CSC4008 Assignment5

118010045

(1) Given a transactions/items dataset in a supermarket, use a programming language that you are familiar with, such as Python, C++, and R language, to implement Apriori algorithm for mining frequent itemsets. Use it to find at least 10 most frequent 2-itemset in the supermarket dataset. Also explore at least 10 most frequent 3-itemset and 4-itemset. (60%)

frequent 2-itemsets	support
[' bread and cake ', ' milk-cream ']	0.5050788848065701
[' bread and cake ', ' fruit ']	0.5024854117138534
[' bread and cake ', ' vegetables ']	0.49665009725524095
[' vegetables ', ' fruit ']	0.4769829263021396
[' baking needs ', ' bread and cake ']	0.4735249621785174
[' bread and cake ', ' frozen foods ']	0.4601253511994813
[' biscuits ', ' bread and cake ']	0.4501837043440674
[' milk-cream ', ' fruit ']	0.44045818024637995
[' milk-cream ', ' vegetables ']	0.4376485843959369
[' baking needs ', ' vegetables ']	0.42122325480873135

frequent 3-itemsets	support
[' bread and cake ', ' vegetables ', ' fruit ']	0.38707585908796194
[' bread and cake ', ' milk-cream ', ' fruit ']	0.3639507240112384
[' bread and cake ', ' milk-cream ', ' vegetables ']	0.3583315323103523
[' baking needs ', ' bread and cake ', ' vegetables ']	0.3427706937540523
[' baking needs ', ' bread and cake ', ' milk-cream ']	0.34147395720769397
[' milk-cream ', ' vegetables ', ' fruit ']	0.3395288523881565
[' baking needs ', ' bread and cake ', ' fruit ']	0.33801599308407176
[' bread and cake ', ' frozen foods ', ' vegetables ']	0.33455802896044956
[' bread and cake ', ' frozen foods ', ' fruit ']	0.33455802896044956
[' biscuits ', ' bread and cake ', ' fruit ']	0.3330451696563648

frequent 4-itemsets	support
[' bread and cake ', ' milk-cream ', ' vegetables ', ' fruit ']	0.28333693537929544
[' bread and cake ', ' baking needs ', ' vegetables ', ' fruit ']	0.27123406094661767
[' bread and cake ', ' frozen foods ', ' vegetables ', ' fruit ']	0.2684244650961746
[' bread and cake ', ' biscuits ', ' vegetables ', ' fruit ']	0.26280527339528853
[' bread and cake ', ' baking needs ', ' milk-cream ', ' vegetables ']	0.2526475037821483
[' bread and cake ', ' baking needs ', ' milk-cream ', ' fruit ']	0.25091852172033713
[' bread and cake ', ' frozen foods ', ' biscuits ', ' fruit ']	0.24702831208126216
[' bread and cake ', ' frozen foods ', ' milk-cream ', ' fruit ']	0.24594769829263022
[' bread and cake ', ' frozen foods ', ' milk-cream ', ' vegetables ']	0.24443483898854548
[' bread and cake ', ' biscuits ', ' milk-cream ', ' fruit ']	0.242489734169008

(3) Use a programming language to implement FP-growth algorithm. Compare the performance of FP-growth algorithm with Apriori algorithm. Use it to find at least 10 most frequent 2-itemset in the supermarket dataset. Also explore at least 10 most frequent 3-itemset and 4-itemset. (40%)

```
frequent 2-itemset
{' bread and cake ', ' milk-cream '}                                0.5050788848065701
{' fruit ', ' bread and cake '}                                     0.5024854117138534
{' vegetables ', ' bread and cake '}                                0.49665009725524095
{' vegetables ', ' fruit '}                                         0.4769829263021396
{' baking needs ', ' bread and cake '}                              0.4735249621785174
{' frozen foods ', ' bread and cake '}                              0.4601253511994813
{' biscuits ', ' bread and cake '}                                  0.4501837043440674
{' fruit ', ' milk-cream '}                                         0.44045818024637995
{' vegetables ', ' milk-cream '}                                    0.4376485843959369
{' baking needs ', ' vegetables '}                                  0.42122325480873135

frequent 3-itemset
{' vegetables ', ' fruit ', ' bread and cake '}                     0.38707585908796194
{' bread and cake ', ' fruit ', ' milk-cream '}                     0.3639507240112384
{' bread and cake ', ' vegetables ', ' milk-cream '}                0.3583315323103523
{' baking needs ', ' vegetables ', ' bread and cake '}              0.3427706937540523
{' bread and cake ', ' baking needs ', ' milk-cream '}              0.34147395720769397
{' vegetables ', ' fruit ', ' milk-cream '}                         0.3395288523881565
{' baking needs ', ' fruit ', ' bread and cake '}                   0.33801599308407176
{' frozen foods ', ' fruit ', ' bread and cake '}                   0.33455802896044956
{' frozen foods ', ' vegetables ', ' bread and cake '}              0.33455802896044956
{' biscuits ', ' fruit ', ' bread and cake '}                       0.3330451696563648

frequent 4-itemset
{' bread and cake ', ' vegetables ', ' fruit ', ' milk-cream '}     0.28333693537929544
{' bread and cake ', ' baking needs ', ' fruit ', ' vegetables '}   0.27123406094661767
{' frozen foods ', ' vegetables ', ' fruit ', ' bread and cake '}   0.2684244650961746
{' biscuits ', ' vegetables ', ' fruit ', ' bread and cake '}       0.26280527339528853
{' bread and cake ', ' baking needs ', ' vegetables ', ' milk-cream '}  0.2526475037821483
{' bread and cake ', ' baking needs ', ' fruit ', ' milk-cream '}   0.25091852172033713
{' biscuits ', ' frozen foods ', ' fruit ', ' bread and cake '}     0.24702831208126216
{' frozen foods ', ' bread and cake ', ' fruit ', ' milk-cream '}   0.24594769829263022
{' frozen foods ', ' vegetables ', ' bread and cake ', ' milk-cream '}  0.24443483898854548
{' biscuits ', ' bread and cake ', ' fruit ', ' milk-cream '}       0.242489734169008
```