Business Context:
(Reference: Description and Documentation for the Multi-Channel Gift Company Dataset Version 1.0; Richard J. Courtheoux, President - Marketing Analysis Applications, Inc.; rcourtheoux@maa-online.com)

A multichannel company with sales of several hundred million dollars per year has provided the Direct Marketing Educational Foundation (DMEF) with a dataset. The multichannel company has a network of retail stores, a well-established traditional catalog channel and a website. Its brand is very well known nationally, and it has a strong positive reputation along with very good long term customer loyalty. The core of the company's business consists of food products which are often purchased as gifts during the Christmas season. Their marketing challenges include:

- Customer relationship management, including contact management, across multiple channels.
- A very competitive marketplace for Christmas season gifts
- Generating business during the non-Christmas months of the year

The DMEF dataset consists of 4 files that have been extracted from the company's marketing database and that can be linked based on customer ID numbers. You will be working with just 1 of the 4 databases. This is the customer records file, with 100,051 customer records that have a customer ID, buying activity summarized by channel for 8 seasons, channel summary Recency/Frequency/Monetary information prior to the 8 seasons, gift recipients counts by season, catalog and e-mail circulation counts by season for 6 seasons, first order characteristics, overlay demographic and psychographic data, estimated distance to a retail store and the sectional center code (i.e., the first 3 digits of the customer's ZIPCode).

Detailed Data Description

The data file is provided in CSV format. The first field in the file is a customer ID and the data is sorted by customer ID.

Customer File: DMEFExtractSummaryV01.csv

Cust_ID is the customer ID. The file is sorted on this variable in ascending order.

The purchasing behaviour variables have a naming scheme composed of three components:

- Channel – options are Ret, Int and Cat
- Season – these usually begin with either F (fall) or S (spring) and are followed by the last 2 digits of the year (e.g., '07' for the year 2007). Summaries for the period prior to 2004 have the designation Pre04.
- Measure – options are Dollars, GDollars, NGDollars, Orders, Trips and Lines. Retail sales are not allocated to gift or non-gift so they are just "Dollars" whereas the Internet and catalog sales are reported separately for gift and non-gift. In the Pre04 variables there is also a Recency measure which is the recency of last purchase measured in number of seasons prior to 2004 (e.g., a value 2 indicates that the most recent purchase prior to 2004 was in Spring 2003); a value of 99 indicates that the customer had not purchased in the channel prior to 2004.

So, for example, the variable IntF06GDollars has a customer's Fall 2006 Internet gift sales dollar total.

Variables FirstYYMM, FirstChannel, and FirstDollar give characteristics of the first purchase found in the purchase data.

AcqDate is the date (years and months only) that the company originally added the customer to its database. In some cases, the company originally acquired the customer prior to the earliest retained transactional data (i.e., prior to 2001).

StoreDist gives the estimated distance from the customer's address to the nearest company store in miles. SCF_Code gives the first three digits of the customer's ZIPCode.

Demographic and psychographic variables were obtained through an overlay with a data vendor company. Not all customer records could be matched in the vendor company's database, so many of these fields are blank.

- 'Y' and 'N' codes are used to indicate whether a customer has an interest in a particular topic. These codes pertain to variables Travel, CurrAff, CurrEv, Wines, FineArts, Exercise, SelfHelp, Collect, Needle, Sewing, DogOwner, CarOwner, Cooking, Pets, Fashion, Camping, Hunting and Boating.
- 'Y' and 'N' codes are also used to indicate presence of children of various age ranges in a household. These variables are Child0_2, Child3_5, Child6_11, Child12_16 and Child17_18.
- 'Y' and 'N' codes for variable Email indicate whether or not the company has permission to send e-mails to the customer.
- The HomeValue variable simply gives the estimated value of a customer's home in dollars.
- Specific coding schemes are used for the AgeCode, IncCode, HomeCode, Dwelling, LengthRes and OccupCd variables. Reference tables for these codes are in separate tabs of the spreadsheet DMEF Demo Codes Reference which is provided along with this document.

Using this file, you should address the following marketing issues (resolve at least 4 of the 8 analytics challenges):

- Building predictive segmentation systems. You can predict Fall 2007 sales or response rates by channel using only sales data from Spring 2007 and earlier as well as the demographic, psychographic and store distance variables. Predictions might be done using a Recency/Frequency/Monetary cell approach or a multivariate statistical model.
- Tracking customer value over time ("Customer Lifetime Value"). You can isolate groups of customers based on different characteristics (e.g., initial purchase channel, initial purchase amount, demographics, psychographics or store distance) and see how total value to the company builds up over time. To create a real Lifetime Value calculation, assumptions would need to be made regarding cost of goods, contact costs for both catalogs and e-mails and a discount rate to be applied across seasons.
- Channel loyalty or switching over time. You can examine how initial channel of purchase is predictive of channel usage for subsequent purchases. The period 2001 – 2007 covered by the sales data was one in which the Internet channel matured and became a large part

of the company's sales. Retail channel usage depends to a large extent on distance to a retail store.

- Demographic or psychographic profiling. You can look for relationships between purchasing or channel usage and the various available demographic and psychographic variables. The degree to which prior purchasing activity is a superior predictor of future behaviour as compared to demographics and psychographics might be an important lesson for students.
- Marketing contact strategy analysis. You can relate the intensity of each type of marketing contact to the prior buying activity by channel. If cost assumptions are made about each type of marketing contact, you can look at marketing cost to sales ratios for segments of the customer base.
- Store trade area effects and impacts on all channels. You can look at retail sales by distance to store. You can also look at the channel mix of buying as customers live farther away from a store. Using the demographic and psychographic data students could do profiles of customers based on their distance from a retail store.
- Gift giving behaviour consistency and predictability. Since Christmas season gift giving is a large part of this company's business, gift and non-gift revenue is reported separately for the Internet and catalog channels. (The company's retail systems do not identify gift and non-gift sales). Also, the number of distinct gift recipients is provided by season; a separate count of new gift recipients by season is provided as well. Customers may be segmented by how much of their purchasing is for gift purposes.
- Purchasing by season effects. This company's sales are predominately in the fall (Christmas) season, so it may be important to look at the degree to which customers buy in the spring season and if those spring season buyers have distinct characteristics. The degree of seasonality in customer purchases is related to the customer's purchase channel choices.

The company which provided the raw data for this dataset wishes to remain anonymous. In order to respect that wish some adjustments to the raw data have been made to make it less likely that the company can be successfully identified based on the dataset; also, \different sampling rates based on customer characteristics were applied to select customers for this sample. If someone were to make a lucky guess about the identity of the company, the basic business ratios in this dataset would not accurately match the real numbers which characterize this company's business. The use of these adjustment and sampling procedures should not reduce the utility of the dataset as a resource for analytical modelling.