Data Mining Project 1

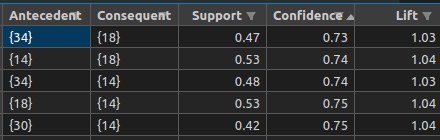
姓名：羅子翔

學號：p76114171

* High support, High confidence (support=0.4 (1200), confidence=0.7)

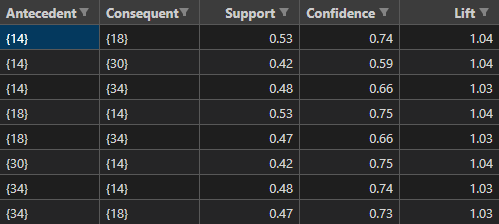
在high support的情況下，出現次數不足的item會在apriori scan的步驟被刪除，而在fp grouth裡就是在first scan的時候會被刪除，這會讓之後處理的步驟，如下次的scan或是在尋找pattern建立combination tree遞迴找frequent item set時大大的縮小時間。

而high confidence 雖然對運算的時間節省幫助不大，但是會對將要output的csv的資料再做篩選，以同一個association rule來說，antecedent item越多，則consequent item越少，confidence就會越高，所以在設定high confidence 的時候，output會比較多較長的antecedent的association rule。



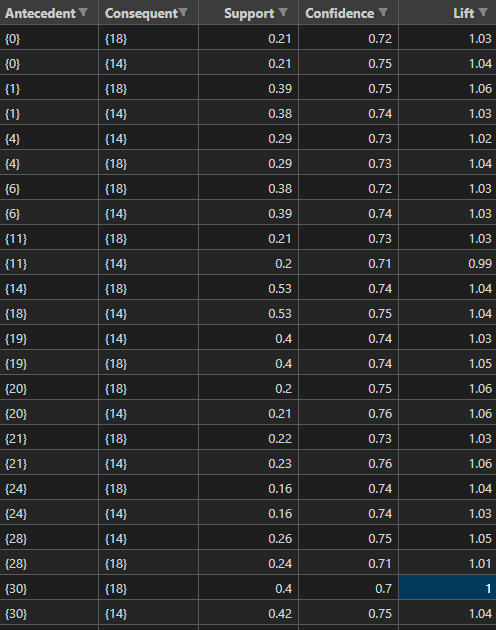
* High support , Low confidence (support=0.4 (1200), confidence=0.1)

因為是high support，所以計算的時間仍然很久，但在low confidence的情況下，會保留比較多大小較小的antecedent，這表示在條件機率的部分的標準比較寬鬆，類似{A} imply {B, C, D, E}的情形會越來越常見，但這不代表標準放的太過寬鬆，因為high support在事先已經先幫我們篩選掉不夠常出現的transaction 了。



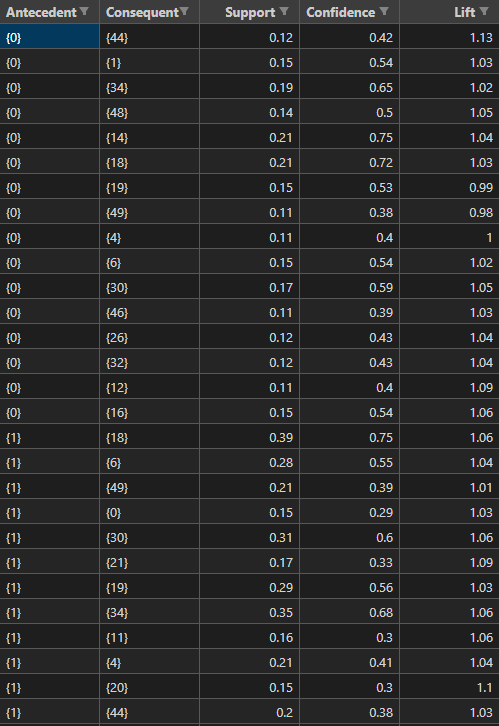
* Low support, High confidence (support=0.1 (297), confidence=0.7)

在Low support的情況下，無論是在apriori的scan，還是在fp grouth的first scan的步驟，都會保留很多的item，因此在運算的時間會成指數成長，我觀察在非常低的support情況下，apriori甚至速度會比fp grouth快上不少，在寫code的時候也有盡量注意不要在多餘地方使用for loop或遞迴，但結果還是不慎理想，我想這是因為在建立fp tree時因為保留了太多的item，導致tree太大棵，因此在後續利用header table尋找的pattern會太多，大部分的時候都在跑遞迴尋找frequent item set，而high confidence 會刪除掉一大部份的結果，所以可能最後的output數量其實不會跟花費的時間成正比。

 (共1107筆)

* Low support, Low confidence (support=0.1 (297), confidence=0.1)

因為是low support，時間花費同樣跟上點所述一樣久，但low confidence會使output的資料很多，因為被尋找到frequent item set已經很多了，他們會產生出更多的association rule，而且大部分的rule都不會被拒絕。

 (共10287筆)