KING SAUD UNIVERSITY.
CCIS.
IS DEPARTMENT.
T.A: Mr. Mourad Benchikh.

IS463 – Introduction to Data Mining
Project
2nd Semester 1438-1439

Given Date: Week 9
Due Date: Week11
The project will be carried out by a 3-students group using their assigned dataset (DataSetX.txt)

The dataset you have been assigned to analyze represents a part of items sold in hypermarket's transactions (available on the LMS: DataSets). Your group's assigned dataset contains your assigned number X.

1) Using excel, create a BAR chart (similar to the figure below) which will display the frequency distributions of items sold in transactions.
    - Because the dataset is huge, use Excel capabilities to help you in this task: the command "*Remove Duplicate*" (in the Data Tab), the Excel function *CountIF* which counts the number of cells that meet a certain condition, etc.
2) Load your assigned dataset into Weka and explore it.
3) Report any preprocessing you will make in order to prepare for a market-basket analysis.
4) Report the most important experiments you will make related to the setting of the Apriori's parameter values and how you will proceed to reach your optimal configuration.
    - We will consider an optimal configuration as the one which delivers some **FEW** strong rules.
5) Discuss the found association rules, obtained using your optimal parameter values, in terms of their significance or interesting associations (ex. uninteresting/interesting, un-useful/useful, unexpected/expected, unobvious/obvious, etc.)
6) Could you have predicted these results based only on the bar chart obtained in 1?
7) Suppose that you are requested to give some recommendations to the hypermarket board that may help them to increase their sales. What are the main recommendations you'll provide to them?

Note: You are requested to make a screen-shoot (when necessary) and highlight the important things to support your comments.

Deliverables:
- Report (hardcopy).
- You arff file saved with your optimal parameter values, as well as you Excel file, will be emailed to me (benchikhm@ksu.edu.sa) with an email subject containing your dataset#, ex. dataset#1.

	TID	Product1	Product2	Product3	Product4
1	1	meat	root vegetables	other vegetables	dessert
2	2	Halawa	Frozen beef slices	other vegetables	processed cheese
3	3	whole milk	brown bread		
4	5	dish cleaner	Halawa	frozen vegetables	
5	6	chicken	other vegetables	bottled water	chewing gum
6	7	Canned Pepsi	specialty chocolate		
7	8	frozen vegetables	frozen meals	shopping bags	
8	10	Ginger powder	chicken		

The items should NOT be duplicated

Item	Nb Occurrences
whole milk	1
Ginger powder	1
shopping bags	1
chicken	2
meat	1
specialty chocolate	1
other vegetables	3
frozen vegetables	2
bottled water	1
Canned Pepsi	1
root vegetables	1
Frozen beef slices	1
Halawa	2
chewing gum	1
dessert	1
brown bread	1
processed cheese	1
frozen meals	1
dish cleaner	1