Data Sets Used in Chapter Examples

Data Set ABSTRACT

The variables in this data set are as follows. There are 1,238 records (rows) in this data set.

Variable Name	Type	Description
Title	Text	Title of abstract
Text	Text	Abstract text

Data Set ASSOCS

This data set contains 7,007 records.

Variable Name	Type	Description
Customer	Numeric	Customer ID number
Time	Numeric	Numeric sequence of time in discrete numbers
Product	Text	Product description

Data Set BUYTEST

The variables in this data set are as follows. There are 10,000 records (rows) in this data set.

Variable Name	Type	Description
Age	Numeric	Age in years
Income	Numeric	Yearly income in thousands of dollars
Married	Numeric	(binary) 1 if married, 0 otherwise
Sex	Category	(binary) M, F
Coa6	Numeric	(binary) 1 if change of address in last 6 months, 0 otherwise
Ownhome	Numeric	(binary) 1 if own home, 0 otherwise
Loc	Category	Location of residence code: A-H
Climate	Category	Climate code for residence, 10, 20, and 30
Buy6	Numeric	Number of purchases in last 6 months
Buy12	Numeric	Number of purchases in last 12 months
Buy18	Numeric	Number of purchases in last 18 months
Value24	Numeric	Total value of purchases in past 24 months
Fico	Numeric	Credit score

Orgsrc	Category	Original customer source code: (C, D, I, O, P, R, U)
Discbuy	Numeric	(binary) 1 if a discount buyer, 0 otherwise
Return24	Numeric	(binary) 1 if product was returned in past 24 months, 0 otherwise
Respond	Numeric	(binary) 1 if responder to test mailing, 0 otherwise
Purchtot	Numeric	Test mailing purchase total
C1 – C7	Numeric	Test mailing total by product category
ID	Category	Unique ID number for each customer

Data Set CUSTOMERS

The two data sets used in the example are CUST_NEW and CUST_NEWSCORE. Both of these data sets are direct derivatives of the CUSTOMERS data set with no changes in the fields.

Variable Name	Type	Description
Channel	Numeric	Channel purchase preference (0=none, 1=indirect, 2=direct, 3=both)
City	Category	City name of customer site
Corp_Rev	Numeric	Corporate revenue generated in last fiscal year incorporated
Customer	Category	(A=acquisition, R=Returning (continued purchasing), C=Churn (no purchase)
Cust_flag	Category	Empty
Cust_id	Category	Customer ID number; unique identifier for each customer
Est_spend	Numeric	Estimated Product-service spend in $ for the next full year
Loc_employee	Numeric	Number of local employees at the customer site
Prod_A	Numeric	Quantity of Product A purchased
Prod_A_Opt	Numeric	Quantity of Product A options purchased
Prod_B	Numeric	Quantity of Product B purchased
Prod_C	Numeric	Quantity of Product C purchased
Prod_D	Numeric	Quantity of Product D purchased
Prod_E	Numeric	Quantity of Product E purchased
Prod_F	Numeric	Quantity of Product F purchased
Prod_G	Numeric	Quantity of Product G purchased
Prod_H	Numeric	Quantity of Product H purchased
Prod_I	Numeric	Quantity of Product I purchased
Prod_I_Opt	Numeric	Quantity of Product I options purchased
Prod_J	Numeric	Quantity of Product J purchased
Prod_J_Opt	Numeric	Quantity of Product J options purchased
Prod_K	Numeric	Quantity of Product K purchased
Prod_L	Numeric	Quantity of Product L purchased
Prod_L_Opt	Numeric	Quantity of Product L options purchased
Prod_M	Numeric	Quantity of Product M purchased

Prod_N	Numeric	Quantity of Product N purchased
Prod_O	Numeric	Quantity of Product O purchased
Prod_O_Opt	Numeric	Quantity of Product O options purchased
Prod_P	Numeric	Quantity of Product P purchased
Prod_Q	Numeric	Quantity of Product Q purchased
Public_sector	Numeric	0 = No, 1 = Yes-public sector
Purchfst	Numeric	first year customer purchased
Purchlst	Numeric	Last year customer purchased
Rev_class	Cateogry	Revenue Class code; A-H (see the following explanation)
Rev_lastyr	Numeric	Last fiscal year's net operating revenues
Rev_thisyr	Numeric	This fiscal year's net operating revunes
RFM	Cateogry	Recency, Frequency, and Monetary value cell code
Seg	Category	Industry Segment code
State	Category	State where customer site is located
Tot_Revenue	Numeric	Total net operating revenues from customer all years
US_region	Category	U.S. Region of customer location
Yrs_purchase	Numeric	Number of years customer has purchased in the past

Variable Name	Type	Description
Channel	Numeric	Channel purchase preference (0=none, 1=indirect, 2=direct, 3=both)
City	Category	City name of customer site
Corp_Rev	Numeric	Corporate revenue generated in last fiscal year incorporated
Customer	Category	(A=acquisition, R=Returning (continued purchasing), C=Churn (no purchase)
Cust_flag	Category	Empty
Cust_id	Category	Customer ID number; unique identifier for each customer
Est_spend	Numeric	Estimated Product-service spend in $ for the next full year
Loc_employee	Numeric	Number of local employees at the customer site
Prod_A	Numeric	Quantity of Product A purchased
Prod_A_Opt	Numeric	Quantity of Product A options purchased
Prod_B	Numeric	Quantity of Product B purchased
Prod_C	Numeric	Quantity of Product C purchased
Prod_D	Numeric	Quantity of Product D purchased
Prod_E	Numeric	Quantity of Product E purchased
Prod_F	Numeric	Quantity of Product F purchased
Prod_G	Numeric	Quantity of Product G purchased
Prod_H	Numeric	Quantity of Product H purchased
Prod_I	Numeric	Quantity of Product I purchased
Prod_I_Opt	Numeric	Quantity of Product I options purchased
Prod_J	Numeric	Quantity of Product J purchased
Prod_J_Opt	Numeric	Quantity of Product J options purchased

Prod_K	Numeric	Quantity of Product K purchased
Prod_L	Numeric	Quantity of Product L purchased
Prod_L_Opt	Numeric	Quantity of Product L options purchased
Prod_M	Numeric	Quantity of Product M purchased
Prod_N	Numeric	Quantity of Product N purchased
Prod_O	Numeric	Quantity of Product O purchased
Prod_O_Opt	Numeric	Quantity of Product O options purchased
Prod_P	Numeric	Quantity of Product P purchased
Prod_Q	Numeric	Quantity of Product Q purchased
Public_sector	Numeric	0 = No, 1 = Yes-public sector
Purchfst	Numeric	first year customer purchased
Purchlst	Numeric	Last year customer purchased
Rev_class	Cateogry	Revenue Class code; A-H (see the following explanation)
Rev_lastyr	Numeric	Last fiscal year's net operating revenues
Rev_thisyr	Numeric	This fiscal year's net operating revunes
RFM	Cateogry	Recency, Frequency, and Monetary value cell code
Seg	Category	Industry Segment code
State	Category	State where customer site is located
Tot_Revenue	Numeric	Total net operating revenues from customer all years
US_region	Category	U.S. Region of customer location
Yrs_purchase	Numeric	Number of years customer has purchased in the past

Data Set CUSTOMER_ACCOUNT_TRANS

This data set is similar to the CUSTOMERS data set, however, it has additional transactional data appended. It contains 982 rows.

Variable Name	Type	Description
Channel	Numeric	Channel purchase preference (0=none, 1=indirect, 2=direct, 3=both)
City	Category	City name of customer site
Corp_Rev	Numeric	Corporate revenue generated in last fiscal year incorporated
Customer	Category	(A=acquisition, R=Returning (continued purchasing), C=Churn (no purchase)
Cust_flag	Category	Empty
Cust_id	Category	Customer ID number; unique identifier for each customer
Est_spend	Numeric	Estimated Product-service spend in $ for the next full year
Loc_employee	Numeric	Number of local employees at the customer site
Prod_A	Numeric	Quantity of Product A purchased
Prod_A_Opt	Numeric	Quantity of Product A options purchased
Prod_B	Numeric	Quantity of Product B purchased

Prod_C	Numeric	Quantity of Product C purchased
Prod_D	Numeric	Quantity of Product D purchased
Prod_E	Numeric	Quantity of Product E purchased
Prod_F	Numeric	Quantity of Product F purchased
Prod_G	Numeric	Quantity of Product G purchased
Prod_H	Numeric	Quantity of Product H purchased
Prod_I	Numeric	Quantity of Product I purchased
Prod_I_Opt	Numeric	Quantity of Product I options purchased
Prod_J	Numeric	Quantity of Product J purchased
Prod_J_Opt	Numeric	Quantity of Product J options purchased
Prod_K	Numeric	Quantity of Product K purchased
Prod_L	Numeric	Quantity of Product L purchased
Prod_L_Opt	Numeric	Quantity of Product L options purchased
Prod_M	Numeric	Quantity of Product M purchased
Prod_N	Numeric	Quantity of Product N purchased
Prod_O	Numeric	Quantity of Product O purchased
Prod_O_Opt	Numeric	Quantity of Product O options purchased
Prod_P	Numeric	Quantity of Product P purchased
Prod_Q	Numeric	Quantity of Product Q purchased
Public_sector	Numeric	0 = No, 1 = Yes-public sector
Purchfst	Numeric	first year customer purchased
Purchlst	Numeric	Last year customer purchased
Rev_class	Cateogry	Revenue Class code; A-H (see the following explanation)
Rev_lastyr	Numeric	Last fiscal year's net operating revenues
Rev_thisyr	Numeric	This fiscal year's net operating revunes
RFM	Cateogry	Recency, Frequency, and Monetary value cell code
Seg	Category	Industry Segment code
State	Category	State where customer site is located
Tot_Revenue	Numeric	Total net operating revenues from customer all years
US_region	Category	U.S. Region of customer location
Yrs_purchase	Numeric	Number of years customer has purchased in the past

Variable Name	**Type**	**Description**
cust_id	Character	Customer ID No.
fy_qtr	Character	Fiscal Yr-Qtr
obs	Numeric	Obs
target	Numeric	Target response variable
date	Date	Date
units	Numeric	Unit Quantity
customer	Character	A=New Acquisition, C=Churn (no purch), R=Cont-Purchase
channel	Numeric	Purchase Sales Channel

City	Character	City location
State	Character	State location
Seg	Character	Industry Segment Code
loc_employee	Numeric	No of local employees
corp_rev	Numeric	Corporate Revenue last fiscal yr.
public_sector	Numeric	0-No, 1=Yes
us_region	Character	US Region Location of Business
rev_lastyr	Numeric	Last Years Fiscal Revenue
rev_thisyr	Numeric	This Years Fiscal Revenue YTD
tot_revenue	Numeric	Revenue for All Years
Purchfst	Numeric	Year of 1st Purchase
Purchlst	Numeric	Last Yr of Purchase
yrs_purchase	Numeric	No of Yrs Purchase
rev_class	Character	Revenue Class Code
RFM	Character	Recency, Freq, & Monetary Value Code
est_spend	Currency	Estimated Product-Service Spend
Prod_A	Numeric	Quantity of Product A purchased
Prod_A_Opt	Numeric	Quantity of Product A options purchased
Prod_B	Numeric	Quantity of Product B purchased
Prod_C	Numeric	Quantity of Product C purchased
Prod_D	Numeric	Quantity of Product D purchased
Prod_E	Numeric	Quantity of Product E purchased
Prod_F	Numeric	Quantity of Product F purchased
Prod_G	Numeric	Quantity of Product G purchased
Prod_H	Numeric	Quantity of Product H purchased
Prod_I	Numeric	Quantity of Product I purchased
Prod_I_Opt	Numeric	Quantity of Product I options purchased

Rev_class coding definitions

The revenue class is a coded grouping of revenue in percentile groups as follows.

A   No revenue

B   0 to 25% (bottom quartile)

C   25 to 50%

D   50 to 75%

E   75 to 90%

F   90 to 95%

G   95 to 99%

H   Top 1%

Industry Segment Code definitions:

AER   Aerospace

AUT   Automotive

BKG   Banking

CHM   Chemical

CPG   Consumer Products

ELE   Electronics

FMM   Forest, Mining, and Metals

HCR   Healthcare

INS   Insurance

MED   Media and Communications

NAT   National Government

OIL   Oil and Gas

PHM   Pharmaceuticals

PSV   Professional Services

RTL   Retail/Wholesale

SLE   State/Local, and Education

TEL   Telecom

TRV   Travel

UTL   Utilities

Data Set CUST_SURVEY_SEGMENT

This data set is a sub-set of the Customers Data set however it contains a variable called SURVEY_SEGMENTS. This survey segmentation was appended to these customers and is used in Chapter 14.

Cust_Survey_Segment Data Column Descriptions

Variable	Type	Label
cust_site_id	Character	Customer Identifier
FY1984	Numeric	Fiscal Yr 1984 Revenues
FY1985	Numeric	Fiscal Yr 1985 Revenues
FY1986	Numeric	Fiscal Yr 1986 Revenues
FY1987	Numeric	Fiscal Yr 1987 Revenues
FY1988	Numeric	Fiscal Yr 1988 Revenues
FY1989	Numeric	Fiscal Yr 1989 Revenues
FY1990	Numeric	Fiscal Yr 1990 Revenues
FY1991	Numeric	Fiscal Yr 1991 Revenues
FY1992	Numeric	Fiscal Yr 1992 Revenues
FY1993	Numeric	Fiscal Yr 1993 Revenues
FY1994	Numeric	Fiscal Yr 1994 Revenues
FY1995	Numeric	Fiscal Yr 1995 Revenues
FY1996	Numeric	Fiscal Yr 1996 Revenues
FY1997	Numeric	Fiscal Yr 1997 Revenues
FY1998	Numeric	Fiscal Yr 1998 Revenues
FY1999	Numeric	Fiscal Yr 1999 Revenues
FY2000	Numeric	Fiscal Yr 2000 Revenues
FY2001	Numeric	Fiscal Yr 2001 Revenues
FY2002	Numeric	Fiscal Yr 2002 Revenues
FY2003	Numeric	Fiscal Yr 2003 Revenues
FY2004	Numeric	Fiscal Yr 2004 Revenues
FY2005	Numeric	Fiscal Yr 2005 Revenues
FY2006	Numeric	Fiscal Yr 2006 Revenues
FY2007	Numeric	Fiscal Yr 2007 Revenues
tot_rev_allyrs	Numeric	Total Revenue All Years
first_purch_yr	Numeric	First Yr Customer Purchased
last_purch_yr	Numeric	Last Yr Customer Purchased
years_purchased	Numeric	Number of Yrs Purchased
channel_purchase	Numeric	Channel Customer Purchased
regional_geo	Character	Regional Geography Code
employee_atsite	Numeric	Syndicated Site No of Employees
synd_id2	Character	Syndicated 2nd Level ID
synd_id3	Character	Syndicated 3rd Level ID
synd_id4	Character	Syndicated 4th Level ID

SIC8	Character	Eight Digit Primary Std. Industry Class Code
industry_vert	Character	Aggregated Industry Vertical Code
company_revenue	Numeric	Syndicated Total Company Revenues
total_employees	Numeric	Syndicated Total Employees
STATE	Character	State Customer is Located In
sales_class	Character	Sales Customer Classification Code
rfm_cell	Character	RFM Cell Code A-K
RESTRICT_MAIL	Character	If Direct Mail Restricted (Y/N)
RESTRICT_PHONE	Character	If Phone Contact Restricted (Y/N)
RESTRICT_EMAIL	Character	If Email Contact Restricted (Y/N)
survey_segments	Numeric	Survey Segment Number Response
it_budget	Character	IT Budget Range A-E
it_spending	Currency	Estimated IT Spending in $

Data Set NYTOWNS

This data set comes from Data-Miners.com (used with permission) and the column descriptions are as follows. There are 1,006 records (rows) in this data set.

NY Towns Data Column Descriptions

Variable	Type	Label
AncArab	Num	ANCESTRY (single or multiple); Total ancestries reported; Arab; Percent
AncCzech	Num	ANCESTRY (single or multiple); Total ancestries reported; Czech1; Percent
AncDanish	Num	ANCESTRY (single or multiple); Total ancestries reported; Danish; Percent
AncDutch	Num	ANCESTRY (single or multiple); Total ancestries reported; Dutch; Percent
AncEnglish	Num	ANCESTRY (single or multiple); Total ancestries reported; English; Percent
AncFrCanad	Num	ANCESTRY (single or multiple); Total ancestries reported; French Canadian1; Percent
AncFrench	Num	ANCESTRY (single or multiple); Total ancestries reported; French (except Basque)1; Percent
AncGerman	Num	ANCESTRY (single or multiple); Total ancestries reported; German; Percent
AncGreek	Num	ANCESTRY (single or multiple); Total ancestries reported; Greek; Percent
AncHungary	Num	ANCESTRY (single or multiple); Total ancestries reported; Hungarian; Percent

| AncIrish | Num | ANCESTRY (single or multiple); Total ancestries reported; Irish1; Percent |

Variable	Type	Label
AncItalian	Num	ANCESTRY (single or multiple); Total ancestries reported; Italian; Percent
AncLithu	Num	ANCESTRY (single or multiple); Total ancestries reported; Lithuanian; Percent
AncNorweg	Num	ANCESTRY (single or multiple); Total ancestries reported; Norwegian; Percent
AncOthr	Num	ANCESTRY (single or multiple); Total ancestries reported; Other ancestries; Percent
AncPolish	Num	ANCESTRY (single or multiple); Total ancestries reported; Polish; Percent
AncPortug	Num	ANCESTRY (single or multiple); Total ancestries reported; Portuguese; Percent
AncRepP	Num	ANCESTRY (single or multiple); Total ancestries reported; Percent
AncRussian	Num	ANCESTRY (single or multiple); Total ancestries reported; Russian; Percent
AncScot	Num	ANCESTRY (single or multiple); Total ancestries reported; Scottish; Percent
AncScotIre	Num	ANCESTRY (single or multiple); Total ancestries reported; Scotch-Irish; Percent
AncSlovak	Num	ANCESTRY (single or multiple); Total ancestries reported; Slovak; Percent
AncSubSah	Num	ANCESTRY (single or multiple); Total ancestries reported; Subsaharan African; Percent
AncSwedish	Num	ANCESTRY (single or multiple); Total ancestries reported; Swedish; Percent
AncSwiss	Num	ANCESTRY (single or multiple); Total ancestries reported; Swiss; Percent
AncUS	Num	ANCESTRY (single or multiple); Total ancestries reported; United States or American; Percent
AncUkraine	Num	ANCESTRY (single or multiple); Total ancestries reported; Ukrainian; Percent
AncWIndian	Num	ANCESTRY (single or multiple); Total ancestries reported; West Indian (excluding Hispanic groups); Percent

Variable	Type	Label
AncWelsh	Num	ANCESTRY (single or multiple); Total ancestries reported; Welsh; Percent
AreaLand	Num	
AreaWater	Num	
BadKitchen	Num	Occupied Housing Units; Selected characteristics; Lacking complete kitchen facilities; Percent
BadPlumbing	Num	Occupied Housing Units; Selected characteristics; Lacking complete plumbing facilities; Percent
BoatRVVan	Num	Total housing units; Units in structure; Boat, RV, van, etc.; Percent
BornAfrica	Num	Region of birth of foreign born; Total (excluding born at sea); Africa; Percent
BornAsia	Num	Region of birth of foreign born; Total (excluding born at sea); Asia; Percent
BornEurope	Num	Region of birth of foreign born; Total (excluding born at sea); Europe; Percent
BornLatAmer	Num	Region of birth of foreign born; Total (excluding born at sea); Latin America; Percent
BornNorAmer	Num	Region of birth of foreign born; Total (excluding born at sea); Northern America; Percent
BornOceania	Num	Region of birth of foreign born; Total (excluding born at sea); Oceania; Percent
BuiltAfter40	Num	Cumulative percentage of % built after X
BuiltAfter60	Num	Cumulative percentage of % built after X
BuiltAfter70	Num	Cumulative percentage of % built after X
BuiltAfter80	Num	Cumulative percentage of % built after X
BuiltAfter90	Num	Cumulative percentage of % built after X
BuiltAfter95	Num	Cumulative percentage of % built after X
BuiltAfter99	Num	Cumulative percentage of % built after X
COUNTY	Char	County
CitizenNative	Num	Nativity and place of birth; Total population; Native; Percent

(*continued*)

Variable	Type	Label
CitizenNatr	Num	Nativity and place of birth; Total population; Foreign born; Naturalized citizen; Percent
CitizenNot	Num	Nativity and place of birth; Total population; Foreign born; Not a citizen; Percent
CommuteAtHome	Num	Commuting to work; Workers 16 years and over; Worked at home; Percent
CommuteAvgTravTime	Num	Commuting to work; Workers 16 years and over; Mean travel time to work (minutes); Number
CommuteCarpool	Num	Commuting to work; Workers 16 years and over; Car, truck, or van -- carpooled; Percent
CommuteDrive	Num	Commuting to work; Workers 16 years and over; Car, truck, or van -- drove alone; Percent
CommuteOther	Num	Commuting to work; Workers 16 years and over; Other means; Percent
CommutePubTran	Num	Commuting to work; Workers 16 years and over; Public transportation (including taxicab); Percent
CommuteWalk	Num	Commuting to work; Workers 16 years and over; Walked; Percent
CostDivIncLT15	Num	Specified owner-occupied units; Selected monthly owner costs as a percentage of household income in 1999; Less than 15 percent; Percent
CostDivIncLT20	Num	Owner costs as % of household income less than 19%; Adjust for not computed
CostDivIncLT25	Num	Owner costs as % of household income less than 24%; Adjust for not computed
CostDivIncLT30	Num	Owner costs as % of household income less than 29%; Adjust for not computed
CostDivIncLT35	Num	Owner costs as % of household income less than 34% (1- % 35 or greater); Adjust for not computed
DisAdult	Num	Disability status of the civilian noninstitutionalized population; Population 21 to 64 years; With a disability; Percent
DisEmploy	Num	Disability status of the civilian noninstitutionalized population; Population 21 to 64 years; With a disability; Percent employed; Number

(*continued*)

Variable	Type	Label
DisSenior	Num	Disability status of the civilian noninstitutionalized population; Population 65 years and over; With a disability; Percent
DisYoung	Num	Disability status of the civilian noninstitutionalized population; Population 5 to 20 years; With a disability; Percent
Edu2YDeg	Num	Educational attainment; Population 25 years and over; Associate degree; Percent
Edu4YDeg	Num	Educational attainment; Population 25 years and over; Bachelors degree; Percent
EduBAplus	Num	Educational attainment; Population 25 years and over; Percent bachelors degree or higher; Number
EduColNoDeg	Num	Educational attainment; Population 25 years and over; Some college, no degree; Percent
EduGradDeg	Num	Educational attainment; Population 25 years and over; Graduate or professional degree; Percent
EduHSDip	Num	Educational attainment; Population 25 years and over; High school graduate (includes equivalency); Percent
EduHSNoDip	Num	Educational attainment; Population 25 years and over; 9th to 12th grade, no diploma; Percent
EduHSplus	Num	Educational attainment; Population 25 years and over; Percent high school graduate or higher; Number
EduLT9	Num	Educational attainment; Population 25 years and over; Less than 9th grade; Percent
FSChilPP	Num	POVERTY STATUS IN 1999 (below poverty level); Families; With related children under 18 years; With related children under 5 years; Percent below poverty level; Percent
FamIncLT100K	Num	Income in 1999; Families; Less than $100,000; Percent
FamIncLT10K	Num	Income in 1999; Families; Less than $10,000; Percent
FamIncLT150K	Num	Income in 1999; Families; Less than $150,000; Percent
FamIncLT15K	Num	Income in 1999; Families; Less than $15,000; Percent
FamIncLT200K	Num	Income in 1999; Families; Less than $200,000; Percent
FamIncLT25K	Num	Income in 1999; Families; Less than $25,000; Percent

(*continued*)

Variable	Type	Label
FamIncLT35K	Num	Income in 1999; Families; Less than $35,000; Percent
FamIncLT50K	Num	Income in 1999; Families; Less than $50,000; Percent
FamIncLT75K	Num	Income in 1999; Families; Less than $25,000; Percent
GEO_ID	Char	Geographic key from census bureau
GEO_NAME	Char	Geography
GPGuardP	Num	Grandparents as caregivers; Grandparent living in household with one or more own grandchildren under 18 years; Grandparent responsible for grandchildren; percent
HouseIncLT100K	Num	Income in 1999; Households; Less than $100,000; Percent
HouseIncLT10K	Num	Income in 1999; Households; Less than $10,000; Percent
HouseIncLT150K	Num	Income in 1999; Households; Less than $150,000; Percent
HouseIncLT15K	Num	Income in 1999; Households; Less than $15,000; Percent
HouseIncLT200K	Num	Income in 1999; Households; Less than $200,000; Percent
HouseIncLT25K	Num	Income in 1999; Households; Less than $25,000; Percent
HouseIncLT35K	Num	Income in 1999; Households; Less than $35,000; Percent
HouseIncLT50K	Num	Income in 1999; Households; Less than $50,000; Percent
HouseIncLT75K	Num	Income in 1999; Households; Less than $75,000; Percent
HouseMultiFamily	Num	% of housing units that are multifamily
IncAvgEarnings	Num	Income in 1999; Households; With earnings; Mean earnings (dollars); Number
IncAvgPubAss	Num	Income in 1999; Households; With public assistance income; Mean public assistance income (dollars); Number
IncAvgRetire	Num	Income in 1999; Households; With retirement income; Mean retirement income (dollars); Number
IncAvgSocSec	Num	Income in 1999; Households; With Social Security income; Mean Social Security income (dollars); Number
IncAvgSupSec	Num	Income in 1999; Households; With Supplemental Security Income; Mean Supplemental Security Income (dollars); Number

Variable	Type	Label
IncEarnings	Num	Income in 1999; Households; With earnings; Percent
IncMedEarnings	Num	Income in 1999; Households; Median household income (dollars); Number
IncMedFamily	Num	Income in 1999; Families; Median family income (dollars); Number
IncMedFemales	Num	Income in 1999; Median income for Females (dollars); Number
IncMedMales	Num	Income in 1999; Median income for Males (dollars); Number
IncPerCapita	Num	Income in 1999; Families; Per capita income (dollars); Number
IncPubAss	Num	Income in 1999; Households; With public assistance income; Percent
IncRatioM2F	Num	Male to Female Median Income Ratio
IncRetirement	Num	Income in 1999; Households; With retirement income; Percent
IncSecSec	Num	Income in 1999; Households; With Social Security income; Percent
IncSupSec	Num	Income in 1999; Households; With Supplemental Security Income; Percent
IndAgric	Num	Employed civilian population 16 years and over; Industry; Agriculture, forestry, fishing and hunting, and mining; Percent
IndArtsEntertainment	Num	Employed civilian population 16 years and over; Industry; Arts, entertainment, recreation, accommodation and food services; Percent
IndConstruction	Num	Employed civilian population 16 years and over; Industry; Construction; Percent
IndEduHealthSocServ	Num	Employed civilian population 16 years and over; Industry; Educational, health and social services; Percent
IndFinanceRealEstate	Num	Employed civilian population 16 years and over; Industry; Finance, insurance, real estate, and rental and leasing; Percent
IndInformation	Num	Employed civilian population 16 years and over; Industry; Information; Percent
IndManufacturing	Num	Employed civilian population 16 years and over; Industry; Manufacturing; Percent

(*continued*)

Variable	Type	Label
IndOtherService	Num	Employed civilian population 16 years and over; Industry; Other services (except public administration); Percent
IndProfScienceWaste	Num	Employed civilian population 16 years and over; Industry; Professional, scientific, management, administrative, and waste management services; Percent
IndPublicAdmin	Num	Employed civilian population 16 years and over; Industry; Public administration; Percent
IndRetail	Num	Employed civilian population 16 years and over; Industry; Retail trade; Percent
IndTransport	Num	Employed civilian population 16 years and over; Industry; Transportation and warehousing, and utilities; Percent
IndWholesale	Num	Employed civilian population 16 years and over; Industry; Wholesale trade; Percent
JobAgriculture	Num	Employed civilian population 16 years and over; Occupation; Farming, fishing, and forestry occupations; Percent
JobConstruct	Num	Employed civilian population 16 years and over; Occupation; Construction, extraction, and maintenance occupations; Percent
JobManageProf	Num	Employed civilian population 16 years and over; Occupation; Management, professional, and related occupations; Percent
JobOfficeSales	Num	Employed civilian population 16 years and over; Occupation; Sales and office occupations; Percent
JobService	Num	Employed civilian population 16 years and over; Occupation; Service occupations; Percent
JobTransport	Num	Employed civilian population 16 years and over; Occupation; Production, transportation, and material moving occupations; Percent
LabCivilEmployed	Num	
LabFCivilEmployed	Num	
LaborAll	Num	Employment status; Population 16 years and over; In labor force; Percent
LaborArmedForces	Num	Employment status; Population 16 years and over; In labor force; Armed Forces; Percent
LaborCivilian	Num	Employment status; Population 16 years and over; In labor force; Civilian labor force; Percent

(*continued*)

Variable	Type	Label
LaborFemale	Num	Employment status; Females 16 years and over; In labor force; Percent
LaborUnder6ParentsWork		Num Employment status; Own children under 6 years; All parents in family in labor force; Percent
LangAsiaPoorEng	Num	Language spoken at home; Population 5 years and over; Language other than English; Asian and Pacific Island languages; Speak English less than very well; Percent
LangAsian	Num	Language spoken at home; Population 5 years and over; Language other than English; Asian and Pacific Island languages; Number
LangIEPoorEng	Num	Language spoken at home; Population 5 years and over; Language other than English; Other Indo-European languages; Speak English less than very
LangIndoEuro	Num	Language spoken at home; Population 5 years and over; Language other than English; Other Indo-European languages; Percent
LangNotEng	Num	Language spoken at home; Population 5 years and over; Language other than English; Percent
LangOnlyEng	Num	Language spoken at home; Population 5 years and over; English only; Percent
LangPoorEng	Num	Language spoken at home; Population 5 years and over; Language other than English; Speak English less than very well; Percent
LangSpan	Num	Language spoken at home; Population 5 years and over; Language other than English; Spanish; Percent
LangSpanPoorEng	Num	Language spoken at home; Population 5 years and over; Language other than English; Spanish; Speak English less than very well; Percent
Latitude	Num	Latitude of town location
LessEq1Room	Num	% with 1 room
LessEq2Rooms	Num	% with 2 or fewer rooms
LessEq3Rooms	Num	% with 3 or fewer rooms
LessEq4Rooms	Num	% with 4 or fewer rooms
LessEq5Rooms	Num	% with 5 or fewer rooms

Variable	Type	Label
LessEq6Rooms	Num	% with 6 or fewer rooms
LessEq7Rooms	Num	% with 7 or fewer rooms
LessEq8Rooms	Num	% with 8 or fewer rooms (1-% with 9 or more)
Longitude	Num	Longitude of town location
MarDivorce	Num	Marital status; Population 15 years and over; Divorced; Percent
MarDivorcedFemales	Num	Marital status; Population 15 years and over; Divorced; Female; Percent
MarFemaleDivorcees	Num	% of marriage aged females who are divorced
MarFemaleWidows	Num	Percent of widows that are female
MarNever	Num	Marital status; Population 15 years and over; Never married; Percent
MarSep	Num	Marital status; Population 15 years and over; Separated; Percent
MarWidow	Num	Marital status; Population 15 years and over; Widowed; Percent
MarWidowedFemales	Num	Marital status; Population 15 years and over; Widowed; Female; Percent
Married	Num	Marital status; Population 15 years and over; Now married, except separated; Percent
MedHomeValue	Num	Specified owner-occupied units; Value; Median (dollars); Number
MedOwnerCostWMort	Num	Median Monthly ownership cost with mortgage
MedOwnerCostWOMort	Num	Median Monthly ownership cost with no mortgage
MedRoom	Num	Total housing units; Rooms; Median (rooms); Number
MedianRent	Num	Specified renter-occupied units; Gross rent; Median (dollars); Number
Mobile	Num	Total housing units; Units in structure; Mobile home; Percent
MortageLT300	Num	Percent of mortgages with payment less than $300
MortageLT500	Num	Percent of mortgages with payment less than $500
MortageLT1000	Num	Percent of mortgages with payment less than $1,000

(*continued*)

Variable	Type	Label
MortageLT1500	Num	Percent of mortgages with payment less than $1,500
MortageLT2000	Num	Percent of mortgages with payment less than $2,000
MortgageLT700	Num	Percent of mortgages with payment less than $700
NAME	Char	Area Name-Legal/Statistical Area Description (LSAD) Term-Part Indicator
NativityDifState	Num	Nativity and place of birth; Total population; Native; Born in United States; Different state; Percent
NativityForeign	Num	Nativity and place of birth; Total population; Foreign born; Percent
NativityNY	Num	Nativity and place of birth; Total population; Native; Born in United States; State of residence; Percent
NativityNewFor	Num	Nativity and place of birth; Total population; Foreign born; Entered 1990 to March 2000; Percent
NativityOutUS	Num	Nativity and place of birth; Total population; Native; Born outside United States; Percent
NativityUS	Num	Nativity and place of birth; Total population; Native; Born in United States; Percent
NoCashRent	Num	Specified renter-occupied units; Gross rent; No cash rent; Percent
NoDisEmploy	Num	Disability status of the civilian noninstitutionalized population; Population 21 to 64 years; No disability; Percent employed; Number
NoPhone	Num	Occupied Housing Units; Selected characteristics; No telephone service; Percent
OwnWMortgage	Num	Specified owner-occupied units; Mortgage status and selected monthly owner costs; With a mortgage; Percent
OwnWOMortgage	Num	Specified owner-occupied units; Mortgage status and selected monthly owner costs; Not mortgaged; Percent
Penetration	Num	Product penetration (percent of households)
PopRural	Num	Number of rural population
PopTotal	Num	Number of total population
PopUrban	Num	Number of urban population

Variable	Type	Label
PovFamChildNoHusband	Num	POVERTY STATUS IN 1999 (below poverty level); Families with female householder, no husband present; With related children under 18 years; Number
PovFamNoHusband	Num	POVERTY STATUS IN 1999 (below poverty level); Families with female householder, no husband present; With related children under 18 years; Percent
PovFamWBaby	Num	POVERTY STATUS IN 1999 (below poverty level); Families with female householder, no husband present; Percent below poverty level; Percent
PovFamWChild	Num	POVERTY STATUS IN 1999 (below poverty level); Families; With related children under 18 years; Percent below poverty level; Percent
PovFamilies	Num	POVERTY STATUS IN 1999 (below poverty level); Families; Percent below poverty level; Percent
PovIndAdults	Num	POVERTY STATUS IN 1999 (below poverty level); Individuals; 18 years and over; Percent below poverty level; Percent
PovIndChild	Num	POVERTY STATUS IN 1999 (below poverty level); Individuals; Related children under 18 years; Percent below poverty level; Percent
PovIndSeniors	Num	POVERTY STATUS IN 1999 (below poverty level); Individuals; 18 years and over; 65 years and over; Percent below poverty level; Percent
PovIndividual	Num	POVERTY STATUS IN 1999 (below poverty level); Individuals; Percent below poverty level; Percent
RentDivIncLT15	Num	Specified renter-occupied units; Gross rent as a percentage of household income in 1999; Less than 15 percent; Percent
RentDivIncLT20	Num	Rent as % of income less than 20%; adjust for not computed
RentDivIncLT25	Num	Rent as % of income less than 25%; adjust for not computed
RentDivIncLT30	Num	Rent as % of income less than 30%; adjust for not computed
RentDivIncLT35	Num	Rent as % of income less than 35% (1- % 35 or over); adjust for not computed
RentLT200	Num	% rent less than $200
RentLT300	Num	% rent less than $300

Variable	Type	Label
RentLT500	Num	% rent less than$ 500
RentLT750	Num	% rent less than $750
RentLT1000	Num	% rent less than $1,000
RentLT1500	Num	% rent less than $1,500
ResDfCnty	Num	Residence in 1995; Population 5 years and over; Different house in the U.S. in 1995; Different county; Percent
ResDfHome	Num	Residence in 1995; Population 5 years and over; Different house in the U.S. in 1995; Percent
ResDfState	Num	Residence in 1995; Population 5 years and over; Different house in the U.S. in 1995; Different county; Different state; Percent
ResElse	Num	Residence in 1995; Population 5 years and over; Elsewhere in 1995; Percent
ResSame	Num	Residence in 1995; Population 5 years and over; Same house in 1995; Percent
ResSmCnty	Num	Residence in 1995; Population 5 years and over; Different house in the U.S. in 1995; Same county; Percent
ResSmState	Num	Residence in 1995; Population 5 years and over; Different house in the U.S. in 1995; Different county; Same state; Percent
Resid10Yrs	Num	Resident for 10 or more years
Resid1Yr	Num	Resident for 1 or more years
Resid20Yrs	Num	Resident for 20 or more years
Resid30Yrs	Num	Resident for 30 or more years
Resid5Yrs	Num	Resident for 5 or more years
SchElem	Num	School enrollment; Population 3 years and over enrolled in school; Elementary school (grades 1-8); Percent
SchHS	Num	School enrollment; Population 3 years and over enrolled in school; High school (grades 9-12); Percent
SchKind	Num	School enrollment; Population 3 years and over enrolled in school; Kindergarten; Percent
SchNurs	Num	School enrollment; Population 3 years and over enrolled in school; Nursery school, preschool; Percent

(*continued*)

Variable	Type	Label
SchUniv	Num	School enrollment; Population 3 years and over enrolled in school; College or graduate school; Percent
ValueLT100K	Num	% value less than $100,000
ValueLT150K	Num	% value less than $150,000
ValueLT1Mil	Num	% value less than $1,000,000 (1-% over $1,000,000)
ValueLT200K	Num	% value less than $200,000
ValueLT300K	Num	% value less than $300,000
ValueLT500K	Num	% value less than $500,000
ValueLT50K	Num	% value less than $50,000
VehicGT0	Num	At least one vehicle available
VehicGT1	Num	More than one vehicle available
VehicGT2	Num	More than two vehicles available
VeteranC	Num	Veteran status; Civilian population 18 years and over; Civilian veterans; Percent
WorkClassFamily	Num	Employed civilian population 16 years and over; Class of worker; Unpaid family workers; Percent
WorkClassGovt	Num	Employed civilian population 16 years and over; Class of worker; Government workers; Percent
WorkClassSalary	Num	Employed civilian population 16 years and over; Class of worker; Private/wage & salary workers; Percent
WorkClassSelf	Num	Employed civilian population 16 years and over; Class of worker; Self-employed workers in own not incorporated business; Percent

Data Set NEWS

This data set contains 600 records (rows).

Variable Name	Type	Description
TEXT	Text	Contains the text of a sample of news stories.
Graphics	Numeric	Binary (0 and 1) – 1 means the text is about graphics; 0 means not about graphics.
Hockey	Numeric	Binary (0 and 1) – 1 means the text is about hockey; 0 means not about hockey.
Medical	Numeric	Binary (0 and 1) – 1 meaning the text is about a medical topic; 0 means not about a medical topic.

The data for the Web crawling is derived from DMReview.com Web site and the content changes over time.

Data Set PRODSALES

This data set contains 27 records (rows).

Variable Name	Type	Description
Customer	Text	Customer name
Item	Text	Retail grocery product item description

Data Set SATISFACTION

This data set contains 641 records (rows).

Variable Name	Type	Description
Comment	Text	General Comment Text
Rating	Numeric	Customer Rating Metric
Mrating	Numeric	Major Rating (Binary: 1,2)

Data Set ENSEMBLE_SEGMENTS_TRAIN

This data set contains 39,109 records (rows).

Variable Name	Type	Description
cust_site_id	Character	Customer Identifier
FY1984	Numeric	Fiscal Yr 1984 Revenues
FY1985	Numeric	Fiscal Yr 1985 Revenues
FY1986	Numeric	Fiscal Yr 1986 Revenues
FY1987	Numeric	Fiscal Yr 1987 Revenues
FY1988	Numeric	Fiscal Yr 1988 Revenues
FY1989	Numeric	Fiscal Yr 1989 Revenues
FY1990	Numeric	Fiscal Yr 1990 Revenues
FY1991	Numeric	Fiscal Yr 1991 Revenues
FY1992	Numeric	Fiscal Yr 1992 Revenues
FY1993	Numeric	Fiscal Yr 1993 Revenues
FY1994	Numeric	Fiscal Yr 1994 Revenues
FY1995	Numeric	Fiscal Yr 1995 Revenues

FY1996	Numeric	Fiscal Yr 1996 Revenues
FY1997	Numeric	Fiscal Yr 1997 Revenues
FY1998	Numeric	Fiscal Yr 1998 Revenues
FY1999	Numeric	Fiscal Yr 1999 Revenues
FY2000	Numeric	Fiscal Yr 2000 Revenues
FY2001	Numeric	Fiscal Yr 2001 Revenues
FY2002	Numeric	Fiscal Yr 2002 Revenues
FY2003	Numeric	Fiscal Yr 2003 Revenues
FY2004	Numeric	Fiscal Yr 2004 Revenues
FY2005	Numeric	Fiscal Yr 2005 Revenues
FY2006	Numeric	Fiscal Yr 2006 Revenues
FY2007	Numeric	Fiscal Yr 2007 Revenues
tot_rev_allyrs	Numeric	Total Revenue All Years
first_purch_yr	Numeric	First Yr Customer Purchased
last_purch_yr	Numeric	Last Yr Customer Purchased
years_purchased	Numeric	Number of Yrs Purchased
channel_purchase	Numeric	Channel Customer Purchased
regional_geo	Character	Regional Geography Code
synd_id2	Character	Syndicated 2nd Level ID
synd_id3	Character	Syndicated 3rd Level ID
synd_id4	Character	Syndicated 4th Level ID
industry_vert	Character	Aggregated Industry Vertical Code
STATE	Character	State Customer is Located In
sales_class	Character	Sales Customer Classification Code
rfm_cell	Character	RFM Cell Code A-K
RESTRICT_MAIL	Character	If Direct Mail Restricted (Y/N)
RESTRICT_PHONE	Character	If Phone Contact Restricted (Y/N)
RESTRICT_EMAIL	Character	If Email Contact Restricted (Y/N)
survey_segments	Numeric	Survey Segment Number Response
it_budget	Character	IT Budget Range A-E
PWR_company_revenue	Numeric	Transformed: Syndicated Total Company Revenues
TRANS_0	Numeric	Log(IT Spend)
TRANS_1	Numeric	Log(Employee @ Site)
TRANS_2	Numeric	Log(Total Employees)
_SEGMENT_LABEL_	Character	Segment Description
SOM_SEGMENT	Numeric	SOM Segment ID
WARN	Character	Warnings
F_SOM_SEGMENT	Character	From: SOM_SEGMENT
U_SOM_SEGMENT	Numeric	Unnormalized Into: SOM_SEGMENT
R_SOM_SEGMENT16	Numeric	Residual: SOM_SEGMENT=16
R_SOM_SEGMENT15	Numeric	Residual: SOM_SEGMENT=15
R_SOM_SEGMENT14	Numeric	Residual: SOM_SEGMENT=14
R_SOM_SEGMENT13	Numeric	Residual: SOM_SEGMENT=13

R_SOM_SEGMENT12	Numeric	Residual: SOM_SEGMENT=12
R_SOM_SEGMENT11	Numeric	Residual: SOM_SEGMENT=11
R_SOM_SEGMENT10	Numeric	Residual: SOM_SEGMENT=10
R_SOM_SEGMENT9	Numeric	Residual: SOM_SEGMENT=9
R_SOM_SEGMENT8	Numeric	Residual: SOM_SEGMENT=8
R_SOM_SEGMENT7	Numeric	Residual: SOM_SEGMENT=7
R_SOM_SEGMENT6	Numeric	Residual: SOM_SEGMENT=6
R_SOM_SEGMENT5	Numeric	Residual: SOM_SEGMENT=5
R_SOM_SEGMENT4	Numeric	Residual: SOM_SEGMENT=4
R_SOM_SEGMENT3	Numeric	Residual: SOM_SEGMENT=3
R_SOM_SEGMENT2	Numeric	Residual: SOM_SEGMENT=2
R_SOM_SEGMENT1	Numeric	Residual: SOM_SEGMENT=1
EM_EVENTPROBABILITY	Numeric	Probability for level 16 of SOM_SEGMENT
EM_PROBABILITY	Numeric	Probability of Classification
EM_CLASSIFICATION	Character	Prediction for SOM_SEGMENT
EM_CLASSTARGET	Character	Target Variable: SOM_SEGMENT