CanaCode

– Lifestyle Clusters of Canadian Consumers

220 Duncan Mill Road, Suite 519
Toronto, ON M3B 3J5
CANADA
Tel.: 416-760-8828 Fax: 416-760-8826
Email: info@manifolddataming.com
Http: www.manifolddatamining.com

CanaCode – Lifestyle Clusters of Canadian Consumers

Consumer Lifestyle
A consumer's lifestyle is the outward expression of his or her values and needs, reflected through patterns in his/her activities, interests and opinions. A consumer's lifestyle can change over time because of changing environmental influences, i.e.:

- Life stage, education and culture;
- The reference groups to which the consumer belongs or aspires;
- Social class;
- Family's beliefs;
- Income and economic conditions which limit purchasing power and capacity;
- The information sources available to the consumer;
- Education level attained.

With the rapid development of information technology, consumers are becoming more sophisticated and diversified. New lifestyle segments are forming and old ones are shrinking. Identifying market segments, anticipating demand of different lifestyles, and being the first to proactively engage with the consumers will lead to competitive edge, successful product and service offering, as well as prosperous relationship with the consumers.

Cluster Analysis
Cluster analysis is a process of identifying patterns in consumer lifestyles and classifying consumers into different groups. It is based on the premise that people sharing similar lifestyles have similar consumption behaviors. The strategy of target marketing with lifestyle clustering is to capture consumer's lifestyle, cluster by cluster, and secure the clusters with greatest potential before competitors do. Based on lifestyle clusters, marketers can attain better penetration into the larger, higher value clusters as well as those clusters with the greatest propensity to purchase their products or services.

There are many clustering algorithms for various types of data, for examples,
- Hierarchical clustering techniques;
- Fuzzy Clustering algorithms;
- Center-based clustering algorithms;
- Search-based clustering algorithms;
- Graph-based clustering algorithms;
- Grid-based clustering algorithms;
- Density-based clustering algorithms;
- Model-based clustering algorithms;
- Subspace clustering.

In developing our CanaCode Lifestyle Clusters we took a hybrid approach, i.e., we combined hierarchical clustering techniques with fuzzy logics and implemented them into a center-based classification procedure.

Lifestyle Marketing

Lifestyle Marketing is about meeting the consumer's needs and desires by tailoring products/services to his/her lifestyles. Marketers nowadays can leverage the widely available comprehensive and precise consumer data, and advanced data mining techniques to predict consumer's lifestyle and select the right micro-communication channels to reach them.

Focus and power have shifted from brands to consumers. Now consumers have unprecedented information to make their purchase decisions. The growing number of choices and channels has fueled directly in to more demand in communications, products and services. Consumers are no longer passive receptors of advertising and promotional messages. In the dynamic lifestyle and business environment they become active participants and influencers. They demand communications, products and services that suit their own lifestyles and are willing to spend more on products/services that emotionally engage them, such as food, wine, travel, cars, toys, sporting goods, etc. This demand for lifestyle underscores the need for marketers to tailor their marketing plans so as to reach the desired consumers across the multi-communication channels. Moreover, marketers need to continuously measure and optimize their marketing programs to acquire, retain and increase loyalty and lifetime value of the empowered, perpetual fast-moving consumers.

Case Study 1: A multinational consumer goods manufacturer wants to introduce a new type of toilet paper to the Canadian market. Though this kind of product is used in every household, a consumer profile analysis of several existing products reveals that among thousands of factors that might be related to the consumers' purchase decision, one is particularly important, namely households with hockey fans and baseball fans. These two types of households have significantly different consumer behavior in toilet papers. Therefore, the analysis directs the sponsorship and advertisement of the relevant products for the right teams and through the right channels to the right audiences that reduce costs and increase revenue for the company.

Case study 2: A national charity organization wants to quantify donor's commitment and acquire more new donors. Combining survey responses with donor's lifestyle analysis reveals distinct attributes of different segments of donors, e.g. value focused and image oriented, and their lifestyles. With such knowledge the organization is able to address concerns and interests of different lifestyle clusters of donors with targeted communications. This has reduced significantly the attrition rate and increased response rate in their acquisition programs.

CanaCode Lifestyle Clusters

CanaCode is a two-tier lifestyle cluster system at the 6-digit postal code level. The first tier consists of 17 distinct lifestyle clusters and is designed for high level description of consumer lifestyles. The second tier has 109 niches and reveals more patterns in the lifestyles. It is designed for assisting predictive analytics.

CanaCode identifies distinct Canadian lifestyles. Each lifestyle cluster consists of consumers with similar needs and characteristics that lead them to consume and respond in a more similar way to a particular campaign, product or service than the Canadian average.

The following table shows the distribution of households among the 17 lifestyle clusters in Canada, Ontario, QC and BC.

Name of CanaCode Clusters	% Household Canada	% Household Ontario	% Household BC	% Household QC
A: Affluents	3.21%	5.14%	2.22%	1.31%
B: Elite Professionals	6.55%	8.61%	6.31%	3.28%
C: Ethnic Cruisers	3.96%	6.83%	6.84%	0.65%
D: Nest Builders	5.79%	7.43%	6.11%	2.92%
E: Buy Me a New Home	11.32%	12.82%	10.08%	10.49%
F: Empty Nesters	5.65%	6.62%	7.96%	3.06%
G: Up the Ladder	16.62%	14.23%	17.08%	18.54%
H: High Trades	6.89%	6.09%	10.19%	7.09%
I: Urban Life in Small Town	4.09%	3.45%	6.78%	1.77%
J: Joyful Country	3.68%	3.32%	1.73%	4.66%
K: Rural Handymen	3.64%	2.09%	1.82%	5.01%
L: Comfortable Apartment Dwellers	11.24%	12.56%	8.30%	13.39%
M: Singles	5.18%	1.90%	3.68%	10.42%
N: New Canadians	3.19%	3.44%	4.02%	2.02%
O: Renters	3.39%	1.75%	3.07%	5.60%
P: One Parent Families	2.11%	1.00%	1.74%	3.50%
Q: Thrifty	3.49%	2.72%	2.08%	6.30%

Data Sources

The input data for developing CanaCode lifestyle clusters were extracted from the following data products of the year 2015.

1. **SuperDemographics:** Current year estimates of population statistics including age, dwelling, household, family, education, immigration, ethnicity, minority, home language, mother tongue, knowledge of other languages, labour force, employment, occupation, mobility, income and religion. It has over 2,000 variables.

2. **Household Spending Patterns:** Estimated dollar amount of household annual spending on food, clothing, shelter, transportation, household operation, household furnishings, art and antiques, equipment, health care, personal care, recreation, reading materials, education, tobacco products and alcoholic beverages, financial services and insurance, gifts and contributions as well as RRSP, etc. It has over 500 variables.

3. **Estimates and Projections**: Current year population and income estimates, and projections of population by age group, household and family for the year 2020 and 2025. It has over 350 variables.

4. **Daytime Market**: Current year Population statistics of daytime, night time, working, student and stay-at-home including 0-14 and 65+ age groups, and working at home population. It has 25 variables.

5. **Business Patterns**: Current year estimates of the business statistics, e.g., number of business establishments, supermarkets, department stores, pharmacies. It has over 100 variables.

6. **Geographic Patterns**: Urban/rural indicators, proximity to parks, shopping malls, pharmacies, community centers and schools.

7. **Consumer Purchase Behavioral, Product Usage, Lifestyle and Psychographic Patterns:** This data product is based on the Return-To-Sample (RTS) Survey from Numeris, formerly Bureau of Broadcast and Measurement (BBM). It has over 8,000 variables including consumer food consumption, leisure activities, psychographic patterns, beverage consumption, snack consumption, restaurant visits, shopping patterns, travelling patterns purchasing, spending and loyalty patterns, household telecom usage patterns, energy conservation patterns, financial patterns, home improvement patterns, consumer health care patterns and work patterns.

   The RTS survey is conducted twice a year and the sample size is over 100,000. The RTS survey data has been widely used by Canadian media operators, agencies and advertisers. We have been partnering with NUMERIS/BBM

Canada for 10 years and have in total over a million responders in our database. They are well stratified by geography and demographics. They represent Canadian consumers across the country.

In addition to the Numeris/BBM RTS data we sourced the data from Statistics Canada*, Citizenship and Immigration Canada, Health Canada, Industry Canada, Provincial Ministries of Health, Canada Postal Corporation, Canadian Mortgage and Housing Corporation and Manifold proprietary research, consumer surveys and databases.

A holistic view of consumers
We fused the geographic, demographic, household spending, daytime population, population projection, consumer consumption and purchase behavioral, attitude and psychographic data at the 6-digit postal code level and created a 360-degree view of consumers with over 10,000 variables.

Methodology for Creating CanaCode Lifestyle Clusters

Multi-Scaling and Dimension Reduction
Using a multi-scale normalization procedure we transform all variables into comparable scales. Thereafter, we performed a systematic principal component analysis of the input database, where we considered the categories, e.g., demographics, household spending, consumer behavior and psychographics separately. This enables us to reduce dimension of the data significantly on the one hand, and on the other hand to ensure each category is represented properly of the input data. We selected the top principal components as input variables for clustering analysis. For example, 5 principal components from the demographic data, 3 from spending data and 10 from consumer behavior data. Each combination may lead to different set of clusters. The final choice of principal components was determined in an optimization procedure based the variance they represent, their relative strength to the other categories and profile analysis of the preliminary clusters. In this way we consolidated the input data into a low dimensional dataset consisting of highly predictive factors.

In order to reflect the Canadian multi-cultural landscape, we considered the following categories of demographic data: consumers' ethnic origins, languages they speak at home, their mother tongues, other language they speak, when they immigrated to Canada, generational status in Canada, their belief and religion. Each component is represented by a proper number of principal components.

We also divided the consumers' psychographics into the following categories: attitude toward advertising, health consciousness, opinion about new products, brand loyalty, cost sensitivity, social network, lifestyle, social activity, family and self-esteem. Again each component is represented by a proper number of principal components.

* No confidential information about an individual, family, household, organization or business has been obtained from Statistics Canada.

Similarly, we divided spending data, consumer behavioral data into categories and performed the dimension reduction separately. The top principal components were selected to form the input data for clustering analysis. In fact, by weighting different categories with number of principal components, we can custom the cluster by industry, e.g., retail, insurance, finance, package goods, …

A Hybrid Clustering Process

To cluster the 6-digit postal codes into homogeneous lifestyle groups, we combined a hierarchical clustering technique with fuzzy logics and incorporated them into an adaptive K-means method. The fuzzy clustering allows an object can to belong to one or more clusters with probabilities, which reflects the behavior of many consumers, particularly those similar to the average Canadians.

We started with a set of principal components and ran a fuzzy K-means method for a series of number of clusters iteratively. At each step the initial seeds of clusters were carefully selected so that outliers and small clusters are separated in the clustering process.

To determine the optimal number of clusters, we examined the local peaks of the pseudo F statistic: ratio of between cluster variance and within cluster variance. We also looked at peaks of the cubic clustering criterion (CCC) and developed our own statistical measures. For details we refer to our publications[1,2].

We validated the clusters with both in-time and out-time surveys and historical Census data. Our object was to ensure that CanaCode lifestyle cluster can differentiate Canadian neighborhoods substantially and generate a consistent and significant lift when applied to target marketing programs. To this end, we also incorporated the geographic information and neighborhood knowledge into our clustering process.

We maintain ongoing research projects on developing efficient and effective clustering algorithms[3] with researchers at York University. Our research project has been endorsed by the Natural Sciences and Engineering Research Council of Canada (NSERC).

Contact: Dr. Zhen Mei or Thomas Ding
Manifold Data Mining Inc.
220 Duncan Mill Road, Suite 519
Toronto, ON M3B 3J5
T: 416-760-8828
F: 416-760-8826
E: zhen@manifolddatamining.com

[1] H. Sun, S. Wang, and Z. Mei: A fuzzy clustering based algorithm for feature selection. Machine Learning and Cybernetics, 2002. Page(s): 1993 - 1998 vol.4 4-5 Nov. 2002

[2] H. Sun, M. Sun and Z. Mei, “Feature Selection via Fuzzy Clustering”, in Proceedings of IEEE ICMLC 2006, pp. 1400-1405. (EI)

[3] G. Gan, C. Ma and J. Wu: Data Clustering: Theory, Algorithms and Applications. SIAM and ASA, 2007.