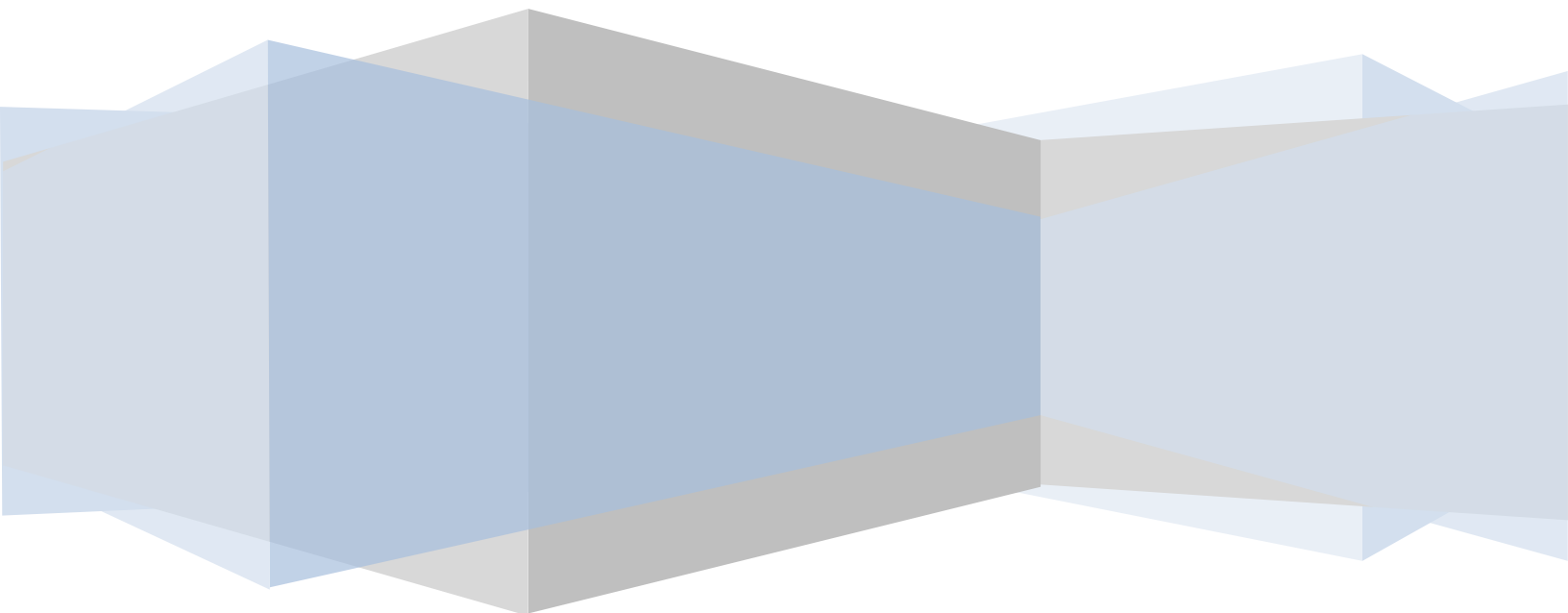


PSYTE HD Canada

Methodology Statement



PSYTE HD Canada: a high definition view of Canadian Neighbourhoods

Introduction

The development of PSYTE HD Canada builds on a long tradition within Pitney Bowes Business Insight (formerly Troy, NY-based MapInfo Corporation and Toronto-based Compusearch Micromarketing Data & Systems). The PBBI team that built PSYTE HD has a strong track record of pioneering the evolving methodologies used to develop geo-demographic clustering systems for both Canada and the United States. In fact, the primary methodologist on the PBBI team was directly involved with the development of the first and subsequent PSYTE Canada products from Compusearch. While the history of clustering small area units of analysis for marketing and site location applications goes back at least to the 1970's, methodologies and data sources continue to evolve along with the computational power available. For example, during the 1990's the use of neural network techniques became practical for the development of PSYTE US Advantage. Furthermore, multi-stage techniques to generate atoms and clusters were used to develop PSYTE Canada Advantage following the release of Census 2001 data¹. The approach described here basically makes optimal use of traditional objectives and strategies while also taking advantage of advances in statistical procedures and computing power.

As stated many times, clustering is part Art as well as part Science. The Art derives from the exercise of deciding what variables are selected as that involves crafting a set of implicit weights that will shape the final clusters. Judgment is also required in evaluating available statistical routines and designing the sequence of steps that applies those routines to actual data to produce a set of clusters that meet design criteria. The Science consists of the decision-making process used within the bounds of the data and technique to know when a sound segmentation system has been created. In the end, in order to "let the data speak for itself," the system must be statistically rigorous even as it allows for an artistic component.

Contemporary clustering methods must use robust statistical methods to capture and measure the qualities of reliability, validity, accuracy and precision. While the application of the scientific method may take place within a controlled environment, scientists still need to decide what to put under the microscope. The social-cultural-economy is too complex to simply put "everything" into a classifier. The result would be "mush" with no way of knowing which result is better than the others. While an "objective function" is not easily articulated in this discipline, it is possible to point the classifier in the right direction by selecting and distilling the data down to what may be considered essential, practical, and indeed, intriguing. In the end, a proper geo-demographic cluster system optimizes the nexus of consumer lifestyle and location.

¹ Exter, T. and Mosley, I. (2005). *Geodemographic Segmentation: PSYTE Canada and PSYTE Quebec*. Paper presented at the XXV International Population Conference of the IUSSP, Tours, France.



PSYTE HD Canada: a high definition view of Canadian Neighbourhoods

Data Selection

Several categories of data can go into the clustering process, but they must be selected with care and with an eye to their statistical validity. In the case of PSYTE HD Canada (Census 2006-based) as with PSYTE Canada Advantage (Census 2001-based), census data at the dissemination area level were used in the first stage of clustering. Dissemination Areas (DA's) are the lowest level for which reliable census data are published. No other small area unit of analysis contains the wealth of data available from the census for DA's. Since literally hundreds of variables are available, however, the key to their initial selection is their relationship to consumer behaviour and their ability to characterize neighbourhoods. Key groups of variables were selected, including life stage variables (for example: age, marital status, household type and size, and presence of children); variables related to education and employment (for example: educational attainment, occupation, industry of employment); variables related to immigration and cultural background (for example: ancestry, country of birth, year of immigration, and language spoken at home); socio-economic status variables (for example: household income, dwelling value, and home ownership); and, variables describing the predominant type of housing unit in the neighbourhood (for example: owner occupancy, rental occupancy, single-detached, walk-up, or high rise). Updated variables such as current estimates and projections as well as household wealth variables can be used to describe final clusters but are generally not used in the initial phase.

The best tool for data reduction – that is, reducing a large number of variables down to a manageable set – is Principal Components Analysis (PCA) as it provides an excellent overview of the general structure of variance within the data and what variables correlate with these variance components. The data for initially clustering are thus reduced to a set of relatively uncorrelated themes that essentially have the ability to distinguish each initial set of neighbourhoods in a robust way.

The Clustering Process

The clustering process starts with the grouping of DA's into 300 mini-clusters or “atoms” based on key demographic themes or dimensions such as: age, dwelling type, family structure, education, employment characteristics, ethnicity, income, and mode of travel to work. The initial phase involves a sophisticated clustering algorithm which maximizes the similarities of DA's within each atom while simultaneously maximizing the differences among DA's across the atoms. Thus, the atoms represent the fundamental geo-demographic structure of Canada whereby each atom contains DA's that are similar with respect to principal characteristics of demography and areal situation. The atoms become the key building blocks for the final geo-demographic clusters.

The second stage of the clustering process was executed on the atoms. This stage used hierarchical clustering techniques to group the 300 atoms into 59 clusters and one “unclassified” cluster which includes essentially unpopulated DA's. Additional data inputs for the second stage are significant. PBBi has established a working relationship with Toronto-based ICOM, Inc. to utilize an extensive set of household-level inputs in the second stage of the clustering process. Essentially, ICOM provides an extremely detailed dataset on the interests and consumer behavioural activities of several million



PSYTE HD Canada: a high definition view of Canadian Neighbourhoods

Canadian households based on a long-established and ongoing omnibus survey. Also available are a set of household-level “cluster codes” from ICOM’s own segmentation product. Together, these inputs represent an unprecedented level of detail and behaviour-based data that provided a significant enhancement to the PSYTE HD Canada cluster system. Additionally, as in prior generations of PSYTE Canada, extensive auto registration data from Polk Canada were used to discriminate among the atoms in the agglomeration to cluster stage. The auto data is an input that reports direct consumer behaviour – car purchases – which are noteworthy for their reflections of significant consumer interests, values, and lifestyles.

At this point it is important to assess the clustering outcome relative to its ultimate application. Dissemination areas can be grouped in many ways for many purposes. However, the key for PSYTE HD Canada as for its predecessors given its intended use in marketing, site location research, and the geo-demography of consumer markets is its ability to “discriminate” one neighbourhood market from another along important dimensions, that is, dimensions that drive consumer behaviour and business decisions. Several tests using actual consumer behaviour and sales data are used to confirm that the ultimate clusters provide statistically significant “lift” in the sense of increased ability to discriminate behaviours across a set of neighbourhoods.

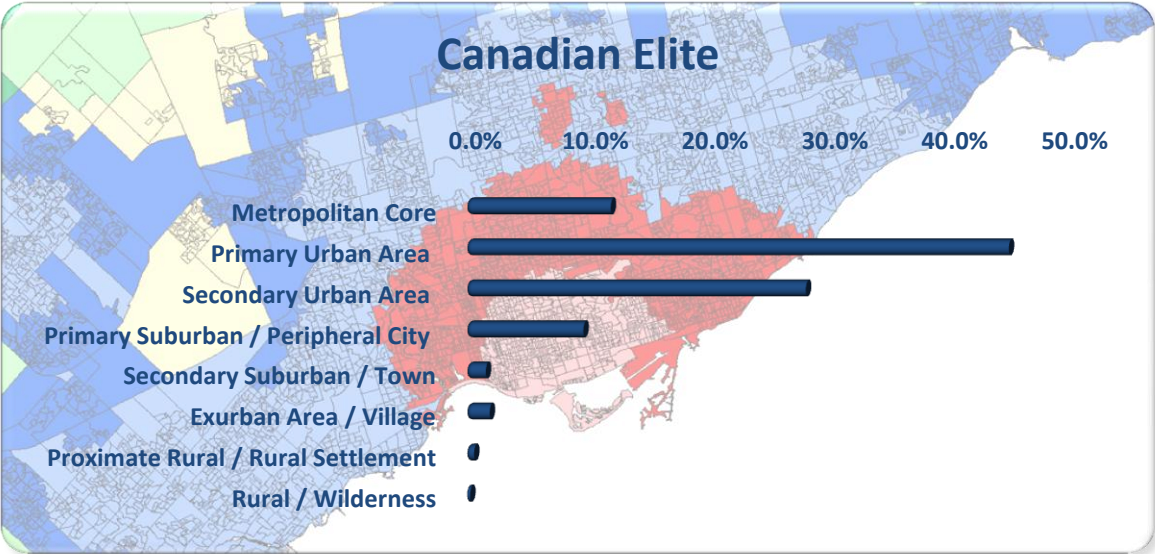
Settlement Context

Settlement context refers to broad categories of the physical/socio-economic landscape that describe an important dimension of human settlements. One of the objectives in the development of PSYTE HD Canada was to capture the current reality as well as the dynamic of change in the Canadian context. After extensive testing of several spatial analytic routines, a method emerged that basically involves a combination of two dimensions. The first is a scale of urbanity or urbanization that measures relative levels of industrialization, commercial development, road type and street density, housing types, and distance from major urban centres. The second dimension measures the intensity of human settlements in terms of density classes. The density measure is continuous in the sense that it controls for the irregular boundaries of census dissemination areas. The combination of these two dimensions provides a single settlement context “assignment” for each dissemination area.



PSYTE HD Canada: a high definition view of Canadian Neighbourhoods

There are eight settlement context levels as noted in the following chart, which shows the distribution of “Canadian Elite” households across the settlement context schema:



One important implication of the greater analytical attention paid to settlement context is the allowance for a distribution of contexts within a single PSYTE HD Canada cluster. In the past, each cluster was given a single settlement context assignment. With PSYTE HD Canada, in contrast, the set of dissemination areas belonging to each cluster may be distributed across several settlement contexts. One key advantage of this approach is that consumer households in each cluster can be further divided by one or more settlement context levels to provide more focus to the end-user analysis.

A second advantage of this approach to settlement context is the ability of analysts to better visualize the dynamic nature of human settlements. Just as the earliest suburbs arose from the city environs of the last century, so new suburbs, exurbs, and mid-sized cities grow and extend their influence within a micro-region and beyond. Over time, as commuting patterns and employment hubs form more complex networks, the nature and extent of commercial activity evolves. The settlement context descriptions below provide a sense of this dynamic of change across the different areas as well as the “view from the front porch” of a typical consumer’s home.

- ✓ Metropolitan Core – This settlement context forms the heart of the largest metropolitan areas. By definition a metropolitan area encompasses both cities and suburbs. The metropolitan core, therefore, is not just an “urban core” but the centre of a broad, indeed international, regional economy. That is, its significance extends well beyond the immediate, large city. Daytime populations swell due to in-commuting.



PSYTE HD Canada: a high definition view of Canadian Neighbourhoods

- ✓ Primary Urban Area – This settlement context forms either the first ring of intense urbanization around a Metropolitan Core or is, in fact, a primary urban area in its own right. Relatively high density living, access to mass transit, older street networks, and significant commercial hubs characterize this settlement type. Daytime populations increase but can be offset by commuting to the Metropolitan Core.
- ✓ Secondary Urban Area – This settlement context, perhaps once called “urban fringe,” is not really “fringe” to urban living. New high rise apartments, condominiums, as well as detached housing developments characterize this settlement type. Often, to accommodate the extending urban area, commercial hubs and nearby employment centres evolve and grow over time. New commuting patterns develop that are both multi-directional and multi-modal.
- ✓ Primary Suburban / Peripheral City – This settlement context contains older suburban areas with long histories of out-commuting to major employment centres. However, also found within these areas are cities that contain their own economic and industrial base. As some employers have sought or needed more land for their developing operations, these areas have established a momentum of their own for economic development.
- ✓ Secondary Suburban / Town – This settlement context contains newer suburban developments driven by the extended reach of the Metropolitan Core or Primary Urban Area. Initial developments are typically, primarily residential and commercial. As “suburban,” however, these areas may also experience more diversified employment growth. Towns that were historically “stand-alone” villages have been encompassed or now evolve within a more “suburban” context. This process may occur near a major metropolitan area or within a smaller regional context.
- ✓ Exurban Area / Village – This settlement context represents the furthest reaches of feasible commuting to a Metropolitan Core or Primary Urban Area. Towns in these areas may have a significant agricultural history but now find themselves encompassed by new, exurban developments and commuter populations. These areas represent a mix of old and new. They may be challenged to preserve traditions even as they are attracted to the opportunities of change.
- ✓ Proximate Rural / Rural Settlement – The term “rural” has generally received less attention than “urban,” despite significant differences across the rural landscape. Proximate Rural refers to those areas near suburban and exurban developments that may, in fact, have the potential for future residential development. “The view from the front porch” is often bifurcated between the serene agricultural vista and the highways and cul-de-sacs of the evolving suburbs and exurbs. Rural Settlements within this context are those smaller towns, villages, and even “crossroads” that lie beyond the reach of the metro commuters but still contain some commercial and primary industry activity.
- ✓ Rural / Wilderness – Extensive areas, some touched by agriculture, forestry, fishing, or mining are included here. Other areas remain largely untouched by such activities and may perhaps best remain so.



PSYTE HD Canada: a high definition view of Canadian Neighbourhoods

Summary and Conclusion

PSYTE HD Canada is fundamentally a geo-demographic cluster system. Geo-demographic cluster systems, in contrast to household-based systems or hybrid systems, use the smallest area for which census data are published – in Canada, the census dissemination area – as a de facto neighbourhood base. Nevertheless, diversity within dissemination areas as “neighbourhoods” exists and will likely increase. Users will find, however, that PSYTE HD Canada captures much of that diversity and still provides a sound basis for sensible market segmentation strategies. The methodological approach described here generates a stable and statistically reliable set of cluster assignments. The neighbourhood clusters of PSYTE HD Canada provide a unique and identifiable description of geographic target markets that permits marketers and site location specialists to use the system effectively. PSYTE HD Canada provides a multidimensional framework that allows analysts to capture the complexity of Canadian consumer culture without having to manipulate literally hundreds of census variables, enabling analysts to locate their prime markets “on the ground.”

