Consumer Intelligence Group

Data Development Methodology

Overview:

CiG's geo-demographic projected data is developed to address and transform the outdated Canadian data development and licensing model for marketers. We focus on providing current, robust and diverse data for defining, understanding and precisely targeting audiences at scale.

The only true way of having confidence in data is to try it in your applications, models and audience definitions. CiG has partnered with Snowflake Marketplace®, a leading cloud-based data warehousing platform, to provide trial access to a full complement of intelligentVIEW's data themes.

❯ Contact us to get a data sample to test

Want to see the data in action in real-time? Contact us to set-up a 30 min demo of our intelligentVIEW platform using your customer data.

❯ Request a custom demo

Approach:

CiG's unique approach to developing a data ecosystem provides a solution to:

- Data Latency – methodology allows for the delivery of quarterly and semi-annual data updates through intelligentVIEW, our cloud-based insights platform.
- Resource Intensive Data Processing – leverage the rapid growth of statistical and modelling software and the increase in processing speeds to automate the process.
- Restrictive Usage – Data is included in all of CiG's plans, with flexibility to download and use data, as needed, removing the prohibitive usage allowances found in the traditional data licensing model.
- Cost – Data included in all SaaS pricing plans, with add-ons available to download data on demand, pay for what you need.
- Coverage – Data are developed with the latest geographic files updated quarterly, providing maximum access to addressable market audiences where they live.

CiG is transforming the way that companies build, maintain and validate their first, second, and third-party data. We deliver this through a combination of:

- An ever expanding list of new, privacy compliant data sources becoming available in the Canadian marketplace.
- New tools and technology that allow for the automation of data processing, decreasing the reliance on existing traditional data providers and costly consultants.
- Distribution through the cloud.

CiG has automated and modernized the access to, and usage of, consumer and market data to drive insights for Canadian businesses.

Some of Our Data Sources:

- **Statistics Canada Data** (under licence) – This data includes Census, Wealth and Assets and Household Expenditures data that is supplied and continually updated at the various levels of Statistics Canada geographies for census and intercensal periods.
- **Canada Post Householder Plus (HHP) Data** (under licence) – Leveraged to facilitate unaddressed advertising mail, the HHP contains the most up to date list of active Canadian postal codes updated monthly. Up to date addressable postal codes and current vintage Canada Post Smart Mail monthly data feeds, ensures maximum access to highly targetable audiences in a timely fashion.
- **Mover Listing Data** – Publicly available real estate listings data used for micro economic financial indicators and updated monthly. Real estate values are a strong indication and highly predictive of overall household wealth characteristics.
- **Numeris RTS Survey Data** (under licence) – Collected semi-annually, the Numeris RTS is a national research study containing demographics, media preferences and consumer brand and product behaviours across a survey sample of approximately 45,000 Canadian households.
- **Postal Code Conversion File** – Built from a variety of publicly available sources, CiG's PCCF file is an up to date and accurate link between Canada Post postal codes and other geographies including the Canadian Census geographies with a monthly update schedule.

Our Methodology:

CiG's methodology begins with our unique approach that pools our data sources into an ecosystem where sources are used to cross-validate and contribute to each other's robustness.

Treating our data as a part of an ecosystem, rather than independent tables or data products, means that statistical relationships are exploited, bases (Populations, etc.) are consistent, and each data source is a valuable contributor to the maintenance of the many.

Maintaining the CiG data ecosystem is solving challenges with data sources released at a different frequency with various lag times from monthly, to semi-annually, annually and, as with the Canadian Census, every five to seven years.

With the CiG data ecosystem approach, marketers and analysts see the benefit of the most current release of source data soon after it is released, rather than waiting to gather all the data sources over a year, collating them and spending months projecting them. Without the ecosystem, marketers see data projections based on source data that is several releases old.

Historically CiG's experienced data team has focussed on the use and creation of postal code level projections. With the development of the new CiG core data assets and data ecosystem, the team started with the basics.

CiG designed processes to bring older releases of data, such as the 2011 and 2016 Census, up to date, project them to six-digit postal code, and then leverage those patterns to project other attributes from samples (i.e. Numeris RTS study) to populations.

It is well known that Census population does not reflect actual population. By leveraging data sources including historical real estate patterns, housing starts, and Canada Post monthly data releases, we can accurately adjust population estimates at both the micro and macro geographic levels.

Statistics Canada releases adjusted quarterly population estimates. Click here to read more.

As per CIG's latest data on April 1, 2022, when aggregating all postal code populations reveals that Canada's population is estimated at **38,722,936 covering 15,926,330 households**.

	Q1 2021	Q2 2021	Q3 2021	Q4 2021	Q1 2022
Geography	**Persons**				
	38,068,872	38,153,211	38,246,108	38,469,348	38,722,936

Keeping in mind that the main purpose of intelligentVIEW data are to create, understand and precisely target audiences for marketing at scale, we need to accurately project the characteristics panel responders to the Canadian population.

Panel data, such as Numeris RTS and Vividata have research samples that are carefully collected and weighted to represent the Canadian population. We developed an approach that clusters matching geodemographic characteristics from postal code demographics to the weighting attributes collected on panel responders.

This approach effectively creates estimates at six-digit postal code level for the entire country, allowing marketers to profile customers, markets and prospects across thousands of consumer characteristics with a near 100 per cent match rate.

Accurate geographic allocation of data are integral not only to the execution of marketer programs, but the integration and cross validation of many data sources.

CiG has carefully built and maintains an accurate location database for postal codes and their many one/one to many relationships with Statistics Canada standard geographies. These files are often referred to as postal code conversion files. We have further enhanced our PCCF with a comprehensive urbanity classification with the following eleven classes:

- **MOSTLY DOWNTOWN APARTMENTS**
- **BIG CITY SUBURBS**
- **MOSTLY URBAN APARTMENTS**
- **COMMERCIAL/RESIDENTIAL**
- **URBAN RESIDENTIAL MIX**
- **DOWNTOWN RESIDENTIAL MIX**
- **RURAL**
- **COMMERCIAL**
- **MOSTLY URBAN HOMES**
- **SMALL CITY, TOWNS, AND SUBURBS**
- **MOSTLY DOWNTOWN HOMES**

CiG's data ecosystem is now used to expedite the production of estimates across all attributes at postal code level, add new attributes from sources and validate outputs against sources and geographies.

Some of the processes and techniques leveraged in the construction of the original base data sources and in the maintenance of the ecosystem are:

- Principle components analysis
- Correlation analysis
- Linear Regression
- Logistic Regression
- Other Non-linear regression techniques
- Clustering (mostly K-means / K-median)

A part of CiG's investment in developing core data assets and our data ecosystem, is the release of our proprietary segmentation system, intelligentSEGMENTS.

intelligentSEGMENTS are built using k-median and k-means techniques that first creates a base of 20 large clusters across core socio-economic characteristics at postal code level. Candidates within each of the large clusters are then clustered further into 5 clusters each with a focus on behavioural, ethnicity, housing type, education, and hundreds of other characteristics.

The resulting 100 clusters are re-ranked first on socioeconomics at the group level, and then again at the subgroup level. Small clusters are analyzed relative to their neighbouring clusters and collapsed resulting in a system of 94 robust segments.

Penetration and index values of thousands of additional variables are used to further describe and bring the 94 intelligentSEGMENTS to life. Automated processes identify attributes of significance and generate and compile data driven statements from these profiling tables to form initial segment descriptions.

CiG data team analysts, with the assistance of marketing, further enhance the written descriptions, generate core profile statistics and assemble imagery that portray the households and families that populate each intelligentSEGMENT.