



análise de dados

# E-COMMERCE

*Insights estratégicos utilizando-se de uma  
base de dados transacional*

# Fundamentação

O principal objetivo da análise de Dados é possibilitar acesso interativo aos dados, permitindo sua manipulação e oferecendo a analistas e tomadores de decisão a capacidade de conduzir análises apropriadas.

Com a análise de dados, situações e desempenhos, obtêm-se vislumbres valiosos para a tomada de decisões de maneira mais embasada e assertiva. Tal processo se baseia na transformação de dados em informações, depois em decisões, finalizando-se em ações.



Etapas:

- 1 – Dados
- 2 – Clientes
- 3 – Produtos

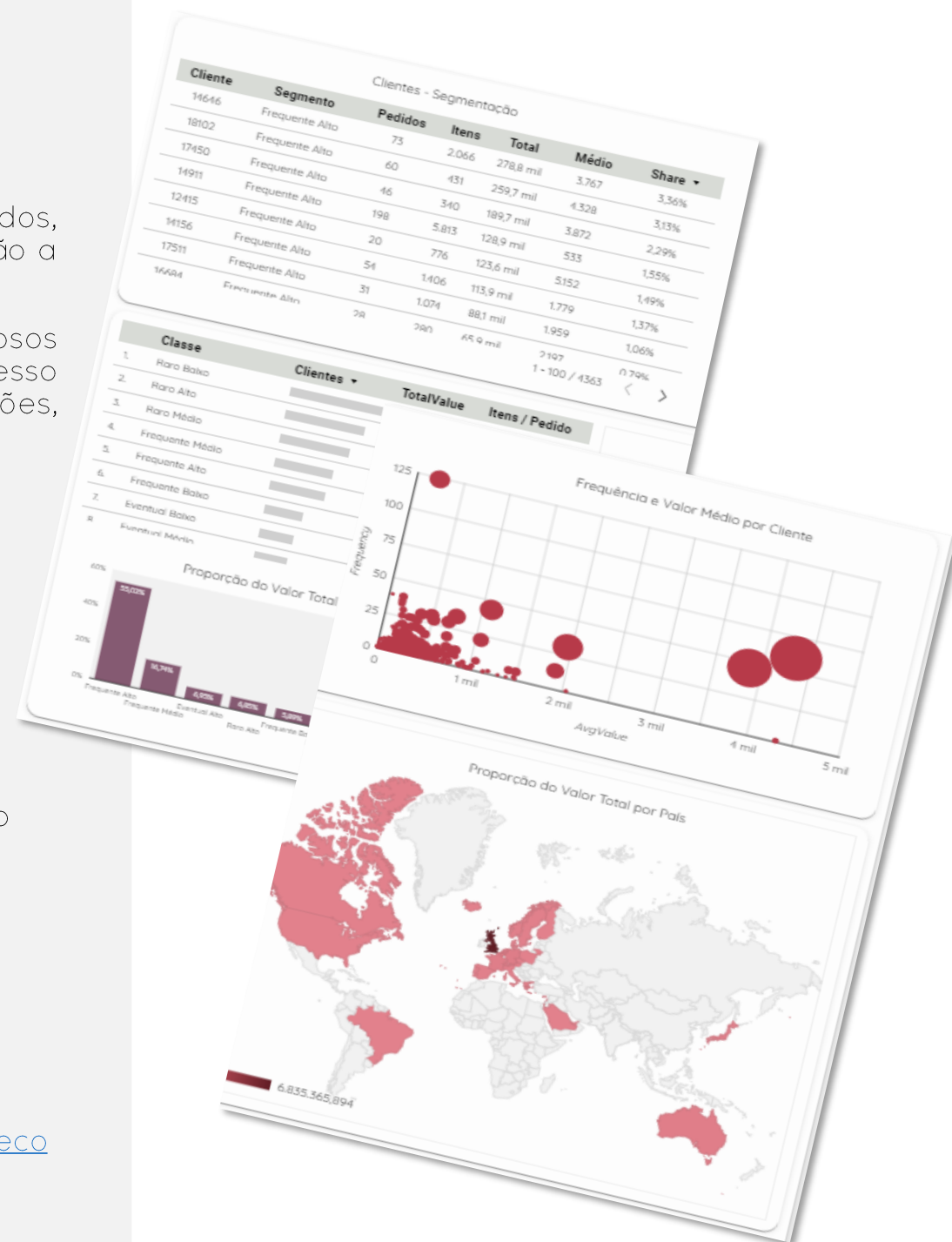
Conjuntamente à apresentação, foi criada uma dashboard e um notebook para a documentação do código utilizado para a elaboração da análise.

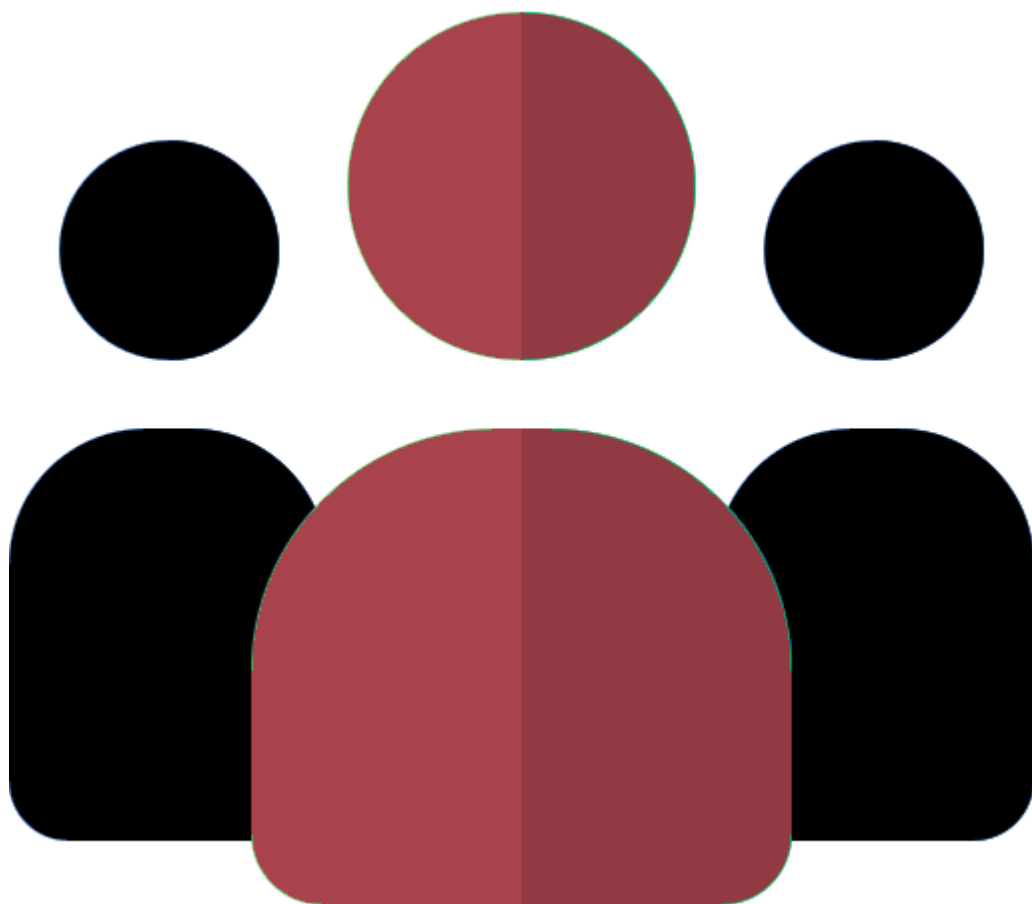
Dashboard:

<https://datastudio.google.com/reporting/0a10920d-faf6-4a36-ad01-979ebeb66595/page/levWC>

Códigos (linguagem R):

[https://github.com/viniciusbelchior0/dataanalytics\\_ecommerce/blob/main/Ecommerce\\_analysis.ipynb](https://github.com/viniciusbelchior0/dataanalytics_ecommerce/blob/main/Ecommerce_analysis.ipynb)





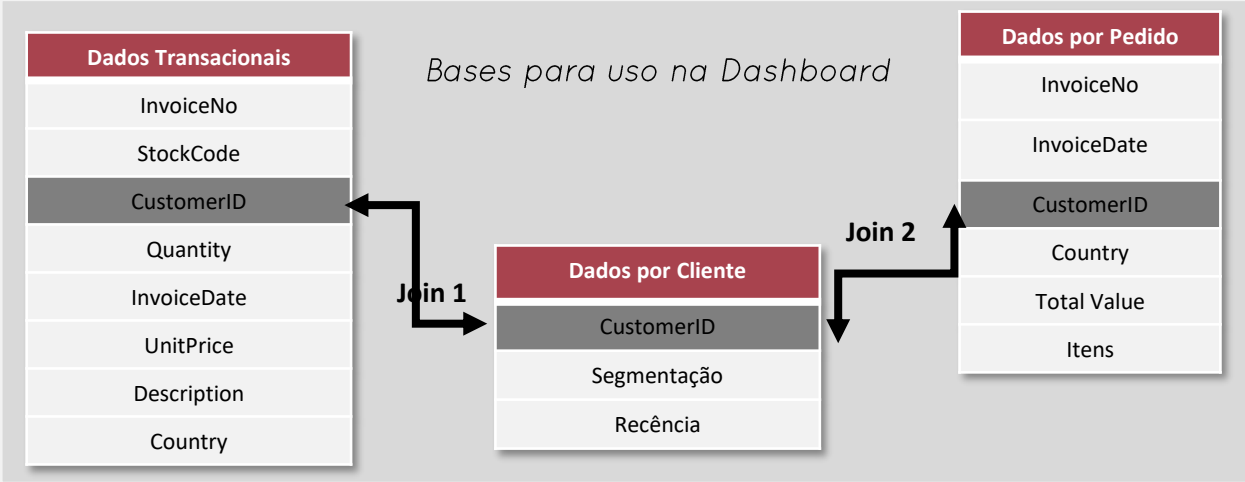
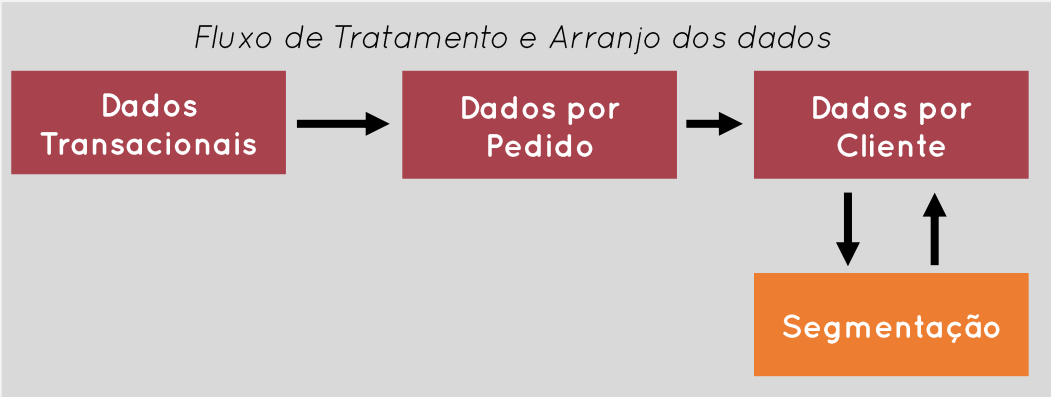
# Dados

O objetivo dessa análise é de estender uma base de dados transacional - com registros representando itens de determinado pedido - para outras bases de modo a transformar os dados para cada análise que possa ser interessante para o negócio.

A extensão dessa base em outras permite a extração de novas informações que anteriormente não estavam disponíveis. As múltiplas bases são capazes de trazer insights valiosos para cada olhar de análise do negócio.

A base original de transações permite-se ter noções dos padrões de consumo granulares do cliente. Sua agregação na base de clientes, traz consigo a perda dessas informações granulares, mas ganha ao adicionar à informação referente ao padrão geral dos consumidores, não disponível anteriormente; o mesmo argumento vale para a base dos produtos e dos pedidos.

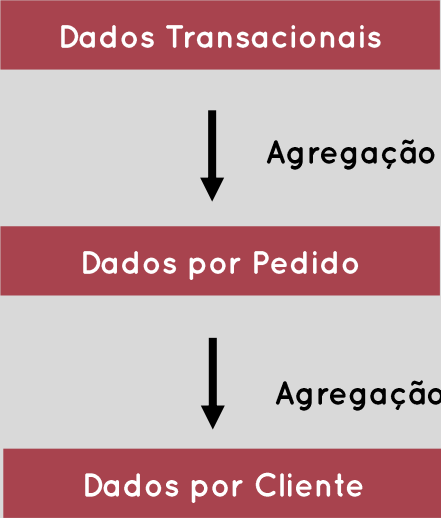
# Etapas da análise



Os dados provenientes das transações serão processados e modificados durante um fluxo para que novas medidas e informações possam ser obtidas e utilizadas para análise posterior.

Variáveis	Descrição	Tipo	Insights
InvoiceNo	ID do Pedido	Categórica	Realizar agregações e filtragens com base nos diferentes pedidos
StockCode	ID do Produto	Categórica	Realizar agregações e filtragens com base nos diferentes produtos
Description	Descrição do Produto	Categórica	Descrição do produto; sem utilidade analítica
Quantity	Quantidade	Contínua	Medidas descreiivas; agregações
InvoiceDate	Data do Pedido	Data	Comportamento temporal
CustomerID	ID do Cliente	Categórica	Realizar agregações e filtragens com base nos diferentes pedidos
Country	País	Categórica/Geográfica	Comportamento geográfico

Dados originalmente de: <https://archive.ics.uci.edu/ml/datasets/online%26Bretail>

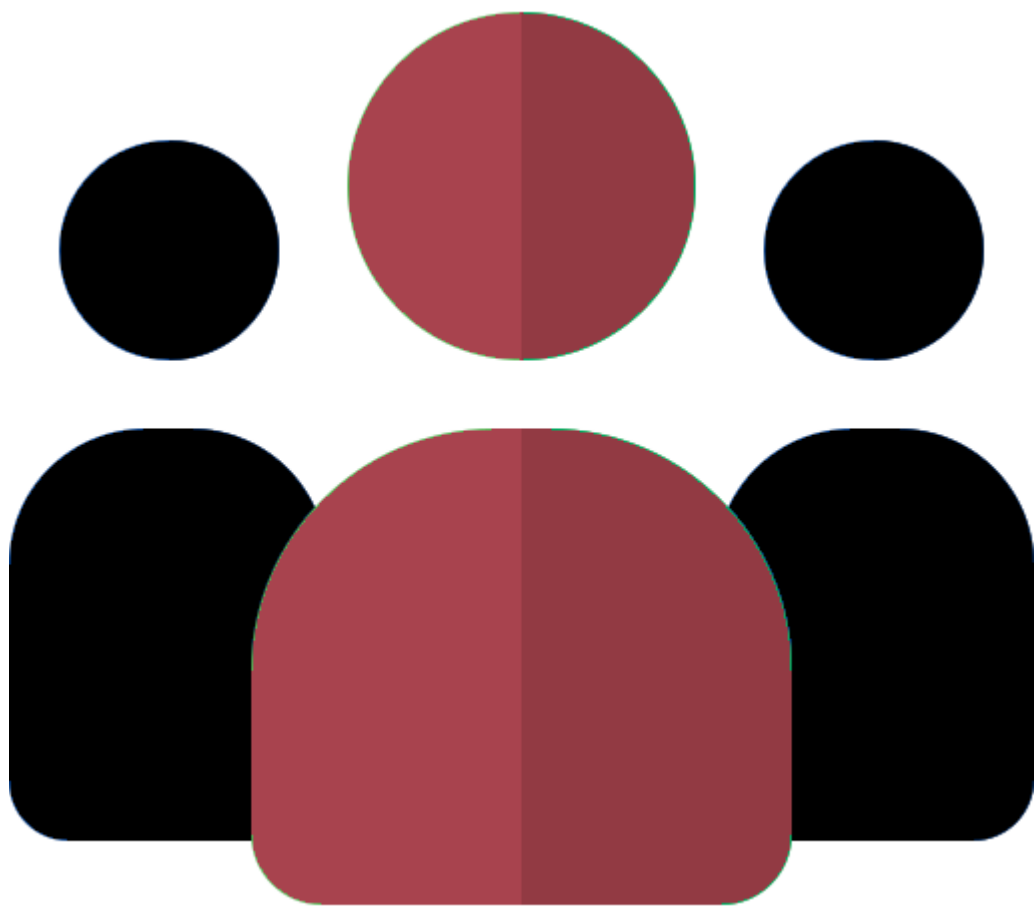


Dados por Pedido

Variáveis	Descrição	Tipo	Insights
InvoiceNo	ID do Pedido	Categórica	Realizar agregações e filtrações com base nos diferentes pedidos
InvoiceDate	Data do Pedido	Data	Comportamento temporal
CustomerID	ID do Cliente	Categórica	Realizar agregações e filtrações com base nos diferentes clientes
Country	País	Categórica	Comportamento geográfico
Valor Total	Valor Total do Pedido	Contínua/Geográfica	Medidas Descritivas (Soma, média, variação); Índice para gastos do cliente
Transações/Itens	Número de Itens Diferentes	Discreta	Medidas descritivas (Soma, média, variação); Índice para diversidade de produtos de um cliente

Dados por Cliente

Variáveis	Descrição	Tipo	Insights
CustomerID	ID do Cliente	Categórica	Realizar agregações e filtrações com base nos diferentes clientes
NInvoices	Quantidade de pedidos	Discreta	Medidas Descritivas; Indica a frequência do cliente
Total Value	Valor Total gasto pelo cliente	Contínua	Medidas descritivas; Indicativo do potencial financeiro absoluto
Total Transactions/Itens	Total de Itens/Transações do cliente	Discreta	Medidas descritivas; Indicativo do potencial financeiro relativo
AvgValueInvoice	Valor Médio por Pedido	Contínua	Medidas Descritivas (Soma, média, variação); Índice para gastos do cliente
DaysLastInvoice	Quantidade de dias desde a última compra	Discreta	Índice para o status do cliente (ativo ou inativo)



# Clientes

A segmentação dos clientes é de extrema importância para os negócios, para que os mesmos possam reconhecer pessoas/tipos em seus clientes e com isso possam direcionar suas ações de maneira assertiva para cada grupo de cliente.

Para que isso possa ser realizado é necessário o conhecimento de características (variáveis) desses clientes. Perfis demográficos e de consumo são extensamente utilizados. Para nossa análise, utilizaremos a base agregada de clientes, com informações à respeito de seus hábitos de consumo.

# Segmentação 1: Clusterização

Essa técnica é extensamente utilizada quando há a intenção de verificar a existência de comportamentos semelhantes entre observações. Seu principal objetivo é a alocação de observações em uma quantidade relativamente pequena de agrupamentos homogêneos internamente e heterogêneos entre si e que representem o comportamento conjuntos das observações a partir de determinadas variáveis.

As técnicas de agrupamento são técnicas exploratórias, sem caráter preditivo e/ou inferencial. O método não é generalizável, uma vez que a inclusão de novas observações na amostra torna necessária a reaplicação da modelagem, para que sejam gerados novos agrupamentos. Em um ambiente com mudanças muito dinâmicas, talvez seja necessária a sucessiva aplicação dessas modelagens, que podem se tornar insustentáveis devido às suas incessantes mudanças.

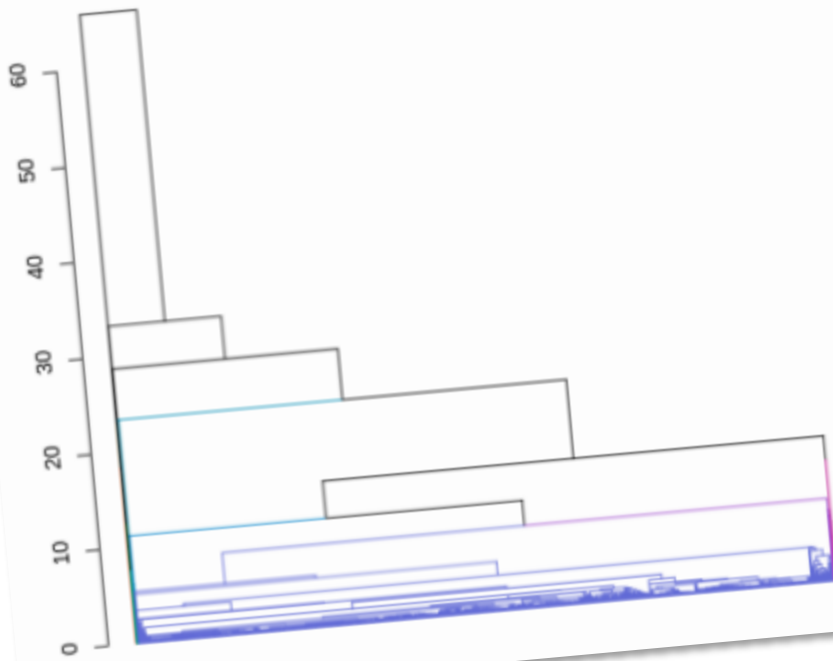
A atribuição dos rótulos foi feita utilizando-se a distância euclidiana, com método hierárquico de ligação média (average linkage).

Pode-se observar na imagem e na tabela ao lado a elaboração dos agrupamentos realizada pelo método. Valores discrepantes (outliers) influenciam a técnica e acabam por trazer resultados insatisfatórios para a tomada de decisão. Nota-se que os cluster podem ser interpretados como: 2 – “clientes gerais”; 3 – “clientes especiais”; 1 – “caso especial”.

Explicação mais detalhada no código:

[https://github.com/viniciusbelchior0/dataanalytics\\_ecommerce/blob/main/Ecommerce\\_analysis.ipynb](https://github.com/viniciusbelchior0/dataanalytics_ecommerce/blob/main/Ecommerce_analysis.ipynb)

Processo para estabelecimento dos Clusters



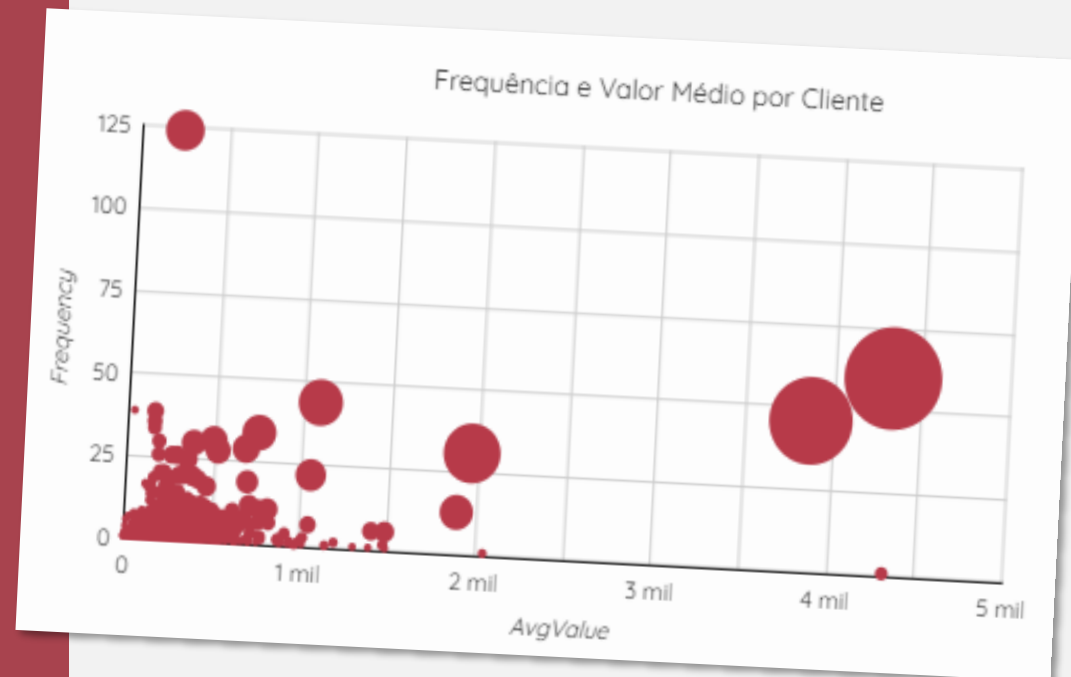
Cluster	Clientes	Pedidos (Média)	Valor Total	Valor Médio	Dias da última compra (Média)
2	4329	4	379	379	92
3	4	136	1252	1252	1
1	2	2	80710	80710	163

# Segmentação 2: RFM

Os métodos de agrupamento necessitam de reaplicação quando da inclusão de novos registros e variáveis. Utilizar um score determinado pode-ser considerada uma alternativa para lidar com esse problema. Três variáveis são interessantes para essa maneira de segmentação: **Recência**, **Frequência** e **Valor Médio**.

- **Recência:** data da última compra realizada pelo cliente. É um indicativo de atividade do cliente. Os clientes podem ser considerados ativos ou inativos. Tal métrica é necessária para que se possa filtrar e analisar mais assertivamente apenas os clientes ativos para tomada de ações;
- **Frequência:** o número de pedidos/compras realizados pelo cliente; É um indicativo da fidelidade do cliente
- **Valor médio:** o valor médio gasto nos pedidos realizados pelo cliente. É um indicativo do potencial financeiro do cliente.

A atribuição dos rótulos foi feita através dos valores apresentados pelos clientes. Tais valores foram determinados pela distribuição dos valores (quantis) das variáveis. Foram usados 3 quantis para a Frequência e Valor Médio e apenas 2 para a Recência.



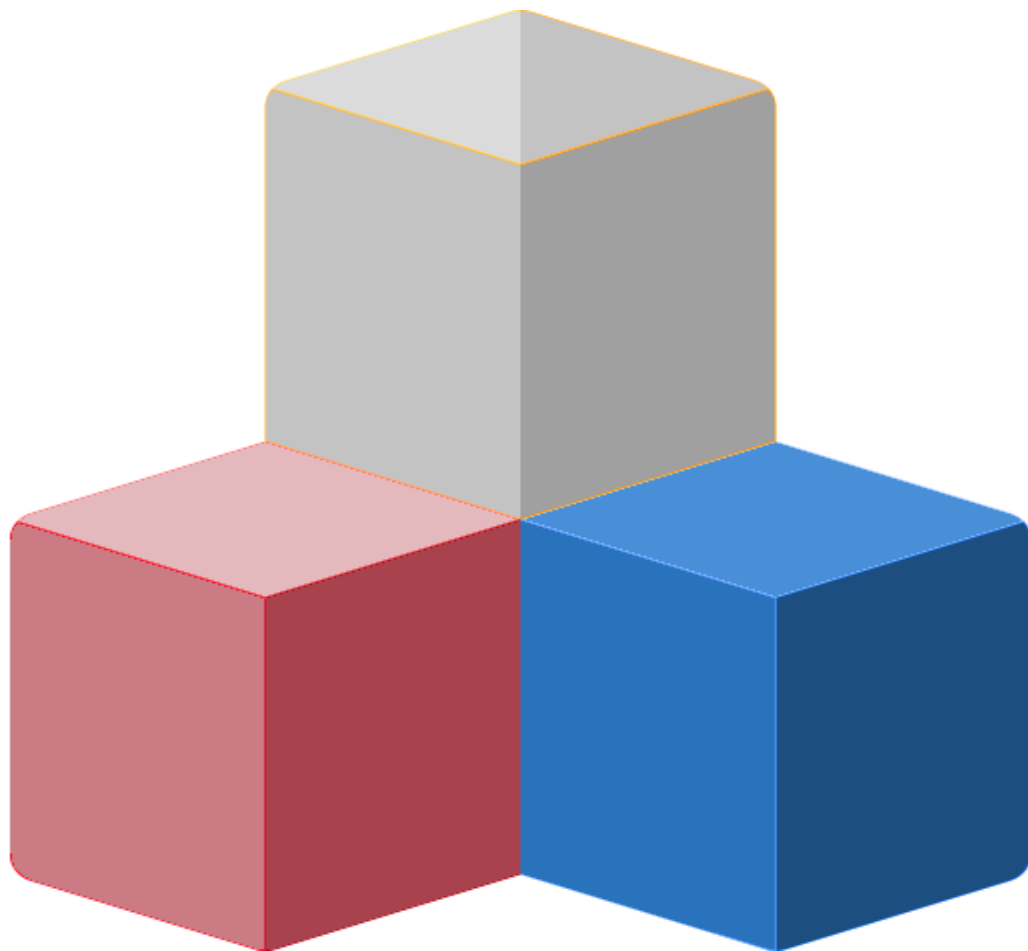
Intervalo dos Dados	Valor (Dias)	Rótulo
0 -50%	$0 \leq x \leq 50$	Ativo
51 – 100%	$50 < x$	Inativo

Intervalo dos Dados	Valor (Pedidos)	Rótulo
0 -33%	$x \leq 1$	Raro
34 – 66%	$1 < x \leq 4$	Eventual
67 – 100%	$4 < x$	Frequente

Intervalo dos Dados	Valor (Valor Médio)	Rótulo
0 -33%	$0 \leq x \leq 209$	Baixo
34 – 66%	$209 < x \leq 363$	Médio
67 – 100%	$363 < x$	Alto



# Produtos



Analisar o portfólio de produtos – o mix total de produtos oferecidos pela organização – é essencial para se tomar decisões de posicionamento .

O mix pode ser caracterizado de acordo com sua participação de mercado, receitas e potencial de crescimento. Relações de consumo entre os itens também trazem conhecimentos úteis para a elaboração de novas estratégias.

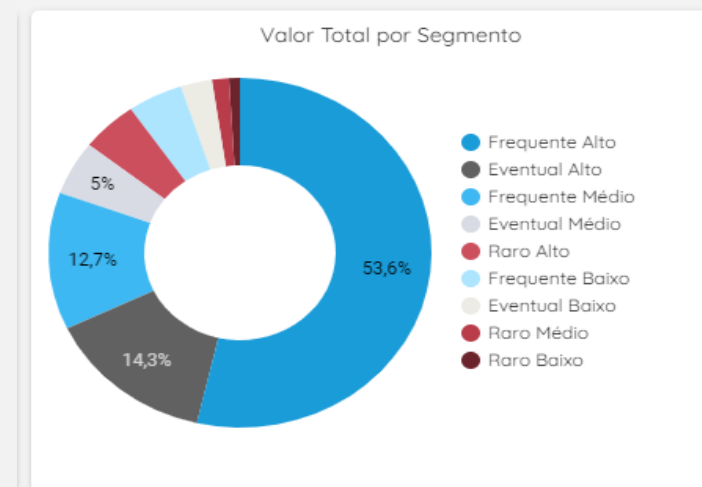
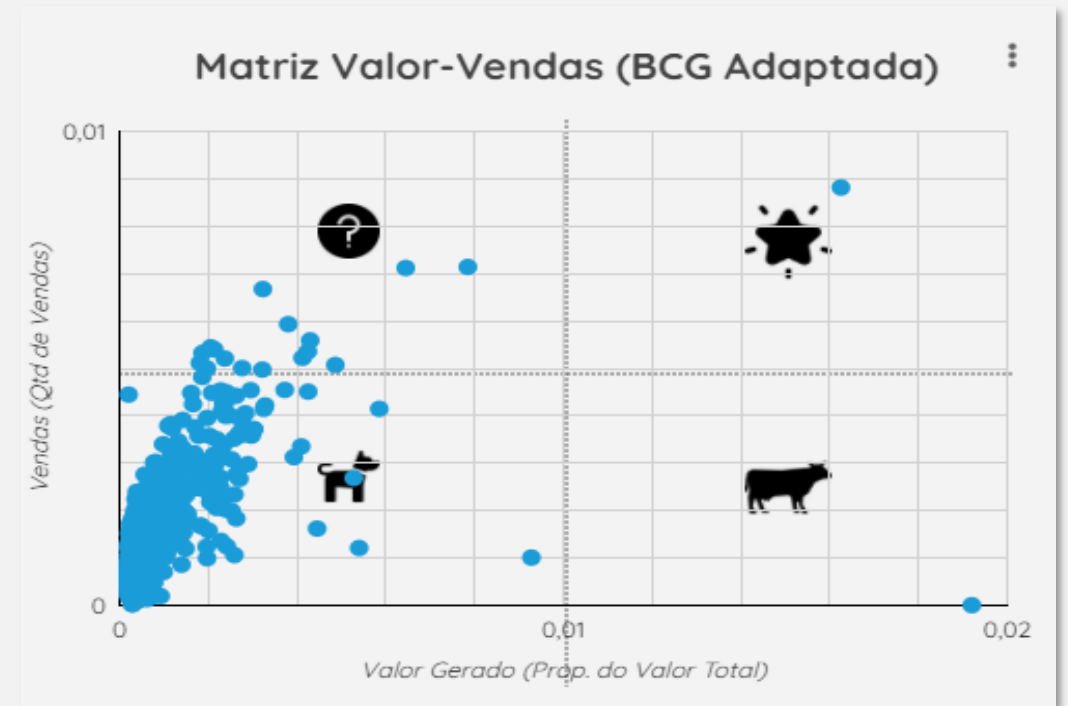
# Matriz Valor-Vendas

Uma das maneiras interessantes de se olhar para os produtos é em relação às suas vendas e ao valor com o qual produto contribui para a empresa. Um das técnicas mais amplamente utilizadas é a Matriz BCG – usada para categorizar produtos em termos de crescimento e participação de mercado. Como não temos informações à respeito dos valores totais do mercado, é necessária uma adaptação.

A participação de mercado será substituída pela proporção do valor total gerado e o crescimento do mercado será substituído pelo número de vendas do produto. Através dessa adaptação, os produtos terão um rótulo.

1. **Cão:** baixa participação nas vendas e baixo número de vendas. Não são produtos que merecem atenção individual e tratamento especial.
2. **Interrogações:** Possuem baixo valor gerado, mas alta quantidade de vendas. Podem ter potencialmente de crescimento ou podem decair.
3. **Vaca:** Baixo número de vendas e respondem por uma boa proporção do valor total. Possuem forte presença e geram receita sólidas.
4. **Estrela:** produtos com muitas vendas e alta participação no valor gerado. São os produtos especiais que merecem atenção individual.

Também é interessante analisar os segmentos de clientes que mais consomem tal produto. São informações que podem ser redirecionadas para ações específicas a cada tipo de cliente.



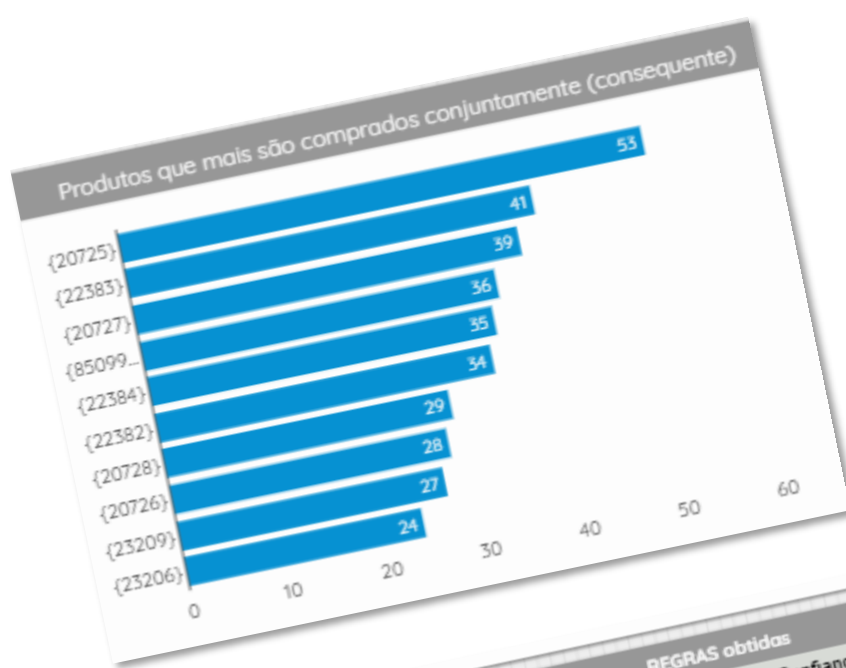
# Market Basket Analysis

A principal ideia na análise de cesta de mercado é identificar relações entre produtos que usualmente são comprados em conjunto.

A análise de cesta de mercado auxilia na necessidade de inferir necessidades e preferências dos clientes a partir de seus padrões de consumo. Ela pode ser usada como uma ferramenta para o planejamento de ações de cross-selling, up-selling, para melhorar a organização do espaço físico das lojas, entre outras aplicações.

Para a obtenção das associações entre os produtos, são utilizados três parâmetros:

- **Suporte:** Quantidade de vezes que os produtos são comprados conjuntamente em relação à todas as compras(transações). É usada para melhor alocação da técnica, mas não representa muito valor informacional já que em mix com muitos produtos, a proporção das associações será muito baixa.
- **Confiança:** É a probabilidade de encontrar o Item 2 em transações em que o Item 1 esteja presente.
- **Elevação:** mede o quão provável o item 2 será comprado com o Item 1. Métrica de grande valor informacional, quando os valores são maiores que 1, é um indicativo de que os itens são muito prováveis de serem comprados conjuntamente.



ID_Asc	X (Antecedente)	link	REGRAS obtidas		Elevação
			Y (Consequente)	Confiança	
675	{23170,23171}	=>	(23172)	<div></div>	66,41
674	(23170,23172)	=>	(23171)	<div></div>	64,04
1	(23171)	=>	(23171)	<div></div>	61,5
2	(22745,22748)	=>	(23172)	<div></div>	61,5
678	(23175)	=>	(22746)	<div></div>	53,51
5	(23174)	=>	(23174)	<div></div>	52,03
6	(23171,23172)	=>	(23175)	<div></div>	52,03
673	(22746,22748)	=>	(23170)	<div></div>	51,41
677	(23172)	=>	(22745)	<div></div>	50,41
3	(23170)	=>	(23170)	<div></div>	49,37
4	(22745,22746)	=>	(23172)	<div></div>	49,37
676	(23170)	=>	(22748)	<div></div>	48,28
32	(23170)	=>	(23171)	<div></div>	47,61