Wisconsin Illinois SAS Users Conference

Visual Market Basket Analysis with SAS and MS Excel

John Xu
1ST Consulting LLC

November 12, 2012
IN THE PRESENTATION

- What is Market Basket Analysis – Theory and Terminology.
- Implementation Difficulty.
- Matt Redlon’s SUGI Paper
- John’s Macro.
- Excel Visual Analysis
WHAT IS MARKET BASKET ANALYSIS?

Walmart Story:

Diaper and Beer purchased together.

Young dads were sent to store to buy diaper.

What to do?
WHAT IS MARKET BASKET ANALYSIS?

What to do:

Put Premium Beer and Diaper together.

For those already purchased them: Up Sale.
For those new purchasers: Cross Sale.
WHAT IS MARKET BASKET ANALYSIS?

Real World Application:
Market Basket Analysis is used to determine the likelihood that a set of items will be bought together (in the same basket).
WHAT IS MARKET BASKET ANALYSIS?

% “Learning” Buyer among “Little” Customers % “Learning” buyer among ALL customers

<table>
<thead>
<tr>
<th></th>
<th>ANALYSIS UNIT</th>
<th>ANALYSIS UNIT FREQ</th>
<th>ASSOC_ANALYSIS_UNIT</th>
<th>ASSOC_ANALYSIS_UNIT_FREQ</th>
<th>FREQ_CO_OCCUR</th>
<th>TOT_BASKET_DIMENSIONS</th>
<th>SUPPORT</th>
<th>CONFIDENCE</th>
<th>EXPECTED_CONFIDENCE</th>
<th>LIFT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>21,407</td>
<td>4,975</td>
<td>3,578</td>
<td>1,333</td>
<td>241,720</td>
<td>0.55%</td>
<td>6.23%</td>
<td>1.48%</td>
<td>4.21</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>21,407</td>
<td>6,214</td>
<td>15,932</td>
<td>4,191</td>
<td>241,720</td>
<td>1.70%</td>
<td>19.24%</td>
<td>6.25%</td>
<td>3.08</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>21,407</td>
<td>5,073</td>
<td>2,165</td>
<td>568</td>
<td>241,720</td>
<td>0.24%</td>
<td>2.65%</td>
<td>0.90%</td>
<td>2.96</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>21,407</td>
<td>8,985</td>
<td>5,497</td>
<td>1,231</td>
<td>241,720</td>
<td>0.51%</td>
<td>5.75%</td>
<td>2.27%</td>
<td>2.53</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>21,407</td>
<td>5,600</td>
<td>4,797</td>
<td>1,004</td>
<td>241,720</td>
<td>0.42%</td>
<td>4.69%</td>
<td>1.98%</td>
<td>2.37</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>21,407</td>
<td>4,323</td>
<td>16,641</td>
<td>3,414</td>
<td>241,720</td>
<td>1.41%</td>
<td>15.95%</td>
<td>6.88%</td>
<td>2.32</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>21,407</td>
<td>6,306</td>
<td>10,871</td>
<td>2,213</td>
<td>241,720</td>
<td>0.92%</td>
<td>10.34%</td>
<td>4.50%</td>
<td>2.30</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>21,407</td>
<td>6,763</td>
<td>6,513</td>
<td>1,307</td>
<td>241,720</td>
<td>0.54%</td>
<td>6.11%</td>
<td>2.69%</td>
<td>2.27</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>21,407</td>
<td>5,305</td>
<td>2,954</td>
<td>579</td>
<td>241,720</td>
<td>0.24%</td>
<td>2.71%</td>
<td>1.22%</td>
<td>2.21</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>21,407</td>
<td>6,533</td>
<td>24,089</td>
<td>4,365</td>
<td>241,720</td>
<td>1.89%</td>
<td>21.33%</td>
<td>9.97%</td>
<td>2.14</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>21,407</td>
<td>6,678</td>
<td>14,644</td>
<td>2,759</td>
<td>241,720</td>
<td>1.14%</td>
<td>12.89%</td>
<td>6.06%</td>
<td>2.13</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>21,407</td>
<td>4,725</td>
<td>3,289</td>
<td>619</td>
<td>241,720</td>
<td>0.26%</td>
<td>2.89%</td>
<td>1.36%</td>
<td>2.13</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>21,407</td>
<td>104,891</td>
<td>3,136</td>
<td>576</td>
<td>241,720</td>
<td>0.24%</td>
<td>2.69%</td>
<td>1.30%</td>
<td>2.07</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>21,407</td>
<td>4,468</td>
<td>17,951</td>
<td>3,292</td>
<td>241,720</td>
<td>1.35%</td>
<td>15.38%</td>
<td>7.43%</td>
<td>2.07</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>21,407</td>
<td>6,345</td>
<td>22,466</td>
<td>4,057</td>
<td>241,720</td>
<td>1.66%</td>
<td>18.95%</td>
<td>9.28%</td>
<td>2.04</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>21,407</td>
<td>6,278</td>
<td>5,976</td>
<td>1,036</td>
<td>241,720</td>
<td>0.43%</td>
<td>4.84%</td>
<td>2.47%</td>
<td>1.96</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>21,407</td>
<td>5,085</td>
<td>4,957</td>
<td>845</td>
<td>241,720</td>
<td>0.35%</td>
<td>3.95%</td>
<td>2.05%</td>
<td>1.92</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>21,407</td>
<td>6,316</td>
<td>2,347</td>
<td>397</td>
<td>241,720</td>
<td>0.16%</td>
<td>1.86%</td>
<td>0.97%</td>
<td>1.91</td>
</tr>
</tbody>
</table>

1=“The Little SAS Book”, 21,407=# customers bought “Little” 4,975=“Learning SAS by Example” 3,578=# customers bought “Learning” 1,333=# customers bought both 241,720=Total # of Customers
WHAT IS MARKET BASKET ANALYSIS?

A) Analysis Unit = The Little SAS Book
B) Analysis Unit Frequency = The number of “Little” sold
C) Associated Analysis Unit = Learning SAS by Example
D) Associated Analysis Unit Frequency = The number of “Learning” sold
E) Co-Occurrence = # of customer bought both books
F) Total Basket Dimension = Total # of Customers
WHAT IS MARKET BASKET ANALYSIS?

G) SUPPORT = CO OCCURRENCE # / TOTAL BASKET DIMENSIONS

H) CONFIDENCE = CO OCCURRENCE # / ANALYSIS UNIT #

I) EXPECTED_CONFIDENCE = ASSOC ANALYSIS UNIT # / TOTAL BASKET DIMENSIONS

J) LIFT = CONFIDENCE / EXPECTED_CONFIDENCE
WHAT WE LEARNED FROM THE EXAMPLE?

Customers who bought “The little SAS Book” are over 4 times more likely buy “Learning SAS by Example”.

* We are using fake data. The actual data will likely show the lift is much higher than 4.
Some association is obvious, such as shampoo and Conditioner.

Others can be hidden, such as Beer and Diaper in Walmart example.
NOTE 2

The Analysis Unit can be created per analysis need.

For example, all SAS books can be a single unit. Or even all books can be a single unit.
Similarly, the association (Basket) can be different. You can look at:

- One Transaction
- All the transaction in a day/week/month
- You can also look into the association between one transaction and next transaction etc.
NOTE 4

We can do MBA at multiple levels.

Level 2: Product A is associated with Product B1, B2, B3, ranked by Lift.

Level 3: Product A and B1 are associated with Product C1, C2, C3, ranked by Lift.

Level 4: ....
Level 5: ....

More levels will uncover some deep, hard to find association.
IMPLEMENTATION OF 2-LEVEL MBA

Matthew Redlon’s SUGI 28 paper:

A SAS® Market Basket Analysis Macro: The “Poor Man’s Recommendation Engine”
IMPLEMENTATION OF MULTI-LEVEL MBA

Too many combinations
If there are only 100 product:
Level 2: $A \times B = 100 \times 99 = 9,900$ (10K)
Level 3: $A \times B \times C = 100 \times 99 \times 98 = 979,200$ (1 million)
Level 4: $A \times B \times C \times D = 100m$
Level 5: $A \times B \times C \times D \times E$ (10 Billion)
COMPARE TO OLAP

Market Analysis look at the same dimension.

OLAP drill down look at different dimension. Product -> Demographic -> ...
IMPLEMENTATION DIFFICULTY.

How to go down to the higher level?

If a customer purchased A, B, C, D, E, ..., what he most likely to purchase together?

Hint: Think about Support.
NEW IMPLEMENTATION

USE SQL to calculate Co Occurrence at level 2
Calculate other variables for level 2, drop the combination when support under threshold.

Use the remaining 2 product combination as Analysis Unit for Level 3, repeat the above process for Level 3.

Continue to level 4, 5, ... until desired level or all combinations are under the threshold.
SAS CODE REVIEW

See SAS windows.

Sample usage:
```
%mba(indset=test.sample,
     outlib=test,
     analysis_unit=product,
     basket_dimension=customerkey,
     level=6);
```

Note the SQL MxM join in the code.
Demo in Excel

Use Filter on Result

Dynamic Look up.
QUESTIONS?
Thanks!