Contextualized Market Basket Analysis

How to learn more from your Point of Sale Data

Presentation to: Hanesbrands

Andrew Kramer, Louisiana State University

Agenda

1. Project Overview
2. History of Market Basket Analysis
3. Downsides for retailers
4. Contextualized Market Baskets
5. SAS Demo & Code Analysis
6. Conclusion and Q&A

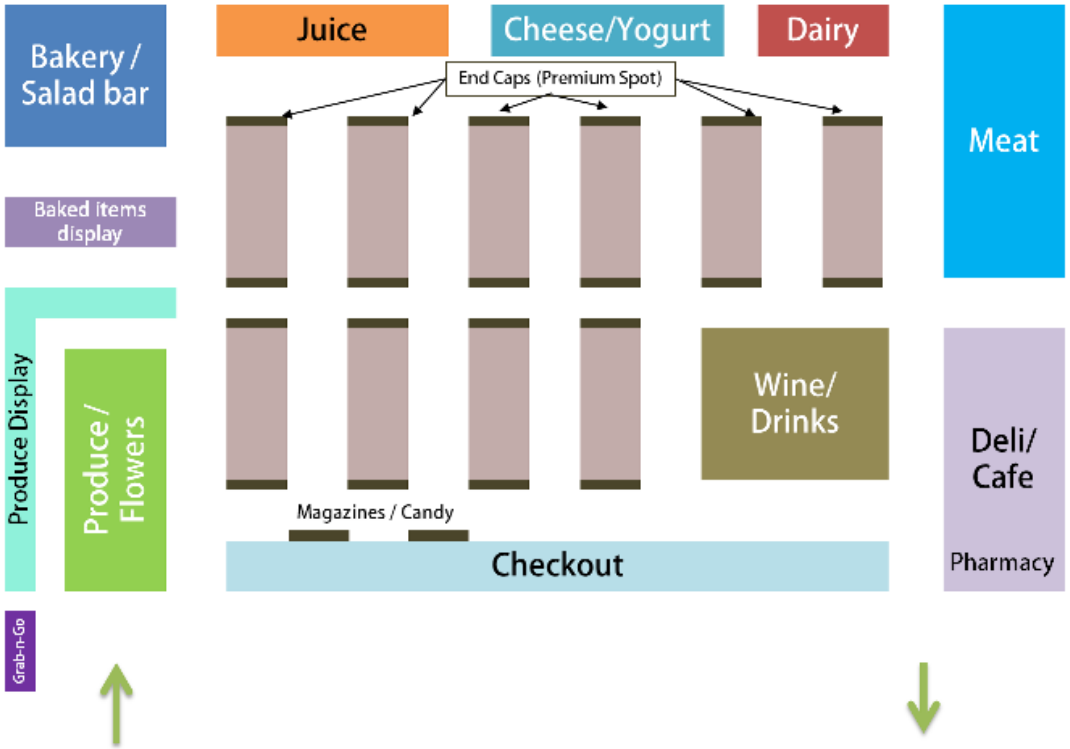
Project Overview

- Experience in retail: desire to learn more about Market Basket Analysis
- Saw an analytical need not covered in current SAS products
  - What are the profit implications of an association?
- Created SAS Enterprise Miner Macro to add onto the results of a Market Basket Analysis:
  - Proc SQL
  - SAS Macros
  - Database Steps
- Presented in Student Symposium at 2015 SAS South Central User Group Conference

Market Basket Analysis

- "Find joint values of the variables that appear frequently in the database"
  -*Elements of Statistical Learning*
- Which products are commonly purchased together?
- Apriori Algorithm: Only Consider relationships between commonly occurring items in dataset
- Development of sophisticated predictive models such as recommender systems

Downsides for Retailers



- Fundamental Question: Do the associations create profitable, long-term relationships with customers?
- Traditional Algorithms lack key data points such as Customer Id, Date, Quantity, Purchase Price, Store Location

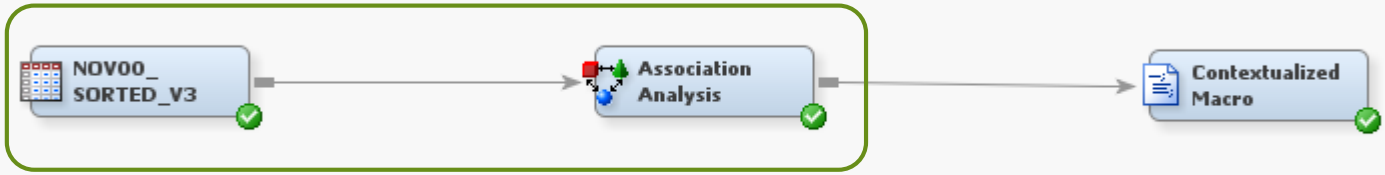
Sample Dataset

- Sample Taiwanese data set from anonymous retailer
- Data from November 2000 – December 2000
- Transaction #: ID Variable in Enterprise Miner
- Prodid: Target in Enterprise Miner

Sample Taiwanese Data Set

Obs	Transaction #	Date	Customer_ID	Age	Area	subclass	ProdId	AMT	Cost	SalePrice
1	1	01NOV2000	00038317	J	E	130315	4714981010038	2	56	48
2	1	01NOV2000	00038317	J	E	120105	4719090105002	1	28	28
3	2	01NOV2000	00045902	H	E	100304	4710147100018	1	24	28
4	2	01NOV2000	00045902	H	E	130204	4710088434692	1	114	119
5	2	01NOV2000	00045902	H	E	100511	4710594912028	6	210	313
6	2	01NOV2000	00045902	H	E	100113	4710285000126	1	112	95
7	3	01NOV2000	00045957	G	E	110217	4710265849066	1	180	133
8	4	01NOV2000	00046855	D	E	110411	4710085120468	3	51	57
9	4	01NOV2000	00046855	D	E	110401	4710088410139	1	43	39
10	4	01NOV2000	00046855	D	E	110117	4710063031878	1	77	89

Market Basket Analysis in EM

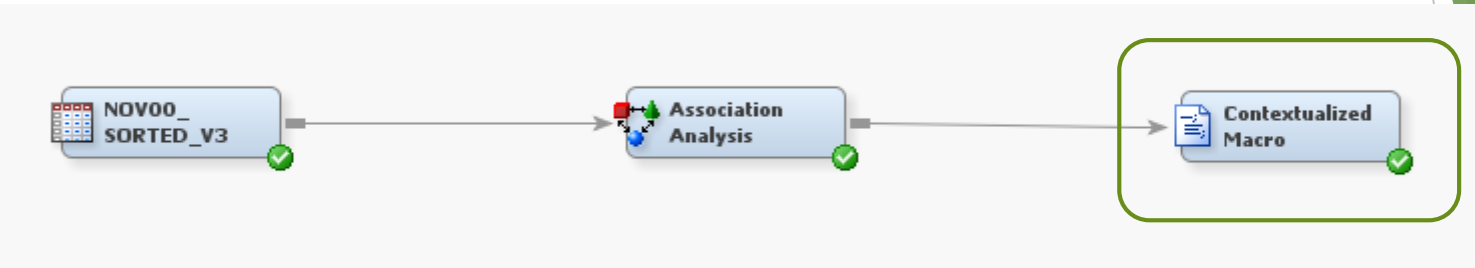


EXP_CONF	CONF	SUPPORT	LIFT	COUNT	RULE	_LHAND	_RHAND
1.35	76.00	0.66	56.18	209.00	4710085120697 ==> 4710085120680	4710085120697	4710085120680
0.86	48.49	0.66	56.18	209.00	4710085120680 ==> 4710085120697	4710085120680	4710085120697
10.76	47.03	0.72	4.37	230.00	4711663700010 ==> 4714981010038	4711663700010	4714981010038
2.82	12.04	0.89	4.26	284.00	4711271000014 ==> 4710421090059	4711271000014	4710421090059
7.40	31.56	0.89	4.26	284.00	4710421090059 ==> 4711271000014	4710421090059	4711271000014

Association Analysis Definitions

- Expected Confidence(%): Percent of transactions that the right-hand side of the rule appears in
- Confidence(%): Given that a transaction has purchased the left-hand side of the rule, what is the probability the transaction will also contain the right-hand side?
- Support: Percent of transactions that contains both items in the rule
- Lift(%): Confidence/Expected Confidence: Lift > 1 evidence of non-independence
- Count: Number of transactions in which both the left-hand and right-hand side of the rule were purchased

Contextualized Market Baskets



- Appends the output from the Association Analysis node based on cost using a Custom Node
- Analyze the market baskets as a whole
  - Item-level & market-basket profitability
  - How size and makeup of market baskets change
- Goal: Profit implications of an association
- Code uses PROC SQL, Database steps, and SAS Macros to crawl the dataset and create output

Macro Output

- Appended Association Analysis Output

RULE	profitable	Not Profitable	Percent Profitable	VarOneProfit	VarTwoProfit	Association Profit	MBProfit
4710085120697 ==> 4710085120680	116	93	0.5550	$3,526	$1,035	$4,561	$66,341
4710085120680 ==> 4710085120697	116	93	0.5550	$1,035	$3,526	$4,561	$66,341
4711663700010 ==> 4714981010038	4	226	0.0174	$134	$-11,510	$-11,376	$12,661
4711271000014 ==> 4710421090059	2	282	0.0070	$-5,084	$-8,213	$-13,297	$30,708
4710421090059 ==> 4711271000014	2	282	0.0070	$-8,213	$-5,084	$-13,297	$30,708

- Contextual output for each Rule (Rule 5 Shown)

Profitable Transaction	N Obs	Variable	N	Mean	Std Dev	Minimum	Maximum
0	123	TotalSales	123	-38.89	21.09	-130.00	-1.00
		VariableInQuestionQty	123	5.60	1.24	1.00	12.00
		VariableInQuestionQty2	123	1.96	0.30	1.00	4.00
		BasketSize	123	10.83	4.17	4.00	28.00
		UniqueItemsPur	123	4.49	2.74	2.00	14.00
		VariableInQuestionProfit	123	-31.73	9.77	-36.00	0.00
		VariableInQuestionProfit2	123	-19.67	6.51	-80.00	-5.00
1	161	TotalSales	161	220.44	300.48	1.00	1946.00
		VariableInQuestionQty	161	5.12	1.79	1.00	12.00
		VariableInQuestionQty2	161	1.78	0.46	1.00	4.00
		BasketSize	161	23.64	11.24	5.00	60.00
		UniqueItemsPur	161	13.91	8.14	2.00	40.00
		VariableInQuestionProfit	161	-26.77	19.31	-36.00	144.00
		VariableInQuestionProfit2	161	-16.55	6.75	-40.00	2.00

ProfitableTransaction	TotalProfit
0	-4783
1	35491

Rule 5 Analysis

- Eggs ==> Butter

Rule	# Times sold at Profit	# Times sold at loss	Association Profit/loss
Eggs ==> Butter	2	282	$13,297

- Total Market Basket Analysis

Rule	# Profitable Transactions	Average Profit	# Not Profitable Transactions	Average Loss	Total Market Basket Profit/Loss
Eggs ==> Butter	161	$220.44	123	$38.89	$30,708

- On Average, Profitable transactions:
  - Buy 13 more items
  - Are $260 more profitable per transaction
  - Purchase fewer items sold at a loss

Code Breakdown

- Utilizes PROC SQL, Data Steps, and SAS Macros
  - Can use strategies to implement in Base SAS

```
PROC SQL;
             SELECT count(_LHAND)
             INTO :rules
             FROM &EM_IMPORT_RULES;
QUIT;
```

```
PROC SQL;
            CREATE Table varlist AS
            SELECT _LHAND, _RHAND
            FROM &EM_IMPORT_RULES;
QUIT;

DATA _NULL_;
            SET VARLIST;
            id=_N_;
            CALL SYMPUT('ids'||left(put(id,20.)), _LHAND);
            CALL SYMPUT('idst'||left(put(id,20.)), _RHAND);
RUN;
```

Macro variables to extract the number of rules and the SKUs that make up each rule

Code Breakdown

```
DATA Nov00_Analyzed;
        SET &EM_IMPORT_TRANSACTION;
        BY ID;
        IF first.ID THEN DO;
                    TotalProfit=0;
                    BasketSize=0;
1                   VariableInQuestionProfit=0;
                    VariableInQuestionProfit2=0;
                    VariableInQuestionQty=0;
                    VariableInQuestionQty2=0;
                    END;

        IF prodid="&&ids&i" THEN DO;
                    VariableInQuestionProfit+(AMT*(SalePrice-Asset));
                    VariableInQuestionQty+1;
2                   END;

        IF prodid = "&&idst&i" THEN DO;
                    VariableInQuestionProfit2+(AMT*(SalePrice-Asset));
                    VariableInQuestion2Qty+1;
                    END;

        TotalProfit+(Amt*(SalePrice-Asset));
                    Basketsize+AMT;

3       IF TotalProfit GT 0 THEN ProfitableTransaction=1;
                    ELSE ProfitableTransaction=0;


4       IF last.ID and VariableInQuestionQty GT 0 AND VariableInQuestionQty2 GT 0;
```

Code Breakdown

- 1. Set counter variables to zero at the start of each new transaction ID
- 2. Evoke the variables “ids” and “idst” for the correct loop count as defined by “i”. If the pointer reads a row showing the purchase of one of the two SKUs evoked by the macro, the code will generate summary statistics for the SKU
- 3. Defines variables representing the profitability and basket size of each transaction
- 4. Extracts the calculated fields into a new data set with one row for each transaction in which both SKUs were listed

Code Breakdown

```
%MACRO AllSubs;
        %DO i=1 %TO &rules;
                %DsCrawl
                %ProfitTrans
                %DatasetUse
                %REPORT
        %END;
%MEND AllSubs;
```

Run the DO loop once for each rule and appends the results to an empty dataset. The Report feature merges this new dataset with the results from the EM Association Analysis

Conclusion

- This macro is designed to show the profit implications of an association
- Currently only works with two SKUs, but could be modified to handle a larger number of SKUs
- Importance of analyst and managerial input
- Any Questions?

Contact Information:

Andrew.kramer526@gmail.com