Paper 10481 - 2016

Product Purchase Sequence Analyses By Using Horizontal Data Sorting Technique

Justin Jia, TransUnion Canada, Burlington, Ontario, Canada
Amanda Lin, CIBC, Toronto, Ontario, Canada

ABSTRACT

Horizontal data sorting is a very useful SAS® technique in advanced data analysis using SAS programming. Three years ago (Paper 376-2013), we presented and illustrated various methods and approaches to perform horizontal data sorting, and demonstrated its valuable application in strategic data reporting. However, this technique can also be employed as a creative analytic method in advanced business analytics. This paper presents and discusses its innovative and insightful applications in product purchase sequence analyses, such as product opening sequence analysis, product affinity analysis, next best offer analysis, time span analysis etc. Compared to other analytic approaches, horizontal data sorting technique has the distinct advantages of being straightforward, simple, convenient to use and producing easy-to-interpret analytic results. Therefore the technique can have a wide variety of applications in customer data analysis and business analytics fields.

INTRODUCTION

Sorting and ordering data are fundamental skills in SAS® data analysis. Data sorting can be vertical sorting, across rows, or horizontal sorting, across columns. Compared to vertical sort, horizontal sort is used less frequently, and it requires the user to employ multiple sophisticated SAS skills such as Transpose, Rotate, Array, Macro, etc. Utilization of this technique can significantly enhance the format and layout of data reporting, and thus provide informative insights into data. For example, in a cross-sell marketing campaign, we have customers' purchase data as shown in Table 1. The arrangement of products and purchase quantities is not in a defined order, which makes it very inconvenient for us to look up or compare.

Table 1. Sample Raw Purchase Data of A Cross-Sell Campaign

Client ID	Product 1	Qty	Product 2	Qty	Product 3	Qty	Product 4	Qty
101	Printer	10	Computer	25	Games	6	Software	8
107	TV	17	Software	32	Fax Machine	11		
216	Clothes	16	Book	53	Phone	9	Glass Ware	24

As illustrated in Table 2, a better way of reporting the purchase data is to sort and present the data in a descending sequence of the purchase quantity. This kind of data sorting is horizontal data sorting because it sorts data horizontally across table columns or data set variables. As can be seen from Table 2, this horizontal sorting and presentation of data will not only make the report easy to read, much more important, it also provides insights into customers' purchase tendency and preferences. Therefore, horizontal data sorting is a very useful SAS technique in insightful and strategic reporting of customer and business data.

Table 2. Horizontally Sorted Purchase Data: Order By Descending Purchase Quantity.

Client ID	Product 1	Qty	Product 2	Qty	Product 3	Qty	Product 4	Qty
101	Computer	25	Printer	10	Software	8	Games	6
107	Software	32	TV	17	Fax Machine	11		
216	Book	53	Glass Ware	24	Clothes	16	Phone	9

In a previous paper[1], we presented and discussed various approaches and methods to achieve horizontal data sorting through SAS programming, including Bubble Sort Method, Rotate and Transpose Method, Call Sort Routine Method etc. We have also shown that horizontal data sorting is very useful in insightful data presentation and reporting, for example, in RFM analyses of retail customer data. In that publication, the paper mainly focused on the technical aspects of horizontal data sorting and its application in data reporting. However, this technique is not only limited to its use in data reporting, it can also be employed as a creative means in data analysis. We can use it to leverage data and perform deep-dive business analytics.

In this paper, we explore this SAS technique further and present its application in business intelligence and business analytics. For example, we can use it in product purchase sequence analysis, product affinity analysis and affinity marketing, next best offer analysis, time span analysis and so on. With the advantages of being straightforward, quick and easy to interpret analysis results, this technique can have a wide variety of uses in business analytics fields. We can apply it to probe and understand customers' business behaviours, develop competitive insights and design effective marketing strategies.

CASE STUDY: FINANCIAL PRODUCT PURCHASE SEQUENCE ANALYSES

In customer marketing, a common practice is to cross sell or up sell products to existing customers in order to deepen customer relationships, improve customer loyalty and increase sales revenue. For these purposes, we need conduct various deep-dive analyses to probe and understand the behavior, preference and tendency of our customers. For example, product purchase sequence analysis can provide insights into customers' purchase patterns in terms of time sequences. Product affinity analysis can give information to identify the products with a high tendency of being purchased together. Through these data analyses, we can then develop keen business insights and design effective marketing strategies. Although there are some statistical methods to achieve these goals such as clustering analysis, predictive modeling etc., however, these statistical methods are usually high-costing, expertise-demanding and time taking.

As an alternative to the statistical methods, we can conduct these analyses through horizontal data sorting technique, which is quick, low-costing and easy to implement. Hereby we use a financial business case as an example to illustrate the analytic technique and approaches. Table 3 demonstrates the attributes and structure of the raw financial customer data for our case study. This raw data set contains 2.56 million of observations randomly selected from a huge financial customer database within a specified time window.

Table 3. Attributes of the raw financial customer data.

Client_ID	Account_ID	Account_Type	Opened_Date
1001	601152	Chequing	2011-05-26
1001	103759	Savings	2011-05-26
1001	726789	TFSA	2011-06-01
1002	632975	Chequing	2012-06-11
1002	363108	Credit Card	2012-10-24
1002	101492	Credit Insurance	2012-10-24
1002	161079	Savings	2012-11-07
1003	987783	Credit Card	2010-11-29
1003	230104	Savings	2012-08-09
1003	599067	Credit Card	2012-09-18
1003	385030	Chequing	2012-06-05
1003	632316	Mortgage	2013-07-11
1003	281144	TFSA	2012-12-22
1003	369267	Savings	2013-02-17
1003	443686	Loan	2013-10-30

Client_ID: numeric, N8, which is unique to identify an individual customer.

Account_ID: character, N8, the banking account number. A financial customer may have multiple banking accounts in different types.

Account_Type: character, $10, the type of a banking account , such as Chequing, Savings, Credit Card, Credit Insurance, Mortgage, Line of Credit(PLC), Tax Free Savings Account(TFSA) etc.

Opened_Date: numeric, in YYMMDD10. format, the opened date of a banking account.

Analysis Requests:

Based on the raw financial customer data demonstrated in Table 3, our data analyses are aimed to probe and understand the banking preferences and business patterns of financial customers, and then to develop competitive strategies for customer marketing. We want to determine what is the client penetration of different banking products, what is the next product to offer our clients and when is the correct time to approach them. We will illustrate how to achieve these goals through horizontal data sorting technique and frequency distribution studies.

PART I. HORIZONTAL SORTING OF RAW CUSTOMER DATA

The first step is to horizontally sort the raw data in the account opening sequence. We can utilize several different approaches to fulfill it, such as Bubble Sort Method, Rotate and Transpose Method, Call Sort Method etc. Our previous paper presented and illustrated these methods in detail already[1]. Hereby we choose to use PROC SORT and ARRAY approaches to do it attributed to the big flexibility. Below is the SAS code.

```
************************** Horizontally Sort Data. **********************;

libname A "file path\WLMU2015\Product Sequence";

proc sort data= A.Raw ;
by Client_ID Account_Type Opened_Date;
run;

proc sort data=A.Raw  out=Raw_Deduped  nodupkey;
by Client_ID Account_Type ;
run;

proc sql;
create table Prod_Count as
select count(*) as Count
from Raw_Deduped
group by Client_ID;

select max(Count) into : N trimmed
from Prod_Count;
quit;

%let N=&N;

proc sort data= Raw_Deduped  ;
by Client_ID Opened_Date;
run;
```

```
data Open_Sequence;
set Raw_Deduped ;
by Client_ID Opened_Date;

retain  Prod_Type_1- Prod_Type_&N
        Purchase_Date_1 - Purchase_Date_&N;

array Prod(&N) $20 Prod_Type_1- Prod_Type_&N;
array Open(&N) Purchase_Date_1 - Purchase_Date_&N;

if first.Client_ID then do I=1 to &N;
CNT=0;
call missing(Prod(I), Open(I));
end;

CNT+1;
Prod(CNT)=Account_Type;
Open(CNT)=Opened_Date;


if last.Client_ID then output;
drop I CNT Account_ID Account_Type Opened_Date;
format Purchase_Date_1 - Purchase_Date_&N yymmdd10. ;
run;


data Final;
retain Client_ID
Prod_Type_1   Purchase_Date_1
Prod_Type_2   Purchase_Date_2
Prod_Type_3   Purchase_Date_3
Prod_Type_4   Purchase_Date_4
Prod_Type_5   Purchase_Date_5
Prod_Type_6   Purchase_Date_6
Prod_Type_7   Purchase_Date_7
Prod_Type_8   Purchase_Date_8;
set Open_Sequence;

label
Prod_Type_1="1st Product" Purchase_Date_1="Purchase Date"
Prod_Type_2="2nd Product" Purchase_Date_2="Purchase Date"
Prod_Type_3="3rd Product" Purchase_Date_3="Purchase Date"
Prod_Type_4="4th Product" Purchase_Date_4="Purchase Date"
Prod_Type_5="5th Product" Purchase_Date_5="Purchase Date"
Prod_Type_6="6th Product" Purchase_Date_6="Purchase Date" ;
run;

Title "Banking Product Purchase Sequences By Financial Customers" ;
proc print data=Final noobs label;
run;
```

In this business case, our analysis is focused on the product purchase sequences. Therefore, for each banking customer, we only keep the earliest opened account for the same account type. Hence, as shown above, we first sort the raw financial customer data by Client_ID, Account_Type and Opened_Date, and then sort it again by Client_ID and Account_Type with the NODUPKEY option. Consequently, the sorting will remove duplicates and only retain the earliest opened account for each client and each account type. This deduping is critically important in this product purchase sequence analysis. After deduping, we apply PROC SQL to count the maximum accounts that a client may have,

which is needed for defining the number of maximum elements of the arrays in the following data step. This number of maximum accounts is calculated and assigned to a macro variable N. Then we sort the deduped data set again by Client_ID and Opened_Date, and transpose the vertically sorted data into a horizontal sort through the use of SAS arrays in the next DATA step. In this DATA step, we SET the vertically sorted data by Client_ID and Opened_Date, and define a character array Prod (length=$20) to hold the values of account types and a numeric array Open to keep the values of account opened dates. The number of elements of the two arrays is determined by the macro variable N. The use of the RETAIN statement is necessary to retain the values across data step iterations, and a counter variable CNT is created by the SUM statement to ensure the correct assignments of variable values into the corresponding elements in arrays. The following DATA step is aimed to arrange the variables in the desired order by the RETAIN statement[2], and to label the variables with appropriate headings. The above data manipulations will sort the data horizontally in the order of opened dates. Table 4 shows the structure of the generated data set Final.

Table 4. Partial Print Out of the Final Date Set: In An Ascending Order Of Purchase Date

Client_ID	1st Product	Purchase Date	2nd Product	Purchase Date	3rd Product	Purchase Date	4th Product	Purchase Date	5th Product	Purchase Date	6th Product	Purchase Date
1001	Chequing	2011-05-26	Savings	2011-05-26	TFSA	2011-06-01		.		.		.
1002	Chequing	2012-06-11	Credit Card	2012-10-24	Credit Insurance	2012-10-24	Savings	2012-11-07		.		.
1003	Credit Card	2010-11-29	Chequing	2012-06-05	Savings	2012-08-09	TFSA	2012-12-22	Mortgage	2013-07-11	Loan	2013-10-30

As shown in Table 4, the banking products of each financial customer are sorted and presented in their purchase sequences. For example, Customer 1001 has 3 types of banking products with the financial institution, his first, second and third purchased products are Chequing, Savings and TFSA respectively. This horizontal data sort is not only a good way for insightful data reporting, more important, it will also render us to conduct product purchase sequence analyses conveniently.

PART II. PRODUCT PUCHASE SEQUENCE ANALYSES

After horizontally sorting the customer data, we can perform product purchase sequence analyses easily through frequency distribution studies. These analyses allow us to leverage the customer data and develop keen insights into customers with respect to their business behaviors.

Overall Product Purchase Sequence Analysis

As illustrated below, a simple one-way frequency distribution by PROC FREQ procedure gives us insights into the financial customers relating to their purchase sequences. Here we only show the first three purchased products and the top 5 ranking products. The following is the SAS code and analytic results.

```
****************** Product Purchase Sequence Analysis *****************;

proc freq data=Final order=freq nlevels;
tables Prod_Type_1  Prod_Type_2 Prod_Type_3/missing;
run;
```

Table 5. Banking Product Purchase Patterns of Financial Customers.

Ranking	1st Purchased Product	Clients(#)	Clients(%)	2nd Purchased Product	Clients(#)	Clients(%)	3rd Purchased Product	Clients(#)	Clients(%)
1	Chequing	580,597	22.7%	Savings	318,171	12.8%	Credit Card	181,995	7.1%
2	Credit Card	304,540	13.3%	Credit Card	291,198	11.6%	Savings	173,087	6.5%
3	Savings	304,121	13.2%	Credit Insurance	280,253	11.3%	TFSA	160,478	6.3%
4	TFSA	281,670	12.2%	TFSA	249,403	10.5%	Credit Insurance	136,779	5.4%
5	Mortgage	239,208	10.7%	Chequing	158,380	8.3%	Chequing	87,606	4.1%

As shown in Table 5, among the 2.56 million financial customers under study, Chequing, Credit Card and Savings are the three most common products in their first purchase. The need of Chequing product is much higher than the Credit Card and Savings products (22.7% vs. 13.3%). Therefore, if we want to acquire new customers, Chequing, Credit Card and Savings are the top products for our promotions. In their next purchase, the top 3 preferred products are Savings, Credit Card and Credit Insurance respectively. As for the third purchased products, Credit Card is the most popular one followed by Savings and TFSA. If needed, we can extend this kind of analyses to the fourth, fifth and sixth purchased products and look into their purchase patterns. Therefore, we can probe and learn about the banking behavior and purchase tendency of our financial customers through this purchase sequence analysis.

Product Affinity Analysis

Affinity analysis is a data analysis and data mining technique that discovers co-occurrence relationships among activities performed by (or recorded about) specific individuals or groups[3]. For example, in retail industry, product affinity analysis helps one to detect and identify the products which have a high probability of being purchased together. This information can then be used to design profitable product bundles for cross-selling and up-selling purposes. In addition to increasing sales revenue, the information can also provide insights for customer loyalty programs, store design, and discount plans. This analysis has been given many other names such as Market Basket analysis, Anchor Attach analysis, Cross Shop analysis etc.

Similarly, in financial customer marketing, for cross- sell or up-sell marketing campaigns, one useful strategy is to design and promote profitable product bundles or product packages. This strategy is very effective to deepen customer relationships and increase marketing productivity. To implement this strategy, we need product affinity analysis to investigate and find out what are the best selling product bundles, and then design profitable product bundles and campaign offers accordingly. We can achieve this goal through horizontal data sorting technique too, which is illustrated as follows.

```
************************ Product Affinity Analysis. ********************;
***************** Single Product Customers.*************;
%let Start= Date1;          *To specify a start date. ;
%let End =  Date2;          *To specify an end date.

data Single_Prod_Clients;
set Final;
where not missing(Prod_Type_1) and missing(Prod_Type_2);
Bundle_1=Prod_Type_1;

if Purchase_Date_1 >= &Start and Purchase_Date_1 <= &End;
run;

proc freq data=Single_Prod_Clients order=freq;
tables Bundle_1/missing;
run;
```

```
***************** Two Product Customers.************;
data Two_Prod_Clients;
set Final;
where not missing(Prod_Type_1) and not missing(Prod_Type_2)
      and missing(Prod_Type_3) ;

if Purchase_Date_1 >= &Start and Purchase_Date_2 <= &End;
Bundle_2=catX("/", Prod_Type_1, Prod_Type_2 );
run;

proc freq data=Two_Prod_Clients order=freq;
tables Bundle_2/missing;
run;

***************** Three Product Customers.************;
data Three_Prod_Clients;
set Final;
where not missing(Prod_Type_1) and not missing(Prod_Type_2)
      and not missing(Prod_Type_3) and missing(Prod_Type_4);
Bundle_3=catX("/", Prod_Type_1, Prod_Type_2, Prod_Type_3);

if Purchase_Date_1 >= &Start and Purchase_Date_3 <= &End;
run;

proc freq data=Three_Prod_Clients order=freq;
tables Bundle_3/missing;
run;
```

As everyone knows, financial product purchases are usually not concurrent and time-taking in nature, which are very different from retail product purchases. Therefore, in financial product affinity analysis, it is a better study approach to define and specify a prolonged time duration as the purchase window based on business insights. For example, we can specify the time duration of a promotion campaign as our purchase study window. The products purchased by a customer during this time window are considered co-occurrence events and the event probabilities are then studied.

As demonstrated above, to look into the business behaviors of financial customers with different banking products, we first use a DATA step to filter and segment them, and then apply PROC FREQ to study the structure and distributions of banking products and product bundles. It is worthy to note that the frequency distribution also represents the client penetration of banking products and product bundles. Table 6 demonstrates the analysis results, here we only show the top 3 preferred products.

Table 6. Structure and Penetration of Banking Products and Product Bundles

	Structure	Ranking	Client Penetration (%)
Single Product	Credit Card	1	19.8%
	Savings	2	14.9%
	Chequing	3	13.3%
Two Products	Credit Card / Credit Insurance	1	9.3%
	Chequing / Savings	2	7.9%
	Chequing / TFSA	3	7.2%
Three Products	Chequing / Savings / Credit Card	1	5.6%
	Chequing / Credit Card / Mortgage	2	5.2%
	Chequing / Savings / TFSA	3	3.7%

Table 6 shows the structure and penetration of banking products in financial customers. Among single product customers, Credit Card, Chequing and Savings accounts are the top 3 most popular products, their client penetrations are 19.8%, 14.9% and 13.3% respectively. Among two product customers, the three most common product bundles are Credit Card / Credit Insurance, Chequing / Savings, Chequing / TFSA respectively. Please note that a higher client penetration also suggests a higher affinity of the bundled products. As for three product bundles, Chequing / Savings / Credit Card bundle has the highest client penetration, followed by the Chequing / Credit Card / Mortgage and Chequing / Savings / TFSA bundles. We can continue the above analyses to more than three products bundles if needed.

Through above penetration analysis, we can learn customers' banking preferences and gain insights for affinity marketing, because the client penetration is an indication of product affinity. For example, to acquire new customers, Credit Card, Chequing and Savings products are presumably the attractive ones for them. If we want to design cross-sell product bundles, Chequing, Savings, Credit Card and TFSA are the high-affinity elements for bundling together according to our analysis.

Next Best Offer Analysis

For up-sell and cross-sell campaigns, it is important to understand our existing customers' needs and offer them the products that really interest them. In marketing research, there exist several methods to serve this purpose, such as association analysis, clustering analysis and predictive modeling etc. Under most circumstances, these methods require advanced statistical expertise and techniques, and they are high-costing and time-taking in nature. As an alternative, we can achieve these analysis goals easily using a two-way frequency distribution study based on the horizontally sorted data. The approach and results are exampled as follows.

```
************************ Next Best Offer Analyses. *********************;

data Second_Prod_Sequence;
set Final;
where not missing(Prod_Type_1) and not missing(Prod_Type_2);
run;

proc freq data=Second_Prod_Sequence order=freq nlevels;
tables Prod_Type_1*Prod_Type_2/missing;
run;

data Third_Prod_Sequence;
set Final;
where not missing(Prod_Type_1) and not missing(Prod_Type_2)
      and not missing(Prod_Type_3);
Bundle_2=catX("/", Prod_Type_1, Prod_Type_2 );
run;

proc freq data=Third_Prod_Sequence order=freq nlevels;
tables Bundle_2*Prod_Type_3/missing;
run;
```

Table 7. Next Product Purchase Patterns Of Single Product Customers

Current Product	Next Purchased Product (Top 5)					
Chequing	Product	Savings	Credit Card	Credit Insurance	TFSA	Loan
	% of Clients	32.8%	20.7%	11.2%	10.9%	7.5%
Credit Card	Product	Chequing	Savings	TFSA	Mortgage	Line of Credit
	% of Clients	28.9%	20.1%	15.7%	13.8%	6.3%
Savings	Product	Chequing	TFSA	Mortgage	GIC	Credit Card
	% of Clients	40.7%	19.4%	13.5%	9.8%	5.2%

Table 7 presents the frequency distributions of next products purchased by single product customers, arranged in the descending order of purchase frequencies. It should be noted that the frequency can also be interpreted as the probability of a customer to purchase his next product. For example, for all the customers starting with Chequing account as his current product, most of them (32.8%) opened Savings account in their second purchase. It suggests that a customer with Chequing as his current product is most likely to open a Savings account in his next purchase. The probability of this event to occur is 32.8%, provided that the events are independent of one another. Furthermore, we can see that the product type and event likelihood of next purchases are largely dependent on the current products. Chequing customers' next favorite products are Savings, Credit Card and Credit Insurance and so on, but Savings customers are most interested in Chequing, TFSA and Mortgage products in their next purchase.

In a similar way, we can apply this analysis to the financial customers with two current products and investigate their next purchase tendency. Table 8 gives the analysis results.

Table 8. Next Product Purchase Patterns Of Two Product Customers

Current Products	Next Purchased Product (Top 3)			
Chequing + Savings	Product	TFSA	Credit Card	Credit Insurance
	% of Clients	29.1%	15.9%	11.3%
Chequing + Credit Card	Product	Savings	TFSA	Mortgage
	% of Clients	31.5%	28.7%	15.9%
Savings + Credit Card	Product	Chequing	TFSA	Line of Credit
	% of Clients	42.7%	21.8%	10.5%

It can be seen from Table 8 that a customer's purchase behavior depends on what products he currently has. If a customer currently has Chequing and Savings accounts, his next purchase is most likely to be TFSA with a probability of 29.1%. However, a customer with Chequing and Credit Card accounts may be interested in both Savings and TFSA products since the probabilities of the two next opened products are very comparable (31.5% vs. 28.7%). The Savings and Credit Card customers are most likely to get Chequing in their next purchase (probability=42.7%), which is notably higher than that for TFSA (21.8%) and Line of Credit (10.5%).

Therefore, the Next Best Offer analysis can provide us with valuable information of customers' banking patterns and purchase tendencies. We can thus develop competitive marketing strategies accordingly.

When To Provide The Next Offer? Purchase Time Span Analysis

In customer marketing, we not only need to know what is the best product to offer our customers, we also need understand when we should provide them the offer. In short, for successful marketing, we should give our customers the right offer at the right time. They are both critically important in developing marketing strategies. Therefore, it is essential to study the banking pattern of financial customers in terms

of the time duration and figure out the right time to contact them. Horizontal data sorting technique enables us to conduct this time span analysis too, below gives an example of this time span analysis.

```
********************* Purchase Time Span Analysis. *********************;

data Duration;
set Final;
where not missing(Prod_Type_1) and not missing(Prod_Type_2);
Duration_1= Opened_Date_2 - Opened_Date_1;
run;

proc means data= Duration N Min Max Mean STD maxdec=1;
class Prod_Type_1;
var Duration_1;
run;

proc means data= Duration Nway N Min Max Mean STD maxdec=1;
class Prod_Type_1 Prod_Type_2;
var Duration_1;
run;
```

Table 9. Purchase Time Span Analysis of Financial Customers

Current Product	Next Product	Time Span To Purchase Next Product (days)				
		N	Min	Max	Mean	Standard Deviation
Chequing	**Overall**	2,135,673	0	636	**183.8**	221.3
	Savings	704,772	1	507	118.2	85.7
	Credit Card	491,205	2	325	**93.5**	81.6
	TFSA	234,924	26	629	254.6	193.3
Credit Card		N	Min	Max	Mean	Standard Deviation
	Overall	2,236,539	0	568	**257.6**	221.7
	Chequing	626,231	0	512	**192.1**	104.9
	Savings	469,673	1	533	216.0	92.4
	TFSA	351,137	68	554	235.8	162.6

As shown above, the horizontally sorted data enables us to conduct the time span analysis easily. In the DATA step, we first select all the customers who have at least two products using the WHERE statement, and then calculate the time duration between the first and second purchases. PROC MEANS is then employed to calculate the statistics of the time duration to purchase next products. The analysis results are given in Table 9.

Table 9 reveals that for financial customers with Chequing as his current product, the overall average time duration for them to purchase next products is approximately 6 months (183 days). Therefore, the appropriate time to contact these existing customers for cross-sell purposes is probably 6 months after they purchase their first product. However, customers with Credit Card as their current product have longer time duration to purchase their next products, which is 257 days on average. We can also see that, the average time spans differ remarkably for purchasing different next products. It usually takes Chequing customers 93 days to open a Credit Card account, while it is much longer for them to purchase a TFSA product (254 days). Through these analyses, we can find the time span patterns and determine the correct time to contact them for marketing different products.

Therefore, as shown above, we can perform time span analyses through the horizontal data sorting technique. Although we only illustrate it with a very simple example here, however, this technique can be easily extended to apply in more complicated cases based on different analysis requests.

CONCLUSION

As discussed in this paper, horizontal data sorting is a very useful SAS technique. Utilization of this technique can significantly enhance the format and layout of data analysis and presentation, which is very helpful in data presentation and insightful data reporting. Much more important, it is also an innovative and valuable technique in advanced data analysis and business analytics. We can use it in product sequence analyses and other business analytics fields. This data levering technique has the distinct advantages of being simple, straightforward, convenient to use, and producing easy-to-interpret analysis results etc. It can find important applications in a wide variety of business analytics and business intelligence fields.

REFERENCES

[1] Justin Jia and Amanda Lin, "**Horizontal Data Sorting and Insightful Reporting: A Useful SAS® Technique**", 376-2013, 2013 SAS Global Forum, San Francisco, CA, USA.

[2] SAS Support Website,
http://support.sas.com/documentation/cdl/en/lrdict/64316/HTML/default/viewer.htm#a000214163.htm

[3] Jiawei Han & Micheline Kamber, "**Data Mining: Concepts and Techniques**", ISBN 978-1-55860-901-3, Elsevier Inc , 2006.

ACKNOWLEDGEMENTS

Special thanks to Dr. Arthur Tabachneck (Analyst Finder) for his valuable feedback and suggestions during reviewing this paper. We are very grateful to him for his great help and support.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Justin Jia TransUnion Canada, Burlington, Ontario, Canada

Amanda Lin CIBC, Toronto, Ontario, Canada