Domas Budrys – Assignment 4 CSCI5080

Question 1 (6.6) (Time spent: 4h 30min):

A)

Apriori

C1

	count
M	3
O	4
N	2
K	5
E	4
Y	3
D	1
A	1
U	1
C	2
I	1

L1

	count
M	3
O	4
K	5
E	4
Y	3

C2

	count
MO	1
MK	3
ME	2
MY	2
OK	3
OE	3
OY	2
KE	4
KY	3
EY	2

L2

	count
MK	3
OK	3
OE	3
KE	4
KY	3

C3

	count
MOK	1
OKE	3
MKY	2
KYO	2

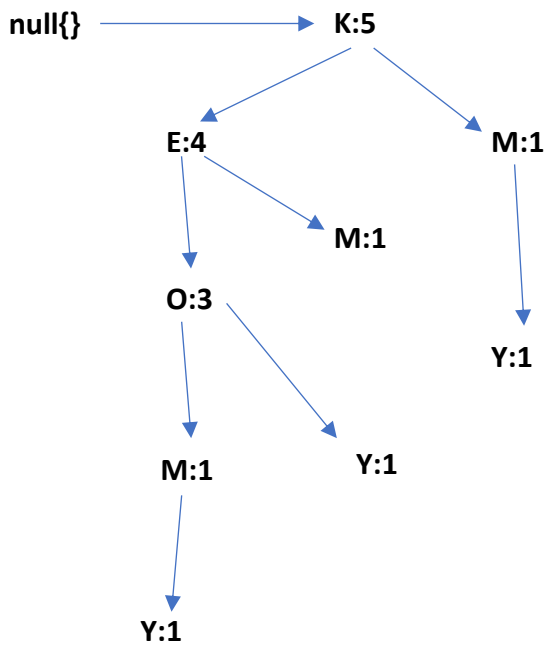
L3

	count
OKE	3

FP-Growth:

	Support
K	5
E	4
O	3
M	3
Y	3
~~N~~	~~2~~
~~C~~	~~2~~
~~D~~	~~1~~
~~A~~	~~1~~
~~U~~	~~1~~
~~I~~	~~1~~

	Support
T100	K, E, O, M, Y
T200	K, E, O, Y
T300	K, E, M
T400	K, M, Y
T500	K, E, O



Item	Conditional Pattern Base	Conditional FP Tree	Frequent Patterns
Y	[[K E O M:1], [K E O:1], [K M:1]]	[K:3]	[K,Y:3]
M	[[K E O:1], [K E:1], [K:1]]	[K:3]	[K,M:3]
O	[[K E:3]]	[K E:3]	[[K,O:3, [E,O: 3], [K,E,O:3]]]
E	[[K:4]]	[K:4]	[K E:4]
K	------	------	------

Comparison:

When these algorithms are applied to large databases, FP-Growth algorithm should perform better. It is because Apriori algorithm requires to create new candidate sets that are self-joined, which is a not computational efficient process. On the other hand, FP-Growth algorithm does not create any additional candidate sets.

B)

Buys (X, K) and Buys(X, O) $\Rightarrow$ Buys (X, E) [60%, 100%]
Buys (X, E) and Buys(X, O) $\Rightarrow$ Buys (X, K) [60%, 100%]

Question 2 (6.14) (Time spent: 1h 30min):

A)

$$Support(hot\ dogs \rightarrow hamburgers) = \frac{2000}{5000} = .4$$

$$Confidence(hot\ dogs \rightarrow hamburgers) = \frac{2000}{3000} = .667$$

Conclusion:

Association rule is strong because support and confidence percentage is above minimum threshold requirements

B)

$$P(hot\ dog) = \frac{3000}{5000} = .6$$

$$P(hamburger) = \frac{2500}{5000} = .5$$

$$P(hot\ dog, hamburger) = \frac{2000}{5000} = .4$$

$$lift(hot\ dog, hamburger) = \frac{P(hot\ dog,\ \ hamburger)}{P(hot\ dog)\ X\ P(hamburger)} = \frac{.4}{.6\ X\ .5} = \mathbf{1.33(3)}$$

Conclusion:

Since **1.33 > 1** we can conclude that buying hot dogs is not independent of buying hambugers and that there is a positive correlation between these two items.