

## Insurance Data Set Description

A bank seeks to increase sales of a variable annuity product. To do this, the bank will send product offers to existing banking customers. To maximize profits, however, the bank wants to be selective about whom it targets. This selectivity will be achieved by constructing a predictive model.

To achieve their analytic objective, an analysis data set was assembled. The data set contains 10,619 records and 48 variables, assembled from several source tables within the bank's data warehouse. The source tables include the customer master table, the transaction detail table, the product detail table, and a third party demographic overlay table. The variables describe each customer's demographics and usage of other banking products prior to acquisition of the variable annuity. Two of the variables are nominally scaled; the remainder are binary or interval. A summary of the interval and binary inputs in the analysis data set is provided below.

Variable	Label	N	Number Missing
ACCTAGE	Age of Oldest Account	9941	678
DDA	Checking Account	10619	0
DDABAL	Checking Balance	10619	0
DEP	Checking Deposits	10619	0
DEPAMT	Amount Deposited	10619	0
CASHBK	Number Cash Back	10619	0
CHECKS	Number of Checks	10619	0
DIRDEP	Direct Deposit	10619	0
NSF	Number Insufficient Fund	10619	0
NSFAMT	Amount NSF	10619	0
PHONE	Number Telephone Banking	9286	1333
TELLER	Teller Visits	10619	0
SAV	Saving Account	10619	0
SAVBAL	Saving Balance	10619	0
ATM	ATM	10619	0
ATMAMT	ATM Withdrawal Amount	10619	0
POS	Number Point of Sale	9286	1333
POSAMT	Amount Point of Sale	9286	1333
CD	Certificate of Deposit	10619	0
CDBAL	CD Balance	10619	0

IRA	Retirement Account	10619	0
IRABAL	IRA Balance	10619	0
LOC	Line of Credit	10619	0
LOCBAL	Line of Credit Balance	10619	0
INV	Investment	9286	1333
INVBAL	Investment Balance	9286	1333
ILS	Installment Loan	10619	0
ILSBAL	Loan Balance	10619	0
MM	Money Market	10619	0
MMBAL	Money Market Balance	10619	0
MMCRED	Money Market Credits	10619	0
MTG	Mortgage	10619	0
MTGBAL	Mortgage Balance	10619	0
CC	Credit Card	9286	1333
CCBAL	Credit Card Balance	9286	1333
CCPURC	Credit Card Purchases	9286	1333
SDB	Safety Deposit Box	10619	0
INCOME	Income	8683	1936
HMOWN	Owens Home	8774	1845
LORES	Length of Residence	8683	1936
HMVAl	Home Value	8683	1936
AGE	Age	8478	2141
CRSCORE	Credit Score	10373	246
MOVED	Recent Address Change	10619	0
INAREA	Local Address	10619	0

About half of the variables have some missing values. Many of the variables, especially those relating to monetary amounts, have an extremely large range and highly skewed distribution.

A summary of the nominal variables and the target variable (INS) is provided by the FREQ procedure.

Insurance Product					
INS	Frequency	Percent	Cumulative Frequency	Cumulative Percent	
0	6959	65.53	6959	65.53	
1	3660	34.47	10619	100.00	

Branch of Bank				
BRANCH	Frequency	Percent	Cumulative Frequency	Cumulative Percent
B1	922	8.68	922	8.68
B10	98	0.92	1020	9.61
B11	74	0.70	1094	10.30
B12	178	1.68	1272	11.98
B13	184	1.73	1456	13.71
B14	336	3.16	1792	16.88
B15	708	6.67	2500	23.54
B16	494	4.65	2994	28.19
B17	259	2.44	3253	30.63
B18	196	1.85	3449	32.48
B19	93	0.88	3542	33.36
B2	1744	16.42	5286	49.78
B3	920	8.66	6206	58.44
B4	1876	17.67	8082	76.11
B5	932	8.78	9014	84.89
B6	480	4.52	9494	89.41
B7	476	4.48	9970	93.89
B8	461	4.34	10431	98.23
B9	188	1.77	10619	100.00

  

Area Classification				
RES	Frequency	Percent	Cumulative Frequency	Cumulative Percent
R	2672	25.16	2672	25.16
S	3753	35.34	6425	60.50
U	4194	39.50	10619	100.00

The BRANCH variable, a nominal input with 19 distinct levels, indicates the branch in which the customer's initial account was opened. The RES variable, a nominal input with three distinct levels, classifies the customer's primary residence as rural, suburban, and urban.

The target variable for this analysis, INS, indicates acquisition of the variable annuity over a fixed period of time. While overall acquisition rate is about 2%, the acquisition rate in the raw analysis data is more than 34%. This reflects the separate sampling used to generate the raw data.

The bank expects to realize an average short-term revenue of about \$100 from each customer who purchase the annuity product. It is expected to cost the bank about \$4 per solicitation (which involves an initial mail solicitation with a telephone follow-up) to carry out the campaign.