ENVIRONICS ANALYTICS

Technical Documentation

General

MAY 2025

TRUSTe Data Certification Powered by TrustArc

environicsanalytics.com

AICPA SOC

To produce SocialValues, CannabisInsights Powered by Vividata, CommunityHealth, MoneyMatters Powered by Canadian Financial Monitor, and other behavioural databases, we use one of several types of modelling frameworks: PRIZM coefficients, K-Nearest Neighbours (K-NN), or Fusion.

Modelling Framework by Product

Product	Coefficient	K-NN	Fusion+KNN
SocialValues	X		
MoneyMatters	X		
ShopperChoice	X		
Opticks Mobile		X	
Opticks Social		X	
Opticks eShopper		X	
Opticks Numeris		X	
Opticks Media		X	
Opticks Vividata		X	
Opticks Vividata (Gender)		X	
CommunityHealth			X
GreenLiving			X
GivingBack			X
CommunityLife			X

For our coefficient-based datasets, we assign each record in the survey a PRIZM® segment. Depending on the survey, we may append PRIZM® at the dissemination area (DA) or postal code level. We summarize the PRIZM®-coded survey to the 67 segments. At the PRIZM level, we calculate empirical coefficients, which can either be a rate or an average. Rates are the total number of people who do the activity divided by the total universe for the activity. Averages are the total number of units consumed/spent/held divided by the total number of people who consume/spend/hold. The empirical coefficients are rolled out to geography using FSALDU PRIZM® assignments and by multiplying the coefficient for each variable against our DemoStats current-year universes. Common universes for our surveys are households, household population 12+, and household population 15+.

The K-Nearest Neighbour (K-NN) methodology matches the collection of records from surveys to our postal code roster, otherwise referred to as target data. From the collection of survey records, we calculate weighted average coefficients. This means that each postal code geography has a unique set of coefficients. The coefficients are rolled out against our DemoStats current year universes, and we use 50 variables to "match" survey records to our postal code roster. These 50 variables come from our DemoStats database and Equifax Neighbourhood View™, and they are appended to postal code-level data from the surveys. Using these 50 variables allows us to calculate multi-dimensional differences between each survey and target postal code. The survey postal codes with the smallest difference compared to the target postal code are used to generate the coefficients for the target postal code. It is worth noting that we restrict matches between survey and target postal codes by geographic region (Atlantic, Quebec, Ontario, Prairies, B.C./Territories).

Fusion is a technique similar to K-NN. We use this technique as a pre-processing step for K-NN when working with public-use micro-files (PUMFs). PUMF data comes with few, if any, geographic indicators. The Fusion technique is used to match PUMF records to another dataset, otherwise known as the source dataset. The source dataset will have postal codes as a geographic identifier. We generate these matches based on variables that are available in the PUMF data and our source dataset. This approach allows us to append postal codes to PUMF records. Note: the

appended postal code is one of many possible postal codes that can be appended. Once we have postal codes appended to the PUMF data, we can proceed with the K-NN methodology as described above.

Modelling Framework Summary

Coefficient

- Modeled to the PRIZM level
- Coefficients equal empirical propensity by segment
- PRIZM coefficients are rolled out to geography
- All geographies with the same PRIZM segment assignment have the same propensity

K-NN and/or Fusion (Modelled)

- Data are modelled directly to the postal code level
- The models are constrained by region
- In theory, each postal code obtains a unique set of propensities