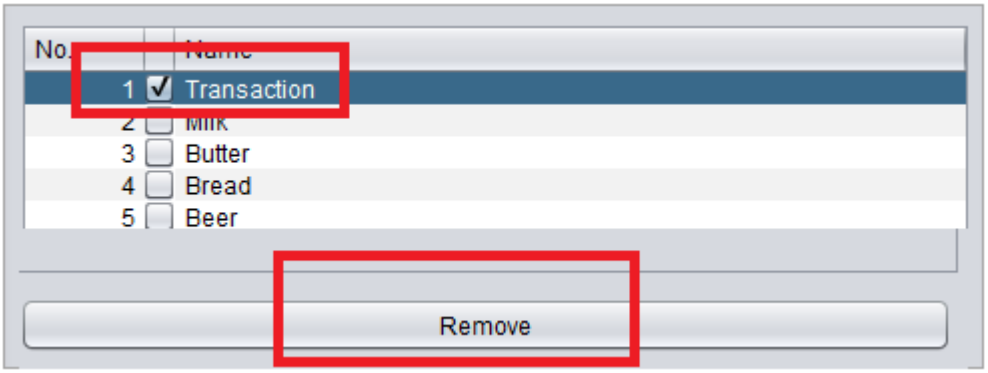
Association Rule Mining Using WEKA Explorer

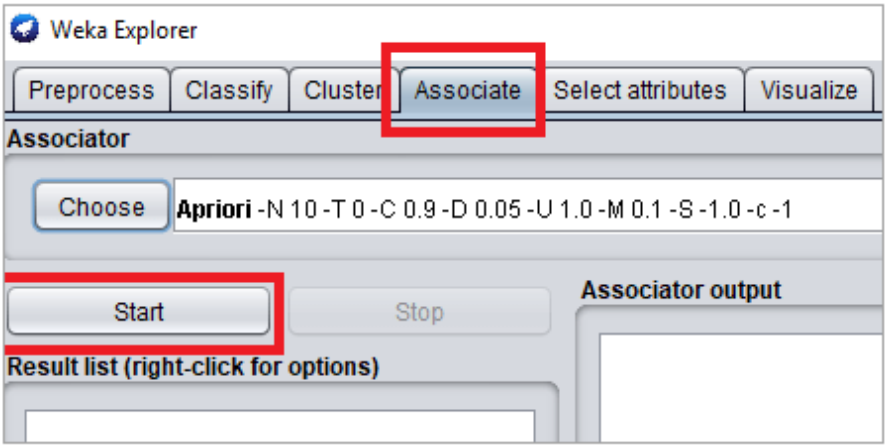
Implementation Using WEKA Explorer

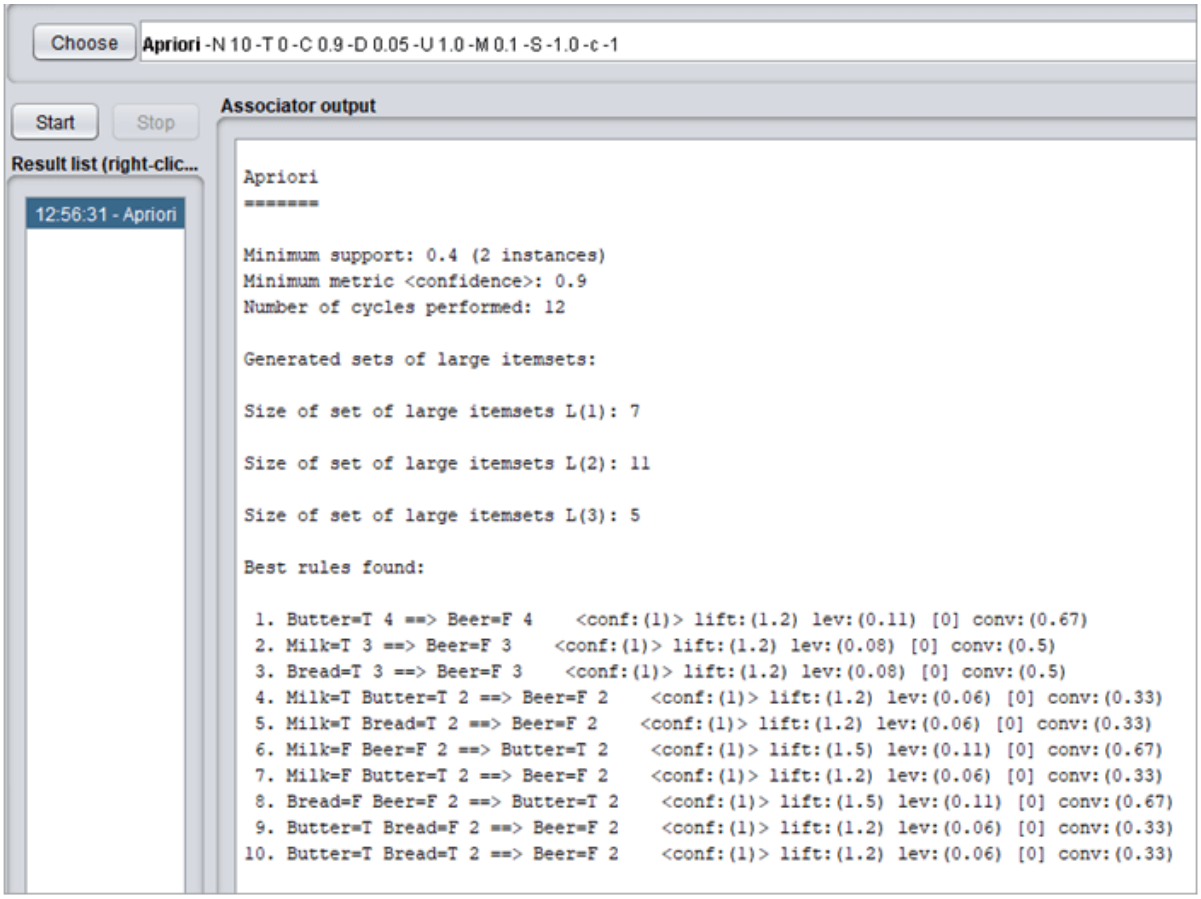
WEKA contains an implementation of the **Apriori algorithm** for learning association rules. Apriori works only with binary attributes, categorical data (nominal data) so, if the data set contains any numerical values convert them into nominal first.
Apriori finds out all rules with minimum support and confidence threshold.

Follow the steps below:
#1) Prepare an excel file dataset and name it as "**apriori.csv**"

	A	B	C	D	E
1	Transactic	Milk	Butter	Bread	Beer
2	1	T	T	F	F
3	2	F	T	T	F
4	3	F	F	F	T
5	4	T	T	T	F
6	5	F	T	F	F
7	6	T	F	T	F
8					






Choose
Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Start
Stop
Result list (right-clic...
12:56:31 - Apriori
Associator output
Apriori
=======
Minimum support: 0.4 (2 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 12
Generated sets of large itemsets:
Size of set of large itemsets L(1): 7
Size of set of large itemsets L(2): 11
Size of set of large itemsets L(3): 5
Best rules found:
1. Butter=T 4 ==> Beer=F 4 <conf:(1)> lift:(1.2) lev:(0.11) [0] conv:(0.67)
2. Milk=T 3 ==> Beer=F 3 <conf:(1)> lift:(1.2) lev:(0.08) [0] conv:(0.5)
3. Bread=T 3 ==> Beer=F 3 <conf:(1)> lift:(1.2) lev:(0.08) [0] conv:(0.5)
4. Milk=T Butter=T 2 ==> Beer=F 2 <conf:(1)> lift:(1.2) lev:(0.06) [0] conv:(0.33)
5. Milk=T Bread=T 2 ==> Beer=F 2 <conf:(1)> lift:(1.2) lev:(0.06) [0] conv:(0.33)
6. Milk=F Beer=F 2 ==> Butter=T 2 <conf:(1)> lift:(1.5) lev:(0.11) [0] conv:(0.67)
7. Milk=F Butter=T 2 ==> Beer=F 2 <conf:(1)> lift:(1.2) lev:(0.06) [0] conv:(0.33)
8. Bread=F Beer=F 2 ==> Butter=T 2 <conf:(1)> lift:(1.5) lev:(0.11) [0] conv:(0.67)
9. Butter=T Bread=F 2 ==> Beer=F 2 <conf:(1)> lift:(1.2) lev:(0.06) [0] conv:(0.33)
10. Butter=T Bread=T 2 ==> Beer=F 2 <conf:(1)> lift:(1.2) lev:(0.06) [0] conv:(0.33)

From the given data, they wish to find the items that were purchased most frequently. They also wish to determine the item(s) which encouraged the customer to purchase additional item(s). Such analysis is commonly termed as **Market Basket Analysis**, where the interesting associations between various items are determined. The analysis that leads to determining purchase behaviour of customers arises from the items attribute. The marketing team seeks to study the items attribute more closely to determine associations between various items.

From the given sample data set, the most frequently purchased item can be determined using a frequency table, as shown below

Transaction	Seat Cover	Audio system	Car cover	Steering Cover	Toolbox	Foot mats	Mud flaps	Window tint
1	Y		Y			Y	Y	
2	Y	Y		Y		Y		
3			Y					Y
4		Y					Y	Y
5					Y			
6	Y					Y		
7		Y			Y			Y
8							Y	
9				Y		Y		
10		Y			Y			Y
11			Y			Y		
12				Y				
13	Y				Y			
14						Y		
15			Y					
Total	**4**	**4**	**4**	**3**	**4**	**6**	**3**	**4**

3. Apply the FP_Growth Algorithm to the following dataset of transactions using WEKA.

1	Bread	Milk	Diaper	Beer	Eggs	Coke
2	Yes	Yes	?	?	?	?
3	Yes	?	Yes	Yes	Yes	?
4	?	Yes	Yes	Yes	?	Yes
5	Yes	Yes	Yes	Yes	?	
6	Yes	Yes	Yes	?	?	Yes

^^^^^^^^^^^^^