

ANÁLISE COHORT, RFV E SEGMENTAÇÃO COM K MEANS

Tópicos

01

Introdução

02

Cohort

03

Segmentação RFV

04

Clusterização K-Means

05

Conclusões

06

Recomendações

Introdução

Este projeto visa realizar uma análise de cohorts e implementar uma segmentação RFV em um e-commerce. Ao explorar as transações ao longo do tempo, são buscados insights sobre padrões de retenção e identificação de grupos de clientes. Com essas abordagens, pretende-se personalizar estratégias de marketing, impulsionar a fidelidade do cliente e maximizar o valor gerado pelo negócio online. O objetivo é oferecer uma visão estratégica para otimizar a experiência do cliente e impulsionar o sucesso do e-commerce.

Objetivos

O objetivo deste projeto é realizar uma análise detalhada de cohorts e implementar uma segmentação baseada nos critérios de Recência, Frequência e Valor (RFV) para um e-commerce.

- Através da análise de cohorts, buscamos compreender o comportamento dos clientes ao longo do tempo, identificando padrões de retenção e evolução.
- A segmentação RFV permitirá categorizar os clientes em grupos específicos, proporcionando insights valiosos para a personalização de estratégias de marketing, aprimoramento da fidelidade do cliente e otimização do valor total gerado pelo e-commerce.

O objetivo final é impulsionar o desempenho do negócio, adaptando as abordagens de acordo com as características e necessidades específicas de cada segmento de clientes.

Origem dos dados

Este conjunto de dados é composto por registro de transações, abrangendo todas as compras realizadas entre 12/01/2010 e 12/09/2011, provenientes de um varejo online sem loja física, registrado e sediado no Reino Unido.

Link de acesso: <https://archive.ics.uci.edu/dataset/352/online+retail>

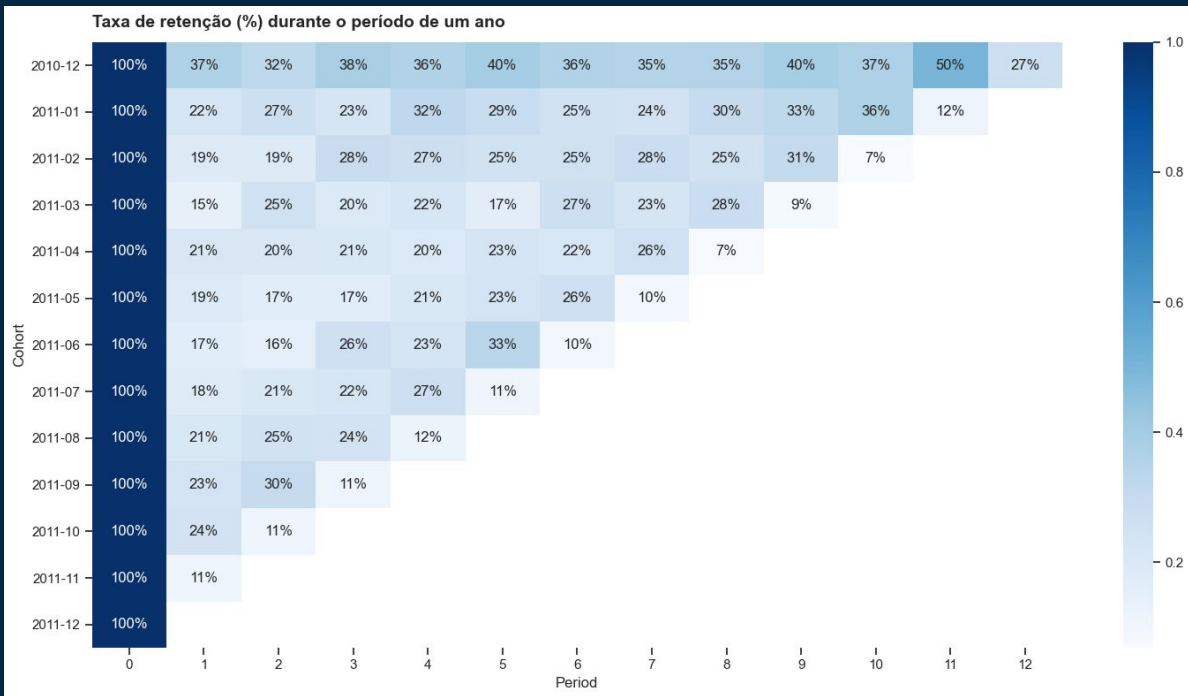
Cohort

Em janeiro de 2011, a empresa conquistou novos clientes. No entanto, após o primeiro mês, apenas 22% desses clientes continuaram a realizar compras ou retornaram. Posteriormente, observou-se um aumento para 27%, indicando que alguns clientes voltaram para efetuar novas compras.



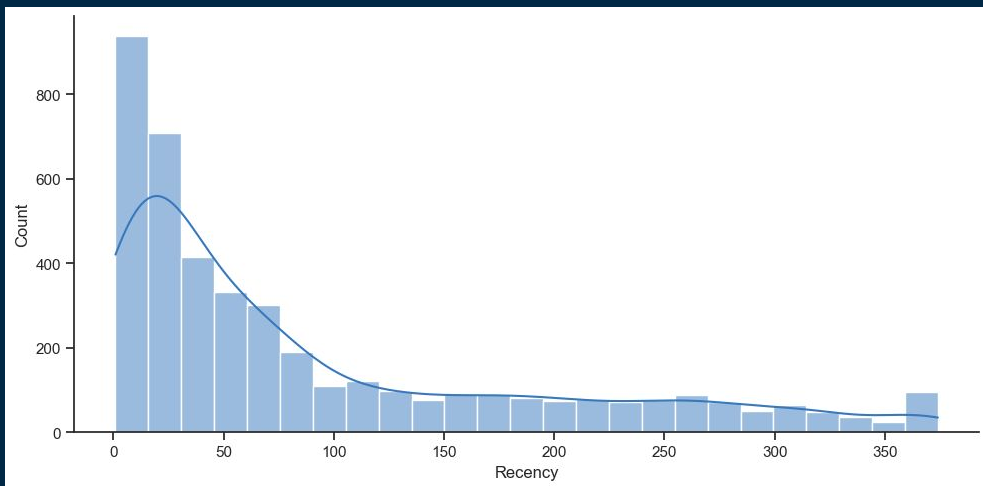
Oportunidades de Melhoria:

Para cohorts com queda acentuada na retenção, pode ser necessário revisar estratégias de retenção de clientes, entender os motivos por trás da perda de clientes e implementar ações corretivas.



RFV

Recência (Recency): Refere-se ao tempo desde a última compra de um cliente. Clientes que compraram recentemente podem estar mais propensos a comprar novamente do que aqueles que não o fizeram há muito tempo.



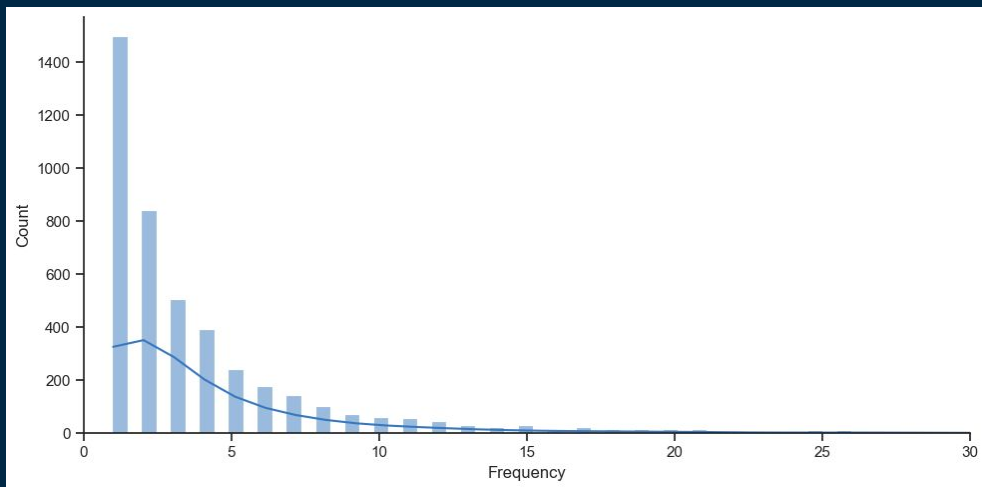
Score*	Nº de Clientes	%
1	432	9.97
2	646	14.91
3	1073	24.76
4	1059	24.44
5	1123	25.92

Assim, um cliente com **Score 1**, cuja última interação ocorreu entre **1 e 18 dias atrás**, pode ser abordado com um tipo de email diferente em comparação ao cliente com Score 2, por exemplo.

*Baseada no calculo estatístico de quintil, que consiste na ordenação dos valores e separação dos dados em 5 grupos que contenham igualmente 20% da base total.

RFV

Frequência (Frequency): Indica a frequência com que um cliente faz compras. Clientes que compram com mais frequência podem ser considerados mais engajados e leais.

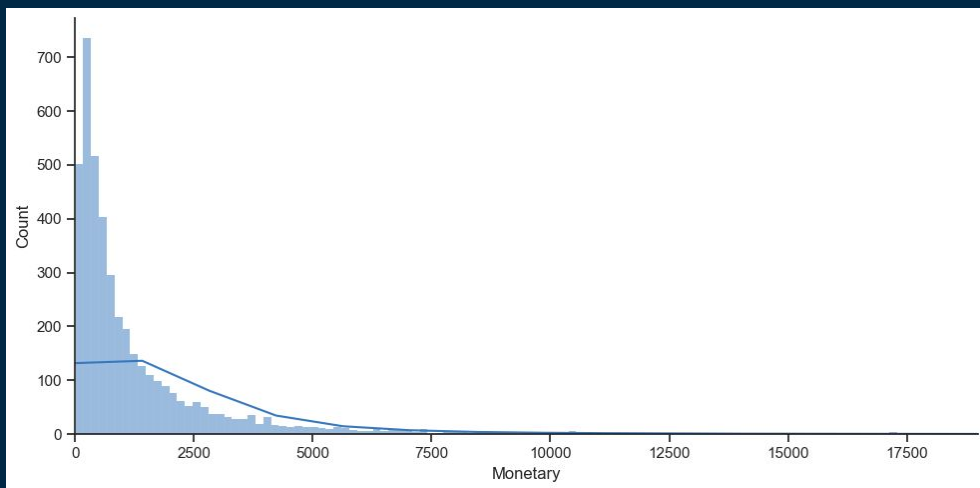


Score	Nº de Clientes	%
1	1496	34.53
2	837	19.32
3	1131	26.10
4	481	11.10
5	388	8.95

O **Score 5** refere-se a clientes com **5 ou mais pedidos únicos**, sendo considerados clientes valiosos. Essa categoria representa **8,95%** da base de dados.

RFV

Valor (Monetary Value): Relacionado ao valor total gasto por um cliente em um determinado período. Isso ajuda a identificar clientes de alto valor que podem merecer estratégias de marketing específicas.

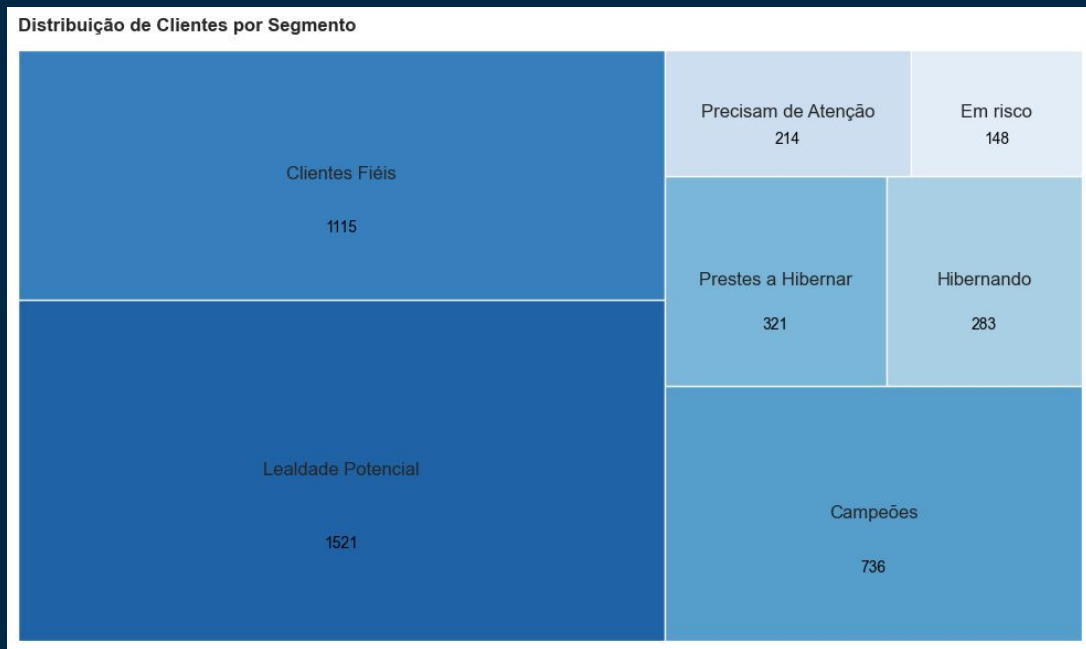


Score	Nº de Clientes	%	Valor Total
1	1083	24.99	192094.71
2	1084	25.02	499527.81
3	1083	24.99	1150786.01
4	650	15.00	1563563.94
5	433	9.99	5180771.29

O **Score 5** é atribuído a clientes que efetuam compras com valores entre 3 mil e 290 mil. Esta categoria compreende **10% da base de clientes** e contribui com **20% do valor total** gasto pelos clientes. O valor total é calculado somando-se as contribuições de cada categoria de Score.

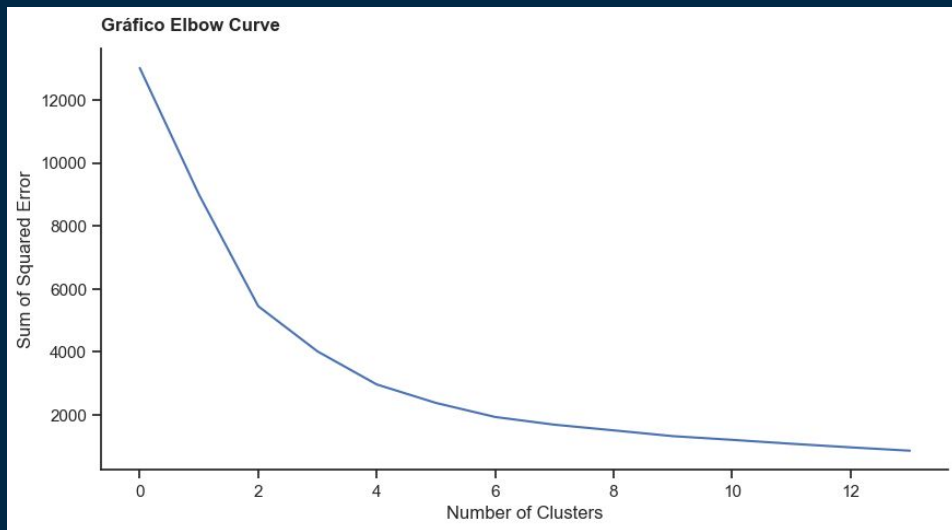
Segmentação RFV

Os grupos mais expressivos incluem "Lealdade Potencial" e "Clientes Fiéis", indicando oportunidades de fortalecimento da fidelidade. Por outro lado, os grupos "Hibernando" e "Em Risco" sugerem áreas que podem exigir estratégias específicas de reativação e prevenção. A presença de "Campeões" representa um segmento valioso que pode ser alvo de ações personalizadas para otimizar seu valor.



Clusterização K-Means

Dentro do escopo da análise de cohorts e segmentação RFV para o e-commerce, a utilização do algoritmo K-Means se destaca como uma ferramenta crucial. O K-Means nos permite agrupar clientes de acordo com suas características de recência, frequência e valor, criando perfis distintos que serão fundamentais para nossa compreensão do comportamento dos consumidores.

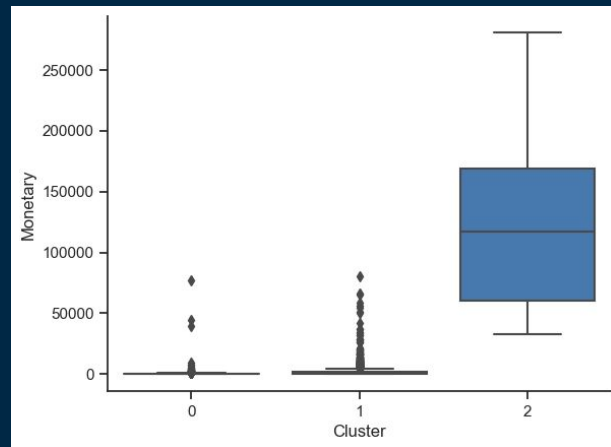
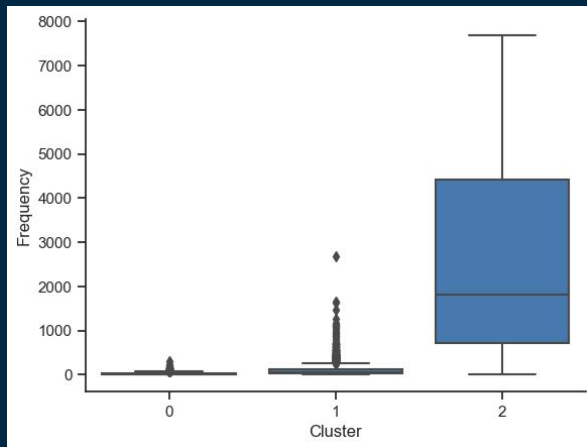
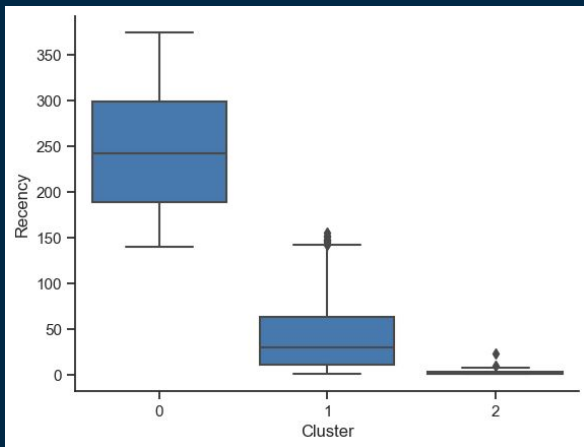


Definição de Clusters: Para determinar o número de clusters, foi utilizado o método Elbow Curve. Após uma análise cuidadosa, foi decidido optar por **3 clusters** como a configuração mais apropriada.

Clusterização K-Means

A segmentação de cluster feita com K Means dividiu os clientes em três grupos distintos:

- O **Cluster 0** representa os clientes menos engajados com a empresa. Eles compram com pouca frequência e gastam pouco dinheiro.
- O **Cluster 1** representa os clientes neutros. Eles compram com frequência regular e gastam uma quantia moderada de dinheiro.
- O **Cluster 2** representa os clientes mais engajados com a empresa. Eles compram com frequência e gastam muito dinheiro.



Conclusões

O projeto revelou-se essencial para a compreensão detalhada do comportamento dos clientes em um e-commerce. Ao explorar padrões de retenção, identificar grupos distintos com base em recência, frequência e valor, foram obtidos insights valiosos sobre o perfil e a interação dos clientes.

A análise de cohorts permitiu uma visão temporal do envolvimento do cliente, destacando pontos críticos de retenção e evolução ao longo do tempo. A segmentação RFV categorizou eficientemente os clientes, proporcionando uma compreensão mais refinada de suas características e necessidades.

A proposta de implementação do algoritmo K-Means para clusterização fortalece a personalização das estratégias de marketing, direcionando-as de maneira mais eficaz a grupos específicos. Isso torna a análise robusta, aproveitando todos os recursos disponíveis.

Recomendações

A próxima fase do projeto no que diz respeito ao modelo K-Means, sugere-se expandir a gama de dados de clientes.

Ao incluir dados adicionais, almeja-se não apenas escalabilidade, mas também uma melhoria substancial na eficiência operacional dos negócios. Essa abordagem visa aprimorar a capacidade do modelo de capturar nuances e padrões mais complexos no comportamento dos clientes, conduzindo a insights mais profundos e estratégias mais eficazes.



Sugestão de dados adicionais

- Perfil Demográfico:
 - Idade
 - Gênero
 - Estado civil
 - Localização geográfica
- Informações Socioeconômicas:
 - Renda
 - Educação
 - Profissão
- Histórico de Interações:
 - Respostas a campanhas anteriores
 - Participação em programas de fidelidade
- Canais de Comunicação Preferidos:
 - Preferência por e-mails, mensagens de texto, notificações push, etc.
- Atividade Online:
 - Navegação no site
 - Interações em redes sociais
 - Avaliações e comentários

Maison Henrique

