Rafael Guimaraes and Brian Kangogo  
 Marketing Analytics

3/16/2017  
 **Market Basket Analysis: Miller Chapter 9**

Github Link: https://github.com/bmainaa/MarketBasket

1. **(10 pts) Modify Miller’s script to focus to identify which items are categorized as “dairy produce”. Report the specific items that make up the category.**



1. **(10) Miller decides to use a support cutoff of 0.025, which yields 344 rules. Modify the code to create approximately 200 rules. Report on the support cutoff that accomplishes this and also the number of resulting rules.**

Support refers to the number of transactions that include all items in the antecedent and consequent parts of the rule. In this case miller set the proportion to 0.025. To get 200 rules, we would need to increase the support. As we would expect, higher support yields lower rules. In our case:  
Support: 0.0343  
Rules: 201

1. **(10) Working from the same cutoff as you found in step 2, suppose we tighten the confidence level to 0.01 rather than 0.05. Describe the impact of these changes on the set of rules.**Confidence refers to the ratio of the number of transaction that include all items in the consequent and the antecedent (support) to the number of transactions that include all items in the antecedent. By definition above, we would expect that tightening confidence level would increase the number of rules. As we would expect, the number of rules increased  
   Confidence: 0.01  
   Rules: 206
2. **(25) Now, using the cutoff from step 2 and a confidence level of 0.05, modify the script one last time to generate Miller’s analysis, but for dairy produce rather than vegetables. This should create 5 pdfs (rename the last one from “farmer rules” to “dairy rules”).**

Uploaded on latte

1. **(25) In your document, write a short commentary and explanation for each of the five graphs. Write as if you are presenting this to a merchandising team in charge of store layout as well as promotions for various food items.**

**Graph1**

Graph1 shows the proportion of shopping basket with specific items with a support of 0.0343. This graph is very important because it gives us most purchased items in the store. As highlighted, the results are as we expected, whole milk, vegetables and soda are bought by many customers. Its interesting that not many people don’t buy onions.

**Graph2**

Graph2 is almost similar to Graph1 but the difference is that Graph2 is an aggregate of similar items eg, dairy produce which contains whole milk, yogurt UHT-milk etc. This graph shows the proportion of different item category in a shopping basket with a support of 0.0343. As expected, dairy produce, vegetables and bread/backed goods are included in many shopping baskets. It is also interesting that not many people buy canned fruits or vegetables.

**Graph3**

Graph3 is a scatter plot of 201 association rules obtained by confidence of 0.05 and support of 0.0343. In the graph support is on the horizontal axis and confidence on the vertical axis. As expected (by definition), a higher confidence translates to lower support. This is highlighted by lift which is a measure of relative predictive confidence.

**Graph4**

Graph4 gives us a clear view of the different association rules. The matrix shows the antecedent at the top and the consequent on the right. The size of the bubble represents support while the color intensity represents lift (measure of confidence).

As shown in the matrix, baskets with dairy products is associated with many other items (support) and also has higher lift as compared to beer which has higher support but smaller lift.

**Graph5**

Graph5 shows top 10 rules that list dairy produce as the consequent. This answer the question, what products are commonly purchased with dairy produce? As shown in the graph, bread and backed goods,fruit and vegetables have the highest confidence which shows that these are items that are commonly purchased with dairy produce.

1. **(10) Make specific recommendations aimed at cross-selling to customers who purchase dairy produce – what other items might we encourage them to buy, and how might we accomplish that based on the intelligence revealed in your analysis.**

Our analysis shows that people who buy dairy produce also buy bread and backed goods, cheese , fruit and vegetables. Because this are mostly bought together, we can get customers to buy them more by;

* Placing dairy produce and items above close together
* Giving discounts that link the two products, e.g., 10% bread when you buy a given dairy produce
* It is also important for us to try and sell the less obvious products to these consumers e.g. books, chocolate and perfumery. We can do this by offering discounts like 10% on perfumery when you purchase dairy produce.

1. **(10) (reflection—no wrong answers here!) In a corporate setting, version control is quite common and serves important functions. Comment on your impressions of this approach. Did it facilitate or impede your workflow? How might it be helpful going forward?**

Github really helped us to manage our different versions of our code. One could easily delete or add code regardless of where they are. It also keeps track of different version of the code. This made work easier because without this, we would have to share code via email and incase of any errors it would take us longer to figure out the error.

It first, it was hard to understand how it works but after a while playing around with push, pull and commit our work became easier.