

DATA MINING



FUNDAÇÃO INSTITUTO
DE ADMINISTRAÇÃO



LÁBDATA

Disciplina: Estatística 1

Tema da Aula: Análise de Cesto de Compras

Coordenação:

Prof. Dr. Adolpho Walter
Pimazzi Canton

Profa. Dra. Alessandra de
Ávila Montini

Profa. Daniel Martins

Sobre Daniel Martins

Breve Descrição Profissional

Frequente atuação em projetos analíticos para empresas dos seguimentos: financeiro, varejo e Telecomunicação. Implementação de soluções de Risco (Banco do Brasil, Bradesco e Itaú), Prevenção a Fraude (SEFAZ e Santander), Forecasting (Prismah e Natura), Retenção de clientes (Claro), Big Data Analytics(Petrobrás). Além de atuar, eventualmente, como pré-vendas realizando apresentações, Workshops e PoC's para clientes de diversos segmentos de mercado.



Educação

- Graduação em Estatística (USP)
- Professor - Data Mining (FIA)
- Pós G. em Economia (FGV)
- Mestrando(USP)

Data Scientist Business Development

Objetivo

A análise de Cesto de Compras ajuda a detectar associação entre os dados.

Esta técnica é muito utilizada para detectar Padrão em bancos de dados grandes.

Muito utilizada para recomendar ações.

Análise de Associação

Exemplo 1

Análise de Associação

Um cliente que vai ao mercado e coloca em seu carrinho,



então com grande chance ele vai colocar também

Tomada de decisão

- Colocar o carvão próximo à picanha.
- Colocar o carvão longe da picanha.
- Vender o carvão e a picanha de forma casada.
- Vender o carvão, a picanha e mais alguns produtos de forma casada.

Exemplo 2

Análise de Associação

Um cliente de telecom que

Reduz o uso de voz

Reduz o uso de dados

Então com grande chance ele vai



Tomada de decisão

- Verificar se o cliente efetuou alguma reclamação
- Oferecer um novo plano ao cliente
- Oferecer um combo ao cliente
- Oferecer um pacote de produtos e serviços a família

Exemplo 3



Análise de Associação

Um funcionário que

Chega atraso

Falta sempre

Então com grande chance ele vai



Tomada de decisão



Exemplo 4

Análise de Associação

Após a análise dos dados pode-se concluir que, para um dado segmento, um cliente que possui

CDB

Previdência

Renda Fixa



Então com grande chance ele vai adquirir um seguro.

Análise de Associação

Após a análise dos dados pode-se concluir que, para um dado segmento, um cliente que possui

Empréstimo pessoal

Usa o rotativo do cartão

Então com grande chance ele vai adquirir um automóvel financiado.



Exemplo 5

Análise de Associação

Pode-se verificar que um cliente que adquiri no site
Geladeira
Micro-ondas
com grande chance ele vai adquirir um fogão.

Após visitar este item, o que os clientes acabam comprando?

65% compram



Fogão Consul 4 Bocas Facilité
CF450BB

De: R\$ 699,90

Por: R\$ 479,90

ou 12X de R\$ 39,99
sem juros

13% compram



Fogão Consul 4 Bocas Erva Doce
CF350BB

De: R\$ 599,90

Por: R\$ 449,90

ou 12X de R\$ 37,49
sem juros

12% compram



Fogão Consul 4 Bocas Facilité
CF550BB c/ Timer

De: R\$ 799,90

Por: R\$ 549,90

ou 12X de R\$ 45,82
sem juros

5% compram



Fogão Consul 4 Bocas Facilité
CF550BR c/ Timer - Inox

De: R\$ 799,90

Por: R\$ 699,90

ou 12X de R\$ 58,32
sem juros

Exemplo 6

Análise de Associação

Pode-se verificar que um paciente que

É fumante

Possui colesterol alto

Não faz exercícios físicos

com grande chance ele pode ter um infarto.



Exemplo 7

A Netflix usa algoritmos para recomendar filmes aos clientes.



NETFLIX

Entrar

Assista a filmes e séries de TV quando quiser, onde quiser.

Comece o seu mês grátis

The image is a screenshot of a Netflix advertisement. It features a family of five (a man, a woman, and three children) sitting on a light-colored sofa in a modern living room, watching a large television. The TV screen displays the Netflix interface with a grid of movie and TV show thumbnails. A blue bicycle is leaning against the wall behind the sofa. In the foreground, a low wooden coffee table holds a bowl of popcorn, a smartphone, and some papers. The overall atmosphere is cozy and relaxed. The text 'Assista a filmes e séries de TV quando quiser, onde quiser.' is overlaid on the left side of the image. A red banner at the top left contains the 'NETFLIX' logo, and a red button at the top right says 'Entrar'. A blue button at the bottom left says 'Comece o seu mês grátis'.

Exemplo 8



Oferecer produtos com base no rastro digital



Medidas de Associação

Medidas de Associação:

- Support
- Confidence
- Expeted Confidence
- Lift

Estudo de Caso



Neste **estudo de Caso** será apresentado o cálculo das medidas de associação para compras na Loja Renner.

LOJAS RENNER S.A.					
AV. ASSIS BRASIL, 1010 - PORTO ALEGRE - RS					
CNPJ:92.754.738/0013-04			IE:096/0372202		
03/05/2013 10:05:40		CCF:002260		C00:010169	
CUPOM FISCAL					
ITEM	CODIGO	DESCRIÇÃO			
QTD.	UN.	VL.UNIT(R\$)	ST	IAT	VL.ITEM(R\$)
001	530236310	Blusa n3 4			
	1,000Un x	89,90	T04	A	89,90%
002	520196220	PERFUME PA			
	1,000Un x	199,00	F1	A	199,00%
TOTAL		R\$		288,90	
Dinheiro				288,90	
T04=17,00%					

Deve-se verificar
quais itens foram
comprados em cada
Cupom Fiscal.

Para exemplificar o cálculo das medidas serão considerados 5 cupons em que os clientes comprem meias.



A



C



B



D



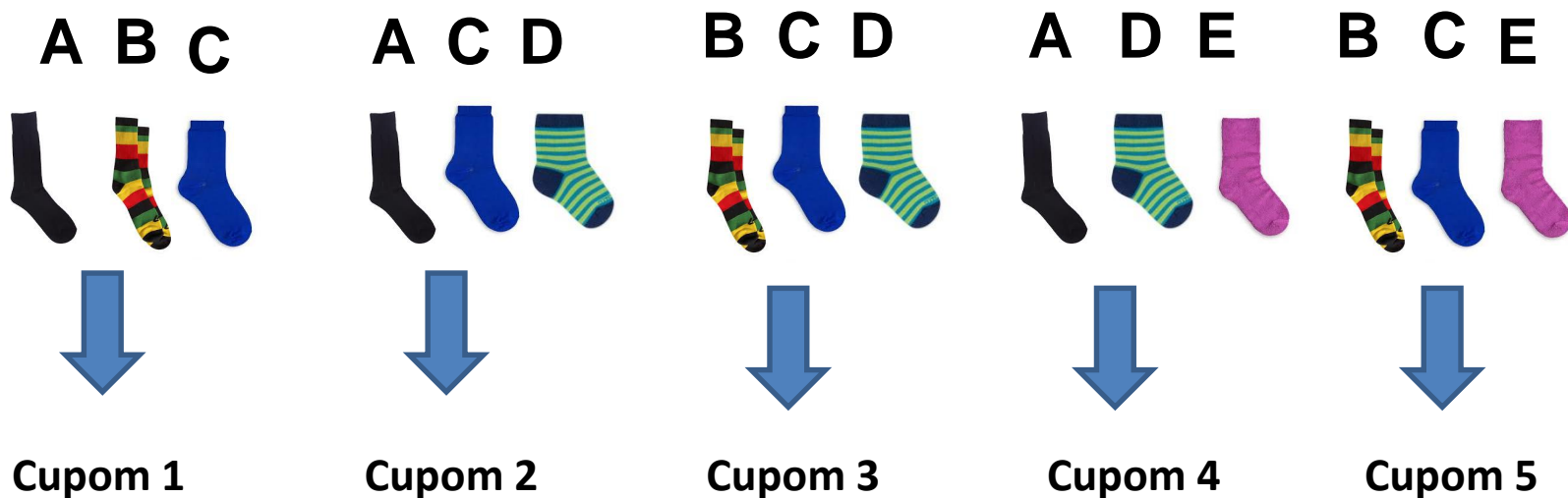
E



Support



Support: Mede a frequência que uma combinação ocorre considerando-se **todos os cupons fiscais**.



As meias A e C foram vendidas de forma conjunta em quantos cupons ?



As meias A e C foram vendidas de forma conjunta em 2 cupons.



Desta forma considerando os 5 cupons a combinação A e C foram vendidas em 2/5 das compras.

As meias A, D e E foram vendidas de forma conjunta em 1 cupom.



Desta forma considerando os 5 cupons a combinação A, D e E foram vendidas em $1/5$ das compras.

Calculo de algumas frequências.

Produto(S)	Produto(S)	Frequência
A	C	2 / 5
A	D	2 / 5
A D	E	1 / 5

Regra

A → C

A → D

A & D → E

Support

2/5

2/5

1/5

Confidence



Confidence: Mede a frequência de vezes em que os 2 itens foram comprados juntos, em relação as vezes que o item da esquerda foi comprado.



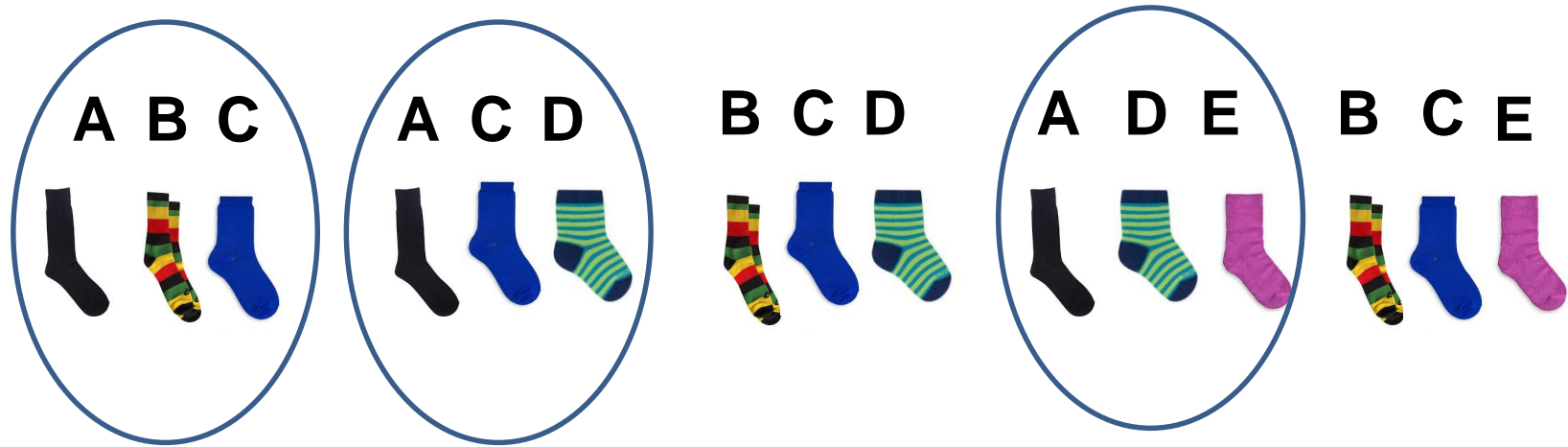
Obtenção do Confidence para a Regra: A \rightarrow D

Número de vezes em que os 2 itens foram comprados (A e D)
juntos = 2.



Obtenção do Confidence para a Regra: A D

Número de vezes em que o item da esquerda foi comprado (A) = 3



Confidence: Mede a frequência de vezes em que os 2 itens foram comprados juntos, em relação as vezes que o item da esquerda foi comprado.

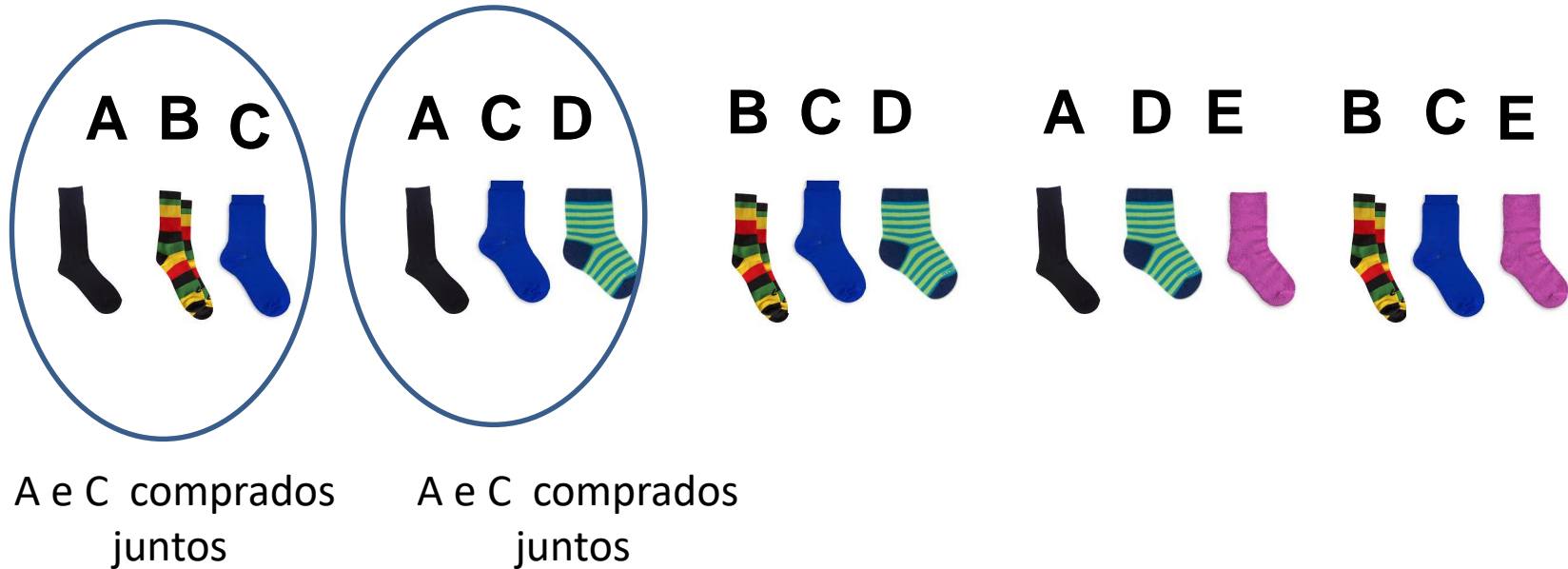


Obtenção do Confidence para a Regra: A → D

$$\text{Confidence} = 2 / 3$$

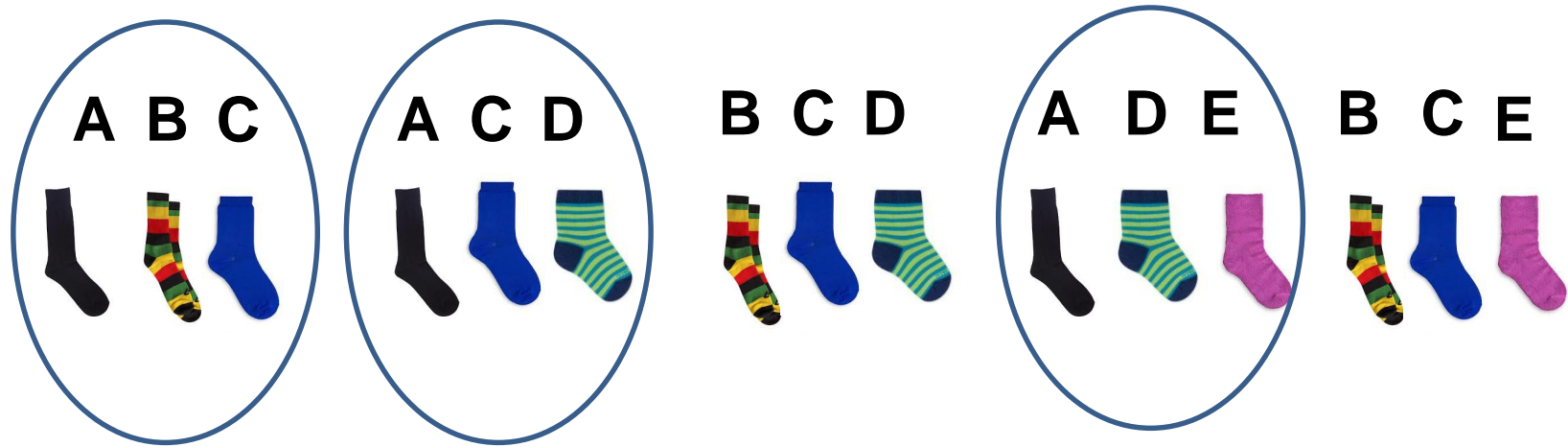
Obtenção do Confidence para a Regra: A \rightarrow C

Número de vezes em que os 2 itens foram comprados (A e C) juntos = 2.



Obtenção do Confidence para a Regra: A \rightarrow C

Número de vezes em que o item da esquerda foi comprado (A) = 3



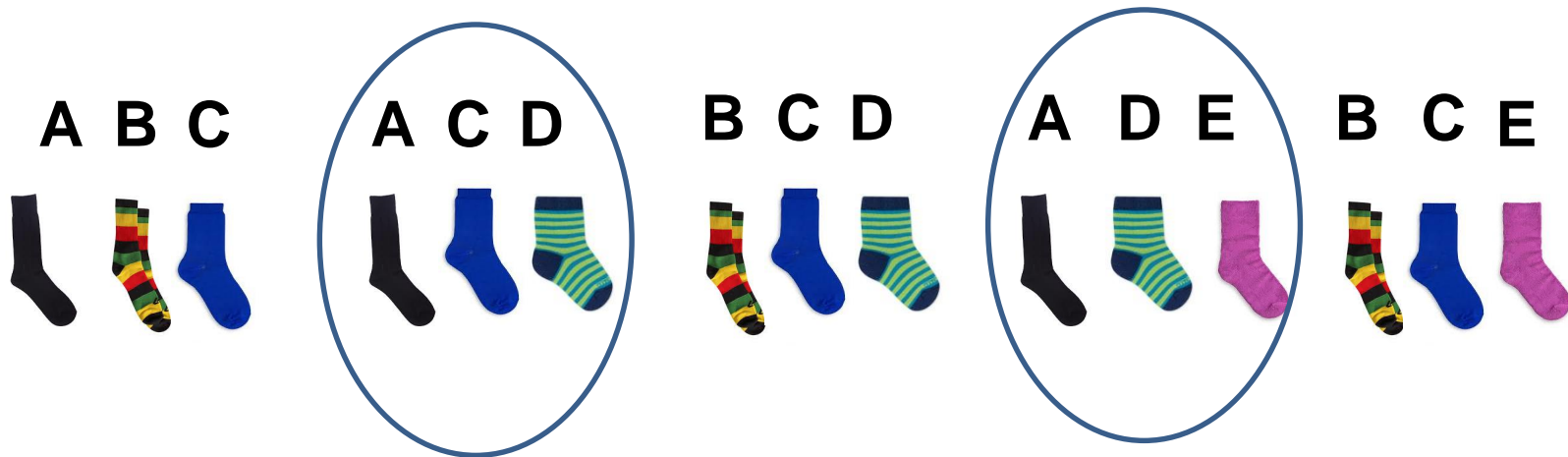
Obtenção do Confidence para a Regra: A & D → E

Número de vezes em que os itens foram comprados (A , D e E)
juntos = 1.



Obtenção do Confidence para a Regra: A & D → E

Número de vezes em que os itens da esquerda foram comprados (A & D) = 2.



Confidence: Mede a frequência de vezes em que os 2 itens foram comprados, em relação as vezes que o item da esquerda foi comprado.



Regra

A → C

A → D

A & D → E

Confidence

$2/3$

$2/3$

$1/2$

Expected Confidence

Expected Confidence: Mede a frequência de vezes em que o item da direita é comprado em relação ao total.



Regra

E. Confidence

A	→	C	4/5
A	→	D	3/5
A & D	→	E	2/5

Lift

Lift: quociente entre o **Confidence** e o **Expected Confidence**



<u>Regra</u>	<u>Confidence</u>	<u>E. Confidence</u>	<u>Lift</u>
A → C	$2/3 = 0.666$	$4/5 = 0.800$	$0.666 / 0.800 = 0.8325$
A → D	$2/3 = 0.666$	$3/5 = 0.600$	$0.666 / 0.6000 = 1.11$
A & D → E	$1/2 = 0.500$	$2/5 = 0.400$	$0.500 / 0.400 = 1.25$

Regra

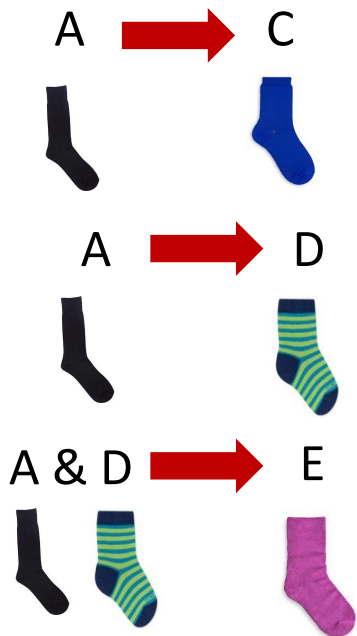
Confidence

Probabilidade da compra com relação ao número de clientes que compraram o **conjunto da esquerda**

E. Confidence

Probabilidade do item da direita com relação ao **total de compras**

Lift



0.666

0.800

0.8325

0.666

0.600

1.11

0.500

0.400

1.25

Regra

Confidence

Probabilidade da compra **dos produtos** com relação ao número de clientes que compraram o **conjunto da esquerda**

E. Confidence

Probabilidade do item da direita com relação ao **total de compras**

Lift

A → C	0.666	0.800	0.8325
A → D	0.666	0.600	1.11
A & D → E	0.500	0.400	1.25

Quando o lift < 1 existe aversão entre os produtos.

Quando o lift > 1 existe atração entre os produtos.

Pensando na Regra **A**  **C**

$$\text{Suporte} = \frac{\text{número de vezes que ocorre A e D}}{\text{total de cupons}}$$

$$\text{Expected Confidence} = \frac{\text{número de vezes que ocorre D}}{\text{total de cupons}}$$

$$\text{Confidence} = \frac{\text{número de vezes que ocorre A e D}}{\text{número de vezes que ocorre A}}$$

$$\text{Lift} = \frac{\text{Confidence}}{\text{Expected Confidence}}$$

Exercício

Obtenha o Confidence, Expected Confidence e Lift para as regras:



Regra

Confidence

Probabilidade da compra **dos produtos**
com relação ao número de clientes que
compraram o **conjunto da esquerda**

E. Confidence

Probabilidade do item **da**
direita com relação ao
total de compras

Lift

B → C

C → D

B & D → C

Exemplo

Exemplo

Distribuição de clientes de uma determinada agência que possuem CDB e/ou fundo de renda fixa.

		B = CDB		
		Não	Sim	Total
A = Fundo de Renda Fixa	Não	500	3500	4000
	Sim	2000	4000	6000
	Total	2500	7500	10000

Qual a proporção de clientes que possuem CDB ?

		B = CDB		Total
		Não	Sim	
A = Fundo de Renda Fixa	Não	500	3500	4000
	Sim	2000	4000	6000
	Total	2500	7500	10000

A proporção de clientes que possuem CDB é $\frac{7500}{10000} = 0,75$

Expected Confidence = 0,75

Qual a proporção de clientes que possuem CDB dentro dos clientes que possuem Fundo de Renda Fixa?

		B = CDB		Total
		Não	Sim	
A = Fundo de Renda Fixa	Não	500	3500	4000
	Sim	2000	4000	6000
	Total	2500	7500	10000

A proporção de clientes é $\frac{4000}{6000} = 0,6667$

Confidence = 0,6667

Expected Confidence

A proporção de clientes que possuem CDB é 0,75.

Confidence

A proporção de clientes que possuem CDB dentro dos clientes que possuem Fundo de Renda Fixa é 0,6667.

Neste caso existe aversão entre os produtos pois $0,6667 < 0,75$.

O lift é dado por: $\frac{\text{Confidence}}{\text{Expected Confidence}}$

O lift é dado por: $\frac{0.6667}{0.7500} = 0.8889$

Como o lift < 1 existe aversão entre os produtos.

Exemplo

Exemplo

Distribuição de clientes de uma determinada agência que possuem CDB e/ou fundo de renda fixa.

		B = CDB		
		Não	Sim	Total
A = Fundo de Renda Fixa	Não	3500	500	4000
	Sim	50	5950	6000
	Total	3550	6450	10000

Qual a proporção de clientes que possuem CDB ?

		B = CDB		
		Não	Sim	Total
A = Fundo de Renda Fixa	Não	3500	500	4000
	Sim	50	5950	6000
	Total	3550	6450	10000

A proporção de clientes que possuem CDB é $\frac{6450}{10000} = 0,6450$

Expected Confidence = 0,6450

Qual a proporção de clientes que possuem CDB dentro dos clientes que possuem Fundo de Renda Fixa?

		B = CDB		
		Não	Sim	Total
A = Fundo de Renda Fixa	Não	3500	500	4000
	Sim	50	5950	6000
	Total	3550	6450	10000

A proporção de clientes é $\frac{5950}{6000} = 0,9917$

Confidence = 0,9917

Expected Confidence

A proporção de clientes que possuem CDB é 0,9917.

Confidence

A proporção de clientes que possuem CDB dentro dos clientes que possuem Fundo de Renda Fixa é 0,6450.

Neste caso existe atração entre os produtos pois $0,9917 > 0,6450$.



O lift é dado por: $\frac{\text{Confidence}}{\text{Expected Confidence}}$

O lift é dado por: $\frac{0.9917}{0.6450} = 1,5375$

Como o lift > 1 existe atração entre os produtos.

Exercícios



1 - Distribuição de clientes que compram blusa e/ou perfume.

		B = BLUSA		Total
		Não	Sim	
A = PERFUME	Não	100	5000	5100
	Sim	2000	4000	6000
	Total	2100	9000	11100

Obtenha:

Expected Confidence

Confidence

Lift

2 - Distribuição de clientes que compram blusa e/ou perfume.

		B = BLUSA		
		Não	Sim	Total
A = PERFUME	Não	4000	1500	5500
	Sim	1200	6000	7200
	Total	5200	7500	12700

Obtenha:

Expected Confidence

Confidence

Lift

Aplicação no SAS MINER

Base de Dados

A base de dados é formada por transações de compras considerando 20 cupons. Nestas transações foram adquiridos 12 produtos diferentes, podendo conter mais de um produto por transação.

Tabela: Rener1.xls

A tabela apresenta os produtos adquiridos em cada cupom.

Cupon	Transacao
1	MEIA
1	TENIS
1	PERFUME
2	CALCA
2	SAPATO
2	PERFUME
3	TENIS
3	MEIA
4	MEIA
4	BOLSA
4	CAMISA
4	GRAVATA
5	PERFUME
5	BLUSA
5	CALCA
5	MEIA
5	TENIS

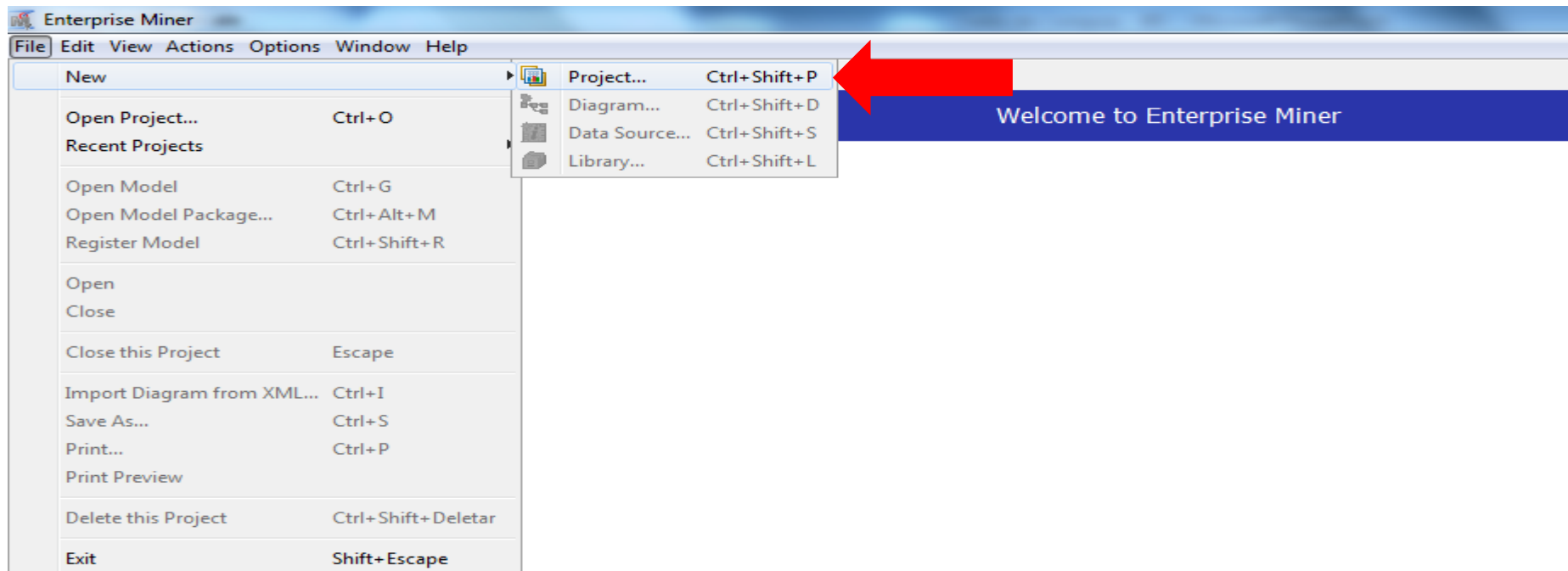
A tabela apresenta o número de produtos adquiridos nas 20 compras.

MEIA	11
GRAVATA	9
TENIS	8
CALCA	5
PERFUME	5
BATON	4
CAMISA	4
BOLSA	3
BASE	2
SOMBRA	2
BLUSA	1
SAPATO	1

SAS MINER



Criação de um projeto .



Definição do nome do projeto e do diretório de trabalho.

Create New Project -- Etapa 3 de 4 Register the Project

SAS®
Enterprise
Miner™ 6.2

Specify a project name and directory on the SAS Server for this project. All SAS data sets and files will be written to this location.

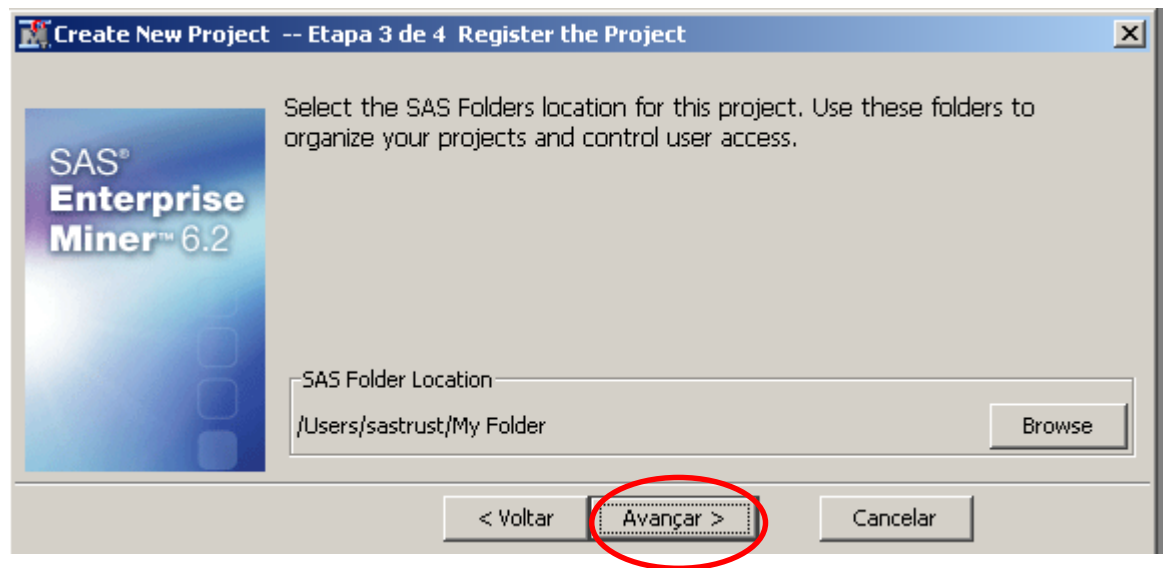
Project Name
Aula_Cesto

SAS Server Directory
C:\

< Voltar Avançar > Cancelar

Insira um nome para o projeto

Escolha um diretório de trabalho onde o Miner irá salvar os dados e arquivos.



Criação de um projeto .

Create New Project -- Etapa 4 de 4 New Project Information

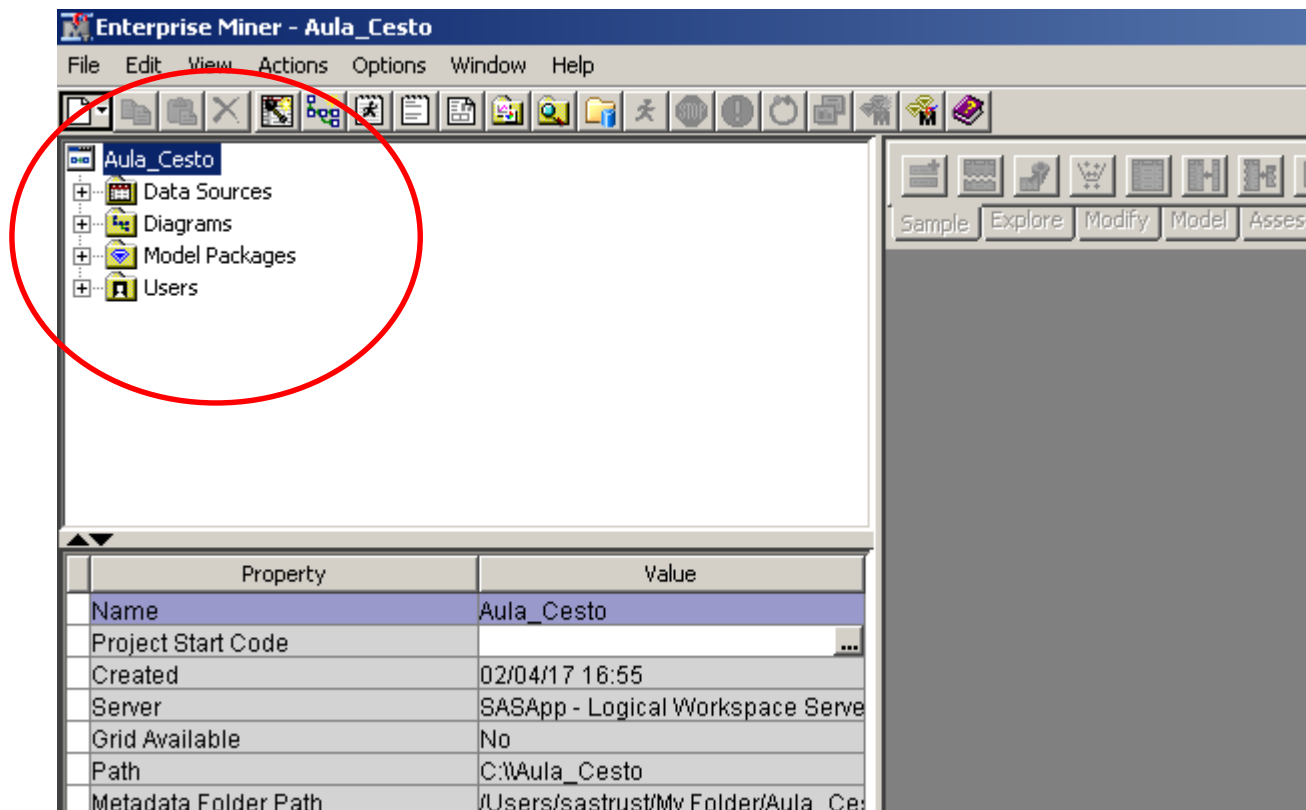
SAS® Enterprise Miner™ 6.2

New Project Information

Name	Aula_Cesto
SAS Metadata location	/Users/sastrust/My Folder
SAS Application server	SASApp - Logical Workspace Server
Server Directory	C:\

< Voltar **Concluir** Cancelar

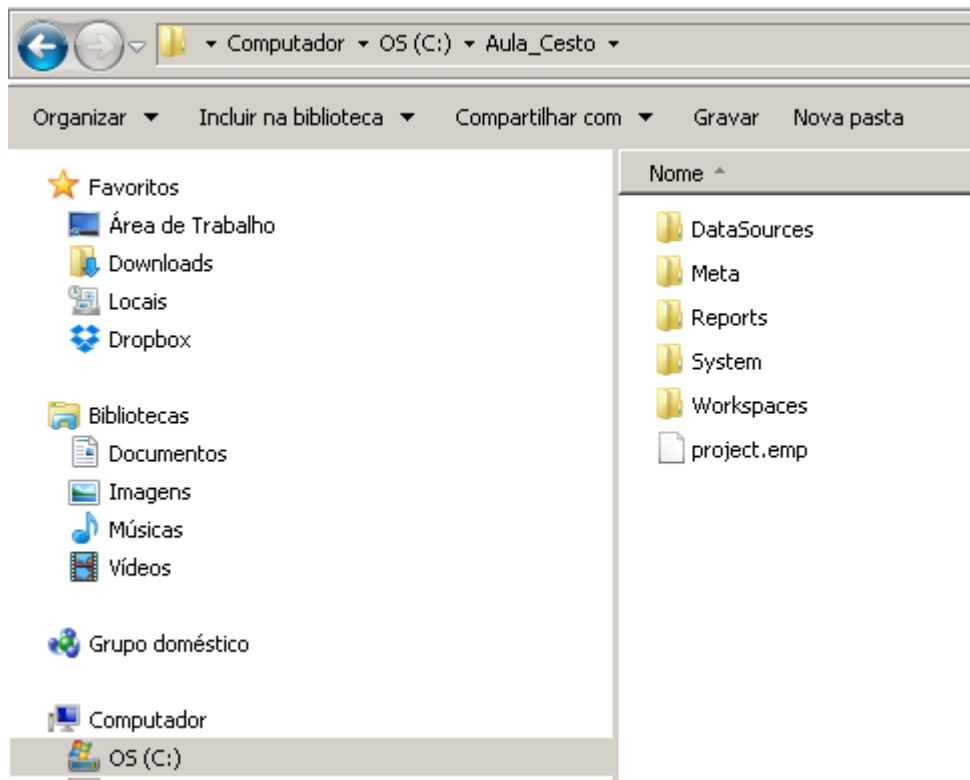
Criação de um projeto .



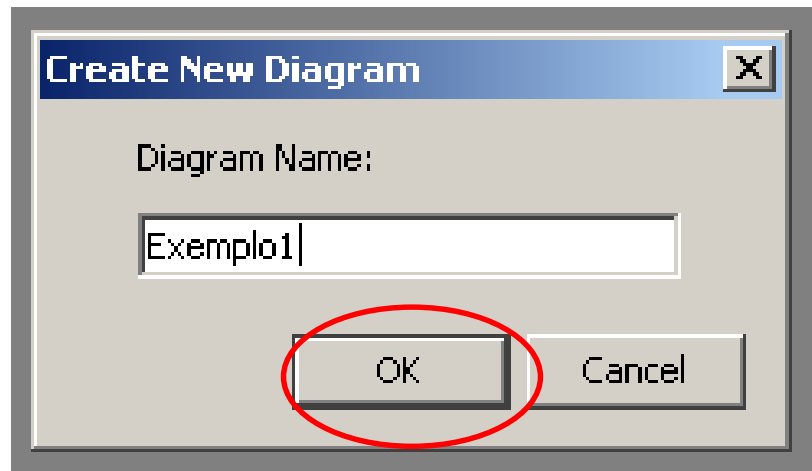
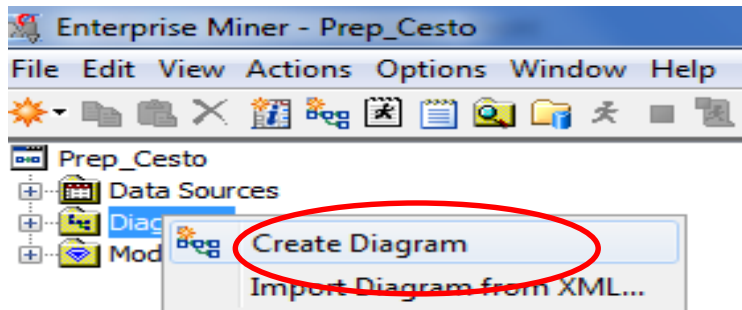
The screenshot shows the Enterprise Miner software interface. The title bar reads "Enterprise Miner - Aula_Cesto". The menu bar includes File, Edit, View, Actions, Options, Window, and Help. The toolbar contains various icons for file operations and project management. The left pane displays a tree view with the following items: Data Sources, Diagrams, Model Packages, and Users. The "Aula_Cesto" project is selected and highlighted. The right pane displays a table of project properties.

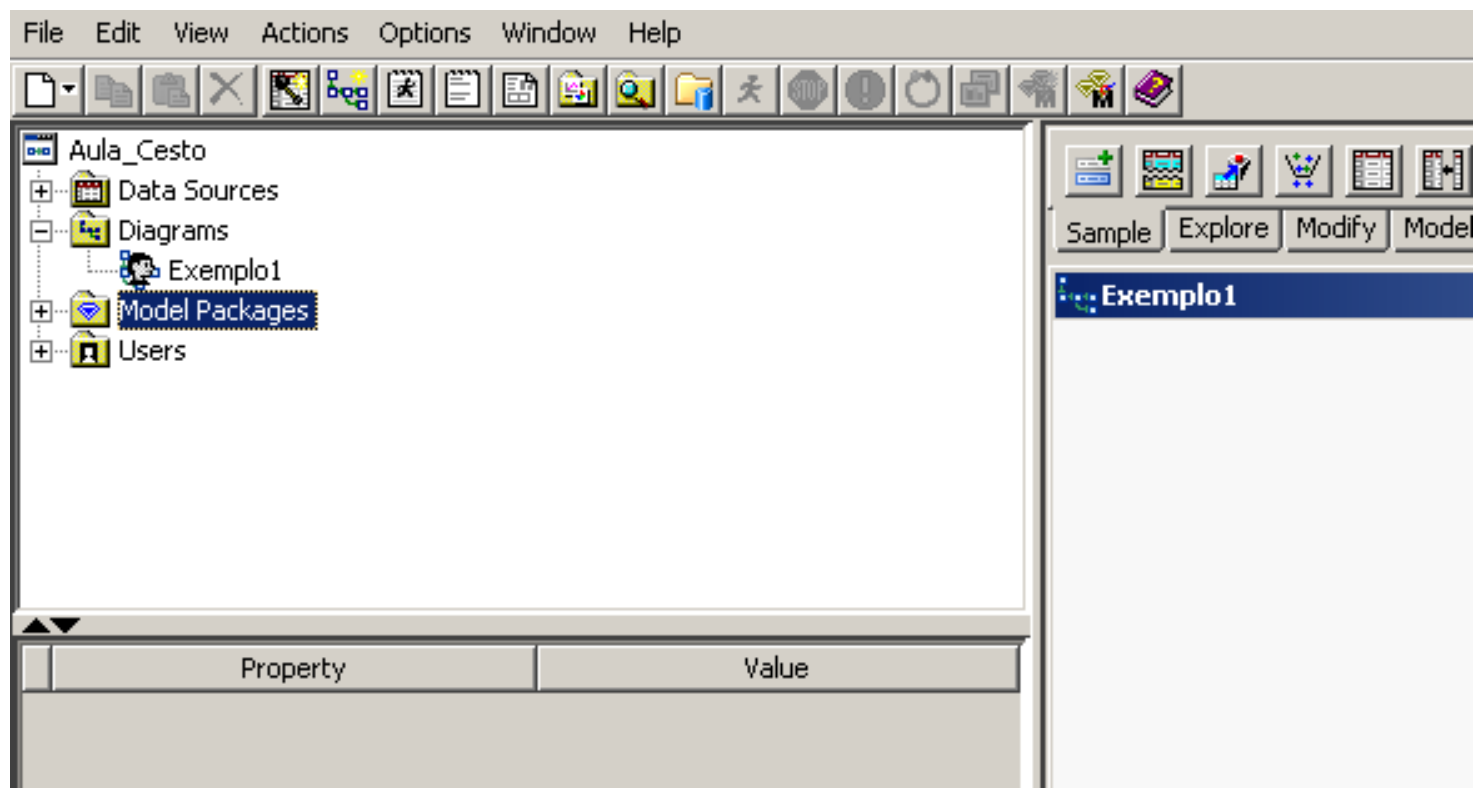
Property	Value
Name	Aula_Cesto
Project Start Code	
Created	02/04/17 16:55
Server	SASApp - Logical Workspace Serve
Grid Available	No
Path	C:\Aula_Cesto
Metadata Folder Path	/Users/sastrust/Mv Folder/Aula_Ce

O SAS cria um diretório de trabalho para salvar os arquivos.

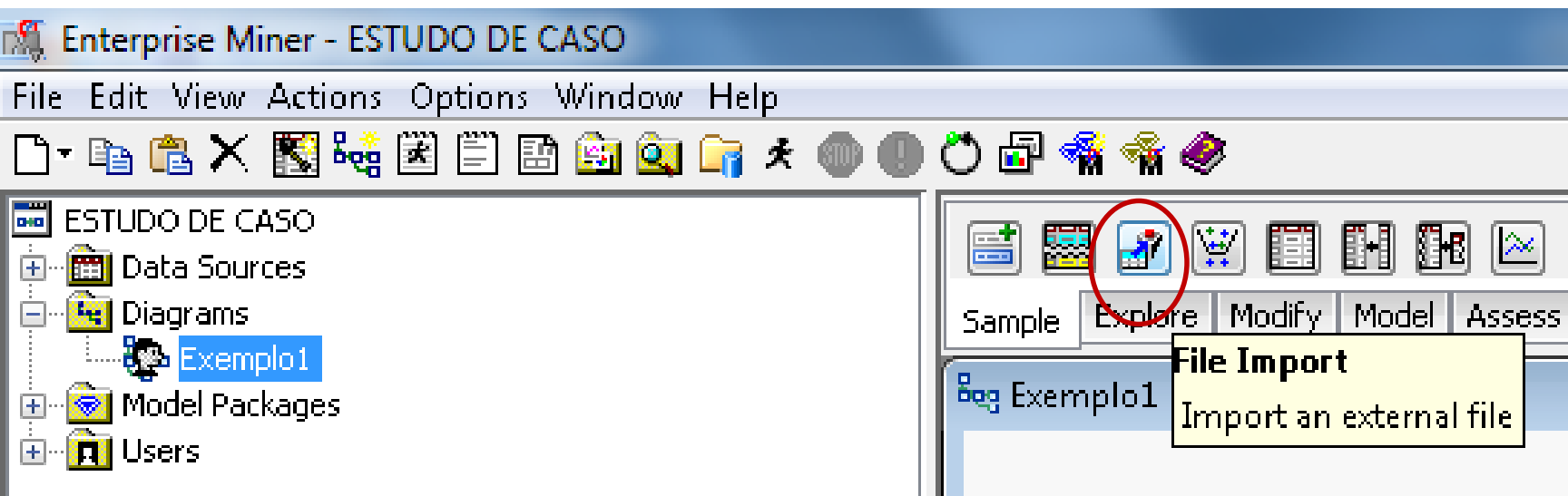


Criação de um diagrama para as análises. Clicar em diagrams com o botão direito.





Importar a base de dados

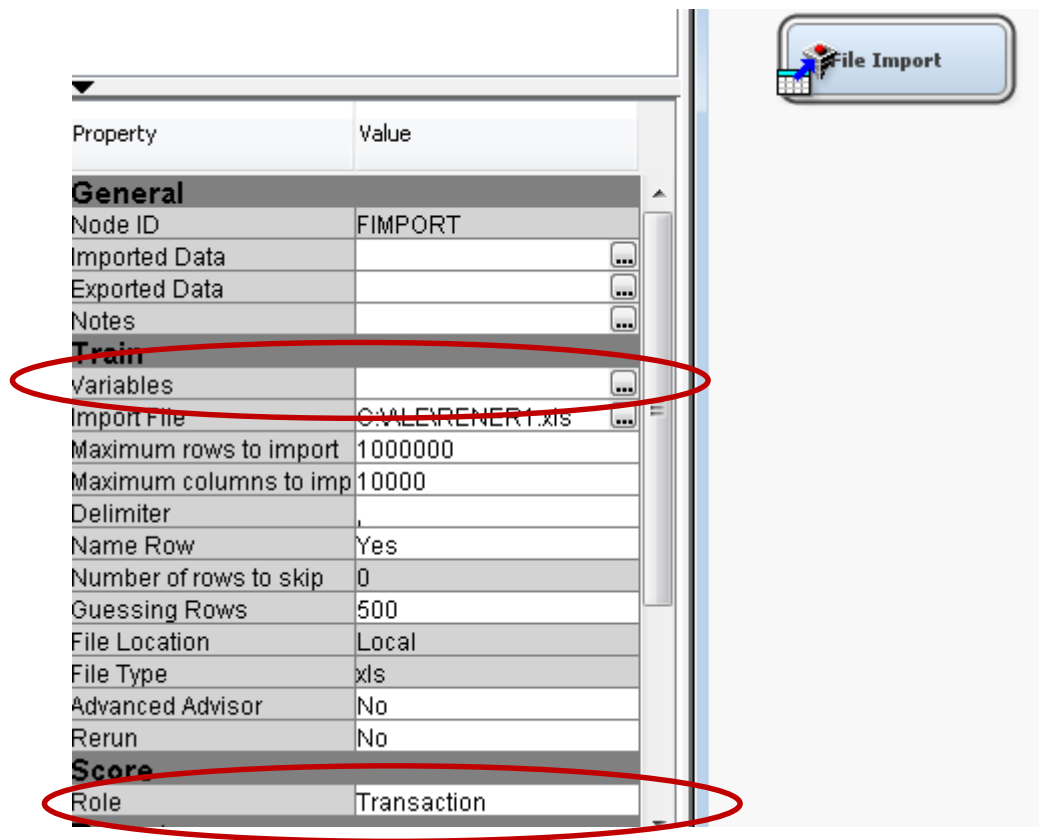


Importar a base de dados

File Import

Property	Value
General	
Node ID	FIMPORT
Imported Data	...
Exported Data	...
Notes	...
Train	
variables	...
Import File	C:\ALE\RENER1.xls
Maximum rows to import	1000000
Maximum columns to import	10000
Delimiter	,
Name Row	Yes
Number of rows to skip	0
Guessing Rows	500
File Location	Local
File Type	xls
Advanced Advisor	No
Rerun	No
Score	
Role	Transaction

O banco de dados deve estar como Transacional. Definir o tipo de variável



File Import

Property	Value
General	
Node ID	FIMPORT
Imported Data	...
Exported Data	...
Notes	...
Train	
Variables	...
Import File	C:\ALEXRENER1.xls
Maximum rows to import	1000000
Maximum columns to import	10000
Delimiter	,
Name Row	Yes
Number of rows to skip	0
Guessing Rows	500
File Location	Local
File Type	xls
Advanced Advisor	No
Rerun	No
Score	
Role	Transaction

Definir o tipo de variável

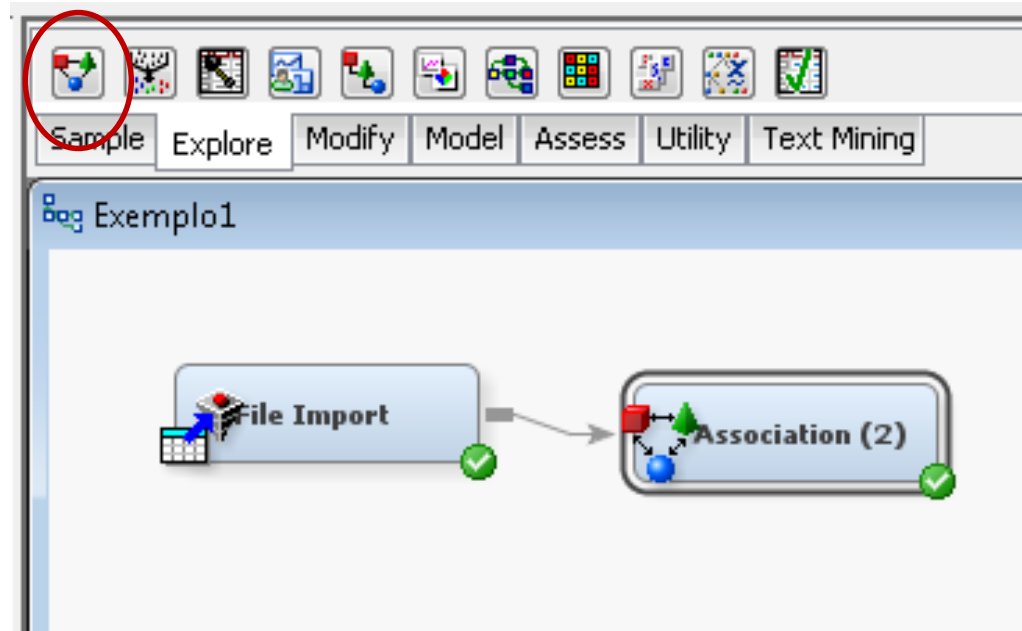
(none) ☐ not Equal to ...

Columns: ☐ Label ☐ Mining

Name	Role	Level	Report	Drop	Order
Cupon	ID	Interval	No	No	
Transacao	Target	Nominal	No	No	

Análise de Cesto de Compras

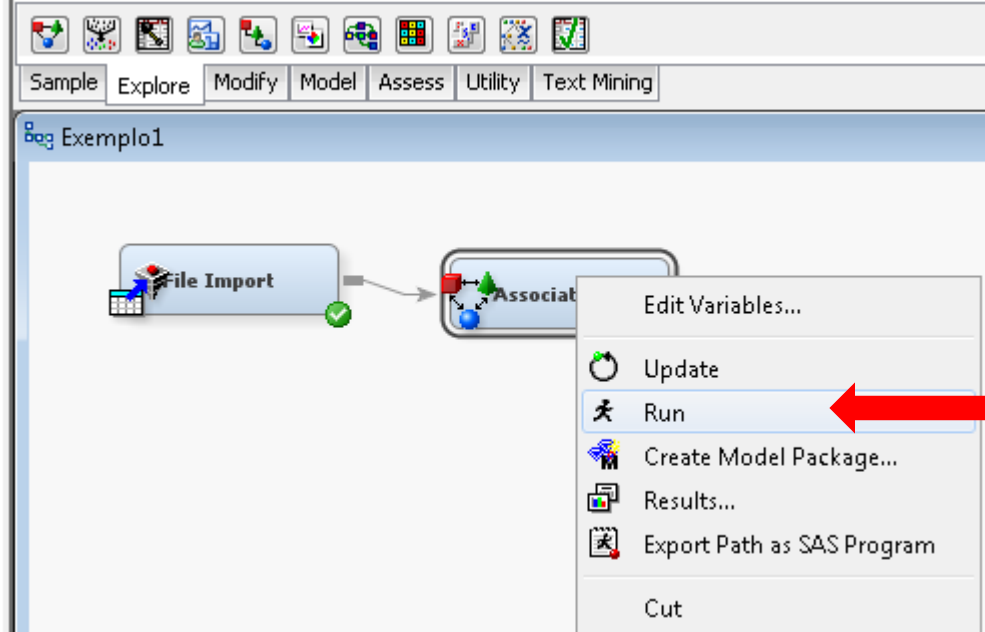
Análise de Cesto de Compras



Clicar em Association e alterar a propriedade **Maximum Items** para 2.

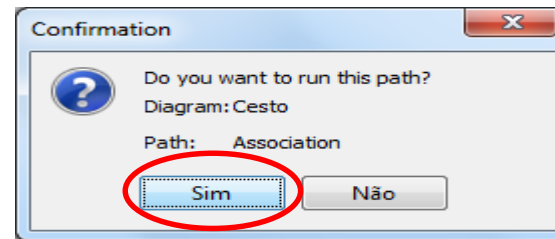
The screenshot displays a software interface with a left-hand pane containing a tree view of settings for a node named 'Assoc2'. The 'Association' section is expanded, and the 'Maximum Items' property is highlighted with a red circle. The value for 'Maximum Items' is set to 2. Other visible properties include 'Maximum Number of Item' (100000), 'Rules', 'Support Type' (Percent), 'Support Count' (1), and 'Support Percentage' (5.0). The right-hand pane shows a workflow diagram with a 'File Import' node connected to an 'Association (2)' node.

General	
Node ID	Assoc2
Imported Data	
Exported Data	
Notes	
Train	
Variables	
Maximum Number of Item	100000
Rules	
Association	
Maximum Items	2
Minimum Confidence Level	10
Support Type	Percent
Support Count	1
Support Percentage	5.0

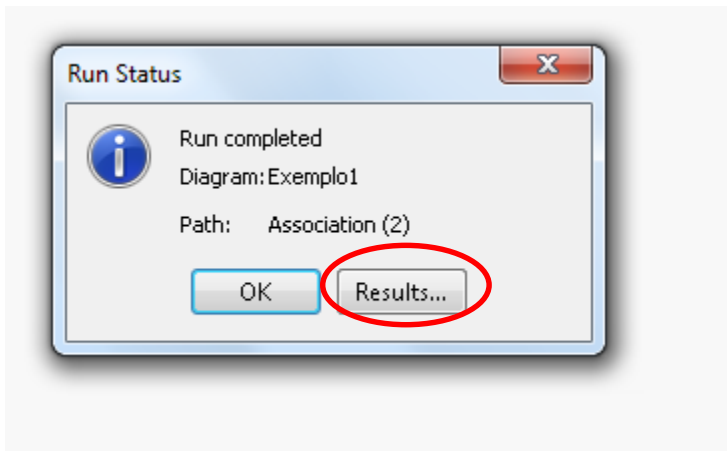


Clique com o botão direito em Association e depois em “Run”

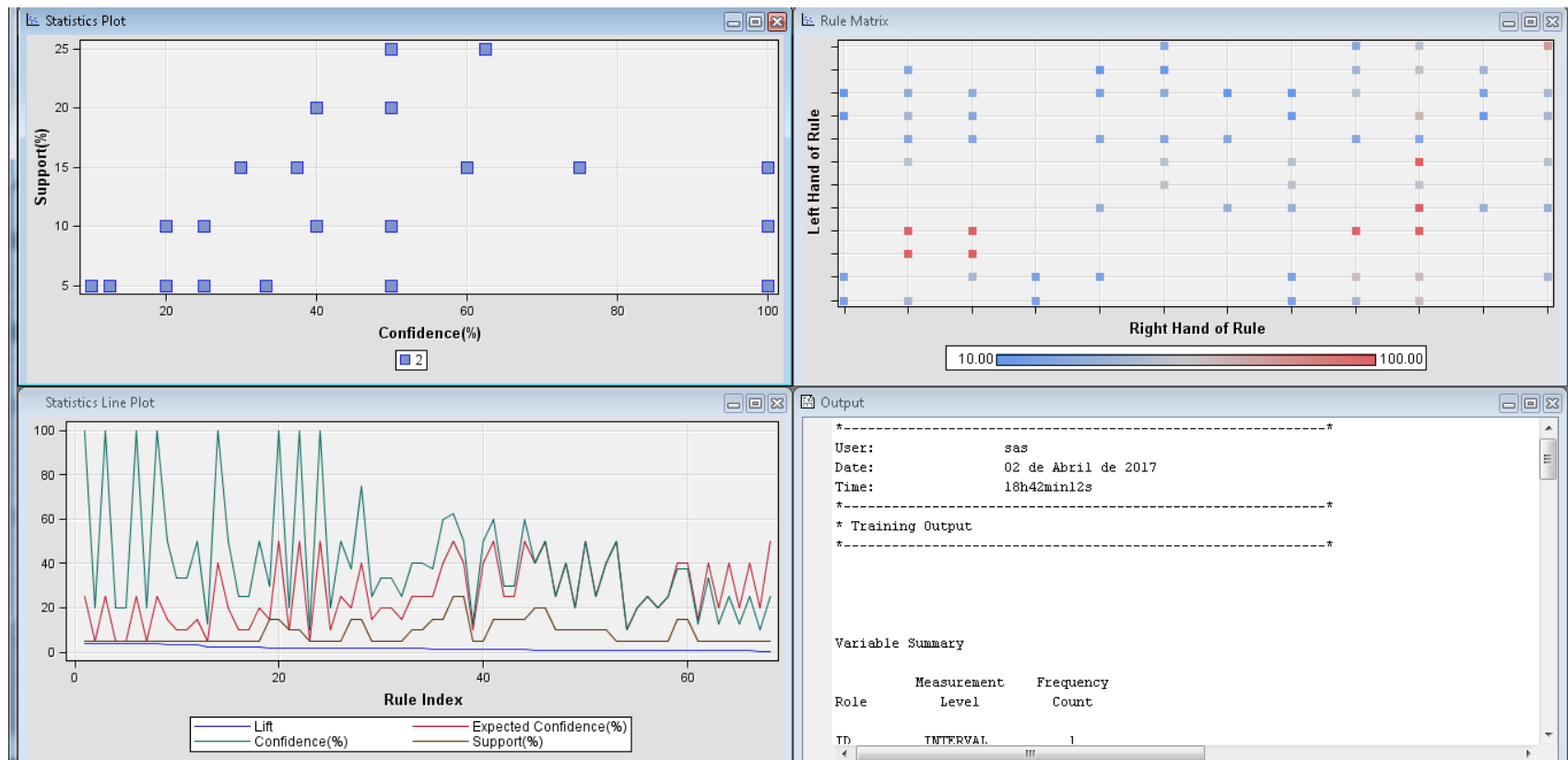
Depois, na caixa de mensagem seguinte, clique em “sim”



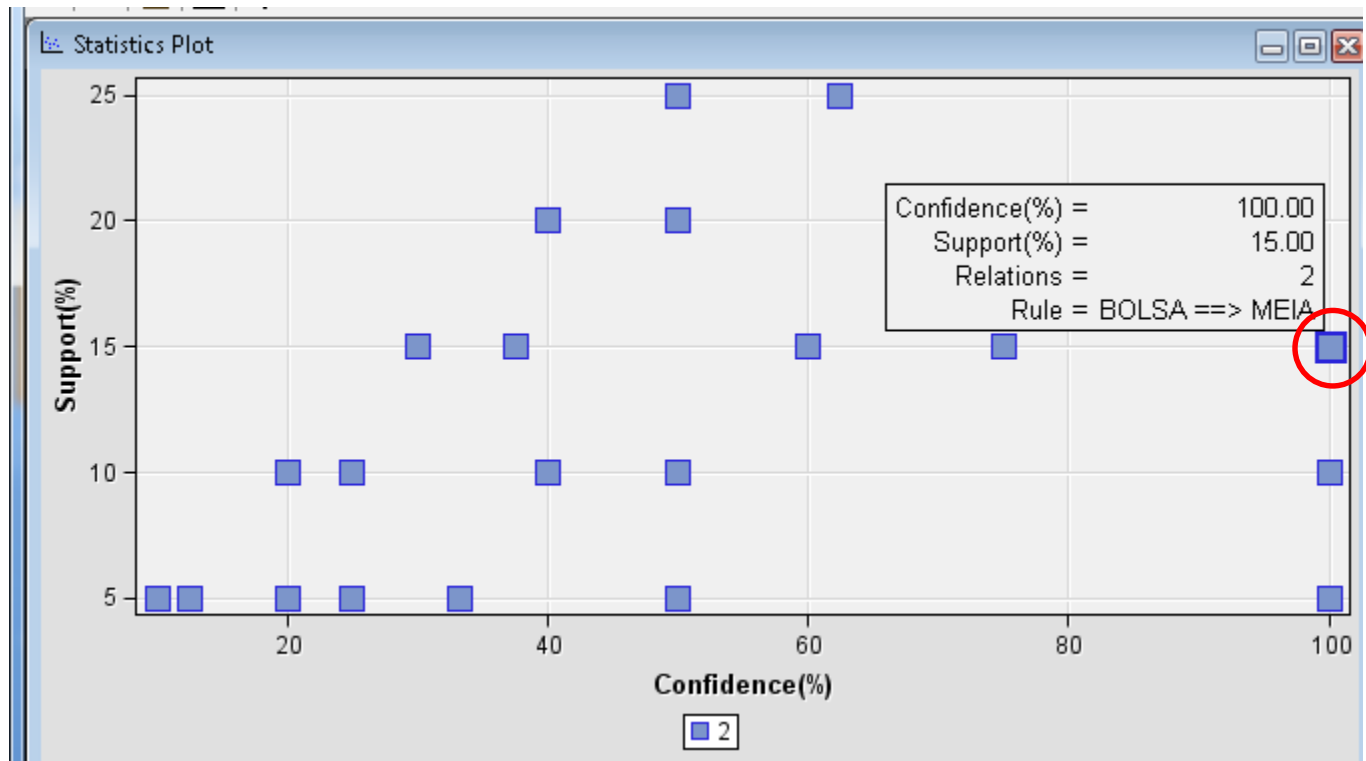
Após a execução, aparecerá uma caixa de mensagens informando o final do processamento. Caso o processamento tenha terminado com sucesso, aparecerão símbolos verdes nos componentes. Após isso, clique em “Results”



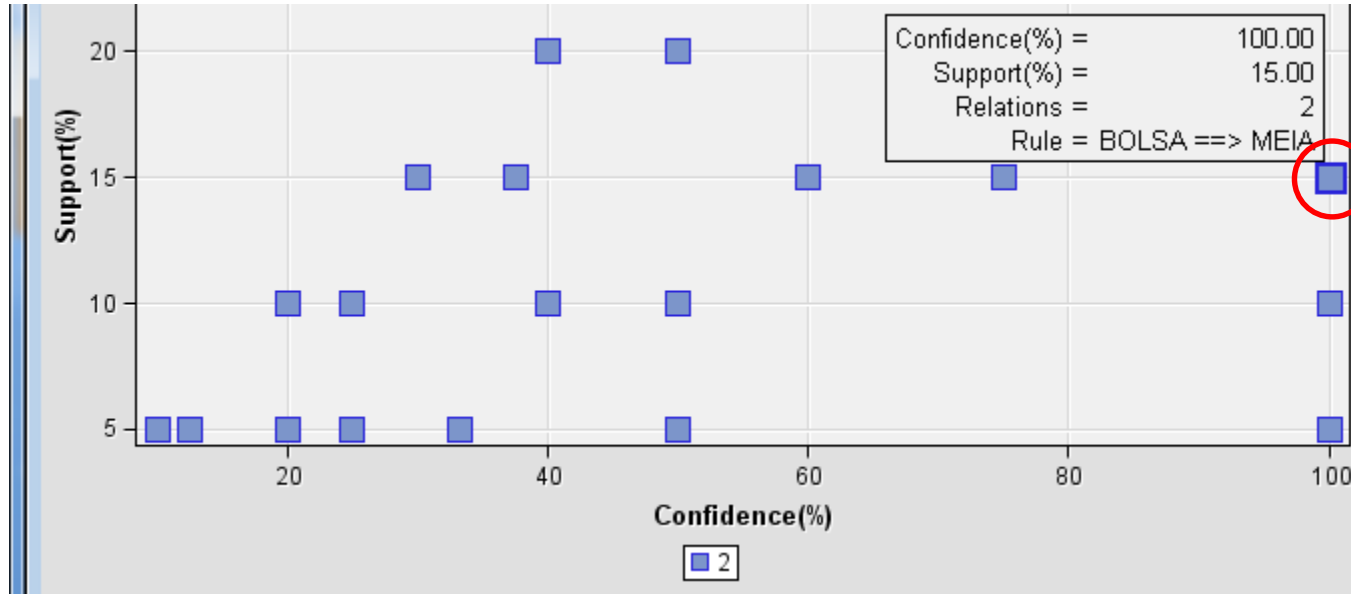
O Miner possui inúmeras ferramentas gráficas para analisar as quatro medidas.



Este gráfico apresenta o Support e o Confidence para as regras.



Este gráfico apresenta o Support e o Confidence para as regras.



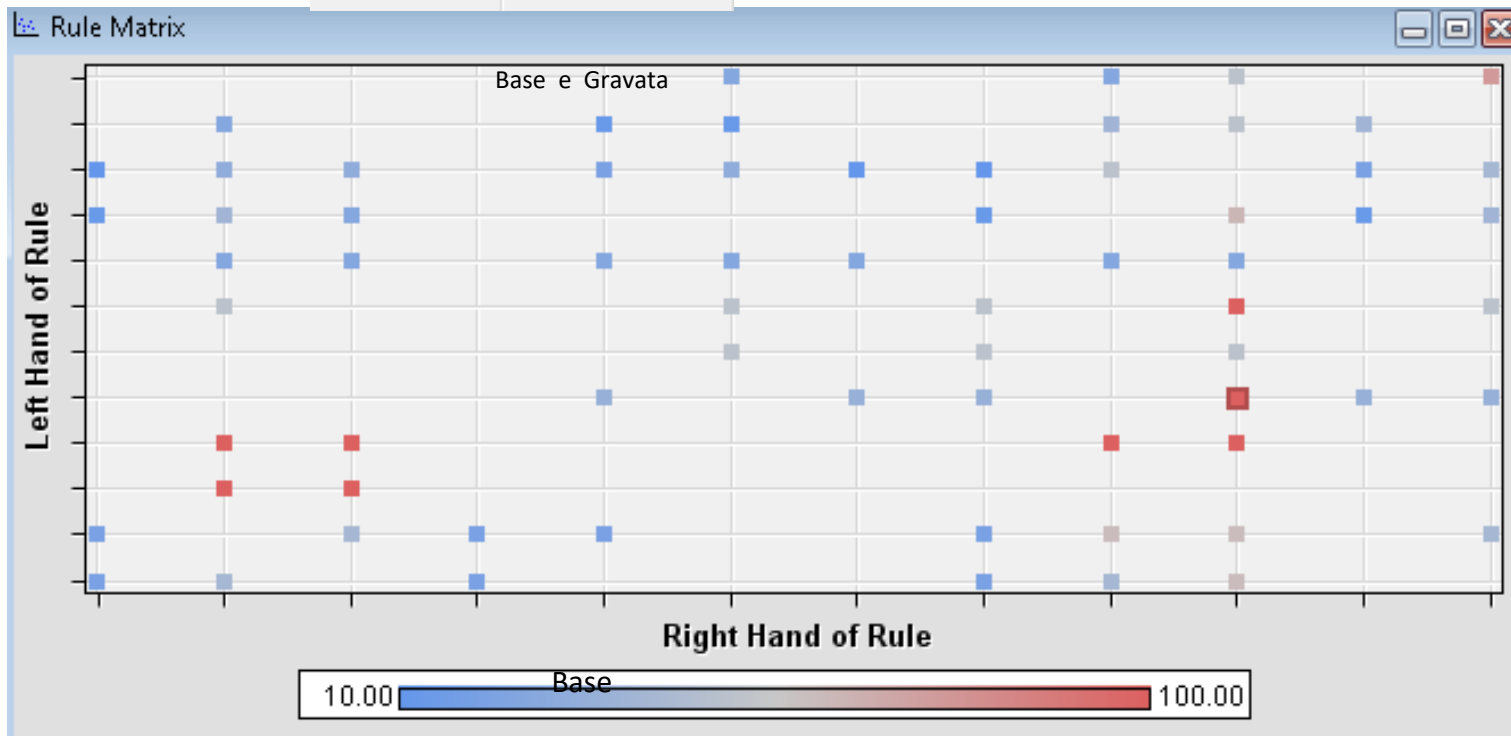
Considerando todas as compras, 15 % das vezes a Renner **vende Bolsa e Meia (Suport)**.

Considerando todas as compras de bolsa, 100 % das vezes a Renner **vende Meia (Confidence)**.

Este gráfico apresenta o **Confidence** para todas o regras em forma de matriz.

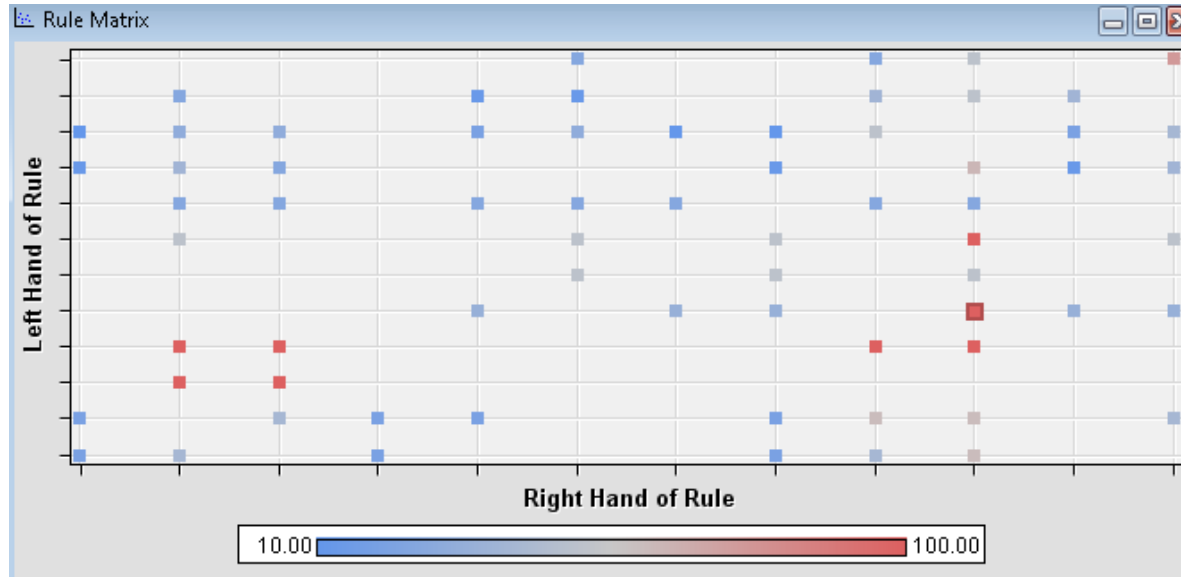
Right Hand of Rule = BASE
Left Hand of Rule = GRAVATA
Confidence(%) = 12.50

Gravata

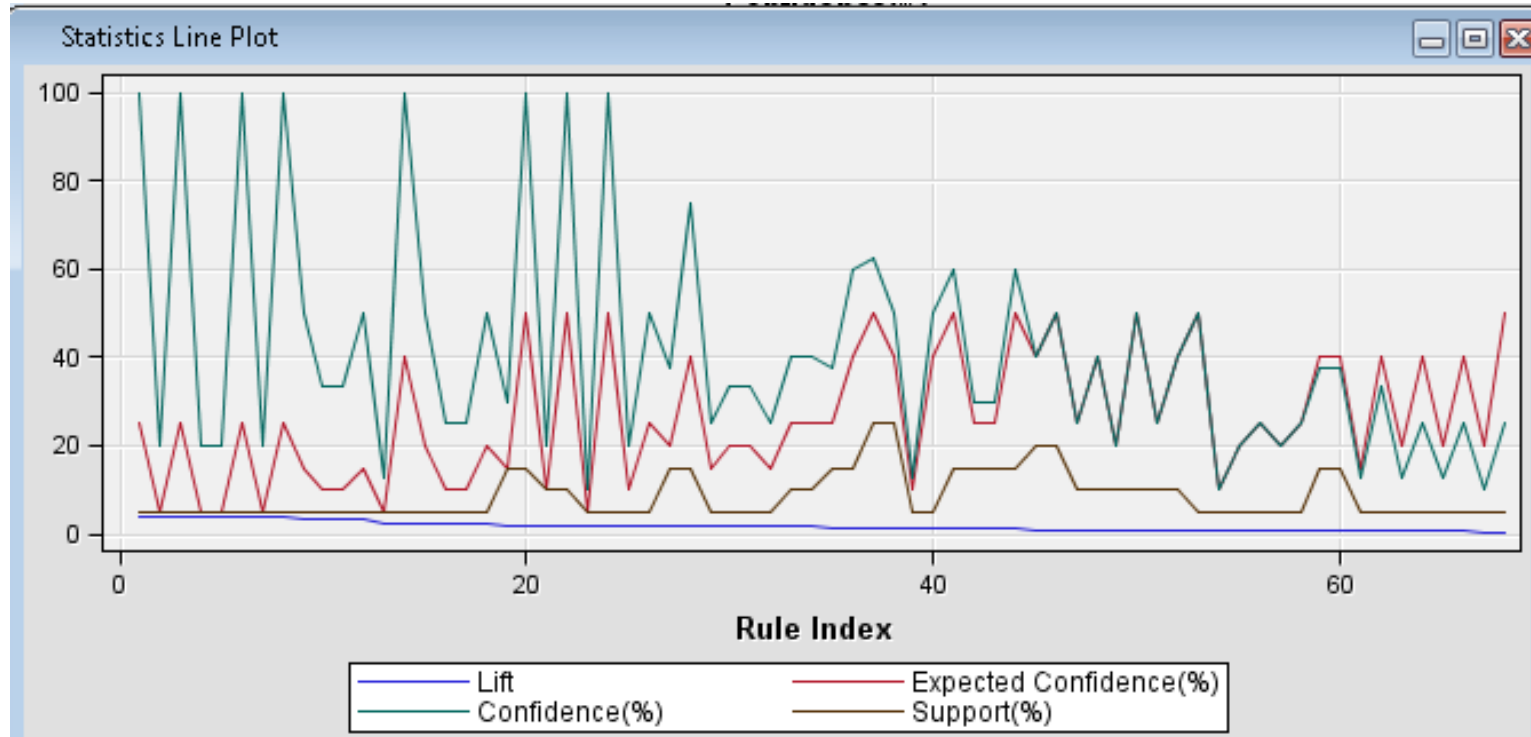


Este gráfico apresenta o **Confidence** para todas o regras em forma de matriz.

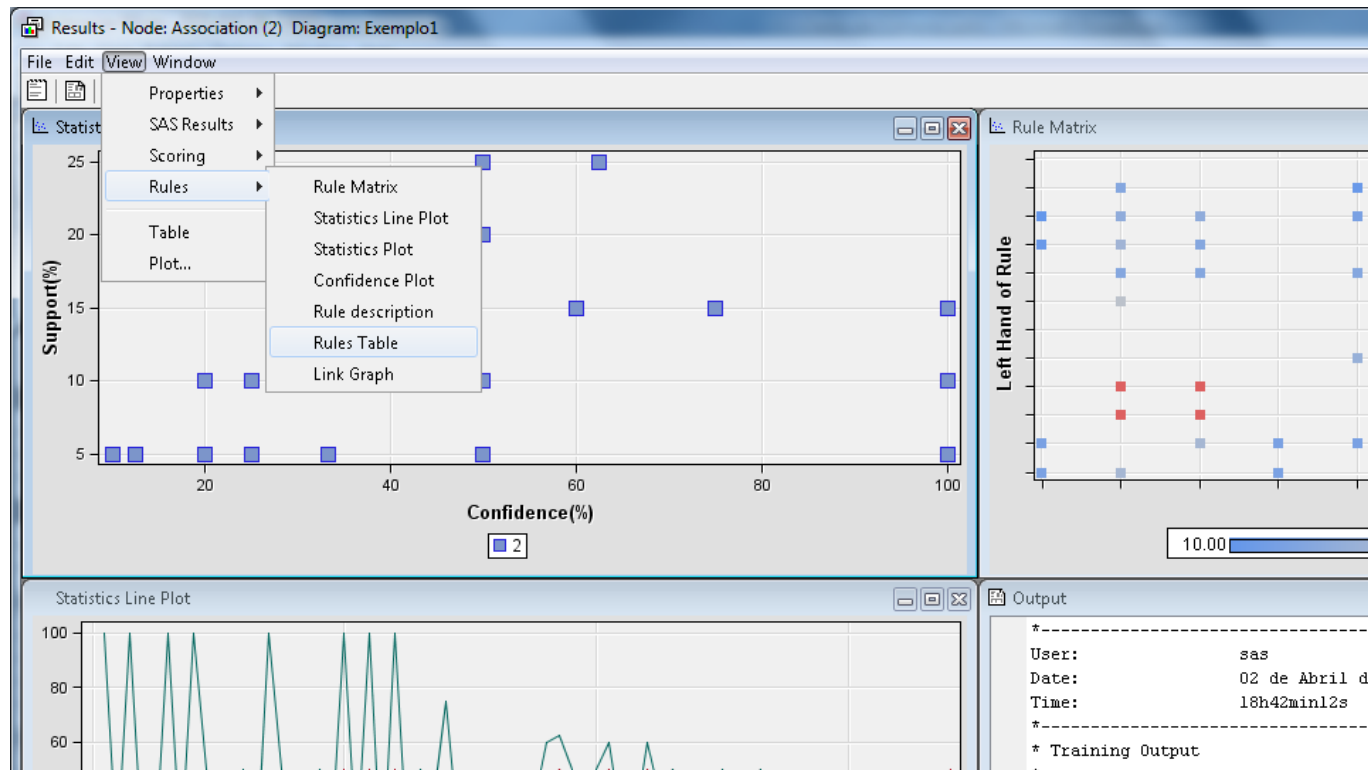
Quanto mais vermelho mais forte o confidence.



Este gráfico apresenta o Lift, Expected Confidence e Confidence de todas as regras.



Clique em “View”, depois “Rules” e depois em “Rules Table” para visualizar a tabela com os resultados.



Para ordenar uma coluna clicar na coluna.

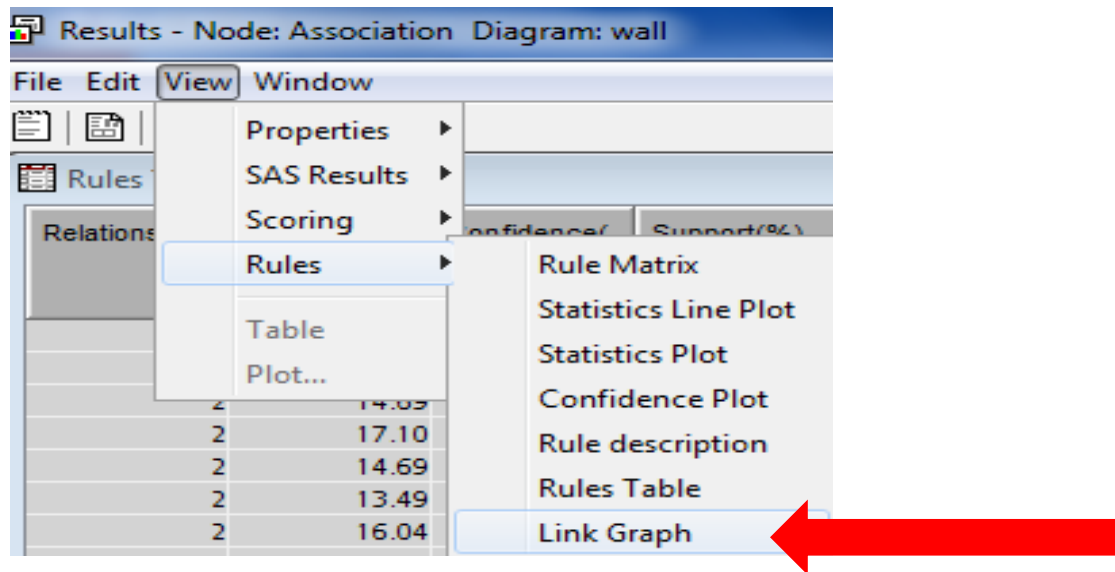
Rules Table

Relations	Expected Confidence(%)	Confidence(%)	Support(%)	Lift ▼	Transaction Count	Rule	Left Hand of Rule	Right Hand of Rule	Rule Item 1	
2	5.00	20.00	5.00	4.00	1.00	PERFUME ==> BLUSA	PERFUME	BLUSA	PERFUME	=
2	5.00	20.00	5.00	4.00	1.00	CALCA ==> BLUSA	CALCA	BLUSA	CALCA	=
2	25.00	100.00	5.00	4.00	1.00	SAPATO ==> CALCA	SAPATO	CALCA	SAPATO	=
2	25.00	100.00	5.00	4.00	1.00	BLUSA ==> CALCA	BLUSA	CALCA	BLUSA	=
2	25.00	100.00	5.00	4.00	1.00	SAPATO ==> PERFUME	SAPATO	PERFUME	SAPATO	=
2	25.00	100.00	5.00	4.00	1.00	BLUSA ==> PERFUME	BLUSA	PERFUME	BLUSA	=
2	5.00	20.00	5.00	4.00	1.00	PERFUME ==> SAPATO	PERFUME	SAPATO	PERFUME	=
2	5.00	20.00	5.00	4.00	1.00	CALCA ==> SAPATO	CALCA	SAPATO	CALCA	=
2	10.00	33.33	5.00	3.33	1.00	BOLSA ==> BASE	BOLSA	BASE	BOLSA	=
2	15.00	50.00	5.00	3.33	1.00	SOMBRA ==> BOLSA	SOMBRA	BOLSA	SOMBRA	=
2	15.00	50.00	5.00	3.33	1.00	BASE ==> BOLSA	BASE	BOLSA	BASE	=
2	10.00	33.33	5.00	3.33	1.00	BOLSA ==> SOMBRA	BOLSA	SOMBRA	BOLSA	=
2	10.00	25.00	5.00	2.50	1.00	BATON ==> BASE	BATON	BASE	BATON	=
2	20.00	50.00	5.00	2.50	1.00	SOMBRA ==> BATON	SOMBRA	BATON	SOMBRA	=
2	20.00	50.00	5.00	2.50	1.00	BASE ==> BATON	BASE	BATON	BASE	=
2	5.00	12.50	5.00	2.50	1.00	TENIS ==> BLUSA	TENIS	BLUSA	TENIS	=
2	10.00	25.00	5.00	2.50	1.00	BATON ==> SOMBRA	BATON	SOMBRA	BATON	=

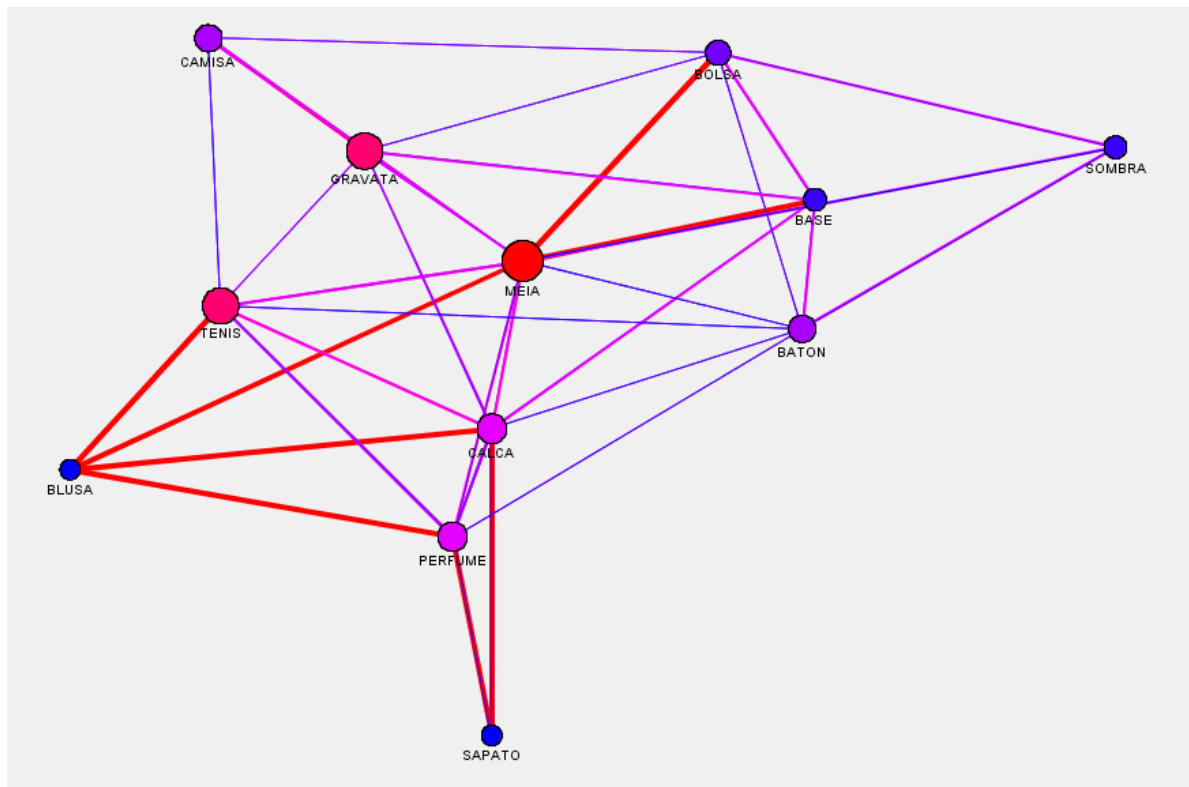
Rules Table

Relations	Expected Confidence(%)	Confidence(%)	Support(%)	Lift ▲	Transaction Count	Rule	Left Hand of Rule	Right Hand of Rule	Rule Item 1	Rule Item 2	Rule Item 3	Rule
2	20.00	10.00	5.00	0.50	1.00	MEIA ==> BATON	MEIA	BATON	MEIA	=====...	BATON	
2	50.00	25.00	5.00	0.50	1.00	BATON ==> MEIA	BATON	MEIA	BATON	=====...	MEIA	
2	20.00	12.50	5.00	0.63	1.00	TENIS ==> BATON	TENIS	BATON	TENIS	=====...	BATON	
2	20.00	12.50	5.00	0.63	1.00	TENIS ==> CAMISA	TENIS	CAMISA	TENIS	=====...	CAMISA	
2	40.00	25.00	5.00	0.63	1.00	CAMISA ==> TENIS	CAMISA	TENIS	CAMISA	=====...	TENIS	
2	40.00	25.00	5.00	0.63	1.00	BATON ==> TENIS	BATON	TENIS	BATON	=====...	TENIS	
2	15.00	12.50	5.00	0.83	1.00	GRAVATA ==> BOLSA	GRAVATA	BOLSA	GRAVATA	=====...	BOLSA	
2	40.00	33.33	5.00	0.83	1.00	BOLSA ==> GRAVATA	BOLSA	GRAVATA	BOLSA	=====...	GRAVATA	
2	40.00	37.50	15.00	0.94	3.00	TENIS ==> GRAVATA	TENIS	GRAVATA	TENIS	=====...	GRAVATA	
2	40.00	37.50	15.00	0.94	3.00	GRAVATA ==> TENIS	GRAVATA	TENIS	GRAVATA	=====...	TENIS	
2	20.00	20.00	5.00	1.00	1.00	PERFUME ==> BATON	PERFUME	BATON	PERFUME	=====...	BATON	
2	20.00	20.00	5.00	1.00	1.00	CALCA ==> BATON	CALCA	BATON	CALCA	=====...	BATON	
2	25.00	25.00	10.00	1.00	2.00	GRAVATA ==> CALCA	GRAVATA	CALCA	GRAVATA	=====...	CALCA	
2	25.00	25.00	5.00	1.00	1.00	BATON ==> CALCA	BATON	CALCA	BATON	=====...	CALCA	
2	20.00	20.00	10.00	1.00	2.00	MEIA ==> CAMISA	MEIA	CAMISA	MEIA	=====...	CAMISA	
2	40.00	40.00	20.00	1.00	4.00	MEIA ==> GRAVATA	MEIA	GRAVATA	MEIA	=====...	GRAVATA	
2	40.00	40.00	10.00	1.00	2.00	CALCA ==> GRAVATA	CALCA	GRAVATA	CALCA	=====...	GRAVATA	
2	50.00	50.00	20.00	1.00	4.00	GRAVATA ==> MEIA	GRAVATA	MEIA	GRAVATA	=====...	MEIA	

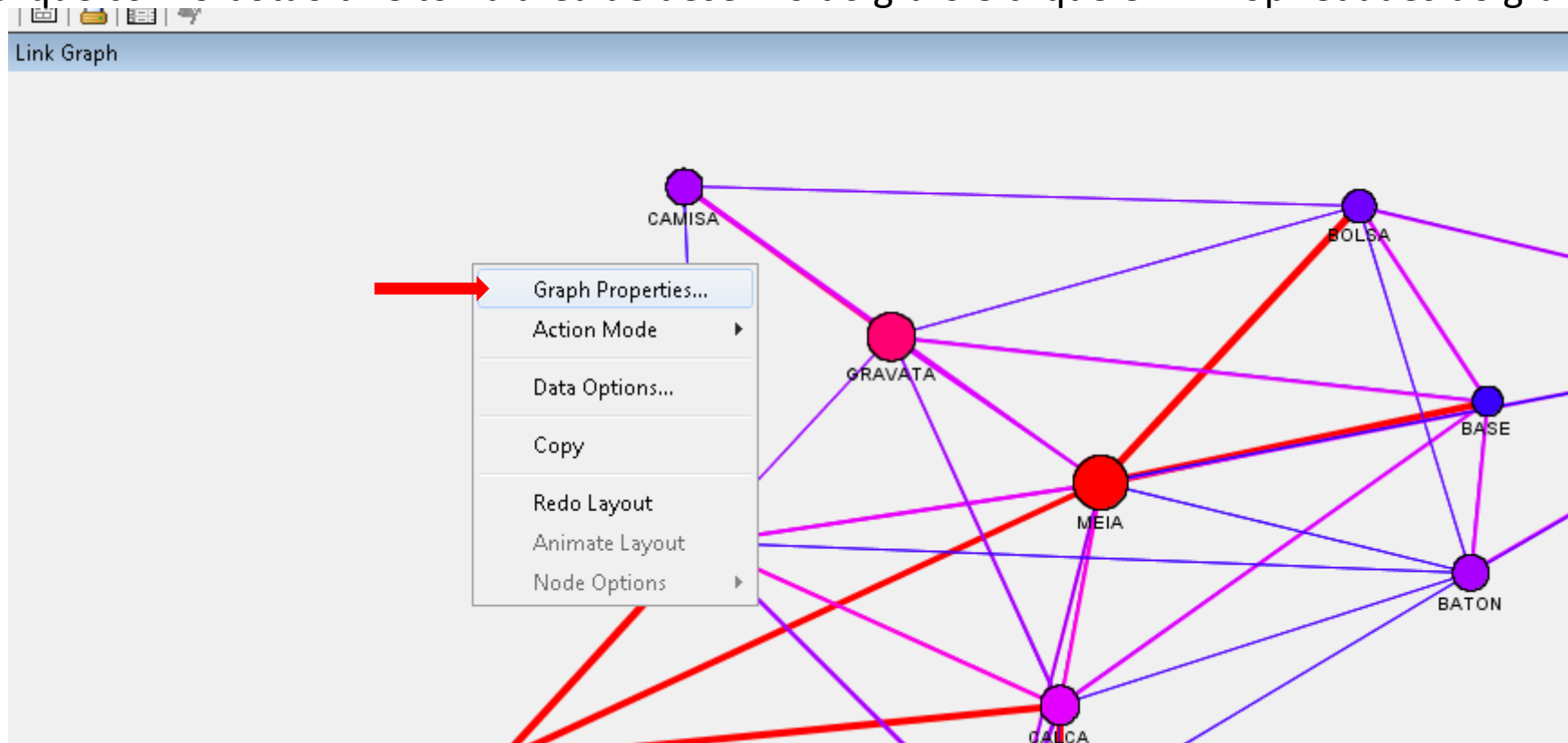
Analise as relações entre o produto Perfume e os demais produtos.

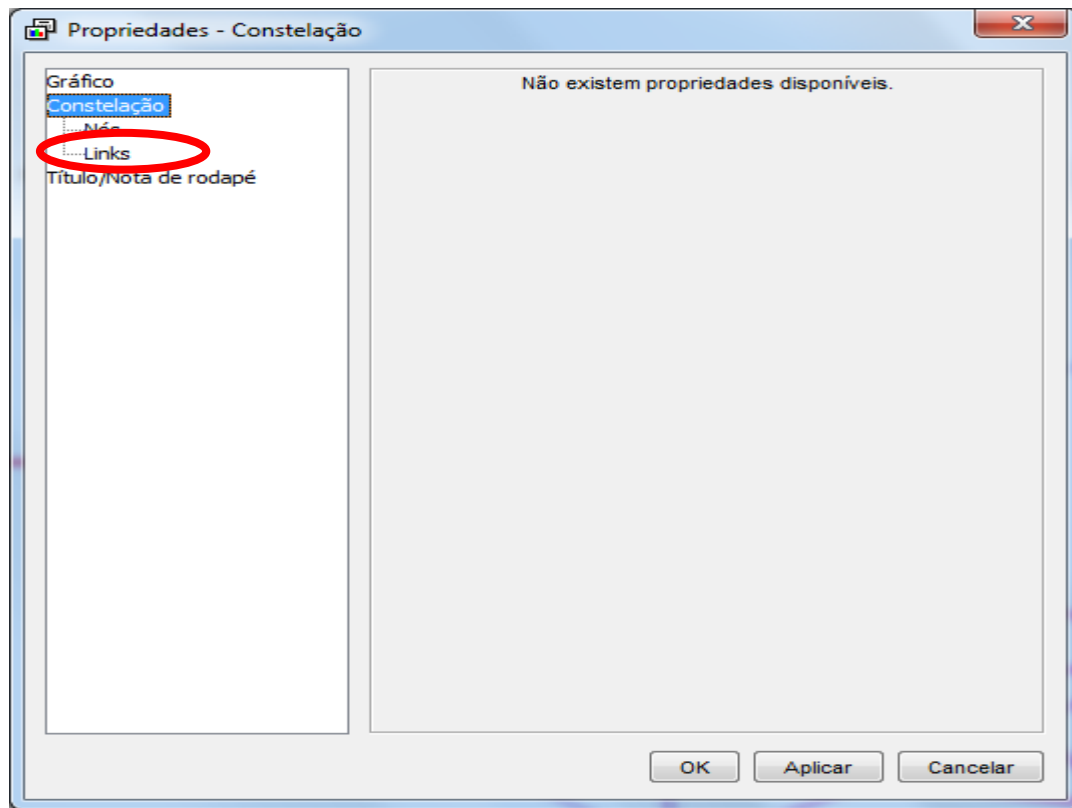


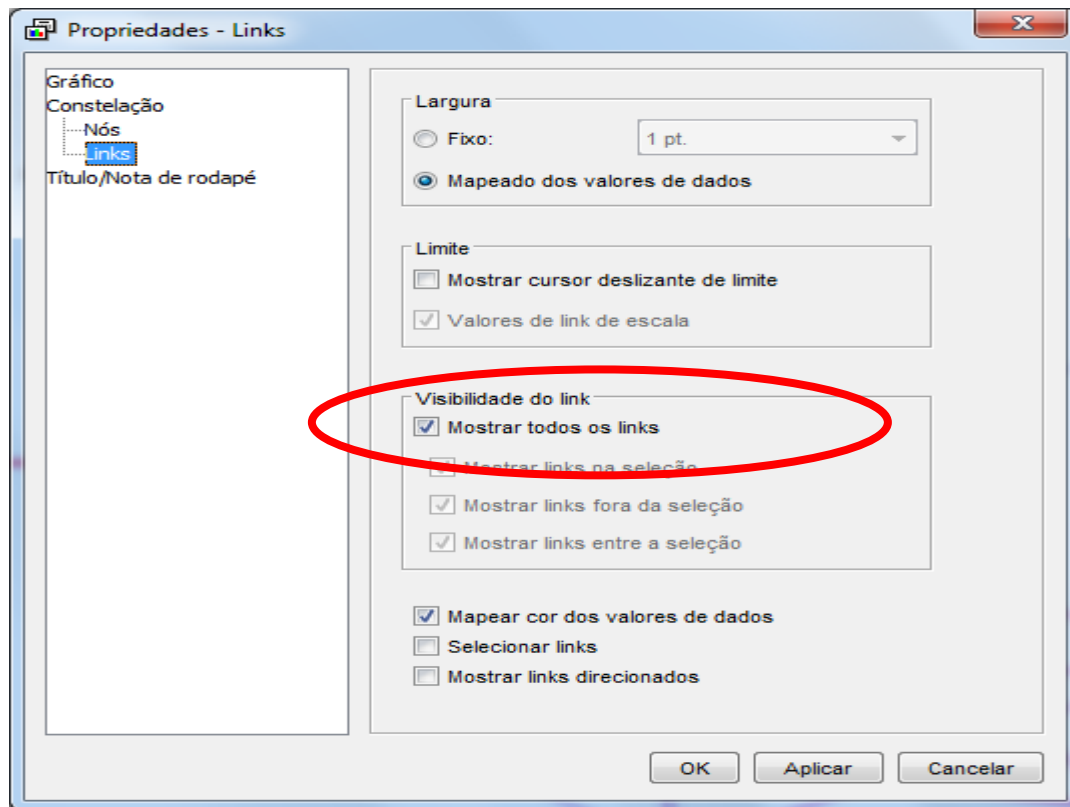
Interpretação: Quem compra Camisa compra Gravata
Quem compra Sombra compra Base

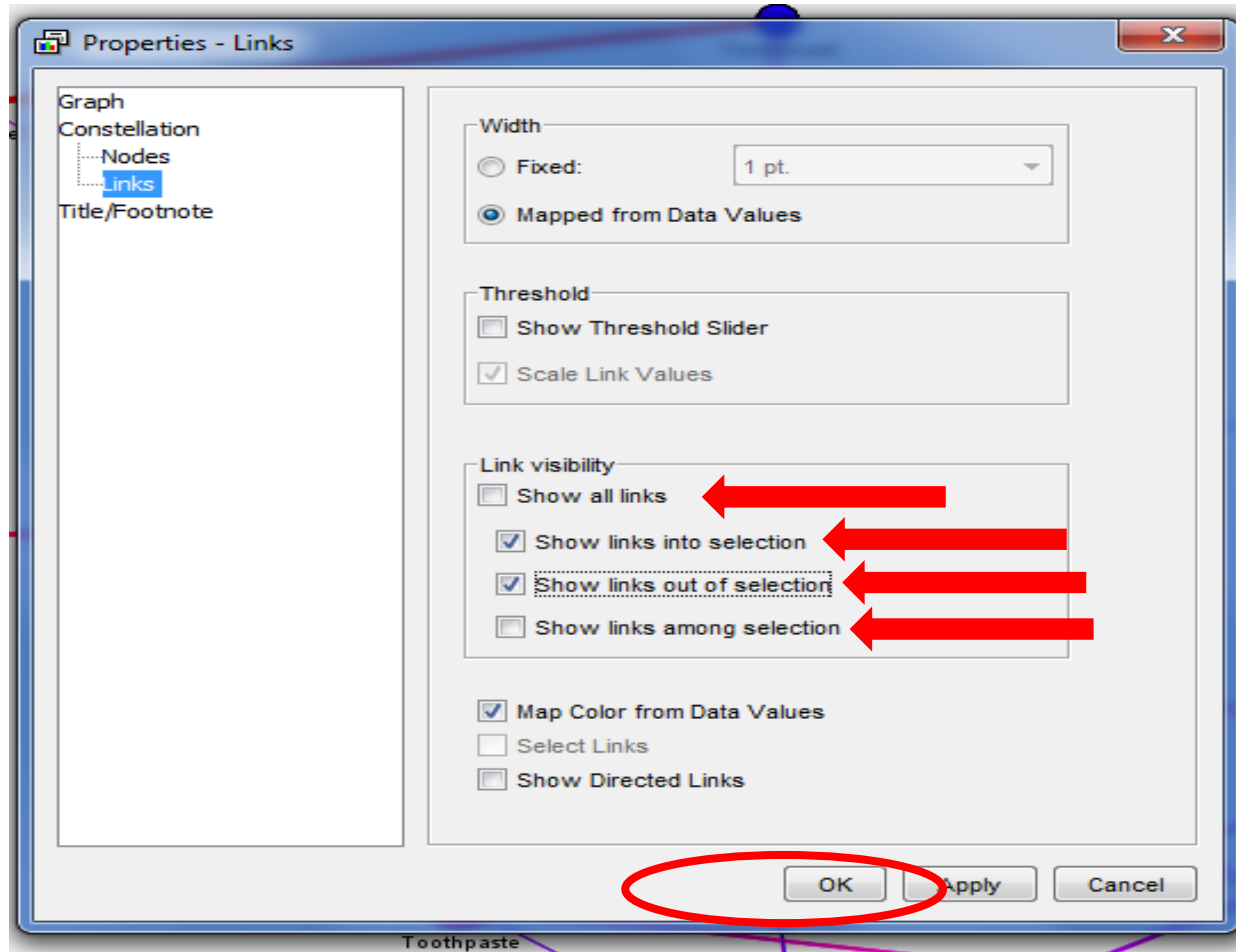


Clique com o botão direito na área de desenho do grafo e clique em “Propriedades do gráfico”









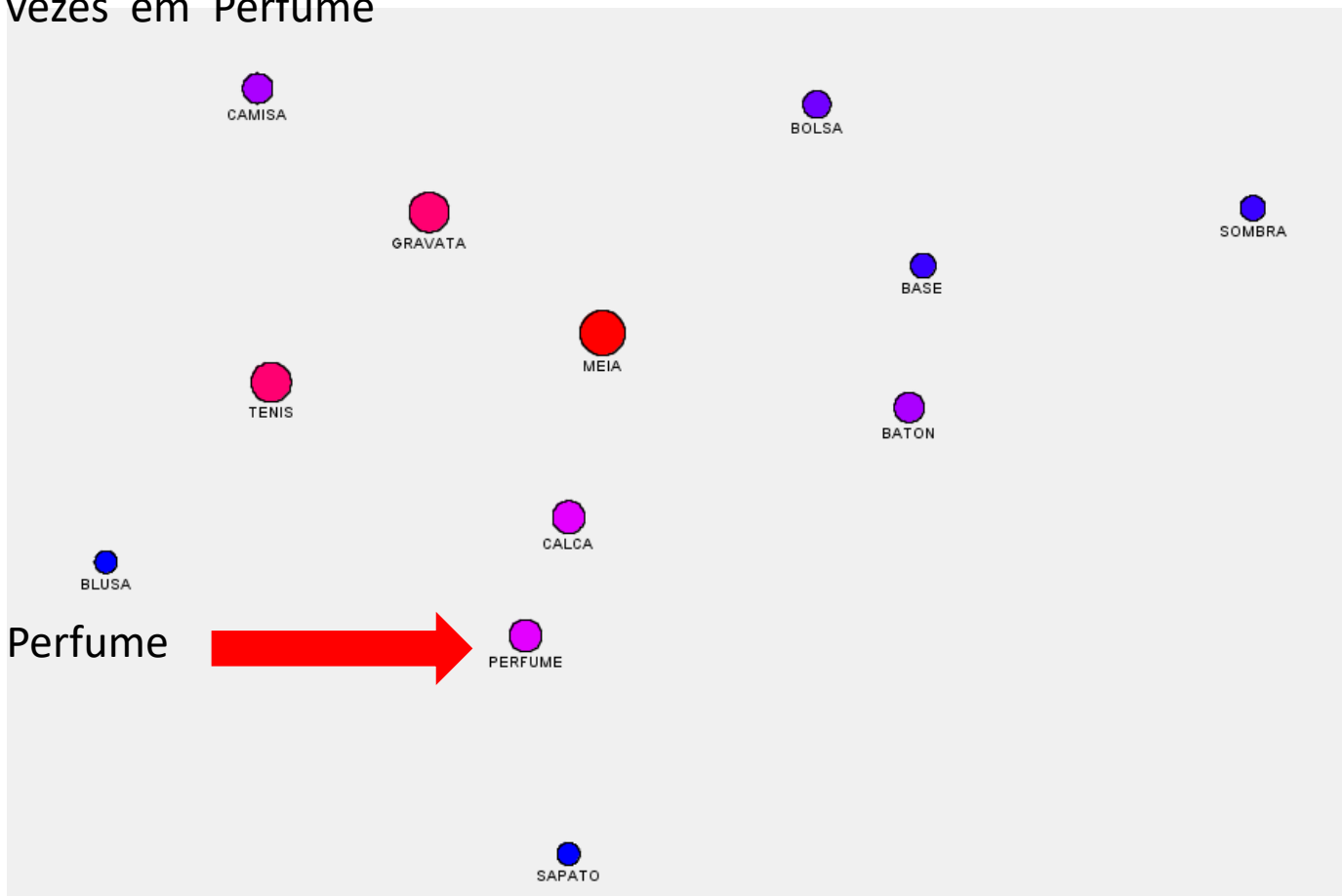
Toothpaste



LABDATA



Clicar duas vezes em Perfume

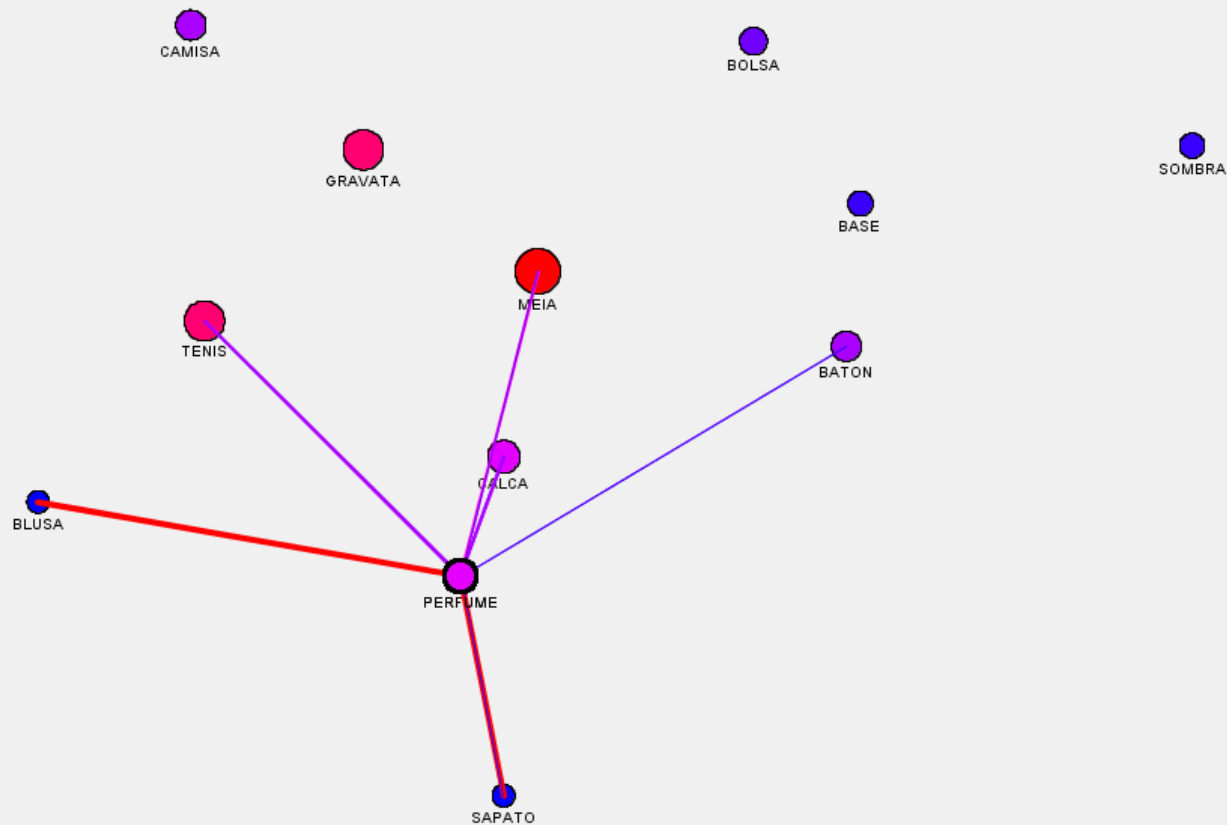


Clique em Perfume

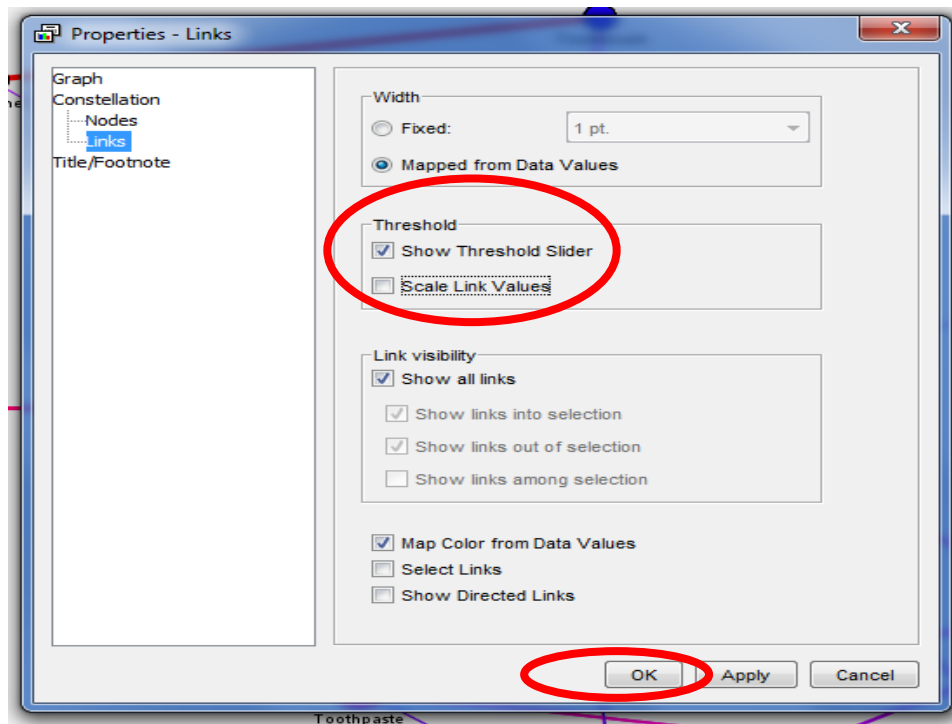


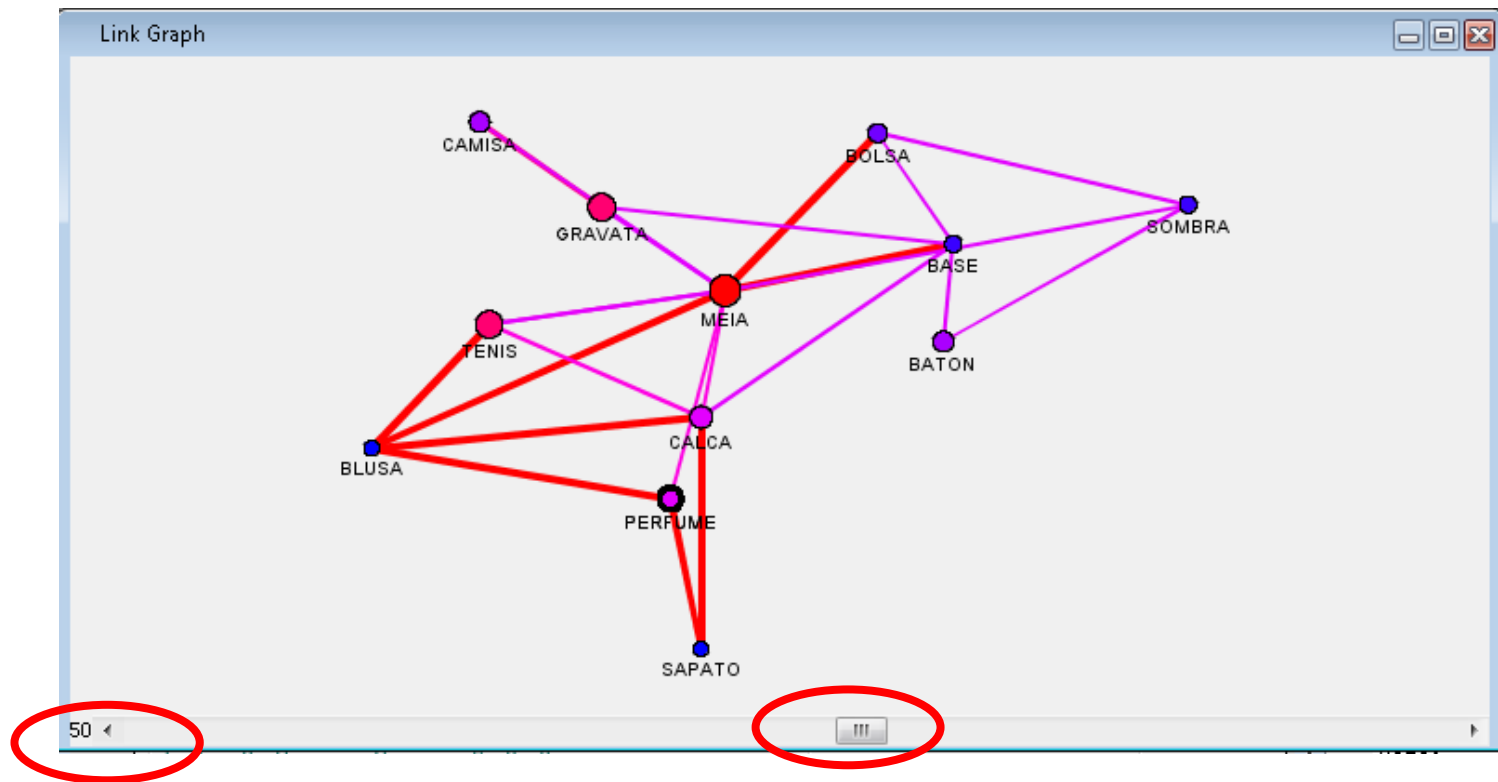
PERFUME

Apresenta os links
com Perfume



Regras com confidence 50%.





Deslize o cursor até que o número à esquerda tenha o valor 50

Exercício



Base de Dados

A base de dados é formada por transações de compras considerando 20 cupons. Nestas transações foram adquiridos 12 produtos diferentes, podendo conter mais de um produto por transação.

Tabela: Rener1.xls

Clicar em Association e alterar a propriedade **Maximum Items** para 3.

Property	Value
General	
Node ID	Assoc2
Imported Data	...
Exported Data	...
Notes	...
Train	
Variables	...
Maximum Number of Item	100000
Rules	...
Association	
Maximum Items	3
Minimum Confidence Level	1.0
Support Type	Percent
Support Count	1
Support Percentage	5.0
Sequence	
Chain Count	3
Consolidate Time	0.0
Maximum Transaction Duration	0.0
Support Type	Percent
Support Count	1

Considerando o Lift responda as questões:

1 – Qual foi o maior Lift ?

2 – Quantas regras tiveram o maior lift ?

3 – Tennis e perfume ajudam a vender que produto ?

4 – O cliente que compra bolsa e baton também compra qual produto ?

Por meio da análise do Link Graph qual produto não pode faltar ?

Exercício



Base de Dados

A base de dados utilizada nesta primeira demonstração trata de um arquivo contendo dados de compras de produtos de uma loja de conveniência, na qual possui 200.000 transações e 17 produtos diferentes, podendo conter mais de um produto por transação.

Tabela: Compras.sas7bdat

Produtos da base de dados e número de transações com o produto.

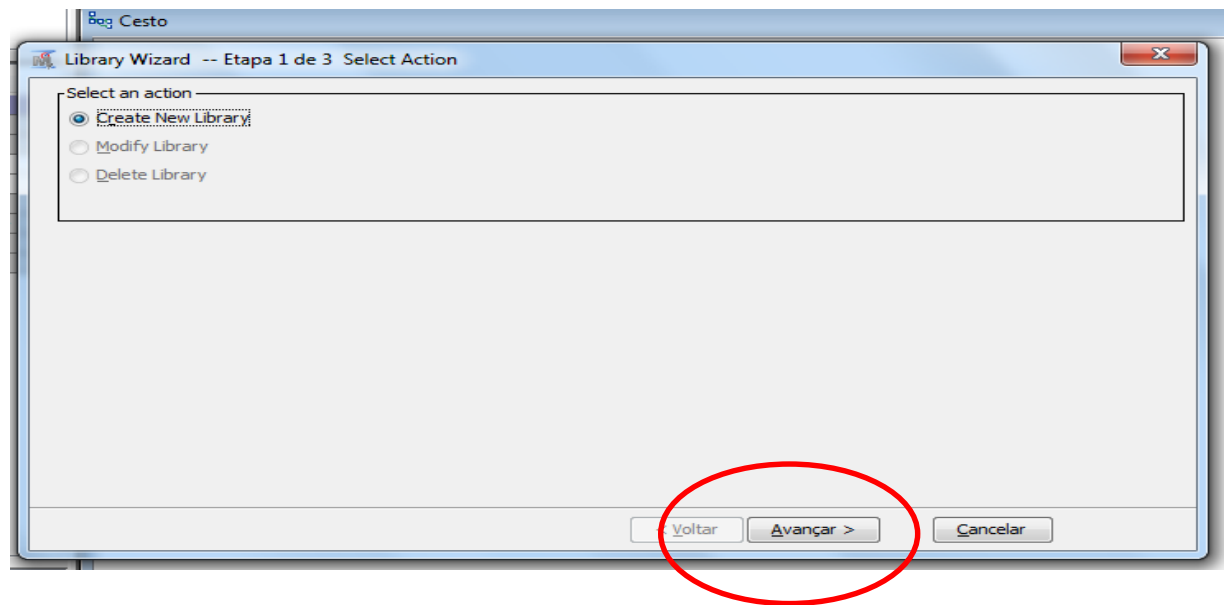
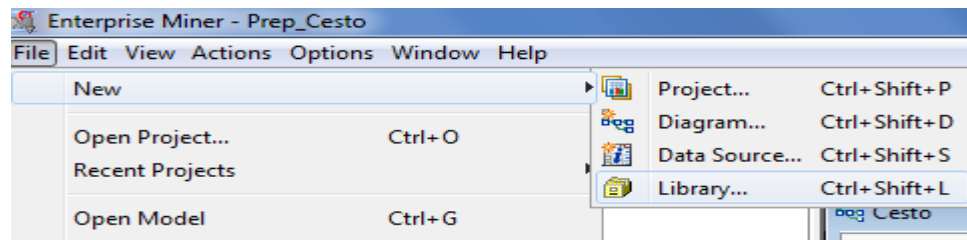
Magazine	73951
Toothpaste	63805
Candy Bar	62375
Greeting Cards	40292
Pencils	38123
Pens	36032
Perfume	25142
Bow	24554
Wrapping Paper	17227
Photo Processing	16386
Toothbrush	15774
Soap	14415
Shampoo	12452
Pain Reliever	8833
Prescription Med	5702
Markers	2605
Deodorant	1590

Tabela Transacional

	Input Data	Code	Log	Output Data
	Modify Task	Filter and Sort	Query Builder	
	transacao	produto		
1	12359	Candy Bar		
2	12362	Pain Reliever		
3	12362	Pain Reliever		
4	12365	Toothpaste		
5	12371	Bow		
6	12380	Greeting Cards		
7	12383	Pain Reliever		
8	12383	Pain Reliever		
9	12386	Pain Reliever		
10	12386	Pain Reliever		
11	12392	Shampoo		
12	12392	Magazine		
13	12401	Candy Bar		
14	12401	Candy Bar		
15	12401	Pencils		
16	12401	Magazine		
17	12404	Candy Bar		
18	12410	Pens		
19	12413	Magazine		
20	12470	Candy Bar		
21	12473	Shampoo		
22	12476	Shampoo		
23	12482	Toothpaste		
24	12563	Deodorant		
25	12563	Deodorant		
26	12563	Deodorant		
27	12563	Deodorant		
28	12563	Deodorant		
29	12572	Candy Bar		
30	12572	Candy Bar		

Candy Bar

Criar uma biblioteca apontando para o diretório onde está a tabela **Compras.sas7bdat.**



Library Wizard -- Etapa 2 de 3 Create or Modify

Name: Cesto

Engine: BASE

Library information

Path: Browse...

Options:

< Voltar Avançar > Cancelar

Apontar para o diretório em que está a tabela **Compras.sas7bdat**

Library Wizard -- Etapa 2 de 3 Create or Modify

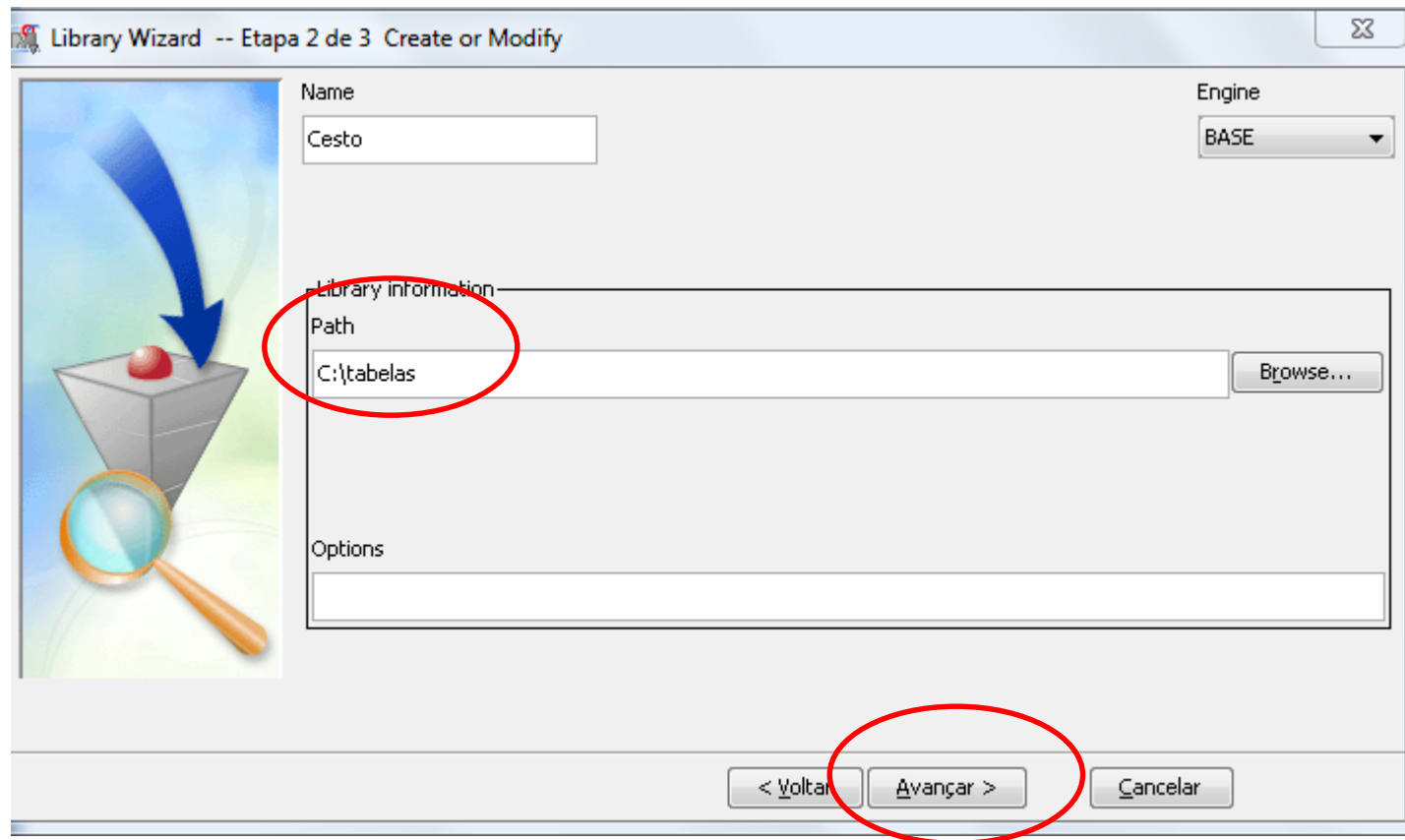
Name: Cesto Engine: BASE

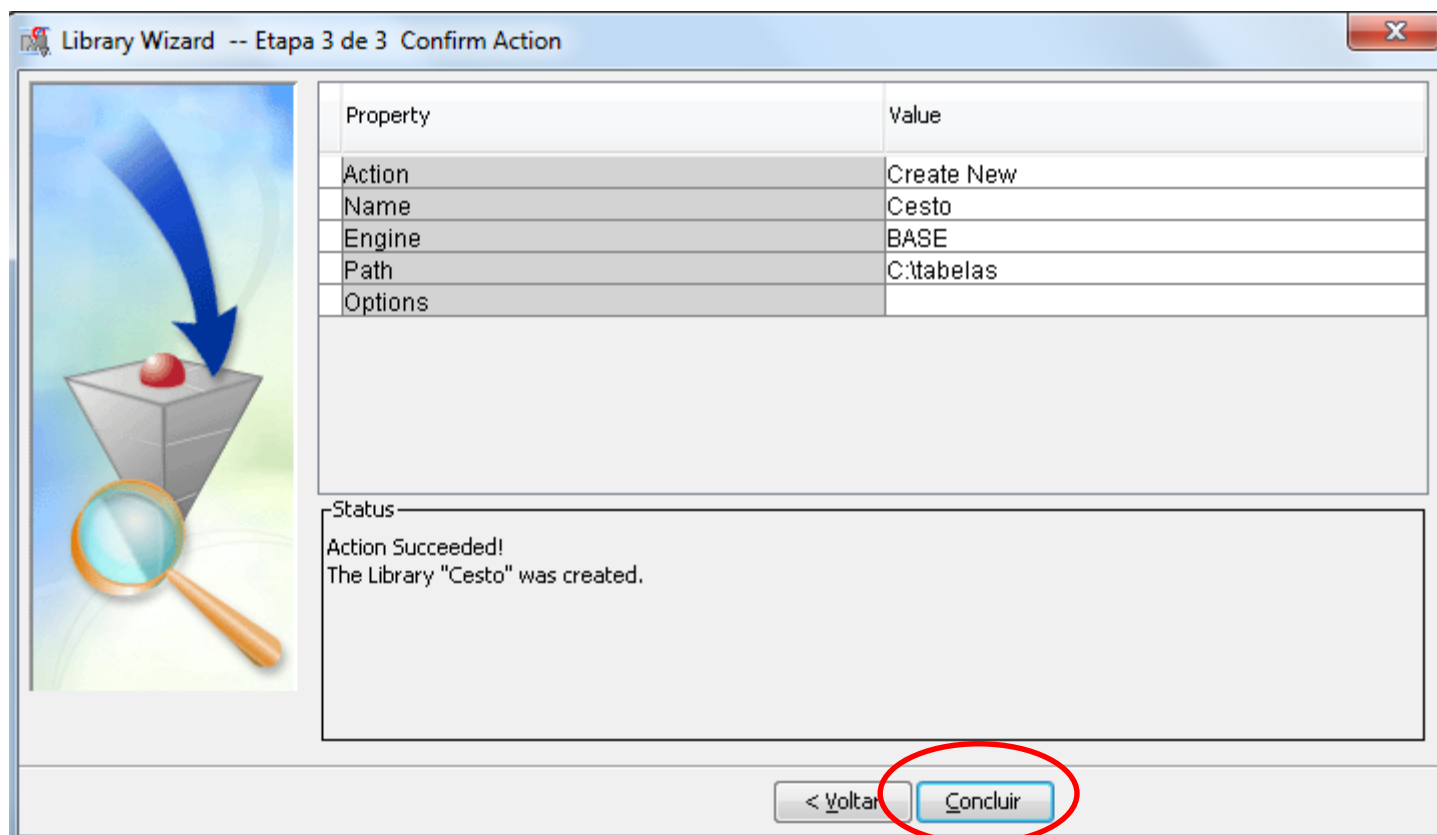
Library information

Path: C:\tabelas Browse...

Options:

< Voltar Avançar > Cancelar

The image shows a 'Library Wizard' dialog box, specifically 'Etapa 2 de 3 Create or Modify'. It has a sidebar on the left with a graphic of a cube and a magnifying glass. The main area contains fields for 'Name' (Cesto), 'Engine' (BASE), 'Path' (C:\tabelas), and 'Options'. The 'Path' field and the 'Avançar >' button at the bottom are circled in red. The title bar indicates it's a 'Library Wizard' window.



Enterprise Miner - ESTUDO DE CASO

File Edit View Actions Options Window Help

Sample Explore Modify Model Assess Utility Text Mining

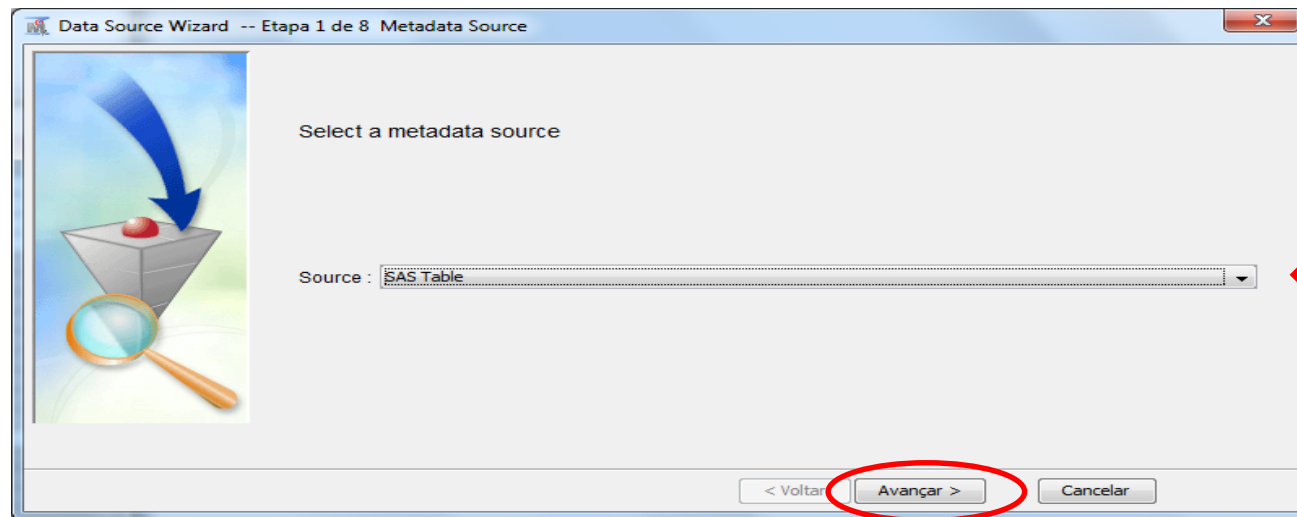
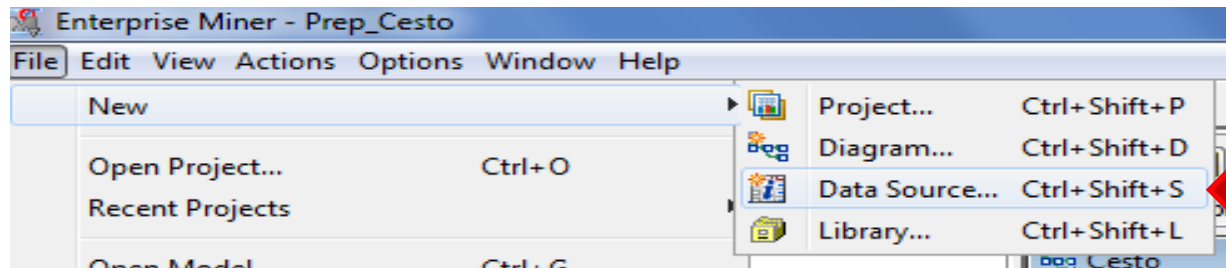
ESTUDO DE CASO

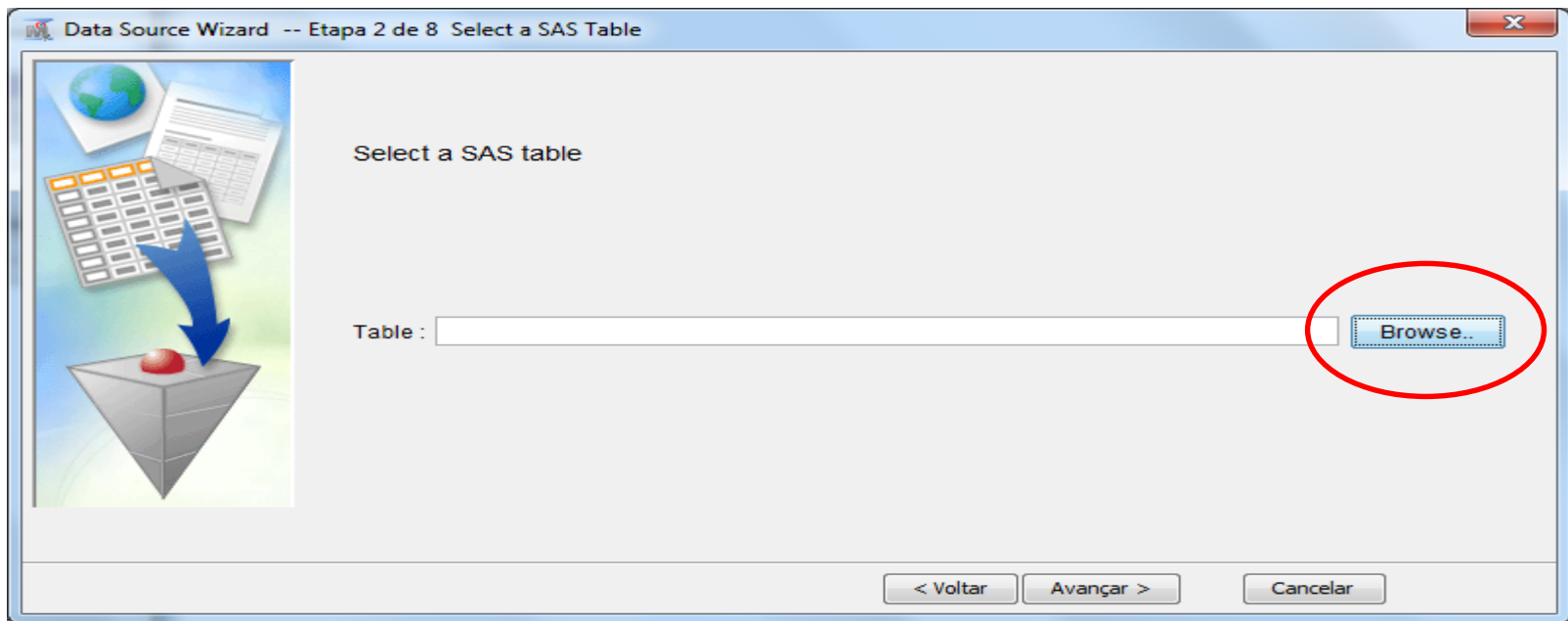
- Data Sources
- Diagrams
 - Exemplo1
- Model Packages
- Users

Property	Value
ID	EMWS2
Name	Exemplo1
Status	Open
Notes	...
History	...

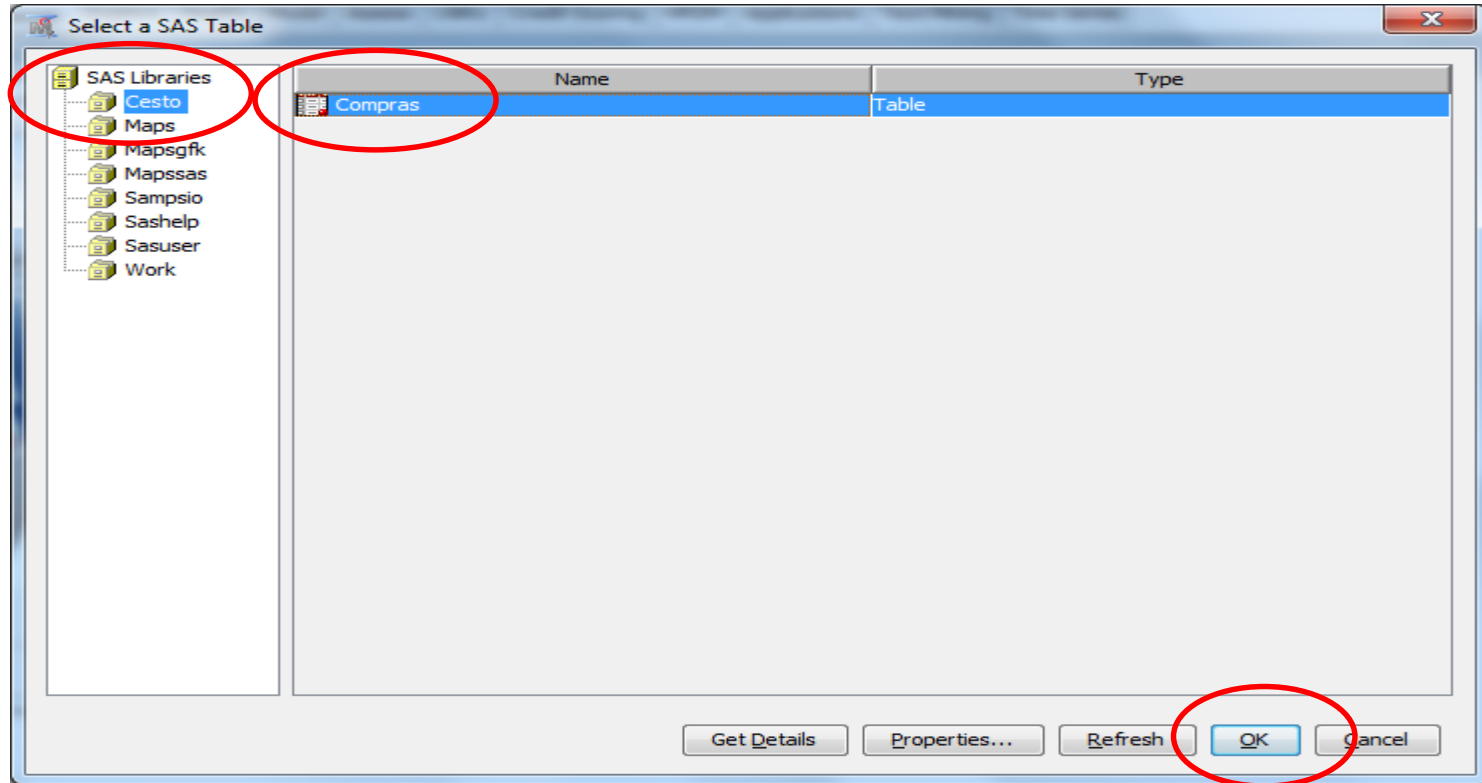
Exemplo1

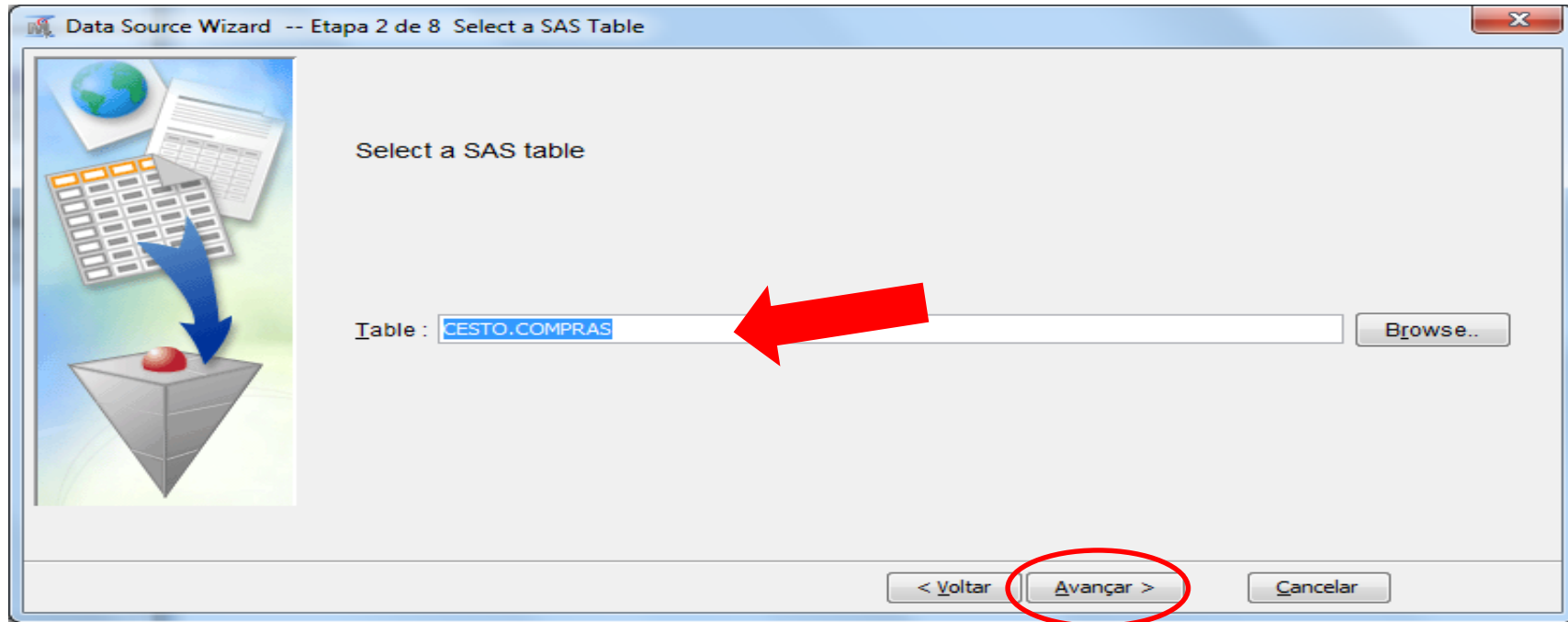
Criação de um Data Source para a tabela **Compras.sas7bdat**





Clicar em Cesto , clicar em Compras e clicar em OK



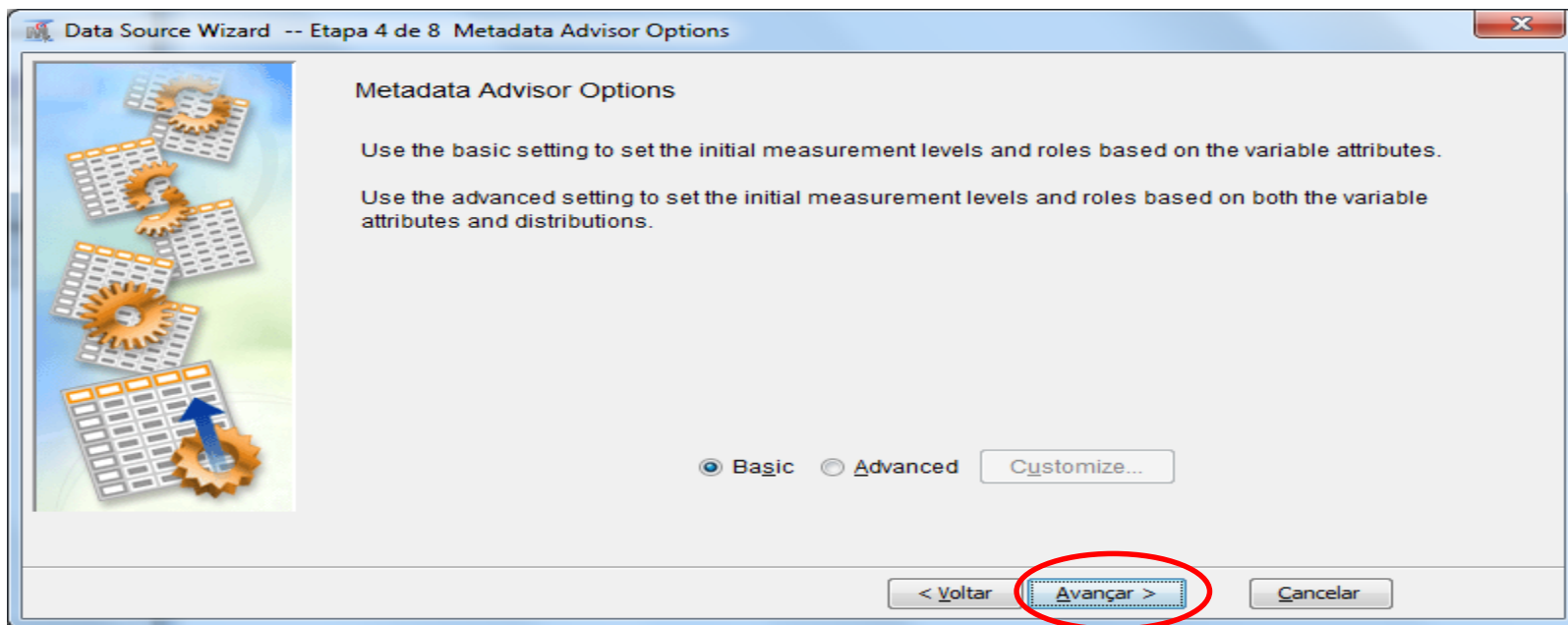


Data Source Wizard -- Etapa 3 de 8 Table Information

Table Properties

Property	Value
Table Name	CESTO.COMPRAS
Description	
Member Type	DATA
Data Set Type	DATA
Engine	BASE
Number of Variables	2
Number of Observations	459258
Created Date	4 de Junho de 2013 12h23min16s BRT
Modified Date	4 de Junho de 2013 12h23min16s BRT

< Voltar Avançar > Cancelar




Data Source Wizard -- Etapa 5 de 8 Column Metadata

(nenhum) ☐ não Igual a ...

Columns: ☐ Label ☐ Mining ☐ Basic ☐ Statistics

Name	Role	Level	Report	Order	Drop	Lower Limit	Upper Limit
produto	Input	Nominal	No		No	.	
transacao	Input	Interval	No		No	.	

Data Source Wizard -- Etapa 6 de 8 Create Sample



Do you wish to create a sample data set?

☒ No ☐ Yes

Table Info

Columns 2

Rows 459258

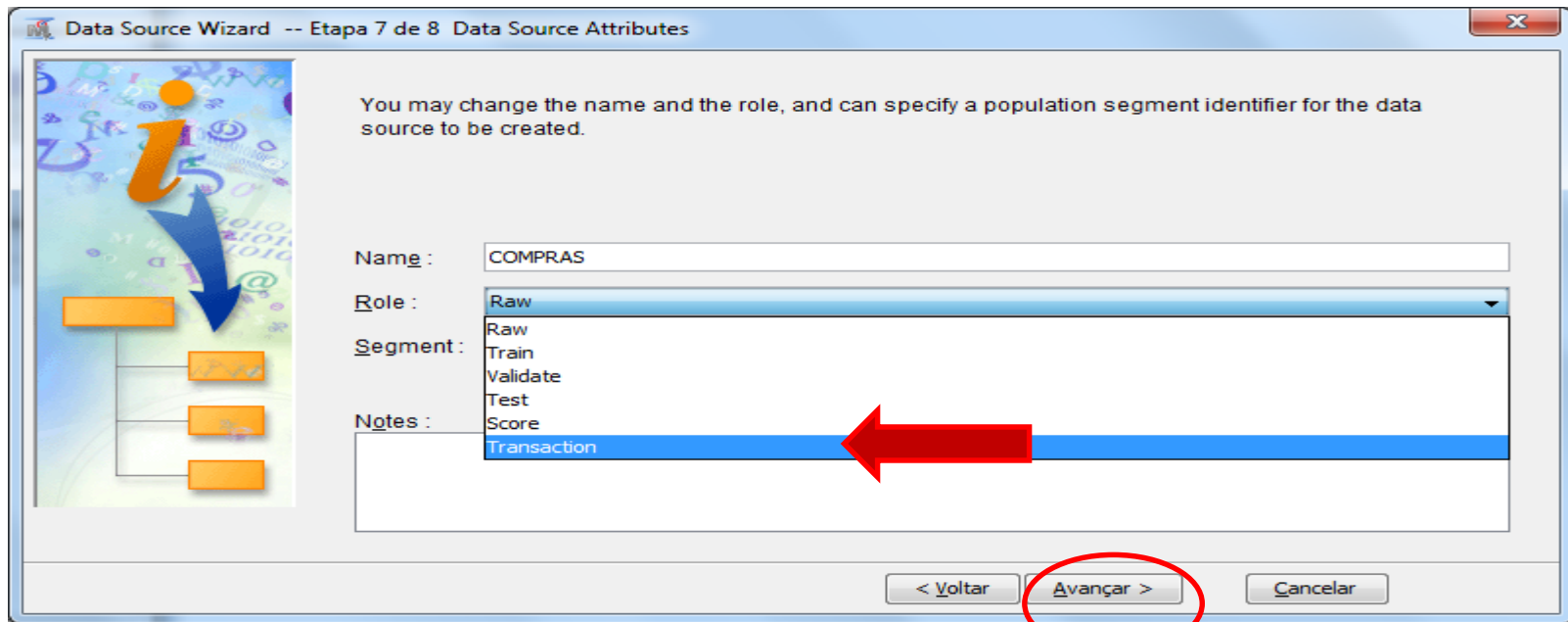
Sample Size

Type

Percent

Rows

< Voltar **Avançar >** Cancelar



Atenção! O número de passos a serem concluídos podem mudar de acordo com as opções e configurações definidas nas telas.

Data Source Wizard -- Step 8 of 9 Data Source Attributes

You may change the name and the role, and can specify a population segment identifier for the data source to be created.

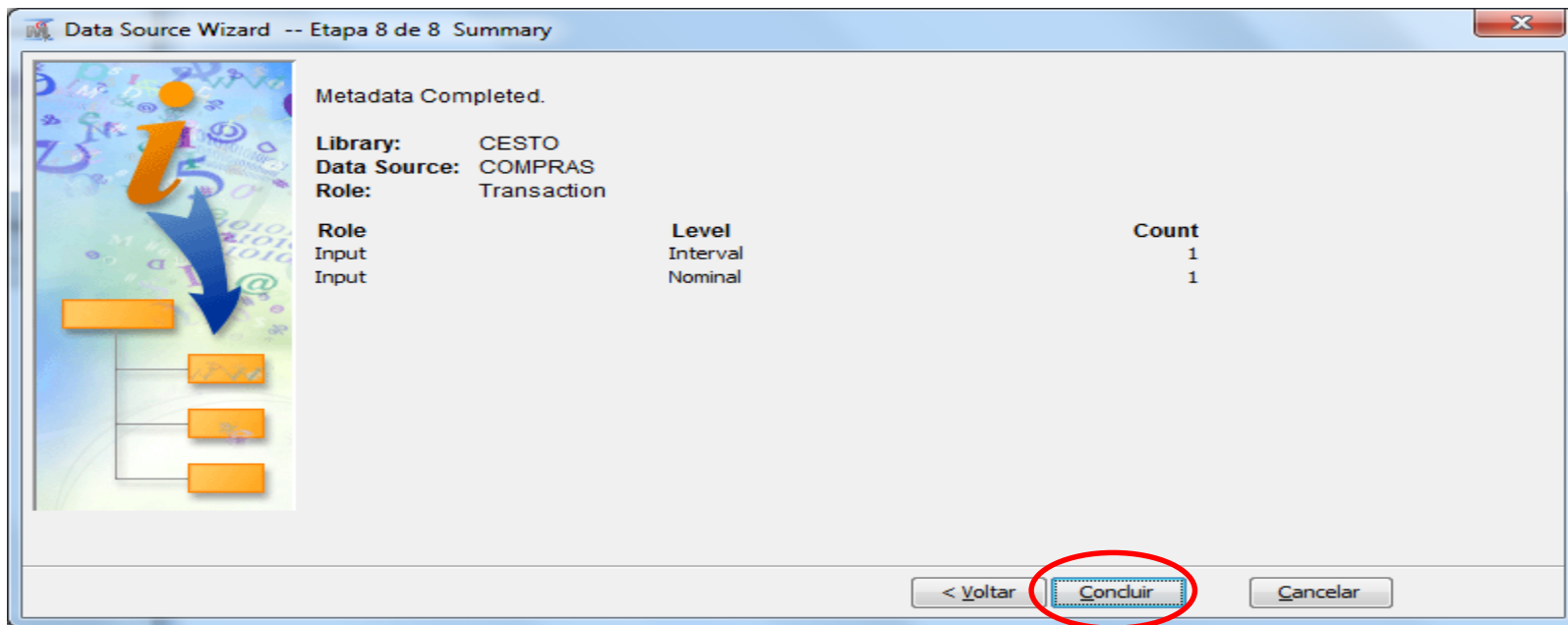
Name :

Role :

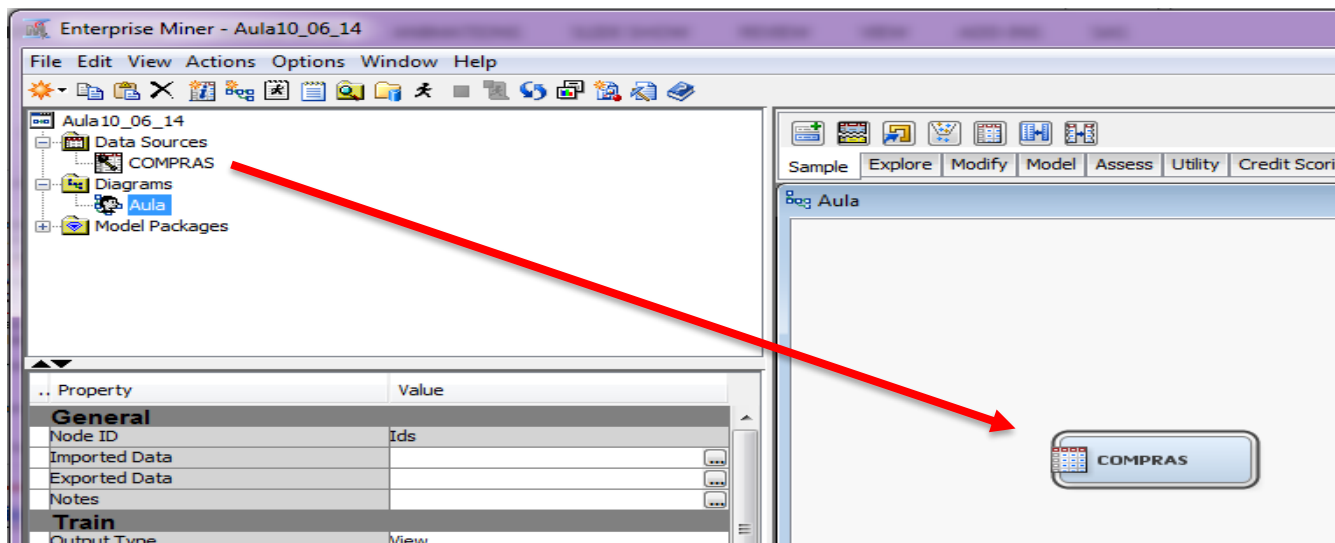
Segment :

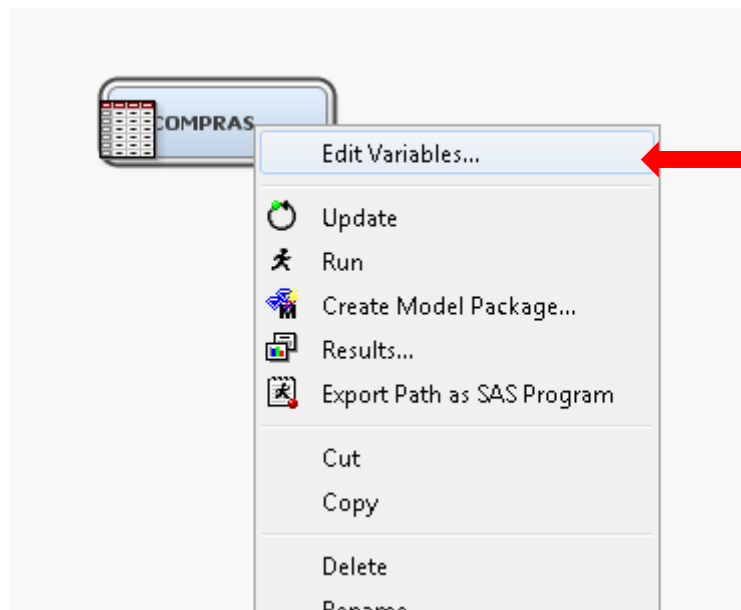
Notes :

< Back **Next >** Cancel



Clicar e arrastar com o mouse exatamente como a seta vermelha indica.





Clique com o botão direito em compras e depois em “Edit Variables”

Variables - Ids

(nenhum) ☐ não Igual a ...

Columns: ☐ Label ☐ Mining ☐ Basic ☐ Statistics

Name	Role	Level	Report	Order	Drop	Lower Limit	Upper Limit
produto	Input	Nominal	No		No	.	.
transacao	Input	Interval	No		No	.	.

Explore... OK Cancel

Variables - Ids

(nenhum) ☐ não Igual a

Columns: ☐ Label ☐ Mining ☐ Basic ☐ Statistics

Name	Role	Level	Report	Order	Drop	Lower Limit	Upper Limit
produto	Target	Nominal	No		No	.	.
transacao	Label	Interval	No		No	.	.

Prediction
Referrer
Rejected
Residual
Segment
Sequence
Target

Explore... OK Cancel

Variables - Ids

(nenhum) ☐ não Igual a ...

Aplicar Redefinir

Columns: ☐ Label ☐ Mining ☐ Basic ☐ Statistics

Name	Role	Level	Report	Order	Drop	Lower Limit	Upper Limit
produto	Target	Nominal	No		No	.	.
transacao	ID	Interval	No		No	.	.

Assessmen
Censor
Classificati
Cost
Cross ID
Decision
Frequency
ID

Explore... OK Cancel

Variables - Ids

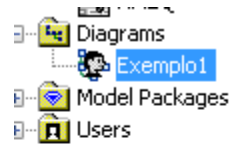
(nenhum) ☐ não Igual a ...

Columns: ☐ Label ☐ Mining ☐ Basic ☐ Statistics

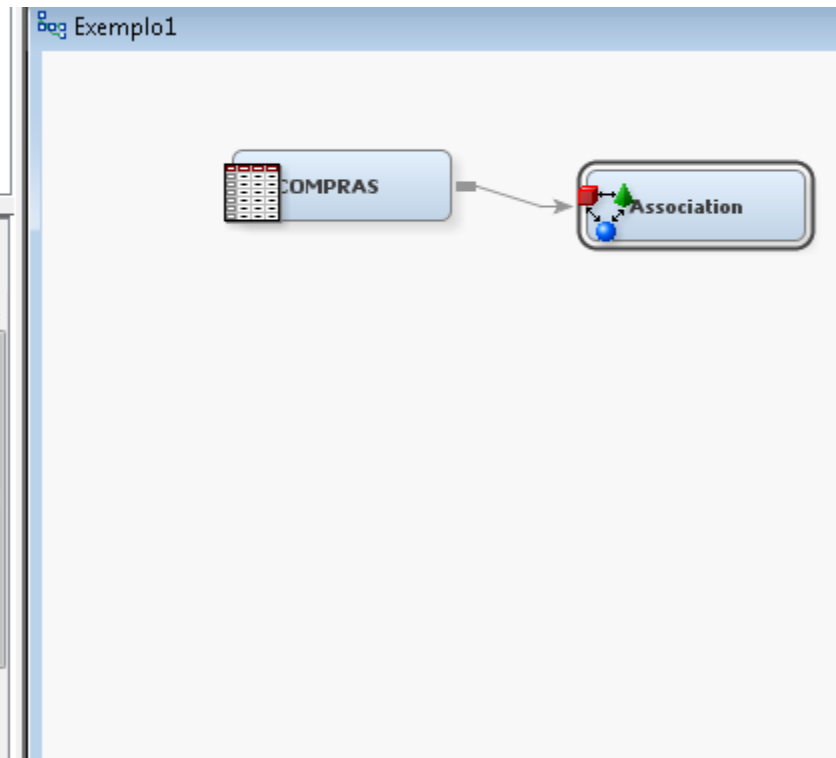
Name	Role	Level	Report	Order	Drop	Lower Limit	Upper Limit
produto	Target	Nominal	No		No	-	-
transacao	ID	Interval	No		No	-	-

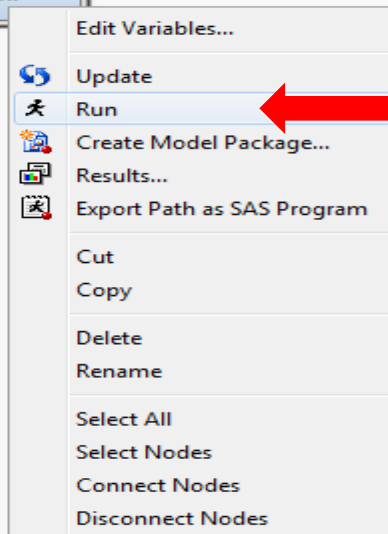
Explore... **OK** Cancel





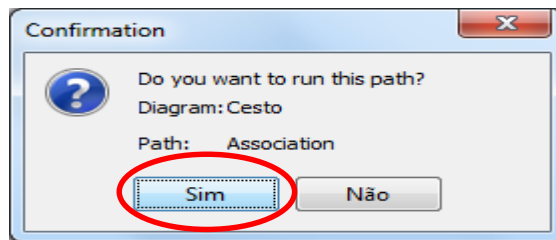
Property	Value
General	
Node ID	Assoc
Imported Data	...
Exported Data	...
Notes	...
Train	
Variables	...
Maximum Number of Item	100000
Rules	...
Association	
Maximum Items	2
Minimum Confidence Level	0
Support Type	Percent
Support Count	1
Support Percentage	5.0



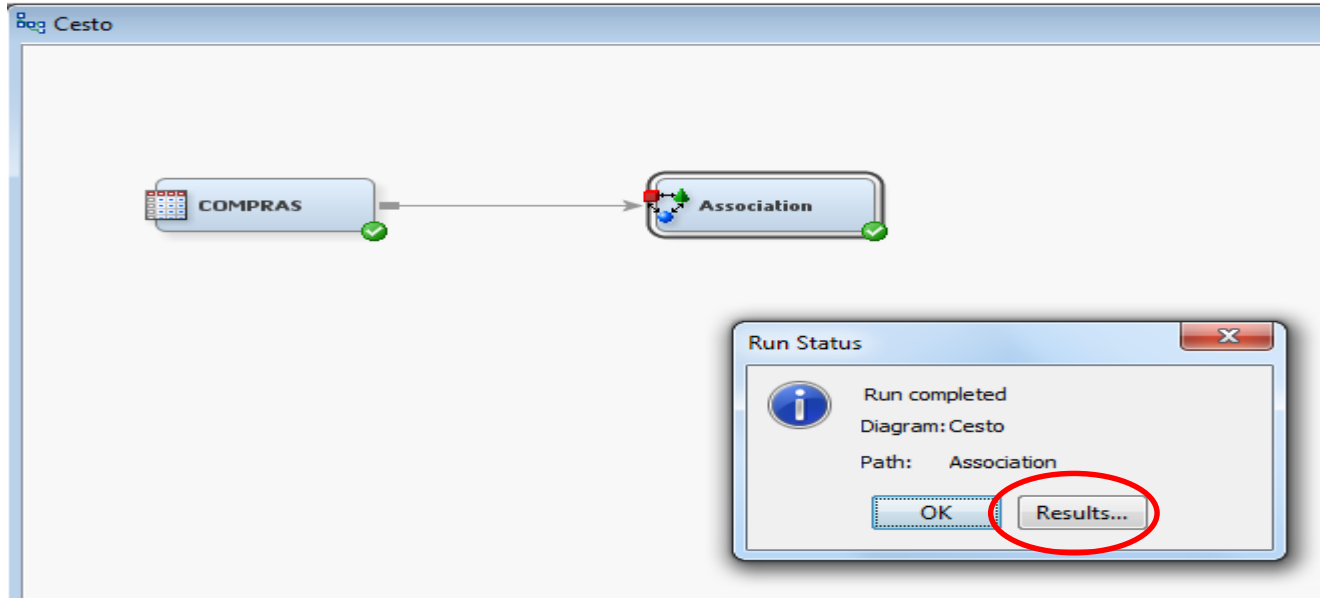


Clique com o botão direito em Association e depois em “Run”

Depois, na caixa de mensagem seguinte, clique em “sim”



Após a execução, aparecerá uma caixa de mensagens informando o final do processamento. Caso o processamento tenha terminado com sucesso, aparecerão símbolos verdes nos componentes. Após isso, clique em “Results”



- 1) Quantas regras de 2 itens passaram pelos filtros?
- 2) Qual é a regra filtrada com a maior aversão?
- 3) Qual é a regra filtrada de maior atração ?
- 4) Faça o Link Graph. Quem compra Magazine e Greeting Cards também compra qual produto ?
- 5) Qual produto você colocaria perto do Perfume ?

Rode a análise para regras de até 3 itens.

- 1) Quantas regras de 3 itens passaram pelos filtros?
- 2) Quantas regras possuem lift maior do que 1 ?
- 3) Qual é a regra filtrada com a maior aversão?
- 4) Qual é a regra filtrada de maior atração ?
- 5) Faça o Link Graph. Quem compra Magazine e Greeting Cards também compra qual produto ?
- 6) Qual produto você colocaria perto de Toothbrush ?
- 7) Analisando o Link Graph qual é o item central do gráfico? O que isso representa?

Rode a análise considerando regras de até 4 itens.

1) Comente a regra com maior Lift

2) Comente a regra com menor Lift

Exercício



Um executivo de uma instituição financeira deseja oferecer cestas de produtos financeiros.

Ajuste uma análise de cesto de compras para a tabela: **prod_financei.sas7bdat**

Produto	Número de Transações
Ações	2517
Capitalização	8614
Cartão de Crédito	2157
Cheque Especial	3254
Consortio	5336
Conta Corrente	9233
Ouro	970
Previdencia	4799
Renda Fixa	3561
Renda Variável	740
Seguro Auto	4864
Seguro Residência	8755
Seguro Vida	5500

Variables - Ids

(nenhum) ☐ não Igual a ...

Columns: ☐ Label ☐ Mining ☐ Basic ☐ Statistics

Name	Role	Level	Report	Order	Drop	Lower Limit	Upper Limit
produto	Target	Nominal	No		No	-	-
transacao	ID	Interval	No		No	-	-

Explore... OK Cancel

- 1) Quantas regras com 2 produtos foram geradas ?
- 2) Quantas regras com 3 produtos foram geradas ?
- 3) Qual é a regra filtrada de maior atração ?
- 4) Qual produto você pode sugerir para um cliente que possui ações ?
- 5) Qual produto você pode sugerir para um cliente que possui seguro de auto e residência ?
- 6) Qual produto você pode sugerir para um cliente que possui seguro de auto e capitalização ?
- 7) Os clientes que possuem conta corrente e previdência compram mais algum produto ?