

Aprendizado de Máquina

Prof. Aran Morales, UNISUL – Anima Educação

Apostila 9: Agrupamentos

Análises Preditivo: Clusters



Análises Preditivo: Clusters

Aprendizagem supervisionada:

o sistema é apresentado com exemplos de “insumos” e os resultados desejados, dados por um “target”, e o objetivo é aprender um modelo (regras) que mapeia entradas para saídas.

Aprendizagem não supervisionada:

nenhum rótulo é dado ao algoritmo de aprendizagem, deixando-o sozinho para encontrar a estrutura em sua entrada.

A aprendizagem não supervisionada pode ser um objetivo em si (descobrir padrões ocultos em dados).

Análises Preditivo: Clusters

Problemas com a aprendizagem não supervisionada:

A aprendizagem não supervisionada é mais difícil em comparação com tarefas de aprendizagem supervisionada.

Como sabemos se os resultados são significativos, pois nenhum rótulo de resposta está disponível?

Deixar o “especialista” analisar os resultados (avaliação externa)

Definir uma “função objetivo” no clustering (avaliação interna)

Agrupamentos



Agrupamentos: Conceitos

Conjunto de métodos usados para a construção de **grupos de objetos** com base nas **semelhanças** e **diferenças** entre os objetos, de tal maneira que os grupos obtidos são os mais homogêneos e bem separados possíveis.

A Clusterização é uma tarefa prévia à classificação. Sem classes, não se pode determinar a pertinência de um objeto em determinado contexto.

Agrupamentos: Conceitos

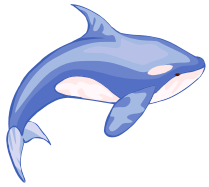
O problema de **Clustering** é descrito como:

tendo um conjunto de dados, de objetos, tentar agrupá-los de forma que os elementos que compõem cada grupo sejam mais parecidos entre si do que parecidos com os elementos dos outros grupos.

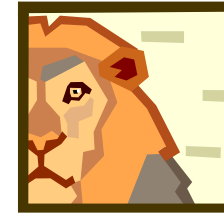
Em resumo, é colocar os iguais (ou quase iguais) juntos num mesmo grupo e os desiguais em grupos distintos.

Agrupamentos: Conceitos

- Como agrupar os animais seguintes?



Com bico



Sem bico



Ovíparo



Mamífero

Água

Terra

Agrupamentos: Conceitos

Como **medir** a **similaridade** entre os itens? (como qualificar os itens)

Como formar os **agrupamentos**?

(que variáveis fazem parte da geração dos agrupamentos)

Quantos grupos devem ser formados?

(como definir o número de agrupamentos, ou o raio de abrangência do agrupamento).

Agrupamentos: Conceitos

Trabalha com dados **categóricos** e **numéricos** e é de **fácil aplicação**;

Alta dependência na escolha da **métrica** de **similaridade**;

Sensibilidade aos **parâmetros** iniciais e **tipos de dados**;

Pode ser difícil **interpretar** os **resultados** alcançados;

Agrupamentos: Conceitos

Uma forma de **classificar** os **algoritmos** de **Clustering** é:

- **Métodos por particionamento;**
- **Métodos hierárquicos;**
- Métodos baseados em densidade;
- Métodos baseados em grades;
- Métodos baseados em modelos.

Medidas de similaridade



Agrupamentos: Similaridade

Cluster baseado em distância .

Dado um conjunto de pontos, e a noção (definição) de distância entre pontos, devemos agrupar os pontos em grupos (*clusters*), de modo que as distâncias internas (dentro do cluster) devem ser pequenas, ou seja, membros de clusters são próximos / similares.

As distâncias externas (intracluster) devem ser grandes, ou seja, os membros de diferentes clusters são diferentes.

Agrupamentos: Similaridade

As **medidas** de **similaridade** fornecem valores numéricos que expressam a “**distância**” (correlação ou associação) entre dois objetos.

Quanto **menor** o valor da “distancia”, **mais semelhantes** serão os objetos e deverão estes ficarem no **mesmo cluster**.

Não há uma **medida** de **similaridade** que sirva para todos os **tipos** de **variáveis** que podem existir numa base de dados.

A **similaridade** pode ser medida de diversas formas:

- Medidas Correlacionais (e.g., correlação de Pearson);
- Medidas de Distância (e.g., distância euclidiana);
- Medidas de Associação (e.g., índice de Jaccard)

Agrupamentos: Similaridade

Uma **função de distância** deve ser tal que:

- não assuma valores negativos (o menor valor é **0**);
- ser simétrica (a distância do objeto **i** ao **j** tem que ser igual à distância do objeto **j** ao **i**);
- forneça o valor **0** quando calculada a distância do objeto a si mesmo;
- respeitar a desigualdade triangular, (dados **3** objetos, a distância entre dois deles tem que ser menor ou igual a soma das distâncias entre esses dois objetos e o terceiro).

Agrupamentos: Similaridade

Variáveis numéricas:

A medida é normalmente usada para computar as dissimilaridades de objetos descritos por tais variáveis é

a: **Distancia Euclidiana**

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2}$$

Agrupamentos: Similaridade

- Representam a similaridade como a proximidade entre observações (instâncias) ao longo dos atributos (As medidas de distância são, na verdade, uma medida de dissimilaridade, em que os valores maiores denotam menor similaridade);

	X1	X2	X3	X4	X5
Cliente_1	7,000	10,000	9,000	7,000	10,000
Cliente_2	9,000	9,000	8,000	9,000	9,000
Cliente_3	5,000	5,000	6,000	7,000	7,000
Cliente_4	6,000	6,000	3,000	3,000	4,000
Cliente_5	1,000	2,000	2,000	1,000	2,000
Cliente_6	4,000	3,000	2,000	3,000	3,000
Cliente_7	2,000	4,000	5,000	2,000	5,000

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2}$$

	Cliente_1	Cliente_2	Cliente_3	Cliente_4	Cliente_5	Cliente_6	Cliente_7
Cliente_1	0,00						
Cliente_2	3,32	0,00					
Cliente_3	6,86	6,63	0,00				
Cliente_4	10,24	10,20	6,00	0,00			
Cliente_5	15,78	16,19	10,10	7,07	0,00		
Cliente_6	13,11	13,00	7,28	3,87	3,87	0,00	
Cliente_7	11,27	12,16	6,32	5,10	4,90	4,36	0,00

Agrupamentos: Similaridade

$x = (\langle \text{idade}(x), \text{altura}(x), \text{peso}(x) \rangle)$, Exemplo:

joão = $(\langle 36, 1.80, 76 \rangle)$

josé = $(\langle 30, 1.78, 72 \rangle)$

maria = $(\langle 25, 1.65, 60 \rangle)$

Calculo da distância euclidiana: $d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2}$

$$d(\text{joão}, \text{josé}) = [(36-30)^2 + (1.80-1.78)^2 + (76-72)^2]^{1/2}$$

$$= (36+0.0004+16)^{1/2} = \mathbf{7,21 \text{ (são os mais similares)}}$$

$$d(\text{joão}, \text{maria}) = (121+0.0225+256)^{1/2} = 19,41$$

$$d(\text{josé}, \text{maria}) = (25+0.0169+144)^{1/2} = 13,00$$

Algoritmos por Particionamento

Algoritmo K-means



Agrupamentos: k-means

O algoritmo **k-means** toma um parâmetro de entrada, **k**, e particiona um conjunto de **n** objetos em **k** clusters tal que, a similaridade dos elementos “entre grupos” resultante é alta, mas a similaridade dos elementos “de grupos diferentes” é baixa.

O algoritmo **k-means**, trabalha bem quando os clusters são densos e compactos e bem separados uns dos outros.

Agrupamentos: k-means

O Algoritmo **k-means** para particionamento de objetos baseia-se no valor médio das distâncias dos objetos no cluster.

Entrada: O número de clusters, **k**, e a base de dados com **n** objetos.

Saída: Um conjunto de **k** clusters que minimizam o critério do erro quadrado.

Método:

1. Escolha arbitrariamente **k** objetos da base de dados como os centros iniciais dos clusters;

2. Repita

- 2.1 (Re)atribua cada objeto ao cluster ao qual o objeto é mais similar, de acordo com o valor médio dos objetos no cluster;
- 2.2. Calcule o valor médio dos objetos para cada cluster;

Até que não haja mudança de objetos de um cluster para outro.

Agrupamentos: k-means

Dataset a ser agrupado

Item	Variáveis	
	x1	x2
A	5	3
B	-1	1
C	1	-2
D	-3	-2

Agrupamentos: k-means

Passo 1

Particiona-se os itens em dois *clusters* (AB) e (CD) e calcula-se a coordenada (x_1, x_2) do centróide do *cluster*.

Cluster	Coordenadas dos centros	
	x_1	x_2
C1=(AB)	$(5 + (-1)) / 2 = 2$	$(3 + 1) / 2 = 2$
C2=(CD)	$(1 + (-3)) / 2 = -1$	$(-2 + (-2)) / 2 = -2$

Agrupamentos: k-means

Passo 2

Calcula-se a similaridade de cada item em relação ao centroide e em relação a cada item no grupo mais próximo. Se um item é movido o centroide do *cluster* dever ser atualizado.

$$d(A;C1) = [(5 - 2)^2 + (3 - 2)^2]^{1/2} = (10)^{1/2} = \mathbf{3,16}$$

$$d(A;C2) = [(5 - (-1))^2 + (3 - (-2))^2]^{1/2} = (61)^{1/2} = 7,81$$

O item A esta no cluster “certo”

$$d(B;C1) = [(-1 - 2)^2 + (1 - 2)^2]^{1/2} = (10)^{1/2} = 3,16$$

$$d(B;C2) = [(-1 - (-1))^2 + (1 - (-2))^2]^{1/2} = (9)^{1/2} = \mathbf{3,00}$$

O item B esta no cluster “errado”

Agrupamentos: k-means

Ocorre o deslocamento do item (B) para o segundo *cluster* e calcula-se novamente as coordenadas.

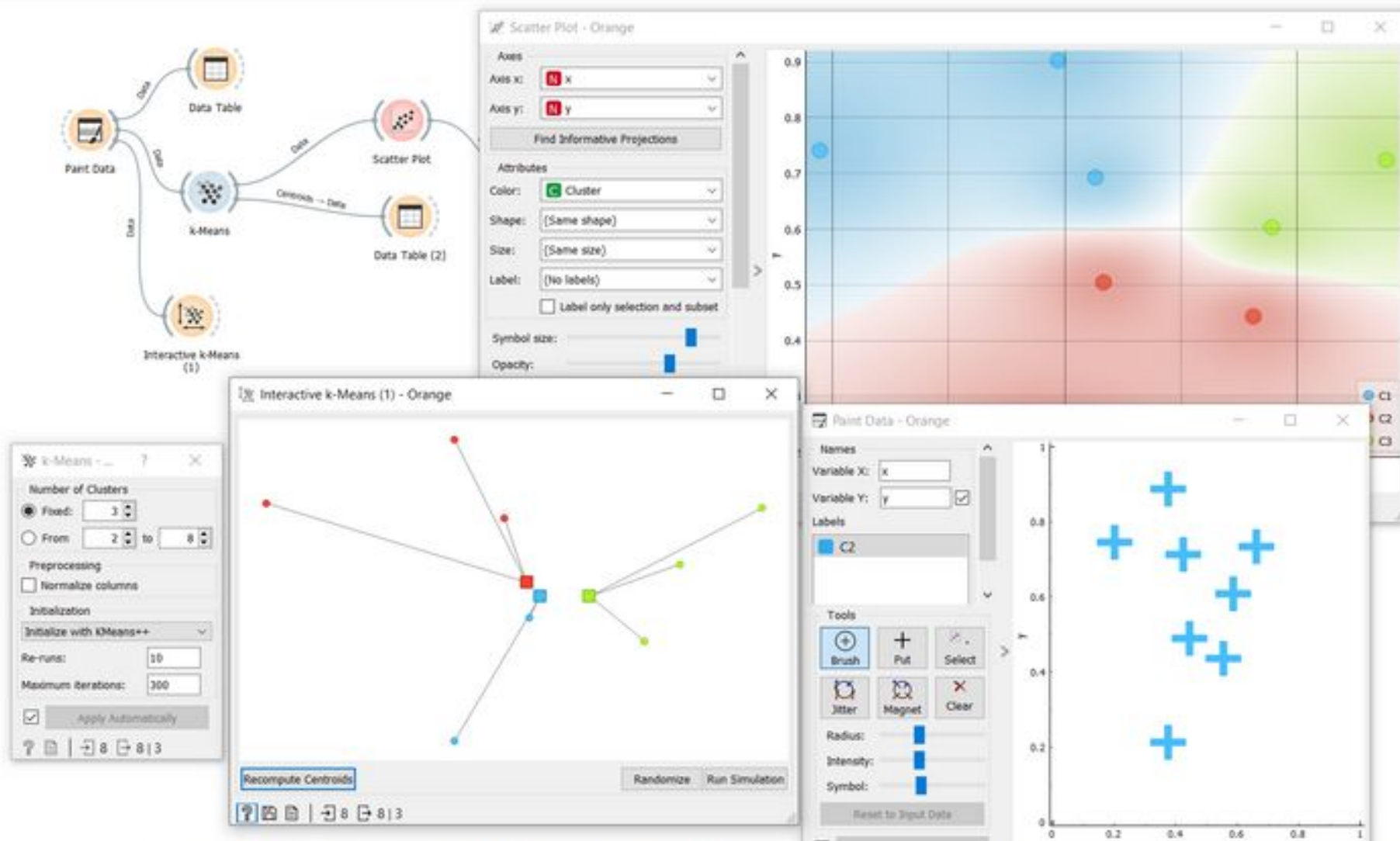
Cluster	Coordenadas dos centros	
	x1	x2
C1=(A)	5	3
C2=(BCD)	-1	-1

Agrupamentos: k-means

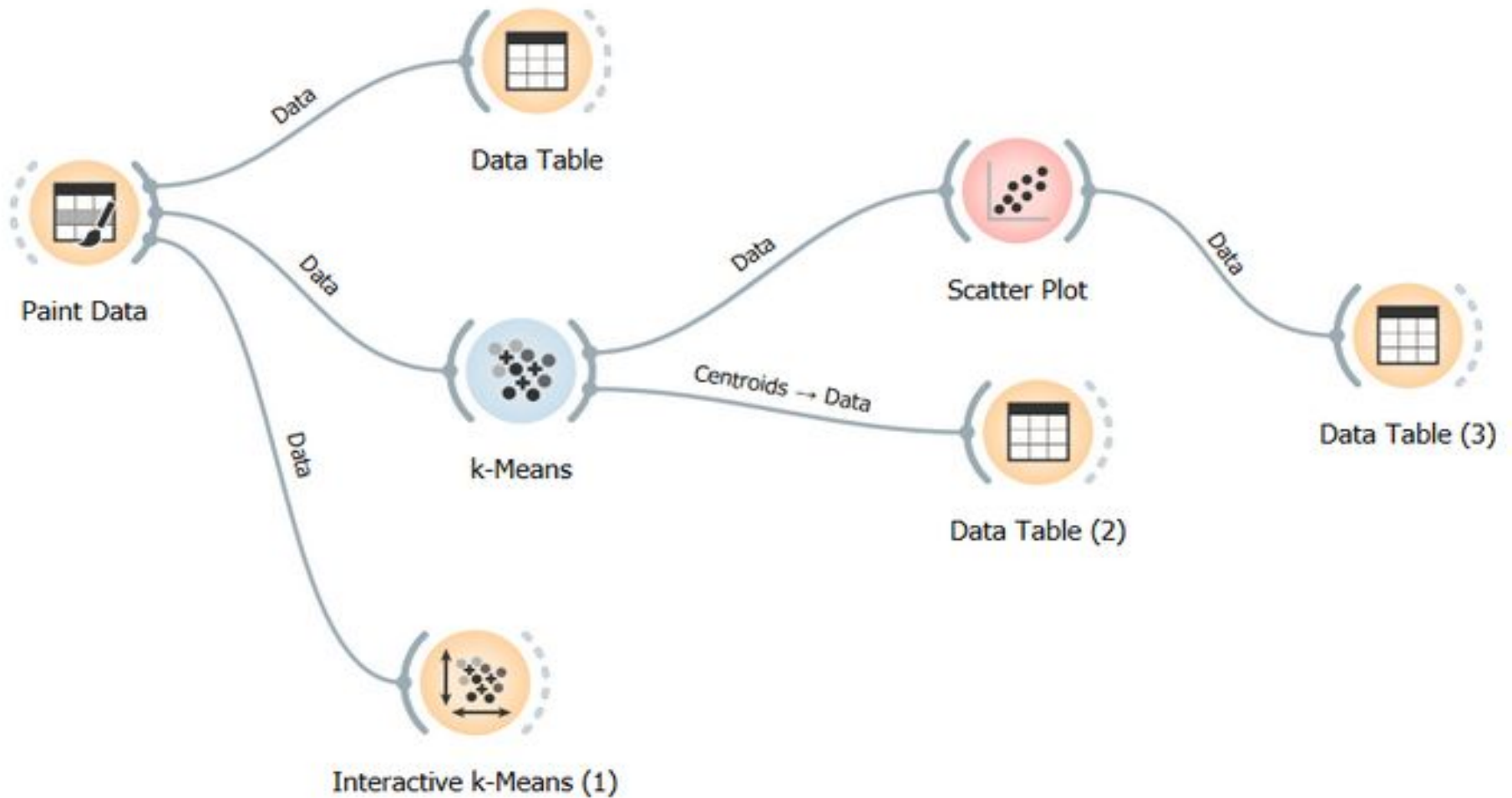
Calcula-se a distância dos itens em relação ao cluster, para verificar a parada do algoritmo.

Cluster	Distâncias dos centróides			
	A	B	C	D
C1=(A)	0	40	41	89
C2=(BCD)	52	4	5	5

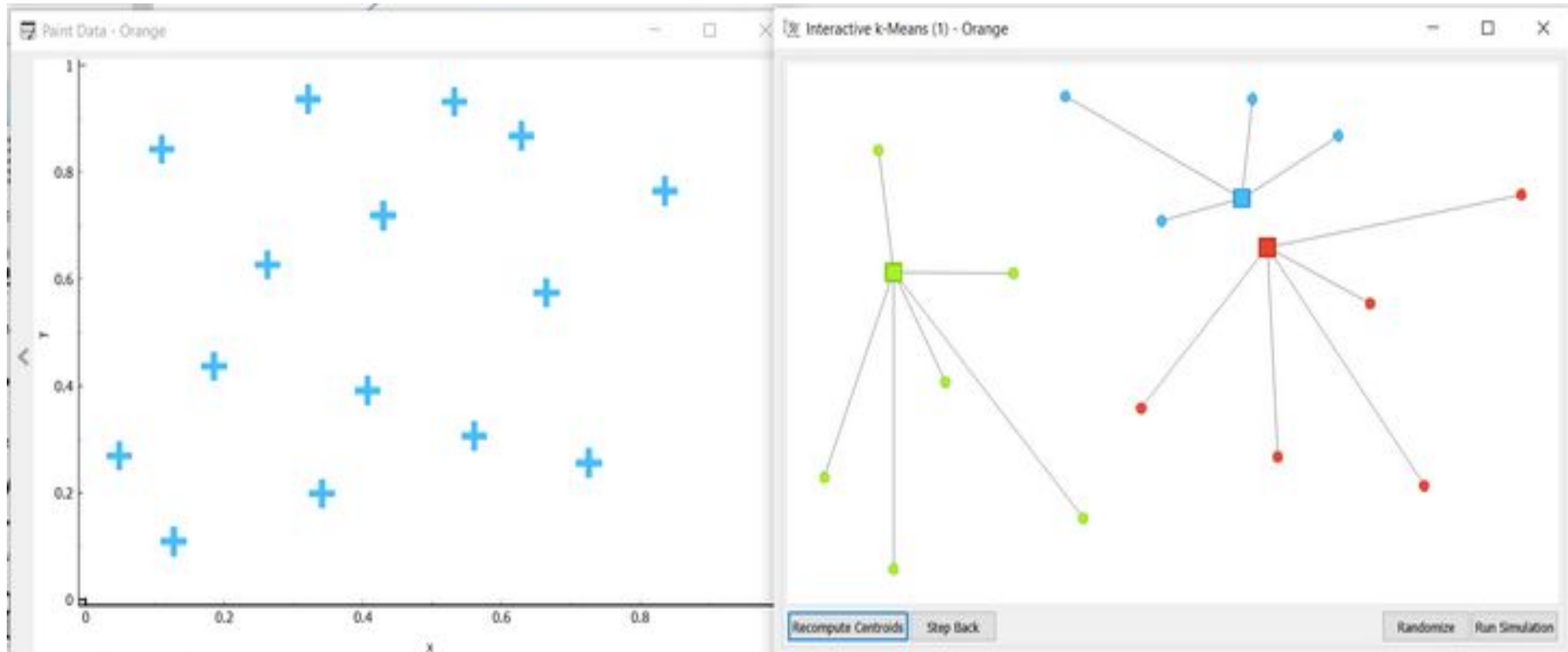
K-means: Exemplo



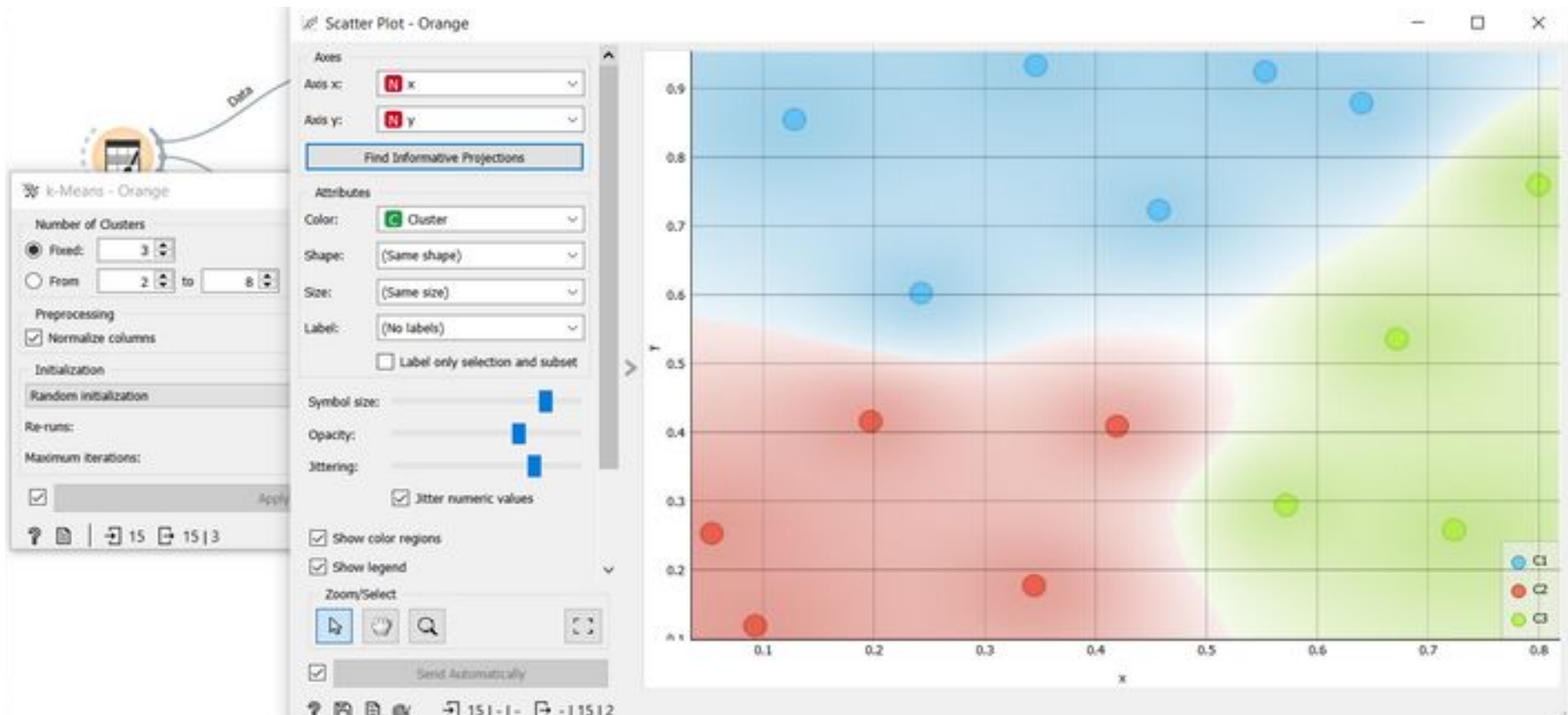
K-means: Exemplo



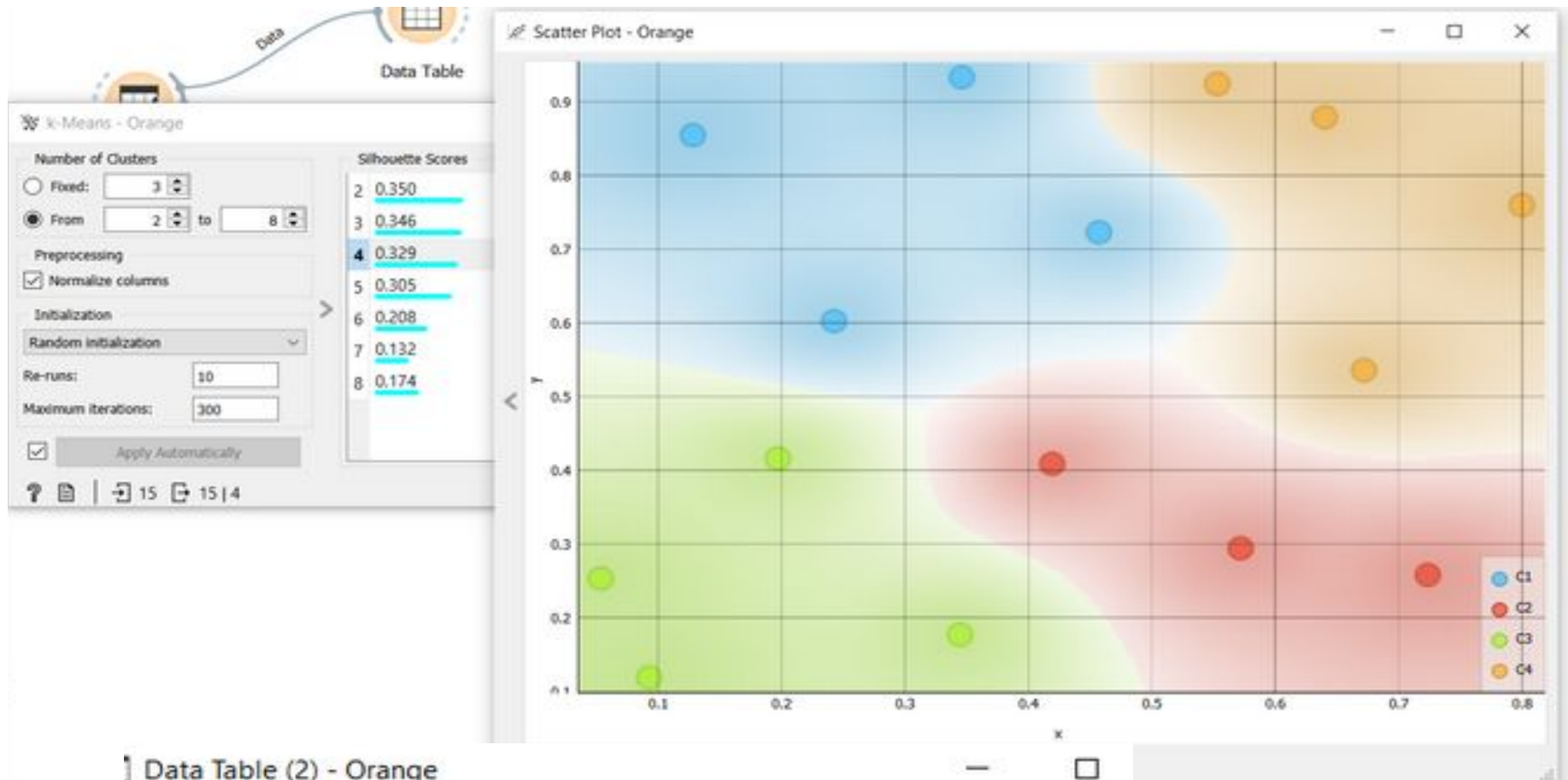
K-means: Exemplo



K-means: Exemplo



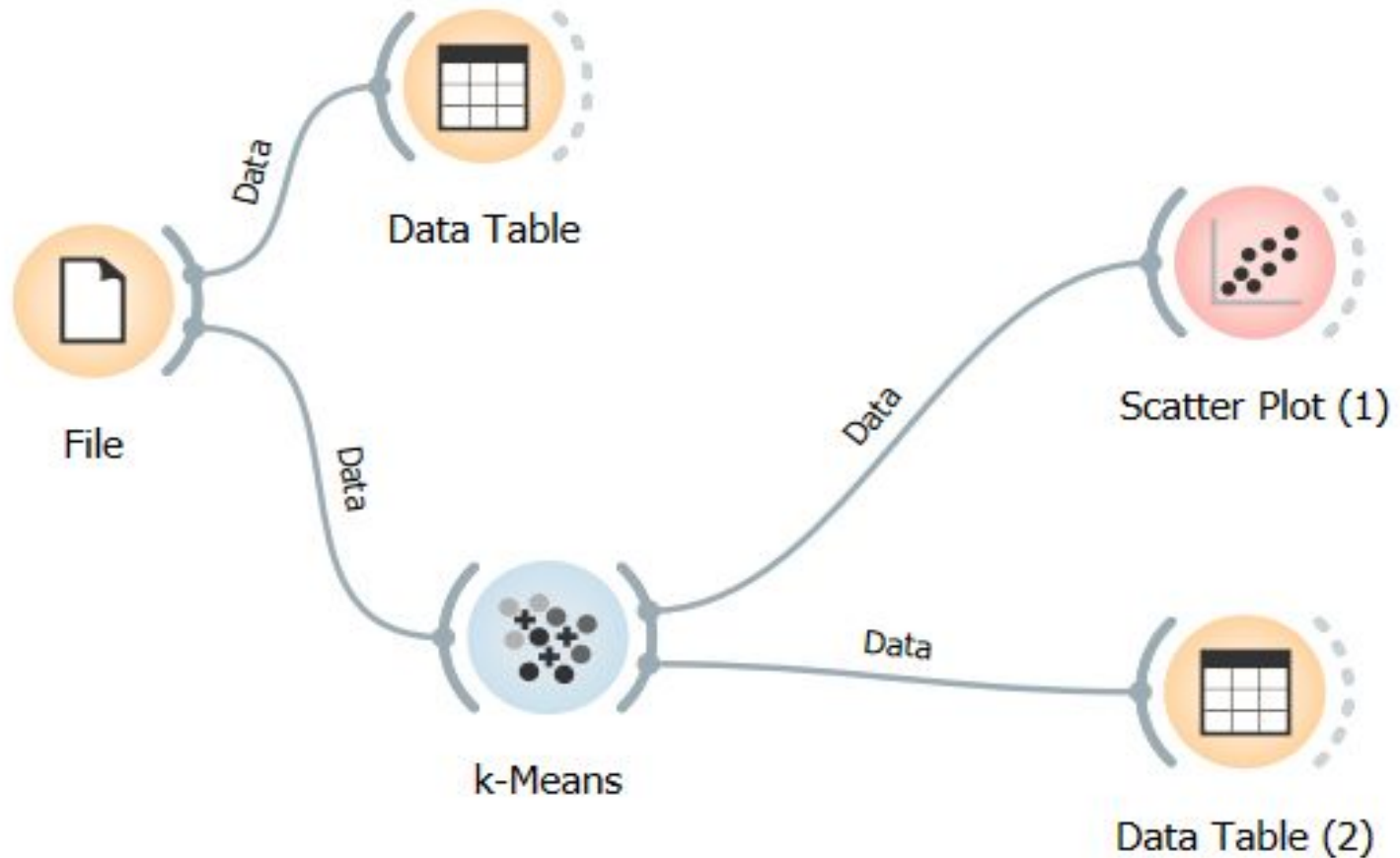
K-means: Exemplo



Data Table (2) - Orange

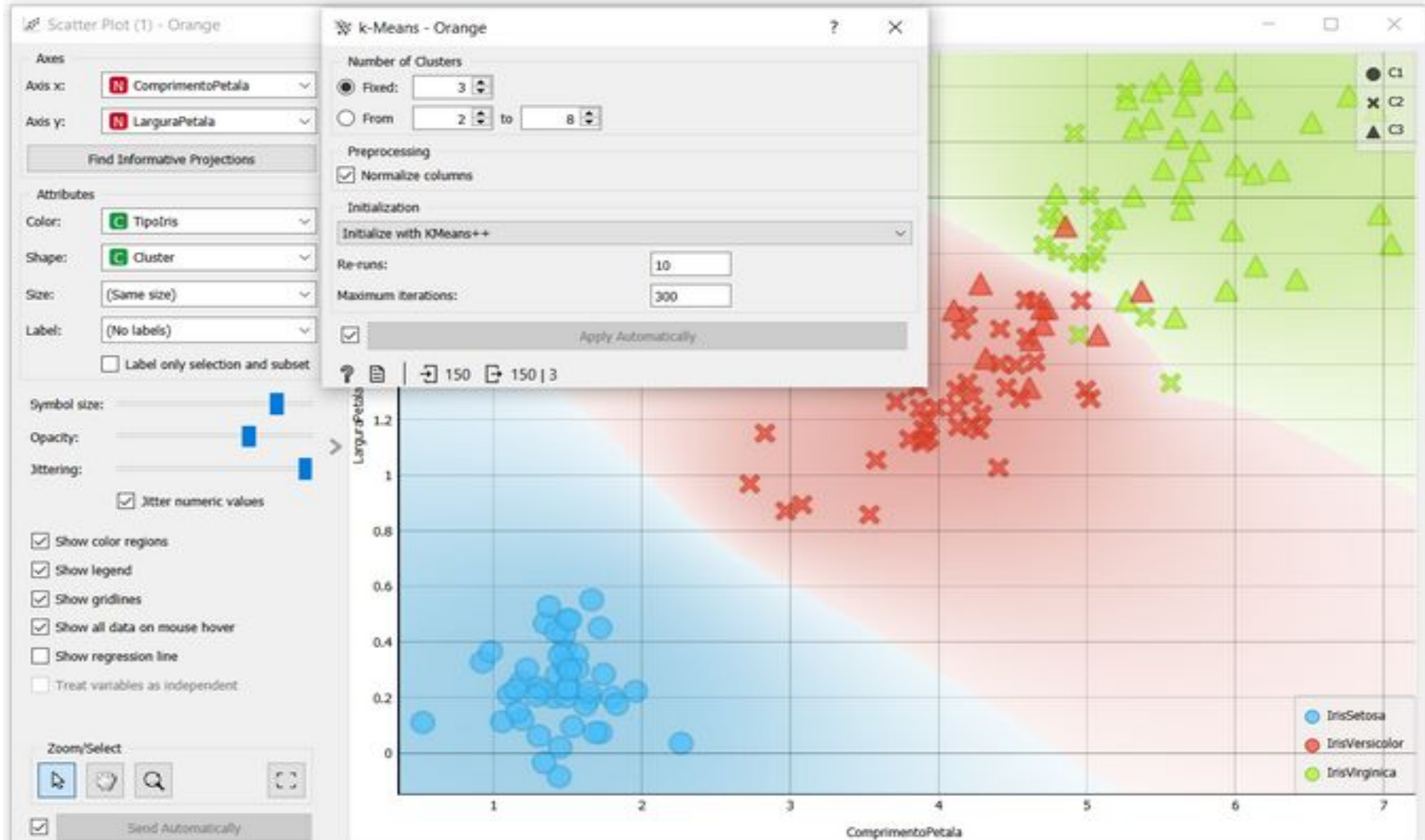
	Cluster	Silhouette	x	y
1	C1	0.599884	-0.5623	0.8421
2	C2	0.61142	0.6528	-0.8367
3	C3	0.602502	-1.0129	-1.0692
4	C4	0.582201	1.0856	0.8546

K-means: Flores Iris



K-means: Flores Iris

Algoritmo k-means: $k = 3$



K-means: Python



Exemplo_Kmeans_1.ipynb ☆

Arquivo Editar Ver Inserir Ambiente de execução Fe

+ Código + Texto

Exemplo do Algoritmo K-means

Importação de bibliotecas

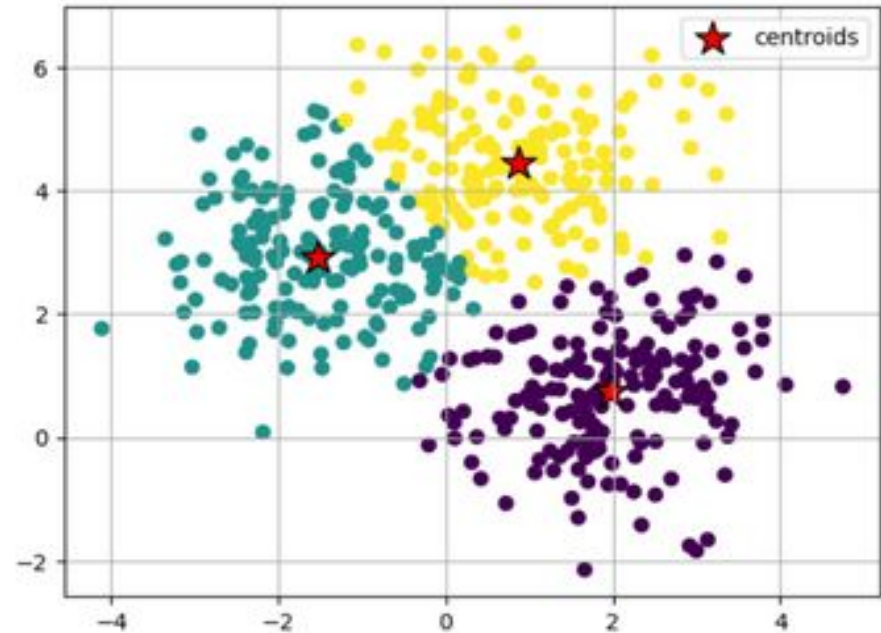
```
[ ] import matplotlib.pyplot as plt
    from sklearn.datasets import make_blobs
    from sklearn.cluster import KMeans
```

Criação do DataSet (dados) e visualização dos dados

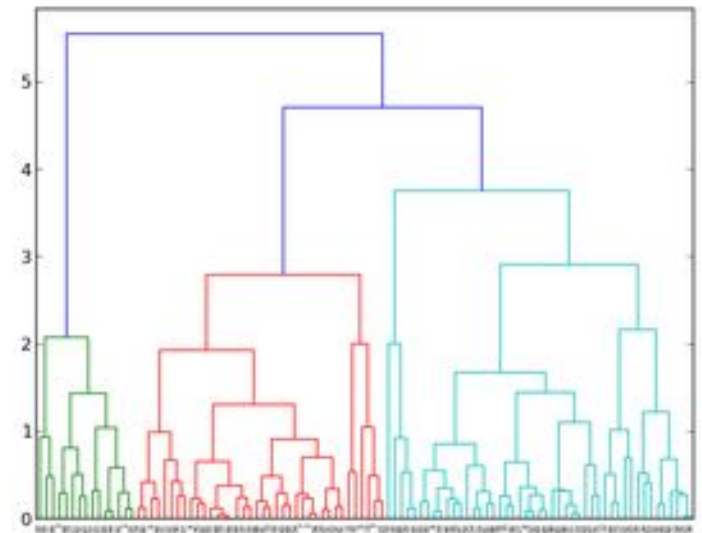
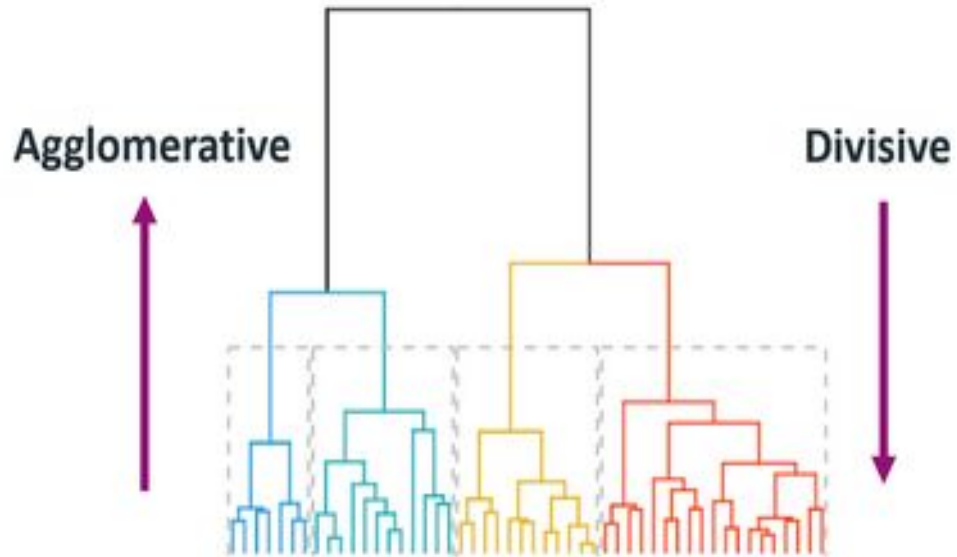
```
[ ] X, y = make_blobs(n_samples=500, n_features=2, random_state=0)

# plotamos os pontos
plt.scatter(
    X[:, 0], X[:, 1],
    c='white', marker='o',
    edgecolor='black')

plt.show()
```



Métodos Hierárquicos



Métodos Hierárquicos

Algoritmos hierárquicos criam uma decomposição hierárquica da base de dados, que é representada por um **dendrograma**, uma árvore que iterativamente divide a base de dados em subconjuntos menores até que cada subconjunto consista de somente um objeto.

O **dendrograma** pode ser criado de duas formas:

- **Abordagem aglomerativa** (*bottom-up*): parte-se das folhas para a raiz.
- **Abordagem divisiva** (*top-down*): parte-se da raiz para as folhas.

Métodos Hierárquicos

1. **Abordagem aglomerativa** (bottom-up): parte-se das folhas superiores para a raiz.
 - Começamos por colocar cada objeto em seu próprio cluster (ou seja, todos os objetos estão separados), totalizando n clusters.
 - Em cada etapa, calculamos a distância entre cada par de Clusters (matriz de distâncias).
 - Então, escolhemos **2** clusters com a distância mínima que serão transformados em um único cluster.
 - A seguir, atualizamos a matriz de distâncias. Este processo continua até que todos os objetos estejam em um único cluster (o nível mais alto da hierarquia), ou até que uma condição de término ocorra.

Métodos Hierárquicos

2. Abordagem divisiva (top-down): parte-se da raiz para as folhas.

- Invertemos o processo por começar com todos os objetos em um único cluster.
- Em cada etapa, um cluster é escolhido e dividido em dois clusters menores.
- Este processo continua até que tenhamos n clusters ou até que uma condição de término aconteça.

Métodos Hierárquicos

Algoritmo Geral de Agrupamento Hierárquico Aglomerativo

Passo 1: Iniciar o agrupamento formado por grupos unitários

Passo 2: Encontre o par de grupos de dissimilaridade mínima

Passo 3: Construa um novo grupo pela fusão desse par de grupos de dissimilaridade mínima

Passo 4: Atualize a matriz de dissimilaridades: suprima as linhas e as colunas correspondentes aos grupos fusionados e adicione uma linha e uma coluna correspondente as dissimilaridades entre o novo grupo e os grupos antigos

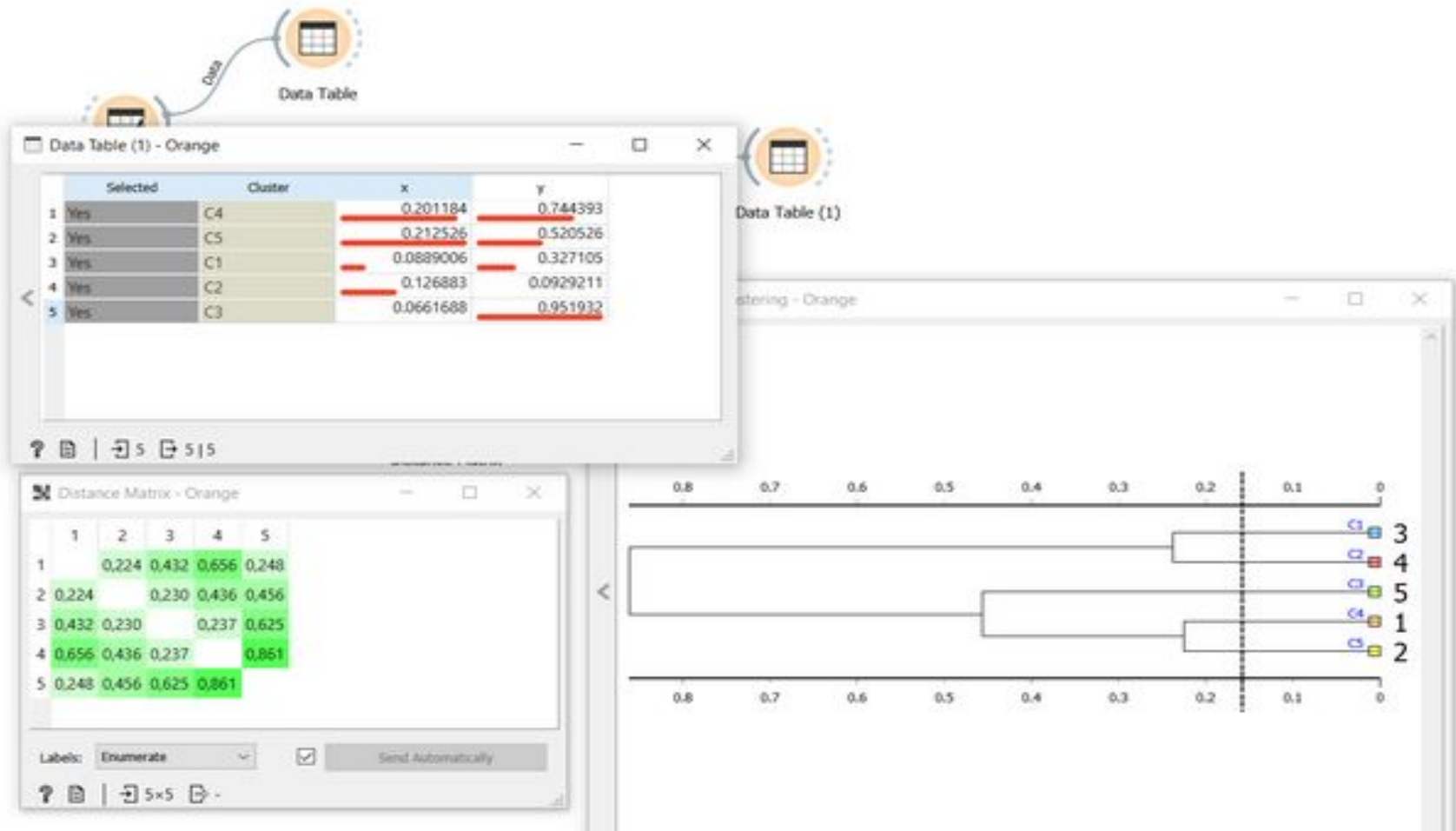
Passo 5: Se todos os objetos estão agrupados, pare; senão vá para o passo 2

Métodos Hierárquicos: Exemplo



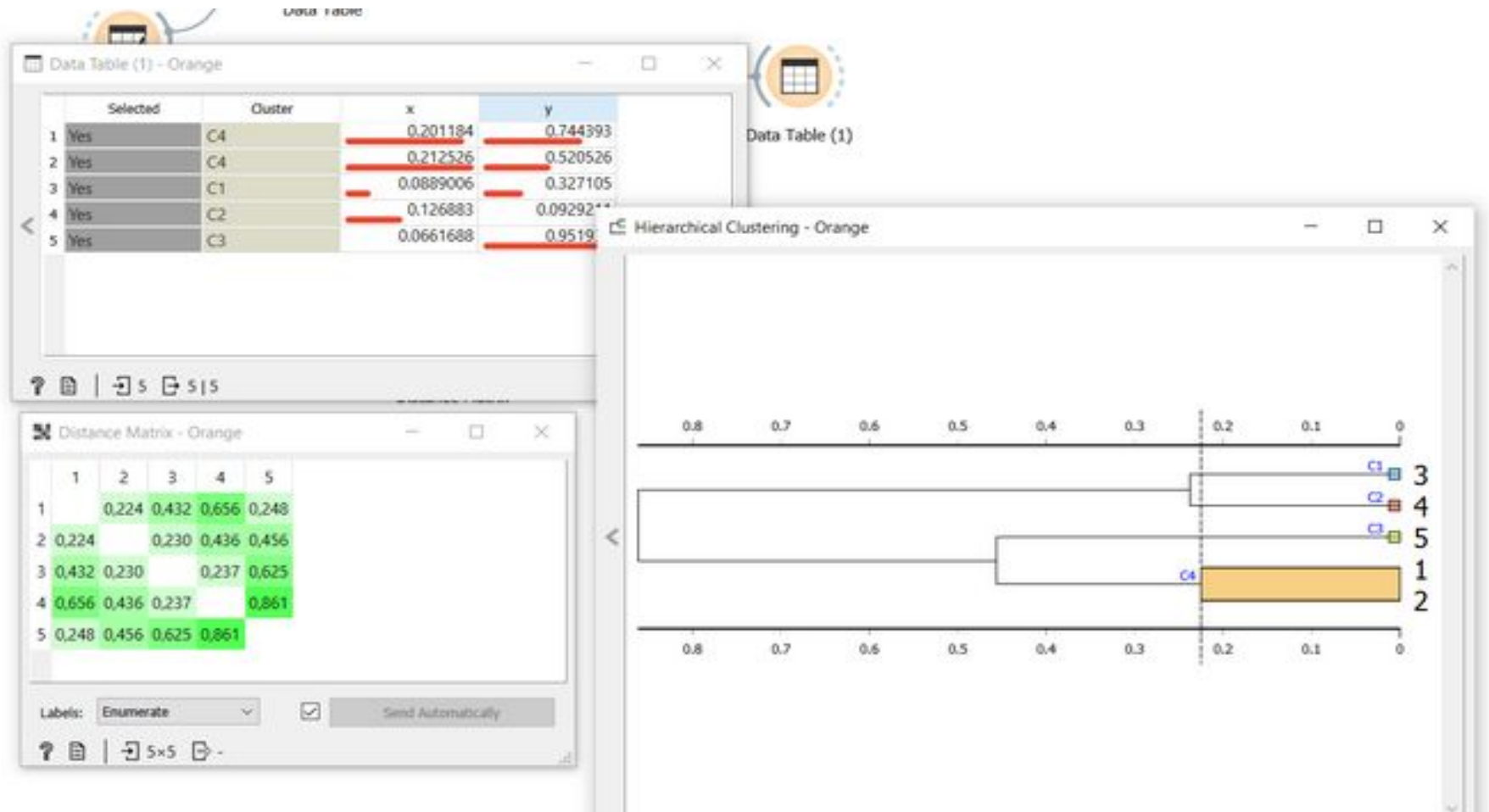
Métodos Hierárquicos: Exemplo

Observamos a matriz de distancia, e os agrupamentos formados pelos itens de forma individual (sozinhos).



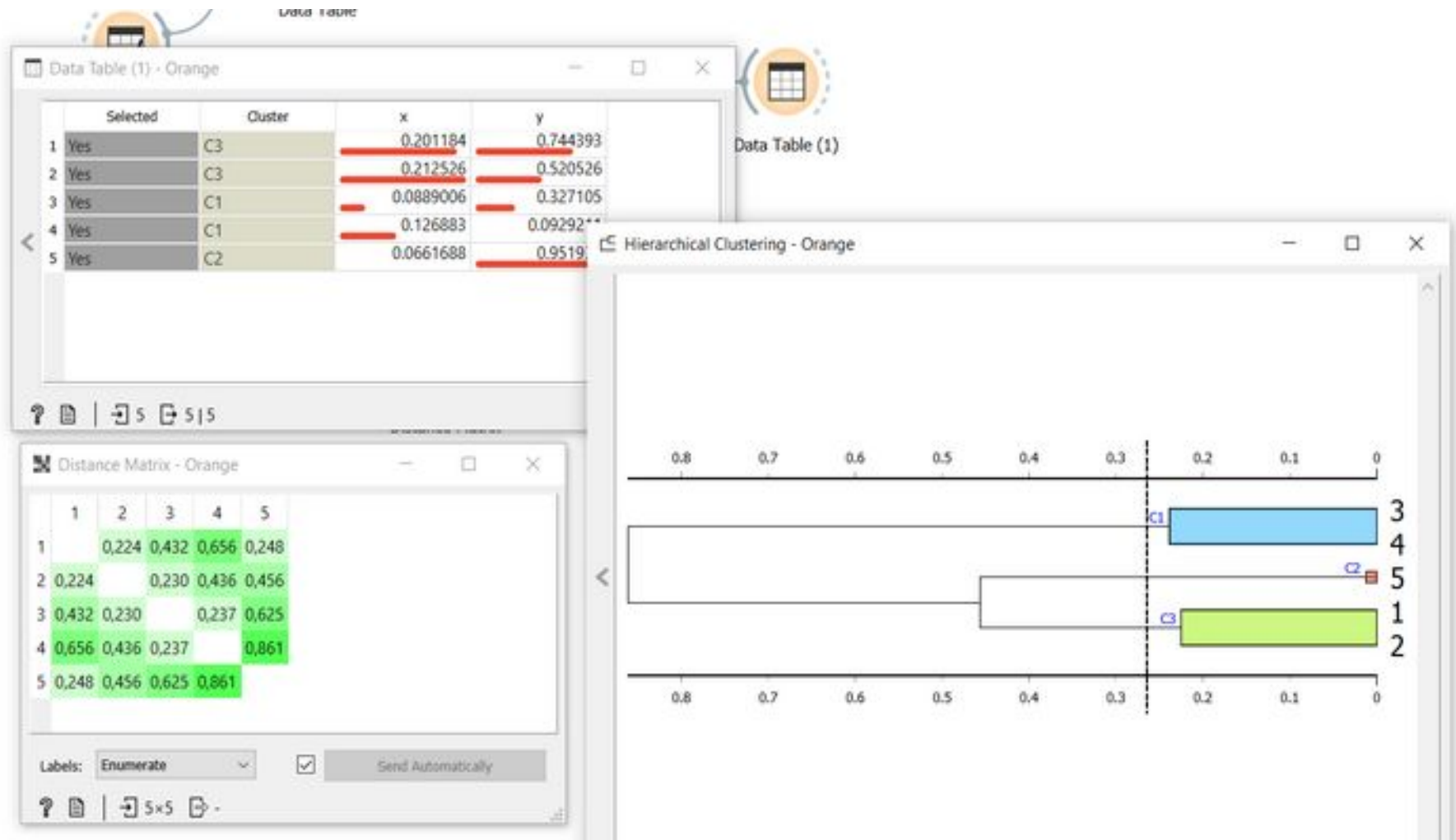
Métodos Hierárquicos: Exemplo

Os itens que tem menor distancia, foram o próximo agrupamento, itens 1 e 2 forma C4, os outros itens continuam sozinhos.



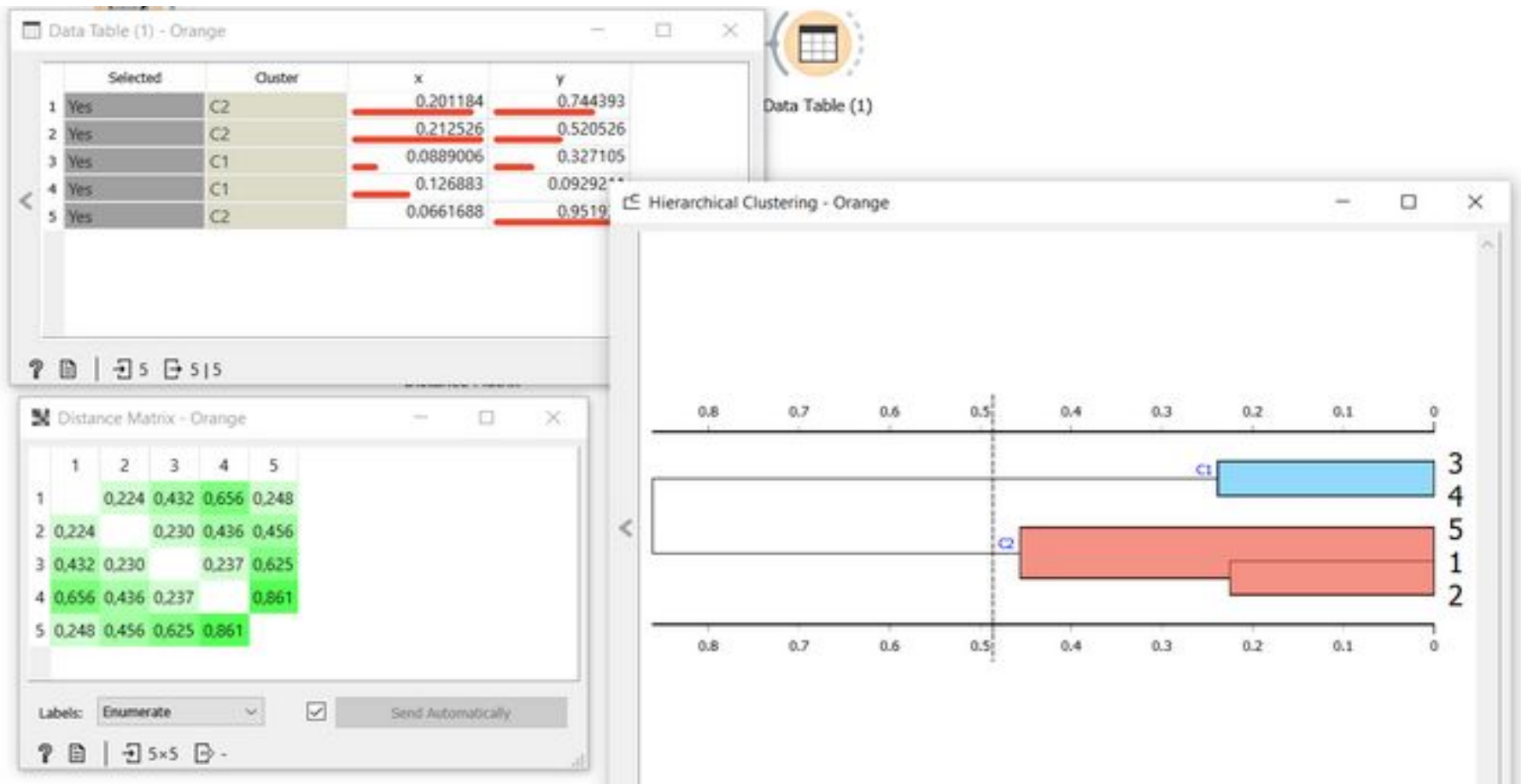
Métodos Hierárquicos: Exemplo

O próximo grupo é formado pelos itens 3 e 4 (C1).

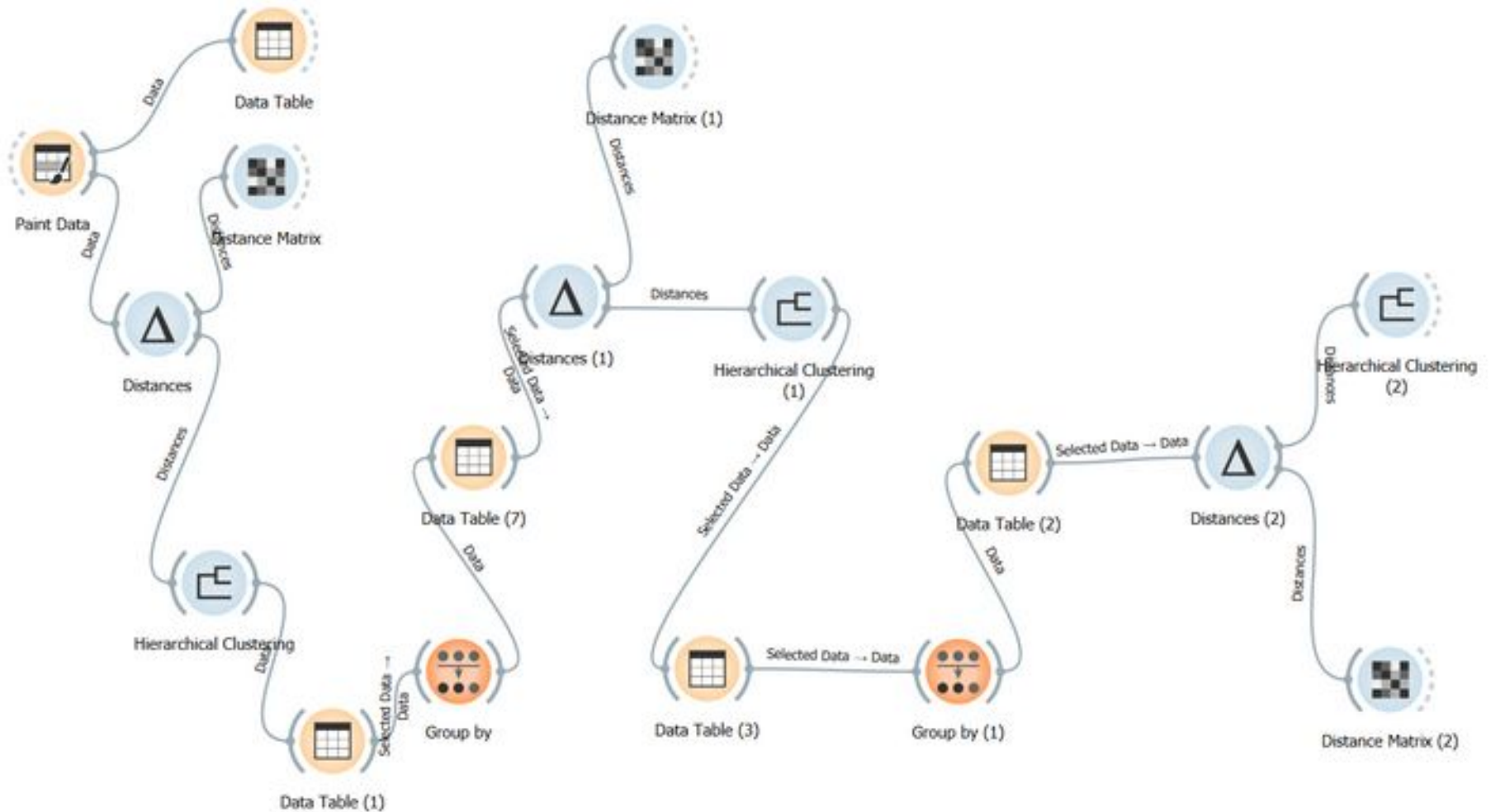


Métodos Hierárquicos: Exemplo

Continuando no mesmo raciocínio, o próximo grupo é formado pelos itens 1, 2 e 5 chamado agora de C2.

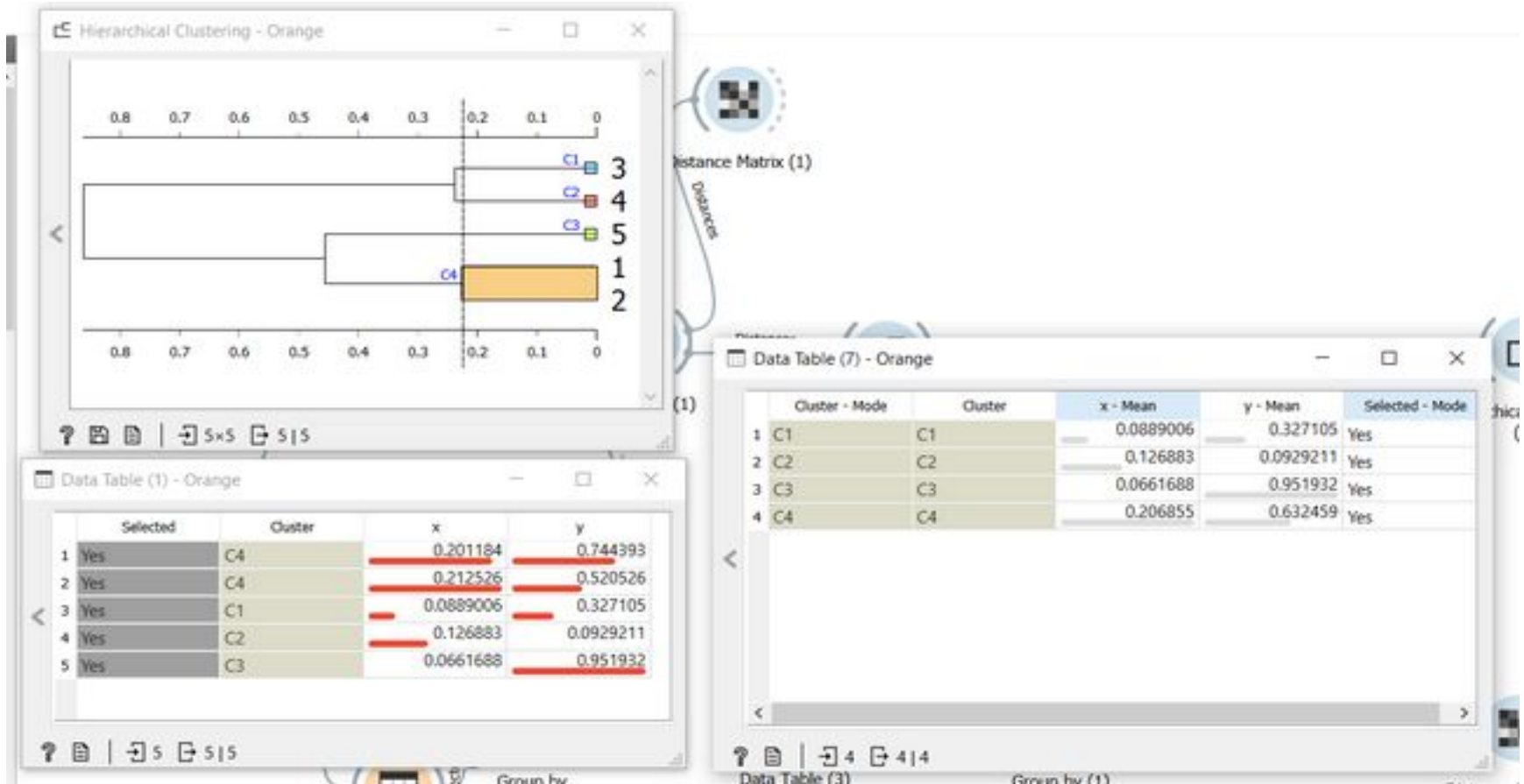


Exemplo: Passo a passo



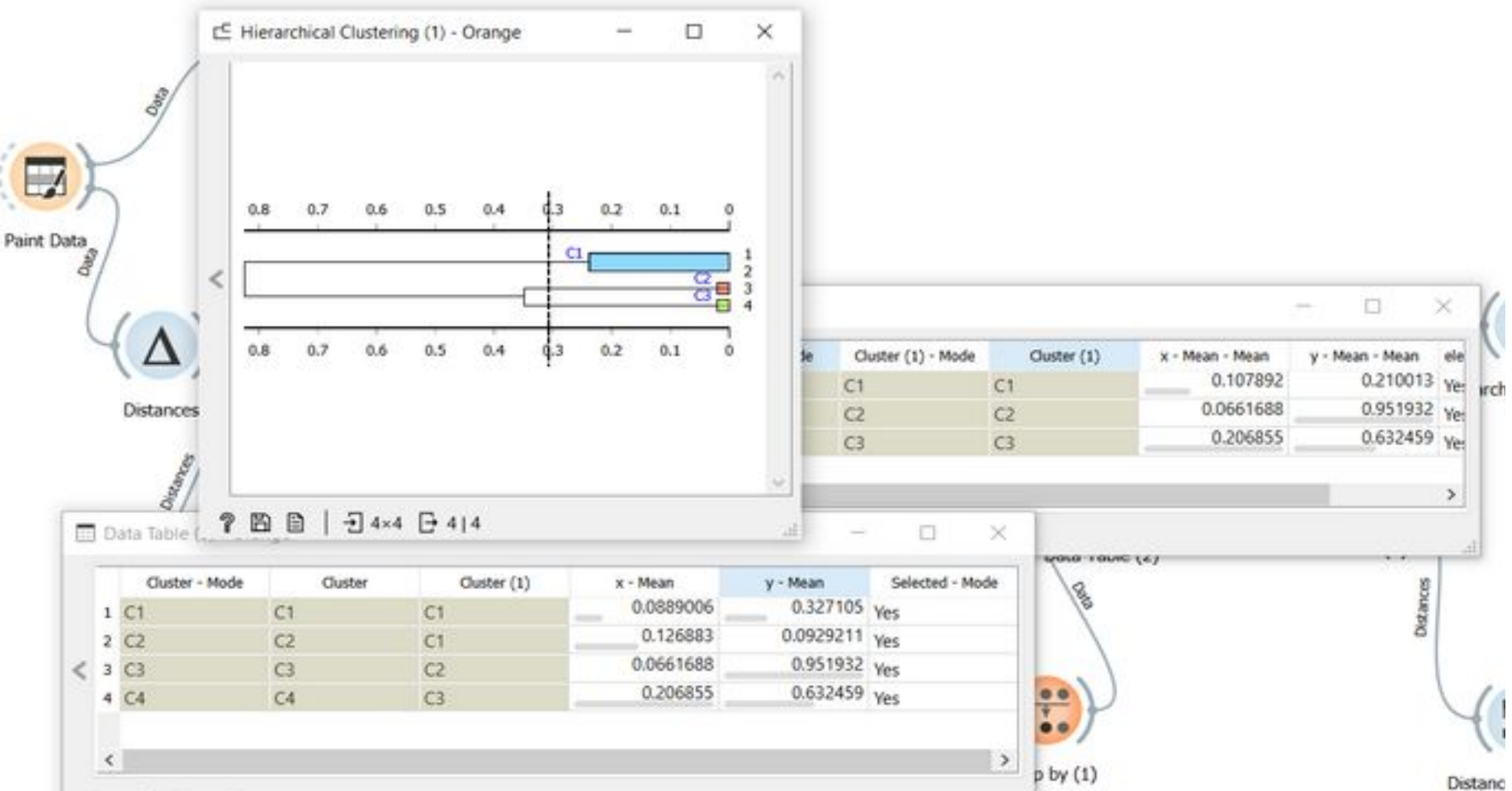
Exemplo: Passo a passo

Instancias 1 e 2 são agrupadas no C4, que formam um novo conjunto de coordenadas



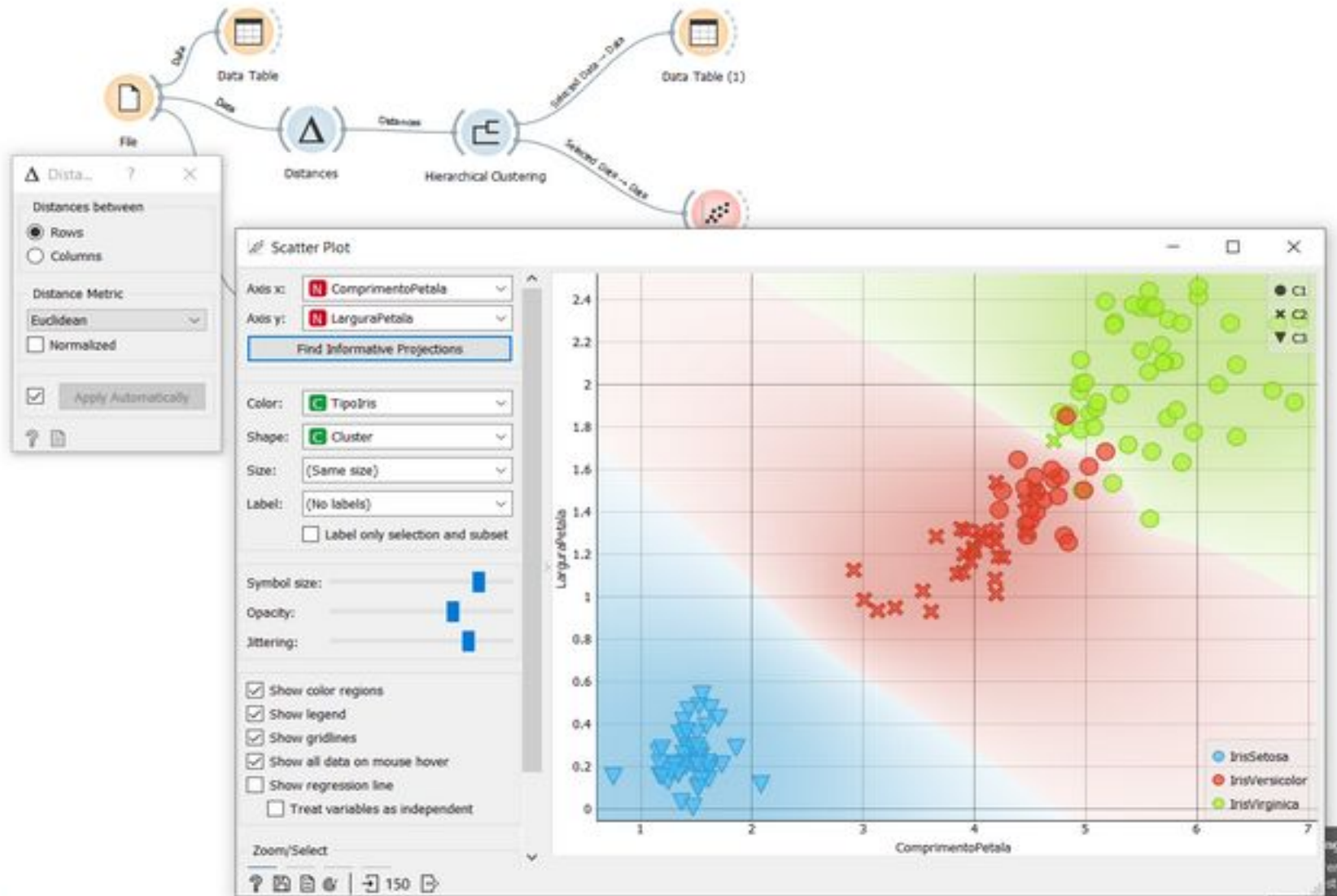
Exemplo: Passo a passo

Instancias 1 e 2 são agrupadas no C1, que formam um novo conjunto de coordenadas



Métodos Hierárquicos: Flores Iris

Algoritmo hierárquico



Exemplo: Segmentação de Mercado



Segmentação de Mercado

Definimos Segmentação de Mercado como a:

"Divisão do mercado-alvo ou clientes com base em algumas características significativas que poderiam ajudar uma empresa a vender (ou oferecer) mais produtos em menos despesas de marketing."

A "segmentação de mercado" é uma análise detalhada dos clientes de uma empresa.

O processo de "segmentação", ajuda uma empresa a **entender melhor seus clientes** e facilita a "**modificação**" de produtos de acordo com as necessidades, **direciona as campanhas de marketing** em função dos comportamentos e preocupações específicas de diferentes tipos de clientes.

Segmentação de Mercado

Por exemplo, em vez de gastar dinheiro para comercializar um novo produto para cada cliente no banco de dados da empresa, uma empresa pode analisar qual segmento de cliente tem mais probabilidade de comprar o produto e, em seguida, comercializar o produto apenas nesse segmento específico.

Segmentação de Mercado

A segmentação de mercado é um problema de aprendizado não supervisionado.

Isso significa que você não tem uma variável dependente.

Na segmentação de mercado as empresas, procuram dividir os clientes em diferentes segmentos para que diferentes estratégias de marketing possam ser aplicadas a segmentos distintos para retê-los.

Segmentação de Mercado

Atributos do Cliente

ID do cliente; ano de nascimento; nível de educação; estado civil; renda familiar anual; número de crianças na casa; número de adolescentes na casa do cliente; data do cadastro do cliente na empresa; número de dias desde a última compra do cliente; reclamação: 1 se o cliente reclamou nos últimos 2 anos, 0 caso contrário

Atributos dos Produtos comprados pelo cliente

valor gasto em vinho nos últimos 2 anos;
valor gasto em frutas nos últimos 2 anos;
valor gasto em carne nos últimos 2 anos;
valor gasto em pescado nos últimos 2 anos;
valor gasto em doces nos últimos 2 anos;
valor gasto em “outros” nos últimos 2 anos

Segmentação de Mercado

Atributos do Local da Compra realizada pelo Cliente

número de compras realizadas através do site da empresa;

número de compras feitas usando um catálogo;

número de compras feitas diretamente nas lojas;

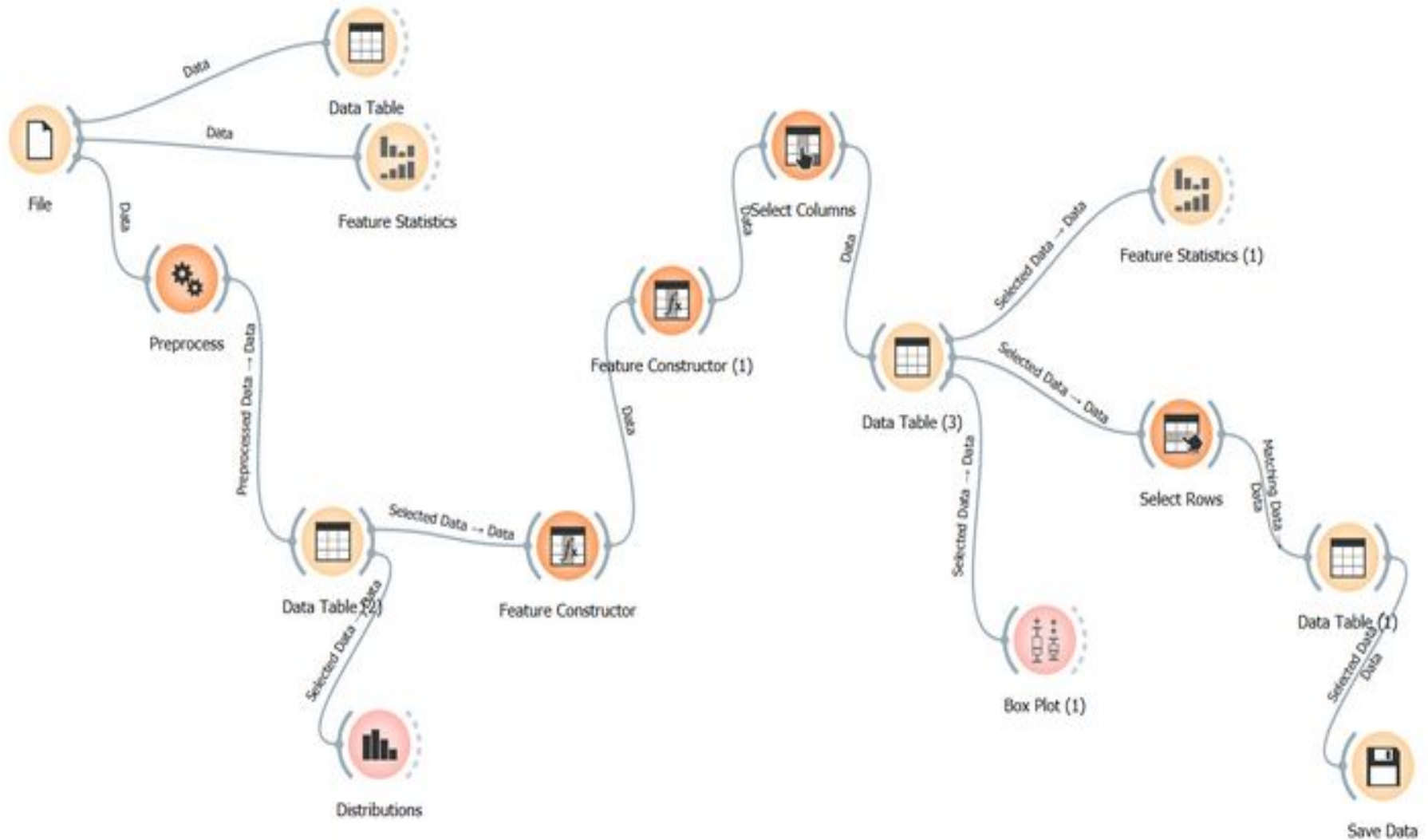
número de visitas ao site da empresa no último mês;

Atributos das promoções

número de compras feitas com desconto; 1ª campanha: 1 se o cliente aceitou e 0 em não aceito; idêntico para as 2ª campanha, 3ª campanha, 4ª campanha, 5ª campanha; resposta: 1 se o cliente aceitou a oferta na última campanha, 0 caso contrário

Segmentação de Mercado

Tratamento e Limpeza de Dados.



Segmentação de Mercado

Tratamento e Limpeza de Dados.

Temos 2.240 instancias e 28 variáveis.

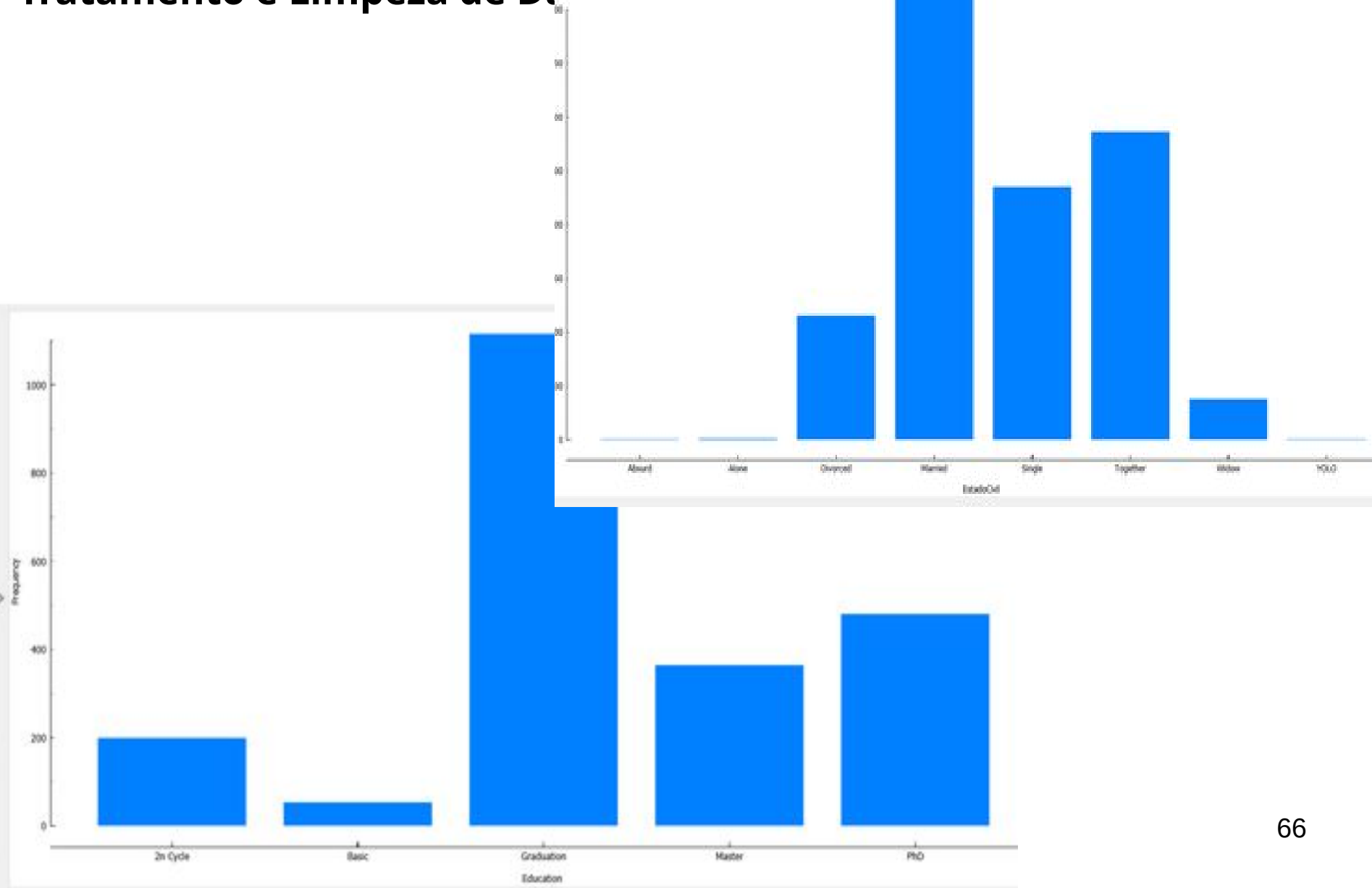
O atributo RendaAnual tem instancias sem dados, vamos retirar essas instancias da amostra.

The screenshot displays the Orange3 interface. On the left, a workflow includes a 'File' widget connected to a 'Preprocess' widget. The 'Preprocess' widget's 'Preprocessors' list has 'Impute Missing Values' selected, with the option 'Remove rows with missing values' chosen. A red circle highlights this option. In the top-left corner, a red circle highlights the data summary: '2240 instances', '28 features (0.0 % missing data)', 'No target variable', and 'No meta attributes'. The 'Feature Statistics - Orange' window shows a table with columns: Name, Distribution, Mean, Median, Dispersion, Min., Max., and Missing. The 'RendaAnual' row shows 24 missing values (1%). A red circle highlights this '24 (1%)' value.

Name	Distribution	Mean	Median	Dispersion	Min.	Max.	Missing
Ano_Nascimento		1968.806	1970.0	0.006	1893.0	1996.0	0 (0 %)
RendaAnual		11.5	0.482	1730.0	666666.0	24 (1 %)	
...	...	0.0	1.212	0.0	2.0	0 (0 %)	

Segmentação de Mercado

Tratamento e Limpeza de Dados



Segmentação de Mercado

Tratamento e Limpeza de Dados: vamos criar novas variáveis: idade; nova classificação de educação e estado civil, gastos (somando todas as despesas), nro. de filhos (crianças + adolescentes); clientes aceita ofertas (somando todas as variáveis de ofertas), nro. de compras (somando todos os canais de venda), tamanho da família e se o cliente é pai com filhos;

The screenshot shows the 'Variable Definitions' window in Orange Data Mining. It contains a list of newly created variables and their formulas:

- Nr_Compras_Total**: $Nr_ComprasWeb + Nr_ComprasCatalogo + Nr_ComprasLoja + Nr_ComprasDescontos$
- Idade**: $2022 - Ano_Nascimento$
- Gasto_Total**: $DespesasVinhos + DespesasFrutas + DespesasCarnes + DespesasPeixe + DespesasDoce + DespesasLacteos$
- NovoEstadoCivil**: "Parceiro" if (EstadoCivil == "Married" or EstadoCivil == "Together") else "Sozinho"
- Nr_Filhos**: $Nr_Crianças + Nr_Adolescentes$
- NovoEducacao**: "PosGraduado" if (Education == "PhD" or Education == "Master") else "Sem_Graduacao" if (Education == "2n Cycle" or Education == "Basic") else "Graduacao"
- ClienteAceitaOfertas**: "1" if (ClienteAceitoOferta_1 == "1" or ClienteAceitoOferta_2 == "1" or ClienteAceitoOferta_3 == "1" or ClienteAceitoOferta_4 == "1" or ClienteAceitoOferta_5 == "1" or ClienteAceitoOferta_Ultima == "1") else "0"
- Nr_Compras_Total**: $Nr_ComprasWeb + Nr_ComprasCatalogo + Nr_ComprasLoja + Nr_ComprasDescontos$

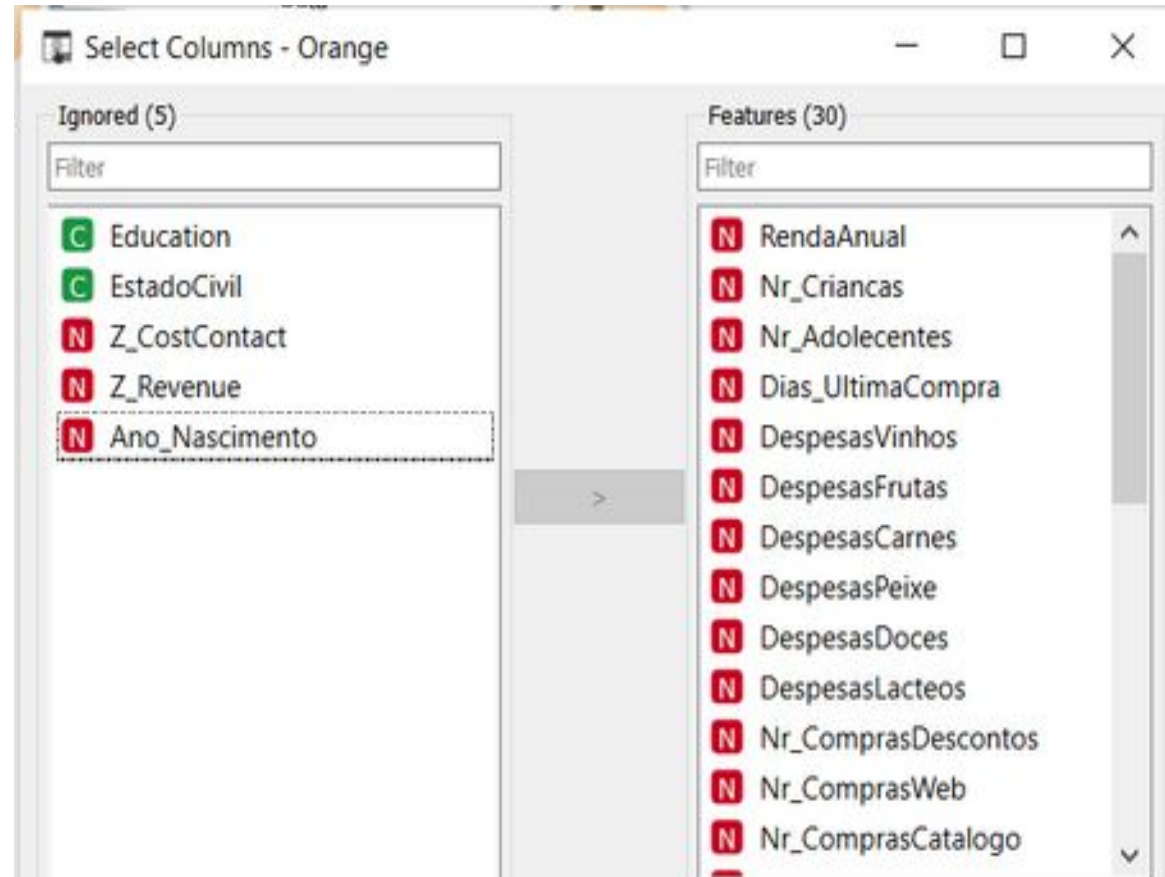
The screenshot shows the 'Feature Constructor (1) - Orange' window. It contains a list of newly created features and their formulas:

- Pai_Filhos**: 0 if (Nr_Filhos == 0) else 1
- Tamanho_Familia**: $(1 + Nr_Filhos)$ if (NovoEstadoCivil == "Sozinho") else $(2 + Nr_Filhos)$
- Pai_Filhos**: 0 if (Nr_Filhos == 0) else 1

Segmentação de Mercado

Tratamento e Limpeza de Dados.

Retiramos as variáveis: Education; EstadoCivil; Z_CostContact; Z_Revenue e Ano_Nascimento;

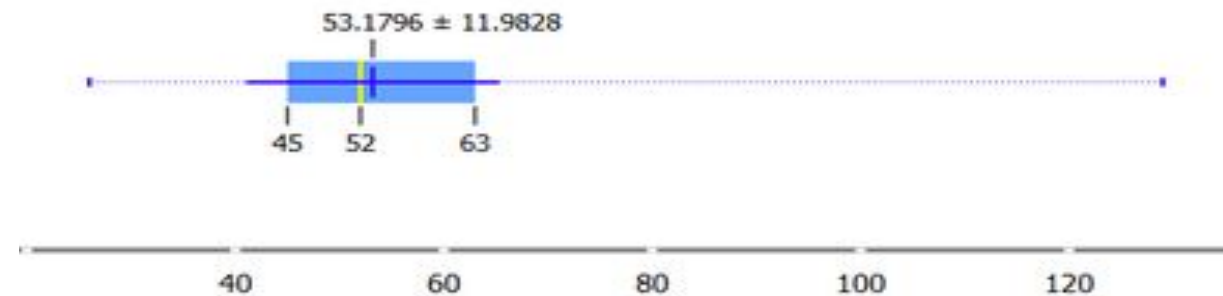
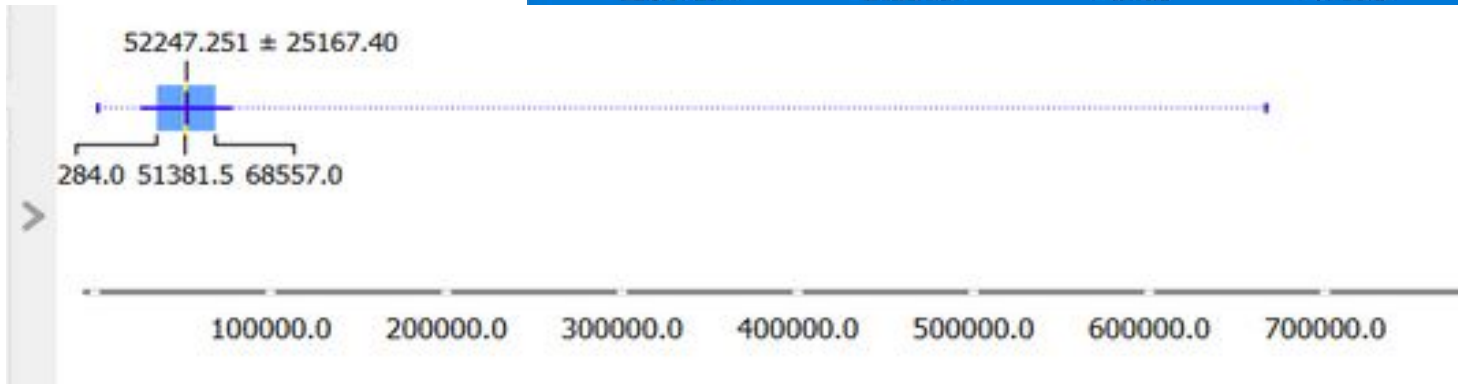


Segmentação de Mercado

Tratamento e Limpeza de Dados.

Temos outlier nas variáveis Idade e RendaAnual, retiramos essas instancias.

Mean	Median	Dispersion	Min.	Max.	Missing
53.1796	52	0.225328	26	129	0 (0 %)
52247.251	51381.5	0.482	1730.0	666666.0	0 (0 %)



Segmentação de Mercado

Tratamento e Limpeza de Dados.

Temos agora 2.212 instancias;

Select Rows - Orange

Conditions

RendaAnual	is between	0,000000	and	200.000,000000
Idade	is between	0,000000	and	99,000000

Data Table (1) - Orange

Info
2212 instances (no missing data)
30 features
No target variable.
No meta attributes

Variables
 Show variable labels (if present)
 Visualize numeric values
 Color by instance classes

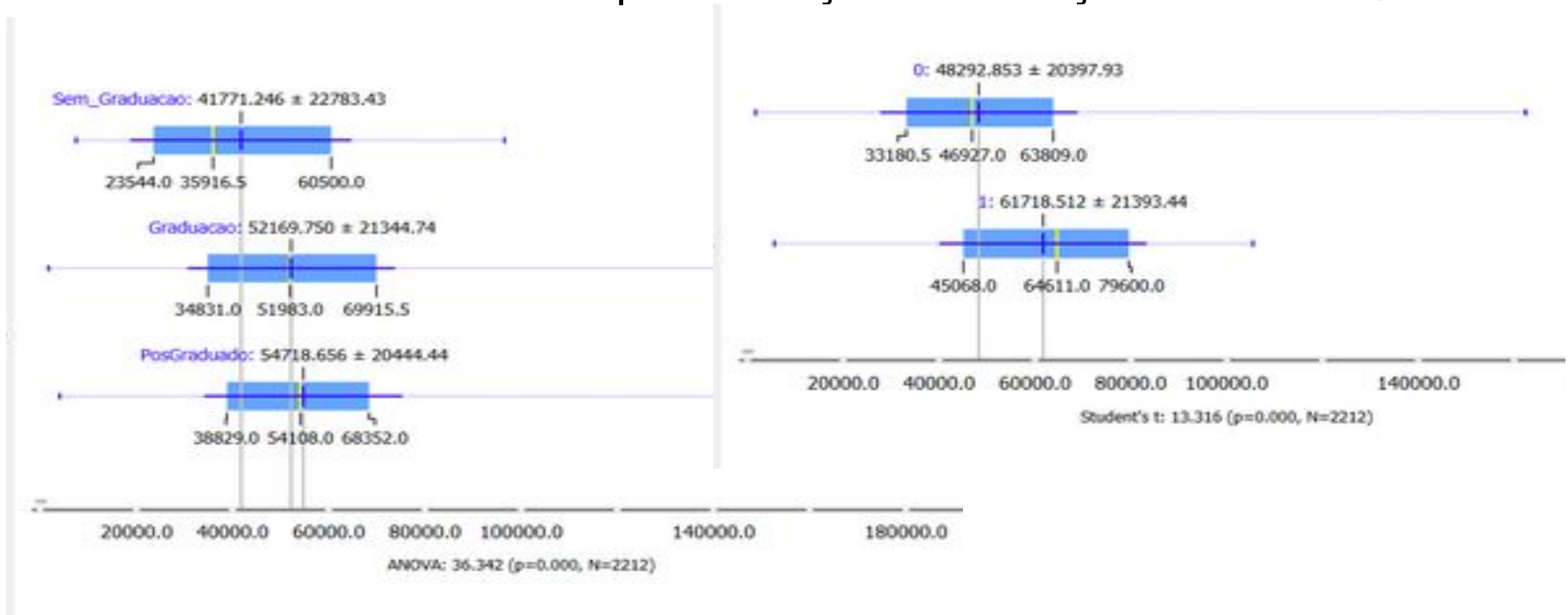
Selection
 Select full rows

	RendaAnual	Nr_Crianças	Nr_Adolecentes	Dias_UltimaCompra	DespesasVinhos
1	70951.0	0.0	0.0	66.0	239.0
2	57091.0	0.0	0.0	0.0	464.0
3	74881.0	1.0	1.0	48.0	505.0
4	7500.0	1.0	0.0	98.0	5.0
5	36957.0	1.0	1.0	43.0	100.0
6	54466.0	1.0	1.0	78.0	12.0
7	16927.0	1.0	1.0	50.0	20.0
8	12571.0	1.0	0.0	86.0	3.0
9	31163.0	1.0	0.0	54.0	2.0
10	74293.0	0.0	0.0	66.0	375.0
11	72282.0	0.0	0.0	70.0	503.0
12	59462.0	0.0	1.0	29.0	724.0
13	15315.0	0.0	0.0	27.0	7.0
14	25176.0	1.0	1.0	79.0	4.0
15	13724.0	1.0	0.0	43.0	2.0

Segmentação de Mercado

Análises exploratória.

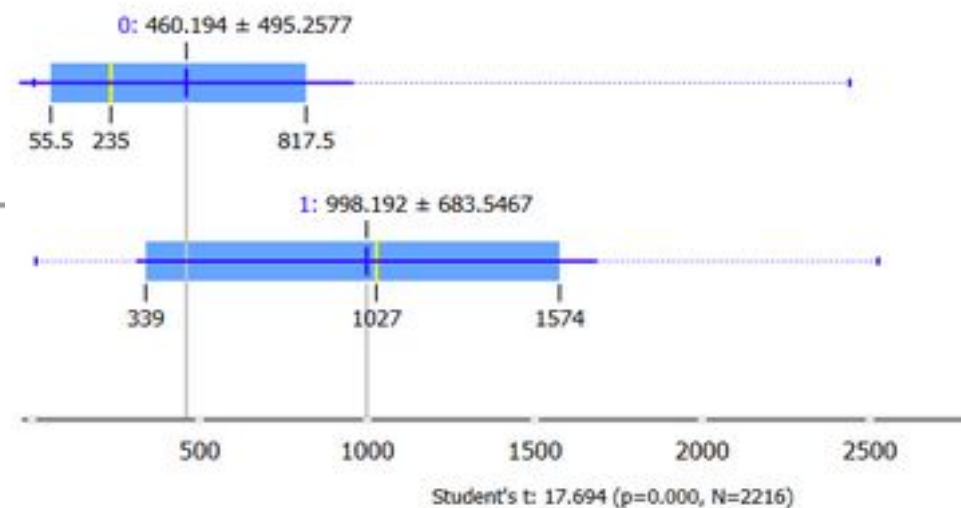
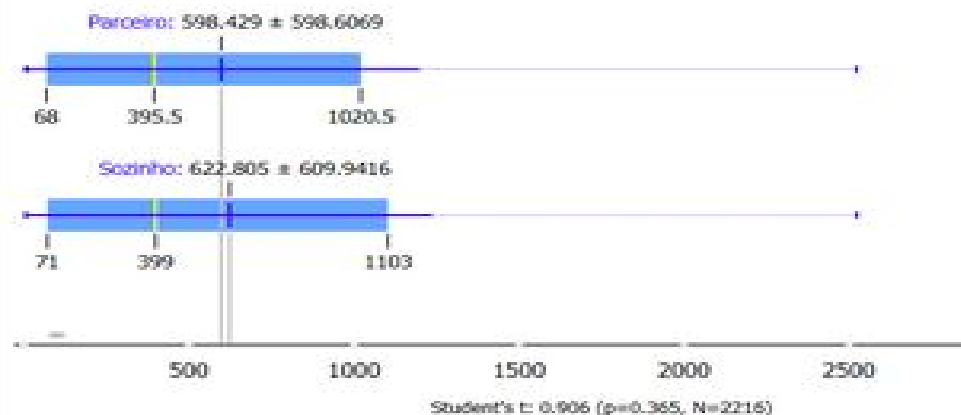
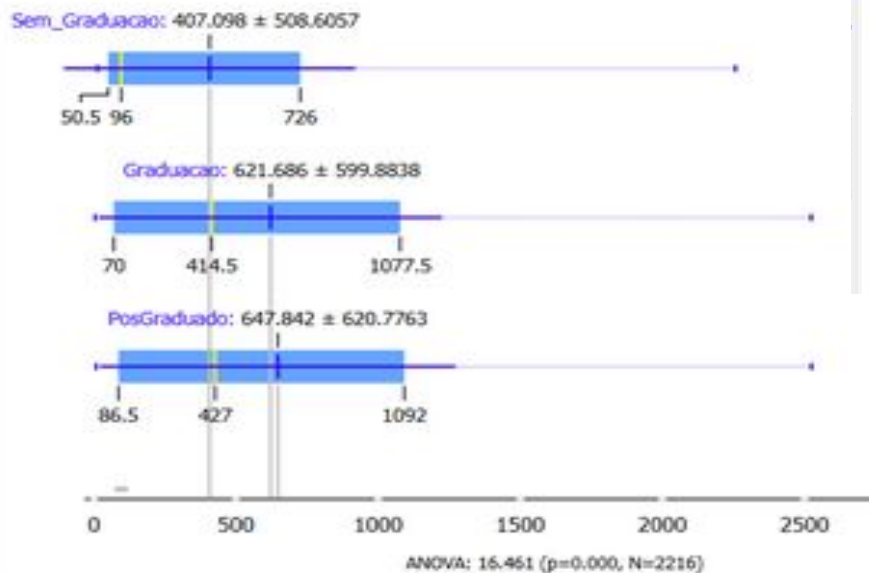
Observar os valores de renda por titulação e aceitação de ofertas;



Segmentação de Mercado

Análises exploratória.

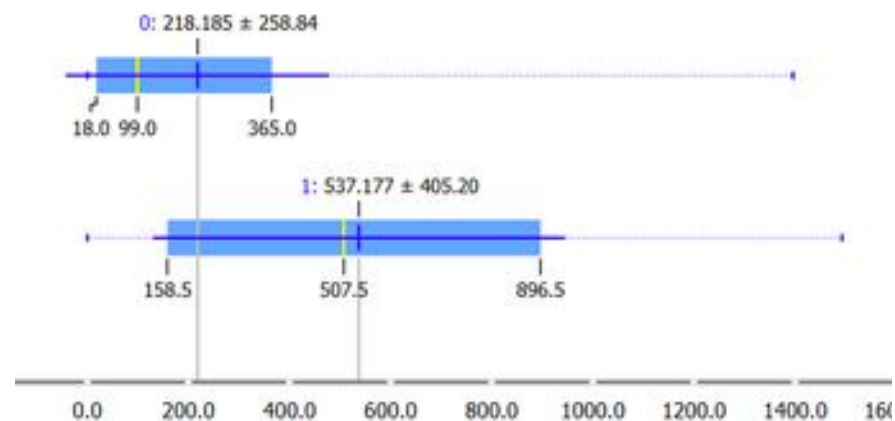
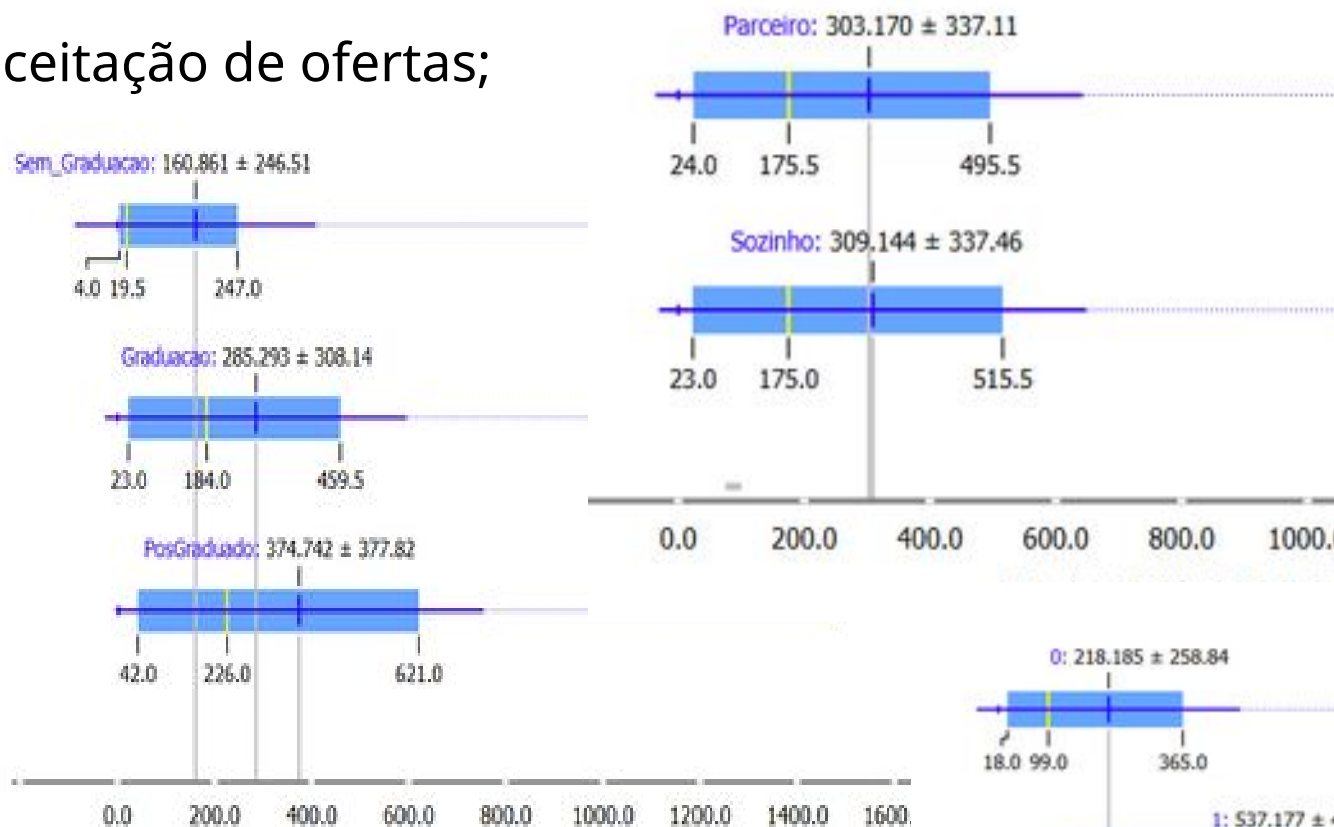
Observar a mediana de gastos por titulação, estado civil e aceitação de ofertas;



Segmentação de Mercado

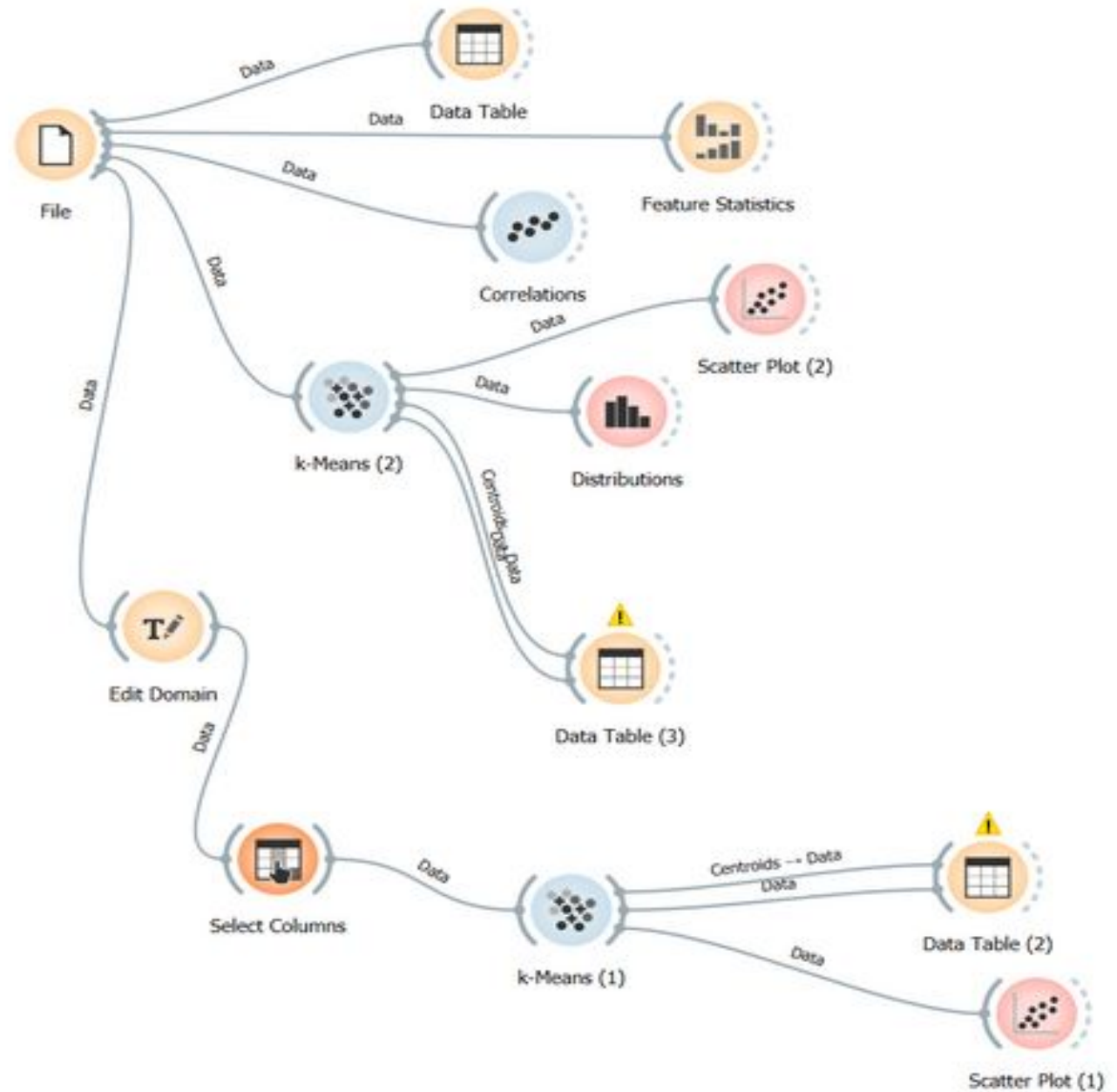
Análises exploratória.

Observar a mediana de gastos em vinhos por titulação, estado civil e aceitação de ofertas;



Segmentação de Mercado

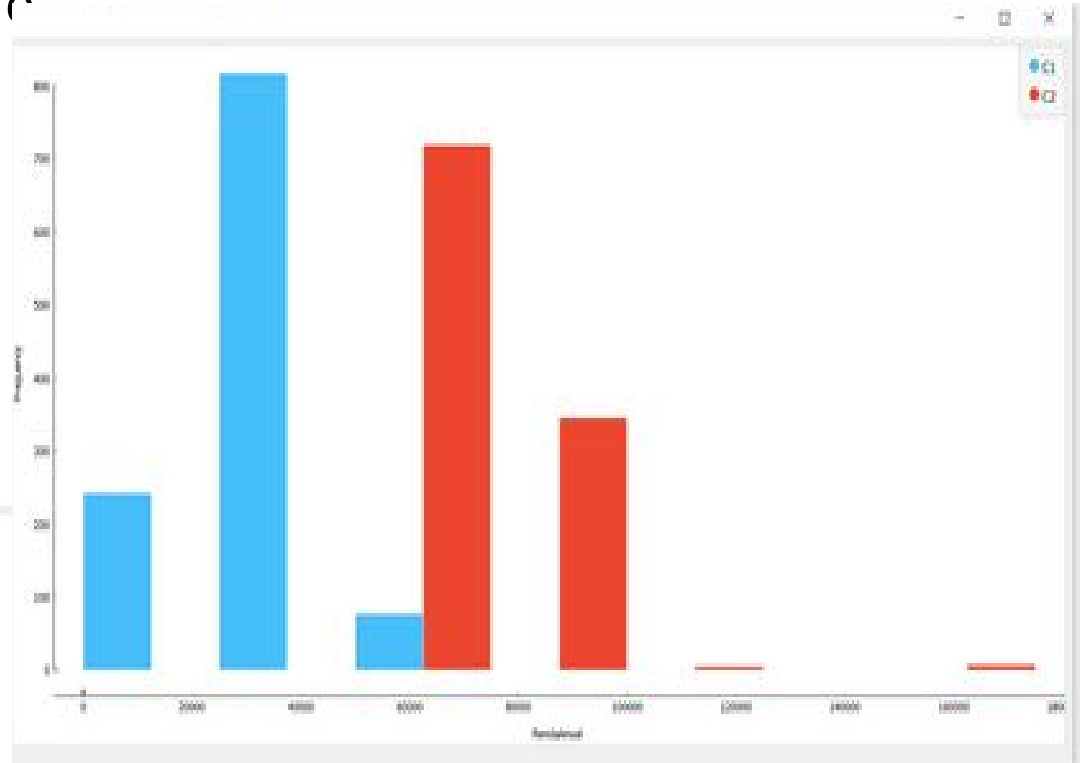
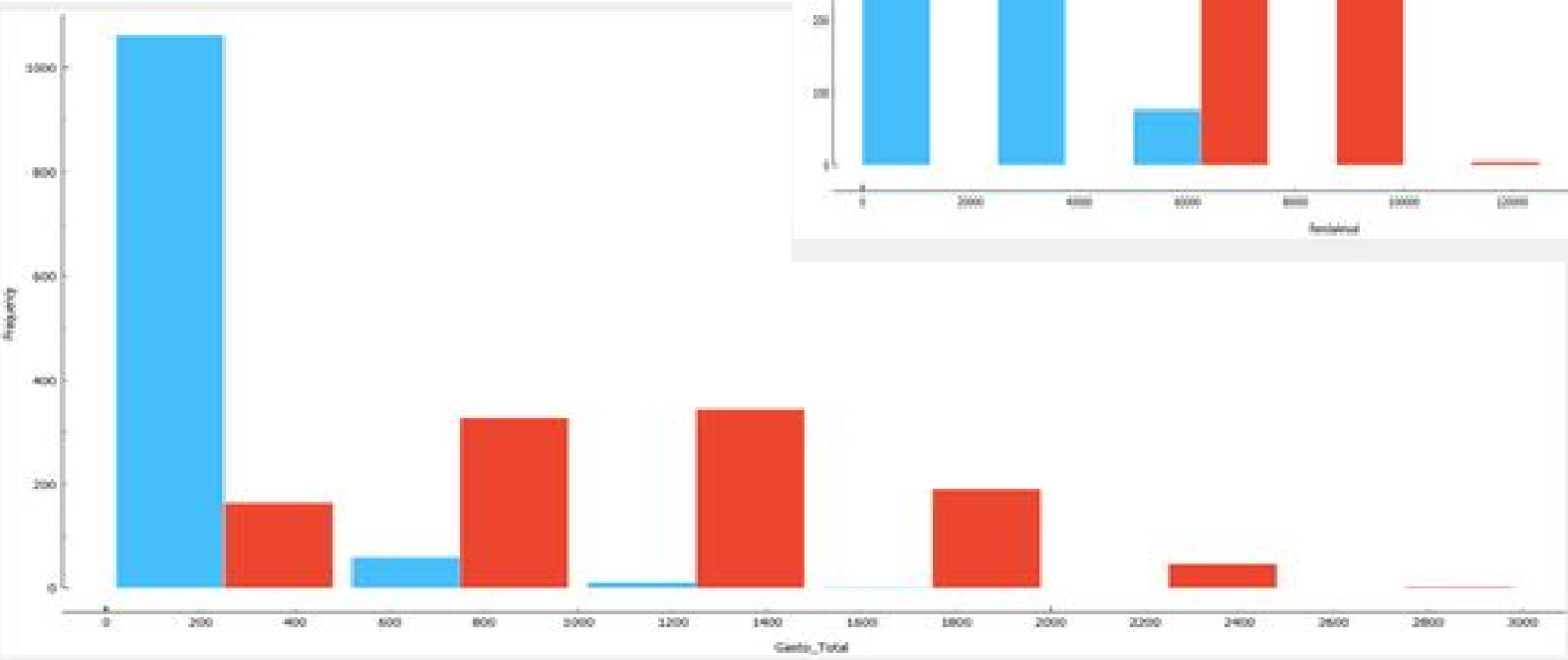
Agrupamentos



Segmentação de Mercado

Agrupamentos

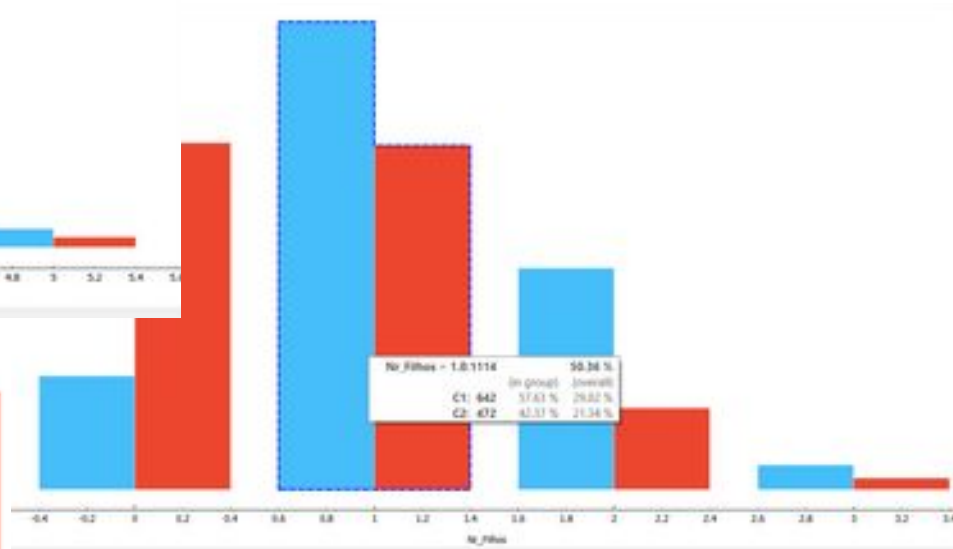
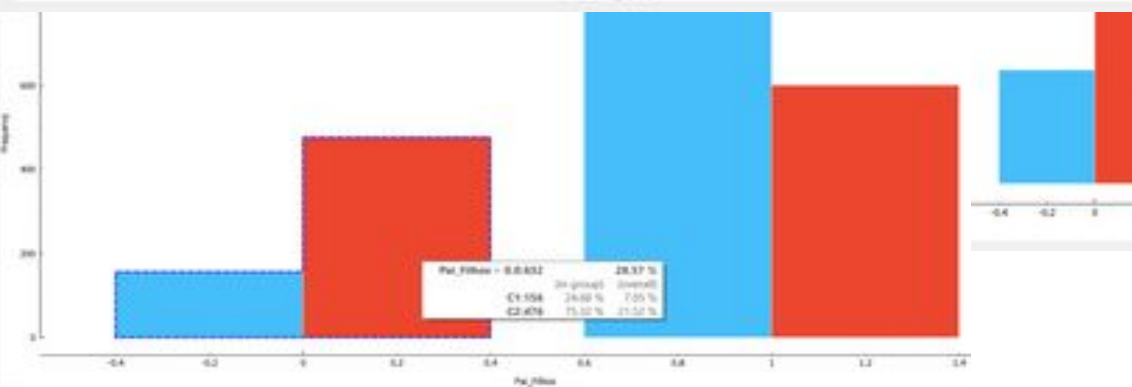
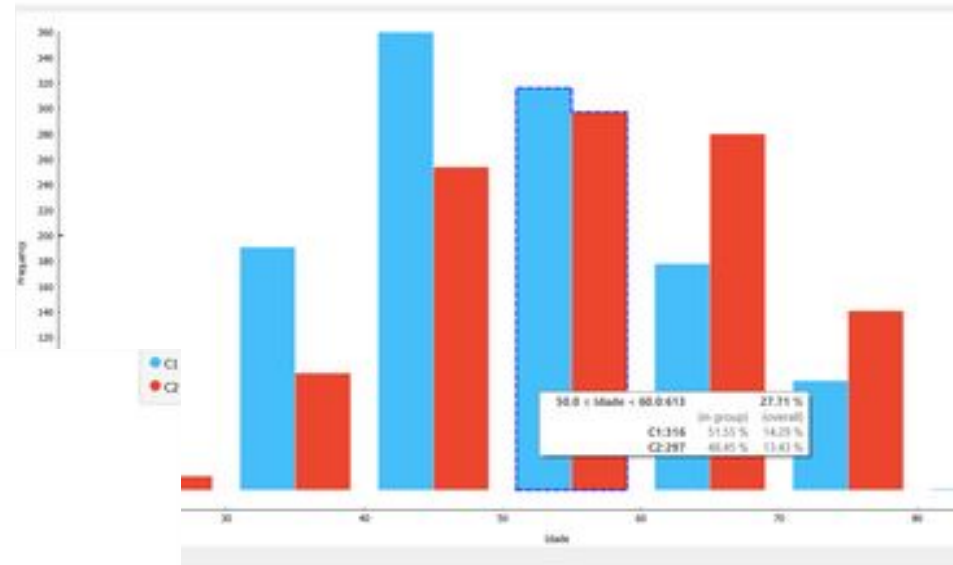
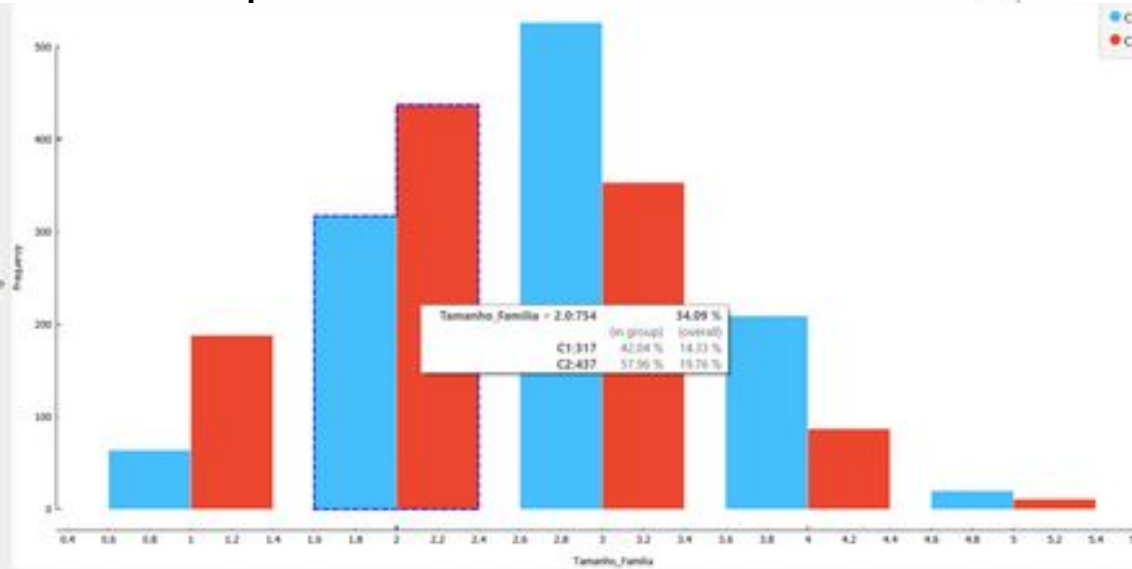
Divisão dos grupos por Renda e
Gastos;



Segmentação de Mercado

Agrupamentos

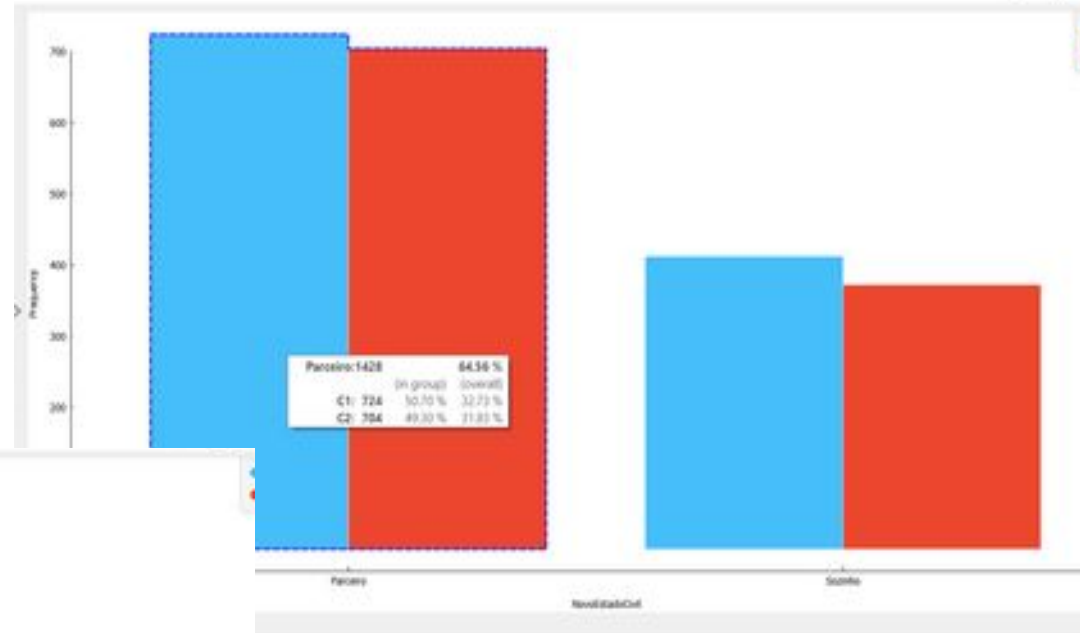
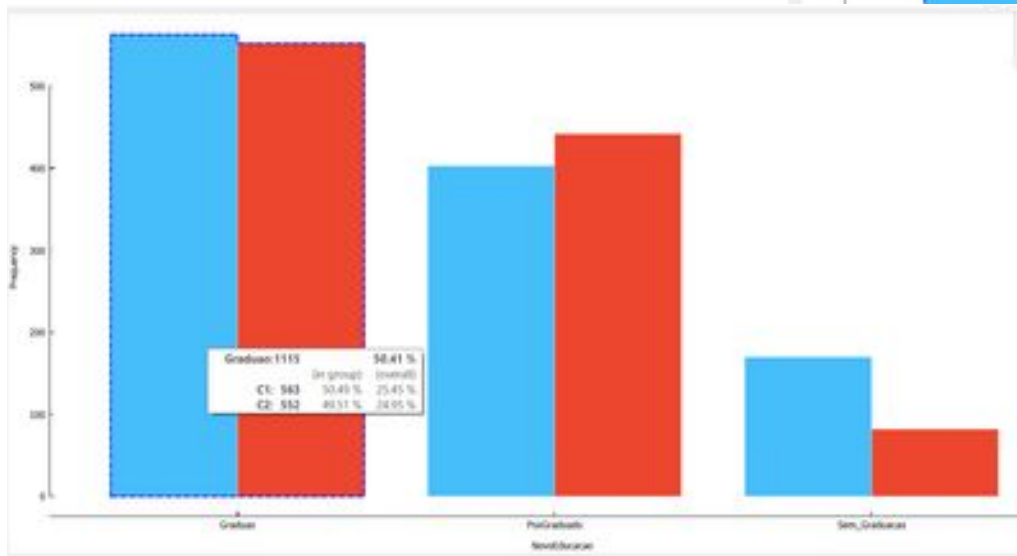
Divisão dos grupos por Idade e Tamanho da família, número de Filhos e quem tem filhos;



Segmentação de Mercado

Agrupamentos

Divisão dos grupos por estado civil e nível educacional;



Segmentação de Mercado

Resumo dos clusters:

Clusters 1:

Renda abaixo de 60 mil

Gastos abaixo de 500

Predominância de pai com filhos

Clientes de 60 anos ou menos

Clusters 2:

Renda acima de 60 mil

Gastos acima de 500

Predominância de pai sem filhos

Clientes acima de 60 anos

Aprendizado de Máquina

Prof. Aran Morales, UNISUL – Anima Educação

Apostila 9: Agrupamentos