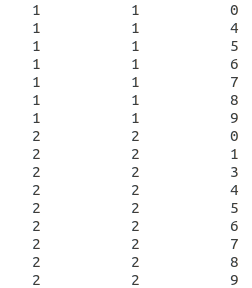
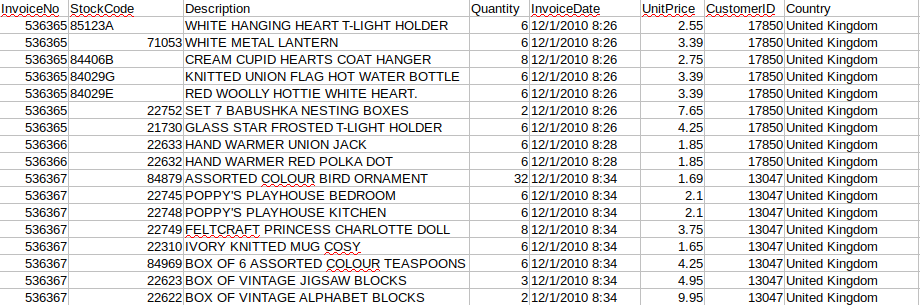
Data Mining Project 1 Report

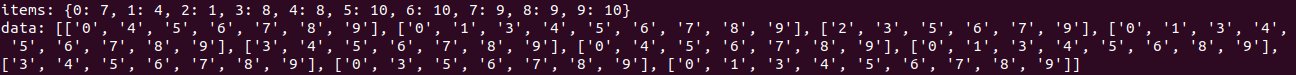
* Dataset
  + IBM Quest data Generator
    - IBM Quest data Generator lit -ntrans 0.01 -nitems 0.01
    - transactions：交易數 ex. 10
    - items ：物品種類 ex. 10
    - exapmle：



* + E-Commerce Data - Actual transactions from UK retailer from Kaggle
    - <https://www.kaggle.com/carrie1/ecommerce-data/home>
    - 跨國交易資料集，其中包含2010年1月12日到2011年12月9日期間在英國註冊的非商店在線零售業務中所發生的所有交易，該公司主要銷售各種禮品
    - example：

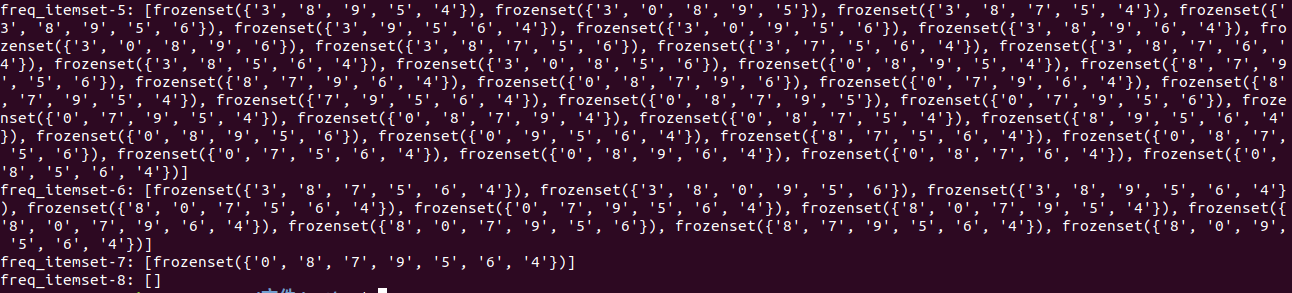


* Algorithm
  + Apriori Algorithm on IBM Quest data Generator

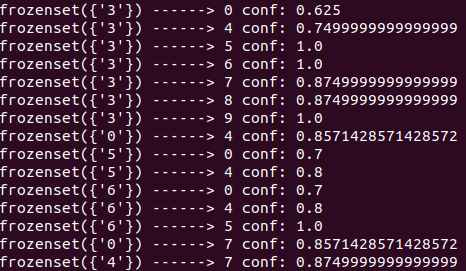


* items：儲存各item出現次數 ; data：所有itemset





* frequent itemset，minisupport：0.5



* rule，miniconf：0.6
  + FP Growth
* Experience

關聯法則在大數據分析上是一個很好用的工具，Apriori演算法為關聯法則中較經典的演算法，利用循序漸進的方式找出資料中項目之間的關係，關聯法則的生成可分為以下兩步驟：

1. 對資料進行預處理
2. 找關聯法則
   1. 得到frequent itemset
   2. 關聯規則生成

我在資料預處理的部份將所有品項種類出現的次數統計後保存，且將每筆itemset整理存入list中以方便後續找關聯，Apriori作為Association rule mining的代表演算法，概念相當直覺好懂，將出現大於minsupport的set留下，剔除較無關聯性的set，以此類推，雖然概念較簡單，但是在資料量很大時，效能上卻比不上其他演算法。

* Reference
  + <https://blog.csdn.net/u010665216/article/details/78470307>
  + <https://blog.xuite.net/coke750101/coketech/20478312-Apriori%E6%BC%94%E7%AE%97%E6%B3%95>