

# Clusterização de dados: segmentação de clientes

# Segmentação de clientes

## Fatores que influenciam na venda de um produto:

- Localização e tamanho da loja;
- Visibilidade do produto;
- Preço;
- Promoções;
- Disponibilidade.

O marketing de sucesso sempre foi oferecer o produto certo para a pessoa certa no momento certo.

## Types of customer segmentation



Demographic



Geographic



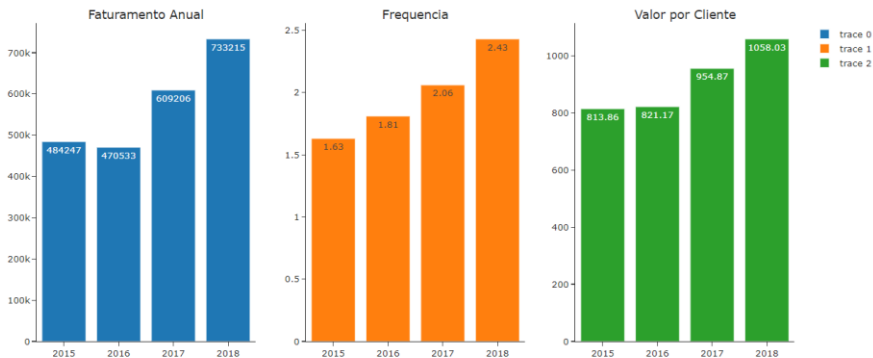
By behaviour



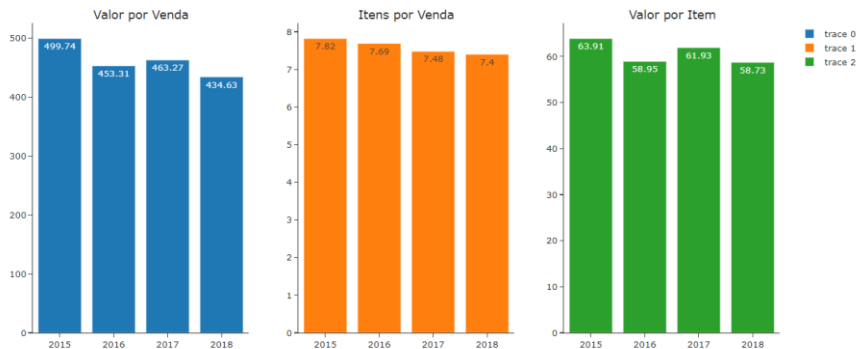
Customer Journey

# Overview

Comparação 2015 - 2018



Comparação 2015 - 2018



## Dataset

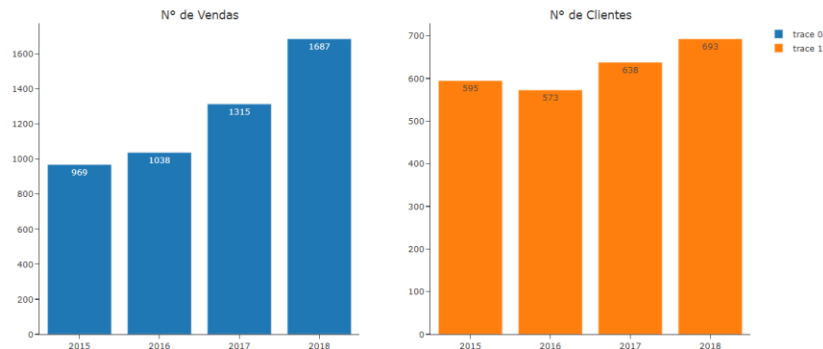
Dados de 4 anos de vendas identificadas de uma empresa de varejo de materiais de escritório, móveis e eletrônicos.

## Análise dos Indicadores

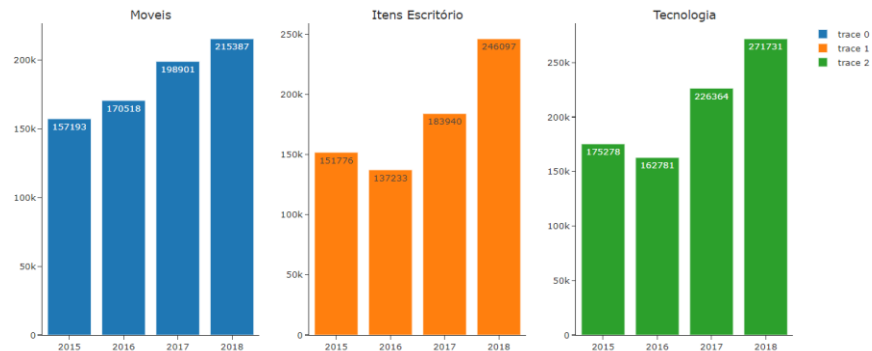
- Receita anual cresceu impulsionada por aumento de gasto por cliente, que por sua vez é impulsionada pelo aumento da frequência de compra dos clientes;
- Valor por venda está caindo devido queda de itens por venda e pelo preço médio dos itens;

# Overview

Comparação 2015 - 2018



Comparação 2015 - 2018 por Categoria



## Análise dos Indicadores

- Número de clientes está subindo anualmente, mas não está na mesma velocidade do aumento de faturamento;
- Móveis possui crescimento constante, entretanto, itens de escritório e de tecnologia foram os responsáveis pelo maior crescimento.

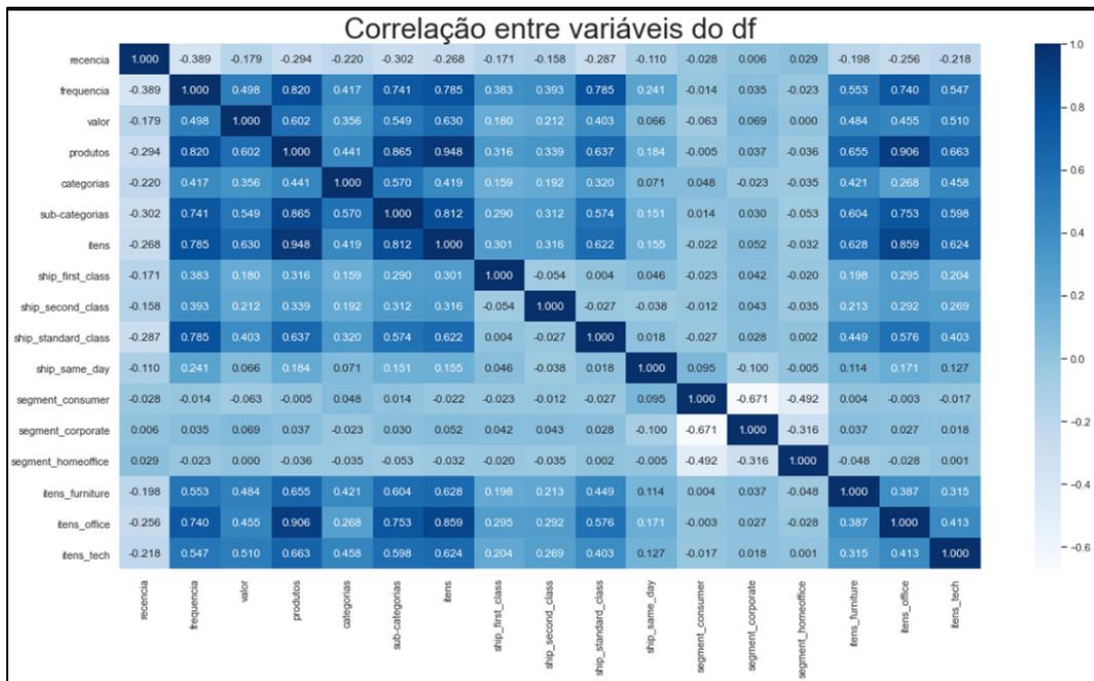
## Problemas

- Nem todas categorias acompanharam o crescimento de faturamento da empresa, havendo oportunidades em algumas delas;
- O número de clientes e a frequência de compras vem aumentando, porém o valor por compra vem caindo devido redução da quantidade de itens por venda e também do preço médio dos itens.

## Próximos passos

- Aumentar o valor por compra dos clientes, gerando maior receita para a empresa com o mesmo número de clientes através de campanhas personalizadas de relacionamento com o cliente;
- Produtização do modelo: cliente receber campanhas personalizadas ao entrar no site da empresa.

# Correlação entre as Variáveis

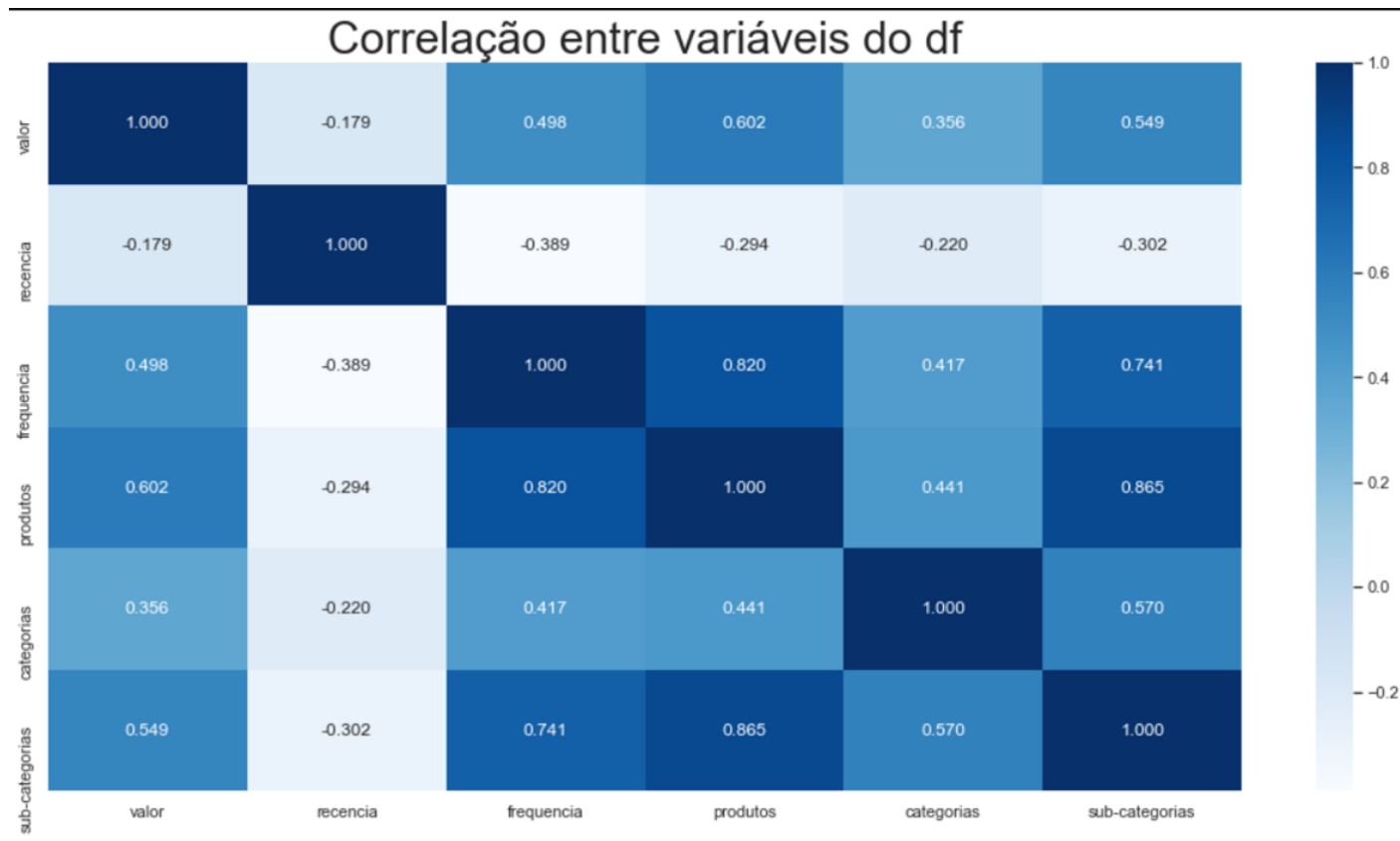


Altas correlações entre valor & frequência, nº produtos, nº categorias, nº sub-categorias, nº itens, frete tipo padrão.

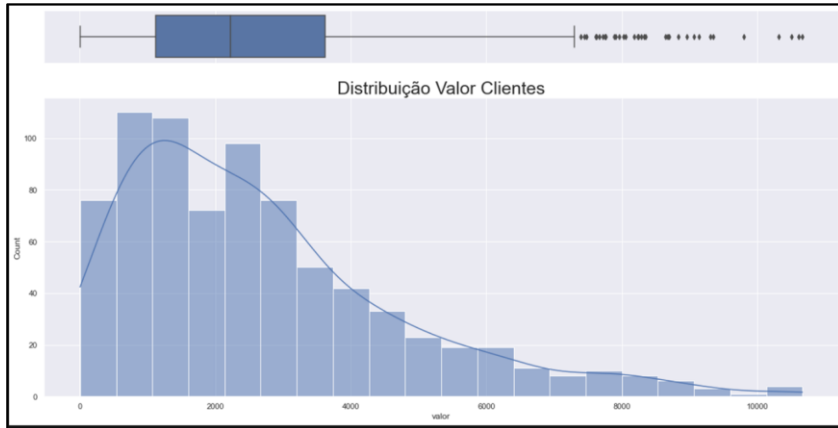
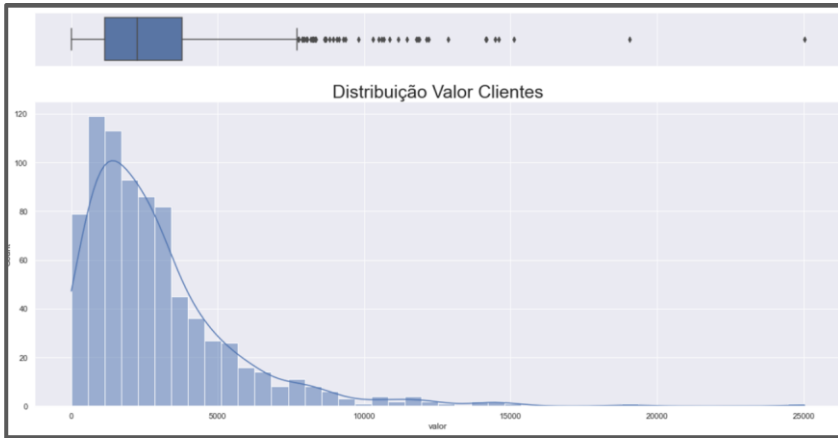
Valor com correlação inversa com a recência (tempo em dias desde a última compra), ou seja, clientes que compraram recentemente tem maior tendência a terem maior valor gasto.

Baixa correlação de valor com fretes primeira classe, segunda classe, mesmo dia e também com o tipo de cliente.

# Correlação entre as Variáveis



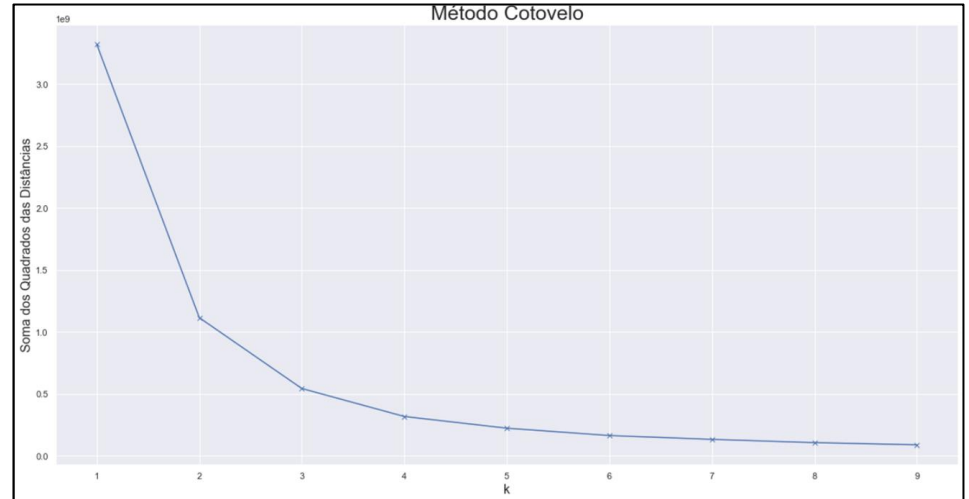
# Tratamento de Dados e Modelagem



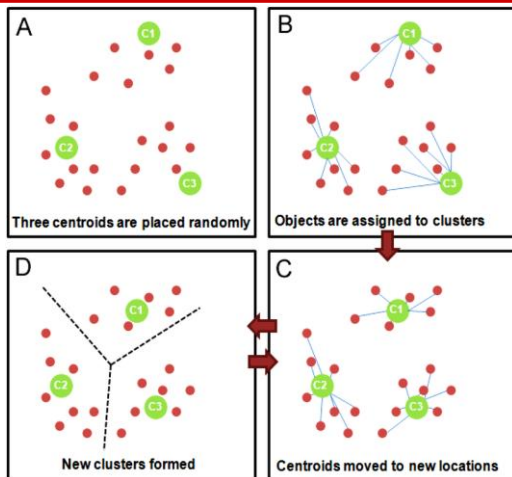
Remoção de outliers - clientes com gasto muito elevado.

Tratamento de variáveis categóricas.

Definição K: verificação qualitativa dos dados dos clusters (ponto de vista de negócio) e k otimizado pelo Método do Cotovelo.



# Algoritmos de Segmentação

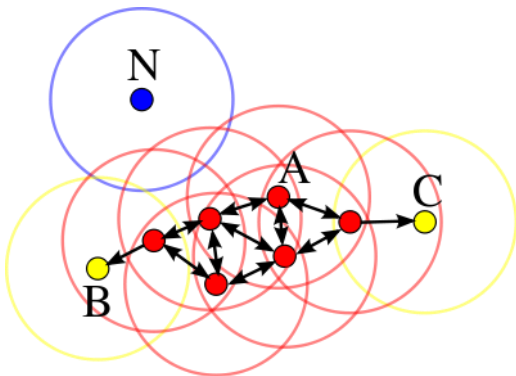


## KMeans

- Segmentação de dados em K clusters (K é escolhido ao treinar o modelo)
- Maximização de distância inter clusters
- Minimização de distância intra clusters
- Principal Vantagem: escolha do número de clusters ao treinar o modelo

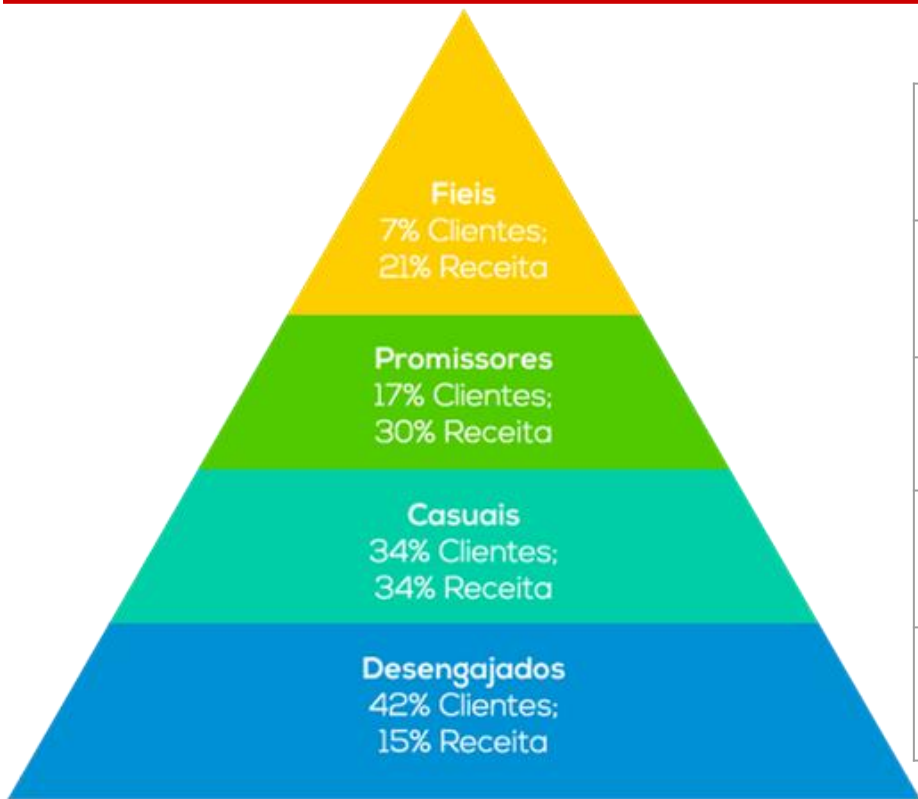
## DBSCAN

- Baseado em densidade espacial dos dados.
- Encontra clusters de formato arbitrário e diferentes tamanhos.
- Separa os ruídos dos dados



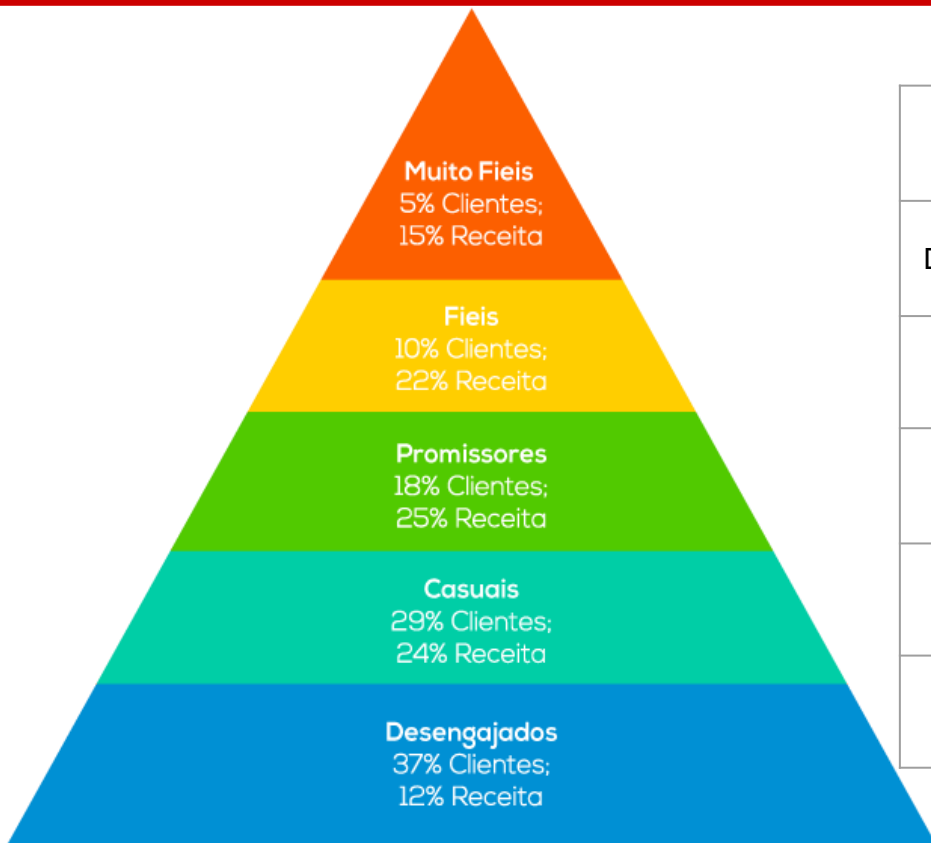


# Clusterização: KMeans com K = 4



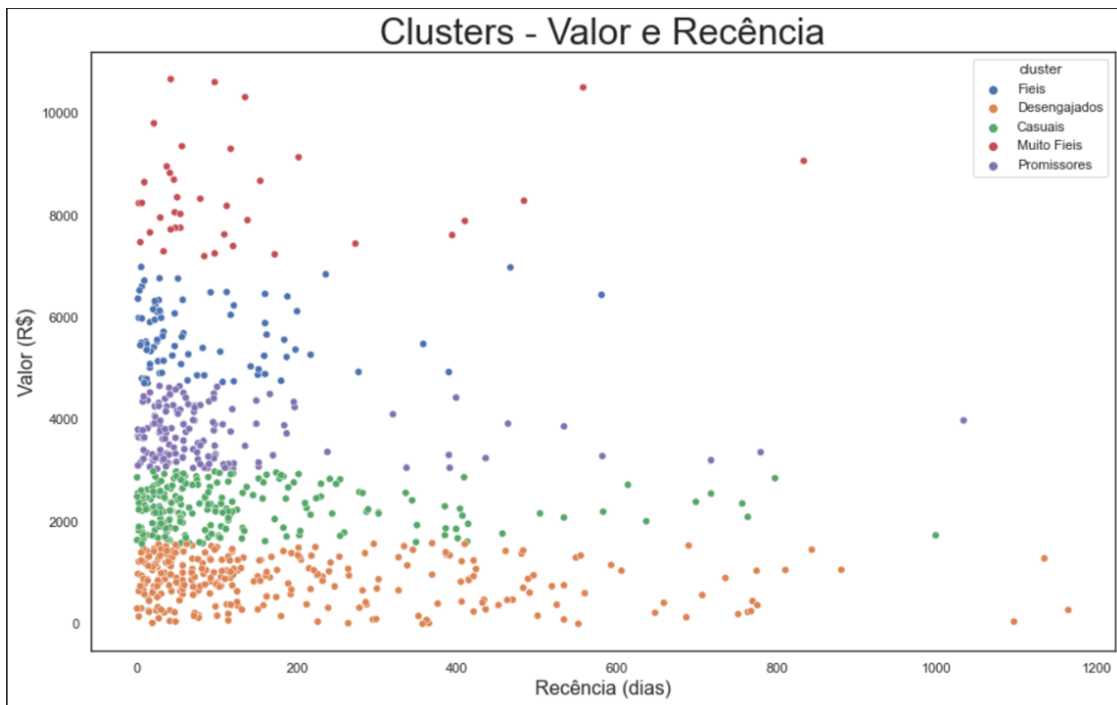
Cluster	Recência (d)	Frequência (n° compras)	Valor (R\$)
Desengajados	194	4,85	R\$ 951
Casuais	121	6,91	R\$ 2.656
Promissores	90	7,70	R\$ 4.745
Fiéis	131	8,31	R\$ 7.841

# Clusterização: KMeans com K = 5



Cluster	Recência (d)	Frequência (n° compras)	Valor (R\$)
Desengajados	200	4,71	R\$ 859
Casuais	132	6,60	R\$ 2.289
Promissores	101	7,40	R\$ 3.704
Fiéis	88	8,00	R\$ 5.639
Muito Fiéis	137	8,52	R\$ 8.405

# Escolha do Modelo



Escolhido o algoritmo KMeans e  $K = 5$

## Método Cotovelo:

- Um número maior de clusters representaria pouca redução de erro.
- Para  $k < 4$  há um grande aumento do erro.

## Negócios:

- Clusters com características distintas entre si considerando RFV (recência, frequência e valor)
- Cada cluster possui diferentes categorias “favoritas”, possibilitando recomendação personalizada.

## DBSCAN:

- Os clusters não fizeram sentido do ponto de vista de negócios.

# Produtização

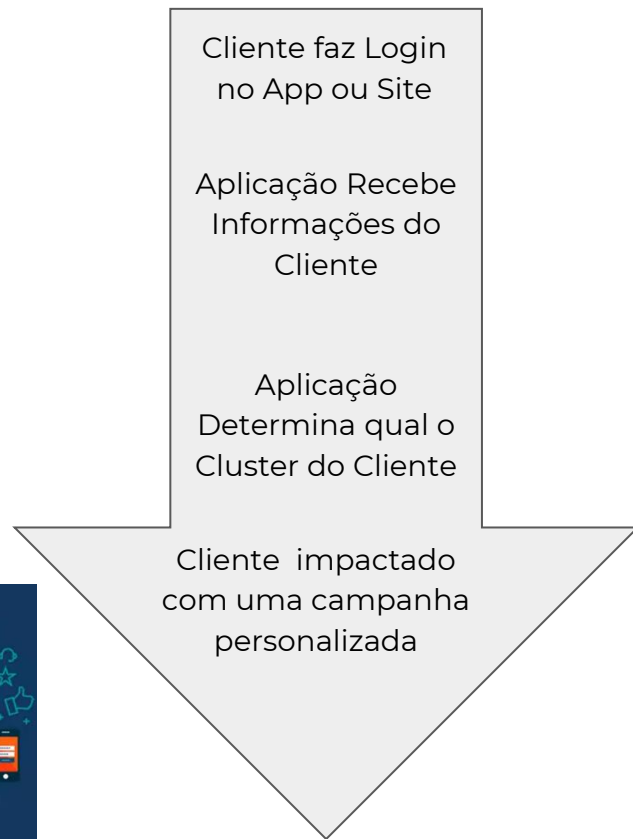
Criação de API usando o FLASK

- 1) Fazer a requisição do serviço FLASK
- 2) Importação de informações do cliente
- 3) O modelo determina qual cluster o cliente pertence
- 4) Uma mensagem personalizada de marketing é direcionada ao cliente.

## Aplicações:

Recomendações de produtos durante a navegação do cliente no site da empresa.

Envio de campanhas personalizadas via mídias digitais (exemplo: e-mail, sms, notificação).



# Produtização - Recomendação Produtos/Cupom

Produto escolhido para recomendação: produto mais comprado do Cluster

Dinâmica de Cupom: de acordo com as características de cada Cluster

Cluster	Recomendações
Desengajados	Promoção cadeira de escritório com 15% OFF usando o cupom CADEIRA15
Casuais	Filtros para purificador de ar com 10% OFF com o cupom FILTRO10, venha conferir!
Promissores	Confira os novos modelos de viva-voz de conferência, não perca tempo!
Fiéis	Mesa com altura regulável - Compre 3 e ganhe 20% de desconto no terceiro item
Muito Fiéis	Carrinho organizador para escritório com 5% OFF usando o cupom CARRINHO5, confira!