IEMS 308

Jay Jung

11/1/2016

**Association Rule**

**Executive Summary**

Daniel Kahneman and Amos Tversky stated that the psychology of human being affects the economic behavior of people so that people’s economic behavior does not always follow the economic theory. In the book “Nudge”, economists from University of Chicago, Richard Thaler and Cass Sunstein, stated that stakeholders in economic behavior such as business owners should lead other partakers to the desired outcome by giving them nudge; this includes arranging products in supermarket in an effective way so that customers impulsively purchase other items as well.   
 Association rule is a data analysis and data mining technique that discovers co-occurrence relationships among activities performed by specific individual or groups. Market Basket Analysis is one of the practical uses of Affinity Analysis. It reveals affinities between individual products or product groupings.

Our client is a retailer having 453 stores over the United States. It offers 104876 SKUs. These days the company. The company is trying to relocate the product inside their stores to trigger impulsive purchase and increase the profit.

Out of 100 million transaction data provided by our client, about 20% of transaction data(24million data) were randomly collected by sampling transaction data from subset of stores. Due to the constraint of staff, only 20 moves and total 100 items can be rearranged in the store.

By performing association rules, 50 rules were found to address 100 potential candidate SKUs. For final 20 moves, 22 rules among SKUs were observed. Total 17 moves of items were expected. A lot of those relationship had high confidence and lift, which means that two purchases were not coincidental.

**Problem Statement**

‘Invisible Hand’ by Adam Smith does not always work correctly. Economic behavior is largely dependent on psychology of partakers. The company wants to increase the profit on their franchises without adding new SKU or having promotion. It rather wants to relocate the products within their store so that the store can have customer impulsively purchases more items. Due to budget constraint, only 20 moves can be made, and the client is interested in potential 100 items that can be rearranged in the store.

**Assumption**

* Consumer behaviors of all the region across the United States were assumed to be same. Since the transaction data set was extremely huge, transaction data from subset of overall stores were collected. Nevertheless, this data can represent the general consumer behavior because of this assumption
* Even if some purchased products were returned, consumers, at the beginning, had intentions to purchase the items. Thus, even if the product was eventually returned, that transaction was considered for analysis.
* The focus on the move was more about the odd of impulsive purchase rather than profitability. Some relationship between(among) items is less likely but can trigger extra purchase with high profitability. However, the company’s business focuses on selling quantity of products rather than selling some high profitable products. Thus, we focused on the likelihood of relationship.
* The planogram and the interior of all the stores are similar. Since we are sampling subset of all stores, it should be assumed that the original planogram and interior of all stores are same. Thus, transaction data from sampled store can represent the affinity rules in other stores as well.

**Methodology**

The software ‘R’ was used for data analysis. Our first step to begin association rule was to select about 20%(24million data) from the transaction data set. The whole data set was loaded with function ‘fread’. Then transaction data from stores that had store number larger than 7500 were selected.

The transaction data were explored. After exploring data, we figured out the criteria for the same basket(basket ID). In order to consider multiple items to be purchased within the same basket, Transaction Number(Trannum), Register Number(Register), Sale Date, Store Number should be matched. Using “paste” function, SKUs from transactions that have certain Transaction Number(Trannum), Register Number(Register), Sale Date and Store Number were drawn. Afterward, we forced this data frame into transaction data where on the first column they have Transaction Number(Trannum), Register Number(Register), Sale Date and Store Number, which are the criteria for Basket ID, while, on the second column, there were lists of SKUs that belonged to that basket ID.

Then the association ‘arules’ package was used to find affinity between each item purchased. First, 100 candidate SKUs for potential moves was calculated. For 100 candidates, 50 rules were expected:50\*2. Larger support(0.1,0.01,0.001) hardly rendered any relation while smaller minimum supports increased the number of relationships. For 100 candidates, minimum support of 0.00009, confidence of 0.25 and minlen of 2 was used. For final 20 moves, minimum support of 0.0001 was used with Minimum confidence of 0.2 and Minlens of 2. As the chart goes down, the confidence suddenly dropped. Afterward, redundant relationship was pruned in order to curtail our findings. Too many relationships will complicate the understanding of the data.

Discovered relationship was mapped in order to increase the understanding of findings by clients. First, the data was visualized in terms of confidence, lift and support. Second, the data was mapped with arrows so that the client better understand where to locate some items close to each other to increase sales.

**Analysis**

**100 Candidates**

lhs rhs support

[1] {,"4462521"} => {,"4512521"} 9.280373e-05

[2] {,"4512521"} => {,"4462521"} 9.280373e-05

[3] {,"4142521"} => {,"4512521"} 9.036929e-05

[4] {,"4512521"} => {,"4142521"} 9.036929e-05

[5] {,"6490353"} => {,"6470353"} 9.044306e-05

[6] {,"6470353"} => {,"6490353"} 9.044306e-05

[7] {,"6642521"} => {,"6752521"} 9.391029e-05

[8] {,"6752521"} => {,"6642521"} 9.391029e-05

[9] {,"6032521",,"6062521"} => {,"6072521"} 9.147585e-05

[10] {,"6742521"} => {,"6752521"} 1.049021e-04

[11] {,"6752521"} => {,"6742521"} 1.049021e-04

[12] {,"6642521"} => {,"6742521"} 9.154962e-05

[13] {,"6742521"} => {,"6642521"} 9.154962e-05

[14] {,"6032521",,"6072521"} => {,"6062521"} 9.147585e-05

[15] {,"6062521",,"6072521"} => {,"6032521"} 9.147585e-05

[16] {,"6062521"} => {,"6072521"} 1.251154e-04

[17] {,"6072521"} => {,"6062521"} 1.251154e-04

[18] {,"6032521"} => {,"6072521"} 1.155989e-04

[19] {,"6072521"} => {,"6032521"} 1.155989e-04

[20] {,"6062521"} => {,"6032521"} 1.080005e-04

[21] {,"6032521"} => {,"6062521"} 1.080005e-04

[22] {,"5369905"} => {,"6349904"} 9.752506e-05

[23] {,"6349904"} => {,"5369905"} 9.752506e-05

[24] {,"5309905"} => {,"7029904"} 1.040169e-04

[25] {,"7029904"} => {,"5309905"} 1.040169e-04

[26] {,"5109905"} => {,"5749904"} 1.091071e-04

[27] {,"5749904"} => {,"5109905"} 1.091071e-04

[28] {,"5189905"} => {,"6949904"} 1.054923e-04

[29] {,"6949904"} => {,"5189905"} 1.054923e-04

[30] {,"2716578",,"3908011"} => {,"3988011"} 1.251154e-04

[31] {,"2716578",,"3988011"} => {,"3908011"} 1.251154e-04

[32] {,"2726578",,"3998011"} => {,"3908011"} 1.197301e-04

[33] {,"3908011",,"3988011"} => {,"2716578"} 1.251154e-04

[34] {,"7596135"} => {,"6656135"} 1.379515e-04

[35] {,"3690654",,"3898011"} => {,"3968011"} 1.614106e-04

[36] {,"2726578",,"3908011"} => {,"3998011"} 1.197301e-04

[37] {,"3908011",,"3998011"} => {,"2726578"} 1.197301e-04

[38] {,"3690654",,"3968011"} => {,"3898011"} 1.614106e-04

[39] {,"3908011"} => {,"3988011"} 1.690828e-04

[40] {,"3898011",,"3968011"} => {,"3690654"} 1.614106e-04

[41] {,"3524026",,"3978011"} => {,"3898011"} 2.253699e-04

[42] {,"3988011"} => {,"2716578"} 2.255913e-04

[43] {,"3908011"} => {,"2716578"} 1.604516e-04

[44] {,"3524026",,"3898011"} => {,"3978011"} 2.253699e-04

[45] {,"3908011"} => {,"3998011"} 1.721074e-04

[46] {,"3908011"} => {,"2726578"} 1.682713e-04

[47] {,"3898011"} => {,"3968011"} 3.884773e-04

[48] {,"3968011"} => {,"3898011"} 3.884773e-04

[49] {,"3898011",,"3978011"} => {,"3524026"} 2.253699e-04

[50] {,"3898011"} => {,"3978011"} 4.002069e-04

confidence lift

[1] 0.7443787 5225.48835

[2] 0.6514759 5225.48835

[3] 0.7020057 4928.03297

[4] 0.6343863 4928.03297

[5] 0.6637791 4763.28797

[6] 0.6490206 4763.28797

[7] 0.7080089 4690.81503

[8] 0.6221896 4690.81503

[9] 0.8469945 4599.93027

[10] 0.6882865 4560.14729

[11] 0.6950147 4560.14729

[12] 0.6902113 4528.63165

[13] 0.6006776 4528.63165

[14] 0.7913210 4462.03987

[15] 0.7311321 4284.84805

[16] 0.7054908 3831.43996

[17] 0.6794872 3831.43996

[18] 0.6774751 3679.28986

[19] 0.6278045 3679.28986

[20] 0.6089850 3568.99715

[21] 0.6329442 3568.99715

[22] 0.4242619 1704.02299

[23] 0.3917037 1704.02299

[24] 0.4554264 1579.31119

[25] 0.3607061 1579.31119

[26] 0.4377035 1547.94806

[27] 0.3858596 1547.94806

[28] 0.4203410 1338.17001

[29] 0.3358384 1338.17001

[30] 0.7797701 1038.73487

[31] 0.5546109 896.06936

[32] 0.5097362 823.56659

[33] 0.7399651 816.82325

[34] 0.4764331 809.91778

[35] 0.8284741 583.94197

[36] 0.7115300 494.39429

[37] 0.6956708 475.07097

[38] 0.6841776 449.49173

[39] 0.2731824 363.90731

[40] 0.4154956 349.87246

[41] 0.5156989 338.80437

[42] 0.3005110 331.72426

[43] 0.2592372 286.16344

[44] 0.7251365 233.07204

[45] 0.2780691 193.21151

[46] 0.2718713 185.65987

[47] 0.2552222 179.89092

[48] 0.2738145 179.89092

[49] 0.5631336 144.20342

[50] 0.2629283 84.50994

**Final 20 Moves**

lhs rhs support

[1] {,"6742521"} => {,"6752521"} 0.0001049021

[2] {,"6062521"} => {,"6072521"} 0.0001251154

[3] {,"6032521"} => {,"6072521"} 0.0001155989

[4] {,"6062521"} => {,"6032521"} 0.0001080005

[5] {,"5309905"} => {,"7029904"} 0.0001040169

[6] {,"5109905"} => {,"5749904"} 0.0001091071

[7] {,"5189905"} => {,"6949904"} 0.0001054923

[8] {,"2716578",,"3908011"} => {,"3988011"} 0.0001251154

[9] {,"2726578",,"3998011"} => {,"3908011"} 0.0001197301

[10] {,"7596135"} => {,"6656135"} 0.0001379515

[11] {,"3690654",,"3898011"} => {,"3968011"} 0.0001614106

[12] {,"3908011"} => {,"3988011"} 0.0001690828

[13] {,"3524026",,"3978011"} => {,"3898011"} 0.0002253699

[14] {,"3988011"} => {,"2716578"} 0.0002255913

[15] {,"3908011"} => {,"2716578"} 0.0001604516

[16] {,"3908011"} => {,"3998011"} 0.0001721074

[17] {,"3908011"} => {,"2726578"} 0.0001682713

[18] {,"3898011"} => {,"3968011"} 0.0003884773

[19] {,"8798636"} => {,"2783996"} 0.0002139355

[20] {,"3898011"} => {,"3978011"} 0.0004002069

[21] {,"8798636"} => {,"5528349"} 0.0002497143

[22] {,"3898011"} => {,"3524026"} 0.0003107966

confidence lift

[1] 0.6882865 4560.14729

[2] 0.7054908 3831.43996

[3] 0.6774751 3679.28986

[4] 0.6089850 3568.99715

[5] 0.4554264 1579.31119

[6] 0.4377035 1547.94806

[7] 0.4203410 1338.17001

[8] 0.7797701 1038.73487

[9] 0.5097362 823.56659

[10] 0.4764331 809.91778

[11] 0.8284741 583.94197

[12] 0.2731824 363.90731

[13] 0.5156989 338.80437

[14] 0.3005110 331.72426

[15] 0.2592372 286.16344

[16] 0.2780691 193.21151

[17] 0.2718713 185.65987

[18] 0.2552222 179.89092

[19] 0.2059220 85.92280

[20] 0.2629283 84.50994

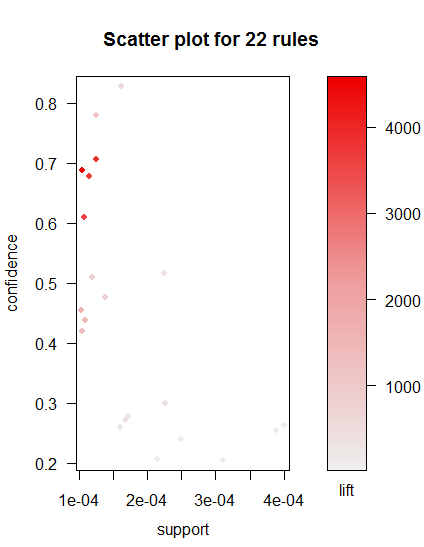
[21] 0.2403607 71.67511

[22] 0.2041875 52.28693

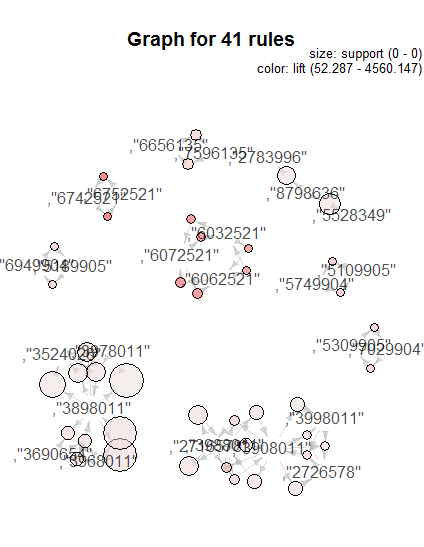
22 Rules were observed by association rules. There were several lhs that contained two SKUs. All the rhs has one SKU. 12 The minimum support for the data set was pretty small. In order to come up with decent number of transactions, I had to come up with minimum support of 0.0001. 12 Items had confidence larger than 0.4. That means conditional probability of rhs given lhs were above 40percent, which is pretty probable. 8 Rules had 4 digits lift, which means high correlation between two purchases.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Item 1 | Department | Item 2 | Department | Item 1 | Department |
| **6742521** | KORET OF | >=6696651, KORET OF |  | **6752521** | KORET OF |
| 6062521 | NOBLE EX |  |  | 6072521 | NOBLE EX |
| 6032521 | NOBLE EX |  |  | 6072521 | NOBLE EX |
| 6062521 | NOBLE EX |  |  | 6032521 | NOBLE EX |
| 5309905 | MILCO IN |  |  | **7029904** | KORET OF |
| 5109905 | MILCO IN |  |  | 5749904 | MILCO IN |
| 5189905 | MILCO IN |  |  | **6949904** | KORET OF |
| 2716578 | CLINIQUE | 3908011 | CLINIQUE | 3988011 | CLINIQUE |
| 2726578 | CLINIQUE | 3998011 | CLINIQUE | 3908011 | CLINIQUE |
| **7596135** | KORET OF |  |  | 6656135 | HUE/KAYS |
| 3690654 | CLINIQUE | 3898011 | CLINIQUE | 3968011 | CLINIQUE |
| 3908011 | CLINIQUE |  |  | 3988011 | CLINIQUE |
| 3524026 | CLINIQUE | 3978011 | CLINIQUE | 3898011 | CLINIQUE |
| 3988011 | CLINIQUE |  |  | 2716578 | CLINIQUE |
| 3908011 | CLINIQUE |  |  | 2716578 | CLINIQUE |
| 3908011 | CLINIQUE |  |  | 3998011 | CLINIQUE |
| 3908011 | CLINIQUE |  |  | 2726578 | CLINIQUE |
| 3898011 | CLINIQUE |  |  | 3968011 | CLINIQUE |
| **879863**6 | KORET OF |  |  | 2783996 | LANCOME |
| 3898011 | CLINIQUE |  |  | 3978011 | CLINIQUE |
| **8798636** | KORET OF |  |  | 5528349 | LANCOME |
| 3898011 | CLINIQUE |  |  | 3524026 | CLINIQUE |

This is the table that displays which department each item in lhs and rhs belongs to. ‘VLOOKUP’ function was used to locate the department of each SKU. The excel could not load the whole transaction data because the data set was extremely huge. Excel did not display SKUs with SKU number greater than 6696651. It was not sure if VLOOKUP function displayed right department for those SKUs, or it just located KORET OF, which was the department of the last SKUs on the spreadsheet. The colored relationships signified interdepartment move while other relationships merely signified intradepartment movements.



Majority of rules had high confidence, higher than 0.4. This means that it is very likely for one of those items in lhs will trigger impulsive purchase of other item. Support for all rules were generally very low. It is because the data set was huge, and, thereby, the probability for certain item’s purchase was very minute. Majority of rules had higher lift. It signified that the purchase of those groups of items were not coincidental but had cause &effect relationship.



This data maps clearly presents how SKUs should be placed in order to trigger impulsive purchases and increase the profit. 6656135(Hue/Kays) should be moved to KORET OF department, near SKU 7596135. Intradepartment move of 675251(KORET OF) should me made to 674251(KORET OF). 6949904(KORET OF) should be moved to 5189906(MILCO IN). Four items from CLINIQUE Department(3524026,3978011,3968011,3968011) should be moved close to 3898011(CLINIQUE). 4 SKUs from CLINIQUE(3998011,272658,2716578,3988011) should be placed closed to SKU 3908011. SKU 5309905(MILCO IN) should be moved to 7029904(KORET OF). 5109905(MILCO IN) has to be placed close to 5749904(MILCO IN). 6032521 and 6062521 from NOBLE EX should be placed next to 6072521(NOBLE EX). 278396 and 5528349 from LANCOME should be placed next to 8798636(KORET OF). Thus, total 17 moves of item including intradepartment and interdepartment have to be made in order to trigger impulsive purchases.

**Conclusion**

Little bit more SKUs than 100 were selected using association rules. However, due to budgetary restriction, we could not move everything on the list. List was shortened to about 20 rules by running association rules with higher minimum support. Total 22 rules were selected, and 26 items were chosen to improve the planogram of the store. A lot of those relationships had higher confidence, which was larger than 40%. Total 17 moves can be made with those 22 rules found.

Products from NOBLE EX and CLINIQUE had affinity relationships within department. SKUs from MILCO IN, LANCOME, HUE/KAYS and KORET OF had affinity relationships with other products from other department.

**NEXT STEP**

Possible next step is to move items as suggested on the map. For relationships within same department, items can be located close to each other. Customers might need other items at the time of their purchases, and it will lead to extra sales.

For items that are in different department, it might takes a lot of time and labor to change the location of the items. Thus, the store can perform promotion or issue a coupon that gives discounts if two items are purchased together. Customer might travel to other department to get the deal, and this might increase the shopping time and lead to another impulse transaction.