



# Desafio: Crie um modelo de análise das métricas RFV

☰ Escola	Dados
☷ Habilidades	Clustering
# N°	6



## Desafio



### Arquivo do Desafio:

<https://s3-us-west-2.amazonaws.com/secure.notion-static.com/e73c934e-cd3a-4e00-85f6-b704762fec77/data.csv>

- **Python Graph Gallery**: repositório com o passo a passo de como gerar gráficos utilizando as principais bibliotecas de Python
- **SciKit Learn**: documentação com os principais modelos utilizados para clusterização
- **ChatGPT**: Pode ser útil para iniciar sua pesquisa!

## Elabore um modelo de Clustering

Você deverá utilizar um modelo de Clustering para criar um sistema de agrupamento de perfis de clientes para um e-commerce. Deverá analisar o modelo mais eficiente, capaz de tornar as análises mais simplificadas para a empresa.

## Contexto

Você foi contratado por uma empresa de e-commerce que está buscando entender melhor o comportamento de seus clientes para personalizar as suas campanhas de marketing. Para isso, a empresa disponibilizou uma base de dados em csv contendo dados sobre clientes, produtos e transações da loja realizadas entre os anos de 2010 e 2011.

Com base nesses dados, você precisa agrupar os clientes em clusters com base em seu comportamento de compra. Isso irá permitir identificar padrões e características em comum entre os clientes, como:

- Clientes que compram os mesmos produtos;
- Clientes que possuem a mesma frequência de compras;
- Clientes que gastam mais dinheiro em suas compras.

A partir desses clusters, gere insights para que a empresa possa segmentar melhor a sua base de clientes e personalizar as suas campanhas de marketing, direcionando promoções e ofertas aos clientes com base no comportamento de compras.

## Sobre os dados

Acesse os dados aqui: <https://www.kaggle.com/datasets/carrie1/ecommerce-data>

Os dados fornecidos possuem informações de transações de compras de uma loja de e-commerce em 38 países e territórios, com mais de 4.000 clientes únicos e mais de 540.000 transações.

Atenção: as datas estão no formato MM/DD/YYYY HH:mm:ss; existem dados nulos que precisam ser tratados antes da realização da análise; por mais que os códigos de identificação sejam numéricos, o modelo não pode considerá-los como grandezas numéricas.

Coluna	Descrição	Tipo
<i>InvoiceNo</i>	Identificação da transação	Int
<i>StockCode</i>	Código de estoque do produto	String
<i>Description</i>	Descrição do produto	String
<i>Quantity</i>	Quantidade de produtos por transação	Int
<i>InvoiceDate</i>	Data da transação	Datetime
<i>UnitPrice</i>	Preço unitário do produto	Float
<i>CustomerID</i>	Identificação do cliente	Int
<i>Country</i>	País de origem da transação	String

## Como começar?

Desenvolva um modelo de clusterização que seja capaz de agrupar os clientes conforme o seu comportamento de compras levando em consideração o RFM calculado no desafio 4.

Analise os clusters obtidos para identificar o perfil de cliente, como padrões e características em comum para determinar o seu comportamento de compra. Utilize gráficos e visualizações para auxiliar na análise.

Não se esqueça de documentar cada etapa, justificando as escolhas realizadas. É essencial informar os insights obtidos e as recomendações de ações futuras para que a empresa possa colocá-los em prática. Boa sorte!

## Etapas de Desenvolvimento

### Etapa 01) Análise exploratória dos dados

- Carregue a base de dados;
- Realize uma descrição estatística dos dados;
- Visualize as distribuições e identifique a relevância das colunas para a análise;
- Verifique a presença de dados nulos, duplicados, outliers e demais inconsistências nos dados.



**Dica:** certifique-se de que os dados estejam no formato ideal para as análises; por mais que os códigos de identificação sejam numéricos, sua análise não pode considerá-los como grandezas numéricas.

## Etapa 02) Pré-processamento dos dados

- Realize a normalização dos dados
- Faça uma seleção das variáveis mais relevantes para o modelo;
- Remova os dados nulos, duplicados, outliers e inconsistentes

## Etapa 03) Selecione um algoritmo de clusterização

- Escolha um algoritmo adequado para base de dados, como o KMeans, DBSCAN, Hierarquia ou Mean Shift
- Encontre a quantidade ideal de clusters através dos métodos de Elbow ou Silhouette Score
- Implemente o algoritmo escolhido

## Etapa 04) Analise os clusters obtidos

- Identifique os padrões e características em comum entre os clientes
- Plote gráficos para auxiliar na análise

## Etapa 05) Interpretação dos resultados obtidos

- Descreva o perfil de compras dos clientes de cada cluster
- Justifique como essa análise pode ser útil para empresa para segmentação de seus clientes e personalização das campanhas de marketing
- Sugira ações possíveis com base nas ações realizadas



## Critérios de Avaliação

Os critérios de avaliação mostram como você será avaliado em relação ao seu desafio.

Critérios	Atendeu às Especificações	Pontos
<b>Organização e legibilidade do código</b>	O código precisa estar organizado de forma clara e legível, com comentários que expliquem as ações tomadas e divisão em sessões. Essa é a documentação do projeto apresentado para o cliente.	20
<b>Justificativa das tomadas de decisão</b>	É importante que as decisões tomadas durante todo o processo de análise sejam devidamente justificadas, desde a escolha das variáveis até a escolha do algoritmo de clusterização.	20
<b>Qualidade da análise exploratória</b>	A análise deve ser completa e detalhada, mostrando a relação entre as variáveis e possíveis insights encontrados.	20
<b>Qualidade da análise dos clusters</b>	Os clusters devem ser bem definidos e interpretáveis, identificando suas características que os definem.	20
<b>Apresentação de proposta coerente com os clusters encontrados</b>	A proposta deve ser coerente com os clusters encontrados, sugerindo ações que possam ser tomadas para melhorar o desempenho do e-commerce com base nas suas características.	20



## Entrega



**Como entregar:** Você deverá submeter o link compartilhável do colab!

Evite sequências longas de código sem explicação do que está sendo realizado.



**Dica:** pense que essa documentação está sendo apresentada para o cliente final, seu modelo não pode ser uma “caixa preta”.