ENVIRONICS
ANALYTICS

TECHNICAL DOCUMENTATION
PRIZM 2021
APRIL 2021

Creating the PRIZM segmentation system involved more than a year of planning and development. The system categorizes every Canadian neighbourhood and postal code into one of 67 distinct lifestyle segments based on the characteristics of households. PRIZM examined five years of neighbourhood formation, change and, in numerous cases, stability to better understand the demographic, socioeconomic and behavioural characteristics of consumers.

Every edition of PRIZM is developed by a team of experts familiar with the demographics and geography of Canada at both the city and regional levels. The system is built from the ground up using authoritative data from recognized suppliers like Statistics Canada, Canada Post, Canada Revenue Agency, Equifax, TomTom, Environics Research and others.

The vast majority of targeting and marketing in Canada is still done based on age, sex and income. Examples of typical target groups using these characteristics are:

- Women aged 18-34 for fashion (clothing)
- Baby Boomers aged 55-70 who travel frequently
- Households with income over $200,000 for prestige vehicles

Geodemographic systems permit refinement of this typical targeting method. Small neighbourhoods are assigned to household segments based on similarities of demographic attributes and general lifestyle behaviour. This approach is a well-established method for segmenting and identifying target groups for numerous products and industries.

DATA

The primary foundation for any geodemographic system is high-quality demographic data. Determining which variables to include in the final clustering is a matter of science and art. The science element captures variables that are significant in their ability to differentiate neighbourhoods in ways that are important for marketing. The art component falls into two categories: a) understanding that users of the final system expect specific variables to be included, and b) determining the final selection of weights for each variable. When going to market, customer expectation and demand also play an important role. At the same time, there is a need to respect and adhere to thorough scientific methods. Successfully balancing these many requirements makes a product credible and effective.

The number of variables in the final clustering should be kept to a minimum (for scientific parsimony) yet, at the same time, must include all of the important demographic characteristics. There is no known acceptable number of variables to include in clustering; it is up to the experienced analyst to select the optimal set of variables and experiment with weights to find the optimal mix.

Creating PRIZM begins with CensusPlus, a database derived from Statistics Canada's census, which has been enhanced by our modellers to fill in missing values. The core data are available at the dissemination area (DA) level, the smallest unit of geography for which any significant demographic and socioeconomic data are released. There are 56,590 DAs in Canada. CensusPlus contains about 850 variables for each of these DAs covering numerous themes.

We take CensusPlus and combine it with the latest vintage of DemoStats, a proprietary database that reflects our estimates of current-year demographics and socioeconomic characteristics at the neighbourhood

level[1], to select the final demographic and socioeconomic variables. The variables are categorized in the following 18 themes:

Age	Dwelling Value
Household Size	Mother Tongue/Home Language
Marital Status	Ethnic Origin
Households with Children	Visible Minority
Migration	Aboriginal Identity
Immigration	Education
Dwelling Type	Labour Force/Occupation/Work Place
Dwelling Tenure	Mode of Transport
Dwelling Period of Construction	Income

In addition to the core demographic and socioeconomic data, other data are used as basic ingredients of PRIZM. Data describing the settlement context—the geographic location of neighbourhoods—are fundamental to understanding where the resulting segments are situated geographically. Are the segments predominantly found in large urban centres, small suburban towns or sparse rural communities? Proximity to major retail centres is another measure we use to classify established urban cores differently from suburban, town and rural neighbourhoods.

Another important source of input data comes from our SocialValues database, which is derived from data supplied by our sister company, Environics Research. Every year, Environics Research conducts a nationwide survey that measures human motivation and social relations, employing advanced techniques to understand the mindset of Canadians. From these data, they create "social constructs" that identify common trends in views and attitudes among Canadians.

In addition, we use aggregated, anonymized small-area credit data from Equifax Canada. These data measure credit worthiness, credit usage and credit default rates. Additional variables capturing the financial theme are also included in creating PRIZM.

From these data sources, we select more than 80 variables, including at least one variable for each of the above themes.

PROCESS

At the outset, our analysts determined that the PRIZM segmentation system was going to consist of three different products. The segments had to be available for populated residential DAs and postal codes, plus they had to operate at two levels: One with a conventional number of segments between 60 and 70 and a more detailed version with 150 segments, which we call Delta. This approach was divided into three stages.

STAGE ONE

The initial phase of stage one involved creating a set of variables that captured the settlement context of the DAs. Settlement context is a scaled measure of urbanity, from the dense urban core of large cities to the most

[1] For more information on the development of DemoStats 2020, refer to the DemoStats technical document available at environicsanalytics.com.

sparse, uninhabited rural parts of the countryside. These are key variables that serve in the initial segmentation process.

The next phase involved assessing and selecting CensusPlus and DemoStats variables from the more than 1,400 variables available for the creation of the atoms. We selected variables that we know from experience to be significant for differentiation among the DAs.

In the final step we selected the clustering algorithm for creating the atoms. Numerous algorithms are available for cluster analysis and the method used is critical to the success of PRIZM. Based on our research, we selected the K-medians algorithm, an iterative partitioning approach commonly used for these types of applications. The greatest strength of this algorithm is its ability to find similar patterns that maximize within-segment uniformity while differentiating between all the identified segments. Additionally, the resulting segments are not as influenced by extreme values ("outliers") as many other traditional methods.

To determine the best segmentation solution, we tested thousands of weighted data combinations. Every run was informed by the previous one, and with each subsequent run, adjustments to variables and variable weights were made. The runs that offered the greatest differentiation between segments were examined and systematically tested. The best solution offered the greatest discrimination of segments against actual consumer behaviour (more on this later). Finally, we produced what we consider to be an optimal 150-atom segment solution.

STAGE TWO

The focus of this stage was to link SocialValues data to the 150 atoms, and then aggregate them to create a system between 60 and 70 segments. We resolved to look for fewer than 70 segments, to make the system more manageable and maintain the greatest differentiation between segments. Using our estimated demographic and socioeconomic data, along with settlement context, financial credit data and SocialValues, the atoms from the DA system were aggregated using several clustering algorithms.

Our analysts identified many segment solutions by applying different weights to a variety of variable subsets. In reviewing the solutions that were automatically generated through the clustering processes, 67 segments offered the greatest variety in neighbourhood and SocialValues types, while meeting our minimum cluster size as represented by the number of Canadian households assigned to it (0.50 percent of Canadian households). The 150 atoms nest perfectly within the final 67 segments.

TESTING SEGMENTS

We tested each solution using a variety of data supplied by several of our data partners as well as with an analysis of key products and services. From the thousands of clustering runs, three solutions emerged as leading contenders.

Key survey providers for the testing exercise were:

- Vividata
- Numeris
- AskingCanadians™ (Social Media, Mobile and eShopper surveys)
- IHS Markit™
- Select client data from different industries

More than 3,000 variables were selected from the various surveys covering such topics as category and product usage, media preferences, leisure activities and attitudes. The survey data were aggregated to the 67 segments and each variable was compared to its Canadian average. A review of each segment's demographics, socioeconomics, settlement context and behavioural survey variables served as the method for analyzing and comparing the numerous cluster solutions.

These solutions were tested against one another to help identify the single best segmentation system solution. In reviewing the solutions, we examined the following for each of the 67 segments:

1) Demographic reports with more than 400 variables summarized at the segment level showing percent of a segment having the attribute and an index showing its relation to the Canadian average.
2) Geographic maps showing the segment distribution and whether the segment was concentrated in a few markets or dispersed across the country.
3) Survey reports with selected variables indexed against Canada for each segment on a large selection of category, product, behavioural and attitudinal variables.

STAGE THREE

With the 67 segments finalized at the DA-neighbourhood level, the next task was to assign all residential postal codes to the final solution of 150 atoms and its 67 segments. This stage involved combining DemoStats data with the Equifax credit data, all at the postal code level.

A set of demographic and socioeconomic variables were selected from the more than 500 available in DemoStats. Added into the mix were the settlement context data that were assigned to all postal codes (based on the DA they fall within) and a small set of variables from the Equifax data. This information was assembled for the complete roster of residential postal codes.

We then created 150 cluster centroids, the statistics reflecting multi-dimensional segment profiles—the basic building blocks of segments—using the atoms created at the DA-neighbourhood level. In addition to these DA-level centroids, we developed a new version using only postal code-level data. Several versions of the centroids were created and tested to ensure they captured the fundamental characteristics that describe each atom at the DA level. Did family-based segments have the correct ages of children? Were culturally diverse segments showing high concentrations of the relevant groups? Were urban segments found in urban areas and rural segments in rural areas?

We selected the centroid that depicted the 150 atoms the best. Following this process, all postal codes were assigned to the closest atom based on statistical proximity to guarantee the optimal assignment for all selected variables.

In addition, there were a few final manual adjustments to the automated cluster solutions. These adjustments were made to preserve, as much as possible, the settlement context structure of Urban, Urban Fringe, Suburban, Town and Rural. Other important considerations in the clean-up phase ensured that the wealthiest segments were captured appropriately, that key culturally diverse segments were identified correctly as speaking dialects of Chinese and South Asian languages, and that francophone segments had a minimum number of French-speaking populations.

RESULT

PRIZM consists of a whole new set of geodemographic segments for Canada reflecting the most recent and reliable data. There are 67 segments, made up of 150 atoms, which capture all dimensions of the Canadian landscape. PRIZM is available for both DAs and postal codes.

SOCIOECONOMIC STATUS INDICATOR (SESI)

With the final segmentation system created, we had to decide how to number and rank the segments. A proprietary score was developed to characterize each segment using a Socioeconomic Status Indicator (SESI). This SESI score reflects a variety of factors such as average household income, discretionary income, education attainment, the value of private dwellings, average net worth and household size.

As a result, a blue-collar, high school-educated segment whose residents earn above-average incomes may rank lower on the SESI ladder than an educated, up-and-coming younger segment whose residents have average household incomes. A segment with an older population, many of whom are on fixed incomes, may rise in the ranking if their net worth is significant. And a segment earning $120,000 on average will rise or fall in the ranking, depending on whether the household is composed of dual-income couples or families with several young children.

The 67 segments have been ranked from 01 to 67 on the SESI scale, with 01 classified as the highest. Because this ranking reflects more than income alone, most of the segments have a SESI score that is different from their average household income ranking.

SOCIAL AND LIFESTAGE GROUPS

The 67 PRIZM segments were assigned to one of 20 Social Groups and 8 Lifestage Groups. The Social Groups consider the urban-rural context, home language (English, French and non-official), affluence, family status, age of maintainer and ethnicity. Each segment was assigned to one, and only one, Social Group. The Social Groups reflect various groupings, patterns and trends. A critical issue concerned the urban-rural dimension, which is neither linear nor one-dimensional. Each segment was assigned to one of five settlement types to form the Social Groups: Urban, Urban Fringe, Suburban, Town or Rural. Urban Fringe segments reflect once-suburban areas that, over the last 30 years, have been absorbed by urban sprawl. In general, urban segments are found in large- and medium-sized cities. Suburban segments tend to consist of communities located on the outskirts of cities and can often be found in the core neighbourhoods of smaller cities and larger towns. Town neighbourhoods are found in smaller towns across the country. Rural neighbourhoods reflect areas that are smaller than towns and include very small towns, villages, hamlets, and rural farms and isolated areas.

The final segmentation solution features many francophone-based segments, a variety of culturally diverse segments and many segments that represent important combinations of age, lifestage and family status—from young singles living on their own up to widowed seniors in apartments. These were essential inputs into the creation of the Social Groups. The ranking of Social Groups is based on average income (not a SESI ranking). Groups have a letter and number combination. The letters U, F, S, T and R stand for Urban, Urban Fringe, Suburban, Town or Rural, while the numbers refer to income, with 1 indicating the highest average income for the group and 7 the lowest.

The Lifestage Groups categorize household composition according to the presence of singles, couples and families. The major grouping divides the 67 segments into Young, Family and Mature. These groups are then further subdivided by analyzing the commonality among the segments. The Young group is divided into three subgroups according to the presence of singles, couples or starter families. Families are split into three sets based on the age of children: the very young, tweens, teens and twenty-somethings. The Mature group is divided into two based on the age of maintainers and the presence of children at home.

ANNUAL UPDATE

Each year, when a new edition of DemoStats is completed, PRIZM is updated. The update reflects the most recent estimates of Canada's demographic and socioeconomic characteristics, along with updated Social Values data and financial credit data from Equifax. Both DA and postal code PRIZM assignments are reassessed and updated.

We recognize that changes and shifts occur in Canada's demographic landscape on a daily basis. However, we are only concerned with large systematic change that can be measured. During the development of DemoStats, we examined the themes that are most important, such as age structure, income distribution, housing stock and tenure, diverse populations, family structure and the continuing urbanization of Canada.

For the 2021 update, our objective was to retain existing PRIZM segment assignments, except in cases where solid empirical evidence indicated a neighbourhood has changed significantly. We wrote routines to look at data compiled for each postal code and DA, calculating the "fit" to all segments to confirm any significant change. For those areas whose previous year's segment assignment was no longer the best fit, we changed the assignment. And, of course, new postal codes were assigned to the segment that was the best fit based on measuring distance to all segments.

In the 2021 update, 22 percent of postal codes underwent a change in their PRIZM assignment—a shift that affected 13 percent of households. The three most stable segments—that is, those that experienced the least geographic change—were Keep on Trucking(37), Turbo Burbs (4) and Wealthy & Wise (2). The three segments that experienced the greatest geographic change were Came From Away (61), Les Énerjeunes (40) and Enclaves Multiethniques (55). Meanwhile, the three segments that experienced the highest growth in households were Banlieues Tranquilles (42), Country Traditions (26) and Wealthy & Wise (2); the three segments that declined the most in households were Came From Away (61), Country & Western (50) and Stressed in Suburbia (38). But it is important to note that, even with these changes, PRIZM can be described as a very stable segmentation system, reflecting an expected level of geodemographic change over the last year.