Market Basket Analysis and Algorithms

Q. 3a

{1,2,3,4}: merging {1,2,3} and {4}
{1,2,3,5}: merging {1,2,3} and {5}
{1,2,4,5}: merging {1,2,4} and {5}
{1,3,4,5}: merging {1,3,4} and {5}
{2,3,4,5}: merging {2,3,4} and {5}

Q. 3b

First, find all frequent 2-itemsets: {1,2},{1,3},{1,4},{1,5},{2,3},{2,4},{2,5},{3,4},{3,5},{4,5}
Then find their union sets: {1,2,3},{1,2,4},{1,2,5},{1,3,4},{1,3,5},{1,4,5},{2,3,4},{2,3,5},{2,4,5},{3,4,5}
Using Fk-1*F1 merging strategy: {1,2,3,4}, {1,2,3,5}, {1,2,4,5}, {1,3,4,5}, {2,3,4,5}

Q. 3c

For {1,2,3,4}, {1,2,3,5}, {1,2,4,5}, {1,3,4,5}, {2,3,4,5}, all of their supports are 1. If we assume the minimum support to be 1, then all of them survive.

Q. 4a

The distinct items in data sets:
{Beer, diapers, milk, bread, butter, cookies, eggs}
The number of distinct items: 7
The maximum number of association rules: 3^7 - 2^(7+1)+1 = 1932

Q. 4b

Confidence of {milk, diapers} => {butter} = 2/4 = 50%

Q. 4c

Support of {milk, diapers} => {butter} = 2/10 = 20%

Q. 4d

True. Because {a, b} is a subset of {a,b,c,d}, it always have no-less support than {a,b,c,d}. If support of {a,b,c,d} is higher than the threshold, then {a, b} must also be higher.

Q. 4e

False. {a,b,c} could have lower support than any of its subsets, which can make its support lower than the threshold value. So {a,b,c} is not certainly a frequent set.

Q. 4f

False. Given that {b} is a subset of {b,c}, its support must be higher than {b,c}, which is higher than 30.

Q. 4g

False. The maximum number would be 5*(5-1)/1*2 = 10.

Q. 4h