



Instituto Federal de Educação, Ciência e Tecnologia da Bahia
Curso de Tecnologia em Análise e Desenvolvimento de Sistemas
Prof.: Grinaldo Lopes de Oliveira
Disciplina: Banco de Dados 2

Atividade de Laboratório

Membros da Dupla: _____

Objetivo do Laboratório:

Fixar conceitos relativos à **Mineração de Dados** aprendido em sala de aula através de uma abordagem prática utilizando o programa WEKA escrito em linguagem JAVA. Neste laboratório teremos a oportunidade de analisar uma base de dados de pedidos e tirar algumas conclusões com o WEKA;

INICIANDO OS MOTORES



Iniciando as Atividades

Este laboratório deverá ser feito **OBRIGATORIAMENTE** em dupla. Iniciaremos nossas atividades praticando nossos conhecimentos através de um programa que procura representar o conhecimento acerca de uma árvore genealógica.

Inicialmente, baixe o WEKA-COMPACTO no site do MOODLE, execute o programa **rode.bat**. Pressione o botão EXPLORER logo em seguida.



Este laboratório será feito com o banco de dados NorthWind que acompanha o banco de dados Excel.

Baixe o arquivo NWIND.MDB que está zipado no moodle. Descompacte-o e abra-o.

Procure entender o banco de dados. Quais tipos de informação ele possui?

Abra a consulta **Pedidos Trimestrais por Produto**

Como você entende esta consulta?

Vamos realizar uma mineração de dados que estude o comportamento de vendas ao longo de trimestres a partir de um algoritmo de associação.

Copie e cole todo o conteúdo da consulta em uma planilha excel. INCLUINDO OS TITULOS DOS CAMPOS.

Vamos substituir as ocorrências de valores em cada trimestre pelas palavras FEZ e NAO-FEZ, onde FEZ significa que houve pedido em um trimestre (célula com valores) e NÃO-FEZ o inverso.

Para isto, crie mais 4 colunas na planilha e com os nomes TRIM1, TRIM2, TRIM3 e TRIM4. Você terá a seguinte disposição de linhas:

Nome do Produto	Cliente	Trim 1	Trim 2	Trim 3	Trim 4	Trim1	Trim2	Trim3	Trim4
Alice Mutton	Antonio Moreno Taquería		R\$702,00						

Associe a cada uma a seguinte fórmula a cada célula nova criada:

=SE(XX="";"nao-fez";"fez")

onde XX é a célula de mesmo nome à célula criada.

Você terá a seguinte disposição de células.

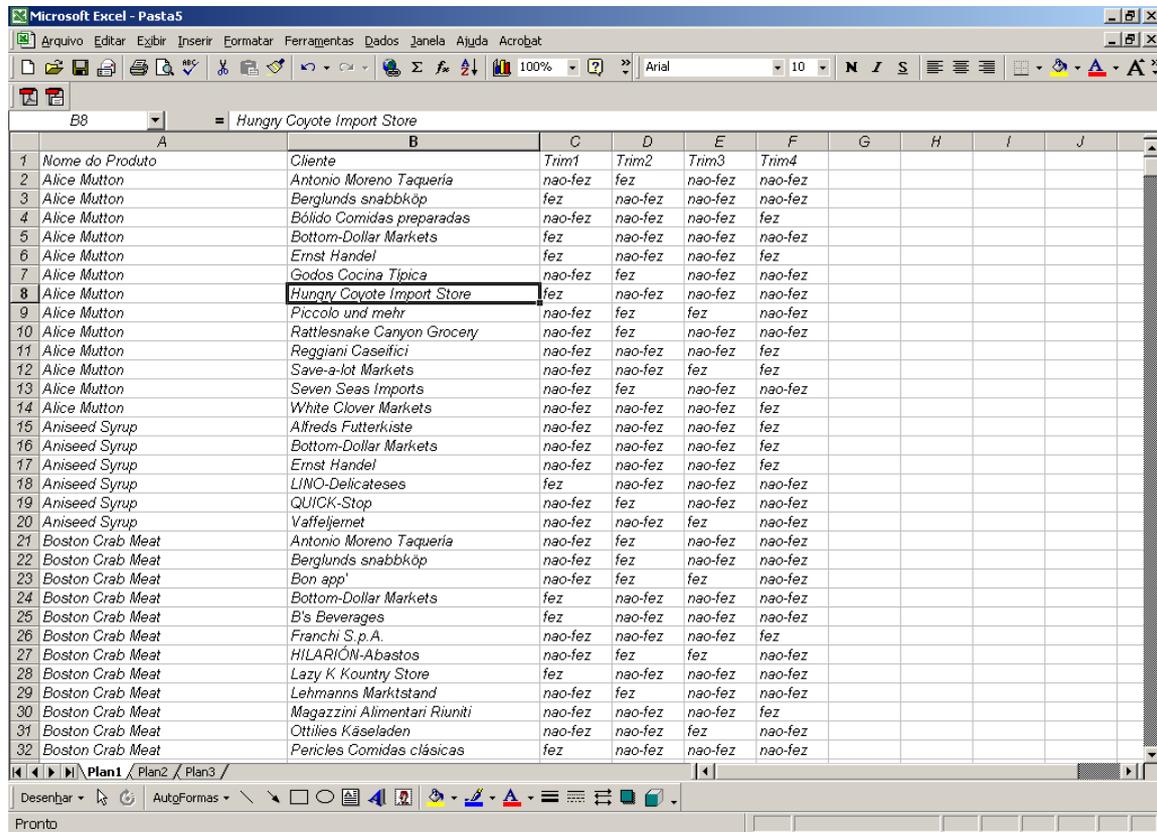
	A	B	C	D	E	F	G	H	I	J	K
1	Nome do Produto	Cliente	Trim 1	Trim 2	Trim 3	Trim 4	Trim1	Trim2	Trim3	Trim4	
2	Alice Mutton	Antonio Moreno Taquería		R\$702,00			nao-fez	fez	nao-fez	nao-fez	
3	Alice Mutton	Berglunds snabbköp	R\$312,00				fez	nao-fez	nao-fez	nao-fez	
4	Alice Mutton	Bólido Comidas preparadas				R\$1.170,00	nao-fez	nao-fez	nao-fez	fez	
5	Alice Mutton	Bottom-Dollar Markets	R\$1.170,00				fez	nao-fez	nao-fez	nao-fez	
6	Alice Mutton	Ernst Handel	R\$1.123,20			R\$2.607,15	fez	nao-fez	nao-fez	fez	
7	Alice Mutton	Godos Cocina Típica		R\$280,80			nao-fez	fez	nao-fez	nao-fez	
8	Alice Mutton	Hungry Coyote Import Store	R\$62,40				fez	nao-fez	nao-fez	nao-fez	
9	Alice Mutton	Piccolo und mehr		#####	R\$936,00		nao-fez	fez	fez	nao-fez	
10	Alice Mutton	Rattlesnake Canyon Grocery		R\$592,80			nao-fez	fez	nao-fez	nao-fez	
11	Alice Mutton	Reggiani Caseifici				R\$741,00	nao-fez	nao-fez	nao-fez	fez	

Uma vez feita a colocação de uma fórmula de uma linha, basta replicar a mesma para todas as células.

Vamos agora retirar as células que tem informação numérica dos trimestres.

Copie toda a planilha. Crie uma nova planilha. Cole ela de forma especial, usando o menu EDIT->COLAR ESPECIAL->VALORES.

Apague as células trimestrais que possui valores. Você terá a seguinte planilha.



Salve esta planilha como arquivo CSV – Separado por vírgulas. Dê o nome de NORTH-ASSOCIA.CSV. Responda OK e SIM às perguntas feitas.

Depois de salvo, feche a planilha e diga NÃO à pergunta feita.

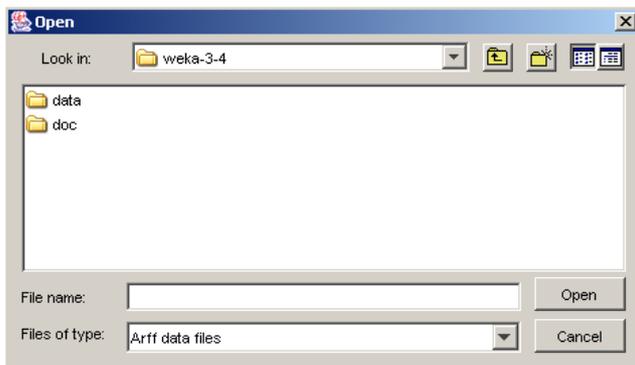
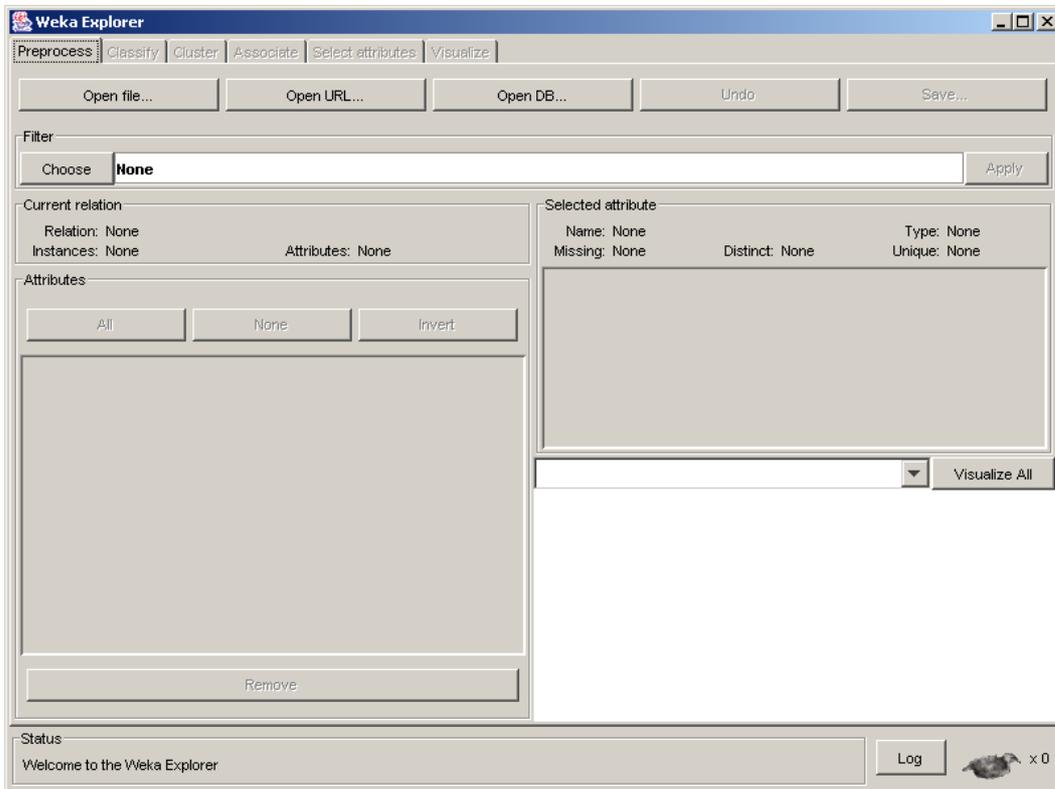
Abra o arquivo NORTH-ASSOCIA.CSV no WORDPAD.

Substitua os seguintes caracteres conforme ordem e tabela abaixo:

Original	Trocar por
Ponto (.)	Nada (deixar em branco)
Apóstrofo (')	Nada (deixar em branco)
Ponto e virgula (;)	Virgula (,)

Salve o novo arquivo modificado.

No WEKA, Clique no botão “Open file...”, abrirá uma janela de abertura de arquivo:



Escolha o tipo CSV DATA FILES na caixa de seleção **FILES OF TYPE**. Coloque a localização do arquivo NORTH-ASSOCIA.CSV e clique no botão “open”, a janela WEKA Explorer mostra os campos, marcamos todos os campos para serem minerados:

