

Wisconsin Illinois SAS Users Conference

Visual Market Basket Analysis with SAS and MS Excel

John Xu
1ST Consuting LLC

November 12, 2012

IN THE PRESENTATION

- What is Market Basket Analysis ? – Story and Real World Applications.
- 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:

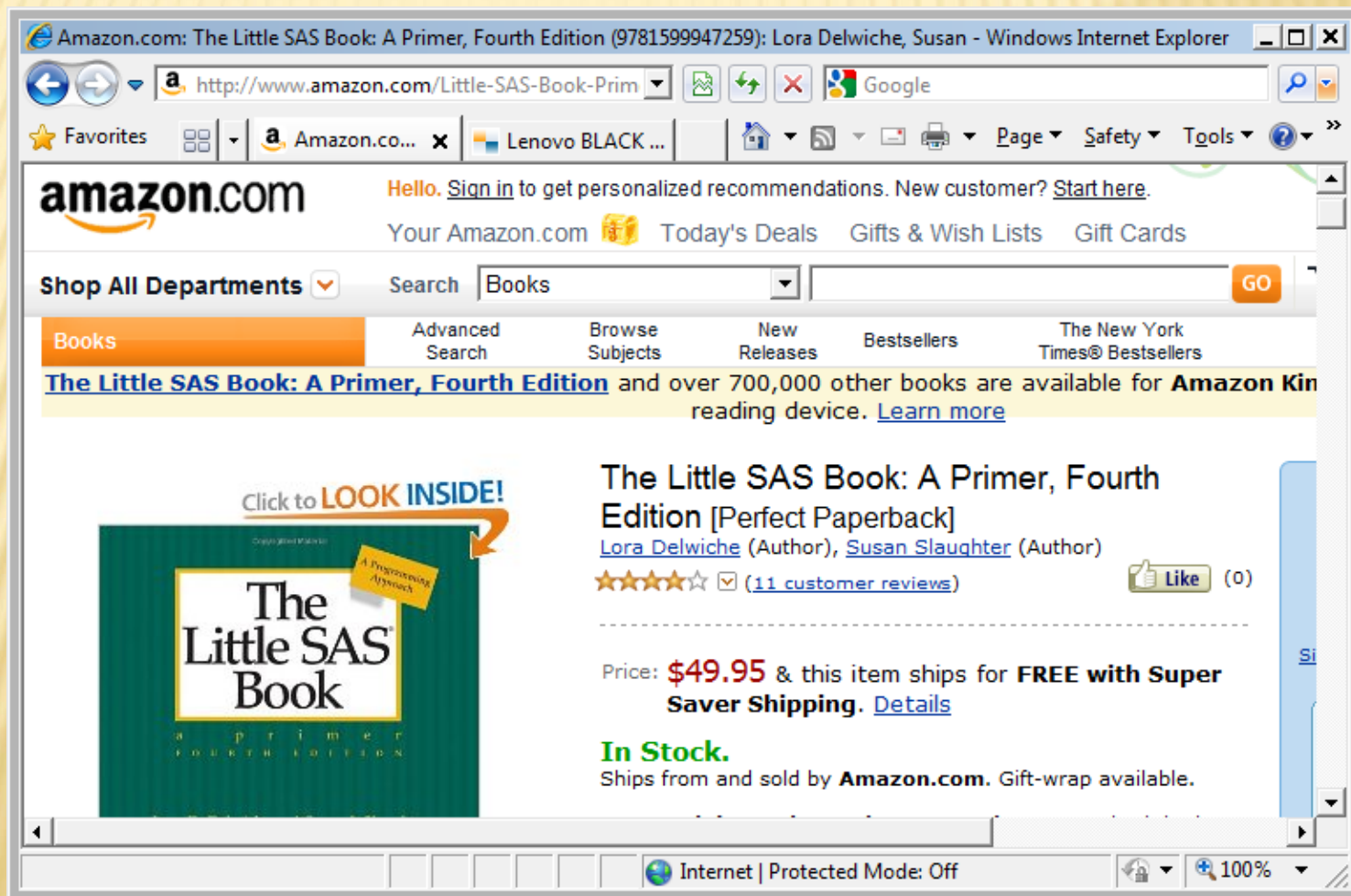
Put Premium Beer and Diaper together.

For those already purchased them: Up Sale.

For those new purchasers: Cross Sale.

WHAT IS MARKET BASKET ANALYSIS?

Real World Application:



The screenshot shows a Windows Internet Explorer browser window displaying the Amazon.com product page for "The Little SAS Book: A Primer, Fourth Edition" by Lora Delwiche and Susan Slaughter. The browser's address bar shows the URL: <http://www.amazon.com/Little-SAS-Book-Prim>. The Amazon.com logo is visible at the top left, along with navigation links for "Hello. Sign in to get personalized recommendations. New customer? Start here.", "Your Amazon.com", "Today's Deals", "Gifts & Wish Lists", and "Gift Cards". The search bar contains the word "Books". Below the search bar, there are navigation tabs for "Books", "Advanced Search", "Browse Subjects", "New Releases", "Bestsellers", and "The New York Times® Bestsellers". The product title is "The Little SAS Book: A Primer, Fourth Edition" and it is noted that over 700,000 other books are available for Amazon Kindle reading device. A "Click to LOOK INSIDE!" button is positioned above the book cover image. The book cover is green with a white box containing the title "The Little SAS Book" and "a primer FOURTH EDITION". To the right of the cover, the product title is repeated: "The Little SAS Book: A Primer, Fourth Edition [Perfect Paperback]" by Lora Delwiche (Author) and Susan Slaughter (Author). The product has a star rating of 4.5 out of 5 stars based on 11 customer reviews. The price is listed as \$49.95, and it is noted that the item ships for FREE with Super Saver Shipping. The product is currently "In Stock" and is shipped from and sold by Amazon.com. The browser's status bar at the bottom shows "Internet | Protected Mode: Off" and a zoom level of 100%.


WHAT IS MARKET BASKET ANALYSIS?

Amazon.com: The Little SAS Book: A Primer, Fourth Edition (9781599947259): Lora Delwiche, Susan - Windows Internet Explorer

http://www.amazon.com/Little-SAS-Book-Primer-Fourth/dp/1599947250/ref=sr_1_2?i

Amazon.com: The Little SAS Book: A Primer, Fou...

Frequently Bought Together



Price For All Three: \$186.01

[Add all three to Cart](#) [Add all three to Wish List](#)


[Show availability and shipping details](#)

- This item:** The Little SAS Book: A Primer, Fourth Edition by Lora Delwiche Perfect Paperback **\$49.95**
- Learning SAS by Example: A Programmer's Guide by Ron Cody Paperback **\$69.25**
- Applied Statistics and the SAS Programming Language (5th Edition) by Ron P. Cody Paperback **\$66.81**

Customers Who Bought This Item Also Bought

Page 1 of 25



 <p>Learning SAS by Example: A Programmer's Guide by Ron Cody ★★★★☆ (15) \$69.25</p>	 <p>Applied Statistics and the SAS Programming Language by Ron P. Cody ★★★★☆ (39) \$66.81</p>	 <p>SAS For Dummies by Stephen McDaniel ★★★★☆ (11) \$19.79</p>	 <p>Epidemiology: Beyond the Basics by Moyses Szklo ★★★★☆ (16) \$76.50</p>
---	--	---	---

Internet | Protected Mode: Off 100%

WHAT IS MARKET BASKET ANALYSIS?

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

Obs	ANALYSIS_UNIT	ANALYSIS_UNIT_FREQ	ASSOC_ANALYSIS_UNIT	ASSOC_ANALYSIS_UNIT_FREQ	FREQ_CO_OCCUR	TOT_BASKET_DIMENSIONS	SUPPORT	CONFIDENCE	EXPECTED_CONFIDENCE	LIFT
1	1	21,407	4,975	3,578	1,333	241,720	0.55%	6.23%	1.48%	4.21
2	1	21,407	6,224	15,003	4,111	241,720	1.70%	19.24%	6.25%	3.08
3	1	21,407	5,072	2,165	568	241,720	0.24%	2.65%	0.90%	2.96
4	1	21,407	8,985	5,497	1,231	241,720	0.51%	5.75%	2.27%	2.53
5	1	21,407	5,600	4,787	1,004	241,720	0.42%	4.69%	1.98%	2.37
6	1	21,407	4,323	16,641	3,414	241,720	1.41%	15.95%	6.88%	2.32
7	1	21,407	6,306	10,879	2,213	241,720	0.92%	10.34%	4.50%	2.30
8	1	21,407	6,763	6,513	1,307	241,720	0.54%	6.11%	2.69%	2.27
9	1	21,407	5,305	2,954	579	241,720	0.24%	2.71%	1.22%	2.21
10	1	21,407	6,533	24,089	4,565	241,720	1.89%	21.33%	9.97%	2.14
11	1	21,407	6,678	14,644	2,759	241,720	1.14%	12.89%	6.06%	2.13
12	1	21,407	4,725	3,289	619	241,720	0.26%	2.89%	1.36%	2.13
13	1	21,407	104,891	3,136	576	241,720	0.24%	2.69%	1.30%	2.07
14	1	21,407	4,468	17,951	3,292	241,720	1.36%	15.38%	7.43%	2.07
15	1	21,407	6,345	22,436	4,057	241,720	1.68%	18.95%	9.28%	2.04
16	1	21,407	6,278	5,975	1,036	241,720	0.43%	4.84%	2.47%	1.96
17	1	21,407	5,085	4,957	845	241,720	0.35%	3.95%	2.05%	1.92
18	1	21,407	6,316	2,347	397	241,720	0.16%	1.86%	0.97%	1.91

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) $\text{SUPPORT} = \text{CO OCCURRENCE \#} / \text{TOTAL BASKET DIMENSIONS}$

H) $\text{CONFIDENCE} = \text{CO OCCURENCE \#} / \text{ANALYSIS UNIT \#}$

I) $\text{EXPECTED_CONFIDENCE} = \text{ASSOC ANALYSIS UNIT \#} / \text{TOTAL BASKET DIMENSIONS}$

J) $\text{LIFT} = \text{CONFIDENCE} / \text{EXPECTED_CONFIDENCE}$

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.

NOTE 1

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.

NOTE 3

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: $A \times B = 100 * 99 = 9,900$ (10K)

Level 3: $A \times B \times C = 100 * 99 * 98 = 979,200$ (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.

EXCEL FOR VISUAL ANALYSIS

Demo in Excel

Use Filter on Result

Dynamic Look up.

QUESTIONS?

Thanks!