```
ods pdf file="/home/u58545683/SAS project/donation.pdf";

libname pmlr '/home/u58545683/SAS project/class1';
proc contents data=pmlr.pva_raw_data;
run;

data pva(drop=Control_number);
 set pmlr.pva_raw_data;
run;

/*create a macro variable: ex_inputs, including all
independent variables needed*/

%let ex_inputs= MONTHS_SINCE_ORIGIN
DONOR_AGE IN_HOUSE INCOME_GROUP PUBLISHED_PHONE
MOR_HIT_RATE WEALTH_RATING MEDIAN_HOME_VALUE
MEDIAN_HOUSEHOLD_INCOME PCT_OWNER_OCCUPIED
PER_CAPITA_INCOME PCT_MALE_MILITARY
PCT_MALE_VETERANS PCT_VIETNAM_VETERANS
PCT_WWII_VETERANS PEP_STAR RECENT_STAR_STATUS
FREQUENCY_STATUS_97NK RECENT_RESPONSE_PROP
RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE_PROP
RECENT_AVG_CARD_GIFT_AMT RECENT_RESPONSE_COUNT
RECENT_CARD_RESPONSE_COUNT LIFETIME_CARD_PROM
LIFETIME_PROM LIFETIME_GIFT_AMOUNT
LIFETIME_GIFT_COUNT LIFETIME_AVG_GIFT_AMT
LIFETIME_GIFT_RANGE LIFETIME_MAX_GIFT_AMT
LIFETIME_MIN_GIFT_AMT LAST_GIFT_AMT
CARD_PROM_12 NUMBER_PROM_12 MONTHS_SINCE_LAST_GIFT
MONTHS_SINCE_FIRST_GIFT;
```

The CONTENTS Procedure

Data Set Name	PMLR.PVA_RAW_DATA	**Observations**	19372
Member Type	DATA	**Variables**	50
Engine	V9	**Indexes**	0
Created	2005-01-11 18:34:30	**Observation Length**	368
Last Modified	2005-01-11 18:34:30	**Deleted Observations**	0
Protection		**Compressed**	NO
Data Set Type		**Sorted**	NO
Label			
Data Representation	WINDOWS_32		
Encoding	wlatin1 Western (Windows)		

Engine/Host Dependent Information	
Data Set Page Size	16384
Number of Data Set Pages	441
First Data Page	1
Max Obs per Page	44
Obs in First Data Page	27
Number of Data Set Repairs	0
Filename	/home/u58545683/SAS project/class1/pva_raw_data.sas7bdat
Release Created	9.0101M3
Host Created	XP_PRO
Inode Number	19532338509
Access Permission	rw-r--r--
Owner Name	u58545683
File Size	7MB
File Size (bytes)	7226368

Alphabetic List of Variables and Attributes			
#	**Variable**	**Type**	**Len**
4	CARD_PROM_12	Num	8

4			
9	CLUSTER_CODE	Char	2
3	CONTROL_NUMBER	Char	8
5	DONOR_AGE	Num	8
11	DONOR_GENDER	Char	3
48	FILE_AVG_GIFT	Num	8
49	FILE_CARD_GIFT	Num	8
27	FREQUENCY_STATUS_97NK	Num	8
10	HOME_OWNER	Char	3

The CONTENTS Procedure

Alphabetic List of Variables and Attributes			
#	**Variable**	**Type**	**Len**
12	INCOME_GROUP	Num	8
6	IN_HOUSE	Num	8
43	LAST_GIFT_AMT	Num	8
39	LIFETIME_AVG_GIFT_AMT	Num	8
35	LIFETIME_CARD_PROM	Num	8
37	LIFETIME_GIFT_AMOUNT	Num	8
38	LIFETIME_GIFT_COUNT	Num	8
40	LIFETIME_GIFT_RANGE	Num	8
41	LIFETIME_MAX_GIFT_AMT	Num	8
42	LIFETIME_MIN_GIFT_AMT	Num	8
36	LIFETIME_PROM	Num	8
17	MEDIAN_HOME_VALUE	Num	8
18	MEDIAN_HOUSEHOLD_INCOME	Num	8
4	MONTHS_SINCE_FIRST_GIFT	Num	8

7			
46	MONTHS_SINCE_LAST_GIFT	Num	8
34	MONTHS_SINCE_LAST_PROM_RESP	Num	8
4	MONTHS_SINCE_ORIGIN	Num	8
15	MOR_HIT_RATE	Num	8
45	NUMBER_PROM_12	Num	8
14	OVERLAY_SOURCE	Char	1
20	PCT_MALE_MILITARY	Num	8
21	PCT_MALE_VETERANS	Num	8
19	PCT_OWNER_OCCUPIED	Num	8
22	PCT_VIETNAM_VETERANS	Num	8
23	PCT_WWII_VETERANS	Num	8
24	PEP_STAR	Num	8
50	PER_CAPITA_INCOME	Num	8
13	PUBLISHED_PHONE	Num	8
26	RECENCY_STATUS_96NK	Char	5
31	RECENT_AVG_CARD_GIFT_AMT	Num	8
29	RECENT_AVG_GIFT_AMT	Num	8
33	RECENT_CARD_RESPONSE_COUNT	Num	8
30	RECENT_CARD_RESPONSE_PROP	Num	8
32	RECENT_RESPONSE_COUNT	Num	8
28	RECENT_RESPONSE_PROP	Num	8
25	RECENT_STAR_STATUS	Num	8

The CONTENTS Procedure

Alphabetic List of Variables and Attributes			
#	**Variable**	**Type**	**Len**
8	SES	Char	4
1	TARGET_B	Num	8
2	TARGET_D	Num	8
7	URBANICITY	Char	4
16	WEALTH_RATING	Num	8

```
/* check statistical descriptive analysis for numerical and categorical
variable */

/*Check the mean, minimum, maximum, and count of
missing for each numeric input*/ proc means data=pva
n nmiss mean std min max;
 var &ex_inputs;
run;


/*check frequency for character variable: target_b*/
proc freq data=pva;
 tables _character_ target_b / missing;
run;
```

The MEANS Procedure

Variable	N	N Miss	Mean	Std Dev	Minimum	Maximum
MONTHS_SINCE_ORIGIN	19372	0	73.4099732	41.2555742	5.0000000	137.0000000
DONOR_AGE	14577	479	58.9190506	16.6693824	0	87.0000000
IN_HOUSE	19372	0	0.0731984	0.2604687	1.0000000	1.0000000
PUBLISHED_PHONE	19372	2	3.9075434	1.8647962	0	7.0000000
MOR_HIT_RATE	14980	0	0.4977287	0.5000077	0	1.0000000
WEALTH_RATING	19372	881	5.0053967	2.8153860	0	241.0000000
MEDIAN_HOME_VALUE	19372	0	1079.87	960.7534	0	9.0000000
MEDIAN_HOUSEHOLD_INCOME	19372	0	341.9702	484	0	6000.00
PCT_OWNER_OCCUPIED	19372	0	147	164.2078	0	1500.00
PER_CAPITA_INCOME	10562	0	69.6989074	21.71101	0	99.0000000
PCT_MALE_MILITARY	19372	0	15857.33	86	0	0
PCT_MALE_VETERANS	19372	0	1.0290109	8710.63	0	174523.00
PCT_VIETNAM_VETERANS	19372	0	30.573921	4.9182974	0	97.0000000
PCT_WWII_VETERANS	19372	0	1	11.421471	0	0
PEP_STAR	19372	0	29.603293	4	0	99.0000000
RECENT_STAR_STATUS	19372	0	4	15.120359	1.000000	0
FREQUENCY_STATUS_97NK	19372	0	32.852467	8	0	99.0000000
RECENT_RESPONSE_PROP	19372	0	5	17.839764	0	0
RECENT_AVG_GIFT_AMT	19372	0	0.5044394	8	0	99.0000000
RECENT_CARD_RESPONSE_PROP	19372	0	0.9311377	0.4999932	0	0
RECENT_AVG_CARD_GIFT_AMT	19372	0	1.9839975	2.5455850	0	1.0000000
RECENT_RESPONSE_COUNT	19372	0	0.1901275	1.0993458	0	22.0000000
RECENT_CARD_RESPONSE_COUNT	19372	0	15.365395	0.1139467	0	0
LIFETIME_CARD_PROM	19372	0	9	10.167484	2.0000000	4.0000000
LIFETIME_PROM	19372	0	0.2308077	9	0	1.0000000
LIFETIME_GIFT_AMOUNT	19372	0	11.685470	0.1862301	5.0000000	260.0000000
LIFETIME_GIFT_COUNT	19372	0	3	10.834120	0	000
LIFETIME_AVG_GIFT_AMT	19372	0	3.0431034	2	15.000	1.0000
LIFETIME_GIFT_RANGE	19372	0	1.7305389	2.0464006	0000	000
LIFETIME_MAX_GIFT_AMT	19372	0	18.668077	1.5355208	1.0000	300.0000
LIFETIME_MIN_GIFT_AMT	19372	0	6	8.5587782	000	000
LAST_GIFT_AMT	19372	0	47.570514	22.950158	1.360000	16.0000
CARD_PROM_12	19372	0	1	1	0	000
NUMBER_PROM_12	19372	0	104.425	105.722	0	9.0000000
MONTHS_SINCE_LAST_GIFT	19372	0	7165	4599	5.000000	56.000000
MONTHS_SINCE_FIRST_GIFT	19372	0	9.9797646	8.6881633	0	0
	19372		12.858338	8.7877579	0	194.000000
	19372		3	15.116892	0	0
	19372		11.587875	9	2.000000	377
	19372		8	16.101127	0	5.00
	19372		19.208808	8	4.000000	95.000000
	19372		1	7.9597857	0	0
	19372		7.6209323	11.977557	15.000000	450.0000
	19372		16.584198	7	0	000
	19372		8	1.2642046		997.0000
	19372		5.3671278	4.6420721		000
	19372		12.901868	4.0330648		1000.00
	19372		7	37.568169		450.0000
	19372		18.191152	3		000
	19372		2			450.0000
	19372		69.482087			000
	19372		5			17.000000
	19372					00
	19372					64.000000
	19372					0
	19372					27.000000
	19372					0
	19372					260.0000000
	19372					0
	19372					
	19372					

	19372 19372 19372					

The FREQ Procedure

URBANICITY	Frequency	Percent	Cumulative Frequency	Cumulative Percent
?	454	2.34	454	2.34
C	4022	20.76	4476	23.11
R	4005	20.67	8481	43.78
S	4491	23.18	12972	66.96
T	3944	20.36	16916	87.32
U	2456	12.68	19372	100.00

SES	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	5924	30.58	5924	30.58
2	9284	47.92	15208	78.51
3	3323	17.15	18531	95.66
4	387	2.00	18918	97.66
?	454	2.34	19372	100.00

The FREQ Procedure

CLUSTER_CODE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	454	2.34	454	2.34
01	239	1.23	693	3.58
02	380	1.96	1073	5.54
03	300	1.55	1373	7.09
04	113	0.58	1486	7.67
05	199	1.03	1685	8.70
06	123	0.63	1808	9.33

07	184	0.95	1992	10.28
08	378	1.95	2370	12.23
09	153	0.79	2523	13.02
10	387	2.00	2910	15.02
11	484	2.50	3394	17.52
12	631	3.26	4025	20.78
13	579	2.99	4604	23.77
14	454	2.34	5058	26.11
15	223	1.15	5281	27.26
16	384	1.98	5665	29.24
17	349	1.80	6014	31.04
18	619	3.20	6633	34.24
19	98	0.51	6731	34.75
20	317	1.64	7048	36.38
21	353	1.82	7401	38.20
22	251	1.30	7652	39.50
23	293	1.51	7945	41.01
24	795	4.10	8740	45.12
25	273	1.41	9013	46.53
26	202	1.04	9215	47.57
27	666	3.44	9881	51.01
28	343	1.77	10224	52.78
29	170	0.88	10394	53.65
30	519	2.68	10913	56.33
31	249	1.29	11162	57.62
32	152	0.78	11314	58.40
33	109	0.56	11423	58.97
34	284	1.47	11707	60.43
35	727	3.75	12434	64.19

The FREQ Procedure

CLUSTER_CODE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
36	716	3.70	13150	67.88

37	204	1.05	13354	68.93
38	240	1.24	13594	70.17
39	512	2.64	14106	72.82
40	830	4.28	14936	77.10
41	431	2.22	15367	79.33
42	284	1.47	15651	80.79
43	468	2.42	16119	83.21
44	383	1.98	16502	85.18
45	482	2.49	16984	87.67
46	369	1.90	17353	89.58
47	185	0.95	17538	90.53
48	180	0.93	17718	91.46
49	675	3.48	18393	94.95
50	156	0.81	18549	95.75
51	460	2.37	19009	98.13
52	60	0.31	19069	98.44
53	303	1.56	19372	100.00

HOME_OWNER	Frequency	Percent	Cumulative Frequency	Cumulative Percent
H	10606	54.75	10606	54.75
U	8766	45.25	19372	100.00

DONOR_GENDER	Frequency	Percent	Cumulative Frequency	Cumulative Percent
A	1	0.01	1	0.01
F	10401	53.69	10402	53.70
M	7953	41.05	18355	94.75
U	1017	5.25	19372	100.00

OVERLAY_SOURCE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
B	8732	45.08	8732	45.08

M	1480	7.64	10212	52.72
N	4392	22.67	14604	75.39
P	4768	24.61	19372	100.00

The FREQ Procedure

RECENCY_STATUS_96NK	Frequency	Percent	Cumulative Frequency	Cumulative Percent
A	11918	61.52	11918	61.52
E	427	2.20	12345	63.73
F	1521	7.85	13866	71.58
L	93	0.48	13959	72.06
N	1192	6.15	15151	78.21
S	4221	21.79	19372	100.00

TARGET_B	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	14529	75.00	14529	75.00
1	4843	25.00	19372	100.00

```
/*create new arrays: mi_DONOR_AGE
mi_INCOME_GROUP mi_WEALTH_RATING  use 0/1 to
represent missing and nonmissing values*/
data pva(drop=i);
 set pva;
 array mi{*} mi_DONOR_AGE mi_INCOME_GROUP
 mi_WEALTH_RATING;
 array x{*} DONOR_AGE INCOME_GROUP
WEALTH_RATING;  do i=1 to dim(mi);
 mi{i}=(x{i}=.);
 end;
run;


/*Impute missing value to a new dataset pva1
 use median of the variable to replace NAs*/
proc stdize data=pva
      method=median
 reponly
 out=pva1;
 var DONOR_AGE INCOME_GROUP WEALTH_RATING;
run;


/*Split the imputed data set into training
and test data sets.   Use 70% of the data
for each data
 set role. Stratify on the target variable.*/
proc sort data=pva1 out=pva1;
 by target_b;
run;


proc surveyselect noprint
 data=pva1
 samprate=.7
 out=pva2
 seed=27513
 outall;
 strata target_b;
run;
data pva_train pva_test;
 set pva2;
 if selected then output pva_train;
/*if select =1 */  else output
pva_test; /*else select = 0*/

run;
```

```
/* use proc freq to check whether train is 70% and 30% for test,
and response rate is the same*/ proc freq data=pva_train;
table target_b;
run;

proc freq data=pva_test;
table target_b;
run;
```

The FREQ Procedure

TARGET_B	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	10171	75.00	10171	75.00
1	3391	25.00	13562	100.00

The FREQ Procedure

TARGET_B	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	4358	75.01	4358	75.01
1	1452	24.99	5810	100.00

```
/*create macro variable:ex_screened which has all the
independent variables we need*/ %let ex_screened=
LIFETIME_CARD_PROM LIFETIME_MIN_GIFT_AMT
PER_CAPITA_INCOME mi_INCOME_GROUP
RECENT_RESPONSE_COUNT PCT_MALE_MILITARY
DONOR_AGE PCT_VIETNAM_VETERANS MOR_HIT_RATE
PCT_OWNER_OCCUPIED PCT_MALE_VETERANS PUBLISHED_PHONE
WEALTH_RATING MONTHS_SINCE_LAST_GIFT
RECENT_STAR_STATUS LIFETIME_GIFT_RANGE INCOME_GROUP
IN_HOUSE
RECENT_AVG_GIFT_AMT PCT_WWII_VETERANS
LIFETIME_GIFT_AMOUNT PEP_STAR mi_DONOR_AGE
RECENT_AVG_CARD_GIFT_AMT RECENT_CARD_RESPONSE_PROP
;

/*Use the Spearman correlation coefficients to
screen the inputs with the  least evidence of a
relationship with the target*/
proc corr data=pva_train spearman rank;
 var &ex_screened;
 with target_b;
run;
```

The CORR Procedure

1 With Variables:	TARGET_B
25 Variables:	LIFETIME_CARD_PROM LIFETIME_MIN_GIFT_AMT PER_CAPITA_INCOME mi_INCOME_GROUP RECENT_RESPONSE_COUNT PCT_MALE_MILITARY DONOR_AGE PCT_VIETNAM_VETERANS MOR_HIT_RATE PCT_OWNER_OCCUPIED PCT_MALE_VETERANS PUBLISHED_PHONE WEALTH_RATING MONTHS_SINCE_LAST_GIFT RECENT_STAR_STATUS LIFETIME_GIFT_RANGE INCOME_GROUP IN_HOUSE RECENT_AVG_GIFT_AMT PCT_WWII_VETERANS LIFETIME_GIFT_AMOUNT PEP_STAR mi_DONOR_AGE RECENT_AVG_CARD_GIFT_AMT RECENT_CARD_RESPONSE_PROP

Simple Statistics

Variable	N	Mean	Std Dev	Median	Minimum	Maximum
TARGET_B	13562	0.25004	0.43305	0	0	1.00000
LIFETIME_CARD_PROM	13562	18.62439	8.54080	18.00000	2.00000	56.00000
LIFETIME_MIN_GIFT_AMT	13562	7.64615	7.45707	5.00000	0	201.67000
PER_CAPITA_INCOME	13562	15889	8849	13791	0	174523
mi_INCOME_GROUP	13562	0.22843	0.41984	0	0	1.00000
RECENT_RESPONSE_COUNT	13562	3.04262	2.05699	3.00000	0	16.00000
PCT_MALE_MILITARY	13562	1.00391	4.81273	0	0	97.00000
DONOR_AGE	13562	59.19370	14.39887	60.00000	2.00000	87.00000
PCT_VIETNAM_VETERANS	13562	29.61886	15.08804	29.00000	0	99.00000
MOR_HIT_RATE	13562	3.34147	8.81530	0	0	241.00000
PCT_OWNER_OCCUPIED	13562	69.56673	21.88518	75.00000	0	99.00000
PCT_MALE_VETERANS	13562	30.60905	11.35025	31.00000	0	85.00000
PUBLISHED_PHONE	13562	0.50310	0.50001	1.00000	0	1.00000
WEALTH_RATING	13562	4.99771	2.07834	5.00000	0	9.00000
MONTHS_SINCE_LAST_GIFT	13562	18.21037	4.06264	18.00000	4.00000	27.00000
RECENT_STAR_STATUS	13562	0.93003	2.53571	0	0	22.00000
LIFETIME_GIFT_RANGE	13562	11.53045	15.57212	10.00000	0	997.00000
INCOME_GROUP	1356	3.93637	1.63987	4.0000	1.0000	7.00000

	2			0	0	
IN_HOUSE	13562	0.07418	0.26207	0	0	1.00000
RECENT_AVG_GIFT_AMT	13562	15.33553	9.97628	14.00000	0	250.00000
PCT_WWII_VETERANS	13562	32.84987	17.61161	32.00000	0	99.00000
LIFETIME_GIFT_AMOUNT	13562	103.83724	100.87617	79.00000	15.00000	2605
PEP_STAR	13562	0.50442	0.50000	1.00000	0	1.00000
mi_DONOR_AGE	13562	0.25277	0.43461	0	0	1.00000
RECENT_AVG_CARD_GIFT_AMT	13562	11.71288	10.65755	10.33000	0	300.00000
RECENT_CARD_RESPONSE_PROP	13562	0.23239	0.18760	0.20000	0	1.00000

The CORR Procedure

Spearman Correlation Coefficients, N = 13562
Prob > |r| under H0: Rho=0

TARGET_B	RECENT_RESPONSE_COUNT	RECENT_AVG_GIFT_AMT	PEP_STAR	LIFETIME_MIN_GIFT_AMT	RECENT_CARD_RESPONSE_PROP
	0.13342	-0.11514	0.10234	-0.09888	0.09586
	<.0001	<.0001	<.0001	<.0001	<.0001

Spearman Correlation Coefficients, N = 13562
Prob > |r| under H0: Rho=0

TARGET_B	MONTHS_SINCE_LAST_GIFT	LIFETIME_CARD_PROM	RECENT_STAR_STATUS	LIFETIME_GIFT_AMOUNT	PER_CAPITA_INCOME
	-0.08156	0.06945	0.06733	0.06033	0.04617
	<.0001	<.0001	<.0001	<.0001	<.0001

Spearman Correlation Coefficients, N = 13562
Prob > |r| under H0: Rho=0

TARGET_B	IN_HOUSE	INCOME_GROUP	PCT_MALE_VETERANS	MOR_HIT_RATE	RECENT_AVG_CARD_GIFT_AMT	LIFETIME_GIFT_RANGE
	0.03669	0.03648	0.02910	0.02748	-0.02708	-0.02571
	<.0001	<.0001	0.0007	0.0014	0.0016	0.0027

Spearman Correlation Coefficients, N = 13562
Prob > |r| under H0: Rho=0

TARGET_B	DONOR_AGE	WEALTH_RATING	PCT_VIETNAM_VETERANS	PCT_WWII_VETERANS	PCT_OWNER_OCCUPIED	mi_DONOR_AGE
	0.02 449 0.0043	0.02 233 0.0093	-0.01 747 0.0420	0.01 659 0.0534	0.01 390 0.1054	-0.01 102 0.1994

Spearman Correlation Coefficients, N = 13562 Prob > |r| under H0: Rho=0			
TARGET_B	PUBLISHED_PHONE -0.00 238 0.7813	PCT_MALE_MILITARY -0.00 223 0.7952	mi_INCOME_GROUP 0.00 056 0.9478

The LOGISTIC Procedure

Model Information	
Data Set	WORK.PVA_TRAIN
Response Variable	TARGET_B
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	13562
Number of Observations Used	13562

Response Profile		
Ordered Value	**TARGET_B**	**Total Frequency**
1	1	3391
2	0	10171

Probability modeled is TARGET_B=1.

Backward Elimination Procedure

Step 0. The following effects were entered:

Intercept LIFETIME_CARD_PROM LIFETIME_MIN_GIFT_AM PER_CAPITA_INCOME mi_INCOME_GROUP

RECENT_RESPONSE_COUN PCT_MALE_MILITARY DONOR_AGE PCT_VIETNAM_VETERANS MOR_HIT_RATE PCT_OWNER_OCCUPIED PCT_MALE_VETERANS PUBLISHED_PHONE WEALTH_RATING MONTHS_SINCE_LAST_GI RECENT_STAR_STATUS LIFETIME_GIFT_RANGE INCOME_GROUP IN_HOUSE RECENT_AVG_GIFT_AMT PCT_WWII_VETERANS LIFETIME_GIFT_AMOUNT PEP_STAR mi_DONOR_AGE RECENT_AVG_CARD_GIFT RECENT_CARD_RESPONSE

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	15255.877	14883.080
SC	15263.392	15078.471
-2 Log L	15253.877	14831.080

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	422.7971	25	<.0001
Score	425.0920	25	<.0001
Wald	408.1974	25	<.0001

```
/*Fit a logistic regression model with the FAST BACKWARD
method. Use the macro variable ex_screened to represent
all independent variables*/

proc logistic data=pva_train des;
 model target_b = &ex_screened
 /selection=backward fast;
run;
```

Summary of Backward Elimination					
Step	Effect Removed	DF	Number In	Wald Chi-Square	Pr > ChiSq
1	PCT_WWII_VETERANS	1	24	0.0000	0.9966
1	mi_INCOME_GROUP	1	23	0.0010	0.9752
1	LIFETIME_GIFT_AMOUNT	1	22	0.0085	0.9263
1	IN_HOUSE	1	21	0.0668	0.7960
1	RECENT_AVG_CARD_GIFT	1	20	0.0782	0.7797
1	MOR_HIT_RATE	1	19	0.0977	0.7546
1	PCT_MALE_MILITARY	1	18	0.1240	0.7247
1	PCT_OWNER_OCCUPIED	1	17	0.2944	0.5874
1	WEALTH_RATING	1	16	0.2829	0.5948
1	LIFETIME_GIFT_RANGE	1	15	1.0878	0.2970
1	mi_DONOR_AGE	1	14	1.1062	0.2929
1	PUBLISHED_PHONE	1	13	0.8280	0.3629
1	DONOR_AGE	1	12	1.7407	0.1871
1	LIFETIME_MIN_GIFT_AM	1	11	3.1335	0.0767
1	PCT_VIETNAM_VETERANS	1	10	3.2119	0.0731
1	RECENT_STAR_STATUS	1	9	3.8243	0.0505

```
/*Fit a logistic regression model with the FAST Stepwise
method. Use the macro variable ex_screened to represent
all independent variables*/

proc logistic data=pva_train des;
 model target_b = &ex_screened
 /selection=stepwise fast best=1;
run;
/* use backward and stepwise selection to select final variables */
```

Summary of Stepwise Selection							
Step	Effect		DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
	Entered	Removed					
1	RECENT_RESPONSE_COUN		1	1	230.9811		<.0001
2	MONTHS_SINCE_LAST_GI		1	2	54.2230		<.0001
3	PER_CAPITA_INCOME		1	3	37.5667		<.0001
4	PEP_STAR		1	4	36.0959		<.0001
5	INCOME_GROUP		1	5	18.5258		<.0001
6	RECENT_AVG_GIFT_AMT		1	6	18.1926		<.0001
7	RECENT_CARD_RESPONSE		1	7	11.4091		0.0007
8	LIFETIME_CARD_PROM		1	8	8.2996		0.0040
9	PCT_MALE_VETERANS		1	9	4.1210		0.0424
10	RECENT_STAR_STATUS		1	10	3.8433		0.0499
11		RECENT_STAR_STATUS	1	9		3.8341	0.0502

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	-1.4173	0.1453	95.1080	<.0001
LIFETIME_CARD_PROM	1	0.00932	0.00325	8.2253	0.0041
PER_CAPITA_INCOME	1	0.000010	2.3E-6	20.6604	<.0001
RECENT_RESPONSE_COUN	1	0.0509	0.0135	14.2044	0.0002
PCT_MALE_VETERANS	1	0.00362	0.00179	4.1195	0.0424

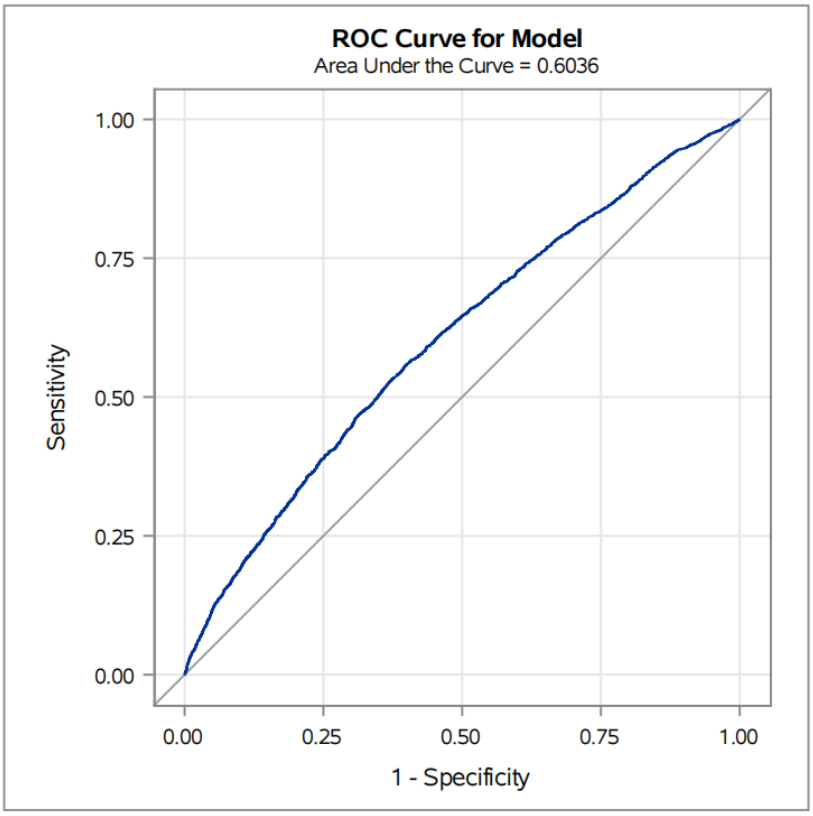
```
/*Final variables will be:
RECENT_RESPONSE_COUN: COunt of responses to promotions since June 1994
MONTHS_SINCE_LAST_GIFT: months since most recent donation
PER_CAPITA_INCOME: Census data
PEP_STAR:flag to identify consecutive donors
INCOME_GROUP: income bracket, from 1 to 7
RECENT_CARD_RESPONSE_PROP: proportion of responses to promotions
LIFETIME_CARD_PROM: Number of card promotions, ever
PCT_MALE_VETERANS: Census data

Variables with significant p-values from both selection methods are included,
while variables with insignificant p-values from either selection method is
excluded.

Check ROC curve for this model and use this model to score pva_test
Also, check average of p_1 in scored_test*/
proc logistic data=pva_train des outest=betas1 plots=ROC;
model target_b =
RECENT_RESPONSE_COUNT
MONTHS_SINCE_LAST_GIFT
PER_CAPITA_INCOME
PEP_STAR
INCOME_GROUP
RECENT_CARD_RESPONSE_PROP
LIFETIME_CARD_PROM
PCT_MALE_VETERANS;
run;
```

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	60.4	**Somers' D**	0.207
Percent Discordant	39.6	**Gamma**	0.207
Percent Tied	0.0	**Tau-a**	0.078
Pairs	34489861	**c**	0.604

The LOGISTIC Procedure



/*The result of ROC AUC is 0.6110, which means the model’s performance at distinguishing between the positive and negative cl However, the Somers'D is 0.222 which is too small. The model has a poor predictor.*/

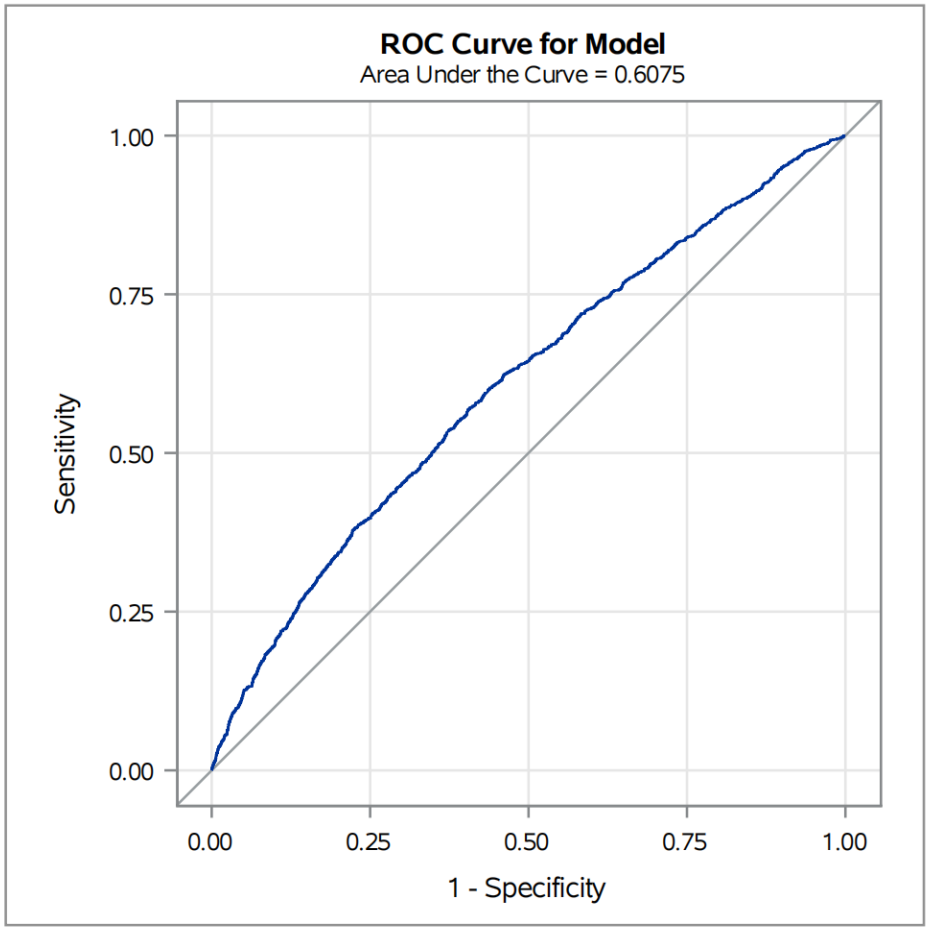
```
/*score pva_test data*/
proc logistic data=pva_train des plots=ROC;
model target_b =
MONTHS_SINCE_LAST_GIFT
INCOME_GROUP
RECENT_AVG_GIFT_AMT
PEP_STAR
RECENT_CARD_RESPONSE_PROP;
score data = pva_test out=scored_test;
run;

proc means data=scored_test;
var p_1;
run;

/* check the performance on Test dataset*/
proc logistic data=scored_test des plots=ROC;
model target_b = p_1;
run;

ods pdf close;
```

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	60.7	**Somers' D**	0.215
Percent Discordant	39.2	**Gamma**	0.215
Percent Tied	0.0	**Tau-a**	0.081
Pairs	6327816	**c**	0.608



/*Result shows that ROC AUC of both train and test are

0.6110, so there is no overfitting*/