Advanced Business Decision Support Systems

Professor Deepu Philip
Department of Industrial Engineering and Management Engineering
Indian Institute of Technology, Kanpur
Professor Amandeep Singh
Imagineering Laboratory
Dr. Prabal Pratap Singh
Indian Institute of Technology, Kanpur

Lecture 42

Market Basket Analysis: MS Excel Demonstration

Welcome to the third lecture of week 12. We discussed the Market Basket Analysis in the last lecture. And, I discussed about an algorithm to understand and calculate the lift from the support and confidence to understand or to select the baskets which are most suitable for our client range. Now, this algorithm is known as Apriori algorithm. Apriori algorithm which is most widely used and most well-known for the association rules. It is a very popular algorithm.

Other algorithms are also there. I will just mention the names here like the AIS algorithm is there. Then, we have the SETM algorithm and we have an algorithm for Frequent Pattern growth known as FP growth algorithm. In the Apriori algorithm, as I discussed in the last lecture, we have to support confidence and lift. This helps us to find the frequent item sets in transaction and identify the rules between these items. The limitation of this algorithm is the frequent item set generation which I have generated here some data.

Not generated, I have taken it from the very famous data set available online that is the Titanic data. I have converted that into a CSTR data. I will just explain what data set or the table I have here. So, this algorithm needs to scan the database many times leading to increase time and reduce performances and computational cost is also sometimes high because we need to use the concepts of first support and confidence and do not only we can understand or calculate lift.

Next comes the AIS algorithm, this creates multiple passes of the entire database or the transactional data, during each pass it scans all the transactions. So, first pass, second pass, it goes in certain passes and we try to generate a large item set in a transaction database.

It is focused on enhancement of the databases with necessary performance to process decision support. This algorithm finds whether an association between items exists or not even. So, only again the disadvantage of this algorithm is that it generates too many candidates, which means too many sets that turn out to be small later. So, also the data structure is to be maintained for this. Then, comes the SHEM algorithm. It is similar to what AIS algorithm does, that is it also creates a collective pass over databases and tries to generate small item sets.

Now, the data sets are a large term which I menti1d in the previous lecture as well TID. TIDs are Transaction IDs of the generated transactions that we are generating for the candidate item sets. Benefits of this algorithm is that the arrangement of candidates or item sets together with TIDs is in a sequential manner. But again a huge number of TIDs are generated, so the data set becomes larger. So, there is an alternate algorithm which is an advancement of Apriori algorithm that is FP growth.

FP growth is Frequent Pattern growth. So, it represents the FP tree or we call it simply Frequent Pattern only in which there is no need for candidate generation to generate a Frequent Pattern. This Frequent Pattern tree maintains the association between the item sets and this Frequent Pattern tree, that is the FP tree is a structure that is made with the earlier item sets of the data.

The association of these nodes with the lower nodes, with the higher nodes, a lot of connections are made and we try to create a tree that is kept as a record and that could be also used for the future calculations again. So, these algorithms, as I talked about the major applications of them, I will now use them to have a POS decision that is a Point-of-Sale decision for stationary data.

I will only try to use the Apriori algorithm, in which support, confidence and lift would be taken. How are these taken let us try to see. So, let us recall what is support? Support is just the plain basic probability of an event to occur, if out of the 10000 items sold, 2000 items are of our interest, 2000 by 10000 this becomes, 20 percent becomes our support. It is simply the probability that we are talking about. So, if we analyze the transaction table here, we will try to see what the support is and also, we will try to put the key element that we are trying to work upon.

This is the purchase, the sales, if I put the station data here, the sales of pencil, notepad or so. And, also, we will try to see, the left-hand side and right-hand side are equal or not. And, we will try to compare that, and we will try to see 2-way lift and 3-way lift. I will put certain formulas in MS Excel. There could be even 4-way lifts, it could be 5-way lifts, multiple lifts could be there. That means, 1 item is being compared with the other 2 parameters. The other 2 parameters, I will try to explain the parameters, then I will discuss them. So, this is support.

Then comes Confidence. Confidence is a consequent event, in which the antecedent event, antecedent eventis the 1 that I am trying to call as key event. It can be described by the conditional probability, that is the probability that if event A is happening given that event B has already happened, that is, they are purchasing the erasers provided they have already taken the pencils. So, that is combined purchases are there. So, based upon both the confidence and support, we try to get the value of the lift, which I said should be greater than 1.

Lift more than 1 means there are more chances of the items being purchased together than just occurred by the simple probability. Simple probability is a support or the confidence that we get, but lift is based upon the simple probability that we have taken from the past data. What is the higher probability of being purchased together of item, that is item 2 based upon the antecedent event. Antecedent item that is the prior item or the key item that I am trying to say. So, lift is trying to say the consequent event,

consequent event and in the support, we were really trying to see the antecedent event. So, lift is mainly the measure of how likely an item is purchased when another item is purchased, already controlling how popular both items are. So, these were all calculated in the previous lecture.

Let us now try to understand the data set I have here. So, I have critiqued or maybe tweaked the data that is the Titanic data available online. I will put the link in the slides in the notes of this week from which the data was taken and this data is now tweaked to the stationary sales, I have named it as stationary data.

You can see the sheet name here. So, what is the data here, there are 3 schools 1, 2, 3, maybe let us try to rank the schools based upon the school fee they have, maybe school 1 might be having a fee more than 6 to 7000 per month, school 2 might be having around 3000, school 3 is might having maybe rupees 1000 fee, maybe I could say I am putting the schools into the classes. So, school 1, 2, 3 are there, and the candidate who is coming for purchase is male or female as you can see these are all put into the binary figures. Everything is binary here 0 or 1.

Here, 0 means it is not existing. 1 is the item existing, that is wherever you find 1, in school 1 at school 2 or school 3, that means this value would be possible either in school 1, 2 or 3 only. You can see any of the rows here. Let me say suppose row 9 is heavy. It is from school 3 or so. Also, when we say 0 or 1, it is also termed as true and false, in the terms of excel. I will try to come, when I will put it in the formula.

So, school 1, 2, 3 are there, whether it is a female or not. If it is 0, it is male. If it is 1, it is female. The age of the person is noted when they come for purchase, maybe they are given some gift card where they might have to put their date of birth or so, the age is noted there, then whether the person who has come is child or adult or elderly, we can put the age ranges here. For instance, I can say any1 lower than the age of 16 is a child here, from 16 to 45 might be an adult, 45 or why I am calling it elderly only.

So, these 3 ranges are there. Then, do they come single, double or triple or in a group? Means, this could be a group of friends coming together, different children are coming together to purchase stati1ry items. Stationary items, as I talked about some examples, in the month of April, when sales of notepads, sales of pencils, sales of pens and erasers are more.

So, are they coming in groups to purchase or are they coming single or are they bringing some other person with them. Single, double, triple or more than 3 in group, how are they coming for the purchase. Then, if the child is coming, whether the child is accompanied by a parent or not, this is child parent. This is a child parent and I have a parent child as well. If a parent is coming whether they are bringing a kid with them or not, this is a parent child.

So, then we have no ages and finally, what we have the column P which is purchased that is, they have purchased the items of our interest or may be the items which are sold together that we have put in the baskets. So, purchase means it is a success. So, it is again 0 and 1 here as I said success and failure. So, it is again 0 and 1 here as I said failure and success here.

Let us now talk to see or create the pivot table. Not exactly the pivot table, but the formula table that we will try to use in the solver to have or understand the associations between different parameters here. It is now first to see what are the total number of transactions. I can even count it, but I am trying to just show you physically, there are 892 rows filled here. The first row is only the header row. So, that means, total number of transactions here are 891. I would say total transactions these are 891.

Then, let me just put the data in the feature frequency lookup table. So, what we have here is starting from column A to P which means, we have 16 different features here and the feature that we have taken as our antecedent feature is purchased whether the item is purchased by any of these groups of the other features or the single feature. If it is a group of the features, I will put them into left hand side that could be 2 way. If 1 is compared with 1, it is 2 way, if 1 is compared with 2 other items, that becomes 3 way, if 1 is compared with 3 other items that becomes a 4 way and, so on.

For instance, if I put purchased in the right-hand side that is my antecedent event and the consequent events is only maybe if the child has come for it, this becomes a 2-way lift and if I put purchased as my antecedent event and I select more than 1 feature, more than 1 means I select 2 features at least that is, I say from school 3 and child and I try to compare them, this becomes a 3 way lift, it is antecedent is compared with 2 consequent events.

Now, let us try to put the data in the table. I will first call it feature index which is just a number and from A to P that means, we have 16 features, I will push the number 1, 2 and, so on. I just drag it down to the number 60, fine.

So, what is this feature? The feature name only I need to put here, so I can just say, I am selecting specific column, I will say here this is equal to the header name. I will select only A1 that is, school 1 and say is equal to, it has selected this feature. You can see, this in this sheet titanic data that is, a sheet name it has selected header school 1 that is the first header in the column A is selected.

Similarly, I can select put escape the other feature here this is equal to second column this is equal to column 3 or column C here this is equal to column D, so on. So, fine, feature names have now come here. Now, let us try to see the frequency. These features are available frequency means, out of the total items that is total transaction which has happened in this stationary store, how many people have come from school 1, how many have come from school 2 and school 3 and, so on that means school 1, school 2 school 3 should be equal to 100 percent of the people.

So, frequency is just a percentage as I said, so it is taken out of the total transaction 891. I have just copied the formula here, some of the formulas which are little large this is some indirect from the data sheet that is a titanic data. I will select the respective column and divided by the total number of transactions; I will just copy this formula here. So, I will just come here, copy this formula and paste it here for the frequency.

So, it has selected the S7 item. School 1 is trying to understand the frequency of it has to be divided by S3 here, fine. Now, it is correct, you can see the formula closely here it is some of the indirect from the

titanic data sheet, the column or the item put in the cell S7 that is, school 1 from the titanic data, how many times it has come and it is divided by the total transactions that is the cell S3.

So, similarly let me see if I drag it down, yes it has copied the whole formula because we have locked already the cell S3 here while putting dollar sign, it is locked already and it is only selecting the respective cells put as features here, I will just try to bold this row control + B. It has come now frequency here.

Now, let us try to see the 2-way lift while putting on right hand side, only the items are purchased or not. This is my antecedent event and the consequent events; we will select 1 of those available in the columns. So, in the 2-way lift, I say let me see what is my left-hand side that is my consequent event, then what is on right hand side, that is my antecedent event. Right hand side is only the items which are purchased.

So, I will just put this value here, this is equal to or I will just write purchased here and this is true for all the items, "you can pay a small attention here". I have dragged it down to only 15 items. If you see this table here, I have dragged it down to 15 items because out of the 16 features which are available, one has G1 to right hand side and remaining with whom they are being compared are only 15 left.

So, I have left one column or one cell blank here right only 15 items are taken what are these 15 items. I will just copy this column here and these are the left-hand side. If school 1 people have purchased, school 2 people have purchased, school 3 people have purchased, female have purchased. So, these are the combined effects of the lift that we are going to take in the forthcoming columns here.

Next is occurrences at how many times the specific item has occurred for this also the formula was long, I have copied it here. So, I will just copy this formula from here and paste it here, it is missing something, so what is it missing? It is missing the reference values reference values are nothing but the two columns that we have selected the left-hand side column and the right-hand side column, right. So, here the left-hand side column the reference is to be given which is my column, this W7 and for the right-hand side the column is X7.

So, occurrences which are being shown here are 136 times. The school 1 persons have purchased out of the available data. They have purchased 136 times the set of the items that we have put in the specific basket. So, because I have not locked any of the items here neither W7 nor X7 if I just drag this formula, it should be copied for all the 15 features yet it has come here. So, occurrences are there, right. So, the 2 way lift basic data table is ready, I will put them into borders, right.

So, let us try to see the calculations for 2 way lift further. So, what is the left-hand side frequency for 2-way lift? So, this because column would be used for 2-way lift and 3-way lift all of them. I will try to just drag it here and call it 2 way lift specific calculations.

So, what is the left-hand side frequency? Left hand side frequency and similarly, I would have right hand side frequency and I will see number of transactions that is total number of transactions as you remember in the support, we divide everything by total number of transactions. Then, we say predicted transactions

and we will calculate lift from here, right. This table would be used for the 15 features left hand side frequency. When I say frequency, number of times it has occurred that will be proportional to the percentages it has come.

So, it is just a frequency, the percentages which are available here if I divide it by 100, then also the frequency, the probability would come or I can say or put the V look up loop here. I would say this is equal to V look up between W7 and S7 to T22 and I will say out of two, I am putting here false. So, it is given here true means, approximate match false means the exact match. So, it has given the frequency.

You can see the formula once again W7, how many times it has occurred W7 that is, this item; how many times has it occurred and S7 to T22 that is, in this set of the 2 columns and I have selected 2 that is, the feature which I am trying to select is 2 of the features I need to see proofing true and true and false, here it is false.

The difference here is when I say true in V look up formula, it is the approximate match. It will try to see but we are not talking about any approximation, we are talking about the deterministic model that is, why we say false which means, the exact match could be given, right. So, this is the V look up loop that we have put here and I have not locked any cell here and if I drag it down, it should show me the whole frequency, yes, it is proportional to whatever percentage is what for instance, you can say it is 24.2 percent here and it is exactly showing 0.242 and maybe for another, I think if I say for female, it is 35.2 percent. It is showing 0.352 is the frequency in the left-hand side.

Let us try to see the right-hand side frequency also accordingly, so let me just copy this formula and only change the parameters for V look up. So, we need to select from X7 and between the columns again S6 and S7; true and false, it is showing the frequency as 38 percent which is also given here 38.4 percent here.

You can also see in the cell T22, 38.4 percent, this is 0.383 which is equivalent to that and I am going to lock the values here because we are not going to change anything for the further rows but anyway, it is given per changed in all the cells if I do not even lock, it will copy the same formula and show me the right-hand side frequency, yes, I have got the right-hand side frequency and number of transactions are 891. I am just putting it directly here and copying it here. Then, comes the Predictive Transactions and lift.

What is Predictive Transaction? Simply, what we are trying to calculate here is the number of times the specific event has occurred that is, we are trying to multiply the left-hand frequency with the right-hand frequency and with the number of transactions.

So, that means, Predictive Transaction is equal to the product of the left-hand frequency into right hand frequency into the number of times the transactions have occurred. So, this is 82.90, I can say it is the probability or the number of transactions it has come as 82.90 or almost 83 percent. What was Lift? Lift was the number of occurrences and the total predictive transactions that they have taken.

So, this means, this becomes the number of times it has occurred I will say this is column Y7 divided by column predictive transactions. So, here the lift value is 1.64 which is more than 1 and when the lift is more than 1, that means, just while comparing it to the chance, it is 1, it is having more probability of being lifted or being purchased. if I compare this tool, 1 percent people as purchased the specific set of items that were put in the basket. So, similarly, I can calculate the lift for the all other features or all other parameters. Here, the different parameters are there and we can calculate the lift accordingly. I will just drag this form down here and you can see the lift has come in different ranges here.

So, it has value more than around 1.7 is there, 1.93 value is also there which is large value that is around 2 lift value has come and it is as low as 0.70 or may be the lowest is 0.363. I can also see or maybe I can look for the maximum of these values and minimum of these values, this is equal to maximum of these set of values, it is at is the maximum lift that I have gotten is 1.93 and I will say max, I will say min, the minimum of these is equal to minimum of these values. I

can even put the first max, second max and often or not or I am let me say, I can even color them accordingly. The way it has come, you can see the maximum value is 1.93, I will just color it may be little dark green. After that, what we have is 1.70, I can may be also color it is the second anything more than 1.7 I am trying to color it dark green.

So, anything more than 1.67, I am trying to color it little less or little lighter green. So, on I can even color the items which are least as red that is, these should not be put in the basket at all. So, this is just for my visualization for trying to understand that these are the items or these are the pair that should not be put that is, for school 3, it is very less likely to be purchased because as I said, the standards of the schools are there school 1, school 2, school 3.

The people who have come to a specific store, it may say, it is a store where only high-quality items are available and the people who have come from may be mediocre schools or lower schools do come less to these kinds of stores.

So, it is 0.63 is marked red here. Similarly, 0.76 could also be marked red or, so in between I can put yellow colors may be a value close to exactly 1 on 1.2 I can put an intermediate color between red and green. So, I am putting it as yellow here, so on I can color them accordingly. So, all the items could be colored, so on it is only not 1 or 2 items, let me only complete it for you, I will color all the items.

So, anything less than 1.4, I will term it as yellow. This is yellow, this is yellow, right. Then, these two are both less than 1. Anything less than 1 has to be red. So, I am something very close to 1, this also I am putting red only. So, this is the 2-way lift, it has come similarly, we can also calculate the 3-way lift let us try to see the 3-way lift and I will not try to use my MS Excel solver for calculating the 3-way lift.

Before that, let us try to prepare the table for the 3-way lift. I will say 3-way lift and let me try to prepare the table for that the left-hand side right hand side and I can select the feature index here, I will say feature index or I would call it as feature 1 index. Similarly, I would have feature 2 index that means, 2

items are taken here. Now, I will number the columns between the 3 and I will select the columns between 1 to 15 or 1 to 16. All the columns will be numbered.

So, let me say, I will put some column here may be 3 and 4, how do I connect these numbers to the number of the column here. So, let us talk about feature one because we are talking about 3 way lift here. Now, we will talk about 3 features: feature 1, feature 2, feature 3 and we will also need to have the occurrences of them.

Along with this, we will have the left-hand side frequency, the right-hand side frequency and lift. So, this would make my table ready once I put the formulas here for putting the values into solver. So, let me try to understand this, so feature 1, I will say, yes feature 1 has to be these numbers that I have put here, these 2 numbers, I have just put some random numbers here. These number would be taken from this set here. You can see feature index; the feature index numbers are their feature name is there, right.

From here, we will pick it and we will put the V look up loop once again, here we will say V look up this cell that is W1 and W2. So, this cell W28 comma among these two columns these 2 columns are selected, I will say exact match that means, I will put false here, yes. So, number 3 is for scubel 3, it is correct if I change this number 5 let me say, whether it changes, yes number of changes is for when age is missing for them, it is correct now the same formula I can copy it here. So, here I need to select S8, these two columns are changed because these one is locked. So, I will again select these columns and say yes feature 2 is also selected, now feature 3, I can select it from there or I can just only put because my antecedent event is only purchased.

So, I will say it is only purchased when you put it physically or manually, make sure that the spellings are correct now, then come the occurrences is a long formula. For that, I have put that formula already here, I just copy and paste this formula here.

Now, it needs to take the values correctly that means, number of times the things have occurred. So, it has to be the two features that we have put here. So, here, I will put the value as this feature and here it would be the number of times that I have put it, this feature and here I have the third feature, let me see whether it works. Now, still something is missing.

So, it is some indirect attending data. So, it is still not showing the exact value. Something is missing, it is divided by this spell cell S2, it should be S3 cell. So, I am putting the cell S3 yes, the currencies have come now because the S3 cell is containing the total number of transactions, this is cell number S3 here. Now, left hand side frequency also, I have put a formula.

For that, here left-hand side frequency has this formula that is two features. I am putting on the left-hand side now and on the right-hand side I am putting only 1 feature. So, I can say or I can remove may be right hand side from here. So, the two features which are put in left hand side are feature 1 and feature 2.

Fine, left-hand side frequency has come. So, which means, on the right-hand side, I am putting purchase on the left-hand side, I am putting two features which are here, feature 1 and feature 2 are selected as age missing and female. If I change this number, the index number, this feature could be changed for instance I will say let me compare 3 and 7 enter.

So, you can see the 3 is school 3 and 7 is whether the person is adult or not accordingly occurrences are left hand frequencies are changed. Right hand frequency would always remain same that is, for purchased which is the number of times, we have purchased item. I will just copy the right-hand side frequency from here only.

So, I can just lock this formula for right hand frequency here. I will put all dollar signs and could select the formula from here only because this is the fixed value this be selected and I am pasting it here. So, right hand frequency is also there. Now, lift is my occurrences divided by the product of the left-hand side and right-hand frequency is.

So, this is equal to occurrences divided by (I will put it in a bracket) the product of left-hand side frequency and right-hand side frequency right. So, lift here has come 0.53 this is my table ready to go as an input to the left-hand side frequency. So, now, I will put this in the table ready to go as an input to the solver. Let me try to now come to the solver. So, let me only compare these three only from school 3 and adult whether they have purchased the item and the how and what is the maximum chances of the lift here.

Let us try to see this in the solver now. I will come to data and I will come to data, I will first delete or reset all of them. So, I need to first set the objective. So, what is the objective function? I have here objective function is my lift. So, this value is lift and this is to be maximized, right. Maximize the cell AC31 that is a lift value subject to the constraints. The constraints here are that the 2 values, 2 numbers that I have selected here. First, I need to select the variable cells here.

The variable cells are only these two index numbers that I have selected here, these are to be only varied, right. Now, there are certain constraints which are need to be added here. Ideal constraint 1 that is, these two numbers which I am varying are to be less than equal to 15 that is, maximum number of comparisons that I am trying to do in the 3-way lift is these 2 numbers on the left-hand side.

These 2 numbers cannot take any value, it cannot take value 100, it cannot take value more than 15 even 16 values could not come because 16th value has G1 as the right-hand side or the antecedent that is purchased. So, these are to be less than or equal to 15. I will say ok, then second constraint I will add is that. these 2 values are to be only the integer.

So, this is integer. this is like a just an equal to 15. Also, these values cannot be 0. So, let me add another constraint that these two values or the features which I am varying should be greater than or equal to 1, right. You can see here the which means minimum value is 1, maximum value is 15 for any of these 2 values, right.

Now, 1 more constraint I will put here let me say, the occurrences that I have should be more than some value let me say, the occurrences which are there should be greater than or equal to at least 2 percent. I will pick not the linear programming here, the way I did in the last lecture demonstration simplex and lim program, you were selected.

I will now select the evolutionary and I will try to see whether what options are there, you can say in the evolutionary, the convergence is 0.0001 that means, number of times it has occurred, then mutation rate is 0.075, I can say the number of times it has taken, I will just make it little larger 0.5 and yes population size is 100 random seed is 0, it is taking seed by itself in maximum time without improvement is 30.

I think now my excel sheet is ready now, when I say solve the excel sheet, it will try to solve it two times you can see first for the feature 1, second for the feature 2 and it will compare with the feature 3. Feature 3 is my right-hand side which is my antecedent event and for each feature, it will try to see the values in all the rows that is, 891 times will calculate for feature 1 and 891 times will calculate for feature 2.

Let me try to solve this excel solver, so it is now calculating if you can see here, you can see the sub problem, it is running here down left corner, it will take some time. It has G1 through one set of the iterations that is 891 into 2 times, it will go that means, around 1900 calculations will be taken. Excel will show up the lift that is, maximum value that we are going to take and the results would be shown here. Now, the solver has completed the solution.

Now, you can see keep the solver solution, I will say yes, so this is what we have gotten from the solver solution. So, I am just selecting these rows or I am coloring them because these are taken into calculations. These 2 are taken into calculations. These occurrences are G1 into calculations for the solver and we have finally taken this objective function which is my lift value, so maximum lift is 2.5 which is high value.

So, it is showing that for school 1 that is, feature 1 and feature 4, 2.5 is the maximum lift. If you put them together, so this is the way it should go but it is just showing that maximum lift has come for this. Similarly, we can change these features and try to may be calculate from something else. Let me say, if not for school 1, I am trying to calculate it for school 2 and only the female people there, so it is also showing lift for the school 2 and female.

Similarly, if I say if not female, let me say, fully child for children comes for purchase, so number for that is child number here is 6, so I will say, 2 and 6 children from school 2 come only for purchase if I say I say enter it is showing lift 2.37.

So, all the lift value which are showing here are more than 1 that means, these baskets are good to be put and the maximum lift it is calculating. Also, it can show further if you try to use the python program or may be the R program, it can show further options of the lift or the basket, you can set for instance, for one set it is 2.3.78 for another set it could be 2.1, then it could come 2.0 or, so that order of the lift can also be taken using some other program as well, so here we have used the MS excel solver program to understand how do we conduct the Market Basket Analysis.

The data sheet that I have generated here that will be shown you or given to you in the lecture notes and also, we will try to see the course summary in the last lecture where we will try to just go through whatever we have covered in the course, so that you prepare well for your exam thank you.

	School 1	School 2	School 3	Female	AgeMissing	Child	Adult	Elderly	Single	Double	Triple	Group	Child parent	Parent child	NoAges	Purchased
2	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0
3	1	0	0	1	0	0	1	0	0	1	0	0	0	0	0	1
4	0	0	1	1	0	0	1	0	1	0	0	0	0	0	0	1
5	1	0	0	1	0	0	1	0	0	1	0	0	0	0	0	1
6	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0
7	0	0	1	0	1	0	0	0	1	0	0	0	0	0	1	0
8	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0
9	0	0	1	0	0	1	0	0	0	0	0	1	1	0	0	0
10	0	0	1	1	0	0	1	0	0	0	1	0	1	0	0	1
11	0	1	0	1	0	1	0	0	0	1	0	0	1	0	0	1
12	0	0	1	1	0	1	0	0	0	0	1	0	1	0	0	1
13	1	0	0	1	0	0	0	1	1	0	0	0	0	1	0	1
14	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0
15	0	0	1	0	0	0	1	0	0	0	0	1	1	0	0	0
16	0	0	1	1	0	1	0	0	1	0	0	0	1	0	0	0
17	0	1	0	1	0	0	0	1	1	0	0	0	0	1	0	1
18	0	0	1	0	0	1	0	0	0	0	0	1	1	0	0	0
19	0	1	0	0	1	0	0	0	1	0	0	0	0	0	1	1
20	0	0	1	1	0	0	1	0	0	1	0	0	0	0	0	0
21	0	0	1	1	1	0	0	0	1	0	0	0	0	0	1	1
22	0	1	0	0	0	0	1	0	0	1	0	0	1	0	0	0
23	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	1
24	0	0	1	1	0	1	0	0	1	0	0	0	1	0	0	1
25	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	1
26	0	0	1	1	0	1	0	0	0	0	0	1	1	0	0	0
27	0	0	1	1	0	0	1	0	0	0	0	1	1	0	0	1
28	0	0	1	0	1	0	0	0	1	0	0	0	0	0	1	0
29	1	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0
30	0	0	1	1	1	0	0	0	1	0	0	0	0	0	1	1
31	0	0	1	0	1	0	0	0	1	0	0	0	0	0	1	0
32	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0
33	1	0	0	1	1	0	0	0	0	0	1	0	0	1	0	1
34	0	0	1	1	1	0	0	0	1	0	0	0	0	0	1	1
35	0	1	0	0	0	0	0	1	1	0	0	0	0	1	0	0
36	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0
37	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0

Total transactions 891

Feature Frequency Look-up table

Feature Index	Feature	Frequency
1	School 1	24.2%
2	School 2	20.7%
3	School 3	55.1%
4	Female	35.2%
5	AgeMissing	19.9%
6	Child	12.7%
7	Adult	62.7%
8	Elderly	4.7%
9	Single	54.0%
10	Double	20.3%
11	Triple	11.3%
12	Group	14.4%
13	Child-parent	23.6%
14	Parent-child	8.5%
15	NoAges	16.4%
16	Purchased	38.4%

TRUE: Approximate match
FALSE: Exact Match

2-way lift

LHS (consequent event)	RHS (antecedent event)	Occurrences	LHS frequency	RHS frequency	Transactions	Predicted Transactions	Lift
School 1	Purchased	136	0.242424242	0.383838384	891	82.90909091	1.640350877
School 2	Purchased	87	0.20650954	0.383838384	891	70.62626263	1.231836384
School 3	Purchased	119	0.551066218	0.383838384	891	188.4646465	0.631418158
Female	Purchased	233	0.352413019	0.383838384	891	120.5252525	1.933204827
AgeMissing	Purchased	52	0.198653199	0.383838384	891	67.93939394	0.765388046
Child	Purchased	61	0.126823793	0.383838384	891	43.37373737	1.406380997
Adult	Purchased	216	0.627384961	0.383838384	891	214.5656566	1.00668487
Elderly	Purchased	13	0.047138047	0.383838384	891	16.12121212	0.806390977
Single	Purchased	130	0.539842873	0.383838384	891	184.6262626	0.704125178
Double	Purchased	93	0.203142536	0.383838384	891	69.47474747	1.338615877
Triple	Purchased	66	0.11335578	0.383838384	891	38.76767677	1.702449192
Group	Purchased	53	0.14365881	0.383838384	891	49.13131313	1.078741776
Child-parent	Purchased	103	0.235690236	0.383838384	891	80.60606061	1.277819549
Parent-child	Purchased	41	0.085297419	0.383838384	891	29.17171717	1.405470914
NoAges	Purchased	40	0.163860831	0.383838384	891	56.04040404	0.713770728

Max	Min
1.933205	0.631418

3-way lift

LHS	RHS
Feature 1 Index	Feature 2 Index
2	6

Feature 1	Feature 2	Feature 3	Occurrences	LHS frequency	RHS frequency	Lift
School 2	Child	Purchased	0.023569024	0.025813692	0.383838384	2.378718535

Formulae		
Freq	Occurrenc	LHS Freq
#DIV/0!	#VALUE!	
	#DIV/0!	#DIV/0!

1. Apriori – Support (Antecedent event), Confidense, Lift (Consequent event) — PoS
2. AIS
3. SETM (TIDs) – sequential manner
4. FP Growth: Frequent Pattern Growth, FP Tree