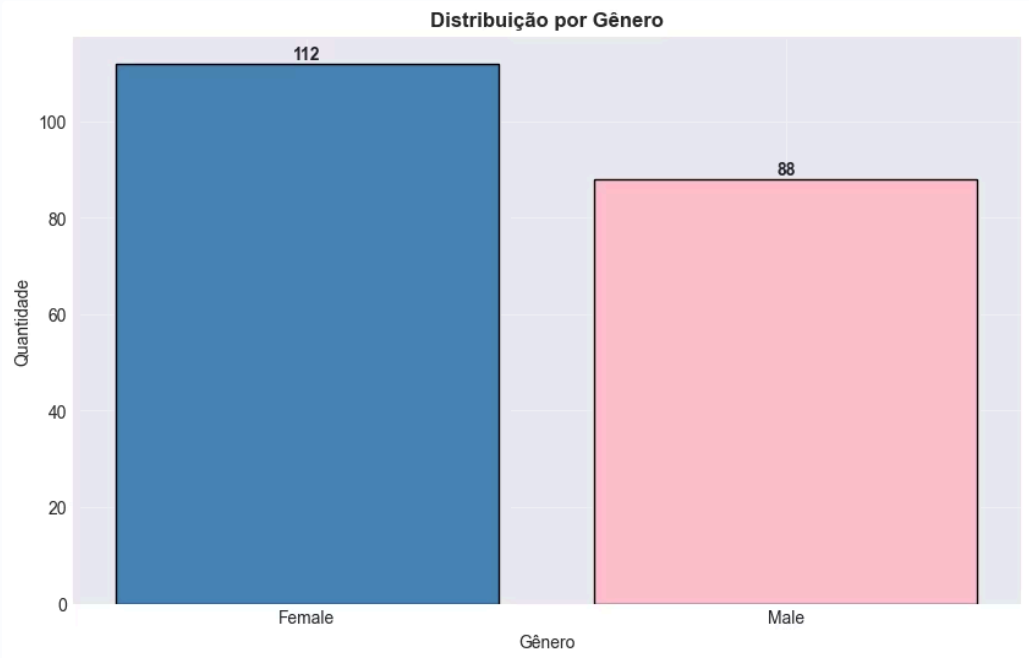
Média de renda: \$60.6k. Concentração na faixa de renda média.

## Distribuição de Spending Score



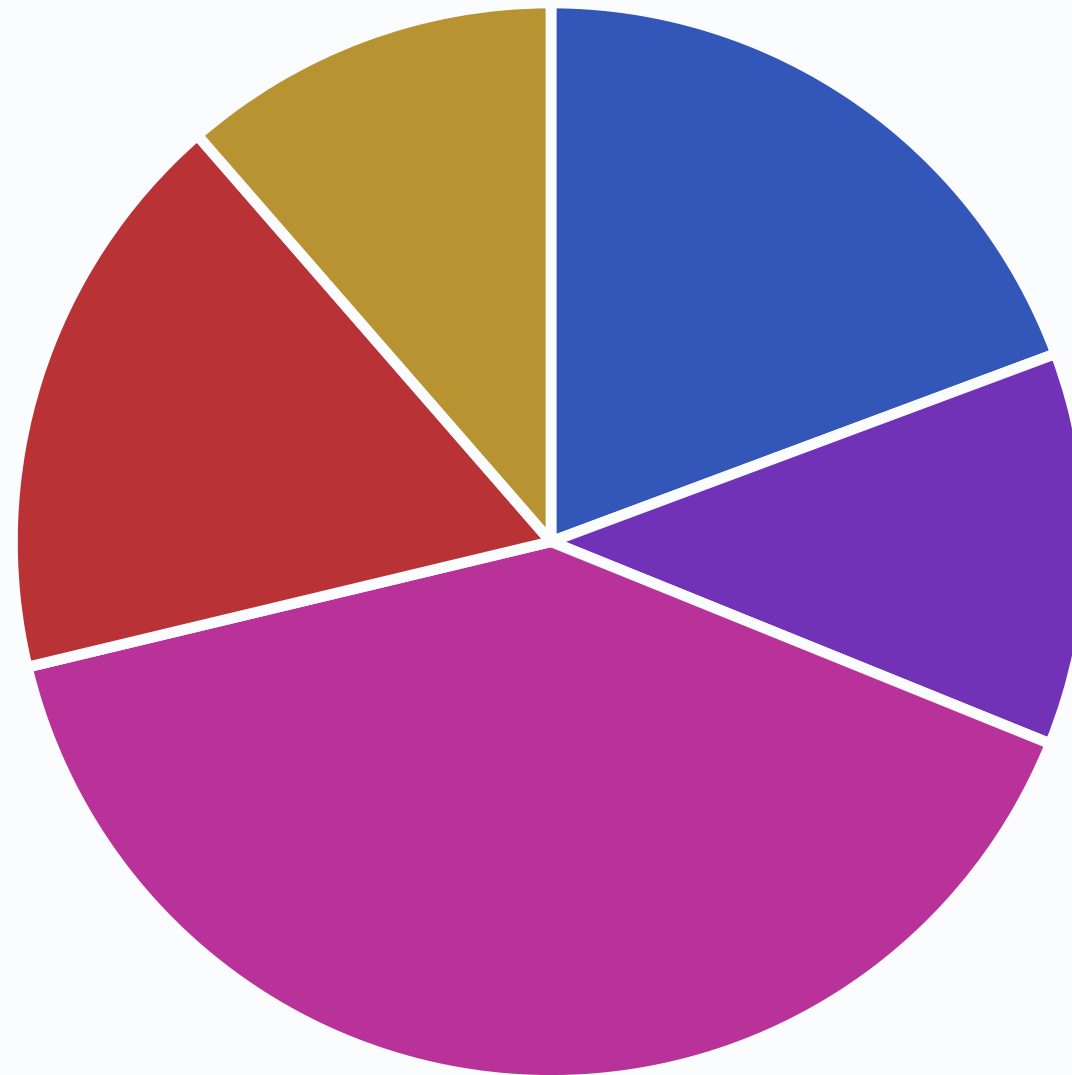
Média de score: 50.2. Distribuição relativamente uniforme, com picos nos extremos.

## Distribuição por Gênero



56% Feminino (112 clientes) e 44% Masculino (88 clientes).

# Agrupamento K-means: Identificando 5 Segmentos



■ Premium

■ Entusiastas

■ Moderados

■ Conservadores

■ Econômicos



# Regras de Associação (Apriori)

Descoberta de padrões de comportamento com confiança mínima de 60% e  $\text{lift} > 1.0$ .

## Regra 1 (Lift: 5.466)

**[Cluster\_1] → [Score\_Baixo, Renda\_Média]**

62.9% dos clientes do Cluster 1 também possuem Score Baixo e Renda Média.

## Regra 2 (Lift: 5.466)

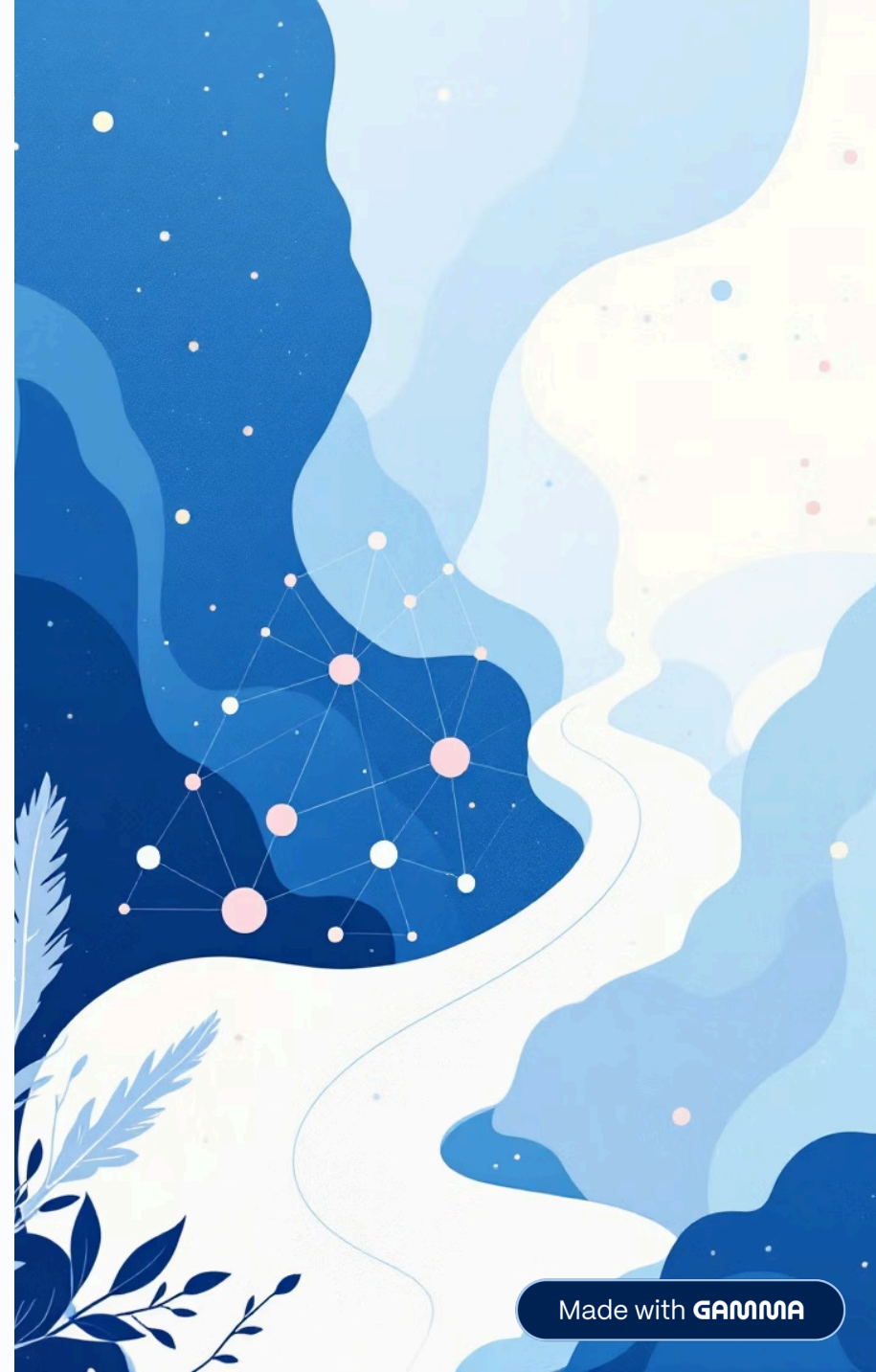
**[Score\_Baixo, Renda\_Média] → [Cluster\_1]**

95.7% dos clientes com Score Baixo e Renda Média pertencem ao Cluster 1.

## Regra 3 (Lift: 5.211)

**[Adulto, Cluster\_2] → [Score\_Alto, Renda\_Média]**

67.7% dos clientes adultos do Cluster 2 também possuem Score Alto e Renda Média.





# Resultados e Conclusões

1

## Segmentos Claros

5 clusters distintos identificados, com perfis de idade, renda e score de gastos bem definidos.

2

## Padrões de Comportamento

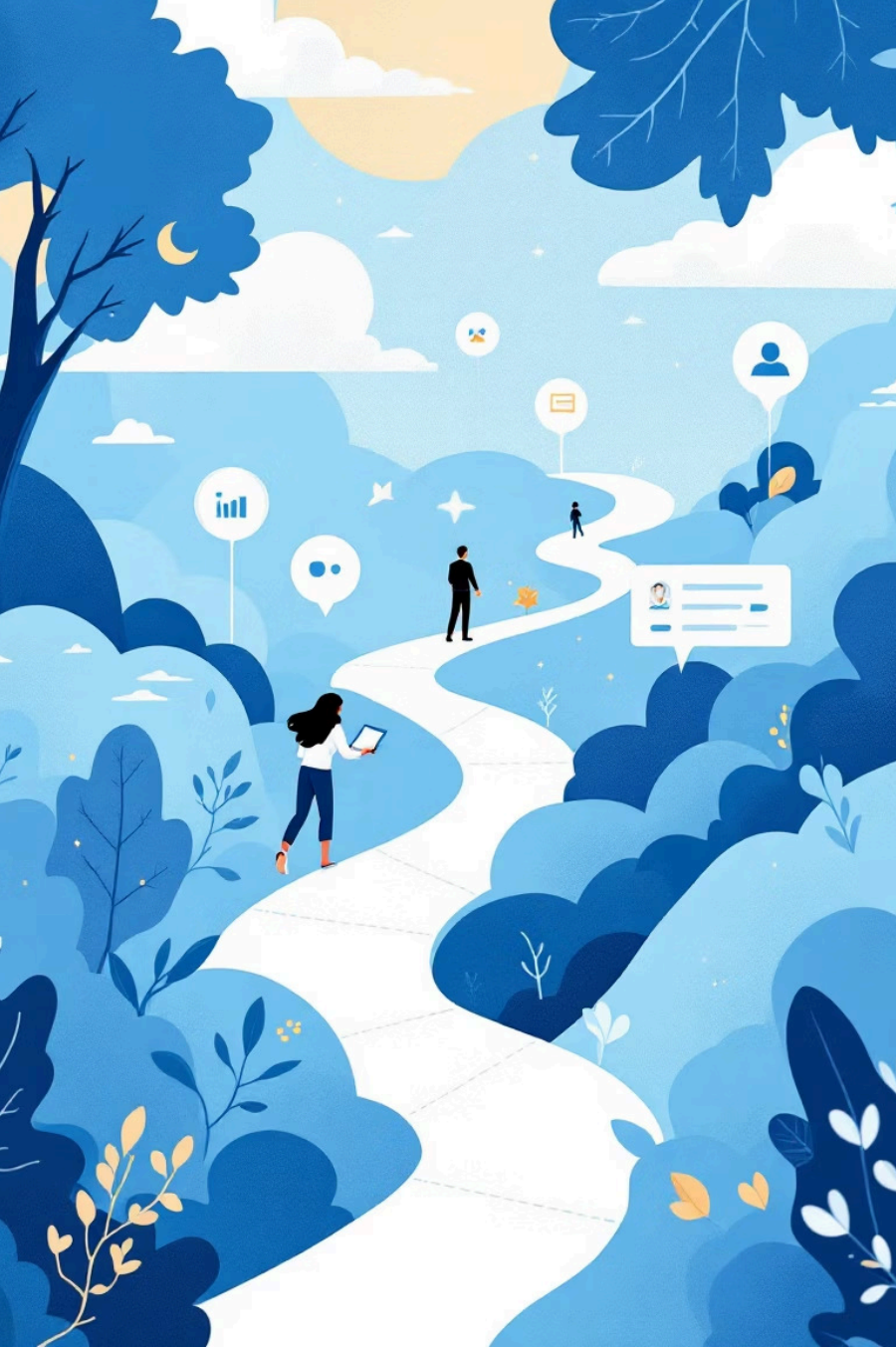
Regras de associação revelam conexões fortes entre características demográficas e de cluster.

3

## Insights Acionáveis

Base sólida para personalização de marketing, otimização de recursos e melhoria da experiência do cliente.





# Recomendações Estratégicas

## Marketing Personalizado

- Campanhas específicas por segmento.
- Cross-selling baseado em regras de associação.
- Comunicação direcionada por perfil.

## Otimização de Recursos

- Alocação de orçamento por potencial de cluster.
- Priorização de segmentos de alto valor.
- Ajuste do mix de produtos por perfil.

## Experiência do Cliente

- Personalização da jornada de compra.
- Sistema de recomendação inteligente.
- Programas de fidelidade customizados.

# Ações Táticas por Segmento

1

## Entusiastas (Cluster 0)

Opções de parcelamento, programa de pontos e cashback, promoções frequentes.

2

## Conservadores (Cluster 1)

Produtos de qualidade premium, garantias estendidas, comunicação sobre durabilidade.

3

## Premium (Cluster 2)

Lançamentos exclusivos, atendimento VIP, programa de fidelidade exclusivo.

4

## Moderados (Cluster 3)

Mix balanceado de produtos, promoções sazonais, cross-selling.

5

## Econômicos (Cluster 4)

Campanhas de cupons e descontos, produtos de entrada, comunicação focada em valor.

# Referências e Alinhamento com a Literatura

## Referências Chave

KUMAR, AMIT (2022): Customer Segmentation of Shopping Mall Users Using K-Means Clustering.

AGRAWAL & SRIKANT (1994): Fast Algorithms for Mining Association Rules.

HAN, KAMBER & PEI (2011): Data Mining - Concepts And Techniques.

CHEN et al. (1996): Data mining: an overview from a database perspective.

KUMAR, V.; RAJAN, B. (2012): Social coupons as a marketing strategy...

## Análise de Alinhamento (Comparação)

### Coerência no Número de Clusters (KUMAR, 2022)

O estudo de Kumar (2022), que segmenta o mesmo tipo de cliente, obteve **5 clusters ótimos** para Renda vs. Score. O resultado do seu estudo (K=5) é **quase idêntico**, validando a estabilidade e a relevância desse número de segmentos.

### Validação Metodológica (AGRAWAL & SRIKANT, 1994)

O uso de **K-Means** (WEDEL & KAMAKURA, 2000) e **Apriori** (AGRAWAL & SRIKANT, 1994) em conjunto, conforme o ciclo de Mineração de Dados (HAN, KAMBER & PEI, 2011), demonstra que o estudo está **firmemente ancorado** nos fundamentos teóricos da área.

### Insights Similares

- ❗ A identificação de perfis como "**Conservadores**" (Alta Renda, Baixo Gasto) e "**Entusiastas**" (Baixa Renda, Alto Gasto) reflete achados comuns em estudos com o dataset *Mall Customers*, confirmando a validade dos perfis encontrados.