SAS Data Set pva_raw_data

A national veterans' organization seeks to better target its solicitations for donation. By soliciting only the most likely donors, the organization can spend less money on solicitation efforts and more money on charitable concerns. Solicitations involve sending a small gift to an individual with a request for donation.

Gifts include mailing labels and greeting cards.

The organization has more than 3.5 million individuals in its mailing database. These individuals have been classified by their response behavior to previous solicitation efforts. Of particular interest is the class of individuals identified as lapsing donors. These individuals made their most recent donation between 12 and 24 months ago. The organization found that by predicting the response behavior of this group, they could use the model to rank all 3.5 million individuals in their database. With this ranking, a decision can be made to either solicit or ignore an individual in the current solicitation campaign. The current campaign refers to a greeting card mailing sent in June of 1997. It is identified in the raw data as the 97NK campaign.

The **pva_raw_data** data set is a subset of the raw analysis data (the mailing database) that consists of slightly more than 19,000 records. This subset is a biased sample that was created by selecting all of the events (individuals who donated in the last mailing campaign) and a sample of the nonevents (individuals who did not donate in the last mailing campaign). The proportion of events in the population is estimated at 0.05.

The 481 fields in the mailing database were reduced to 50 in **pmlr.pva_raw_data**. Some of the fields were eliminated when their potential association with the analysis objective was considered (for example, it is doubtful that CD player ownership is strongly correlated with donation potential). Other fields were combined to form summaries of a particular customer behavior. In general, if the variable has PROM in its name, then the variable is related to the number of items the organization sent to the customer. However, if the variable has GIFT in its name, then the variable is related to the amount of money the customer sent to the organization. The remaining input variables include demographics of the individual, the number of mailings the organization sent to the customer, and the amount of money the individual sent to the organization.

Description

#	**Variable**	**Type**	**Len**	**Description**
44	CARD_PROM_12	Num	8	Count of card promotions in the last 12 months
9	CLUSTER_CODE	Char	2	Socio-Economic Cluster Code
3	CONTROL_NUMBER	Char	8	Control number
5	DONOR_AGE	Num	8	Donor Age
11	DONOR_GENDER	Char	3	Donor Gender
48	FILE_AVG_GIFT	Num	8	[No description available]
49	FILE_CARD_GIFT	Num	8	[No description available]
27	FREQUENCY_STATUS_97NK	Num	8	Count of Donations between June 1995 and June 1996 (capped at 4)
10	HOME_OWNER	Char	3	Home Owner flag
12	INCOME_GROUP	Num	8	Income Bracket, from 1 to 7
6	IN_HOUSE	Num	8	Flag for *In-House* donor program
43	LAST_GIFT_AMT	Num	8	Amount of most recent donation
39	LIFETIME_AVG_GIFT_AMT	Num	8	Average donation amount, ever
35	LIFETIME_CARD_PROM	Num	8	Number of card promotions, ever
37	LIFETIME_GIFT_AMOUNT	Num	8	Total donation amount, ever
38	LIFETIME_GIFT_COUNT	Num	8	Total number of donations, ever
40	LIFETIME_GIFT_RANGE	Num	8	Maximum gift amount less minimum gift amount
41	LIFETIME_MAX_GIFT_AMT	Num	8	Maximum gift amount, ever
42	LIFETIME_MIN_GIFT_AMT	Num	8	Minimum gift amount, ever
36	LIFETIME_PROM	Num	8	Count of solicitations ever sent
17	MEDIAN_HOME_VALUE	Num	8	Census data
18	MEDIAN_HOUSEHOLD_INCOME	Num	8	Census data
47	MONTHS_SINCE_FIRST_GIFT	Num	8	Months since first donation
46	MONTHS_SINCE_LAST_GIFT	Num	8	Months since most recent donation
34	MONTHS_SINCE_LAST_PROM_RESP	Num	8	Months since last promotion response
4	MONTHS_SINCE_ORIGIN	Num	8	Months since entry onto the file
15	MOR_HIT_RATE	Num	8	Data recorded by a third party-Mail Order Response rate
45	NUMBER_PROM_12	Num	8	Count of promotions in the last 12 months
14	OVERLAY_SOURCE	Char	1	Source of Demographic overlay

#	Variable	Type	Len	Description
	Alphabetic List of Variables and Attributes			
20	PCT_MALE_MILITARY	Num	8	Census data
21	PCT_MALE_VETERANS	Num	8	Census data
19	PCT_OWNER_OCCUPIED	Num	8	Census data
22	PCT_VIETNAM_VETERANS	Num	8	Census data
23	PCT_WWII_VETERANS	Num	8	Census data
24	PEP_STAR	Num	8	Flag to identify consecutive donors
50	PER_CAPITA_INCOME	Num	8	Census data
13	PUBLISHED_PHONE	Num	8	Flag
26	RECENCY_STATUS_96NK	Char	5	Categorization of donation patterns
31	RECENT_AVG_CARD_GIFT_AMT	Num	8	Average donation amount in response to card promotions since June 1994
29	RECENT_AVG_GIFT_AMT	Num	8	Average donation amount in response to promotions since June 1994
33	RECENT_CARD_RESPONSE_COUNT	Num	8	Count of responses to card promotions since June 1994
30	RECENT_CARD_RESPONSE_PROP	Num	8	Proportion of responses to card promotions since June 1994
32	RECENT_RESPONSE_COUNT	Num	8	Count of responses to promotions since June 1994
28	RECENT_RESPONSE_PROP	Num	8	Proportion of responses to promotions since June 1994
25	RECENT_STAR_STATUS	Num	8	STAR status flag, since June 1994
8	SES	Char	4	A clustering of the levels of **CLUSTER_CODE**
1	TARGET_B	Num	8	B=Binary, flag for response to 97NK—Target Variable
2	TARGET_D	Num	8	Dollar amount of response to 97NK
7	URBANICITY	Char	4	Categorization of residency
16	WEALTH_RATING	Num	8	Measures wealth relative to others within state

Sample Data

Obs	TARGET_B	TARGET_D	CONTROL_NUMBER	MONTHS_SINCE_ORIGIN	DONOR_AGE	IN_HOUSE	URBANICITY	SES	CLUSTER_C(
1	0	.	00000005	101	87	0	?	?	.
2	1	10.00	00000012	137	79	0	R	2	45
3	0	.	00000037	113	75	0	S	1	11
4	0	.	00000038	92	.	0	U	2	04
5	0	.	00000041	101	74	0	R	2	49
6	0	.	00000052	101	63	0	U	3	08
7	0	.	00000053	89	71	0	R	3	50
8	1	5.00	00000067	89	79	0	C	2	28
9	0	.	00000070	89	41	0	C	3	30
10	1	16.00	00000071	101	63	0	R	2	43