IE School of Human Sciences & Technology

Machine Learning I

Machine Learning Unsupervised Learning
Association Analysis for a grocery store

Author: Krishna Agrawal

Course

Master's in Business Analytics
and Big Data

Executive Summary:

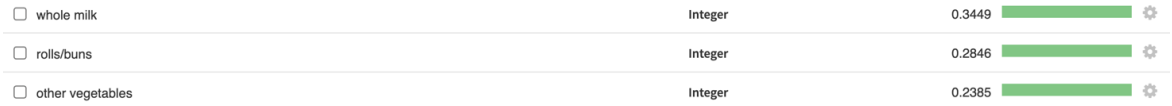
We developed the following analysis with the main objective of assisting your store in finding insights on interesting combinations of products. These findings are helpful in order to develop relevant marketing offers and decide on the placement of the products in your store. In order to accomplish this, we conducted association analysis based on the dataset provided. The primary goal with association analysis is to find combinations of products that are purchased together more frequently than purchased separately in order to assist you with the placement of products and potential promotions. Association analysis also allows us to identify negative relationships between products in order to recommend which products should not be placed near each other. We used the following metrics: Lift, Confidence Difference, and Confidence Ratio when running the Apriori Algorithm to derive our conclusions and offer recommendations we believe would be beneficial for the management of your store. As a final step, we also conducted prediction analysis in order to generate interesting insights at the customer level. These findings can assist you with implementing promotions or offers that are relevant to your individual customers on their next visits. In the following sections, you will find all the details of our approach and analysis along with the rules (product combinations) that we found most relevant for your store.

Analysis Approach:

We would like to walk you through our analysis to explain the steps taken when conducting association analysis and running the Apriori Algorithm (refer to **Annex 1c** for the definition) to reach our conclusions and recommendations for your store. In the next section, we will describe the steps taken in the analysis, the metrics and parameters selected as well as the conclusions and recommendations.

Sample Data Analysis:

- Firstly, we looked at the sample data of 1,000 customers provided to us and realized there were customers without any purchases included in the dataset. In order to get better conclusions, we decided to clean the data by removing empty rows with no purchases. This decreased the total sample from 1,000 to 780 shopping trips from customers; 220 rows were removed (refer to **Annex 2** for additional details). We also noticed that the majority of customers only purchased 1 or 2 products. For this reason, it's likely we wouldn't see large combinations of antecedents (products already in the customer's basket) leading to the purchase of a consequent (product we are studying) in the results.
- Secondly, we wanted to understand the sample data before running any further analysis. We looked at the prior confidence (mean purchasing frequency) for each product:
    - From the sample dataset, there are 14 products total which will be analyzed.
    - The top 3 products bought the most are the following as evidenced by the screenshot below: whole milk (34%), rolls/buns (28%) and other vegetables (24%).



    - The bottom 3 products bought the least are the following: whipped/sour cream (9%), bottled beer (9%) and sausage/pastry/shopping bags (each at 10%).

- Finally, for our association analysis of the products we decided to look at three metrics: **Lift, Confidence Difference and Confidence Ratio.** Analyzing the data with these three different metrics would allow us to select different types of rules (showing both positive and negative relationships between products) which can derive useful information from a store management perspective.
- We would like to describe all the metrics and parameters of our analysis for your understanding. We believe this would be useful, as these concepts will be used in the remainder of the document to explain the methodology and recommendations (please read **Annex 1**).

Association Analysis Metrics and Conclusions:

Taking into consideration the definitions provided (refer to **Annex 1**), we will start by describing our approach to gather interesting rules for the relationship between products using **Lift** as an initial metric. Our objective was to find rules with high values for this metric. The higher the value, the greater the association between antecedent and consequent.

Below is an explanation of the metric and the parameters selected to derive interesting rules using Lift:

- **Lift** is the ratio between posterior confidence and prior confidence. It explains the likelihood of purchasing a product if another product (or combination of products) is in the basket.
    - Lift = Posterior Confidence / Prior Confidence
    - The parameters below were selected to derive rules using this metric:
        - **Minimum threshold** = 1
            - We selected a minimum threshold of 1. If lift is equal to 1, it means that the antecedent and consequent are independent. If lift is greater than 1 there is association between the antecedent and consequent which are the rules we are looking for.
        - **Minimum confidence**: 25%
            - We ran the analysis with different confidence values and decided to leave it at 25% to have a decent number of rules generated based on our dataset.
        - **Minimum rule support:** 0%
            - We decided not to include any restriction on rule support.
        - **Minimum support:** 10%
            - We decided to have instances of at least 10% of all shopping trips.
        - **Maximum antecedents:** 5
            - We allowed a maximum combination of 5 antecedents to be produced.
- **Conclusions and Recommendations**: 19 rules were generated with these parameters (refer to **Annex 4a** for all rules). We sorted the results by lift to look at the ones with the strongest relationships.
- Out of the 19, we selected the 10 rules below which we consider most relevant from a store management perspective and we decided to group them in different categories.

antecedents	consequent	instance	antecedent suppor	consequent suppor	confidence	rule support	lift	conf.difference	conf.ratio	rule
'tropical fruit	'root vegetables	97	0.124358974	0.141025641	0.268041237	0.033333333	1.900656045	0.127015596	0.473865878	18
'root vegetables	'other vegetables	110	0.141025641	0.238461538	0.390909091	0.055128205	1.639296188	0.152447552	0.389982111	19
'sausage	'rolls/buns	78	0.1	0.284615385	0.461538462	0.046153846	1.621621622	0.176923077	0.383333333	5
'tropical fruit	'yogurt	97	0.124358974	0.162820513	0.257731959	0.032051282	1.582920692	0.094911446	0.36825641	12
'shopping bags	'soda	78	0.1	0.203846154	0.307692308	0.030769231	1.509433962	0.103846154	0.3375	15
'tropical fruit	'other vegetables	97	0.124358974	0.238461538	0.329896907	0.041025641	1.383438643	0.091435369	0.277163462	16
'root vegetables	'whole milk	110	0.141025641	0.344871795	0.445454545	0.062820513	1.291652585	0.100582751	0.225798012	2
'yogurt	'whole milk	127	0.162820513	0.344871795	0.440944882	0.071794872	1.278576237	0.096073087	0.217880037	3
'sausage	'soda	78	0.1	0.203846154	0.256410256	0.025641026	1.257861635	0.052564103	0.205	17
'citrus fruit	'other vegetables	94	0.120512821	0.238461538	0.287234043	0.034615385	1.204529856	0.048772504	0.16980057	10

- Based on our lift analysis, the above 10 rules for the specified antecedents and consequent combinations have the greatest lift values.
- The way to interpret these results, taking the first row as example, is that the likelihood of a customer purchasing root vegetables almost doubles when a customer purchases tropical fruits. (Lift=1.9)
- ***Light Green Rows***: What we can see from the results and we recommend to you, is to maintain fruits (tropical fruit, citrus fruit) next to vegetables (root vegetables, other vegetables) along with keeping vegetables next to each other within your store. Rows highlighted in light green support this conclusion. Based on lift, following this recommendation will increase the chances that a customer purchases the consequent if they are also purchasing the antecedent.
- ***Light Orange Rows:*** Another recommendation based on the analysis is to keep some of the diary products (yogurt, whole milk) next to the fruits and vegetables along with keeping dairy products together (whole milk, yogurt). Rows highlighted in light orange support this conclusion. Again, lift shows a positive relationship between these products.
- ***Dark Orange Rows:*** Additionally, shopping bags and sausage have a relationship with soda. Placing these products closer to each other could have a positive impact in the purchase of soda.
- ***Dark Green Row:*** Finally, there is a positive relationship between sausage (antecedent) and rolls/buns (consequent). The Lift is 1.6 meaning that the chances of a customer

purchasing sausage and rolls/buns is 1.6 times more than the chances of the customer only buying sausage. We recommend keeping these products close to each other in the store.

Following the Lift metric, the next association metric we used to extract rules is **Confidence Difference** (refer to **Annex 1b** for the definition). The goal of using Confidence Difference is to extract positive but also negative rules from the dataset.
Below is an explanation of the metric and the parameters selected to derive interesting rules:

- **Confidence Difference** is the Posterior Confidence deducted by the Prior Confidence in absolute value. It sets the minimum lower bound for the difference between the Confidences.
    - Confidence Difference = |Posterior - Prior | > Confidence Difference lower bound
    - The parameters below were selected to arrive at significant rules using this metric:
        - **Minimum threshold** = 5%
            - After looking at the means from our product list (prior conf.) in order to have a feeling for this threshold, we selected a 5% Confidence Difference lower bound. In other terms, the confidence difference is not allowed to be below 5%.
        - **Minimum confidence**: 15%
            - We run our Apriori script with a minimum confidence level of 15% in order to extract significant rules beside the ones generated by lift.
        - The parameters **minimum rule support, minimum support, and maximum antecedents** were left unchanged.
- **Conclusions and Recommendations:** 27 Rules were generated with the parameters above (refer to **Annex 4b** for all rules). We sorted the result first by the Confidence Difference and then a second time using Lift, allowing for easier comparison with the previous analysis.

antecedents	consequent	instanc	antecedent suppo	consequent suppo	confidence	rule support	lift	conf.difference	conf.ratio	rule
other vegetables	rolls/buns	186	0,238461538	0,284615385	0,231182796	0,055128205	0,812263877	0,053432589	0,187736123	27
shopping bags	whole milk	78	0,1	0,344871795	0,269230769	0,026923077	0,780669145	0,075641026	0,219330855	12
soda	whole milk	159	0,203846154	0,344871795	0,238993711	0,048717949	0,692992916	0,105878084	0,307007084	26
pastry	other vegetables	78	0,1	0,238461538	0,153846154	0,015384615	0,64516129	0,084615385	0,35483871	6
tropical fruit	citrus fruit	97	0,124358974	0,120512821	0,195876289	0,024358974	1,625356438	0,075363468	0,384750337	7
citrus fruit	tropical fruit	94	0,120512821	0,124358974	0,20212766	0,024358974	1,625356438	0,077768685	0,384750337	8

    - As a result from our Confidence Difference analysis we added the 6 rules stated above to our general rule selection as they are not duplicates from the lift analysis.
    - ***Light Orange Rows*** show product combinations which should not be placed together given that negative rules were generated with a lift below 1. We see that the store should **not** group other vegetables with rolls/buns, shopping bags with whole milk, soda with whole milk or place pastry close to other vegetables.
    - ***Light Green Rows*** show the product combinations which should be grouped together, adding to our positive rules generated by lift. The store should place tropical fruit with citrus fruit, as it has a lift of 1.6.

Next, we decided to focus on the **Confidence Ratio** (refer to **Annex 1b** for the definition). The purpose was to find new rules with low prior confidence products that could be of value to our analysis. We wanted to see which products were bought the least to understand if the rules would give us products that should or should not be placed together.
Below is an explanation of the metric and the parameters selected to derive interesting rules:

- **Confidence ratio** is the ratio of the prior and posterior confidence (minimum over maximum) subtracted from 1. It is more sensitive to ratios in the lower confidence regions.
    - Confidence Ratio = 1 - (min (post conf, prior conf)/max (post conf, prior conf))
    - The parameters below were selected to derive rules using this metric:
        - **Minimum threshold** = 20%
            - We decided to select a 20% Confidence Ratio lower bound. The confidence ratio is not allowed to be below 20%, which we thought would give us significant rules.
        - **Minimum confidence**: 15%
            - Similar to the Confidence Difference, we decided to leave this parameter at 15% to have significant rules generated in addition to rules generated for other metrics.
        - The parameters **minimum rule support**, **minimum support**, and **maximum antecedents** were left unchanged.

- **Conclusions and Recommendations**: 33 rules were generated with these parameters (refer to **Annex 4c** for all rules) We sorted the result first by Confidence Ratio and then a second time using Lift, allowing for easier comparison to the previous analysis.

antecedents	consequent	instance	antecedent suppor	consequent suppor	confidence	rule support	lift	conf.difference	conf.ratio	ruleid
other vegetables	whipped/sour cream	186	0.238461538	0.094871795	0.155913978	0.037179487	1.643417611	0.061042184	0.391511936	1
rolls/buns	sausage	222	0.284615385	0.1	0.162162162	0.046153846	1.621621622	0.062162162	0.383333333	5
soda	shopping bags	159	0.203846154	0.1	0.150943396	0.030769231	1.509433962	0.050943396	0.3375	15
pastry	soda	78	0.1	0.203846154	0.153846154	0.015384615	0.754716981	0.05	0.245283019	8

    - Out of the 33, we selected the above 4 rules which we consider most relevant for the store management, which have a low prior confidence (products bought the least).
    - Thus, we added these 4 rules to the Lift and Confidence Difference analysis that was previously done, since we believe these rules are of value for the store manager's perspective in terms of relationships between products.
  - Based on the 4 rules, these are our recommendations:
    - The relationship between pastry and soda is negative, since the lift is below 1 (negative rule), thus placing these products together could have a negative impact on the purchase of soda. **We recommend keeping these products away from each other in the store.**
    - The other relationships between antecedent and consequent are positive (positive rules), so placing these products together will have a positive impact on the least bought products. For instance, placing soda close to shopping bags, rolls/buns close to sausage, or other vegetables close to the whipped/sour cream will have a positive impact in the purchase of the consequent in these cases. **We recommend keeping these products together in the store.**

Additional Recommendations:

After running the Apriori Algorithm to find relevant rules for your store, we also conducted prediction analysis to find the best rules for each customer. We used the tabular dataset to run the analysis and these are the conclusions:

- Predictions for 641 out of the total 780 customers were created, based on 18 different rules generated by our lift analysis.
- As you can see in the table below (consequent shown first, followed by antecedents), the majority of predictions are for the purchase of whole milk (270 predictions; with an average prediction confidence of 40%), followed by other vegetables (196 predictions; with an average prediction confidence of 29%) and rolls/buns (132 predictions; with an average prediction confidence of 36%).
- Aside from the recommendations provided in the section above deriving interesting combinations of products for your marketing offers, these additional conclusions should also be taken into consideration. This additional analysis is helpful to develop promotions or real time offers for a given customer's future visit. Given that this analysis selects the rule with the highest confidence for each customer based on the combination of products they have purchased, we recommend using the table below and these additional findings to customize marketing offers that are specifically relevant for a given customer.

Prediction Analysis table:

Row Labels	Count of predruleid	Average predconf	Average lift
whole milk	**270**	**0.406**	**1.18**
other vegetables	75	0.387	1.12
root vegetables	59	0.445	1.29
yogurt	53	0.441	1.28
pastry	30	0.359	1.04
citrus fruit	29	0.383	1.11
tropical fruit	24	0.381	1.11
other vegetables	**196**	**0.291**	**1.22**
whole milk	96	0.268	1.12
bottled water	34	0.265	1.11
root vegetables	28	0.391	1.64
tropical fruit	15	0.330	1.38
yogurt	15	0.276	1.16
citrus fruit	8	0.287	1.20
rolls/buns	**132**	**0.358**	**1.26**
soda	75	0.308	1.08
sausage	42	0.462	1.62
citrus fruit	15	0.319	1.12
soda	**37**	**0.297**	**1.46**
shopping bags	29	0.308	1.51
sausage	8	0.256	1.26
root vegetables	**6**	**0.268**	**1.90**
tropical fruit	6	0.268	1.90
Grand Total	**641**	**0.354**	**1.23**

As a conclusion, by analysing these three metrics we were able to extract 20 rules that we thought were relevant and helped us to find interesting combinations of products. We believe that if you implement our recommendations you will be able to achieve your goal of creating a set of relevant marketing offers as well as increasing the purchase of some products in your store.

Technical Annex:

Annex 1 : Tables explaining all the parameters and metrics used for our analysis

a. Explanation of the parameters

Parameter	Definition
Antecedent	An antecedent is an element found in the data sample.
Consequent	The consequent is found in combination with the antecedent.
Instances	The instance is the number of records that match the antecedent.
Antecedent Support	The antecedent support is the percentage of records that match the antecedents.
Consequent Support (Prior Confidence)	The consequent support is the relative frequency of the consequent.
Confidence (Posterior Confidence)	The confidence is the percentage of records matching the antecedent that also match the consequent.
Rule Support	The rule support is the percentage of the total records that match the entire rule.

b. Explanation of the metrics

Metrics	Definition
Lift	Lift describes the likelihood of purchasing one product if another product (or combination of products) is in the basket. It is the ratio of the posterior confidence to the prior confidence
Confidence Difference	The confidence difference is based on the absolute difference between the posterior and prior confidence, setting a lower bound.
Confidence Ratio	The confidence ratio compares the posterior and prior confidence computing their ratio (minimum over maximum) and subtracting it from 1. It is more sensitive to ratios in the lower confidence regions.

c. Explanation of the algorithm used

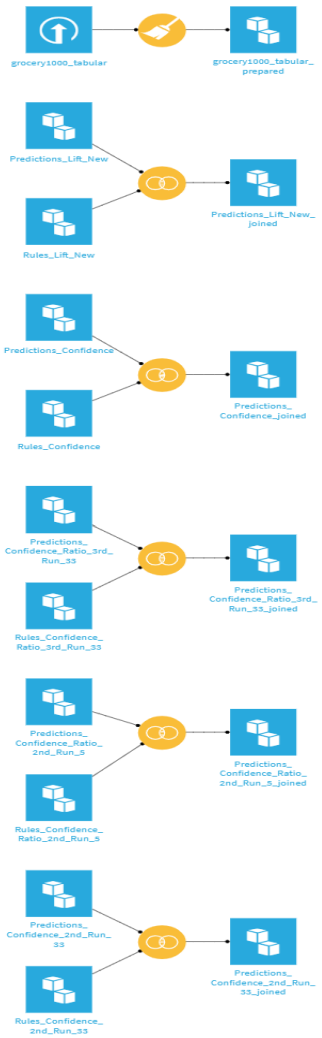
Algorithm for Association Analysis	Definition
Apriori Algorithm	The apriori algorithm is a foundational algorithm in market basket analysis. The algorithm is for finding frequent item sets in a dataset for boolean association rules. Sources: Apriori Algorithm - GeeksforGeeks Apriori Algorithm : Know How to Find Frequent Itemsets | Edureka
Association Rule Mining	Association rule mining is a procedure which aims to observe frequently occurring patterns, correlations, or associations from datasets. Association rules may be thought of as an IF-THEN relationship.

Annex 2 : Steps for Data cleansing using Dataiku software

The cleaning of data was done through the software **Dataiku**, which helped us in removing the rows we decided were not relevant for our analysis. Since we are trying to find interesting combinations of products to customize offers, we figured using customers that did not buy any product would be less relevant to our analysis. Thus, we removed these rows, by creating a **recipe** (going to **actions** on our dataset, and choosing the broom option, called **'Prepare'**). After doing so, we added a **new step** to our dataset, called **'filter rows/cells on numerical range'** in order to filter these rows, and used the option **'remove matching rows'** to remove the rows that were not relevant. Thus, comparing all the columns, we decided to remove rows with values only between -0.2 and 0.2 (since a range is required) so that rows with only 0 as values would be removed, or in other words rows with customers that made no purchases. **Running the script** was the last step, which provided us with a new dataset, including **780** customers of the total 1000, since 220 customers made no purchases at all.

Annex 3 : Flow chart dataset

This is the flow of our Datasets, having run two different scripts for confidence difference and confidence ratio.



Annex 4 : All the rules predicted for every metric

a. Here are the 19 rules generated using the metric 'Lift'

antecedents	consequent	instance	antecedent suppor	consequent suppor	confidence	rule support	lift	conf.difference	conf.ratio	ruleid
'tropical fruit	'root vegetables	97	0.124358974	0.141025641	0.268041237	0.033333333	1.900656045	0.127015596	0.473865878	18
'root vegetables	'other vegetables	110	0.141025641	0.238461538	0.390909091	0.055128205	1.639296188	0.152447552	0.389982111	19
'sausage	'rolls/buns	78	0.1	0.284615385	0.461538462	0.046153846	1.621621622	0.176923077	0.383333333	5
'tropical fruit	'yogurt	97	0.124358974	0.162820513	0.257731959	0.032051282	1.582920692	0.094911446	0.36825641	12
'shopping bags	'soda	78	0.1	0.203846154	0.307692308	0.030769231	1.509433962	0.103846154	0.3375	15
'tropical fruit	'other vegetables	97	0.124358974	0.238461538	0.329896907	0.041025641	1.383438643	0.091435369	0.277163462	16
'root vegetables	'whole milk	110	0.141025641	0.344871795	0.445454545	0.062820513	1.291652585	0.100582751	0.225798012	2
'yogurt	'whole milk	127	0.162820513	0.344871795	0.440944882	0.071794872	1.278576237	0.096073087	0.217880037	3
'sausage	'soda	78	0.1	0.203846154	0.256410256	0.025641026	1.257861635	0.052564103	0.205	17
'citrus fruit	'other vegetables	94	0.120512821	0.238461538	0.287234043	0.034615385	1.204529856	0.048772504	0.16980057	10
'yogurt	'other vegetables	127	0.162820513	0.238461538	0.275590551	0.044871795	1.155702311	0.037129013	0.134725275	8
'whole milk	'other vegetables	269	0.344871795	0.238461538	0.267657993	0.092307692	1.122436743	0.029196454	0.109081197	13
'other vegetables	'whole milk	186	0.238461538	0.344871795	0.387096774	0.092307692	1.122436743	0.042224979	0.109081197	14
'citrus fruit	'rolls/buns	94	0.120512821	0.284615385	0.319148936	0.038461538	1.1213341	0.034533552	0.108205128	9
'bottled water	'other vegetables	132	0.169230769	0.238461538	0.265151515	0.044871795	1.111925709	0.026689977	0.100659341	1
'citrus fruit	'whole milk	94	0.120512821	0.344871795	0.382978723	0.046153846	1.110495927	0.038106929	0.099501425	4
'tropical fruit	'whole milk	97	0.124358974	0.344871795	0.381443299	0.047435897	1.106043767	0.036571504	0.095876646	7
'soda	'rolls/buns	159	0.203846154	0.284615385	0.308176101	0.062820513	1.082780894	0.023560716	0.076452119	11
'pastry	'whole milk	78	0.1	0.344871795	0.358974359	0.035897436	1.040892193	0.014102564	0.039285714	6

b. Here are the 27 rules generated using the metric 'Confidence Difference'

antecedents	consequent	instance	antecedent suppor	consequent suppor	confidence	rule support	lift	conf.difference	conf.ratio	rulei
root vegetables	tropical fruit	110	0.141025641	0.124358974	0.236363636	0.033333333	1.900656045	0.112004662	0.473865878	14
tropical fruit	root vegetables	97	0.124358974	0.141025641	0.268041237	0.033333333	1.900656045	0.127015596	0.473865878	13
other vegetables	whipped/sour cream	186	0.238461538	0.094871795	0.155913978	0.037179487	1.643417611	0.061042184	0.391511936	1
root vegetables	other vegetables	110	0.141025641	0.238461538	0.390909091	0.055128205	1.639296188	0.152447552	0.389982111	20
other vegetables	root vegetables	186	0.238461538	0.141025641	0.231182796	0.055128205	1.639296188	0.090157155	0.389982111	21
tropical fruit	citrus fruit	97	0.124358974	0.120512821	0.195876289	0.024358974	1.625356438	0.075363468	0.384750337	7
citrus fruit	tropical fruit	94	0.120512821	0.124358974	0.20212766	0.024358974	1.625356438	0.077768685	0.384750337	8
sausage	rolls/buns	78	0.1	0.284615385	0.461538462	0.046153846	1.621621622	0.176923077	0.383333333	4
rolls/buns	sausage	222	0.284615385	0.1	0.162162162	0.046153846	1.621621622	0.062162162	0.383333333	5
yogurt	tropical fruit	127	0.162820513	0.124358974	0.196850394	0.032051282	1.582920692	0.072491419	0.36825641	16
tropical fruit	yogurt	97	0.124358974	0.162820513	0.257731959	0.032051282	1.582920692	0.094911446	0.36825641	15
soda	shopping bags	159	0.203846154	0.1	0.150943396	0.030769231	1.509433962	0.050943396	0.3375	11
shopping bags	soda	78	0.1	0.203846154	0.307692308	0.030769231	1.509433962	0.103846154	0.3375	10
sausage	yogurt	78	0.1	0.162820513	0.243589744	0.024358974	1.496062992	0.080769231	0.331578947	2
yogurt	root vegetables	127	0.162820513	0.141025641	0.204724409	0.033333333	1.451682176	0.063698768	0.311143984	18
root vegetables	yogurt	110	0.141025641	0.162820513	0.236363636	0.033333333	1.451682176	0.073543124	0.311143984	19
yogurt	bottled water	127	0.162820513	0.169230769	0.236220472	0.038461538	1.395848246	0.066989703	0.283589744	23
bottled water	yogurt	132	0.169230769	0.162820513	0.227272727	0.038461538	1.395848246	0.064452214	0.283589744	24
tropical fruit	other vegetables	97	0.124358974	0.238461538	0.329896907	0.041025641	1.383438643	0.091435369	0.277163462	17
citrus fruit	yogurt	94	0.120512821	0.162820513	0.223404255	0.026923077	1.372089127	0.060583742	0.271184371	9
root vegetables	whole milk	110	0.141025641	0.344871795	0.445454545	0.062820513	1.291652585	0.100582751	0.225798012	22
yogurt	whole milk	127	0.162820513	0.344871795	0.440944882	0.071794872	1.278576237	0.096073087	0.217880037	25
sausage	soda	78	0.1	0.203846154	0.256410256	0.025641026	1.257861635	0.052564103	0.205	3
other vegetables	rolls/buns	186	0.238461538	0.284615385	0.231182796	0.055128205	0.812263877	0.053432589	0.187736123	27
shopping bags	whole milk	78	0.1	0.344871795	0.269230769	0.026923077	0.780669145	0.075641026	0.219330855	12
soda	whole milk	159	0.203846154	0.344871795	0.238993711	0.048717949	0.692992916	0.105878084	0.307007084	26
pastry	other vegetables	78	0.1	0.238461538	0.153846154	0.015384615	0.64516129	0.084615385	0.35483871	6

c. Here are the 33 rules generated using the metric 'Confidence Ratio'

antecedents	consequent	instance	antecedent suppor	consequent suppor	confidence	rule support	lift	conf.difference	conf.ratio	ruleid
tropical fruit	root vegetables	97	0.124358974	0.141025641	0.268041237	0.033333333	1.900656045	0.127015596	0.473865878	17
root vegetables	tropical fruit	110	0.141025641	0.124358974	0.236363636	0.033333333	1.900656045	0.112004662	0.473865878	18
other vegetables	whipped/sour cream	186	0.238461538	0.094871795	0.155913978	0.037179487	1.643417611	0.061042184	0.391511936	1
root vegetables	other vegetables	110	0.141025641	0.238461538	0.390909091	0.055128205	1.639296188	0.152447552	0.389982111	25
other vegetables	root vegetables	186	0.238461538	0.141025641	0.231182796	0.055128205	1.639296188	0.090157155	0.389982111	26
citrus fruit	tropical fruit	94	0.120512821	0.124358974	0.20212766	0.024358974	1.625356438	0.077768685	0.384750337	11
tropical fruit	citrus fruit	97	0.124358974	0.120512821	0.195876289	0.024358974	1.625356438	0.075363468	0.384750337	10
sausage	rolls/buns	78	0.1	0.284615385	0.461538462	0.046153846	1.621621622	0.176923077	0.383333333	4
rolls/buns	sausage	222	0.284615385	0.1	0.162162162	0.046153846	1.621621622	0.062162162	0.383333333	5
yogurt	tropical fruit	127	0.162820513	0.124358974	0.196850394	0.032051282	1.582920692	0.072491419	0.36825641	20
tropical fruit	yogurt	97	0.124358974	0.162820513	0.257731959	0.032051282	1.582920692	0.094911446	0.36825641	19
shopping bags	soda	78	0.1	0.203846154	0.307692308	0.030769231	1.509433962	0.103846154	0.3375	14
soda	shopping bags	159	0.203846154	0.1	0.150943396	0.030769231	1.509433962	0.050943396	0.3375	15
sausage	yogurt	78	0.1	0.162820513	0.243589744	0.024358974	1.496062992	0.080769231	0.331578947	2
yogurt	root vegetables	127	0.162820513	0.141025641	0.204724409	0.033333333	1.451682176	0.063698768	0.311143984	23
root vegetables	yogurt	110	0.141025641	0.162820513	0.236363636	0.033333333	1.451682176	0.073543124	0.311143984	24
bottled water	yogurt	132	0.169230769	0.162820513	0.227272727	0.038461538	1.395848246	0.064452214	0.283589744	30
yogurt	bottled water	127	0.162820513	0.169230769	0.236220472	0.038461538	1.395848246	0.066989703	0.283589744	29
other vegetables	tropical fruit	186	0.238461538	0.124358974	0.172043011	0.041025641	1.383438643	0.047684036	0.277163462	22
tropical fruit	other vegetables	97	0.124358974	0.238461538	0.329896907	0.041025641	1.383438643	0.091435369	0.277163462	21
citrus fruit	yogurt	94	0.120512821	0.162820513	0.223404255	0.026923077	1.372089127	0.060583742	0.271184371	13
yogurt	citrus fruit	127	0.162820513	0.120512821	0.165354331	0.026923077	1.372089127	0.04484151	0.271184371	12
whole milk	root vegetables	269	0.344871795	0.141025641	0.182156134	0.062820513	1.291652585	0.041130493	0.225798012	27
root vegetables	whole milk	110	0.141025641	0.344871795	0.445454545	0.062820513	1.291652585	0.100582751	0.225798012	28
yogurt	whole milk	127	0.162820513	0.344871795	0.440944882	0.071794872	1.278576237	0.096073087	0.217880037	31
whole milk	yogurt	269	0.344871795	0.162820513	0.208178439	0.071794872	1.278576237	0.045357926	0.217880037	32
pastry	citrus fruit	78	0.1	0.120512821	0.153846154	0.015384615	1.276595745	0.033333333	0.216666667	6
pastry	yogurt	78	0.1	0.162820513	0.205128205	0.020512821	1.25984252	0.042307692	0.20625	7
sausage	soda	78	0.1	0.203846154	0.256410256	0.025641026	1.257861635	0.052564103	0.205	3
shopping bags	whole milk	78	0.1	0.344871795	0.269230769	0.026923077	0.780669145	0.075641026	0.219330855	16
pastry	soda	78	0.1	0.203846154	0.153846154	0.015384615	0.754716981	0.05	0.245283019	8
soda	whole milk	159	0.203846154	0.344871795	0.238993711	0.048717949	0.692992916	0.105878084	0.307007084	33
pastry	other vegetables	78	0.1	0.238461538	0.153846154	0.015384615	0.64516129	0.084615385	0.35483871	9

Annex 5 : Prediction Analysis Table by rule id

Row Labels	Count of predruleid	Average predconf	Average lift
whole milk	270	0.406	1.18
other vegetables	75	0.387	1.12
19	75	0.387	1.12
root vegetables	59	0.445	1.29
13	59	0.445	1.29
yogurt	53	0.441	1.28
16	53	0.441	1.28
pastry	30	0.359	1.04
3	30	0.359	1.04
citrus fruit	29	0.383	1.11
6	29	0.383	1.11
tropical fruit	24	0.381	1.11
11	24	0.381	1.11
other vegetables	196	0.291	1.22
whole milk	96	0.268	1.12
18	96	0.268	1.12
bottled water	34	0.265	1.11
14	34	0.265	1.11
root vegetables	28	0.391	1.64
12	28	0.391	1.64
tropical fruit	15	0.330	1.38
10	15	0.330	1.38
yogurt	15	0.276	1.16
15	15	0.276	1.16
citrus fruit	8	0.287	1.20
4	8	0.287	1.20
rolls/buns	132	0.358	1.26
soda	75	0.308	1.08
17	75	0.308	1.08
sausage	42	0.462	1.62
2	42	0.462	1.62
citrus fruit	15	0.319	1.12
5	15	0.319	1.12
soda	37	0.297	1.46
shopping bags	29	0.308	1.51
7	29	0.308	1.51
sausage	8	0.256	1.26
1	8	0.256	1.26
root vegetables	6	0.268	1.90
tropical fruit	6	0.268	1.90
8	6	0.268	1.90
Grand Total	641	0.354	1.23