BU692A Individual Assignment 3: **Market Basket Analysis**

I use Market Basket Analysis as the key technique to uncover three promising product bundles I plan to promote on Black Friday. This dataset consists 537577 observations and 12 variables

I use split to create a list of basket with Produc\_ID and User\_ID and then use AS function to show the large transactions with a unique User\_ID and a subset of the products. Then, I use the function of itemFrequency to list the most frequently products and plot the top 10 products.

Next step, I will define a rule as an implication form X ⇒ Y

•X (left-hand-side or LHS) is called antecedent

•Y (right-hand-side or RHS) is called consequent

Before we generate the rules, we need to understand the definition and implications of association rules: Support, Confidence and Lift. Therefore, we can know **our criteria to promote three bundles.** The **Support** of one product is the number of transactions containing that product. Within the MBA macro, support can be expressed as a count or as a percentage of all transactions. Support of X supp( X)=n(X)/N(T); Support of X and Y supp(X ⇒ Y ) =n(X and Y)/N(T). **Confidence** is defined as the conditional probability that a transaction containing the LHS will also contain the RHS. Confidenc Formula: conf(X ⇒ Y ) = supp(X ∪ Y )/supp(X). **The ideal confidence is above 80%,** which means when you buy X, you are very likely to buy B. If Confidence is 1, then it indicates that people who buy A must buy B. **Lift** is a measure of the improvement in the occurrence of the RHS given the LHS: it is the ratio of the conditional probability of the RHS given the LHS, divided by the unconditional probability of the RHS. A **lift** greater than 1 suggests that the presence of the antecedent increases the chances that the consequent will occur in a given transaction.Lift below 1 indicates that purchasing the antecedent reduces the chances of purchasing the consequent in the same transaction. So, **we are usually interested in a lift value greater than one.**

Therefore, I set 0.001 as Min Support, 0.8 as Min Confidence. The length of rules is 2860576. **I would like to raise Min Support to be 0.003 (LHS and RHS appear in 0.3% of the transactions) and keep the same value of Confidence,** then length of rules significantly was decreased to 8583. Next, I sort out rules by confidence and create a list of 'high-confidence' rules. I found that the top 7 rules have confidence of 1, so I would not recommend those products bundles, just like what I have explained, people who buy A must buy B, so there is no need to promote them. And, the remaining 3 confidence levels are below 1. Besides, I **create a list of ‘high-lift ‘rules** and show up the top 3 rules by lift with support, confidence levels and lift levels, which all meet the criteria of bundles.

In the following, I scatterplot 1000 rules based on confidence and support dimension, also on confidence and lift dimension. From the scatterplots, I can see confident levels are mostly scattered below 0.95, support levels are concentrated under 0.004, and most lift levels are below 20.

I plot the top rules based on the confidence levels which are all above 0.95. I don’t want to choose these products to bundle as even I don’t promote them, they still have the great probability to be purchased together.

Next, I plot the top 3 rules as below based on the Lift levels which I am satisfied with all of the following levels for bundles.

Table 1:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | LHS | RHS | Support | Confidence | Lift |
| 1 | {P00142142, P00180142} => | {P00180042} | 0.0031 | 0.8182 | 73.0289 |
| 2 | {P00129842, P00368042} => | {P00368142} | 0.0031 | 0.9000 | 27.4710 |
| 3 | {P00226642, P00324642} => | {P00329342} | 0.0031 | 0.8500 | 25.8945 |

Table 2:

A picture containing bird

Description automatically generated

Hence, I can choose these three most promising product bundles to promote - {P00142142, P00180142} =>{P00180042}, {P00129842, P00368042} =>{P00368142}, {P00226642, P00324642} =>{P00329342}, as they have top lift levels, good confidence levels (0.8-0.9) and reasonable support levels (0.0031).

(Note: I also tried the products of best sellers to find out related bundles which could meet the rules. Actually, I can also consider them to promote if I am allowed to promote more.)

|  |
| --- |
|  |
|  | |  |