Kennesaw State University
IS 8935 Business Intelligence: Traditional & Big Data Analytics
Dr. Reza Vaezi
Assignment 4
February 3, 2019

Bakery Sales Association Rules Analysis

By Karis Kim

Executive Summary

The goal of this report is to uncover meaningful association rules among the bakery sales data and offer management insights to potentially increase sales. The analysis revealed the following top three affinity products: (1) *bread5* and *Dissert2*, (2) *Dissert1* and *coffee2*, (3) *bread2* and *Dissert2*. Thus, the bakery is advised to capitalize on cross-selling opportunities when *bread5, Dissert1, bread2* are purchased. *Dissert2, bread2, coffee2, Dissert1, bread5* and *tea2* are the top most frequently sold individual items, so inventory and production of these items should be constantly maintained to meet the demand in sales.

CONTENTS

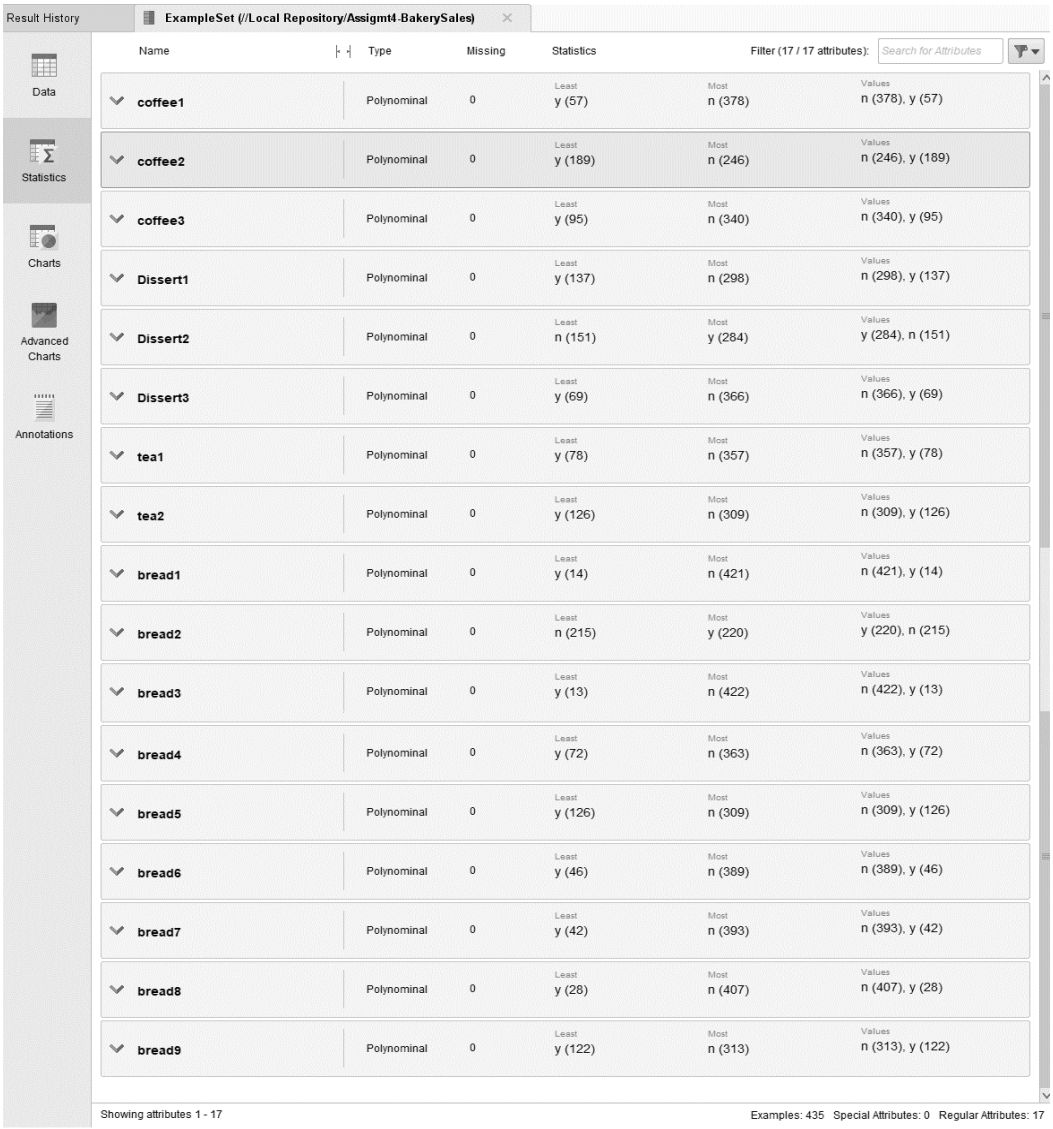
BUSINESS UNDERSTANDING

A small bakery sells 3 types of coffee, 3 types of desserts, 2 types of tea and 9 types of bread and seeks to increase their average order value per visit. This report aims to uncover meaningful association rules among the sales data for this bakery and offer management insights to potentially increase their sales. By revealing affinity products, this report will equip the bakery with necessary information for cross-selling or up-selling.

DATA UNDERSTANDING

The Bakery Sales data set contains 17 regular attributes, where each attribute represents a unique menu item sold at the bakery. The dataset represents sales data for this small bakery over a given week. There are 435 examples (rows) or total transactions. All values are in the form of Y for yes and N for No. Initial data set values are in the data type polynominal. For each attribute, the count of Y plus the count of N equals 435.

Figure 1. Descriptive statistics overview of raw data set



DATA PREPARATION

Data Type Transformation: Using RapidMiner operator *Nominal to Binominal*, the data type of all attributes has been converted from polynominal to binominal, since Association Rules requires data types to be binominal.

Figure 2. RapidMiner Process for data type conversion to binominal

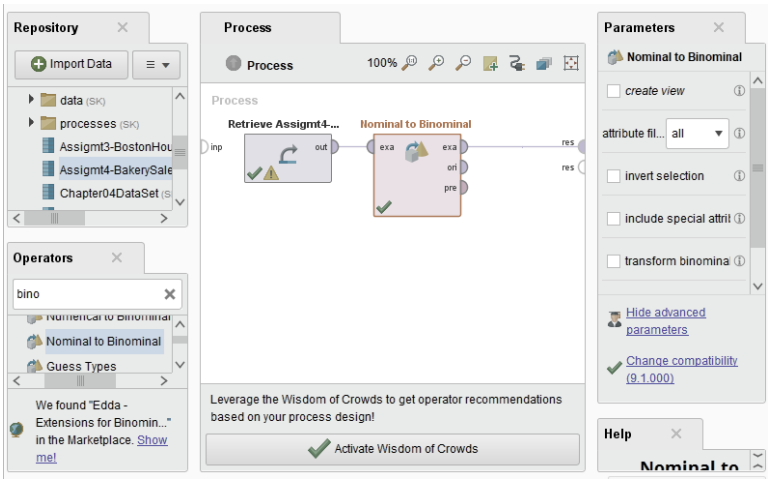
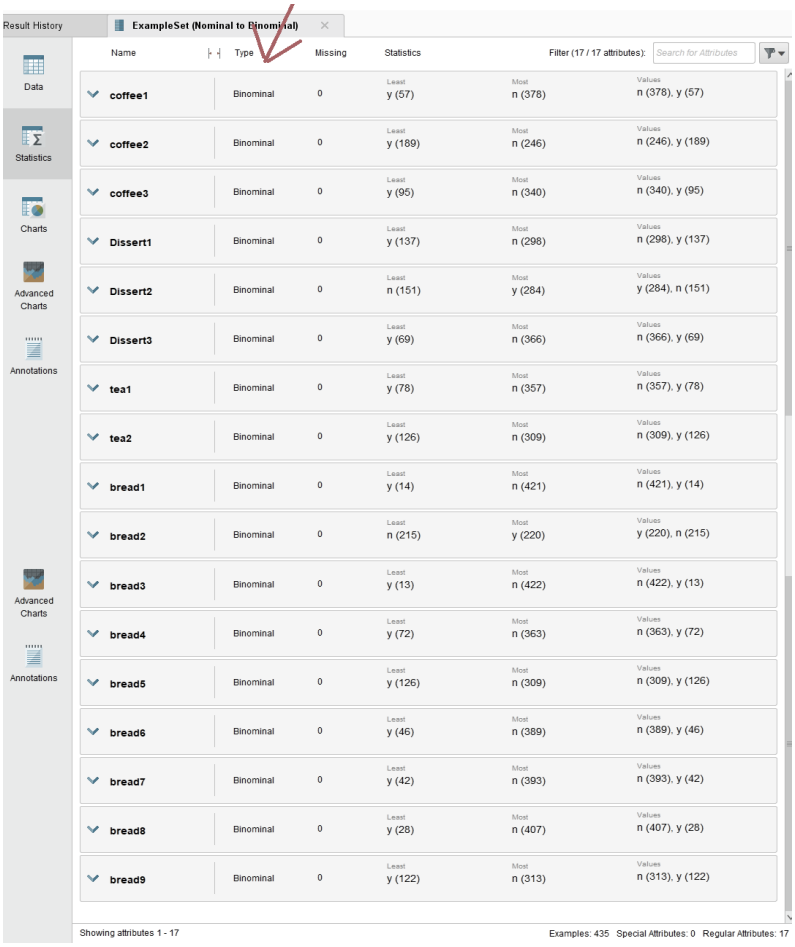


Figure 3. RapidMiner Result with all data types in binominal and no missing values



Data Preparation of Missing Values: No missing values were found in the data set.
Data Preparation of Inconsistent Values: No inconsistent values or outliers were found in the data set.

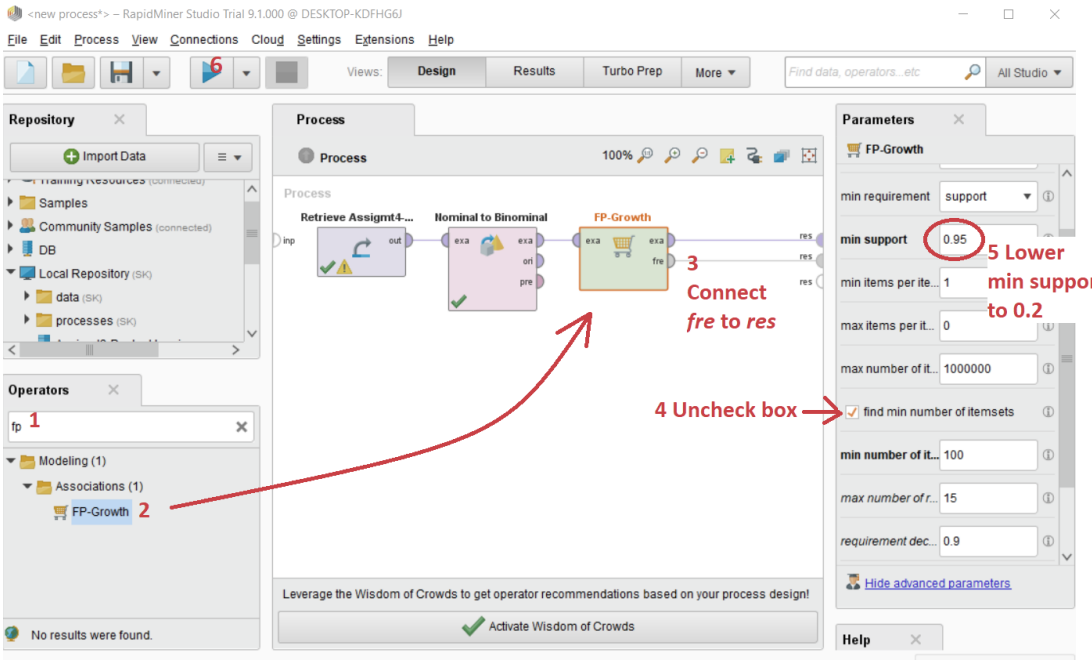
MODELING

The two-step process of generating meaningful association rules involves: finding all frequent item sets and extracting the rules from frequent item sets.

1. FP-GROWTH

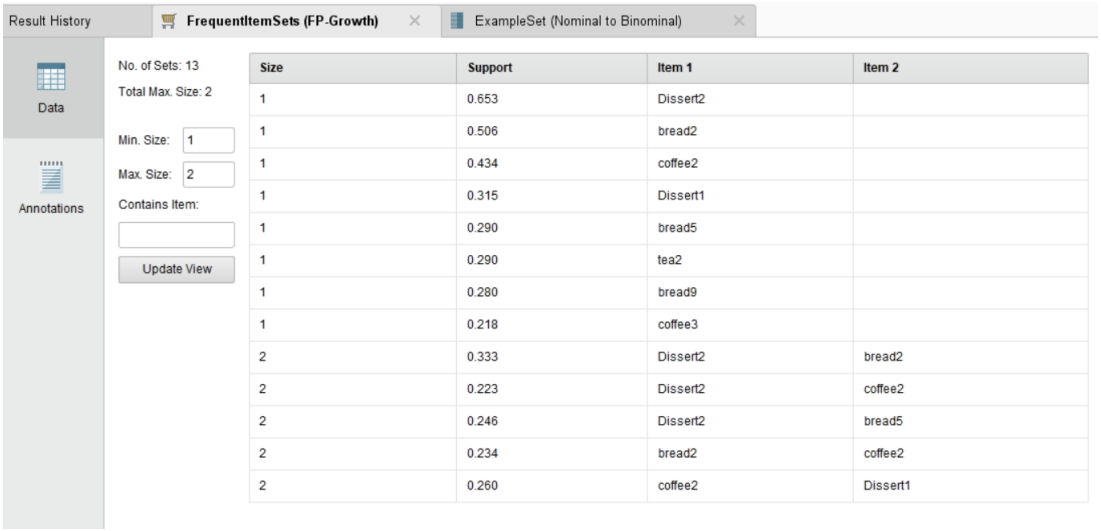
First, the FP-Growth operator in RapidMiner is used generate all frequent item sets meeting the minimum support criteria of 0.2. Support measures the frequency that an item(s) occurs in the data set. Support is calculated by the number of occurrences of an item divided by the total number of transactions in the data set. By setting the minimum support to 0.2, the results will show all bakery menu item(s) that occur in at least 20% of the 435 transactions. Setting the minimum support to a higher number yielded not enough items in the result, so the threshold has been set to 0.2.

Figure 4. RapidMiner Frequent Pattern-Growth operator process



The result (in figure 5 below) shows 8 items that appeared in at least 20% of the transactions, with Dissert2 appearing the most at 65.3% of the transactions. In other words, Dissert2 was the most frequently purchased menu item at the bakery in this given week of sales. The second most frequently purchased menu item at the bakery was bread2, appearing in 50.6% of the transactions in this given week. Lowering the minimum support parameter also yielded 5 item pairs that appeared together in at least 20% of the transactions, with Dissert2 and bread2 at the top of those pairs, appearing together in 33.3% of the transactions. Dissert2 and bread2 pair, which showed as the top first and second most frequently purchased single menu item, was also the most frequently purchased pair of items at the bakery in this given week of sales.

Figure 5. RapidMiner Result with minimum support over 0.2

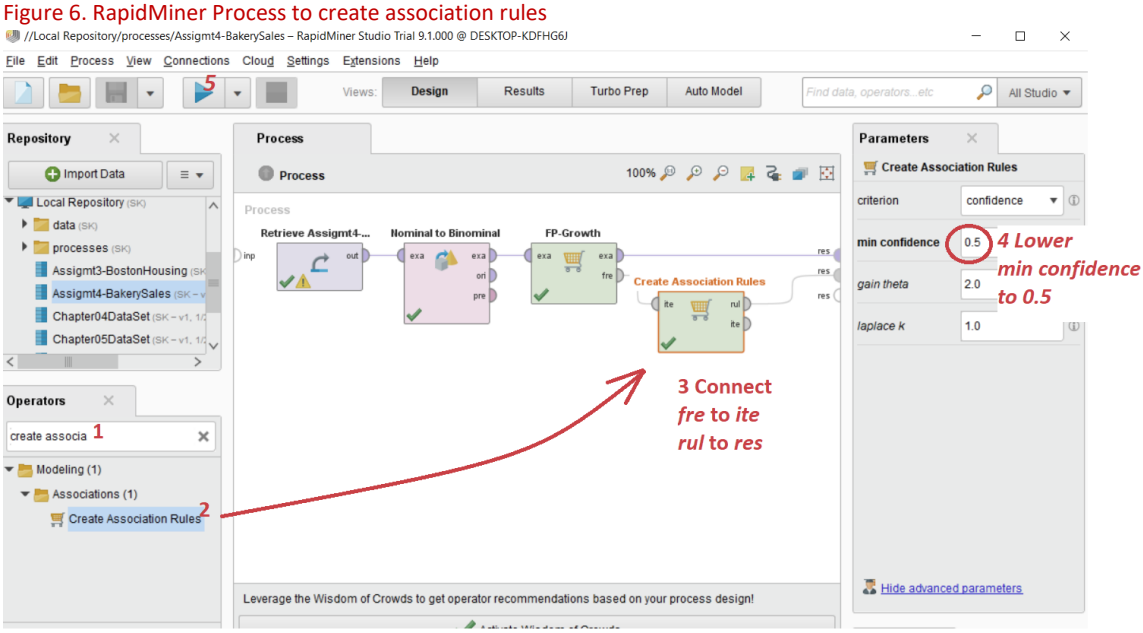


Size	Support	Item 1	Item 2
1	0.653	Dissert2	
1	0.506	bread2	
1	0.434	coffee2	
1	0.315	Dissert1	
1	0.290	bread5	
1	0.290	tea2	
1	0.280	bread9	
1	0.218	coffee3	
2	0.333	Dissert2	bread2
2	0.223	Dissert2	coffee2
2	0.246	Dissert2	bread5
2	0.234	bread2	coffee2
2	0.260	coffee2	Dissert1

2. CREATE ASSOCIATION RULES

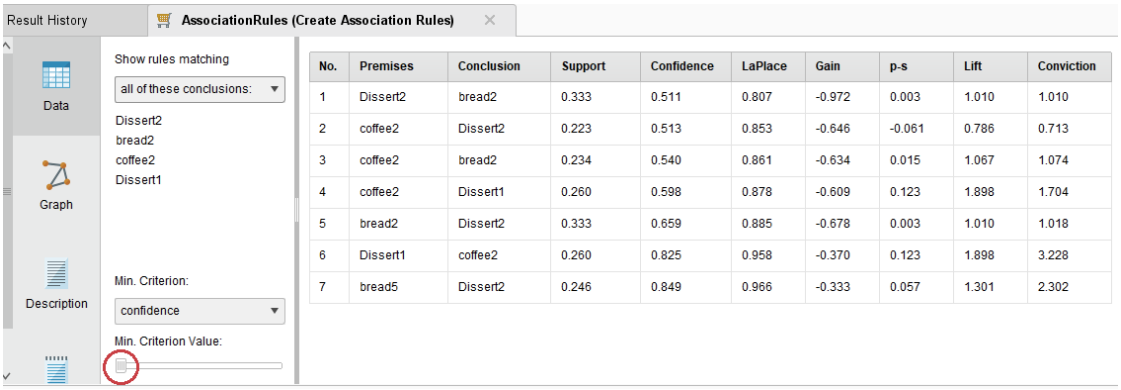
Next, the RapidMiner Create Association Rules operator is used to take frequency pattern data results and seek out any patterns that occur frequently enough to generate association rules. For this, the minimum confidence is set to 0.5, so that results will show association rules with at least 50% confidence of the rule. Confidence is the reliability measure of an association rule, and tells the likelihood of conclusion y from all transactions containing y. Confidence of a rule x → y is calculated by the support (x U y) divided by support (x).

Figure 6. RapidMiner Process to create association rules



The result (in figure 7 below) shows that the association rule with the highest confidence is *bread5 → Dissert2*. In other words, when bread5 was purchased, Dissert2 was also purchased. This rule has an 84.9% confidence, which means that 84.9% of customers who purchased bread5 also purchased Dissert2, and we are 84.9% confident that the purchase of bread5 will also yield a purchase of Dissert 2.

Figure 7. RapidMiner Result with association rules for 0.5 minimum confidence



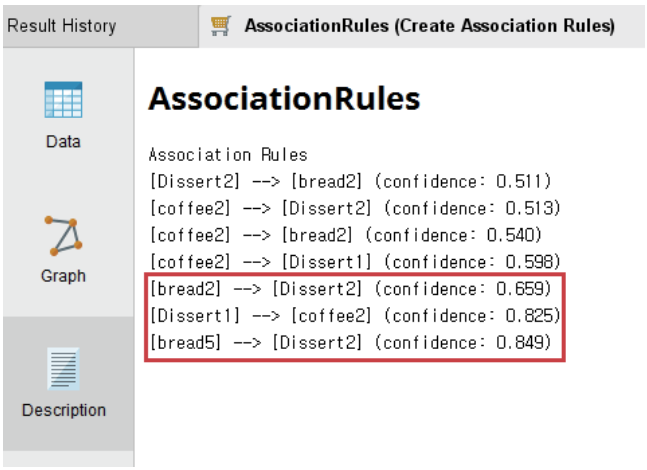
No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift	Conviction
1	Dissert2	bread2	0.333	0.511	0.807	-0.972	0.003	1.010	1.010
2	coffee2	Dissert2	0.223	0.513	0.853	-0.646	-0.061	0.786	0.713
3	coffee2	bread2	0.234	0.540	0.861	-0.634	0.015	1.067	1.074
4	coffee2	Dissert1	0.260	0.598	0.878	-0.609	0.123	1.898	1.704
5	bread2	Dissert2	0.333	0.659	0.885	-0.678	0.003	1.010	1.018
6	Dissert1	coffee2	0.260	0.825	0.958	-0.370	0.123	1.898	3.228
7	bread5	Dissert2	0.246	0.849	0.966	-0.333	0.057	1.301	2.302

EVALUATION OF FINDINGS

1. SUMMARY OF FINDINGS

7 Association Rules Yielded: From the Bakery Sales data set with support minimum at 20% and confidence minimum at 50%, the following 7 association rules (in figure 8) were discovered. The top 3 premise → conclusion rules (in red box below) show that 85% of customers who purchased bread5 also purchased Dissert2, 82.5% of customers who purchased Dissert1 also purchased coffee2, and 66% of customers who purchased bread2 also purchased Dissert2.

Figure 8. RapidMiner Description of association rules with 0.5 minimum confidence

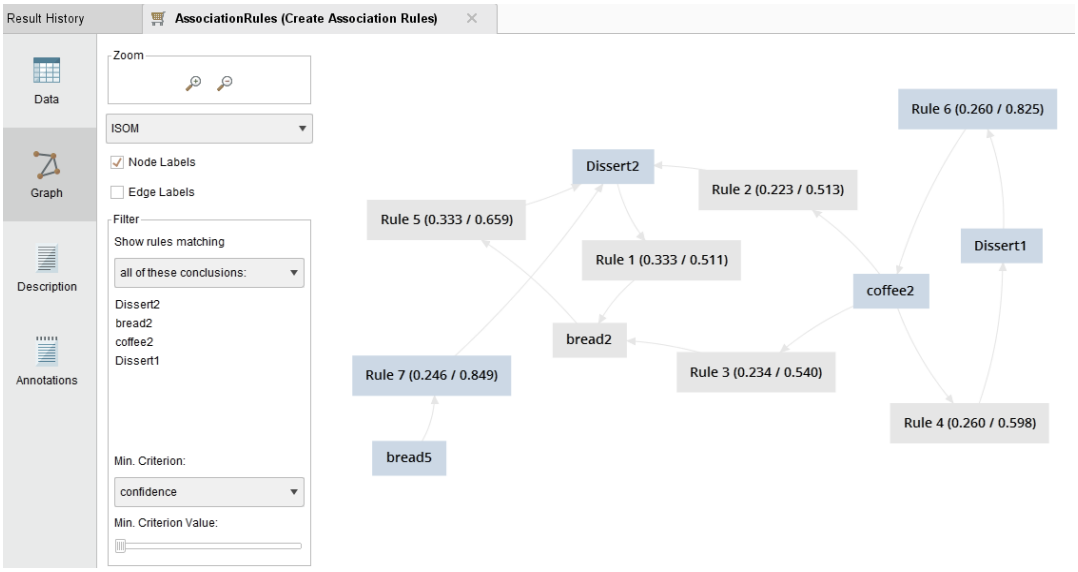


Association Rules with Highest Confidence: In the following graph of the association rules (figure 9), we see in Rule 7 that even though the bread5 and Dissert2 pair only appeared in 24.6% of the transactions, 85% of customers who purchased bread5 also purchased Dissert2. Similarly, in Rule 6, the Dissert1 and coffee2 pair only appeared in 26% of the transactions, but 82.5% of customers who purchased Dissert1 also purchased coffee2. Rule 7 and Rule 6 had the two highest confidence levels.

When Premise-Conclusion Reversed: For the bread5 and Dissert2 pair in Rule 7, no significant association rule was found for the reverse premise and conclusion; that is, Dissert2 purchases did not result in bread5 purchases

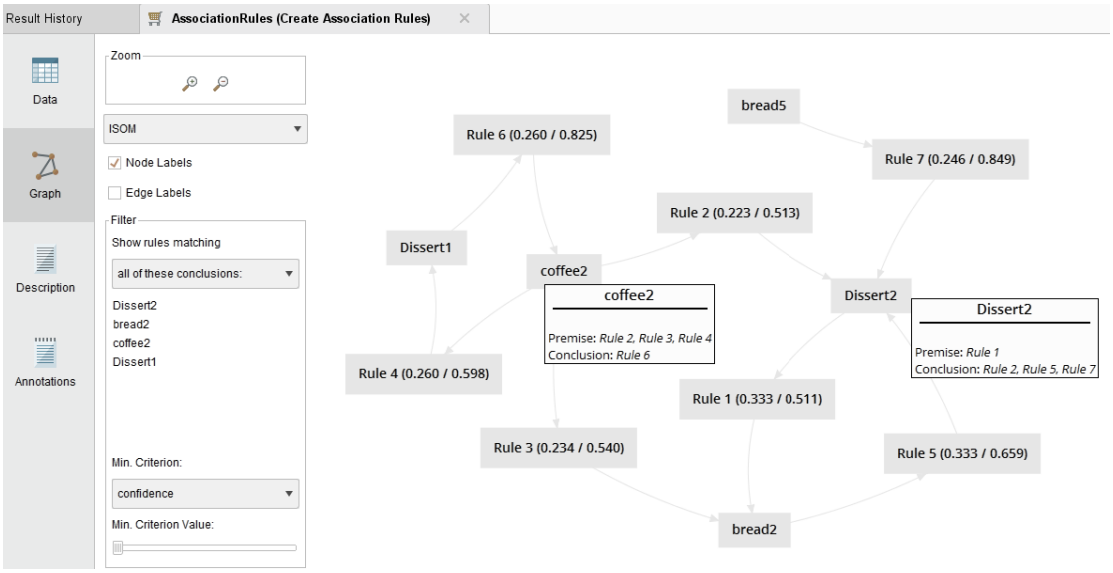
with a confidence above our minimum 50%. In fact, Dissert2→ bread5 only had a confidence of 37.7% (not shown in figure9), so this premise-conclusion was dropped from our association rule set.
For the reverse premise and conclusion of the Dissert1 and coffee2 pair in Rule 6, coffee2 → Dissert1 also had a lower confidence than Dissert1 → coffee2, but the reversed rule did yield a confidence of 59.8%, so the reversed premise-conclusion of Rule 6 did make the top 7 association rules, holding the fourth highest confidence.

Figure 9. RapidMiner Graph of association rules wit 0.5 minimum confidence with top 2 rules highlighted



Items Occurring Most Often in Rules: In the following graph of association rules (figure 10), two menu items can be found occurring most often in the 7 association rules; Dissert2 and coffee2. These two items each appear in four different rules. Dissert2 appears in one premise and three conclusions, while coffee2 appears in three premises and one conclusion. In other words, customers who buy coffee2 are also likely buy another item (namely Dissert1, bread2, Dissert2- shown in figure 8), and customers who buy another item (namely bread5, bread2, coffee2- shown in figure 8) are also likely to buy Dissert2.

Figure 10. RapidMiner Graph of association rules with top 2 items occurring the most in rules



2. SUPPORT, CONFIDENCE, LIFT

The following figure shows a side-by-side listing of the top items by support, confidence, and lift with brief explanations below.

Figure 11. Side-by-side list of top items by support, confidence, lift

Top 5 by Support		Top 5 by Confidence		Top 3 by Lift	
Dissert2	65.3%	**bread5 → Dissert2** Rule 7	84.9%	**coffee2 → Dissert1** Rule 4	1.898
bread2	50.6%	**Dissert1 → coffee2** Rule 6	82.5%	**Dissert1 → coffee2** Rule 6	1.898
coffee2	43.4%	**bread2 → Dissert2** Rule 5	65.9%	**bread5 → Dissert2** Rule 7	1.301
Dissert1	31.5%	**coffee2 → Dissert1** Rule 4	59.8%		
bread5, tea2	29%	**coffee2 → bread2** Rule 3	54%		

Items Occurring Most Often in Purchases: During this given week, the top 5 menu items in the order of popularity as indicated by the support percents are Dissert2, bread2, coffee2, Dissert1, bread5, tea2 (bread 5 and tea2 are tied for 5th place).

Highest Confidence and Highest Support: The association rule with the highest confidence is *bread5 → Dissert2*. Dissert2 is also the single most frequently purchased item, showing up in 65.3% of all purchases, but bread 5, which is the premise of the top association rule, only shows up in 29% of purchases. Bread5 and Dissert2 as a pair only appeared in 24.6% of the transactions, but had the highest confidence and thereby the highest potential for cross-selling.

The association rule with the second highest confidence is *Dissert1 → coffee2*, with 82.5% confidence that customers who purchase Dissert1 will also purchase coffee2. Dissert1 only appeared in 31.5% of transactions when purchased alone and coffee2 in only 43.4% of transactions when purchased alone, and Dissert1 plus coffee2 as a pair only appeared in 26% of transactions. Thus, the analysis indicates that high support does not necessarily point to high confidence.

High Lift: In order to avert the frequency of occurrence of a rule conclusion from influencing the confidence of an association rule, the lift is considered. The lift is the ratio of the confidence of the rule and the expected confidence of the rule [1]. In other words, lift is the ratio of support of the union of the premise and conclusion with what is expected, if the premise and conclusion were completely independent [2]. Both Rule 7 and Rule 6, the top two rules in terms of confidence, have a lift value above 1. Lift value greater than 1 indicates that the premise and the conclusion appear more often together than expected, which means that the occurrence of the premise has a positive effect on the occurrence of the conclusion [1].

Interestingly, association Rule 6 (*Dissert1 → coffee2)*, which was second highest in confidence, has a higher Lift value than Rule 7 (*bread5 → Dissert2*), even though Rule 7 was the first highest in confidence. This higher lift value for Rule 6 means that the Dissert1 premise and coffee2 conclusion appear more often together than expected, and the occurrence of Dissert1 has a positive effect on the occurrence of coffee2.

BUSINES RECOMMENDATIONS

1. CROSS-SELLING/UP-SELLING

This analysis found 7 association rules which determined that *bread5, Dissert1, bread2, coffee2* and *Dissert2* are the best candidates for cross-selling and/or up-selling. Training of bakery sales staff is advised, so that they can be vigilant for any customers purchasing these items and ask if the customer would also like to purchase the conclusion item to these premise items as identified in the Association Rules in figure 8 and figure 11. Bakery management is advised to develop some program in the cash register computer that alerts the sales staff that a premise item has been entered and to remind them to cross-sell/up-sell.

In addition, the Bakery management should stock or present these 5 premise items next to the conclusion items, to ensure ease of access to these conclusion items that are highly likely to be purchased together. If resources permit, the management should create combo menus comprised of these items and further encourage cross-selling of these high confidence items sets. For menu items with lower support or sales, a triple combo menu comprised of the high confidence premise → conclusion item and the third lower selling item with a discount as an incentive could be created to boost potential sales.

2. INVENTORY MANAGEMENT

Of the 17 items on the menu at the Bakery, this analysis revealed that *Dissert2, bread2, coffee2, Dissert1, bread5* and *tea2* were the most frequently sold items. These items were also appeared in many association rules which revealed opportunities for cross-selling and/or up-selling. In particular, *Dissert2* and *coffee2* appeared most frequently in the association rules. So, the bakery management is advised to ensure that inventory and production of these top selling items is constantly maintained in preparation to meet the demand in sales, as well as to ensure that inventory is ready for successful cross-selling attempts.

REFERENCES

1 Lift in an association rule. (n.d.). Retrieved from https://www.ibm.com/support/knowledgecenter/SSEPGG_10.5.0/com.ibm.im.model.doc/c_lift_in_an_association_rule.html

2 Kotu, V., & Deshpande, B. (2015). *Predictive analytics and data mining: Concepts and practice with RapidMiner*. Amsterdam: Morgan Kaufmann.