

=====

EPSILON CONFIDENTIAL EPSILON CONFIDENTIAL EPSILON CONFIDENTIAL

INFORMATION LISTED BELOW IS AVAILABLE UNDER THE TERMS OF THE
CONFIDENTIALITY AGREEMENT

=====

PARALYZED VETERANS OF AMERICA (PVA)
DATA DICTIONARY TO ACCOMPANY

KDD-CUP-98

The Second International Knowledge Discovery and
Data Mining Tools Competition

Held in Conjunction with KDD-98

The Fourth International Conference on Knowledge
Discovery and Data Mining

Sponsored by the

American Association for Artificial Intelligence (AAAI)
Epsilon Data Mining Laboratory
Paralyzed Veterans of America (PVA)

Created: 7/20/98
Last update: 7/20/98
Edited by Dean Abbott: 10/1/14
file name: cup98DIC edited.txt

Variable	Description
STATE	State abbreviation (a nominal/symbolic field)
ZIP	Zipcode (a nominal/symbolic field)
MAILCODE	Mail Code " " = Address is OK B = Bad Address
PVASTATE	EPVA State or PVA State Indicates whether the donor lives in a state served by the organization's EPVA chapter P = PVA State E = EPVA State (Northeastern US)
DOB	Date of birth (YYMM, Year/Month format.)
NOEXCH	Do Not Exchange Flag (For list rental) _ = can be exchanged X = do not exchange
RECINHSE	In House File Flag _ = Not an In House Record X = Donor has given to PVA's In House program
RECP3	P3 File Flag _ = Not a P3 Record X = Donor has given to PVA's P3 program
RECPGVG	Planned Giving File Flag _ = Not a Planned Giving Record X = Planned Giving Record
RECSWEEP	Sweepstakes file flag

_ = Not a Sweepstakes Record
X = Sweepstakes Record

MDMAUD

The Major Donor Matrix code
The codes describe frequency and amount of giving for donors who have given a \$100+ gift at any time in their giving history.
An RFA (recency/frequency/monetary) field.

The (current) concatenated version is a nominal or symbolic field. The individual bytes could separately be used as fields and refer to the following:

First byte: Recency of Giving

C=Current Donor

L=Lapsed Donor

I=Inactive Donor

D=Dormant Donor

2nd byte: Frequency of Giving

1=One gift in the period of recency

2=Two-Four gifts in the period of recency

5=Five+ gifts in the period of recency

3rd byte: Amount of Giving

L=Less than \$100(Low Dollar)

C=\$100-499(Core)

M=\$500-999(Major)

T=\$1,000+(Top)

4th byte: Blank/meaningless/filler

'X' indicates that the donor is not a major donor.

For more information regarding the RFA codes, see the promotion history field definitions.

DOMAIN

DOMAIN/Cluster code. A nominal or symbolic field. could be broken down by bytes as explained below.

1st byte = Urbanicity level of the donor's neighborhood

U=Urban

C=City

S=Suburban

T=Town

R=Rural

2nd byte = Socio-Economic status of the neighborhood

1 = Highest SES

2 = Average SES

3 = Lowest SES (except for Urban communities, where

1 = Highest SES, 2 = Above average SES,

3 = Below average SES, 4 = Lowest SES.)

CLUSTER

CLUSTER

Code indicating which cluster group the donor falls into. Each cluster is unique in terms of socio-economic status, urbanicity, ethnicity and a variety of other demographic characteristics. A nominal or symbolic field.

AGE

Overlay Age

0 = missing

AGEFLAG

Age Flag

E = Exact

I = Inferred from Date of Birth Field

HOMEOWNR	Home Owner Flag H = Home owner U = Unknown
CHILD03	Presence of Children age 0-3 B = Both, F = Female, M = Male
CHILD07	Presence of Children age 4-7
CHILD12	Presence of Children age 8-12
CHILD18	Presence of Children age 13-18
NUMCHLD	NUMBER OF CHILDREN
INCOME	HOUSEHOLD INCOME
GENDER	Gender M = Male F = Female U = Unknown J = Joint Account, unknown gender
WEALTH1	Wealth Rating
HIT	MOR Flag # HIT (Mail Order Response) Indicates total number of known times the donor has responded to a mail order offer other than PVA's.

	The following variables indicate the number of known times the donor has responded to other types of mail order offers.
MBCRAFT	Buy Craft Hobby
MBGARDEN	Buy Gardening
MBBOOKS	Buy Books
MBCOLECT	Buy Collectables
MAGFAML	Buy General Family Mags
MAGFEM	Buy Female Mags
MAGMALE	Buy Sports Mags
PUBGARDN	Gardening Pubs
PUBCULIN	Culinary Pubs
PUBHLTH	Health Pubs
PUBDOITY	Do It Yourself Pubs
PUBNEWFN	News / Finance Pubs
PUBPHOTO	Photography Pubs
PUBOPP	Opportunity Seekers Pubs

DATASRCE	Source of Overlay Data Indicates which third-party data source the donor matched against 1 = MetroMail 2 = Polk 3 = Both
MALEMILI	% Males active in the Military
MALEVET	% Males Veterans
VIETVETS	% Vietnam Vets
WWIIVETS	% WWII Vets
LOCALGOV	% Employed by Local Gov
STATEGOV	% Employed by State Gov
FEDGOV	% Employed by Fed Gov
SOLP3	SOLICIT LIMITATION CODE P3 = can be mailed (Default) 00 = Do Not Solicit or Mail 01 = one solicitation per year 02 = two solicitations per year

03 = three solicitations per year
04 = four solicitations per year
05 = five solicitations per year
06 = six solicitations per year
12 = twelve solicitations per year

SOLIH SOLICITATION LIMIT CODE IN HOUSE
= can be mailed (Default)
00 = Do Not Solicit
01 = one solicitation per year
02 = two solicitations per year
03 = three solicitations per year
04 = four solicitations per year
05 = five solicitations per year
06 = six solicitations per year
12 = twelve solicitations per year

MAJOR Major (\$\$) Donor Flag
_ = Not a Major Donor
X = Major Donor

WEALTH2 Wealth Rating
Wealth rating uses median family income and population statistics from each area to index relative wealth within each state. The segments are denoted 0-9, with 9 being the highest income group and zero being the lowest. Each rating has a different meaning within each state.

GEOCODE Geo Cluster Code indicating the level geography at which a record matches the census data.
A nominal or symbolic field.
Blank=No code has been assigned or did not match at any level.

The following variables reflect donor interests, as collected from third-party data sources

COLLECT1	COLLECTABLE (Y/N)
VETERANS	VETERANS (Y/N)
BIBLE	BIBLE READING (Y/N)
CATLG	SHOP BY CATALOG (Y/N)
HOME	WORK FROM HOME (Y/N)
PETS	HOUSEHOLD PETS (Y/N)
CDPLAY	CD PLAYER OWNERS (Y/N)
STEREO	STEREO/RECORDS/TAPES/CD (Y/N)
PCOWNERS	HOME PC OWNERS/USERS
PHOTO	PHOTOGRAPHY (Y/N)
CRAFTS	CRAFTS (Y/N)
FISHER	FISHING (Y/N)
GARDENIN	GARDENING (Y/N)
BOATS	POWER BOATING (Y/N)
WALKER	WALK FOR HEALTH (Y/N)
KIDSTUFF	BUYS CHILDREN'S PRODUCTS (Y/N)
CARDS	STATIONARY/CARDS BUYER (Y/N)
PLATES	PLATE COLLECTOR (Y/N)

LIFESRC LIFE STYLE DATA SOURCE
Indicates source of the lifestyle variables listed above
1 = MATCHED ON METRO MAIL ONLY
2 = MATCHED ON POLK ONLY
3 = MATCHED BOTH MM AND POLK

PEPSTRFL

Indicates PEP Star RFA Status
blank = Not considered to be a PEP Star
'X' = Has PEP Star RFA Status

The fields listed below are from the promotion history file.

PROMOTION CODES:

The following lists the promotion codes and their respective field names (where XXXX refers to ADATE, RFA, RDATE and RAMNT.)

'97NK' ==> xxxx_2 (mailing was used to construct
the target fields)
'96NK' ==> xxxx_3
'96TK' ==> xxxx_4
'96SK' ==> xxxx_5
'96LL' ==> xxxx_6
'96G1' ==> xxxx_7
'96GK' ==> xxxx_8
'96CC' ==> xxxx_9
'96WL' ==> xxxx_10
'96X1' ==> xxxx_11
'96XK' ==> xxxx_12
'95FS' ==> xxxx_13
'95NK' ==> xxxx_14
'95TK' ==> xxxx_15
'95LL' ==> xxxx_16
'95G1' ==> xxxx_17
'95GK' ==> xxxx_18
'95CC' ==> xxxx_19
'95WL' ==> xxxx_20
'95X1' ==> xxxx_21
'95XK' ==> xxxx_22
'94FS' ==> xxxx_23
'94NK' ==> xxxx_24

1st 2 bytes of the code refers to the year of the mailing while 3rd and 4th bytes refer to the following promotion codes/types:

LL mailings had labels only
WL mailings had labels only
CC mailings are calendars with stickers but do not have labels
FS mailings are blank cards that fold into thirds with labels
NK mailings are blank cards with labels
SK mailings are blank cards with labels
TK mailings have thank you printed on the outside with labels
GK mailings are general greeting cards (an assortment of birthday, sympathy, blank, & get well) with labels
XK mailings are Christmas cards with labels
X1 mailings have labels and a notepad
G1 mailings have labels and a notepad

This information could certainly be used to calculate several summary variables that count the number of occurrences of various types of promotions received in the most recent 12-36 months, etc.

RFA (RECENCY/FREQUENCY/AMOUNT)

The RFA (recency/frequency/amount) status of the donors (as of the promotion dates) is included in the RFA fields.

The (current) concatenated version is a nominal or symbolic field. The individual bytes could separately be used as fields and refer to the following:

First Byte of code is concerned with RECENCY based on Date of the last Gift

F=FIRST TIME DONOR Anyone who has made their first donation in the last 6 months and has made just one donation.

N=NEW DONOR Anyone who has made their first donation in the last 12 months and is not a First time donor. This is everyone who made their first donation 7-12 months ago, or people who made their first donation between 0-6 months ago and have made 2 or more donations.

A=ACTIVE DONOR Anyone who made their first donation more than 12 months ago and has made a donation in the last 12 months.

L=LAPSING DONOR A previous donor who made their last donation between 13-24 months ago.

I=INACTIVE DONOR A previous donor who has not made a donation in the last 24 months. It is people who made a donation 25+ months ago.

S=STAR DONOR STAR Donors are individuals who have given to 3 consecutive card mailings.

Second Byte of code is concerned with FREQUENCY based on the period of recency. The period of recency for all groups except L and I is the last 12 months. For L it is 13-24 months ago, and for I it is 25-36 months ago. There are four valid frequency codes.

1=One gift in the period of recency
2=Two gift in the period of recency
3=Three gifts in the period of recency
4=Four or more gifts in the period of recency

Third byte of the code is the Amount of the last gift.

A=\$0.01 - \$1.99
B=\$2.00 - \$2.99
C=\$3.00 - \$4.99
D=\$5.00 - \$9.99
E=\$10.00 - \$14.99
F=\$15.00 - \$24.99
G=\$25.00 and above

RFA_2	Donor's RFA status as of 97NK promotion date
RFA_3	Donor's RFA status as of 96NK promotion date
RFA_4	Donor's RFA status as of 96TK promotion date
RFA_5	Donor's RFA status as of 96SK promotion date
RFA_6	Donor's RFA status as of 96LL promotion date
RFA_7	Donor's RFA status as of 96G1 promotion date
RFA_8	Donor's RFA status as of 96GK promotion date
RFA_9	Donor's RFA status as of 96CC promotion date
RFA_10	Donor's RFA status as of 96WL promotion date
RFA_11	Donor's RFA status as of 96X1 promotion date
RFA_12	Donor's RFA status as of 96XK promotion date
RFA_13	Donor's RFA status as of 95FS promotion date
RFA_14	Donor's RFA status as of 95NK promotion date
RFA_15	Donor's RFA status as of 95TK promotion date
RFA_16	Donor's RFA status as of 95LL promotion date
RFA_17	Donor's RFA status as of 95G1 promotion date
RFA_18	Donor's RFA status as of 95GK promotion date
RFA_19	Donor's RFA status as of 95CC promotion date
RFA_20	Donor's RFA status as of 95WL promotion date
RFA_21	Donor's RFA status as of 95X1 promotion date
RFA_22	Donor's RFA status as of 95XK promotion date
RFA_23	Donor's RFA status as of 94FS promotion date
RFA_24	Donor's RFA status as of 94NK promotion date

The following fields are summary variables from the promotion history file.

CARDPROM	Lifetime number of card promotions received to date. Card promotions are promotion type FS, GK, TK, SK, NK, XK, UF, UU.
MAXADATE	Date of the most recent promotion received (in YYYY, Year/Month format)
NUMPROM	Lifetime number of promotions received to date
CARDPM12	Number of card promotions received in the last 12 months (in terms of calendar months translates into 9603-9702)
NUMPRM12	Number of promotions received in the last 12 months (in terms of calendar months translates into 9603-9702)

The following fields are from the giving history file.

RDATE_3	Date the gift was received for 96NK
RDATE_4	Date the gift was received for 96TK
RDATE_5	Date the gift was received for 96SK
RDATE_6	Date the gift was received for 96LL
RDATE_7	Date the gift was received for 96G1
RDATE_8	Date the gift was received for 96GK
RDATE_9	Date the gift was received for 96CC
RDATE_10	Date the gift was received for 96WL
RDATE_11	Date the gift was received for 96X1
RDATE_12	Date the gift was received for 96XK
RDATE_13	Date the gift was received for 95FS
RDATE_14	Date the gift was received for 95NK
RDATE_15	Date the gift was received for 95TK
RDATE_16	Date the gift was received for 95LL
RDATE_17	Date the gift was received for 95G1
RDATE_18	Date the gift was received for 95GK
RDATE_19	Date the gift was received for 95CC
RDATE_20	Date the gift was received for 95WL
RDATE_21	Date the gift was received for 95X1
RDATE_22	Date the gift was received for 95XK
RDATE_23	Date the gift was received for 94FS

RDATE_24 Date the gift was received for 94NK

RAMNT_3	Dollar amount of the gift for 96NK
RAMNT_4	Dollar amount of the gift for 96TK
RAMNT_5	Dollar amount of the gift for 96SK
RAMNT_6	Dollar amount of the gift for 96LL
RAMNT_7	Dollar amount of the gift for 96G1
RAMNT_8	Dollar amount of the gift for 96GK
RAMNT_9	Dollar amount of the gift for 96CC
RAMNT_10	Dollar amount of the gift for 96WL
RAMNT_11	Dollar amount of the gift for 96X1
RAMNT_12	Dollar amount of the gift for 96XK
RAMNT_13	Dollar amount of the gift for 95FS
RAMNT_14	Dollar amount of the gift for 95NK
RAMNT_15	Dollar amount of the gift for 95TK
RAMNT_16	Dollar amount of the gift for 95LL
RAMNT_17	Dollar amount of the gift for 95G1
RAMNT_18	Dollar amount of the gift for 95GK
RAMNT_19	Dollar amount of the gift for 95CC
RAMNT_20	Dollar amount of the gift for 95WL
RAMNT_21	Dollar amount of the gift for 95X1
RAMNT_22	Dollar amount of the gift for 95XK
RAMNT_23	Dollar amount of the gift for 94FS
RAMNT_24	Dollar amount of the gift for 94NK

The following fields are summary variables from
the giving history file.

RAMNTALL	Dollar amount of lifetime gifts to date
NGIFTALL	Number of lifetime gifts to date
CARDGIFT	Number of lifetime gifts to card promotions to date
MINRAMNT	Dollar amount of smallest gift to date
MINRDATE	Date associated with the smallest gift to date
MAXRAMNT	Dollar amount of largest gift to date
MAXRDATE	Date associated with the largest gift to date
LASTGIFT	Dollar amount of most recent gift
LASTDATE	Date associated with the most recent gift
FIRSTDATE	Date of first gift
NEXTDATE	Date of second gift
TIMELAG	Number of months between first and second gift
AVGGIFT	Average dollar amount of gifts to date

CONTROLN Control number (unique record identifier)

TARGET_B Target Variable: Binary Indicator for Response to
97NK Mailing

TARGET_D Target Variable: Donation Amount (in \$) associated
with the Response to 97NK Mailing

HPHONE_D Indicator for presence of a published home phone #

(See the section on RFA for meaning of the codes)

RFA_2R	Recency code for RFA_2
RFA_2F	Frequency code for RFA_2
RFA_2A	Donation Amount code for RFA_2
MDMAUD_R	Recency code for MDMAUD
MDMAUD_F	Frequency code for MDMAUD
MDMAUD_A	Donation Amount code for MDMAUD

CLUSTER2 Classic Cluster Code (a nominal symbolic field)
GEOCODE2 County Size Code