Association Rule Mining.

Trace the results of using the Apriori algorithm on the grocery store example with support threshold *s*=33.34% and confidence threshold *c*=60%. Show the candidate and frequent itemsets for each database scan. Enumerate all the final frequent itemsets. Also indicate the association rules that are generated and highlight the strong ones, sort them by confidence.

Transaction ID	Items
T1	HotDogs, Buns, Ketchup
T2	HotDogs, Buns
T3	HotDogs, Coke, Chips
T4	Chips, Coke
T5	Chips, Ketchup
T6	HotDogs, Coke, Chips

Solution:

Support threshold =33.34% => threshold is at least 2 transactions.

Applying Apriori

Pass (k)	Candidate k-itemsets and their support	Frequent k-itemsets
k=1	HotDogs(4), Buns(2), Ketchup(2), Coke(3), Chips(4)	HotDogs, Buns, Ketchup, Coke, Chips
k=2	{HotDogs, Buns}(2), ~~{HotDogs, Ketchup}~~(1), {HotDogs, Coke}(2), {HotDogs, Chips}(2), ~~{Buns, Ketchup}~~(1), ~~{Buns, Coke}~~(0), ~~{Buns, Chips}~~(0), ~~{Ketchup, Coke}~~(0), ~~{Ketchup, Chips}~~(1), {Coke, Chips}(3)	{HotDogs, Buns}, {HotDogs, Coke}, {HotDogs, Chips}, {Coke, Chips}
k=3	{HotDogs, Coke, Chips}(2)	{HotDogs, Coke, Chips}
k=4	{}	

Note that {HotDogs, Buns, Coke} *and* {HotDogs, Buns, Chips} *are not candidates when* k=3 *because their subsets* {Buns, Coke} *and* {Buns, Chips} *are not frequent.*

Note also that normally, there is no need to go to k=4 *since the longest transaction has only 3 items.*

All Frequent Itemsets: {HotDogs}, {Buns}, {Ketchup}, {Coke}, {Chips}, {HotDogs, Buns}, {HotDogs, Coke}, {HotDogs, Chips}, {Coke, Chips}, {HotDogs, Coke, Chips}.

Association rules:

{HotDogs, Buns} would generate: HotDogs → Buns (2/6=0.33, 2/4=0.5) and
Buns → HotDogs (2/6=0.33, 2/2=1);

{HotDogs, Coke} would generate: HotDogs → Coke (0.33, 0.5) and
Coke → HotDogs (2/6=0.33, 2/3=0.66);

{HotDogs, Chips} would generate: HotDogs → Chips (0.33, 0.5) and
Chips → HotDogs (2/6=0.33, 2/4=0.5);

{Coke, Chips} would generate: **Coke → Chips (3/6=0.5, 3/3=1)** and
Chips → Coke (3/6=0.5, 3/4=0.75);

{HotDogs, Coke, Chips} would generate: HotDogs → Coke ^ Chips (2/6=0.33, 2/4=0.5),
Coke → Chips ^ HotDogs (2/6=0.33, 2/3=0.66),
Chips → Coke ^ HotDogs (2/6=0.33, 2/4=0.5),
HotDogs ^ Coke → Chips(2/6=0.33, 2/2=1),
HotDogs ^ Chips → Coke(2/6=0.33, 2/2=1) and
Coke ^ Chips → HotDogs(2/6=0.33, 2/3=0.66).

With the confidence threshold set to 60%, the Strong Association Rules are (sorted by confidence):

1. Coke → Chips (0.5, 1)
2. Buns → HotDogs (0.33, 1);
3. HotDogs ^ Coke → Chips(0.33, 1)
4. HotDogs ^ Chips → Coke(0.33, 1)
5. Chips → Coke (0.5, 0.75);
6. Coke → HotDogs (0.33, 0.66);
7. Coke → Chips ^ HotDogs (0.33, 0.66)
8. Coke ^ Chips → HotDogs(0.33, 0.66).