Paper 042-2010

Using Base SAS® and SAS® Enterprise Miner™ to Develop Customer Retention Modeling

Rex Pruitt, PREMIER Bankcard, LLC, Sioux Falls, SD

ABSTRACT

In this paper I will describe how to develop the components necessary using SAS tools and business analytics to effectively identify a “Good Customer”.

The resulting “Good Customer Score” is intended for use in modeling exercises designed to help improve the cost effectiveness and development of Acquisition, Rewards, Retention, and Recovery efforts at PREMIER.

“Good Customer Score” has been added to PREMIER’s Data Warehouse and is being used to develop and implement specific targeted Retention and other strategies with an estimated $50+ Million in annual revenue lift identified thus far.

Estimated Opportunity Value:
For example, reduce the attrition of PREMIER’s “Top Good Customers” >= 2 Years on Book = $15+ Million annually (see also SAS Global Forum Paper 278-2009.pdf). The total of all opportunities noted in this paper exceeds an estimated $70 Million in annual revenue lift.

Portfolio Scoring & Ranking Specifics:
The accuracy of the new “Good Customer Score” is supported by the statistical correlation to Behavior Score (3rd party score), as well as other scores, when identifying those customers who will perform in the top 25% of the Portfolio ranked by Good Customer Score (Target). The strength of like scores is noted in the Chi-Square correlation results. Additionally, the statistical soundness of the score comparison exercise performed using modeling in SAS Enterprise Miner is supported by a KS Statistic of 58 and a target prediction accuracy of 85%.

INTRODUCTION

Business analytics using Base SAS and predictive modeling using SAS Enterprise Miner is very powerful and capable of generating significant lifts in revenue for the organization. The example illustrated in the context of this paper is a clear depiction of the benefits that result from the application of business analytics and predictive modeling to solve "Customer Intelligence" business problems. Using SAS tools in customer retention is having a significant impact on our Company's business. Venturing into the huge amounts of internal customer data and information can be a daunting task. However, by employing SAS Enterprise Miner, coupled with some Base SAS techniques, gold nuggets (sometimes chunks) can be identified.

In this paper I will describe how to develop the components necessary using SAS tools and business analytics to effectively identify a "Good Customer." This "Good Customer Score" can then be used in predictive modeling exercises designed to help improve the cost effectiveness and development of Retention efforts, as well as, other customer focused programs.

METHODS

"Good Customer Score" Development – A Base SAS program was written that applies data step logic against internal customer data that has been cleansed and adjusted for outliers. Key customer performance measures were used in a weighting process to generate a ratio representative of a "Good Customer" by our company's definition. The attributes used are noted in the "Definition Matrix" (see Figure 1).

Measure	*Portfolio Mean*	*Definition*
1) FeeBalRatio	0.63	**Revenue** Value Measure: Internal calculation is proprietary. Measure of the outstanding balance of billed fees on the customer's account in relation to the annualized collected fee value.
2) TotCreditUtilRatio	0.20	**Risk** Exposure Value Measure: Internal calculation is proprietary. Measure of the total credit utilization on the customer's account.
3) PrinBalExposureRatio	0.29	**Risk** Exposure Value Measure: Internal calculation is proprietary. Measure of the principal balance exposure on the customer's account.
4) OCLOccuredRatio	0.75	**Behavior** Experience Measure: Internal calculation is proprietary. Measure representing the number of times a customer has exceeded the credit limit.
5) OCLThresholdRatio	0.75	**Behavior** Experience Measure: Internal calculation is proprietary. Measure representing the number of times a customer has exceeded the credit limit threshold.
6) DelRatio	0.65	**Behavior** Experience Measure: Internal calculation is proprietary. Measure representing the number of times a customer has been delinquent.
7) MOBRatio	0.65	**Loyalty** Measure: Internal calculation is proprietary. Measure representing the longevity of the customer's account with PREMIER.
GoodCustRatio	0.56	Mean value of the 7 individual customer performance measures.
GoodCustRankNum	10.50	Mean rank value for the portfolio.
GoodCustScore	561	Conversion of Good Customer Ratio to a representative whole number score value.

Figure 1. Definition Matrix - Good Customer Score (GCS)

Portfolio Scoring & Ranking – The “Good Customer Score” was then used to rank the portfolio of customers and bin them into 20 different buckets. Doing so isolates the customer profile characteristics that define our best vs. worst customers. This sets the stage for model and program development. It becomes clear how to “target” the specific segments for respective treatment depending on the customer’s needs (e.g., acquire more good customer “look-a-likes,” retain and/or reward the best performers, take corrective action on poor performers).

SAS PROGRAMMING LOGICAL FLOW

While I cannot share the actual data step programming, variables, or formulas used in full context to generate the Good Customer Score (GCS); I can provide the basic logical flow and abstract examples of the various Base SAS coding methods including some sample code and output (see Figure 2).

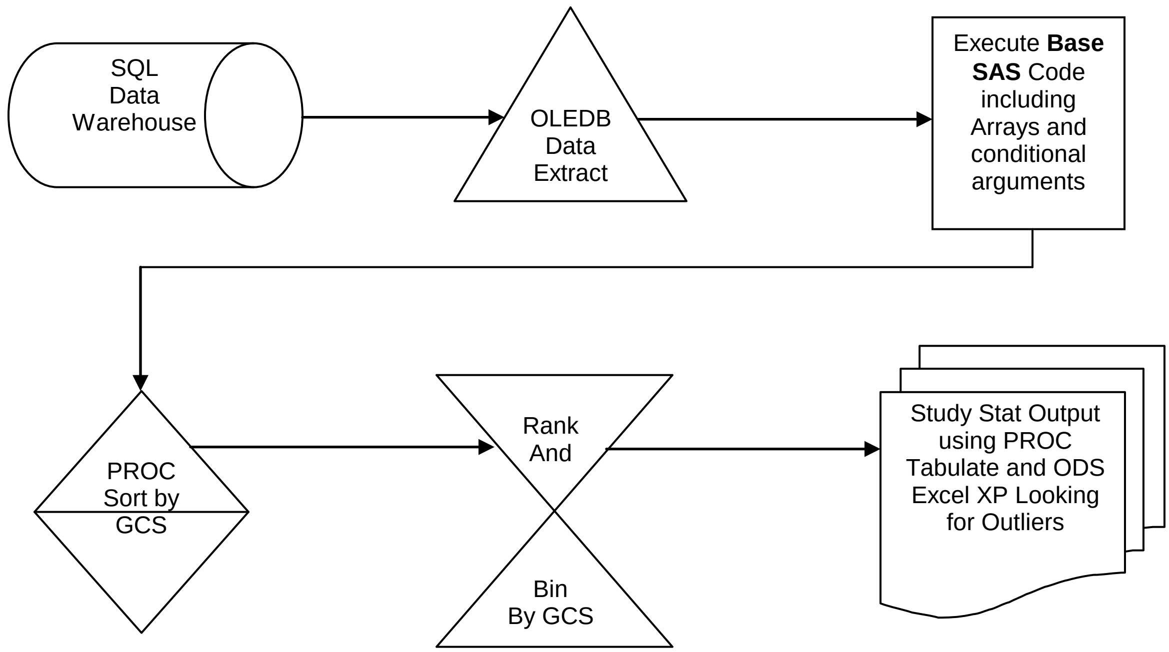

Figure 2. Base SAS Programming Flow Diagram

BASE SAS CODE DESIGN TO IDENTIFY "GOOD CUSTOMERS"

In order to identify the "Good Customers (GC)" within there has to be a clear understanding of what a GC looks like. To the CFO, a GC is a customer that generates high revenues and does not create long term maintenance issues. To the Operations Executives, a GC is a customer that handles the details of their account very well. The challenge is combining these two spectrums into a composite view that describes the best of both worlds. This is not an uncommon problem and there are several books describing the need for customer focus. In fact, Customer Relationship Management (CRM) is a business model that has been touted since the early 1990's (maybe earlier). There are even companies that have been formed that will provide the service of implementing the CRM model and/or corresponding metrics in any organization for a price.

This paper is focused on simply "how to develop customer retention modeling" through the development of an internal score. In order to target the GC effectively for a "Retention" program (or other business problem for that matter), the problem noted in the prior paragraph must be answered using data contained within the organization's customer portfolio. I would suggest beginning with a simple series of questions focused on identifying the data attributes used to describe the customer's Revenue, Risk, Behavior, and Loyalty. The resulting components I used are noted in Figure 1. By taking these specific representative data elements and applying the logical programming to create the resulting "Good Customer Ratio (GCR)," I was able to calculate a measure that can be used to identify the best and worst customers.

The following Base SAS code illustrates the flow of this methodology programmatically within the data step program data vector. I'm sorry, but I cannot share the actual formulas for the individual components.

```
/* Calculate the individual component Ratio Values */
        /* Calculate the Fee Balance Ratio */
              FeeBalRatio=…
        /* Calculate the Credit Utilization Ratio */
               TotCreditUtilRatio=…
        /* Calculate the Principle Balance Exposure Ratio */
               PrinBalExposureRatio=…
        /* Calculate the OCL Occurred Ratio */
               OCLOccuredRatio=…
        /* Calculate the OCL Threshold Ratio */
               OCLThresholdRatio…
        /* Calculate the Delinquency Ratio */
               DelRatio…
        /* Calculate the Vintage Months on Book (MOB) Ratio */
               MOBRatio…
        /* Calculate the Non-Weighted Average GoodCustRatio */
               GoodCustRatio=Sum(of
                       FeeBalRatio,
                       TotCreditUtilRatio,
                       PrinBalExposureRatio,
                       OCLOccuredRatio,
                       OCLThresholdRatio,
                       DelRatio,
                       MOBRatio
                       )/7;
```

RANK AND BIN BY GCR

Once the GCR has been established at the customer record level, the entire portfolio can be ranked and binned accordingly. The GCR is transformed to Good Customer Score (GCS) by simply multiplying the value by 1,000 and rounding it to a whole number. This was done to represent the GCS as a value that is familiar to the business sector for consistency sake. Credit and/or performance viability measures are usually cast as a 2 or 3 digit number.

```
/* BEGIN Sorting & Ranking process */
Proc Sort
       Data=PSP04RM.PSP04_RM_GoodCustRatioAll_&ccyymm;
       By descending GoodCustRatio;
       run;

Proc Means Noprint Data=PSP04RM.PSP04_RM_GoodCustRatioAll_&ccyymm;
       Output
              Out=RankedTotal (rename=(_freq_=RankedTotal))
              ;
              run;

Data _Null_;
       Set RankedTotal  (Where=(_Type_=0));
       Call Symput('RankedTotal',RankedTotal);
       run;

Proc Format;
       Value DecileF
              Low-.05='01'
              .05-.1='02'
              .1-.15='03'
              .15-.2='04'
              .2-.25='05'
              .25-.3='06'
              .3-.35='07'
              .35-.4='08'
              .4-.45='09'
              .45-.5='10'
              .5-.55='11'
              .55-.6='12'
              .6-.65='13'
              .65-.7='14'
              .7-.75='15'
              .75-.8='16'
              .8-.85='17'
              .85-.9='18'
              .9-.95='19'
              .95-1='20'
              ;
              run;

Data
       PSP04RM.PSP04_RM_GoodCustRatioAll_&ccyymm
       PSP04RM.PSP04_RM_GoodCustRatio_&ccyymm
       (Keep=
              DebtDimId
              LastAcctId
              GoodCustRank
              GoodCustRatio
              GoodCustScore
              DataSetDt
              Owner
              UpdatedDt
              )
              ;
       Length
              GoodCustRankNum 8.
              ;
       Set PSP04RM.PSP04_RM_GoodCustRatioAll_&ccyymm;
```

```
            Rank=_n_/&RankedTotal;
            GoodCustRank=Put(Rank,DecileF.);
            GoodCustRankNum=GoodCustRank;
            GoodCustScore=Round(GoodCustRatio*1000);
      Output PSP04RM.PSP04_RM_GoodCustRatioAll_&ccyymm;
      Output PSP04RM.PSP04_RM_GoodCustRatio_&ccyymm;
      run;

/* END Sorting & Ranking process */
```

NOTE: There are several methods available in SAS to perform the binning and ranking process. The intent of this paper is not to explore every possible method. However, the method that I used seems to have generated effective results and is easy to understand and communicate.

STUDY STAT OUTPUT

While creating the GCS is the most critical step in the process, the results must be studied effectively in order to make any sense of the GCS' viability and potential used in solving any business problems. Of course, in the end, the objective is to recommend a program to business managers that they will trust and, ultimately, embrace the solution. Tools, such as PROC Tabulate and Output Delivery System (ODS) directed to the Excel XP Tagset were very nice for examining the results of the data step and procedure output. Hopefully, the example code and corresponding output will demonstrate how these SAS tools were utilized.

The following code steps are used to push the output to an MS Excel spreadsheet using the ODS ExcelXP Tagset. Certain SAS System Option and Tagset Option defaults need to be overridden in order to get the desired formatted output. Placing the output in a tool such as this can enrich the review and discovery process. It also makes it easier to deliver the results to "non-SAS" analysts or members of management for review.

```
Options NoDate Label Missing='0' Orientation=Landscape SPOOL;
%inc '\\pbidelprd042\DM_Inputs\01_Shared_SAS_Code\ExcelXP_Tagset.sas';
options
      topmargin=.25in
      bottommargin=.25in
      leftmargin=.25in
      rightmargin=.25in
      ;

ods listing close;
ODS Tagsets.ExcelXP
      style=Journal
      file="\\pbidelprd042\DM_Inputs\rpruitt\PSP04\PSP04_RM_GoodCustRatio_&ccyymm..xls
"
      options (
            doc='help'
            default_column_width='20'
            Orientation="landscape"
            AUTOFIT_HEIGHT="no"
            CENTER_HORIZONTAL="yes"
            EMBEDDED_TITLES="yes"
            EMBEDDED_FOOTNOTES="yes"
            FITTOPAGE="yes"
            Frozen_Headers='5'
            Row_Repeat='1-5'
            THOUSANDS_SEPARATOR=','
            CURRENCY_SYMBOL='$'
            CURRENCY_FORMAT='Currency'
            DECIMAL_SEPARATOR='.'
            EMBED_TITLES_ONCE='yes'
            )
      ;
```

This code was used to extract 50 random records for review out of the over 3 million scored accounts. PROC Surveyselect is a very useful tool for reducing the time it takes to perform a results review of the detail. It is very important to understand how the logical application of programming code used to manipulate the data within the program data vector is performing.

```
ODS Tagsets.ExcelXP Options (sheet_name='SampleSelectAllCols' SHEET_INTERVAL='none');
Proc Surveyselect
       data=PSP04RM.PSP04_RM_GoodCustRatioAll_&ccyymm
       out=SamplePrintAll
       SAMPSIZE=50
       Method=SRS
       Seed=12345
       ;
       %title;
       %Footnote;
       run;
```

PORTFOLIO STRATIFICATION PROJECT (PSP)

PSP04 - Retention Model (RM) Good Customer Identification Ratio

Portfolio Data as of 200812 with a Run Date of   January 12, 2009 - 21:56:34

Selection Method	Simple Random Sampling

NOTE: This information is CONFIDENTIAL and intended for internal MIS Analytics use only.

Please contact Rex Pruitt x3810 with questions

file=\\pbidelprd042\DM_Inputs\rpruitt\PSP04\PSP04_RM_GoodCustRatio_200812xls

Input Data Set	PSP04_RM_GOODCUSTRATIOALL_200812
Random Number Seed	12345
Sample Size	**50**
Selection Probability	0.000016
Sampling Weight	60733.72
Output Data Set	SAMPLEPRINTALL

Figure 3. PROC Surveyselect ODS Output

NOTE: Each PROC Step that produces printed output is written to a new tab in the Excel spreadsheet.

A simple PROC Print is used to display variables of interest from the 50 random records out of the over 3 million scored accounts. These are not the only values that were evaluated within the context of this project, but should sufficiently illustrate the review concept.

```
ODS Tagsets.ExcelXP Options (sheet_name='RatioPrint' SHEET_INTERVAL='none');
Proc Print data=SamplePrintAll
       noobs label n
       ;
       Var
              DebtDimId
              VintMos
              GoodCustRatio
              GoodCustRank
              GoodCustScore
              / style(header data) = {just=C}
              ;
       %title;
       %footnote;
       run;
```

PORTFOLIO STRATIFICATION PROJECT (PSP)

PSP04 - Retention Model (RM) Good Customer Identification Ratio

Portfolio Data as of 200812 with a Run Date of January 12, 2009 - 21:56:34

DebtDimId	*VintMos*	*GoodCustRatio*	*GoodCustRank*	*GoodCustScore*
13230495	27	0.97568	1	976
16712333	9	0.92603	2	926
13481835	25	0.92394	2	924
15426230	16	0.88055	2	881
9697616	46	0.80986	3	810
14542701	20	0.79891	4	799
15226659	17	0.79877	4	799
... 8 to 45 removed …	…	…	…	…
18079383	1	0	17	0
17799485	3	0	18	0
17682248	3	0	19	0
18173309	1	0	20	0
18341623	0	0	20	0

N = 50

NOTE: This information is CONFIDENTIAL and intended for internal MIS Analytics use only.

Please contact Rex Pruitt x3810 with questions

file=\\pbidelprd042\DM_Inputs\rpruitt\PSP04\PSP04_RM_GoodCustRatio_200812xls

Figure 4. Proc Print ODS Output

PROC Tabulate was then used to display specified statistics for variables of interest for the over 3 million scored accounts. Reviewing these statistical results is important in understanding the viability of how the resulting measure and its component variables relate to the overall population.

```
ODS Tagsets.ExcelXP Options (sheet_name='DataStatsAllCols' SHEET_INTERVAL='none');
Proc Tabulate
       data=PSP04RM.PSP04_RM_GoodCustRatioAll_&ccyymm
       Missing
       ;
       Var
              VintMos
              Num030DayDelTot
              Num060DayDelTot
              Num090DayDelTot
              NumIADayDelTot
              OCL12MnthOccurred
              OCL12MnthThreshold
              NetFeeRev12Mo
              FeeBal
              FeeBalRatio
              TotCreditUtilRatio
              PrinBalExposureRatio
              OCLOccuredRatio
              OCLThresholdRatio
              DelRatio
              MOBRatio
              GoodCustRatio
              GoodCustRankNum
              GoodCustScore
              ;
       Table
              (VintMos
              Num030DayDelTot
              Num060DayDelTot
              Num090DayDelTot
              NumIADayDelTot
              OCL12MnthOccurred
              OCL12MnthThreshold
              NetFeeRev12Mo
              FeeBal
              FeeBalRatio
              TotCreditUtilRatio
              PrinBalExposureRatio
              OCLOccuredRatio
              OCLThresholdRatio
              DelRatio
              MOBRatio
              GoodCustRatio
              GoodCustRankNum
              GoodCustScore
                     )
                     ,
                     (N NMISS)
                            *f=comma12.*[Style=[tagattr='format:#,##0' Just=C]]
                     (Mean Min Max)
                            *f=comma12.2*[Style=[tagattr='format:#,##0.00' Just=C]]
                     (T PROBT VAR STDDEV)
                            *f=comma14.4*[Style=[tagattr='format:#,##0.0000' Just=C]]
                     ;
       %Title;
       %Footnote;
       run;

ODS _ALL_ CLOSE;
ODS Listing;
```

PORTFOLIO STRATIFICATION PROJECT (PSP)

PSP04 - Retention Model (RM) Good Customer Identification Ratio

Portfolio Data as of 200812 with a Run Date of   January 12, 2009 - 21:56:34

	N	*NMiss*	*Mean*	*Min*	*Max*	*t*	*Probt*	*Var*	*StdDev*
VintMos	2,853,975	182,711	24.18	1.00	243.00	1,642.7147	<.0001	618.5651	24.8710
Num030DayDelTot	3,036,686	0	1.33	0.00	12.00	1,543.9612	<.0001	2.2645	1.5048
Num060DayDelTot	3,036,686	0	0.36	0.00	12.00	885.2232	<.0001	0.5132	0.7164
Num090DayDelTot	3,036,686	0	0.29	0.00	12.00	605.9204	<.0001	0.7007	0.8371
NumIADayDelTot	3,036,686	0	0.00	0.00	12.00	56.2776	<.0001	0.0139	0.1178
OCL12MnthOccurred	3,036,670	16	1.86	0.00	41.00	1,616.7081	<.0001	4.0282	2.0070
OCL12MnthThreshold	3,036,670	16	1.17	0.00	29.00	1,408.5088	<.0001	2.0868	1.4446
NetFeeRev12Mo	3,036,686	0	209.21	0.00	1,731.99	2,781.6611	<.0001	17,178.0983	131.0652
FeeBal	2,806,113	230,573	97.41	0.00	8,663.48	1,538.6169	<.0001	11,246.9727	106.0517
FeeBalRatio	3,036,686	0	0.63	0.00	1.00	2,886.3393	<.0001	0.1437	0.3790
TotCreditUtilRatio	3,036,686	0	0.16	0.00	1.00	1,049.2629	<.0001	0.0681	0.2610
PrinBalExposureRatio	3,036,686	0	0.29	0.00	1.00	1,650.5651	<.0001	0.0943	0.3070
OCLOccuredRatio	3,036,671	15	0.79	0.00	1.00	3,457.1176	<.0001	0.1593	0.3991
OCLThresholdRatio	3,036,671	15	0.79	0.00	1.00	3,457.3011	<.0001	0.1601	0.4002
DelRatio	3,036,686	0	0.69	0.00	1.00	3,259.9843	<.0001	0.1372	0.3704
MOBRatio	3,036,686	0	0.69	0.00	1.00	3,001.9947	<.0001	0.1616	0.4020
GoodCustRatio	3,036,686	0	0.58	0.00	1.00	3,299.4908	<.0001	0.0932	0.3053
GoodCustRankNum	3,036,686	0	10.50	1.00	20.00	3,173.1709	<.0001	33.2500	5.7663
GoodCustScore	3,036,686	0	578.13	0.00	1,000.00	3,299.4904	<.0001	93,231.3025	305.3380

NOTE: This information is CONFIDENTIAL and intended for internal MIS Analytics use only.

Please contact Rex Pruitt x3810 with questions

file=\\pbidelprd042\DM_Inputs\rpruitt\PSP04\PSP04_RM_GoodCustRatio_200812xls

Figure 5. Proc Tabulate ODS Output

"Good Customer Score" Validation – The accuracy of the new "Good Customer Score" is supported by the statistical correlation to Behavior Score (3rd party score), as well as other scores, when identifying those customers who will perform in the top 25% of the Portfolio ranked by Good Customer Score. The strength of various performance related scores is demonstrated in the Chi-Square correlation table (see Figure 7). Additionally, the score comparison exercise performed using modeling in E-Miner was validated with a KS Statistic of 58 and a prediction accuracy of 85% (see Figure 8).

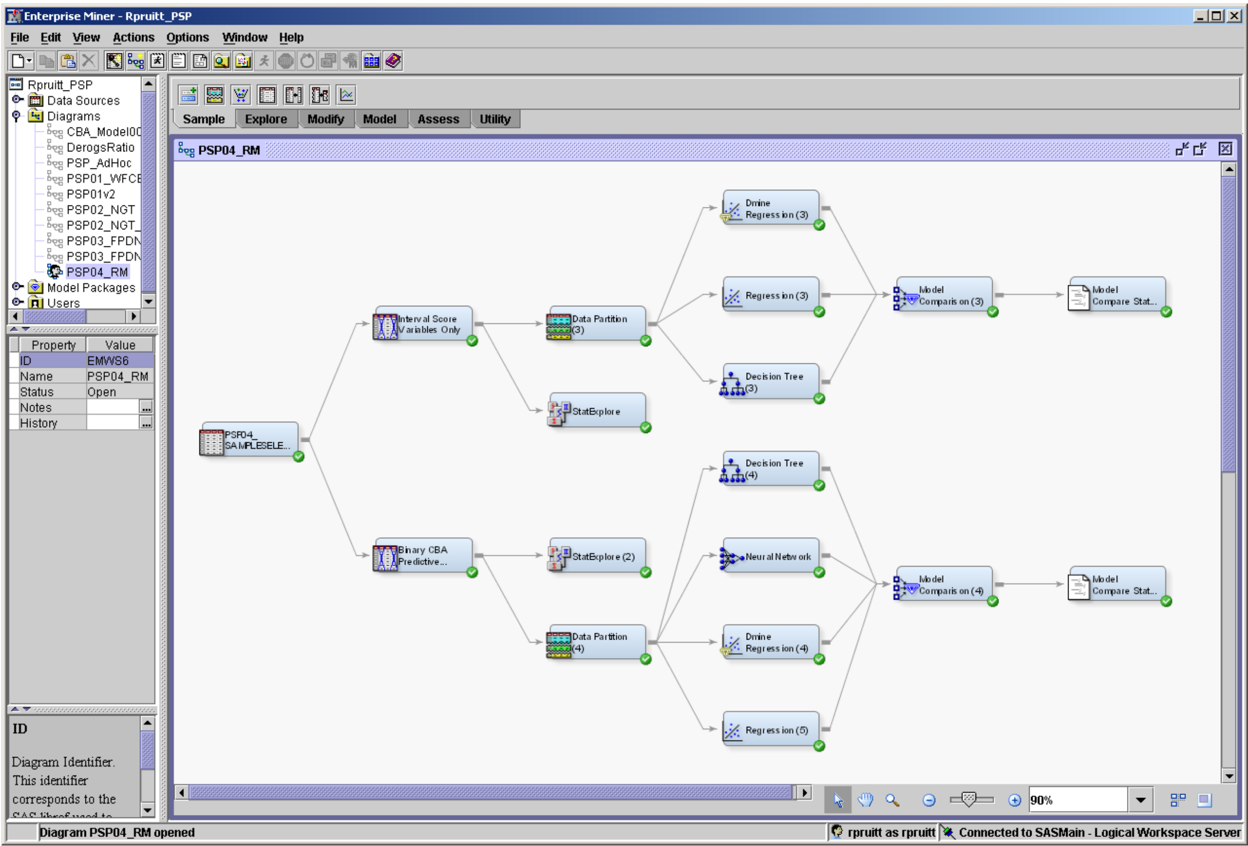

Figure 6. Enterprise Miner v5.3 Diagram Display

Input	Chi-Square	Df	Prob
BehavScore	6730.4037	289	<.0001
QFICO	1952.8028	369	<.0001
Fraud	1830.2613	98	<.0001
SAS_JM1	1470.0137	93	<.0001
SAS_JK2	1452.6677	97	<.0001
ThinDex	896.1424	331	<.0001
OriginalFICO	495.5547	325	<.0001
Experian	442.7340	82	<.0001
DMS	257.4743	100	<.0001
AustinITA	237.6065	71	<.0001
PreScr	203.4391	88	<.0001
ITA	196.3312	78	<.0001
AustinINT	148.2257	70	<.0001
IntNet	143.3236	61	<.0001

Chi-Square Statistics

Figure 7. Enterprise Miner v5.3 Chi-Square correlation table results from the Stat Explore Node

Test Type	Fit Statistic	DmineReg3	Tree3	Reg3
0-Use Indicator	Model Selected (1-Yes, 0=No)	**1.00**	0.00	0.00
1-KS	Bin-Based Two-Way Kolmogorov-Smirnov Statisti	0.58	0.49	0.57
1-KS	Kolmogorov-Smirnov Statistic	**0.58**	0.51	0.57
2-GINI	Gini Coefficient	0.74	0.54	0.64
4-Classification	Frequency of Classified Cases	8000.00	.	.
4-Classification	Misclassification Rate	**0.15**	0.16	0.16
4-Classification	Number of Wrong Classifications	1209.00	.	.
5-Error	Average Error Function	0.36	.	0.45
5-Error	Average Squared Error	0.11	0.12	0.14
5-Error	Degrees of Freedom for Error	.	.	7985.00
5-Error	Error Function	5741.28	.	7123.64
5-Error	Final Prediction Error	.	.	0.14
5-Error	Maximum Absolute Error	0.98	0.89	1.00
5-Error	Mean Square Error	.	.	0.14
5-Error	Root Average Squared Error	0.33	0.35	0.37
5-Error	Root Final Prediction Error	.	.	0.37
5-Error	Root Mean Squared Error	.	.	0.37
5-Error	Sum of Squared Errors	1767.04	1939.25	2164.64
6-Other	Akaike's Information Criterion	.	.	7153.64
6-Other	Divisor for ASE	16000.00	16000.00	16000.00
6-Other	Gain	270.06	260.30	251.42
6-Other	Lift	3.34	3.50	3.12
6-Other	Model Degrees of Freedom	.	.	15.00
6-Other	Number of Estimate Weights	.	.	15.00
6-Other	Percent Capture Response	16.69	17.50	15.61
6-Other	Percent Response	75.00	78.60	70.13
6-Other	Roc Index	0.87	0.77	0.82
6-Other	Schwarz's Bayesian Criterion	.	.	7258.45
6-Other	Sum of Case Weights Times Freq	.	16000.00	16000.00
6-Other	Sum of Frequencies	8000.00	8000.00	8000.00
6-Other	Total Degrees of Freedom	.	8000.00	8000.00

Figure 8. Enterprise Miner v5.3 Statistics from the Model Comparison Node

GCS PROJECTS UNDER CONSTRUCTION

Customer Retention Program – The “Good Customer Score” was used right away in the development of an initial customer retention recommendation. The program recommended will reduce the attrition of PREMIER’s “Top Good Customers” >= 2 Years on Book and generate an excess of $15+ Million (see Figure 9) in annual revenue. The analysis was performed using a combination of Base SAS code including PROC Tabulate and ODS Output similar to that noted in Figures 3-5 previously.

	1-Top Quartile								
		Annualized NCFR		*Retention of Best Customers*				*Good Customer Score*	
Top 50% by MOB by Annualized NCFR & Good Customer Score	*Count*	*Sum*	*Col % Sum*	*% Loss*	*Count*	*$ Loss*	*Save Value at 10% (est.)*	*Minimum Value*	*Average Value*
00-06 MOB	42,258	$9,345,268	3.88%					762	855
07-12 MOB	137,738	$43,181,777	17.92%					762	869
13-24 MOB	260,416	$82,776,641	34.36%					762	873
25-36 MOB	155,398	$43,154,629	17.91%	16.45%	105,018	$39,622,011	$3,962,201	762	877
37-48 MOB	91,715	$23,472,496	9.74%	8.17%	63,683	$19,682,134	$1,968,213	762	878
49-60 MOB	56,743	$13,878,324	5.76%	3.98%	34,972	$9,594,171	$959,417	762	878
61+ MOB	132,197	$25,117,773	10.43%					762	878
Total	876,465	$240,926,907	100.00%		203,673		**$6,889,832**	762	874

	2-Upper Middle Quartile								
		Annualized NCFR		*Retention of Best Customers*				*Good Customer Score*	
Top 50% by MOB by Annualized NCFR & Good Customer Score	*Count*	*Sum*	*Col % Sum*	*% Loss*	*Count*	*$ Loss*	*Save Value at 10% (est.)*	*Minimum Value*	*Average Value*
00-06 MOB	36,098	$7,960,884	3.06%					685	716
07-12 MOB	178,393	$55,892,757	21.46%					685	716
13-24 MOB	280,937	$95,678,907	36.73%					685	722
25-36 MOB	136,760	$41,142,443	15.79%	20.94%	144,177	$54,536,463	$5,453,646	685	723
37-48 MOB	79,039	$21,974,973	8.44%	7.36%	57,721	$19,167,470	$1,916,747	685	723
49-60 MOB	47,422	$12,623,341	4.85%	3.59%	31,617	$9,351,632	$935,163	685	724
61+ MOB	117,816	$25,226,084	9.68%					685	723
Total	876,465	$260,499,390	100.00%		233,515		**$8,305,557**	685	721

Total Opportunity Estimated Annual Value: **$ 15,195,388**

Figure 9. GCS Retention Program Customer “Save” Opportunity

Predictive Modeling for Pre-Screen Mail Program – A preliminary model has been developed that has a prediction accuracy of > 81%. The KS Statistic is > 44 using the Decision Tree training results. This project is attempting to assess the potential for prescribing a pre-screen mail test that would target "Good" customers in the top 25% of PREMIER's portfolio (see Figure 10) while venturing into a mailing universe where prior programs have failed to succeed. This project has huge revenue and profit potential as a virtually "untapped" market.

Sample=1-Train

Test Type	Fit Statistic	Tree4	DmineReg4	Reg5	Neural
0-Use Indicator	Model Selected (1-Yes, 0=No)	1.00	0.00	0.00	0.00
1-KS	Bin-Based Two-Way Kolmogorov-Smirnov Statisti	0.44	0.47	0.20	0.21
1-KS	Kolmogorov-Smirnov Statistic	0.44	0.47	0.20	0.22
2-GINI	Gini Coefficient	0.58	0.60	0.19	0.21
4-Classification	Frequency of Classified Cases	.	8000.00	.	.
4-Classification	Misclassification Rate	0.19	0.20	0.22	0.22
4-Classification	Number of Wrong Classifications	.	1582.00	.	1778.00
5-Error	Average Error Function	.	0.43	0.54	0.54
5-Error	Average Squared Error	0.14	0.14	0.18	0.18
5-Error	Degrees of Freedom for Error.	.	.	7984.00	7948.00
5-Error	Error Function	.	6835.68	8710.72	8632.60
5-Error	Final Prediction Error.	.	.	0.18	0.18
5-Error	Maximum Absolute Error	0.95	0.97	0.97	0.98
5-Error	Mean Squared Error.	.	.	0.18	0.18
5-Error	Root Average Squared Error	0.37	0.37	0.42	0.42
5-Error	Root Final Prediction Error.	.	.	0.42	0.42
5-Error	Root Mean Squared Error.	.	.	0.42	0.42
5-Error	Sum of Squared Errors	2179.71	2183.66	2837.83	2815.32
6-Other	Akaike's Information Criterion.	.	.	8742.72	8736.60
6-Other	Divisor for ASE	16000.00	16000.00	16000.00	16000.00
6-Other	Gain	188.41	182.63	21.87	30.77
6-Other	Lift	2.65	2.45	0.85	0.95
6-Other	Model Degrees of Freedom.	.	.	16.00	52.00
6-Other	Number of Estimated Weights.	.	.	16.00	52.00
6-Other	Percent Capture Response	13.25	12.27	4.23	4.73
6-Other	Percent Response	59.50	55.14	19.00	21.25
6-Other	Roc Index	0.79	0.80	0.59	0.61
6-Other	Schwarz's Bayesian Criterion.	.	.	8854.52	9099.94
6-Other	Sum of Case Weights Times Freq	16000.00	.	16000.00	16000.00
6-Other	Sum of Frequencies	8000.00	8000.00	8000.00	8000.00
6-Other	Total Degrees of Freedom	8000.00	.	8000.00	8000.00

Figure 10. GCS Pre-Screen Predictive Mail Program Statistics

NOTE:
This project is attempting to assess the potential for prescribing a pre-screen mail test that would target "Good" customers in the top 25% of PREMIER's portfolio. The inputs are made up of specific Credit Bureau Attributes and 3rd party credit scores with the intention to tap new markets.

Customer Cross-Sell Program – As an enhancement to the existing cross-sell program, this program is designed to offer qualified candidates a second product similar to the original. The “Good Customer Score” will be used to more accurately target qualified customers. Using the new score enabled the identification of a 2% increase in qualified candidates which translates into an opportunity lift of over $2.3 Million in annual revenue. Additionally, using the Good Customer Score to target qualified candidates more accurately, PREMIER has an opportunity to realize a conservative $24 Million lift in annual revenue (see Figure 11).

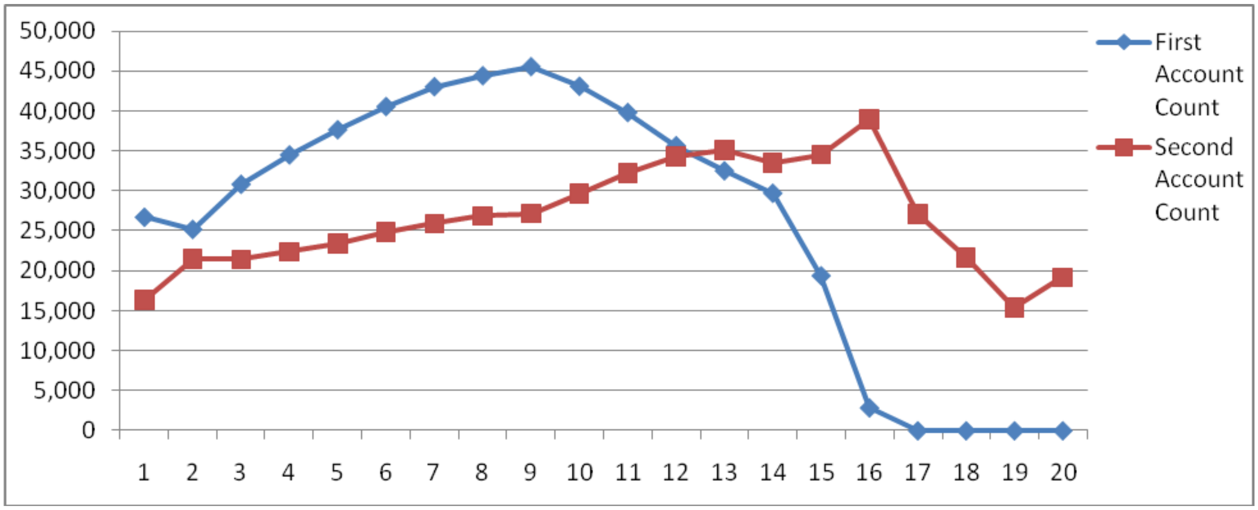

Figure 11. Customer Score “Cross-Sell” Opportunity Graph

NOTE:
In order to get the "Second Account" to track closer to the "First Account," offer the "Second Account" to customers with a Good Customer Score ranking higher than 10. This will result in the mitigation of lost revenue on over 120,000 accounts. Thus, generating an annual revenue gain of $24+ Million as a by-product of more accurately targeting “Good Customers.”

RESULTS

The “Good Customer Score” has been added to the production Data Warehouse and we are using it to develop specific targeted Retention strategies. Additionally, we are in the process of using the “Good Customer Score” to complete further tests that will definitively influence its use in analytically supported business decisions at PREMIER.

Estimated Opportunity Value:
The initial recommendation will reduce the attrition of PREMIER’s “Top **Good** Customers” >= 2 Years on Book and generate an excess of **$15+ Million** (see Figure 9) in annual revenue.

Future projects that will be engaged include:

1. More accurately identify customers qualified for specific retention strategies = **$8+ Million annually**
2. Reduce Mailing Costs by prescribing targeted mailing programs focused on the Good Customer profile of our “Top Customers” **= $12+ Million annually**
3. Improve customer programs through better targeting and providing a viable measure for offering proactive up-sell/cross-sell opportunities = **$24+ Million annually**
4. Add Net Income to PREMIER by Lowering Delinquency & Charge-Off experience in the overall Portfolio = **$10+ Million annually**
5. Eliminate 3rd party purchased score and dependant software = **$1.5+ Million annually**

CONCLUSION

By using predictive modeling with SAS Enterprise Miner coupled with Base SAS programming for Customer Intelligence portfolio segmentation, I have been able to generate huge lifts in revenue opportunity for the organization. My experience, as demonstrated by the example illustrated in the context of this paper, is a clear depiction of the benefits resulting from the use of SAS tools in business analytics, specifically “Customer Intelligence.” Venturing into the huge amounts of internal customer data and information, or any other huge set of data, can be a daunting task. However, by employing SAS Enterprise Miner, coupled with some Base SAS techniques, gold nuggets can be identified.

Contact Information

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

Rex Pruitt
PREMIER Bankcard
PO Box 5114; Mail Drop #113
3820 N. Louise Ave.
Sioux Falls, SD 57117-5114
(605) 575-9810 - Office
(605) 575-9866 - Fax
rpruitt@premierbankcard.com