**Assignment 1**

**Team Number : 04**

**Team Members: NetId**

Akanksha Mishra axm170031

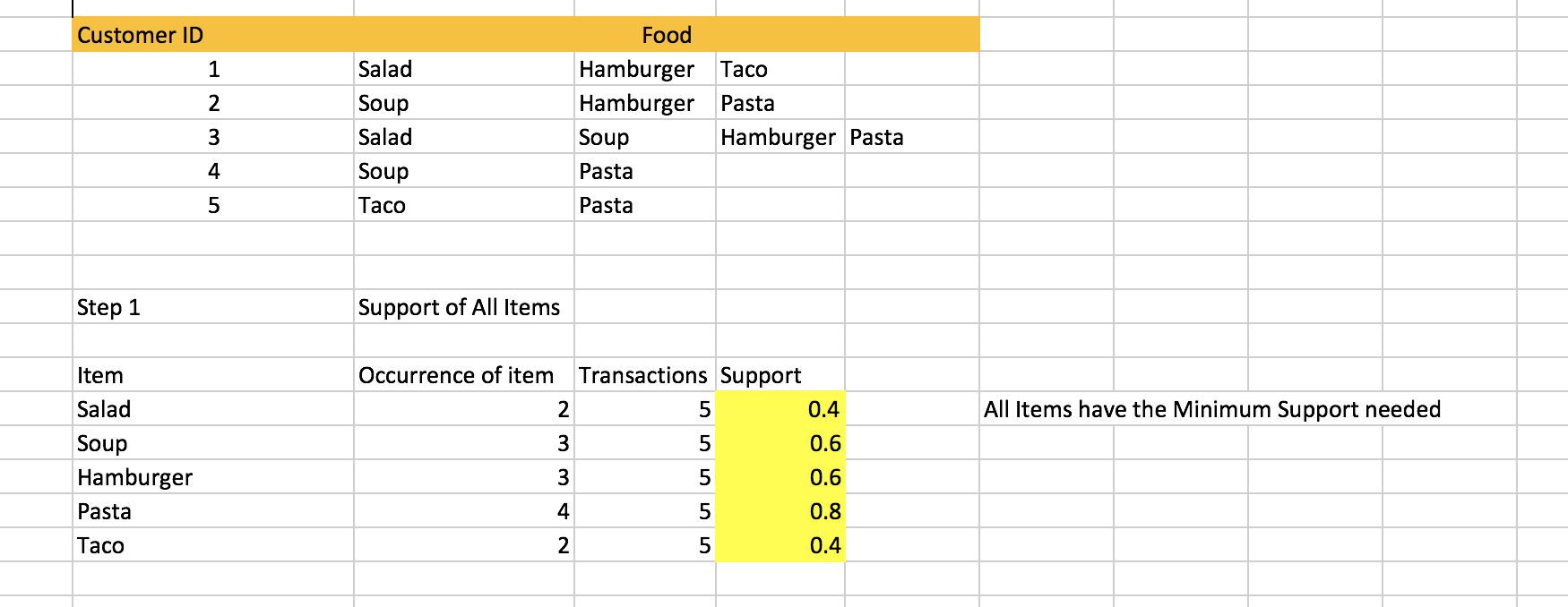
Harman Singh hxs134930

Lokesh Kumar Chaturvedi Lolla lxl163230

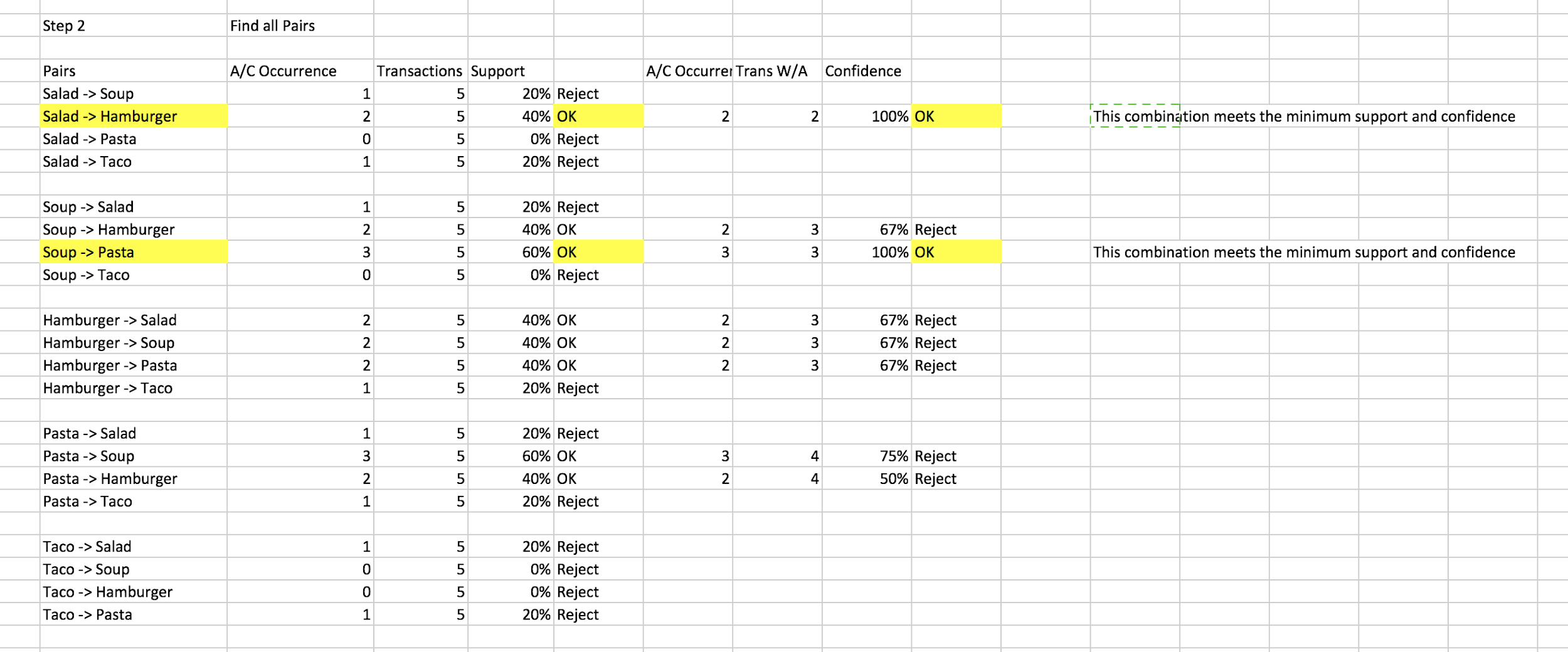
Yugandhara Rawool yer170030

**Solution 1:** Below are the screenshots of the calculation performed using excel.

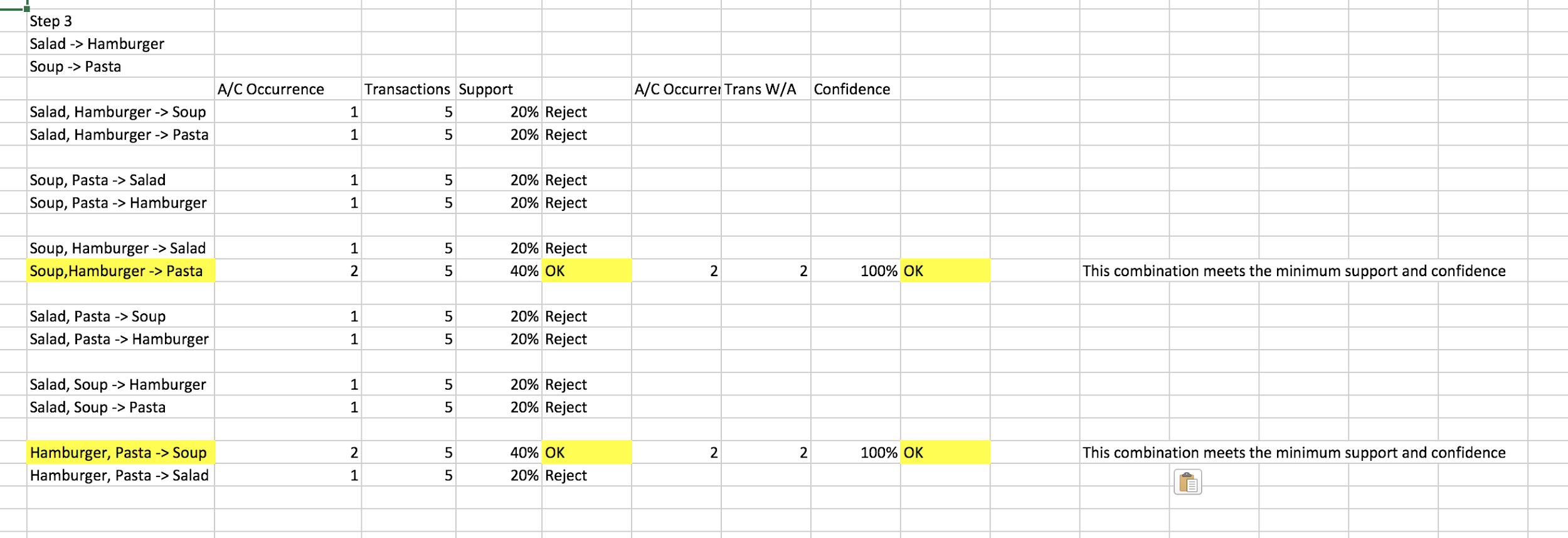
**Step 1**: Calculating Support of the Individual Item in all transactions.



**Step 2:** Calculating support and confidence for multiple rules



**Step 3:**



**Rule 1:** **Salad ⇒ Hamburger**

**Support** = No. of transactions containing all items in antecedent and consequent/No. of transactions in database

So, Support of Rule 1 = 2/5 = 0.4 = 40%

**Confidence** = No. of transactions containing all items in antecedent and consequent/No. of transactions containing items in antecedent

So, Confidence of Rule 1 = 2/2 = 1 = 100%

**Lift** = Confidence of the rule 1/Support of consequent

So, Lift of Rule 1 = 1/(3/5) = 1.67, which is greater than 1 and it is significant.

**Rule 2: Hamburger ⇒ Salad**

**Support** = No. of transactions containing all items in antecedent and consequent/No. of transactions in database

So, Support of Rule 1 = 2/5 = 0.4 = 40%

**Confidence** = No. of transactions containing all items in antecedent and consequent/No. of transactions containing items in antecedent

So, Confidence of Rule 1 = 2/3 = 0.67 = 67% and it is less that minimum confidence (80%).

**Rule 3: Soup ⇒ Pasta**

**Support** = No. of transactions containing all items in antecedent and consequent/No. of transactions in database

So, Support of Rule 1 = 3/5 = 0.6 = 60%

**Confidence** = No. of transactions containing all items in antecedent and consequent/No. of transactions containing items in antecedent

So, Confidence of Rule 1 = 3/3 = 1 = 100%

**Lift** = Confidence of the rule 3/Support of consequent

So, Lift of Rule 3 = 1/(4/5) = 1.25 which is greater than 1 and it is significant.

**Rule 4: Pasta ⇒ Soup**

**Support** = No. of transactions containing all items in antecedent and consequent/No. of transactions in database

So, Support of Rule 1 = 3/5 = 0.6 = 60%

**Confidence** = No. of transactions containing all items in antecedent and consequent/No. of transactions containing items in antecedent

So, Confidence of Rule 1 = 3/4 = 0.75 = 75% and it is less that minimum confidence (80%).

**Rule 5: Soup, Hamburger ⇒ Pasta**

**Support** = No. of transactions containing all items in antecedent and consequent/No. of transactions in database

So, Support of Rule 5 = 2/5 = 0.4 = 40%

**Confidence** = No. of transactions containing all items in antecedent and consequent/No. of transactions containing items in antecedent

So, Confidence of Rule 5 = 2/2 = 1 = 100%

**Lift** = Confidence of the rule 5/Support of consequent

So, Lift of Rule 5 = 1/(4/5) = 1.25 which is greater than 1 and it is significant.

**Rule 6: Hamburger, Pasta ⇒ Soup**

**Support** = No. of transactions containing all items in antecedent and consequent/No. of transactions in database

So, Support of Rule 6 = 2/5 = 0.4 = 40%

**Confidence** = No. of transactions containing all items in antecedent and consequent/No. of transactions containing items in antecedent

So, Confidence of Rule 6 = 2/2 = 1 = 100%

**Lift** = Confidence of the rule 5/Support of consequent

So, Lift of Rule 6 = 1/(3/5) = 1.67 which is greater than 1 and it is significant.

So the valid association rules that satisfy the minimum support and confidence thresholds are 1,3 and 5.

The final Support, Confidence and Lift values of valid rules, which are calculated above, are mentioned in the following table.

|  |  |  |  |
| --- | --- | --- | --- |
| **Valid Rules** | **Support** | **Confidence** | **Lift** |
| **Rule 1:** Salad ⇒ Hamburger | 40% | 100% | 1.67 |
| **Rule 3:** Soup ⇒ Pasta | 60% | 100% | 1.25 |
| **Rule 5:** Soup, Hamburger ⇒ Pasta | 40% | 100% | 1.25 |
| **Rule 6:** Hamburger, Pasta ⇒ Soup | 40% | 100% | 1.67 |

**Solution 2:**

**(a)** For the given table of data and the association rule “hot dogs ⇒ hamburgers”, calculation of support, confidence and lift will prove if the rule is valid.

**Support** = No. of transactions containing all items in antecedent and consequent/No. of transactions in database

So, Support of Rule = 2000/5000 = 0.4 = 40%

**Confidence** = No. of transactions containing all items in antecedent and consequent/No. of transactions containing items in antecedent

So, Confidence of Rule = 2000/3000 = 0.67 = 67%

**Lift** = Confidence of the rule /Support of consequent

So, Lift of Rule = (2000/3000)/(2500/5000) = 4/3 = 1.33

The support and confidence satisfy the minimum thresholds (25% and 50% respectively). The lift is greater than 1, so the relationship between hotdogs and hamburgers is significant and the rule is valid.

**(b)** To find whether the purchase of hot dogs is independent of the purchase of hamburgers, we should see how many times hamburgers is bought out of all the transactions i.e calculate the support of hamburgers.

**Support** = transactions containing hamburgers/total transactions

= 2500/5000 = 0.5 = 50%.

So, 50% of the transactions contain hamburgers with or without hot dogs.

Now, for **hot dogs -> Hamburgers**,

**Support** = 2000/5000 =40 % transactions contain Hamburgers and hot dogs

**Confidence** =2000/3000= 67%

which means it is more likely that 67% of transactions that contain hotdogs will also contain hamburgers.

**Lift**=(2000/3000)/(2500/5000) =1.33

Now, for **No hot dogs ->Hamburgers**,

**Support** = 500/5000 = 0.1 = 10% transactions contain only Hamburgers(without hot dogs).

**Confidence** =500/2000 =0.25 =25% which means that it is more likely that 25% of transactions that contain hamburgers will not have hot dogs.

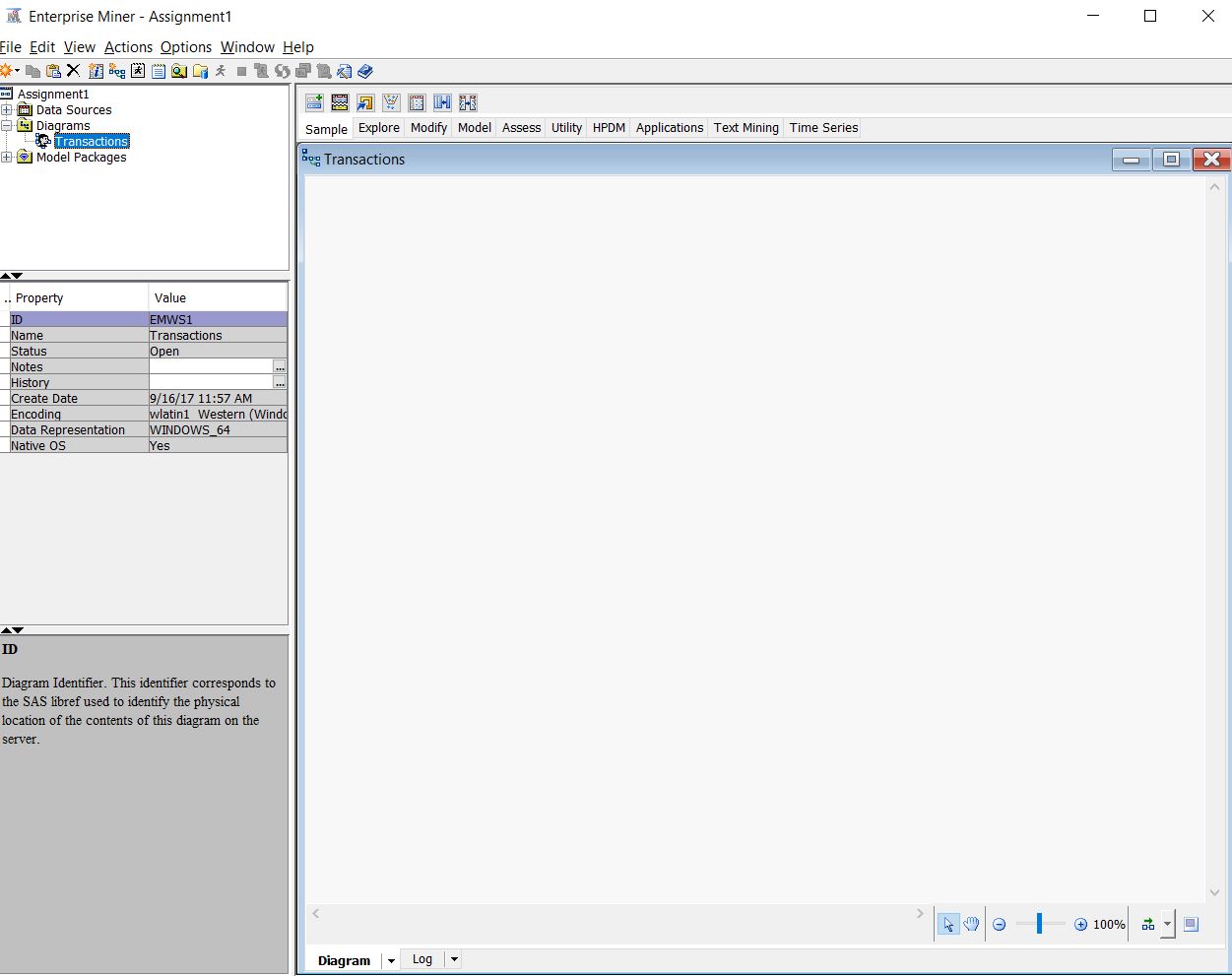
**Lift** = (500/2000) / (2500/5000) = 0.5

Thus, we see that purchase of hot dog increase the chances of purchasing hamburger by 1.33 times. Also, the support for Hamburger is reduced from 50% to 10 % if there is no purchase of hot dogs.

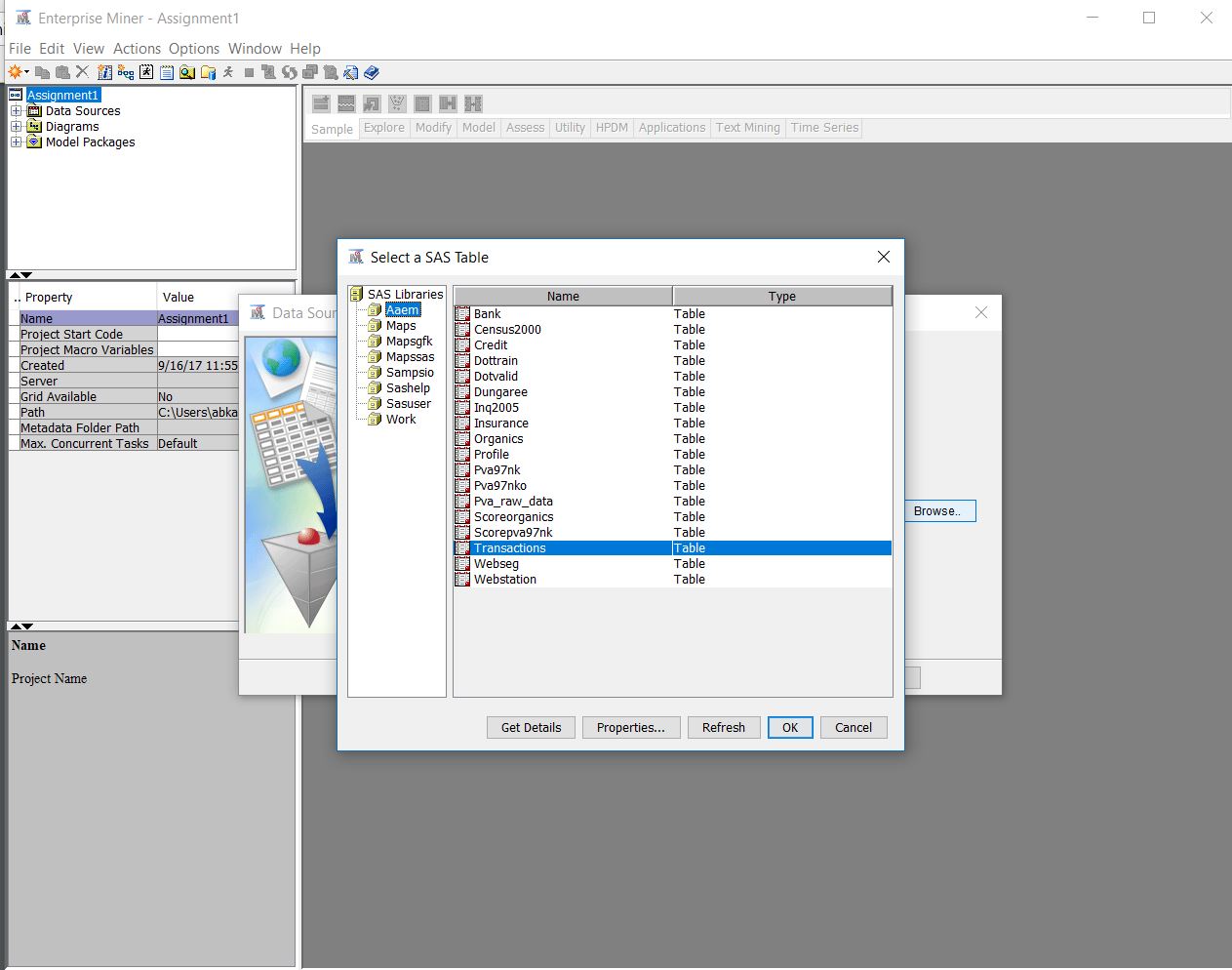
**Solution 3:**

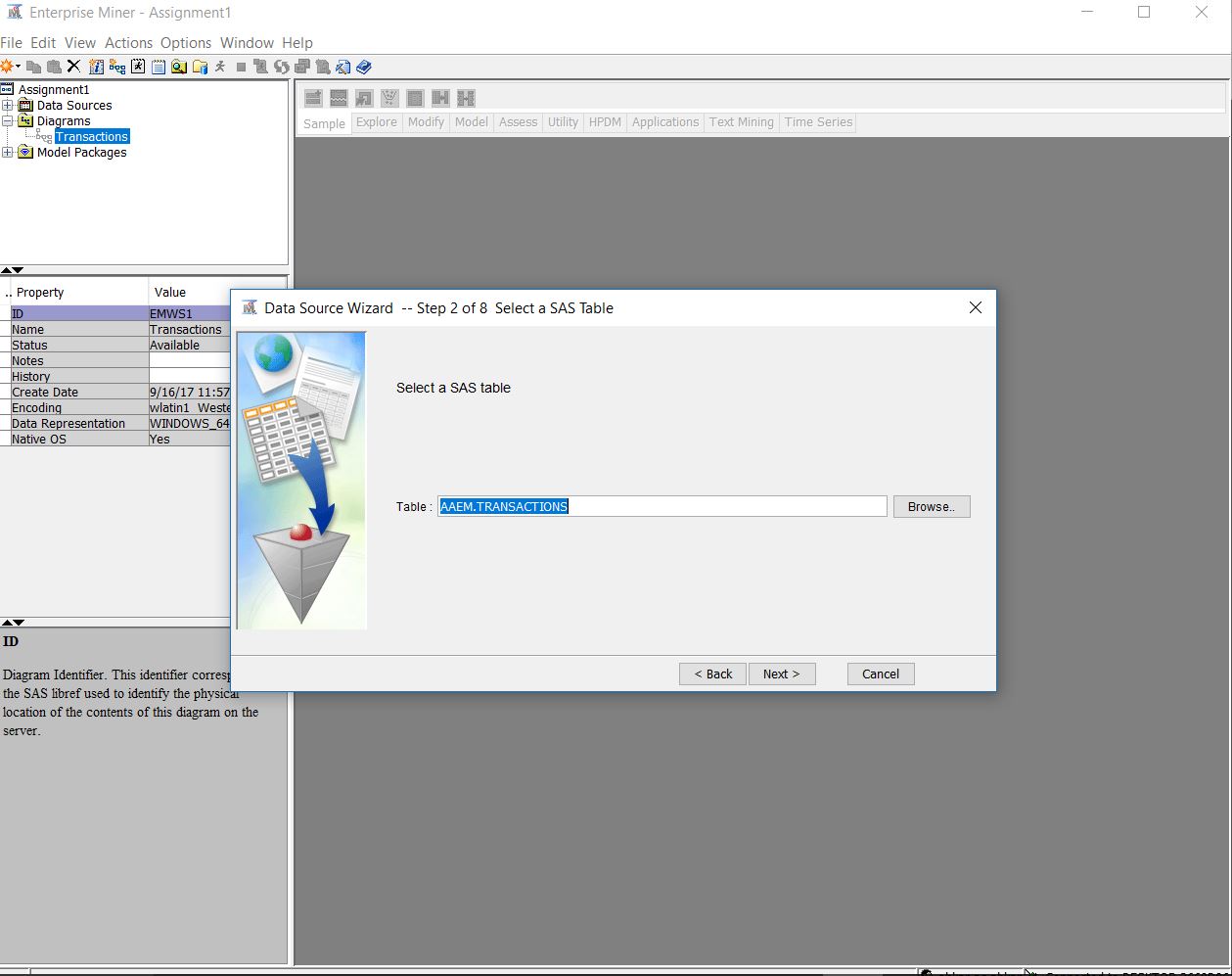
The association analysis using the SAS Enterprise Miner is depicted below by appropriate screenshots.

1. A new diagram, Transactions is created in our project titled Assignment 1.

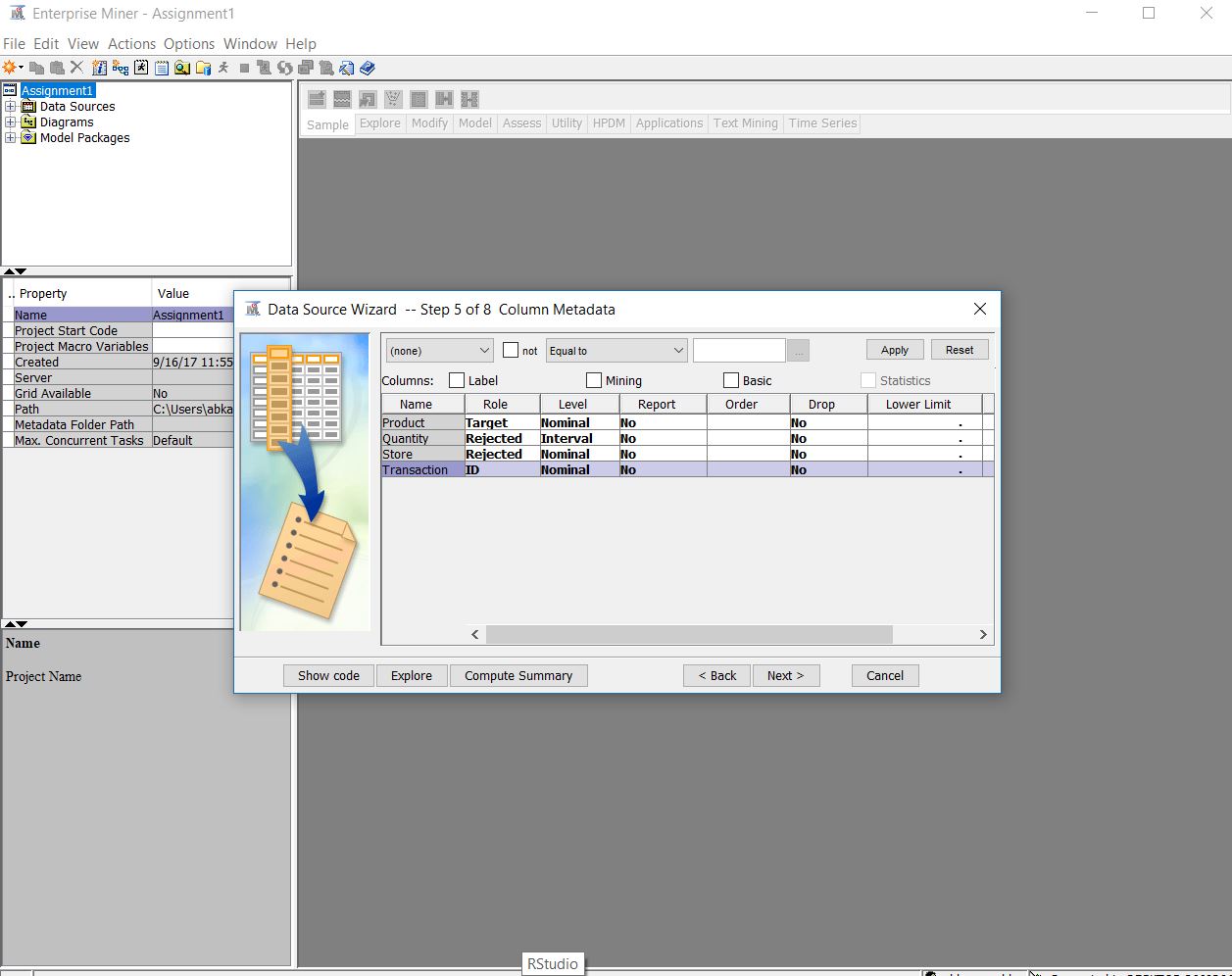


**(b)** A new library is setup and directed to AAEM folder. A new datasource is created using the data set AAEM.TRANSACTIONS as shown below.

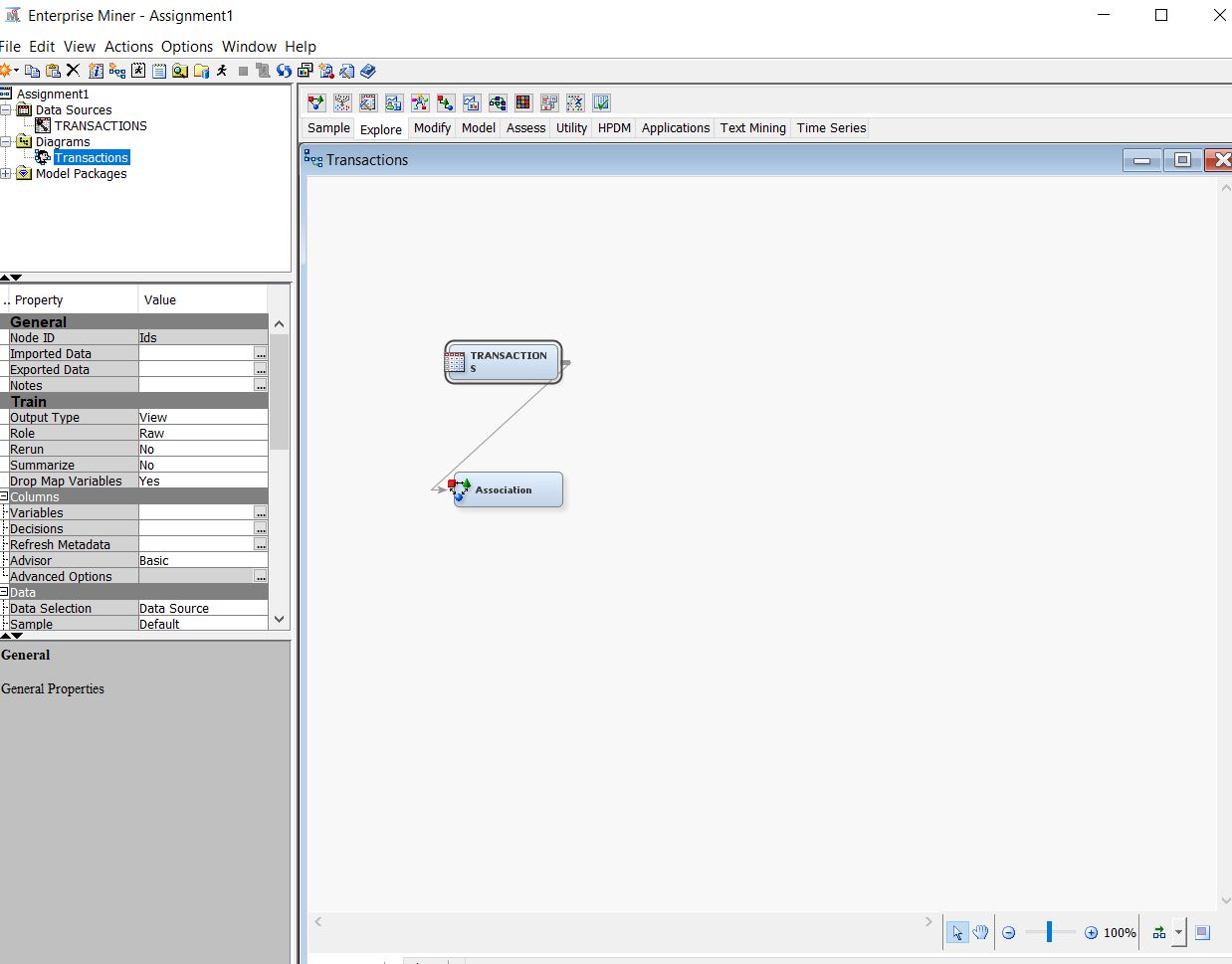




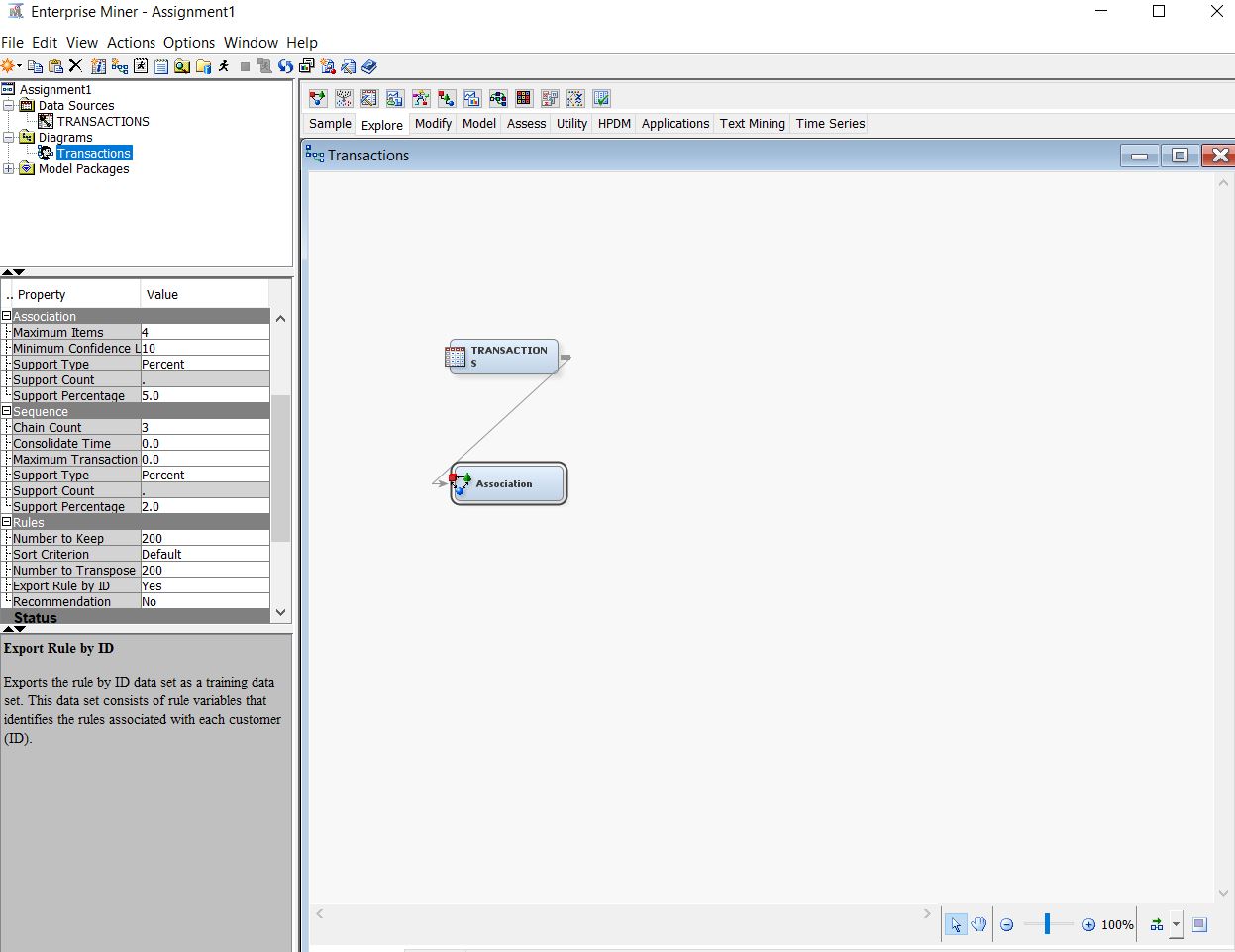
**(c)** For the variables STORE and QUANTITY, the model role is assigned as Rejected. The model role ID is assigned to the variable TRANSACTIONS. For the variable PRODUCT, the model rule target is assigned.



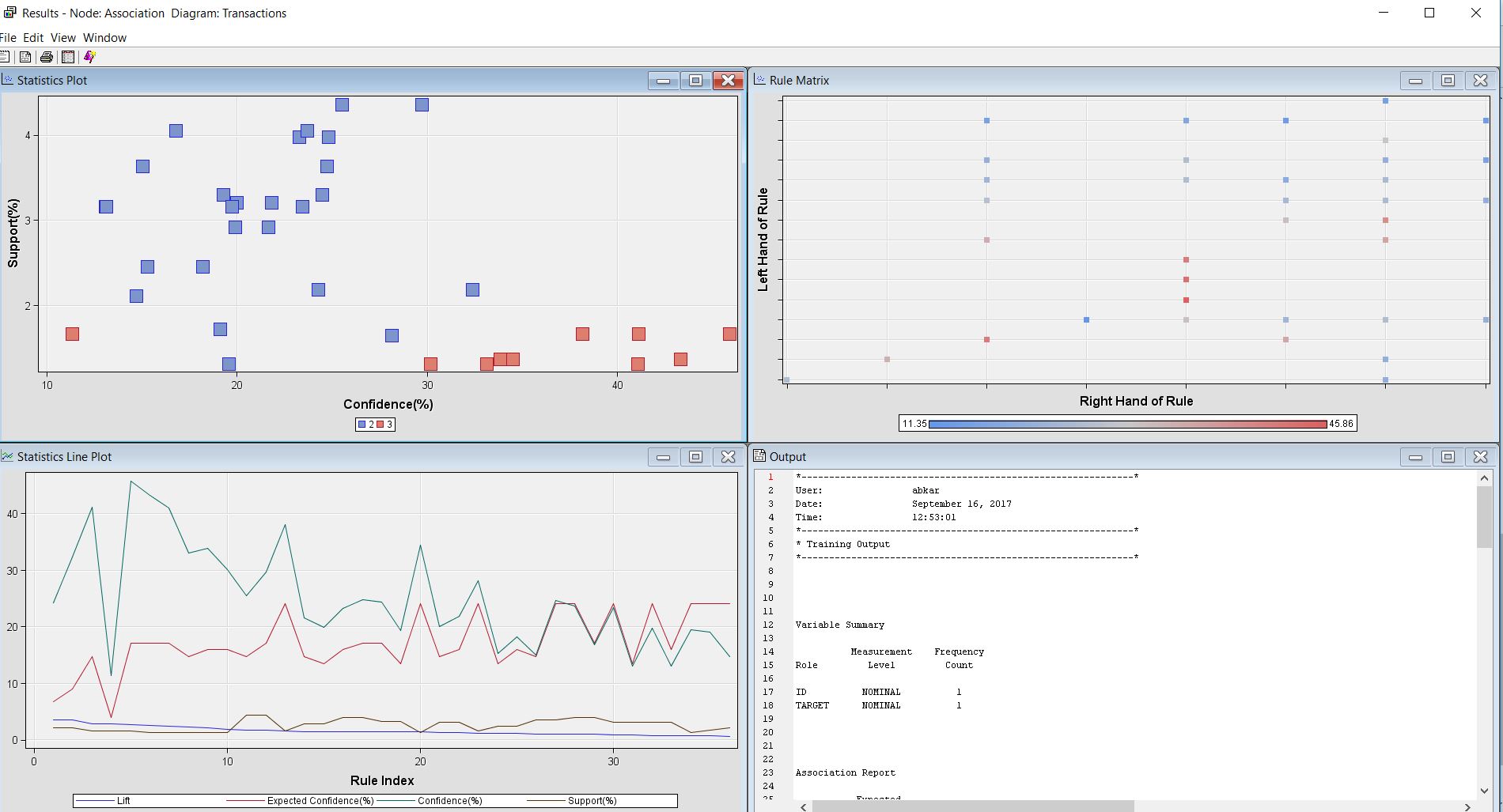
**(d)** The node for TRANSACTIONS data set and the association node to the diagram are added and shown below.



**(e)** The export rule by ID is set to YES as below.



**(f)** The default settings for association node are left alone and analysis is performed.



**(g)**

**i)** The highest lift value is 3.60 for the following two rules namely,

* Perfume ⇒ Toothbrush
* Toothbrush ⇒ Perfume

**ii)** The Lift values for Perfume ⇒ Toothbrush and Toothbrush ⇒ Perfume is 3.60 and it is calculated using the following formulae.

**Support** = No. of transactions containing all items in antecedent and consequent/No. of transactions in database

As per the results, Support of Rule = 2.18

**Confidence** = No. of transactions containing all items in antecedent and consequent/No. of transactions containing items in antecedent

As per the results, Confidence of Rule = 24.26

**Lift** = Confidence of the rule/Support of consequent

**iii)**

**Rule 1:**  Perfume ⇒ Toothbrush

The Lift of this rule suggests that, whenever a perfume is purchased, it increases the chances of toothbrush being bought by 3.6 times.

**Rule 2:** Toothbrush ⇒ Perfume

The Lift of this rule suggests that, whenever a toothbrush is purchased, it increases the chances of perfume being bought by 3.6 times.

**Rule 3:** Magazine & Candy bar ⇒ Greeting cards

The Lift of this rule suggests that, whenever Magazine & Cardy bar are purchased together, it increases the chances of Greeting cards being bought by 2.8 times.

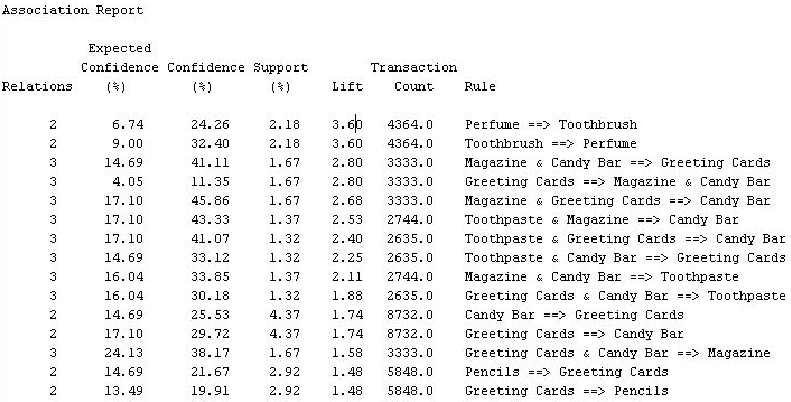
**Rule 4:** Greeting Card ⇒ Magazine & Candy bars

The Lift of this rule suggests that, whenever a Greeting Card is purchased, it increases the chances of Magazine and candy bars being bought by 2.8 times.

**Rule 5:** Magazine & Greeting cards ⇒ Candy bar

The Lift of this rule suggests that, whenever Magazine & Greeting card are purchased, it increases the chances of Candy bar being bought by 2.68 times.

**iv)** The first few rules are captured below as shown



**Rules 1 & 2:** The support of the first two rules involving Perfume and Toothbrush is 2.18 with a lift of 3.60 which infers that, the number of times perfume or toothbrush purchased in the given transactions is same and thus they both are purchased together. So, over here the first rule Perfume⇒ Toothbrush is redundant and can be eliminated as it has lower confidence level.

**Rule 3, 4, 5 & 13:** For these four rules, we see that the support is same as 1.67 for Greeting Cards, Magazine and Candy Bar. By comparing these rules it can be inferred that the lift for the rules is almost same which is around 2.80 except for one rule Greeting Cards & Candy Bar ⇒ Magazine whose lift is 1.58 which implies that at least 1.58 times these 3 products are purchased together, thereby making these rules redundant. So by keeping all the values into consideration we can proceed with Rule 5: Magazine & Greeting card ⇒ Candy Bar as it has the highest confidence Level value of 45.86.

**Rule 6 &9:** The support of the two rules involving Toothpaste, Magazine and Candybar are 1.37 with a lift of 2.53 and 2.11 which infers that at least 2.11 times these 3 products are purchased together, thereby making rule 6 & 9 redundant. So, we can take rule Toothpaste,Magazine ⇒ Candybar into consideration as it has higher confidence level over the rule Magazine, CandyBar⇒ Toothpaste.

**Rules 11 & 12:** The support of the these two rules involving Greeting card and Candy Bar is 4.37 with a lift of 1.74 which infers that, the number of times perfume or toothbrush purchased in the given transactions are equal and thus they both are almost always purchased together, which makes the two stated rules 11 and 12 as redundant. Rule 12 which is Candy Bar ⇒ Greeting Card can be eliminated as it has lower confidence level of 25.53 as compared to Rule 11 Greeting Card ⇒ Candy Bar confidence level of 29.72

**Rule 14 & 15:** The support of the two rules involving Pencil and Greeting cards are 2.92 with a lift of 1.48 which infers that, the number of times pencil or greeting cards purchased in the given transactions is same and thus they both are purchased together. So, over here the rule Greeting card ⇒ Pencils redundant and can be eliminated as it has lower confidence level.