

ANÁLISE DE CLIENTES

UMA ABORDAGEM COM MODELOS DE
APRENDIZADO DE MÁQUINA

PROF. DR. MURILO COELHO NALDI

BRUNO LEANDRO PEREIRA
RA 791067

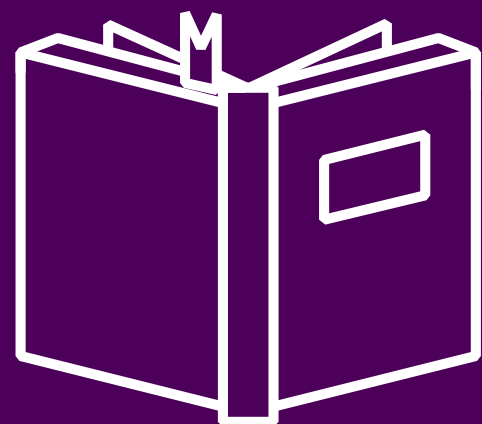
CARLOS EDUARDO FONTANELI
RA 769949

IVAN DUARTE CALVO
RA 790739



Problemática

Escolha e estudo de conjunto de dados, gerar e avaliar um modelo de agrupamento de dados



Conjunto de Dados

O conjunto de dados escolhido é relativo a dados de clientes de um determinado empresa em relação a seus hábitos de consumo e suas respostas as campanhas de marketing elaboradas pela empresa.



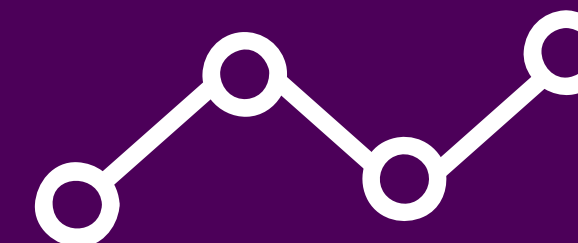
Atributos

- Estado Civil
- Gastos
- Compras
- Promoções
- Educação
- etc...



Estudo dos Dados e Engenharia de Atributos

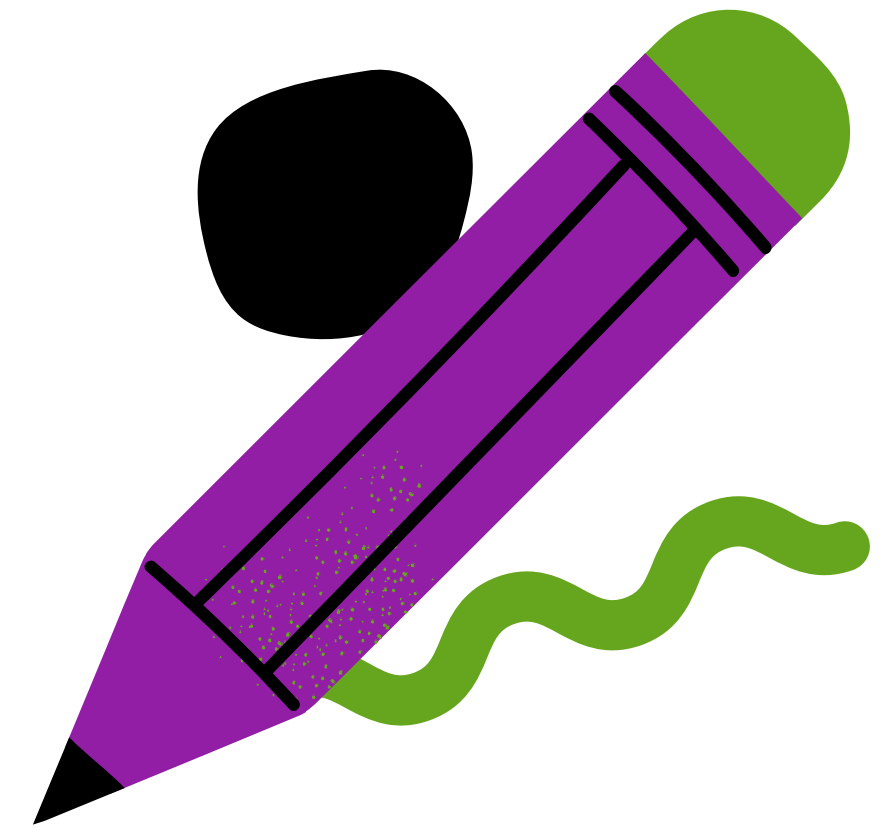
Realização da análise inicial e exploratória dos dados e alterações nos atributos.



Aplicação e Avaliação de Modelos

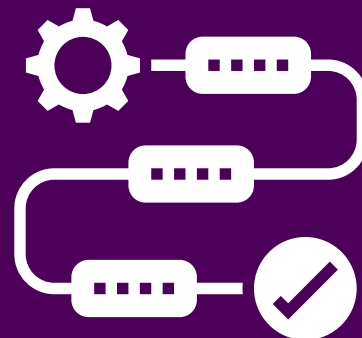
Implementação dos modelos de KMeans, DBSCAN e Agrupamento Hierárquico e análise dos resultados.

Objetivos & Metodologia



Objetivos

O principal objetivo é entender o comportamento dos clientes , para que assim os produtos e camapanhas sejam direcionados aos clientes de acordo com suas necessidades.



Metodologia

- Análise exploratória dos dados.
- Pré-processamento dos dados e engenharia de atributos;
- Aplicação de modelos de agrupamento;
- Avaliação dos resultados obtidos.



Ferramentas

Linguagem de programação multi-paradigma orientada a objetos: Python.

Jupyter Notebooks para realização do relatório.

Canva para elaboração da apresentação.

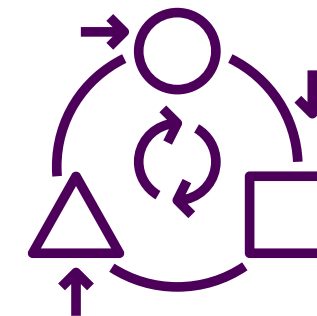
Análise Inicial



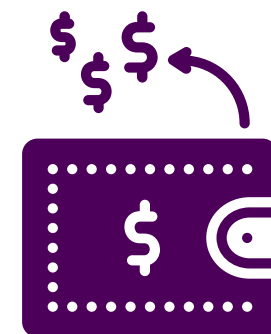
Obtenção de informações mais palpáveis e interpretativas



Interpretação de como os dados estão armazenados

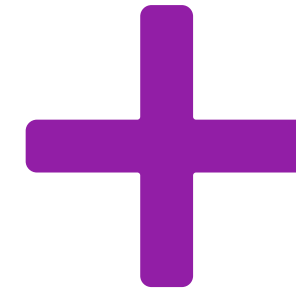


Atributos nominais com muitas classes distintas (Educação e Estado Civil);

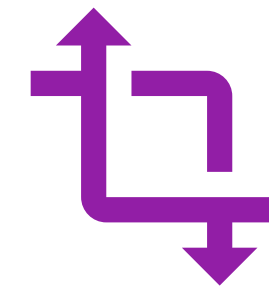


Atributos faltantes em renda, compras e gastos estão subdivididos em diversas categorias;

Engenharia de Atributos

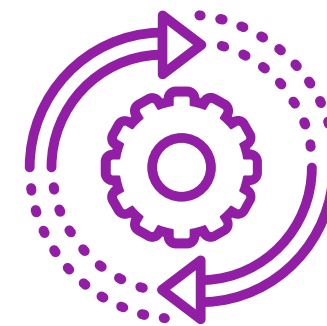


União de atributos
relativos a compra de
itens;



Simplificação de atributos
nominais para apenas 2
classes possíveis;

Gasto, Compras, Filhos



Transformação de
atributos;

Nascimento -> Idade;

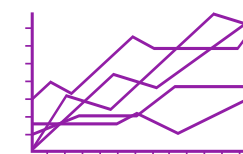
Análise Exploratória



Obtenção de informações
mais palpáveis e
interpretativas



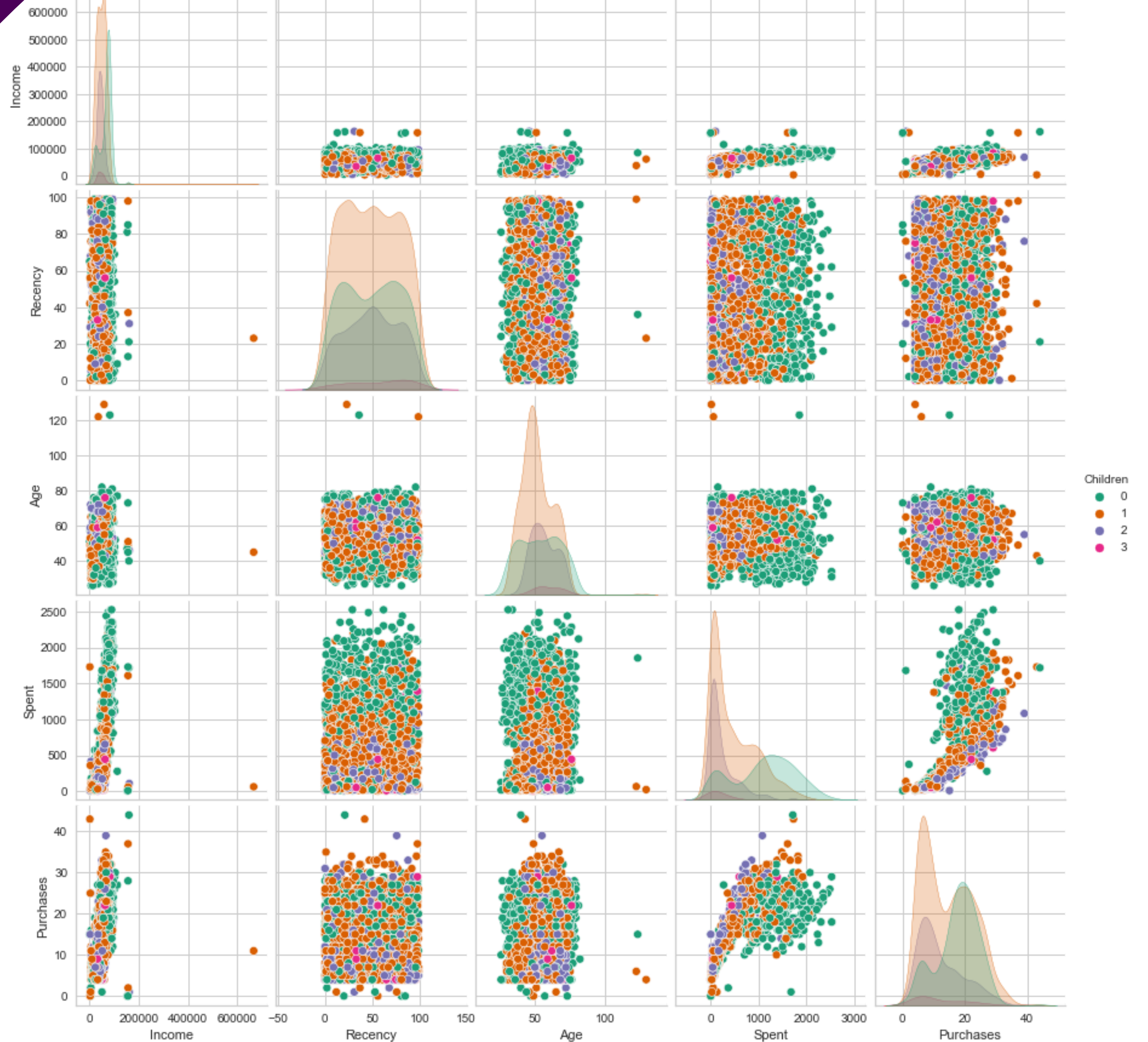
Busca de possíveis
outliers ;



Representação visual dos
dados através de gráficos

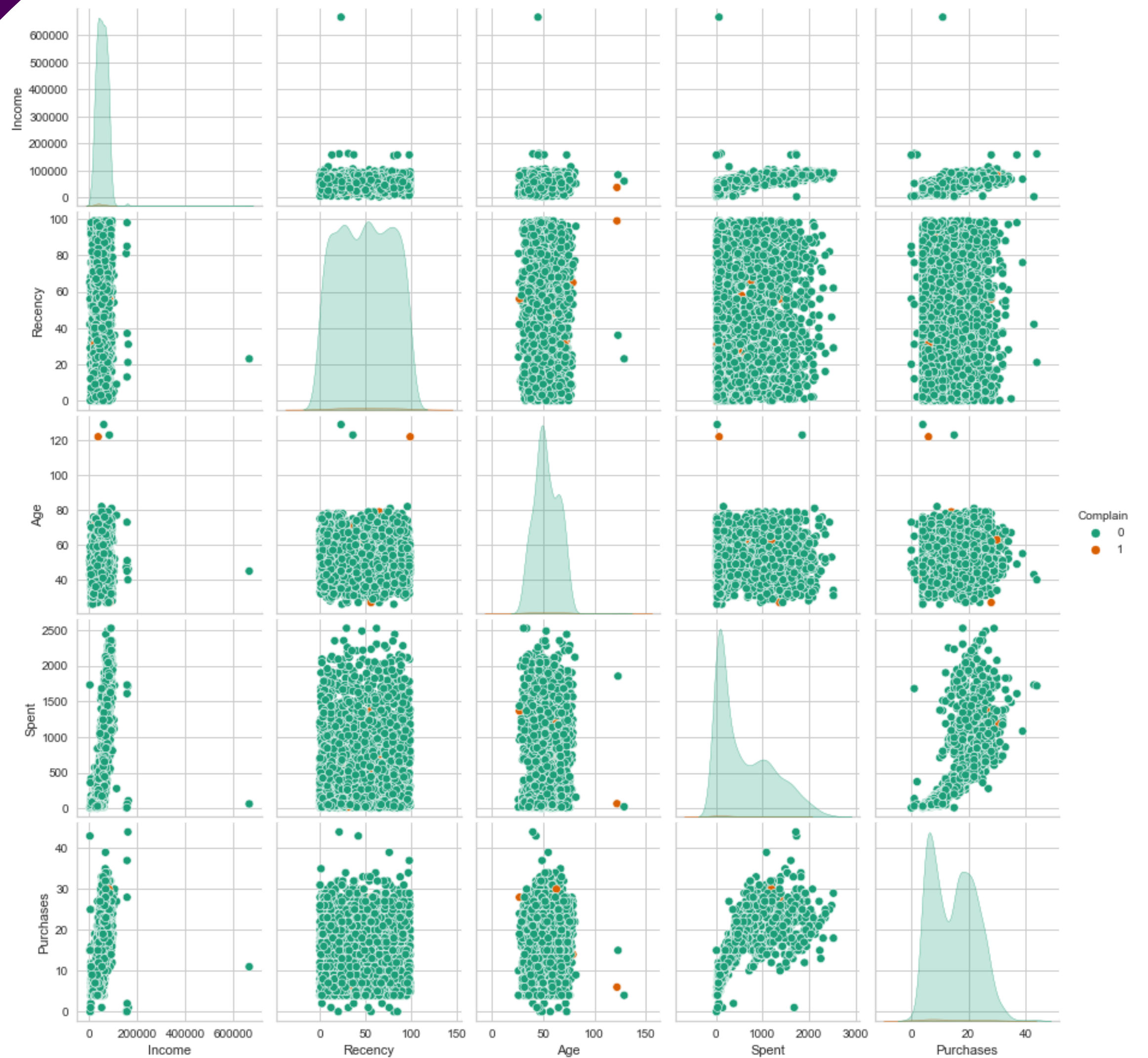
Dados x Filhos

- Maioria tem entre 0-1 filhos;
- Possíveis outliers em idade e renda;



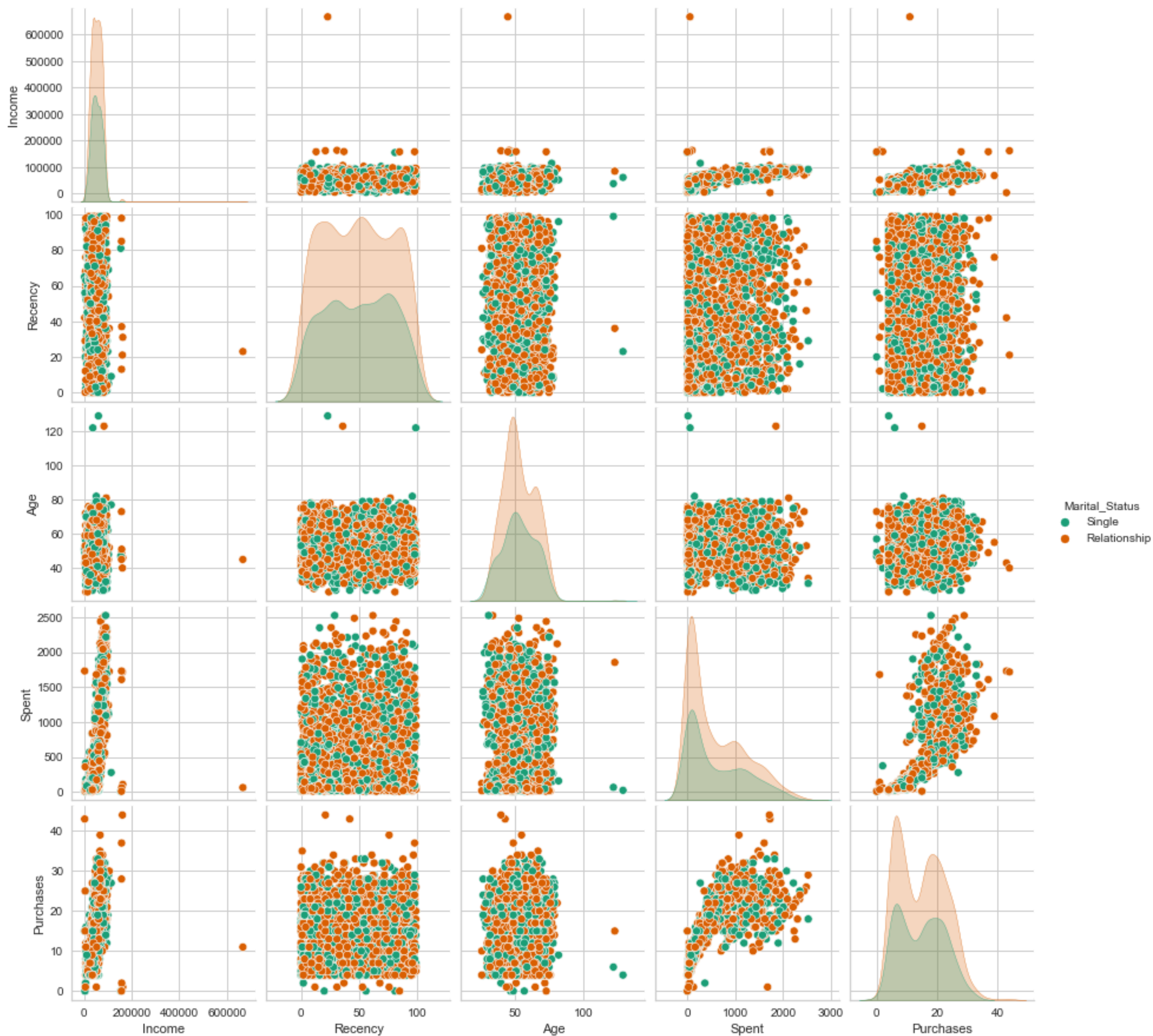
Dados x Reclamou

- Grande maioria nunca reclamou;

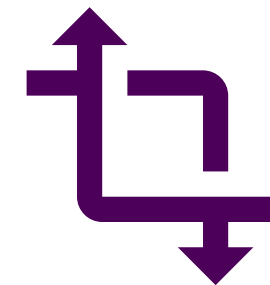


Dados x Estado Civil

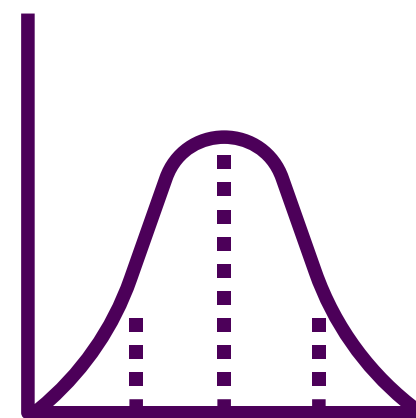
- A maioria está em um relacionamento;



Pré-Processamento dos Dados



Discretização dos atributos nominais;

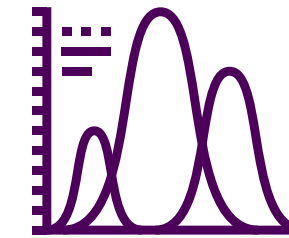


Normalização dos atributos contínuos, redução de dimensionalidade através da técnica **PCA**;



Remoção de outliers, exclusão de tuplas com valores muito discrepantes para o atributo idade e renda;

DBSCAN



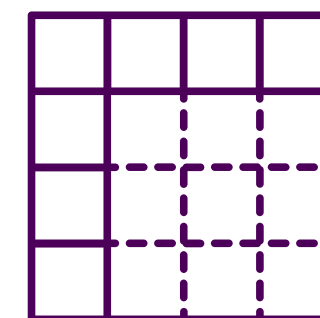
Modelo não supervisionado baseado na densidade



Aplicação do modelo



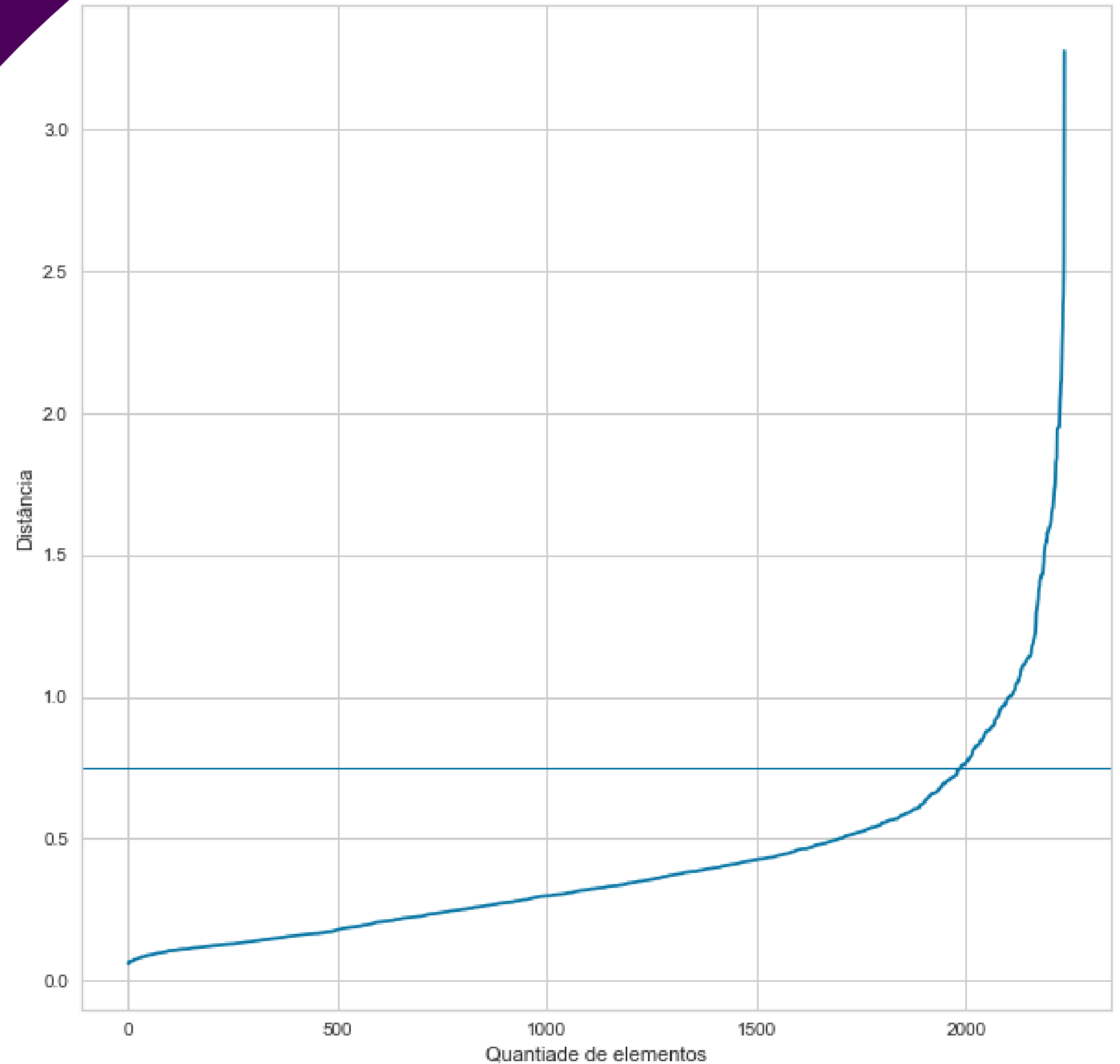
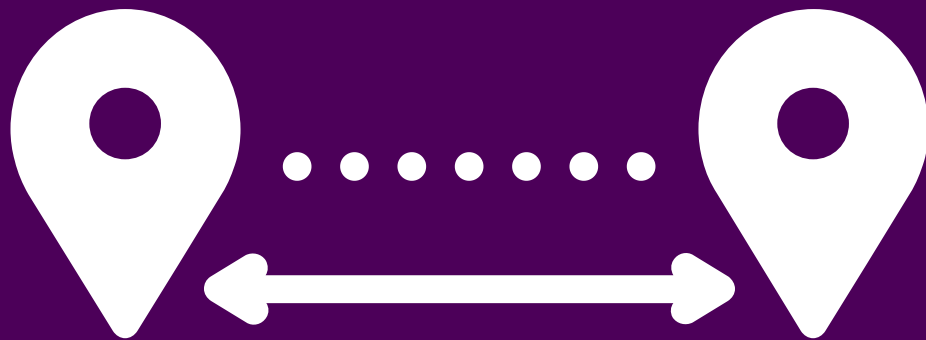
Otimização dos parâmetros



Análise visual dos resultados através de gráficos

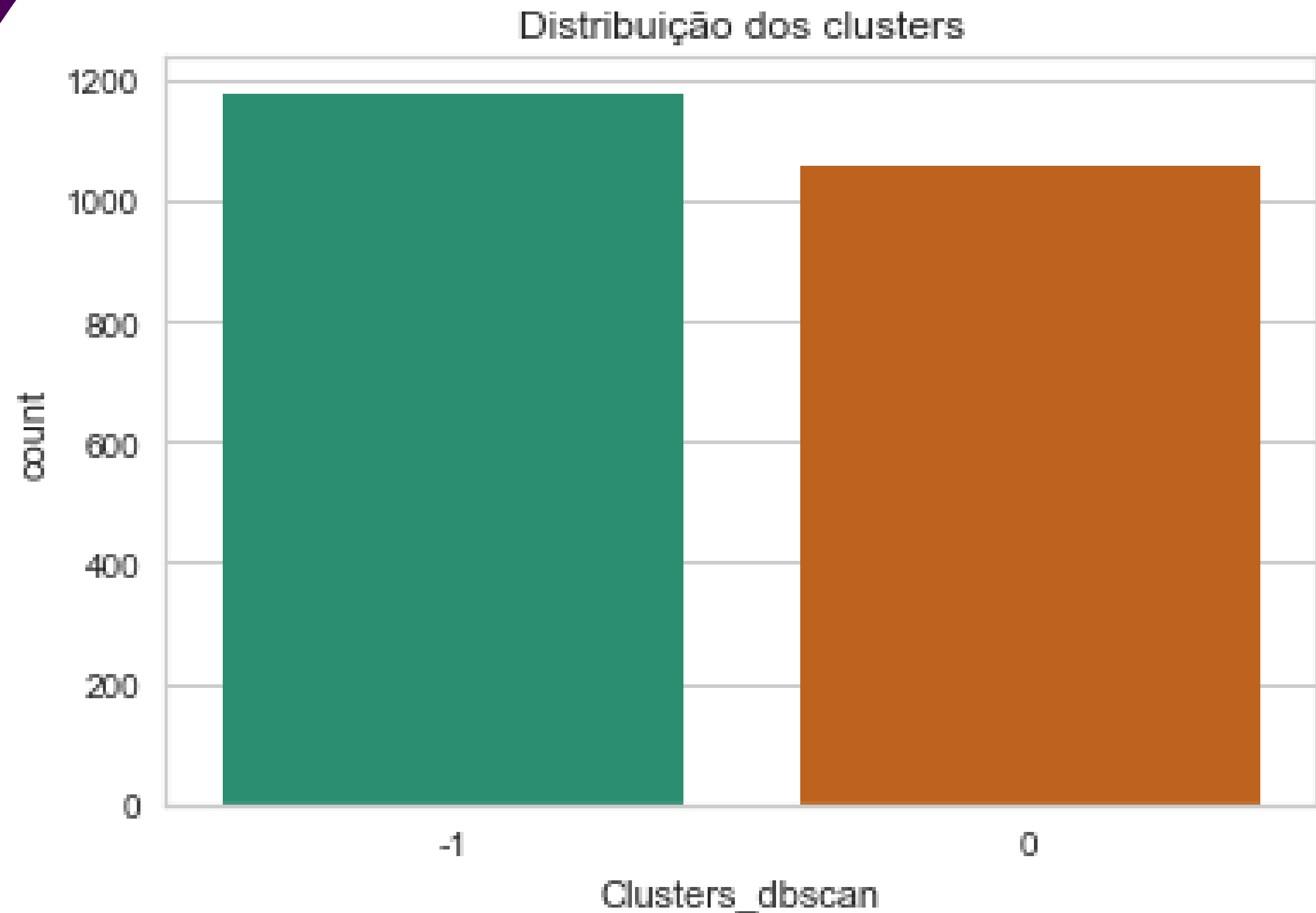
Otimização₅

- KNN para determinação de distância de pontos no mesmo cluster;
-
- $2 * \text{dimensão do dataset}$ para quantidade pontos mínimos por cluster;



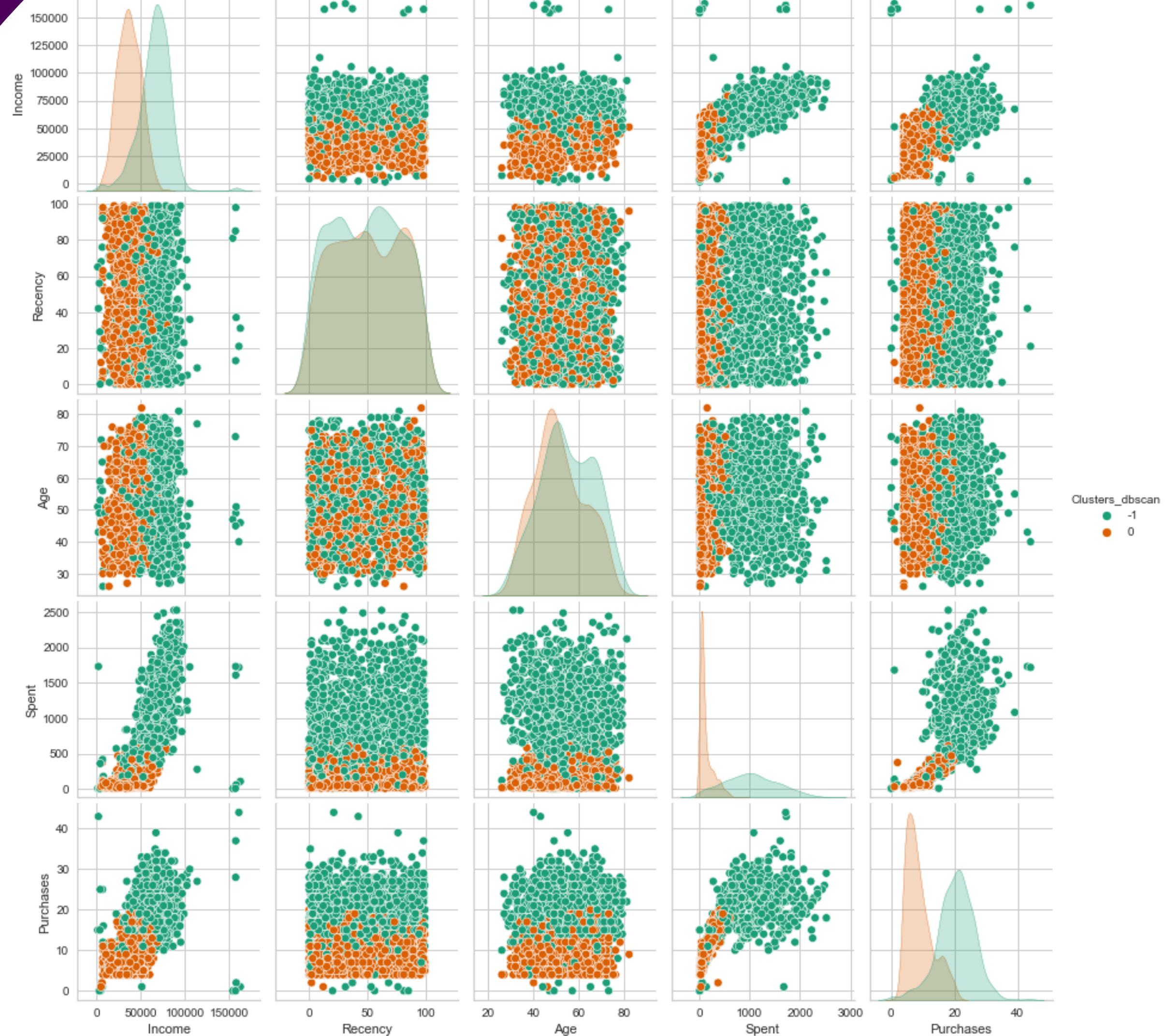
Distribuição dos Elementos

- Dividido igualmente porém apenas 2 clusters;

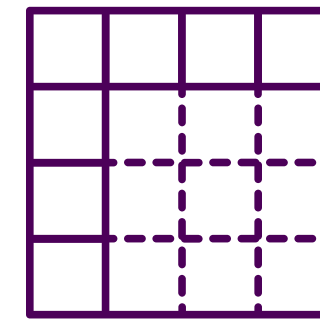


Análise dos Clusters

- Divisão de certa forma perceptível;
- Para alguns atributos não foi possível dividir bem;



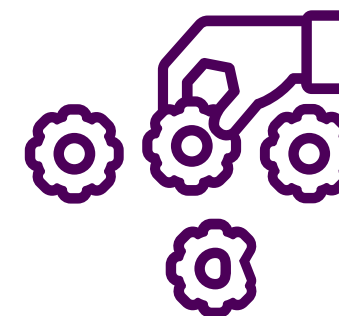
KMeans



Gera um número K de Centróides



Distribui os centros em relação às distâncias dos centróides



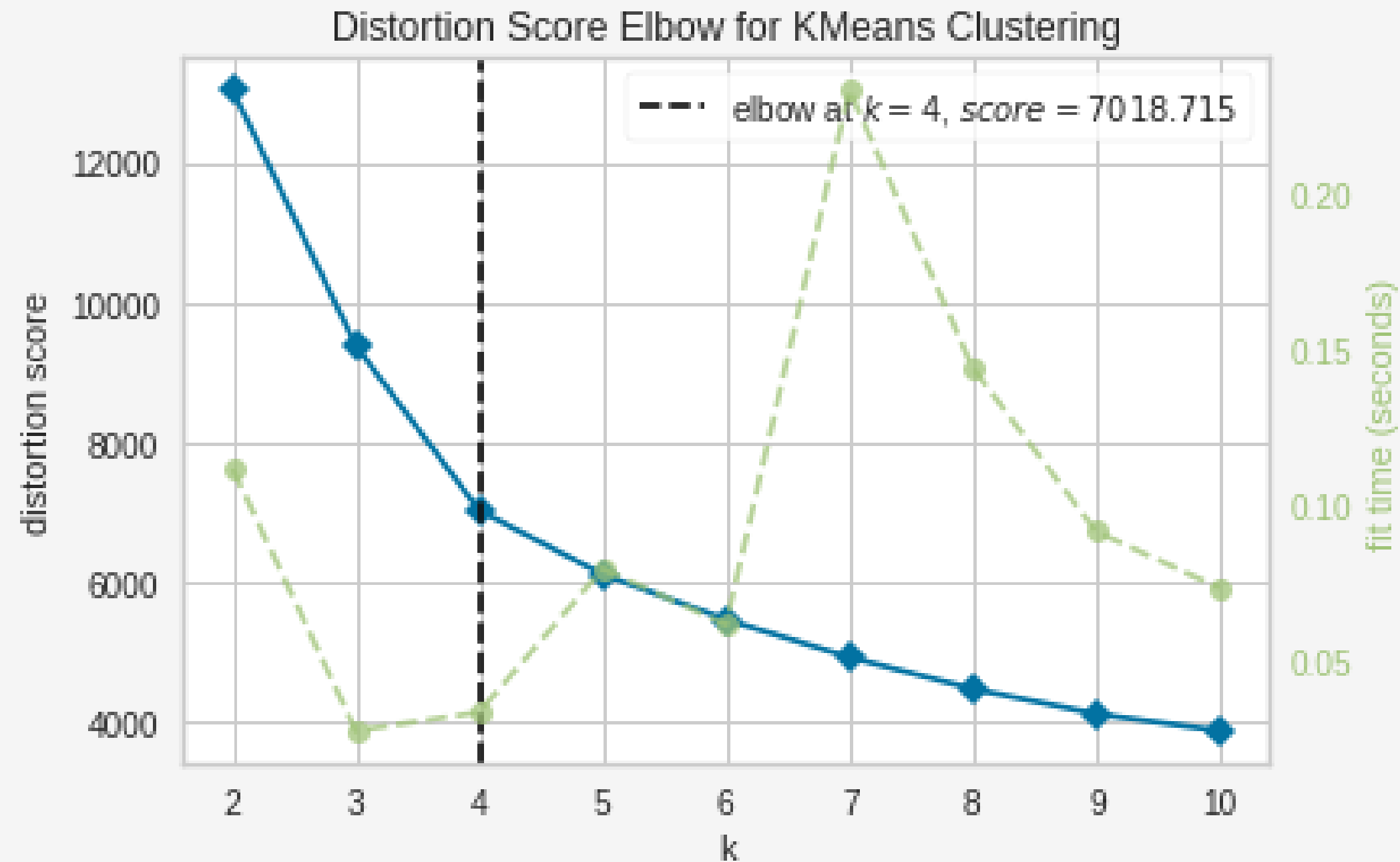
Reposiciona os centróides com base na média das posições dos pontos



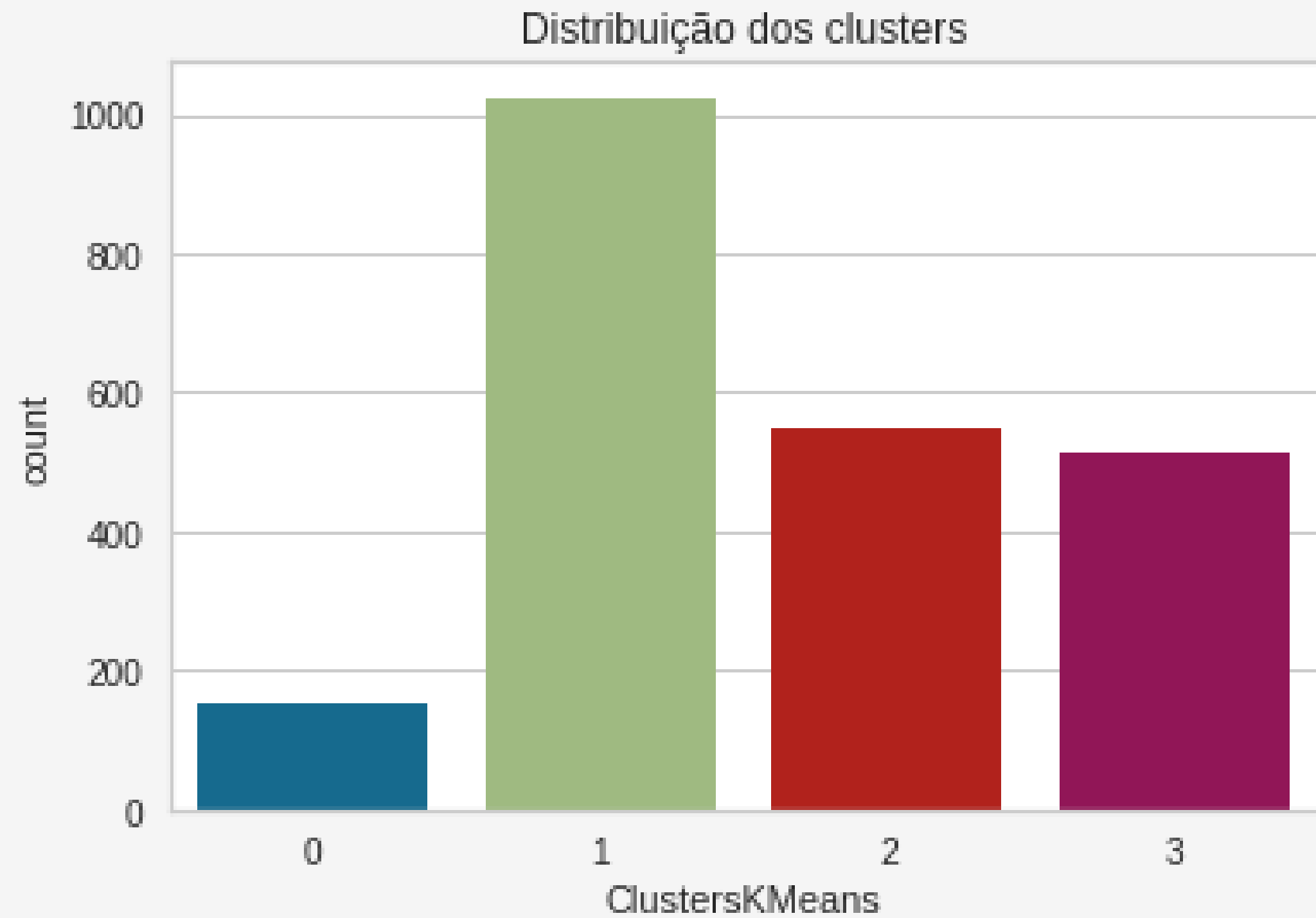
Repete

Escolhendo o valor ideal de K

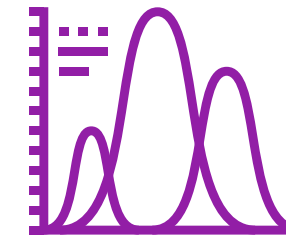
- Cálculo da distorção através do KEWlbowVisualizer e silhouette_score



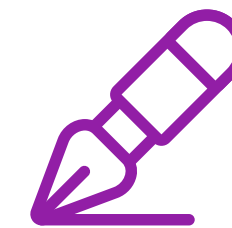
Distribuição dos Clusters



Agrupamento Hierárquico



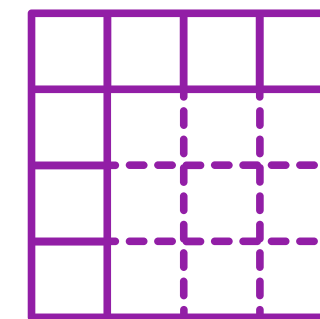
Modelo não supervisionado



Aplicação do modelo



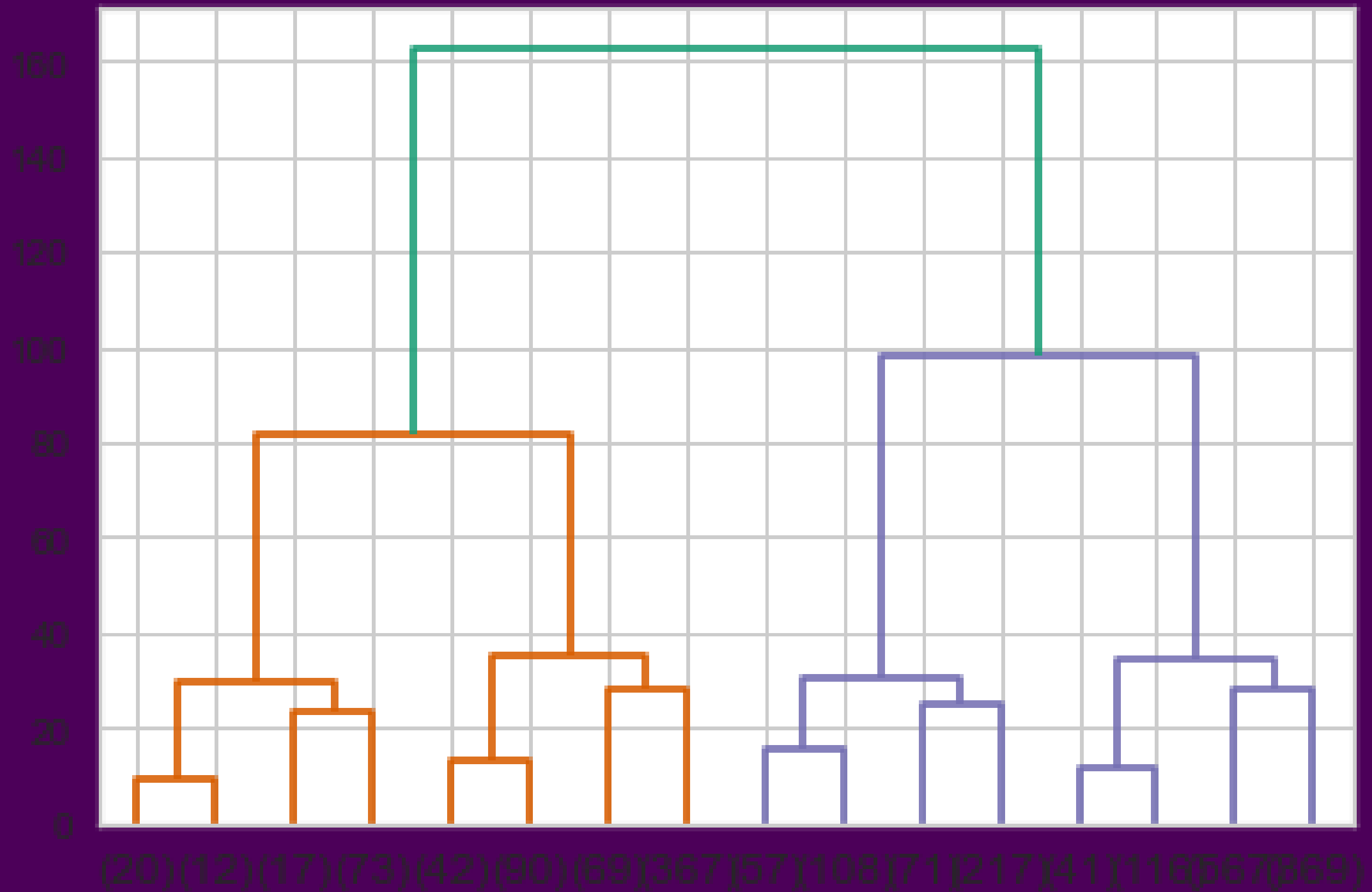
Otimização dos parâmetros



Análise visual dos resultados através de gráficos

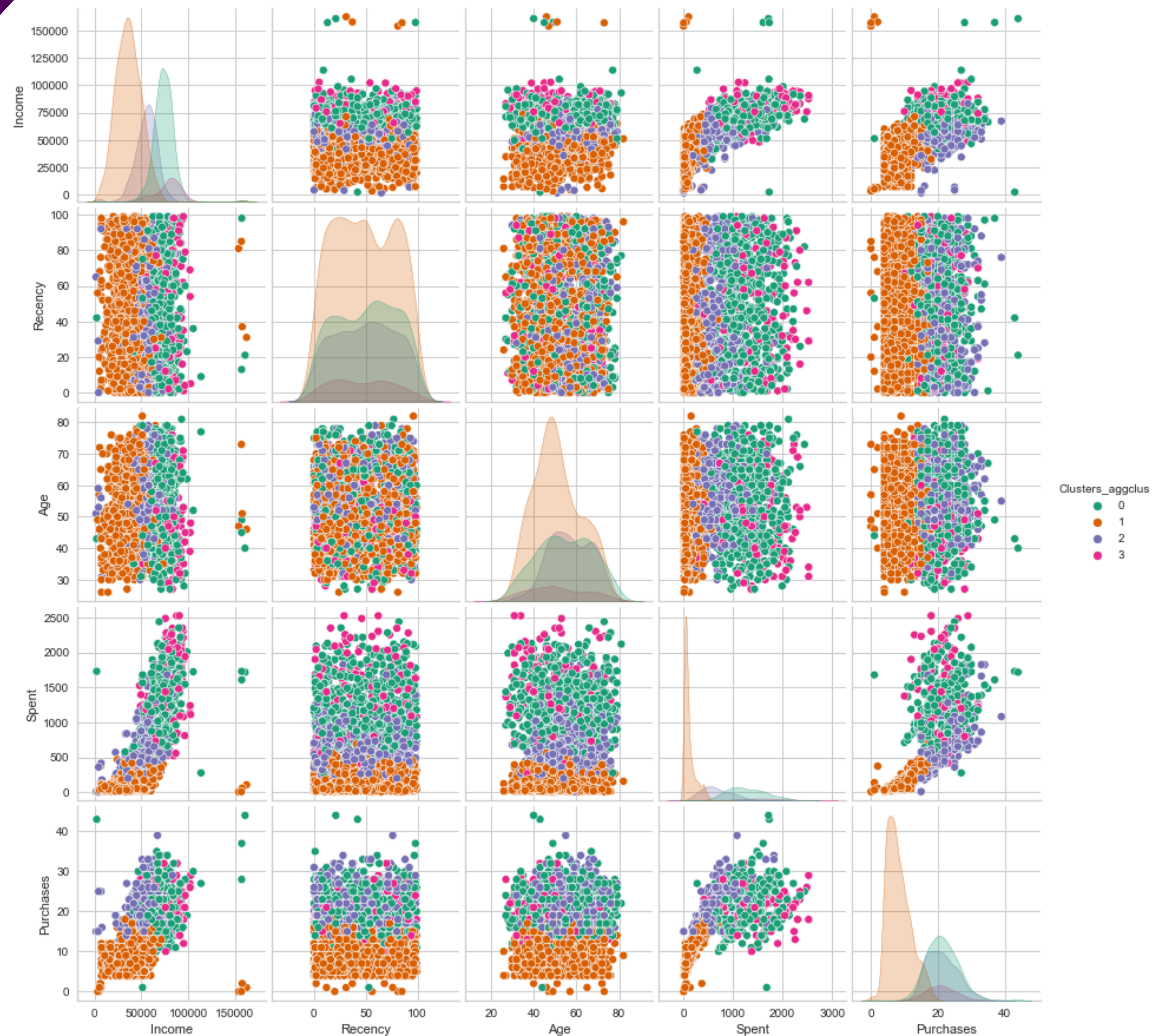
Dendrograma

Para análise e ajuda na seleção da quantidade de clusters



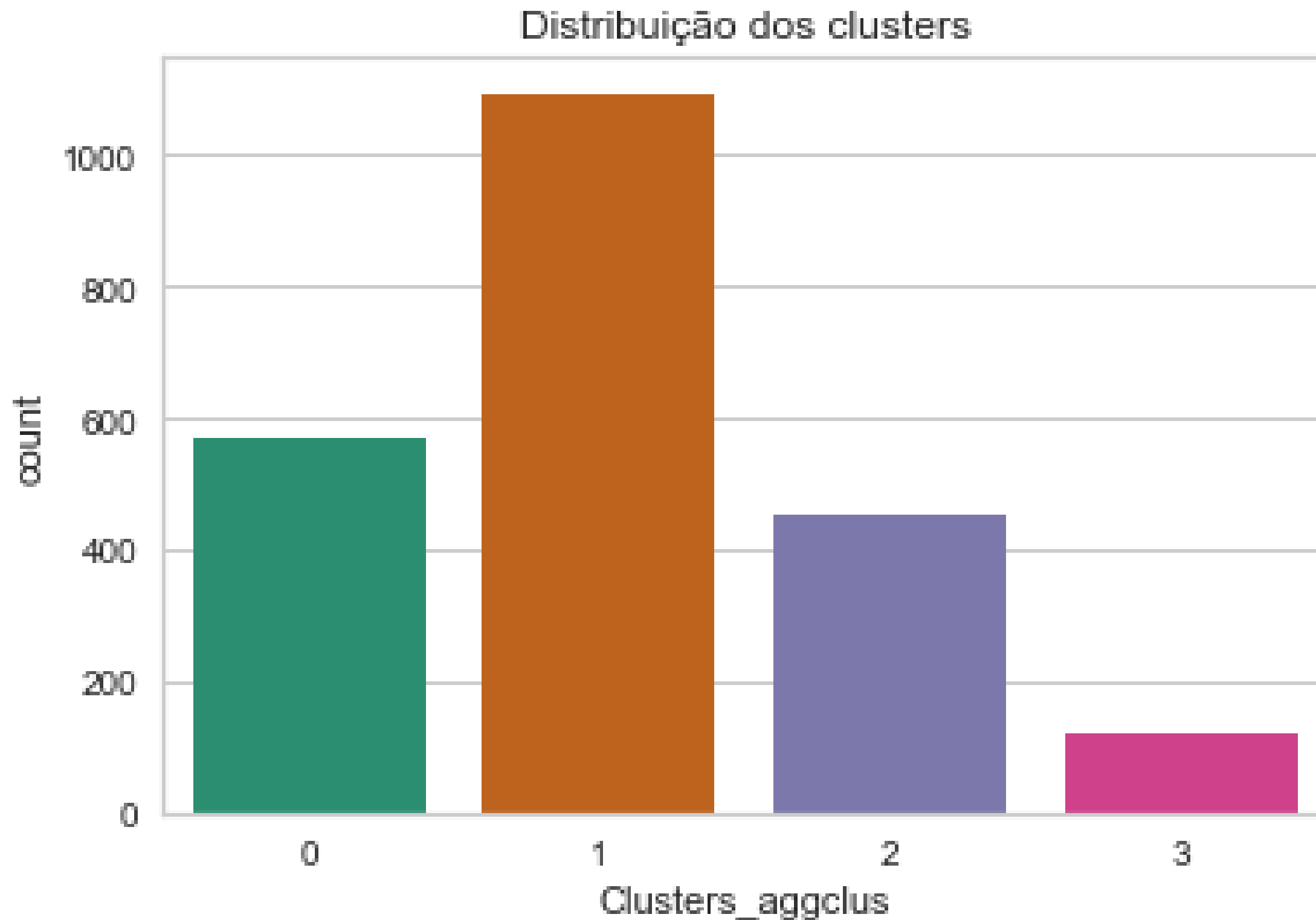
Análise dos Clusters

- Divisão de certa forma perceptível;
- Para alguns atributos não foi possível dividir bem;

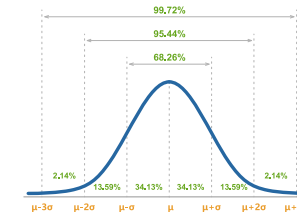
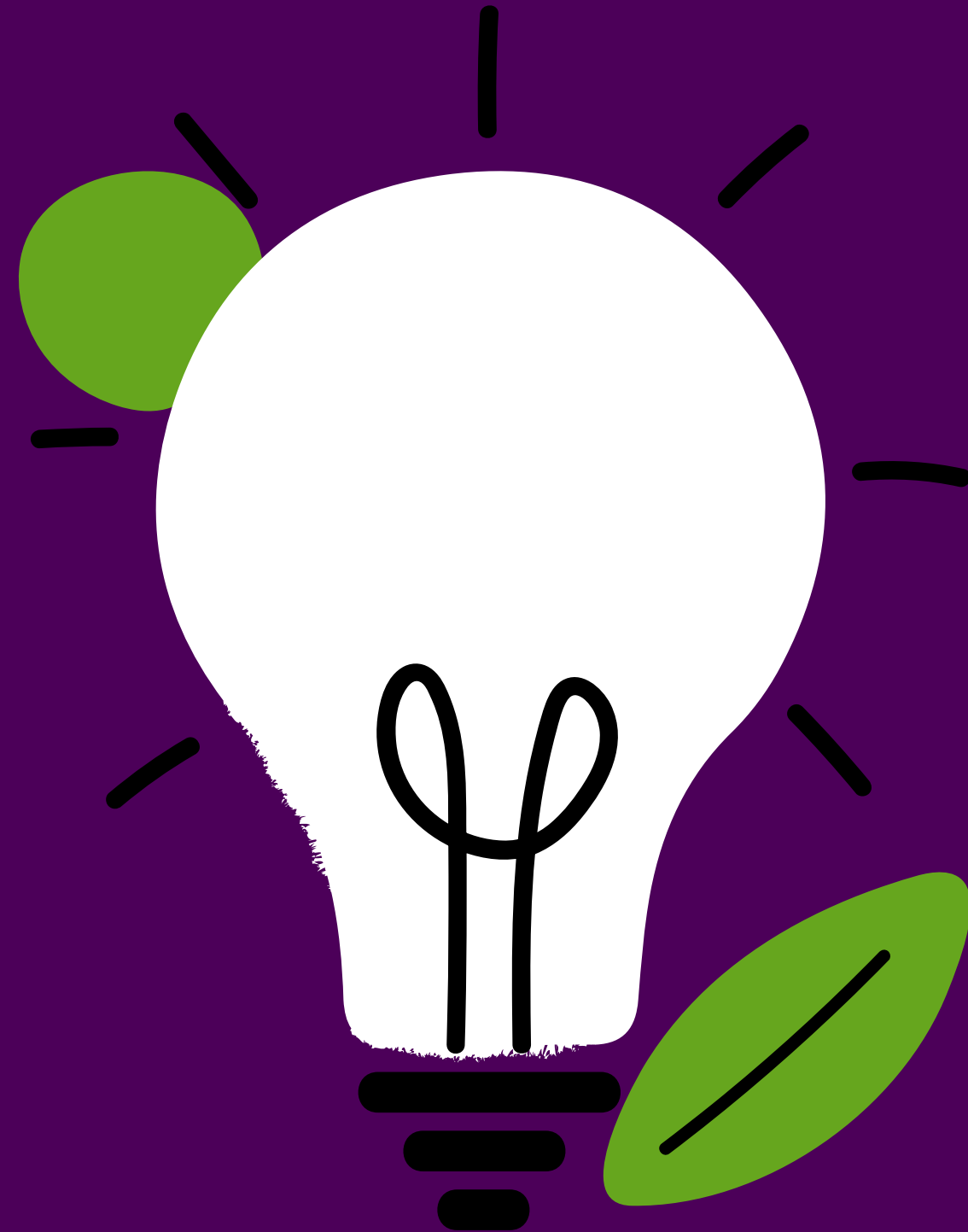


Distribuição dos Elementos

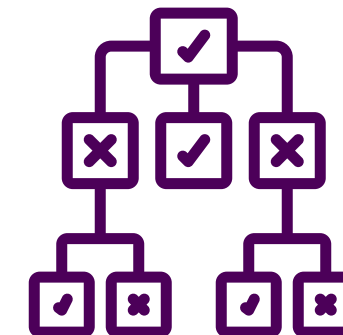
- Dividido em 4 clusters, após análise do dendograma e gráficos.



Extração de Características dos Clusters obtidos com KMeans



Gráficos de densidade para análise de distribuição;



Análise da aceitação de promoções;

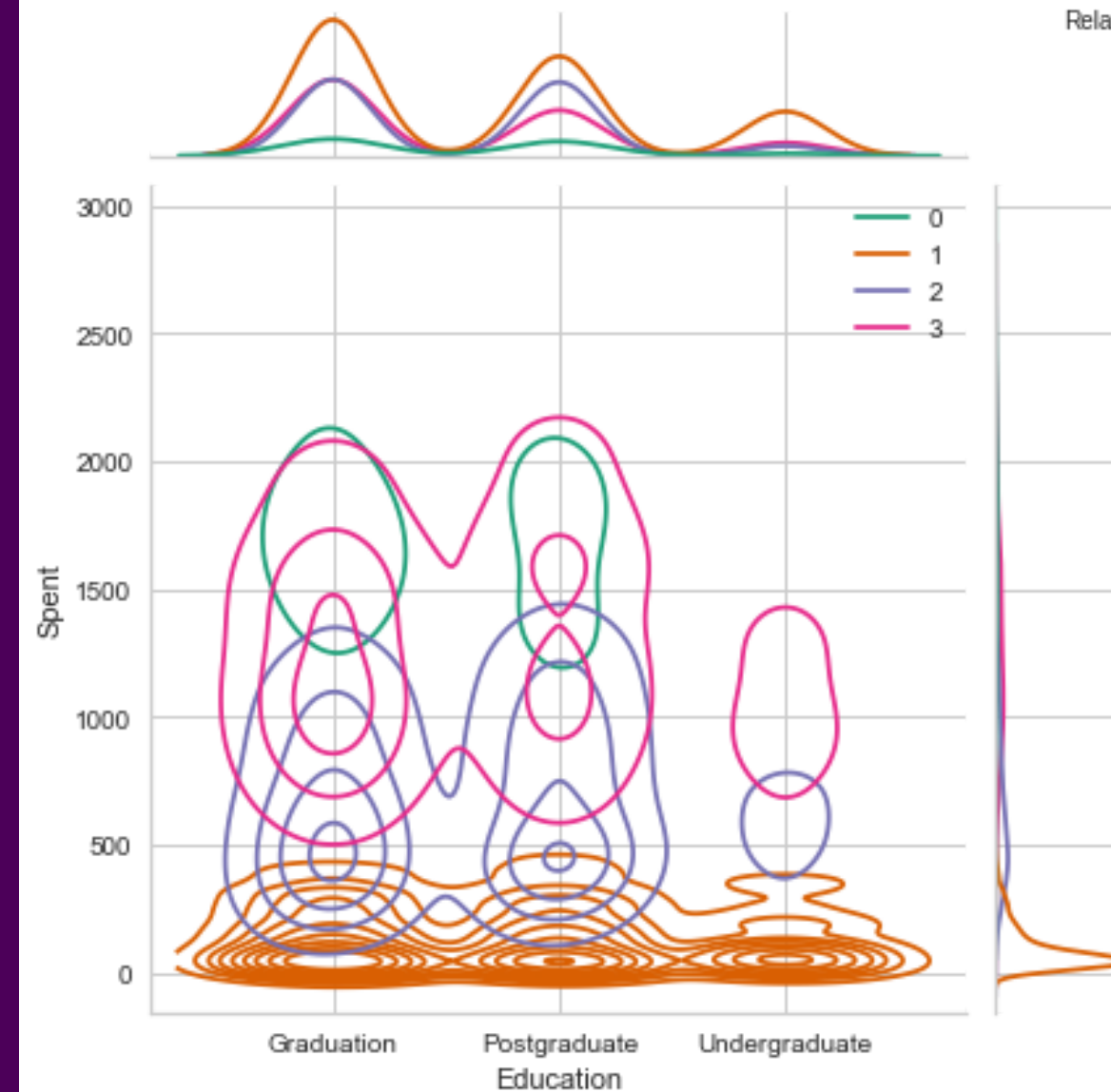
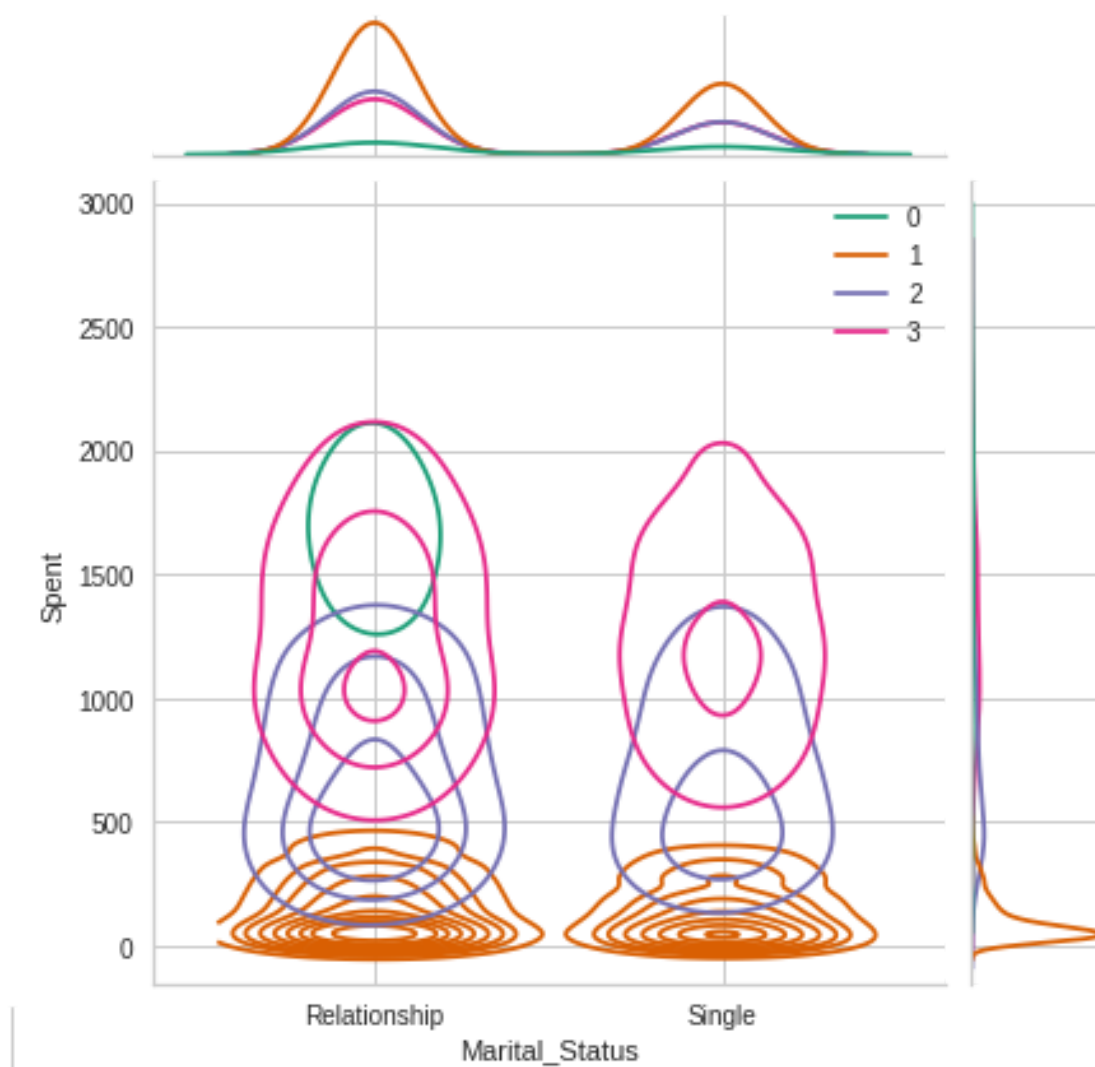


Distribuição dos dados conforme os clusters;



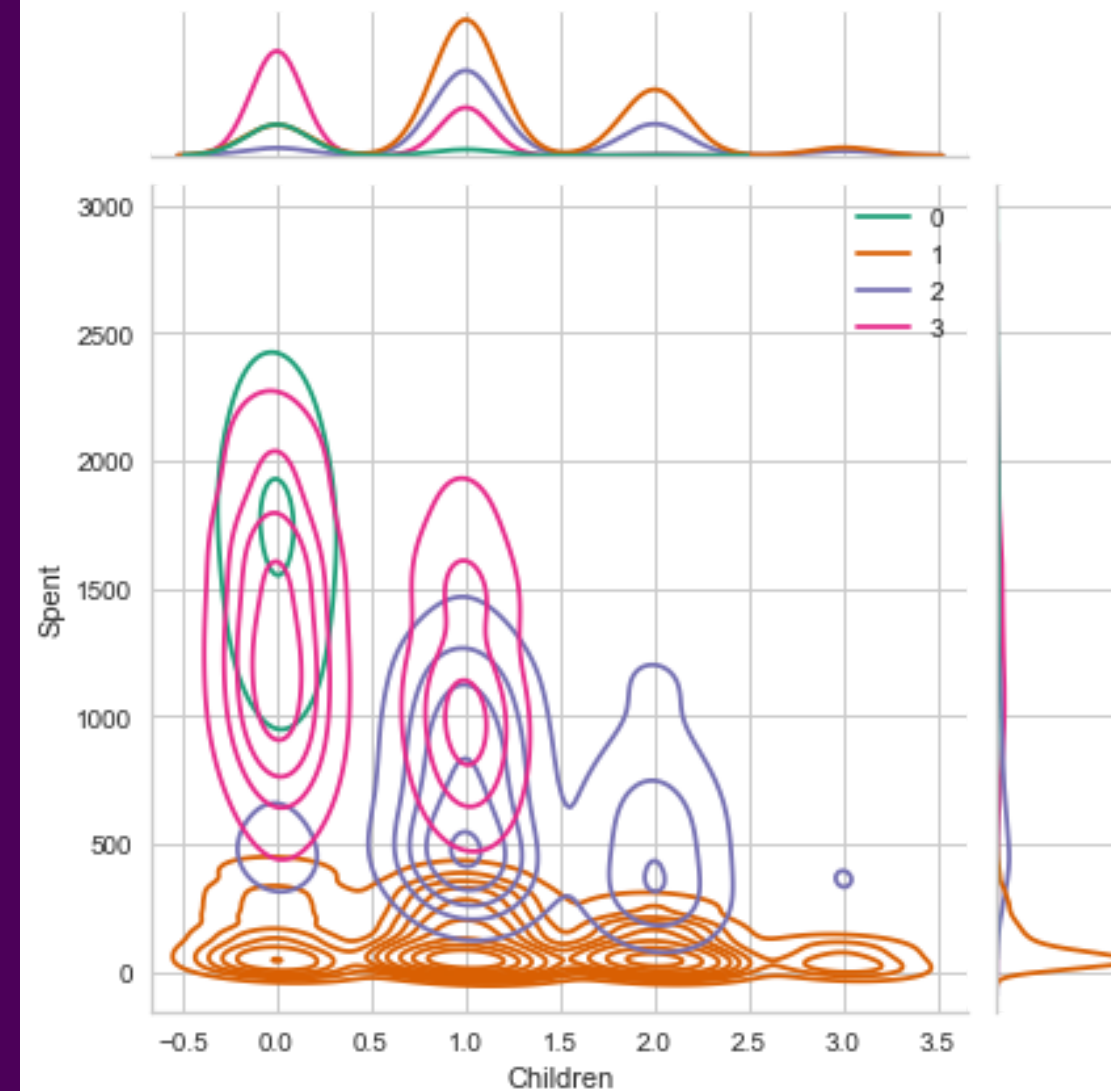
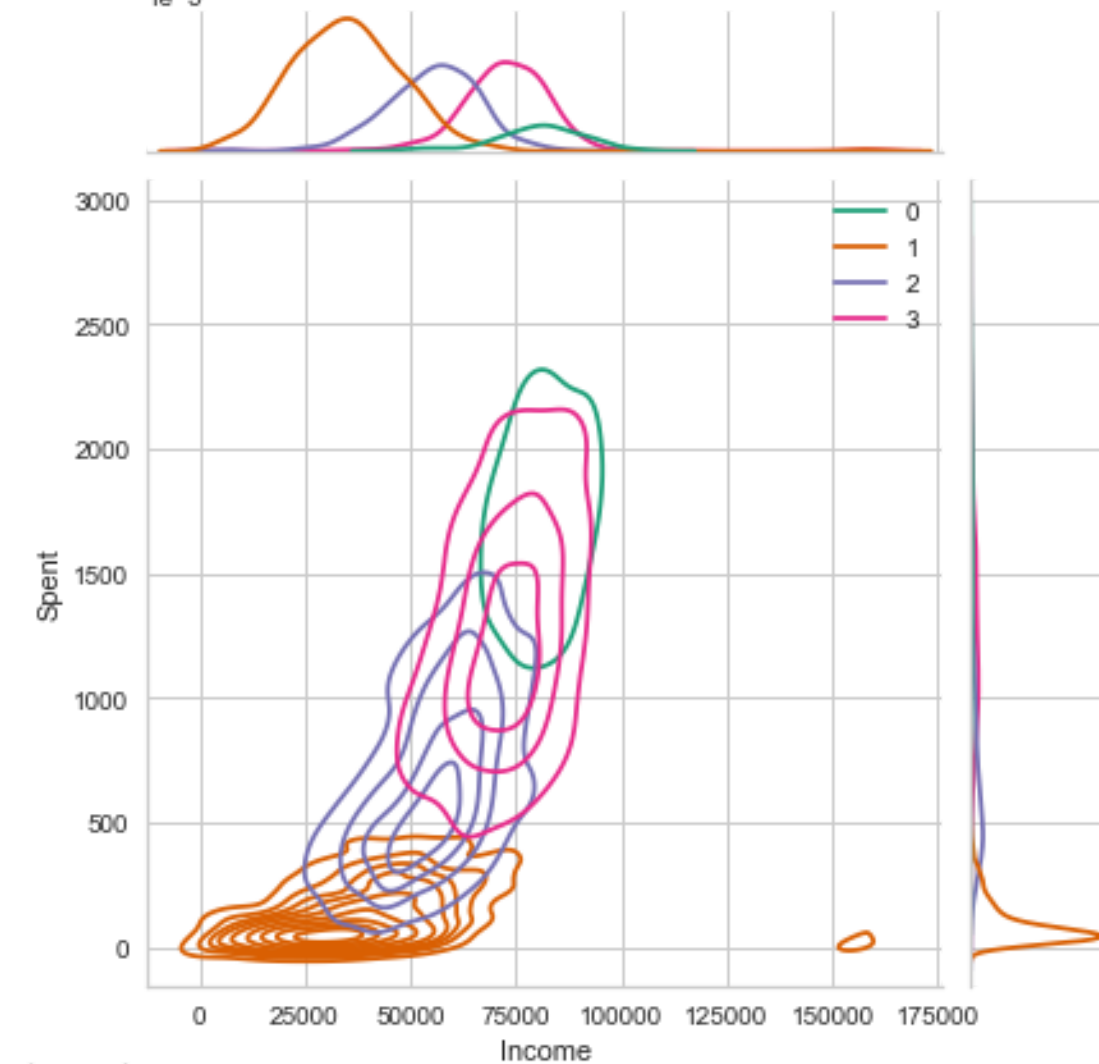
Cluster 0

- **Mostrou ser o menor cluster;**
- **Gasto relativamente alto, compras elevadas;**
- **Idade entre 40 e 50 anos;**
- **Sem filhos;**
- **Uma das rendas mais alta;**
- **Possui diploma de nível superior, pelo menos;**
- **Estão em um relacionamento;**



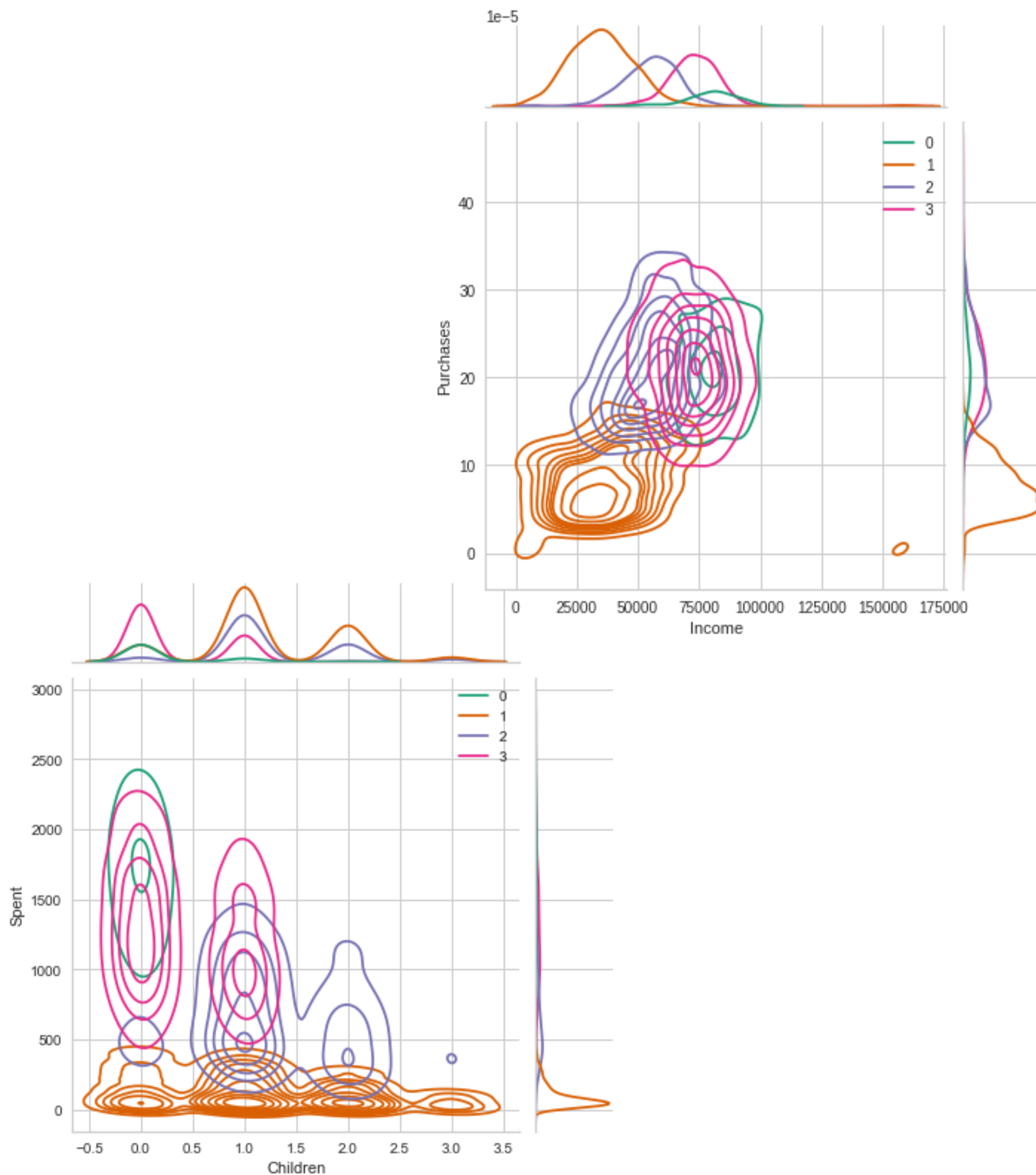
Cluster 1

- Cluster com o menor gasto;
- Em sua maioria, os indivíduos possuem baixa escolaridade;
- Possuem uma menor renda;
- Em sua maioria possui ao menos 1 filho;
- É o maior cluster dentre os criados;



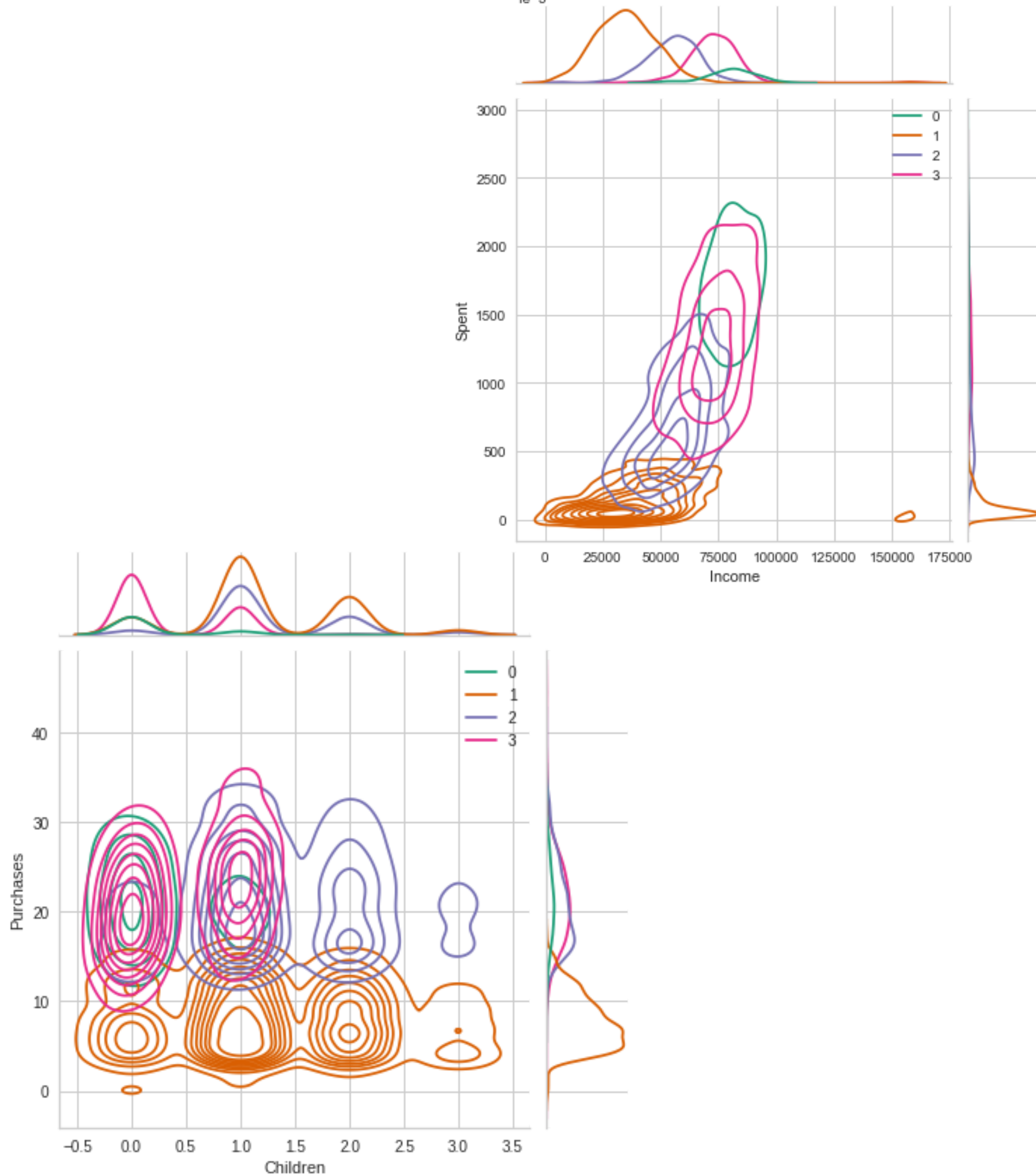
Cluster 2

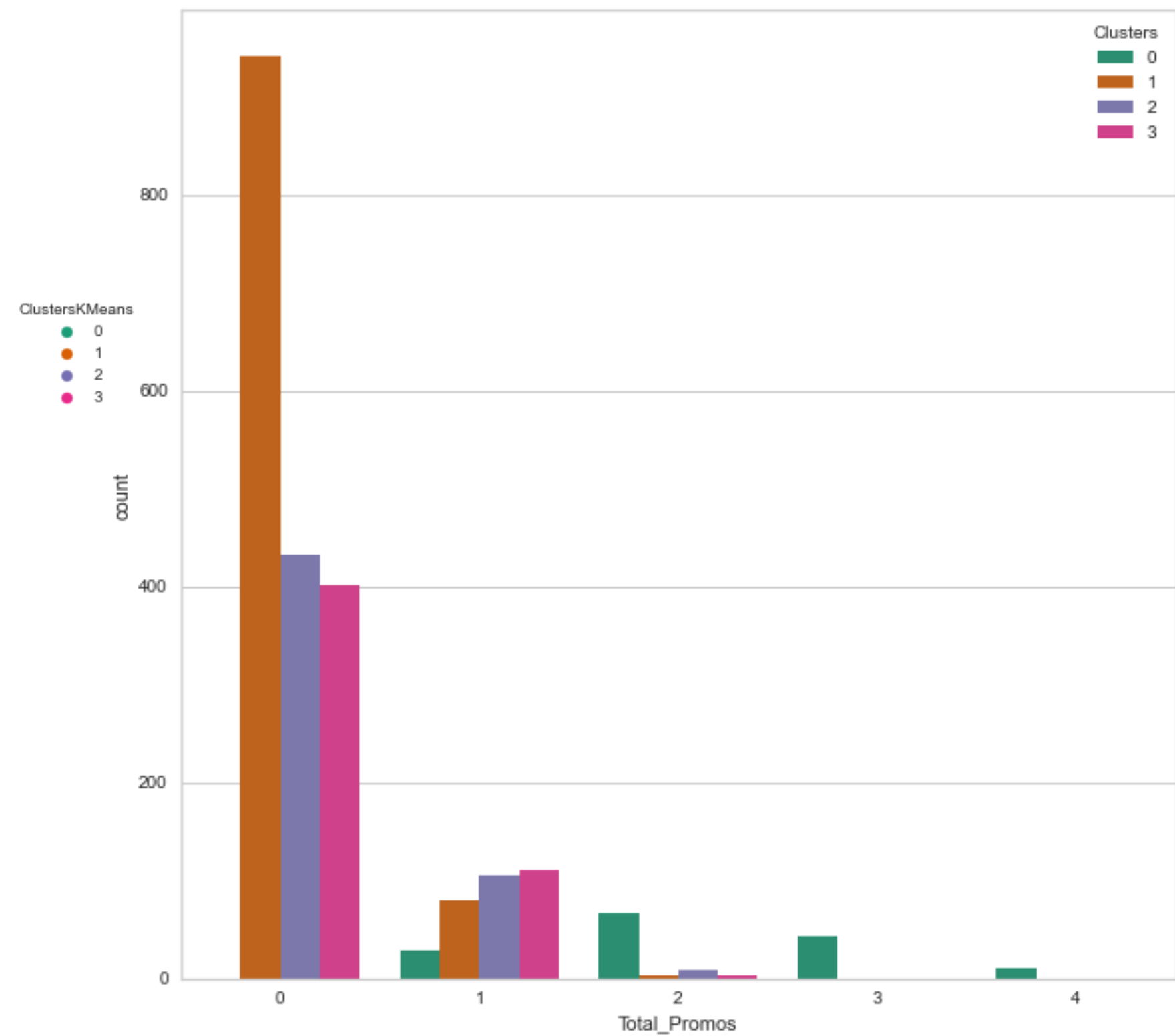
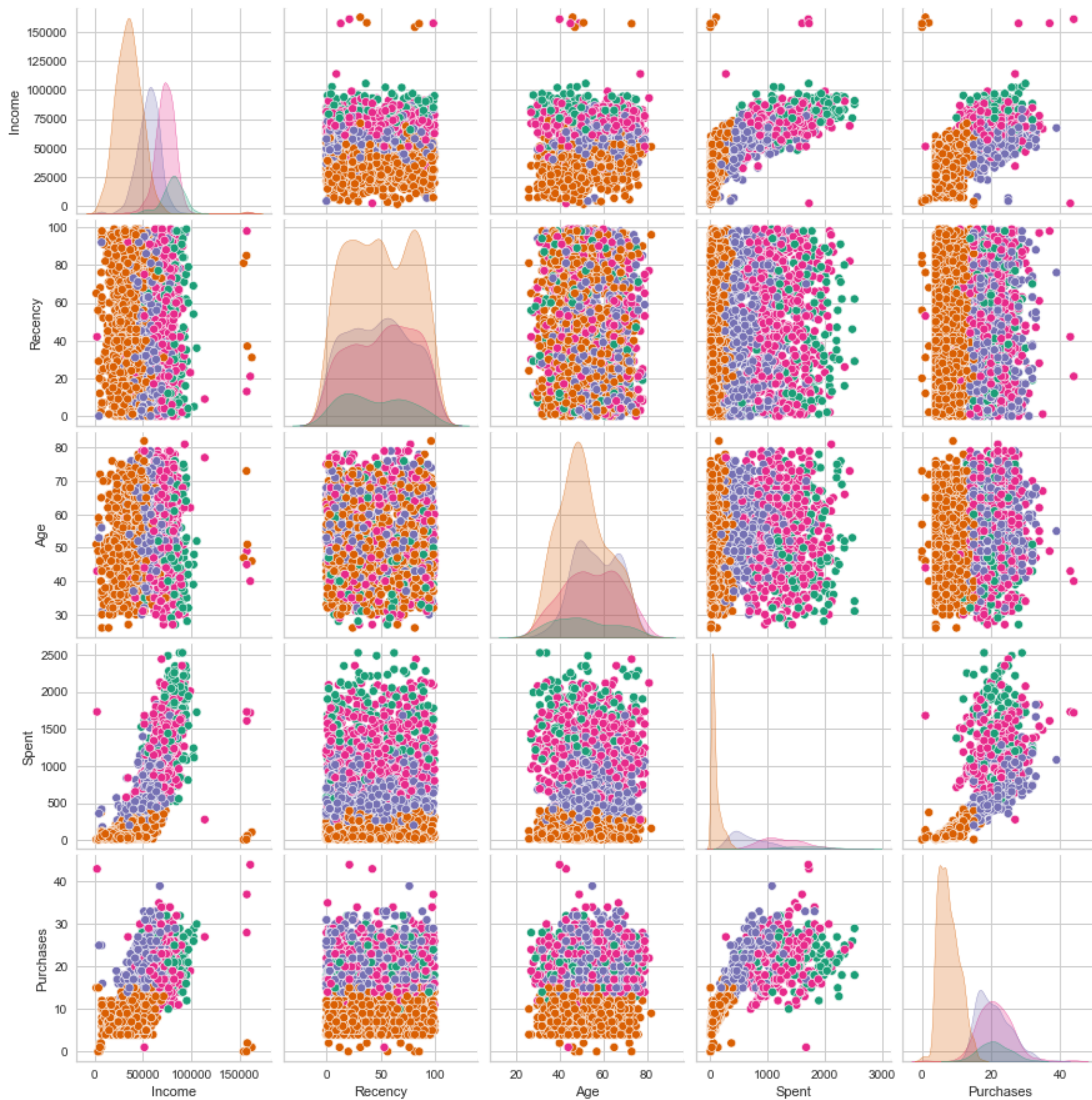
- **Gastos elevados;**
- **Maioria possui pelo menos um filho;**
- **Renda mediana, menor que os cluster 0 e 3;**



Cluster 3

- Gastos maiores que o cluster 2;
- Renda elevada, aproximando-se do cluster 0;
- Maior parte não possui filhos, alguns possuem um;





Conclusões

Os clientes agregados no **cluster 0** apesar de possuírem comportamento semelhante aos dos **clusters 2 e 3** no que tange à quantidade de item comprados, mostram-se aptos à gastarem uma quantidade maior em suas compras.

O **cluster 0** é um cliente em propenso a comprar itens mais caros e em maior quantidade dado seu gasto elevado, talvez seja interessante direcionar campanhas a tal grupo com itens mais sofisticados e/ou com uma temática romântica adulta.

O **cluster 1**, é um grupo que pode ser mais objetivo focar em produtos mais baratos, talvez algo voltado ao público infantil, visto que a maioria de seus indivíduos são pais e possuem uma renda mais baixa

Para o **cluster 2** campanhas voltadas a crianças e/ou com itens de valor mediano pode surtir um efeito positivo

Para o **cluster 3** campanhas e produtos com valor médio para alto, sem enfoque infantil, demonstra ser algo plausível de aceitação por tal grupo.

